

Capacity of coherent-state adaptive decoders with interferometry and single-mode detectors

Matteo Rosati, Andrea Mari, and Vittorio Giovannetti

NEST, Scuola Normale Superiore, Istituto Nanoscienze-CNR, I-56127 Pisa, Italy

(Received 4 April 2017; published 12 July 2017)

A class of adaptive decoders (ADs) for coherent-state sequences is studied, including in particular the most common technology for optical-signal processing, e.g., interferometers, coherent displacements, and photon-counting detectors. More generally we consider ADs comprising adaptive procedures based on passive multimode Gaussian unitaries and arbitrary single-mode destructive measurements. For classical communication on quantum phase-insensitive Gaussian channels with a coherent-state encoding, we show that the AD's optimal information transmission rate is not greater than that of a single-mode decoder. Our result also implies that the ultimate classical capacity of quantum phase-insensitive Gaussian channels is unlikely to be achieved with the considered class of ADs.

DOI: 10.1103/PhysRevA.96.012317

I. INTRODUCTION

Quantum communication theory is a promising field for the application of quantum technology since its predictions could be applied in the short term in several settings of practical relevance. An important example is communication on free-space or optical-fiber links, which are well described theoretically by quantum phase-insensitive Gaussian channels [1–3], e.g., the lossy bosonic channel [4].

The maximum transmission rate of classical information on a quantum channel, known as its capacity, is provided by the Holevo-Schumacher-Westmoreland (HSW) theorem [5–9]. In particular for quantum phase-insensitive Gaussian channels the capacity at constrained average input energy can be achieved [10–13] by a simple *separable encoding*, i.e., sending sequences of coherent states [14], each of them constituting a letter for a single use of the channel or communication mode. This fact may seem surprising at first since coherent states are among the simplest states of the electromagnetic field and often are regarded as fundamentally classical. Nevertheless they are sufficient to achieve the maximum communication rate allowed by quantum mechanics on a broad class of channels of considerable practical relevance. Unfortunately the truly quantum challenge posed by these systems seems to reside in the decoding procedures since all known capacity-achieving measurements require *joint decoding* operations [8,9,15–26], i.e., reading out entire blocks of letters at once by projecting onto arbitrary entangled superpositions of the codewords. Hence even the classical coherent-state encoding requires a highly nontrivial quantum decoding to achieve capacity. Such joint quantum measurements are difficult to design with current technology [27–35] so that the quest for an optimal decoder of separable coherent-state codewords that would finally trigger practical applications is still open. Given the difficulty of implementing truly joint quantum measurements, research has then mainly focused on decoding coherent states with the general class of adaptive decoders (ADs) depicted in Fig. 1(a). The latter combines the available *single-mode* technology, e.g., photodetectors and local transformations, with *multimode passive* interferometers and *classical feed-forward control*. The rationale behind this choice is that introducing correlations between modes during the decoding procedure may increase the transmission rate of simple

