The recursive representation of Gaussian quantum mechanics

Yuan Yao,1,2,3, * Filippo Miatto,3,1,2,† and Nicolás Quesada4,‡

1 Télécom Paris, LTCI, 20 Place Marguerite Perny, 91120 Palaiseau, France
2 Institut Polytechnique de Paris, 5 Av. Le Chatelier, 91764 Palaiseau, France
3 Xanadu, 777 Bay St. M5G 2C8 Toronto, ON, Canada
4 Department of Engineering Physics, École Polytechnique de Montréal, Montréal, QC, H3T 1JK, Canada

(Dated: November 14, 2022)

We introduce a unified and differentiable Fock space representation of Gaussian objects, namely, pure and mixed states, unitaries and channels in terms of a single linear recurrence relation that can generate their Fock space amplitudes recursively. Due to its recursive and differentiable nature, it makes for a simple and fast computational implementation that enables calculating the gradient of the Fock amplitudes with respect to parametrizations. To demonstrate the flexibility and the generality of the gradient calculation, we show how to optimize M-mode Gaussian objects without the need to express them using fundamental components, by performing an optimization directly on the manifold of the symplectic group (or the unitary group for M-mode interferometers). We also find the composition rule of Gaussian operations expressed as the parameters of the recurrence relation, which allows us to obtain the correct global phase when composing Gaussian operations (which normally is absent from the description of Gaussian objects), and therefore to extend our model to states that can be written as linear combinations of Gaussians. We implemented all of these methods in the freely available open-source library MrMustard [1] and we show an example in which we find 3-mode circuits to generate cat states with unprecedented success probability and fidelity.

I. INTRODUCTION

Gaussian quantum mechanics is a subset of quantum mechanics that finds applications in several fields of quantum physics, such as quantum optics [2], quantum key distribution [3], optomechanical systems [4], quantum chemistry [5], condensed matter systems [6]. In the context of quantum optics, many of the available states (e.g. coherent, squeezed, thermal), transformations (e.g. beam splitter, squeezer, displacement, attenuator, amplifier), and measurements (e.g. homodyne, heterodyne) are Gaussian, i.e. characterized by a Gaussian phase space representation (be it the Wigner function, characteristic function, Husimi Q function, etc...). Gaussian objects are easy to manipulate, but in order to access a broader (in fact, universal) set of states and transformations, one needs to include non-Gaussian effects. One way to take into account non-Gaussian effects (e.g. photon-number-resolving detection), is to transform from the Gaussian phase space representation to the Fock space representation. Hence, studying the Fock space representation of Gaussian objects plays an important role in optical quantum simulation and optical quantum information processing [7–10].

The Fock space representation of Gaussian objects has been studied in different communities: in chemical physics, one studies vibronic transitions using the Hermite polynomials as a computational tool [11–14], and the matrix elements of unitary Gaussian and non-Gaussian transformations have been evaluated in [15] by using the multimode Bogoliubov transformation. In the mathematical physics context, these transformations correspond to the Bargmann-Fock representation of the symplectic group (also known as a metaplectic representation or oscillator representation), which we can understand as the Fock space representation of the group of Gaussian transformations [16].

In [17], we introduced a method to compute the Fock space amplitudes of Gaussian unitary transformations using a generating function. Part of the present work extends that method to cover Gaussian pure states, mixed states, and Gaussian channels as well. While many libraries exist to simulate quantum optical circuits [18–23], none of them so far have the dual features of fully exploiting the properties of Gaussian quantum mechanics and being differentiable. Thus, we implemented all of the methods and algorithms derived here in the open-source library MrMustard [1].

Besides an open-source library, we present three significant results, significantly extending what we did in [17]: (1) In Sec. III we introduce a unified, differentiable recursive representation of pure and mixed Gaussian states, Gaussian unitaries, and Gaussian channels in Fock space. (2) In Sec. IV we compute the global phase of the composition of Gaussian operations, which allows our method to be extended to states and transformations beyond the Gaussian ones as proposed in [24]. (3) In Sec. V we show how to perform a Riemannian optimization of M-modes Gaussian objects directly on the underlying symmetry group, bypassing the need to decompose them into some arrangement of fundamental elements and therefore allowing us to optimize a Gaussian quantum circuit as an entire block. We illustrate the utility of our methods and

* yuan.yao@telecom-paris.fr
† filippo@xanadu.ai
‡ nicolas.quesada@polymtl.ca
library by finding simple Gaussian circuits for the heralded preparation of cat states with mean photon number 4, fidelity 99.38%, and success probability 7.39%.

We will adopt the following notation conventions: The transposition and Hermitian conjugation operations are denoted as $^T$ and $^\dagger$. We use boldface for vectors $\mathbf{r}$ and matrices $\mathbf{S}$ but denote their components as $r_i$ and $S_{ij}$ respectively. We use $0_M$ for the $M \times M$ null matrix, $\mathbf{0}$ for a zero vector, and 0 for a scalar zero. $\mathbb{1}_M$ denotes for the $M \times M$ identity matrix.

Given a vector of integers $\mathbf{n} = (n_1, \ldots, n_M)$ we write $\mathbf{n}! = \prod_{i=1}^M n_i!$, $|\mathbf{n}\rangle = |n_1\rangle \otimes |n_2\rangle \otimes \cdots \otimes |n_M\rangle$ and given a complex or real vector $\alpha = (\alpha_1, \ldots, \alpha_M)^T$ we write $\alpha^n = \prod_{i=1}^M \alpha_i^n$ and $\partial^n = \prod_{i=1}^M \partial_i^n$. We write H.c. for Hermitian conjugate term.

**II. GAUSSIAN FORMALISM**

A. Commutation relations

Given an $M$-mode quantum continuous variables system, the field operators (i.e. annihilation and creation operators) $a_j, a_j^\dagger: j \in \{1, 2, \ldots, M\}$ satisfy the canonical commutation relation [25]:

$$[a_i, a_j^\dagger] = \delta_{ij}, \quad [a_i, a_j] = [a_i^\dagger, a_j^\dagger] = 0. \quad (1)$$

We can express these relations in a compact way by defining a vector of annihilation and creation operators $\mathbf{z} = (a_1, a_M, a_1^\dagger, a_M^\dagger)$, so that we can write

$$[z_i, z_j^\dagger] = Z_{ij}, \quad (2)$$

with

$$Z = \begin{pmatrix} \mathbb{1}_M & 0_M \\ 0_M & -\mathbb{1}_M \end{pmatrix}. \quad (3)$$

An alternative way to describe continuous-variable systems is obtained by defining the hermitian position $q$ and momentum $p$ operators:

$$q_j = \sqrt{\frac{\hbar}{2}}(a_j^\dagger + a_j), \quad p_j = i\sqrt{\frac{\hbar}{2}}(a_j^\dagger - a_j). \quad (4)$$

We can group these operators into a quadrature vector $\mathbf{r} = (q_1, \ldots, q_M, p_1, \ldots, p_M)$ so that $\mathbf{r}$ is related to $\mathbf{z}$ by the unitary matrix $\mathbf{W}$:

$$\mathbf{r} = \sqrt{\hbar} \mathbf{W} \mathbf{z}, \quad (5)$$

where

$$\mathbf{W} = \frac{1}{\sqrt{2}} \begin{pmatrix} \mathbb{1}_M & \mathbb{1}_M \\ -i\mathbb{1}_M & i\mathbb{1}_M \end{pmatrix}, \quad (6)$$

where $i = \sqrt{-1}$ is the imaginary unit.

Combining Eq. (2) and Eq. (5), we have:

$$[r_j, r_k^\dagger] = \hbar (\mathbf{W}^\dagger \mathbf{Z} \mathbf{W})_{jk} = i\hbar \Omega_{jk}, \quad (7)$$

where $\Omega$ is the skew-symmetric matrix:

$$\Omega = \begin{pmatrix} 0_M & \mathbb{1}_M \\ -\mathbb{1}_M & 0_M \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \otimes \mathbb{1}_M, \quad (8)$$

which is central to the description of the symplectic group (see section V). A brief summary of properties of the symplectic group can be found in Appendix A.

B. Gaussian states

A Gaussian state is any state whose characteristic functions and quasi-probability distributions are Gaussian functions in phase space [25]. Some well-known examples are coherent states, squeezed states, thermal states, and the vacuum state (which is the only state which is at the same time Gaussian and a number eigenstate).

The characteristic function of a state with density matrix $\rho$ is defined as:

$$\chi(\mathbf{s}; \rho) = \text{Tr}(\mathcal{D}_s \rho), \quad (9)$$

where $\mathcal{D}_s = \exp(-i s^T \Omega r/\hbar)$ is the Weyl, or displacement, operator and $\mathbf{s} \in \mathbb{R}^{2M}$ is a real vector in phase space.

For a Gaussian state we write the characteristic function in terms of its mean vector $\bar{\mathbf{r}}$ and covariance matrix $\mathbf{V}$ as [26]

$$\chi(\mathbf{s}; \rho) = \exp \left[ -\frac{1}{2} \mathbf{s}^T \Omega^T \mathbf{V} \mathbf{s} - i \bar{\mathbf{r}}^T \Omega \mathbf{s} \right], \quad (10)$$

where

$$\bar{r}_i = \langle r_i \rangle, \quad (11)$$

$$V_{ij} = \frac{1}{2} \langle r_i r_j + r_j r_i \rangle - \bar{r}_i \bar{r}_j, \quad (12)$$

Note that the covariance matrix $\mathbf{V}$ is a real, symmetric, positive definite matrix.

If we use the covariance matrix $\mathbf{V}$, we find the mean vector $\bar{\mathbf{r}}$ and the covariance matrix $\mathbf{\sigma}$:

$$\bar{\mu}_i = \langle z_i \rangle = \frac{1}{\sqrt{\hbar}} (\mathbf{W}^\dagger \mathbf{r})_i, \quad (13)$$

$$\sigma_{ij} = \frac{1}{2} \langle z_i z_j^\dagger + z_j z_i^\dagger \rangle - \bar{\mu}_i \bar{\mu}_j = \frac{1}{\hbar} (\mathbf{W}^\dagger \mathbf{W})_{ij}. \quad (14)$$

Compared with the real covariance matrix $\mathbf{V}$, we denote the $\mathbf{\sigma}$ as the complex covariance matrix.

In the remainder of this paper we will write the phase space description of a Gaussian state as the pair $(\mathbf{V}, \bar{\mathbf{r}})$ or $(\mathbf{\sigma}, \bar{\mathbf{\mu}})$ depending on which basis we use. For example, the vacuum state $|0\rangle$, which satisfies $a_j|0\rangle = 0$, has a zero mean vector and covariance matrix $\mathbf{V} = \frac{\hbar}{2} \mathbb{1}$ or $\mathbf{\sigma} = \frac{\hbar}{2} \mathbb{1}$.
C. Gaussian transformations

Gaussian unitary transformations are those that map Gaussian states to Gaussian states [26], thus in the Schrödinger picture, an input Gaussian state $\rho$ is mapped to an output Gaussian state

$$\rho \mapsto \rho' = U_G \rho U_G^\dagger.$$  \hspace{1cm} (15)

Gaussian unitaries have as generators polynomials of at most degree 2 in the quadratures (or equivalently in the creation and annihilation operators).

In the Heisenberg picture a Gaussian unitary (parameterized by a $2M \times 2M$ matrix $S$ and a real vector $d$ with size $2M$) transforms the quadrature operators as follows

$$r \mapsto r' = U_G^\dagger r U_G = Sr + d.$$  \hspace{1cm} (16)

Since $r'$ is obtained from $r$ by unitary conjugation, it must satisfy the canonical commutation relations in Eq. (7). This implies that the matrix $S$ satisfies

$$S \Omega S^T = \Omega,$$  \hspace{1cm} (17)

that is, $S$ must be an element of the (real) symplectic group, $S \in \text{Sp}(2M, \mathbb{R})$.

A $M$-mode Gaussian unitary generated by a second-degree polynomial in the quadratures can be decomposed into a $M$-mode displacement $D_d$ and a $M$-mode unitary generated by a strictly quadratic unitary that is responsible for the symplectic matrix appearing in Eq. (16) and thus can be written

$$U_G = D_d U(S),$$  \hspace{1cm} (18)

where $D_d$ is the displacement operator, parameterized by a real vector $d$ of size $2M$. We can also express the $M$-mode displacement operator as the tensor product of the single-mode displacement operator, with a complex vector $\gamma$ of size $M$. The relation between the vector $d$ and $\gamma$ can be derived from Eq. (4):

$$d = \sqrt{2\hbar} [\Re(\gamma), \Im(\gamma)].$$  \hspace{1cm} (19)

The single-mode displacement operator is defined as

$$D(\gamma) = \exp [\gamma a^\dagger - \gamma^* a].$$  \hspace{1cm} (20)

We will also give the definitions of other single-mode Gaussian unitaries, noting that the multi-mode version is just the tensor product extension of their single-mode version.

The single-mode rotation operator

$$R(\phi) = \exp [i\phi a^\dagger a],$$  \hspace{1cm} (21)

which has $d_{\text{rot}} = 0$ and

$$S_{\text{rot}} = \begin{bmatrix} \cos \phi & -\sin \phi \\ \sin \phi & \cos \phi \end{bmatrix}.$$  \hspace{1cm} (22)

The single-mode squeezing operator is defined as

$$S(\zeta) = \exp \left[ \frac{i}{2} \zeta^* a^2 - \text{H.c.} \right],$$  \hspace{1cm} (23)

where $\zeta = re^{i\delta}$, and it has $d_{\text{sq}} = 0$ and

$$S_{\text{sq}} = S_{\text{rot}}(\delta/2) \begin{bmatrix} e^{-r} & 0 \\ 0 & e^r \end{bmatrix} S_{\text{rot}}(\delta/2)^T.$$  \hspace{1cm} (24)

An $M$-mode interferometer with Hilbert space operator [28]

$$\mathcal{W}(J) = \exp \left[ i \sum_{k,l=1}^M J_{kl} a_k^\dagger a_l \right],$$  \hspace{1cm} (25)

which has $d_{\text{intf}} = 0$, and

$$S_{\text{intf}} = \begin{bmatrix} \Re(U) & -\Im(U) \\ \Im(U) & \Re(U) \end{bmatrix}. $$  \hspace{1cm} (26)

where $U = \exp [iJ]$ is a unitary matrix (since $J = J^\dagger$).

