

Out-of-Distribution Generalization for Learning Quantum Channels with Low-Energy Coherent States

Jason L. Pereira^{1,2,3,*}, Quntao Zhuang^{2,4} and Leonardo Banchi^{5,1}

¹*Istituto Nazionale di Fisica Nucleare (INFN) Sezione di Firenze, via G. Sansone 1, Sesto Fiorentino (FI) I-50019, Italy*

²*Ming Hsieh Department of Electrical and Computer Engineering, University of Southern California, Los Angeles, California 90089, USA*

³*Sorbonne Université, CNRS, LIP6, 4 Place Jussieu, Paris F-75005, France*

⁴*Department of Physics and Astronomy, University of Southern California, Los Angeles, California 90089, USA*

⁵*Department of Physics and Astronomy, University of Florence, via G. Sansone 1, Sesto Fiorentino (FI) I-50019, Italy*



(Received 12 February 2025; revised 9 July 2025; accepted 13 August 2025; published 8 October 2025)

When experimentally learning the action of a continuous-variable quantum process by probing it with inputs, there will often be some restriction on the input states used. One experimentally simple way to probe a quantum channel is to use low-energy coherent states. Learning a quantum channel in this way presents difficulties, due to the fact that two channels may act similarly on low-energy inputs but very differently for high-energy inputs. They may also act similarly on coherent-state inputs but differently on nonclassical inputs. Extrapolating the behavior of a channel for more general input states from its action on the far more limited set of low-energy coherent states is a case of out-of-distribution generalization. To be sure that such generalization gives meaningful results, one needs to relate error bounds for the training set to bounds that are valid for all inputs. We show that for any pair of channels that act sufficiently similarly on low-energy coherent-state inputs, one can bound how different the input-output relations are for any (high-energy or highly nonclassical) input. This proves that out-of-distribution generalization is always possible for learning quantum channels using low-energy coherent states, as long as enough samples are used.

DOI: [10.1103/5m7p-kbf3](https://doi.org/10.1103/5m7p-kbf3)

I. INTRODUCTION

Many physical systems can be modeled as quantum channels that map one (input) quantum state into another (the output). Learning about a physical process can therefore be regarded as the task of finding the input-output relations of the enacted quantum channel [1]. Experimentally, we would do this by sending probe states through the channel and characterizing the outputs. However, a full characterization can be very “expensive” in terms of the number of experiments required. This is especially true for continuous-variable (CV) systems [2,3]. Since the dimension is infinite, it is not possible to, e.g., probe a channel with every basis state individually.

Instead, we must probe the channel with some finite set of input states and then extrapolate the results to all possible inputs. One particularly simple set of input states is the coherent states. These are classical states that can easily be generated experimentally. However, they form an over-complete basis and so are sufficient to characterize a CV channel [4–7]. In any real experiment, we will also have a maximum energy for our probe states, so the inputs will be restricted to coherent states of bounded energy.

Whenever we extrapolate from limited data, we need to understand how reliable our learned answer is outside of our limited data set, i.e., whether it generalizes. The problem is compounded if we want our solution to hold even outside of the parameter region in which our data set lies (i.e., outside the space spanned by inputs in the training distribution). If our data set consists of only the outputs for low-energy classical inputs, we may wonder how reliable our input-output relations are for higher-energy or nonclassical systems. This is called the out-of-distribution generalization problem.

*Contact author: Jason.Pereira@lip6.fr

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.

The problem of out-of-distribution generalization for learning quantum channels has been investigated in the discrete-variable (DV) unitary case [8,9]. In Ref. [8], it is shown that by probing a unitary with a restricted set of input states, one can learn how it acts on a very different set of inputs. The authors also state that their proofs can be extended to a class of channels called doubly stochastic channels. While many follow-on works about out-of-distribution generalization in the DV setting have largely focused on the unitary setting, the noisy versions of certain unitary-learning problems can also be considered as specific instances of channel learning [10]. Other works about incoherent learning of quantum dynamics can be seen as out-of-distribution generalization results for specific classes of quantum to classical channels [11,12]. Similarly, in Ref. [13], the authors have shown that observables learned on one distribution can be generalized to other distributions, as long as both distributions have a particular property called local flatness.

In the CV case, it is known that we can extrapolate the input-output relations for coherent states to general states [4,5,14] but not how an uncertainty in those relations for coherent states propagates when applied out-of-distribution to general states. Since we will never learn the action of a process perfectly (with finite samples), even on a limited set of inputs, an understanding of out-of-distribution error propagation is crucial.

In particular, understanding of out-of-distribution generalization for CV channels impacts the study of many different quantum systems. In terms of quantum sensing and communication systems, CV channels are important for optomechanical systems [15–17], microwave cavity quantum electrodynamics (cQED) systems [18], radio-frequency photonic sensors [19], radar and lidar detection [20], optical spectroscopy processes [21–24], and optical communication and quantum networks [25,26]. At the same time, CV channels also model nanophotonic chips for machine learning [27] and optical quantum computers [28].

In this work, we present a general formalism for understanding out-of-distribution generalization in CV channels. We show that if any pair of CV channels has sufficiently close outputs for low-energy coherent-state inputs, it is possible to construct a bound on the output distance for any input state. Hence, we prove that out-of-distribution generalization is always possible for learning quantum channels using low-energy coherent states.

We start, in Sec. II, by introducing the problem and giving an overview of our main results. In Sec. III, we go into more detail about the generalization from low-energy coherent states to higher-energy coherent states. In Sec. IV, we investigate some examples of this type of generalization. In Sec. V, we show the extension from coherent states to general states, and in Sec. VI, we give examples of this extension. In Sec. VII, we discuss how our work can be

framed in terms of quantum process tomography, quantum metrology, and quantum channel discrimination. In Sec. VIII, we present our conclusions and discuss possible future research directions.

II. OVERVIEW

Learning a quantum channel means understanding how any input state will be mapped to an output state by that channel. We can express our knowledge of the target channel as a “learned” channel that mimics as closely as possible the input-output relations of the target. This knowledge could (nonexhaustively) take the form of a classical description of the input-output relations, the parameter values for a specific form of parametrized channel, or some set of settings for a physical device that lets us enact the learned channel. The better we learn the target channel, the closer the two channels will be.

Our aim is to bound the error in learning the action of a CV channel on an arbitrary state after learning on samples from a restricted set of input coherent states. Specifically, we want to bound the distance (trace norm) between the output of a target channel, Ψ , and our learned channel, Φ , for test state ρ , with average photon number $\bar{n}$. To be useful (nontrivial), the bound on $\|\Psi[\rho] - \Phi[\rho]\|$ should approach 0 as we learn Ψ better over our restricted set of inputs. We will occasionally refer to the “trace distance,” which is just half of the trace norm of the state difference.

If such a bound exists, this tells us that *out-of-distribution generalization is possible*, in the sense that learning the input-output relations for a very restricted set of classical state inputs is sufficient to mimic them for completely general (including highly nonclassical) input states. In some cases, this may require us to learn the channel very well for our restricted set of input states. The tightness of the bound depends on the class of channels to which Ψ and Φ belong, e.g., Gaussian [3,29], unitary operation, etc. If ρ has a known finite negativity (of its P representation) $\mathcal{N}$ [30], we may also wish to take this into account to obtain a tighter bound. A bound of this type can be constructed in three stages (as illustrated in Fig. 1):

- (1) Given N samples of low-energy coherent states sent through a channel Ψ , bound the error in learning the action of this channel on the same distribution of low-energy coherent states. As sketched in Fig. 1(c), we have N samples of the form $\Psi[|re^{i\phi}\rangle\langle re^{i\phi}|_{\text{coh}}]$, for $r \leq \tau$. We find a protocol that uses them to learn a channel Φ , such that (with high probability) $\|(\Psi - \Phi)[|re^{i\phi}\rangle\langle re^{i\phi}|_{\text{coh}}]\| \leq \epsilon_0$ for all $r \leq \tau$, where ϵ_0 is a decreasing function of N (so that more samples let us learn the channel better). In other words, we bound the “in-distribution” error for learning the action of channel Ψ on coherent states with

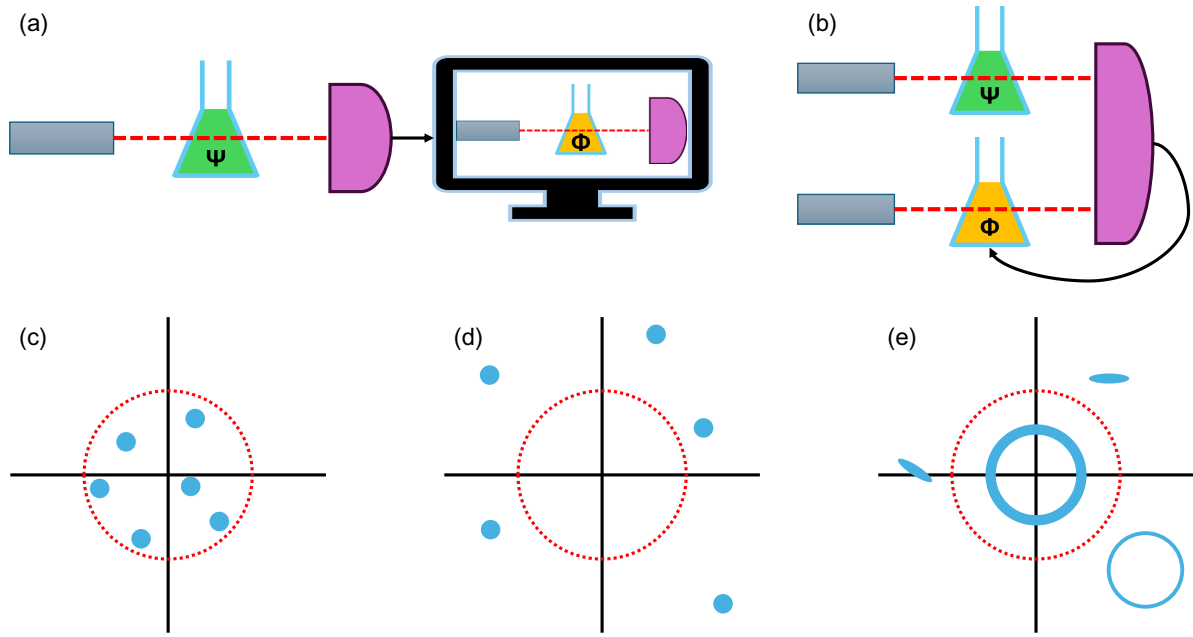


FIG. 1. (a),(b) Two different ways in which we could learn a channel, Φ , that mimics the action of a target process, Ψ , for low-energy coherent states: (a) refers to “classical” learning of a quantum channel, while (b) shows a form of quantum machine learning. In this illustration, we depict probing the optical properties of an unknown (green) sample, although the process could be a magnetic field, a nonlinear medium, a reflective cavity, or any other transformation that could be applied to a state of light. We search through a set of possible (yellow) optical media to find the one that best imitates the green substance. In (a), we use a (finite-energy) laser to send low-energy coherent states through the target and characterize the outputs with measurements. The results are postprocessed (e.g., by feeding them into a classical computer, per the diagram) to obtain classical knowledge of the input-output relations. We can then “simulate” the action of Ψ on an unknown state, using our classical computer and the transformation Φ . In (b), we have a tunable quantum device (such as a different substance with tunable optical properties, a tunable magnetic field, or some parametrized optical circuit) that we want to use to imitate the target process. We probe both processes with the same coherent states (we show two lasers, but a real implementation could use a single laser and a balanced beam splitter, to ensure that the probing states are identical). Instead of characterizing both outputs, we measure them jointly, to determine their trace distance. We could then use the measurement results to tune the quantum device and so reduce the distance. Once Φ is tuned, we hope that an unknown state sent through the quantum device will be transformed similarly to if it were sent through Ψ . (c)–(e) The three types of input for which we can bound the closeness of two channels. (c) This type corresponds to the input states for (a) and (b); the probes used in our physical measurements are low-energy coherent states. The dotted line shows the maximum average energy of the inputs. We can construct a bound on the output distance between our target and learned processes, Ψ and Φ , based on our measurement results. (d) Here, we consider higher-energy coherent-state inputs. These are outside of the class of inputs for which we have actual measurement results; rather, we want to be able to trust that our simulation, Φ , of Ψ is still accurate for these inputs [in the scenario of (a)] or that our tuned quantum device still mimics our target well [in the scenario of (b)]. We use our bound for the inputs shown in (c) to construct a bound on the inputs shown in (d). The scaling of this bound depends on the class of channels to which the target and learned channel belong. (e) Finally, we consider other types of input states, including nonclassical states such as squeezed vacuums and Fock states. We again want to trust that Φ is a good simulation of Ψ for these types of input and so we extend our bounds for the inputs in (d) to the inputs in (e).

an average photon number of $\leq \tau^2$. Decreasing ϵ_0 means using more samples to find a new channel Φ , such that the output distance (for our subset of coherent states) is smaller. We require that $\epsilon_0 \rightarrow 0$ as $N \rightarrow \infty$.

- (2) Assume that stage 1 was successful. In stage 2, depicted in Fig. 1(d), we aim to solve the restricted case of out-of-distribution generalization in which we only consider coherent states beyond the energy of the actual probe states from stage 1. More precisely, we need to find a function $\epsilon(\epsilon_0, r^2)$ such that $\|(\Psi - \Phi)[|re^{i\phi}\rangle\langle re^{i\phi}|_{\text{coh}}]\| \leq \epsilon(\epsilon_0, r^2)$ for all r . We

also require that $\epsilon(\epsilon_0, r^2)$ is a decreasing function of ϵ_0 that reaches 0 for $\epsilon_0 = 0$, ensuring that our out-of-distribution error reaches 0 if our in-distribution error does.

- (3) Assume that stage 2 was successful. In the final stage, depicted in Fig. 1(e), we generalize to arbitrary states: for an arbitrary input ρ , we need to bound $\|(\Psi - \Phi)[\rho]\|$ in terms of $\bar{n}(\rho)$, $\epsilon(\epsilon_0, r)$ and possibly $\mathcal{N}(\rho)$ (if finite). This gives us a complete bound for out-of-distribution generalization in the general case and we require that it reaches 0 for $\epsilon = 0$.

The first stage involves using a limited set of inputs to probe a quantum process and applying in-distribution generalization to bound the error over this set. Learning the input-output relations for a channel over a limited set of states is a somewhat under-specified problem, since any concrete statements about it would strongly depend on the learning process and the classes of channel under consideration. This is intentional, since the aim of this paper is to be as general as possible and to give a formalism for extending a bound for low-energy coherent state inputs to general states, regardless of how the initial bound has been obtained. However, in Sec. VII, we discuss briefly how this task has been addressed in the frameworks of quantum process tomography, quantum metrology, and quantum machine learning. Elsewhere, we do not focus on this step but, rather, assume that we already have the bound $\|(\Psi - \Phi)[|re^{i\phi}\rangle\langle re^{i\phi}|_{\text{coh}}]\| \leq \epsilon_0$ (for $r \leq \tau$) and want to extend it to more general inputs. ϵ_0 lies in the range $[0, 2]$ (with a value of 2 being trivial).

We focus on the remaining two stages. Our goals are twofold: we want to make some general statements that hold for all classes of channels and all input states, and we also want to give tighter bounds for cases in which we have more information about the channels and/or input states. Our main results can be informally summarized as follows:

- (1) If two channels have exactly the same output for low-energy coherent-state inputs, they also have the same output for high-energy coherent-state inputs. If two channels have sufficiently similar outputs (i.e., a tightly bounded output distance) for low-energy coherent-state inputs, they also have a bounded output distance for high-energy coherent-state inputs. As the channel outputs become more similar for low-energy inputs, they also become more similar for high-energy inputs (as one bound converges to 0, the other does too). For particular example classes of channels, we derive explicit bounds.
- (2) If two channels have similar outputs for coherent-state inputs, they also have a bounded output distance for any input state. As the channel outputs become more similar for coherent states, they also become more similar for general inputs. We give explicit expressions for particular examples of non-classical states.

Together, these results show that *out-of-distribution generalization is possible for every input state and all target channels*, in the sense that a sufficiently small in-distribution error bound can always be translated to an out-of-distribution error bound. It is therefore possible to achieve an arbitrarily small error in reproducing the input-output relations for any input state, as long as sufficient

samples are taken from the set of low-energy coherent states.

A. Main results: Out-of-distribution generalization for coherent states

We start by investigating how the output distance between a pair of channels, for a coherent-state input, scales with the energy of the input. The question we want to answer is: *If the outputs of two channels are close for low-energy pure coherent-state inputs, how quickly do they become far apart as the input energy increases?* The first observation that we make (proven in Appendix B) is as follows.

Lemma 1. Suppose that $\|(\Psi - \Phi)[|re^{i\phi}\rangle\langle re^{i\phi}|_{\text{coh}}]\| = 0$ for all ϕ and all $r \leq \tau$, for some $\tau > 0$. Then, for all ϕ and all r , $\|(\Psi - \Phi)[|re^{i\phi}\rangle\langle re^{i\phi}|_{\text{coh}}]\| = 0$.

In other words, if we learn the action of a channel exactly on an (infinite but compact) subset of the pure coherent states, we also learn it exactly on all coherent states. While intuitive, this is important to state, as otherwise one might think there are pairs of channels that exactly coincide for low-energy coherent states but that diverge elsewhere in phase space. This follows from the over-completeness of the coherent states or, alternatively, from the fact that every differential of every matrix element (in any basis) of $(\Psi - \Phi)[|re^{i\phi}\rangle\langle re^{i\phi}|_{\text{coh}}]$ is identically zero and is in line with what we might expect from Refs. [5] and [14].

On the other hand, even if we can only bound the error on the subset with some $\epsilon_0 > 0$, we can still bound the error for all coherent states. In Appendix I, we show the following theorem.

Theorem 1. Suppose that $\|(\Psi - \Phi)[|re^{i\phi}\rangle\langle re^{i\phi}|_{\text{coh}}]\| \leq \epsilon_0$ for all ϕ and all $r \leq \tau$, for some $\tau > 0$. Then, there exists a concave (in r^2) function $\epsilon(\epsilon_0, r^2)$ such that $\|(\Psi - \Phi)[|re^{i\phi}\rangle\langle re^{i\phi}|_{\text{coh}}]\| \leq \epsilon(\epsilon_0, r^2)$ for all ϕ and r , with the following properties:

- (1) For any r and for $\epsilon'_0 < \epsilon_0$, $\epsilon(\epsilon'_0, r^2) \leq \epsilon(\epsilon_0, r^2)$.
- (2) For any r , $\epsilon(\epsilon_0, r^2) \rightarrow 0$ as $\epsilon_0 \rightarrow 0$.

We express ϵ as a function of r^2 , rather than r , since r^2 is the mean photon number of the coherent state input and so is a more natural variable to use. This theorem shows that if we learn a channel sufficiently well on a subset of the coherent states, our target and learned channels will also have similar outputs for coherent states outside of that subset.

Although Theorem 1 shows the existence of a bounding function $\epsilon(\epsilon_0, r^2)$ that is nontrivial for sufficiently small ϵ_0 (and, in fact, the proof is constructive), the function found in Appendix I requires very small ϵ_0 to be nontrivial for

large r . It is therefore better to find channel-specific ϵ functions. In Sec. IV, we use specific examples to demonstrate how we can obtain analytical expressions for the function $\epsilon(\epsilon_0, r^2)$ in certain cases, e.g., if our target channel is in a known class.

B. Main results: Out-of-distribution generalization for general states

Next, we address more general states. The question now is: *If the outputs of two channels are within a bounded distance of each other for coherent-state inputs, how far apart are they for other types of input states?* We now assume that we have a function $\epsilon(\epsilon_0, r^2)$ that upper bounds the trace norm between channel outputs for a coherent-state input. Starting from this function, the goal is to prove a bound on the trace norm between channel outputs for any bounded-energy input state ρ .

Such a bound can be obtained using the P representation of CV states and the convexity of the trace norm. The calculations are outlined in Sec. V and presented in greater detail in the appendixes. In the most general case, the output distance for any nonclassical input can be bounded with the following theorem.

Theorem 2. Let Ψ and Φ be a pair of quantum channels for which $\|(\Psi - \Phi)[|re^{i\phi}\rangle\langle re^{i\phi}|_{\text{coh}}]\| \leq \epsilon(\epsilon_0, r^2)$ for all r and for some concave (in r^2) function ϵ . Let ρ be any input state with average photon number $\bar{n}$. Then, for any positive integer M and any $0 \leq s < \frac{1}{2}$ obeying $((1-s)(1-2s))/(s(M-1)) > 1$,

$$\|(\Psi - \Phi)[\rho]\| \leq \frac{2(1-s)^M M}{s^{M-1}(1-2s)} \epsilon\left(\epsilon_0, \frac{s(1-s)(M+1)}{1-2s}\right) + 4\sqrt{s(1+2\bar{n})} + 4\sqrt{\frac{\bar{n}}{M}}. \quad (1)$$

This theorem has two free parameters, s and M , which must be tuned to attain the tightest bound. It has been obtained by bounding the output distance for an input state close to ρ , denoted $\sigma_{s,M}$, and applying a continuity bound. The physical meaning of the free parameters is as follows: to obtain $\sigma_{s,M}$ from ρ , apply a hard energy cutoff of $M-1$ photons, then pass the resulting state through a Gaussian additive noise channel, with a noise parametrized by s . We can fix particular relationships between s , M , and $\bar{n}$ to study specific cases, though any such choice may not be optimal (in terms of giving the tightest bound). For instance, we can study the scaling with the input energy, $\bar{n}$, by making s proportional to M^{-1} , with some constant κ , obtaining the following corollary.

Corollary 1. Let ρ be any input state with average photon number $\bar{n}$. For any $\kappa > 1$ and positive integer M ,

$$\|(\Psi - \Phi)[\rho]\| \leq \mathcal{O}[e^{M \log(M\kappa)} \kappa^{-1} \epsilon(\epsilon_0, \kappa^{-1})] + \mathcal{O}\left[\bar{n}^{\frac{1}{2}} M^{-\frac{1}{2}}\right]. \quad (2)$$

Equation (2) allows us to make a few observations. First, if we fix κ , then as ϵ_0 approaches 0, $\epsilon(\epsilon_0, \kappa^{-1})$ will also approach zero and we will be able to send $M \rightarrow \infty$, so that Eq. (1) approaches 0 too. Hence, as long as we are able to keep decreasing ϵ_0 , by taking more samples, we are also able to achieve an arbitrarily small output trace distance for any input state. Second, fixing $M \sim \bar{n}$, we see that the leading-order term scales at most polyexponentially in the input energy (approximately $\mathcal{O}[e^{\bar{n} \log(\bar{n})} \epsilon(\epsilon_0, \kappa^{-1})]$). A linear increase in the allowed $\bar{n}$ (for fixed error) can therefore always be achieved with an (at most) exponential decrease in $\epsilon(\epsilon_0, \kappa^{-1})$.

In the specific case of a classical (but not necessarily coherent-state) input, we get the following corollary (see Appendix D).

Corollary 2. Let ρ_{class} be a classical input state with average photon number $\bar{n}$. Then,

$$\|(\Psi - \Phi)[\rho_{\text{class}}]\| \leq \epsilon(\epsilon_0, \bar{n}). \quad (3)$$

In the more general, but still not universal, case in which the input state has a finite negativity, $\mathcal{N}$, and a known P representation, $P(re^{i\phi})$, we define the quantities

$$\mu_P = 1 + 2\mathcal{N} = \iint |P(re^{i\phi})| r d\phi dr, \quad v_P = \iint |P(re^{i\phi})| r^3 d\phi dr, \quad \bar{n}_{\pm} = \frac{\iint P_{\pm}(re^{i\phi}) r^3 d\phi dr}{\iint P_{\pm}(re^{i\phi}) r d\phi dr}, \quad (4)$$

where $P_{\pm}$ means that we integrate over the positive or negative domain of P . Then, we have the following corollary.

Corollary 3. Let ρ_P be an input state with P representation $P(re^{i\phi})$ and finite negativity $\mathcal{N}$. Then,

$$\|(\Psi - \Phi)[\rho_P]\| \leq (1 + \mathcal{N})\epsilon(\epsilon_0, \bar{n}_+) + \mathcal{N}\epsilon(\epsilon_0, \bar{n}_-) \leq \mu_P \epsilon\left(\epsilon_0, \frac{v_P}{\mu_P}\right). \quad (5)$$

III. OUT-OF-DISTRIBUTION GENERALIZATION FOR COHERENT STATES

Suppose that two channels, Ψ and Φ , output similar output states for input coherent states in a given region. This similarity is quantified by $\|(\Psi - \Phi)[|re^{i\phi}\rangle\langle re^{i\phi}|_{\text{coh}}]\| \leq \epsilon_0$

for $re^{i\phi}$ lying in some closed subset. We want to generalize this bound to every complex value of $re^{i\phi}$. For simplicity (and physical relevance), we assume that the subset of coherent states over which we know Ψ and Φ are similar is a circle centered on the origin, so that the radius of the subset corresponds to the maximum average energy of the coherent states lying in it. Hence, our starting point is $\|(\Psi - \Phi)[|re^{i\phi}\rangle\langle re^{i\phi}|_{\text{coh}}]\| \leq \epsilon_0$ for $r^2 \leq \tau^2$ and our goal is to find $\epsilon(\epsilon_0, r^2)$ such that $\|(\Psi - \Phi)[|re^{i\phi}\rangle\langle re^{i\phi}|_{\text{coh}}]\| \leq \epsilon(\epsilon_0, r^2)$ for all r (with the inequality as tight as possible).

