Quantum-optimal information encoding using noisy passive linear optics

Andrew Tanggara[1,2], Ranjith Nair[2], Syed Assad[3,4], Varun Narasimhachar[5,2], Spyros Tserkis[3], Jayne Thompson[5], Ping Koy Lam[3,4], and Mile Gu[2,1,6]

[1]Centre for Quantum Technologies, National University of Singapore, 3 Science Drive 2, Singapore 117543.

[2]Nanyang Quantum Hub, School of Physical and Mathematical Sciences, Nanyang Technological University, Singapore 639673.

[3]Centre for Quantum Computation and Communication Technology, Department of Quantum Science, Research School of Physics and Engineering, Australian National University, Canberra ACT, Australia 2601.

[4]A*STAR Quantum Innovation Centre (Q.InC), Institute of Materials Research and Engineering (IMRE), Agency for Science Technology and Research (A*STAR), 2 Fusionopolis Way, 08-03 Innovis, Singapore 138634.

[5]A*STAR Quantum Innovation Centre (Q.InC), Institute of High Performance Computing (IHPC), Agency for Science, Technology and Research (A*STAR), Singapore.

[6]CNRS-UNS-NUS-NTU International Joint Research Unit, UMI 3654, Singapore 117543.

The amount of information that a noisy channel can transmit has been one of the primary subjects of interest in information theory. In this work we consider a practically-motivated family of optical quantum channels that can be implemented without an external energy source. We optimize the Holevo information over procedures that encode information in attenuations and phase-shifts applied by these channels on a resource state of finite energy. It is shown that for any given input state and environment temperature, the maximum Holevo information can be achieved by an encoding procedure that uniformly distributes the channel's phase-shift parameter. Moreover for large families of input states, any maximizing encoding scheme has a finite number of channel attenuation values, simplifying the codewords to a finite number of rings around the origin in the output phase space. The above results and numerical evidence suggests that this property holds for all resource states. Our results are directly applicable to the quantum reading of an optical memory in the presence of environmental thermal noise.

Andrew Tanggara: andrew.tanggara@gmail.com

Mile Gu: mgu@quantumcomplexity.org

1 Introduction

For a classical channel between a sender (Alice) and receiver (Bob), the rate at which information can be noiselessly transmitted from Alice and Bob for a given probability distribution of the input symbols equals the Shannon mutual information between the input and output [1]. Maximizing the mutual information over all input probability distributions gives the capacity of the channel, which can be achieved using encoding and decoding over long codes. Quantum-mechanically, the analog of an input distribution is an ensemble of quantum states indexed by Alice's input alphabet, and the Holevo information of the ensemble [2] replaces the mutual information in the classical case. If Alice encodes her messages in suitable sequences of states and Bob can make measurement on all received output states at once, communication at a rate equals to the Holevo information is possible with arbitrarily small error probability [3, 4]. Optimizing over the input probability distribution thus gives the ultimate capacity of encoding information in the given set of quantum states.

Optical quantum channels are an important class of quantum channels for which the Holevo information has been very well studied in the context of different tasks, such as communication [5], optical memory reading [6, 7, 8, 9, 10, 11], algorithmic cooling [12], as well as in pure-loss quantum optical channels [13]. Given its many

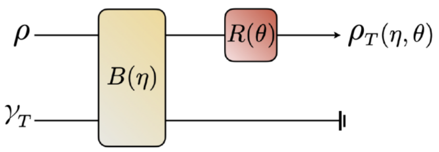

Figure 1: Thermal channel circuit representation. Input "resource" state ρ is mixed with an "environment" in a thermal state γ_T by a beamsplitter B_η of transmittance η and then undergoes a θ phase-rotation operation R_θ, giving an output codeword state $\rho_T(\eta, \theta)$.

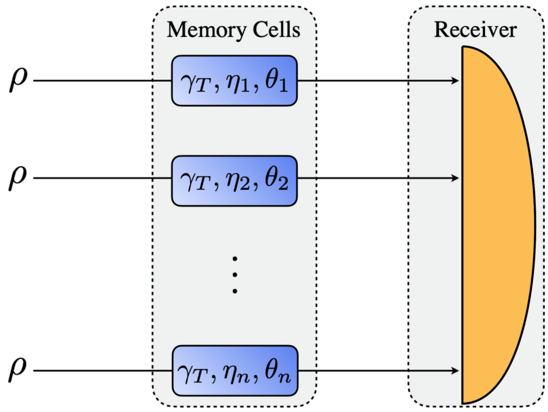

Figure 2: Thermal channel memory cell reading. Each memory cell is a thermal channel with temperature T parameterized by attenuation η_j and phase θ_j, transforming an input resource state ρ to an output state to be measured by the receiver.

applications, we focus on optical quantum channels that separates out the sources of energy, thus rendering energy as a resource in performing the task at hand. This notion of energy as resource has been studied in [14] for thermodynamic tasks.

In what follows, we consider encoding information by optimally applying passive linear-optical operations on a given finite-energy resource state ρ. Such operations, which we call *thermal channels* (illustrated in Fig 1) are optical quantum channels that can be constructed simply by mixing the given input state with the environment at a certain temperature, followed by a phase-shift operation. The average energy of every state of such output ensembles is bounded above. Such peak energy constraints have physical meaning as the greatest amount of energy that can be tolerated by a given device or channel. In addition, technological limitations may impose such constraints – for example, it is very difficult to generate Fock states of large occupation number. Even for coherent-state sources, such constraints can be severe in some applications. For example, a satellite-based laser communication system may be highly constrained in its energy output by payload limitations and energy scarcity in space. Note that a peak energy constraint on an ensemble is more restrictive than an overall average energy constraint, which has been the subject of earlier work. Indeed, the capacity-achieving ensembles for communication on a large class of bosonic Gaussian channels are circular Gaussian distributions of coherent states in phase space [5, 13, 15], which are are clearly not peak-constrained.

In addition, we also note that by adopting the thermal channel model we obtain a finite-temperature generalization of existing results that assumed a zero-temperature (i.e. vacuum) environment. These include past studies of tasks such as the quantum reading of optical memory [10, 11] (illustrated in Fig. 2) and communication over the pure-loss channel [13].

In this work, we investigate the information capacity of such a thermal encoding protocol as quantified by its Holevo information. First we show that given an arbitrary input state and an arbitrary environment temperature, the encoding that maximizes the Holevo information is characterized by an independent distribution of the attenuation coefficients and the phase shift values, with the latter being uniformly distributed – we call such encodings circularly symmetric encodings. We then derive necessary and sufficient information-theoretic conditions that characterize the optimal distribution of the attenuation coefficient for such circularly symmetric encodings. For the case of a coherent-state resource at zero temperature and a thermal state at any temperature, we show analytically that the optimal encoding involves only a finite number of attenuation values. In addition, numerical results based on the aforementioned information-theoretic optimality conditions also show that this property holds in various cases where the output states are displaced thermal states. Based on this combination of analytical and numerical evidence, we conjecture that circular symmetric encodings with a finite number of attenuation values maximize the Holevo information for any resource state.

This paper is organized as follows: In section 2, we lay down the formal definition of a thermal channel and then present the general properties of an optimal encoding. In Section 3, we focus on the case of a zero-temperature environment and obtain optimal encodings when the resource is a coherent state or a displaced thermal state. In Section 4, we discuss a nonzero-temperature environment with a coherent-state or thermal-state resource. We close with Section 5, which briefly discusses the implications of our results. The technical proofs are relegated to the appendices.

2 Thermal encodings: definitions and optimization

2.1 Thermal operations

In this section, we first define *thermal channels* whose output states constitute the potential codewords for the channel. Consider a given input *resource state* with energy $E < \infty$ represented by density operator ρ in Hilbert space $\mathcal{H}_A$ and thermal state with temperature T as the *environment state* represented by density operator $\gamma_T = \frac{e^{-\beta H}}{\mathrm{Tr}(e^{-\beta H})}$ in Hilbert space $\mathcal{H}_E$ for a Hamiltonian $H = \hat{a}_E^\dagger \hat{a}_E$ (where $\hat{a}_E^\dagger$ and $\hat{a}_E$ are the creation and annihilation operators on $\mathcal{H}_E$, respectively) and inverse temperature $\beta = \frac{1}{kT}$. Where it is convenient to express a thermal state in terms of its mean photon number $N = \langle \hat{a}_E^\dagger \hat{a}_E \rangle$, we write $\rho_{\mathrm{th}}(N) = \sum_{n=0}^{\infty} \frac{N^n}{(N+1)^{n+1}} |n\rangle\langle n|$. We consider only phase-rotation operations and mixing operations with a thermal state. These operations do not require an external energy source and have been studied in [14, 12]. The thermal channel consists of exactly each of these two operations on the resource state: an operation mixing ρ and γ_T followed by a noiseless phase-rotation operation.

Definition 1 (Thermal channel)**.** A *thermal channel* with attenuation $\eta \in [0, 1]$, phase rotation $\theta \in [-\pi, \pi)$, and environment temperature $T \geq 0$ is a quantum channel $\mathcal{T}_{\eta,\theta}^{(T)}$ whose action on a given input state $\rho \in \mathcal{H}_A$ is defined as

$$\mathcal{T}_{\eta,\theta}^{(T)}(\rho) = \mathrm{Tr}_E \left(R_\theta \, B_\eta (\rho \otimes \gamma_T) B_\eta^\dagger \, R_\theta^\dagger \right)$$

where $B_\eta = e^{\varphi_\eta (\hat{a}_A^\dagger \hat{a}_E - \hat{a}_E^\dagger \hat{a}_A)}$ is a beamsplitter operation with transmissivity $\eta = \cos^2 \varphi_\eta$ and $R_\theta = e^{-i\theta \hat{a}_A^\dagger \hat{a}_A}$ is a phase-rotation of θ on $\mathcal{H}_A$. For a circuit representation, see Fig. 1.

Given thermal channel with temperature T, attenuation η, and phase θ, its output is called a codeword and is denoted by $\rho_T(\eta, \theta) = \mathcal{T}_{\eta,\theta}^{(T)}(\rho)$ or simply $\rho(\eta, \theta)$ when temperature T is clear from the context. A *thermal encoding* is defined by a cumulative distribution function $F(\eta, \theta)$ over attenuation η and phase θ of the thermal channel which generates the ensemble of codewords $\{\rho(\eta, \theta), dF(\eta, \theta)\}_{\eta,\theta}$. For a thermal encoding F with environment temperature T acting on resource state ρ, the amount of classical information that can be encoded onto the output quantum state is given by the Holevo information [2, 3, 4]:

$$\chi_{te}[F] = S(\rho_{\mathrm{ave}}) - \int dF(\eta, \theta) S(\rho_T(\eta, \theta)) \, , \quad (1)$$

where S is the von Neumann entropy and $\rho_{\mathrm{ave}} = \int dF(\eta, \theta) \rho_T(\eta, \theta)$ is the averaged state of the ensemble. All information quantities throughout this paper are in bits, hence all log functions are base 2. Given a resource state ρ and temperature T, our task is to find the thermal encoding F^* that maximises the Holevo information $\chi_{te}[F]$. We note that χ_{te} is an adaptation of the thermal information capacity proposed in [16] which quantifies the optimal Holevo information over distribution of passive thermal operations.

2.2 Characterizations of optimal thermal encodings

In this section, we state the properties of thermal encodings F that maximize information capacity of the thermal channel $\chi_{te}[F]$ given resource state ρ and environment temperature T. Interestingly, it turns out that these properties share many features with the optimal encodings for classical additive white Gaussian noise (AWGN) channels where the signals satisfy a *peak* energy constraint E (see discussions in Section 5). Before we proceed to state these properties, we define circularly symmetric encodings which play an important role in the optimality conditions.

Definition 2. A thermal encoding F' such that $F'(\eta, \theta) = F(\eta) \frac{\theta + \pi}{2\pi}$ is a *circularly symmetric encoding.* Given a resource state ρ, for each η, we call $\rho(\eta) = \int \frac{d\theta}{2\pi} \rho(\eta, \theta)$ a *ring state.*

A ring state $\rho(\eta)$ is a mixture of codewords $\{\rho(\eta, \theta)\}_\theta$ in which the phase θ is distributed uniformly around the origin of the phase space, thus forming a ring. Hence we can write the average output state as a mixture of ring states

$$\rho_{\text{ave}} = \int dF(\eta)\rho(\eta)\ .$$

Both $\rho(\eta)$ and ρ_{ave} are diagonal in the Fock basis with their n-th diagonal entry being $P_n[\rho(\eta)] = \langle n|\rho(\eta)|n\rangle$ and

$$P_n[F] = \langle n|\rho_{\text{ave}}|n\rangle = \int dF(\eta)\, P_n[\rho(\eta)]\ , \quad (2)$$

respectively. For a proof, see Lemma 1. This family of encodings is central to the first optimality condition.

Proposition 1. *Given a resource state ρ and channel temperature T, for any encoding F, there exists a circularly symmetric encoding F' such that $\chi_{te}[F'] \geq \chi_{te}[F]$.*

This implies that for any optimal encoding $F^* = \arg\max_F \chi_{te}[F]$ there exists a circularly symmetric encoding F' that attains that same optimal capacity (i.e. $\chi_{te}[F'] = \chi_{te}[F^*]$). This property is mentioned and shown in a plot by Guha et al. [9] comparing capacity of different quantum reading settings, but was missing a formal proof. We complete this missing piece with a proof in Appendix A.1. Thanks to Proposition 1, we can focus our search for optimal encodings to circularly symmetric encodings, which are completely characterized by their cumulative distribution function (cdf) over the attenuation coefficent η alone. To avoid notational complexity, we will hereafter use $F(\eta)$ to also denote the cdf of η alone when dealing with circulary symmetric encodings.

We now discuss the second property of the optimal thermal encoding, a necessary and sufficient condition for an optimal encoding F^*. We first note that the Holevo information of a circularly symmetric encoding F can be written as an average of a function that depends on η that indicates information content of the ring at η by expressing χ_{te} in terms of relative entropy $S(\cdot||\cdot)$, i.e.

$$\begin{aligned}\chi_{te}[F] &= \int dF(\eta,\theta)S(\rho(\eta,\theta)||\rho_{\text{ave}}) \\ &= \int dF(\eta,\theta)\Big(- \text{Tr}\left[\rho(\eta,\theta)\log\rho_{\text{ave}}\right] + \\ &\quad \text{Tr}\left[\rho(\eta,\theta)\log\rho(\eta,\theta)\right]\Big) \\ &= \int dF(\eta)\Big(- \text{Tr}\left[\rho(\eta)\log\rho_{\text{ave}}\right] - S(\rho(\eta,0))\Big)\ ,\end{aligned} \quad (3)$$

where we have used Fubini's theorem (see, e.g. Chapter 2.3 of [17]) to swap the sum and integral in $S(\rho_{\text{ave}})$ and used the fact that entropy $S(\rho(\eta,\theta))$ is independent of the phase θ, allowing us to put $\rho(\eta,0)$ in place of $\rho(\eta,\theta)$. Now, define the *marginal information density* of circularly symmetric encoding F at η as

$$i[\eta,F] = -\sum_n P_n[\rho(\eta)]\log P_n[F] - S(\rho(\eta,0)) \quad (4)$$

$$\text{so,}\quad \chi_{te}[F] = \int dF(\eta)i[\eta,F]\ . \quad (5)$$

Since the relative entropy $S(\rho(\eta,\theta)||\rho_{\text{ave}})$ is always non-negative, the marginal information density $i[\eta,F]$ is also non-negative, indicating that it is an information measure of ring $\rho(\eta)$ given circularly symmetric encoding F. We note that the marginal information density also plays an integral part in the results for the capacity of the classical amplitude-constrained additive white-Gaussian-noise (AWGN) channel by Smith [18] and the direct-detection photon channel by Shamai [19], particularly on the properties of the optimal input alphabet distribution. The set of attenuation-coefficients η where F is increasing is of particular importance for this second property. This set has been called the *points of increase (POI)* (e.g. [20, 18, 21]) and *support* (e.g. [22, 23]) of the encoding in the literature, depending on whether the encoding is defined as a cumulative distribution function or a probability distribution over the attenuation-coefficients.

Definition 3. Given a probability distribution on $[0,1]$ and its cdf $G(\eta) := \Pr[x \leq \eta]$. The *support* of the distribution is defined as the smallest closed set $\mathcal{S} \subseteq [0,1]$ such that $\Pr[\mathcal{S}] = 1$. A point $\eta \in \mathcal{S}$ is also called a *point of increase* of the cdf G. We will call $\mathcal{I}_G$ the set of points of increase (equivalently, the support set of the probability distribution) the *POI* of G.

Now we are ready to state our second property.

Proposition 2. *For any given resource state ρ, a circularly symmetric encoding $F^*(\eta)$ is uniquely optimal if and only if*

$$i[\eta,F^*] \leq \chi_{te}[F^*] \quad \text{for all } \eta \in [0,1] \quad (6a)$$

$$i[\eta,F^*] = \chi_{te}[F^*] \quad \text{if and only if } \eta \in \mathcal{I}_{F^*} \quad (6b)$$

for POI $\mathcal{I}_{F^*}$.

The complete proof is provided in Appendix A.3. This property establishes a necessary and sufficient condition for a circularly symmetric encoding F^* that achieves the optimal Holevo information $\sup_F \chi_{te}[F]$. In particular, the first statement can be understood as the Holevo information $\chi_{te}[F^*]$ being an upper bound to the marginal information $i[\eta, F^*]$ regardless the value of η. Whereas, the second statement says that for any point η for which $F^*(\eta)$ is increasing (either in a continuous or discontinuous manner), its marginal information $i[\eta, F^*]$ is equal to the Holevo information. Because an encoding is a cumulative distribution function it may be more intuitive for one to interpret this as a characterization of the support of probability distribution dF^* induced by optimal encoding F^*.

Supported by analytical and numerical results we present in later sections for some large classes of Gaussian resource states, we conjecture that the Holevo-optimal distribution F^* over attenuation η for *any* resource state ρ has a finite support by showing that its corresponding POI $\mathcal{I}_{F^*}$ must have a finite cardinality.

Conjecture 1. For any resource state ρ and temperature $T \geq 0$, the set $\mathcal{I}_{F^*}$ of points of increase of the attenuation-coefficient distribution $F^*(\eta)$ associated with the optimal circularly-symmetric encoding has finite cardinality.

