



CHALMERS
UNIVERSITY OF TECHNOLOGY

Classical simulation and quantum resource theory of non-Gaussian optics

Downloaded from: <https://research.chalmers.se>, 2026-05-21 07:44 UTC

Citation for the original published paper (version of record):

Hahn, O., Takagi, R., Ferrini, G. et al (2025). Classical simulation and quantum resource theory of non-Gaussian optics. *Quantum*, 9. <http://dx.doi.org/10.22331/q-2025-10-13-1881>

N.B. When citing this work, cite the original published paper.

Classical simulation and quantum resource theory of non-Gaussian optics

Oliver Hahn^{1,2}, Ryuji Takagi², Giulia Ferrini¹, and Hayata Yamasaki^{3,4}

¹Wallenberg Centre for Quantum Technology, Department of Microtechnology and Nanoscience, Chalmers University of Technology, Sweden, SE-412 96 Göteborg, Sweden

²Department of Basic Science, The University of Tokyo, 3-8-1 Komaba, Meguro-ku, Tokyo, 153-8902, Japan

³Department of Physics, Graduate School of Science, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo, 113-0033, Japan

⁴Department of Computer Science, Graduate School of Information Science and Technology, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan

We propose efficient algorithms for classically simulating Gaussian unitaries and measurements applied to non-Gaussian initial states. The constructions are based on decomposing the non-Gaussian states into linear combinations of Gaussian states. We use an extension of the covariance matrix formalism to efficiently track relative phases in the superpositions of Gaussian states. We get an exact simulation algorithm, which costs quadratically with the number of Gaussian states required to represent the initial state, and an approximate simulation algorithm, which costs linearly with the l_1 norm of the coefficients associated with the superposition. We define measures of non-Gaussianity quantifying this simulation cost, which we call the Gaussian rank and the Gaussian extent. From the perspective of quantum resource theories, we investigate the properties of this type of non-Gaussianity measure and compute optimal decompositions for states relevant to continuous-variable quantum computing.

1 Introduction

Quantum computing holds the promise of solving problems that are currently beyond the reach of classical methods. In recent years, continuous-variable (CV) systems have emerged as a prime candidate to implement quantum computation

and have received increased attention due to their inherent noise resilience, showcased by a series of breakthrough experiments [1, 2]. CV systems cannot be described within a finite-dimensional Hilbert space and naturally appear in optical [3], microwave radiation [4, 5], and optomechanical systems [6, 7].

In practical implementations, some operations are easier to perform experimentally than others. While Gaussian operations are relatively straightforward to implement in CV systems, they are not sufficient for achieving universal quantum computing. Moreover, when restricted to Gaussian operations and states, efficient classical simulation becomes possible [8] and tasks such as error correction become impossible [9]. Thus, the incorporation of non-Gaussian resources, such as Fock states, becomes necessary for achieving universal quantum computing and opens the possibility for any potential quantum advantage.

This fact introduces a big challenge in assessing and designing new quantum information protocols, such as error correction schemes. Indeed, quantum mechanical systems are notoriously difficult to classically simulate in general, even more so in the case of CV systems. Aside from such practical use cases, classical simulation algorithms are useful in tackling fundamental problems. They provide a lens to directly study the boundary between the computational power of classical and quantum computational models and help to identify the origin of potential quantum advantages.

The other side of the coin is that adding non-

Oliver Hahn: hahn@g.ecc.u-tokyo.ac.jp

Gaussian resources to otherwise classically simulatable Gaussian systems in turn makes it hard to simulate them classically. Thus, any classical simulation algorithm will be restricted in some way. One way to simulate a CV system is to use the Wigner function [10, 11]. The simulation time scales with the amount of negative parts of the Wigner function. The negativity of the Wigner function [12–14] is a measure of non-Gaussianity, meaning that Gaussian states and processes can be represented with strictly positive Wigner function and thus sampled from efficiently. Hence, this method is intrinsically constrained to circuits with very low or no Wigner negativity. The simulation of specific circuits with large negativity has been considered in Refs. [15–18]. The approaches therein are restricted to the simulation of a Gaussian circuit with input infinitely squeezed GKP “states” encoding stabilizer states, and are not directly applicable to the case of more general input non-Gaussian states, including realistic (finitely-squeezed) GKP states. A further approach relies on decomposing states in the Fock basis with bounded support [19, 20]. The non-Gaussian measure quantifying the simulation overhead of non-Gaussian states using such a decomposition is the stellar rank [21]. However, many states that are of utmost relevance for CV quantum computing, such as GKP and cat states, require a large number or even infinitely many Fock states, leading to scaling issues. The simulator [22] circumvents some of these issues, but its performance is not studied analytically, making an analysis difficult.

In this work, we introduce two classical algorithms for the simulation of CV systems, which are inherently connected with two measures for the resource theory of non-Gaussianity. The idea behind the simulators is to decompose non-Gaussian states into a superposition of Gaussian states and use an efficient Gaussian subroutine. For the subroutine, we develop a technique to compute the relative phase of these Gaussian states using an extension of the covariance matrix formalism. The first algorithm is exact and scales with the number of Gaussian states in the superposition—the Gaussian rank—quadratically. We improve this algorithm by approximating the input state by sparsifying the decomposition, yielding the second algo-

rithm. For the latter, we also develop a method for the fast estimation of the norm of the non-Gaussian state, bringing the scaling down to linear in the l_1 norm of the coefficients in the superposition, or equivalently, the Gaussian extent. Hence, in both algorithms, the simulation cost increases with the amount of non-Gaussianity in the input state, measured by the two measures of non-Gaussianity we introduce.

Our simulation algorithms are particularly suitable for quantum optical setups, where non-Gaussian states are generated and then transformed via Gaussian operations. Although we consider primarily the scenario of input non-Gaussian states followed by Gaussian operations and measurements, the simulators also allow for the simulation of other non-Gaussian circuits through gadgetisation.

The idea of the simulator, through decomposing non-Gaussian states into superpositions of Gaussian states, is based on analogous ideas that were used for the simulation of magic states in qubit systems [23], fermionic [24], and passive linear optics [25].

In the second part, we formally introduce the two measures, the Gaussian rank and the Gaussian extent. These measures quantify the smallest cost to run the two algorithms. This fact gives the Gaussian rank and the Gaussian extent an immediate operational interpretation. Additionally, we thoroughly investigate the properties of the measures. The Gaussian extent is connected to the generalized robustness of non-Gaussianity [26, 27]. This allows us to derive a useful condition for the optimal witness of the generalized robustness. At last, we compute optimal decompositions of states of interest for bosonic error correction, as well as derive bounds for CV protocols such as Gaussian boson sampling and cat state breeding. Our work connects the practical usefulness of an efficient classical simulator with the fundamental investigation of resources required for CV quantum computing.

The rest of this paper is organized as follows. In Sec. 2, we provide an introduction to Gaussian quantum optics. In Sec. 3, we develop a phase-sensitive simulator for Gaussian states and operations based on the covariant matrix formalism. This allows us to compute the overlap between pure Gaussian states, including the phase information, which is not available using the standard

formalism based on the covariance matrix. In Sec. 4, we provide two classical simulation algorithms for non-Gaussian optics. The first algorithm performs exact simulation by using the decomposition of non-Gaussian states into a superposition of Gaussian states. This allows the usage of continuous decompositions as well. The second algorithm improves the runtime scaling by using a low-rank approximation of non-Gaussian states by sparsifying their decomposition in Gaussian states. These low-rank approximations are generally not normalized, so we introduce a fast norm estimation algorithm. In Sec. 5 we introduce two measures in the resource theory of non-Gaussianity — the Gaussian rank and the Gaussian extent — and prove their properties. The simulation cost of the two simulation algorithms we introduced scales with the resource measured by the two new measures. We characterize the Gaussian extent by a robustness measure and find a condition that needs to be obeyed by the optimal decomposition in Gaussian states. Examples and applications of this measure are given in Sec. 6, such as optimal decompositions and a resource-theoretic analysis using the Gaussian extent of Gaussian boson sampling and cat state breeding. Finally, our conclusions are given in Sec. 7.

2 Gaussian quantum optics

Gaussian quantum optics offers a rich area of research. Beyond its vast range of applicability, which spans from quantum metrology to quantum key distribution, an appealing feature of Gaussian quantum optics is that several analytical techniques are available in this regime. Here, we introduce the main notations and formalism used in this paper regarding Gaussian quantum optics.

In this work, we will use the canonical operators q, p with

$$[q, p] = i. \quad (1)$$

Gaussian unitaries are defined as

$$U = e^{iH} \quad (2)$$

$$H = \frac{1}{2} \mathbf{r}^T \mathbf{H} \mathbf{r} + \bar{\mathbf{r}} \mathbf{r} \quad (3)$$

where we use the shorthand notation $\mathbf{r} = (q_1, p_1, \dots, q_n, p_n)^T$ for the vector of canonical operators and $\bar{\mathbf{r}} = (r_{q_1}, r_{p_1}, \dots, r_{q_n}, r_{p_n})^T$ a vector of

real numbers. The matrix \mathbf{H} is a symmetric matrix, while H denotes the operator.

This allows us to define Gaussian states as

$$\rho_G = \frac{e^{-\beta H}}{\text{Tr}[e^{-\beta H}]} \quad (4)$$

including the case $\beta \rightarrow \infty$ being pure states. We let \mathcal{G} denote the set of Gaussian states.

The advantage of Gaussian states is that they are fully determined by the mean

$$\bar{\mathbf{r}} = \text{Tr}[\mathbf{r} \rho] \quad (5)$$

and the covariance matrix

$$\sigma = \text{Tr}[\{(\mathbf{r} - \bar{\mathbf{r}}), (\mathbf{r} - \bar{\mathbf{r}})^T\} \rho] \quad (6)$$

where $\{, \}$ is the anti-commutator, and $\mathbf{r} \mathbf{r}^T$ is the outer product. The requirement on all covariance matrices is

$$\sigma \pm i\Omega \geq 0, \quad (7)$$

where $\Omega = \bigoplus_{j=1}^N \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$. Gaussian operations refer to any operations composed of the preparation of Gaussian states, applications of Gaussian unitaries, and measurement by homodyne or heterodyne detection.

Using the covariance matrix formalism, one can easily compute the dynamics of a quantum state as long as the operations involved are Gaussian. This fact means that Gaussian quantum optics can be efficiently simulated on a classical computer. This implies, in turn, that no exponential quantum advantage can be achieved with only Gaussian quantum circuitry. Below, we summarize a list of ingredients used for the Gaussian simulator in the covariance matrix formalism [28].

Gaussian unitaries Displacement operations by $\bar{\mathbf{r}}'$ have the following action on the mean and the covariance matrix:

$$\bar{\mathbf{r}} \rightarrow \bar{\mathbf{r}} + \bar{\mathbf{r}}', \quad (8)$$

$$\sigma \rightarrow \sigma. \quad (9)$$

A general symplectic transformation S acts as

$$\bar{\mathbf{r}} \rightarrow S\bar{\mathbf{r}}, \quad (10)$$

$$\sigma \rightarrow S\sigma S^T. \quad (11)$$

There exists a parameterization for symplectic transformations [29]. Every symplectic transformation can be decomposed as [30]

$$S = O_1 Z O_2, \quad (12)$$

$$Z = \bigoplus_{i=1}^n \begin{pmatrix} z_i & 0 \\ 0 & z_i^{-1} \end{pmatrix}, \quad (13)$$

with $O_1, O_2 \in SP_{2n, \mathbb{R}} \cap SO(2n)$, which are the passive symplectic transformations. So we can decompose all symplectic transformations in terms of single-mode squeezers, phase shifters, and two-mode beam splitters. Note that $SP_{2n, \mathbb{R}} \cap SO(2n)$ is isomorphic to $U(n)$

$$O_i = \bar{U}'^\dagger \begin{pmatrix} U^* & 0_n \\ 0_n & U \end{pmatrix} \bar{U}' \quad (14)$$

with $U \in U(n)$ and

$$\bar{U}' = \frac{1}{\sqrt{2}} \begin{pmatrix} \mathbf{1}_n & i\mathbf{1}_n \\ \mathbf{1}_n & -i\mathbf{1}_n \end{pmatrix}. \quad (15)$$

Tensor product and partial trace Suppose that we have a two-mode Gaussian state with

$$\bar{\mathbf{r}} = \bar{\mathbf{r}}_A \oplus \bar{\mathbf{r}}_B = \begin{pmatrix} \bar{\mathbf{r}}_A \\ \bar{\mathbf{r}}_B \end{pmatrix}, \quad (16)$$

$$\sigma = \sigma_A \oplus \sigma_B = \begin{pmatrix} \sigma_A & 0 \\ 0 & \sigma_B \end{pmatrix}. \quad (17)$$

Then, the partial trace on the second subsystem yields

$$\bar{\mathbf{r}} \rightarrow \bar{\mathbf{r}}_A, \quad (18)$$

$$\sigma \rightarrow \sigma_A. \quad (19)$$

Gaussian completely positive and trace-preserving (CPTP) map Given a Gaussian initial state with covariance σ and mean $\bar{\mathbf{r}}$, the evolution of a Gaussian CPTP map is characterized by two real matrices X, Y and a vector \mathbf{D} as

$$\bar{\mathbf{r}} \rightarrow X\bar{\mathbf{r}} + \mathbf{D}, \quad (20)$$

$$\sigma \rightarrow X\sigma X^T + Y, \quad (21)$$

with the requirement that

$$Y + i\Omega \geq iX\Omega X^T. \quad (22)$$

Gaussian measurements Measurement results of homodyne detection are determined by the mean and covariance matrix, e.g.,

$$p(x) = \frac{e^{-\frac{(x-\bar{x})^2}{\sigma_1}}}{\sqrt{\pi\sigma_1}}. \quad (23)$$

More generally, the POVMs of a general-dyne detection are given with the completeness condition

$$\mathbb{1}_n = \frac{1}{(2\pi)^{2n}} \int_{\mathbb{R}^{2n}} d\mathbf{r}_m D(-\mathbf{r}_m) S^\dagger |0\rangle\langle 0| SD(\mathbf{r}_m) \quad (24)$$

with the outcomes \mathbf{r}_m . The probability density $p(\mathbf{r}_m)$ of measuring a Gaussian state ρ with σ and $\bar{\mathbf{r}}$ can then be inferred from the expression of Gaussian overlaps

$$p(\mathbf{r}_m) = \frac{\langle \psi_G | \rho | \psi_G \rangle}{(2\pi)^n} = \frac{e^{-(\mathbf{r}_m - \bar{\mathbf{r}})^T (\sigma + SS^T)^{-1} (\mathbf{r}_m - \bar{\mathbf{r}})}}{\pi^n \sqrt{\det(\sigma + SS^T)}}, \quad (25)$$

where $|\psi_G\rangle = D(-\mathbf{r}_m) S^\dagger |0\rangle$ from the POVM. One retrieves homodyne detection in certain limits; e.g, for $n = 1$, one measures q in the limit $SS^T = \lim_{z \rightarrow 0} \text{diag}(z^2, 1/z^2)$ and p for $z \rightarrow \infty$.

One can generalize this notion to noisy measurements, such as measurements with finite detection efficiency. The POVMs are then given with the completeness condition

$$\mathbb{1}_n = \frac{1}{(2\pi)^{2n}} \int_{\mathbb{R}^{2n}} d\mathbf{r}_m D(-\mathbf{r}_m) \rho_m D(\mathbf{r}_m) \quad (26)$$

for the outcomes \mathbf{r}_m and ρ_m having covariance matrix σ_m . The probability density of obtaining result \mathbf{r}_m is then

$$p(\mathbf{r}_m) = \frac{1}{(2\pi)^n} \text{Tr} [\rho D(-\mathbf{r}_m) \rho_m D(\mathbf{r}_m)] \quad (27)$$

or equivalently if ρ is a Gaussian state with covariance σ and mean $\bar{\mathbf{r}}$

$$p(\mathbf{r}_m) = \frac{e^{-(\mathbf{r}_m - \bar{\mathbf{r}})^T (\sigma + \sigma_m)^{-1} (\mathbf{r}_m - \bar{\mathbf{r}})}}{\pi^n \sqrt{\det[\sigma + \sigma_m]}}. \quad (28)$$

Conditional Gaussian dynamics We consider a bipartite Gaussian state

$$\bar{\mathbf{r}} = \begin{pmatrix} \bar{\mathbf{r}}_A \\ \bar{\mathbf{r}}_B \end{pmatrix}, \quad (29)$$

$$\sigma = \begin{pmatrix} \sigma_A & \sigma_{AB} \\ \sigma_{AB}^T & \sigma_B \end{pmatrix}. \quad (30)$$

The n -mode subsystem of an $(n+m)$ -mode Gaussian state with the covariance and mean defined above undergoes the following mapping upon general-dyne measurement of the m modes characterized by measurement outcome \mathbf{r}_m and covariance σ_m :

$$\bar{\mathbf{r}}_A \rightarrow \bar{\mathbf{r}}_A - \sigma_{AB} \frac{1}{\sigma_B + \sigma_m} (\mathbf{r}_m - \bar{\mathbf{r}}_B), \quad (31)$$

$$\sigma_A \rightarrow \sigma_A - \sigma_{AB} \frac{1}{\sigma_B + \sigma_m} \sigma_{AB}^T. \quad (32)$$

3 Phase-sensitive simulator for Gaussian states and operations

First, we consider the simulation when all states and operations are Gaussians. The task is to compute the Born probability of obtaining a sample x while performing heterodyne detection of the Gaussian state $|G\rangle$

$$P(x) = |\langle G|x\rangle|^2. \quad (33)$$

In principle, this can easily be computed using the covariance formalism, as it is well known that Gaussian states and operations are classically simulable efficiently.

However, the issue is that the conventional way of simulating Gaussians as summarized in Sec. 2 is based on the covariance matrix formalism and thus does not compute global phases. By contrast, for our non-Gaussian simulator, as we will see later, we need to exploit a Gaussian simulator that keeps track of the phases, which we will de-

velop here inspired by Ref. [31]. A more in-depth derivation can be found in Appendix A.

The most important ingredient is to compute the inner product of Gaussian pure states

$$\langle G_i|G_j\rangle \quad \text{for } |G_i\rangle, |G_j\rangle \in \mathcal{G}, \quad (34)$$

in a phase-sensitive way; i.e., our method does not only provide $|\langle G_i|G_j\rangle|$ but also clarifies $\theta \in [0, 2\pi)$ for $\langle G_i|G_j\rangle = e^{i\theta} |\langle G_i|G_j\rangle|$. We fix a Gaussian reference state $G_0 = |G_0\rangle\langle G_0|$, which can be an arbitrary Gaussian state.

Given the reference state $G_0 = |G_0\rangle\langle G_0|$ and our two Gaussian states of interest $G_1 = |G_1\rangle\langle G_1|$ and $G_2 = |G_2\rangle\langle G_2|$, we deduce the phase sensitive overlap between $\langle G_1|G_2\rangle$ via

$$\text{Tr}(G_0 G_1 G_2) = \text{Tr}(|G_0\rangle\langle G_0| |G_1\rangle\langle G_1| |G_2\rangle\langle G_2|) \quad (35)$$

$$= \langle G_2|G_0\rangle \langle G_1|G_2\rangle \langle G_0|G_1\rangle. \quad (36)$$

The fidelity between an n -mode Gaussian state ρ_0 with covariance σ_0 and mean \mathbf{x}_0 and an n -mode pure Gaussian state $\rho_1 = |\phi\rangle\langle\phi|$ with covariance σ_1 and mean \mathbf{x}_1 is given by [32]

$$\mathcal{F}(\rho_0, |\phi\rangle\langle\phi|) = \langle\phi|\rho_0|\phi\rangle = 2^n \frac{e^{-(\mathbf{d}^T(\sigma_0 + \sigma_1)^{-1}\mathbf{d})}}{\sqrt{\det[\sigma_0 + \sigma_1]}} \quad (37)$$

with $\mathbf{d} = \mathbf{x}_0 - \mathbf{x}_1$. Then, as shown in Appendix A, the right-hand side of (36) can be calculated using the covariance matrices σ_i and means $\boldsymbol{\mu}_i$ by

$$\langle G_2|G_0\rangle \langle G_1|G_2\rangle \langle G_0|G_1\rangle = \frac{1}{\sqrt{\det(\sigma_1 + \sigma_2)}} \frac{1}{\sqrt{\det(\sigma_0 + \Delta)}} e^{-\frac{1}{4}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^T (\sigma_1 + \sigma_2)^{-1} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)} \quad (38)$$

$$e^{-\frac{1}{4}(\boldsymbol{\mu}_0 - \boldsymbol{\mu}_\Delta)^T (\sigma_0 + \Delta)^{-1} (\boldsymbol{\mu}_0 - \boldsymbol{\mu}_\Delta)}, \quad (39)$$

where we write

$$\Delta = \sigma_2 - \frac{(\sigma_2 - i\Omega)^T}{2} (\sigma_1 + \sigma_2)^{-1} \frac{(\sigma_2 - i\Omega)}{2}, \quad (40)$$

$$\boldsymbol{\mu}_\Delta = \boldsymbol{\mu}_2 + \frac{1}{2}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)(\sigma_1 + \sigma_2)^{-1} \frac{(\sigma_2 - i\Omega)}{2}. \quad (41)$$

In order to obtain the desired inner product $\langle G_1|G_2\rangle$, we need to use the inner products with the reference states. Therefore, we not only spec-

ify all Gaussian states with covariance matrix σ_i and mean $\boldsymbol{\mu}_i$, but we also use the inner product with the reference state, which we indicate as o_i ;

as a whole, we will henceforth represent a pure Gaussian state as

$$|G_i\rangle = |\sigma_i, \boldsymbol{\mu}_i, o_i\rangle \quad (42)$$

when necessary. We can then use the tools presented in Sec. 2 to update the covariance and mean of the Gaussian state and then perform a Gaussian measurement by computing the overlap with the corresponding Gaussian state in a phase-sensitive manner. We can directly give the covariance matrices and means to compute the overlap with the reference state for a single mode; more generally, we give an algorithm to do this computation in Appendix B. Also in Appendix C, we give based on the article [33] an alternative way to simulate Gaussian overlaps in a phase sensitive way using the stellar formalism.

