Quantum error cancellation in photonic systems: Undoing photon losses

Adam Taylor ⬤,[1,*] Gabriele Bressanini ⬤,[1] Hyukjoon Kwon ⬤,[2] and M. S. Kim[1,2]

[1]*Blackett Laboratory, Imperial College London, London SW7 2AZ, England, United Kingdom*
[2]*Korea Institute for Advanced Study, Seoul 02455, South Korea*

Real photonic devices are subject to photon losses that can decohere quantum information encoded in the system. In the absence of full fault tolerance, quantum error mitigation techniques have been introduced to help manage errors in noisy quantum devices. In this paper, we introduce an error mitigation protocol inspired by probabilistic error cancellation (a popular error mitigation technique in discrete variable systems) for continuous variable systems. We show that our quantum error cancellation protocol can undo photon losses in expectation value estimation tasks. To do this, we analytically derive the (nonphysical) inverse photon loss channel and decompose it into a sum over physically realizable channels with potentially negative coefficients. The bias of our ideal expectation value estimator can be made arbitrarily small at the cost of increasing the sampling overhead. The protocol requires a noiseless amplification followed by a series of photon subtractions. While these operations can be implemented probabilistically, for certain classes of initial state one can avoid the burden of carrying out the amplification and photon subtractions by leveraging Monte Carlo methods to give an unbiased estimate of the ideal expectation value. We validate our proposed mitigation protocol by simulating the scheme on squeezed vacuum states, cat states, and entangled coherent states.

I. INTRODUCTION

Determining whether quantum advantage—defined as the ability of a quantum computer to solve a useful problem significantly faster than any classical method—can be achieved in the near term constitutes a major open problem. The fragile nature of quantum systems makes them susceptible to errors that could corrupt the computation. Long term, we expect that quantum error correction will allow useful quantum algorithms with a proven speedup over their classical counterparts to be implemented reliably [1–5]. However, the large hardware overhead and small error threshold requirements [6,7] necessary for quantum error correction to be feasible mean that fault-tolerant universal computation has yet to be achieved.

Currently available quantum devices are characterized by noisy operations and limited number of qubits. [8–11]. Consequently, these noisy devices are suitable to execute limited classes of quantum algorithms, including variational quantum algorithms [12–14] and quantum simulations [15–17]. In the absence of full fault tolerance, reducing the impact of noise is an essential ingredient in achieving quantum advantage in the near term. Quantum error mitigation (QEM) is one potential route to tackling this challenge [10,18,19].

The key insight underlying QEM protocols is that a noise-free expectation value can be extracted from a noisy device by sampling from multiple circuits with suitable postprocessing of the noisy measurement data [20,21]. Removing unavoidable noise from quantum expectation values is clearly a task with a domain of interest that also extends beyond quantum computing. Therefore, while all these QEM strategies were initially tailored towards the quantum computing community, the techniques may be useful more generally.

QEM techniques have been shown to improve the computational capabilities of noisy devices [17,22–25], with notable examples including zero-noise extrapolation [26,27] and virtual distillation [28,29]. Another promising QEM protocol is probabilistic error cancellation (PEC) [24–26,30–32]. Within this framework, one decomposes the ideal system dynamics into a sum of quantum channels (corresponding to noisy evolutions) with potentially negative coefficients. One then repeatedly probabilistically selects one of these noisy channels, runs the experiment, and records the measurement outcomes. These are then appropriately linearly combined to obtain an unbiased estimator of the ideal expectation value. The wide range of applicability and lack of qubit overhead in PEC, combined with the fact it provides an unbiased estimator, make it one of the most promising QEM protocols for potentially achieving quantum advantage in the near term.

The effectiveness of PEC under a sparse Pauli-Lindblad noise model was demonstrated in Ref. [25], where a sparse learning protocol was proposed that allowed the necessary noise parameters to be accurately estimated. While a significant improvement over unmitigated results was demonstrated, the sampling overhead for this noise model scaled exponentially in qubit number, noise strength, and circuit depth.

Building on this work, PEC has recently been extended to circuits that include measurement-based operations such as midcircuit projective measurements and potential Pauli feedforward operations dependent on those measurement outcomes [33].

So far, most QEM literature has focused on discrete variable (DV) quantum systems. The continuous variable (CV) QEM literature [34,35] has mostly been targeted towards Gaussian boson sampling, a nonuniversal form of quantum computation in which one is interested in sampling from the output photon number distribution obtained by sending squeezed light into a multimode passive linear optical interferometer, a task that is believed to be classically intractable under standard complexity theoretic conjectures [36–38]. The authors of Ref. [34] proposed a PEC-inspired protocol to mitigate against the effects of losses in a Gaussian boson sampler. In particular, their "loss cancellation" scheme does not require changes to the circuit and involves only classical postprocessing of the noisy sampling probabilities to infer their ideal counterparts. However, the protocol is limited to photon number detection and its convergence to the noiseless photon sampling distribution is, in general, difficult to assess and only guaranteed in the small-to-intermediate loss regime. Similarly, the authors of Ref. [39] recently introduced a family of strategies for mitigating losses in DV linear optical quantum computers based around "recycled probabilities." While their approach outperforms postselection in the large loss regime, postselection appears to achieve better results for small-to-intermediate losses. It is also limited to DV input states and again to photon number measurements.

In this paper, we develop a PEC-inspired scheme to mitigate against photon losses in (either DV or CV) photonic systems in expectation value estimation tasks. To achieve this, we first derive the inverse of the photon loss map. While such a map is not completely positive (CP) [5], we show that it can be expressed as a linear combination of physically realizable (i.e., CP) noisy quantum operations. We then suggest how these operations can be (probabilistically) implemented experimentally. These operations include both an amplification and, counterintuitively, a series of photon subtractions. Sampling from these circuits and applying simple postprocessing weights to the measurement outcomes results in an improved estimate of the noise-free ideal expectation value over running the unmitigated circuit. The bias of such an estimate can be made arbitrarily small, at the cost of an increasing the sampling overhead. We also present a Monte Carlo based approach that involves sampling from different initial states and linearly combining the measurement outcomes with suitable weights to obtain an unbiased estimator. This approach scales better to many modes than the first approach and can often be implemented deterministically, but only works for specific classes of input state. We highlight both these methods with clear examples. Our main contribution is an analytic form of the inverse photon loss channel, as well as possible methods to implement this channel on average, which can thus be used as a quantum error mitigation technique which we refer to as quantum error cancellation.

This paper is structured as follows. In Sec. II, we introduce the basics of quantum error mitigation and outline the probabilistic error cancellation framework. We also describe the noise model for photon losses used throughout this paper and argue that, while not universally applicable, it is relevant and accurate for a variety of important experimental setups. In Sec. III B, we present our main results by introducing a quantum error cancellation scheme based around a quasiprobability decomposition of the inverse photon loss map. We discuss its implementation, the classical postprocessing cost, and sampling overhead, and provide theoretical guarantees of accuracy. In Sec. III E we numerically investigate the performance of our protocol via paradigmatic examples involving single mode squeezed vacuum states and cat states. This section considers just a single mode in order to highlight the key ideas behind the proposed mitigation strategy. In Sec. IV, we extend the scheme to the multimode setting where we again discuss implementation, postprocessing costs, and sampling overhead and provide a thorough error analysis. In Sec. IV E, we provide numerical simulations of our scheme being applied to multimode entangled states, namely two mode squeezed vacua (TMSVs) and entangled coherent states. Finally, we offer concluding remarks in Sec. V.

II. PRELIMINARIES

A. Quantum error mitigation

In this paper, we are concerned with weak quantum error mitigation [21]. Given an initial state ρ_0, and a generic ideal unitary evolution $\mathcal{U}_{\text{ideal}}$, the aim of weak quantum error mitigation is to estimate the expectation value $\langle O \rangle_{\text{ideal}} = \text{Tr}[O\,\mathcal{U}_{\text{ideal}}[\rho_0]]$ for a generic observable O. Any attempt to perfectly implement the unitary dynamics ultimately fails due to unavoidable interactions between the system and the environment. As a result, one is only able to implement a noisy evolution, which we denote with $\mathcal{U}_{\text{noisy}}$. Consequently, the noisy dynamics leads to a measured expectation value that reads $\langle O \rangle_{\text{noisy}} = \text{Tr}[O\,\mathcal{U}_{\text{noisy}}[\rho_0]]$. QEM protocols sample from collections of noisy circuits and combine their measurement outcomes with classical postprocessing to give a better estimate of $\langle O \rangle_{\text{ideal}}$, with respect to the unmitigated scenario.

The quality of an estimator $\bar{O}$ of $\langle O \rangle_{\text{ideal}}$ can be assessed by the mean square error (MSE), which is given by

$$\text{MSE}[\bar{O}] = \mathbb{E}[(\bar{O} - \langle O \rangle_{\text{ideal}})^2] \qquad (1)$$

$$= (\text{Bias}[\bar{O}])^2 + \text{Var}[\bar{O}], \qquad (2)$$

where $\text{Bias}[\bar{O}] = |\mathbb{E}[\bar{O}] - \langle O \rangle_{\text{ideal}}|$ and $\text{Var}[\bar{O}] = \mathbb{E}[\bar{O}^2] - \mathbb{E}[\bar{O}]^2$. For example, consider taking N measurements of O on the noisy state $\mathcal{U}_{\text{noisy}}[\rho_0]$. The N measurement outcomes, $(O_1^{(\text{noisy})}, O_2^{(\text{noisy})}, \ldots, O_N^{(\text{noisy})})$, may be averaged over to give the noisy estimator $\bar{O}_{\text{noisy}} = \frac{1}{N}\sum_{i=1}^{N} O_i^{(\text{noisy})}$. The variance of that estimator (often called shot noise) reads $\text{Var}[\bar{O}_{\text{noisy}}] = \frac{1}{N}\text{Var}[O_{\text{noisy}}]$, where $\text{Var}[O_{\text{noisy}}] = \langle O^2 \rangle_{\text{noisy}} - \langle O \rangle_{\text{noisy}}^2$ is the single shot variance of O with respect to the noisy output state $\mathcal{U}_{\text{noisy}}[\rho_0]$. Therefore, in the large N limit, the dominant source of error is the bias, $|\langle O \rangle_{\text{noisy}} - \langle O \rangle_{\text{ideal}}|$.

QEM protocols aim to provide an estimator of $\langle O \rangle_{\text{ideal}}$, denoted with $\bar{O}_{\text{mit}}$, with a smaller bias than $\bar{O}_{\text{noisy}}$. However, this comes at the cost of an increased variance and hence to obtain a mitigated estimator with the same shot noise as the unmitigated estimator will require more data samples to

be taken. This *sampling overhead* approximately amounts to $\mathrm{Var}[\bar{O}_{\mathrm{mit}}]/\mathrm{Var}[\bar{O}_{\mathrm{noisy}}]$ [18]. While the exact details of the bias-sampling overhead tradeoff relation depend on the particular mitigation scheme, the sampling overhead typically scales exponentially with noise strength, circuit depth, and number of qubits [20,21,40,41].

Probabilistic error cancellation

In Ref. [26], Temme *et al.* demonstrated that, given a quasiprobability decomposition of the ideal dynamics $\mathcal{U}_{\mathrm{ideal}}$ in terms of some arbitrary physically realizable (CP) noisy quantum operations $\{\mathcal{F}_j\}$ [42],

$$\mathcal{U}_{\mathrm{ideal}} = \sum_j \omega_j \, \mathcal{F}_j, \tag{3}$$

it is possible to produce an unbiased estimate of $\langle O \rangle_{\mathrm{ideal}}$. Here, $\{\omega_j\}$ is a set of real coefficients that, in order for $\mathcal{U}_{\mathrm{ideal}}$ to be trace preserving, satisfy $\sum_j \omega_j = 1$. We refer to ω_j as a quasiprobability distribution and to Eq. (3) as a quasiprobability representation of $\mathcal{U}_{\mathrm{ideal}}$.

In practice, given a noise completely positive trace-preserving map, one computes the mathematical inverse of such a map (assuming such an inverse exists). This is typically a non-CP map and hence cannot be physically realized. However, the inverse map always admits a quasiprobability representation in terms of physically realizable quantum operations. This can then be combined with the full circuit dynamics to write the ideal circuit dynamics in the quasiprobability representation over noisy CP channels. We refer to Ref. [25] for a detailed demonstration of how one can obtain a quasiprobability decomposition of the ideal dynamics, starting from the noise model. Note that by doing this, one is implicitly assuming that the noise channel and ideal dynamics can be separated (see, e.g., Ref. [43] for a detailed discussion on this point).

We then use the quasiprobability distribution $\{\omega_j\}$ to define a new set of coefficients $q_j := |\omega_j|/S$, where $S = \sum_j |\omega_j|$. These coefficients satisfy $0 \leqslant q_j \leqslant 1$ and $\sum_j q_j = 1$, and hence form a valid probability distribution. The quasiprobability representation of $\mathcal{U}_{\mathrm{ideal}}$ can then be rewritten as

$$\mathcal{U}_{\mathrm{ideal}} = S \sum_j \frac{|\omega_j|}{S} \mathrm{sgn}(\omega_j) \mathcal{F}_j. \tag{4}$$

Via linearity, it is clear that

$$\langle O \rangle_{\mathrm{ideal}} = S \sum_j q_j \, \mathrm{sgn}(\omega_j) \langle O \rangle_{\mathcal{F}_j}, \tag{5}$$

where $\langle O \rangle_{\mathcal{F}_j} = \mathrm{Tr}[O \, \mathcal{F}_j[\rho_0]]$. Using a Monte Carlo approach, the ideal expectation value can be estimated by sampling from the channels $\{\mathcal{F}_j\}$ according to the probability distribution $\{q_j\}$, and performing a single shot measurement of O—this probabilistic choice of $\mathcal{F}_j$ is where probabilistic error cancellation gets its name. We thus obtain a list of measurement outcomes $(O_1^{(\mathcal{F}_{j_1})}, O_2^{(\mathcal{F}_{j_2})}, \ldots, O_N^{(\mathcal{F}_{j_N})})$, where the superscript $(\mathcal{F}_{j_\ell})$ tells us which channel was implemented in the ℓth run of the experiment. These N samples may then be linearly

combined to obtain an *unbiased* mitigated estimator, namely

$$\bar{O}_{\mathrm{mit}} = \frac{S}{N} \sum_{\ell=1}^{N} \mathrm{sgn}(\omega_{j_\ell}) O_\ell^{\mathcal{F}_{j_\ell}}. \tag{6}$$

Under the assumption that $\mathrm{Var}[O_{\mathcal{F}_j}]/\mathrm{Var}[O_{\mathrm{noisy}}] \approx \mathcal{O}(1)$ for all j, the sampling overhead $\mathrm{Var}[\bar{O}_{\mathrm{mit}}]/\mathrm{Var}[\bar{O}_{\mathrm{noisy}}]$ can easily be shown to scale with S^2.

B. Noise model

In photonic platforms, the two dominant sources of noise are photon losses and dephasing errors, while thermal noise is negligible for systems at visible or infrared frequencies near room temperature. In this paper, we focus on the photon loss channel because it is the most significant issue impacting photonic platforms. In Appendix D, we both discuss a strategy for mitigating against dephasing errors and also show that the introduction of a dephasing error channel has no impact on our photon loss mitigation procedure.

Let Λ_γ denote the single mode pure Markovian photon loss channel with noise parameter $\gamma \in [0, 1]$. We model these losses via a beam splitter acting on a single mode input state ρ and an ancillary vacuum state, before tracing over the environmental degrees of freedom, namely [44]

$$\Lambda_\gamma[\rho] = \mathrm{Tr}_{\mathrm{env}}[U_{\mathrm{BS},\gamma} (\rho \otimes |0\rangle\langle 0|) U_{\mathrm{BS},\gamma}^\dagger]. \tag{7}$$

Here, $U_{\mathrm{BS},\gamma} = e^{\arccos(\sqrt{1-\gamma})(\hat{a}_{\mathrm{env}}^\dagger \hat{a} - \hat{a}^\dagger \hat{a}_{\mathrm{env}})}$ denotes the beam splitter unitary, and $1 - \gamma$ its transmissivity, while $\hat{a}$ and $\hat{a}_{\mathrm{env}}$ are the bosonic operators of the system and environment, respectively. For future convenience, we exploit the state's Glauber-Sudarshan P representation [45]

$$\rho = \int_{\mathbb{C}} d^2\alpha \, P_\rho(\alpha) |\alpha\rangle\langle\alpha|, \tag{8}$$

where $|\alpha\rangle$ is a single mode coherent state and P_ρ is the P function of ρ. Hence, we may express Eq. (7) as

$$\Lambda_\gamma[\rho] = \int_{\mathbb{C}} d^2\alpha P_\rho(\alpha) |\sqrt{1-\gamma}\,\alpha\rangle\langle\sqrt{1-\gamma}\,\alpha|. \tag{9}$$

One can also use Eq. (7) to obtain a Kraus representation of the loss map, namely

$$\Lambda_\gamma[\rho] = \sum_{j=0}^{\infty} K_j(\gamma) \, \rho \, K_j^\dagger(\gamma), \tag{10}$$

where the Kraus operators read

$$K_j(\gamma) = \sqrt{\frac{\gamma^j}{j!}}(1-\gamma)^{\frac{\hat{n}}{2}}\hat{a}^j = \sqrt{\frac{1}{j!}\left(\frac{\gamma}{1-\gamma}\right)^j}\hat{a}^j(1-\gamma)^{\frac{\hat{n}}{2}}. \tag{11}$$

We prove the above identity in Appendix A. As the channel is trace preserving, the Kraus operators satisfy $\sum_j K_j^\dagger K_j = \mathbb{I}$, where $\mathbb{I}$ denotes the identity operator on the Hilbert space. Equivalently, it can be seen that Eq. (7) is a solution to the Lindblad master equation

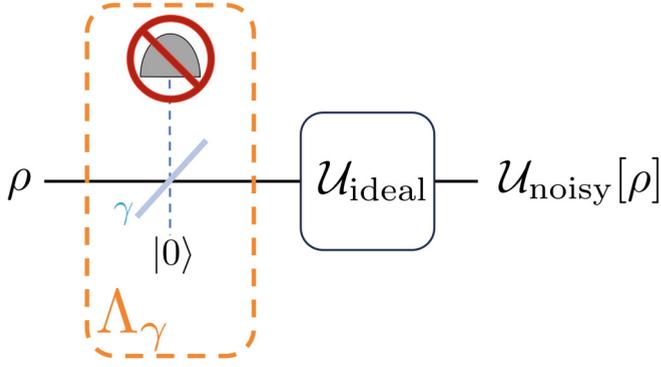$$\frac{d}{dt}\rho(t) = \kappa \mathcal{D}[\hat{a}](\rho(t)), \tag{12}$$

FIG. 1. Graphical representation of the decomposition of the noisy dynamics into a pure loss map followed by an ideal unitary evolution for a single mode, i.e., $\mathcal{U}_{\text{noisy}} = \mathcal{U}_{\text{ideal}} \circ \Lambda_\gamma$. The loss map is modeled via a beam splitter interaction between the system and an environmental mode initialized in the vacuum state, followed by a trace over the environmental degrees of freedom. The beam splitter has a transmissivity of $1 - \gamma$, where $\gamma \in [0, 1)$ is the loss parameter.

where $\mathcal{D}[\hat{a}](\rho) = \hat{a}\rho\hat{a}^\dagger - \frac{1}{2}\{\hat{a}^\dagger\hat{a}, \rho\}$ is the dissipation superoperator and $\kappa > 0$ is the dissipation rate, if we make the identification $\gamma = 1 - e^{-\kappa t}$.

