Superadditivity of the Classical Capacity with Limited Entanglement Assistance

Elton Yechao Zhu, Quntao Zhuang, and Peter W. Shor

Phys. Rev. Lett. **119**, 040503 — Published 27 July 2017

DOI: 10.1103/PhysRevLett.119.040503

Superadditivity of the classical capacity with limited entanglement assistance

Elton Yechao Zhu,[1,2,*] Quntao Zhuang,[1,3] and Peter W. Shor[2,4]

[1]Department of Physics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[2]Center For Theoretical Physics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[3]Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[4]Department of Mathematics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

(Dated: June 20, 2017)

Finding the optimal encoding strategies can be challenging for communication using quantum channels, as classical and quantum capacities may be superadditive. Entanglement assistance can often simplify this task, as the entanglement-assisted classical capacity for any channel is additive, making entanglement across channel uses unnecessary. If the entanglement assistance is limited, the picture is much more unclear. Suppose the classical capacity is superadditive, then the classical capacity with limited entanglement assistance could retain superadditivity by continuity arguments. If the classical capacity is additive, it is unknown if superadditivity can still be developed with limited entanglement assistance. We show this is possible, by providing an example. We construct a channel for which, the classical capacity is additive, but that with limited entanglement assistance can be superadditive. This shows entanglement plays a weird role in communication and we still understand very little about it.

In Shannon's classical information theory [1], a classical (memoryless) channel is a probabilistic map from input states to output states. This has been extended to the quantum world. A (memoryless) quantum channel is a time-invariant completely positive trace preserving (CPTP) linear map from input quantum states to output quantum states [2]. A classical channel can only transmit classical information, and the maximum communication rate is fully characterized by its capacity. A quantum channel can be used to transmit not only classical information but also quantum information. Hence, there are different types of capacity, such as classical capacity $\mathcal{C}$ for classical communication [3, 4], and quantum capacity $\mathcal{Q}$ for quantum communication [5–7].

Since quantum channels transmit quantum states, and quantum states can be entangled with other parties, it is natural to ask if entanglement can assist the communication. This was first considered by Bennett *et al.*, who showed that unlimited pre-shared entanglement could improve the classical capacity of a noisy channel [8, 9]. Shor examined the case where only finite pre-shared entanglement is available and obtained a trade-off curve that illustrates how the optimal rate of classical communication depends on the amount of entanglement assistance (CE tradeoff) [10]. One can also consider how entanglement (E), classical communication (C), or quantum communication (Q) can trade off against each other as resources. The tradeoff capacity of almost any two resources was studied by Devetak *et al.* [11, 12], such as entanglement-assisted quantum capacity (QE tradeoff). Subsequently, the triple resource (CQE) tradeoff capacity was also characterized [13–15].

However, almost all the capacity formulae above are given by regularized expressions. They are difficult to evaluate, because they require an optimization over an infinite number of channel uses, which is typically intractable. The existence of this regularization is because entanglement across different channel uses can sometimes protect information against noise and improve the communication rate, a phenomenon often called superadditivity. Superadditivity has long been known to be the case for quantum capacity [16, 17], but remained undiscovered for classical capacity until Hastings gave an example [18]. One exception is the entanglement-assisted classical capacity $\mathcal{C}_E$ [9, 19]. An intuitive understanding of the additivity of $\mathcal{C}_E$ is that the best way to use entanglement is to pre-share it to the receiver, but not across different channels. The need for regularization for various capacity formulae represents our incomplete understanding of quantum channels, as one cannot find the optimal

