Gaussian benchmark for optical communication aiming towards ultimate capacity

Jaehak Lee,[1, *] Se-Wan Ji,[1] Jiyong Park,[1] and Hyunchul Nha[1, 2, †]

[1] *Department of Physics, Texas A & M University at Qatar, P.O. Box 23874, Doha, Qatar*
[2] *School of Computational Sciences, Korea Institute for Advanced Study, Seoul 130-722, Korea*

We establish the fundamental limit of communication capacity within Gaussian schemes under phase-insensitive Gaussian channels, which employ *multimode* Gaussian states for encoding and *collective* Gaussian operations and measurements for decoding. We prove that this Gaussian capacity is additive, i.e., its upper bound occurs with separable encoding and separable receivers so that a single-mode communication suffices to achieve the largest capacity under Gaussian schemes. This rigorously characterizes the gap between the ultimate Holevo capacity and the capacity within Gaussian communication, showing that Gaussian regime is not sufficient to achieve the Holevo bound particularly in the low-photon regime. Furthermore the Gaussian benchmark established here can be used to critically assess the performance of non-Gaussian protocols for optical communication. We move on to identify non-Gaussian schemes to beat the Gaussian capacity and show that a non-Gaussian receiver recently implemented by Becerra *et al.* [Nat. Photon. **7**, 147 (2013)] can achieve this aim with an appropriately chosen encoding strategy.

PACS numbers: 03.67.Hk, 42.50.Lc, 42.50.Ex

Introduction—Sending and receiving signals via optical channels, e.g. optical fiber networks, is a crucial basis of communication. Employing protocols like intensity modulation and phase-shifting in optical communication [1–4], there eventually arises a question of fundamental importance—how quantum mechanics sets bound on communication capacity achievable using light beams. A remarkable result was recently established by proving the minimum output entropy conjecture [5, 6], i.e., the ultimate capacity under phase-insensitive Gaussian channels is achieved by using coherent states as information carriers (encoding). However, there still exists an outstanding problem on what quantum receivers (decoding) can practically be used to obtain (near) ultimate capacity. The Holevo-Schumacher-Westmoreland theorem states that the ultimate capacity [7] can be achieved *asymptotically* (using infinitely many channels) with a certain joint measurement [8, 9], which however requires highly nonlinear, so very demanding, operations. It is therefore important to identify quantum receivers achieving high communication rates practically.

Numerous studies on quantum receivers focused mostly on distinguishing a finite set of coherent states with error rate below standard quantum limit (SQL) [10–14]. For binary coherent states, the Dolinar receiver [11, 12] among them achieves the minimum error rate (Helstrom bound) [15]. Extending to M-ary signals ($M > 2$), several receivers have also been proposed [16–20] and experimentally demonstrated e.g. by adaptive phase nulling and photon counting [21, 22]. While the performance of these receivers was evaluated by error rate below SQL under specific codewords [14, 22], it is critically important to see how those *non-Gaussian* receivers manifest advantages in terms of capacity (mutual information). We thus need to identify the capacity achievable within *Gaussian communication schemes* employing Gaussian states, op-

erations, and measurements [23] readily available in laboratory [24]. It is unknown to what extent general Gaussian schemes particularly using entangling operations can improve capacity in contrast to separable schemes.

So far there are two well-known Gaussian communication schemes, coherent-state scheme with heterodyne detection and squeezed-state scheme with homodyne detection, studied under an ideal situation [1, 3] or channel noises [25, 26]. We recently extended study to general *single*-channel Gaussian communications with arbitrary inputs and measurements and showed that the optimal strategy among them is either coherent-state scheme or squeezed-state scheme depending on channel parameters [27]. As for multimode scenario, with inputs restricted to coherent states under thermal dissipative channels, Takeoka and Guha showed that the optimal Gaussian receiver is a separable one [28]. Since Gaussian receivers with coherent-state inputs do not saturate the ultimate channel capacity although the channel capacity is obtained with coherent-state inputs, their work provides an evidence for the gap between the capacity of Gaussian schemes and the ultimate channel capacity. However, the restriction to coherent-state inputs is not sufficient as other inputs (squeezed state) can yield higher capacity under some Gaussian channels [27].

In this paper we establish the ultimate limit of Gaussian schemes under phase-insensitive Gaussian channels in a general multimode scenario using arbitrary N-mode Gaussian input states and collective Gaussian measurements. We prove that its upper bound is achieved by separable inputs and separable measurements (additivity of Gaussian communication). The highest capacity of Gaussian schemes is thus obtained by the optimal *single-channel* protocol, i.e., either coherent-state scheme or squeezed-state scheme [27]. As the capacities of those two schemes do not achieve the Holevo bound [25, 27], we

characterize the exact gap between the ultimate channel capacity and the capacity within Gaussian communication. Our results shed light on optical communication in several aspects. First, it identifies an optimal protocol when resources are confined to Gaussian operations and Gaussian receivers. Until now, coherent-state and squeezed-state schemes were used as standard protocols due to simple applicability. We now show that they actually attain the upper limit of capacity within Gaussian resources. Second, it establishes a benchmark to rigorously assess enhanced performance of non-Gaussian schemes in terms of mutual information—a central quantity of interest in communication theory. Furthermore, we suggest a non-Gaussian receiver of [18, 21] combined with an appropriate encoding method as a feasible scheme for higher communication rate than Gaussian limit.

Gaussian Communication—An N-mode Gaussian state is fully characterized by its first moments (averages) and second moments (variances) of position and momentum operators $\hat{\boldsymbol{\xi}} = (\hat{x}_1, \hat{p}_1, \hat{x}_2, \hat{p}_2, \cdots, \hat{x}_N, \hat{p}_N)^T$ [24]. The second-order moments are given by a covariance matrix (CM) $\boldsymbol{\gamma}$ with elements $\gamma_{ij} = \frac{1}{2}\left\langle \hat{\xi}_i\hat{\xi}_j + \hat{\xi}_j\hat{\xi}_i \right\rangle - \left\langle \hat{\xi}_i \right\rangle \left\langle \hat{\xi}_j \right\rangle$ $(i,j = 1,\cdots,2N)$. Let Alice prepare an N-mode Gaussian state ρ_0 with mean values $\langle \hat{\xi}_i \rangle = 0$ and CM $\boldsymbol{\gamma}_{\text{in}}$. She encodes random variables $\boldsymbol{d}_\alpha = \sqrt{2}(\text{Re}\alpha_1, \text{Im}\alpha_1, \text{Re}\alpha_2, \text{Im}\alpha_2, \cdots, \text{Re}\alpha_N, \text{Im}\alpha_N)^T$ by performing displacements as $\rho_{\text{in}} = \bigotimes_{j=1}^N \hat{D}_j(\alpha_j)\rho_0 \bigotimes_{j=1}^N \hat{D}_j^\dagger(\alpha_j)$, with $\hat{D}_j(\alpha) = \exp\left(\alpha\hat{a}_j^\dagger - \alpha^*\hat{a}_j\right)$ a displacement operator on jth mode. While Gaussian schemes use Gaussian operations and measurements, no restriction is given to the probability distribution $P(\boldsymbol{d}_\alpha)$ of encoded variables (continuous or discrete). In the classical information theory [29], it is known that if the added noise is Gaussian, the input distribution must also be Gaussian to optimize capacity. In our Gaussian communication scenario, both of the noise emerging from the internal fluctuation of output states and the noise induced by measurement are Gaussian. Therefore, mutual information is maximized by a Gaussian distribution

