Unbounded number of channel uses may be required to detect quantum capacity

Toby Cubitt,[1] David Elkouss,[2,*] William Matthews,[1,3] Maris Ozols,[1] David Pérez-García,[2] and Sergii Strelchuk[1]

[1] *Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Cambridge CB3 0WA, U.K.*
[2] *Departamento de Análisis Matemático and Instituto de Matemática Interdisciplinar, Universidad Complutense de Madrid, 28040 Madrid, Spain*
[3] *Statistical Laboratory, University of Cambridge, Wilberforce Road, Cambridge CB3 0WB, U.K.*

Transmitting data reliably over noisy communication channels is one of the most important applications of information theory, and is well understood for channels modelled by classical physics. However, when quantum effects are involved, we do not know how to compute channel capacities. This is because the formula for the quantum capacity involves maximising the coherent information over an unbounded number of channel uses. In fact, entanglement across channel uses can even increase the coherent information from zero to non-zero! Here we study the number of channel uses necessary to detect positive coherent information. In all previous known examples, two channel uses already sufficed. Could it be that a finite number of channel uses is always sufficient? We show that this is not the case: for any number of uses, there are channels for which the coherent information is zero, but which nonetheless have capacity.

In the classical case, not only can we exactly characterise the maximum rate of communication over any channel – its capacity – we also have practical error-correcting codes that attain this theoretical limit. It is instructive to review why the capacity of classical channels is a solved problem. Even though optimal communication over a discrete, memoryless classical channel involves encoding the information across many uses of the channel, Shannon showed that a channel's capacity is given by optimising an entropic quantity (the mutual information) over a single use of the channel. This follows immediately from the fact that mutual information is additive.

It is for this reason that additivity questions for quantum channel capacities took on such importance, and why the major recent breakthroughs proving that additivity is violated [1, 2] had such an impact. A regularised expression for the quantum capacity has been known for some time [3–5] – the optimisation of an entropic quantity (the coherent information I_{coh}) in the limit of arbitrarily many uses of the channel:

$$Q^{(n)}(\mathcal{N}) := \frac{1}{n} \max_{\rho^{(n)}} I_{\text{coh}}(\mathcal{N}^{\otimes n}, \rho^{(n)}), \tag{1}$$

$$Q(\mathcal{N}) := \lim_{n\to\infty} Q^{(n)}(\mathcal{N}). \tag{2}$$

Here

$$I_{\text{coh}}(\mathcal{N}^{\mathsf{A}\to\mathsf{B}}, \rho^{\mathsf{A}}) := S\big(\mathcal{N}(\rho^{\mathsf{A}})\big) - S\big(\mathcal{N}(\rho^{\mathsf{AR}})\big) \tag{3}$$

where $\mathcal{N}^{\mathsf{A}\to\mathsf{B}}$ is a channel from A to B, ρ^{AR} is a purification of ρ^{A}, and S denotes the von Neumann entropy. However, the regularisation renders computing the quantum capacity infeasible; it involves an optimisation over an infinite parameter space.

Were the coherent information additive, so that $Q^{(n)}(\mathcal{N}) = Q^{(1)}(\mathcal{N})$, the regularisation could be removed and the quantum capacity could be computed by a single optimization, similarly to classical channels. However, this is not the case. The first explicit examples of superadditivity were given by Di Vincenzo et al. [6], and extended by Smith et al. [7]. For these examples (where $\mathcal{N}$ is a particular depolarising channel) it was shown (numerically) that $0 \leq Q^{(1)}(\mathcal{N}) < Q^{(n)}(\mathcal{N})$ for small values of $n \leq 33$.

While the classical capacity of quantum channels also involves a regularised formula [2], we at least know precisely in which cases it is zero: simply for those channels whose output is completely independent of the input. The set of channels with zero quantum capacity is much richer. Indeed, the complete characterisation of such channels is unknown. To date, we know of only two kinds of channels with zero quantum capacity: antidegradable channels [8, 9] and entanglement-binding channels [10]. The former has the property that the environment can reproduce the output, thus $Q = 0$ by the no-cloning theorem [11]. The latter can only distribute PPT entanglement, which cannot be distilled by local operations and classical communication [12], again implying $Q = 0$.

This has dramatic consequences. It is possible to take two channels with no quantum capacity whatsoever ($Q(\mathcal{N}_1) = Q(\mathcal{N}_2) = 0$), $\mathcal{N}_1$ antidegradable and $\mathcal{N}_2$ entanglement-binding, which, when used together, do have quantum capacity ($Q(\mathcal{N}_1 \otimes \mathcal{N}_2) > 0$). This "superactivation" phenomenon was discovered recently by Smith and Yard [1] (and extended in [13]). They also show that a single channel $\mathcal{N}$, which can be 'switched' between acting like $\mathcal{N}_1$ or $\mathcal{N}_2$, exhibits an extreme form of superadditivity of the coherent information, where $0 = Q^{(1)}(\mathcal{N}) < Q^{(2)}(\mathcal{N})$. Even stronger superactivation phenomena have been shown in the context of zero-error communication [14–18].

These recent additivity violation results demonstrate how much we still do not understand about the behaviour of information in quantum mechanical systems. On the

* delkouss@ucm.es

one hand, it means that the known formula for the quantum capacity must be regularised, hence cannot be used to compute the capacity. On the other hand, since the coherent information is additive for unentangled input states, additivity violation also implies that entanglement can protect information from noise in a way that is not possible classically.

But just how badly can additivity be violated? One might hope that, at least in determining whether the quantum capacity is non-zero, one need only consider a finite number of uses of a channel. Indeed, since the Smith and Yard construction relies on combining the only two known types of zero-capacity channels, one might dare to hope that even two uses suffice. (Similarly, for the classical capacity of quantum channels the only known method for constructing examples of additivity violation [2, 19] cannot give a violation for more than two uses of a channel, and there is some evidence that this may be more than just a limitation of the proof techniques [20].) Were this the case, one could decide if a channel has quantum capacity by optimising the coherent information over two uses of the channel, which is not substantially more difficult than the optimisation over a single channel use.

In this paper, we show for the first time that this is not the case: additivity violation is essentially as bad as it could possible be. More precisely (see Fig. 1), we prove that for any n one can construct a channel $\mathcal{N}$ for which the coherent information of n uses is zero ($Q^{(n)}(\mathcal{N}) = 0$), yet for a larger number of uses the coherent information is strictly positive, implying that the channel has non-zero quantum capacity ($Q(\mathcal{N}) > 0$). This is also the first proof that there can be a gap between $Q^{(n)}(\mathcal{N})$ and the quantum capacity for an arbitrarily large number n of uses of the channel. Our result implies that, in general, one must consider an arbitrarily large number of uses of the channel just to determine whether the channel has any quantum capacity at all!

RESULTS

A channel with zero coherent information but positive capacity

Perhaps the earliest indication that deciding whether a channel has quantum capacity may be difficult comes from the work of Watrous [21], who showed that an arbitrarily large number of copies of a bipartite quantum state can be required for entanglement distillation assisted by two-way classical communication. Our result can be regarded as the counterpart of [21] for the quantum capacity (which is mathematically equivalent to entanglement distillation assisted by one-way communication). However, since the proof ideas and techniques of [21] assume two-way communication, they do not apply in our setting. Our result is instead based on the ideas of Smith and Yard, in particular the intuition provided by Oppenheim's commentary thereon [22].

This intuition comes from a class of bipartite quantum states called pbits (private bits) [23], together with the standard equivalences between quantum capacity (sending entanglement over a channel) and distilling entanglement from the state obtained by sending one half of a maximally entangled state through the channel (its Choi-Jamiołkowski state). A pbit ρ_{aAbB} is a state shared between Alice (who holds aA) and Bob (who holds bB), where the ab part of the system is usually called the "key", AB the "shield" (see Fig. 1 for a graphical representation and refer to "Channel Construction" in the Supplementary Note 1 for the mathematical details of the pbit construction). For concreteness let us consider a state ρ_{aAbB} of the following form:

$$\rho_{\mathsf{aAbB}} = \tfrac{1}{2}(|\phi^+\rangle\langle\phi^+|_{\mathsf{ab}} \otimes \sigma^+_{\mathsf{AB}} + |\phi^-\rangle\langle\phi^-|_{\mathsf{ab}} \otimes \sigma^-_{\mathsf{AB}}). \quad (4)$$

That is, they hold one of the two states $|\phi^\pm\rangle\langle\phi^\pm|_{\mathsf{ab}} \otimes \sigma^\pm_{\mathsf{AB}}$ with equal probability. Here, $|\phi^\pm\rangle$ are Bell states and $\sigma^\pm$ are hiding states [24] encoding the identity of the Bell state. Hiding states are perfectly distinguishable globally, but cannot be distinguished locally using local operations and classical communication (LOCC). If Alice and Bob knew which Bell state they held, they would have at least one ebit of shared entanglement. But this information, and hence the entanglement, is inaccessible to them unless one party is given the whole shield AB.