4 Classical simulation algorithm for non-Gaussian optics

In this section, we present our non-Gaussian simulation algorithm using the phase-sensitive Gaussian simulator introduced in Sec. 3. Our interest is the classical simulation of representative classes of operations in non-Gaussian optics that can be written as combinations of Gaussian operations with auxiliary non-Gaussian states, e.g., quantum computation using Gottesman-Kitaev-Preskill (GKP) codes [34–36] and cubic phase gates implemented by gate teleportation [34, 37].

As a simple case, we begin by considering non-Gaussian states decomposed into a sum of Gaussian states as

$$|\psi\rangle = \sum_{i=1}^x c_i |G_i\rangle, \quad (43)$$

where we used the notation $|G_i\rangle = |\sigma_i, \boldsymbol{\mu}_i, o_i\rangle$ as in Eq. (42). The Born probability that we want to estimate then has the form

$$P(x) = \frac{\langle\psi|x\rangle\langle x|\psi\rangle}{\|\psi\|^2} = \frac{\sum_{i,j=1}^x c_j^* c_i \langle G_j|x\rangle\langle x|G_i\rangle}{\|\psi\|^2}, \quad (44)$$

where we perform heterodyne detection and $|x\rangle$ is then just a tensor product of coherent states. We will then argue how to approximate arbitrary non-Gaussian states by such a finite sum.

4.1 Exact simulation

The first simulation algorithm computes the Born probability exactly by computing all overlaps in Eq. (44). Thus, to simulate Gaussian operations on non-Gaussian states, we update the individual Gaussian states using the tools of Sec. 2. Then we apply the phase-sensitive Gaussian simulator and the tools of Sec. 3 for each term $\langle x|G_i\rangle$ in Eq. (44), where we need to take into account the relative phase between the terms. The cost of exactly estimating the Born probability requires $\mathcal{O}(\chi^2)$ inner products.

4.2 Approximate simulation

We can improve this simulation method by introducing an approximate version of the latter, by considering sampling terms $|\sigma_i, \boldsymbol{\mu}_i, o_i\rangle$ in the decomposition Eq. (43) to sparsify it. The intuition behind this is to find low rank approximations and therefore speed up the computation. Using a sparsified state allows one to use continuous decompositions of the form

$$|\psi\rangle = \int dk c(k) |\sigma_k, \boldsymbol{\mu}_k, o_k\rangle \quad (45)$$

as well. The sampling procedure becomes apparent if we rewrite the decomposition of the state ψ as follows

$$|\psi\rangle = \sum_{i=1}^x c_i |\sigma_i, \boldsymbol{\mu}_i, o_i\rangle \quad (46)$$

$$= \|c\|_1 \sum_{i=1}^x p(i) |\tilde{\sigma}_i, \tilde{\boldsymbol{\mu}}_i, \tilde{o}_i\rangle, \quad (47)$$

where $|\tilde{\sigma}_i, \tilde{\boldsymbol{\mu}}_i, \tilde{o}_i\rangle := (c_i/|c_i|) |\sigma_i, \boldsymbol{\mu}_i, o_i\rangle$ with the probability distribution

$$p(i) := \frac{|c_i|}{\|c\|_1}. \quad (48)$$

The same holds for the continuous decomposition as well, with

$$|\psi\rangle = \|c\|_1 \int dk p(k) |\tilde{\sigma}_k, \tilde{\boldsymbol{\mu}}_k, \tilde{o}_k\rangle \quad (49)$$

with the probability distribution

$$p(k) := \frac{|c(k)|}{\|c\|_1}. \quad (50)$$

Thus if we sample Gaussian state $|\tilde{\sigma}_i, \tilde{\boldsymbol{\mu}}_i, \tilde{\delta}_i\rangle$ from the decomposition of $|\psi\rangle$ with probability $p(i)$, then the expectation is the state $|\psi\rangle$

$$|\psi\rangle = \|c\|_1 \mathbb{E}_{i \sim p(i)} [|\tilde{\sigma}_i, \tilde{\boldsymbol{\mu}}_i, \tilde{\delta}_i\rangle], \quad (51)$$

where we may omit the subscript of the expectation value $\mathbb{E}_{i \sim p(i)}$ to write \mathbb{E} if it is obvious from the context.

Sampling from this probability distribution k times, we get a sparsified state—a low rank approximation of state $|\psi\rangle$ —

$$|\Omega\rangle = \frac{\|c\|_1}{k} \sum_{i=1}^k |\tilde{\sigma}_i, \tilde{\boldsymbol{\mu}}_i, \tilde{\delta}_i\rangle. \quad (52)$$

As a consequence, we have that

$$\mathbb{E}[\langle\Omega|\psi\rangle] = \mathbb{E}[\langle\psi|\Omega\rangle] = 1. \quad (53)$$

If we use the sparsified state $|\Omega\rangle$ instead of the state $|\psi\rangle$ in the computation the error is upper bounded by

$$\begin{aligned} & \mathbb{E}[|\langle\psi|\psi\rangle - \langle\Omega|\Omega\rangle|^2] \\ &= \mathbb{E}[\langle\Omega|\Omega\rangle] + \mathbb{E}[\langle\psi|\psi\rangle] - \mathbb{E}[\langle\psi|\Omega\rangle] - \mathbb{E}[\langle\Omega|\psi\rangle] \end{aligned} \quad (54)$$

$$\leq \frac{\|c\|_1^2}{k}. \quad (55)$$

Thus, by choosing the number

$$k = \left(\frac{\|c\|_1}{\delta} \right)^2 \quad (56)$$

of sampled Gaussian states, we can upper-bound

$$\mathbb{E}[\|\langle\psi|\psi\rangle - \langle\Omega|\Omega\rangle\|^2] \leq \delta^2. \quad (57)$$

The minimum sampling cost for a given state $|\psi\rangle$ and error δ in Eq. (56) for each non-Gaussian state is characterized by the Gaussian extent in Eq. (81). Using the sparsified state in estimating the Born probability scales linearly with χ instead of quadratically in the exact case [38]. For further details and an alternative sampling strategy introduced in Ref. [30], please consult Appendix D.

However, in order to compute the Born probability in Eq. (44) approximately by using a sparsified state $|\Omega\rangle$ defined in Eq. (52), we need to compute the norm of the state. This is necessary because the sparsification will lead to an unnormalized state $|\Omega\rangle$. Given a decomposition (52) of

$|\Omega\rangle$, the required time steps for a straightforward computation of its norm would grow quadratically in the number of summands of the decomposition. By contrast, the procedure here, called the fast norm estimation, computes the norm of a sparsified state with a cost that scales linearly rather than quadratically in the number of Gaussian states in the superposition. This allows for an overall linear scaling in the χ —the number of Gaussian states in the decomposition—for the non-Gaussian simulator. In the following, we present the procedure of fast norm estimation for non-Gaussian states. A more detailed derivation can be found in Appendix E.

It is known that coherent states and displacement operators in CV cases can be used analogously to the 1-design of multiqubit cases. We can represent the identity using coherent states or displacement operators as

$$\mathbb{1} = \frac{1}{\pi} \int_{\mathbb{C}} d\alpha |\alpha\rangle\langle\alpha| = \frac{1}{\pi} \int_{\mathbb{C}} d\alpha D(\alpha) |0\rangle\langle 0| D^\dagger(\alpha). \quad (58)$$

This means that, using a uniformly weighted integral over all coherent states $|\boldsymbol{\xi}\rangle = D(\boldsymbol{\xi})|0\rangle$ for X defined as

$$X := \pi^{-n} |\langle\boldsymbol{\xi}|\Omega\rangle|^2, \quad (59)$$

we have that

$$\int_{\mathbb{C}^n} d\boldsymbol{\xi} X = \langle\Omega|\Omega\rangle, \quad (60)$$

where $D(\boldsymbol{\xi})$ is integrated over the non-weighted choices of $\boldsymbol{\xi}$. Consequently, if we can approximate this integral by some quadrature, computing the overlap between a random coherent state and the state $|\Omega\rangle$ can be accomplished in time scaling linearly with the number of Gaussian states in the decomposition of $|\Omega\rangle$.

Yet problematically, the distribution of all displacements is not compact, and it is therefore impossible to sample from a uniform distribution over it. To resolve this problem, similar to Ref. [39], we use sampling from a Gaussian ensemble

$$\mathcal{D}_N = \left\{ D(\boldsymbol{\xi}) : \boldsymbol{\xi} \sim P_D^G(\boldsymbol{\xi}, N) = \frac{e^{-|\boldsymbol{\xi}|^2/N}}{\pi^n N^n} \right\}, \quad (61)$$

which reproduces the identity in the limit of

$$\lim_{N \rightarrow \infty} N^n \int_{\mathbb{C}^n} d\boldsymbol{\xi} P_D^G(\boldsymbol{\xi}, N) D(\boldsymbol{\xi}) |0\rangle\langle 0| D^\dagger(\boldsymbol{\xi}) = \mathbb{1}, \quad (62)$$

and it also holds that

$$\begin{aligned} & \lim_{N \rightarrow \infty} N^n \int_{\mathbb{C}^n} d\xi P_D^G(\xi, N) D(\xi)^{\otimes 2} |0 \otimes 0\rangle \langle 0 \otimes 0| D^\dagger(\xi)^{\otimes 2} \\ &= \frac{1}{\pi^n} \int_{\mathbb{C}^n} d\xi D(\xi)^{\otimes 2} |0 \otimes 0\rangle \langle 0 \otimes 0| D^\dagger(\xi)^{\otimes 2}. \end{aligned} \quad (63)$$

In particular, these limits mean that, for any $\delta >$

0, there is a sufficiently large N_δ such that, for a state $|\Omega\rangle$ of interest,¹

$$\left| N_\delta^n \int_{\mathbb{C}^n} d\xi P_D^G(\xi, N_\delta) \langle \Omega | D(\xi) |0\rangle \langle 0| D^\dagger(\xi) | \Omega \rangle - \langle \Omega | \Omega \rangle \right| \leq \delta \langle \Omega | \Omega \rangle, \quad (64)$$

and

$$\begin{aligned} & \left| \pi^n N_\delta^n \int_{\mathbb{C}^n} d\xi P_D^G(\xi, N_\delta) \langle \Omega \otimes \Omega | D(\xi) \otimes D(\xi) |0 \otimes 0\rangle \langle 0 \otimes 0| D^\dagger(\xi) \otimes D^\dagger(\xi) | \Omega \otimes \Omega \rangle \right. \\ & \left. - \int_{\mathbb{C}^n} d\xi \langle \Omega \otimes \Omega | (D(\xi) \otimes D(\xi)) |0 \otimes 0\rangle \langle 0 \otimes 0| (D^\dagger(\xi) \otimes D^\dagger(\xi)) | \Omega \otimes \Omega \rangle \right| \leq \delta \langle \Omega \otimes \Omega | \Omega \otimes \Omega \rangle. \end{aligned} \quad (65)$$

Based on Eq. (64), we define the random variable as

$$X = N_\delta^n |\langle \xi | \Omega \rangle|^2. \quad (66)$$

Thus, if we sample random displacements or, equivalently, coherent states from the ensemble in Eq. (61) with $N = N_\delta$ and compute the over-

lap with the state of interest $|\Omega\rangle$, we can estimate the norm of $|\Omega\rangle$ within the desired target error δ .

In order to bound the number of samples needed to estimate the norm, we will use Chebyshev's inequality and thus need to bound the variance of X . Due to (65), the variance is upper-bounded by

$$\text{Var}[X] \leq \mathbb{E}[X^2] \quad (67)$$

$$= N_\delta^{2n} \int_{\mathbb{C}^n} d\xi \frac{e^{-|\xi|^2/N_\delta}}{\pi^n N_\delta^n} \langle \Omega \otimes \Omega | D(\xi) \otimes D(\xi) |0 \otimes 0\rangle \langle 0 \otimes 0| D^\dagger(\xi) \otimes D^\dagger(\xi) | \Omega \otimes \Omega \rangle \quad (68)$$

$$\leq \frac{N_\delta^n}{\pi^n} 2^{-n} \langle \Omega \otimes \Omega | \Pi | \Omega \otimes \Omega \rangle + \delta \langle \Omega \otimes \Omega | \Omega \otimes \Omega \rangle \quad (69)$$

$$\leq \frac{2^{-n} N_\delta^n + \delta \pi^n}{\pi^n} \|\Omega\|^4. \quad (70)$$

where Π is a projector.

Thus, by performing IID sampling of L coherent states from the Gaussian ensemble in Eq. (61), we can estimate the overlap with the state $|\Omega\rangle$ for sufficiently large N_δ . In particular, we can define the estimator as

$$\eta = \frac{1}{L} \sum_{i=1}^L N_\delta^n |\langle \alpha_i | \Omega \rangle|^2, \quad (71)$$

¹A rigorous mathematical proof of the interchange of limit and integral in Eqs. (62) and (63) is left for future work. The scope of this paper is a physical regime where Eqs. (62) and (63) can be assumed, as in Ref. [39].

where $|\alpha_i\rangle$ is the coherent state for the i th sampling. Then, η estimates the expectation value $\|\Omega\|^2$ and has a variance that is upper bounded by $\sigma^2 \leq L^{-1}(2^{-n} N_\delta^n + \delta/\pi^n)/\pi^n \|\Omega\|^4$. Using Chebyshev's inequality and choosing the number of samples as

$$L = \frac{2^{-n} N_\delta^n + \delta \pi^n}{\pi^n} \epsilon^{-2} p_f^{-1}, \quad (72)$$

we can conclude that the estimator is bounded, with a probability of at least $1 - p_f$, by

$$(1 - \epsilon - \delta) \|\Omega\|^2 \leq \eta \leq (1 + \epsilon + \delta) \|\Omega\|^2. \quad (73)$$

See also Appendix E for more details.

The bounds we investigated can be improved using physical insights in the states we want to simulate. By taking into account the mean photon number of the sparsified state $|\Omega\rangle$, that is $N_\Omega = \frac{\langle \Omega | n | \Omega \rangle}{\langle \Omega | \Omega \rangle} = \frac{1}{\pi^n \langle \Omega | \Omega \rangle} \int_{\mathbb{C}^n} d\xi |\xi|^2 |\langle \xi | \Omega \rangle|^2$, we can improve the fast norm algorithm. See Appendix F for more details. By explicitly using the mean photon number N_Ω of the state $|\Omega\rangle$, we can bound the expectation value of X by

$$\langle \Omega | \Omega \rangle \left(1 - \frac{N_\Omega}{N} \right) \leq \mathbb{E}(X) \leq \langle \Omega | \Omega \rangle, \quad (74)$$

and the variance by

$$\text{Var}[X] \leq \mathbb{E}[X^2] \leq \frac{N^n}{2^n \pi^n} \|\Omega\|^4. \quad (75)$$

Using the same techniques as before, we obtain that with using

$$L = \frac{2^{-n} N_\Omega^n}{\pi^n} \epsilon^{-2} p_f^{-1} \quad (76)$$

samples the estimator η is bounded with a probability of at least $1 - p_f$ by

$$(1 - \epsilon - \frac{N_\Omega}{N}) \|\Omega\|^2 \leq \eta \leq (1 + \epsilon) \|\Omega\|^2. \quad (77)$$

In the end, the fastnorm algorithm scales linearly in the number of terms required to represent the non-Gaussian state; however, the runtime now depends on the mean photon number of the state of interest.

Note that whereas we presented our non-Gaussian simulator for pure states, the simulation works for the mixed states with minor modifications; that is for a state $\rho = \sum_i p_i |\psi_j\rangle\langle\psi_j|$ one starts by sampling $|\psi_j\rangle$ with p_i and then runs the simulation for the pure state $|\psi_j\rangle$.

4.3 Comparison

In this subsection, we compare the algorithms presented in this manuscript with other state-of-the-art methods.

The most commonly used simulation technique to treat infinite-dimensional systems is based on Fock space expansions. These methods are very versatile since states and operators can be expressed as matrices, and then standard linear algebra tools can be applied. The big downside with this approach is that the matrices tend

to get large, especially when one needs to treat multi-mode systems. Using this approach, even Gaussian states become intractable very quickly. Additionally, one makes approximations from the start of the simulation. Our simulation algorithm circumvents these problems. The algorithm in this manuscript can always treat Gaussian states efficiently, and there is no *a priori* increase in simulation complexity by considering more modes.

The algorithm presented in [25] is similar to the exact algorithm presented in this manuscript. Reference [25] considers decompositions in coherent states and allows passive Gaussian unitaries. Our algorithms allow for a larger class of operations that do not increase the simulation cost. Furthermore, the decomposition of states using Gaussian states instead of coherent states will lead to the same or smaller simulation cost since coherent states are a subset of Gaussian states. Furthermore, the approximate simulation algorithm leads to an improved scaling to linear in the number of terms in the decomposition.

5 Measures of non-Gaussianity based on Gaussian decomposition

It is known that the coherent states and, therefore, the Gaussian state form an over-complete basis. The references [40, 41] study which subset of the coherent states form a complete basis. With such a basis, we can expand an arbitrary state $|\psi\rangle$ as

$$|\psi\rangle = \sum_{m,n} c_{m,n} |\alpha_{m,n}\rangle \quad (78)$$

for a coherent state $|\alpha_{m,n}\rangle$ with $\alpha = m\omega_1 + n\omega_2$, where the complex numbers ω_1, ω_2 span a cell with area π . The simplest example is $\alpha_{m,n} = \sqrt{\pi}(m + in)$. This countably infinite subset of the coherent state is complete. Thus, one can, in principle, express a quantum state $|\psi\rangle$ in a countable infinite basis. We relax this notation and not only optimize over the set of coherent states, but arbitrary Gaussian states. Since the set of coherent states is a proper subset of Gaussian states, expanding states in a superposition of Gaussian states is possible, but may be impractical for some states/hard to find.

We can therefore represent any arbitrary state

as a sum of Gaussian states

$$|\psi\rangle = \sum_{k=1}^{\chi} c_k |\sigma_k, \boldsymbol{\mu}_k\rangle \quad (79)$$

where we label each Gaussian state with covariance σ_k and mean $\boldsymbol{\mu}_k$, and $\chi \in \{1, 2, \dots, \infty\}$. We have seen before that the exact simulation algorithms scale with χ^2 , while the approximate algorithm linearly in $\|c\|_1^2$. Decomposing arbitrary states in superposition of Gaussian states allows us to define quantifiers of non-Gaussianity as in the framework of quantum resource theories [42, 43]. In analogy to the resource theory of magic [23, 38, 44], we define the Gaussian rank as

$$\chi(|\psi\rangle) := \inf \left\{ \chi : |\psi\rangle = \sum_{k=1}^{\chi} c_k |\sigma_k, \boldsymbol{\mu}_k\rangle \right\}, \quad (80)$$

and the Gaussian extent as

$$\xi(|\psi\rangle) := \inf \left\{ \|c\|_1^2 : |\psi\rangle = \sum_k c_k |\sigma_k, \boldsymbol{\mu}_k\rangle \right\}. \quad (81)$$

We allow in the definition of the Gaussian rank and the Gaussian extent countably infinitely many terms in the decomposition. A decomposition of quantum states in coherent states was instead considered in the context of the resource theory of non-classicality [45].