$$P(\boldsymbol{d}_\alpha) = \frac{1}{(2\pi)^N \sqrt{\det P_{\text{in}}}} \exp\left(-\frac{1}{2}\boldsymbol{d}_\alpha^T P_{\text{in}}^{-1}\boldsymbol{d}_\alpha\right), \quad (1)$$

with P_{in} a $2N \times 2N$ real positive matrix characterizing the range (variance) of encoded variables.

Alice sends the N-mode state to Bob via N independent Gaussian channels, which gives an output ρ_{out} with amplitudes $\boldsymbol{d}'_\alpha = \sqrt{\tau}\boldsymbol{d}_\alpha$ and CM

$$\boldsymbol{\gamma}_{\text{out}} = T\boldsymbol{\gamma}_{\text{in}}T^T + M, \quad (2)$$

where $T = \sqrt{\tau}I^{\otimes N}$ and $M = mI^{\otimes N}$ (I: 2×2 identity matrix). Note that an arbitrary phase-insensitive Gaussian channel can be understood as a concatenation of loss and amplification channels, fully characterized by

two parameters τ and m satisfying $m \geq |\tau - 1|$[30]. For instance, a loss channel with transmittance $\eta \leq 1$ and thermal photons n_{th} is characterized by $\tau = \eta$ and $m = (1-\eta)(n_{\text{th}} + \frac{1}{2})$, while an amplification channel with gain $g \geq 1$ is by $\tau = g$ and $m = (g-1)\left(n_{\text{th}} + \frac{1}{2}\right)$.

Bob finally obtains outcomes $\boldsymbol{d}_\beta$ using an N-mode Gaussian measurement. An arbitrary N-mode Gaussian measurement can be described by a projection onto a Gaussian state ρ_M with mean values $\boldsymbol{d}_\beta$ and CM $\boldsymbol{\gamma}_M$. The conditional probability for $\boldsymbol{d}_\beta$, given an input $\boldsymbol{d}_\alpha$, then follows as a Gaussian distribution centered at $\boldsymbol{d}'_\alpha$ with its second moments determined by both the internal fluctuation of ρ_{out} and the added noise from ρ_M,

$$P(\boldsymbol{d}_\beta|\boldsymbol{d}_\alpha) = \frac{1}{(2\pi)^N \sqrt{\det(\boldsymbol{\gamma}_{\text{out}} + \boldsymbol{\gamma}_M)}} \quad (3)$$
$$\times \exp\left[-\frac{1}{2}(\boldsymbol{d}_\beta - \boldsymbol{d}'_\alpha)^T(\boldsymbol{\gamma}_{\text{out}} + \boldsymbol{\gamma}_M)^{-1}(\boldsymbol{d}_\beta - \boldsymbol{d}'_\alpha)\right].$$

Additionally, one might construct a receiver employing partial measurement on some (ancillary) modes and classical feedforward. However, the feedforward scheme does not make improvement to our general Gaussian scenario, like the case of coherent-state input [28], since any Gaussian measurement on Gaussian states can be transformed into a deterministic Gaussian operation [23, 31, 32]. It thus suffices to consider our settings without ancillary modes or feedforward scheme.

As for the calculation of capacity, our case resembles the classical Gaussian communication where signal power is given by $P_{\text{out}} = \tau P_{\text{in}}$ and noise by $\boldsymbol{\gamma}_{\text{out}} + \boldsymbol{\gamma}_M$ [29]. The mutual information measured in bits between Alice and Bob is given by

$$I(A:B) = \int d^{2N}\boldsymbol{d}_\alpha d^{2N}\boldsymbol{d}_\beta \ P(\boldsymbol{d}_\beta, \boldsymbol{d}_\alpha)\log_2 \frac{P(\boldsymbol{d}_\beta|\boldsymbol{d}_\alpha)}{P(\boldsymbol{d}_\beta)}$$
$$= \frac{1}{2}\log_2 \frac{\det(P_{\text{out}} + \boldsymbol{\gamma}_{\text{out}} + \boldsymbol{\gamma}_M)}{\det(\boldsymbol{\gamma}_{\text{out}} + \boldsymbol{\gamma}_M)}. \quad (4)$$

It grows indefinitely with the signal power P_{out}, so energy constraint is introduced for a realistic consideration. The average photon number per channel is bounded by $\bar{n}$, i.e. $n_0 + n_s \leq N\bar{n}$ where $n_0 = \frac{1}{2}(\text{tr}\boldsymbol{\gamma}_{\text{in}} - N)$ is from input state and $n_s = \frac{1}{2}\text{tr}P_{\text{in}}$ from signal encoding.

