

1-1-2014

## Strong converse for the classical capacity of the pure-loss bosonic channel

M. M. Wilde  
*Louisiana State University*

A. Winter  
*Universitat Autònoma de Barcelona*

Follow this and additional works at: [https://repository.lsu.edu/physics\\_astronomy\\_pubs](https://repository.lsu.edu/physics_astronomy_pubs)

---

### Recommended Citation

Wilde, M., & Winter, A. (2014). Strong converse for the classical capacity of the pure-loss bosonic channel. *Problems of Information Transmission*, 50 (2), 117-132. <https://doi.org/10.1134/S003294601402001X>

This Article is brought to you for free and open access by the Department of Physics & Astronomy at LSU Scholarly Repository. It has been accepted for inclusion in Faculty Publications by an authorized administrator of LSU Scholarly Repository. For more information, please contact [ir@lsu.edu](mailto:ir@lsu.edu).

# Strong converse for the classical capacity of the pure-loss bosonic channel

Mark M. Wilde\*      Andreas Winter<sup>†‡</sup>

5 November 2013

## Abstract

This paper strengthens the interpretation and understanding of the classical capacity of the pure-loss bosonic channel, first established in [Giovannetti *et al.*, *Physical Review Letters* **92**, 027902 (2004), arXiv:quant-ph/0308012]. In particular, we first prove that there exists a trade-off between communication rate and error probability if one imposes only a mean-photon number constraint on the channel inputs. That is, if we demand that the mean number of photons at the channel input cannot be any larger than some positive number  $N_S$ , then it is possible to respect this constraint with a code that operates at a rate  $g(\eta N_S/(1-p))$  where  $p$  is the code's error probability,  $\eta$  is the channel transmissivity, and  $g(x)$  is the entropy of a bosonic thermal state with mean photon number  $x$ . We then prove that a strong converse theorem holds for the classical capacity of this channel (that such a rate-error trade-off cannot occur) if one instead demands for a maximum photon number constraint, in such a way that mostly all of the “shadow” of the average density operator for a given code is required to be on a subspace with photon number no larger than  $nN_S$ , so that the shadow outside this subspace vanishes as the number  $n$  of channel uses becomes large. Finally, we prove that a small modification of the well-known coherent-state coding scheme meets this more demanding constraint.

## Introduction

The pure-loss bosonic channel is one of the most important communication channels studied in quantum information theory [18, 20]. It has acquired this elevated status because it is a simple model for free-space communication or transmission over fiber optic cables. Indeed, a normal-mode decomposition of a quantized propagating electromagnetic field in free space leads naturally to a set of orthogonal spatio-temporal modes [18], such that the  $k$ th output mode can be expressed in terms of the  $k$ th input and environment modes as follows:

$$\hat{b}_k = \sqrt{\eta_k} \hat{a}_k + \sqrt{1 - \eta_k} \hat{e}_k, \quad (1)$$

where  $\hat{a}_k$ ,  $\hat{b}_k$ , and  $\hat{e}_k$  are the annihilation operators corresponding to the  $k$ th field mode of the sender, receiver, and environment, respectively. For the pure-loss channel, the environment modes

---

\*Department of Physics and Astronomy, Center for Computation and Technology, Louisiana State University, Baton Rouge, Louisiana 70803, USA

<sup>†</sup>ICREA & Física Teòrica: Informació i Fenòmens Quàntics, Universitat Autònoma de Barcelona, ES-08193 Bellaterra (Barcelona), Spain

<sup>‡</sup>School of Mathematics, University of Bristol, Bristol BS8 1TW, United Kingdom

are originally prepared in the vacuum state. The transmissivity parameter  $\eta_k \in [0, 1]$  characterizes (roughly) the fraction of photons that make it through one of these channels on average to the receiver, and the state prepared at each environment mode is the vacuum state. When attempting to gain an information-theoretic understanding of free space communication, it is often more convenient and simpler to focus on a single-mode channel of the form in (1), rather than the full normal mode decomposition, and this is what we refer to as the pure-loss bosonic channel.

If we allow for signal states at the input of the pure-loss bosonic channel that have an arbitrarily large number of photons, then the classical capacity of this channel is infinite. This is because, with infinite-energy signal states, one can space them out in such a way that there are an infinite number of them that are all arbitrarily well distinguishable from one another, at the channel output. Thus, in order to have a sensible notion of classical capacity for this channel, we should impose a constraint on the photon number  $a^\dagger a$  (or, equivalently, energy  $H = a^\dagger a + \frac{1}{2}$  of the signal states). One natural constraint is on the mean photon number—i.e., we demand that the mean number of photons in any codeword transmitted through the channel should be no larger than some number  $N_S \geq 0$ . With such a constraint, the classical capacity of this channel is equal to  $g(\eta N_S)$  [6], where

$$g(x) \equiv (x + 1) \log_2(x + 1) - x \log_2 x \quad (2)$$

is the entropy of a bosonic thermal state with mean photon number  $x$ . This result follows from a proof that there exists a coding scheme that can achieve this rate [10], and a matching converse proof that demonstrates it is impossible to have perfectly reliable communication if the rate exceeds  $g(\eta N_S)$  [6].<sup>1</sup>

Although Ref. [6] proved that  $g(\eta N_S)$  is equal to the classical capacity of the pure-loss bosonic channel, the converse theorem used there is only a “weak converse,” meaning that the upper bound on the rate  $R$  of any coding scheme for this channel with error probability  $\varepsilon$  is of the following form:

$$R \leq \frac{1}{1 - \varepsilon} [g(\eta N_S) + h_2(\varepsilon)], \quad (3)$$

where  $h_2(\varepsilon)$  is the binary entropy, with the property that  $\lim_{\varepsilon \rightarrow 0} h_2(\varepsilon) = 0$ . Thus, in order to establish  $g(\eta N_S)$  as the capacity, one really needs to take the limit in (3) as the error probability  $\varepsilon \rightarrow 0$ . This in fact is the hallmark of a weak converse—it leaves room for a trade-off between rate and error, and suggests that one might be able to attain a higher communication rate by allowing for some error.

A strong converse theorem demonstrates that there is no such trade-off in the limit of large blocklength. That is, a strong converse theorem holds if the error probability converges to one in the limit of many channel uses when the communication rate of a coding scheme exceeds the classical capacity. Thus, such a theorem improves our understanding of the capacity as a sharp dividing line between what communication rates are possible or impossible, and in this sense, it is analogous to a phase transition in statistical physics. Furthermore, there are applications of strong converse theorems in establishing security for particular models of cryptography [12].

Several prior works have established the strong converse for the classical capacity of certain quantum channels. There are two independent proofs [23] and [14] that the strong converse theorem holds for any discrete memoryless channel with a classical input and finite-dimensional quantum

---

<sup>1</sup>The reader might also consider related work on classical communication over noiseless bosonic channels [26], classical communication over pure-state [7] and general quantum channels [17, 9], and other schemes for decoding the pure-loss bosonic channel [21].

output, so-called “cq-channels.” More generally, it holds for arbitrary finite dimensional quantum channels with product state encoding [14, 24]. Many years later, it was shown that the strong converse holds for covariant channels for which their maximum output  $p$ -norm is multiplicative [11]. Finally, recent work has shown that the strong converse holds for all entanglement-breaking and Hadamard channels [22].

## Summary of results

This paper establishes several facts regarding classical communication over the pure-loss bosonic channel:

1. If we demand only that the mean photon number is no larger than some number  $N_S \geq 0$ , then we show that the strong converse does not hold (see Section 1). We can show that this is the case even if we restrict the codewords to be pure states. In some sense, this latter result provides a distinction between the classical and quantum theories of information for continuous variables, but it does bear some similarities with the observations in Theorem 77 of Ref. [16] and we remark on this point further in Section 1.
2. In light of the above result, we can only hope to prove that the strong converse holds under some alternate photon number constraint. Let  $\rho_m$  denote an  $n$ -mode codeword in a given codebook, so that the average code density operator is given by  $\frac{1}{M} \sum_m \rho_m$ , where  $M$  is the total number of messages. We instead demand that the average code density operator satisfies a maximum photon number constraint, such that it should have a large “shadow” onto a subspace of photon number no more than  $\lceil nN_S \rceil$ :

$$\frac{1}{M} \sum_m \text{Tr}\{\Pi_{\lceil nN_S \rceil} \rho_m\} \geq 1 - \delta(n). \quad (4)$$

In the above,  $\Pi_{\lceil nN_S \rceil}$  is the projector onto a subspace with photon number no larger than  $\lceil nN_S \rceil$  and  $\delta(n)$  decreases to zero with increasing  $n$ . Under such a constraint, we prove in Section 2 that the strong converse holds for the classical capacity of the pure-loss bosonic channel.

