

Single-shot Quantum State Classification via Nonlinear Quantum Amplification

Elif Cüce,¹ Saeed A. Khan,² Boris Mesits,^{3,4,5} Michael Hatridge,^{4,5} and Hakan E. Türeci¹

¹*Department of Electrical and Computer Engineering,
Princeton University, Princeton, NJ 08544, USA*

²*School of Applied and Engineering Physics, Cornell University, Ithaca, NY 14853, USA*

³*Department of Physics and Astronomy, University of Pittsburgh, Pittsburgh, PA 15260, USA*

⁴*Department of Applied Physics, Yale University, New Haven, CT 06511, USA*

⁵*Yale Quantum Institute, Yale University, New Haven, Connecticut 06520, USA*

(Dated: March 16, 2026)

Quantum amplifiers are intrinsically nonlinear systems whose performance limits are set by quantum mechanics. In quantum measurement, amplifier operation is conventionally optimized in the linear regime by maximizing signal-to-noise ratio, an objective that is well-suited to parameter estimation but is typically insufficient for more general tasks such as arbitrary quantum state discrimination. Here we show that single-shot quantum state classification can benefit from operating a quantum amplifier outside the linear regime, when the measurement chain is optimized end-to-end for a task-specific cost function. We analyze a realistic superconducting readout architecture that includes state preparation, cryogenic nonlinear amplification, and room-temperature detection with finite noise. By introducing performance metrics tailored to state discrimination, we identify operating regimes in which nonlinear amplification provides a measurable advantage and clarify the trade-offs that ultimately limit classification fidelity. Building on these results, we propose a qubit readout architecture without cavity displacement that exploits nonlinear amplification to enhance single-shot state discrimination performance. Our results establish the practical value of nonlinear quantum amplifiers for quantum state discrimination and lay the foundation for a broader program to develop a general, end-to-end framework for resource-constrained optimization of nonlinear amplification in quantum information processing tasks.

I. INTRODUCTION

Rapid, high-fidelity quantum nondemolition measurement requires amplification of an electromagnetic probe field near the limits set by quantum mechanics. In superconducting-qubit readout, where the probe lies in the microwave domain and the qubits are only weakly nonlinear, this is achieved by pre-amplifying the outgoing field with a cryogenic amplifier at base temperature, prior to further amplification and digital processing at room temperature. [1]. Although many cryogenic amplifiers are intrinsically nonlinear, they are conventionally operated in their linear regime, where the relevant quantum limits are well understood [2–6]. Can operating a cryogenic quantum amplifier in a nonlinear regime provide a fundamental performance advantage beyond what is achievable in linear, quantum-limited operation? We address this question through a hardware-realistic analysis of a class of quantum state–discrimination tasks in which operating an experimentally available parametric oscillator in a genuinely nonlinear regime is optimal. In contrast to the prevailing emphasis on linear, Kerr-mitigated amplification [7, 8], we identify operating regimes where nonlinear amplification yields a measurable classification advantage and quantify how nonlinear distortion, effective noise, and downstream added noise set the ultimate limits on classification fidelity.

In classical signal processing, nonlinear amplification has long been used as a deliberate functional element for tasks where linear gain is suboptimal. The underlying principle is that when the objective is not faithful

waveform reproduction but rather feature extraction, detection, or classification, nonlinear transformations can dramatically improve performance under resource constraints [9, 10]. By comparison, the deliberate use of nonlinear amplification for processing quantum-domain signals has received relatively little attention to date. Ref. [11] examined an optical amplifier based on a nonlinear interferometric setup. In microwave quantum circuits, non-linear dynamical transduction either provided by a resonant cavity [12] or by a Josephson junction tuned close to its ‘bifurcation point’ [13] has been explored to attain high-sensitivity and speed in qubit readout. A comprehensive analysis of the quantum limits of a broad class of nonlinear amplifiers was recently presented in Ref. [14], and stands out in the contemporary literature.

When the objective is rapid, high-fidelity measurement for quantum state discrimination rather than an analog task, quantum limits on amplification alone are not sufficient to determine performance. Additional considerations, including amplifier nonlinearity, demodulation strategy, integration time, and the noise incurred upon extraction of the signal to room temperature, play a critical role [4, 6, 15, 16]. Moreover, through a concrete example, we show that the optimal strategy depends explicitly on the measurement objective and, more generally, on the specific cost function being maximized.

The purpose of this article is to present a systematic study of the quantum-limited operating regimes of a specific class of nonlinear amplifiers, with an eye on end-to-end resource-aware optimization of a task-relevant cost-function. We highlight the central role of task-specific optimization by focusing on a quantum state discrimi-

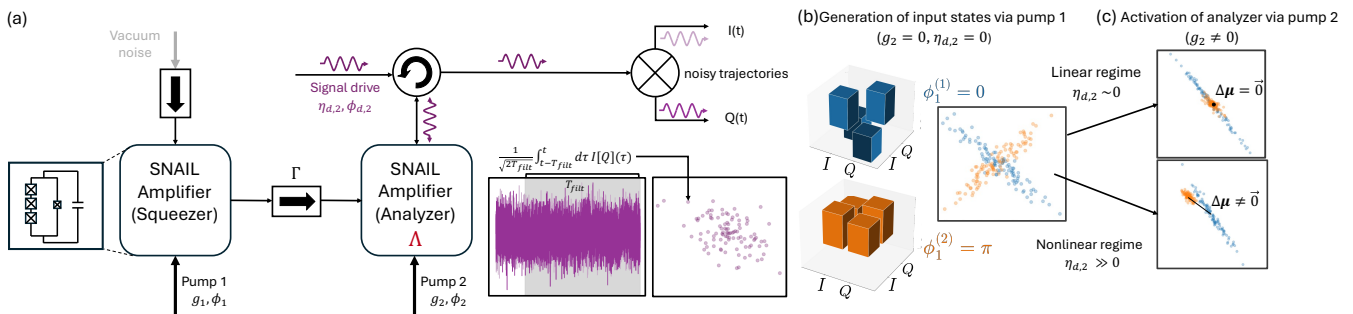


Figure 1. (a) The composite two-SNAIL system for state generation and discrimination. The pump applied to the *squeezer* (pump 1) sets which input state is generated, (1-blue) or (2-orange). Two additional drives at the pump frequency ω_p (pump 2) and at $\omega_p/2$ (signal drive with strength $\eta_{d,2}$) act on the *analyzer* and impact the overall classification performance. The two SNAIL amplifiers are identical except that the analyzer has a non-zero Kerr nonlinearity (Δ) (b) Intracavity quadrature covariance matrix elements and quadrature measurements for two input states with the analyzer pump off ($g_2 = 0$). (c) When the analyzer is on, the output-state means coincide for linear operation ($\eta_{d,2} \approx 0$) and the classification fidelity is 0.84, but are separated for nonlinear operation with a sufficiently-large signal drive ($\eta_{d,2} \gg 0$) and the classification fidelity is 1 (this is the same operating point as in Fig. 2(d-2)).

nation problem that is poorly served by amplifiers optimized for linear, quantum-limited operation followed by linear detection and post-processing. Our analysis identifies the operating regimes in which nonlinear amplification yields a measurable performance advantage and clarifies the trade-offs that ultimately set the limits on achievable measurement fidelity.

Our work is motivated by recent proposals in quantum computational sensing [17], which articulate a framework for formulating quantum measurement and sensing tasks in close analogy with conventional computational sensing and imaging approaches. This line of work takes the perspective that the detection of weak signals is often performed with the ultimate aim of performing a specific computation on the signal, in service of completing a given task. Consequently, weak-signal detection or sensing can be combined with processing by a suitably-matched quantum mechanical system prior to detection. This line of research fuses concepts from quantum sensing and quantum machine learning, and has demonstrated that the resulting quantum computational sensors can provide advantages for tasks such as classification [18–21] and function approximation [17, 22] in comparison to sensing schemes alone. For example, within this framework, circuit-based quantum computational sensors have been shown to approach the fundamental performance bounds imposed by quantum mechanics for a broad class of sensing tasks [23].

The present work lies in the same domain, but differs in scope and methodology. Firstly, rather than adopting a gate-based and lossless quantum computational sensor, we consider a continuous-variable, dynamical, driven-dissipative sensor. Prior work has shown that nonlinear resources available in such cryogenic sensors can be used for certain quantum information processing applications [24]. Here, we consider a sensor that is closely-modeled on practical quantum amplifiers, and

therefore features amplification and dissipation as key components. Using a model-based approach in which the full measurement chain, including signal generation, propagation, amplification, and detection, is described numerically, we assess performance and identify optimal operating regimes for such a device for quantum state discrimination tasks under constrained resources.

We next examine the optimal operating conditions for nonlinear amplification in dispersive qubit readout without cavity displacement, and present a detailed analysis of an architecture incorporating these findings. We show that the optimal operating point is architecture-dependent and, in this setting, lies in the regime $\chi < \kappa/2$, where χ is the qubit-induced dispersive shift and κ is the resonator linewidth. This stands in contrast to conventional dispersive readout with linear, phase-preserving amplification, where maximizing measurement rate and SNR typically favors operation in or near the strong-dispersive regime, $\chi = \kappa/2$, for fixed photon number and coherence constraints.

The rest of this paper is organized as follows. In Sec. II, we describe the mathematical model of the composite system that describes both the generation of states to be discriminated and the measurement chain including the nonlinear amplifier. In Sec. III, we define metrics to quantify the classification performance and optimize the externally tunable parameters of the system. Next, in Sec. IV, we analyze the effect of the noise from the electronic amplification on the performance metrics. Finally, in Sec. V, we propose a scheme to perform dispersive qubit readout in a regime without cavity displacement, and find the optimal encoding parameters for this scheme. In Sec. VI, we summarize our conclusions and outline directions for future work.

II. DESCRIPTION OF THE COMPOSITE SYSTEM AND THE CLASSIFICATION TASK

We consider the task of discriminating two states of a resonator by amplifying the radiation that leaks out of the resonator with an amplifier that intrinsically is nonlinear. To this end, we introduce a suitable objective function of the single-shot sampled output of the amplifier that we require to be maximized. We consider maximization with respect to parameters of the amplifier that can be parametrically controlled during the experiment.

For a single shot, there is a fundamental physical limit to the expected maximum of the objective function that is set by the quantum theory, regardless of whether the amplifier is operated in the linear or the nonlinear regime. We are interested in the optimal achievable objective function within a given class of amplifiers. We note that the achievable maximum heavily depends on the objective function itself and the states to be discriminated. In the case of dispersive qubit readout, the task can be framed as a quantum state discrimination task where two distinct pointer states of a readout resonator are generated conditional on the two possible states of the qubit. Here the physical limits are well understood when the downstream amplifier and the subsequent detector is linear [2–5, 15, 16].

In this work, we address a more challenging problem in which the two pointer states are resonator states with identical mean amplitudes and are distinguished solely by their second moments. We do not specify a particular physical mechanism for generating these conditional pointer states, and focus on a regime in which quantum back-action on the qubit, outside the scope of the present study, is expected to be minimal. Specifically, we consider the limit in which the mean fields of the pointer states vanish, so that the influence of the qubit can be neglected and the problem reduces to the discrimination of two zero-mean resonator states differing only in their fluctuations. Earlier work [14] has shown that an ideal nonlinear amplifier in principle allows one to measure any normal operator (of its input) with a linear detector while adding a half-quantum of vacuum fluctuations as noise at the output, but what types of measurements can be enabled by common experimentally realizable nonlinearities was not analyzed. Here, for the first time to our knowledge, we consider the end-to-end optimization of an experimentally realizable superconducting readout chain, including state preparation, directional coupling to a cryogenic nonlinear amplifier, and the noise floor of a subsequent room-temperature detection stage. This optimization requires the quantum dynamical modeling of the entire measurement chain and the sampling of measurement-conditioned trajectories observed at the linear detector, which we analyze using a recently developed cumulant expansion technique [24, 25]. We derive analytical performance metrics that guide our understanding of the behavior of the amplifier in certain well-understood operational limits. We then analyze the

performance across an experimentally relevant parameter space and identify optimal, in situ tunable operating points. Finally, we provide intuitive physical reasoning for these optimal choices, highlighting the properties that underlie the classification mechanism.

A schematic of the complete measurement chain, including preparation of the two pointer states and the nonlinear amplification stage, is shown in Fig. 1. The setup employs two Superconducting Nonlinear Asymmetric Inductive Element (SNAIL) oscillators, which serve as the fundamental building blocks of state-of-the-art cryogenic microwave amplifiers [7]. The first SNAIL oscillator, referred to as the *squeezer*, is used to generate two distinct pointer states, which are squeezed vacuum states with orthogonal squeezing directions. These states are subsequently discriminated by the second SNAIL oscillator, referred to as the *analyzer*. Both SNAILs are identical in that they share the same bare resonance frequency and the same third- and fourth-order (Kerr) nonlinearity strengths (g_3 and g_4 in Eq. (2)). The difference between the two lies in their operating conditions. The *analyzer* SNAIL is driven coherently at the signal port, which displaces the mode and leads to an effective contribution of the Kerr term to its dynamics. In contrast, the *squeezer* SNAIL is not subject to a coherent signal drive, and the Kerr term does not contribute appreciably to its dynamics. We therefore neglect the Kerr term for the *squeezer* SNAIL to simplify the model. We note that this set up has a simple linear limit for the *analyzer*, achieved by either turning off the additional coherent drive or setting its nonlinearity $\Lambda = 0$, when the analyzer becomes a phase-sensitive linear amplifier. In our numerical analysis, we model all essential elements of a typical superconducting measurement chain, including directional propagation between the *squeezer* and *analyzer* SNAILs, the demodulation at the mixer and the integration of the down-converted measurement trajectory at the room temperature.

A. The Model

We describe the measurement-conditioned evolution of the system with the following stochastic quantum master equation (SQME), assuming that the *analyzer* mode undergoes continuous linear measurement:

$$\dot{\hat{\rho}}^c = \mathcal{L}\hat{\rho}^c + \mathcal{S}_{an}\hat{\rho}^c. \quad (1)$$

The Liouvillian of the system, which we describe in detail in this section, is derived from a general model of two SNAILs, representing the *squeezer* and the *analyzer*, that are unidirectionally coupled. The Hamiltonian of a single SNAIL oscillator representing either the *squeezer* or the *analyzer*, with resonance frequency ω_s , incorporating both third- and fourth-order nonlinearities and driven simultaneously with two tones at frequencies ω_p

and $\omega_p/2 = \omega_s$, is given by

$$\begin{aligned} \hat{\mathcal{H}}_{SNAIL} = & \omega_s \hat{a}^\dagger \hat{a} + g_3 (\hat{a} + \hat{a}^\dagger)^3 + g_4 (\hat{a} + \hat{a}^\dagger)^4 \\ & + \sqrt{\kappa_s} \epsilon_p (e^{-i\phi_p} e^{-i\omega_p t} + e^{i\phi_p} e^{i\omega_p t}) (\hat{a} + \hat{a}^\dagger) \\ & + \sqrt{\kappa_s} \eta_{sig} (e^{-i\phi_{sig}} e^{-i\frac{\omega_p}{2} t} + e^{i\phi_{sig}} e^{i\frac{\omega_p}{2} t}) (\hat{a} + \hat{a}^\dagger). \end{aligned} \quad (2)$$

Here, the first line includes the bare resonance term, third- and fourth-order nonlinearities of the SNAIL that are non-zero. The second and third lines represent the pump drive and signal drive, respectively. We assume that the *squeezer* has zero Kerr nonlinearity ($g_4 = 0$) and is not driven at resonance frequency ($\eta_{sig} = 0$). Due to the presence of the pump and the signal drive on the SNAIL, there will be non-zero population at ω_p (pump mode) and $\omega_p/2$ (resonant mode). To account for both modes, we introduce the following multimode field ansatz:

$$\hat{a} = \hat{s} e^{i\frac{\omega_p}{2} t} + \hat{p} e^{i\omega_p t}, \quad (3)$$

where \hat{s} represents the signal mode and \hat{p} represents the pump mode. As long as the dynamics of the new operators have bandwidth much smaller than $\omega_p/2$, it is a good approximation to assume \hat{s} and \hat{p} represent independent dynamical degrees of freedom [26, 27]. After plugging in this ansatz and assuming the pump is stiff, we end up with the Liouvillian that we introduce in this section. More details of this derivation is in Appendix A. The Liouvillian \mathcal{L} of the system is composed of three components pertaining to the *squeezer*, nonreciprocal coupler and the *analyzer*:

$$\mathcal{L} = \mathcal{L}_{sq} + \mathcal{L}_{an} + \mathcal{L}_{coup}. \quad (4)$$