Without loss of generality, we only consider a pure-state input (encoding) and a projection onto a pure state (decoding). For a mixed state ρ_{in} or ρ_M, it is always possible to find a pure state leading to a higher capacity. Any pure Gaussian state can be obtained by applying on a vacuum state a unitary operation decomposed as USV, with U and V passive transformations (energy conserving) and S single-mode squeezing operations [33]. Using $S_V S_V^T = I$ with S_V a symplectic matrix for a passive unitary operation V [34], we can diagonalize CMs as

$$\boldsymbol{\gamma}_{\text{in}} = S_{U_0}\boldsymbol{\gamma}_{\text{in}}^{(D)}S_{U_0}^T \equiv S_{U_0}\bigoplus_{j=1}^N \begin{pmatrix} \frac{1}{2}e^{-2r_j} & 0 \\ 0 & \frac{1}{2}e^{2r_j} \end{pmatrix} S_{U_0}^T, \quad (5)$$
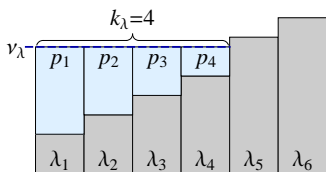
FIG. 1. Illustration of water-filling solution (optimal encoding) for a given noise vector $\vec{\lambda}$.

$$\gamma_{\mathrm{M}} = S_{U_M}\gamma_M^{(D)}S_{U_M}^T \equiv S_{U_M}\bigoplus_{j=1}^{N}\left(\begin{array}{cc}\frac{1}{2}e^{-2s_j} & 0 \\ 0 & \frac{1}{2}e^{2s_j}\end{array}\right)S_{U_M}^T. \tag{6}$$

Here S_{U_0} and S_{U_M} are symplectic matrices for passive unitary operations U_0 and U_M, and r_j and s_j single-mode squeezing parameters. Using Eqs. (2) and (5), the output CM is diagonalized as

$$\gamma_{\mathrm{out}} = S_{U_0}\left(T\gamma_{\mathrm{in}}^{(D)}T^T + M\right)S_{U_0}^T = S_{U_0}\gamma_{\mathrm{out}}^{(D)}S_{U_0}^T, \tag{7}$$

with the same S_{U_0} as applying on the input. We first investigate the mutual information with fixed r_j and s_j in decreasing order $r_1 \geq r_2 \geq \cdots \geq r_N \geq 0$ and $s_1 \geq s_2 \geq \cdots \geq s_N \geq 0$ and may later optimize r_j and s_j.

Optimization—For a given noise matrix $\gamma_{\mathrm{out}} + \gamma_M$, the mutual information (4) can be maximized when signal P_{out} is given by the so-called *water-filling* solution known in classical information theory [29] (Fig. 1). Let $\vec{\lambda} \equiv (\lambda_1, \lambda_2, \cdots, \lambda_{2N})^T$ be the eigenvalues (not sympletic ones) of $\gamma_{\mathrm{out}} + \gamma_M$ in increasing order $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_{2N}$, with its diagonalization $\Lambda = \mathrm{diag}(\lambda_1, \lambda_2, \cdots, \lambda_{2N}) = R^{-1}(\gamma_{\mathrm{out}} + \gamma_M)R$ (R: similarity transformation) [34]. The mutual information is then maximized when P_{out} is diagonalized by the same transformation, i.e., $R^{-1}P_{\mathrm{out}}R = \mathrm{diag}(p_1, p_2, \cdots, p_{2N})$. The optimal signal power p_j assigned against the noise λ_j is given by $p_j = \max\{\nu_\lambda - \lambda_j, 0\}$, where the "water" level ν_λ is determined to saturate the energy constraint, $\sum_{j=1}^{2N}\max\{\nu_\lambda - \lambda_j, 0\} = \mathrm{tr}P_{\mathrm{out}} = 2\tau n_s = 2\tau(N\bar{n} - n_0)$. Let k_λ denote the number of nonzero signals i.e. $\lambda_j < \nu_\lambda$ for $j \leq k_\lambda$ and $\lambda_j \geq \nu_\lambda$ for $j > k_\lambda$. Then, the mutual information via water-filling solution is given by

$$I(A:B) = \frac{1}{2}\sum_{j=1}^{2N}\log_2\left(1 + \frac{p_j}{\lambda_j}\right) = \frac{1}{2}\sum_{j=1}^{k_\lambda}\log_2\frac{\nu_\lambda}{\lambda_j}$$
$$= \frac{1}{2}\sum_{j=1}^{k_\lambda}\log_2\frac{2\tau n_s + \sum_{i=1}^{k_\lambda}\lambda_i}{k_\lambda\lambda_j} \equiv f(\lambda_1, \lambda_2\cdots, \lambda_{k_\lambda}). \tag{8}$$

In [35], we prove that $S_{U_0} = S_{U_M} = I$, i.e. CMs already diagonalized $\gamma_{\mathrm{in}} = \gamma_{\mathrm{in}}^{(D)}$ and $\gamma_M = \gamma_M^{(D)}$ representing product Gaussian states, gives the maximum in Eq. (8). Our proof relies on the majorization theory [36, 37] and the fact that $f(\lambda_1, \lambda_2\cdots, \lambda_{k_\lambda})$ in Eq. (8) is Schur-convex. It shows that a separable encoding/decoding is
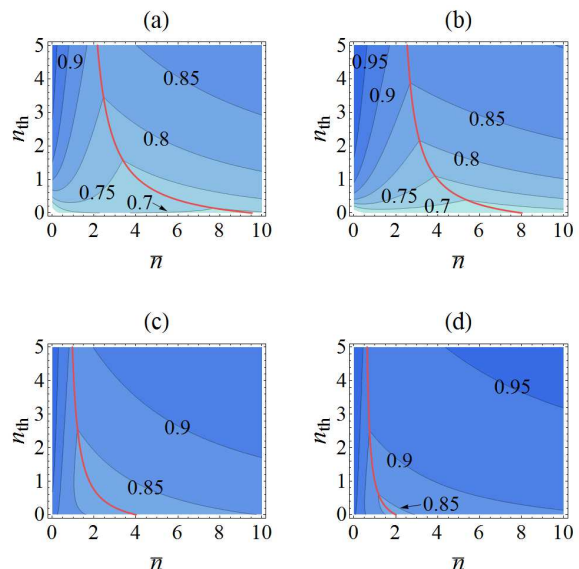

FIG. 2. Ratio of the Gaussian communication capacity to the ultimate Holevo bound against input energy $\bar{n}$ and thermal noise n_{th} under a lossy channel with (a) $\tau = 0.7$, (b) $\tau = 0.5$, and an amplification channel with (c) $\tau = 1.5$ and (d) $\tau = 2$. The red curves represent the crossover from C^{sq} (left region) to C^{coh} (right region) for the optimal Gaussian capacity.

the optimal strategy, reducing the problem to finding the optimal *single-channel* Gaussian scheme.

