Quantum Coding with Finite Resources

Marco Tomamichel,[1] Mario Berta,[2] and Joseph M. Renes[3]

[1]*School of Physics, University of Sydney, Sydney, NSW 2006, Australia*
[2]*Institute for Quantum Information and Matter, Caltech, Pasadena, CA 91125, USA*
[3]*Institute for Theoretical Physics, ETH Zurich, 8093 Zürich, Switzerland*

The quantum capacity of a memoryless channel is often used as a single figure of merit to characterize its ability to transmit quantum information coherently. The capacity determines the maximal rate at which we can code reliably over asymptotically many uses of the channel. We argue that this asymptotic treatment is insufficient to the point of being irrelevant in the quantum setting where decoherence severely limits our ability to manipulate large quantum systems in the encoder and decoder. For all practical purposes we should instead focus on the trade-off between three parameters: the rate of the code, the number of coherent uses of the channel, and the fidelity of the transmission. The aim is then to specify the region determined by allowed combinations of these parameters.

Towards this goal, we find approximate and exact characterizations of the region of allowed triplets for the qubit dephasing channel and for the erasure channel with classical post-processing assistance. In each case the region is parametrized by a second channel parameter, the quantum channel dispersion. In the process we also develop several general inner (achievable) and outer (converse) bounds on the coding region that are valid for all finite-dimensional quantum channels and can be computed efficiently. Applied to the depolarizing channel, this allows us to determine a lower bound on the number of coherent uses of the channel necessary to witness super-additivity of the coherent information.

I. INTRODUCTION

One of the quintessential topics in quantum information theory is the study of reliable quantum information transmission over a noisy quantum channel. Here the word "channel" simply refers to a description of a physical evolution (by means of a completely positive trace-preserving map on density operators). Traditionally one considers point-to-point communication settings where a memoryless channel can be used many times in sequence. The sender (often called Alice) first encodes a quantum state into a sequence of registers and then sends them one by one trough the channel to the receiver (often called Bob). Bob collects these registers and then attempts to decode the quantum state. Alternatively, consider a collection of physical qubits that are exposed to independent noise. The goal is then to encode quantum information (logical qubits) into this system so that the quantum information can be decoded with high fidelity at a later stage. One of the primary goals of information theory is to find fundamental limits imposed on any coding scheme that tries to accomplish such tasks.

Following a tradition going back to Shannon's groundbreaking work [32], this problem is usually studied asymptotically: the *quantum capacity* of a channel is defined as the *optimal rate* (in qubits per use of the channel) at which we can transmit quantum information *with vanishing error* as the number of channel uses *goes to infinity*. In the context of information storage, the rate simply corresponds to the ratio of logical to physical qubits, and the number of channel uses corresponds to the number of physical qubits. The quantum capacity of arbitrary channels has been determined in a series of works, an upper bound to the capacity shown in [2, 3, 31] and achievability of that bound shown in [9, 22, 33].

However, in any application of the theory resources are *finite* and the number of channel uses is necessarily limited. More importantly, at least for the near future it appears unrealistic to expect that encoding and decoding circuits can coherently manipulate large numbers of qubits. Restricting the size of the quantum devices used for encoding the channel inputs and decoding its outputs is

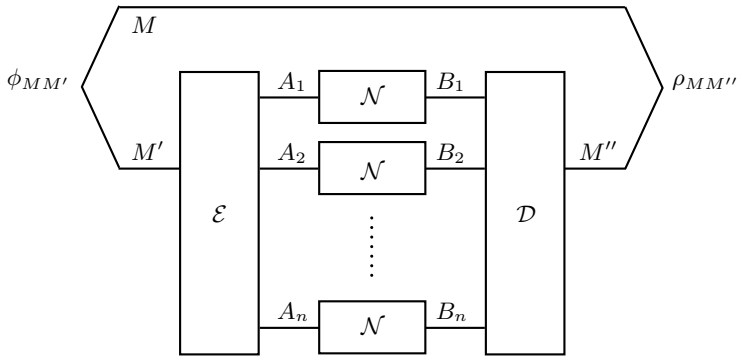


FIG. 1: Coding Scheme for entanglement transmission over n uses of a channel $\mathcal{N} \equiv \mathcal{N}_{A\to B}$. The systems M, M' and M'' are isomorphic. The encoder $\mathcal{E} \equiv \mathcal{E}_{M'\to A^n}$ encodes the part M' of the maximally entangled state $\phi_{MM'}$ into the channel input systems. Later, the decoder $\mathcal{D} \equiv \mathcal{D}_{B^n\to M''}$ recovers the state from the channel output systems. The performance of the code is measured using the fidelity $F(\phi_{MM'}, \rho_{MM''})$.

tantamount to considering communication with only a fixed number of channel uses. This then raises the question whether an asymptotic approach where this number goes to infinity — which has proven to be very successful for the analysis of classical communication systems — is equally suitable for the quantum setting. Clearly, what we really want to understand is how well we can transmit quantum information in a realistic setting where the number of channel uses and the size of quantum devices is limited. The quantum capacity is at most a proxy for the answer to this question, and in this article we argue that it is often not a very good one.

The study of such non-asymptotic scenarios has recently garnered significant attention in classical information theory [16, 27, 38] as well as in quantum information theory [8, 21, 40, 41]. Here we extend these considerations to the setting of quantum communication.

Outline: The remainder of the paper is structured as follows. In Section II we discuss our main results detail. In Section III we formally introduce relevant notation and information measures. In Section IV A we derive our converse (outer) bounds and in Section IV B we derive our achievability (inner) bounds. Finally, Section V discusses the specific examples presented below as Results 1, 2, and 3.

II. DISCUSSION OF RESULTS

In this work we focus on codes enabling a state entangled with a reference system to be reliably transmitted through the channel. This is a strong requirement: reliable entanglement transmission implies reliable transmission, on average, of all non-entangled input states. The coding scheme is depicted in Figure 1. We are given a quantum channel $\mathcal{N} \equiv \mathcal{N}_{A\to B}$ and denote by $\mathcal{N}^{\otimes n}$ the n-fold parallel repetition of this channel. An *entanglement transmission code* for $\mathcal{N}^{\otimes n}$ is given by a triplet $\{|M|, \mathcal{E}, \mathcal{D}\}$, where $|M|$ is the local dimension of a maximally entangled state $|\phi\rangle_{MM'}$ that is to be transmitted over $\mathcal{N}^{\otimes n}$. The quantum channels $\mathcal{E} \equiv \mathcal{E}_{M'\to A^n}$ and $\mathcal{D} \equiv \mathcal{D}_{B^n\to M''}$ are encoding and decoding operations, respectively. (A more formal treatment will follow in Section III B.) With this in hand, we now say that a triplet $\{R, n, \varepsilon\}$ is *achievable* on the channel $\mathcal{N}$ if there exists an entanglement transmission code satisfying

$$\frac{1}{n}\log|M| \geq R \qquad \text{and} \qquad F\Big(\phi_{MM'},\, (\mathcal{D}\circ\mathcal{N}^{\otimes n}\circ\mathcal{E})(\phi_{MM'})\Big) \geq 1-\varepsilon\,. \tag{1}$$

Here, R is the *rate* of the code, n is the number of channel uses, and ε is the tolerated error measured in terms of the fidelity F.

The non-asymptotic *achievable region* of a quantum channel $\mathcal{N}$ is then given by the union of all achievable triplets $\{R, n, \varepsilon\}$. The goal of (non-asymptotic) information theory is to find tight bounds on this achievable region, in particular to determine if certain triplets are outside the achievable region and thus *forbidden*. For this purpose, we define its boundary

$$\hat{R}_{\mathcal{N}}(n;\varepsilon) := \max\left\{R : (R, n, \varepsilon) \text{ is achievable on } \mathcal{N}\right\}, \tag{2}$$

and investigate it as a function of n for a fixed value of ε.[1] We will often drop the subscript $\mathcal{N}$ if it is clear which channel is considered.

To begin, let us rephrase the seminal capacity results [2, 3, 11, 22, 31, 33] in this language. The quantum capacity is defined as the asymptotic limit of $\hat{R}_{\mathcal{N}}(n;\varepsilon)$ when n (first) goes to infinity and ε vanishes. The capacity can be expressed in terms of a regularized coherent information:

$$Q(\mathcal{N}) := \lim_{\varepsilon\to 0}\lim_{n\to\infty} \hat{R}_{\mathcal{N}}(n;\varepsilon) = \lim_{\ell\to\infty} \frac{I_{\mathrm{c}}\left(\mathcal{N}^{\otimes\ell}\right)}{\ell}, \tag{3}$$

where the coherent information I_{c} will be defined in Section IV B. This result is highly unsatisfactory, not least because the regularization makes its computation intractable.[2] Worse, the statement is not as strong as we would like it to be because it does not give any indication of the fundamental limits for finite ε or finite n.

For example, even sticking to the asymptotic limit for now, we might be willing to admit a small but nonzero error in our recovery. Formally, instead of requiring that the error vanishes asymptotically, we only require that it does not exceed a certain threshold, ε. Can we then achieve a higher asymptotic rate in the above sense? Surprisingly, the answer to this question is not known in general. Recent work [42] at least settles the question in the negative for a class of generalized dephasing channels and in particular for the qubit dephasing channel

$$\mathcal{Z}_\gamma : \rho \mapsto (1-\gamma)\rho + \gamma Z\rho Z, \tag{4}$$

where $\gamma \in [0,1]$ is a parameter and Z is the Pauli Z operator. Dephasing channels are particularly interesting examples because dephasing noise is dominant in many physical implementations of qubits. The results of [42] thus allow us to fully characterize the achievable region in the limit $n \to \infty$ for such channels, and in particular ensure that

$$\lim_{n\to\infty} \hat{R}_{\mathcal{Z}_\gamma}(n;\varepsilon) = I_c(\mathcal{Z}_\gamma), \tag{5}$$

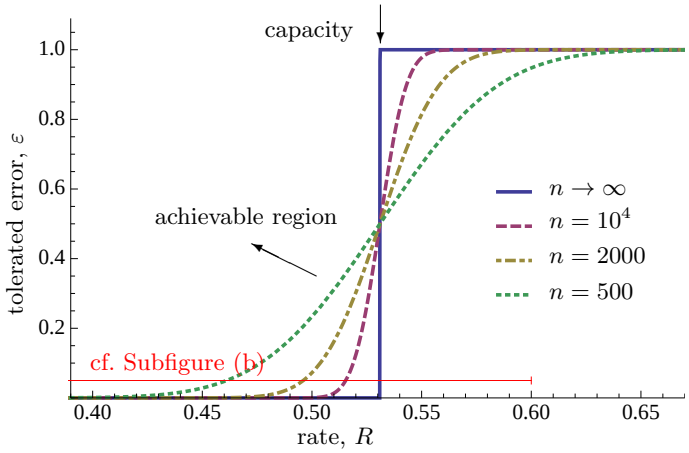
independent of the value of $\varepsilon \in (0,1)$. Note also that the regularization is not required here since these channels are degradable [10].

Here we go beyond studying the problem in the asymptotic limit and develop characterizations of the achievable region for finite values of n. We find inner (achievability) and outer (converse) bounds on the boundary of the achievable region. These do not agree for general channels (which is unsurprising given the fact that such an agreement has not even been established asymptotically for nonzero error), but they do coincide for certain important examples.

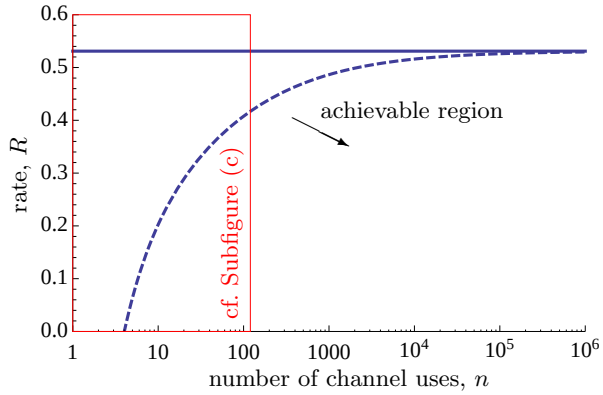
---

[1] An alternative approach would be to investigate the boundary $\hat{\varepsilon}_{\mathcal{N}}(n;R) := \max\{\varepsilon : (R, n, \varepsilon) \text{ is achievable}\}$. This leads to the study of error exponents (and the reliability function) as well as strong converse exponents. We will not discuss this here since such an analysis usually does not yield sufficiently tight bounds for small values of n.

