Superadditivity of Quantum Channel Coding Rate with Finite Blocklength Quantum Measurements

Hye Won Chung Saikat Guha Lizhong Zheng

Abstract—We investigate superadditivity in the maximum achievable rate of reliable classical communication over a quantum channel with a general pure-state alphabet. We define superadditivity as the phenomenon that the maximum accessible information per channel use strictly increases as the number of channel outputs jointly measured at the receiver increases. We analyze this trade-off between information rate and receiver complexity by considering the capacity of the classical discrete memoryless superchannel induced under a concatenated coding scheme, where the quantum joint detection measurement acts exclusively on the length-$N$ inner code words, but unfettered from any assumption on the complexity of the classical outer code. We give a general lower bound on the maximum accessible information per channel use for a finite-length joint measurement, and we further refine the bound purely in terms of $V$, the quantum version of channel dispersion, and $C$, the capacity of the classical-quantum channel. We also observe a similar phenomenon of superadditivity even in the channel capacity of a classical discrete memoryless channel (DMC) in a concatenated coding scheme, due to a loss of information from hard-decisions by the inner decoder over a finite blocklength, $N$. Finally, we develop a unifying framework, within which the superadditivity in capacity of the classical DMC and that of a pure-state classical-quantum channel can both be expressed with a parameter $V/C^2$, a quantity that we show is proportional to the inner-code measurement length $N$ that is sufficient to achieve a given fraction $0 < \alpha \leq 1$ of the capacity $C$.

I. BACKGROUND

How many classical bits per channel use can be reliably communicated over a quantum channel? This has been a central question in quantum information theory in an effort to understand the intrinsic limit on the classical capacity of physical quantum channels such as the optical fiber or free-space optical channels. The Holevo limit of a quantum channel is an upper bound to the Shannon capacity of the classical channel induced by pairing the quantum channel with any specific receiver measurement \cite{Holevo, Holevo2}. The Holevo limit is in principle also an achievable information rate, and is known for several important practical channels, such as the lossy bosonic channel \cite{LossyBosonic}. However, a receiver that attains the Holevo capacity, must in general make joint (collective) measurements over long codeword blocks. Such measurements cannot be realized by detecting single modulation symbols followed by classical post processing. We call this phenomenon of a joint-detection receiver (JDR) being able to yield a higher information rate (in error-free bits communicated per channel use) than what is possible by any single-symbol receiver measurement, as superadditivity of capacity.\footnote{There are several JDR measurements that are known to achieve the Holevo capacity—the square-root measurement (SRM) \cite{SRM}, Helstrom’s minimum probability of error measurement \cite{Helstrom}, the sequential-decoding measurement \cite{SequentialDecoding}, the successive-cancellation decoder for the quantum polar code \cite{PolarCode}, \cite{PolarCode2}, and a two-stage near-unambiguous-detection receiver \cite{NearUnambiguous}. There are a few characteristics that are common to each one of these measurements. First, the size of the joint-detection measurement is tied to the blocklength of the code, i.e., the measurement must act on the entire codeword and hence its size must increase with the length of the codeword. Second, none of these measurement specifications translate readily to a realizable receiver in the context of optical communication. Since it is known that a simple laser-light (coherent-state) modulation achieves the Holevo capacity of the lossy bosonic channel \cite{LossyBosonic}, almost all the complexity in achieving the ultimate limit to the reliable communication rate lies at the receiver. Finally, none of these capacity-achieving measurements tell us how the achievable information rate increases with the size of the receiver measurement. Since the complexity of implementing a joint quantum measurement over $N$ channel symbols in general grows exponentially with $N$, it is of great practical interest to find how the maximum achievable information rate (error-free bits per channel use) scales with the size $N$ of the joint-detection receiver (while imposing no constraint whatsoever on the classical code complexity). In this paper, we shed some light on this, for classical communication over a quantum channel using a pure-state alphabet—the so-called pure-state classical-quantum (cq) channel (the lossy bosonic channel being an example)—by proving a general lower bound on the finite-measurement-length capacity.

Finally, we would like to remark on an important difference between our results in this paper and the finite-
The classical 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 at the receiver with a set of quantum measurements as the number of transmissions $N_r$ goes to infinity. Consider a pure-state classical-quantum (cq) channel $W : x \rightarrow |\psi_x\rangle$, where $x \in X$ is the classical input, and $\{|\psi_x\rangle\} \in \mathcal{H}$ are corresponding modulation symbols at the output of the channel. One practical example of a pure-state cq channel is the single-mode lossy optical channel $\eta : \alpha \rightarrow |\sqrt{\eta}\alpha\rangle$, where $\alpha \in \mathbb{C}$ is the complex field amplitude at the input of the channel, $\eta \in (0, 1]$ is the transmissivity (the fraction of input power that appears at the output), and $|\sqrt{\eta}\alpha\rangle$ is the quantum description of an ideal laser-light pulse, a coherent state $|\alpha\rangle$.

In refs. [1], [2], it was shown that the classical capacity of a cq channel $W$ is given by

$$C = \max_{P_X} \text{Tr}(-\rho \log \rho),$$

where $\rho = \sum_{x \in X} P_X(x)|\psi_x\rangle\langle\psi_x|$. The states $|\psi_x\rangle$, $x \in X$, are normalized vectors in a complex Hilbert space $\mathcal{H}$, $\langle\psi_x|$ is the Hermitian conjugate vector of $|\psi_x\rangle$, and $\rho$ is a density operator, a linear combination of the outer products $|\psi_x\rangle\langle\psi_x|$ with weights $P_X(x)$. The capacity can also be written as $C = \max_{P_X} S(\rho)$, where $S(\rho) = \text{Tr}(-\rho \log \rho)$ is the von Neumann entropy of the density operator $\rho$.

For an input codeword $[x_1, \cdots, x_{N_c}]$, the received codeword is a tensor product state, $|\psi_{x_1}\rangle \otimes \cdots \otimes |\psi_{x_{N_c}}\rangle$, which is jointly detected by an orthogonal projective measurement in the $N_c$-fold Hilbert space $\mathcal{H}^\otimes N_c$. When the received codeword is projected into the orthogonal measurement vectors $\{|\Phi_k\rangle\}$, $k \in K$, which resolves the identity, i.e., $\sum_k |\Phi_k\rangle\langle\Phi_k| = 1$, in $\mathcal{H}^\otimes N_c$, the classical output $k$ is observed with probability equal to the magnitude squared of the inner product between the received codeword state and the measurement vector $|\Phi_k\rangle$ corresponding to the output $k$. The orthogonal projective measurement is designed to decode the received codewords with as small error probability as possible. For any rate $R < C$, a block code of length $N_c$ and rate $R$, generated by picking each of the $N_c$ symbols of $\rho^{N_cR}$ from the distribution $P_X$ that attains the maximum in Eq. (1), paired with an appropriate quantum measurement acting jointly on the received codeword in $\mathcal{H}^\otimes N_c$, can attain an arbitrarily small probability of error as $N_c \to \infty$ [1], [2], [7], [11].