3. Finally, we prove in Section 3 that there exist codes for the pure-loss bosonic channel that meet the constraint in (4) while having an error probability that can be less than an arbitrarily small constant for a sufficiently large number of channel uses. Indeed, we show that the usual coherent state encoding scheme essentially satisfies the constraint.

## 1 No strong converse under a mean photon number constraint

We first prove that a strong converse does not hold for the classical capacity of the pure-loss bosonic channel if we impose only a mean photon number constraint. Indeed, a method for proving the existence of a code that achieves the classical capacity of the pure-loss bosonic channel is first to sample coherent-state codewords independently from a circularly-symmetric complex Gaussian distribution with variance  $N_S$  [10, 6]. Let

$$|\alpha^n(m)\rangle \equiv |\alpha_1(m)\rangle \otimes \cdots \otimes |\alpha_n(m)\rangle$$

denote each of the  $n$ -mode coherent state codewords (with the dependence on the message  $m$  explicitly indicated) and let  $[M]$  denote a message set of size  $M$ . Then one can prove that there exists a choice of codebook such that every codeword in the resulting codebook  $\{|\alpha^n(m)\rangle\}_{m \in [M]}$  satisfies the mean photon number constraint. Let  $|\beta^n(m)\rangle$  denote the state resulting from sending the codeword  $|\alpha^n(m)\rangle$  through the pure-loss bosonic channel, so that

$$|\beta^n(m)\rangle \equiv |\sqrt{\eta}\alpha_1(m)\rangle \otimes \cdots \otimes |\sqrt{\eta}\alpha_n(m)\rangle, \quad (5)$$

and  $\eta$  is the channel transmissivity parameter. Furthermore, this choice of codebook is such that the receiver can decode the transmitted codewords with arbitrarily high success probability by performing a square-root measurement [7] or a sequential decoding measurement [21], for example. That is, as long as  $(\log_2 M)/n \approx g(\eta N_S)$  and the number  $n$  of channel uses is sufficiently large, there exists a measurement  $\{\Lambda_m\}_{m \in [M]}$  and codebook such that

$$\forall m \in [M] : \text{Tr}\{\Lambda_m |\beta^n(m)\rangle\langle\beta^n(m)|\} \geq 1 - \varepsilon,$$

where  $\varepsilon$  is an arbitrarily small positive number. For finite  $n$ ,  $M$ , and  $\varepsilon$ , we call such a code an  $(n, M, \varepsilon)$  code.

### 1.1 No strong converse with mixed-state codewords

Now, consider a pure-loss bosonic channel with transmissivity parameter  $\eta$  and mean photon number constraint  $N_S$ . As we have stated before, the classical capacity of this channel is equal to  $g(\eta N_S)$  [6]. We show that a strong converse theorem cannot hold—our approach is to employ a codebook of the above form at a rate larger than  $g(\eta N_S)$ . Indeed, let  $\{|\alpha^n(m)\rangle\}_{m \in [M]}$  now denote a codebook such that each codeword has mean photon number  $P > N_S$ , let the rate of the code be  $(\log_2 M)/n \approx g(\eta P) > g(\eta N_S)$ , and let  $\{\Lambda_m\}_{m \in [M]}$  be a decoding measurement for this codebook. We then modify the codewords as follows:

$$\rho(m) \equiv (1 - p)|\alpha^n(m)\rangle\langle\alpha^n(m)| + p(|0\rangle\langle 0|)^{\otimes n}, \quad (6)$$

where  $p$  is such that  $0 \leq p \leq 1$  and  $(1 - p)P = N_S$  and  $|0\rangle^{\otimes n}$  is the  $n$ -fold tensor product vacuum state. Observe that the mean photon number of each codeword  $\rho(m)$  is equal to  $(1 - p)P = N_S$  (because the vacuum contributes nothing to the photon number). Then the state resulting from transmitting these codewords through the pure-loss bosonic channel is as follows:

$$(1 - p)|\beta^n(m)\rangle\langle\beta^n(m)| + p(|0\rangle\langle 0|)^{\otimes n}, \quad (7)$$

with  $|\beta^n(m)\rangle$  defined in the same way as in (5). The success probability for every codeword is as follows:

$$\text{Tr}\{\Lambda_m ((1 - p)|\beta^n(m)\rangle\langle\beta^n(m)| + p(|0\rangle\langle 0|)^{\otimes n})\} \geq (1 - p)\text{Tr}\{\Lambda_m |\beta^n(m)\rangle\langle\beta^n(m)|\} \quad (8)$$

$$\geq (1 - p)(1 - \varepsilon), \quad (9)$$

simply by using the fact that the decoding measurement  $\{\Lambda_m\}$  decodes the codewords  $\{|\beta^n(m)\rangle\}$  with success probability larger than  $1 - \varepsilon$ . Thus, the success probability need not converge to zero in the limit of many channel uses if the rate of the code exceeds the classical capacity and we impose only a mean photon number constraint on each codeword in the codebook.

**Remark 1** *There is nothing particularly “quantum” about the above argument, other than the fact that we employ mixtures of quantum states as codewords and a quantum measurement for decoding. Indeed, the same argument demonstrates that a strong converse does not hold when we impose only a mean power constraint for each classical codeword transmitted over the classical additive white Gaussian noise channel but allow for probabilistic encodings. That is, for a given codebook where each codeword has mean power  $P$ , we can always produce a new codebook such that each codeword is a  $(1-p, p)$  Bernoulli mixture of a codeword from the original codebook and the all-zero signal. One then produces a codebook where each codeword has mean power  $(1-p)P$ , and we can violate the strong converse using an argument very similar to the above one. However, if we restrict to deterministic or classical “pure-state” encodings, then a strong converse theorem does hold as shown in Ref. [16].*

## 1.2 No strong converse with pure-state codewords

This section demonstrates an important distinction between the classical and quantum theories of information for continuous variables: we show that a strong converse does not hold when imposing a mean photon-number constraint and even when we restrict to pure-state codewords (see Remark 1 above). Our argument is similar to the one in the previous section.

We are again considering the pure-loss bosonic channel with transmissivity parameter  $\eta$  and mean photon number constraint  $N_S$ . We will show the existence of a codebook with pure-state codewords each with mean photon number  $N_S$  such that the rate is strictly larger than  $g(\eta N_S)$  while the success probability of decoding is bounded from below by a constant number between zero and one.

The idea is similar to that in the previous section, except we make our codewords be superpositions of the following form:

$$|\gamma_p(m)\rangle \equiv \sqrt{1-p}|\alpha^n(m)\rangle|0\rangle + \sqrt{p}|0\rangle^{\otimes n}|1\rangle, \quad (10)$$

so that we add just one more mode to send through the channel that has a negligible effect on the parameters of the code (an  $(n, M, \varepsilon)$  code becomes an  $(n+1, M, \varepsilon)$  code, which is a negligible change for large  $n$ ). This extra mode uses a single photon to purify the state in (6). The mean photon number of the codeword  $|\gamma_p(m)\rangle$  is equal to

$$\begin{aligned} & \text{Tr} \left\{ \frac{1}{n+1} \sum_{i=1}^{n+1} \hat{a}_i^\dagger \hat{a}_i |\gamma_p(m)\rangle \langle \gamma_p(m)| \right\} \\ &= \text{Tr} \left\{ \frac{1}{n+1} \sum_{i=1}^{n+1} \hat{a}_i^\dagger \hat{a}_i \left[ \begin{array}{l} (1-p)|\alpha^n(m)\rangle \langle \alpha^n(m)| \otimes |0\rangle \langle 0| + \sqrt{(1-p)p} |\alpha^n(m)\rangle \langle 0|^{\otimes n} \otimes |0\rangle \langle 1| \\ + \sqrt{(1-p)p} |0\rangle^{\otimes n} \langle \alpha^n(m)| \otimes |1\rangle \langle 0| + p |0\rangle \langle 0|^{\otimes n} \otimes |1\rangle \langle 1| \end{array} \right] \right\} \\ &= (1-p) \frac{nP}{n+1} + p \frac{1}{n+1}. \end{aligned}$$

