A neural processing approach to quantum state discrimination

Saeed A. Khan,[1] Fangjun Hu,[1] Gerasimos Angelatos,[1] Michael Hatridge,[2] and Hakan E. Türeci[1]

[1]*Department of Electrical and Computer Engineering, Princeton University, Princeton, NJ 08544*
[2]*Department of Physics and Astronomy, University of Pittsburgh, Pittsburgh*
(Dated: September 10, 2024)

Although linear quantum amplification has proven essential to the processing of weak quantum signals, extracting higher-order quantum features such as correlations in principle demands nonlinear operations. However, nonlinear processing of quantum signals is often associated with non-idealities and excess noise, and absent a general framework to harness nonlinearity, such regimes are typically avoided. Here we present a framework to uncover general quantum signal processing principles of a broad class of bosonic quantum nonlinear processors (QNPs), inspired by a remarkably analogous paradigm in nature: the processing of environmental stimuli by nonlinear, noisy neural ensembles, to enable perception. Using a quantum-coherent description of a QNP monitoring a quantum signal source, we show that quantum nonlinearity can be harnessed to calculate higher-order features of an incident quantum signal, concentrating them into linearly-measurable observables, a transduction not possible using linear amplifiers. Secondly, QNPs provide coherent nonlinear control over quantum fluctuations including their own added noise, enabling noise suppression in an observable without suppressing transduced information, a paradigm that bears striking similarities to optimal neural codings that allow perception even under highly stochastic neural dynamics. Unlike the neural case, we show that QNP-engineered noise distributions can exhibit non-classical correlations, providing a new means to harness resources such as entanglement. Finally, we show that even simple QNPs in realistic measurement chains can provide enhancements of signal-to-noise ratio for practical tasks such as quantum state discrimination. Our work provides pathways to utilize nonlinear quantum systems as general computation devices, and enables a new paradigm for nonlinear quantum information processing.

I. INTRODUCTION

While engineering quantum systems hinges on isolating quantum components from their noisy classical environment, the observation of any such system necessitates extracting its emitted weak quantum signals back to the classical world. It is therefore essential to bolster quantum signals prior to their interaction with noisy classical modes and the readout noise they introduce, using signal processors that are themselves quantum-mechanical [1]. The most widely-used such processors are quantum amplifiers [2–5], which provide linear gain to an input signal and have been critical to quantum signal processing applications such as high-fidelity quantum state readout [6, 7] and generation of non-classical light [8, 9]. In spite of this success, the linearity of quantum amplifiers means they are only able to reveal a subset of the features of a complex quantum signal. A well-known limitation is the estimation of higher-order correlations. When only linear readout is available, the nonlinearity demanded by correlation calculations must be provided by classical post-processing: this process amplifies readout noise, leading to a signal-to-noise ratio that degrades exponentially with the order of the desired correlation function [10–13]. For signals that exhibit non-classical, higher-order quantum correlations [12, 14], linear processing restricts our ability to fully resolve the quantum domain, thereby limiting our capacity to control it.

Interestingly, even though a rich expanse of quantum devices can be realized beyond the confines of linearity [15], the use of nonlinear quantum systems as general purpose quantum signal processors has been limited. Partly, this is because there are no established general principles of *quantum* information processing with nonlinear quantum systems. The very fact that nonlinear systems are *a priori* not bound by constraints of a linear dependence between input and output - which could hold the key to new ways of processing quantum signals - makes their operation much more complicated to analyze. More importantly, a complete description of the operation of any general quantum processor must be constrained by quantum mechanics, namely the uncertainty principle [16]. For linear quantum amplifiers, this translates to well-known limits on noise added during the amplification process [1, 3, 17]. Analogous limits are much less explored for nonlinear amplifiers (with some exceptions [18–20]), and nonlinear regimes are often even associated with excess noise [12, 21]. As a result, in spite of some measurement schemes effectively utilizing nonlinear systems [22–24], nonlinearity is often considered an inconvenience in quantum signal processing, being the source of gain saturation [4, 12, 25, 26] and operational instabilities [4], which has led to significant efforts to mitigate nonlinear effects [27, 28].

Remarkably, although the processing of weak stochastic signals using nonlinear devices is still not thoroughly explored in the quantum domain, this paradigm is fundamental to perhaps the most well-known information processor: the brain. It is understood that neurons, the signalling cells that constitute all nervous matter, function as part of large neural ensembles whose collective firing properties can be highly nonlinear [29] and exhibit great stochasticity [30, 31]. Experiments recording the activ-

ity of neural ensembles in the visual cortex of animals have provided robust evidence that neuronal response exhibits significant variability even under repeated presentations of identical stimuli (see Fig. 1(a)), with the total noise power even increasing with the size of the neural ensemble [32]. Nevertheless, such noisy ensembles are able to successfully process distinct stimuli to enable perception [33, 34]. Research over the past few decades has begun to establish an explanation [35]. Nature appears to prefer encoding cortical signals in the response of collective neuron modes, as opposed to highly-variable single neuron dynamics (referred to as *population* coding [36–38]). Furthermore, by exploiting the nonlinear response of neurons [39], the encoding of sensory information in collective neuron modes (or coding directions) that have the largest noise power is avoided [32, 40]. While neural noise still places limits on sensory signal discrimination [41–43], this coding principle allows the perception of weaker stimuli than would be resolvable if the largest noise mode was used for coding. This mechanism provides an example of how Nature is able to harness nonlinearity to facilitate signal processing in the presence of seemingly overwhelming stochasticity.

Motivated by the example of stochastic neural processing, we therefore ask the question: can *nonlinear*, fundamentally stochastic quantum systems provide an advantage in the detection and processing of weak quantum signals *beyond the paradigm of linear amplification*? If so, can any general principles be established? Answering this question in its most general setting requires a description of the joint dynamics of the quantum nonlinear processor (QNP) and the signal generating quantum system (QS), often excluded in descriptions of linear quantum amplifiers [1] (with exceptions e.g. Ref. [44]). Here we develop the necessary framework to model a QNP as an *in situ* processor of signals generated by a quantum system (QS) it is coherently linked to (see Fig. 1(b)); our approach thus accounts for the QNP nonlinearity (with some restrictions), the QS-QNP interaction (including the possibility of their entanglement), but most importantly for not only the quantum fluctuations of the QNP itself, but also those due to the quantum nature of the input signals. Building on this framework, our analysis is able to identify two extremely general aspects of quantum information processing enabled by QNPs, which we summarize here. First, by characterizing their now nonlinear input-output map, we show that QNPs enable *higher-order transduction*: the mapping of higher-order observables, such as the correlation of quantum signals, to lower-order observables such as quadrature means, making them accessible using much simpler linear readout protocols. Secondly, QNPs are able to *coherently manipulate* incident quantum fluctuations so as to direct amplified noise to observables that do not encode the transduced signal. This feature, which is particularly reminiscent of population coding in noisy neural ensembles, relies on both the QNP nonlinearity and its deployment as a *quantum* processor.
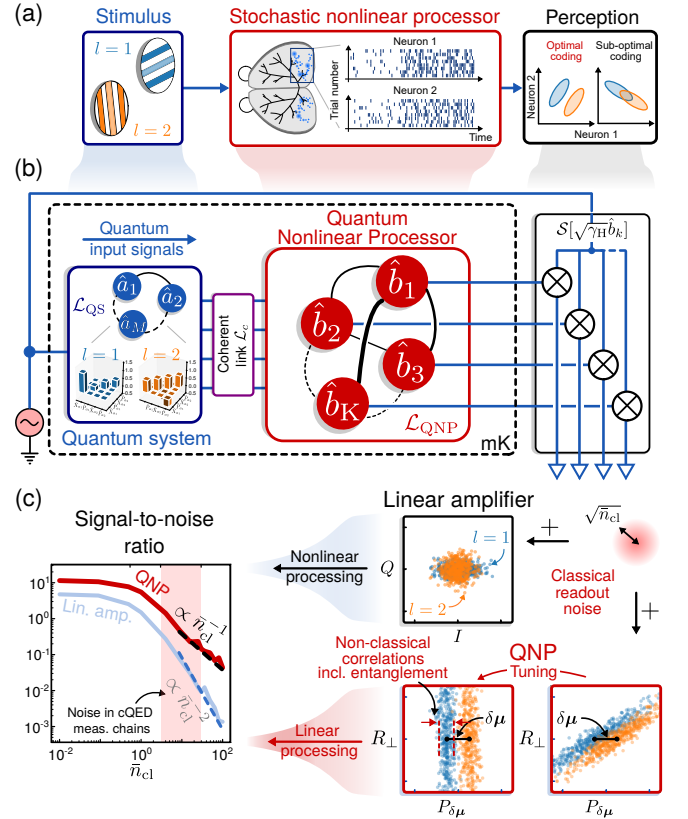

Figure 1. (a) Schematic representation of processing of visual signals by the brain. Visual stimuli lead to firing activity in neural ensembles in the visual cortex; the activity is highly stochastic even for identical stimuli, but can be strongly correlated for distinct neurons. Optimal neural codings encode visual information to avoid large noise in neuronal dynamics, whereas a sub-optimal encoding is more susceptible to noise. (b) Schematic representation of processing of quantum signals using a quantum nonlinear processor (QNP). Quantum signals are generated by a quantum system (QS) governed by Liouvillian $\mathcal{L}_{\mathrm{QS}}$, and are incident on the QNP via a coherent link $\mathcal{L}_c$. We demonstrate the QNP for the task of discriminating two states $\hat{\rho}_{\mathrm{QS}}^{(l)}$ of the QS using heterodyne readout of the QNP, described by $\mathcal{S}[\sqrt{\gamma_{\mathrm{H}}}\hat{b}_k]$. (c) The QNP enables such tasks using only linear processing of measured outputs, in contrast to linear quantum amplifiers, which require nonlinear post-processing.

To demonstrate the above results, we show QNPs at work for a practical quantum information processing task inspired by Gaussian state tomography [45]: the binary discrimination of states for which the quadratures of constituent QS modes possess identical mean values, but differ only in their quantum fluctuation signatures (Fig. 1(b)). These signatures may be evident in the quantum fluctuations of signals emanating from a single QS mode, or may only be revealed in non-classical noise correlations across distinct modes, such as for entangled states. When signals from a QS in such states are processed using standard linear amplifiers and heterodyne readout, the limitations of a linear input-output

map ensure that obtained output distributions exhibit no difference in their mean values (Fig. 1(c), top). As a result, the distributions are impossible to distinguish perfectly using only a linear decision boundary.

In contrast, we show that QNPs are able to transduce the quantum fluctuation signatures of incident signals into displacements observable via heterodyne measurement, effectively functioning as quantum 'cross-correlators' across multiple QS modes. As a result the obtained output distributions (Fig. 1(c), bottom) exhibit a nonzero mean separation $\delta\boldsymbol{\mu}$, enabling a linear classifier to separate them perfectly. Most importantly, QNPs enable all the nonlinear processing demanded by this discrimination task to be performed on pristine quantum signals, prior to their corruption by classical readout noise $\bar{n}_{\rm cl}$. Consequently, the signal-to-noise ratio for QNP-enabled discrimination degrades only as $\bar{n}_{\rm cl}^{-1}$. In contrast, with standard linear amplifiers a nonlinear post-processing step must be performed on the final classical signals; this worsens readout noise, furnishing a signal-to-noise ratio that scales as $\bar{n}_{\rm cl}^{-2}$. We note that the concept of transduction using nonlinear systems has been explored previously [19], often via a lossless Hamiltonian description; here we consider a much more general setting, and account for practical constraints of losses and measurement. Secondly, here QNPs are deployed as nonlinear processors of stochastic *quantum* signals; this paradigm is therefore distinct from the more routinely explored use of quantum systems to compute nonlinear functions of deterministic *classical* signals.

Finally, we show how QNPs can control quantum fluctuations for information processing in ways unavailable to their linear counterparts. Linear amplifiers are restricted to perform the same transformation on an incident signal and its noise; for a fixed input signal, therefore, a decrease in the magnitude of the amplifier output noise is accompanied by a decrease in the magnitude of the output signal, and vice versa. Unconstrained by linearity, QNPs enable a much more complex interplay between the transduced signal and the output noise. In particular, we show how tuning QNP parameters alone, the output noise in the signal carrying quadrature $P_{\delta\boldsymbol{\mu}}$ (the unique quadrature parallel to $\delta\boldsymbol{\mu}$) can be decreased while the transduced signal magnitude $||\delta\boldsymbol{\mu}||$ remains fixed (Fig. 1(c), bottom). Most strikingly, a tuning condition can be reached such that the signal carrying quadrature is parallel to the quadrature with minimal output noise, and thus orthogonal to all other quadrature combinations which possess larger noise, a scenario that appears very similar to optimal coding in neural circuits. Crucially, for QNPs processing quantum input signals, the large noise modes include quantum fluctuations generated by the upstream QS, but also fluctuations amplified by the QNP during processing itself, referred to as 'added noise'. The minimum noise mode can even reach non-classical (i.e. sub-vacuum) levels, in the presence of squeezing or entanglement. Therefore the ability to control noise using QNPs provides a practical means to harness non-classical

correlations for quantum state discrimination.

A quantum-coherent description of a QS and QNP in the same measurement chain, valid for multi-mode systems, and across a range of excitation conditions and interaction types, demands multiple theoretical techniques. Our key analytic results are built upon a nonlinear van Kampen expansion [46] of the Fokker-Planck equation associated with the quantum state of the complete measurement chain. At the expense of being restricted to weakly nonlinear QNPs, this approach has the advantage of applying to arbitrarily-multimode bosonic quantum systems. As such, we expect the uncovered operating principles to hold for general QNPs under heterodyne monitoring, especially relevant to cQED. These results are numerically verified and extended to stronger nonlinearities using a truncated cumulants approach that simulates measurement-conditioned dynamics of the measurement chain, with a complexity that is only quadratic in the number of total modes. Supplementing these results via full (stochastic) master equation simulations and exact results for select few-mode, multi-system measurement chains [47], we are able to provide a comprehensive description of quantum information processing using QNPs.

The remainder of this paper is organized as follows. In Sec. II, we introduce our model for processing quantum signals using a QNP, provide a summary of the key results of our theoretical approach, and introduce the particular classification tasks we analyze. Sec. III then analyzes quantum state classification using a single-mode QNP, based on a model that can be easily realized in cQED experiments. We highlight the crucial role of QNP nonlinearity and discuss how QNP parameters enable control over quantum fluctuations, including the ability to harness non-classical correlations. In Sec. IV we provide a comparison of quantum state measurement using a QNP against standard linear quantum amplifiers, finding improved robustness to added classical noise for the QNP-based scheme. Sec. V then considers quantum state classification tasks that require the non-local processing of quantum fluctuations, where a multi-mode QNP becomes essential. Here, we demonstrate how output field entanglement can be engineered using a QNP to produce sub-vacuum noise in a multi-mode quadrature. The paper concludes with a discussion of possible applications of QNPs in the quantum information processing landscape.

II. QUANTUM NONLINEAR PROCESSORS FOR QUANTUM SIGNALS

A. Quantum measurement chain for processing quantum signals

The complete measurement chain we consider is depicted in Fig. 1(b). The conditional evolution of the quantum state $\hat{\rho}^c$ of the chain under continuous measurement is formally described by the stochastic master

equation (SME)

$$d\hat{\rho}^c = (\mathcal{L}_{\mathrm{QS}} + \mathcal{L}_c + \mathcal{L}_{\mathrm{QNP}}) \hat{\rho}^c \, dt + \sum_k \mathcal{S}[\sqrt{\gamma_{\mathrm{H}}}\hat{b}_k]\hat{\rho}^c. \quad (1)$$

The superoperator $\mathcal{L}_{\mathrm{QS}}$ describes a quantum system (QS) that is the source of quantum signals to be processed for a given quantum information processing task. We consider the form

$$\mathcal{L}_{\mathrm{QS}}\hat{\rho} = -i\big[\hat{\mathcal{H}}_{\mathrm{QS}}^{(l)}, \hat{\rho}\big] + \sum_m \kappa_m \mathcal{D}[\hat{a}_m]\hat{\rho}, \quad (2)$$

where $\hat{\mathcal{H}}_{\mathrm{QS}}$ is a linear Hamiltonian governing the dynamics of M bosonic modes $\hat{a}_m, m \in [1, M]$, and experiencing damping $\{\kappa_m\}$. The superscript (l) indexes distinct states of the QS, to be distinguished in a classification task.

This QS is coupled to the quantum nonlinear processor (QNP) in the *same measurement chain* (via a non-reciprocal coupling described by $\mathcal{L}_c$, see Appendix A), enabling the QNP to monitor the QS as an *in situ* measurement apparatus. The QNP is governed by $\mathcal{L}_{\mathrm{QNP}}$,

$$\mathcal{L}_{\mathrm{QNP}}\hat{\rho} = -i\big[\hat{\mathcal{H}}_{\mathrm{QNP}} + \hat{\mathcal{N}}_{\mathrm{QNP}}, \hat{\rho}\big], \quad (3)$$

where $\hat{\mathcal{H}}_{\mathrm{QNP}}$ is a general linear Hamiltonian governing the dynamics of K bosonic modes $\hat{b}_k, k \in [1, K]$. The QNP nonlinearity can then be furnished by any general bosonic nonlinear interaction; for simplicity we consider a Hamiltonian describing K Kerr modes with common nonlinearity strength Λ, $\hat{\mathcal{N}}_{\mathrm{QNP}} = -\sum_k \frac{\Lambda}{2}\hat{b}_k^\dagger\hat{b}_k^\dagger\hat{b}_k\hat{b}_k$.

The chosen measurement scheme is of the standard heterodyne type: it comprises continuous monitoring of the QNP decay channels (with unit efficiency). The resulting measurement-conditioned evolution of the *complete* measurement chain in accordance with quantum measurement theory [48] is governed by the stochastic heterodyne measurement superoperator $\mathcal{S}[\sqrt{\gamma_{\mathrm{H}}}\hat{b}_k]$, with a rate γ_{H} assumed equal for all modes (see Appendix A). This yields heterodyne records $\mathcal{I}_k(t), \mathcal{Q}_k(t)$ for each monitored QNP mode [48]:

$$\mathcal{I}_k(t) = \xi_{\mathcal{I}_k}(t) + \sqrt{\gamma_{\mathrm{H}}}\left[\langle \hat{X}_k \rangle + \xi_{\mathcal{I}_k}^{\mathrm{qm}}(t)\right] + \sqrt{\bar{n}_{\mathrm{cl}}}\,\xi_{\mathcal{I}_k}^{\mathrm{cl}}(t),$$
$$\mathcal{Q}_k(t) = \xi_{\mathcal{Q}_k}(t) + \sqrt{\gamma_{\mathrm{H}}}\left[\langle \hat{P}_k \rangle + \xi_{\mathcal{Q}_k}^{\mathrm{qm}}(t)\right] + \sqrt{\bar{n}_{\mathrm{cl}}}\,\xi_{\mathcal{Q}_k}^{\mathrm{cl}}(t).$$
$$(4)$$

Heterodyne monitoring probes the canonically-conjugate QNP quadratures $\hat{X}_k = \frac{1}{\sqrt{2}}(\hat{b}_k + \hat{b}_k^\dagger)$, $\hat{P}_k = -\frac{i}{\sqrt{2}}(\hat{b}_k - \hat{b}_k^\dagger)$, but a given measurement record also contains contributions from multiple noise sources ξ. The terms $\xi_{\mathcal{I}_k}, \xi_{\mathcal{Q}_k}$ describe vacuum noise that would be present even if no signal was emanating from the monitored quantum modes (i.e. $\gamma_{\mathrm{H}} \to 0$). The much more interesting terms are marked 'qm': these describe noise contributions of a quantum origin, such as non-classical correlations due to squeezing or entanglement, or added noise by quantum

dynamics, and are contingent on the measurement superoperator $\mathcal{S}$ (see Appendix A). In contrast, the terms 'cl' define classical readout noise in the measurement chain; these are not associated with a stochastic measurement superoperator and hence have no backaction on the quantum measurement chain. Equivalently, $\bar{n}_{\mathrm{cl}}$ quantifies noise added after the so-called *Heisenberg-von Neumann cut* [48, 49].

These stochastic records are thus often filtered to obtained heterodyne quadratures $(I_k, Q_k) \equiv \int_{t_0}^{t_0+\mathcal{T}} d\tau \, \mathcal{K}(\tau) \times (\mathcal{I}_k(\tau), \mathcal{Q}_k(\tau))$, where $\mathcal{K}(\tau)$ is the filter function (we assume a boxcar filter $\mathcal{K}(\tau) = \frac{1}{\sqrt{2\mathcal{T}}} \, \forall \, \tau$) over a window of length $\mathcal{T}$ starting from an initial time t_0. The quadratures can be compactly represented via the vector $\boldsymbol{x} = (I_1, Q_1, \ldots, I_K, Q_K)^T \in \mathbb{R}^{2K}$. For any quantum information processing task, the data $\boldsymbol{x}$ is typically further processed to obtain a vector $\boldsymbol{y}$ of output features, a step we define generally via $\boldsymbol{y} = F[\boldsymbol{x}]$, which can include ensemble averaging over distinct shots, but also nonlinear processing to be clarified in due course. The complete measurement chain thus described is designed to measure QS properties via the obtained outputs $\boldsymbol{y}$; note that our description subsumes the standard paradigm of linear quantum amplification plus heterodyne measurement if $\hat{\mathcal{N}}_{\mathrm{QNP}} \to 0$ and an appropriate choice of $\hat{\mathcal{H}}_{\mathrm{QNP}}$ is made.

We can now specialize to the consideration of binary quantum state discrimination (although QNPs can be deployed for more general processing tasks as well). Binary discrimination has a simple objective: to distinguish QS state $\rho_{\mathrm{QS}}^{(l)}$ from $\rho_{\mathrm{QS}}^{(p)}$, where $\rho_{\mathrm{QS}}^{(l)} \equiv \mathrm{tr}_{\mathrm{QNP}}[\hat{\rho}^{(l)}]$, based on the corresponding outputs $\boldsymbol{y}^{(l)}$ and $\boldsymbol{y}^{(p)}$ from the measurement chain. To this end, we introduce a standard measure of the distinguishability of the two measured distributions of $\boldsymbol{y}^{(l)}$ and $\boldsymbol{y}^{(p)}$: Fisher's discriminant D_{F}, defined as

$$D_{\mathrm{F}}(\boldsymbol{y}) = \delta\boldsymbol{\mu}^T \cdot \mathbf{V}^{-1} \cdot \delta\boldsymbol{\mu}, \quad (5)$$

where $\delta\boldsymbol{\mu} = \boldsymbol{\mu}^{(l)} - \boldsymbol{\mu}^{(p)}$ is the difference of means of the two measured distributions, while $\mathbf{V} = \frac{1}{2}\left(\boldsymbol{\Sigma}^{(l)} + \boldsymbol{\Sigma}^{(p)}\right)$ is a measure of their combined variance:

$$\boldsymbol{\mu}^{(l)} = \mathbb{E}\left[\boldsymbol{y}^{(l)}\right], \ \boldsymbol{\Sigma}^{(l)} = \mathbb{E}\left[\boldsymbol{y}^{(l)}\boldsymbol{y}^{(l)T}\right] - \mathbb{E}\left[\boldsymbol{y}^{(l)}\right]\mathbb{E}\left[\boldsymbol{y}^{(l)}\right]^T.$$
$$(6)$$

Note that D_{F} has the intuitive form of a generalized signal-to-noise ratio (SNR). Furthermore, the fidelity $\mathcal{C}$ of classifying two Gaussian distributions with identical covariance matrices $\boldsymbol{\Sigma}^{(1)} = \boldsymbol{\Sigma}^{(2)}$, is simply $\mathcal{C} = \frac{1}{2}(1 + \mathrm{erf}\frac{D_{\mathrm{F}}}{2\sqrt{2}})$, where $\mathrm{erf}\,z$ is the standard Gaussian error function; as $D_{\mathrm{F}} \to \infty$, $\mathcal{C} \to 1$. For binary QS state discrimination, the D_{F} of measured distributions is correlated with the fidelity of discriminating the two QS states that give rise to these distributions.

Using such binary classification tasks, where the concepts of signal $\delta\boldsymbol{\mu}$ and noise $\mathbf{V}$ are precisely defined, we will demonstrate how QNPs can process quantum signals

in situ for practical quantum information processing applications.

B. Approximate input-output map for quantum nonlinear processors

In the context of binary state discrimination, the role of QNPs can be characterized by evaluating the central quantities $\delta\boldsymbol{\mu}$ and $\mathbf{V}$ identified in the previous section. However, the nonlinearity introduced by the QNP - central to its operation - generally excludes exact theoretical treatments that would be valid if the measurement chain comprised only linear bosonic modes and interactions. Furthermore, allowing the QS and QNP to comprise arbitrary numbers of modes renders exact numerical integration of Eq. (1) unfeasible for all but the lowest excitation numbers.

Our solution begins by introducing an alternate set of dynamical variables to describe the quantum state $\hat{\rho}$ of the entire nonlinear measurement chain: *quantum cumulants* (see Appendix B). Formally, cumulants are an infinite set of dynamical variables parameterizing $\hat{\rho}$, indexed by an integer order $n_{\mathrm{ord}} \in \mathbb{Z}^+$. Crucially, however, we show that retaining cumulants only up to a certain *finite* order $n_{\mathrm{ord}} \leq n_{\mathrm{trunc}}$ can very accurately describe the nonlinear measurement chains we consider, contingent on the nonlinearity strength. In this work, we show that $n_{\mathrm{trunc}} = 2$ is an excellent approximation provided certain well-understood and achievable constraints on the nonlinearity strength are met. In this case, defining $\hat{\boldsymbol{b}} \equiv (\hat{b}_1, \hat{b}_1^\dagger, \ldots, \hat{b}_K, \hat{b}_K^\dagger)^T$ and $\hat{\boldsymbol{a}}$ analogously for QS operators, the retained first-order cumulants are simply single-operator expectation values $\begin{pmatrix} \langle \hat{\boldsymbol{a}} \rangle \\ \langle \hat{\boldsymbol{b}} \rangle \end{pmatrix}$, while second-order cumulants are *normal-ordered* covariances, $\mathbf{C} = \langle : \begin{pmatrix} \hat{\boldsymbol{a}} \\ \hat{\boldsymbol{b}} \end{pmatrix} (\hat{\boldsymbol{a}}^T \ \hat{\boldsymbol{b}}^T) : \rangle - \langle \begin{pmatrix} \hat{\boldsymbol{a}} \\ \hat{\boldsymbol{b}} \end{pmatrix} \rangle \langle (\hat{\boldsymbol{a}}^T \ \hat{\boldsymbol{b}}^T) \rangle$. This constitutes a *truncated cumulants* ansatz for $\hat{\rho}$: an efficient description of multimode nonlinear quantum dynamics, with the number of retained cumulants scaling only quadratically, and not exponentially, with the total number of modes $R = K + M$ for $n_{\mathrm{trunc}} = 2$.

We then develop an approximation that allows the truncated cumulants to be solved for analytically, formally employing a nonlinear van Kampen (NVK) expansion [46] in the QNP nonlinearity (also referred to as a 'system size' expansion, see full details in SI Sec. G). Under the NVK approximation, first-order cumulants are written as $\begin{pmatrix} \langle \hat{\boldsymbol{a}} \rangle \\ \langle \hat{\boldsymbol{b}} \rangle \end{pmatrix} = \bar{\Lambda}^{-\frac{1}{2}} \begin{pmatrix} \langle \bar{\hat{\boldsymbol{a}}} \rangle \\ \langle \bar{\hat{\boldsymbol{b}}} \rangle \end{pmatrix} + \begin{pmatrix} \langle \delta \hat{\boldsymbol{a}} \rangle \\ \langle \delta \hat{\boldsymbol{b}} \rangle \end{pmatrix}$ where we have introduced the dimensionless nonlinearity $\bar{\Lambda} = \frac{\Lambda}{\gamma}$, and γ, the QNP mode decay rate, serves as a normalization factor. $\bar{\Lambda} \to 0$ describes the classical (large occupation) limit, where $\langle \bar{\hat{\boldsymbol{a}}} \rangle$, $\langle \bar{\hat{\boldsymbol{b}}} \rangle$ become the dominant contributions to first-order cumulants; these define the expansion point

and satisfy the equations of motion:

$$d_t \langle \bar{\hat{\boldsymbol{a}}} \rangle = \mathbf{L}_a^{(l)} \langle \bar{\hat{\boldsymbol{a}}} \rangle + \bar{\Lambda}^{\frac{1}{2}} \vec{\eta}^{(l)} \tag{7a}$$

$$d_t \langle \bar{\hat{\boldsymbol{b}}} \rangle = \mathbf{L}_b \langle \bar{\hat{\boldsymbol{b}}} \rangle + \vec{N}_b(\langle \bar{\hat{\boldsymbol{b}}} \rangle) - \vec{\eta}_b^{(l)} \tag{7b}$$

These equations immediately provide useful insight. Being linear, the QS response is governed entirely by $\mathbf{L}_a$ and possible coherent drives $\vec{\eta}^{(l)}$, as dictated by $\hat{\mathcal{H}}_{\mathrm{QS}}^{(l)}$ (and dissipative terms). The non-reciprocal QS→QNP interaction defined by $\mathcal{L}_c$ ensures that the QS drives the QNP via the coupling $\boldsymbol{\Gamma}$ and not vice-versa, leading to an effective QS state-dependent drive $\vec{\eta}_b^{(l)} \equiv \boldsymbol{\Gamma} \langle \bar{\hat{\boldsymbol{a}}}^{(l)} \rangle$ on the QNP. The QNP dynamics, in contrast, contain both a linear contribution $\mathbf{L}_b$ as well as a nonlinear contribution $\vec{N}_b(\langle \bar{\hat{\boldsymbol{b}}} \rangle)$, the latter determined by the nonlinear Hamiltonian $\hat{\mathcal{N}}_{\mathrm{QNP}}$. Thus Eq. (7b) is nonlinear in $\langle \bar{\hat{\boldsymbol{b}}} \rangle$.

The full quantum state further requires specifying the deviation $\langle \delta \hat{\boldsymbol{b}} \rangle$ (we can show that $\langle \delta \hat{\boldsymbol{a}} \rangle \to 0$) and the second-order cumulants $\mathbf{C}$. Starting with the latter and introducing the block form $\mathbf{C} = \begin{pmatrix} \mathbf{C}_a & \mathbf{C}_{ab} \\ \mathbf{C}_{ab}^T & \mathbf{C}_b \end{pmatrix}$, we show that second-order cumulants satisfy the Lyapunov differential equation:

$$d_t \mathbf{C} = \left[\mathbf{C} \begin{pmatrix} \mathbf{L}_a & \mathbf{0} \\ -\boldsymbol{\Gamma} & \bar{\mathbf{J}}_b \end{pmatrix} + m.t. \right] + \begin{pmatrix} \mathbf{D}_a^{(l)} & \mathbf{0} \\ \mathbf{0} & \bar{\mathbf{D}}_b \end{pmatrix}, \tag{8}$$

where $m.t.$ is the (unconjugated) matrix transpose. Here $\bar{\mathbf{J}}_b$ is the Jacobian matrix of the QNP, $[\bar{\mathbf{J}}_b]_{ij} = [\mathbf{L}_b]_{ij} + \frac{\partial [\vec{N}_b]_i}{\partial \langle \bar{\hat{\boldsymbol{b}}} \rangle_j}$, with $[\cdot]_{ij}$ indicating tensor notation, and $(\bar{\cdot})$ specifying evaluation at the expansion point $\langle \bar{\hat{\boldsymbol{b}}} \rangle$. Then $\mathbf{D}_a$, $\bar{\mathbf{D}}_b$ are diffusion matrices for the QS and QNP respectively; these describe incident fluctuations - quantum or classical - *beyond* vacuum fluctuations, and must be non-vanishing to yield nontrivial $\mathbf{C}$.

Finally, we introduce the most important dynamical equation in the NVK approximation, for $\langle \delta \hat{\boldsymbol{b}} \rangle$:

$$d_t \langle \delta \hat{\boldsymbol{b}} \rangle = \bar{\mathbf{J}}_b \langle \delta \hat{\boldsymbol{b}} \rangle + \bar{\Lambda}^{\frac{1}{2}} \bar{\mathbf{H}}_b : \mathbf{C}_b^{(l)}. \tag{9}$$

Crucially, the change in first-order QNP cumulants depends on the second-order cumulants, via the final term. This dependence is quantified by the *Hessian tensor* $\bar{\mathbf{H}}_b$ of the QNP, defined as $[\bar{\mathbf{H}}_b]_{ijk} = \frac{\partial [\vec{N}_b]_i}{\partial \langle \bar{\hat{\boldsymbol{b}}} \rangle_j \partial \langle \bar{\hat{\boldsymbol{b}}} \rangle_k}$. The Hessian operates on matrices - here on $\mathbf{C}_b$ - to return a vector, via the tensor double contraction (:) over pairs of indices. The Hessian contribution is often neglected in standard linearization schemes; we will see its crucial role in quantum nonlinear processing. Note that if the QNP is linear, $\vec{N}_b \to 0$, and the Hessian vanishes.

The NVK approximation, combined with input-output theory [50], allows us to analytically obtain $\boldsymbol{\mu}^{(l)}$ and $\boldsymbol{\Sigma}^{(l)}$ as defined in Eq. (6), and eventually D_{F}. The quadrature

means after a filtering time $\mathcal{T}$ are given by

$$\boldsymbol{\mu}^{(l)} = \sqrt{\frac{\gamma_{\mathrm{H}}\mathcal{T}}{2}}\, \mathbf{U}_K \left(\bar{\Lambda}^{-\frac{1}{2}} \langle \hat{\bar{\boldsymbol{b}}}^{(l)} \rangle - \bar{\Lambda}^{\frac{1}{2}} \bar{\mathbf{J}}_b^{-1} \bar{\mathbf{H}}_b : \mathbf{C}_b^{(l)} \right),$$
(10)

where $\mathbf{U}_K$ is the K-mode quadrature change-of-basis matrix. The second term demonstrates transduction: the sensitivity of linearly-measurable QNP quadratures to its higher-order (here, second-order) quantum cumulants.

The covariance matrix in the long-filter limit ($\gamma\mathcal{T} \gg 1$, see SI Sec. F 5) is given by:

$$\boldsymbol{\Sigma}^{(l)} \overset{\gamma\mathcal{T}\gg1}{=} \sigma_{\mathrm{vac}}^2 (\bar{n}_{\mathrm{cl}}+1)\mathbf{I}_K$$
$$+ \sigma_{\mathrm{vac}}^2 \frac{\gamma_{\mathrm{H}}}{\gamma} \mathbf{U}_K \bar{\mathbf{J}}_b^{-1} \Big[\bar{\mathbf{D}}_b + \boldsymbol{\Gamma}\mathbf{L}_a^{-1}\mathbf{D}_a^{(l)}(\mathbf{L}_a^{-1})^T\boldsymbol{\Gamma}^T \Big] (\bar{\mathbf{J}}_b^{-1})^T \mathbf{U}_K^T.$$
(11)

The first term describes output vacuum fluctuations $\sigma_{\mathrm{vac}}^2 = \frac{1}{2}$ and classical readout noise $\propto \bar{n}_{\mathrm{cl}}$, the latter also in units of σ_{vac}^2 ($\mathbf{I}_K \in \mathbb{R}^{2K}$ is the identity matrix). The second line then describes all other contributions from quantum noise. The term $\propto \bar{\mathbf{D}}_b$ describes noise added by the QNP itself. The term $\propto \mathbf{D}_a^{(l)}$, on the other hand, arises due to our quantum-coherent description of both the QNP and QS: it describes noise originating from the QS, which arrives at the QNP via the coupling $\boldsymbol{\Gamma}$, after undergoing QS evolution via $\mathbf{L}_a^{-1}$. Both noise terms are processed by the QNP, as indicated by the appearance of the Jacobian $\bar{\mathbf{J}}_b^{-1}$.

The above expressions can be used to understand quantum state classification using general weakly-nonlinear bosonic quantum systems. First, we consider the limit of ideal linear quantum amplifiers, processing input drive signals ($\mathbf{D}_a \to 0$). In this case $\bar{\mathbf{H}}_b \to 0$, $\bar{\mathbf{J}}_b^{-1} \to \mathbf{L}_b^{-1}$, and $\bar{\Lambda}^{-\frac{1}{2}}\langle\hat{\bar{\boldsymbol{b}}}\rangle \to -\mathbf{L}_b^{-1}(\mathbf{L}_a^{(l)})^{-1}\vec{\eta}^{(l)}$; hence both the mean and covariance are determined entirely by the matrix $\mathbf{L}_b$, as must be the case for a linear system. The resulting expressions can be used to obtain the standard quantum limits [1, 3] on amplification (see SI Sec. H).

Practical quantum amplifiers, in contrast, can exhibit nonlinear behaviour for sufficiently strong input signals. Then, the leading contribution $\langle\hat{\bar{\boldsymbol{b}}}\rangle$ to $\boldsymbol{\mu}$ is determined by the nonlinear Eq. (7b), while $\boldsymbol{\Sigma}$ is determined by $\bar{\mathbf{J}}_b$, an indicator of the difference in response of nonlinear quantum systems to signal and noise. However, if the aim is still to process input drives $\vec{\eta}^{(l)}$, the large drives typically needed to reach such regimes lead to high signal-to-noise ratios where this difference is often ignored, with more attention instead paid to mitigating the nonlinear response of $\langle\hat{\bar{\boldsymbol{b}}}\rangle$ (to alleviate gain compression, for example).

However, in other quantum information processing tasks, noise takes center stage. A simple case arises when discriminating quantum states l and p such that $\langle\hat{\bar{\boldsymbol{a}}}^{(l)}\rangle = \langle\hat{\bar{\boldsymbol{a}}}^{(p)}\rangle$, so that the leading term $\propto \langle\hat{\bar{\boldsymbol{b}}}\rangle$ in $\boldsymbol{\mu}$ makes no contribution to $\delta\boldsymbol{\mu}$. Then, quantum state classification becomes the task of processing quantum fluctuations

encoded in $\mathbf{D}_a^{(l)}$, and consequently in $\mathbf{C}_b^{(l)}$. The QNP nonlinearity enables such processing *in situ*, via the Hessian tensor. Furthermore, while the covariance is still determined by $\bar{\mathbf{J}}_b^{-1}$ alone, $\delta\boldsymbol{\mu}$ is now determined both by $\bar{\mathbf{J}}_b^{-1}$ and by the Hessian tensor $\bar{\mathbf{H}}_b$. This enables a complex interplay of signal and noise only possible using nonlinear quantum systems. We demonstrate the implications of this result using various examples in this paper.

While these expressions are derived using the NVK approximation, we show that the quantum information processing principles they reveal persist more generally. Importantly, the truncated cumulants ansatz holds past the NVK regime; we use this to develop a computational approach, the Stochastic Truncated Equations of Motion (STEOMs), that can be used to simulate measurement-conditioned dynamics of the nonlinear multimode measurement chain beyond the NVK approximation (see SI Sec. C). The STEOMs allow us to account for classification under finite sampling as in real experiments, numerically verify NVK results, and analyze performance in strongly-nonlinear QNP regimes. Finally, we provide select comparisons using exact (S)ME integration to qualitatively verify our results without any assumptions on nonlinearity strength (see SI Sec. I).

C. Quantum state discrimination tasks

We begin by defining the QS of Eq. (2) whose states we wish to classify. Precisely, we consider the QS to be a general M-mode *linear* quantum system under coherent driving at frequencies $\{\omega_{dm}\}$; written in the interaction picture and assuming resonant driving, it is described by the Hamiltonian

$$\hat{\mathcal{H}}_{\mathrm{QS}}^{(l)} = \Big(\frac{1}{2}\sum_m G_m^{(l)} e^{-i\phi_m^{(l)}} \hat{a}_m^2 + \sum_{n\neq m} G_{nm}^{(l)} e^{-i\phi_{nm}^{(l)}} \hat{a}_n \hat{a}_m + h.c. \Big)$$
$$+ \sum_m \eta_m^{(l)}(-i\hat{a}_m + i\hat{a}_m^\dagger).$$
(12)

The integer l then indexes the states we wish to classify, generated by distinct choices of parameters also indexed by l. Such a Hamiltonian can be realized in the cQED architecture using tunable parametric drives [51–53].

We are ultimately interested in coupling the QS directly to the QNP - a nonlinear quantum device - for quantum signal processing. However, to provide a benchmark for later comparisons, we begin by introducing the classification tasks and how they may be performed in a simpler, more usual context: heterodyne readout of the QS using *linear* quantum amplifiers. Recall that this simplified measurement chain, depicted in Fig. 2(a) for $M = 2$, is still described by the same general form of SME, Eq. (1), provided $\hat{\mathcal{N}}_{\mathrm{QNP}} \to 0$ and $\hat{\mathcal{H}}_{\mathrm{QNP}} \to \hat{\mathcal{H}}_{\mathrm{PP}}$, where the latter Hamiltonian specifically describes the phase-preserving (PP) style of linear amplifier (for full model, see Appendix D).

Table I. QS parameters defining the two classification tasks considered, in units of common QS mode loss rate κ. Note that we set $\Delta_m^{(l)} = \phi_m^{(l)} = 0 \; \forall \; m, \; l$.

		Task I						Task II			
l	$G_1^{(l)}$	$G_{12}^{(l)}$	$\phi_{12}^{(l)}$	$\eta_1^{(l)}$	$\eta_2^{(l)}$	l	$G_1^{(l)}$	$G_{12}^{(l)}$	$\phi_{12}^{(l)}$	$\eta_1^{(l)}$	$\eta_2^{(l)}$
1	0.3	0.0	0.0	$0.20\mathcal{A}$	0.0	3	0.0	0.3	$-\frac{\pi}{2}$	$0.8\mathcal{A}$	$0.8\mathcal{A}$
2	0.0	0.3	0.0	$0.32\mathcal{A}$	0.0	4	0.0	0.3	$+\frac{\pi}{2}$	$0.2\mathcal{A}$	$0.2\mathcal{A}$

We can finally define the binary quantum state discrimination tasks we wish to analyze: our work considers the discrimination of QS states which possess the unique feature that any monitored modes (i.e. modes that are coupled to the downstream processor) have identical steady-state quadrature expectation values for both states. More precisely, for two distinct states indexed by l and p, $\langle \hat{a}_m^{(l)} \rangle = \langle \hat{a}_m^{(p)} \rangle = \mathcal{A} \; \forall \; m$ s.t. $\Gamma_m \neq 0$, where $\mathcal{A}$ defines a common amplitude of monitored modes. Such states can then only be distinguished on the basis of their quantum fluctuation statistics, *if* these are distinct.

For Task I of this type, we consider distinguishing a single-mode squeezed state of mode $\hat{a}_1$ ($l = 1$) from a two-mode squeezed state ($l = 2$), by monitoring only mode $\hat{a}_1$ ($\Gamma_1 \neq 0, \Gamma_2 = 0$). The QS parameters used to realize these states are summarized in Table I. The top plot in Fig. 2(b) shows the steady-state quadrature expectation values of the monitored mode $\hat{a}_1$, which are equal by construction. In contrast, the depicted full QS quadrature covariance matrices, related to the second-order QS cumulants $\mathbf{C}_a^{(l)}$ via a simple change-of-basis (see SI), clearly indicate the differences in quantum fluctuations of the two QS states. The discrimination task requires extracting these differences from measurement records $\boldsymbol{x}$, typically following some processing to obtain readout features $\boldsymbol{y} = F[\boldsymbol{x}]$. If we restrict this processing to only linear operations, the obtained distributions of features $\boldsymbol{y} = F_{\mathrm{L}}[\boldsymbol{x}] = (I_1, Q_1)$ (simulated using the STEOMs) are shown in the first panel of Fig. 2(c) for single-shot ($S = 1$) readout of different realizations of each QS state. Clearly, the mean values of both measured feature distributions overlap; note that this mean would be unchanged by averaging over repeated shots ($S > 1$). Hence $||\delta\boldsymbol{\mu}|| \to 0$ and therefore $D_{\mathrm{F}} \to 0$, leading to the conclusion that the two states cannot be distinguished in the space of (I_1, Q_1).