Conjecture 1 states that given an optimal circularly symmetric encoding, the average output state of the thermal channel is a finite mixture of ring states, which can be nicely visualized as discrete rings in its phase-space representation as presented in Fig. 3. We show this property rigorously for coherent state resource in zero-temperature and thermal state resource in any temperature as stated in Proposition 3 and Proposition 4, respectively, and discussed in the corresponding sections they are in. Numerical evidence for more general cases where the channel output is a displaced thermal state are also given in the following sections.

3 Zero temperature environment

In this section we discuss finding the optimal encoding when the environment has temperature $T = 0$, which corresponds to the environment state being the vacuum state $\gamma_0 = |0\rangle\langle 0|$. Given a resource state ρ with energy $E_{\max}$, the codeword $\rho(\eta, \theta)$ has average energy $\eta E_{\max}$ as a consequence of γ_0 environment having a 0 energy. Therefore, the average state ρ_{ave} has energy at most $E_{\max}$. As such, the Holevo information of any encoding and any resource state is bounded above by the quantity [5, 15]

$$g(E_{\max}) := (1 + E_{\max}) \log(1 + E_{\max}) - E_{\max} \log E_{\max} . \tag{7}$$

We use this universal capacity upper-bound for $T = 0$ environment to benchmark the encoding capacity of our optimal schemes – see Appendix D for more details.

3.1 Coherent state resource

Given a coherent state $\rho = |\alpha_{\max}\rangle\langle\alpha_{\max}|$ as resource (hence $E_{\max} = |\alpha_{\max}|^2$), the possible codewords $\rho(\eta, \theta)$ are coherent states $|\alpha\rangle\langle\alpha|$ with amplitude $|\alpha| = \sqrt{\eta}|\alpha_{\max}| \leq |\alpha_{\max}|$ for η distributed by the choice of encoding F. Since the entropy of every codeword is zero (the codewords are pure states), the Holevo information equals the entropy of the averaged state, namely, $\chi_{tc}[F] = S(\rho_{\text{ave}})$. The problem of finding the optimal encoding is then equivalent to finding the encoding F that maximises the entropy of the mixture distribution over $n \in \mathbb{N}$

$$\int dF(\eta)\, \mathcal{P}_n\left(\eta|\alpha_{\max}|^2\right) , \tag{8}$$

where $\mathcal{P}_n(\lambda) = e^{-\lambda}\frac{\lambda^n}{n!}$ is a Poisson probability distribution over outcome n with mean λ.

The optimal circularly-symmetric encoding finiteness property stated in Conjecture 1 holds for coherent states at $T = 0$, as shown in Appendix B.

Proposition 3. *For any coherent-state resource $\rho = |\alpha\rangle\langle\alpha|$ and $T = 0$, the attenuation-coefficients η of the optimal circularly-symmetric encoding takes a finite number of values in $[0, 1]$.*

We plot this finite optimal distribution for resource state energy $|\alpha_{\max}|^2 \leq 20$ in Fig. 4(a) using numerical optimization[1] to find the optimal encoding scheme and provide ring-shaped phase

[1] We used a generic Matlab constrained optimisation routine which worked well enough for our purposes. Algorithms such as in [24] may yield more efficient runtime.

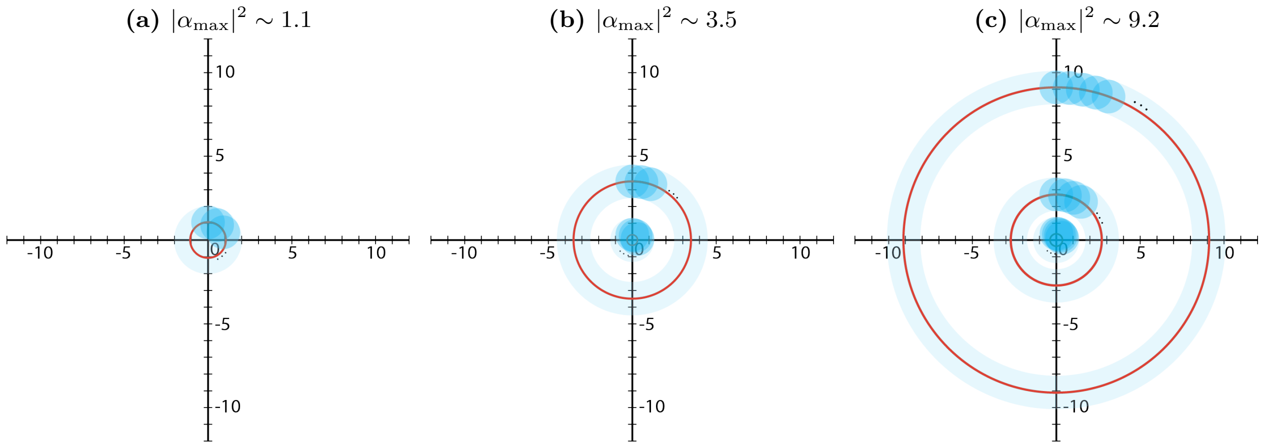

Figure 3: Phase-space visualization of optimal circularly symmetric encoding with coherent state resource at different energies in zero-temperature environment. These are phase-space representations of some instances of optimal encodings plotted in Fig. 4(a),(b). Radius of each red circle indicate energy of its corresponding ring state determined by points-of-increase while, the blue circles indicate attenuated and phase-shifted coherent state codewords that are in each ring state mixture. The coherent state codewords are centered on the circumference of each red circle, with the light blue rings providing a depiction of the ring state corresponding to each point of increase. **(a)** Optimal encoding for input energy $|\alpha_{\max}|^2 \sim 1.1$ with capacity $\chi_{te} \sim 2$ and one ring with energy ~ 1.1. **(b)** Optimal encoding for input energy $|\alpha_{\max}|^2 \sim 3.5$ with capacity $\chi_{te} \sim 3$ and two rings with energies ~ 0.24 and ~ 3.5. **(c)** Optimal encoding for input energy $|\alpha_{\max}|^2 \sim 9.2$ with capacity $\chi_{te} \sim 4$ and three rings with energies ~ 0.3, ~ 2.7, and ~ 9.2.

space representations of these optimal encodings at some fixed values of $|\alpha_{\max}|^2$ in Fig. 3. As the reader may notice in the figure, for any given amplitude $|\alpha_{\max}|$ of the input coherent state, the support of optimal encoding over attenuation η is finite. Fig. 4(a) also shows that the optimal encoding always includes an outermost ring at $|\alpha| = |\alpha_{\max}|$ and that the probability mass of this outermost ring is non-increasing as the $|\alpha_{\max}|$ increases. For small $|\alpha_{\max}|$, it is optimal to encode purely in phase, i.e. using states that forms just one ring with Holevo information given by

$$-\sum_{n=0}^{\infty} \frac{e^{-E_{\max}} E_{\max}^n}{n!} \log \frac{e^{-E_{\max}} E_{\max}^n}{n!} , \tag{9}$$

which is the Shannon entropy of Poisson distributed random variable with values $n \in \mathbb{N}$ with mean $E_{\max} = |\alpha_{\max}|^2$ (see Appendix D.2.1). This is line with the findings of [9], where this ring encoding was shown to be close to the upper bound $g(E_{\max})$ for $E_{\max} \ll 1$ (See Fig. 4(b)).

The threshold at which this single-ring encoding becomes suboptimal can be obtained by solving for $|\alpha_{\max}|$ when adding a vacuum state to the mixture of coherent states begins to increase the information content of the total state, as we know that the optimal encoding is finite over the channel attenuation. That is, we want to solve for $\alpha_{\max}$ that makes

$$\frac{\partial}{\partial \epsilon} S\left(\epsilon \rho_0 + (1-\epsilon)\rho_{\max}\right)\Big|_{\epsilon=0} = 0 , \tag{10}$$

where $\rho_{\max}$ is the ring state at amplitude $\alpha_{\max}$,

$$\rho_{\max} = \int_{-\pi}^{\pi} \frac{d\theta}{2\pi} R_\theta \, |\alpha_{\max}\rangle\langle\alpha_{\max}| \, R_\theta^\dagger , \tag{11}$$

and vacuum state $\rho_0 = |0\rangle\langle 0|$. Interestingly, the solution to this is when the entropy of the encoded states is proportional to its energy

$$S(\rho_{\max}) = \frac{|\alpha_{\max}|^2}{\ln 2} , \tag{12}$$

where we get $|\alpha_{\max}|^2 = 1.25034$ (see Appendix C). It can be observed further in Fig. 4(a) that as $|\alpha_{\max}|^2$ slightly exceeds 1.25034, the optimal encoding now includes an additional codeword $|0\rangle\langle 0|$. As $|\alpha_{\max}|$ increases further, this vacuum codeword gains probability mass until it transforms into a second ring as the first ring codeword diameter further increase. More and more rings are introduced in this way as $E_{\max}$ increases (see Fig. 3). We also plot the encoding capacity of the single-ring encoding and the uniform "flat" encoding (see Appendix D) in Fig. 4(b). For large $|\alpha_{\max}|^2$, the flat encoding is not far from the optimal capacity derived here.

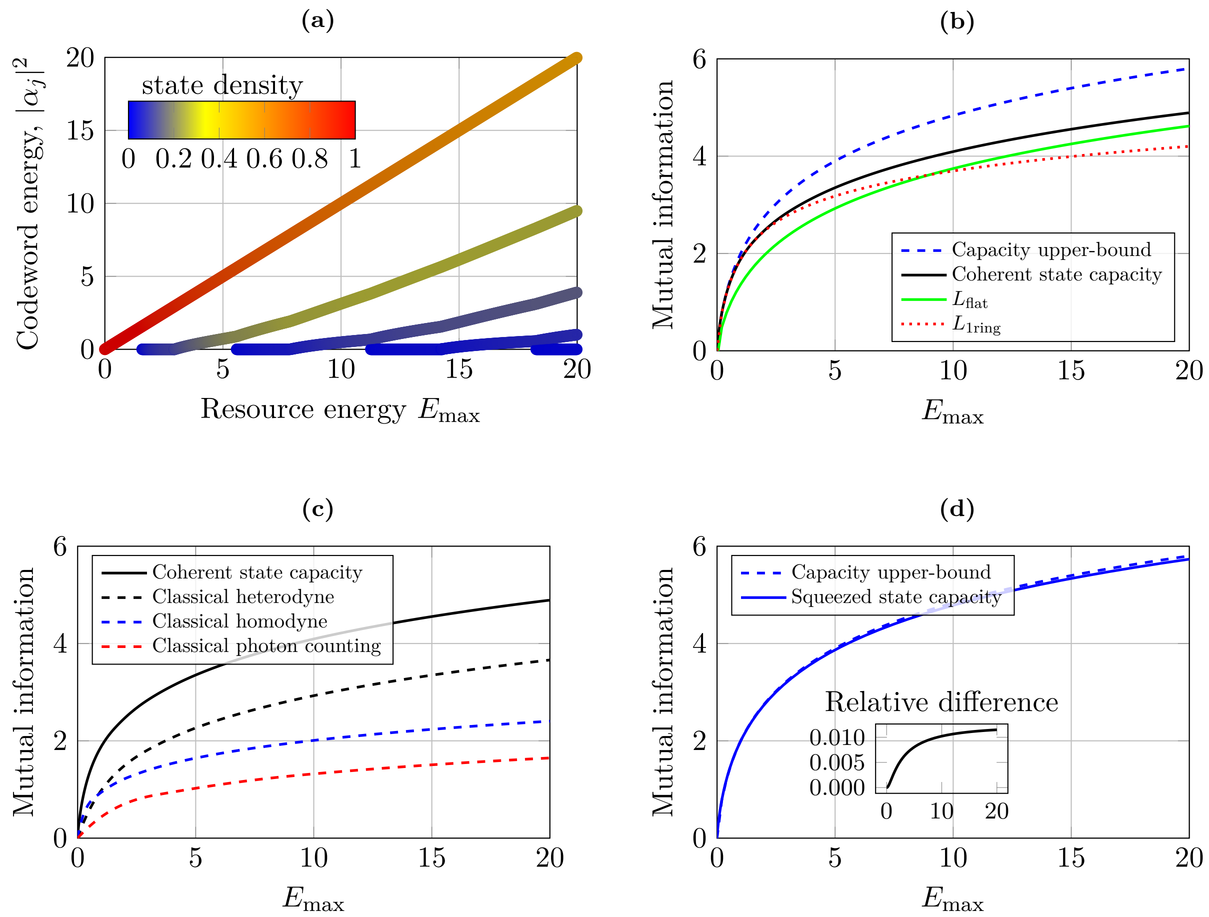

Figure 4: (a) For an optimal circularly symmetric encoding F^* given a coherent-state resource $\rho = |\sqrt{E_{\max}}\rangle\langle\sqrt{E_{\max}}|$ in a $T = 0$ environment, there are a finite number of points-of-increase (POI) $\mathcal{I}_{F^*} = \{\eta_j\}_j$ (see Definition 3). The optimal encoding F^* at a given $E_{\max}$ is specified in the plot by the intersection(s) of the vertical line at $E_{\max}$ with the colored lines. The y-axis value of each colored point corresponds to a codeword with energy $|\alpha_j|^2 = \eta_j E_{\max}$ which correspond to attenuation η_j in the POI of encoding F^* with its density $dF^*(\eta_j)$ color encoded. For example, given energy $E_{\max} \sim 3.5$, there are two points of increase $\eta_0 > \eta_1$ in an optimal encoding F^*. They correspond to codewords with energy $|\alpha_0|^2 = \eta_0 E_{\max} \sim 3.5$ and $|\alpha_1|^2 = \eta_1 E_{\max} \sim 0.24$ with density $dF^*(\eta_0) \sim 0.87$ and $dF^*(\eta_1) \sim 0.13$, respectively. **(b)** Thermal encoding capacity with resource state energy $E_{\max}$ for the capacity upper-bound (7) and coherent resource state with: optimal encoding F^*, one-ring encoding (L_{1ring}), and flat encoding (L_{flat}). One-ring encoding is optimal when $E_{\max} < 1.25034$ and flat encoding is almost optimal when $E_{\max}$ is large (see Appendix D.2.1 and Appendix D.2.2). **(c)** The information capacities over resource energy $E_{\max}$ using homodyne [18], heterodyne [21] and photon counting [19] in comparison with the optimal capacity χ_{tc}. **(d)** The channel capacity for displaced squeezed state resource approximates the capacity upper bound (7), which is achieved by the non-Gaussian state (14) (see Section 3.3). The inset plot shows that their relative difference is at most ~ 1 percent for resource energy $E_{\max} \leq 20$.

The Wigner function of the average state $\rho_{\text{ave}} = \int dF^*(\eta)\rho(\eta)$ where F^* is optimal encoding and $\rho(\eta)$ is a ring state for resource $|\alpha_{\max}\rangle$ is plotted in Fig. 10 in Appendix H for multiple values of $|\alpha_{\max}|$. As $|\alpha_{\max}|$ gets larger, one can observe from Fig. 10 and Fig. 11 that the Wigner function of the average state of the optimal encoding tends to a flat distribution.

For a given $|\alpha_{\max}|$, the one-ring encoding capacity L_{1ring} and the flat distribution capacity L_{flat} (corresponding to a uniform probability distribution on the disk with $\eta \in [0, 1]$ and $\theta \in [0, 2\pi]$ – see also [9]) are clearly lower bounds on the thermal encoding capacity. These bounds are visualized in Fig. 4(b), showing that the former is tighter for small $|\alpha_{\max}|$ and the latter tighter for larger $|\alpha_{\max}|$. As our codewords are pure states, the one-ring capacity is the von Neumann entropy of the ring state at $|\alpha_{\max}|$, given by (9). See Appendix D.2.1 for a simple derivation. For

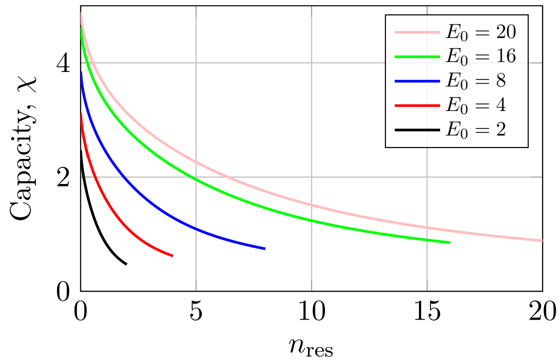

Figure 5: Capacity for a mixed state resource with fixed energy E_0 and mean photon number $n_{\rm res}$ at $T = 0$. For each line, the total energy of the resource is fixed at E_0. The left end of each line (with $n_{\rm res} = 0$) corresponds to a pure coherent state resource, while the right end is a thermal state resource with mean photon number $n_{\rm res} = E_0$. For each E_0, the capacity χ decreases as the resource mean photon number $n_{\rm res}$ increases.

the flat distribution, its capacity is the von Neumann entropy of the average state with encoding $F(\eta) = \eta$, which is given by

$$-\sum_{n=1}^{\infty} P_n[F] \log P_n[F] \tag{13}$$

for $P_n[F] = \frac{n! - \Gamma(1+n, E_{\max})}{n! E_{\max}}$. For a derivation and further discussion, see Appendix D.2.2.

3.2 Mixed state resource

As so far we only discuss pure resource states in this section, we will now consider a mixed Gaussian state as a resource state, namely the displaced thermal state $\rho_{\rm dts}(n_{\rm res}, \alpha) = D(\alpha)\rho_{\rm th}(n_{\rm res})D^\dagger(\alpha)$ where D is the displacement operator and $\rho_{\rm th}(N)$ is a thermal state with mean photon number N. We numerically optimize the capacities over encodings for displaced thermal state resource with different mean photon number $n_{\rm res}$ and fixed total energy E_0 and obtained Fig. 5. All of the optimal encodings obtained from this numerical optimization – which satisfy the optimality cosnditions in Proposition 2 – has a finite POI, supporting Conjecture 1.

Note that when $n_{\rm res} = 0$, we have a pure coherent state resource $\rho_{\rm dts}(0, \alpha) = |\alpha\rangle\langle\alpha|$, and when $n_{\rm res} = E_0$, a thermal state with zero amplitude $\rho_{\rm dts}(n_{\rm res}, 0) = \rho_{\rm th}(n_{\rm res})$. A thermal state with zero amplitude can still be used to transmit information as long as its energy $E_0 = n_{\rm res}$ is greater than zero. However given a fixed energy E_0, having a coherent state as a resource is more optimal compared to a resource thermal state with the same energy. Hence, we can transmit more information as the resource state gets closer to a pure coherent state, as shown in Fig. 5.