Thus, again by tuning  $p$  and choosing  $n$  large enough, we can set  $(1-p) \frac{nP}{n+1} + p \frac{1}{n+1} = N_S$ . The receiver operates by tracing over the very last mode, which dephases the outputs to be of the form in (7), and he then uses the decoding measurement for the codebook  $\{|\alpha^n(m)\rangle\}$ . More explicitly, the receiver measures the  $n+1$  output modes with the decoding POVM  $\{\Lambda_m \otimes I\}$ , where the operators  $\Lambda_m$  are the same as those from the previous section (acting on the first  $n$  modes) and

the identity operator acts on the very last mode. We then find the following bound on the success probability for every codeword:

$$\begin{aligned} \text{Tr}\{(\Lambda_m \otimes I)\mathcal{N}^{\otimes(n+1)}(|\gamma_p(m)\rangle\langle\gamma_p(m)|)\} &= \text{Tr}\{\Lambda_m\mathcal{N}^{\otimes n}(\rho(m))\} \\ &= \text{Tr}\{\Lambda_m((1-p)|\beta^n(m)\rangle\langle\beta^n(m)| + p(|0\rangle\langle 0|)^{\otimes n})\} \\ &\geq (1-p)(1-\varepsilon). \end{aligned}$$

The first equality follows from the fact that the channel is completely positive and trace preserving and from the fact that the state  $|\gamma_p(m)\rangle$  defined in (10) is a purification of the state  $\rho(m)$  defined in (6). The next equality is from the definition of  $\rho(m)$ . The last inequality follows from the argument in (8)-(9). The resulting code operates at a rate  $\approx g(\eta P)$  while meeting the mean photon number constraint. Thus, a strong converse theorem cannot hold for the pure-loss bosonic channel with only a mean photon number constraint and when restricting to pure-state codewords.

**Remark 2** *The coding scheme given both in this section and the previous one demonstrate that we can achieve a rate-error trade-off of  $(g(\eta N_S/(1-p)), p)$  for all  $0 \leq p \leq 1$ , where  $p$  is the error probability and  $N_S$  is the mean photon number constraint. This result complements the classical result in Theorem 77 of Ref. [16]. However, in our case, we have not proved that this trade-off is optimal, merely that it is achievable.*

## 2 Strong converse under a maximum photon number constraint

In this section, we prove that the strong converse holds when imposing a particular maximum photon number constraint. Our approach for proving the strong converse theorem is a simulation based argument, similar to that used in Refs. [25, 1, 2, 3]. We can illustrate the main idea behind this argument by recalling a simple proof of the strong converse theorem for the noiseless qubit channel [11, 13]. Consider that any scheme for classical communication over  $n$  noiseless qubit channels consists of an encoding of a message  $m$  as a quantum state  $\rho_m$  on  $n$  qubits, followed by a decoding measurement  $\{\Lambda_m\}$  to recover the message. Let  $M$  be the total number of messages, so that the rate of the code is  $R \equiv \frac{1}{n} \log_2(M)$ . The average success probability of this scheme is bounded as follows:

$$\begin{aligned} \frac{1}{M} \sum_m \text{Tr}\{\Lambda_m \rho_m\} &\leq \frac{1}{M} \sum_m \text{Tr}\{\Lambda_m\} \\ &= M^{-1} 2^n \\ &= 2^{-n(R-1)}, \end{aligned} \tag{11}$$

where the first inequality follows from the operator inequality  $\rho_m \leq I$ , which holds for any density operator, and the first equality follows because  $\sum_m \Lambda_m = I^{\otimes n}$  (the identity operator on  $n$  qubits). Thus, it is clear that if the rate  $R$  exceeds one, then the average success probability of any communication scheme decreases exponentially fast to zero with increasing blocklength.

Our proof of the strong converse for the classical capacity of the pure-loss bosonic channel is similar in spirit to the above argument, but it requires some nontrivial additions. First, we show that there is a simple protocol using approximately  $ng(\eta N_S)$  noiseless qubit channels to faithfully simulate the action of  $n$  instances of the pure-loss bosonic channel with transmissivity  $\eta$  when

acting on a space with total photon number less than  $nN_S$ , such that the simulation error becomes arbitrarily small as  $n$  becomes large. So we can combine this simulation protocol with any classical code  $\mathcal{C}$  for the pure-loss bosonic channel in which each codeword has almost all of its “shadow” on a space with total photon number less than  $nN_S$ . Letting  $\mathcal{N}$  denote the pure-loss bosonic channel, we can phrase this simulation argument in the language of resource inequalities [5] as follows:

$$g(\eta N_S)[q \rightarrow q] \geq \langle \mathcal{N} : \mathcal{C} \rangle.$$

The meaning of the above resource inequality is that one can simulate the action of  $n$  instances of the pure-loss bosonic channel on the codewords in  $\mathcal{C}$  by exploiting noiseless qubit channels at a rate equal to  $g(\eta N_S)$ , and this simulation becomes perfect in the limit as  $n$  becomes large. If one could send classical information over the pure-loss bosonic channel at a rate  $R$  larger than  $g(\eta N_S)$ , then it would be possible to serially concatenate the above protocol with a classical coding scheme for the channel  $\mathcal{N}$  and achieve the following resource inequality:

$$g(\eta N_S)[q \rightarrow q] \geq \langle \mathcal{N} : \mathcal{C} \rangle \geq R[c \rightarrow c],$$

where  $R[c \rightarrow c]$  denotes noiseless classical communication at rate  $R$ . Since the above protocol would give a strong violation of the Holevo bound for  $R > g(\eta N_S)$  (in particular, the refinement given in (11)), it must not be possible. In fact, essentially the same argument as in (11) demonstrates that the error probability of the classical communication protocol goes exponentially fast to one if the rate  $R$  is strictly larger than the classical capacity of this channel. Figure 1 depicts this simulation argument.

Let  $\Pi_L$  denote the projector onto a subspace of  $n$  bosonic modes, such that the total photon number is no larger than  $L$ :

$$\Pi_L \equiv \sum_{a_1, \dots, a_n: \sum_i a_i \leq L} |a_1\rangle\langle a_1| \otimes \dots \otimes |a_n\rangle\langle a_n|,$$

where  $|a_i\rangle$  is a photon number state of photon number  $a_i$ . We call  $\Pi_L$  the “photon number cutoff projector” in what follows.

We begin by proving two important lemmas.

**Lemma 3** *The rank of the photon number cutoff projector  $\Pi_{\lceil nN_S \rceil}$ , where  $\lceil \cdot \rceil$  denotes the ceiling function, is no larger than*

$$2^{n[g(N_S) + \delta]},$$

where

$$\delta \geq \frac{1}{n} \left( \log e + \log \left( 1 + \frac{1}{N_S} \right) \right).$$

(Clearly, we can pick  $\delta$  to be an arbitrarily small positive constant by taking  $n$  to be large enough.)

