Superadditivity of Quantum Channel Coding Rate with Finite Blocklength Quantum Measurements

Hye Won Chung Lizhong Zheng
EECS, MIT, Cambridge, MA 02139, USA. Email: {hwchung,lizhong}@mit.edu

***Abstract*—This paper investigates the superadditivity of the maximum achievable information rate in a quantum channel. The superadditivity is a phenomenon that the maximum accessible information per channel strictly increases as the number of quantum states jointly measured increases. We give an explanation of this phenomenon by comparing the quantum channel with the classical discrete memoryless channel (DMC) under the concatenated codes, and give a lower bound on the maximum accessible information per channel at a finite length of quantum measurements in terms of V, which is the quantum version of channel dispersion, and C, the capacity of the quantum channel.**

I. Introduction and Problem Definition

How many classical bits can be reliably communicated through a quantum channel has been an interesting question in quantum information theory in the effort to understand the intrinsic limit on the classical capacity of a quantum channel such as an optical fiber. In this paper, we will investigate this fundamental question with consideration of the trade-off between the maximum achievable information rate and the complexity of quantum processing for a finite blocklength of measurements.

The capacity of a quantum channel is defined as the maximum number of information bits that can be modulated into the input quantum states and reliably decoded with a set of quantum measurements as the number of transmissions N_c goes to infinity. In [1], [2], authors showed that the capacity of a quantum channel with its input states $\{|\psi_x\rangle\}$, $x \in \mathcal{X}$, is

$$C = \max_{P_X} \mathrm{Tr}(-\rho \log \rho) \tag{1}$$

where $\rho = \sum_{x\in\mathcal{X}} P_X(x)|\psi_x\rangle\langle\psi_x|$.

The input states $|\psi_x\rangle$, $x \in \mathcal{X}$, are normalized vectors in a complex Hilbert space $\mathcal{H}$, and when $\langle\psi_x|$ is the Hermitian conjugate vector of $|\psi_x\rangle$, ρ is a *density operator* that is the linear combination of the outer product of each state $|\psi_x\rangle\langle\psi_x|$ with $P_X(x)$ as a coefficient. Using the definition of the von Neumann entropy $S(\rho) = \mathrm{Tr}(-\rho\log\rho)$ for a density operator ρ, the capacity can also be written as $C = \max_{P_X} S(\rho)$.

For input codeword $[x_1, \cdots, x_{N_c}]$, the sequence of quantum states $|\psi_{x_1}\rangle, \cdots, |\psi_{x_{N_c}}\rangle$ is transmitted through a lossless quantum channel and arrives at the receiver. In the design of the codeword, it is assumed that no *entangled* codeword is used, which means that the length-N_c codeword is mapped to the N_c-fold tensor product of the input states, but not to any linear combinations of the product states. The received sequence of quantum states, which can be written as a tensor product of the states, $|\psi_{x_1}\rangle \otimes \cdots \otimes |\psi_{x_{N_c}}\rangle$, is jointly detected by orthogonal projective measurements in the N_c-fold Hilbert space $\mathcal{H}^{\otimes N_c}$. When the received codeword is projected into the orthogonal measurements $\{|\Phi_k\rangle\}$, $k \in \mathcal{K}$, which resolves identity, i.e., $\sum_k |\Phi_k\rangle\langle\Phi_k| = \mathbb{1}$, in $\mathcal{H}^{\otimes N_c}$, the classical output k is observed with the probability equal to the magnitude squared of the inner product between the received codeword and the measurement vector $|\Phi_k\rangle$ corresponding to the output k. The orthogonal projective measurements are designed to decode the received codewords with as small error probability as possible. In [1], [2], it was shown that for any rate $R < C$, there exists a block code of length N_c and rate R such that the codewords can be decoded by a set of quantum measurements with arbitrarily small probability of error as N_c goes to infinity.

To achieve this capacity, however, a joint detection receiver (JDR) needs to be implemented, which can measure the length-N_c sequence of states jointly and decode it reliably among $e^{N_c \cdot C}$ possible messages. The number of different measurements incorporated in the receiver increases exponentially in N_c. Moreover, since it is very costly to devise mathematically described quantum measurements with physical components, complexity of communication over a quantum channel can be measured by the number of different measurements implemented in the receiver for the detection of the quantum states. With consideration of the complexity, the receiver can restrict the maximum length of the sequence of states to be jointly detected as a finite number independent of the length of a codeword. However, there is no guarantee that such quantum measurements of a fixed blocklength can still achieve the ultimate capacity of the quantum channel.

Therefore, to understand the trade-off between performance and complexity of communication over a quantum channel, this paper investigates the maximum number of classical information bits that can be reliably decoded when the quantum states of a finite blocklength N, which can be shorter than the length of the codeword N_c, i.e., $N \leq N_c$, are jointly measured. As the number of quantum states jointly measured increases, the maximum number of classical bits extracted per quantum state increases, and this phenomenon is called *superadditivity* of the maximum achievable information rate in a quantum channel. After the receiver detects the quantum states, it can collect all the classical information extracted from each block of length N, and then apply any decoding algorithm over the collected information to decode the transmitted message reliably. To explain how it works, we introduce the architecture of concatenation over a quantum channel in Fig. 1.

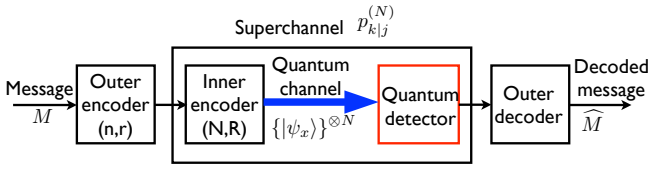

Fig. 1. Concatenated codes over a quantum channel

In the communication system depicted in Fig. 1, a *concatenated code* is used to transmit the message M over the quantum channel. For an inner code of length N and rate R, there can be total e^{NR} inputs to the inner encoder and the input is denoted as $J \in \{1, \ldots, e^{NR}\}$. Depending on the input J, the inner encoder generates a length-N sequence of quantum states, and the quantum states are transmitted through the quantum channel. The quantum detector measures the length-N quantum states jointly and generates an estimate $K \in \{1, \ldots, e^{NR}\}$ of the encoded message J. Generally, the estimate would match the input, but for a fixed N, the error probability may not be close to 0. Therefore, with the transition probability $p_{k|j}^{(N)} := \mathsf{Pr}(K = k|J = j)$, the operation by the inner encoder-quantum channel-quantum detector can be viewed as a discrete memoryless *superchannel*. The maximum mutual information of the superchannel for a finite inner code blocklength N is defined as

$$C_N := \max_{p_j} \max_{\{\text{Length-}N \text{ inner code-measurements}\}} I(p_j, p_{k|j}^{(N)}). \quad (2)$$

Note that the channel distribution $p_{k|j}^{(N)}$ depends on the set of measurements at the quantum detector as well as the inner code of a blocklength N.