Note that $S_{\text{intf}} \in \text{Sp}(2n, \mathbb{R}) \cap O(2n) \cong U(n)$ where $\text{Sp}(2n, \mathbb{R})$ is the symplectic group, $O(2n)$ is the orthogonal group and $U(n)$ is the unitary group.

A particular instance of an interferometer is the one beamsplitter, parametrized in terms of transmission angle $\theta$ and a phase $\phi$ (the energy transmission is given by $\cos^2 \theta$). In this case we have

$$J = i \begin{pmatrix} 0 & \theta e^{-i\phi} \\ -\theta e^{i\phi} & 0 \end{pmatrix}, \quad U = \begin{pmatrix} \cos \theta & -e^{-i\phi} \sin \theta \\ e^{i\phi} \sin \theta & \cos \theta \end{pmatrix}. $$  \hspace{1cm} (27)

Note that our definition of interferometer immediately implies that $\mathcal{W}(J)|0\rangle = |0\rangle$ without any ambiguity in the global phase of the state on the right hand side.

Gaussian unitaries transform the mean vector $\bar{r}$ and the covariance matrix $V$ of a Gaussian state as:

$$(V, \bar{r}) \mapsto (V', \bar{r}') = (SVS^T, S\bar{r} + d).$$  \hspace{1cm} (28)

A deterministic Gaussian channel is the most general trace-preserving map between Gaussian states. It is characterized by two matrices $X, Y$ and a vector $d$ [26]. The action of the channel on a Gaussian state $(V, \bar{r})$ is

$$(V, \bar{r}) \mapsto (XVX^T + Y, X\bar{r} + d).$$  \hspace{1cm} (29)

where the matrices $X$ and $Y$ need to satisfy

$$Y + i\frac{\hbar}{2} \Omega \geq i\frac{\hbar}{2} XX^T.$$  \hspace{1cm} (30)

More generally, the action of a Gaussian channel on the characteristic function of an arbitrary state amounts to

$$\chi(s) \mapsto \chi'(s) = \chi(\Omega^T X^T \Omega s) \exp \left( -\frac{1}{2} s^T \Omega^T Y \Omega s - id^T \Omega s \right).$$  \hspace{1cm} (31)
Note that unitary channels such as Eq. (28) are special cases of a Gaussian channels where \( Y = 0_{2M} \) and \( X \) is symplectic. More generally, when \( X \) is not symplectic and thus the channel is not unitary, the matrix \( Y \) represents added noise in the state.

Examples of single-mode Gaussian channels are the pure loss channel (defined in Eq. (5.77) in the book [26]) with energy gain and the amplification channel (defined in Eq. (5.87) in the book [26]) with energy transmission \( 0 \leq \eta \leq 1 \), which has

\[
X = \sqrt{\eta} \mathbb{1}_2, \quad Y = \frac{\hbar}{2} (1 - \eta) \mathbb{1}_2, \quad d = 0,
\]

and the amplification channel (defined in Eq. (5.87) in the book [26]) with energy gain \( g \geq 1 \), which has

\[
X = \sqrt{g} \mathbb{1}_2, \quad Y = \frac{\hbar}{2} (g - 1) \mathbb{1}_2, \quad d = 0.
\]

An example of a multi-mode Gaussian channel is the lossy interferometer parametrized in terms of a transmission matrix \( T \) with singular values bounded from above by 1. For this channel, we find

\[
X = \begin{bmatrix} \Re(T) & -\Im(T) \\ \Im(T) & \Re(T) \end{bmatrix},
\]

\[
Y = \frac{\hbar}{2} (\mathbb{1}_{2M} - XX^T),
\]

\[
d = 0.
\]

Note that in the case where \( T \) is unitary, then \( X \) is symplectic and orthogonal, and thus \( Y = 0_{2M} \) recovering the results from the previous subsection.

### III. ONE RECURRENCE RELATION TO RULE THEM ALL

We can write \( M \)-mode pure states, mixed states, unitaries, and channels in the Fock space representation as

\[
|\psi\rangle = \sum_k \psi_k |k\rangle,
\]

\[
\rho = \sum_{j,k} \rho_{j,k} |j\rangle\langle k|,
\]

\[
U = \sum_{j,k} U_{j,k} |j\rangle\langle k|,
\]

\[
\Phi(|j\rangle\langle i|) = \sum_{k,l} \Phi_{k,l,i,j} |i\rangle\langle k|,
\]

where the Fock space indices are expressed as a multi-index \( k = (k_1, k_2, \ldots, k_M) \). We now simplify the notation by considering the collections of amplitudes \( \psi_k, \rho_{j,k}, U_{j,k} \) and \( \Phi_{i,j,k,l} \) as instances of a tensor \( \mathcal{G}_k \) where \( k \) is \( M \)-dimensional for pure states, \( 2M \)-dimensional for mixed states and unitary transformations, and \( 4M \)-dimensional for channels.

One way to produce the Fock space amplitudes of a Gaussian object is to start from a generating function \( \Gamma(\alpha) \) and then compute its derivatives. The generating function \( \Gamma(\alpha) \) is also known as the stellar function [29] or the Bargmann function [16]. To obtain the generating function, one needs to contract each index of a Gaussian object with a rescaled multi-mode coherent state

\[
e^{\frac{1}{2}||\alpha||^2} |\alpha\rangle.
\]

For example, for a pure state, we have

\[
\Gamma_\psi(\alpha) = e^{\frac{1}{2}||\alpha||^2} \sum_k \psi_k |\alpha^*\rangle |k\rangle = \sum_k \psi_k \frac{\alpha^k}{\sqrt{k!}}
\]

\[
= \psi_0 \exp \left( \alpha^T b_\psi + \frac{1}{2} \alpha^T A_\psi \alpha \right),
\]

where \( A_\psi \) is an \( M \times M \) complex symmetric matrix, \( b_\psi \) is an \( M \)-dimensional complex vector and \( \psi_0 \) is the vacuum amplitude.

In the case of density matrices, we obtain an analogous exponential as in (42), except that \( A_\rho \) are of size \( 2M \times 2M \) and \( 2M \times 2M \) respectively. For unitaries, \( A_U \) and \( b_U \) are of size \( 2M \times 2M \) and \( 2M \), and for channels \( A_\Phi \) and \( b_\Phi \) are of size \( 4M \times 4M \) and \( 4M \), respectively. Therefore, all Gaussian objects are characterized by a complex symmetric matrix \( A \), a complex vector \( b \) and a complex scalar \( c = G_0 \), or conversely given valid \( A \) and \( b \) and \( c \) we can calculate the coefficients \( G_k \) by computing derivatives of the appropriate order of the generating function \( \Gamma(\alpha) \):

\[
G_k = G_0 \frac{\partial_k^k}{\sqrt{k!}} \exp \left( \alpha^T b + \frac{1}{2} \alpha^T A \alpha \right) \bigg|_{\alpha=0}.
\]

In this way we unify the calculation of the Fock space amplitudes of Gaussian objects into a single method that works in all cases, depending on which triple \((A, b, c)\) one is considering.

In practice (as we will do in the following sections), it is sufficient to apply this method to the case of mixed states only, as the expressions split naturally thanks to the properties of the Hermite polynomials (see Eq. (61) to (65)), and one obtains the case of Gaussian pure states. For transformations, using the Choi-Jamiołkowski duality we can treat channels as mixed states, and if a channel is unitary, the expressions split in the same way as they do for states (see Eq. (86) to (90)), and one obtains the case of Gaussian unitaries.

Multivariate derivatives of the exponential of a function can be computed with a linear recurrence formula [30], and in the case the function is a polynomial of degree \( D \), the recurrence relation has order \( D \). In our case, the polynomial has degree 2, which means we can write a linear recurrence relation of order 2 between the Fock space amplitudes:

\[
G_{k+1} = \frac{1}{\sqrt{k+1}!} \left( b_k G_k + \sum_j \sqrt{k+1} A_{ij} G_{k-1,j} \right),
\]

with the vacuum amplitude initialized as \( G_0 = c \). In this recurrence relation, \( k + 1 \) is like \( k \) but the \( i \)-th index has
been increased by 1 (and similarly for \( k - 1, j \), where it is decreased by 1). We refer to \( w = \sum k_i \) as the weight of the index. In essence, the recurrence relation allows us to write amplitudes of weight \( w + 1 \) as linear combinations of amplitudes of weight \( w \) and \( w - 1 \). By applying it repeatedly, one can reach any Fock space amplitude (in practice, one eventually reaches a numerical precision horizon [31]).

A. Multidimensional Hermite Polynomials

Before studying the mapping between phase space and Fock space, we recall the definition of the multidimensional Hermite polynomials as the Taylor series of a multidimensional Gaussian function

\[
K^A(y, b) = \exp \left( y^T b + \frac{1}{2} y^T A y \right) = \sum_{k \geq 0} \frac{G_k^A(b)}{k!} y^k.
\]

(46)

Note the sign of the quadratic term in the exponential, which can differ from other conventions. In the last equation \( b \in \mathbb{C}^\ell \) is a complex vector, \( A = A^T \in \mathbb{C}^{\ell \times \ell} \) is a complex symmetric matrix and \( k \in \mathbb{Z}_{\geq 0}^\ell \) is a vector of non-negative integers. This notation makes it explicit that

\[
\left[ \prod_{i=1}^{\ell} \left( \frac{\partial}{\partial y_i} \right)^{k_i} \right] K^A(y, b) \bigg|_{y=0} = G_k^A(b).
\]

(47)

These polynomials satisfy the recurrence relation

\[
G_{k+1,i}(b) - b_i G_k^A(b) - \sum_{j=1}^{M} A_{i,j} p_j G_{k-1,j}(b) = 0,
\]

(48)

where \( 1_i \) is a vector that has a 1 in the \( i \)-th entry and 0s elsewhere. Note that \( G_0^A(b) = 1 \), \( G_1^A(b) = b_i \) and that \( G_{1,i}^A(b) = b_i b_j + A_{i,j} \). The multidimensional Hermite polynomial is related to the loop-hafnian function introduced in Ref. [32] which counts the number of perfect matchings of weighted graphs, including self-loops. They are related as follows

\[
G_k^A(b) = \text{lhaf}(\text{diag}(A_k, b_k)),
\]

(49)

where \( \text{diag} \) fills the diagonal of the matrix in the first argument using the vector in the second argument. Note that \( A_k \) is the matrix obtained from \( A \) by repeating its \( i \)-th row and column \( k_i \) times. Similarly, \( b_k \) is the vector obtained from \( b \) by repeating its \( i \)-th entry \( k_i \) times. Note that when \( k_i = 0 \) the relevant row and column of \( A \) and entry of \( b \) are deleted. The best known methods to calculate the single loop-hafnian in Eq. (49) requires \( O(C^3 \sqrt{\prod_{i=1}^{\ell} (1 + k_i)}) \) steps where \( C \) is the number of nonzero entries in the vector \( k \) [33].

We will show below that the Fock representation of a pure Gaussian state, a mixed Gaussian state, a Gaussian unitary, or a Gaussian channel can all be written as

\[
e \times \frac{G_k^A(b)}{\sqrt{k!}},
\]

(50)

where \( e \) is a scalar, \( b \) is a vector of dimension \( \ell \), \( A \) is a square matrix of size \( \ell \times \ell \) and \( k \in \mathbb{Z}_{\geq 0}^\ell \). The integer \( \ell \) equals \( M, 2M, 2M, 4M \) for pure states, mixed states, unitaries or channels on \( M \) modes respectively.

Note that the quantity in Eq. (50) is potentially the ratio of two large numbers. In particular, since this quantity represents a probability or a probability amplitude it should be bounded in absolute value by 1. Thus it is often convenient, especially for numerical purposes, to introduce renormalized multidimensional Hermite polynomials as

\[
G_k^A(b) = \sum_{i=1}^{\ell} \sqrt{k_i} \delta_{ij} G_{k-1,j}^A(b).
\]

(51)

which satisfy the recurrence relation in Eq. (45).

Using results from Ref. [17] we can also find the differential of the matrix elements:

\[
dG_k^A(b) = \frac{[dc]}{c} G_k^A(b) + \sum_{i=1}^{\ell} \sqrt{k_i} \delta_{ij} G_{k-1,j}^A(b)
\]

(52)

\[
+ \frac{1}{2} \sum_{i,j=1}^{\ell} [dA_{i,j}] \sqrt{k_i(k_j - \delta_{ij}) G_{k-1-i,j}^A(b)}.
\]

We can use this relation to write a new differential formula for the loop-hafnian with arbitrary repetitions that generalizes the results in Ref. [34]

\[
d[\text{lhaf}(\text{diag}(A_k, b_k))] =
\]

(53)

\[
\sum_{i=1}^{\ell} [db_i] k_i \text{lhaf}(\text{diag}(A_{k-1,i}, b_{k-1,i}))
\]

\[
+ \frac{1}{2} \sum_{i,j=1}^{\ell} [dA_{i,j}] k_i(k_j - \delta_{ij}) \text{lhaf}(\text{diag}(A_{k-1-i,j}, b_{k-1-i,j}))
\]

Note that in the limit of no loops \( b = \mathbf{0} \) and no repetitions \( k_i \in \{0, 1\} \) the last equation reproduces precisely Eq. (A12) of Ref. [34].

B. States

In this subsection, we show how to turn the symplectic representation of a Gaussian state into the metaplectic or Fock space representation of the same object [16]. This follows the developments in Refs. [7–10, 35, 36].

To compute the Fock space amplitudes of a Gaussian pure state we need the triple \((A_\psi, b_\psi, c_\psi)\) where \(A_\psi\) and \(b_\psi\) are \(M\)-dimensional. If the state is mixed, we need the
triple \((A_p, b_p, c_p)\) where \(A_p\) and \(b_p\) are 2\(M\)-dimensional. We are now going to show how to obtain these triples.

