

# Exact Quantum Sensing Limits for Bosonic Dephasing Channels

Zixin Huang<sup>1,2,\*</sup>, Ludovico Lami<sup>3,4,5,†</sup> and Mark M. Wilde<sup>6,‡</sup>

<sup>1</sup>*School of Mathematical and Physical Sciences, Macquarie University, New South Wales 2109, Australia*

<sup>2</sup>*Centre for Quantum Software and Information, Faculty of Engineering and Information Technology, University of Technology Sydney, Sydney, New South Wales, Australia*

<sup>3</sup>*QuSoft, Science Park 123, Amsterdam 1098 XG, the Netherlands*

<sup>4</sup>*Korteweg–de Vries Institute for Mathematics, University of Amsterdam, Science Park 105-107, Amsterdam 1098 XG, the Netherlands*

<sup>5</sup>*Institute for Theoretical Physics, University of Amsterdam, Science Park 904, Amsterdam 1098 XH, the Netherlands*

<sup>6</sup>*School of Electrical and Computer Engineering, Cornell University, Ithaca, New York 14850, USA*



(Received 16 February 2024; accepted 8 May 2024; published 6 June 2024)

Dephasing is a prominent noise mechanism that afflicts quantum information carriers, and it is one of the main challenges toward realizing useful quantum computation, communication, and sensing. Here, we consider discrimination and estimation of bosonic dephasing channels, when using the most general adaptive strategies allowed by quantum mechanics. We reduce these difficult quantum problems to simple classical ones based on the probability densities defining the bosonic dephasing channels. By doing so, we rigorously establish the optimal performance of various distinguishability and estimation tasks and construct explicit strategies to achieve this performance. To the best of our knowledge, this is the first example of a non-Gaussian bosonic channel for which there are exact solutions for these tasks.

DOI: [10.1103/PRXQuantum.5.020354](https://doi.org/10.1103/PRXQuantum.5.020354)

## I. INTRODUCTION

An important channel to consider in the context of quantum technologies is the bosonic dephasing channel (BDC). A single-mode BDC  $\mathcal{D}_p$  is characterized by a probability density function  $p(\phi)$ , where  $\phi \in [-\pi, \pi]$  represents the random angle of phase-space rotation induced by the channel [1]. Accordingly, the action of  $\mathcal{D}_p$  on an input density operator  $\rho$  is given by

$$\mathcal{D}_p(\rho) := \int_{-\pi}^{\pi} d\phi p(\phi) e^{-i\hat{n}\phi} \rho e^{i\hat{n}\phi}, \quad (1)$$

where  $\hat{n}$  is the photon-number operator [2]. Dephasing is a major noise mechanism that afflicts quantum information carriers [3] and it is one of the main challenges toward realizing useful quantum computation, communication, and

sensing. In a dephasing noise process, the relative phase information between different photon-number components of a superposed state is lost; for quantum communication, such a process can be understood as arising from, e.g., temperature fluctuations of the environment that stretch or contract the length of a fiber [4]. As such, the problem of understanding the ultimate quantum limits for quantum information tasks using such channels has received considerable attention recently [1,5–12].

Two important tasks for characterizing the capabilities of BDCs are channel discrimination (quantum hypothesis testing) and parameter estimation (quantum metrology). For hypothesis testing, the task is to distinguish between models describing different physical processes. The most basic setting involves a binary decision, for which the goal is to distinguish between two hypotheses, commonly called the null hypothesis and the alternative hypothesis. Quantum state discrimination is crucial in several applications (e.g., quantum communication [13], astronomical sensing [14,15], and spectroscopy [16]) and it has been extensively studied [17,18]. Quantum channel discrimination, a generalization of state discrimination, has been studied less; however, there is an increasing body of literature on this topic [19–36]. In channel discrimination, the unknown channel is called  $n$  times and the goal is to perform a measurement on the final state to determine which

\*Corresponding author: [zixin.huang@mq.edu.au](mailto:zixin.huang@mq.edu.au)

†Corresponding author: [ludovico.lami@gmail.com](mailto:ludovico.lami@gmail.com)

‡Corresponding author: [wilde@cornell.edu](mailto:wilde@cornell.edu)

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/) license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

channel was called. This setting is more complex than state discrimination because one can optimize over various strategies in order to make the error probabilities as low as possible. One can either employ an adaptive strategy or a nonadaptive parallel strategy. It is known that adaptive strategies possess advantages over nonadaptive ones in the nonasymptotic regime [23,32]; however, these advantages come with limits [34,36]. In the setting of symmetric error, adaptive strategies can also outperform parallel ones in the asymptotic regime [35] but they do not in the asymptotic setting of asymmetric error [29–31,36].

Quantum metrology deals with the optimal estimation of parameters encoded in quantum states and quantum channels, and the typical goal is to minimize the variance of the parameter of interest. Quantum strategies for estimation involve nonclassical effects such as entanglement to achieve precision limits beyond those that are allowed by classical physics. The ultimate quantum limit for quantum metrology using adaptive strategies is in general exceedingly difficult to characterize, as the nested optimizations over each step of an adaptive protocol often lead to a mathematically intractable problem. When estimating parameters encoded into quantum channels, one can again consider parallel and adaptive strategies [37], with there being differences in their performance [38]. Previous works have considered bounds for parameter estimation of unitary channels [39–41], for teleportation-covariant and Gaussian channels [26,27], and for general channels [42].

In this paper, we consider several tasks associated with BDCs:

- (1) channel discrimination of two BDCs in the symmetric error setting of hypothesis testing
- (2) channel discrimination of two BDCs in the asymmetric error setting of hypothesis testing
- (3) noise parameter estimation of BDCs

We consider several other discrimination tasks and discuss the multimode generalization of our results in some of the appendixes. By making a connection to the physics (“environmental state”) that gives rise to the channel processes, for (1), (2), and (3), we quantify the largest distinguishability or estimability that can be realized between BDCs using the most general strategy achievable by adaptive protocols. We do so by proving optimality bounds and showing their attainability. In the latter case, we provide a fully rigorous proof of the convergence of the guessed probability density to the original one. To the best of our knowledge, our results here constitute the first example of a class of non-Gaussian bosonic channels for which we have exact solutions for these tasks.

The structure of our paper is as follows. In Secs. II A–II B, we introduce the theoretical frameworks of quantum channel discrimination and estimation. We present our main results in Sec. III; the optimality parts of

our proofs are given in Sec. V, and the attainability parts in Sec. VI. We show that in the energy-unconstrained limit, the corresponding figures of merit match, thus leading to exact solutions for channel discrimination and estimation tasks involving BDCs. In Sec. VII, we generalize our findings when photon loss is present in addition to dephasing, showing that the same fundamental limits apply. We finally conclude in Sec. VIII with a summary and some directions for future research. In Appendix E, we discuss a variety of other scenarios to which our results apply, including the strong converse exponent, the error exponent, multiple channel discrimination, and antidistinguishability. In Appendix F, we briefly touch upon how our results generalize to multimode BDCs and in Appendix G, how our findings generalize to multiparameter estimation.

## II. QUANTUM CHANNEL DISCRIMINATION AND ESTIMATION

### A. Quantum channel discrimination

The goal of quantum channel discrimination is to distinguish one quantum channel  $\mathcal{N}_0$  from another channel  $\mathcal{N}_1$  by calling the unknown channel  $n$  times. A strategy for doing so is abbreviated by  $\mathcal{A}$ , which denotes the initial state  $\tau$  prepared, the  $n - 1$  adaptive operations between every call to the unknown channel, and the final measurement, denoted by  $\mathcal{Q} := (Q_0, Q_1)$ . The conditions  $Q_0, Q_1 \geq 0$ , and  $Q_0 + Q_1 = I$  hold, so that  $\mathcal{Q}$  is a positive operator-valued measure (POVM). The measurement outcome  $Q_0$  corresponds to deciding  $\mathcal{N}_0$  and the outcome  $Q_1$  corresponds to deciding  $\mathcal{N}_1$ . An example of such a quantum channel discrimination protocol with  $n = 3$  is depicted in Figure 1. The type-I error probability is the probability of deciding  $\mathcal{N}_1$  when the actual channel is  $\mathcal{N}_0$  and the type-II error probability is the probability of deciding  $\mathcal{N}_0$  when the actual channel is  $\mathcal{N}_1$ . We denote these error probabilities, respectively, as follows:

$$\alpha_n(\mathcal{A}) := \text{Tr} \left[ Q_1 \rho_0^{(n)} \right], \quad \beta_n(\mathcal{A}) := \text{Tr} \left[ Q_0 \rho_1^{(n)} \right], \quad (2)$$

where  $\rho_0^{(n)}$  denotes the final state of the protocol in the case that the unknown channel is  $\mathcal{N}_0$  and  $\rho_1^{(n)}$  denotes the final state of the protocol in the case that the unknown channel is  $\mathcal{N}_1$ . Additionally, in the notations  $\alpha_n(\mathcal{A})$  and  $\beta_n(\mathcal{A})$ , we have left the dependence on the channels  $\mathcal{N}_0$  and  $\mathcal{N}_1$  implicit but we explicitly denote the dependence on the number  $n$  of calls to the unknown channel and the strategy  $\mathcal{A}$ . In radar applications,  $\alpha_n$  is known as the false-alarm probability and  $\beta_n$  is called the missed-detection probability [43]. For more details of quantum channel discrimination, see Ref. [29, Section III-A].

The difficulty of quantum channel discrimination arises because most operational quantities of interest involve

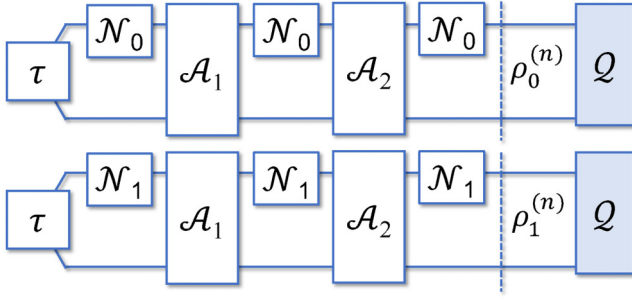


FIG. 1. A general adaptive protocol for channel discrimination and parameter estimation, when either  $\mathcal{N}_0$  or  $\mathcal{N}_1$  is called three times. The initial input state is  $\tau$ , the adaptive operations are  $\mathcal{A}_1$  and  $\mathcal{A}_2$ , and the final measurement is  $\mathcal{Q}$ . The final states are denoted by  $\rho_0^{(n)}$  and  $\rho_1^{(n)}$ , and  $n = 3$  in this case.

optimizations of these error probabilities over every possible adaptive strategy  $\mathcal{A}$ . For the finite-dimensional case, these optimizations can be phrased as semidefinite programs [32] but their computational complexity grows exponentially in  $n$ . For the infinite-dimensional case, the resulting optimization problem is an infinite-dimensional semidefinite program and, in general, is even more difficult to solve.

Classical hypothesis testing, on the other hand, is much simpler to handle computationally. In this scenario, a sample  $\phi$  is selected from a probability density function  $p$  or  $q$ , and the goal is to identify whether  $\phi$  came from  $p$  or  $q$ . This setting can be generalized to the multiple sample setting, in which  $n$  samples, abbreviated as  $\phi^n \equiv (\phi_1, \dots, \phi_n)$ , are selected from  $p^{\otimes n}$  or  $q^{\otimes n}$ . In order to decide whether  $p^{\otimes n}$  or  $q^{\otimes n}$  is the underlying density, the most general procedure one can perform on  $\phi^n$  is a randomized test  $t$  [44, Section 2.1], which is the classical version of a POVM (i.e.,  $t(\phi^n) \in [0, 1]$  is the probability of deciding  $p^{\otimes n}$  for all  $\phi^n$ ). The error probabilities of classical hypothesis testing are then given by

$$\alpha_n(t) := 1 - \int d\phi^n t(\phi^n) p^{\otimes n}(\phi^n), \quad (3)$$

$$\beta_n(t) := \int d\phi^n t(\phi^n) q^{\otimes n}(\phi^n), \quad (4)$$

where, in the notation for  $\alpha_n(t)$  and  $\beta_n(t)$ , we have left the dependence on the probability densities  $p$  and  $q$  implicit.

## B. Quantum channel estimation

The goal of quantum channel estimation is to estimate a parameter  $\theta$  encoded in a family  $(\mathcal{N}_\theta)_{\theta \in \Theta}$  of quantum channels. The main difference with quantum channel discrimination is that  $\theta$  is selected from a continuous set  $\Theta$  and thus we can only obtain an estimate of it, rather than identify it exactly. However, one can think of channel discrimination as being a special case of channel estimation

in which the set  $\Theta$  is a finite set consisting of just two elements (i.e.,  $\Theta = \{0, 1\}$  for channel discrimination).

The most general protocol for channel estimation is adaptive, similar to that discussed for channel discrimination. As such, we use the same notation  $\mathcal{A}$  to refer to an adaptive strategy for channel estimation. However, the main difference is that the final POVM of a channel-estimation protocol outputs an estimate  $\hat{\theta}$  of the unknown parameter  $\theta$  and, thus, the final POVM is of the form  $(Q_{\hat{\theta}})_{\hat{\theta} \in \Theta}$ , where  $Q_{\hat{\theta}} \geq 0$  for all  $\hat{\theta}$  and  $\int d\hat{\theta} Q_{\hat{\theta}} = I$ . Denoting the final state of an  $n$ -shot protocol by  $\rho_\theta^{(n)}$  whenever the underlying channel is  $\mathcal{N}_\theta$ , the conditional probability density of observing  $\hat{\theta}$  is  $\text{Tr}[Q_{\hat{\theta}} \rho_\theta^{(n)}]$ .

To quantify the performance of a channel-estimation protocol, we employ a cost function  $c(\hat{\theta}, \theta)$  that measures the deviation of the estimate  $\hat{\theta}$  from the true value  $\theta$ . The basic properties for such a cost function are as follows [45, Section 2.1]:

- (1)  $c(\hat{\theta}, \theta) = 0$  if  $\hat{\theta} = \theta$
- (2)  $c(\hat{\theta}, \theta) = c(\theta, \hat{\theta})$
- (3)  $c(\hat{\theta}', \theta) \geq c(\hat{\theta}, \theta)$  for  $\hat{\theta}' \geq \hat{\theta} \geq \theta$
- (4)  $c(\hat{\theta}, \theta)$  is not identically equal to zero
- (5) for some constants  $k, a > 0$ , the following bound holds:  $c(\hat{\theta}, \theta) \leq k(1 + |\hat{\theta} - \theta|^a)$  for all  $\hat{\theta}, \theta$

Beyond basic properties expected for such a function, we require it to be continuous. Common choices when  $\Theta = \mathbb{R}$  include the absolute deviation  $c(\hat{\theta}, \theta) = |\hat{\theta} - \theta|$  and quadratic cost  $c(\hat{\theta}, \theta) = (\hat{\theta} - \theta)^2$ . We then define the risk of an  $n$ -round adaptive strategy  $\mathcal{A}$  to be the expected cost:

$$r_n(\theta, \mathcal{A}) := \int d\hat{\theta} \text{Tr}[Q_{\hat{\theta}} \rho_\theta^{(n)}] c(\hat{\theta}, \theta). \quad (5)$$

As with channel discrimination, in the notation  $r_n(\theta, \mathcal{A})$ , we leave the dependence on  $\mathcal{N}_\theta$  implicit. If we choose  $c(\hat{\theta}, \theta) = (\hat{\theta} - \theta)^2$ , the expected cost is the mean squared error or variance, which is the standard error metric for parameter estimation. The difficulty in channel estimation also lies in the complexity of adaptive strategies that can be considered.