3.3 Nearly-optimal squeezed-state resource

Now we go back to the capacity upper-bound $g(E_{\max})$ in (7) for resource state energy $E_{\max}$. It is shown in [9, 10] that the optimal resource state that achieves this is the pure non-Gaussian state $\rho = |\phi\rangle\langle\phi|$ where

$$|\phi\rangle = \frac{1}{\sqrt{E_{\max}+1}} \sum_{n=0}^{\infty} |n\rangle \sqrt{\frac{E_{\max}^n}{(E_{\max}+1)^n}} . \tag{14}$$

This state is optimal because resource state $|\phi\rangle$ and encoding with a uniform phase distribution gives us

$$\chi_{te} = (E_{\max}+1)\log(E_{\max}+1) - E_{\max}\log E_{\max} , \tag{15}$$

which is the maximum capacity for an average-energy constrained channel without any additional constraint on the encoding. This capacity (in the average-energy constrained case) can be achieved using coherent state codewords with a Gaussian distribution encoding (see Appendix D.1).

The state (14) has very high fidelity with respect to a displaced phase-squeezed state even for large $E_{\max} \sim 20$. This motivates us to look at maximum χ_{te} achievable using a ring of displaced phase-squeezed state. We plot the capacity for a displaced squeezed state in Fig. 4(d) which shows that it performs almost as good as (14). The displacement is chosen to match the quadrature expectation value of (14) while the squeezing is set such that the energy of the state is fixed at $E_{\max}$ [2]. The Wigner function and photon number distribution of the average state using the displaced phase-squeezed state and the optimal state for $E_{\max} = 3$ is plotted in Fig. 9 of Appendix H.

4 Non-zero temperature environment

In this section we consider the case of $T > 0$, i.e. the environment is in a thermal state $\gamma_T =$

[2]Optimising the displacement gives only a slight improvement which is indistinguishable in the plot.

$\rho_{\rm th}(n_{\rm env})$ with mean photon number $n_{\rm env} > 0$. We will first discuss the encoding capacity when the resource state is a thermal state and show that the optimal encoding satisfies Conjecture 1. We then proceed to discuss numerical results for coherent state resource $|\alpha\rangle$.

4.1 Thermal state resource

When the resource state ρ is a thermal state $\rho_{\rm th}(n_{\rm res})$ with mean photon number $n_{\rm res}$, the output codewords are also thermal states of the form $\rho_{\rm th}(n_\eta)$ where $n_\eta = \eta n_{\rm res} + (1-\eta)n_{\rm env}$ is a mixture of the resource mean photon number and environment mean photon number. Therefore, for a given encoding F the output average state is a mixture of thermal states $\rho_{\rm ave} = \int dF(\eta)\rho_{\rm th}(n_\eta)$. Since thermal states are invariant under phase shifts, it is useless to encode any information in the phase θ [3]. While this is not ideal for maximizing information transfer, it also means that the optimum decoding measurement is just a measurement of the photon number, which is readily made in practice. In this scenario, the cardinality of the distribution support corresponding to an optimal encoding is finite, as proven in Appendix E.

Proposition 4. *For a thermal-state resource $\rho = \rho_{\rm th}(n_{\rm res})$ and any $T \geq 0$, the attenuation-coefficients η of the optimal circularly-symmetric encoding takes a finite number of values in* $[0,1]$.

The special case when the resource state has mean photon number $n_{\rm res} = 0$ (i.e. a vacuum state) reduces to the scenario in section 3.2. This is due to the codeword corresponding to the transmittance value $\tilde\eta$ being equal to the codeword corresponding to $\eta = 1-\tilde\eta$ in section 3.2 with $\rho = \rho_{\rm th}(n_{\rm env})$ (the displacement is set to zero). The optimal encoding in this case is plotted in Fig. 6(a). We find that for $n_{\rm env} < 8.67754$, the optimal encoding has exactly two codewords: the resource vacuum state ($\eta = 0$) and the thermal state ($\eta = 1$). In Appendix F we obtain an analytical form of the Holevo information

$$\chi_{te} = \log\left(1 + \frac{n_{\rm env}}{(1+n_{\rm env})^{\frac{1+n_{\rm env}}{n_{\rm env}}}}\right) . \qquad (16)$$

[3] There is no contradiction with Proposition 1, however, since for any encoding F, the circularly symmetrized encoding F' has the same Holevo information.

This quantity tends to 1 as $n_{\rm env} \to \infty$. As $n_{\rm env}$ increases beyond 8.67754, it can be seen in Fig. 6(a) that adding more codewords is necessary to achieve the thermal encoding capacity (black line in Fig. 6(b)). Thus interestingly, it is possible to encode information using a vacuum state resource whenever the environmental temperature is non-zero.

4.2 Coherent state resource

Finally, we consider the capacity of a coherent state resource $|\alpha_{\rm max}\rangle$ (resource energy $E_{\rm max} = |\alpha_{\rm max}|^2$) when the environment is in a thermal state with mean photon number $n_{\rm env} > 0$. The numerically optimized thermal encoding capacity is shown in Fig. 7. For a fixed $n_{\rm env}$, a larger $E_{\rm max}$ gives a higher capacity. However when we vary $E_{\rm max}$, a lower $n_{\rm env}$ has higher capacity for larger $E_{\rm max}$, whereas larger $n_{\rm env}$ has higher capacity for smaller $E_{\rm max}$ (see Fig. 7 inset), which is akin to the phenomenon seen in the previous subsection for a vacuum-state resource. In fact, when $n_{\rm env} \gg 1$ and $\alpha_{\rm max} = 0$, we can get $\chi_{te} = 1$ using a two-codeword distribution (Appendix F).

We now look into a communication scenario over a lossy channel which mixes a low-energy $E_{\rm max} \ll 1$ coherent state resource at a fixed attenuation-coefficient $\eta_{\rm ch}$. In this scenario we have a single-ring circularly symmetric encoding and displaced thermal state codewords

$$\rho(\eta_{\rm ch}, \theta) = R_\theta D(\sqrt{\eta_{\rm ch}}\alpha)\rho_{\rm th}(n_{\rm ch})D^\dagger(\sqrt{\eta_{\rm ch}}\alpha)R_\theta^\dagger , \qquad (17)$$

where $\rho_{\rm th}(n_{\rm ch})$ is a thermal state with mean photon number $n_{\rm ch} = (1-\eta_{\rm ch})n_{\rm env}$, we call $n_{\rm ch}$ the codeword *thermal photon number*. The thermal encoding Holevo information is therefore given by

$$\chi[E_{\rm max}] = S(\rho_{\rm ave}) - S(\rho_{\rm th}(n_{\rm ch})) , \qquad (18)$$

as the codeword entropy is equal for all phase θ.

For $E_{\rm max} \ll 1$, it is shown in Fig. 8 that by using a coherent state resource we can almost reach the upper bound of the lossy-channel capacity [5, supplementary information, eqn.(12)], given by

$$g(\eta_{\rm ch}E_{\rm max} + (1-\eta_{\rm ch})n_{\rm env}) - g((1-\eta_{\rm ch})n_{\rm env}) . \qquad (19)$$

In fact for $E_{\rm max} \ll 1$, we can approximate $\chi[E_{\rm max}]$ by

$$\tilde\chi_{te}[E_{\rm max}] = \eta_{\rm ch}E_{\rm max}\log\frac{1+n_{\rm ch}}{n_{\rm ch}} , \qquad (20)$$

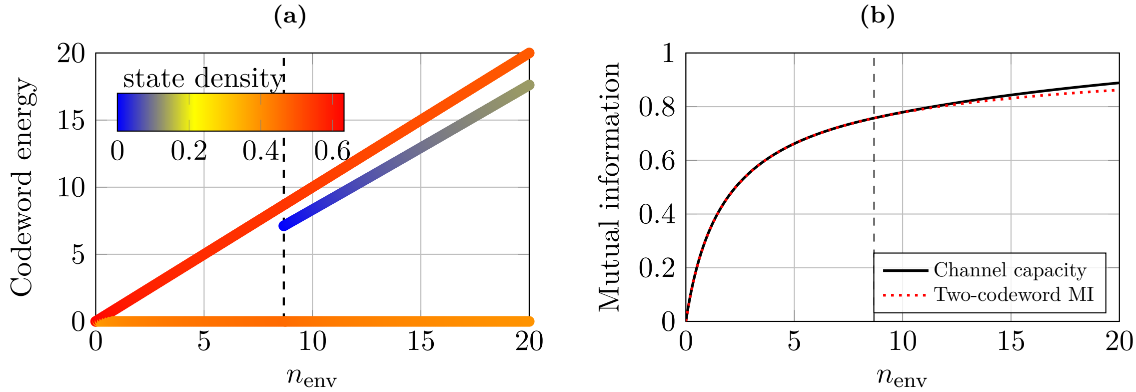

Figure 6: Encoding using vacuum state resource at temperature $T > 0$ (i.e. environment mean photon number $n_{\mathrm{env}} > 0$). (a) Optimal encoding using vacuum state resource with $n_{\mathrm{env}} > 0$. The shading indicates the probability p_j of the j-th codeword. For $n_{\mathrm{env}} < 8.67754$, the two-codeword encoding consisting of the vacuum resource and the thermal environment is optimal (Appendix F). When n_{env} increases beyond 8.67754 (vertical line), the optimal encoding includes a third codeword. **(b)** Capacity and bounds for vacuum state resource as a function of n_{env}. The mutual information from a two-codeword encoding (dotted red line) is optimal when $n_{\mathrm{env}} < 8.67754$, coinciding with the numerically computed capacity (black line).

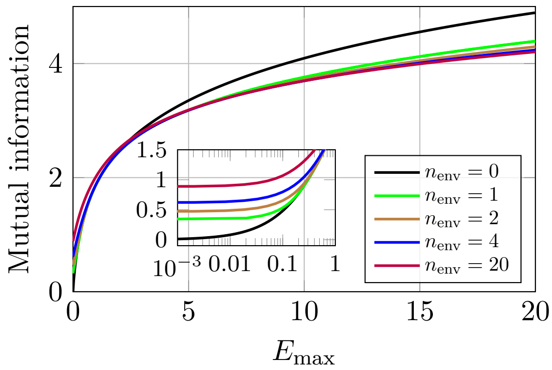

Figure 7: Thermal-encoding capacity for coherent state resource $|\alpha_{\max}\rangle$ at $T > 0$ environment. Capacity varies with environment mean photon number n_{env} (see Section 4.2).

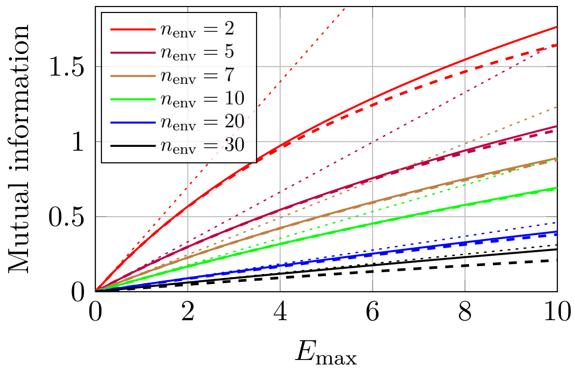

Figure 8: Lossy-channel with attenuation $\eta_{\mathrm{ch}} = 0.4$ at $T > 0$ environment. For $E_{\max} \ll 1$ and various environment mean photon-number n_{env}, 1-ring encoding capacity with coherent state resource (18) (dashed lines) almost reaches the lossy-channel capacity upper-bound (19) (solid lines) and can be approximated by (20) (dotted lines), see Section 4.2.

which is derived in Appendix G and shown as the dotted line in Fig. 8. However as the dotted lines goes above the solid lines for large $E_{\max}$, indicating that (20) is not a good approximation of (18) as $E_{\max}$ increases. Interestingly as n_{ch} increases, indicating more thermal noise (see blue and black lines in Fig. 8), (20) is a better approximation of (19) as the two lines are close to each other even up to the values $E_{\max} > 1$. We also found that this phenomena occurs as η_{ch} decreases.

Approximation (20) implies that the information capacity per unit photon tends to the value $\lim_{E_{\max}\to 0} \tilde{\chi}_{te}[E_{\max}]/E_{\max} = \eta_{\mathrm{ch}} \log \frac{1+n_{\mathrm{ch}}}{n_{\mathrm{ch}}}$, which is finite because $n_{\mathrm{ch}} > 0$. Interestingly, a result by Guha and Shapiro [11] shows that a coherent state resource and binary phase-shift-keying encoding in a vacuum environment ($n_{\mathrm{ch}} \to 0$) has $\chi_{te}[E]/E \to \infty$ as $E \to 0$, allowing transmission of unlimited bits per photon when the average photon in the coherent state is small.

5 Discussion

In this work, we studied the encoding of information by modulating the phase and transmittance of a peak-energy constrained resource state using passive linear optical channels. The thermal channel family – which has been used to model various tasks such as optical memory reading, communication, and algorithmic cooling – is simple to implement and requires no external energy source. We showed that the maximum amount of information that can be encoded by such channels applied to a given resource state can always be achieved by an encoding scheme that uniformly distributes the phase introduced by the channel (Proposition 1). Among such encodings, the unique optimal encoding is characterized by information-theoretic conditions of how the thermal channel transmittance is distributed (Proposition 2).

We conjectured that all optimal circularly symmetric encoding schemes involve only a finite number of values of the channel transmittance. This conjecture is supported by showing that this condition holds in the case of coherent state resource in zero-temperature environments (Proposition 3) and in the case of thermal state resource interacting with an environment at any temperature (Proposition 4). We also supported this conjecture by numerical evidence based on the aforementioned encoding optimality conditions for cases where the channel output is a displaced thermal state. An intriguing question that one may ask is whether there is a fundamental reason why the optimal circularly-symmetric encoding has a discrete attenuation distribution.

Interestingly, Proposition 1 finds an analog in the optimality of such encodings for the classical peak-constrained 2-dimensional AWGN channel [21], while the properties of optimal distribution of the attenuation coefficient in Proposition 2 is similar to the seminal result of Smith [20, 18] on the capacity of the real-valued AWGN channel under a peak energy constraint. These similarities extend to more recent work on n-dimensional AWGN channel capacity under a peak energy constraint (see [22, 23] and references therein). In the quantum domain, our work opens the question of the ultimate capacity of bosonic Gaussian channels under a peak energy constraint on the input ensemble, i.e., to the problem of optimizing the Holevo information over such ensembles generated by thermal encodings and beyond.

Acknowledgement

This work is supported by the Singapore Ministry of Education Tier 2 Grant MOE-T2EP50221-0005, The Singapore Ministry of Education Tier 1 Grants RG146/20 and RG77/22, grant no. FQXi-RFP-1809 (The Role of Quantum Effects in Simplifying Quantum Agents) from the Foundational Questions Institute and Fetzer Franklin Fund (a donor-advised fund of Silicon Valley Community Foundation) and the National Research Foundation, Singapore, and Agency for Science, Technology and Research (A*STAR) under its QEP2.0 programme (NRF2021-QEP2-02-P06). R.N. thanks Saikat Guha for useful discussions on quantum reading capacity. V.N. acknowledges support from the Lee Kuan Yew Endowment Fund (Postdoctoral Fellowship). A.T. acknowledges support from CQT PhD scholarship.

References

[1] Claude Elwood Shannon. "A mathematical theory of communication". The Bell system technical journal **27**, 379–423 (1948).

[2] Alexander Semenovich Holevo. "Bounds for the quantity of information transmitted by a quantum communication channel". Problemy Peredachi Informatsii **9**, 3–11 (1973). url: `http://mi.mathnet.ru/ppi903`.

[3] A.S. Holevo. "The capacity of the quantum channel with general signal states". IEEE Transactions on Information Theory **44**, 269–273 (1998).

[4] Benjamin Schumacher and Michael D. Westmoreland. "Sending classical information via noisy quantum channels". Physical Review A **56**, 131–138 (1997).

[5] V. Giovannetti, R. García-Patrón, N. J. Cerf, and A. S. Holevo. "Ultimate classical communication rates of quantum optical channels". Nature Photonics **8**, 796–800 (2014).

[6] Hsu, Delaubert, Bowen, Fabre, Bachor, and Koy Lam. "A quantum study of multibit phase coding for optical storage". IEEE Journal of Quantum Electronics **42**, 1001–1007 (2006).

[7] Stefano Pirandola. "Quantum reading of a classical digital memory". Physical Review Letters**106** (2011).

[8] Stefano Pirandola, Cosmo Lupo, Vittorio Giovannetti, Stefano Mancini, and Samuel L

Braunstein. “Quantum reading capacity”. New Journal of Physics **13**, 113012 (2011).

[9] Saikat Guha, Zachary Dutton, Ranjith Nair, Jeffrey H. Shapiro, and Brent J. Yen. “Information capacity of quantum reading”. Frontiers in Optics 2011/Laser Science XXVII (2011).

[10] Mark M Wilde, Saikat Guha, Si-Hui Tan, and Seth Lloyd. “Explicit capacity-achieving receivers for optical communication and quantum reading”. In 2012 IEEE International Symposium on Information Theory Proceedings. Pages 551–555. IEEE (2012).

[11] Saikat Guha and Jeffrey H Shapiro. “Reading boundless error-free bits using a single photon”. Physical Review A **87**, 062306 (2013).

[12] A Serafini, M Lostaglio, S Longden, U Shackerley-Bennett, C-Y Hsieh, and G Adesso. “Gaussian thermal operations and the limits of algorithmic cooling”. Physical Review Letters **124**, 010602 (2020).

[13] 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**, 027902 (2004).

[14] Varun Narasimhachar, Syed Assad, Felix C Binder, Jayne Thompson, Benjamin Yadin, and Mile Gu. “Thermodynamic resources in continuous-variable quantum systems”. npj Quantum Information **7**, 9 (2021).

[15] Alexander S Holevo. “Quantum systems, channels, information”. In Quantum Systems, Channels, Information. de Gruyter (2019).

[16] Varun Narasimhachar, Jayne Thompson, Jiajun Ma, Gilad Gour, and Mile Gu. “Quantifying memory capacity as a quantum thermodynamic resource”. Physical Review Letters **122**, 060601 (2019).

[17] Elias M Stein and Rami Shakarchi. “Real analysis: measure theory, integration, and hilbert spaces”. Princeton University Press. (2009).