\mathcal{L}_{sq} describes the *squeezer*,

$$\mathcal{L}_{sq} \hat{\rho}^c = -i \left[\hat{\mathcal{H}}_{sq}, \hat{\rho}^c \right] + \kappa_1 \mathcal{D}[\hat{s}_1] \hat{\rho}^c, \quad (5)$$

with

$$\hat{\mathcal{H}}_{sq} = -\Delta_1 \hat{s}_1^\dagger \hat{s}_1 + \frac{g_1}{2} (e^{-i\phi_1} \hat{s}_1 \hat{s}_1 + h.c.). \quad (6)$$

The Hamiltonian for the *squeezer* is expressed in a frame rotating at $\omega_p/2$, where ω_p is the pump tone frequency. This is the frame in which the squeezing terms are resonant. We assume the system is being pumped on resonance for amplification ($\omega_1 = \omega_2 = \omega_p/2$), which makes $\Delta_1 = 0$. We assume that the pump is stiff and characterized with real-valued constants g_1 and ϕ_1 for the strength and phase respectively of this stiff pump. We verify the validity of this approximation within the operating regime considered in the paper by comparing it to simulations performed without the stiff-pump approximation. Further, \mathcal{L}_{an} describes the *analyzer*,

$$\mathcal{L}_{an} \hat{\rho}^c = -i \left[\hat{\mathcal{H}}_{an}, \hat{\rho}^c \right] + \kappa_2 \mathcal{D}[\hat{s}_2] \hat{\rho}^c, \quad (7)$$

with

$$\begin{aligned} \hat{\mathcal{H}}_{an} = & -\Delta_2 \hat{s}_2^\dagger \hat{s}_2 + \frac{g_2}{2} (e^{-i\phi_2} \hat{s}_2 \hat{s}_2 + h.c.) - \frac{\Lambda}{2} \hat{s}_2^\dagger \hat{s}_2^2 \\ & + \sqrt{\kappa_2} \eta_{d,2} (i e^{-i\phi_{d,2}} \hat{s}_2 e^{-i(\omega_{d,2} - \omega_p/2)t} + h.c.). \end{aligned} \quad (8)$$

The Hamiltonian for the *analyzer* is also expressed in the frame rotating at $\omega_p/2$ for the same reason. With this choice and the assumption $\omega_2 = \omega_p/2$, Δ_2 only includes the cross-Kerr shift due to the pump. We present this calculation in Appendix A. Again, we assume that the *analyzer* pump is stiff, and is characterized by constants g_2 and ϕ_2 for its strength and phase. Different from the *squeezer*, $\hat{\mathcal{H}}_{an}$ contains the term with coefficient Λ that corresponds to the intrinsic Kerr nonlinearity of the SNAIL device. In the context of parametric amplifiers this fourth-order nonlinearity generally acts as a parasitic effect that limits dynamic range and leads to gain compression and unwanted frequency shifts, and is therefore typically treated as an imperfection to be minimized in linear-amplification regimes in order to improve dynamic range and approach quantum-limited performance [7, 8, 28]. In Appendix A we present a derivation of this effective SQME starting with the two-SNAIL Hamiltonian, the conditions under which this model accurately describes the dynamics of the system, as well as the relationship of the model parameters to the original physical parameters of the SNAILS. The final term represents the signal drive on the *analyzer*. We choose a resonant drive on the analyzer mode $\omega_{d,2} = \omega_2 = \omega_p/2$.

The coupling interaction with rate Γ is described by the coupling Liouvillian \mathcal{L}_{coup} ,

$$\mathcal{L}_{coup} \hat{\rho}^c = -i \left[\left(i \frac{\Gamma}{2} \hat{s}_1^\dagger \hat{s}_2 + h.c. \right), \hat{\rho}^c \right] + \Gamma \mathcal{D}[\hat{s}_1 + \hat{s}_2] \hat{\rho}^c. \quad (9)$$

This term balances a coherent hopping interaction with a dissipative hopping interaction to ensure nonreciprocal transmission from the *squeezer* mode \hat{s}_1 to the *analyzer* mode \hat{s}_2 only [24, 29].

Lastly, the measurement operator models a weak measurement of only the *analyzer* mode (\hat{s}_2) as described by the measurement operator \mathcal{S}_{an} [30]:

$$\begin{aligned} \mathcal{S}_{an} \hat{\rho}^c = & \sqrt{\frac{\kappa_2}{2}} \left(\hat{s}_2 \hat{\rho}^c + \hat{\rho}^c \hat{s}_2^\dagger - \langle \hat{s}_2 + \hat{s}_2^\dagger \rangle \frac{dW_I}{dt} \right) \\ & + \sqrt{\frac{\kappa_s}{2}} \left(-i \hat{s}_2 \hat{\rho}^c + i \hat{\rho}^c \hat{s}_2^\dagger - i \langle \hat{s}_2 - \hat{s}_2^\dagger \rangle \right) \frac{dW_Q}{dt}. \end{aligned} \quad (10)$$

Here, $dW_{I,Q} = dW_{I,Q}(t)$ are independent Wiener increments, that describe the backaction on the system dynamics conditioned on the homodyne I-Q quadrature measurements of the *analyzer* mode. We include the measurement terms in the description of the system since we want to account for finite measurement resources available in an experimental setting such as a finite integration time in a single run of the experiment.

B. System dynamics and measurement

The numerical solution of Eq. (1) necessitates a truncation scheme that remains accurate across widely disparate oscillator occupations within the measurement chain. In the regime considered here, a Fock-space truncation is impractical due to resource constraints, as the *analyzer* can reach large occupations for certain parameter regimes analyzed. Here we employ a cumulant-based approach, which we have extensively benchmarked and found to remain accurate over the range of nonlinearities and dissipation strengths relevant to this work. Cumulant- (or cluster-) based moment closures are a standard tool in quantum optics for reducing the infinite hierarchy of operator moments to a finite, tractable set by discarding connected correlators above a chosen order [31]. When higher-order connected correlations remain parametrically small e.g., in near-Gaussian, weakly correlated, or mean-field-like regimes, this truncation provides a controlled and numerically stable approximation that has been extensively benchmarked across representative light-matter and nonlinear optical settings [31]. Building on this foundation, we employ a truncated-cumulant scheme recently adapted to efficiently sample stochastic measurement trajectories produced by nonlinear measurement chains in Refs. [24, 25].

In this section, we present the equations of motion for the lowest-order operators to illustrate the approach and to discuss important details of the numerical pipeline used to generate the I - Q plots (such as e.g. Fig. 1(b,c)) employed throughout the analysis. For details, see Appendix B.

The equation of motion of the *analyzer* mode expectation value conditioned on the measurement $\langle \hat{s}_2 \rangle^c$ is given by:

$$\begin{aligned}
 d\langle \hat{s}_2 \rangle^c = dt & \left\{ -\Gamma_1 \langle \hat{s}_1 \rangle^c + \left(i\Delta_2 - \frac{\kappa_2 + \Gamma_1}{2} \right) \langle \hat{s}_2 \rangle^c \right. \\
 & - i g_2 e^{i\phi_2} \langle \hat{s}_2^\dagger \rangle^c - \sqrt{\kappa_2} \eta_{d,2} e^{i\phi_{d,2}} \\
 & \left. + i\Lambda \left[2C_{s_2^\dagger s_2}^c \langle \hat{s}_2 \rangle^c + \langle \hat{s}_2^\dagger \rangle^c \left(\langle \hat{s}_2 \rangle^{c^2} + C_{s_2 s_2}^c \right) \right] \right\} \\
 & + \sqrt{\frac{\kappa_2}{2}} \left(C_{s_2 s_2}^c + C_{s_2^\dagger s_2}^c \right) dW_I \\
 & - i \sqrt{\frac{\kappa_2}{2}} \left(C_{s_2 s_2}^c - C_{s_2^\dagger s_2}^c \right) dW_Q,
 \end{aligned} \tag{11}$$

where $C_{o_1 o_2}^c = \langle \hat{o}_1 \hat{o}_2 \rangle^c - \langle \hat{o}_1 \rangle^c \langle \hat{o}_2 \rangle^c$ are the conditional second-order cumulants and $\langle \cdot \rangle^c$ represents the conditional average, due to the presence of the measurement operator in the SQME.

The Heisenberg equations of motion, a generally infinite hierarchy of *linear* equations expressed in moments, when reorganized in cumulants, becomes a set of coupled nonlinear differential equations. The non-zero Kerr nonlinearity of the *analyzer* mode manifests itself in the

appearance of nonlinear deterministic terms at lowest order. If $\Lambda = 0$, noise-averaged first order equations decouple from equations of motion of second-order cumulants (see Appendix B Eqs. (B6a-B6d)). Note however that the conditional evolution has noise terms that couple to second order cumulants and generate what would be considered multiplicative noise (see Appendix B Eqs. B5). We will show that the residual nonlinearity of the amplifier, as captured here, enables the transduction of information encoded in higher-order features of the measured quantum system, specifically the information contained in the second-order correlations of the two pointer states of the *squeezer*. The versatility of STEOM is critical as we are interested in a regime where the *analyzer* mode is sufficiently populated for the Kerr nonlinearity to be effective.

Once the complete set of nonlinear STEOMs are solved to a certain order, the I - Q plots typically measured in an experiment are obtained after demodulation at $\omega_p/2$ and integration with a box-car filter. First step in this procedure is the generation of quantum trajectories of the outgoing *analyzer* field which represent the noisy measurement acquired after demodulation:

$$I(t) = \xi_I(t) + \sqrt{\frac{\kappa_2}{2}} \langle \hat{s}_2 + \hat{s}_2^\dagger \rangle^c + \sqrt{\bar{n}_{cl}} \xi_I^{cl}(t), \tag{12a}$$

$$Q(t) = \xi_Q(t) - i \sqrt{\frac{\kappa_2}{2}} \langle \hat{s}_2 - \hat{s}_2^\dagger \rangle^c + \sqrt{\bar{n}_{cl}} \xi_Q^{cl}(t), \tag{12b}$$

where $\xi_{I,Q}(t) = \frac{dW_{I,Q}}{dt}$, with $dW_{I,Q}$ as in Eq. (10), and $\xi_{I,Q}^{cl}(t)$ are independent white noise processes modelling electronic amplification noise with strength \bar{n}_{cl} . Moreover, we define intracavity quadrature variables $\hat{I}_2 = (\hat{s}_2 + \hat{s}_2^\dagger) / \sqrt{2}$ and $\hat{Q}_2 = -i(\hat{s}_2 - \hat{s}_2^\dagger) / \sqrt{2}$ for later use.

The noisy trajectories in Eq. (12) are generally filtered over a filtering period \mathcal{T} , to obtain single-shot measurement results:

$$I = \frac{1}{\sqrt{2\mathcal{T}}} \int_{t-\mathcal{T}}^t d\tau \mathcal{I}(\tau), \tag{13a}$$

$$Q = \frac{1}{\sqrt{2\mathcal{T}}} \int_{t-\mathcal{T}}^t d\tau \mathcal{Q}(\tau). \tag{13b}$$

The scaling with \mathcal{T} is chosen here to keep variance independent of \mathcal{T} in the long filtering time limit, up to corrections that scale as $\mathcal{O}(\mathcal{T}^{-1})$, whereas the mean value of the filtered quadrature scales with $\sqrt{\mathcal{T}}$. See Appendix C for more details on the filtered single-shot measurement results.

C. Generation of squeezed vacuum states for the example task

The *squeezer* generates the squeezed vacuum states to be classified. Its output state is controlled by the pump

tone: the pump amplitude sets the degree of squeezing, while the pump phase ϕ_1 fixes the squeezing orientation in phase space (see Fig. 1). These properties can be determined analytically by deactivating the *analyzer* in the model, i.e., by setting $g_2 = 0$, $\eta_{d,2} = 0$ and $\Lambda = 0$, so that the measured output of the *analyzer* is a proxy for the measured statistical distribution of the *squeezer* state. In this limit the entire system is linear, allowing an exact derivation of the quadrature covariance matrix that is measured at the output of the *analyzer*. The squeezing axis in the measured I - Q plane can be shown to be given by $\phi_{\text{sq}} = \frac{\pi}{4} - \frac{\phi_1}{2}$, where the I quadrature defines the x -axis and the Q quadrature the y -axis, with I and Q defined as in Eq. (13).

For the classification task, we choose the same pump strength $g_1 = 0.8g_1^{\text{th}}$, corresponding to operation at 80% of the instability threshold. This ensures that the two classes of quantum states exhibit identical squeezing strength. We set the phase of the pump tone $\phi_1^{(1)} = 0$ and $\phi_1^{(2)} = \pi$ to generate type-1 and type-2 states, respectively, such that the corresponding squeezed quadratures are orthogonal. A representative sampling of the measured outcomes in the I - Q plane for these two parameter settings is shown in Fig. 1(b), yielding quadrature angles $\pi/4$ and $-\pi/4$, as designed. The task is to distinguish these two squeezed states based only on the direction (or phase) of squeezing, and not the amount of squeezing. The simplicity of this example allows us to build intuition for how a quantum nonlinear amplifier performs the task and how its operating parameters can be optimized to maximize classification performance.

III. OPTIMAL SYSTEM PARAMETERS FOR THE CLASSIFICATION TASK

A. Performance metrics

We first introduce two metrics to quantify classification performance based on the measured quadrature distributions (e.g., Fig. 1(c)). The first performance metric is the mean separation between the distributions corresponding to the two types of states on the measured I - Q quadrature plane ($\Delta\boldsymbol{\mu}$):

$$\Delta\boldsymbol{\mu} = \begin{bmatrix} \Delta I \\ \Delta Q \end{bmatrix} = \begin{bmatrix} \mathbb{E}[I^{(1)}] - \mathbb{E}[I^{(2)}] \\ \mathbb{E}[Q^{(1)}] - \mathbb{E}[Q^{(2)}] \end{bmatrix}, \quad (14)$$

where I and Q are defined in Eq. (13), and $\|\Delta\boldsymbol{\mu}\|$ is the ℓ^2 -norm of $\Delta\boldsymbol{\mu}$

$$\|\Delta\boldsymbol{\mu}\| = \sqrt{\Delta I^2 + \Delta Q^2}, \quad (15)$$

For the classification of weak signals such as the ones we consider here, where the noise in the quadrature distributions is comparable to the mean separation, $\|\Delta\boldsymbol{\mu}\|$ alone may not be an accurate proxy of empirical classification performance under finite sampling. In this case, which

we will encounter for our particular task here, a more accurate metric of empirical classification performance is given by the Fisher discriminant, D_F [32]. It is defined as follows:

$$D_F = \Delta\boldsymbol{\mu}^T \cdot \mathbf{V}^{-1} \cdot \Delta\boldsymbol{\mu}. \quad (16)$$

This metric also represents a natural multivariate generalization of the one-dimensional signal-to-noise ratio. The spread of the two distributions is captured by the combined variance of output fields of different types of states (\mathbf{V}),

$$\mathbf{V} = \frac{1}{2} \left(\boldsymbol{\Sigma}_{I,Q}^{(1)} + \boldsymbol{\Sigma}_{I,Q}^{(2)} \right), \quad (17)$$

with

$$\boldsymbol{\Sigma}_{I,Q}^{(i)} = \begin{bmatrix} \sigma_{I^{(i)}}^2 & \sigma_{I,Q^{(i)}} \\ \sigma_{I,Q^{(i)}} & \sigma_{Q^{(i)}}^2 \end{bmatrix}, \quad (18)$$

where

$$\sigma_{I^{(i)}}^2 = \mathbb{E}[I^{(i)2}] - \mathbb{E}[I^{(i)}]^2, \quad (19a)$$

$$\sigma_{I,Q^{(i)}}^2 = \mathbb{E}[I^{(i)}Q^{(i)}] - \mathbb{E}[I^{(i)}]\mathbb{E}[Q^{(i)}], \quad (19b)$$

$$\sigma_{Q^{(i)}}^2 = \mathbb{E}[Q^{(i)2}] - \mathbb{E}[Q^{(i)}]^2. \quad (19c)$$