Crucially, we further assume that the noisy dynamics $\mathcal{U}_{\text{noisy}}$ can be decomposed into a pure loss map taking place *before* the system undergoes an ideal evolution, i.e., we write $\mathcal{U}_{\text{noisy}} = \mathcal{U}_{\text{ideal}} \circ \Lambda_\gamma$, as depicted in Fig. 1. A more precise modeling of the noisy dynamics would necessitate solving the full Lindblad master equation in Eq. (12) with an additional term on the right-hand side corresponding to the ideal dynamics. Nonetheless, while the assumption above is not universally applicable, we point out that it is generally justified in a number of relevant cases, particularly in the multimode setting, which we discuss below.

We now extend the above noise model to multiple modes, under the assumption that each mode is subject to independent single mode Markovian photon losses parametrized by some learnable loss parameters. For M modes, let $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, \ldots \gamma_M)$ denote these loss parameters in each mode. The full M-mode loss channel, $\Lambda_{\boldsymbol{\gamma}}^{(M)}$, can then be written as a tensor product over the single mode loss channels, i.e.,

$$\Lambda_{\boldsymbol{\gamma}}^{(M)} = \bigotimes_{i=1}^{M} \Lambda_{\gamma_i, i} \tag{13}$$

where $\Lambda_{\gamma_i, i}$ only acts nontrivially on the ith mode. In the P representation, an M-mode state can be written as

$$\rho = \int_{\mathbb{C}^M} d^{2M}\boldsymbol{\alpha} P_\rho(\boldsymbol{\alpha}) |\alpha_1, \ldots, \alpha_M\rangle\langle\alpha_1, \ldots, \alpha_M|, \tag{14}$$

where $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_M)$ and $|\alpha_1, \ldots, \alpha_M\rangle = \bigotimes_{i=1}^{M} |\alpha_i\rangle$. Photon losses then take the familiar form

$$\Lambda_{\boldsymbol{\gamma}}^{(M)}[\rho] = \int_{\mathbb{C}^M} d\boldsymbol{\alpha} P_\rho(\boldsymbol{\alpha}) |\sqrt{1-\gamma_1}\alpha_1\rangle\langle\sqrt{1-\gamma_1}\alpha_1|$$

$$\otimes \ldots \otimes |\sqrt{1-\gamma_M}\alpha_M\rangle\langle\sqrt{1-\gamma_M}\alpha_M|. \tag{15}$$

It can be seen that uniform losses commute with linear optical elements [46]. In this case, losses occurring at any point in

the dynamics (including uniform detector inefficiencies) can be "shifted" to the front of the circuit. Therefore, a noisy passive linear optical network in which each path is subject to approximately the same level of loss is well modeled by $\mathcal{U}_{\text{noisy}} = \mathcal{U}_{\text{ideal}} \circ \Lambda_{\boldsymbol{\gamma}}^{(M)}$.

In Ref. [47], Clements *et al.* provided an optimal decomposition of any M-mode unitary $U \in SU(2M)$ using only passive linear optical components. Their decomposition is *balanced* because each path within the interferometer passes through the same number of beam splitters. This in turn implies that photon loses are well modeled as being uniform, thus matching our noise model.

Nonuniform (i.e., path-dependent) losses cannot be perfectly separated from the ideal passive linear optical dynamics. However, the authors of Ref. [48] showed that *some* losses can still be shifted to the front of the circuit, leaving behind a *new* noisy passive linear optical network (described by $\mathcal{U}'_{\text{noisy}}$) with less noise, namely $\mathcal{U}_{\text{noisy}} = \mathcal{U}'_{\text{noisy}} \circ \Lambda_{\boldsymbol{\gamma}'}^{(M)}$. In principle, one could then use our protocol to partially mitigate the effect of losses, i.e., those coming from $\Lambda_{\boldsymbol{\gamma}'}^{(M)}$. However, we point out that determining the individual loss parameters $(\gamma_1, \gamma_2, \ldots, \gamma_M)$ and the exact form of the new $\mathcal{U}'_{\text{noisy}}$ may be challenging.

In realistic settings, one would expect the loss parameter to fluctuate slightly between shots. This can be modeled by selecting γ from a Gaussian distribution centered on $\bar{\boldsymbol{\gamma}} := \mathbb{E}[\gamma]$ with standard deviation $\boldsymbol{\sigma}_{\boldsymbol{\gamma}}$. Over longer time periods, the average loss parameter $\bar{\boldsymbol{\gamma}}$ could vary and hence one should periodically remeasure the loss parameter. In practice however, the high shot rate associated with photonic platforms means we do not expect $\bar{\boldsymbol{\gamma}}$ to appreciably vary over the timescale of a single experimental run.

While our discussion is focused on photonic platforms, our results are applicable to generic bosonic systems described by the canonical commutation relations $[\hat{a}_i, \hat{a}_j^\dagger] = \delta_{ij}\mathbb{I}$. Consequently, our mitigation protocol can be applied in a diverse range of bosonic architectures.

Learning the loss parameter

The average loss parameter can be estimated by inputting a multimode coherent state $|\boldsymbol{\alpha}\rangle = |\alpha_1, \alpha_2, \ldots, \alpha_M\rangle$ and measuring the output intensity, under the assumption of uniform losses, noisy dynamics that can be decomposed as in Fig. 1, and passive ideal dynamics. The input intensity is proportional to the expected number of input photons, $\langle\hat{N}\rangle_{\text{in}} = \sum_{i=1}^{M} |\alpha_i|^2$, where $\hat{N} = \sum_{i=1}^{M} \hat{n}_i$. The multimode photon loss channel transforms these coherent states into $|\boldsymbol{\alpha}\rangle \xrightarrow{\Lambda_\gamma} |\sqrt{1-\gamma}\alpha_1, \sqrt{1-\gamma}\alpha_2, \ldots, \sqrt{1-\gamma}\alpha_M\rangle$. Via energy conservation, the expected output number of photons will therefore be given by $\langle\hat{N}\rangle_{\text{out}} = \text{Tr}[\hat{N}\mathcal{U}_{\text{ideal}} \circ \Lambda_\gamma[|\boldsymbol{\alpha}\rangle\langle\boldsymbol{\alpha}|]] = \sum_{i=1}^{M} |\sqrt{1-\gamma}\alpha_i|^2 = (1-\gamma)\langle\hat{N}\rangle_{\text{in}}$. Denoting our estimate of γ as $\tilde{\gamma}$, we have

$$\tilde{\gamma} = 1 - \frac{\langle\hat{N}\rangle_{\text{out}}}{\langle\hat{N}\rangle_{\text{in}}} = 1 - \frac{I_{\text{out}}}{I_{\text{in}}} \tag{16}$$

where $I_{\text{in/out}}$ is the average total intensity summed across all modes going in and out of the circuit. Due to the fact that coherent states are well approximated by laser fields, which

are among the most stable states one can prepare in quantum optical experiments, this reduces the impact of state preparation errors on our estimate of γ. State preparation errors can be further alleviated by varying across a variety of different choices $|\alpha\rangle$ and taking the average estimated loss parameter. Inefficient detectors in quantum optics are modeled as a photon loss occurring immediately before detection, and can there be incorporated into our estimate of γ (and hence mitigated against by our error cancellation protocol).

If $\mathcal{U}_{\text{ideal}}$ is a passive linear optical network, then the output state will be the multimode coherent state $|\boldsymbol{\beta}\rangle = |\beta_1, \beta_2, \ldots, \beta_M\rangle$. When estimating the output intensity by making N_{shots} measurements, the variance of $\tilde{\gamma}$ will be

$$\text{Var}[\tilde{\gamma}] = \frac{1}{N_{\text{shots}}}\text{Var}\left[1 - \frac{I_{\text{out}}}{I_{\text{in}}}\right] = \frac{1}{N_{\text{shots}}}\text{Var}\left[\frac{\langle\hat{N}\rangle_{\text{out}}}{\langle\hat{N}\rangle_{\text{in}}}\right]$$

$$= \frac{1}{N_{\text{shots}}\langle\hat{N}\rangle_{\text{in}}^2}\text{Var}\sum_{i=1}^{M}\langle n_i\rangle_{\text{out}} = \frac{(1-\gamma)^2}{N_{\text{shots}}\sum_i|\alpha_i|^2}. \quad (17)$$

One can reduce this variance by increasing the number of shots or increasing the total intensity of the initial coherent state $|\alpha\rangle$. This also provides guarantees about the level of accuracy via Chebyshev's inequality; for example, to have a >99% probability of estimating γ to within an accuracy <0.01 when $\sum_{i=1}^{M}|\alpha_i|^2 = M$ requires only $N_{\text{shots}} > (1 \times 10^6)/M$. Given that coherent states can be easily produced, this number of shots can be achieved in less than 1 s for many optical platforms.

Alternatively, for passive linear optical networks one could incorporate learning the losses with learning the exact ideal unitary being implemented by using one of the methods proposed in Refs. [49,50]. This has the advantage of also giving information about exactly what $\mathcal{U}_{\text{ideal}}$ is, but is a more involved process.

III. QUANTUM ERROR CANCELLATION

In this section, our proposed quantum error cancellation scheme is presented for a single mode.

A. Inverse loss map

In this section, we construct the inverse Λ_γ^{-1} to the pure loss map. In particular, Eq. (9) implies that the inverse photon loss map, expressed using the P representation, must read

$$\Lambda_\gamma^{-1}[\rho] = \int_{\mathbb{C}} d^2\alpha\, P_\rho(\alpha)|g_0\alpha\rangle\langle g_0\alpha|, \quad (18)$$

where $g_0 = 1/\sqrt{1-\gamma} > 1$. This immediately leads to $\Lambda_\gamma[\Lambda_\gamma^{-1}[\rho]] = \Lambda_\gamma^{-1}[\Lambda_\gamma[\rho]] = \rho$. In other words, Λ_γ^{-1} effectively implements a noiseless amplification on every constituent coherent state in the P representation [51] *without changing their relative weights*.

An operator-sum representation of this inverse loss map is given by

$$\Lambda_\gamma^{-1}[\rho] = \sum_{j=0}^{\infty}\frac{(-\gamma)^j}{j!}\,\hat{a}^j g_0^{\hat{n}}\rho\, g_0^{\hat{n}}(\hat{a}^\dagger)^j. \quad (19)$$

The equivalence between Eqs. (18) and (19) is shown in Appendix B. It can be easily seen that Λ_γ^{-1} is Hermitian preserving and trace preserving, but not completely positive [as evident by the $(-\gamma)^j$ term], and hence is nonphysical. Nevertheless, Eq. (19) may be expressed as a linear combination of CP maps $\{\mathcal{E}_j\}$, namely

$$\Lambda_\gamma^{-1}[\rho] = \sum_{j=0}^{\infty}\omega_j(\rho)\mathcal{E}_j[\rho], \quad (20)$$

where

$$\mathcal{E}_j[\rho] = \frac{1}{\mathcal{N}_j(\rho)}\hat{a}^j g_0^{\hat{n}}\rho\, g_0^{\hat{n}}(\hat{a}^\dagger)^j, \quad (21)$$

with

$$\mathcal{N}_j(\rho) = \text{Tr}[\hat{a}^j g_0^{\hat{n}}\rho\, g_0^{\hat{n}}(\hat{a}^\dagger)^j], \quad (22)$$

and the (possibly negative) coefficients are given by

$$\omega_j(\rho) = \frac{(-\gamma)^j}{j!}\mathcal{N}_j(\rho). \quad (23)$$

Note that while these maps are not trace preserving, we will use a slight abuse of notation and refer to them as channels.

These channels can be, at least *probabilistically*, physically realized. Subject to the noise model assumptions discussed in the previous section, we conclude that any ideal system evolution may be decomposed into a weighted sum of noisy, physically realizable channels, i.e.,

$$\mathcal{U}_{\text{ideal}}[\rho] = \sum_j \omega_j(\rho)(\mathcal{U}_{\text{noisy}}\circ\mathcal{E}_j)[\rho]. \quad (24)$$

By making the identification $\mathcal{F}_j = (\mathcal{U}_{\text{noisy}}\circ\mathcal{E}_j)$, it can be seen that this is a quasiprobability representation of the ideal dynamics Eq. (3).

We point out that for certain states Eq. (19) can be rewritten to contain only a finite number of terms. For example, for states containing at most $K-1$ photons, it is clear that only the first K terms of the operator-sum representation of Λ_γ^{-1} act nontrivially. Similarly, if $\rho = |\psi\rangle\langle\psi|$ and $g^{\hat{n}}|\psi\rangle$ is an eigenstate of $\hat{a}^K$, then Eq. (19) can be rewritten such that it contains only K terms.

Intuition behind the inverse photon loss map

We now aim to briefly provide some physical intuition behind the—somewhat unexpected—presence of photon-subtracting terms $\hat{a}^j$ in the inverse of the loss map Eq. (19). For simplicity, let us consider the photon loss map and its inverse acting on the Fock state $|m\rangle\langle m|$, i.e.,

$$(\Lambda_\gamma\circ\Lambda_\gamma^{-1})[|m\rangle\langle m|] = \sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\frac{1}{i!}\frac{1}{j!}\left(\frac{\gamma}{1-\gamma}\right)^i\left(\frac{-\gamma}{1-\gamma}\right)^j$$

$$\times\,\hat{a}^{i+j}|m\rangle\langle m|(\hat{a}^\dagger)^{i+j}. \quad (25)$$

We can see that the noiseless amplification term $g^{\hat{n}}$ appearing in Λ_γ^{-1} cancels out with the $g^{-\hat{n}}$ term present in the loss map expansion, leaving behind a series of photon subtraction terms only. Using the fact that at most m photons may be subtracted

from $|m\rangle\langle m|$, one can show that Eq. (25) can be expressed as

$$\sum_{n=0}^{m}\left[\sum_{j=0}^{m-n}(-1)^{j}\binom{m-n}{j}\right]\binom{m}{n}\left(\frac{\gamma}{1-\gamma}\right)^{m-n}|n\rangle\langle n|. \quad (26)$$

For $n < m$, the alternating sign causes the coefficient inside the square brackets to vanish. We remind the reader that the alternating sign originates from the operator-sum representation of the inverse loss map. On the other hand, for $n = m$, said coefficient is equal to 1 and hence $(\Lambda_{\gamma}\circ\Lambda_{\gamma}^{-1})[|m\rangle\langle m|] = |m\rangle\langle m|$ as expected.

Physically, we can understand these cancellations in terms of quantum trajectories. If we consider just the photon loss map, then there is clearly only one trajectory in which the input Fock state loses k photons (i.e., via the action of the Kraus operator K_{k}). On the other hand, we can see from Eq. (25) that the presence of the inverse loss maps allows the existence of *multiple* trajectories in which the system loses $k > 0$ photons, namely every i and j in Eq. (25) such that $i + j = k$. The alternating sign in Eq. (19) leads to the cancellation of these trajectories. Conversely, there is only one trajectory in which the system does not lose any photons (i.e., $i + j = 0$), hence in this case the cancellation cannot take place and the original noise-free state is returned.

B. Quantum error cancellation protocol

Our mitigation scheme places all interventions before the state undergoes its noisy dynamics $\mathcal{U}_{\text{ideal}}\circ\Lambda_{\gamma}$.

To realize the channels $\{\mathcal{E}_{j}\}$, one needs to implement an *amplification* operation, $g^{\hat{n}}$, and a j-photon subtraction, $\hat{a}^{j}$. The operator $g^{\hat{n}}$ is said to implement a (noiseless) amplification operation because, up to normalization, it has the effect of amplifying a coherent state's amplitude by a factor of $g > 1$. The standard implementation of photon subtraction requires sending the input state through a beam splitter which further attenuates the state [52]. However, we can compensate for this spurious effect in the amplification step by considering a gain factor $g_{\mu} > g_{0}$ that depends on the beam splitter transmissivity. In the following sections we discuss these aspects in detail.

1. Amplification

Currently available protocols for noiseless amplification are notoriously challenging to implement [51,53–55]. This is primarily due to their inherent probabilistic nature and to their demanding resources requirements, such as multiple single-photon Fock states, that limits scalability to many modes.