Classical parameter estimation is again simpler. In this scenario, there is a parametrized family  $(p_\theta)_{\theta \in \Theta}$  of probability densities. Given access to  $n$  independent samples selected from  $p_\theta^{\otimes n}$ , labeled by  $\phi^n$ , an estimate  $\hat{\theta}$  is output according to the conditional probability density  $t(\hat{\theta}|\phi^n)$ . Then, the conditional probability density for outputting  $\hat{\theta}$  is

$$s(\hat{\theta}|\theta) := \int d\phi^n t(\hat{\theta}|\phi^n) p_\theta^{\otimes n}(\phi^n) \quad (6)$$

and thus the risk of a classical strategy  $t$  for estimating  $\theta$  is

$$r_n(\theta, t) := \int d\hat{\theta} s(\hat{\theta}|\theta) c(\hat{\theta}, \theta). \quad (7)$$

Here, in the notation  $r_n(\theta, t)$ , we again leave the dependence on the underlying probability density  $p_\theta$  implicit.

### III. MAIN RESULTS

One of the main insights of our paper is that, when restricting the underlying channels to BDCs  $\mathcal{D}_p$  and  $\mathcal{D}_q$  of the form in Eq. (1) and with respective probability densities  $p$  and  $q$ , various optimized functions of the error probabilities  $\alpha_n(\mathcal{A})$  and  $\beta_n(\mathcal{A})$ , as defined in Eq. (2), are equal to the same optimized functions of the classical error probabilities  $\alpha_n(t)$  and  $\beta_n(t)$ , as defined in Eqs. (3)–(4). Our first main result is that the following equality holds for all  $\lambda \in (0, 1)$ :

$$\begin{aligned} & \inf_{\mathcal{A}} \{ \lambda \alpha_n(\mathcal{A}) + (1 - \lambda) \beta_n(\mathcal{A}) \} \\ &= \inf_t \{ \lambda \alpha_n(t) + (1 - \lambda) \beta_n(t) \}, \end{aligned} \quad (8)$$

where the optimization on the left is over all adaptive strategies and that on the right is over all randomized tests (we adopt this same abbreviated notation in all related statements that follow). Our second main result is that the following equality holds for all  $\varepsilon \in (0, 1)$ :

$$\inf_{\mathcal{A}} \{ \beta_n(\mathcal{A}) : \alpha_n(\mathcal{A}) \leq \varepsilon \} = \inf_t \{ \beta_n(t) : \alpha_n(t) \leq \varepsilon \}. \quad (9)$$

In Appendix A, we prove that these claims are actually a consequence of the fact that the sets of achievable pairs,  $\{(\alpha_n(\mathcal{A}), \beta_n(\mathcal{A}))\}_{\mathcal{A}}$  and  $\{(\alpha_n(t), \beta_n(t))\}_t$ , coincide.

A related insight of our paper concerns parametrized families  $(\mathcal{D}_{p_\theta})_{\theta \in \Theta}$  of BDCs, for which the probability density  $p_\theta$  underlying  $\mathcal{D}_{p_\theta}$  is parametrized. In this case, our third main result is that the following equality holds for an arbitrary risk function for which the underlying cost function is continuous:

$$\inf_{\mathcal{A}} r_n(\theta, \mathcal{A}) = \inf_t r_n(\theta, t), \quad (10)$$

where, similar to Eq. (8), the optimization on the left is over all adaptive strategies for channel estimation of  $\mathcal{D}_{p_\theta}$  and the optimization on the right is over all classical estimation strategies for  $p_\theta$ .

The equalities in Eqs. (8)–(10) are some of the main results of our paper and as we will see, they give a complete understanding of the fundamental limits of channel discrimination and parameter estimation for BDCs. The equalities in Eqs. (8) and (9) thus represent a significant reduction in the difficulty of channel discrimination for BDCs, i.e., reducing it to a classical hypothesis testing

problem, for which there is a wealth of knowledge that we can apply. A similar statement applies for Eq. (10) and channel estimation of BDCs. In the subsections that follow, we discuss various scenarios of interest in more detail.

#### A. Symmetric hypothesis testing

In the symmetric setting of channel discrimination, the goal is to find an optimal strategy that attains the minimum average error probability, defined as

$$\inf_{\mathcal{A}} \{ \lambda \alpha_n(\mathcal{A}) + (1 - \lambda) \beta_n(\mathcal{A}) \}, \quad (11)$$

where  $\lambda \in (0, 1)$  is the prior probability that channel  $\mathcal{N}_0$  is selected and  $\alpha_n(\mathcal{A})$  and  $\beta_n(\mathcal{A})$  are defined in Eq. (2). Our result here establishes, for BDCs  $\mathcal{D}_p$  and  $\mathcal{D}_q$ , that the equality in Eq. (8) holds. As a consequence of a well known result from statistics (see, e.g., Ref. [46, Lemma 1.4]), we have the following explicit form for the error probability:

$$\begin{aligned} & \inf_t \{ \lambda \alpha_n(t) + (1 - \lambda) \beta_n(t) \} \\ &= \frac{1}{2} (1 - \| \lambda p^{\otimes n} - (1 - \lambda) q^{\otimes n} \|_1), \end{aligned} \quad (12)$$

where  $\|f\|_1 := \int d\phi^n |f(\phi^n)|$  is the  $\ell_1$ -norm of a function  $f$ .

The nonasymptotic error exponent for channel discrimination in the symmetric setting is defined as

$$C_n(\lambda, \mathcal{N}_0, \mathcal{N}_1) := \sup_{\mathcal{A}} \left\{ -\frac{1}{n} \ln (\lambda \alpha_n(\mathcal{A}) + (1 - \lambda) \beta_n(\mathcal{A})) \right\}. \quad (13)$$

By appealing to the Chernoff theorem from probability theory [47] and defining the Chernoff divergence of two probability densities  $p$  and  $q$  as

$$C(p\|q) := -\ln \inf_{s \in [0, 1]} \int_{-\pi}^{\pi} d\phi p(\phi)^s q(\phi)^{1-s}, \quad (14)$$

our result implies the following simple expression for the asymptotic error exponent of BDCs  $\mathcal{D}_p$  and  $\mathcal{D}_q$ :

$$\lim_{n \rightarrow \infty} C_n(\lambda, \mathcal{D}_p, \mathcal{D}_q) = C(p\|q). \quad (15)$$

This is because

$$\begin{aligned} & \lim_{n \rightarrow \infty} C_n(\lambda, \mathcal{D}_p, \mathcal{D}_q) \\ &= \lim_{n \rightarrow \infty} \sup_t \left\{ -\frac{1}{n} \ln (\lambda \alpha_n(t) + (1 - \lambda) \beta_n(t)) \right\} \\ &= C(p\|q), \end{aligned} \quad (16)$$

where the first equality follows from Eqs. (8) and (13) and the second from Chernoff's theorem.

## B. Asymmetric hypothesis testing

In the asymmetric setting of channel discrimination, the goal is to minimize the type-II error probability subject to a constraint on the type-I error probability. One example is in radar applications or quantum illumination [48–50], where one is willing to tolerate a certain rate of false alarms but then desires to minimize the chances of missed detections [51]. Indeed, the asymmetric error setting is the right setting to focus on for such applications and leads to a curve known as the receiver operating characteristic [43].

More formally, for  $\varepsilon \in (0, 1)$ , the goal is to optimize over every adaptive strategy  $\mathcal{A}$  in order to minimize the type-II error probability  $\beta_n$ :

$$\inf_{\mathcal{A}} \{\beta_n(\mathcal{A}) : \alpha_n(\mathcal{A}) \leq \varepsilon\}, \quad (18)$$

where  $\alpha_n(\mathcal{A})$  and  $\beta_n(\mathcal{A})$  are defined in Eq. (2). Our result here is that, for BDCs  $\mathcal{D}_p$  and  $\mathcal{D}_q$ , the equality in Eq. (9) holds, where  $\alpha_n(t)$  and  $\beta_n(t)$  are defined in Eqs. (3) and (4).

The nonasymptotic error exponent for channel discrimination in the asymmetric setting is defined as

$$Z_n(\varepsilon, \mathcal{N}_0, \mathcal{N}_1) := \sup_{\mathcal{A}} \left\{ -\frac{1}{n} \ln \beta_n(\mathcal{A}) : \alpha_n(\mathcal{A}) \leq \varepsilon \right\}. \quad (19)$$

By appealing to Strassen's theorem [52, Theorem 3.1], itself a refinement of the classical Stein's lemma [53, 54], our result implies the following expansion of  $Z_n(\varepsilon, \mathcal{D}_p, \mathcal{D}_q)$ :

$$\begin{aligned} Z_n(\varepsilon, \mathcal{D}_p, \mathcal{D}_q) \\ = D(p \| q) + \sqrt{\frac{V(p \| q)}{n}} \Phi^{-1}(\varepsilon) + \frac{\ln n}{2n} + O\left(\frac{1}{n}\right). \end{aligned} \quad (20)$$

(see also Ref. [44, Proposition 2.3]). In the above, the relative entropy, the relative entropy variance, and the inverse cumulative distribution function of a standard normal random variable are, respectively, defined as

$$D(p \| q) := \int_{-\pi}^{\pi} d\phi p(\phi) \ln \left( \frac{p(\phi)}{q(\phi)} \right), \quad (21)$$

$$V(p \| q) := \int_{-\pi}^{\pi} d\phi p(\phi) \left[ \ln \left( \frac{p(\phi)}{q(\phi)} \right) - D(p \| q) \right]^2, \quad (22)$$

$$\Phi^{-1}(\varepsilon) := \sup \{a \in \mathbb{R} : \Phi(a) \leq \varepsilon\}, \quad (23)$$

where

$$\Phi(a) := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^a dx \exp(-x^2/2). \quad (24)$$

The equality in Eq. (20) follows from the equality in Eq. (9) and from the following expansion, a direct application of Strassen's theorem:

$$\begin{aligned} \sup_t \left\{ -\frac{1}{n} \ln \beta_n(t) : \alpha_n(t) \leq \varepsilon \right\} \\ = D(p \| q) + \sqrt{\frac{V(p \| q)}{n}} \Phi^{-1}(\varepsilon) + \frac{\ln n}{2n} + O\left(\frac{1}{n}\right). \end{aligned} \quad (25)$$

Note that  $\Phi^{-1}(\varepsilon) < 0$  for  $\varepsilon < 1/2$  and  $\Phi^{-1}(\varepsilon) > 0$  for  $\varepsilon > 1/2$ . As such, by inspecting Eq. (20), we see that, as a function of the number  $n$  of channel uses, the error exponent  $Z_n(\varepsilon, \mathcal{D}_p, \mathcal{D}_q)$  approaches the optimal asymptotic value  $D(p \| q)$  from below for  $\varepsilon < 1/2$  and from above for  $\varepsilon > 1/2$ , at a speed determined by the relative entropy variance  $V(p \| q)$ . The regime of practical interest occurs when  $\varepsilon < 1/2$ , for which the error exponent  $Z_n(\varepsilon, \mathcal{D}_p, \mathcal{D}_q)$  thus approaches  $D(p \| q)$  from below.

## C. Quantum metrology

In the setting of quantum metrology (an umbrella term containing quantum channel estimation), the goal is to minimize the risk over all possible adaptive strategies. Our main result for channel estimation, which applies to a continuous family  $(\mathcal{D}_{p_\theta})_\theta$  of BDCs, states that the equality in Eq. (10) holds. In the case that the underlying cost function is the quadratic cost function and restricting the optimization to be over adaptive strategies that lead to an unbiased estimator, a consequence of our finding and the classical Cramér-Rao bound is the following inequality:

$$\inf_{\mathcal{A}} r_n(\theta, \mathcal{A}) \geq \frac{1}{nF(\theta)}, \quad (26)$$

where the Fisher information of the parametrized family  $(p_\theta)_\theta$  is defined as follows:

$$F(\theta) := \int_{-\pi}^{\pi} d\phi p_\theta(\phi) \left( \frac{d}{d\theta} \ln p_\theta(\phi) \right)^2. \quad (27)$$

The inequality in Eq. (26) follows directly from Eq. (10) and the well known Cramér-Rao bound  $\inf_t r_n(\theta, t) \geq [nF(\theta)]^{-1}$  [45, Corollary 1.9].

## IV. EXAMPLES OF BOSONIC DEPHASING CHANNELS

To gain some intuition about our findings, let us consider the specific examples of BDCs studied in Ref. [1]. We plot the Chernoff divergence, the relative entropy, and the Fisher information of certain instances of these channels, which are the main quantities of interest in

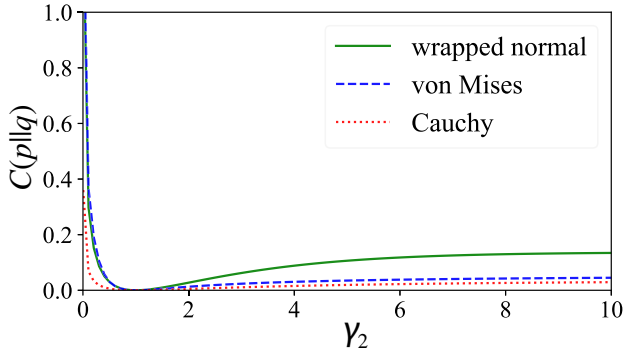


FIG. 2. The Chernoff exponent of different BDCs as a function of  $\gamma_2 = \lambda_2 = \kappa_2$ , for  $\gamma_1 = \lambda_1 = \kappa_1 = 1$ .

the asymptotic settings of symmetric channel discrimination, asymmetric channel discrimination, and channel estimation, respectively.

As stressed in Ref. [1] and previous works [5,6], perhaps the most important class of BDCs are those resulting from setting the probability density  $p(\phi)$  in Eq. (1) to be the wrapped normal distribution:

$$p_\gamma(\phi) := \frac{1}{\sqrt{2\pi\gamma}} \sum_{k=-\infty}^{+\infty} e^{-\frac{1}{2\gamma}(\phi+2\pi k)^2}, \quad (28)$$

where  $\gamma > 0$  is the variance. This probability density results from picking  $\phi$  according to a mean-zero normal distribution of variance  $\gamma$  but then outputting a value in  $[-\pi, \pi]$  modulo  $2\pi$ . Physically, as considered in Refs. [5,6], it corresponds to interacting the channel input mode with an environmental mode prepared in the vacuum state, according to the Hamiltonian  $\hat{n} \otimes (\hat{e} + \hat{e}^\dagger)$ , where  $\hat{e}$  is the annihilation operator for the environmental mode. It can alternatively be realized in terms of a Lindbladian evolution for a time  $\gamma$  according to the single Lindblad operator  $\hat{n}$ .

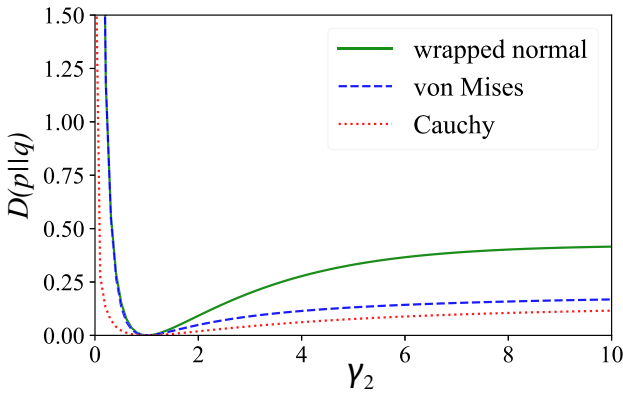


FIG. 3. The relative entropy of different BDCs as a function of  $\gamma_2 = \lambda_2 = \kappa_2$ , for  $\gamma_1 = \lambda_1 = \kappa_1 = 1$ .

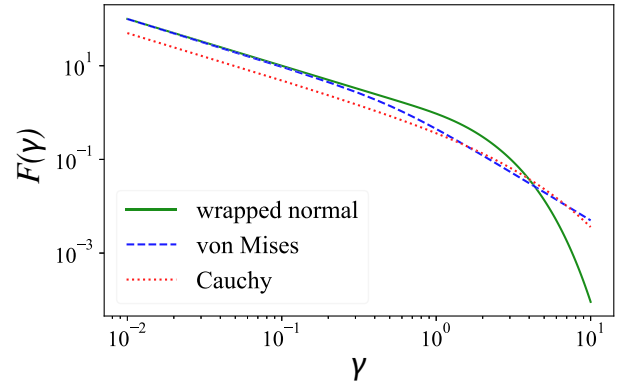


FIG. 4. The Fisher information of different BDCs as a function of  $\gamma = \kappa = \lambda$ , shown as a log-log plot. The Fisher information quickly drops to zero as  $\gamma$  increases.