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 $e^{N_cR}$ possible messages. The number of measurement outcomes thus scales exponentially with the length of the codeword $N_c$. Hence, the complexity of physical implementation (in terms of number of elementary finite-length quantum operations) of the receiver in general also grows exponentially in $N_c$. Considering this exponential growth in complexity, one might want to limit the maximum length $N \leq N_c$ of the sequence of states to be jointly detected at the receiver, independent of the length of the codeword $N_c$. However, there is no guarantee that such quantum measurements of fixed blocklengths can still achieve the ultimate capacity of the quantum channel.

Our goal in this paper is to study the trade-off between information rate and receiver complexity, for classical communication over a quantum channel. This paper investigates the maximum number of classical information bits that can be reliably decoded per use of the quantum channel, when quantum states of a finite blocklength $N \leq N_c$ are jointly measured, with no restriction on the complexity of the overall classical code ($N_c \to \infty$). As the number of channel outputs $N$ jointly measured increases, the maximum number of classical bits extracted per use of the channel increases.

We call this phenomenon superadditivity of the maximum achievable classical information rate over 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.

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 a total of $e^{NR}$ inputs to the inner encoder, $J \in \{1, \ldots, e^{NR}\}$. The inner encoder maps each input $J$ to a length-$N$ classical codeword, which maps to a length-$N$ sequence of quantum states at the output of the channel. The quantum joint-detection receiver measures the length-$N$ quantum codeword and generates an estimate $K \in \{1, \ldots, e^{NR}\}$ of the encoded message $J$. For a good inner code and joint measurement with $N$ large, the estimate would generally match the input message. But...
for a fixed $N$, the error probability may not be close to 0. The inner encoder, the classical-quantum channel, and the inner decoder (the joint detection receiver), collectively forms a discrete memoryless superchannel $\text{Superchannel}$, with transition probabilities $p_{k|j}(N) := \Pr(K = k | J = j)$. We define the maximum mutual information of this superchannel, over all choices of inner codes of blocklength $N$, and over all choices of inner-decoder joint measurements of length $N$ as:

$$C_N := \max_{p_j} \max \{N\text{-symbol inner code-measurement pairs}\} I(p_j, p_{k|j}^{(N)}).$$

(2)

A classical Shannon-capacity-achieving 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 $e^{nr} = e^{n(r/N)}$. 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 made arbitrarily small as $n \to \infty$. Therefore, the maximum information rate achievable by the concatenated code per use of the 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. This implies the existence of the limit $\lim_{N \to \infty} C_N/N$. Holevo [18] showed that the limit is equal to the ultimate capacity of the quantum channel, $\lim_{N \to \infty} C_N/N = C = \max_x S(p_x)$. 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 quantum processing occurs only at the quantum decoder 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 (rate) performance and the (quantum) complexity can be captured by how fast $C_N/N$ increases with $N$. It is known that for some examples of input states, strict superadditivity of $C_N$ can be demonstrated [18], [19], [20]. However, the calculation of $C_N$, even for a pure-state binary alphabet, is extremely hard for $N > 1$ because the complexity of optimization increases exponentially with $N$. The concatenated coding scheme described above was considered in [21] for the lossy bosonic channel, where some examples of inner codes and structured optical joint-detection receivers were found for which $C_N/N > C_1$ holds for a binary coherent-state modulation alphabet.

Instead of aiming to calculate the exact $C_N$, in this paper, a lower bound on $C_N/N$, which becomes tight for large enough $N$, will be derived. From this 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 [22].

The rest of the paper is outlined as follows. In Section III, examples of quantum channels where strict superadditivity $C_1 < C$ holds, will be demonstrated. The main theorem that states a lower bound on $C_N/N$ will be given in Section IV 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 V. Thereafter in Section VI, the effect of superadditivity due to finite-blocklength inner-code measurements in a concatenated coding architecture, will be compared between a quantum channel and a classical discrete memoryless channel. An approximation of the lower bound of $C_N/N$ will also be provided by introducing a quantum version of channel dispersion $V$, with a unifying picture encompassing quantum and classical channels. Section VII will conclude the paper.

III. STRICT SUPERADDITIVITY OF $C_N$

Before investigating how $C_N/N$ increases with $N$, we will show examples where strict superadditivity of $C_N$ can be shown, i.e., $C_1 < C$. As discussed before, given a set of input states $\{|\psi_x\}\}, x \in \mathcal{X}$, $C$ can be calculated from Holevo’s result by finding the optimum input distribution that maximizes the von Neumann entropy $S(\rho) = \text{Tr}(\rho \log \rho)$ where $\rho = \sum_x P_x(x) |\psi_x\rangle \langle \psi_x|$. Calculating $C_1$ on the other hand requires finding a set of measurements as well as an input distribution to maximize the resulting mutual information, where the measurement acts on one channel symbol at a time. For general input states, this is a hard optimization problem, since the measurement that maximizes $C_1$ may not be a projective measurement, and could be a Positive Operator Valued Measure (POVM)—the most general description of a quantum measurement—and furthermore the optimum POVM could have up to $|\mathcal{X}|(|\mathcal{X}| + 1)/2$ outcomes [15], [16], [17]. However, for binary pure states $\{|\psi_0\rangle, |\psi_1\rangle\}$ of inner product $|\langle \psi_0 | \psi_1 \rangle| = \gamma$, $C_1$ and $C$ can be calculated as simple functions of the inner product $\gamma = |\langle \psi_0 | \psi_1 \rangle|$, and strict superadditivity can be shown, as summarized below [18].

The first step to calculate $C$ for the binary input channel is to find the eigenvalues of $\rho$ 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 $\rho$ have a form of $|\psi_0\rangle + \beta |\psi_1\rangle$ with some $\beta$ that satisfies

$$\rho(|\psi_0\rangle + \beta |\psi_1\rangle) = \sigma(|\psi_0\rangle + \beta |\psi_1\rangle)$$

(3)
with eigenvalues $\sigma$. By solving the equation, we obtain the two eigenvalues as:

$$\sigma_1 = \frac{1}{2} \left( 1 - \sqrt{1 - 4q(1-q)(1-\gamma^2)} \right), \text{ and}$$

$$\sigma_2 = \frac{1}{2} \left( 1 + \sqrt{1 - 4q(1-q)(1-\gamma^2)} \right),$$

and the resulting von Neumann entropy,

$$S(\rho) = \text{Tr}(\rho \log \rho) = -\sigma_1 \log \sigma_1 - \sigma_2 \log \sigma_2,$$

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,

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

For the binary channel, $C_1$ is attained by the equiprior input distribution and a binary-valued projective measurement in the span of $\{|\psi_0\rangle, |\psi_1\rangle\}$—the same measurement that minimizes the average probability of discriminating between equally-likely states $|\psi_0\rangle$ and $|\psi_1\rangle$. The derivation of $C_1$ for the binary case can be found in [18], and is given by:

$$C_1 = \frac{1 - \sqrt{1 - \gamma^2}}{2} \log \left( 1 - \sqrt{1 - \gamma^2} \right) + \frac{1 + \sqrt{1 - \gamma^2}}{2} \log \left( 1 + \sqrt{1 - \gamma^2} \right).$$