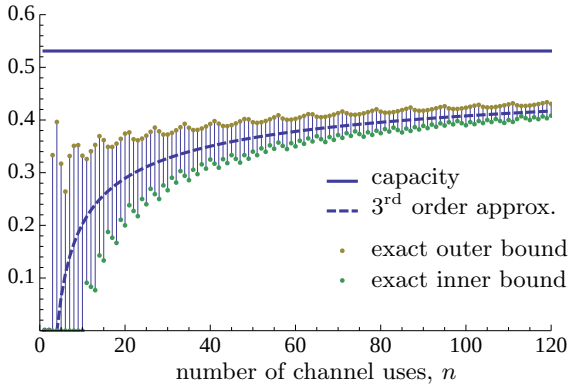
[2] It is not clear if the limit $\ell \to \infty$ is necessary for any fixed channel, but it was recently shown that there does not exist a universal constant ℓ_0 such that $C(\mathcal{N}) \leq \frac{1}{\ell_0} I_{\mathrm{c}}(\mathcal{N}^{\ell_0})$ for all channels $\mathcal{N}$ [7].



(a) Boundary of the achievable region for different values of n (second order approximation).



(b) Boundary of the achievable region for $\varepsilon = 5\%$ (third order approximation in (6)).



(c) Comparison of strict bounds with third order approximation for $\varepsilon = 5\%$.

FIG. 2: Approximation of the non-asymptotic achievable rate region of a qubit dephasing channel with $\gamma = 0.1$ (see Result 1).

A. Qubit Dephasing Channel

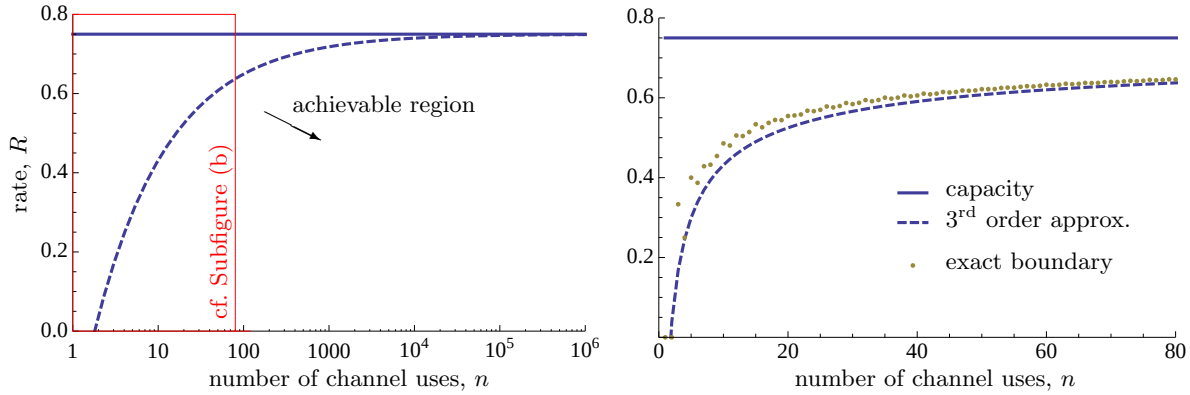
The first example is the qubit dephasing channel. Building on recent work that established the strong converse for this channel [42], we will show that its non-asymptotic achievable region is equivalent to the corresponding region of a (classical) binary symmetric channel. This allows us to employ results from classical information theory and establish the following characterization of the achievable region for the qubit dephasing channel.

Result 1. *For the qubit dephasing channel $\mathcal{Z}_\gamma$ with $\gamma \in [0,1]$, the boundary $\hat{R}(n;\varepsilon)$ satisfies*

$$\hat{R}(n;\varepsilon) = 1 - h(\gamma) + \sqrt{\frac{v(\gamma)}{n}}\,\Phi^{-1}(\varepsilon) + \frac{\log n}{2n} + O\Big(\frac{1}{n}\Big)\,, \tag{6}$$

where Φ is the cumulative standard Gaussian distribution, $h(\gamma) := -\gamma\log\gamma - (1-\gamma)\log(1-\gamma)$ denotes the binary entropy and $v(\gamma)$ the corresponding variance, $v(\gamma) := \gamma(\log\gamma + h(\gamma))^2 + (1-\gamma)(\log(1-\gamma) + h(\gamma))^2$.

The expression without the remainder term $O(\frac{1}{n})$ is called the third order approximation of the (boundary of the) non-asymptotic achievable region. It is visualized in Figures 2 for an example



(a) Boundary of the achievable region.

(b) Comparison of exact bounds with third order approximation.

FIG. 3: Approximation of the non-asymptotic achievable rate region (with classical post-processing assistance) of a qubit erasure channel with $\beta = 0.25$ and error parameter $\varepsilon = 1\%$ (see Result 2).

channel with $\gamma = 0.1$. In Figure 2(a) we plot the smallest achievable error ε as a function of the rate R. Here we use the second order expansion without the term $\frac{1}{2n}\log n$ since it can conveniently be solved for ε. In the limit $n \to \infty$ we see an instantaneous transition of ε from 0 to 1, the signature of a strong converse: coding below the capacity $Q(\mathcal{Z}_\gamma) = 1 - h(\gamma)$ is possible with perfect fidelity whereas coding above the capacity will necessarily result in a vanishing fidelity.

In Figure 2(b) we plot the third order approximation in (6) for the highest achievable rate, $\hat{R}(n;\varepsilon)$, as a function of n for a fixed fidelity of 95% (i.e. we set $\varepsilon = 5\%$). For example, this allows us to calculate how many times we need to use the channel in order to approximately achieve the quantum capacity. The third order approximation shows that we need approximately 850 channel uses to achieve 90% of the quantum capacity. Note that a coding scheme achieving this would probably require us to coherently manipulate 850 qubits in the decoder, which appears to be a quite challenging task. This example shows that the capacity does not suffice to characterize the ability of a quantum channel to transmit information, and further motivates the study of the achievable region for finite n.

Finally, we remark that the third order approximation is quite strong even for small n. To prove this we compare it to concrete upper and lower bounds on $\hat{R}(n;\varepsilon)$ in Figure 2(c) and see that the remainder term $O(\frac{1}{n})$ becomes negligible for fairly small $n \approx 100$ for the present values of γ and ε.

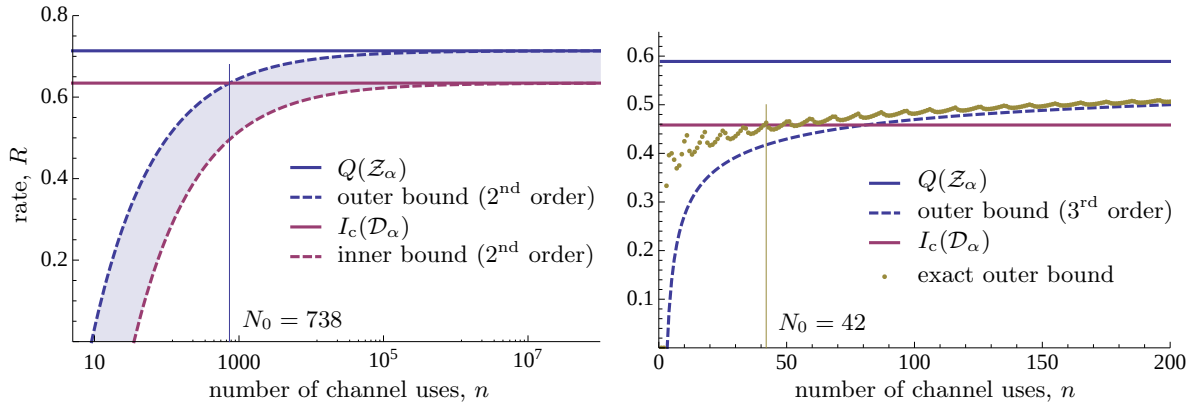
B. Qubit Erasure Channel

Another channel we can analyze in this manner is the qubit erasure channel, given by the map

$$\mathcal{E}_\beta : \rho \mapsto (1-\beta)\rho + \beta\,|e\rangle\langle e|\,, \tag{7}$$

where $\beta \in [0,1]$ is a parameter and $|e\rangle\langle e|$ is a pure state orthogonal to ρ. Here we investigate coding schemes that allow classical post-processing (cpp) between the sender and receiver after the quantum transmission (see also Figure 5 in Section III). We denote the corresponding boundary of the achievable region by $\hat{R}^{\text{cpp}}(n;\varepsilon)$. Since this includes all codes that do not take advantage of cpp, we clearly have $\hat{R}(n;\varepsilon) \leq \hat{R}^{\text{cpp}}(n;\varepsilon)$ for all channels.

Here we can determine the boundary $\hat{R}^{\text{cpp}}(n;\varepsilon)$ exactly, again by generalizing [42] and relating the problem to that of the classical erasure channel.



(a) Inner and outer bounds for $\alpha = 0.05$ and $\varepsilon = 1\%$. (b) Exact outer bound for $\alpha = 0.0825$ and $\varepsilon = 5.5\%$.

FIG. 4: Approximate inner and outer bounds on the non-asymptotic achievable rate region for the depolarizing channel (see Results 3 and 5).

Result 2. *For the qubit erasure channel $\mathcal{E}_\beta$ with $\beta \in [0,1]$, the boundary $\hat{R}^{\text{cpp}}(n;\varepsilon)$ satisfies*

$$\varepsilon = \sum_{l=n-k+1}^{n} \binom{n}{l} \beta^l (1-\beta)^{n-l} \left(1 - 2^{n\left(1-\hat{R}^{\text{cpp}}(n;\varepsilon)\right)-l}\right) . \tag{8}$$

Moreover, for large n, we have the expansion

$$\hat{R}^{\text{cpp}}(n;\varepsilon) = 1 - \beta + \sqrt{\frac{\beta(1-\beta)}{n}}\Phi^{-1}(\varepsilon) + O\left(\frac{1}{n}\right) . \tag{9}$$

The latter expression is a third order approximation of the achievable region, but this time the term proportional to $\frac{1}{n}\log n$ vanishes. In Figure 3 we show this approximation for a qubit erasure channel with $\beta = 0.25$ and fidelity 99%. In Figure 3(a) we see that the non-asymptotic achievable region reaches 90% of the channel capacity for $n \approx 180$. Again, this confirms that the non-asymptotic treatment is crucial in the quantum setting. In Figure 3(b) we compare the third order approximation with the exact boundary of the achievable region in (8). We see that the approximation is already very precise (and the term $O(\frac{1}{n})$ thus negligible) for fairly small $n \approx 50$.

C. Depolarizing Channel

Another basic channel of interest is the qubit depolarizing channel. It is given by the map

$$\mathcal{D}_\alpha : \rho \mapsto (1-\alpha)\rho + \frac{\alpha}{3}\left(X\rho X + Y\rho Y + Z\rho Z\right) , \tag{10}$$

where $\alpha \in [0,1]$ is a parameter and X, Y, Z are the Pauli operators. For this channel not even a closed formula for the quantum capacity $Q(\mathcal{D}_\alpha)$ is known, and the non-regularized coherent information

$$I_c(\mathcal{D}_\alpha) = 1 - h(\alpha) - \alpha \log 3 \tag{11}$$

is only a strict lower bound on it [12] (where $h(\alpha)$ again denotes the binary entropy). However, various upper bounds on the quantum capacity have been established as well [29, 34, 35, 37]. For example, in [34] it is shown that $Q(\mathcal{D}_\alpha) \leq Q(\mathcal{Z}_\alpha) = 1 - h(\alpha)$, the quantum capacity of the qubit dephasing channel with dephasing parameter α. Here we extend this result to the non-asymptotic setting and find the following outer (converse) bound for the achievable rate region.

Result 3. *For the qubit depolarizing channel $\mathcal{D}_\alpha$ with $\alpha \in [0,1]$, the boundary $\hat{R}^{\text{cpp}}(n;\varepsilon)$ satisfies*

$$\hat{R}_{\mathcal{D}_\alpha}(n;\varepsilon) \leq \hat{R}_{\mathcal{Z}_\alpha}(n;\varepsilon)\,, \tag{12}$$

where $\hat{R}_{\mathcal{Z}_\alpha}(n;\varepsilon)$ denotes the boundary of the achievable rate region for the qubit dephasing channel with dephasing parameter α as in Result 1.