It is convenient to introduce the \(s\)-parametrized complex covariance matrix

\[
\sigma_s = \sigma + \frac{s}{2}\sigma_{2M},
\]

by definition \(\sigma_0 \equiv \sigma\) and moreover we use the shorthand notation \(\sigma_\pm \equiv \sigma_{\pm 1}\).

We recall the results derived in Ref. [9]. An expression for the metaplectic representation of the Gaussian state is

\[
\langle m|\rho|n\rangle = c_p \times \prod_{s=1}^{M} \frac{\partial_{a_s} \partial_{a_s}^*}{\sqrt{n_s! m_s!}} \exp \left[ \frac{i}{2} y^T A_p y + y^T b_p \right],
\]

where, relative to Eq. (47), we identified \(y = [\sigma^s]^{-1}, k = n \otimes m, \ell = 2M\) and used the results from Refs. [7, 36, 37] to write together with the definitions in Eqs. (13) (14)

\[
A_p = P_M [I - \sigma_+^{-1}] = P_M \sigma_- \sigma_+^{-1} = P_M \sigma_{-1} \sigma_-, \quad (56)
\]

\[
b_p = (\sigma_+^{-1} \mu) = P_M \sigma_- \mu, \quad (57)
\]

\[
c_p = (0|\rho|=0) = \exp \left[ -\frac{i}{2} \mu_+ \sigma_+^{-1} \mu_+ \right] \frac{1}{\sqrt{\det(\sigma_+)}}, \quad (58)
\]

\[
P_M = \left[ \frac{I_{2M}^*}{I_{2M}} \right], \quad (59)
\]

to finally write

\[
\langle m|\rho|n\rangle = c_p \times \frac{G_{\sigma_+ n \otimes m} (b_p)}{\sqrt{n! m!}}. \quad (60)
\]

The map \(\sigma \rightarrow \sigma_{-1} \sigma_-\) in Eq. (56) is the Cayley transform [38, 39]. In the case where \(\rho = |\Psi\rangle\langle\Psi|\) is a pure state it is easy to show that

\[
A_p = A_p^* \oplus A_p, \quad (61)
\]

\[
b_p = b_p^* \oplus b_p, \quad (62)
\]

and then we can write

\[
G_{\sigma_+ n \otimes m} (b_p) = G_{\sigma^s n \otimes m} (A_p \oplus A_p)\]

\[
= G_{\sigma^s n \otimes m} (b_p^* \oplus b_p)\]

\[
= [G_{\sigma^s n \otimes m} (b_p)]^* \otimes G_{\sigma^s m \otimes n} (b_p), \quad (63)
\]

which allows us to write the probability amplitude of a pure state

\[
\langle m|\Psi\rangle = c_p \frac{G_{\sigma_+ n \otimes m} (b_p)}{\sqrt{m!}}, \quad c_p = e^{i\varphi_p} \sqrt{c_p},
\]

up to a global phase \(\varphi\) that cannot be determined from the covariance matrix and vector of means of the pure Gaussian state. This will be discussed in a later section.

Note that the last equation can be used to write the Hilbert-space ket representing the state as [8]

\[
|\Psi\rangle = c_p \exp \left[ \frac{M}{2} (b_p \sigma_1^1 + \frac{1}{2} \sum_{i,j=1}^M (A_p)_{ij} a_i^a a_j^\dagger) \right]|0\rangle, \quad (67)
\]

thus showing that this formalism reduces to the one introduced by Krenn et al. in Refs. [40–43] when the displacements are zero. Moreover, the Gaussian formulation allows us to easily include the most common form of decoherence for bosonic modes, namely loss, since this process is a Gaussian channel.

We now give a few examples. The recursive representation of a single-mode coherent state of amplitude \(\alpha\) is given by \(A_p = 0, b_p = \alpha\) and \(c_p = \psi_0 = e^{-\frac{1}{2}|\alpha|^2}\):

\[
\psi_{k+1}^{coh} = \frac{1}{\sqrt{k+1}} \alpha \psi_k^{coh}. \quad (68)
\]

A squeezed state with squeezing parameter \(r\) and angle \(\phi\) is given by \(A_p = \tanh(r)e^{i\phi}, b_p = 0\) and \(c_p = \psi_0 = \sqrt{\text{sech}(r)}\):

\[
\psi_{k+1}^{sq} = \frac{1}{\sqrt{k+1}} \tanh(r)e^{i\phi} \psi_{k-1}^{sq}. \quad (69)
\]

Note that, as expected, this recurrence relation skips odd indices. A displaced squeezed state (which is the most general pure single-mode Gaussian state) is given by \(A_p = \tanh(r)e^{i\phi}, b_p = \alpha\) and \(c_p = \sqrt{\text{sech}(r)}e^{-\frac{1}{2}|\alpha|^2}\):

\[
\psi_{k+1}^{dsq} = \frac{1}{\sqrt{k+1}} (\alpha \psi_k^{dsq} + \sqrt{k} \tanh(r)e^{i\phi} \psi_{k-1}^{dsq}). \quad (70)
\]

For the simple case of \(M\)-mode squeezed states with parameters \(r_i\) sent into an interferometer with unitary \(U\) we have that \(A_p = -U [\otimes_{l=1}^M \tanh r_l] U^T\).

The thermal state is given by \(A_p = \frac{n}{\pi^\frac{1}{2}}, b_p = 0\) and \(c_p = \frac{1}{\sqrt{\text{coth}(n)}}\), where \(\bar{n}\) is the average photon number, giving rise to the recurrence relations:

\[
\rho_{k+1, k+2}^{th} = \sqrt{\frac{k_2}{k_1 + 1}} \frac{\bar{n}}{\bar{n} + 1} \rho_{k+1, k}^{th}, \quad (71)
\]

\[
\rho_{k+1, k+2}^{th} = \sqrt{\frac{k_1}{k_2 + 1}} \frac{\bar{n}}{\bar{n} + 1} \rho_{k+1, k}^{th}. \quad (72)
\]

For a squeezed state along the \(q\)-quadrature with \(r > 0\) (the symplectic matrix \(S\) can be found in Eq. (24)) that undergoes loss by transmission \(\eta\) (defined in Eq. (32)), we start from the vacuum state with \(V = \frac{1}{\sqrt{2}}\), we apply the squeezing operator \(V' = SVS^T\), we make the pass through the lossy channel \(V'' = XV'X'^T + Y\), and we obtain its covariance matrix \(\sigma = \frac{1}{2} WV'V'W\). Then it is easy to find \(A_p\) from Eq. (56) that

\[
A_p = \frac{\eta}{\text{coth}^2 r - (\eta - 1)^2} \left[ -\text{coth} r 1 - \eta - \text{coth} r \right]. \quad (73)
\]

In the limit of no loss we find \(A_p = -[\tanh r \oplus \tanh r]\) while in the limit of zero transmission we retrieve the single-mode vacuum, \(A_p = 0\).
C. Transformations

We can lift the description of states in the previous section to describe transformations via the Choi-Jamiolkowski isomorphism in phase space, which allows us to faithfully map a channel by applying it over one-half of a full-rank entangled state. A Gaussian channel $\Phi[\cdot]$ is uniquely determined by the triplet $X, Y, d$ and acts on a Gaussian state as $(V, r) \mapsto ( XVX^T + Y, Xr + d )$. We can then write (see Appendix C and Appendix D for details)

$$
\langle i | \Phi [ j ] \langle l | \rangle | k \rangle = c_k \times \frac{G_{A^\phi} b_k}{\sqrt{2^j j!}}.
$$

(74)

where

$$
A = P_{2M} R \left[ \begin{array}{ccc}
\mathbb{1}_{2M} & -\xi^{-1}X & \xi^{-1}X \\
X^T & \mathbb{1}_{2M} & -X^T \xi^{-1}X
\end{array} \right] R^\dagger,
$$

(75)

$$
b = \frac{1}{\sqrt{h}} R^\dagger \left[ \begin{array}{c}
\mathbb{1}_{2M} \xi^{-1}d \\
-\xi^{-1}X \\
-X^T \xi^{-1}d
\end{array} \right],
$$

(76)

and

$$
R = \frac{1}{\sqrt{2^j j!}} \left[ \begin{array}{ccc}
1 & 1 & 1 \\
1 & 0 & 1 \\
0 & 1 & 0
\end{array} \right].
$$

(78)

For example, for a single-mode amplifier channel with gain $g \geq 1$, we find

$$
A = \left[ \begin{array}{ccc}
0 & \frac{1}{\sqrt{g}} & 0 \\
\frac{1}{\sqrt{g}} & 0 & 0 \\
0 & 0 & \frac{1}{\sqrt{g}}
\end{array} \right],
$$

(80)

For the case of the $M$-mode lossy interferometer with transmission matrix $T$ we find

$$
A = \left[ \begin{array}{ccc}
0_M & T & 0_M \\
T^\dagger & 0_M & 0_M \\
0_M & 0_M & 0_M \\
0_M & 0_M & 0_M
\end{array} \right],
$$

(81)

$$
b = 0,
$$

(82)

$$
c = 1.
$$

(83)

This identity allows us to find the probability of measuring an outcome photon number pattern $j = (j_1, \ldots, j_M)$ when the multimode Fock state $|i\rangle = |i_1\rangle \otimes \cdots \otimes |i_M\rangle$ is sent into a lossy interferometer with transmission matrix $T$ (cf. Appendix F)

$$
\langle j | \Phi_T [ i ] \langle i | \rangle | j \rangle = \frac{1}{i^j j!} \text{perm} \left[ \begin{array}{ccc}
\mathbb{1}_M - T^\dagger T & T^\dagger \\
T & 0
\end{array} \right],
$$

(84)

where perm is the permanent. The last equation reduces to the well-known lossless [44–46] case when $\mathbb{1}_M - T^\dagger T = 0$. Moreover, this equation can be used to obtain marginals of the Fock state Boson Sampling distribution, since a lossy interferometer can always be understood as part of a larger unitary interferometer.

In the case where the channel is unitary, we can write $\Phi[\cdot] = U [ \cdot ] U^\dagger$ and then we obtain

$$
\langle i | \Phi [ j ] \langle l | \rangle | k \rangle = \langle iU | j \langle l | U^\dagger | k \rangle.
$$

(85)

This corresponds to the case where $Y = 0_{2M}$ and $X = S$ is symplectic. As we show in the Appendix E, we can then write

$$
A_U = A_U^* \oplus A_U,
$$

(86)

$$
b_U = b_U^* \oplus b_U,
$$

(87)

and then we have

$$
\langle i | \Phi [ j ] \langle l | \rangle | k \rangle = G_{k \oplus l}^U (b_U \oplus b_U^*)
$$

(88)

$$
= G_{k \oplus l}^U (b_U) \times G_{k \oplus l}^U (b_U^*)
$$

(89)

$$
= \left[ G_{k \oplus l}^U (b_U^*) \right]^* \times G_{k \oplus l}^U (b_U).
$$

(90)

Comparing Eq. (85) and the last equation we easily identify

$$
\langle i U | j \rangle = c_{U} \times G_{k \oplus l}^U (b_U) \sqrt{2^j j!}, \quad c_{U} = \sqrt{c_{d} \exp i \varphi_{U}},
$$

(91)

where $\varphi_{U}$ is a phase that will be discussed in the next section. Note that the quantities $c_{U}, b_{U}^*$ and $A_{U}$ correspond to the $C, \mu, - \Sigma$ introduced in Eq. (26) of Ref. [17]. This comparison also allows us to conclude that $A_{U}$ is not only symmetric but also unitary (this can also be seen by inspecting the form of $A_{U}$ in Eq. (E10) in the Appendix).

IV. GLOBAL PHASE OF THE FOCK REPRESENTATION

In the Gaussian representation, transformations are specified by a symplectic matrix and a displacement vector. However, these two quantities do not uniquely specify the evolution of a quantum state. For example, when two displacement operators with parameters $d_1$ and $d_2$ are composed in the Gaussian representation, their effect is just another displacement with parameter $d = d_1 + d_2$. However, the unitary representation acquires a global phase:

$$
\mathcal{D}(\alpha)\mathcal{D}(\beta) = e^{i (\alpha \beta^* - \alpha^* \beta)/2} \mathcal{D}(\alpha + \beta).
$$

(92)
i.e., we do not only add up both displacement parameters $D(\alpha+\beta)$ here, but also get an extra part $e^{(\alpha^*\beta-\alpha^*\beta)/2}$, which is a \textit{global phase}. Such a global phase is important when evolving linear combinations of Gaussian states with Gaussian operations \cite{24}. This section will compute this global phase and provide some examples.

We know that the Fock representation of an arbitrary Gaussian unitary transformation is parametrized by the triple $(A_U, b_U, c_U)$. The unitary representation of the combination of two Gaussian transformations may have an additional global phase:

$$G_1(A_{U_1}, b_{U_1}, c_{U_1})G_2(A_{U_2}, b_{U_2}, c_{U_2}) = G_f(A_{U_f}, b_{U_f}, c_{U_f})$$

(93)

where $U_f = U_1U_2$. The question now is how to find the phase of $G_f$ as a function of $A_{U_1}, b_{U_1}, c_{U_1}$ and $A_{U_2}, b_{U_2}, c_{U_2}$.