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, we will consider the superadditivity of $C_N$ in quantum channels with an input constraint, in the context of optical communication. The constraint will be the mean energy of input states. A coherent state $|\alpha\rangle$ is the quantum description of a single spatio-temporal-polarization mode of a classical optical-frequency electromagnetic (ideal laser-light) field, where $\alpha \in \mathbb{C}$ is the complex amplitude, and $|\alpha|^2$ is the mean photon number of the mode. Since the energy of a photon with angular frequency $\omega$ is $E = h\omega$ with $h = h/2\pi$ where the Planck constant $h = 6.63 \times 10^{-34}\text{m}^2\text{kg/s}$, the average energy (in Watts) of the coherent state $|\alpha\rangle$ of a quasi-monochromatic field mode of center frequency $\omega$, is $h|\alpha|^2\omega$. Note that the mean photon number $|\alpha|^2$ is a dimensionless quantity. Therefore, for quasi-monochromatic propagation at a fixed center frequency $\omega$, an average energy constraint on the input states (or equivalently, an average power constraint with a fixed time-slot width) can be represented as a constraint on the average photon number per transmitted state. For example, if the modulation constellation comprises of the set of input states $\{|\alpha_1\rangle, |\alpha_2\rangle, \ldots, |\alpha_K\rangle\}$, an average energy constraint $h\omega\mathcal{E}$ per symbol transmission can be expressed as a constraint on the prior distribution $\{p_1, \ldots, p_K\}$, with

$$\sum_{i=1}^{K} p_i |\alpha_i|^2 \leq \mathcal{E},$$

where $\mathcal{E}$ is the maximum mean photon number per symbol.

The important question of how many bits can be reliably communicated per use (i.e., per transmitted mode) of a pure-loss optical channel of power transmissivity $\eta \in (0, 1)$, under the constraint on the average photon number per transmitted mode $\mathcal{E}$, was answered in [13]. It was also shown that product coherent state inputs are sufficient to achieve the Holevo capacity of this quantum channel. Since a coherent state $|\alpha\rangle$ of mean photon number $\mathcal{E} = |\alpha|^2$ transforms into another coherent state $|\sqrt{\eta}|\alpha\rangle$ of mean photon number $\eta\mathcal{E}$ over the lossy channel, we will henceforth, without loss of generality, subsume the channel loss in the energy constraint, and pretend that we have a lossless channel ($\eta = 1$) with a mean photon number constraint $\mathbb{E}[|\alpha|^2] \leq \mathcal{E}$ per mode (or per ‘channel use’). The capacity of this channel is given by [3]

$$C(\mathcal{E}) = (1 + \mathcal{E}) \log(1 + \mathcal{E}) - \mathcal{E} \log \mathcal{E} \text{ [nats/mode]},$$

and it is achievable with a coherent-state random code with the amplitude $\alpha$ chosen from a circulo-complex Gaussian distribution with variance $\mathcal{E}$, $p(\alpha) = \exp[-|\alpha|^2/\mathcal{E}]/(\pi\mathcal{E})$.

The number of information bits that can be reliably communicated per photon—the photon information efficiency (PIE)—under a mean photon number constraint per mode, $\mathcal{E}$, is given by $C(\mathcal{E})/\mathcal{E}$ (nats/photon). From [8], it can be shown that in order to achieve high PIE, $\mathcal{E}$ must be small. In the $\mathcal{E} \to 0$ regime, the capacity (9) can be approximated as

$$C(\mathcal{E}) = \mathcal{E} \log \frac{1}{\mathcal{E}} + \mathcal{E} + o(\mathcal{E}),$$

which shown that PIE $\sim -\log \mathcal{E}$ for $\mathcal{E} \ll 1$. Thus there is no upper limit in principle to the photon information efficiency.

We will now show that in the high-PIE (low photon number) regime, this ultimate capacity is achievable closely even with a simple Binary Phase Shift Keying (BPSK) coherent state constellation $\{|\sqrt{\mathcal{E}}\rangle, |\sqrt{\mathcal{E}}\rangle\}$, which satisfies the energy constraint with any prior distribution. The inner product between the two states, $\gamma = |\langle \alpha | \beta \rangle| = \exp[-|\alpha - \beta|^2/2]$. Therefore,

$$|\langle \sqrt{\mathcal{E}} | \sqrt{\mathcal{E}} \rangle| = \exp[-2\mathcal{E}].$$

By plugging $\gamma$ into (6), we obtain the capacity of the BPSK input constellation,

$$C_{\text{BPSK}}(\mathcal{E}) = \mathcal{E} \log \frac{1}{\mathcal{E}} + \mathcal{E} + o(\mathcal{E}),$$

which is equal to $C(\mathcal{E})$ for the first- and second-order terms in the limit $\mathcal{E} \to 0$.

We now ask, for binary coherent-state inputs under the same constraint on the mean photon number per mode $\mathcal{E}$, how high an information rate is achievable when each mode is detected one-by-one, i.e., $N = 1$. The maximum capacity of the bosonic channel with a binary-input with mean photon number constraint $\mathcal{E}$, and a $N = 1$ measurement, will be denoted as $C_{1, \text{Binary}}(\mathcal{E})$. For BPSK input states, by using (7), the maximum achievable rate at $N = 1$ is

$$C_{1, \text{BPSK}}(\mathcal{E}) = 2\mathcal{E} + o(\mathcal{E}).$$

Thus, PIE of the BPSK channel caps off at 2 nats/photon for $N = 1$, while for $N$ large, achievable PIE $\to \infty$ as $\mathcal{E} \to 0$. 
$C_{1, \text{Binary}}(\mathcal{E})$ can be calculated in the regime $\mathcal{E} \to 0$ by finding the optimum binary inputs $\{(|\alpha_0\rangle, |\alpha_1\rangle\}$ with distribution $[1-q,q]$ that satisfies the average constraint,

$$(1-q)|\alpha_0|^2 + q|\alpha_1|^2 \leq \mathcal{E}. \quad (14)$$

The following lemma summarizes the result.

**Lemma 1:** The optimum binary inputs for $N = 1$, are $\alpha_0 = \sqrt{\mathcal{E}} \cdot \frac{q^*}{1-q^*}$ and $\alpha_1 = -\sqrt{\mathcal{E}} \cdot (1-q^*/q^*)$ with

$$q^* = \frac{\mathcal{E}}{2} \log \frac{1}{\mathcal{E}}. \quad (15)$$

and the resulting $C_{1, \text{Binary}}(\mathcal{E})$ is

$$C_{1, \text{Binary}}(\mathcal{E}) = \mathcal{E} \log \frac{1}{\mathcal{E}} - \mathcal{E} \log \log \frac{1}{\mathcal{E}} + O(\mathcal{E}). \quad (16)$$

**Proof:** Appendix A

Compared to the ultimate capacity (10) with the same energy constraint, the first-order term of $C_{1, \text{Binary}}(\mathcal{E})$ is the same as that of $C(\mathcal{E})$. But, the difference in the second-order term shows how much less capacity is achievable at $N = 1$ even with the optimized input states. In [23], we showed that (16) can be achieved using an on-off keying modulation $\{\mathcal{E}, 0\}$ and a simple on-off direct-detection (photon counting) receiver. Therefore, in the context of optical communication in the high-PIE regime, all of the performance gain from the complex quantum processing in the JDR is captured by the difference between the second-order terms of Eqs. (10) and (16). In the low photon number regime, this difference in the second-order term can have a significant impact on the practical design of an optical communication system [21]. It would therefore be interesting to ask how large a JDR length $N$ is needed to bridge the gap in the second-order term. To answer this question, we will need the general lower bound on $C_{N}$ that we develop in the following section.