An outer code can be used to reliably communicate information through the superchannel of the maximum mutual information C_N. When an outer code of length n and rate r is adopted, the total number of messages transmitted by the code is $e^{nr} = e^{nN(r/N)}$. Since the overall length of the concatenated code, which is composed of the inner code and the outer code, is $N_c = nN$, the total rate of the concatenated code is $R_c = r/N$. By Shannon's coding theorem, for any rate $r < C_N$, there exists an outer code of length n and rate r such that the decoding error can be arbitrarily small as $n \to \infty$. Therefore, the maximum information rate achievable by the concatenated code *per use of the quantum channel* can approach C_N/N.

From the definition of C_N, superadditivity of the quantity, i.e., $C_{N_1} + C_{N_2} \leq C_{N_1+N_2}$, can be shown, which implies the existence of the $\lim_{N\to\infty} C_N/N$. Holevo [3] showed that the limit is equal to the ultimate capacity of the quantum channel, $\lim_{N\to\infty} C_N/N = C = \max_{P_X} S(\rho)$. Therefore, C_N/N is an increasing sequence in N with its limit equal to the capacity.

The question we want to answer is: how does the maximum achievable information rate C_N/N increase as the length of the quantum measurement, N, increases? On the receiver side, since the quantum processing occurs only at the quantum detector for the inner code of a finite length N, the complexity of the quantum processing only depends on N, but not on the outer code length n. Therefore, the trade-off between the performance and the complexity can be captured by how fast C_N/N increases with N. In [3], [4], for certain sets of input states, the exact C_1 is calculated and compared with the capacity C where strict superadditivity is demonstrated by showing $C_1 < C$. In the next section, we will provide the examples. The calculation of C_N, however, has not been able to be generalized for a finite $N > 1$ because the complexity of optimization increases exponentially in N.

Instead of aiming to calculate the exact C_N, in this paper, a lower bound of C_N/N, which becomes tight for large enough N, is derived. From the bound, it will be possible to calculate the inner code blocklength N at which a given fraction of the ultimate capacity is achievable. A new framework for understanding the strict superadditivity of C_N in quantum channels will also be provided, which is different from the previous explanation of the phenomenon by *entangled measurements* and the resulting memory in the quantum channel [5]. Moreover, under the new framework, the superadditivity of C_N, which has been mostly thought to be a unique property observed in quantum channels but not in classical DMCs, can be understood as a more general phenomenon that happens even in classical DMCs when the concatenated code is used with an inner decoder that makes a hard-decision at a finite inner code blocklength.

The rest of the paper is outlined as follows. In Section II, examples of quantum channels where strict superadditivity $C_1 < C$ can be shown will be offered. The main theorem that states a lower bound of C_N/N will be introduced in Section III with examples to show how to use the main theorem to calculate a blocklength N to achieve a given fraction of the capacity. The theorem will be proved in Section IV. The superadditivity will be explained by comparing the quantum channel and the classical DMC under the architecture of concatenation in Section V. The approximation of the lower bound of C_N/N will also be provided by introducing a quantum version of *channel dispersion* V. In terms of V and C, the superadditivity in the quantum channel and the classical channel will be compared. Section VI will give the conclusion of this paper.

II. Strict Superadditivity of C_N

Before investigating how C_N/N increases with N, we will give examples with which strict superadditivity of the quantity C_N can be shown from $C_1 < C$. As discussed before, given a set of input states $\{|\psi_x\rangle\}$, C can be calculated from Holevo's result by finding the optimum input distribution that maximizes the von Neumann entropy $S(\rho) = \mathrm{Tr}(-\rho \log \rho)$ where $\rho = \sum_x P_X(x)|\psi_x\rangle\langle\psi_x|$. Calculating C_1 requires finding a set of measurements as well as an input distribution to maximize the resulting mutual information. For general input states, this can already be a hard optimization problem, but for binary inputs, C_1 can be calculated and represented as a function of the inner product between the two input states. In [3], for binary input states $\{|\psi_0\rangle, |\psi_1\rangle\}$ of the inner product $|\langle\psi_0|\psi_1\rangle| = \gamma$, C_1 and C are calculated and the strict

superadditivity of the binary input channel is demonstrated. The result is summarized below.

The first step to calculate C for the binary input channel is to find two eigenvalues of ρ under an input distribution $[1-q, q]$. For $\rho = (1-q)|\psi_0\rangle\langle\psi_0| + q|\psi_1\rangle\langle\psi_1|$, the eigenvectors of ρ have a form of $|\psi_0\rangle + \beta|\psi_1\rangle$ with some β that satisfies

$$\rho(|\psi_0\rangle + \beta|\psi_1\rangle) = \sigma(|\psi_0\rangle + \beta|\psi_1\rangle) \tag{3}$$

with eigenvalues σ. By solving the equation, the two eigenvalues are

$$\begin{aligned}\sigma_1 =& \frac{1}{2}\left(1 - \sqrt{1 - 4q(1-q)(1-\gamma^2)}\right)\\ \sigma_2 =& \frac{1}{2}\left(1 + \sqrt{1 - 4q(1-q)(1-\gamma^2)}\right),\end{aligned} \tag{4}$$

and the resulting $S(\rho)$ is

$$S(\rho) = \mathrm{Tr}(-\rho\log\rho) = -\sigma_1\log\sigma_1 - \sigma_2\log\sigma_2. \tag{5}$$

where $|\langle\psi_0|\psi_1\rangle| = \gamma$. From this equation, it can be shown that $S(\rho)$ for the binary inputs is maximized at $q = 1/2$, and the resulting capacity of the binary channel

$$C = \max_{P_X} S(\rho) = -\frac{1-\gamma}{2}\log\frac{1-\gamma}{2} - \frac{1+\gamma}{2}\log\frac{1+\gamma}{2}. \tag{6}$$

Compared to C_1, whose derivation can be found in [3],

$$\begin{aligned}C_1 =& \frac{1-\sqrt{1-\gamma^2}}{2}\log\left(1-\sqrt{1-\gamma^2}\right)\\ &+ \frac{1+\sqrt{1-\gamma^2}}{2}\log\left(1+\sqrt{1-\gamma^2}\right),\end{aligned} \tag{7}$$

the capacity C is strictly greater than C_1 for all $0 < \gamma < 1$, which demonstrates the strict superadditivity of C_N for all binary input quantum channels.