We begin by calculating the Husimi $Q(\beta, \beta')$ function of the composition of $G_1$ and $G_2$ and we use a resolution of the identity in terms of coherent states to write:

$$\langle \beta^* | G_1G_2 | \beta' \rangle = \frac{1}{\pi^M} \int_{-\infty}^{\infty} d^2M \alpha \langle \beta^* | G_1 | \alpha \rangle \langle \alpha | G_2 | \beta' \rangle,$$

(94)

(95)

where we can replace the Husimi $Q$ functions for generic Gaussian transformations $\langle \beta^* | G_1 | \alpha \rangle$ and $\langle \alpha | G_2 | \beta' \rangle$. After integrating $\alpha$, the $Q$ function of the composite operator $G_f$ is obtained, which is characterized by:

$$A_{U_f} = \begin{bmatrix} B_1 - C_1D_1X^{-1}C_1^T & -C_1X^{-1}C_2 \\ -C_2^T \left( X^T \right)^{-1}C_1^T & D_2 - C_2^T D_1 \left( X^T \right)^{-1} C_2 \end{bmatrix},$$

(96)

$$b_{U_f} = \begin{bmatrix} c_1 - C_1D_1X^{-1}d_1 - C_1 \left( X^T \right)^{-1} c_2 \\ d_2 - C_2^T \left( X^T \right)^{-1} d_1 - C_2^T D_1 \left( X^T \right)^{-1} c_2 \end{bmatrix},$$

(97)

$$c_{U_f} = \frac{c_{U_1}c_{U_2}}{\sqrt{-1)^M \det(D_1B_2 - 1)}} e^{-\frac{1}{2} d_1^T D_1X^{-1}d_1 + c_2^T \left( X^T \right)^{-1} d_1} d_2^T X^{-1} c_2 + c_2^T D_1 \left( X^T \right)^{-1} c_2,$$

(98)

where $b_{U_f}^T$ and $A_{U_f}$ are written in block form:

$$b_{U_f}^T = [c_1^T, d_1^T],$$

(99)

$$A_{U_f} = \begin{bmatrix} B_1 & C_1 \\ C_1^T & D_1 \end{bmatrix},$$

(100)

and we define a new matrix $X$:

$$X = B_2D_1 - 1.$$  

(101)

Eq. (98) gives the global phase for the composite Gaussian operator. The details of this calculation can be found in Appendix G.

As examples, we show the composition of two single-mode displacements and the composition of two single-mode squeezers.

For displacement operators $D(\alpha), D(\beta)$, we have

$$\det(D_1B_2 - 1) = 1, \quad \chi = -1. \quad (102)$$

We then obtain the global phase:

$$\alpha_{U_f} = \alpha_{U_1}\alpha_{U_2} \exp(-\alpha^*\beta) = \alpha_{U(\alpha+\beta)} \exp\left(\frac{1}{2} \alpha^*\beta - \frac{1}{2} \beta^*\alpha^*\right), \quad (103)$$

recovering Eq. (92).

For two squeezers $S(\zeta_1), S(\zeta_2)$, since $b_U$ is zero, we have

$$\det(D_1B_2 - 1) = 1 - e^{i(\delta_2-\delta_1)} \tanh r_1 \tanh r_2, \quad (104)$$

and in turn, we get

$$\alpha_{U_f} = \frac{c_{U_1}c_{U_2}}{\sqrt{-1)^M \det(D_1B_2 - 1)}} \frac{1 + e^{i(\delta_2-\delta_1)} \tanh r_1 \tanh r_2}{\sqrt{\text{sech} r_1 \text{sech} r_2}}, \quad (105)$$

which coincides with the results from Refs \cite{47, 48}.

Finally, note that when composing two passive Gaussian unitaries we already know that there is no extra phase since by construction (cf. Eq. (25)) $\langle 0 | W(J) | 0 \rangle = 1$.

V. LEARNING GAUSSIAN STATES AND TRANSFORMATIONS

Differentiability is a desirable property for a computational model, as it enables gradient descent optimization. This section summarizes the basic ideas of gradient descent on Riemannian manifolds, particularly on the manifold of symplectic matrices.

Note that in the first four subsections below, the symbols $A, B, M, p, R, W, X, Y, Z, \gamma$ are defined locally and do not correspond to previous uses.

A. The symplectic group

We describe the manifold of real symplectic $2n \times 2n$ matrices as an embedded sub-manifold of $\mathbb{R}^{2n \times 2n}$:

$$\text{Sp}(2n, \mathbb{R}) = \{ S \in \mathbb{R}^{2n \times 2n} | S \Omega S^T = \Omega \}, \quad (107)$$

where $\Omega$ is defined in Eq. (8). Given that the condition $S \Omega S^T = \Omega$ is quadratic in $S$, the manifold of symplectic matrices is not a linear subspace of $\mathbb{R}^{2n \times 2n}$, which means that we likely leave the manifold after a naive straight step of gradient descent. In this section, we explain how to overcome this difficulty.

Note that unless details are relevant, we abbreviate $\text{Sp}(2n, \mathbb{R})$ with $\text{Sp}$. 
B. Tangent and Normal spaces

If we differentiate the quadratic condition $S\Omega S^T = \Omega$ we obtain the linear tangency condition $X\Omega S^T + S\Omega X^T = 0$. All the matrices $X$ that satisfy the new condition form a linear subspace of $\mathbb{R}^{2n \times 2n}$ called the tangent space of $Sp$ at the point $S$:

$$T_{S}Sp = \{X \in \mathbb{R}^{2n \times 2n} | X\Omega S^T + S\Omega X^T = 0_{2n}\} \quad (108)$$

$$= \{\Omega A | A = A^T\}. \quad (109)$$

Eq. (109) is a compact way of parametrizing the tangent space at $S$ using symmetric matrices. It can be found by imposing $X = S\Omega A$ in the tangency condition.

As a special case, the Lie algebra of $Sp$ is the tangent space at the identity; i.e.

$$sp(2n, \mathbb{R}) = T_eSp(2n, \mathbb{R}) \quad (110)$$

$$= \{X \in \mathbb{R}^{2n \times 2n} | X\Omega + \Omega X^T = 0_{2n}\} \quad (111)$$

$$= \{\Omega A | A = A^T\}. \quad (112)$$

We can then define the normal space at $S$ as the linear space containing all the elements that are orthogonal to $T_{S}Sp$:

$$N_{S}Sp = \{W \in \mathbb{R}^{2n \times 2n} | \text{Tr}(W^T X) = 0_{2n}, X \in T_{S}Sp\} \quad (113)$$

$$= \{\Omega SB | B = -B^T\}, \quad (114)$$

with Eq. (114) showing that we can parametrize the normal space at each point in $Sp$ using anti-symmetric matrices.

C. Riemannian metric on $Sp(2n)$

A Riemannian manifold such as $Sp(2n, \mathbb{R})$ comes equipped with an inner product $\langle \cdot, \cdot \rangle_S$ on the tangent space $T_{S}Sp$ at each point $S \in Sp$. The family of inner products forms the Riemannian metric tensor. The inner product in $T_{S}Sp$ is defined as

$$\langle X, Y \rangle_S = \langle S^{-1}X, S^{-1}Y \rangle = \langle RX, Y \rangle, \quad (115)$$

where $R = S^{-T}S^{-1} = \Omega SS^T\Omega^T$ and note that $R^{-1} = SS^T$.

Consider now a cost function $L : Sp \rightarrow \mathbb{R}$. The Euclidean gradient $\partial L$ at the point $S$ (which is computed using the embedding coordinates in $\mathbb{R}^{2n \times 2n}$) is related to the Riemannian gradient $\nabla L \in T_{S}Sp$ by the compatibility condition

$$\langle \nabla L, X \rangle_S = \langle \partial L, X \rangle \quad \forall X \in T_{S}Sp. \quad (116)$$

After rearranging the terms, the condition is equivalent to

$$\langle R\nabla L - \partial L, X \rangle = 0 \quad \forall X \in T_{S}Sp. \quad (117)$$

This means that $R\nabla L - \partial L \in N_{S}Sp$ and therefore, it must be possible to write

$$R\nabla L - \partial L = \Omega SB, \quad (118)$$

for some anti-symmetric matrix $B$. At the same time we have the tangency condition $\nabla L SS^T + S\Omega S^T L = 0$. If we replace $\nabla L$ from Eq. (118) into the tangency condition, we obtain an expression for $B$ and we can finally write the Riemannian gradient on the symplectic group:

$$\nabla L = \frac{S}{2} (Z + \Omega Z^T\Omega), \quad (119)$$

where $Z = S^T \partial L$.

The symplectic matrix that describes an interferometer belongs to the intersection of the orthogonal group $O(2n)$ and the symplectic group $Sp(2n)$, which is a Unitary group $U(n)$:

$$U(n) = \{M \in \mathbb{C}^{n \times n} | M^\dagger M = MM^\dagger = 1_n\}. \quad (120)$$

We can go through the same arguments as with the symplectic group and obtain the Riemannian gradient in the unitary group (More calculation details are in Appendix B):

$$\nabla L = \frac{M}{2} (Z - Z^\dagger), \quad (121)$$

where $Z = M^\dagger \partial L$.

D. Geodesic optimization on $Sp(2n)$ and $U(n)$

The shortest curve connecting two points on a Riemannian manifold $M$ is called a geodesic, and it can be defined by the starting point $\gamma(0) = p$ and its velocity on the tangent space at that point: $V = \dot{\gamma}(0) \in T_pM$. For the symplectic and unitary groups, geodesics take the following form (which can be found by minimizing a variational formulation of the path length between two points $[49, 50]$):

$$\gamma_{Sp(2n)}(t) = Se^{t(S^{-1}V)}e^{t(S^{-1}V - (S^{-1}V)^T)}, \quad (122)$$

$$\gamma_{U(n)}(t) = Me^{t(M^\dagger V)^T}. \quad (123)$$

By using a geodesic, we guarantee that each update step remains on the manifold.

For gradient descent, we use $V = -\nabla L$:

$$\gamma_{Sp(2n)}(t) = Se^{-tY}e^{-t(Y-Y^T)}, \quad (124)$$

with $Y = S^{-1}\nabla L = \frac{1}{2}(Z + \Omega Z^T\Omega)$. For the unitary group, we obtain

$$\gamma_{U(n)}(t) = Me^{-tY}, \quad (125)$$
with $Y = M^d \nabla L = \frac{1}{2}(Z - Z^T) = \frac{1}{2}(M^d \partial L - (\partial L)^d M)$. We now have a geodesic update formula that we can apply in place of the usual gradient descent step. The parameter $t$ takes the role of the learning rate (which we fix depending on the application). For the symplectic group, we have

$$Z_k \leftarrow S_k^T \partial L,$$

$$Y_k \leftarrow \frac{1}{2}(Z_k + \Omega Z_k \Omega),$$

$$S_{k+1} \leftarrow S_k e^{-iY_k} e^{-i(t Y_k - Y_k^T)}.$$  \hspace{1cm} (126)

For the unitary group, we have

$$Z_k \leftarrow M_k^d \partial L,$$

$$Y_k \leftarrow \frac{1}{2}(Z_k - Z_k^T),$$

$$M_{k+1} \leftarrow M_k e^{-iY_k}.$$  \hspace{1cm} (129)

Finally, we obtain the orthogonal matrix of the interferometer using Eq. (34).

E. The Riemannian update step in practice

All Gaussian objects can be updated in a learning step on the symplectic group, on the displacement parameters, or on the symplectic eigenvalues. As the latter two are Euclidean updates, we will not describe them in great detail. In fact, once the relevant Euclidean gradient has been computed, the update rule can be taken as a single step of gradient descent or one of its variants (e.g. using momentum). For instance, the update of the displacement parameter could simply follow the rule

$$d \leftarrow d - t \frac{\partial L}{\partial d},$$  \hspace{1cm} (132)

using the Euclidean gradient.

We will concentrate then on detailing the update on the symplectic group and we will take Gaussian unitaries as a basic example (pure states, mixed states, and channels can have a symplectic matrix among their parameters, via the Choi-Jamiolkowski isomorphism).

The backpropagation procedure is shown in Fig. 1. The Euclidean gradient of the symplectic matrix can be calculated via the chain rule:

$$\frac{\partial L}{\partial S} = 2R \left[ \sum_{X=A,b,c} \sum_k \frac{\partial L}{\partial G_k} \frac{\partial X}{\partial S} \frac{\partial G_k}{\partial X} \right].$$  \hspace{1cm} (133)

\begin{figure}[h]
\centering
\begin{tikzpicture}
\node at (0,0) {\text{Sp}(2n, \mathbb{R}) \times \mathbb{R}^{2n} (S, d)};
\node at (2,0) {\text{Wigner} (V, \bar{r})};
\node at (4,0) {\text{Bargmann} (A, b, c)};
\node at (6,0) {\text{Fock} (G_k)};
\node at (8,0) {\mathbb{R}};
\node at (10,0) {\text{Cost Function}};
\node at (12,0) {\text{Recurrence relations}};
\node at (14,0) {\text{Gaussian evolution}};
\node at (16,0) {\text{Complexify} + \text{Cayley transform}};
\node at (18,0) {\text{R}};
\node at (19,0) {\text{Open-source library \text{TheWalrus} [51] and \text{MrMustard} [1]}.
\end{tikzpicture}
\caption{The detailed forward and backward passes. The Riemannian gradient $\nabla L$ for the geodesic update is calculated via the chain rule and Eq. (119), which backpropagates the gradient of the cost function with respect to the Fock amplitudes $\partial L/S$, all the way to $\nabla L$, while the gradient $\partial L/S$ is used directly to optimize $d$ on $\mathbb{R}^{2n}$. The backpropagation steps can be left to an Automatic Differentiation framework, except for the Fock to Bargmann step and the conversion between Euclidean and Riemannian gradient, which we implement ourselves.}
\end{figure}

FIG. 1. The detailed forward and backward passes. The Riemannian gradient $\nabla L$ for the geodesic update is calculated via the chain rule and Eq. (119), which backpropagates the gradient of the cost function with respect to the Fock amplitudes $\partial L/S$, all the way to $\nabla L$, while the gradient $\partial L/S$ is used directly to optimize $d$ on $\mathbb{R}^{2n}$. The backpropagation steps can be left to an Automatic Differentiation framework, except for the Fock to Bargmann step and the conversion between Euclidean and Riemannian gradient, which we implement ourselves.

Then, we can write the update rule for the real symplectic matrix $S$ to follow a geodesic path starting at $S$ with a velocity $\nabla L$ defined by its Riemannian gradient and guarantee the updated matrix is still on Sp(2n).

VI. NUMERICAL EXPERIMENTS

In this section, we showcase the optimization methods introduced in the previous sections with two examples. The recurrent methods presented here paper are implemented in the open-source library TheWalrus [51] and they are integrated with the optimization methods in the open-source library MrMustard [1].
A. Maximizing the entanglement in Gaussian Boson Sampling

We first analyze high-dimensional Gaussian Boson Sampling (GBS) instances similar to the 216-mode circuit of the photonic processor Borealis [52]. In a $D$-dimensional high-dimensional GBS instance with $M = d^D$ modes, a set of $K \leq M$ squeezed modes are sent into an interferometer composed of layers of beamsplitter gates (with a local rotation gate in the first mode) between modes $i$ and $i + \tau$ with $\tau \in \{1, d, d^2, \ldots, d^{D-1}\}$ as shown schematically for $d = 6$ and $D = 3$ in Fig. 2.