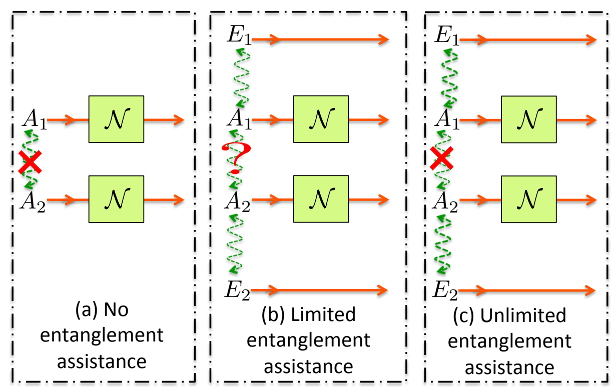

FIG. 1. Consider a channel $\mathcal{N}$ for classical communication, with additive classical capacity. We have the following three scenarios. (a) Entanglement across channel uses does not help if we do not have any assistance. (c) Entanglement across channel uses also does not help if we have unlimited entanglement assistance (this is always true regardless of the channel). The question addressed is case (b), whether entanglement across channel uses can help if we have some entanglement assistance.

transmission rate and best encoding strategies. Thus, an important goal in quantum Shannon theory is to characterize quantum channels with additive capacities. For classical capacity, many such channels are known, including unital qubit channels [20], entanglement-breaking channels [21], etc. For quantum capacity, there are also examples like degradable channels [11]. Additivity for the double or triple resource tradeoff capacity has also been considered, but many fewer examples are known [22].

One can also ask if it's possible to characterize the additivity of a capacity region (*e.g.* CE tradeoff) from some of its subregions (*e.g.* $\mathcal{C}$). This has been shown to be possible for QE tradeoff, as additivity of $\mathcal{Q}$ implies the additivity of quantum capacity with any amount of entanglement assistance [12]. However, the same problem is open for CE tradeoff. This question has only been recently explored [23], where one can restrict the encoding and constraint on entanglement to make it additive.

In this work, we consider the implication of additivity of the classical capacity on the CE tradeoff region. Suppose $\mathcal{C}$ is additive, this means we can look at each channel separately and entangled input states do not help (Fig. 1a). The same is true if there is unlimited entanglement assistance (Fig. 1c). But with limited entanglement assistance, it is unclear whether entangled input states could help (Fig. 1b). We answer the above question affirmatively. We show that there exists a channel $\mathcal{N}$ such that the classical capacity is additive, but with some entanglement assistance P, it becomes superadditive. We give a schematic plot of our CE tradeoff curve in Fig. 2a.

To describe our results precisely, we need to first review a few key notions and results in classical capacity. To transmit classical information, Alice picks a set of signal states ρ_i with probability p_i (denoted as $\{p_i, \rho_i\}$), and sends them through the channel Φ to Bob. The 1-shot classical capacity (*i.e.* Holevo capacity) [3, 4] of Φ is

$$\mathcal{C}^{(1)}(\Phi) = \max_{\{p_i,\rho_i\}} S\left(\sum_i p_i \Phi(\rho_i)\right) - \sum_i p_i S(\Phi(\rho_i)),$$

where $S(\rho) = -\mathrm{tr}(\rho\log(\rho))$ is the von Neumann entropy. This is the maximal rate of reliable classical information transmission achieved using tensor products of states ρ_i, hence the "1-shot" classical capacity [24] . If we can use input states which are entangled across n channel uses, we obtain the n-shot classical capacity $\mathcal{C}^{(n)}(\Phi) = \mathcal{C}^{(1)}(\Phi^{\otimes n})/n$. $\mathcal{C}(\Phi) = \lim_{n\to\infty} \mathcal{C}^{(n)}(\Phi)$ denotes the (regularized) classical capacity, and is the ultimate limit of reliable classical information transmission through Φ. If $\mathcal{C}(\Phi)$ is additive for channel Φ, *i.e.* $\mathcal{C}^{(n)}(\Phi) = \mathcal{C}^{(1)}(\Phi)$ for all n, then we use $\mathcal{C}(\Phi)$ in place of $\mathcal{C}^{(n)}(\Phi)$.