It is not necessarily the case that the channel outputs grow further apart as r increases. For instance, if Ψ and Φ are both replacement channels (i.e., channels that ignore the input and prepare some fixed output), the output distance will be constant for any input. For some channels, the channel outputs could even converge for large r . However, in either of these scenarios, out-of-distribution generalization is trivial, as we can simply define $\epsilon(\epsilon_0, r^2) = \epsilon_0$. We therefore focus on scenarios in which we expect $\epsilon(\epsilon_0, r^2)$ to grow with r . For instance, if Ψ and Φ are lossy channels, a small difference in the loss parameters will still result in similar outputs for low-energy inputs but the difference will grow with the energy of the inputs.

Trivially, we could choose the step function, $\epsilon_{\text{step}}(\epsilon_0, r^2)$, defined by

$$\epsilon_{\text{step}}(\epsilon_0, r^2) = \begin{cases} \epsilon_0, & r \leq \tau, \\ 2, & r > \tau; \end{cases} \quad (6)$$

however, this clearly gives us no out-of-distribution information. A slightly better choice would be to interpolate between the two regions, using the data-processing inequality, which tells us that changing the input state by a distance of δ cannot increase the output distance by more than 2δ . Nonetheless, this improved form still lacks an important property: it is trivial for large r , even when $\epsilon_0 \rightarrow 0$. Hence, even if we gain perfect information about the subset of coherent states, the step function tells us nothing about large r . Lemma 1, which is a simple consequence of expressing the input state in a fixed basis (see Appendix B), tells us that at $\epsilon_0 = 0$, we should have $\epsilon(0, r^2) = 0$.

To be useful, the function $\epsilon(\epsilon_0, r^2)$ should be a decreasing function of ϵ_0 for every value of r^2 [i.e., we require that for any r and for $\epsilon'_0 < \epsilon_0$, $\epsilon(\epsilon'_0, r^2) \leq \epsilon(\epsilon_0, r^2)$]. We also require it to approach 0 as ϵ_0 does so [i.e., for any r , $\epsilon(\epsilon_0, r^2) \rightarrow 0$ as $\epsilon_0 \rightarrow 0$]. In theory, $\epsilon(\epsilon_0, r^2)$ could have an arbitrary dependence on r , since we have only required it to be a decreasing function of ϵ_0 . However, given such a function with an arbitrary r dependence, we can always construct $\epsilon'(\epsilon_0, r^2)$ as an upper concave hull of $\epsilon(\epsilon_0, r^2)$, i.e., a concave function (in terms of r^2) such that $\epsilon'(\epsilon_0, r^2) \geq \epsilon(\epsilon_0, r^2)$ for every value of r^2 . We use the concavity of $\epsilon(\epsilon_0, r^2)$ in Appendix D. Thus, we will henceforth assume that our upper bound on the output trace norm for

coherent states is an increasing and concave function of r^2 . Theorem 1 (proven in Appendix I, using various concepts from later sections) tells us that it is always possible for us to find such a function.

If we know the classes of channel to which Ψ and Φ belong (e.g., if we know the general form of Ψ but are trying to learn specific channel parameters), we can find a function $\epsilon(\epsilon_0, r^2)$ by finding the worst-case-scenario set of errors in the parameters such that we still meet the constraint $\|(\Psi - \Phi)[|re^{i\phi}\rangle\langle re^{i\phi}|_{\text{coh}}]\| \leq \epsilon_0$ for $r^2 \leq \tau^2$. In Sec. IV, we demonstrate this for three example classes of channel.

In Fig. 2, we illustrate how we can construct various bounding functions based on the degree of knowledge we have about the channels. The trivial bound holds for every channel but does not approach 0 for large $\bar{n}$, no matter how small we make ϵ_0 . If our target and learned channels are Gaussian, we can apply a tighter and always nontrivial bound (see Sec. IV A). If we know that we are looking for a phase-rotation channel (i.e., the task is to learn the rotation parameter), then the bound becomes exact. Note that the nontrivial Gaussian bound is looser than the trivial bound in some regions; this is a consequence of the way in which it has been constructed (the looseness of the Fuchs–van de Graaf inequalities and taking multiple worst-case values that cannot all be attained simultaneously). Comparing the plots for $\epsilon_0 = 0.3$ and $\epsilon_0 = 0.1$, we see that improving the in-distribution error can greatly improve the out-of-distribution error. While Fig. 2 refers to a specific scenario in which the channel we want to learn is a phase rotation, Theorem 1 tells us that a nontrivial bounding function always exists for small enough ϵ_0 .

Without knowing the specific form of $\epsilon(\epsilon_0, r^2)$, it is difficult to say anything general about its scaling with ϵ_0 and τ ; however, we can calculate the scaling for particular channel classes. For instance (as we will see in Sec. IV A), for Gaussian channels it is approximately given by $\epsilon(\epsilon_0, r^2) \sim \mathcal{O}[\sqrt{\epsilon_0} r \tau^{-1}]$. Even if we know the scaling with ϵ_0 and τ , it is not possible to say anything definitive about the scaling with the number of samples, N , since this depends on the learning method employed in stage 1 of the process detailed in Sec. II, about which we are intentionally agnostic. Hence, without specifying the learning method, we cannot say whether it is “better” (in terms of sample efficiency) to increase τ or to decrease ϵ_0 .

As an example, if we were to employ a tomography-based method (discussed in Appendix A), we might expect each measurement of an output to require between $\mathcal{O}[E\epsilon_0^{-2}]$ and $\mathcal{O}[E^2\epsilon_0^{-3}]$ copies of the state, where E is the average energy of the outputs. One method of guaranteeing the validity of the output-distance bound over the entire $r \leq \tau$ region is to take evenly distributed samples, so that nowhere in the region is more than $\mathcal{O}[\epsilon_0]$ from a sampled point. In that case, the number of different coherent-state inputs would be $\mathcal{O}[\epsilon_0^{-2}\tau^2]$, so the overall

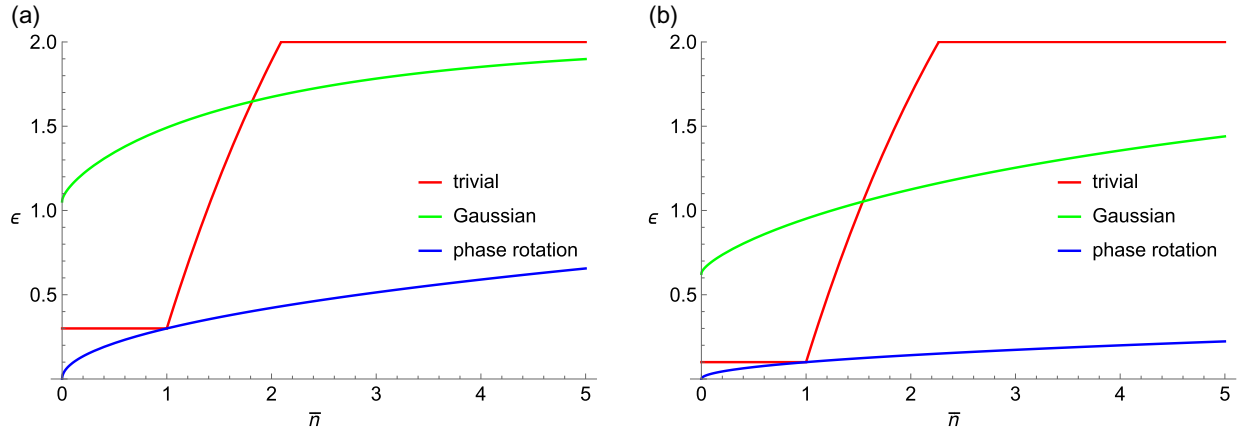


FIG. 2. The trace-norm bounds as a function of the energy of the input coherent state using a variety of techniques: (a) $\epsilon_0 = 0.3$, $\tau^2 = 1$; (b) $\epsilon_0 = 0.1$, $\tau^2 = 1$. The red lines show the trivial piecewise step-function bound, the green lines show the bound for any pair of Gaussian channels [from Eq. (11)], and the blue lines show the bound when both channels are phase rotations [from Eq. (14)].

number of samples, N , would be between $\mathcal{O}[E\epsilon_0^{-4}\tau^2]$ and $\mathcal{O}[E^2\epsilon_0^{-5}\tau^2]$. However, if we are dealing specifically with Gaussian channels (and perhaps other channel classes with a fixed known form), the structure of the problem should allow us to sample significantly fewer (different) input states and still achieve the same guarantees over the $r \leq \tau$ region, improving the sample efficiency. In Ref. [14], it has been shown that Gaussian unitaries can be learned by sampling only two distinct coherent-state inputs, so for certain channel classes, the number of different input states required could even be independent of ϵ_0 and τ , meaning that the required number of samples would depend only on the efficiency of state tomography. One subtlety is that in many physically relevant scenarios, the output energy, E , depends on the input energy, which is proportional to τ^2 , introducing a “hidden” dependence on τ . We therefore emphasize that any statement about whether it is better to increase τ or decrease ϵ_0 must be specific to one model of channel learning and one class of target channels.

IV. EXAMPLES: OUT-OF-DISTRIBUTION GENERALIZATION FOR COHERENT STATES

We want to give analytical expressions for the function $\epsilon(\epsilon_0, r^2)$ for the case in which we have some prior knowledge about the target channel. The expressions necessarily depend on the specific forms of the target and learned channels, so to give more concrete statements, we focus on cases in which both belong to a particular parametrized class of channels. We therefore have target parameters and learned parameters and the closer the two sets of parameters are, the closer the channels (and hence the channel outputs) will be. As specific examples, we will consider Gaussian channels, the cubic phase unitary, and the Kerr interaction.

A. Gaussian channels

A Gaussian channel transforms a Gaussian input with first-moment vector q and covariance matrix V according to

$$q \rightarrow Mq + d, \quad V \rightarrow MVM^T + N, \quad (7)$$

where d is a two-element real vector, M and N are 2×2 real matrices, $N = N^T \geq 0$, and $\det[N] \geq (\det[M] - 1)^2$ (we have adopted the convention that $\hbar = 2$) [3,29]. For the coherent state $|re^{i\phi}\rangle$, q has elements $2\text{Re}[re^{i\phi}] = 2r\cos(\phi)$ and $2\text{Im}[re^{i\phi}] = 2r\sin(\phi)$ and V is the identity matrix. For Gaussian channels $\mathcal{G}_1$ and $\mathcal{G}_2$, characterized by d_i , M_i , and N_i , the output fidelity is [31]

$$F^2 = \frac{2 \exp[-\frac{1}{2}\mu^T(V_1 + V_2)^{-1}\mu]}{\sqrt{\Delta + \delta} - \sqrt{\delta}}, \quad \mu = (M_2 - M_1)q + (d_2 - d_1), \quad V_i = M_i M_i^T + N_i, \quad (8)$$

where $\delta = (\det[V_1] - 1)(\det[V_2] - 1)$ and $\Delta = \det[V_1 + V_2]$. Examining the r dependence, we can see that F^2 is the exponential of a quadratic in r . Letting x , y , and z be some unknown functions of d_i , M_i , N_i , and ϕ , the output fidelity can be expressed as

$$F^2 = xe^{-(yr^2 + zr)} \leq xe^{-yr^2 + |z|r}, \quad (9)$$

where we have folded the angular dependence of the outputs into the definitions of parameters x , y , and z . This form holds for every pair of Gaussian channels, so our goal is simply to bound x , y , and z .

If we know $\|(\mathcal{G}_1 - \mathcal{G}_2)[|re^{i\phi}\rangle\langle re^{i\phi}|_{\text{coh}}]\| \leq \epsilon_0$ for all ϕ and for $r \leq \tau$, we can lower bound x and upper bound y and $|z|$. Hence, we can lower bound F^2 and so, via

the Fuchs–van de Graaf inequality [32], upper bound the output distance. Specifically,

$$\begin{aligned} x &\geq \left(\frac{2-\epsilon_0}{2}\right)^2, \quad y \leq \frac{2}{\tau^2} \log \left[\frac{2}{2-\epsilon_0} \right], \\ |z| &\leq \frac{1}{\tau} \log \left[\frac{2}{2-\epsilon_0} \right], \end{aligned} \quad (10)$$

where we obtain the bounds by assessing Eq. (9) at $r = 0$ and $r = \pm\tau$ (strictly, $r \geq 0$ but for simplicity we treat $\phi \rightarrow \phi + \pi$ as $r \rightarrow -r$) and comparing the resulting expressions to our condition. Note that the three expressions in Eq. (10) cannot all be saturated simultaneously for our condition to hold but they allow us to lower bound Eq. (9) for every value of r . We obtain

$$\begin{aligned} F^2 &\leq \left(\frac{2-\epsilon_0}{2}\right)^{2\frac{r^2}{\tau^2} + \frac{r}{\tau} + 2}, \\ \epsilon_{\text{Gaussian}}(\epsilon_0, r^2) &= 2 \sqrt{1 - \left(\frac{2-\epsilon_0}{2}\right)^{2\frac{r^2}{\tau^2} + \frac{r}{\tau} + 2}}, \\ &\geq \|(\mathcal{G}_1 - \mathcal{G}_2)[|re^{i\phi}\rangle\langle re^{i\phi}|_{\text{coh}}]\| \end{aligned} \quad (11)$$

which we immediately see obeys the property of convergence to 0 as $\epsilon_0 \rightarrow 0$. Notably, it is never trivial (greater than 2) for large r and it is concave in r^2 (though not in r). However, it can be significantly higher than ϵ_0 even in the $r < \tau$ region (and so is trivial in this region). This is the price that we pay for bounding x, y , and z individually; we could potentially get a tighter bound by working directly with Eq. (9) to construct a (possibly piecewise) bound. For small ϵ_0 , the leading-order term is $\mathcal{O}[\sqrt{\epsilon_0} r \tau^{-1}]$. Further calculation details are given in Appendix C.

If we specify more details about the forms of the Gaussian channels under consideration, we may be able to do better. For instance, for an unknown displacement, the output fidelity is independent of the input coherent state, so $\epsilon_{\text{dis}}(\epsilon_0, r^2) = \epsilon_0$. As another example, suppose that the channels are rotationally symmetric (no angular dependence), as is the case for a range of channel classes, including lossy channels and quantum-limited amplifiers. Then we have $z = 0$, so we obtain the tighter bound

$$\epsilon_{\text{sym}}(\epsilon_0, r^2) = 2 \sqrt{1 - \left(\frac{2-\epsilon_0}{2}\right)^{2\left(\frac{r^2}{\tau^2} + 1\right)}}. \quad (12)$$

In the case of many single-parameter Gaussian unitaries, it is the case that a particular value of ϵ_0 , for a given value of τ , is one to one with a specific value of the parameter difference, so that $\epsilon(\epsilon_0, r^2)$ can be tight. For example, for phase-rotation channels Ψ_{PR} and Φ_{PR} , with phases differing by θ , the output trace norm for a coherent-state input is

$$\begin{aligned} &\|(\Psi_{\text{PR}} - \Phi_{\text{PR}})[|re^{i\phi}\rangle\langle re^{i\phi}|]\| \\ &= 2\sqrt{1 - |\langle re^{i\phi}|re^{i(\phi+\theta)}\rangle|^2} = 2\sqrt{1 - e^{-2r^2(1-\cos(\theta))}}, \end{aligned} \quad (13)$$

so we can write the exact expression

$$\epsilon_{\text{PR}}(\epsilon_0, r^2) = 2 \sqrt{1 - \left(1 - \frac{1}{4}\epsilon_0^2\right)^{\frac{r^2}{\tau^2}}}. \quad (14)$$

Similarly, for single-mode squeezing,

$$\epsilon_{\text{sq}}(\epsilon_0, r^2) = 2 \sqrt{1 - \frac{1}{2\tau^2} W_0 \left[e^{2\tau^2} \tau^2 \left(2 - \frac{1}{2}\epsilon_0^2 \right) \right] \exp \left[r^2 \left(\frac{1}{\tau^2} W_0 \left[e^{2\tau^2} \tau^2 \left(2 - \frac{1}{2}\epsilon_0^2 \right) \right] - 2 \right) \right]}, \quad (15)$$

where W_0 is the Lambert W function. We note that obtaining an exact expression for the output trace norm for coherent states, based on ϵ_0 and τ , is not possible for all single-parameter unitaries. The reason why it is possible for the cases addressed here is because the expressions for the coherent-state output trace norm satisfy the condition for a useful $\epsilon(\epsilon_0, r^2)$ stated in Sec. III: namely, that they are uniformly nonincreasing in ϵ_0 . In other words, choosing closer parameter values for the target and learned channels, Ψ and Φ , and so decreasing ϵ_0 , always results in a

smaller (or equal) value of $\epsilon(\epsilon_0, r^2)$ for every r . For non-Gaussian unitaries, this is not always the case (even when learning over a single parameter). We discuss an example of a unitary without this property in Sec. IV C.

B. Cubic phase unitaries

The cubic phase gate is an important gate in CV quantum computing, because it is one option for the non-Gaussian operation required to make CV quantum

computing universal [3]. It applies the unitary $V_\gamma = e^{i\gamma\hat{q}^3}$, where q is one of the quadratures of a CV mode. To understand how it transforms a coherent state, it is easiest to work in the $\hat{q}$ basis. Specifically, we write

$$|\alpha\rangle = \frac{1}{\sqrt[4]{2\pi}} \int_{-\infty}^{\infty} \exp\left[-\frac{(q - 2\text{Re}[\alpha])^2}{4} + i(q - 2\text{Re}[\alpha])\text{Im}[\alpha]\right] |q\rangle dq, \quad (16)$$

where the $\{|q\rangle\}$ are the (unnormalizable) eigenstates of the $\hat{q}$ operator. The output fidelity between operations V_β and V_γ is then

$$|\langle\alpha|V_\beta^\dagger V_\gamma|\alpha\rangle| = \frac{1}{\sqrt{2\pi}} \left| \int_{-\infty}^{\infty} e^{i(\gamma-\beta)q^3 - \frac{1}{2}(q-2\text{Re}[\alpha])^2} dq \right|. \quad (17)$$

This integral is difficult to evaluate analytically but numerically we find that it is a decreasing function of $\Delta_\gamma = |\gamma - \beta|$ and of the magnitude of the real part of α . To understand why, consider that Eq. (17) is the convolution of the Gaussian function $(2\pi)^{-\frac{1}{2}} e^{-\frac{1}{2}q^2}$ (which integrates to 1) and the oscillatory function $e^{i\Delta_\gamma q^3}$ at the point $2\text{Re}[\alpha]$. As $|q|$ (or Δ_γ) gets larger, the oscillations become more frequent, so the integral over a small locality becomes approximately zero (the positive and negative contributions cancel out). If the Gaussian function is centered in the slowly oscillating region (i.e., if $\text{Re}[\alpha]$ is small), then the contributions will not cancel out much and the magnitude of the integral will be close to 1. For large $\text{Re}[\alpha]$, the integral will be closer to 0. As we may expect, the output fidelity is independent of $\text{Im}[\alpha]$ but since we

want to bound the output distance over all coherent states in the region, we must choose $\alpha = \pm\text{Re}[\alpha]$. The output trace norm for coherent-state inputs is plotted for various values of Δ_γ in Fig. 3.

If we are guaranteed that, for some τ , $\|(\Psi - \Phi)[|\tau e^{i\phi}\rangle\langle\tau e^{i\phi}|_{\text{coh}}]\| \leq \epsilon_0$, we can numerically upper bound Δ_γ and so we can construct a function $\epsilon(\epsilon_0, r^2)$ such that $\|(\Psi - \Phi)[|re^{i\phi}\rangle\langle re^{i\phi}|_{\text{coh}}]\| \leq \epsilon(\epsilon_0, r^2)$ for all r . Note, however, that we have made the assumption that Eq. (17) is a decreasing function of the parameter difference at every point (value of $|\alpha|$), which is suggested by the numerics, but not analytically proven. It would therefore be desirable to find either a more mathematically rigorous proof that this is the case or an analytically solvable bound on Eq. (17).

C. Kerr unitaries

The Kerr interaction is another example of a physically relevant non-Gaussian process. A Kerr gate applies the unitary $K_\kappa = e^{i\kappa\hat{n}^2}$. Using the Fock basis decomposition of coherent states, we can calculate

$$|\langle re^{i\phi}|K_\beta^\dagger K_\gamma|re^{i\phi}\rangle| = e^{-r^2} \left| \sum_{m=0}^{\infty} \frac{r^{2m}}{m!} e^{i(\gamma-\beta)m^2} \right|. \quad (18)$$

This infinite sum is difficult to evaluate in general but it can be solved, analytically or numerically, for specific values of $\kappa = |\gamma - \beta|$. In particular, in Fig. 3, we have set $\kappa = \pi/L$ for $L = 2, 4, 8, 16$, and 32 . We observe that Kerr unitaries are an interesting counterexample to the idea that, for single-parameter unitaries, reducing the parameter difference should reduce the coherent-state output distance for every value of $\bar{n}$.

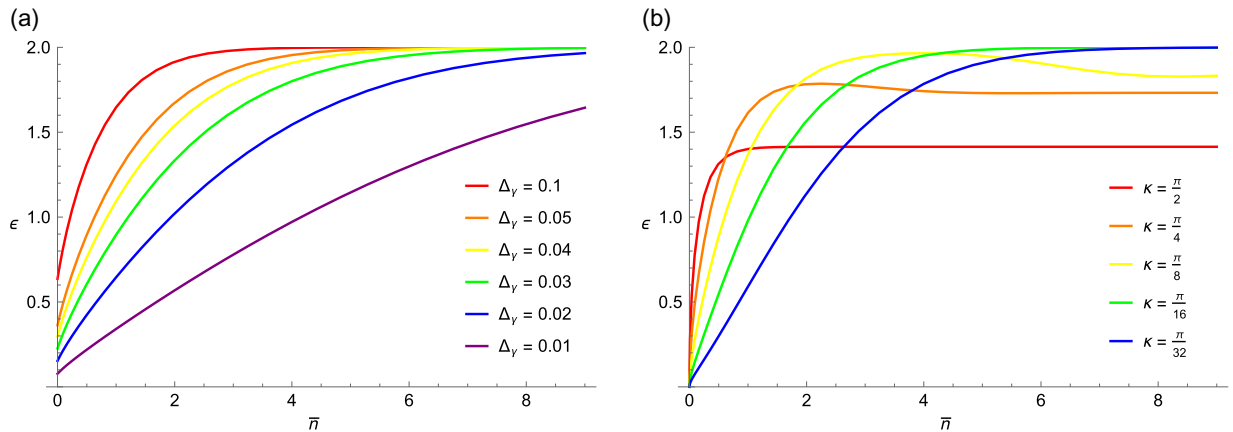


FIG. 3. The output trace norm for a coherent-state input for (a) a pair of cubic phase unitaries and (b) a pair of Kerr unitaries. The curves are plotted for various values of the parameter difference between the target and learned channels. In the case of cubic phase unitaries, the output trace norm appears to be an increasing function of the parameter difference for all values of $\bar{n}$, while for Kerr unitaries, we observe that many of the curves have inflection points and cross each other, so there are points at which a lower value of the parameter difference gives a higher output trace norm.

In fact, Fig. 3 might initially seem quite counterintuitive, since smaller values of κ result in larger asymptotic (in $\bar{n}$) values of the coherent-state output distance. This counterintuitiveness highlights the value of Theorem 1, as without it, one might suspect that there exist particularly “weirdly behaved,” but physically valid, sets of channels for which it is not always possible to construct a bounding function, $\epsilon(\epsilon_0, r^2)$, of the form that we require.

The “solution” to this counterintuitiveness can be seen by considering which nonclassical probes can perfectly discriminate the target and learned channels. Defining a general pure probe (with no idler) as $|\Omega\rangle = \sum_{m=0}^{\infty} \omega_m |m\rangle$, with $\sum_{m=0}^{\infty} |\omega_m|^2 = 1$, the output fidelity for Kerr unitaries with parameter difference κ is

$$|\langle \Omega | K_{\beta}^{\dagger} K_{\gamma} | \Omega \rangle| = \left| \sum_{m=0}^{\infty} |\omega_m|^2 e^{i\kappa m^2} \right|. \quad (19)$$

In order for this sum to be 0, we need there to exist at least two nonzero components, ω_a and ω_b , such that $e^{i\kappa a^2}$ and $e^{i\kappa b^2}$ lie in opposite quadrants of the complex plane. In particular, if we have nonzero ω_0 and want to construct a probe that discriminates the two channels perfectly, we need a nonzero ω_m for some $m \geq \sqrt{\pi\kappa^{-1}}$. Picking, for simplicity, $\kappa = \pi/L$ and restricting to cases in which L is a square number, one can see that the state $|\Omega\rangle = (|0\rangle + |\sqrt{L}\rangle)/\sqrt{2}$, with average photon number $L/2$, achieves an output fidelity of 0 (and is the lowest-energy state to do so). Thus, a smaller parameter difference (and so a larger L) means that the minimum energy single-mode probe that can perfectly discriminate the two unitaries has a higher average photon number, resolving the apparent paradox that closer channels are easier to discriminate using coherent-state probes (and further highlighting the importance of understanding out-of-distribution generalization from classical to nonclassical inputs).