One desirable property of any GBS instance is that its adjacency matrix, which corresponds to $A_\psi$ in our notation, should not have any special property like being banded, sparse, or low-rank. This is because these types of properties can be exploited to speed-up the classical simulation of GBS.

For high-dimensional GBS instances like the one implemented in Borealis, it is known that the $A_\psi$ is full-rank (since every input is squeezed) and not banded (due to the long-ranged gates). However, one needs to judiciously choose the parameters of the beamsplitter so that the distribution of its entries is not heavily dominated by just a few of them. For example, if the one chooses the rotation gates and the transmission angles of the beamsplitters to be uniformly random in $[-\frac{\pi}{2}, \frac{\pi}{2}]$ one obtains the distribution shown in blue bars in Fig. 3c and the $A_\psi$ matrix shown in Fig. 3a. For these results and following Ref. [52] we fix the phase angle of the beamsplitter to be $\pi/2$, we set the input squeezing parameter in all the modes to be $r = \text{arcsinh} \, 1 \approx 0.8813736$ and take $D = 3$, $d = 6$ and thus a total of $M = 6^3 = 216$ modes. Note that the values of the matrix are heavily concentrated, i.e., for each row and column a few values are overwhelmingly larger than the rest.

We can now use the methods we developed to try to spread-out as much as possible the entries of the matrix $A_\psi$. We can then optimize the cost function

$$\min \sum_{ij} (|A_\psi|_{ij} - \text{mean}|A_\psi|^2)^2. \quad (134)$$

We perform this optimization obtaining the distribution shown with the orange bars in Fig. 3c and the matrix shown in Fig. 3b. Notice that now the values are more evenly distributed.

B. State preparation

In this section, we find explicit circuits that prepare cat states for fault-tolerant quantum computation. Our examples contain different ingredients: optical components with trainable parameters, a suitable cost function and optimization routines for updating the trainable parameters.

We consider three kinds of circuits:

1. Gaussian gates with symplectic optimization;
2. Squeezing gate and multi-mode interferometers
with unitary optimization;

3. Squeezing gate and beamsplitter with Euclidean optimization.

In the experiments, we show the first and third (as in this case, the beamsplitter is equivalent to a two-mode interferometer) using the Euclidean optimization for the basic optical components (single-mode squeezing, rotation operator, beamsplitter, etc.) and the symplectic optimization for Gaussian objects.

Cat states are superpositions of coherent states which we write as

\[ |\text{cat}_{\pm}\rangle = \mathcal{N} (|\alpha\rangle \pm |\alpha^*\rangle) , \]

where \( \mathcal{N} \) is a normalization constant and \(|\alpha\rangle = D(\alpha)|0\rangle \) is a coherent state. In the last equation, the plus sign corresponds to even cat states and the minus corresponds to odd cat states. The normalization constant is never \( 1/\sqrt{2} \) as the two coherent states are never fully orthogonal, but they approach orthogonality exponentially fast in \(|\alpha|\).

For this example, we will target the generation of an odd cat state with \( \alpha = 2 \) and will employ the symplectic and the euclidean optimizers of MrMustard (version 0.3.0).

The first circuit (shown in Fig. 3a) consists of a Gaussian gate and photon-number measurement on the first mode and generates the (approximate) cat state in the second mode. We use the symplectic optimizer to train the Gaussian gate. The result is shown in Fig. 4b with a fidelity of 99.42% and 5.40% success probability.

The code snippet below corresponds to the circuit shown in Fig. 3a:

```python
import numpy as np
from mrmustard.lab import *
from mrmustard.physics import fidelity, normalize
from mrmustard.training import Optimizer

#Target cat state: Normalized(|alpha> - |-alpha>)
alpha = 2.0
cutoff = 50

cat_amps = (Coherent(alpha).ket([cutoff])
    - Coherent(-alpha).ket([cutoff]))
cat_target = normalize(State(ket=cat_amps))
np.random.seed(1)
circ = Ggate(num_modes=2, symplectic_trainable=True)
def output():
    return Vacuum(num_modes=2) >> circ
<< Fock(3, modes=[0])

def cost_fn():
    return -fidelity(normalize(output()), cat_target)

opt = Optimizer(symplectic_lr = 0.005)
opt.minimize(cost_fn, by_optimizing=[circ],
    max_steps=2000)
```

In the second circuit (shown in Fig. 3b), we use a simple two-mode circuit, starting with two single-mode squeezers followed by a beamsplitter and ending with a Fock measurement of 3 photons on the first mode. We train the circuit from random initial parameters for two squeezers and a beamsplitter. We obtain Fig. 4c as a result with a fidelity of 99.38% and 7.39% success probability.

Another code snippet below corresponds to the circuit shown in Fig. 3b:
(a) Target cat state defined in Eq. (135) with $\alpha = 2$.

(b) Optimization of the cat state with the circuit in Fig. 3a.

(c) Optimization of the cat state with the circuit in Fig. 3b.

FIG. 4. The target cat state and the optimized cat state.

VII. EXTENSIONS TO LINEAR COMBINATIONS OF GAUSSIANS

While the set of Gaussian states is rather restrictive, many non-Gaussian states of interest, such as cat states, Gottesman-Kitaev-Preskill (GKP) states [53], or Fock states, can be written as linear combinations of Gaussians in phase space [24]. This representation has the nice property that any Gaussian channel can act on these states directly in phase space, i.e., without requiring to write their Fock representation explicitly. Because of linearity, we can simply obtain the Fock representation of any states expressible as a linear combination of Gaussians by obtaining the Fock representation of each Gaussian component. This argument is equally valid for pure and mixed states. For the case of pure states, it is important to correctly account for the global phase as described in the previous sections. This phase will be important for states for which the coefficients $c_{\psi}$ have non-trivial dependence on the displacement and squeezing that describes each individual component, as it is apparent in squeezed-comb states defined as [54]

$$|0_{\text{Comb}}\rangle = \frac{1}{N_{\text{Comb}}} \sum_{n=1}^{N} |\psi_{n}\rangle, \quad |\psi_{n}\rangle = D(q_{n})S(r)|0\rangle,$$

(136)

where recall $D(\cdot)$ and $S(\cdot)$ are the single-mode displacement and squeezing operator defined in Sec. II C, $q_{n} = -(N+1)(d/2) + nd$. Note that squeezed-comb states have as limit both cat states (when the squeezing parameters are zero and $N = 2$) and the GKP states (when $r > 0$ and $N$ is large). Note that each element in the linear combination will have a non-trivial phase that appears in a linear superposition and thus cannot be factored out as a global phase, making clear the relevance of the results in Sec. IV.
Consider now the density matrix associated with the state above
\[
\rho_{\text{Comb}} = |0_{\text{Comb}}\rangle \langle 0_{\text{Comb}}| = \sum_{n=1}^{N} |\psi_n\rangle \langle \psi_n| + \sum_{n=1}^{N} \sum_{m=1, m \neq n}^{N} |\psi_n\rangle \langle \psi_m|.
\]

(138)

On the one hand, the “diagonal” terms $|\psi_n\rangle \langle \psi_n|$ correspond to positive semi-definite operators with Gaussian characteristic functions. On the other hand, the “off-diagonal” terms $|\psi_n\rangle \langle \psi_m|$ do not represent positive semi-definite operators but they still have complex-Gaussian characteristic functions as shown in Appendix A of Ref. [24]. This implies that the recursion relations derived in this manuscript still hold for each term in the equation above.

VIII. CONCLUSION

In this work we have presented a linear recurrence relation that connects the phase space and the Fock space representations of Gaussian pure and mixed states, as well as Gaussian unitary and non-unitary transformations. While working with Gaussian gates within the phase space representation is easily achieved using symplectic algebra, it is valuable to implement fast numerical simulations in Fock representation, in order to include non-Gaussian effects. Moreover, the recurrence relation is exact and differentiable, which enables accurate gradient computations and gradient-based optimization.

Since the covariance matrix of Gaussian objects is parametrized by symplectic matrices that live in a Riemannian manifold, a geodesics-based optimization method is proposed in this paper. We show some optimization examples using the open-source library MrMustard, where we implemented our methods. In particular, we optimized the adjacency matrix of a high-dimensional Gaussian Boson Sampling instance with 216 modes directly in phase space to highlight the euclidean optimization functionality of our library. We then obtained new circuits to generate mesoscopic cat states with unprecedented success probability. On the theory side, we also showed how to keep track of the global phase induced by Gaussian unitary transformations. This paves the way to simulate and optimize non-Gaussian objects by writing them as linear combinations of Gaussians [24]. Dealing with non-Gaussian simulation and optimization is a significant challenge in the optical information processing community [9, 55]. Our methods offer a promising avenue to address this challenge.

ACKNOWLEDGEMENTS

N.Q. thanks R. Chadwick and T. Kalajdzievski for valuable discussions and the Ministère de l’Economie et de l’Innovation du Québec and the Natural Sciences and Engineering Research Council of Canada for financial support.

[1] XanaduAI, Mrmustard, https://github.com/XanaduAI/MrMustard (2021).
[2] C. Gerry and P. Knight, Introductory Quantum Optics (Cambridge University Press, 2004).
[3] G. Carioleto, Quantum Communications (Springer International Publishing, 2015).
[4] M. Amazigou, M. Nassik, and N. Habiballah, Entanglement, EPR steering and gaussian geometric discord in a double cavity optomechanical systems, The European Physical Journal D 72, 10.1140/epjd/e2018-90151-6 (2018).
[5] J. R. McClean, J. Romero, R. Babbush, and A. Aspuru-Guzik, The theory of variational hybrid quantum-classical algorithms, New Journal of Physics 18, 023023 (2016).
[6] M. W. Matsen, The standard gaussian model for block copolymer melts, Journal of Physics: Condensed Matter 14, R21 (2001).
[7] C. S. Hamilton, R. Kruse, L. Sansoni, S. Barkhofen, C. Silberhorn, and I. Jex, Gaussian boson sampling, Physical review letters 119, 170501 (2017).
[8] N. Quesada, Franck-Condon factors by counting perfect matchings of graphs with loops, J. Chem. Phys. 150, 164113 (2019).
[9] N. Quesada, L. Helt, J. Izaac, J. Arrazola, R. Shahrokhshahi, C. Myers, and K. Sabapathy, Simulating realistic non-Gaussian state preparation, Phys. Rev. A 100, 022341 (2019).
[10] V. V. Dodonov, O. V. Man'ko, and V. I. Man'ko, Multidimensional hermite polynomials and photon distribution for polynode mixed light, Phys. Rev. A 50, 813 (1994).
[11] E. Doktorov, I. Malkin, and V. Man’ko, Dynamical symmetry of vibronic transitions in polyatomic molecules and the franck-condon principle, Journal of Molecular Spectroscopy 56, 1 (1975).
[12] R. Berger, C. Fischer, and M. Klessinger, Calculation of the vibronic fine structure in electronic spectra at higher temperatures. 1. benzene and pyrazine, The Journal of Physical Chemistry A 102, 7157 (1998).
[13] D. Gruner and P. Brumer, Efficient evaluation of harmonic polyatomic franck-condon factors, Chemical Physics Letters 138, 310 (1987).
[14] S. M. Rabidoux, V. Eijkhout, and J. F. Stanton, A highly-efficient implementation of the doktorov recurrence equations for franck-condon calculations, Journal of Chemical Theory and Computation 12, 728 (2016).
[15] J. Huh, Multimode bogoliubov transformation and husimi’s q-function, Journal of Physics: Conference Series 1612, 012015 (2020).
Appendix A: Review of the symplectic formalism

The real symplectic group is defined as

\[ \text{Sp}(2n, \mathbb{R}) = \{ S \in \mathbb{R}^{2n \times 2n} | S\Omega S^T = \Omega \} \]  

(A1)

where \( \Omega \) is defined in Eq. (8).

Some properties of this group:

\[ \Omega \in \text{Sp}(2n, \mathbb{R}) \]  

(A2)

\[ \Omega^{-1} = \Omega^T = -\Omega \in \text{Sp}(2n, \mathbb{R}) \]  

(A3)

\[ S^{-1} = -\Omega S^T \Omega \in \text{Sp}(2n, \mathbb{R}) \]  

(A4)

A real symplectic matrix \( S \) can be decomposed as

\[ S = O_1 \Lambda O_2, \]  

(A5)

with \( O_1, O_2 \in C(n) \) and

\[ \Lambda = \Lambda_x \otimes \Lambda_x^{-1}, \]  

(A6)

with \( \Lambda_x = \text{diag}(\lambda_1, \ldots, \lambda_n) \) and \( \lambda_j > 0 \forall j \in [1, \ldots, n] \). \( C(n) \) denotes the compact subgroup and \( C(n) = \text{Sp}(2n, \mathbb{R}) \cap \text{O}(2n) \). It means that any symplectic matrix can be decomposed into a diagonal and positive semi-definite matrix \( \Lambda \) with two orthogonal groups \( O_1 \) and \( O_2 \), which stands for the passive transformation (interferometer).