Another probability density of interest for the BDC is based on the von Mises distribution:

$$p_\lambda(\phi) := \frac{e^{\cos(\phi)/\lambda}}{2\pi I_0(1/\lambda)}, \quad (29)$$

where  $I_n$  denotes a modified Bessel function of the first kind. The parameter  $\lambda$  determines the spread of the distribution, analogous to  $\gamma$  for the wrapped normal. For  $\lambda \rightarrow \infty$ , it converges to the uniform density, while it becomes highly peaked at zero in the limit  $\lambda \rightarrow 0$ .

The final circular distribution that we consider is the wrapped Cauchy distribution, given by

$$p_\kappa(\phi) := \frac{1}{2\pi} \frac{\sinh \sqrt{\kappa}}{\cosh \sqrt{\kappa} - \cos(\phi)}. \quad (30)$$

The parameter  $\kappa > 0$  again determines the spread of the distribution.

Figure 2 plots the Chernoff divergence of a pair of BDCs for each kind of distribution, with one spread parameter fixed at a value of  $\gamma = \lambda = \kappa = 1$  and the other spread parameter varied. Figure 3 does the same for the relative entropy. Figure 4 plots the Fisher information of the underlying channel parameter  $\gamma$ ,  $\lambda$ , or  $\kappa$ . We find similar qualitative behavior for all three kinds of probability densities.

## V. OPTIMALITY

In this section, we prove one side of the equalities in Eqs. (8)–(10) (called the “optimality” part), based on a simple observation about all BDCs of the form in Eq. (1). Namely, they can be simulated by the method discussed in Ref. [55, Section 3.3], that is, by means of adjoining a parametrized environment state followed by the action of an unparametrized channel. After Ref. [55] appeared, similar observations were made for other channels in several subsequent works, including Refs. [26,27,29,41], and

here our contribution is to make a similar observation for BDCs. Namely, all BDCs can be simulated by composing the following two processes:

- (1) A classical background phase  $\phi$  is chosen randomly according to the probability density  $p(\phi)$  in Eq. (1).
- (2) The input system has the phase operator  $e^{-i\hbar\phi}$  applied to it, based on the value of  $\phi$  chosen, and the value  $\phi$  is subsequently discarded.

More formally, first we define an environment state  $\sigma_p$  that encodes the probability density  $p(\phi)$  in Eq. (1) as follows:

$$\sigma_p := \int_{-\pi}^{\pi} d\phi p(\phi) |\phi\rangle\langle\phi|, \quad (31)$$

where, in the physics literature,  $\{|\phi\rangle\}_\phi$  is usually interpreted as a set of ‘‘eigenkets’’ obeying the ‘‘orthogonality relation’’  $\langle\phi'|\phi\rangle = \delta(\phi - \phi')$  (i.e.,  $\{|\phi\rangle\}_\phi$  can be seen as an orthogonal basis for the phase  $\phi$ ). We may also interpret Eq. (31) as a representation of a random variable on  $[-\pi, \pi]$  with probability density  $p$ . Then,  $\mathcal{D}_p$  decomposes as

$$\mathcal{D}_p = \mathcal{G} \circ \mathcal{F}_p, \quad (32)$$

where

$$\mathcal{F}_p(\rho) := \rho \otimes \sigma_p, \quad (33)$$

$$\mathcal{G}(\rho' \otimes \rho'') := \int_{-\pi}^{\pi} d\phi' e^{-i\hbar\phi'} \rho' e^{i\hbar\phi'} \text{Tr}[|\phi'\rangle\langle\phi'|\rho'']. \quad (34)$$

The first channel  $\mathcal{F}_p$  appends the environment state  $\sigma_p$  to the input state  $\rho$ , while  $\mathcal{G}$  measures  $\sigma_p$  and, based on the measured phase  $\phi$ , applies the unitary phase operator  $e^{-i\hbar\phi}$  to  $\rho$ . The action of  $\mathcal{D}_p$  on an arbitrary input state  $\rho$  is thus as follows:

$$\mathcal{D}_p(\rho) = \mathcal{G}(\rho \otimes \sigma_p). \quad (35)$$

The implications of the composition in Eq. (32) are far reaching, indeed leading to our optimality bounds. The idea is that when we decompose the channel this way, we can ‘‘pull back’’ the environmental state from our analysis of an adaptive strategy  $\mathcal{A}$ , as depicted in Figure 5. Then, a quantum channel discrimination or estimation task can be recast as a classical state discrimination or estimation task, respectively. Given that the operations in the adaptive strategy  $\mathcal{A}$  composed with  $n$  instances of the channel  $\mathcal{G}$  are independent of  $p$  and  $q$ , the optimality of the distinguishability task is then limited by the distinguishability of the environmental states. More formally, every  $n$ -round adaptive strategy  $\mathcal{A}$  for channel discrimination, when applied to the BDC  $\mathcal{D}_p$ , can be composed with

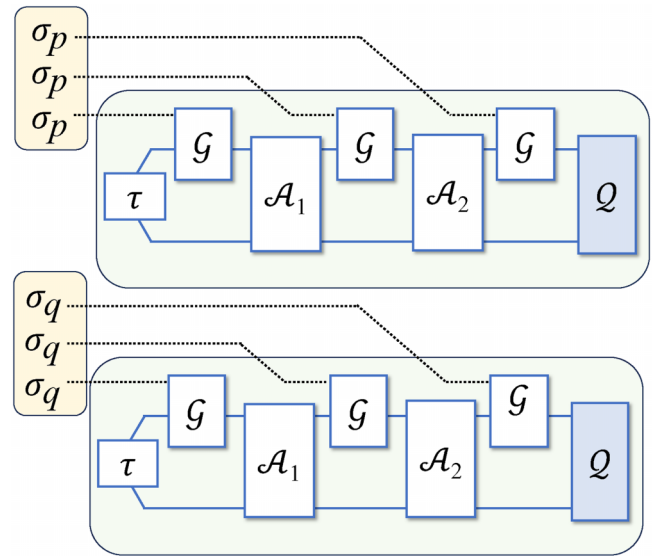


FIG. 5. Channel discrimination and parameter estimation for environment-parametrized channels  $\mathcal{D}_p$  and  $\mathcal{D}_q$ , where the underlying environment states are  $\sigma_p$  and  $\sigma_q$ , respectively. The yellow-shaded boxes denote the underlying environmental states to which we do not have access.

$n$  calls of the  $p$ -independent channel  $\mathcal{G}$  to view the resulting strategy as a particular classical test  $t$  performed on the probability density  $p^{\otimes n}$ . This reasoning then implies the following inequality, which holds for every adaptive strategy  $\mathcal{A}$  applied to BDCs  $\mathcal{D}_p$  and  $\mathcal{D}_q$ :

$$\lambda\alpha_n(\mathcal{A}) + (1 - \lambda)\beta_n(\mathcal{A}) \geq \inf_t \{\lambda\alpha_n(t) + (1 - \lambda)\beta_n(t)\}. \quad (36)$$

Since this inequality holds for every adaptive strategy  $\mathcal{A}$ , we conclude the inequality ‘‘ $\geq$ ’’ in Eq. (8) for BDCs  $\mathcal{D}_p$  and  $\mathcal{D}_q$ . Furthermore, for every adaptive strategy  $\mathcal{A}$  satisfying  $\alpha_n(\mathcal{A}) \leq \varepsilon$ , we conclude, by the same reasoning, for BDCs  $\mathcal{D}_p$  and  $\mathcal{D}_q$  that

$$\beta_n(\mathcal{A}) \geq \inf_t \{\beta_n(t) : \alpha_n(t) \leq \varepsilon\}. \quad (37)$$

Since this inequality holds for every adaptive strategy  $\mathcal{A}$  satisfying  $\alpha_n(\mathcal{A}) \leq \varepsilon$ , we conclude the inequality ‘‘ $\geq$ ’’ in Eq. (9) for BDCs  $\mathcal{D}_p$  and  $\mathcal{D}_q$ .

The same reasoning applies for channel estimation, with respect to a continuous family  $(\mathcal{D}_{p_\theta})_\theta$  of parametrized BDCs. Indeed, every  $n$ -round adaptive strategy  $\mathcal{A}$  for channel estimation can be composed with  $n$  calls of the  $\theta$ -independent channel  $\mathcal{L}$  to view the resulting strategy as a particular randomized estimator  $t$  performed on the probability density  $p_\theta^{\otimes n}$ . We then conclude the following inequality for every continuous family  $(\mathcal{D}_{p_\theta})_\theta$  of

parametrized BDCs:

$$r_n(\theta, \mathcal{A}) \geq \inf_t r_n(\theta, t). \quad (38)$$

Since this inequality holds for every adaptive strategy  $\mathcal{A}$ , we conclude the inequality “ $\geq$ ” in Eq. (10) for every continuous family  $(\mathcal{D}_{p_\theta})_\theta$  of parametrized BDCs.

## VI. ATTAINABILITY

We now prove the other side (“attainability”) of the equalities in Eqs. (8)–(10). Here, again, the basic principle behind our reasoning is simple. As we will show, for a BDC  $\mathcal{D}_p$ , it is possible to input a sequence  $(\rho_\nu)_{\nu \in \mathbb{N}}$  of states to it and perform a POVM  $(M_\phi)_\phi$  such that, for all  $\phi \in [-\pi, \pi]$ ,

$$p(\phi) = \lim_{\nu \rightarrow \infty} p_\nu(\phi), \quad (39)$$

where

$$p_\nu(\phi) := \text{Tr}[M_\phi \mathcal{D}_p(\rho_\nu)]. \quad (40)$$

In the above formulation, we have used  $\nu$  as an abstract index for a sequence of states. In Secs. VI A–VI B, we provide concrete examples in which  $\nu$  is replaced by photon number or used as an index for a sequence of coherent states with increasing energy. A channel satisfying Eqs. (39) and (40) is said to be environment seizable [29, Definition 36] because it is possible to perform pre- and postprocessing of the channel in order to “seize” the background environment state. In this case, we can recover the probability density  $p(\phi)$ , characterizing a BDC  $\mathcal{D}_p$ , exactly in the  $\nu \rightarrow \infty$  limit and process it directly. It is similarly possible to do this for all  $n \in \mathbb{N}$  because, for all  $\phi^n \in [-\pi, \pi]^n$ ,

$$p^{\otimes n}(\phi^n) = \lim_{\nu \rightarrow \infty} p_\nu^{\otimes n}(\phi^n), \quad (41)$$

as a direct consequence of Eq. (39). Thus, a particular sequence of strategies for channel discrimination is to input the state  $\rho_\nu$  to every channel use, followed by the measurement  $M_\phi$ , leading to the density  $p_\nu^{\otimes n}(\phi)$ . We then process the resulting densities with a classical test  $t$ . As we will see shortly, such a sequence of strategies is optimal in the limit  $\nu \rightarrow \infty$ .

In the case that Eq. (39) holds, it directly follows that the type-I and type-II error probabilities under an arbitrary

test  $t$  obey the following equalities:

$$\alpha_n(t) = \lim_{\nu \rightarrow \infty} \alpha_{n,\nu}(t), \quad \beta_n(t) = \lim_{\nu \rightarrow \infty} \beta_{n,\nu}(t), \quad (42)$$

where

$$\alpha_{n,\nu}(t) := 1 - \int d\phi^n t(\phi^n) p_\nu^{\otimes n}(\phi^n), \quad (43)$$

$$\beta_{n,\nu}(t) := \int d\phi^n t(\phi^n) q_\nu^{\otimes n}(\phi^n), \quad (44)$$

and with  $q_\nu(\phi)$  defined as in Eq. (40) but with  $\mathcal{D}_p$  replaced by  $\mathcal{D}_q$ .

As a consequence of Eq. (42), if we can show that there exists a sequence  $(\rho_\nu)_{\nu \in \mathbb{N}}$  of states and a measurement  $M_\phi$  such that Eq. (39) holds, then the desired attainability claims hold because the aforementioned strategy is a particular kind of adaptive strategy  $\mathcal{A}$ ; that is, for every test  $t$  and for every test  $t'$  such that  $\alpha_n(t') \leq \varepsilon$ ,

$$\begin{aligned} \inf_{\mathcal{A}} \{ \lambda \alpha_n(\mathcal{A}) + (1 - \lambda) \beta_n(\mathcal{A}) \} &\leq \lambda \alpha_n(t) + (1 - \lambda) \beta_n(t), \\ \inf_{\mathcal{A}} \{ \beta_n(\mathcal{A}) : \alpha_n(\mathcal{A}) \leq \varepsilon \} &\leq \beta_n(t'). \end{aligned} \quad (45)$$

Since the inequalities hold for every test  $t$  and for every test  $t'$  such that  $\alpha_n(t') \leq \varepsilon$ , we conclude that the same inequalities hold with infima taken on the right-hand side. Combining this claim with the optimality results from the previous section concludes the proof of the desired equalities in Eqs. (8) and (9).

We can make similar conclusions for channel estimation for a continuous family  $(\mathcal{D}_{p_\theta})_\theta$  of parametrized BDCs. Indeed, in the case that Eq. (39) holds and the estimator  $t$  satisfies

$$\int d\hat{\theta} \int d\phi^n t(\hat{\theta}|\phi^n) c(\hat{\theta}, \theta) < \infty, \quad (46)$$

a simple application of Lebesgue’s dominated convergence theorem shows that

$$r_n(\theta, t) = \lim_{\nu \rightarrow \infty} r_{n,\nu}(\theta, t), \quad (47)$$

where

$$r_{n,\nu}(\theta, t) := \int d\hat{\theta} \int d\phi^n t(\hat{\theta}|\phi^n) p_{\theta,\nu}^{\otimes n}(\phi^n) c(\hat{\theta}, \theta), \quad (48)$$

$$p_{\theta,\nu}(\phi) := \text{Tr}[M_\phi \mathcal{D}_{p_\theta}(\rho_\nu)]. \quad (49)$$

As a consequence of Eq. (48), and similar reasoning used above in Eq. (45), we conclude the following attainability inequality:

$$\inf_{\mathcal{A}} r_n(\theta, \mathcal{A}) \leq \inf_t r_n(\theta, t), \quad (50)$$

which thus finishes the proof of our main channel estimation result in Eq. (10).



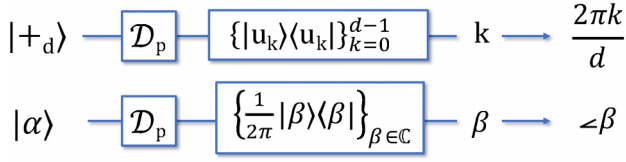


FIG. 6. Two measurement methods that achieve the optimality bounds. The method on the top (see Sec. VIA) involves preparing a uniform superposition of photon-number states, transmitting through the BDC  $\mathcal{D}_p$ , and performing a Fourier transform, followed by photodetection and classical postprocessing. The method on the bottom (see Sec. VIB) involves preparing a coherent state  $|\alpha\rangle$ , where  $\alpha \in \mathbb{R}_+$ , transmitting through the BDC  $\mathcal{D}_p$ , and performing heterodyne detection, followed by classical postprocessing.

In the two subsections that follow, we exhibit two specific schemes for which the needed equality in Eq. (39) holds. Moreover, the methods are simple to describe in physical terms, involving either (1) preparation of a uniform superposition of photon-number states at the input and a quantum Fourier transform followed by photon-number measurement and classical postprocessing at the output or (2) preparation of a coherent state at the input and heterodyne detection followed by classical postprocessing at the output. For a visual depiction of the two methods, see Fig. 6. The latter method is robust to loss in the channel in addition to dephasing, due to the fact that the purity of coherent states is retained under a pure-loss channel (for more discussions of this point, see Sec. VII). The first scheme is similar to that introduced in Ref. [56] and the measurement used in the first scheme can be considered an approximation of the canonical phase measurement, also discussed in Ref. [57]. The second scheme has been considered in Ref. [57].