Now consider the scenario where the purifications of the states ρ_i are pre-shared to Bob, who can use them together with the states he receives through Φ for decoding. If we restrict the average amount of pre-shared entanglement to be P ebits per channel use, we arrive at the 1-shot classical capacity with entanglement assistance P [10], denoted as $\mathcal{C}_P^{(1)}(\Phi)$,

$$\mathcal{C}_P^{(1)}(\Phi) = \max_{\substack{\{p_i,\rho_i\} \\ \sum_i p_i S(\rho_i)\le P}} \sum_i p_i S(\rho_i) + S\left(\Phi\left(\sum_i p_i\rho_i\right)\right) - \sum_i p_i S(\Phi\otimes\mathcal{I}(\phi_i)),$$

where $\phi_i := |\phi_i\rangle\langle\phi_i|$ is the density matrix of ρ_i together with a purification. This is also achieved using inputs which are tensor products of states ρ_i. Similar to classical capacity, there is $\mathcal{C}_P^{(n)}(\Phi) = \mathcal{C}_{nP}^{(1)}(\Phi^{\otimes n})/n$ and $\mathcal{C}_P(\Phi)$. Note that the above formula works for any P. In particular, when $P = 0$, we get $\mathcal{C}^{(1)}(\Phi)$. When P is maximal, we get $\mathcal{C}_E(\Phi)$.

Now we are ready to state our main result.

Theorem 1 *(Main Theorem) There exists a channel $\mathcal{N}$ such that*

$$\mathcal{C}(\mathcal{N}) = \mathcal{C}^{(1)}(\mathcal{N})$$

i.e. its classical capacity is additive. However, there exists P such that

$$\mathcal{C}_P(\mathcal{N}) > \mathcal{C}_P^{(1)}(\mathcal{N})$$

i.e. its classical capacity with limited entanglement assistance can be superadditive.

This additivity to superadditivity transition in classical capacity is illustrated in Fig. 2a. This is in sharp contrast with the QE tradeoff curve (Fig. 2b), as $\mathcal{Q}_P^{(n)}$ grows linearly in P with gradient 1. Additivity of $\mathcal{Q}_p$ follows from the additivity of $\mathcal{Q}$.

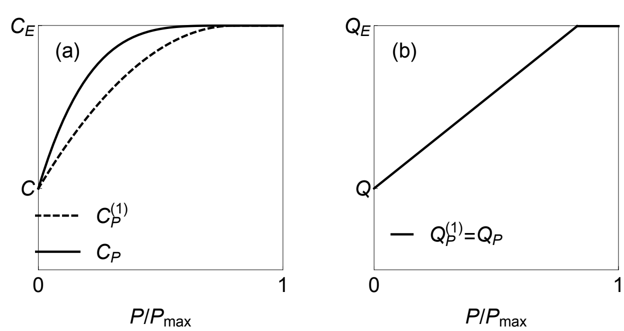

FIG. 2. (a) Schematic plot of the superadditivity in CE tradeoff for our channel. $\mathcal{C}_P = \mathcal{C}_P^{(1)}$ at $P = 0$ and $P_{\max}$, but not for all values in between. (b) QE tradeoff curve for channels with additive quantum capacity. $P_{\max}$ is the maximum amount of available entanglement assisntance.

Our channel $\mathcal{N}$ is a conditional quantum channel $\mathcal{N}^{MA\to B}$ [25], where register M determines whether

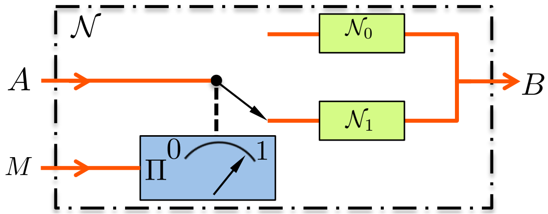

FIG. 3. diagrammatic representation of $\mathcal{N}$

$\mathcal{N}_0^{A\to B}$ or $\mathcal{N}_1^{A\to B}$ is used (see Fig. 3 for a diagrammatic representation). Explicitly, on any input state ρ^{MA} [26],

$$\mathcal{N}(\rho^{MA}) = \mathcal{N}_0\left(\langle 0|\,\rho^{MA}\,|0\rangle^M\right) + \mathcal{N}_1\left(\langle 1|\,\rho^{MA}\,|1\rangle^M\right). \tag{1}$$

This construction is similar to the one in Ref. [27]. However, their construction does not directly apply to our case since M is kept and contains classical information.