Appendix B: Riemannian gradient of the unitary group

The Riemannian metric of the unitary group at point \( A \) is:

\[ \langle X, Y \rangle_A = \langle A^{-1}X, A^{-1}Y \rangle_{12n} = \text{Tr} \left( (A^{-1}X)^\dagger A^{-1}Y \right), \quad X, Y \in T_A \text{U}(n, \mathbb{C}). \]  

(B1)

The Riemannian gradient \( \nabla_A f \) at point \( A \) of a sufficiently regular function \( f : \text{U}(n, \mathbb{C}) \to \mathbb{C} \) associated to the Riemannian metric satisfies

\[ \nabla_A f = \frac{1}{2} \left( \partial_A f - A \partial_A^\dagger f A \right). \]  

(B2)

Proof. According to the compatibility of the Riemannian gradient with the Riemannian metric (defined in Eq. (B1)), we have:

\[ \langle \nabla_A f, T \rangle_A = \langle \partial_A f, T \rangle_{\text{euc}}, \quad \forall T \in T_A \text{Sp}, \]  

(B3)

that it,

\[ \langle \partial_A f - A^{-\dagger} A^{-1} \nabla_A f, T \rangle_{\text{euc}} = 0. \]  

(B4)
This implies that $\partial_A f - A^\dagger A^{-1} \nabla_A f \in N_A U$. So we have, with $AA^\dagger = \mathbb{1}$:

$$\partial_A f - A^{-1} A^\dagger \nabla_A f = AN \quad \text{(B5)}$$

$$\partial_A f = \nabla_A f + AN. \quad \text{(B6)}$$

Using the tangency condition $\nabla_A f \in T_A U$, we know

$$(\nabla_A f)^\dagger A + A^\dagger \nabla_A f = 0_n. \quad \text{(B7)}$$

Together Eq. (B7) and Eq. (B6) with $N = N^\dagger$, we solve

$$N = \frac{1}{2} \left(A^\dagger \partial_A f + \partial_A^\dagger f A\right). \quad \text{(B8)}$$

Thus we obtain

$$\nabla_A f = \partial_A f - A^{-1} \left(A^\dagger \partial_A f + \partial_A^\dagger f A\right) \quad \text{(B9)}$$

$$= \frac{1}{2} \left(\partial_A f - A \partial_A^\dagger f A\right). \quad \text{(B10)}$$

**Appendix C: Choi-Jamiolkowski isomorphism**

**FIG. 5.** 2M-mode circuit for implementing the Choi-Jamiolkowski isomorphism. $\Phi$ is the channel that is applied on the first half $M$ modes and the two dots represent a two-mode squeezing operator connecting two modes: one comes from the first $M$ modes and the other one comes from the second $M$ modes.

In this section, we employ the Choi-Jamiolkowski isomorphism [26, 56, 57] to reduce the calculation of the matrix elements of an arbitrary Gaussian channel in $M$ to the calculation of the matrix element of a Gaussian state with $2M$. We first consider a collection of systems with arbitrary, but identical, dimensionality $N$.

We write the state right before the channel $\Phi$ is applied to the first half of the modes in Fig. 5 as

$$|\Psi\rangle = \sqrt{N} \sum_{n=0}^{N-1} \tau^n \langle n | \otimes | n \rangle, \quad \text{(C1)}$$

where $\sum_{n=0}^{N-1} \equiv \sum_{n_1=0}^{N-1} \cdots \sum_{n_M=0}^{N-1}$. $N$ is a normalization constant to be determined in a moment and $\tau$ is the squeezing parameter of the two-mode squeezing operator connecting the first $M$ modes and the second $M$ modes. The density matrix of the state $|\Psi\rangle$ is simply

$$|\Psi\rangle\langle\Psi| = N \sum_{m=0}^{N-1} \sum_{n=0}^{N-1} \tau^{n+m} |n \rangle \otimes |n \rangle \langle m | \otimes |m \rangle. \quad \text{(C2)}$$

We can now write the output of the circuit after the application of the channel $\Phi$ as

$$\rho = (\Phi \otimes \mathbb{1}) |\Psi\rangle\langle\Psi| = N \sum_{m=0}^{N-1} \sum_{n=0}^{N-1} \tau^{n+m} \Phi |n \rangle \langle m | \otimes |n \rangle \langle m |. \quad \text{(C3)}$$
We can premultiply the equation above by \( \langle i \rangle \otimes \langle j \rangle \) and postmultiply by \( |k\rangle \otimes |l\rangle \) to obtain
\[
(\langle i \rangle \otimes \langle j \rangle) \rho (|k\rangle \otimes |l\rangle) = N \tau^{j+l} \langle i \rangle (\Phi |j\rangle \langle l|) |k\rangle.
\] (C4)

In finite-dimensional systems it is convenient to pick \( \tau = (1, \ldots, 1) \) and the normalization \( N \) is simply given by the dimensionality of the system \( N^M \). For infinite dimensional systems, if one were to try to pick the same normalization as for a finite-dimensional, one would obtain a non-normalizable state \( |\Psi\rangle \). Thus it is convenient to pick \( \tau = (\tau, \ldots, \tau) \) with \( \tau = \tanh t < 1 \) and then
\[
N = (1 - \tau^2)^M = (1 - \tanh^2 t)^M, \quad \tau^{j+l} = (\tanh t)^{\sum_{i=1}^M l_i + j_i}.
\] (C5)

For a rigorous justification of this derivation see sec 5.5 of Serafini [26]. Now consider the case where the channel \( \Phi \) is Gaussian parametrized by
\[
X = \begin{bmatrix} X_{qq} & X_{qp} \\ X_{pq} & X_{pp} \end{bmatrix}, \quad Y = \begin{bmatrix} Y_{qq} & Y_{qp} \\ Y_{pq} & Y_{pp} \end{bmatrix}, \quad d = \begin{bmatrix} d_q \\ d_p \end{bmatrix}.
\] (C7)

Then the output state is also Gaussian since the input state to the channel is nothing but one-half of a two-mode squeezed state. In this case, we can write the quadrature covariance matrix and vector of means of the output state as
\[
V = \tilde{X} \mathcal{T}(t) \left( \frac{\hbar}{2} \mathbb{I}_{4M} \right) (\mathcal{T}(t))^T \tilde{X}^T + \tilde{Y} = \frac{\hbar}{2} \tilde{X} \mathcal{T}(2t) \tilde{X}^T + \tilde{Y}, \quad \bar{r} = \begin{bmatrix} d_q \\ d_p \end{bmatrix},
\] (C8)

where
\[
\tilde{X} = \begin{bmatrix} X_{qq} & 0_M & X_{qp} & 0_M \\ 0_M & \mathbb{I}_M & 0_M & 0_M \\ X_{pq} & 0_M & X_{pp} & 0_M \\ 0_M & 0_M & \mathbb{I}_M & 0_M \end{bmatrix}, \quad \tilde{Y} = \begin{bmatrix} Y_{qq} & 0_M & Y_{qp} & 0_M \\ 0_M & 0_M & 0_M & 0_M \\ Y_{pq} & 0_M & Y_{pp} & 0_M \\ 0_M & 0_M & 0_M & 0_M \end{bmatrix},
\] (C9)
\[
\mathcal{T}(t) = \begin{bmatrix} \cosh t \mathbb{I}_M & \sinh t \mathbb{I}_M & 0_M & 0_M \\ \sinh t \mathbb{I}_M & \cosh t \mathbb{I}_M & 0_M & 0_M \\ 0_M & 0_M & \cosh t \mathbb{I}_M & -\sinh t \mathbb{I}_M \\ 0_M & 0_M & -\sinh t \mathbb{I}_M & \cosh t \mathbb{I}_M \end{bmatrix},
\] (C10)

and we used the fact that \( \mathcal{T}(t) (\mathcal{T}(t))^T = \mathcal{T}(t) (\mathcal{T}(t)) = \mathcal{T}(2t) \).

In the next appendix, we show that we can associate with the 2\( M \)-Gaussian Choi-Jamiołkowski the following
quantities

\[ A_\rho = E(t)A_\Phi E(t), \quad (C11) \]
\[ A_\Phi = P_{2M}R \begin{bmatrix} 1 & -X \xi^{-1} & X \xi^{-1} \\ X^T \xi^{-1} & 1 & X^T \xi^{-1} \end{bmatrix} R^\dagger, \quad (C12) \]
\[ = P_{2M}R \begin{bmatrix} \xi^{-1} & 0 \\ 0 & -X^T \xi^{-1} \end{bmatrix} R^\dagger, \quad (C13) \]
\[ b_\rho = E(t)b_\Phi, \quad (C14) \]
\[ b_\Phi = \frac{1}{\sqrt{h}} R^* \begin{bmatrix} \xi^{-1}d \\ -X^T \xi^{-1}d \end{bmatrix}, \quad (C15) \]
\[ c_\rho = (1 - \tanh^2 t)^M c_\Phi, \quad (C16) \]
\[ c_\Phi = \frac{\exp \left[ -\frac{1}{2\pi} d^T \xi^{-1}d \right]}{\sqrt{\det(\xi)}}, \quad (C17) \]

where \( E(t) = 1_M \oplus (\tanh t 1_M) \oplus 1_M \oplus (\tanh t 1_M) \), \( P_M = \begin{bmatrix} 0_M & 1_M \\ 1_M & 0_M \end{bmatrix} \) and

\[ R = \frac{1}{\sqrt{2}} \begin{bmatrix} 1_M & i1_M & 0_M & 0_M \\ 0_M & 0 & 1_M & -i1_M \\ 1_M & -i1_M & 0_M & 0_M \\ 0_M & 0_M & 1_M & i1_M \end{bmatrix}, \quad \xi = \frac{1}{2} \left( 1_M + XXT + \frac{2}{h} Y \right). \quad (C18) \]

Note that \( \xi \) is nothing but the \( qp \)-Husimi covariance matrix (in units where \( h = 1 \)) of the state obtained by sending the \( M \) mode vacuum state in the process specified by \( \Phi \).

With these results we can write

\[ \langle i | \otimes \langle j | \rangle (|k\rangle \otimes |l\rangle) = c_\rho \times \frac{G_{n}^{A_\rho}(b_\rho)}{\sqrt{i!j!k!l!}}. \quad (C19) \]

Now we recall a fundamental property that multidimensional Hermite polynomials inherit from loop-hafnians [32], namely that if \( E = \bigoplus_{i=1}^{\ell} E_i \) is a diagonal matrix then

\[ G_{n}^{EAE}(Eb) = \left( \prod_{i=1}^{\ell} E_{ni} \right) G_{n}^{A_n}(b), \quad (C20) \]

We can use the definitions from Eq. (C11) to Eq. (C17) together with the Eq. (C4) and the relation Eq. (C19) to find

\[ \langle i | \Phi | j \rangle |l\rangle \rangle |k\rangle = \frac{\langle i | \otimes \langle j | \rangle (|k\rangle \otimes |l\rangle) = c_\rho \times \frac{G_{k_{0}1_{i+1}}^{A_\rho}(b_\rho)}{\sqrt{i!j!k!l!} \times c_\Phi \times \frac{G_{k_{0}1_{i+1}}^{A_\Phi}(b_\Phi)}{\sqrt{i!j!k!l!}}, \quad (C21) \]

which allows us to find the matrix elements of the channel without any reference to the specific amount of squeezing used to create the two-mode squeezed vacuum.

Appendix D: Description of the Choi-Jamiołkowski isomorphism in Phase-Space

The (complex) covariance matrix \( \sigma \) of the Gaussian state obtained by sending \( M \) halves of \( M \) two-mode squeezed vacuum states through the channel \( \Phi \) is given by

\[ \sigma = W \left( \frac{1}{2} \hat{X} \mathcal{T}(2t) \hat{X}^T + \frac{Y}{h} \right) W^\dagger. \quad (D1) \]
Note that \((\mathbf{T}(t))^T = \mathbf{T}(t)\) is symmetric, \(\mathbf{X}\) is symplectic if \(X = \begin{bmatrix} X_{qq} & X_{qp} \\ X_{pq} & X_{pp} \end{bmatrix}\) is symplectic and \(\mathbf{W}\) is unitary. Let

\[
Q' = \left(\frac{1}{4M} + \frac{1}{2} \mathbf{X} \mathbf{T}(2t) \mathbf{X}^T + \frac{Y}{\hbar}\right),
\]

then \((\sigma + \frac{1}{4M})^{-1} = W(Q')^{-1}W^\dagger\). Now we define

\[
Q = LQ'L^T,
\]

with

\[
L = \begin{bmatrix} 1_M & 0_M & 0_M & 0_M \\ 0_M & 1_M & 0_M & 0_M \\ 0_M & 0_M & 1_M & 0_M \\ 0_M & 0_M & 0_M & 1_M \end{bmatrix},
\]

Then we have that \(Q^{-1} = L(Q')^{-1}L^T\), which implies that \(L^TQ^{-1}L = (Q')^{-1}\). So calculating \(Q^{-1}\) gives \((Q')^{-1}\) and therefore \((\sigma + \frac{1}{4M})^{-1}\).

Expressing \(Q\) as a block matrix \(Q = \begin{bmatrix} A & B \\ C & D \end{bmatrix}\), we can write \(Q^{-1}\) using Schur complements as [26]

\[
Q^{-1} = \begin{bmatrix} \xi^{-1} & -\xi^{-1}BD^{-1} \\ -D^{-1}C\xi^{-1} & D^{-1} + D^{-1}C\xi^{-1}BD^{-1} \end{bmatrix},
\]

where \(\xi = \mathbf{A} - BD^{-1}\mathbf{C}\). The blocks \(A, B, C,\) and \(D\), are given by

\[
A = \frac{Y}{\hbar} + \frac{1}{2M} + \frac{1}{2} \cosh(2t)XX^T,
\]

\[
B = \frac{1}{2} \sinh 2t \begin{bmatrix} X_{qq} & -X_{qp} \\ X_{pq} & -X_{pp} \end{bmatrix} = \frac{1}{2} \sinh 2tXX^T,
\]

\[
C = \frac{1}{2} \sinh 2t \begin{bmatrix} X^T_{qq} & X^T_{qp} \\ -X^T_{pq} & -X^T_{pp} \end{bmatrix} = B^T = \frac{1}{2} \sinh 2tZX^T,
\]

\[
D = \cosh^2(t) \begin{bmatrix} 1_M & 0_M \\ 0_M & 1_M \end{bmatrix},
\]

where \(Z = \begin{bmatrix} 1_M & 0_M \\ 0_M & -1_M \end{bmatrix}\). We now use these to calculate the blocks of \(Q^{-1}\) starting with \(\xi\),

\[
\xi = A - BD^{-1}C = \frac{1}{2} \left( \mathbb{1}_{2M} + XX^T + \frac{2Y}{\hbar} \right) = \xi^T,
\]

which turns out to be independent of \(t\). Next, we find

\[
-\xi^{-1}BD^{-1} = -\tanh(t)\xi^{-1}XZ,
\]

\[
-D^{-1}C\xi^{-1} = -\tanh(t)ZX^T\xi^{-1}.
\]

Finally, the bottom right block, which can be simplified by substituting the other three blocks, is given by

\[
D^{-1} + D^{-1}C\xi^{-1}BD^{-1} = (1 - \tanh^2(t)) \mathbb{1}_{2M} + \tanh^2(t)ZX^T\xi^{-1}XZ,
\]

\[
= \mathbb{1}_{2M} + \tanh^2(t)Z (X^T\xi^{-1}X - \mathbb{1}_{2M}) Z.
\]
Putting these blocks together, we get the expanded form of $Q^{-1}$

$$Q^{-1} = \begin{bmatrix} \xi^{-1} & -\tanh(t)\xi^{-1}XZ \\ -\tanh(t)ZX^{T}\xi^{-1} & \mathbb{1}_{2M} + \tanh^{2}(t)Z(X^{T}\xi^{-1}X - \mathbb{1}_{2M})Z \end{bmatrix}. \tag{D15}$$

Now with the form of the inverse known, we can multiply by the remaining matrices to get the final form of $\sigma^{-1}_{+} = (\sigma + \frac{1}{2}\mathbb{1}_{2M})^{-1} = WL^{T}Q^{-1}LW^{T}$, with $L$ as in Eq. (D4).