One approach by Zavatta *et al.* [53] achieves noiseless amplification in the Hilbert subspace spanned by the Fock states $\{|0\rangle, |1\rangle\}$ by using a nonlinear optical element to probabilistically implement the operator $\mathbb{I} + c\hat{n}$. Looping this scheme N times with different constants c_{i} leads to the implementation of the operator $\prod_{i=1}^{N}(\mathbb{I} + c_{i}\hat{n})$. By carefully tuning the parameters c_{i}, one can implement exact noiseless amplification within the subspace spanned by $\{|0\rangle, |1\rangle, \ldots, |N\rangle\}$. Alternatively, if the amplification factor g_{μ} is close to 1, one could use this same approach to implement the truncated Taylor series, $g_{\mu}^{\hat{n}} \approx \sum_{i=0}^{N}\ln(g_{\mu})^{i}\hat{n}^{i}/i!$. Similarly, the teleamplification scheme by Guanzon *et al.* [55]

is a generalized quantum scissor operation that transforms a infinite-dimensional state $|\psi\rangle = \sum_{n=0}^{\infty} b_{n}|n\rangle$ into a finite-dimensional state $|\psi\rangle \rightarrow \sum_{n=0}^{N} g^{n}b_{n}|n\rangle$ in the Hilbert space spanned by $\{|0\rangle, |1\rangle, \ldots, |N\rangle\}$. While this technique does not require nonlinear optical elements and has a larger success probability than Zavatta *et al.*'s scheme (and in fact asymptotically saturates the maximal possible success probability when amplifying in the two-dimensional subspace, see Ref. [56]), it requires a resource state of many single-photon Fock states.

If the amplification operation takes place within the circuit, it could then be advantageous to first do the photon subtraction as the latter operation could force a larger fraction of the state ρ_{0} into the subspace spanned by $\{|0\rangle, \ldots, |N\rangle\}$ which would lead to the amplification being applied with a better fidelity. One can easily account for this new ordering of the operations with the commutation relation in Eq. (A1).

We remark that the probabilistic nature of any noiseless amplification protocol would dramatically increase the sampling overhead of our scheme. One way around this involves conducting the amplification offline and storing the successfully amplified state until it is required.

Alternatively, for many experimentally relevant classes of states, one can avoid the burden of actually implementing the amplification step, provided that the protocol's input state is the following amplified version of ρ_{0}, namely

$$\rho_{\text{amp}} = \frac{1}{\mathcal{N}_{\text{amp}}(\rho_{0})}g_{\mu}^{\hat{n}}\rho_{0}g_{\mu}^{\hat{n}}, \quad (27)$$

where $\mathcal{N}_{\text{amp}}(\rho_{0}) = \text{Tr}[g_{\mu}^{\hat{n}}\rho_{0}g_{\mu}^{\hat{n}}]$ is a normalizing factor. In particular, if the amplified state ρ_{amp} belongs to the same class of states as ρ_{0}, then noiseless amplification simply corresponds to tuning the parameter that characterizes such class. For example, if ρ_{0} is a cat state of magnitude α_{0}, then ρ_{amp} is a cat state of magnitude $\alpha_{\text{amp}} = g_{\mu}\alpha_{0} > \alpha_{0}$. Similarly, noiselessly amplifying a squeezed vacuum state is equivalent to increasing its squeezing parameter.

In this setting, we may also change perspective and, instead of using ρ_{0} to define $\rho_{\text{amp}} \sim g_{\mu}^{\hat{n}}\rho_{0}g_{\mu}^{\hat{n}}$, we start from the amplified state ρ_{amp} and use it to determine $\rho_{0} \sim g_{\mu}^{-\hat{n}}\rho_{\text{amp}}g_{\mu}^{-\hat{n}}$, i.e., the state whose noise-free expectation value we are estimating. This approach is particularly helpful when experimental limitations constrain the range of easily accessible states, as illustrated in the following example. Suppose that the most energetic coherent state we have experimental access to is $|\alpha_{\text{max}}\rangle$. It follows that, using $\rho_{\text{amp}} = |\alpha_{\text{max}}\rangle\langle\alpha_{\text{max}}|$ as an initial state, we can estimate noise-free expectation values associated with the state $|\alpha_{\text{max}}/g_{\mu}\rangle$.

2. Photon subtraction

Photon subtraction can be probabilistically implemented using a beam splitter and photon number resolving detector [52]. Consider a beam splitter of transmissivity $1 - \mu$, its unitary operation defined by $U_{\text{BS},\mu} = e^{\arccos(\sqrt{1-\mu})(\hat{a}_{A}^{\dagger}\hat{a} - \hat{a}^{\dagger}\hat{a}_{A})}$, where the A subscript denotes the ancillary mode. If ρ_{amp} is sent through the beam splitter together with a vacuum state in the second input port, then a measurement of j photons in the output port of the ancillary mode will herald the state (details
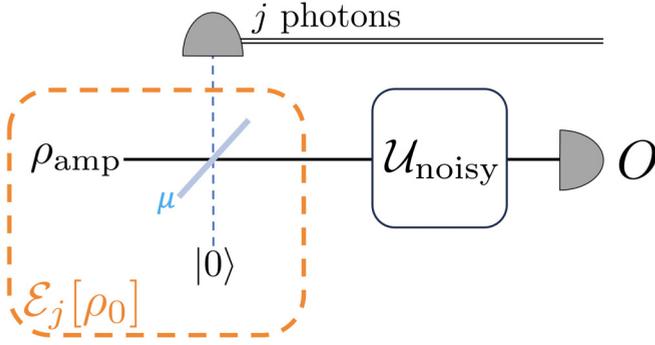
FIG. 2. Schematic diagram depicting the full circuit used in the single mode mitigation scheme as described in Sec. III B. If one wants to estimate the noise-free expectation value $\langle O \rangle_{\text{ideal}} = \text{Tr}[O \mathcal{U}_{\text{ideal}}[\rho_0]]$, one first prepares the amplified state $\rho_{\text{amp}} \sim g_\mu^{\hat{n}} \rho_0 g_\mu^{\hat{n}}$. This state is sent through the depicted beam splitter with a vacuum in the ancilla input port. A measurement of j photons from the ancilla output port will herald the state $\mathcal{E}_j[\rho_0]$—this is as if we physically started with ρ_0 and implemented the CP channel $\mathcal{E}_j$. Which channel is implemented in each individual shot is determined by the number of detected photons j. Following this, the noisy dynamics $\mathcal{U}_{\text{noisy}}$ are run and a measurement of O is made. This circuit is run many times to obtain estimates of $\langle O \rangle_{j,\text{noisy}} = \text{Tr}[O (\mathcal{U}_{\text{noisy}} \circ \mathcal{E}_j)[\rho_0]]$ for different values of j. These results are then linearly combined with the relevant coefficients $\omega_j(\rho_0)$ to give the mitigated estimator $\bar{O}_{\text{mit}}$. All interventions take place before the state undergoes its noisy dynamics and so it is not necessary to reconfigure or change the circuit between experimental shots.

can be found in Appendix C)

$$\rho_{\text{amp}} \rightarrow \frac{1}{p_j} K_j(\mu) \rho_{\text{amp}} K_j^\dagger(\mu), \tag{28}$$

where $K_j(\mu)$ are the Kraus operators defined in Eq. (11), and $p_j = \text{Tr}[K_j(\mu) \rho_{\text{amp}} K_j^\dagger(\mu)]$ is the probability of having measured j photons. Finally, comparing Eq. (28) with Eq. (21) reveals that, for the heralded state to be equal to $\mathcal{E}_j[\rho_0]$, the amplification factor must be set to $g_\mu = 1/\sqrt{(1-\gamma)(1-\mu)}$. Hence, assuming that we start our protocol with the amplified state ρ_{amp}, the probabilistic nature of the photon subtraction operation naturally allows us to sample from the set of states $\{\mathcal{E}_j[\rho_0]\}$ according to the probability distribution $\{p_j\}$. This is depicted graphically in Fig. 2.

3. Postprocessing

Since the implementation of the maps $\{\mathcal{E}_j\}$ is inherently probabilistic, Monte Carlo methods cannot be straightforwardly used to estimate $\langle O \rangle_{\text{ideal}}$. Nevertheless, the linearity of the expectation value lets us write

$$\langle O \rangle_{\text{ideal}} = \sum_{j=0}^{\infty} \omega_j(\rho_0) \langle O \rangle_{j,\text{noisy}}, \tag{29}$$

where $\langle O \rangle_{j,\text{noisy}} = \text{Tr}[O (\mathcal{U}_{\text{noisy}} \circ \mathcal{E}_j)[\rho_0]]$ is the noisy expectation value of O given that the jth channel was implemented. In other words, a suitable linear combination of the noisy expectation values $\{\langle O \rangle_{j,\text{noisy}}\}$ allows us to estimate the noise-free expectation value $\langle O \rangle_{\text{ideal}}$. In N runs of the circuit depicted in Fig. 2, we collect N measurement outcomes

$\{O_\ell^{(j_\ell)}\}_{\ell=1}^N$, where the superscript j_ℓ denotes how many photons were measured by the photon number resolving detector (i.e., which particular channel $\mathcal{E}_{j_\ell}$ was implemented) on the ℓth run. The noisy expectation value $\langle O \rangle_{k,\text{noisy}}$ is estimated via the sample average

$$\bar{O}_{k,\text{noisy}} = \frac{1}{N_k} \sum_{\ell | j_\ell = k} O_\ell^{(j_\ell)}, \tag{30}$$

where N_k is the number of times $\mathcal{E}_k$ was implemented. Notice that the probability of subtracting $j + 1$ photons (if possible) is typically a factor of $\mathcal{O}(\frac{\mu}{(j+1)(1-\mu)})$ smaller than subtracting j photons. We will therefore assume to have a set of noisy expectation values of increasing j up to some maximum cutoff, J_{max}, i.e., we have $\{\bar{O}_{j,\text{noisy}}\}_{j=0}^{J_{\text{max}}}$. We then linearly combine these expectations values to obtain the mitigated estimator of $\langle O \rangle_{\text{ideal}}$, namely

$$\bar{O}_{\text{mit}} = \sum_{j=0}^{J_{\text{max}}} \omega_j(\rho_0) \bar{O}_{j,\text{noisy}}, \tag{31}$$

which has an expected value of

$$\mathbb{E}[\bar{O}_{\text{mit}}] = \sum_{j=0}^{J_{\text{max}}} \omega_j(\rho_0) \langle O \rangle_{j,\text{noisy}}. \tag{32}$$

We may formally express the previous equation as $\mathbb{E}[\bar{O}_{\text{mit}}] = \text{Tr}[O \rho_{\text{mit}}]$, where

$$\rho_{\text{mit}} = \sum_{j=0}^{J_{\text{max}}} \omega_j(\rho_0) (\mathcal{U}_{\text{noisy}} \circ \mathcal{E}_j)[\rho_0] \tag{33}$$

is a *pseudostate*, as it may not be a valid density matrix. Nevertheless, the fidelity between ρ_{mit} and $\mathcal{U}_{\text{ideal}}[\rho_0]$ quantifies the ability of our error mitigation protocol to recover the noise-free expectation value for generic observables [57].

C. Error analysis

The bias and variance of the mitigated estimator Eq. (31) for a particular choice of J_{max} are respectively given by

$$\text{Bias}[\bar{O}_{\text{mit}}; J_{\text{max}}] = \left| \sum_{j=J_{\text{max}}+1}^{\infty} \omega_j(\rho_0) \langle O \rangle_{j,\text{noisy}} \right|, \tag{34}$$

$$\text{Var}[\bar{O}_{\text{mit}}; J_{\text{max}}] = \sum_{j=0}^{J_{\text{max}}} \frac{|\omega_j(\rho_0)|^2}{N_j} \text{Var}[O_{j,\text{noisy}}], \tag{35}$$

where $\text{Bias}[\bar{O}_{\text{mit}}; J_{\text{max}}] := |\mathbb{E}[\bar{O}_{\text{mit}|J_{\text{max}}}] - \langle O \rangle_{\text{ideal}}|$ and $\text{Var}[O_{j,\text{noisy}}] = \langle O^2 \rangle_{j,\text{noisy}} - \langle O \rangle_{j,\text{noisy}}^2$.

We point out that both the bias and variance fundamentally arise from the finite number of samples considered and therefore vanish in the infinite sampling limit. For a fixed (and suitably large) choice of N, the value of J_{max} can be increased by increasing μ. As J_{max} increases, the bias tends to decrease while the variance (and hence sampling overhead) monotonically increases—this accuracy-overhead tradeoff is characteristic of many QEM schemes.

In practice, J_{max} is implicitly defined as the smallest integer J that satisfies $\text{Bias}[\bar{O}_{\text{mit}}; J] < \vartheta_{\text{bias}}$, where ϑ_{bias} is the

chosen acceptable level of bias that we want to achieve. As the channel $\mathcal{E}_{J_{\max}}$ is implemented with probability $p_{J_{\max}}$, it follows that the number of samples required to achieve ϑ_{bias} scales as $\mathcal{O}(p_{J_{\max}}^{-1})$. Once $J_{\max}$ is fixed, any measurement result corresponding to the kth channel for $k > J_{\max}$ should be discarded because these results will unnecessarily increase the shot noise.

It is clear that in experimental settings Eq. (34) cannot be used to estimate the bias, since its expression explicitly depends on unknown noisy expectation values. Nonetheless, for nonzero $\langle O \rangle_{\text{ideal}}$ the *fractional* bias (or equivalently percentage bias), $\text{Bias}[\bar{O}_{\text{mit}}; J]/\langle O \rangle_{\text{ideal}}$, may still be approximately estimated if we can assume that $\langle O \rangle_{j,\text{noisy}}/\langle O \rangle_{\text{ideal}} \approx \mathcal{O}(1)$. With this simplification, the fractional bias reads

$$\frac{\text{Bias}[\bar{O}_{\text{mit}}; J]}{\langle O \rangle_{\text{ideal}}} \approx \left| \sum_{j=J_{\max}+1}^{\infty} \omega_j(\rho_0) \right|, \quad (36)$$

and we can therefore estimate the $J_{\max}$ necessary to achieve a target bias using the coefficients $\{\omega_j(\rho_0)\}$ only. Finally, assuming $N_j \approx p_j N$, the sampling overhead reads

$$\text{sampling overhead} \approx \sum_{j=0}^{J_{\max}} \frac{|\omega_j(\rho_0)|^2}{p_j} \frac{\text{Var}[O_{j,\text{noisy}}]}{\text{Var}[O_{\text{noisy}}]}. \quad (37)$$

This can also be estimated using only $\{\omega_j\}$ and $\{p_j\}$ under the assumption that $\text{Var}[O_{j,\text{noisy}}]/\text{Var}[O_{\text{noisy}}] \approx \mathcal{O}(1)$.

D. Monte Carlo approach

For specific classes of input states ρ_0 it may be possible to efficiently prepare each state $\mathcal{E}_j[\rho_0]$ appearing in the inverse noise channel decomposition Eq. (20), without having to directly implement the amplification and the photon subtraction (probabilistic) operations. In this scenario, our protocol can be implemented using the standard Monte Carlo technique. Explicitly, in each experimental shot one selects an initial state from the ensemble $\{\mathcal{E}_j[\rho_0]\}$ with probability $\{q_j = |\omega_j|/S\}$ where $S = \sum_{j=0}^{\infty} |\omega_j|$. The measurement outcome in each shot is then multiplied $S\,\text{sgn}(\omega_j)$ before being averaged over. This provides an unbiased estimate of $\langle O \rangle_{\text{ideal}}$ because there is no finite cutoff acting on the operator-sum representation of the inverse loss channel in Eq. (20). This Monte Carlo approach can then be easily generalized to the multimode setting (see Sec. IV D for details).

In Sec. III E 2, we provide a detailed example of the Monte Carlo approach for the two-component cat state—for these states, Eq. (20) can be rewritten in such a way that it contains only two terms and hence we just have to sample from two different initial states, $\{\mathcal{E}_0[\rho_0], \mathcal{E}_1[\rho_0]\}$.

Another notable example is given by the single-photon Fock state $\rho_0 = |1\rangle\langle 1|$. In this case, one easily shows that $\Lambda_\gamma^{-1}[\rho_0] = \frac{1}{1-\gamma}|1\rangle\langle 1| - \frac{\gamma}{1-\gamma}|0\rangle\langle 0|$. Hence, one can implement error cancellation by sending into the noisy circuit $|1\rangle$ ($|0\rangle$) with probability $1/(1-\gamma)$ [$\gamma/(1-\gamma)$] and applying a postprocessing weight of $S = \frac{1+\gamma}{1-\gamma}$ ($-S$) to the measurement outcome.

E. Examples

In this section, we provide explicit examples of our error cancellation protocol to showcase its capabilities and assess its performance. In particular, we will consider the following classes of input states: single mode squeezed vacuum states and two-component cat states.

1. Squeezed states

We first consider the input state to be a single mode squeezed vacuum state, i.e., $\rho_0 = |S(r_0)\rangle\langle S(r_0)|$, defined as

$$|S(r_0)\rangle = S(r_0)\,|0\rangle$$
$$= \frac{1}{\sqrt{\cosh(r_0)}} \sum_{n=0}^{\infty} \left(\frac{\tanh(r)}{2}\right)^n \frac{\sqrt{(2n)!}}{n!} |2n\rangle. \quad (38)$$

Here, $r_0 \geq 0$ is the squeezing parameter and $S(r_0) = e^{\frac{r_0}{2}(\hat{a}^2 - \hat{a}^{\dagger 2})}$ is the single mode squeezing operator. As previously mentioned, noiselessly amplifying $|S(r_0)\rangle$ by a factor of $g > 1$ produces a new squeezed vacuum state $|S(r_{\text{amp}})\rangle \propto g^{\hat{n}} S(r_0)\,|0\rangle$, where $\tanh(r_{\text{amp}}) = g^2 \tanh(r_0)$. Notice that r_{amp} is not well defined when $g^2 \tanh(r_0) \geq 1$. If this is the case then it is not physically meaningful to exactly amplify $|S(r_0)\rangle$ by a factor g [58]. Therefore, in order to exactly amplify by increasing the squeezing of the initial state, one has to either decrease μ or redefine r_0 to $r_0 \to \tilde{r}_0 < r_0$, both of which come with potential consequences. Alternatively, one could amplify by g within some subspace, as discussed in Sec. III B 1. For example, in Ref. [55], the authors show that their noiseless linear teleamplification scheme is able to approximately amplify single mode squeezed vacua beyond this $g^2 \tanh(r_0) < 1$ limit.