The intuition why a channel like $\mathcal{N}$ will work is that without entanglement, we are only using the classical channel $\mathcal{N}_0$, hence its classical capacity is additive. As one increases entanglement assistance, one starts using the quantum channel $\mathcal{N}_1$, where superadditivity kicks in.

Our construction is generic and does not depend on the specific forms of $\mathcal{N}_0$ and $\mathcal{N}_1$. Hence we give the properties of $\mathcal{N}_0$ and $\mathcal{N}_1$ that are required for our argument to work, and will give a construction of $\mathcal{N}_1$ later. An example of $\mathcal{N}_0$ is given in the supplemental material.

We require the classical channel $\mathcal{N}_0$ to have the following properties:

0.1 $\mathcal{C}\left(\mathcal{N}_0\right) = \log(|B|) - \min S\left(\mathcal{N}_0\left(\rho\right)\right)$.

0.2 It has a noise parameter η which can be tuned, such that $\mathcal{C}\left(\mathcal{N}_0\right)$ varies from 0 to $\log(|B|)$ continuously.

We require the quantum channel $\mathcal{N}_1$ to have the following properties:

1.1 It has a superadditive classical capacity, *i.e.* $\mathcal{C}\left(\mathcal{N}_1\right) > \mathcal{C}^{(1)}\left(\mathcal{N}_1\right)$.

1.2 For any n and P,

$$\mathcal{C}_P^{(n)}(\mathcal{N}_1) = \log(|B|) - \min_{S(\rho)\le nP} \frac{1}{n}\left(S\left(\mathcal{N}_1^{\otimes n}\otimes\mathcal{I}\left(\phi_\rho\right)\right) - S\left(\rho\right)\right).$$

1.3 There exists $P > 0$ such that $\mathcal{C}_P\left(\mathcal{N}_1\right) > \mathcal{C}_P^{(1)}\left(\mathcal{N}_1\right)$ and $\mathcal{C}_P\left(\mathcal{N}_1\right)$ is strictly concave at P.

Here by saying a function f is strictly concave at y, we mean $f(y) > (1-p)f(v) + pf(w)$ for all $v < y < w$ satisfying $(1-p)v + pw = y$, with $p \in (0,1)$. It is clear that $\mathcal{C}_P(\Phi)$ is always concave in P. If $P = (1-p)P_1 + pP_2$, then $\mathcal{C}_P(\Phi) \ge p\mathcal{C}_{P_1}(\Phi) + (1-p)\mathcal{C}_{P_2}(\Phi)$, as one can always just use entanglement P_1 for p fraction of the channel uses, and entanglement P_2 for the other fraction.

The rest of the paper is organized as follows. We first state Lemma 2 about the classical capacity with limited entanglement assistance, of partial cq channels (defined in Lemma 2). This lemma together with the properties above lead to the simplification of capacity formulae, as we show in Lemmas 3 and 4. We'll prove our main theorem in the main text, and leave the proofs of various lemmas to the supplemental material.

Lemma 2 *Suppose a channel Ψ has input Hilbert space $\mathcal{H}_R\otimes\mathcal{H}_C$. If there exists a noiseless classical channel Π on $\mathcal{H}_R$ with orthonormal basis $\{|j\rangle\}$, such that*

$$\Psi = \Psi\circ\left(\Pi\otimes\mathcal{I}_C\right),$$

then $\mathcal{C}_P^{(1)}(\Psi)$ can be achieved with an input ensemble $\{p_{ij}, |j\rangle\langle j|\otimes\rho_{ij}\}$, where ρ_{ij} are states of C.