V. OUT-OF-DISTRIBUTION GENERALIZATION FOR GENERAL STATES

Now assume that we have a concave function $\epsilon(\epsilon_0, r^2)$ that bounds the distance between two channel outputs for any coherent-state input. Our goal is to extend the bound to a general state input. Any state ρ can be represented as

$$\begin{aligned} \rho &= \int P(\alpha) |\alpha\rangle \langle \alpha|_{\text{coh}} d^2\alpha \\ &= \int_0^{\infty} \int_0^{2\pi} P(re^{i\phi}) |re^{i\phi}\rangle \langle re^{i\phi}|_{\text{coh}} r d\phi dr, \end{aligned} \quad (20)$$

where we use $d^2\alpha$ to emphasize that this is a double integration over both the real and the imaginary components of α . The P representation takes both positive and negative values, and is only positive semidefinite for classical states.

We begin by considering a classical state with average photon number $\bar{n}$; i.e., we have a state of the form in Eq. (20) but where the P representation, $P_{\text{class}}(re^{i\phi})$, is constrained to be positive and to obey $\int_0^{\infty} r^2 \int_0^{2\pi} P_{\text{class}}(re^{i\phi}) r d\phi dr = \bar{n}$. In Appendix D, we apply the convexity of the trace norm, the concavity of $\epsilon(\epsilon_0, r^2)$ (it is worth noting that concavity in r^2 is a looser condition than concavity in r), and Jensen’s inequality to obtain the upper bound from Corollary 2:

$$\|(\Psi - \Phi)[\rho_{\text{class}}]\| \leq \epsilon(\epsilon_0, \bar{n}).$$

A. States with finite negativity

Now consider the more general case, in which the P representation can take negative values. We can no longer immediately apply the convexity of the output trace norm. However, we can split ρ into two contributions:

$$\begin{aligned} \rho &= \rho_+ - \rho_-, \quad \rho_+ = \int_{P(\alpha) > 0} P(\alpha) |\alpha\rangle \langle \alpha|_{\text{coh}} d^2\alpha, \\ \rho_- &= \int_{P(\alpha) < 0} -P(\alpha) |\alpha\rangle \langle \alpha|_{\text{coh}} d^2\alpha, \end{aligned} \quad (21)$$

where ρ_+ and ρ_- are given by integrating over the positive and negative domains, respectively, of the P representation. They represent sub- or supernormalized classical states. Defining $\mathcal{N} = \text{Tr}[\rho_-]$ (the negativity), we define

$$\sigma_+ = \frac{1}{1 + \mathcal{N}} \rho_+, \quad \sigma_- = \frac{1}{\mathcal{N}} \rho_-. \quad (22)$$

Except for in the case of infinite negativity, these are now valid classical quantum states, as they are normalized and have a positive overlap with every coherent state. From the linearity of quantum channels and using the triangle inequality, we write

$$\begin{aligned} \|(\Psi - \Phi)[\rho]\| &= \|(1 + \mathcal{N})(\Psi - \Phi)[\sigma_+] - \mathcal{N}(\Psi - \Phi)[\sigma_-]\| \\ &\leq (1 + \mathcal{N})\|(\Psi - \Phi)[\sigma_+]\| + \mathcal{N}\|(\Psi - \Phi)[\sigma_-]\|. \end{aligned} \quad (23)$$

We can then bound the trace norm of the outputs of the classical input states σ_+ and σ_- . Note that the average photon numbers of $\sigma_{\pm}$, $\bar{n}_{\pm}$, need not equal $\bar{n}$, the average photon number of ρ ; we only require $\bar{n} = (1 + \mathcal{N})\bar{n}_+ - \mathcal{N}\bar{n}_-$. Applying Eq. (3),

$$\|(\Psi - \Phi)[\rho]\| \leq (1 + \mathcal{N})\epsilon(\epsilon_0, \bar{n}_+) + \mathcal{N}\epsilon(\epsilon_0, \bar{n}_-). \quad (24)$$

Hence, from the P representation of ρ , we can bound the distance of the learned channel output from the true channel output, using $\mathcal{N}$ (which can be used as a measure of

the nonclassicality of the state [30]) and $\bar{n}_\pm$. Applying the concavity of function ϵ ,

$$\begin{aligned} & (1 + \mathcal{N})\epsilon(\epsilon_0, \bar{n}_+) + \mathcal{N}\epsilon(\epsilon_0, \bar{n}_-) \\ &= (1 + 2\mathcal{N}) \left(\frac{1 + \mathcal{N}}{1 + 2\mathcal{N}} \epsilon(\epsilon_0, \bar{n}_+) + \frac{\mathcal{N}}{1 + 2\mathcal{N}} \epsilon(\epsilon_0, \bar{n}_-) \right) \\ &\leq (1 + 2\mathcal{N}) \epsilon \left(\epsilon_0, \frac{(1 + \mathcal{N})\bar{n}_+ + \mathcal{N}\bar{n}_-}{1 + 2\mathcal{N}} \right) = \mu \epsilon \left(\epsilon_0, \frac{\nu}{\mu} \right), \end{aligned} \quad (25)$$

where we have defined the quantities $\mu = 1 + 2\mathcal{N}$ and $\nu = (1 + \mathcal{N})\bar{n}_+ + \mathcal{N}\bar{n}_-$. Equation (24) provides a tighter bound but Eq. (25) only requires knowledge of two quantities, rather than three, and will prove useful in the next subsection. Together, Eqs. (24) and (25) give Corollary 3.

B. States with infinite negativity

In general, $\mathcal{N}$ can be infinite. Examples include Fock states and squeezed vacuum states. In this case, the bound in Eq. (24) becomes infinite and hence trivial. The intuition that we use to get around this is that every state with infinite negativity is δ -close to a state with finite (but potentially large) negativity. Suppose that we have a state ρ with $\mathcal{N}(\rho) = \infty$. Then, for any small δ , we can find another state, σ , with finite negativity, such that $\|\rho - \sigma\| \leq \delta$. By the data-processing inequality,

$$\|(\Psi - \Phi)[\rho]\| \leq \|(\Psi - \Phi)[\sigma]\| + 2\delta. \quad (26)$$

We can calculate $\|(\Psi - \Phi)[\sigma]\|$ using Eq. (25). As $\epsilon_0 \rightarrow 0$, $\|(\Psi - \Phi)[\sigma]\| \rightarrow 0$ but δ will remain fixed, so we will need to find another state with smaller δ but higher negativity.

Suppose that we have a parametrized sequence of states $\{\sigma_s\}$, for $s \geq 0$, with distance δ_s from ρ and negativity $\mathcal{N}_s$, so that δ_s is an increasing function of s that starts at 0 for $s = 0$ and $\mathcal{N}_s$ is a decreasing function of s that is finite for any $s \neq 0$. The average photon number for σ_s , $\bar{n}_s$, may differ from $\bar{n}$. Then, we can decrease s as $\epsilon_0 \rightarrow 0$, so that we eventually converge to 0. For all s ,

$$\|(\Psi - \Phi)[\rho]\| \leq \mu_s \epsilon \left(\epsilon_0, \frac{\nu_s}{\mu_s} \right) + 2\delta_s. \quad (27)$$

The remaining task is therefore to find an appropriate sequence of states close to ρ . Taking the convolution of the P representation of any state, $P(\alpha)[\rho]$, with $(1/s\pi)e^{-(1/s)|\alpha|^2}$ results in a new valid state (for $s > 0$). In fact, if we set $s = \frac{1}{2}$ or $s = 1$, we obtain the W and Q representations, respectively, of the original state. Since the Q representation is always non-negative, this suggests that

the sequence of states with P representations given by

$$P_s(\alpha)[\rho] = \rho_s = \frac{1}{s\pi} e^{-\frac{1}{s}|\alpha|^2} \star P(\alpha)[\rho] \quad (28)$$

could be useful for our purposes. For convenience, we define $\mathcal{C}_s$ as the channel enacted by this convolution (equivalent to a Gaussian additive noise channel). In Appendix G, we show that for this sequence, δ_s is upper bounded by

$$\delta_s \leq \delta_s^{(\text{UB})} = 2\sqrt{s(1 + 2\bar{n})}. \quad (29)$$

We know that for the Fock states, any nonzero value of s has a finite negativity. However, this is not true in general: for squeezed vacuums, the negativity remains infinite up to some (squeezing-dependent) threshold value, at which point it becomes 0. This threshold value is the nonclassical depth for a squeezed vacuum state.

To make our sequence of states universal, we can therefore combine an energy truncation with application of the channel $\mathcal{C}_s$. That is, we define $\sigma_{s,M} = \mathcal{C}_s[\rho^{(M)}]$, with P representation $P_{s,M}$, where $\rho^{(M)}$ is the (normalized) truncation of ρ to an M -dimensional qudit state (i.e., applying a hard energy cutoff of $M - 1$ photons and renormalizing). This adds an extra distance from the true state, $\delta_M = \|\rho - \rho^{(M)}\|$. We are free to choose any relationship between parameters s and M (so as to combine them into a single parameter) as long as $s \rightarrow 0$ when $M \rightarrow \infty$ (some later bounds will take $s \sim M^{-1}$). In Appendix G, we show that for pure states, $\delta_M = \sqrt{1 - \eta_M}$, where η_M is the probability that a photon-counting measurement on ρ would have an outcome less than M , while in general, δ_M is bounded by

$$\delta_M \leq 2\sqrt{\frac{\bar{n}}{M}}. \quad (30)$$

Finally, we can write

$$\|(\Psi - \Phi)[\rho]\| \leq \mu_{s,M} \epsilon \left(\epsilon_0, \frac{\nu_{s,M}}{\mu_{s,M}} \right) + 2\delta_s + 2\delta_M. \quad (31)$$

In Appendixes E and F, we show that the term $\mu_{s,M} \epsilon(\epsilon_0, \nu_{s,M}/\mu_{s,M})$ can be upper bounded as

$$\mu_{s,M} \epsilon \left(\epsilon_0, \frac{\nu_{s,M}}{\mu_{s,M}} \right) \leq \mu_{s,M}^{(\text{UB})} \epsilon \left(\epsilon_0, \frac{s(1-s)(M+1)}{1-2s} \right), \quad (32)$$

where $\mu_{s,M}^{(\text{UB})}$ has an analytical expression, based on the number-state decomposition of ρ , given in Eqs. (F15) and (F16). Crucially, Eq. (32) means that for an appropriate scaling of s with M ($s \lesssim M^{-1}$), we can fix the argument of

ϵ to a constant, eliminating our dependence on the specific form of function ϵ entirely.

On the other hand, if we have less information about the state (i.e., we only know the average energy and nothing about the form of the state), we can write a looser but even more general bound. In Appendix H, we show that

$$\begin{aligned} \|(\Psi - \Phi)[\rho]\| &\leq \frac{2(1-s)^M M}{s^{M-1}(1-2s)} \epsilon \left(\epsilon_0, \frac{s(1-s)(M+1)}{1-2s} \right) \\ &\quad + 4\sqrt{s(1+2\bar{n})} + 4\sqrt{\frac{\bar{n}}{M}}, \end{aligned}$$

per Theorem 2. The condition, given in Theorem 2, that $((1-s)(1-2s))/(s(M-1)) > 1$ comes from how we have bounded $\mu_{s,M}$ over all states. Per Corollary 1, we can choose particular relationships between s , M , and $\bar{n}$, to find the worst-case scaling of the error. While this bound may be extremely loose (and exponential in M , so that even a fairly low M results in an error bound much larger than ϵ_0), it shows that out-of-distribution generalization is always possible, for any input state and with almost no prior knowledge about the input state (except for $\bar{n}$).

The physical meaning of the $M = 1$ case can be a little confusing, since it means that we have truncated to a one-dimensional state. In fact, what this means is that we have set $\sigma_{0,1} = |0\rangle\langle 0|$, i.e., to the vacuum, and the $\sigma_{s,1}$, for $s > 0$, are thermal states. Equation (1) still holds but is unnecessarily loose, since the vacuum is classical ($\mathcal{N} = 0$), so we can write

$$\|(\Psi - \Phi)[\rho]\| \leq \epsilon(\epsilon_0, 0) + 4\sqrt{\bar{n}}. \quad (33)$$

We may be interested in how small we require $\epsilon(\epsilon_0, r^2)$ to be in order to obtain a nontrivial bound (i.e., a value less than 2). When using Eq. (32), this is highly state dependent; however, the parameter region in which Eq. (1) is nontrivial is an important question, both for understanding the usefulness of the most general bound and because of the physical implications. An exact characterization of this region depends on the form of the function ϵ and would require optimization over both s and M . However, one interesting case is obtained by setting $(s(1-s)(M+1))/(1-2s) = \bar{n}$ (thus fixing s as a function of M). In this case, we have a nontrivial bound if and only if

$$\begin{aligned} \epsilon(\epsilon_0, \bar{n}) &< \frac{2(\sqrt{(M+1)^2 + 4\bar{n}^2} - 2\bar{n})(M+1+2\bar{n} - \sqrt{(M+1)^2 + 4\bar{n}^2})^{M-1}}{M(M+1-2\bar{n} + \sqrt{(M+1)^2 + 4\bar{n}^2})^M} \\ &\quad \times \left(1 - 2\sqrt{\frac{\bar{n}}{M}} - \sqrt{\frac{2(2\bar{n}+1)(M+1+2\bar{n} - \sqrt{(M+1)^2 + 4\bar{n}^2})}{M+1}} \right). \end{aligned} \quad (34)$$

In Fig. 4, we plot the required value of $\epsilon(\epsilon_0, \bar{n})$ in terms of $\bar{n}$ and M . We note the caveat that this is for a particular relationship between s and M , so Eq. (1) can be made nontrivial in a broader region by optimally tuning s and M . In general, we want M to be as small as possible, since Eq. (1) is exponential in M . The plot in Fig. 4(b) uses the tighter bound for the $M = 1$ case, coming from Eq. (33). Although the $M = 1$ curve is lower than the $M = 2$, $M = 3$, and $M = 4$ curves, we should be careful not to interpret this as meaning that setting $M = 1$ is always the best choice. The curves only show the region in which the bounds are nontrivial, not the comparative values of the bounds for each parameter choice; for small values of $\epsilon(\epsilon_0, \bar{n})$, the term coming from δ_M (i.e., the error due to approximating ρ with the vacuum state) will dominate. It is also not surprising that the one-dimensional case, for which we know that $\mathcal{N} = 0$, since σ is always the vacuum, sometimes gives a tighter bound than bounding over all d -dimensional states, for $d > 1$.

Figure 4 also has an interesting interpretation in terms of channel discrimination. Equation (1) can be interpreted as a bound on the maximum quantum advantage for channel-discrimination tasks in the following sense: if we set $\bar{n} = (s(1-s)(M+1))/(1-2s)$, then we can directly compare the (upper bound on the) best single-shot discrimination probability using any finite-energy probe (proportional to the output trace distance) to the best single-shot discrimination probability using a classical probe of the same energy (proportional to $\epsilon(\epsilon_0, \bar{n})$). Since Fig. 4 depicts the values of $\epsilon(\epsilon_0, \bar{n})$ for which Theorem 2 provides a nontrivial (< 2) bound on the output trace norm, it can be interpreted as the region in which we can meaningfully bound quantum advantage in channel discrimination. We can, e.g., say that (per the dashed red line in Fig. 4) any pair of channels for which the maximum output trace norm using a classical probe of average photon number 0.297 is less than approximately 10^{-9} (corresponding to a discrimination success probability of approximately

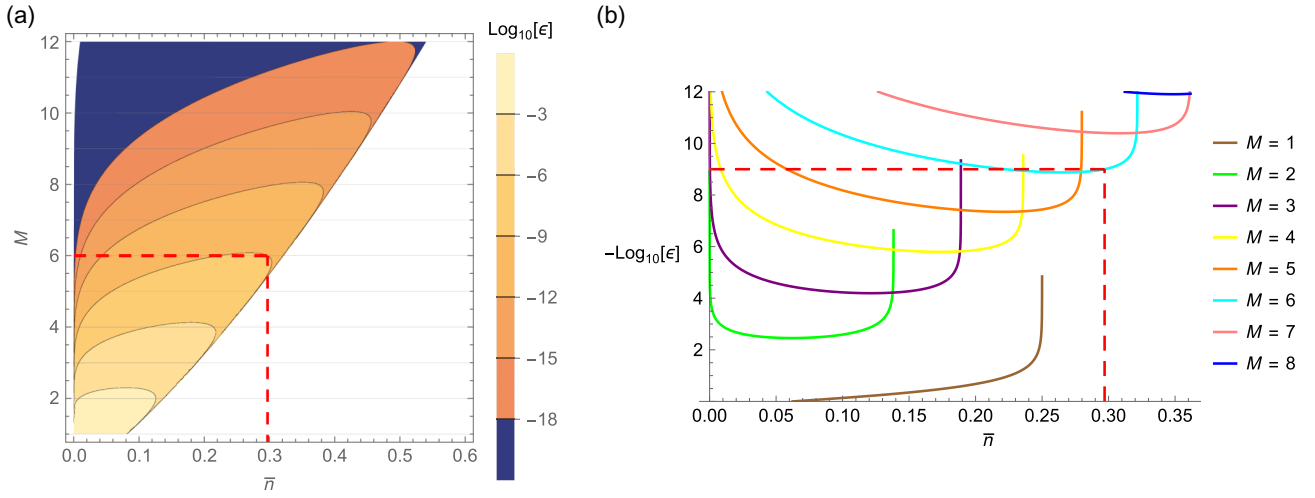


FIG. 4. The minimum value of $\epsilon(\epsilon_0, \bar{n})$ for which the bound from Theorem 2 is nontrivial, when we fix $(s(1-s)(M+1))/(1-2s) = \bar{n}$ (given on a logarithmic scale). (a) For a given value of $\bar{n}$, we can choose any integer value for M (indicated by the horizontal lines). Then, the plot can be interpreted as follows: if $\epsilon(\epsilon_0, \bar{n})$ (the bound on the output trace norm for classical inputs) is less than the value indicated by the color, then the bound will be nontrivial. The white region (in the bottom right) does not admit a nontrivial bound (for this choice of relationship between s and M), while the navy region (in the top left) has an extremely small value for the required $\epsilon(\epsilon_0, \bar{n})$. (b) This plot shows the same thing from a slightly different perspective: each curve is for a fixed value of M . As an example, by following the dashed red line, we can see that for any input with average photon number $\bar{n} \leq 0.297$, we always have a nontrivial bound if $\epsilon(\epsilon_0, \bar{n}) \leq 10^{-9}$.

$\frac{1}{2} + 2.5 \times 10^{-10}$) cannot be perfectly discriminated by using any nonclassical probe of the same energy. We revisit this interpretation in Sec. VII, while a more extensive discussion of when the bounds are nontrivial and whether they can be saturated is provided in Appendix L.

VI. EXAMPLES: OUT-OF-DISTRIBUTION GENERALIZATION FOR GENERAL STATES

We will consider three different types of nonclassical input state, to demonstrate how the various bounds can be applied: single-photon-added thermal states (SPATs), Fock states, and one-mode squeezed vacuums. SPATs have finite negativity. Fock states have infinite negativity for any nonzero energy but their negativity becomes finite with convolution alone (without the need for an energy truncation). Squeezed vacuums require both truncation and convolution to reach finite negativity.

A. Single-photon-added thermal states

SPATs have the P representation [33]

$$P_{\text{SPAT}}(re^{i\phi}) = \frac{1+q}{\pi q^3} \left(r^2 - \frac{q}{1+q} \right) e^{-\frac{r^2}{q}}, \quad (35)$$

where $q > 0$ is the average photon number of the thermal state before the photon addition (a conditional process, involving postselection). q is connected to the average photon number of the SPATs by $\bar{n} = 1 + 2q$. From Eq. (35),

we can see that the P representation has no angular dependence and starts negative for small r before becoming positive for $r^2 > q/(1+q) = (\bar{n}-1)/(\bar{n}+1)$. We can explicitly calculate $\mathcal{N}$ and $\bar{n}_{\pm}$:

$$\begin{aligned} \mathcal{N} &= e^{-\frac{1}{1+q}} \left(1 + \frac{1}{q} \right) - 1, \\ \bar{n}_{-} &= 1 + 2q - \frac{1}{1 + \left(1 - e^{\frac{1}{1+q}} \right) q}, \\ \bar{n}_{+} &= 1 + 2q - \frac{1}{1 + q}. \end{aligned} \quad (36)$$

We can now use the bound in Eq. (24) or we can simplify our calculations by instead using the bound in Eq. (25), with

$$\begin{aligned} \mu &= 2e^{-\frac{1}{1+q}} \frac{1+q}{q} - 1, \\ \frac{\nu}{\mu} &= 1 + 2q - \frac{2}{2 + 2q - e^{\frac{1}{1+q}} q} < 1 + 2q. \end{aligned} \quad (37)$$

The bound becomes

$$\begin{aligned} \|(\Psi - \Phi)[\rho_{\text{SPAT}}]\| &\leq \left(2e^{-\frac{1}{1+q}} \frac{1+q}{q} - 1 \right) \\ &\times \epsilon \left(\epsilon_0, 1 + 2q - \frac{2}{2 + 2q - e^{\frac{1}{1+q}} q} \right). \end{aligned} \quad (38)$$

Since $\bar{n}_- < \nu/\mu < \bar{n}_+$, Eq. (38) is no more than a factor of $\mu/(\mu - 1) = 2 - e^{1/(1+q)q}/(1+q)$ looser than the bound from Eq. (24).

The quantity ν/μ increases approximately linearly with $\bar{n}$ (or k). Since the function ϵ is concave, $\epsilon(\epsilon_0, \nu/\mu)$ increases sublinearly with the energy of the SPATS. On the other hand, the negativity diverges for small $\bar{n}$; this is to be expected, as the limiting case of $\bar{n} \rightarrow 1$ ($q \rightarrow 0$) is the Fock state $|1\rangle\langle 1|$. The overall bound on the closeness of the two channel outputs for an input SPATS is therefore larger for lower energies (but higher nonclassicalities). It should be noted that the bound being larger does not necessarily imply that the channel outputs are further away; the bound may simply be looser for such inputs. As briefly discussed in Appendix J, it may be possible to apply the technique from Sec. VB of using a sequence of states with lower negativities in order to tighten the bounds somewhat.

B. Fock states

The Fock state $|m\rangle\langle m|$ has the P representation

$$P_{s,\text{Fock}}(re^{i\phi}) = \frac{(-1)^m (1-s)^m}{\pi s^{m+1}} e^{-\frac{1}{s}r^2} L_m \left[\frac{r^2}{s(1-s)} \right]. \quad (39)$$

In Appendixes G and F, we show that the values of the relevant quantities are

$$\begin{aligned} \mu_s^{(\text{UB})} &= 2 \frac{(1-s)^{m+1}}{s^m(1-2s)}, \quad \nu_s^{(\text{UB})} = 4 \frac{(1-s)^{m+2}}{s^{m-1}(1-2s)^2}, \\ \delta_s &= 2 \sqrt{1 - \frac{G[m, m, 0, s]}{(1+s)^{2m+1}}}, \\ G[m, m, 0, s] &= \sum_{k=0}^m \binom{m}{k}^2 s^{2k}, \end{aligned} \quad (40)$$

where the function G is a polynomial defined in Appendix G [Eq. (G16)], so that the bound in Eq. (27) becomes

$$\begin{aligned} \|(\Psi - \Phi)[|m\rangle\langle m|]\| &\leq 2 \frac{(1-s)^{m+1}}{s^m(1-2s)} \epsilon \left(\epsilon_0, \frac{2s(1-s)}{1-2s} \right) \\ &\quad + 4 \sqrt{1 - \frac{G[m, m, 0, s]}{(1+s)^{2m+1}}}. \end{aligned} \quad (41)$$

The first term is exponential in the energy of the Fock state, since it is proportional to $((1-s)/s)^m$, while the second term is sublinear in m . The s dependence is less straightforward and depends on the form of ϵ , but the multiplicative factor quickly becomes large for small s and $m > 0$.