Clearly this allows us to recycle the bounds in Result 1 and use them as outer bounds for the achievable region. This is done in Figure 4 for two example channels. In Figure 4(a) we plot the second order approximation of the outer bound for a depolarizing channel with $\alpha = 0.05$ and 99% fidelity. We can see that in order to exceed the coherent information, we will need to code for at least $N_0 = 738$ channel uses. This indicates that the question of whether the coherent information is a good or bad lower bound on the asymptotic quantum capacity is not practically relevant as long as we do not have a quantum computer that is able to perform a decoding operation on many hundreds of qubits. We also show a second order approximation for a general inner bound which is given in Result 5 below.

In Figure 4(b) we examine a channel with parameters $\alpha = 0.0825$ and $\varepsilon = 5.5\%$. Instead of using an approximation for the outer bound we use the exact outer bound to give the answer (it is 42) to the question of how many channel uses we need at minimum to exceed the coherent information. However, note that this does not give us any indication of what code (in particular if it is assisted or not) one would need to use for this purpose.

D. General Bounds

We have so far focused our attention on two specific (albeit very important) examples of channels. However, many of the results derived in this paper also hold more generally. For example, we find the following outer (converse) bound for coding schemes that allow classical post-processing.

Result 4. *For any quantum channel $\mathcal{N}$, the boundary $\hat{R}^{\text{cpp}}(n;\varepsilon)$ satisfies*

$$\hat{R}^{\text{cpp}}(n;\varepsilon) \leq -\log f\big(\mathcal{N}^{\otimes n},\varepsilon\big)\,, \tag{13}$$

where $f(\mathcal{N},\varepsilon)$ is the solution to a semidefinite optimization program. Moreover, if $\mathcal{N}$ is covariant we find the asymptotic expansion

$$\hat{R}^{\text{cpp}}(n;\varepsilon) \leq \hat{R}^{\text{cpp}}_{\text{outer}}(n;\varepsilon), \quad \textit{with} \quad \hat{R}^{\text{cpp}}_{\text{outer}}(n;\varepsilon) = I_R(\mathcal{N}) + \sqrt{\frac{V_R^\varepsilon(\mathcal{N})}{n}}\,\Phi^{-1}(\varepsilon) + O\left(\frac{\log n}{n}\right), \tag{14}$$

where the Rains information, $I_R(\mathcal{N})$, and its variance, $V_R^\varepsilon(\mathcal{N})$, are defined in Theorem 3.

The semidefinite optimization program $f(\mathcal{N},\varepsilon)$ (see Section IV A for details) is similar in spirit to the metaconverse for classical channel coding by Polyanskiy *et al.* [27], formulated as a linear program by Matthews [23].[3] Note that the bound (14) is tight up to the second order asymptotically for the qubit dephasing channel (Result 1) and the erasure channel with classical post-processing assistance (Result 2). However, in the generic covariant case the bound is in general not expected to be tight even in first order. Moreover, if the channel is not covariant we cannot find any non-trivial outer bounds on the non-asymptotic achievable region that allows for an asymptotic expansion in the above sense.

Moreover, an inner (achievability) bound of the form shown in Result 1 also holds generally for all quantum channels.

[3] For quantum coding, Matthews and Leung [20] also give semidefinite optimization program lower bounds on the error boundary $\hat{\varepsilon}_{\mathcal{N}}(n;R) := \max\{\varepsilon : (R,n,\varepsilon) \text{ is achievable}\}$ for fixed rate R.

Result 5. *For any quantum channel $\mathcal{N}$, the boundary $R^*(n;\varepsilon)$ satisfies*

$$R^*(n;\varepsilon) \geq R^*_{\text{inner}}(n;\varepsilon), \quad \textit{with} \quad R^*_{\text{inner}}(n;\varepsilon) = I_c(\mathcal{N}) + \sqrt{\frac{V_c^\varepsilon(\mathcal{N})}{n}}\Phi^{-1}(\varepsilon) + O\left(\frac{\log n}{n}\right), \tag{15}$$

where the coherent information, $I_c(\mathcal{N})$, and its variance, $V_c^\varepsilon(\mathcal{N})$, are defined in Theorem 6.

Note that the bound (15) is tight up to the second order asymptotically for the qubit dephasing channel (Result 1). However, the bound does not tightly characterize the achievable region of general channels, although we have reasons to conjecture that it does for degradable channels. In fact, this bound is a direct consequence of an inner bound due to Morgan and Winter [24] together with a second order expansion of smooth entropies in [40].

III. NOTATION, INFORMATION MEASURES, AND CODES

In this paper log denotes the binary logarithm. To express the second order expansion of the non-asymptotic quantities we need the *cumulative standard Gaussian distribution* function

$$\Phi(x) := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{y^2}{2}} \mathrm{d}y\,. \tag{16}$$

We denote *finite-dimensional Hilbert spaces* by capital letters. In particular, we use A and B to model the channel input and output space, whereas M and the isomorphic spaces M' and M'' are used to model the quantum systems containing the maximally entangled state to be transmitted. We also use A^n to denote the n-fold tensor product of A for any $n \in \mathbb{N}$. The dimension of A is denoted by $|A|$, we use $[A]$ to denote the set $\{1, 2, \ldots, |A|\}$, and fix a standard orthonormal basis $\{|x\rangle_A\}_{x\in[A]}$.

We use $\mathcal{L}(A)$ to denote the set of *linear operators* on A, $\mathcal{P}(A)$ to denote the set of *positive semi-definite operators* on A, and $\mathcal{S}(A) := \{\rho \in \mathcal{P}(A) : \mathrm{tr}(\rho) = 1\}$ to denote *quantum states* with unit trace on A. A quantum state is called *pure* if it has rank one. We write $\rho \ll \sigma$ if the support of ρ is contained in the support of σ. For general positive operators $\rho, \sigma \in \mathcal{P}(A)$, we define Uhlmann's *fidelity* [43] as

$$F(\rho,\sigma) := \left(\left\|\sqrt{\rho}\sqrt{\sigma}\right\|_1\right)^2, \tag{17}$$

where $\|X\|_1 := \mathrm{tr}\left(\sqrt{XX^\dagger}\right)$ is the trace norm. If one of the states is pure, this expression simplifies to $F(|\psi\rangle\langle\psi|, \sigma) = \langle\psi|\sigma|\psi\rangle$.

We often use subscripts to clarify which Hilbert spaces an operator acts on. Let A' be isomorphic to A. Throughout the manuscript we denote the *maximally entangled state* on AA' by $\phi_{AA'} = |\phi\rangle\langle\phi|_{AA'}$ with $|\phi\rangle_{AA'} = |A|^{-1/2}\sum_{x\in[A]} |x\rangle_A \otimes |x\rangle_{A'}$. For a general state $\rho_A \in \mathcal{S}(A)$, its *canonical purification* is $|\psi^\rho\rangle_{AA'} = |A|\sqrt{\rho_A}\otimes 1_{A'}|\phi\rangle_{AA'}$. We clearly have $\mathrm{tr}_{A'}(\psi^\rho_{AA'}) = \rho_A$ where $\mathrm{tr}_{A'}$ denotes the partial trace over A'.

Quantum channels are completely positive and trace preserving maps between operators and denoted by calligraphic letters. In particular, we investigate channels $\mathcal{N}_{A\to B}$ that map $\mathcal{P}(A)$ to $\mathcal{P}(B)$. The *Choi state* of a quantum channel $\mathcal{N}_{A\to B}$ is defined using the corresponding non-calligraphic letter as $N_{AB} = (\mathcal{I}_A \otimes \mathcal{N}_{A'\to B})(|A|\phi_{AA'})$.

A. Information Measures

Our asymptotic results are stated in terms of the following quantities. For $\rho \in \mathcal{S}(\mathcal{H})$ and $\sigma \in \mathcal{P}(\mathcal{H})$ with $\rho \ll \sigma$, Umegaki's *relative entropy* [44] and the quantum *relative entropy variance* [21, 40] are given by

$$D(\rho\|\sigma) := \operatorname{tr}\left[\rho\left(\log\rho - \log\sigma\right)\right] \quad \text{and} \quad V(\rho\|\sigma) := \operatorname{tr}\left[\rho\left(\log\rho - \log\sigma - D(\rho\|\sigma)\right)^2\right], \tag{18}$$

respectively. The *conditional entropy* and the *conditional entropy variance* [40] of a state $\rho_{AB} \in \mathcal{S}(AB)$ are given as

$$H(A|B)_\rho := -D(\rho_{AB}\|1_A \otimes \rho_B) \quad \text{and} \quad V(A|B)_\rho := V(\rho_{AB}\|1_A \otimes \rho_B), \tag{19}$$

respectively. Related to this we define the *coherent information* of ρ_{AB} as $I(A\rangle B)_\rho := -H(A|B)_\rho$ and its corresponding variance $V(A\rangle B)_\rho := V(A|B)_\rho$.

For our non-asymptotic results, we require the following quantities. For $\rho \in \mathcal{S}(\mathcal{H})$ and $\sigma \in \mathcal{P}(\mathcal{H})$ the *hypothesis testing relative entropy* [45] is defined as

$$D_H^\varepsilon(\rho\|\sigma) := -\log\beta_{1-\varepsilon}(\rho\|\sigma) \quad \text{with} \quad \beta_{1-\varepsilon}(\rho\|\sigma) := \min_{\substack{0\leq\Lambda\leq 1 \\ \operatorname{tr}[\Lambda\rho]\geq 1-\varepsilon}} \operatorname{tr}[\Lambda\sigma]. \tag{20}$$

The *hypothesis testing Rains relative entropy* of a quantum channel $\mathcal{N}_{A\to B}$ is defined as (following the generalized divergence framework discussed in [42]),

$$I_R^\varepsilon(\mathcal{N}_{A\to B}) := \sup_{\rho_A\in\mathcal{S}(A)} I_R^\varepsilon(A:B)_{\mathcal{N}_{A'\to B}(\psi^\rho_{AA'})} \quad \text{with} \quad I_R^\varepsilon(A:B)_\rho := \inf_{\sigma_{AB}\in\mathrm{PPT}'(A:B)} D_H^\varepsilon(\rho_{AB}\|\sigma_{AB}), \tag{21}$$

where $\mathrm{PPT}'(A\!:\!B)$ is the *Rains set* [1, 29], a superset of the set of positive partial transpose (PPT) states. It is defined as

$$\mathrm{PPT}'(A\!:\!B) := \left\{\tau_{AB} \in \mathcal{P}(AB) \,\middle|\, \left\|T_B(\tau_{AB})\right\|_1 \leq 1\right\}, \tag{22}$$

where T_B denotes the partial transpose map on B. In particular, we have the following inequality [28, Lm. 2]. For every $\sigma_{AB} \in \mathrm{PPT}'(A:B)$, we have

$$\langle\phi|\sigma_{AB}|\phi\rangle_{AB} \leq \frac{1}{|M|} \tag{23}$$

for all maximally entangled states $|\phi\rangle_{AB}$ of local dimension $|M|$. Finally, a quantum channel $\mathcal{N}_{A\to B}$ is called PPT preserving if a PPT state input necessarily results in a PPT state output. It turns out that PPT-preserving channels output PPT states for any input, since they have PPT Choi states [29] (see the discussion after Eq. 4.13). Channels with PPT Choi states were also called PPT-binding in [19].

B. Codes for Entanglement Transmission Assisted by Classical Post-Processing

We have defined unassisted entanglement-transmission codes in Section II and in Figure 1. Let us reintroduce them in the context of codes assisted by classical post-processing.