Now imagine they have access to a quantum erasure channel $\mathcal{E}_{\frac{1}{2}}$ which with probability $1/2$ transmits its input perfectly, but otherwise completely erases it. It is well known that such a channel cannot be used to transmit any entanglement [25]. However, if they also share ρ_{aAbB}, Alice can use the erasure channel to send her part A of the shield to Bob. If the erasure channel transmits, Bob now holds the entire AB system and can now distinguish $\sigma^\pm$. Thus, with probability $1/2$, Alice and Bob can now extract the entanglement from ρ_{aAbB}.

Instead of supplying Alice and Bob with the state ρ_{aAbB} and an erasure channel, we supply them with a switched channel. This has an auxiliary classical input that controls whether the channel acts as $\mathcal{E}_{\frac{1}{2}}$ or Γ, where Γ is the channel with Choi-Jamiołkowski state ρ_{aAbB}. The above argument then implies that no quantum information can be sent over a single use of the channel, but it can be sent using two uses, by switching one to $\mathcal{E}_{\frac{1}{2}}$ and the other to Γ.

This is the intuition behind the Smith and Yard construction [22]. However, it is constructed out of two very particular types of quantum channels, so this idea does not seem to extend to more than two channel uses. Nonetheless, the intuition behind our result is based on a refinement of these ideas, which we now sketch.

Sketch of the general construction

We want to achieve two seemingly contradictory goals: Firstly, to prevent Alive from sending any quantum information to Bob over n of uses of the channel, and secondly

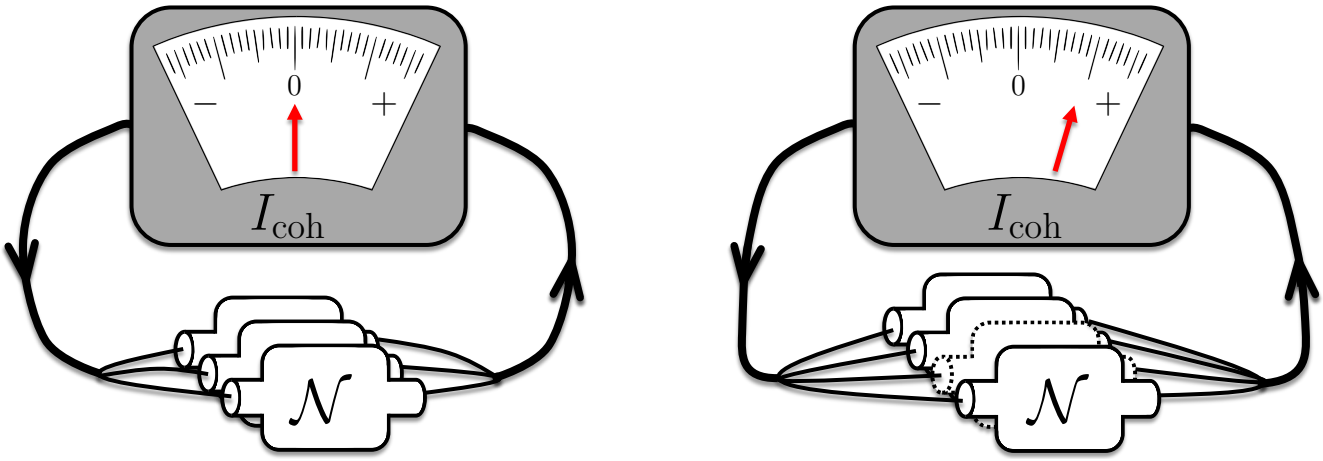

FIG. 1. Positive coherent information can be detected after unbounded uses. Checking the coherent information (I_{coh}) for $n = 3$ uses fails to reveal that channel $\mathcal{N}$ has quantum capacity. However, the channel has capacity and this can be shown by checking some larger number of uses of the channel. We show that for any number of uses of the channel n there are channels with this behavior.

to permit this when Alice has access to some larger number of uses $N > n$. We can achieve the first goal by increasing the erasure probability of the erasure channel to something much closer to one, and also adding noise to the Γ channel; the noise then swamps any entanglement. The problem is that this seems to render the second goal impossible. If the channel is so noisy that it destroys all entanglement sent through it, then no amount of coding over multiple uses of the channel can transmit any quantum information.

However, note that the information that Alice needs to send to Bob in order to extract entanglement from the pbit ρ_{aAbB} is essentially classical. Bob just needs to know one classical bit of information to distinguish the two hiding states. This suggests that classical error correction might help Alice send this information to Bob, even when the channel is very noisy. The intuition behind our proof is that a simple classical repetition code suffices. Instead of the pbit ρ_{aAbB}, we use a pbit

$$\frac{1}{2}\Big(|\phi^+\rangle\langle\phi^+|_{\mathsf{ab}} \otimes \sigma^+_{\mathsf{A}_1\mathsf{B}_1} \otimes \cdots \otimes \sigma^+_{\mathsf{A}_N\mathsf{B}_N} + |\phi^-\rangle\langle\phi^-|_{\mathsf{ab}} \otimes \sigma^-_{\mathsf{A}_1\mathsf{B}_1} \otimes \cdots \otimes \sigma^-_{\mathsf{A}_N\mathsf{B}_N}\Big) \tag{5}$$

that contains N copies of the shield. For Bob to distinguish the hiding states, it suffices for a single copy to make it through the erasure channel. Alice now tries to send all of the copies of the shield through many uses of the erasure channel. However high the erasure probability, the probability that at least one will get through becomes arbitrarily close to one for sufficiently many attempts.

Making the above intuition rigorous is non-trivial: first, we must prove that the coherent information of n uses of the channel is strictly zero, for any input to the channel (not just the input states from the above intuition). To this end, we cannot directly use a pbit with N-copy shield of the form given above, as it would have distillable entanglement. We must instead adapt an approximate pbit construction from [23]. However, we must then take this approximation into account in the proof that the channel does have capacity. This requires a delicate analysis of the various parameters of our channel to show that both of the desired properties can hold simultaneously, which requires a more technical argument described in the Methods section (with full technical details in the Supplementary Note 1).

DISCUSSION

A natural question, which we leave open, is whether a stronger form of the result holds, which gives a constant upper bound on the channel dimension. It is even conceivable that the presence of quantum capacity is undecidable, which would imply the stronger form of result mentioned. It would also be interesting to see if one can obtain a result analogous to ours for the private capacity of quantum channels.

METHODS

Channel description

First, let us give a more precise description of our channel. The erasure channel with erasure probability p is

$$\mathcal{E}_p^{\mathsf{A}\to\mathsf{FB}} := (1-p)|0\rangle\langle 0|^{\mathsf{F}} \otimes \mathcal{I}^{\mathsf{A}\to\mathsf{B}} + p|1\rangle\langle 1|^{\mathsf{F}} \otimes \mathbb{1}^{\mathsf{B}}/\dim(\mathsf{B}), \tag{6}$$

where $\mathcal{I}^{\mathsf{A}\to\mathsf{B}}$ is the identity channel from A to B, and F is the erasure flag. The channel $\Gamma^{\mathsf{aA}\to\mathsf{bB}}$ will belong to

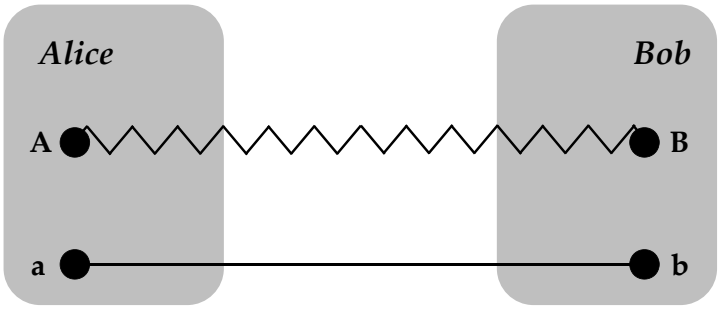

FIG. 2. Pbit representation. A pbit is a bipartite state with subsystems ab called the "key" and AB called the "shield". One party, Alice, holds the subsystems aA and the other party, Bob, holds the subsystems bB.