In the rest of this section, the superadditivity of C_N in quantum channels *with input constraint* will be considered. The constraint on the input will be described by the average energy of the input states. To discuss an average energy of the *coherent state*, which is a quantum state from a laser beam, we will use a bold face $|\boldsymbol{\alpha}\rangle$, $\boldsymbol{\alpha} \in \mathbb{C}$, to denote a coherent state, compared to a general quantum state $|\psi\rangle$. The average photon number in the coherent state $|\boldsymbol{\alpha}\rangle$ is $|\boldsymbol{\alpha}|^2$. Since the energy of *a photon* with angular frequency ω is $E = \hbar\omega$ with $\hbar = h/2\pi$ where the Planck constant $h = 6.63\times10^{-34}\mathrm{m}^2\mathrm{kg/s}$, the average energy in the coherent state $|\boldsymbol{\alpha}\rangle$ with a fixed angular frequency ω is $\hbar|\boldsymbol{\alpha}|^2\omega$. Note that the average photon number $|\boldsymbol{\alpha}|^2$ is dimensionless. For the fixed angular frequency, the average energy constraint on the input states can be represented as a condition on the average photon numbers per transmission. For example, for a fixed set of input states $\{|\boldsymbol{\alpha}_1\rangle, |\boldsymbol{\alpha}_2\rangle, \ldots, |\boldsymbol{\alpha}_K\rangle\}$ with the angular frequency ω, when the average energy per transmission of an input state should be less than $\hbar\omega\mathcal{E}_{\mathsf{ch}}$, the condition can be written as a constraint on the input distribution $[p_1, \ldots, p_K]$ such that

$$\sum_{i=1}^{K} p_i|\boldsymbol{\alpha}_i|^2 \le \mathcal{E}_{\mathsf{ch}} \tag{8}$$

with the maximum average photon number per transmission, denoted as $\mathcal{E}_{\mathsf{ch}}$.

The important question of how many bits can be reliably communicated through a quantum channel under the constraint on the average photon number per transmission has been answered in [6]. For continuous complex input of coherent states $|\boldsymbol{\alpha}\rangle$ under the energy constraint of $\mathbb{E}[|\boldsymbol{\alpha}|^2] \le \mathcal{E}_{\mathsf{ch}}$, the capacity of the channel $C(\mathcal{E}_{\mathsf{ch}})$ is

$$C(\mathcal{E}_{\mathsf{ch}}) = (1+\mathcal{E}_{\mathsf{ch}})\log(1+\mathcal{E}_{\mathsf{ch}}) - \mathcal{E}_{\mathsf{ch}}\log\mathcal{E}_{\mathsf{ch}} \text{ [nats/ch.]}, \tag{9}$$

and it is achievable with the complex Gaussian distribution with variance $\mathcal{E}_{\mathsf{ch}}$.

How many information bits can be reliably communicated *per use of a photon* with the constraint of the average photon number per channel, $\mathcal{E}_{\mathsf{ch}}$, can be calculated from $C(\mathcal{E}_{\mathsf{ch}})/\mathcal{E}_{\mathsf{ch}}$. From (9), it can be shown that to achieve high photon information efficiency (PIE) [nats/photon], the average photon number transmitted per channel should be small, and therefore a new input state should be modulated in every small time interval. In the high PIE regime where $\mathcal{E}_{\mathsf{ch}} \to 0$, the capacity (9) can be approximated as

$$C(\mathcal{E}_{\mathsf{ch}}) = \mathcal{E}_{\mathsf{ch}}\log\frac{1}{\mathcal{E}_{\mathsf{ch}}} + \mathcal{E}_{\mathsf{ch}} + o(\mathcal{E}_{\mathsf{ch}}). \tag{10}$$

Under this scaling law, it can be shown that the ultimate capacity is achievable even with the simple binary input states under the same energy constraint. For example, Binary Phase Shift Keying (BPSK) coherent states $\{|\sqrt{\boldsymbol{\mathcal{E}_{\mathsf{ch}}}}\rangle, |-\sqrt{\boldsymbol{\mathcal{E}_{\mathsf{ch}}}}\rangle\}$, which satisfy the energy constraint with any input distribution, can achieve (10) at input distribution $[1/2, 1/2]$. From a property of coherent states that the inner product between two coherent states is $|\langle\boldsymbol{\alpha}|\boldsymbol{\beta}\rangle| = \exp[-|\boldsymbol{\alpha}-\boldsymbol{\beta}|^2/2]$, the inner product between the BPSK inputs can be written as a function of $\mathcal{E}_{\mathsf{ch}}$

$$|\langle\sqrt{\boldsymbol{\mathcal{E}_{\mathsf{ch}}}}|-\sqrt{\boldsymbol{\mathcal{E}_{\mathsf{ch}}}}\rangle| = \exp[-2\mathcal{E}_{\mathsf{ch}}]. \tag{11}$$

By plugging it into the capacity equation for binary input channel (6), the capacity of BPSK input is

$$C_{\mathsf{BPSK}}(\mathcal{E}_{\mathsf{ch}}) = \mathcal{E}_{\mathsf{ch}}\log\frac{1}{\mathcal{E}_{\mathsf{ch}}} + \mathcal{E}_{\mathsf{ch}} + o(\mathcal{E}_{\mathsf{ch}}), \tag{12}$$

which is equal to $C(\mathcal{E}_{\mathsf{ch}})$ for the first- and second-order terms as $\mathcal{E}_{\mathsf{ch}} \to 0$. This result clarifies that the binary input states are good enough to communicate information bits with the ultimate photon information efficiency in the low photon number regime.

Now, we ask, for binary inputs under the same energy constraint, how high an information rate is achievable when each quantum state is measured one-by-one, i.e., $N = 1$. The maximum rate of binary channels at $N = 1$ will be denoted as $C_{1,\mathsf{Binary}}(\mathcal{E}_{\mathsf{ch}})$. For the BPSK input states, by using (7), the maximum achievable rate at $N = 1$ is

$$C_{1,\mathsf{BPSK}}(\mathcal{E}_{\mathsf{ch}}) = 2\mathcal{E}_{\mathsf{ch}} + o(\mathcal{E}_{\mathsf{ch}}) \tag{13}$$

Note that PIE of only 2 nats/photon is achievable by the BPSK channel at $N = 1$, while, for an arbitrarily large N, the achievable photon efficiency diverges as $\mathcal{E}_{\mathsf{ch}} \to 0$.

The $C_{1,\mathsf{Binary}}(\mathcal{E}_{\mathsf{ch}})$ can also be calculated in the regime of low $\mathcal{E}_{\mathsf{ch}}$ by finding the optimum binary inputs $\{|\boldsymbol{\alpha}_0\rangle, |\boldsymbol{\alpha}_1\rangle\}$ with distribution $[1-q, q]$ that satisfy the average photon number constraint.

$$(1-q)|\boldsymbol{\alpha}_0|^2 + q|\boldsymbol{\alpha}_1|^2 \leq \mathcal{E}_{\mathsf{ch}}. \tag{14}$$

The following lemma summarizes the result.

