Capacity of optical communication in loss and noise with general quantum Gaussian receivers

Masahiro Takeoka[1] and Saikat Guha[2]
[1]*National Institute of Information and Communications Technology, Koganei, Tokyo 184-8795, Japan*
[2]*Quantum Information Processing Group, Raytheon BBN Technologies, Cambridge, Massachusetts 02138, USA*
(Received 9 February 2014; published 10 April 2014)

Laser-light (coherent-state) modulation is sufficient to achieve the ultimate (Holevo) capacity of classical communication over a lossy and noisy optical channel, but requires a receiver that jointly detects long modulated code words with highly nonlinear quantum operations, which are near-impossible to realize using current technology. We analyze the capacity of the lossy-noisy optical channel when the transmitter uses coherent-state modulation but the receiver is restricted to a general quantum-limited *Gaussian* receiver, i.e., one that may involve arbitrary combinations of Gaussian operations [passive linear optics: beam splitters and phase shifters; second-order nonlinear optics (or active linear optics): squeezers, along with homodyne or heterodyne detection measurements] and any amount of classical feedforward within the receiver. Under these assumptions, we show that the Gaussian receiver that attains the maximum mutual information is either homodyne detection, heterodyne detection, or time sharing between the two, depending upon the received power level. In other words, our result shows that to exceed the theoretical limit of conventional coherent optical communication, one has to incorporate non-Gaussian, i.e., third- or higher-order nonlinear operations in the receiver. Finally we compare our Gaussian receiver limit with experimentally feasible non-Gaussian receivers and show that in the regime of low received photon flux, it is possible to overcome the Gaussian receiver limit by relatively simple non-Gaussian receivers based on photon counting.

DOI: 10.1103/PhysRevA.89.042309 PACS number(s): 03.67.Hk, 42.50.Lc

I. INTRODUCTION

The bosonic channel plays a crucial role in classical and quantum information theory applied to optical communication. Classical capacity of the bosonic channel is particularly important since it determines the ultimate performance limit of conventional optical communication technology. Derivation of this ultimate limit, the Holevo capacity, is a nontrivial theoretical problem on which much effort has been spent [1–11]. If the channel is lossless and noiseless, capacity can be attained by modulating photon number states and an ideal photon number counting receiver [1,2]. For a pure-loss channel, the capacity was found in [3]:

$$C_{\rm loss} = g(\eta\bar{N}), \tag{1}$$

where η is the channel's power transmissivity, $\bar{N}$ is the average input photon number per mode, and $g(x) = (x+1)\log(x+1) - x\log x$ (throughout the paper, we choose $\log \equiv \log_2$). It was also shown [3] that capacity can be attained by a coherent-state modulation, but the optimal receiver must use joint measurements over many channel uses, which are very hard to realize physically [4–7]. For lossy bosonic channel with additive thermal noise from the environment, it was conjectured [8] that the capacity is given by

$$C_{\rm thermal} = g[\eta\bar{N} + (1-\eta)N_{\rm th}] - g[(1-\eta)N_{\rm th}], \tag{2}$$

where $N_{\rm th}$ is the average number of noise photons per mode. The above conjecture relied on an unproven minimum output entropy conjecture, which was recently proven [11] to be true, hence confirming that (2) indeed is the capacity of the lossy-noisy bosonic channel, and that it can be achieved using a coherent-state modulation.

One of the practically important things to note is that the capacities $C_{\rm loss}$, and $C_{\rm thermal}$ can be achieved by encoding information in coherent states with a Gaussian distribution, i.e., one does not need to use nonclassical states such as entangled or squeezed states. Unlike the simple transmitter, the receiver must use a decoding strategy that is highly nonclassical. In the proof of achievability in the original Holevo-Schumacher-Westmoreland theorem [12–14], they employed the *square-root measurement* (SRM) over an infinite sequence of signal states (code word). This is in general a *collective* measurement which may include entangling operations between modulation symbols across channel uses and thus is a highly nonclassical operation. Since there is a clear gap between the Holevo capacity $C_{\rm loss}$ and the Shannon capacities of conventional optical receivers [3,6], i.e., homodyne, heterodyne, or direct-detection receivers, a natural question arises: how to design receiver measurements that can achieve a reliable communication rate beyond the Shannon limits of conventional receivers using a receiver that is more practical than the SRM, the mathematical specification of which unfortunately sheds no light on a structured receiver design. This question has been explored theoretically [5,15–17] and experimentally [18,19], but a truly realizable receiver specification that outperforms conventional optical receivers still eludes us.

In this paper, we restrict ourselves to receiver measurements consisting of a limited class of quantum operations, called Gaussian operations, and allow any amount of classical feedback or feedforward (FF) within the receiver. This class of operations can be constructed by combining passive linear optics (a network of beam splitters and phase shifters), second-order nonlinear optical processes (squeezers), along with homodyne and heterodyne detection measurements, all of which are now routinely implemented [20,21]. On the other hand, it has also been found that some of the important quantum protocols cannot be performed with Gaussian operations and classical processing alone. Examples of those are universal quantum computing [22], entanglement distillation of Gaussian states [23–25], optimal cloning of coherent states [26], optimal discrimination of coherent states [27–30],

and Gaussian quantum error correction [31]. Here we ask the question whether the classical information transmission capacity for bosonic channels can benefit from a receiver that is restricted to Gaussian quantum operations and FF. The main result of our paper is an addition to the long list of the above "no-go theorems", i.e., we show that Gaussian operations and classical FF processing cannot exceed the Shannon limit of homodyne and heterodyne detection for the lossy-noisy bosonic channel with coherent-state inputs. Our result has a practical importance: To exceed the theoretical limit of the conventional coherent optical communication system via a collective quantum decoder, one has to incorporate non-Gaussian operations in the receiver. As a by-product of the analysis, we propose a hybrid receiver that time shares between homodyne and heterodyne, which we show to slightly surpass the envelope of the homodyne and heterodyne capacities in a certain regime of average signal photon number per mode.

Finally we compare our Gaussian receiver limit with experimentally feasible non-Gaussian receivers based on photon counting detectors. We show that for extremely low signal power, it is possible to overcome the Gaussian receiver limit via relatively simple non-Gaussian receivers.

II. CHANNEL MODEL

The sender encodes messages in coherent states and sends them via multiple uses of a lossy-noisy bosonic channel $\mathcal{N}_{\rm B}$, each use of which can be modeled by a beam splitter with transmissivity $\eta \in (0,1]$ with a zero-mean thermal state injected into the other input port (see Fig. 1). Let $\tilde{N}_{\rm th}$ be the mean photon number of the thermal state. For a coherent state input $|\beta\rangle$, $\beta \in \mathbb{C}$, the output is a displaced thermal state:

$$\mathcal{N}_{\rm B}(|\beta\rangle\langle\beta|) = D(\alpha)\rho_{\rm th}(N_{\rm th})D^\dagger(\alpha) \equiv \rho(\alpha), \tag{3}$$