**Proof.** Consider that the rank of the photon number cutoff projector is exactly equal to

$$\sum_{j=0}^{\lceil nN_S \rceil} \binom{j+n-1}{n-1} = \binom{\lceil nN_S \rceil + n}{n}.$$



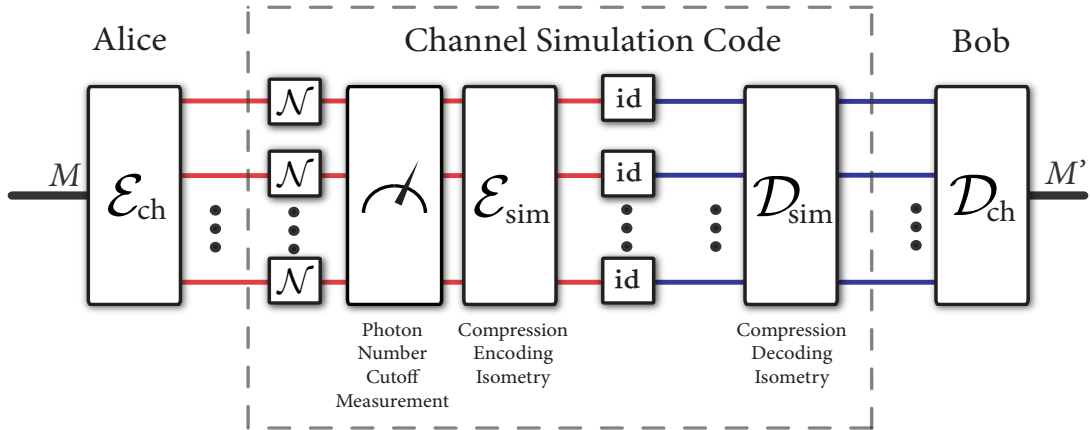


Figure 1: The simulation argument for the strong converse of the classical capacity of the pure-loss bosonic channel with transmissivity  $\eta \in [0, 1]$ . If an  $n$ -mode quantum state input to  $n$  uses of the pure-loss bosonic channel has nearly all of its “shadow” on a subspace with photon number no larger than  $nN_S$ , then one can simulate this channel with high fidelity on any such input at a rate of  $g(\eta N_S)$  qubits per mode. If it were possible to send classical information over the pure-loss bosonic channel at a rate  $R > g(\eta N_S)$ , then Alice and Bob could combine a channel code with the simulation code and violate the Holevo bound. Analyzing this contradiction in more detail allows us to conclude the strong converse theorem. The simulation protocol in the figure begins with Alice encoding a message  $M$  using the channel encoder  $\mathcal{E}_{\text{ch}}$ . Alice proceeds to simulate the channel by first actually performing it on all of the input modes and then projecting onto a subspace with photon number no larger than  $\approx n\eta N_S$ . This measurement succeeds with high probability and then she can compress the quantum data at the output of the measurement at a rate of  $g(\eta N_S)$  qubits per mode (sending this quantum data over noiseless qubit channels, denoted by “id” in the figure for “identity” channels). Bob then decompresses the quantum data, completing the channel simulation, and he finally decodes the classical message using the channel decoder  $\mathcal{D}_{\text{ch}}$ .

This is because we can enumerate the photon number eigenstates as tuples  $(a_1, \dots, a_n)$  of non-negative integers such that  $\sum_i a_i \leq \lceil nN_S \rceil$ , which equals the number of tuples  $(a_0, a_1, \dots, a_n)$  of  $a_i \geq 0$  such that  $\sum_{i=0}^n a_i = \lceil nN_S \rceil$ . In other words, we wish to count the unordered partitions of  $\lceil nN_S \rceil$  into  $n + 1$  non-negative integer pieces, and these are in turn in one-to-one correspondence with selecting  $n$  “separator” positions in  $\lceil nN_S \rceil + n$  to break a block of  $\lceil nN_S \rceil + n$  into  $n + 1$  pieces of positive integer size, which is given by the binomial coefficient on the right hand side above. We can then bound the rank of  $\Pi_{\lceil nN_S \rceil}$  as follows:

$$\begin{aligned}
 \text{rank } \Pi_{\lceil nN_S \rceil} &= \sum_{j=0}^{\lceil nN_S \rceil} \binom{j+n-1}{n-1} = \binom{\lceil nN_S \rceil + n}{n} \\
 &\leq 2^{(\lceil nN_S \rceil + n)h_2(n/(\lceil nN_S \rceil + n))} \\
 &\leq 2^{n[g(N_S) + \delta]}
 \end{aligned}$$

where  $\delta \geq \frac{1}{n} \left( \log e + \log \left( 1 + \frac{1}{N_S} \right) \right)$ . The first line of identities we have just argued, and the inequality in the second line is a well known combinatorial result (see Example 11.1.3 of Ref. [4]), where  $h_2(p) \equiv -p \log_2 p - (1-p) \log_2 (1-p)$ . The last inequality follows from

$$\begin{aligned}
& (\lceil nN_S \rceil + n) h_2(n/(\lceil nN_S \rceil + n)) \\
&= (\lceil nN_S \rceil + n) \left[ -\frac{n}{\lceil nN_S \rceil + n} \log \left( \frac{n}{\lceil nN_S \rceil + n} \right) - \frac{\lceil nN_S \rceil}{\lceil nN_S \rceil + n} \log \left( \frac{\lceil nN_S \rceil}{\lceil nN_S \rceil + n} \right) \right] \\
&= -n \log \left( \frac{n}{\lceil nN_S \rceil + n} \right) - \lceil nN_S \rceil \log \left( \frac{\lceil nN_S \rceil}{\lceil nN_S \rceil + n} \right) \\
&= n \log \left( \frac{\lceil nN_S \rceil + n}{n} \right) + \lceil nN_S \rceil \log \left( 1 + \frac{n}{\lceil nN_S \rceil} \right) \\
&\leq n \log \left( \frac{nN_S + n + 1}{n} \right) + (nN_S + 1) \log \left( 1 + \frac{1}{N_S} \right) \\
&\leq n \log(N_S + 1) + \log e + nN_S \log \left( 1 + \frac{1}{N_S} \right) + \log \left( 1 + \frac{1}{N_S} \right) \\
&= ng(N_S) + \log e + \log \left( 1 + \frac{1}{N_S} \right).
\end{aligned}$$

The first few equalities are simple algebra. The first inequality follows from the definition of the ceiling function. The second inequality follows because  $\ln(x+y) \leq \ln(x) + y$  for  $x \geq 1$  and  $y \geq 0$ . The last equality uses the definition of  $g(x)$  in (2). ■

**Lemma 4** *Let  $\rho^{(n)}$  be a density operator on  $n$  modes such that*

$$\text{Tr} \left\{ \Pi_{\lceil nN_S \rceil} \rho^{(n)} \right\} \geq 1 - \delta_1,$$

for some small  $\delta_1 > 0$ . Then

$$\text{Tr} \left\{ \Pi_{\lceil n(\eta N_S + \delta_2) \rceil} \mathcal{N}^{\otimes n}(\rho^{(n)}) \right\} \geq 1 - 2\sqrt{\delta_1} - \delta_1 - \exp\left\{ -(2\delta_3^2 \eta N_S)n \right\},$$

where  $\mathcal{N}^{\otimes n}$  represents  $n$  instances of the pure-loss bosonic channel with transmissivity  $\eta$ , and  $\delta_2$  and  $\delta_3$  are fixed positive constants such that  $0 < \delta_3 \leq \frac{1}{\lceil nN_S \rceil} (n\delta_2 - \eta)$  for  $n$  large enough.

**Proof.** We begin by observing that

$$\begin{aligned}
& \text{Tr} \left\{ \Pi_{\lceil n(\eta N_S + \delta_2) \rceil} \mathcal{N}^{\otimes n}(\rho^{(n)}) \right\} \\
&\geq \text{Tr} \left\{ \Pi_{\lceil n(\eta N_S + \delta_2) \rceil} \mathcal{N}^{\otimes n} \left( \Pi_{\lceil nN_S \rceil} \rho^{(n)} \Pi_{\lceil nN_S \rceil} \right) \right\} - \left\| \mathcal{N}^{\otimes n}(\rho^{(n)}) - \mathcal{N}^{\otimes n} \left( \Pi_{\lceil nN_S \rceil} \rho^{(n)} \Pi_{\lceil nN_S \rceil} \right) \right\|_1 \\
&\geq \text{Tr} \left\{ \Pi_{\lceil n(\eta N_S + \delta_2) \rceil} \mathcal{N}^{\otimes n} \left( \Pi_{\lceil nN_S \rceil} \rho^{(n)} \Pi_{\lceil nN_S \rceil} \right) \right\} - \left\| \rho^{(n)} - \Pi_{\lceil nN_S \rceil} \rho^{(n)} \Pi_{\lceil nN_S \rceil} \right\|_1 \\
&\geq \text{Tr} \left\{ \Pi_{\lceil n(\eta N_S + \delta_2) \rceil} \mathcal{N}^{\otimes n} \left( \Pi_{\lceil nN_S \rceil} \rho^{(n)} \Pi_{\lceil nN_S \rceil} \right) \right\} - 2\sqrt{\delta_1}.
\end{aligned} \tag{12}$$