Lemma 1: The optimum binary inputs occur at $\boldsymbol{\alpha}_0 = \sqrt{\mathcal{E}_{\mathsf{ch}} \cdot q^*/(1-q^*)}$ *and* $\boldsymbol{\alpha}_1 = -\sqrt{\mathcal{E}_{\mathsf{ch}} \cdot (1-q^*)/q^*}$ *with*

$$q^* = \frac{\mathcal{E}_{\mathsf{ch}}}{2} \log \frac{1}{\mathcal{E}_{\mathsf{ch}}}, \tag{15}$$

and the resulting $C_{1,\mathsf{Binary}}(\mathcal{E}_{\mathsf{ch}})$ *is*

$$C_{1,\mathsf{Binary}}(\mathcal{E}_{\mathsf{ch}}) = \mathcal{E}_{\mathsf{ch}} \log \frac{1}{\mathcal{E}_{\mathsf{ch}}} - \mathcal{E}_{\mathsf{ch}} \log \log \frac{1}{\mathcal{E}_{\mathsf{ch}}} + O(\mathcal{E}_{\mathsf{ch}}). \tag{16}$$

Compared to the ultimate capacity (10) with the same energy constraint, the first-order term of $C_{1,\mathsf{Binary}}(\mathcal{E}_{\mathsf{ch}})$ is the same as that of $C(\mathcal{E}_{\mathsf{ch}})$, but the difference in the second-order term shows how less capacity is achievable at $N=1$ even with the optimized input states. In [7], it was shown that (16) can be achievable by a simple direct photon detector. Therefore, all the performance gain from the complex quantum processing by the JDR is captured by the difference between the second-order terms of (10) and (16). In the low photon number regime, the difference in the second-order term matters greatly in the practical design of the communication system over the quantum channel. Therefore, it would be interesting to ask how large N is needed to bridge the gap in the second-order term. To answer such a question, in the next section, a lower bound on the general C_N will be provided.

III. Lower Bound on C_N

In this section, a lower bound is derived for the maximum achievable information rate at a finite blocklength N of quantum measurements. By using the result, it will be possible to calculate a blocklength N at which a given fraction of the capacity can be achieved. Therefore, the result will provide a framework to understand the trade-off between performance and complexity in the transmission of classical information over the quantum channel.

Theorem 1: For a given set of input states $\{|\psi_x\rangle\}$*, the maximum achievable information rate at a blocklength* N *of quantum measurements, which is* C_N/N *in (2), is lower bounded by*

$$\frac{C_N}{N} \geq \max_R \left((1-2e^{-NE(R)})R - \frac{\log 2}{N} \right) \tag{17}$$

where

$$E(R) = \max_{0 \leq s \leq 1} \left(\max_{P_X} \left(-\log \mathrm{Tr}(\rho^{1+s}) \right) - sR \right) \tag{18}$$

with $\rho = \sum_x P_X(x) |\psi_x\rangle\langle\psi_x|$.

By using this theorem, for the BPSK input channel, a blocklength N can be calculated at which the lower bound of (17) exceeds certain targeted rates below its capacity. In the previous section, it was shown that there is a gap between $C_{1,\mathsf{BPSK}}(\mathcal{E}_{\mathsf{ch}})/\mathcal{E}_{\mathsf{ch}}$ in (13) and $C_{\mathsf{BPSK}}(\mathcal{E}_{\mathsf{ch}})/\mathcal{E}_{\mathsf{ch}}$ in (12) as $\mathcal{E}_{\mathsf{ch}} \to 0$. The BPSK channel works as well as the optimum continuous input channel as N goes to infinity, i.e., $C_{\mathsf{BPSK}}(\mathcal{E}_{\mathsf{ch}})$ is the same as $C(\mathcal{E}_{\mathsf{ch}})$ for the first two order terms; but at the blocklength $N=1$, it cannot even achieve the maximum mutual information of the optimum binary input channel, $C_{1,\mathsf{Binary}}(\mathcal{E}_{\mathsf{ch}})$ in (16). Therefore, the performance of the BPSK channel depends significantly on the regime of N. Now, we will find how much quantum processing is enough in order to communicate through the BPSK channel at rates close to its capacity.

The following corollary summarizes an answer for the question. Note that for the BPSK inputs $\{|\sqrt{\boldsymbol{\mathcal{E}_{\mathsf{ch}}}}\rangle, |-\sqrt{\boldsymbol{\mathcal{E}_{\mathsf{ch}}}}\rangle\}$, any input distribution satisfies the energy constraint of $\mathcal{E}_{\mathsf{ch}}$. Consequently, we can directly apply the main theorem to the BPSK channel while satisfying the energy constraint, even though the theorem itself does not consider the energy constraint.

Corollary 1: For the BPSK input channel, in the regime $N \geq \mathcal{E}_{\mathsf{ch}}^{-1}(\log(1/\mathcal{E}_{\mathsf{ch}}))$*,*

$$\frac{C_{N,\mathsf{BPSK}}}{N} \geq \left((1-2e^{-NE_{\mathsf{BPSK}}(R^*)})R^* - \frac{\log 2}{N} \right) \tag{19}$$

where

$$R^* = \mathcal{E}_{\mathsf{ch}} \log \frac{1}{\mathcal{E}_{\mathsf{ch}}} \left(1 - \sqrt{\frac{\log\left(N\mathcal{E}_{\mathsf{ch}} \log(N\mathcal{E}_{\mathsf{ch}})\right)}{N\mathcal{E}_{\mathsf{ch}}}} \right) + \mathcal{E}_{\mathsf{ch}},$$

$$E_{\mathsf{BPSK}}(R) = -\log\left(\left(\frac{1+e^{-2\mathcal{E}_{\mathsf{ch}}}}{2} \right)^{1+s} + \left(\frac{1-e^{-2\mathcal{E}_{\mathsf{ch}}}}{2} \right)^{1+s} \right) - sR$$

$$\textit{for } \; s = \frac{\log\log(1/\mathcal{E}_{\mathsf{ch}}) - \log(R-\mathcal{E}_{\mathsf{ch}})}{\log(1/\mathcal{E}_{\mathsf{ch}})} - 1.$$

We assume that $\mathcal{E}_{\mathsf{ch}}$ *is small enough to make the corresponding* s *to be* $0 \leq s \leq 1$.