C. One-mode squeezed vacuums

A one-mode squeezed vacuum with squeezing parameter $r = \text{arctanh}(\lambda)$ (and average photon number $\bar{n} = \sinh(r)^2 = \lambda^2/(1-\lambda^2)$) has density matrix [3]

$$\rho_{\text{sq}} = \sqrt{1-\lambda^2} \sum_{p,q=0}^{\infty} \frac{\lambda^{p+q} \sqrt{(2p)!(2q)!}}{2^{p+q} p! q!} |2p\rangle\langle 2q|. \quad (42)$$

λ lies in the range $(-1, 1)$ but, for convenience, we assume positive values (similar results can be derived for $\lambda < 0$). Note that we should also restrict to odd M (even-dimensional truncated systems). In Appendix K, we show that

$$\delta_s = 2 \sqrt{1 - \left(1 + s^2 + 2s \frac{1+\lambda^2}{1-\lambda^2} \right)^{-\frac{1}{2}}}. \quad (43)$$

If $s > \lambda/(1+\lambda)$, then $\mathcal{N} = 0$. Otherwise, since squeezed vacuums are pure, δ_M is bounded by $\delta_M = \sqrt{1-\eta_M}$, where η_M is the probability that a photon-counting measurement on ρ_{sq} gives a result of $M-1$ or less. η_M obeys

$$\begin{aligned} \eta_M &= \sqrt{1-\lambda^2} \sum_{p=0}^{\frac{M-1}{2}} \frac{\lambda^{2p} (2p)!}{4^p (p!)^2} \\ &= 1 - \frac{\beta \left[\lambda^2; \frac{M+1}{2}, \frac{1}{2} \right] \Gamma \left[\frac{M}{2} + 1 \right]}{\sqrt{\pi} \left(\frac{M-1}{2} \right)!} \geq 1 - \frac{\lambda^2}{M(1-\lambda^2)}, \end{aligned} \quad (44)$$

where β is the incomplete beta function. The exact value is easy to calculate numerically for a specific value of M , while the bound can be used for a simpler analytical expression. We find that μ is upper bounded by

$$\begin{aligned} \mu_{s,M}^{(\text{UB})} &= \frac{2(1-s)\sqrt{1-\lambda^2}}{\eta_M(1-2s)} \\ &\quad \times \left(\frac{4 \left(y_1^{\frac{M+1}{2}} - 1 \right)}{\pi \sqrt{1+x}(y_1-1)} + \frac{y_2^{\frac{M+1}{2}} - 1}{y_2 - 1} \right), \quad (45) \\ y_1 &= \frac{\lambda(1-s)(1+x)}{s(1-2s)}, \quad y_2 = \frac{\lambda(1-s)x}{s(1-2s)}, \\ x &= \frac{(1-2s)(1-s)\lambda}{s}. \end{aligned} \quad (46)$$

The approximate scaling of the negativity with s , M , and λ (for $s \ll \lambda/(1+\lambda)$) is $\mathcal{O}[\lambda^{M-1} s^{1-M}]$.

We can then construct a piecewise bound, with

$$\begin{aligned} \|(\Psi - \Phi)[\rho_{\text{sq}}]\| &\leq \epsilon \left(\epsilon_0, \frac{\lambda^2}{1 - \lambda^2} + s \right) \\ &\quad + 4 \sqrt{1 - \left(1 + s^2 + 2s \frac{1 + \lambda^2}{1 - \lambda^2} \right)^{-\frac{1}{2}}}, \\ &\text{for } s \geq \frac{\lambda}{1 + \lambda}, \end{aligned} \quad (47)$$

and Eq. (31), substituting in Eqs. (43)–(45) otherwise. This bound is discontinuous in s , so exact optimization over s and M may be tricky; often, it may be easier to set $s = \lambda/(1 + \lambda) = \sqrt{\bar{n}(\bar{n} + 1)} - \bar{n}$ and use Eq. (47).

D. Tightness of the bounds for specific examples

By looking numerically at some specific examples of channels and input states, we can investigate whether and when the bounds can be tight. If there exist examples of channels and input states that come close to saturating the various bounds, it would suggest that those bounds cannot be greatly improved while maintaining full generality. Using the input states studied previously in this section as examples, we have performed a brief numerical investigation of this type, the full details of which are presented in Appendix L.

Using SPATS inputs, we have been able to find a class of channels, which we have dubbed parity channels, for which the upper bound, found by using Eq. (24) and the quantities from Sec. VIA, is close to the exact output trace norm. We depict this in Fig. 5, in which we plot

the output trace norm and the bounds, up to a multiplicative factor of γ (a parameter of the target and learned channels). We also plot the ratio between the true value and the bound from Eq. (24) as a function of the negativity of the corresponding SPATS. We observe that the ratio never drops below approximately 0.788, suggesting that the bound from Corollary 3 can be fairly tight.

On the other hand, using one-mode squeezed vacuum inputs and focusing on phase-rotation channels, we find that the upper bound from Sec. VIC is much larger than the exact output trace norm. This does not necessarily mean that the bound for the case of infinite negativity is loose in general; we could simply have picked a poor example for the channel class. Indeed, we might suspect that a pair of non-Gaussian channels could have a much larger output trace norm, with the same behavior for coherent states, than phase-rotation channels (which are a particularly simple class of Gaussian channels).

E. When to apply the various bounds

We have presented a variety of bounds for different kinds of input states, ρ . The more information we have about ρ , the tighter our bounds can be, as demonstrated previously in this section, where we have calculated state-specific bounds. As a rule, explicit calculation of, or state-specific bounds on, the various quantities ($\mathcal{N}$, δ_s , δ_M , etc.) results in tighter bounds, while more general bounds on the quantities give looser bounds (for more discussion of the tightness of the bounds, see Appendix L). In Table I, we present a brief summary of when and how the various bounds should be used.

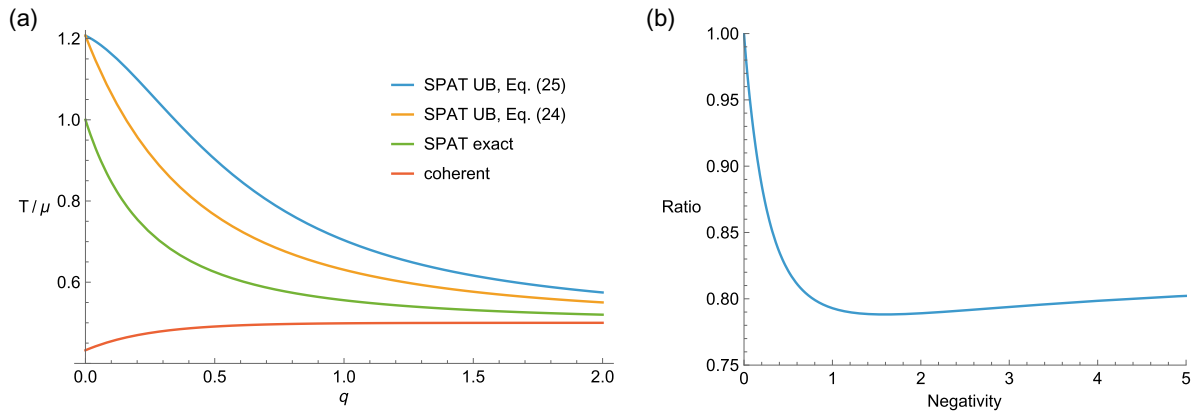


FIG. 5. (a) The output trace norm, T , up to a factor of γ , for a pair of parity channels. The green line shows the exact value for SPATSs, while the orange and blue lines show the bounds coming from Eqs. (24) and (25), respectively. The red line shows the output trace norm for a coherent state of the same energy. Note that the x axis tracks q , the average photon number of the thermal state before the photon addition, which is connected to the average photon number of the SPATS by $\bar{n} = 1 + 2q$, so that $q = 0$ corresponds to $\bar{n} = 1$. (b) A plot of the negativity of SPATSs against the ratio between the exact output trace norm and the upper bound (UB) from Eq. (24).

TABLE I. A summary of the various bounds and when each should be applied.

Type of state	Examples	$\mathcal{N}$ and $\bar{n}_\pm$ or μ and ν	δ_s	δ_M
Classical	Mixture of Gaussians	$\mathcal{N} = 0$	$\delta_s = 0$	$\delta_M = 0$
Finite negativity	SPATSS	Calculate $\mathcal{N}$ and $\bar{n}_\pm$ (analytically or numerically)	$\delta_s = 0$	$\delta_M = 0$
Infinite negativity, finite after any convolution	Fock states	Calculate analytically or use Eqs. (F15), (F16), and (F18)	Calculate analytically or use Eq. (29)	$\delta_M = 0$
Infinite negativity, still infinite after small convolutions	One-mode squeezed vacuums	Calculate analytically or use Eqs. (F15), (F16), and (F18)—for larger s , $\mathcal{N}$ may be 0	Calculate analytically or use Eq. (29)	Calculate analytically or use Eq. (30)—for larger s , we may not need to truncate
General or unknown states	Arbitrary P representation	Use Eq. (H7)	Use Eq. (29)	Use Eq. (30)

VII. QUANTUM PROCESS TOMOGRAPHY, METROLOGY, AND MACHINE-LEARNING INTERPRETATIONS OF CHANNEL LEARNING AND COMPARISON WITH EXISTING WORKS

The language and framing of our theorems is more typical of a “learning” framework than of metrology or process tomography; however, this same problem is an important task in all three fields. Our results have general applicability and so we set out here the connections between the various frameworks. However, this is not intended to be an exhaustive review of quantum machine learning, CV channel tomography, or quantum metrology.

First, let us describe the task of finding a bound of the form $\|(\Psi - \Phi)[|re^{i\phi}\rangle\langle re^{i\phi}|_{\text{coh}}]\| \leq \epsilon_0$ from a learning perspective. Φ could take the form of a physical device or process for which we can choose optimal parameters but it could also simply represent our state of knowledge about the target. For example, it could be a quantum process matrix or an equation, stored on a classical computer, that we can use to calculate the output state for a given input. The goal is to learn such a channel using a finite number, N , of sample states of the form $\Psi[|re^{i\phi}\rangle\langle re^{i\phi}|_{\text{coh}}]$ (where $r \leq \tau$). A learning algorithm might need multiple samples for each value of $re^{i\phi}$ as well as channel outputs for many different values of $re^{i\phi}$. On a quantum device, an algorithm of this type could work by using samples of both $\Psi[|re^{i\phi}\rangle\langle re^{i\phi}|_{\text{coh}}]$ and $\Phi[|re^{i\phi}\rangle\langle re^{i\phi}|_{\text{coh}}]$ (at the same points) to assess and bound $\|(\Psi - \Phi)[|re^{i\phi}\rangle\langle re^{i\phi}|_{\text{coh}}]\|$, over the region $r \leq \tau$, for a particular Φ , and then evolving Φ to a new channel Φ' , with a smaller output distance (using further samples from the target channel at each update step). This would be a form of quantum machine learning. If, instead, we only have a classical computer, we might perform tomography on all of our samples of $\Psi[|re^{i\phi}\rangle\langle re^{i\phi}|_{\text{coh}}]$ and then classically reconstruct Ψ , by finding the learned channel, Φ , that best reproduces

the measurement results amongst all channels in a certain class. These two scenarios are depicted in Figs. 1(a) and 1(b). In Appendix A, we take a brief look at the sample efficiency of possible measurements.

Working in the learning framework can provide us with useful tools for formulating guarantees on the output trace distance. If the “cost function” that we minimize is the maximum value of the output distance, evaluated over the training region, then we are immediately working in the formalism of our theorems. If the inputs are sampled randomly from the training region, we can apply in-distribution generalization techniques [34–37] to bound (with high probability) the difference between the observed and true cost functions. Alternatively, we could choose the sampled inputs according to some fixed scheme that guarantees that nowhere in the training region is far from a sampled point; bounds over the region would then follow from the data-processing inequality. In Ref. [14], it has been shown that certain cost functions based on Bell state or coherent state inputs are “faithful,” for the problem of learning unitaries, in the sense that they only reach 0 when the learned and target unitaries coincide. This is in line with our Lemma 1.

In the language of quantum process tomography, characterizing a quantum process involves learning the quantum process matrix that relates the elements of the input density matrix to the elements of the output density matrix, in some basis. However, for a completely arbitrary CV channel, this matrix has infinite elements. A number of works have therefore developed the theory of coherent-state quantum process tomography (csQPT) [4,5,38]. We will look in a little more detail at two foundational works.

The idea (as in this work) is to exploit the P representation of Fock states in order to construct the process matrix. In Ref. [4], the process matrix is constructed by characterizing the outputs for coherent-state inputs ($\Psi[|\alpha\rangle\langle\alpha|_{\text{coh}}]$)

and directly substituting them into the P representations of the Fock-state elements. The authors deal with the singularity of the P representation using the Klauder approximation, which works similarly to the convolution with a Gaussian that we apply in Sec. VB. In Ref. [5], on the other hand, the authors differentiate the Fock-basis elements of $\Psi[|\alpha\rangle\langle\alpha|_{\text{coh}}]$ with regard to (the real and imaginary parts of) α . In both cases, a Fock-basis truncation must be applied to the input.

In both cases, the goal has been to write the output of the channel as a function of its inputs. Here, instead, we assume that we have such a function and want to rigorously bound its error. Indeed, the function could have been obtained by one of these methods. Hence, although a Fock decomposition of the input we want to know about may tighten our bounds, we do not need one in order to apply our theorems. Although the method in Sec. VB also uses a Fock-basis truncation, we also show how this truncation can be taken to infinity as our learning error, ϵ_0 , approaches 0.

Both papers discuss the errors in their output states; however, they do so in the context of the additional errors introduced by their approximations or truncations. That is, they discuss the additional error introduced by using a finite Fock cutoff and (in the case of Ref. [4]) Klauder approximation, but assuming the channel is perfectly understood over the sampled input states. Hence, they do not account for the fact that any measurements of the output states will themselves have nonzero errors; here, we examine and bound how our imperfect understanding of the channel, even for the set of input states we do test, propagates as we look at inputs outside of that set. Since our work compares the input-output relations of the target and learned channels, it automatically accounts for all sources of error.

Further, in Ref. [4], interpolation is applied between the input states, to understand the behavior for states that lie in-distribution but that have not previously been sampled. While this may be physically justified in many cases, a rigorous treatment of errors should bound the worst-case scenario; our previous discussion of the learning framework gives examples of how this can be accomplished (in-distribution generalization bounds or uniform sampling).

Our work could be applied to understand the error in quantum process tomography using, e.g., the method from Ref. [5]. In this case, many of the bounds could be tightened by combining the error analysis from Ref. [5] with our own work, since the learned channel takes a specific form. This could be an interesting topic for future study.

Another physically relevant framework is to see channel learning as a problem of parameter estimation. In many cases, the channel that we want to learn about is not some completely arbitrary transformation. Rather, we may know that it belongs to some parametrized class of channels and

want to learn the (potentially many) parameters, which could represent some important information about a physical process. Examples could include the loss of a medium or the strength of some particular nonlinear interaction. Quantum metrology then involves probing the channels in order to learn the parameter values. Errors in parameter estimation of $\mathcal{O}[N^{-\frac{1}{2}}]$ can be achieved using coherent states and better error scalings are possible using clever strategies.

For the important class of Gaussian channels, protocols have been proposed that use csQPT to perform parameter estimation [39,40]. One-mode Gaussian channels are fully characterized by a limited number of parameters and it is possible to learn all of them to within a small uncertainty.

The difficulty, from our perspective, is understanding how errors in the parameters translate to errors in output distance for arbitrary inputs. A small change in displacement is likely to have a drastically different effect on an arbitrary state than a similar-magnitude change in a squeezing parameter. This is especially true if we consider non-Gaussian effects. Thus, a condition on the variance of whichever parameter we want to learn (often the goal in parameter estimation) does not automatically translate to a useful bound on the output distance of our target and learned channels. In Sec. IV, we illustrate via examples how knowledge about the output distance can be converted to bounds on the parameter difference and used to construct $\epsilon(\epsilon_0, r^2)$. This works both ways: for many channel classes, we may be able to convert a bound on the difference in parameters into a bound on the coherent-state output distance and then apply our bounds for general states. This is the case regardless of how we have estimated the parameters (i.e., even if we have learned them using a different limited set of inputs to the low-energy coherent states, such as a small number of noncoherent Gaussian probes [41,42]), so our results are not limited to a specific model of quantum metrology.

We can connect parameter estimation to the idea of quantum machine learning. If we have an optical circuit that we want to imitate our target channel as closely as possible, we would do this by tuning the parameters of our circuit so that the outputs are as close to the target outputs as possible. We can view this as estimating the optimal parameters.

The precise form of the bound for low-energy coherent-state inputs that we use as our basic ingredient may not be the format in which the error of a sensing protocol is presented elsewhere. However, we have chosen this form for its generality, as any similar bound on the output distance can be rewritten in this way.

Finally, we note that our results are also of interest from a purely channel-discrimination perspective (i.e., separately from both learning and quantum metrology). The output trace distance, $\|(\Psi - \Phi)[\rho]\|$, corresponds to the

optimal probability of successfully deciding whether an unknown channel is Ψ or Φ by inputting ρ and measuring the output. We are often interested in the success probability optimized over both probes and measurements, but in the CV setting, we must generally impose an energy constraint on the input state. Our results can therefore be interpreted as bounds on the probability of successfully discriminating between two channels, given any probe of a certain energy. One difference between this setting and ours is that, typically, in channel-discrimination tasks, we know the two channels between which we want to choose. Here, we give bounds for all pairs of channels constrained only by a certain maximum success probability with low-energy coherent-state probes.

As discussed at the end of Sec. VB, our results can therefore be used to bound quantum advantage in channel-discrimination tasks. Per Fig. 4, if the maximum discrimination probability using a coherent-state probe is close enough to $\frac{1}{2}$, then no nonclassical probe is able to discriminate with a success probability of 1 (we could draw a similar region for a different maximum success probability for nonclassical probes, e.g., 0.9). This is not a particularly strong statement, as it may be intuitive that there exists some limit to the benefit of nonclassicality, but it is nontrivial that such a limit should exist for every possible pair of channels that are very close for classical inputs, even if we pick extremely non-Gaussian channels and input states that are tailored to show as much quantum advantage as possible.

Further, as a counterpoint to the intuition that there should exist a limit to the benefit of nonclassicality, we note that when using nonclassical states for quantum metrology and using the quantum Fisher information (QFI) as our metric, no such limit exists. At any finite energy, there exist “infinite-precision” probe states that result in an output with infinite QFI, even for a Gaussian- (phase-rotation) encoding channel (strictly speaking, infinite QFI is obtained as the limit of a sequence of physically achievable fixed-energy probes with finite and increasing QFI) [43]. As such, for metrology, there is no upper bound on the benefit of nonclassicality (at least in terms of QFI), so it is nontrivial that such a bound should exist for channel discrimination. We also note that the topic of ultimate bounds for channel discrimination has largely been broached in the more limited cases of DV or teleportation-covariant channels, rather than the extremely general case of arbitrary CV channels [44,45].

In many channel-discrimination settings, we are also allowed an idler system, i.e., a system that does not pass through the unknown channel but that may be entangled with the probe. In such contexts, we are often interested in the diamond norm, which is the output trace norm maximized over all probes, including those entangled with idler systems. In CV systems, we generally need to apply an energy constraint, arriving at the energy-constrained

diamond norm [46–48]. If our results were extended to multimode channels, it would be possible to also bound the energy-constrained diamond norm in terms of the output distance for classical states. This would provide a fully general bound on the quantum advantage of every channel-discrimination task and would be an interesting extension of this work.

VIII. CONCLUSIONS

When learning the action of a quantum channel, it is not possible to test every possible input state. Instead, we can aim to find a learned channel that reproduces the input-output relations for simple low-energy classical probes. We want to be sure that this learned channel also reproduces the input-output relations for higher-energy and nonclassical inputs and to understand how an error in learning the channel over low-energy coherent states propagates when considering inputs outside of the sampled distribution.

In this paper, we have shown that a bound on the distance between the outputs of two channels for low-energy coherent-state inputs can always give rise to a bound on the output distance for higher-energy coherent-state inputs, as long as the initial bound is tight enough. Such a bound, in turn, gives rise to a bound on the output distance for all input states. This bound converges to 0 as the output distance for coherent-state inputs converges to 0. This is a useful result—with implications for subjects such as quantum machine learning and quantum metrology, amongst others—as it shows that it is sufficient to probe a process with experimentally simple classical probes. We have given general statements that hold for all classes of channel and for all input states, as well as looking at specific useful examples of channel classes and types of input state.

One avenue for future research is to extend this result from one-mode channels to multimode channels. This is likely to be a simple extension, since all multimode CV states also have a P representation (i.e., a quasiprobability distribution over multimode coherent states); however, it would require recalculating the distances δ_s and δ_M and reformulating the bound on $\mu_{s,M}$ for this more general setting. The energy truncation would also need to be handled with care, since there are different ways of applying an energy truncation in the multimode case (local or global). This research direction would be interesting, as it would allow one to formulate bounds that account for idler modes and hence bound the diamond norm as well as the output trace norm.

Another option would be to look at other specific classes of channels or input states. We have looked at some examples in this work but bounds for specific physically relevant channels or states could allow our results to be applied more broadly.

Many of the existing bounds could also be tightened. From numerically studying some examples (see

Appendix L), it appears that at least Eqs. (1) and (45) can be significantly tightened. It is also possible that there is a more optimal sequence of states that we could use in the infinite-negativity case (rather than $\sigma_{s,M} = C_s[\rho^{(M)}]$).

Finally, it would be useful to bound the worst-case scaling of the function ϵ with r^2 ; in this work, we have only demonstrated that there exists an ϵ that approaches 0 as ϵ_0 approaches 0.

ACKNOWLEDGMENTS

We acknowledge financial support from: CLUSTEC, which has received funding from the European Union (EU) Horizon Europe program, under Grant Agreement No. 101080173 (J.L.P.); the Next Generation Internet (NGI) Enrichers fellowship program, which has received funding from the EU Horizon Europe research and innovation program, under Grant Agreement No. 101070125 (J.L.P.); the U.S. Department of Energy, Office of Science, National Quantum Information Science Research Centers, Superconducting Quantum Materials and Systems Center (SQMS), under Contract No. DE-AC02-07CH11359 (J.L.P. and Q.Z.); and the EU Horizon Europe research and innovation program, under EPIQUE Project Grant Agreement No. 101135288 (L.B.). Q.Z. also acknowledges support from the National Science Foundation (NSF) (Grants No. CCF-2240641, No. OMA-2326746, and No. 2350153), the Office of Naval Research (ONR) (Grant No. N00014-23-1-2296), the Air Force Office of Scientific Research (AFOSR) Multidisciplinary Research Program of the University Research Initiative (MURI) Grant No. FA9550-24-1-0349, and the Defense Advanced Research Projects Agency (DARPA) [Measurement-Based Quantum Information and Transduction (MeasQUIT) Grants No. HR0011-24-9-0362, No. HR00112490453, and No. D24AC00153-02]. Part of the work was carried out during J.L.P.'s research visit to the University of Southern California. J.L.P. also thanks Matteo Paris and Sean Moore for helpful discussions.

DATA AVAILABILITY

No data were created or analyzed in this study.

APPENDIX A: MEASUREMENTS OF THE OUTPUTS

Let us consider how we can assess $\|(\Psi - \Phi)[|re^{i\phi}\rangle\langle re^{i\phi}|_{\text{coh}}]\|$ for a particular value of $re^{i\phi}$. One option is to carry out tomography on our copies of $\Psi[|re^{i\phi}\rangle\langle re^{i\phi}|_{\text{coh}}]$ (up to some trace-norm error η) and then assess the output distance classically. Another option, if we are carrying out the learning algorithm using a quantum device capable of producing copies of $\Phi[|re^{i\phi}\rangle\langle re^{i\phi}|_{\text{coh}}]$, is to directly measure $\|(\Psi - \Phi)[|re^{i\phi}\rangle\langle re^{i\phi}|_{\text{coh}}]\|$, without explicitly obtaining classical descriptions of the output states.

There exist a variety of methods for performing tomography on CV states, some of which are specific to certain types of state. Some more general methods involve the reconstruction of the Husimi Q function or the Wigner function by applying homodyne, heterodyne, or other measurements, after displacing, squeezing, or otherwise operating on the state [49–51]. If we know that the output states have a property called reflection symmetry, we can estimate their characteristic function at several points [52]. If we know that the outputs are Gaussian states, we need only accurately learn their first and second moments in order to fully characterize them. In Ref. [53], it is shown that we need only learn these moments up to $\mathcal{O}[\eta^2]$ and explicit bounds on the trace distance between Gaussian states are given, based on the difference in first and second moments and the average energies of the states.

The downside of tomography-based methods is that they can be extremely inefficient for CV states. Even for pure states, one requires $S = \mathcal{O}[E\eta^{-2}]$ copies for tomography of a non-Gaussian output and at least $S = \mathcal{O}[E^2\eta^{-2}]$ copies (at most $S = \mathcal{O}[E^2\eta^{-3}]$ copies) are required for mixed states [53]. Here, E is the average energy of the output that we want to characterize. Crucially, this is not necessarily the same as r^2 , the energy of the input coherent state and could, depending on the target channel, be much higher (and also may not be known *a priori*). Note that the error in estimation, η , must be added to the assessed value of $\|(\Psi - \Phi)[|re^{i\phi}\rangle\langle re^{i\phi}|_{\text{coh}}]\|$, when formulating the bound.