The first inequality is a special case of the inequality

$$\text{Tr} \{ \Lambda \rho \} \geq \text{Tr} \{ \Lambda \sigma \} - \|\rho - \sigma\|_1, \tag{13}$$

which holds for  $\Lambda, \rho, \sigma$  such that  $0 \leq \Lambda \leq I$ ,  $0 \leq \rho, \sigma$ , and  $\text{Tr}\{\rho\}, \text{Tr}\{\sigma\} \leq 1$ . The second inequality follows from the monotonicity of trace distance under quantum operations. The third inequality is a consequence of the Gentle Operator Lemma [23, 15], which states that  $\|\rho - \sqrt{\Lambda}\rho\sqrt{\Lambda}\|_1 \leq 2\sqrt{\varepsilon}$  if  $\text{Tr}\{\Lambda\rho\} \geq 1 - \varepsilon$  for  $0 \leq \varepsilon \leq 1$ ,  $0 \leq \Lambda \leq I$ ,  $\rho \geq 0$ ,  $\text{Tr}\{\rho\} \leq 1$ . We can then expand  $\Pi_{\lceil nN_S \rceil} \rho^{(n)} \Pi_{\lceil nN_S \rceil}$  as follows in the number-state basis:

$$\sum_{\substack{a^n, b^n: \\ \sum_i a_i, \sum_i b_i \leq \lceil nN_S \rceil}} \rho_{a^n, b^n} |a^n\rangle\langle b^n|,$$

where  $a^n \equiv a_1 \cdots a_n$ ,  $b^n \equiv b_1 \cdots b_n$ , and  $\rho^{(n)} = \sum_{a^n, b^n} \rho_{a^n, b^n} |a^n\rangle\langle b^n|$ . Observe that the hypothesis of the theorem is equivalent to

$$\sum_{a^n: \sum_i a_i \leq \lceil nN_S \rceil} \rho_{a^n, a^n} \geq 1 - \delta_1.$$

The remaining term in (12) is equal to

$$\begin{aligned} & \text{Tr} \left\{ \Pi_{\lceil n(\eta N_S + \delta_2) \rceil} \mathcal{N}^{\otimes n} \left( \Pi_{\lceil nN_S \rceil} \rho^{(n)} \Pi_{\lceil nN_S \rceil} \right) \right\} \\ &= \text{Tr} \left\{ \left( \Pi_{\lceil n(\eta N_S + \delta_2) \rceil} \otimes I \right) U^{\otimes n} \left( \Pi_{\lceil nN_S \rceil} \rho^{(n)} \Pi_{\lceil nN_S \rceil} \otimes |0\rangle\langle 0|^{\otimes n} \right) \left( U^\dagger \right)^{\otimes n} \right\} \\ &= \sum_{\substack{a^n, b^n: \\ \sum_i a_i, \sum_i b_i \leq \lceil nN_S \rceil}} \rho_{a^n, b^n} \text{Tr} \left\{ \left( \Pi_{\lceil n(\eta N_S + \delta_2) \rceil} \otimes I \right) U^{\otimes n} \left( |a^n\rangle\langle b^n| \otimes |0\rangle\langle 0|^{\otimes n} \right) \left( U^\dagger \right)^{\otimes n} \right\}, \quad (14) \end{aligned}$$

where  $U$  is the unitary transformation corresponding to a beamsplitter with transmissivity  $\eta$ . Now, we recall briefly how a beamsplitter of transmissivity  $\eta$  acts on the joint state  $|n\rangle_A |0\rangle_E$ :

$$\begin{aligned} |n\rangle_A |0\rangle_E &= \frac{(\hat{a}^\dagger)^n}{\sqrt{n!}} |0\rangle_A |0\rangle_E \\ &\mapsto \frac{(\sqrt{\eta}\hat{a}^\dagger + \sqrt{1-\eta}\hat{e}^\dagger)^n}{\sqrt{n!}} |0\rangle_A |0\rangle_E \\ &= \frac{1}{\sqrt{n!}} \sum_{k=0}^n \binom{n}{k} \sqrt{\eta}^k \sqrt{1-\eta}^{n-k} (\hat{a}^\dagger)^k (\hat{e}^\dagger)^{n-k} |0\rangle_A |0\rangle_E \\ &= \sum_{k=0}^n \sqrt{\binom{n}{k}} \sqrt{\eta}^k \sqrt{1-\eta}^{n-k} |k\rangle_A |n-k\rangle_E, \end{aligned}$$

so that

$$\begin{aligned} & U^{\otimes n} \left( |a^n\rangle\langle b^n| \otimes |0\rangle\langle 0|^{\otimes n} \right) \left( U^\dagger \right)^{\otimes n} \\ &= \sum_{k_1=0}^{a_1} \cdots \sum_{k_n=0}^{a_n} \sum_{l_1=0}^{b_1} \cdots \sum_{l_n=0}^{b_n} \sqrt{\binom{a_1}{k_1} \cdots \binom{a_n}{k_n} \binom{b_1}{l_1} \cdots \binom{b_n}{l_n}} \times \\ & \sqrt{\eta}^{\sum_{i=1}^n k_i + l_i} \sqrt{1-\eta}^{\sum_{i=1}^n a_i - k_i + b_i - l_i} |k_1\rangle\langle l_1|_{A_1} \otimes \cdots \otimes |k_n\rangle\langle l_n|_{A_n} \otimes \\ & |a_1 - k_1\rangle\langle b_1 - l_1|_{E_1} \otimes \cdots \otimes |a_n - k_n\rangle\langle b_n - l_n|_{E_n} \end{aligned}$$

To evaluate the trace in (14), we can perform it with respect to the number basis, calculating

$$\sum_{\sum_i a_i, \sum_i b_i \leq [nN_S]} \rho_{a^n, b^n} \sum_{\sum_i c_i \leq [n(\eta N_S + \delta_2)]} \langle c^n |_{B^n} \langle d^n |_{E^n} U^{\otimes n} (|a^n\rangle \langle b^n|_{A^n} \otimes (|0\rangle \langle 0|_{E^n})^{\otimes n}) (U^\dagger)^{\otimes n} |c^n\rangle_{B^n} |d^n\rangle_{E^n}$$

as

$$\begin{aligned} & \sum_{\sum_i a_i, \sum_i b_i \leq [nN_S]} \rho_{a^n, b^n} \sum_{\sum_i c_i \leq [n(\eta N_S + \delta_2)]} \sum_{k_1=0}^{a_1} \cdots \sum_{k_n=0}^{a_n} \sum_{l_1=0}^{b_1} \cdots \sum_{l_n=0}^{b_n} \sqrt{\binom{a_1}{k_1} \cdots \binom{a_n}{k_n} \binom{b_1}{l_1} \cdots \binom{b_n}{l_n}} \times \\ & \sqrt{\eta^{\sum_{i=1}^n k_i + l_i}} \sqrt{1 - \eta}^{\sum_{i=1}^n a_i - k_i + b_i - l_i} \langle c_1 | k_1 \rangle \langle l_1 | c_1 \rangle_{A_1} \times \cdots \times \langle c_n | k_n \rangle \langle l_n | c_n \rangle_{A_n} \times \\ & \langle d_1 | a_1 - k_1 \rangle \langle b_1 - l_1 | d_1 \rangle_{E_1} \times \cdots \times \langle d_n | a_n - k_n \rangle \langle b_n - l_n | d_n \rangle_{E_n} \end{aligned}$$