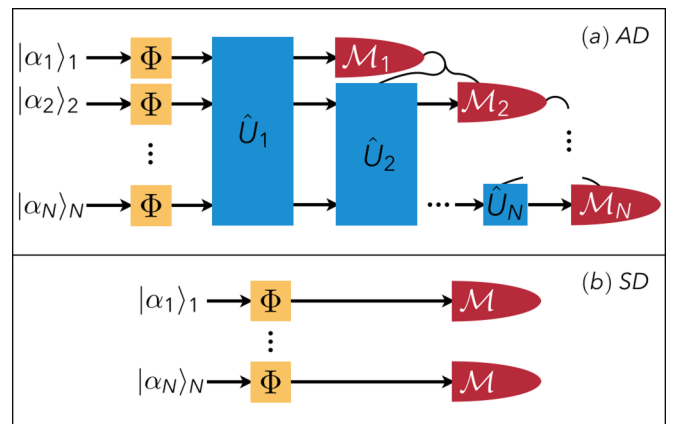


FIG. 1. Schematic of the class of (a) adaptive decoders (ADs) and (b) separable decoders (SDs) considered, whose maximum information transmission rate is proved to be equal. (a) In the AD case the sender, Alice, encodes the message into separable sequences of coherent states $|\alpha_1\rangle_1 \otimes \dots \otimes |\alpha_N\rangle_N$ and sends it to the receiver, Bob, with N distinct uses of a quantum phase-insensitive Gaussian channel Φ [yellow (light-gray) boxes], Eq. (2). Bob's AD comprises multimode passive Gaussian interferometers \hat{U}_j [blue (gray) boxes] “Eq. (3)” and arbitrary destructive single-mode measurements \mathcal{M}_j [red (dark-gray) shapes] “Eq. (4)” adaptively dependent on the measurement results of previous modes and applied successively on the remaining modes. (b) In the SD case Alice uses the same encoding, but Bob performs the same measurement \mathcal{M} on each mode and cannot use adaptive procedures.

separable measurements getting closer to the structure of joint quantum measurements that seems to be ultimately necessary to achieve the capacity of phase-insensitive Gaussian channels.

On the contrary, in this paper we prove that the maximum information transmission rate of such channels with coherent-state encoding and AD is equal to that obtained with a SD employing the same measurement on each mode as shown in Fig. 1(b). The general idea behind our proof is to map the quantum AD into an effective classical programmable channel with feedback to the encoder. Then we obtain our results by extending Shannon's feedback theorem [36,37] to this kind of channel.

Our paper gives several major contributions: (i) It implies the conjecture by Chung *et al.* [38,39], namely, that adaptive passive Gaussian interactions, single-mode displacements, and photodetectors do not increase the optimal transmission rate; (ii) if the HSW capacity of phase-insensitive Gaussian channels is achieved only by joint measurements as the evidence suggests so far, then it cannot be achieved with our AD scheme; (iii) it extends the results of Takeoka and Guha [30], who considered only Gaussian measurements; (iv) it extends the analysis made by Shor [40] in the context of trine states to coherent states and passive interactions. Our results, although already envisaged in previous works on the subject, have strong relevance for future research on practical decoders: (i) they extend the study of decoders by considering arbitrary single-mode manipulations before measurement, including non-Gaussian and nonunitary ones; (ii) they exclude a decoding advantage of adaptive passive Gaussian interactions, which are the easiest to realize in practice, suggesting that more difficult interactions are necessary to achieve capacity. Furthermore the possibility of employing ancillary states is partially included in our AD scheme: This is the case if each ancilla is allowed to interact just with one mode before being measured; otherwise, i.e., if the ancillae can interact with several modes, the problem of determining the decoder's optimal rate remains open and could give a practical advantage over SDs [41].

The article is structured as follows: In Sec. II we describe in detail the communication protocol and the class of decoders considered; in Sec. III we demonstrate that the AD's optimal rate is equal to the SD's one; in Sec. IV we discuss implications and draw our conclusions.

II. THE ADAPTIVE DECODER

Let us suppose that the sender, Alice, wants to transmit a classical message on N -independent communication modes employing coherent states of the electromagnetic field. The latter are defined in terms of the field's annihilation and creation operators \hat{a} , \hat{a}^\dagger as displaced vacuum states of phase-space amplitude $\alpha \in \mathbb{C}$, i.e., $|\alpha\rangle = \hat{D}(\alpha)|0\rangle$ with $\hat{D}(\alpha) = \exp[\alpha\hat{a}^\dagger - \alpha^*\hat{a}]$ as the displacement operator. The messages, represented by the sequence of classical input random variables $A_{(1,N)}$ with letters $A_j = \{\alpha_j \in \mathbb{C}\}$ for each $j = 1, \dots, N$, are encoded into a separable sequence of optical coherent states $|\alpha_{(1,N)}\rangle = |\alpha_1\rangle_1 \otimes \dots \otimes |\alpha_N\rangle_N$, one for each mode $|\cdot\rangle_j$ where we have used the compact notation $c_{(j,\ell)}$, $j \leq \ell$ to indicate a sequence of quantities c_j, \dots, c_ℓ on different modes from the j th to the ℓ th one. Each message is chosen according to a joint probability distribution $P_{A_{(1,N)}}(\alpha_{(1,N)}; E)$ at constrained average input energy per mode E , i.e.,