For this, we again consider any quantum channel $\mathcal{N} \equiv \mathcal{N}_{A\to B}$ and its n-fold extension $\mathcal{N}^{\otimes n}$ that maps states on A^n to states on B^n. An *entanglement transmission code assisted by classical*

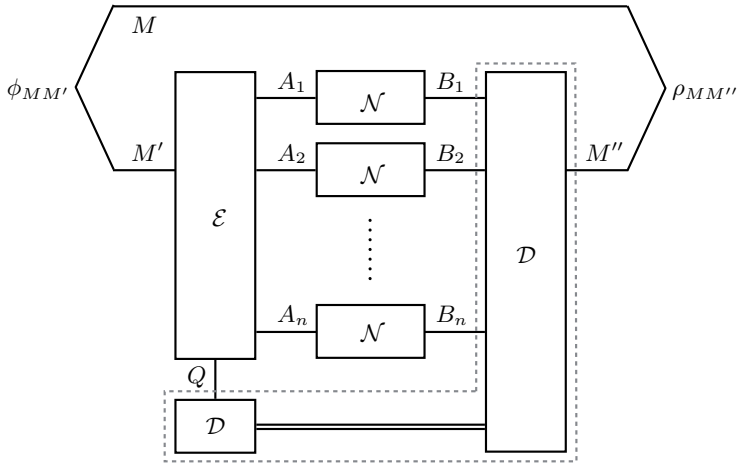


FIG. 5: Coding Scheme for entanglement transmission over n uses of a channel $\mathcal{N}_{A\to B}$ with classical post-processing. The encoder $\mathcal{E} \equiv \mathcal{E}_{M'\to A^nQ}$ encodes M' into the channel input systems and a local memory Q. Later, the decoder $\mathcal{D} \equiv \mathcal{D}_{QB^n\to M''}$ recovers the maximally entangled state from the channel output systems and the memory Q using classical communication and local operations. The performance of the code is measured using the fidelity $F(\phi_{MM'}, \rho_{MM''})$.

post-processing for $\mathcal{N}^{\otimes n}$ is given by a triplet $\{|M|, \mathcal{E}, \mathcal{D}\}$, as depicted in Figure 5. Here, $|M|$ is the local dimension of a maximally entangled state $|\phi\rangle_{MM'}$ that is to be transmitted over $\mathcal{N}^{\otimes n}$. The encoder $\mathcal{E}_{M'\to A^nQ}$ is a completely positive trace-preserving map that prepares the channel inputs $A_1, A_2, \ldots A_n$ and a local memory system, which we denote by Q. The decoder $\mathcal{D}_{QB^n\to M''}$ is a completely positive trace-preserving map that is restricted to local operations and classical communication with regards to the bipartition $Q : B^n$ and outputs M'' on the receiver's side.

Finally, we note that unassisted codes are recovered if we choose Q to be trivial. Hence, unassisted codes are contained in the set of assisted codes.

IV. PROOFS: GENERAL BOUNDS

It is convenient to first discuss the general bounds discussed in Results 4 and 5.

A. General Outer Bounds on the Achievable Region

In this section we derive the general outer bounds from Result 4, precisely stated as Corollary 2 and Theorem 3 below. Our results are inspired by the strong converse results for generalized dephasing channels from [42] and the metaconverse for classical channel coding [27].

We first formulate a general metaconverse bounding possible rates R given a tolerated error ε for single uses of a fixed channel $\mathcal{N}$. This bound has the useful property that channel symmetries can be used to simplify its form. Nevertheless, when applied to n instances of $\mathcal{N}$, the bound is not efficiently computable. Loosening the bound produces a more computationally tractable convex optimization, specifically a semidefinite program.

Lemma 1. *Let $\mathcal{N}_{A\to B}$ be a quantum channel. Then, for any fixed $\varepsilon \in (0, 1)$, the achievable region*

with cpp-assistance satisfies,[4]

$$\hat{R}^{\mathrm{cpp}}(1;\varepsilon) \leq I_R^{\varepsilon}\big(\mathcal{N}\big)\,. \tag{24}$$

Proof. First, observe that the encoding operation $\mathcal{E}_{M'\to AQ}$ can be chosen to be an isometry without loss of generality, because we may include any extension systems needed for the Stinespring dilation into Q. Then we may express the entanglement fidelity as follows

$$F = \mathrm{tr}[\phi_{MM'}\mathcal{D}_{BQ\to M'}\circ\mathcal{N}_{A\to B}\circ\mathcal{E}_{M'\to AQ}(\phi_{MM'})] \tag{25}$$

$$= \mathrm{tr}[\mathcal{E}_{M\to\bar{A}\bar{Q}}\otimes\mathcal{D}^{\dagger}_{BQ\to M'}(\phi_{MM'})\mathcal{N}_{A\to B}(\mathcal{E}_{M'\to AQ}\otimes\mathcal{E}_{M\to\bar{A}\bar{Q}}(\phi_{MM'}))]\,. \tag{26}$$

Since $\mathcal{E}$ is an isometry, the state $\rho_{A\bar{A}Q\bar{Q}} = \mathcal{E}_{M'\to AQ}\otimes\mathcal{E}_{M\to\bar{A}\bar{Q}}(\phi_{MM'})$ is pure, and therefore there exists an isometry $W_{A'\to\bar{A}Q\bar{Q}}$ such that $|\rho\rangle_{A\bar{A}Q\bar{Q}} = W_{A'\to\bar{A}Q\bar{Q}}\,|\psi^{\rho}\rangle_{AA'}$. Thus,

$$F = \mathrm{tr}[W^{\dagger}_{A'\to\bar{A}Q\bar{Q}}(\mathcal{E}_{M\to\bar{A}\bar{Q}}\otimes\mathcal{D}^{\dagger}_{BQ\to M'}(\phi_{MM'}))W_{A'\to\bar{A}Q\bar{Q}}\mathcal{N}_{A\to B}(\psi^{\rho}_{AA'})]\,. \tag{27}$$

Now consider the entanglement fidelity of any PPT$'$ state $\sigma_{A'B}$ instead of $\mathcal{N}_{A\to B}(\psi^{\rho}_{AA'})$. By (23) we have

$$\mathrm{tr}[\phi_{MM'}(\mathcal{E}^{\dagger}_{M\to\bar{A}\bar{Q}}\otimes\mathcal{D}_{BQ\to M'}(W_{A'\to\bar{A}Q\bar{Q}}\sigma_{A'B}W^{\dagger}_{A'\to\bar{A}Q\bar{Q}}))] \leq \frac{1}{M}\,, \tag{28}$$

as the operations on $\sigma_{A'B}$ are all PPT-preserving. We may write this bound in terms of the hypothesis-testing relative entropy, because

$$\Lambda_{A'B} := W^{\dagger}_{A'\to\bar{A}Q\bar{Q}}(\mathcal{E}_{M\to\bar{A}\bar{Q}}\otimes\mathcal{D}^{\dagger}_{BQ\to M'}(\phi_{MM'}))W_{A'\to\bar{A}Q\bar{Q}} \tag{29}$$

is a feasible test to discriminate between $\mathcal{N}_{A\to B}(\psi^{\rho}_{A'A})$ and $\sigma_{A'B}$. That is, $\Lambda_{A'B}$ satisfies $0 \leq \Lambda_{A'B} \leq 1_{A'B}$ and $\mathrm{tr}[\Lambda_{A'B}\mathcal{N}_{A\to B}(\psi^{\rho}_{A'A})] \geq 1-\varepsilon$, the former since $\mathcal{D}$ is completely-positive and trace-preserving and $\mathcal{E}$ and W are isometries, the latter by assumption that $F \geq 1-\varepsilon$. From (28) we then obtain

$$\hat{R}^{\mathrm{cpp}}(1;\varepsilon) \leq D_H^{\varepsilon}(\mathcal{N}_{A\to B}(\psi^{\rho}_{A'A}))\,\|\,\sigma_{A'B})\,. \tag{30}$$

Since the bound holds for all PPT$'$ $\sigma_{A'B}$, we may take the infimum over this set to obtain

$$\hat{R}^{\mathrm{cpp}}(1;\varepsilon) \leq I_R^{\varepsilon}(A':B)_{\mathcal{N}_{A\to B}(\psi^{\rho}_{A'A})}\,. \tag{31}$$

This bound depends on the precise channel input $\rho_A \in \mathcal{S}(A)$ used by the code, but we can remove the dependence by taking the supremum over all possible inputs. The result is (24). □

Applied to the channel $\mathcal{N}^{\otimes n}$ we immediately get for any fixed $\varepsilon \in (0,1)$,

$$\hat{R}^{\mathrm{cpp}}(n;\varepsilon) \leq I_R^{\varepsilon}\big(\mathcal{N}^{\otimes n}\big)\,. \tag{32}$$

This bound forms a counterpart to Lemma 4, but suffers from the same weakness. It is generally hard to evaluate this bound even for moderately large n. However, we may relax the bound from Lemma 1 to a convex optimization by restricting the form of the possible states σ_{AB} in the definition of the hypothesis testing Rains relative entropy I_R^{ε}.

[4] Indeed, the outer bound also holds for coding schemes with (unphysical) positive partial transpose (PPT) assistance and this includes in particular classical pre- and post-processing assistance (see, e.g., [20] for a precise definition of PPT assisted codes).

Corollary 2. *Let $\mathcal{N}_{A\to B}$ be a quantum channel. We define the function*

$$f(\mathcal{N},\varepsilon) := \inf_{\rho_A\in\mathcal{S}(A)} \inf_{\Lambda_{AB}\in\Gamma(\rho_A,\mathcal{N},\varepsilon)} \sup_{\mathcal{M}_{A\to B}\in PPT} \operatorname{tr}[\Lambda_{AB}M_{AB}]\,, \tag{33}$$

with the set $\Gamma(\rho_A,\mathcal{N},\varepsilon) := \{\Lambda_{AB} : 0\leq\Lambda_{AB}\leq\rho_A^T\otimes 1_B, \operatorname{tr}[\Lambda_{AB}N_{AB}]\geq 1-\varepsilon\}$, and the Choi states M_{AB} of $\mathcal{M}_{A\to B}$ and N_{AB} of $\mathcal{N}_{A\to B}$. Then, for any fixed $\varepsilon\in(0,1)$, the achievable region satisfies

$$\hat{R}^{\mathrm{cpp}}(1;\varepsilon) \leq -\log f(\mathcal{N},\varepsilon)\,. \tag{34}$$

Proof. Suppose that $\sigma_{AB} = (\mathcal{I}_A\otimes\mathcal{M}_{A'\to B})(\psi^\rho_{AA'})$ for some PPT-preserving (PPT-binding) channel $\mathcal{M}_{A\to B}$. Writing out the right-hand side of (24) using this relaxation and notation gives

$$\hat{R}^{\mathrm{cpp}}(1;\varepsilon) \leq -\log \inf_{\rho_A\in\mathcal{S}(A)} \sup_{\mathcal{M}_{A\to B}\in\mathrm{PPT}} \inf_{\substack{0\leq\Lambda'\leq 1\\ \operatorname{tr}[\Lambda'\rho]\geq 1-\varepsilon}} \operatorname{tr}\left[\Lambda'_{AB}(\mathcal{I}_A\otimes\mathcal{M}_{A'\to B})(\psi^\rho_{AA'})\right]. \tag{35}$$

Now we may define $\Lambda_{AB} = (\rho_A^T)^{1/2}\Lambda'_{AB}(\rho_A^T)^{1/2}$ and find

$$\hat{R}^{\mathrm{cpp}}(1;\varepsilon) \leq -\log \inf_{\rho_A\in\mathcal{S}(A)} \sup_{\mathcal{M}_{A\to B}\in\mathrm{PPT}} \inf_{\Lambda_{AB}\in\Gamma(\rho_A,\mathcal{N},\varepsilon)} \operatorname{tr}[\Lambda_{AB}M_{AB}]\,. \tag{36}$$

Finally for fixed channel input ρ_A, we can reverse the order of the inner optimizations in (36) by von Neumann's minimax theorem, since the objective function is linear and the sets are both convex and compact. This concludes the proof. □

Furthermore, we show in Appendix A that $f(\mathcal{N},\varepsilon)$ can be expressed as a semidefinite optimization program that satisfies strong duality.

Group Covariant Channels: In the following we show that symmetries of the channel can further simplify the outer bounds. First let us state precisely what we mean by symmetries. Suppose G is a group represented by unitary operators U_g on A and V_g on B. A quantum channel $\mathcal{N}_{A\to B}$ is covariant with respect to G when

$$V_g\mathcal{N}(\cdot)V_g^\dagger = \mathcal{N}(U_g\cdot U_g^\dagger), \quad \forall g\in G\,. \tag{37}$$

Alternatively we can also write this as an invariance of the channel

$$\mathcal{N}(\cdot) = V_g^\dagger\mathcal{N}(U_g\cdot U_g^\dagger)V_g, \quad \forall g\in G\,. \tag{38}$$

Now the main workhorse to simplify our outer bounds for channels with symmetries is [42, Prop. 2], which states that we may restrict the optimization in Lemma 1 to covariant input states. Due to the form of the hypothesis testing Rains relative entropy, we may then also choose group covariant PPT′ states σ and test operators Λ to obtain the tightest bound. Note that the convex optimization outer bound in Corollary 2 inherits these symmetry simplifications.

For general tensor product channels, which are invariant to permutation of the inputs and outputs, this allows us to restrict attention to pure states that are permutation invariant. Moreover, if the channel is covariant, that is, covariant with respect to the full unitary group, then the channel input state can be chosen to be maximally mixed.