We can now go back and write the quantity of interest

$$\mathbb{1}_{4M} - \left( \sigma + \frac{1}{2}\mathbb{1}_{4M} \right)^{-1} = WL^{T}\left( \mathbb{1}_{4M} - \begin{bmatrix} \xi^{-1} & -\tanh(t)\xi^{-1}XZ \\ -\tanh(t)ZX^{T}\xi^{-1} & \mathbb{1}_{2M} + \tanh^{2}(t)Z(X^{T}\xi^{-1}X - \mathbb{1}_{2M})Z \end{bmatrix} \right)LW^{T},$$

$$= WL^{T}\begin{bmatrix} \mathbb{1}_{2M} - \xi^{-1} & \tanh(t)\xi^{-1}Z \\ \tanh(t)ZX^{T}\xi^{-1} & \mathbb{1}_{2M} - X^{T}\xi^{-1}X \end{bmatrix}LW^{T}. \tag{D16}$$

Defining the matrix $F = \begin{bmatrix} \mathbb{1}_{2M} & 0_{2M} \\ 0_{2M} & z_{\text{tanh}(t)} \end{bmatrix}$, we can rewrite the last equation as

$$\mathbb{1}_{4M} - \left( \sigma + \frac{1}{2}\mathbb{1}_{4M} \right)^{-1} = WL^{T}F\begin{bmatrix} \mathbb{1}_{2M} - \xi^{-1} & \xi^{-1}X \\ X^{T}\xi^{-1} & \mathbb{1}_{2M} - X^{T}\xi^{-1}X \end{bmatrix}F^{T}LW^{T}, \tag{D17}$$

$$= E(t)R\begin{bmatrix} \mathbb{1}_{2M} - \xi^{-1} & \xi^{-1}X \\ X^{T}\xi^{-1} & \mathbb{1}_{2M} - X^{T}\xi^{-1}X \end{bmatrix}R^{T}E(t), \tag{D18}$$

where we noted that $WL^{T}F = WL^{*} = E(t)R$ (cf. Eq. (C18)). To arrive at the expression for $A_{\rho}$ we simply note $[E(t), P_{2M}] = 0$.

We would also like to find

$$b_{\rho} = (\sigma^{-1}_{+}\tilde{\mu})^{*} = \left( WLQ^{-1}LW^{T}\left[ \frac{1}{\sqrt{\hbar}}W_{\tilde{r}} \right] \right)^{*} = \frac{1}{\sqrt{\hbar}}(WLQ^{-1}L\tilde{r})^{*} = \frac{1}{\sqrt{\hbar}}(WL)^{*}Q^{-1}\begin{bmatrix} d \\ 0 \end{bmatrix}, \tag{D19}$$

$$= \frac{1}{\sqrt{\hbar}}(WL)^{*}\begin{bmatrix} \xi^{-1}d \\ -\tanh(t)ZX^{T}\xi^{-1}d \end{bmatrix} = \frac{1}{\sqrt{\hbar}}(WL)^{*}\begin{bmatrix} \xi^{-1}d \\ -X^{T}\xi^{-1}d \end{bmatrix} = \frac{1}{\sqrt{\hbar}}E(t)R^{*}\begin{bmatrix} \xi^{-1}d \\ -X^{T}\xi^{-1}d \end{bmatrix}. \tag{D20}$$

Finally, we can obtain

$$c_{\rho} = (\langle 0 | \otimes | 0 \rangle \rho (| 0 \rangle \otimes | 0 \rangle)) = \mathcal{N} \langle 0 | (\Phi | 0 \rangle | 0 \rangle) | 0 \rangle = \mathcal{N}c_{\Phi}. \tag{D21}$$

The Husimi covariance matrix of the state $\Phi | 0 \rangle | 0 \rangle$ is simply $\hbar\xi$ and its vector of means is $d$ and thus we can write

$$\langle 0 | (\Phi | 0 \rangle | 0 \rangle) | 0 \rangle = \frac{\exp\left[-\frac{1}{\hbar}d^{T}(\hbar\xi)^{-1}d\right]}{\sqrt{\det(\xi)}}. \tag{D22}$$

**Appendix E: Unitary Processes**

Now consider a unitary process. In this case we know that $Y = 0$ and that $X = S \in Sp_{2M}$ where $Sp_{2M}$ is the Symplectic group. Since $S$ is symplectic, then we can write a symplectic singular-value decomposition

$$S = \begin{bmatrix} \Re(U_{1}) & -\Im(U_{1}) \\ \Im(U_{1}) & \Re(U_{1}) \end{bmatrix} \begin{bmatrix} e^{-r} & 0_{M} \\ 0_{M} & e^{r} \end{bmatrix} \begin{bmatrix} \Re(U_{2}) & -\Im(U_{2}) \\ \Im(U_{2}) & \Re(U_{2}) \end{bmatrix} = O_{1}\lambda O_{2}, \tag{E1}$$
where \( U_1, U_2 \) are \( M \times M \) unitaries and \( r = \otimes_{i=1}^M r_i \) represents squeezing.

We can now calculate the Schur complement to find

\[
\xi = \frac{1}{2} (1_{2M} + SS^T), \quad (E2)
\]

\[
\xi^{-1} = 2O_1 \frac{1_{2M}}{1_{2M} + \lambda^2} O_1^T, \quad (E3)
\]

and then we find

\[
\begin{bmatrix}
1_{2M} - \xi^{-1} & \xi^{-1} X \\
X^T \xi^{-1} & 1_{2M} - X^T \xi^{-1} X
\end{bmatrix} = \begin{bmatrix}
O_1 \frac{\lambda^2 - 1_{2M}}{\lambda^2 + 1_{2M}} O_1^T & O_1 \frac{2\lambda}{\lambda^2 + 1_{2M}} O_2 \\
O_2^T \frac{2\lambda}{\lambda^2 + 1_{2M}} O_2^T & -O_2^T \frac{\lambda^2 - 1_{2M}}{\lambda^2 + 1_{2M}} O_2
\end{bmatrix}, \quad (E4)
\]

\[
= \begin{bmatrix}
O_1 & 0_{2M} \\
0_{2M} & O_2
\end{bmatrix} \begin{bmatrix}
\lambda^2 - 1_{2M} \\
\lambda^2 + 1_{2M}
\end{bmatrix} \begin{bmatrix}
O_1^T & 0_{2M} \\
0_{2M} & O_2
\end{bmatrix}. \quad (E5)
\]

Note that

\[
\frac{\lambda^2 - 1_{2M}}{\lambda^2 + 1_{2M}} = \begin{bmatrix}
-\tanh \lambda & 0_{M} \\
0_{M} & \tanh \lambda
\end{bmatrix}, \quad \frac{2\lambda}{\lambda^2 + 1_{2M}} = \begin{bmatrix}
\sech \lambda & 0_{M} \\
0_{M} & \sech \lambda
\end{bmatrix}. \quad (E6)
\]

We can now calculate

\[
R \begin{bmatrix}
1_{2M} - \xi^{-1} & \xi^{-1} X \\
X^T \xi^{-1} & 1_{2M} - X^T \xi^{-1} X
\end{bmatrix} R^\dagger = \begin{bmatrix}
O_1 & 0_{M} \\
0_{M} & O_2
\end{bmatrix} R^\dagger R \begin{bmatrix}
\lambda^2 - 1_{2M} \\
\lambda^2 + 1_{2M}
\end{bmatrix} R^\dagger \begin{bmatrix}
O_1^T & 0_{2M} \\
0_{2M} & O_2
\end{bmatrix} R^\dagger, \quad (E7)
\]

\[
= \begin{bmatrix}
U_1 & 0_{M} & 0_{M} \\
0_{M} & U_2^T & 0_{M} \\
0_{M} & U_1 & 0_{M}
\end{bmatrix} \begin{bmatrix}
0_{M} & 0_{M} & -\tanh \lambda \sech \lambda \\
0_{M} & 0_{M} & \sech \lambda \tanh \lambda \\
-\tanh \lambda \sech \lambda & \sech \lambda \tanh \lambda & 0_{M}
\end{bmatrix} \begin{bmatrix}
U_1 & 0_{M} & 0_{M} \\
0_{M} & U_2^T & 0_{M} \\
0_{M} & U_1 & 0_{M}
\end{bmatrix}
\]

\[
= -\begin{bmatrix}
0_{M} & A_U \\
A_U^* & 0_{M}
\end{bmatrix}, \quad (E8)
\]

where

\[
A_U = \begin{bmatrix}
U_1 & 0_{M} \\
0_{M} & U_2^T
\end{bmatrix} \begin{bmatrix}
\tanh \lambda & -\sech \lambda \\
-\sech \lambda & -\tanh \lambda
\end{bmatrix} \begin{bmatrix}
U_1 & 0_{M} \\
0_{M} & U_2^T
\end{bmatrix}^T = A_U^T. \quad (E9)
\]

Appendix F: Passive processes

Now consider the case of non-unitary passive process described by a transfer matrix \( T \), \( T^\dagger T \leq 1_M \). For this process

\[
X = \begin{bmatrix}
R(T) & \Im(T) \\
\Im(T) & R(T)
\end{bmatrix} \quad \text{and} \quad Y = \frac{1}{2} (1_{2M} - XX^T).
\]

Since the process is passive we know that \( \xi = 1_{2M} \). We can simplify the expression to obtain

\[
P_{2M} \left[ \frac{1_{4M}}{2} \right]^{-1} \otimes \left[ \sigma + \frac{1_{4M}}{2} \right]^{-1} = E(t) \begin{bmatrix}
0_{M} & T^* & 0_{M} & 0_{M} \\
T^\dagger & 0_{M} & 0_{M} & 1_M - T^\dagger T \\
0_{M} & 0_{M} & 0_{M} & T \\
0_{M} & 1_M - T^T T^* & T^T & 0_{M}
\end{bmatrix} E(t). \quad (F1)
\]
Following the Choi-Jamiołkowski relation we gave in Eq. (C11), the $A_\Phi$ for the lossy interferometer is
\[
A_\Phi = \begin{bmatrix}
0_M & T^* & 0_M & 0_M \\
T & 0_M & 0_M & 1_M - T^T T \\
0_M & 0_M & 0_M & T \\
0_M & 1_M - T^T T^* & T^T & 0_M
\end{bmatrix}.
\] (F2)

If we sandwich $A_\Phi$ with a permutation matrix $P_{4123}$, we would have:
\[
P_{4123} A_\Phi P_{4123}^T = \begin{bmatrix}
0_{2M} & 1_M - T^T T^* & T^T & 0_{2M} \\
T^* & 0_M & T & 0_{2M}
\end{bmatrix}.
\] (F3)

To get the probability with a measurement of photon number pattern $j = (j_1, \ldots, j_M)$ given the input $i = (i_1, \ldots, i_M)$, we let $k = i, l = j$ in (74):
\[
G_{4123}^{A_\Phi} (0) = G_{4123}^{P_{4123} A_\Phi P_{4123}^T} (0) = \frac{1}{i! j!} \text{haf}(P_{4123} A_\Phi P_{4123}^T)_{j \oplus i \oplus j \oplus i}
\] (F4)
\[
= \frac{1}{i! j! \text{perm}} \begin{bmatrix}
1_M - T^T T^* & T^T & 0_{2M}
T & 0_M
\end{bmatrix}_{j \oplus i}.
\] (F5)

**Appendix G: Global phase**

In this section, we are looking for the global phase when we apply two successive Gaussian operators.

The main idea of finding out this global phase is to rewrite the Husimi $Q$ function of these two Gaussian operators $\mathcal{G}_1(A_U, b_{U_1}, c_{U_1})$ and $\mathcal{G}_2(A_U, b_{U_2}, c_{U_2})$ as one single operator $\mathcal{G}_j(A_U, b_{U_j}, c_{U_j})$, and only the constant part $c_{U_j}$ is related to the global phase.

Eq. (23) of Ref. [17] shows that the Husimi $Q$ function of an arbitrary Gaussian unitary can be characterized by three quantities $C$, $\mu$, and $\Sigma$. As we already know the relation between $C$, $\mu$, $\Sigma$ and $c_{U_j}$, $b_{U_j}, A_U$ in the main text, now we will rewrite the Husimi $Q$ function for an arbitrary Gaussian unitary as:
\[
\langle \alpha^* | \mathcal{G} | \beta \rangle = \exp \left( -\frac{1}{2} \left[ ||\alpha||^2 + ||\beta||^2 \right] \right) c_{U} \exp \left( b_{U}^T \nu + \frac{1}{2} \nu^T A_U \nu \right),
\] (G1)

where
\[
\nu = \begin{bmatrix}
\alpha \\
\beta
\end{bmatrix}.
\] (G2)

So we first extend the Husimi $Q$ function of the two successive $M$-mode Gaussian operators by inserting an identity between them and using the completeness of the coherent states
\[
\frac{1}{\pi^M} \int d^{2M} \alpha \langle \alpha | \alpha \rangle = 1.
\] (G3)
to recombine them into two Husimi $Q$ functions:
\[
\langle \beta^* | \mathcal{G}_1 \mathcal{G}_2 | \beta' \rangle = \langle \beta^* | \mathcal{G}_1 I \mathcal{G}_2 | \beta' \rangle = \frac{1}{\pi^M} \int_{-\infty}^{+\infty} d^{2M} \alpha \langle \beta^* | \mathcal{G}_1 | \alpha \rangle \langle \alpha | \mathcal{G}_2 | \beta' \rangle,
\] (G4)
where \( d^{2M} = dR(\alpha) d\mathfrak{F}(\alpha) \).