However, the distributions of measured quadratures *are* visibly distinct, just not in their mean values, but instead in their second-order moments or covariances. To estimate such second-order moments (routinely required for example for Gaussian state tomography [45, 54, 55]), the standard approach is to obtain S shots and estimate the variance of the measurements over the dataset yielding readout features $\boldsymbol{y} = F_{\mathrm{NL}}[\boldsymbol{x}] = \frac{1}{S}\sum_s^S (I_1^2, Q_1^2)$ (shot s dependence of quadratures I_k, Q_k is implied). For Task I, we plot distributions of these nonlinear readout features in the second panel of Fig. 2(c). The mean values of
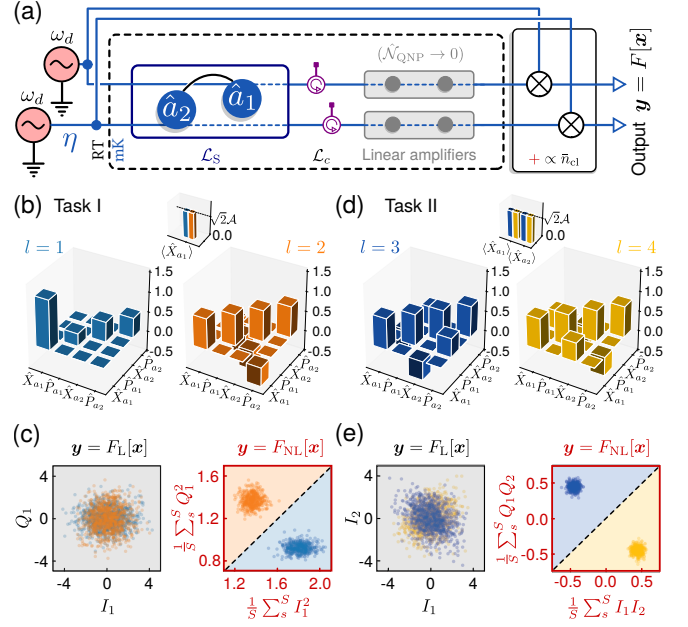

Figure 2. (a) Measurement chain defined by Eq. (1) for heterodyne monitoring of the QS (shown for $M = 2$ modes) using *linear* quantum amplifiers. (b) Task I: distinguishing a single-mode squeezed state from a two-mode squeezed state, using readout of mode $\hat{a}_1$ only. (c) Distributions of measured features for Task I under linear (left) and nonlinear (right) processing (centered by subtracting off mean values) (d) Task II: distinguishing two-mode squeezed states with orthogonal squeezed quadratures, using joint readout of modes $\hat{a}_1$ and $\hat{a}_2$. (e) Same as (c), for Task II.

these features are now estimators of the monitored mode covariances; as these are distinct, the centers of the distributions no longer coincide, rendering them linearly separable. Without this nonlinear processing step, it is impossible to perfectly distinguish the QS states in Task I under QS readout using only linear amplifiers.

For Task II we consider a more complex classification task, which requires both nonlinear and *nonlocal* information processing. Specifically, we wish to distinguish a pair of two-mode squeezed states $l \in (3, 4)$ that experience an identical amount of joint squeezing, but whose two-mode squeezed quadratures are mutually orthogonal. Such states can be generated using two-mode squeezing interactions of equal strength but opposite phase (see Table I). The full QS covariance matrices (see Fig. 2(d)) then differ only in the *sign of cross-correlations* between the two QS modes (off-diagonal blocks); all other covariance metrics are identical.

Measuring such cross-correlations necessitates monitoring both QS modes ($\Gamma_1, \Gamma_2 \neq 0$). We again first show measured features under linear processing alone, now for two distinct modes, $\boldsymbol{y} = F_{\mathrm{L}}[\boldsymbol{x}] = (I_1, I_2)$, in the left panel of Fig. 2(e). Clearly, the distributions have anisotropic profiles that differ in the axis of minimal fluctuations, indicative of two-mode squeezing of distinct joint quadratures. However, the centers of measured distributions

overlap as before, due to their equal mean values. Also as before, nonlinear processing to estimate second-order moments provides the solution, but with an important caveat: all *local* second-order moments are insensitive to the sign of cross-correlations of the QS modes. Instead, the estimation of *nonlocal* second-order moments, $\boldsymbol{y} = F_{\mathrm{NL}}[\boldsymbol{x}] = \frac{1}{S}\sum_s^S (I_1 I_2, Q_1 Q_2)$, is necessary to obtain linearly separable distributions for Task II.

Therefore, both our considered classification tasks - instances of the broad family of tasks that require measuring correlations between observables to distinguish quantum states - demand nonlinear processing of heterodyne records at room temperature. However, such processing can be particularly sensitive to high-temperature noise $\propto \bar{n}_{\mathrm{cl}}$ in the measurement chain. A key result of our work is that using QNPs to process signals from the QS in the same cryogenic environment where they are generated can circumvent the need for nonlinear processing at room temperature. This renders classification schemes incorporating QNPs much more robust to excess noise in the measurement chain (see Sec. IV).

III. QUANTUM STATE DISCRIMINATION USING A SINGLE-MODE QNP

We now demonstrate the paradigm of *in situ* nonlinear processing of quantum signals enabled by the QNP in the context of quantum state discrimination. In particular, we will show that such processing would not be possible using standard linear amplifiers in the same measurement configuration. Our analysis begins with the simpler Task I; Task II is analyzed in Sec. V.

Recalling that only a single QS mode is read out in Task I, we consider a $K = 1$ QNP to monitor this QS mode, which proves sufficient. Our chosen QNP is defined by a single Kerr-nonlinear mode with frequency ω_1 and nonlinearity Λ (for details, see Appendix A). The resulting quantum measurement chain is then depicted in Fig. 3(a). When coupled, the quantum state of the QNP is determined by the state of the QS, as desired: this dependence is shown in Fig. 3(b) via the covariance matrices of the complete measurement chain for the two QS states to be disntinguished. In the NVK approximation, the dependence is quantified by the linear Lyapunov system, Eq. (8), although the exact relationship can be more complex. From Eq. (8), it is also clear that this inter-dependence itself does not require the QNP to be nonlinear; it only requires a nonzero coupling $\boldsymbol{\Gamma}$.

Crucially, we require this dependence to be transduced to QNP readout features to enable successful QS state discrimination by *linear* readout of the QNP alone. It is here that the role of QNP nonlinearity becomes clear. To demonstrate this, we analyze QNP readout distributions for features obtained under linear processing only, $\boldsymbol{y} = F_{\mathrm{L}}[\boldsymbol{x}] = (I_1, Q_1)$, as a function of nonlinearity. Our STEOMs framework enables simulating individual quantum trajectories of the QS and QNP (crucially account-

ing for the latter's nonlinearity), providing the resulting heterodyne measurement records defined by Eq. (4) that are used to construct $\boldsymbol{y}$; typical examples are shown in Fig. 3(c), (d). By repeating over several measurement chain initializations for each QS state, here $l = 1, 2$, we obtain distributions of measured features shown in Fig. 3(e), (f). Then Fisher's discriminant D_{F} computed for the two distributions determines the fidelity of classifying the QS states.

Recall that D_{F} depends on the mean separation $\delta\boldsymbol{\mu}$ of the distributions. The NVK approximation provides a very useful form of $\delta\boldsymbol{\mu}$, Eqs. (10), which specialized to the single-mode Kerr QNP takes the form:

$$\delta\boldsymbol{\mu} = \sqrt{\frac{\gamma_{\mathrm{H}}\mathcal{T}}{2}} \mathbf{U}_K \left[\sqrt{\frac{\Lambda}{\gamma}} \, \bar{\mathbf{J}}_b^{-1} \bar{\mathbf{H}}_b : \left(\mathbf{C}_b^{(l)} - \mathbf{C}_b^{(p)} \right) \right], \quad (13)$$

where $\gamma = \gamma_{\mathrm{H}} + \Gamma_1$ is the total QNP damping rate. The nonlinear dependence is associated with the Hessian $\bar{\mathbf{H}}_b$, which vanishes for linear systems, so that $||\delta\boldsymbol{\mu}|| = 0$. This is exactly what is observed for readout with linear amplifiers in Sec. II C, as well as with vanishing QNP nonlinearity ($\Lambda = 0$), shown in Fig. 3(e). In this case, increasing the filtering time $\mathcal{T}$ also has no effect on $||\delta\boldsymbol{\mu}||$, and consequently on D_{F}.

The measured distributions change qualitatively when the QNP is nonlinear ($\Lambda \neq 0$), as shown in Fig. 3(f). Now $||\delta\boldsymbol{\mu}|| \neq 0$, and the mean separation increases with $\mathcal{T}$ as predicted by Eq. (13). The QNP is able to transduce information encoded in QNP fluctuations $\mathbf{C}_b^{(l)}$, which in turn depend on the QS state, to QNP quadrature mean values, an operation governed by the Hessian $\bar{\mathbf{H}}_b$. Now even with only linear processing $F_{\mathrm{L}}[\cdot]$ of measured quadratures, $D_{\mathrm{F}} \neq 0$: a suitable QNP is thus able to circumvent the need for nonlinear post-processing of measurement records, and the resulting classification accuracy shown in Fig. 3(f) approaches unity. In contrast, for vanishing nonlinearity, perfect classification is impossible using only $F_{\mathrm{L}}[\cdot]$.

This simple example highlights the extremely general principle of transduction using QNPs. However, making successful use of this principle requires asking some more complex questions, each of which we address in this paper. First, one may ask whether *any* QNP works for a given quantum state discrimination task. Unsurprisingly, this is not the case; instead, QNP optimization is important, and features a landscape far richer than that of linear quantum amplifiers. The reason for this is simple: the very nonlinearity that enables a nonzero 'signal' $||\delta\boldsymbol{\mu}||$ also determines the fluctuations $\mathbf{C}_b^{(l)}$ that appear in the 'noise', characterized by the measured covariance matrix $\boldsymbol{\Sigma}^{(l)}$. This leads to a complex interplay between signal and noise for classification using QNPs, where one cannot be optimized without considering the other. Further complicating this interplay is the noise added by the QNP itself, which must also be accounted for in the optimization. While the limits of added noise have been well-established for linear quantum amplifiers [1, 3], much less
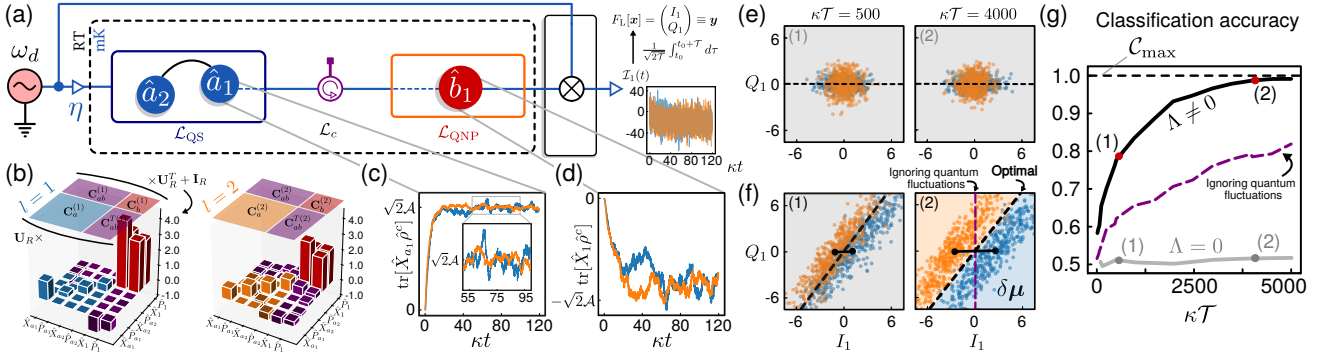
Figure 3. (a) Measurement chain for Task I: using a single-mode QNP to distinguish a single-mode squeezed state $\hat{\rho}_{\mathrm{QS}}^{(1)}$ from a two-mode squeezed state $\hat{\rho}_{\mathrm{QS}}^{(2)}$. (b) Covariance matrices of the joint quantum state of the measurement chain when the QS is in state $\hat{\rho}_{\mathrm{QS}}^{(1)}$ (left) and $\hat{\rho}_{\mathrm{QS}}^{(2)}$ (right); $\mathbf{U}_R, \mathbf{I}_R$ are the R-mode change-of-basis and identity matrix respectively. (c) STEOMs simulations of quantum trajectories and measurement records: for each QS state, we show the expectation value of $\hat{X}$ conditioned on an individual quantum trajectory $\hat{\rho}^c$ for (i) mode $\hat{a}_1$ and (ii) mode $\hat{b}_1$, while the resulting heterodyne record $\mathcal{I}_1(t)$ is shown in (iii). (e) Distribution of readout features $\boldsymbol{y}$ under linear processing $F_{\mathrm{L}}[\cdot]$ for a QNP with $\Lambda = 0$ and for filtering times (1) $\kappa\mathcal{T} = 500$, and (2) $\kappa\mathcal{T} = 4000$. (f) Same as (e) but now for a QNP with $\Lambda \neq 0$. (g) Classification accuracy as a function of $\kappa\mathcal{T}$. Black curve is for the QNP with $\Lambda \neq 0$ and optimal decision boundary in (f), purple curve is obtained if quantum fluctuations are ignored (purple decision boundary in (f)), while the gray curve corresponds to the QNP with $\Lambda = 0$ in (e).

is known about added noise in the context of nonlinear quantum systems, and even less so for driven-dissipative cases relevant in practical settings.

Our framework characterizes this nontrivial relationship between signal and noise in QNPs. Even more importantly, it also accounts for noise added by QNPs during processing. The latter crucially includes the possibility of the noise being correlated, even non-classically, either due to correlations in the incident quantum signals arriving from the QS, or even due to correlations generated by the QNP dynamics itself. Crucially, we find that the nonlinear nature of QNPs implies that the signal and noise contributions exhibit different relative functional dependencies, a freedom not available to linear quantum systems. This distinction provides the ability to manipulate the readout noise for QNPs so as to limit its impact on the signal, by tuning parameters of the QNP alone. This capability proves crucial to the optimization of a QNP for learning, as is discussed in the next subsection.

We emphasize that a precise understanding of the noise physics and optimal operating regimes of QNPs is not merely an academic question, but one that must be addressed to understand the ultimate limits of information processing with such devices. For example, for classification with minimal resources (e.g. filtering time $\mathcal{T}$), information processing must be carried out in regimes where the signal $||\delta\boldsymbol{\mu}||$ is of comparable magnitude to the standard deviation of measured distributions. Such a case is depicted in Fig. 3(f). From the observed non-isotropic distributions, it is clear that QNP readout can exhibit correlated noise statistics. If the correlated nature of fluctuations was ignored (in other words, if the added noise was assumed to be uncorrelated and isotropic), the decision boundary would simply be the bisector of the mean separation $||\delta\boldsymbol{\mu}||$, depicted by the dashed purple line. Unless $\mathcal{T}$ is very large so that $||\delta\boldsymbol{\mu}||$ dominates over noise (a significant increase in measurement resources), the resulting classification accuracy (purple) is substantially lower than the optimal. Therefore the correlation properties of noise in QNP readout - including of noise added by the QNP itself - must be accounted for, as we do here and analyze in detail next.

A. Coherent control of quantum fluctuations using a QNP

We have seen that the QNP nonlinearity is essential to learn QS states using linear readout. However, not all QNPs will be equally effective at learning: performance metrics such as $\mathcal{C}_{\max}$ for a fixed $\mathcal{T}$ are strongly dependent on QNP operating parameters. In this section, we address how QNPs can be optimized for learning. Our analysis centers around understanding how Fisher's discriminant D_{F} for QNP-enabled classification can be maximised, to which end analytic expressions obtained using the NVK approximation are particularly powerful. In particular, while our analysis is presented for Task I using a $K = 1$ mode QNP, the NVK approximation allows us to uncover general principles for QNP operation that apply to more complex tasks with larger QNPs, as we show in Sec. V with Task II.

A nonzero Fisher's discriminant D_{F} requires a non-vanishing $\delta\boldsymbol{\mu}$. Within the NVK approximation, Eq. (13), it is clear that the magnitude $||\delta\boldsymbol{\mu}||$ must be determined by the inverse of the Jacobian $\bar{\mathbf{J}}_b^{-1}$ describing the QNP. This is unsurprising, as the Jacobian typically determines the linearized response of a nonlinear system, with the dimensionless susceptibility $|\chi_b|$ given by the largest eigenvalue of $\bar{\mathbf{J}}_b^{-1}$. As discussed in Appendix G, for a $K = 1$
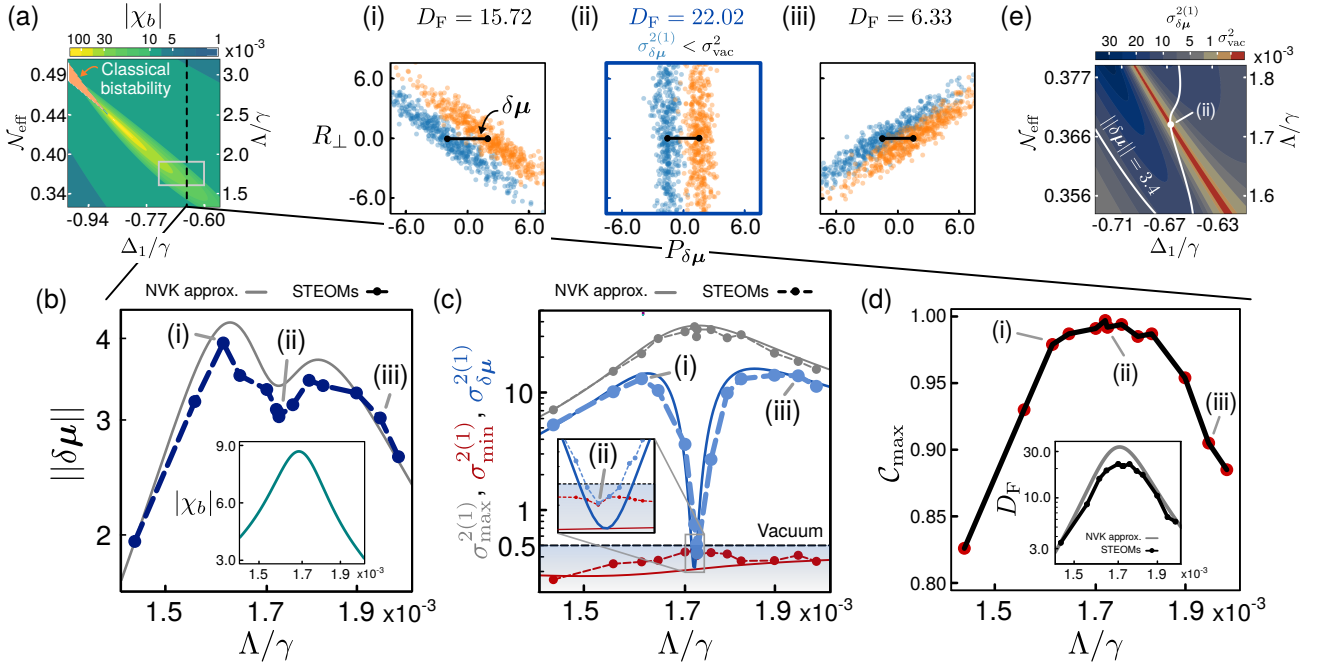
Figure 4. (a) Susceptibility $|\chi_b|$ as a function of nonlinearity Λ/γ and detuning Δ_1/γ, which exhibits an increase as the classical Kerr oscillator bistability is approached. (b) $|\chi_b|$ and $||\delta\boldsymbol{\mu}||$, (c) noise properties, and (d) $\mathcal{C}_{\max}$ as a function of nonlinearity Λ/γ, for the cross-section of (a) at fixed detuning $\Delta_1/\gamma = -0.67$. Inset in (d) indicates Fisher's linear discriminant D_{F}. Results are shown using both the NVK approximation and numerical STEOMs (apart from $\mathcal{C}_{\max}$). (i)-(iii) show measured QNP quadrature distributions under only linear processing $F_{\mathrm{L}}[\cdot]$, simulated using the STEOMs at the indicated nonlinearity values, depicting the ability to manipulate quantum fluctuations. (ii) maximises D_{F} within the NVK approximation by minimizing projected noise. (e) $\sigma_{\delta\boldsymbol{\mu}}^{2(1)}$ as a function of nonlinearity Λ/γ and detuning Δ_1/γ in the NVK approximation. The white curve shows operating conditions with constant $||\delta\boldsymbol{\mu}||$, but with clearly varying projected noise values.

mode Kerr-type QNP this takes the form

$$|\chi_b| = \max\{|\mathrm{eig}\, \bar{\mathbf{J}}_b^{-1}|\} = \frac{\gamma}{\frac{\gamma}{2} - \sqrt{(2\bar{n}_1)^2 - (\Delta_1 + 4\bar{n}_1)^2}}, \tag{14}$$

where $\bar{n}_1 = |\langle \bar{b}_1 \rangle|^2$, obtained by solving Eq. (7b). Importantly, it can be shown that $|\chi_b|$ depends on the QNP nonlinearity only via the dimensionless effective parameter $\mathcal{E}_{\mathrm{eff}}$,

$$\mathcal{E}_{\mathrm{eff}} = \sqrt{\frac{\Lambda}{\gamma}} \cdot \frac{\Gamma_1}{\gamma} \langle \hat{\bar{a}}_1^{(l)} \rangle = \sqrt{\frac{\Lambda}{\gamma}} \cdot \mathcal{A} \frac{\Gamma_1}{\gamma} \tag{15}$$

which is independent of l due to our choice of classification tasks. The above form of $|\chi_b|$ is ubiquitous in studies of the linearized response of a variety of Kerr-based systems, from parametric amplifiers [21] to frequency comb generators [56]. The susceptibility can be made large by suitable parameter choices, including $\mathcal{E}_{\mathrm{eff}}$; typically this is determined by the strength of a separately applied pump tone. Our use of Kerr-based QNPs exhibits a slight difference: $\mathcal{E}_{\mathrm{eff}}$ is instead set by the amplitude $\mathcal{A}$ of signals incident from the QS upstream. The QS state to be measured hence serves to 'pump' the very QNP being used to measure it (although a distinct QNP pump tone could equivalently be applied).

This difference notwithstanding, suitable choices of $(\mathcal{E}_{\mathrm{eff}}, \Delta_1)$ can similarly cause $|\chi_b|$ to become large, increasing the magnitude of QNP response to any input, including to quantum fluctuations from the QS. In Fig. 4(a), we plot $|\chi_b|$ as a function of Δ_1/γ and $\mathcal{E}_{\mathrm{eff}}$ (also Λ/γ for $\mathcal{A} \cdot \frac{\Gamma_1}{\gamma} = \frac{80}{9}$). The orange region, where $|\chi_b|$ diverges, marks the well-known classical bistability of the single Kerr oscillator [57], here brought about by the mean amplitude $\mathcal{A}$ of incident QS signals. Operating near this bistability - and more generally, any instability of the Jacobian $\bar{\mathbf{J}}_b$ - will increase $|\chi_b|$.

We therefore consider QNPs near the bistability, more precisely for the fixed detuning $\Delta_1/\gamma = -0.67$ and for varying $\mathcal{E}_{\mathrm{eff}}$ across the vertical dashed line in Fig. 4(a). Here the susceptibility $|\chi_b|$ is plotted in the inset of Fig. 4(b), which exhibits a Lorentzian-like profile with a single maximum. Note that the important quantity for Fisher's discriminant is instead the magnitude of measured mean separation $||\delta\boldsymbol{\mu}||$; this is also shown in Fig. 4(b), using both the NVK approximation and STEOMs integration. Interestingly, while $||\delta\boldsymbol{\mu}||$ does increase in conjunction with $|\chi_b|$, it also displays a double-peak structure that is manifestly distinct. Thus, while $|\chi_b|$ is clearly important, it does not completely define the physics of QS state learning using a QNP. Here, the difference can be attributed to the fact that $||\delta\boldsymbol{\mu}||$ is not

simply the linearized response to an input field, but the response to quantum fluctuations encoded in $\mathbf{C}_b^{(l)}$, which drive the QNP via the Hessian tensor $\bar{\mathbf{H}}_b$ (which is absent in standard linearization approaches).

The other factor determining D_{F} is the noise in measured quadratures, quantified by the covariance matrices $\boldsymbol{\Sigma}^{(l)}$. To aid our analysis of noise properties, we first introduce the minimum and maximum noise eigenvalues and corresponding eigenvectors of the $2K$-by-$2K$ covariance matrix $\boldsymbol{\Sigma}^{(l)}$ as $\{\sigma_{\min}^{2(l)}, \boldsymbol{v}_{\min}^{(l)}\}$ and $\{\sigma_{\max}^{2(l)}, \boldsymbol{v}_{\max}^{(l)}\}$ respectively. These eigenpairs denote combinations of measured QNP quadratures with minimal and maximal noise, and for $K = 1$, completely define $\boldsymbol{\Sigma}^{(l)}$. We next note that binary classification in an arbitrary-dimensional measured space can be cast into a two-dimensional subspace for visualization. We introduce a vector $\boldsymbol{v}_{\|} = \frac{1}{||\delta\boldsymbol{\mu}||}\delta\boldsymbol{\mu}$ parallel to $\delta\boldsymbol{\mu}$, which is unique upto normalization, and a vector $\boldsymbol{v}_{\perp}$ orthogonal to $\boldsymbol{v}_{\|}$, namely $\boldsymbol{v}_{\|}^T \boldsymbol{v}_{\perp} = 0$, of which there are $2K-1$ choices. These vectors allow us to a define a measured QNP quadrature $P_{\delta\boldsymbol{\mu}}$ that is parallel to $\delta\boldsymbol{\mu}$, and $R_{\perp}$ as one of $2K-1$ measured quadratures orthogonal to $\delta\boldsymbol{\mu}$,

$$P_{\delta\boldsymbol{\mu}} = \boldsymbol{v}_{\|}^T \boldsymbol{y}, \ R_{\perp} = \boldsymbol{v}_{\perp}^T \boldsymbol{y}. \tag{16}$$

The quadrature $P_{\delta\boldsymbol{\mu}}$ has the property that $\mathbb{E}[P_{\delta\boldsymbol{\mu}}^{(l)} - P_{\delta\boldsymbol{\mu}}^{(p)}] = ||\delta\boldsymbol{\mu}||$, as may be readily verified. For isotropic noise distributions where only the mean separation determines distinguishability, $P_{\delta\boldsymbol{\mu}}$ then defines the only feature that need be measured for classification. Of course, noise in our situation is far from isotropic. We therefore also introduce the noise projected along $\delta\boldsymbol{\mu}$,

$$\sigma_{\delta\boldsymbol{\mu}}^{2(l)} = \boldsymbol{v}_{\|}^T \cdot \boldsymbol{\Sigma}^{(l)} \cdot \boldsymbol{v}_{\|}. \tag{17}$$

Note that the noise in the $P_{\delta\boldsymbol{\mu}}$ quadrature is exactly $\sigma_{\delta\boldsymbol{\mu}}^{2(l)}$.

In Fig. 4(c), we now plot this quantity, together with the minimum and maximum noise eigenvalues, for $l = 1$ and for the same parameters as Fig. 4(b). Finally, $\mathcal{C}_{\max}$ and D_{F} are plotted over the same parameter range in Fig. 4(d). Clearly, $\sigma_{\max}^{2(1)}$ increases following $|\chi_b|$, describing the amplification of fluctuations by the QNP. In contrast, $\sigma_{\min}^{2(1)}$, which should represent squeezing, does not undergo a corresponding dip; this is because the Kerr-based QNP does not generate ideal squeezed states. Regardless, we still have $\sigma_{\min}^{2(1)} < \sigma_{\mathrm{vac}}^2$, so the measured distributions exhibit squeezing below vacuum. The projected noise $\sigma_{\delta\boldsymbol{\mu}}^{2(1)}$ varies in between these maximum and minimum noise eigenvalues, with some important features that we now describe.

To illustrate the interplay of mean separation and noise, we consider three QNPs defined by nonlinearity values marked (i)-(iii) in Fig. 4. As seen in Fig. 4(b), measured output distributions from QNP (i) have the largest mean separation. However, the projected noise plotted in Fig. 4(c) is far from minimal. From the resulting measured distributions depicted in Fig. 4(i), it

is visually clear that noise in the direction of $\delta\boldsymbol{\mu}$ is not minimized.

QNP (ii) ends up being the most interesting operating point. While the mean separation of outputs from QNP (ii) is lower than (i), the projected noise is minimized, $\sigma_{\delta\boldsymbol{\mu}}^{2(1)} = \sigma_{\min}^{2(1)}$, as seen in Fig. 4(ii). From the measured distributions, we see that $\delta\boldsymbol{\mu}$ is now aligned with the direction of minimum noise, defined by the noise eigenvector $\boldsymbol{v}_{\min}^{(1)}$. The resulting Fisher's discriminant D_{F} and $\mathcal{C}_{\max}$ are in fact larger than for QNP (i), even though $||\delta\boldsymbol{\mu}||$ is slightly smaller.

Finally, QNP (iii) is specifically chosen so that its output $||\delta\boldsymbol{\mu}||$ is equal to that for QNP (ii); however, the projected noise is much larger. From Fig. 4(iii), it is visually clear that the measured distributions from QNP (iii) are the least distinguishable of the three QNPs considered, leading to the smallest value of D_{F} and hence $\mathcal{C}_{\max}$.

This analysis has several important conclusions. First, it emphasizes the necessity of accounting for quantum fluctuations for optimal classification. This is the only factor that distinguishes QNPs (ii) and (iii). Secondly, we are able to clarify the role of QNP-mediated squeezing of quantum fluctuations in classification. As seen from Fig. 4(d), the measured outputs from all three QNPs exhibit distributions with squeezing below vacuum ($\sigma_{\min}^{2(1)} < \sigma_{\mathrm{vac}}^2$). However, this squeezing does not optimally aid classification unless the specific measured QNP quadrature defined by $P_{\delta\boldsymbol{\mu}}$ is the one undergoing squeezing, here QNP (ii).

Finally, and most remarkably, even a simple, single-mode Kerr QNP provides the ability to manipulate quantum fluctuations separately from mean values, so that the aforementioned optimal squeezing scenario can actually be engineered using suitable parameter choices. This capability is emphasized by QNPs (ii) and (iii), which exhibit the same 'signal' $||\delta\boldsymbol{\mu}||$ but very different projected noise properties, a scenario not possible in, for example, linear amplifiers. The remarkable freedom in adjusting the projected noise independently of the mean separation is demonstrated via the surface plot of $\sigma_{\min}^{2(1)}$ in Fig. 4(e); here the white curve, which defines operating points with fixed $||\delta\boldsymbol{\mu}||$, traverses through regions of both very large projected noise, as well as minimal projected noise that is sub-vacuum. Clearly, optimal performance requires operating in regimes where the projected noise can be lowered (not necessarily always exactly minimized, as D_{F} also depends on $||\delta\boldsymbol{\mu}||$).

Before proceeding, we make two further observations. First, it is clear that $||\delta\boldsymbol{\mu}||$ (and hence D_{F}) exhibits a complex dependence on the QNP nonlinearity, in contrast to the monotonic dependence alluded to by Eq. (13). This is because simply varying Λ also modifies $\mathcal{E}_{\mathrm{eff}}$ via Eq. (15) if $\mathcal{A}$ is fixed, thus changing the QNP operating conditions in a nontrivial fashion. To explore the performance improvement with Λ hinted at by Eq. (13) a more careful analysis is needed; this is presented in

Appendix C. Secondly, note that all quantities calculated using the NVK approximation agree well with their counterparts obtained using exact integration of the STEOMs, especially in qualitative terms. This highlights the utility and validity of the NVK approximation in analyzing QNPs and, more generally, networks of coupled nonlinear quantum systems.

IV. QNP ADVANTAGE FOR QUANTUM SIGNALS

Thus far we have demonstrated that QNPs enable processing of quantum signals in a way that is distinct from linear quantum amplifiers. However, it is not yet clear that this difference will lead to a practical advantage for quantum information processing. We now address this key question.

To do so, we will compare readout using the QNP against standard linear quantum amplifiers for the task of quantum state discrimination. Note that both these schemes have already been discussed separately in Sec. III and Sec. II C respectively, albeit the latter only for phase-preserving amplification. Now, we take advantage of the sufficiently general and comprehensive description of quantum measurement chains enabled by our framework to consider the configuration in Fig. 5, where a general quantum processor follows the QS to be read out. The processor can be the QNP, or one of a phase-preserving (PP) or phase-sensitive (PS) linear quantum amplifier (see Appendix D for exact models). In this way, the individual couplings, losses, and readout rates of modes constituting the quantum processor can be held fixed from one model to the next, ensuring a direct comparison. Secondly, we now also include the effect of excess classical noise $\propto \bar{n}_{\mathrm{cl}}$, which necessitates the use of quantum processors in the first place. We can then identify what practical advantages (if any) are enabled by the QNP.

We briefly note that if one considers only binary quantum state discrimination tasks, provably optimal measurement protocols are known, namely Helstrom measurements [48]. While a sufficiently broad definition of a QNP would also incorporate such schemes, such measurements are specific to the particular states to be distinguished, and are not always straightforward to perform as they typically require non-Gaussian operations or strong constraints on coherence. We envision the QNPs considered in this paper to be akin to linear quantum amplifiers in this respect: they may not necessarily be the optimal choice for a single task, but can provide general quantum processing for quantum signals for much more general tasks. Nevertheless, we will still compare the QNP and linear amplifiers against bounds set by optimal discrimination schemes, to be discussed shortly.

Even for processors in the same measurement chain, one last degree of freedom remains: the operating parameters. For linear PS and PP amplifiers, it is straightforward to identify equivalent operating regimes: we can simply choose them to operate at the same gain, or equivalently the same susceptibility $|\chi_b|$. The PS amplifier has only one additional degree of freedom, namely the phase of the quadrature to be amplified. Unlike these amplifiers, the QNP does not provide only linear amplification, and so its operation is not determined entirely by $|\chi_b|$, as shown earlier. Nevertheless, we choose to operate the QNP at the same fixed $|\chi_b|$ as the linear amplifiers to ensure comparable operating points. These operating points are defined by 'isogain' contours of $|\chi_b|$ in Λ-Δ_1 space; considered examples are shown in Fig. 5(b).

A. Quantum Chernoff Bound and optimal discrimination

In assessing the relative performance of the various processors at quantum state discrimination, we first note that the upper bound on the discrimination accuracy must depend on the fundamental distinguishability of the two reduced QS quantum states $\hat{\rho}_{\mathrm{QS}}^{(1)}, \hat{\rho}_{\mathrm{QS}}^{(2)}$, independent of the specific processing or measurement scheme. One measure of this distinguishability is the Quantum Chernoff Bound (QCB) ζ. The QCB bounds the discrimination accuracy according to $1 - \mathcal{C}_{\max} \sim \exp(-N\zeta)$ where access to N identical copies of each state is assumed, all of which may be measured [58–62]. Clearly, the larger the value of the QCB ζ, the more easily distinguishable the two quantum states are for fixed N. While strictly speaking a bound that holds in the asymptotic limit of many copies N, we use the QCB here due to the ease of its calculation for Gaussian states. We then define ζ_{QS} as the QCB for the QS states $\hat{\rho}_{\mathrm{QS}}^{(1)}, \hat{\rho}_{\mathrm{QS}}^{(2)}$. Similarly, when such states are processed using a downstream QNP, the QCB ζ can also be computed to quantify the distinguishability of the QNP states $\hat{\rho}_{\mathrm{QNP}}^{(1)}, \hat{\rho}_{\mathrm{QNP}}^{(2)}$, where $\hat{\rho}_{\mathrm{QNP}}^{(l)} = \mathrm{tr}_{\mathrm{QS}}[\hat{\rho}^{(l)}]$. We will consider Task I in what follows, so that the QCB is computed for the single mode to be read out (see Appendix E for full details).

We plot the ratio of these two QCB values, $\zeta/\zeta_{\mathrm{QS}}$, for three different $|\chi_b|$ values for the different quantum processors in Fig. 5(c). For the PS amplifier the optimal QCB is chosen by varying the PS amplification phase. For the QNP, the QCB is plotted along the isogain contours in Fig. 5(b), using both TEOMs and within the NVK approximation. Immediately, we note that $\zeta/\zeta_{\mathrm{QS}} < 1$: the use of a quantum processor will generally introduce loss channels that are independent of the states to be distinguished and therefore hinder classification. We also note that for the present task, PS amplification enables a larger QCB than PP amplification: even though the PS amplifier can only amplify information in a single quadrature of the quantum input signal, it does so without adding any noise. Interestingly, we see that the QNP can approach and at certain operating points even exceed the QCB obtainable using linear quantum amplifiers in the same measurement chain. We emphasize that this was not *a priori* guaranteed in the
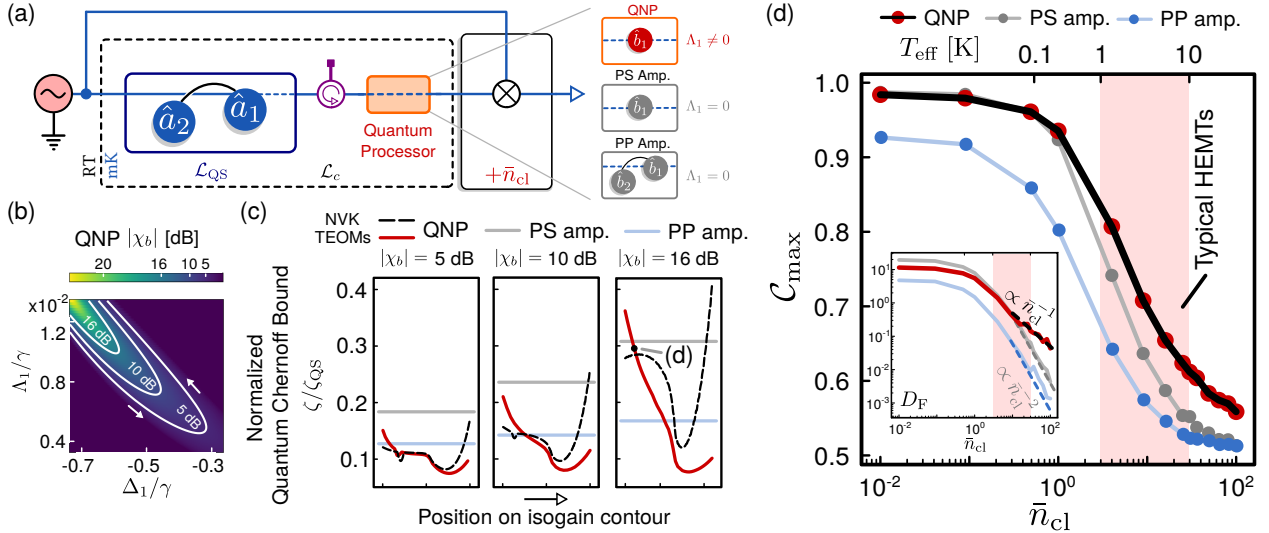
Figure 5. (a) Measurement chain for quantum state discrimination using different quantum processors: phase-sensitive (PS) or phase-preserving (PP) quantum amplifiers, or QNPs. (b) QNP susceptibility $|\chi_b|$ in dB, as a function of operating parameters. Contours of fixed $|\chi_b|$ are shown in white. (c) Normalized QCB $\zeta/\zeta_{\mathrm{QS}}$ for QNP parameters along the three 'isogain' contours in (b), using both the NVK approximation and using TEOMs. Horizontal lines show normalized QCB values at the same $|\chi_b|$ for linear PS and PP amplifiers. (d) $\mathcal{C}_{\max}$ for QNP readout (black), PS amplifier readout (gray), and PP amplifier readout (blue) as a function of excess classical noise power in the measurement chain, all for $|\chi_b| =$16 dB indicated in (c). Top axis calibrates the classical readout noise in terms of an effective temperature T_{eff}. Red shaded region marks the noise added by typical HEMT amplifiers in modern cQED measurement chains. Inset: Fisher's linear discriminant D_{F}, a measure of the *signal-to-noise* ratio for quantum state discrimination, for the QNP and both linear amplifiers as a function of $\bar{n}_{\mathrm{cl}}$.

presence of the added noise of the QNP; hence this observation is promising for the use of QNPs for quantum state discrimination.

B. QNP robustness to readout noise

Given the reduction in the QCB when using a downstream quantum processor, we reiterate why such processors are necessary in the first place. Excess classical noise $\bar{n}_{\mathrm{cl}}$ in the measurement chain can swamp weak quantum signals, such as those emanating from the QS directly, as they are extracted to the classical observer. The QCB is a discrimination bound that assumes optimal measurements, and therefore does not account for the limitations on readout imposed by $\bar{n}_{\mathrm{cl}}$. In practical measurements subject to readout noise, quantum processors such as linear amplifiers then provide the pre-amplification necessary to overcome this noise and enable visibility of weak quantum signals.

To quantify the impact of this practical constraint, we instead compute the classification accuracy for the different quantum processors under heterodyne readout and in the presence of $\bar{n}_{\mathrm{cl}}$. We consider the operating point labelled (d) in Fig. 5(c), where $\zeta/\zeta_{\mathrm{QS}}$ agrees well between the NVK and TEOMs. We also deliberately consider an operating point where the PS amplifier enables a larger QCB ζ than the QNP; as we will show, the particular QNP advantage we wish to highlight prevails in spite of this. Finally, we define readout features

obtained using linear ($F_{\mathrm{L}}[\cdot]$) and nonlinear ($F_{\mathrm{NL}}[\cdot]$) post-processing of measurement records over multiple shots, $\boldsymbol{y} = \frac{1}{S}\sum_s^S (I_1, Q_1, I_1^2, Q_1^2, I_1 Q_1)$. This enables a fair comparison of the QNP against linear amplifiers, as the latter can only perform this discrimination task when nonlinear post-processing is allowed.

We plot the obtained $\mathcal{C}_{\max}$ using the different quantum processors as a function of $\bar{n}_{\mathrm{cl}}$ in Fig. 5(d). For low enough $\bar{n}_{\mathrm{cl}}$, both QNP and PS readout exhibit similar performance while the PP amplifier is worse, consistent with the obtained QCB values. With increasing $\bar{n}_{\mathrm{cl}}$, the performance of all readout schemes degrades as expected. However QNP readout is clearly the most robust, exhibiting the smallest reduction. The red shaded region marks typical $\bar{n}_{\mathrm{cl}}$ values for cQED measurement chains, attributed primarily to noisy HEMT amplifiers (for calibration and mapping to equivalent noise temperatures as shown, see SI Sec. H). QNP readout therefore provides enhanced robustness against experimentally relevant levels of excess classical noise.

The observed advantage stems from a simple principle: QNPs enable nonlinear processing of quantum signals from the QS in the same quantum environment where they are generated, and - crucially - *prior* to corruption by excess readout noise $\bar{n}_{\mathrm{cl}}$. In contrast, linear quantum amplifiers provide only *linear* gain to quantum signals, so that nonlinear processing of the extracted noisy signals is still required. By reducing or even eliminating the need for nonlinear post-processing, QNPs

therefore purvey a fundamentally different scaling of the signal-to-noise ratio with readout noise $\bar{n}_{cl}$. As shown in the inset of Fig. 5(d), Fisher's discriminant (which defines the signal-to-noise ratio for binary discrimination) degrades only linearly with $\bar{n}_{cl}$ for the QNP. This is markedly different to linear amplifiers, in which case D_F degrades *quadratically* with $\bar{n}_{cl}$, as the post-processing required to compute second-order nonlinear features necessary for this task must also amplify the added readout noise. Most importantly, our results emphasize that this QNP processing advantage does not demand large-scale or highly-coherent quantum systems as QNPs: few- and even single-mode nonlinear quantum systems suffice, deployed in lossy quantum measurement chains with minimal additional components.

We conclude with two practical observations. First, while other choices of readout features $\boldsymbol{y}$ (e.g. for different temporal filters $\mathcal{K}(\tau)$) could influence the quantitative performance of either type of processor, the *qualitative* difference in D_F scaling with $\bar{n}_{cl}$, which relies on the *in situ* nonlinear processing paradigm enabled by QNPs, will remain.