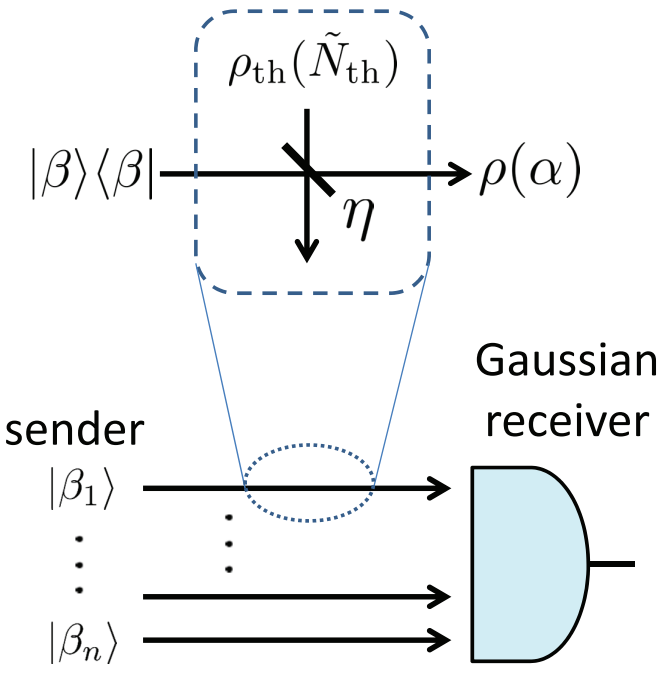

FIG. 1. (Color online) Lossy and noisy bosonic channel model and n use of the channels.

where $D(\alpha) = \exp[\alpha\hat{a}^\dagger - \alpha^*\hat{a}]$ is a displacement operation, $\hat{a}$ and $\hat{a}^\dagger$ are the annihilation and creation operators, respectively, $\rho_{\rm th}(N_{\rm th})$ is a thermal state with mean photon number $N_{\rm th}$, and

$$\alpha = \sqrt{\eta}\beta, \quad N_{\rm th} = (1-\eta)\tilde{N}_{\rm th}. \tag{4}$$

The sender prepares an n-mode coherent state, $\rho_S = \int d^{2n}\underline{\beta}\, P(\underline{\beta})|\underline{\beta}\rangle\langle\underline{\beta}|$, to encode a message, where $|\underline{\beta}\rangle = |\beta_1\rangle \otimes \cdots \otimes |\beta_n\rangle$, with $\underline{\beta} = [\beta_1, \ldots, \beta_n]^T \in \mathbb{C}^n$ and $P(\underline{\beta})$ a probability distribution function. The modulation power constraint translates to

$$\int d^{2n}\underline{\beta}\, P(\underline{\beta})|\underline{\beta}|^2 \leqslant \tilde{N}. \tag{5}$$

The n channel use transmission transforms ρ_S to

$$\rho_R = \int d^{2n}\underline{\alpha}\, P(\underline{\alpha})\rho_r(\underline{\alpha}), \tag{6}$$

where $\underline{\alpha} = [\alpha_1, \ldots, \alpha_n]^T = \sqrt{\eta}\underline{\beta}$ and $\rho_r(\underline{\alpha}) = \underline{D}(\underline{\alpha})\rho_{\rm th}^{\otimes n}(N_{\rm th})\underline{D}^\dagger(\underline{\alpha})$, with $\underline{D}(\underline{\alpha}) = D(\alpha_1) \otimes \cdots \otimes D(\alpha_n)$,

$$\int d^{2n}\underline{\alpha}\, P(\underline{\alpha})|\underline{\alpha}|^2 \leqslant \bar{N}, \tag{7}$$

and $\bar{N} = \eta\tilde{N}$. A *quantum Gaussian receiver* decodes the message from ρ_R, which in general could make a Gaussian collective measurement involving any Gaussian operation, classical FF and postprocessing. Let $\{\Pi(\underline{\alpha}_M)\}$ be a positive operator-valued measure for an n-mode quantum Gaussian receiver with measurement outcome $\underline{\alpha}_M$. The maximum reliable information rate per channel use is given by

$$C(\mathcal{N}) = \lim_{n\to\infty} \frac{1}{n} \max_{P,\,\Pi} I(X^n; Y^n), \tag{8}$$

where $I(X^n; Y^n)$ is the mutual information calculated for priors $P_{X^n}(x^n) = P(\underline{\alpha})$ and transition probabilities $P_{Y^n|X^n}(y^n|x^n) = P(\underline{\alpha}_M|\underline{\alpha}) = \mathrm{Tr}[\rho_r(\underline{\alpha})\Pi(\underline{\alpha}_M)]$.

III. CAPACITY OF A LOSSY-NOISY BOSONIC CHANNEL WITH COHERENT STATES AND GAUSSIAN RECEIVER

In this section, we derive an explicit expression for Eq. (8) and show that the optimal quantum Gaussian receiver is simply given by a homodyne or heterodyne receiver, or an appropriate time sharing between them. Let us consider the general structure of an n-mode Gaussian receiver. As illustrated in Fig. 2(a), it can be decomposed into $(n+m)$-mode-input Gaussian unitary operations U_{Gi}, $i = 1,2,\ldots$ with m-mode ancillary Gaussian input, single-mode Gaussian measurements (without classical FF), and classical FF into subsequent unitaries. Without loss of generality, we consider only "noise-free" operations that map pure states into pure states (we can simulate any noisy process by simply discarding a part of the system in a noise-free operation [25]).

We prove the statement of optimality stated above, by showing (1) the classical FF is not necessary to maximize $I(X^n; Y^n)$, (2) the optimal Gaussian measurement is a separable one, and (3) the optimal separable measurement is given by homodyne, heterodyne, or time sharing between them.

Step 1: Classical feedforward operations. First we show that the classical FF operations in Fig. 2(a) are not necessary.

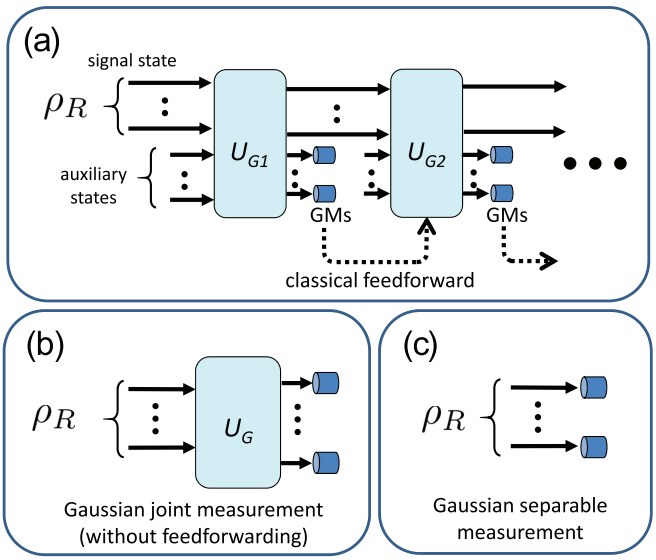

FIG. 2. (Color online) (a) General model of Gaussian receiver, (b) Gaussian joint measurement, and (c) Gaussian separable measurement. U_G: Gaussian unitary operation. GM: Gaussian measurement. ρ_R: signal state after transmitting the lossy and noisy bosonic channels.

One is able to show this by slightly extending the theorem on Gaussian operations shown in [23–25] as a part of the no-go theorem on Gaussian entanglement distillation via Gaussian local operations. Precisely, in [23–25] it was shown that for a Gaussian state input, any trace decreasing Gaussian operation (such as a partial measurement) can always be transformed into a trace preserving (deterministic) operation by adding an appropriate conditional displacement operation. It is thus clear that if the input in Fig. 2(a) is a Gaussian state, the conditional classical FF based on partial measurement outcomes are unnecessary. Note, however, that the received signal ensemble [Eq. (6)] is a convex combination of displaced thermal states which in general could be a non-Gaussian state. Nevertheless, by slightly extending the above theorem, we can show that for any possible convex combination [i.e., for any probability distribution $P(\underline{\alpha})$], any trace decreasing Gaussian operation can be transformed into a trace preserving Gaussian operation. This allows us to conclude that the conditional operations with partial measurement and classical FF are not necessary in our receiver. Though the extension of [23–25] is rather straightforward, we describe it in Appendix B for completeness.