For pure output states, a direct assessment of the output distance can be accomplished using the SWAP test, which is a two-outcome measurement on a pair of states, ψ_1 and ψ_2 , that outputs 0 with probability $\frac{1}{2} + \frac{1}{2}\text{Tr}[\psi_1\psi_2]$. If the outputs, $\Psi[|re^{i\phi}\rangle\langle re^{i\phi}|_{\text{coh}}]$ and $\Phi[|re^{i\phi}\rangle\langle re^{i\phi}|_{\text{coh}}]$, are pure, this immediately tells us their squared fidelity, up to an error of $\mathcal{O}[\sqrt{S^{-1}}]$. Applying the Fuchs–van de Graaf relation, we obtain $S = \mathcal{O}[\eta^2]$. Though this error scaling may seem similar to the tomography expressions, a crucial difference is that it is entirely independent of the energy of the output state.

The SWAP test can be enacted using a controlled beam splitter [54,55]. However, this is a non-Gaussian component and so may be difficult to implement. In Ref. [56], it has been shown that the SWAP test can be implemented in the CV setting using only linear optics and photon counting. Thus, for unitary Ψ and Φ , we can assess the output distance using just a single beam splitter and photon counters (although the finite threshold of any real photon counter will introduce errors).

For mixed-state outputs, the SWAP test only lets us learn the overlap of $\Psi[|re^{i\phi}\rangle\langle re^{i\phi}|_{\text{coh}}]$ and $\Phi[|re^{i\phi}\rangle\langle re^{i\phi}|_{\text{coh}}]$. This would not give a bound on the output distance, which converges to 0 when the target and learned channels coincide. One alternative is to use the implementation of the Helstrom measurement from Ref. [57]. This method is

based on the state-exponentiation algorithm [58] and phase estimation and it uses copies of (unknown) states ψ_1 and ψ_2 to apply the optimal discriminator between ψ_1 and ψ_2 to an arbitrary state. In Ref. [59], it has been shown that enacting the Helstrom measurement on a single state with a failure probability of ζ requires $S = \mathcal{O}[-\log(\zeta)\zeta^{-3}]$ copies of each state. Since the optimal measurement should correctly discriminate between the states with a probability linear in the trace distance (in fact, this is precisely its operational meaning), we could estimate the output distance by applying the Helstrom measurement to multiple copies of the output states and finding the error probability. To estimate it to error probability $\mathcal{O}[\zeta]$ (so that the two contributors to the error are equal), we would carry out the protocol approximately ζ^{-2} times and so the total number of copies required is $S = \mathcal{O}[-\log(\zeta)\zeta^{-5}]$. At first glance, this is quite costly but the complete lack of dependence on the output states themselves (including the lack of dependence on E) could make this algorithm more efficient than tomography, in some circumstances. In terms of implementation, however, this would require many uses of a controlled beam splitter and so would not be practical for near-term applications.

APPENDIX B: OUT-OF-DISTRIBUTION GENERALIZATION FOR COHERENT STATES WHEN $\epsilon_0 = 0$

We want to show that if two channels (i.e., the target channel and the learned channel) have exactly the same output for low-energy coherent-state inputs, then they must also have the same output for any coherent-state input. More precisely, suppose that $\|(\Psi - \Phi)[|re^{i\phi}\rangle\langle re^{i\phi}|_{\text{coh}}]\| = 0$, for $r \leq \tau$. Then, this holds for $r > \tau$ too. While intuitive, this is worth proving, since it guarantees that if we learn a channel exactly over a small area of the phase space, we have also learned it over all of phase space.

If $\|(\Psi - \Phi)[|re^{i\phi}\rangle\langle re^{i\phi}|_{\text{coh}}]\| = 0$ for $r \leq \tau$, then

$$e^{-r^2} \left\| \sum_{m,n=0}^{\infty} \frac{r^{m+n} e^{i(m-n)\phi}}{\sqrt{m!n!}} (\Psi - \Phi)[|m\rangle\langle n|] \right\| = 0 \quad \forall r \leq \tau. \quad (\text{B1})$$

Choosing any (infinite-dimensional) basis, every element of $\sum_{m,n=0}^{\infty} r^{m+n} e^{i(m-n)\phi} / \sqrt{m!n!} (\Psi - \Phi)[|m\rangle\langle n|]$ must equal 0, for any $r \leq \tau$. In particular, we choose the Fock basis. Then, the condition can only be fulfilled if every element of $(\Psi - \Phi)[|m\rangle\langle n|]$ is identically 0, so $\|(\Psi - \Phi)[|re^{i\phi}\rangle\langle re^{i\phi}|_{\text{coh}}]\| = 0$ for all r (including $r > \tau$). This is also in line with what we expect from Ref. [5], since it is shown there that the quantum process matrix of a channel, Ψ , is completely determined by all of the (infinite) derivatives of $\|\Psi[|re^{i\phi}\rangle\langle re^{i\phi}|_{\text{coh}}]\|$ assessed at the origin. Since they must all be the same for both Ψ and Φ over a finite

region that includes the origin, the processes are identical everywhere.

We revisit the case of $\epsilon_0 > 0$ in Appendix I.

APPENDIX C: GAUSSIAN CHANNELS

The output fidelity for Gaussian channels takes the form given in Eq. (9). By bounding x , y , and z individually, we can bound the output fidelity for any r . We have a constraint on the output distance for $r \leq \tau$, which translates into a lower bound on the fidelity via a Fuchs–van de Graaf inequality. Specifically, for all $r \leq \tau$ (and replacing F with F^2 for pure states),

$$F[\mathcal{G}_1[|re^{i\phi}\rangle\langle re^{i\phi}|_{\text{coh}}], \mathcal{G}_2[|re^{i\phi}\rangle\langle re^{i\phi}|_{\text{coh}}]] \geq \frac{2 - \epsilon_0}{2}. \quad (\text{C1})$$

At $r = 0$, the output fidelity is x , so the lower bound on x follows directly. Now consider $r = \pm\tau$, for which we have

$$xe^{-y\tau^2 \mp z\tau} \geq \left(\frac{2 - \epsilon_0}{2}\right)^2. \quad (\text{C2})$$

If we want to maximize y and $|z|$, we should make x as large as possible, i.e., 1. To bound y , we set $z = 0$ and obtain

$$e^{-y\tau^2} \geq \left(\frac{2 - \epsilon_0}{2}\right)^2, \quad (\text{C3})$$

and so recover our bound on y . Finally, we bound $|z|$ by comparing F^2 at $r = 0$ and at $r = \pm\tau$. Concavity can be confirmed by differentiating Eq. (11) twice with regard to r^2 .

Applying single-mode squeezing to $|\alpha\rangle_{\text{coh}}$, the first and second moments become

$$V = \begin{pmatrix} e^{2s} & 0 \\ 0 & e^{-2s} \end{pmatrix}, \quad x = \begin{pmatrix} 2e^s \alpha_R \\ 2e^{-s} \alpha_I \end{pmatrix}, \quad (\text{C4})$$

so the output fidelity between a target and learned squeezing unitary is $F_{\text{sq}}^2 = \text{sech}(s_T - s_L) \exp[2|\alpha|^2(\text{sech}(s_T - s_L) - 1)]$. If we are again given $\|(\Psi - \Phi)[|\tau e^{i\phi}\rangle\langle \tau e^{i\phi}|_{\text{coh}}]\| \leq \epsilon_0$, for some τ , we obtain the condition $\text{sech}(s_T - s_L) \leq W_0[e^{2\tau^2}\tau^2(2 - \frac{1}{2}\epsilon_0^2)]/(2\tau^2)$, where W_0 is the Lambert W function. The right-hand side goes to 1 as $\epsilon_0 \rightarrow 0$, as expected. We therefore recover Eq. (15).

APPENDIX D: OUT-OF-DISTRIBUTION GENERALIZATION FOR CLASSICAL STATES

From the convexity of the trace norm, we can write

$$\|(\Psi - \Phi)[\rho_{\text{class}}]\| \leq \int_0^\infty \epsilon(\epsilon_0, r^2) \int_0^{2\pi} P_{\text{class}}(re^{i\phi}) r d\phi dr = \int_0^\infty \epsilon(\epsilon_0, r^2) p_{\text{class}}(r) dr, \quad (\text{D1})$$

where we have defined $p(r) = \int_0^{2\pi} P(re^{i\phi}) r d\phi$. Using this definition, $\int_0^\infty p(r) dr = 1$ for any normalized P representation and our energy constraint takes the form $\int_0^\infty r^2 p(r) dr \leq \bar{n}$. We now apply Jensen's inequality for concave functions. The right-hand side of Eq. (D1) can be expressed as the expectation value of $\epsilon(\epsilon_0, r^2)$ over the classical probability density function $p_{\text{class}}(r)$. For any valid probability density function

$$\mathbb{E}_{p_{\text{class}}}[\epsilon(\epsilon_0, r^2)] \leq \epsilon(\epsilon_0, \mathbb{E}_{p_{\text{class}}}[r^2]) \leq \epsilon(\epsilon_0, \mathbb{E}_{p_{\text{class}}}[\bar{n}]). \quad (\text{D2})$$

Note that since $\bar{n}$ is the expectation value of r^2 and not of r , we only require concavity in r^2 .

APPENDIX E: SUFFICIENCY OF UPPER BOUNDING μ

Recall that the quantity in Eq. (31) that we want to bound is $\mu\epsilon(v/\mu)$, where $\epsilon(x)$ is concave in x and where we have temporarily dropped the dependence on ϵ_0 . Recall that μ and v are defined by

$$\mu = 1 + 2\mathcal{N}, \quad v = (1 + \mathcal{N})\bar{n}_+ + \mathcal{N}\bar{n}_-. \quad (\text{E1})$$

The negativity of a state ρ is defined as the negative volume of the P representation, i.e.,

$$\mathcal{N} = - \int_{P(\alpha)[\rho] < 0} P(\alpha)[\rho] d^2\alpha, \quad (\text{E2})$$

and the $\bar{n}_\pm$ are defined by

$$\bar{n}_+ = \frac{1}{1 + \mathcal{N}} \int_{P(\alpha) > 0} P(\alpha) |\alpha|^2 d^2\alpha,$$

$$\bar{n}_- = -\frac{1}{\mathcal{N}} \int_{P(\alpha) < 0} P(\alpha) |\alpha|^2 d^2\alpha. \quad (\text{E3})$$

Hence, μ and v can be calculated as

$$\mu = \int |P(\alpha)| d^2\alpha, \quad v = \int |P(\alpha)| |\alpha|^2 d^2\alpha, \quad (\text{E4})$$

where we integrate over both the negative and the positive domains of the P distribution but take the absolute value.

At first glance, upper bounding $\mu\epsilon(v/\mu)$ requires us to find both an upper and a lower bound on μ (and an upper bound on v), since it is used both as a multiplicative factor and as the denominator, inside the function ϵ . However, by applying the concavity, we can show that it suffices to find an upper bound on μ .

Taking the partial derivative of $\mu\epsilon(v/\mu)$ with regard to μ , we obtain

$$\frac{\partial}{\partial \mu} \left[\mu \epsilon \left(\frac{v}{\mu} \right) \right] = \epsilon \left(\frac{v}{\mu} \right) - \frac{v}{\mu} \epsilon' \left(\frac{v}{\mu} \right), \quad (\text{E5})$$

where ϵ' is the partial derivative of ϵ with regard to μ . Recalling that $\epsilon'(v/\mu)$ is the gradient of ϵ at v/μ and applying the concavity of ϵ , we see that $(\partial/\partial \mu) [\mu\epsilon(v/\mu)] \geq 0$ for any (valid) value of μ . Thus, an overestimate of μ always results in an overestimate of $\mu\epsilon(v/\mu)$ and so it suffices to use an upper bound of μ , both inside and outside ϵ , when bounding $\mu\epsilon(v/\mu)$.

APPENDIX F: BOUNDING THE OUTPUT DISTANCE FOR $\sigma_{s,M}$

The P representation, $P_{s,M}$, of $\sigma_{s,M}$ (the state obtained by first truncating ρ to a maximum photon number of $M - 1$ and then convolving the resulting P representation with a Gaussian function of width s^{-1}), can be written as

$$P_{s,M}(\alpha)[\rho] = P_{s,M}(re^{i\phi})[\rho] = \sum_{m,n=0}^{M-1} \langle m | \rho^{(M)} | n \rangle P_s(re^{i\phi})[|m\rangle\langle n|], \quad (\text{F1})$$

where $|m\rangle$ and $|n\rangle$ are Fock states. We specify $\langle m | \rho^{(M)} | n \rangle$ rather than $\langle m | \rho | n \rangle$ to emphasize that $\rho^{(M)}$ is normalized. The contribution to $P_s(re^{i\phi})$ of the terms $|m\rangle\langle n|$ and $|n\rangle\langle m|$, where $m > n$, is [60, Eq. (2.15)]

$$\begin{aligned} P_s(re^{i\phi}) \left[\frac{1}{2} (e^{i\theta} |m\rangle\langle n| + e^{-i\theta} |n\rangle\langle m|) \right] &= P_{s,\theta}^{(m,n)}(re^{i\phi}) \\ &= \cos(\theta - (m - n)\phi) \frac{(-1)^n}{\pi} \sqrt{\frac{n!}{m!}} \frac{(1 - s)^n}{s^{m+1}} e^{-\frac{1}{s} r^2} r^{m-n} L_n^{m-n} \left[\frac{r^2}{s(1-s)} \right], \end{aligned} \quad (\text{F2})$$

where L_x^y is a generalized Laguerre polynomial. We include θ for completeness but we will find it has no effect on our calculations and so we will generally drop it from the subscript. For Fock states, $|m\rangle\langle m|$, we get Eq. (39). Equation (39) has no angular dependence, while the angular component of Eq. (F2) is a cosine function of period $2\pi/(m-n)$. We will upper bound μ and ν by finding and summing the contributions from each term $P_s^{(m,n)}$; this is just an upper bound and is not tight unless the different contributions to the total P representations never cancel each other out (which is not generally the case). Noting that the P representation contributions of the on- and off-diagonal elements of $\rho^{(M)}$ have different properties, it is helpful to separate their contributions to μ and ν .

Starting with the off-diagonals and using Eq. (F2), the exact contributions to μ and ν are

$$\begin{aligned} \mu[P_s^{(m,n)}] &= \frac{1}{\pi} \frac{(1-s)^n}{s^{m+1}} \sqrt{\frac{n!}{m!}} \int_0^{2\pi} |\cos(\theta - (m-n)\phi)| d\phi \\ &\quad \times \int_0^\infty e^{-\frac{1}{s}r^2} r^{m-n+1} \left| L_n^{m-n} \left[\frac{r^2}{s(1-s)} \right] \right| dr, \end{aligned} \quad (\text{F3})$$

$$\begin{aligned} \nu[P_s^{(m,n)}] &= \frac{1}{\pi} \frac{(1-s)^n}{s^{m+1}} \sqrt{\frac{n!}{m!}} \int_0^{2\pi} |\cos(\theta - (m-n)\phi)| d\phi \\ &\quad \times \int_0^\infty e^{-\frac{1}{s}r^2} r^{m-n+3} \left| L_n^{m-n} \left[\frac{r^2}{s(1-s)} \right] \right| dr. \end{aligned} \quad (\text{F4})$$

The integral over ϕ is always equal to 4. Then, applying a bound (by Szegő) on the magnitude of Laguerre polynomials,

$$\begin{aligned} \mu[P_s^{(m,n)}] &\leq \frac{4}{\pi} \frac{(1-s)^n}{s^{m+1}} \frac{(m-n+1)_n}{\sqrt{m!n!}} \\ &\quad \times \int_0^\infty r^{m-n+1} e^{\frac{r^2}{s} \left(\frac{1}{2(1-s)} - 1 \right)} dr, \end{aligned} \quad (\text{F5})$$

$$\begin{aligned} \nu[P_s^{(m,n)}] &\leq \frac{4}{\pi} \frac{(1-s)^n}{s^{m+1}} \frac{(m-n+1)_n}{\sqrt{m!n!}} \\ &\quad \times \int_0^\infty r^{m-n+3} e^{\frac{r^2}{s} \left(\frac{1}{2(1-s)} - 1 \right)} dr, \end{aligned} \quad (\text{F6})$$

where $(x)_n$ is the Pochhammer symbol. These integrals converge for $s < \frac{1}{2}$. Evaluating them [61], we obtain

$$\begin{aligned} \mu[P_s^{(m,n)}] &\leq \frac{2}{\pi} \frac{(1-s)^n}{s^{m+1}} \frac{(m-n+1)_n}{\sqrt{m!n!}} \left(\frac{2s(1-s)}{1-2s} \right)^{1+\frac{m-n}{2}} \\ &\quad \times \Gamma \left[1 + \frac{m-n}{2} \right], \end{aligned} \quad (\text{F7})$$

$$\begin{aligned} \nu[P_s^{(m,n)}] &\leq \frac{2}{\pi} \frac{(1-s)^n}{s^{m+1}} \frac{(m-n+1)_n}{\sqrt{m!n!}} \left(\frac{2s(1-s)}{1-2s} \right)^{2+\frac{m-n}{2}} \\ &\quad \times \Gamma \left[2 + \frac{m-n}{2} \right]. \end{aligned} \quad (\text{F8})$$

After some simplification, we end up with the inequalities

$$\begin{aligned} \mu[P_s^{(m,n)}] &\leq \frac{2^{2+\frac{m-n}{2}} (1-s)^{1+\frac{m+n}{2}} (m-n+1)_n}{\pi s^{\frac{m+n}{2}} (1-2s)^{1+\frac{m-n}{2}}} \frac{1}{\sqrt{m!n!}} \\ &\quad \times \Gamma \left[1 + \frac{m-n}{2} \right], \end{aligned} \quad (\text{F9})$$

$$\begin{aligned} \nu[P_s^{(m,n)}] &\leq \frac{2^{3+\frac{m-n}{2}} (1-s)^{2+\frac{m+n}{2}} (m-n+1)_n}{\pi s^{\frac{m+n}{2}-1} (1-2s)^{2+\frac{m-n}{2}}} \frac{1}{\sqrt{m!n!}} \\ &\quad \times \Gamma \left[2 + \frac{m-n}{2} \right]. \end{aligned} \quad (\text{F10})$$

We now use Eq. (39) to find the contributions from the on-diagonals:

$$\begin{aligned} \mu[P_s^{(m,m)}] &= \frac{1}{\pi} \frac{(1-s)^m}{s^{m+1}} \int_0^{2\pi} 1 d\phi \\ &\quad \times \int_0^\infty e^{-\frac{1}{s}r^2} r \left| L_m \left[\frac{r^2}{s(1-s)} \right] \right| dr, \end{aligned} \quad (\text{F11})$$

$$\begin{aligned} \nu[P_s^{(m,m)}] &= \frac{1}{\pi} \frac{(1-s)^m}{s^{m+1}} \int_0^{2\pi} 1 d\phi \\ &\quad \times \int_0^\infty e^{-\frac{1}{s}r^2} r^3 \left| L_m \left[\frac{r^2}{s(1-s)} \right] \right| dr. \end{aligned} \quad (\text{F12})$$

Again using the bound on the magnitude of the Laguerre polynomials,

$$\begin{aligned} \mu[P_s^{(m,m)}] &\leq 2 \frac{(1-s)^m}{s^{m+1}} \int_0^\infty r e^{\frac{r^2}{s} \left(\frac{1}{2(1-s)} - 1 \right)} dr \\ &\leq 2 \frac{(1-s)^{m+1}}{s^m (1-2s)}, \end{aligned} \quad (\text{F13})$$

$$\begin{aligned} \nu[P_s^{(m,m)}] &\leq 2 \frac{(1-s)^m}{s^{m+1}} \int_0^\infty r^3 e^{\frac{r^2}{s} \left(\frac{1}{2(1-s)} - 1 \right)} dr \\ &\leq 4 \frac{(1-s)^{m+2}}{s^{m-1} (1-2s)^2}. \end{aligned} \quad (\text{F14})$$

We can therefore upper bound $\mu[P_{s,M}] = \mu_{s,M}$, based on the number-state decomposition of $\rho^{(M)}$, as

$$\mu_{s,M}^{(\text{UB})} = \sum_{m,n=0}^{M-1} |\langle m|\rho^{(M)}|n\rangle| \mu_{s,m,n} \geq \mu_{s,M}, \quad (\text{F15})$$

$$\mu_{s,m,m} = \frac{2(1-s)^{m+1}}{s^m(1-2s)}, \quad \mu_{s,m,n \neq m} = \frac{2^{2+\frac{|m-n|}{2}}(1-s)^{1+\frac{m+n}{2}}(|m-n|+1)^{\min[m,n]} \Gamma\left[1+\frac{|m-n|}{2}\right]}{\pi s^{\frac{m+n}{2}}(1-2s)^{1+\frac{|m-n|}{2}} \sqrt{m!n!}}. \quad (\text{F16})$$

By comparing Eqs. (F9) and (F10) and Eqs. (F13) and (F14), we see that, in both the on- and off-diagonal cases, the ratio between (the upper bounds on) the contributions to ν and μ can be expressed (for $m \geq n$) as

$$\frac{\nu^{(\text{UB})}[P_s^{(m,n)}]}{\mu^{(\text{UB})}[P_s^{(m,n)}]} = \frac{s(1-s)}{1-2s}(2+m-n). \quad (\text{F17})$$

We now have two options: we can explicitly write upper bounds on μ and ν , by combining Eqs. (F9), (F10), (F13), and (F14) with a number-state description of ρ , or we can upper bound the ratio between $\nu^{(\text{UB})}$ and $\mu^{(\text{UB})}$ (though perhaps loosely) as

$$\frac{\nu^{(\text{UB})}[P_{s,M}]}{\mu^{(\text{UB})}[P_{s,M}]} \leq \frac{s(1-s)}{1-2s}(M+1). \quad (\text{F18})$$

Equation (F18) is less than approximately $M/2$ times Eq. (F17) and ϵ is concave, so we choose this method and hence arrive at Eq. (32).

APPENDIX G: DISTANCE OF $\sigma_{s,M}$ FROM $\rho^{(M)}$

Our goal is to determine how far each state in our parametrized sequence, $\{\sigma_{s,M}\}$, is from the truncated state, $\rho^{(M)}$. Recall that each state $\sigma_{s,M}$ is obtained from ρ by first truncating it to an M -dimensional representation, $\rho^{(M)}$, and then convolving the P representation of the resulting state, $P_M(\alpha)$, with $e^{-\frac{1}{s}|\alpha|^2}/(s\pi)$ (equivalently, applying channel $\mathcal{C}_s$), to obtain state $\sigma_{s,M}$, with P representation $P_{s,M}(\alpha)$. It is clear that as $M \rightarrow \infty$ and $s \rightarrow 0$, $\sigma_{s,M}$ converges to ρ but we are interested in the rate of convergence. We split $\|\rho - \sigma_{s,M}\|$ into separate contributions from the truncation and from the convolution, δ_M and δ_s .

Some care is required in handling the energy truncation from ρ to $\rho^{(M)}$. We define

$$\rho^{(M)} = \eta_M^{-1} \sum_{m,n=0}^{M-1} \langle m|\rho|n\rangle |m\rangle\langle n|, \quad \eta_M = \sum_{m=0}^{M-1} \langle m|\rho|m\rangle, \quad (\text{G1})$$

where η_M is the probability that a photon-counting measurement on ρ would give a result of $M-1$ or less. We can calculate the fidelity between $\rho^{(M)}$ and $\rho_{s,M}$ for an arbitrary pure state, $|\Gamma\rangle$, by expressing $|\Gamma\rangle$ as $\sum_{m=0}^{\infty} \gamma_m |m\rangle$ for

some complex parameters $\{\gamma_m\}$ such that $\sum_{m=0}^{\infty} |\gamma_m|^2 = 1$. Then,

$$\begin{aligned} F(|\Gamma\rangle\langle\Gamma|, |\Gamma\rangle\langle\Gamma|^{(M)})^2 &= \eta_M^{-1} \sum_{l,p=0}^{\infty} \sum_{m,n=0}^{M-1} \gamma_l^* \gamma_m \gamma_n^* \gamma_p \langle l|m\rangle \langle n|p\rangle \\ &= \eta_M^{-1} \sum_{m,n=0}^{M-1} |\gamma_m|^2 |\gamma_n|^2 = \eta_M, \end{aligned} \quad (\text{G2})$$

since $\sum_{m=0}^M |\gamma_m| = \eta_M$. Then, from the Fuchs–van de Graaf inequality, $\delta_M = 2\sqrt{1-\eta_M}$. η_M can be lower bounded, using the average photon number of the state, as $\eta_M \geq 1 - \bar{n}/M$, so for pure states, $\delta_M = 2\sqrt{1-\eta_M} \leq 2\sqrt{\bar{n}/M}$.

