

A Strong Converse Theorem for Quantum Multiple Access Channels

R. Ahlswede and N. Cai

Abstract. With the wringing technique developed by the first author for classical multiple access channels we show that the strong converse theorem holds also for quantum multiple access channels, if classical messages are transmitted.

Keywords: classical quantum multiple access channel; strong converse; wringing technique; non-stationary memoryless classical quantum channel.

1 Introduction

The coding theorem for memoryless channels, the most fundamental theorem in Information Theory, was presented by C.E. Shannon with a sketch of its proof in his celebrated work [23] of 1948. The first formally satisfactory proof of the coding theorem appeared in 1954 in [12] by A. Feinstein, who attributed the part called weak converse of the theorem to R.M. Fano [11] (in 1952). The strong converse for memoryless channels was first proved by J. Wolfowitz [29] in 1957. A.S. Holevo [15] and B. Schumacher–M. Westmoreland [22] extended the coding theorem to quantum memoryless channels for transmission of classical messages. The theorem is known as HSW Theorem in Quantum Information Theory — named after the authors — and is one of most important results in Quantum Information Theory. With the *concept of typical subspaces* A. Winter [27] successfully extended Wolfowitz’s method to a quantum version and proved the strong converse for quantum memoryless channels. He also presented an elegant new proof for the direct part of the coding theorem. At the same time a proof of the strong converse was given by T. Ogawa and H. Nagaoka [18] and their method of proof may be regarded as an extension of that by S. Arimoto [6]. The coding theorem, direct part and strong converse, for non-stationary memoryless classical quantum channels was proved by A. Winter [28] who extended Ahlswede’s method of [1] to the quantum case.

A significant difference between Shannon’s Coding Theorem and the HSW Coding Theorem is that in the terminology of classical Information Theory the former is a single-letter characterization but the latter is not. Hilbert spaces of unbounded dimensions are involved in the capacity formula in the HSW Theorem. In fact it is one of the most important and challenging problems in Quantum Information Theory to get a computable characterization. A family of quantum memoryless channels, known as classical quantum channels, whose HSW Theorem has single-letter characterization, was introduced by A.S. Holevo [14].

An observation by A.S. Holevo [16] shows that the HWS Theorem for general quantum channels can be easily derived from the HWS Theorem for a classical quantum channel by regarding those channels as a classical quantum channel with a “super alphabet”. So it seems to be natural to consider classical quantum multi–user channel when we try to extend the Classical Multi–user Information Theory to Quantum before a single–letter form of the HWS Theorem is obtained.

C.E. Shannon, the founder of Information Theory, also started Multi–user Information Theory in [24]. The only multi–user channel whose single–letter capacity region is completely known is the multiple access channel (MAC). The Coding Theorem for MAC, the direct part and weak converse, was proved by R. Ahlswede [2]. Its strong converse theorem was shown by G. Dueck [9] (with the Ahlswede, Gacs, Körner Blowing Up Lemma and a wringing technique for mutual information) and R. Ahlswede [3] (with a stronger wringing technique for probability alone). To the best of our knowledge the quantum model of the MAC was first studied by A.E. Allahverdyan and D.B. Saakian [5] and the Coding Theorem for the classical quantum MAC, the direct part and weak converse, was shown by A. Winter [26].

However, till now the strong converse resisted all efforts of proof. Already for the classical MAC the strong converse theorem could not be proved for many years, because average errors are genuinely used and even nowadays there is no way to do this via typical sequences and also not via Arimoto’s method. In the quantum case it gets even worse, because we have no analogue of the Blowing Up Method for operators (instead for words). Consequently Dueck’s approach fails also. Fortunately Ahlswede’s wringing method for probabilities, invented for classification of the role of fundamental methods, works also for the quantum situation. However the Packing Lemma of [3] requires now a more demanding analytical proof than the original application of Chebyshev’s inequality. We present necessary definitions and state formally our results in the next section and provide the plan to prove the main result in Section 3. In Section 4 we show a strong converse for quantum non–stationary memoryless channels (slightly stronger than that in [28]), which we need in the proof to the main theorem. Finally the proof of the main theorem is completed in Section 5.

2 Definitions and Results

Throughout the paper the script letters $\mathcal{X}, \mathcal{Y}, \mathcal{Z}, \dots$ stand for finite sets and $\mathcal{X}^n, \mathcal{Y}^n, \mathcal{Z}^n, \dots$ are their Cartesian powers. We denote their elements and the random variables taking values in them, by corresponding lower letters, $x, y, z, \dots, x^n, y^n, z^n, \dots$ and capital letters $X, X', Y, \hat{Y}, Z, \dots, Z^n, Y^n, \dots$ respectively. The probability distribution of random variable X , the conditional distribution of random variables X given Y, \dots are written as $P_X, P_{X|Y}, \dots$. When we speak of P_X , we mean the underlying random variable X is automatically defined and similarly for $P_{X|Y}$. As in the standard way (e.g., [17], $|i\rangle, |j\rangle, |\alpha\rangle, |\beta\rangle, \dots$ (the “ket”) stand for normalized column vectors in Hilbert spaces and their dual vectors are written as $\langle i|, \langle j|, \langle \alpha|, \langle \beta|, \dots$ (the “bra”). The density operators or states are denoted

by Greek letters e.g. $\rho, \sigma, \varphi, \dots$. We write von Neumann entropy of state ρ and the quantum relative entropy of ρ to σ as $S(\rho)$ and $D(\rho\|\sigma)$, respectively, i.e., $S(\rho) = -th(\rho \log \rho)$ and $D(\rho\|\sigma) = tr(\rho \log \rho) - tr(\rho \log \sigma)$. Here and throughout the paper \log is the logarithm with base 2 whereas \ln is the logarithm with base e . Also throughout the paper we assume all Hilbert spaces in the discussion are over the complex field and have finite dimensions.

One way to define classical quantum channels is to specify them by a set of states $\Phi = \{\varphi(x) : x \in \mathcal{X}\}$ in the same Hilbert space $\mathcal{H}\mathcal{H}$, labelled by a finite set \mathcal{X} . We call Φ a classical quantum channel, or shortly a $c - q$ channel, with input alphabet \mathcal{X} and output space $\mathcal{H}\mathcal{H}$. $x \in \mathcal{X}$ and $\varphi(x)$ are called input letter and output state, respectively. We assume that the receiver, or decoder, of the channel receives the state $\varphi(x)$ from the output of the channel, if the sender, or encoder, of the channel inputs a letter $x \in \mathcal{X}$ to the channel. The Holevo quantity of a $c - q$ channel $\Phi = \{\varphi(x) : x \in \mathcal{X}\}$ with respect to the input distribution P is denoted by $\chi(P; \Phi)$

$$\begin{aligned} \chi(P; \Phi) &= S(\sigma) - \sum_{x \in \mathcal{X}} P(x)S(\varphi(x)) \\ &= \sum_{x \in \mathcal{X}} P(x) [-tr(\varphi(x) \log \sigma) + tr(\varphi(x) \log \varphi(x))] \\ &= \sum_{x \in \mathcal{X}} P(x)D(\varphi(x)\|\sigma), \end{aligned} \tag{2.1}$$

where $\sigma = \sum_{x \in \mathcal{X}} P(x)\varphi(x)$. (Perhaps it should be called ... mutual information.)

A non-stationary memoryless classical quantum channel, or shorter non-stationary memoryless $c - q$ channel, is specified by a sequence of $c - q$ channels $\{\Phi_n\}_{n=1}^\infty$ with common input alphabet \mathcal{X} and output space $\mathcal{H}\mathcal{H}$ such that an output state

$$\varphi^{\otimes n}(x^n) = \varphi_1(x_1) \otimes \varphi_2(x_2) \otimes \dots \otimes \varphi_n(x_n)$$

is output from the channel if the sequence $x^n = (x_1, \dots, x_n) \in \mathcal{X}^n$ is input to the channel, where $\Phi_t = \{\varphi_t(x) : x \in \mathcal{X}\}$ for $t = 1, 2, \dots$, and \otimes is the tensor product. An $(n, M, \bar{\lambda})$ -code for the non-stationary memoryless $c - q$ channel, $(\mathcal{U}, \mathcal{D})$ consists of a subset $\mathcal{U} \subset \mathcal{X}^n$ and a measurement \mathcal{D} on the outputspace $\mathcal{H}\mathcal{H}^{\otimes n}$ described by $\{\mathcal{D}_{u^n} : u^n \in \mathcal{U}\}$, where \mathcal{D}_{u^n} corresponds to input u^n , such that $|\mathcal{U}| = M$ and

$$M^{-1} \sum_{u^n \in \mathcal{U}} tr[\varphi^{\otimes n}(u^n)\mathcal{D}_{u^n}] > 1 - \bar{\lambda}. \tag{2.2}$$

Here \mathcal{U} is called codebook, members of \mathcal{U} are called codewords, n is called the length of the code, or of the codewords, and $\bar{\lambda}$ is called the average probability of error. $\frac{1}{n} \log M$ is called rate of the code. An $(n, M, \bar{\lambda})$ -code is used to transmit a set of classical messages of size M as follows. In the case the encoder has to send the m th message in the set, he sends the m th codeword, say $u^n(m) \in \mathcal{U}$ through the non-stationary memoryless $c - q$ channel. To decode the message, the decoder

performs the measurement \mathcal{D} to the outputstate and he decodes the output to the m' th message of the set if the outcome of measurement \mathcal{D} is $u^n(m')$. Then the average fidelity of transmission is bounded from below by (2.2). A positive real number R is said to be $\bar{\lambda}$ -achievable for a $c-q$ channel if for all $\varepsilon > 0$ there exists an $(n, M, \bar{\lambda})$ -code for the channel with rate $\frac{1}{n} \log M > R - \varepsilon$ if n is sufficiently large. The maximum $\bar{\lambda}$ -achievable rate for a channel is called its $\bar{\lambda}$ -capacity, denoted by $C(\bar{\lambda})$. The capacity of the channel is defined as $C = \inf_{\bar{\lambda} > 0} C(\bar{\lambda})$.

Instead of capacity for average probability of error, we have capacity for maximum probability of error, if we replace (2.2) by

$$\text{tr}[\varphi^{\otimes n}(u)\mathcal{D}_{u^n}] > 1 - \lambda \text{ for all } u^n \in \mathcal{U}. \tag{2.3}$$

However by the pigeon-hole principle we know that the coding problems for average probability and maximum probability of error are equivalent for two terminals channels.

We call a non-stationary memoryless $c-q$ channel specified by $\{\Phi_n\}_{n=1}^\infty$ stationary, or a stationary memoryless $c-q$ channel, if for all $n \Phi_n = \Phi_1$. The coding theorem for stationary memoryless $c-q$ channels is an important special case of the HSW Theorem. Coding theorem with strong converse for non-stationary memoryless $c-q$ channels were shown by A. Winter in [28] and his strong converse has the following form.

Theorem W. (Winter [28]) *Given a non-stationary memoryless $c-q$ channel $\{\Phi_n\}_{n=1}^\infty$, for all $0 < \lambda < 1$ and $\varepsilon > 0$, there is an $n_o = n_o(\lambda, \varepsilon)$ (independent of the channel) such that if there exists an (n, M, λ) -code for the channel with $n > n_o$, then*

$$\frac{1}{n} \log M \leq \frac{1}{n} \sum_{t=1}^n \max_{P_t} \mathcal{X}(P_t; \Phi_t) + \varepsilon.$$

To prove our main result, we need a slightly stronger version of the strong converse for a non-stationary memoryless $c-q$ channel.