Step 2: Joint and separable measurement. Removing the partial measurements and classical FF, the receiver measurement is now as illustrated in Fig. 2(b) where an n-mode Gaussian unitary operation U_G is followed by n heterodyne measurements. We now show that even a collective measurement is not necessary to obtain the maximum mutual information. Such a collective Gaussian detection is described by a set of operators $\{\Pi_G(\Gamma_{\mathcal{M}},d_{\mathcal{M}})\}_{d_{\mathcal{M}}}$ where $\Gamma_{\mathcal{M}}$ and $d_{\mathcal{M}}$ are, respectively, the covariance matrix (CM) and displacement vector (DV) of the characteristic function:

$$\begin{aligned}\chi_G(x) &= \mathrm{Tr}[\Pi_G(\Gamma_{\mathcal{M}},d_{\mathcal{M}})\mathcal{W}(x)]\\ &= \exp\left[-\tfrac{1}{4}x^T\Gamma_{\mathcal{M}}x + id_{\mathcal{M}}^T x\right],\end{aligned} \tag{9}$$

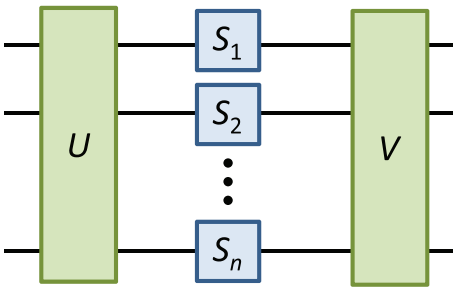

FIG. 3. (Color online) Decomposition of Gaussian unitary operation. U,V: unitary operations with linear optics. S_i: single-mode squeezers.

where $\mathcal{W}(x) = \exp[-ix^T R]$ is the Weyl operator, $x \in \mathbb{R}^{2n}$, and $R = [\hat{x}_1,\ldots,\hat{x}_n,\hat{p}_1,\ldots,\hat{p}_n]^T$ consists of quadrature operators satisfying commutation relation $[\hat{x}_k,\hat{p}_l] = i\delta_{kl}$. Note that the displacement vector $d_{\mathcal{M}}$ corresponds to a set of measurement outcomes. We summarize some basic properties of characteristic functions in Appendix A.

It is known that a Gaussian unitary operation can be decomposed into a passive linear optical unitary operation U (implementable via a network of beam splitters and phase shifters), a set of single-mode squeezers, and another passive linear optic unitary V [32] as illustrated in Fig. 3 [see also Eq. (A9) in Appendix A]. The covariance matrix and the displacement vector for the Gaussian measurement is now constructed as follows. A set of n heterodyne measurements on n modes is given by a correlation matrix $\gamma_{M_0} = I_{2n}$ and a $2n$ real vector d_{M_0}, where I_{2n} is a $2n \times 2n$ identity matrix. Denoting the symplectic matrix for the linear operation V as S_V, the CM and the DV of the measurement consisting of V and the heterodyne receivers are $S_V^T\gamma_{M_0}S_V = I_{2n}$ and $S_V^T d_{M_0}$, respectively. Including the single-mode squeezers, the measurement is described by $S_S^T S_V^T\gamma_{M_0}S_V S_S = \gamma_{\tilde{M}} = \mathrm{diag}[e^{-2r_1}\cdots e^{-2r_n}\,e^{2r_1}\cdots e^{2r_n}]$ and $S_S^T S_V^T d_{M_0} = d_{\tilde{M}}$, where r_i are the squeezing parameters. Finally, adding the linear operation U and defining $S_U^T d_{\tilde{M}} \equiv d_{\tilde{M}}$, the characteristic function of the joint Gaussian measurement is given by

$$\chi_G(x) = \exp\left[-\tfrac{1}{4}x^T S_U^T\gamma_{\tilde{M}}S_U x + id_{\tilde{M}}^T x\right]. \tag{10}$$

Note that the entries of $d_{\tilde{M}}$ can be obtained by applying the linear transformation V, in software, on the measurement outcome d_{M_0}. Since this operation can be performed *after* the measurement we can remove V from the Gaussian unitary operation without loss of generality.

The received ensemble is a set of displaced thermal states (6). In terms of the characteristic functions, denoting $\underline{\alpha} = [\alpha_1,\ldots,\alpha_n]^T$ by a $2n$-length real vector $d_r = \sqrt{2}[-\mathrm{Im}\,\alpha_1,\ldots,-\mathrm{Im}\,\alpha_n,\mathrm{Re}\,\alpha_1,\ldots,\mathrm{Re}\,\alpha_n]^T$, the characteristic function of the displaced thermal state $\rho_r(\underline{\alpha})$ is given by

$$\chi_r(x) = \exp\left[-\tfrac{1}{4}x^T\gamma_{\mathrm{th}}x + id_r^T x\right], \tag{11}$$

where

$$\gamma_{\mathrm{th}} = \begin{bmatrix}1+2N_{\mathrm{th}} & 0\\ 0 & 1+2N_{\mathrm{th}}\end{bmatrix}^{\oplus n}. \tag{12}$$

The conditional probability of obtaining the outcome $d_{\tilde{M}}$ by detecting $\rho_r(\underline{\alpha})$ with the Gaussian measurement in Eq. (10)

is then given by

$$P(d_{\tilde{M}}|d_S) = \left(\frac{1}{2\pi}\right)^{2n} \int dx\, \chi_r(x)\chi_M(-x) = \frac{1}{\sqrt{\det\left(\gamma_{\rm th} + S_U^T \gamma_{\tilde{M}} S_U\right)}} \times \exp\left[-(d_S - d_{\tilde{M}})^T \frac{1}{\gamma_{\rm th} + S_U^T \gamma_{\tilde{M}} S_U}(d_S - d_{\tilde{M}})\right]. \tag{13}$$

The expression in Eq. (13) is equivalent to that of a classical multidimensional Gaussian channel with a correlated noise CM, $\gamma_{\rm th} + S_U^T \gamma_{\tilde{M}} S_U$ [33]. It is well known that the mutual information of a Gaussian channel is maximized by a Gaussian input distribution [33],

$$P(d_r) = \frac{1}{2\pi^n \sqrt{\det P}} \exp\left[-\frac{1}{2} d_r^T \frac{1}{P} d_r\right], \tag{14}$$

where P is a diagonal matrix [34] with the power constraint: $\frac{1}{2n}\sum_{i=1}^{2n} P_{ii} \leqslant \bar{N}$. As a consequence the mutual information per channel use is given by $I(X;Y) = I(X^n;Y^n)/n$ where

$$I(X^n;Y^n) = \frac{1}{2}\log\frac{\det\left(2P + \gamma_{\rm th} + S_U^T \gamma_{\tilde{M}} S_U\right)}{\det\left(\gamma_{\rm th} + S_U^T \gamma_{\tilde{M}} S_U\right)} = \frac{1}{2}\log\frac{\det\left(2P + \gamma_{\rm th} + S_U^T \gamma_{\tilde{M}} S_U\right)}{\det(\gamma_{\rm th} + \gamma_{\tilde{M}})}. \tag{15}$$