By saying Π is a noiseless classical channel with orthonormal basis $\{|j\rangle\}$, we mean $\Pi\left(\rho\right) = \sum_j |j\rangle\langle j|\,\rho\,|j\rangle\langle j|$. This is very intuitive. Entanglement between R and other parties is not useful, as Π destroys it. Since we only have limited entanglement, it is better to use it on C.

Using Lemma 2 and properties of $\mathcal{N}_0$ and $\mathcal{N}_1$, we can simplify the various capacity formulae of $\mathcal{N}$.

Lemma 3

$$\mathcal{C}^{(1)}(\mathcal{N}) = \max\left\{\mathcal{C}\left(\mathcal{N}_0\right), \mathcal{C}^{(1)}\left(\mathcal{N}_1\right)\right\},$$
$$\mathcal{C}(\mathcal{N}) = \max\left\{\mathcal{C}(\mathcal{N}_0), \mathcal{C}(\mathcal{N}_1)\right\}.$$

Lemma 2 ensures that for different uses of the channel, we can choose to use $\mathcal{N}_0$ or $\mathcal{N}_1$ only, without sacrificing the capacity. Lemma 3 simply states that, for all channel uses, we should use either $\mathcal{N}_0$ or $\mathcal{N}_1$.

Lemma 4

$$\mathcal{C}_P^{(1)}(\mathcal{N}) = \max_{\substack{\{q,P'\}\\(1-q)P'=P}} q\mathcal{C}(\mathcal{N}_0) + (1-q)\mathcal{C}_{P'}^{(1)}(\mathcal{N}_1), \tag{2}$$

$$\mathcal{C}_P\left(\mathcal{N}\right) = \max_{\substack{\{q,P'\}\\(1-q)P'=P}} q\mathcal{C}\left(\mathcal{N}_0\right) + (1-q)\mathcal{C}_{P'}\left(\mathcal{N}_1\right). \tag{3}$$

This lemma states that, for entanglement-assisted classical communication, the best strategy is to use $\mathcal{N}_0$ for some fraction of the channel uses and $\mathcal{N}_1$ for the other fractions of the channel uses (*i.e.* time sharing). Since using $\mathcal{N}_0$ does not require entanglement assisntace, we can allocate more of it to $\mathcal{N}_1$.

Now we are ready to prove the main theorem.

Proof of Main Theorem— Choose $\mathcal{N}_0$ such that

$$\mathcal{C}\left(\mathcal{N}_0\right) = \mathcal{C}\left(\mathcal{N}_1\right) > \mathcal{C}^{(1)}\left(\mathcal{N}_1\right). \tag{4}$$

By Lemma 3, the classical capacity of $\mathcal{N}$ is additive, *i.e.*

$$\mathcal{C}\left(\mathcal{N}\right) = \mathcal{C}\left(\mathcal{N}_0\right) = \mathcal{C}^{(1)}\left(\mathcal{N}\right). \tag{5}$$

From Eqs. (3),(4) and concavity of $\mathcal{C}_P(\mathcal{N}_1)$ with respect to P, we have $\mathcal{C}_P(\mathcal{N}) \leq \mathcal{C}_P(\mathcal{N}_1)$. Also, $\mathcal{C}_P(\mathcal{N}) \geq \mathcal{C}_P(\mathcal{N}_1)$ by choosing $q=0$ in Eq. (3). So we have

$$\mathcal{C}_P(\mathcal{N}) = \mathcal{C}_P(\mathcal{N}_1). \tag{6}$$

Choose $P>0$ according to Property 1.3. By Lemma 4, suppose $\mathcal{C}_P^{(1)}(\mathcal{N})$ is achieved at some $\left\{\tilde{q}, \tilde{P}\right\}$ with $(1-\tilde{q})\tilde{P} = P$, *i.e.*

$$\mathcal{C}_P^{(1)}(\mathcal{N}) = \tilde{q}\mathcal{C}(\mathcal{N}_0) + (1-\tilde{q})\mathcal{C}_{\tilde{P}}^{(1)}(\mathcal{N}_1). \tag{7}$$