By inspection, the only terms that survive are those for which  $c_i = k_i = l_i$  and  $d_i = a_i - k_i = b_i - l_i$  (and hence  $a_i = b_i$ ), leaving the above equal to

$$\sum_{\sum_i a_i \leq [nN_S]} \rho_{a^n, a^n} \sum_{k_1=0}^{a_1} \cdots \sum_{k_n=0}^{a_n} \mathcal{I} \left[ \sum_{i=1}^n k_i \leq [n(\eta N_S + \delta_2)] \right] \binom{a_1}{k_1} \cdots \binom{a_n}{k_n} \eta^{\sum_{i=1}^n a_i - k_i} (1 - \eta)^{\sum_{i=1}^n k_i}, \quad (15)$$

where  $\mathcal{I}[\cdot]$  is an indicator function. To find a lower bound on this expression, first we should realize that

$$\sum_{k_1=0}^{a_1} \cdots \sum_{k_n=0}^{a_n} \mathcal{I} \left[ \sum_{i=1}^n k_i \leq [n(\eta N_S + \delta_2)] \right] \binom{a_1}{k_1} \cdots \binom{a_n}{k_n} \eta^{\sum_{i=1}^n a_i - k_i} (1 - \eta)^{\sum_{i=1}^n k_i} \quad (16)$$

is equal to one whenever  $\sum_i a_i \leq [n(\eta N_S + \delta_2)]$  because, in such a case, we are guaranteed that  $\sum_{i=1}^n k_i \leq [n(\eta N_S + \delta_2)]$ . Thus, we should focus on the case in which

$$[n(\eta N_S + \delta_2)] < \sum_i a_i \leq [nN_S].$$

However, the expression in (16) is related to the probability that the average of a large number of independent Bernoulli random variables is no larger than the mean of these random variables plus a small offset. That is, after defining the i.i.d. Bernoulli random variables  $X_{i,j}$  each with parameter  $\eta$ ,<sup>2</sup> such that the binomial random variable  $K_i = \sum_j X_{i,j}$ , we find that the expression above is equal to

$$\Pr \left\{ \sum_{i,j} X_{i,j} \leq [n(\eta N_S + \delta_2)] \right\} \geq \Pr \left\{ \sum_{i,j} X_{i,j} \leq S(\eta + \delta_3) \right\}, \quad (17)$$

where  $S \equiv \sum_i a_i$  is the total number of these Bernoulli random variables (total number of photons)

<sup>2</sup>The nice physical interpretation here is that each of the  $\sum_i a_i$  i.i.d. Bernoulli random variables corresponds to a single photon that has a probability  $\eta$  of making it through the beamsplitter.

and the inequality in (17) follows from

$$\begin{aligned}
\lceil n(\eta N_S + \delta_2) \rceil &\geq n(\eta N_S + \delta_2) \\
&= n\eta N_S + n\delta_2 \\
&\geq \eta(\lceil nN_S \rceil - 1) + n\delta_2 \\
&\geq \lceil nN_S \rceil(\eta + \delta_3) \\
&\geq S(\eta + \delta_3),
\end{aligned}$$

where for large enough  $n$  we can choose a constant  $\delta_3$  such that  $0 < \delta_3 \leq \frac{1}{\lceil nN_S \rceil}(n\delta_2 - \eta)$ . We can then apply the Hoeffding concentration bound [8, 19] to conclude that (17) is larger than

$$1 - \exp\{-2\delta_3^2 S\} \geq 1 - \exp\{-2\delta_3^2 \eta n N_S\},$$

where the inequality follows from the assumption that  $S \geq \lceil n(\eta N_S + \delta_2) \rceil \geq n\eta N_S$ . Then this last expression converges to one exponentially fast with increasing  $n$ . Thus, it follows that the expression in (15) is larger than

$$(1 - \delta_1)(1 - \exp\{-(2\delta_3^2 \eta N_S)n\}) \geq 1 - \delta_1 - \exp\{-(2\delta_3^2 \eta N_S)n\}.$$

Combining the above inequality with the one in (12) gives the statement of the lemma. ■

With Lemmas 3 and 4 in hand, we can now easily prove the strong converse by an approach similar to that in (11). Indeed, let  $\rho_m$  be the codewords of any code for the pure-loss bosonic channel with transmissivity  $\eta$  such that the average codeword density operator has a large projection onto a space with total photon number less than  $\lceil nN_S \rceil$ :

$$\frac{1}{M} \sum_m \text{Tr}\{\Pi_{\lceil nN_S \rceil} \rho_m\} \geq 1 - \delta_1(n), \quad (18)$$

where  $\delta_1(n)$  is a function decreasing to zero with increasing  $n$ . Let  $\{\Lambda_m\}$  denote a decoding POVM acting on the output space of  $n$  instances of the pure-loss bosonic channel.

**Theorem 5 (Strong converse)** *The average success probability of any code satisfying (18) is bounded as follows:*

$$\frac{1}{M} \sum_m \text{Tr}\{\Lambda_m \mathcal{N}^{\otimes n}(\rho_m)\} \leq 2^{-n(R - g(\eta N_S) - \delta_2 - \delta)} + 2\sqrt{\delta_1(n) + \exp\{-(2\delta_3^2 \eta N_S)n\}} + 2\sqrt{\delta_1(n)},$$

where  $\mathcal{N}^{\otimes n}$  denotes  $n$  instances of the pure-loss bosonic channel,  $\delta$  is defined in Lemma 3 (with  $N_S$  replaced by  $\eta N_S$ ),  $\delta_1(n)$  is defined in (18),  $\delta_2$  is an arbitrarily small positive constant, and  $\delta_3$  is defined in Lemma 4. Thus, if  $R > g(\eta N_S)$ , then we can pick  $\delta_2$  and  $\delta$  small enough such that  $R > g(\eta N_S) + \delta_2 + \delta$ , and it follows that the success probability of any family of codes satisfying (18) decreases to zero in the limit of large  $n$ .

**Proof.** Consider that

$$\begin{aligned}
\frac{1}{M} \sum_m \text{Tr}\{\Lambda_m \mathcal{N}^{\otimes n}(\rho_m)\} &\leq \frac{1}{M} \sum_m \text{Tr}\{\Lambda_m \Pi_{\lceil n(\eta N_S + \delta_2) \rceil} \mathcal{N}^{\otimes n}(\rho_m) \Pi_{\lceil n(\eta N_S + \delta_2) \rceil}\} \\
&\quad + \frac{1}{M} \sum_m \|\mathcal{N}^{\otimes n}(\rho_m) - \Pi_{\lceil n(\eta N_S + \delta_2) \rceil} \mathcal{N}^{\otimes n}(\rho_m) \Pi_{\lceil n(\eta N_S + \delta_2) \rceil}\|_1 \\
&\leq \frac{1}{M} \sum_m \text{Tr}\{\Pi_{\lceil n(\eta N_S + \delta_2) \rceil} \Lambda_m \Pi_{\lceil n(\eta N_S + \delta_2) \rceil} \mathcal{N}^{\otimes n}(\rho_m)\} \\
&\quad + 2\sqrt{\delta_1(n) + \exp\{- (2\delta_3^2 \eta N_S) n\} + 2\sqrt{\delta_1(n)}}.
\end{aligned}$$

The first inequality is a consequence of (13). The second inequality follows from a variation of the Gentle Operator Lemma which holds for ensembles [23, 15]. That is, for an ensemble  $\{p_X(x), \rho_x\}$  for which  $\sum_x p_X(x) \text{Tr}\{\Lambda \rho_x\} \geq 1 - \varepsilon$  for  $0 \leq \varepsilon \leq 1$ , the following inequality holds

$$\sum_x p_X(x) \|\rho_x - \sqrt{\Lambda} \rho_x \sqrt{\Lambda}\|_1 \leq 2\sqrt{\varepsilon}.$$

We apply this along with Lemma 4 and the assumption in (18) to arrive at the second inequality. Focusing on the first term in the last expression, we find the following upper bound by a method similar to that in (11):

$$\begin{aligned}
\frac{1}{M} \sum_m \text{Tr}\{\Pi_{\lceil n(\eta N_S + \delta_2) \rceil} \Lambda_m \Pi_{\lceil n(\eta N_S + \delta_2) \rceil} \mathcal{N}^{\otimes n}(\rho_m)\} &\leq \frac{1}{M} \sum_m \text{Tr}\{\Pi_{\lceil n(\eta N_S + \delta_2) \rceil} \Lambda_m \Pi_{\lceil n(\eta N_S + \delta_2) \rceil}\} \\
&= M^{-1} \text{Tr}\{\Pi_{\lceil n(\eta N_S + \delta_2) \rceil}\} \\
&\leq 2^{-n(R - g(\eta N_S) - \delta_2 - \delta)}.
\end{aligned}$$

Here, the first inequality follows because  $\|\mathcal{N}^{\otimes n}(\rho_m)\|_\infty \leq 1$ . The following equality holds because  $\sum_m \Lambda_m = I$ , and the last inequality follows from Lemma 3. Putting everything together, we arrive at the bound in the statement of the theorem. ■

### 3 Capacity-achieving codes meeting the maximum photon number constraint

The demand in (18) seems like it is somewhat stringent. That is, do there actually exist capacity-achieving codes for the pure-loss bosonic channel that meet this constraint while being a reliable scheme for communication?