Sometimes it is more convenient to decompose the state in continuous decomposition instead of sums. We have seen that the approximate simulation algorithm works equivalently if one chooses a continuous decomposition. Thus we can use the integral in place of the sum in the definition of these measures to take into account continuous decomposition, i.e.,

$$\xi'(|\psi\rangle) := \inf \left\{ \|c\|_1^2 : |\psi\rangle = \int dk c_k |\sigma_k, \boldsymbol{\mu}_k\rangle \right\}. \quad (82)$$

We leave the analysis of such measures for future work. The Gaussian extent can always be used to upper-bound the approximate Gaussian rank, which is defined as

$$\chi_\delta(|\psi\rangle) := \inf \{ \chi(|\psi'\rangle) : \|\psi - \psi'\| < \delta \}. \quad (83)$$

The upper bound is then given as

$$\chi_\delta(|\psi\rangle) \leq 1 + \frac{\xi(|\psi\rangle)}{\delta^2}, \quad (84)$$

which follows from Eq. (57) that extends the argument in Ref. [23] for magic states to non-Gaussian states. The two measures for the resource theory of non-Gaussianity we have introduced here have a powerful operational meaning; namely, they quantify the simulation overhead introduced by the non-Gaussian nature. These measures can be naturally extended to density matrices, e.g.,

$$\Xi(\rho) = \inf \left\{ \sum_j p(j) \xi(|\psi_j\rangle) : \rho = \sum_j p(j) |\psi_j\rangle\langle\psi_j| \right\}, \quad (85)$$

where the minimization is taken over all possible ensembles $\{p(j), |\psi_j\rangle\}$ such that $\rho = \sum_j p_j |\psi_j\rangle\langle\psi_j|$ and $|\psi_j\rangle\langle\psi_j|$ are pure quantum states as in the conventional technique called convex roof extensions [42].

It is possible to use the continuous decomposition as well, i.e.,

$$\Xi'(\rho) = \inf \left\{ \int dj p(j) \xi(\psi_j) : \rho = \int dj p(j) |\psi_j\rangle\langle\psi_j| \right\}, \quad (86)$$

while we leave the analysis of the difference in these definitions for future work. It is also straightforward to define the approximate version of these measures in the same way as those in the resource theory of magic [23, 38].

These functions are valid measures of non-Gaussianity satisfying the monotonicity under Gaussian operations as follows. To see the monotonicity, it suffices to check that the composition with Gaussian states, applications of Gaussian unitaries, and measurement by heterodyne or homodyne detection cannot increase the Gaussian rank and the Gaussian extent. It is easy to see that the composition with Gaussian states as well as application of Gaussian unitaries do not increase the Gaussian rank or extent, because of the fact that the Gaussian states in the decomposition will be mapped to a different Gaussian state, while the number of Gaussian states in the superposition is left invariant. What is left to check is monotonicity under heterodyne measurements. Heterodyne measurements on subsystem B consisting of m modes are given by the following POVMs $\Pi_{|\boldsymbol{\alpha}\rangle\langle\boldsymbol{\alpha}|} = \frac{1}{(2\pi)^m} |\boldsymbol{\alpha}\rangle\langle\boldsymbol{\alpha}|$. In order to improve the readability, we will use the notation $|G_k\rangle = |\sigma_k, \boldsymbol{\mu}_k\rangle$ for the Gaussian state k . By

choosing one specific outcome, we get

$$|\psi_{\alpha}\rangle = \frac{1}{\sqrt{p(\alpha)}} \Pi_{|\alpha\rangle\langle\alpha|}^B |\psi\rangle \quad (87)$$

$$= \sum_{k=1}^{\chi} \frac{c_k}{\sqrt{p(\alpha)}} \Pi_{|\alpha\rangle\langle\alpha|} |G_k\rangle^{AB} \quad (88)$$

$$= \sum_{k=1}^{\chi} \frac{c_k^{\alpha}}{\sqrt{p(\alpha)}} |G_k^{\alpha}\rangle^{AB}, \quad (89)$$

where the covariance matrix gets changed according to Sec. 2 and $p(\alpha) = \text{Tr}[\Pi_{|\alpha\rangle\langle\alpha|}^B |\psi\rangle\langle\psi|]$, $c_k^{\alpha} = c_k \sqrt{\tilde{p}_k(\alpha)}$, $|G_k^{\alpha}\rangle^{AB} = \Pi_{|\alpha\rangle\langle\alpha|} |G_k\rangle^{AB} / \sqrt{\tilde{p}_k(\alpha)}$

with $\tilde{p}_k(\alpha) = \text{Tr}[\Pi_{|\alpha\rangle\langle\alpha|}^B |G_k^{AB}\rangle\langle G_k^{AB}|]$. Therefore, we see that the Gaussian rank of the post-measurement state is upper-bounded by the Gaussian rank of the pre-measurement state. A similar argument can be given homodyne and other Gaussian measurements.

Furthermore, we have for the Gaussian extent that

$$\xi(|\psi\rangle) \geq \int_{\mathcal{C}^m} d\alpha p(\alpha) \xi(|\psi_{\alpha}\rangle). \quad (90)$$

We can see this as follows.

The Gaussian extent ξ in (81) can be written by

$$\begin{aligned} \xi(|\psi\rangle) &= \inf \left\{ \left(\sum_i |c_i| \right)^2 \mid |\psi\rangle = \sum_i c_i |\phi_i\rangle, c_i \in \mathbb{C}, |\phi_i\rangle \in f \right\} \\ &= \inf \left\{ \left(\sum_i c_i \right)^2 \mid |\psi\rangle = \sum_i c_i |\phi_i\rangle, c_i \geq 0, |\phi_i\rangle \in f \right\} \\ &= \inf \left\{ \mu^2 \mid |\psi\rangle = \mu \sum_i p_i |\phi_i\rangle, |\phi_i\rangle \in f \right\} \\ &= \inf \left\{ \mu^2 \mid |\psi\rangle \in \mu \text{conv} f \right\} \end{aligned} \quad (91)$$

where in the second line we used the fact that $e^{i\theta} |\phi\rangle \in \mathcal{G}$ if $|\phi\rangle \in \mathcal{G}$, and in the third line $\{p_i\}_i$ denotes a probability distribution.

This means that the optimal decomposition of a state $|\psi\rangle$ is given as

$$|\psi\rangle = \mu \sum_i p_i |G_i\rangle. \quad (92)$$

We will write $|\sigma\rangle = \sum_i p_i |G_i\rangle \in \text{conv} \mathcal{G}$, meaning that $|\sigma\rangle$ is in the convex superposition of Gaussian states with $\sum_i p_i = 1$. By applying a heterodyne measurement $\Pi_{|\alpha\rangle\langle\alpha|}$ or, more generally, a free completely positive map, we know that

$$\Pi_{|\alpha\rangle\langle\alpha|} |\sigma\rangle = \sum_i p_i \Pi_{|\alpha\rangle\langle\alpha|} |G_i\rangle \quad (93)$$

$$= \sum_i p_i |G_i^{\alpha}\rangle =: \eta_{\alpha}, \quad (94)$$

where $|G_i^{\alpha}\rangle = \Pi_{|\alpha\rangle\langle\alpha|} |G_i\rangle$. Then, η_{α} is in $\|\eta_{\alpha}\| \text{conv} \mathcal{G} = \|\Pi_{|\alpha\rangle\langle\alpha|} |\sigma\rangle\| \text{conv} \mathcal{G}$. So we can write that

$$\Pi_{|\alpha\rangle\langle\alpha|} |\psi\rangle = \mu \eta_{\alpha} = \mu \|\eta_{\alpha}\| |\bar{G}\rangle \quad (95)$$

where $|\bar{G}\rangle \in \text{conv} \mathcal{G}$. This is a feasible solution to $\xi(\frac{\Pi_{|\alpha\rangle\langle\alpha|} |\psi\rangle}{\sqrt{p(\alpha)}})$, however not guaranteed to be optimal and thus

$$\xi\left(\frac{\Pi_{|\alpha\rangle\langle\alpha|} |\psi\rangle}{\sqrt{p(\alpha)}}\right) \leq \frac{\mu^2 \|\eta_{\alpha}\|^2}{p(\alpha)}. \quad (96)$$

Furthermore we have that, since $\|\psi\|^2 = 1$ and $\mu^2 \geq 1$ that $\|\sigma\| \leq 1$. This means however that $\int_{\mathcal{C}^m} d\alpha \|\eta_{\alpha}\| \leq 1$, since it holds that $\int_{\mathcal{C}^m} d\alpha \Pi_{|\alpha\rangle\langle\alpha|} = 1$. To conclude, we get that

$$\int_{\mathcal{C}^m} d\alpha p(\alpha) \xi(|\psi_{\alpha}\rangle) \leq \mu^2 \|\eta_{\alpha}\|^2 \leq \mu^2 = \xi(|\psi\rangle). \quad (97)$$

This monotonicity also holds for their extensions to mixed states due to a conventional argument for the monotonicity of convex-roof resource measures [42]. It is easy to see that both measures are faithful, i.e., have the minimal value 1 only for probabilistic mixtures of Gaussian states.

5.1 Characterization of Gaussian extent by robustness measure

In this section, we investigate the properties of the Gaussian extent ξ in Eq. (81). We show a characterization of the Gaussian extent using

$$\underline{R}_{\mathcal{F}}(\rho) = \sup \left\{ \text{Tr}(W\rho) \mid W \geq 0, \text{Tr}(W\sigma) \leq 1, \forall \sigma \in \mathcal{F} \right\}. \quad (98)$$

By choosing \mathcal{F} as the closure of the convex hull of the set of pure Gaussian states [13, 27], i.e., $\mathcal{F} = \text{cl conv}\{|\phi\rangle\langle\phi| \mid |\phi\rangle \in \mathcal{G}\}$, (98) represents the lower semicontinuous robustness of non-Gaussianity.

To see the relation to the Gaussian extent, let $\bar{\xi}$ be the quantity defined by

$$\bar{\xi}(|\psi\rangle) := \inf \left\{ \mu^2 \mid |\psi\rangle \in \mu \text{ cl conv } \mathcal{G} \right\} \quad (99)$$

where $\mu \text{ cl conv } \mathcal{G} = \text{cl conv}\{\mu|\phi\rangle \mid |\phi\rangle \in \mathcal{G}\}$ is the closure of the convex hull of vectors in \mathcal{G} scaled by μ .

Comparing (99) and (91) implies $\bar{\xi}(|\psi\rangle) \leq \xi(|\psi\rangle)$. Together with the identification between $\bar{\xi}(|\psi\rangle)$ and $\underline{R}_{\mathcal{F}}(|\psi\rangle\langle\psi|)$ [27, Proposition 17], we get

$$\xi(|\psi\rangle) \geq \bar{\xi}(|\psi\rangle) = \underline{R}_{\mathcal{F}}(|\psi\rangle\langle\psi|) \quad (100)$$

for an arbitrary pure state ψ .

This means that if we find a decomposition $|\psi\rangle = \sum_i c_i |\phi_i\rangle$ with $|\phi_i\rangle \in \mathcal{G}$ such that $(\sum_i |c_i|)^2$ matches $\underline{R}_{\mathcal{F}}(|\psi\rangle\langle\psi|)$, we can conclude that $\xi(|\psi\rangle) = \bar{\xi}(|\psi\rangle) = \underline{R}_{\mathcal{F}}(|\psi\rangle\langle\psi|)$ for such a state.

On the other hand, we show in Appendix G that whenever the equality $\xi(|\psi\rangle) = \underline{R}_{\mathcal{F}}(|\psi\rangle\langle\psi|)$ holds and the optimal values for both expressions are achieved with a certain decomposition and a witness operator, the optimal decomposition $|\psi\rangle = \sum_i c_i |\phi_i\rangle$, $|\phi_i\rangle \in \mathcal{G}$ such that $(\sum_i |c_i|)^2 = \xi(|\psi\rangle)$ is related to the optimal witness operator W for $\underline{R}(|\psi\rangle\langle\psi|)$ satisfying $\underline{R}_{\mathcal{F}} = \text{Tr}(W|\psi\rangle\langle\psi|)$ as

$$|\text{Tr}(W|\phi_i\rangle\langle\phi_i|)| = 1, \forall i. \quad (101)$$

We prove this for general resource theories defined in infinite-dimensional Hilbert spaces equipped with an arbitrary subset of pure states, extending the corresponding relation established for finite-dimensional theories in the context of the resource theory of magic [46, Lemma 3].

the robustness measure. This characterization provides a powerful condition for finding optimal decomposition for the Gaussian extent. Lower semicontinuous robustness for a set \mathcal{F} of states is given by [27, Corollary 6]

This characterization is a useful guide to finding the optimal decomposition of a state for which we know the optimal witness for robustness. For instance, we know for Fock state $|n\rangle$ that [27, Proposition 28]

$$\underline{R}_{\mathcal{F}}(|n\rangle\langle n|) = \frac{1}{\sup_{\alpha, \xi} |\langle n|\alpha, \xi\rangle|^2}, \quad (102)$$

where ξ is a squeezing parameter. Letting $|\alpha^*, \xi^*\rangle$ be a state that optimizes the right-hand side, we have

$$W = \frac{|n\rangle\langle n|}{|\langle n|\alpha^*, \xi^*\rangle|^2}. \quad (103)$$

In turn, the free states in the optimal decomposition would need to satisfy

$$|\langle n|\phi_i\rangle|^2 = |\langle n|\alpha^*, \xi^*\rangle|^2, \forall i. \quad (104)$$

These states include $|e^{i\theta}\alpha^*, \xi^*\rangle$, i.e., the states obtained by a phase shift from the seed state $|\alpha^*, \xi^*\rangle$. As we will see in the next section (Sec. 6), the optimal decomposition fulfills these requirements.

Finding optimal α^* and ξ^* involves optimization in general. Nevertheless, for the Fock state of $n = 1$, we have an analytical form [27, Proposition 29]

$$\alpha^* = \frac{2}{3}, \quad \xi^* = \ln \sqrt{3}. \quad (105)$$

6 Decompositions and Applications

6.1 Optimal Decompositions for the Gaussian extent

In this section, we give Gaussian decompositions for states that are of interest to the CV quantum computing community and show optimality for some of those.

The decomposition for the Fock state $|n = 1\rangle$ is obtained as follows. In the previous section, we

saw that the optimal value for the Gaussian extent of the Fock state $|1\rangle$ is given with the parameters in Eq. (105). We can construct the corresponding Gaussian decomposition in the following way. For this purpose, we use the projectors defined in Ref. [47], i.e.,

$$\Pi_{2N}^l = \sum_{k=0}^{\infty} |2kN + l\rangle\langle 2kN + l| \quad (106)$$

$$= \frac{1}{2N} \sum_{m=0}^{2N-1} \left(e^{i\pi l/N} e^{i\pi n/N} \right)^m, \quad (107)$$

where n is the photon number operator. In particular, we define a projector into the Fock state $|1\rangle$ as

$$|1\rangle\langle 1| = \lim_{N \rightarrow \infty} \Pi_{2N}^1. \quad (108)$$

Then, we apply this projector to a Gaussian state

$$\lim_{N \rightarrow \infty} \Pi_{2N}^1 |G\rangle = \langle 1|G\rangle |1\rangle. \quad (109)$$

We need a normalization to get $|1\rangle$. In order to minimize $\|c\|_1$, we choose the Gaussian state that has the maximum overlap with the Fock state $|1\rangle$, since it in turn will minimize $\|c\|_1$. Among coherent states, the maximum overlap is achieved by $|\alpha = 1\rangle$ with $\langle 1|\alpha = 1\rangle = \frac{1}{\sqrt{e}}$. Intuitively, the operator $\left(e^{i\pi l/N} e^{i\pi n/N} \right)^m$ is a phase-shifting operator. In the limit $\lim_{N \rightarrow \infty} \Pi_{2N}^1$, it superposes phase-shifted versions of the same state $|G\rangle$ on a circle with the same weight. We can immediately read the coefficients

$$|1\rangle = \frac{1}{\langle 1|G\rangle} \lim_{N \rightarrow \infty} \Pi_{2N}^1 |G\rangle \quad (110)$$

$$= \frac{1}{\langle 1|G\rangle} \lim_{N \rightarrow \infty} \frac{1}{2N} \sum_{m=0}^{2N-1} \left(e^{i\pi l/N} e^{i\pi n/N} \right)^m |G\rangle, \quad (111)$$

then we have $\|c\|_1 = \sqrt{e}$. We see that all coherent states in the decomposition have the same weight. For the more general Gaussian states, the state that maximizes the overlap is $|\alpha = \frac{2}{3}, \xi = \ln \sqrt{3}\rangle$. The overlap between the state $|\alpha = \frac{2}{3}, \xi = \ln \sqrt{3}\rangle$ and the Fock state $|1\rangle$ is $\left| \langle 1|\alpha = \frac{2}{3}, \xi = \ln \sqrt{3}\rangle \right|^2 = \frac{3\sqrt{3}}{4e}$. Using the state $|\alpha = \frac{2}{3}, \xi = \ln \sqrt{3}\rangle$ as the seed state $|G\rangle$, we obtain the decomposition of Fock state $|1\rangle$ with $\|c\|_1^2 = \frac{4e}{3\sqrt{3}}$. By comparing with the reported

optimal result in [27], we see that this decomposition is optimal. To summarize, the optimal decomposition of $|1\rangle$ is then a superposition of the seed state $|\alpha = \frac{2}{3}, \xi = \ln \sqrt{3}\rangle$ on which phase-shift operators have been applied to. This decomposition fulfills Eq. (104), e.g. it holds that for all m that

$$\left| \langle 1| \left(e^{i\pi l/N} e^{i\pi n/N} \right)^m \left| \alpha = \frac{2}{3}, \xi = \ln \sqrt{3}\rangle \right|^2 \quad (112)$$

$$= \left| \langle 1|\alpha = \frac{2}{3}, \xi = \ln \sqrt{3}\rangle \right|^2. \quad (113)$$

A family of states that have many applications for CV quantum computing [18, 48, 49], especially for error correction, is GKP states [34]. A finite-energy GKP state that encodes a computational basis state μ of a d -dimensional qudit can be decomposed into a sum of squeezed states by definition, i.e., [34, 50]

$$|\mu_{\kappa, \Delta}\rangle = \frac{1}{\sqrt{\mathcal{N}}} \sum_{s=-\infty}^{\infty} e^{-\frac{1}{2}\kappa^2 \alpha_d^2 (ds + \mu)^2} \quad (114)$$

$$\times D(\alpha_d(ds + \mu)) S(-\log \Delta) |0\rangle \quad (115)$$

$$= \frac{1}{\sqrt{\mathcal{N}}} \sum_{s=-\infty}^{\infty} e^{-\frac{1}{2}\kappa^2 \alpha_d^2 (ds + \mu)^2} |G_s\rangle, \quad (116)$$

where \mathcal{N} is a normalization constant. The constant κ^{-1} is the width of the Gaussian envelope, Δ describes the individual squeezing parameter with $\log(\Delta^{-1})$ and $\alpha_d = \sqrt{\frac{2\pi}{d}}$ a constant. One recovers the ideal GKP state for $\Delta, \kappa \rightarrow 0$. We use the notation $G_s = D(\alpha_d(ds + \mu)) S(-\log \Delta) |0\rangle$ to show that the GKP state can be written as a superposition of Gaussian states.

We see that states labeled by high s are exponentially suppressed in the superposition, making the simulator based on the Gaussian extent significantly more viable as compared to the one based on the Gaussian rank.

Another family of bosonic codes is the cat code. The cat code is an example of a rotational symmetric code [51]. The codewords of the cat code are written as

$$|\mu = 0/1^M\rangle = \frac{1}{\sqrt{\mathcal{N}}} \sum_{m=0}^{2M-1} (-1)^{\mu m} e^{i\frac{m\pi}{M} n} |\alpha\rangle, \quad (117)$$

where $|\alpha\rangle$ is a coherent state, and \mathcal{N} is a normalization constant. For squeezed cat codes, the

superposition goes over squeezed states instead of coherent states. The most commonly used states of this family are the even/odd cat states

$$|\alpha_{\pm}\rangle = \frac{1}{\sqrt{\mathcal{N}}}(|\alpha\rangle \pm |-\alpha\rangle) \quad (118)$$

with $\mathcal{N} = 2(1 \pm e^{-2|\alpha|^2})$. For this decomposition, we obtain that $\|c\|_1 = \frac{2}{\sqrt{\mathcal{N}}} = \sqrt{\frac{2}{1 \pm e^{-2|\alpha|^2}}}$. It trivially holds that these decompositions are optimal in minimizing the Gaussian rank in Eq. (83) for non-vanishing α because 2 is the smallest rank for a non-Gaussian state.