Remark 1: For a narrower range of N such that

$$\mathcal{E}_{\mathsf{ch}}^{-1}(\log(1/\mathcal{E}_{\mathsf{ch}}))^2 \leq N \leq \mathcal{E}_{\mathsf{ch}}^{-2},$$

the lower bound can be further simplified as

$$\frac{C_{N,\mathsf{BPSK}}}{N} \geq \mathcal{E}_{\mathsf{ch}} \log \frac{1}{\mathcal{E}_{\mathsf{ch}}} \left(1 - \sqrt{\frac{\log\left(N\mathcal{E}_{\mathsf{ch}} \log(N\mathcal{E}_{\mathsf{ch}})\right)}{N\mathcal{E}_{\mathsf{ch}}}} \right) + \mathcal{E}_{\mathsf{ch}} + o(\mathcal{E}_{\mathsf{ch}}).$$

Therefore, at $N = \mathcal{E}_{\mathsf{ch}}^{-1}(\log(1/\mathcal{E}_{\mathsf{ch}}))^2 (\log\log(1/\mathcal{E}_{\mathsf{ch}}))^2$,

$$\frac{C_{N,\mathsf{BPSK}}}{N} \geq \mathcal{E}_{\mathsf{ch}} \log \frac{1}{\mathcal{E}_{\mathsf{ch}}} + \mathcal{E}_{\mathsf{ch}} + o(\mathcal{E}_{\mathsf{ch}}). \tag{20}$$

At the level of N, the lower bound already approaches to $C_{\mathsf{BPSK}}(\mathcal{E}_{\mathsf{ch}})$, which is the maximum achievable information rate of the BPSK channel with an arbitrarily large length of quantum processing.

Using the result of (19), the photon information efficiency (PIE), i.e., the number of information bits transmitted per photon, achievable by the BPSK channel is plotted as a function of N in Fig. 2. When the average photon number transmitted per channel, $\mathcal{E}_{\mathsf{ch}}$, is 0.01, the inner product between

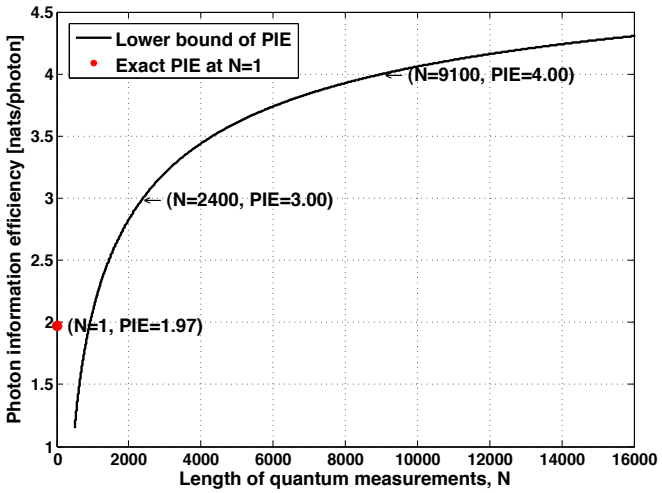

Fig. 2. Photon information efficiency of the BPSK channel, $C_N/(N\mathcal{E}_{\mathsf{ch}})$, at $\mathcal{E}_{\mathsf{ch}} = 0.01$ for the finite blocklength N.

the BPSK input states becomes $\gamma := |\langle\sqrt{\boldsymbol{\mathcal{E}_{\mathsf{ch}}}}|-\sqrt{\boldsymbol{\mathcal{E}_{\mathsf{ch}}}}\rangle| = \exp[-2\mathcal{E}_{\mathsf{ch}}] = e^{-0.02}$. By plugging γ into (6) and (7) and dividing the resulting capacities by $\mathcal{E}_{\mathsf{ch}}$, the exact PIE at an arbitrarily large N is 5.55 nats/photon and at $N = 1$, 1.97 nats/photon. Therefore, as N increases from 1 to ∞, the PIE of the BPSK channel should strictly increase from 1.97 to 5.55 nats/photon. From the lower bound of PIE in Fig. 2, it can be shown that at $N = 2400$, the PIE of 3.0 nats/photon is already achieved, and at $N = 9100$, 4.0 nats/photon. The lower bound is not tight in the regime of very small N, but it gets tighter as N increases, and approaches the ultimate limit of PIE as $N \to \infty$.

IV. Proof of Theorem 1

Theorem 1 will be proved based on two lemmas that will be introduced in this section. Note that in the definition of C_N in (2), sets of quantum measurements as well as input distributions over the length-N inner code need to be investigated to find the maximum mutual information of the superchannel. However, the channel distribution of the superchannel is hard to calculate exactly since it depends on the detailed structure of the inner code and the quantum measurements. Therefore, instead of calculating the exact distribution given a set of measurements and an inner code, a representative characteristic of the channel distribution $p_{k|j}^{(N)}$ is analyzed by calculating the average probability of error under uniform input distribution

$$p_e = e^{-NR} \sum_{j=1}^{e^{NR}} \sum_{k \neq j} p_{k|j}^{(N)}. \tag{21}$$

Among all the possible channel distributions $p_{k|j}^{(N)}$ that have the same average probability of error p_e, an *equierror superchannel* is defined with distribution $\overline{p}_{k|j}^{(N)}$ as

$$\overline{p}_{k|j}^{(N)} := \begin{cases} 1 - p_e, & k = j\,; \\ \left(e^{NR} - 1\right)^{-1} p_e, & k \neq j. \end{cases} \tag{22}$$

This channel is symmetric between all the inputs, and conditioned on an input j, all the possible outputs that result in an error, i.e., $k \neq j$, occur with the same probability. Due to the symmetry, the input distribution that maximizes the mutual information of this channel is uniform. The resulting maximum mutual information of the equierror superchannel is

$$\begin{aligned} \max_{p_j} I(p_j, \overline{p}_{k|j}^{(N)}) &= NR - p_e \log\left(e^{NR} - 1\right) - H_{\mathsf{B}}(p_e) \\ &> (1 - p_e)NR - \log 2 \end{aligned} \tag{23}$$

where $H_{\mathsf{B}}(p) = -p \log p - (1-p)\log(1-p)$.

It will be shown that the mutual information of the equierror channel is smaller than that of any other superchannel with the same average probability of error, p_e.

Lemma 2: For any $p_{k|j}^{(N)}$ with a fixed p_e in (21),

$$\max_{p_j} I(p_j, p_{k|j}^{(N)}) \geq \max_{p_j} I(p_j, \overline{p}_{k|j}^{(N)}) \tag{24}$$

for the equierror superchannel of $\overline{p}_{k|j}^{(N)}$.