Secondly, in the interest of a fair comparison, here we have considered QNPs as a direct replacement for linear quantum amplifiers, demanding QNPs to offer both transduction *and* amplification to overcome readout noise. In this configuration, we are thus not able to take full advantage of the QNP's ability to engineer minimal projected noise $\sigma^2_{\delta\boldsymbol{\mu}} = \sigma^2_{\min}$, as this minimal noise can be swamped by the following readout noise $\bar{n}_{cl} \gg \sigma^2_{\min}$. Using the QNP to control quantum fluctuations is still important, but to instead operate in a regime where $\sigma^2_{\delta\boldsymbol{\mu}} \gtrsim \bar{n}_{cl}$ (to the extent possible for a given $|\chi_b|$). Alternatively, QNPs can be deployed in conjunction with linear amplifiers. Then, a QNP would first be used to engineer a transduced signal quadrature $P_{\delta\boldsymbol{\mu}}$ with minimal projected noise $\sigma^2_{\min}$. Next, a phase-sensitive amplifier would provide noise-free gain $\sqrt{\mathcal{G}_{PS}}$ to this quadrature, yielding $||\delta\boldsymbol{\mu}||^2 = \mathcal{G}_{PS}(\mathbb{E}[P^{(l)}_{\delta\boldsymbol{\mu}} - P^{(p)}_{\delta\boldsymbol{\mu}}])^2$ and $\sigma^2_{\delta\boldsymbol{\mu}} = \mathcal{G}_{PS}\sigma^2_{\min} + (\bar{n}_{cl} + 1)\sigma^2_{vac}$. Provided $\mathcal{G}_{PS}$ is large enough to satisfy $\mathcal{G}_{PS}\sigma^2_{\min} \gg \bar{n}_{cl}$, the output D_F becomes $D_F \simeq (\mathbb{E}[P^{(l)}_{\delta\boldsymbol{\mu}} - P^{(p)}_{\delta\boldsymbol{\mu}}])^2/\sigma^2_{\min}$, taking advantage of both transduction using QNPs (which determines the 'signal'), and their ability to engineer non-classical noise distributions (which determines the 'noise').

V. MULTI-MODE QNPS

In this final section, we analyze the use of a QNP to perform Task II: distinguishing a pair of two mode squeezed states. Here, we see that the need for multi-mode QNPs naturally arises, which then allows us to identify the role of entanglement in quantum information processing tasks using QNPs.
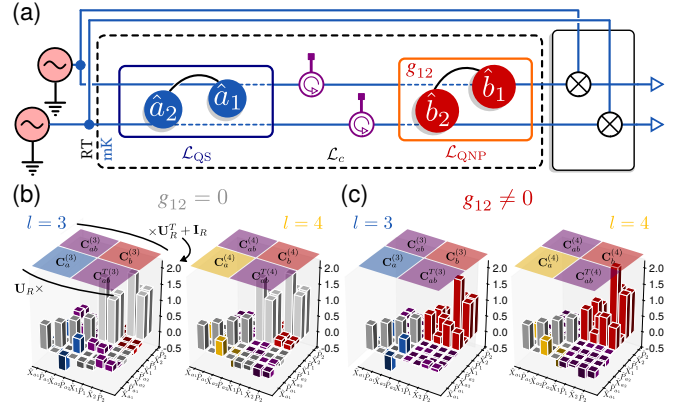

Figure 6. (a) Measurement chain using a two-mode QNP for Task II: distinguishing two-mode squeezed states $\hat{\rho}^{(3)}_{QS}, \hat{\rho}^{(4)}_{QS}$ of a QS. (b) Covariance matrices of the joint quantum state of the chain when the QS is in state $\hat{\rho}^{(3)}_{QS}$ (left) and $\hat{\rho}^{(4)}_{QS}$ (right). Only matrix elements that are distinct between the two cases are shown in color. (c) Same as (b) but now for $g_{12} \neq 0$.

A. Role of coupling

As shown in Sec. II C, successfully performing Task II under direct QS readout requires monitoring of both QS modes, and computation of cross-correlations. For QS modes with disparate frequencies $|\nu_1 - \nu_2| \gg \kappa$ (as required to realize non-degenerate squeezed states), quantum signals from the QS modes at $\omega_{dn} \sim \nu_n$ are also at widely separated frequencies. A single QNP mode with similar bandwidth $\gamma_1 + \Gamma_1 \simeq \kappa_1 + \Gamma_1$, will thus not exhibit a strong response to two signals with such a wide spectral separation. Note that increasing the QNP bandwidth relative to the QS requires $\gamma_1 \gg \Gamma_1$; however, this undercouples the QS to the QNP, reducing the influence of input QS signals and hence the QNP response (see Eq. (15)).

The processing of quantum signals from multiple non-degenerate quantum modes is therefore a situation where the need for a multi-mode QNP naturally emerges. We consider the measurement chain shown in Fig. 6(a) incorporating a $K = 2$ mode QNP, and where each QS mode is monitored by a QNP mode via the non-reciprocal interaction defined by $\boldsymbol{\Gamma}$. A tunable (e.g. parametric) coupling g_{12} between the QNP modes allows us to clearly identify the role played by the multi-mode nature of the QNP.

The full measurement chain covariance matrices are shown in Fig. 6(b) for the two QS states $l = 3, 4$ constituting Task II, first for $g_{12} = 0$. Only matrix elements that are different between the two states are shown in color, to aid visibility. In particular, when $g_{12} = 0$, the local QNP mode covariances are independent of l; this is because the two QS states also have identical local covariances. Instead, even if the QNP modes are uncoupled, non-trivial *correlations* that vary with the QS state are established between the QNP modes. While

this may appear surprising at first glance, it is merely a consequence of the fact that the two uncoupled QNP modes are nevertheless driven by a correlated quantum input signal originating from the QS, which means the two QNP modes' states must also be correlated. However, it is only when the coupling is turned on, Fig. 6(c), that these cross-correlations can be transduced to differences in the local QNP mode covariances. We will see how, in contrast to Task I, this means that now both the nonlinearity and coupling of QNP modes is required for success at Task II.

Unfortunately, with increasing QNP size, the number of tunable parameters and the complexity of their interplay grows, making it *a priori* difficult to isolate the importance of a particular parameter, such as the coupling g_{12}. Here, our analysis of QNP optimization for Task I in Sec. III A proves its generality, and hence its value. For a given $\mathcal{A}$, two principles were found: the importance of $|\chi_b|$ (the largest eigenvalue of the QNP Jacobian) in enhancing the response of the QNP to any input, and the control of quantum fluctuations to minimize projected noise $\sigma_{\delta\boldsymbol{\mu}}^{2(l)}$. Our strategy to isolate the role of coupling is to therefore fix other parameters such that $|\chi_b|$ can be held constant and $\sigma_{\delta\boldsymbol{\mu}}^{2(l)}$ is simultaneously minimized.

The two-mode QNP model we consider is a realization of the Kerr or Bose-Hubbard dimer that has been analyzed in prior work, but only under coherent driving [63], not with both modes driven by quantum signals from a QS upstream as illustrated schematically in Fig. 7(a). By analyzing the Jacobian of this two-mode QNP, we identify regions of parameter space where the susceptibility $|\chi_b|$ grows and diverges, namely near the classical bistability of the Kerr dimer, marked as the orange region in the phase diagram of Fig. 7(a). Then, using the NVK approximation, we identify a trajectory through this parameter space near the classical bistability - labelled the *optimal noise trajectory* (gray curve) - where $|\chi_b| = 9.0$ at every coupling strength, and simultaneously the projected noise is minimized, $\sigma_{\delta\boldsymbol{\mu}}^{2(3)} \simeq \sigma_{\min}^{2(3)}$. An example of how this choice constraints parameters is shown for point (ii) on this trajectory in the lower panel of Fig. 7(a).

We again consider single-shot readout features obtained under linear processing only, now from *both* QNP modes, $\boldsymbol{y} = F_{\mathrm{L}}[\boldsymbol{x}] = (I_1, Q_1, I_2, Q_2)$. Performing Task II for parameters along this trajectory, we obtain $\mathcal{C}_{\max}$ as a function of coupling in Fig. 7(c), which follows a smooth curve. We also plot the corresponding mean separation $||\delta\boldsymbol{\mu}||$ and projected noise in Fig. 7(c). For $g_{12} \to 0$, we see that $||\delta\boldsymbol{\mu}|| \to 0$, even though $\mathbf{C}_b^{(3)} \neq \mathbf{C}_b^{(4)}$ as seen in Fig. 6(b) and the nonlinearity is nonzero. This is because the Hessian tensor in Eq. (13) is still local, as the nonlinearity is on-site, so that $\delta\boldsymbol{\mu}$ is only sensitive to local QNP mode covariances which are independent of l when $g_{12} = 0$. This observation is the result of a much simpler fact: since each QNP mode is coupled to only one QS mode, and the difference between QS states $l = 3, 4$ is present in correlations of *different* QS modes, it is nec-

essary for the two QNP modes to 'communicate' (hence, be coupled) to be able to distinguish the inputs they receive. If this communication is not performed *in situ* using g_{12}, it will have to be performed in post-processing, by computing correlations of measured QNP quadratures for distinct QNP modes explicitly, which constitutes a nonlinear processing step. Turning on g_{12} allows these correlations to be computed via the QNP dynamics, allowing the difference in QS states to translate to a nonzero $||\delta\boldsymbol{\mu}||$ of measured QNP quadratures, enabling $\mathcal{C}_{\max} \to 1$.

B. Engineering entanglement for classification

We are now well-placed to explore the role of entanglement in the processing of quantum signals using a QNP. Importantly, we are interested in output field entanglement, which requires non-classical correlations between *measured* QNP quadratures. For the two-mode QNP, the 4-by-4 measured covariance matrix can be expressed in the block form

$$\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{12}^T & \boldsymbol{\Sigma}_{22} \end{pmatrix} \tag{18}$$

where $\boldsymbol{\Sigma}_{jk}$ is the contribution from covariances between QNP modes j and k (see Appendix F), and we have suppressed the superscript (l) indicating the corresponding QS state. The metric we then use to quantify the degree of entanglement in QNP outputs is the *logarithmic negativity* E_{N} [64], a standard entanglement monotone. In terms of Eq. (18), E_{N} can be defined as

$$E_{\mathrm{N}} = \max\{0, -\ln 2\nu^-\}, \ \nu^- = \sqrt{\frac{d - \sqrt{d^2 - 4\det\boldsymbol{\Sigma}}}{2}},$$
$$d = \det\boldsymbol{\Sigma}_{11} + \det\boldsymbol{\Sigma}_{22} - 2\det\boldsymbol{\Sigma}_{12} \tag{19}$$

In Fig. 7(b), we then plot E_{N} as a function of coupling g_{12} along the optimal noise trajectory, computed using the NVK approximation as well as using simulated covariance matrices via the STEOMs. At first glance, the role of entanglement appears straightforward: increasing coupling coincides with an increase in the degree of entanglement between measured QNP quadratures, and leads to improved classification performance.

However, this observation does not indicate whether entanglement is *always* useful, or if it provides an advantage that goes beyond non-classical correlations of only single-mode variables. The context of binary quantum state discrimination allows us to probe these questions directly. We once again consider the reduced quadrature description introduced in Sec. III A. In Fig. 7(d), we then plot measured QNP distributions in the projected space spanned by $(P_{\delta\boldsymbol{\mu}}, R_\perp)$ for three different operating parameters (i)-(iii), labelled in Fig. 7(a).

By construction, the separation of distribution means for all QNPs (i)-(iii) is entirely confined to the $P_{\delta\boldsymbol{\mu}}$ quadrature. QNPs (i) and (ii) fall along the optimal noise
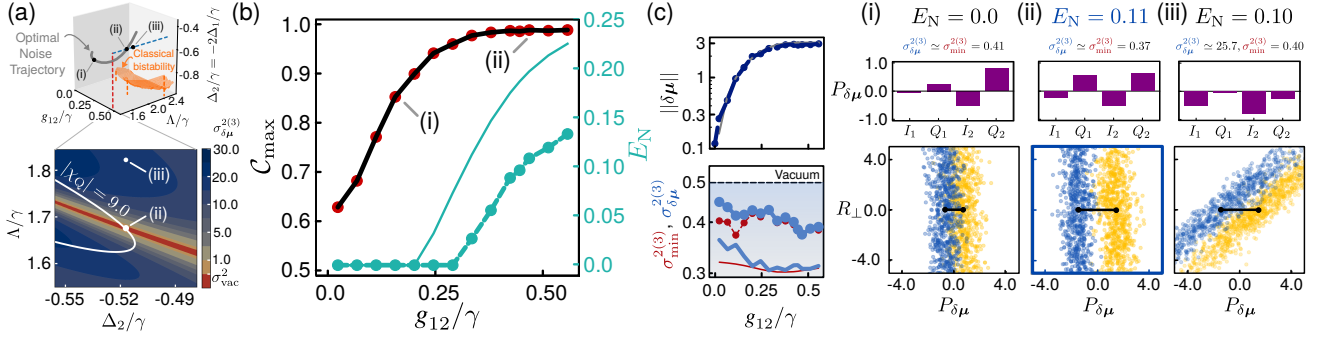
Figure 7. Performance metrics for Task II using a $K = 2$ mode QNP. (a) Top panel: QNP phase diagram in $(g_{12}, \Lambda, \Delta_2)/\gamma$ parameter space, where $\Delta_2 = -2\Delta_1$. Orange region marks the bistability region of Kerr dimer. Gray curve marks the optimal noise trajectory of parameters for which $|\chi_b| = 9.0$ *and* projected noise is minimized. Lower panel: Cross-section at fixed coupling, showing the contour of fixed $|\chi_b| = 9.0$, and surface plot of projected noise. (b) $\mathcal{C}_{\max}$ (black, left-hand axis) and measured output logarithmic negativity E_N (green, right-hand axis) characterizing entanglement between QNP mode outputs, as a function of QNP mode coupling g_{12}/γ. (c) $||\delta\boldsymbol{\mu}||$ and $\sigma_{\delta\boldsymbol{\mu}}^{2(3)}$ as a function of QNP mode coupling along the optimal noise trajectory, using both the NVK approximation and STEOMs. (i)-(iii) show measured distributions in $(P_{\delta\boldsymbol{\mu}}, R_\perp)$ subspace for QNPs with parameters labelled (i)-(iii) in (a). Top bar plots show the respective quadrature composition of $P_{\delta\boldsymbol{\mu}}$.

trajectory and are thus engineered to minimize noise in the $P_{\delta\boldsymbol{\mu}}$ quadrature, as is clearly visible in the distributions. Fig. 7(d). In the top panel of Fig. 7(d), we also show the content of the $P_{\delta\boldsymbol{\mu}}$ quadrature. For (i), $P_{\delta\boldsymbol{\mu}}$ is predominantly comprised of quadratures of only a single QNP mode (here $k = 2$). Hence, the sub-vacuum noise in $P_{\delta\boldsymbol{\mu}}$ is the result of quantum correlations between outputs of only a single QNP mode, indicative of single-mode squeezing. In this weak coupling regime, entanglement is not necessary to obtain sub-vacuum noise; particular, the measured QNP outputs do not exhibit entanglement as $E_N = 0.0$.

However, Task II benefits from increasing coupling between the QNP modes, as observed previously; this not only increases the mean separation $||\delta\boldsymbol{\mu}||$, but also generates output entanglement, as seen for QNP (ii) where $E_N = 0.11$. Now $P_{\delta\boldsymbol{\mu}}$ is comprised of quadratures of *distinct* QNP modes, and thus is a non-local quadrature. Sub-vacuum noise in $P_{\delta\boldsymbol{\mu}}$ therefore must arise due to non-classical correlations amongst these non-local quantum modes, namely entanglement. Here, entanglement of measured quadratures directly improves classification performance by ensuring that non-local quadratures that carry useful information (encoded in $||\delta\boldsymbol{\mu}||$) are also non-classically correlated, such that their fluctuations are reduced below the vacuum limit.

Perhaps most interestingly, just the presence of output field entanglement is not always guaranteed to be beneficial for classification. A simple counterexample is QNP (iii), which demonstrates a nonzero logarithmic negativity $E_N = 0.10$ like QNP (ii), but is not on the optimal noise trajectory as seen from the lower panel of Fig. 7(a). $P_{\delta\boldsymbol{\mu}}$ is still a non-local quadrature, but its noise is clearly not minimal, and is in fact above vacuum. The presence of output field entanglement means that there is some set of non-local quadratures that exhibit non-classical correlations that lead to sub-vacuum noise; however, these

quadratures are not always guaranteed to carry useful information encoded in $||\delta\boldsymbol{\mu}||$, and hence such non-classical correlations may not always be useful for the task at hand.

Crucially, the use of nonlinear QNPs provides the ability to control quantum fluctuations, such that the situation in (ii) can be engineered, and non-classical correlations can be manipulated for computational benefit. For binary classification, this entails finding operating points where the unique quadrature with maximal signal $P_{\delta\boldsymbol{\mu}}$ can simultaneously exhibit minimal, non-classical (e.g. sub-vacuum) fluctuations. The ability to engineer this unique quadrature also opens up avenues to only amplify and measure said quadrature using noiseless phase-sensitive amplification, instead of phase-preserving amplification of all $2K$ quadratures for a K-mode QNP, which necessarily adds noise.

VI. CONCLUSIONS AND OUTLOOK

Nonlinearity is an essential component of signal processing, playing a key role from digital processing to neural circuits. However, the role of general, fundamentally stochastic quantum nonlinear devices in processing quantum signals - beneficial or otherwise - is much less explored. In this paper, we have addressed this limitation by identifying key general principles of quantum information processing enabled by a broad class of nonlinear bosonic quantum systems, which we refer to as QNPs.

Our two main results, which hold beyond the quantum state discrimination tasks we have chosen to demonstrate them, and which make explicit use of QNPs being both quantum *and* nonlinear, can be simply stated. We show that QNPs can be efficient information transducers (cf. Eq. (10)): by processing quantum signals *in situ*, QNPs can render nonlinear properties of quantum sig-

nals such as correlation functions accessible via linear readout schemes like heterodyne monitoring. Secondly, by harnessing a nonlinear input-output map, QNPs can coherently manipulate quantum fluctuations in a manner unavailable to linear amplifiers that are linearly coupled to a quantum signal source: without suppressing the transduced signal magnitude, QNPs can modify its output noise properties. This requires QNPs to control not only quantum fluctuations emanating from the quantum system the QNP is monitoring, but also those resulting from the QNP operation itself, referred to as 'added noise' for linear amplifiers (cf. Eq. (11)). Remarkably, both these capabilities can be accessed using small-scale QNPs, making them relevant for implementations in current experiments. In fact, by analyzing realistic measurement chains, we show that even single-mode QNPs can provide robustness against classical readout noise.

On the other hand, the key theoretical tools these results are built on - the analytic NVK approximation of nonlinear measurement chains and the numerical STEOMs framework to efficiently simulate their conditional dynamics, both verified against exact master equation methods - provide the means to analyze quantum information processing using very general, arbitrarily-multimode nonlinear bosonic quantum systems. As such they can be used to study the processing of signals from many-body quantum systems and non-classical correlations or entanglement across several modes, as we study in Task II. Furthermore, they can enable the exploration of more general paradigms of quantum information processing. One example pointed out in the conclusion of Sec. IV is the use of measurement chains employing both QNPs and phase-*sensitive* amplifiers. Related paradigms enabled by our framework include the possibility of entangling the QNP with the QS, via for example non-reciprocal entangling operations [65, 66].

Taking an even broader view, our work has direct applications to general information processing and computation paradigms such as quantum machine learning and quantum sensing. Nonlinearity is considered essential to the expressive capacity of physical neural networks including quantum systems [67–73]. However, several successful bosonic quantum machine learning platforms are still linear [74–77], instead enabling nonlinear processing by careful use of nonlinear input encoding schemes. Our work provides both tools and possible directions to explore the utility of multi-mode quantum nonlinear devices for learning applications; distinct from other approaches, our framework also describes learning on *quantum* inputs. In particular, it is expected that genuinely quantum advantages in such applications must make use of decidedly quantum properties, such as squeezing and entanglement. QNPs harness nonlinearity to control quantum fluctuations such that non-classical correlations appear in desired observables only. Furthermore, we elucidate general principles that can enable QNPs to operate in regimes to enable such processing. The resulting quantum mechanism to harness quantum correlations for the enhancement of classification accuracy could prove useful in extracting quantum advantages for quantum machine learning and quantum sensing.

Our work also invites exploration of more complex QNPs beyond the weakly-nonlinear regime. Firstly, the scaling advantage with $\bar{n}_{\rm cl}$ can be expected to be even more significant with increasing complexity of the required nonlinear computation, such as the calculation of higher-order or many-body correlations [10]. Such computations will demand the analysis of QNPs with higher-order nonlinearities, or operation in increasingly non-Gaussian regimes. Secondly, for the specific case of quantum state discrimination, we have shown that QNPs can approach and even exceed the optimal discrimination bound, quantified here by the Quantum Chernoff bound (QCB), achievable using linear quantum processors. However, a study of the maximum attainable QCB limits using QNPs requires going beyond the NVK approximation, and could quantify the ultimate constraints on quantum information processing using nonlinear systems. Both these directions are natural extensions that we leave for future work.

Finally, we return full circle to our original motivation: by extracting more information from the quantum domain, we hope QNPs can ultimately improve our ability to control quantum systems. By enabling the efficient simulation of measurement-conditioned dynamics of measurement chains including QNPs, our STEOMs framework provides the necessary first step in the study of quantum feedback and control using QNPs. Such control is necessary for important quantum information processing tasks such as error correction [78–80], either via continuous monitoring and feedback [81, 82] or autonomous protocols via coherent quantum feedback [83].

Appendix A: Measurement chain description and table of symbols

In this Appendix section we expand on details of the measurement chain considered described by Eq. (1) that were omitted for brevity in the main text.

First, we define the interaction between the QS and QNP described by $\mathcal{L}_c$. We engineer this interaction to be non-reciprocal by balancing a coherent and a dissipative hopping interaction with an appropriately-chosen phase [84],

$$\mathcal{L}_c \hat{\rho} = -i \left[\frac{i}{2} \sum_m \Gamma_m \hat{a}_m^\dagger \hat{b}_m + h.c., \hat{\rho} \right] + \sum_m \Gamma_m \mathcal{D}[\hat{a}_m + \hat{b}_m]\hat{\rho}. \tag{A1}$$

Some features of this interaction are of note. Firstly, although the interaction has a dissipative component, it defines a coherent *link*, such that it allows for a non-separable joint quantum state of the QS and QNP. Secondly, for simplicity we require that one mode of the QS couples to at most one mode of the QNP, with strength Γ_m (although the coupling can also vanish if $\Gamma_m = 0$ for a given $\hat{a}_m$). Finally, we note that Eq. (A1) describes a standard circulator [84], and is therefore realized as a matter of course in cQED experiments.

For practical implementations in the cQED architecture, we consider for the QNP a model of K coupled Kerr nonlinear modes $\hat{b}_k, k \in [1, K]$, furnished by capacitively-shunted Josephson junctions [85]. The nonlinear modes have frequencies $\{\omega_k\}$ and nonlinearity strengths $\{\Lambda_k\}$, with linear coupling $\{g_{jk}\}$ between modes i and j. Then, in an appropriate interaction picture at frequencies $\{\omega_{dk}\}$ (close to $\{\omega_k\}$ respectively), the linear QNP Hamiltonian takes the form

$$\hat{\mathcal{H}}_{\text{QNP}} = -\sum_k \Delta_k \hat{b}_k^\dagger \hat{b}_k + \sum_{jk} g_{jk}(\hat{b}_j \hat{b}_k^\dagger + \hat{b}_j^\dagger \hat{b}_k), \tag{A2}$$

with detunings $\Delta_k = \omega_{dk} - \omega_k$, while the nonlinear component of the QNP Hamiltonian is given by

$$\hat{\mathcal{N}}_{\text{QNP}} = -\sum_k \frac{\Lambda_k}{2} \hat{b}_k^\dagger \hat{b}_k^\dagger \hat{b}_k \hat{b}_k. \tag{A3}$$

For simplicity, we assume $\Lambda_k \equiv \Lambda \ \forall \ k$, although we emphasize this is not necessary for QNPs.

Finally, the conditional dynamics of $\hat{\rho}^c$ under heterodyne monitoring are governed by the stochastic measurement superoperator $\mathcal{S}[\sqrt{\gamma_{\text{H}}}\hat{b}_k]$, given by:

$$\mathcal{S}[\sqrt{\gamma_{\text{H}}}\hat{b}_k]\hat{\rho}^c = \sqrt{\frac{\gamma_{\text{H}}}{2}}\left(\hat{b}_k\hat{\rho}^c + \hat{\rho}^c\hat{b}_k^\dagger - \langle \hat{b}_k + \hat{b}_k^\dagger \rangle^c\right)dW_{\mathcal{I}_k}(t)$$
$$+ \sqrt{\frac{\gamma_{\text{H}}}{2}}\left(-i\hat{b}_k\hat{\rho}^c + i\hat{\rho}^c\hat{b}_k^\dagger - \langle -i\hat{b}_k + i\hat{b}_k^\dagger \rangle^c\right)dW_{\mathcal{Q}_k}(t)$$
$$+ \gamma_{\text{H}}\mathcal{D}[\hat{b}_k]\hat{\rho}, \tag{A4}$$

Table II. Definition of measurement chain parameters. Superscripts (l) index parameters defining different quantum states for classification tasks.

Quantum System (QS)	
$\eta_m^{(l)}$	Drive strength on $\hat{a}_m$
ω_{dm}	Drive frequency on $\hat{a}_m$
$G_m^{(l)}$	Single-mode squeezing strength on $\hat{a}_m$
$\phi_m^{(l)}$	Single-mode squeezing phase
$G_{nm}^{(l)}$	Two-mode squeezing strength
$\phi_{nm}^{(l)}$	Two-mode squeezing phase
κ_m	Unmonitored loss for $\hat{a}_m$
κ	Total loss $\kappa_m + \Gamma_m$, equal for all m
Non-reciprocal coupling	
Γ_m	Coupling strength between modes $\hat{a}_m$ and $\hat{b}_m$
Quantum Nonlinear Processor (QNP)	
Λ	Kerr nonlinearity strength, equal for all k
Δ_k	Detuning $\omega_{dk} - \omega_k$
g_{jk}	Linear coupling between modes $\hat{b}_j$ and $\hat{b}_k$
γ_{H}	Monitoring rate for $\hat{b}_k$, equal for all k
γ	Total loss $\gamma_{\text{H}} + \Gamma_k$, equal for all k
Output layer	
$\mathcal{I}_k(t), \mathcal{Q}_k(t)$	Raw heterodyne records for QNP mode $\hat{b}_k$
I_k, Q_k	Filtered records for QNP mode $\hat{b}_k$
$\mathcal{T}$	Filtering window length

where $\langle \hat{o} \rangle^c$ indicates the conditional expectation value of an arbitrary operator $\hat{o}$ with respect to the measurement-conditioned quantum state, $\langle \hat{o} \rangle^c = \text{tr}\{\hat{\rho}^c \hat{o}\}$. Here $dW_{\mathcal{I}_k}, dW_{\mathcal{Q}_k}$ are independent Wiener increments describing measurement noise, satisfying $\langle dW_{\mathcal{I}_k} \rangle = \langle dW_{\mathcal{Q}_k} \rangle = 0$, and $\langle dW_{\mathcal{I}_k} dW_{\mathcal{Q}_{k'}} \rangle = 0, \langle dW_{\mathcal{I}_k} dW_{\mathcal{I}_{k'}} \rangle = \langle dW_{\mathcal{Q}_k} dW_{\mathcal{Q}_{k'}} \rangle = \delta_{k,k'}dt$. The Wiener increments are therefore related to the white noise terms introduced in Eq. (4) via $\xi_{\mathcal{I}_k} \equiv \frac{dW_{\mathcal{I}_k}}{dt}, \xi_{\mathcal{Q}_k} \equiv \frac{dW_{\mathcal{Q}_k}}{dt}$. The classical readout noise terms in Eq. (4) are also taken to obey white noise statistics.

On the other hand, the quantum noise contributions $\xi_{\mathcal{I}_k}^{\text{qm}}, \xi_{\mathcal{Q}_k}^{\text{qm}}$ introduced in Eq. (4) generally do *not* obey white noise statistics. Quantum trajectories $\hat{\rho}^c$ are conditioned on the heterodyne measurement record via the stochastic measurement superoperator $\mathcal{S}$. Quantum noise contributions depend on these quantum trajectories; more concretely, for example, $\xi_{\mathcal{I}_k}^{\text{qm}}(t) = \langle \hat{X}_k \rangle^c - \langle \hat{X}_k \rangle$, and thus describes the stochastic deviation of an observable conditioned on a specific quantum trajectory from the ensemble average. Hence the quantum noise contributions inherit nontrivial correlation and noise properties from quantum trajectories. Lastly, monitoring of QNP modes opens them up to linear damping at the rate γ_{H}, described by the standard dissipative superoperator $\mathcal{D}[\hat{o}] = \hat{o}\hat{\rho}\hat{o}^\dagger - \frac{1}{2}\{\hat{o}^\dagger\hat{o}, \hat{\rho}\}$.

Definitions of all parameters characterizing the general measurement chain are summarized in Table II. Specific parameter values used to generate the various figures in the main text are summarized in Table IV in the SI.

Appendix B: Truncated cumulants approach

For an arbitrary $\hat{\rho}$ describing a quantum measurement chain with $R = M + K$ modes, we can define its associated *characteristic function* $\chi(\vec{w}, \vec{w}^*)$ [86],

$$\chi(\vec{w}, \vec{w}^*) = \text{tr} \left\{ \hat{\rho} \prod_{j=1}^{R} e^{iw_j^* \hat{o}_j^\dagger} \prod_{l=1}^{R} e^{iw_l \hat{o}_l} \right\} \qquad (B1)$$

where $\hat{o}_j \in \{\hat{a}_m, \hat{b}_k\}$ describes any mode of the complete quantum measurement chain, and $\vec{w} = (w_1, \ldots, w_R)$ are auxiliary variables. Then, normal-ordered cumulants of the density matrix are formally defined via the log of the characteristic function (sometimes called the *generating function*),

$$C_{o_1^{\dagger p_1} \ldots o_R^{\dagger p_R} o_1^{q_1} \ldots o_R^{q_R}} \equiv \frac{\partial^{n_{\text{ord}}} \ln \chi(\vec{w}, \vec{w}^*)}{\prod_j \partial(iw_j^*)^{p_j} \prod_l \partial(iw_l)^{q_l}} \bigg|_{\vec{w} = \vec{w}^* = 0} \tag{B2}$$

where $n_{\text{ord}} = p_1 + \ldots + p_J + q_1 + \ldots + q_J$ defines the order of cumulants. Notwithstanding the complex formal definition, cumulants can be transparently related to more familiar operator expectation values: for a general operator $\hat{o}_j$, first-order cumulants $C_{o_j} \equiv \langle \hat{o}_j \rangle$ are simply expectation values, while second-order cumulants $C_{o_j o_l} = \langle \hat{o}_j \hat{o}_l \rangle - \langle \hat{o}_j \rangle \langle \hat{o}_l \rangle$ are their covariances. Expressions for higher-order cumulants become increasingly more involved, but can be systematically obtained, as discussed in the SI [47].

1. Numerical scheme: Stochastic Truncated Equations of Motion

The crucial advantage of cumulants as descriptors of a quantum state is that *specific* multimode quantum states admit particularly efficient representations when expressed in terms of cumulants. In particular, a quantum system in a product of coherent states is described entirely by its nonzero first-order cumulants; all cumulants with $n_{\text{ord}} > 1$ vanish (see SI [47] for derivations). Multimode quantum states that are defined entirely by their first *and* second-order cumulants admit Gaussian phase-space representations, and are thus labelled Gaussian states. States with nonzero cumulants of third or higher-order are thus by definition non-Gaussian states [12, 87].

Our numerical approach leverages this economy of representation by using normal-ordered cumulants as a set of dynamical variables for quantum modes of the measurement chain. For example, for specific systems such as coherently-driven linear bosonic systems initialized in a Gaussian state, cumulants of order $n_{\text{ord}} \leq 2$ are sufficient, as such systems can be shown to persist in Gaussian states. If the QNPs considered in this paper were also linear, our model would satisfy this requirement. However,

our interest is precisely in the role of the QNP nonlinearity. In this case, the cumulants describing the measurement chain and their dependencies are shown schematically in Fig. 8 for $M = K = 1$. In particular, the nonlinearity generates states with nonzero cumulants of $n_{\text{ord}} > 2$. In the most general case, the dynamical equations for these cumulants couple to all orders, forming an infinite hierarchy of equations that does not close.

To obtain a tractable numerical method, we consider an ansatz wherein the quantum state of the complete measurement is described entirely by cumulants up to a finite order $n_{\text{ord}} \leq n_{\text{trunc}}$; all cumulants of order $n_{\text{ord}} > n_{\text{trunc}}$ are thus set to zero, truncating the hierarchy and yielding a closed set of equations for the retained nonzero cumulants. In this paper, we choose $n_{\text{trunc}} = 2$ for a quantum measurement chain defined entirely by its first and second-order cumulants, although the truncation can similarly be carried out at higher order. The resulting Stochastic Truncated Equations Of Motion (STEOMs) form the basis of our numerical approach in this paper.

We emphasize that the truncated cumulants ansatz has some important differences when compared to standard linearization approaches. As seen in Fig. 8, note that both the nonlinearity and measurement terms couple second-order cumulants to first-order cumulants: this fact proves critical for the quantum state classification tasks we consider in this paper.

The utility of the STEOMs is naturally determined by the validity of the truncated cumulants ansatz. Since we are specifically interested in nonlinear quantum systems, which can generate higher-order cumulants in dynamics, one must ask when such an ansatz may hold. In the SI [47], we carry out a detailed analysis of the regimes of validity of this framework, together with benchmarking against exact solutions and full SME integration. We find that the STEOMs provide a very good approximation of full SME integration provided the strength of nonlinearity of QNP modes is weak relative to their loss rates; good agreement remains even up to $\Lambda_k/\gamma \sim 0.1$, which is around an order of magnitude larger than the strongest nonlinearity we consider in this paper.

2. Semi-analytic scheme: Nonlinear van Kampen expansion based of the Fokker-Planck equation

In addition to enabling the (S)TEOMs as a practical numerical method for multimode measurement chains, the truncated cumulants approach is also central to our main analytic tool: a description of quantum dynamics that is perturbative in the nonlinearity of the measurement chain. This analysis is enabled by the close connection between normal-ordered cumulants and the positive-P representation via the characteristic function. More precisely, the positive-P representation is simply
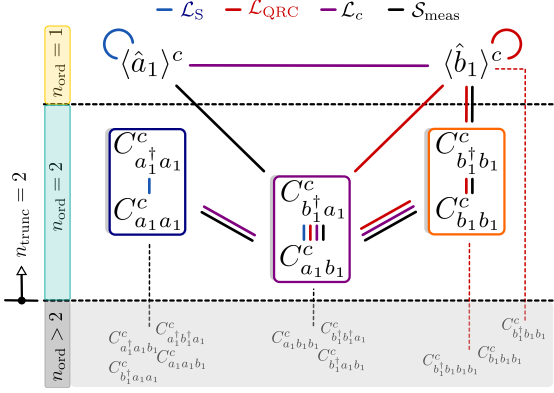
Figure 8. Schematic representation of stochastic truncated equations of motion (STEOMs), shown for simplicity for the $M = K = 1$ case. Black, red, purple, and blue lines indicate dependencies between cumulants due to measurement, QNP nonlinearity, system-QNP coupling, and system dynamics respectively. Dashed lines indicate dependencies on cumulants of order higher than $n_{\rm trunc} = 2$ in this case, which can arise due to the nonlinearity generating non-Gaussian states.

Figure 9. $\mathcal{C}_{\rm max}$ as a function of QNP nonlinearity Λ for instances of Task I with different amplitudes $\mathcal{A}$ as indicated, calculated using numerical integration of the STEOMs. For each task instance, the top panel shows $||\delta\boldsymbol{\mu}||$ and $\sigma^{2(1)}_{\delta\boldsymbol{\mu}}$, calculated using the NVK approximation (solid curves), and STEOMs (dots).

the Fourier transform of the characteristic function [86],

$$
\mathcal{P}(\vec{\mathcal{O}}, \vec{\mathcal{O}}^\dagger) = \frac{1}{\pi^{2J}} \int \prod_{j=1}^{R} d^2 w_j d^2 w_j^* \; e^{-i w_j^* \mathcal{O}_j^\dagger} e^{-i w_j \mathcal{O}_j} \chi(\vec{w}, \vec{w}^*)
$$
(B3)

The dynamics of the positive-P distribution follow a Fokker-Planck equation that is equivalent to Eq. (1), while also employing normal-ordered cumulants as its natural dynamical variable set. We use this connection to first obtain an approximate Fokker-Planck equation for dynamics of the measurement chain in powers of the QNP nonlinearity. This directly allows us to obtain semi-analytic solutions for the TEOMs, Eqs. (10), (11), from which classification metrics such as the Fisher's discriminant can be readily evaluated. Our analysis and its results, which are used throughout the main text, are detailed in the SI, Sec. E and G.

Appendix C: Scaling of QNP performance with nonlinearity for fixed $\mathcal{E}_{\rm eff}$

In Sec. III A, we explored how QNP parameters can be optimized to improve classification performance. However, a plain reading of the expression for the mean separation $\delta\boldsymbol{\mu}$, Eq. (13), would appear to suggest that an even simpler strategy may be to simply increase the QNP nonlinearity strength Λ, which monotonically scales $\delta\boldsymbol{\mu}$. However, in practice this dependence is much more complex: for a classification task defined by a given $\mathcal{A}$, varying Λ also varies $\mathcal{E}_{\rm eff}$ via Eq. (15). This changes the operating conditions of the QNP and directly influences the mean separation and noise properties.
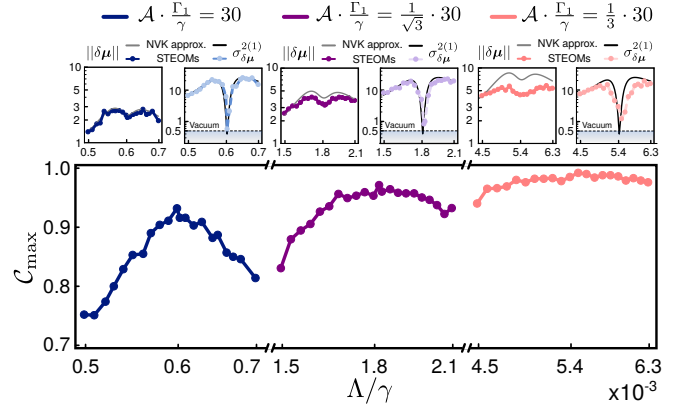
Therefore, to observe and understand the scaling of $||\delta\boldsymbol{\mu}||$ with nonlinearity, we must keep $\mathcal{E}_{\rm eff}$ fixed while varying the nonlinearity. Practically, we do so by using QNPs of varying nonlinearity to perform separate instances of Task I, characterized by different values of $\mathcal{A} \cdot \frac{\Gamma_1}{\gamma}$. For each instance, the achieved $\mathcal{C}_{\rm max}$ as a function of Λ is plotted in Fig. 9.

We immediately note that for decreasing $\mathcal{A} \cdot \frac{\Gamma_1}{\gamma}$, the required QNP nonlinearity to reach the optimal $\mathcal{C}_{\rm max}$ increases, as required by the form of $\mathcal{E}_{\rm eff}$. The crosses indicate three specific QNPs, one for each instance of the considered task, with nonlinearity Λ such that $\mathcal{E}_{\rm eff}$ is the same. Comparing these QNPs, we now clearly see that the optimal $\mathcal{C}_{\rm max}$ increases with increasing Λ. In the top panel of Fig. 9, we plot the mean separation amplitude $||\delta\boldsymbol{\mu}||$ and the projected noise $\sigma^{2(l)}_{\delta\boldsymbol{\mu}}$ for $l = 1$ for each considered instance of Task I, under the NVK approximation and using integration of STEOMs. To lowest order in nonlinearity as captured by the NVK approximation, the projected noise properties remain the same for each instance. In contrast, the mean separation increases with nonlinearity, indicative of the sought-after scaling. Unchanged noise properties with increasing mean separation imply an increase in $D_{\rm F}$ and hence classification accuracy.

Note that with increasing nonlinearity Λ, the agreement between the NVK approximation and exact integration of the STEOMs is reduced. In particular, the exact mean separation is lower than the NVK approximation, while the projected noise is larger. Both effects serve to reduce $D_{\rm F}$ and hence limit the improvement in $\mathcal{C}_{\rm max}$ to below that predicted by the NVK approximation. These observations are an example of saturation that is higher-order in the QNP nonlinearity and hence not captured by the NVK approximation, and are commonplace in strongly-driven nonlinear quantum systems [12].

Appendix D: Models of standard linear quantum amplifiers

In this section we provide the models for both phase-preserving and phase-sensitive linear quantum amplifiers deployed for readout in cQED, and which we use as benchmarks in Sec. IV of the main text. Note that our description of the measurement chain, Eq. (1), is general enough to include this standard paradigm, by neglecting the nonlinear contribution to the Hamiltonian, $\hat{\mathcal{N}}_{\text{QNP}} \to 0$.

In particular, by taking $\hat{\mathcal{H}}_{\text{QNP}} \to \hat{\mathcal{H}}_{\text{PP}}$ where

$$\hat{\mathcal{H}}_{\text{PP}} = -\sum_k \Delta_k \hat{b}_k^\dagger \hat{b}_k + G_{\text{PP}}(-i\hat{b}_1\hat{b}_2 + h.c.), \quad \text{(D1)}$$

we are able to describe QS readout using linear phase-preserving quantum amplifiers.

For phase-sensitive (PS) amplifiers on the other hand, we can set $\hat{\mathcal{H}}_{\text{QNP}} \to \hat{\mathcal{H}}_{\text{PS}}$, where

$$\hat{\mathcal{H}}_{\text{PS}} = -\sum_k \Delta_k \hat{b}_k^\dagger \hat{b}_k + G_{\text{PS}}(-i\hat{b}_1^2 + h.c.). \quad \text{(D2)}$$

Appendix E: Quantum Chernoff Bound: additional details

In this Appendix section we provide some additional details of the calculation of the Quantum Chernoff bound (QCB) used in Sec. IV of the main text to determine the optimal discrimination bounds for quantum state discrimination tasks.

To define the QCB as used in the main text, we first introduce the quantity [58, 59]:

$$Q(\hat{\rho}^{(l)}, \hat{\rho}^{(p)}) = -\min_{s \in [0,1]} \log \text{tr}\left[(\hat{\rho}^{(l)})^s (\hat{\rho}^{(p)})^{1-s}\right] \quad \text{(E1)}$$

We compute the QCB using only the quantum states of the modes that are intended to be monitored. For Task I, this is simply mode $\hat{a}_1$ for the QS. We can then define the QCB for the QS ζ_{QS} as:

$$\zeta_{\text{QS}} \equiv Q\left(\text{tr}_{\hat{a}_2}\left[\hat{\rho}_{\text{QS}}^{(1)}\right], \text{tr}_{\hat{a}_2}\left[\hat{\rho}_{\text{QS}}^{(2)}\right]\right) \quad \text{(E2)}$$

where $\text{tr}_{\hat{o}}[\hat{\rho}]$ is used to trace out the sector corresponding to mode $\hat{o}$ from the density matrix $\hat{\rho}$. For the general quantum processor, only mode $\hat{b}_1$ is monitored, so the QCB ζ analogously becomes:

$$\zeta \equiv Q\left(\text{tr}_{\hat{b}_2}\left[\hat{\rho}_{\text{QNP}}^{(1)}\right], \text{tr}_{\hat{b}_2}\left[\hat{\rho}_{\text{QNP}}^{(2)}\right]\right) \quad \text{(E3)}$$

Note that the QNP and the PS amplifier are both single-mode devices for Task I, so the trace operation has no effect in those cases.

The expressions in Eqs. (E2), (E3) can be computed straightforwardly if the quantum states are Gaussian. These simplified expressions are derived in Ref. [59]; we employ these results in plotting the QCB in the main text, Fig. 5.

Table III. QS parameters ($M = 1$ only) defining the third classification task considered, in units of QS mode total loss rate κ.

	Task III		
l	$G_1^{(l)}$	$n_1^{\text{th}(l)}$	$\eta_1^{(l)}$
5	0.0	0.1	0.50$\mathcal{A}$
6	0.0	0.8	0.50$\mathcal{A}$

Appendix F: Non-classical noise due to the QNP alone: classifying thermal states

In the main text, the considered Tasks I and II both require classification of states that exhibit some degree of quantum correlations, namely at least one state exhibits some degree of single or two-mode squeezing. As a result, the observed quantum (i.e. sub-vacuum) noise observed at the QNP output will in general contain a contribution attributable to the quantum correlations of QS signals, and cannot be attributed to the QNP alone.