Theorem 1. *Let $\{\Phi_n\}_{n=1}^\infty$ be a non-stationary memoryless $c-q$ channel such that for diagonalizations of $\varphi_n(x)$, $n = 1, 2, \dots$, $x \in \mathcal{X}$,*

$$\varphi_n(x) = \sum_{j=1}^d W_n(j|x) |j_n(x)\rangle \langle j_n(x)| \tag{2.4}$$

where $d = \dim \mathcal{H}\mathcal{H}$ and $\mathcal{H}\mathcal{H}$ is the output space,

$$\inf_{n,x,j}^+ W_n(j|x) = w > 0, \tag{2.5}$$

where the \inf^+ is infimum of $W_n(j|x)$ for $1, 2, \dots$, $j = 1, 2, \dots, d$ and $x \in \mathcal{X}$ with $W_n(j|x) > 0$. Then there exists a function h defined on $(0, 1)$ and depending on w in (2.5) such that for all (n, M, λ) -codes for the channel $\{\Phi_n\}_{n=1}^\infty$ and sufficiently large n ,

$$\frac{1}{n} \log M \leq \sum_{t=1}^n \frac{1}{n} \chi(P_{X_t}; \Phi_t) + \frac{1}{\sqrt{n}} h(\lambda), \tag{2.6}$$

where $X^n = (X_1, X_2, \dots, X_n)$ is the random variable uniformly distributed over the codebook \mathcal{U} . In other words P_{X^n} is the empirical distribution over the codebook and P_{X_t} is the marginal distribution of its t -th component.

We shall explain the role of the theorem in the proof of the main result in the next section and we prove it in Section 4.

Remarks

1. As one can always use a finite set of $c - q$ channels to approach an infinite set of $c - q$ channels by a quantum version (see Section VII in [28]) of ε -net techniques in Classical Information Theory and the condition (2.5) automatically holds, in the case that $\Phi_n, n = 1, 2, \dots$ are taken from a finite set of $c - q$ channels, the condition (2.5) can be removed from Theorem 1. Consequently one may replace h by a function independent of w . This can be done in exactly the same way as in [28]. However the current form of the theorem is sufficient for our purposes.
2. By considering the previous remark, Theorem W follows from Theorem 1.

Now let us turn to our main object, (stationary) memoryless classical quantum multiple access channels, or for short $c - q$ MAC. In general, like for classical MAC, a $c - q$ MAC has more than one sender (or encoder), for simplicity in the current paper, we consider $c - q$ MAC with two encoders. Thus a $c - q$ MAC is specified by a set $\Phi = \{\varphi(x, y) : x \in \mathcal{X}, y \in \mathcal{Y}\}$ of states in common Hilbert space $\mathcal{H}\mathcal{H}$ labelled by the elements in a Cartesian product $\mathcal{X} \times \mathcal{Y}$ of two finite sets \mathcal{X} and \mathcal{Y} . Again $\mathcal{H}\mathcal{H}$ is called output space and \mathcal{X} and \mathcal{Y} are called input alphabets. We address the two encoders as \mathcal{X} - and \mathcal{Y} -encoders. The decoder receives the state

$$\varphi^{\otimes n}(x^n, y^n) = \varphi(x_1, y_1) \otimes \varphi(x_2, y_2) \otimes \dots \otimes \varphi(x_n, y_n) \tag{2.7}$$

if the \mathcal{X} -encoder and the \mathcal{Y} -encoder send input sequences $x^n = (x_1, x_2, \dots, x_n) \in \mathcal{X}^n$ and $y^n = (y_1, y_2, \dots, y_n) \in \mathcal{Y}^n$, respectively, over the $c - q$ MAC. An $(n, M, N, \bar{\lambda})$ -code $(\mathcal{U}, \mathcal{V}, \mathcal{D})$ for a $c - q$ MAC Φ consists of a subset $\mathcal{U} \subset \mathcal{X}^n$, a subset $\mathcal{V} \subset \mathcal{Y}^n$, with cardinalities $|\mathcal{U}| = M, |\mathcal{V}| = N$ respectively and a measurement $\mathcal{D} = \{\mathcal{D}_{u^n, v^n} : u^n \in \mathcal{U}, v^n \in \mathcal{V}\}$, where \mathcal{D}_{u^n, v^n} corresponds an outcoming (u^n, v^n) , such that

$$M^{-1}N^{-1} \sum_{u^n \in \mathcal{U}} \sum_{v^n \in \mathcal{V}} \text{tr}[\varphi^{\otimes n}(u^n, v^n)\mathcal{D}_{u^n, v^n}] > 1 - \bar{\lambda}. \tag{2.8}$$

We call \mathcal{U} and \mathcal{V} \mathcal{X} - and \mathcal{Y} -codebooks and their members codewords. $\bar{\lambda}$ and n are called average probability of error and length of code (or of codewords), respectively.

Define $R_1 = \frac{1}{n} \log |M|$ and $R_2 = \frac{1}{n} \log |N|$. (R_1, R_2) is called pair of rates of the code. A pair (R_1, R_2) of positive real numbers is called $\bar{\lambda}$ -achievable if for

all $\varepsilon > 0$ and sufficiently larger n there exists an $(n, M, N, \bar{\lambda})$ -code for the $c - q$ MAC with rates $\frac{1}{n}M > \mathcal{R}_1 - \varepsilon$ and $\frac{1}{n} \log N > \mathcal{R}_2 - \varepsilon$. The set of achievable pairs of real numbers is called $\bar{\lambda}$ -capacity region of the $c - q$ MAC and denoted by $\mathcal{R}(\bar{\lambda})$. Then the capacity region of the $c - q$ MAC Φ is defined as

$$\mathcal{R} = \bigcap_{1 > \bar{\lambda} > 0} \mathcal{R}(\bar{\lambda}). \tag{2.9}$$

As a special case of a $c - q$ MAC $\Phi = \{\varphi(x, y) : x \in \mathcal{X}, y \in \mathcal{Y}\}$ becomes a classical MAC, when $\varphi(x, y)$ $x \in \mathcal{X}, y \in \mathcal{Y}$ can be simultaneously diagonalized by the same basis. One may define the capacity region (λ -capacity region) for maximum probability of error in an analogous way by replacing (2.8) by

$$\text{tr}[\varphi(u^n, v^n)\mathcal{D}_{u^n, v^n}] > 1 - \lambda$$

for all $u^n \in \mathcal{U}, v^n \in \mathcal{V}$.

However, it turned out that unlike for two terminal channels capacity regions for maximum probability of error and average probability of error of the same $c - q$ MAC may be different even for the special case of a classical MAC (c.f. [8]).

Throughout the paper for $c - q$ MAC we only consider the average probability of error and to present the formulas of capacity region of a $c - q$ MAC for it we introduce the following notation.

For a given $c - q$ MAC Φ with input alphabets \mathcal{X} and \mathcal{Y} , and a pair of random variables (X, Y) distributed on $\mathcal{X} \times \mathcal{Y}$ we write

$$\chi(X, Y; \Phi) = \chi(P_{XY}; \Phi), \tag{2.10}$$

where $\chi(\cdot; \cdot)$ is the Holevo quantity defined in (2.1),

$$\begin{aligned} \chi(X : \Phi|Y) &= \sum_{y \in \mathcal{Y}} P_Y(y) \chi(P_{X|Y}(\cdot|y); \Phi(\cdot, y)) \\ &= \sum_{y \in \mathcal{Y}} P_Y(y) \left[S(\sigma_{XY,2}(y)) - \sum_{x \in \mathcal{X}} P_{X|Y}(x|y) S(\varphi(x, y)) \right], \end{aligned} \tag{2.11}$$

where $\Phi(\cdot, y) = \{\varphi(x, y) : x \in \mathcal{X}\}$ for all $y \in \mathcal{Y}$ is a $c - q$ channel with input alphabet \mathcal{X} and $\sigma_{XY,2} = \sum_x P_{X|Y}(x, y) \varphi(x, y)$ for all $y \in \mathcal{Y}$, and

$$\begin{aligned} \chi(Y; \Phi|X) &= \sum_{x \in \mathcal{X}} P_X(x) \chi(P_{Y|X}(\cdot|x); \Phi(x, \cdot)) \\ &= \sum_{x \in \mathcal{X}} P_X(x) \left[S(\sigma_{XY,1}(x)) - \sum_{y \in \mathcal{Y}} P_{Y|X}(y|x) S(\varphi(x, y)) \right] \end{aligned} \tag{2.12}$$

where $\Phi(x, \cdot) = \{\varphi(x, y) : y \in \mathcal{Y}\}$ for all x is a $c - q$ channel with input alphabet \mathcal{Y} and $\sigma_{XY,1}(x) = \sum_y P_{Y|X}(y|x) \varphi(x, y)$.

Then it immediately follows from (2.1), (2.10) – (2.12) that

$$\chi(X, Y; \Phi) = \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} P_{XY}(x, y) D(\varphi(x, y) \| \sigma_{XY}), \tag{2.13}$$

where $\sigma_{XY} = \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} P_{XY}(x, y) \varphi(x, y)$,

$$\chi(X; \Phi|Y) = \sum_{y \in \mathcal{Y}} P_Y(y) \sum_{x \in \mathcal{X}} P_{X|Y}(x|y) D(\varphi(x, y) \| \sigma_{XY,2}(y)), \tag{2.14}$$

and

$$\chi(Y; \Phi|X) = \sum_{x \in \mathcal{X}} P_X(x) \sum_{y \in \mathcal{Y}} P_{Y|X}(y|x) D(\varphi(x, y) \| \sigma_{XY,1}(x)). \tag{2.15}$$

For a pair of random variables (X, Y) with values in $\mathcal{X} \times \mathcal{Y}$ we let $\mathcal{R}(X, Y)$ be the set of pairs of positive real numbers (R_1, R_2) satisfying

$$R_1 + R_2 \leq \chi(X, Y; \Phi), \tag{2.16}$$

$$R_1 \leq \chi(X; \Phi|Y), \tag{2.17}$$

and

$$R_2 \leq \chi(Y; \Phi|X). \tag{2.18}$$

Denote by $\partial(\mathcal{X}) \times \partial(\mathcal{Y})$ the set of pairs of independent random variables X taking values on \mathcal{X} and Y taking values on \mathcal{Y} (i.e. $P_{XY}(x, y) = P_X(x)P_Y(y)$) and

$$\mathcal{R}^* = \text{conv} \left[\bigcup_{(X,Y) \in \partial(\mathcal{X}) \times \partial(\mathcal{Y})} \mathcal{R}(X, Y) \right], \tag{2.19}$$

where $\text{conv}(A)$ stand for the convex closure of the set A .

Then the Coding Theorem for the $c - q$ MAC of A. Winter [26] says that

$$\mathcal{R} = \mathcal{R}^*. \tag{2.20}$$

The main contribution of this paper is the strong converse for all $c - q$ MAC.

Theorem 2. *Given a $c - q$ M Φ , there exists a function \tilde{h} defined on $(0, 1)$ such that for all $(n, M, N, \bar{\lambda})$ -codes with $\bar{\lambda} \in (0, 1)$, there exist random variables $\tilde{X}_n = (\tilde{X}_1, \dots, \tilde{X}_n)$ and $\tilde{Y}_n = (\tilde{Y}_1, \dots, \tilde{Y}_n)$ taking values in \mathcal{X}^n and \mathcal{Y}^n , respectively, with*

$$\frac{1}{n} \log M + \frac{1}{n} \log N \leq \frac{1}{n} \sum_{t=1}^n \chi(\tilde{X}_t, \tilde{Y}_t; \Phi) + \frac{\log n}{\sqrt{n}} \tilde{h}(\bar{\lambda}), \tag{2.21}$$

$$\frac{1}{n} \log M \leq \frac{1}{n} \sum_{t=1}^n \chi(\tilde{X}_t; \Phi|\tilde{Y}_t) + \frac{\log n}{\sqrt{n}} \tilde{h}(\bar{\lambda}), \tag{2.22}$$

and

$$\frac{1}{n} \log N \leq \frac{1}{n} \sum_{t=1}^n \chi(\tilde{Y}_t; \Phi | \tilde{X}_t) + \frac{\log n}{\sqrt{n}} \tilde{h}(\bar{\lambda}), \tag{2.23}$$

and for all t , \tilde{X}_t and \tilde{Y}_t are independent.