Proof: For a random variable X that is uniformly distributed over e^{NR} inputs, and the conditional distribution $P_{Y|X}(k|j) := p_{k|j}^{(N)}$,

$$\max_{p_j} I(p_j, p_{k|j}^{(N)}) \geq I(X;Y) = NR - H(X|Y). \tag{25}$$

From Fano's inequality, we have

$$\begin{aligned} H(X|Y) \leq & H_B\left(\Pr(X \neq Y)\right) + \Pr(X \neq Y)\log(e^{NR} - 1) \\ = & H_B(p_e) + p_e \log(e^{NR} - 1). \end{aligned}$$

By combining the above two inequalities,

$$\begin{aligned} \max_{p_j} I(p_j, p_{k|j}^{(N)}) \geq & NR - p_e \log(e^{NR} - 1) - H_B(p_e) \\ = & \max_{p_j} I(p_j, \overline{p}_{k|j}^{(N)}). \end{aligned} \tag{26}$$

■

Then, by the definition of C_N and Lemma 2, when there exists an inner code of length N and rate R that can be decoded by a set of length N measurements with average error probability p_e,

$$\frac{C_N}{N} \geq \max_{p_j} \frac{I(p_j, p_{k|j}^{(N)})}{N} > (1 - p_e)R - \frac{\log 2}{N}. \tag{27}$$

Holevo in [3] showed an upper bound of p_e for a code of length N and rate R, which is summarized in the following lemma.

Lemma 3: [Holevo] *For a set of input states $\{|\psi_x\rangle\}$, there exists a block code of length N and rate R that can be decoded by a set of measurements with the average probability of error satisfying*

$$p_e \leq 2\exp[-NE(R)] \tag{28}$$

where, for $\rho = \sum_x P_X(x)|\psi_x\rangle\langle\psi_x|$,

$$E(R) = \max_{0 \leq s \leq 1} \left[\max_{P_X} \left(-\log \mathrm{Tr}\left(\rho^{1+s}\right)\right) - sR \right]. \tag{29}$$

By combining this lemma with (27), Theorem 1 is proven.

V. Interpretation of the Superadditivity: Classical DMC vs. Quantum Channel

Strict superadditivity of C_N, i.e., $C_N + C_M < C_{N+M}$, is demonstrated for binary input quantum channels with and without the energy constraint in Section II. The lower bound of C_N/N for a fixed N is provided in Theorem 1, which makes it possible to understand the trade-off between the maximum achievable information rate and the complexity of quantum processing as N increases. Previously, the superadditivity of C_N has been mostly thought of as a unique and attractive property that can be observed in quantum channels, but not in classical DMCs. The interpretation of this phenomenon is that a set of length-N *entangled* quantum measurements can generate some memory in the superchannel, which can increase the number of information bits extractable per quantum state, as a longer block of quantum states is measured jointly. However, this viewpoint does not provide enough insights to fully understand and analyze the phenomenon. In this section, a new aspect to understanding strict superadditivity of C_N will be introduced by comparing performance of the concatenated codes over quantum channels and classical DMCs for a fixed inner code blocklength N.

Fig. 3 illustrates the concatenated codes over a classical DMC. Compared to Fig. 1, the quantum channel is replaced with the classical DMC $P_{Y|X}$, and in the position of a quantum detector, an inner decoder is placed. In [8], the concatenated code over a classical DMC is introduced with consideration of the trade-off between performance and complexity of communication over the classical DMC. In [8], the performance is analyzed by the *error exponent* achievable with the concatenated codes; also examined is how the decoding complexity of the concatenated codes increases as the overall length of the concatenated code $N_c = nN$ increases. It is obvious that when the inner decoder generates a sufficient statistics of the channel output and forwards it to the outer decoder, there is no loss of information, so that the performance of the concatenated codes can be as good as an optimum code, even with the restricted structure of concatenation. However, for this case, the complexity of the decoding increases exponentially with the overall length of the code. In [8], it is shown that even if there is a loss of information at the inner decoder by making a hard-decision on the message of the inner code, as the inner code blocklength N goes to infinity, the capacity of the classical DMC can be achieved with the concatenated code. Moreover, the overall complexity of the decoding algorithm is significantly reduced to be almost linear with the length of the concatenated code. The loss of information at the inner decoder, however, degrades the achievable error exponent over all rates below the capacity.

This result can be proved by analyzing a lower bound on the performance of the concatenated codes over the classical DMC. To get the lower bound, the equierror superchannel defined in (22), whose mutual information is smaller than that of any other superchannel with the same p_e, is used. The average probability of error p_e from decoding at the

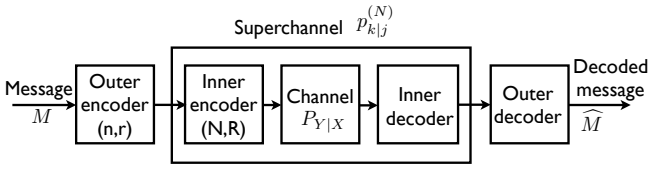

Fig. 3. Concatenated codes over classical DMC

inner decoder can be analyzed by using the error exponent of the classical DMC $P_{Y|X}$ in [9]. It is shown that an optimum inner code with the minimum decoding error can achieve p_e as low as

$$p_e = \exp[-N(E(R) + o(1))] \tag{30}$$

as $N \to \infty$, when

$$E(R) = \max_{0 \le s \le 1} \left(\max_{P_X} (E_0(s, P_X)) - sR \right) \tag{31}$$

with

$$E_0(s, P_X) := -\log \sum_y \left[\sum_x P_X(x) P_{Y|X}(y|x)^{1/(1+s)} \right]^{1+s}. \tag{32}$$

By using the p_e in (30) and analyzing the performance of the equierror channel, it can be shown that the capacity of the DMC, which is $C = \max_{P_X} I(P_X, P_{Y|X})$, can be achievable by the concatenated code *as both the inner code blocklength N and the outer code blocklength n go to infinity*, even when the inner decoder makes the hard-decision for the estimate of the inner code message, and discards all the rest of the information about the channel output.

Let us clarify the difference between the concatenated code over the classical DMC and that over the quantum channel. When a likelihood detector is used at the classical inner decoder, after decoding the most likely codeword given a received channel output, the classical inner decoder can still have the information about which codeword is the second mostly likely one, and how much less likely it is compared to the first one, etc. On the contrary, for the quantum channel, once the quantum states are measured by the quantum detector, it certainly results in a loss of information since after the measurement, the quantum states are changed and the information that was encoded in the quantum states is destroyed. Therefore, different from the classical inner decoder, which has an option to maintain a sufficient statistics of the channel outputs with the cost of complexity, a loss of information at the quantum detector is not avoidable. For the quantum channel, the trade-off between the achievable information rate and the complexity of quantum processing can be analyzed by observing how C_N/N increases with N. In contrast, for the concatenated code over the classical DMC, the trade-off between performance and complexity is analyzed by assuming a certain type of loss of information at the inner decoder that makes the decoding complexity increase almost linearly with the overall blocklength of the code, and by showing how the error exponent of the overall code is degraded by the loss of information at the inner

decoder, *under the assumption of a large enough inner code blocklength* N.