In this Appendix section, we show that non-classical correlations can be generated solely by appropriate QNP models, and they can be manipulated for use in classification in the same way as has been demonstrated for both Task I and II. To this end, we first introduce the modified Liouvillian for the QS:

$$\mathcal{L}_{\text{QS}}\hat{\rho} = -i[\hat{\mathcal{H}}_{\text{QS}}^{(l)}, \hat{\rho}] + \sum_m \kappa_m (n_m^{\text{th}(l)} + 1)\mathcal{D}[\hat{a}_m]\hat{\rho} + \kappa_m n_m^{\text{th}(l)} \mathcal{D}[\hat{a}_m^\dagger]\hat{\rho}, \quad \text{(F1)}$$

where in contrast to Eq. (2) each QS mode is now coupled to a finite temperature bath. For simplicity, we further restrict ourselves to a single-mode QS being monitored by a single-mode QNP, $M = K = 1$; all QS parameters for mode $m = 2$ are set to zero. Secondly, we turn off single-mode squeezing of the QS mode, $G_1^{(l)} = 0 \, \forall \, l$, so that the QS states to be distinguished have no nontrivial quantum correlations. Instead, the states to be distinguished are characterized entirely by different thermal bath temperatures and hence occupation numbers, $n_m^{\text{th}(l)}$. While not straightforward to generate in practice, this choice is ideal to test whether non-classical correlations can be generated by the QNP alone, and whether they can still be manipulated by the QNP. All parameters characterizing Task III are summarized in Table III.

It suffices to consider only the NVK approximation and to then plot both the the mean separation of measured QNP quadratures $||\delta\boldsymbol{\mu}||$ and the projected noise in Fig. 10, as a function of QNP nonlinearity. We note that the qualitative features are as observed in Fig. 4, and therefore reiterate the main results in the paper: the ability to compute correlations *in situ*, and the ability to control noise in measured quadratures differently from their mean value. Crucially, projected noise below vacuum in Fig. 10(b) can now solely be attributed to the
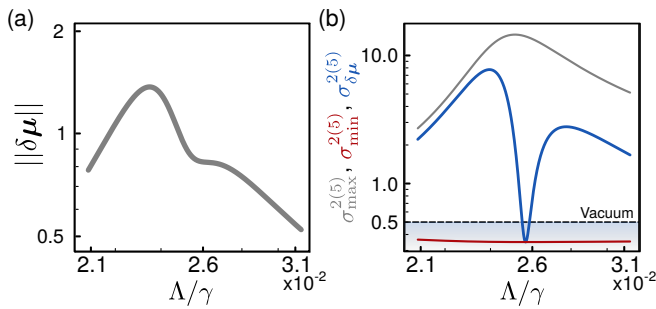
Figure 10. QNP performance metrics for Task III, classification of thermal states, within the NVK approximation. (a) Mean separation $||\delta\boldsymbol{\mu}||$ and (b) noise properties as a function of nonlinearity Λ/γ. The reduction of projected noise to below vacuum levels where input signals are classical (originating from a thermal state) indicates that the QNP alone can give rise to non-classical correlations in the measured outputs.

action of the QNP, as the QS states are purely thermal and do not demonstrate such non-classical correlations; in fact, they exhibit fluctuations above the vacuum value due to the coupling to a finite temperature bath. This emphasizes that the QNP can provide the ability to not just manipulate existing quantum correlations, but also provide useful quantum correlations itself.

We emphasize that the QNP model we consider here is based on coherently-driven Kerr oscillators, which by themselves are not ideal squeezers. However, our general approach applies to much more general QNP models, where more non-classical correlations could be generated.

[1] C. M. Caves, Quantum limits on noise in linear amplifiers, Physical Review D **26**, 1817 (1982).

[2] T. Yamamoto, K. Inomata, M. Watanabe, K. Matsuba, T. Miyazaki, W. D. Oliver, Y. Nakamura, and J. S. Tsai, Flux-driven Josephson parametric amplifier, Applied Physics Letters **93**, 042510 (2008).

[3] A. A. Clerk, M. H. Devoret, S. M. Girvin, F. Marquardt, and R. J. Schoelkopf, Introduction to quantum noise, measurement, and amplification, Reviews of Modern Physics **82**, 1155 (2010).

[4] A. Roy and M. Devoret, Introduction to parametric amplification of quantum signals with Josephson circuits, Comptes Rendus Physique Quantum microwaves / Micro-ondes quantiques, **17**, 740 (2016).

[5] J. Aumentado, Superconducting Parametric Amplifiers: The State of the Art in Josephson Parametric Amplifiers, IEEE Microwave Magazine **21**, 45 (2020).

[6] N. Bergeal, R. Vijay, V. E. Manucharyan, I. Siddiqi, R. J. Schoelkopf, S. M. Girvin, and M. H. Devoret, Analog information processing at the quantum limit with a Josephson ring modulator, Nature Physics **6**, 296 (2010).

[7] C. Macklin, K. O'Brien, D. Hover, M. E. Schwartz, V. Bolkhovsky, X. Zhang, W. D. Oliver, and I. Siddiqi, A near–quantum-limited Josephson traveling-wave parametric amplifier, Science **350**, 307 (2015).

[8] M. A. Castellanos-Beltran, K. D. Irwin, G. C. Hilton, L. R. Vale, and K. W. Lehnert, Amplification and squeezing of quantum noise with a tunable Josephson metamaterial, Nature Physics **4**, 929 (2008).

[9] J. Y. Qiu, A. Grimsmo, K. Peng, B. Kannan, B. Lienhard, Y. Sung, P. Krantz, V. Bolkhovsky, G. Calusine, D. Kim, A. Melville, B. M. Niedzielski, J. Yoder, M. E. Schwartz, T. P. Orlando, I. Siddiqi, S. Gustavsson, K. P. O'Brien, and W. D. Oliver, Broadband squeezed microwaves and amplification with a Josephson travelling-wave parametric amplifier, Nature Physics **19**, 706 (2023).

[10] M. P. da Silva, D. Bozyigit, A. Wallraff, and A. Blais, Schemes for the observation of photon correlation functions in circuit QED with linear detectors, Physical Review A **82**, 043804 (2010).

[11] C. A. Ryan, B. R. Johnson, J. M. Gambetta, J. M. Chow, M. P. da Silva, O. E. Dial, and T. A. Ohki, Tomography via correlation of noisy measurement records, Physical Review A **91**, 022118 (2015).

[12] S. Boutin, D. M. Toyli, A. V. Venkatramani, A. W. Eddins, I. Siddiqi, and A. Blais, Effect of Higher-Order Nonlinearities on Amplification and Squeezing in Josephson Parametric Amplifiers, Physical Review Applied **8**, 054030 (2017).

[13] J. Rovny, Z. Yuan, M. Fitzpatrick, A. I. Abdalla, L. Futamura, C. Fox, M. C. Cambria, S. Kolkowitz, and N. P. de Leon, Nanoscale covariance magnetometry with diamond quantum sensors, Science **378**, 1301 (2022).

[14] T. Schweigler, M. Gluza, M. Tajik, S. Sotiriadis, F. Cataldini, S.-C. Ji, F. S. Møller, J. Sabino, B. Rauer, J. Eisert, and J. Schmiedmayer, Decay and recurrence of non-Gaussian correlations in a quantum many-body system, Nature Physics **17**, 559 (2021).

[15] A. Blais, S. M. Girvin, and W. D. Oliver, Quantum information processing and quantum optics with circuit quantum electrodynamics, Nature Physics **16**, 247 (2020).

[16] W. Heisenberg, Über den anschaulichen Inhalt der quantentheoretischen Kinematik und Mechanik, Zeitschrift fur Physik **43**, 172 (1927).

[17] H. A. Haus and J. A. Mullen, Quantum Noise in Linear Amplifiers, Physical Review **128**, 2407 (1962).

[18] R. S. Bondurant, Quantum noise properties of a nonlinear amplifier, Physical Review Letters **71**, 1709 (1993).

[19] J. M. Epstein, K. B. Whaley, and J. Combes, Quantum limits on noise for a class of nonlinear amplifiers, Physical Review A **103**, 052415 (2021).

[20] R. Yanagimoto, R. Nehra, R. Hamerly, E. Ng, A. Marandi, and H. Mabuchi, Quantum Nondemolition Measurements with Optical Parametric Amplifiers for Ultrafast Universal Quantum Information Processing, PRX Quantum **4**, 010333 (2023).

[21] C. Laflamme and A. A. Clerk, Quantum-limited amplifi-

cation with a nonlinear cavity detector, Physical Review A **83**, 033803 (2011).

[22] R. Vijay, M. H. Devoret, and I. Siddiqi, Invited review article: The josephson bifurcation amplifier, Review of Scientific Instruments **80**, 111101 (2009), https://doi.org/10.1063/1.3224703.

[23] M. D. Reed, L. DiCarlo, B. R. Johnson, L. Sun, D. I. Schuster, L. Frunzio, and R. J. Schoelkopf, High-fidelity readout in circuit quantum electrodynamics using the jaynes-cummings nonlinearity, Phys. Rev. Lett. **105**, 173601 (2010).

[24] Y. Sunada, K. Yuki, Z. Wang, T. Miyamura, J. Ilves, K. Matsuura, P. A. Spring, S. Tamate, S. Kono, and Y. Nakamura, Photon-Noise-Tolerant Dispersive Readout of a Superconducting Qubit Using a Nonlinear Purcell Filter, PRX Quantum **5**, 010307 (2024).

[25] G. Liu, T.-C. Chien, X. Cao, O. Lanes, E. Alpern, D. Pekker, and M. Hatridge, Josephson parametric converter saturation and higher order effects, Applied Physics Letters **111**, 202603 (2017), publisher: American Institute of Physics.

[26] C. Liu, T.-C. Chien, M. Hatridge, and D. Pekker, Optimizing Josephson-ring-modulator-based Josephson parametric amplifiers via full Hamiltonian control, Physical Review A **101**, 042323 (2020), publisher: American Physical Society.

[27] N. E. Frattini, U. Vool, S. Shankar, A. Narla, K. M. Sliwa, and M. H. Devoret, 3-wave mixing Josephson dipole element, Applied Physics Letters **110**, 222603 (2017).

[28] V. Sivak, N. Frattini, V. Joshi, A. Lingenfelter, S. Shankar, and M. Devoret, Kerr-Free Three-Wave Mixing in Superconducting Quantum Circuits, Physical Review Applied **11**, 054060 (2019), publisher: American Physical Society.

[29] E. M. Izhikevich, *Dynamical Systems in Neuroscience: The Geometry of Excitability and Bursting* (The MIT Press, 2006).

[30] A. A. Faisal, L. P. J. Selen, and D. M. Wolpert, Noise in the nervous system, Nature Reviews Neuroscience **9**, 292 (2008).

[31] G. Deco, E. T. Rolls, and R. Romo, Stochastic dynamics as a principle of brain function, Progress in Neurobiology **88**, 1 (2009).

[32] O. I. Rumyantsev, J. A. Lecoq, O. Hernandez, Y. Zhang, J. Savall, R. Chrapkiewicz, J. Li, H. Zeng, S. Ganguli, and M. J. Schnitzer, Fundamental bounds on the fidelity of sensory cortical coding, Nature **580**, 100 (2020).

[33] W. T. Newsome, K. H. Britten, and J. A. Movshon, Neuronal correlates of a perceptual decision, Nature **341**, 52 (1989).

[34] K. H. Britten, M. N. Shadlen, W. T. Newsome, and J. A. Movshon, The analysis of visual motion: a comparison of neuronal and psychophysical performance, Journal of Neuroscience **12**, 4745 (1992).

[35] S. Panzeri, M. Moroni, H. Safaai, and C. D. Harvey, The structures and functions of correlations in neural population codes, Nature Reviews Neuroscience **23**, 551 (2022).

[36] L. F. Abbott and P. Dayan, The Effect of Correlated Variability on the Accuracy of a Population Code, Neural Computation **11**, 91 (1999).

[37] S. Panzeri, F. Montani, G. Notaro, C. Magri, and R. S. Peterson, Population Coding, in *Analysis of Parallel Spike Trains*, edited by S. Grün and S. Rotter (Springer

US, Boston, MA, 2010) pp. 303–319.

[38] J. S. Montijn, G. T. Meijer, C. S. Lansink, and C. M. A. Pennartz, Population-Level Neural Codes Are Robust to Single-Neuron Variability from a Multidimensional Coding Perspective, Cell Reports **16**, 2486 (2016).

[39] M. Shamir and H. Sompolinsky, Nonlinear population codes, Neural Computation **16**, 1105 (2004).

[40] R. Romo, A. Hernández, A. Zainos, and E. Salinas, Correlated Neuronal Discharges that Increase Coding Efficiency during Perceptual Discrimination, Neuron **38**, 649 (2003).

[41] R. Moreno-Bote, J. Beck, I. Kanitscheider, X. Pitkow, P. Latham, and A. Pouget, Information-limiting correlations, Nature Neuroscience **17**, 1410 (2014).

[42] I. Kanitscheider, R. Coen-Cagli, and A. Pouget, Origin of information-limiting noise correlations, Proceedings of the National Academy of Sciences **112**, E6973 (2015).

[43] R. Bartolo, R. C. Saunders, A. R. Mitz, and B. B. Averbeck, Information-Limiting Correlations in Large Neural Populations, Journal of Neuroscience **40**, 1668 (2020).

[44] M. Silveri, E. Zalys-Geller, M. Hatridge, Z. Leghtas, M. H. Devoret, and S. M. Girvin, Theory of remote entanglement via quantum-limited phase-preserving amplification, Physical Review A **93**, 062310 (2016).

[45] C. W. Sandbo Chang, M. Simoen, J. Aumentado, C. Sabín, P. Forn-Díaz, A. M. Vadiraj, F. Quijandría, G. Johansson, I. Fuentes, and C. M. Wilson, Generating Multimode Entangled Microwaves with a Superconducting Parametric Cavity, Physical Review Applied **10**, 044019 (2018).

[46] N. G. Van kampen, *Stochastic Processes in Physics and Chemistry (Third Edition)*, North-Holland Personal Library (Elsevier, Amsterdam, 2007).

[47] See supplementary information available online., .

[48] H. M. Wiseman and G. J. Milburn, *Quantum Measurement and Control* (Cambridge University Press, Cambridge, 2009).

[49] W. Heisenberg, C. Eckart, and F. C. Hoyt, *The physical principles of the quantum theory,* (The University of Chicago Press, Chicago, Ill., 1930).

[50] C. W. Gardiner and M. J. Collett, Input and output in damped quantum systems: Quantum stochastic differential equations and the master equation, Physical Review A **31**, 3761 (1985).

[51] B. Abdo, K. Sliwa, F. Schackert, N. Bergeal, M. Hatridge, L. Frunzio, A. D. Stone, and M. Devoret, Full Coherent Frequency Conversion between Two Propagating Microwave Modes, Physical Review Letters **110**, 173902 (2013).

[52] B. Abdo, A. Kamal, and M. Devoret, Nondegenerate three-wave mixing with the Josephson ring modulator, Physical Review B **87**, 014508 (2013).

[53] K. Sliwa, M. Hatridge, A. Narla, S. Shankar, L. Frunzio, R. Schoelkopf, and M. Devoret, Reconfigurable Josephson Circulator/Directional Amplifier, Physical Review X **5**, 041020 (2015).

[54] D. T. Smithey, M. Beck, M. G. Raymer, and A. Faridani, Measurement of the Wigner distribution and the density matrix of a light mode using optical homodyne tomography: Application to squeezed states and the vacuum, Physical Review Letters **70**, 1244 (1993).

[55] S. Kotler, G. A. Peterson, E. Shojaee, F. Lecocq, K. Cicak, A. Kwiatkowski, S. Geller, S. Glancy, E. Knill, R. W. Simmonds, J. Aumentado, and J. D. Teufel, Direct ob-

servation of deterministic macroscopic entanglement, Science **372**, 622 (2021).

[56] S. Khan and H. E. Türeci, Frequency Combs in a Lumped-Element Josephson-Junction Circuit, Physical Review Letters **120**, 153601 (2018), publisher: American Physical Society.

[57] M. Dykman, *Fluctuating Nonlinear Oscillators: From Nanomechanics to Quantum Superconducting Circuits* (OUP Oxford, 2012).

[58] K. M. R. Audenaert, J. Calsamiglia, R. Muñoz-Tapia, E. Bagan, L. Masanes, A. Acin, and F. Verstraete, Discriminating States: The Quantum Chernoff Bound, Physical Review Letters **98**, 160501 (2007).

[59] J. Calsamiglia, R. Muñoz-Tapia, L. Masanes, A. Acin, and E. Bagan, Quantum Chernoff bound as a measure of distinguishability between density matrices: Application to qubit and Gaussian states, Physical Review A **77**, 032311 (2008).

[60] H. Chernoff, A Measure of Asymptotic Efficiency for Tests of a Hypothesis Based on the sum of Observations, The Annals of Mathematical Statistics **23**, 493 (1952).

[61] S. Pirandola and S. Lloyd, Computable bounds for the discrimination of Gaussian states, Physical Review A **78**, 012331 (2008).

[62] F. Nielsen, Revisiting Chernoff Information with Likelihood Ratio Exponential Families, Entropy **24**, 1400 (2022).

[63] D. Sarchi, I. Carusotto, M. Wouters, and V. Savona, Coherent dynamics and parametric instabilities of microcavity polaritons in double-well systems, Physical Review B **77**, 125324 (2008).

[64] G. Vidal and R. F. Werner, Computable measure of entanglement, Physical Review A **65**, 032314 (2002).

[65] Y.-X. Wang, C. Wang, and A. A. Clerk, Quantum Nonreciprocal Interactions via Dissipative Gauge Symmetry, PRX Quantum **4**, 010306 (2023).

[66] L. Orr, S. A. Khan, N. Buchholz, S. Kotler, and A. Metelmann, High-Purity Entanglement of Hot Propagating Modes Using Nonreciprocity, PRX Quantum **4**, 020344 (2023).

[67] D. Marković, N. Leroux, M. Riou, F. Abreu Araujo, J. Torrejon, D. Querlioz, A. Fukushima, S. Yuasa, J. Trastoy, P. Bortolotti, and J. Grollier, Reservoir computing with the frequency, phase, and amplitude of spin-torque nano-oscillators, Applied Physics Letters **114**, 012409 (2019).

[68] G. Angelatos, S. A. Khan, and H. E. Türeci, Reservoir Computing Approach to Quantum State Measurement, Physical Review X **11**, 041062 (2021).

[69] L. C. G. Govia, G. J. Ribeill, G. E. Rowlands, H. K. Krovi, and T. A. Ohki, Quantum reservoir computing with a single nonlinear oscillator, Physical Review Research **3**, 013077 (2021).

[70] R. A. Bravo, K. Najafi, X. Gao, and S. F. Yelin, Quantum Reservoir Computing Using Arrays of Rydberg Atoms, PRX Quantum **3**, 030325 (2022).

[71] P. Mujal, R. Martínez-Peña, G. L. Giorgi, M. C. Soriano, and R. Zambrini, Time-series quantum reservoir computing with weak and projective measurements, npj Quantum Information **9**, 1 (2023).

[72] F. Hu, S. A. Khan, N. T. Bronn, G. Angelatos, G. E. Rowlands, G. J. Ribeill, and H. E. Türeci, Overcoming the coherence time barrier in quantum machine learning on temporal data, Nature Communications **15**, 7491 (2024).

[73] A. Senanian, S. Prabhu, V. Kremenetski, S. Roy, Y. Cao, J. Kline, T. Onodera, L. G. Wright, X. Wu, V. Fatemi, and P. L. McMahon, Microwave signal processing using an analog quantum reservoir computer, Nature Communications **15**, 7490 (2024).

[74] J. Nokkala, R. Martínez-Peña, G. L. Giorgi, V. Parigi, M. C. Soriano, and R. Zambrini, Gaussian states of continuous-variable quantum systems provide universal and versatile reservoir computing, Communications Physics **4**, 10.1038/s42005-021-00556-w (2021).

[75] J. Nokkala, R. Martínez-Peña, R. Zambrini, and M. C. Soriano, High-Performance Reservoir Computing With Fluctuations in Linear Networks, IEEE Transactions on Neural Networks and Learning Systems , 1 (2021).

[76] J. García-Beni, G. L. Giorgi, M. C. Soriano, and R. Zambrini, Scalable Photonic Platform for Real-Time Quantum Reservoir Computing, Physical Review Applied **20**, 014051 (2023).

[77] J. Dudas, B. Carles, E. Plouet, F. A. Mizrahi, J. Grollier, and D. Marković, Quantum reservoir computing implementation on coherently coupled quantum oscillators, npj Quantum Information **9**, 1 (2023).

[78] C. Ahn, A. C. Doherty, and A. J. Landahl, Continuous quantum error correction via quantum feedback control, Physical Review A **65**, 042301 (2002).

[79] M. Gregoratti and R. F. Werner, On quantum error-correction by classical feedback in discrete time, Journal of Mathematical Physics **45**, 2600 (2004).

[80] J. Cramer, N. Kalb, M. A. Rol, B. Hensen, M. S. Blok, M. Markham, D. J. Twitchen, R. Hanson, and T. H. Taminiau, Repeated quantum error correction on a continuously encoded qubit by real-time feedback, Nature Communications **7**, 11526 (2016).

[81] A. C. Doherty and K. Jacobs, Feedback control of quantum systems using continuous state estimation, Physical Review A **60**, 2700 (1999).

[82] Z. K. Minev, S. O. Mundhada, S. Shankar, P. Reinhold, R. Gutiérrez-Jáuregui, R. J. Schoelkopf, M. Mirrahimi, H. J. Carmichael, and M. H. Devoret, To catch and reverse a quantum jump mid-flight, Nature **570**, 200 (2019).

[83] S. Lloyd, Coherent quantum feedback, Physical Review A **62**, 022108 (2000).

[84] A. Metelmann and A. A. Clerk, Nonreciprocal photon transmission and amplification via reservoir engineering, Phys. Rev. X **5**, 021025 (2015).

[85] J. Clarke and A. I. Braginski, *The SQUID Handbook* (Wiley-VCH, Weinheim, 2003).

[86] H. J. Carmichael, *Statistical Methods in Quantum Optics 1 - Master Equations and Fokker-Planck Equations | Howard J. Carmichael | Springer* (2002).

[87] Y.-S. Ra, A. Dufour, M. Walschaers, C. Jacquard, T. Michel, C. Fabre, and N. Treps, Non-Gaussian quantum states of a multimode light field, Nature Physics **16**, 144 (2020).

[88] R. J. Glauber, Coherent and Incoherent States of the Radiation Field, Physical Review **131**, 2766 (1963).

[89] E. C. G. Sudarshan, Equivalence of Semiclassical and Quantum Mechanical Descriptions of Statistical Light Beams, Physical Review Letters **10**, 277 (1963).

[90] J. Koch, T. M. Yu, J. Gambetta, A. A. Houck, D. I. Schuster, J. Majer, A. Blais, M. H. Devoret, S. M. Girvin, and R. J. Schoelkopf, Charge-insensitive qubit design de-

rived from the Cooper pair box, Physical Review A **76**, 042319 (2007).

[91] P. D. Drummond and D. F. Walls, Quantum theory of optical bistability. I. Nonlinear polarisability model, Journal of Physics A: Mathematical and General **13**, 725 (1980).

[92] M. I. Dykman and M. A. Krivoglaz, Theory of fluctuational transitions between stable states of a nonlinear oscillator, Soviet Journal of Experimental and Theoretical Physics **77**, 30 (1979).

[93] J. K. Breslin and G. J. Milburn, Conditional variance reduction by measurements on correlated field modes, Physical Review A **55**, 1430 (1997).

[94] O. Černotík, D. V. Vasilyev, and K. Hammerer, Adia-

batic elimination of Gaussian subsystems from quantum dynamics under continuous measurement, Physical Review A **92**, 012124 (2015).

[95] J. Zhang and K. Mølmer, Prediction and retrodiction with continuously monitored Gaussian states, Physical Review A **96**, 062131 (2017).

[96] J. R. Johansson, P. D. Nation, and F. Nori, QuTiP 2: A Python framework for the dynamics of open quantum systems, Computer Physics Communications **184**, 1234 (2013).

[97] C. Gardiner, *Stochastic Methods: A Handbook for the Natural and Social Sciences*, 4th ed., Springer Series in Synergetics (Springer-Verlag, Berlin Heidelberg, 2009).

SUPPLEMENTARY INFORMATION

Appendix A: Table of simulation parameters

Simulation parameters that are used to generate plots in the various figures in the main text are summarized in Table IV.

Table IV. Specific parameters for classification tasks considered in this paper. Brackets $[\cdot,\cdot]$ indicate range of values, while slashes (/) indicate discrete choices.

Parameter	Task I				Task II
	Fig. 3	Fig. 4	Fig. 5(d)	Fig. 9	Fig. 7
$\mathcal{A}$	10	80	10	10/17.3/30	80
κ_1	0.5	0.5	0.5	0.5	0.5
κ_2	1.0	1.0	1.0	1.0	0.5
κ	1.0	1.0	1.0	1.0	1.0
Γ_1	0.5	0.5	0.5	0.5	0.5
Γ_2	0.0	0.0	0.0	0.0	0.5
Λ_1 $(\times 10^{-3})$	5.5	[6.5,9.0]	7.4	$(\frac{30}{\mathcal{A}})^2[0.5,0.6]$	[6.6,7.7]
Λ_2 $(\times 10^{-3})$	–	–	–	–	[6.6,7.7]
Δ_1	-0.67	-3.0	-0.83	-0.67	[0.8,1.5]
Δ_2	–	–	–	–	[-1.6,-3.0]
g_{12}	–	–	–	–	[0.1,2.5]
γ_{H}	0.5	4.0	0.5	0.5	4.0
γ	1.0	4.5	1.0	1.0	4.5

Appendix B: Quantum cumulants

The use of quantum cumulants as an efficient representation of certain QNP states is essential to our work. In this section, we provide a simplified introduction to quantum cumulants as a description of quantum states.

1. Definition in terms of the generating function

The starting point for our discussion of cumulants is via the characteristic function $\chi(z, z^*)$, defined in terms of the density operator $\hat{\rho}$ for an arbitrary quantum state as

$$\chi(z, z^*) = \text{tr}\left\{\hat{\rho} e^{iz^* \hat{b}^\dagger} e^{iz\hat{b}}\right\}. \tag{B1}$$

The characteristic function as defined in Eq. (B1) has the property that it defines all *normal-ordered* operator averages with respect to the quantum state $\hat{\rho}$; simply evaluating the derivative of the characteristic function [86] at $z = z^* = 0$ yields

$$\frac{\partial^{p+q}}{\partial(iz^*)^p \partial(iz)^q}(\text{tr}\{\rho e^{iz^*\hat{b}^\dagger} e^{iz\hat{b}}\})\bigg|_{z=z^*=0} = \langle \hat{b}^{\dagger p} \hat{b}^q \rangle. \tag{B2}$$

It is then possible to define a distribution function for calculating these normal-ordered moments,

$$P(\beta, \beta^\dagger) = \frac{1}{\pi^2} \int d^2 z \; \chi(z, z^*) e^{-iz^* \beta^*} e^{-iz\beta} \tag{B3}$$

for which we use the suggestive notation $P(\beta, \beta^\dagger)$, as it is easily shown to be none other than the Glauber-Sudarshan P distribution [86, 88, 89]:

$$\hat{\rho} = \int d^2\beta \; P(\beta, \beta^\dagger)|\beta\rangle\langle\beta| \tag{B4}$$

Having recalled the connection of the characteristic function to the quantum state defined by $\hat{\rho}$ and its normal-ordered moments, we now define a quantum generating function $G : \mathbb{C} \times \mathbb{C} \to \mathbb{R}$,

$$G(z, z^*) = \ln \text{tr}\{\chi(z, z^*)\} = \ln \text{tr}\{\hat{\rho} e^{iz^*\hat{b}^\dagger} e^{iz\hat{b}}\}. \tag{B5}$$

The *quantum cumulants* are defined through Eq. (B5) by the series expansion of the generating function $G(z, z^*)$ which yields all quantum cumulants $C_{a^{\dagger p} a^q}$ for any $p, q \in \mathbb{N}^2$

$$\ln \text{tr}\{\hat{\rho} e^{iz^*\hat{b}^\dagger} e^{iz\hat{b}}\} = \sum_{n=1}^{\infty} \sum_{\substack{p, q \in \mathbb{N} \\ p+q=n}} \frac{1}{p!q!} C_{b^{\dagger p} b^q}(iz^*)^p (iz)^q, \tag{B6}$$

or equivalently,

$$C_{b^{\dagger p} b^q} \equiv \frac{\partial^{p+q} G}{\partial(iz^*)^p \partial(iz)^q}(z, z^*)\bigg|_{z=z^*=0} \tag{B7}$$

So far, cumulants may simply appear to be a mathematical construct. However, it can be shown that normal-ordered cumulants and normal-ordered moments can be mapped to each other one-to-one. To see this, we can think of $G(z, z^*)$ as the composite of the logarithm function $\log(\chi) = \ln \chi$ and $\chi(z, z^*) = \text{tr}\left\{\hat{\rho} e^{iz^*\hat{b}^\dagger} e^{iz\hat{b}}\right\}$, and employ multi-variable version of Faà di Bruno's formula for the n-th order partial derivative where $n = p + q$,

$$\frac{\partial^{p+q} G}{\partial(iz^*)^p \partial(iz)^q}(z, z^*) = \sum_{\pi} \left(\log^{(|\pi|)}(\chi(z, z^*)) \prod_{B \in \pi} \frac{\partial^{p'+q'} \chi}{\partial(iz^*)^{p'} \partial(iz)^{q'}}(z, z^*) \right), \tag{B8}$$

where π defines the set of possible partitions of the ordered set $\{iz^*, \cdots, iz^*, iz, \cdots iz\}$ with iz^* appearing p times and iz appearing q times. B indicates elements in this set of partitions, and p' and q' respectively counts the number

of iz^* and iz in each partition element B. Finally, $f^{(|\pi|)}(\cdot)$ denotes the $|\pi|$-order derivative of the function $f(\cdot)$, where $|\pi|$ denotes the cardinality of the set π. By noting that $\chi(0,0) = \text{tr}(\hat{\rho}) = 1$, and then using $\log^{(|\pi|)}(\chi(0,0)) = (\ln \chi)^{(|\pi|)}|_{\chi=1} = (-1)^{|\pi|-1}(|\pi|-1)!$ we obtain

$$
\begin{aligned}
C_{b^{\dagger p}b^q} &= \left.\frac{\partial^{p+q}G}{\partial(iz^*)^p\partial(iz)^q}(z,z^*)\right|_{z=z^*=0} \\
&= \sum_\pi \left((-1)^{|\pi|-1}(|\pi|-1)! \prod_{B\in\pi} \langle \hat{b}^{\dagger p'}\hat{b}^{q'}\rangle \right).
\end{aligned}
\tag{B9}
$$

As such, cumulants can be thought of as a reparameterization of the normal-ordered moments that typically appear in observables measured in an experiment. We will see in the next two appendices that this reparameterization leads to a much more efficient description of certain quantum states, which can then be leveraged to construct an efficient computational framework.

For our purposes, we will find it more useful to express quantum moments in terms of quantum cumulants. This inverse transformation can be performed by thinking of $\chi(z,z^*) = \text{tr}\left\{\hat{\rho}e^{iz^*\hat{b}^\dagger}e^{iz\hat{b}}\right\}$ as the composite of the exponential function $\exp(G) = e^G$ and $G(z,z^*) = \ln\text{tr}\{\hat{\rho}e^{iz^*\hat{b}^\dagger}e^{iz\hat{b}}\}$. Again, Faà di Bruno's formula gives

$$
\frac{\partial^{p+q}\chi}{\partial(iz^*)^p\partial(iz)^q}(z,z^*) = \sum_\pi \left(\exp^{(|\pi|)}(G(z,z^*)) \prod_{B\in\pi} \frac{\partial^{p'+q'}G}{\partial(iz^*)^{p'}\partial(iz)^{q'}}(z,z^*) \right)
\tag{B10}
$$

where π defines the set of possible partitions of the ordered set $\{iz^*, \cdots, iz^*, iz, \cdots iz\}$ with iz^* appearing p times and iz appearing q times, B indicates elements in this set of partitions, and p' and q' respectively counts the number of iz^* and iz in each partition element B. Again, noting that $G(0,0) = \ln\text{tr}(\hat{\rho}) = 0$, and then using $\exp^{(|\pi|)}(G(0,0)) = (e^G)^{(|\pi|)}|_{G=0} = 1$ we get for n-th order partial derivative where $n = p + q$:

$$
\begin{aligned}
\langle \hat{b}^{\dagger p}\hat{b}^q\rangle &= \left.\frac{\partial^{p+q}\chi}{\partial(iz^*)^p\partial(iz)^q}(z,z^*)\right|_{z=z^*=0} \\
&= \left.\sum_\pi \prod_{B\in\pi} \frac{\partial^{p'+q'}G}{\partial(iz^*)^{p'}\partial(iz)^{q'}}(z,z^*)\right|_{z=z^*=0} \\
&= \sum_\pi \prod_{B\in\pi} C_{b^{\dagger p'}b^{q'}}.
\end{aligned}
\tag{B11}
$$

Finally, it is straightforward to generalize the definition of quantum cumulants to the bosonic N-mode case. For a complex-valued vector $\boldsymbol{z} = (z_1, z_2, \cdots, z_N)$, we can define the generating function $G: \mathbb{C}^N \times \mathbb{C}^N \to \mathbb{R}$ as:

$$
G(\boldsymbol{z}, \boldsymbol{z}^*) = \ln\text{tr}\left\{\hat{\rho}e^{iz_1^*\hat{b}_1^\dagger}\cdots e^{iz_N^*\hat{b}_N^\dagger}e^{iz_1\hat{b}_1}\cdots e^{iz_N\hat{b}_N}\right\}.
\tag{B12}
$$

The series expansion of $G(\boldsymbol{z}, \boldsymbol{z}^*)$ yields all quantum cumulants

$$
\left.\frac{\partial^n G}{\partial(iz_1^*)^{p_1}\cdots\partial(iz_N^*)^{p_N}\partial(iz_1)^{q_1}\cdots\partial(iz_N)^{q_N}}(\boldsymbol{z}, \boldsymbol{z}^*)\right|_{\boldsymbol{z}=\boldsymbol{z}^*=\boldsymbol{0}} = C_{b_1^{\dagger p_1}\cdots b_N^{\dagger p_N}b_1^{q_1}\cdots b_N^{q_N}}.
\tag{B13}
$$

where $p_1 + \cdots + p_N + q_1 + \cdots + q_N = n$. The transformation between moments and cumulants in multi-mode case can be obtained by simply replacing all single-mode moments and cumulants in Eq.(B9), (B11) with multi-mode moments and cumulants corresponding to normal-ordered operator $\hat{o}_1\hat{o}_2\cdots\hat{o}_n = \hat{b}_1^{\dagger p_1}\cdots\hat{b}_N^{\dagger p_N}\hat{b}_1^{q_1}\cdots\hat{b}_N^{q_N}$:

$$
C_{o_1 o_2 \cdots o_n} = \sum_\pi \left((-1)^{|\pi|-1}(|\pi|-1)! \prod_{B\in\pi} \langle \hat{o}_i : i \in B\rangle \right),
\tag{B14}
$$

$$
\langle \hat{o}_1\hat{o}_2\cdots\hat{o}_n\rangle = \sum_\pi \prod_{B\in\pi} C_{o_i : i\in B},
\tag{B15}
$$

where π defines the set of possible partitions of operators in the n-order moment, and B indicates elements in this set of partitions.

2. Cumulants for some simple quantum states

Eqs. (B9), (B11) and their multi-mode generalization Eqs. (B14), (B15) show that quantum cumulants and quantum moments have a one-to-one correspondence through logarithm and exponential. On the other hand, for two complex-valued vectors $\boldsymbol{z} = (z_1, z_2, \cdots, z_N)$ and $\boldsymbol{\beta} = (\beta_1, \beta_2, \cdots, \beta_N)$, the P-representation of quantum state and quantum characteristic function are mutually related by Fourier transformation [86]:

$$P(\boldsymbol{\beta}, \boldsymbol{\beta}^*) = \frac{1}{\pi^2} \int \chi(\boldsymbol{z}, \boldsymbol{z}^*) e^{-i\boldsymbol{z} \cdot \boldsymbol{\beta}^*} e^{-i\boldsymbol{z}^* \boldsymbol{\beta}} \; d^2\boldsymbol{z}, \tag{B16}$$

$$\chi(\boldsymbol{z}, \boldsymbol{z}^*) = \int P(\boldsymbol{\beta}, \boldsymbol{\beta}^*) e^{i\boldsymbol{z} \cdot \boldsymbol{\beta}^*} e^{i\boldsymbol{z}^* \boldsymbol{\beta}} \; d^2\boldsymbol{\beta}. \tag{B17}$$

where the inner product is conventionally $\boldsymbol{z} \cdot \boldsymbol{\beta}^* = z_1^* \beta_1^* + \cdots + z_N^* \beta_N^*$ and $\boldsymbol{z}^* \cdot \boldsymbol{\beta} = z_1 \beta_1 + \cdots + z_N \beta_N$. This result ensures the one-to-one correspondence between a quantum state and the quantum cumulants.

Importantly, the truncation of cumulants to certain discrete orders naturally characterizes some special types of quantum states. In this section, we will show that states with only nonzero first order cumulants correspond to coherent states, while states with only nonzero first and second cumulants orders correspond to Gaussian states.

We first consider a product of coherent states $|\boldsymbol{\beta}_0\rangle \equiv |\beta_1, \beta_2, \cdots, \beta_N\rangle$. In Eq. (B12), we then set $\hat{\rho} = |\boldsymbol{\beta}_0\rangle \langle \boldsymbol{\beta}_0|$, following which the RHS gives

$$\begin{aligned} &\ln \text{tr} \left\{ |\boldsymbol{\beta}_0\rangle \langle \boldsymbol{\beta}_0| \, e^{iz_1^* \hat{b}_1^\dagger} \cdots e^{iz_N^* \hat{b}_N^\dagger} e^{iz_1 \hat{b}_1} \cdots e^{iz_N \hat{b}_N} \right\} \\ &= \ln \left(\langle \boldsymbol{\beta}_0 | \, e^{iz_1^* \hat{b}_1^\dagger} \cdots e^{iz_N^* \hat{b}_N^\dagger} e^{iz_1 \hat{b}_1} \cdots e^{iz_N \hat{b}_N} \, |\boldsymbol{\beta}_0\rangle \right) \\ &= \ln e^{iz_1^* \beta_1^* + \cdots + iz_N^* \beta_N^* + iz_1 \beta_1 + \cdots + iz_N \beta_N} \\ &= iz_1^* \beta_1^* + \cdots + iz_N^* \beta_N^* + iz_1 \beta_1 + \cdots + iz_N \beta_N, \end{aligned} \tag{B18}$$

From Eq. (B13), we obtain non-trivial derivative contributions only for first-order derivatives. Thus for all $k \in \{1, 2, \cdots, N\}$, we have $C_{b_k} = \beta_k$ and $C_{b_k^\dagger} = \beta_k^*$, so that

$$(C_{b_1}, C_{b_2}, \cdots, C_{b_N}) = \boldsymbol{\beta}_0, \tag{B19a}$$

$$\left(C_{b_1^\dagger}, C_{b_2^\dagger}, \cdots, C_{b_N^\dagger} \right) = \boldsymbol{\beta}_0^* \tag{B19b}$$

We can also show that physical states with vanishing cumulants of order greater than one are necessarily coherent states. We begin with the normal-ordered characteristic function for such states, which is given by $\chi(\boldsymbol{z}, \boldsymbol{z}^*) = e^{G(i\boldsymbol{z}^*, i\boldsymbol{z})} = e^{i\boldsymbol{z} \cdot \boldsymbol{\beta}_0^* + i\boldsymbol{z}^* \cdot \boldsymbol{\beta}_0}$. The corresponding P-representation is simply the Fourier transform of the characteristic function, and is given by

$$\begin{aligned} P(\boldsymbol{\beta}, \boldsymbol{\beta}^*) &= \frac{1}{\pi^2} \int \chi(\boldsymbol{z}, \boldsymbol{z}^*) e^{-i\boldsymbol{z} \cdot \boldsymbol{\beta}^*} e^{-i\boldsymbol{z}^* \cdot \boldsymbol{\beta}} \; d^2\boldsymbol{z} \\ &= \frac{1}{\pi^2} \int e^{i\boldsymbol{z} \cdot \boldsymbol{\beta}_0^* + i\boldsymbol{z}^* \cdot \boldsymbol{\beta}_0} e^{-i\boldsymbol{z} \cdot \boldsymbol{\beta}^*} e^{-i\boldsymbol{z}^* \cdot \boldsymbol{\beta}} \; d^2\boldsymbol{z} \\ &= \delta(\boldsymbol{\beta} - \boldsymbol{\beta}_0), \end{aligned} \tag{B20}$$

which describes a product of coherent states. As a result, the states with all cumulants or order greater than one vanishing are exactly the collection of all coherent states.

Next, we consider quantum states whose quasi-probability distribution is Gaussian. Recall that the P, Q, and Wigner phase-space representations are related by the Weierstrass transformation, which always maps one Gaussian distribution to another Gaussian distribution [86]:

$$W(\boldsymbol{\beta}, \boldsymbol{\beta}^*) = \frac{2}{\pi} \int d^2\widetilde{\boldsymbol{\beta}} \; e^{-2|\boldsymbol{\beta} - \widetilde{\boldsymbol{\beta}}|^2} P(\widetilde{\boldsymbol{\beta}}, \widetilde{\boldsymbol{\beta}}^*), \tag{B21}$$

$$Q(\boldsymbol{\beta}, \boldsymbol{\beta}^*) = \frac{2}{\pi} \int d^2\widetilde{\boldsymbol{\beta}} \; e^{-2|\boldsymbol{\beta} - \widetilde{\boldsymbol{\beta}}|^2} W(\widetilde{\boldsymbol{\beta}}, \widetilde{\boldsymbol{\beta}}^*). \tag{B22}$$

The relation between the P-representation and characteristic function in Eqs. (B16), (B17) respectively is via Fourier transform, which also maps one Gaussian distribution to another Gaussian distribution.

Via this series of transformations, we therefore see that a quantum state with a Gaussian Wigner distribution will also have a Gaussian characteristic function, and hence must be entirely defined by cumulants of up to second-order (by definition of a Gaussian distribution).

Figure 11. Schematic showing a single-mode QNP as part of a simplified measurement chain: a single frequency coherent drive is incident on the QNP via an input channel that is not measured, while the second port of the QNP is monitored. The QNP can be realized in cQED as a tunable-frequency Kerr oscillator using a capacitively shunted SQUID loop [90].

Appendix C: Truncated Equations of Motion approach applied to the single-mode Kerr QNP

To benchmark the computational approach developed in this paper based on quantum cumulants, we will use a simple system as a testbed: a single coherently-drive Kerr oscillator. This system also forms a single mode of the QNPs we employ.

The resulting system schematic is shown in Fig. 11. We begin by analyzing the *unconditional* dynamics of the resulting system, described by the quantum master equation

$$d\hat{\rho} \equiv \mathcal{L}\hat{\rho} \, dt = \mathcal{L}_{\text{QNP}}\hat{\rho} \, dt - i[\hat{\mathcal{H}}_d, \hat{\rho}] \, dt \tag{C1}$$

where the drive Hamiltonian in the frame rotating at the drive frequency ω_d is then

$$\hat{\mathcal{H}}_d = \eta_b(\hat{b} + \hat{b}^\dagger), \tag{C2}$$

and we have defined for convenience $\hat{b} \equiv \hat{b}_1$. The coherent drive is incident on the QNP directly, via a separate input port with loss rate Γ_1. This situation describes QNP processing of classical signals encoded in the generally time-dependent coherent drive amplitude and phase. The Liouvillian $\mathcal{L}_{\text{QRC}}$ describing the single-mode QNP takes the form

$$\mathcal{L}_{\text{QNP}}\hat{\rho} = -i\left[-\Delta_1\hat{b}^\dagger\hat{b} - \frac{\Lambda_1}{2}\hat{b}^\dagger\hat{b}^\dagger\hat{b}\hat{b}, \hat{\rho}\right] + (\gamma_1 + \Gamma_1)\mathcal{D}[\hat{b}]\hat{\rho} \tag{C3}$$

where $\Delta = \omega_d - \omega_1$ is the detuning of the drive frequency from the bare QNP mode frequency ω. For simplicity of notation, in what follows we introduce for convenience the total linear damping $\gamma = \gamma_1 + \Gamma_1$ of the nonlinear mode.