$$\int d^{2N} \alpha_{(1,N)} P_{A_{(1,N)}}(\alpha_{(1,N)}; E) \sum_{j=1}^N |\alpha_j|^2 \leq NE. \quad (1)$$

Let us also suppose that the transmission medium is well described by a quantum phase-insensitive Gaussian channel, represented by a linear completely positive and trace-preserving map Φ on the Hilbert space of a single mode and completely defined by its action on the displacement operator,

i.e.,

$$\hat{D}(\alpha) \xrightarrow{\Phi} \hat{D}(\mu_1\alpha)e^{-\mu_2\frac{|\alpha|^2}{2}}, \quad (2)$$

in terms of two parameters $\mu_i \geq 0$ satisfying the constraint $\mu_2 \geq |1 - (\mu_1)^2|$ [2]. As shown in Refs. [10–13], the separable coherent-state encoding discussed above achieves the classical capacity of Φ when its probability distribution is independent and identically distributed and Gaussian on each mode.

The receiver, Bob, has an AD that outputs the sequence of classical random variables $Y_{(1,N)}$, where $Y_j = \{y_j \in \mathcal{I}\}$ for all modes $j = 1, \dots, N$ and \mathcal{I} is the set of possible single-mode outcomes, which can be discrete or continuous, e.g., $\mathcal{I} = \mathbb{R}$ for homodyne detection. The probability distribution of the output variables can be computed from the conditional probability of obtaining an outcome sequence $y_{(1,N)}$ if the input sequence $\alpha_{(1,N)}$ was sent, i.e., $P_{Y_{(1,N)}|A_{(1,N)}}(y_{(1,N)}|\alpha_{(1,N)})$. The latter is determined by the specific decoding operations of the AD, Fig. 1(a), comprising for all $j = 1, \dots, N$:

(1) a multimode passive Gaussian unitary $\hat{U}_j(y_{(1,j-1)})$, i.e., a network of beam splitters and phase shifters conditioned on the outcomes of previous measurements, acting on the set of modes from the j th to the N th as

$$\hat{U}_j(y_{(1,j-1)})|\alpha_{(j,N)}\rangle = |U_j(y_{(1,j-1)})\alpha_{(j,N)}\rangle, \quad (3)$$

where U_j is the $(N - j + 1)$ -dimensional unitary matrix representing \hat{U}_j in phase space, applied directly to $\alpha_{(j,N)}$ as a phase-space vector;

(2) single-mode operations and a final destructive measurement, altogether represented by a local positive operator-valued measure (POVM) $\mathcal{M}_j(\lambda(y_{(1,j-1)}))$ chosen among a set of possible POVMs that are labeled by the (discrete or continuous) index $\lambda \in \Lambda$ conditioned on the outcomes of previous modes. Each POVM is defined by a collection of positive operators corresponding to the possible single-mode outcomes,

$$\mathcal{M}_j(\lambda(y_{(1,j-1)})) = \{\hat{E}_{y_j}(\lambda(y_{(1,j-1)}))\}_{y_j \in \mathcal{I}}, \quad (4)$$

where the operators \hat{E}_{y_j} sum up to the identity on the Hilbert space of a single mode.