### IV. 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 Holevo capacity of a pure-state cq channel can be achieved. Therefore, this result will provide a framework to understand the trade-off between the (rate) performance and the (quantum) receiver complexity, for reliable transmission of classical information over a quantum channel.

**Theorem 1:** For a given set of pure input states $\{|\psi_x\rangle\}$, $x \in \mathcal{X}$ the maximum achievable information rate using quantum measurements of blocklength $N$, which is $C_{N}/N$ as defined in (2), is lower bounded by

$$\frac{C_{N}}{N} \geq \max_{\mathcal{R}} \left( (1-2e^{-NE(R)})R - \log \frac{2}{N} \right), \quad (17)$$

where

$$E(R) = \max_{0 \leq s \leq 1} \left( \log \frac{\text{Tr}(\rho^{1+s})}{P_\mathcal{X}} - sR \right), \quad (18)$$

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

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 capacity. In the previous section, it was shown that there is a gap between $C_{\text{BPSK}}(\mathcal{E})/\mathcal{E}$ in (13) and $C_{\text{BPSK}}(\mathcal{E})/\mathcal{E}$ in (12) as $\mathcal{E} \to 0$. We saw that the capacity of the BPSK alphabet is as good as that of the optimum continuous Gaussian-distributed input $N$ goes to infinity, i.e., $C_{\text{BPSK}}(\mathcal{E})$ is the same as $C(\mathcal{E})$ in the first two order terms. However, at the measurement blocklength $N = 1$, a BPSK channel cannot even achieve the maximum mutual information of the optimum binary input channel, $C_{1, \text{Binary}}(\mathcal{E})$ in (16), and the PIE caps off at 2 nats/photon. Therefore, the performance of the BPSK channel depends significantly on the regime of $N$. We will now find how much quantum processing is sufficient in order to communicate using the BPSK alphabet at rates close to its capacity.

The following corollary summarizes an answer for the question. Note that for the BPSK inputs $\{|\sqrt{\mathcal{E}}\rangle, |\sqrt{\mathcal{E}}\rangle\}$, any input distribution satisfies the energy constraint of $\mathcal{E}$. Consequently, we can directly apply the Theorem 1 to the BPSK channel—while automatically satisfying the energy constraint—even though the theorem itself does not assume any energy constraint.

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

$$\frac{C_{N, \text{BPSK}}}{N} \geq \left( 1 - 2e^{-N\mathcal{E}(R^*)} \right) R^* - \log \frac{2}{N}, \quad (19)$$

where

$$R^* = \mathcal{E} \log \frac{1}{\mathcal{E}} \left( 1 - \sqrt{\log \frac{N\mathcal{E} \log(N\mathcal{E})}{N\mathcal{E}}} \right) + \mathcal{E}, \quad (20)$$

for $\mathcal{E} = \log(\log(1/\mathcal{E}) - \log(R - \mathcal{E})) - 1.$

We assume that $\mathcal{E}$ is small enough to make the corresponding $s$ be in the range of $0 \leq s \leq 1$.

**Remark 1:** For a narrower range of $N$ such that

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

the lower bound can be further simplified as

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

**Proof:** Appendix B

Note that for the above range of $N$, the lower bound already approaches $C_{\text{BPSK}}(\mathcal{E})$ to the first two orders, which is
the maximum information rate achievable using BPSK with an arbitrarily large length of quantum processing.

Using the result of [19], the photon information efficiency achievable by the BPSK channel is plotted as a function of $N$ in Fig. 2. When the average photon number transmitted per symbol, $\mathcal{E}$, is 0.01, the inner product $\gamma := |(\sqrt{\mathcal{E}} - \sqrt{\mathcal{E}})| = \exp[-2\mathcal{E}] = e^{-0.02}$. By plugging $\gamma$ into (6) and (7), and dividing the resulting capacities by $\mathcal{E}$, the PIE at an arbitrarily large $N$ is 5.55 nats/photon, and at $N = 1$, is 1.97 nats/photon. Therefore, as $N$ increases from 1 to $\infty$, 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 seen that at $N = 2400$, a PIE of 3.0 nats/photon can be achieved, and at $N = 9100$, 4.0 nats/photon is achievable. 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 $N \rightarrow \infty$.

V. 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), both the choice of the JDR measurement as well as joint input distributions over all length-$N$ inner codes must be optimized, in order to find the maximum mutual information of the superchannel. However, the transition probabilities of the superchannel is hard to calculate exactly since it depends on the detailed structure of the inner code and the quantum measurement. Therefore, instead of calculating the exact distribution given a set of measurements and an inner code, we analyze a representative characteristic of the channel distribution $p_{k|j}^{(N)}$ by calculating the average probability of decoding error for the inner decoder, under a uniform distribution over the inner codewords,

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

Now, among superchannels that have the same value of $p_e$, we find a superchannel whose mutual information is the smallest. An equierror superchannel, which was first introduced in [24], is defined with the following distribution:

$$I_{k|j}^{(N)} := \begin{cases} 1 - p_e, & k = j; \\ \frac{1}{e^{NR} - 1} p_e, & k \neq j. \end{cases}$$  \hfill (22)

This channel assumes that the probability of making an error is equal for every input, and when an error occurs, all wrong estimates $k \neq j$ are equally likely. Therefore, this channel is symmetric between inputs, and is symmetric between outputs except for the right estimate, i.e., $k = j$. 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,

$$\max_{p_j} I(p_j, \overline{p}_{k|j}^{(N)}) = NR - p_e \log(e^{NR} - 1) - H_B(p_e) > (1 - p_e)NR - \log 2,$$

where $H_B(p) = -p \log p - (1 - p) \log(1 - p)$.

We will now show 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 the same $p_e$, defined in (21),

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

for the equierror superchannel, $\overline{p}_{k|j}^{(N)}$ with the same $p_e$.

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

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

From Fano’s inequality, we have

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

By combining the above two inequalities,

$$\max_{p_j} I(p_j, \overline{p}_{k|j}^{(N)}) \geq NR - p_e \log(e^{NR} - 1) - H_B(p_e) = \max_{p_j} I(p_j, \overline{p}_{k|j}^{(N)}).$$  \hfill (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 an average error probability $p_e$,

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

In ref. [18], Holevo showed the following upper bound on $p_e$ for a code of length $N$ and rate $R$.

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
\[ p_e \leq 2 \exp[-NE(R)], \] (28)
where, for \( \rho = \sum_x P_X(x)|\psi_x\rangle\langle\psi_x|, \)
\[ E(R) = \max_{0 \leq s \leq 1} \left[ \max_{P_X} \left( -\log \text{Tr} \left( \rho^{1+s} \right) \right) - sR \right]. \] (29)

By combining Lemma 3 with (27), Theorem 1 can be proven.