1. Obtaining TEOMs for a single coherently-driven Kerr oscillator

Our approach requires the expression of moments of a multimode quantum system in terms of their associated cumulants. Moments and cumulants are related via their generating functions, which allows one to write a general prescription for writing an arbitrary normal-ordered joint system moment of order n in terms of a series of cumulants of order 1 up to n. This takes the specific form given by Eq. (B15),

$$\langle \hat{o}_1\hat{o}_2 \ldots \hat{o}_n \rangle = \sum_\pi \prod_{B \in \pi} C_{o_i : i \in B}, \tag{C4}$$

where π defines the set of possible partitions of operators in the n-order moment, and B indicates elements in this set of partitions. The rather opaque form can be simplified by considering a specific case. We will use the above prescription to obtain the TEOMs for the single-mode QNP.

This minimal system already includes the key nonlinear element of the QNP, and therefore the most complex terms that need to be analyzed will arise already in these equations. We start by writing down the equation for $\langle \hat{b} \rangle = \text{tr}\{\mathcal{L}\hat{\rho}\hat{b}\}$,

$$\langle \dot{\hat{b}} \rangle = \left(i\Delta_1 - \frac{\gamma}{2}\right)\langle \hat{b} \rangle + i\Lambda_1\langle \hat{b}^\dagger\hat{b}\hat{b} \rangle - i\eta_b. \tag{C5}$$

To rewrite the above equation in terms of cumulants, we must express the third-order moment $\langle \hat{b}^\dagger \hat{b} \hat{b} \rangle$ in terms of cumulants. The possible partitions of the third-order moment are given by: $\pi \in \{(\hat{b}^\dagger \hat{b} \hat{b}), (\hat{b}^\dagger \hat{b}, \hat{b}) \times 2, (\hat{b} \hat{b}, \hat{b}^\dagger), (\hat{b}^\dagger, \hat{b}, \hat{b})\}$. Then, using Eq. (C4), we obtain

$$\langle \hat{b}^\dagger \hat{b} \hat{b} \rangle = C_{b^\dagger bb} + 2C_{b^\dagger b}\langle \hat{b} \rangle + C_{bb}\langle \hat{b}^\dagger \rangle + \langle \hat{b}^\dagger \rangle\langle \hat{b} \rangle^2. \tag{C6}$$

Using the above, we can now rewrite Eq. (C5) in terms of cumulants as

$$\langle \dot{\hat{b}} \rangle = \left(i\Delta_1 - \frac{\gamma}{2} \right) \langle \hat{b} \rangle + i\Lambda_1 \langle \hat{b}^\dagger \rangle\langle \hat{b} \rangle^2 - i\eta_b + i\Lambda_1 \left(C_{bb}\langle \hat{b}^\dagger \rangle + 2C_{b^\dagger b}\langle \hat{b} \rangle + C_{b^\dagger bb} \right) \tag{C7}$$

To similarly obtain dynamical equations for cumulants $C_{b^\dagger b}, C_{bb}$, we write equations of motion for the second-order moments $\langle \hat{b}^\dagger \hat{b} \rangle$ and $\langle \hat{b} \hat{b} \rangle$ respectively. Starting with the former, we find:

$$\frac{d}{dt}\langle \hat{b}^\dagger \hat{b} \rangle = -\gamma\langle \hat{b}^\dagger \hat{b} \rangle + i\eta_b\langle \hat{b} \rangle - i\eta_b\langle \hat{b}^\dagger \rangle$$
$$= -\gamma C_{b^\dagger b} - \gamma\langle \hat{b}^\dagger \rangle\langle \hat{b} \rangle + i\eta_b\langle \hat{b} \rangle - i\eta_b\langle \hat{b}^\dagger \rangle \tag{C8}$$

where in the second line we rewrite the second-order moment in terms of cumulants. From this we can easily write $dC_{b^\dagger b} = d\langle \hat{b}^\dagger \hat{b} \rangle - (d\langle \hat{b}^\dagger \rangle)\langle \hat{b} \rangle - (d\langle \hat{b} \rangle)\langle \hat{b}^\dagger \rangle$, which yields the dynamical equation for $C_{b^\dagger b}$:

$$\dot{C}_{b^\dagger b} = -\gamma C_{b^\dagger b} - i\Lambda_1 \left(C_{bb}\langle \hat{b}^\dagger \rangle^2 - C_{bb}^*\langle \hat{b} \rangle^2 \right) - i\Lambda_1 \left(C_{b^\dagger bb}\langle \hat{b}^\dagger \rangle - C_{b^\dagger bb}^*\langle \hat{b} \rangle \right) \tag{C9}$$

Finally, we write the equation of motion for $\langle \hat{b} \hat{b} \rangle$,

$$\frac{d}{dt}\langle \hat{b} \hat{b} \rangle = (2i\Delta_1 - \gamma)\langle \hat{b} \hat{b} \rangle - 2i\eta_b\langle \hat{b} \rangle + i\Lambda_1\langle \hat{b} \hat{b} \rangle + i2\Lambda_1\langle \hat{b}^\dagger \hat{b} \hat{b} \hat{b} \rangle$$
$$= (2i\Delta_1 - \gamma + i\Lambda_1)\left(C_{bb} + \langle \hat{b} \rangle\langle \hat{b} \rangle \right) - 2i\eta_b\langle \hat{b} \rangle + i2\Lambda_1\langle \hat{b}^\dagger \hat{b} \hat{b} \hat{b} \rangle \tag{C10}$$

where we once again rewrite second-order moments in terms of cumulants in the second line. We now need to express the remaining fourth-order moment $\langle \hat{b}^\dagger \hat{b} \hat{b} \hat{b} \rangle$ in terms of cumulants. The possible partitions of this fourth-order moment are given by $\pi \in \{(\hat{b}^\dagger \hat{b} \hat{b} \hat{b}), (\hat{b}^\dagger \hat{b} \hat{b}, \hat{b}) \times 3, (\hat{b} \hat{b} \hat{b}, \hat{b}^\dagger), (\hat{b}^\dagger \hat{b}, \hat{b} \hat{b}) \times 3, (\hat{b}^\dagger \hat{b}, \hat{b}, \hat{b}) \times 3, (\hat{b} \hat{b}, \hat{b}^\dagger, \hat{b}) \times 3, (\hat{b}^\dagger, \hat{b}, \hat{b}, \hat{b})\}$, so that

$$\langle \hat{b}^\dagger \hat{b} \hat{b} \hat{b} \rangle = C_{b^\dagger bbb} + 3\langle \hat{b} \rangle C_{b^\dagger bb} + \langle \hat{b}^\dagger \rangle C_{bbb}$$
$$+ 3C_{b^\dagger b}C_{bb} + 3C_{b^\dagger b}\langle \hat{b} \rangle^2 + 3C_{bb}|\langle \hat{b} \rangle|^2 + \langle \hat{b}^\dagger \rangle\langle \hat{b} \rangle\langle \hat{b} \rangle\langle \hat{b} \rangle \tag{C11}$$

From this we can finally write $dC_{bb} = d\langle \hat{b} \hat{b} \rangle - 2(d\langle \hat{b} \rangle)\langle \hat{b} \rangle$, which yields the dynamical equation for C_{bb}:

$$\dot{C}_{bb} = (2i\Delta_1 - \gamma + i\Lambda_1)\left(C_{bb} + \langle \hat{b} \rangle^2 \right) - 2i\eta_b\langle \hat{b} \rangle$$
$$+ i2\Lambda_1 \left(C_{b^\dagger bbb} + 3\langle \hat{b} \rangle C_{b^\dagger bb} + \langle \hat{b}^\dagger \rangle C_{bbb} \right)$$
$$+ i2\Lambda_1 \left(3C_{b^\dagger b}C_{bb} + 3C_{b^\dagger b}\langle \hat{b} \rangle^2 + 3C_{bb}|\langle \hat{b} \rangle|^2 + \langle \hat{b}^\dagger \rangle\langle \hat{b} \rangle^3 \right)$$
$$- 2\left(i\Delta_1 - \frac{\gamma}{2} \right)\langle \hat{b} \rangle^2 - i2\Lambda_1\langle \hat{b}^\dagger \rangle\langle \hat{b} \rangle^3 + 2i\eta_b\langle \hat{b} \rangle$$
$$- i2\Lambda_1 \left(C_{bb}|\langle \hat{b} \rangle|^2 + 2C_{b^\dagger b}\langle \hat{b} \rangle^2 + C_{b^\dagger bb}\langle \hat{b} \rangle \right) \tag{C12}$$

which finally simplifies to:

$$\dot{C}_{bb} = (2i\Delta_1 - \gamma + i\Lambda_1)C_{bb} + i\Lambda_1\langle \hat{b} \rangle^2 + i4\Lambda_1|\langle \hat{b} \rangle|^2 C_{bb} + i6\Lambda_1 C_{b^\dagger b}C_{bb}$$
$$+ i2\Lambda_1 C_{b^\dagger b}\langle \hat{b} \rangle^2 + i2\Lambda_1 C_{b^\dagger bbb} + i4\Lambda_1 C_{b^\dagger bb}\langle \hat{b} \rangle + i2\Lambda_1 C_{bbb}\langle \hat{b}^\dagger \rangle \tag{C13}$$

Clearly, second-order cumulants couple to cumulants of third and fourth order. This is a general feature of nonlinear quantum systems: moments and thus cumulants of a certain order can couple to those of higher-order, leading to an infinite hierarchy of equations that do not form a closed set.

However, an infinite set of normal-ordered cumulants is not necessarily required to describe all multimode quantum states. In particular, as we have seen in Sec. B of the SI, a multimode quantum system in a product of coherent states

is described entirely by its nonzero first-order cumulants; all cumulants of order $p+q > 1$ vanish. Multimode quantum states that are defined entirely by their first and second-order cumulants admit Gaussian phase-space representations, and are labelled Gaussian states. States with nonzero cumulants of third or higher-order are thus by definition non-Gaussian states [12, 87].

Our numerical approach leverages this efficient representation of specific multimode quantum states in terms of cumulants. In particular, we consider an ansatz wherein the quantum state of the complete measurement is described entirely by cumulants up to a finite order $p + q \leq n_{\text{trunc}}$; all cumulants of order $p + q > n_{\text{trunc}}$ are thus set to zero, truncating the hierarchy and yielding a closed set of equations for the retained nonzero cumulants. In this work, we choose $n_{\text{trunc}} = 2$ for a quantum measurement chain defined entirely by its first and second-order cumulants, although the truncation can similarly be carried out at higher order. The resulting Truncated Equations Of Motion (TEOMs), and their generalization to *conditional* dynamics of multimode quantum systems, the Stochastic Truncate Equations of Motion (STEOMs), form the basis of our computational approach in this paper. We have developed a computer algebra approach that automates the above process of calculating contributions to equations of motion and expressing arbitrary moments in terms of cumulants, thus allowing the systematic truncation necessary to arrive at TEOMs (and STEOMs) introduced in the main text.

2. Calculating STEOMs - stochastic measurement contributions to truncated cumulant dynamics

The measurement-conditioned dynamics of a quantum system under heterodyne monitoring is given by a stochastic master equation (SME). We can now calculate the contributions to equations of motion for cumulants due to stochastic contributions via an SME. To illustrate the salient features of the calculation, a single homodyne measurement superoperator for monitoring a general mode $\hat{b}_k$ suffices:

$$\mathcal{S}_{\text{hom}}[\sqrt{\gamma_k}\hat{b}_k]\hat{\rho} = \sqrt{\frac{\gamma_k}{2}}\left(\hat{b}_k\hat{\rho} + \hat{\rho}\hat{b}_k^\dagger - \langle\hat{b}_k + \hat{b}_k^\dagger\rangle\right)dW_{\mathcal{I}_k}(t) \tag{C14}$$

For convenience, we will then write a general stochastic master equation in the form:

$$d\hat{\rho}^c = \mathcal{L}\hat{\rho}^c \, dt + \mathcal{S}_{\text{hom}}[\sqrt{\gamma_k}\hat{b}_k]\hat{\rho}^c \tag{C15}$$

where $\mathcal{L}$ defines all contributions to the master equation governing deterministic evolution. We can thus write for the differential of conditional expectation of an arbitrary operator $\hat{o}$:

$$d\langle\hat{o}\rangle^c = \text{tr}\{\mathcal{L}\hat{\rho}^c\hat{o}\} \, dt + \sqrt{\frac{\gamma_k}{2}}\left(C_{ob_k}^c + C_{b_k^\dagger o}^c\right)dW_{\mathcal{I}_k}(t) \tag{C16}$$

Analogously, we can write for arbitrary second-order moments:

$$d\langle\hat{o}_1\hat{o}_2\rangle^c = \text{tr}\{\mathcal{L}\hat{\rho}^c\hat{o}_1\hat{o}_2\} \, dt + \sqrt{\frac{\gamma_k}{2}}\left(\langle\hat{o}_1\hat{o}_2\hat{b}_k\rangle^c + \langle\hat{b}_k^\dagger\hat{o}_1\hat{o}_2\rangle^c - \langle\hat{o}_1\hat{o}_2\rangle^c\langle\hat{b}_k\rangle^c - \langle\hat{b}_k^\dagger\rangle^c\langle\hat{o}_1\hat{o}_2\rangle^c\right)dW_{\mathcal{I}_k}(t) \tag{C17}$$

We will now use the above results obtain the equation of motion for an arbitrary *normal-ordered* cumulant $C_{o_1o_2}^c = \langle\hat{o}_1\hat{o}_2\rangle^c - \langle\hat{o}_1\rangle^c\langle\hat{o}_2\rangle^c$. The important step arises in writing the differential of this cumulant using Ito's lemma:

$$dC_{o_1o_2}^c = d\langle\hat{o}_1\hat{o}_2\rangle^c - (d\langle\hat{o}_1\rangle^c)\langle\hat{o}_2\rangle^c - \langle\hat{o}_1\rangle^c(d\langle\hat{o}_2\rangle^c)$$
$$- (d\langle\hat{o}_1\rangle^c)(d\langle\hat{o}_2\rangle^c) \tag{C18}$$

where the term on the second line arises from the modified chain rule in Ito calculus. We can now proceed to obtaining the individual contributions to the above equation.

We start with the first term in Eq. (C18), which was calculated in Eq. (C17). To proceed, it will prove convenient to write moments higher than first-order in terms of their corresponding cumulants. We begin with the second-order moments:

$$d\langle\hat{o}_1\hat{o}_2\rangle^c = \text{tr}\{\mathcal{L}\hat{\rho}^c\hat{o}_1\hat{o}_2\} \, dt +$$
$$\sqrt{\frac{\gamma_k}{2}}\left(\langle\hat{o}_1\hat{o}_2\hat{b}_k\rangle^c + \langle\hat{b}_k^\dagger\hat{o}_1\hat{o}_2\rangle^c - C_{o_1o_2}^c\langle\hat{b}_k\rangle^c - \langle\hat{b}_k^\dagger\rangle^c C_{o_1o_2}^c \right.$$
$$\left. - \langle\hat{o}_2\rangle^c\langle\hat{o}_1\rangle^c\langle\hat{b}_k\rangle^c - \langle\hat{o}_2\rangle^c\langle\hat{o}_1\rangle^c\langle\hat{b}_k^\dagger\rangle^c\right)dW_{\mathcal{I}_k}(t) \tag{C19}$$

Next, we write down the expressions relating third-order moments to their corresponding cumulants using Eq. (C4):

$$
\begin{aligned}
C^c_{o_1 o_2 b_k} = \langle \hat{o}_1 \hat{o}_2 \hat{b}_k \rangle^c &- C^c_{o_1 o_2} \langle \hat{b}_k \rangle^c - C^c_{o_1 b_k} \langle \hat{o}_2 \rangle^c - C^c_{o_2 b_k} \langle \hat{o}_1 \rangle^c \\
&- \langle \hat{o}_2 \rangle^c \langle \hat{o}_1 \rangle^c \langle \hat{b}_k \rangle^c
\end{aligned}
\tag{C20a}
$$

$$
\begin{aligned}
C^c_{b^\dagger_k o_1 o_2} = \langle \hat{b}^\dagger_k \hat{o}_1 \hat{o}_2 \rangle^c &- C^c_{o_1 o_2} \langle \hat{b}^\dagger_k \rangle^c - C^c_{b^\dagger_k o_1} \langle \hat{o}_2 \rangle^c - C^c_{b^\dagger_k o_2} \langle \hat{o}_1 \rangle^c \\
&- \langle \hat{o}_2 \rangle^c \langle \hat{o}_1 \rangle^c \langle \hat{b}^\dagger_k \rangle^c
\end{aligned}
\tag{C20b}
$$

employing which, Eq. (C18) becomes:

$$
\begin{aligned}
d\langle \hat{o}_1 \hat{o}_2 \rangle^c = \mathrm{tr}\{\mathcal{L}\hat{\rho}^c \hat{o}_1 \hat{o}_2\} \, dt + \\
\sqrt{\frac{\gamma_k}{2}} \left(C^c_{o_1 o_2 b_k} + C^c_{o_1 b_k} \langle \hat{o}_2 \rangle^c + C^c_{o_2 b_k} \langle \hat{o}_1 \rangle^c \right) dW_{\mathcal{I}_k}(t) \\
+ \sqrt{\frac{\gamma_k}{2}} \left(C^c_{b^\dagger_k o_1 o_2} + C^c_{b^\dagger_k o_1} \langle \hat{o}_2 \rangle^c + C^c_{b^\dagger_k o_2} \langle \hat{o}_1 \rangle^c \right) dW_{\mathcal{I}_k}(t)
\end{aligned}
\tag{C21}
$$

where terms involving only first-order moments cancel. Note that the above includes terms of up to $O(dt)$, as $dW_{\mathcal{I}_k}(t)$ is formally $O(dt^{1/2})$.

Next, we can calculate the second and third terms in Eq. (C18), using Eq. (C16). These take the simple forms:

$$
(d\langle \hat{o}_1 \rangle^c)\langle \hat{o}_2 \rangle^c = \langle \hat{o}_2 \rangle^c \mathrm{tr}\{\mathcal{L}\hat{\rho}^c \hat{o}_1\} \, dt + \sqrt{\frac{\gamma_k}{2}} \left(C^c_{o_1 b_k} \langle \hat{o}_2 \rangle^c + C^c_{b^\dagger_k o_1} \langle \hat{o}_2 \rangle^c \right) dW_{\mathcal{I}_k}(t)
\tag{C22a}
$$

$$
(d\langle \hat{o}_2 \rangle^c)\langle \hat{o}_1 \rangle^c = \langle \hat{o}_1 \rangle^c \mathrm{tr}\{\mathcal{L}\hat{\rho}^c \hat{o}_2\} \, dt + \sqrt{\frac{\gamma_k}{2}} \left(C^c_{o_2 b_k} \langle \hat{o}_1 \rangle^c + C^c_{b^\dagger_k o_2} \langle \hat{o}_1 \rangle^c \right) dW_{\mathcal{I}_k}(t)
\tag{C22b}
$$

Both equations above once again contain terms up to $O(dt)$.

Finally, we can write down the term arising from Ito's lemma. The requiring differential is that given by Eq. (C16). However, we need only retain the lowest $O(dt)$ term here, which is given by:

$$
(d\langle \hat{o}_1 \rangle^c)(d\langle \hat{o}_2 \rangle^c) = \frac{\gamma_k}{2} \left(C^c_{o_1 b_k} + C^c_{b^\dagger_k o_1} \right) \left(C^c_{o_2 b_k} + C^c_{b^\dagger_k o_2} \right) \, dt + O(dt^{3/2})
\tag{C23}
$$

where we have used $(dW_{\mathcal{I}_k})^2 = dt$.

Finally, we can write down Eq. (C18) by combining Eqs. (C21), (C22a), (C22b), and (C23). This finally yields:

$$
\begin{aligned}
dC^c_{o_1 o_2} = \mathrm{tr}\{\mathcal{L}\hat{\rho}^c \hat{o}_1 \hat{o}_2 - \langle \hat{o}_2 \rangle^c \mathcal{L}\hat{\rho}^c \hat{o}_1 - \langle \hat{o}_1 \rangle^c \mathcal{L}\hat{\rho}^c \hat{o}_2\} dt \\
- \frac{\gamma_k}{2} \left(C^c_{o_1 b_k} + C^c_{b^\dagger_k o_1} \right) \left(C^c_{o_2 b_k} + C^c_{b^\dagger_k o_2} \right) dt \\
+ \sqrt{\frac{\gamma_k}{2}} \left(C^c_{o_1 o_2 b_k} + C^c_{b^\dagger_k o_1 o_2} \right) dW_{\mathcal{I}_k}(t)
\end{aligned}
\tag{C24}
$$

The first term on the right hand side describes deterministic dynamics governed by $\mathcal{L}$. The second term is due to the measurement, but does not depend on the stochastic Wiener increment $dW_{\mathcal{I}_k}(t)$. The last term is a stochastic contribution; however note that it is related only to third-order cumulants. Within the ansätz used in this paper, this explicitly stochastic term vanishes.

3. Numerical tests: unconditional dynamics and comparison to exact results

We will use the single coherently-driven QNP mode modeled as a Kerr oscillator to benchmark the TEOMs approach. Using the results of Sec. C 1 of the SI, and following truncation with $n_{\mathrm{trunc}} = 2$, the single Kerr oscillator is defined by the first order cumulant $C_b \equiv \langle \hat{b} \rangle$ governed by the equation,

$$
\langle \dot{\hat{b}} \rangle = \left(i\Delta_1 - \frac{\gamma}{2} \right) \langle \hat{b} \rangle + i\Lambda_1 \langle \hat{b}^\dagger \rangle \langle \hat{b} \rangle \langle \hat{b} \rangle - i\eta_b + i\Lambda_1 \left(C_{bb} \langle \hat{b}^\dagger \rangle + 2 C_{b^\dagger b} \langle \hat{b} \rangle \right)
\tag{C25}
$$

as well as the second-order cumulants $C_{bb} = \langle \hat{b}\hat{b} \rangle - \langle \hat{b} \rangle^2$, $C_{b^\dagger b} = \langle \hat{b}^\dagger \hat{b} \rangle - \langle \hat{b}^\dagger \rangle \langle \hat{b} \rangle$, defined by

$$
\dot{C}_{b^\dagger b} = -\gamma C_{b^\dagger b} - i\Lambda_1 (C_{bb} \langle \hat{b}^\dagger \rangle^2 - C^*_{bb} \langle \hat{b} \rangle^2),
\tag{C26a}
$$

$$
\dot{C}_{bb} = \left(i2\Delta_1 - \gamma + i\Lambda_1(1 + 4|\langle \hat{b} \rangle|^2 + 6C_{b^\dagger b}) \right) C_{bb} + i\Lambda_1 \langle \hat{b} \rangle^2 (1 + 2C_{b^\dagger b}).
\tag{C26b}
$$

Eqs. (C25)-(C26b) together comprise the TEOMs for the single Kerr oscillator with $n_{\text{trunc}} = 2$.

It is reasonable to assume that such a truncation must be valid only in a certain subset of the full parameter space defining the system. To identify the general constraints on such a truncation, we make extensive use of the exact quantum steady-state solution of the coherently-driven Kerr oscillator, obtained via the complex-$\mathcal{P}$ representation [91]; details can be found in SI Sec. J. This provides access to exact steady-state cumulants of arbitrary order, which help us benchmark the accuracy of the TEOM approach.

One such regime should be furnished by the *classical limit* of the single-mode Kerr QNP, where the QNP is always in a coherent state and thus described entirely by its first-order cumulants. To identify this classical limit, and thus explore deviations from it due to quantum effects, it proves convenient to work with specific scaled quantities: after introducing dimensionless time $t' = \gamma t$ and energy scales $(\Delta_1', \Lambda_1', \eta_b') = (\Delta_1, \Lambda_1, \eta_b)/\gamma$, we scale $\langle \hat{b} \rangle' \to \sqrt{\Lambda_1'} \langle \hat{b} \rangle$, and impose our truncated ansatz, following which Eq. (C25) becomes

$$\frac{d}{dt'}\langle \hat{b} \rangle' = \left(i\Delta_1' - \frac{1}{2} \right) \langle \hat{b} \rangle' + i\langle \hat{b}^\dagger \rangle' \langle \hat{b} \rangle' \langle \hat{b} \rangle' - i\mathcal{E} + i\Lambda_1' \left(C_{bb} \langle \hat{b}^\dagger \rangle' + 2C_{b^\dagger b} \langle \hat{b} \rangle' \right). \tag{C27}$$

The first line in Eq. (C27) makes no reference to second-order cumulants of the QNP mode state; it thus describes dynamics when the QNP is in a coherent state (for which such cumulants vanish), corresponding to the classical limit. Dynamics in this limit depend on the drive and the nonlinearity via a single dimensionless parameter [57, 92],

$$\mathcal{E} = \eta' \sqrt{\Lambda'} = \frac{\eta_b}{\gamma} \sqrt{\frac{\Lambda_1}{\gamma}}. \tag{C28}$$

The second line of Eq. (C27) does however depend on cumulants; these terms describe quantum dynamics that lead to a deviation from purely coherent state evolution. In these scaled units, the TEOMs for the second-order cumulants obtained from Eq. (C26a), (C26b) take the form

$$\frac{d}{dt'}C_{b^\dagger b} = -C_{b^\dagger b} - i(C_{bb}(\langle \hat{b}^\dagger \rangle')^2 - C_{bb}^*(\langle \hat{b} \rangle')^2), \tag{C29a}$$

$$\frac{d}{dt'}C_{bb} = \left(i2\Delta_1' - 1 + i(\Lambda_1' + 4|\langle \hat{b} \rangle'|^2 + 6\Lambda_1' C_{b^\dagger b}) \right) C_{bb} + i(\langle \hat{b} \rangle')^2 (1 + 2C_{b^\dagger b}). \tag{C29b}$$

Consider now the transformation where $\Lambda_1' \to 0$, while $\mathcal{E}$ is held fixed (by increasing the drive strength η_b' simultaneously). Terms in the first line of Eq. (C27) describing classical dynamics remain unchanged, while those in the second line describing quantum deviations become smaller, *provided* second-order cumulants do not grow with Λ_1'. From Eqs. (C29a), (C29b), it is clear that $C_{b^\dagger b}$ is independent of Λ_1', while C_{bb} has a dependence that is negligible as $\Lambda_1' \to 0$ for $|\langle \hat{b} \rangle'| \neq 0$. Hence by keeping $\mathcal{E}$ fixed while decreasing the nonlinearity strength, the influence of second-order cumulants on dynamics of the QNP mode amplitude is suppressed, describing the classical limit of QNP operation. Physically, this transformation leads to an increase in the unscaled drive η' and QNP mode amplitude $|\langle \hat{b} \rangle|$, and hence occupation number, decreasing the relative impact of quantum fluctuations in agreement with conventional notions of classicality. Conversely, increasing nonlinearity strength for fixed $\mathcal{E}$ should enhance the impact of quantum fluctuations, marked by a systematic increase in second-order cumulants, and taking the Kerr QNP gradually towards a quantum regime of operation.

To explore this transition, we start with the classical limit of the QNP as determined by terms in the first line of Eq. (C27), for which the steady-state phase diagram can be easily computed (see SI [47]). This phase diagram, shown in $(\Delta_1, \mathcal{E})$-space in the center panel of Fig. 12(a), is the well-known result for a coherently-driven classical Kerr oscillator [57, 92]: the orange region depicts the classical bistability, which emerges for sufficiently negative detuning given our choice of the sign of the Kerr nonlinearity.

This bistability is a feature specific to nonlinear systems; in its vicinity, we can expect an enhanced nonlinear response of the QNP observables to QNP inputs. This is explicitly observed in the scaled field amplitude $|\langle \hat{b} \rangle'|$, plotted in Fig. 12(b) as a function of $\mathcal{E}$ for fixed $\Delta/\gamma_1 = -1.0$ (that is, across the indicated vertical dashed line of the classical phase diagram). The solid blue curve is the exact complex-P result for $\Lambda/\gamma_1 = 0.005$, while blue dots are obtained by evolving the TEOMs, Eqs. (C25), (C26a), (C26b), to their steady state numerically. For small $\mathcal{E}$, the response follows that of a linear oscillator (dashed black line), but becomes nonlinear with increasing $\mathcal{E}$. Also shown in dashed orange is the response of the classical nonlinear oscillator (explicitly dropping second-order cumulants), which does not depend on the actual nonlinearity strength but only on $\mathcal{E}$.

The influence of nonlinearity is similarly evident when plotting $|\langle \hat{b}' \rangle|$ as a function of Δ, for fixed $\mathcal{E} = 0.385$ in Fig. 12(c) (across the horizontal dashed line of the classical phase diagram). The black dashed curve indicates the standard Lorentzian response of a linear oscillator, centered around the bare frequency ω_b. The response of the
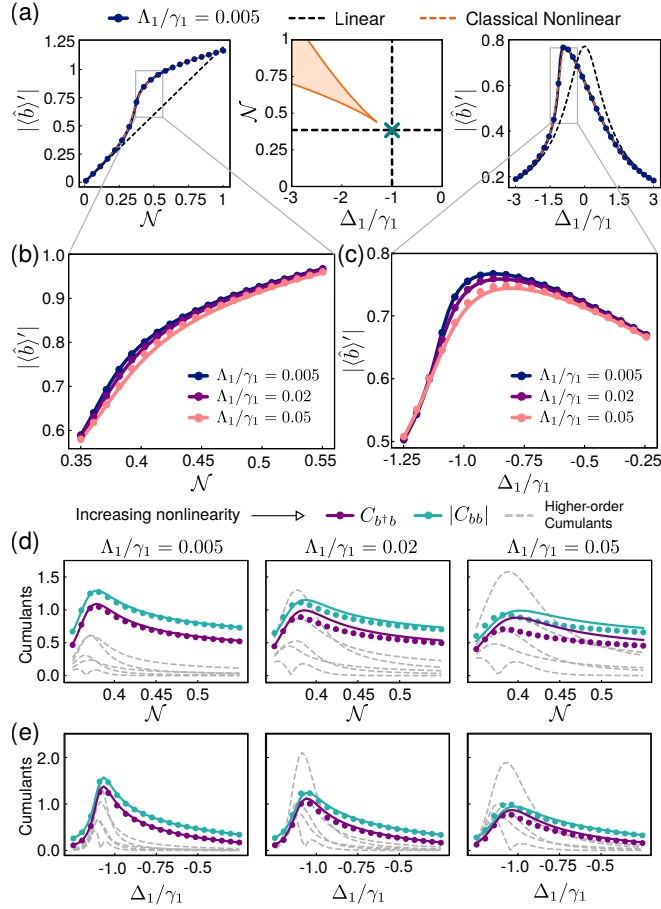
Figure 12. Classical and quantum dynamics of a single Kerr QNP mode. In this plot, colored solid curves are exact complex-$\mathcal{P}$ results, colored dots are TEOMs simulations in the long-time limit, and dashed orange curves are classical results. Other curves are specified when relevant. (a) Center panel: *Classical* phase diagram of the single coherently-driven Kerr QNP mode in $(\Delta_1, \mathcal{E})$ space. Left panel: QNP amplitude as a function of $\mathcal{E}$ for fixed detuning $\Delta_1/\gamma_1 = -1.0$ (across indicated vertical cross-section of the phase diagram). Note that the total damping rate $\gamma = \gamma_1 + \Gamma_1$, and $\Gamma_1 = \gamma_1/2$. Right panel: QNP amplitude as a function of detuning Δ_1 for fixed $\mathcal{E} = 0.385$ (across indicated horizontal cross-section of the phase diagram). Dashed black lines in both plots indicate the response for a linear oscillator. (b) Steady-state scaled amplitude $|\langle \hat{b}' \rangle|$ for nonlinearity strengths $\Lambda_1/\gamma_1 \in [0.005, 0.02, 0.05]$, as a function of $\mathcal{E}$ for fixed $\Delta_1/\gamma_1 = -1.0$ (region within the gray box in (a), left panel). (c) Same as (b) but as a function of detuning Δ_1 for fixed $\mathcal{E} = 0.385$ (region within the gray box in (a), right panel). (d) Steady-state second-order cumulants ($C_{b^\dagger b}$, purple, and $|C_{bb}|$, green) and higher (third, fourth) order cumulants (dashed gray) across the cross-section in (b); nonlinearity increases from left to right. (e) Same as (d) but now for the cross-section in (c).

nonlinear QNP mode is clearly that of a Lorentzian deformed towards lower frequencies due to the Kerr-induced frequency shift.

In both results, we find excellent agreement between the exact quantum results, the TEOMs, *and* classical nonlinear dynamics. This is indicative of operating parameters where not only is a truncated description using first and second-order cumulants sufficient, but the influence of second-order cumulants on first-order cumulants is negligible as well. Consequently, we expect the QNP mode here to function as a classical reservoir mode, still allowing for nonlinear processing of information encoded in the drive amplitude, phase, and/or frequency.

If we now keep $\mathcal{E}$ fixed but increase the nonlinearity strength Λ, the QNP can be moved to a quantum regime of operation. We plot the QNP response as a function of $\mathcal{E}$ for $\Delta/\gamma_1 = -1.0$ in Fig. 12(b), and as a function of Δ for $\mathcal{E} = 0.385$ in Fig. 12(c), for three different nonlinearity strengths $\Lambda/\gamma_1 \in [0.005, 0.02, 0.05]$. In both cases, we see that with increasing nonlinearity, the response deviates from that of the classical nonlinear oscillator (which recall is essentially equivalent to the quantum result for weak $\Lambda/\gamma_1 = 0.005$). From Eq. (C27), it is clear that this deviation is due to the coupling of first-order moments to second-order cumulants; this is precisely the effect accounted for by the TEOMs, which thus agree with the full quantum result.

Also shown in Figs. 12(d) and (e) are steady-state cumulants plotted across the same cross-sections as Figs. 12(b)

and (c) respectively, with panels from left to right indicating stronger nonlinearities. Second-order cumulants are plotted in green and purple (lines are exact complex-$\mathcal{P}$ results, dots are TEOMs); they clearly display non-monotonic behaviour as a function of $\mathcal{E}$ and Δ, acquiring large magnitudes for specific operating parameters in the vicinity of the classical bistability. Physically, second-order cumulants are related to the maximum and minimum (dimensionless) quadrature variances of the QNP modes: $\frac{1}{2} + C_{b^\dagger b} \pm |C_{bb}|$ respectively, and their magnitudes are thus measures of amplification and squeezing of quantum fluctuations due to the underlying Kerr nonlinearity.

Third and fourth order cumulants are shown in the same plots in dashed gray (exact complex-$\mathcal{P}$ only). There are two features of note. Firstly, the overall magnitude of these higher-order cumulants relative to second-order cumulants appears to increase for stronger nonlinearities. This is an indicator of the emergence of non-Gaussian features with increasing Λ, as expected, and generally leads to greater deviation between TEOMs and exact quantum results for first and second-order cumulants. Secondly, even for a fixed nonlinearity strength, the magnitude of cumulants depends on operating parameters ($\mathcal{E}$ and Δ) that lead to operation in the vicinity of the classical bistability. Hence, while results obtained from TEOMs for such operating parameters can deviate from the exact quantum solution, away from these regions the agreement improves, even for strong nonlinearities.

4. Numerical tests: conditional dynamics

We have so far used steady-state, unconditional quantities to analyze the nonlinear processing capabilities of a single Kerr-based QNP mode, and identified parameter regimes where an approach based on truncated EOMs is valid. However, to allow for processing of time-dependent information by the QNP, and to analyze the output from the QNP obtained as *individual* measurement records, an approach is needed that is able to accurately capture the *conditional, dynamical* evolution of the QNP mode. This is described by the SME

$$d\hat{\rho}^c = \mathcal{L}\hat{\rho}^c \, dt + \mathcal{S}[\sqrt{\gamma_1}\hat{b}]\hat{\rho}^c \tag{C30}$$

where $\mathcal{S}[\sqrt{\gamma_1}\hat{b}]$ is defined as,

$$\mathcal{S}[\sqrt{\gamma_1}\hat{b}_k]\hat{\rho}^c = \sqrt{\frac{\gamma_1}{2}}\left(\hat{b}_k\hat{\rho}^c + \hat{\rho}^c\hat{b}_k^\dagger - \langle\hat{b}_k + \hat{b}_k^\dagger\rangle^c\right)dW_{\mathcal{I}_k}(t)$$

$$+ \sqrt{\frac{\gamma_1}{2}}\left(-i\hat{b}_k\hat{\rho}^c + i\hat{\rho}^c\hat{b}_k^\dagger - \langle -i\hat{b}_k + i\hat{b}_k^\dagger\rangle^c\right)dW_{\mathcal{Q}_k}(t) \tag{C31}$$

and our current system assumes the $K = 1$ case. Note the slight differences to Eq. (A4) of the main text, which are only for purposes of exposition: the dissipator in Eq. (A4) is absorbed into $\mathcal{L}$ here, to aid the discussion in the previous subsection.

Expectation values with respect to the conditional quantum state under measurement $\hat{\rho}^c$ are now conditional, which we indicate by the superscript c. For example, the variation $d\langle\hat{b}\rangle^c$ in a time dt is now given by $\text{tr}\{(\mathcal{L} \, dt + \mathcal{S}[\hat{b}])\hat{\rho}^c\hat{b}\}$, which takes the form (as derived in Sec. C 2 of the SI)

$$d\langle\hat{b}\rangle^c = \text{tr}\{(\mathcal{L}\hat{\rho}^c\hat{b}\} \, dt +$$

$$\sqrt{\frac{\gamma_1}{2}}\left(C_{b^\dagger b}^c + C_{bb}^c\right)dW^X(t) + i\sqrt{\frac{\gamma_1}{2}}\left(C_{b^\dagger b}^c - C_{bb}^c\right)dW^P(t). \tag{C32}$$

The first line $\propto \mathcal{L}$ simply includes terms from Eq. (C25), with expectation values replaced by their conditional counterparts. Terms in the second line are due to the measurement superoperator, and render the equation of motion for $\langle\hat{b}\rangle^c$ *stochastic*, conditioned on the trajectory-specific realizations of $dW^{X,P}(t)$. Dropping cumulants higher than second order in Eq. (C32), arising in the first line, yields the *stochastic* truncated EOM (STEOM) for $\langle\hat{b}\rangle^c$.

We can similarly obtain truncated equations of motion for the conditional cumulants C_{bb}^c, $C_{b^\dagger b}^c$ under heterodyne measurement (as derived in Sec. C 2 of the SI):

$$dC_{b^\dagger b}^c = \left[\text{tr}\{\mathcal{L}\hat{\rho}^c\hat{b}^\dagger\hat{b}\} - \langle\hat{b}^\dagger\rangle^c\text{tr}\{\mathcal{L}\hat{\rho}^c\hat{b}\} - \langle\hat{b}\rangle^c\text{tr}\{\mathcal{L}\hat{\rho}^c\hat{b}^\dagger\}\right]dt$$

$$- \gamma_1\left[(C_{b^\dagger b}^c)^2 + C_{bb}^c(C_{bb}^c)^*\right] \, dt, \tag{C33a}$$

$$dC_{bb}^c = \left[\text{tr}\{\mathcal{L}\hat{\rho}^c\hat{b}\hat{b}\} - 2\langle\hat{b}\rangle^c\text{tr}\{\mathcal{L}\hat{\rho}^c\hat{b}\}\right]dt$$

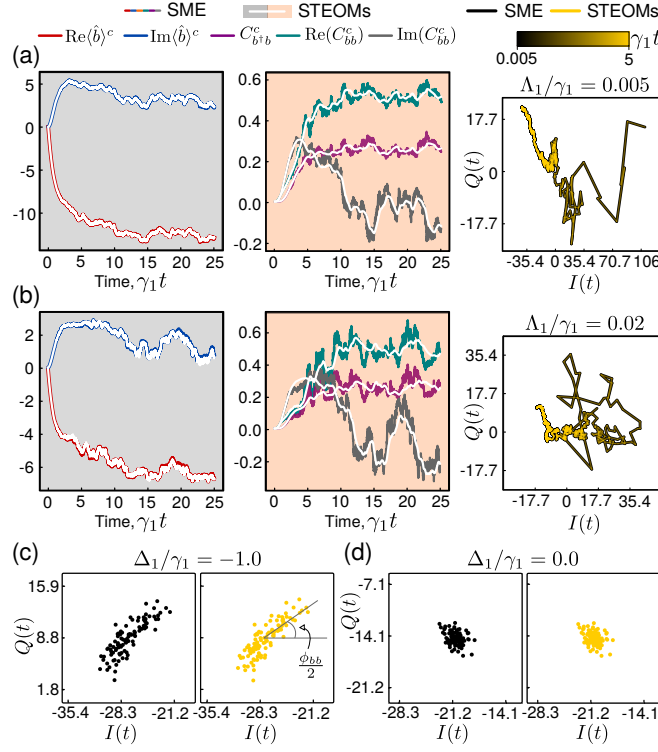$$- 2\gamma_1 C_{bb}^c C_{b^\dagger b}^c \, dt. \tag{C33b}$$

Figure 13. Benchmarking STEOMs against full SME simulations. Simulated parameters are as marked by the teal 'X' in the phase diagram of Fig. 12(a): $\mathcal{E} = 0.385$, $\Delta_1/\gamma_1 = -1.0$, for (a) $\Lambda_1/\gamma_1 = 0.005$ and (b) $\Lambda_1/\gamma_1 = 0.02$. Note that the total damping rate $\gamma = \gamma_1 + \Gamma_1$, and $\Gamma_1 = \gamma_1/2$. Left panel shows real and imaginary parts of the conditional expectation value $\langle \hat{b} \rangle$, and center panel shows conditional second-order cumulants $C_{bb}, C_{b^\dagger b}$; colored (white) curves are SME (STEOMs) simulations. Right panel shows measured quadratures $I(t)$, $Q(t)$: the SME result is plotted in solid black, while the STEOMs result is plotted in yellow, with a lighter shade indicating earlier times. (c) Measured quadratures $\{I(t), Q(t)\}$ at $\gamma t = 20.0$ obtained using SME (left panel) and STEOMs (right panel), for $\mathcal{E} = 0.385$, $\Lambda_1/\gamma_1 = 0.02$, and $\Delta_1/\gamma_1 = -1.0$. (d) Same as (c) but for $\Delta_1/\gamma_1 = 0.0$.

Again, terms $\propto \mathcal{L}$ are as found in Eqs. (C26a), (C26b) post-truncation, with expectation values replaced by their conditional counterparts. The second line of each equation describes the evolution due to measurement, which at first glance appears deterministic: note that Wiener increments $dW^{X,P}(t)$ make no appearance. For *linear* quantum systems under continuous weak measurement, this is in fact the case: second-order cumulants form a closed set described by the above equations, and no stochastic terms arise [81, 93–95]. However, for *nonlinear* quantum systems of interest here, second-order cumulants can couple to the *stochastic* first-order moments (here, via terms $\propto \mathcal{L}$), rendering the conditional evolution of second-order cumulants generally stochastic as well.

Eqs. (C32), (C33a), (C33b) define the STEOMs for the single Kerr QNP mode under continuous heterodyne monitoring. To assess their validity, we compare their simulation results against integration of the SME, Eq. (C30). For concreteness, we consider the operating point marked by the teal 'X' in the phase diagram of Fig. 12(a): $\mathcal{E} = 0.385$, $\Delta/\gamma = -1.0$. Note that the *same noise realizations* $dW^{X,P}(t)$ must always be used for both simulations to ensure the compared trajectories are conditioned on the same measurement record. Results are included in Fig. 13(a), (b) for $\Lambda/\gamma \in [0.005, 0.02]$ respectively, showing first-order moments (left panel) and second-order cumulants (center panel), with STEOMs results in white, and SME results obtained using QuTiP [96] in color. Both methods agree very well, especially for first-order cumulants. In particular, note the stochastic nature of second-order cumulants, which is a direct signature of the nonlinearity of the quantum system under study, and is further enhanced at strong nonlinearities. However, for the same temporal resolution, SME simulations with a Hilbert space cutoff of 625 photons (around 3 times larger than the maximum occupation number, a typical conservative value) demand several orders of magnitude more simulation time than the STEOMs.