But the dependence is not necessary. It is a consequence of the dependence of h on w in (2.5) in Theorem 1. So by Remark 1 and on inspection of the proof for Theorem 2, one may obtain an \tilde{h}^* independent of the channel.

3 Outline of Ideas to Extend Ahlswede’s Strong Converse Proof for Classical MAC to $c - q$ MAC

Roughly speaking, his proof is divided into three steps.

In the first step Ahlswede extended Augustin’s strong converse theorem in [7], which says that for an arbitrary (n, M, λ) code for a (classical) non-stationary memoryless channel $\{W_n\}_{n=1}^\infty$

$$\log M \leq \sum_{t=1}^n I(X_t; Y_t) + \frac{3}{1-\lambda} |\mathcal{X}| \sqrt{n},$$

where $X^n = (X_1, \dots, X_n)$ is random variable uniformly distributed on the codebook. $Y^n = (Y_1, \dots, Y_n)$ is the output random variable for input X^n , and I is Shannon’s mutual information.

By applying the strong converse in the first step to the classical MAC, one may obtain an outer bound for the achievable rate pair (R_1, R_2) such that

$$\begin{aligned} R_1 + R_2 &\lesssim \frac{1}{n} \sum_{t=1}^n I(X_t Y_t; Z_t) \\ R_1 &\lesssim \frac{1}{n} \sum_{t=1}^n I(X_t; Z_t | Y_t) \end{aligned}$$

and

$$R_2 \leq \frac{1}{n} \sum_{t=1}^n I(Y_t; Z_t | X_t),$$

formally like the capacity region. But now (X_t, Y_t) may not be independent! So the outer bound is not tight.

The reason is due to application of the strong converse of non-stationary channels in step 1 requiring the maximum error criterion, whereas for MAC the capacity regions of maximum and average error criteria may be different [8]. By the Markov inequality one may obtain a subcode \mathcal{A} from the original code for MAC. But the uniform distributions over \mathcal{A} are not independent. To solve the problem R. Ahlswede discovered the following technique for the second step.

Lemma A. (Wringing Technique [3]) *Let P and Q be probability distributions over a Cartesian power Z^n of a finite set Z such that for a positive constant c*

$$P(z^n) \leq (1 + c)Q(z^n) \text{ for all } z^n \in \mathcal{Z}^n,$$

then for any $0 < \gamma < c$, $0 \leq \varepsilon < 1$ there exist $t_1, \dots, t_k \in \{1, 2, \dots, n\}$, where $0 \leq k \leq \frac{c}{\gamma}$ such that for some $\bar{z}_{t_1}, \dots, \bar{z}_{t_k}$,

$$P(z_t | \bar{z}_{t_1}, \dots, \bar{z}_{t_k}) \leq \max[(1 + \gamma)Q(z_t | \bar{z}_{t_1}, \dots, \bar{z}_{t_k}), \varepsilon]$$

and

$$P(\bar{z}_{t_1}, \dots, \bar{z}_{t_k}) \geq \varepsilon^k.$$

With the lemma Ahlswede obtained a further subcode \mathcal{B} from the subcode \mathcal{A} (without loosing too much rate) such that the uniform distribution over the codebook for it is nearly componentwise independent. Notice that the uniform distribution for \mathcal{B} is not necessarily independent, but componentwise independence is sufficient for the purpose.

In the third step Ahlswede first combined the results in the first two steps to obtain an outer bound in terms of nearly componentwise independent input distributions and then by some calculation based on the continuity of Shannon information quantities showed the outer bound is arbitrarily close to the capacity region of the classical MAC.

We now plan to finish our proof for Theorem 2 paralleling these three steps. We first inspect the second step, the wringing technique, and find that we can take it into our proof almost without doing additional work, because the wringing technique is only applied in the input space and by definition our input space is classical. That means the only thing, which we need to do, is replace the codebook for the classical MAC by one for the $c - q$ MAC, and then consider the consequence at the output.

As by Fannes inequality [10] von Neumann entropy is continuous the extension to the quantum version in the third step is not so difficult. The only difficult part for the extension is the first part. Winter’s strong converse (Theorem W) in [28] for non-stationary $c - q$ channels is in terms of optimal input distribution whereas the strong converse, which we need, is in terms of empirical distributions over codebooks. Hayashi and Nagaoka [13] extended a general capacity formula due to Verdú and Han [25] to classical quantum channels. But it turned out that they obtained a “ratewise” strong converse for stationary $c - q$ channels in terms of optimal input distributions by applying their formula.

So to prove our strong converse for $c - q$ MAC, Theorem 2, we have to show a strong converse for non-stationary $c - q$ channels in terms of the uniform distribution over the codebook i.e. our Theorem 1. In summary we state our plan as follows.

In the first step we prove Theorem 1 as an auxiliary result.

In the second step we apply the wringing technique to our codebook to obtain a subcode whose uniform distribution $P_{\tilde{X}^n \tilde{Y}}$ is nearly componentwise independent (i.e. $P_{\tilde{X}_t \tilde{Y}_t}(x, y)$ is arbitrarily close to $P_{\tilde{X}_t}(x)P_{\tilde{Y}_t}(y)$ for all $x \in \mathcal{X}$, $y \in \mathcal{Y}$).

Finally we finish our proof in the third step by calculation based on continuity of von Neumann entropy.

4 A Strong Converse for Classical Quantum Non-stationary Channels

We begin with our proof to Theorem 1 with a lemma which can be considered as a quantum analogue of Lemma 1 in [3].

Let \mathcal{U} be a finite index set and let

$$\{\rho(u) : u \in \mathcal{U}\} \tag{4.1}$$

be a set of states labelled by indices in \mathcal{U} in a given complex Hilbert space $\mathcal{H}\mathcal{H}$ of finite dimension d . Further let σ be a state in $\mathcal{H}\mathcal{H}$. For a given real number $r(u)$ we denote the projector of $\mathcal{H}\mathcal{H}$ onto subspace

$$\rho(u) - 2^{r(u)}\sigma \geq 0 \tag{4.2}$$

by $\mathcal{P}^+(u, \sigma, r(u))$. That is, $\mathcal{P}^+(u, \sigma, r)$ is the projector with

$$\mathcal{P}^+(u, \sigma, r(u)) = \sum_{\alpha_j(u) \geq 0} |e_j(u)\rangle\langle e_j(u)| \tag{4.3}$$

if $\rho(u) - 2^{r(u)}\sigma$ is diagonalized as

$$\rho(u) - 2^{r(u)}\sigma = \sum_{j=0}^{d-1} \alpha_j(u) |e_j(u)\rangle\langle e_j(u)|. \tag{4.4}$$

Then we have

Lemma 1. *Let \mathcal{U} , $\{\rho(u) : u \in \mathcal{U}\}$, σ , and $\mathcal{P}^+(u, \sigma, r(u))$ be defined as above and let $\{\mathcal{D}_u : u \in \mathcal{U}\}$ be a measurement in $\mathcal{H}\mathcal{H}$ (i.e., $0 \leq \mathcal{D}_u \leq I$ for all $u \in \mathcal{U}$, $\sum_{u \in \mathcal{D}_u} \mathcal{D}_u = I$, and \mathcal{D}_u corresponds to the outcome “ u ”, where I is the identity operator in $\mathcal{H}\mathcal{H}$) such that for a positive real number δ*

$$\text{tr}\rho(u)\mathcal{D}_u - \text{tr}\rho(u)\mathcal{P}^+(u, \sigma, r(u)) \geq \delta \tag{4.5}$$

for all $u \in \mathcal{U}$. Then

$$|\mathcal{U}| \leq \delta^{-1} 2^{|\mathcal{U}|^{-1} \sum_{u \in \mathcal{U}} r(u)}. \tag{4.6}$$

Proof: For all $u \in \mathcal{U}$, by (4.3), (4.4) and (4.5),

$$\begin{aligned} 2^{r(u)}\text{tr}\sigma\mathcal{D}_u &= \text{tr}\rho(u)\mathcal{D}_u - \text{tr}[\rho(u) - 2^{r(u)}\sigma]\mathcal{D}_u \\ &\geq \text{tr}\rho(u)\mathcal{D}_u - \text{tr}[\rho(u) - 2^{r(u)}\sigma]\mathcal{P}^+(u, \sigma, r(u)) \\ &\geq \text{tr}\rho(u)\mathcal{D}_u - \text{tr}\rho(u)\mathcal{P}^+(u, \sigma, r(u)) \geq \delta, \end{aligned} \tag{4.7}$$

where the first inequality follows from (4.3), (4.4), and the fact that for any operator π with diagonalization $\pi = \sum_{i=0}^{d-1} p_i |e_i\rangle\langle e_i|$, the projector $\mathcal{P}_\pi^+ = \sum_{p_i \geq 0} |e_i\rangle\langle e_i|$ maximizes $\text{tr}(\pi\tau)$ among the operators τ with $0 \leq \tau \leq I$, and the last inequality holds by condition (4.5).

Next notice that $\delta 2^{-|\mathcal{U}|^{-1} \sum_{u \in \mathcal{U}} r(u)} = \left[\prod_{u \in \mathcal{U}} \delta 2^{-r(u)} \right]^{\frac{1}{|\mathcal{U}|}}$. Then (4.7) is followed by

$$\delta 2^{-|\mathcal{U}|^{-1} \sum_{u \in \mathcal{U}} r(u)} \leq \left[\prod_{u \in \mathcal{U}} tr(\sigma \mathcal{D}_u) \right]^{\frac{1}{|\mathcal{U}|}}. \tag{4.8}$$

However geometric means may not exceed arithmetic means and so we can continue to bound (4.8) by

$$\delta 2^{-|\mathcal{U}|^{-1} \sum_{u \in \mathcal{U}} r(u)} \leq \frac{1}{|\mathcal{U}|} \sum_{u \in \mathcal{U}} tr(\sigma \mathcal{D}_u) = |\mathcal{U}|^{-1} tr \left(\sigma \sum_{u \in \mathcal{U}} \mathcal{D}_u \right) = |\mathcal{U}|^{-1} tr(\sigma I) = |\mathcal{U}|^{-1}.$$

That is (4.6). □

To apply Lemma 1 in the proof of Theorem 1 we have to estimate $[tr \rho \mathcal{P}^+(\rho, \sigma, r)]$ for projector

$$\mathcal{P}^+(\rho, \sigma, r) = \sum_{\alpha_j \geq 0} |e_j \rangle \langle e_j| \tag{4.9}$$

onto the subspace $\rho - 2^r \sigma \geq 0$, where ρ and σ are states and the diagonalization of $\rho - 2^r \sigma$ is

$$\rho - 2^r \sigma = \sum_{j=0}^{d-1} \alpha_j |e_j \rangle \langle e_j|. \tag{4.10}$$

For $j = 0, 1, 2, \dots, d - 1$, let $P(j) = \langle e_j | \rho | e_j \rangle$ and $Q(j) = \langle e_j | \sigma | e_j \rangle$. Then P and Q are probability distributions on $\{0, 1, \dots, d - 1\}$ since ρ and σ are states: Thus by (4.9) and (4.10) one may rewrite

$$tr \rho \mathcal{P}^+(\rho, \sigma, r) = \sum_{j: P(j) - 2^r Q(j) \geq 0} P(j) = \sum_{j: \log \frac{P(j)}{Q(j)} \geq r} P(j). \tag{4.11}$$

A natural way to estimate $[tr \rho \mathcal{P}^+(\rho, \sigma, r)]$ is applying Chebyshev’s inequality to (4.11). But it will lead us to an estimation in terms of Shannon information quantities, which is not what we want to have.