Now, we ask a new question for the concatenated code over the classical DMC, similar to that asked for the quantum channel: when the inner decoder makes an estimate for the message of the inner code *at a finite blocklength* N, how does the maximum achievable information rate of the concatenated code increase as N increases?

For the inner decoder that makes the hard-decision at a finite blocklength N of the inner code, the maximum achievable information rate by the concatenated code is C_N/N where

$$C_N = \max_{p_j} \max_{\{\text{Length N inner code-decoder}\}} I(p_j, p_{k|j}^{(N)}) \tag{33}$$

for the superchannel distribution $p_{k|j}^{(N)}$, which is determined by the decoding algorithm, given an inner code. By using Lemma 2, it can also be shown that when there exists a code of length N and rate R whose probability of decoding error is p_e,

$$\frac{C_N}{N} > (1-p_e)R - \frac{\log 2}{N}. \tag{34}$$

Moreover, in [9], it is shown that for the classical DMC $P_{Y|X}$, there exists a code of length N and rate R whose probability of error p_e is bounded by

$$p_e \leq \exp[-NE(R)] \tag{35}$$

with $E(R)$ in (31). By combining (34) and (35), the following theorem can be proved for the maximum achievable information rate of the concatenated codes over the classical DMC at a finite N.

Theorem 2: With a fixed inner code blocklength N*,*

$$\frac{C_N}{N} \geq \max_R \left(\left(1 - e^{-NE(R)}\right) R - \frac{\log 2}{N} \right) \tag{36}$$

with $E(R)$ *in (31).*

Note that the lower bound of C_N/N in (36) strictly increases with N, and it has exactly the same form as that for the quantum channel in (17) except the difference in $E(R)$ and a constant multiplied to $e^{-NE(R)}$. As a result, we can observe a phenomenon similar to the superadditivity of C_N in the quantum channel, even in the classical DMC when the inner decoder makes the hard-decision at a finite blocklength. The reason why C_N is away from the capacity of the channel C for a finite inner code blocklength N is because the hard-decision at the inner decoder results in a significant amount of loss of information that even hurts the rate of the communication. Moreover, as N increases, the quality of the hard-decision is improved, which makes it possible to achieve a higher information rate. Therefore, the superadditivity of C_N can now be interpreted as a degradation of the performance by the loss of information at the inner decoder that makes the hard-decision at a finite blocklength. This new aspect can also be applied to explain the same phenomenon observed in the quantum channel by replacing the role of inner decoder with the quantum detector, which also makes a hard-decision on the block of quantum states with a finite blocklength.

In the rest of this section, we will simplify the lower bound of C_N by finding an approximation of the error exponent $E(R)$ for the quantum channel and for the classical DMC. Using the simplified lower bound, it will be possible to compare the quantum channel and the classical channel by calculating the inner code blocklength N required to achieve a given fraction of the ultimate capacity of each channel. To avoid confusion, from this point on, a function for the quantum channel will be written with a superscript (q) and that for the classical DMC with a superscript (c), for example, $E^{(q)}(R)$ and $E^{(c)}(R)$.

The error exponent of the classical DMC, $E^{(c)}(R)$ in (31), can be approximated by the Taylor expansion at the rate R close to the capacity C as

$$E^{(c)}(R) = \frac{1}{2V^{(c)}}(R-C)^2 + O\left((R-C)^3\right) \tag{37}$$

with a parameter $V^{(c)}$

$$V^{(c)} = \sum_{x,y} p_x p_{y|x} \left[\left(\log \frac{p_{y|x}}{p_y} - \sum_{x,y} p_x p_{y|x} \log \frac{p_{y|x}}{p_y} \right)^2 \right] \tag{38}$$

for the capacity achieving input distribution $p_x := P_X^*(x)$ and the corresponding output distribution $p_y := P_Y^*(y)$ according to the channel $p_{y|x} := P_{Y|X}(y|x)$. In (38), $V^{(c)}$ is the variance of $\log(p_{y|x}/p_y)$ under the distribution $p_x p_{y|x}$, and is termed the *channel dispersion* in [10].

Similarly, the error exponent of the quantum channel, $E^{(q)}(R)$ in (18), can be approximated with a parameter $V^{(q)}$, which is a characteristic of the quantum channel similar to the channel dispersion of the classical DMC. The definition of $V^{(q)}$ depends on the density operator ρ, which fully characterizes the capacity of the quantum channel. For a set of input states $\{|\psi_x\rangle\}$, when P_X^* is the optimum input distribution that maximizes the capacity of the quantum channel $C = \text{Tr}(-\rho \log \rho)$ where $\rho = \sum_x P_X(x)|\psi_x\rangle\langle\psi_x|$, the parameter $V^{(q)}$ is defined by the eigenvalues of the density operator ρ at $P_X = P_X^*$. Let us denote the eigenvalues of ρ by σ_i, $i = 1, \ldots, J$ where J is the dimension of the space spanned by the input states $\{|\psi_x\rangle\}$. From the fact that ρ is a positive operator and $\text{Tr}(\rho) = 1$, it can be shown that each $\sigma_i \geq 0$ for all i and $\sum_{i=1}^J \sigma_i = 1$. Then, $V^{(q)}$ is defined as a variance of the random variable $-\log\sigma$ where $\sigma \in \{\sigma_i\}$ with probability distribution $[\sigma_1, \ldots, \sigma_J]$, i.e.,

$$V^{(q)} = \sum_{i=1}^J \sigma_i (-\log \sigma_i)^2 - \left(\sum_{i=1}^J \sigma_i \left(-\log \sigma_i\right) \right)^2. \tag{39}$$

By the Taylor expansion of $E^{(q)}(R)$ in (18) at the rate R close to C, it can be shown that

$$E^{(q)}(R) = \frac{1}{2V^{(q)}}(R-C)^2 + O\left((R-C)^3\right). \tag{40}$$

Therefore, both the error exponent of the classical DMC and that of the quantum channel can be approximated as a quadratic term in the rate R with the quadratic coefficient

inversely proportional to the dispersion of the channel. Since the lower bound of C_N as well as the approximated error exponent $E(R)$ have the same forms both for the classical DMC and for the quantum channel, it is possible to compare the classical DMC and the quantum channel by a common simplified lower bound of C_N, which can be written with the parameter V and C as follows.