For our results to hold, a crucial assumption is that the single-mode POVMs completely destroy the measured state before any information is sent to the rest of the system; if instead Bob can perform partial measurements the AD's rate may increase, see Ref. [40]. Let us also note that the generic set of allowed POVMs described above can be restricted case by case by properly choosing the \hat{E}_{y_j} . For example the simplest toolbox for optical-signal processing is that of the Kennedy receiver [42] with POVMs of the form

$$\mathcal{M}^{\text{ken}}(\lambda) = \{E_0(\lambda), \mathbf{1} - E_0(\lambda)\}, \quad (5)$$

$$E_0(\lambda) = \hat{D}^\dagger(\lambda)|0\rangle\langle 0|\hat{D}(\lambda), \quad (6)$$

where the index $\lambda \in \mathbb{C}$ is the amplitude of a phase-space displacement in this case. Since the latter depends adaptively on previous outcomes, the AD with a single-mode Kennedy structure behaves similarly to a Dolinar receiver [43].

III. THE OPTIMAL RATE

The performance of a quantum decoder for the transmission of classical information can be evaluated by computing the mutual information of its classical input and output random variables. The latter is defined for our AD as

$$I(A_{(1,N)}:Y_{(1,N)}) = H(Y_{(1,N)}) - H(Y_{(1,N)}|A_{(1,N)}), \quad (7)$$

i.e., the difference of the Shannon entropy [37] of $Y_{(1,N)}$ and the Shannon conditional entropy of $Y_{(1,N)}$ given $A_{(1,N)}$. The AD's optimal information transmission rate then is obtained by maximizing the mutual information (7) over the input distribution with energy constraint E and the decoding operations and regularizing it as a function of the number of uses N , i.e.,

$$\mathcal{R}_{\text{AD}}(E) = \lim_{N \rightarrow \infty} \max_{\substack{P_{A_{(1,N)}}(\alpha_{(1,N)}; E), \\ \hat{U}_j(Y_{(1,j-1)}), \\ \lambda(Y_{(1,j-1)}) \in \Lambda}} \frac{I(A_{(1,N)}:Y_{(1,N)})}{N}. \quad (8)$$

We want to compare the AD with the SD of Fig. 1(b), comprising for each use of the channel Φ only a single-mode POVM $\mathcal{M}(\lambda)$ chosen from the same set of those in the AD parametrized by $\lambda \in \Lambda$ "Eq. (4)" but without any interaction or classical communication between modes. Obviously, the optimal rate of this SD is obtained by maximizing the mutual information of the single-mode input and output variables A_1 and Y_1 over the input distribution at constrained energy E and the POVM's parameter, i.e.,

$$\mathcal{R}_{\text{SD}}(E) = \max_{P_{A_1}(\alpha_1; E), \lambda \in \Lambda} I(A_1:Y_1). \quad (9)$$

In order to show that the optimization (8) reduces to (9), we find it useful to consider a more general decoder comprising the AD and a classical feedback link from Bob to Alice, that certainly cannot decrease the optimal rate (8). Exploiting this feedback and the phase-insensitive property of Φ , Alice can always perform the \hat{U}_j instead of Bob. Hence all the AD's interactions are represented by a classical feedback to the encoder that rearranges the remaining sequences $\alpha_{(j,N)} \in A_{(j,N)}$ into new sequences $\beta_{(j,N)} \in B_{(j,N)}$ with $B_j = \{\beta_j \in \mathbb{C}\}$ for all modes $j = 1, \dots, N$, before transmission on the channel. Crucially, each choice of \hat{U}_j corresponds to a different rearrangement performed by the encoder in such a way that the total average-energy constraint (1) still is respected by the joint probability distribution $P_{B_{(1,N)}}(\beta_{(1,N)}; E)$ of the new messages $B_{(1,N)}$.