Based on the works [20], [21] by Petz, T. Ogawa and H. Nagaoka provided an estimation in [19].

Lemma ON ([19]). For all $s \in [0, 1)$

$$\log [tr \rho \mathcal{P}^+(\rho, \sigma, r)] \leq 2^{-rs} tr(\rho^{1+s} \sigma^{-s}). \tag{4.12}$$

It can be shown that

$$\frac{1}{s} tr \rho^{1+s} \sigma^{-s} \rightarrow D(\rho || \sigma) \text{ as } s \rightarrow 0, \tag{4.13}$$

which was used by M. Hayashi and H. Nagaoka to obtain a strong converse theorem for stationary memoryless $c - q$ channels in terms of optimal input distributions.

However the convergence in (4.13) is not sufficient for us to obtain a strong converse theorem for non-stationary channel and, roughly saying, we need that

$$\frac{1}{s} \log [\text{tr}(\rho_t(u)^{1+s} \sigma_t^{-s})] \rightarrow D(\rho_t(u) \parallel \sigma_t) \text{ as } s \rightarrow 0 \tag{4.14}$$

uniformly in t for sets $\{\rho_t(u) : u \in \mathcal{U}\}$ and $\sigma_t, t = 1, 2, \dots, n$ as states. Precisely our estimation is done by the following lemma.

Lemma 2. *For a given positive real number w there exists a positive $a = a(w)$ such that for all states ρ whose minimum positive eigenvalue is not smaller than w , all states σ , and all $s \in [0, \frac{1}{2})$*

$$D(\rho \parallel \sigma) s \leq \log [\text{tr} \rho^{1+s} \sigma^{-s}] \leq D(\rho \parallel \sigma) s + d^{\frac{1}{2}} [a + \text{tr}(\rho \sigma^{-1})] s^2, \tag{4.15}$$

where d is the dimension of the Hilbert space $\mathcal{H}H$.

Proof: Let ρ and σ have diagonalizations

$$\rho = \sum_j W(j) |j\rangle \langle j| \tag{4.16}$$

and

$$\sigma = \sum_y Q(y) |y\rangle \langle y|, \tag{4.17}$$

respectively. We may assume that the support of ρ contains the support of σ , because otherwise all terms in (4.15) are infinity and we need to do nothing. So in the following we assume all summations run over positive eigenvalues of σ and do not worry about “zero denominators”. Then by (4.15) and (4.16)

$$\begin{aligned} \text{tr} \rho^{1+s} \sigma^{-s} &= \sum_j \sum_y \text{tr} \left[W(j) \left(\frac{W(j)}{Q(y)} \right)^s |j\rangle \langle j| |y\rangle \langle y| \right] \\ &= \sum_j \sum_y W(j) \left(\frac{W(j)}{Q(y)} \right)^s |j|y|^2. \end{aligned} \tag{4.18}$$

Let $f(s) = \text{tr}(\rho^{1+s} \sigma^{-s})$ and $g(s) = \log f(s)$. Then by Taylor expansion we have for $s \in [0, \frac{1}{2})$ the estimate

$$g(s) = g(0) + g'(0)s + g''(s_0)s^2 \tag{4.19}$$

for some $s_0 \in [0, s]$. Next we have to calculate the derivations of g . By simple calculation we obtain from (4.18)

$$f'(s) = \sum_j \sum_y W(j) |j|y|^2 \left(\frac{W(j)}{Q(y)} \right)^s \ln \frac{W(j)}{Q(y)}, \tag{4.20}$$

and

$$f''(s) = \sum_j \sum_y W(j)|\langle j|y\rangle|^2 \left(\frac{W(j)}{Q(y)}\right)^s \left(\ln \frac{W(j)}{Q(y)}\right)^2. \tag{4.21}$$

Next we substitute $f(s) = \text{tr}(\rho^{1+s}\sigma^{-s})$ and (4.20) into

$$g'(s) = \frac{f'(s)}{f(s)} \cdot \log e, \tag{4.22}$$

and $f(s) = \text{tr}(\rho^{1+s}\sigma^{-s})$, (4.20) and (4.21) into

$$g''(s) = \left[\frac{f''(s)}{f(s)} - \frac{(f'(s))^2}{f^2(s)} \right] \log e, \tag{4.23}$$

respectively.

Then we obtain

$$g'(s) = \frac{\sum_j \sum_y W(j)|\langle j|y\rangle|^2 \left(\frac{W(j)}{Q(y)}\right)^s \log \frac{W(j)}{Q(y)}}{\text{tr}(\rho^{1+s}\sigma^{-s})} \tag{4.24}$$

and

$$g''(s) = \left[\frac{\sum_j \sum_y W(j)|\langle j|y\rangle|^2 \left(\frac{W(j)}{Q(y)}\right)^s \left(\ln \frac{W(j)}{Q(y)}\right)^2}{\text{tr}(\rho^{1+s}\sigma^{-s})} - \left(\frac{\sum_j \sum_y W(j)|\langle j|y\rangle|^2 \left(\frac{W(j)}{Q(y)}\right)^s \ln \frac{W(j)}{Q(y)}}{\text{tr}(\rho^{1+s}\sigma^{-s})} \right)^2 \right] \log e. \tag{4.25}$$

We are ready to see from (4.25) that for all s

$$g''(s) \geq 0. \tag{4.26}$$

Indeed by (4.18), $Q_s = \left\{ Q_s(j, y) = \frac{W(j)|\langle j|y\rangle|^2 \left(\frac{W(j)}{Q(y)}\right)^s}{\text{tr}(\rho^{1+s}\sigma^{-s})} \right\}_{(j,y)}$ is a probability

distribution. So we may define a random variable $Z(s)$ taking value $\frac{W(j)}{Q(y)}$ with probability $Q_s(j, y)$ and rewrite (4.25) as $g''(s) = \text{Var}[\ln Z(s)] \log e$. Moreover by (4.24), (4.16), and (4.17) we have that

$$\begin{aligned} g'(0) &= \sum_j \sum_y W(j)|\langle j|y\rangle|^2 \log W(j) - \sum_j \sum_y W(j)|\langle j|y\rangle|^2 \log Q(y) \\ &= \text{tr}(\rho \log \rho) - \text{tr}(\rho \log \sigma) = D(\rho\|\sigma), \end{aligned} \tag{4.27}$$

Which with (4.19) and (4.26) and $g(0) = 0$ yields the first inequality in (4.15).

To obtain the second inequality in (4.15), we have to choose a at the right hand side of (4.15) according to w . We observe that $\lim_{x \rightarrow \infty} \frac{x^{\frac{1}{2}}(\ell n x^2) \log e}{x} = 0$ and so there exists an $A > 1$ such that for all $s \in [0, \frac{1}{2}]$, $x \in [A, \infty)$

$$x^s (\ell n x)^2 \log e \leq x^{\frac{1}{2}} (\ell n x)^2 \log e \leq x, \tag{4.28}$$

We choose

$$a = \max \left\{ \max_{x \in [w, 1]} (\ell n x)^2 \log e, \max_{x \in [1, A]} x^{\frac{1}{2}} (\ell n x)^2 \log e \right\}$$

and then for all $s \in [0, \frac{1}{2}]$, $x \in [w, A]$,

$$x^s (\ell n x)^2 \log e \leq a. \tag{4.29}$$

By (4.28) and (4.29) we upperbound

$$x^s (\ell n x)^2 \log e \leq x + a \tag{4.30}$$

for $s \in [0, \frac{1}{2}]$ and $x \in [w, \infty)$. Notice that by our assumption that for all j $w \leq W(j)$, for all j, y , $\frac{W(j)}{Q(y)} \in [w, \infty)$.

Thus we may apply the upper bound in (4.30) with $x = \frac{W(j)}{Q(y)}$ to (4.25). This gives us that for all $s \in [0, \frac{1}{2}]$

$$\begin{aligned} g''(s) &\leq \frac{1}{\text{tr}(\rho^{1+s} \sigma^{-s})} \sum_j \sum_y |\langle j|y \rangle|^2 \left(\frac{W(j)}{Q(y)} \right)^s \left(\ell n \frac{W(j)}{Q(y)} \right)^2 \log e \\ &\leq \frac{1}{\text{tr}(\rho^{1+s} \sigma^{-s})} \sum_j \sum_y |\langle j|y \rangle|^2 \left(\frac{W(j)}{Q(y)} + a \right) = \frac{\text{tr}(\rho \sigma^{-1}) + a}{\text{tr}(\rho^{1+s} \sigma^{-s})}. \end{aligned} \tag{4.31}$$

Since for $s > 0$ $\sigma^{-s} \geq I$ and $x^{\frac{3}{2}}$ is convex \cup , for $s \in [0, \frac{1}{2}]$

$$\text{tr}(\rho^{1+s} \sigma^{-s}) \geq \text{tr} \rho^{1+s} = \sum_j W(j)^{1+s} \geq \sum_j W(j)^{\frac{3}{2}} \geq d \left[\sum_j \frac{1}{d} W(j) \right]^{\frac{3}{2}} = d^{-\frac{1}{2}}. \tag{4.32}$$

Finally the second inequality in (4.15) follows from (4.19), (4.27), (4.31), (4.32) and $g(0) = 0$. □

Proof of Theorem 1. To conclude the section we prove Theorem 1. Let $\{\Phi\}_{n=1}^\infty$ be a non-stationary $c - q$ and let $(\mathcal{U}, \mathcal{D})$ be an (n, M, λ) -code for it. Without loss of generality, we assume it is a code for the maximum probability of error because by Markov's inequality one always may obtain an $(n, \lfloor \frac{\lambda - \bar{\lambda}}{\lambda} M \rfloor, \lambda)$ -code for maximum probability of error from an $(n, M, \bar{\lambda})$ -code with average probability of error for all $\lambda \in (\bar{\lambda}, 1)$. So, we have for all $u^n \in \mathcal{U}$ that (2.3) holds.

Let $X^n = (X_1, \dots, X_n)$ be the sequence of random variables uniformly distributed on the codebook \mathcal{U} and for $t = 1, 2, \dots, n$

$$\sigma_t = E\varphi_t(X_t) = \sum_{x \in \mathcal{X}} P_{X_t}(x)\varphi_t(x) \tag{4.33}$$

where $\varphi_t(x) \in \Phi_t = \{\varphi_t(x) : x \in \mathcal{X}\}$. Let

$$\sigma^{\otimes n} = \sigma_1 \otimes \sigma_2 \otimes \dots \otimes \sigma_n. \tag{4.34}$$

To prove the theorem, we should apply Lemma 1 to $\{\varphi^{\otimes n}(u^n) : u^n \in \mathcal{U}\}$ and $\sigma^{\otimes n}$, where

$$\varphi^{\otimes n}(u^n) = \varphi_1(u_1) \otimes \varphi_2(u_2) \otimes \dots \otimes \varphi_1(u_n) \tag{4.35}$$

for $u^n = (u_1, \dots, u_n)$. To this end we set for all $u^n \in (u_1, \dots, u_n) \in \mathcal{U}$

$$r(u^n) = \sum_{t=1}^n \left\{ D(\varphi_t(u_t) \parallel \sigma_t) + \frac{d^{\frac{1}{2}}[a + \text{tr}(\varphi_t(u_t)\sigma_t^{-1})] + \log \frac{2}{1-\lambda}}{\sqrt{n}} \right\}, \tag{4.36}$$

where a is the constant in (4.15), Lemma 2 (defined in the proof). We have to verify (4.5) for $\delta = \frac{1-\lambda}{2}$, which will be done by applying Lemma ON and Lemma 2. By Lemma ON, we have that for all $u^n \in \mathcal{U}$ and $s \in [0, 1)$

$$\begin{aligned} \text{tr}[(\varphi^{\otimes n}(u^n)\mathcal{P}^+(u^n, \sigma^{\otimes n}r(u^n)))] &\leq 2^{-r(u^n)s} \text{tr}([\varphi^{\otimes n}(u^n)]^{1+s} [\sigma^{\otimes n}]^{-s}) \\ &= 2^{-r(u^n)s} \prod_{t=1}^n \text{tr}[\varphi_t^{1+s}(u_t)\sigma_t^{-s}] \\ &= 2^{-r(u^n)s} 2^{\sum_{t=1}^n \log \text{tr}[\varphi_t^{1+s}(u_t)\sigma_t^{-s}]}, \end{aligned} \tag{4.37}$$

where the first equality follows from (4.34) and (4.35). Next we bound

$$\log \text{tr}[\varphi_t^{1+s}(u_t)\sigma_t^{-s}]$$

by the second inequality in (4.15), Lemma 2 and notice that by condition (2.5) of Theorem 1, for all t , and all $u \in \mathcal{X}$, the minimum positive eigenvalue of $\varphi_t(u)$ is not smaller than w .