[18] Joel G. Smith. “The information capacity of amplitude- and variance-constrained sclar gaussian channels”. Information and Control **18**, 203–219 (1971).

[19] S.(Shitz) Shamai. “Capacity of a pulse amplitude modulated direct detection photon channel”. IEE Proceedings I Communications, Speech and Vision **137**, 424 (1990).

[20] Joel Gorham Smith. “On the information capacity of peak and average power constrained gaussian channels”. University of California, Berkeley. (1969). url: https://www.proquest.com/openview/b433005905c536863fa5c50e1d91b0c6/1?pq-origsite=gscholar&cbl=18750&diss=y.

[21] S. Shamai and I. Bar-David. “The capacity of average and peak-power-limited quadrature gaussian channels”. IEEE Transactions on Information Theory **41**, 1060–1071 (1995).

[22] Alex Dytso, Mert Al, H. Vincent Poor, and Shlomo Shamai Shitz. “On the capacity of the peak power constrained vector gaussian channel: An estimation theoretic perspective”. IEEE Transactions on Information Theory **65**, 3907–3921 (2019).

[23] Semih Yagli, Alex Dytso, H. Vincent Poor, and Shlomo Shamai Shitz. “An upper bound on the number of mass points in the capacity achieving distribution for the amplitude constrained additive gaussian channel”. 2019 IEEE International Symposium on Information Theory (ISIT) (2019).

[24] J. Huang and S.P. Meyn. “Characterization and computation of optimal distributions for channel coding”. IEEE Transactions on Information Theory **51**, 2336–2351 (2005).

[25] Alexander S Holevo and Reinhard F Werner. “Evaluating capacities of bosonic gaussian channels”. Physical Review A **63**, 032312 (2001).

[26] Filippo Caruso, Vittorio Giovannetti, and Alexander S Holevo. “One-mode bosonic gaussian channels: a full weak-degradability classification”. New Journal of Physics **8**, 310 (2006).

[27] Maksim E Shirokov. “On the energy-constrained diamond norm and its application in quantum information theory”. Problems of Information Transmission **54**, 20–33 (2018).

[28] Simon Becker and Nilanjana Datta. "Convergence rates for quantum evolution and entropic continuity bounds in infinite dimensions". Communications in Mathematical Physics **374**, 823–871 (2020).

[29] Ranjith Nair. "Quantum-limited loss sensing: Multiparameter estimation and bures distance between loss channels". Physical Review Letters **121**, 230801 (2018).

[30] Ranjith Nair, Guo Yao Tham, and Mile Gu. "Optimal gain sensing of quantum-limited phase-insensitive amplifiers". Physical Review Letters **128**, 180506 (2022).

[31] Ronald J Evans and J Boersma. "The entropy of a poisson distribution (c. robert appledorn)". SIAM Review **30**, 314–317 (1988).

[32] Jerrold E Marsden, Michael J Hoffman, Terry Marsden, et al. "Basic complex analysis". Macmillan. (1999).

[33] Borzoo Rassouli and Bruno Clerckx. "An upper bound for the capacity of amplitude-constrained scalar awgn channel". IEEE Communications Letters **20**, 1924–1926 (2016).

[34] A.L. McKellips. "Simple tight bounds on capacity for the peak-limited discrete-time channel". International Symposium onInformation Theory, 2004. ISIT 2004. Proceedings. (2004).

[35] Andrew Thangaraj, Gerhard Kramer, and Georg Bocherer. "Capacity bounds for discrete-time, amplitude-constrained, additive white gaussian noise channels". IEEE Transactions on Information Theory **63**, 4172–4182 (2017).

[36] Amos Lapidoth and Stefan M. Moser. "On the capacity of the discrete-time poisson channel". IEEE Transactions on Information Theory **55**, 303–322 (2009).

[37] Amos Lapidoth, Stefan M. Moser, and MichÈle A. Wigger. "On the capacity of free-space optical intensity channels". IEEE Transactions on Information Theory **55**, 4449–4461 (2009).

A Proofs of general properties of optimal thermal encoding

A.1 Proof of proposition 1

Consider a thermal encoding F and its corresponding codewords $\{\rho(\eta,\theta)\}_{\eta,\theta}$. The Holevo information of this encoding F is given by

$$\chi_{tc}[F] = S(\rho_{\text{ave}}) - \int dF(\eta,\theta) S(\rho(\eta,\theta)) \,, \tag{21}$$

where S is the von Neumann entropy and $\rho_{\text{ave}} = \int dF(\eta,\theta)\rho(\eta,\theta)$ is the average density operator with respect to F. First we will show that any state that is an average of codewords over uniformly distributed phase is diagonal in Fock basis.

Lemma 1. *For any state $\rho = \int dF(x)\rho_x$ with pure ρ_x's, its average state with uniform phase distribution $\tilde{\rho} = \frac{1}{2\pi}\int d\phi \; R_\phi \rho R_\phi^\dagger$ for $\phi \in [-\pi,\pi]$ phase rotation operator R_ϕ is diagonal in the Fock basis.*

Proof. This can be shown by writing each pure state in terms of Fock basis $\rho_x =$

$\sum_{n,m} z_n(x)z_m(x)^* |n\rangle\langle m|$ for $z_n(x), z_m(x) \in \mathbb{C}$, then we get

$$
\begin{aligned}
\tilde{\rho} &= \int dF(x) \int \frac{d\phi}{2\pi} R_\phi \rho_x R_\phi^\dagger \\
&= \int dF(x) \sum_{n,m} z_n(x) z_m(x)^* \int \frac{d\phi}{2\pi} R_\phi |n\rangle\langle m| R_\phi^\dagger \\
&= \int dF(x) \sum_{n,m} z_n(x) z_m(x)^* \int \frac{d\phi}{2\pi} e^{i(n-m)\phi} |n\rangle\langle m| \\
&= \sum_{n,m} \int dF(x) z_n(x) z_m(x)^* \delta_{n,m} |n\rangle\langle m| = \sum_n |n\rangle\langle m| \int dF(x) |z_n(x)|^2 ,
\end{aligned}
$$

which shows that $\tilde{\rho}$ is diagonal in Fock basis $\{|n\rangle\}_{n=0}^\infty$ with real diagonal elements $\{\int dF(x)|z_n(x)|^2\}_{n=0}^\infty$. □

Now we are ready for the proof of Proposition 1. Namely, we show that for an arbitrary thermal encoding F, there is a circularly symmetric encoding $\tilde{F}$ giving an average state $\tilde{\rho}_{\text{ave}} = \int dF(\eta) \int \frac{d\phi}{2\pi} \rho(\eta, \phi)$ with at least the same entropy. The first term of (21) is upper bounded by the entropy of $\tilde{\rho}_{\text{ave}}$

$$S(\rho_{\text{ave}}) \overset{(1)}{=} \int \frac{d\phi}{2\pi} S\Big(\int dF(\eta,\theta) R_\phi \rho(\eta,\theta) R_\phi^\dagger\Big) \tag{22}$$

$$\overset{(2)}{\leq} S\Big(\int dF(\eta,\theta) \int \frac{d\phi}{2\pi} R_\phi \rho(\eta,\theta) R_\phi^\dagger\Big) \tag{23}$$

$$\overset{(3)}{=} S\Big(\int dF(\eta,\theta) \int \frac{d\phi}{2\pi} \rho(\eta,\phi)\Big) \tag{24}$$

$$\overset{(4)}{=} S\Big(\int dF(\eta) \int \frac{d\phi}{2\pi} \rho(\eta,\phi)\Big) = S(\tilde{\rho}_{\text{ave}}) \tag{25}$$

where (1) is because phase rotation R is a unitary operation, thus does not change the entropy of a state, and (2) by the concavity of the von Neumann entropy, and (3) by defining $\rho(\eta,\phi) = R(\phi)\rho(\eta,\theta)R^\dagger(\phi)$ because state $\int \frac{d\phi}{2\pi} \rho(\eta,\phi)$ is diagonal in Fock basis by Lemma 1, and hence independent of phase θ, and lastly (4) by marginalizing $\int dF(\eta,\theta)$ over phase θ.

Now we look at the second term of (21). In the case that all codeword $\rho(\eta,\theta)$ is a pure state, the second term is zero and therefore $\chi_{tc}[\tilde{F}] \geq \chi_{tc}[F]$ holds because $S(\tilde{\rho}_{\text{ave}}) \geq S(\rho_{\text{ave}})$. In the case of general codewords $\rho(\eta,\theta)$, the second term of (21) is equal to $\int d\tilde{F}(\eta,\theta) S(\rho(\eta,\theta))$ because $S(\rho(\eta,\theta))$ is independent of phase θ. Hence we constructed a circularly symmetric encoding $\tilde{F}$ such that $\chi_{tc}[\tilde{F}] \geq \chi_{tc}[F]$ for an arbitrary encoding F.

A.2 Levy metric

For completeness, in this section we define the Levy metric which we will use as a metric on the space of encodings (i.e. cumulative distribution functions) in order to prove some of the properties of an optimal circularly symmetric encoding. Here we specifically define the metric on $\mathscr{F}$, the space of encodings (i.e. cumulative distribution over the interval $[0,1]$). The Levy metric $d_{Le} : \mathscr{F} \times \mathscr{F} \to \mathbb{R}$ is defined as

$$d_{Le}(F,G) = \frac{1}{\sqrt{2}} \sup_{x,x' : F(x)+x = G(x')+x'} d\Big((x, F(x)), (x', G(x'))\Big) .$$

Here $d(\cdot,\cdot) : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$ is the euclidean distance in $\mathbb{R}^2$ and we maximize the distance between two points on the graphs F and G intersecting a diagonal line $a = x + y$ (on an (x,y)-coordinate plane), over all $a \in \mathbb{R}$. (for discussions on its geometric interpretations, see Chapter 2 of [20]). We show that d_{Le} is indeed a metric as $d(r,s) = d(s,r)$ for all $r, s \in \mathbb{R}^2$, and $d((x, F(x)), (x', G(x'))) = 0$ for a maximizing x, x' iff $F(x) = G(x)$ for all x, and satisfies the triangle inequality

$$\begin{aligned} d_{Le}(F,G) + d_{Le}(G,H) &= \frac{1}{\sqrt{2}} \sup_{x,x'} d\Big((x, F(x)), (x'G(x'))\Big) + \frac{1}{\sqrt{2}} \sup_{z,z'} d\Big((z, G(z)), (z', H(z'))\Big) \\ &\geq \frac{1}{\sqrt{2}} \sup_{x,x',z,z'} \Big(d\Big((x, F(x)), (x'G(x'))\Big) + d\Big((z, G(z)), (z', H(z'))\Big)\Big) \\ &\geq \frac{1}{\sqrt{2}} \sup_{x,z'} d\Big((x, F(x)), (z', H(z'))\Big) = d_{Le}(F,H) \end{aligned}$$

where the maximization in the second line is over x, x', z, z' such that $F(x) + x = G(x') + x' = G(z) + z = H(z') + z'$. The last line is obtained by simply noting that $x' = z$ must be true since $G(x') + x' = G(z) + z$ and for any given three points $c_F = (x, F(x)), c_G = (x', G(x')), c_H = (z', H(z'))$ in $\mathbb{R}^2$ on a line, distance between c_F and c_H simply cannot exceed the sum of distances between c_F and c_G and between c_G and c_H.

A.3 Proof of proposition 2

We give a proof of Proposition 2 analogous to the optimization argument given by Smith [18] about the optimality of AWGN channel encoding. First, we show that the thermal channel Holevo information χ_{tc} is concave (Lemma 3) and weakly differentiable (Lemma 4) and continuous (Lemma 6) over the space of circularly symmetric encoding F equipped with the Levy metric (for definition, see Appendix A.2), then argue that $\mathscr{F}$ is convex and compact. If χ_{tc} and $\mathscr{F}$ satisfy these properties, we may use the optimization theorem to obtain conditions for an optimal circularly symmetric encoding.

Lemma 2 (Optimization theorem)**.** *Consider mapping f from compact and convex topological space Ω to real numbers $\mathbb{R}$. If f is continuous, concave, and weakly differentiable, then there exists a $x^* \in \Omega$ such that $f(x^*) = \sup_{x\in\Omega} f(x)$, and it is necessary and sufficient for such x^* that the weak derivative of f at x^* in the direction of x, defined by*

$$f'_{x^*}(x) = \lim_{\epsilon\to 0} \frac{f(x^* + \epsilon(x - x^*)) - f(x^*)}{\epsilon}$$

for $\epsilon > 0$, is non-positive for all $x \in \Omega$. Moreover if f is strictly concave, then such x^* is unique.

A proof of this optimization theorem can be found in [20, p.15]. After establishing the conditions required to use Lemma 2, we will later conclude the proof by showing that

$$i[\eta, F^*] \leq \chi_{tc}[F^*] \quad \text{for all } \eta \in [0,1] \tag{26a}$$

$$i[\eta, F^*] = \chi_{tc}[F^*] \quad \text{if and only if } \eta \in \mathcal{I} \tag{26b}$$

hold for encoding F^* if and only if F^* has non-positive weak derivative in all directions, thus (26a) and (26b) hold if and only if $\chi_{tc}[F^*] = \sup_F \chi_{tc}[F]$.

Note that all encodings F in $\mathscr{F}$ distribute the channel phase uniformly and independently of the distribution over the channel attenuation, therefore optimization of $\chi_{tc}[F]$ is only over encodings over the attenuation, i.e. space of circularly symmetric encodings $\mathscr{F}$ is the space of cumulative distribution functions over the interval $[0,1]$. Space $\mathscr{F}$ is convex because an encoding $F_\lambda = \lambda F_1 + (1-\lambda)F_2$ is clearly in $\mathscr{F}$ for any $\lambda \in [0,1]$ and $F_1, F_2, \in \mathscr{F}$. The compactness of $\mathscr{F}$ follows from the compactness proof of the space of cumulative distribution functions over interval $[-a, a]$ for some real a in the Levy metric by Smith [20, pp.21-25] where he shows that this space is totally bounded, from which compactness follows. Smith's proof for total boundedness of encodings over $[-a, a]$ applies to space of encodings over any bounded and connected interval of $\mathbb{R}$, thus works for $\mathscr{F}$.

Lemma 3. *Thermal channel Holevo information χ_{tc} is a strictly concave function of encodings in $\mathscr{F}$.*

Proof. For $\lambda \in [0,1]$ and an arbitrary pair of encodings F_1 and F_2, let $F_\lambda = \lambda F_1 + (1-\lambda)F_2$ for $\lambda \in (0,1)$. We will show that $\chi_{tc}[F_\lambda] \geq \lambda\chi_{tc}[F_1] + (1-\lambda)\chi_{tc}[F_2]$ with equality iff $F_1 = F_2$. By the definition of χ_{tc} of F_λ,

$$\chi_{tc}[F_\lambda] = S\left(\int dF_\lambda(\eta,\theta)\rho(\eta,\theta)\right) - \int dF_\lambda(\eta,\theta)S(\rho(\eta,\theta)). \tag{27}$$

Now let $\rho_{\text{ave}j} = \int dF_j(\eta,\theta)\rho(\eta,\theta)$ for $j \in \{1,2\}$. Because for any vector $|k\rangle$ from basis $\{|k\rangle\}_k$ the integral $\int dF\langle k|\rho|k\rangle$ is finite for any F, so we have

$$\int dF_\lambda(\eta,\theta)\rho(\eta,\theta) = \lambda\rho_{\text{ave1}} + (1-\lambda)\rho_{\text{ave2}}\ . \tag{28}$$

Hence by the strict concavity of the von Neumann entropy, we obtain

$$S\left(\int dF_\lambda(\eta,\theta)\rho(\eta,\theta)\right) \geq \lambda S(\rho_{\text{ave1}}) + (1-\lambda)S(\rho_{\text{ave2}})\ , \tag{29}$$

where equality holds iff $\rho_{\text{ave1}} = \rho_{\text{ave2}}$, which happens iff $F_1 = F_2$.

As for the second term of $\chi_{tc}[F_\lambda]$, since the entropy $S(\rho(\eta,\theta))$ is integrable with respect to any encoding, we have

$$\int dF_\lambda(\eta,\theta)S(\rho(\eta,\theta)) = \lambda\int dF_1(\eta,\theta)S(\rho(\eta,\theta)) + (1-\lambda)\int dF_2(\eta,\theta)S(\rho(\eta,\theta)). \tag{30}$$

By applying the equations (29) and (30) to (27), the required concavity condition $\chi_{tc}[F_\lambda] \geq \lambda\chi_{tc}[F_1] + (1-\lambda)\chi_{tc}[F_2]$ is therefore shown. □

Now we show the weak-differentiability of χ_{tc}.