The optimal decomposition that minimizes the Gaussian extent in Eq. (81) requires more careful analysis. For the resource theory of non-classicality, the decomposition Eq. (118) is optimal for the case of $|\alpha_+\rangle$; for $|\alpha_-\rangle$, the same is the case for approximately $\alpha > 1$. Both cases yield $\|c\|_1 = \sqrt{2}$ for $\alpha \rightarrow \infty$. For the resource theory of non-Gaussianity, the previous decomposition is an upper bound, and the optimality is unknown in general. To gain further insight, consider the extreme cases $\lim_{\alpha \rightarrow 0} |\alpha_+\rangle = |0\rangle$ and $\lim_{\alpha \rightarrow 0} |\alpha_-\rangle = |1\rangle$, which yield $\|c\|_1^2 = 1$ and $\|c\|_1^2 = \frac{4e}{3\sqrt{3}}$, respectively. In the opposite case of $\alpha \rightarrow \infty$, it should converge to $\|c\|_1 = \sqrt{2}$ even if one exchanges the coherent states with squeezed states (squeezed along the correct axis) as for the non-classicality case; after all, they asymptotically become orthogonal. Therefore, the decomposition above is optimal for large enough cat states (and/or squeezing). It could even be that for $|\alpha_+\rangle$, the optimal decomposition is the same for non-Gaussianity and non-classicality. Note here that either way, the scaling is better than the one proposed in [22], reducing the simulation overhead.

In general, for non-classicality, the norm $\|c\|_1$ and thus the optimal decomposition is multiplicative for single-mode states [27, Proposition 24]. However, this is not known in the case of non-Gaussianity. In Ref. [24, Theorem 6.6], it is proved that the multiplicativity of the \mathcal{D} -fidelity implies multiplicativity of the \mathcal{D} -extent for a possibly infinite dictionary $\mathcal{D} \subset \mathcal{H}$ if there exists a finite ϵ -net in \mathcal{D} that contains an orthonormal basis of \mathcal{H} . A sufficient condition for this is that the subset $\mathcal{D} \subset \mathcal{H}$ is compact. However, in our case, the set of Gaussian states is *not* compact, and thus their proof is not applicable to this case.

We will nonetheless report here a counterexample for the multiplicativity of the Gaussian extent. As reported before in Eq. (102), the lower semicontinuous robustness of non-Gaussianity is the inverse maximal fidelity of a Fock state $|n\rangle$ and pure Gaussian state $|\alpha, \xi\rangle$. Since the optimal value for $\underline{R}_{\mathcal{F}}(|1\rangle\langle 1|) = \xi(|1\rangle) = \frac{4e}{3\sqrt{3}}$, this, together with Eq. (102), implies that the maximal overlap between a Fock state and a Gaussian state is $|\langle 1|G\rangle|^2 = \frac{3\sqrt{3}}{4e} \approx 0.47789$, where $|G\rangle$ is the Gaussian state closest to the state $|1\rangle$. Since we do not have provable optimal results for the multimode case, we use a numerical approach. Details as well as the optimization parameters can be found in Appendix H. The highest fidelity we numerically found between the tensor product of two Fock states $|1 \otimes 1\rangle$ and a two-mode Gaussian state $|G'\rangle$ is $|\langle 1 \otimes 1|G'\rangle|^2 = \frac{1}{4}$. Note here that this fidelity is not provably maximum, but is nonetheless a lower bound on the maximal fidelity between two Fock states $|1 \otimes 1\rangle$ and the set of Gaussian states. It thus holds that

$$|\langle 1 \otimes 1|G'\rangle|^2 > |\langle 1|G\rangle|^2 |\langle 1|G\rangle|^2, \quad (119)$$

which shows that the Gaussian fidelity is not multiplicative for the state $|1\rangle$. Since the Gaussian extent coincides with the Gaussian fidelity for Fock states as in Eq. (102), we conclude that the Gaussian extent is not multiplicative, even for single-mode states. This means, in terms of classical simulation, that finding multi-mode expansions can improve the performance of the simulation algorithm as the simulation cost is sub-multiplicative.

6.2 Applications

In this section, we discuss the applications of our non-Gaussian simulator and the bounds that we derive using the Gaussian extent.

Gaussian boson sampling: We start by investigating sampling models. There are several versions of boson sampling. All can be implemented using linear optics transformations. Standard boson sampling [52] has several Fock states as input and measures photon number

$$|1\rangle^N \otimes |0\rangle^{M-N} \rightarrow \text{Measure } n. \quad (120)$$

Gaussian boson sampling [53] changes the input to Gaussian states and measures photon number

$$|G\rangle^N \rightarrow \text{Measure } n. \quad (121)$$

Equivalently, this can be reversed, and instead, we can have Fock states in the input and measure heterodyne or, more generally, a Gaussian measurement [19, 54]

$$|1\rangle^N \rightarrow \text{Measure } G. \quad (122)$$

All variants of boson sampling are based on the fact that it is hard to compute the output probabilities of measuring a certain photon pattern in the high mode regime, where the number of modes scales faster than the number of photons.

Among these variants, Gaussian boson sampling and the reversed variant are especially suited for our approach. The probability of measuring a certain output pattern in a Gaussian boson sample experiment is

$$P(\vec{n}) = \text{Tr} \left[\bigotimes_{j=1}^M |n_j\rangle\langle n_j| U |G\rangle\langle G| U^\dagger \right], \quad (123)$$

where we start with a Gaussian pure state $|G\rangle\langle G|$ and then evolve it under a Gaussian (passive) unitary U with $n_j = 0/1$. This is equivalent to the time-reversed version

$$P(\vec{n}) = \text{Tr} \left[\bigotimes_{j=1}^M |n_j\rangle\langle n_j| U |G\rangle\langle G| U^\dagger \right] \quad (124)$$

$$= \text{Tr} \left[U^\dagger \bigotimes_{j=1}^M |n_j\rangle\langle n_j| U |G\rangle\langle G| \right], \quad (125)$$

where we start with some Fock states $|1\rangle$ and $|0\rangle$, evolve them with the inverse unitary, which is again a Gaussian unitary, and then finally measure by a certain Gaussian measurement. Interestingly, the hardness proofs require that there are more modes than photons in the system. Our simulator has, in some sense, the converse behavior. Adding Gaussian auxiliary modes in the vacuum states does not really increase the simulation cost significantly, but adding single-photon states does.

As we have shown before, the Gaussian extent is not multiplicative; therefore, we can give only an upper bound on the simulation cost and not the lowest one possible. Note that the simulation algorithm still works—it just does not have

optimal performance. The upper bound of the simulation cost is then just $(\|c\|_1^2)^{\bar{M}} = \left(\frac{4e}{3\sqrt{3}}\right)^{\bar{M}}$, where \bar{M} is the number of Fock state $|1\rangle$ in the input. Therefore, our results provide an algorithm whose simulator runtime scales exponentially in the number of Fock states, allowing one to classically simulate logarithmically many Fock states efficiently. This is consistent with [19].

It is insightful to compare the case of the resource theory of non-classicality since all the boson sampling models only involve passive symplectic unitary operations. In this case, the corresponding extent is multiplicative, and the optimal decomposition is known with

$$\underline{R}_{\mathcal{F}}(|n\rangle\langle n|) = e^n \frac{n!}{n^n}. \quad (126)$$

Therefore, to compute the probability of obtaining a specific sampling pattern that involves \bar{M} single photons, the simulation cost scales with $e^{\bar{M}} > \left(\frac{4e}{3\sqrt{3}}\right)^{\bar{M}}$. We see that even though the dynamics and measurements are fully included in the set of free operations, using the decomposition obtained for the resource theory of non-Gaussianity improves the simulation capacity, even if we only compare it with the upper bound.

Cat state breeding: The generation of grid states is of wide interest in the continuous-variable quantum information community due to their usefulness for fault-tolerant quantum computing [34]. The generation of such states remains challenging. One of the most promising ways is to use cat states to breed grid states [55]. The breeding protocol takes in squeezed cat states and uses Gaussian operations to generate a grid state of higher quality. The targeted grid states are the sensor states, i.e.,

$$|\Psi_\Delta\rangle \propto \sum_{t=-\infty}^{\infty} e^{-\pi\Delta^2 t^2} D\left(t\sqrt{\frac{\pi}{2}}\right) S(\Delta) |0\rangle \quad (127)$$

$$\propto \sum_{t=-\infty}^{\infty} e^{-\pi\Delta^2 t^2} |G_t\rangle. \quad (128)$$

We can use the Gaussian extent to derive a lower bound on how many cat states are needed to obtain one sensor state with a certain squeezing Δ . Cat state breeding protocols require large enough displacement in the input [55], so we assume that the cat states are in the limit of

Δ	$\xi(\Psi_\Delta\rangle)$	n
0.3	2.797	2
0.2	3.969	2
0.1	7.496	3
0.05	14.562	4
0.025	28.701	5
0.01	71.126	7

Table 1: Numerical values of the Gaussian extent $\xi(|\Psi_\Delta\rangle)$ of the grid state $|\Psi_\Delta\rangle$ with squeezing Δ in Eq.(128) and the lower bound on the required number n of cat states for obtaining $|\Psi_\Delta\rangle$.

near-orthogonality of $|\alpha\rangle$ and $|\alpha_+\rangle$ leading to $\xi(|\alpha_+\rangle) \approx 2$. We do not require this fact in order to derive a lower bound for the required copies of cat states. Since we have optimal values of the robustness for non-classicality, we have upper bounds on the Gaussian extent, which allows us to compute a lower bound on the number of copies required. The minimal required number n of cat states $|\alpha_+\rangle$ for generating a grid states $|\Psi_\Delta\rangle$ is bounded by using the monotonicity of the Gaussian extent ξ as

$$\xi(|\alpha_+^{\otimes n}\rangle) \geq \xi(|\Psi_\Delta\rangle). \quad (129)$$

Since it holds that

$$\xi(|\alpha_+^{\otimes n}\rangle) \leq \xi(|\alpha_+\rangle)^n, \quad (130)$$

we can find the lower bound by

$$n \geq \left\lceil \frac{\log \xi(|\Psi_\Delta\rangle)}{\log \xi(|\alpha_+\rangle)} \right\rceil = \left\lceil \frac{\log \xi(|\Psi_\Delta\rangle)}{\log 2} \right\rceil. \quad (131)$$

For high enough squeezing, i.e., small enough Δ , the Gaussian states $|G_t\rangle$ of the grid states are nearly orthogonal. So we compute $\xi(|\Psi_\Delta\rangle)$ using the coefficients of Eq. (128). The results can be found in Table 1. These bounds are far away from the results reported in Ref. [55]. To obtain a squeezing level of $\Delta = 0.1$, the authors of [55] report $M = 6$ rounds of breeding using $2^M = 64$ cat states. However, they fix the input squeezing and displacement of their cat states, while we essentially leave it indeterminate. So our bounds are valid for all input cat states, independent of their magnitude.

We here compare these bounds with other state-of-the-art techniques. The most commonly studied measures in the resource theory of non-Gaussianity are the stellar rank [21] and the

Wigner negativity [13, 14]. In the following, we will compare state transformation bounds that one can obtain using the stellar rank and Wigner negativity with the bounds we obtain using the Gaussian extent. If we consider the state conversion from many cat states to one GKP state, the stellar rank cannot be used as cat and GKP states have infinite stellar rank. The Wigner negativity of cat states with large spacing is upper-bounded [12]. The number of such cat states with large magnitude to GKP states with $\Delta = 0.3, 0.2, 0.1$ is equivalent to the one obtained with the Gaussian extent presented in Tab. 1. However, if we consider the state transformation from cat states with large magnitude to a single Fock state, Wigner negativity allows this conversion. Even more, as long as the magnitude α is approximately larger than $\sqrt{2}$, the transformation is possible. The stellar rank allows the state conversion even for $\alpha = \epsilon$. The bounds using the Gaussian extent, however, tell us that we need at least two cat states to obtain the Fock state one. Generally, one can say that since the Gaussian extent is sub-multiplicative and the Wigner negativity is multiplicative, the Gaussian extent will have tighter bounds in most multi-copy scenarios.

7 Conclusion

In this work, we introduced two efficient algorithms for the classical simulation of non-Gaussian optics. Inherently connected to these algorithms are two measures of non-Gaussianity, i.e., the Gaussian rank and the Gaussian extent, which quantify the computational cost of the simulation. The simulator, whose cost scales quadratically with the Gaussian rank, is exact, while the other one uses a sparsification of the non-Gaussian states and is thus approximate. This approximation allows for reducing the scaling of the simulation cost to linear in the Gaussian extent. The algorithms use extensions of the standard covariance matrix formalism that include phase-sensitive overlaps and a routine for fast norm estimation that allows for linear scaling. Furthermore, we investigated the properties of the Gaussian extent. Employing the fact that the Gaussian extent is connected to the lower semicontinuous robustness of non-Gaussianity, we were able to find optimal decompositions for

states that are of interest to the continuous variable quantum computing community. Using the direct connection between the Gaussian extent and the Gaussian fidelity that exists for Fock states, we give a counterexample showing that the Gaussian extent is, in general, not multiplicative. We applied the simulator and the tools we developed in this work to boson sampling. We gave an upper bound to the simulation cost and showed that even when the dynamics are fully included within the free set of operations of the resource theory of non-classicality, it is beneficial to use the more general decompositions using Gaussian states. At last, we made a resource-theoretic analysis of cat state breeding and derived bounds using the Gaussian extent. The bounds are fundamental lower bounds on how many cat states are required, independent of their magnitude.

Although we have focused the presentation on Gaussian circuits with input non-Gaussian states, our method might also be useful to simulate circuits with non-Gaussian operations by recasting the latter into the former with techniques similar to those used for Clifford + T circuits [20, 56, 57]. Furthermore, in order to compute bounds on state conversions within the resource theory of non-Gaussianity, one requires provably optimal decompositions. These decompositions, in turn, improve the scaling of the simulation algorithm. Therefore, it is an interesting path to find a provably optimal decomposition for more states that are relevant for CV quantum computing and to compute bounds. Our work connects the practical usefulness of an efficient classical simulator with the fundamental investigation of resources required for CV quantum computing.

Note: During the preparation of this manuscript, we became aware of related work by B. Dias and R. Koenig [58]. We independently arrived at similar results on defining a simulator for non-Gaussian states using similar ideas to compute the overlap and estimate the norm.

Acknowledgments

G.F. acknowledges support from the Swedish Research Council (Vetenskapsrådet) through the project grant DAIQUIRI, as well as from the HORIZON-EIC-2022-PATHFINDERCHALLENGES-01 programme

under Grant Agreement Number 101114899 (Veriqub). G.F. and O.H. acknowledge support from the Knut and Alice Wallenberg Foundation through the Wallenberg Center for Quantum Technology (WACQT). R.T. acknowledges the support of JSPS KAKENHI Grant Number JP23K19028, JP24K16975, JST, CREST Grant Number JPMJCR23I3, Japan, and MEXT KAKENHI Grant-in-Aid for Transformative Research Areas A “Extreme Universe” Grant Number JP24H00943. H.Y. was supported by JST PRESTO Grant Number JPMJPR201A, JPMJPR23FC, JSPS KAKENHI Grant Number JP23K19970, and MEXT Quantum Leap Flagship Program (MEXT QLEAP) JPMXS0118069605, JPMXS0120351339.

A Inner product formula

We present the details of the way of computing overlaps of two pure Gaussian states $|G_i\rangle, |G_j\rangle \in \mathcal{G}$, i.e.,

$$\langle G_i | G_j \rangle. \quad (132)$$

For this purpose, we introduce a Gaussian reference state $G_0 = |G_0\rangle\langle G_0|$, which can be a random Gaussian state or the vacuum state. As presented in the main text, we want to calculate overlaps between Gaussian states $\langle G_1 | G_2 \rangle$ using

$$\text{Tr}(G_0 G_1 G_2) = \text{Tr}(|G_0\rangle\langle G_0| |G_1\rangle\langle G_1| |G_2\rangle\langle G_2|) \quad (133)$$

$$= \langle G_2 | G_0 \rangle \langle G_1 | G_2 \rangle \langle G_0 | G_1 \rangle. \quad (134)$$

To obtain the desired inner product $\langle G_1 | G_2 \rangle$ from (133), we need to specify Gaussian states not only using the covariance matrix and the mean but also by specifying the inner product with the reference state.

To compute (133), we represent the Gaussian states $|G\rangle$ in terms of the characteristic function as

$$G = \frac{1}{(2\pi)^n} \int_{\mathbb{R}^{2n}} d\mathbf{r} e^{-\frac{1}{4}\mathbf{r}^T \sigma \mathbf{r} + i\boldsymbol{\mu}^T \mathbf{r}} D(\Omega^T \mathbf{r}). \quad (135)$$

Using this representation, we have that

$$\text{Tr}(G_0 G_1 G_2) = \frac{1}{(2\pi)^{3n}} \int_{\mathbb{R}^{2n}} d\mathbf{r}_0 d\mathbf{r}_1 d\mathbf{r}_2 e^{-\frac{1}{4}\mathbf{r}_0^T \sigma_0 \mathbf{r}_0 + i\boldsymbol{\mu}_0^T \mathbf{r}_0} e^{-\frac{1}{4}\mathbf{r}_1^T \sigma_1 \mathbf{r}_1 + i\boldsymbol{\mu}_1^T \mathbf{r}_1} e^{-\frac{1}{4}\mathbf{r}_2^T \sigma_2 \mathbf{r}_2 + i\boldsymbol{\mu}_2^T \mathbf{r}_2} \quad (136)$$

$$\times \text{Tr} \left[D(\Omega^T \mathbf{r}_0) D(\Omega^T \mathbf{r}_1) D(\Omega^T \mathbf{r}_2) \right]. \quad (137)$$

Then, using

$$D(\Omega^T \mathbf{r}_1) D(\Omega^T \mathbf{r}_2) = D(\Omega^T (\mathbf{r}_1 + \mathbf{r}_2)) e^{-\frac{i}{2}(\Omega \mathbf{r}_1)^T \Omega (\Omega \mathbf{r}_2)} \quad (138)$$

$$= D(\Omega^T (\mathbf{r}_1 + \mathbf{r}_2)) e^{\frac{i}{2}\mathbf{r}_1^T \Omega \mathbf{r}_2} \quad (139)$$

and

$$\text{Tr} \left[D(\Omega^T r) D(\Omega^T s) \right] = (2\pi)^n \delta^{2n}(\Omega^T (r + s)), \quad (140)$$

we obtain

$$\frac{1}{(2\pi)^{2n}} \int_{\mathbb{R}^{2n}} d\mathbf{r}_0 d\mathbf{r}_1 d\mathbf{r}_2 e^{-\frac{1}{4}\mathbf{r}_0^T \sigma_0 \mathbf{r}_0 + i\boldsymbol{\mu}_0^T \mathbf{r}_0} e^{-\frac{1}{4}\mathbf{r}_1^T \sigma_1 \mathbf{r}_1 + i\boldsymbol{\mu}_1^T \mathbf{r}_1} e^{-\frac{1}{4}\mathbf{r}_2^T \sigma_2 \mathbf{r}_2 + i\boldsymbol{\mu}_2^T \mathbf{r}_2} \delta^{2n}(\Omega^T (\mathbf{r}_0 + \mathbf{r}_1 + \mathbf{r}_2)) e^{\frac{i}{2}\mathbf{r}_1^T \Omega \mathbf{r}_2} \quad (141)$$

$$= \frac{1}{(2\pi)^{2n}} \int_{\mathbb{R}^{2n}} d\mathbf{r}_0 d\mathbf{r}_1 e^{-\frac{1}{4}\mathbf{r}_0^T \sigma_0 \mathbf{r}_0 + i\boldsymbol{\mu}_0^T \mathbf{r}_0} e^{-\frac{1}{4}\mathbf{r}_1^T \sigma_1 \mathbf{r}_1 + i\boldsymbol{\mu}_1^T \mathbf{r}_1} e^{-\frac{1}{4}(\mathbf{r}_0^T + \mathbf{r}_1^T) \sigma_2 (\mathbf{r}_0 + \mathbf{r}_1) - i\boldsymbol{\mu}_2^T (\mathbf{r}_0 + \mathbf{r}_1)} e^{-\frac{i}{2}\mathbf{r}_1^T \Omega (\mathbf{r}_0 + \mathbf{r}_1)} \quad (142)$$

$$= \frac{1}{(2\pi)^{2n}} \int_{\mathbb{R}^{2n}} d\mathbf{r}_0 d\mathbf{r}_1 e^{-\frac{1}{4}\mathbf{r}_0^T \sigma_0 \mathbf{r}_0 + i\boldsymbol{\mu}_0^T \mathbf{r}_0} e^{-\frac{1}{4}\mathbf{r}_1^T \sigma_1 \mathbf{r}_1 + i\boldsymbol{\mu}_1^T \mathbf{r}_1} \quad (143)$$

$$\times e^{-\frac{1}{4}(\mathbf{r}_0^T \sigma_2 \mathbf{r}_0 + \mathbf{r}_1^T \sigma_2 \mathbf{r}_0 + \mathbf{r}_0^T \sigma_2 \mathbf{r}_1 + \mathbf{r}_1^T \sigma_2 \mathbf{r}_1) - i\boldsymbol{\mu}_2^T \mathbf{r}_0 - i\boldsymbol{\mu}_2^T \mathbf{r}_1} e^{-\frac{i}{2}\mathbf{r}_1^T \Omega (\mathbf{r}_0 + \mathbf{r}_1)} \quad (144)$$