To assess the performance of our mitigation protocol, we consider the scenario where the ideal unitary evolution $\mathcal{U}_{\text{ideal}}$ is the identity and the measured observable corresponds to a projector onto the (pure) noiseless reference state, ρ_0, i.e., we consider the observable $O = \rho_0$. Clearly, the expectation value of ρ_0 with respect to an arbitrary state σ corresponds to their quantum fidelity, namely $\text{Tr}[\rho_0 \sigma] = F(\rho_0, \sigma)$ [5,57]. Hence, in this scenario, the mitigated expectation value corresponds to the fidelity $F(\rho_0, \rho_{\text{mit}})$ between the target state ρ_0 and the pseudostate ρ_{mit} defined in Eq. (33). Intuitively, we expect that if the state σ has a large fidelity with ρ_0, then $\text{Tr}[A\sigma]$ provides a good estimate of $\text{Tr}[A\rho_0]$ for most observables A. A value of $F(\rho_0, \rho_{\text{mit}})$ close to 1 then implies that (for most choices of A and generic unitary dynamics $\mathcal{U}$) $\text{Tr}[A\,\mathcal{U}[\rho_{\text{mit}}]]$ will be a good estimator of $\text{Tr}[A\,\mathcal{U}[\rho_0]]$. We therefore can use $F(\rho_0, \rho_{\text{mit}})$ as a metric of the mitigation scheme's performance, independent of any particular dynamics or observables.

In Fig. 3, we plot the exact percentage bias of $\bar{O}_{\text{mit}}$ with $O = |S(r_0)\rangle\langle S(r_0)|$ as a function of $J_{\max}$, for different values of loss, γ, and squeezing, r_0. We see significant improvements with respect to the unmitigated results (horizontal lines), indicating that the pseudostate ρ_{mit} has a higher fidelity with ρ_0 than $\Lambda_\gamma[\rho_0]$. For example, to ensure $\vartheta_{\text{bias}} < 10^{-3}$ (corresponding to a percentage bias smaller than 0.1%) with $\gamma = 0.1$ and $r_0 = 1.0$ only requires $J_{\max} = 3$. Even when $J_{\max} = 0$, which corresponds to performing a zero photon subtraction
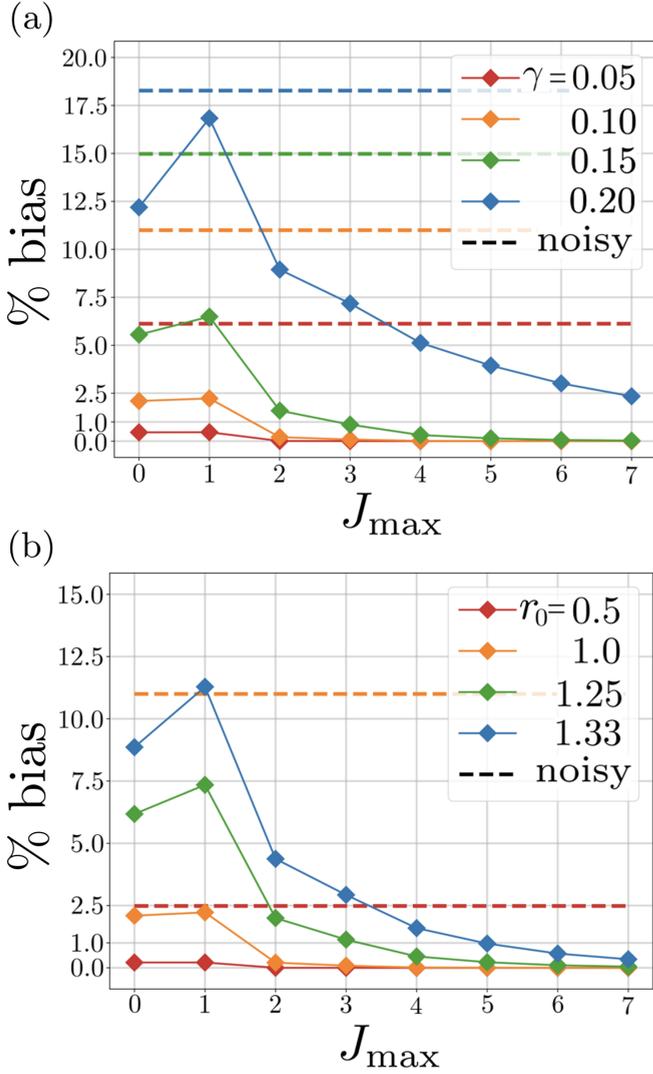
(a)

(b)

FIG. 3. Percentage bias as a function of J_{max} for different values of r_0 and γ, for the estimation of the fidelity with respect to the noiseless target state. The dashed horizontal lines represent the percentage bias of the *unmitigated* estimators. In (a), we vary γ at fixed $r_0 = 1$. On the other hand, in (b) we fix $\gamma = 0.1$ and consider different values of r_0. The unmitigated percentage biases for $r_0 = 1$ and 1.33 lie outside the graph, but are respectively given by 18.0 and 20.6%. In all displayed cases, we see an improvement over the unmitigated result, even for modest values of J_{max}.

[59] on the amplified state, we see an improvement in the bias for all plotted cases.

In Fig. 4, we plot the sampling overhead and r_{amp} as a function of μ for different values of r_0 and fixed γ. We choose the acceptable level of bias, ϑ_{bias}, to be 1%. For suitably small μ, we observe that increasing μ tends to decrease the sampling overhead. This is because it increases the probability that $j \neq 0$ photons are subtracted which in turn will tend to decrease the shot noise in $\bar{O}_{j\neq0,noisy}$. This comes at the cost of increasing the necessary level of squeezing, r_{amp}. Finding the optimal value of μ in the tradeoff between sampling overhead and r_{amp} will depend on experimental considerations such as

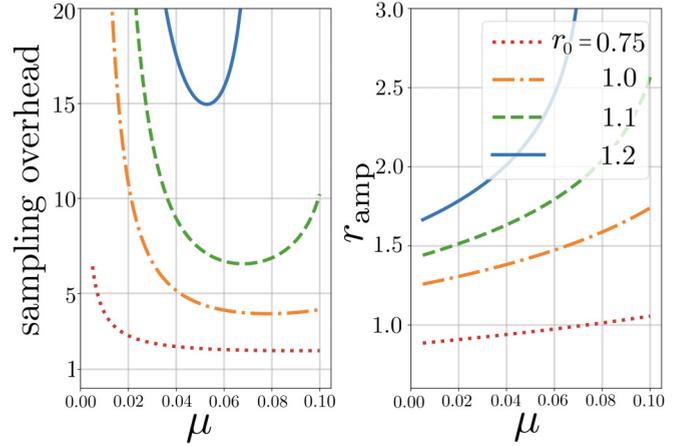FIG. 4. Sampling overhead and r_{amp} as a function of μ for varied r_0 with fixed $\gamma = 0.1$. For $r_0 = (0.75, 1.0, 1.1, 1.2)$, we choose $J_{max} = (1, 2, 3, 3)$ which gives the percentage biases as $(0.72\%, 0.21\%, 0.23\%, 0.65\%) < 1\%$. The corresponding *unmitigated* biases are $(5.86\%, 11.0\%, 13.58\%, 16.46\%)$. We observe a tradeoff between minimizing the sampling overhead and minimizing the required amount of squeezing. The optimal choice of μ will depend on the individual experimental details such as the difficulty in preparing highly squeezed states and circuit shot rate. We point out that we did not include the effect of discarding detection outcomes corresponding to $J > J_{max}$ in the computation of the sampling overhead and hence we are slightly underestimating its true value. Fixing r_0 and varying γ produces analogous looking behavior.

the difficulty of increasing the squeezing parameter and the circuit shot rate.

Finally, we investigate the stability of our error cancellation protocol in the presence of imperfect noise strength parameter estimation. In particular, we imagine that the experimenter has incorrectly estimated the loss parameter to be $\tilde{\gamma}$ and then study the effects that implementing the mitigation scheme using loss parameter $\tilde{\gamma} \neq \gamma$ has on the bias. In Fig. 5, we plot $\mathbb{E}[\bar{O}_{mit}] = F(\rho_0, \rho_{mit})$ as a function of $\tilde{\gamma}$ with $r_0 = 1$, $\gamma = 0.1$ and $J_{max} = 3$—for this choice of parameters, when $\tilde{\gamma} = \gamma$ the percentage bias is $<0.1\%$.

We see that the percentage bias then grows approximately linearly as $\tilde{\gamma}$ moves away from the true loss parameter, and that the gradient of this curve is favorable—for example, a 25% error in estimation of the noise parameter leads to only a 4% bias in the expected value of the mitigated estimator. In order for the mitigated bias to match the noisy expectation value bias (which for these parameters is around 11%), one would have to estimate $\tilde{\gamma} = 0$ or 0.16. Therefore, for these parameters as long as the experimental error in γ estimation is below 60%, the bias of the mitigated estimator outperforms the noisy results. Notably, for all $\tilde{\gamma} < \gamma$ our mitigation scheme outperforms the noisy result and therefore even partially undoing the losses is always better (from a bias perspective) than running the noisy dynamics.

This slow growth of the bias with respect to $|\gamma - \tilde{\gamma}|/\gamma$ constitutes a strong indication that our mitigation scheme is robust against imperfect characterization of the noise strength. This is because the impact of estimating $\tilde{\gamma} \neq \gamma$ presents itself
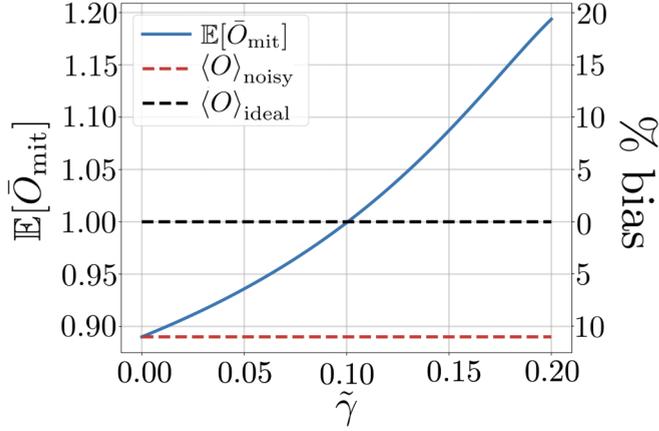
FIG. 5. $\mathbb{E}[\bar{O}_{\text{mit}}] = F(\rho_0, \rho_{\text{mit}})$ as a function of $\tilde{\gamma}$, with $\gamma = 0.1$, $r_0 = 1$, and $J_{\max} = 3$. These parameters gives a mitigated percentage bias of 0.1% when we have perfect knowledge of the loss parameter, i.e., $\tilde{\gamma} = \gamma$. We observe that our mitigation scheme is stable under imperfect knowledge of γ: for example, $\tilde{\gamma} = 0.075$ (a 25% error on the noise parameter characterization) only leads to a modest 4% bias of the mitigated estimator.

as a moderate bias in the end result, but this bias is relatively small for reasonable estimates of γ, such as within 10%.

2. Cat states

In this section, we provide an example of the Monte Carlo approach previously outlined in Sec. III D. We do so by considering two-component cat states $\rho_\phi(\alpha) = |\mathcal{C}_\phi(\alpha)\rangle\langle\mathcal{C}_\phi(\alpha)|$ as input states for our mitigation scheme. These are defined as $|\mathcal{C}_\phi(\alpha)\rangle = (|\alpha\rangle + e^{i\phi}|-\alpha\rangle)/\mathcal{A}_\phi(\alpha)$, where $\mathcal{A}_\phi(\alpha) = \sqrt{2(1 + \cos(\phi)e^{-2|\alpha|^2})}$ denotes the normalization coefficient. Since the two-component cat states are eigenstates of $\hat{a}^2$, it follows that we only need the channels $\mathcal{E}_0$ and $\mathcal{E}_1$ to fully describe the action of the inverse noise channel Λ_γ^{-1}. One can easily show that

$$\mathcal{E}_j[\rho_\phi(\alpha)] = \begin{cases} \mathcal{E}_0[\rho_\phi(\alpha)] & j \text{ even} \\ \mathcal{E}_1[\rho_\phi(\alpha)] & j \text{ odd} \end{cases} \tag{39}$$

which further implies that we can write

$$\Lambda_\gamma^{-1}[\rho_\phi(\alpha)] = W_{\text{even}}\rho_\phi(g_0\alpha) + W_{\text{odd}}\rho_{\phi+\pi}(g_0\alpha), \tag{40}$$

where the weights are given by

$$W_{\text{even}} = \sum_{j=0}^{\infty} \omega_{2j} = e^{\frac{\gamma|\alpha|^2}{1-\gamma}} \cosh\left(\frac{\gamma|\alpha|^2}{1-\gamma}\right)\left|\frac{\mathcal{A}_\phi(g_0\alpha)}{\mathcal{A}_\phi(\alpha)}\right|^2,$$

$$W_{\text{odd}} = \sum_{j=0}^{\infty} \omega_{2j+1} = -e^{\frac{\gamma|\alpha|^2}{1-\gamma}} \sinh\left(\frac{\gamma|\alpha|^2}{1-\gamma}\right)\left|\frac{\mathcal{A}_{\phi+\pi}(g_0\alpha)}{\mathcal{A}_\phi(\alpha)}\right|^2. \tag{41}$$

We remind the reader that $g_0 = 1/\sqrt{1-\gamma}$ and point out that W_{odd} is negative. If one is able to directly prepare the states $|\mathcal{C}_\phi(g_0\alpha)\rangle$ and $|\mathcal{C}_{\phi+\pi}(g_0\alpha)\rangle$, then the mitigation scheme may be implemented simply by sending these two states through the circuit with probabilities respectively given by $q_{\text{even}} =$

$|W_{\text{even}}|/S$ and $q_{\text{odd}} = |W_{\text{odd}}|/S$, where

$$S = |W_{\text{even}}| + |W_{\text{odd}}| = \frac{e^{\frac{2\gamma}{1-\gamma}|\alpha|^2} + \cos(\phi)e^{-\frac{2}{1-\gamma}|\alpha|^2}}{1 + \cos(\phi)e^{-2|\alpha|^2}}. \tag{42}$$

The mitigated estimator is then given by

$$\bar{O}_{\text{mit}} = S(\bar{O}_{\text{even}} - \bar{O}_{\text{odd}}). \tag{43}$$

The latter is unbiased because we did not include a cutoff in the operator-sum representation of Λ_γ^{-1}. The associated sampling overhead reads

$$\text{sampling overhead} \approx S\left(|W_{\text{even}}|\frac{\text{Var}[O_{\text{even}}]}{\text{Var}[O_{\text{noisy}}]}\right.$$

$$\left. + |W_{\text{odd}}|\frac{\text{Var}[O_{\text{odd}}]}{\text{Var}[O_{\text{noisy}}]}\right). \tag{44}$$

When $\text{Var}[O_{\text{even}}] \approx \text{Var}[O_{\text{odd}}] \approx \text{Var}[O_{\text{noisy}}]$, the sampling overhead is approximately S^2, exactly as in the DV PEC case. This scaling arises because the task has been reduced to a Monte Carlo estimation task. Notice that the leading term in Eq. (42) is exponential in $\gamma|\alpha|^2/(1-\gamma)$. In Fig. 6 we plot S^2 (i.e., the approximate sampling overhead) as a function of α and γ.

IV. EXTENSION TO MULTIPLE MODES

In this section, we extend our error cancellation protocol to M modes. Conceptually, the multimode protocol is exactly analogous to the single mode case. In order to highlight this similarity, we introduce two new operators corresponding to a multimode photon subtraction and a multimode amplification or attenuation:

$$\hat{\mathcal{A}}_{\boldsymbol{j}}^{(M)} := \bigotimes_{i=1}^{M} \hat{a}_i^{j_i}, \quad \hat{\mathcal{G}}_{\boldsymbol{g}}^{(M)} := \bigotimes_{i=1}^{M} g_i^{\hat{n}_i}. \tag{45}$$

Here, $\boldsymbol{j} = (j_1, \ldots, j_M)$ is a vector of photon subtractions and $\boldsymbol{g} = (g_1, \ldots, g_M)$ is a vector of the amplification factors g_i in the ith mode.