VI. INTERPRETATION OF SUPERADDITIVITY: CLASSICAL DMC VS. QUANTUM CHANNEL

In Section III, we demonstrated strict superadditivity of \( C_N \), i.e., \( C_N + C_M < C_{N+M} \), for binary-input quantum channels with and without an energy constraint. We provided a general lower bound on \( C_N / N \) for a fixed \( N \) in Theorem 1 which made it possible for us to understand the trade-off between the maximum achievable information rate and the complexity of quantum processing at the receiver as \( N \), the length of the JDR, increases. Previously, the superadditivity of \( C_N \) has been thought of as a unique property that can be observed only in quantum channels, but not in classical DMCs. One popular interpretation of this phenomenon is that a set of length-\( N \) entangling quantum measurements can induce a classical superchannel that has memory over the \( N \) channel uses (despite the fact that the underlying cq channel \( x \to |\psi_x\rangle \) is memoryless over each channel use). The Shannon capacity of this induced classical channel (with memory) can be higher than \( N \) times the Shannon capacity of the classical memoryless channel induced by any symbol-by-symbol receiver measurement. This capability of inducing a classical superchannel by harnessing the optimally-correlated quantum noise in the \( N \)-fold Hilbert space is what increases the number of information bits extractable per modulation symbol, when a longer block of symbols is detected collectively while still in the quantum domain.

Despite the fact that the above intuition of why superadditivity appears in the capacity of classical-quantum channels is somewhat satisfying, this viewpoint does not provide enough quantitative insight to fully understand the phenomenon. In this section, we will introduce a new aspect on understanding strict superadditivity of \( C_N \) by comparing the performance of concatenated coding over quantum channels as we analyze here, and concatenated coding over classical DMCs as studied by Forney [24], for a fixed inner code length \( N \).

Fig. 3 illustrates a concatenated coding architecture over a classical DMC. Compared to Fig. 1, the quantum channel is replaced with a classical DMC \( P_{Y|X} \), and in place of a quantum joint-detection receiver, we now have a classical inner decoder. Forney introduced this concatenated coding architecture for a classical DMC to analyze the trade-off between (rate) performance and (coding) complexity for communication over a classical DMC [24]. Forney analyzed the performance by evaluating the error exponent achievable with a concatenated code, and also examined how the decoding complexity increases as the overall length of the concatenated code, \( N_c = nN \) increases. It is obvious that when the inner decoder generates a sufficient statistic of the channel output and forwards it to the outer decoder, there is no loss of information, so that the performance of the concatenated code can be as good as an optimum code, even within the restricted structure of code concatenation. Despite the fact that the performance remains intact, the decoding complexity increases exponentially with the overall length of the code. On the other hand, it was shown in [24] that even if there is some 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 underlying classical DMC can be achieved with the concatenated code. Moreover, the overall complexity of the decoding algorithm is significantly reduced to be almost linear in the length of the concatenated code. The loss of information at the inner decoder, however, degrades the achievable error exponent over all rates below capacity.

The above 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 inner decoder can be analyzed by using the error exponent of the classical DMC \( P_{Y|X} \) in [25]. It is shown that an optimum inner code with the minimum decoding error can achieve \( p_e \) as low as
\[ p_e = \exp[-N(E(R) + o(1))] \] (30)
as \( N \to \infty \), when
\[ E(R) = \max_{0 \leq s \leq 1} \left( \max_{P_X} \left( E_0(s, P_X) \right) - sR \right) \] (31)
with
\[ E_0(s, P_X) := -\log \sum_y \left[ \sum_x P_X(x) P_{Y|X}(y|x)^{1/(1+s)} \right]^{1+s}. \] (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 hard-decisions on estimating the inner code messages, 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 state of the inner codeword is lost, and in turn all the information that was encoded in the quantum states is destroyed, except for the hard guess of the inner codeword message that the JDR generated. 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 $N$.

We now ask a new question for the concatenated code over the classical DMC, similar to the one we asked for the quantum channel: When the inner decoder makes a hard estimate of messages of the inner code at a finite blocklength $N$, how does the maximum achievable information rate (error-free bits per use of the underlying DMC) with the concatenated code increase as $N$ increases (with no restriction on the complexity of the outer code)?

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,

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

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 $p_e$,

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

Moreover, in [25], 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

$$
p_e \leq \exp[-N E(R)] \quad (35)
$$

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 $N$,

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

with $E(R)$ as defined in [31].

Note that the lower bound on $C_N/N$ in (36) strictly increases with $N$, and it has exactly the same form as that for the quantum channel in (17) except for the difference in $E(R)$ and a constant multiplying $e^{-N E(R)}$. 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 hard-decisions 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, which 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 no 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 understanding can also be applied to explain the same phenomenon observed in the quantum channel by replacing the role of inner code with a quantum joint-detection receiver that makes hard-decisions on finite blocks of quantum states.

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

$$
E^{(c)}(R) = \frac{1}{2V^{(c)}} (R - C)^2 + O ((R - C)^3), \quad (37)
$$
with a parameter $V^{(c)}$, where

$$V^{(c)} = \sum_{x,y} p_x p_y |x \log \frac{p_{y|x}}{p_y} |^2,$$

for the capacity achieving input distribution $p_x := P_X^c(x)$ and the corresponding output distribution $p_y := P_{Y|X}^c(y|x)$ 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 was termed the channel dispersion in [23].

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 average density operator $\rho$, which fully characterizes the classical capacity of the pure-state quantum channel. For a set of input states $\{|\psi_x\rangle\}$, when $P_X$ is the optimum input distribution that attains the capacity of the quantum channel $C = \max_{P_X} \text{Tr}(-\rho \log \rho)$ 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 $\rho$ at $P_X = P_X^c$.

Let us denote the eigenvalues of $\rho$ by $\sigma_i$, $i = 1, \ldots, J$, where $J$ is the dimension of the space spanned by the input states $\{|\psi_x\rangle\}$. From the fact that $\rho$ 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 the variance of the random variable $-\log \sigma$ where $\sigma \in \{\sigma_i\}$ with probability distribution $[\sigma_1, \ldots, \sigma_J]$, i.e.,

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

(39)

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

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

(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 on $C_N$ as well as the approximated error exponent $E(R)$ have similar forms 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 on $C_N$, which can be written with the parameter $V$ and $C$ as follows.

**Theorem 3:** For both a classical DMC and a pure-state classical-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

$$\frac{C_N}{N} \geq C \cdot \left( 1 - \sqrt{\frac{V}{NC^2} \log \left( \frac{NC^2}{V} \right)} \right) - \frac{\log 2}{N} + O \left( \frac{V}{N \log \frac{NC^2}{V} \log \left( \frac{NC^2}{V} \right)} \right),$$

(41)

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

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

(42)

using Theorems 1 and 2. Then, for a fixed rate $R^*$

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

(43)

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

$$E(R^*) = \frac{1}{2V^{(q)}} \log \left( \frac{NC^2}{V} \log \left( \frac{NC^2}{V} \right) \right) + O \left( \frac{V}{NC^2 \log \left( \frac{NC^2}{V} \right)} \right)^{3/2}.$$

(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

$$e^{-NE(R^*)} = \sqrt{\frac{V}{NC^2 \log \left( \frac{NC^2}{V} \right)}} \left( 1 + O(1) \right).$$

(45)

By plugging (43) and (45) into the lower bound (42), $C_N/N$ can be bounded as shown in (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 channels, using the parameter $V/C^2$, we can compare the behaviors of the quantum channel and the classical DMC. For the BPSK quantum channel, by using the two eigenvalues of $\rho$ at $P_X^c$, which are $|\sigma_1 = (1 - e^{-2\bar{\xi}})/2$ and $\sigma_2 = (1 + e^{-2\bar{\xi}})/2$, the channel dispersion and the capacity can be calculated as

$$V_{\text{BPSK}} = C \left( \log \frac{1}{\bar{\xi}} \right)^2 (1 + O(\bar{\xi})), \quad \text{and}$$