B. Method of computation

To compute the mean and covariance of the output field for the different input states, we solve the truncated equations of motion (TEOMs) derived from the quantum master equation (QME), which corresponds to the infinite-shot limit of the stochastic quantum master equation (SQME) [30]. The means of the single-shot measurement outcomes are directly related to the quantum expectation values obtained from the TEOMs, as described below (see Appendix C for the derivation):

$$\mathbb{E}[I] = \sqrt{\mathcal{T}}\langle\hat{I}_2\rangle, \quad \mathbb{E}[Q] = \sqrt{\mathcal{T}}\langle\hat{Q}_2\rangle. \quad (20)$$

We estimate the measured quadrature covariance in terms of intracavity covariance of the output field, as if we have knowledge of the output states at the infinite shot limit,

$$\boldsymbol{\Sigma}_{I,Q} \approx \begin{bmatrix} \Delta I_2 & \Delta I Q_2 \\ \Delta I Q_2 & \Delta Q_2 \end{bmatrix}, \quad (21)$$

with

$$\Delta I_2 = \langle\hat{I}_2^2\rangle - \langle\hat{I}_2\rangle^2, \quad (22a)$$

$$\Delta I Q_2 = \langle\hat{I}_2\hat{Q}_2 + \hat{Q}_2\hat{I}_2\rangle/2 - \langle\hat{I}_2\rangle\langle\hat{Q}_2\rangle, \quad (22b)$$

$$\Delta Q_2 = \langle\hat{Q}_2^2\rangle - \langle\hat{Q}_2\rangle^2. \quad (22c)$$

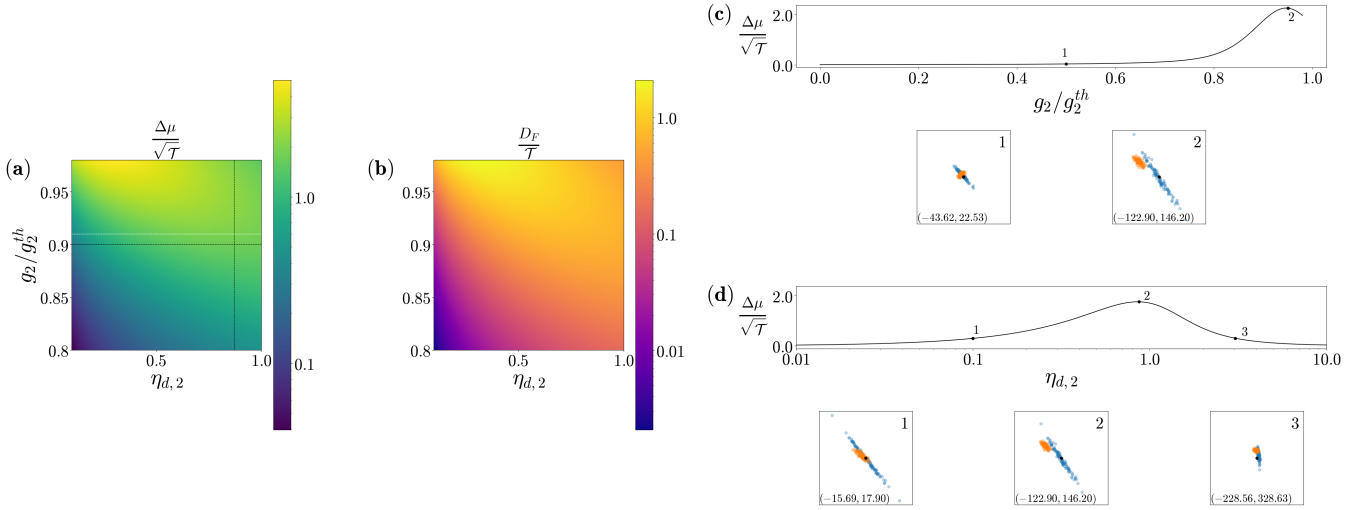


Figure 2. Variation of (a) normalized mean separation and (b) normalized Fisher discriminant versus pump strength g_2 (y-axis) and signal drive strength $\eta_{d,2}$ (x-axis). The vertical and horizontal black dashed lines in (a) indicate the cross sections shown in insets (c) and (d), respectively. The white dashed line marks the pump value giving 20 dB analyzer gain under the stiff-pump approximation. (c) Normalized mean separation versus g_2 , along with measured quadratures for the two quantum-state types at two pump operating points shown in the insets below. The classification fidelity for instances 1 and 2 are 0.855 and 0.995 respectively. (d) Normalized mean separation versus $\eta_{d,2}$, with measured quadratures at three signal drive operating points shown in the insets below. The classification fidelity for instances 1, 2 and 3 are 0.845, 1 and 0.9 respectively. The coordinates at the bottom left corner of the instances in insets (c) and (d) correspond to the center of the plotted region of the I - Q plot, all plots in each inset use identical axis ranges. The data shown is obtained by solving Eqs. B6 and the data shown in the insets of (c) and (d) is obtained by solving Eqs. B5 with 100 sample trajectories for each type of input state with integration time $800/\kappa_2$.

This does not fully agree with the single-shot measurement covariance, but gives a good idea of where the performance optimum will occur, as they agree qualitatively.

Before presenting the optimization results, we derive analytical expressions for the mean separation $\Delta\mu$ and the combined variance matrix \mathbf{V} that determine the Fisher discriminant D_F .

$$\Delta\mu = i\sqrt{\Lambda\mathcal{T}}\mathbf{U}\bar{\mathbf{J}}^{-1} \begin{bmatrix} 2\Delta\bar{C}_{s_2^\dagger s_2} & \Delta\bar{C}_{s_2 s_2} \\ -\Delta\bar{C}_{s_2^\dagger s_2^\dagger} & -2\Delta\bar{C}_{s_2^\dagger s_2} \end{bmatrix} \begin{bmatrix} \bar{s}_2 \\ \bar{s}_2^\dagger \end{bmatrix}, \quad (23)$$

where $\Delta\bar{C}_{o_1 o_2} = \bar{C}_{o_1 o_2}^{(2)} - \bar{C}_{o_1 o_2}^{(1)}$ and with

$$\bar{\mathbf{J}} = \begin{bmatrix} \left(i\Delta_2 - \frac{\kappa_2 + \Gamma}{2} + i2\bar{s}_2 \bar{s}_2^\dagger \right) & (-ig_2 e^{i\phi_2} + i\bar{s}_2^2) \\ \left(ig_2 e^{-i\phi_2} - i\bar{s}_2^{\dagger 2} \right) & (-i\Delta_2 - \frac{\kappa_2 + \Gamma}{2} - i2\bar{s}_2 \bar{s}_2^\dagger) \end{bmatrix}, \quad (24)$$

and

$$\mathbf{U} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ -i & i \end{bmatrix}. \quad (25)$$

This result is obtained in the weak-nonlinearity limit using a nonstandard perturbative treatment of the TEOMs, with details provided in Appendix D. The terms \bar{s}_2 and $\bar{C}_{s_2 s_2}, \bar{C}_{s_2^\dagger s_2}, \bar{C}_{s_2^\dagger s_2^\dagger}$ are zeroth-order contributions in the perturbative expansion for the mean and cumulant respectively. This analytical closed-form expression of the

mean separation in a weak nonlinearity limit demonstrates some properties clearly. First, we see that it depends on $\sqrt{\Lambda}$ and $\sqrt{\mathcal{T}}$ linearly. That gives correctly in the limit that Λ goes to zero, we have zero mean separation, thus classification fails. Secondly, the mean separation depends on the difference of the second order cumulants of the output distributions of different types. This shows how the nonlinear amplifier is able to compute the difference of cumulants via its quantum dynamics, and map them to first-order observables that are accessible via linear measurements. One can also calculate the Fisher discriminant under the perturbative approximation by making use of the approximate value of the combined variance matrix \mathbf{V} , which is obtained from Eq. (17) and Eq. (21):

$$\mathbf{V} = \frac{1}{4}\mathbf{U} \begin{bmatrix} \bar{C}_{s_2 s_2}^{(1)} + \bar{C}_{s_2 s_2}^{(2)} & \bar{C}_{s_2^\dagger s_2}^{(1)} + \bar{C}_{s_2^\dagger s_2}^{(2)} \\ \bar{C}_{s_2^\dagger s_2}^{(1)} + \bar{C}_{s_2^\dagger s_2}^{(2)} & \bar{C}_{s_2^\dagger s_2^\dagger}^{(1)} + \bar{C}_{s_2^\dagger s_2^\dagger}^{(2)} \end{bmatrix} \mathbf{U}^T. \quad (26)$$

Lastly, in all the results we present, we analyze mean separation ($\Delta\mu/\sqrt{\mathcal{T}}$) and the Fisher discriminant (D_F/\mathcal{T}) normalized by filtering time. For steady-state operation, as the filtering time is increased, both of these performance metrics will increase. The scaling is clearly shown in Eq. (23). This is also an intuitive result, as the noisy trajectories are integrated for a longer time, more information is collected, and one has more distinguishable output distributions.

C. Performance metrics under varying system parameters

In this section, we focus on how the externally-adjustable system parameters affect the classification performance. These adjustable system parameters are identified as the pump strength (g_2) and phase (ϕ_2), and the signal drive strength ($\eta_{d,2}$) and phase ($\phi_{d,2}$), where both drives are acting on the *analyzer*. An arbitrary choice of parameters will not work, as the analyzer needs to be finely-tuned to a regime where nonlinear processing can be achieved.

This section shows that optimal performance requires the joint optimization of the pump and signal drive strengths to identify the optimal operating points. Even with optimal drive strengths, an unfavorable choice of phase parameters can strongly degrade performance, making phase optimization equally essential. The remainder of this section provides an intuitive framework for selecting these optimal parameters by relating them directly to the classification task.

We separate the parameters into two groups, the strength parameters and the phase parameters, and show their effects separately. Let us start with the effect of the strength parameters on performance metrics. We show that the dependence of the mean separation, in Fig. 2(a), and the Fisher discriminant, in Fig. 2(b), are qualitatively very similar on these parameters. There is no global optimum in this two-dimensional parameter space. For an ideal linear amplifier, increasing the pump strength g_2 leads to improved performance, and the signal drive strength $\eta_{d,2}$ has no influence on performance. However, the Kerr nonlinearity of the practical *analyzer* changes this notion. At any gain value, a certain drive strength is needed for the Kerr nonlinearity to have an effect; at larger gains, a smaller drive strength suffices as the amplification can increase the *analyzer* mode occupation. However, it is also possible to have too strong a drive: then, the frequency shift due to the Kerr nonlinearity becomes too large and suppresses the useful nonlinear operation of the *analyzer* mode by making it non-resonant. As a result, for every pump strength there is an optimum drive strength and vice versa, as can be seen clearly from the cross-sections of the 2D plots shown in Fig 2c, d.

We now focus on the phase parameters and their effect on the performance metrics. The dependence of the mean separation, shown in Fig. 3(a), and of the Fisher discriminant, shown in Fig. 3(b), is more complex than their respective dependence on strength parameters. First, there are multiple optimal points of operation in the parameter space of the two phases. Moreover, as in the case of strength parameters, the trends for the mean separation and the Fisher discriminant are similar: peak performance in terms of both metrics are achieved around similar parameter values. To analyze and understand the relation of these optimal phase parameters with the given task and the performance metrics, we look at the

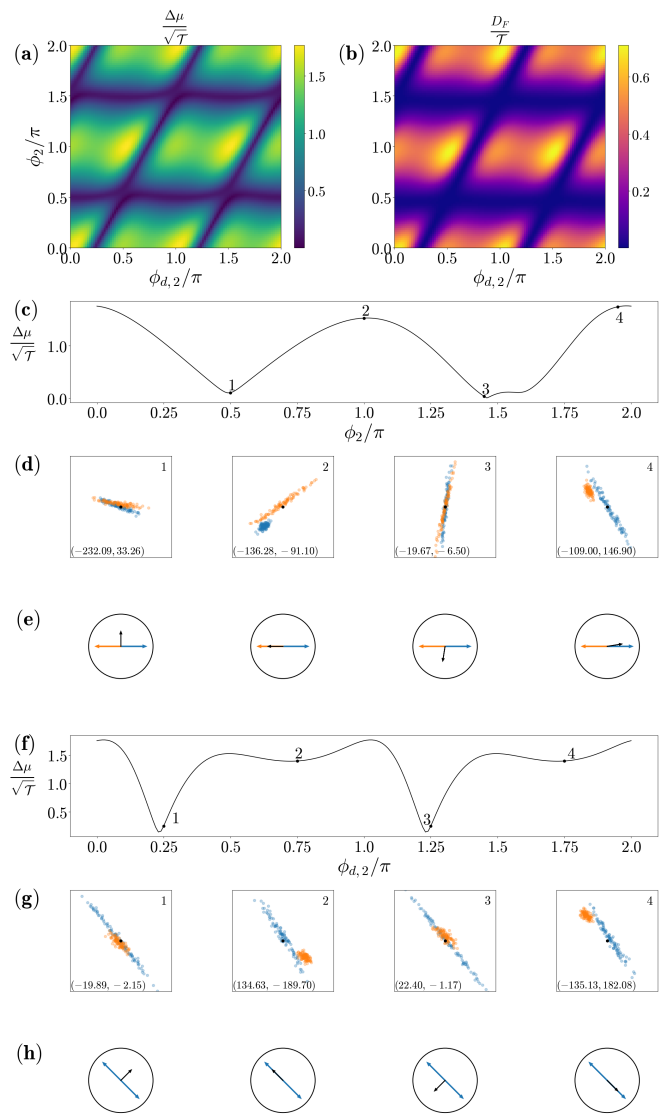


Figure 3. Variation of (a) normalized mean separation and (b) normalized Fisher discriminant versus pump phase ϕ_2 (y-axis) and signal drive phase $\phi_{d,2}$ (x-axis). Inset (c) shows the normalized mean separation at fixed $\phi_{d,2} = 0$ versus ϕ_2 , with inset (d) displaying corresponding measured quadratures for the two quantum state types at the operating points marked in (c). The classification fidelity for instances 1, 2, 3 and 4 are 0.81, 1, 0.785 and 1 respectively. Inset (e) illustrates the ϕ_1 values on the unit circle generating type-1 and type-2 states (blue and orange), with the black arrow indicating ϕ_2 at the chosen operating point. Inset (f) shows normalized mean separation at fixed $\phi_2 = 0$ versus $\phi_{d,2}$, and inset (g) shows the measured quadratures at the operating points marked in (f). The classification fidelity for instances 1, 2, 3 and 4 are 0.84, 0.99, 0.87 and 0.995 respectively. Inset (h) depicts the anti-squeezing (amplification) direction set by the fixed pump phase ($\phi_2 = 0$), aligned with the type-1 state (blue), with the black arrow marking the corresponding $\phi_{d,2}$. The coordinates at the bottom left corner of insets (d) and (g) correspond to the center of the plotted region of the I - Q plot, all plots in each inset use identical axis ranges. The data shown is obtained by solving Eqs. B6 and the data shown in the insets (d) and (g) is obtained by solving Eqs. B5 with 100 sample trajectories for each type of input state with integration time $800/\kappa_2$.

variation with respect to one of them at a time.

Looking at the mean separation against pump phase plot in Fig. 3(c), we see that the mean separation maxima, thus the best performance, are achieved when the phase of the *analyzer* pump is the same as the phase of the *squeezer* pump for either the type-1 (Fig. 3(e-4)) or type-2 (Fig. 3(e-2)) state ($\phi_2 = \phi_1^{(1)} = 0$ or $\phi_2 = \phi_1^{(2)} = \pi$), and the worst performance occurs when ϕ_2 differs by $\pi/2$ from $\phi_1^{(1)}$ and $\phi_1^{(2)}$ [Fig. 3(e-1,3)]. This can be further quantified by the classification fidelity, which is unity at operating points 2 and 4, indicating no overlap between the two state distributions. In contrast, at operating points 1 and 3 the fidelity is 0.81 and 0.785, similar to the fidelity under linear operation, 0.84, as shown in Fig. 1(c). The classification fidelity is evaluated using quadratic discriminant analysis on two distributions of 100 shots each at the corresponding operating point [33].