$$= \frac{1}{(2\pi)^{2n}} \int_{\mathbb{R}^{2n}} d\mathbf{r}_0 d\mathbf{r}_1 e^{-\frac{1}{4}\mathbf{r}_0^T \sigma_0 \mathbf{r}_0 + i\boldsymbol{\mu}_0^T \mathbf{r}_0} e^{-\frac{1}{4}\mathbf{r}_0^T \sigma_2 \mathbf{r}_0 - i\boldsymbol{\mu}_2^T \mathbf{r}_0} e^{-\frac{1}{4}\mathbf{r}_1^T \sigma_1 \mathbf{r}_1 + i\boldsymbol{\mu}_1^T \mathbf{r}_1} \quad (145)$$

$$\times e^{-\frac{1}{4}(\mathbf{r}_1^T \sigma_2 \mathbf{r}_0 + \mathbf{r}_0^T \sigma_2 \mathbf{r}_1 + \mathbf{r}_1^T \sigma_2 \mathbf{r}_1) - i\boldsymbol{\mu}_2^T \mathbf{r}_1} e^{-\frac{i}{2}\mathbf{r}_1^T \Omega \mathbf{r}_0} e^{-\frac{i}{2}\mathbf{r}_1^T \Omega \mathbf{r}_1} \quad (146)$$

$$= \frac{1}{(2\pi)^{2n}} \int_{\mathbb{R}^{2n}} d\mathbf{r}_0 e^{-\frac{1}{4}\mathbf{r}_0^T \sigma_0 \mathbf{r}_0 + i\boldsymbol{\mu}_0^T \mathbf{r}_0} e^{-\frac{1}{4}\mathbf{r}_0^T \sigma_2 \mathbf{r}_0 - i\boldsymbol{\mu}_2^T \mathbf{r}_0} \int_{\mathbb{R}^{2n}} d\mathbf{r}_1 e^{-\frac{1}{4}\mathbf{r}_1^T \sigma_1 \mathbf{r}_1 + i\boldsymbol{\mu}_1^T \mathbf{r}_1} \quad (147)$$

$$\times e^{\frac{i}{2}\mathbf{r}_0^T \Omega \mathbf{r}_1} e^{-\frac{1}{4}(\mathbf{r}_1^T \sigma_2 \mathbf{r}_1 + \mathbf{r}_1^T \sigma_2 \mathbf{r}_0 + \mathbf{r}_0^T \sigma_2 \mathbf{r}_1) - i\boldsymbol{\mu}_2^T \mathbf{r}_1}, \quad (148)$$

where (142) is obtained by using the Dirac distribution $\delta^{2n}(\Omega^T(\mathbf{r}_0 + \mathbf{r}_1 + \mathbf{r}_2))$ to replace $r_2 = -r_0 - r_1$, and the last line follows from the fact that $\mathbf{r}_1^T \Omega \mathbf{r}_1 = 0$ since Ω is anti-symmetric. Using the Gaussian integral formula

$$\int_{\mathbb{R}^{2n}} d\mathbf{r} e^{-\frac{1}{4}\mathbf{r}^T A \mathbf{r} + \mathbf{B}^T \mathbf{r}} = \sqrt{\frac{(2\pi)^{2n}}{\det A}} e^{\frac{1}{4}\mathbf{B}^T A^{-1} \mathbf{B}},$$

we rewrite the integral over \mathbf{r}_1 into

$$\frac{1}{(2\pi)^n} \int_{\mathbb{R}^{2n}} d\mathbf{r}_1 e^{-\frac{1}{4}\mathbf{r}_1^T (\sigma_1 + \sigma_2) \mathbf{r}_1 + (-\frac{1}{2}\mathbf{r}_0^T \sigma_2 + i[\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2 + \frac{1}{2}\mathbf{r}_0^T \Omega]) \mathbf{r}_1} = \frac{1}{\sqrt{\det A}} e^{\frac{1}{4}\mathbf{B}^T A^{-1} \mathbf{B}}, \quad (149)$$

where we use $\mathbf{r}_1^T \sigma_2 \mathbf{r}_0 = \mathbf{r}_0^T \sigma_2 \mathbf{r}_1$, and we take

$$A = \sigma_1 + \sigma_2, \quad (150)$$

$$B^T = -\frac{1}{2}\mathbf{r}_0^T \sigma_2 + i \left[\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2 + \frac{1}{2}\mathbf{r}_0^T \Omega \right]. \quad (151)$$

The right-hand side is further simplified by

$$\begin{aligned} e^{\frac{1}{4}\mathbf{B}^T A^{-1} \mathbf{B}} &= e^{\frac{1}{4} \left[(-\frac{1}{2}\mathbf{r}_0^T \sigma_2 + i(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2 + \frac{1}{2}\mathbf{r}_0^T \Omega)) (\sigma_1 + \sigma_2)^{-1} (-\frac{1}{2}\mathbf{r}_0^T \sigma_2 + i(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2 + \frac{1}{2}\mathbf{r}_0^T \Omega))^T \right]} \\ &= e^{\frac{1}{4} \left[\left(\frac{\mathbf{r}_0^T}{2} (-\sigma_2 + i\Omega) + i(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) \right) (\sigma_1 + \sigma_2)^{-1} \left((-\sigma_2 + i\Omega)^T \frac{\mathbf{r}_0}{2} + i(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^T \right) \right]} \\ &= e^{\frac{1}{4} \left[\mathbf{r}_0^T \frac{(\sigma_2 - i\Omega)^T}{2} (\sigma_1 + \sigma_2)^{-1} \frac{(\sigma_2 - i\Omega)}{2} \mathbf{r}_0 \right]} e^{\left(\frac{-i}{2} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^T (\sigma_1 + \sigma_2)^{-1} \frac{(\sigma_2 - i\Omega)}{2} \right) \mathbf{r}_0} e^{-\frac{1}{4} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^T (\sigma_1 + \sigma_2)^{-1} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)}. \end{aligned}$$

Thus, we have for all integrals

$$\begin{aligned} &\frac{1}{(2\pi)^n} \frac{1}{\sqrt{\det(\sigma_1 + \sigma_2)}} e^{-\frac{1}{4} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^T (\sigma_1 + \sigma_2)^{-1} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)} \\ &\quad \times \int_{\mathbb{R}^{2n}} d\mathbf{r}_0 e^{-\frac{1}{4}\mathbf{r}_0^T \left[\sigma_0 + \sigma_2 - \frac{(\sigma_2 - i\Omega)^T}{2} (\sigma_1 + \sigma_2)^{-1} \frac{(\sigma_2 - i\Omega)}{2} \right] \mathbf{r}_0} e^{-i \left[\boldsymbol{\mu}_0 - \boldsymbol{\mu}_2 + \frac{1}{2} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) (\sigma_1 + \sigma_2)^{-1} \frac{(\sigma_2 - i\Omega)}{2} \right] \mathbf{r}_0} \quad (152) \end{aligned}$$

$$= \frac{1}{\sqrt{\det(\sigma_1 + \sigma_2)}} \frac{1}{\sqrt{\det(\sigma_0 + \Delta)}} e^{-\frac{1}{4} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^T (\sigma_1 + \sigma_2)^{-1} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)} e^{-\frac{1}{4} (\boldsymbol{\mu}_0 - \boldsymbol{\mu}_\Delta)^T (\sigma_0 + \Delta)^{-1} (\boldsymbol{\mu}_0 - \boldsymbol{\mu}_\Delta)}, \quad (153)$$

where we take

$$\Delta = \sigma_2 - \frac{(\sigma_2 - i\Omega)^T}{2} (\sigma_1 + \sigma_2)^{-1} \frac{(\sigma_2 - i\Omega)}{2}, \quad (154)$$

$$\boldsymbol{\mu}_\Delta = \boldsymbol{\mu}_2 + \frac{1}{2} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2) (\sigma_1 + \sigma_2)^{-1} \frac{(\sigma_2 - i\Omega)}{2}. \quad (155)$$

Consequently, we obtain

$$\begin{aligned} \langle G_2 | G_0 \rangle \langle G_1 | G_2 \rangle \langle G_0 | G_1 \rangle &= \frac{1}{\sqrt{\det(\sigma_1 + \sigma_2)}} \frac{1}{\sqrt{\det(\sigma_0 + \Delta)}} e^{-\frac{1}{4} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)^T (\sigma_1 + \sigma_2)^{-1} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)} \\ &\quad \times e^{-\frac{1}{4} (\boldsymbol{\mu}_0 - \boldsymbol{\mu}_\Delta)^T (\sigma_0 + \Delta)^{-1} (\boldsymbol{\mu}_0 - \boldsymbol{\mu}_\Delta)}. \end{aligned}$$

B Phases and reference states

As we have shown in Appendix A, we can compute the inner product between two pure Gaussian states using the covariance matrix, the mean, and a reference state. Any n -mode Gaussian unitary can be decomposed as

$$G = US(\gamma)D(\beta)V, \quad (156)$$

where U, V are passive symplectic transformations, S the single-mode squeezing, and D the displacement [59]. Any pure Gaussian state can be consequently written as

$$|G\rangle = US(\gamma)D(\beta)V|0\rangle \quad (157)$$

$$= U\left(\bigotimes_{i=1}^n S(\gamma_i)D(\beta_i)\right)|0\rangle \quad (158)$$

$$= U\bigotimes_{i=1}^n |G_i\rangle, \quad (159)$$

where we use $V|0\rangle = |0\rangle$. We choose the vacuum $|0\rangle$ as our reference state. We thus want to compute the overlap

$$\langle 0|G\rangle = \langle 0|U\bigotimes_{i=1}^n |G_i\rangle = \bigotimes_{i=1}^n \langle 0|G_i\rangle. \quad (160)$$

We can compute this overlap by using the stellar representation [59]; in particular with [59, Lemma 7], we see that

$$\bigotimes_{i=1}^n \langle 0|G_i\rangle = \prod_{i=1}^n (1 - |t_i|^2)^{-\frac{1}{4}} e^{C_i} \quad (161)$$

with $t_i = e^{i\theta_i} \tanh(r_i)$, $\gamma_i = e^{i\theta_i} r_i$, and $C_i = \frac{1}{2}(t_i^* \beta_i^2 - |\beta_i|^2)$ for $\gamma_i = r_i e^{i\theta_i}$.

We can retrieve these coefficients from a given covariance matrix and mean. For this retrieval, our simulator keeps track of the mean and the covariance matrix during the applications of the Gaussian unitary. The mean and covariance matrix before the application of U are those of single-mode Gaussians

$$\mu = \bigoplus_{i=1}^n S_i \beta_i, \quad (162)$$

$$\sigma = \bigoplus_{i=1}^n S_i \mathbb{1}_i S_i^T = \bigoplus_{i=1}^n \sigma_i, \quad (163)$$

where S_i is the symplectic matrix associated to $S(\gamma_i)$. Then, the application of U yields

$$\mu = \bar{U} \bigoplus_{i=1}^n S_i \beta_i, \quad (164)$$

$$\sigma = \bar{U} \bigoplus_{i=1}^n \sigma_i \bar{U}^T, \quad (165)$$

where \bar{U} is the orthogonal symplectic matrix associated with the unitary U . We can thus use σ to find \bar{U} by block-diagonalizing σ and consequently S_i from σ_i . Having access to \bar{U} and S_i , we can also compute β from the mean μ . Numerical recipes can be found in [60].

C Alternative phase sensitive simulator

In this section, we give an alternative phase-sensitive simulator for Gaussian overlaps. We base it on [33]. We use the stellar formalism directly in the computation. The stellar function of a Gaussian pure state $|\psi\rangle\langle\psi|$ consisting of M modes is given by

$$\Gamma(\boldsymbol{\alpha}) = c_\psi \exp\left(\boldsymbol{\alpha}^T \mathbf{b}_\psi + \frac{1}{2} \boldsymbol{\alpha}^T A_\psi \boldsymbol{\alpha}\right), \quad (166)$$

where A_ψ is an $M \times M$ complex symmetric matrix, \mathbf{b}_ψ is a M -dimensional complex vector, c_ψ the vacuum amplitude. The same can be done for Gaussian mixed states ρ with A_ρ , \mathbf{b}_ρ , c_ρ where A_ρ , \mathbf{b}_ρ are $2M$ dimensional. In this case, we can compute these parameters directly from the covariance matrix and mean of the Gaussian state ρ . We will use the s -parameterized complex covariance matrix $\bar{\sigma}_S$

$$\bar{\sigma}_S = \bar{\sigma} + \frac{s}{2} \mathbb{1}_{2M} \quad (167)$$

with

$$\bar{\sigma} = \bar{U}^\dagger \sigma \bar{U} \quad (168)$$

and the complex mean

$$\bar{\boldsymbol{\mu}} = \bar{U}^\dagger \bar{\mathbf{r}}. \quad (169)$$

We use the matrix

$$\bar{U} = \bigoplus_{j=1}^M \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & i \\ 1 & -i \end{pmatrix} \quad (170)$$

to go between the quadrature operators \mathbf{r} and the annihilation and creation operators a, a^\dagger . We will use the matrix W to switch between the ordering of the quadratures operators $r_{q_1}, r_{p_1}, \dots, r_{q_M}, r_{p_M}$ to $r_{q_1}, \dots, r_{q_M}, r_{p_1}, \dots$

Then we can compute the matrix A_ρ , the vector \mathbf{b}_ρ and c_ρ using the covariance matrix and the mean as follows:

$$A_\rho = P_M W \bar{\sigma}_{+1}^{-1} \bar{\sigma}_{-1} W^\dagger, \quad (171)$$

$$\mathbf{b}_\rho = P_M W \eta_{+1}^{-1} \bar{\boldsymbol{\mu}}, \quad (172)$$

$$c_\rho = \frac{\exp\left(-\frac{1}{2} \bar{\boldsymbol{\mu}}^\dagger \sigma_{+1}^{-1} \bar{\boldsymbol{\mu}}\right)}{\sqrt{\det(\sigma_{+1})}}, \quad (173)$$

$$P_M = \begin{pmatrix} 0_M & \mathbb{1}_M \\ \mathbb{1}_M & 0_M \end{pmatrix}. \quad (174)$$

Note that the map $\bar{\sigma} \rightarrow \bar{\sigma}_{+1}^{-1} \bar{\sigma}_{-1}$ is known as the Caley transform and for a pure state $|\psi\rangle\langle\psi|$. It holds that

$$A_\rho = A_\psi^* \oplus A_\psi, \quad (175)$$

$$\mathbf{b}_\rho = \mathbf{b}_\psi^* \oplus \mathbf{b}_\psi. \quad (176)$$

So we see that for pure Gaussian states, we only need dimensionality M instead of $2M$.

For a single mode Gaussian state $|\psi\rangle = D(\alpha)S(re^{i\phi})|0\rangle$ we can give these parameters directly

$$A_\psi = -\tanh(r)e^{i\phi}, \quad (177)$$

$$\mathbf{b}_\psi = \alpha + \alpha^* e^{i\phi} \tanh(r), \quad (178)$$

$$c_\psi = \frac{\exp\left(-\frac{1}{2} [|\alpha|^2 + \alpha^{*2} e^{i\phi} \tanh(r)]\right)}{\sqrt{\cosh(r)}}. \quad (179)$$

If we want to compute the overlap between Gaussian states, we want to compute quantities like

$$\langle \boldsymbol{\alpha}^* | U | \boldsymbol{\beta} \rangle = \exp\left(-\frac{1}{2} [|\boldsymbol{\alpha}|^2 + |\boldsymbol{\beta}|^2]\right) c_U \exp\left(\mathbf{b}_U^T \boldsymbol{\nu} + \frac{1}{2} \boldsymbol{\nu}_U^A \boldsymbol{\nu}\right) \quad (180)$$

with $\boldsymbol{\nu} = (\boldsymbol{\alpha}, \boldsymbol{\beta})^T$, where $|\alpha\rangle, |\beta\rangle$ are multimode coherent states, and U a Gaussian unitary. These parameters can be computed directly from the symplectic matrix S and the displacement \mathbf{d} corresponding to U as follows:

$$A_U = P_{2M} R \begin{pmatrix} \mathbb{1}_{2M} - \boldsymbol{\xi}^{-1} & \boldsymbol{\xi}^{-1} S \\ S^T \boldsymbol{\xi}^{-1} & \mathbb{1}_{2M} - S^T \boldsymbol{\xi}^{-1} S \end{pmatrix} R^\dagger, \quad (181)$$

$$\mathbf{b}_U = R^* \begin{pmatrix} \boldsymbol{\xi}^{-1} \mathbf{d} \\ -S^T \boldsymbol{\xi}^{-1} \mathbf{d} \end{pmatrix}, \quad (182)$$

$$c_U = \frac{\exp\left(-\frac{1}{2} \mathbf{d}^T \boldsymbol{\xi}^{-1} \mathbf{d}\right)}{\sqrt{\det(\boldsymbol{\xi})}}, \quad (183)$$

with

$$R = \frac{1}{\sqrt{2}} \begin{pmatrix} \mathbb{1}_M & i\mathbb{1}_M & 0_M & 0_M \\ 0_M & 0_M & \mathbb{1}_M & -i\mathbb{1}_M \\ \mathbb{1}_M & -i\mathbb{1}_M & 0_M & 0_M \\ 0_M & 0_M & \mathbb{1}_M & i\mathbb{1}_M \end{pmatrix}, \quad (184)$$

$$\boldsymbol{\xi} = \frac{1}{2} (\mathbb{1}_{2M} + S S^T). \quad (185)$$

This can be computed faster by simplifying it further, see [33].

If two unitaries are applied one after another $U_f = U_1 U_2$, it is not enough to just consider $S = S_1 S_2$ and $\mathbf{d} = \mathbf{d}_1 + \mathbf{d}_2$ as the phase information is lost this way. However, one can update the parameters $A_{U_f}, \mathbf{b}_{U_f}, c_{U_f}$ to compute the overlap in a phase sensitive way. For two Gaussian unitaries $U_f = U_1 U_2$, the parameters are updated as follows:

$$A_{U_f} = B_1 \oplus D_2 + \{C_1 \oplus C_2^T\} \mathcal{Z} \{C_1^T \oplus C_2\}, \quad (186)$$

$$\mathbf{b}_{U_f}^T = [c_1^T, \mathbf{d}_2^T] [d_1^T, c_2^T] \mathcal{Z} \{C_1^T \oplus C_2\}, \quad (187)$$

$$c_{U_f} = \frac{c_{U_1} c_{U_2}}{\sqrt{\det(Y)}} \exp\left(\frac{1}{2} [d_1^T, c_2^T] \mathcal{Z} [d_1, c_2]^T\right), \quad (188)$$

where we used the notation

$$\mathbf{b}_{U_i}^T = [c_i^T, \mathbf{d}_i^T], \quad (189)$$

$$A_{U_i} = \begin{pmatrix} B_i & C_i \\ C_i^T D_i & \end{pmatrix}, \quad (190)$$

and the auxiliary quantities

$$Y = \mathbb{1}_M - D_1 B_2, \quad (191)$$

$$\mathcal{Z} = \begin{pmatrix} Y^{-1} B_2 & Y^{-1} \\ (Y^T)^{-1} & D_1 Y^{-1} \end{pmatrix}. \quad (192)$$

If more Gaussian unitaries are applied one just updates the parameters consecutively. Using Eq. (180) one can then compute the overlap using the updated parameters.

D Simulation algorithm based on Gaussian extent

D.1 Finite rank approximations

If the given non-Gaussian state $|\psi\rangle$ has an infinite Gaussian rank, such as the Fock state $|1\rangle$,

$$|\psi\rangle = \sum_{i=0}^{\infty} c_i |G_i\rangle, \quad |c_0| \geq |c_1| \geq |c_2| \geq \dots, \quad (193)$$

then we consider a state with a cut-off

$$|\tilde{\psi}_m\rangle = \sum_{i=0}^m c_i |G_i\rangle, \quad (194)$$

$$|\psi_m\rangle = \frac{1}{N_m} |\tilde{\psi}_m\rangle, \quad (195)$$

with normalization

$$N_m = \sqrt{\langle \tilde{\psi}_m | \tilde{\psi}_m \rangle} = \sqrt{\sum_{i,j=0}^m c_i^* c_j \langle G_i | G_j \rangle}. \quad (196)$$

We can check that

$$\lim_{m \rightarrow \infty} \langle \psi | \psi_m \rangle = 1, \quad (197)$$

$$\lim_{m \rightarrow \infty} N_m = 1. \quad (198)$$

In the following, we will work on this finite-rank approximation.