The second equality follows from $\gamma_{\rm th} = (1 + 2N_{\rm th})I_{2n}$ and unitarity of S_U. To maximize $I(X^n;Y^n)$, we need to optimize S_U such that $\det(2P + \gamma_{\rm th} + S_U^T \gamma_{\tilde{M}} S_U)$ is maximized. According to the Hadamard inequality:

$$\det(X) \leqslant \prod_i X_{ii}, \tag{16}$$

where X is a positive matrix and equality holds iff X is diagonal, the maximum is obtained when $2P + \gamma_{\rm th} + S_U^T \gamma_{\tilde{M}} S_U$ is diagonal. Since P and $\gamma_{\rm th}$ are diagonal and the trace of $S_U^T \gamma_{\tilde{M}} S_U$ is invariant under any S_U, we conclude that $S_U = I_{2n}$ is optimal. As a consequence, the passive-linear-optic unitary U is unnecessary. Since we have already concluded that the other unitary V is also unnecessary, we conclude that it is sufficient for the optimal quantum Gaussian receiver to make a set of separable measurements [Fig. 2(c)].

Step 3: Optimization of the separable receiver. We split this part into the following two steps. We first consider a fixed receiver for all n channel uses and show that the optimal measurement is given either by a homodyne or a heterodyne measurement depending on the value of $\bar{N}$. Next we show that in a given range of $\bar{N}$ values, one can further optimize the mutual information by sharing the channel uses between homodyne and heterodyne measurements with an optimal power allocation across the channel uses.

As a first step, let us consider the maximization of the single-mode mutual information

$$I(X;Y) = \frac{1}{2}\log\frac{\det\left(2P^{(1)} + \gamma_{\rm th}^{(1)} + \gamma_M^{(1)}\right)}{\det\left(\gamma_{\rm th}^{(1)} + \gamma_M^{(1)}\right)}, \tag{17}$$

by optimizing a single-mode measurement $\gamma_M^{(1)}$ and power distribution $P^{(1)}$ (the superscripts denote $n = 1$). As mentioned above, $I(X;Y)$ is maximized with diagonal $P^{(1)}$ and $\gamma_M^{(1)}$. General expressions of diagonal $P^{(1)}$ and $\gamma_M^{(1)}$ are given by

$$P^{(1)} = \begin{bmatrix} N_1 & 0 \\ 0 & N_2 \end{bmatrix}, \quad \gamma_M^{(1)} = \begin{bmatrix} e^{-2r} & 0 \\ 0 & e^{2r} \end{bmatrix}, \tag{18}$$

where the power constraint is $(N_1 + N_2)/2 = \bar{N}$. Note that $r = \pm\infty$ and $r = 0$ correspond to homodyne and heterodyne measurement, respectively. Substituting Eq. (18) and $N_2 = 2\bar{N} - N_1$ into Eq. (17), we have

$$I(X;Y) = \frac{1}{2}\log\left[\frac{(2N_1 + 2N_{\rm th} + e^{-2r})}{(1 + 2N_{\rm th} + e^{-2r})} \times \frac{(4\bar{N} - 2N_1 + 2N_{\rm th} + e^{2r})}{(1 + 2N_{\rm th} + e^{2r})}\right] \equiv f(N_1, r),$$

which we want to maximize over $0 \leqslant N_1 \leqslant 2\bar{N}$ and $r \in (-\infty,\infty)$. By evaluating $\partial f(N_1,r)/\partial N_1$ and $\partial f(N_1,r)/\partial r$, we find that the extremum could exist only at $(N_1,r) = (\bar{N},0)$ with $f(\bar{N},0) = \log[(1 + N_{\rm th} + \bar{N})/(1 + N_{\rm th})]$. On the other hand, for $r \to \pm\infty$ or $N_1 = 0, 2\bar{N}$, the maximum f is obtained at $(N_1,r) = (2\bar{N},\infty)$ and $(0,-\infty)$ with

$$f(2\bar{N},\infty) = f(0,-\infty) = \frac{1}{2}\log\frac{1 + 2N_{\rm th} + 4\bar{N}}{1 + 2N_{\rm th}}. \tag{19}$$

Therefore, the maximum mutual information is given by

$$I(X;Y) = \max\left[\frac{1}{2}\log\frac{1 + 2N_{\rm th} + 4\bar{N}}{1 + 2N_{\rm th}}, \log\frac{1 + N_{\rm th} + \bar{N}}{1 + N_{\rm th}}\right] = \begin{cases} \frac{1}{2}\log\frac{1+2N_{\rm th}+4\bar{N}}{1+2N_{\rm th}}, & 0 \leqslant \bar{N} \leqslant \frac{2(1+N_{\rm th})}{1+2N_{\rm th}} \\ \log\frac{1+N_{\rm th}+\bar{N}}{1+N_{\rm th}}, & \bar{N} \geqslant \frac{2(1+N_{\rm th})}{1+2N_{\rm th}} \end{cases}, \tag{20}$$

which implies that if we fix the measurement on each mode, homodyne or heterodyne measurement is optimal. In other words, squeezing cannot increase the mutual information.

The above result is slightly improved around $\bar{N} = 2(1 + N_{\rm th})/(1 + 2N_{\rm th})$ by optimizing the power allocation between channel uses. In the following, to simplify the expressions, we set $N_{\rm th} = 0$ although the same approach works for finite $N_{\rm th}$.

Consider n channel uses (modes) and suppose homodyne detection is used on the first t uses and heterodyne detection on the rest. We can optimize the power allocation for these n modes under the condition $\sum_i N_i \leqslant n\bar{N}$ where N_i is the average photon number of the ith mode. This optimization can be carried out by the Lagrange multiplier method, defining

$$F(N_1,\ldots,N_n) = \sum_{i=1}^{t}\frac{1}{2}\log(1 + 4N_i) + \sum_{i=t+1}^{n}\log(1 + N_i) + \lambda\left(\sum_{i=1}^{n} N_i - \bar{N}\right), \tag{21}$$

where λ is a Lagrange multiplier. Solving $(dF)/(dN_i) = 0$ for $i = 1,\ldots,n$, we find the optimal photon numbers N_i as

$$N_i = \frac{1}{4}(\nu - 1) \quad (1 \leqslant i \leqslant t), \tag{22}$$

$$N_i = \frac{\nu}{2} - 1 \quad (t < i \leqslant n), \tag{23}$$

where $\nu = (4\bar{N} + 1 + 3x)/(1 + x)$ and $x = (n - t)/n$. The mutual information is then given by

$$\frac{1}{n} I(X^n; Y^n) = \frac{1}{2}(1 + x) \log \frac{4\bar{N} + 1 + 3x}{1 + x} - x. \quad (24)$$

We can further optimize x (or equivalently ν) and obtain

$$\frac{1}{n} I(X^n; Y^n) = \frac{1}{\nu^* - 3}(\log\, \nu^* - 2)(2\bar{N} - 1) + 1, \quad (25)$$

where ν^* is one of the solutions of $\nu(1 + 2\,\ln 2 - \ln \nu) = 3$, which satisfies $N_i > 0$ for all i. Numerically, this is $\nu^* = 7.145\ldots$. The optimal x is then given in terms of ν^* as $x = (4\bar{N} + 1 - \nu^*)/(\nu^* - 3)$, which yields $0 < x \leqslant 1$ when $(\nu^* - 1)/4 < \bar{N} \leqslant (\nu^* - 2)/2$. Therefore, in summary, the capacity of the pure-loss optical channel with laser-light modulation and a general quantum Gaussian receiver is given by

$$C = \begin{cases} \frac{1}{2}\log(1 + 4\bar{N}), & 0 \leqslant \bar{N} \leqslant \frac{\nu^* - 1}{4} \quad \text{(homodyne)} \\ \frac{\log\, \nu^* - 2}{\nu^* - 3}(2\bar{N} - 1) + 1, & \frac{\nu^* - 1}{4} < \bar{N} \leqslant \frac{\nu^* - 2}{2} \quad \text{(homodyne + heterodyne)} \\ \log(1 + \bar{N}), & \frac{\nu^* - 2}{2} < \bar{N} \quad \text{(heterodyne)}, \end{cases} \quad (26)$$

where $(\nu^* - 1)/4 = 1.536\ldots$ and $(\nu^* - 2)/2 = 2.572\ldots$. This is illustrated in Fig. 4.