However, in a real experiment, such information about the QNP state must be extracted from measurement records. For the case of heterodyne measurement of the single Kerr oscillator, the obtained records $\mathcal{I}(t) \equiv \mathcal{I}_1(t)$, $\mathcal{Q}(t) \equiv \mathcal{Q}_1(t)$ are defined in Eqs. (4) of the main text (here we take $\bar{n}_{\text{cl}} = 0$, ignoring the role of classical readout noise). These records are typically processed to reduce noise, most commonly via temporal filtering; the processed records define
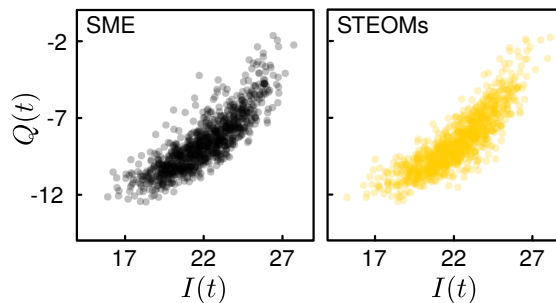
Figure 14. Non-Gaussian signatures in single Kerr oscillator heterodyne measurements. Measured quadratures $\{I(t), Q(t)\}$ at $\gamma_1 t = 60.0$ obtained using SME (left panel) and STEOMs (right panel), for $\mathcal{E} = 0.369$, $\Lambda_1/\gamma_1 = 0.0295$, and $\Delta_1/\gamma_1 = -1.0$. Note that the total damping rate $\gamma = \gamma_1 + \Gamma_1$, and $\Gamma_1 = \gamma_1/2$.

measured quadratures as in the main text,

$$I(t) = \frac{1}{\sqrt{2\mathcal{T}}} \int_{t-\mathcal{T}}^{t} d\tau \ \mathcal{I}(\tau); \ Q(t) = \frac{1}{\sqrt{2\mathcal{T}}} \int_{t-\mathcal{T}}^{t} d\tau \ \mathcal{Q}(\tau). \tag{C34}$$

which here describe processing of single-shot readout records. In this section, we take $t - \mathcal{T} = 0$ for filtering over the entire collection window. Measured quadratures obtained using a single set of measurement records are shown in the measured phase space given by $(I(t), Q(t))$ in the right panel of Fig. 13(a), (b) $\Lambda/\gamma \in [0.005, 0.02]$ as before; SME results are shown in solid black, with the STEOMs results plotted on top, using a colorscale going from black to yellow indicating increasing time. Both show excellent agreement; this is unsurprising since $I(t), Q(t)$ yield information about the underlying conditional dynamics of $\langle \hat{b} \rangle^c$, which also agree very well between the two methods.

Note that the measured quadratures are themselves stochastic quantities, whose statistics are correlated with the underlying measured system state. This can be seen via the measured quadrature distributions in Fig. 13(c), obtained by simulating several measurement records (here 100 in total) and plotting $\{I(t), Q(t)\}$ at $\gamma t = 20.0$, for $\Delta_1/\gamma_1 = -1.0$ as before. At this operating point near the classical bistable region, the Kerr nonlinearity amplifies the magnitude of cumulants $C_{bb} \equiv |C_{bb}| e^{i\phi_{bb}}$ [see Fig. 12 (d)] leading to squeezing of the internal QNP field along the axis determined by $\phi_{bb}/2$ (and amplification along the orthogonal quadrature). This internal QNP squeezing manifests as squeezing of the measured quadrature distribution; the squeezing axis is unchanged since the temporal filter defining measured quadratures via Eq. (C34) is quadrature-agnostic, and thus only reduces the overall noise power while preserving its relative strength amongst measured quadratures. On the other hand, for $\Delta_1/\gamma_1 = 0.0$, further away from the bistable region, C_{bb} is much smaller in magnitude and the measured quadrature distribution in Fig. 13(d) exhibits no squeezing. Excellent agreement between SME (left) and STEOMs (right) is observed for both cases.

5. Non-Gaussian signatures captured by STEOM simulations

The (S)TEOMs approach to nonlinear quantum systems in this work uses $n_{\text{trunc}} = 2$, so that only cumulants of up to second-order are retained. Gaussian states have at most second-order cumulants as well. However, the TEOMs approach is distinct from other methods of analyzing nonlinear quantum systems where fluctuations are treated within a Gaussian approximation, such as standard linearization. There are two important indicators of this difference. First is the deviation of first-order cumulants from their classical values, observed in Fig. 12, which would not happen in standard linearization schemes. The second indicator appears in stochastic dynamics of second-order cumulants, as seen in Fig. 13, which would be perfectly deterministic for a quantum system whose state is defined *exactly* by a Gaussian description.

These features indicate the ability of (S)TEOMs to simulate dynamics beyond those of just Gaussian states. In fact, we find that the STEOMs can exhibit non-Gaussian characteristics in measured heterodyne outputs that are also captured by SME simulations. An example is depicted in Fig. 14, where we show measured distributions obtained using the SME and STEOMs for a Kerr oscillator operated in the vicinity of the classical Kerr bistability (for parameters, see caption). Both measured distributions clearly display 'crescent'-shaped deformations typical of non-Gaussian characteristics observed in transient dynamics of Kerr-type nonlinear oscillators [12].

6. Validity and scalability of (S)TEOMs

The key observations from our analysis of Kerr QNP operating regimes and the generation of higher-order cumulants (summarized in Fig. 12) are twofold. Primarily, the magnitude of nonzero higher-order cumulants increases with the strength of the Kerr nonlinearity relative to the QNP mode damping rate, Λ/γ (where γ is the total QNP mode loss rate), which coincides with the QNP under coherent driving transitioning from classical to quantum regimes. However, operating regimes in the vicinity of the classical bistability also lead to an enhancement in higher-order cumulants, regardless of nonlinearity strength. To be confident of the validity of the truncated cumulants approach, we therefore analyze QNPs with nonlinearity strengths $\Lambda/\gamma \leq 0.02$, and operate *near* but not within the classically-bistable region.

This truncated cumulants approach provides a highly efficient mathematical description when applied to multimode quantum systems. For a measurement chain comprising an M-mode quantum system coupled to a K-mode QNP, the composite system of $R = M + K$ total quantum modes is described by $2R^2 + 3R$ unknowns using (S)TEOMs, scaling quadratically with R instead of the exponential growth in required Hilbert space size for full (S)ME simulations. This description also places no constraints on modal occupation numbers, enabling our exploration of a well-defined classical limit of the coherently-driven Kerr QNP. Our (S)TEOMs approach is built to be a scalable theoretical framework via an efficient computer-algebra implementation, which allows the calculation of equations of motion of retained cumulants up to order n_{trunc} (here, second-order) for nonlinear quantum systems comprising arbitrary numbers of bosonic modes under continuous measurement. Supplementary benchmarking simulations for systems with more than one mode are included in Sec. I of the SI. In the main text, we employ (S)TEOMs to simulate dynamics of measurement chains comprising measured quantum systems coupled to multimode QNPs, while operating within the aforementioned constraints.

Appendix D: Quantum properties of the quantum system (QS) states to be distinguished

In this section we analyze the quantum system (QS) defined by Eq. (12) of the main text, whose states we wish to distinguish in classification tasks. The TEOMs that describe the dynamics of the QS are simply given by the QS sector of Eqs. (7a), (8), and are *exact* for $n_{\text{trunc}} = 2$, owing to the QS linearity under coherent driving, and its non-reciprocal coupling to the QNP. For convenience, we reproduce the TEOMs below,

$$\frac{d}{dt}\langle\hat{\boldsymbol{a}}\rangle = \mathbf{L}_a\langle\hat{\boldsymbol{a}}\rangle + \vec{\eta}, \tag{D1a}$$

$$\frac{d}{dt}\mathbf{C}_a = \mathbf{L}_a\mathbf{C}_a + \mathbf{C}_a\mathbf{L}_a^T + \mathbf{D}_a, \tag{D1b}$$

where recall that $\hat{\boldsymbol{a}} \equiv (\hat{a}_1, \hat{a}_1^\dagger, \ldots, \hat{a}_M, \hat{a}_M^\dagger)^T$ and $\mathbf{C}_a = \langle:\hat{\boldsymbol{a}}\hat{\boldsymbol{a}}^T:\rangle - \langle\hat{\boldsymbol{a}}\rangle\langle\hat{\boldsymbol{a}}^T\rangle$. Crucially, we emphasize that the above TEOMs describe the unconditional dynamics of QS state variables *irrespective* of the QNP's precise nonlinearity, due to the non-reciprocal coupling which ensures the QS drives the QNP, but not vice-versa.

In what follows, we specialize to the $M = 2$ mode system and its specified parameters from Sec. II C, namely $G_2 = 0$, $\phi_1 = \frac{\pi}{2}$, and $\Gamma_m + \kappa_m \equiv \kappa \; \forall \; m$. Then, the drift matrix $\mathbf{L}_a$ and diffusion matrix $\mathbf{D}_a$ defining Eqs. (D1a), (D1b) take the respective forms:

$$\mathbf{L}_a = \begin{pmatrix} -\frac{\kappa}{2} & -iG_1 & 0 & -ie^{i\phi_{12}}G_{12} \\ -iG_1 & -\frac{\kappa}{2} & ie^{-i\phi_{12}}G_{12} & 0 \\ 0 & -ie^{i\phi_{12}}G_{12} & -\frac{\kappa}{2} & 0 \\ ie^{-i\phi_{12}}G_{12} & 0 & 0 & -\frac{\kappa}{2} \end{pmatrix}, \tag{D2a}$$

$$\mathbf{D}_a = \begin{pmatrix} -iG_1 & 0 & -ie^{i\phi_{12}}G_{12} & 0 \\ 0 & iG_1 & 0 & ie^{-i\phi_{12}}G_{12} \\ -ie^{i\phi_{12}}G_{12} & 0 & 0 & 0 \\ 0 & ie^{-i\phi_{12}}G_{12} & 0 & 0 \end{pmatrix}, \tag{D2b}$$

while the coherent drive vector is $\vec{\eta} = (\eta_1, \eta_1, \eta_2, \eta_2)^T$.

The linearity of Eqs. (D1a), (D1b) renders them straightforward to solve. In particular, as we are interested in classification in the long-time limit, we simply evaluate the steady-state first-order cumulants describing the quantum system by setting $\frac{d}{dt}\langle\hat{\boldsymbol{a}}\rangle = 0$ in Eq. (D1a), obtaining

$$\langle\hat{\boldsymbol{a}}\rangle = -\mathbf{L}_a^{-1}\vec{\eta}. \tag{D3}$$

Similarly, second-order cumulants in the long-time limit can be obtained by setting $\frac{d}{dt}\mathbf{C}_a = 0$ in Eq. (D1b); the steady state value of $\mathbf{C}_a$ then satisfies the continuous Lyapunov equation,

$$\mathbf{L}_a\mathbf{C}_a + \mathbf{C}_a\mathbf{L}_a^T + \mathbf{D}_a = 0. \tag{D4}$$

In lieu of writing down the exact expressions for steady-state cumulants of the quantum system in complete generality – which can be unwieldy – we analyze the specific states indexed by l and summarized in Fig. 2 that we wish to distinguish in classification tasks. Starting with Task 1, recall that $\Gamma_2 = 0$, so that $\hat{a}_2$ is not read out, and is consequently also not coupled to the QNP. Hence the value of first-order cumulants defining its quantum states does not contribution to classification. The first-order cumulants for $\hat{a}_1$ are then given by:

$$\langle\hat{a}_1^{(1)}\rangle = -\frac{2\eta_1^{(1)}}{\kappa - 2G_1^{(1)}}, \;\; \langle\hat{a}_1^{(2)}\rangle = -\frac{2\eta_1^{(2)}}{\kappa^2 - 4(G_{12}^{(2)})^2} \tag{D5}$$

By suitable choices of $\eta_1^{(1,2)}$, an example of which is provided in Fig. 2(b), we can realize states for which $\langle\hat{a}_1^{(1)}\rangle = \langle\hat{a}_1^{(2)}\rangle = \mathcal{A}$. However, the realized states do differ in their second-order cumulants, which take the forms:

$$\mathbf{C}_a^{(1)} = \frac{1}{\kappa^2 - 4(G_1^{(1)})^2}\begin{pmatrix} \kappa G_1^{(1)} & 2(G_1^{(1)})^2 & 0 & 0 \\ 2(G_1^{(1)})^2 & \kappa G_1^{(1)} & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \;\; \mathbf{C}_a^{(2)} = \frac{1}{\kappa^2 - 4(G_{12}^{(2)})^2}\begin{pmatrix} 0 & 2(G_{12}^{(2)})^2 & -i\kappa G_{12}^{(2)} & 0 \\ 2(G_{12}^{(2)})^2 & 0 & 0 & i\kappa G_{12}^{(2)} \\ -i\kappa G_{12}^{(2)} & 0 & 0 & 2(G_{12}^{(2)})^2 \\ 0 & i\kappa G_{12}^{(2)} & 2(G_{12}^{(2)})^2 & 0 \end{pmatrix}$$

$$\tag{D6}$$

It is clear that the structure of covariance matrices is very distinct, in particular as seen in the top-left sector corresponding to mode $\hat{a}_1$, which is the only mode read out in Task 1.

For Task 2, we set $\Gamma_2 = \Gamma_1 = \kappa/2$ so that both system modes $\hat{a}_1$ and $\hat{a}_2$ are read out. In this case, the first-order cumulants are given by

$$\langle \hat{a}_1^{(3,4)} \rangle = -\frac{2(\kappa \eta_1^{(3,4)} - i2G_{12}^{(l)} \eta_2^{(3,4)} e^{i\phi_{12}^{(l)}})}{\kappa^2 - 4(G_{12}^{(l)})^2} \tag{D7a}$$

$$\langle \hat{a}_2^{(3,4)} \rangle = -\frac{2(\kappa \eta_2^{(3,4)} - i2G_{12}^{(l)} \eta_1^{(3,4)} e^{i\phi_{12}^{(l)}})}{\kappa^2 - 4(G_{12}^{(l)})^2} \tag{D7b}$$

Again, by appropriate choice of $\eta_1^{(3,4)}$, an example of which is provided in Fig. 2(c), we can realize states for which $\langle \hat{a}_{1,2}^{(3)} \rangle = \langle \hat{a}_{1,2}^{(4)} \rangle = \mathcal{A}$. The second-order cumulants are now given by

$$\mathbf{C}_a^{(3,4)} = \frac{1}{\kappa^2 - 4(G_{12}^{(3,4)})^2} \begin{pmatrix} 0 & 2(G_{12}^{(3,4)})^2 & -i\kappa G_{12}^{(3,4)} e^{i\phi_{12}^{(3,4)}} & 0 \\ 2(G_{12}^{(3,4)})^2 & 0 & 0 & i\kappa G_{12}^{(3,4)} e^{-i\phi_{12}^{(3,4)}} \\ -i\kappa G_{12}^{(3,4)} e^{i\phi_{12}^{(3,4)}} & 0 & 0 & 2(G_{12}^{(3,4)})^2 \\ 0 & i\kappa G_{12}^{(3,4)} e^{-i\phi_{12}^{(3,4)}} & 2(G_{12}^{(3,4)})^2 & 0 \end{pmatrix} \tag{D8}$$

Note that states $l = 3, 4$ differ only in their two-mode interaction phase ϕ_{12}. Furthermore, this appears only in the off-diagonal sector of the covariance matrix, and therefore is a property of the joint cumulants of the quantum system.

The above completes our characterization of the QS states we wish to classify in Tasks I and II considered in the main text. In a measurement chain where the signals from the QS are processed by a downstream QNP, the QS state determines the QNP state; this relationship is quantified in SI Sec. E.

Appendix E: Nonlinear van Kampen expansion of the Fokker-Planck equation describing the complete measurement chain

The (S)TEOMs, an example of which was derived in SI Sec. C, provide a useful approximate numerical scheme to simulate the dynamics of measurement chains with QNPs. However the close connection between quantum cumulants that underpin the (S)TEOMs framework and the positive-P representation (discussed in SI Sec. B) allows us to further develop an approximate *analytic* approach to describing the dynamics of QNPs under quantum input signals. Arriving at this analytic description is the subject of the present section.

1. Fokker-Planck equation in the Positive-P representation

Our approach will begin by considering an alternative representation of the general quantum state $\hat{\rho}$ of the complete measurement chain: the positive-P phase space representation [86]. The positive-P representation is obtained via a projection of $\hat{\rho}$ onto a non-diagonal basis of coherent state projectors:

$$\hat{\rho}(t) = \int d\vec{Z}\ \mathcal{P}(\vec{Z},t) \bigotimes_i \frac{|\alpha_i\rangle\langle\alpha_i^{\dagger *}|}{e^{\alpha_i\alpha_i^\dagger}} \bigotimes_j \frac{|\beta_j\rangle\langle\beta_j^{\dagger *}|}{e^{\beta_j\beta_j^\dagger}} \tag{E1}$$

The positive-P representation is defined in a phase space spanned by $\mathbb{C}$-number variables $\vec{Z} = (\alpha_1, \alpha_1^\dagger, \ldots, \alpha_M, \alpha_M^\dagger, \beta_1, \beta_1^\dagger, \ldots, \beta_K, \beta_K^\dagger)^T$, which are associated with the respective quantum operators $\hat{\boldsymbol{z}} = (\hat{a}_1, \hat{a}_1^\dagger, \ldots, \hat{a}_M, \hat{a}_M^\dagger, \hat{b}_1, \hat{b}_1^\dagger, \ldots, \hat{b}_K, \hat{b}_K^\dagger)^T$. The computational importance of this correspondence is that it enables *normal-ordered* expectation values with respect to the quantum state to be computed by instead evaluating averages of phase space variables with respect to the positive-P distribution:

$$\text{tr}\{(:\hat{z}_i \cdots \hat{z}_k:)\hat{\rho}(t)\} = \langle:\hat{z}_i \cdots \hat{z}_k:\rangle = \langle Z_i \cdots Z_k\rangle = \int d\vec{\zeta}\ (Z_i \cdots Z_k)\mathcal{P}(\vec{Z},t) \tag{E2}$$

Evaluating expectation values via averages computed against the positive-P representation can in many cases be extremely efficient, circumventing the Hilbert space restrictions that constrain evolution of $\hat{\rho}$.

However, using Eq. (E2) requires knowledge of the positive-P representation $\mathcal{P}(\vec{Z},t)$, which in turn requires a governing equation for $\mathcal{P}(\vec{Z},t)$ that could formally be solved for its time evolution and steady-state, the analog to the master equation for the quantum state $\hat{\rho}$. Using Eq. (E1), the master equation can be mapped to this desired equation, which for a large class of bosonic nonlinear systems takes the form of a Fokker-Planck equation (FPE). The general form of an FPE for the positive-P representation $\mathcal{P}$ is given by:

$$d_\tau \mathcal{P}(\vec{Z},\tau) = \left(-\frac{\partial}{\partial Z_i}A_i + \frac{1}{2}\frac{\partial^2}{\partial Z_i \partial Z_j}\mathbf{D}_{ij}\right)\mathcal{P}(\vec{Z},\tau) \tag{E3}$$

where we have for convenience also introduced dimensionless time and nonlinearity,

$$\tau = \gamma t,\ \bar{\Lambda}_k = \Lambda_k/\gamma \tag{E4}$$

where γ is the common monitored channel loss rate of all QNP modes, $\gamma_k = \gamma\ \forall\ k$. Then the dynamics of the FPE are entirely governed by the dimensionless drift vector $\vec{A}$ with elements A_i, and the dimensionless diffusion matrix $\mathbf{D}$ with elements $\mathbf{D}_{ij}$; repeated indices are summed over. Both $\vec{A}$ and $\mathbf{D}$ are completely defined by the underlying master equation governing $\hat{\rho}$, from the transformation of which the FPE is obtained.

For a measurement chain of comprising multiple, possibly nonlinear bosonic modes as described by Eq. (1), we can further write general forms for the drift vector and diffusion matrix:

$$\vec{A}(\vec{Z}) = \mathbf{A}^L\vec{Z} + \bar{\mathbf{\Lambda}}\vec{A}^N(\vec{Z}) + \vec{f}, \tag{E5a}$$

$$\mathbf{D}(\vec{Z}) = \mathbf{D}^L + \bar{\mathbf{\Lambda}}\mathbf{D}^N(\vec{Z}). \tag{E5b}$$

Here, the matrices $\mathbf{A}^L$ and $\mathbf{D}^L$ define the respective and sole contributions to the drift and diffusion terms of Eq. (E3) if all bosonic modes in the measurement chain were linear. The inhomogeneous terms $\vec{f}$ further includes the entire effect of coherent driving. Then, $\bar{\mathbf{\Lambda}}$ is a matrix describing the nonlinearity; here it is a diagonal (due to the local nature of the nonlinearities) matrix of the QNP mode Kerr nonlinearity strengths, and hence only vanishes if $\bar{\Lambda}_k = 0\ \forall\ k$.

The vector $\vec{A}^N$ and matrix $\mathbf{D}^N$ then respectively describe the nonlinear contributions to the drift and diffusion terms respectively; these have a more complex dependence on the phase space variables $\vec{Z}$ than their corresponding linear counterparts do, and will be the focus of our analysis here.

In what follows, we will assume $\bar{\Lambda}_k \equiv \bar{\Lambda} \ \forall \ k$, the case considered in the main text. This is not necessary for the remainder of the calculation, or for QNPs in general, but enables us to simplify our notation by foregoing a matrix representation of the nonlinearity. We thus let $\bar{\mathbf{\Lambda}} \to \bar{\Lambda}$ from here on.

We next introduce a displacement of the phase space variables $\vec{Z}$, around a quantum state of the measurement chain where the phase space variables take the expectation values $\langle \vec{Z}(t) \rangle$. Formally, we write the original phase space variables in the form

$$\vec{Z} = \bar{\Lambda}^{-\frac{1}{2}} (\langle \vec{Z}(\tau) \rangle + \bar{\Lambda}^q \vec{\zeta}) \tag{E6}$$

where the variables $\vec{\zeta}$ describe the quantum statistics of fluctuations around the quantum state we are expanding about. We can therefore define a positive-P distribution $\bar{\mathcal{P}}$ in the phase space of fluctuations. which is related to the original positive-P distribution via,

$$\bar{\mathcal{P}}(\vec{\zeta}, \tau) \equiv \bar{\Lambda}^{q-\frac{1}{2}} \cdot \mathcal{P}(\bar{\Lambda}^{-\frac{1}{2}} \langle \vec{Z}(\tau) \rangle + \bar{\Lambda}^{q-\frac{1}{2}} \vec{\zeta}, \tau) \tag{E7}$$

where the prefactor ensures the correct normalization of the positive-P distribution of fluctuations.

Our aim now is to convert Eq. (E3) to an equivalent FPE for in the phase space of *fluctuations* alone. To this end, we first make use of the chain rule as implied by Eq. (E6),

$$\frac{\partial}{\partial \langle \bar{Z}_j \rangle} = \frac{\partial \zeta_j}{\partial \langle \bar{Z}_j \rangle} \frac{\partial}{\partial \zeta_j} = \bar{\Lambda}^{-q} \frac{\partial}{\partial \zeta_j}, \tag{E8a}$$

$$\frac{\partial}{\partial Z_j} = \frac{\partial \zeta_j}{\partial Z_j} \frac{\partial}{\partial \zeta_j} = \bar{\Lambda}^{-(q-\frac{1}{2})} \frac{\partial}{\partial \zeta_j}. \tag{E8b}$$

Then, to determine the left hand side of Eq. (E3), we first compute the time derivative of Eq. (E7),

$$d_\tau \bar{\mathcal{P}} = \bar{\Lambda}^{q-\frac{1}{2}} \cdot \left(\frac{\partial \mathcal{P}}{\partial \langle \bar{Z}_i \rangle} \frac{\partial \langle \bar{Z}_i \rangle}{\partial \tau} + \frac{\partial \mathcal{P}}{\partial \tau} \right)$$
$$= \left(\bar{\Lambda}^{-q} \frac{\partial \langle \bar{Z}_i \rangle}{\partial \tau} \right) \frac{\partial \bar{\mathcal{P}}}{\partial \zeta_i} + \frac{\partial \bar{\mathcal{P}}}{\partial \tau} \tag{E9}$$

Now, the final term on the right hand side defines the partial derivative of $\bar{\mathcal{P}}$ with respect to time; using Eq. (E6), this can directly be related to the FPE for the original positive-P distribution, Eq. (E3) under the expansion of Eq. (E6). We first replace phase space derivatives in Eq. (E3) via the chain rule, Eq. (E8b), and then simplify to arrive at:

$$d_\tau \bar{\mathcal{P}}(\vec{Z}, \tau) = \left(\bar{\Lambda}^{-q} \frac{\partial \langle \bar{Z}_i \rangle}{\partial \tau} \frac{\partial}{\partial \zeta_i} - \bar{\Lambda}^{-(q-\frac{1}{2})} \frac{\partial}{\partial \zeta_i} A_i + \frac{1}{2} \bar{\Lambda}^{-2(q-\frac{1}{2})} \frac{\partial^2}{\partial \zeta_i \partial \zeta_j} \mathbf{D}_{ij} \right) \bar{\mathcal{P}}(\vec{Z}, \tau) \tag{E10}$$

The above takes the form of an FPE, now for the positive-P distribution of fluctuations, $\bar{\mathcal{P}}(\vec{Z}, \tau)$. However, an important complexity is hidden in the notation. Note that both A_i and $\mathbf{D}_{ij}$ depend on phase space variables, and therefore must also be evaluated under the expansion of Eq. (E6) to arrive at the final desired FPE.

2. Expanding the drift term

We first write the drift term following the expansion of Eq. (E6). For simplicity, we introduce the right-bar notation,

$$F \rbrack \equiv \left(F(\vec{Z}) \Big|_{\vec{Z} \to \bar{\Lambda}^{-\frac{1}{2}} \langle \vec{Z} \rangle} \right) \tag{E11}$$

to describe the evaluation of any quantity F (including vectors and matrices) of phase space variables $\vec{Z}$ around the expansion point defined by Eq. (E6).

In particular we will consider a Taylor expansion of the drift vector,

$$\bar{\Lambda}^{\frac{1}{2}} A_i = \bar{\Lambda}^{\frac{1}{2}} A_i \rbrack + \sum_j \bar{\Lambda}^{\frac{1}{2}} \frac{\partial A_i}{\partial Z_j} \rbrack \bar{\Lambda}^{q-\frac{1}{2}} \zeta_j + \frac{1}{2} \sum_{jk} \bar{\Lambda}^{\frac{1}{2}} \frac{\partial^2 A_i}{\partial Z_j \partial Z_k} \rbrack \bar{\Lambda}^{2(q-\frac{1}{2})} \zeta_j \zeta_k + \dots \tag{E12}$$

Using Eq. (E5a), the above can be written to explicitly extract the dependence on nonlinearity:

$$\bar{\Lambda}^{\frac{1}{2}} A_i = \left(\sum_j \bar{\Lambda}^{\frac{1}{2}} \mathbf{A}_{ij}^L \bar{\Lambda}^{-\frac{1}{2}} \langle \bar{Z}_j \rangle \right) + \bar{\Lambda}^{\frac{3}{2}} A_i^N \rfloor + \bar{\Lambda}^{\frac{1}{2}} f_i$$

$$+ \bar{\Lambda}^q \left\{ \sum_j \bar{\Lambda}^{\frac{1}{2}} \mathbf{A}_{ij}^L \bar{\Lambda}^{-\frac{1}{2}} \zeta_j + \sum_j \bar{\Lambda}^{\frac{3}{2}} \frac{\partial A_i^N}{\partial Z_j} \rfloor \bar{\Lambda}^{-\frac{1}{2}} \zeta_j \right\}$$

$$+ \bar{\Lambda}^{2q} \left\{ \frac{1}{2} \sum_{jk} \bar{\Lambda}^{\frac{3}{2}} \frac{\partial^2 A_i^N}{\partial Z_j \partial Z_k} \rfloor \bar{\Lambda}^{-1} \zeta_j \zeta_k \right\} + \dots \tag{E13}$$

The evaluation ($\rfloor$) of nonlinear terms A_i^N and their derivatives at the expansion point will return additional factors of Λ, subject to how the A_i^N depend on phase space variables $\vec{Z}$. This is ultimately determined by the order of nonlinearity under consideration. For example, as we will see for the Kerr model, A_i^N are *cubic* in phase space variables Z_j; as a result, evaluating the nonlinear drift terms yields factors of $(\bar{\Lambda}^{-\frac{1}{2}})^3$,

$$A_i^N(\vec{Z}) \rfloor = \bar{\Lambda}^{-\frac{3}{2}} A_i^N(\langle \vec{\bar{Z}} \rangle) \equiv \bar{\Lambda}^{-\frac{3}{2}} \bar{A}_i^N \tag{E14}$$

so that we have now defined $\bar{A}_i^N$ as the nonlinear drift term only in terms of $\langle \vec{\bar{Z}} \rangle$. Analogous relations can be obtained if the nature of nonlinearity is different from the Kerr model considered here.

We can straightforwardly compute derivatives of the left hand side with respect to phase space variables, as required in Eq. (E13). First, we note that since A_i^N depends at most cubically on the phase space variables, derivatives of order $n > 3$ with respect to these variables must vanish. Derivatives of lower order can be evaluated with the help of the chain rule,

$$\frac{\partial^{(n)} A_i^N}{\prod_j^n \partial Z_j} \rfloor = \begin{cases} \bar{\Lambda}^{-\frac{3}{2}} \prod_j^n \frac{\partial \langle \bar{Z}_j \rangle}{\partial Z_j} \frac{\partial^{(n)} \bar{A}_i^N}{\prod_j^n \partial \langle \bar{Z}_j \rangle}, & 0 \le n \le 3 \\ 0, & \text{otherwise} \end{cases} \tag{E15}$$

Here, we note from Eq. (E6) that $\prod_j^n \frac{\partial \langle \bar{Z}_j \rangle}{\partial Z_j} = \bar{\Lambda}^{\frac{n}{2}}$. Simplifying and rearranging, we can write derivatives of order n entirely as derivatives of the scaled nonlinear drift terms $\bar{A}_i^N$,

$$\bar{\Lambda}^{\frac{3}{2}} \cdot \frac{\partial^{(n)} A_i^N}{\prod_j^n \partial Z_j} \rfloor = \begin{cases} \bar{\Lambda}^{\frac{n}{2}} \frac{\partial^{(n)} \bar{A}_i^N}{\prod_j^n \partial \langle \bar{Z}_j \rangle}, & 0 \le n \le 3 \\ 0, & \text{otherwise} \end{cases} \tag{E16}$$

Substituting Eq. (E16) into Eq. (E13), we finally obtain the Taylor expansion of the drift vector elements, written explicitly up to order Λ^{2q},

$$\bar{\Lambda}^{\frac{1}{2}} A_i = \left[\sum_j \mathbf{A}_{ij}^L \langle \bar{Z}_i \rangle + \bar{A}_i^N + \bar{f}_i \right] + \bar{\Lambda}^q \sum_j \left(\mathbf{A}_{ij}^L + \frac{\partial \bar{A}_i^N}{\partial \langle \bar{Z}_j \rangle} \right) \zeta_j + \bar{\Lambda}^{2q} \left(\frac{1}{2} \sum_{jk} \frac{\partial^2 \bar{A}_i^N}{\partial \langle \bar{Z}_j \rangle \partial \langle \bar{Z}_k \rangle} \right) \zeta_j \zeta_k + O(\Lambda^{3q}) \tag{E17}$$

where we have also introduced the scaled inhomogeneous terms $\bar{f}_i \equiv \bar{\Lambda}^{\frac{1}{2}} f_i$.

3. Expanding the diffusion term

Analogously, we consider the Taylor expansion of the diffusion matrix terms around the expansion point,

$$\mathbf{D}_{ij} = \mathbf{D}_{ij} \rfloor + \sum_k \frac{\partial \mathbf{D}_{ij}}{\partial Z_k} \rfloor \bar{\Lambda}^{q-\frac{1}{2}} \zeta_k + \dots \tag{E18}$$

Using Eq. (E5b), the above can be written explicitly as well:

$$\mathbf{D}_{ij} = \mathbf{D}_{ij}^L + \bar{\Lambda} \mathbf{D}_{ij}^N \rfloor + \bar{\Lambda}^q \left\{ \sum_k \bar{\Lambda} \frac{\partial \mathbf{D}_{ij}^N}{\partial Z_k} \rfloor \bar{\Lambda}^{-\frac{1}{2}} \zeta_k \right\} + \dots \tag{E19}$$

Once more, the nonlinear diffusion terms can be written by taking into account their dependence on the phase space variables. For the Kerr model, the diffusion matrix depends *quadratically* on the phase space variables; as a result, evaluation at the expansion point yields factors of $(\Lambda^{-\frac{1}{2}})^2$,

$$\mathbf{D}_{ij}^N(\vec{Z})\Big] = \bar{\Lambda}^{-1}\mathbf{D}_{ij}^N(\langle\vec{\bar{Z}}\rangle) \equiv \Lambda^{-1}\bar{\mathbf{D}}_{ij}^N \tag{E20}$$

and so we have again defined $\bar{\mathbf{D}}_{ij}^N$ as the nonlinear diffusion matrix elements in terms of $\langle\vec{\bar{Z}}\rangle$ alone.

The evaluation of derivatives proceeds as before: derivatives of order $n > 2$ vanish since the diffusion matrix elements depend at most quadratically on phase space variables, while derivatives of lower order can be written as derivatives with respect to $\langle\bar{Z}_j\rangle$ variables using the chain rule,

$$\frac{\partial^{(n)}\mathbf{D}_{ij}^N}{\prod_k^n \partial Z_k}\Bigg] = \begin{cases} \bar{\Lambda}^{-1}\prod_k^n \frac{\partial\langle\bar{Z}_k\rangle}{\partial Z_k}\frac{\partial^{(n)}\bar{\mathbf{D}}_{ij}^N}{\prod_k^n \partial\langle\bar{Z}_k\rangle}, & 0 \le n \le 2 \\ 0, & \text{otherwise} \end{cases} \tag{E21}$$

Rearranging and simplifying, we arrive at the general form of derivatives of order n of the diffusion matrix elements,

$$\bar{\Lambda} \cdot \frac{\partial^{(n)}\mathbf{D}_{ij}^N}{\prod_k^n \partial Z_k}\Bigg] = \begin{cases} \bar{\Lambda}^{\frac{n}{2}}\frac{\partial^{(n)}\bar{\mathbf{D}}_{ij}^N}{\prod_k^n \partial\langle\bar{Z}_k\rangle}, & 0 \le n \le 2 \\ 0, & \text{otherwise} \end{cases} \tag{E22}$$

Finally, substituting Eq. (E22) into Eq. (E19) yields the Taylor expansion of the diffusion matrix elements, which we write explicitly up to order Λ^q,

$$\mathbf{D}_{ij} = \left[\mathbf{D}_{ij}^L + \bar{\mathbf{D}}_{ij}^N\right] + \bar{\Lambda}^q\left\{\sum_k \frac{\partial\bar{\mathbf{D}}_{ij}^N}{\partial\langle\bar{Z}_k\rangle}\zeta_k\right\} + O(\bar{\Lambda}^{2q}) \tag{E23}$$

4. Fokker-Planck equation of fluctuations to lowest nontrivial order in QNP nonlinearity

We can now substitute the final expressions for the Taylor expansions of the drift and diffusion matrices, Eqs. (E17), (E23) respectively, into Eq. (E10). This finally yields the FPE for fluctuations in increasing powers of $\bar{\Lambda}^q$,

$$\begin{aligned}
d_\tau \bar{\mathcal{P}}(\vec{\zeta}, \tau) = {} & \bar{\Lambda}^{-q}\left(\frac{\partial\langle\bar{Z}_i\rangle}{\partial\tau} - \left[\sum_j \mathbf{A}_{ij}^L\langle\bar{Z}_j\rangle + \bar{A}_i^N + \bar{f}_i\right]\right)\frac{\partial\bar{\mathcal{P}}}{\partial\zeta_i} \\
& - \frac{\partial}{\partial\zeta_i}\left[\bar{\Lambda}^0\sum_j\left(\mathbf{A}_{ij}^L + \frac{\partial\bar{A}_i^N}{\partial\langle\bar{Z}_j\rangle}\right)\zeta_j + \bar{\Lambda}^q\left(\frac{1}{2}\sum_{jk}\frac{\partial^2\bar{A}_i^N}{\partial\langle\bar{Z}_j\rangle\partial\langle\bar{Z}_k\rangle}\zeta_j\zeta_k\right) + O(\bar{\Lambda}^{2q})\right]\bar{\mathcal{P}}(\vec{\zeta}, \tau) \\
& + \frac{1}{2}\frac{\partial^2}{\partial\zeta_i\partial\zeta_j}\left[\bar{\Lambda}^{-2(q-\frac{1}{2})}\left(\mathbf{D}_{ij}^L + \bar{\mathbf{D}}_{ij}^N\right) + \bar{\Lambda}^{1-q}\left(\sum_k\frac{\partial\bar{\mathbf{D}}_{ij}^N}{\partial\langle\bar{Z}_k\rangle}\zeta_k\right) + O(\bar{\Lambda})\right]\bar{\mathcal{P}}(\vec{\zeta}, \tau)
\end{aligned} \tag{E24}$$

The leading term of order $\bar{\Lambda}^{-q}$ will diverge as $\bar{\Lambda} \to 0$ and therefore its coefficient must be set to zero; this yields a set of TEOMs for first-order cumulants defining the expansion point around which we are considering fluctuations,

$$\frac{\partial}{\partial\tau}\langle\vec{\bar{Z}}\rangle = \mathbf{A}^L\langle\vec{\bar{Z}}\rangle + \vec{A}^N(\langle\vec{\bar{Z}}\rangle) + \vec{\bar{f}} \tag{E25}$$

Recalling that $\vec{A}^N$ are nonlinear functions of $\langle\vec{\bar{Z}}\rangle$, the above system of equations is clearly nonlinear.

Satisfying Eq. (E39) ensures that the $\bar{\Lambda} \to 0$ limit does not diverge. However, to obtain a well-defined distribution function in this limit, with both finite drift and diffusion contributions, we require the leading order drift term, $O(\bar{\Lambda}^0)$, and the leading diffusion term, $O(\bar{\Lambda}^{-2(q-\frac{1}{2})})$, to be of the same order. This naturally enforces $q = \frac{1}{2}$ as the choice of the expansion parameter.

Our analysis thus far is exact, since all terms in the expansion are retained. We can now make an approximation in the strength of the nonlinearity by dropping terms in Eq. (E24). However, it is clear that under the choice $q = \frac{1}{2}$, the second order drift and diffusion terms both appear at the same order of nonlinearity.

The resulting approximate FPE of fluctuations then takes the compact form,

$$d_\tau \bar{\mathcal{P}}(\vec{\zeta}, \tau) \simeq \left(-\frac{\partial}{\partial \zeta_i} \left[\bar{\Lambda}^0 \sum_j \bar{\mathbf{J}}_{ij} \zeta_j + \bar{\Lambda}^{\frac{1}{2}} \sum_{jk} \bar{\mathbf{H}}_{ijk} \zeta_j \zeta_k \right] + \frac{1}{2} \frac{\partial^2}{\partial \zeta_i \partial \zeta_j} \left[\bar{\Lambda}^0 \bar{\mathbf{B}}_{ij} + \sum_k O(\bar{\Lambda}^{\frac{1}{2}}) \zeta_k \right] \right) \bar{\mathcal{P}}(\vec{\zeta}, \tau) \tag{E26}$$

where we have defined $\bar{\mathbf{J}}$ as the familiar Jacobian describing the measurement chain,

$$\bar{\mathbf{J}}_{ij} = \mathbf{A}_{ij}^L + \frac{\partial \bar{A}_i^N}{\partial \langle \bar{Z}_j \rangle} \tag{E27}$$

which clearly includes contributions from the first derivative of the nonlinearity vector $\vec{\bar{A}}^N$. Secondly, we have also introduced the linearized diffusion matrix $\bar{\mathbf{B}}$ evaluated at the expansion point,

$$\bar{\mathbf{B}}_{ij} = \mathbf{D}_{ij}^L + \bar{\mathbf{D}}_{ij}^N \tag{E28}$$

The most interesting term is the final drift term of $O(\bar{\Lambda}^{\frac{1}{2}})$, where we have introduced $\bar{\mathbf{H}}_{ijk}$ as the *Hessian tensor* of the nonlinearity vector $\vec{\bar{A}}^N$; its elements are given by

$$\mathbf{H}_{ijk} \equiv \frac{1}{2} \frac{\partial^2 \bar{A}_i^N}{\partial \langle \bar{Z}_j \rangle \partial \langle \bar{Z}_k \rangle} \tag{E29}$$

Here, it plays the role of mapping second-order cumulants to first-order cumulants.

5. TEOMs to lowest order in the QNP nonlinearity using the NVK approximation

Under the ansatz of Eq. (E6), the desired first-order cumulants in the original basis can be related to cumulants of the fluctuation variables:

$$\langle Z_i \rangle = \bar{\Lambda}^{-\frac{1}{2}} \langle \bar{Z}_i \rangle + \langle \zeta_i \rangle \tag{E30}$$

where we have used $q = \frac{1}{2}$. Similarly, second-order cumulants in the original basis can be related to cumulants of the fluctuation variables:

$$\begin{aligned}
\mathbf{C}_{ij} &= \langle Z_i Z_j \rangle - \langle Z_i \rangle \langle Z_j \rangle \\
&= \bar{\Lambda}^{-1} \langle \bar{Z}_i \rangle \langle \bar{Z}_j \rangle + \bar{\Lambda}^{-\frac{1}{2}} \langle \bar{Z}_i \rangle \langle \zeta_j \rangle + \bar{\Lambda}^{-\frac{1}{2}} \langle \bar{Z}_j \rangle \langle \zeta_i \rangle + \langle \zeta_i \zeta_j \rangle - \left(\bar{\Lambda}^{-1} \langle \bar{Z}_i \rangle \langle \bar{Z}_j \rangle + \bar{\Lambda}^{-\frac{1}{2}} \langle \bar{Z}_i \rangle \langle \zeta_j \rangle + \bar{\Lambda}^{-\frac{1}{2}} \langle \bar{Z}_j \rangle \langle \zeta_i \rangle + \langle \zeta_i \rangle \langle \zeta_j \rangle \right) \\
&= \langle \zeta_i \zeta_j \rangle - \langle \zeta_i \rangle \langle \zeta_j \rangle \tag{E31}
\end{aligned}$$

Therefore, if we can obtain TEOMS for the fluctuation variables, these can be solved to obtain cumulants in the original basis. Our work in deriving the FPE for fluctuation variables, Eq. (E26), provides us exactly the path to deriving such simplified TEOMs within the NVK approximation. In particular, we note that the positive-P distribution that governs fluctuations can be used to compute expectation values of any function O of phase space fluctuation variables ζ_i via the definition $\langle O \rangle = \int d\vec{\zeta}\, O \bar{\mathcal{P}}(\vec{\zeta}, \tau)$. Computing time derivatives of this definition, we find:

$$\frac{d}{d\tau} \langle O \rangle = \int d\vec{\zeta}\, O\, d_\tau \bar{\mathcal{P}}(\vec{\zeta}, \tau) \tag{E32}$$

The time derivative in the integrand is simply the definition of the FPE governing fluctuations, Eq. (E26). Making use of the FPE, the time derivative in the integrand can be eliminated in favour of derivatives of $\bar{\mathcal{P}}$ with respect to phase space variables. Standard applications of integration by parts can then be used until the integrand is defined only in terms of $\bar{\mathcal{P}}$ and not its derivatives; for full details we refer the reader to standard texts such as Ref. [86].