Then by Lemma 2, we have that for all $s \in [0, \frac{1}{2})$

$$\log [\text{tr}\varphi_t^{1+s}(u_t)\sigma_t^{-s}] \leq D(\varphi_t(u_t) \parallel \sigma_t)s + d^{\frac{1}{2}} [a + \text{tr}(\varphi_t(u_t)\sigma_t^{-1})] s^2, \tag{4.38}$$

By substitution of (4.38) into (4.37), we further bound $\text{tr}(\varphi^{\otimes n}(u^n)\mathcal{P}^+(u^n, \sigma^{\otimes n}, r(u^n)))$ as follows:

$$\text{tr}(\varphi^{\otimes n}(u^n)\mathcal{P}^+(u^n, \sigma^{\otimes n}, r(u^n))) \leq$$

$$\exp_2 \left\{ -r(u^n)s + \sum_{t=1}^n [D(\varphi_t(u_t) \parallel \sigma_t)s + d^{\frac{1}{2}} [a + \text{tr}(\varphi_t(u_t)\sigma_t^{-1})] s^2] \right\} \tag{4.39}$$

for all $s \in (0, \frac{1}{2}]$. We choose $s = \frac{1}{\sqrt{n}}$ in (4.39) and then substitute (4.36) into it.

Thus we obtain

$$\begin{aligned}
 & \text{tr}(\varphi^{\otimes n}(u^n)\mathcal{P}^+(u^n, \sigma^{\otimes n}r(u^n))) \\
 & \leq \exp_2 \left\{ \frac{-1}{\sqrt{n}} \sum_{t=1}^n \left[D(\varphi_t(u_t)\|\sigma_t) + \frac{d^{\frac{1}{2}}[a + \text{tr}(\varphi_t(u_t)\sigma_t^{-1})] + \log \frac{2}{1-\lambda}}{\sqrt{n}} \right] \right\} \\
 & + \frac{1}{\sqrt{n}} \sum_{t=1}^n D(\varphi_t(u_t)\|\sigma_t) + \frac{1}{n} \sum_{t=1}^n d^{\frac{1}{2}} [a + \text{tr}(\varphi_t(u_t)\sigma_t^{-1})] \\
 & = 2^{-\frac{1}{n} \sum_{t=1}^n \log \frac{2}{1-\lambda}} = 2^{-\log \frac{2}{1-\lambda}} = \frac{1-\lambda}{2}, \tag{4.40}
 \end{aligned}$$

which with assumption (2.3) yields that (4.5) holds for $\varphi^n(u^n)$, σ^n , $r(u^n)$, $u^n \in \mathcal{U}$, and $\delta = \frac{1-\lambda}{2}$. That is, the conditions of Lemma 1 are satisfied, and thus (for $\{\rho(u) : u \in \mathcal{U}\} = \{\varphi^{\otimes n}(u^n) : u \in \mathcal{U}\}$, $\sigma = \sigma^{\otimes n}$, $u = u^n$) get

$$\log M \leq -M^{-1} \sum_{u^n \in \mathcal{U}} r(u^n) - \log \frac{1-\lambda}{2}. \tag{4.41}$$

Finally we recall the definition of random variables $X^n = (X_1, \dots, X_n)$, the definition of $r(u^n)$ in (4.36) and treat $r(u^n)$, $D(\varphi_t(u)\|\sigma_t)$ and $\text{tr}(\varphi_t(u)\sigma_t^{-1})$ as functions of u^n and u . Then it follows from (4.41), (4.36), (4.33), and (2.1) that

$$\begin{aligned}
 & \frac{1}{n} \log M \leq \frac{1}{n} \sum_{x^n \in \mathcal{X}^n} Pr(X^n = x^n)r(x^n) - \frac{1}{n} \log \frac{1-\lambda}{2} = \frac{1}{n} Er(X^n) - \frac{1}{n} \log \frac{1-\lambda}{2} \\
 & = \frac{1}{n} \sum_{t=1}^n E[D(\varphi_t(X_t)\|\sigma_t)] + n^{-\frac{3}{2}} \sum_{t=1}^n d^{\frac{1}{2}} [a + E[\text{tr}(\varphi_t(X_t)\sigma_t^{-1})]] \\
 & - \log \frac{1-\lambda}{2} (n^{-1} + n^{-\frac{1}{2}}) \\
 & = \frac{1}{n} \sum_{t=1}^n \sum_{x \in \mathcal{X}} P_{X_t}(x) D(\varphi_t(x)\|\sigma_t) + n^{-\frac{3}{2}} s^{\frac{1}{2}} \sum_{t=1}^n [a + \text{tr}[E(\varphi_t(X_t))\sigma_t^{-1}]] \\
 & - \log \frac{1-\lambda}{2} (n^{-1} + n^{-\frac{1}{2}}) \\
 & = \frac{1}{n} \sum_{t=1}^n \chi(P_{X_t}; \Phi_t) + n^{-\frac{3}{2}} d^{\frac{1}{2}} \sum_{t=1}^n [a + \text{tr}(\sigma_t \cdot \sigma_t^{-1})] \\
 & - \log \frac{1-\lambda}{2} (n^{-1} + n^{-\frac{1}{2}}) \\
 & \leq \frac{1}{n} \sum_{t=1}^n \chi(P_{X_t}; \Phi_t) + n^{-\frac{1}{2}} d^{\frac{1}{2}} (a + d) - 2n^{-\frac{1}{2}} \log \frac{1-\lambda}{2} \\
 & = \sum_{t=1}^n \frac{1}{n} \chi(P_{X_t}; \Phi_t) + \frac{1}{\sqrt{n}} \left[d^{\frac{1}{2}} (a + d) - 2 \log \frac{1-\lambda}{2} \right]. \tag{4.42}
 \end{aligned}$$

Finally we complete our proof by choosing $h(\lambda) = d^{\frac{1}{2}}(a + d) - 2 \log \frac{1-\lambda}{2}$ in (4.42). \square

5 The Proof of the Main Result

In previous section we have shown Theorem 1, which completed the first step of the plan to prove Theorem 2. In this section we shall finish our proof according to the plan in Section 3. The second step is to apply wringing technique Lemma A and it directly follows from the Lemma the

Corollary A. *For given finite sets \mathcal{X} and \mathcal{Y} , $\mathcal{U} \subset \mathcal{X}^n$ and $\mathcal{V} \subset \mathcal{Y}^n$ and a subset $\mathcal{A} \subset \mathcal{U} \times \mathcal{V}$ with cardinality $|\mathcal{A}| \geq \beta|\mathcal{U}| \times |\mathcal{V}|$ for a $\beta \in (0, 1)$, a $\gamma \in (0, \beta^{-1} - 1)$, and $\varepsilon > 0$ there exists $t_1, t_2, \dots, t_k \in \{1, 2, \dots, n\}$ and $(x_{t_1}, y_{t_2}), \dots, (x_{t_k}, y_{t_k})$ for a $k \leq \frac{\beta^{-1}-1}{\gamma}$ such that the section of \mathcal{A} at $(x_{t_1}, y_{t_1}), (x_{t_2}, y_{t_2}), \dots, (x_{t_k}, y_{t_k})$,*

$$\mathcal{B} = \{(u^n, v^n) \in \mathcal{A} : (u_{t_i}, v_{t_i}) = (x_{t_i}, y_{t_i}), i = 1, 2, \dots, k\}$$

has the following properties.

$$|\mathcal{B}| \geq \varepsilon^k |\mathcal{A}|, \tag{5.1}$$

and for the pair (\bar{X}^n, \bar{Y}^n) of sequences of random variables uniformly distributed on \mathcal{B} ,

$$\begin{aligned} & (1 + \gamma)Pr(\bar{X}_t = x)Pr(\bar{Y}_t = y) - \gamma - |\mathcal{X}||\mathcal{Y}|\varepsilon \\ & \leq Pr(\bar{X}_t = x, \bar{Y}_t = y) \leq \max((1 + \gamma)Pr(\bar{X}_t = x)Pr(\bar{Y}_t = y), \varepsilon) \end{aligned} \tag{5.2}$$

for all $x \in \mathcal{X}$, $y \in \mathcal{Y}$ and $t = 1, 2, \dots, n$.

The corollary is actually Corollary 2 in [3]. (The only difference is that we now remove the assumption in [3] that $\mathcal{U} \times \mathcal{V}$ is a codebook for the MAC but this is obviously not an essential assumption.) (5.1) and the first inequality in (5.2) are a simple consequence of Lemma 1 with the choice of P as uniform distribution on \mathcal{A} and Q as uniform distribution of $\mathcal{U} \times \mathcal{V}$. The second inequality in (5.2) follows the first inequality in (5.2) and the fact that $\sum_{(x,y) \in \mathcal{X} \times \mathcal{Y}} Pr(\bar{X}_t = x)Pr(\bar{Y}_t = y) = 1$.

So we omit the details. Readers can make them by themselves or read [3].