A. Inverse loss map

Our proposed mitigation scheme extends naturally from a single mode to multiple modes when the photon losses are of the form presented in Eq. (13). By comparison with the single mode case, one can see from the P representation of $\Lambda_{\boldsymbol{\gamma}}^{(M)}$ in Eq. (15) that the multimode inverse loss map is given by a tensor product of M single mode inverse loss maps with loss parameters $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, \ldots, \gamma_M)$:

$$\Lambda_{\boldsymbol{\gamma}}^{-1(M)} = \bigotimes_{i=1}^{M} \Lambda_{\gamma_i, i}^{-1}. \tag{46}$$

Once again, this map is Hermitian preserving and trace preserving but not completely positive and hence nonphysical. Nevertheless, these operators can be expressed as a sum over physically realizable operations with potentially negative coefficients. We consider the following set of maps $\{\mathcal{E}_{\boldsymbol{j}}^{(M)}\}$:

$$\mathcal{E}_{\boldsymbol{j}}^{(M)}[\rho] = \frac{1}{\tilde{\mathcal{N}}_{\boldsymbol{j}}(\rho)} \hat{\mathcal{A}}_{\boldsymbol{j}}^{(M)} \hat{\mathcal{G}}_{\boldsymbol{g_0}}^{(M)} \rho \, \hat{\mathcal{G}}_{\boldsymbol{g_0}}^{(M)} \hat{\mathcal{A}}_{\boldsymbol{j}}^{(M)\dagger}, \tag{47}$$
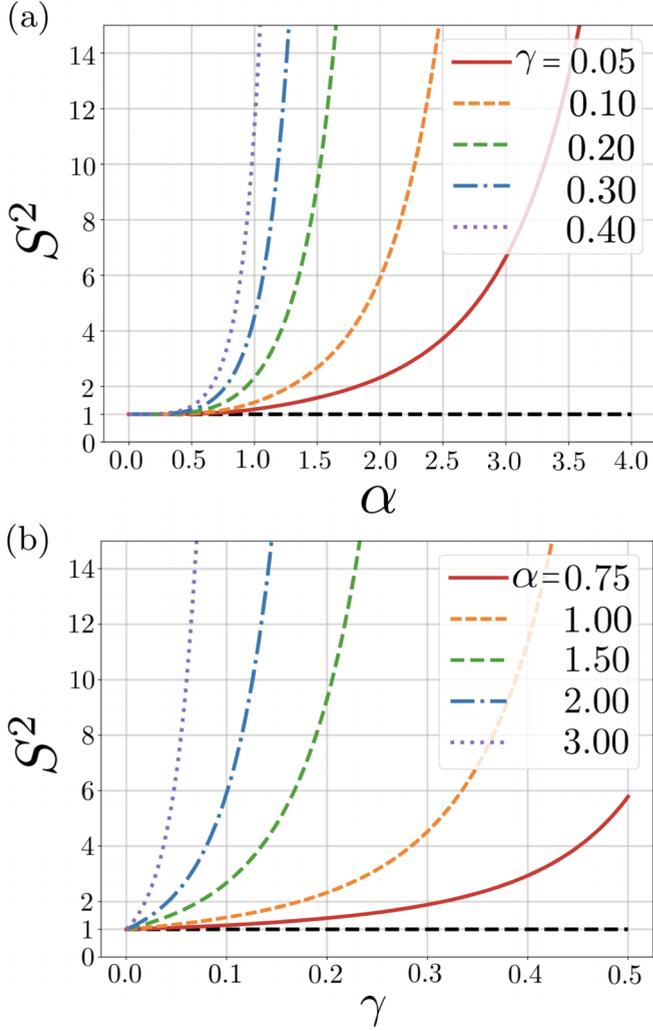
(a)

(b)

FIG. 6. $S^2 \approx$ sampling overhead as a function of (a) γ and (b) $|\alpha|$. When $\text{Var}[O_{\text{even}}] \approx \text{Var}[O_{\text{odd}}] \approx \text{Var}[O_{\text{noisy}}]$, we can approximate the sampling overhead with S^2, which is plotted. Despite the exponential growth with respect to $|\alpha|$ and γ, there is a large and relevant parameter space in which the sampling overhead is sufficiently modest to be experimentally reasonable.

where $j_i \in \{0, 1, 2, \dots\}$ denotes how many photons are subtracted from mode i, $\boldsymbol{g_0} = (g_{0,0}, g_{0,1}, \dots, g_{0,M})$ with $g_{0,i} = 1/\sqrt{1-\gamma_i}$, and $\tilde{\mathcal{N}}_{\boldsymbol{j}}(\rho) = \text{Tr}[\hat{\mathcal{A}}_{\boldsymbol{j}}^{(M)} \hat{\mathcal{G}}_{\boldsymbol{g_0}}^{(M)} \rho \, \hat{\mathcal{G}}_{\boldsymbol{g_0}}^{(M)} \hat{\mathcal{A}}_{\boldsymbol{j}}^{(M)\dagger}]$ is the normalization. We again refer to these maps as channels for simplicity. If the initial state ρ can be written as a tensor product, $\rho_{\text{prod}} = \bigotimes_{i=1}^{M} \rho_i$, then the normalization reduces to $\tilde{\mathcal{N}}_{\boldsymbol{j}}(\rho_{\text{prod}}) = \prod_{i=1}^{M} \mathcal{N}_{j_i}(\rho_i)$ where $\mathcal{N}_{j_i}(\rho_i)$ is defined in Eq. (22).

The M mode inverse loss map can then be decomposed into physically realizable channels with potentially negative coefficients via

$$\Lambda_{\boldsymbol{\gamma}}^{-1\,(M)}[\rho] = \sum_{j_1=0}^{\infty} \cdots \sum_{j_M=0}^{\infty} \Omega_{\boldsymbol{j}}(\rho) \mathcal{E}_{\boldsymbol{j}}^{(M)}[\rho]. \quad (48)$$

Once again, $\Omega_{\boldsymbol{j}} = (\prod_{i=1}^{M} \frac{(-\gamma_i)^{j_i}}{j_i!}) \tilde{\mathcal{N}}_{\boldsymbol{j}}(\rho)$ is a quasiprobability distribution and hence satisfies $\sum_{j_1, j_2, \dots j_M=0}^{\infty} \Omega_{\boldsymbol{j}} = 1$. If ρ is a

product state then the coefficients reduce to a product of those defined in Eq. (23), i.e., $\Omega_{\boldsymbol{j}}(\rho_{\text{prod}}) = \prod_{i=1}^{M} \omega_{j_i}(\rho_i)$.

B. Quantum error cancellation protocol

In order to physically implement the set of channels $\{\mathcal{E}_{\boldsymbol{j}}^{(M)}\}$, we need to have access to (noiseless) amplification and multiphoton subtractions. Photon subtractions are routinely implemented via a beam splitter interaction that further attenuates the state. In order to compensate for this unwanted effect, we need to amplify each mode by a factor $g_{\mu,i} > g_{0,i}$, where i denotes the mode and $g_{\mu,i}$ is a new parameter that depends on the transmissivity of the beam splitter in the ith mode. We then redefine the vector of amplification factors as $\boldsymbol{g_\mu} = (g_{\mu,1}, g_{\mu,2}, \dots, g_{\mu,M})$. For an ideal initial state ρ_0, the amplified initial state, ρ_{amp}, is given by

$$\rho_{\text{amp}} = \frac{1}{\tilde{\mathcal{N}}_{\text{amp}}} \hat{\mathcal{G}}_{\boldsymbol{g_\mu}}^{(M)} \rho_0 \hat{\mathcal{G}}_{\boldsymbol{g_\mu}}^{(M)}, \quad (49)$$

where $\tilde{\mathcal{N}}_{\text{amp}} = \text{Tr}[\hat{\mathcal{G}}_{\boldsymbol{g_\mu}}^{(M)} \rho_0 \hat{\mathcal{G}}_{\boldsymbol{g_\mu}}^{(M)}]$ is the normalization. As in the single mode case, we recommend (where possible) placing this amplification in the state preparation stage—the discussion of this amplification within the single mode setting in Sec. III B 1 is still highly relevant, particularly for product states. For entangled initial states, there are still many experimentally relevant states that maintain their functional form under amplification. For example, amplifying a two mode squeezed vacuum state of squeezing parameter r_0 produces a two mode squeezed vacuum of squeezing parameter $r_{\text{amp}} > r_0$. Similarly, amplification of two mode entangled coherent state $\sim |\alpha_0\rangle |\beta_0\rangle + |-\alpha_0\rangle |-\beta_0\rangle$ implements the transformations $\alpha_0 \to \alpha_{\text{amp}} > \alpha_0$ and $\beta_0 \to \beta_{\text{amp}} > \beta_0$. The photon subtraction term can be implemented with M beam splitters and photon number resolving detectors. In each mode, a beam splitter of transmissivity $1 - \mu_i$ for $i \in \{1, 2, \dots, M\}$ is used with a vacuum in the second entry port. The photon number resolving detectors then measure the output of the second exit port. A measurement of $\boldsymbol{j} = (j_1, j_2, \dots, j_M)$ photons with the photon number resolving detectors will implement a tensor product of Kraus operators,

$$\mathcal{K}_{\boldsymbol{j},\boldsymbol{\mu}}^{(M)} = K_{j_1}(\mu_1) \otimes K_{j_2}(\mu_2) \otimes \cdots \otimes K_{j_M}(\mu_M), \quad (50)$$

on ρ_{amp}, where $K_{j_i}(\mu_i)$ are defined in Eq. (10). This action heralds the state

$$\rho_{\text{amp}} \to \frac{1}{\tilde{p}_{\boldsymbol{j}}} \mathcal{K}_{\boldsymbol{j},\boldsymbol{\mu}}^{(M)} \rho_{\text{amp}} \mathcal{K}_{\boldsymbol{j},\boldsymbol{\mu}}^{(M)\dagger} = \mathcal{E}_{\boldsymbol{j}}^{(M)}[\rho_0], \quad (51)$$

where $\tilde{p}_{\boldsymbol{j}} = \text{Tr}[\mathcal{K}_{\boldsymbol{j},\boldsymbol{\mu}}^{(M)} \rho_{\text{amp}} \mathcal{K}_{\boldsymbol{j},\boldsymbol{\mu}}^{(M)\dagger}]$ is the probability that $\boldsymbol{j}$ photons were subtracted. The equality in Eq. (51) is easily be seen by expanding both ρ_{amp} and $\mathcal{K}_{\boldsymbol{j},\boldsymbol{\mu}}^{(M)}$. We therefore have a way of sampling from $\{\mathcal{E}_{\boldsymbol{j}}^{(M)}[\rho_0]\}$ with corresponding probabilities $\{\tilde{p}_{\boldsymbol{j}}\}$.

We can now consider measuring a generic M-mode observable O. It can be seen that the ideal expectation value, $\langle O \rangle_{\text{ideal}} = \text{Tr}[O \mathcal{U}_{\text{ideal}}[\rho_0]]$, can be decomposed into

$$\langle O \rangle_{\text{ideal}} = \sum_{\boldsymbol{j}} \Omega_{\boldsymbol{j}}(\rho_0) \langle O \rangle_{\boldsymbol{j},\text{noisy}}, \quad (52)$$

where $\langle O \rangle_{j,\text{noisy}} = \text{Tr}[O\,(\mathcal{U}_{\text{noisy}} \circ \mathcal{E}_j^{(M)})[\rho_0]]$ is the noisy expectation value of O given that the channel $\mathcal{E}_j$ was implemented and $\sum_j$ symbolically represents a sum over all possible $\boldsymbol{j} = (j_1, j_2, \ldots, j_M)$.

Now imagine running the circuit N times. The channel that was implemented in each shot is stochastically determined by the number and distribution of detected photons in the photon subtraction stage. We bin the results based on $\boldsymbol{j}$—the set of different channels implemented is denoted $\boldsymbol{J}$. For each element of $\boldsymbol{j} \in \boldsymbol{J}$, the associated noisy expectation value $\langle O \rangle_{j,\text{noisy}}$ is estimated by the arithmetic mean of the O measurement outcomes in which the channel $\mathcal{E}_j^{(M)}$ was implemented—we denote this $\bar{O}_{j,\text{noisy}}$. The mitigated estimator of $\langle O \rangle_{\text{ideal}}$ is then

$$\bar{O}_{\text{mit}} = \sum_{j \in \boldsymbol{J}} \Omega_j(\rho_0)\bar{O}_{j,\text{noisy}}. \tag{53}$$

C. Error analysis

The bias and variance of the mitigated estimator $\bar{O}_{\text{mit}}$ are given by

$$\text{Bias}[\bar{O}_{\text{mit}}; \boldsymbol{J}] = \left| \sum_{j \notin \boldsymbol{J}} \Omega_j(\rho_0)\langle O \rangle_{j,\text{noisy}} \right| \tag{54}$$

and

$$\text{Var}[\bar{O}_{\text{mit}}; \boldsymbol{J}] = \sum_{j \in \boldsymbol{J}} \frac{|\Omega_j(\rho_0)|^2}{N_j} \text{Var}[O_{j,\text{noisy}}], \tag{55}$$

where $\text{Var}[O_{j,\text{noisy}}] = \langle O^2 \rangle_{j,\text{noisy}} - \langle O \rangle_{j,\text{noisy}}^2$ is the single shot variance of O given the channel $\mathcal{E}_j^{(M)}$ was implemented and N_j is the number of measurement outcomes associated with $\mathcal{E}_j^{(M)}$. Once again, the fractional bias can be estimated if one assumes $\langle O \rangle_{j,\text{noisy}}/\langle O \rangle_{\text{ideal}} \approx \mathcal{O}(1)$. Given an acceptable bias ϑ_{bias}, a suitable set $\boldsymbol{J}$ can be chosen—this choice can have a significant impact on bias. The variance grows monotonically with $|\boldsymbol{J}|$ which leads to an accuracy-overhead tradeoff that one has to optimize. There also exists an amplification-overhead tradeoff governed by $\boldsymbol{\mu}$. Under the approximation $N_j = \tilde{p}_j N$, the sampling overhead of the estimated observable O is approximately given by

$$\text{sampling overhead}[O] \approx \sum_{j \in \boldsymbol{J}} \frac{|\Omega_j(\rho_0)|^2}{\tilde{p}_j} \frac{\text{Var}[O_{j,\text{noisy}}]}{\text{Var}[O_{\text{noisy}}]}, \tag{56}$$

where $\text{Var}[O_{\text{noisy}}] = \langle O^2 \rangle_{\text{noisy}} - \langle O \rangle_{\text{noisy}}^2$ is the single shot variance of O given that only the noisy dynamics, $\mathcal{U}_{\text{noisy}}$, were run. If one is measuring a collection of local observables simultaneously, such as $O_1 := O_1 \otimes \mathbb{I}_2 \otimes \cdots \otimes \mathbb{I}_M$, $O_2 := \mathbb{I} \otimes O_2 \otimes \cdots \otimes \mathbb{I}_M$, then the sampling overhead is given by the largest individual sampling overhead,

$$\text{sampling overhead} \approx \max_i [\text{sampling overhead}[O_i]]. \tag{57}$$

D. Monte Carlo approach

As in the single mode case, if each *state* in the operator-sum representation of $\Lambda_{\boldsymbol{\gamma}}^{-1(M)}$, namely $\mathcal{E}_j^{(M)}[\rho_0]$, can be efficiently prepared, then our scheme can be reduced to Monte

Carlo estimation. By sampling from $\{\mathcal{E}_j^{(M)}[\rho_0]\}$ with probabilities $\{|\Omega_j|/S\}$ where $S = \sum_j |\Omega_j|$, an unbiased estimator of $\langle O \rangle_{\text{ideal}}$ can be obtained by multiplying the measurement outcomes associated with $\boldsymbol{j}$ by $S\,\text{sgn}[\Omega_j(\rho_0)]$—this is exactly analogous to the single mode Monte Carlo approach discussed in Sec. III D.

We provide an explicit example of this approach for two modes in Sec. IV E. This approach is also particularly useful for systems in the dual-rail representation [60], where logical qubits are encoded by a single photon shared between two modes via

$$|0\rangle_L := |0\rangle\,|1\rangle, \quad |1\rangle_L := |1\rangle\,|0\rangle. \tag{58}$$

If one wanted to implement our error cancellation scheme using the Monte Carlo approach for an initial state $|0\rangle_L^{\otimes N} = (|0\rangle\,|1\rangle)^{\otimes N}$ then, under the assumption of uniform losses, S can be shown to be

$$S = \left(\frac{1+\gamma}{1-\gamma} \right)^N. \tag{59}$$

Intriguingly, this means one can undo the photon losses by stochastically sending *fewer* single photons into the circuit. We provide some intuition for why this is in Sec. III A.

E. Examples

In this section we look at two classes of entangled initial states, namely two mode squeezed vacua and entangled coherent states. As for the single mode squeezed vacuum case, we again use the observable $O = \rho_0$; for the pure initial states considered here, this expectation value estimation task corresponds to estimating the fidelity between ρ_{noisy} and ρ_{mit} with respect to ρ_0. Furthermore, as argued in Sec. III E 1, this provides an observable and unitary-dynamics independent benchmark for our mitigation scheme. We also consider the covariance matrix as an observable for the two mode squeezed vacua example.

We allow the loss parameter to fluctuate between shots in Figs. 8 and 9 in order to create a more realistic noise model, as discussed in Sec. II B. Recovery of the ideal expectation value in both of these simulations is strong evidence that our mitigation scheme is robust against a fluctuating loss parameter, and hence demonstrates the applicability of quantum error cancellation realistic scenarios. In Fig. 10, we plot an archetypal histogram of different loss parameters used in a 1×10^6 shot experiment with average losses exceeding 50%.