As a function of the encoded variables β_j , the rest of the AD scheme can be rewritten as a single-mode classical *programmable* channel, i.e., a channel with memory λ that can be chosen adaptively depending on previous outcomes. The corresponding conditional probability at the j th use then is

$$\begin{aligned} P_{Y_j|B_j, Y_{(1,j-1)}}(y_j|\beta_j, \lambda(Y_{(1,j-1)})) \\ = \text{Tr}[\hat{E}_{y_j}(\lambda(Y_{(1,j-1)}))\Phi(|\beta_j\rangle\langle\beta_j|)], \end{aligned} \quad (10)$$

where $\hat{E}_{y_j}(\lambda)$'s are the elements of the POVM $\mathcal{M}_j(\lambda)$ as in Eq. (4).

In light of the previous observations we can conclude that the AD of Fig. 1(a) with additional classical communication from Bob to Alice is equivalent to the classical programmable

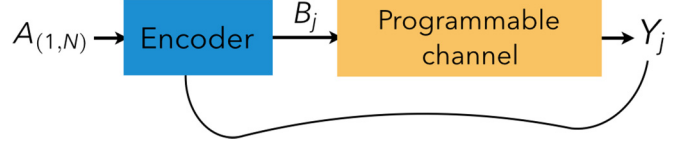


FIG. 2. Schematic of the classical communication scheme induced by the quantum AD, Fig. 1(a). The input sequence $\alpha_{(1,N)} \in A_{(1,N)}$ is encoded [blue (gray box)] into single letters $\beta_j \in B_j$ that are sent one by one on a classical memory channel [yellow (light-gray) box] with output $y_j \in Y_j$ for each $j = 1, \dots, N$. The adaptive passive interactions \hat{U}_j of the quantum scheme here correspond to a classical feedback to the encoder, i.e., the encoding function that generates each β_j depends on the input message $\alpha_{(1,N)}$ and on all previous results $y_{(1,j-1)}$. The single-mode phase-insensitive channel Φ and adaptive POVM \mathcal{M}_j employed on each letter β_j instead correspond to several uses of a classical programmable channel, whose memory at each use depends only on previous outcomes through the parameter λ characterizing the measurement as in Eq. (10).

channel (10) with feedback as shown in Fig. 2. Hence the AD's optimal rate Eq. (8) is upper bounded by the feedback capacity of (10). Similarly, the capacity of the programmable channel without feedback for a single use is equal to the SD's optimal rate Eq. (9). Eventually, the two classical capacities just defined are related via the following theorem, which is a generalization of Shannon's feedback theorem [36,37] to the class of programmable channels considered:

Theorem 1. The feedback capacity of a classical programmable channel is equal to its capacity without feedback, and it is additive.

Proof. Suppose we employ the channel to transmit a classical message $w \in W$ with probability distribution $P_W(w)$, outputting $y_j \in Y_j$ for each use j ; the most general technique allows a feedback to the sender, who encodes the input message into a sequence of letters $\beta_j \in B_j$ through an encoding function $\beta_j = f(w, y_{(1,j-1)})$ for each use j . If β represents the complex amplitude of a signal we must impose a total average-energy constraint as in Eq. (1). The feedback capacity of this classical programmable channel at constrained total average energy per mode E is obtained by maximizing the mutual information over the input distribution, the encoding functions, and the programmable parameters $\lambda(Y_{(1,j-1)})$ for each use,

$$C_{\infty}^{fb}(E) = \lim_{N \rightarrow \infty} \max_{\substack{P_W(w), \\ f(w, Y_{(1,j-1)}), \\ \lambda(Y_{(1,j-1)}) \in \Lambda}} \frac{I(W:Y_{(1,N)})}{N}. \quad (11)$$

Similarly, for independent uses of the channel without feedback, the capacity at constrained average-energy E can be defined as

$$C_1(E) = \max_{\substack{P_{B_1}(\beta_1; E), \\ \lambda \in \Lambda}} I(B_1:Y_1). \quad (12)$$