We illustrate the mechanism using operating point 4 in Fig. 3(c), with output distributions shown in Fig. 3(d-4). Here, the analyzer pump phase matches the squeezer pump phase for the type-1 state and differs by π for the type-2 state [Fig. 3(e-4)]. Consequently, the analyzer squeezes the type-1 state along the same quadrature, yielding a sharp elliptical output, while it acts as an anti-squeezer for the type-2 state, producing an almost circular distribution. This large covariance mismatch leads to a correspondingly large separation of the output means, since the mean separation is proportional to the difference in covariance of the two types of output states, as shown in Eq. (23). Operating point 2 follows the same logic, where the analyzer pump phase aligns with the squeezer pump phase of the type-2 state instead of type-1.

In contrast, at operating points 1 and 3 the analyzer phase is offset by $\pi/2$ from both squeezer phases [Fig. 3(e-1,3)], giving rise to similar output distributions for the two input states [Fig. 3(d-1,3)]. The resulting small covariance difference leads to a smaller mean separation.

Now, looking at the dependence of the mean separation on the signal drive phase in Fig. 3(f), we observe minima at $\phi_{d,2} = \pi/4$ and $\phi_{d,2} = 5\pi/4$ [Fig. 3(g-1,3)]. This occurs because this cross-section is taken at a pump phase of $\phi_2 = 0$, for which the anti-squeezing (amplification) axis lies along the $-\pi/4$ direction (and equivalently $3\pi/4$, as two ends of a line point in directions that are π apart) in the I - Q plane. At these signal drive phases, the direction of the signal drive is orthogonal to the amplification axis [Fig. 3(h-1,3)], resulting in a very small effective gain. Consequently, the input field is weakly amplified, leading to a vanishingly small mean separation, as the *analyzer* mode population is insufficient for the Kerr nonlinearity to significantly affect the system dynamics. Consistent with this behavior, the classification fidelities at operating points 1 and 3 are 0.84 and 0.87, respectively, well below those at the optimal operating points 2 and 4 (0.99 and 0.995).

On the other hand, when the signal drive phase is aligned with the amplification direction (anti-squeezing

direction) $\phi_{d,2} = 3\pi/4$ or $\phi_{d,2} = -\pi/4$ [Fig. 3(h-2,4)], we see a large value of the mean separation at those values and best performance is achieved around these points, with a twin-peak structure. The response $||\Delta\mu||$ does not depend only on the linear response function of the analyzer, in which case it would exhibit only a single peak, but has a more complicated dependence on the covariance terms too, as indicated by Eq. (23); this leads to a double-peak structure that has also been observed in previous work [24].

Lastly, one can see in the two-dimensional plots in Fig. 3(a) and Fig. 3(b), that all the optimal points maximizing both metrics occur around $\phi_2 = n\pi$. Around $\phi_2 = (2n+1)\pi/2$, both performance metrics take on their smallest values (close to zero), and there is not a big variation as the signal drive phase is varied. This shows that the optimization of phase parameters should start with identifying the optimal pump phase, followed by finding the optimal signal drive phase for that pump phase. In summary, therefore, the pump and signal drive strengths set the response of the *analyzer* mode, with larger responses generally yielding improved classification performance. However, even for optimally-chosen amplitudes, a poor choice of phase parameters can drive both performance metrics close to zero, resulting in poor classification. Therefore a combined optimization of *analyzer* pump phase and signal drive phase is crucial.

IV. EFFECT OF THE CLASSICAL MEASUREMENT NOISE

To obtain a more robust readout with larger signal amplitude, the output of the quantum amplifier is further amplified by electronic amplifiers, such as HEMTs, later in the measurement chain. This electronic amplification introduces additional noise to the measured signal, which degrades the classification performance metrics. We parametrize the strength of this added noise by \bar{n}_{cl} in Eq. (12). The resulting covariance matrix of the measured quadratures under this noise is:

$$\Sigma_{I,Q}^n = \bar{n}_{cl}\mathbb{I} + \Sigma_{I,Q}. \quad (27)$$

We refer to this contribution as *classical noise* since it does not induce backaction on the quantum system and enters solely as additive noise on the measured quantum trajectories. Using this covariance definition with added classical noise, we calculate the Fisher discriminant value for different noise contributions coming from the electronic amplification stage. We have plotted the variation of mean separation and the Fisher discriminant with different values of \bar{n}_{cl} with respect to the pump phase in Fig. 4. We note that the variation of the performance metrics with respect to parameters other than the pump phase is similar, and here we show results as a function of the pump phase for illustrative purposes.

The first observation one can make by looking at Fig. 4 is that the value of the Fisher discriminant decreases with

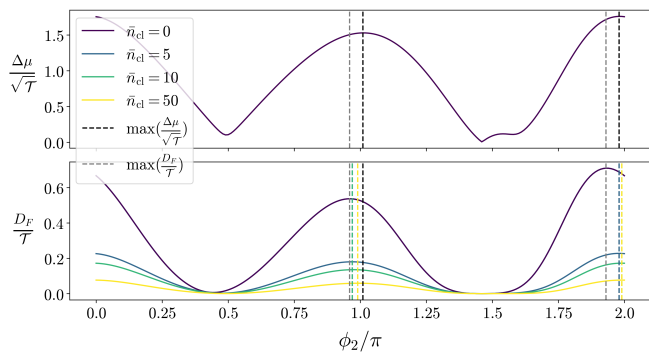


Figure 4. Normalized mean separation and the Fisher discriminant versus pump phase for different classical noise strengths (darker colors indicate larger noise). Black dashed lines mark the mean separation maxima, and grey dashed lines mark the Fisher discriminant maxima in the zero classical noise case. Colored dashed lines indicate the Fisher discriminant maxima for nonzero noise, using matching colors to the corresponding solid curves.

increasing classical noise at all pump phase values. This is an expected result as this classical noise adds uncertainty equally along all directions in the I - Q plane, thus reducing the signal-to-noise ratio (SNR) of the output independently of the system parameters.

We also see that the peaks of mean separation and the Fisher discriminant with zero classical noise, occur in similar neighborhoods of the pump phase but at different values. As the directionally uniform classical noise component in the system is increased, the pump phase value where the peak of the Fisher discriminant is achieved, shifts towards the one where mean separation peak is achieved. As this noise becomes the dominant noise in the system, the two different types of output distributions inevitably overlap with each other more, unless the mean separation between them is much greater than the variance of the added classical noise. In the presence of classical noise, therefore, to maximize the Fisher discriminant one should aim to maximize the mean separation.

V. APPLICATION TO QUBIT READOUT WITHOUT CAVITY DISPLACEMENT

In this section, we discuss a possible application of our *squeezer-analyzer* architecture to qubit readout. We discuss a scenario where the information of the qubit state is imprinted onto squeezed vacuum states by dispersively coupling the qubit to the *squeezer* as shown in Fig. 5(a). The *analyzer* performs state discrimination on these squeezed vacuum states via nonlinear amplification, effectively performing qubit readout. We consider this scenario to demonstrate how qubit readout can be performed by harnessing squeezing and nonlinearity as the main resources for quantum information processing, by considering simple modifications of the *squeezer-analyzer*

system discussed in previous sections.

For the dispersive qubit readout application without cavity displacement, we consider a simple setup in which the only change from our *squeezer-analyzer* setup is that the *squeezer* is coupled to a qubit. In this setting, the detuning of the squeezer mode (Δ_1) now takes on the value of the dispersive shift of the cavity resonance (χ) imposed by the qubit state:

$$\Delta_1 = \begin{cases} \chi, & \text{qubit in excited state } (|1\rangle) \\ -\chi, & \text{qubit in ground state } (|0\rangle). \end{cases} \quad (28)$$

With this change, the pointer states to be classified by the *analyzer* will now be conditioned on the state of the qubit, instead of being set by an external pump phase, as we considered previously. As a result, classifying these pointer states will enable qubit state readout.

The advantage of this qubit readout scheme is that there is no displaced field in the resonator mode, which is coupled to the qubit; instead the resonator is squeezed by the *squeezer* pump. Compared to standard dispersive qubit readout, this scheme offers potential advantages. In particular, operating with squeezed vacuum states can reduce the average intracavity photon number, therefore mitigating measurement-induced backaction [1]. Beyond that, engineered squeezing may also positively influence qubit coherence as coupling a transmon to a squeezed field has been shown to enhance its transverse coherence time [34].

In this section, we analyze the optimal parameter regime for encoding the qubit state onto squeezed vacuum states. We analyze the effects of the dispersive shift (χ) and the *squeezer* pump phase (ϕ_1) on optimal encoding that maximizes the classification performance when the analyzer is acting on the new pointer states generated by the qubit-*squeezer* coupling.

In order to do so, we choose fixed pump strengths $g_1 = 0.9g_1^{th}$ and $g_2 = 0.9g_2^{th}$, we should note that the pump strength we have chosen for the *squeezer* approximately corresponds to a photon number of 2 ($\langle \hat{s}_1^\dagger \hat{s}_1 \rangle = 2.13$) verifying that the setup operates in a very low-photon regime. We also fix the *analyzer* pump phase at $\phi_2 = \pi/2$ and the analyzer signal drive phase at a value matching the amplification direction of the *analyzer* amplifier ($\phi_{d,2} = -\pi/4 + \phi_2/2 = 0$). For a fixed signal drive strength on the analyzer ($\eta_{d,2} = 1$), we then sweep ϕ_1 and χ and calculate the mean separation and Fisher discriminant semi-analytically using the perturbative expressions (Eqs. 23-26). We present the results in insets (b-c) of Fig. 5. Moreover, we present the I - Q plots of the input states (obtained by turning the analyzer off, i.e. setting $g_2 = \eta_{d,2} = 0$) and output states at various operating points in the insets (d-g) of Fig. 5.

Inspecting Fig. 5(b-c), we notice that the maxima for both performance metrics are achieved in the same region of the parameter space around $\phi_1 \approx 0$ and $\chi \approx 0.2\kappa$, where $\kappa = \kappa_1 + \Gamma$ is the total loss rate of the *squeezer* mode including the bare resonator decay rate and the decay due to unidirectional coupler.

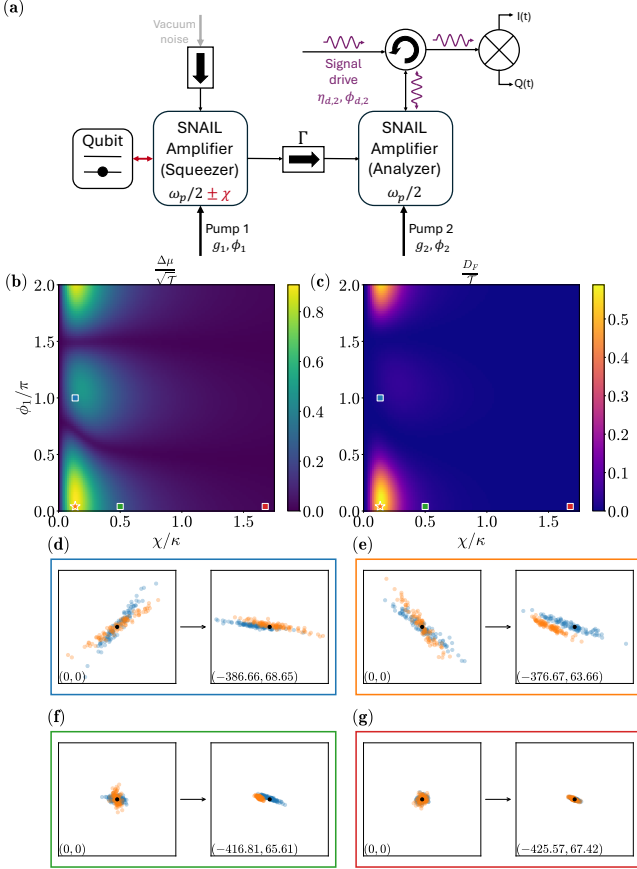


Figure 5. (a) Schematic of the modified *squeezer-analyzer* setup for qubit readout application without cavity displacement. The qubit is dispersively coupled to the *squeezer* leading to a resonance shift χ . Variation of (b) normalized mean separation and (c) normalized Fisher discriminant versus *squeezer* pump phase (ϕ_1) and qubit induced detuning (χ). The input (on the left) and output (on the right) I - Q plots as operating points marked in (b) and (c) are shown underneath with the matching colors in (d-g). The optimal operating point is shown with a star in orange and the corresponding I - Q plots are shown in inset (e). The coordinates shown at the bottom-left corner of insets (d-g) correspond to the midpoint of the plotting window, all plots corresponding to input(output) states use identical axis ranges. The classification fidelity for output distributions in (d-g) are 0.905, 0.995, 0.995 and 0.61 respectively. The data shown in (b) and (c) are obtained by using the analytical approximation in the perturbative limit in Eqs. (23-26) and the data shown in the insets (d-g) is obtained by solving Eqs. B5 with 100 sample trajectories for each type of input state with integration time $4000/\kappa_2$.

First, the optimal value of ϕ_1 depends on the fixed value of ϕ_2 , therefore it is better to address this as an optimal value for pump phase difference $\Delta\phi = \phi_2 - \phi_1$. We have verified this by running the same simulation with different ϕ_2 values while keeping the remaining fixed variables constant. In this case, the optimal value of ϕ_1 shifts with ϕ_2 , with the optimal value of $\Delta\phi$ remaining

the same. One can see how the performance deteriorates away from the optimal $\Delta\phi$, comparing the optimal $\Delta\phi$ point in inset (e) to the non-optimal $\Delta\phi$ point in inset (d) in Fig. 5.

Second, we observe that the optimal value of χ is smaller than the optimal dispersive shift for the standard dispersive qubit readout scheme at $\chi_{opt,disp} = \kappa/2$ [1](shown in inset Fig. 5(f)) for phase-readout. The reason behind this is that in this scheme, as χ is increased, the *squeezer* mode is tuned away from the single-mode-squeezed frequency, leading to a decrease in the squeezing strength. This can be seen from the input states in the Fig. 5(e-g). At the same time as χ increases, the angle between the squeezing axes of the input states gets closer to being orthogonal. This feature of the encoding imposes an important trade-off between the degree of squeezing and the separation of the squeezing axes of the pointer states as χ is varied.

For two single-mode Gaussian (squeezed) states with equal squeezing strength, their quantum distinguishability e.g. as quantified by the Helstrom bound or state overlap, is maximized when their squeezing ellipses are oriented along orthogonal quadrature axes [35]. We have analyzed such a binary discrimination task earlier in this paper. For the current qubit readout scheme where the qubit is dispersively coupled to the *squeezer* mode, we can also find an operating regime where the two qubit-conditioned states of the squeezer have orthogonal squeezing axes. An example operating point in (ϕ_1, χ) space where this can be achieved is marked in Fig. 5(b) with a red marker corresponding to inset (g). However, for qubit-state encoding optimization, we find that the degree of squeezing and the separation of squeezing axes become competing factors that both affect classification accuracy. We see that if the states have orthogonal squeezing axes but the squeezing strength is not large enough, one ends up with poor classification as in Fig. 5(g). We deduce that in this case, the orthogonal squeezing axes for the input states does not yield the optimal encoding, as the degree of squeezing to achieve the orthogonal separation becomes too small for effective state discrimination.

Lastly, the scheme we present here for encoding qubit states onto squeezed vacuum relies on a non-zero detuning of the squeezed mode, which fundamentally reduces the gain (equivalently the squeezing strength) in the *squeezer*. Therefore, both normalized mean separation and normalized Fisher discriminant at the optimal performance points attain lower values for the qubit readout scheme without cavity displacement than for our earlier scheme. Therefore, the quantum trajectories need to be integrated for a longer time to achieve similar classification fidelity. Alternatively, one can increase the *squeezer* pump strength (g_1), to increase performance metrics and therefore achieve similar fidelity with smaller integration time. However, we should note that increasing the squeezing strength also increases the average photon number in the readout cavity, therefore

posing a trade-off.