Lemma 4 (Weak differentiability of χ_{tc})**.** *Thermal channel Holevo information χ_{tc} is weakly differentiable. Particularly, given any pair of encodings F_0 and F, for weak derivative of χ_{tc} at F_0 in the direction of F*

$$\chi'_{F_0}[F] := \lim_{\epsilon\to 0}\frac{\chi_{tc}[F_0+\epsilon(F-F_0)] - \chi_{tc}[F_0]}{\epsilon} \tag{31}$$

for $\epsilon > 0$, we have

$$\chi'_{F_0}[F] = \left(\int dF(\eta)\, i[\eta,F_0]\right) - \chi_{tc}[F_0]\ . \tag{32}$$

Proof. First, let $F_\epsilon = (1-\epsilon)F_0 + \epsilon F$ so $dF_\epsilon = (1-\epsilon)dF_0 + \epsilon dF$. Then we have

$$\chi'_{F_0}[F] = \lim_{\epsilon\to 0}\frac{1}{\epsilon}\left(\chi_{tc}[F_\epsilon] - \chi_{tc}[F_0]\right) \tag{33}$$

$$= \lim_{\epsilon\to 0}\frac{1}{\epsilon}\left((1-\epsilon)\int dF_0(\eta)\ i[\eta,F_\epsilon] + \epsilon\int dF(\eta)\ i[\eta,F_\epsilon] - \chi_{tc}[F_0]\right) \tag{34}$$

$$= \int dF(\eta)i[\eta,F_0] - \int dF_0(\eta)i[\eta,F_0] + \lim_{\epsilon\to 0}\frac{1}{\epsilon}\left(\int dF_0(\eta)i[\eta,F_\epsilon] - \chi_{tc}[F_0]\right)\ , \tag{35}$$

where we used the expression of the Holevo information in terms of the marginal information density

$$\chi_{tc}[F] = \int dF(\eta)i[\eta,F] \tag{36}$$

as in (5) to obtain the second equality and note that $\lim_{\epsilon\to 0} i[\eta, F_\epsilon] = i[\eta, F_0]$ in the last line. Noting that $\int dF_0(\eta) i[\eta, F_0] = \chi_{tc}[F_0]$ by (36), it remains for us to show that the limit term equals to 0. We do this by again using (36) as follows

$$\lim_{\epsilon\to 0}\frac{1}{\epsilon}\left(\int dF_0(\eta) i[\eta, F_\epsilon] - \chi_{tc}[F_0]\right) = \lim_{\epsilon\to 0}\frac{1}{\epsilon}\int dF_0(\eta)\left(i[\eta, F_\epsilon] - i[\eta, F_0]\right) \tag{37}$$

$$= \lim_{\epsilon\to 0}\frac{1}{\epsilon}\int dF_0(\eta)\left(\sum_n P_n[\rho(\eta)]\log\frac{P_n[F_0]}{P_n[F_\epsilon]}\right) \tag{38}$$

$$= \lim_{\epsilon\to 0}\frac{1}{\epsilon}\left(\sum_n P_n[F_0]\log\frac{P_n[F_0]}{P_n[F_\epsilon]}\right) . \tag{39}$$

By the virtue of the L'Hôpital's rule, we evaluate the limit of each terms in the sum to obtain

$$\lim_{\epsilon\to 0}\frac{P_n[F_0]}{\epsilon}\log\frac{P_n[F_0]}{P_n[F_\epsilon]} = \lim_{\epsilon\to 0}\frac{1}{\frac{d}{d\epsilon}\epsilon}\left(\frac{d}{d\epsilon}P_n[F_0]\log\frac{P_n[F_0]}{P_n[F_\epsilon]}\right) \tag{40}$$

$$= P_n[F_0]\lim_{\epsilon\to 0}\frac{P_n[F] - P_n[F_0]}{((1-\epsilon)P_n[F_0] + \epsilon P_n[F])\ln 2} \tag{41}$$

$$= \frac{P_n[F] - P_n[F_0]}{\ln 2} . \tag{42}$$

Substituting this back into the sum gives us

$$\lim_{\epsilon\to 0}\frac{1}{\epsilon}\left(\int dF_0(\eta) i[\eta, F_\epsilon] - \chi_{tc}[F_0]\right) = \frac{1}{\ln 2}\left(\sum_n P_n[F] - \sum_m P_m[F_0]\right) = 0 , \tag{43}$$

concluding the proof for (32). □

Here we will give a lemma to show weak continuity of $\rho(\eta)$, i.e. that $\langle\varphi'|\rho(\eta)|\varphi\rangle$ is a continuous function over η for any $|\varphi'\rangle$ and $|\varphi\rangle$, which we will need later in showing that Holevo information χ_{tc} is continuous over the space of circularly symmetric encodings equipped with the Levy metric (for definition, see Appendix A.2).

Lemma 5. *Given a resource state ρ with energy $E < \infty$, its ring state $\rho(\eta)$ is weakly continuous over $\eta \in [0, 1]$; namely for any pair of states $|\varphi'\rangle$ and $|\varphi\rangle$, $\langle\varphi'|\rho(\eta)|\varphi\rangle$ is continuous.*

Proof. We will show that $\lim_{\eta'\to\eta} ||\rho(\eta) - \rho(\eta')||_1 = 0$ where $||\cdot||_1$ is the trace norm, which implies that $\langle\varphi'|\rho(\eta)|\varphi\rangle$ is a continuous function over η for any pair of states $|\varphi'\rangle$ and $|\varphi\rangle$. First, we consider amplifier channel $\mathcal{A}_G$ with gain $G = (1-\eta)n_{\mathrm{env}} + 1$ and attenuator channel $\mathcal{L}_{\tilde\eta}$ with attenuation $\tilde\eta = \frac{\eta}{G}$, defined by unitaries U ad V over the input system and environment system as

$$\mathcal{A}_G(\rho) = \mathrm{tr}_E(U\rho\otimes\rho_{\mathrm{env}}U^\dagger) \quad\text{and}\quad \mathcal{L}_{\tilde\eta}(\rho) = \mathrm{tr}_E(V\rho\otimes\rho_{\mathrm{env}}V^\dagger) ,$$

respectively. Unitaries U and V are defined by their actions (in the Heisenberg picture) on the annihilation operator $\hat a$ of the input system [25, 26]

$$U^\dagger(\hat a\otimes I_{\mathrm{env}})U = \sqrt{G}(\hat a\otimes I_{\mathrm{env}}) + \sqrt{G-1}(I\otimes\hat a^\dagger_{\mathrm{env}})$$

and

$$V^\dagger(\hat a\otimes I_{\mathrm{env}})V = \sqrt{\tilde\eta}(\hat a\otimes I_{\mathrm{env}}) + \sqrt{1-\tilde\eta}(I\otimes\hat a_{\mathrm{env}})$$

where $\hat a_{\mathrm{env}}$ and I_{env} are the annihilation operator and identity operator of the environment, respectively. Similarly for attenuation η' we define $G' = (1-\eta')n_{\mathrm{env}} + 1$ and $\tilde\eta' = \frac{\eta'}{G'}$.

Using the one-mode channel decomposition result in [26], we may write a channel $\mathcal{N}_\eta^{(n_{\mathrm{env}})}$ that mixes an input state with attenuation η with a thermal environment with mean photon number n_{env} as a composition of an attenuation and an amplifier channel

$$\mathcal{N}_\eta^{(n_{\mathrm{env}})} = \mathcal{A}_G\circ\mathcal{L}_{\tilde\eta} .$$

Therefore we may write the thermal channel output state $\rho(\eta) = \mathcal{A}_G \circ \mathcal{L}_{\tilde{\eta}}(\rho)$ (similarly with attenuation η'), so we get

$$\begin{aligned} ||(\mathcal{A}_G \circ \mathcal{L}_{\tilde{\eta}} - \mathcal{A}_{G'} \circ \mathcal{L}_{\tilde{\eta}'})(\rho)||_1 &= ||(\mathcal{A}_G \circ \mathcal{L}_{\tilde{\eta}} - \mathcal{A}_{G'} \circ \mathcal{L}_{\tilde{\eta}} + \mathcal{A}_{G'} \circ \mathcal{L}_{\tilde{\eta}} - \mathcal{A}_{G'} \circ \mathcal{L}_{\tilde{\eta}'})(\rho)||_1 \\ &\overset{(1)}{\leq} ||(\mathcal{A}_G \circ \mathcal{L}_{\tilde{\eta}} - \mathcal{A}_{G'} \circ \mathcal{L}_{\tilde{\eta}})(\rho)||_1 + ||(\mathcal{A}_{G'} \circ \mathcal{L}_{\tilde{\eta}} - \mathcal{A}_{G'} \circ \mathcal{L}_{\tilde{\eta}'})(\rho)||_1 \\ &= ||(\mathcal{A}_G - \mathcal{A}_{G'}) \circ \mathcal{L}_{\tilde{\eta}}(\rho)||_1 + ||\mathcal{A}_{G'} \circ (\mathcal{L}_{\tilde{\eta}} - \mathcal{L}_{\tilde{\eta}'})(\rho)||_1 \\ &\overset{(2)}{\leq} ||(\mathcal{A}_G - \mathcal{A}_{G'}) \circ \mathcal{L}_{\tilde{\eta}}(\rho)||_1 + ||(\mathcal{L}_{\tilde{\eta}} - \mathcal{L}_{\tilde{\eta}'})(\rho)||_1 \end{aligned}$$

where we again used the triangle inequality and data processing inequality to obtain (1) and (2), respectively. Because ρ has energy E, we can use the energy-constrained diamond norm [27, 28] on the amplifier channels and attenuator channels in the last line, defined as $||\mathcal{N} - \mathcal{M}||_{\diamond E} = \sup_\rho ||(\mathcal{N} - \mathcal{M}) \otimes \mathcal{I}_C(\rho)||_1$ where $\mathcal{I}_C$ is the identity channel for some ancilla system C and the supremum is taken over all states ρ on the product space of the input system and the ancilla system C with energy constraint E on the input system. Thus we obtain the following upper bound

$$\begin{aligned} ||(\mathcal{A}_G \circ \mathcal{L}_{\tilde{\eta}} - \mathcal{A}_{G'} \circ \mathcal{L}_{\tilde{\eta}'})(\rho)||_1 &\leq ||\mathcal{A}_G - \mathcal{A}_{G'}||_{\diamond E\tilde{\eta}} + ||\mathcal{L}_{\tilde{\eta}} - \mathcal{L}_{\tilde{\eta}'}||_{\diamond E} \\ &\leq 2\beta_{E\eta}(\mathcal{A}_G, \mathcal{A}_{G'}) + 2\beta_E(\mathcal{L}_{\tilde{\eta}}, \mathcal{L}_{\tilde{\eta}'}) \end{aligned}$$

as attenuator channel $\mathcal{L}_{\tilde{\eta}}$ scales the energy constraint by $\tilde{\eta}$ for the first term, and for the second inequality we use [15, Lemma 10.17] to bound the diamond norms by energy-constrained Bures distance β_E which can be expressed as [29, 30]

$$\begin{aligned} \beta_{E\eta}(\mathcal{A}_G, \mathcal{A}_{G'}) &= \sqrt{2}\sqrt{1 - (1 - \{E\eta\})\nu^{\lfloor E\eta \rfloor} - \{E\eta\}\nu^{\lfloor E\eta \rfloor + 1}} \\ \beta_E(\mathcal{L}_{\tilde{\eta}}, \mathcal{L}_{\tilde{\eta}'}) &= \sqrt{2}\sqrt{1 - (1 - \{E\})\mu^{\lfloor E \rfloor} - \{E\}\mu^{\lfloor E \rfloor + 1}} \end{aligned}$$

where $\lfloor x \rfloor$ and $\{x\}$ are the integer part and fractional part of x, respectively, and $\mu = \sqrt{\tilde{\eta}\tilde{\eta}'} + \sqrt{(1-\tilde{\eta})(1-\tilde{\eta}')}$ and $\nu = (\sqrt{GG'} + \sqrt{(G-1)(1-G'-1)})^{-1}$.

As $\eta' \to \eta$ we have $\tilde{\eta}' \to \tilde{\eta}$ and $G' \to G$ which implies $\mu \to 1$ and $\nu \to 1$. Hence both $\beta_{E\eta}(\mathcal{A}_G, \mathcal{A}_{G'})$ and $\beta_E(\mathcal{L}_{\tilde{\eta}}, \mathcal{L}_{\tilde{\eta}'})$ limit as $\eta' \to \eta$ is 0, implying that $\lim_{\eta' \to \eta} ||\rho(\eta) - \rho(\eta')||_1 = 0$, and therefore as a consequence $\rho(\eta)$ is weakly continuous. $\square$

Lemma 6. *Given resource state ρ with energy $E < \infty$, the thermal channel Holevo information χ_{tc} is continuous over space of circularly symmetric encodings $\mathscr{F}$ equipped with the Levy metric d_{Le}.*

We emphasize that this continuity condition partially rests on our assumption that the resource state has a finite energy $E_{\max}$ as indicated in the beginning of Section 2.1. Here we use the continuity of von Neumann entropy S over $\mathscr{G}_E$, which is the set of states with energy at most $E_{\max}$, shown in [15, Lemma 11.8] to show the continuity of χ_{tc} over $\mathscr{F}$ in 4 steps:

1. Convergence of a sequence of encodings $\{F_n\}_{n\in\mathbb{N}} \subseteq \mathscr{F}$ in the Levy metric to some $F \in \mathscr{F}$ implies that the corresponding sequence of average states $\{\rho_{F_n}\}_{n\in\mathbb{N}} \subseteq \mathscr{G}_E$ weakly converges to $\rho_F \in \mathscr{G}_E$.

2. Weak convergence (for detailed discussions see [15, Section 11.1]) of $\{\rho_{F_n}\}_{n\in\mathbb{N}}$ to some ρ_F implies that $\lim_{n\to\infty} S(\rho_{F_n}) = S(\rho_F)$ because S is continuous (by [15, Lemma 11.8]). Therefore, $S(\rho_F)$ is continuous over $F \in \mathscr{F}$.

3. The second term of χ_{tc} , $\int dF'(\eta)\rho(\eta)$ is continuous over $F' \in \mathscr{F}$, i.e.

$$\lim_{n\to\infty} \int dF_n(\eta) S(\rho(\eta)) = \int dF(\eta) S(\rho(\eta))$$

   if $\lim_{n\to\infty} F_n = F$ in the Levy metric.

4. Putting the previous two steps together, the Holevo information

$$\chi_{tc}[F] = S(\rho_F) - \int dF(\eta) S(\rho(\eta))$$

is continuous over $F \in \mathscr{F}$.

Proof. First we show that if a sequence of encodings $\{F_n\}_{n\in\mathbb{N}} \subseteq \mathscr{F}$ converging to some $F \in \mathscr{F}$ under the Levy metric, then sequence of states $\{\rho_{F_n}\}_{n\in\mathbb{N}}$ is weakly converging to some ρ_F, where $\rho_{F'} = \int dF'(\eta)\rho(\eta)$ with $F' \in \mathscr{F}$. More precisely, for all encoding sequence $\{F_n\}_{n\in\mathbb{N}} \subseteq \mathscr{F}$ such that $\lim_{n\to\infty} d_{Le}(F_n, F) = 0$ for some $F \in \mathscr{F}$, we have $\lim_{n\to\infty}\langle\varphi'|\rho_{F_n}|\varphi\rangle = \langle\varphi'|\rho_F|\varphi\rangle$ for all states $|\varphi'\rangle, |\varphi\rangle$. First, we note that for any $F' \in \mathscr{F}$

$$\langle\varphi'|\rho_{F'}|\varphi\rangle = \int dF'(\eta)\langle\varphi'|\rho(\eta)|\varphi\rangle \ .$$

Given that a convergence in the Levy metric implies a weak convergence of distributions (defined as $\lim_{n\to\infty}\int dF_n(x)f(x) = \int dF(x)f(x)$ for continuous and bounded f) in compact metric space $\mathscr{F}$ (hence is separable), and since $\langle\varphi'|\rho(\eta)|\varphi\rangle$ is continuous (by Lemma 5) and bounded (as $\rho(\eta)$ is a density matrix hence is a bounded operator) over $\eta \in [0,1]$ for all $|\varphi'\rangle, |\varphi\rangle$, we have

$$\lim_{n\to\infty}\langle\varphi'|\rho_{F_n}|\varphi\rangle = \lim_{n\to\infty}\int dF_n(\eta)\langle\varphi'|\rho(\eta)|\varphi\rangle = \int dF(\eta)\langle\varphi'|\rho(\eta)|\varphi\rangle = \langle\varphi'|\rho_F|\varphi\rangle \ .$$

Therefore, Levy-metric convergence $\lim_{n\to\infty} d_{Le}(F_n, F) = 0$ implies weak convergence of states (defined as $\lim_{n\to\infty}\langle\varphi'|\rho_{F_n}|\varphi\rangle = \langle\varphi'|\rho_F|\varphi\rangle$).

By [15, Lemma 11.8], weak convergence of a sequence of bounded-energy states $\{\rho_n\}_{n\in\mathbb{N}} \subseteq \mathscr{G}_E$ to some $\rho \in \mathscr{G}_E$ implies the convergence of their von Neumann entropy. More precisely, $\lim_{n\to\infty}\langle\varphi'|\rho_n|\varphi\rangle = \langle\varphi'|\rho|\varphi\rangle$ implies

$$\limsup_{n\to\infty} S(\rho_n) \le S(\rho) \le \liminf_{n\to\infty} S(\rho_n) \ .$$

Now, in order to show that the second term of χ_{tc} is continuous over $\mathscr{F}$, we first note that $S(\rho(\eta))$ is continuous and bounded over $\eta \in [0,1]$ (as the thermal channel does not increase energy, hence [15, Lemma 11.8] applies) and that the convergence of a sequence of encodings $\{F_n\}_{n\in\mathbb{N}} \subseteq \mathscr{F}$ in the Levy metric to $F \in \mathscr{F}$ implies weak convergence of distributions (again, because $\mathscr{F}$ is a compact metric space). Therefore, $\lim_{n\to\infty} F_n = F$ in the Levy metric implies that

$$\lim_{n\to\infty}\int dF_n(\eta)S(\rho(\eta)) = \int dF(\eta)S(\rho(\eta)) \ .$$

Hence the limit of a sequence of thermal channel Holevo information with sequence encodings $\{F_n\}_{n\in\mathbb{N}}$ converging to F in the Levy metric is

$$\lim_{n\to\infty}\chi_{tc}[F_n] = \lim_{n\to\infty} S(\rho_{F_n}) - \int dF_n(\eta)S(\rho(\eta)) = S(\rho_F) - \int dF(\eta)S(\rho(\eta)) = \chi_{tc}[F] \ .$$

Which implies that χ_{tc} is continuous over $\mathscr{F}$. □

Now we have shown that χ_{tc} is strictly concave, weakly differentiable, and continuous (because von Neumann entropy is continuous) over $\mathscr{F}$, also that the space of circularly symmetric encodings $\mathscr{F}$ is convex and compact. Therefore, χ_{tc} and $\mathscr{F}$ satisfy all conditions of the optimization theorem (Lemma 2) to guarantee the existence of a unique optimal encoding F^*, i.e $\chi_{tc}[F^*] = \sup_F \chi_{tc}[F]$ if and only if $\chi'_{F^*}[F] \le 0$ for all F. In other words, the Holevo function is non-increasing in any direction F at F^*. We will use this fact to show that F^* is optimal iff (26a) and (26b) holds for F^*.

First we will show that if (26a) and (26b) holds for encoding F^*, then F^* must satisfy $\chi'_{F^*}[F] \le 0$ for all F. This can be shown by simply noting that if (26a) holds, then for any encoding F

$$\int dF(\eta)i[\eta, F^*] \le \int dF(\eta)\chi_{tc}[F^*] = \chi_{tc}[F^*] \ . \tag{44}$$

Hence $\chi'_{F^*}[F] \leq 0$ for all F by lemma 4. Now recall the set of points of increase $\mathcal{I}$ (see definition 3). Then by noting that $\int_{\mathcal{I}} dF^*(\eta) = 1$ we have

$$\int dF^*(\eta) i[\eta, F^*] = \int_{\mathcal{I}} dF^*(\eta) i[\eta, F^*] + \int_{[0,\infty)\backslash\mathcal{I}} dF^*(\eta) i[\eta, F^*] = \int_{\mathcal{I}} dF^*(\eta) \chi_{tc}[F^*] = \chi_{tc}[F^*] \; , \quad (45)$$

showing that equality $\chi'_{F^*}[F^*] = 0$ is achieved for encoding F^*.