(46)

$$C_{\text{BPSK}} = C \log \frac{1}{\bar{\xi}} + C + O(\bar{\xi}).$$

Then, $V_{\text{BPSK}}/C_{\text{BPSK}} \approx 1/\bar{\xi}$ for the low photon number regime where $\bar{\xi} \to 0$. For the classical AWGN channel in the low-power regime where $\text{SNR} \to 0$, $V_{\text{AWGN}}/C_{\text{AWGN}}^2$ can be calculated by using the result of (28), and it is $4/\text{SNR}$. For both 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 meet a targeted fraction of capacity, it is necessary to adopt a longer inner code.
VII. CONCLUSION

The Holevo capacity of a classical-quantum channel, i.e., the ultimate rate of reliable communication for sending classical data over a quantum channel, is a doubly-asymptotic result; meaning the achievability of the capacity $C$ has been proven so far for the case when the transmitter is allowed to code over an arbitrarily large sequence of quantum states (spanning $N_c$ channel uses), and when the receiver is assumed to be able to jointly measure quantum states of the received codewords, also over $N_c$ channel uses, while $N_c \to \infty$. However, the assumption that arbitrarily large number of quantum states can be jointly measured is the primary barrier prohibiting practical implementations of joint detection receivers—particularly in the context of optical communication. Our goal in this paper was to separate these two infinities: the coding blocklength $N_c$ (a relatively inexpensive resource), and the length of the joint detection receiver, $N \leq N_c$ (a far more expensive resource), and to evaluate how the capacity $C_N$, constrained to length-$N$ joint measurements (but no restrictions on the classical code complexity), grows with $N$. We analyzed superadditivity in classical capacity of a pure-state quantum channel while focusing on the quantitative trade-off between reliable-rate performance and quantum-decoding complexity. In order to analyze this trade-off, we adopted a concatenated coding scheme where a quantum joint-detection receiver acts on finite-blocklength quantum codewords of the inner code, and we found a lower bound on the maximum achievable information rate as a function of the length $N$ of the quantum measurement that decodes the inner code. We also observed a similar phenomenon for a classical discrete memoryless channel (DMC), and explained how a classical superadditivity in channel capacity occurs due to a loss of information from the hard-decision at the inner decoder of finite blocklength $N$. We developed a unifying framework, within which the superadditivity in capacity of the classical DMC and that of the pure-state quantum channel can be compared with a parameter $V/C^2$ (where $V$ is the channel dispersion, and $C$ is channel capacity), which is proportional to the inner-code measurement $N$ that is sufficient to achieve a given fraction of the capacity.

ACKNOWLEDGMENTS

This research was supported by the DARPA Information in a Photon (InPho) program under contract number HR0011-10-C-0159.

APPENDIX A

PROOF OF LEMMA \[ \]

$C_{1, \text{Binary}}(\mathcal{E})$ is the maximum mutual information of the binary input channels $\{\{\alpha_0, \alpha_1\}\}$ under the average photon number constraint, $(1 - q)|\alpha_0|^2 + q|\alpha_1|^2 \leq \mathcal{E}$. In [29], it is shown that

$$C_{1, \text{Binary}}(\mathcal{E}) = \max_{(1-q)|\alpha_0|^2 + q|\alpha_1|^2 = \mathcal{E}} H_B(q) - H_B(p)$$

(47)

where $H_B(x) = -x \log x - (1-x) \log (1-x)$, and

$$p = \frac{1 - \sqrt{1 - 4q(1-q)e^{-|\alpha_0 - \alpha_1|^2/2}}}{2}$$

(48)

We first find the optimum input states $\{\{\alpha_0, \alpha_1\}\}$ for a fixed $q$ and then will find $q$ that maximizes (47). In [29], it was shown how to find the optimum input states given a fixed $q$, but we summarize the steps again for reader’s convenience.

For a fixed $q$, the input states that maximize $C_{1, \text{Binary}}(\mathcal{E})$ should minimize $H_B(p)$. Since $H_B(p)$ is an increasing function in $p$ for $0 \leq p \leq 1/2$, and $p$ in (48) gets smaller as $|\alpha_0 - \alpha_1|^2$ increases, we need find inputs $\alpha_0, \alpha_1 \in \mathbb{C}$ that maximizes $|\alpha_0 - \alpha_1|^2$ under the energy constraint $(1 - q)|\alpha_0|^2 + q|\alpha_1|^2 \leq \mathcal{E}$ for a given $q$. To maximize $|\alpha_0 - \alpha_1|^2$ under the photon number constraint,

$$\alpha_1 = -k \alpha_0$$

(49)

for a real number $k \geq 0$ that satisfies

$$(1 - q)|\alpha_0|^2 + q|\alpha_1|^2 = (1 - q + k^2 \cdot q)|\alpha_0|^2 = \mathcal{E}.$$  (50)

The optimum $k$ that maximizes $f(k) := |\alpha_0 - \alpha_1|^2 = (1 + k^2)|\alpha_0|^2 = \left(1 + k^2\mathcal{E}\right)/(1 - q + k^2 \cdot q)$ can be found from

$$\frac{\partial f(k)}{\partial k} = 0,$$

(51)

and the solution is $k^* = (1 - q)/q$. Therefore, the optimum inputs and the resulting $p$ in (48) become

$$\alpha_0^* = \sqrt{\mathcal{E} - q/(1 - q)}$$

$$\alpha_1^* = -\sqrt{\mathcal{E} \cdot (1 - q)/q}$$

(52)

$$p^* = \left(1 - \sqrt{1 - 4q(1-q) \exp\left(-\frac{\mathcal{E}}{q(1-q)}\right)}\right)/2$$

Now, $C_{1, \text{Binary}}(\mathcal{E})$ can be written as

$$C_{1, \text{Binary}}(\mathcal{E}) = \max_q H_B(q) - H_B(p^*)$$

(53)

for $0 \leq q \leq 1/2$.

When we define $I(q) := H_B(q) - H_B(p^*)$, the optimum $q^*$ should satisfy $\partial I(q)/\partial q|_{q=q^*} = 0$. However,

$$\frac{\partial I(q)}{\partial q} = \log \frac{1 - q}{q} - \log \frac{1 - p^*}{p^*} \times$$

$$\left(1 - \frac{2q}{1 - 2p^*} \left(1 + \frac{\mathcal{E}}{q(1-q)}\right) \exp\left(-\frac{\mathcal{E}}{q(1-q)}\right)\right),$$

(54)

and $\partial I(q)/\partial q = 0$ does not have a closed form solution. Instead, by focusing on the low photon number regime where $\mathcal{E} \to 0$, we can find an approximated solution for $q^*$, and calculate $C_{1, \text{Binary}}(\mathcal{E})$ for the first and second order terms. First, it will be shown that

$$\left.\frac{\partial I(q)}{\partial q}\right|_{q = \frac{\mathcal{E}}{2} \log \frac{1}{\mathcal{E}}} > 0 \ \text{and} \ \left.\frac{\partial I(q)}{\partial q}\right|_{q = \frac{\mathcal{E}}{2} \left(\log \frac{1}{\mathcal{E}}\right)} < 0,$$

(55)

which implies that the optimum $q^*$ is between

$$\frac{\mathcal{E}}{2} \log \frac{1}{\mathcal{E}} \leq q^* \leq \frac{\mathcal{E}}{2} \left(\log \frac{1}{\mathcal{E}}\right),$$

(56)
Let us write down the approximation of $\partial I(q)/\partial q$ as $E \to 0$.