Now let $\mathcal{N}_{A\to B}$ be a covariant quantum channel and $\phi_{AA'}$ a maximally entangled state. Then, we bound

$$\hat{R}^{\mathrm{cpp}}(n;\varepsilon) \leq \min_{\sigma_{AB}\in\mathrm{PPT}'(A:B)} \frac{1}{n} D_H^\varepsilon\left(\mathcal{N}_{A'\to B}(\phi_{AA'})^{\otimes n}\middle\|\sigma_{AB}^{\otimes n}\right), \tag{39}$$

where we voluntarily restricted the minimization to product states $\sigma_{AB}^{\otimes n}$ in $\mathrm{PPT}'(A\!:\!B)$. Moreover, since these states have tensor product structure, the outer bound can be expanded using [21, 40]

$$\frac{1}{n}D_H^\varepsilon(\rho^{\otimes n}\|\sigma^{\otimes n}) = D(\rho\|\sigma) + \sqrt{\frac{V(\rho\|\sigma)}{n}}\,\Phi^{-1}(\varepsilon) + O\left(\frac{\log n}{n}\right). \tag{40}$$

This leads to the following theorem.

Theorem 3. *Let $\mathcal{N} \equiv \mathcal{N}_{A\to B}$ be a quantum channel. We define its Rains information as*

$$I_R(\mathcal{N}) := \min_{\sigma_{AB}\in \mathrm{PPT}'(A:B)} D(\mathcal{N}_{A'\to B}(\phi_{AA'})\|\sigma_{AB})\,. \tag{41}$$

and let $\Pi \subset \mathrm{PPT}'(A:B)$ be the set of states that achieve the minimum. The variance of the channel Rains information is

$$V_R^\varepsilon(\mathcal{N}) := \begin{cases} \max\limits_{\sigma_{AB}\in\Pi} V(\mathcal{N}_{A'\to B}(\phi_{AA'})\|\sigma_{AB}) & \text{for} \quad \varepsilon < \frac{1}{2} \\ \min\limits_{\sigma_{AB}\in\Pi} V(\mathcal{N}_{A'\to B}(\phi_{AA'})\|\sigma_{AB}) & \text{for} \quad \varepsilon \geq \frac{1}{2} \end{cases}. \tag{42}$$

If $\mathcal{N}$ is covariant, then for any fixed $\varepsilon \in (0,1)$, the achievable region with cpp-assistance satisfies

$$\hat{R}^{\mathrm{cpp}}(n;\varepsilon) \leq \hat{R}^{\mathrm{cpp}}_{\mathrm{outer}}(n;\varepsilon), \quad \textit{with} \quad \hat{R}^{\mathrm{cpp}}_{\mathrm{outer}}(n;\varepsilon) = I_R(\mathcal{N}) + \sqrt{\frac{V_R^\varepsilon(\mathcal{N})}{n}}\,\Phi^{-1}(\varepsilon) + O\left(\frac{\log n}{n}\right). \tag{43}$$

Since we are here interested in outer bounds, we are also free to chose a potentially sub-optimal $\sigma_{AB} \in \mathrm{PPT}'(A:B)$ to further relax this bound. As we will see in Section V for the qubit dephasing channel and the erasure channel with classical post-processing assistance the bound from Theorem 3 is tight up to the second order asymptotically.

B. General Inner Bounds on the Achievable Region

In this section we derive the general inner bound from Result 5, stated as Theorem 6 below. We use the decoupling approach [13, 14, 18], and in particular a one-shot bound by Morgan and Winter [24] which is a tighter version of previous bounds [4, 6].

To reproduce their result we need the following additional notation. Sub-normalized quantum states are collected in the set $\mathcal{S}_\bullet(A) := \{\rho \in \mathcal{P}(A) : \mathrm{tr}(\rho) \leq 1\}$. The purified distance [39] ε-ball around $\rho \in \mathcal{S}(A)$ is then defined as $\mathcal{B}^\varepsilon(\rho) := \left\{\bar{\rho} \in \mathcal{S}_\bullet(\mathcal{H}) \,\middle|\, F(\bar{\rho},\rho) \geq (1-\varepsilon)^2\right\}$. Finally, for $\rho_{AB} \in \mathcal{S}(AB)$ and $\varepsilon \geq 0$ the *smooth conditional min-entropy* [30, 39] is defined as

$$H_{\min}^\varepsilon(A|B)_\rho := \sup_{\bar{\rho}_{AB}\in\mathcal{B}^\varepsilon(\rho_{AB})} \sup_{\sigma_B\in\mathcal{S}(B)} \sup\left\{\lambda \in \mathbb{R} \,\middle|\, \bar{\rho}_{AB} \leq 2^{-\lambda}\cdot 1_A \otimes \sigma_B\right\}. \tag{44}$$

Let us now restate Morgan and Winter's result expressed in terms of the non-asymptotic achievable region as introduced in Section II.

Lemma 4. *[24, Prop. 20] Let $\mathcal{N}_{A\to B}$ be a quantum channel with complementary channel $\mathcal{N}^c_{A\to E}$. Then $(R,1,\varepsilon)$ is achievable if, for any $\eta \in (0,\varepsilon]$ and any state $\rho_A \in \mathcal{S}_\circ(A)$, we have*

$$R \leq H_{\min}^{\sqrt{\varepsilon}-\eta}(A|E)_\omega - 4\log\frac{1}{\eta}\,, \tag{45}$$

where $\omega_{AE} = \left(\mathcal{I}_A \otimes \mathcal{N}^c_{A'\to E}\right)(\psi^\rho_{AA'})$.

Note that Morgan and Winter use the purified distance as their figure of merit whereas we use the fidelity criterion (1). This accounts for the square root in the smoothing parameter of the conditional min-entropy. Also Morgan and Winter state their result for the special case $n = 1$, but this can be generalized to arbitrary $n \in \mathbb{N}$ if we simply consider $\mathcal{N}_{A\to B}^{\otimes n}$ as a single channel. This leads immediately to the following inner bound on the achievable region:

Corollary 5. *Using the notation of Lemma 4 with* $\omega_{A^nE^n} = \big(\mathcal{I}_{A^n} \otimes (\mathcal{N}^c_{A'\to E})^{\otimes n}\big)(\psi^\rho_{A^nA'^n})$, *we have*

$$\hat{R}(n;\varepsilon) \geq \sup_{\eta\in(0,\varepsilon)} \sup_{\rho_{A^n}\in\mathcal{S}(A^n)} \frac{1}{n}\big(H_{\min}^{\sqrt{\varepsilon}-\eta}(A^n|E^n)_\omega - 4\log\frac{1}{\eta} - 1\big)\,. \tag{46}$$

The problem with this bound is that it is generally hard to evaluate, even for moderately large values of n. Hence we are interested to further simplify the expression on the right-hand side in this regime. To do so, we choose $\eta = 1/\sqrt{n}$ and use input states of the form $\rho_A^{\otimes n}$. This yields the following relaxation, which holds if $n > \frac{1}{\varepsilon}$:

$$\hat{R}(n;\varepsilon) \geq \sup_{\rho_A\in\mathcal{S}(A)} \frac{1}{n}\big(H_{\min}^{\varepsilon_n}(A^n|E^n)_{\omega^{\otimes n}} - 2\log n - 1\big)\,. \tag{47}$$

Here we introduced $\varepsilon_n = \sqrt{\varepsilon} - \frac{1}{\sqrt{n}}$ and ω_{AE} as in Lemma 4. Using standard second order expansion methods [40], we can give an asymptotic expansion of $R^*_{\text{inner}}(n;\varepsilon)$ in (47) as follows.

Theorem 6. *Let* $\mathcal{N} \equiv \mathcal{N}_{A\to B}$ *be a quantum channel. We define its coherent information as*

$$I_c(\mathcal{N}) := \max_{\rho_A\in\mathcal{S}(A)} I(A\rangle B)_\omega\,, \quad \text{with} \quad \omega_{AB} = \big(\mathcal{I}_A \otimes \mathcal{N}_{A'\to B}\big)(\psi^\rho_{AA'}) \tag{48}$$

and let $\Pi \subset \mathcal{S}_\circ(A)$ *be the set of states that achieve the maximum. Define*

$$V_c^\varepsilon(\mathcal{N}) := \begin{cases} \min\limits_{\rho_A\in\Pi} V(A\rangle B)_\omega & \text{for} \quad \varepsilon < \frac{1}{2} \\ \max\limits_{\rho_A\in\Pi} V(A\rangle B)_\omega & \text{for} \quad \varepsilon \geq \frac{1}{2} \end{cases}. \tag{49}$$

Then, for any fixed $\varepsilon \in (0,1)$, *the achievable region satisfies*

$$\hat{R}(n;\varepsilon) \geq \hat{R}_{\text{inner}}(n;\varepsilon), \quad \text{with} \quad \hat{R}_{\text{inner}}(n;\varepsilon) = I_c(\mathcal{N}) + \sqrt{\frac{V_c^\varepsilon(\mathcal{N})}{n}}\,\Phi^{-1}(\varepsilon) + O\left(\frac{\log n}{n}\right). \tag{50}$$

Proof. We analyze the expression in (47) using the following asymptotic expansion of the smooth conditional min-entropy [40],

$$\frac{1}{n}H_{\min}^\varepsilon(A|B)_{\rho^{\otimes n}} = H(A|B)_\rho + \sqrt{\frac{V(A|B)_\rho}{n}}\,\Phi^{-1}\big(\varepsilon^2\big) + O\left(\frac{\log n}{n}\right). \tag{51}$$

This yields that for any $\rho_A \in \mathcal{S}(A)$, we have

$$\hat{R}(n;\varepsilon) \geq H(A|E)_\omega + \sqrt{\frac{V(A|E)_\omega}{n}}\,\Phi^{-1}(\varepsilon) + O\left(\frac{\log n}{n}\right), \tag{52}$$

and then by duality of the conditional entropy we find $H(A|E)_\omega = I(A\rangle B)_\omega$. Furthermore, it is easy to verify that $V(A\rangle E)_\omega = V(A\rangle B)_\omega$ (see, e.g., [17]). We conclude the proof by choosing an optimal state $\rho_A \in \Pi$ depending on the sign of $\Phi^{-1}(\varepsilon)$. □

As we will see in Section V for the qubit dephasing channel the bound in Theorem 6 agrees with the outer bound stated in Result 4 up to the second order asymptotically.

V. PROOFS: EXAMPLES

In this section we derive our results concerning the qubit dephasing channel, the erasure channel, and the qubit depolarizing channel as announced in Results 1, 2, and 3.

A. Covariant Generalized Dephasing Channels

First we consider covariant generalized dephasing channels, which have the (additional) property that $\mathcal{Z}_{A'\to B}(\psi_{AA'})$ has full support on the projector $\Pi = \sum_x |x\rangle\langle x|_A \otimes |x\rangle\langle x|_B$ in some basis. In that case, starting from (39), we can use the data-processing inequality for a map $\mathcal{E}(\cdot) = \Pi\cdot\Pi + (1-\Pi)\cdot(1-\Pi)$ to write

$$\hat{R}^{\mathrm{cpp}}(n;\varepsilon) \leq \min_{\sigma_{AB}\in \mathrm{PPT}'(A:B)} \frac{1}{n} D_H^{\varepsilon}\big((\mathcal{Z}_{A'\to B}(\phi_{AA'}))^{\otimes n}\big\|\sigma_{AB}^{\otimes n}\big) \tag{53}$$

$$\leq \min_{\sigma_B\in\mathcal{S}(B)} \frac{1}{n} D_H^{\varepsilon}\big((\mathcal{Z}_{A'\to B}(\phi_{AA'}))^{\otimes n}\big\|1_{A^n}\otimes\sigma_B^{\otimes n}\big) \tag{54}$$

$$\leq \frac{1}{n} D_H^{\varepsilon}\big((\mathcal{Z}_{A'\to B}(\phi_{AA'}))^{\otimes n}\big\|(1_A\otimes\mathcal{Z}_{A'\to B}(\phi_{A'}))^{\otimes n}\big)\,. \tag{55}$$

To show the second inequality we apply $\mathcal{E}^{\otimes n}$ to both states and employ the data-processing inequality for the hypothesis testing relative entropy. Note that the map $\mathcal{E}$ keeps $\mathcal{Z}(\phi_{AA'})$ invariant and maps $1_A\otimes\sigma_B$ to a normalized state σ_{AB} that is classically correlated and thus in particular in $\mathrm{PPT}'(A:B)$. The bound in (55) has the form of a conditional entropy or coherent information, and it can be expanded again using (40) to find

$$\hat{R}^{\mathrm{cpp}}(n;\varepsilon) \leq I(A\rangle B)_{\mathcal{Z}(\phi)} + \sqrt{\frac{V(A\rangle B)_{\mathcal{Z}(\phi)}}{n}}\,\Phi^{-1}(\varepsilon) + O\left(\frac{\log n}{n}\right). \tag{56}$$

So for covariant generalized dephasing channels this now agrees with the inner bound from Theorem 6 up to the second order asymptotically (an example being the qubit dephasing channel as discussed in Section V B). Given the recent results [42] the outer bound (56) might also hold for generalized dephasing channels (without assuming covariance) and then agree with the inner bound from Theorem 6 up to the second order asymptotically.