For mixed states, using the joint concavity of quantum fidelity, $F(\rho, \rho^{(M)}) \geq \sum p_i \sqrt{\eta_{M,i}} \geq \sum p_i \eta_{M,i} = \eta_M$, where $\eta_{M,i}$ are the values of η_M calculated on each of the individual states $|\Gamma_i\rangle$ in the decomposition of ρ and p_i are their respective probabilities (replacing the sum with an integral and the probabilities with probability densities in the case of a continuous distribution). Hence, $\delta_M \leq 2\sqrt{1-\eta_M^2}$. Alternatively, we can use the convexity of the trace norm, the concavity of the square root function, and Jensen's inequality to obtain $\delta_M \leq (2/\sqrt{M}) \sum p_i \sqrt{\bar{n}_i} \leq (2/\sqrt{M}) \sqrt{\sum p_i \bar{n}_i} = 2\sqrt{\bar{n}/M}$, where $\bar{n}_i$ are the average photon numbers of the states $|\Gamma_i\rangle$, recovering the same bound as in the pure case. The bound that we choose depends on which is tighter, in a particular case.

As for $\|\rho - \rho^{(M)}\|$, we bound δ_s by calculating the fidelity between $\rho^{(M)}$ and $\sigma_{s,M}$ and applying a Fuchs–van de Graaf inequality. The overlap between the states, $\text{Tr}[\rho^{(M)} \sigma_{s,M}]$, is given by

$$\begin{aligned} \text{Tr}[\rho^{(M)} \sigma_{s,M}] &= \pi \int P_s(\alpha) [\rho^{(M)}] Q(\alpha) [\rho^{(M)}] d^2\alpha \\ &= \pi \int \left(\frac{1}{s\pi} e^{-\frac{1}{s}|\alpha|^2} \star P(\alpha) [\rho^{(M)}] \right) \\ &\quad \times \left(\frac{1}{\pi} e^{-|\alpha|^2} \star P(\alpha) [\rho^{(M)}] \right) d^2\alpha, \end{aligned} \quad (\text{G3})$$

where $Q(\alpha)$ is the Husimi Q representation. Since $F(A, B)^2 \geq \text{Tr}[AB]$, we can use this overlap to bound the

fidelity (sometimes defined as the square of the quantity that we call fidelity here). Some care is required here, since our lower bound on the squared fidelity, based on the overlap, is not tight and does not converge to 1 as $s \rightarrow 0$ unless the state is pure.

First, we assess $\text{Tr}[|\psi\rangle\langle\psi| \mathcal{C}_s[|\psi\rangle\langle\psi|]]$ for an arbitrary pure state $|\psi\rangle$ (lying in the Hilbert space of $\rho^{(M)}$). Expressing $|\psi\rangle$ as $\sum_{m=0}^{M-1} a_m |m\rangle$ for some $\{a_m\}$ such that $\sum_{m=0}^{M-1} a_m a_m^* = 1$, the P representation of $\mathcal{C}_s[|\psi\rangle\langle\psi|]$ is

$$P_s[|\psi\rangle\langle\psi|] = \sum_{m,n=0}^{M-1} |a_m a_n^*| P_{s,\theta(m,n)}^{(m,n)},$$

$$\theta(m,n) = \arg[a_{\max[m,n]} a_{\min[m,n]}^*], \quad (\text{G4})$$

where the expressions for $P_s^{(m,n)}$ are given by Eqs. (F2) and (39) and where we have again included the subscript θ for completeness. The linearity of the trace means that we can decompose $\text{Tr}[|\psi\rangle\langle\psi| \mathcal{C}_s[|\psi\rangle\langle\psi|]]$ into a sum of contributions:

$$\begin{aligned} \text{Tr}[|\psi\rangle\langle\psi| \mathcal{C}_s[|\psi\rangle\langle\psi|]] &= \pi \int P_s(\alpha)[|\psi\rangle\langle\psi|] Q(\alpha)[|\psi\rangle\langle\psi|] d^2\alpha \\ &= \sum_{m_1,n_1=0}^{M-1} \sum_{m_2,n_2=0}^{M-1} |a_{m_1} a_{n_1}^* a_{m_2} a_{n_2}^*| \gamma_s(m_1, n_1, \theta_1, m_2, n_2, \theta_2), \end{aligned} \quad (\text{G5})$$

where each contribution $\gamma_s(m_1, n_1, \theta_1, m_2, n_2, \theta_2)$ is defined by

$$\begin{aligned} \gamma_s(m_1, n_1, \theta_1, m_2, n_2, \theta_2) &= \pi \int_0^\infty \int_0^{2\pi} P_{s,\theta_1}^{(m_1,n_1)}(re^{i\phi}) Q_{\theta_2}^{(m_2,n_2)}(re^{i\phi}) r d\phi dr, \end{aligned} \quad (\text{G6})$$

$$\begin{aligned} Q_\theta^{(m,n)}(re^{i\phi}) &= \lim_{s \rightarrow 1} P_{s,\theta}^{(m,n)}(re^{i\phi}) = \cos(\theta - (m-n)\phi) \frac{e^{-r^2} r^{m+n}}{\pi \sqrt{m!n!}}, \end{aligned} \quad (\text{G7})$$

$$Q^{(m,m)}(re^{i\phi}) = \lim_{s \rightarrow 1} P_s^{(m,m)}(re^{i\phi}) = \frac{e^{-r^2} r^{2m}}{\pi m!}. \quad (\text{G8})$$

From the angular dependence of the P and Q representations, it is immediate that $\gamma_s(m_1, n_1, \theta_1, m_2, n_2, \theta_2)$ is only nonzero if $|m_1 - n_1| = |m_2 - n_2|$. This is because, for $|m_1 - n_1| \neq |m_2 - n_2|$,

$$\int_0^{2\pi} \cos(\theta_1 - (m_1 - n_1)\phi) \cos(\theta_2 - (m_2 - n_2)\phi) d\phi = 0. \quad (\text{G9})$$

Note too that $\gamma_s(m_1, n_1, \theta_1, m_2, n_2, \theta_2)$ is symmetric under swapping m and n , so we can set $m \geq n$.

Starting by looking at the off-diagonals and setting $\Delta = m - n$ and $m > n$,

$$\begin{aligned} \gamma_s(m_1, m_1 - \Delta, \theta_1, m_2, m_2 - \Delta, \theta_2) &= \sqrt{\frac{(m_1 - \Delta)!}{m_1! m_2! (m_2 - \Delta)!}} \frac{(s-1)^{m_1 - \Delta}}{\pi s^{m_1 + 1}} \\ &\times \int_0^{2\pi} \cos(\theta_1 - \Delta\phi) \cos(\theta_2 - \Delta\phi) d\phi \\ &\times \int_0^\infty e^{-\frac{1+s}{s} r^2} r^{2m_2+1} L_{m_1 - \Delta}^\Delta \left[\frac{r^2}{s(1-s)} \right] dr. \end{aligned} \quad (\text{G10})$$

Then, by carrying out the integration over ϕ and explicitly expanding the Laguerre polynomial,

$$\begin{aligned} \gamma_s(m_1, m_1 - \Delta, \theta_1, m_2, m_2 - \Delta, \theta_2) &= \cos(\theta_1 - \theta_2) \sqrt{\frac{(m_1 - \Delta)!}{m_1! m_2! (m_2 - \Delta)!}} \frac{(s-1)^{m_1 - \Delta}}{s^{m_1 + 1}} \\ &\times \sum_{i=0}^{m_1 - \Delta} \frac{(-1)^i}{s^i (1-s)^i i!} \binom{m_1}{\Delta + i} \int_0^\infty e^{-\frac{1+s}{s} r^2} r^{2(m_2+i)+1} dr. \end{aligned} \quad (\text{G11})$$

Next, we calculate that for any non-negative integer x ,

$$\int_0^\infty e^{-\frac{1+s}{s} r^2} r^{2x+1} dr = \frac{x!}{2} \left(\frac{s}{1+s} \right)^{1+x}, \quad (\text{G12})$$

where, per convention, $0! = 1$ [61]. Using Eq. (G12), we obtain

$$\begin{aligned} \gamma_s(m_1, m_1 - \Delta, \theta_1, m_2, m_2 - \Delta, \theta_2) &= \cos(\theta_1 - \theta_2) \frac{s^{m_2 - m_1}}{2} \sqrt{\frac{(m_1 - \Delta)!}{m_1! m_2! (m_2 - \Delta)!}} \frac{(s-1)^{m_1 - \Delta}}{(1+s)^{m_2 + 1}} \sum_{i=0}^{m_1 - \Delta} \frac{(m_2 + i)!}{(s^2 - 1)^i i!} \binom{m_1}{\Delta + i} \\ &= \cos(\theta_1 - \theta_2) \frac{s^{m_2 - m_1}}{2} \sqrt{\binom{m_1}{\Delta} \binom{m_2}{\Delta}} \frac{{}_2F_1(\Delta - m_1, m_2 + 1, \Delta + 1, (1-s^2)^{-1})}{(1+s)^{m_2 + 1} (s-1)^{\Delta - m_1}}, \end{aligned} \quad (\text{G13})$$

where ${}_2F_1(a, b, c, z)$ is the hypergeometric function. Applying Pfaff's transformation rule for hypergeometric functions,

$$\gamma_s(m_1, m_1 - \Delta, \theta_1, m_2, m_2 - \Delta, \theta_2) = \frac{\cos(\theta_1 - \theta_2)}{2} \sqrt{\binom{m_1}{\Delta} \binom{m_2}{\Delta}} \frac{{}_2F_1(\Delta - m_1, \Delta - m_2, \Delta + 1, s^{-2})}{s^{2\Delta - m_1 - m_2} (1 + s)^{m_1 + m_2 + 1 - \Delta}}. \quad (\text{G14})$$

This rearrangement is useful because it highlights the fact that γ_s is symmetric under swapping m_1 and m_2 (and θ_1 and θ_2) and also because ${}_2F_1(\Delta - m_1, \Delta - m_2, \Delta + 1, s^{-2})$ can be expressed as a sum over only positive terms (rather than a sum of terms with alternating signs). Since γ_s is symmetric, we set $m_1 \leq m_2$. Then, since $\Delta - m_1 = -n_1$ and $\Delta - m_2 = -n_2$ are both negative (we will substitute back and forth between m and n depending on what is easiest),

$$\begin{aligned} & s^{m_1 + m_2 - 2\Delta} {}_2F_1(\Delta - m_1, \Delta - m_2, \Delta + 1, s^{-2}) \\ &= s^{n_1 + n_2} \sum_{k=0}^{n_1} (-1)^k \binom{n_1}{k} \frac{(-n_2)_k}{(\Delta + 1)_k} s^{-2k} \\ &= s^{n_2 - n_1} \sum_{k=0}^{n_1} \binom{n_1}{k} \frac{n_2!}{(n_2 - k)!} \frac{\Delta!}{(\Delta + 1)!} s^{2(n_1 - k)}. \quad (\text{G15}) \end{aligned}$$

We define the polynomial

$$G[m_1, m_2, \Delta, s] = \sum_{k=0}^{n_1} \binom{n_1}{k} \frac{n_2!}{(n_2 - k)!} \frac{\Delta!}{(\Delta + k)!} s^{2(n_1 - k)}, \quad (\text{G16})$$

where we assume that $m_2 \geq m_1 \geq \Delta$. Crucially, this is a polynomial in s that only has positive coefficients, so we are guaranteed that it will evaluate to a finite positive value for any value of s . Finally, we can rewrite Eq. (G14) as

$$\begin{aligned} & \gamma_s(m_1, m_1 - \Delta, \theta_1, m_2, m_2 - \Delta, \theta_2) \\ &= \frac{\cos(\theta_1 - \theta_2)}{2} \sqrt{\binom{m_1}{\Delta} \binom{m_2}{\Delta}} \frac{s^{m_2 - m_1} G[m_1, m_2, \Delta, s]}{(1 + s)^{m_1 + m_2 + 1 - \Delta}}. \quad (\text{G17}) \end{aligned}$$

For on-diagonals ($m = n$), we have

$$\begin{aligned} \gamma_s(m_1, m_1, m_2, m_2) &= \frac{1}{m_2!} \frac{(s-1)^{m_1}}{\pi s^{m_1+1}} \int_0^{2\pi} 1 d\phi \int_0^\infty e^{-\frac{1+s}{s} r^2} r^{2m_2+1} L_{m_1} \left[\frac{r^2}{s(1-s)} \right] dr \\ &= \frac{2}{m_2!} \frac{(s-1)^{m_1}}{s^{m_1+1}} \sum_{i=0}^{m_1} \binom{m_1}{i} \frac{(-1)^i}{s^i (1-s)^i i!} \int_0^\infty e^{-\frac{1+s}{s} r^2} r^{2(m_2+i)+1} dr \\ &= \frac{s^{m_2-m_1}}{m_2!} \frac{(s-1)^{m_1}}{(1+s)^{m_2+1}} \sum_{i=0}^{m_1} \binom{m_1}{i} \frac{(m_2+i)!}{(s^2-1)^i i!} \\ &= s^{m_2-m_1} \frac{(s-1)^{m_1}}{(1+s)^{m_2+1}} {}_2F_1(-m_1, m_2+1, 1, (1-s^2)^{-1}) \\ &= s^{m_1+m_2} \frac{{}_2F_1(-m_1, -m_2, 1, s^{-2})}{(1+s)^{m_1+m_2+1}}, \quad (\text{G18}) \end{aligned}$$

where the only difference from the off-diagonal case is in the prefactor, coming from the integration of the angular part. We have dropped the θ dependence because, per Eq. (G4), for $m = n$, $\theta = \arg[a_m a_m^*] = 0$. Again, the expression is symmetric under interchange of m_1 and m_2 , so we again set $m_2 \geq m_1$. Then,

$$\gamma_s(m_1, m_1, m_2, m_2) = s^{m_2-m_1} \sum_{k=0}^{m_1} \binom{m_1}{k} \frac{m_2!}{(m_2-k)! k!} s^{2(m_1-k)} = \frac{s^{m_2-m_1} G[m_1, m_2, 0, s]}{(1+s)^{m_1+m_2+1}}. \quad (\text{G19})$$

Recall that we are only interested in the fidelity for small s . We will therefore construct a small- s approximation of the squared fidelity. From Eqs. (G17) and (G19), we can see that the gamma functions go to 0 as $s \rightarrow 0$ except in the case of $m_1 = m_2$, due to the prefactor of $s^{m_2-m_1}$. For $m_1 = m_2$, Eqs. (G17) and (G19) become

$$\gamma_s(m, n, \theta, m, n, \theta) = \frac{1}{2} \binom{m}{n} \frac{G[m, m, m-n, s]}{(1+s)^{m+n+1}}, \quad (\text{G20})$$

$$\gamma_s(m, m, m, m) = \frac{G[m, m, 0, s]}{(1+s)^{2m+1}}. \quad (\text{G21})$$

As an aside, Eq. (G21) is exactly the squared fidelity if our input, ρ , is a Fock state and so can be used directly in this case (as we do in Sec. VIB). From Eq. (G16), we can see that only the $k = n_1$ term survives as $s \rightarrow 0$, so Eqs. (G20) and (G21) go to $\frac{1}{2}$ and 1, respectively, as $s \rightarrow 0$. Assessing Eq. (G5) for $s = 0$ (and again dropping unnecessary θ dependence), we therefore obtain

$$\begin{aligned} \text{Tr}[\psi \langle \psi | \mathcal{C}_0[\psi] \langle \psi |] \\ = \sum_{m=0} |a_m^4| \gamma_0(m, m, m, m) + 2 \sum_{\substack{m, n=0, \\ m \neq n}} |a_m^2 a_n^2| \gamma_0(m, n, m, n) \\ = \sum_{m, n=0} |a_m^2 a_n^2| = 1, \end{aligned} \quad (\text{G22})$$

where the factor of 2 comes from summing both $\gamma_0(m, n, m, n)$ and $\gamma_0(m, n, n, m)$.

To lower bound the fidelity for nonzero (but small) s , we take the lowest-ordered terms of $\gamma_s(m, m, m, m)$ and $\gamma_s(m, n, m, n)$, i.e., we take the s^0 terms of Eqs. (G20) and (G21) and neglect all other terms. This is valid because, as we will now show, the sum over all terms of the form $\gamma_s(m, n, m+q, n+q)$ (for $q > 0$) is positive. That is,

$$\begin{aligned} \sum_{q=1}^{M-2} \sum_{m, n=0}^{M-1-q} (2 - \delta_{\text{kron}}(m, n)) |a_{m+q} a_{n+q} a_m a_n| \\ \times \gamma_s(m, n, \theta_1, m+q, n+q, \theta_2) \geq 0, \end{aligned} \quad (\text{G23})$$

where δ_{kron} is the Kronecker delta and the prefactor of $(2 - \delta_{\text{kron}}(m, n))$ is so that we count both $\gamma_s(m, n, m+q, n+q)$ and $\gamma_s(n, m, m+q, n+q)$ for $m \neq n$. For completeness, we should also sum over $\gamma_s(m+q, n+q, m, n)$ and $\gamma_s(m+q, n+q, n, m)$, but since this only results in a prefactor of 2 on all terms, we can neglect this. We rewrite

Eq. (G17) (for $q > 0$ and $m \neq n$) as

$$\begin{aligned} \gamma_s(m, n, \theta_1, m+q, n+q, \theta_2) \\ = \frac{\cos(\theta_1 - \theta_2)}{2} \frac{\sqrt{m!n!(m+q)!(n+q)!}}{(1+s)^{m+n+q+1}} \\ \times \sum_{j=0}^{\min[m, n]} \frac{s^{q+2j}}{j!(q+j)!(n-j)!(m-j)!} \\ = \frac{1}{4|a_{m+q} a_{n+q} a_m a_n|} \sum_{j=0}^{\min[m, n]} \psi_{qj, m} \psi_{qj, n}^* + \psi_{qj, m}^* \psi_{qj, n}, \end{aligned} \quad (\text{G24})$$

where we define

$$\psi_{qj, r} = \frac{a_r a_{r+q}^* s^{j+\frac{q}{2}} \sqrt{r!(r+q)!}}{(1+s)^{r+\frac{q+1}{2}} (r+q)! \sqrt{j!(q+j)!}} \quad (\text{G25})$$

and we use the fact that

$$\begin{aligned} a_m a_{m+q}^* a_n^* a_{n+q} + a_m^* a_{m+q} a_n a_{n+q}^* \\ = 2|a_{m+q} a_{n+q} a_m a_n| \cos\left(\arg[a_m a_n^*] - \arg[a_{m+q} a_{n+q}^*]\right). \end{aligned} \quad (\text{G26})$$

Similarly, Eq. (G19) becomes

$$\gamma_s(m, m, m+q, m+q) = \frac{1}{|a_{m+q}^2 a_m^2|} \sum_{j=0}^m \psi_{qj, m} \psi_{qj, m}^*. \quad (\text{G27})$$

Thus, the condition from Eq. (G23), which we want to prove, becomes

$$\begin{aligned} \sum_{q=1}^{M-2} \sum_{m, n=0}^{M-1-q} \sum_{j=0}^{\min[m, n]} \psi_{qj, m} \psi_{qj, n}^* \\ = \sum_{q=1}^{M-2} \sum_{j=0}^{M-1-q} \sum_{m, n=j}^{M-1-q} \psi_{qj, m} \psi_{qj, n}^* \geq 0. \end{aligned} \quad (\text{G28})$$

Defining ψ_{qj} as the $1 \times (M-j-q)$ vector with entries $\psi_{qj, r}$ for r ranging from j to $M-1-q$, we note that the left-hand side of Eq. (G28) is the sum over q and j of the sum of all entries of the matrices $\psi_{qj} \psi_{qj}^\dagger$. Summing all entries of a positive-semidefinite matrix gives a number that is ≥ 0 and since $\psi_{qj}^\dagger \psi_{qj} \geq 0$, $\psi_{qj} \psi_{qj}^\dagger$ is positive semidefinite.

Hence, we can validly lower bound the fidelity by summing only the s^0 terms in Eqs. (G20) and (G21).

Specifically,

$$\text{Tr}[|\psi\rangle\langle\psi| \mathcal{C}_s[|\psi\rangle\langle\psi|]] \geq \sum_{m,n=0}^{M-1} \frac{|a_m^2 a_n^2|}{(1+s)^{m+n+1}}, \quad (\text{G29})$$

where we only use the first term in $G[m, m, n, 0]$. This is a convex function in s , so for pure states, we can write

$$\begin{aligned} F^2(|\psi\rangle\langle\psi|, \mathcal{C}_s[|\psi\rangle\langle\psi|]) &\geq 1 - s \sum_{m,n=0}^{M-1} |a_m^2 a_n^2| (m+n+1) \\ &\geq 1 - s \left(1 + 2 \sum_{m=0}^{M-1} |a_m^2| m \right), \end{aligned} \quad (\text{G30})$$

where we have lower bounded it using the first-order Taylor expansion around $s = 0$. Using the definition of the average energy,

$$F^2(|\psi\rangle\langle\psi|, \mathcal{C}_s[|\psi\rangle\langle\psi|]) \geq 1 - s(1 + 2\bar{n}(|\psi\rangle\langle\psi|)), \quad (\text{G31})$$

and from the Fuchs–van de Graaf relations,

$$\delta_s(|\psi\rangle\langle\psi|) \leq 2\sqrt{s(1 + 2\bar{n}(|\psi\rangle\langle\psi|))}. \quad (\text{G32})$$

Using the convexity of the trace norm and Jensen's inequality, we arrive at Eq. (29).

Finally, we note that we are free to decompose $\|\rho - \sigma_{s,M}\|$ in a different way, namely,

$$\begin{aligned} \|\rho - \sigma_{s,M}\| &= \|\rho - \mathcal{C}_s[\rho^{(M)}]\| \\ &= \|\rho - \mathcal{C}_s[\rho] + \mathcal{C}_s[\rho] - \mathcal{C}_s[\rho^{(M)}]\| \\ &\leq \|\rho - \mathcal{C}_s[\rho]\| + \|\mathcal{C}_s[\rho] - \mathcal{C}_s[\rho^{(M)}]\| \\ &\leq \|\rho - \mathcal{C}_s[\rho]\| + \|\rho - \rho^{(M)}\| \\ &\leq \|\rho - \mathcal{C}_s[\rho]\| + \delta_M, \end{aligned} \quad (\text{G33})$$

where, on the fourth line, we have used the data-processing inequality. Where convenient, we can therefore calculate $\|\rho - \mathcal{C}_s[\rho]\|$ instead of $\|\rho^{(M)} - \mathcal{C}_s[\rho^{(M)}]\|$. We will call the quantity δ_s in both cases, to avoid unnecessary extra notation.

APPENDIX H: BOUNDING THE OUTPUT DISTANCE OVER ALL STATES

To upper bound $\mu^{(\text{UB})}$ over all states $\rho^{(M)}$, we use the fact $\rho^{(M)}$ is a positive-semidefinite matrix, and hence

$$|\langle m|\rho^{(M)}|n\rangle| \leq \frac{\langle m|\rho^{(M)}|m\rangle + \langle n|\rho^{(M)}|n\rangle}{2}. \quad (\text{H1})$$

Upper bounding the upper bound on μ from Eq. (F15), we obtain

$$\begin{aligned} \mu_{s,M}^{(\text{UB})} &\leq \sum_{m,n=0}^{M-1} \frac{\langle m|\rho^{(M)}|m\rangle + \langle n|\rho^{(M)}|n\rangle}{2} \mu_{s,m,n} \\ &= \sum_{m=0}^{M-1} \langle m|\rho^{(M)}|m\rangle \sum_{n=0}^{M-1} \mu_{s,m,n} \leq \max_m \left[\sum_{n=0}^{M-1} \mu_{s,m,n} \right]. \end{aligned} \quad (\text{H2})$$

For small s , we will show that this maximum is achieved by setting $m = M - 1$. We must note that we are taking an upper bound on an upper bound here, so the resulting bound on μ may be very loose. However, a loose bound is to be expected, since our goal is to make an extremely general statement that holds even without having any description of the input state at all, other than knowing its average photon number. We could improve this bound somewhat by using the average photon number (i.e., by accounting for the fact that $\sum_m \langle m|\rho^{(M)}|m\rangle m = \bar{n}$) but this would result in more complicated expressions.