When $\frac{\xi}{2} \leq \frac{1}{2} \leq \frac{\xi}{2}$, by Taylor expansion, each term in (54) can be approximated as

$$\exp \left( -\frac{E}{q(1-q)} \right) = 1 - E/q + E^2/(2q^2) + O(E^3/q^3)$$

$p^* = q (1 - E/q + E^2/(2q^2)) + O(E^3/q^3)$

$$\log \frac{1-p^*}{q} = \log(1/q) - q + O(q^2)$$

$$\log \frac{1-p^*}{q} = \log(1/q) + E/q - E^2/(2q^2) + O(E^3/q^3)$$

$$\frac{1}{E} - \frac{q}{q(1-q)} = 1 + O(E).$$

(57)

By using these approximations, it can be shown that $\partial I(q)/\partial q$ in (54) can be written as

$$\frac{\partial I(q)}{\partial q} = \frac{E^2}{2q^2} \log \frac{1}{q} - \frac{E}{q} + O \left( \frac{E^3}{q^3} \log \frac{1}{q} \right).$$

(58)

Now, it can be checked that when $q = \left( E \log(1/E) \right)/2$,

$$\frac{\partial I(q)}{\partial q} = \frac{2}{\log 2} \left( \log \frac{2}{E} - \frac{1}{2} \log \log \frac{1}{E} \right) + O \left( \frac{1}{\sqrt{\log \frac{1}{E}}} \right),$$

$$=-2 + o(1) > 0,$$

and when $q = \left( \frac{E}{2} \log(1/E) \right)/2$,

$$\frac{\partial I(q)}{\partial q} = \frac{2}{\log 2} \left( \log \frac{2}{E} - \log \log \frac{1}{E} \right) - \frac{2}{\log E}$$

$$+ O \left( \frac{1}{\log \frac{1}{E}} \right)^2$$

$$= -2 \log \log \frac{1}{E} \frac{1}{\log \frac{1}{E}} + O \left( \frac{1}{\log \frac{1}{E}} \right)^2 < 0.$$ 

(60)

Therefore, (56) is verified.

Now, we will write $q^* = \frac{\xi}{2} \left( \log \frac{1}{E} \right)^\alpha$ for some $1/2 \leq \alpha \leq 1$, and then find $\alpha^*$ that maximizes $I(q)$. Let us find the approximation of $I(q)$ as $E \to 0$. By using the Taylor expansion for $H_B(x) = -x \log x + x + O(x^2)$ as $x \to 0$,

$$H_B(p^*) = -q \log q + E \log q - \left( E^2 \log q \right)/(2q) + q + O(E),$$

and thus

$$I(q) = H_B(q) - H_B(p^*)$$

$$= -E \log q + (E^2 \log q)/(2q) + O(E).$$

(61)

At $q = q^* = \frac{\xi}{2} \left( \log \frac{1}{E} \right)^\alpha$,

$$I(q) = E \log \frac{1}{E} - E \left( \log \frac{1}{E} \right)^{1-\alpha} - \alpha E \log \log \frac{1}{E} + O(E).$$

(63)

It can be easily checked that $\partial I(q)/\partial \alpha = 0$ when $\alpha = 1$ from

$$\partial I(q)/\partial \alpha = E \log \frac{1}{E} \left( 1 - \left( \log \frac{1}{E} \right)^{1-\alpha} \right).$$

(64)

Therefore, the optimum $q^* = \frac{\xi}{2} \left( \log \frac{1}{E} \right)$ and by plugging $\alpha = 1$ in (63),

$$C_{\text{Binary}} = \max_q I(q) = I(q)|_{\alpha=1}$$

$$= E \log \frac{1}{E} - E \log \log \frac{1}{E} + O(E).$$

(65)

**APPENDIX B**

**PROOF OF COROLLARY**

For the BPSK inputs $\{ |\sqrt{E} \rangle, -|\sqrt{E} \rangle \}$ with input distribution $\rho = (1-q) |\sqrt{E} \rangle \langle \sqrt{E} | + q |\sqrt{C} \rangle \langle \sqrt{C} |$ are

$$\sigma_1 = \left( 1 - \sqrt{1 - 4q(1-q)(1-e^{-2E})} \right)/2$$

$$\sigma_2 = \left( 1 + \sqrt{1 - 4q(1-q)(1-e^{-2E})} \right)/2$$

(66)

from (4) and $r = | \langle \sqrt{E} | - | \sqrt{C} \rangle | = e^{-2E}$.

It can be easily checked that the optimum $q$ that maximizes

$$-\log \text{Tr}(\rho^{1+s}) = - \log (\sigma_1^{1+s} + \sigma_2^{1+s})$$

is equal to $1/2$ by using the symmetry between $q$ and $1-q$ in (60).

When $\sigma_1$ and $\sigma_2$ at $q = 1/2$ are denoted as $\sigma_1^*$ and $\sigma_2^*$,

$$\sigma_1^* = (1 - e^{-2E})/2$$

$$\sigma_2^* = (1 + e^{-2E})/2.$$ 

(68)

Then, the error exponent $E(R)$ in (18) for the BPSK inputs can be written as

$$E(R) = \max_{0 \leq s \leq 1} \left( \max_{P_S} \left( -\log \text{Tr}(\rho^{1+s}) - sR \right) \right)$$

$$= \max_{0 \leq s \leq 1} \left( -\log (\sigma_1^{1+s} + \sigma_2^{1+s}) - sR \right)$$

(69)

A closed form solution for the optimum $s$ that achieves $E(R)$ cannot be found. Instead, by using the assumption of the low photon number regime, i.e., $E \ll 1$, we pick the following $s'$ and find a lower bound of $E(R)$.

$$s' := \begin{cases} \frac{\log \log(1/E) - \log(R-E)}{\log(1/E)}, & R_e < R < C; \\ 1, & R \leq R_e; \\ 0, & R \geq C. \end{cases}$$

(70)

where $R_e = E + E^2 \log(1/E)$ and $C = E \log(1/E) + E$. When we define

$$\tilde{E}(R) := \left( -\log \left( (\sigma_1^*)^{1+s'} + (\sigma_2^*)^{1+s'} \right) - s'R \right),$$

(71)

the error exponent $E(R)$ is lower bounded by $\tilde{E}(R)$ for every $R$, i.e.,

$$E(R) \geq \tilde{E}(R).$$

(72)
Therefore, the lower bound of $C_N/N$ in (17) from Theorem 1 can be further lower bounded by using $\tilde{E}(R)$ as follows.

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

A closed form solution of the optimum $R$ that maximizes the lower bound in (73) cannot be found, but we pick

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

for $N \geq \mathcal{E}^{-1} \log(1/\mathcal{E})$. It can be shown that there exists $\mathcal{E}_0 > 0$ such that when $\mathcal{E} < \mathcal{E}_0$, the chosen rate $R^*$ is in $R_c < R^* < C$ where $N \geq \mathcal{E}^{-1} \log(1/\mathcal{E})$. Therefore, for small enough $\mathcal{E}$ to make $R^*$ be in the range $R_c < R^* < C$,

$$
\frac{C_N}{N} \geq (1 - 2e^{-NE(R^*)})R^* - \frac{\log 2}{N}
$$

in the range of $N \geq \mathcal{E}^{-1} \log(1/\mathcal{E})$. The corollary is proven.