Single-channel communication—In [27], we proved that the optimal Gaussian communication under a single-channel use is one of two well-known Gaussian schemes, coherent-state scheme or squeezed-state scheme depending on channel parameters [38]. In [35], we give details for their capacities, C^{coh} (coherent-state scheme) and C^{sq} (squeezed-state scheme), compared with the ultimate Holevo bound C^{Holevo} for completeness. Precisely, the Gaussian capacity approaches the Holevo bound in the high-energy limit, but does not so in low-photon regime of practical importance. Nevertheless, it attains a considerably high capacity in broad parameter regions. In Fig. 2, we plot the ratio of Gaussian communication capacity to the Holevo capacity, $\max\{C^{\mathrm{coh}}, C^{\mathrm{sq}}\}/C^{\mathrm{Holevo}}$. With input energy $\bar{n}$ increasing, there generally occurs a crossover from squeezed-state scheme to coherent-state scheme for the highest Gaussian capacity, with critical $\bar{n}_c = \frac{1+2m+\tau}{2m\tau}$. For $n < n_c$, the squeezed-state scheme achieves a high efficiency ($> 90\%$) for small $\bar{n}$ and large n_{th}. As $\bar{n}$ increases, the efficiency of squeezed-state scheme drops significantly and the optimal strategy turns into the coherent-state scheme for $n > n_c$. The coherent-state scheme becomes efficient with $\bar{n}$ and n_{th} increasing. Under amplification channel [Fig. 2(c,d)], a high efficiency ($> 90\%$) is achieved broadly with gain g increasing. Under loss channel [Fig. 2 (a,b)], such a high efficiency hardly appears with n_{th} small. Aiming at 80% (90%) efficiency under pure-loss channel (n_{th}=0), a large

input energy $\tau\bar{n} \gtrsim 52$ (8098) is required.

Beyond Gaussian limit—We have established the Gaussian benchmark for the capacity limit of Gaussian communication under general Gaussian settings, which turns out to be below the ultimate Holevo capacity particularly in the low-photon regime. It is then interesting to identify non-Gaussian protocols to beat this Gaussian limit. A simple non-Gaussian receiver using photon counting was proposed for binary coherent inputs beyond Gaussian limit in the extremely low photon-number regime [28]. Because a binary input carries at most one bit of information, we extend to an M-ary signal modulation ($M > 2$) for higher communication rate. While some studies were done for M-ary signals [16–22], the analysis was made on error rate against SQL. In contrast, capacity in bits is an important quantity in communication and we are now able to rigorously assess merits of non-Gaussian receivers against the Gaussian-capacity benchmark.

We here investigate a non-Gaussian receiver recently demonstrated by Becerra *et al.* [18, 21] using on-off detection without photon-number resolving for easy implementation, together with quadrature amplitude modulation (QAM) of coherent states as encoding. The Becerra receiver takes recursive steps of (i) splitting the coherent signal into L pulses and (ii) carrying out signal nulling (via inverse displacement) and photon detection sequentially. For each pulse, the receiver chooses a most likely hypothetical coherent state among inputs to displace input to a vacuum and then detects photons. If the hypothesis is correct, no click occurs at the detector. Looking into the detection history of previous stages, the receiver updates the hypothesis via Bayesian conditional probability, determining the displacement at next stage. The whole detection outcomes are used to guess an input.

Under QAM encoding, Alice prepares one of the coherent states at lattice points with spacing δ in phase space (Fig. 3(a,b)). In Fig. 3(c), we show the mutual information attained by the Becerra receiver for 4-QAM under a loss channel, compared with the Gaussian capacity limit and the Holevo bound [39]. Its capacity is improved by increasing the number of stages L. While it cannot beat the coherent-state scheme with $L = 4$, it slightly does so with $L = 16$ and 64. However, it is hard to beat the squeezed-state scheme, which is highly efficient for a large η and a small $\bar{n}$. For a small transmittance $\eta \lesssim 0.2$, the Becerra receiver begins to beat both Gaussian communication schemes.

To find enhanced performance of this non-Gaussian scheme by increasing the number of inputs, we move on to 16-QAM. With the pulse splitting to $L = 64$, we find that the Becerra receiver cannot beat the Gaussian limit with a standard 16-QAM where all input states are prepared with uniform probability (Fig. 3(d)). This is because a coherent state with a larger displacement carries more energy, which needs to be probabilistically suppressed as we evaluate capacity under energy constraint. We thus
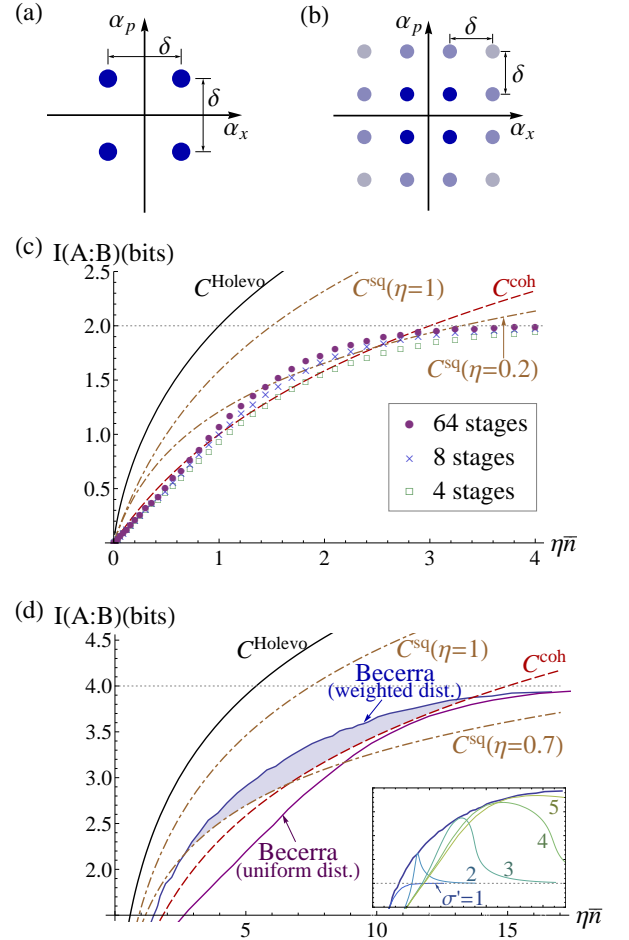

FIG. 3. (a) 4-QAM and (b) 16-QAM with more probable inputs in darker blue. Mutual information with (c) 4-QAM and (d) 16-QAM using the Becerra receiver, which beats all Gaussian schemes in the shaded region. Inset: capacity of Becerra receiver with $\sigma' = \{1, 2, 3, 4, 5\}$ (thin curves) and optimal σ (thick curve). Horizontal dotted line represents the maximum possible capacity 2-bits (4-bits) for 4-QAM (16-QAM).