VI. CONCLUSION

We have shown that quantum amplifiers operated beyond the linear regime can yield a measurable advantage for single-shot quantum state classification when the measurement chain is optimized for task-specific objectives rather than SNR alone. Our analysis is grounded in a realistic superconducting readout architecture comprising experimentally accessible components, including a SNAIL-based oscillator for state preparation and a second SNAIL oscillator functioning as a nonlinear cryogenic amplifier.

Within this framework, we identify optimal operating points for the externally tunable parameters of the nonlinear amplifier and provide a physical interpretation of these optima, clarifying the mechanisms that enhance classification fidelity. We further analyze the impact of downstream classical noise from an electronic amplifier and outline strategies to mitigate its effect, thereby preserving the performance gains enabled by nonlinear amplification.

Our results motivate deliberate operation of quantum amplifiers in the nonlinear regime and establish their potential as functional primitives for quantum information processing. While the present analysis focuses on performance metrics derived from second-order moments, it naturally extends to the processing of information encoded in higher-order moments of quantum signals, a regime of direct relevance to non-Gaussian quantum information protocols.

We further establish the feasibility of qubit readout without cavity displacement enabled by nonlinear amplification, showing that two zero-mean cavity states can be discriminated via phase information contained in their higher-order statistical moments, even when their first moments are identical. Building on this observation, we

introduce a readout protocol that operates without cavity displacement and encodes the qubit state directly in these higher-order moments, thereby enabling discrimination in the zero-cavity-displacement regime. The present treatment does not account for measurement backaction, motivating future work that incorporates a realistic model of qubit–squeezer coupling. Moreover, the architecture we propose for qubit readout without cavity displacement is preliminary and could be further optimized, providing a natural direction for future work. More broadly, our findings provide a foundation for systematic, end-to-end, resource-constrained optimization of nonlinear quantum amplifiers for quantum information processing applications.

ACKNOWLEDGMENTS

We thank Leon Bello for insightful discussions. We acknowledge support from the AFOSR under Grant No. FA9550-20-1-0177 and the Army Research Office under Grant No. W911NF18-1-0144. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Office or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes, notwithstanding any copyright notation herein. Simulations in this paper were performed using the Princeton Research Computing resources at Princeton University, which is a consortium of groups led by the Princeton Institute for Computational Science and Engineering (PICSciE) and Office of Information Technology’s Research Computing. B.M. acknowledges support from the NSF Graduate Research Fellowship Program Fellow No. 2022338601. **Conflict of Interest:** Michael Hatridge serves as a consultant for Quantum Circuits, Inc., receiving remuneration in the form of consulting fees, and hold equity in the form of stock options.

-
- [1] A. Blais, A. L. Grimsmo, S. Girvin, and A. Wallraff, *Reviews of Modern Physics* **93**, 025005 (2021).
 - [2] H. A. Haus and J. A. Mullen, *Physical Review* **128**, 2407 (1962).
 - [3] C. M. Caves, *Physical Review D* **26**, 1817 (1982).
 - [4] A. A. Clerk, M. H. Devoret, S. M. Girvin, F. Marquardt, and R. J. Schoelkopf, *Reviews of Modern Physics* **82**, 1155 (2010).
 - [5] J. Aumentado, *IEEE Microwave Magazine* **21**, 45 (2020).
 - [6] S. A. Khan, R. Kaufman, B. Mesits, M. Hatridge, and H. E. Türeci, *PRX Quantum* **5**, 020364 (2024).
 - [7] N. E. Frattini, V. V. Sivak, A. Lingenfelter, S. Shankar, and M. H. Devoret, *Physical Review Applied* **10**, 054020 (2018).
 - [8] V. Sivak, N. Frattini, V. Joshi, A. Lingenfelter, S. Shankar, and M. Devoret, *Physical Review Applied* **11**, 054060 (2019).
 - [9] S. M. Kay, *Fundamentals of Statistical Signal Processing: Detection Theory* (Prentice Hall PTR, 1998).
 - [10] H. L. V. Trees, *Detection, Estimation, and Linear Modulation Theory* (Wiley, 1968).
 - [11] R. S. Bondurant, *Physical Review Letters* **71**, 1709 (1993).
 - [12] M. D. Reed, L. DiCarlo, B. R. Johnson, L. Sun, D. I. Schuster, L. Frunzio, and R. J. Schoelkopf, *Physical Review Letters* **105**, 173601 (2010).
 - [13] R. Vijay, M. H. Devoret, and I. Siddiqi, *Review of Scientific Instruments* **80**, 111101 (2009).
 - [14] J. M. Epstein, K. B. Whaley, and J. Combes, *Physical Review A* **103**, 052415 (2021).
 - [15] J. Gambetta, W. A. Braff, A. Wallraff, S. M. Girvin, and R. J. Schoelkopf, *Phys. Rev. A* **76**, 012325 (2007).
 - [16] M. Silveri, E. Zaly-Geller, M. Hatridge, Z. Leghtas, M. H. Devoret, and S. M. Girvin, *Phys. Rev. A* **93**,

- 062310 (2016).
- [17] S. A. Khan, S. Prabhu, L. G. Wright, and P. L. McMahon, *Quantum Computational-Sensing Advantage* (2025), arXiv:2507.16918 [quant-ph].
 - [18] Q. Zhuang and Z. Zhang, *Physical Review X* **9**, 041023 (2019).
 - [19] J. Sinanan-Singh, G. L. Mintzer, I. L. Chuang, and Y. Liu, *Quantum* **8**, 1427 (2024).
 - [20] Z. E. Chin, D. R. Leibbrandt, and I. L. Chuang, arXiv:2405.05888 (2024).
 - [21] R. R. Allen, F. Machado, I. L. Chuang, H.-Y. Huang, and S. Choi, *Quantum computing enhanced sensing* (2025), arXiv:2501.07625 [quant-ph].
 - [22] Z. Eldredge, M. Foss-Feig, J. A. Gross, S. L. Rolston, and A. V. Gorshkov, *Physical Review A* **97**, 042337 (2018).
 - [23] T. Ilias, F. Hu, M. Vives, and H. E. Türeci, *End-to-end Optimization of Single-Shot Quantum Machine Learning for Bayesian Inference* (2025), arXiv:2512.20492 [quant-ph].
 - [24] S. A. Khan, F. Hu, G. Angelatos, M. Hatridge, and H. E. Türeci, *A neural processing approach to quantum state discrimination* (2025), arXiv:2409.03748 [quant-ph].
 - [25] S. A. Khan, F. Hu, G. Angelatos, and H. E. Türeci, *Physical reservoir computing using finitely-sampled quantum systems* (2021), arXiv:2110.13849 [quant-ph].
 - [26] A. Roy and M. Devoret, *Comptes Rendus Physique* **17**, 740 (2016), quantum microwaves / Micro-ondes quantiques.
 - [27] A. Roy and M. Devoret, *Phys. Rev. B* **98**, 045405 (2018).
 - [28] S. Boutin, D. M. Toyli, A. V. Venkatramani, A. W. Edkins, I. Siddiqi, and A. Blais, *Physical Review Applied* **8**, 054030 (2017).
 - [29] A. Metelmann and A. A. Clerk, *Phys. Rev. X* **5**, 021025 (2015).
 - [30] H. M. Wiseman and G. J. Milburn, *Quantum Measurement and Control* (Cambridge University Press, Cambridge, 2009).
 - [31] R. Schack and A. Schenzle, *Phys. Rev. A* **41**, 3847 (1990).
 - [32] R. A. Fisher, *Annals of Eugenics* **7**, 179 (1936), <https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1469-1809.1936.tb02137.x>.
 - [33] T. Hastie, R. Tibshirani, and J. Friedman, *The elements of statistical learning: data mining, inference and prediction*, 2nd ed. (Springer, 2009).
 - [34] K. W. Murch, S. J. Weber, K. M. Beck, E. Ginossar, and I. Siddiqi, *Nature* **499**, 62 (2013).
 - [35] E. M. F. Curado, S. Faci, J.-P. Gazeau, and D. Noguera, *Entropy* **24**, 220 (2022).

Appendix A: Derivation of the system Hamiltonian in the most general form - Hamiltonian including third- and fourth-order nonlinearities of SNAIL

Here, we want to establish the most general Hamiltonian to model a SNAIL amplifier including all effects of third- and fourth-order nonlinearities, driven at resonance frequency of the SNAIL (ω_s) and twice of resonance frequency ($\omega_p = 2\omega_s$). We start with the Hamiltonian of the single SNAIL:

$$\begin{aligned}\hat{\mathcal{H}}_{SNAIL} = & \omega_s \hat{a}^\dagger \hat{a} + g_3 (\hat{a} + \hat{a}^\dagger)^3 + g_4 (\hat{a} + \hat{a}^\dagger)^4 \\ & + \sqrt{\kappa_s} \epsilon_p (e^{-i\phi_p} e^{-i2\omega_s t} + e^{i\phi_p} e^{i2\omega_s t}) (\hat{a} + \hat{a}^\dagger) \\ & + \sqrt{\kappa_s} \eta_{sig} (e^{-i\phi_{sig}} e^{-i\omega_s t} + e^{i\phi_{sig}} e^{i\omega_s t}) (\hat{a} + \hat{a}^\dagger).\end{aligned}\quad (\text{A1})$$

Then, we represent both the resonant SNAIL mode at ω_s and the pump mode at $2\omega_s$ within the mode \hat{a} .

$$\hat{a} = \hat{s} + \hat{p}. \quad (\text{A2})$$

We plug in this ansatz and go to a rotating frame with $\hat{s} = \hat{s} e^{-i\omega_s t}$, $\hat{p} = \hat{p} e^{-i2\omega_s t}$, where $\omega_s = \omega_p/2$, and keep only non-rotating terms, and we end up with the following Hamiltonian:

$$\begin{aligned}\hat{\mathcal{H}}_{3-4} = & (-\omega_s + 24g_4) \hat{p}^\dagger \hat{p} + (24g_4) \hat{s}^\dagger \hat{s} + 3g_3 (\hat{p}^\dagger \hat{s} \hat{s} + \hat{p} \hat{s}^\dagger \hat{s}^\dagger) + 6g_4 (\hat{p}^\dagger \hat{p}^\dagger \hat{p} \hat{p} + \hat{s}^\dagger \hat{s}^\dagger \hat{s} \hat{s}) + 24g_4 (\hat{p}^\dagger \hat{p} \hat{s}^\dagger \hat{s}) \\ & + \sqrt{\kappa_s} \epsilon_p (e^{-i\phi_p} \hat{p}^\dagger + e^{i\phi_p} \hat{p}) + \sqrt{\kappa_s} \eta_{sig} (e^{-i\phi_{sig}} \hat{s}^\dagger + e^{i\phi_{sig}} \hat{s}).\end{aligned}\quad (\text{A3})$$

If we assume the pump is stiff, the Hamiltonian will be simplified in the following way:

$$\hat{\mathcal{H}}_{3-4}^{stiff} = (24g_4 (1 + |\bar{\mathcal{P}}|^2)) \hat{s}^\dagger \hat{s} + 3g_3 (\bar{\mathcal{P}}^* \hat{s} \hat{s} + \bar{\mathcal{P}} \hat{s}^\dagger \hat{s}^\dagger) + 6g_4 \hat{s}^\dagger \hat{s}^\dagger \hat{s} \hat{s} + \sqrt{\kappa_s} \eta_{sig} (e^{-i\phi_{sig}} \hat{s}^\dagger + e^{i\phi_{sig}} \hat{s}). \quad (\text{A4})$$

with $\bar{\mathcal{P}} = \frac{i\sqrt{\kappa_s} \epsilon_p e^{-i\phi_p}}{(i\omega_s - \frac{\kappa_s}{2})}$.

When we match the parameters here with the *analyzer* in the main text we have:

$$\Delta_2 = -24g_4 (1 + |\bar{\mathcal{P}}|^2), \quad (\text{A5a})$$

$$\Lambda = -12g_4, \quad (\text{A5b})$$

$$\epsilon_p = \left| -\frac{g_2 e^{\phi_2}}{i6g_3 \sqrt{\kappa_2} \chi_2} \right|, \quad (\text{A5c})$$

$$\phi_p = -\arg \left\{ -\frac{g_2 e^{\phi_2}}{i6g_3 \sqrt{\kappa_j} \chi_2} \right\}, \quad (\text{A5d})$$

$$\eta_{d,2} = \eta_{sig}, \quad (\text{A5e})$$

$$\phi_{sig} = \phi_{d,2} - \pi/2, \quad (\text{A5f})$$

with $\chi_2^{-1} = -i\Delta_2 + \frac{\kappa_2}{2}$.

We can also model the system in the main text, consisting of an *squeezer* SNAIL unidirectionally coupled to an *analyzer* SNAIL, in this way. The Liouvillian can be split into three parts as follows:

$$\mathcal{L}_{non-stiff} = \mathcal{L}_{sq-non-stiff} + \mathcal{L}_{an-non-stiff} + \mathcal{L}_{coup}, \quad (\text{A6})$$

and \mathcal{L}_{coup} is as in Eq. (9). The *squeezer* Liouvillian is modified to include non-stiff dynamics to:

$$\mathcal{L}_{sq}^{full} \hat{\rho} = -i[\hat{\mathcal{H}}_{sq}^{full}, \hat{\rho}] + \kappa_1 \mathcal{D}[\hat{s}_1 + \hat{p}_1] \hat{\rho}, \quad (\text{A7})$$

with

$$\hat{\mathcal{H}}_{sq}^{full} = -\omega_{s,1} \hat{p}_1^\dagger \hat{p}_1 + 3g_{3,1} (\hat{p}_1^\dagger \hat{s}_1 \hat{s}_1 + \hat{p}_1 \hat{s}_1^\dagger \hat{s}_1^\dagger) + \sqrt{\kappa_1} \epsilon_{p,1} (e^{-i\phi_{p,1}} \hat{p}_1^\dagger + e^{i\phi_{p,1}} \hat{p}_1). \quad (\text{A8})$$

Finally, the *analyzer* Liouvillian is as follows:

$$\mathcal{L}_{an}^{full} \hat{\rho} = -i[\hat{\mathcal{H}}_{an}^{full}, \hat{\rho}] + \kappa_2 \mathcal{D}[\hat{s}_2 + \hat{p}_2] \hat{\rho}, \quad (\text{A9})$$

with

$$\begin{aligned} \hat{\mathcal{H}}_{an}^{full} = & (-\omega_{s,2} + 24g_4) \hat{p}_2^\dagger \hat{p}_2 + (24g_4) \hat{s}_2^\dagger \hat{s}_2 + 3g_{3,2} \left(\hat{p}_2^\dagger \hat{s}_2 \hat{s}_2 + \hat{p}_2 \hat{s}_2^\dagger \hat{s}_2 \right) + 6g_4 \left(\hat{p}_2^\dagger \hat{p}_2^\dagger \hat{p}_2 \hat{p}_2 + \hat{s}_2^\dagger \hat{s}_2^\dagger \hat{s}_2 \hat{s}_2 \right) \\ & + 24g_4 \left(\hat{p}_2^\dagger \hat{p}_2 \hat{s}_2^\dagger \hat{s}_2 \right) + \sqrt{\kappa_2} \epsilon_{p,2} \left(e^{-i\phi_{p,2}} \hat{p}_2^\dagger + e^{i\phi_{p,2}} \hat{p}_2 \right) + \sqrt{\kappa_2} \eta_{d,2} \left(e^{-i\phi_{sig,2}} \hat{s}_2^\dagger + e^{i\phi_{sig,2}} \hat{s}_2 \right). \end{aligned} \quad (\text{A10})$$

One can similarly match the pump strength and phase for the *squeezer* with the stiff-pump approximation as in Eq. (A5c) and Eq. (A5d), simply by changing indices 2 to 1.