Now let us note that $C_{\infty}^{fb}(E) \geq C_1(E)$ since among all adaptive schemes involved in the optimization (11) there is one which employs no feedback and the same single-mode measurements that are optimal for Eq. (12). To prove the opposite consider

the following:

$$\begin{aligned}
 I(W:Y_{(1,N)}) &= \sum_{j=1}^N I(B_j:Y_j|Y_{(1,j-1)}) \\
 &\leq \left\langle \sum_{j=1}^N C_1[E_j(Y_{(1,j-1)})] \right\rangle_{P_{Y_{(1,N)}}(Y_{(1,N)})} \\
 &\leq NC_1(E),
 \end{aligned} \tag{13}$$

where the first equality follows from the chain rule of mutual information and the fact that conditioning over W and $Y_{(1,j-1)}$ is equivalent to conditioning over B_j and $Y_{(1,j-1)}$ thanks to the encoding functions, i.e., $H(Y_j|W, Y_{(1,j-1)}) = H(Y_j|B_j, Y_{(1,j-1)})$. The first inequality instead is obtained by employing the definition of Eq. (12) as an upper bound on each mutual information term in the sum and writing explicitly the average over the output distribution; the last inequality follows from concavity of the classical capacity as a function of the energy and the total average energy per mode constraint, i.e., $\sum_j E_j(Y_{(1,j-1)}) = NE$. Eventually by plugging Eq. (13) into the definition (11) we obtain the upper bound $C_\infty^{fb}(E) \leq C_1(E)$. ■

This implies that the AD’s optimal rate is not greater than the SD’s one. Since the former is certainly not smaller than the latter, we conclude $\mathcal{R}_{AD}(E) = \mathcal{R}_{SD}(E)$.

IV. IMPLICATIONS AND CONCLUSIONS

Our analysis implies that a broad class of adaptive decoders for coherent communication on phase-insensitive Gaussian channels, including a majority of those most easily realizable with current technology, cannot beat the optimal single-mode-measurement rate of information transmission. This in turn seems to suggest that such decoders cannot achieve the HSW capacity of phase-insensitive Gaussian channels; however there is no actual proof that joint decoders are really necessary for the task so that this possibility remains open. In any case our result does not mean that block-coding techniques and adaptive receivers are completely useless for practical applications; indeed in general there may exist specific AD schemes that are more convenient to implement than SD ones and perform equally well, e.g., see the Hadamard codes [32–35].

Let us also note that, despite the fact that our result is very powerful in decoupling the AD’s multimode structure for any kind of single-mode POVM, still the difficult optimization of the SD rate of Eq. (9) is left if one wants an explicit expression of the rate for any set of POVMs. For example, we can simplify this calculation for the set of single-mode receivers comprising

a coherent displacement followed by any other kind of single-mode operation [the Kennedy receiver of Eq. (5) belongs to this set]. Indeed let us define the variance of a single-mode input probability distribution $P_A(\alpha)$ over coherent states as $V = \langle |\alpha|^2 \rangle_{P_A(\alpha)} - |\langle \alpha \rangle_{P_A(\alpha)}|^2$; the energy is instead $E = \langle |\alpha|^2 \rangle_{P_A(\alpha)}$. One can decide to put a constraint either on the energy or on the variance of the input signals, and the former is stricter than the latter. It can then be shown that the net effect of the displacement in a coherent-state receiver is simply to enlarge the family of allowed input distributions from the energy- to the variance-constrained ones so that the optimal rate (9) can be computed on a shrunken set of allowed POVMs.

A particularly useful kind of single-mode receiver is that of Kennedy, defined by Eqs. (5) and (6), employing a coherent displacement and an on-off photodetector. The SD’s optimal rate for this receiver has been computed in the low-energy limit $E \ll 1$ in Refs. [38,39], showing that it equals

$$\mathcal{R}_{SD}^{\text{ken}}(E) = E \log \frac{1}{E} - E \log \log \frac{1}{E} + O(E). \tag{14}$$

Moreover the same authors have shown that an AD scheme without unitaries has the same optimal rate and conjectured that adaptive unitaries also do not help. Our result exactly implies the validity of this conjecture for the particular choice of POVMs (5),(6).