1. Two mode squeezed vacua

Consider the TMSV states given by

$$|S_2(r)\rangle = e^{r(\hat{a}_1^\dagger \hat{a}_2^\dagger - \hat{a}_1 \hat{a}_2)}|00\rangle$$

$$= \cosh[-1](r) \sum_{n=0}^\infty \tanh[n](r)|nn\rangle. \tag{60}$$

Let $\rho_0 = |S_2(r_0)\rangle\langle S_2(r_0)|$. Amplifying $|S_2(r_0)\rangle$ by $\mathcal{G}_{\boldsymbol{g}=(g_1,g_2)}^{(M)}$ gives $|S_2(r_{\text{amp}})\rangle$ up to normalization, where $\tanh(r_{\text{amp}}) = g_1 g_2 \tanh(r_0)$ and the amplification normalization is $\tilde{\mathcal{N}}_{\text{amp by }\boldsymbol{g}} = \frac{\cosh[2](r_{\text{amp}})}{\cosh[2](r_0)}$. As in the single mode squeezed
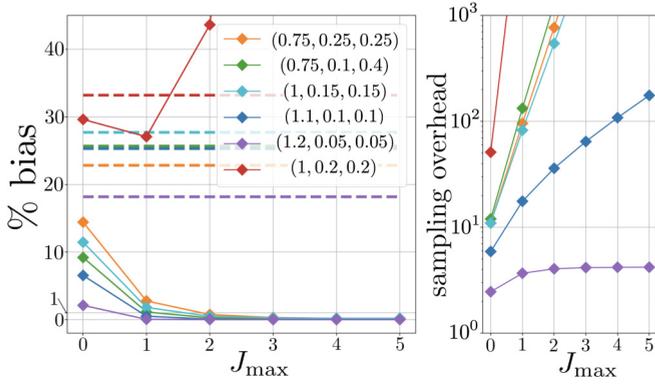
FIG. 7. On the left, we plot the bias of $\mathbb{E}[\bar{O}_{\mathrm{mit}}] = F(\rho_0, \rho_{\mathrm{mit}})$ as a function of $J_{\max}$ for the set $\boldsymbol{J}^{(\mathrm{local})}(J_{\max})$. Each curve is labeled by $(r_0, \gamma_1, \gamma_2)$. The dashed lines give the noisy expectation values, $\langle O \rangle_{\mathrm{noisy}} = F(\rho_0, \Lambda_{\boldsymbol{\gamma}}[\rho_0])$. With the exception of the red curve, we see a rapid convergence to the ideal expectation value as $J_{\max}$ grows, and even the zero-photon subtraction channel yields a drastic improvement over the unmitigated results. Subtracting just up to one photon from each mode is sufficient to achieve $\mathcal{O}(1\%)$ bias and in the small loss regime this is improved to $\mathcal{O}(0.1\%)$. The red curve diverges because $\Omega_{\boldsymbol{j}}$ increases as j_1 and j_2 increase in the regime where $\boldsymbol{j}$ photon losses are non-negligible—this behavior can be calculated *a priori* and hence can be avoided by redefining r_0. On the right, we plot the approximate sampling overhead, $\sum_{\boldsymbol{j} \in \boldsymbol{J}} |\Omega_{\boldsymbol{j}}|^2 / \tilde{p}_{\boldsymbol{j}}$ as a function of $J_{\max}$ under the same choice of parameters. We see a clear exponential growth with respect to $J_{\max}$ with a gradient that has a larger dependence on the strength of noise than the initial squeezing parameter. For each $(r_0, \boldsymbol{\gamma})$, the beam splitter parameters $\boldsymbol{\mu} = (\mu, \mu)$ (beam splitter parameters are chosen to the nearest 0.005) minimize sampling overhead when $J_{\max} = 1$. The squeezing parameters associated with $\hat{\mathcal{G}}_{\boldsymbol{g_0}}[\rho_0]$ and $\hat{\mathcal{G}}_{\boldsymbol{g_\mu}}[\rho_0]$ for each curve (orange, green, ..., purple, red) are given by $(1.245, 1.746)$, $(1.310, 1.734)$, $(1.452, 1.924)$, $(1.420, 1.794)$, $(1.365, 1.552)$, and $(1.856, 2.335)$. These are all accessible with current devices. One can decrease the necessary amplification $\boldsymbol{g_\mu}$ by accepting a larger sampling overhead.

vacuum case, we require $g_1 g_2 \tanh(r_0) < 1$ for the amplification to converge and be physically meaningful.

One can show that the quasiprobability distribution $\Omega_{\boldsymbol{j}}$ is given by

$$\Omega_{\boldsymbol{j}}(|S_2(r_0)\rangle\langle S_2(r_0)|) = \frac{(-\gamma_1)^{j_1}(-\gamma_2)^{j_2}}{\cosh[2](r_0)} \sum_{n=\max[j_1, j_2]}^{\infty} \binom{n}{j_1}\binom{n}{j_2}$$

$$\times \left(\frac{\tanh[2](r_0)}{(1-\gamma_1)(1-\gamma_2)} \right)^n. \quad (61)$$

Again, this converges when $\tanh(r_0)/\sqrt{(1-\gamma_1)(1-\gamma_2)} < 1$. For fixed choices of r_0 and $\boldsymbol{\gamma}$, if increasing j_1 and/or j_2 leads to $\Omega_{\boldsymbol{j}}$ increasing, then the quasiprobability weight is pushed towards the channels $\mathcal{E}_{\boldsymbol{j}}^{(M)}$ that cannot be implemented. This leads to mitigated expectation values that can look divergent in j_1, j_2 and places further constraints on what values r_0 and $\boldsymbol{\gamma}$ can take. An example of this is seen in the red curve of Fig. 7. This has parameters $(r_0, \gamma_1, \gamma_2) = (1, 0.2, 0.2)$ and starts rapidly diverging away from the ideal result—in principle, we know that $\langle O \rangle_{\mathrm{ideal}} = \sum_{\boldsymbol{j}} \Omega_{\boldsymbol{j}} \langle O \rangle_{\boldsymbol{j}}$ but in practice, any numerical cutoff we tested led to divergent behavior.
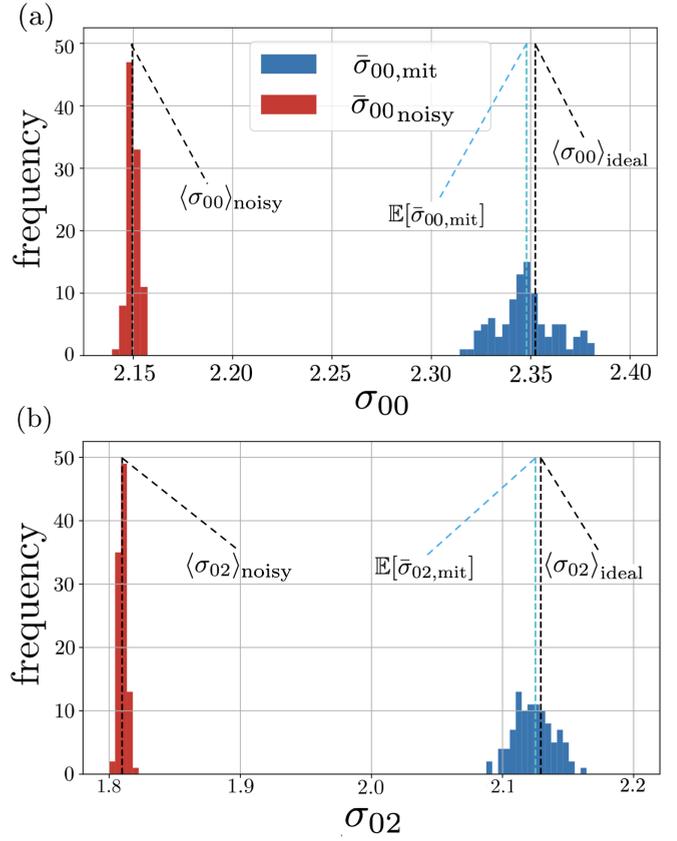

FIG. 8. Two histograms illustrating the distribution of results from 100 simulated experiments in which (a) σ_{00} and (b) σ_{02} were measured. The initial state, $|S_2(0.75)\rangle$, was subject to losses stochastically chosen from a discretized Gaussian distribution centered on $\bar{\boldsymbol{\gamma}} = (0.15, 0.15)$ with standard deviation $\boldsymbol{\sigma_\gamma} = 0.1\bar{\boldsymbol{\gamma}}$ in each shot. We measured 1×10^6 shots in each experiment to give an estimate of the dimensionless $\sigma_{k\ell}$ and the photon subtraction beam splitters were parametrized by $\boldsymbol{\mu} = (0.105, 0.105)$. The red bars show the unmitigated results, the blue bars show the mitigated results, and the labeled vertical dashed lines give the ideal, noisy, and mitigated expectation values. The percentage bias decreases from $(8.6\%, 15.0\%)$ to $(0.19\%, 0.19\%)$ while the sampling overhead is a modest ≈ 4.4 and hence our mitigation scheme gives a vast improvement over the unmitigated results, as clearly demonstrated by the histograms.

Numerically, we found that keeping $\tanh(r_0)/\sqrt{(1-\gamma_1)(1-\gamma_2)} \lesssim 0.9$ tends to ensure that, for the $\boldsymbol{j}$ we are interested in, the expectation value converges to the ideal result. This limit can be increased slightly if $\gamma_1 \neq \gamma_2$ and $\gamma_i \ll 1$.

Interestingly, $\Omega_{\boldsymbol{j}}$ can initially decrease with $\boldsymbol{j}$ for "small" values of j_1 and j_2 before beginning to increase with $\boldsymbol{j}$ for "large" values of j_1 and j_2 in such a way that these higher photon subtraction terms dominate the quasiprobability. Fortunately, this effect is negated by the fact that within the Kraus representation of $\Lambda_{\boldsymbol{\gamma}}[\rho_0]$, the terms corresponding to high photon subtraction terms are themselves vanishingly small and therefore the net effect is no appreciable impact on the expectation value.

In Fig. 7, we plot the expected percentage bias of the mitigated estimator, $F(\rho_0, \rho_{\mathrm{mit}})$, as well as the sampling overhead
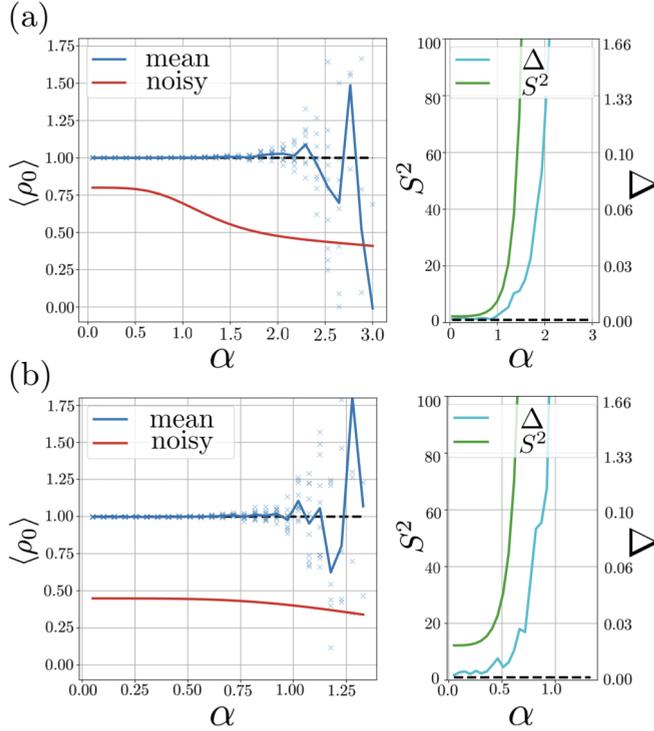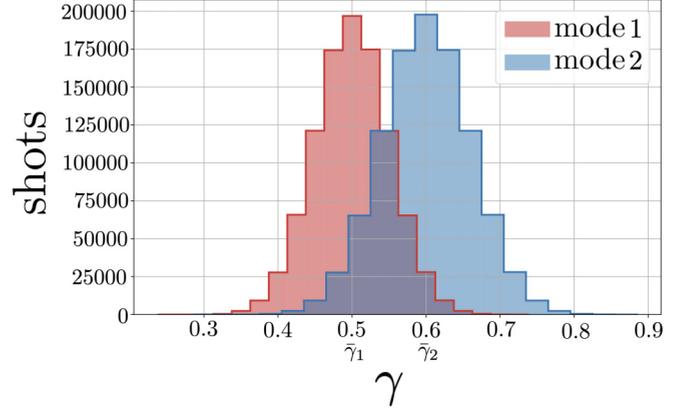
(a)

(b)

FIG. 10. A histogram representative of the γ values sampled in Fig. 9(b) for a particular experiment with 1×10^6 shots. The discretized Gaussian distribution depicted here has mean $\bar{\gamma} = (0.5, 0.6)$ and standard deviation $\sigma_{\bar{\gamma}} = 0.1\bar{\gamma}$. Despite the relatively large spread in possible loss parameters in a particular shot (ranging all the way from 30 to 90% in mode 2), the fidelity recovery seen in Fig. 9 demonstrates that our mitigation scheme is robust against fluctuating losses and hence can be applied successfully in real experiments. In each mode, the Gaussian distribution was linearly discretized into 21 possible values. The range of possible γ values was truncated at the tails such that the probability of γ lying outside the discretized distribution was $\approx 10^{-7}$.

FIG. 9. Simulation of $\langle \rho_0 \rangle = F(\rho_0, \rho_{\mathrm{mit}})$ against α for the entangled coherent states $\propto |\alpha, \alpha\rangle - |-\alpha, -\alpha\rangle$ subject to losses using our mitigation scheme. In each shot, the loss parameter was stochastically selected from a discretized Gaussian distribution centered on $\bar{\gamma}$ with width $\sigma_\gamma = 0.1\bar{\gamma}$. In (a), the average losses are $\bar{\gamma} = (0.2, 0.2)$ while in (b) the average losses are $\bar{\gamma} = (0.5, 0.6)$. For each value of α, we conduct ten experiments with 1×10^6 shots per experiment—each experiment is marked by a tick on the graph and the blue curve is the average over the ten experiments. We see a clear recovery of the ideal expectation value for $\alpha \lesssim 2.25$ in (a) and $\alpha \lesssim 1$ in (b). Beyond these points, the large sampling requirements make individual experiments unreliable and can give fidelities that drastically exceed 1. The graphs on the right plot $S^2 \approx$ sampling overhead and Δ, which is defined as the difference between the maximum and minimum values of $\bar{O}_{\mathrm{mit}}$ for each α. We see a clear relationship between the two which indicates that the sampling overhead is accurately characterizing the shot noise.

for various choices of (r_0, γ, μ). We choose the set $\boldsymbol{J} = \boldsymbol{J}^{(\mathrm{local})}$ which is defined in terms of a parameter J_{max} by

$$\boldsymbol{J}^{(\mathrm{local})}(J_{\mathrm{max}}) = \{(j_1, j_2) \mid j_1 \leqslant J_{\mathrm{max}}, \ j_2 \leqslant J_{\mathrm{max}}\}. \quad (62)$$

In other words, J_{max} labels the maximum number of photons subtracted *locally* in each mode. We note in passing that the choice of $\boldsymbol{J}$ can have a large impact on the bias—for example, the set $\boldsymbol{J} = \boldsymbol{J}^{(\mathrm{global})}(J_{\mathrm{max}}) := \{(j_1, j_2) \mid j_1 + j_2 \leqslant J_{\mathrm{max}}\}$ is often outperformed by $\boldsymbol{J}^{(\mathrm{local})}$, even in cases where $\boldsymbol{J}^{(\mathrm{local})}(J_{\mathrm{max},1}) \subset \boldsymbol{J}^{(\mathrm{global})}(J_{\mathrm{max},2})$.

The results show a fast convergence to the ideal expectation value for all the results except the red curve, where $(r_0, \gamma_1, \gamma_2) = (1, 0.2, 0.2)$. This red diverging curve is explained by the fact that Ω_j increases with $\boldsymbol{j}$—the factor $\tanh(r_0)/\sqrt{(1-\gamma_1)(1-\gamma_2)} \approx 0.952 > 0.9$. Decreasing γ_i from $= 0.2$ to 0.15 changes the factor to $\tanh(r_0)/\sqrt{(1-\gamma_1)(1-\gamma_2)} \approx 0.896 < 0.9$ which does

converge correctly as discussed above. The sampling overhead is also plotted. We see that it grows exponentially with J_{max} and the gradient is most significantly impacted by the level of noise γ.

We will now briefly discuss the covariance matrix before using it as the observable in simulations plotted in Fig. 8. For further details on the covariance matrix, refer to Refs. [61,62].

Two mode continuous variable systems can be described by the dimensionless quadrature operators $\hat{x}_i = \frac{1}{\sqrt{2}}(\hat{a}_i + \hat{a}_i^\dagger)$ and $\hat{p}_i = \frac{1}{i\sqrt{2}}(\hat{a}_i - \hat{a}_i^\dagger)$ for $i \in \{1, 2\}$. These can be collected into single four-component vector $\hat{\boldsymbol{r}} = (\hat{x}_1, \hat{p}_1, \hat{x}_2, \hat{p}_2)^T$. The covariance matrix $\boldsymbol{\sigma}$ has matrix elements [61]

$$\sigma_{k\ell} = \langle \{\hat{r}_k, \hat{r}_\ell\} \rangle - 2\langle \hat{r}_k \rangle \langle \hat{r}_\ell \rangle \quad (63)$$

where $\{A, B\} := AB + BA$ is the anticommutator. States that admit a Gaussian Wigner function are called Gaussian states and they can be fully characterized by $\langle \hat{\boldsymbol{r}} \rangle$ and $\boldsymbol{\sigma}$ [61]. TMSV states and TMSV states that have been subject to loss are both examples of Gaussian states with $\langle \hat{\boldsymbol{r}} \rangle = 0$ and are hence fully characterized by their respective covariance matrices—by measuring $\boldsymbol{\sigma}$, we are essentially carrying out state tomography on $|S_2(r)\rangle\langle S_2(r)|$ and $\Lambda_\gamma^{(2)}[|S_2(r)\rangle\langle S_2(r)|]$.

For both of these states, of the 16 matrix elements in $\boldsymbol{\sigma}$, it can be shown that (up to a sign) there are only two unique nonzero terms:

$$\boldsymbol{\sigma}(|S_2(r)\rangle\langle S_2(r)|) = \begin{pmatrix} \cosh(2r)\boldsymbol{I} & \sinh(2r)\boldsymbol{Z} \\ \sinh(2r)\boldsymbol{Z} & \cosh(2r)\boldsymbol{I} \end{pmatrix} \quad (64)$$

where $\boldsymbol{I}$ is the 2×2 identity matrix and $\boldsymbol{Z}$ is the Z-Pauli matrix. Hence if one measures σ_{00} and σ_{02}, one has fully characterized the state.