Appendix B: Cumulant expansion method and STEOMs of the system

In this section, we explain the relation between the moments of a multimode quantum system and their cumulants and then show how we use this relation to derive a finite set of equations of motions of the cumulants of our system. An n-order moment of an arbitrary multimode system can be written in terms of cumulants as follows:

$$\langle o_1 o_2 \cdots o_n \rangle = \sum_{\pi} \prod_{B \in \pi} C_{o_i: i \in B}, \quad (\text{B1})$$

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

We show this on an example derived from the SQME describing our system, which is introduced in the main text in Eq. (11). If we derive the equation of motion for $\langle \hat{s}_2 \rangle^c = \text{tr}\{\mathcal{L}\hat{\rho}^c\}$ from the SQME we end up with the following equation:

$$\begin{aligned} d\langle \hat{s}_2 \rangle^c = & \left(-\Gamma_1 \langle \hat{s}_1 \rangle^c + \left(i\Delta_2 - \frac{\kappa_2 + \Gamma_1}{2} \right) \langle \hat{s}_2 \rangle^c + i\Lambda \langle \hat{s}_2^\dagger \hat{s}_2 \hat{s}_2 \rangle^c - ig_2 e^{i\phi_2} \langle \hat{s}_2^\dagger \rangle^c - \sqrt{\kappa_2} \eta_{d,2} e^{i\phi_{d,2}} \right) dt \\ & + \sqrt{\frac{\kappa_2}{2}} \left(\langle \hat{s}_2 \hat{s}_2 \rangle^c + \langle \hat{s}_2^\dagger \hat{s}_2 \rangle^c - \langle \hat{s}_2 \rangle^c \langle \hat{s}_2 \rangle^c - \langle \hat{s}_2^\dagger \rangle^c \langle \hat{s}_2 \rangle^c \right) dW_I \\ & - i\sqrt{\frac{\kappa_2}{2}} \left(\langle \hat{s}_2 \hat{s}_2 \rangle^c - \langle \hat{s}_2^\dagger \hat{s}_2 \rangle^c - \langle \hat{s}_2 \rangle^c \langle \hat{s}_2 \rangle^c + \langle \hat{s}_2^\dagger \rangle^c \langle \hat{s}_2 \rangle^c \right) dW_Q. \end{aligned} \quad (\text{B2})$$

We notice two things from here. First, the measurement operator couples this first-order moment equation to second-order moments. We know $C_{o_1 o_2} = \langle o_1 o_2 \rangle - \langle o_1 \rangle \langle o_2 \rangle$, so we can replace the terms of second-order moments with cumulants in the measurement terms.

The second and more important observation is that the Kerr term (associated with Λ), couples this first-order moment equation to a third-order moment. This coupling to the higher-order moments will lead to an unclosed set of equations as the equations of motions lower-order moments include terms with higher-order moments. In order to prevent this and end up with finite number of equations, we truncate our cumulants at second-order, meaning we assume cumulants of order three and higher are all zero. This approximation means that we assume that our states are Gaussian, which is a good approximation in the case of weakly nonlinear systems with Gaussian inputs, as in our case. We now show how this is done using the third-order moment associated with the Kerr term. First, we write it in terms of cumulants:

$$\langle \hat{s}_2^\dagger \hat{s}_2 \hat{s}_2 \rangle^c = C_{\hat{s}_2^\dagger \hat{s}_2 \hat{s}_2}^c + 2C_{\hat{s}_2^\dagger \hat{s}_2}^c \langle \hat{s}_2 \rangle^c + C_{\hat{s}_2 \hat{s}_2}^c \langle \hat{s}_2^\dagger \rangle^c + \langle \hat{s}_2^\dagger \rangle^c \langle \hat{s}_2 \rangle^c \langle \hat{s}_2 \rangle^c. \quad (\text{B3})$$

Then we assume that the third-order moment is zero ($C_{\hat{s}_2^\dagger \hat{s}_2 \hat{s}_2}^c = 0$), thus we can write the third-order cumulant in terms of second and first-order cumulants. After doing these conversions we end up with the following equation which only contains first and second-order cumulants:

$$\begin{aligned} d\langle \hat{s}_2 \rangle^c = & dt \left(-\Gamma_1 \langle \hat{s}_1 \rangle^c + \left(i\Delta_2 - \frac{\kappa_2 + \Gamma_1}{2} \right) \langle \hat{s}_2 \rangle^c + i\Lambda \left(2C_{\hat{s}_2^\dagger \hat{s}_2}^c \langle \hat{s}_2 \rangle^c + \langle \hat{s}_2^\dagger \rangle^c \left(\langle \hat{s}_2 \rangle^c \right) + C_{\hat{s}_2 \hat{s}_2}^c \right) - ig_2 e^{i\phi_2} \langle \hat{s}_2^\dagger \rangle^c \right. \\ & \left. - \sqrt{\kappa_2} \eta_{d,2} e^{i\phi_{d,2}} \right) + \sqrt{\frac{\kappa_2}{2}} \left(C_{\hat{s}_2 \hat{s}_2}^c + C_{\hat{s}_2^\dagger \hat{s}_2}^c \right) dW_I - i\sqrt{\frac{\kappa_2}{2}} \left(C_{\hat{s}_2 \hat{s}_2}^c - C_{\hat{s}_2^\dagger \hat{s}_2}^c \right) dW_Q. \end{aligned} \quad (\text{B4})$$

We apply the same procedure to all the other first- and second-order moments' equations of motion and end up with 15 coupled closed set of equations, STEOMs, for this system:

$$d\langle\hat{s}_1\rangle^c = dt\left(-\frac{\Gamma_1 + \kappa_1}{2}\langle\hat{s}_1\rangle^c + i\Delta_1\langle\hat{s}_1\rangle^c - ig_1e^{i\phi_1}\langle\hat{s}_1^\dagger\rangle^c\right) + \sqrt{\frac{\kappa_2}{2}}\left(C_{s_1s_2}^c + C_{s_2^\dagger s_1}^c\right)dW_I - i\sqrt{\frac{\kappa_2}{2}}\left(C_{s_1s_2}^c - C_{s_2^\dagger s_1}^c\right)dW_Q, \quad (\text{B5a})$$

$$d\langle\hat{s}_2\rangle^c = dt\left(-\Gamma_1\langle\hat{s}_1\rangle^c + \left(i\Delta_2 - \frac{\kappa_2 + \Gamma_1}{2}\right)\langle\hat{s}_2\rangle^c + i\Lambda\left(2C_{s_2^\dagger s_2}^c\langle\hat{s}_2\rangle^c + \langle\hat{s}_2^\dagger\rangle^c\left(\langle\hat{s}_2\rangle^c + C_{s_2s_2}^c\right)\right) - ig_2e^{i\phi_2}\langle\hat{s}_2^\dagger\rangle^c - \sqrt{\kappa_2}\eta_{d,2}e^{i\phi_{d,2}}\right) + \sqrt{\frac{\kappa_2}{2}}\left(C_{s_2s_2}^c + C_{s_2^\dagger s_2}^c\right)dW_I - i\sqrt{\frac{\kappa_2}{2}}\left(C_{s_2s_2}^c - C_{s_2^\dagger s_2}^c\right)dW_Q, \quad (\text{B5b})$$

$$d\langle\hat{s}_1^\dagger\rangle^c = \left(-\frac{\Gamma_1 + \kappa_1}{2}\langle\hat{s}_1^\dagger\rangle^c - i\Delta_1\langle\hat{s}_1^\dagger\rangle^c + ig_1e^{-i\phi_1}\langle\hat{s}_1\rangle^c\right)dt + \sqrt{\frac{\kappa_2}{2}}\left(C_{s_1^\dagger s_2}^c + C_{s_1^\dagger s_2^\dagger}^c\right)dW_I - i\sqrt{\frac{\kappa_2}{2}}\left(C_{s_1^\dagger s_2}^c - C_{s_1^\dagger s_2^\dagger}^c\right)dW_Q, \quad (\text{B5c})$$

$$d\langle\hat{s}_2^\dagger\rangle^c = dt\left(-\Gamma_1\langle\hat{s}_1^\dagger\rangle^c - \left(i\Delta_2 + \frac{\kappa_2 + \Gamma_1}{2}\right)\langle\hat{s}_2^\dagger\rangle^c - i\Lambda\left(2C_{s_2^\dagger s_2}^c\langle\hat{s}_2^\dagger\rangle^c + \langle\hat{s}_2\rangle^c\left(\langle\hat{s}_2^\dagger\rangle^c + C_{s_2^\dagger s_2}^c\right)\right) + ig_2e^{-i\phi_2}\langle\hat{s}_2\rangle^c - \sqrt{\kappa_2}\eta_{d,2}e^{-i\phi_{d,2}}\right) + \sqrt{\frac{\kappa_2}{2}}\left(C_{s_2^\dagger s_2}^c + C_{s_2^\dagger s_2^\dagger}^c\right)dW_I - i\sqrt{\frac{\kappa_2}{2}}\left(C_{s_2^\dagger s_2}^c - C_{s_2^\dagger s_2^\dagger}^c\right)dW_Q, \quad (\text{B5d})$$

$$C_{s_1s_1}^{\dot{c}} = (-\Gamma_1 - \kappa_1 + i2\Delta_1)C_{s_1s_1}^c - 2\kappa_2C_{s_1s_2}^cC_{s_2^\dagger s_1}^c - ig_1e^{i\phi_1}(1 + 2C_{s_1^\dagger s_1}^c), \quad (\text{B5e})$$

$$C_{s_1s_2}^{\dot{c}} = -\Gamma_1C_{s_1s_1}^c + \left(-\Gamma_1 - \frac{\kappa_1 + \kappa_2}{2} + i(\Delta_1 + \Delta_2)\right)C_{s_1s_2}^c - \kappa_2C_{s_1s_2}^cC_{s_2^\dagger s_2}^c - \kappa_2C_{s_2s_2}^cC_{s_2^\dagger s_1}^c + i\Lambda\left(C_{s_2^\dagger s_1}^cC_{s_2s_2}^c + 2C_{s_1s_2}^cC_{s_2^\dagger s_2}^c + C_{s_2^\dagger s_1}^c\langle\hat{s}_2\rangle^c + 2C_{s_1s_2}^c\langle\hat{s}_2\rangle^c\langle\hat{s}_2^\dagger\rangle^c\right) - ig_1e^{i\phi_1}C_{s_1^\dagger s_2}^c - ig_2e^{i\phi_2}C_{s_2^\dagger s_1}^c, \quad (\text{B5f})$$

$$C_{s_2s_2}^{\dot{c}} = -2\Gamma_1C_{s_1s_2}^c + (-\Gamma_1 - \kappa_2 + i2\Delta_2)C_{s_2s_2}^c - 2\kappa_2C_{s_2^\dagger s_2}^cC_{s_2s_2}^c + i\Lambda\left(C_{s_2s_2}^c + 6C_{s_2s_2}^cC_{s_2^\dagger s_2}^c + \langle\hat{s}_2\rangle^c + 2C_{s_2^\dagger s_2}^c\langle\hat{s}_2\rangle^c + 4C_{s_2s_2}^c\langle\hat{s}_2\rangle^c\langle\hat{s}_2^\dagger\rangle^c\right) - ig_2e^{i\phi_2}\left(1 + 2C_{s_2^\dagger s_2}^c\right), \quad (\text{B5g})$$

$$C_{s_1^\dagger s_1}^{\dot{c}} = (-\Gamma_1 - \kappa_1)C_{s_1^\dagger s_1}^c - \kappa_2C_{s_1^\dagger s_2}^cC_{s_2^\dagger s_1}^c - \kappa_2C_{s_1s_2}^cC_{s_1^\dagger s_2^\dagger}^c - ig_1e^{i\phi_1}C_{s_1^\dagger s_1^\dagger}^c + ig_1e^{-i\phi_1}C_{s_1s_1}^c, \quad (\text{B5h})$$

$$C_{s_1^\dagger s_2}^{\dot{c}} = -\Gamma_1C_{s_1^\dagger s_1}^c + \left(-\Gamma_1 - \frac{\kappa_1 + \kappa_2}{2} + i(-\Delta_1 + \Delta_2)\right)C_{s_1^\dagger s_2}^c - \kappa_2C_{s_1^\dagger s_2}^cC_{s_2^\dagger s_2}^c - \kappa_2C_{s_1^\dagger s_2^\dagger}^cC_{s_2s_2}^c + i\Lambda\left(2C_{s_1^\dagger s_2}^cC_{s_2^\dagger s_2}^c + C_{s_1^\dagger s_2^\dagger}^cC_{s_2s_2}^c + C_{s_1^\dagger s_2^\dagger}^c\langle\hat{s}_2\rangle^c + 2C_{s_1^\dagger s_2}^c\langle\hat{s}_2\rangle^c\langle\hat{s}_2^\dagger\rangle^c\right) + ig_1e^{-i\phi_1}C_{s_1s_2}^c - ig_2e^{i\phi_2}C_{s_1^\dagger s_2^\dagger}^c, \quad (\text{B5i})$$

$$C_{s_2^\dagger s_1}^{\dot{c}} = -\Gamma_1C_{s_1^\dagger s_1}^c + \left(-\Gamma_1 - \frac{\kappa_1 + \kappa_2}{2} + i(\Delta_1 - \Delta_2)\right)C_{s_2^\dagger s_1}^c - \kappa_2C_{s_2^\dagger s_1}^cC_{s_2^\dagger s_2}^c - \kappa_2C_{s_1s_2}^cC_{s_2^\dagger s_2^\dagger}^c - i\Lambda\left(2C_{s_2^\dagger s_1}^cC_{s_2^\dagger s_2}^c + C_{s_1s_2}^cC_{s_2^\dagger s_2^\dagger}^c + 2C_{s_2^\dagger s_1}^c\langle\hat{s}_2\rangle^c\langle\hat{s}_2^\dagger\rangle^c + C_{s_1s_2}^c\langle\hat{s}_2^\dagger\rangle^c\right) + ig_2e^{-i\phi_2}C_{s_1s_2}^c - ig_1e^{i\phi_1}C_{s_1^\dagger s_2^\dagger}^c, \quad (\text{B5j})$$

$$C_{s_2^\dagger s_2}^{\dot{c}} = -\Gamma_1C_{s_1^\dagger s_2}^c - \Gamma_1C_{s_2^\dagger s_1}^c + (-\Gamma_1 - \kappa_2)C_{s_2^\dagger s_2}^c - \kappa_2C_{s_2^\dagger s_2}^cC_{s_2^\dagger s_2}^c - \kappa_2C_{s_2s_2}^cC_{s_2^\dagger s_2^\dagger}^c + i\Lambda\left(C_{s_2^\dagger s_2^\dagger}^c\langle\hat{s}_2\rangle^c - C_{s_2s_2}^c\langle\hat{s}_2^\dagger\rangle^c\right) - ig_2e^{i\phi_2}C_{s_2^\dagger s_2^\dagger}^c + ig_2e^{-i\phi_2}C_{s_2s_2}^c, \quad (\text{B5k})$$

$$C_{s_1^\dagger s_1^\dagger}^c = (-\Gamma_1 - \kappa_1 - i2\Delta_1)C_{s_1^\dagger s_1^\dagger}^c - 2\kappa_2 C_{s_1^\dagger s_2}^c C_{s_1^\dagger s_2}^c + ig_1 e^{-i\phi_1} (1 + 2C_{s_1^\dagger s_1}^c), \quad (\text{B5l})$$

$$C_{s_1^\dagger s_2}^c = -\Gamma_1 C_{s_1^\dagger s_1}^c + \left(-\Gamma_1 - \frac{\kappa_1 + \kappa_2}{2} - i(\Delta_1 + \Delta_2)\right) C_{s_1^\dagger s_2}^c - \kappa_2 C_{s_1^\dagger s_2}^c C_{s_2^\dagger s_2}^c - \kappa_2 C_{s_1^\dagger s_2}^c C_{s_2^\dagger s_2}^c \\ - i\Lambda \left(2C_{s_1^\dagger s_2}^c C_{s_2^\dagger s_2}^c + C_{s_1^\dagger s_2}^c C_{s_2^\dagger s_2}^c + 2C_{s_1^\dagger s_2}^c \langle \hat{s}_2 \rangle^c \langle \hat{s}_2^\dagger \rangle^c + C_{s_1^\dagger s_2}^c \langle \hat{s}_2^\dagger \rangle^{c^2}\right) + ig_2 e^{-i\phi_2} C_{s_1^\dagger s_2}^c + ig_1 e^{-i\phi_1} C_{s_2^\dagger s_1}^c, \quad (\text{B5m})$$

$$C_{s_2^\dagger s_2}^c = -2\Gamma_1 C_{s_1^\dagger s_2}^c + (-\Gamma_1 - \kappa_2 - i2\Delta_2)C_{s_2^\dagger s_2}^c - 2\kappa_2 C_{s_2^\dagger s_2}^c C_{s_2^\dagger s_2}^c \\ - i\Lambda \left(C_{s_2^\dagger s_2}^c + 6C_{s_2^\dagger s_2}^c C_{s_2^\dagger s_2}^c + 4C_{s_2^\dagger s_2}^c \langle \hat{s}_2 \rangle^c \langle \hat{s}_2^\dagger \rangle^c + \langle \hat{s}_2^\dagger \rangle^{c^2} + 2C_{s_2^\dagger s_2}^c \langle \hat{s}_2^\dagger \rangle^{c^2}\right) + ig_2 e^{-i\phi_2} (1 + 2C_{s_2^\dagger s_2}^c). \quad (\text{B5n})$$