This final section argues that there exist codes for the pure-loss bosonic channel that meet the following two constraints:

1. The maximum photon number constraint in (18) is satisfied with  $\delta_1(n)$  a function exponentially decreasing to zero in  $n$ .
2. The error probability is less than an arbitrarily small constant for sufficiently large  $n$ .

So our development here is a refinement of the coding theorem given in Ref. [10, 6].

To be precise, for a given code, we would like for the following two conditions to hold

$$\frac{1}{M} \sum_m \text{Tr}\{\Pi_{\lceil nN_S \rceil} \rho_m\} \geq 1 - \delta_1(n), \quad (E_1)$$

$$\frac{1}{M} \sum_m \text{Tr}\{\Lambda_m \mathcal{N}^{\otimes n}(\rho_m)\} \geq 1 - \varepsilon, \quad (E_2)$$

for an arbitrarily small positive number  $\varepsilon$  and sufficiently large  $n$ . In order to prove the existence of codes satisfying these two constraints, we pick codewords as  $n$ -fold tensor products of coherent states, randomly chosen according to a complex, isotropic Gaussian distribution with variance  $N_S - \delta$  where  $\delta$  is an arbitrarily small positive constant (the same approach as given in Ref. [6], following [10]). We can then analyze the probability that the constraints  $E_1$  and  $E_2$  above are not met for a randomly chosen code:

$$\begin{aligned} & \Pr\{(E_1 \cap E_2)^c\} \\ & \leq \Pr\{E_1^c\} + \Pr\{E_2^c\} \\ & = \Pr\left\{1 - \frac{1}{M} \sum_m \text{Tr}\{\Pi_{\lceil nN_S \rceil} \rho_m\} > \delta_1(n)\right\} + \Pr\left\{1 - \frac{1}{M} \sum_m \text{Tr}\{\Lambda_m \mathcal{N}^{\otimes n}(\rho_m)\} > \varepsilon\right\} \\ & \leq \frac{1}{\delta_1(n)} \mathbb{E}\left\{1 - \frac{1}{M} \sum_m \text{Tr}\{\Pi_{\lceil nN_S \rceil} \rho_m\}\right\} + \frac{1}{\varepsilon} \mathbb{E}\left\{1 - \frac{1}{M} \sum_m \text{Tr}\{\Lambda_m \mathcal{N}^{\otimes n}(\rho_m)\}\right\}. \end{aligned} \quad (19)$$

The first inequality is a consequence of the union bound and the second follows from the Markov inequality. We know that for every constant  $\varepsilon^2$ , the expectation of the average error probability of a randomly chosen code is less than  $\varepsilon^2$  as long as the rate is no larger than  $g(\eta N_S)$  and  $n$  is sufficiently large [7, 21]. This means that

$$\mathbb{E}\left\{1 - \frac{1}{M} \sum_m \text{Tr}\{\Lambda_m \mathcal{N}^{\otimes n}(\rho_m)\}\right\} \leq \varepsilon^2. \quad (20)$$

We can analyze the other term in (19) as follows:

$$\mathbb{E}\left\{1 - \frac{1}{M} \sum_m \text{Tr}\{\Pi_{\lceil nN_S \rceil} \rho_m\}\right\} = 1 - \frac{1}{M} \sum_m \text{Tr}\{\Pi_{\lceil nN_S \rceil} \mathbb{E}\{\rho_m\}\}. \quad (21)$$

The expected density operator  $\mathbb{E}\{\rho_m\}$  is a thermal state of mean photon number  $N'_S \equiv N_S - \delta$  since we are choosing codewords as coherent states according to a complex Gaussian distribution with variance  $N'_S$ :

$$\mathbb{E}\{\rho_m\} = \theta(N'_S)^{\otimes n},$$

where

$$\theta(N'_S) \equiv \int d^2\alpha \frac{1}{\pi N'_S} \exp\{-|\alpha|^2/N'_S\} |\alpha\rangle\langle\alpha| = \frac{1}{N'_S + 1} \sum_{l=0}^{\infty} \left(\frac{N'_S}{N'_S + 1}\right)^l |l\rangle\langle l|.$$

Observe that the distribution of a thermal state with respect to the photon number basis is a geometric distribution with mean  $N'_S$ . Then (21) reduces to

$$1 - \frac{1}{M} \sum_m \text{Tr}\{\Pi_{\lceil nN_S \rceil} \mathbb{E}\{\rho_m\}\} = 1 - \text{Tr}\{\Pi_{\lceil nN_S \rceil} \theta(N'_S)^{\otimes n}\}.$$

Finally, the expression  $1 - \text{Tr}\left\{\Pi_{\lceil nN_S \rceil} \theta(N'_S)^{\otimes n}\right\}$  is equal to the probability that the average of a large number of independent geometric random variables deviate from their mean by more than  $\delta$ , which is exponentially decreasing in  $n$  as

$$1 - \text{Tr}\left\{\Pi_{\lceil nN_S \rceil} \theta(N'_S)^{\otimes n}\right\} \leq [C(\delta, N'_S)]^n,$$

where  $C(\delta, N'_S)$  is a constant strictly less than one (see the appendix for an explicit proof). Thus, we can choose  $\delta_1(n) = [C(\delta, N'_S)]^{(n/2)}$  (for example) and combine the above bound with (19) and (20) to arrive at the following upper bound

$$\Pr\{(E_1 \cap E_2)^c\} \leq [C(\delta, N'_S)]^{(n/2)} + \varepsilon.$$

So, for  $n$  large enough and since  $\delta$  can be an arbitrarily small positive constant, this proves the existence of a code that satisfies the two constraints given at the beginning of this section with a rate equal to  $g(\eta N_S)$ . (In fact, the overwhelming fraction of codes selected in this way satisfy these constraints for sufficiently large  $n$ , while operating at the aforementioned rate.)

## 4 Conclusion

This paper has broadened the understanding of the classical capacity of the pure-loss bosonic channel by determining conditions under which a strong converse theorem can and cannot hold. First, we proved that there is a rate-error trade-off whenever there is only a mean photon number constraint on the codewords input to the channels, so that a strong converse theorem does not hold under such a constraint. One can even use pure-state codewords to achieve this trade-off, which is an important distinction between the classical and quantum theories of information for continuous variables. Next, we proved that a strong converse theorem holds under a particular maximum photon number constraint. Our proof was a simulation-based argument: we first showed that it is possible to faithfully simulate the action of  $n$  instances of the pure-loss bosonic channel with transmissivity parameter  $\eta \in [0, 1]$ , at a rate of  $g(\eta N_S)$  whenever it is guaranteed that the input to the channel has nearly all of its shadow on a subspace having no more than  $nN_S$  photons. By concatenating a channel code with the simulation code, we showed that it would be possible to strongly violate the Holevo bound if one could send classical data over the pure-loss bosonic channel at a rate  $R > g(\eta N_S)$ . Finally, we refined the coding theorem of [10] and [6], to show that there exist coherent-state codes that achieve the classical capacity while satisfying our maximum photon number constraint.

**Acknowledgements.** MMW is grateful to the quantum information theory group at the Universitat Autònoma de Barcelona for hosting him for a research visit during April-May 2013, during which some of the work for this research was completed. He is also grateful to the quantum information groups at MIT and Raytheon BBN Technologies for hosting him as a visitor during June-August 2013. AW's work is supported by the European Commission (STREP "QCS"), the European Research Council (Advanced Grant "IRQUAT") and the Philip Leverhulme Trust; furthermore by the Spanish MINECO, project FIS2008-01236, with the support of FEDER funds.