For the converse, we will show that both (26a) and (26b) must hold for any optimal encoding F^* by showing that if either (26a) or (26b) does not hold, we will get a contradiction. Suppose there exist $\eta^* \in [0, 1]$ such that $i[\eta^*, F^*] > \chi_{tc}[F^*]$. By defining an encoding

$$G(\eta) = \begin{cases} 1, & \text{for } \eta \geq \eta^* \\ 0, & \text{for } \eta < \eta^* \end{cases} \; , \quad (46)$$

we can write $i[\eta^*, F^*] = \int dG(\eta) i[\eta, F^*] > \chi[F^*]$ which contradicts the necessary and sufficient weak derivative condition for optimal F^* that $\chi'_{F^*}[F] \leq 0$ for all F.

Similarly for (26b), suppose that there exist some points of increase $C \subset \mathcal{I}$ such that $\int_C dF^*(\eta) = \delta > 0$ and $i[\eta, F^*] < \chi[F^*]$ for all $\eta \in C$. For the remaining points of increase $\eta \in \mathcal{I}\backslash C$, we have $\int_{\mathcal{I}\backslash C} dF^*(\eta) = 1 - \delta$ and $i[\eta, F^*] = \chi_{tc}[F^*]$. Because $\chi'_{F^*}(F^*) = 0$ and $F^*([0, \infty)\backslash\mathcal{I}) = 0$, we have

$$\chi[F^*] = \int dF^*(\eta) i[\eta, F^*] = \int_{\mathcal{I}\backslash C} dF^*(\eta) i[\eta, F^*] + \int_C dF^*(\eta) i[\eta, F^*] < (1 - \delta)\chi[F^*] + \delta\chi[F^*] = \chi[F^*] \quad (47)$$

giving us a contradiction $\chi[F^*] < \chi[F^*]$, hence both (26a) and (26b) must hold for any optimal encoding F^*. Therefore, concluding the proof of Proposition 2.

B Proof of finite points-of-increase for coherent state resource at temperature $T = 0$

In this appendix, we give a proof of proposition 3. This closely follows the proof for the finiteness of the set of points-of-increase for optimal photon channel by Shamai [19], which uses the Bolzano-Weierstrass theorem and the identity theorem of analytic function to show that infinite points-of-increase implies that the marginal information density $i[\lambda, F^*]$ is equal to the mutual information $I(F^*)$ for all photon channel intensity $\lambda \geq 0$. Shamai then shows that this leads to a contradiction. We adopt the same reasoning to show that infinite points-of-increase implies that $i[\eta, F^*] = \chi[F^*]$ for all positive reals η, which leads to a contradiction. Because $\rho(\eta, 0)$ is a pure state, which implies that its entropy is 0, the marginal information density is

$$i[\eta, F^*] = -\sum_n P_n[\rho(\eta)] \log P_n[F^*] \; . \quad (48)$$

So, to establish the contradiction if the points-of-increase is assumed to has infinite cardinality we show that $i[\cdot, F^*]$ is analytic on positive reals.

Lemma 7. *Marginal information density $i[\eta, F^*]$ for coherent thermal state resource state in zero-temperature environment is analytic for all $\eta \in (0, \infty)$.*

Proof. Marginal information density in the case of direct detection channel

$$i_{dd}[x, F] = -\sum_n e^{-x} \frac{x^n}{n!} \log \frac{\int dF(\tilde{x}) e^{-\tilde{x}} \frac{\tilde{x}^n}{n!}}{e^{-x} \frac{x^n}{n!}}$$

is analytic for all real $x > 0$ as argued in [19]. Setting $x = \eta|\alpha|^2$ we can write

$$i[\eta, F^*] = i_{dd}[\eta|\alpha|^2, F^*] \underbrace{- \sum_n P_n[\rho(\eta)] \log P_n[\rho(\eta)]}_{(1)}$$

where $P_n[\rho(\eta)] = \mathrm{Poi}_n(\eta|\alpha|^2) = e^{-\eta|\alpha|^2}\frac{(\eta|\alpha|^2)^n}{n!}$ therefore term (1) is $-H(\mathrm{Poi}(\eta|\alpha|^2))$, the negative of Shannon entropy of Poisson random variable with mean $\eta|\alpha|^2$. It is shown in [31] that

$$\begin{aligned} H(\mathrm{Poi}(\eta|\alpha|^2)) &= \eta|\alpha|^2(1 - \log \eta|\alpha|^2) + \sum_n P_n[\rho(\eta)] \log n! \\ &= \eta|\alpha|^2(1 - \log \eta|\alpha|^2) + \mathbb{E}_{\mathrm{Poi}(\eta|\alpha|^2)}[\log N!] \end{aligned} \tag{49}$$

where N is a Poisson random variable with mean $\eta|\alpha|^2$. The first term $\eta|\alpha|^2(1 - \log \eta|\alpha|^2)$ is clearly analytic, therefore we only need to show that $\mathbb{E}_{\mathrm{Poi}(\eta|\alpha|^2)}[\log N!]$ is analytic for all $\eta > 0$.

Now for the rest of the proof we will show that the second term of (49) is analytic in the set $\mathbb{C}_\delta = \{z \in \mathbb{C} : \mathrm{Re}(z) > \delta \wedge |\mathrm{Im}(z)| < \delta\}$ for all $\delta > 0$ by showing that the sum is uniformly convergent in any closed disk in $\mathbb{C}_\delta$. Its analyticity over $\mathbb{R}_{>0}$ then follows because it is analytic over all $\mathbb{C}_\delta$ for $\delta > 0$ and because $\mathbb{R}_{>0} \subseteq \mathbb{C}_\delta$.

Let us now extend $P_n[\rho(\eta)]$ to all $\eta \in \mathbb{C}_\delta$ for some small $\delta > 0$, which is an open subset of $\mathbb{C}$ containing all real numbers in (δ, ∞). Now let $\eta = x + iy$ and consider

$$|P_n[\rho(\eta)]| = \left| e^{-\eta|\alpha|^2} \frac{(\eta|\alpha|^2)^n}{n!} \right| \tag{50}$$

$$= e^{-x|\alpha|^2} \frac{(\sqrt{x^2+y^2}|\alpha|^2)^n}{n!} \tag{51}$$

$$\leq e^{-x|\alpha|^2} \frac{(x|\alpha|^2)^n + (|y||\alpha|^2)^n}{n!} = P_n[\rho(x)] + e^{-x|\alpha|^2} \frac{(|y||\alpha|^2)^n}{n!} \,. \tag{52}$$

So we can now upper bound the modulus of the expected value of $\log N!$ for the extended Poisson distribution function $P_n[\rho(\eta)]$

$$\begin{aligned} |\mathbb{E}_{\mathrm{Poi}(\eta|\alpha|^2)}[\log N!]| &= \Big| \sum_n P_n[\rho(\eta)] \log n! \Big| \\ &\leq \sum_n |P_n[\rho(\eta)]| \log n! \\ &\leq \mathbb{E}_{\mathrm{Poi}(x|\alpha|^2)}[\log N!] + e^{-x|\alpha|^2} \sum_n \frac{(|y||\alpha|^2)^n}{n!} \log n! \\ &< 2\mathbb{E}_{\mathrm{Poi}(x|\alpha|^2)}[\log N!] \end{aligned}$$

where the last inequality is because $|y| < \delta < x$ for any $\eta = x + iy \in \mathbb{C}_\delta$.

Now we have $\log n! \leq \log n^n = n \log n \leq n(n-1)$, and therefore by considering a closed disk $D_r(z_0) = \{z \in \mathbb{C}_\delta : |z - z_0| \leq r\}$ with $z_0 \in \mathbb{C}_\delta$ and $r > 0$ and ξ the maximum of the real part of any $z \in D_r(z_0)$ we get upper bound

$$\mathbb{E}_{\mathrm{Poi}(x|\alpha|^2)}[\log N!] \leq \mathbb{E}_{\mathrm{Poi}(x|\alpha|^2)}[N \log N] \leq \mathbb{E}_{\mathrm{Poi}(x|\alpha|^2)}[N(N-1)] \leq (x|\alpha|^2)^2 \leq (\xi|\alpha|^2)^2 \,.$$

Now we will show that $\mathbb{E}_{\mathrm{Poi}(\eta|\alpha|^2)}[\log N!]$ uniformly converges in any closed disk $D_r(z_0) = \{z \in \mathbb{C}_\delta : |z - z_0| \leq r\}$ with $z_0 \in \mathbb{C}_\delta$ and $r > 0$ such that $D_r(z_0) \subseteq \mathbb{C}_\delta$. Using a standard theorem of complex analysis (see e.g. [32, Theorem 3.1.8]) this implies that $\mathbb{E}_{\mathrm{Poi}(\eta|\alpha|^2)}[\log N!]$ is analytic in $\mathbb{C}_\delta$ because each of the term in the sum is analytic.

To show uniform convergence of $\mathbb{E}_{\mathrm{Poi}(\eta|\alpha|^2)}[\log N!]$ in $D_r(z_0)$ we will use the Weierstrass M-test (see e.g. [32, Theorem 3.1.7]) which states that f_n uniformly converges to f in some set A if there is some real numbers $\{M_n\}_n$ such that $\sum_n M_n < \infty$ and $|f_n(z)| \leq M_n$ for all n and all $z \in A$. Let

$\xi = \max\{\mathrm{Re}(z) : z \in D_r(z_0)\}$ (i.e. the largest real part of complex numbers in disk $D_r(z_0)$). By (52) we know that $|P_n[\rho(\eta)]| \log n! \leq 2P_n[\rho(x)] \log n!$ and for sufficiently large N we have $P_n[\rho(x)] \leq P_n[\rho(\xi)]$ for all $n > N$. So, let

$$M_n = \begin{cases} 2P_n[\rho(\xi)] \log n! & \text{if } n > N \\ (\xi|\alpha|^2)^2 & \text{if } n \leq N \end{cases} .$$

Given that $\mathbb{E}_{\mathrm{Poi}(x|\alpha|^2)}[\log N!] \leq (\xi|\alpha|^2)^2$ implies that $\sum_{n>N} M_n \leq 2(\xi|\alpha|^2)^2$ and $\sum_{n\leq N} M_n = N(\xi|\alpha|^2)^2$ we have $\sum_n M_n \leq (N+1)(\xi|\alpha|^2)^2 < \infty$ with $|P_n[\rho(\eta)]| \leq M_n$ for all n.

Hence, since the second term of (49) uniformly converges in any closed disk $D_r(z_0) \subseteq \mathbb{C}_\delta$ and each of its term $P_n[\rho(\eta)] \log n!$ is analytic in $\mathbb{C}$, it is analytic in $\mathbb{C}_\delta$. Consequently the restriction of $i[\cdot, F^*]$ over (δ, ∞) is analytic for any $\delta > 0$, concluding the proof. □

Note that the transmissivity can only physically takes value in $[0, 1]$, nevertheless because $i[\eta, F^*]$ is analytic on $(0, \infty)$, a contradiction will follow from assuming that there are infinite POIs. First we consider a lower bound

$$\begin{aligned} i[\eta, F^*] &= -\sum_n P_n[\rho(\eta)] \log\Big(\int dF^*(\tilde\eta) e^{-\tilde\eta|\alpha|^2} \frac{(\tilde\eta|\alpha|^2)^n}{n!}\Big) \\ &\overset{(1)}{\geq} \sum_n e^{-\eta|\alpha|^2} \frac{(\eta|\alpha|^2)^n}{n!} \Big(n \log \frac{1}{|\alpha|^2} + \log n!\Big) \\ &\overset{(2)}{=} \eta|\alpha|^2 \log \frac{1}{|\alpha|^2} + \sum_n e^{-\eta|\alpha|^2} \frac{(\eta|\alpha|^2)^n}{n!} \log n! \; , \end{aligned} \tag{53}$$

where (1) is because $\log \int dF^*(\tilde\eta) e^{-\tilde\eta|\alpha|^2} \frac{(\tilde\eta|\alpha|^2)^n}{n!} \leq n \log |\alpha|^2 - \log n!$ as the integral is only over $[0, 1]$, and (2) is because $\sum_n \frac{(\eta|\alpha|^2)^n}{n!} n = \eta|\alpha|^2 e^{\eta|\alpha|^2}$.

The second term in (53) is an expectation of $\log N!$ where N is a Poisson random variable with mean $\eta|\alpha|^2$

$$\mathbb{E}_{\mathrm{Poi}(\eta|\alpha|^2)}[\log N!] = \sum_n e^{-\eta|\alpha|^2} \frac{(\eta|\alpha|^2)^n}{n!} \log n! \; . \tag{54}$$

By Stirling's approximation for log of factorials we can use big omega notation to obtain lower bound $\log n! \geq \Omega(n \log n)$, so

$$\mathbb{E}_{\mathrm{Poi}(\eta|\alpha|^2)}[\log N!] \geq \Omega\Big(\mathbb{E}_{\mathrm{Poi}(\eta|\alpha|^2)}[N \log N]\Big) \tag{55}$$

$$\geq \Omega\Big(\mathbb{E}_{\mathrm{Poi}(\eta|\alpha|^2)}[N] \log(\mathbb{E}_{\mathrm{Poi}(\eta|\alpha|^2)}[N])\Big) = \Omega(\eta|\alpha|^2 \log \eta|\alpha|^2) \tag{56}$$

as the Poisson-distributed expectation $\mathbb{E}_{\mathrm{Poi(z)}}[N] = z$. Hence applying this lower bound to (53) gives

$$i[\eta, F^*] \geq \eta|\alpha|^2 \log \frac{1}{|\alpha|^2} + \mathbb{E}_{\mathrm{Poi}(\eta|\alpha|^2)}[\log N!] \geq \Omega\Big(\eta|\alpha|^2 \log \eta\Big)$$

which implies that $i[\eta, F^*]$ grows arbitrarily large as $\eta \to \infty$. Hence for any resource energy $|\alpha|^2$ and any optimal thermal encoding F^*, an infinite points-of-increase leads to a contradiction as Holevo information $\chi_{tc}[F^*] = i[\eta, F^*]$ goes to infinity as $\eta \to \infty$, so the cardinality of the points-of-increase must be finite.

C Threshold when one ring is no longer optimal

Here for ring state at amplitude $\alpha_{\max}$, $\rho_\alpha = \int_{-\pi}^{\pi} \frac{d\theta}{2\pi} R_\theta \left|\alpha\middle\rangle\!\middle\langle\alpha\right| R_\theta^\dagger$ and vacuum state $\rho_0 = \left|0\middle\rangle\!\middle\langle 0\right|$ we will solve for α that satisfies

$$\frac{\partial}{\partial\epsilon} S(\epsilon\rho_0 + (1-\epsilon)\rho_\alpha)\Big|_{\epsilon=0} = 0 \tag{57}$$

which can be written more explicitly in terms of limit

$$\lim_{\epsilon\to 0} \frac{S(\rho_\alpha + \epsilon(\rho_0 - \rho_\alpha)) - S(\rho_\alpha)}{\epsilon} = 0 \ . \tag{58}$$

Using the Taylor expansion of $S(A+\epsilon B)$ for two commuting matrices A and B,

$$\begin{aligned} S(A+\epsilon B) &= \frac{1}{\ln 2}\mathrm{Tr}\left[(A+\epsilon B)\ln(A+\epsilon B)\right] && (59)\\ &= \frac{1}{\ln 2}\mathrm{Tr}\left[(A+\epsilon B)\left(\ln A + \frac{\epsilon B}{A}\right)\right] && (60)\\ &= S(A) + \epsilon \mathrm{Tr} B\log A + \frac{\epsilon}{\ln 2}\mathrm{Tr} B \ , && (61) \end{aligned}$$

we get

$$\begin{aligned} \mathrm{Tr}\left[(\rho_0 - \rho_\alpha)\log\rho_\alpha\right] + \frac{1}{\ln 2}\mathrm{Tr}\left[\rho_0 - \rho_\alpha\right] &= 0 && (62)\\ \mathrm{Tr}\left[\rho_0 \log\rho_\alpha\right] + S(\rho_\alpha) &= 0 && (63)\\ \log e^{-|\alpha|^2} + S(\rho_\alpha) &= 0 && (64)\\ \frac{|\alpha|^2}{\ln 2} &= S(\rho_\alpha) && (65) \end{aligned}$$

as claimed. The solution to this is $|\alpha| = 1.25034$.

D Bounds on the capacity for coherent state resource at $T = 0$ and average energy constraint

A lot of work have been done to derive bounds for capacities of classical channels. For example, see [33, 34, 35] for bounds for the AWGN channels and [35] for quadrature modulations and measurements. Lapidoth derives bounds for a discrete time Poissonian channel [36] and intensity modulation with additive Gaussian noise [37]. There are also bounds on the number of rings for the n-dimensional AWGN channel by Dytso [22, 23]. In this section, we will present the upper and lower bounds for the Holevo information χ when we have a coherent state $|\alpha_{\max}\rangle$ resource with environment temperature $T = 0$.

D.1 Upper bounds

If we relax the channel by allowing all encoding F such that the average energy (i.e: energy of ρ_{ave}) to be bounded by $E_{\max} = |\alpha_{\max}|^2$ instead bounding the energy of individual codewords $\rho(\eta,\theta)$, we may use the result by Giovannetti, et al. [5] to get an upper bound to Holevo information χ. For the average-energy constrained channel, χ is attained by a Gaussian distribution encoding and is given in the supplemental material of [5]

$$\begin{aligned} \chi_{\mathrm{GDE}}(\alpha_{\max}) &= (E_{\max}+1)\log(E_{\max}+1) - E_{\max}\log E_{\max}\\ &= -E_{\max}\log_2\frac{E_{\max}}{1+E_{\max}} + \log_2(1+E_{\max}) \ . && (66) \end{aligned}$$

D.2 Lower bounds

While the classical mutual information gives a lower bound to its quantum analogue χ, this lower bound can be obtained by a single-ring encoding and an encoding correspond to a "flat" distribution.