Eventually our result intersects with those of Refs. [30,40], expanding the set of adaptive receivers whose optimal rate is equal to that of separable ones. Indeed Ref. [30] computes the capacity of coherent communication with arbitrary adaptive Gaussian measurements showing it is separable; here instead we considered a restricted interaction set, i.e., passive Gaussians, but an extended single-mode measurement one, i.e., arbitrary POVMs. As for Ref. [40], it is stated there that adaptive schemes based on partial single-mode measurements of all the modes may increase the optimal rate; here we considered only destructive single-mode measurements but included the simplest kind of interactions and still could not surpass separable decoding rates. In particular, as stated in Sec. I, our AD includes the use of ancillary systems if they interact with just one of the received modes since this process can be thought of as a part of the single-mode destructive measurements. Unfortunately the interaction of ancillary systems with multiple modes is not included since it results in nondestructive measurements that could provide an advantage over SDs. Future lines of research could be as follows: studying the less-known, interesting class of nondestructive adaptive decoders, computing explicitly the optimal rate for other classes of POVMs, and exploring the potential of squeezing and non-Gaussian interactions.

[1] C. M. Caves and P. D. Drummond, *Rev. Mod. Phys.* **66**, 481 (1994).
 [2] A. S. Holevo, *Quantum Systems, Channels, Information*, De Gruyter Studies in Mathematical Physics (De Gruyter, Berlin, 2012).
 [3] A. S. Holevo and V. Giovannetti, *Rep. Prog. Phys.* **75**, 046001 (2012).

[4] V. Giovannetti, S. Guha, S. Lloyd, L. Maccone, J. H. Shapiro, and H. P. Yuen, *Phys. Rev. Lett.* **92**, 027902 (2004).
 [5] A. S. Holevo, *Probl. Peredachi Inf.* **9**, 3 (1973) [*Probl. Inf. Transm.* (Engl. Transl.) **9**, 110 (1973)].
 [6] A. S. Holevo, *IEEE Trans. Inf. Theory* **44**, 269 (1998).
 [7] A. S. Holevo, [arXiv:quant-ph/9809023](https://arxiv.org/abs/quant-ph/9809023) (see also Tamagawa University Research Review, No. 4) (1998).