The capacities for the channel with finite N_{th} can be derived from Eq. (20) by the same optimization procedure. We plot the numerical results in Fig. 5.

IV. NON-GAUSSIAN RECEIVERS BASED ON PHOTON COUNTING

In the previous sections, we showed that one has to incorporate non-Gaussian operations in the receiver in order to exceed the theoretical limit of conventional coherent optical communication. In this section, we compare that limit with the performance theoretically achievable using currently known structured non-Gaussian optical receivers, and also with the ultimate (Holevo) capacity limit. We only consider separable non-Gaussian receivers, i.e., receivers that detect each modulation symbol one at a time. One of the practical, but highly non-Gaussian, operations for an optical communication receiver is photon counting. This falls under the category of *direct detection* receivers (unlike homodyne and heterodyne detection, which collectively fall under *coherent detection* receivers). Direct detection receivers include the photon number resolving detector (PNRD) and the on-off single-photon detection (SPD) receiver, where the latter can discriminate only between zero and nonzero photons.

It has been shown theoretically that combining an ideal single-photon detector with the phase-space displacement operation (implementable using a highly transmissive beam splitter and a strong coherent-state local oscillator), and potentially also classical feedback (e.g., subsequent single-photon detection events triggering real-time updates to the amplitude and phase of a local oscillator that in turn is mixed with, to coherently *null*, the received pulse before it is incident on the detector's active surface) can beat the coherent-detection (homodyne and heterodyne) limit of discriminating two or more coherent states [27,35–38]. This was experimentally verified, without correcting for imperfections, for the binary phase-shift keyed (BPSK) signal ($\{|\alpha\rangle, |-\alpha\rangle\}$) [28] and for the quadrature phase-shift keyed signal ($\{|\alpha\rangle, |i\alpha\rangle, |-\alpha\rangle, |-i\alpha\rangle\}$) [39]. It was recently suggested that a PNRD receiver could also

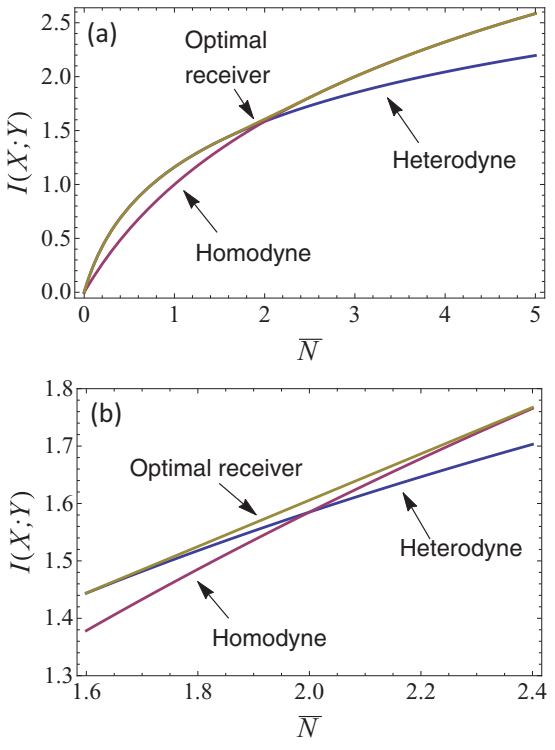

FIG. 4. (Color online) Capacity of homodyne (violet), heterodyne (blue), and optimal (yellow) receivers for a pure-loss optical channel as a function of the average photon number at the receiver $\bar{N} = \eta\bar{N}$. (a) and (b) show the same plots with different x-y ranges.

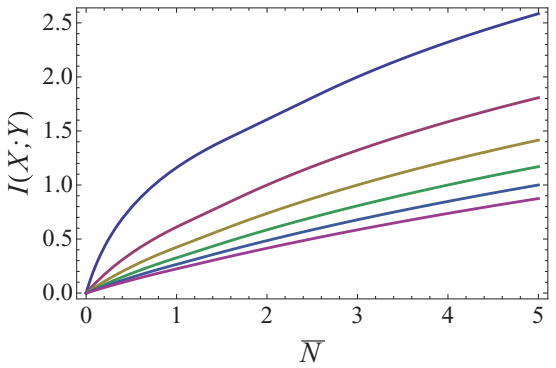

FIG. 5. (Color online) Capacity of optimal Gaussian receiver for optical channels with loss and thermal noise. The solid lines from top to bottom correspond to $N_{\text{th}} = 0,1,2,3,4,5$.

be useful for designing a nulling-based receiver to discriminate M-ary PSK signals at an error rate below the heterodyne detection limit [40]. In addition, it should be noted that a similar technique is useful to reduce the discrimination error below what is possible with standard optical receivers, not only for phase modulated signals but also for intensity modulated signals such as on-off keying (OOK) [41,42] and pulse-position modulation (PPM) [43,44]. Finally, a general design of a sequential-nulling receiver was recently proposed, which outperforms heterodyne detection for discriminating *any* M spatiotemporal coherent-state wave forms by a factor of 4 in the error-probability exponent [45]. These results on improved structured receivers for coherent-state discrimination suggests that receivers based on photon counting could also be useful to go beyond the capacity limit of Gaussian receivers. This is because the task of a communication receiver is essentially to discriminate between 2^{nR} modulated code words, each of which is a n-mode coherent-state wave form.

For optical communication system designers, a popular way to assess the performance of a transceiver is to plot the trade-off between spectral efficiency [expressed in bits per symbol, or (bits/s)/Hz] and the photon information efficiency (expressed in bits per received photon), for a given modulation format and a receiving strategy. For instance, for a pure-loss channel with $\bar{n}$ mean received photon number per mode (or per time slot), using an optimal code and a Holevo-capacity-achieving receiver, the spectral efficiency (SE) is $g(\bar{n})$ (bits/s)/Hz and the photon information efficiency (PIE) is $g(\bar{n})/\bar{n}$ bits/photon. When $\bar{n}$ is small, PIE is high and SE is small (this is the regime interesting for deep-space communication where every received photon is very precious), and in the high $\bar{n}$ regime, PIE is low and SE is high (this is the regime of interest for fiber-optic communication where the primary goal is to maximize the data rate). In Fig. 6, we plot the PIE-SE tradeoff for the Gaussian receiver limit (homodyne, heterodyne, and time sharing between the two), the ultimate Holevo limit, and several different modulation and receiver strategies. For this plot, we chose the lossy optical channel ($N_{\rm th} = 0$) for simplicity.