D.2 Standard Approach

This argument is based on Ref. [23]. We write our input state as

$$|\psi\rangle = \sum_{i=1}^{\chi} c_i |G_i\rangle,$$

where $|G_i\rangle$ is a Gaussian state characterized by a covariance matrix σ_i and a mean μ_i , so we will use the notation $|G_i\rangle = |\sigma_i, \mu_i\rangle$. We want to estimate the Born probability

$$P(x) = \frac{\langle \psi | x \rangle \langle x | \psi \rangle}{\|\psi\|^2} \quad (199)$$

$$= \frac{\sum_{i,j=1}^{\chi} \langle \sigma_j, \mu_j | x \rangle \langle x | \sigma_i, \mu_i \rangle}{\|\psi\|^2}. \quad (200)$$

Estimating the Born probability takes $\mathcal{O}(\chi^2)$ inner products. This can be improved by sampling terms $|\sigma_i, \mu_i\rangle$ in the decomposition to sparsify $|\psi\rangle$ according to

$$|\psi\rangle = \sum_{i=1}^{\chi} c_i |\sigma_i, \mu_i\rangle \quad (201)$$

$$= \|c\|_1 \sum_{i=1}^{\chi} \frac{|c_i|}{\|c\|_1} \frac{c_i}{|c_i|} |\sigma_i, \mu_i\rangle \quad (202)$$

$$= \|c\|_1 \sum_{i=1}^{\chi} p(i) \frac{c_i}{|c_i|} |\sigma_i, \mu_i\rangle \quad (203)$$

$$= \|c\|_1 \sum_{i=1}^{\chi} p(i) |\tilde{\sigma}_i, \tilde{\mu}_i\rangle. \quad (204)$$

We now consider sampling the states $|\tilde{\sigma}_i, \tilde{\mu}_i\rangle$ with probability $p(i)$. To do this, we define a random variable $|\omega_i\rangle$ that is $|\tilde{\sigma}_i, \tilde{\mu}_i\rangle$ with probability $p(i)$. By definition, it holds that $\mathbb{E}[|\omega_i\rangle] = \frac{|\psi\rangle}{\|c\|_1}$ and consequently

$$|\psi\rangle = \|c\|_1 \mathbb{E}[|\omega_i\rangle]. \quad (205)$$

Performing this sampling k times, we obtain a sparsified state

$$|\Omega\rangle = \frac{\|c\|_1}{k} \sum_{i=1}^k |\omega_i\rangle. \quad (206)$$

By definition, it holds that

$$\mathbb{E}[\langle\Omega|\psi\rangle] = \mathbb{E}[\langle\psi|\Omega\rangle] = 1. \quad (207)$$

The norm of $|\Omega\rangle$ is

$$\mathbb{E}[\langle\Omega|\Omega\rangle] = \frac{\|c\|_1^2}{k^2} \left(\mathbb{E} \left[\sum_{\alpha=1}^k \langle\omega_\alpha|\omega_\alpha\rangle \right] + \mathbb{E} \left[\sum_{\alpha \neq \beta} \langle\omega_\alpha|\omega_\beta\rangle \right] \right) \quad (208)$$

$$= \frac{\|c\|_1^2}{k} \mathbb{E}[\langle\omega_\alpha|\omega_\alpha\rangle] + \frac{1}{k^2} \sum_{\alpha \neq \beta} \mathbb{E}[\|c\|_1^2 \langle\omega_\alpha|\omega_\beta\rangle] \quad (209)$$

$$= \frac{\|c\|_1^2}{k} + \frac{1}{k^2} \sum_{\alpha \neq \beta} \langle\psi|\psi\rangle \quad (210)$$

$$= \frac{\|c\|_1^2}{k} + \frac{k(k-1)}{k^2} \quad (211)$$

$$= \frac{\|c\|_1^2}{k} + 1 - \frac{1}{k}. \quad (212)$$

The bound of the approximation is

$$\mathbb{E}[|\langle\psi|\psi\rangle - \langle\Omega|\Omega\rangle|^2] = \mathbb{E}[\langle\Omega|\Omega\rangle] + \mathbb{E}[\langle\psi|\psi\rangle] - \mathbb{E}[\langle\psi|\Omega\rangle] - \mathbb{E}[\langle\Omega|\psi\rangle] \quad (213)$$

$$= \frac{\|c\|_1^2}{k} - 1 \leq \frac{\|c\|_1^2}{k}. \quad (214)$$

To guarantee the upper bound $\mathbb{E}[|\langle\psi|\psi\rangle - \langle\Omega|\Omega\rangle|^2] \leq \delta^2$, it suffices to sample k times with

$$k = \left(\frac{\|c\|_1}{\delta} \right)^2. \quad (215)$$

In Ref. [23], a sparsification tail bound is given. If we choose $k \geq \left(\frac{\|c\|_1}{\delta} \right)^2$, then we have

$$\mathbb{E}[\langle\Omega|\Omega\rangle - 1] \leq \delta^2. \quad (216)$$

Using the triangle inequality, we get

$$\|\psi - \Omega\|^2 \leq \langle\Omega|\Omega\rangle - 1 + 2 \times |1 - \text{Re}(\langle\psi|\Omega\rangle)|. \quad (217)$$

We define a random variable

$$X_\alpha = \|c\|_1 \text{Re}(\langle\psi|\omega_\alpha\rangle), \quad (218)$$

$$\tilde{X} = \frac{1}{k} \sum_{\alpha=1}^k X_\alpha = \text{Re}(\langle\psi|\Omega\rangle). \quad (219)$$

Then, we have

$$|\text{Re}(\langle\psi|\Omega\rangle) - 1| = |\tilde{X} - \mathbb{E}[\tilde{X}]| \quad (220)$$

Here, \tilde{X} is a sample mean of k IID random variables X_α , satisfying

$$|X_\alpha| \leq \|c\|_1 |\langle \psi | w_\alpha \rangle| \leq \|c\|_1 \sqrt{F(\psi)}, \quad (221)$$

with F being the Gaussian fidelity $F(\psi) = \sup_{|G\rangle \in \mathcal{G}} \{|\langle \psi | G \rangle|^2\}$. We can then apply Hoeffding's inequality to obtain

$$\begin{aligned} & \Pr \left[\left| \text{Re}(\langle \psi | \Omega \rangle) - 1 \right| \geq \frac{\delta^2}{2} \right] \\ & \leq 2 \exp \left(-\frac{2k(\delta^2/2)^2}{(2\|c\|_1 \sqrt{F(\psi)})^2} \right) \end{aligned} \quad (222)$$

$$\leq 2 \exp \left(-\frac{\delta^2}{8F(\psi)} \right). \quad (223)$$

Consequently, it holds that

$$\Pr[\|\psi - \Omega\|^2 \leq \langle \Omega | \Omega \rangle - 1 + \delta^2] \geq 1 - 2 \exp \left(-\frac{\delta^2}{8F(\psi)} \right). \quad (224)$$

D.3 Approach by Seddon et al.

We describe another approach based on Ref. [30]. If we use the approach described in the previous section, post-selection is needed for states with a norm close to 1, but Ref. [30] proposes a sampling strategy that avoids post-selection. This strategy renormalizes and bounds the error between $|\psi\rangle$ and the ensemble

$$\rho_1 = \mathbb{E} \left(\frac{|\Omega\rangle\langle\Omega|}{\langle\Omega|\Omega\rangle} \right) = \sum_{\Omega} \Pr(\Omega) \frac{|\Omega\rangle\langle\Omega|}{\langle\Omega|\Omega\rangle} \quad (225)$$

of sparsified states. We then want to bound

$$\|\rho_1 - |\psi\rangle\langle\psi|\|_1. \quad (226)$$

First, we define

$$\rho_2 = \frac{1}{\mu} \mathbb{E}[|\Omega\rangle\langle\Omega|]. \quad (227)$$

with $\mathbb{E}[\langle\Omega|\Omega\rangle]$. The triangle inequality yields

$$\|\rho_1 + \rho_2 - \rho_2 - |\psi\rangle\langle\psi|\|_1 \leq \|\rho_1 - \rho_2\|_1 + \|\rho_2 - |\psi\rangle\langle\psi|\|_1. \quad (228)$$

The first term on the right-hand side is given by

$$\|\rho_1 - \rho_2\|_1 = \left\| \mathbb{E} \left[|\Omega\rangle\langle\Omega| \left(\frac{1}{\langle\Omega|\Omega\rangle} - \frac{1}{\mu} \right) \right] \right\|_1. \quad (229)$$

We then use Jensen's inequality to obtain

$$\|\rho_1 - \rho_2\|_1 \leq \mathbb{E} \left[\left\| |\Omega\rangle\langle\Omega| \left(\frac{1}{\langle\Omega|\Omega\rangle} - \frac{1}{\mu} \right) \right\|_1 \right] \quad (230)$$

$$= \mathbb{E} \left[\left| \langle\Omega|\Omega\rangle \left(\frac{1}{\langle\Omega|\Omega\rangle} - \frac{1}{\mu} \right) \right| \right] \quad (231)$$

$$= \frac{1}{\mu} \mathbb{E}[|\mu - \langle\Omega|\Omega\rangle|] \quad (232)$$

$$\leq \sqrt{\mathbb{E}[\mu - \langle\Omega|\Omega\rangle]^2} \quad (233)$$

$$= \sqrt{\text{Var}[\langle\Omega|\Omega\rangle]}. \quad (234)$$

To evaluate the second term on the right-hand side of (228), we use

$$\mu\rho_2 = \frac{\|c\|_1^2}{k^2} \left(\sum_{\alpha \neq \beta} \mathbb{E}[|w_\alpha\rangle\langle w_\beta|] + \sum_{\alpha} |\alpha\rangle\langle\alpha| \right) \quad (235)$$

$$= \frac{k(k-1)}{k^2} |\psi\rangle\langle\psi| + \frac{\|c\|_1^2}{k} \sigma \quad (236)$$

with $\sigma = |w_\alpha\rangle\langle w_\alpha|$. Then, we have

$$\|\rho_2 - |\psi\rangle\langle\psi|\|_1 = \frac{1}{\mu} \left\| (1 - k^{-1} - \mu) |\psi\rangle\langle\psi| + \frac{\|c\|_1^2}{k} \sigma \right\|_1 \quad (237)$$

$$= \frac{\|c\|_1^2}{k\mu} \|\sigma - |\psi\rangle\langle\psi|\| \quad (238)$$

$$\leq 2 \frac{\|c\|_1^2}{k}, \quad (239)$$

where we use $\mu^{-1} \leq 1$, the triangle inequality, and $\|\sigma\|_1 = 1$.

As for the bounds of the variance, we have

$$\text{Var}(\langle\Omega|\Omega\rangle) = \mathbb{E}[\langle\Omega|\Omega\rangle^2] - \mathbb{E}[\langle\Omega|\Omega\rangle]^2. \quad (240)$$

The right-hand side is evaluated by

$$\langle\Omega|\Omega\rangle^2 = \frac{\|c\|_1^4}{k^4} (k^2 + 2kB + B^2) \quad (241)$$

with $B = \sum_{\alpha} \sum_{\alpha \neq \beta} \langle w_\alpha | w_\beta \rangle$. Thus, we obtain

$$\mathbb{E}[\langle\Omega|\Omega\rangle^2] = \frac{\|c\|_1^4}{k^4} (k^2 + 2k\mathbb{E}[B] + \mathbb{E}[B^2]). \quad (242)$$

Therefore, we have

$$\text{Var}(\langle\Omega|\Omega\rangle) = \frac{\|c\|_1^4}{k^4} (\mathbb{E}[B^2] - \mathbb{E}[B]^2). \quad (243)$$

Following the same calculation for matching the coefficients as that found in the appendix of Ref. [30], we arrive at

$$\text{Var}(\langle\Omega|\Omega\rangle) \leq \frac{4(C-1)}{k} + 2 \left(\frac{\|c\|_1^2}{k} \right)^2 + \mathcal{O}\left(\frac{C}{k^3}\right) \quad (244)$$

with $C = \|c\|_1 \sum_i |c_i| |\langle\psi|\tilde{\sigma}_i, \tilde{\mu}_i\rangle|^2$.

Then, there is a critical precision $\delta_c = 8(C-1)/\|c\|_1^2$ such that for every target precision δ_S for which $\delta_c \leq \delta_S$, we can sample pure states from an ensemble ρ_1 , where every pure state drawn from ρ_1 has Gaussian rank at most $\lceil 4\|c\|_1^2/\delta_S \rceil$. This bound provides a critical threshold, where we get an improvement by a factor of $1/\delta_S$. For higher precision, there is no improvement obtained from this sampling strategy outside of avoiding post-selection.

E Fast norm estimation

E.1 Description of the core idea

A naive estimation of the norm of a decomposed non-Gaussian state would require a quadratic cost of the number of Gaussian states in the decomposition. To achieve better efficiency, we can estimate

the norm probabilistically by sampling random coherent states and computing the overlap. This procedure, called the fast norm estimation, results in a linear cost of the number of Gaussian states in the decomposition. To achieve this scaling, we sample a coherent state $|\alpha\rangle$ and use the completeness condition

$$\mathbb{1} = \frac{1}{\pi} \int_{\mathbb{C}} d\alpha |\alpha\rangle\langle\alpha|. \quad (245)$$

A non-trivial part of the analysis arises from the fact that the right-hand side of this completeness condition uses the integral over all α , but it is not straightforward to sample α from a uniform distribution since the displacement operators are a non-compact group and thus impossible to uniformly sample. To address this point, we need to sample α from a weighted distribution in reality. Nevertheless, in this section, we use the integral without weight to describe the core idea of our fast norm estimation procedure, while we will explain the sampling from the weighted distribution in the next section.

Using the uniform integral of a random coherent state $|\xi\rangle = D(\xi)|0\rangle$, we can define a variable X and $|\psi\rangle$

$$X = \pi^{-n} |\langle\xi|\psi\rangle|^2. \quad (246)$$

In this case, we have

$$\int_{\mathbb{C}} d\xi X = \frac{1}{\pi} \int_{\mathbb{C}} d\alpha \langle\psi|\alpha\rangle \langle\alpha|\psi\rangle = \langle\psi|\psi\rangle. \quad (247)$$

Similarly, we can write for X^2 as

$$\int_{\mathbb{C}^n} d\xi X^2 = \pi^{-2n} \int_{\mathbb{C}^n} d\xi \langle\psi \otimes \psi| D(\xi) \otimes D(\xi) |0 \otimes 0\rangle \langle 0 \otimes 0| D^\dagger(\xi) \otimes D^\dagger(\xi) |\psi \otimes \psi\rangle. \quad (248)$$

So we want to investigate

$$\mathcal{T}(|0 \otimes 0\rangle\langle 0 \otimes 0|) := \int_{\mathbb{C}^n} d\xi D(\xi) \otimes D(\xi) |0 \otimes 0\rangle\langle 0 \otimes 0| D^\dagger(\xi) \otimes D^\dagger(\xi) \quad (249)$$

in order to get further insights into X^2 . It holds in general for displacement operators that

$$\begin{aligned} D(\xi)D(\alpha)D(\xi)^\dagger \\ = D(\xi)D(\alpha)D(-\xi) \end{aligned} \quad (250)$$

$$= e^{\frac{1}{2}(\xi\alpha^* - \xi^*\alpha)} D(\xi + \alpha)D(-\xi) \quad (251)$$

$$= e^{\frac{1}{2}(\xi\alpha^* - \xi^*\alpha)} e^{\frac{1}{2}((\xi+\alpha)(-\xi)^* - (\xi+\alpha)^*(-\xi))} D(\alpha) \quad (252)$$

$$= e^{\frac{1}{2}(\xi\alpha^* - \xi^*\alpha)} e^{\frac{1}{2}(-\xi\xi^* - \alpha\xi^* + \xi^*\xi + \alpha^*\xi)} D(\alpha) \quad (253)$$

$$= e^{\xi\alpha^* - \xi^*\alpha} D(\alpha). \quad (254)$$

We can expand every state in terms of the basis of displacement operators with the characteristic function χ_ρ as

$$\rho = \frac{1}{\pi^{2n}} \int_{\mathbb{C}^n} d\alpha d\beta \chi_\rho(\alpha, \beta) D(\alpha) \otimes D(\beta). \quad (255)$$

So $\mathcal{T}(|0 \otimes 0\rangle\langle 0 \otimes 0|)$ can be computed using the following property

$$\int_{\mathbb{C}^n} d\xi [D(\xi) \otimes D(\xi)] [D(\alpha) \otimes D(\beta)] [D^\dagger(\xi) \otimes D^\dagger(\xi)] \quad (256)$$

$$= \int_{\mathbb{C}^n} d\xi (D(\xi)D(\alpha)D(\xi)^\dagger) \otimes (D(\xi)D(\beta)D(\xi)^\dagger) \quad (257)$$

$$= \int_{\mathbb{C}^n} d\xi e^{\xi\alpha^* - \xi^*\alpha} e^{\xi\beta^* - \xi^*\beta} D(\alpha) \otimes D(\beta) \quad (258)$$

$$= \int_{\mathbb{C}^n} d\xi e^{\xi(\alpha^* + \beta^*) - \xi^*(\alpha + \beta)} D(\alpha) \otimes D(\beta) \quad (259)$$

$$= \pi^n \delta(\alpha + \beta) D(\alpha) \otimes D(\beta) \quad (260)$$

and then

$$\mathcal{T}(|0 \otimes 0\rangle\langle 0 \otimes 0|) = \pi^{-2n} \int_{\mathbb{C}^n} d\boldsymbol{\alpha} d\boldsymbol{\beta} \chi_{|0 \otimes 0\rangle}(\boldsymbol{\alpha}, \boldsymbol{\beta}) \mathcal{T}(D(\boldsymbol{\alpha}) \otimes D(\boldsymbol{\beta})) \quad (261)$$

$$= \pi^{-n} \int_{\mathbb{C}^n} d\boldsymbol{\alpha} \chi_{|0 \otimes 0\rangle}(\boldsymbol{\alpha}, -\boldsymbol{\alpha}) D(\boldsymbol{\alpha}) \otimes D(-\boldsymbol{\alpha}). \quad (262)$$

We will follow the reasoning of Ref. [31] and show that that $\mathcal{T}(|0 \otimes 0\rangle\langle 0 \otimes 0|)$ is proportional to a projector. This will allow us to bound $\langle \psi \otimes \psi | \mathcal{T}(|0 \otimes 0\rangle\langle 0 \otimes 0|) | \psi \otimes \psi \rangle$ from above. So we need to show that $\mathcal{T}(|0 \otimes 0\rangle\langle 0 \otimes 0|) = \mathcal{T}(|0 \otimes 0\rangle\langle 0 \otimes 0|)^\dagger$ and $\mathcal{T}(|0 \otimes 0\rangle\langle 0 \otimes 0|) \propto \mathcal{T}(|0 \otimes 0\rangle\langle 0 \otimes 0|)^2$. By definition of \mathcal{T} in Eq. (249), we have that

$$\mathcal{T}(|0 \otimes 0\rangle\langle 0 \otimes 0|) = \mathcal{T}(|0 \otimes 0\rangle\langle 0 \otimes 0|)^\dagger. \quad (263)$$

So we just need to show that $\mathcal{T}(|0 \otimes 0\rangle\langle 0 \otimes 0|) = \mathcal{T}(|0 \otimes 0\rangle\langle 0 \otimes 0|)^2$.