D.2.1 Single-ring encoding

We can obtain a lower bound by considering a distribution with just one ring at $\alpha_{\max}$. For $E_{\max} = |\alpha_{\max}|^2$, this is given by the von Neumann entropy of the ring state at amplitude $\alpha_{\max}$, which is a uniformly random phase-shifted mixture of $|\alpha_{\max}\rangle$

$$
\begin{aligned}
L_{1\text{ring}}(\alpha_{\max}) &= \chi[F_{1\text{ring}}] && (67)\\
&= S\left(\int_{-\pi}^{\pi} \frac{d\theta}{2\pi} R_\theta \, |\alpha_{\max}\rangle\langle\alpha_{\max}| \, R_\theta^\dagger\right) && (68)\\
&= -\sum_{n=0}^{\infty} \frac{e^{-E_{\max}} E_{\max}^n}{n!} \log_2 \frac{e^{-E_{\max}} E_{\max}^n}{n!} \, . && (69)
\end{aligned}
$$

D.2.2 Flat encoding

Consider an encoding that uniformly distributes the attenuation, i.e. $F(\eta) = \eta$. This distribution will (by construction) has a flat Wigner function in the middle and drops off at the edge (Figure 11). Curiously, the photon number distribution is also rather flat at low photons numbers when energy $E_{\max}$ is large enough (see Figure 11). This encoding has a Fock state distribution

$$
\begin{aligned}
P_n[F] &= \int dF(\eta) \, P_n[\alpha_{\max}(\eta)] && (70)\\
&= \int d\eta \, e^{-\eta E_{\max}} \frac{(\eta E_{\max})^n}{n!} && (71)\\
&= \frac{1}{E_{\max}} - \frac{1}{n! E_{\max}} \Gamma(1+n, E_{\max}) \, . && (72)
\end{aligned}
$$

The lower bound (which we denote as L_{flat}) can then be computed by taking the entropy of this distribution. This bound becomes more informative than $L_{1\text{ring}}$ as $E_{\max}$ increases. These bounds and the numerically computed capacity are plotted in Figure 4(b). The information per unit energy diverges as the input $E_{\max}$ tends to zero.

Now consider the Poisson distribution parameter of $P_n[\alpha_{\max}(\eta)]$, which is $\eta E_{\max}$ and denote a random variable $\mathcal{X}$ for such parameter. Such random variables have been studied in the classical context of discrete-time Poisson channels where the communication channel consists of a coherent state intensity modulation followed by direct detection at the receiver [36]. Proposition 11 of [36] gives the lower bound on the entropy of such channel

$$
H \geq h[\mathcal{X}] + (1 + \mathbb{E}[\mathcal{X}]) \log\left(1 + \frac{1}{\mathbb{E}[\mathcal{X}]}\right) - \frac{1}{\ln 2} \, , \tag{73}
$$

where $h[\mathcal{X}]$ is the differential entropy of $\mathcal{X}$ and $\mathbb{E}[\mathcal{X}]$ is the mean of $\mathcal{X}$. Using this result on the flat distribution $F(\eta) = \eta$ for $0 \leq \eta \leq 1$, we could have a uniform encoding for random variable $\mathcal{X}$, $F(x) = x/E_{\max}$ so that $h[\mathcal{X}] = \log E_{\max}$ and $\mathbb{E}[\mathcal{X}] = E_{\max}/2$. This gives a lower bound for the capacity L_{flat} for flat encoding

$$
L_{\text{flat}} \geq \log E_{\max} + \left(1 + \frac{E_{\max}}{2}\right) \log\left(1 + \frac{2}{E_{\max}}\right) - \frac{1}{\ln 2} \, . \tag{74}
$$

E Proof of finite points-of-increase for thermal state resource

In this appendix we give a proof of proposition 4. Here we employ similar argument to the case of coherent state input in Appendix B where we show that the marginal information density function for the thermal state is analytic for all positive reals then use the identity theorem of analytic functions along with the Bolzano-Weierstrass theorem to obtain $i[\eta, F^*] = \chi_{tc}[F^*]$ for all positive real η from assuming that the points-of-increase is infinite. We again show that this leads to a contradiction to conclude that the assumption of the cardinality of the set of points-of-increase being infinite cannot be true. Here we assume that both mean photon number of the input state n_{res} and mean photon number of the environment n_{env} are strictly larger than zero. We also assume that $n_{\text{res}} \neq n_{\text{env}}$ because otherwise we have $\rho(\eta) = \rho_{\text{ave}}$ for all η hence we cannot encode any information as reflected by $\chi_{tc}[F] = \int dF(\eta) S(\rho(\eta)||\rho_{\text{ave}}) = 0$ for any choice of encoding F. For simplicity, the above assumptions will be taken into account in the proofs below, but we will give an argument later that the same technique can still be used in the case that either n_{res} or n_{env} is zero (i.e. vacuum).

Noting that the $S(\rho(\eta, 0))$ term in the marginal information density (4) is the entropy of thermal state with mean photon number $n_\eta = \eta n_{\text{res}} + (1-\eta) n_{\text{env}}$, we can rewrite it as

$$i[\eta, F^*] = \sum_k P_k[\rho(\eta)] \log \frac{1}{P_k[F^*]} - \Big((n_\eta + 1)\log(n_\eta + 1) - n_\eta \log n_\eta \Big) \tag{75}$$

where $P_k[F^*] = \int dF^*(\eta') P_k[\rho(\eta')]$ and

$$P_k[\rho(\eta)] = \frac{n_\eta^k}{(n_\eta + 1)^{k+1}} \, . \tag{76}$$

Now we will show that $i[\cdot, F^*]$ is analytic over all positive reals.

Lemma 8. *Marginal information density $i[\eta, F^*]$ for thermal state input is analytic for all $\eta \in (0, \infty)$.*

Before proving Lemma 8, we first need the following technical lemma.

Lemma 9. *Let $Q_k(z) = \frac{z^k}{(z+1)^{k+1}}$ and $n_{\max} = \max\{n_{\text{res}}, n_{\text{env}}\}$ and $n_{\min} = \min\{n_{\text{res}}, n_{\text{env}}\}$. The following inequality holds for all circularly symmetric encoding F and all $k > n_{\max}$*

$$\log \frac{1}{P_k[F]} \leq \log \frac{1}{Q_k(n_{\min})} = \log(n_{\min} + 1) + k \log \frac{n_{\min} + 1}{n_{\min}} \, .$$

Proof. We can write $P_k[F] = \int_0^1 dF(\eta) Q_k(n_\eta)$. Note that the derivative of Q_k is

$$\frac{d}{dz} Q_k(z) = (k - z) z^{k-1} (z+1)^{-k-2} \, .$$

So by solving $\frac{d}{dz} Q_k(z) = 0$ for z we can find that Q_k achieves its maximum at $z = k$. Moreover, one can observe from the $(k - z)$ term that Q_k is monotonically increasing for $z < k$ and monotonically decreasing for $z > k$ (i.e. $\forall z' < k, z' > z \Rightarrow Q_k(z') > Q_k(z)$ and $\forall z > k, z' > z \Rightarrow Q_k(z') < Q_k(z)$). Therefore

$$P_k[F] = \int_0^1 dF(\eta) Q_k(n_\eta) \geq Q_k(n_{\min})$$

if $k > n_{\max}$. This implies that for all $k > n_{\max}$ we have $\frac{1}{P_k[F]} \leq \frac{1}{Q_k(n_{\min})}$, and thus

$$\log \frac{1}{P_k[F]} \leq \log \frac{1}{Q_k(n_{\min})} = \log \frac{(n_{\min}+1)^{k+1}}{n_{\min}^k} = \log(n_{\min} + 1) + k \log \frac{n_{\min} + 1}{n_{\min}}$$

because log is a monotonically increasing function. □

Now we are set to show the analiticity of the marginal information density.

proof of Lemma 8. First, let $h[\eta, F^*] = \sum_{k=0}^{\infty} P_k[\rho(\eta)] \log \frac{1}{P_k[F^*]}$ so that

$$i[\eta, F^*] = h[\eta, F^*] - \Big((n_\eta + 1)\log(n_\eta + 1) - n_\eta \log n_\eta\Big) . \tag{77}$$

The second term is clearly analytic as it is a finitely many compositions of log functions, multiplications, and additions over $z \in \mathbb{R}_{>0}$, hence it remains to show that $h[\cdot, F^*]$ is analytic over $\mathbb{R}_{>0}$. Again by considering $\mathbb{C}_\delta = \{z \in \mathbb{C} : \mathrm{Re}(z) > \delta \wedge |\mathrm{Im}(z)| < \delta\}$, we will show that $h[\cdot, F^*]$ is analytic over $\mathbb{C}_\delta$ for any $\delta > 0$ by showing that $h[\cdot, F^*]$ converges uniformly in any closed disk $D_r(z_0) = \{z \in \mathbb{C}_\delta : |z - z_0| \leq r\} \subseteq \mathbb{C}_\delta$. Denoting $h_N[\eta, F^*] = \sum_{k>N} P_k[\rho(\eta)] \log \frac{1}{P_k[F^*]}$, it suffices to show that for any $\varepsilon > 0$ there exists $M \in \mathbb{N}$ such that for all $N \geq M$, $|h_N[\eta, F^*]| < \varepsilon$ for all $\eta \in D_r(z_0)$ because the partial sum $h[\eta, F^*] - h_N[\eta, F^*]$ is a composition of additions, multiplications, and log functions, hence is analytic.

By using Lemma 9 we have an upper bound

$$\begin{aligned}
|h_{n_{\max}}[\eta, F^*]| &= \Big| \sum_{k>n_{\max}} P_k[\rho(\eta)] \log \frac{1}{P_k[F^*]} \Big| \\
&\leq \sum_{k>n_{\max}} |P_k[\rho(\eta)]| \Big(\log(n_{\min} + 1) + k \log \frac{n_{\min} + 1}{n_{\min}} \Big) \\
&= \frac{1}{|n_\eta + 1|} \sum_{k>n_{\max}} \frac{|n_\eta|^k}{|n_\eta + 1|^k} \Big(\log(n_{\min} + 1) + k \log \frac{n_{\min} + 1}{n_{\min}} \Big) \\
&= \frac{\log(n_{\min} + 1)}{|n_\eta + 1|} \sum_{j>n_{\max}} \frac{|n_\eta|^j}{|n_\eta + 1|^j} + \frac{1}{|n_\eta + 1|} \log \frac{n_{\min} + 1}{n_{\min}} \sum_{k>n_{\max}} \frac{|n_\eta|^k}{|n_\eta + 1|^k}\, k .
\end{aligned} \tag{78}$$

From the series identity $\sum_{k=0}^{\infty} k c^k = \frac{c}{(c-1)^2}$ we have

$$\sum_{k=0}^{\infty} \frac{|n_\eta|^k}{|n_\eta + 1|^k}\, k = \frac{|n_\eta||n_\eta + 1|}{(|n_\eta| - |n_\eta + 1|)^2} ,$$

and for any $N \in \mathbb{N}$, the partial series identity $\sum_{k=0}^{N} k c^k = \frac{(cN - N - 1)c^{N+1} + c}{(1-c)^2}$ gives

$$\begin{aligned}
\sum_{k=0}^{N} \frac{|n_\eta|^k}{|n_\eta + 1|^k}\, k &= \frac{|n_\eta + 1|^2}{(|n_\eta + 1| - |n_\eta|)^2} \Big(\Big(\frac{|n_\eta|}{|n_\eta + 1|} N - N - 1 \Big) \Big(\frac{|n_\eta|}{|n_\eta + 1|} \Big)^{N+1} + \frac{|n_\eta|}{|n_\eta + 1|} \Big) \\
&= \frac{|n_\eta + 1||n_\eta|}{(|n_\eta + 1| - |n_\eta|)^2} \Big(\frac{|n_\eta|}{|n_\eta + 1|} \Big)^N \Big(\frac{|n_\eta|}{|n_\eta + 1|} N - N - 1 \Big) + \frac{|n_\eta + 1||n_\eta|}{(|n_\eta + 1| - |n_\eta|)^2} .
\end{aligned}$$

Hence we get

$$\sum_{k>N} \frac{|n_\eta|^k}{|n_\eta + 1|^k}\, k = \frac{|n_\eta + 1||n_\eta|}{(|n_\eta + 1| - |n_\eta|)^2} \Big(\frac{|n_\eta|}{|n_\eta + 1|} \Big)^N \Big(N + 1 - \frac{|n_\eta|}{|n_\eta + 1|} N \Big) , \tag{79}$$

which goes to 0 as $N \to \infty$ because $\frac{|n_\eta|}{|n_\eta+1|} < 1$. Clearly $\sum_{j>N} \frac{|n_\eta|^j}{|n_\eta+1|^j}$ also goes to 0 as $N \to \infty$ because $(\frac{|n_\eta|}{|n_\eta+1|})^k \leq k(\frac{|n_\eta|}{|n_\eta+1|})^k$ for all $k \in \mathbb{N}$. So by setting $N = n_{\max}$ and then plugging it into (78) we get the bound

$$\Big| h_{n_{\max}}[\eta, F^*] \Big| \leq \frac{2\log(n_{\min} + 1) - \log n_{\min}}{|n_\eta + 1|} \sum_{k>n_{\max}} \frac{|n_\eta|^k}{|n_\eta + 1|^k}\, k . \tag{80}$$

Now consider a disk $D_r(z_0)$ and a complex number in it $\xi = \arg\max_{\eta \in D_r(z_0)} \sum_{k>N} \frac{|n_\xi|^k}{|n_\xi+1|^{k+1}}$. This can be obtained by setting N sufficiently large $N > \max\{n_{\Re(\eta)+\delta} + 1 : \eta \in D_r(z_0)\}$. So for all $\eta \in D_r(z_0)$ we can find an integer M larger than such N and larger than $n_{\max}$ to obtain

$$\left| h_M[\eta, F^*] \right| \leq \frac{2\log(n_{\min}+1) - \log n_{\min}}{|n_\xi + 1|} \sum_{k>M} \frac{|n_\xi|^k}{|n_\xi+1|^k}\, k \ . \tag{81}$$

Because the right hand side goes to 0 as $M \to \infty$, for any $\varepsilon > 0$ we can find sufficiently large M such that for all $l > M$ we have $|h_l[\eta, F^*]| < \varepsilon$ for all $\eta \in D_r(z_0)$. As this holds for any closed disk in $\mathbb{C}_\delta$ for any $\delta > 0$, therefore $h[\cdot, F^*]$ is analytic over all positive reals $\mathbb{R}_{>0}$. $\square$

Note that the analiticity of the marginal information density function still holds even if one of $n_{\text{res}}, n_{\text{env}}$ is zero by the following argument. Without loss of generality assume that $n_{\text{env}} = 0$ (i.e. vacuum environment) and suppose that the optimal encoding is F^*, hence n_η can only takes up value in $[0, n_{\text{res}}]$. In the above's proof this leads to a pathological case where $n_{\min} = n_{\text{env}} = 0$. However we can still find a $z \in (0, n_{\text{res}})$ such that $Q_k(z) \leq \int_0^1 dF^*(\eta) Q_k(n_\eta) = P_k[F^*]$ for all $k > n_{\text{res}}$ because Q_k is monotonically increasing and it cannot be the case that $dF^*(\eta) = 1$ for $\eta = 0$ (or any other $\eta \in [0,1]$ for that matter), namely where the only codeword is the vacuum state of the environment as this leads to zero capacity. One can see this by noting that $dF(0) = 1$ implies that $\chi = S(\rho_{\text{ave}}) - \int dF(\eta) S(\rho(\eta)) = S(\rho(0)) - S(\rho(0)) = 0$ because phase shift operation by the channel leaves the output state invariant for a thermal state input. This means that $P_k[F^*]$ must take value in the interval $(0, Q_k(n_{\text{res}}))$ and therefore for any F^* we can always find some $z > 0$ such that $Q_k(z) < P_k[F^*]$ for all $k > n_{\text{res}}$. Hence, we can still use the technique in the proof of Lemma 8 and Lemma 9 to upper bound the h_N function to obtain the same analiticity result.

Now, as $i[\cdot, F^*]$ is analytic over positive reals by Lemma 8, by the identity theorem of analytic functions and the Bolzano-Weierstrass theorem, infinite points of increase over $[0,1]$ implies that $i[\eta, F^*] = \chi[F^*]$ for all positive real η. Now we show that this leads to a contradiction, then conclude that it is not possible to have an infinite points-of-increase.

First, consider the case of $n_{\text{res}} < n_{\text{env}}$. As we can write $n_\eta = n_{\text{env}} + \eta(n_{\text{res}} - n_{\text{env}})$, then we have a negative number inside the log function in the $-(n_\eta+1)\log(n_\eta+1) + n_\eta \log n_\eta$ term of $i[\eta, F^*]$ as written in (75) for sufficiently large η. This implies that $i[\eta, F^*] = \chi[F^*]$ is not real for sufficiently large η, which is not possible. On the other hand, for the case of $n_{\text{res}} > n_{\text{env}}$ we have $\lim_{\eta\to\infty} i[\eta, F^*] < 0$ because

$$\lim_{\eta\to\infty} -(1+n_\eta)\log(1+n_\eta) + n_\eta \log n_\eta = \lim_{\eta\to\infty} -\log(1+n_\eta) - n_\eta \log \frac{1+n_\eta}{n_\eta} < 0$$

because $\lim_{\eta\to\infty} n_\eta = \infty$ and $\lim_{x\to\infty} x \log \frac{1+x}{x} = 1$, and because $\lim_{\eta\to\infty} P_n[\rho(\eta)] = 0$ for each $n \in \mathbb{N}$, implying that $-\sum_n P_n[\rho(\eta)] \log P_n[F^*] = 0$. As an infinite points-of-increase over $[0,1]$ leads to contradictions, therefore there can only be finitely many points-of-increase.

F Capacity of two-codeword encoding with a vacuum resource at $T > 0$.