We can also derive the complete set equations of motion for the unconditional dynamics of the cumulants, i.e. noise-averaged equations of motion, as follows:

$$\langle \dot{\hat{s}}_1 \rangle = -\frac{\Gamma_1 + \kappa_1}{2} \langle \hat{s}_1 \rangle + i\Delta_1 \langle \hat{s}_1 \rangle - ig_1 e^{i\phi_1} \langle \hat{s}_1^\dagger \rangle, \quad (\text{B6a})$$

$$\langle \dot{\hat{s}}_2 \rangle = -\Gamma_1 \langle \hat{s}_1 \rangle + \left(i\Delta_2 - \frac{\kappa_2 + \Gamma_1}{2}\right) \langle \hat{s}_2 \rangle + i\Lambda \left(2C_{s_2^\dagger s_2} \langle \hat{s}_2 \rangle + \langle \hat{s}_2^\dagger \rangle \left(\langle \hat{s}_2 \rangle^2 + C_{s_2 s_2}\right)\right) - ig_2 e^{i\phi_2} \langle \hat{s}_2^\dagger \rangle - \sqrt{\kappa_2} \eta_{d,2} e^{i\phi_{d,2}}, \quad (\text{B6b})$$

$$\langle \dot{\hat{s}}_1^\dagger \rangle = -\frac{\Gamma_1 + \kappa_1}{2} \langle \hat{s}_1^\dagger \rangle - i\Delta_1 \langle \hat{s}_1^\dagger \rangle + ig_1 e^{-i\phi_1} \langle \hat{s}_1 \rangle, \quad (\text{B6c})$$

$$\langle \dot{\hat{s}}_2^\dagger \rangle = -\Gamma_1 \langle \hat{s}_1^\dagger \rangle - \left(i\Delta_2 + \frac{\kappa_2 + \Gamma_1}{2}\right) \langle \hat{s}_2^\dagger \rangle - i\Lambda \left(2C_{s_2^\dagger s_2} \langle \hat{s}_2^\dagger \rangle + \langle \hat{s}_2 \rangle \left(\langle \hat{s}_2^\dagger \rangle^2 + C_{s_2^\dagger s_2^\dagger}\right)\right) + ig_2 e^{-i\phi_2} \langle \hat{s}_2 \rangle - \sqrt{\kappa_2} \eta_{d,2} e^{-i\phi_{d,2}}, \quad (\text{B6d})$$

$$C_{s_1 s_1}^\cdot = (-\Gamma_1 - \kappa_1 + i2\Delta_1)C_{s_1 s_1} - ig_1 e^{i\phi_1} (1 + 2C_{s_1^\dagger s_1}), \quad (\text{B6e})$$

$$C_{s_1 s_2}^\cdot = -\Gamma_1 C_{s_1 s_1} + \left(-\Gamma_1 - \frac{\kappa_1 + \kappa_2}{2} + i(\Delta_1 + \Delta_2)\right) C_{s_1 s_2} \\ + i\Lambda \left(C_{s_2^\dagger s_1} C_{s_2 s_2} + 2C_{s_1 s_2} C_{s_2^\dagger s_2} + C_{s_2^\dagger s_1} \langle \hat{s}_2 \rangle^2 + 2C_{s_1 s_2} \langle \hat{s}_2 \rangle \langle \hat{s}_2^\dagger \rangle\right) - ig_1 e^{i\phi_1} C_{s_1^\dagger s_2} - ig_2 e^{i\phi_2} C_{s_2^\dagger s_1}, \quad (\text{B6f})$$

$$C_{s_2 s_2}^\cdot = -2\Gamma_1 C_{s_1 s_2} + (-\Gamma_1 - \kappa_2 + i2\Delta_2)C_{s_2 s_2} \\ + i\Lambda \left(C_{s_2 s_2} + 6C_{s_2 s_2} C_{s_2^\dagger s_2} + \langle \hat{s}_2 \rangle^2 + 2C_{s_2^\dagger s_2} \langle \hat{s}_2 \rangle^2 + 4C_{s_2 s_2} \langle \hat{s}_2 \rangle \langle \hat{s}_2^\dagger \rangle\right) - ig_2 e^{i\phi_2} (1 + 2C_{s_2^\dagger s_2}), \quad (\text{B6g})$$

$$C_{s_1^\dagger s_1}^\cdot = (-\Gamma_1 - \kappa_1)C_{s_1^\dagger s_1} - ig_1 e^{i\phi_1} C_{s_1^\dagger s_1} + ig_1 e^{-i\phi_1} C_{s_1 s_1}, \quad (\text{B6h})$$

$$C_{s_1^\dagger s_2}^\cdot = -\Gamma_1 C_{s_1^\dagger s_1} + \left(-\Gamma_1 - \frac{\kappa_1 + \kappa_2}{2} + i(-\Delta_1 + \Delta_2)\right) C_{s_1^\dagger s_2} \\ + i\Lambda \left(2C_{s_1^\dagger s_2} C_{s_2^\dagger s_2} + C_{s_1^\dagger s_2} C_{s_2 s_2} + C_{s_1^\dagger s_2} \langle \hat{s}_2 \rangle^2 + 2C_{s_1^\dagger s_2} \langle \hat{s}_2 \rangle \langle \hat{s}_2^\dagger \rangle\right) + ig_1 e^{-i\phi_1} C_{s_1 s_2} - ig_2 e^{i\phi_2} C_{s_1^\dagger s_2^\dagger}, \quad (\text{B6i})$$

$$C_{s_2^\dagger s_1}^\cdot = -\Gamma_1 C_{s_1^\dagger s_1} + \left(-\Gamma_1 - \frac{\kappa_1 + \kappa_2}{2} + i(\Delta_1 - \Delta_2)\right) C_{s_2^\dagger s_1} \\ - i\Lambda \left(2C_{s_2^\dagger s_1} C_{s_2^\dagger s_2} + C_{s_1 s_2} C_{s_2^\dagger s_2} + 2C_{s_2^\dagger s_1} \langle \hat{s}_2 \rangle \langle \hat{s}_2^\dagger \rangle + C_{s_1 s_2} \langle \hat{s}_2^\dagger \rangle^2\right) + ig_2 e^{-i\phi_2} C_{s_1 s_2} - ig_1 e^{i\phi_1} C_{s_1^\dagger s_2^\dagger}, \quad (\text{B6j})$$

$$\begin{aligned}
\dot{C}_{s_2^\dagger s_2} &= -\Gamma_1 C_{s_1^\dagger s_2} - \Gamma_1 C_{s_2^\dagger s_1} + (-\Gamma_1 - \kappa_2) C_{s_2^\dagger s_2} \\
&\quad + i\Lambda \left(C_{s_2^\dagger s_2} \langle \hat{s}_2 \rangle^2 - C_{s_2 s_2} \langle \hat{s}_2^\dagger \rangle^2 \right) - ig_2 e^{i\phi_2} C_{s_2^\dagger s_2} + ig_2 e^{-i\phi_2} C_{s_2 s_2},
\end{aligned} \tag{B6k}$$

$$\dot{C}_{s_1^\dagger s_1} = (-\Gamma_1 - \kappa_1 - i2\Delta_1) C_{s_1^\dagger s_1} + ig_1 e^{-i\phi_1} (1 + 2C_{s_1^\dagger s_1}), \tag{B6l}$$

$$\begin{aligned}
\dot{C}_{s_1^\dagger s_2} &= -\Gamma_1 C_{s_1^\dagger s_1} + \left(-\Gamma_1 - \frac{\kappa_1 + \kappa_2}{2} - i(\Delta_1 + \Delta_2) \right) C_{s_1^\dagger s_2} \\
&\quad - i\Lambda \left(2C_{s_1^\dagger s_2} C_{s_2^\dagger s_2} + C_{s_1^\dagger s_2} C_{s_2^\dagger s_2} + 2C_{s_1^\dagger s_2} \langle \hat{s}_2 \rangle \langle \hat{s}_2^\dagger \rangle + C_{s_1^\dagger s_2} \langle \hat{s}_2^\dagger \rangle^2 \right) + ig_2 e^{-i\phi_2} C_{s_1^\dagger s_2} + ig_1 e^{-i\phi_1} C_{s_2^\dagger s_1},
\end{aligned} \tag{B6m}$$

$$\begin{aligned}
\dot{C}_{s_2^\dagger s_2} &= -2\Gamma_1 C_{s_1^\dagger s_2} + (-\Gamma_1 - \kappa_2 - i2\Delta_2) C_{s_2^\dagger s_2} \\
&\quad - i\Lambda \left(C_{s_2^\dagger s_2} + 6C_{s_2^\dagger s_2} C_{s_2^\dagger s_2} + 4C_{s_2^\dagger s_2} \langle \hat{s}_2 \rangle \langle \hat{s}_2^\dagger \rangle + \langle \hat{s}_2^\dagger \rangle^2 + 2C_{s_2^\dagger s_2} \langle \hat{s}_2^\dagger \rangle^2 \right) + ig_2 e^{-i\phi_2} (1 + 2C_{s_2^\dagger s_2}).
\end{aligned} \tag{B6n}$$

Appendix C: Filtered single-shot measurement mean and variances for linearized systems

Here we present the calculation of filtered mean and covariance results for a more general system with k modes, starting from the measured trajectories. The mean value of filtered single-shot measurements in the infinite-shot limit is given by:

$$\mu_k(t) = \mathbb{E} \left[\begin{pmatrix} \mathcal{I}_k(t) \\ \mathcal{Q}_k(t) \end{pmatrix} \right] = \frac{1}{\sqrt{2\mathcal{T}}} \int_{t-\mathcal{T}}^t d\tau \mathbb{E} \left[\begin{pmatrix} \mathcal{I}_k(\tau) \\ \mathcal{Q}_k(\tau) \end{pmatrix} \right], \tag{C1}$$

with

$$\mathcal{I}_k(t) = \xi_{\mathcal{I}_k}(t) + \sqrt{\gamma_H} \left(\langle \hat{I}_k(t) \rangle + \xi_{\mathcal{I}_k}^{qm}(t) \right) + \sqrt{\bar{n}_{cl}} \xi_{\mathcal{I}_k}^{cl}(t), \tag{C2a}$$

$$\mathcal{Q}_k(t) = \xi_{\mathcal{Q}_k}(t) + \sqrt{\gamma_H} \left(\langle \hat{Q}_k(t) \rangle + \xi_{\mathcal{Q}_k}^{qm}(t) \right) + \sqrt{\bar{n}_{cl}} \xi_{\mathcal{Q}_k}^{cl}(t). \tag{C2b}$$

We should note that $\xi_{\mathcal{I}, \mathcal{Q}_k}(t) = \frac{dW_{\mathcal{I}, \mathcal{Q}}}{dt}$ where $dW_{\mathcal{I}, \mathcal{Q}}$ are the Wiener increments that we introduce in the SQME of the system. Secondly, $\xi_{\mathcal{I}, \mathcal{Q}_k}^{cl}$ are white noise random variables which represent classical noise contribution to the system after Heisenberg-von Neumann cut. Then we plug in Eqs. C2 into Eq. C1 and get:

$$\mu_k(t) = \sqrt{\frac{\gamma_H}{2\mathcal{T}}} \int_{t-\mathcal{T}}^t d\tau \begin{pmatrix} \langle \hat{I}_k(\tau) \rangle \\ \langle \hat{Q}_k(\tau) \rangle \end{pmatrix}. \tag{C3}$$

In the steady-state limit ($t \rightarrow \infty$), $\langle (\hat{I}, \hat{Q})_k(t) \rangle$ will settle to a fixed value $\langle (\hat{I}, \hat{Q})_k \rangle$:

$$\mu_k = \lim_{t \rightarrow \infty} \mu_k(t) = \sqrt{\frac{\gamma_H \mathcal{T}}{2}} \begin{pmatrix} \langle \hat{I}_k \rangle \\ \langle \hat{Q}_k \rangle \end{pmatrix}. \tag{C4}$$

The covariance matrix for modes j and k of filtered single-shot measurements in the infinite-shot limit is as follows:

$$\Sigma_{jk}(t) = \frac{1}{2\mathcal{T}} \int_{t-\mathcal{T}}^t d\tau \int_{t-\mathcal{T}}^t d\tau' \left\{ \mathbb{E} \left[\begin{pmatrix} \mathcal{I}_j(\tau') \\ \mathcal{Q}_j(\tau') \end{pmatrix} \begin{pmatrix} \mathcal{I}_k(\tau) \\ \mathcal{Q}_k(\tau) \end{pmatrix}^T \right] - \mathbb{E} \left[\begin{pmatrix} \mathcal{I}_j(\tau') \\ \mathcal{Q}_j(\tau') \end{pmatrix} \right] \mathbb{E} \left[\begin{pmatrix} \mathcal{I}_k(\tau) \\ \mathcal{Q}_k(\tau) \end{pmatrix} \right]^T \right\} \tag{C5}$$

$$= \frac{1}{2\mathcal{T}} \int_{t-\mathcal{T}}^t d\tau \int_{t-\mathcal{T}}^t d\tau' \hat{s}_{jk}(\tau', \tau), \tag{C6}$$

where

$$\hat{s}_{jk}(\tau', \tau) = \delta_{jk} \delta(\tau' - \tau) (1 + \bar{n}_{cl}) \mathbb{I}_2 + M_j C(\tau', \tau) M_k^T, \quad (C7)$$

with

$$[C(\tau', \tau)]_{ij} = \langle : z_i(\tau') z_j(\tau) : \rangle - \langle z_i(\tau') \rangle \langle z_j(\tau) \rangle, \quad (C8)$$

$$\vec{z} = (b_1 \ b_1^\dagger \ \dots \ b_k \ b_k^\dagger \ \dots \ b_K \ b_K^\dagger)^T, \quad (C9)$$

$$M_k = \sqrt{\gamma_H} \begin{pmatrix} 0 & \dots & \underbrace{U}_{k^{th} \text{ elem}} & \dots & 0 \end{pmatrix}, \quad (C10)$$

$$U = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ -i & i \end{pmatrix}, \quad (C11)$$

and $b_k^{(\dagger)}$ are creation and annihilation operators for intracavity modes, K is the total number of modes in the system. Plugging these in, we get:

$$\Sigma_{jk}(t) = \frac{1}{2} \delta_{jk} (1 + \bar{n}_{cl}) \mathbb{I}_2 + \frac{1}{2\mathcal{T}} \int_{t-\mathcal{T}}^t d\tau \int_{t-\mathcal{T}}^t d\tau' M_j C(\tau', \tau) M_k^T. \quad (C12)$$

For linear and linearized systems the following statements hold:

$$C(\tau + \theta, \tau) = e^{J\theta} C(\tau), \quad (C13)$$

$$C(\tau - \theta, \tau) = C(\tau - \theta) e^{J^T \theta}. \quad (C14)$$

We can break down the double integral into two as:

$$\int_{t-\mathcal{T}}^t d\tau \int_{t-\mathcal{T}}^t d\tau' = \int_{t-\mathcal{T}}^t d\tau \int_{t-\mathcal{T}}^\tau d\tau' + \int_{t-\mathcal{T}}^t d\tau \int_\tau^t d\tau'. \quad (C15)$$