the class of PPT entanglement-binding channels whose Choi-Jamiołkowski state is an approximate pbit [23]. We show that Γ can be constructed with $\mathsf{A} := \mathsf{A}_1 \dots \mathsf{A}_N$ and $\mathsf{B} := \mathsf{B}_1 \dots \mathsf{B}_N$ consisting of N parts, such that even if Bob only receives part A_i of Alice's shield for any i, they obtain close to one ebit of one-way distillable entanglement. With the shorthand $\tilde{\mathsf{A}} := \mathsf{aA}$, and $\tilde{\mathsf{B}} := \mathsf{bB}$, let $\tilde{\Gamma}_\kappa$ be a noisy version of the channel Γ. More precisely, a composition of Γ with an erasure channel:

$$\tilde{\Gamma}_\kappa^{\tilde{\mathsf{A}}\to\mathsf{F}\tilde{\mathsf{B}}} := \mathcal{E}_\kappa^{\tilde{\mathsf{B}}\to\mathsf{F}\tilde{\mathsf{B}}} \circ \Gamma^{\tilde{\mathsf{A}}\to\tilde{\mathsf{B}}}. \tag{7}$$

Our construction uses channels of the form

$$\mathcal{M}^{\mathsf{S}\tilde{\mathsf{A}}\to\mathsf{S}\mathsf{F}\tilde{\mathsf{B}}} := \mathcal{P}_0^{\mathsf{S}\to\mathsf{S}} \otimes \tilde{\Gamma}_\kappa^{\tilde{\mathsf{A}}\to\mathsf{F}\tilde{\mathsf{B}}} + \mathcal{P}_1^{\mathsf{S}\to\mathsf{S}} \otimes \mathcal{E}_p^{\tilde{\mathsf{A}}\to\mathsf{F}\tilde{\mathsf{B}}}. \tag{8}$$

Here $\mathcal{P}_i^{\mathsf{S}\to\mathsf{S}}$ projects onto the i-th computational basis vector of the qubit system S which thereby acts as a classical switch allowing Alice to choose whether the channel acts as $\mathcal{E}_p$ or $\tilde{\Gamma}_\kappa$ on the main input $\tilde{\mathsf{A}}$. S is retained in the output, which lets Bob learn which choice was made.

Proof outline

We now state and outline the proof of our main result – for any number of channel uses there exists a channel with positive capacity but zero coherent information. Formally, we prove the following:

Theorem. Let $\mathcal{M}$ be the channel defined in Eq. (8). For any positive integer n, if $\kappa \in (0, 1/2)$ and $p \in [(1+\kappa^n)^{-1/n}, 1]$ then we can choose N and Γ such that:

1. $Q^{(n)}(\mathcal{M}) = 0$ and
2. $Q(\mathcal{M}) > 0$.

The proof is divided in two parts. We first prove that, given n and κ, for any Γ with zero capacity there is a range of p that makes the coherent information of $\mathcal{M}^{\otimes n}$ zero. In the second part we prove that there exists Γ with zero capacity such that $\mathcal{M}$ has positive capacity.

For the first part we can simplify the analysis of $\mathcal{M}^{\otimes n}$ by showing that it is optimal to make a definite choice (that is, a computational basis state input) for each of the n switch registers. For each possible setting of the n switches, the coherent information is a convex combination of the coherent information for three cases, weighted by their probabilities: every channel erases, all of the $\mathcal{E}_p$ erase but not all $\tilde{\Gamma}$ erase and at least one of the $\mathcal{E}_p$ does not erase (and therefore acts as the identity channel). The coherent information for second and third cases can be upper bounded respectively by zero and $H(\mathsf{R})$, where R is a system that purifies the input. For the first case it is bounded above by $-H(\mathsf{R})$. Weighting by the probabilities, we find that the total coherent information is upper-bounded by $\big(1-(1+\kappa^n)p^n\big)H(\mathsf{R})$. For any n and κ we can therefore find p such that this upper-bound is zero.

To prove the second part, we show that for fixed κ, p we can find a Γ with an N-copy shield such that the coherent information of $N+1$ uses of the channel $\mathcal{M}$ is positive for some $N+1 > n$. We number the channel uses $0, \dots, N$ and label the systems involved in the i-th use of the channel with superscript i. Consider the following input. The switch registers are set to select $\tilde{\Gamma}_\kappa$ for use 0 and $\mathcal{E}_p$ for the remaining uses $1, \dots, N$. We maximally entangle subsystem A_i^0 of $\tilde{\mathsf{A}}^0$ (which is acted on by $\tilde{\Gamma}_\kappa$) with subsystem A_1^i of $\tilde{\mathsf{A}}^i$ (acted on by an erasure channel). We also maximally entangle subsystem a^0 of $\tilde{\mathsf{A}}^0$ with a purifying reference system a which is retained by Alice. The remaining input subsystems are set to an arbitrary pure state. The resulting coherent information is a convex combination of cases where $\tilde{\Gamma}_\kappa$ erases, $\tilde{\Gamma}_\kappa$ does not erase but all the $\mathcal{E}_p$ erase, and $\tilde{\Gamma}_\kappa$ and at least one $\mathcal{E}_p$ do not erase. The first case contributes coherent information -1 weighted by its probability κ. The second case contributes approximately zero coherent information (due to a standard property of pbits). In the third case, after channel use 0, Alice and Bob share the Choi-Jamiołkowski state of Γ on systems $\mathsf{ab}^0\mathsf{A}_1^1\mathsf{B}_1^0 \dots \mathsf{A}_1^N\mathsf{B}_N^0$, and after the N uses of $\mathcal{E}_p$ at least one of $\mathsf{A}_1^1 \dots \mathsf{A}_1^N$ reaches Bob unerased. They then share a state with close to one ebit of one-way distillable entanglement (coherent information $+1$). This contribution is weighted by the probability $(1-\kappa)(1-p^N)$. We show that for $p \in (0,1)$, $\kappa \in (0, 1/2)$, we can find a Γ with large enough N for which the overall coherent information is

positive, proving that $Q(\mathcal{M}) > 0$. Further mathematical details are given in the Supplementary Note 1.

ACKNOWLEDGEMENTS

DE and DP acknowledge financial support from the European CHIST-ERA project CQC (funded partially by MINECO grant PRI-PIMCHI-2011-1071) and from Comunidad de Madrid (grant QUITEMAD+-CM, ref. S2013/ICE-2801). TSC is supported by the Royal Society. MO acknowledges financial support from European Union under project QALGO (Grant Agreement No. 600700). SS acknowledges the support of Sidney Sussex College.

This work was made possible through the support of grant #48322 from the John Templeton Foundation. The opinions expressed in this publication are those of the authors and do not necessarily reflect the views of the John Templeton Foundation.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

CONTRIBUTIONS

DE, DPG and TSC discussed and worked jointly on this result in Madrid; DE, SS, WM, MO and TSC discussed and worked jointly on this work in Cambridge; all authors helped to write the article.

REFERENCES

[1] G. Smith and J. Yard, *Quantum Communication with Zero-Capacity Channels.* Science **321**, 1812–1815 (2008).

[2] M.B. Hastings, *Superadditivity of communication capacity using entangled inputs.* Nature Physics **5**, 255–257 (2009).

[3] S. Lloyd, *Capacity of the noisy quantum channel.* Phys. Rev. A **55**, 1613-1622 (1997).

[4] P. Shor (Lecture Notes, MSRI Workshop on Quantum Computation, 2002).

[5] I. Devetak, *The private classical capacity and quantum capacity of a quantum channel.* Information Theory, IEEE Transactions on **51**, 44–55 (2005).

[6] D. P. DiVincenzo, P. W. Shor, and J. A. Smolin, *Quantum-channel capacity of very noisy channels.* Phys. Rev. A **57**, 830–839 (1998).

[7] G. Smith and J. A. Smolin, *Degenerate Quantum Codes for Pauli Channels.* Phys. Rev. Lett. **98**, 030501 (2007).

[8] C. H. Bennett, D. P. DiVincenzo, and J. A. Smolin, *Capacities of Quantum Erasure Channels.* Phys. Rev. Lett. **78**, 3217–3220 (1997).

[9] T. S. Cubitt, M. B. Ruskai, and G. Smith, *The structure of degradable quantum channels.* Journal of Mathematical Physics **49**, 102104 (2008).