B. The Qubit Dephasing Channel

The qubit dephasing channel is defined as

$$\mathcal{Z}_\gamma : \rho \mapsto (1-\gamma)\rho + \gamma Z\rho Z\,, \tag{57}$$

where $\gamma\in[0,1]$ is a parameter and Z is the Pauli Z operator. This channel is covariant since it is a qubit Pauli channel. Now to determine the second order asymptotic performance it is sufficient to specialize the outer bound from (56) and to apply Theorem 6 for the inner bound. It is then easily seen that

$$I(R\rangle B)_{\mathcal{Z}_\gamma(\psi)} = 1 - h(\gamma) \quad \text{with} \quad h(\gamma) = 1 - \gamma\log\gamma - (1-\gamma)\log(1-\gamma) \tag{58}$$

$$V(R\rangle B)_{\mathcal{Z}_\gamma(\psi)} = v(\gamma) \quad \text{with} \quad v(\gamma) = \gamma(\log\gamma + h(\gamma))^2 + (1-\gamma)(\log(1-\gamma) + h(\gamma))^2\,, \tag{59}$$

and hence we deduce

$$\hat{R}(n;\varepsilon) = 1 - h(\gamma) + \sqrt{\frac{v(\gamma)}{n}}\,\Phi^{-1}(\varepsilon) + O\left(\frac{\log n}{n}\right). \tag{60}$$

However, we can refine (60) and determine the third order asymptotic performance. We do this by directly obtaining the finite block length behavior of the qubit dephasing channel from that of the classical binary symmetric channel (BSC). First, consider the converse, particularly that of (39), applied to the channel $\mathcal{Z}_\gamma^{\otimes n}$. Using the Bell states $\phi^+_{AB} = \phi_{AA'}$ and $\phi^-_{AB} = (1_A \otimes Z_{A'})\phi_{AA'}(1_A \otimes Z_{A'})$, we immediately find

$$\omega_{AB} := \mathcal{N}_{A'\to B}(\phi_{AA'}) = (1-\gamma)\phi^+_{AB} + \gamma\phi^-_{AB}\,. \tag{61}$$

Now, in (39) we are free to pick any PPT$'$ state to obtain a bound. Pick $\sigma_{AB} = \frac{1}{2}(\phi^+_{AB} + \phi^-_{AB})$, which gives[5]

$$\hat{R}(n;\varepsilon) \leq \hat{R}^{\text{cpp}}(n;\varepsilon) \leq \frac{1}{n} D_H^\varepsilon\big(\omega_{AB}^{\otimes n} \big\| \sigma_{AB}^{\otimes n}\big)\,. \tag{62}$$

To connect to the finite block length bounds of the BSC, consider measuring both A and B in the Pauli x basis, and let X and Y be the output random variables for A and B, respectively. For the state ω_{AB}, this results in the distribution P_{XY} in which P_X is uniformly-distributed and $P[Y = X] = 1 - \gamma$. For σ_{AB}, the distribution is of product form $P_X Q_Y$ with Q_Y also uniform. Moreover, the original quantum states can be reconstructed from the classical random variables X and Y by the map which outputs ϕ^+_{AB} when $X = Y$ and ϕ^-_{AB} otherwise. Therefore, the bound becomes

$$\hat{R}(n;\varepsilon) \leq \frac{1}{n} D_H^\varepsilon\big(P_{XY}^{\times n} \| P_X^{\times n} \times Q_Y^{\times n}\big)\,, \tag{63}$$

which is precisely the bound obtained by Polyanskiy *et al.* for the BSC [27, Theorem 26], which is equivalent to the classical sphere-packing bound [15, Eq. 5.8.19]. This establishes one inequality (upper bound) in (6).

For the achievability, we may directly employ linear codes for the classical BSC to the qubit dephasing channel. Specifically, any linear $\{R, n, \varepsilon\}$ code for the BSC (which recovers the input with probability at least $1 - \varepsilon$, averaged over a uniform choice of inputs), can be converted into an $\{R, n, \varepsilon\}$ Calderbank-Shor-Steane (CSS) code for entanglement transmission over the dephasing channel. This is possible since, for a linear code, the action of the channel is a mapping among the orthogonal Bell states, which is essentially a classical action.

To formalize the connection, begin with the description of the classical linear code by its $(\log n - \log|M|) \times n$ parity check matrix H. Each row $r_j \in \{0,1\}^n$ defines a parity function and the codewords c_k of the code must satisfy $c_k \cdot r_j = 0$ for all j. The associated CSS code can be defined as the simultaneous $+1$ eigenspace of the "stabilizer" operators X^{r_j}, where $X^{r_j} = X^{r_{j,1}} \otimes \cdots \otimes X^{r_{j,n}}$.[6] Crucially, the action of the channel is to apply an operator of the form Z^u, with $u \in \{0,1\}^n$, according to the distribution P_U. At the output, the receiver can simultaneously determine the eigenvalues of all the of the stabilizer operators. This information is precisely equivalent to determining the value of the parity checks of the classical linear code, called the

[5] The choice of σ_{AB} is equivalent to using the convex relaxation of the bound, Corollary 2, and choosing $\mathcal{M} = \mathcal{Z}_{1/2}$ in (33).

[6] Generically, a CSS code has stabilizers of both X-type, as here, and of Z-type, i.e. composed of products of Pauli Z operators.

syndrome s. Given the syndrome, the decoder of the classical code determines a guess as to the input codeword, which is equivalent to a guess $u'(s)$ of the actual channel error.

We may also utilize this algorithm (whatever its precise details) in the quantum case, and attempt to correct the error by applying $Z^{u'(s)}$. When $u'(s)$ is the true error pattern, the quantum state is properly recovered, and the entanglement fidelity is unity. On the other hand, if $u'(s)$ is incorrect, then in the worst case the action $Z^{u'(s)+u}$ is a logical operation on the code subspace, which results in a state orthogonal to the desired entangled state. Therefore, the error probability of the classical code translates directly into the entanglement fidelity of the quantum code. Thus, we may apply finite-block length bounds for linear codes, particularly the bound of Poltyrev [25] (see also [27, Eq. 65]). This establishes the other inequality (lower bound) in (6).

C. The Erasure Channel

The qubit erasure channel is defined as

$$\mathcal{E}_\beta : \rho \mapsto (1-\beta)\rho + \beta \left|e\right\rangle\!\left\langle e\right| , \tag{64}$$

where $\beta \in [0,1]$ is a parameter and $\left|e\right\rangle\!\left\langle e\right|$ is a quantum state orthogonal to ρ. Using the covariance of the channel, we could first obtain a second order asymptotic similar to that of the dephasing channel in (60). However, it is not too difficult to directly derive an outer bound and an explicit coding scheme leading to an inner bound, which precisely match for all n.

Let us begin with the outer bound. Again we may relate the finite block length performance to a classical coding problem, namely the classical binary erasure channel (BEC). The argument for the outer bound proceeds very similarly to the dephasing example. The optimal channel input state corresponds to the maximally entangled state $\phi_{AA'}$, and the state produced by the channel is now

$$\omega_{AB} = (1-\beta)\phi_{AB} + \beta\pi_A \otimes \left|e\right\rangle\!\left\langle e\right|_B , \tag{65}$$

where π_A denotes the maximally-mixed state. Measurement of A in the Pauli x basis and B in the basis $\{\left|+\right\rangle, \left|-\right\rangle, \left|e\right\rangle\}$ produces the distribution P_{XY} with P_X uniform and $Y = X$ with probability $1-\beta$ and $Y = e$ with probability $1-\beta$. The original state can be reconstructed using the map which sends (X,Y) to ϕ^+_{AB} when $X = Y$, ϕ^-_{AB} when $X \neq Y \neq e$, and to $\pi_A \otimes \left|e\right\rangle\!\left\langle e\right|_B$ when $Y = e$ otherwise. As before, we make a specific choice of PPT$'$ state in (39), but this time not a product state. Instead, consider the classical distribution $P_X^{\times n} \times Q_{Y^n}$ given in [27, Eq. 168]. The Q_{Y^n} distribution has the property that any two y^n with the same number of erasure symbols e have the same probability, i.e. there is no dependence on the number of 0s versus 1s. The aforementioned map takes the distribution to a quantum state which is diagonal in the standard bases $\{\left|0\right\rangle, \left|1\right\rangle\}$ for A and $\{\left|0\right\rangle, \left|1\right\rangle, \left|e\right\rangle\}$ for B, and is therefore a PPT state. This can be seen as follows. Consider a fixed position j in a given a pair (x^n, y^n). If $y_j = e$, the state of the jth pair of systems AB is manifestly diagonal in the standard basis. On the other hand, if $y_j \neq e$, then the state is mapped to either ϕ^+_{AB} or ϕ^-_{AB} depending on the value of x_j. But the sequence in which y_j takes the other value has identical probability, meaning the two Bell states occur with equal probability, making the AB state diagonal. Since we may map $\omega_{AB}^{\otimes n}$ and $\sigma_{A^nB^n}$ to the associated classical distributions and back, the following converse holds for the qubit erasure channel,

$$\hat{R}^{\text{cpp}}(n;\varepsilon) \leq \frac{1}{n} D_H^\varepsilon\big(P_{XY}^{\times n} \big\| P_X^{\times n} \times Q_{Y^n}\big) . \tag{66}$$

By design in the choice of σ_{AB},[7] this is precisely the bound for the BEC reported by Polyanskiy *et al.* [27, Thm. 38], as discussed in more detail by Polyanskiy [26].

Next, we construct an explicit coding scheme, involving classical post-processing including communication from the receiver to the sender, which matches the outer bound exactly. (See Figure 5 for schematic description of this code.) The strategy of the coding scheme is to generate maximally entangled qubit states using the quantum channels and then use the successfully transmitted (i.e. not erased) maximally entangled qubit states to distill a an entangled state of local dimension $|M|$, as required. (Note that the number $|M|$ is fixed at the outset of the code, i.e. the entanglement transmission scheme must deliver a maximally entangled state with local dimension $|M|$, possibly at the expense of low fidelity, rather than outputting a variable number of certifiably high fidelity entangled pairs.)

First, the encoder prepares n maximally entangled qubit states $|\psi\rangle$ and sends one half of each over the channel. The other halves, together with the untouched system M', are stored in the memory register Q. The decoder now works as follows. The receiver determines which qubits have not been erased and informs the sender of their locations. Let L be the random variable indicating the total number of erasures and note that L follows a binomial distribution with parameters n and β. Let us also fix $k = \lceil \log |M| \rceil$ and consider the following two cases:

1. If $L = l \leq n - k$ the decoder can extract a maximally entangled state with unit fidelity. To do so, it selects k perfectly transmitted entangled qubits at the sender and receiver. Let us assume (without loss of generality) that these are in a state $|\phi^+\rangle^{\otimes k} = \frac{1}{\sqrt{2^k}} \sum_{i=1}^{2^k} |ii\rangle$.

   The receiver then prepares a maximally entangled state of local dimension $|M|$ by measuring the k qubits with the projective measure

   $$\left\{ \frac{1}{\binom{2^k-1}{|M|-1}} \sum_{i \in \mathcal{S}} |i\rangle\langle i| : \mathcal{S} \subseteq \left[2^k\right] \wedge |\mathcal{S}| = |M| \right\} . \tag{67}$$

   The outcome, a subset $\mathcal{S}$ of cardinality $|M|$, is transmitted to the sender so that both sender and receiver now share a maximally entangled state on the subspace determined by $\mathcal{S}$.