### A. Photon-number-superposition method

As stated above, this method involves preparing a uniform superposition of photon-number states at the input and performing a quantum Fourier transform, followed by photon detection and classical postprocessing at the output. Photon-number superposition states have been well investigated in the context of optical phase estimation (see Refs. [58–62]). The particular input state and measurement basis that we use have been proposed in Ref. [63] and the scheme we consider below has some similarities with that proposed in Ref. [56] and is identical to that given in Refs. [64,65].

Let us begin by defining a  $d$ -level uniform superposition of photon-number states:

$$|+_d\rangle := \frac{1}{\sqrt{d}} \sum_{n=0}^{d-1} |n\rangle, \quad (51)$$

where  $|n\rangle$  is a photon-number state [2]. A property of  $|+_d\rangle$  is that phases become encoded into it as follows:

$$e^{-i\hat{n}\phi} |+_d\rangle = \frac{1}{\sqrt{d}} \sum_{n=0}^{d-1} e^{-in\phi} |n\rangle. \quad (52)$$

Such encoded phases can be recovered approximately by performing a measurement in the Fourier basis, which is defined for all  $k \in \{0, \dots, d-1\}$  as

$$|u_k\rangle := \frac{1}{\sqrt{d}} \sum_{n=0}^{d-1} e^{-2\pi i kn/d} |n\rangle. \quad (53)$$

Indeed, as shown in Appendix B, we find that the probability of measuring  $k \in \{0, \dots, d-1\}$  is as follows:

$$\begin{aligned} p_d(k|\phi) &:= \left| \langle u_k | e^{-i\hat{n}\phi} |+_d\rangle \right|^2 \\ &= \frac{1}{d^2} \frac{\sin^2\left(\pi k - \frac{d\phi}{2}\right)}{\sin^2\left(\frac{\pi k}{d} - \frac{\phi}{2}\right)}. \end{aligned} \quad (54)$$

The function on the last line above is proportional to the Fejér kernel, a well-known object in Fourier analysis [66]; as a function of  $\phi$ , it is peaked at  $\phi = 2\pi k/d$ . By classical postprocessing applied to the value  $2\pi k/d$  (i.e., performing the shift  $2\pi k/d - 2\pi$  if  $2\pi k/d \in (\pi, 2\pi)$ ), adding uniform noise chosen from the interval  $[0, 2\pi/d]$  of size  $2\pi/d$ , the discrete probability distribution  $p_d(k|\phi)$  can be smoothed into a continuous probability density. Mathematically, this corresponds to the probability mass function  $p_d(k|\phi)$  being convolved with a rectangle function  $\Pi_d(x)$  associated with a uniform density of width  $2\pi/d$ , leading to the probability density

$$p_d(\hat{\phi}|\phi) := \sum_{k=0}^{d-1} p_d(k|\phi) \Pi_d\left(\hat{\phi} - \frac{2\pi k}{d} \bmod 2\pi\right), \quad (55)$$

where  $x \bmod 2\pi := \min\{x + 2\pi k : x + 2\pi k \geq 0, k \in \mathbb{Z}\}$ , and  $\Pi_d(x)$  is defined as

$$\Pi_d(x) := \begin{cases} 0 & : x < 0 \text{ or } x \geq \frac{2\pi}{d}, \\ \frac{d}{2\pi} & : x \geq 0 \text{ and } x < \frac{2\pi}{d}. \end{cases} \quad (56)$$

It then follows for every probability density  $p(\phi)$  that

$$\lim_{d \rightarrow \infty} \int_{-\pi}^{\pi} d\hat{\phi} \left| p(\hat{\phi}) - \int_{-\pi}^{\pi} d\phi p(\phi) p_d(\hat{\phi}|\phi) \right| = 0, \quad (57)$$

which is essentially equivalent to the less formal statement that  $p_d(\hat{\phi}|\phi)$  converges to the Dirac delta function  $\delta(\hat{\phi} - \phi)$  in the  $d \rightarrow \infty$  limit. The above convergence statement

in Eq. (57) is proved in a rigorous way in Lemma 1 in Appendix B.

Now applying this reasoning to the BDC  $\mathcal{D}_p$ , we find from a direct application of Eq. (54) that

$$\text{Tr}[|u_k\rangle\langle u_k|\mathcal{D}_p(|+_d\rangle\langle+_d|)] = \int_{-\pi}^{\pi} d\phi p(\phi)p_d(k|\phi). \quad (58)$$

Let us then denote by  $(M_\phi)_\phi$  the measurement that (1) performs a Fourier-basis measurement  $\{|u_k\rangle\langle u_k|\}_k$  with outcome  $k$ , (2) calculates the value  $\phi = 2\pi k/d$  and shifts by  $-2\pi$  if  $\phi \in (\pi, 2\pi)$ , and (3) finally adds to this value uniform noise selected from an interval of size  $2\pi/d$  to produce an outcome  $\phi$ . It then follows as a consequence of Eqs. (57) and (58) that

$$p(\phi) = \lim_{d \rightarrow \infty} \text{Tr}[M_\phi \mathcal{D}_p(|+_d\rangle\langle+_d|)], \quad (59)$$

concluding our proof of Eq. (39) for this scheme.

### B. Coherent-state method

This method involves preparing a coherent state at the input and performing heterodyne detection and classical postprocessing at the output, which is a routine method for optical phase estimation (see, e.g., Refs. [57,67–70]). The coherent-state method is easier to implement in practice than the photon-number-superposition method. In this approach, the initial state is given by the following coherent state:

$$|\alpha\rangle := e^{-\frac{1}{2}|\alpha|^2} \sum_{n=0}^{\infty} \frac{\alpha^n}{\sqrt{n!}} |n\rangle, \quad (60)$$

where  $\alpha \in \mathbb{C}$ . For the scheme we use here, we fix  $\alpha \in \mathbb{R}_+$ . After the phase rotation  $e^{-i\hat{n}\phi}$  acts, the state becomes

$$e^{-i\hat{n}\phi} |\alpha\rangle = |\alpha e^{-i\phi}\rangle, \quad (61)$$

as reviewed in Appendix C. Performing a heterodyne measurement (with POVM elements  $\{\frac{1}{\pi}|\beta\rangle\langle\beta|\}_{\beta \in \mathbb{C}}$ ) on the state in Eq. (61) then leads to the measurement outcome  $\beta$ . The final step (classical postprocessing) is to compute the argument of  $\beta$ , i.e.,  $\hat{\phi} := \arg(\beta)$ , as an estimate of the phase  $\phi$ ; the probability density for  $\hat{\phi}$  is known as the Rician phase distribution and is given by [71, Eqs. (10) & (20)]

$$p_\alpha(\hat{\phi}|\phi) := \frac{e^{-|\alpha|^2}}{2\pi} + \frac{1}{2} \frac{|\alpha|}{\sqrt{\pi}} \cos(\hat{\phi} - \phi) e^{-|\alpha|^2 \sin^2(\hat{\phi} - \phi)} \times \left[ 1 + \text{erf}(|\alpha| \cos(\hat{\phi} - \phi)) \right]. \quad (62)$$

For a derivation of the Rician phase probability density, provided for convenience, see Appendix D. Notably,

this probability density is highly peaked at  $\hat{\phi} = \phi$  and converges to a Dirac delta function in the following sense:

$$\lim_{\alpha \rightarrow \infty} \int_{-\pi}^{\pi} d\hat{\phi} \left| p(\hat{\phi}) - \int_{-\pi}^{\pi} d\phi p_\alpha(\hat{\phi}|\phi) p(\phi) \right|, \quad (63)$$

where  $p(\phi)$  is an arbitrary probability density defined on the interval  $[-\pi, \pi]$ . We provide a rigorous statement of the above convergence in Lemma 2 in Appendix C.

Finally, denoting by  $(M_\phi)_\phi$  the measurement that (1) performs heterodyne detection with outcome  $\beta$  and (2) outputs the value  $\phi = \arg(\beta)$ , it follows as a consequence of Eq. (63) that

$$p(\phi) = \lim_{\alpha \rightarrow \infty} \text{Tr}[M_\phi \mathcal{D}_p(|\alpha\rangle\langle\alpha|)], \quad (64)$$

concluding our proof of Eq. (39) for this scheme. Let us remark that an explicit form for the POVM  $(M_\phi)_\phi$  has been obtained in Ref. [57, Eq. (3.10)] and is as follows:

$$M_\phi = \frac{1}{2\pi} \sum_{m,n=0}^{\infty} |m\rangle\langle n| e^{i\phi(m-n)} \frac{\Gamma(\frac{n+m}{2} + 1)}{\sqrt{n!m!}}. \quad (65)$$

### C. Comparison of methods for finite energy

Let us compare the performance of the photon-number-superposition and coherent-state methods to the fundamental limit when there is an energy constraint in place. In particular, let us consider the channel discrimination (asymmetric error) of two bosonic dephasing channels  $\mathcal{D}_{p_{\gamma_1}}$  and  $\mathcal{D}_{p_{\gamma_2}}$ , for which the underlying probability densities  $p_{\gamma_1}$  and  $p_{\gamma_2}$  are wrapped normal distributions with respective variances  $\gamma_1$  and  $\gamma_2$ . Figure 7 illustrates how quickly the relative entropy of these schemes converges to the optimal relative entropy. For the photon-number-superposition scheme, the probability density as a function of  $d$  is given by

$$p_{d,\gamma}(\hat{\phi}) := \int_{-\pi}^{\pi} d\phi p_\gamma(\phi) p_d(\hat{\phi}|\phi), \quad (66)$$

where  $p_d(\hat{\phi}|\phi)$  is defined in Eq. (55), from which we can calculate the relative entropy  $D(p_{d,\gamma_1} \| p_{d,\gamma_2})$  of this scheme as a function of  $d$ . For the coherent-state scheme, the probability density as a function of  $\alpha$  is given by

$$p_{\alpha,\gamma}(\hat{\phi}) := \int_{-\pi}^{\pi} d\phi p_\alpha(\hat{\phi}|\phi) p_\gamma(\phi), \quad (67)$$

from which we can calculate the relative entropy  $D(p_{\alpha,\gamma_1} \| p_{\alpha,\gamma_2})$  for this scheme as a function of  $\alpha$ . Interestingly, in Fig. 7 it is indicated that these schemes in practice come close to achieving the fundamental limit, and we also see that the coherent-state scheme has an advantage over the photon-number scheme for the same fixed energy, given that the mean photon number is 9.5 for the state  $|+_d\rangle$  in Eq. (51) when  $d = 20$ .

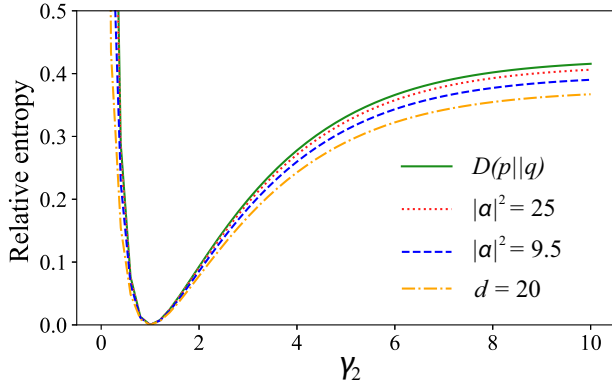


FIG. 7. A comparison of the photon-number-superposition and coherent-state schemes to the fundamental limit, when considering channel discrimination in the setting of asymmetric error. In the figure, we plot the various relative entropies when  $\gamma_1 = 1$ ; here,  $d = 20$  for the photon-number-superposition scheme and  $|\alpha|^2 \in \{9.5, 25\}$  for the coherent-state scheme. Note that the average energy of the input state is the same for  $|\alpha|^2 = 9.5$  and  $d = 20$ .

## VII. BOSONIC DEPHASING AND LOSS

Our results apply more generally to a scenario that involves photon loss in addition to dephasing. This indicates a certain robustness of our results, since we expect to encounter photon loss in any realistic scenario. Namely, suppose that the two channels to distinguish are  $\mathcal{L}_\eta \circ \mathcal{D}_p$  and  $\mathcal{L}_\eta \circ \mathcal{D}_q$ , where  $\mathcal{L}_\eta$  is a pure-loss bosonic channel of transmissivity  $\eta \in (0, 1]$ . This composite channel has been studied in the context of quantum communication, under the name bosonic loss-dephasing channel [72,73]. Our main observation here is that the distinguishability of these channels in all scenarios considered is no different from the distinguishability of  $\mathcal{D}_p$  and  $\mathcal{D}_q$ . Thus, all results stated above for  $\mathcal{D}_p$  and  $\mathcal{D}_q$  also hold for  $\mathcal{L}_\eta \circ \mathcal{D}_p$  and  $\mathcal{L}_\eta \circ \mathcal{D}_q$ .

The optimality part of this claim follows by similar reasoning used in Sec. V. That is, since the pure-loss channel  $\mathcal{L}_\eta$  is common to both  $\mathcal{L}_\eta \circ \mathcal{D}_p$  and  $\mathcal{L}_\eta \circ \mathcal{D}_q$ , it can be considered as part of an adaptive strategy used to discriminate these channels and so their distinguishability is still limited by the classical environmental states  $\sigma_p$  and  $\sigma_q$ . The attainability part follows by using the scheme from Sec. VI and the fact that coherent states retain their purity after the action of a pure-loss channel. That is,  $\mathcal{L}_\eta(|\alpha\rangle\langle\alpha|) = |\sqrt{\eta}\alpha\rangle\langle\sqrt{\eta}\alpha|$ . Then, the following limit holds by applying the same reasoning used to justify Eq. (64):

$$\begin{aligned} p(\phi) &= \lim_{\alpha \rightarrow \infty} \text{Tr}[M_\phi(\mathcal{L}_\eta \circ \mathcal{D}_p)(|\alpha\rangle\langle\alpha|)], \\ &= \lim_{\alpha \rightarrow \infty} \text{Tr}[M_\phi \mathcal{D}_p(|\sqrt{\eta}\alpha\rangle\langle\sqrt{\eta}\alpha|)]. \end{aligned} \quad (68)$$

Here, we have also used the fact that dephasing channels and pure-loss channels commute; i.e.,  $\mathcal{L}_\eta \circ \mathcal{D}_p = \mathcal{D}_p \circ \mathcal{L}_\eta$ .

Similarly, all estimation results stated above for the parametrized family  $(\mathcal{D}_{p_\theta})_{\theta \in \Theta}$  also hold for the parametrized family  $(\mathcal{L}_\eta \circ \mathcal{D}_{p_\theta})_{\theta \in \Theta}$ . This follows from the same reasoning given for the discrimination setting. Namely, the optimality part follows because all estimation strategies for the family  $(\mathcal{L}_\eta \circ \mathcal{D}_{p_\theta})_{\theta \in \Theta}$  are limited by those of the family  $(p_\theta)_{\theta \in \Theta}$  of probability densities. Then, for the attainability part, the equality in Eq. (68) applies, allowing us to apply the reasoning in Sec. VI again.

Finally, numerical estimates using the probability distribution derived in Appendix H indicate that the photon-number-superposition method from Sec. VIA might also be optimal in the presence of loss, provided that one considers the limit of infinite energy. That is, although the uniform superposition state in Eq. (51) is affected detrimentally by loss, it seems to retain sufficiently high coherence to effectively detect a phase-space rotation. The coherent-state scheme from Sec. VIB, however, might still also have an advantage over the photon-number-superposition method in the presence of loss if one considers the finite-energy setting; it is illustrated in Fig. 7 that this is indeed the case for channel discrimination in the setting of asymmetric error.