[10] P. Horodecki, M. Horodecki, and R. Horodecki, *Binding entanglement channels.* Journal of Modern Optics **47**, 347-354 (2000).

[11] W. K. Wootters and W. H. Zurek, *A single quantum cannot be cloned.* Nature **299**, 802–803 (1982).

[12] M. Horodecki, P. Horodecki, and R. Horodecki, *Mixed-State Entanglement and Distillation: Is there a "Bound" Entanglement in Nature?* Phys. Rev. Lett. **80**, 5239–5242 (1998).

[13] F. G. S. L. Brandão, J. Oppenheim, and S. Strelchuk, *When does noise increase the quantum capacity?* Phys. Rev. Lett. **108**, 040501 (2012).

[14] N. Alon, *The Shannon capacity of a union.* Combinatorica **18**, 301–310 (1998).

[15] T. S. Cubitt, J. Chen, and A. W. Harrow, *Superactivation of the asymptotic zero-error classical capacity of a quantum channel.* Information Theory, IEEE Transactions on **57**, 8114–8126 (2011).

[16] J. Chen, T. S. Cubitt, A. W. Harrow, and G. Smith, *Entanglement can completely defeat quantum noise.* Phys. Rev. Lett. **107**, 250504 (2011).

[17] T. S. Cubitt and G. Smith, *An extreme form of superactivation for quantum zero-error capacities.* Information Theory, IEEE Transactions on **58**, 1953–1961 (2012).

[18] M. E. Shirokov, *On channels with positive quantum zero-error capacity having no n-shot capacity.* 1407.8524.

[19] P. Hayden and A. Winter, *Counterexamples to the maximal p-norm multiplicativity conjecture for all $p > 1$.* Comm. Math. Phys. **284**, 263–280 (2008).

[20] A. Montanaro, *Weak multiplicativity for random quantum channels.* Comm. Math. Phys. **319**, 535–555 (2013).

[21] J. Watrous, *Many Copies May Be Required for Entanglement Distillation.* Phys. Rev. Lett. **93**, 010502 (2004).

[22] J. Oppenheim, *For Quantum Information, Two Wrongs Can Make a Right.* Science **321**, 1783-1784 (2008).

[23] K. Horodecki, M. Horodecki, P. Horodecki, and J. Oppenheim, *General Paradigm for Distilling Classical Key From Quantum States.* Information Theory, IEEE Transactions on **55**, 1898–1929 (2009).

[24] T. Eggeling, and R.F. Werner, Physical Review Letters *Hiding Classical Data in Multipartite Quantum States.* **89**, 097905 (2002).

[25] M. M. Wilde, *Quantum Information Theory.* (Cambridge University Press, 2013).

SUPPLEMENTARY NOTE 1

Preliminaries

In the following, each system Q is associated to a Hilbert space of finite dimension $\dim(\mathsf{Q})$, and the Hilbert space has an orthonormal computational basis $\{|i\rangle^{\mathsf{Q}} : i \in \{0, \ldots, \dim(\mathsf{Q}) - 1\}\}$. For any system Q, let $\tau^{\mathsf{Q}} := \mathbb{1}^{\mathsf{Q}}/\dim(\mathsf{Q})$ denote its maximally mixed state.

We write $\mathcal{N}^{\mathsf{A}\to\mathsf{B}}$ to denote a channel from A to B. Let ρ^{A} be an input state, ρ^{AR} its purification, and $\rho^{\mathsf{BR}} := \mathcal{N}^{\mathsf{A}\to\mathsf{B}}(\rho^{\mathsf{AR}})$. Then the coherent information of $\mathcal{N}^{\mathsf{A}\to\mathsf{B}}$ on input ρ^{A} is

$$I_{\mathrm{coh}}(\mathcal{N}^{\mathsf{A}\to\mathsf{B}}, \rho^{\mathsf{A}}) := I(\mathsf{R}\rangle\mathsf{B})_{\rho^{\mathsf{BR}}} := S(\rho^{\mathsf{B}}) - S(\rho^{\mathsf{BR}}), \tag{1}$$

where $S(\cdot)$ is the von Neumann entropy. $I(\mathsf{R}\rangle\mathsf{B})_{\rho^{\mathsf{BR}}}$ is the coherent information of the state ρ^{BR}. A trivial, but useful, upper bound for the coherent information is

$$I_{\mathrm{coh}}(\mathcal{N}, \rho^{\mathsf{A}}) \leq S(\rho^{\mathsf{A}}). \tag{2}$$

Let A and B be two systems of equal dimension, $\mathcal{I}^{\mathsf{A}\to\mathsf{B}}$ denote the identity channel between them, and F be a binary erasure flag. The total erasure channel $\mathcal{E}_1^{\mathsf{A}\to\mathsf{FB}}$ maps any input state to $|1\rangle\langle 1|^{\mathsf{F}} \otimes \tau^{\mathsf{B}}$, while $\mathcal{E}_p^{\mathsf{A}\to\mathsf{FB}} := (1-p)|0\rangle\langle 0|^{\mathsf{F}} \otimes \mathcal{I}^{\mathsf{A}\to\mathsf{B}} + p\mathcal{E}_1^{\mathsf{A}\to\mathsf{FB}}$ denotes the erasure channel with erasure probability p. For any number of uses of $\mathcal{E}_1$ and any input state ρ we have

$$I_{\mathrm{coh}}(\mathcal{E}_1^{\otimes n}, \rho) = -S(\rho). \tag{3}$$

For any register F, a flagged channel is of the form $\mathcal{N}^{\mathsf{A}\to\mathsf{FB}} = \sum_{i=0}^{\dim(\mathsf{F})-1} p_i |i\rangle\langle i|^{\mathsf{F}} \otimes \mathcal{N}_i^{\mathsf{A}\to\mathsf{B}}$ where each $\mathcal{N}_i$ is a quantum channel. An example is $\mathcal{E}_p^{\mathsf{A}\to\mathsf{FB}}$. For any flagged channel we have

$$I_{\mathrm{coh}}(\mathcal{N}^{\mathsf{A}\to\mathsf{FB}}, \rho^{\mathsf{A}}) = \sum_i p_i I_{\mathrm{coh}}(\mathcal{N}_i^{\mathsf{A}\to\mathsf{B}}, \rho^{\mathsf{A}}), \tag{4}$$

which follows easily from

$$I(\mathsf{R}\rangle\mathsf{BF})_{\sum_i p_i \rho_i^{\mathsf{RB}} \otimes |i\rangle\langle i|^{\mathsf{F}}} = \sum_i p_i I(\mathsf{R}\rangle\mathsf{B})_{\rho_i^{\mathsf{RB}}}. \tag{5}$$

For any $i \in \{0, \ldots, \dim(\mathsf{S}) - 1\}$, let $\mathcal{P}_i^{\mathsf{S}\to\mathsf{S}}$ denote the completely positive map $X^{\mathsf{S}} \mapsto |i\rangle\langle i|^{\mathsf{S}} X^{\mathsf{S}} |i\rangle\langle i|^{\mathsf{S}}$. A switched channel is a channel of the form $\sum_{i=0}^{\dim(\mathsf{S})-1} \mathcal{P}_i^{\mathsf{S}\to\mathsf{S}} \otimes \mathcal{N}_i^{\mathsf{A}\to\mathsf{B}}$ where each $\mathcal{N}_i$ is a quantum channel. The register S acts as a classical switch allowing the sender to choose between different channels $\mathcal{N}_i$ to be applied on the "main input" A to produce a state of the "main output" B. We will need the following simple lemma regarding switched channels. This result had been proved previously in [1].

Lemma 1. *For any switched channel,*

$$\max_{\rho^{\mathsf{SA}}} I_{\mathrm{coh}}(\mathcal{N}^{\mathsf{SA}\to\mathsf{SB}}, \rho^{\mathsf{SA}}) = \max_i \max_{\rho^{\mathsf{A}}} I_{\mathrm{coh}}(\mathcal{N}_i^{\mathsf{A}\to\mathsf{B}}, \rho^{\mathsf{A}}) \tag{6}$$

where $0 \leq i \leq \dim(\mathsf{S}) - 1$.