In Fig. 6, all the lines plotted with nonblack colors correspond to structured receivers, i.e., optical receivers the designs of which are fully specifiable in terms of standard optical and electrical elements. The majority of plots in the figure pertain to discrete modulation formats (e.g., BPSK, OOK, and PPM). For computing the highest capacities attainable by coherent-detection (homodyne, heterodyne, and the optimal time sharing between the two) receivers, as well as for evaluating the ultimate Holevo limit, we assume the optimal modulation, which for all those cases, is the continuous Gaussian modulation (i.e., when each symbol $|\alpha\rangle$ of a code word is chosen independent and identically distributed from the distribution $p(\alpha) = \exp[-|\alpha|^2/\bar{n}]/\pi\bar{n}$, $\alpha \in \mathbb{C}$).

The two non-Gaussian receivers we consider are single-photon detection (on-off direct detection) and the Dolinar receiver, a structured non-Gaussian receiver that can discriminate between any two coherent states at the minimum average error rate allowed by quantum mechanics [36,41]. The first observation to make is that in the low $\bar{n}$ (low SE, high PIE) regime, both the aforementioned non-Gaussian

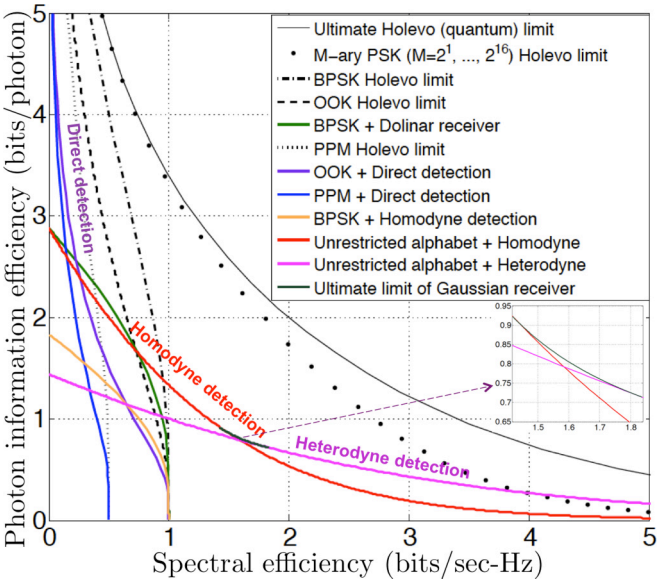

FIG. 6. (Color online) The trade-off between PIE and SE for various choices of modulation formats and receivers. All nonblack plots correspond to structured optical receivers, whereas the black lines are the Holevo capacities constrained to different modulation formats (i.e., with no restrictive assumption on the receiver). The thin black line on the top is the ultimate (Holevo) limit—no constraint on modulation and receiver—the highest capacity attainable over a pure-loss optical channel.

receivers—the Dolinar receiver (along with BPSK modulation) and the SPD receiver (along with either OOK or PPM modulation)—outperform the performance attainable by the general Gaussian receiver we studied in this paper. In the $\bar{n} \ll 1$ regime, the exact scaling of the Holevo limit [$C_{\rm ultimate}(\bar{n}) = -\bar{n}\,\ln\,\bar{n} + \bar{n} + o(\bar{n})$ (nats/s)/Hz], and the capacity achievable by a single-photon-detection receiver [$C_{\rm SPD}(\bar{n}) = -\bar{n}\,\ln\,\bar{n} - \bar{n}\,\ln\,\ln(1/\bar{n}) + O(\bar{n})$ (nats/s)/Hz] show that the gap between the two vanishes (i.e., their ratio goes to 1) as $\bar{n} \to 0$ [46]. Similarly, in the high $\bar{n}$ (low PIE, high SE) regime, the ratio of capacity attained by a coherent detection receiver (heterodyne detection) and the Holevo limit goes to 1 as $\bar{n} \to \infty$. Despite this, it is evident from Fig. 6 that there is a substantial gap between the attainable performance by known structured receivers (all of which admit physical realizations via a symbol-by-symbol detection of the received modes) and the Holevo limit, even at moderate to high spectral efficiency. This gap in capacity is even more amplified when more than one spatial mode is employed, such as in a diffraction-limited near-field free-space optical channel [47]. Even though some recent progress has been made on codes [6] and receiver designs [7] that could in principle attain the Holevo limit, a fully structured, and a practically feasible, design of such a non-Gaussian optical receiver still eludes us. Finally, note that it is not just the receiver choice, but an appropriate choice of the modulation constellation commensurate with the photon number level, is important as well. Figure 6 shows that in the high photon number (high SE) regime, the capacity attained by heterodyne detection (with an optimally chosen modulation) becomes higher than the envelope of the Holevo rates of an M-ary PSK constellation

(for $M = 2^1, 2^2, \ldots, 2^{16}$). This is not surprising since in the high photon number regime, the "circle" distribution does not approximate the circularly symmetric Gaussian distribution $p(\alpha)$ well. This is why the envelope of the Holevo rates of M-ary PSK modulation is very close to the ultimate Holevo capacity at low $\bar{n}$ (since one circle in the phase space very well approximates the Gaussian) and peels off from it at higher $\bar{n}$ values.

V. CONCLUSIONS

The ultimate classical capacity of a lossy, noisy bosonic channel is attained by a transmitter that modulates coherent-state (ideal laser-light) signals, albeit requiring a joint-detection receiver, which is hard to construct. In this paper, we restricted the receiver to a general quantum Gaussian receiver, which is made up of arbitrary quantum Gaussian operations (passive linear optics, squeezing, and homodyne or heterodyne measurements) along with classical feedforward operations. We showed that the optimal Gaussian receiver strategy that maximizes the information rate is simply given by either homodyne or heterodyne detection, or time sharing thereof. In other words, it was shown that any nontrivial Gaussian operation such as squeezing, partial measurements, and conditional feedforward, do not help increase the communication performance over conventional homodyne and heterodyne detection receivers. In order to bridge the gap between the Shannon capacity limit of homodyne and/or heterodyne detection, and the ultimate Holevo capacity [3], the receiver must use non-Gaussian operations. We showed that in the low-photon flux regime, the direct detection receiver (a practical non-Gaussian receiver) as well as the Dolinar receiver (another structured non-Gaussian receiver) can outperform the capacity attained by Gaussian measurements. We quantified the gap between the Shannon capacity limits of all the above known structured optical receivers, and the Holevo limit—the maximum capacity attainable with any receiver structure permissible by physics—in terms of the trade space between photon information efficiency (bits per received photon) and spectral efficiency [(bits/s)/Hz]. In order to attain the Holevo limit, the receiver must make collective measurements over long code word blocks [4–7], which must include non-Gaussian elements (such as Kerr interactions, photon counting, or interactions with non-Gaussian states). Heralded realization of non-Gaussian states and measurements is an active area of theoretical and experimental study. An important theoretical question is to conceive of an experimentally feasible design of a non-Gaussian receiver that can attain the Holevo capacity, the maximum rate at which classical data can be reliably transmitted over an optical communication channel.

ACKNOWLEDGMENTS

This work was performed while M.T. was on sabbatical from NICT at BBN. M.T. acknowledges the kind hospitality of BBNers during his stay at BBN. S.G. was supported by the Defense Advanced Research Projects Agency's (DARPA) Information in a Photon (InPho) program, under Contract No. HR0011-10-C-0159.