## VIII. CONCLUSIONS

In conclusion, we have determined the fundamental limits of discrimination and estimation for BDCs, complementing the recent results of Ref. [1] on communication. Not only have we accomplished this for asymptotic quantities such as the symmetric and asymmetric error exponents for channel discrimination but we have also done so for the underlying fundamental operational quantities such as the symmetric and asymmetric error probabilities of an arbitrary  $n$ -round adaptive strategy [see Eqs. (8) and (9), respectively]. We have done the same for the main operational quantity in channel estimation, the risk of an  $n$ -round adaptive strategy [see Eq. (10)]. The main ideas for these results relied on the method of simulation from Ref. [55], for the optimality part, and to exhibit a sequence of strategies that pre- and postprocess a BDC to recover its underlying probability density, for the attainability part. This is similar in spirit to previous results of Refs. [26,27].

Going forward from here, the main pressing open question is to determine the limits for these tasks whenever there is a realistic energy constraint in place. More specifically, we think it is interesting to determine which scheme, either the photon-number-superposition scheme from Sec. VIA or the coherent-state scheme from Sec. VIB, performs better in the finite-energy regime, as well as in the case that there is photon loss in addition to dephasing. There are certainly other schemes besides these two to consider as well. Furthermore, given that our findings in Sec. VII only apply when the transmissivity parameter  $\eta$  is fixed, it is open to determine the limits

of discrimination and estimation when the transmissivity parameter varies in addition to the dephasing channel.

Another natural generalization of our results is to the case of an arbitrary random unitary channel of the form

$$\mathcal{N}_{p,H}(\rho) := \int_{-\infty}^{+\infty} dt p(t) e^{-iHt} \rho e^{iHt}, \quad (69)$$

where  $p$  is a probability density on the real line and  $H$  is a general Hamiltonian. The same simulation arguments from Sec. V allow for concluding optimality bounds, that all adaptive strategies for discriminating or estimating channels from this class are limited by the underlying classical probability densities. Based on the insights from Ref. [74, Proposition 2], we expect that seizing the underlying probability density  $p$  might be possible for a large class of Hamiltonians. If that is the case, then our results could be extended far beyond the setting we considered here.

All codes used to generate the figures in this paper are available with the arXiv posting of this paper as arXiv ancillary files.

## ACKNOWLEDGMENTS

We thank Ben Q. Baragiola, Francisco Elohim Becerra, Maison Clouâtre, Bunyamin Kartal, Ufuk Keskin, Marco A. Rodríguez-García, and Moe Win for helpful comments. Z.H. is supported by a Sydney Quantum Academy Postdoctoral Fellowship and an ARC DECRA Fellowship (DE230100144) ‘‘Quantum-enabled super-resolution imaging.’’ M.M.W. acknowledges support from the National Science Foundation under Grant No. 2304816 and is grateful to CWI Amsterdam for hospitality during a research visit when this work was finalized.

## APPENDIX A: EQUIVALENCE OF HYPOTHESIS TESTING REGIONS

In this appendix, we prove that the following equality holds for every pair of bosonic dephasing channels  $\mathcal{D}_p$  and  $\mathcal{D}_q$ :

$$\{(\alpha_n(\mathcal{A}), \beta_n(\mathcal{A}))\}_{\mathcal{A}} = \{(\alpha_n(t), \beta_n(t))\}_t, \quad (A1)$$

where  $\alpha_n(\mathcal{A})$  and  $\beta_n(\mathcal{A})$  are defined in Eq. (2) and  $\alpha_n(t)$  and  $\beta_n(t)$  are defined in Eqs. (3) and (4). These sets are known as hypothesis-testing regions and have been studied for a long time in statistics [75] (see also Ref. [76,77] for more recent works in quantum information). Furthermore, the quantities in Eqs. (8) and (9) can be understood as various boundary points of this hypothesis-testing region. As such, the equalities in Eqs. (8) and (9) follow as a consequence of Eq. (A1).

The containment

$$\{(\alpha_n(\mathcal{A}), \beta_n(\mathcal{A}))\}_{\mathcal{A}} \subseteq \{(\alpha_n(t), \beta_n(t))\}_t \quad (A2)$$

follows as a consequence of the same reasoning used to establish Eqs. (36) and (37) and can again be understood by examining Fig. 5. Indeed, every adaptive strategy  $\mathcal{A}$  for distinguishing the BDCs  $\mathcal{D}_p$  and  $\mathcal{D}_q$  can be understood as a particular classical test  $t$  for distinguishing the underlying densities  $p$  and  $q$ . As such, the region of achievable pairs using quantum adaptive strategies is contained in the region of achievable error pairs for the underlying densities.

The other containment,

$$\{(\alpha_n(t), \beta_n(t))\}_t \subseteq \{(\alpha_n(\mathcal{A}), \beta_n(\mathcal{A}))\}_{\mathcal{A}}, \quad (A3)$$

follows by employing either one of the two strategies from Secs. VIA or VIB. Indeed, in the large-energy limit, it is possible to employ either one of these two strategies and obtain  $n$  samples of the underlying densities  $p^{\otimes n}$  or  $q^{\otimes n}$ . Once the samples are in hand, one can then perform an arbitrary classical test  $t$  on them.

## APPENDIX B: CALCULATIONS FOR PHOTON-NUMBER-SUPERPOSITION METHOD

Recall the definitions of  $|+_d\rangle$  and  $|u_k\rangle$  in Eqs. (51) and (53), respectively. Measuring the state  $e^{-i\hat{n}\phi}|+_d\rangle$  in the Fourier basis leads to the following outcome probabilities:

$$\begin{aligned} & \left| \langle u_k | e^{-i\hat{n}\phi} |+_d\rangle \right|^2 \\ &= \left| \left( \frac{1}{\sqrt{d}} \sum_{n'=0}^{d-1} e^{2\pi i k n' / d} \langle n' | \right) \left( \frac{1}{\sqrt{d}} \sum_{n=0}^{d-1} e^{-i n \phi} |n\rangle \right) \right|^2 \end{aligned} \quad (B1)$$

$$= \frac{1}{d^2} \left| \sum_{n',n=0}^{d-1} e^{2\pi i k n' / d} e^{-i n \phi} \langle n' | n \rangle \right|^2 \quad (B2)$$

$$= \frac{1}{d^2} \left| \sum_{n=0}^{d-1} \exp \left( i n \left( \frac{2\pi k}{d} - \phi \right) \right) \right|^2 \quad (B3)$$

$$= \frac{1}{d^2} \left| \frac{1 - \exp(i(2\pi k - d\phi))}{1 - \exp(i(\frac{2\pi k}{d} - \phi))} \right|^2 \quad (B4)$$

$$= \frac{1}{d^2} \frac{2(1 - \cos(2\pi k - d\phi))}{2(1 - \cos(\frac{2\pi k}{d} - \phi))} \quad (B5)$$

$$= \frac{1}{d^2} \frac{\sin^2 \left( \pi k - \frac{d\phi}{2} \right)}{\sin^2 \left( \frac{\pi k}{d} - \frac{\phi}{2} \right)}, \quad (B6)$$

thus justifying Eq. (54).

The following lemma rigorously justifies the convergence statement asserted in Eq. (57).

*Lemma 1.* Let  $p : [-\pi, \pi] \rightarrow \mathbb{R}_+$  be a continuous non-negative function with  $p(-\pi) = p(\pi)$  and  $\int_{-\pi}^{\pi} d\phi p(\phi) = 1$ . For every positive integer  $d$ , all  $k \in \{0, 1, \dots, d-1\}$ , and all  $\hat{\phi} \in [-\pi, \pi]$ , set

$$q_d(\hat{\phi}|k) := \Pi_d \left( \hat{\phi} - \frac{2\pi k}{d} \bmod 2\pi \right), \quad (\text{B7})$$

where  $\Pi_d$  is defined by Eq. (56), and

$$x \bmod 2\pi := \min \{x + 2\pi k : x + 2\pi k \geq 0, k \in \mathbb{Z}\}. \quad (\text{B8})$$

Then the function  $p'_d : [-\pi, \pi] \rightarrow \mathbb{R}_+$  defined by

$$p'_d(\hat{\phi}) := \sum_{k=0}^{d-1} q_d(\hat{\phi}|k) \langle u_k | \mathcal{D}_p(|+\rangle\langle +|) | u_k \rangle, \quad (\text{B9})$$

where  $\mathcal{D}_p$  is the bosonic dephasing channel given by Eq. (1), satisfies

$$p'_d \xrightarrow{d \rightarrow \infty} p \quad (\text{B10})$$

uniformly on  $[-\pi, \pi]$  and, furthermore,

$$\lim_{d \rightarrow \infty} \int_{-\pi}^{\pi} d\phi |p(\phi) - p'_d(\phi)| = 0. \quad (\text{B11})$$

*Proof.* We start by observing that, due to the calculation in the first part of this appendix,

$$p'_d(\hat{\phi}) \stackrel{(i)}{=} \frac{1}{d} \sum_{k=0}^{d-1} \Pi_d \left( \hat{\phi} - \frac{2\pi k}{d} \bmod 2\pi \right) \times \int_{-\pi}^{\pi} d\phi p(\phi) F_d \left( \frac{2\pi k}{d} - \phi \right) \quad (\text{B12})$$

$$\stackrel{(ii)}{=} \int_{-\pi}^{\pi} \frac{d\phi}{2\pi} p(\phi) F_d \left( \frac{2\pi}{d} \left\lfloor \frac{d\hat{\phi}'}{2\pi} \right\rfloor - \phi \right) \quad (\text{B13})$$

$$\stackrel{(iii)}{=} \int_{-\pi}^{\pi} \frac{d\phi}{2\pi} p(\phi) F_d \left( \frac{2\pi}{d} \left\lfloor \frac{d\hat{\phi}}{2\pi} \right\rfloor - \phi \right) \quad (\text{B14})$$

$$\stackrel{(iv)}{=} (p \star F_d) \left( \frac{2\pi}{d} \left\lfloor \frac{d\hat{\phi}}{2\pi} \right\rfloor \right) \quad (\text{B15})$$

Here, in (i) we have introduced the Fejér kernel  $F_d(x) := \frac{\sin^2(dx/2)}{d \sin^2(x/2)}$ , while in (ii) we have observed that the only nonzero term in the sum is for  $k = \left\lfloor d\hat{\phi}'/2\pi \right\rfloor$ , where

$\hat{\phi}' := \hat{\phi} \bmod 2\pi$ ; indeed, since changing  $k \mapsto k+1$  displaces the point  $\hat{\phi} - (2\pi k/d)$  by exactly  $-2\pi/d$  and the function  $\Pi_d$  is nonzero in an interval of length precisely equal to  $2\pi/d$ , there can be only one nonzero term in the sum; using that  $\hat{\phi}' = \hat{\phi} + 2\pi p$ , where  $p \in \{0, 1\}$ , we can also verify that

$$\hat{\phi} - \frac{2\pi}{d} \left\lfloor \frac{d\hat{\phi}'}{2\pi} \right\rfloor = \hat{\phi} - \frac{2\pi}{d} \left\lfloor \frac{d\hat{\phi}}{2\pi} \right\rfloor - 2\pi p. \quad (\text{B16})$$

Using  $x - 1 < \lfloor x \rfloor \leq x$ , we now note that

$$0 \leq \hat{\phi} - \frac{2\pi}{d} \left\lfloor \frac{d\hat{\phi}}{2\pi} \right\rfloor < \hat{\phi} - \frac{2\pi}{d} \left( \frac{d\hat{\phi}}{2\pi} - 1 \right) = \frac{2\pi}{d}, \quad (\text{B17})$$

implying that, indeed,

$$\Pi_d \left( \hat{\phi} - \frac{2\pi}{d} \left\lfloor \frac{d\hat{\phi}}{2\pi} \right\rfloor \bmod 2\pi \right) = \Pi_d \left( \hat{\phi} - \frac{2\pi}{d} \left\lfloor \frac{d\hat{\phi}}{2\pi} \right\rfloor \right) \quad (\text{B18})$$

$$= \frac{d}{2\pi}. \quad (\text{B19})$$

Continuing with the justification of the first chain of identities, in (iii) we have used the periodicity of  $F_d$  to substitute  $\hat{\phi}'$  with  $\hat{\phi}$  and, finally, in (iv) we have introduced the notation

$$(p \star F_d)(\xi) := \int_{-\pi}^{\pi} \frac{d\phi}{2\pi} p(\phi) F_d(\xi - \phi). \quad (\text{B20})$$

Now, calling  $\tilde{p}$  the periodic extension of  $p$  to the whole real line, for all  $\xi \in \mathbb{R}$  one sees that

$$(p \star F_d)(\xi) = \int_{-\pi+\xi}^{\pi+\xi} \frac{d\theta}{2\pi} p(\xi - \theta) F_d(\theta) \quad (\text{B21})$$

$$= \int_{-\pi}^{\pi} \frac{d\theta}{2\pi} \tilde{p}(\xi - \theta) F_d(\theta). \quad (\text{B22})$$

Note that since  $p$  is continuous on the compact set  $[-\pi, \pi]$ , it is also uniformly continuous. Due to the fact that  $p(-\pi) = p(\pi)$ , its extension  $\tilde{p}$  can also be shown to be uniformly continuous. Let  $\omega$  be the modulus of continuity of  $\tilde{p}$ . This means that  $\omega : [0, \infty) \rightarrow [0, \infty)$  is a nondecreasing continuous function, with  $\omega(0) = 0$ , such that for

all  $\xi, \xi' \in \mathbb{R}$ , it holds that

$$|\tilde{p}(\xi) - \tilde{p}(\xi')| \leq \omega(|\xi - \xi'|). \quad (\text{B23})$$

Now, for  $\xi, \xi' \in \mathbb{R}$ , we can write that

$$\begin{aligned} & |(p \star F_d)(\xi) - (p \star F_d)(\xi')| \\ &= \left| \int_{-\pi}^{\pi} \frac{d\theta}{2\pi} (\tilde{p}(\xi - \theta) - \tilde{p}(\xi' - \theta)) F_d(\theta) \right| \end{aligned} \quad (\text{B24})$$

$$\leq \int_{-\pi}^{\pi} \frac{d\theta}{2\pi} |\tilde{p}(\xi - \theta) - \tilde{p}(\xi' - \theta)| F_d(\theta) \quad (\text{B25})$$

$$\leq \omega(|\xi - \xi'|) \int_{-\pi}^{\pi} \frac{d\theta}{2\pi} F_d(\theta) \quad (\text{B26})$$

$$= \omega(|\xi - \xi'|), \quad (\text{B27})$$

where in the last line we have leveraged the fact that  $\int_{-\pi}^{\pi} (d\theta/2\pi) F_d(\theta) = 1$  for all  $d$ . In other words,  $p \star F_d$  is also uniformly continuous and it has the same modulus of continuity as  $\tilde{p}$ .

We are finally ready to put everything together and prove the first half of the claim. We write that

$$\begin{aligned} & |p'_d(\hat{\phi}) - p(\hat{\phi})| \\ &= \left| (p \star F_d) \left( \frac{2\pi}{d} \left\lfloor \frac{d\hat{\phi}}{2\pi} \right\rfloor \right) - p(\hat{\phi}) \right| \end{aligned} \quad (\text{B28})$$

$$\begin{aligned} & \leq \left| (p \star F_d) \left( \frac{2\pi}{d} \left\lfloor \frac{d\hat{\phi}}{2\pi} \right\rfloor \right) - (p \star F_d)(\hat{\phi}) \right| \\ & \quad + \left| (p \star F_d)(\hat{\phi}) - p(\hat{\phi}) \right| \end{aligned} \quad (\text{B29})$$

$$\leq \omega \left( \left| \frac{2\pi}{d} \left\lfloor \frac{d\hat{\phi}}{2\pi} \right\rfloor - \hat{\phi} \right| \right) + \left| (p \star F_d)(\hat{\phi}) - p(\hat{\phi}) \right| \quad (\text{B30})$$

$$\leq \omega(2\pi/d) + \left| (p \star F_d)(\hat{\phi}) - p(\hat{\phi}) \right|, \quad (\text{B31})$$

where in the last line we have noted that  $\hat{\phi} \leq 2\pi/d \left\lfloor d\hat{\phi}/2\pi \right\rfloor \leq \hat{\phi} + (2\pi/d)$ , because of the elementary properties of the floor function. Now, since  $\lim_{d \rightarrow \infty} \omega(2\pi/d) = 0$ , to establish Eq. (B10) we only need to check that  $p \star F_d$  converges uniformly to  $p$  as  $d \rightarrow \infty$ ; and this is well known to follow from the continuity of  $p$ , due to Fejér's theorem [78, Theorem 3.4].