2. On the other hand, if $L = l > n - k$ sender and receiver simply select the successfully transmitted qubits and embed them in a space of local dimension $|M|$. The fidelity with the target state $|\phi\rangle = \frac{1}{\sqrt{|M|}} \sum_{i=1}^{|M|} |ii\rangle$ is given by

   $$F\big(|\phi^+\rangle\langle\phi^+|^{\otimes(n-l)}, \phi\big) = \frac{1}{|M|} \sum_{i,j=1}^{|M|} \langle i| \Big(|\phi^+\rangle\langle\phi^+|^{\otimes(n-l)} \Big) |j\rangle = \frac{2^{n-l}}{|M|} . \tag{68}$$

To complete the decoding operation, the sender and receiver perform quantum teleportation to teleport M' to the receiver, using the maximally entangled state prepared above as a resource. The fidelity of the state prepared above with the target state $\phi_{MM'}$ is then just the expected fidelity over L, which evaluates to

$$F = \sum_{l=0}^{n-k} \binom{n}{l} \beta^l (1-\beta)^{n-l} + \sum_{l=n-k+1}^{n} \binom{n}{l} \beta^l (1-\beta)^{n-l} \frac{2^{n-l}}{|M|} \tag{69}$$

$$= 1 - \sum_{l=n-k+1}^{n} \binom{n}{l} \beta^l (1-\beta)^{n-l} \left(1 - \frac{2^{n-l}}{|M|} \right) . \tag{70}$$

[7] This corresponds to using Corollary 2 with $\mathcal{M}$ the channel which ignores its input and prepares σ_{B^n} at the output.

This is exactly the expression reported in the aforementioned outer bound in [27, Thm. 38], meaning the inner bound coincides with the outer bound when we allow classical post-processing and communication from the receiver to the sender.

D. The Qubit Depolarizing Channel

The qubit depolarizing channel is defined as

$$\mathcal{D}_\alpha : \rho \mapsto (1-\alpha)\rho + \frac{\alpha}{3}\left(X\rho X + Y\rho Y + Z\rho Z\right) , \tag{71}$$

where $\alpha \in [0,1]$ is a parameter and X, Y, Z are the Pauli operators. This channel is covariant since it is a qubit Pauli channel. Using the Bell states $\phi^+_{AB} = \phi_{AA'}$, $\phi^-_{AB} = (1_A \otimes Z_{A'})\phi_{AA'}(1_A \otimes Z_{A'})$, $\psi^+_{AB} = (1_A \otimes X_{A'})\phi_{AA'}(1_A \otimes X_{A'})$, and $\psi^-_{AB} = (1_A \otimes Y_{A'})\phi_{AA'}(1_A \otimes Y_{A'})$, we immediately find

$$\omega_{AB} := (\mathcal{I}_A \otimes \mathcal{D}_\alpha)(\phi_{AA'}) = (1-\alpha)\phi^+_{AB} + \frac{\alpha}{3}\left(\phi^-_{AB} + \psi^+_{AB} + \psi^-_{AB}\right) . \tag{72}$$

Now choosing $\sigma_{AB} = \frac{1}{2}\phi^+_{AB} + \frac{1}{6}(\phi^-_{AB} + \psi^+_{AB} + \psi^-_{AB})$ in (39) gives the outer bound

$$\hat{R}_{\mathcal{D}_\alpha}(n;\varepsilon) \leq \hat{R}^{\text{cpp}}_{\mathcal{D}_\alpha}(n;\varepsilon) \leq \frac{1}{n} D^\varepsilon_H\left(\omega^{\otimes n}_{AB} \middle\| \sigma^{\otimes n}_{AB}\right) . \tag{73}$$

As in the case of the qubit dephasing channel, we can convert the hypothesis test between ω_{AB} and σ_{AB} into a test between classical distributions, in fact precisely those distributions which were used in the dephasing example. This follows by considering the map which generates ϕ^+_{AB} when $X = Y$ and otherwise randomly generates one of the other Bell states when $X \neq Y$. Therefore, we obtain the same outer bound for the qubit depolarization channel as for the qubit dephasing channel. This raises the question of whether cpp assistance (or more generally PPT assistance, cf. Footnote 4) can turn the qubit depolarizing channel into the qubit dephasing channel.

VI. CONCLUSION

We gave inner (achievability) and outer (converse) bounds on the boundary of the achievable region for quantum communication with finite resources. We showed that these bounds agree for the qubit dephasing channel and qubit erasure channel with classical post-processing assistance up to the second order asymptotically. Moreover, we even gave a third order characterization for these specific examples by employing finite block length bounds of the binary symmetric channel and the binary erasure channel, respectively.

However, many questions remain open. For example, we would like to understand if the inner bound in Result 5 characterizes the achievable region for all degradable channels [10] (cf. the open questions in [42]). Also it would also be interesting to explore higher order refinements for channels with zero quantum capacity (e.g., for the erasure channel with $\beta \geq 1/2$ and no assistance). This might lead to a better understanding of super activation of the quantum capacity [36].

Finally, we would like to note that the recent results about finite resource entanglement assisted classical communication [8] can immediately be transformed to entanglement assisted quantum communication (and this then also gives outer bounds on the achievable rate region for unassisted codes). This is accomplished by using the equivalence results in [20, App. B] which make use of quantum teleportation and superdense coding. In particular, one finds that for covariant channels

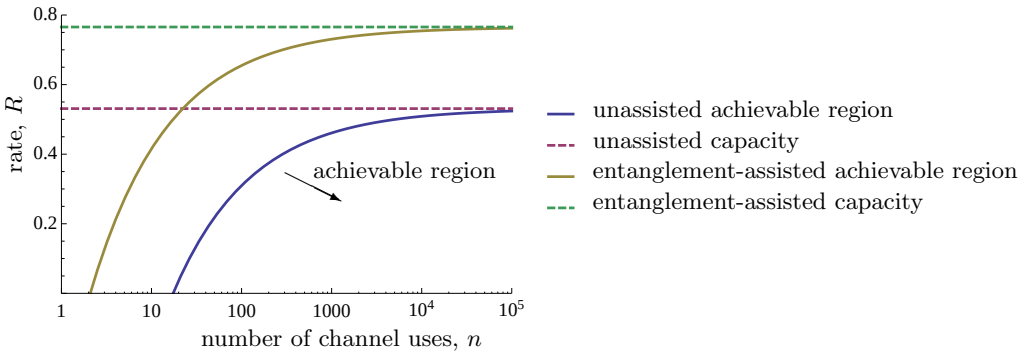


FIG. 6: Second order approximation of the achievable region of a qubit dephasing channel with $\varepsilon = 1\%$ and $\gamma = 0.1$; the achievable region is enlarged in the presence of entanglement [8].

$\mathcal{N}$ (which includes the qubit dephasing channel and the erasure channel) the boundary of the entanglement assisted achievable region $\hat{R}^{\mathrm{E}}(n;\varepsilon)$ satisfies

$$\hat{R}^{\mathrm{E}}(n;\varepsilon) = \frac{I(\mathcal{N})}{2} + \sqrt{\frac{V^{\varepsilon}(\mathcal{N})}{4n}}\Phi^{-1}(\varepsilon) + O\left(\frac{\log n}{n}\right), \tag{74}$$

with the mutual information of the channel, $I(\mathcal{N})$, and its variance, $V^{\varepsilon}(\mathcal{N})$, as defined in [8]. As an example, we mention again the qubit dephasing channel $\mathcal{Z}_{\gamma}$ for which

$$\hat{R}^{\mathrm{E}}(n;\varepsilon) = 1 - 2h(\gamma) + \sqrt{\frac{v(\gamma)}{4n}}\Phi^{-1}(\varepsilon) + O\left(\frac{\log n}{n}\right). \tag{75}$$

where $h(\gamma)$ denotes the binary entropy and $v(\gamma)$ the corresponding variance as defined in Result 1.

Acknowledgments. MT is funded by an University of Sydney Postdoctoral Fellowship and acknowledges support from the ARC Centre of Excellence for Engineered Quantum Systems (EQUS). JMR was supported by the Swiss National Science Foundation (through the National Centre of Competence in Research Quantum Science and Technology) and by the European Research Council (grant No. 258932). We thank Chris Ferrie, Chris Granade, William Matthews, David Sutter, and Mark Wilde for helpful discussions.

Appendix A: Semidefinite Optimization

In this section we describe how to formulate the outer bound from Corollary 2 as a semidefinite optimization program satisfying strong duality.

A semidefinite program (SDP) is simply an optimization of a linear function of a matrix or operator over a feasible set of inputs defined by positive semidefinite constraints. We give only the bare essentials here, for more detail see [5, 46]. The maximization form of an SDP is defined by a Hermiticity-preserving superoperator $\mathcal{E}_{A\to B}$ taking $\mathcal{L}(A)$ to $\mathcal{L}(B)$, a constraint operator $C \in \mathcal{L}(B)$, and an operator $K \in \mathcal{L}(A)$ which defines the objective function. The SDP is the following optimization, which we will also refer to as the primal form,

$$\begin{aligned} \alpha = \text{supremum} \quad & \mathrm{tr}[KX] \\ \text{subject to} \quad & \mathcal{E}(X) \leq C \\ & X \geq 0\,. \end{aligned} \tag{A1}$$

When the feasible set is empty, i.e. no X satisfy the constraints, we set $\alpha = -\infty$. The dual form arises as the optimal upper bound to the primal form, and takes the form

$$\begin{aligned} \beta = \text{infimum} \quad & \text{tr}[CY] \\ \text{subject to} \quad & \mathcal{E}^*(Y) \geq K \\ & Y \geq 0\,. \end{aligned} \tag{A2}$$

Again, when the set of feasible Y is empty, $\beta = \infty$. Weak duality is the statement that $\alpha \leq \beta$, that indeed the dual form gives upper bounds to the primal (or that the primal lower bounds the dual). Strong duality is the statement that the optimal upper bound equals the value of the primal problem, $\alpha = \beta$. This state of affairs often holds in problems of interest, and can be established by either of the following Slater conditions. In the first, called strict primal feasibility, strong duality holds if β is finite and there exists an $X > 0$ such that $\mathcal{E}(X) < C$. Contrariwise, under strict dual feasibility strong duality holds when α is finite and there exists a $Y > 0$ such that $\mathcal{E}^*(Y) > K$. For strongly dual SDPs we also have the so-called complementary slackness conditions $\mathcal{E}^*(Y)X = KX$ and $\mathcal{E}(X)Y = CY$ that relate the primal and dual optimizers.

Theorem 7. *With the notation from Corollary 2, the outer bound $f(\mathcal{N}, \varepsilon)$ can be written as*

$$\begin{aligned} f(\mathcal{N}, \varepsilon) \;=\; \text{infimum} \quad & \text{tr}[\xi_A] \\ \text{subject to} \quad & \xi_A \otimes 1_B \geq \Lambda_{AB} + \Theta_{AB}^{T_A} \\ & \Lambda_{AB} \in \Gamma(\rho_A, \mathcal{N}, \varepsilon); \rho_A \in \mathcal{S}(A) \\ & \xi_A, \Theta_{AB} \geq 0\,, \end{aligned} \tag{A3}$$

or, equivalently,

$$\begin{aligned} f(\mathcal{N}, \varepsilon) \;=\; \text{supremum} \quad & \mu(1-\varepsilon) - \nu \\ \text{subject to} \quad & \mu N_{AB} \leq M_{AB} + R_{AB} \\ & \text{tr}_B[R_{AB}] \leq n 1_A \\ & M_{AB} \in \text{PPT}; \mu, \nu, R_{AB} \geq 0\,. \end{aligned} \tag{A4}$$

Proof. The proof is straightforward: we simply use the dual of the inner optimization in (33) to obtain the minimization problem (A3). Then we use Slater's condition to show that strong duality holds and obtain (A4).

Consider the function

$$f_0(O_{AB}) := \sup_{\mathcal{M}_{A\to B} \in \text{PPT}} \text{tr}[O_{AB} M_{AB}]\,, \tag{A5}$$

and observe that f_0 is a semidefinite program. In particular, it is a primal problem as we have defined it, with $X = M_{AB}$, $K = O_{AB}$, $C = (0, 1_A)$, and $\mathcal{E}(X) = (-X^{T_A}, \text{tr}_B[X])$. Choosing for the dual variables $Y = (\Gamma_{AB}, \xi_A)$, the dual of f_0 is

$$\begin{aligned} \tilde{f}_0(O_{AB}) := \quad \text{infimum} \quad & \text{tr}[\xi_A] \\ \text{subject to} \quad & \xi_A \otimes 1_B \geq O_{AB} + \Gamma_{AB}^{T_A} \\ & \Gamma_{AB}, \xi_A \geq 0\,. \end{aligned} \tag{A6}$$

Combining this with the outer optimization in (33) gives the minimization program (A3). The equality statement is precisely strong duality of the primal and dual forms of the inner optimization. By Slater's condition, strong duality holds if f_0 is finite and there exists a strictly feasible set of dual variables. Observe that $f_0(O_{AB}) \leq |A|$, since for the optimal M_{AB} we have $f_0(O_{AB}) = \text{tr}[M_{AB} O_{AB}] \leq \text{tr}[M_{AB}] \leq \text{tr}_A[1_A] = |A|$. Here we have used the upper bounds $O_{AB} \leq 1_{AB}$ and

$\mathrm{tr}_B[M_{AB}] \leq 1_A$. Thus, the first condition is fulfilled. Meanwhile, $\Gamma_{AB} = 1_{AB}$ and $\xi_A = 3 \cdot 1_A$ are a strictly feasible pair. Thus, $\tilde{f}_0 = f_0$ over the domain of interest.