Now let Φ be a $c - q$ MAC. We have to find an \tilde{h} such that for all $(n, M, N, \bar{\lambda})$ -codes and the channel (2.23) - (2.25) hold. Suppose that we are given an $(n, M, N, \bar{\lambda})$ -code $(\mathcal{U}, \mathcal{V}, \mathcal{D})$ for the $c - q$ MAC Φ . Let $\lambda = \frac{1+\bar{\lambda}}{2}$ and $\mathcal{A} = \{(u^n, v^n) : tr[\varphi(u^n, v^n)\mathcal{D}_{u^n, v^n}] > 1 - \lambda, u^n \in \mathcal{U}, v^n \in \mathcal{V}\}$. Then we obtain that

$$|\mathcal{A}| > \left(1 - \frac{\bar{\lambda}}{\lambda}\right) MN = \frac{1 - \bar{\lambda}}{1 + \bar{\lambda}} MN \tag{5.3}$$

by applying Shannon's well known approach to (2.8). By definition for all $(u^n, v^n) \in \mathcal{A}$

$$tr[\varphi(u^n, v^n)\mathcal{D}_{u^n, v^n}] > 1 - \lambda = \frac{1 - \bar{\lambda}}{2}. \tag{5.4}$$

Let $\beta = \frac{1-\bar{\lambda}}{1+\bar{\lambda}}$, $\gamma = \frac{1}{\sqrt{n}}$, and $\varepsilon = \frac{1}{\sqrt{n}}$. Then corollary A is followed by that there exists a $\mathcal{B} \subset \mathcal{A}$ in Corollary A such that for a $k \leq \frac{2\bar{\lambda}}{1-\bar{\lambda}}\sqrt{n}$,

$$|\mathcal{B}| \geq n^{-\frac{\bar{\lambda}}{1-\bar{\lambda}}\sqrt{n}}|\mathcal{A}| \geq \frac{1-\bar{\lambda}}{1+\bar{\lambda}}n^{-\frac{\bar{\lambda}}{1-\bar{\lambda}}\sqrt{n}}MN, \tag{5.5}$$

where the last inequality follows from (5.3), and (5.2) holds for the pair of sequences of random variables (\bar{X}^n, \bar{Y}^n) uniformly distributed on \mathcal{B} . That is, for all $x \in \mathcal{X}$, $y \in \mathcal{Y}$, and $t = 1, 2, \dots, n$,

$$\begin{aligned} & \frac{1}{\sqrt{n}} [Pr(\bar{X}_t = x)Pr(\bar{Y}_t = y) - 1 - |\mathcal{X}||\mathcal{Y}|] \\ & \leq Pr(\bar{X}_t = x, \bar{Y}_t = y) - Pr(\bar{X}_t = x)Pr(\bar{Y}_t = y) \\ & \leq \frac{1}{\sqrt{n}} [Pr(\bar{X}_t = x)Pr(\bar{Y}_t = y) + 1], \end{aligned} \tag{5.6}$$

where we use the assumption $\gamma = \varepsilon = \frac{1}{\sqrt{n}}$ and to obtain the last inequality we use the obvious inequality

$$\max \left\{ \left(1 + \frac{1}{\sqrt{n}} \right) Pr(\bar{X}_t = x)Pr(\bar{Y}_t = y), \frac{1}{\sqrt{n}} \right\} \leq \left(1 + \frac{1}{\sqrt{n}} \right) Pr(\bar{X}_t = x)Pr(\bar{Y}_t = y) + \frac{1}{\sqrt{n}}.$$

We first treat our $c-q$ MAC channel as a stationary memoryless $c-q$ channel with input alphabet $\mathcal{X} \times \mathcal{Y}$ and it is obvious that $(\mathcal{B}, \mathcal{D})$ is an $(n, |\mathcal{B}|, \lambda)$ -code for it, where $\lambda = \frac{1+\bar{\lambda}}{2}$.

So we may apply Theorem 1 to it and obtain

$$|\mathcal{B}| \leq \sum_{t=1}^n \frac{1}{n} \chi(P_{\bar{X}_t \bar{Y}_t}; \Phi) + \frac{1}{\sqrt{n}} h(\lambda) = \frac{1}{n} \sum_{t=1}^n \chi(\bar{X}_t, \bar{Y}_t; \Phi) + \frac{1}{\sqrt{n}} h(\lambda), \tag{5.7}$$

which together with (5.5) implies

$$\frac{1}{n} \log M + \frac{1}{n} \log N \leq \frac{1}{n} \sum_{t=1}^n \chi(\bar{X}_t, \bar{Y}_t; \Phi) + \frac{1}{\sqrt{n}} h(\lambda) + \frac{1}{\sqrt{n}} \frac{\bar{\lambda}}{1-\bar{\lambda}} \log n + \frac{1}{n} \log \frac{1+\bar{\lambda}}{1-\bar{\lambda}}. \tag{5.8}$$

Next we get for $y^n \in \mathcal{Y}^n$

$$\mathcal{B}(y^n) = \{(u^n, y^n) : (u^n, y^n) \in \mathcal{B}\}. \tag{5.9}$$

Then by our definitions

$$\mathcal{B}(y^n) = \phi \text{ if } y^n \notin \mathcal{V}^n, \tag{5.10}$$

$$Pr(\bar{Y}^n = y^n) = \frac{|\bar{\mathcal{B}}(y^n)|}{|\mathcal{B}|}, \tag{5.11}$$

and $\{\mathcal{B}(v^n) : v^n \in \mathcal{V}, \mathcal{B}(v^n) \neq \emptyset\}$ is a partition of \mathcal{B} . We partition \mathcal{V} into two parts according to $P_{\bar{Y}^n}$,

$$\mathcal{V}^+ = \{v^n \in \mathcal{V} : Pr(\bar{Y}^n = v^n) \geq (nN)^{-1}\} \tag{5.12}$$

and

$$\mathcal{V}^- = \{v^n \in \mathcal{V} : Pr(\bar{Y}^n = v^n) < (nN)^{-1}\}. \tag{5.13}$$

Then by (5.10) and (5.11)

$$1 = Pr(\bar{Y}^n \in \mathcal{V}) < Pr(\bar{Y}^n \in \mathcal{V}^+) + (nN)^{-1}|\mathcal{V}| = Pr(\bar{Y}^n \in \mathcal{V}^+) + \frac{1}{n},$$

or

$$Pr(\bar{Y}^n \in \mathcal{V}^+) > 1 - \frac{1}{n}. \tag{5.14}$$

Now we combine (5.11) and (5.12) with (5.5) and obtain that for all $v^n \in \mathcal{V}^+$

$$|\mathcal{B}(v^n)| \geq (nN)^{-1}|\mathcal{B}| \geq \frac{1 - \bar{\lambda}}{1 + \bar{\lambda}} n^{-\left(\frac{\bar{\lambda}}{1 - \bar{\lambda}} \sqrt{n} + 1\right)} M, \tag{5.15}$$

which together with (5.14) yields

$$\begin{aligned} \sum_{v^n \in \mathcal{V}^n} Pr(\bar{Y}^n = v^n) \log |\mathcal{B}(v^n)| &\geq \sum_{v^n \in \mathcal{V}^+} Pr(\bar{Y}^n = v^n) \log |\mathcal{B}(v^n)| \\ &\geq \left(1 - \frac{1}{n}\right) \left[\log M - \left(\frac{\bar{\lambda}}{1 - \bar{\lambda}} \sqrt{n} + 1\right) \log n + \log \frac{1 - \bar{\lambda}}{1 + \bar{\lambda}} \right] \\ &\geq \left(1 - \frac{1}{n}\right) \left[\log M - \sqrt{n} \log n \left(\frac{1}{1 - \bar{\lambda}} + 1 + \log \frac{1 + \bar{\lambda}}{1 - \bar{\lambda}}\right) \right] \\ &\geq \log M - \frac{1}{n} \log |\mathcal{X}^n| - \sqrt{n} \log n \left(\frac{1}{1 - \bar{\lambda}} + 1 + \log \frac{1 + \bar{\lambda}}{1 - \bar{\lambda}}\right) \\ &\geq \log M - \sqrt{n} \log n \left(\frac{2 - \bar{\lambda}}{1 - \bar{\lambda}} + \log \left(\frac{1 + \bar{\lambda}}{1 - \bar{\lambda}} |\mathcal{X}|\right)\right), \end{aligned}$$

i.e.,

$$\frac{1}{n} \log M \leq \frac{1}{n} \sum_{v^n \in \mathcal{V}^n} Pr(\bar{Y}^n = v^n) \log |\mathcal{B}(v^n)| + \frac{\log n}{\sqrt{n}} \left(\frac{2}{1 - \bar{\lambda}} + \log \left(\frac{1 + \bar{\lambda}}{1 - \bar{\lambda}} |\mathcal{X}|\right)\right). \tag{5.16}$$

Here we use the convention “ $0 \log 0 = 0$ ”. For $v^n \in \mathcal{V}$, we let

$$\mathcal{U}_{\mathcal{B}}(v^n) = \{u^n \in \mathcal{U} : (u^n, v^n) \in \mathcal{B}(v^n)\}. \tag{5.17}$$

Then obviously by definition $(\mathcal{U}_{\mathcal{B}}(v^n), \mathcal{D})$ is an $(n, |\mathcal{B}(v^n)|, \lambda)$ -code for non-stationary memoryless $c - q$ channel $\{\Phi_t(\cdot, v_t)\}_t$ where for $v^n = (v_1, v_2, \dots, v_n)$, $\Phi = \{\varphi(x, y) = x \in \mathcal{X}, y \in \mathcal{Y}\}$, $\Phi_t(\cdot, v_t) = \{\varphi(x, v_t) : x \in \mathcal{X}\}$, and $P_{\bar{X}^n|\bar{Y}^n}(\cdot|v^n)$ is the uniform distribution on $\mathcal{B}(v^n)$. Denote by $w(x, y)$, the minimum positive eigenvalue of $\varphi(x, y)$ and by $w = \min_{x,y} w(x, y)$. Then $w > 0$ since $|\mathcal{X}||\mathcal{Y}| < \infty$, and (2.5) holds. Consequently by Theorem 1 we have that

$$\begin{aligned} \frac{1}{n} \log |\mathcal{B}(v^n)| &\leq \sum_{t=1}^n \frac{1}{n} \left[S \left(\sum_{x \in \mathcal{X}} P_{\bar{X}_t|\bar{Y}_t^n}(x|v^n) \varphi(x, v^n) \right) \right. \\ &\quad \left. - \sum_{x \in \mathcal{X}} P_{\bar{X}_t|\bar{Y}_t^n}(x|v^n) S(\varphi(x, v_t)) \right] + \frac{h(\lambda)}{\sqrt{n}}. \end{aligned} \tag{5.18}$$

Notice that $\varphi(x_t, v_t)$ depends on $x^n = (x_1, \dots, x_n)$ through x_t for fixed v . One may rewrite (5.18) as

$$\begin{aligned} \frac{1}{n} \log |\mathcal{B}(v^n)| &\leq \sum_{t=1}^n \frac{1}{n} \left[S \left(\sum_{x^n \in \mathcal{X}^n} P_{\bar{X}_t^n|\bar{Y}_t^n}(x^n|v^n) \varphi(x_t, v_t) \right) \right. \\ &\quad \left. - \sum_{x^n \in \mathcal{X}^n} P_{\bar{X}_t^n|\bar{Y}_t^n}(x^n|v^n) S(\varphi(x_t, w_t)) \right] + \frac{h(\lambda)}{\sqrt{n}}. \end{aligned} \tag{5.19}$$

Next by the concavity (\cap) of von Neumann entropy and Jensens inequality we have

$$\begin{aligned} &\sum_{y^n \in \mathcal{Y}^n} P_{\bar{Y}^n}(y^n) S \left(\sum_{x^n \in \mathcal{X}^n} P_{\bar{X}^n|\bar{Y}^n}(x^n|y^n) \varphi(x_t, y_t) \right) \\ &= \sum_{y_t \in \mathcal{Y}} P_{\bar{Y}_t}(y) \sum_{\substack{i \neq t \\ y_i \in \mathcal{Y}}} \sum Pr(\tilde{Y}_i = y_i, i \neq t | \tilde{Y}_t = y_t) S \left(\sum_{x^n \in \mathcal{X}^n} P_{\bar{X}^n|\bar{Y}^n}(x^n, y^n) \varphi(x_t, y_t) \right) \\ &\leq \sum_{y_t \in \mathcal{Y}} P_{\bar{Y}_t}(y_t) S \\ &\quad \left[\sum_{i \neq t} \sum_{y_i \in \mathcal{Y}} Pr(\bar{Y}_i = y_i, i \neq t | \bar{Y}_t = y_t) \sum_{x^n \in \mathcal{X}^n} Pr(\bar{X}^n = x^n | \bar{Y}^n = y^n) \varphi(x_t, y_t) \right] \\ &= \sum_{y_t \in \mathcal{Y}} P_{\bar{Y}_t}(y_t) S \left(\sum_{x_t \in \mathcal{X}} P_{\bar{X}_t|\bar{Y}_t}(x_t|y_t) \varphi(x_t, y_t) \right) = \sum_{y_t \in \mathcal{Y}} P_{\bar{Y}_t}(y) S(\sigma_{\bar{X}_t\bar{Y}_t,2}(y)) \end{aligned} \tag{5.20}$$

for $\sigma_{\bar{X}_t\bar{Y}_t,2}(y) = \sum_{x \in \mathcal{X}} P_{\bar{X}_t|\bar{Y}_t}(x|y) \varphi(x, y)$.