Here we will derive equation (16), which is the Holevo information over two-codeword encodings when the resource state is a vacuum state and the environment temperature is $T > 0$ (i.e. $n_{\text{env}} > 0$). For convenience, we restate it below

$$\chi = \log\left(1 + \frac{n_{\text{env}}}{(1+n_{\text{env}})^{\frac{1+n_{\text{env}}}{n_{\text{env}}}}}\right) \tag{82}$$

where thermal photon number n_{env} is equal to energy E_0, which is achieved when encoding $F(\eta)$ has two points of increase at $\eta = 0$ and $\eta = 1$. We denote the prior probabilities of sending the thermal state as q_0 (namely when transmissivity $\eta = 0$) and of sending the vacuum state as q_1 (when transmissivity $\eta = 1$) with $q_0 + q_1 = 1$. So, we define an encoding

$$F_{q_0}(\eta) = \begin{cases} q_0, & \text{for } 0 \leq \eta < 1 \ , \\ 1, & \text{for } \eta = 1 \end{cases} . \tag{83}$$

We want to find q_0 and q_1 that maximise the Holevo quantity for such an encoding,

$$\chi[F_{q_0}] = S(\rho_{\text{ave}}) + q_0 S(\rho(0)) + q_1 \underbrace{S(\rho(1))}_{0} \tag{84}$$

where $S(\rho(1)) = 0$ because $\rho(1)$ is a vacuum state, which is a pure state. The averaged state is diagonal in the Fock basis with

$$P_n[\rho_{\text{ave}}] = \begin{cases} q_0 P_0[\rho(0)] + q_1, & \text{for } n = 0 \ , \\ q_0 P_n[\rho(0)], & \text{for } n \geq 1 \ . \end{cases} \tag{85}$$

By direct computation, we have

$$S(\rho_{\text{ave}}) = q_0 S(\rho(0)) - (q_0 P_0[\rho(0)] + q_1) \log (q_0 P_0[\rho(0)] + q_1) + (q_0 P_0[\rho(0)]) \log (q_0 P_0[\rho(0)]) - q_0 \log q_0 \ , \tag{86}$$

which gives

$$\chi[F_{q_0}] \ = \ -(q_0 P_0[\rho(0)] \ + \ q_1) \log (q_0 P_0[\rho(0)] \ + \ q_1) \ + \ (q_0 P_0[\rho(0)]) \log (q_0 P_0[\rho(0)]) \ - \ q_0 \log q_0 \ . \tag{87}$$

To find the maximum value of χ, we differentiate it with respect to q_0 and solve for q_0 in

$$\frac{d\chi}{dq_0} = 0 \ . \tag{88}$$

The value of q_0 that solves this equation is

$$q_0 = \frac{1}{1 - P_0[\rho(0)] + P_0[\rho(0)]^{\left(\frac{P_0[\rho(0)]}{P_0[\rho(0)]-1}\right)}} \tag{89}$$

$$= \frac{1 + n_{\text{env}}}{n_{\text{env}} + (1 + n_{\text{env}})^{\frac{1+n_{\text{env}}}{n_{\text{env}}}}} \ , \tag{90}$$

where we used $P_0[\rho(0)] = \frac{1}{1+n_{\text{env}}}$ to arrive at the last line. Substituting this into (87) gives the desired result. As $n_{\text{env}} \to \infty$, the optimal q_0 tends to 0.5 and $\chi \to 1$.

G Lossy channel with low-energy coherent state resource

Here, we consider a coherent state resource $|\alpha\rangle$ which passes through a lossy channel, i.e. with an environment mean photon number $n_{\text{env}} > 0$ and a fixed attenuation η_{ch}. We will derive an analytic expression of an approximation of the Holevo information (20) in terms of attenuation η_{ch} and thermal photon number $n_{\text{ch}} = (1-\eta_{\text{ch}})n_{\text{env}}$ for a single-ring circularly symmetric encoding given a small energy $E = |\alpha|^2 \ll 1$.

Note that this is equivalent to a scenario where we have a thermal state with mean photon number n_{ch} displaced by α as a resource state and an environment in thermal state with the same mean photon number n_{ch}. We can disregard the phase θ and simply write (17) as

$$\rho(\eta_{\text{ch}}) = D(\sqrt{\eta_{\text{ch}}}\alpha)\rho_{\text{th}}(n_{\text{ch}})D^{\dagger}(\sqrt{\eta_{\text{ch}}}\alpha) \ . \tag{91}$$

because the entropy of $\rho(\eta_{\rm ch},\theta)$ is equal to the entropy of $\rho_{\rm th}(n_{\rm ch})$. This is due to identical photon number distribution between codeword $\rho(\eta_{\rm ch},\theta)$ and thermal state $\rho_{\rm th}(n_{\rm ch})$, which is given by

$$P_n[\rho(\eta_{\rm ch})] = \frac{1}{1+n_{\rm ch}}\left(\frac{n_{\rm ch}}{1+n_{\rm ch}}\right)^n , \tag{92}$$

and the averaged state $\rho_{\rm ave}$ photon number distribution given by

$$P_n[\rho_{\rm ave}] = \langle n|D(\sqrt{\eta_{\rm ch}}\alpha)\rho_{\rm th}(n_{\rm ch})D^\dagger(\sqrt{\eta_{\rm ch}}\alpha)|n\rangle . \tag{93}$$

Hence we may write the Holevo information of the lossy channel as

$$\chi[E] = S(\rho_{\rm ave}) - S(\rho(\eta_{\rm ch})) . \tag{94}$$

When $|\alpha|$ is small, we can approximate the displacement operator as

$$D(\alpha) \approx 1 + \alpha(a^\dagger - a) + \frac{|\alpha|^2}{2}\left(a^\dagger - a\right)^2 , \tag{95}$$

allowing us to obtain the approximate probability of the average output

$$P_n[\rho_{\rm ave}] \approx P_n + \eta_{\rm ch}|\alpha|^2 nP_{n-1} + \eta_{\rm ch}|\alpha|^2(n+1)^2P_{n+1} - \eta_{\rm ch}|\alpha|^2(1+2n)P_n , \tag{96}$$

where P_n is short for $P_n[\rho(\eta_{\rm ch})]$. Substituting this into (94), we get

$$\chi[E] \approx \eta_{\rm ch}|\alpha|^2\sum_n\left(nP_{n-1} + (n+1)^2P_{n+1} - (1+2n)P_n\right)\log P_n . \tag{97}$$

Finally, substituting Eq. (92) for P_n to the right-hand side, and after some simplification, we arrive at

$$\chi[E] \approx \eta_{\rm ch}|\alpha|^2\log\frac{1+n_{\rm ch}}{n_{\rm ch}} , \tag{98}$$

which is precisely the approximation $\tilde{\chi}_{\rm tc}[E]$ in (20).

The same result is obtained for the optimal resource state (14), restated below for convenience

$$|\phi\rangle = \frac{1}{\sqrt{E+1}}\sum_{n=0}^{\infty}|n\rangle\sqrt{\frac{E^n}{(E+1)^n}} . \tag{99}$$

One way to see this is by considering the displaced squeezed state $D(\alpha)S(r)|0\rangle$ where D is the displacement operator and S is the squeezing operator with $\alpha=\sqrt{E}$ and $r=-(\sqrt{2}-1)E$, hence

$$\begin{aligned} D(\alpha)S(r)|0\rangle &= |0\rangle + |1\rangle\alpha + |2\rangle\frac{|\alpha|^2 - r}{2} && (100)\\ &= |0\rangle + |1\rangle\sqrt{E} + |2\rangle E , && (101)\end{aligned}$$

which is a good approximation for the optimal transmitter codeword given a small E. Transmitting this (now Gaussian state) through the noisy channel, we have a squeezed Gaussian state at the output with amplitude $\sqrt{\eta_{\rm ch}E}$, thermal photon expectation $n_{\rm ch}$, and squeezing factor $-\frac{\eta_{\rm ch}(\sqrt{2}-1)E}{1+2n_{\rm ch}}$. The entropy of each codeword is the same as the entropy of a thermal state with $n_{\rm ch}$ photons. We also find that the photon number distribution of the averaged state is given by (96) with α replaced by $\sqrt{E}$, so that we arrive at same result (98) for the Holevo's quantity. We leave the derivations as an exercise for the interested reader. Note that although the state $|\phi\rangle$ is a resource state that maximizes χ when $T=0$ (see Section 3.3), it is not known whether it maximizes χ at $T>0$ environment nor when the channel is lossy. The fact we established above that χ for both resource states $|\phi\rangle$ and $|\alpha\rangle$ coincide at $E\ll 1$, could however be a good starting point for future work on studying the thermal channel capacity at $T>0$.

G.1 Upper bound on the lossy channel capacity

Here we will proceed to derive an upper bound for χ for the lossy channel scenario which coincides with (20). If each codeword is a Gaussian distribution of coherent states with average energy constraint E, this Gaussian distribution is optimal for a channel with no peak power constraint [5]. At the output of a thermal channel, the averaged state is also a Gaussian state with mean photon number $n_{\text{ave}} = \eta_{\text{ch}} E + n_{\text{ch}}$ and codewords with mean photon number n_{ch}. If we define $\delta_n = P_n[\rho_{\text{ave}}] - P_n[\rho(\eta_{\text{ch}})]$ which is small when E is small, we can approximate $S(\rho_{\text{ave}})$ as

$$S(\rho_{\text{ave}}) = -\sum_n P_n[\rho_{\text{ave}}] \log P_n[\rho_{\text{ave}}] \tag{102}$$

$$= -\sum_n \left(P_n[\rho(\eta_{\text{ch}})] + \delta_n\right) \log\left(P_n[\rho(\eta_{\text{ch}})] + \delta_n\right) \tag{103}$$

$$= -\sum_n \left(P_n[\rho(\eta_{\text{ch}})] + \delta_n\right) \left(\log P_n[\rho(\eta_{\text{ch}})] + \frac{\delta_n}{P_n[\rho(\eta_{\text{ch}})] \ln 2}\right) \tag{104}$$

$$= S(\rho(\eta_{\text{ch}})) - \sum_n \delta_n \log P_n[\rho(\eta_{\text{ch}})] \,. \tag{105}$$

Using (94) and the expression for $P_n[\rho(\eta_{\text{ch}})]$ from (92), the Holevo information is

$$S(\rho_{\text{ave}}) - S(\rho(\eta_{\text{ch}})) = -\sum_n \delta_n \left(n \log \frac{n_{\text{ch}}}{1+n_{\text{ch}}} - \log(1+n_{\text{ch}})\right) \tag{106}$$

$$= \sum_n n\delta_n \log \frac{1+n_{\text{ch}}}{n_{\text{ch}}} \tag{107}$$

$$= (n_{\text{ave}} - n_{\text{ch}}) \log \frac{1+n_{\text{ch}}}{n_{\text{ch}}} \tag{108}$$

$$= \eta_{\text{ch}} E \log \frac{1+n_{\text{ch}}}{n_{\text{ch}}} \,, \tag{109}$$

which gives a tight upper bound for the lossy thermal channel capacity given resource state energy E.

H Wigner functions and photon number distribution plots

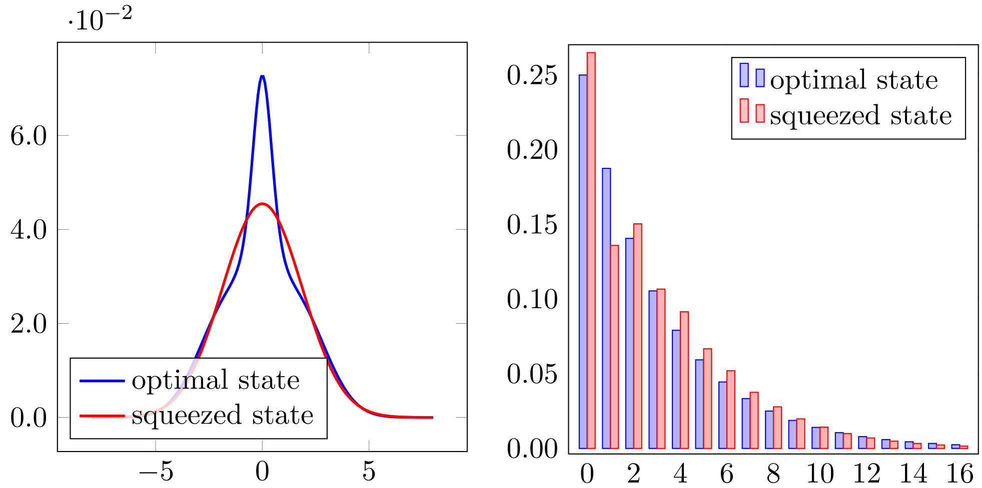

Figure 9: (left) Wigner function of the averaged state with an optimal state in Eq. (14) and a displaced squeezed state resource at $E = 3$. Each displaced squeezed state codeword has 5.40 dB of squeezing and an amplitude of 1.60. (right) Photon number distribution for the average state. The optimal state's Fock state population follows a geometric distribution by construction. The fidelity between the optimal resource and the displaced squeezed state is 0.996.

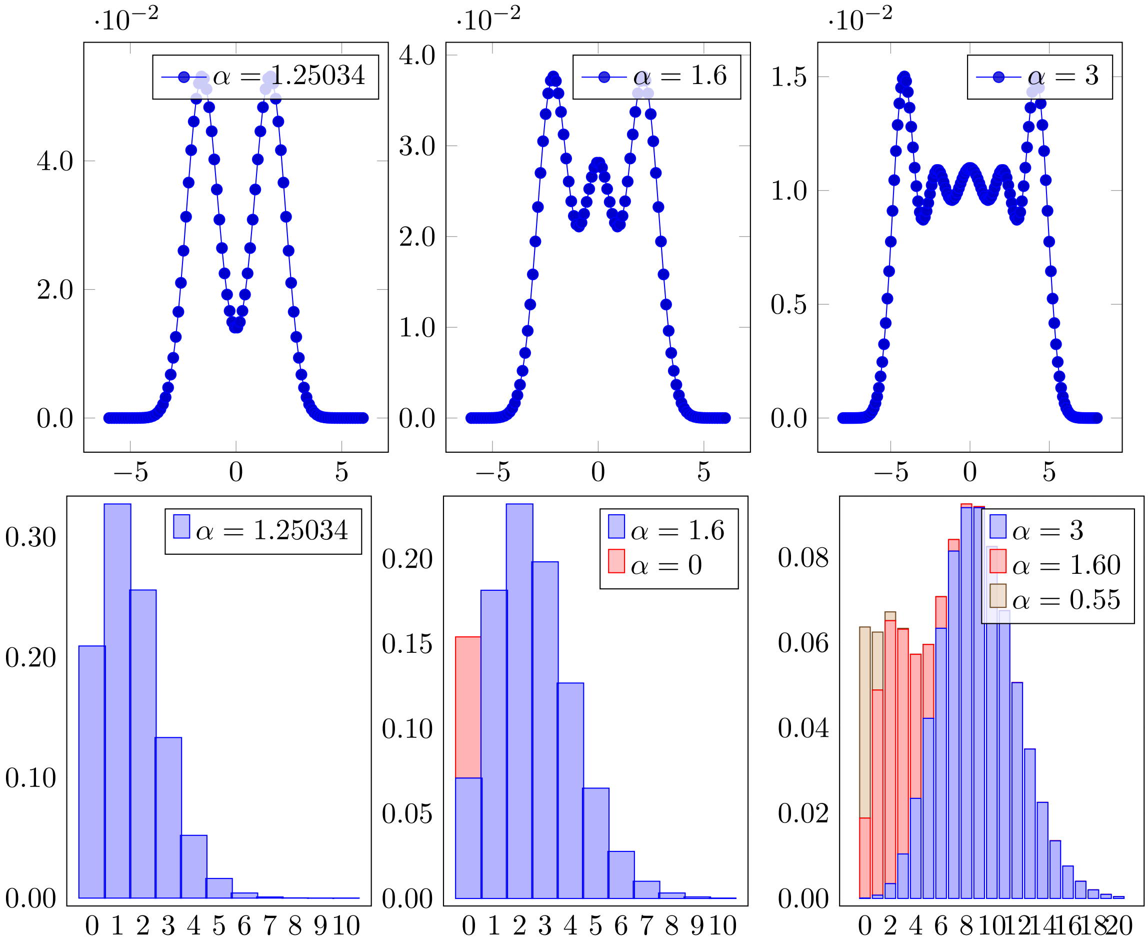

Figure 10: (top) Wigner function for the average state of optimal encoding at $T = 0$. (bottom) Photon number distribution for the average state of optimal encoding.

We plot out the cross-section (at $p = 0$) of the Wigner function $W(x, p)$ for some of the average states in the main text in Fig 10 (top). When $|\alpha_{\max}|$ is less than 1.25034 the optimal encoding is to use one ring at $\eta = 1$. When $|\alpha_{\max}|$ is larger than 1.25034 (but only up to a certain point), the optimal encoding now includes adding a little bit of the vacuum state. As $|\alpha_{\max}|$ increases further, the optimal encoding will include more rings and tends to become flatter. Figure 10 (bottom) shows the corresponding photon number distribution of the average state. However, the flat distribution (Figure 11) is not optimal. It performs worse than the 3-ring distribution.

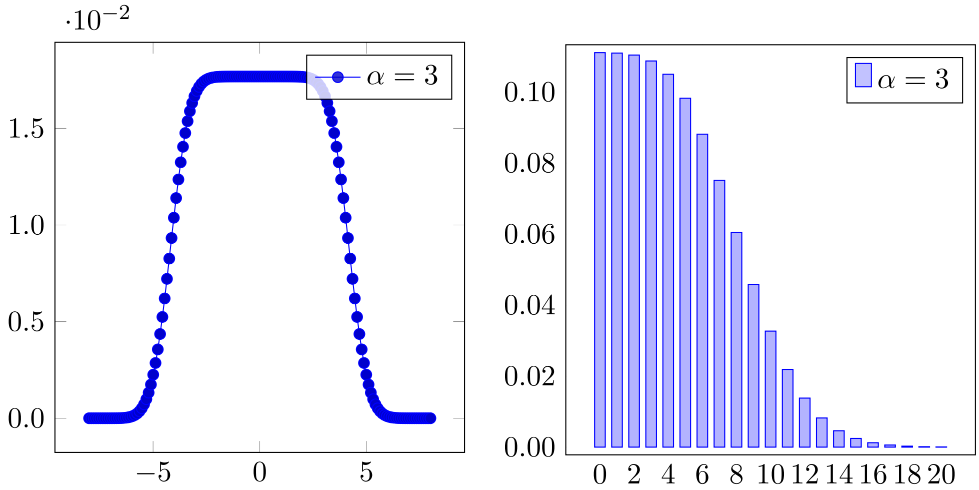

Figure 11: Wigner function (left) and photon number distribution of a flat distribution (with the resource state as the coherent state $|3\rangle$). But this is not optimal and does not perform as well as the 3-ring distribution in Figure 10.