APPENDIX A: CHARACTERISTIC FUNCTIONS AND GAUSSIAN STATES

Here we briefly summarize the characteristic function formalism for Gaussian states and operations. More details can be found, for instance, in Refs. [21,25].

Characteristic function. Let us consider an n-mode bosonic system associated with an infinite-dimensional Hilbert space $\mathcal{H}^{\otimes n}$ and N pairs of annihilation and creation operators, $\{\hat{a}_i, \hat{a}_i^\dagger\}_{i=1,\ldots,n}$, respectively, which satisfy the commutation relations

$$[\hat{a}_i, \hat{a}_j^\dagger] = \delta_{ij}. \tag{A1}$$

From these, one may construct the quadrature field operators:

$$\hat{x}_i = \frac{1}{\sqrt{2}}(\hat{a}_i^\dagger + \hat{a}_i), \quad \text{and} \tag{A2}$$

$$\hat{p}_i = \frac{i}{\sqrt{2}}(\hat{a}_i^\dagger - \hat{a}_i). \tag{A3}$$

It is easy to verify that the commutation relations now translate to $[\hat{x}_i, \hat{p}_j] = i\delta_{ij}$. In the n-mode bosonic system, a quantum state with density operator ρ is described by its characteristic function

$$\chi(x) = \mathrm{Tr}[\rho \mathcal{W}(x)], \tag{A4}$$

where

$$\mathcal{W}(x) = \exp[-i x^T R] \tag{A5}$$

is a Weyl operator, $R = [\hat{x}_1, \ldots, \hat{x}_n, \hat{p}_1, \ldots, \hat{p}_n]^T$ is a $2n$ vector consisting of quadrature operators, and $x = [x_1, \ldots, x_{2n}]$ is a $2n$ real vector. The overlap between any two operators O_1 and O_2 is described by their characteristic functions as

$$\mathrm{Tr}[O_1 O_2] = \left(\frac{1}{2\pi}\right)^n \int dx\, \chi_{O_1}(x)\chi_{O_2}(-x). \tag{A6}$$

Gaussian states and operations. The characteristic function for any Gaussian state is represented by

$$\chi(x) = \exp\left[-\tfrac{1}{4} x^T \gamma x + i d^T x\right], \tag{A7}$$

where $2n \times 2n$ matrix γ and $2n$ vector d are called the covariance matrix and the displacement vector, respectively. Also, a Gaussian unitary operation is defined as a unitary operation that transforms Gaussian states to other Gaussian states. Any Gaussian unitary operation acting on a Gaussian state can be described by symplectic transformations of the covariance matrix and the displacement vector as

$$\gamma \to S^T \gamma S, \quad d \to S^T d, \tag{A8}$$

where S is a symplectic matrix. For any covariance matrix, there exists a symplectic transformation that diagonalizes the covariance matrix (symplectic diagonalization). If the unitary operation includes only linear optical process (beam splitters and phase shifts), then $S^T = S^{-1}$ and such a matrix S is called an orthogonal symplectic matrix.

Decomposition of Gaussian unitary operation. A symplectic matrix S can always be decomposed as

$$S = O \begin{pmatrix} M & 0 \\ 0 & M^{-1} \end{pmatrix} O', \tag{A9}$$

where M is a positive diagonal matrix and O, O' are orthogonal symplectic matrices. The physical meaning of the decomposition is that any Gaussian unitary circuit can be described by a sequential operation of linear optic circuit O, a product of single-mode squeezing operations, and another linear optic circuit O'.

APPENDIX B: CLASSICAL FEEDFORWARD OPERATION FOR A GAUSSIAN STATE ENSEMBLE WITH A NON-GAUSSIAN DISTRIBUTION

In this Appendix, we show that for a convex combination of Gaussian states, any trace decreasing Gaussian operation can be transformed into a trace preserving Gaussian operation. To this end, we use the characteristic functions whose basic properties were mentioned above in Appendix A.

Recall that the received state [Eq. (6)] is given by

$$\rho_R = \int d^{2n}\underline{\alpha} P(\underline{\alpha})\rho_r(\underline{\alpha}).$$

Its characteristic function is given by a convex combination of the characteristic functions of $\rho_r(\underline{\alpha})$:

$$\begin{aligned}\chi_R(x) &= \mathrm{Tr}\left[\rho_R \mathcal{W}(x)\right] \\ &= \int d^{2n}\underline{\alpha} P(\underline{\alpha})\mathrm{Tr}[\rho_r(\underline{\alpha})\mathcal{W}(x)] \\ &= \int d^{2n}\underline{\alpha} P(\underline{\alpha})\chi_{\rho_r(\underline{\alpha})}(x), \end{aligned} \tag{B1}$$

where $\chi_{\rho_r(\underline{\alpha})}(x)$ is the characteristic function of $\rho_r(\underline{\alpha})$.

Consider the trace decreasing operation consisting of a single step FF operation as illustrated in Fig. 7, where an n-mode state ρ_R is incident into an $(n+m)$-mode Gaussian unitary operation U_G with an m-mode auxiliary Gaussian state ρ_{aux} and a part of the output (system B) is measured by an m-mode Gaussian measurement. Let (γ_r, d_r) and $(\gamma_{\mathrm{aux}}, d_{\mathrm{aux}})$ be sets of the covariance matrices and displacement vectors for $\rho_r(\underline{\alpha})$ and ρ_{aux}, respectively. Let S_G be a symplectic matrix for U_G. Then after operating U_G, the covariance matrix and displacement vector of the $(n+m)$-mode output are respectively given by

$$S_G^T(\gamma_r \oplus \gamma_{\mathrm{aux}})S_G \equiv \begin{bmatrix} A & C \\ C^T & B \end{bmatrix} \tag{B2}$$

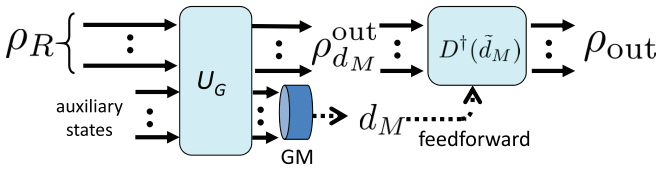

FIG. 7. (Color online) Gaussian measurement with a single step feedforward (see the text for details).

and

$$S(d_r \oplus d_{\mathrm{aux}}) \equiv \begin{bmatrix} d_A \\ d_B \end{bmatrix}. \tag{B3}$$

After measuring the system B by a Gaussian measurement with the covariance matrix γ_M and displacement (measurement outcome) d_M, we obtain the output in A conditioned on d_M, whose covariance matrix and displacement vector are respectively given by [25,27]

$$\Gamma_{\mathrm{out}} = A - C^T \frac{1}{B+\gamma_M} C \tag{B4}$$

and

$$\begin{aligned} d_{\mathrm{out}} &= \left(d_A - C^T \frac{1}{B+\gamma_M} d_B\right) - C^T \frac{1}{B+\gamma_M} d_M \\ &\equiv \tilde{d}_{\mathrm{out}} + \tilde{d}_M, \end{aligned} \tag{B5}$$

where without loss of generality we assumed that U_G does not include displacement operations (it can easily be canceled by the inverse operation). Although this is a conditional operation on the measurement outcome d_M, one can easily eliminate d_M in the output state by adding an additional displacement operation $D^\dagger(\tilde{d}_M)$ which results in the output state to be deterministically given by Γ_{out} and d_{out} those that are independent of $\tilde{d}_M$. As a consequence, for input state ρ_R, we have the output

$$\chi_{\mathrm{out}} = \int d^{2n}\underline{\alpha} P(\underline{\alpha})\chi_{\underline{\alpha}}(x), \tag{B6}$$

where $\chi_{\underline{\alpha}}(x)$ is a Gaussian characteristic function with Γ_{out} and d_{out}. It does not include d_M and thus is independent of the partial measurement outcome. Thus the total operation is deterministic.