Proof. To see this, note that any purification ρ^{SAR} of ρ^{SA} can be written in the form

$$|\rho\rangle^{\mathsf{SAR}} = \sum_i \sqrt{p_i} |i\rangle^{\mathsf{S}} \otimes |\rho_i\rangle^{\mathsf{AR}}. \tag{7}$$

Here p_i is the probability that the switch is set to i, and $|\rho_i\rangle^{\mathsf{AR}}$ is a purification of the channel input state ρ_i^{A} conditioned on that setting. Conversely, given probabilities p_i and states ρ_i^{A} for each switch value, we can always find $|\rho\rangle^{\mathsf{SAR}}$ satisfying (7). Given this, we see that

$$\mathcal{N}^{\mathsf{SA}\to\mathsf{SB}}(\rho^{\mathsf{SAR}}) = \sum_i p_i |i\rangle\langle i|^{\mathsf{S}} \otimes \mathcal{N}_i^{\mathsf{A}\to\mathsf{B}}(\rho_i^{\mathsf{AR}}) \tag{8}$$

where $\rho_i^{\mathsf{AR}} := |\rho_i\rangle\langle\rho_i|^{\mathsf{AR}}$. Applying (4) to (8) it follows that

$$I_{\mathrm{coh}}(\mathcal{N}^{\mathsf{SA}\to\mathsf{SB}}, \rho^{\mathsf{SA}}) = \sum_i p_i I_{\mathrm{coh}}(\mathcal{N}_i^{\mathsf{A}\to\mathsf{B}}, \rho_i^{\mathsf{A}}) \tag{9}$$

$$\leq \sum_i p_i \max_{\rho_i^{\mathsf{A}}} I_{\mathrm{coh}}(\mathcal{N}_i^{\mathsf{A}\to\mathsf{B}}, \rho_i^{\mathsf{A}}) \tag{10}$$

$$\leq \max_i \max_{\rho_i^{\mathsf{A}}} I_{\mathrm{coh}}(\mathcal{N}_i^{\mathsf{A}\to\mathsf{B}}, \rho_i^{\mathsf{A}}) \tag{11}$$

which completes the proof. □

We will also require some basic facts about pbits ("private bits") [2], which we gather here. Given a bipartite system ab with $\dim \mathsf{a} = \dim \mathsf{b} = 2$ and a bipartite system AB with $\dim \mathsf{A} = \dim \mathsf{B}$, a perfect pbit with key ab and shield AB is a state γ^{abAB} of the form

$$\gamma^{\mathsf{abAB}} := U^{\mathsf{abAB}}\big(\phi^{\mathsf{ab}} \otimes \sigma^{\mathsf{AB}}\big)(U^\dagger)^{\mathsf{abAB}}, \tag{12}$$

where ϕ^{ab} is the projector onto $|\phi\rangle^{\mathsf{ab}} := \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)^{\mathsf{ab}}$, σ^{AB} is some mixed state, and

$$U^{\mathsf{abAB}} := \sum_{i,j=0}^{1} |i\rangle\langle i|^{\mathsf{a}} \otimes |j\rangle\langle j|^{\mathsf{b}} \otimes U_{ij}^{\mathsf{AB}} \tag{13}$$

is a twisting unitary controlled by the key ab and acting on the shield AB as some unitary U_{ij}^{AB} for each i and j. Note that, due to the form of ϕ^{ab} and U^{abAB}, we have

$$\gamma^{\mathsf{abAB}} = \frac{1}{2}\sum_{k,l=0}^{1} |k,k\rangle^{\mathsf{ab}}\langle l,l|^{\mathsf{ab}} \otimes U_{kk}^{\mathsf{AB}} \sigma^{\mathsf{AB}} (U_{ll}^\dagger)^{\mathsf{AB}}. \tag{14}$$

Let us define $U^{\mathsf{bAB}} := \sum_{j=0}^{1} |j\rangle\langle j|^{\mathsf{b}} \otimes U_{jj}^{\mathsf{AB}}$. If Bob has access to b and the whole shield AB then he can apply the unitary operation $(U^\dagger)^{\mathsf{bAB}}$ to these systems, yielding a 2-qubit maximally entangled state on ab. Therefore,

$$I(\mathsf{a}\rangle\mathsf{bAB})_{\gamma^{\mathsf{abAB}}} = I(\mathsf{a}\rangle\mathsf{bAB})_{\phi^{\mathsf{ab}}\otimes\sigma^{\mathsf{AB}}} = 1. \tag{15}$$

We will often use the quantum data processing inequality for coherent information [3]:

$$I_{\mathrm{coh}}(\mathcal{N}_1, \rho) \geq I_{\mathrm{coh}}(\mathcal{N}_2 \circ \mathcal{N}_1, \rho), \tag{16}$$

where $\mathcal{N}_2 \circ \mathcal{N}_1$ denotes the channel composed of $\mathcal{N}_1$ followed by $\mathcal{N}_2$. This implies, in particular, that the coherent information cannot increase under Bob's local operations. For example, consider again the scenario when Bob has the whole shield system AB of γ^{abAB}, and assume the following sequence of local maps is applied: first, the system A is discarded and replaced by the maximally mixed state τ^{A}, then the whole shield AB is discarded, and finally Bob dephases locally his key system in the standard basis. This gives the following sequence of inequalities:

$$I(\mathsf{a}\rangle\mathsf{bAB})_{\gamma^{\mathsf{abAB}}} \geq I(\mathsf{a}\rangle\mathsf{bAB})_{\gamma^{\mathsf{abB}}\otimes\tau^{\mathsf{A}}} \geq I(\mathsf{a}\rangle\mathsf{b})_{\gamma^{\mathsf{ab}}} \geq I(\mathsf{a}\rangle\mathsf{b})_{\delta^{\mathsf{ab}}} = 0, \tag{17}$$

obtained by repeated application of (16). Here the final state $\delta^{\mathsf{ab}} := (|00\rangle\langle 00| + |11\rangle\langle 11|)^{\mathsf{ab}}/2$ corresponds to a perfectly random classical bit shared between Alice and Bob, as can be seen from (14). The final equality in (17) is obtained by direct calculation and it represents the fact that the value of the shared random bit δ^{ab} is not private (it can be recovered from the discarded shield AB possessed by the environment).

Channel construction

Our construction is a switched channel

$$\mathcal{M}^{\mathsf{S}\tilde{\mathsf{A}}\to\mathsf{SF}\tilde{\mathsf{B}}} := \mathcal{P}_0^{\mathsf{S}\to\mathsf{S}} \otimes \tilde{\Gamma}_\kappa^{\tilde{\mathsf{A}}\to\mathsf{F}\tilde{\mathsf{B}}} + \mathcal{P}_1^{\mathsf{S}\to\mathsf{S}} \otimes \mathcal{E}_p^{\tilde{\mathsf{A}}\to\mathsf{F}\tilde{\mathsf{B}}} \tag{18}$$

where F is a qubit system called the "erasure flag", and we define $\tilde{\Gamma}_\kappa^{\tilde{\mathsf{A}}\to\mathsf{F}\tilde{\mathsf{B}}}$ to be the composite channel

$$\tilde{\Gamma}_\kappa^{\tilde{\mathsf{A}}\to\mathsf{F}\tilde{\mathsf{B}}} := \mathcal{E}_\kappa^{\tilde{\mathsf{B}}\to\mathsf{F}\tilde{\mathsf{B}}} \circ \Gamma^{\tilde{\mathsf{A}}\to\tilde{\mathsf{B}}}. \tag{19}$$

We will now describe the input and output systems of our channel $\mathcal{M}$. It depends on parameters $N, r, m \in \mathbb{N} := \{1, 2, \ldots\}$ and $p, \kappa, q \in [0, 1]$, where q is an implicit parameter of $\tilde{\Gamma}_\kappa$. Let $\tilde{\mathsf{A}} := \mathsf{aA}$, and $\tilde{\mathsf{B}} := \mathsf{bB}$. We call the two two-dimensional systems (qubits) ab the "key". We define composite system $\mathsf{A} := \{\mathsf{A}_i : i \in [N]\}$ for Alice, and $\mathsf{B} := \{\mathsf{B}_i : i \in [N]\}$ for Bob, where $[n] := \{1, \ldots, n\}$. We call AB the "shield" and call A_i "Alice's i-th share of the shield".