This and the fact

$$\begin{aligned} & \sum_{y^n \in \mathcal{Y}^n} P_{\bar{Y}^n}(y^n) \sum_{x^n \in \mathcal{X}} P_{\bar{X}^n | \bar{Y}^n}(x^n | y^n) S(\varphi(x_t, y_t)) \\ &= \sum_{y \in \mathcal{Y}} P_{\bar{Y}_t}(y) \sum_{x \in \mathcal{X}} P_{\bar{X}_t | \bar{Y}_t}(x | y) S(\varphi(x_t, y_t)) \end{aligned}$$

together with (2.11) imply that

$$\begin{aligned} & \frac{1}{n} \sum_{t=1}^n P_{\bar{Y}^n}(y^n) \log |\mathcal{B}(y^n)| \\ & \leq \frac{1}{n} \sum_{t=1}^n \left[\sum_{y \in \mathcal{Y}} P_{\bar{Y}_t}(y) \left(S(\sigma_{\bar{X}_t \bar{Y}_t, 2}(y)) - \sum_{x \in \mathcal{X}} P_{\bar{X}_t | \bar{Y}_t}(x | y) S(\varphi(x_t, y_t)) \right) \right] + \frac{h(\lambda)}{\sqrt{n}} \\ & = \frac{1}{n} \sum_{t=1}^n \chi(\bar{X}_t; \Phi | \bar{Y}_t) + \frac{h(\lambda)}{\sqrt{n}}. \end{aligned} \tag{5.21}$$

Recalling that $P(\bar{Y}^n = y^n) = 0$, if $y^n \notin \mathcal{V}^n$, by combining (5.16) with (5.21), we have that

$$\begin{aligned} \frac{1}{n} \log M & \leq \frac{1}{n} \sum_{t=1}^n \chi(\bar{X}_t; \Phi | \bar{Y}_t) + \frac{\log n}{\sqrt{n}} \left(\frac{2}{1-\lambda} + \log \left(\frac{1+\bar{\lambda}}{1-\bar{\lambda}} |\mathcal{X}| \right) \right) + \frac{h(\lambda)}{\sqrt{n}} \\ & \leq \frac{1}{n} \sum_{t=1}^n \chi(\bar{X}_t; \Phi | \bar{Y}_t) + \frac{\log n}{\sqrt{n}} \left(\frac{2}{1-\bar{\lambda}} + \log \left(\frac{1+\bar{\lambda}}{1-\bar{\lambda}} |\mathcal{X}| \right) + h \left(\frac{1+\bar{\lambda}}{2} \right) \right), \end{aligned} \tag{5.22}$$

where in the last step we use our choice $\lambda = \frac{1+\bar{\lambda}}{2}$. By interchanging the roles of \bar{X}^n and \bar{Y}^n , we obtain in the same way that

$$\frac{1}{n} \log N \leq \frac{1}{n} \sum_{t=1}^n \chi(\bar{Y}_t | \bar{X}_t) + \frac{\log n}{\sqrt{n}} \left(\frac{2}{1-\bar{\lambda}} + \log \left(\frac{1+\bar{\lambda}}{1-\bar{\lambda}} |\mathcal{Y}| \right) + h \left(\frac{1+\bar{\lambda}}{2} \right) \right). \tag{5.23}$$

So for our proof has not been done and we only finished the second step of our plan although (5.8), (5.22), and (5.23) have the same form as (2.23), (2.24), and (2.25), because (\bar{X}_t, \bar{Y}_t) may not be independent. We have to replace (\bar{X}_t, \bar{Y}_t) by a pair of random variables $(\tilde{X}_t, \tilde{Y}_t)$ with distribution

$$P_{\tilde{X}_t \tilde{Y}_t}(x, y) = P_{\bar{X}_t}(x) P_{\bar{X}_t}(y) \tag{5.24}$$

for $x \in \mathcal{X}, y \in \mathcal{Y}$, for $t = 1, 2, \dots, n$. This is finished in the last step.

In the calculation we need two basic inequalities from Quantum Information Theory.

Strong Convexity of the Trace Distance (P. 407 [17])

Let $\{P_i\}$ and $\{Q_i\}$ be two probability distributions and let $\{\rho_i\}$ and $\{\sigma_i\}$ be two sets of states. Then

$$\text{tr} \left(\left| \sum_i P_i \rho_i - \sum_i Q_i \sigma_i \right| \right) \leq \sum_i |P_i - Q_i| + \sum_i P_i \text{tr} |\rho_i - \sigma_i|. \tag{5.25}$$

Fanne’s Inequality (Continuity of von Neumann Entropy ([10] also P. 512 [17])

For two states ρ and σ

$$|S(\rho) - S(\sigma)| \leq \frac{1}{2} \text{tr} |\rho - \sigma| \log \frac{2d}{\text{tr} |\rho - \sigma|}. \tag{5.26}$$

In the following let us denote by $PQ^+ = \{z \in Z : P(z) \geq Q(z)\}$ for two probability distributions P and Q on the same set Z . Then the second inequality in (5.6) implies that

$$\begin{aligned} & \sum_{x \in \mathcal{X} y \in \mathcal{Y}} |P_{\bar{X}_t \bar{Y}_t}(x, y) - P_{\bar{X}_t}(x) P_{\bar{Y}_t}(y)| \\ & \leq 2 \sum_{(x, y) \in P_{\bar{X}_t \bar{Y}_t}(P_{\bar{X}_t} P_{\bar{Y}_t})} + P_{\bar{X}_t \bar{Y}_t}(x, y) - P_{\bar{X}_t}(x) P_{\bar{Y}_t}(y) \\ & \leq \frac{2}{\sqrt{n}} (1 + |\mathcal{X}| |\mathcal{Y}|). \end{aligned} \tag{5.27}$$

So by (5.25) and (5.27) we have

$$\begin{aligned} & \text{tr} \left(\left| \sum_{x \in \mathcal{X} y \in \mathcal{Y}} P_{\bar{X}_t \bar{Y}_t}(x, y) \varphi(x, y) - \sum_{x \in \mathcal{X} y \in \mathcal{Y}} P_{\bar{X}_t}(x) P_{\bar{Y}_t}(y) \varphi(x, y) \right| \right) \\ & \leq \sum_{x \in \mathcal{X} y \in \mathcal{Y}} |P_{\bar{X}_t \bar{Y}_t}(x, y) - P_{\bar{X}_t}(x) P_{\bar{Y}_t}(y)| \\ & \leq \frac{2}{\sqrt{n}} (1 + |\mathcal{X}| |\mathcal{Y}|) \end{aligned} \tag{5.29}$$

Moreover, by (5.27)

$$\begin{aligned} & \left| \sum_{x, y} P_{\bar{X}_t \bar{Y}_t}(x, y) S(\varphi(x, y)) - \sum_{x, y} P_{\bar{X}_t}(x) P_{\bar{Y}_t}(y) S(\varphi(x, y)) \right| \\ & \leq \sum_{x, y} |P_{\bar{X}_t \bar{Y}_t}(x, y) - P_{\bar{X}_t}(x) P_{\bar{Y}_t}(y)| S(\varphi(x, y)) \\ & \leq \log d \sum_{x, y} |P_{\bar{X}_t \bar{Y}_t}(x, y) - P_{\bar{X}_t}(x) P_{\bar{Y}_t}(y)| \\ & \leq \frac{2}{\sqrt{n}} (1 + |\mathcal{X}| |\mathcal{Y}|) \log d. \end{aligned} \tag{5.30}$$

Now (5.26) and (5.30) imply that

$$\begin{aligned}
 & \left| \chi(\bar{X}_t, \bar{Y}_t; \Phi) - \chi(\tilde{X}_t, \tilde{Y}_t; \Phi) \right| \\
 &= \left| \left[S \left(\sum_{x,y} P_{\bar{X}_t \bar{Y}_t}(x,y) \varphi(x,y) \right) - \sum_{x,y} P_{\bar{X}_t \bar{Y}_t}(x,y) S(\varphi(x,y)) \right] \right. \\
 &\quad \left. - \left[S \left(\sum_{x,y} P_{\bar{X}_t}(x) P_{\bar{Y}_t}(y) \varphi(x,y) \right) - \sum_{x,y} P_{\bar{X}_t}(x) P_{\bar{Y}_t}(y) S(\varphi(x,y)) \right] \right| \\
 &\leq \left| S \left(\sum_{x,y} P_{\bar{X}_t \bar{Y}_t}(x,y) \varphi(x,y) \right) - S \left(\sum_{x,y} P_{\bar{X}_t}(x) P_{\bar{Y}_t}(y) \varphi(x,y) \right) \right| \\
 &\quad + \left| \sum_{x,y} P_{\bar{X}_t \bar{Y}_t}(x,y) S(\varphi(x,y)) - \sum_{x,y} P_{\bar{X}_t}(x) P_{\bar{Y}_t}(y) S(\varphi(x,y)) \right| \\
 &\leq \frac{1}{\sqrt{n}} (1 + |\mathcal{X}||\mathcal{Y}|) \log \frac{\sqrt{nd}}{1 + |\mathcal{X}||\mathcal{Y}|} + \frac{2}{\sqrt{n}} (1 + |\mathcal{X}||\mathcal{Y}|) \log d \\
 &< \frac{1}{\sqrt{n}} (1 + |\mathcal{X}||\mathcal{Y}|) \log \frac{\sqrt{nd}^3}{1 + |\mathcal{X}||\mathcal{Y}|} \\
 &< \frac{1}{\sqrt{n}} (1 + |\mathcal{X}||\mathcal{Y}|) \log \sqrt{nd}^3. \tag{5.31}
 \end{aligned}$$

Next let us turn to estimate the difference $\left| \chi(\bar{X}_t; \Phi | \bar{Y}_t) - \chi(\tilde{X}_t; \Phi | \tilde{Y}_t) \right|$. To this end we have to upper bound the difference $\left| \sum_{y \in \mathcal{Y}} P_{\bar{Y}_t}(y) S(\sigma_{\bar{X}_t \bar{Y}_t, 2}(y)) - \sum_{y \in \mathcal{Y}} P_{\tilde{Y}_t}(y) S(\sigma_{\tilde{X}_t \tilde{Y}_t, 2}(y)) \right|$ for

$$\sigma_{\bar{X}_t \bar{Y}_t, 2}(y) = \sum_{x \in \mathcal{X}} P_{\bar{X}_t | \bar{Y}_t}(x|y) \varphi(x,y) \text{ and } \sigma_{\tilde{X}_t \tilde{Y}_t, 2}(y) = \sum_{x \in \mathcal{X}} P_{\tilde{X}_t}(x) \varphi(x,y).$$

Since for all $y \sum_x P_{\bar{X}_t \bar{Y}_t}(x,y) = \sum_x P_{\bar{X}_t}(x) P_{\bar{Y}_t}(y) = P_{\bar{Y}_t}(y)$, if $P_{\bar{Y}_t}(y) \neq 0$, by (5.6)

$$\begin{aligned}
 P_{\bar{Y}_t}(y) \sum_{x \in \mathcal{X}} |P_{\bar{X}_t | \bar{Y}_t}(y|x) - P_{\bar{X}_t}(x)| &= \sum_{x \in \mathcal{X}} |P_{\bar{X}_t \bar{Y}_t}(x,y) - P_{\bar{X}_t}(x) P_{\bar{Y}_t}(y)| \\
 &= 2\Sigma^+(P_{\bar{X}_t \bar{Y}_t}(x,y) - P_{\bar{X}_t}(x) P_{\bar{Y}_t}(y)) \\
 &\leq \frac{2}{\sqrt{n}} (P_{\bar{Y}_t}(y) + |\mathcal{X}|) \\
 &\leq \frac{2}{\sqrt{n}} (1 + |\mathcal{X}|), \tag{5.32}
 \end{aligned}$$

where the sum Σ^+ is taken over all $x \in \mathcal{X}$ with $P_{\bar{X}_t | \bar{Y}_t}(x|y) \geq P_{\bar{X}_t}(x)$.