investigate a modified 16-QAM where an input state is prepared according to a Gaussian-like distribution of deviation σ (Fig. 3(b)), i.e. more weighted towards smaller amplitudes. Bob receives one of 16-QAM signals with $\delta' = \sqrt{\eta}\delta$ and $\sigma' = \sqrt{\eta}\sigma$ due to channel loss. We now find that the Becerra receiver with this modified 16-QAM beats the Gaussian limit under a loss channel with moderate loss ($\eta = 0.7$). For a fixed σ, the mutual information first increases and then decreases with δ ($\bar{n}$). When δ is too large, only 4 points near the origin contribute to inputs so the mutual information decreases to only $\log_2 4 = 2$ bits. The optimal σ leading to the maximum mutual information increases with $\bar{n}$ (inset in Fig. 3(d)). For a small $\bar{n}$, it is hard to beat the Gaussian limit (squeezed-state scheme). On the other hand, with a moderate value of $\bar{n}$, the Becerra receiver outperforms both Gaussian schemes (shaded region) attaining a high

communication rate ($\gtrsim 3$ bits per channel use) [40].

Conclusion—We identified the capacity limit of general Gaussian settings with multimode Gaussian states and collective Gaussian measurements. We prove the additivity of Gaussian communication, i.e. single-mode communication is optimal without entangled states and joint-measurements. Our finding clarifies the optimal strategy in Gaussian protocols: squeezed-state (coherent-state) scheme in small (large) input-energy regime. Furthermore, such a fundamental Gaussian benchmark can be used to critically assess the advantages of non-Gaussian receivers in view of capacity. We investigated the Becerra receiver with QAM coherent-states manifesting high communication rate beyond the Gaussian limit remarkably with an appropriate encoding strategy. This seems feasible within current technology considering its recent realization for error rate below SQL.

Our Gaussian benchmark can be very useful in identifying and assessing other non-Gaussian schemes as well. One direction to pursue is a joint-measurement receiver showing superadditive capacity like the case of binary inputs [41, 42]. We addressed the importance of appropriately choosing an input-state distribution for a given receiver setting to achieve enhanced capacity, which shall be crucially incorporated in future works.

* jaehak.lee@qatar.tamu.edu

† hyunchul.nha@qatar.tamu.edu

[1] Y. Yamamoto and H. A. Haus, Rev. Mod. Phys. **58**, 1001 (1986).
[2] H. P. Yuen and M. Ozawa, Phys. Rev. Lett. **70**, 363 (1993).
[3] C. M. Caves and P. D. Drummond, Rev. Mod. Phys. **66**, 481 (1994).
[4] R. M. Gagliardi and S. Karp, *Optical Communications*, 2nd ed. (John Wiley & Sons, New York, 2000).
[5] V. Giovannetti, R. García-Patrón, N. J. Cerf, and A. S. Holevo, Nat. Photon. **8**, 796 (2014).
[6] A. Mari, V. Giovannetti, and A.S. Holevo, Nat. Commun. **5**, 3826 (2014).
[7] A.S. Holevo, Prob. Peredachi Inf. **9**, 3 (1973) [Probl. Inf. Transm. **9**, 177 (1973)].
[8] A.S. Holevo, IEEE Trans. Inf. Theory **44**, 269 (1998).
[9] B. Schumacher and M.D. Westmoreland, Phys. Rev. A **56**, 131 (1997).
[10] R.S. Kennedy, Research Laboratory of Electronics, MIT, Quarterly Progress Report No. 108, 1973, p. 219.
[11] S. J. Dolinar, Research Laboratory of Electronics, MIT, Quarterly Progress Report No. 111, 1973, p. 115.
[12] R. L. Cook, P. J. Martin, and J. M. Geremia, Nature **446**, 774 (2007).
[13] M. Takeoka and M. Sasaki, Phys. Rev. A **78**, 022320 (2008).
[14] J. Chen, J. L. Habif, Z. Dutton, R. Lazarus, and S. Guha, Nat. Photon. **6**, 374 (2012).
[15] C. W. Helstrom, *Quantum Detection and Estimation Theory* (Academic Press, New York, 1976).
[16] R. S. Bondurant, Opt. Lett. **18**, 1896 (1993).
[17] C. R. Müller, M. A. Usuga, C. Wittmann, M. Takeoka, C. Marquardt, U. L. Andersen, and G. Leuchs, New J. Phys. **14**, 083009 (2012).
[18] F. E. Becerra, J. Fan, G. Baumgartner, S. V. Polyakov, J. Goldhar, J. T. Kosloski, and A. Migdall, Phys. Rev. A **84**, 062324 (2011).
[19] S. Izumi, M. Takeoka, M. Fujiwara, N. D. Pozza, A. Assalini, K. Ema, and M. Sasaki, Phys. Rev. A **86**, 042328 (2012).
[20] R. Nair, S. Guha, and S.-H. Tan, Phys. Rev. A **89**, 032318 (2014).
[21] F. E. Becerra, J. Fan, G. Baumgartner, J. Goldhar, J. T. Kosloski, and A. Migdall, Nat. Photon. **7**, 147 (2013).
[22] F. E. Becerra, J. Fan, and A. Migdall, Nat. Photon. **9**, 48 (2015).
[23] G. Giedke and J. I. Cirac, Phys. Rev. A **66**, 032316 (2002).
[24] C. Weedbrook, S. Pirandola, R. García-Patrón, N. J. Cerf, T. C. Ralph, J. H. Shapiro, and S. Lloyd, Rev. Mod. Phys. **84**, 621 (2012).
[25] M. J. W. Hall, Phys. Rev. A **50**, 3295 (1994).
[26] G. M. D'Ariano and M. F. Sacchi, Opt. Commun. **149**, 152 (1998).
[27] J. Lee, S.-W. Ji, J. Park, and H. Nha, Phys. Rev. A **91**, 042336 (2015).
[28] M. Takeoka and S. Guha, Phys. Rev. A **89**, 042309 (2014).
[29] T. M. Cover and J. A. Thomas, *Elements of Information Theory* (J. Wiley, New York, 1991).
[30] R. García-Patrón, C. Navarrete-Benlloch, S. Lloyd, J. H. Shapiro, and N. J. Cerf, Phys. Rev. Lett. **108**, 110505 (2012).
[31] J. Eisert, S. Scheel, and M. B. Plenio, Phys. Rev. Lett. **89**, 137903 (2002).
[32] J. Fiurasek, Phys. Rev. Lett. **89**, 137904 (2002).
[33] S. L. Braunstein, Phys. Rev. A. **71**, 055801 (2005).
[34] R. Simon, N. Mukunda, and B. Dutta, Phys. Rev. A **49**, 1567 (1994).
[35] See Supplemental Material for details.
[36] A. W. Marshall, I. Olkin, B. C. Arnold, *Inequalities: Theory of majorization and its applications* (Springer, New York, 2011), 2nd ed.
[37] R. A. Horn and C. R. Johnson, *Matrix Analysis* (Cambridge University Press, Cambridge, 1985).
[38] Although the proof in Ref. [27] was given only for loss channels with $0 < \tau < 1$, it is straightforward to extend to all phase-insensitive Gaussian channels.
[39] Under a pure-loss channel, the capacity of a scheme employing coherent-state inputs is given as a function of output photon number $\eta\bar{n}$. This is because a pure-loss channel transforms a coherent state into another coherent state with only amplitude decreased by η.
[40] As another practical issue, it is not possible to implement a truly continuous modulation of input states for encoding, and one prepares only a finite (however large) number of coherent states. It is then interesting to know whether a scheme based on finite inputs can attain capacity emerging under the condition of continuous modulation, e.g. Holevo capacity and Gaussian communication