We define $\Gamma^{\tilde{\mathsf{A}}\to\tilde{\mathsf{B}}}$ by giving its Choi-Jamiołkowski state, which depends on the parameters N, r, m, and q. It is proportional to

$$\begin{aligned}
\zeta^{\mathsf{abAB}} := (|00\rangle\langle 00| + |11\rangle\langle 11|)^{\mathsf{ab}} \otimes \left(\bigotimes_{k=1}^{m} \left[\frac{q}{2}(\omega + \sigma)\right]\right)^{\mathsf{AB}} \\
+ (|00\rangle\langle 11| + |11\rangle\langle 00|)^{\mathsf{ab}} \otimes \left(\bigotimes_{k=1}^{m} \left[\frac{q}{2}(\omega - \sigma)\right]\right)^{\mathsf{AB}} \\
+ (|01\rangle\langle 01| + |10\rangle\langle 10|)^{\mathsf{ab}} \otimes \left(\bigotimes_{k=1}^{m} \left[(\tfrac{1}{2} - q)\sigma\right]\right)^{\mathsf{AB}},
\end{aligned} \tag{20}$$

where $\omega := \bigotimes_{i=1}^{N} \bigotimes_{j=1}^{r} \frac{1}{2}(\tau_+ + \tau_-)$, and $\sigma := \bigotimes_{i=1}^{N} \bigotimes_{j=1}^{r} \tau_+$ are the Eggeling-Werner data hiding states [4]. Here τ_+ and τ_- are the states proportional to the symmetric and anti-symmetric projectors of a $d \times d$-Hilbert space, respectively.

In Eq. (139) of [2], a state $\rho^{rec}_{(p,d,k)}$ is defined. Apart from p, d and k, it also implicitly depends on a parameter m, so we will denote it by $\rho^{rec}_{(p,d,k;m)}$. Our ζ^{abAB} is precisely $\rho^{rec}_{(q,d,rN;m)}$. From Sections X-A (in particular Lemma 5) and X-B of [2] we see that $\rho^{rec}_{(q,d,rN;m)}$ is PPT if

$$0 < q \leq 1/3 \quad \text{and} \quad \frac{1-q}{q} \geq \left(\frac{d}{d-1}\right)^{rN}. \tag{21}$$

Since a channel is PPT-binding if and only if the matrix of its Choi-Jamiołkowski is PPT, the same conditions suffice for Γ to be PPT-binding. This condition is key to our subsequent analysis.

We will now derive from [2] another important fact about ζ^{abAB}. Defining

$$\zeta^{\mathsf{abA}_1\mathsf{B}_1} := \mathrm{Tr}_{\mathsf{A}_2\mathsf{B}_2\cdots\mathsf{A}_N\mathsf{B}_N} \zeta^{\mathsf{abAB}}, \tag{22}$$

for an appropriate choice of parameters, $\zeta^{\mathsf{abA}_1\mathsf{B}_1}$ can be made arbitrarily close to a perfect pbit $\gamma^{\mathsf{abA}_1\mathsf{B}_1}$ with key ab and shield $\mathsf{A}_1\mathsf{B}_1$. In particular, we will use

Lemma 2. *Let $q := 1/3$ and $r := 2m + \lceil \log_2 m \rceil$. Then $\mu := \|\rho^{\mathsf{abA}_1\mathsf{B}_1} - \gamma^{\mathsf{abA}_1\mathsf{B}_1}\|_1 \leq 16 m^{1/2} 2^{-m/4}$ for some perfect pbit $\gamma^{\mathsf{abA}_1\mathsf{B}_1}$, where $\|\cdot\|_1$ denotes the trace norm.*

Proof. First note that the $\rho^{\mathsf{abA}_1\mathsf{B}_1}$ is simply $\rho^{rec}_{(q,d,r;m)}$. Adopting the notation of [2], let $\|A_{0011}\|_1$ be the norm of the upper right block of the matrix $\rho^{rec}_{(q,d,r;m)}$ expanded in the computational basis of the key system ab. In Proposition 4 of [2], it is shown that if $1/2 - \|A_{0011}\|_1 < \epsilon < 1/8$ then $\mu \leq \delta(\epsilon)$ for some function $\delta(\epsilon)$. The function δ is given in Eq. (70) of [2] as

$$\delta(\epsilon) := 2\big(8\sqrt{2\epsilon} + h(2\sqrt{2\epsilon})\big)^{1/2} + 2\sqrt{2\epsilon} \tag{23}$$

where $h(x) := -x \log_2 x - (1-x)\log_2(1-x)$ is the binary entropy function. Provided $0 \leq x \leq 1/2$, $h(x)$ is an increasing function of x and

$$h(x) \leq x \log_2\left(\frac{1}{x^2}\right). \tag{24}$$

In particular, if we assume that $0 < 2\sqrt{2\epsilon} < 1/2$, i.e.

$$0 < \epsilon < 1/32, \tag{25}$$

then $h(2\sqrt{2\epsilon}) \leq \sqrt{8\epsilon}\log_2 \frac{1}{8\epsilon}$ and thus

$$\delta(\epsilon) \leq 2\left(4\sqrt{8\epsilon} + \sqrt{8\epsilon}\log_2 \frac{1}{8\epsilon}\right)^{1/2} + \sqrt{8\epsilon}. \tag{26}$$

From Eq. (25) we also get $\log_2 \frac{1}{8\epsilon} > 1$. By inserting this extra factor next to $4\sqrt{8\epsilon}$ in Eq. (26) we obtain

$$\delta(\epsilon) \leq 2\left(5\sqrt{8\epsilon}\log_2 \frac{1}{8\epsilon}\right)^{1/2} + \sqrt{8\epsilon}. \tag{27}$$

We can upper bound the last term as $\sqrt{8\epsilon} < (\sqrt{8\epsilon})^{1/2} < (\sqrt{8\epsilon}\log_2 \frac{1}{8\epsilon})^{1/2}$ and the whole expression as

$$\delta(\epsilon) \leq 2^{5/2}\left(\sqrt{8\epsilon}\log_2 \frac{1}{8\epsilon}\right)^{1/2}. \tag{28}$$

Thus, we have the bound

$$\mu \leq 2^{5/2}\left(\sqrt{8\epsilon}\log_2 \frac{1}{8\epsilon}\right)^{1/2}. \tag{29}$$

Rearranging Eq. (142) in the proof of Theorem 6 of [2] we find $1/2 - \|A_{0011}\|_1 = \frac{1}{2}\left(1 - \frac{(1-2^{-r})^m}{1+\left(\frac{1-2q}{2q}\right)^m}\right)$. By omitting the factor $1/2$ we have

$$\frac{1}{2} - \|A_{0011}\|_1 = \frac{1}{2}\left(1 - \frac{(1-2^{-r})^m}{1+\left(\frac{1}{2q}-1\right)^m}\right) \tag{30}$$

$$< \frac{1+\left(\frac{1}{2q}-1\right)^m - (1-2^{-r})^m}{1+\left(\frac{1}{2q}-1\right)^m}. \tag{31}$$

Setting $q = 1/3$ and using

$$(1-x)^m \geq 1 - mx \tag{32}$$

for all $m \in \mathbb{N}$ and $x \in (0,1)$ leads to

$$\frac{1}{2} - \|A_{0011}\|_1 < \frac{1+2^{-m} - (1-2^{-r})^m}{1+2^{-m}} \tag{33}$$

$$< \frac{1+2^{-m} - (1-m2^{-r})}{1+2^{-m}} \tag{34}$$

$$= \frac{2^{-m} + m2^{-r}}{1+2^{-m}} \tag{35}$$

which is a decreasing function of r. Setting $r = 2m + \lceil \log_2 m \rceil$ we get

$$\frac{1}{2} - \|A_{0011}\|_1 < \frac{2^{-m} + m2^{-(2m+\log_2 m)}}{1+2^{-m}} \tag{36}$$

$$= \frac{2^{-m} + 2^{-2m}}{1+2^{-m}} = 2^{-m}. \tag{37}$$

Therefore, for any $m > 5$, substituting $\epsilon = 2^{-m}$ into (29) we obtain

$$\mu \leq 2^{5/2}\left(\sqrt{8\epsilon}\log_2 \frac{1}{8\epsilon}\right)^{1/2} \tag{38}$$

$$= 2^{5/2}\left(\sqrt{2^{3-m}}(m-3)\right)^{1/2} \tag{39}$$

$$\leq 16 \times 2^{-m/4} m^{1/2} \tag{40}$$

as desired. □

Main result

The proof of our main result is based on the following two key lemmas. The first proves the coherent information is zero up to n uses of the channel. The second proves that it is non-zero for some larger number of uses, hence the quantum capacity is positive.