If $\tilde{P} = P$, we have

$$\mathcal{C}_P(\mathcal{N}) = \mathcal{C}_P(\mathcal{N}_1) > \mathcal{C}_P^{(1)}(\mathcal{N}_1) = \mathcal{C}_P^{(1)}(\mathcal{N}), \tag{8}$$

where the inequality follows from Property 1.3.
If $\tilde{P} > P$ and thus $\tilde{q} > 0$,

$$\begin{aligned}\mathcal{C}_P(\mathcal{N}) &= \mathcal{C}_P(\mathcal{N}_1) > \tilde{q}\mathcal{C}(\mathcal{N}_1) + (1-\tilde{q})\mathcal{C}_{\tilde{P}}(\mathcal{N}_1) \\ &\geq \tilde{q}\mathcal{C}(\mathcal{N}_0) + (1-\tilde{q})\mathcal{C}_{\tilde{P}}^{(1)}(\mathcal{N}_1) = \mathcal{C}_P^{(1)}(\mathcal{N}),\end{aligned} \tag{9}$$

where the first inequality follows from Property 1.3.

Construction of $\mathcal{N}_1$— The first two properties for $\mathcal{N}_1$ can be easily satisfied. One can take a channel with a subadditive minimum output entropy [18] and unitally extend it to a channel with a superadditive classical capacity, via Shor's construction [28, 29]. Unfortunately, such channels are poorly understood, and we do not know if it satisfies Property 1.3. We argue that if it doesn't, we can tensor product a dephasing channel that will guarantee it is satisfied, without sacrificing the other properties.

We quote the following property about concave functions [30]: A concave function $u(y)$ is continuous, differentiable from the left and from the right. The derivative is decreasing, *i.e.* for $x<y$ we have $u'(x-) \geq u'(x+) \geq u'(y-) \geq u'(y+)$. We use "$\pm$" to denote the right and left derivatives when needed.

Let $\mathcal{E}^{C\to C}$ be a random orthogonal channel with subadditive minimum output entropy [18] and $\mathcal{F}^{RC\to C}$ (with $|R| = |C|^2$) be a conditional quantum channel of the form

$$\mathcal{F}\left(\rho^{RC}\right) = \sum_{j=1}^{|C|^2} X_j \mathcal{E}\left(\langle j|\,\rho^{RC}\,|j\rangle^R\right) X_j^\dagger, \tag{10}$$

where X_j's are the Heisenberg-Weyl operators on C [6]. This ensures $\mathcal{F}$ satisfies Properties 1.1 and 1.2 [31].

Due to Lemma 2, the useful entanglement assistance is at most $\log(|C|)$. Thus we restrict to $0 \leq P \leq \log(|C|)$.

Let

$$\epsilon = \mathcal{C}(\mathcal{F}) - \mathcal{C}^{(1)}(\mathcal{F}) > 0. \tag{11}$$

Since

$$\mathcal{C}_P^{(1)}(\mathcal{F}) \leq \mathcal{C}^{(1)}(\mathcal{F}) + P, \tag{12}$$

$$\mathcal{C}_E(\mathcal{F}) \leq \mathcal{C}(\mathcal{F}) + \log(|C|) - \epsilon. \tag{13}$$

This implies $d\mathcal{C}_P(\mathcal{F})/dP$ cannot always be 1. Thus there exists $\bar{P} \in [0, \log(|C|))$ such that

$$d\mathcal{C}_P(\mathcal{F})/dP = 1, \quad \forall\, 0 \leq P \leq \bar{P} \tag{14}$$

and

$$d\mathcal{C}_P(\mathcal{F})/dP < 1, \quad \forall P > \bar{P}. \tag{15}$$

Next we discuss the few different cases.