In Fig. 8, we simulate measuring σ_{00} and σ_{02} from a TMSV state with parameter $r_0 = 0.75$ subject to losses centered on $\bar{\gamma} = (0.15, 0.15)$. In each shot, the true photon loss parameter is stochastically taken from a discretized Gaussian distribution of standard deviation $0.1\bar{\gamma}$. To carry out the photon subtractions, two beam splitters of reflectivities $\boldsymbol{\mu} = (0.105, 0.105)$ were chosen [63]. This gave an amplified squeezing parameter of $r_{\text{amp by } g_\mu} \approx 1.2041$ (compare this with the minimum necessary squeezing for mitigation, $r_{\text{amp by } g_0} \approx 0.9666$). We simulated 100 experiments with 1×10^6 shots per experiment and chose $\boldsymbol{J} = \boldsymbol{J}^{(\text{local})}(3)$.

Our error cancellation protocol successfully recovers the noise-free results to within a good approximation, and vastly outperforms the noisy results. The ideal expectation value recovery takes place in spite of a fluctuating $\boldsymbol{\gamma}$ which is evidence that our scheme is robust in realistic settings.

Because σ_{00} and σ_{02} are sufficient to fully characterize $|S_2(r_0)\rangle$, the recovery of the ideal expectation value means we are able to conduct tomography on a Gaussian state that has passed through an unwanted photon loss channel. One can generalize this to tomography of an arbitrary state that has undergone unwanted losses by measuring, for example, the Wigner function as an expectation value [64].

2. Entangled coherent states

Consider now the two mode entangled coherent states given by

$$|\psi_\pm(\alpha, \beta)\rangle = \frac{1}{\mathcal{B}_\pm(\alpha, \beta)}(|\alpha\rangle |\beta\rangle \pm |-\alpha\rangle |-\beta\rangle), \quad (65)$$

where $\mathcal{B}_\pm(\alpha, \beta) = \sqrt{2\{1 \pm \exp[-2(|\alpha|^2 + |\beta|^2)]\}}$ is the necessary normalization. We define $\rho_\pm(\alpha, \beta) = |\psi_\pm(\alpha, \beta)\rangle\langle\psi_\pm(\alpha, \beta)|$. If the system is subject to losses $\boldsymbol{\gamma} = (\gamma_1, \gamma_2)$, then it can be shown that the operator-sum representation of the inverse map acting on $\rho_-(\alpha, \alpha)$ can be rewritten using only two terms, i.e.,

$$\Lambda_{\boldsymbol{\gamma}}^{-1(M)}[\rho_-(\alpha, \alpha)] = \tilde{W}_{\text{even}} \, \rho_-(\tilde{\alpha}_1, \tilde{\alpha}_2) + \tilde{W}_{\text{odd}} \, \rho_+(\tilde{\alpha}_1, \tilde{\alpha}_2). \quad (66)$$

Here, $\tilde{\alpha}_i = \alpha/\sqrt{1 - \gamma_i}$ and the coefficients $\tilde{W}_{\text{even/odd}}$ are given by

$$
\begin{aligned}
\tilde{W}_{\text{even}} = {}& e^{\gamma_1 |\tilde{\alpha}_1|^2 + \gamma_2 |\tilde{\alpha}_2|^2} \left| \frac{\mathcal{B}_-(\tilde{\alpha}_1, \tilde{\alpha}_2)}{\mathcal{B}_-(\alpha, \alpha)} \right|^2 \\
& \times [\cosh(\gamma_1 |\tilde{\alpha}_1|^2) \cosh(\gamma_2 |\tilde{\alpha}_2|^2) \\
& + \sinh(\gamma_1 |\tilde{\alpha}_1|^2) \sinh(\gamma_2 |\tilde{\alpha}_2|^2)],
\end{aligned} \quad (67)
$$

$$
\begin{aligned}
\tilde{W}_{\text{odd}} = {}& -e^{\gamma_1 |\tilde{\alpha}_1|^2 + \gamma_2 |\tilde{\alpha}_2|^2} \left| \frac{\mathcal{B}_+(\tilde{\alpha}_1, \tilde{\alpha}_2)}{\mathcal{B}_-(\alpha, \alpha)} \right|^2 \\
& \times [\cosh(\gamma_1 |\tilde{\alpha}_1|^2) \sinh(\gamma_2 |\tilde{\alpha}_2|^2) \\
& + \sinh(\gamma_1 |\tilde{\alpha}_1|^2) \cosh(\gamma_2 |\tilde{\alpha}_2|^2)].
\end{aligned} \quad (68)
$$

Notice the key negative sign on $\tilde{W}_{\text{odd}}$. If one can produce the entangled coherent states $|\psi_\pm(\tilde{\alpha}_1, \tilde{\alpha}_2)\rangle$, then the Monte Carlo approach from Sec. IV D can be used to give an unbiased estimator of $\langle O \rangle_{\text{ideal}}$. The sampling overhead is approximately given by S^2, where $S = |\tilde{W}_{\text{even}}| + |\tilde{W}_{\text{odd}}|$.

In Fig. 9, we plot simulated experiments in which the fidelity between ρ_0 and ρ_{mit} was measured as an expectation value. As there is no cutoff in the inverse channel [i.e., Eq. (48)] with this implementation, this provides an unbiased estimate for all values of α, β, and γ. On the right, we plot $S^2 \sim$ sampling overhead, and the difference between the maximum and minimum values of $\bar{O}_{\text{mit}}$ for a given α, labeled Δ. In both (a) and (b), the sampling overhead grows exponentially with α, which leads to a drastic increase in the shot noise. Beyond some critical coherent state magnitude, α_{crit}, the shot noise dominates and makes individual experiments unreliable. Nonetheless, there is a large and interesting parameter regime in which losses can be mitigated, with just a modest number of shots—for example, the ideal result is consistently recovered for an entangled coherent state of magnitude $\leqslant 1$ even when total losses exceed 50%.

In each shot of the simulation, we stochastically select the specific loss parameter γ from a discretized Gaussian distribution centered on $\bar{\gamma}$ with standard deviation $\sigma_{\bar{\gamma}} = 0.1\bar{\gamma}$. The recovery of the fidelity then indicates that our scheme is robust against a randomly fluctuating loss parameter. In particular, the results in Figs. 9(b) and 10 demonstrate that recovery is robust even when the loss parameters are sampled from surprisingly wide distributions. In fact, when comparing these results with results in which the loss parameter was equal to $\bar{\gamma}$ for all shots, the authors could not distinguish a noticeable difference.

V. DISCUSSION

In this paper, we have proposed a quantum error mitigation protocol applicable to discrete or continuous variable photonic systems subject to excitation losses, inspired by DV probabilistic error cancellation. We have analytically inverted the photon loss channel and expressed it using a quasiprobability distribution over physically realizable CP channels. These CP channels involve both a noiseless amplification and multiphoton subtractions and we have proposed viable experimental protocols for implementing both these operations. Following this, we have then shown how the noisy dynamics can be run and the measurement outcomes combined to construct an estimator of $\langle O \rangle_{\text{ideal}}$ with a bias that can be made arbitrarily small. Furthermore, we have presented a second approach based on Monte Carlo estimation for certain classes of input states—single photon states, cat states, and entangled coherent states—and proven that this produces an unbiased estimator of $\langle O \rangle_{\text{ideal}}$. In both cases, an unavoidable sampling overhead is introduced. Similar to DV PEC, where the sampling overhead scales exponentially with qubit number and circuit depth, here it scales exponentially with the number of modes and noise strength. A new behavior observed is the exponential dependence on the *energy* of the system through either the squeezing parameter for squeezed states, or the coherent state magnitude for cat states and entangled coherent states. Nevertheless, the high shot rate which is characteristic of photonic platforms can ease the impact of this cost.

We have assessed the protocol's performance via several examples. The squeezed state example displays how the mitigation scheme will typically be employed, and shows that the percentage bias can be made less than 1% within currently

experimentally accessible parameter regimes. The protocol was shown to be robust against imperfect characterization of the loss parameter and a fluctuating loss parameter. On the other hand, the cat state example illustrates a special case where one can exploit the state's property of being an eigenstate of $\hat{a}^2$ to make the estimate's bias vanish entirely. This Monte Carlo approach has a better scaling to multiple modes and can be implemented fairly straightforwardly for single photon states, cat states, and entangled coherent states.

We hope this paper encourages new work on QEM in CV systems, whether that be adapting existing DV techniques to the CV setting or finding entirely new error mitigation strategies. A pertinent research direction involves extending this protocol to include dynamic circuits with midcomputation projective measurements—this will be particularly relevant due to the necessity of midcircuit measurements for universal linear optical quantum computing. This protocol would also work particularly well with states that are naturally robust against dephasing errors in noise-biased situations. Finally, we hope this paper inspires experimental work in this area.

APPENDIX A: USEFUL COMMUTATION RELATION

In this section, we prove the following identity:

$$g^{\hat{n}}\hat{a}^j = g^{-j}\hat{a}^j g^{\hat{n}}. \tag{A1}$$

We start by showing via induction that the commutation relation

$$[\hat{a}^j, \hat{n}] = j\hat{a}^j \tag{A2}$$

holds for all j. The $j = 0$ case is trivial and the $j = 1$ case can be seen by using the canonical commutation relation $[\hat{a}, \hat{a}^\dagger] = \mathbb{I}$. The induction step reads

$$\begin{aligned}
[\hat{a}^{j+1}, \hat{n}] &= \hat{a}\hat{a}^j\hat{n} - \hat{n}\hat{a}^j\hat{a} \\
&= \hat{a}^{j+1}\hat{n} - \hat{a}^j\hat{n}\hat{a} + [\hat{a}^j, \hat{n}]\hat{a} \\
&= \hat{a}^{j+1}\hat{n} - \hat{a}^j(\hat{a}\hat{n} - \hat{a}) + j\hat{a}^{j+1} \\
&= (j+1)\hat{a}^{j+1},
\end{aligned} \tag{A3}$$

concluding our proof of Eq. (A2). Next, we prove that

$$[\hat{a}^j, \hat{n}^k] = \hat{a}^j[\hat{n}^k - (\hat{n} - j\mathbb{I})^k] \tag{A4}$$

for all j and k by induction. For $k = 0$ this trivially holds for arbitrary j. For $k = 1$, the right-hand side reduces to $j\hat{a}^j$—we have just proven that $[\hat{a}^j, \hat{n}] = j\hat{a}^j$ for arbitrary j. Now, we show that if Eq. (A4) holds for arbitrary j and a particular k, then it must also hold for arbitrary j and $k + 1$:

$$\begin{aligned}
[\hat{a}^j, \hat{n}^{k+1}] &= \hat{a}^j\hat{n}^{k+1} - \hat{n}\hat{n}^k\hat{a}^j \\
&= \hat{a}^j\hat{n}^{k+1} - \hat{n}(\hat{a}^j\hat{n}^k - [\hat{a}^j, \hat{n}^k]) \\
&= \hat{a}^j\hat{n}^{k+1} - \hat{n}\hat{a}^j(\hat{n} - j\mathbb{I})^k \\
&= \hat{a}^j\hat{n}^{k+1} - \hat{a}^j(\hat{n} - j\mathbb{I})^{k+1} \\
&= \hat{a}^j[\hat{n}^{k+1} - (\hat{n} - j\mathbb{I})^{k+1}].
\end{aligned} \tag{A5}$$

By induction, we have therefore shown that Eq. (A4) holds for arbitrary non-negative integers j and k.

We now Taylor expand $g^{\hat{n}} = \sum_{k=0}^{\infty} \frac{\ln(g)^k}{k!}\hat{n}^k$ and look at $g^{\hat{n}}\hat{a}^j$:

$$g^{\hat{n}}\hat{a}^j = \sum_{k=0}^{\infty} \frac{\ln(g)^k}{k!}\hat{n}^k\hat{a}^j = \hat{a}^j\sum_{k=0}^{\infty} \frac{\ln(g)^k}{k!}(\hat{n} - j\mathbb{I})^k = \hat{a}^j g^{\hat{n}-j}. \tag{A6}$$

Here, we have used the commutation relation in Eq. (A4) in the second line. This proves the useful relation between $g^{\hat{n}}$ and $\hat{a}^j$ in Eq. (A1). This allows us to exchange the order in which the attenuation or amplification and the photon subtractions occur in the loss map or inverse loss map by simply introducing a c number.

APPENDIX B: INVERSE PHOTON LOSS MAP PROOF

In this section, we prove that the P representation and operator-sum representation of the inverse loss map are equivalent. Let $\Lambda_\gamma^{-1,(\text{op.sum})}[\rho]$ be the operator-sum representation in Eq. (19). Via linearity, it is sufficient to show that $\Lambda_\gamma^{-1,\text{op.sum}}[|\alpha\rangle\langle\alpha|] = |g_0\alpha\rangle\langle g_0\alpha|$ because then every coherent state in the integrand of Eq. (18) will be correctly amplified:

$$\begin{aligned}
\Lambda_\gamma^{-1,\text{op.sum}}[|\alpha\rangle\langle\alpha|] &= \sum_{j=0}^{\infty} \frac{(-\gamma)^j}{j!}\hat{a}^j g_0^{\hat{n}}|\alpha\rangle\langle\alpha|g_0^{\hat{n}}(\hat{a}^\dagger)^j \\
&= e^{(g_0^2-1)|\alpha|^2}\sum_{j=0}^{\infty}\hat{a}^j|g_0\alpha\rangle\langle g_0\alpha|(\hat{a}^\dagger)^j \\
&= e^{\frac{\gamma}{1-\gamma}|\alpha|^2}\sum_{j=0}^{\infty}\frac{1}{j!}\left(\frac{-\gamma|\alpha|^2}{1-\gamma}\right)^j|g_0\alpha\rangle\langle g_0\alpha| \\
&= |g_0\alpha\rangle\langle g_0\alpha|.
\end{aligned} \tag{B1}$$

In the third line we used the fact that $g_0^2 - 1 = \gamma/(1-\gamma)$. Therefore, applying Λ_γ^{-1} in the operator-sum representation to ρ in the P representation will return Λ_γ^{-1} in the P representation. This proves that these two representations are equivalent.

APPENDIX C: PHOTON SUBTRACTION

In this section, we prove that sending an arbitrary state ρ through a beam splitter with a vacuum in the second entry port and measuring j photons heralds a state conjugated by the jth Kraus operators defined in Eq. (10).

Let the input and output modes be described by the annihilation operators $\hat{a}_i$ and $\hat{b}_i$ respectively, where $i \in \{\text{system, ancilla}\}$. We let the input system state be pure and given by $|\psi\rangle = \sum_{n=0}^{\infty} c_n |n\rangle = \sum_{n=0}^{\infty} c_n \frac{\hat{a}_S^n}{\sqrt{n!}} |0\rangle$ while the ancilla input state is a vacuum, $|0\rangle$. The order of the tensor product is system-ancilla (e.g., $|0\rangle |0\rangle = |0\rangle_S |0\rangle_A$).

In the Heisenberg picture, a beam splitter of transmissivity $1 - \mu$, described by the unitary $U_{\text{BS},\mu}$ defined in Sec. III B 2, will transform $\hat{a}_S^\dagger \rightarrow \sqrt{1-\mu} \, \hat{b}_S^\dagger - \sqrt{\mu} \, \hat{b}_A^\dagger$. This acts on $|\psi\rangle |0\rangle$ to produce the state

$$
U_{\text{BS},\mu} |\psi\rangle |0\rangle = \sum_{n=0}^{\infty} c_n \frac{(\sqrt{1-\mu} \, \hat{b}_S^\dagger - \sqrt{\mu} \, \hat{b}_A^\dagger)^n}{n!} |0\rangle |0\rangle ,
$$

$$
= \sum_{n=0}^{\infty} \sum_{m=0}^{n} c_n (\sqrt{1-\mu})^{n-m} (-\sqrt{\mu})^m
$$

$$
\times \binom{n}{m}^{\frac{1}{2}} |n-m\rangle |m\rangle , \tag{C1}
$$

where $\binom{n}{m} = n!/(m!(n-m)!)$ is the binomial coefficient.