Lemma 3. *If Γ is PPT-binding, then for $\kappa \in [0,1]$, $p \in [(1+\kappa^n)^{-1/n}, 1]$, the coherent information of n uses of the channel $\mathcal{M}$ is zero: $Q^{(n)}(\mathcal{M}) = 0$.*

Proof. Using (6) from Lemma 1 and the general fact that

$$\max_{\rho} I_{\mathrm{coh}}(\mathcal{N} \otimes \mathcal{M}, \rho) = \max_{\rho} I_{\mathrm{coh}}(\mathcal{M} \otimes \mathcal{N}, \rho), \tag{41}$$

we have $Q^{(n)}(\mathcal{M}) = \frac{1}{n} \max_{0 \le l \le n} I_l$ where l is the number of switches set to use $\tilde{\Gamma}_\kappa$ and

$$I_l := I_{\mathrm{coh}}\big(\tilde{\Gamma}_\kappa^{\otimes l} \otimes \mathcal{E}_p^{\otimes(n-l)}, \rho_l^{\tilde{\mathsf{A}}^1 \cdots \tilde{\mathsf{A}}^n}\big). \tag{42}$$

Here $\rho_l^{\tilde{\mathsf{A}}^1 \cdots \tilde{\mathsf{A}}^n}$ is an input state for n uses of the channel that maximises the RHS of (42), where $\tilde{\mathsf{A}}^i := \mathsf{a}^i \mathsf{A}_1^i \cdots \mathsf{A}_N^i$ is the main input system for the i-th use of the channel.

From the definition (19) and Eq. (4), we see that I_l can be written as a sum of 2^n terms, each corresponding to a possible setting of the n erasure flags. Hence

$$I_l \le \kappa^l p^{n-l}(-S(\rho_l)) + (1-\kappa^l)p^{n-l} I_{\mathrm{coh}}(\Gamma^{\otimes l} \otimes \mathcal{E}_1^{\otimes n-l}, \rho_l) + (1-p^{n-l})S(\rho_l). \tag{43}$$

The first term in this bound is the case where all n channel uses erase, and follows from (3). The second term upper bounds the cases where all of the $\mathcal{E}_p$ uses erase but not all of the $\tilde{\Gamma}_\kappa$ channels do, obtained via (16). The final term upper bounds the contribution from the remaining cases using the trivial bound from (2).

Using (16) and the fact that Γ is PPT-binding, we obtain $I_{\mathrm{coh}}(\Gamma^{\otimes l} \otimes \mathcal{E}_1^{\otimes n-l}, \rho_l) \le I_{\mathrm{coh}}(\Gamma^{\otimes n}, \rho_l) \le 0$ and thus we can drop the second term in (43):

$$I_l \le \big(-\kappa^l p^{n-l} + 1 - p^{n-l}\big) S(\rho_l) \le \big(1 - (1+\kappa^n)p^n\big) S(\rho_l), \tag{44}$$

where the second inequality follows from $p, \kappa \in [0,1]$. We find that $I_l \le 0$ provided that

$$p \ge (1+\kappa^n)^{-1/n}. \tag{45}$$

On the other hand, $I_l \ge 0$ since we can always choose ρ_l to be pure. This implies $I_l = 0$ and thus $Q^{(n)}(\mathcal{M}) = 0$, which completes the proof. □

Lemma 4. *For $p \in (0,1)$, $\kappa \in (0, 1/2)$, we can choose the parameters q, N, r, m, d such that the PPT condition (21) holds and $Q^{(N+1)}(\mathcal{M}) > 0$.*

Proof. Our proof has two parts. In part (i) we prove a lower bound on $Q^{(N+1)}(\mathcal{M})$ by analysing a particular input to the channel. In part (ii) we show that the channel parameters can be chosen to make this lower bound strictly positive while, at the same time, satisfying (21).

(i) We number the $N+1$ channel uses by $\{0, 1, \ldots, N\}$, and label the systems involved in the i-th channel use with superscript i. The switch systems are set so that the first use of the channel acts as $\tilde{\Gamma}_\kappa$ on its main input, and the remaining N uses act as $\mathcal{E}_p$.

If X and Y are two systems of equal dimensions, we use $\phi^{\mathsf{XY}} := |\phi\rangle\langle\phi|^{\mathsf{XY}}$ to denote the maximally entangled state on XY where $|\phi\rangle^{\mathsf{XY}} := \sum_{i=0}^{\dim(\mathsf{X})-1} |i\rangle^{\mathsf{X}} |i\rangle^{\mathsf{Y}} / \sqrt{\dim(\mathsf{X})}$. Alice prepares maximally entangled states on subsystems $\mathsf{a}^0\mathsf{a}$ and on $\mathsf{A}_i^0 \mathsf{A}_1^i$ for all $i \in [N]$. The purification of the overall input to the $N+1$ uses of the channel $\mathcal{M}$ is

$$|\nu\rangle := |0\rangle^{\mathsf{S}^0} |\phi\rangle^{\mathsf{a}\mathsf{a}^0} \bigotimes_{i=1}^{N} \left(|1\rangle^{\mathsf{S}^i} |\alpha\rangle^{\mathsf{a}^i} |\phi\rangle^{\mathsf{A}_i^0 \mathsf{A}_1^i} \bigotimes_{j=2}^{N} |\beta\rangle^{\mathsf{A}_j^i} \right) \tag{46}$$

where $|\alpha\rangle$ and $|\beta\rangle$ are arbitrary pure states, a is a reference system, and S^i and $\tilde{\mathsf{A}}^i = \mathsf{a}^i \mathsf{A}_1^i \cdots \mathsf{A}_N^i$ are the switch and main input systems for the i-th use of $\mathcal{M}$, respectively.

The switch settings cause the first use of $\mathcal{M}$ to act as $\tilde{\Gamma}_\kappa$ on $\tilde{\mathsf{A}}^0 = \mathsf{a}^0 \mathsf{A}_1^0 \cdots \mathsf{A}_N^0$ (see (19)). With probability κ, $\tilde{\Gamma}_\kappa$ erases, yielding $|1\rangle\langle 1|^{\mathsf{F}^0} \otimes \tau^{\tilde{\mathsf{B}}^0}$. With probability $1-\kappa$, it sets the erasure flag to $|0\rangle\langle 0|^{\mathsf{F}^0}$ and acts as Γ on $\tilde{\mathsf{A}}^0$, producing

$\tilde{\mathsf{B}}^0$. At this point the state of $\mathsf{a}\tilde{\mathsf{B}}^0\mathsf{A}_1^1\cdots\mathsf{A}_1^N$ is just the Choi-Jamiołkowski state ζ^{abAB} defined in (20) with its systems relabelled as follows: $\tilde{\mathsf{B}}\to\tilde{\mathsf{B}}^0$ and $\mathsf{A}_j\to\mathsf{A}_1^j$ for all $j\in[N]$. The switches are set so that the remaining N uses of $\mathcal{M}$ apply $\mathcal{E}_p$ to the each of the systems $\tilde{\mathsf{A}}^j$ for each $j\in[N]$. Bob now applies a simple post-processing operation to the output systems of all $N+1$ channel uses to obtain a state of a system $\mathsf{bA}_1'\mathsf{B}_1\mathsf{GF}^0$: he first measures the erasure flags $\mathsf{F}^1\cdots\mathsf{F}^N$. With probability $1-p^N$, at least one of these flags, say F^j, will be in the state $|0\rangle\langle 0|^{\mathsf{F}^j}$, and the state of A_1^j has been perfectly transferred to his system B_1^j. Otherwise, with probability p^N, Bob picks an arbitrary $j\in[N]$. In this case the state of F^j is $|1\rangle\langle 1|^{\mathsf{F}^j}$ and the state of B_1^j is maximally mixed and uncorrelated with any other system. Now, Bob transfers the state of F^j to a system G, the state of B_1^j to A_1', the state of B_j^0 to B_1 and b^0 to b. Bob then discards all of his systems except for $\mathsf{bA}_1'\mathsf{B}_1\mathsf{GF}^0$, which are now in the state

$$\begin{aligned}\eta^{\mathsf{abA}_1'\mathsf{B}_1\mathsf{GF}^0} &:= \kappa\tau^{\mathsf{a}}\otimes\sigma^{\mathsf{bA}_1'\mathsf{B}_1\mathsf{G}}\otimes|1\rangle\langle 1|^{\mathsf{F}^0}\\ &+(1-\kappa)p^N\zeta^{\mathsf{abB}_1}\otimes\tau^{\mathsf{A}_1'}\otimes|1\rangle\langle 1|^{\mathsf{G}}\otimes|0\rangle\langle 0|^{\mathsf{F}^0}\\ &+(1-\kappa)(1-p^N)\zeta^{\mathsf{abA}_1'\mathsf{B}_1}\otimes|0\rangle\langle 0|^{\mathsf{G}}\otimes|0\rangle\langle 0|^{\mathsf{F}^0}.\end{aligned}\tag{47}$$

Here $\zeta^{\mathsf{abA}_1'\mathsf{B}_1}:=\mathcal{I}^{\mathsf{A}_1\to\mathsf{A}_1'}(\zeta^{\mathsf{abA}_1\mathsf{B}_1})$, where $\zeta^{\mathsf{abA}_1\mathsf{B}_1}$ is the state (22) from Lemma 2. The details of $\sigma^{\mathsf{bA}_1'\mathsf{B}_1\mathsf{G}}$ are unimportant. The first term in (47) corresponds to the case where the first channel use erases. When the first use does not erase, the case where all other uses erase yields the second term, and the case where at least one of the other uses does not erase gives the third term.