[1] C. M. Caves and P. D. Drummond, Rev. Mod. Phys. **66**, 481 (1994).
[2] H. P. Yuen and M. Ozawa, Phys. Rev. Lett. **70**, 363 (1993).
[3] V. Giovannetti, S. Guha, S. Lloyd, L. Maccone, J. H. Shapiro, and H. P. Yuen, Phys. Rev. Lett. **92**, 027902 (2004).
[4] S. Lloyd, V. Giovannetti, and L. Maccone, Phys. Rev. Lett. **106**, 250501 (2011).
[5] S. Guha, Phys. Rev. Lett. **106**, 240502 (2011).
[6] S. Guha and M. Wilde, in *Proceedings of the 2012 IEEE International Symposium on Information Theory (ISIT 2012), Cambridge, MA, USA* (IEEE, Piscataway, NJ, 2012), pp. 546–550.
[7] M. Wilde, S. Guha, S. H. Tan, and S. Lloyd, in *Proceedings of the 2012 IEEE International Symposium on Information Theory (ISIT 2012), Cambridge, MA, USA* (IEEE, Piscataway, NJ, 2012), pp. 551–555.
[8] V. Giovannetti, S. Guha, S. Lloyd, L. Maccone, and J. H. Shapiro, Phys. Rev. A **70**, 032315 (2004).
[9] R. König and G. Smith, Phys. Rev. Lett. **110**, 040501 (2013).
[10] V. Giovannetti, S. Lloyd, L. Maccone, and J. H. Shapiro, Nat. Photon. **7**, 834 (2013).
[11] V. Giovannetti, R. Garcia-Patron, N. J. Cerf, and A. S. Holevo, arXiv:1312.6225.
[12] A. S. Holevo, IEEE Trans. Inf. Theory **44**, 269 (1998).

[13] P. Hausladen, R. Jozsa, B. Schumacher, M. Westmoreland, and W. K. Wootters, Phys. Rev. A **54**, 1869 (1996).
[14] B. Schumacher and M. D. Westmoreland, Phys. Rev. A **56**, 131 (1997).
[15] M. Sasaki, K. Kato, M. Izutsu, and O. Hirota, Phys. Lett. A **236**, 1 (1997).
[16] M. Sasaki, K. Kato, M. Izutsu, and O. Hirota, Phys. Rev. A **58**, 146 (1998).
[17] J. R. Buck, S. J. van Enk, and C. A. Fuchs, Phys. Rev. A **61**, 032309 (2000).
[18] M. Fujiwara, M. Takeoka, J. Mizuno, and M. Sasaki, Phys. Rev. Lett. **90**, 167906 (2003).
[19] M. Takeoka, M. Fujiwara, J. Mizuno, and M. Sasaki, Phys. Rev. A **69**, 052329 (2004).
[20] S. L. Braunstein and P. van Loock, Rev. Mod. Phys. **77**, 513 (2005).
[21] C. Weedbrook, S. Pirandola, R. García-Patrón, N. Cerf, T. Ralph, J. Shapiro, and S. Lloyd, Rev. Mod. Phys. **84**, 621 (2012).
[22] S. D. Bartlett, B. C. Sanders, S. L. Braunstein, and K. Nemoto, Phys. Rev. Lett. **88**, 097904 (2002).
[23] J. Eisert, S. Scheel, and M. B. Plenio, Phys. Rev. Lett. **89**, 137903 (2002).
[24] J. Fiurasek, Phys. Rev. Lett. **89**, 137904 (2002).
[25] G. Giedke and J. I. Cirac, Phys. Rev. A **66**, 032316 (2002).
[26] N. J. Cerf, O. Krüger, P. Navez, R. F. Werner, and M. M. Wolf, Phys. Rev. Lett. **95**, 070501 (2005).
[27] M. Takeoka and M. Sasaki, Phys. Rev. A **78**, 022320 (2008).
[28] K. Tsujino, D. Fukuda, G. Fujii, S. Inoue, M. Fujiwara, M. Takeoka, and M. Sasaki, Phys. Rev. Lett. **106**, 250503 (2011).
[29] C. Wittmann, U. L. Andersen, M. Takeoka, D. Sych, and G. Leuchs, Phys. Rev. Lett. **104**, 100505 (2010).
[30] C. Wittmann, U. L. Andersen, M. Takeoka, D. Sych, and G. Leuchs, Phys. Rev. A **81**, 062338 (2010).
[31] J. Niset, J. Fiurášek, and N. J. Cerf, Phys. Rev. Lett. **102**, 120501 (2009).
[32] S. L. Braunstein, Phys. Rev. A **71**, 055801 (2005).
[33] T. M. Cover and J. A. Thomas, *Elements of Information Theory* (Wiley, New York, 1991).
[34] This does not lose the generality. Because if P is nondiagonal, one can always diagonalize it at the beginning of the decoding process. The diagonalization can be done by some linear optics circuit.
[35] R. S. Kennedy, Research Laboratory of Electronics, MIT, Quarterly Progress Report No. 108, 1973 (unpublished), p. 219.
[36] S. Dolinar, Research Laboratory of Electronics, MIT, Quarterly Progress Report No. 111, 1973 (unpublished), p. 115.
[37] R. S. Bondurant, Opt. Lett. **18**, 1896 (1993).
[38] M. P. da Silva, S. Guha, and Z. Dutton, Phys. Rev. A **87**, 052320 (2013).
[39] F. E. Becerra, J. Fan, G. Baumgartner, J. Goldhar, J. T. Kosloski, and A. Migdal, Nat. Photon. **7**, 147 (2013).
[40] S. Izumi, M. Takeoka, K. Ema, and M. Sasaki, Phys. Rev. A **87**, 042328 (2013).
[41] R. L. Cook, P. J. Martin, and J. M. Geremia, Nature (London) **446**, 774 (2007).
[42] K. Tsujino, D. Fukuda, G. Fujii, S. Inoue, M. Fujiwara, M. Takeoka, and M. Sasaki, Opt. Express **18**, 8107 (2010).
[43] S. Guha, J. L. Habif, and M. Takeoka, J. Mod. Opt. **58**, 257 (2011).
[44] J. Chen, J. L. Habif, Z. Dutton, R. Lazarus, and S. Guha, Nat. Photon. **6**, 374 (2012).
[45] R. Nair, S. Guha, and S.-H. Tan, Phys. Rev. A **89**, 032318 (2014).
[46] H.-W. Chung, S. Guha, and L. Zheng, in *Proceedings of the 2011 IEEE International Symposium on Information Theory (ISIT 2011), St. Petersburg, Russia* (IEEE, Piscataway, NJ, 2011), pp. 284–288.
[47] S. Guha, Z. Dutton, and J. H. Shapiro, in *Proceedings of the 2011 IEEE International Symposium on Information Theory (ISIT 2011), St. Petersburg, Russia* (IEEE, Piscataway, NJ, 2011), pp. 274–278.