capacity. In [35], we show that our modified QAM, i.e. using a finite number of coherent states, can achieve the capacity of Gaussian communication by using a Gaussian receiver (heterodyne detection).

[41] M. Sasaki, K. Kato, M. Izutsu, and O. Hirota, Phys. Rev. A **58**, 146 (1998).

[42] S. Guha, Phys. Rev. Lett. **106**, 240502 (2011).

[43] V. Giovannetti, S. Lloyd, L. Maccone, and J. H. Shapiro, Nat. Photon. **7**, 834 (2013).

SUPPLEMENTAL MATERIAL

Proof for the optimality of separable encoding and decoding

First we briefly introduce the basics of majorization theory [36] used in our proof. When two d-dimensional vectors $\vec{x}$ and $\vec{y}$ sorted in increasing order satisfy the relation $\sum_{i=1}^{j} x_i \geq \sum_{i=1}^{j} y_i$ for all $j = 1, 2, \cdots, d-1$ and $\sum_{i=1}^{d} x_i = \sum_{i=1}^{d} y_i$, we say that $\vec{y}$ majorizes $\vec{x}$, with notation $\vec{x} \prec \vec{y}$. On the other hand, when the inequality $\sum_{i=1}^{j} x_i \geq \sum_{i=1}^{j} y_i$ is satisfied for all $j = 1, 2, \cdots, d$, instead of the equality for $j = d$, we say that $\vec{y}$ weakly majorizes $\vec{x}$, with notation $\vec{x} \prec^w \vec{y}$.

We call a function $f(\vec{x})$ Schur-convex if $f(\vec{x}) \leq f(\vec{y})$ for all majorized pairs $\vec{x} \prec \vec{y}$. On the other hand, for a weak version of majorization, $\vec{x} \prec^w \vec{y}$ implies $f(\vec{x}) \leq f(\vec{y})$ if and only if f is Schur-convex and decreasing with respect to all arguments of the function. A symmetric function $f(\vec{x})$, i.e. invariant under permutation of x_i and x_j, is provably Schur-convex if and only if $(x_j - x_i)\left(\frac{\partial f}{\partial x_j} - \frac{\partial f}{\partial x_i}\right) \geq 0$. By examining the derivative of the mutual information in Eq. (8) of main text, $\frac{\partial f}{\partial \lambda_j} = \frac{1}{2}\left(\frac{1}{\nu_\lambda} - \frac{1}{\lambda_j}\right)$, it is straightforward to show that $f(\lambda_1, \lambda_2, \cdots, \lambda_{k_\lambda})$ is decreasing and Schur-convex with respect to the arguments $0 \leq \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_{k_\lambda} \leq \nu_\lambda$.

In the main text, we have introduced CMs $\boldsymbol{\gamma}_{\text{out}} = S_{U_0} \boldsymbol{\gamma}_{\text{out}}^{(D)} S_{U_0}^T$ and $\boldsymbol{\gamma}_{\text{M}} = S_{U_M} \boldsymbol{\gamma}_{M}^{(D)} S_{U_M}^T$ representing the output state and the Gaussian measurement, respectively, at Bob's station. For the case of $S_{U_0} = S_{U_M} = I$, the eigenvalues of $\boldsymbol{\gamma}_{\text{out}}^{(D)} + \boldsymbol{\gamma}_{M}^{(D)}$ are simply obtained by the sum of their diagonal elements as $\vec{\mu} \equiv \frac{1}{2}(e^{-2r_1} + e^{-2s_1}, e^{-2r_2} + e^{-2s_2}, \cdots, e^{-2r_N} + e^{-2s_N}, e^{2r_N} + e^{2s_N}, e^{2r_{N-1}} + e^{2s_{N-1}}, \cdots, e^{2r_1} + e^{2s_1})^T$ in increasing order. Denoting the eigenvalues of $\boldsymbol{\gamma}_{\text{out}} + \boldsymbol{\gamma}_{\text{M}}$ by a vector $\vec{\lambda} \equiv (\lambda_1, \lambda_2, \cdots, \lambda_{2N})^T$, we have the majorization relation $\vec{\lambda} \prec \vec{\mu}$ for any choices of S_{U_0} and S_{U_M}. This is because for two Hermitian matrices X and Y with their eigenvalues $\vec{x}$ and $\vec{y}$, respectively, $\vec{x} + \vec{y}$ majorizes the eigenvalue vector of $X + Y$ [37]. Even though we have above shown that the mutual information in Eq. (8) of main text is Schur-convex, the relation $\vec{\lambda} \prec \vec{\mu}$ does not immediately guarantee

$$f(\lambda_1, \lambda_2, \cdots, \lambda_{k_\lambda}) \leq f(\mu_1, \mu_2 \cdots, \mu_{k_\mu}), \quad (9)$$

as the left- and the right-hand sides of Eq. (9) may have different number of arguments, i.e. $k_\lambda \neq k_\mu$. Nevertheless, we now show that the inequality (9) is valid. Recall that k_μ and k_λ are the numbers of nonzero signal assigned by water-filling solution to the noise vectors $\vec{\mu}$ and $\vec{\lambda}$, respectively.