See in the first integral $\tau' < \tau$ so we can define $\tau' = \tau - \theta$ and in the second integral $\tau' > \tau$ so we can define $\tau' = \tau + \theta$ and rewrite the integrals in terms of τ and θ .

$$\begin{aligned} \int_{t-\mathcal{T}}^t d\tau \int_{t-\mathcal{T}}^t d\tau' M_j C(\tau', \tau) M_k^T &= \int_{t-\mathcal{T}}^t d\tau \int_0^{\tau-(t-\mathcal{T})} d\theta M_j C(\tau - \theta, \tau) M_k^T + \int_{t-\mathcal{T}}^t d\tau \int_0^{t-\tau} d\theta M_j C(\tau + \theta, \tau) M_k^T \\ &= \int_{t-\mathcal{T}}^t d\tau \int_0^{\tau-(t-\mathcal{T})} d\theta M_j C(\tau - \theta) e^{J^T \theta} M_k^T + \int_{t-\mathcal{T}}^t d\tau \int_0^{t-\tau} d\theta M_j e^{J\theta} C(\tau) M_k^T. \end{aligned} \quad (C16)$$

Then, we go to the steady-state limit as $t \rightarrow \infty$, $C(t) = C(t - \theta) = C$

$$\int_{t-\mathcal{T}}^t d\tau \int_{t-\mathcal{T}}^t d\tau' M_j C(\tau', \tau) M_k^T = \int_{t-\mathcal{T}}^t d\tau \int_0^{\tau-(t-\mathcal{T})} d\theta M_j C e^{J^T \theta} M_k^T + \int_{t-\mathcal{T}}^t d\tau \int_0^{t-\tau} d\theta M_j e^{J\theta} C M_k^T. \quad (C17)$$

We use the diagonal representation of matrix J to continue our calculation, $J = PNP^{-1}$. With this representation then we rewrite $e^{J\theta} = P e^{N\theta} P^{-1}$ where N is a diagonal matrix

$$\begin{aligned} \int_{t-\mathcal{T}}^t d\tau \int_0^{t-\tau} d\theta e^{J\theta} &= P \left[\int_{t-\mathcal{T}}^t d\tau \int_0^{t-\tau} d\theta e^{N\theta} \right] P^{-1} \\ &= PN^{-1} \int_{t-\mathcal{T}}^t d\tau \left[e^{N(t-\tau)} - \mathbb{I} \right] P^{-1} \\ &= PN^{-2} \left[e^{N\mathcal{T}} - \mathbb{I} \right] P^{-1} - \mathcal{T} J^{-1} \\ &= J^{-2} \left[e^{J\mathcal{T}} - \mathbb{I} \right] - \mathcal{T} J^{-1}, \end{aligned} \quad (C18)$$

where we used $J^n = PN^nP^{-1}$ for all $n \in \mathbb{Z}$. Next, we can use $J^{T^n} = P^{T^{-1}}N^nP^T$.

$$\begin{aligned}
\int_{t-\mathcal{T}}^t d\tau \int_0^{\tau-(t-\mathcal{T})} d\theta e^{J^T\theta} &= (P^T)^{-1} \int_{t-\mathcal{T}}^t d\tau \int_0^{\tau-(t-\mathcal{T})} d\theta e^{N\theta} P^T \\
&= (P^T)^{-1} N^{-1} \int_{t-\mathcal{T}}^t d\tau \left[e^{N(\tau-(t-\mathcal{T}))} - \mathbb{I} \right] P^T \\
&= (P^T)^{-1} N^{-2} \left[e^{N\mathcal{T}} - \mathbb{I} \right] P^T - \mathcal{T}(J^T)^{-1} \\
&= (J^T)^{-2} \left[e^{J^T\mathcal{T}} - \mathbb{I} \right] - \mathcal{T}(J^T)^{-1}.
\end{aligned} \tag{C19}$$

Plugging these in, we get:

$$\Sigma_{jk} = \frac{1}{2} \delta_{jk} (1 + \bar{n}_{cl}) \mathbb{I}_2 - \frac{1}{2} M_j \left[C J^{T^{-1}} + J^{-1} C + \frac{1}{\mathcal{T}} \left(C J^{T^{-2}} \left(\mathbb{I} - e^{J^T\mathcal{T}} \right) + J^{-2} \left(\mathbb{I} - e^{J^T\mathcal{T}} \right) C \right) \right] M_k^T. \tag{C20}$$

In the limit where the integration time is very long ($\mathcal{T} \rightarrow \infty$), we end up with the following expression for the covariance matrix (as the eigenvalues of J have all real negative parts)

$$\begin{aligned}
\lim_{\mathcal{T} \rightarrow \infty} \Sigma_{jk} &= \frac{1}{2} \delta_{jk} (1 + \bar{n}_{cl}) \mathbb{I}_2 - \frac{1}{2} M_j \left[C J^{T^{-1}} + J^{-1} C + \frac{1}{\mathcal{T}} \left(C J^{T^{-2}} + J^{-2} C \right) \right] M_k^T \\
&= \frac{1}{2} \delta_{jk} (1 + \bar{n}_{cl}) \mathbb{I}_2 - \frac{1}{2} M_j \left[C J^{T^{-1}} + J^{-1} C \right] M_k^T + \mathcal{O} \left(\frac{1}{\mathcal{T}} \right).
\end{aligned} \tag{C21}$$

Appendix D: A non-traditional perturbative approach to evaluate cumulants in weak nonlinearity limit

In this section, we present a perturbative treatment on the TEOMs. The correct ansatz of the perturbative expansion depends only on the type of nonlinearity in the system, so in order to find the correct expansion coefficients we start with the TEOMs of a single Kerr oscillator, which is much simpler than the system that we work with but still gives us the correct expansion coefficients:

$$\dot{s} = (i\Delta - \kappa/2) s - \sqrt{\kappa}\eta e^{i\phi_d} + i\Lambda (s^\dagger s s + C_{ss} s^\dagger + 2C_{s^\dagger s} s), \tag{D1a}$$

$$\dot{C}_{ss} = (i2\Delta - \kappa) C_{ss} + i\Lambda (C_{ss} (1 + 6C_{s^\dagger s} + 4s s^\dagger) + s^2 (1 + 2C_{s^\dagger s})), \tag{D1b}$$

$$\dot{C}_{s^\dagger s} = -\kappa C_{s^\dagger s} + i\Lambda (C_{s^\dagger s} s^2 - C_{ss} s^{\dagger 2}). \tag{D1c}$$

Then, we expand the first order cumulants with the following ansatz:

$$s = \Lambda^\alpha s^{(0)} + \Lambda^\beta s^{(1)}, \tag{D2}$$

with $\beta > \alpha$, which guarantees that $s^{(0)}$ is the leading order term. Then we expand the second order cumulants with the following ansatz:

$$C = \Lambda^p C^{(0)} + \Lambda^q C^{(1)}, \tag{D3}$$

similarly with $q > p$, which guarantees that $C_{ss}^{(0)}$ is the leading order term. Now, we plug in these expressions in Eq. (D1):

$$\begin{aligned}
\Lambda^\alpha \dot{s}^{(0)} + \Lambda^\beta \dot{s}^{(1)} &= (i\Delta - \kappa/2) \left(\Lambda^\alpha s^{(0)} + \Lambda^\beta s^{(1)} \right) - \sqrt{\kappa}\eta e^{i\phi_d} \\
&+ i \left[\Lambda^{1+3\alpha} s^{(0)2} s^{\dagger(0)} + \Lambda^{1+2\alpha+\beta} \left(2s^{\dagger(0)} s^{(0)} s^{(1)} + s^{(0)2} s^{\dagger(1)} \right) + \Lambda^{1+\alpha+2\beta} \left(2s^{(0)} s^{(1)} s^{\dagger(1)} + s^{\dagger(0)} s^{(1)2} \right) + \Lambda^{1+3\beta} s^{(1)2} s^{\dagger(1)} \right] \\
&+ i \left[\Lambda^{1+p+\alpha} \left(C_{ss}^{(0)} s^{\dagger(0)} + 2C_{s^\dagger s}^{(0)} s^{(0)} \right) + \Lambda^{1+p+\beta} \left(C_{ss}^{(0)} s^{\dagger(1)} + 2C_{s^\dagger s}^{(0)} s^{(1)} \right) + \Lambda^{1+q+\alpha} \left(C_{ss}^{(1)} s^{\dagger(0)} + 2C_{s^\dagger s}^{(1)} s^{(0)} \right) \right. \\
&\left. + \Lambda^{1+q+\beta} \left(C_{ss}^{(1)} s^{\dagger(1)} + 2C_{s^\dagger s}^{(1)} s^{(1)} \right) \right].
\end{aligned} \tag{D4}$$

The lowest-order term for this equation has Λ^α coefficient and the terms on this order only include leading-order terms for first-order cumulants and a drive term and is equivalent to the normalized equation of motion of the Kerr oscillator in the classical limit:

$$\Lambda^\alpha \left[\dot{s}^{(0)} = (i\Delta - \kappa/2) s^{(0)} - \sqrt{\kappa}\eta\Lambda^{-\alpha} e^{i\phi_d} + i\Lambda^{1+2\alpha} s^{(0)2} s^{\dagger(0)} \right]. \quad (\text{D5})$$

In order to have all the prefactors matching, we need $1 + 2\alpha = 0$, which fixes prefactor of the leading-order mean value:

$$\alpha = -1/2. \quad (\text{D6})$$

Notice that, in this case, the drive is normalized by the nonlinearity $\eta' = \eta\sqrt{\Lambda}$. This means that for a fixed value of η' , one ends up with identical dynamics even if the nonlinearity (or drive) is changed. We should also notice that this coefficient will always be decided by the type of nonlinearity one has in the system dynamics.

Moving forward, by plugging in $\alpha = -1/2$, we write down the second largest terms in this expansion, which are scaled with Λ^β :

$$\Lambda^\beta \left[\dot{s}^{(1)} = (i\Delta - \kappa/2) s^{(1)} + i2s^{\dagger(0)} s^{(0)} s^{(1)} + i s^{(0)2} s^{\dagger(1)} + i\Lambda^{\frac{1}{2}+p-\beta} \left(C_{ss}^{(0)} s^{\dagger(0)} + 2C_{s^\dagger s}^{(0)} s^{(0)} \right) \right]. \quad (\text{D7})$$

Here, we keep the lowest-order term coming from the second-order cumulant contribution, which forces $\frac{1}{2} + p - \beta = 0$. We have a new condition for the perturbative coefficients:

$$\beta = p + \frac{1}{2}. \quad (\text{D8})$$

Now, we plug in our perturbative ansatz into the second-order cumulant equation of motion:

$$\begin{aligned} \Lambda^p \dot{C}_{ss}^{(0)} + \Lambda^q \dot{C}_{ss}^{(1)} = & (i2\Delta - \kappa) \left(\Lambda^p C_{ss}^{(0)} + \Lambda^q C_{ss}^{(1)} \right) + i \left[\Lambda^{p+1} C_{ss}^{(0)} + \Lambda^{q+1} C_{ss}^{(1)} + 6\Lambda^{1+2p} C_{ss}^{(0)} C_{s^\dagger s}^{(0)} \right. \\ & + 6\Lambda^{1+p+q} \left(C_{ss}^{(0)} C_{s^\dagger s}^{(1)} + C_{s^\dagger s}^{(0)} C_{ss}^{(1)} \right) + 6\Lambda^{1+2q} C_{ss}^{(1)} C_{s^\dagger s}^{(1)} + 4\Lambda^p C_{ss}^{(0)} s^{(0)} s^{\dagger(0)} \\ & + 4\Lambda^{\frac{1}{2}+p+\beta} \left(C_{ss}^{(0)} \left(s^{(0)} s^{\dagger(1)} + s^{\dagger(0)} s^{(1)} \right) \right) + 4\Lambda^{1+p+2\beta} C_{ss}^{(0)} s^{(1)} s^{\dagger(1)} + 4\Lambda^q C_{ss}^{(1)} s^{(0)} s^{\dagger(0)} \\ & + 4\Lambda^{\frac{1}{2}+q+\beta} \left(C_{ss}^{(1)} \left(s^{(0)} s^{\dagger(1)} + s^{\dagger(0)} s^{(1)} \right) \right) + 4\Lambda^{1+q+2\beta} C_{ss}^{(1)} s^{(1)} s^{\dagger(1)} \left. \right] + i \left[s^{(0)2} + 2\Lambda^{\beta+\frac{1}{2}} s^{(0)} s^{(1)} \right. \\ & + \Lambda^{1+2\beta} s^{(1)2} + \Lambda^p 2C_{ss}^{(0)} s^{(0)2} + \Lambda^{\frac{1}{2}+p+\beta} 2C_{ss}^{(0)} s^{(0)} s^{(1)} + \Lambda^{1+p+2\beta} 2C_{ss}^{(0)} s^{(1)2} \\ & \left. + \Lambda^q 2C_{ss}^{(1)} s^{(0)2} + \Lambda^{\frac{1}{2}+q+\beta} 2C_{ss}^{(1)} s^{(0)} s^{(1)} + \Lambda^{1+q+2\beta} 2C_{ss}^{(1)} s^{(1)2} \right]. \end{aligned} \quad (\text{D9})$$

We write down the lowest-order components:

$$\Lambda^p \left[\dot{C}_{ss}^{(0)} = (i2\Delta - \kappa) C_{ss}^{(0)} + 4C_{ss}^{(0)} s^{(0)} s^{\dagger(0)} + \Lambda^{-p} s^{(0)2} + 2C_{ss}^{(0)} s^{(0)2} \right]. \quad (\text{D10})$$

This equation fixes the value of p and β , by forcing $-p = 0$:

$$p = 0, \quad (\text{D11})$$

$$\beta = \frac{1}{2}. \quad (\text{D12})$$

Finally, to find q we write down the second lowest-order terms in this equation:

$$\begin{aligned} \Lambda^q \left[\dot{C}_{ss}^{(1)} = & (i2\Delta - \kappa) C_{ss}^{(1)} + i\Lambda^{1-q} 6C_{ss}^{(0)} C_{s^\dagger s}^{(0)} + i\Lambda^{1-q} 4C_{ss}^{(0)} \left(s^{(0)} s^{\dagger(1)} + s^{\dagger(0)} s^{(1)} \right) + i4s^{(0)} s^{\dagger(0)} C_{ss}^{(1)} \right. \\ & \left. + i\Lambda^{1-q} 2s^{(0)} s^{(1)} + i\Lambda^{1-q} 4C_{ss}^{(0)} s^{(0)} s^{(1)} + i2s^{(0)2} C_{ss}^{(1)} \right]. \end{aligned} \quad (\text{D13})$$

This finally fixes q by forcing $1 - q = 0$:

$$q = 1. \quad (\text{D14})$$

We plug these coefficients into Eq. (D2,D3), and then use this ansatz to evaluate the TEOMs for the system in the main text.

Notice that this perturbative expansion gives us hierarchical equations of motion starting from determining the zeroth-order solution of the mean value, which is equivalent to the equation of motion of the system in the classical limit, where the mode amplitude is normalized by the nonlinearity. Also, notice that this is the only nonlinear equation that needs to be solved. Compared to the traditional perturbative expansions, this is a major difference, as this derivation takes into account the effect of nonlinearity in the zeroth-order terms for mode means.

The solution of this equation is then used to determine the linear equations of motion for the zeroth-order contribution to the second-order cumulants. This solution, along with the previous solution, is then used to determine the linear equations of motion for the first-order contribution to the mean. All of these solutions are then used to determine the linear equations of motion for the first-order contribution to the second-order cumulants. This is where we stopped our expansion but one can start with an ansatz including higher-order terms to extend this approximation. The same procedure needs to be done to find the coefficients for the higher-order perturbative terms and the same pattern would apply to those higher-order terms in the same hierarchical fashion.

In the main text result for the performance metrics (Eq. (23,24 and 26)), second-order cumulants are approximated by their zeroth-order contribution only ($C \approx C^{(0)}$), and we denote $s^{(0)}$ as \bar{s} and $C^{(0)}$ as \bar{C} in order to avoid confusion with the superscript denoting the input state type.