To deduce Eq. (B11) from Eq. (B10) it suffices to note that

$$\int_{-\pi}^{\pi} d\phi |p(\phi) - p'_d(\phi)| \leq 2\pi \sup_{\phi} |p(\phi) - p'_d(\phi)|, \quad (\text{B32})$$

and the right-hand side tends to 0 as  $d \rightarrow \infty$  due to Eq. (B10). ■

### APPENDIX C: CALCULATIONS FOR COHERENT-STATE METHOD

Let us first justify the equality in Eq. (61). After the phase rotation  $e^{-i\hat{n}\phi}$  acts, the state becomes

$$e^{-i\hat{n}\phi} |\alpha\rangle = e^{-\frac{1}{2}|\alpha|^2} \sum_{n=0}^{\infty} \frac{\alpha^n}{\sqrt{n!}} e^{-i\hat{n}\phi} |n\rangle \quad (\text{C1})$$

$$= e^{-\frac{1}{2}|\alpha|^2} \sum_{n=0}^{\infty} \frac{\alpha^n}{\sqrt{n!}} e^{-in\phi} |n\rangle \quad (\text{C2})$$

$$= e^{-\frac{1}{2}|\alpha|^2} \sum_{n=0}^{\infty} \frac{(\alpha e^{-i\phi})^n}{\sqrt{n!}} |n\rangle \quad (\text{C3})$$

$$= |\alpha e^{-i\phi}\rangle. \quad (\text{C4})$$

After performing heterodyne detection, and as discussed in the main text, we compute the argument of  $\beta$  as the estimate of  $\phi$ , i.e.,  $\hat{\phi} := \arg(\beta)$ . The induced probability density function for  $\hat{\phi}$  is known as the Rician phase distribution (see Ref. [71, Eqs. (10) & (20)]). In particular, we can model the random process by which  $\hat{\phi}$  is generated as being like that in Ref. [71, Eq. (3)], given by

$$\beta = \alpha \exp(-i\phi) + n, \quad (\text{C5})$$

where  $n$  is a complex Gaussian random variable  $\mathcal{CN}(0, 1)$  (such that the variance for each of the real and imaginary parts is  $1/2$ , i.e.,  $\sigma^2 = 1/2$ , using the notation of [71, Eq. (3)]). We can restrict  $\alpha$  to be a positive real number and, in this case, we have that  $A = \alpha$  and  $B = -1$ , using the notation of [71, Eq. (3)]. Following Ref. [71, Eqs. (10) & (20)], we find that the probability density  $p_{\alpha}(\hat{\phi}|\phi)$  for  $\hat{\phi} \in [-\pi, \pi]$  is given by

$$\begin{aligned} p_{\alpha}(\hat{\phi}|\phi) &:= \frac{e^{-|\alpha|^2}}{2\pi} + \frac{1}{2} \frac{|\alpha|}{\sqrt{\pi}} \cos(\hat{\phi} - \phi) e^{-|\alpha|^2 \sin^2(\hat{\phi} - \phi)} \\ & \quad \times \left[ 1 + \operatorname{erf}(|\alpha| \cos(\hat{\phi} - \phi)) \right]. \end{aligned} \quad (\text{C6})$$

We now show that this probability density converges to a Dirac delta at  $\phi$  in the limit as  $\alpha \rightarrow \infty$ , in the sense stated in Eq. (63).

*Lemma 2.* Let  $p : [-\pi, \pi] \rightarrow \mathbb{R}_+$  be a continuous non-negative function with  $p(-\pi) = p(\pi)$  and  $\int_{-\pi}^{\pi} d\phi p(\phi) = 1$ . For all  $\alpha \geq 0$ , let  $p_{\alpha}(\hat{\phi}|\phi)$  be the Rician probability

distribution defined by Eq. (62) and set

$$p'_\alpha(\hat{\phi}) := \int_{-\pi}^{\pi} d\phi p(\phi) p_\alpha(\hat{\phi}|\phi). \quad (\text{C7})$$

Then,

$$p'_\alpha \xrightarrow{\alpha \rightarrow \infty} p \quad (\text{C8})$$

pointwise on  $[-\pi, \pi]$  and, furthermore,

$$\lim_{\alpha \rightarrow \infty} \int_{-\pi}^{\pi} d\phi |p(\phi) - p'_\alpha(\phi)| = 0. \quad (\text{C9})$$

*Proof.* For this proof, it is ideal to work with the integral representation of the Rician probability density given in Eq. (D5); namely,

$$p_\alpha(\hat{\phi}|\phi) = \int_0^\infty db \frac{b}{\pi} e^{-|\alpha - be^{-i(\hat{\phi}-\phi)}|^2}. \quad (\text{C10})$$

Substituting Eq. (C10) into Eq. (C7), we now have that

$$p'_\alpha(\hat{\phi}) \stackrel{(i)}{=} \int_{-\pi}^{\pi} d\phi \tilde{p}(\phi) \int_0^\infty db \frac{b}{\pi} e^{-|\alpha - be^{-i(\hat{\phi}-\phi)}|^2} \quad (\text{C11})$$

$$\stackrel{(ii)}{=} \int_{\mathbb{C}} \frac{d^2\gamma}{\pi} \tilde{p}(\hat{\phi} + \arg(\gamma)) e^{-|\alpha - \gamma|^2} \quad (\text{C12})$$

$$\stackrel{(iii)}{=} \int_{\mathbb{C}} \frac{d^2z}{\pi} \tilde{p}(\hat{\phi} + \arg(z + \alpha)) e^{-|z|^2}. \quad (\text{C13})$$

The justification of the above chain of identities is as follows: in (i), we have used the above integral representation of the Rician probability density and introduced the periodic extension  $\tilde{p}$  of  $p$  to the whole real line; in (ii), we have used Fubini's theorem, changing variables to  $\gamma := be^{-i(\hat{\phi}-\phi)}$ ; and in (iii), we have changed variables again, setting  $z := \gamma - \alpha$ .

Now, since  $\alpha \geq 0$  is real, we have that

$$\arg(z + \alpha) = \arctan\left(\frac{z_I}{\alpha + z_R}\right) \quad (\text{C14})$$

up to multiples of  $2\pi$  and hence

$$\lim_{\alpha \rightarrow \infty} \tilde{p}(\hat{\phi} + \arg(z + \alpha)) = \tilde{p}(\hat{\phi}) \quad (\text{C15})$$

due to the continuity of  $\tilde{p}$  and the arctan function. We can then write

$$\lim_{\alpha \rightarrow \infty} \int_{\mathbb{C}} \frac{d^2z}{\pi} \tilde{p}(\hat{\phi} + \arg(z + \alpha)) e^{-|z|^2} \stackrel{(iv)}{=} \int_{\mathbb{C}} \frac{d^2z}{\pi} \lim_{\alpha \rightarrow \infty} \tilde{p}(\hat{\phi} + \arg(z + \alpha)) e^{-|z|^2} \quad (\text{C16})$$

$$= \int_{\mathbb{C}} \frac{d^2z}{\pi} \tilde{p}(\hat{\phi}) e^{-|z|^2} \quad (\text{C17})$$

$$= \tilde{p}(\hat{\phi}) = p(\hat{\phi}), \quad (\text{C18})$$

where in (iv) we have employed Lebesgue's dominated-convergence theorem, which is applicable because, due to its periodicity and continuity,  $\tilde{p}$  is a bounded function, which means that  $\tilde{p}(\hat{\phi} + \arg(z + \alpha)) e^{-|z|^2} \leq M e^{-|z|^2}$  for some constant  $M > 0$  and the right-hand side is an absolutely integrable function of  $z$ . This completes the proof of Eq. (C8).

To deduce Eq. (C9), we first note that if  $\tilde{p}(\xi) \leq M$  for all  $\xi \in \mathbb{R}$ , then also

$$p'_\alpha(\phi) = \int_{\mathbb{C}} \frac{d^2z}{\pi} \tilde{p}(\phi + \arg(z + \alpha)) e^{-|z|^2} \quad (\text{C19})$$

$$\leq M \int_{\mathbb{C}} \frac{d^2z}{\pi} e^{-|z|^2} = M, \quad (\text{C20})$$

implying that  $|p'_\alpha(\phi) - p(\phi)| \leq M$  for all  $\phi \in [-\pi, \pi]$  and again by Lebesgue's dominated convergence (since we integrate on a finite-measure space),

$$\lim_{\alpha \rightarrow \infty} \int_{-\pi}^{\pi} d\phi |p(\phi) - p'_\alpha(\phi)| = \int_{-\pi}^{\pi} d\phi \lim_{\alpha \rightarrow \infty} |p(\phi) - p'_\alpha(\phi)| \quad (\text{C21})$$

$$= 0. \quad (\text{C22})$$

This establishes Eq. (C9) and thereby concludes the proof.  $\blacksquare$

#### APPENDIX D: DERIVATION OF RICIAN PHASE PROBABILITY DENSITY FUNCTION

Here, for completeness, we provide a derivation of the Rician phase probability density function. Consider that the probability density for obtaining the outcome  $\beta \in \mathbb{C}$  when performing heterodyne detection on a coherent state  $|\alpha\rangle$ , where  $\alpha \in \mathbb{C}$ , is as follows [79, Eqs. (4.7) & (5.122)]:

$$p(\beta|\alpha) = \frac{1}{\pi} e^{-|\alpha - \beta|^2}. \quad (\text{D1})$$

Letting  $\alpha = re^{-i\hat{\phi}}$  and  $\beta = be^{-i\hat{\phi}}$ , with  $r, b \geq 0$  and  $\phi, \hat{\phi} \in [-\pi, \pi]$ , we find that

$$p(\beta|\alpha) d^2\beta = \frac{1}{\pi} e^{-|\alpha - \beta|^2} d^2\beta \quad (\text{D2})$$

$$= \frac{1}{\pi} \exp\left(-|re^{-i\hat{\phi}} - be^{-i\hat{\phi}}|^2\right) b db d\hat{\phi}. \quad (\text{D3})$$

Then, we obtain the marginal probability density for the phase  $\hat{\phi}$  by integrating over the magnitude  $b$ :

$$p(\hat{\phi}|r, \phi) = \int_0^\infty db \frac{b}{\pi} \exp\left(-\left|re^{-i\phi} - be^{-i\hat{\phi}}\right|^2\right) \quad (\text{D4})$$

$$= \int_0^\infty db \frac{b}{\pi} \exp\left(-\left|r - be^{-i(\hat{\phi}-\phi)}\right|^2\right). \quad (\text{D5})$$

Considering that

$$\begin{aligned} \left|r - be^{-i(\hat{\phi}-\phi)}\right|^2 &= r^2 - 2rb \cos(\hat{\phi} - \phi) + b^2 \quad (\text{D6}) \\ &= \left(r \sin(\hat{\phi} - \phi)\right)^2 + \left(b - r \cos(\hat{\phi} - \phi)\right)^2, \quad (\text{D7}) \end{aligned}$$

we find that

$$\begin{aligned} p(\hat{\phi}|r, \phi) &= \frac{e^{-(r \sin(\hat{\phi}-\phi))^2}}{\pi} \int_0^\infty db b e^{-(b-r \cos(\hat{\phi}-\phi))^2} \quad (\text{D8}) \\ &= \frac{e^{-(r \sin(\hat{\phi}-\phi))^2}}{\pi} \\ &\quad \times \int_{-r \cos(\hat{\phi}-\phi)}^\infty d\bar{b} \left(\bar{b} + r \cos(\hat{\phi} - \phi)\right) e^{-\bar{b}^2}. \quad (\text{D9}) \end{aligned}$$

Now consider that

$$\begin{aligned} &\int_{-r \cos(\hat{\phi}-\phi)}^\infty d\bar{b} \left(\bar{b} + r \cos(\hat{\phi} - \phi)\right) e^{-\bar{b}^2} \\ &= \int_{-r \cos(\hat{\phi}-\phi)}^\infty d\bar{b} \bar{b} e^{-\bar{b}^2} \\ &\quad + r \cos(\hat{\phi} - \phi) \int_{-r \cos(\hat{\phi}-\phi)}^\infty d\bar{b} e^{-\bar{b}^2} \quad (\text{D10}) \end{aligned}$$

$$\begin{aligned} &= -\frac{1}{2} \int_{-r \cos(\hat{\phi}-\phi)}^\infty \frac{d}{d\bar{b}} e^{-\bar{b}^2} \\ &\quad + r \cos(\hat{\phi} - \phi) \frac{\sqrt{\pi}}{2} \left(1 + \operatorname{erf}(r \cos(\hat{\phi} - \phi))\right) \quad (\text{D11}) \end{aligned}$$

$$\begin{aligned} &= \frac{1}{2} e^{-(r \cos(\hat{\phi}-\phi))^2} \\ &\quad + r \cos(\hat{\phi} - \phi) \frac{\sqrt{\pi}}{2} \left(1 + \operatorname{erf}(r \cos(\hat{\phi} - \phi))\right). \quad (\text{D12}) \end{aligned}$$

In the above, we have made use of the error function

$$\operatorname{erf}(x) := \frac{2}{\sqrt{\pi}} \int_0^x dt e^{-t^2}, \quad (\text{D13})$$

and some of its properties:  $\operatorname{erf}(+\infty) = 1$  and  $\operatorname{erf}(x) = -\operatorname{erf}(-x)$ . Thus, we finally conclude that

$$\begin{aligned} p(\hat{\phi}|r, \phi) &= \frac{e^{-(r \sin(\hat{\phi}-\phi))^2}}{2\pi} \left( e^{-(r \cos(\hat{\phi}-\phi))^2} \right. \\ &\quad \left. + r \cos(\hat{\phi} - \phi) \sqrt{\pi} \left(1 + \operatorname{erf}(r \cos(\hat{\phi} - \phi))\right) \right) \quad (\text{D14}) \end{aligned}$$

$$\begin{aligned} &= \frac{e^{-r^2}}{2\pi} + \frac{e^{-(r \sin(\hat{\phi}-\phi))^2}}{2\sqrt{\pi}} r \cos(\hat{\phi} - \phi) \\ &\quad \times \left(1 + \operatorname{erf}(r \cos(\hat{\phi} - \phi))\right). \quad (\text{D15}) \end{aligned}$$

## APPENDIX E: OTHER SCENARIOS—STRONG CONVERSE EXPONENT, ERROR EXPONENT, MULTIPLE CHANNEL DISCRIMINATION, AND ANTIDISTINGUISHABILITY

In this appendix, we discuss various other scenarios to which our results apply. The first two are known as the strong converse exponent and error exponent, which also go by the names Han-Kobayashi [80] and Hoeffding [81], respectively. These are settings related to binary hypothesis testing. The other two scenarios are multiple channel discrimination and antidistinguishability.