Thus by (5.26), (5.25), and (5.32), we have that

$$\begin{aligned}
 & \left| \sum_{y \in \mathcal{Y}} P_{\bar{Y}_t}(y) S(\sigma_{\bar{X}_t|\bar{Y}_t,2}(y)) - \sum_{y \in \mathcal{Y}} P_{\tilde{Y}_t}(y) S(\sigma_{\tilde{X}_t\tilde{Y}_t,2}(y)) \right| \\
 & \leq \sum_{y \in \mathcal{Y}} P_{\bar{Y}_t}(y) |S(\sigma_{\bar{X}_t\bar{Y}_t,2}(y)) - S(\sigma_{\tilde{X}_t\tilde{Y}_t,2}(y))| \\
 & \leq \sum_{y \in \mathcal{Y}} P_{\bar{Y}_t}(y) \cdot \frac{1}{2} \text{tr} |\sigma_{\bar{X}_t\bar{Y}_t,2}(y) - \sigma_{\tilde{X}_t\tilde{Y}_t,2}(y)| \log \frac{2d}{\text{tr} |\sigma_{\bar{X}_t\bar{Y}_t,2}(y) - \sigma_{\tilde{X}_t\tilde{Y}_t,2}(y)|} \\
 & \leq \frac{1}{2} \sum_{y \in \mathcal{Y}} P_{\tilde{X}_t}(y) \text{tr} |\sigma_{\bar{X}_t\bar{Y}_t,2}(y) - \sigma_{\tilde{X}_t\tilde{Y}_t,2}(y)| \log \frac{2d}{P_{\tilde{X}_t}(y) \text{tr} |\sigma_{\bar{X}_t\bar{Y}_t,2}(y) - \sigma_{\tilde{X}_t\tilde{Y}_t,2}(y)|} \\
 & \leq \frac{1}{2} \sum_{y \in \mathcal{Y}} P_{\tilde{Y}_t}(y) \sum_{x \in \mathcal{X}} |P_{\bar{X}_t|\bar{Y}_t}(x|y) - P_{\tilde{X}_t}(x)| \log \frac{2d}{P_{\tilde{Y}_t}(y) \sum_{x \in \mathcal{X}} |P_{\bar{X}_t|\bar{Y}_t}(x|y) - P_{\tilde{X}_t}(x)| \varphi(x, y)} \\
 & \leq \frac{1}{2} \sum_{y \in \mathcal{Y}} \frac{2}{\sqrt{n}} (1 + |\mathcal{X}|) \log \frac{d\sqrt{n}}{1 + |\mathcal{X}|} \\
 & = \frac{1}{\sqrt{n}} (1 + |\mathcal{X}|) |\mathcal{Y}| \log \frac{d\sqrt{n}}{1 + |\mathcal{X}|} \\
 & \leq \frac{1}{\sqrt{n}} (1 + |\mathcal{X}|) |\mathcal{Y}| \log d \sqrt{n} \tag{5.33}
 \end{aligned}$$

where the second inequality holds by (5.26); the fourth inequality follows from (5.25) and the monotonicity of $z \log \frac{2d}{z}$ in the interval $[0, \frac{2d}{e}]$; and the fifth inequality follows from (5.32).

Considering that by (2.11)

$$\chi(\bar{X}_t; \Phi|\bar{Y}_t) = \sum_{y \in \mathcal{Y}} P_{\bar{Y}_t} S(\sigma_{\bar{X}_t\bar{Y}_t,2}(y)) - \sum_{x,y} P_{\bar{X}_t\bar{Y}_t}(x, y) S(\varphi(x, y))$$

and $\chi(\tilde{X}_t; \Phi|\tilde{Y}_t) = \sum_{y \in \mathcal{Y}} P_{\tilde{Y}_t} S(\sigma_{\tilde{X}_t\tilde{Y}_t,2}(y)) - \sum_{x,y} P_{\tilde{X}_t} P_{\tilde{Y}_t}(y) S(\varphi(x, y))$, we add up (5.30) and (5.33),

$$\begin{aligned}
 & |\chi(\bar{X}_t; \Phi|\bar{Y}_t) - \chi(\tilde{X}_t; \Phi|\tilde{Y}_t)| \\
 & \leq \left| \sum_{y \in \mathcal{Y}} P_{\bar{Y}_t}(y) S(\sigma_{\bar{X}_t\bar{Y}_t,2}(y)) - \sum_{y \in \mathcal{Y}} P_{\tilde{Y}_t}(y) S(\sigma_{\tilde{X}_t\tilde{Y}_t,2}(y)) \right| \\
 & \quad + \left| \sum_{x,y} P_{\bar{X}_t\bar{Y}_t}(x, y) S(\varphi(x, y)) - \sum_{x,y} P_{\tilde{X}_t}(x) P_{\tilde{Y}_t}(y) S(\varphi(x, y)) \right| \\
 & \leq \frac{1}{\sqrt{n}} (2 + |\mathcal{Y}| + 3|\mathcal{X}||\mathcal{Y}|) \log d + \frac{1}{2\sqrt{n}} (1 + |\mathcal{X}|) |\mathcal{Y}| \log n. \tag{5.34}
 \end{aligned}$$

Next, we exchange the roles of \bar{X}_t and \bar{Y}_t in (5.34), in the same way we obtain that

$$|\chi(\bar{Y}_t; \Phi|\bar{X}_t) - \chi(\tilde{Y}_t; \Phi|\tilde{X}_t)| \leq \frac{1}{\sqrt{n}} (2 + |\mathcal{X}| + 3|\mathcal{X}||\mathcal{Y}|) \log d + \frac{1}{2\sqrt{n}} (1 + |\mathcal{Y}|) |\mathcal{X}| \log n. \tag{5.35}$$

Finally we set

$$\tilde{h}(\bar{\lambda}) = h\left(\frac{1+\bar{\lambda}}{2}\right) + \frac{2+\bar{\lambda}}{1-\bar{\lambda}} + \log\left(\frac{1+\bar{\lambda}}{1-\bar{\lambda}}|\mathcal{X}||\mathcal{Y}|\right) + (3+6|\mathcal{X}||\mathcal{Y}|)\log d$$

and combine (5.8) with (5.31), (5.22) with (5.34), and (5.23) with (5.35), respectively. Then (2.23) – (2.25) follow. \square

References

1. R. Ahlswede, Beiträge zur Shannonschen Informationstheorie im Falle nichtstationärer Kanäle, *Z. Wahrsch. und verw. Gebiete*, Vol. 10, 1–42, 1968.
2. R. Ahlswede, Multi-way communication channels, 2nd Intern. Symp. on Inform. Theory, Thakadsor, 1971, Publ. House of the Hungarian Acad. of Sci., 23–52, 1973.
3. R. Ahlswede, An elementary proof of the strong converse theorem for the multiple-access channel, *J. Comb. Inform. Sys. Sci.*, Vol. 7, 216–230, 1982.
4. R. Ahlswede, Performance parameters for channels, this volume.
5. A.E. Allahverdyan and D.B. Saakian, Multi-access channels in quantum information theory, <http://xxx.lanl.gov/abs/quant-ph/9712034V1>, 1997.
6. S. Arimoto, On the converse to the coding theorem for discrete memoryless channels, *IEEE Trans. Inform. Theory*, Vol. 19, 375–393, 1963.
7. U. Augustin, Gedächtnisfreie Kanäle für diskrete Zeit, *Z. Wahrscheinlichkeitstheorie u. verw. Gebiete*, Vol. 6, 10–61, 1966.
8. G. Dueck, Maximal error capacity regions are smaller than average error capacity regions for multi-user channels, *Problems of Control and Information Theory*, Vol. 7, 11–19, 1978.
9. G. Dueck, The strong converse to the coding theorem for the multiple-access channel, *J. Comb. Inform. Syst. Sci.*, Vol. 6, 187–196, 1981.
10. M. Fannes, A continuity property of the entropy density for spin lattice systems, *Commun. Math. Phys.* 31, 291–294, 1973.
11. R.M. Fano, *Class Notes for Transmission of Information*, Course 6.574, MIT, Cambridge, Mass., 1952.
12. A. Feinstein, A new basic theorem of information theory, *IRE trans., Inform. Theory*, Vol. 4, 2–22, 1954.
13. M. Hayashi and H. Nagaoka, General formulas for capacity of classical–quantum channels, <http://xxx.lanl.gov/abs/quant-ohy/0206186V1>, 2002.
14. A.S. Holevo, Problems in the mathematical theory of quantum communication channels, *Rep. Math. Phys.* 12, 2, 273–278, 1977.
15. A.S. Holevo, The capacity of the quantum channel with general signal states, *IEEE Trans. Inform Theory*, Vol. 44, 269–273, 1998.
16. A.S. Holevo, Coding theorems for quantum channels, <http://xxx.lanl.gov/abs/quant-ph/9809023V1>, 1998.
17. M.A. Nielsen and I. Chuang, *Quantum Computation and Quantum Information*, Cambridge University Press, 2000.
18. T. Ogawa and H. Nagaoka, Strong converse to the quantum channel coding theorem, *IEEE Trans. Inform. Theory*, Vol. 45, 2486–2489, 1999.
19. T. Ogawa and H. Nagaoka, Strong converse and Stein’s lemma in quantum hypothesis testing, *IEEE Trans. Inform. Theory*, Vol. 46, 2428–2433, 2000.

20. D. Petz, Quasientropies for states of a von Neumann algebra, Publ. RIMS, Kyoto Univ. Vol. 21, 787–800, 1985.
21. D. Petz, Quasientropies for finite quantum systems, Rep. Math. Phys., Vol. 23, 57–65, 1986.
22. B. Schumacher and M. Westmoreland, Sending classical information via noisy quantum channel, Phys. Rev. A, Vol. 56, 1, 131–138, 1997.
23. C.E. Shannon, A mathematical theory of communication, Bell Sys. Tech. Journal, Vol. 27, 379–423, 1948.
24. C.E. Shannon, Two-way communication channels, Proc. 4th Berkeley Symp. Math. Statist. and Prob., Unvi. of Calif. Press, Berkeley, Vol. 1, 611–644, 1961.
25. S. Verdú and T.S. Han, A general formula for channel capacity, IEEE Trans. Inform. Theory, Vol. 40, 1147–1157, 1994.
26. A. Winter, The capacity region of the quantum multiple access channel, <http://xxx.lanl.gov/abs/quant-ph/9807019>, 1998.
27. A. Winter, Coding theorem and strong converse for quantum channels, IEEE Trans. Inform. Theory, Vol. 45, 2481–2485, 1999.
28. A. Winter, Coding theorem and strong converse for nonstationary quantum channels, Preprint 99–033, SFB 343 “Diskrete Strukturen in der Mathematik”, Universität Bielefeld, 1999.
29. J. Wolfowitz, The coding of message subject to chance errors, Illinois J. Math. 1, 591–606, 1957.
30. J. Wolfowitz, Coding Theorems of Information Theory, Springer-Verlag, Berlin, Heidelberg, New York, 3-rd edition, 1978.