This approach can be used to obtain equations of motion for first-order cumulants of fluctuations, namely $O \to \zeta_i$:

$$\frac{\partial}{\partial \tau} \langle \zeta_i \rangle \simeq \sum_j \bar{\mathbf{J}}_{ij} \langle \zeta_j \rangle + \bar{\Lambda}^{\frac{1}{2}} \sum_{jk} \bar{\mathbf{H}}_{ijk} \langle \zeta_j \zeta_k \rangle. \tag{E33}$$

Similarly, we can obtain equations of motion for the second-order quantities by setting $O \to \zeta_i \zeta_j$. To lowest nontrivial order in the nonlinearity, we obtain:

$$\frac{d}{d\tau}\langle\zeta_i\zeta_j\rangle = \sum_k \bar{\mathbf{J}}_{ik}\langle\zeta_k\zeta_j\rangle + \sum_k \bar{\mathbf{J}}_{jk}\langle\zeta_k\zeta_i\rangle + \bar{\mathbf{B}}_{ij} + O(\bar{\Lambda}^{\frac{1}{2}}). \tag{E34}$$

From the above equations, we are able to make some useful deductions. First, we see that $\langle\zeta_i\zeta_j\rangle$ will in general be at least $O(\bar{\Lambda}^0)$, provided $\bar{\mathbf{B}}$ is nonvanishing. Secondly, we see that $\langle\zeta_i\rangle$ is at least $O(\bar{\Lambda}^{\frac{1}{2}})$; dropping the term $\propto \bar{\Lambda}^{\frac{1}{2}}$ yields $\langle\zeta_i\rangle \to 0$. Consequently, $\mathbf{C}_{jk} = \langle\zeta_j\zeta_k\rangle + O(\bar{\Lambda}^1)$. Therefore, when retaining terms up to order $O(\bar{\Lambda}^{\frac{1}{2}})$, we can with reasonable confidence approximate $\mathbf{C}_{jk} \simeq \langle\zeta_j\zeta_k\rangle$. Under this approximation, we can therefore write for the equations of motion of first-order cumulants of fluctuations:

$$\frac{\partial}{\partial\tau}\langle\zeta_i\rangle \simeq \sum_j \bar{\mathbf{J}}_{ij}\langle\zeta_j\rangle + \bar{\Lambda}^{\frac{1}{2}}\sum_{jk}\bar{\mathbf{H}}_{ijk}\mathbf{C}_{jk} + O(\bar{\Lambda}^{\frac{3}{2}}) \tag{E35}$$

and similarly for the second-order cumulants:

$$\frac{d}{d\tau}\langle\zeta_i\zeta_j\rangle \simeq \frac{d}{d\tau}\mathbf{C}_{ij} = \sum_k \bar{\mathbf{J}}_{ik}\mathbf{C}_{kj} + \sum_k \bar{\mathbf{J}}_{jk}\mathbf{C}_{ki} + \bar{\mathbf{B}}_{ij} + O(\bar{\Lambda}^{\frac{1}{2}}) \tag{E36}$$

The first set of equations can be compactly expressed in matrix form as:

$$\frac{\partial}{\partial\tau}\langle\vec{\zeta}\rangle \simeq \bar{\mathbf{J}}\langle\vec{\zeta}\rangle + \bar{\Lambda}^{\frac{1}{2}}\bar{\mathbf{H}} : \mathbf{C}, \tag{E37}$$

where (:) represents the tensor double contraction, while the second set takes the form of a Lyapunov system,

$$\frac{d}{d\tau}\mathbf{C} = \bar{\mathbf{J}}\mathbf{C} + \mathbf{C}\bar{\mathbf{J}}^T + \bar{\mathbf{B}}. \tag{E38}$$

Solving the above sets of equations requires solving for the expectation value $\langle\vec{\bar{Z}}\rangle$, governed by the system of equations,

$$\frac{\partial}{\partial\tau}\langle\vec{\bar{Z}}\rangle = \mathbf{A}^L\langle\vec{\bar{Z}}\rangle + \vec{A}^N(\langle\vec{\bar{Z}}\rangle) + \vec{f}. \tag{E39}$$

Eqs. (E37), (E38), and (E39) together allow us to solve for the first and second-order cumulants of fluctuations within the NVK approximation. Finally, using Eq. (E30), (E31), these solutions allow us to determine the first and second-order cumulants in the original basis to lowest order in the nonlinearity strength. In SI Sec. G, we use the solution of these TEOMs for the calculation of quantities that are relevant to quantum state classification considered in the main text.

Finally, we will also require the unequal-time correlations of fluctuation variables, $\mathbf{C}_{ij}(\tau+\theta,\tau) = \langle Z_i(\tau+\theta)Z_j(\tau)\rangle - \langle Z_i(\tau+\theta)\rangle\langle Z_j(\tau)\rangle$, for the calculation of measured covariances (see SI Sec. F 2). For linear(ized) systems governed by $\bar{\mathbf{J}}$, the required correlation matrix can be written entirely in terms of its equal-time counterpart [86], again to lowest nontrivial order in the nonlinearity,

$$\mathbf{C}(\tau+\theta,\tau) = e^{\bar{\mathbf{J}}\theta}\mathbf{C}(\tau) \tag{E40a}$$

$$\mathbf{C}(\tau-\theta,\tau) = \mathbf{C}(\tau-\theta)e^{\bar{\mathbf{J}}^T\theta} \tag{E40b}$$

where the second equation is obtained from the first by computing the transpose, replacing $\tau \to \tau - \theta$, and recalling that the normal-ordered covariance matrix is a symmetric matrix.

6. Comparison of full (S)TEOMs and the NVK approximation

The NVK approximation we have developed solves for the first and second-order cumulants of the complete quantum state of the measurement chain to lowest order in the QNP nonlinearity. This approximation drops certain terms from the full (S)TEOMs even at the same truncation order n_{ord}. Such terms are clearly seen even in the simple example of TEOMs of a single Kerr oscillator in SI Sec. C: here, TEOMs for second-order cumulants, e.g. Eqs. (C29a), (C29b), include terms nonlinear in the cumulants that are absent from Eqs. (E38). Therefore, we find that the NVK approximation and full (S)TEOMs agree for weaker nonlinearity strengths, while for increasing nonlinearity the (S)TEOMs provide a numerical scheme that is valid beyond the NVK approximation. A clear example of the comparison between the NVK approximation and full (S)TEOMs is seen in the exploration of QNP performance against nonlinearity strength in Fig. 9 of the main text.

Appendix F: Quantum statistics of measured quadratures under continuous heterodyne measurement

The STEOMs framework developed in SI Sec. C 4 provides an efficient numerical scheme to sample measurement records obtained from measurement chains deploying QNPs as processors of quantum signals from a QS. Consequently, it provides an approach to determining the quantum statistics of these measurement records by computing empirical averages and variances, analogous to how such statistics may be obtained in any real experiment. The quantum statistics of these measurement records must be related to the quantum state of the QNP, and ultimately to the upstream QS whose signals are being processed by the QNP. In this section, we clarify the first of these connections. In doing so, we also build towards a semi-analytic scheme that allows determining the quantum statistics of measurement records by solving the deterministic TEOMs, as opposed to the more computationally expensive stochastic integration of the STEOMs.

We will begin by identifying precisely the information accessible to an experimentalist performing heterodyne measurements on quantum modes in a measurement chain. Then, we show how these quantities can be related to expectation values of appropriate dynamical variables computed with respect to the quantum state $\hat{\rho}$ of the measurement chain. From a theoretical standpoint, this is simply an application of quantum input-output theory, presented here in the language of quantum trajectories in the Schrödinger picture, as opposed to the perhaps more common description in the Heisenberg representation.

1. Quadrature expectation values

Recall that the heterodyne measured quadratures extracted from the measurement chain are as defined in Eq. (4) of the main text. We are interested in the mean and variance of these measured quadratures. It is sufficient to consider the measured quadratures of just one mode,

$$\boldsymbol{\mu}_k(t) \equiv \mathbb{E}\left[\begin{pmatrix} I_k(t) \\ Q_k(t) \end{pmatrix}\right] = \frac{1}{\sqrt{2\mathcal{T}}}\int_{t-\mathcal{T}}^{t} d\tau \; \mathbb{E}\left[\begin{pmatrix} \mathcal{I}_k(\tau) \\ \mathcal{Q}_k(\tau) \end{pmatrix}\right]. \tag{F1}$$

The required expectation value of measured quadratures can be computed using Eqs. (4) of the main text,

$$\mathbb{E}\left[\begin{pmatrix} \mathcal{I}_k(\tau) \\ \mathcal{Q}_k(\tau) \end{pmatrix}\right] = \mathbb{E}\left[\begin{pmatrix} \xi_{\mathcal{I}_k}(\tau) + \sqrt{\gamma_{\mathrm{H}}}\left[\langle\hat{X}_k(\tau)\rangle + \xi_{\mathcal{I}_k}^{\mathrm{qm}}(\tau)\right] + \sqrt{\bar{n}_{\mathrm{cl}}}\;\xi_{\mathcal{I}_k}^{\mathrm{cl}}(\tau) \\ \xi_{\mathcal{Q}_k}(\tau) + \sqrt{\gamma_{\mathrm{H}}}\left[\langle\hat{P}_k(\tau)\rangle + \xi_{\mathcal{Q}_k}^{\mathrm{qm}}(\tau)\right] + \sqrt{\bar{n}_{\mathrm{cl}}}\;\xi_{\mathcal{Q}_k}^{\mathrm{cl}}(\tau) \end{pmatrix}\right]$$
$$= \sqrt{\gamma_{\mathrm{H}}}\begin{pmatrix} \langle\hat{X}_k(\tau)\rangle \\ \langle\hat{P}_k(\tau)\rangle \end{pmatrix} \tag{F2}$$

where we have used the fact that the expectation values of all the noise terms vanish; for the quantum noise contributions, this follows since the expectation value of conditional quantum trajectories is equal to their unconditional expectation value, $\mathbb{E}[\langle\hat{O}(t)\rangle^c] = \mathbb{E}[\mathrm{tr}\{\hat{O}\hat{\rho}^c(t)\}] = \mathrm{tr}\{\hat{O}\mathbb{E}[\hat{\rho}^c(t)]\} = \mathrm{tr}\{\hat{O}\hat{\rho}(t)\} = \langle\hat{O}(t)\rangle$.

The calculation thus far is general; we now specialize to the requirement of steady-state measured quadratures for the quantum state classification task under consideration. For long enough t, the right hand side above reaches steady state. Substituting the result into Eq. (F1), the time integral can be trivially computed, yielding

$$\boldsymbol{\mu}_k = \mathbb{E}\left[\begin{pmatrix} I_k \\ Q_k \end{pmatrix}\right] = \sqrt{\frac{\mathcal{T}}{2}} \cdot \sqrt{\gamma_{\mathrm{H}}}\begin{pmatrix} \langle\hat{X}_k\rangle \\ \langle\hat{P}_k\rangle \end{pmatrix} \tag{F3}$$

where for compactness we suppress time labels to denote steady state (time-independent) quantities.

We see that expectation values of measured quadratures directly probe internal quadratures of the modes being monitored, which are themselves related to first-order cumulants. However, note that the modes that are monitored are often a subset of the complete set of modes comprising a quantum measurement chain. It therefore proves useful to relate the measured expectation values to the complete set of first-order cumulants describing the quantum measurement chain. To do so, we introduce the vector defining operators for the $R = M + K$ modes of the complete measurement chain, $\hat{\boldsymbol{z}} = (\hat{a}_1, \hat{a}_1^\dagger, \ldots, \hat{a}_M, \hat{a}_M^\dagger, \hat{b}_1, \hat{b}_1^\dagger, \ldots, \hat{b}_K, \hat{b}_K^\dagger)^T$. Then the vector of first-order cumulants is simply $\langle\hat{\boldsymbol{z}}\rangle$. Then, the mode quadratures can be written in terms of first-order cumulants via

$$\sqrt{\gamma_{\mathrm{H}}}\begin{pmatrix} \langle\hat{X}_k\rangle \\ \langle\hat{P}_k\rangle \end{pmatrix} = \mathbf{M}_k\langle\hat{\boldsymbol{z}}\rangle \tag{F4}$$

where we have introduced the *heterodyne measurement matrix* $\mathbf{M}_k$, which simply serves to extract the appropriate subset of mode quadratures in the canonical basis. It is a 2-by-2R matrix with nonzero entries only in columns corresponding to the kth mode in $\hat{\boldsymbol{z}}$:

$$\mathbf{M}_k = \sqrt{\gamma_{\mathrm{H}}} \begin{pmatrix} \mathbf{0} & \cdots & \underbrace{\mathbf{U}}_{k\text{th mode}} & \cdots & \mathbf{0} \end{pmatrix}, \tag{F5}$$

where $\mathbf{U}$ is the standard unitary matrix defining a change-of-basis from canonical quadratures to normal-ordered first-order cumulants,

$$\mathbf{U} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ -i & i \end{pmatrix}. \tag{F6}$$

We can hence finally write the expectation values of measured quadratures in terms of first-order cumulants obtained via TEOMs,

$$\boldsymbol{\mu}_k = \mathbb{E}\left[\begin{pmatrix} I_k \\ Q_k \end{pmatrix} \right] = \sqrt{\frac{\mathcal{T}}{2}} \cdot \mathbf{M}_k \langle \hat{\boldsymbol{z}} \rangle \tag{F7}$$

2. Quadrature noise and correlations

We are also interested in the noise of measured quadratures, including correlations between any pair of distinct quadratures. This information can be accessed by computing quadrature variances. To this end, we introduce the measured quadrature covariance matrix $\boldsymbol{\Sigma}_{jk}$ which describes the correlations between the measured quadratures of any pair (j, k) of modes:

$$\boldsymbol{\Sigma}_{jk}(t) \equiv \mathbb{E}\left[\begin{pmatrix} I_j(t) \\ Q_j(t) \end{pmatrix} \begin{pmatrix} I_k(t) \\ Q_k(t) \end{pmatrix}^T \right] - \mathbb{E}\left[\begin{pmatrix} I_j(t) \\ Q_j(t) \end{pmatrix} \right] \mathbb{E}\left[\begin{pmatrix} I_j(t) \\ Q_j(t) \end{pmatrix} \right]^T \tag{F8}$$

Using the definition of measured quadratures, we can write the above as:

$$\boldsymbol{\Sigma}_{jk}(t) = \frac{\bar{n}_{\mathrm{cl}}}{2}\mathbf{I}_2\delta_{jk} + \frac{1}{2\mathcal{T}} \int_{t-\mathcal{T}}^{t} \int_{t-\mathcal{T}}^{t} d\tau d\tau' \times \left\{ \mathbb{E}\left[\begin{pmatrix} \mathcal{I}_j(\tau') \\ \mathcal{Q}_j(\tau') \end{pmatrix} \begin{pmatrix} \mathcal{I}_k(\tau) \\ \mathcal{Q}_k(\tau) \end{pmatrix}^T \right] - \mathbb{E}\left[\begin{pmatrix} \mathcal{I}_j(\tau') \\ \mathcal{Q}_j(\tau') \end{pmatrix} \right] \mathbb{E}\left[\begin{pmatrix} \mathcal{I}_k(\tau) \\ \mathcal{Q}_k(\tau) \end{pmatrix} \right]^T \right\}$$

$$\equiv \frac{\bar{n}_{\mathrm{cl}}}{2}\mathbf{I}_2\delta_{jk} + \frac{1}{2\mathcal{T}} \int_{t-\mathcal{T}}^{t} \int_{t-\mathcal{T}}^{t} d\tau d\tau' \, \boldsymbol{\mathcal{S}}_{jk}(\tau', \tau) \tag{F9}$$

where we have introduced $\boldsymbol{\mathcal{S}}_{jk}(t', t)$ as the correlation matrix of measured heterodyne records prior to any post-processing. A standard calculation [48] can be used to relate this measured correlation matrix to correlations of mode operators:

$$\boldsymbol{\mathcal{S}}_{jk}(\tau', \tau) = \delta_{jk}\delta(\tau - \tau')\mathbf{I}_2 + \mathbf{M}_j \mathbf{C}(\tau', \tau)\mathbf{M}_k^T \tag{F10}$$

Here, the first term arises simply due to the δ-correlation of measurement noise, which only contributes when autocorrelations are being calculated. The second term is the important contribution that depends on the state of the quantum modes being probed, via $\mathbf{C}(\tau', \tau)$ which is the unconditional normal-ordered correlation matrix of the complete measurement chain. The elements of the correlation matrix are defined explicitly as

$$[\mathbf{C}(\tau', \tau)]_{ij} = \langle :\hat{z}_i(\tau')\hat{z}_j(\tau): \rangle - \langle \hat{z}_i(\tau') \rangle \langle \hat{z}_j(\tau) \rangle \tag{F11}$$

where $\langle :\hat{f}: \rangle$ implies an expectation value where the arbitrary operator $\hat{f}$ is written in normal-ordered form. The measurement matrices in Eq. (F10) then once again extract the sectors of the correlation matrix pertinent to the correlations of mode pair (j, k).

Then, substituting Eq. (F10) into Eq. (F9), the measured covariance matrix takes the form:

$$\boldsymbol{\Sigma}_{jk}(t) = \sigma_{\mathrm{vac}}^2\mathbf{I}_2(\bar{n}_{\mathrm{cl}} + 1)\delta_{jk} + \frac{1}{2\mathcal{T}}\mathbf{M}_j \left[\int_{t-\mathcal{T}}^{t} \int_{t-\mathcal{T}}^{t} d\tau d\tau' \, \mathbf{C}(\tau', \tau) \right] \mathbf{M}_k^T \tag{F12}$$

where $\sigma_{\text{vac}}^2 = \frac{1}{2}$ as introduced in the main text. The first term then describes the vacuum noise contribution that is always present. The second term is the filtered contribution from monitoring the quantum system. In particular, it depends non-trivially on the filter window $\mathcal{T}$.

It is clear that for the completely general filter window $\mathcal{T}$, we will also require an expression for the correlation matrix $\mathbf{C}(\tau', \tau)$ for *unequal* times $\tau \neq \tau'$. We first consider a special case of the measured covariances for short filter windows, where Eq. (F12) simplifies considerably. We then consider approximations of measurement chains under which simple analytic expressions for the correlation matrix can be found.

3. Steady-state measured covariances in the short-filter limit

A simple case of heterodyne monitoring involves a short filter window $\mathcal{T} \to 0$. We can then make the assumption that the correlation matrix $\mathbf{C}(\tau', \tau)$ is effectively unchanged over the course of the short window $[t - \mathcal{T}, t]$. We can then approximate $\tau \simeq \tau' \to t$ over the integration window, thereby replacing the correlation matrix by its equal-time counterpart, the *covariance* matrix $\mathbf{C}(t)$. Extracting the covariance matrix outside of the integral in Eq. (F12) (in analogy with standard Markov-like approximations) yields:

$$\mathbf{\Sigma}_{jk}(t) \overset{\mathcal{T} \to 0}{\simeq} \sigma_{\text{vac}}^2 \mathbf{I}_2 (\bar{n}_{\text{cl}} + 1) \delta_{jk} + \frac{1}{2\mathcal{T}} \mathbf{M}_j \mathbf{C}(t) \left[\int_{t-\mathcal{T}}^t \int_{t-\mathcal{T}}^t d\tau d\tau' \right] \mathbf{M}_k^T, \tag{F13}$$

which immediately simplifies to:

$$\mathbf{\Sigma}_{jk}(t) \overset{\mathcal{T} \to 0}{\simeq} \sigma_{\text{vac}}^2 \mathbf{I}_2 (\bar{n}_{\text{cl}} + 1) \delta_{jk} + \frac{\mathcal{T}}{2} \cdot \mathbf{M}_j \mathbf{C}(t) \mathbf{M}_k^T \tag{F14}$$

Therefore, for short enough filter windows, the measured covariances can be related entirely to the measured mode covariance matrix $\mathbf{C}(t)$. We emphasize that Eq. (F24) holds generally for *any* quantum measurement chain under heterodyne monitoring. In the next subsection, we will make use of the properties of linear(ized) systems to obtain expressions for the measured covariances beyond the assumption of short filter window. As we will show, for long enough filter windows, even equal-time covariances of measured quadratures will include a temporal dependence on correlations of the measured modes at different times.

4. Steady-state measured covariances for linear(ized) systems under the NVK approximation

In SI Sec. E, we have shown that for linear(ized) systems under an appropriate assumption of weak nonlinearity - made precise within the NVK approximation - the correlation matrix $\mathbf{C}(\tau', \tau)$ is governed by the linearized Jacobian matrix $\bar{\mathbf{J}}$ and the steady-state equal-time covariance matrix $\mathbf{C}(\tau)$ by Eqs. (E40a), (E40b). For convenience, we reproduce these equations below:

$$\mathbf{C}(\tau + \theta, \tau) = e^{\bar{\mathbf{J}}\theta} \mathbf{C}(\tau) \tag{F15a}$$

$$\mathbf{C}(\tau - \theta, \tau) = \mathbf{C}(\tau - \theta) e^{\bar{\mathbf{J}}^T \theta} \tag{F15b}$$

This fact allows us to simplify the expression for measured covariances, Eq. (F12), for such linear(ized) systems. In particular, we can first rewrite the integral in Eq. (F12) as two terms:

$$\int_{t-\mathcal{T}}^t \int_{t-\mathcal{T}}^t d\tau d\tau' \, \mathbf{C}(\tau', \tau) = \int_{t-\mathcal{T}}^t \int_{t-\mathcal{T}}^\tau d\tau d\tau' \, \mathbf{C}(\tau', \tau) + \int_{t-\mathcal{T}}^t \int_\tau^t d\tau d\tau' \, \mathbf{C}(\tau', \tau) \tag{F16}$$

where in the first term $\tau' < \tau$, while in the second term $\tau' > \tau$. Next, we introduce an alternate, positive integration variable θ to denote the temporal separation of τ and τ': precisely, we make the substitution $\tau - \tau' = \theta$ in the first term and $\tau' - \tau = \theta$ in the second, to obtain

$$\int_{t-\mathcal{T}}^t \int_{t-\mathcal{T}}^t d\tau d\tau' \, \mathbf{C}(\tau', \tau) = \int_{t-\mathcal{T}}^t \int_0^{\tau - (t-\mathcal{T})} d\tau d\theta \, \mathbf{C}(\tau - \theta, \tau) + \int_{t-\mathcal{T}}^t \int_0^{t-\tau} d\tau d\theta \, \mathbf{C}(\tau + \theta, \tau) \tag{F17}$$

Substituting Eqs. (F15a), (F15b) into Eq. (F17) and simplifying thus allows us to write the correlation matrix contribution entirely in terms of covariance matrices,

$$\int_{t-\mathcal{T}}^t \int_{t-\mathcal{T}}^t d\tau d\tau' \, \mathbf{C}(\tau', \tau) = \left(\int_{t-\mathcal{T}}^t \int_0^{\tau - (t-\mathcal{T})} d\tau d\theta \, \mathbf{C}(\tau - \theta) e^{\bar{\mathbf{J}}^T \theta} \right) + \left(\int_{t-\mathcal{T}}^t \int_0^{t-\tau} d\tau d\theta \, e^{\bar{\mathbf{J}}\theta} \mathbf{C}(\tau) \right) \tag{F18}$$

We can now specialize to the case of steady-state covariances. In particular, the minimal value of integration variables τ, θ is $t - \mathcal{T}$. Provided $t - \mathcal{T}$ is long enough in comparison to the relaxation rates of the measurement chain (by choosing large enough t), the time dependence of the covariance matrix $\mathbf{C}(\tau)$ itself can be neglected. In this case, we can replace $\mathbf{C}(\tau) \to \mathbf{C}$, and simplify the filtered correlation matrix,

$$\int_{t-\mathcal{T}}^{t} \int_{t-\mathcal{T}}^{t} d\tau d\tau' \, \mathbf{C}(\tau', \tau) = \mathbf{C} \left(\int_{t-\mathcal{T}}^{t} \int_{0}^{\tau-(t-\mathcal{T})} d\tau d\theta \, e^{\bar{\mathbf{J}}^T \theta} \right) + \left(\int_{t-\mathcal{T}}^{t} \int_{0}^{t-\tau} d\tau d\theta \, e^{\bar{\mathbf{J}} \theta} \right) \mathbf{C} \tag{F19}$$

Hence we are only required to compute integrals over the $e^{\bar{\mathbf{J}} \theta}$ and its transpose. To this end, the diagonal representation of the matrix $\bar{\mathbf{J}} = \mathbf{P} \mathbf{N} \mathbf{P}^{-1}$ proves very useful, as $e^{\mathbf{A} \theta} = \mathbf{P} e^{\mathbf{N} \theta} \mathbf{P}^{-1}$ where $\mathbf{N}$ is a diagonal matrix of the eigenvalues of $\bar{\mathbf{J}}$. With this result, we can simplify the second of the two integrals above,

$$\int_{t-\mathcal{T}}^{t} \int_{0}^{t-\tau} d\tau d\theta \, e^{\bar{\mathbf{J}} \theta} = \mathbf{P} \int_{t-\mathcal{T}}^{t} d\tau \left[\int_{0}^{t-\tau} d\theta \, e^{\mathbf{N} \theta} \right] \mathbf{P}^{-1}$$

$$= \mathbf{P} \mathbf{N}^{-1} \int_{t-\mathcal{T}}^{t} d\tau \left[e^{\mathbf{N}(t-\tau)} - \mathbf{I} \right] \mathbf{P}^{-1}$$

$$= \mathbf{P} \mathbf{N}^{-2} \left[e^{\mathbf{N}\mathcal{T}} - \mathbf{I} \right] \mathbf{P}^{-1} - \mathcal{T} \bar{\mathbf{J}}^{-1}$$

$$= \bar{\mathbf{J}}^{-2} \left[e^{\bar{\mathbf{J}} \mathcal{T}} - \mathbf{I} \right] - \mathcal{T} \bar{\mathbf{J}}^{-1} \tag{F20}$$

where we have used $\bar{\mathbf{J}}^n = \mathbf{P} \mathbf{N}^n \mathbf{P}^{-1}$ for all $n \in \mathbb{Z}$. The other required integral can be analogously computed, using $(\bar{\mathbf{J}}^T)^n = (\mathbf{P}^T)^{-1} \mathbf{N}^n \mathbf{P}^T$ for all $n \in \mathbb{Z}$,

$$\int_{t-\mathcal{T}}^{t} \int_{0}^{\tau-(t-\mathcal{T})} d\tau d\theta \, e^{\bar{\mathbf{J}}^T \theta} = (\mathbf{P}^T)^{-1} \int_{t-\mathcal{T}}^{t} d\tau \left[\int_{0}^{\tau-(t-\mathcal{T})} d\theta \, e^{\mathbf{N} \theta} \right] \mathbf{P}^T$$

$$= (\mathbf{P}^T)^{-1} \mathbf{N}^{-1} \int_{t-\mathcal{T}}^{t} d\tau \left[e^{\mathbf{N}(\tau-(t-\mathcal{T}))} - \mathbf{I} \right] \mathbf{P}^T$$

$$= (\mathbf{P}^T)^{-1} \mathbf{N}^{-2} \left[e^{\mathbf{N}\mathcal{T}} - \mathbf{I} \right] \mathbf{P}^T - \mathcal{T} (\bar{\mathbf{J}}^T)^{-1}$$

$$= (\bar{\mathbf{J}}^T)^{-2} \left[e^{\bar{\mathbf{J}}^T \mathcal{T}} - \mathbf{I} \right] - \mathcal{T} (\bar{\mathbf{J}}^T)^{-1} \tag{F21}$$

The computed integrals thus allow us to determine the integral of the general correlation matrix as defined in Eq. (F17). This finally allows us to write down, after some rearrangement, the measured covariance matrix for mode pair (j, k) for linear(ized) systems via Eq. (F12):

$$\boldsymbol{\Sigma}_{jk}(t) = \sigma_{\text{vac}}^2 (\bar{n}_{\text{cl}} + 1) \mathbf{I}_2 \delta_{jk} - \frac{1}{2} \mathbf{M}_j \left[\mathbf{C} (\bar{\mathbf{J}}^T)^{-1} + \bar{\mathbf{J}}^{-1} \mathbf{C} \right] \mathbf{M}_k^T + \frac{1}{2\mathcal{T}} \mathbf{M}_j \left[\mathbf{C} (\bar{\mathbf{J}}^T)^{-2} \left(e^{\bar{\mathbf{J}}^T \mathcal{T}} - \mathbf{I} \right) + \bar{\mathbf{J}}^{-2} \left(e^{\bar{\mathbf{J}} \mathcal{T}} - \mathbf{I} \right) \mathbf{C} \right] \mathbf{M}_k^T \tag{F22}$$

5. Short and long filter limits of measured covariances under the NVK approximation

Eq. (F22) is somewhat unwieldy; we therefore evaluate some useful limits depending on the filter window length $\mathcal{T}$. First we consider $\mathcal{T} \to 0$, which corresponds to a very short filter window. In this case, the exponential terms in Eq. (F22) can be expanded to lowest order,

$$\boldsymbol{\Sigma}_{jk}(t) \overset{\mathcal{T} \to 0}{\simeq} \sigma_{\text{vac}}^2 (\bar{n}_{\text{cl}} + 1) \mathbf{I}_2 \delta_{jk} - \frac{1}{2} \mathbf{M}_j \left[\mathbf{C} (\bar{\mathbf{J}}^T)^{-1} + \bar{\mathbf{J}}^{-1} \mathbf{C} \right] \mathbf{M}_k^T$$

$$+ \frac{1}{2\mathcal{T}} \mathbf{M}_j \left[\mathbf{C} (\bar{\mathbf{J}}^T)^{-2} \left(\mathbf{I} + \bar{\mathbf{J}}^T \mathcal{T} + \frac{1}{2} (\bar{\mathbf{J}}^T)^2 \mathcal{T}^2 + \ldots - \mathbf{I} \right) + \bar{\mathbf{J}}^{-2} \left(\mathbf{I} + \bar{\mathbf{J}} \mathcal{T} + \frac{1}{2} \bar{\mathbf{J}}^2 \mathcal{T}^2 + \ldots - \mathbf{I} \right) \mathbf{C} \right] \mathbf{M}_k^T \tag{F23}$$

Expanding the second line, it becomes immediately clear that contributions of $O(\mathcal{T}^0)$ cancel contributions of the same order in the first line. As a result, the lowest order system-dependent contribution to measured covariances is $O(\mathcal{T})$,

$$\boldsymbol{\Sigma}_{jk}(t) \overset{\mathcal{T} \to 0}{\simeq} \sigma_{\text{vac}}^2 (\bar{n}_{\text{cl}} + 1) \mathbf{I}_2 \delta_{jk} + \frac{\mathcal{T}}{2} \cdot \mathbf{M}_j \mathbf{C} \mathbf{M}_k^T . \tag{F24}$$

which is exactly the steady-state version of the result obtained in Eq. (F14). In the short filter limit, the covariances of measured quadratures are therefore only sensitive to covariances of the measured modes, and not to their temporal correlations.

We now consider the opposite limit of $\gamma\mathcal{T} \gg 1$. The requirement of stability of the quantum measurement chain requires $\bar{\mathbf{J}}$ to have negative-definite eigenvalues $\mathbf{N}_{jj} \ \forall \ j$. Hence, provided $\mathcal{T} \gg \max_j\{\mathbf{N}_{jj}^{-1}\}$, $e^{\bar{\mathbf{J}}\mathcal{T}} \to 0$. For simplicity, we simply use the shorthand $\gamma\mathcal{T} \gg 1$ to refer to this regime, although formally $\max_j\{\mathbf{N}_{jj}^{-1}\}$ could be smaller than γ. Regardless, we find,

$$\mathbf{\Sigma}_{jk}(t) \stackrel{\gamma\mathcal{T}\gg 1}{\simeq} \sigma_{\mathrm{vac}}^2(\bar{n}_{\mathrm{cl}}+1)\mathbf{I}_2\delta_{jk} - \frac{1}{2}\mathbf{M}_j\left[\mathbf{C}(\bar{\mathbf{J}}^T)^{-1}+\bar{\mathbf{J}}^{-1}\mathbf{C} + \frac{1}{\mathcal{T}}\left(\mathbf{C}(\bar{\mathbf{J}}^T)^{-2}+\bar{\mathbf{J}}^{-2}\mathbf{C}\right)\right]\mathbf{M}_k^T, \tag{F25}$$

which is now effectively an expansion in $\frac{1}{\mathcal{T}}$. For such filters, the measured covariances are no longer given only by mode covariances $\mathbf{C}$ as in Eq. (F24), but also by their correlation times via the propagator $\bar{\mathbf{J}}^{-1}$. Furthermore, we find that that for the filter windows considered in this work, the contributions of $O(\frac{1}{\mathcal{T}})$ are non-negligible and must be taken into account.

The final expression can be further simplified using the Lyapunov equation that defines the steady-state $\mathbf{C}$, namely by setting the time-derivative in Eq. (E38) to zero. This yields:

$$\mathbf{0} = \bar{\mathbf{J}}\mathbf{C} + \mathbf{C}\bar{\mathbf{J}}^T + \bar{\mathbf{B}} \tag{F26}$$

Multiplying the above through on the left by $\bar{\mathbf{J}}^{-1}$ and on the right by $(\bar{\mathbf{J}}^T)^{-1}$ and rearranging, we arrive at:

$$\mathbf{C}(\bar{\mathbf{J}}^T)^{-1} + \bar{\mathbf{J}}^{-1}\mathbf{C} = -\bar{\mathbf{J}}^{-1}\bar{\mathbf{B}}(\bar{\mathbf{J}}^T)^{-1} \tag{F27}$$

If we consider the measured covariances Eq. (F25) to lowest order in $O(\frac{1}{\mathcal{T}})$, Eq. (F27) can be used to obtain a simplified form:

$$\mathbf{\Sigma}_{jk}(t) \stackrel{\gamma\mathcal{T}\gg 1}{\simeq} \sigma_{\mathrm{vac}}^2(\bar{n}_{\mathrm{cl}}+1)\mathbf{I}_2\delta_{jk} - \frac{1}{2}\mathbf{M}_j\left[\mathbf{C}(\bar{\mathbf{J}}^T)^{-1}+\bar{\mathbf{J}}^{-1}\mathbf{C}\right]\mathbf{M}_k^T$$

$$= \sigma_{\mathrm{vac}}^2(\bar{n}_{\mathrm{cl}}+1)\mathbf{I}_2\delta_{jk} + \frac{1}{2}\mathbf{M}_j\left[\bar{\mathbf{J}}^{-1}\bar{\mathbf{B}}(\bar{\mathbf{J}}^T)^{-1}\right]\mathbf{M}_k^T \tag{F28}$$

so that the measured covariances in the $\gamma\mathcal{T} \gg 1$ can be related directly to the linear(ized) drift (i.e. Jacobian) matrix $\bar{\mathbf{J}}$ and linear(ized) diffusion matrix $\bar{\mathbf{B}}$ alone.

Appendix G: Quantum state discrimination metrics under the NVK approximation

In this section, we will derive analytic expressions to quantify the fidelity of classification of QS states using the QNP. Our results here will make use of the work carried out in all three prior sections. We first require the properties of the general QS states, determined in SI Sec. D). Then, we use the results of SI Sec. E) to determine the QNP dynamics under input signals from the QS in the same measurement chain to lowest nontrivial order in nonlinearity. Finally, we must calculate the quantum statistics of measured outputs from this measurement chain (SI Sec. F) to determine classification fidelity metrics, more precisely Fisher's discriminant.

1. QS and QNP TEOMs under the NVK approximation

We begin by explicitly specifying the FPE of the measurement chain we analyze in the main text, Eq. (1). As given by Eq. (E3), this is parameterized by the drift vector and diffusion matrix introduced in Eqs. (E5a), (E5b). For convenience, we will separate these matrices into QS and QNP sectors, as will be relevant for the final expressions presented in the main text. Starting with the linear and nonlinear drift contributions respectively, these take the form:

$$\mathbf{A}^L = \begin{pmatrix} \mathbf{L}_a & \mathbf{0} \\ -\mathbf{\Gamma} & \mathbf{L}_b \end{pmatrix}, \ \vec{f} = \begin{pmatrix} \vec{\eta} \\ \vec{0} \end{pmatrix},$$

$$\vec{A}^N = \begin{pmatrix} \vec{0} \\ \vec{N}_b \end{pmatrix}, \ \vec{N}_b = \begin{pmatrix} +i\beta_1^\dagger(\beta_1)^2 \\ -i(\beta_1^\dagger)^2\beta_1 \\ \vdots \\ +i\beta_K^\dagger(\beta_K)^2 \\ -i(\beta_K^\dagger)^2\beta_K \end{pmatrix} \tag{G1}$$

Here $\mathbf{L}_a$ describes the linear dynamics of the QS, as introduced in SI Sec. D. Then, $\mathbf{L}_b$ describes the corresponding linear dynamics of the QNP, where $\mathbf{L}_b = \frac{1}{\gamma}\mathrm{diag}(\chi_1^{-1}, \chi_1^{*-1}, \ldots, \chi_K^{-1}, \chi_K^{*-1})$, and $\chi_k^{-1} = i\Delta_k - \frac{\gamma_k + \Gamma_k}{2}$ defines the linear susceptibility of the QNP modes. Then, the nonlinear terms are entirely localized to the QNP modes via the vector $\vec{N}_b$, and depends cubically on phase space variables due to the Kerr nonlinearity. Finally, $\mathbf{\Gamma}$ defines the non-reciprocal coupling between the QS and QNP, introduced in the main text via Eq. (A1).

Similarly, we can define the constant and nonlinear contributions to the diffusion matrix respectively,

$$\mathbf{D}^L = \begin{pmatrix} \mathbf{D}_a & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix},$$

$$\mathbf{D}^N = \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{D}_b \end{pmatrix}, \ \mathbf{D}_b = \begin{pmatrix} i\beta_1^2 & & & \\ & -i\beta_1^{\dagger 2} & & \mathbf{0} \\ & & \ddots & \\ & \mathbf{0} & & i\beta_K^2 \\ & & & & -i\beta_K^{\dagger 2} \end{pmatrix} \tag{G2}$$

Here, $\mathbf{D}_a$ defines the linear part of the diffusion matrix due to the QS, and is the same as that defined in Eq. (E5b). In contrast, $\mathbf{D}_b$ is the nonlinear part of the diffusion matrix that arises due to the Kerr nonlinearity of the QNP modes, and as mentioned earlier depends quadratically on phase space variables.

Using the above, we can now use the results of SI Sec. E to simply write down the TEOMs for quantum cumulants. First, we transition from the general notation used in SI Sec. E, valid for general bosonic nonlinear systems, to the measurement chains we consider, with a QS and QNP. In particular, we have for the expectation values at the expansion point, $\langle \vec{\hat{Z}} \rangle = \begin{pmatrix} \langle \hat{\vec{a}} \rangle \\ \langle \hat{\vec{b}} \rangle \end{pmatrix}$. The expectation values of fluctuation variables are equivalently represented as $\langle \vec{\zeta} \rangle = \begin{pmatrix} \langle \delta\hat{a} \rangle \\ \langle \delta\hat{b} \rangle \end{pmatrix}$.

We can immediately write down the TEOMs for the first-order cumulants defining the expansion point, Eq. (E39),

$$\frac{d}{dt}\begin{pmatrix} \langle \hat{\vec{a}} \rangle \\ \langle \hat{\vec{b}} \rangle \end{pmatrix} = \begin{pmatrix} \mathbf{L}_a & \mathbf{0} \\ -\mathbf{\Gamma} & \mathbf{L}_b \end{pmatrix}\begin{pmatrix} \langle \hat{\vec{a}} \rangle \\ \langle \hat{\vec{b}} \rangle \end{pmatrix} + \begin{pmatrix} \vec{0} \\ \vec{N}_b \end{pmatrix} + \begin{pmatrix} \vec{\eta} \\ \vec{0} \end{pmatrix} \tag{G3}$$

Explicitly writing out the above immediately yields Eqs. (7a), (7b) of the main text. The above also determines the values of $\langle \hat{\bar{\boldsymbol{b}}} \rangle$, which are necessary to determine all the subsequent quantities in the linearized analysis, as we will see shortly.

We next define the TEOMs for the second-order cumulants of the measurement chain. To do so we must obtain the Jacobian of the measurement chain introduced in Eq. (E27); using Eqs. (G1), (G2), this is straightforwardly calculated:

$$\bar{\mathbf{J}} = \begin{pmatrix} \mathbf{L}_a & \mathbf{0} \\ -\boldsymbol{\Gamma} & \bar{\mathbf{J}}_b \end{pmatrix},$$

$$\bar{\mathbf{J}}_b \equiv \mathbf{L}_b + \begin{pmatrix} i2|\langle \hat{\bar{b}}_1 \rangle|^2 & i\langle \hat{\bar{b}}_1 \rangle^2 & & & \\ -i\langle \hat{\bar{b}}_1^\dagger \rangle^2 & -i2|\langle \hat{\bar{b}}_1 \rangle|^2 & & \mathbf{0} & \\ & & \ddots & & \\ & \mathbf{0} & & i2|\langle \hat{\bar{b}}_K \rangle|^2 & i\langle \hat{\bar{b}}_K \rangle^2 \\ & & & -i\langle \hat{\bar{b}}_K^\dagger \rangle^2 & -i2|\langle \hat{\bar{b}}_K \rangle|^2 \end{pmatrix} \tag{G4}$$

We can also write down the linearized diffusion matrix $\bar{\mathbf{B}}$ from Eq. (E28) as:

$$\bar{\mathbf{B}} = \begin{pmatrix} \mathbf{D}_a & \mathbf{0} \\ \mathbf{0} & \bar{\mathbf{D}}_b \end{pmatrix}, \quad \bar{\mathbf{D}}_b = \begin{pmatrix} i\langle \hat{\bar{b}}_1 \rangle^2 & & & & \\ & -i\langle \hat{\bar{b}}_1^\dagger \rangle^2 & & \mathbf{0} & \\ & & \ddots & & \\ & \mathbf{0} & & i\langle \hat{\bar{b}}_K \rangle^2 & \\ & & & & -i\langle \hat{\bar{b}}_K^\dagger \rangle^2 \end{pmatrix} \tag{G5}$$

With these expressions, we can now write down Eq. (E38), the TEOMs for second-order cumulants to lowest order in the nonlinearity, as

$$\frac{d}{dt}\mathbf{C} = \frac{d}{dt}\begin{pmatrix} \mathbf{C}_a & \mathbf{C}_{ab} \\ \mathbf{C}_{ab}^T & \mathbf{C}_b \end{pmatrix} = \begin{pmatrix} \mathbf{L}_a & \mathbf{0} \\ -\boldsymbol{\Gamma} & \bar{\mathbf{J}}_b \end{pmatrix}\begin{pmatrix} \mathbf{C}_a & \mathbf{C}_{ab} \\ \mathbf{C}_{ab}^T & \mathbf{C}_b \end{pmatrix} + \begin{pmatrix} \mathbf{C}_a & \mathbf{C}_{ab} \\ \mathbf{C}_{ab}^T & \mathbf{C}_b \end{pmatrix}\begin{pmatrix} \mathbf{L}_a^T & -\boldsymbol{\Gamma}^T \\ \mathbf{0} & \bar{\mathbf{J}}_b^T \end{pmatrix} + \begin{pmatrix} \mathbf{D}_a & \mathbf{0} \\ \mathbf{0} & \bar{\mathbf{D}}_b \end{pmatrix} \tag{G6}$$

which is simply Eq. (8) of the main text.

Finally, once the steady-state first and second-order cumulants of the measurement chain are known to lowest order in nonlinearity, we can calculate the first-order cumulants of fluctuations by solving Eq. (E37). First, this system can be compactly written in matrix form as

$$\frac{d}{dt}\begin{pmatrix} \langle \boldsymbol{\delta\hat{a}} \rangle \\ \langle \boldsymbol{\delta\hat{b}} \rangle \end{pmatrix} = \begin{pmatrix} \mathbf{L}_a & \mathbf{0} \\ -\boldsymbol{\Gamma} & \bar{\mathbf{J}}_b \end{pmatrix}\begin{pmatrix} \langle \boldsymbol{\delta\hat{a}} \rangle \\ \langle \boldsymbol{\delta\hat{b}} \rangle \end{pmatrix} + \bar{\Lambda}^{\frac{1}{2}}\begin{pmatrix} \vec{0} \\ \vec{h}_b \end{pmatrix}. \tag{G7}$$

First we immediately note that the steady-state expression for first-order cumulants of QS fluctuations simply vanishes, $\langle \boldsymbol{\delta\hat{a}} \rangle = \vec{0}$, due to the absence of an inhomogeneous term on the right hand side. QNP first-order cumulants, on the other hand, have such a term, which is $\propto \bar{\Lambda}^{\frac{1}{2}}$ and arises from the Hessian tensor (see Eq. (E29)) contracted with the covariance matrix. This term is localized to the QNP sector via the vector $\vec{h}_b$ as expected, since this is where the nonlinearity is present. Then, the steady-state for fluctuations in the first-order cumulants of the QNP is given by

$$\langle \boldsymbol{\delta\hat{b}} \rangle = -\bar{\Lambda}^{\frac{1}{2}}\bar{\mathbf{J}}_b^{-1}\vec{h}_b. \tag{G8}$$

Furthermore, for the Kerr model, the vector $\vec{h}_b$ takes the simple form,

$$\vec{h}_b = \bar{\mathbf{H}}_b : \mathbf{C}_b = \begin{pmatrix} \bar{\mathbf{H}}_{b_1} : \mathbf{C}_{b_1} \\ \bar{\mathbf{H}}_{b_1^\dagger} : \mathbf{C}_{b_1} \\ \vdots \\ \bar{\mathbf{H}}_{b_K} : \mathbf{C}_{b_K} \\ \bar{\mathbf{H}}_{b_K^\dagger} : \mathbf{C}_{b_K} \end{pmatrix} \tag{G9}$$

where $(:)$ denotes the standard tensor double contraction. Each element in $\vec{h}_b$ is hence obtained via the double contraction of a pair of 2-by-2 matrices, $\bar{\mathbf{H}}_{b_k}$ or $\bar{\mathbf{H}}_{b_k^\dagger}$,

$$\bar{\mathbf{H}}_{b_k} = \begin{pmatrix} +i\langle \hat{\bar{b}}_k^\dagger \rangle & +i\langle \hat{\bar{b}}_k \rangle \\ +i\langle \hat{\bar{b}}_k \rangle & 0 \end{pmatrix}, \quad \bar{\mathbf{H}}_{b_k^\dagger} = \begin{pmatrix} 0 & -i\langle \hat{\bar{b}}_k^\dagger \rangle \\ -i\langle \hat{\bar{b}}_k^\dagger \rangle & -i\langle \hat{\bar{b}}_k \rangle \end{pmatrix}, \tag{G10}$$

and $\mathbf{C}_{b_k}$ simply defines the local second-order cumulants for the kth QNP mode:

$$\mathbf{C}_{b_k} = \begin{pmatrix} C_{b_k b_k} & C_{b_k^\dagger b_k} \\ C_{b_k^\dagger b_k} & C_{b_k^\dagger b_k^\dagger} \end{pmatrix}. \tag{G11}$$

This form indicates that the dynamics of the kth QNP mode is "driven" by the covariances of the kth mode. Cross-cumulants such as $C_{b_k b_j}$ for $j \neq k$ do not appear in $\vec{h}_b$ since the nonlinear drift vector $\vec{A}_b^N$ that determines the Hessian tensor involves no nonlinear coupling terms between distinct QNP modes.