(i) For the case of $k_\mu = k_\lambda$, $\vec{\lambda} \prec \vec{\mu}$ directly implies the weak majorization relation of truncated vec-
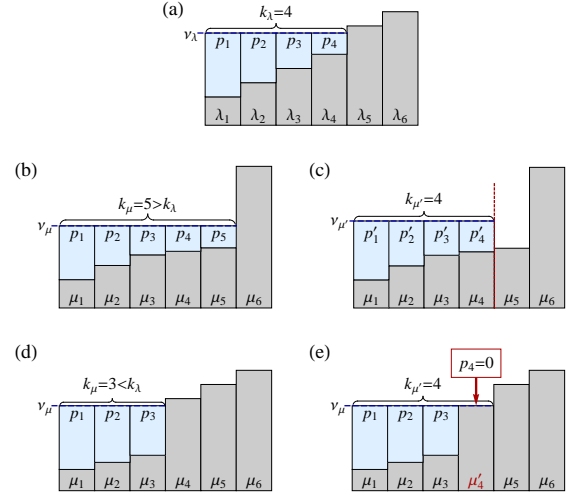

FIG. 4. Illustration of water-filling solutions (optimal encoding) for two different noise vectors (a) $\vec{\lambda}$ (case of generalized inputs and decoding) and (b,d) $\vec{\mu}$ (case of separable inputs and decoding). For the case of $k_\mu > k_\lambda$, we show (b) the optimal water-filling solution for a noise vector $\vec{\mu}$ and (c) an alternative encoding scheme assigning only $k_{\mu'} = k_\lambda$ number of signals. For the case of $k_\mu < k_\lambda$, we show (d) the optimal water-filling solution and (e) an encoding achieving the same capacity while the noise is adjusted so that $k_{\mu'} = k_\lambda$ including null signal $s_j = 0$.

tors, $(\lambda_1, \lambda_2, \cdots, \lambda_{k_\lambda})^T \prec^w (\mu_1, \mu_2, \cdots, \mu_{k_\mu})^T$ and we thus find $f(\lambda_1, \lambda_2, \cdots, \lambda_{k_\lambda}) \leq f(\mu_1, \mu_2 \cdots, \mu_{k_\mu})$.

(ii) When $k_\mu > k_\lambda$, let us consider an alternative encoding where we only assign k_λ signal to the noise vector $\vec{\mu}$ [Fig. 4(c)], instead of the optimal encoding with k_μ signal [Fig. 4(b)]. Under this restriction, the mutual information becomes maximal when the signal is assigned by water-filling method to the noise vector $(\mu_1, \mu_2, \cdots, \mu_{k_\lambda})^T$ as depicted in Fig. 4(c). Note that the new "water level" $\nu_{\mu'}$ is higher than ν_μ due to the energy constraint. Of course, this is not the optimal strategy for the given noise vector $\vec{\mu}$, however, we have $f(\lambda_1, \lambda_2, \cdots, \lambda_{k_\lambda}) \leq f(\mu_1, \mu_2 \cdots, \mu_{k_\lambda})$ due to the majorization relation $(\lambda_1, \lambda_2, \cdots, \lambda_{k_\lambda})^T \prec^w (\mu_1, \mu_2, \cdots, \mu_{k_\lambda})^T$. It thus gives the desired relation $f(\lambda_1, \lambda_2, \cdots, \lambda_{k_\lambda}) \leq f(\mu_1, \mu_2 \cdots, \mu_{k_\lambda}) \leq f(\mu_1, \mu_2 \cdots, \mu_{k_\mu})$.

(iii) When $k_\mu < k_\lambda$, we define a k_λ-dimensional vector $\vec{\mu}' = (\mu_1, \mu_2, \cdots, \mu_{k_\mu}, \nu_\mu, \nu_\mu, \cdots, \nu_\mu)^T$, where ν_μ is the water level for the original vector $\vec{\mu}$. Because $\mu_j \geq \nu_\mu$ for $j > k_\mu$, we have the majorization relation $(\lambda_1, \lambda_2, \cdots, \lambda_{k_\lambda})^T \prec^w (\mu_1, \mu_2, \cdots, \mu_{k_\lambda})^T \prec^w \vec{\mu}'$. Using the noise vector $\vec{\mu}'$, the optimal encoding is the same as that for $\vec{\mu}$ [Fig. 4 (e)], thus the mutual information does not change. In fact, we only add null signal $p_j = 0$ for $j = k_\mu + 1, k_\mu + 2, \cdots, k_\lambda$ [Fig. 4(e)]. Because we set elements of $\vec{\mu}'$ such that $0 \leq \mu'_1 \leq \mu'_2 \leq \cdots \leq \mu'_{k_\lambda} \leq \nu_\mu$, $f(\vec{\mu}')$ is again decreasing and Schur-

convex within this range. It thus gives the desired relation $f\left(\lambda_1, \lambda_2, \cdots, \lambda_{k_\lambda}\right) \leq f\left(\bar{\mu}'\right) = f\left(\mu_1, \mu_2 \cdots, \mu_{k_\mu}\right)$.

Gaussian communication schemes

A coherent-state scheme employs symmetric two-quadrature encoding on a coherent-state input and balanced heterodyne detection, that is, $\boldsymbol{\gamma}_{\text{in}} = \frac{1}{2}I$, $P_{\text{in}} = \bar{n}I$, and $\boldsymbol{\gamma}_M = \frac{1}{2}I$. Its capacity is given by

$$C^{\text{coh}} = \log_2\left(1 + \frac{2\tau\bar{n}}{1 + \tau + 2m}\right). \quad (10)$$

Squeezed-state scheme employs single-quadrature encoding on a squeezed-state input and homodyne detection, that is, $\boldsymbol{\gamma}_{\text{in}} = \frac{1}{2}\text{diag}\left(e^{-2r}, e^{2r}\right)$, $P_{\text{in}} = \text{diag}\left(2(\bar{n} - \sinh^2 r), 0\right)$, and $\boldsymbol{\gamma}_M = \frac{1}{2}\text{diag}\left(e^{-2s}, e^{2s}\right)$ with $s \to \infty$. With the choice of optimal squeezing $\exp(2r) = \frac{-\tau + \sqrt{8\tau m\bar{n} + (\tau + 2m)^2}}{2m}$ [25–27], the capacity of squeezed-state scheme becomes

$$C^{\text{sq}} = \log_2\left(\frac{-\tau + \sqrt{8\tau m\bar{n} + (\tau + 2m)^2}}{2m}\right). \quad (11)$$