Thus, it follows that

$$\mathcal{T}(|0 \otimes 0\rangle\langle 0 \otimes 0|) \mathcal{T}(|0 \otimes 0\rangle\langle 0 \otimes 0|) = \pi^{-4n} \int_{\mathbb{C}^n} d\boldsymbol{\alpha} d\boldsymbol{\beta} d\boldsymbol{c} d\boldsymbol{d} \chi_{|0 \otimes 0\rangle}(\boldsymbol{\alpha}, \boldsymbol{\beta}) \chi_{|0 \otimes 0\rangle}(\boldsymbol{c}, \boldsymbol{d}) \quad (264)$$

$$\times \mathcal{T}(D(\boldsymbol{\alpha}) \otimes D(\boldsymbol{\beta})) \mathcal{T}(D(\boldsymbol{c}) \otimes D(\boldsymbol{d})) \quad (265)$$

$$= \pi^{-2n} \int_{\mathbb{C}^n} d\boldsymbol{\alpha} d\boldsymbol{c} \chi_{|0 \otimes 0\rangle}(\boldsymbol{\alpha}, -\boldsymbol{\alpha}) \chi_{|0 \otimes 0\rangle}(\boldsymbol{c}, -\boldsymbol{c}) [D(\boldsymbol{\alpha}) \otimes D(-\boldsymbol{\alpha})] [D(\boldsymbol{c}) \otimes D(-\boldsymbol{c})] \quad (266)$$

$$= \pi^{-2n} \int_{\mathbb{C}^n} d\boldsymbol{\alpha} d\boldsymbol{c} \chi_{|0 \otimes 0\rangle}(\boldsymbol{\alpha}) \chi_{|0 \otimes 0\rangle}(\boldsymbol{c}) e^{\boldsymbol{\alpha} \boldsymbol{c}^* - \boldsymbol{\alpha}^* \boldsymbol{c}/2} D(\boldsymbol{\alpha} + \boldsymbol{c}) \otimes \chi_{|0 \otimes 0\rangle}(-\boldsymbol{\alpha}) \chi_{|0 \otimes 0\rangle}(-\boldsymbol{c}) e^{\boldsymbol{\alpha} \boldsymbol{c}^* - \boldsymbol{\alpha}^* \boldsymbol{c}/2} D(-\boldsymbol{\alpha} - \boldsymbol{c}) \quad (267)$$

$$= \pi^{-2n} \int_{\mathbb{C}^n} d\boldsymbol{\alpha} d\boldsymbol{\mu} \chi_{|0 \otimes 0\rangle}(\boldsymbol{\alpha}) \chi_{|0 \otimes 0\rangle}(\boldsymbol{\mu} - \boldsymbol{\alpha}) e^{\boldsymbol{\alpha} \boldsymbol{\mu}^* - \boldsymbol{\alpha}^* \boldsymbol{\mu}/2} D(\boldsymbol{\mu}) \otimes \chi_{|0 \otimes 0\rangle}(-\boldsymbol{\alpha}) \chi_{|0 \otimes 0\rangle}(-\boldsymbol{\mu} + \boldsymbol{\alpha}) e^{\boldsymbol{\alpha} \boldsymbol{\mu}^* - \boldsymbol{\alpha}^* \boldsymbol{\mu}/2} D(-\boldsymbol{\mu}) \quad (268)$$

$$= \pi^{-2n} \int_{\mathbb{C}^n} d\boldsymbol{\mu} D(\boldsymbol{\mu}) \otimes D(-\boldsymbol{\mu}) \left[\int_{\mathbb{C}^n} d\boldsymbol{\alpha} \chi_{|0 \otimes 0\rangle}(\boldsymbol{\alpha}) \chi_{|0 \otimes 0\rangle}(-\boldsymbol{\alpha}) \chi_{|0 \otimes 0\rangle}(\boldsymbol{\mu} - \boldsymbol{\alpha}) \chi_{|0 \otimes 0\rangle}(-\boldsymbol{\mu} + \boldsymbol{\alpha}) e^{\boldsymbol{\alpha} \boldsymbol{\mu}^* - \boldsymbol{\alpha}^* \boldsymbol{\mu}} \right] \quad (269)$$

$$= \pi^{-2n} \int_{\mathbb{C}^n} d\boldsymbol{\mu} D(\boldsymbol{\mu}) \otimes D(-\boldsymbol{\mu}) \left[\int_{\mathbb{C}^n} d\boldsymbol{\alpha} |\chi_{|0 \otimes 0\rangle}(\boldsymbol{\alpha})|^2 |\chi_{|0 \otimes 0\rangle}(\boldsymbol{\mu} - \boldsymbol{\alpha})|^2 e^{\boldsymbol{\alpha} \boldsymbol{\mu}^* - \boldsymbol{\alpha}^* \boldsymbol{\mu}} \right] \quad (270)$$

$$= \pi^{-n} 2^{-n} \int_{\mathbb{C}^n} d\boldsymbol{\mu} D(\boldsymbol{\mu}) \otimes D(-\boldsymbol{\mu}) |\chi_{|0 \otimes 0\rangle}(\boldsymbol{\mu})|^2 \quad (271)$$

$$= \pi^{-n} 2^{-n} \int_{\mathbb{C}^n} d\boldsymbol{\mu} \chi_{|0 \otimes 0\rangle}(\boldsymbol{\mu}) \chi_{|0 \otimes 0\rangle}(-\boldsymbol{\mu}) D(\boldsymbol{\mu}) \otimes D(-\boldsymbol{\mu}). \quad (272)$$

We used that

$$\int_{\mathbb{C}^n} d\boldsymbol{\alpha} |\chi_{|0 \otimes 0\rangle}(\boldsymbol{\alpha})|^2 |\chi_{|0 \otimes 0\rangle}(\boldsymbol{\mu} - \boldsymbol{\alpha})|^2 e^{(\boldsymbol{\alpha} \boldsymbol{\mu}^* - \boldsymbol{\alpha}^* \boldsymbol{\mu})} \quad (273)$$

$$= \int_{\mathbb{C}^n} d\boldsymbol{\alpha} e^{-|\boldsymbol{\alpha}|^2} e^{-|\boldsymbol{\mu} - \boldsymbol{\alpha}|^2} e^{(\boldsymbol{\alpha} \boldsymbol{\mu}^* - \boldsymbol{\alpha}^* \boldsymbol{\mu})} \quad (274)$$

$$= e^{-|\boldsymbol{\mu}|^2} \int_{\mathbb{C}^n} d\boldsymbol{\alpha} e^{-2|\boldsymbol{\alpha}|^2 + 2\boldsymbol{\mu}^* \boldsymbol{\alpha}} \quad (275)$$

$$= e^{-|\boldsymbol{\mu}|^2} \int_{\mathbb{R}^n} d\boldsymbol{\alpha}_r d\boldsymbol{\alpha}_i e^{-2(\boldsymbol{\alpha}_r^2 + \boldsymbol{\alpha}_i^2 + \boldsymbol{\alpha}_r \boldsymbol{\mu}_r - \boldsymbol{\alpha}_i \boldsymbol{\mu}_i + i\boldsymbol{\alpha}_r \boldsymbol{\mu}_i - i\boldsymbol{\mu}_r \boldsymbol{\alpha}_i)} \quad (276)$$

$$= e^{-|\boldsymbol{\mu}|^2} \int_{\mathbb{R}^n} d\boldsymbol{\alpha}_r e^{-2\boldsymbol{\alpha}_r^2 + \boldsymbol{\alpha}_r (2\boldsymbol{\mu}_r - 2i\boldsymbol{\mu}_i)} \int_{\mathbb{R}^n} d\boldsymbol{\alpha}_i e^{-2\boldsymbol{\alpha}_i^2 + \boldsymbol{\alpha}_i (2\boldsymbol{\mu}_i + 2i\boldsymbol{\mu}_r)} \quad (277)$$

$$= e^{-|\boldsymbol{\mu}|^2} \frac{\pi^n}{2^n} e^{(\boldsymbol{\mu}_r - i\boldsymbol{\mu}_i)^2/2} e^{(\boldsymbol{\mu}_i + i\boldsymbol{\mu}_r)^2/2} \quad (278)$$

$$= |\chi_{|0 \otimes 0\rangle}(\boldsymbol{\mu})|^2 \frac{\pi^n}{2^n} \quad (279)$$

and $\chi_{|0 \otimes 0\rangle}(a) = e^{-\frac{1}{2}|a|^2}$. By comparing the results of $\mathcal{T}(|0 \otimes 0\rangle\langle 0 \otimes 0|)\mathcal{T}(|0 \otimes 0\rangle\langle 0 \otimes 0|)$ with

$$\mathcal{T}(|0 \otimes 0\rangle\langle 0 \otimes 0|) = \pi^{-n} \int_{\mathbb{C}^n} d\boldsymbol{\alpha} \chi_{|0 \otimes 0\rangle}(\boldsymbol{\alpha}, -\boldsymbol{\alpha}) D(\boldsymbol{\alpha}) \otimes D(-\boldsymbol{\alpha}), \quad (280)$$

we observe that they are the same up to a scaling factor

$$\mathcal{T}(|0 \otimes 0\rangle\langle 0 \otimes 0|)^2 = 2^{-n} \mathcal{T}(|0 \otimes 0\rangle\langle 0 \otimes 0|). \quad (281)$$

By multiplying $\mathcal{T}(|0 \otimes 0\rangle\langle 0 \otimes 0|)$ with 2^n , we get the projector $\Pi = 2^n \mathcal{T}(|0 \otimes 0\rangle\langle 0 \otimes 0|)$. This approach is, however, not realistic. The displacement operators are a non-compact group and thus impossible to sample uniformly. We address this problem in the next section.

E.2 Fast norm estimation by sampling from Gaussian ensemble

We can make it possible to sample from a distribution of displacements by introducing a Gaussian ensemble of displacements [39]. The group of all displacements is not compact, and we cannot sample it uniformly. Reference [39] uses sampling from a Gaussian ensemble

$$\mathcal{D}_N = \left\{ D(\boldsymbol{\xi}) : \boldsymbol{\xi} \sim P_D^G(\boldsymbol{\xi}, N) = \frac{e^{-|\boldsymbol{\xi}|^2/N}}{\pi^n N^n} \right\}, \quad (282)$$

which can be used in CV state tomography and reproduces the identity in the limit

$$\lim_{N \rightarrow \infty} N^n \int_{\mathbb{C}^n} d\boldsymbol{\xi} P_D^G(\boldsymbol{\xi}, N) D(\boldsymbol{\xi}) |0\rangle\langle 0| D^\dagger(\boldsymbol{\xi}) = \mathbf{1}. \quad (283)$$

It also holds that

$$\begin{aligned} & \lim_{N \rightarrow \infty} N^n \int_{\mathbb{C}^n} d\boldsymbol{\xi} P_D^G(\boldsymbol{\xi}, N) D(\boldsymbol{\xi})^{\otimes 2} |0 \otimes 0\rangle\langle 0 \otimes 0| D^\dagger(\boldsymbol{\xi})^{\otimes 2} \\ &= \frac{1}{\pi^n} \int_{\mathbb{C}^n} d\boldsymbol{\xi} D(\boldsymbol{\xi})^{\otimes 2} |0 \otimes 0\rangle\langle 0 \otimes 0| D^\dagger(\boldsymbol{\xi})^{\otimes 2}. \end{aligned} \quad (284)$$

Our contribution here is to propose sampling from the Gaussian ensemble as an approximation of the procedure in the previous section. We present an analysis in the following.

The above limits ensure that for any $\delta > 0$, there is a sufficiently large N_δ such that, for a state $|\Omega\rangle$ of interest,

$$\left| N_\delta^n \int_{\mathbb{C}^n} d\boldsymbol{\xi} P_D^G(\boldsymbol{\xi}, N_\delta) \langle \Omega | D(\boldsymbol{\xi}) |0\rangle\langle 0| D^\dagger(\boldsymbol{\xi}) | \Omega \rangle - \langle \Omega | \Omega \rangle \right| \leq \delta \langle \Omega | \Omega \rangle \quad (285)$$

and

$$\begin{aligned} & \left| \pi^n N_\delta^n \int_{\mathbb{C}^n} d\boldsymbol{\xi} P_D^G(\boldsymbol{\xi}, N_\delta) \langle \Omega \otimes \Omega | D(\boldsymbol{\xi}) \otimes D(\boldsymbol{\xi}) |0 \otimes 0\rangle\langle 0 \otimes 0| D^\dagger(\boldsymbol{\xi}) \otimes D^\dagger(\boldsymbol{\xi}) | \Omega \otimes \Omega \rangle \right. \\ & \quad \left. - \int_{\mathbb{C}^n} d\boldsymbol{\xi} \langle \Omega \otimes \Omega | (D(\boldsymbol{\xi}) \otimes D(\boldsymbol{\xi})) |0 \otimes 0\rangle\langle 0 \otimes 0| (D^\dagger(\boldsymbol{\xi}) \otimes D^\dagger(\boldsymbol{\xi})) | \Omega \otimes \Omega \rangle \right| \leq \delta \langle \Omega \otimes \Omega | \Omega \otimes \Omega \rangle. \end{aligned} \quad (286)$$

Equivalently to the previous section, using Eq. (285), we define

$$\mathcal{T}_G(|0 \otimes 0\rangle\langle 0 \otimes 0|) := \pi^n N_\delta^n \int_{\mathbb{C}^n} d\xi \frac{e^{-|\xi|^2/N_\delta}}{\pi^n N_\delta^n} D(\xi) \otimes D(\xi) |0 \otimes 0\rangle\langle 0 \otimes 0| D^\dagger(\xi) \otimes D^\dagger(\xi) \quad (287)$$

$$= \frac{\pi^n N_\delta^n}{\pi^{2n}} \int_{\mathbb{C}^n} d\alpha d\beta d\xi \chi_{|0 \otimes 0\rangle}(\alpha, \beta) \frac{e^{-|\xi|^2/N_\delta}}{\pi^n N_\delta^n} (D(\xi) \otimes D(\xi))(D(\alpha) \otimes D(\beta))(D^\dagger(\xi) \otimes D^\dagger(\xi)) \quad (288)$$

$$= \frac{\pi^n N_\delta^n}{\pi^{2n}} \int_{\mathbb{C}^n} d\alpha d\beta d\xi \chi_{|0 \otimes 0\rangle}(\alpha, \beta) \frac{e^{-|\xi|^2/N_\delta}}{\pi^n N_\delta^n} e^{\xi(\alpha^* + \beta^*) - \xi^*(\alpha + \beta)} D(\alpha) \otimes D(\beta) \quad (289)$$

$$= \frac{\pi^n N_\delta^n}{\pi^{2n}} \int_{\mathbb{C}^n} d\alpha d\beta \chi_{|0 \otimes 0\rangle}(\alpha, \beta) D(\alpha) \otimes D(\beta) \quad (290)$$

$$\times \int_{\mathbb{C}^n} d\xi_r d\xi_i \frac{e^{-|\xi_r|^2/N_\delta} e^{-|\xi_i|^2/N_\delta}}{\pi^n N_\delta^n} e^{\xi_r(\alpha^* + \beta^*) - \xi_r(\alpha + \beta)} e^{i\xi_i(\alpha^* + \beta^*) + i\xi_i(\alpha + \beta)} \quad (291)$$

$$= \frac{\pi^n N_\delta^n}{\pi^{2n}} \int_{\mathbb{C}^n} d\alpha d\beta \chi_{|0 \otimes 0\rangle}(\alpha, \beta) e^{-N_\delta |\alpha + \beta|^2} D(\alpha) \otimes D(\beta), \quad (292)$$

which will approach the result in the previous section in the limit of $N_\delta \rightarrow \infty$.

Importantly, we can use the results from the previous section that

$$\mathcal{T}(|0 \otimes 0\rangle\langle 0 \otimes 0|) = \int_{\mathbb{C}^n} d\xi D(\xi) \otimes D(\xi) |0 \otimes 0\rangle\langle 0 \otimes 0| D^\dagger(\xi) \otimes D^\dagger(\xi) \quad (293)$$

is proportional to a projector $\Pi = 2^n \mathcal{T}(|0 \otimes 0\rangle\langle 0 \otimes 0|)$. In particular, Eq. (286) immediately implies that

$$\left| \pi^n N_\delta^n \int_{\mathbb{C}^n} d\xi P_D^G(\xi, N_\delta) \langle \Omega \otimes \Omega | D(\xi)^{\otimes 2} |0 \otimes 0\rangle\langle 0 \otimes 0| D^\dagger(\xi)^{\otimes 2} | \Omega \otimes \Omega \rangle - 2^{-n} \langle \Omega \otimes \Omega | \Pi | \Omega \otimes \Omega \rangle \right| \quad (294)$$

$$\leq \delta \langle \Omega \otimes \Omega | \Omega \otimes \Omega \rangle. \quad (295)$$

We can then define a random variable X as

$$X = N_\delta^n |\langle \xi | \Omega \rangle|^2. \quad (296)$$

Thus, by sampling coherent states from the Gaussian ensemble, we can resolve the identity and estimate the norm

$$|\mathbb{E}[X] - \langle \Omega | \Omega \rangle| = \left| N_\delta^n \int_{\mathbb{C}^n} d\xi \frac{e^{-|\xi|^2/N_\delta}}{\pi^n N_\delta^n} \langle \Omega | D(\xi) |0\rangle\langle 0| D^\dagger(\xi) | \Omega \rangle - \langle \Omega | \Omega \rangle \right| \leq \delta \langle \Omega | \Omega \rangle, \quad (297)$$

where the inequality follows from Eq. (285).

Due to Eq. (294), the variance can be bounded by

$$\text{Var}[X] \leq \mathbb{E}[X^2] \quad (298)$$

$$= N_\delta^{2n} \int_{\mathbb{C}^n} d\xi \frac{e^{-|\xi|^2/N_\delta}}{\pi^n N_\delta^n} \langle \Omega \otimes \Omega | D(\xi) \otimes D(\xi) |0 \otimes 0\rangle\langle 0 \otimes 0| D^\dagger(\xi) \otimes D^\dagger(\xi) | \Omega \otimes \Omega \rangle \quad (299)$$

$$\leq \frac{N_\delta^n}{\pi^n} 2^{-n} \langle \Omega \otimes \Omega | \Pi | \Omega \otimes \Omega \rangle + \delta \langle \Omega \otimes \Omega | \Omega \otimes \Omega \rangle \quad (300)$$

$$\leq \frac{2^{-n} N_\delta^n + \delta \pi^n}{\pi^n} \|\Omega\|^4. \quad (301)$$

Now, if we define an estimator as

$$\eta = \frac{1}{L} \sum_{i=1}^L N_{\delta}^n |\langle \xi | \Omega \rangle|^2. \quad (302)$$

Then, η has expectation value $\|\Omega\|^2$ up to $\pm\delta$ and variance $\sigma^2 \leq L^{-1}(2^{-n}N_{\delta}^n + \delta\pi^n)/\pi^n\|\Omega\|^4$. Using Chebyshev's inequality and choosing the number of samples as

$$L = \frac{2^{-n}N_{\delta}^n + \delta\pi^n}{\pi^n} \epsilon^{-2} p_f^{-1}, \quad (303)$$

we have that, with a probability of at least $1 - p_f$,

$$(1 - \epsilon - \delta)\|\Omega\|^2 \leq \eta \leq (1 + \epsilon + \delta)\|\Omega\|^2. \quad (304)$$

F Refined Bounds

The output of the channel

$$\int_{\mathbb{C}^n} d\xi P_D^G(\xi, M) D(\xi) |0\rangle\langle 0| D^\dagger(\xi) = \rho_M^{th} \quad (305)$$

is the thermal state with mean photon number M

$$\rho_M^{th} = \frac{1}{M+1} \sum_{n=0}^{\infty} \left(\frac{M}{M+1}\right)^n |n\rangle\langle n|. \quad (306)$$

This channel is known in the literature as the classical-noise channel [61]. Therefore, the integral is upper bounded by

$$\int_{\mathbb{C}^n} d\xi P_D^G(\xi, M) \langle \Omega | D(\xi) |0\rangle\langle 0| D^\dagger(\xi) | \Omega \rangle \leq \langle \Omega | \Omega \rangle. \quad (307)$$

Furthermore it holds for $\langle \Omega | \Omega \rangle = \sum_{n=0}^{\infty} |\Omega_n|^2$. Thus

$$M \langle \Omega | \rho_M^{th} | \Omega \rangle = \frac{M}{M+1} \sum_{n=0}^{\infty} |\Omega_n|^2 \left(\frac{M}{M+1}\right)^n \leq \sum_{n=0}^{\infty} |\Omega_n|^2 = \langle \Omega | \Omega \rangle. \quad (308)$$

We can lower bound this integral using the mean photon number of the state $|\Omega\rangle$ $N_{\Omega} = \frac{\langle \Omega | n | \Omega \rangle}{\langle \Omega | \Omega \rangle} = \frac{1}{\pi^n \langle \Omega | \Omega \rangle} \int_{\mathbb{C}^n} d\xi |\xi|^2 |\langle \xi | \Omega \rangle|^2$. It holds for a Gaussian function that $e^{-x^2} \geq (1 - x^2)$. So by expanding $P_D^G(\xi, N)$ to the first order we get

$$\begin{aligned} & \int_{\mathbb{C}^n} d\xi P_D^G(\xi, N) \langle \Omega | D(\xi) |0\rangle\langle 0| D^\dagger(\xi) | \Omega \rangle = \int_{\mathbb{C}^n} d\xi \frac{e^{-|\xi|^2/N}}{\pi^n N^n} |\langle \xi | \Omega \rangle|^2 \\ & \geq \frac{1}{\pi^n N^n} \int_{\mathbb{C}^n} d\xi \left(1 - \frac{|\xi|^2}{N}\right) |\langle \xi | \Omega \rangle|^2 \\ & = \frac{1}{N^n} \langle \Omega | \Omega \rangle \left(1 - \frac{N_{\Omega}}{N}\right). \end{aligned} \quad (309)$$

We can then define a random variable X as

$$X = N^n |\langle \xi | \Omega \rangle|^2. \quad (310)$$

We can then estimate the norm of $|\Omega\rangle$ by sampling coherent states from the Gaussian ensemble, since

$$\mathbb{E}(X) = N^n \int_{\mathbb{C}^n} d\xi P_D^G(\xi, N) |\langle \xi | \Omega \rangle|^2 \quad (311)$$

with

$$\langle \Omega | \Omega \rangle \left(1 - \frac{N_\Omega}{N} \right) \leq \mathbb{E}(X) \leq \langle \Omega | \Omega \rangle. \quad (312)$$

The variance is bounded in the following way.

$$\text{Var}[X] \leq \mathbb{E}[X^2] \quad (313)$$

$$= N^{2n} \int_{\mathbb{C}^n} d\xi \frac{e^{-|\xi|^2/N}}{\pi^n N^n} \langle \Omega \otimes \Omega | D(\xi) \otimes D(\xi) | 0 \otimes 0 \rangle \langle 0 \otimes 0 | D^\dagger(\xi) \otimes D^\dagger(\xi) | \Omega \otimes \Omega \rangle \quad (314)$$

$$\leq \frac{N^n}{\pi^n} \int_{\mathbb{C}^n} d\xi |\langle \xi \otimes \xi | \Omega \otimes \Omega \rangle|^2 \quad (315)$$

$$= \frac{N^n}{\pi^n} 2^{-n} \langle \Omega \otimes \Omega | \Pi | \Omega \otimes \Omega \rangle \quad (316)$$

$$\leq \frac{N^n}{2^n \pi^n} \|\Omega\|^4. \quad (317)$$

G Proof of Eq. (101)

Let $f \subset \mathcal{H}$ be a subset of separable Hilbert space \mathcal{H} . Consider the extent measure with respect to the set f defined by

$$\xi_f(|\psi\rangle) = \inf \left\{ \left(\sum_i c_i \right)^2 \mid |\psi\rangle = \sum_i c_i |\phi_i\rangle, |\phi_i\rangle \in f \right\}. \quad (318)$$

We also define

$$\bar{\xi}_f(|\psi\rangle) := \inf \left\{ \mu^2 \mid |\psi\rangle \in \mu \text{ cl conv } f \right\}. \quad (319)$$

This allows for an alternative expression [27, 62]

$$\bar{\xi}_f(|\psi\rangle) = \sup \left\{ |\langle w | \psi \rangle|^2 \mid |\langle w | \phi \rangle| \leq 1, \forall |\phi\rangle \in f \right\}. \quad (320)$$

As shown in Ref. [27], this quantity coincides with the lower semicontinuous robustness

$$\bar{\xi}_f(|\psi\rangle) = \underline{R}_{\mathcal{F}}(|\psi\rangle\langle\psi|) \quad (321)$$

for $\mathcal{F} = \text{cl conv} \left\{ |\phi\rangle\langle\phi| \mid |\phi\rangle \in f \right\}$.