1. $\bar{P} > 0$. Then $\mathcal{C}_P(\mathcal{F})$ is strictly concave at $\bar{P}$ by definition.
   Note that $\mathcal{C}_{\bar{P}}(\mathcal{F}) = \mathcal{C}(\mathcal{F}) + \bar{P}$ but $\mathcal{C}_{\bar{P}}^{(1)}(\mathcal{F}) \leq \mathcal{C}^{(1)}(\mathcal{F}) + \bar{P}$, thus $\mathcal{C}_{\bar{P}}(\mathcal{F}) - \mathcal{C}_{\bar{P}}^{(1)}(\mathcal{F}) \geq \epsilon$ and $\mathcal{N}_1 = \mathcal{F}$ satisfies Property 1.3.

2. $\bar{P} = 0$. Let $\mathcal{N}_1 = \mathcal{F} \otimes \Delta_\lambda^Z$, where Δ_λ^Z is the qubit dephasing channel $\Delta_\lambda^Z(\rho) = (1-\lambda)\rho + \lambda Z\rho Z$. The CQE tradeoff region is additive for $\Phi \otimes \Delta_\lambda^Z$, for any channel Φ, thus $\mathcal{N}_1$ satisfies Property 1.1. Δ_λ^Z satisfies Property 1.2, and by arguments similar to Appendix B of Ref. [22], one can show $\mathcal{N}_1$ also satisfies Property 1.2.

   Since $d\mathcal{C}_P(\mathcal{F})/dP|_{0+} < 1$, choose $\lambda > 0$ small such that $d\mathcal{C}_P\left(\Delta_\lambda^Z\right)/dP|_{1-} > d\mathcal{C}_P(\mathcal{F})/dP|_{0+}$. This ensures that when $0 < P \leq 1$,

   $$\mathcal{C}_P(\mathcal{N}_1) = \mathcal{C}(\mathcal{F}) + \mathcal{C}_P\left(\Delta_\lambda^Z\right). \tag{16}$$

   Since $\mathcal{C}_P\left(\Delta_\lambda^Z\right)$ is strictly concave with respect to P when $\lambda < 1/2$ [13], $\mathcal{C}_P(\mathcal{N}_1)$ is also strictly concave with respect to P, for $0 < P \leq 1$. Also, when $P < \epsilon$,

   $$\begin{aligned}\mathcal{C}_P(\mathcal{N}_1) &\geq \mathcal{C}(\mathcal{F}) + \mathcal{C}\left(\Delta_\lambda^Z\right) \\ &> \mathcal{C}^{(1)}(\mathcal{F}) + \mathcal{C}\left(\Delta_\lambda^Z\right) + P \geq \mathcal{C}_P^{(1)}(\mathcal{N}_1),\end{aligned}$$

   where the first inequality comes from Eq.(16), the second one comes from our assumption $P < \epsilon$ and Eq. (11) and the last one comes from Eq.(12).
   This ensures that $\mathcal{C}_P(\mathcal{N}_1)$ is superadditive. Thus when $0 < P < \min\{1, \epsilon\}$, $\mathcal{C}_P(\mathcal{N}_1)$ is strictly concave and superadditive, satisfying Property 1.3.

Conclusion— Our work unveils the complications in characterizing the additivity of the CE capacity region. In fact, the only known channels that admit an additive CE capacity region are the quantum erasure channels [13] and Hadamard channels [22], many fewer than the class of channels with an additive classical capacity. Coincidentally, these two classes of channels also admit an additive CQE tradeoff capacity, suggesting a non-trivial connection [13, 22, 32].

Also, we do not know the number of shots at which the superadditivity occurs. However, it is very likely that our $\mathcal{N}_1$ only has superadditivity in classical capacity up to 2 shots[33]. In that case, the superadditivity in classical capacity with limited entanglement will appear at 2 shots.

Acknowledgement—EYZ and QZ would like to thank Min-Hsiu Hsieh for many insightful discussions. EYZ and PWS are supported by the National Science Foundation under grant Contract Number CCF-1525130. QZ is supported by the Claude E. Shannon Research Assistantship. PWS is supported by the NSF through the STC for Science of Information under grant number CCF0-939370.