To construct the maximization program, we simply dualize the minimization program. In particular, $f(\mathcal{N}, \varepsilon)$ is a dual-form semidefinite program in the variable $Y = (\phi_A, \Lambda_{AB}, \Gamma_{AB}, \xi_A)$ with $C = (0, 0, 0, 1_A)$, $K = (1 - \varepsilon, -1, 0, 0)$, and

$$\mathcal{E}^*(Y) = (\mathrm{tr}[N_{AB}\Lambda_{AB}], -\mathrm{tr}[\phi_A], \phi_A^T \otimes 1_B - \Lambda_{AB}, \xi_A \otimes 1_B - \Lambda_{AB} - \Gamma_{AB}^{T_A}) \,. \tag{A7}$$

Choosing primal variables $X = (m, n, R_{AB}, M_{AB})$ leads to the maximization in (A4). Equality again follows from Slater's condition: f is finite (in particular the bound on f_0 used above), while a feasible choice of dual variables is given by $M_{AB} = R_{AB} = \frac{1}{2|B|} 1_{AB}$, $n = 1$, and $m = \frac{1}{2|A||B|}$. The choice of m ensures the first constraint holds strictly, since any Choi operator of a trace-preserving map satisfies $\|N_{AB}\|_\infty = |A|$ (largest singular value). □

No discussion of strong duality of semidefinite programs is complete until the complementary slackness conditions have been formulated. Often, these give considerable insight into the form and properties of the optimizing variables. First observe that

$$\mathcal{E}(X) = (-n1_A + \mathrm{tr}_B[R_{AB}^{T_A}],\, mN_{AB} - M_{AB} - R_{AB},\, -M_{AB}^{T_A},\, \mathrm{tr}_B[M_{AB}]) \,. \tag{A8}$$

Then the conditions are easy to read off from the form of C and K. They are

$$\mathrm{tr}[\phi_A] = 1 \tag{A9}$$
$$\mathrm{tr}[\Lambda_{AB} N_{AB}] = 1 - \varepsilon \tag{A10}$$
$$\phi_A^T R_{AB} = \Lambda_{AB} R_{AB} \tag{A11}$$
$$\xi_A M_{AB} = (\Lambda_{AB} + \Gamma_{AB}^{T_A}) M_{AB} \tag{A12}$$
$$n\phi_A = \mathrm{tr}_B[R_{AB}^{T_A}]\phi_A \tag{A13}$$
$$M_{AB}^{T_A} \Gamma_{AB} = 0 \tag{A14}$$
$$\mathrm{tr}_B[M_{AB}]\xi_A = \xi_A \tag{A15}$$
$$mN_{AB}\Lambda_{AB} = (M_{AB} + R_{AB})\Lambda_{AB} \,. \tag{A16}$$

[1] K. Audenaert, B. De Moor, K. Vollbrecht, and R. Werner. Asymptotic relative entropy of entanglement for orthogonally invariant states. *Phys. Rev. A*, 66(3):032310, 2002. DOI: 10.1103/PhysRevA.66.032310.

[2] H. Barnum, E. Knill, and M. A. Nielsen. On Quantum Fidelities and Channel Capacities. *IEEE Trans. on Inf. Theory*, 46:1317–1329, 2000.

[3] H. Barnum, M. A. Nielsen, and B. Schumacher. Information transmission through a noisy quantum channel. *Physical Review A*, 57(6):4153, 1998. DOI: 10.1103/PhysRevA.57.4153.

[4] M. Berta. Single-Shot Quantum State Merging. Master's thesis, ETH Zurich, 2008. arXiv: 0912.4495.

[5] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

[6] F. Buscemi and N. Datta. The Quantum Capacity of Channels With Arbitrarily Correlated Noise. *IEEE Trans. on Inf. Theory*, 56(3):1447–1460, 2010. DOI: 10.1109/TIT.2009.2039166.

[7] T. Cubitt, D. Elkouss, W. Matthews, M. Ozols, D. Perez-Garcia, and S. Strelchuk. Unbounded Number of Channel Uses are Required to See Quantum Capacity. *Nat. Commun.*, 6:6739, 2015. DOI: 10.1038/ncomms7739.

[8] N. Datta, M. Tomamichel, and M. M. Wilde. Second-Order Coding Rates for Entanglement-Assisted Communication. 2014. arXiv: 1405.1797.

[9] I. Devetak. The Private Classical Capacity and Quantum Capacity of a Quantum Channel. *IEEE Trans. on Inf. Theory*, 51(1):44–55, 2005. DOI: 10.1109/TIT.2004.839515.

[10] I. Devetak and P. W. Shor. The Capacity of a Quantum Channel for Simultaneous Transmission of Classical and Quantum Information. *Commun. Math. Phys.*, 256(2):287–303, 2005. DOI: 10.1007/s00220-005-1317-6.

[11] I. Devetak and A. Winter. Distillation of Secret Key and Entanglement from Quantum States. *Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.*, 461(2053):207–235, 2005.

[12] D. DiVincenzo, P. Shor, and J. Smolin. Quantum-channel capacity of very noisy channels. *Phys. Rev. A*, 57(2):830–839, 1998. DOI: 10.1103/PhysRevA.57.830.

[13] F. Dupuis. *The Decoupling Approach to Quantum Information Theory*. PhD thesis, Université de Montréal, 2009. arXiv: 1004.1641.

[14] F. Dupuis, M. Berta, J. Wullschleger, and R. Renner. One-Shot Decoupling. *Commun. Math. Phys.*, 328(1):251, 2014. DOI: 10.1007/s00220-014-1990-4.

[15] R. G. Gallager. *Information Theory and Reliable Communication*. Wiley, New York, 1968.

[16] M. Hayashi. Information Spectrum Approach to Second-Order Coding Rate in Channel Coding. *IEEE Trans. on Inf. Theory*, 55(11):4947–4966, 2009. DOI: 10.1109/TIT.2009.2030478.

[17] M. Hayashi and M. Tomamichel. Correlation Detection and an Operational Interpretation of the Renyi Mutual Information. 2014. arXiv: 1408.6894.

[18] P. Hayden, M. Horodecki, J. Yard, and A. Winter. A Decoupling Approach to the Quantum Capacity. *Open Syst. Inf. Dyn.*, 15(01):7, 2008. DOI: 10.1142/S1230161208000043.

[19] P. Horodecki, M. Horodecki, and R. Horodecki. Binding entanglement channels. *Journal of Modern Optics*, 47(2):347, 2000. DOI: 10.1080/09500340008244047.

[20] D. Leung and W. Matthews. On the Power of PPT-Preserving and Non-Signalling Codes. 2014. arXiv: 1406.7142.

[21] K. Li. Second-Order Asymptotics for Quantum Hypothesis Testing. *Ann. Stat.*, 42(1):171–189, 2014. DOI: 10.1214/13-AOS1185.

[22] S. Lloyd. The Capacity of The Noisy Quantum Channel. *Phys. Rev. A*, 55(3):1613–1622, 1996. DOI: 10.1103/PhysRevA.55.1613.

[23] W. Matthews. A linear program for the finite block length converse of Polyanskiy-Poor-Verd\'u via non-signalling codes. page 8, 2011. arXiv: 1109.5417.

[24] C. Morgan and A. Winter. Pretty Strong Converse for the Quantum Capacity of Degradable Channels. *IEEE Trans. on Inf. Theory*, 60(1):317–333, 2014. DOI: 10.1109/TIT.2013.2288971.

[25] G. Poltyrev. Bounds on the Decoding Error Probability of Binary Linear Codes via Their Spectra. *IEEE Trans. on Inf. Theory*, 40(4):1284–1292, 1994. DOI: 10.1109/18.335935.

[26] Y. Polyanskiy. Saddle Point in the Minimax Converse for Channel Coding. *IEEE Trans. on Inf. Theory*, 59(5):2576–2595, 2013. DOI: 10.1109/TIT.2012.2236382.

[27] Y. Polyanskiy, H. V. Poor, and S. Verdú. Channel Coding Rate in the Finite Blocklength Regime. *IEEE Trans. on Inf. Theory*, 56(5):2307–2359, 2010. DOI: 10.1109/TIT.2010.2043769.

[28] E. Rains. Bound on Distillable Entanglement. *Phys. Rev. A*, 60(1):179, 1999. DOI: 10.1103/PhysRevA.60.179.

[29] E. Rains. A semidefinite program for distillable entanglement. *IEEE Trans. on Inf. Theory*, 47(7):2921–2933, 2001. DOI: 10.1109/18.959270.

[30] R. Renner. *Security of Quantum Key Distribution*. PhD thesis, ETH Zurich, 2005. arXiv: quant-ph/0512258.

[31] B. Schumacher and M. A. Nielsen. Quantum data processing and error correction. *Physical Review A*, 54(4):2629, 1996. DOI: 10.1103/PhysRevA.54.2629.

[32] C. Shannon. A Mathematical Theory of Communication. *Bell Syst. Tech. J.*, 27:379–423, 1948.

[33] P. W. Shor. The Quantum Channel Capacity and Coherent Information. In *Lectures Notes, MSRI Workshop on Quantum Computation*, 2002.

[34] G. Smith and J. A. Smolin. Additive Extensions of a Quantum Channel. In *Proc. IEEE ITW*, pages 368–372. IEEE, 2008. DOI: 10.1109/ITW.2008.4578688.

[35] G. Smith, J. A. Smolin, and A. Winter. The Quantum Capacity with Symmetric Side Channels. *IEEE Trans. on Inf. Theory*, 54(9):4208–4217, 2008. DOI: 10.1109/TIT.2008.928269.

[36] G. Smith and J. T. Yard. Quantum Communication with Zero-Capacity Channels. *Science*, 321(5897):1812–1815, 2008. DOI: 10.1126/science.1162242.

[37] D. Sutter, V. B. Scholz, and R. Renner. Approximate Degradable Quantum Channels. 2014. arXiv: 1412.0980.

[38] V. Y. F. Tan. Asymptotic Estimates in Information Theory with Non-Vanishing Error Probabilities. *Found. Trends Commun. Inf. Theory*, 10(4):1–184, 2014. DOI: 10.1561/0100000086.

[39] M. Tomamichel, R. Colbeck, and R. Renner. Duality Between Smooth Min- and Max-Entropies. *IEEE Trans. on Inf. Theory*, 56(9):4674–4681, 2010. DOI: 10.1109/TIT.2010.2054130.

[40] M. Tomamichel and M. Hayashi. A Hierarchy of Information Quantities for Finite Block Length Analysis of Quantum Tasks. *IEEE Trans. on Inf. Theory*, 59(11):7693–7710, 2013. DOI: 10.1109/TIT.2013.2276628.

[41] M. Tomamichel and V. Y. F. Tan. On the Gaussian Approximation for the Classical Capacity of Quantum Channels. 2014. arXiv: 1308.6503.

[42] M. Tomamichel, M. M. Wilde, and A. Winter. Strong Converse Rates for Quantum Communication. 2014. arXiv: 1406.2946.

[43] A. Uhlmann. The Transition Probability for States of Star-Algebras. *Ann. Phys.*, 497(4):524–532, 1985.

[44] H. Umegaki. Conditional Expectation in an Operator Algebra. *Kodai Math. Sem. Rep.*, 14:59–85, 1962.

[45] L. Wang and R. Renner. One-Shot Classical-Quantum Capacity and Hypothesis Testing. *Phys. Rev. Lett.*, 108(20):200501, 2012. DOI: 10.1103/PhysRevLett.108.200501.

[46] J. Watrous. Theory of Quantum Information, Lecture Notes, 2011. Available online: https://cs.uwaterloo.ca/~watrous/quant-info/.