Let us now assume that a decomposition $|\psi\rangle = \sum_i \tilde{c}_i |\tilde{\phi}_i\rangle$ with $|\tilde{\phi}_i\rangle \in f$ satisfies

$$\left(\sum_i |\tilde{c}_i| \right)^2 = \xi_f(|\psi\rangle) = \bar{\xi}_f(|\psi\rangle) = \underline{R}_{\mathcal{F}}(|\psi\rangle\langle\psi|) = \text{Tr}(\tilde{W} |\psi\rangle\langle\psi|) \quad (322)$$

for an optimal witness operator \tilde{W} , which appears in the definition of $\underline{R}_{\mathcal{F}}$. Note that

$$\begin{aligned} \underline{R}_{\mathcal{F}}(|\psi\rangle\langle\psi|) &= \sup \left\{ \text{Tr}(W\psi) \mid W \geq 0, \text{Tr}(W\sigma) \leq 1, \sigma \in \mathcal{F} \right\} \\ &\geq \sup \left\{ |\langle w | \psi \rangle|^2 \mid |\langle w | \phi \rangle| \leq 1, \phi \in f \right\} \\ &= \bar{\xi}_f(|\psi\rangle) \end{aligned} \quad (323)$$

where in the second line we restricted W to the form $W = |w\rangle\langle w|$ for some unnormalized vector $|w\rangle$, and the third line is because of (320). This, together with (321), ensures that the optimal witness \tilde{W}

in (322) takes the form $\tilde{W} = |\tilde{w}\rangle\langle\tilde{w}|$ for some unnormalized vector $|\tilde{w}\rangle$ satisfying $|\langle\tilde{w}|\phi\rangle| \leq 1$ for every $|\phi\rangle \in f$. Therefore, we get

$$\begin{aligned}\bar{\xi}_f(|\psi\rangle) &= \left| \sum_i \tilde{c}_i \langle\tilde{w}|\tilde{\phi}_i\rangle \right|^2 \\ &\leq \left(\sum_i |\tilde{c}_i| |\langle\tilde{w}|\tilde{\phi}_i\rangle| \right)^2 \\ &\leq \left(\sum_i |\tilde{c}_i| \right)^2 \\ &= \xi_f(|\psi\rangle)\end{aligned}\tag{324}$$

where in the first line we wrote $|\psi\rangle = \sum_i \tilde{c}_i |\tilde{\phi}_i\rangle$ and used that (320) is achieved with $|\tilde{w}\rangle$ due to (323), the second line is due to the triangle inequality, and the third line is because $|\langle\tilde{w}|\phi\rangle| \leq 1, \forall |\phi\rangle \in f$. Then, (322) implies that these all coincide, and in particular,

$$\langle\tilde{w}|\tilde{\phi}_i\rangle = \frac{\tilde{c}_i}{|\tilde{c}_i|}, \quad \forall \tilde{c}_i \neq 0.\tag{325}$$

This gives

$$\left| \text{Tr}(\tilde{W} |\tilde{\phi}_i\rangle\langle\tilde{\phi}_i|) \right| = 1, \quad \forall i\tag{326}$$

concluding the proof.

H Numerical evaluation of maximal fidelity between Gaussian states and two Fock states

We want to compute the maximal overlap between a two-mode Gaussian state and two Fock states $|1\rangle \otimes |1\rangle$. We numerically find an approximate solution of an optimization problem of maximizing the fidelity by parameterizing the unitaries in $|G'\rangle = U(\phi, \xi) S(r_1 e^{i\theta_1}) \otimes S(r_2 e^{i\theta_2}) D(\alpha_1) \otimes D(\alpha_2) |0\rangle \otimes |0\rangle$, where U is a passive symplectic unitary. In this optimization, we numerically optimize the overlap between $|G'\rangle$ and $|1\rangle \otimes |1\rangle$. We used the parameterization used in Ref. [63]. The parameters that we numerically found are

$$\alpha_1 = 0,\tag{327}$$

$$\alpha_2 = 0,\tag{328}$$

$$r_1 = 0.8814,\tag{329}$$

$$\theta_1 = 0.609,\tag{330}$$

$$r_2 = 0.8814,\tag{331}$$

$$\theta_2 = 1.107,\tag{332}$$

$$\phi = -1.322,\tag{333}$$

$$\xi = 1.571.\tag{334}$$

This yields

$$|\langle 1 \otimes 1 | G' \rangle|^2 = 0.25 = \frac{1}{4},\tag{335}$$

which coincides with the result reported in Ref. [64].

References

- [1] Nissim Ofek, Andrei Petrenko, Reinier Heeres, Philip Reinhold, Zaki Leghtas, Brian Vlastakis, Yehan Liu, Luigi Frunzio, S. M. Girvin, Liang Jiang, Mazyar Mirrahimi, M. H. Devoret, and R. J. Schoelkopf. “Extending the lifetime of a quantum bit with error correction in superconducting circuits”. *Nature* **536**, 441–445 (2016).
- [2] VV Sivak, Alec Eickbusch, Baptiste Royer, Shraddha Singh, Ioannis Tsioutsios, Suhas Ganjam, Alessandro Miano, BL Brock, AZ Ding, Luigi Frunzio, et al. “Real-time quantum error correction beyond break-even”. *Nature* **616**, 50–55 (2023).
- [3] Olivier Pfister. “Continuous-variable quantum computing in the quantum optical frequency comb”. *Journal of Physics B: Atomic, Molecular and Optical Physics* **53**, 012001 (2020).
- [4] Alexandre Blais, Arne L. Grimsmo, S. M. Girvin, and Andreas Wallraff. “Circuit quantum electrodynamics”. *Rev. Mod. Phys.* **93**, 025005 (2021).
- [5] Arne L. Grimsmo and Alexandre Blais. “Squeezing and quantum state engineering with josephson traveling wave amplifiers”. *npj Quantum Information* **3**, 20 (2017).
- [6] Michael Schmidt, Max Ludwig, and Florian Marquardt. “Optomechanical circuits for nanomechanical continuous variable quantum state processing”. *New Journal of Physics* **14**, 125005 (2012).
- [7] Oussama Houhou, Habib Aissaoui, and Alessandro Ferraro. “Generation of cluster states in optomechanical quantum systems”. *Physical Review A* **92**, 063843 (2015).
- [8] A. Mari and J. Eisert. “Positive wigner functions render classical simulation of quantum computation efficient”. *Phys. Rev. Lett.* **109**, 230503 (2012).
- [9] Julien Niset, Jaromír Fiurášek, and Nicolas J. Cerf. “No-go theorem for gaussian quantum error correction”. *Phys. Rev. Lett.* **102**, 120501 (2009).
- [10] Victor Veitch, Nathan Wiebe, Christopher Ferrie, and Joseph Emerson. “Efficient simulation scheme for a class of quantum optics experiments with non-negative wigner representation”. *New Journal of Physics* **15**, 013037 (2013).
- [11] Hakop Pashayan, Joel J. Wallman, and Stephen D. Bartlett. “Estimating outcome probabilities of quantum circuits using quasiprobabilities”. *Phys. Rev. Lett.* **115**, 070501 (2015).
- [12] Anatole Kenfack and Karol Życzkowski. “Negativity of the wigner function as an indicator of non-classicality”. *J. Opt. B: Quantum Semiclass. Opt.* **6**, 396 (2004).
- [13] Ryuji Takagi and Quntao Zhuang. “Convex resource theory of non-Gaussianity”. *Phys. Rev. A* **97**, 062337 (2018).
- [14] Francesco Albarelli, Marco G. Genoni, Matteo G. A. Paris, and Alessandro Ferraro. “Resource theory of quantum non-gaussianity and wigner negativity”. *Phys. Rev. A* **98**, 052350 (2018).
- [15] Laura García-Álvarez, Cameron Calcluth, Alessandro Ferraro, and Giulia Ferrini. “Efficient simulatability of continuous-variable circuits with large Wigner negativity”. *Phys. Rev. Research* **2**, 043322 (2020).
- [16] Cameron Calcluth, Alessandro Ferraro, and Giulia Ferrini. “Efficient simulation of Gottesman-Kitaev-Preskill states with Gaussian circuits”. *Quantum* **6**, 867 (2022).
- [17] Cameron Calcluth, Alessandro Ferraro, and Giulia Ferrini. “Vacuum provides quantum advantage to otherwise simulatable architectures”. *Phys. Rev. A* **107**, 062414 (2023).
- [18] Oliver Hahn, Giulia Ferrini, and Ryuji Takagi. “Bridging magic and non-gaussian resources via gottesman-kitaev-preskill encoding”. *PRX Quantum* **6**, 010330 (2025).
- [19] Ulysse Chabaud, Giulia Ferrini, Frédéric Grosshans, and Damian Markham. “Classical simulation of gaussian quantum circuits with non-gaussian input states”. *Phys. Rev. Res.* **3**, 033018 (2021).
- [20] Ulysse Chabaud and Mattia Walschaers. “Resources for bosonic quantum computational advantage”. *Phys. Rev. Lett.* **130**, 090602 (2023).
- [21] Ulysse Chabaud, Damian Markham, and Frédéric Grosshans. “Stellar representation of non-gaussian quantum states”. *Phys. Rev. Lett.* **124**, 063605 (2020).

- [22] J. Eli Bourassa, Nicolás Quesada, Ilan Tzitrin, Antal Száva, Theodor Isacsson, Josh Izaac, Krishna Kumar Sabapathy, Guillaume Dauphinais, and Ish Dhand. “Fast simulation of bosonic qubits via gaussian functions in phase space”. *PRX Quantum* **2**, 040315 (2021).
- [23] Sergey Bravyi, Dan Browne, Padraic Calpin, Earl Campbell, David Gosset, and Mark Howard. “Simulation of quantum circuits by low-rank stabilizer decompositions”. *Quantum* **3**, 181 (2019).
- [24] Beatriz Dias and Robert Koenig. “Classical simulation of non-Gaussian fermionic circuits”. *Quantum* **8**, 1350 (2024).
- [25] Jeffrey Marshall and Namit Anand. “Simulation of quantum optics by coherent state decomposition”. *Optica Quantum* **1**, 78 (2023).
- [26] Bartosz Regula, Ludovico Lami, Giovanni Ferrari, and Ryuji Takagi. “Operational quantification of continuous-variable quantum resources”. *Phys. Rev. Lett.* **126**, 110403 (2021).
- [27] Ludovico Lami, Bartosz Regula, Ryuji Takagi, and Giovanni Ferrari. “Framework for resource quantification in infinite-dimensional general probabilistic theories”. *Phys. Rev. A* **103**, 032424 (2021).
- [28] Alessio Serafini. “Quantum continuous variables: a primer of theoretical methods”. *CRC press*. (2017).
- [29] Froilán M Dopico and Charles R Johnson. “Parametrization of the matrix symplectic group and applications”. *SIAM journal on matrix analysis and applications* **31**, 650–673 (2009).
- [30] James R. Seddon, Bartosz Regula, Hakop Pashayan, Yingkai Ouyang, and Earl T. Campbell. “Quantifying quantum speedups: Improved classical simulation from tighter magic monotones”. *PRX Quantum* **2**, 010345 (2021).
- [31] Sergey Bravyi and David Gosset. “Complexity of Quantum Impurity Problems”. *Communications in Mathematical Physics* **356**, 451–500 (2017).
- [32] Gaetana Spedalieri, Christian Weedbrook, and Stefano Pirandola. “A limit formula for the quantum fidelity”. *Journal of Physics A: Mathematical and Theoretical* **46**, 025304 (2012).
- [33] Yuan Yao, Filippo Miatto, and Nicolás Quesada. “Riemannian optimization of photonic quantum circuits in phase and Fock space”. *SciPost Phys.* **17**, 082 (2024).
- [34] Daniel Gottesman, Alexei Kitaev, and John Preskill. “Encoding a qubit in an oscillator”. *Phys. Rev. A* **64**, 012310 (2001).
- [35] Ben Q. Baragiola, Giacomo Pantaleoni, Rafael N. Alexander, Angela Karanjai, and Nicolas C. Menicucci. “All-gaussian universality and fault tolerance with the gottesman-kitaev-preskill code”. *Phys. Rev. Lett.* **123**, 200502 (2019).
- [36] Hayata Yamasaki, Takaya Matsuura, and Masato Koashi. “Cost-reduced all-gaussian universality with the gottesman-kitaev-preskill code: Resource-theoretic approach to cost analysis”. *Phys. Rev. Res.* **2**, 023270 (2020).
- [37] Shohini Ghose and Barry C. Sanders. “Non-gaussian ancilla states for continuous variable quantum computation via gaussian maps”. *Journal of Modern Optics* **54**, 855–869 (2007).
- [38] Sergey Bravyi and David Gosset. “Improved classical simulation of quantum circuits dominated by clifford gates”. *Phys. Rev. Lett.* **116**, 250501 (2016).
- [39] Quntao Zhuang, Thomas Schuster, Beni Yoshida, and Norman Y. Yao. “Scrambling and complexity in phase space”. *Phys. Rev. A* **99**, 062334 (2019).
- [40] V. Bargmann, P. Butera, L. Girardello, and John R. Klauder. “On the completeness of the coherent states”. *Reports on Mathematical Physics* **2**, 221–228 (1971).
- [41] A M Perelomov. “On the completeness of a system of coherent states”. *Theoretical and Mathematical Physics* **6**, 156–164 (1971).
- [42] Eric Chitambar and Gilad Gour. “Quantum resource theories”. *Rev. Mod. Phys.* **91**, 025001 (2019).
- [43] Kohdai Kuroiwa and Hayata Yamasaki. “General Quantum Resource Theories: Distillation, Formation and Consistent Resource Measures”. *Quantum* **4**, 355 (2020).
- [44] Sergey Bravyi, Graeme Smith, and John A. Smolin. “Trading classical and quantum computational resources”. *Phys. Rev. X* **6**, 021043 (2016).

- [45] M. Mraz, J. Sperling, W. Vogel, and B. Hage. “Witnessing the degree of nonclassicality of light”. *Phys. Rev. A* **90**, 033812 (2014).
- [46] Arne Heimendahl, Felipe Montealegre-Mora, Frank Vallentin, and David Gross. “Stabilizer extent is not multiplicative”. *Quantum* **5**, 400 (2021).
- [47] Arne L. Grimsmo, Joshua Combes, and Ben Q. Baragiola. “Quantum computing with rotation-symmetric bosonic codes”. *Phys. Rev. X* **10**, 011058 (2020).
- [48] Shunya Konno, Warit Asavanant, Fumiya Hanamura, Hironari Nagayoshi, Kosuke Fukui, Atsushi Sakaguchi, Ryuhoh Ide, Fumihiko China, Masahiro Yabuno, Shigehito Miki, Hirotaka Terai, Kan Takase, Mamoru Endo, Petr Marek, Radim Filip, Peter van Loock, and Akira Furusawa. “Logical states for fault-tolerant quantum computation with propagating light”. *Science* **383**, 289–293 (2024).
- [49] Oliver Hahn, Alessandro Ferraro, Lina Hultquist, Giulia Ferrini, and Laura García-Álvarez. “Quantifying qubit magic resource with Gottesman-Kitaev-Preskill encoding”. *Phys. Rev. Lett.* **128**, 210502 (2022).
- [50] Takaya Matsuura, Hayata Yamasaki, and Masato Koashi. “Equivalence of approximate Gottesman-Kitaev-Preskill codes”. *Phys. Rev. A* **102**, 032408 (2020).
- [51] Arne L. Grimsmo and Shruti Puri. “Quantum error correction with the Gottesman-Kitaev-Preskill code”. *PRX Quantum* **2**, 020101 (2021).
- [52] Scott Aaronson and Alex Arkhipov. “The computational complexity of linear optics”. In Proceedings of the Forty-Third Annual ACM Symposium on Theory of Computing. Page 333–342. STOC ’11 New York, NY, USA (2011). Association for Computing Machinery.
- [53] Craig S. Hamilton, Regina Kruse, Linda Sansoni, Sonja Barkhofen, Christine Silberhorn, and Igor Jex. “Gaussian boson sampling”. *Phys. Rev. Lett.* **119**, 170501 (2017).
- [54] U. Chabaud, T. Douce, D. Markham, P. van Loock, E. Kashefi, and G. Ferrini. “Continuous-variable sampling from photon-added or photon-subtracted squeezed states”. *Phys. Rev. A* **96**, 062307 (2017).
- [55] Daniel J. Weigand and Barbara M. Terhal. “Generating grid states from Schrödinger-cat states without postselection”. *Phys. Rev. A* **97**, 022341 (2018).
- [56] Vlad Gheorghiu, Michele Mosca, and Priyanka Mukhopadhyay. “T-count and t-depth of any multi-qubit unitary”. *npj Quantum Information* **8**, 141 (2022).
- [57] Francesco Anna Mele, Antonio Anna Mele, Lennart Bittel, Jens Eisert, Vittorio Giovannetti, Ludovico Lami, Lorenzo Leone, and Salvatore FE Oliviero. “Learning quantum states of continuous variable systems” (2024). url: <https://arxiv.org/abs/2405.01431>.
- [58] Beatriz Dias and Robert König. “Classical simulation of non-gaussian bosonic circuits”. *Phys. Rev. A* **110**, 042402 (2024).
- [59] Ulysse Chabaud and Saeed Mehraban. “Holomorphic representation of quantum computations”. *Quantum* **6**, 831 (2022).
- [60] Martin Houde, Will McCutcheon, and Nicolás Quesada. “Matrix decompositions in quantum optics: Takagi/autonne, Bloch–Messiah/Euler, Iwasawa, and Williamson”. *Canadian Journal of Physics* **102**, 497–507 (2024).
- [61] Vittorio Giovannetti, Saikat Guha, Seth Lloyd, Lorenzo Maccone, and Jeffrey H. Shapiro. “Minimum output entropy of bosonic channels: A conjecture”. *Phys. Rev. A* **70**, 032315 (2004).
- [62] William Arveson. “Maximal vectors in Hilbert space and quantum entanglement”. *J. Funct. Anal.* **256**, 1476–1510 (2009).
- [63] Ulysse Chabaud, Ganaël Roeland, Mattia Walschaers, Frédéric Grosshans, Valentina Parigi, Damian Markham, and Nicolas Treps. “Certification of non-gaussian states with operational measurements”. *PRX Quantum* **2**, 020333 (2021).
- [64] Ulysse Chabaud, Ganaël Roeland, Mattia Walschaers, Frédéric Grosshans, Valentina Parigi, Damian Markham, and Nicolas Treps. “Erratum: Certification of non-gaussian states with operational measurements [PRX Quantum 2, 020333 (2021)]”. *PRX Quantum* **6**, 010902 (2025).