2. Semi-analytic expressions for measured quadrature means, covariances, and Fisher's discriminant

Having obtained the TEOMs to lowest nontrivial order in the nonlinearity for cumulants defining the internal dynamics of the measurement chain, we can finally compute the mean and covariances of measured output quadratures for steady-state quantum state classification. The steady-state mean value of measured QNP quadratures is simply given by Eq. (F7):

$$\boldsymbol{\mu} \equiv \begin{pmatrix} \boldsymbol{\mu}_1 \\ \vdots \\ \boldsymbol{\mu}_K \end{pmatrix} = \mathbb{E}\left[\begin{pmatrix} I_1 \\ Q_1 \\ \vdots \\ I_K \\ Q_K \end{pmatrix} \right] = \sqrt{\frac{\mathcal{T}}{2}} \begin{pmatrix} \mathbf{M}_1 \\ \vdots \\ \mathbf{M}_K \end{pmatrix} \langle \vec{Z} \rangle$$

$$\implies \boldsymbol{\mu} = \sqrt{\frac{\gamma_{\mathrm{H}}\mathcal{T}}{2}} \left(\bar{\Lambda}^{-\frac{1}{2}} \langle \hat{\bar{\boldsymbol{b}}} \rangle + \langle \boldsymbol{\delta}\hat{\boldsymbol{b}} \rangle \right) \tag{G12}$$

where the second line is obtained since we are only reading out quadratures of QNP modes using heterodyne readout. The mean separation that is required to calculate Fisher's discriminant is then immediately given by:

$$\delta\boldsymbol{\mu} = \boldsymbol{\mu}^{(l)} - \boldsymbol{\mu}^{(p)} = \sqrt{\frac{\gamma_{\mathrm{H}}\mathcal{T}}{2}} \left(\langle \boldsymbol{\delta}\hat{\boldsymbol{b}}^{(l)} \rangle - \langle \boldsymbol{\delta}\hat{\boldsymbol{b}}^{(p)} \rangle \right) \tag{G13}$$

where we make use of the fact that $\langle \hat{\bar{\boldsymbol{b}}}^{(l)} \rangle = \langle \hat{\bar{\boldsymbol{b}}}^{(p)} \rangle$ for the binary state classification tasks we have considered. Hence Fisher discriminant is nonzero only due to the mean values of fluctuation variables. Using Eq. (G8), we can finally write for the mean separation,

$$\delta\boldsymbol{\mu} = -\sqrt{\frac{\gamma_{\mathrm{H}}\mathcal{T}}{2}} \cdot \sqrt{\frac{\Lambda}{\gamma}}\, \bar{\mathbf{J}}_b^{-1} \bar{\mathbf{H}}_b : [\mathbf{C}_b^{(l)} - \mathbf{C}_b^{(p)}] = -\sqrt{\frac{\gamma_{\mathrm{H}}\mathcal{T}}{2}} \cdot \sqrt{\frac{\Lambda}{\gamma}}\, \bar{\mathbf{J}}_b^{-1} \begin{pmatrix} \mathbf{H}_{b_1} : [\mathbf{C}_{b_1}^{(l)} - \mathbf{C}_{b_1}^{(p)}] \\ \mathbf{H}_{b_1^\dagger} : [\mathbf{C}_{b_1}^{(l)} - \mathbf{C}_{b_1}^{(p)}] \\ \vdots \\ \mathbf{H}_{b_K} : [\mathbf{C}_{b_K}^{(l)} - \mathbf{C}_{b_K}^{(p)}] \\ \mathbf{H}_{b_K^\dagger} : [\mathbf{C}_{b_K}^{(l)} - \mathbf{C}_{b_K}^{(p)}] \end{pmatrix} \tag{G14}$$

Of course Fisher's discriminant also requires us to obtain the noise properties of the measured quadratures in the steady-state. For simplicity, we consider the long-filter limit $\gamma\mathcal{T} \gg 1$ of measured covariances between QNP modes $\hat{b}_k$ and $\hat{b}_j$ to lowest order in $\frac{1}{\mathcal{T}}$ via Eq. (F28),

$$\boldsymbol{\Sigma}_{jk} \overset{\gamma\mathcal{T}\gg 1}{\simeq} \sigma_{\mathrm{vac}}^2 (\bar{n}_{\mathrm{cl}} + 1)\mathbf{I}_2 \delta_{jk} + \frac{1}{2}\mathbf{M}_j \left[\bar{\mathbf{J}}^{-1} \bar{\mathbf{B}} (\bar{\mathbf{J}}^T)^{-1} \right] \mathbf{M}_k^T \tag{G15}$$

Note that we are ultimately interested in the measured covariance matrix over all K modes of the QNP corresponding to the QS state l; this can be constructed from Eq. (G15) as

$$\boldsymbol{\Sigma}^{(l)} = \begin{pmatrix} \boldsymbol{\Sigma}_{11}^{(l)} & \cdots & \boldsymbol{\Sigma}_{1K}^{(l)} \\ \vdots & \ddots & \vdots \\ \boldsymbol{\Sigma}_{K1}^{(l)} & \cdots & \boldsymbol{\Sigma}_{KK}^{(l)} \end{pmatrix}, \tag{G16}$$

Recalling that the measurement matrices serve to extract the specific modes being observed, $\boldsymbol{\Sigma}^{(l)}$ will depend only on the QNP sector of the matrix $\left[\bar{\mathbf{J}}^{-1}\bar{\mathbf{B}}(\bar{\mathbf{J}}^T)^{-1} \right]$. We now proceed to simplify this term in particular, using the block form of the Jacobian matrix, Eq. (G4). More precisely, the inverse of the Jacobian matrix can be written explicitly in block form as:

$$\bar{\mathbf{J}}^{-1} = \begin{pmatrix} \mathbf{L}_a^{-1} & \mathbf{0} \\ \bar{\mathbf{J}}_b^{-1}\boldsymbol{\Gamma}\mathbf{L}_a^{-1} & \bar{\mathbf{J}}_b^{-1} \end{pmatrix} \tag{G17}$$

Making further use of the block form of $\bar{\mathbf{B}}$, Eq. (G5), we find:

$$\begin{aligned}
\bar{\mathbf{J}}^{-1}\bar{\mathbf{B}}(\bar{\mathbf{J}}^T)^{-1} &= \begin{pmatrix} \mathbf{L}_a^{-1} & \mathbf{0} \\ \bar{\mathbf{J}}_b^{-1}\boldsymbol{\Gamma}\mathbf{L}_a^{-1} & \bar{\mathbf{J}}_b^{-1} \end{pmatrix} \begin{pmatrix} \mathbf{D}_a & \mathbf{0} \\ \mathbf{0} & \bar{\mathbf{D}}_b \end{pmatrix} \begin{pmatrix} (\mathbf{L}_a^T)^{-1} & (\mathbf{L}_a^T)^{-1}\boldsymbol{\Gamma}^T(\bar{\mathbf{J}}_b^T)^{-1} \\ \mathbf{0} & (\bar{\mathbf{J}}_b^T)^{-1} \end{pmatrix} \\
&= \begin{pmatrix} \mathbf{L}_a^{-1}\mathbf{D}_a & \mathbf{0} \\ \bar{\mathbf{J}}_b^{-1}\boldsymbol{\Gamma}\mathbf{L}_a^{-1}\mathbf{D}_a & \bar{\mathbf{J}}_b^{-1}\bar{\mathbf{D}}_b \end{pmatrix} \begin{pmatrix} (\mathbf{L}_a^T)^{-1} & (\mathbf{L}_a^T)^{-1}\boldsymbol{\Gamma}^T(\bar{\mathbf{J}}_b^T)^{-1} \\ \mathbf{0} & (\bar{\mathbf{J}}_b^T)^{-1} \end{pmatrix} \\
&= \begin{pmatrix} \mathbf{L}_a^{-1}\mathbf{D}_a(\mathbf{L}_a^T)^{-1} & \mathbf{L}_a^{-1}\mathbf{D}_a(\mathbf{L}_a^T)^{-1}\boldsymbol{\Gamma}^T(\bar{\mathbf{J}}_b^T)^{-1} \\ \bar{\mathbf{J}}_b^{-1}\boldsymbol{\Gamma}\mathbf{L}_a^{-1}\mathbf{D}_a(\mathbf{L}_a^T)^{-1} & \bar{\mathbf{J}}_b^{-1}\bar{\mathbf{D}}_b(\bar{\mathbf{J}}_b^T)^{-1} + \bar{\mathbf{J}}_b^{-1}\boldsymbol{\Gamma}\mathbf{L}_a^{-1}\mathbf{D}_a(\mathbf{L}_a^T)^{-1}\boldsymbol{\Gamma}^T(\bar{\mathbf{J}}_b^T)^{-1} \end{pmatrix}
\end{aligned} \tag{G18}$$

The measurement matrices - encoding the fact that we are only monitoring QNP modes - will serve to extract only the lower-right block in the above matrix. We can therefore write Eq. (G16) explicitly as:

$$\boldsymbol{\Sigma}^{\gamma\mathcal{T} \gg 1}_{\simeq} \sigma_{\mathrm{vac}}^2(\bar{n}_{\mathrm{cl}}+1)\mathbf{I}_K + \sigma_{\mathrm{vac}}^2\frac{\gamma_{\mathrm{H}}}{\gamma}\mathbf{U}_K\left[\bar{\mathbf{J}}_b^{-1}\bar{\mathbf{D}}_b(\bar{\mathbf{J}}_b^T)^{-1} + \bar{\mathbf{J}}_b^{-1}\boldsymbol{\Gamma}\mathbf{L}_a^{-1}\mathbf{D}_a(\mathbf{L}_a^T)^{-1}\boldsymbol{\Gamma}^T(\bar{\mathbf{J}}_b^T)^{-1}\right]\mathbf{U}_K^T \tag{G19}$$

where the factor $\frac{\gamma_{\mathrm{H}}}{\gamma}$ arises due to the measurement matrices, and we have rewritten $\sigma_{\mathrm{vac}}^2 = \frac{1}{2}$. The above can be written to the form in Eq. (11) of the main text. This finally allows us to construct the averaged covariance matrix $\mathbf{V} = \frac{1}{2}\left(\boldsymbol{\Sigma}^{(l)} + \boldsymbol{\Sigma}^{(p)}\right)$.

Finally, substituting Eqs. (G14), (G19) into Eq. (5) of the main text, reproduced below for convenience:

$$D_{\mathrm{F}} = \delta\boldsymbol{\mu}^T \cdot \mathbf{V}^{-1} \cdot \delta\boldsymbol{\mu}, \tag{G20}$$

yields the semi-analytic expression for the Fisher's discriminant for quantum state classification under heterodyne measurement of QNP modes, to lowest nontrivial order in the nonlinearity Λ. This expression is used to calculate the approximate results labelled 'NVK' presented throughout the main text.

Appendix H: Characterizing noise in a measurement chain using a phase-preserving parametric amplifier

In Sec. IV of the main text, we analyze how nonlinear QNPs incorporated in quantum measurement chains enable quantum state classification schemes that are more robust to added classical noise power than comparable approaches using linear amplifiers and nonlinear post-processing. This analysis requires a calibration of the added classical noise power, which we present in this section using a standard quantum measurement chain with phase-preserving amplifiers.

For concreteness, we consider the measurement chain shown in Fig. 15(a), which consists of a single-port phase-preserving (i.e. non-degenerate) parametric amplifier, one mode of which is monitored via homodyne measurement. The measurement chain is described by the SME

$$d\hat{\rho}^c = -i[\hat{H}_d, \hat{\rho}^c]dt + \sum_n \gamma_{dn}\mathcal{D}[\hat{d}_n]\hat{\rho}^c dt + \sqrt{\gamma_{d1}}\mathcal{S}_{\text{hom}}[\hat{d}_1]\hat{\rho}^c \tag{H1}$$

where $\hat{H}_d$ is the Hamiltonian describing the phase-preserving amplifier,

$$\hat{H}_d = -\sum_n \Delta_{dn}\hat{d}_n^\dagger\hat{d}_n + \eta_1(-i\hat{d}_1 + i\hat{d}_1^\dagger) + G_{12}(-i\hat{d}_1\hat{d}_2 + h.c.), \tag{H2}$$

while the stochastic measurement superoperator for homodyne measurement is given by:

$$\mathcal{S}_{\text{hom}}[\hat{d}_1]\hat{\rho}^c = \left(\hat{d}_1\hat{\rho}^c + \hat{\rho}^c\hat{d}_1^\dagger - \langle\hat{d}_1 + \hat{d}_1^\dagger\rangle^c\right)dW_{\mathcal{I}_d}(t) \tag{H3}$$

For simplicity, we consider the case where $\Delta_{d1} = \Delta_{d2} = 0$, and $\gamma_{d1} = \gamma_{d2} \equiv \gamma_d$.

Note that being a linear system, the Hamiltonian in Eq. (H2) that governs the phase-preserving amplifier is simply a special case of the general linear QS we have analyzed in SI Sec. D. The correspondence simply requires setting $\kappa \to \gamma_d$, $\phi_{12} \to \frac{\pi}{2}$, and $\eta_2, G_1, G_2 \to 0$. This allows us to directly use solutions for the steady-state first and second-order cumulants derived earlier. Defining the corresponding vector of mode operators $\hat{\boldsymbol{d}} = (\hat{d}_1, \hat{d}_1^\dagger, \hat{d}_2, \hat{d}_2^\dagger)^T$, the solution for steady-state first-order cumulants is given by:

$$\langle\hat{\boldsymbol{d}}\rangle = -\mathbf{L}_d^{-1}\vec{\eta}$$
$$= \frac{1}{\gamma_d^2 - 4G_{12}^2}\begin{pmatrix} 2\gamma_d & 0 & 0 & 4G_{12} \\ 0 & 2\gamma_d & 4G_{12} & 0 \\ 0 & 4G_{12} & 2\gamma_d & 0 \\ 4G_{12} & 0 & 0 & 2\gamma_d \end{pmatrix}\begin{pmatrix} \eta_1 \\ \eta_1 \\ 0 \\ 0 \end{pmatrix} \tag{H4}$$

Similarly, the matrix of second-order cumulants is given by:

$$\mathbf{C}_d = \frac{1}{\gamma_d^2 - 4G_{12}^2}\begin{pmatrix} 0 & 2G_{12}^2 & \gamma_d G_{12} & 0 \\ 2G_{12}^2 & 0 & 0 & \gamma_d G_{12} \\ \gamma_d G_{12} & 0 & 0 & 2G_{12}^2 \\ 0 & \gamma_d G_{12} & 2G_{12}^2 & 0 \end{pmatrix} \tag{H5}$$

To characterize the noise in measured quadratures, we must compute the quantum statistics of the obtained measurements as described in SI Sec. F. However, the use of homodyne and not heterodyne detection means that expressions presented in that section for quantum statistics of measurements are modified, albeit in a simple way which we now explain.

1. Measured homodyne records and amplifier gain

For homodyne measurements described by the stochastic measurement superoperator in Eq. (H3), and using a single port used for both inputs and outputs, the obtained homodyne measurement records are given by:

$$\mathcal{I}_d(t) = \xi_{\mathcal{I}_d} + \sqrt{\gamma_d}\left[-\mathcal{I}_d^{\text{in}} + \sqrt{2}(\langle\hat{X}_{d_1}\rangle + \xi_{\mathcal{I}_d}^{\text{qm}})\right] + \sqrt{\bar{n}_{\text{cl}}}\xi_{\mathcal{I}_d}^{\text{cl}}, \tag{H6}$$

The white noise terms ξ are defined as in the main text (e.g. $\xi_{\mathcal{I}_d} = \frac{dW_{\mathcal{I}_d}}{dt}$), and $\mathcal{I}_d^{\text{in}} = 2\eta_1/\gamma_d$ defines the reflected part of the coherent input drive that also appears as part of the output since we are now considering monitoring in
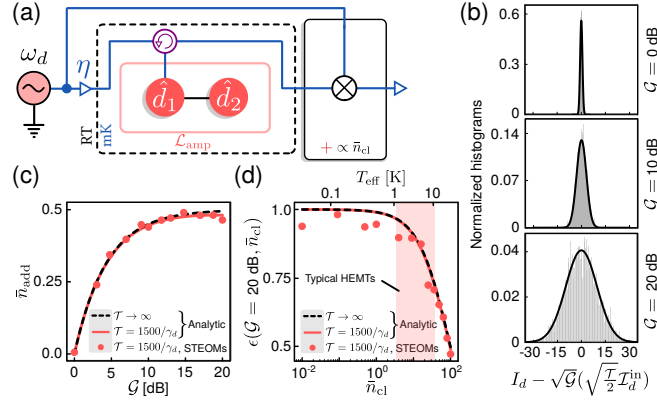
Figure 15. (a) Measurement chain for homodyne measurement of a phase-preserving amplifier. (b) Histograms of steady-state measured homodyne quadratures for different values of phase-preserving reflection gain $\mathcal{G}$, indicating increase in noise with gain. Here $\mathcal{T} = 1500/\gamma_d$. (c) Added noise power referred to the input, $\bar{n}_{\mathrm{add}}$, as a function of phase-preserving reflection gain $\mathcal{G}$, computed using both variances estimated using STEOM simulations (dots), as well as analytic calculations (solid and dashed lines) for different filter lengths. Here excess classical noise power $\bar{n}_{\mathrm{cl}} = 0$. (d) Quantum efficiency of the measurement chain as a function of excess classical noise power $\bar{n}_{\mathrm{cl}}$ and effective noise temperature T_{eff}. Ranges of these two quantities for typical measurement chains in cQED are illustrated by the red shaded region.

reflection. The multiplicative factor of $\sqrt{2}$ enhances the signal in comparison to heterodyne monitoring; this simply emphasizes that homodyne readout is able to provide a higher signal-to-noise ratio than heterodyne readout, at the expense of monitoring only a single quadrature. Measured quadratures defined by filtering homodyne measurement records are defined as before, $I_d(t) = \frac{1}{\sqrt{2\mathcal{T}}} \int_{t_0}^{t_0+\mathcal{T}} d\tau \, \mathcal{I}_d(\tau)$.

In comparison to the analysis in SI Sec. F, the difference in homodyne monitoring is simply that only a single quadrature of a given mode is monitored. This difference is easily accounted by defining a new *homodyne* measurement matrix for mode $\hat{d}_1$,

$$\mathbf{M}_{d_1} = \sqrt{\gamma_d} \begin{pmatrix} 1 & 1 & 0 & 0 \end{pmatrix} \tag{H7}$$

Clearly $\mathbf{M}_{d_1}$ is a 1-by-2R matrix for a system of R modes (here, $R = 2$), describing the extraction of a single measured quadrature from the quantum system.

The expectation value of measurement records is then given by an expression analogous to Eq. (F7), except using the homodyne measurement matrix,

$$\mu_d = \mathbb{E}[I_d] = \sqrt{\frac{\mathcal{T}}{2}} \left(-\sqrt{\gamma_d} \cdot \mathcal{I}_d^{\mathrm{in}} + \mathbf{M}_{d_1} \langle \hat{\boldsymbol{d}} \rangle \right) \tag{H8}$$

Using Eq. (H4), we therefore obtain the steady-state reflection gain $\mathcal{G}$ of the phase-preserving amplifier,

$$\sqrt{\mathcal{G}} \equiv \frac{\mu_d}{\sqrt{\frac{\mathcal{T}}{2}} \mathcal{I}_d^{\mathrm{in}}} = \frac{\gamma_d^2 + 4G_{12}^2}{\gamma_d^2 - 4G_{12}^2} \tag{H9}$$

2. Covariances of measured quadratures and added quantum noise

We can now calculate the variance of the measured quadrature, now for homodyne measurement, in which case the covariance "matrix" is now just a single number Σ_{d_1}, and is defined via the homodyne measurement matrix. For simplicity we consider the expression analogous to Eq. (F25), namely the $\mathcal{T} \to \infty$ limit often considered in the noise analysis of quantum amplifiers (referred to as calculating the noise at zero frequency). Then, to lowest order in $\frac{1}{\mathcal{T}}$,

$$\begin{aligned}
\Sigma_{d_1} &= \mathbb{E}[I_{d_1}^2] - (\mathbb{E}[I_{d_1}])^2 \\
&\stackrel{\mathcal{T} \to \infty}{=} \sigma_{\mathrm{vac}}^2 (\bar{n}_{\mathrm{cl}} + 1) - \frac{1}{2} \mathbf{M}_{d_1} \left[\mathbf{C}_d (\mathbf{L}_d^T)^{-1} + \mathbf{L}_d^{-1} \mathbf{C}_d \right] \mathbf{M}_{d_1}^T \\
&= \sigma_{\mathrm{vac}}^2 (\bar{n}_{\mathrm{cl}} + 1) + \gamma_d \left(\frac{4\gamma_d C_{d_1^\dagger d_1} + 4G_{12}(C_{d_1 d_2} + C_{d_1^\dagger d_2^\dagger})}{\gamma_d^2 - 4G_{12}^2} \right)
\end{aligned} \tag{H10}$$

where we have used Eqs. (H4), (H5) in obtaining the second line. We immediately note that the measured variance does not only depend on covariances of mode $\hat{d}_1$; it also contains contributions due to correlations between modes $\hat{d}_1$ and $\hat{d}_2$, which are nonzero when the modes are coupled via a non-degenerate squeezing interaction, necessary for phase-preserving amplification. This contribution is the source of added noise during phase-preserving amplification.

Using Eqs. (H5), together with the expression for $\mathcal{G}$ in Eq. (H9), we can simplify the expression for the measured variance further to:

$$
\begin{aligned}
\Sigma_{d_1} &= \sigma_{\text{vac}}^2(\bar{n}_{\text{cl}} + 1) + 2\sigma_{\text{vac}}^2(\mathcal{G} - 1) \\
&= \sigma_{\text{vac}}^2 \bar{n}_{\text{cl}} + \mathcal{G}\sigma_{\text{vac}}^2 + (\mathcal{G} - 1)\sigma_{\text{vac}}^2
\end{aligned}
\tag{H11}
$$

where we have rearranged the expression into a more standard form in the second line. When $\bar{n}_{\text{cl}} = 0$ and for unit gain, the measured quadrature variance simply reduces to the vacuum value, σ_{vac}^2. Contributions to the quadrature variance beyond this value arise from phase-preserving gain, as well as the classical contribution $\propto \bar{n}_{\text{cl}}$. In Fig. 15(b), we plot histograms of steady-state measured quadratures I_d with a filtering window $\mathcal{T} = 1500/\gamma_d$, for a selection of increasing values of gain $\mathcal{G}$. The histograms are centred by subtracting off the mean, which from Eq. (H9) is simply equal to the amplified coherent input signal, $\sqrt{\mathcal{G}}(\sqrt{\frac{\mathcal{T}}{2}}\mathcal{I}_d^{\text{in}})$. Moving from the top panel to the bottom panel, the increase in gain clearly leads to an increase in the added noise power, growing the variance of the measured amplified quadratures (indicated by the width of the histograms).

A standard metric to parameterize this added noise is by "referring it to the amplifier's input", namely normalizing the total noise by the phase-preserving gain and subtracting off the vacuum contribution to compute the total excess. The resulting added noise referred to the input $\bar{n}_{\text{add}}$ is then given by:

$$
\bar{n}_{\text{add}}(\mathcal{G}, \bar{n}_{\text{cl}}) \equiv \frac{\Sigma_{d_1}}{\mathcal{G}} - \sigma_{\text{vac}}^2 = \sigma_{\text{vac}}^2 \left[\frac{\bar{n}_{\text{cl}}}{\mathcal{G}} + \left(1 - \frac{1}{\mathcal{G}} \right) \right]
\tag{H12}
$$

In the absence of excess classical noise ($\bar{n}_{\text{cl}} = 0$) the above expression for $\bar{n}_{\text{add}}$ is plotted as a function of gain $\mathcal{G}$ in Fig. 15(c), together with its numerical calculation from measured quadratures obtained by integrating STEOMs of the phase-preserving amplifier. In the large $\mathcal{G}$ limit, $\bar{n}_{\text{add}}$ approaches $\sigma_{\text{vac}}^2 = \frac{1}{2}$, often referred to as the standard quantum limit of added noise for phase-preserving amplification: "half a photon of noise referred to the input".

3. Quantifying excess classical noise power

The general form of measured noise from a measurement chain with a phase-preserving amplifier enables us to characterize the excess classical noise power $\bar{n}_{\text{cl}}$ in a number of different ways. First, we note from Eq. (H11) that $\bar{n}_{\text{cl}}$ can be interpreted as excess noise power in units of vacuum state variance σ_{vac}^2. Hence $\bar{n}_{\text{cl}}$ units of classical noise added increases the variance of measured quadratures by $\bar{n}_{\text{cl}}$ times the vacuum noise power.

The units of excess classical noise are also commonly interpreted as a "photon" number, and converted to an effective temperature T_{eff} via the Bose-Einstein distribution at a representative system frequency Ω, namely $\bar{n}_{\text{cl}} \simeq \frac{k_b T_{\text{eff}}}{\hbar \Omega}$. Typical HEMT amplifiers have noise temperatures T_{eff} of around 1-10 K, which corresponds to an effective classical noise power $\bar{n}_{\text{cl}}$ of around 10-30 photons at microwave (GHz) frequencies.

The efficiency $\epsilon(\mathcal{G}, \bar{n}_{\text{cl}})$ of the measurement chain when incorporating a phase-preserving amplifier with gain $\mathcal{G}$ is defined as

$$
\epsilon(\mathcal{G}, \bar{n}_{\text{cl}}) = \frac{\bar{n}_{\text{add}}(\mathcal{G}, \bar{n}_{\text{cl}} = 0)}{\bar{n}_{\text{add}}(\mathcal{G}, \bar{n}_{\text{cl}})}
\tag{H13}
$$

Unit efficiency is achieved when the amount of noise added is the minimum required by quantum mechanics, $\bar{n}_{\text{add}}(\mathcal{G}, \bar{n}_{\text{cl}} = 0) \to \frac{1}{2}$ in the large gain limit. The quantum efficiency is plotted as a function of the classical noise power in Fig. 15(d) for a typical phase-preserving gain of $\mathcal{G} = 20$ dB. The dashed vertical line marks $\bar{n}_{\text{cl}} = 30$, which corresponds to a typical quantum efficiency of $\epsilon \simeq 0.77$, and to an added noise referred to the input $\bar{n}_{\text{add}} \simeq 0.65$. These are typical numbers for a modern measurement chain in cQED, and is what we use to label the typical excess classical noise powers $\bar{n}_{\text{cl}}$ and effective temperatures T_{eff} in the red shaded region of Fig. 5 in the main text.

Appendix I: Supplementary benchmarking master equation simulations: Multimode measurement chain

In SI Sec. C, we provided results benchmarking the performance of (S)TEOMs against full (S)ME simulations for a single coherently-driven QNP mode undergoing heterodyne measurement. In this section, we present results for a more complex measurement chain that includes the key features common to the classification tasks considered in the main text: the inclusion of a measured QS and its coupling to the QNP.

The specific measurement chain we consider is a simplified version of the full measurement chain employed for quantum state classification tasks in Fig. 3(a) of the main text, and is depicted in Fig. 16(a). More precisely, the QS is as defined in Eq. (12) of the main text but with $G_{12} = 0$ and $\Gamma_2 = 0$, thereby reducing the QS to a single mode $\hat{a}_1$. The rest of the measurement chain is unchanged: the mode $\hat{a}_1$ is coupled to the single QNP mode via a non-reciprocal hopping interaction Γ_1, and the QNP mode undergoes continuous heterodyne measurement. We wish to use this system to again classify two distinct states with equal means and distinct covariance, which we call Task IV; the states to be distinguished are indexed $l = 5$ and $l = 6$, and are defined by the QS parameters in Table V.

Table V. QS parameters ($M = 1$ only) defining the fourth classification task considered, in units of QS mode total loss rate κ.

	Task IV		
l	$G_1^{(l)}$	$\phi_1^{(l)}$	$\eta_1^{(l)}$
7	0.3	$+\frac{\pi}{2}$	$0.20\mathcal{A}$
8	0.3	$-\frac{\pi}{2}$	$0.80\mathcal{A}$

1. Individual quantum trajectories and measurement records

We begin by simulating an individual quantum trajectory for the modes of the measurement chain for the state $l = 5$ as defined in Table V. For fair benchmarking comparisons, we choose parameters where the QNP enables classification of the QS states; from results in the main text, we choose $\Delta_1/\gamma_{\rm H} = -\frac{2}{3}$ and $\mathcal{E}_{\rm eff} = 0.385$ (see Eq. 15). Unfortunately, increasing the number of modes immediately makes the SME simulations much more computationally taxing. To lower the modal occupation numbers, we keep $\mathcal{E}_{\rm eff}$ fixed while using a relatively strong nonlinearity strength of $\Lambda/\gamma = 0.067$, which requires choosing a low $\mathcal{A} \simeq 4.5$, corresponding to the drive strength $\eta_1/\gamma \simeq 0.6$. We note that this nonlinearity strength is about five times stronger than the largest nonlinearity used in Sec. C of the SI. We simulate trajectories for a total time $\gamma_1 t = 20.0$. All other parameters of the measurement chain are summarized in the caption of Fig. 16.

The complete measurement chain of $R = K + M = 2$ quantum modes is determined by $2R^2 + 3R = 14$ independent degrees of freedom: $\{\langle \hat{b}_1 \rangle^c, \langle \hat{a}_1 \rangle^c, C_{b_1 b_1}^c, C_{a_1 a_1}^c, C_{b_1 a_1}^c, C_{b_1^\dagger a_1}^c\}$ and their complex conjugates, as well as the real-valued cumulants $\{C_{b_1^\dagger b_1}^c, C_{a_1^\dagger a_1}^c\}$. We plot these quantities in Fig. 16(b)-(e) for a single measurement trajectory obtained using the STEOMs (white) and full SME simulations (color). We find that even though the nonlinearity is much stronger than those considered in the main text, there is in general good agreement between the methods. We also emphasize that for the same time step (here, $\gamma_1 \Delta t = 5e^{-5}$), SME simulations with a Hilbert space cutoff of 40 photons per quantum mode take about 37 hours to complete [96] for each quantum trajectory; the STEOMs require 4 minutes.

2. Mean displacement and projected noise for quantum state classification

We now consider the full classification task, Task IV, defined in Table V. In particular, we wish to evaluate the quantities that define Fisher's discriminant for classification using both the TEOMs approach and its analytic NVK approximation, as done in Fig. 4 of the main text; furthermore, we now wish to compare these results against exact ME and SME integration. Evaluating the required quantities of the mean separation and the projected noise as a function of QNP parameters such as nonlinearity strength requires several independent simulations, and hence places an even larger computational overhead. We therefore consider an even lower $\mathcal{A} = 3.5$ than in the previous subsection, to minimize modal occupation numbers. For (S)ME integration, we also consider a Hilbert space of 42 photons for $\hat{a}_1$ and 17 photons for mode $\hat{b}_1$.

The steady-state mean separation $||\delta\boldsymbol{\mu}||$ only requires single operator, equal-time expectation values and therefore an efficient exact evaluation scheme can integrate the ME directly. In Fig. 17(a), we plot $||\delta\boldsymbol{\mu}||$ using the NVK approximation, the TOEMs, as well as the full ME simulation as a function of nonlinearity strength over a large range. We note the existence of peaks in the mean separation where the QNP response increases. The TEOMs exhibit
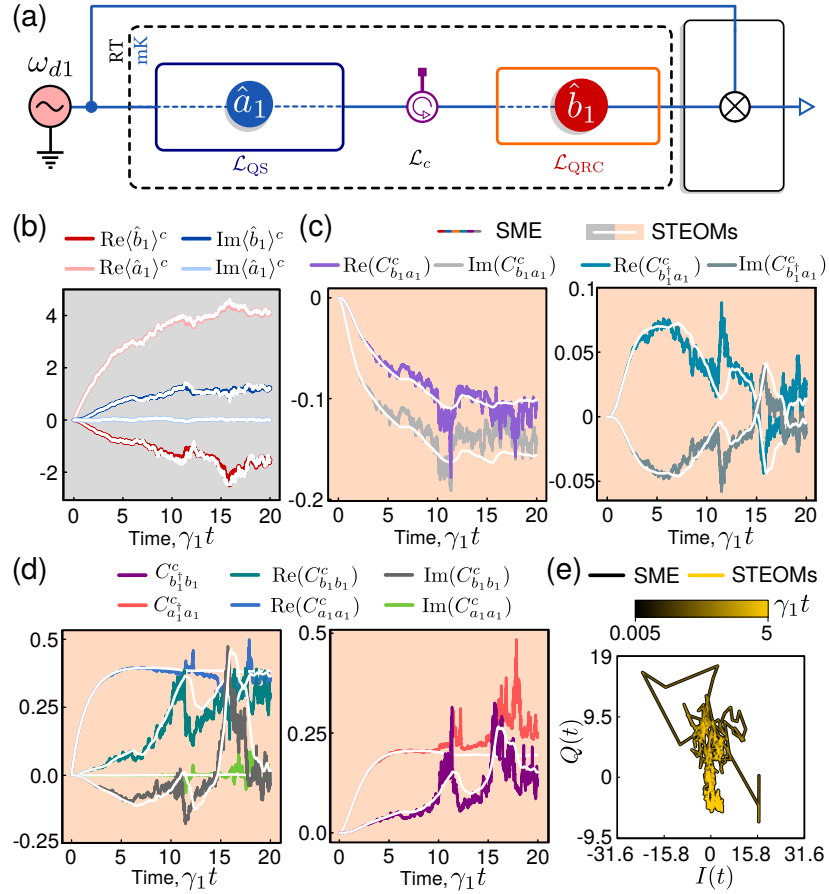
Figure 16. Benchmarking STEOMs against full SME simulations for the two-mode measurement chain depicted in (a). The single-mode measured quantum system is initialized in state $l = 5$ as defined in Table V, and with $\mathcal{A} \simeq 4.5$. We choose $\kappa_1 = \Gamma_1 = \frac{1}{2}\gamma_{\mathrm{H}}$, and $\gamma_{\mathrm{H}} = \frac{2}{3}\gamma$ (hence $\kappa = \kappa_1 + \Gamma_1 = \frac{2}{3}\gamma$). The QNP parameters are given by $\Delta_1/\gamma = -\frac{2}{3}$, and $\Lambda_1/\gamma = 0.067$ such that $\mathcal{E}_{\mathrm{eff}} = 0.385$ (see Eq. 15). Simulation results for a single measurement trajectory showing (b) first-order cumulants, (c) second-order cross-cumulants, (d) second-order self-cumulants, and (e) QNP measured quadratures $\{I(t), Q(t)\}$.
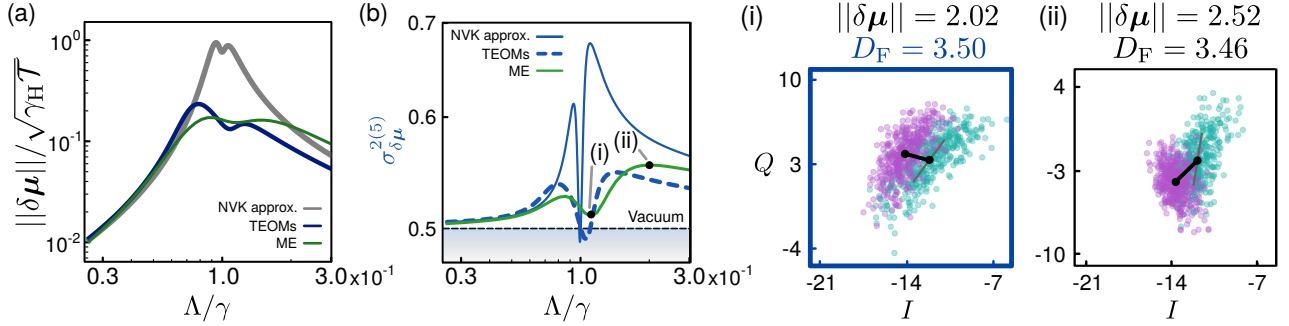

Figure 17. (a) Mean displacement $||\delta\boldsymbol{\mu}||$ and (b) Projected noise $\sigma_{\delta\boldsymbol{\mu}}^{2(7)}$ in the short-filter limit for Task IV, calculated using TEOMs, ME integration, as well as the NVK approximation as a function of nonlinearity $\Lambda/\gamma_{\mathrm{H}}$ for fixed detuning $\Delta_1/\gamma_{\mathrm{H}} = -\frac{2}{3}$. (c) Measured distributions obtained using SME simulations for nonlinearity values marked (i) and (ii) in (b).

the same qualitative behaviour, but with lower and broader peaks, due to the saturating effect of the nonlinearity, as also observed in Fig. 9. The ME simulations agree qualitatively with the TEOMs, although presenting slightly broader features. However, the most important feature of exhibiting a non-zero $||\delta\boldsymbol{\mu}||$ is robustly observed in the exact ME simulations, validating the result of the NVK approximation even at strong nonlinearities.

The calculation of projected noise $\sigma_{\delta\boldsymbol{\mu}}^{2(l)}$ is slightly more involved: depending on the length of the filter window $\mathcal{T}$

in comparison to the timescale of dynamics $\sim 1/\gamma$, the measured covariance matrix $\boldsymbol{\Sigma}^{(l)}$ will have contributions from correlations at different times. However, in the short filter limit, $\boldsymbol{\Sigma}^{(l)}$ and hence $\sigma^{2(l)}_{\delta\boldsymbol{\mu}}$ is defined entirely by equal time covariances (see Sec. F 3 of the SI), and can again be efficiently calculated using ME simulations. We plot $\sigma^{2(l)}_{\delta\boldsymbol{\mu}}$ for $l = 5$ in Fig. 17(b) using the NVK approximation, TEOMs, and ME simulations, assuming $\gamma\mathcal{T} = 0.1 \ll 1$. We again see qualitatively similar behaviour across all three methods: the projected noise achieves a minimum for a specific nonlinearity value while also exhibiting nonzero $||\delta\boldsymbol{\mu}||$. The projected noise can be manipulated by adjusting the QNP nonlinearity alone, again validating a key principle uncovered by the NVK approximation in this paper. We again observe the broadening of features via the TEOMs and ME approaches in comparison to the NVK approximation.

Finally, we would also like to evaluate the long-filter limit of readout, as $||\delta\boldsymbol{\mu}||$ scales directly with $\sqrt{\gamma\mathcal{T}}$. However, long filters probe the steady-state temporal correlation functions of the monitored modes (see Sec. F 4 of the SI). To evaluate via ME simulations, this requires the use of the quantum regression theorem and places a high demand on computational memory. An alternative approach is to integrate the SME to obtain measurement records, which provide the statistics of the filtered measured distributions directly as in the STEOMs approach. We plot measured distributions for the two marked nonlinearity values (i) and (ii) using 400 simulated measurement records per state, and show the results in Fig. 17(c). We note that here we use $t - \mathcal{T} = 0$ to maximize $||\delta\boldsymbol{\mu}||$ without having to simulate the SME for even longer, due to computational constraints. This leads to non-Gaussian signatures in transient dynamics appearing in the observed output, similar to Fig. 14 in SI Sec. C.

The value (i) is a local minimum of projected noise, according to the ME results in Fig. 17(b); we find that the position of the projected noise minimum is generally not changed by the filter window length. Here, we see that the two distributions obtained via SME simulations are almost parallel, so that the noise along the mean separation direction is close to minimal, consistent with the ME results. On the other hand, the value (ii) has a larger projected noise. We note that the measured distributions are no longer parallel. Crucially, we note that in (ii), the obtained $||\delta\boldsymbol{\mu}||$ is larger than in (i), by a factor of 1.25. If the projected noise was unchanged, this would lead to a larger value of Fisher's discriminant, by a factor of $(1.25)^2$. However, the obtained value of Fisher's discriminant D_{F} is in fact *smaller* than in (i). This is because the mean separation is not entirely along the minimum noise eigenvector. The SME simulations again validate the QRC's ability to coherently manipulate quantum fluctuations.

Appendix J: Complex-$\mathcal{P}$ representation of single coherently-driven Kerr oscillator

Recall that a single mode of our QNP model - a single coherently-driven Kerr oscillator coupled to a zero temperature bath - is defined by the master equation

$$\mathcal{L}\hat{\rho} = -i\left[-\Delta_1\hat{b}^\dagger\hat{b} - \frac{\Lambda_1}{2}\hat{b}^\dagger\hat{b}^\dagger\hat{b}\hat{b} + \eta_b(e^{-i\varphi_\eta}\hat{b} + e^{i\varphi_\eta}\hat{b}^\dagger), \hat{\rho}\right] + \gamma\mathcal{D}[\hat{b}]\hat{\rho}, \tag{J1}$$

where $\gamma = \gamma_1 + \Gamma_1$. For this system, it can be shown that the steady-state complex-$\mathcal{P}$ distribution $\mathcal{P}_{\mathrm{ss}}(\beta, \beta^\dagger)$, in the phase space of variables $\beta, \beta^\dagger$ associated with operators $\hat{b}, \hat{b}^\dagger$ respectively, can be found exactly by the method of potentials [91, 97]. To arrive at the desired result, we note first that the drive phase φ_η can simply be absorbed into the definition of the operators $\hat{b} \to \hat{b}e^{i\varphi_\eta}, \hat{b}^\dagger \to \hat{b}^\dagger e^{-i\varphi_\eta}$, while leaving their commutator unchanged. The drive amplitude can thus be chosen to be completely real. Following this transformation, the steady-state complex-$\mathcal{P}$ distribution takes the form

$$\mathcal{P}_{\mathrm{ss}}(\beta, \beta^\dagger) = (\beta)^{(c-2)}(\beta^\dagger)^{(c^*-2)}\exp\left\{\frac{2\eta_b}{\Lambda_1}\left(\frac{1}{\beta} + \frac{1}{\beta^\dagger}\right) + 2\beta^\dagger\beta\right\} \tag{J2}$$

where

$$c = \frac{-i\Delta_1 + \frac{\gamma}{2}}{-i\Lambda_1}. \tag{J3}$$

Knowledge of the exact steady-state complex $\mathcal{P}$-distribution allows one to calculate arbitrary moments of the quantum steady-state of the driven nonlinear mode by integrating over complex phase space. We find for arbitrary normal-ordered steady-state moments [91]

$$\langle(\hat{b}^\dagger)^j(\hat{b})^i\rangle = e^{-i\varphi_\eta(i-j)}\left|\frac{2\eta_b}{\Lambda_1}\right|^2\frac{\Gamma(c)\Gamma(c^*)}{\Gamma(c+i)\Gamma(c^*+j)}\frac{h(c+i, c^*+j, 8|\eta_b/\Lambda_1|^2)}{h(c, c^*, 8|\eta_b/\Lambda_1|^2)}, \tag{J4}$$

where we have reintroduced the drive phase by simply undoing the earlier transformation on operators $\hat{b}, \hat{b}^\dagger$. Here the function $h(x, y, z)$ is given by the hypergeometric series

$$h(x, y, z) = \sum_{n=0}^{\infty}\frac{z^n}{n!}\frac{\Gamma(x)\Gamma(y)}{\Gamma(x+n)\Gamma(y+n)} \tag{J5}$$

and $\Gamma(x)$ is the Gamma function.

Eq. (J4) is used to calculate exact steady-state first-order moments for the coherently-driven single-mode Kerr QNP in SI Sec. C.