* eltonzhu@mit.edu

[1] C. E. Shannon, The Bell System Technical Journal **27**, 379 (1948).
[2] M.-D. Choi, Linear Algebra and its Applications **10**, 285 (1975).
[3] A. S. Holevo, IEEE Transactions on Information Theory **44**, 269 (1998).
[4] B. Schumacher and M. D. Westmoreland, Phys. Rev. A **56**, 131 (1997).
[5] S. Lloyd, Phys. Rev. A **55**, 1613 (1997).
[6] P. W. Shor, MSRI Workshop on Quantum Computation (2002).
[7] I. Devetak, IEEE Transactions on Information Theory **51**, 44 (2005).
[8] C. H. Bennett, P. W. Shor, J. A. Smolin, and A. V. Thapliyal, Phys. Rev. Lett. **83**, 3081 (1999).
[9] C. H. Bennett, P. W. Shor, J. A. Smolin, and A. V. Thapliyal, IEEE Transactions on Information Theory **48**, 2637 (2002).
[10] P. W. Shor, Quantum Information and Computation **4**, 537 (2004).
[11] I. Devetak and P. W. Shor, Communications in Mathematical Physics **256**, 287 (2005).
[12] I. Devetak, A. W. Harrow, and A. J. Winter, IEEE Transactions on Information Theory **54**, 4587 (2008).
[13] M.-H. Hsieh and M. M. Wilde, IEEE Transactions on Information Theory **56**, 4682 (2010).
[14] M.-H. Hsieh and M. M. Wilde, IEEE Transactions on Information Theory **56**, 4705 (2010).
[15] M. M. Wilde and M.-H. Hsieh, Quantum Information Processing **11**, 1431 (2012).
[16] D. P. DiVincenzo, P. W. Shor, and J. A. Smolin, Phys. Rev. A **57**, 830 (1998).
[17] G. Smith and J. A. Smolin, Phys. Rev. Lett. **98**, 030501 (2007).
[18] M. B. Hastings, Nat Phys **5**, 255 (2009).
[19] Here we are only talking about the case of a single sender and a single receiver. If there are multiple senders, then the entanglement-assisted classical capacity requires regularization [35].
[20] C. King, Journal of Mathematical Physics **43**, 4641 (2002).
[21] P. W. Shor, Journal of Mathematical Physics **43**, 4334 (2002).
[22] K. Brádler, P. Hayden, D. Touchette, and M. M. Wilde, Phys. Rev. A **81**, 062312 (2010).
[23] Q. Zhuang, E. Y. Zhu, and P. W. Shor, "The additive classical capacity of quantum channels assisted by noisy entanglement," (2016), arXiv:1609.08592.
[24] This is not to be confused with the ϵ-1-shot classical capacity, which is the amount of information transmitted through a single use of the channel, with the average error probability less than ϵ.
[25] J. Yard, "Simultaneous classical-quantum capacities of quantum multiple access channels," (2005), arXiv:quant-ph/0506050.
[26] One can also normalize the input states to $\mathcal{N}_0$ and $\mathcal{N}_1$.
[27] D. Elkouss and S. Strelchuk, Phys. Rev. Lett. **115**, 040501 (2015).
[28] P. W. Shor, Communications in Mathematical Physics **246**, 453 (2004).
[29] M. Fukuda, Quantum Information Processing **6**, 179 (2007).
[30] W. Rudin, *Principles of Mathematical Analysis* (McGraw-Hill Education, 1976).
[31] See supplemental material.
[32] E. Y. Zhu, Q. Zhuang, M.-H. Hsieh, and P. W. Shor, (2017).
[33] A. Montanaro, Communications in Mathematical Physics **319**, 535 (2013).
[34] M. M. Wilde, *Quantum Information Theory* (Cambridge University Press, 2013).
[35] M.-H. Hsieh, I. Devetak, and A. Winter, IEEE Transactions on Information Theory **54**, 3078 (2008).