We can upper bound the row sum $\sum_{n=0}^{M-1} \mu_{s,m,n}$ by bounding the ratio between neighboring terms. We start by calculating

$$\begin{aligned} \frac{\mu_{s,m,n}}{\mu_{s,m,n-1}} &= \sqrt{\frac{(1-s)(1-2s)}{2s}} \frac{m-n+1}{\sqrt{n}} \frac{\Gamma\left[\frac{m-n+2}{2}\right]}{\Gamma\left[\frac{m-n+3}{2}\right]} \\ &> \sqrt{\frac{(1-s)(1-2s)}{s}} \frac{m-n+1}{\sqrt{n(m-n+3)}}, \quad \text{for } m > n, \end{aligned} \quad (\text{H3})$$

where we bound the ratio between gamma functions using Gautschi's inequality. Similarly,

$$\begin{aligned} \frac{\mu_{s,m,n+1}}{\mu_{s,m,n}} &= \sqrt{\frac{2(1-s)}{s(1-2s)}} \frac{\sqrt{n+1}}{n-m+1} \frac{\Gamma\left[\frac{n-m+3}{2}\right]}{\Gamma\left[\frac{n-m+2}{2}\right]} \\ &> \sqrt{\frac{1-s}{s(1-2s)}} \sqrt{\frac{n+1}{n-m+1}}, \quad \text{for } m < n. \end{aligned} \quad (\text{H4})$$

Next, we calculate

$$\begin{aligned} \frac{\mu_{s,m,m}}{\mu_{s,m,m-1}} &= \sqrt{\frac{\pi}{2}} \sqrt{\frac{(1-s)(1-2s)}{sm}}, \\ \frac{\mu_{s,m,m+1}}{\mu_{s,m,m}} &= \sqrt{\frac{2}{\pi}} \sqrt{\frac{(1-s)(m+1)}{s(1-2s)}}, \end{aligned} \quad (\text{H5})$$

where we have used $\Gamma[\frac{3}{2}] = \sqrt{\pi}/2$. Upper and lower bounding the n -dependent terms (for $m \neq 0$),

$$\begin{aligned} \frac{\mu_{s,m,n}}{\mu_{s,m,n-1}} &> \sqrt{\frac{(1-s)(1-2s)}{sm}}, \quad \text{for } m \geq n; \\ \frac{\mu_{s,m,n+1}}{\mu_{s,m,n}} &> \sqrt{\frac{1-s}{s(1-2s)}}, \quad \text{for } m \leq n. \end{aligned} \quad (\text{H6})$$

From Eq. (H6), we can see that, as long as s decreases at least with M^{-1} (more specifically, as long as $((1-s)(1-2s))/(s(M-1)) > 1$), each $\mu_{s,m,n-1} < \mu_{s,m,n}$ (for both $m \geq n$ and $m < n$). Since $\mu_{s,m,n} = \mu_{s,n,m}$, this means that $\sum_{n=0}^{M-1} \mu_{s,m-1,n} < \sum_{n=0}^{M-1} \mu_{s,m,n}$. Hence, the $m = M-1$ row has the largest row sum and thus [using Eq. (F16)]

$$\begin{aligned} \mu_{s,M}^{(\text{UB})} &\leq \sum_{n=0}^{M-1} \mu_{s,M-1,n} < \mu_{s,M-1,M-1} \\ &\times \sum_{k=0}^{M-1} \left(\frac{s(M-1)}{(1-s)(1-2s)} \right)^{\frac{k}{2}} < \frac{2(1-s)^M M}{s^{M-1}(1-2s)}, \end{aligned} \quad (\text{H7})$$

where we could have used the formula for the sum of a geometric sequence to obtain a slightly tighter bound.

Substituting Eq. (H7) into Eq. (31), we recover Eq. (1), i.e.,

$$\begin{aligned} \|(\Psi - \Phi)[\rho]\| &\leq \frac{2(1-s)^M M}{s^{M-1}(1-2s)} \epsilon \left(\epsilon_0, \frac{s(1-s)(M+1)}{1-2s} \right) \\ &+ 4\sqrt{s(1+2\bar{n})} + 4\sqrt{\frac{\bar{n}}{M}}. \end{aligned}$$

Finally, we consider one particular relationship between s and M . To study the scaling with the input energy, we will set $(s(1-s)(M+1))/(1-2s)$ to a constant with regard to $\bar{n}$, allowing us to eliminate any dependence on the specific form of ϵ . For this, we need s to be linear in M^{-1} . If we choose $s = 1/(\kappa(M+3))$, for some $\kappa > 1$, then the argument to ϵ is upper bounded by $(s(1-s)(M+1))/(1-2s) < 1/\kappa$ for all $\kappa > 1$. Consequently, for any $\kappa > 1$,

$$\begin{aligned} \|(\Psi - \Phi)[\rho]\| &\leq 2(M+3)^M \kappa^{M-1} \epsilon \left(\epsilon_0, \frac{1}{\kappa} \right) \\ &+ 4\sqrt{\frac{1+2\bar{n}}{\kappa(M+3)}} + 4\sqrt{\frac{\bar{n}}{M}}. \end{aligned} \quad (\text{H8})$$

This can also hold for $\kappa < 1$ as long as M is not very small [and with some care required with the condition $((1-s)(1-2s))/(s(M-1)) > 1$]. In terms of the

leading-order terms, this can be expressed, per Corollary 1, as

$$\begin{aligned} \|(\Psi - \Phi)[\rho]\| &\leq \mathcal{O}[e^{M \log(M\kappa)} \kappa^{-1} \epsilon(\epsilon_0, \kappa^{-1})] \\ &+ \mathcal{O}\left[\bar{n}^{\frac{1}{2}} M^{-\frac{1}{2}}\right]. \end{aligned}$$

Per the discussion after Corollary 1, we can retrieve an upper bound on the scaling with $\bar{n}$ by setting $M \sim \bar{n}$. Some caution is needed here, as strictly $\bar{n}$ is a continuous variable and M is discrete, but we could get around this by using the ceiling function (and we are focused on the scaling, so the difference is not very important).

APPENDIX I: OUT-OF-DISTRIBUTION GENERALIZATION FOR COHERENT STATES WHEN $\epsilon_0 > 0$

We now use some of the techniques from the previous appendixes to revisit out-of-distribution generalization for coherent states. The aim now is to explicitly show that we can always (i.e., regardless of the class of the target channel) construct a concave function $\epsilon(\epsilon_0, r^2)$ that bounds the output distance for coherent states and that has the properties listed in Theorem 1.

Recall that coherent states have the number-state decomposition

$$|re^{i\phi}\rangle \langle re^{i\phi}|_{\text{coh}} = e^{-r^2} \sum_{m,n=0}^{\infty} \frac{r^{m+n} e^{i(m-n)\phi}}{\sqrt{m!n!}} |m\rangle \langle n|. \quad (\text{I1})$$

We use the same method as in Sec. VB, replacing $|re^{i\phi}\rangle \langle re^{i\phi}|_{\text{coh}}$ with $\mathcal{C}_s[|re^{i\phi}\rangle \langle re^{i\phi}|_{\text{coh}}]$. Per Eq. (29), $|re^{i\phi}\rangle \langle re^{i\phi}|_{\text{coh}}$ has a distance from $\mathcal{C}_s[|re^{i\phi}\rangle \langle re^{i\phi}|_{\text{coh}}]$ of no more than $2\sqrt{s(1+2r^2)}$. We then upper bound $\|(\Psi - \Phi)[\mathcal{C}_s[|re^{i\phi}\rangle \langle re^{i\phi}|_{\text{coh}}]]\|$ by bounding $\|(\Psi - \Phi)[\mathcal{C}_s[|m\rangle \langle m|]]\|$ and $\|(\Psi - \Phi)[\frac{1}{2}\mathcal{C}_s[e^{i\theta}|m\rangle \langle n| + e^{-i\theta}|n\rangle \langle m|]]\|$ for every m, n , and θ .

Note that we have previously upper bounded these same quantities (in Sec. VI B and Appendix F); however, here our starting point is different. Previously, we assumed that we had a known concave function $\epsilon(\epsilon_0, r^2)$, but now finding such a function is the goal. Instead, we only assume that $\|(\Psi - \Phi)[|re^{i\phi}\rangle \langle re^{i\phi}|_{\text{coh}}]\| \leq \epsilon_0$ for $r^2 \leq \tau^2$. Then, we can apply the step function $[\epsilon_{\text{step}}$ from Eq. (6)] that assigns ϵ_0 for $r \leq \tau$ and 2 for $r > \tau$.

Recall that the P representation of $|m\rangle \langle m|$ is given by Eq. (39). $\|(\Psi - \Phi)[\mathcal{C}_s[|m\rangle \langle m|]]\|$ can therefore be upper bounded by

$$\begin{aligned} \|(\Psi - \Phi)[\mathcal{C}_s[|m\rangle \langle m|]]\| &\leq \epsilon_0 \mu_{\tau} [P_s^{(m,m)}] + 2 \left(\mu [P_s^{(m,m)}] - \mu_{\tau} [P_s^{(m,m)}] \right), \end{aligned} \quad (\text{I2})$$

$$\begin{aligned}\mu_\tau[P_s^{(m,m)}] &= \int_0^{2\pi} \int_0^\tau |P_s^{(m,m)}(re^{i\phi})| r d\phi dr, \\ \mu[P_s^{(m,m)}] &= \int_0^{2\pi} \int_0^\infty |P_s^{(m,m)}(re^{i\phi})| r d\phi dr, \quad (I3)\end{aligned}$$

where we have applied the same technique as in Sec. VB of separating the negative and positive parts of the P representation, but applying the step function instead of $\epsilon(\epsilon_0, r^2)$. Since all of the mass of $P_s^{(m,m)}$ concentrates around the origin as $s \rightarrow 0$ (since the decay exponent becomes larger as s decreases), the second term approaches 0 as s does. On the other hand, $\mu_\tau[P_s^{(m,m)}]$ approaches ∞ . Following the methods used in Eq. (F13), we obtain (for $s < \frac{1}{2}$)

$$\mu_\tau[P_s^{(m,m)}] \leq 2 \left(1 - e^{-\tau^2 \frac{1-2s}{2s(1-s)}}\right) \frac{(1-s)^{m+1}}{s^m(1-2s)}, \quad (I4)$$

so we can write the upper bound

$$\begin{aligned}\|(\Psi - \Phi)[\mathcal{C}_s[|m\rangle\langle m|]]\| \\ \leq 2 \frac{(1-s)^{m+1}}{s^m(1-2s)} \left(\epsilon_0 + (2 - \epsilon_0)e^{-\tau^2 \frac{1-2s}{2s(1-s)}}\right). \quad (I5)\end{aligned}$$

Noting that Eq. (I5) is trivial for large m , we provide the nontrivial bound $\|(\Psi - \Phi)[\mathcal{C}_s[|m\rangle\langle m|]]\| \leq \xi_{\tau,s}^{(m,m)}$, where

$$\xi_{\tau,s}^{(m,m)} = \min \left[2 \frac{(1-s)^{m+1}}{s^m(1-2s)} \left(\epsilon_0 + (2 - \epsilon_0)e^{-\tau^2 \frac{1-2s}{2s(1-s)}}\right), 2 \right]. \quad (I6)$$

Following a similar approach for the off-diagonals, we recall that the P representation of $\frac{1}{2}\mathcal{C}_s[e^{i\theta}|m\rangle\langle n| + e^{-i\theta}|n\rangle\langle m|]$ is given by Eq. (F2). Proceeding similarly to the on-diagonal case, we must find $\mu_\tau[P_s^{(m,n)}]$. Following Eq. (F9),

$$\mu_\tau[P_s^{(m,n)}] \leq \frac{2^{2+\frac{m-n}{2}}(1-s)^{1+\frac{m+n}{2}}(m-n+1)_n}{\pi s^{\frac{m+n}{2}}(1-2s)^{1+\frac{m-n}{2}}\sqrt{m!n!}} \left(\Gamma\left[1 + \frac{m-n}{2}\right] - \Gamma\left[1 + \frac{m-n}{2}, \frac{\tau^2(1-2s)}{2s(1-s)}\right] \right), \quad (I7)$$

where $\Gamma[a, b]$ is the incomplete gamma function [61]. Hence, $\frac{1}{2}\|(\Psi - \Phi)[\mathcal{C}_s[e^{i\theta}|m\rangle\langle n| + e^{-i\theta}|n\rangle\langle m|]]\| \leq \xi_{\tau,s}^{(m,n)}$, where

$$\xi_{\tau,s}^{(m,n)} = \min \left[\frac{2^{2+\frac{m-n}{2}}(1-s)^{1+\frac{m+n}{2}}(m-n+1)_n}{\pi s^{\frac{m+n}{2}}(1-2s)^{1+\frac{m-n}{2}}\sqrt{m!n!}} \left(\epsilon_0 \Gamma\left[1 + \frac{m-n}{2}\right] + (2 - \epsilon_0) \Gamma\left[1 + \frac{m-n}{2}, \frac{\tau^2(1-2s)}{2s(1-s)}\right] \right), 2 \right]. \quad (I8)$$

Finally, putting these various elements together, we obtain

$$\|(\Psi - \Phi)[|re^{i\phi}\rangle\langle re^{i\phi}|_{\text{coh}}]\| \leq \epsilon(\epsilon_0, r^2) = \min \left[\inf_{0 < s < \frac{1}{2}} \left\{ e^{-r^2} \sum_{m,n=0}^{\infty} \frac{r^{m+n}}{\sqrt{m!n!}} \xi_{\tau,s}^{(m,n)} + 4\sqrt{s(1+2r^2)} \right\}, 2 \right]. \quad (I9)$$

From Eqs. (I6) and (I8), we see that decreasing ϵ_0 (or increasing τ) for fixed s decreases $\xi_{\tau,s}^{(m,n)}$. For sufficiently small ϵ_0 , we can choose s such that Eq. (I9) is nontrivial for any r . Equation (I9) is not known to be concave but we can always make it concave by taking its upper concave hull. Nonetheless, it is not a practically useful bound, as we require ϵ_0 to be extremely small for it to be nontrivial for large r . However, it is sufficient to prove that a concave bounding function $\epsilon(\epsilon_0, r^2)$ can always be constructed and hence that out-of-distribution generalization is always possible.

APPENDIX J: TIGHTENING THE BOUND FOR SPATSS

For low-energy SPATSS, the negativity is large, so the bound on the distance between the output states, given by Eq. (38), may be loose. To tighten it, we may consider using the same technique as for states with infinite negativity, i.e., we can bound the output distance for the input $\sigma_{s,\text{SPAT}}$, where $\sigma_{s,\text{SPAT}}$ is given by applying the Gaussian additive noise channel $\mathcal{C}_s$ to ρ_{SPAT} . The distance between ρ_{SPAT} and $\sigma_{s,\text{SPAT}}$ is given by

Eq. (29). Parametrizing with q instead of $\bar{n}$, we obtain $\delta_s \leq 2\sqrt{s(3+4q)}$.

Using the P representation of ρ_{SPAT} from Eq. (35), we find that the P representation of $\sigma_{s,\text{SPAT}}$ is

$$\left(\frac{1}{s\pi}e^{-\frac{1}{s}|\alpha|^2}\right) \star P_{\text{SPAT}}(re^{i\phi}) = \frac{1+q}{\pi(q+s)^3} \left(r^2 - \frac{(q+s)(1-s)}{1+q}\right) e^{-\frac{r^2}{q+s}}. \quad (\text{J1})$$

Integrating separately over the negative $[r^2 < ((q+s)(1-s))/(1+q)]$ and positive regions, we find that the values of μ_s and ν_s/μ_s for $\sigma_{s,\text{SPAT}}$ are

$$\mu_s = 2e^{-\frac{1-s}{1+q}} \frac{1+q}{q+s} - 1, \quad \frac{\nu}{\mu} = 1 + 2q + s - \frac{2(1-s)^2}{2+2q - e^{-\frac{1-s}{1+q}}(q+s)} < 1 + 2q + s. \quad (\text{J2})$$

We can then write the bound

$$\|(\Psi - \Phi)[\rho_{\text{SPAT}}]\| \leq \left(2e^{-\frac{1-s}{1+q}} \frac{1+q}{q+s} - 1\right) \times \epsilon \left(\epsilon_0, 1 + 2q + s - \frac{2(1-s)^2}{2+2q - e^{-\frac{1-s}{1+q}}(q+s)}\right) + 4\sqrt{s(3+4q)}, \quad (\text{J3})$$

which reduces to Eq. (38) in the limit of $s \rightarrow 0$. Since the argument to concave function ϵ is approximately linear in s , a small change in s may not increase $\epsilon(\epsilon_0, \nu/\mu)$ by much. The δ_s term is also sublinear in s . On the other hand, for small q , a small change in s can result in a significant change in the multiplicative factor μ_s . Hence, we may obtain tighter bounds by choosing $s > 0$.

APPENDIX K: APPLICATION TO SQUEEZED VACUUMS

Recall, from Eq. (G33), that we are free to calculate δ_s as $\|\rho_{\text{sq}} - \mathcal{C}_s[\rho_{\text{sq}}]\|$, rather than $\|\rho_{\text{sq}}^{(M)} - \mathcal{C}_s[\rho_{\text{sq}}^{(M)}]\|$. Since squeezed vacuums and the channel $\mathcal{C}_s$ are both Gaussian, we can calculate the fidelity between ρ_{sq} and $\mathcal{C}_s[\rho_{\text{sq}}]$ [31]:

$$F(\rho_{\text{sq}}, \mathcal{C}_s[\rho_{\text{sq}}]) = \left(1 + s^2 + 2s \frac{1 + \lambda^2}{1 - \lambda^2}\right)^{-\frac{1}{4}}. \quad (\text{K1})$$

We therefore retrieve Eq. (43).

Using Eqs. (F15), (F16), and (42), we can calculate $\mu_{s,M}$ by calculating the contributions from each element of ρ_{sq} . For any off-diagonal element with $p > q$,

$$|\langle 2p | \rho | 2q \rangle| \mu_{s,2p,2q} = \frac{4\sqrt{1-\lambda^2}(1-s)}{\pi(1-2s)^{p-q+1}} \frac{(1-s)^{p+q} \lambda^{p+q}}{s^{p+q}} \times \frac{(1+2p-2q)_{2q}(p-q)!}{2^{2q}p!q!}, \quad (\text{K2})$$

where we have separated out terms based on their dependence on p and q . For on-diagonal elements,

$$|\langle 2p | \rho | 2p \rangle| \mu_{s,2p,2p} = \frac{2\sqrt{1-\lambda^2}(1-s)}{(1-2s)} \frac{(1-s)^{2p} \lambda^{2p}}{s^{2p}} \times \frac{(2p)!}{2^{2p}(p!)^2}. \quad (\text{K3})$$

Note that if and only if $((1-s)\lambda)/s < 1$, the sum $\sum_q |\langle 2(p_0+q) | \rho | 2q \rangle| \mu_{s,2(p_0+q),2q}$ for q ranging from 0 to ∞ is convergent. This is as expected, since $((1-s)\lambda)/s < 1$ is precisely the condition for the negativity of a squeezed vacuum to be finite (in fact, zero). However, we want to be able to take s to 0 as $\epsilon_0 \rightarrow 0$, so we are interested in the divergent region. Instead, we set $x = ((1-2s)(1-s)\lambda)/s$, so that

$$\begin{aligned} \sum_{q=0}^{p-1} |\langle 2p | \rho | 2q \rangle| \mu_{s,2p,2q} &= \frac{4\sqrt{1-\lambda^2}}{\pi} \frac{\lambda^p (1-s)^{p+1}}{s^p (1-2s)^{p+1}} \sum_{q=0}^{p-1} \frac{x^q (1+2p-2q)_{2q} (p-q)!}{2^{2q}p!q!} \\ &= \frac{4\sqrt{1-\lambda^2}}{\pi} \frac{\lambda^p (1-s)^{p+1}}{s^p (1-2s)^{p+1}} \left((1+x)^{p-\frac{1}{2}} - \frac{x^p {}_2F_1\left(\frac{1}{2}, 1, 1+p, -x\right) (2p)!}{2^{2p}(p!)^2} \right). \end{aligned} \quad (\text{K4})$$

Multiplying by 2 to account for $p < q$ and adding the on-diagonal contribution,

$$\begin{aligned} |\langle 2p | \rho | 2p \rangle| \mu_{s,2p,2p} + 2 \sum_{q=0}^{p-1} |\langle 2p | \rho | 2q \rangle| \mu_{s,2p,2q} &= \frac{8\sqrt{1-\lambda^2}}{\pi} \frac{\lambda^p (1-s)^{p+1}}{s^p (1-2s)^{p+1}} \left((1+x)^{p-\frac{1}{2}} \right. \\ &\quad \left. + \frac{x^p (2p)!}{2^{2p}(p!)^2} \left(\frac{\pi}{4} - {}_2F_1\left(\frac{1}{2}, 1, 1+p, -x\right) \right) \right). \end{aligned} \quad (\text{K5})$$

Applying a Pfaff transformation to the hypergeometric function so that its final argument is $x/(1+x)$ and then writing it in power-series form, we can easily verify that it is always positive. Hence, we can upper bound this expression with

$$\begin{aligned} & |\langle 2p|\rho|2p\rangle|\mu_{s,2p,2p} + 2 \sum_{q=0}^{p-1} |\langle 2p|\rho|2q\rangle|\mu_{s,2p,2q} \\ & \leq \frac{2(1-s)\sqrt{1-\lambda^2}}{(1-2s)} \left(\frac{4}{\pi\sqrt{1+x}} \left(\frac{\lambda(1-s)(1+x)}{s(1-2s)} \right)^p \right. \\ & \quad \left. + \frac{(2p)!}{2^{2p}(p!)^2} \left(\frac{\lambda(1-s)x}{s(1-2s)} \right)^p \right). \end{aligned} \quad (\text{K6})$$

The first term is a geometric sequence in p , so is simple to sum. For the second term, $(2p)!/(2^{2p}(p!)^2) \leq 1$ (this can be verified by taking the ratio between this expression for p and $p+1$ and seeing it is a decreasing function of p). Hence, we can upper bound the second term by replacing $(2p)!/(2^{2p}(p!)^2)$ with its maximum value of 1. Then, both terms are geometric sequences, so we can recover Eq. (45) by summing over p (note that the maximum value of p is $(M-1)/2$, not $M-1$) and then normalizing by dividing by η_M (since we recall that $\rho^{(M)}$ is normalized).

There are two approaches that we can take. If we know the form of $\epsilon(\epsilon_0, r^2)$, we can investigate the region in which a nontrivial bound is possible, in terms of $\bar{n}$ and ϵ_0 . Alternatively, we can eliminate the dependence on the function ϵ by setting $(s(1-s)(M+1))/(1-2s) = \bar{n}$. We must set $s = (M+1+2\bar{n} - \sqrt{(M+1)^2 + 4\bar{n}^2})/(2(M+1))$, so that the condition becomes, per Eq. (34),

$$\begin{aligned} \epsilon(\epsilon_0, \bar{n}) & < \frac{2(\sqrt{(M+1)^2 + 4\bar{n}^2} - 2\bar{n})(M+1+2\bar{n} - \sqrt{(M+1)^2 + 4\bar{n}^2})^{M-1}}{M(M+1-2\bar{n} + \sqrt{(M+1)^2 + 4\bar{n}^2})^M} \\ & \times \left(1 - 2\sqrt{\frac{\bar{n}}{M}} - \sqrt{\frac{2(2\bar{n}+1)(M+1+2\bar{n} - \sqrt{(M+1)^2 + 4\bar{n}^2})}{M+1}} \right). \end{aligned}$$

Note that we do still require $((1-s)(1-2s))/(s(M-1)) > 1$ but this condition is met for all M , for this choice of relationship between s and M , as long as $\bar{n} < 1$, which is the case for Fig. 4; for larger $\bar{n}$, we would need to take care that this still holds. The region in which the bound is nontrivial is then defined by $\bar{n}$ and $\epsilon(\epsilon_0, \bar{n})$, where $\epsilon(\epsilon_0, \bar{n})$ is the maximum output trace norm for a classical probe of average photon number $\bar{n}$. While the function ϵ still appears in this inequality, we can treat $\epsilon(\epsilon_0, \bar{n})$ as a single variable and so carry out a comparison between the classical and non-classical cases. From Eq. (34), we observe that a necessary condition for there to exist a positive $\epsilon(\epsilon_0, \bar{n})$ satisfying the

APPENDIX L: ANALYTICAL AND NUMERICAL INVESTIGATION OF THE TIGHTNESS OF THE BOUNDS FOR NONCLASSICAL STATES

The trace norm of a difference between quantum states is constrained to lie in the range $[0, 2]$, so an upper bound ≥ 2 is trivial. Theorems 1 and 2, taken together, show that for any fixed input state, ρ , we can nontrivially bound the output distance, provided that ϵ_0 is small enough. In a learning setting, we are able to make ϵ_0 arbitrarily small by taking more samples, i.e., by increasing N . If we were to fix a specific learning model and class of channels, it might also be possible to understand the minimum N required to nontrivially bound the output distance for a particular ρ , but this would be very problem specific.

On the other hand, for fixed ϵ_0 , we might wonder for which set of states we are able to nontrivially bound the output distance. In the finite-negativity scenario of Corollary 3, it is simple to check whether the right-hand side of Eq. (5) is ≥ 2 . However, the region of nontriviality for the more general infinite-negativity scenario of Theorem 2 is harder to characterize, since there are two free parameters to tune. For Eq. (1) to be nontrivial, we require there to exist $0 \leq s < \frac{1}{2}$ and integer $M \geq \bar{n}$ [and the condition $((1-s)(1-2s))/(s(M-1)) > 1$ to be met] such that the bound in Eq. (1) is < 2 .

condition is

$$\begin{aligned} & 1 - \sqrt{\frac{2(2\bar{n}+1)(M+1+2\bar{n} - \sqrt{(M+1)^2 + 4\bar{n}^2})}{M+1}} \\ & - 2\sqrt{\frac{\bar{n}}{M}} > 0. \end{aligned} \quad (\text{L1})$$

The minimal value of $(\log_{10} \text{ of }) \epsilon(\epsilon_0, \bar{n})$ required to satisfy Eq. (34) is plotted in Fig. 4. In theory, we could solve for M and so only have a condition in $\bar{n}$ and $\epsilon(\epsilon_0, \bar{n})$, but this is difficult to do analytically. Note that in Fig. 4, M is

treated as a continuous variable but in practice we must constrain it to take integer values. We can see that the largest minimal value of $\epsilon(\epsilon_0, \bar{n})$ is generally obtained by choosing an M as close as possible to the curve defined by Eq. (L1), i.e., the smallest value of M that fulfils the condition, although we can see from the plotted contours that this may not be exactly true in all cases. We could also substitute in a particular expression for $\epsilon(\epsilon_0, \bar{n})$, in order to investigate the dependence on ϵ_0 , for a particular channel model, although in theory we could get slightly tighter bounds by tuning s and M in Eq. (1) directly.