### 1. Strong converse exponent

The nonasymptotic strong converse exponent for channel discrimination is defined for  $r > 0$  as follows:

$$\begin{aligned} H_n(r, \mathcal{N}_0, \mathcal{N}_1) &:= \inf_{\mathcal{A}} \left\{ -\frac{1}{n} \ln(1 - \alpha_n(\mathcal{A})) : \beta_n(\mathcal{A}) \leq e^{-rn} \right\}, \quad (\text{E1}) \end{aligned}$$

where  $\alpha_n(\mathcal{A})$  and  $\beta_n(\mathcal{A})$  are defined in Eq. (2). By applying the same reasoning as given in Secs. V and VI, we conclude for BDCs  $\mathcal{D}_p$  and  $\mathcal{D}_q$  that

$$H_n(r, \mathcal{D}_p, \mathcal{D}_q) = \inf_t \left\{ -\frac{1}{n} \ln(1 - \alpha_n(t)) : \beta_n(t) \leq e^{-rn} \right\}, \quad (\text{E2})$$

where  $\alpha_n(t)$  and  $\beta_n(t)$  are defined in Eqs. (3)–(4) and taken with respect to the probability densities  $p$  and  $q$  defining  $\mathcal{D}_p$  and  $\mathcal{D}_q$ , respectively. By taking the  $n \rightarrow \infty$  limit and applying the classical result of Ref. [80], we conclude that

$$\lim_{n \rightarrow \infty} H_n(r, \mathcal{D}_p, \mathcal{D}_q) = \sup_{\alpha > 1} \frac{\alpha - 1}{\alpha} (r - D_\alpha(p \| q)), \quad (\text{E3})$$



where the Rényi relative entropy  $D_\alpha(p\|q)$  is defined for  $\alpha \in (0, 1) \cup (1, \infty)$  as

$$D_\alpha(p\|q) := \frac{1}{\alpha - 1} \ln \int_{-\pi}^{\pi} d\phi p^\alpha(\phi) q^{1-\alpha}(\phi). \quad (\text{E4})$$

## 2. Error exponent

The nonasymptotic error exponent for channel discrimination is defined for  $r > 0$  as follows:

$$B_n(r, \mathcal{N}_0, \mathcal{N}_1) := \sup_{\mathcal{A}} \left\{ -\frac{1}{n} \ln \alpha_n(\mathcal{A}) : \beta_n(\mathcal{A}) \leq e^{-rn} \right\}, \quad (\text{E5})$$

where  $\alpha_n(\mathcal{A})$  and  $\beta_n(\mathcal{A})$  are defined in Eq. (2). By applying the same reasoning as given in Secs. V and VI, we conclude for BDCs  $\mathcal{D}_p$  and  $\mathcal{D}_q$  that

$$B_n(r, \mathcal{D}_p, \mathcal{D}_q) = \sup_t \left\{ -\frac{1}{n} \ln \alpha_n(t) : \beta_n(t) \leq e^{-rn} \right\}, \quad (\text{E6})$$

where  $\alpha_n(t)$  and  $\beta_n(t)$  are defined in Eqs. (3)–(4) and taken with respect to the probability densities  $p$  and  $q$  defining  $\mathcal{D}_p$  and  $\mathcal{D}_q$ , respectively. By taking the  $n \rightarrow \infty$  limit and applying the classical result of Ref. [81], we conclude that

$$\lim_{n \rightarrow \infty} B_n(r, \mathcal{D}_p, \mathcal{D}_q) = \sup_{\alpha \in (0,1)} \frac{\alpha - 1}{\alpha} (r - D_\alpha(p\|q)), \quad (\text{E7})$$

where the Rényi relative entropy  $D_\alpha(p\|q)$  is defined in Eq. (E4).

## 3. Multiple channel discrimination

The goal of multiple channel discrimination is to decide which channel has been chosen from a tuple of channels. More formally, let  $(\mathcal{N}_i)_{i=1}^\ell$  be a tuple of channels. Then, an adaptive protocol for channel discrimination consists of an adaptive strategy of the form discussed previously in Sec. II A, with the only difference being that the final measurement is  $\mathcal{Q} := (Q_i)_{i=1}^\ell$ . Letting  $\rho_i^{(n)}$  be the final state of such a protocol when the  $i$ th channel has been selected, the success probability of multiple channel discrimination is

$$p_n^s((\mathcal{N}_i)_{i=1}^\ell) := \sup_{\mathcal{A}} \sum_{i=1}^\ell \lambda_i \text{Tr} \left[ Q_i \rho_i^{(n)} \right], \quad (\text{E8})$$

where  $\lambda_i$  is the prior probability that channel  $\mathcal{N}_i$  is selected. (Thus, the following constraints apply:  $\lambda_i \geq 0$  for all  $i \in \{1, \dots, \ell\}$  and  $\sum_{i=1}^\ell \lambda_i = 1$ .)

Now let us consider classical multiple hypothesis testing. Let  $(p_i)_{i=1}^\ell$  be a tuple of probability densities. Here, the goal is to observe a sample  $\phi^n \equiv (\phi_1, \dots, \phi_n)$  from one

of the product densities (i.e., of the form  $p_i^{\otimes n}$ ) and decide the value of  $i$  (i.e., which density generated the sample sequence). The success probability is given by

$$p_n^s((p_i)_{i=1}^\ell) := \sup_t \sum_{i=1}^\ell \lambda_i \int d\phi^n t(i|\phi^n) p_i^{\otimes n}(\phi^n), \quad (\text{E9})$$

where  $\lambda_i$  is a prior probability and  $(t(i|\phi^n))_{i=1}^\ell$  is a conditional probability distribution (i.e., satisfying  $t(i|\phi^n) \geq 0$  for all  $i \in \{1, \dots, \ell\}$  and  $\sum_{i=1}^\ell t(i|\phi^n) = 1$ ).

By the same reasoning from Secs. V and VI, our main result here is that

$$p_n^s((\mathcal{D}_{p_i})_{i=1}^\ell) = p_n^s((p_i)_{i=1}^\ell), \quad (\text{E10})$$

where  $(\mathcal{D}_{p_i})_{i=1}^\ell$  is a tuple of bosonic dephasing channels defined by the corresponding tuple  $(p_i)_{i=1}^\ell$  of probability densities. By employing the known result [82] (see also Refs. [83, Theorem 4.2] and [84–86]) that the asymptotic error exponent for multiple hypothesis testing is equal to the minimum pairwise Chernoff divergence, we conclude the following:

$$\lim_{n \rightarrow \infty} -\frac{1}{n} \ln \left( 1 - p_n^s((\mathcal{D}_{p_i})_{i=1}^\ell) \right) = \min_{i \neq j} C(p_i\|p_j), \quad (\text{E11})$$

where the Chernoff divergence  $C(p_i\|p_j)$  is defined from Eq. (14).

## 4. Antidistinguishability

The problem of antidistinguishability has the same structure as multiple channel discrimination but the goal is the opposite. That is, the goal is to decide which channel was not selected. That is, if the  $i$ th channel is selected, the goal is to report back “not  $i$ .” We can thus adopt all of the notation from the previous section but the error probability for the antidistinguishability problem is given by

$$p_n^e((\mathcal{N}_i)_{i=1}^\ell) := \inf_{\mathcal{A}} \sum_{i=1}^\ell \lambda_i \text{Tr} \left[ Q_i \rho_i^{(n)} \right]. \quad (\text{E12})$$

Similarly, for the classical antidistinguishability problem, the error probability is given by

$$p_n^e((p_i)_{i=1}^\ell) := \inf_t \sum_{i=1}^\ell \lambda_i \int d\phi^n t(i|\phi^n) p_i^{\otimes n}(\phi^n). \quad (\text{E13})$$

Thus, the main difference with multiple hypothesis testing mathematically is to minimize the objective functions

rather than maximize them. By the same reasoning from Secs. V and VI, we conclude that

$$p_n^e((\mathcal{D}_{p_i})_{i=1}^\ell) = p_n^e((p_i)_{i=1}^\ell), \quad (\text{E14})$$

where  $(\mathcal{D}_{p_i})_{i=1}^\ell$  is a tuple of bosonic dephasing channels defined by the corresponding tuple  $(p_i)_{i=1}^\ell$  of probability densities.

As shown recently in Ref. [87], there is a solution for the asymptotic error exponent of antidistinguishability. Namely, the following limit holds:

$$\lim_{n \rightarrow \infty} -\frac{1}{n} \ln p_n^e((p_i)_{i=1}^\ell) = C((p_i)_{i=1}^\ell), \quad (\text{E15})$$

where the multivariate Chernoff divergence  $C((p_i)_{i=1}^\ell)$  is defined as

$$C((p_i)_{i=1}^\ell) := -\ln \inf_s \int d\phi \prod_{i=1}^\ell p_i^{s_i}(\phi), \quad (\text{E16})$$

with the optimization over  $s := (s_i)_{i=1}^\ell$ , a probability vector (satisfying  $s_i \geq 0$  for all  $i \in \{1, \dots, \ell\}$  and  $\sum_{i=1}^\ell s_i = 1$ ). Combining Eqs. (E14) and (E15), we conclude that

$$\lim_{n \rightarrow \infty} -\frac{1}{n} \ln p_n^e((\mathcal{D}_{p_i})_{i=1}^\ell) = C((p_i)_{i=1}^\ell), \quad (\text{E17})$$

where  $(\mathcal{D}_{p_i})_{i=1}^\ell$  is a tuple of bosonic dephasing channels defined by the corresponding tuple  $(p_i)_{i=1}^\ell$  of probability densities.

## APPENDIX F: MULTIMODE BOSONIC DEPHASING CHANNELS

In this appendix, we briefly argue how all of our results apply to multimode BDCs as well. Recall from Ref. [1] that a multimode BDC is defined as

$$\mathcal{D}_p^{(m)}(\rho) := \int_{[-\pi, \pi]^m} d^m \phi p(\phi) e^{-i \sum_j \hat{n}_j \phi_j} \rho e^{i \sum_j \hat{n}_j \phi_j}, \quad (\text{F1})$$

where  $\phi := (\phi_1, \dots, \phi_m)$  is a vector of  $m$  phases and  $p(\phi)$  is a joint probability density function.

This claim holds because the same arguments used in Sec. V for optimality and in Sec. VI for attainability go through. Indeed, the channel  $\mathcal{D}_p^{(m)}$  can be decomposed

similarly to Eq. (35), as

$$\mathcal{D}_p^{(m)} = \mathcal{G}^{(m)} \circ \mathcal{F}_p^{(m)}, \quad (\text{F2})$$

where

$$\mathcal{F}_p^{(m)}(\rho) := \rho \otimes \sigma_p^{(m)}, \quad (\text{F3})$$

$$\sigma_p^{(m)} := \int_{[-\pi, \pi]^m} d^m \phi p(\phi) |\phi\rangle\langle\phi|, \quad (\text{F4})$$

$$|\phi\rangle := |\phi_1\rangle \otimes \dots \otimes |\phi_m\rangle, \quad (\text{F5})$$

$$\begin{aligned} \mathcal{G}^{(m)}(\rho' \otimes \rho'') & := \int_{[-\pi, \pi]^m} d^m \phi e^{-i \sum_j \hat{n}_j \phi_j} \rho' e^{i \sum_j \hat{n}_j \phi_j} \text{Tr}[|\phi\rangle\langle\phi| \rho'']. \end{aligned} \quad (\text{F6})$$

As such, the same simulation argument as before applies, with all distinguishability or estimation tasks limited by the distinguishability or estimability of the classical environmental states of the form  $\sigma_p^{(m)}$ . The attainability part follows because one can simply employ a tensor product of the strategies considered in Sec. VI. That is, for all  $i \in \{1, \dots, m\}$ , suppose that  $(\rho_v^i)_{v \in \mathbb{N}}$  is a sequence of states and  $(M_{\phi_i})_{\phi_i}$  is a corresponding measurement such that Eqs. (39) and (40) hold for a single-mode BDC. Then, it follows immediately that

$$p(\phi) = \lim_{v \rightarrow \infty} p_v(\phi), \quad (\text{F7})$$

for all  $\phi = (\phi_1, \dots, \phi_m)$ , where

$$p_v(\phi) := \text{Tr} \left[ \bigotimes_{i=1}^m M_{\phi_i} \mathcal{D}_p^{(m)} \left( \bigotimes_{i=1}^m \rho_v^i \right) \right]. \quad (\text{F8})$$

Let us finally note that, generalizing the statements in Sec. VII, all of these results for multimode BDCs hold even when a set of multimode BDCs are affected by a common multimode pure-loss channel.

## APPENDIX G: GENERALIZATION TO MULTIPARAMETER CHANNEL ESTIMATION

Our channel estimation results generalize to the setting of multiparameter channel estimation. This follows by considering a cost function of multiple parameters, which results in a risk function. Then, optimizing over all adaptive strategies leads to the same risk function evaluated on the underlying probability density of the BDC. The reasoning here is essentially the same as that used for all other conclusions in our paper: all adaptive strategies for estimation are particular estimation strategies on the underlying probability densities and one can obtain samples from these underlying densities by employing either of the schemes in Secs. VIA or VIB.

### APPENDIX H: PHOTON-NUMBER-SUPERPOSITION METHOD IN THE PRESENCE OF LOSS

In this appendix, we calculate the probability distribution that results when using the photon-number-superposition method if there is photon loss in addition to the action of a phase rotation. By modeling the pure-loss channel as a beam-splitter interaction between the input and an environment mode in the vacuum state [79], we find that the input state transforms as follows:

$$|+_d\rangle \rightarrow \frac{1}{\sqrt{d}} \sum_{n=0}^{d-1} \sum_{\ell=0}^n \sqrt{\eta^{n-\ell} (1-\eta)^\ell} \binom{n}{\ell} |n-\ell\rangle |\ell\rangle. \quad (\text{H1})$$

After the action of the phase rotation  $e^{-i\hat{n}\phi} \otimes I$ , the state becomes

$$|\psi(\phi, \eta)\rangle := \frac{1}{\sqrt{d}} \sum_{n=0}^{d-1} \sum_{\ell=0}^n e^{-i(n-\ell)\phi} \sqrt{\eta^{n-\ell} (1-\eta)^\ell} \binom{n}{\ell} |n-\ell\rangle |\ell\rangle. \quad (\text{H2})$$

The probability for obtaining outcome  $k$  after measuring in the Fourier basis  $\{|u_k\rangle\}_{k=0}^{d-1}$  (defined in Eq. (53)) is then calculated according to the Born rule as follows:

$$\|(\langle u_k | \otimes I) |\psi(\phi, \eta)\rangle\|_2^2 = \left\| \left( \frac{1}{\sqrt{d}} \sum_{m=0}^{d-1} e^{2\pi i m k / d} \langle m | \otimes I \right) \times \left( \frac{1}{\sqrt{d}} \sum_{n=0}^{d-1} \sum_{\ell=0}^n e^{-i(n-\ell)\phi} \sqrt{\eta^{n-\ell} (1-\eta)^\ell} \binom{n}{\ell} |n-\ell\rangle |\ell\rangle \right) \right\|_2^2 \quad (\text{H3})$$

$$= \frac{1}{d^2} \left\| \sum_{n=0}^{d-1} \sum_{\ell=0}^n e^{2\pi i (n-\ell)k/d} e^{-i(n-\ell)\phi} \sqrt{\eta^{n-\ell} (1-\eta)^\ell} \binom{n}{\ell} |\ell\rangle \right\|_2^2 \quad (\text{H4})$$

$$= \frac{1}{d^2} \left\| \sum_{n=0}^{d-1} \sum_{\ell=0}^n e^{i(n-\ell)(2\pi k/d - \phi)} \sqrt{\eta^{n-\ell} (1-\eta)^\ell} \binom{n}{\ell} |\ell\rangle \right\|_2^2 \quad (\text{H5})$$

$$= \frac{1}{d^2} \left( \sum_{m=0}^{d-1} \sum_{\ell'=0}^m e^{-i(m-\ell')(2\pi k/d - \phi)} \sqrt{\eta^{m-\ell'} (1-\eta)^{\ell'}} \binom{m}{\ell'} |\ell'\rangle \right) \times \left( \sum_{n=0}^{d-1} \sum_{\ell=0}^n e^{i(n-\ell)(2\pi k/d - \phi)} \sqrt{\eta^{n-\ell} (1-\eta)^\ell} \binom{n}{\ell} |\ell\rangle \right) \quad (\text{H6})$$

$$= \frac{1}{d^2} \sum_{m,n=0}^{d-1} \sum_{\ell=0}^{\min\{n,m\}} e^{-i(m-n)(2\pi k/d - \phi)} \eta^{(m+n)/2-\ell} (1-\eta)^\ell \sqrt{\binom{m}{\ell} \binom{n}{\ell}}. \quad (\text{H7})$$

We recover the probability distribution in Eq. (54) in the limit as  $\eta \rightarrow 1$ . Numerical experiments indicate that the distribution in Eq. (H7) is highly peaked around  $\phi$  for fixed transmissivity  $\eta \in (0, 1)$  as  $d$  becomes larger.