Now, we will show the approximation of (75) as $\mathcal{E} \to 0$. For $0 < s < 1$, by using Taylor expansion,

$$(\sigma_1)^{1+s} = 1 - (1+s)\mathcal{E} + \frac{(1+s)^2(2+s)}{2} \mathcal{E}^2 + o(\mathcal{E}^3),$$

$$(\sigma_2)^{1+s} = \mathcal{E}^{1+s} - (1+s)\mathcal{E}^2 + o(\mathcal{E}^3).$$

By using these approximations and Taylor expansion of $\log(1+x) = x + O(x^2)$ as $x \to 0$,

$$-\log \left((\sigma_1)^{1+s} + (\sigma_2)^{1+s}\right) = (1+s)\mathcal{E} - \mathcal{E}^{1+s} + O(\mathcal{E}^2).$$

Then, for $s = s'$ in the range of $R_c < R < C$,

$$\tilde{E}(R) = (1+s')\mathcal{E} - \mathcal{E}^{1+s'} - s'R + O(\mathcal{E}^2)$$

$$= \frac{(R - \mathcal{E})}{\log(1/\mathcal{E})} \left(\log(R - \mathcal{E}) + \log \frac{1}{\mathcal{E}} - \log \log \frac{1}{\mathcal{E}} - 1\right)
+ \mathcal{E} + O(\mathcal{E}^2).$$

Now, at $R = R^*$,

$$\tilde{E}(R^*) = (\mathcal{E} \cdot f) / 2 + O(\mathcal{E} \cdot f^{3/2} + \mathcal{E}^2).$$

If we further restrict the range of $N$ such that

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

i.e., $\log(1/\mathcal{E}) \leq NE \leq \mathcal{E}^{-1}$,

$$\tilde{E}(R^*)$$

becomes

$$\tilde{E}(R^*) = \frac{\mathcal{E}}{2} \frac{\log (NE \log(N\mathcal{E}))}{NE} + O\left(\frac{1}{N}\right)$$

as $\mathcal{E} \to 0$. Therefore,

$$e^{-NE(R^*)} = \log \sqrt{\mathcal{E} \mathcal{E} \log(NE) + O(1)}$$

in $\mathcal{E}^{-1} \log(1/\mathcal{E}) \leq N \leq \mathcal{E}^{-2}$. Moreover, for a narrower regime of $N$ where $\mathcal{E}^{-1} \log(1/\mathcal{E}) \leq N \leq \mathcal{E}^{-2}$, the term in $O(\cdot)$ can be further simplified such that

$$\frac{C_N}{N} \geq \mathcal{E} \log \frac{1}{\mathcal{E}} \left(1 - \sqrt{\frac{\log (NE \log(N\mathcal{E}))}{NE}}\right) + \mathcal{E} + o(\mathcal{E}).$$

**REFERENCES**

[1] P. Haussladen, 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] 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.

[4] M. Hastings, Nature Physics, Vol. 5, p. 245–257 (2009).

[5] M. Sasaki, K. Kato, M. Izutsu, O. Hirota, IEEE. A 58, 146 (1998).

[6] C. W. Helstrom, Quantum Detection and Estimation Theory, Academic Press, New York (1976).

[7] S. Lloyd, V. Giovannetti, L. Maccone, Phys. Rev. Lett. 106, 250501 (2011).

[8] M. Wilde, S. Guha, S. H. Tan, S. Lloyd, “Explicit capacity-achieving receivers for optical communication and quantum reading, Proceedings of the 2012 IEEE International Symposium on Information Theory (ISIT), Cambridge, MA, USA, pages 551-555, arXiv:1202.0518v2 [quant-ph] (2012).

[9] M. M. Wilde and S. Guha, “Polar Codes for Classical-Quantum Channels”, IEEE Trans. on Inf. Theory, DOI:10.1109/TIT.2012.2218792, (advanced online publication) Sept. 13, 2012.

[10] S. Guha and M. Wilde, “Polar coding to achieve the Holevo capacity of a pure-loss optical channel”, arXiv:1202.0533v2 [cs.IT], Proc. of ISIT 2012, Cambridge, MA, 2012.

[11] M. Takeoka, H. Krovi, and S. Guha, “Achieving the Holevo Capacity of a Pure State Classical-Quantum Channel via Unambiguous State Discrimination”, Proceedings of the 2013 IEEE International Symposium on Information Theory (ISIT), Istanbul, Turkey (2013).

[12] M. Tomamichel, and V. Y. F. Tan, “Second-Order Asymptotics of Classical-Quantum Channels”, arXiv:1308.6805 [quant-ph] (2013).

[13] W. Matthews, and S. Wehner, “Finite blocklength converse bounds for quantum channels”, arXiv:1210.4722 [quant-ph] (2013).

[14] M. V. Burnashev, and A.S.Holevo, “On Reliability Function of Quantum Communication Channel”, [arXiv:quant-ph/9703008] (1998).

[15] M. Wilde, E. B. Davies, “Information and Quantum Measurement”, Academic Press, New York (1976).

[16] S. Guha, “Untrusted quantum receivers for classical-quantum channels”, arXiv:1306.1660v1 [quant-ph] (2013).

[17] W. Matthews, and S. Wehner, “Finite blocklength converse bounds for quantum channels”, arXiv:1210.4722 [quant-ph] (2013).

[18] M. Wilde, S. Guha, “Polar coding to achieve the Holevo capacity of a pure-loss optical channel”, arXiv:1202.0533v2 [cs.IT], Proc. of ISIT 2012, Cambridge, MA, 2012.

[19] M. Takeoka, H. Krovi, and S. Guha, “Achieving the Holevo Capacity of a Pure State Classical-Quantum Channel via Unambiguous State Discrimination”, Proceedings of the 2013 IEEE International Symposium on Information Theory (ISIT), Istanbul, Turkey (2013).
[21] S. Guha, Z. Dutton, and J. H. Shapiro, “On quantum limit of optical communications: concatenated codes and joint-detection receivers”, Proc. IEEE International Symposium on Information Theory (ISIT), 2011.

[22] 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.

[23] 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.

[24] G. D. Forney Concatenated codes, Citeseer, 1966, vol. 11.

[25] R. G. Gallager, “Information theory and reliable communication,” 1968.

[26] M. M. Wilde, O. Landon-Cardinal, and P. Hayden, “Towards efficient decoding of classical-quantum polar codes”, arXiv:1302.0398 [quant-ph], Presented at 8th Conference on the Theory of Quantum Computation, Communication, and Cryptography (TQC), Guelph, Canada (2013).

[27] C. A. Fuchs and C. M. Caves, “Mathematical techniques for quantum communication theory”, [arXiv:quant-ph/9604001]. Open Systems & Information Dynamics, 3(3):345356, (1995).

[28] 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.

[29] M. Sohma and O. Hirota, “Binary discretization for quantum continuous channels,” Physical Review A, vol. 62, no. 5, p. 52312, 2000.