Theorem 3: For both the classical DMC and the quantum channel, when the channel dispersion V and the capacity C satisfy i) $\sqrt{\frac{V}{NC^2}} \to 0$ as $N \to \infty$ and ii) $V \cdot C$ is finite, the maximum achievable information rate at the inner code blocklength N is lower bounded by

$$\begin{aligned}\frac{C_N}{N} \geq & C \cdot \left(1 - \sqrt{\frac{V}{NC^2} \log\left(\frac{NC^2}{V}\right)}\right) - \frac{\log 2}{N} \\ & + O\left(\sqrt{\frac{V}{N \log \frac{NC^2}{V}}} \log\log\left(\frac{NC^2}{V}\right)\right).\end{aligned} \tag{41}$$

Proof: The quadratic approximation of $E(R)$ will be used to find a simplified form for a lower bound of C_N/N. Both for the quantum channel and the classical channel, C_N is lower bounded by

$$\frac{C_N}{N} \geq \max_R \left(\left(1 - 2e^{-NE(R)}\right) R - \frac{\log 2}{N}\right), \tag{42}$$

from Theorems 1 and 2. Then, for a fixed R^*

$$R^* = C \cdot \left(1 - \sqrt{\frac{V}{NC^2} \log\left(\frac{NC^2}{V} \log \frac{NC^2}{V}\right)}\right), \tag{43}$$

whose derivation is omitted in this paper, the approximated error exponent at R^* is

$$\begin{aligned}E(R^*) = & \frac{1}{2N} \log\left(\frac{NC^2}{V} \log \frac{NC^2}{V}\right) \\ & + O\left(\frac{VC}{N}\sqrt{\frac{V}{NC^2}}\left(\log\left(\frac{NC^2}{V}\log\frac{NC^2}{V}\right)\right)^{3/2}\right).\end{aligned} \tag{44}$$

It can be checked that under the assumptions of i) $\sqrt{\frac{V}{NC^2}} \to 0$ and ii) $V \cdot C$ is finite, the term in $O(\cdot)$ in (44) approaches 0, which results in

$$e^{-NE(R^*)} = \sqrt{\frac{V}{NC^2 \log \frac{NC^2}{V}}}(1 + o(1)). \tag{45}$$

By plugging (43) and (45) into the lower bound (42), C_N/N can be bounded as (41). ∎

Remark 1: From the lower bound of Theorem 3, we can see that the inner code blocklength N at which the lower bound is equal to a given fraction of the capacity is proportional to V/C^2.

Since the same bound on C_N/N as in (41) holds both for the quantum and the classical channel, by the parameter V/C^2, we can compare behavior of the quantum channel and the classical DMC. For the BPSK quantum channel, by using the two eigenvalues of ρ at P_X^*, which are $\sigma_1 = (1 - e^{-2\mathcal{E}_{\mathsf{ch}}})/2$ and $\sigma_2 = (1 + e^{-2\mathcal{E}_{\mathsf{ch}}})/2$, the channel dispersion and the capacity of the channel can be calculated as

$$\begin{aligned}V_{\mathsf{BPSK}} = & \mathcal{E}_{\mathsf{ch}} \left(\log \frac{1}{\mathcal{E}_{\mathsf{ch}}}\right)^2 (1 + O(\mathcal{E}_{\mathsf{ch}})) \\ C_{\mathsf{BPSK}} = & \mathcal{E}_{\mathsf{ch}} \log \frac{1}{\mathcal{E}_{\mathsf{ch}}} + \mathcal{E}_{\mathsf{ch}} + o(\mathcal{E}_{\mathsf{ch}}).\end{aligned} \tag{46}$$

Then, $V_{\mathsf{BPSK}}/C_{\mathsf{BPSK}}^2 \approx 1/\mathcal{E}_{\mathsf{ch}}$ for the low photon number regime where $\mathcal{E}_{\mathsf{ch}} \to 0$. For the classical AWGN channel in the low-power regime where $\mathsf{SNR} \to 0$, $V_{\mathsf{AWGN}}/C_{\mathsf{AWGN}}^2$ can be calculated by using the result of [10], and it is $4/\mathsf{SNR}$. For both of the channels, V/C^2 is inversely proportional to the energy to transmit the information per channel use. This means that as the energy decreases, in order to make the lower bound to meet a targeted fraction of capacity, it is necessary to adopt a longer inner code.

VI. Conclusion

The superadditivity of the quantum channel is analyzed with consideration of the trade-off between performance and complexity for communication over the quantum channels. The concatenated code with an inner code of a finite blocklength N is adopted to observe the trade-off, and a lower bound on the maximum achievable information rate is provided as a function of N. We can observe a similar phenomenon even in the classical DMC, and can understand that the superadditivity is caused by a loss of the information from the hard-decision at the inner decoder with a finite blocklength N. Under this framework, the classical DMC and the quantum channel can be compared with a parameter V/C^2, which is proportional to the N sufficient to achieve a given fraction of the capacity.

References

[1] P. Hausladen, R. Jozsa, B. Schumacher, M. Westmoreland, and W. K. Wootters, "Classical information capacity of a quantum channel," *Phys. Rev. A*, vol. 54, pp. 1869–1876, Sep 1996.

[2] A. Holevo, "The capacity of the quantum channel with general signal states," *Information Theory, IEEE Transactions on*, vol. 44, no. 1, pp. 269–273, 1998.

[3] A. S. Holevo, "Coding theorems for quantum channels," *arXiv preprint quant-ph/9809023*, 1998.

[4] A. Peres and W. K. Wootters, "Optimal detection of quantum information," *Physical Review Letters*, vol. 66, no. 9, pp. 1119–1122, 1991.

[5] M. Sasaki, K. Kato, M. Izutsu, and O. Hirota, "A demonstration of superadditivity in the classical capacity of a quantum channel," *Physics Letters A*, vol. 236, no. 1, pp. 1–4, 1997.

[6] V. Giovannetti, S. Guha, S. Lloyd, L. Maccone, J. H. Shapiro, and H. P. Yuen, "Classical capacity of the lossy bosonic channel: The exact solution," *Physical review letters*, vol. 92, no. 2, p. 027902, 2004.

[7] H. W. Chung, S. Guha, and L. Zheng, "On capacity of optical channels with coherent detection," in *Information Theory Proceedings (ISIT), 2011 IEEE International Symposium on*. IEEE, 2011, pp. 284–288.

[8] G. D. Forney and G. D. Forney, *Concatenated codes*. Citeseer, 1966, vol. 11.

[9] R. G. Gallager, "Information theory and reliable communication," 1968.

[10] Y. Polyanskiy, H. V. Poor, and S. Verdú, "Channel coding rate in the finite blocklength regime," *Information Theory, IEEE Transactions on*, vol. 56, no. 5, pp. 2307–2359, 2010.