- 
- [1] L. Lami and M. M. Wilde, Exact solution for the quantum and private capacities of bosonic dephasing channels, *Nat. Photonics* **17**, 525 (2023).  
 [2] C. C. Gerry and P. L. Knight, *Introductory Quantum Optics* (Cambridge University Press, Cambridge, UK, 2004).  
 [3] D. Suter and G. A. Álvarez, Colloquium: Protecting quantum information against environmental noise, *Rev. Mod. Phys.* **88**, 041001 (2016).

- [4] K. H. Wanser, Fundamental phase noise limit in optical fibres due to temperature fluctuations, *Electron. Lett.* **28**, 53 (1992).  
 [5] L.-Z. Jiang and X.-Y. Chen, in *Quantum and Nonlinear Optics*, edited by Q. Gong, G.-C. Guo, and Y.-R. Shen, International Society for Optics and Photonics (SPIE, 2010), Vol. 7846, p. 244, <https://www.spie.org/Publications/Proceedings/Volume/7846>.  
 [6] A. Arqand, L. Memarzadeh, and S. Mancini, Quantum capacity of a bosonic dephasing channel, *Phys. Rev. A* **102**, 042413 (2020).  
 [7] Q. Zhuang, Quantum-enabled communication without a phase reference, *Phys. Rev. Lett.* **126**, 060502 (2021).  
 [8] M. Fanizza, M. Rosati, M. Skotiniotis, J. Calsamiglia, and V. Giovannetti, Squeezing-enhanced communication without a phase reference, *Quantum* **5**, 608 (2021).

- [9] M. Rexiti, L. Memarzadeh, and S. Mancini, Discrimination of dephasing channels, *J. Phys. A: Math. Theor.* **55**, 245301 (2022).
- [10] A. Arqand, L. Memarzadeh, and S. Mancini, Energy-constrained LOCC-assisted quantum capacity of the bosonic dephasing channel, *Entropy* **25**, 1001 (2023).
- [11] B. M. Terhal, Quantum error correction for quantum memories, *Rev. Mod. Phys.* **87**, 307 (2015).
- [12] G. Q. Ai, Exponential suppression of bit or phase errors with cyclic error correction, *Nature* **595**, 383 (2021).
- [13] J. S. Sidhu, M. S. Bullock, S. Guha, and C. Lupo, Linear optics and photodetection achieve near-optimal unambiguous coherent state discrimination, *Quantum* **7**, 1025 (2023).
- [14] Z. Huang and C. Lupo, Quantum hypothesis testing for exoplanet detection, *Phys. Rev. Lett.* **127**, 130502 (2021).
- [15] Z. Huang, C. Schwab, and C. Lupo, Ultimate limits of exoplanet spectroscopy: A quantum approach, *Phys. Rev. A* **107**, 022409 (2023).
- [16] H. Shi, Z. Zhang, S. Pirandola, and Q. Zhuang, Entanglement-assisted absorption spectroscopy, *Phys. Rev. Lett.* **125**, 180502 (2020).
- [17] J. Bae and L.-C. Kwek, Quantum state discrimination and its applications, *J. Phys. A: Math. Theor.* **48**, 083001 (2015).
- [18] A. Cheffles, in *Quantum State Estimation*, edited by M. Paris and J. Řeháček (Springer-Verlag, Berlin, 2004), p. 467.
- [19] G. Chiribella, G. M. D’Ariano, and P. Perinotti, Memory effects in quantum channel discrimination, *Phys. Rev. Lett.* **101**, 180501 (2008).
- [20] R. Duan, Y. Feng, and M. Ying, Perfect distinguishability of quantum operations, *Phys. Rev. Lett.* **103**, 210501 (2009).
- [21] M. Piani and J. Watrous, All entangled states are useful for channel discrimination, *Phys. Rev. Lett.* **102**, 250501 (2009).
- [22] M. Hayashi, Discrimination of two channels by adaptive methods and its application to quantum system, *IEEE Trans. Inf. Theory* **55**, 3807 (2009).
- [23] A. W. Harrow, A. Hassidim, D. W. Leung, and J. Watrous, Adaptive versus nonadaptive strategies for quantum channel discrimination, *Phys. Rev. A* **81**, 032339 (2010).
- [24] W. Matthews, M. Piani, and J. Watrous, Entanglement in channel discrimination with restricted measurements, *Phys. Rev. A* **82**, 032302 (2010).
- [25] T. Cooney, M. Mosonyi, and M. M. Wilde, Strong converse exponents for a quantum channel discrimination problem and quantum-feedback-assisted communication, *Commun. Math. Phys.* **344**, 797 (2016).
- [26] S. Pirandola and C. Lupo, Ultimate precision of adaptive noise estimation, *Phys. Rev. Lett.* **118**, 100502 (2017).
- [27] M. Takeoka and M. M. Wilde, Optimal estimation and discrimination of excess noise in thermal and amplifier channels, [ArXiv:1611.09165](https://arxiv.org/abs/1611.09165).
- [28] D. Puzzioli and J. Watrous, Ancilla dimension in quantum channel discrimination, *Ann. Henri Poincaré* **18**, 1153 (2017).
- [29] M. M. Wilde, M. Berta, C. Hirche, and E. Kaur, Amortized channel divergence for asymptotic quantum channel discrimination, *Lett. Math. Phys.* **110**, 2277 (2020).
- [30] X. Wang and M. M. Wilde, Resource theory of asymmetric distinguishability for quantum channels, *Phys. Rev. Res.* **1**, 033169 (2019).
- [31] K. Fang, O. Fawzi, R. Renner, and D. Sutter, Chain rule for the quantum relative entropy, *Phys. Rev. Lett.* **124**, 100501 (2020).
- [32] V. Katariya and M. M. Wilde, Evaluating the advantage of adaptive strategies for quantum channel distinguishability, *Phys. Rev. A* **104**, 052406 (2021).
- [33] K. Fang and H. Fawzi, Geometric Rényi divergence and its applications in quantum channel capacities, *Commun. Math. Phys.* **384**, 1615 (2021).
- [34] B. Bergh, N. Datta, R. Salzmänn, and M. M. Wilde, Parallelization of sequential quantum channel discrimination in the non-asymptotic regime, *IEEE Trans. Inf. Theory* **70**, 2617 (2024).
- [35] F. Salek, M. Hayashi, and A. Winter, Usefulness of adaptive strategies in asymptotic quantum channel discrimination, *Phys. Rev. A* **105**, 022419 (2022).
- [36] B. Bergh, J. Kochanowski, R. Salzmänn, and N. Datta, Infinite dimensional asymmetric quantum channel discrimination, [ArXiv:2308.12959](https://arxiv.org/abs/2308.12959).
- [37] S. Zhou and L. Jiang, Asymptotic theory of quantum channel estimation, *PRX Quantum* **2**, 010343 (2021).
- [38] Q. Liu, Z. Hu, H. Yuan, and Y. Yang, Optimal strategies of quantum metrology with a strict hierarchy, *Phys. Rev. Lett.* **130**, 070803 (2023).
- [39] B. M. Escher, R. L. de Matos Filho, and L. Davidovich, General framework for estimating the ultimate precision limit in noisy quantum-enhanced metrology, *Nat. Phys.* **7**, 406 (2011).
- [40] R. Demkowicz-Dobrzański, J. Kolodyński, and M. Guță, The elusive Heisenberg limit in quantum-enhanced metrology, *Nat. Commun.* **3**, 1063 (2012).
- [41] R. Demkowicz-Dobrzański and L. Maccone, Using entanglement against noise in quantum metrology, *Phys. Rev. Lett.* **113**, 250801 (2014).
- [42] V. Katariya and M. M. Wilde, Geometric distinguishability measures limit quantum channel estimation and discrimination, *Quantum Inf. Process.* **20**, 78 (2021).
- [43] H. L. Van Trees, *Detection, Estimation, and Modulation Theory, Part I: Detection, Estimation, and Linear Modulation Theory* (John Wiley & Sons, 2004), <https://onlinelibrary.wiley.com/doi/book/10.1002/0471221082>.
- [44] V. Y. F. Tan, Asymptotic estimates in information theory with non-vanishing error probabilities, *Found. Trends Commun. Inf. Theory* **11**, 1 (2014).
- [45] A. P. Korostelev and O. Korosteleva, *Mathematical Statistics: Asymptotic Minimax Theory* (American Mathematical Society, Providence, Rhode Island, USA, 2011), Vol. 119.
- [46] C. L. Canonne, Topics and techniques in distribution testing: A biased but representative sample, *Found. Trends Commun. Inf. Theory* **19**, 1032 (2022).
- [47] H. Chernoff, A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations, *Ann. Math. Stat.* **23**, 493 (1952).
- [48] M. F. Sacchi, Optimal discrimination of quantum operations, *Phys. Rev. A* **71**, 062340 (2005).
- [49] S. Lloyd, Enhanced sensitivity of photodetection via quantum illumination, *Science* **321**, 1463 (2008).

- [50] S.-H. Tan, B. I. Erkmen, V. Giovannetti, S. Guha, S. Lloyd, L. Maccone, S. Pirandola, and J. H. Shapiro, Quantum illumination with Gaussian states, *Phys. Rev. Lett.* **101**, 253601 (2008).
- [51] M. M. Wilde, M. Tomamichel, S. Lloyd, and M. Berta, Gaussian hypothesis testing and quantum illumination, *Phys. Rev. Lett.* **119**, 120501 (2017).
- [52] V. Strassen, in *Transactions of the Third Prague Conference on Information Theory etc* (C. Academy of Sciences, Prague, 1962), p. 689. English translation available at <https://pi.math.cornell.edu/~pmlut/strassen.pdf>.
- [53] C. Stein, Information and Comparison of Experiments (unpublished), Charles Stein papers (SC1224), Box 12, Folder 7, Department of Special Collections and University Archives, Stanford University Libraries.
- [54] H. Chernoff, Large-sample theory: Parametric case, *Ann. Math. Stat.* **27**, 1 (1956).
- [55] K. Matsumoto, On metric of quantum channel spaces, [ArXiv:1006.0300](https://arxiv.org/abs/1006.0300).
- [56] M. J. Holland and K. Burnett, Interferometric detection of optical phase shifts at the Heisenberg limit, *Phys. Rev. Lett.* **71**, 1355 (1993).
- [57] H. M. Wiseman and R. B. Killip, Adaptive single-shot phase measurements: The full quantum theory, *Phys. Rev. A* **57**, 2169 (1998).
- [58] R. Demkowicz-Dobrzański, M. Jarzyna, and J. Kołodyński, Quantum limits in optical interferometry, *Prog. Opt.* **60**, 345 (2015).
- [59] H. T. Dinani and D. W. Berry, Loss-resistant unambiguous phase measurement, *Phys. Rev. A* **90**, 023856 (2014).
- [60] M. A. Rodríguez-García, M. T. DiMario, P. Barberis-Blostein, and F. E. Becerra, Determination of the asymptotic limits of adaptive photon counting measurements for coherent-state optical phase estimation, *npj Quantum Inf.* **8**, 94 (2022).
- [61] M. Cooper, L. J. Wright, C. Söller, and B. J. Smith, Experimental generation of multi-photon Fock states, *Opt. Express* **21**, 5309 (2013).
- [62] J. P. Dowling, Quantum optical metrology—The lowdown on high-N00N states, *Contemp. Phys.* **49**, 125 (2008).
- [63] D. T. Pegg and S. M. Barnett, Phase properties of the quantized single-mode electromagnetic field, *Phys. Rev. A* **39**, 1665 (1989).
- [64] V. Buzek, R. Derka, and S. Massar, Optimal quantum clocks, *Phys. Rev. Lett.* **82**, 2207 (1999).
- [65] M. Hassani, C. Macchiavello, and L. Maccone, Digital quantum estimation, *Phys. Rev. Lett.* **119**, 200502 (2017).
- [66] K. Hoffman, *Banach Spaces of Analytic Functions*, Dover Books on Mathematics (Dover Publications, 2007).
- [67] S. Olivares, S. Cialdi, F. Castelli, and M. G. A. Paris, Homodyne detection as a near-optimum receiver for phase-shift-keyed binary communication in the presence of phase diffusion, *Phys. Rev. A* **87**, 050303(R) (2013).
- [68] L. S. Martin, W. P. Livingston, S. Hacoheh-Gourgy, H. M. Wiseman, and I. Siddiqi, Implementation of a canonical phase measurement with quantum feedback, *Nat. Phys.* **16**, 1046 (2020).
- [69] M. G. Genoni, S. Olivares, and M. G. A. Paris, Optical phase estimation in the presence of phase diffusion, *Phys. Rev. Lett.* **106**, 153603 (2011).
- [70] J. Trapani, B. Teklu, S. Olivares, and M. G. A. Paris, Quantum phase communication channels in the presence of static and dynamical phase diffusion, *Phys. Rev. A* **92**, 012317 (2015).
- [71] Z. Luo, Y. Zhan, and E. Jonckheere, in *2020 IEEE/CIC International Conference on Communications in China (ICCC) (2020)*, p. 306.
- [72] P. Leviant, Q. Xu, L. Jiang, and S. Rosenblum, Quantum capacity and codes for the bosonic loss-dephasing channel, *Quantum* **6**, 821 (2022).
- [73] F. A. Mele, F. Salek, V. Giovannetti, and L. Lami, Quantum communication on the bosonic loss-dephasing channel, [ArXiv:2401.15634](https://arxiv.org/abs/2401.15634).
- [74] A. Winter, Energy-constrained diamond norm with applications to the uniform continuity of continuous variable channel capacities, [ArXiv:1712.10267](https://arxiv.org/abs/1712.10267).
- [75] J. Neyman and E. S. Pearson, IX. On the problem of the most efficient tests of statistical hypotheses, *Philos. Trans. R. Soc. Lond. A* **231**, 289 (1933).
- [76] J. M. Renes, Relative submajorization and its use in quantum resource theories, *J. Math. Phys.* **57**, 122202 (2016).
- [77] F. Buscemi and G. Gour, Quantum relative Lorenz curves, *Phys. Rev. A* **95**, 012110 (2017).
- [78] A. Zygmund, *Trigonometric Series: Vols. I, II* (Cambridge University Press, Cambridge, United Kingdom, 1968), 2nd ed.
- [79] A. Serafini, *Quantum Continuous Variables: A Primer of Theoretical Methods* (CRC Press, Boca Raton, Florida, USA, 2017).
- [80] T. S. Han and K. Kobayashi, The strong converse theorem for hypothesis testing, *IEEE Trans. Inf. Theory* **35**, 178 (1989).
- [81] W. Hoeffding, Asymptotically optimal tests for multinomial distributions, *Ann. Math. Stat.* **36**, 369 (1965).
- [82] N. P. Salikhov, Asymptotic properties of error probabilities of tests for distinguishing between several multinomial testing schemes, *Dokl. Akad. Nauk.* **209**, 54 (1973).
- [83] E. N. Torgersen, Measures of information based on comparison with total information and with total ignorance, *Ann. Stat.* **9**, 638 (1981).
- [84] C. C. Leang and D. H. Johnson, On the asymptotics of  $m$ -hypothesis Bayesian detection, *IEEE Trans. Inf. Theory* **43**, 280 (1997).
- [85] N. P. Salikhov, On one generalization of Chernov's distance, *Theory Probab. Appl.* **43**, 239 (1999).
- [86] N. P. Salikhov, Optimal sequences of tests for several polynomial schemes of trials, *Theory Probab. Appl.* **47**, 286 (2003).
- [87] H. K. Mishra, M. Nussbaum, and M. M. Wilde, On the optimal error exponents for classical and quantum antidinguishability, [ArXiv:2309.03723](https://arxiv.org/abs/2309.03723).