In Fig. 5, we show the capacities of two Gaussian schemes together with the Holevo bound [5, 43] given by

$$C^{\text{Holevo}} = g\left(\tau\bar{n} + m + \tfrac{\tau-1}{2}\right) - g\left(m + \tfrac{\tau-1}{2}\right), \quad (12)$$
$$g(x) \equiv (1 + x)\log_2(1 + x) - x\log_2 x,$$

where $g(n)$ is the von Neumann entropy of a thermal state with thermal photon number n. We find that there exists a gap (shaded region) between the Holevo bound and the capacity of optimal Gaussian communication. For an ideal channel [Fig. 5(a)], the squeezed-state scheme always beats the coherent-state scheme [3] and attains the capacity close to the Holevo bound. However, as some noise is added [Fig. 5(b)], the capacity of the squeezed-state scheme grows with energy $\bar{n}$ less prominently than the coherent-state scheme, and the coherent-state scheme instead attains the capacity close to the Holevo bound. Except for the ideal case, there always exists a crossover between C^{coh} and C^{sq}. This is because a squeezed state is more fragile against the channel noise than a coherent state [5, 6]. One can readily find that the coherent-state scheme is optimal when $\bar{n} \geq \frac{1+2m+\tau}{2m\tau}$, and otherwise, the squeezed-state scheme is optimal. For the loss channel and the amplification channel with $n_{\text{th}} = 0$, we find a small-energy region where the squeezed-state scheme beats the coherent-state scheme [Fig. 5(c,e)], but this region becomes smaller as n_{th} increases [Fig. 5(d,f)]. The gap between the Holevo bound and the capacity of coherent-state scheme decreases as n_{th} increases, or m increases.
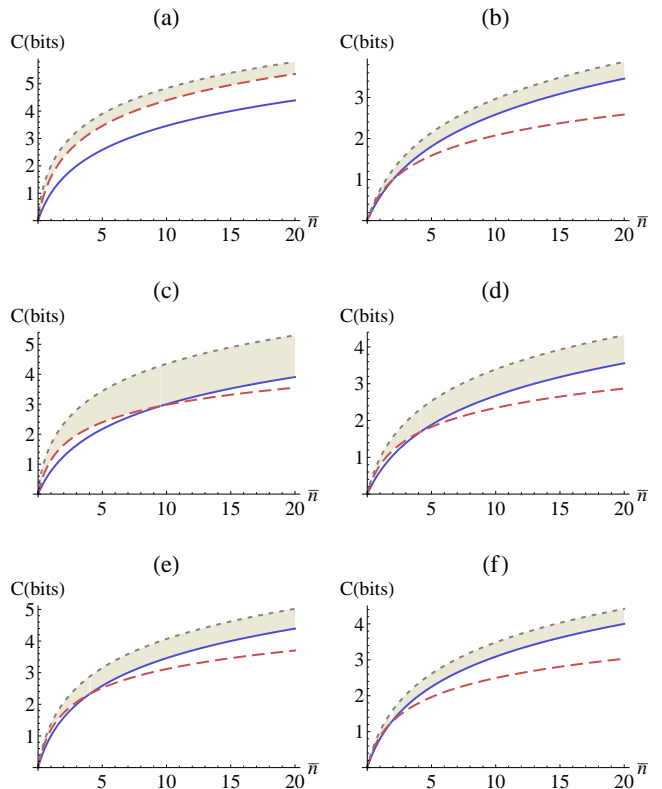

FIG. 5. Plot comparing the capacity of coherent-state scheme (blue solid curves), that of squeezed-state scheme (red dashed curves), and the Holevo bound (gray dotted curves) for different types of channel (a) an ideal channel, (b) a channel with added noise $m = 1$, (c) a pure-loss channel with $\tau = 0.5$ and $n_{\text{th}} = 0$, (d) a loss channel with $\tau = 0.7$ and $n_{\text{th}} = 1$, (e) a quantum-limited amplification channel with $\tau = 1.5$ and $n_{\text{th}} = 0$, and (f) an amplification channel with $\tau = 1.5$ and $n_{\text{th}} = 1$. The shaded region represents the gap between the Holevo bound and the capacity of optimal Gaussian communication.

QAM encoding with heterodyne detection

We here show that it is possible to approach the capacity of coherent-state scheme obtained under the condition of continuous modulation by using a *finite* number of coherent states for encoding. In Fig. 6, we plot the mutual information attained with 16(64)-QAM and balanced heterodyne detection under a pure-loss channel with transmittance η, along with the capacity of coherent-state scheme $C^{\text{coh}} = \log_2(1 + \eta\bar{n})$ based on continuous modulation. We see that the mutual information attained using the M-QAM with Gaussian-like distribution (more weighted towards the origin of phase space) approaches very closely the capacity of coherent-state scheme (optimal strategy with heterodyne receiver) under the same energy, with the condition $\eta\bar{n} \lesssim \sigma'^2$. The maximum mutual information under M-QAM is given by $\log_2 M$ bits, i.e., 4(6) bits for 16(64)-QAM, and is
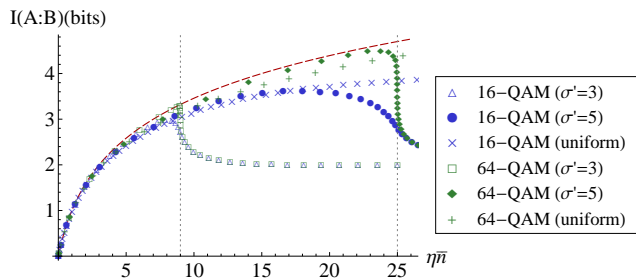
FIG. 6. Mutual information attained with 16(64)-QAM and balanced heterodyne detection. Each point is obtained numerically with different encoding strategies as represented in legends and with different lattice spacing δ. The dashed red curve shows the capacity of coherent-state scheme C^{coh} and vertical dotted lines represent $\eta\bar{n} = \sigma'^2$ for $\sigma' = \{3, 5\}$, respectively.

achieved when the points are perfectly distinguishable. The mutual information drops as the spacing δ becomes large enough as $\eta\bar{n} > \sigma'^2$ (vertical dashed lines). When δ is too large, only 4 points near the origin become significant among the inputs so the mutual information decreases to only $\log_2 4 = 2$ bits. On the other hand, if we employ M-QAM with uniform distribution ($\sigma \to \infty$), $I(A:B)$ shows a prominent gap from C^{coh} with increasing energy.