## Appendix

Here we detail an explicit proof of the following bound:

$$\Pr\left\{\frac{1}{n}\sum_{i=1}^n Z_i \geq \mu + \delta\right\} \leq [C(\delta, p)]^n,$$

where  $C(\delta, p)$  is a constant strictly less than one,  $\delta > 0$ , and each  $Z_i$  is an independent geometric random variable with mean  $\mu = \frac{p}{1-p}$  and probability mass function  $\Pr\{Z_i = k\} = p^k(1-p)$  for  $k \in \{0, 1, 2, \dots\}$ . For this purpose, we use the well known ‘‘Bernstein trick’’ (exponential moment method) [19], according to which

$$\Pr\left\{\frac{1}{n}\sum_{i=1}^n Z_i \geq \mu + \delta\right\} \leq \inf_{t>0} \left(\frac{\mathbb{E}\{\exp\{tZ\}\}}{\exp\{t(\mu + \delta)\}}\right)^n.$$

So we just need to find a  $t$  for which  $\mathbb{E}\{\exp\{tZ\}\} < \exp\{t(\mu + \delta)\}$ . The moment generating function  $\mathbb{E}\{\exp\{tZ\}\}$  of a geometric random variable is

$$\mathbb{E}\{\exp\{tZ\}\} = \frac{1-p}{1-pe^t},$$

where we require  $t$  to be chosen so that  $pe^t < 1$  (i.e.,  $t < -\ln p$ ) in order to ensure convergence of  $\mathbb{E}\{\exp\{tZ\}\}$ . Setting  $x = e^t$ , our problem is to find a value of  $x$  for which

$$\frac{1-p}{1-px} < x^{p/(1-p)+\delta}.$$

At  $x = 1$ , the left hand side is equal to the right hand side, and taking derivatives, we find that

$$\left.\frac{\partial}{\partial x} \left[\frac{1-p}{1-px}\right]\right|_{x=1} = \frac{p}{1-p},$$

while

$$\left.\frac{\partial}{\partial x} \left[x^{p/(1-p)+\delta}\right]\right|_{x=1} = \frac{p}{1-p} + \delta$$

It holds that

$$\frac{p}{1-p} < \frac{p}{1-p} + \delta,$$

so we can conclude that  $x^{p/(1-p)+\delta}$  is growing faster than  $\frac{1-p}{1-px}$  in a neighborhood of 1, so that there exists a value of  $x < 1/p$  (and thus  $t$ ) such that

$$\frac{\mathbb{E}\{\exp\{tZ\}\}}{\exp\{t(\mu + \delta)\}} < 1.$$

We then set  $C(\delta, p) = \mathbb{E}\{\exp\{tZ\}\} / \exp\{t(\mu + \delta)\}$  for this value of  $t$ .

## References

- [1] Charles H. Bennett, Igor Devetak, Aram W. Harrow, Peter W. Shor, and Andreas Winter. Quantum reverse Shannon theorem. 2009. arXiv:0912.5537.
- [2] Mario Berta, Fernando Brandão, Matthias Christandl, and Stephanie Wehner. Entanglement Cost of Quantum Channels. In *Proceedings of the 2012 International Symposium on Information Theory*, pages 900–904, Cambridge, MA, USA, July 2012.
- [3] Mario Berta, Fernando Brandão, Matthias Christandl, and Stephanie Wehner. Entanglement Cost of Quantum Channels. *IEEE Transactions on Information Theory*, 59(10):6779–6795, October 2013. arXiv:1108.5357.
- [4] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory*. Wiley-Interscience, 2nd edition, 2006.
- [5] Igor Devetak, Aram W. Harrow, and Andreas Winter. A Resource Framework for Quantum Shannon Theory. *IEEE Transactions on Information Theory*, 54(10):4587–4618, October 2008. arXiv:quant-ph/0512015.
- [6] Vittorio Giovannetti, Saikat Guha, Seth Lloyd, Lorenzo Maccone, Jeffrey H. Shapiro, and Horace P. Yuen. Classical Capacity of the Lossy Bosonic Channel: The Exact Solution. *Physical Review Letters*, 92(2):027902, January 2004. arXiv:quant-ph/0308012.
- [7] Paul Hausladen, Richard Jozsa, Benjamin Schumacher, Michael Westmoreland, and William K. Wootters. Classical information capacity of a quantum channel. *Physical Review A*, 54(3):1869–1876, September 1996.
- [8] Wassily Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58(301):13–30, March 1963.
- [9] Alexander S. Holevo. The Capacity of the Quantum Channel with General Signal States. *IEEE Transactions on Information Theory*, 44:269–273, 1998.
- [10] Alexander S. Holevo and Reinhard F. Werner. Evaluating capacities of Bosonic Gaussian channels. *Physical Review A*, 63:032312, February 2001.
- [11] Robert Koenig and Stephanie Wehner. A Strong Converse for Classical Channel Coding Using Entangled Inputs. *Physical Review Letters*, 103:070504, August 2009. arXiv:0903.2838.
- [12] Robert Koenig, Stephanie Wehner, and Jürg Wullschlegler. Unconditional Security From Noisy Quantum Storage. *IEEE Transactions on Information Theory*, 58(3):1962–1984, March 2012. arXiv:0906.1030.
- [13] Ashwin Nayak. Optimal lower bounds for quantum automata and random access codes. In *Proceedings of the 40th Annual Symposium on Foundations of Computer Science*, pages 369–376, New York City, NY, USA, October 1999. arXiv:quant-ph/9904093.
- [14] Tomohiro Ogawa and Hiroshi Nagaoka. Strong Converse to the Quantum Channel Coding Theorem. *IEEE Transactions on Information Theory*, 45(7):2486–2489, November 1999. arXiv:quant-ph/9808063.

- [15] Tomohiro Ogawa and Hiroshi Nagaoka. Making Good Codes for Classical-Quantum Channel Coding via Quantum Hypothesis Testing. *IEEE Transactions on Information Theory*, 53(6):2261–2266, June 2007.
- [16] Yury Polyanskiy. *Channel coding: Non-asymptotic fundamental limits*. PhD thesis, Princeton University, November 2010.
- [17] Benjamin Schumacher and Michael D. Westmoreland. Sending classical information via noisy quantum channels. *Physical Review A*, 56(1):131–138, July 1997.
- [18] Jeffrey H. Shapiro. The quantum theory of optical communications. *IEEE Journal of Selected Topics in Quantum Electronics*, 15(6):1547–1569, 2009.
- [19] Terence Tao. *Topics in Random Matrix Theory*, volume 132 of *Graduate Studies in Mathematics*. American Mathematical Society, 2012. See also <http://terrytao.wordpress.com/2010/01/03/254a-notes-1-concentration-of-measure/>.
- [20] Christian Weedbrook, Stefano Pirandola, Raúl García-Patrón, Nicolas J. Cerf, Timothy C. Ralph, Jeffrey H. Shapiro, and Seth Lloyd. Gaussian quantum information. *Reviews of Modern Physics*, 84:621–669, May 2012. arXiv:1110.3234.
- [21] Mark M. Wilde, Saikat Guha, Si-Hui Tan, and Seth Lloyd. Explicit capacity-achieving receivers for optical communication and quantum reading. In *Proceedings of the 2012 International Symposium on Information Theory*, pages 551–555, Boston, Massachusetts, USA, July 2012. arXiv:1202.0518.
- [22] Mark M. Wilde, Andreas Winter, and Dong Yang. Strong converse for the classical capacity of entanglement-breaking and Hadamard channels. June 2013. arXiv:1306.1586.
- [23] Andreas Winter. Coding Theorem and Strong Converse for Quantum Channels. *IEEE Transactions on Information Theory*, 45(7):2481–2485, 1999.
- [24] Andreas Winter. *Coding Theorems of Quantum Information Theory*. PhD thesis, Universität Bielefeld, July 1999. arXiv:quant-ph/9907077.
- [25] Andreas Winter. Compression of sources of probability distributions and density operators. 2002. arXiv:quant-ph/0208131.
- [26] Horace P. Yuen and Masanao Ozawa. Ultimate Information Carrying Limit of Quantum Systems. *Physical Review Letters*, 70:363–366, January 1993.