Per Sec. VID, we now look numerically at some specific examples of channels and input states, in order to investigate whether and when the bounds can be tight. First, we will consider whether Theorem 2 and Corollaries 2 and 3 can be tight for cases in which $\epsilon(\epsilon_0, r^2)$ is known and tight. That is, we will fix a target channel, Ψ , and a learned channel, Φ , for which we can calculate the output distance for coherent states and will investigate whether our bounds on the extension to nonclassical states can be tight. Trivially, Corollary 2 can be saturated by choosing a pure coherent state, so we will start by focusing on Corollary 3, i.e., checking whether Eq. (24) can be tight.

Since SPATs are relatively simple states, with finite P representations, that we have already studied (in Sec. VIA), we will use them as an example. Equation (24) makes use of the triangle inequality, summing separate contributions from the negative and positive supports of the P representation of the input state, so we might expect it to be tight when the action of the channels is very different on these two domains. Consequently, we choose what we call the parity channel, which is defined by the Kraus operators $|\psi_E\rangle\langle 2m|$ and $|\psi_O\rangle\langle 2m+1|$, for m from 0 to ∞ . In other words, the channel performs a photon-number-counting measurement and returns state $|\psi_E\rangle$ if the result is even and $|\psi_O\rangle$ if the result is odd. We choose that for both the target channel and the learned channel, $|\psi_E\rangle = |0\rangle$, but that the outputs of each channel for odd photon numbers, $|\psi_O^\Psi\rangle$ and $|\psi_O^\Phi\rangle$, differ by $\| |\psi_O^\Psi\rangle\langle \psi_O^\Psi| - |\psi_O^\Phi\rangle\langle \psi_O^\Phi| \| = \gamma$. We can immediately see that the Fock state $|1\rangle$ would be an optimal probe for discriminating between channels Ψ and Φ , so we may expect that low-energy, high-negativity SPATs will have a large output distance.

For a coherent-state input, the output distance for channels Ψ_{par} and Φ_{par} is

$$\begin{aligned} \|(\Psi_{\text{par}} - \Phi_{\text{par}})[|re^{i\phi}\rangle\langle re^{i\phi}|]\| &= \gamma e^{-r^2} \sum_{m=0}^{\infty} \frac{r^{4m+2}}{(2m+1)!} \\ &= \gamma e^{-r^2} \sinh(r^2). \end{aligned} \quad (\text{L2})$$

This starts at 0 for $r^2 = 0$ and tends toward $\gamma/2$. SPATs have the Fock-basis decomposition

$$\rho_{\text{SPAT}}(q) = \frac{1}{1+q} \sum_{m=0}^{\infty} \frac{(m+1)q^m}{(1+q)^{m+1}} |m+1\rangle\langle m+1|, \quad (\text{L3})$$

so we can calculate the output distance exactly as

$$\begin{aligned} \|(\Psi_{\text{par}} - \Phi_{\text{par}})[\rho_{\text{SPAT}}(q)]\| &= \frac{\gamma}{1+q} \sum_{m=0}^{\infty} \frac{(2m+1)q^{2m}}{(1+q)^{2m+1}} \\ &= \frac{\gamma}{2} \left(1 + \frac{1}{(1+2q)^2} \right). \end{aligned} \quad (\text{L4})$$

Substituting our expression for $\|(\Psi_{\text{par}} - \Phi_{\text{par}})[|re^{i\phi}\rangle\langle re^{i\phi}|]\|$ into Eqs. (24) and (25) and using the quantities derived in Sec. VIA, we can compare the bounds from Corollary 3 with the exact value of the output distance, from Eq. (L4). This comparison is carried out in Fig. 5, in which we plot the output trace norm and the bounds, up to a multiplicative factor of γ . We can see that the bound from Eq. (24) is not much larger than the true value (although any such assessment is subjective) and the two converge as q becomes large. This latter observation is not surprising, as the negativity tends to 0 as q tends to ∞ . To illustrate this, we also plot the ratio between the true value and the bound from Eq. (24) as a function of the negativity of the corresponding SPATs. Since the ratio never drops below approximately 0.788, this suggests that our bounds can be fairly tight. However, we note that we are only looking here at a particular channel model and that SPATs can only have low energy and high negativity or high energy and low negativity, i.e., we have not looked at states with both high energy and high negativity, to see if the bounds can still be tight. It remains an open question whether it is possible, for every $\bar{n}$ and $\mathcal{N}$, to find a triple of Ψ , Φ , and ρ_{in} such that ρ_{in} has average photon number $\bar{n}$ and negativity $\mathcal{N}$ and Eq. (24) is tight.

Next, we look numerically at an example for the case of input states with infinite negativity. Specifically, we focus on the phase-rotation channel (previously discussed in Sec. IV) for one-mode squeezed vacuum inputs. Recall, from Eq. (13), that the output trace norm for coherent-state inputs is $\|(\Psi_{\text{PR}} - \Phi_{\text{PR}})[|re^{i\phi}\rangle\langle re^{i\phi}|]\| = 2\sqrt{1 - e^{-2r^2(1 - \cos(\theta))}}$, where θ is the difference in the applied phases.

Since squeezed vacuums are Gaussian states and phase rotations are Gaussian unitaries, we can calculate the exact output fidelity for a coherent-state input using the formula for the fidelity of a Gaussian state [31]. Specifically, we want to find the fidelity between two zero-mean Gaussian states with covariance matrices

$$V_\Psi = \begin{pmatrix} e^{-2q} & 0 \\ 0 & e^{2q} \end{pmatrix}, \quad V_\Phi = \begin{pmatrix} e^{-2q} \cos^2(\theta) + e^{2q} \sin^2(\theta) & \sin(2\theta) \sinh(2q) \\ \sin(2\theta) \sinh(2q) & e^{2q} \cos^2(\theta) + e^{-2q} \sin^2(\theta) \end{pmatrix}, \quad (\text{L5})$$

where $q = \operatorname{arctanh}(\sqrt{\bar{n}/(\bar{n}+1)})$, and the resulting fidelity is

$$F_{\text{out}} = (2\bar{n}(\bar{n}+1)(1 - \cos(2\theta)) + 1)^{-\frac{1}{4}}. \quad (\text{L6})$$

Using the Fuchs–van de Graaf relation, $\|(\Psi_{\text{PR}} - \Phi_{\text{PR}})[\rho_{\text{sq}}]\| = 2\sqrt{1 - (2\bar{n}(\bar{n}+1)(1 - \cos(2\theta)) + 1)^{-\frac{1}{2}}}$.

When applying the bound from Sec. VIC, we must minimize over parameters s and M . Numerically (see the

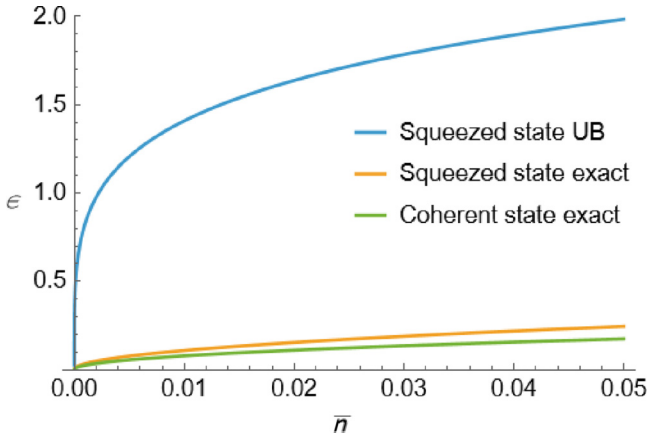


FIG. 6. The output trace norm for a pair of phase-rotation channels. The parameter difference between the target and learned channels is $\frac{\pi}{8}$. We plot the exact value for a coherent-state input, the exact value for a one-mode squeezed vacuum, and the bound for a one-mode squeezed vacuum, coming from Eq. (47). UB, upper bound.

Supplemental Material [61]), we find that the bound from Eq. (47) [i.e., setting $s = \lambda/(1+\lambda)$] is often not much worse than the bound we would obtain from a full minimization using Eq. (1), unless θ is very small. As such, for our comparison, we set $s = \sqrt{\bar{n}(\bar{n}+1)} - \bar{n}$, so that Eq. (47) becomes

$$\|(\Psi_{\text{PR}} - \Phi_{\text{PR}})[\rho_{\text{sq}}]\| \leq \epsilon_{\text{PR}}(\epsilon_0, \sqrt{\bar{n}(\bar{n}+1)}) + 4\sqrt{1 - \frac{\bar{n}+1 + \sqrt{\bar{n}(\bar{n}+1)}}{(\bar{n}+1)\sqrt{4\bar{n}+1 + 4\sqrt{\bar{n}(\bar{n}+1)}}}}. \quad (\text{L7})$$

We then plot this bound, for $\theta = \pi/8$, in Fig. 6.

From Fig. 6, we can see that the upper bound is very loose. There could be a few reasons for this. First, we are looking here at a very simple class of Gaussian channels but our bound holds over all channels that have the same output distance for coherent-state inputs. There could exist non-Gaussian channels with the same coherent-state output distance but a much higher output distance for squeezed states. We have focused on the phase-rotation channels because, due to their simplicity, we can calculate the exact output distance for squeezed states.

Alternatively, our estimation of $\mu_{s,M}$, using Eq. (45), could be too large. If so, a tighter upper bound on $\mu_{s,M}$ could allow us to decrease s , and so δ_s , without making $\mu_{s,M} \in (\epsilon_0, \nu_{s,M}/\mu_{s,M})$ large. In Fig. 7, demonstrates that this is indeed the case. The true values of $\mu_{s,M}$ were calculated by constructing the P representation, using Eqs. (39), (F2), and (42), and then numerically integrating the absolute value of it, for various values of s . The downside of

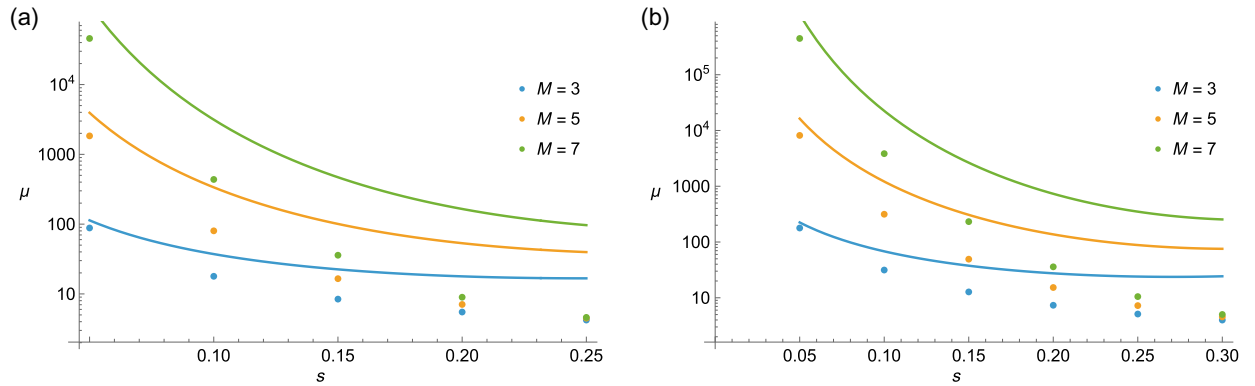


FIG. 7. A comparison of the true value of $\mu_{s,M}$ for a (truncated) squeezed state, $\rho_{\text{sq}}^{(M)}$, and the upper bound on $\mu_{s,M}$ from Eq. (45). The true values are given as points, while the upper bounds are the continuous lines: squeezed states with average photon number (a) $\frac{1}{4}$ and (b) $\frac{1}{10}$.

this method is that we do not get an analytical formula but just a numerical value. However, this is enough to demonstrate that the gap between $\mu_{s,M}$ and $\mu_{s,M}^{\text{UB}}$ can be quite extreme. This suggests a way in which the bounds for squeezed states could be improved.

-
- [1] V. Gebhart, R. Santagati, A. A. Gentile, E. M. Gauger, D. Craig, N. Ares, L. Banchi, F. Marquardt, L. Pezzè, and C. Bonato, Learning quantum systems, *Nat. Rev. Phys.* **5**, 141 (2023).
 - [2] S. L. Braunstein and P. Van Loock, Quantum information with continuous variables, *Rev. Mod. Phys.* **77**, 513 (2005).
 - [3] C. Weedbrook, S. Pirandola, R. García-Patrón, N. J. Cerf, T. C. Ralph, J. H. Shapiro, and S. Lloyd, Gaussian quantum information, *Rev. Mod. Phys.* **84**, 621 (2012).
 - [4] M. Lobino, D. Korystov, C. Kupchak, E. Figueroa, B. C. Sanders, and A. I. Lvovsky, Complete characterization of quantum-optical processes, *Science* **322**, 563 (2008).
 - [5] S. Rahimi-Keshari, A. Scherer, A. Mann, A. T. Reza-khani, A. I. Lvovsky, and B. C. Sanders, Quantum process tomography with coherent states, *New J. Phys.* **13**, 013006 (2011).
 - [6] M. Cooper, E. Slade, M. Karpiński, and B. J. Smith, Characterization of conditional state-engineering quantum processes by coherent state quantum process tomography, *New J. Phys.* **17**, 033041 (2015).
 - [7] M. Kervinen, S. Ahmed, M. Kudra, A. Eriksson, F. Quijandria, A. F. Kockum, P. Delsing, and S. Gasparinetti, Extended quantum process tomography of logical operations on an encoded bosonic qubit, *Phys. Rev. A* **110**, L020401 (2024).
 - [8] M. C. Caro, H.-Y. Huang, N. Ezzell, J. Gibbs, A. T. Sornborger, L. Cincio, P. J. Coles, and Z. Holmes, Out-of-distribution generalization for learning quantum dynamics, *Nat. Commun.* **14**, 3751 (2023).
 - [9] Y. Zhang, R. Wiersema, J. Carrasquilla, L. Cincio, and Y. B. Kim, Scalable quantum dynamics compilation via quantum machine learning, *arXiv:2409.16346*.
 - [10] J. Gibbs, Z. Holmes, M. C. Caro, N. Ezzell, H.-Y. Huang, L. Cincio, A. T. Sornborger, and P. J. Coles, Dynamical simulation via quantum machine learning with provable generalization, *Phys. Rev. Res.* **6**, 013241 (2024).
 - [11] S. Jerbi, J. Gibbs, M. S. Rudolph, M. C. Caro, P. J. Coles, H.-Y. Huang, and Z. Holmes, The power and limitations of learning quantum dynamics incoherently, *arXiv:2303.12834*.
 - [12] Y. Du, M.-H. Hsieh, and D. Tao, Efficient learning for linear properties of bounded-gate quantum circuits, *Nat. Commun.* **16**, 3790 (2025).
 - [13] H.-Y. Huang, S. Chen, and J. Preskill, Learning to predict arbitrary quantum processes, *PRX Quantum* **4**, 040337 (2023).
 - [14] T. Volkoff, Z. Holmes, and A. Sornborger, Universal compiling and (no-)free-lunch theorems for continuous-variable quantum learning, *PRX Quantum* **2**, 040327 (2021).
 - [15] B.-B. Li, L. Ou, Y. Lei, and Y.-C. Liu, Cavity optomechanical sensing, *Nanophotonics* **10**, 2799 (2021).
 - [16] Y. Xia, A. R. Agrawal, C. M. Pluchar, A. J. Brady, Z. Liu, Q. Zhuang, D. J. Wilson, and Z. Zhang, Entanglement-enhanced optomechanical sensing, *Nat. Photonics* **17**, 470 (2023).
 - [17] A. J. Brady, X. Chen, Y. Xia, J. Manley, M. Dey Chowdhury, K. Xiao, Z. Liu, R. Harnik, D. J. Wilson, Z. Zhang *et al.*, Entanglement-enhanced optomechanical sensor array with application to dark matter searches, *Commun. Phys.* **6**, 237 (2023).
 - [18] A. Eickbusch, V. Sivak, A. Z. Ding, S. S. Elder, S. R. Jha, J. Venkatraman, B. Royer, S. M. Girvin, R. J. Schoelkopf, and M. H. Devoret, Fast universal control of an oscillator with weak dispersive coupling to a qubit, *Nat. Phys.* **18**, 1464 (2022).
 - [19] Y. Xia, W. Li, W. Clark, D. Hart, Q. Zhuang, and Z. Zhang, Demonstration of a reconfigurable entangled radio-frequency photonic sensor network, *Phys. Rev. Lett.* **124**, 150502 (2020).
 - [20] Q. Zhuang and J. H. Shapiro, Ultimate accuracy limit of quantum pulse-compression ranging, *Phys. Rev. Lett.* **128**, 010501 (2022).
 - [21] I. Coddington, N. Newbury, and W. Swann, Dual-comb spectroscopy, *Optica* **3**, 414 (2016).
 - [22] H. Shi, Z. Chen, S. E. Fraser, M. Yu, Z. Zhang, and Q. Zhuang, Entanglement-enhanced dual-comb spectroscopy, *npj Quantum Inf.* **9**, 91 (2023).
 - [23] A. Hariri, S. Liu, H. Shi, Q. Zhuang, X. Fan, and Z. Zhang, Entangled dual-comb spectroscopy, *Phys. Rev. X* (2025).
 - [24] D. I. Herman, M. Walsh, M. K. Kreider, N. Lordi, E. J. Tsao, A. J. Lind, M. Heyrich, J. Combes, J. Genest, and S. A. Diddams, Squeezed dual-comb spectroscopy, *Science* **387**, 653 (2025).
 - [25] W. Kozłowski and S. Wehner, in *Proceedings of the Sixth Annual ACM International Conference on Nanoscale Computing and Communication* (Association for Computing Machinery, New York, 2019), pp. 1–7.
 - [26] H. Shi, Z. Zhang, and Q. Zhuang, Practical route to entanglement-assisted communication over noisy bosonic channels, *Phys. Rev. Appl.* **13**, 034029 (2020).
 - [27] Y. Shen, N. C. Harris, S. Skirlo, M. Prabhu, T. Baehr-Jones, M. Hochberg, X. Sun, S. Zhao, H. Larochelle, D. Englund *et al.*, Deep learning with coherent nanophotonic circuits, *Nat. Photonics* **11**, 441 (2017).
 - [28] P. Kok, W. J. Munro, K. Nemoto, T. C. Ralph, J. P. Dowling, and G. J. Milburn, Linear optical quantum computing with photonic qubits, *Rev. Mod. Phys.* **79**, 135 (2007).
 - [29] A. Serafini, *Quantum Continuous Variables: A Primer of Theoretical Methods* (CRC Press, Boca Raton, Florida, USA, 2023).
 - [30] K. C. Tan, S. Choi, and H. Jeong, Negativity of quasiprobability distributions as a measure of nonclassicality, *Phys. Rev. Lett.* **124**, 110404 (2020).
 - [31] L. Banchi, S. L. Braunstein, and S. Pirandola, Quantum fidelity for arbitrary Gaussian states, *Phys. Rev. Lett.* **115**, 260501 (2015).
 - [32] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information: 10th Anniversary Edition* (Cambridge University Press, USA, 2010), 10th ed.

- [33] T. Kiesel, W. Vogel, V. Parigi, A. Zavatta, and M. Bellini, Experimental determination of a nonclassical Glauber-Sudarshan P function, *Phys. Rev. A* **78**, 021804 (2008).
- [34] S. Arunachalam and R. de Wolf, *Guest Column: A Survey of Quantum Learning Theory* (Association for Computing Machinery, New York, NY, USA, 2017), Vol. 48, pp. 41–67.
- [35] H.-Y. Huang, M. Broughton, M. Mohseni, R. Babbush, S. Boixo, H. Neven, and J. R. McClean, Power of data in quantum machine learning, *Nat. Commun.* **12**, 2631 (2021).
- [36] L. Banchi, J. Pereira, and S. Pirandola, Generalization in quantum machine learning: A quantum information standpoint, *PRX Quantum* **2**, 040321 (2021).
- [37] M. C. Caro, H.-Y. Huang, M. Cerezo, K. Sharma, A. Sornborger, L. Cincio, and P. J. Coles, Generalization in quantum machine learning from few training data, *Nat. Commun.* **13**, 4919 (2022).
- [38] A. Anis and A. I. Lvovsky, Maximum-likelihood coherent-state quantum process tomography, *New J. Phys.* **14**, 105021 (2012).
- [39] X.-B. Wang, Z.-W. Yu, J.-Z. Hu, A. Miranowicz, and F. Nori, Efficient tomography of quantum-optical Gaussian processes probed with a few coherent states, *Phys. Rev. A* **88**, 022101 (2013).
- [40] Y. S. Teo, K. Park, S. Shin, H. Jeong, and P. Marek, Highly accurate Gaussian process tomography with geometrical sets of coherent states, *New J. Phys.* **23**, 063024 (2021).
- [41] M. Bina, F. Grasselli, and M. G. A. Paris, Continuous-variable quantum probes for structured environments, *Phys. Rev. A* **97**, 012125 (2018).
- [42] A. Candeloro, S. Razavian, M. Piccolini, B. Teklu, S. Olivares, and M. G. A. Paris, Quantum probes for the characterization of nonlinear media, *Entropy* **23**, 1353 (2021).
- [43] S. Moore, Ph.D. thesis, School of Mathematical and Physical Sciences, University of Sussex, 2024.
- [44] S. Pirandola, R. Laurenza, C. Lupo, and J. L. Pereira, Fundamental limits to quantum channel discrimination, *npj Quantum Inf.* **5**, 50 (2019).
- [45] Q. Zhuang and S. Pirandola, Ultimate limits for multiple quantum channel discrimination, *Phys. Rev. Lett.* **125**, 080505 (2020).
- [46] 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).
- [47] S. Pirandola, R. Laurenza, C. Ottaviani, and L. Banchi, Fundamental limits of repeaterless quantum communications, *Nat. Commun.* **8**, 15043 (2017).
- [48] M. E. Shirokov, On the energy-constrained diamond norm and its application in quantum information theory, *Probl. Inf. Transm.* **54**, 20 (2018).
- [49] A. I. Lvovsky and M. G. Raymer, Continuous-variable optical quantum-state tomography, *Rev. Mod. Phys.* **81**, 299 (2009).
- [50] O. Landon-Cardinal, L. C. G. Govia, and A. A. Clerk, Quantitative tomography for continuous variable quantum systems, *Phys. Rev. Lett.* **120**, 090501 (2018).
- [51] M. Kalash and M. V. Chekhova, Wigner function tomography via optical parametric amplification, *Optica* **10**, 1142 (2023).
- [52] Y.-D. Wu, Y. Zhu, G. Chiribella, and N. Liu, Efficient learning of continuous-variable quantum states, *Phys. Rev. Res.* **6**, 033280 (2024).
- [53] F. A. Mele, A. A. Mele, L. Bittel, J. Eisert, V. Giovannetti, L. Lami, L. Leone, and S. F. E. Oliviero, Learning quantum states of continuous variable systems, [arXiv:2405.01431](https://arxiv.org/abs/2405.01431).
- [54] C.-H. Nguyen, K.-W. Tseng, G. Maslennikov, H. C. J. Gan, and D. Matsukevich, Experimental SWAP test of infinite dimensional quantum states, [arXiv:2103.10219](https://arxiv.org/abs/2103.10219).
- [55] I. Pietikäinen, O. Černotík, S. Puri, R. Filip, and S. M. Girvin, Controlled beam splitter gate transparent to dominant ancilla errors, *Quantum Sci. Technol.* **7**, 035025 (2022).
- [56] T. J. Volkoff and Y. Subaşı, Ancilla-free continuous-variable SWAP test, *Quantum* **6**, 800 (2022).
- [57] S. Lloyd, M. Schuld, A. Ijaz, J. Izaac, and N. Killoran, Quantum embeddings for machine learning, [arXiv:2001.03622](https://arxiv.org/abs/2001.03622).
- [58] S. Lloyd, M. Mohseni, and P. Rebentrost, Quantum principal component analysis, *Nat. Phys.* **10**, 631 (2014).
- [59] L. Banchi, J. L. Pereira, S. T. Jose, and O. Simeone, Statistical complexity of quantum learning, *Adv. Quantum Technol.* **2300311** (2024).
- [60] A. Wünsche, Some remarks about the Glauber-Sudarshan quasiprobability, *Acta Phys. Slovaca* **48**, 385 (1998).
- [61] See the Supplemental Material at <http://link.aps.org/supplemental/10.1103/5m7p-kbf3> for various calculations from the main text, some numerics referred to in the main text, and the code used to generate various figures.