To calculate the state that is heralded by a measurement of j photons in the ancilla output mode, we first apply $\langle j|_A$ to the above state:

$$
\langle j|_A U_{\text{BS},\mu} |\psi\rangle_S |0\rangle_A
$$

$$
= (-1)^j \sqrt{\left(\frac{\mu}{1-\mu}\right)^j} \sum_{n=j}^{\infty} c_n (1-\mu)^{\frac{n}{2}} \binom{n}{j}^{\frac{1}{2}} |n-j\rangle
$$

$$
= (-1)^j \sqrt{\left(\frac{\mu}{1-\mu}\right)^j} \sum_{n=0}^{\infty} c_n (1-\mu)^{\frac{n}{2}}
$$

$$
\times \binom{n}{j}^{\frac{1}{2}} \sqrt{\frac{(n-j)!}{n!}} \hat{a}^j |n\rangle
$$

$$
= (-1)^j \sqrt{\frac{1}{j!} \left(\frac{\mu}{1-\mu}\right)^j} \hat{a}^j \sum_{n=0}^{\infty} c_n (1-\mu)^{\frac{n}{2}} |n\rangle
$$

$$
= (-1)^j \sqrt{\frac{1}{j!} \left(\frac{\mu}{1-\mu}\right)^j} \hat{a}^j (1-\mu)^{\frac{\hat{n}}{2}} |\psi\rangle
$$

$$
= (-1)^j K_j(\mu) |\psi\rangle , \tag{C2}
$$

where in the third line we used $\hat{a}^j |n < j\rangle = 0$, in the fourth line $\binom{n}{j}^{\frac{1}{2}} \sqrt{\frac{(n-j)!}{n!}} = \sqrt{\frac{1}{j!}}$, and in the fifth line $g^n |n\rangle = g^{\hat{n}} |n\rangle$. Using this, we can see that a projective measurement of j photons from the ancilla output port will (up to normalization) herald the state

$$
|\psi\rangle\langle\psi| \rightarrow \text{Tr}[\mathbb{I} \otimes |j\rangle\langle j|] U_{\text{BS},\mu} (|\psi\rangle\langle\psi|_S \otimes |0\rangle\langle 0|_A) U_{\text{BS},\mu}^\dagger]
$$

$$
= \langle j|_A U_{\text{BS},\mu} |\psi\rangle_S |0\rangle_A \quad \langle\psi|_S \langle 0|_A U_{\text{BS},\mu}^\dagger |j\rangle_A
$$

$$
= K_j(\mu) |\psi\rangle\langle\psi| K_j^\dagger(\mu) . \tag{C3}
$$

This proves the statement for pure states. Via linearity, it can easily be seen that for an arbitrary density operator $\rho = \sum_i \lambda_i |\psi_i\rangle\langle\psi_i|$, a measurement of j photons from the ancillary output port will (up to normalization) herald the state

$$
\rho \rightarrow K_j(\mu) \rho K_j^\dagger(\mu) . \tag{C4}
$$

APPENDIX D: DEPHASING ERRORS

In this section, we briefly discuss bosonic dephasing errors, and the impact they have on our loss mitigation scheme, and give a new dephasing error mitigation scheme based around the inverse dephasing error map. The single mode pure bosonic dephasing channel, Λ_{D,γ_D}, is defined by

$$
\Lambda_{D,\gamma_D}[\rho] = \sum_{n=0}^{\infty} \sum_{m=0}^{\infty} e^{-\frac{\gamma_D}{2}(m-n)^2} \langle m|\rho|n\rangle |m\rangle\langle n|, \tag{D1}
$$

where $\gamma_D \in [0, \infty)$ is a parameter that determines the noise strength. There is an exponential damping on the off-diagonal terms that decoheres the state.

This can equivalently be seen to solve the Lindblad master equation

$$
\frac{d}{dt}\rho(t) = \kappa_D \mathcal{D}[\hat{n}][\rho(t)], \tag{D2}
$$

where $\mathcal{D}[\hat{n}](\rho) = \hat{n}\rho\hat{n} - \frac{1}{2}\{\hat{n}^2, \rho\}$ is the dephasing Lindbladian and $\kappa_D > 0$ is the dephasing rate, if we make the identification $\gamma_D = \kappa_D t$.

The joint loss-dephasing channel, denoted $\Lambda_{LD,\gamma,\gamma_D}$, gives the solution to the Lindblad master equation:

$$
\frac{d}{dt}\rho(t) = [\kappa \mathcal{D}[\hat{a}] + \kappa_D \mathcal{D}[\hat{n}]][\rho(t)]. \tag{D3}
$$

The fact that the two Lindbladians commute means that the joint loss-dephasing channel can be decomposed into a pure loss channel of strength γ taking place before or after a pure dephasing channel of strength γ_D; i.e., $\Lambda_{LD,\gamma,\gamma_D} = \Lambda_\gamma \circ \Lambda_{D,\gamma_D} = \Lambda_{D,\gamma_D} \circ \Lambda_\gamma$. This in turn means that if our system is subject to a joint loss-dephasing channel of strength (γ, γ_D), then we can use our quantum error cancellation protocol to mitigate against the losses, leaving behind an expectation value that is equivalent if the system was just subject to the pure dephasing channel of strength γ_D. In other words, our error cancellation scheme for photon losses is robust against the introduction of a dephasing error channel.

We now briefly introduce a scheme for mitigating against the pure dephasing channel. A discrete Kraus representation of the dephasing channel is given by

$$
\Lambda_{D,\gamma_D}[\rho] = \sum_{j=0}^{\infty} \frac{\gamma_D^j}{j!} \hat{n}^j e^{-\frac{\gamma_D}{2}\hat{n}^2} \rho \, e^{-\frac{\gamma_D}{2}\hat{n}^2} \hat{n}^j, \tag{D4}
$$

where we have explicitly expanded out the Kraus operators $D_j(\gamma_D) = \sqrt{\gamma_D^j/j!} \, \hat{n}^j e^{-\frac{\gamma_D}{2}\hat{n}^2}$. Using this, one can then easily show that the inverse dephasing channel has an operator-sum representation of

$$
\Lambda_{D,\gamma_D}^{-1}[\rho] = \sum_{j=0}^{\infty} \frac{-\gamma_D^j}{j!} \hat{n}^j e^{\frac{\gamma_D}{2}\hat{n}^2} \rho \, e^{\frac{\gamma_D}{2}\hat{n}^2} \hat{n}^j. \tag{D5}
$$

This corresponds to simply changing the the sign of $\gamma_D \rightarrow -\gamma_D$ in Eq. (D4). Similarly to the photon loss inverse map, Eq. (D5) can be expressed as a weighted sum (with some

negative coefficients) of CP maps. Consider the channels defined by

$$\mathcal{E}_{D,j}[\rho] = \frac{1}{\mathcal{N}_{D,j}(\rho)} \hat{n}^j e^{\frac{\gamma_D}{2}\hat{n}^2} \rho \, e^{\frac{\gamma_D}{2}\hat{n}^2} \hat{n}^j, \qquad (D6)$$

where $\mathcal{N}_{D,j}(\rho)$ is a normalization coefficient. The inverse can then be written as linear sum over these channels with coefficients $\omega_{D,j}(\rho) = (-\gamma_D)^j \mathcal{N}_{D,j}(\rho)/j!$. Using this, if one can realize the channels $\mathcal{E}_{D,j}$ then one can mitigate against the dephasing channel using similar ideas that have been discussed in this paper for mitigating against the loss channel.

Unfortunately, an experimental realization of the channels $\{\mathcal{E}_{D,j}\}$ seems particularly challenging. In fact, a realization of $\hat{n}^j$ alone requires postselection that makes scaling beyond a couple of modes unfeasible. Additionally, unlike the noiseless amplification term present in the inverse excitation loss map, the operator $e^{\frac{\gamma_D}{2}\hat{n}^2}$ may not be easily incorporated in the state preparation stage. Unless a way around these difficulties is found, the error cancellation protocol presented in this paper is unlikely to be experimentally viable to mitigate against dephasing errors.

[1] A. Montanaro, Quantum algorithms: An overview, npj Quantum Inf. **2**, 15023 (2016).

[2] P. W. Shor, Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer, SIAM Rev. **41**, 303 (1999).

[3] L. K. Grover, Quantum mechanics helps in searching for a needle in a haystack, Phys. Rev. Lett. **79**, 325 (1997).

[4] S. Lloyd, Universal quantum simulators, Science **273**, 1073 (1996).

[5] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, New York, 2010).

[6] D. Gottesman, An introduction to quantum error correction and fault-tolerant quantum computation, in *Quantum Information Science and its Contributions to Mathematics*, Proceedings of Symposia in Applied Mathematics Vol. 68 (American Mathematical Society, Washington, DC, 2010), pp. 13–58.

[7] D. A. Lidar and T. A. Brun, *Quantum Error Correction* (Cambridge University Press, New York, 2013).

[8] J. Preskill, Quantum computing in the NISQ era and beyond, Quantum **2**, 79 (2018).

[9] B. Cheng, X.-H. Deng, X. Gu, Y. He, G. Hu, P. Huang, J. Li, B.-C. Lin, D. Lu, Y. Lu *et al.*, Noisy intermediate-scale quantum computers, Front. Phys. **18**, 21308 (2023).

[10] K. Bharti, A. Cervera-Lierta, T. H. Kyaw, T. Haug, S. Alperin-Lea, A. Anand, M. Degroote, H. Heimonen, J. S. Kottmann, T. Menke, W.-K. Mok, S. Sim, L.-C. Kwek, and A. Aspuru-Guzik, Noisy intermediate-scale quantum algorithms, Rev. Mod. Phys. **94**, 015004 (2022).

[11] S. Chen, J. Cotler, H.-Y. Huang, and J. Li, The complexity of NISQ, Nat. Commun. **14**, 6001 (2023).

[12] M. Cerezo, A. Arrasmith, R. Babbush, S. C. Benjamin, S. Endo, K. Fujii, J. R. McClean, K. Mitarai, X. Yuan, L. Cincio *et al.*, Variational quantum algorithms, Nat. Rev. Phys. **3**, 625 (2021).

[13] J. Tilly, H. Chen, S. Cao, D. Picozzi, K. Setia, Y. Li, E. Grant, L. Wossnig, I. Rungger, G. H. Booth *et al.*, The variational quantum eigensolver: A review of methods and best practices, Phys. Rep. **986**, 1 (2022).

[14] S. Endo, J. Sun, Y. Li, S. C. Benjamin, and X. Yuan, Variational quantum simulation of general processes, Phys. Rev. Lett. **125**, 010501 (2020).

[15] A. J. McCaskey, Z. P. Parks, J. Jakowski, S. V. Moore, T. D. Morris, T. S. Humble, and R. C. Pooser, Quantum chemistry as a benchmark for near-term quantum computers, npj Quantum Inf. **5**, 99 (2019).

[16] A. J. Daley, I. Bloch, C. Kokail, S. Flannigan, N. Pearson, M. Troyer, and P. Zoller, Practical quantum advantage in quantum simulation, Nature (London) **607**, 667 (2022).

[17] Y. Kim, A. Eddins, S. Anand, K. X. Wei, E. Van Den Berg, S. Rosenblatt, H. Nayfeh, Y. Wu, M. Zaletel, K. Temme *et al.*, Evidence for the utility of quantum computing before fault tolerance, Nature (London) **618**, 500 (2023).

[18] Z. Cai, R. Babbush, S. C. Benjamin, S. Endo, W. J. Huggins, Y. Li, J. R. McClean, and T. E. O'Brien, Quantum error mitigation, Rev. Mod. Phys. **95**, 045005 (2023).

[19] S. Endo, Z. Cai, S. C. Benjamin, and X. Yuan, Hybrid quantum-classical algorithms and quantum error mitigation, J. Phys. Soc. Jpn. **90**, 032001 (2021).

[20] R. Takagi, S. Endo, S. Minagawa, and M. Gu, Fundamental limits of quantum error mitigation, npj Quantum Inf. **8**, 114 (2022).

[21] Y. Quek, D. S. França, S. Khatri, J. J. Meyer, and J. Eisert, Exponentially tighter bounds on limitations of quantum error mitigation, Nat. Phys. (2024).

[22] A. Kandala, K. Temme, A. D. Córcoles, A. Mezzacapo, J. M. Chow, and J. M. Gambetta, Error mitigation extends the computational reach of a noisy quantum processor, Nature (London) **567**, 491 (2019).

[23] V. Russo, A. Mari, N. Shammah, R. LaRose, and W. J. Zeng, Testing platform-independent quantum error mitigation on noisy quantum computers, IEEE Trans. Quantum Eng. **4**, 2500318 (2023).

[24] C. Song, J. Cui, H. Wang, J. Hao, H. Feng, and Y. Li, Quantum computation with universal error mitigation on a superconducting quantum processor, Sci. Adv. **5**, eaaw5686 (2019).

[25] E. Van Den Berg, Z. K. Minev, A. Kandala, and K. Temme, Probabilistic error cancellation with sparse Pauli-Lindblad models on noisy quantum processors, Nat. Phys. **19**, 1116 (2023).

[26] K. Temme, S. Bravyi, and J. M. Gambetta, Error mitigation for short-depth quantum circuits, Phys. Rev. Lett. **119**, 180509 (2017).

[27] Y. Li and S. C. Benjamin, Efficient variational quantum simulator incorporating active error minimization, Phys. Rev. X **7**, 021050 (2017).

[28] W. J. Huggins, S. McArdle, T. E. O'Brien, J. Lee, N. C. Rubin, S. Boixo, K. B. Whaley, R. Babbush, and J. R. McClean, Virtual distillation for quantum error mitigation, Phys. Rev. X **11**, 041036 (2021).

[29] B. Koczor, Exponential error suppression for near-term quantum devices, Phys. Rev. X **11**, 031057 (2021).

[30] S. Endo, S. C. Benjamin, and Y. Li, Practical quantum error mitigation for near-future applications, Phys. Rev. X **8**, 031027 (2018).

[31] C. Piveteau, D. Sutter, and S. Woerner, Quasiprobability decompositions with reduced sampling overhead, npj Quantum Inf. **8**, 12 (2022).

[32] Y. Suzuki, S. Endo, K. Fujii, and Y. Tokunaga, Quantum error mitigation as a universal error reduction technique: Applications from the NISQ to the fault-tolerant quantum computing eras, PRX Quantum **3**, 010345 (2022).

[33] R. S. Gupta, E. van den Berg, M. Takita, K. Temme, and A. Kandala, Probabilistic error cancellation for dynamic quantum circuits, Phys. Rev. A **109**, 062617 (2024).

[34] D. Su, R. Israel, K. Sharma, H. Qi, I. Dhand, and K. Brádler, Error mitigation on a near-term quantum photonic device, Quantum **5**, 452 (2021).

[35] T.-Y. Yang, Y.-X. Shen, Z.-K. Cao, and X.-B. Wang, Postselection in noisy Gaussian boson sampling: Part is better than whole, Quantum Sci. Technol. **8**, 045020 (2023).

[36] S. Aaronson and A. Arkhipov, The computational complexity of linear optics, Theor. Comput. **9**, 143 (2013).

[37] A. P. Lund, A. Laing, S. Rahimi-Keshari, T. Rudolph, J. L. O'Brien, and T. C. Ralph, Boson sampling from a Gaussian state, Phys. Rev. Lett. **113**, 100502 (2014).

[38] C. S. Hamilton, R. Kruse, L. Sansoni, S. Barkhofen, C. Silberhorn, and I. Jex, Gaussian boson sampling, Phys. Rev. Lett. **119**, 170501 (2017).

[39] J. Mills and R. Mezher, Mitigating photon loss in linear optical quantum circuits: Classical postprocessing methods outperforming postselection, arXiv:2405.02278.

[40] K. Tsubouchi, T. Sagawa, and N. Yoshioka, Universal cost bound of quantum error mitigation based on quantum estimation theory, Phys. Rev. Lett. **131**, 210601 (2023).

[41] R. Takagi, H. Tajima, and M. Gu, Universal sampling lower bounds for quantum error mitigation, Phys. Rev. Lett. **131**, 210602 (2023).

[42] The set of noisy operators $\{\mathcal{F}_j\}$ should form a noisy operator basis into which $\mathcal{U}_{\text{ideal}}$ can be decomposed.

[43] J. Sun, X. Yuan, T. Tsunoda, V. Vedral, S. C. Benjamin, and S. Endo, Mitigating realistic noise in practical noisy intermediate-scale quantum devices, Phys. Rev. Appl. **15**, 034026 (2021).

[44] M. S. Kim and N. Imoto, Phase-sensitive reservoir modeled by beam splitters, Phys. Rev. A **52**, 2401 (1995).

[45] R. J. Glauber, Coherent and incoherent states of the radiation field, Phys. Rev. **131**, 2766 (1963).

[46] M. Oszmaniec and D. J. Brod, Classical simulation of photonic linear optics with lost particles, New J. Phys. **20**, 092002 (2018).

[47] W. R. Clements, P. C. Humphreys, B. J. Metcalf, W. S. Kolthammer, and I. A. Walmsley, Optimal design for universal multiport interferometers, Optica **3**, 1460 (2016).

[48] D. J. Brod and M. Oszmaniec, Classical simulation of linear optics subject to nonuniform losses, Quantum **4**, 267 (2020).

[49] S. Rahimi-Keshari, M. A. Broome, R. Fickler, A. Fedrizzi, T. C. Ralph, and A. G. White, Direct characterization of linear-optical networks, Opt. Express **21**, 13450 (2013).

[50] I. Dhand, A. Khalid, H. Lu, and B. C. Sanders, Accurate and precise characterization of linear optical interferometers, J. Opt. **18**, 035204 (2016).

[51] T. C. Ralph and A. Lund, Nondeterministic noiseless linear amplification of quantum systems, AIP Conf. Proc. **1110**, 155 (2009).

[52] M. Kim, Recent developments in photon-level operations on travelling light fields, J. Phys. B **41**, 133001 (2008).

[53] A. Zavatta, J. Fiurášek, and M. Bellini, A high-fidelity noiseless amplifier for quantum light states, Nat. Photon. **5**, 52 (2011).

[54] M. S. Winnel, N. Hosseinidehaj, and T. C. Ralph, Generalized quantum scissors for noiseless linear amplification, Phys. Rev. A **102**, 063715 (2020).

[55] J. J. Guanzon, M. S. Winnel, A. P. Lund, and T. C. Ralph, Ideal quantum teleamplification up to a selected energy cutoff using linear optics, Phys. Rev. Lett. **128**, 160501 (2022).

[56] J. J. Guanzon, M. S. Winnel, D. Singh, A. P. Lund, and T. C. Ralph, Saturating the maximum success probability bound for noiseless linear amplification using linear optics, PRX Quantum **5**, 020359 (2024).

[57] Z. Cai, A practical framework for quantum error mitigation, arXiv:2110.05389.

[58] Since $g^{\hat{n}}$ is an unbounded operator, we cannot expect it to produce physical states when acting on arbitrary CV states. This is a generic property of unbounded operators acting on states in an infinite-dimensional Hilbert space.

[59] C. M. Nunn, S. U. Shringarpure, and T. B. Pittman, Transforming photon statistics through zero-photon subtraction, Phys. Rev. A **107**, 043711 (2023).

[60] 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).

[61] A. Serafini, *Quantum Continuous Variables: A Primer of Theoretical Methods* (CRC Press, Boca Raton, FL, 2017).

[62] 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).

[63] This choice of μ was selected because, to the nearest 0.005, it optimized the sampling overhead.

[64] A. Royer, Wigner function as the expectation value of a parity operator, Phys. Rev. A **15**, 449 (1977).