Let us call Bob's post-processing operation $\mathcal{P}$. Using the state $\nu:=|\nu\rangle\langle\nu|$ from (46), we can write

$$(N+1)Q^{(N+1)}(\mathcal{M})\geq I_{\mathrm{coh}}(\mathcal{M}^{\otimes N+1},\nu)\geq I_{\mathrm{coh}}(\mathcal{P}\circ\mathcal{M}^{\otimes N+1},\nu)=I(\mathsf{a}\rangle\mathsf{bA}_1'\mathsf{B}_1\mathsf{GF}^0)_{\eta^{\mathsf{abA}_1'\mathsf{B}_1\mathsf{GF}^0}},\tag{48}$$

where the composition property (16) was used. Given the "flagged" structure of (47), we can use (5):

$$(N+1)Q^{(N+1)}(\mathcal{M})\geq\kappa I(\mathsf{a}\rangle\mathsf{bA}_1'\mathsf{B}_1\mathsf{G})_{\tau^{\mathsf{a}}\otimes\sigma^{\mathsf{bA}_1'\mathsf{B}_1\mathsf{G}}}+(1-\kappa)p^N I(\mathsf{a}\rangle\mathsf{bA}_1'\mathsf{B}_1)_{\zeta^{\mathsf{abB}_1}\otimes\tau^{\mathsf{A}_1'}}+(1-\kappa)(1-p^N)I(\mathsf{a}\rangle\mathsf{bA}_1'\mathsf{B}_1)_{\zeta^{\mathsf{abA}_1'\mathsf{B}_1}}.\tag{49}$$

The first term is $-\kappa S(\tau^{\mathsf{a}})=-\kappa$. If $\mu=\|\zeta^{\mathsf{abA}_1'\mathsf{B}_1}-\gamma^{\mathsf{abA}_1'\mathsf{B}_1}\|_1$ for some perfect pbit $\gamma^{\mathsf{abA}_1'\mathsf{B}_1}$, then by the monotonicity of the trace distance under CPTP maps

$$\mu\geq\|\zeta^{\mathsf{abB}_1}\otimes\tau^{\mathsf{A}_1'}-\gamma^{\mathsf{abB}_1}\otimes\tau^{\mathsf{A}_1'}\|_1.\tag{50}$$

In what follows, we will use the Alicki-Fannes inequality [5]. This states that for ρ^{RB} and σ^{RB} such that $\mu:=\|\rho^{\mathsf{RB}}-\sigma^{\mathsf{RB}}\|_1<1$ we get

$$\left|I(\mathsf{R}\rangle\mathsf{B})_{\rho^{\mathsf{RB}}}-I(\mathsf{R}\rangle\mathsf{B})_{\sigma^{\mathsf{RB}}}\right|\leq 4\mu\log_2\dim(\mathsf{R})+2h(\mu).\tag{51}$$

Using (50) and properties (15), (17), $\dim(\mathsf{a})=2$ together with the Alicki-Fannes inequality we have

$$I(\mathsf{a}\rangle\mathsf{bA}_1'\mathsf{B}_1)_{\zeta^{\mathsf{abA}_1'\mathsf{B}_1}}\geq 1-\Delta,\tag{52}$$

$$I(\mathsf{a}\rangle\mathsf{bA}_1'\mathsf{B}_1)_{\zeta^{\mathsf{abB}_1}\otimes\tau^{\mathsf{A}_1'}}\geq-\Delta,\tag{53}$$

where

$$\Delta:=4\mu+2h(\mu).\tag{54}$$

Therefore, $(N+1)Q^{(N+1)}(\mathcal{M})\geq(1-\kappa)(1-p^N-\Delta)-\kappa$ which is strictly positive if

$$\Delta<1-p^N-\frac{\kappa}{1-\kappa}.\tag{55}$$

(ii) We will now show how the parameters must be chosen. First, to ensure that (21) is satisfied, we specify that $d:=2Nr$ and $q:=1/3$. Now, if $\kappa\in(0,1/2)$ then $\kappa/(1-\kappa)\in(0,1)$, so for any $p\in(0,1)$ we can always choose N large enough to make the RHS of (55) positive. Fixing this value of N, we then must choose m and r to make Δ small enough to satisfy (55). Lemma 2 tells us that with $q=1/3$ and $r=2m+\log_2 m$, we have $\mu\leq 16m^{1/2}2^{-m/4}$. Recall that $\Delta=4\mu+2h(\mu)$, into which we are substituting $\mu\leq 16m^{1/2}2^{-m/4}$. Provided $0\leq x\leq 1/2$, $h(x)$ is an increasing function of x, and $h(x)\leq 2x\log_2\frac{1}{x}$, so $h(\mu)\leq h(16m^{1/2}2^{-m/4})\leq 4m^{3/2}2^{-m/4}$ (provided $16m^{1/2}2^{-m/4}\leq 1/2$). We get

$$\Delta\leq 64m^{1/2}2^{-m/4}+8m^{3/2}2^{-m/4}\leq 72\times 2^{-m/4}m^{3/2}.\tag{56}$$

One can choose m to make this as small as required. □

We can now prove our main result:

Theorem. *Let $\mathcal{M}$ be the channel defined in Eq. (18). For any positive integer n, if $\kappa \in (0, 1/2)$ and $p \in [(1+\kappa^n)^{-1/n}, 1]$ then we can choose q, d, r, N, m such that:*

1. $Q^{(n)}(\mathcal{M}) = 0$ *and*
2. $Q(\mathcal{M}) > 0$.

Proof. In Lemma 3 we show that if Γ is PPT-binding and κ, p satisfy $\kappa \in [0, 1]$ and $p \in [(1+\kappa^n)^{-1/n}, 1]$, then the first statement holds. In Lemma 4 we show that for any $\kappa \in (0, 1/2)$ and $p \in (0, 1)$, we can choose the parameters q, d, r, N, m so that the second statement holds and Γ is PPT-binding. Therefore, for (κ, p) in the intersection of the two regions, the channel $\mathcal{M}$ satisfies both statements. □

To be concrete, we can choose $\kappa = 1/4$ (so that $\kappa/(1-\kappa) = 1/3$) and choose $p = (1+\kappa^n)^{-1/n}$. We can then choose N so that $1 - p^N \geq 2/3$: we require $(1+\kappa^n)^{-N/n} < 1/3$. Taking logs of both sides, rearranging and using $x/\ln(2) \geq \log_2(1+x)$, we have $N > (\log_2 3)(\ln 2) n 4^n$, so let us take $N = 2n4^n$. We must now choose m large enough ($m \geq 68$) that $\Delta < 1/3$ in (56).

SUPPLEMENTARY REFERENCES

[1] M. Fukuda and M. M. Wolf, *Simplifying additivity problems using direct sum constructions.* Journal of Physics A: Mathematical and General **48**, 072101 (2007).

[2] K. Horodecki, M. Horodecki, P. Horodecki, and J. Oppenheim, *General Paradigm for Distilling Classical Key From Quantum States.* Information Theory, IEEE Transactions on **55**, 1898–1929 (2009).

[3] M. M. Wilde, *Quantum Information Theory.* (Cambridge University Press, 2013).

[4] T. Eggeling, and R.F. Werner, Physical Review Letters *Hiding Classical Data in Multipartite Quantum States.* **89**, 097905 (2002).

[5] R. Alicki and M. Fannes, *Continuity of quantum conditional information.* Journal of Physics A: Mathematical and General **37**, L55–L57 (2004).