- [8] B. Schumacher and M. D. Westmoreland, *Phys. Rev. A* **56**, 131 (1997); P. Hausladen, R. Jozsa, B. W. Schumacher, M. Westmoreland, and W. K. Wootters, *ibid.* **54**, 1869 (1996).
- [9] A. Winter, *IEEE Trans. Inf. Theory* **45**, 2481 (1999).
- [10] V. Giovannetti, R. García-Patrón, N. J. Cerf, and A. S. Holevo, *Nat. Photonics* **8**, 796 (2014).
- [11] V. Giovannetti, A. S. Holevo, and R. García-Patrón, *Commun. Math. Phys.* **334**, 1553 (2014).
- [12] A. Mari, V. Giovannetti, and A. S. Holevo, *Nat. Commun.* **5**, 3826 (2014).
- [13] V. Giovannetti, A. S. Holevo, and A. Mari, *Theor. Math. Phys.* **182**, 284 (2015).
- [14] 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).
- [15] T. Ogawa, Ph.D. dissertation, University of Electro-Communications, Tokyo, Japan, 2000 (in Japanese); T. Ogawa and H. Nagaoka, in *Proceedings of the 2002 IEEE International Symposium on Information Theory, Lausanne, Switzerland, 2002* (IEEE, New York, 2002), p. 73; T. Ogawa, *IEEE Trans. Inf. Theory* **45**, 2486 (1999).
- [16] T. Ogawa and H. Nagaoka, *IEEE Trans. Inf. Theory* **53**, 2261 (2007).
- [17] M. Hayashi and H. Nagaoka, *IEEE Trans. Inf. Theory* **49**, 1753 (2003).
- [18] M. Hayashi, *Phys. Rev. A* **76**, 062301 (2007); *Commun. Math. Phys.* **289**, 1087 (2009).
- [19] P. Hausladen and W. K. Wootters, *J. Mod. Opt.* **41**, 2385 (1994).
- [20] S. Lloyd, V. Giovannetti, and L. Maccone, *Phys. Rev. Lett.* **106**, 250501 (2011).
- [21] V. Giovannetti, S. Lloyd, and L. Maccone, *Phys. Rev. A* **85**, 012302 (2012).
- [22] P. Sen, [arXiv:1109.0802](https://arxiv.org/abs/1109.0802).
- [23] E. Arikani, *IEEE Trans. Inf. Theory* **55**, 3051 (2009).
- [24] M. M. Wilde, O. Landon-Cardinal, and P. Hayden, in *8th Conference on the Theory of Quantum Computation, Communication and Cryptography (TQC 2013), Dagstuhl, Germany, 2013*, edited by S. Severini and F. Brandao (Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, Dagstuhl, Germany, 2013), pp. 157–177.
- [25] M. M. Wilde and S. Guha, *IEEE Trans. Inf. Theory* **59**, 1175 (2013).
- [26] M. Rosati and V. Giovannetti, *J. Math. Phys.* **57**, 062204 (2016).
- [27] M. M. Wilde and S. Guha, in *Proceedings of the 2012 International Symposium on Information Theory and its Applications (ISITA2012), Honolulu, 2012* (IEICE, Tokyo, 2012), pp. 303–307.
- [28] M. M. Wilde, S. Guha, S.-H. Tan, and S. Lloyd, in *Proceedings of the 2012 IEEE International Symposium on Information Theory, ISIT 2012* (IEEE, Cambridge, MA, 2012), pp. 551–555.
- [29] S. Guha, J. L. Habif, and M. Takeoka, in *2010 IEEE International Symposium on Information Theory* (IEEE, Piscataway, NJ, 2010), pp. 2038–2042.
- [30] M. Takeoka and S. Guha, *Phys. Rev. A* **89**, 042309 (2014).
- [31] J. Lee, S.-W. Ji, J. Park, and H. Nha, *Phys. Rev. A* **93**, 050302(R) (2016).
- [32] S. Guha, *Phys. Rev. Lett.* **106**, 240502 (2011).
- [33] S. Guha, Z. Dutton, and J. H. Shapiro, in *IEEE International Symposium on Information Theory (ISIT)* (IEEE, Piscataway, NJ, 2011), p. 274.
- [34] A. Klimek, M. Jachura, W. Wasilewski, and K. Banaszek, *J. Mod. Opt.* **63**, 2074 (2016).
- [35] M. Rosati, A. Mari, and V. Giovannetti, *Phys. Rev. A* **94**, 062325 (2016).
- [36] C. Shannon, *IRE Trans. Inf. Theory* **2**, 8 (1956).
- [37] T. M. Cover and J. A. Thomas, *Elements of Information Theory* (Wiley-Interscience, Hoboken, NJ, 2006).
- [38] H. W. Chung, S. Guha, and L. Zheng, in *Proceedings of the 2011 IEEE International Symposium on Information Theory, St. Petersburg, 2011* (IEEE, Piscataway, NJ, 2011), pp. 284–288.
- [39] H. W. Chung, S. Guha, and L. Zheng, [arXiv:1610.07578](https://arxiv.org/abs/1610.07578).
- [40] P. W. Shor, *IBM J. Res. Dev.* **48**, 115 (2004).
- [41] The latter case of arbitrary interactions with the ancillae has been conjectured as well not to give any advantage in an updated version of the article by Chung *et al.* [39].
- [42] R. Kennedy, *MIT Res. Lab. Electron. Quart. Progr. Rep.* **108**, 219 (1973).
- [43] S. Dolinar, *MIT Res. Lab. Electron. Quart. Progr. Rep.* **111**, 115 (1973).