And then we need to solve this integral of \( \alpha \):

\[
\frac{1}{\pi^M} \int_{-\infty}^{\infty} \frac{d^{2M} \alpha}{d\mathfrak{F}(\alpha)} \exp \left( -\frac{1}{2} [||\beta||^2 + 2||\alpha||^2 + ||\beta'||^2] \right) c_{U_1} c_{U_2} \exp \left( b_{U_1}^T \nu_1 + \frac{1}{2} \nu_1^T A_{U_1} \nu_1 + b_{U_2}^T \nu_2 + \frac{1}{2} \nu_2^T A_{U_2} \nu_2 \right). \tag{G5}
\]

where

\[
\nu_1 = [\beta, \alpha], \tag{G6}
\]

\[
\nu_2 = [\alpha^*, \beta^*]. \tag{G7}
\]

Let us calculate the integral and rewrite \( \langle \beta^* | G_1 G_2 | \beta' \rangle \) as \( \langle \beta^* | G_f | \beta' \rangle \), we will accordingly get the equation from \( c_{U_1}, c_{U_2}, b_{U_1}^T, b_{U_2}^T, A_{U_1}, A_{U_2} \) to \( c_{U_f}, b_{U_f}^T, A_{U_f} \), and the global phase will be contained in the term \( c_{U_f} \).

1. Integral \( \alpha \)

In order to integrate the \( \alpha \) with its real and imaginary parts, we introduce the following vectors because the integral of the multi-dimensional Gaussian expression is based on the real parameters:

\[
x_r^T = [\Re(\alpha), \Im(\alpha), \Re(\beta), \Im(\beta), \Re(\beta^*), \Im(\beta^*)], \tag{G8}
\]

\[
x_c^T = [\alpha, \alpha^*, \beta, \beta^*, \beta^*, \beta^*], \tag{G9}
\]

and it is clear to transfer from \( x_r \) to \( x_c \) with the matrix \( M_0 \):

\[
x_c = (M_0 \oplus M_0 \oplus M_0)x_r, \tag{G10}
\]

where

\[
M_0 = \begin{bmatrix}
1 & i1 \\
1 & -i1
\end{bmatrix}. \tag{G11}
\]

Next step, we replace all elements in Eq. (G5) in terms of \( x_c^T \).

So we have:

\[
||\alpha||^2 = \frac{1}{2} x_c^T M_2 x_c,
\]

\[
\nu_1 = M_3 x_c,
\]

\[
\nu_2 = M_4 x_c, \tag{G12}
\]

where

\[
M_2 = \begin{bmatrix}
0 & 1 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0
\end{bmatrix}, \tag{G13}
\]

\[
M_3 = \begin{bmatrix}
0 & 0 & 1 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0
\end{bmatrix}, \tag{G14}
\]

\[
M_4 = \begin{bmatrix}
0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0
\end{bmatrix}. \tag{G15}
\]
We can now rewrite Eq. (G5) as a function of $x_c$:

$$
\langle \beta^* | G_1 G_2 | \beta' \rangle = \frac{1}{\pi M} \int_{-\infty}^{+\infty} d^{2M} \alpha \exp \left(-\frac{1}{2} \left[ ||\beta||^2 + ||\beta'||^2 \right] \right) c_{U_1} c_{U_2} \exp \left( (b_{U_1}^T M_3 + b_{U_2}^T M_4) x_c + \frac{1}{2} x_c^T (M_3^T A_{U_1} M_3 + M_4^T A_{U_2} M_4 - M_2) x_c \right),
$$

\begin{equation}
= \frac{1}{\pi M} \int_{-\infty}^{+\infty} d^{2M} \alpha \exp \left(-\frac{1}{2} \left[ ||\beta||^2 + ||\beta'||^2 \right] \right) c_{U_1} c_{U_2} \exp \left( B_T x_c + \frac{1}{2} x_c^T A x_c \right),
\end{equation}

where

$$
B_T = b_{U_1}^T M_3 + b_{U_2}^T M_4,
\begin{bmatrix}
    d_1^T & c_1^T \\
    c_2^T & d_2^T
\end{bmatrix},
= [B_T', B_T^T].
$$

$$
A = M_3^T A_{U_1} M_3 + M_4^T A_{U_2} M_4 - M_2,
\begin{bmatrix}
    D_1 & C_1^T \\
    -C_1 & B_2
\end{bmatrix},
= \begin{bmatrix}
    A_1 & A_2 \\
    A_3 & A_4
\end{bmatrix}.
$$

and we write $b_{U_1}$ and $A_{U_1}$ in block:

$$
b_{U_1}^T = [c_{i1}^T, d_{i1}^T],
$$

$$
A_{U_1} = \begin{bmatrix}
    B_{i1} & C_{i1} \\
    C_{i1}^T & D_{i1}
\end{bmatrix}.
$$

It is obvious that $A$ is symmetric and $A_1$ and $A_4$ are also symmetric because of the symmetric of $A_{U_1}$ and $A_{U_2}$.

Now we want to separate out the components with $\alpha$ from $\exp (B_T x_c + \frac{1}{2} x_c^T A x_c)$ in order to calculate the integral.

$$
x_{rT} = \begin{bmatrix}
    \Re(\alpha) & \Im(\alpha) \\
    \Re(\beta) & \Im(\beta) & \Re(\beta') & \Im(\beta')
\end{bmatrix} = [x_{rl}, x_{rr}]^T.
$$

where $x_{rl}$ is the term related to $\alpha$. So that we have:

$$
x_c = (M_0 \oplus M_0 \oplus M_0) x_r = \begin{bmatrix}
    M_0 x_{rl} \\
    (M_0 \oplus M_0) x_{rr}
\end{bmatrix} = \begin{bmatrix}
    x_{cl} \\
    x_{cr}
\end{bmatrix}.
$$

Then we rewrite Eq. (G16) with $x_{cl}$ and $x_{cr}$:

$$
\langle \beta^* | G_1 G_2 | \beta' \rangle = \frac{1}{\pi M} \int_{-\infty}^{+\infty} d^{2M} \alpha \exp \left(-\frac{1}{2} \left[ ||\beta||^2 + ||\beta'||^2 \right] \right) c_{U_1} c_{U_2} \exp \left( B_{cl}^T x_{cl} + B_{cr}^T x_{cr} + \frac{1}{2} (x_{cl} A_{cl} x_{cl} + x_{cl} A_{2} x_{cr} + x_{cr} A_{3} x_{cl} + x_{cr} A_{4} x_{cr}) \right),
$$

because of the symmetry of $A$ and $x_{cl} = M_0 x_{rl}$, we have:

$$
\langle \beta^* | G_1 G_2 | \beta' \rangle = \exp \left(-\frac{1}{2} \left[ ||\beta||^2 + ||\beta'||^2 \right] \right) c_{U_1} c_{U_2} \exp \left( B_{cl}^T x_{cl} + \frac{1}{2} x_{cl}^T A_{3} x_{cl} \right) \cdot M_0 x_{rl} + \frac{1}{2} x_{rl}^T M_0^T A_{4} x_{rl}.
$$
We can integrate the multi-dimensional Gaussian expression with $x_r = [\Re(\alpha), \Im(\alpha)]:$
\[
\exp\left(-\frac{1}{2} \left[ ||\beta||^2 + ||\beta'||^2 \right] \right) c_{U_1} c_{U_2} \exp\left( B_r^T x_{cr} + \frac{1}{2} x_{cr}^T A_4 x_{cr} \right) * \frac{1}{\pi^M} \left( 2\pi \right)^M \sqrt{\det(A)} \exp\left(-\frac{1}{2} B'^T (A')^{-1} B' \right), \tag{G25}
\]
where
\[
B'^T = \left( B_r^T + x_{cr}^T A_3 \right) M_0,
\]
\[
A' = M_0^T A_1 M_0. \tag{G26}
\]

2. Derivation of $c_{U_f}, b_{U_f}$ and $A_{U_f}$

So that we have the Husimi Q function of two Gaussian unitaries:
\[
\langle \beta'| G_1 G_2 |\beta' \rangle = \exp\left(-\frac{1}{2} \left[ ||\beta||^2 + ||\beta'||^2 \right] \right) \frac{2^M c_{U_1} c_{U_2}}{\sqrt{\det(A)}^M} \exp\left( B_r^T x_{cr} + \frac{1}{2} x_{cr}^T A_4 x_{cr} - \frac{1}{2} \left( B_r^T + x_{cr}^T A_3 \right) A_1^{-1} \left( B_r^T + x_{cr}^T A_3 \right)^T \right),
\]
\[
= \exp\left(-\frac{1}{2} \left[ ||\beta||^2 + ||\beta'||^2 \right] \right) \frac{c_{U_1} c_{U_2}}{\sqrt{(-1)^M \det(A_1)}} \exp\left(-\frac{1}{2} B_r^T A_1^{-1} B_r + \left( B_r^T - B_r^T A_1^{-1} A_3^T \right) x_{cr} + \frac{1}{2} x_{cr}^T \left( A_4 - A_3 A_1^{-1} A_3^T \right) x_{cr} \right). \tag{G28}
\]

We need to notice that, because of the integral, we are working with $x_{cr} = (M_0 \oplus M_0)[\Re(\beta), \Im(\beta), \Re(\beta'), \Im(\beta')]$, however, we need to go back to the vector with complex expressions $\nu'^T = [\beta, \beta', \beta^*, \beta'^*]$. We have:
\[
\nu' = \begin{bmatrix}
1 & i & 0 & 0 \\
0 & 0 & 1 & i \\
1 & -i & 0 & 0 \\
0 & 0 & 1 & -i \\
\end{bmatrix} x_{rr} = M_5 x_{rr} = M_5 (M_0 \oplus M_0)^{-1} x_{cr}. \tag{G29}
\]

Then we rewrite the expression with $\nu'$:
\[
\langle \beta'| G_1 G_2 |\beta' \rangle = \exp\left(-\frac{1}{2} \left[ ||\beta||^2 + ||\beta'||^2 \right] \right) c_{U_f} \exp(b_{U_f}^T \nu' + \frac{1}{2} \nu'^T A_{U_f}^{-1} \nu'), \tag{G30}
\]
where
\[
c_{U_f} = \frac{c_{U_1} c_{U_2}}{\sqrt{(-1)^M \det(A_1)}} \exp\left(-\frac{1}{2} B_r^T A_1^{-1} B_r \right),
\]
\[
b_{U_f}^T = (B_r^T - B_r^T A_1^{-1} A_3^T)(M_0 \oplus M_0)M_5^{-1},
\]
\[
A_{U_f}' = (M_5^{-1})^T(M_0 \oplus M_0)^T(A_4 - A_3 A_1^{-1} A_3^T)(M_0 \oplus M_0)M_5^{-1}. \tag{G31}
\]

In order to get the explicit formula, we compute:
\[
(M_0 \oplus M_0)M_5^{-1} = \begin{bmatrix}
1 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & 0 & 1 \\
\end{bmatrix}, \tag{G32}
\]
and by using the Schur complement, since $A_1$ is symmetric, $D_1$ and $B_2$ are both symmetric, we can define the inverse of $A_1$:
\[
A_1^{-1} = \begin{bmatrix}
D_1 & X^{-1} & 0 \\
X^{-1} & D_1 & 0 \\
0 & 0 & I \\
\end{bmatrix}, \tag{G33}
\]
\[ \mathbf{X} = \mathbf{B}_2 \mathbf{D}_1 - \mathbf{1}. \]  
\[ \text{(G34)} \]

We also have the following:
\[ \det(\mathbf{A}_1) = \det(\mathbf{D}_1 \mathbf{B}_2 - \mathbf{1}). \]  
\[ \text{(G35)} \]

Then we can get:
\[
c_{\mathbf{U}_f} = \frac{c_{\mathbf{U}_1} c_{\mathbf{U}_2}}{\sqrt{(-1)^M \det(\mathbf{D}_1 \mathbf{B}_2 - \mathbf{1})}} \exp \left[ \frac{1}{2} \left( \mathbf{d}_1^T \mathbf{D}_1 \mathbf{X} \mathbf{X}^{-1} \mathbf{d}_1 + c_2^T (\mathbf{X}^T)^{-1} \mathbf{c}_1 + c_2^T \mathbf{D}_1 \mathbf{X} \mathbf{X}^{-1} \mathbf{c}_2 \right) \right],
\]
\[ \text{(G36)} \]

\[
b_{\mathbf{U}_f} = \begin{bmatrix} c_1 - \mathbf{C}_1 \mathbf{D}_1 \mathbf{X} \mathbf{X}^{-1} \mathbf{d}_1 - \mathbf{C}_1 (\mathbf{X}^T)^{-1} \mathbf{c}_2 \\ \mathbf{d}_2 - \mathbf{C}_2^T \mathbf{X} \mathbf{X}^{-1} \mathbf{d}_1 - \mathbf{C}_2 \mathbf{D}_1 (\mathbf{X}^T)^{-1} \mathbf{c}_2 \\ 0_M \\ 0_M \end{bmatrix} = \begin{bmatrix} b_{\mathbf{U}_f} \\ 0_{2M} \end{bmatrix},
\]
\[ \text{(G37)} \]

\[
\mathbf{A}_{\mathbf{U}_f} = \begin{bmatrix} \mathbf{B}_1 - \mathbf{C}_1 \mathbf{D}_1 \mathbf{X} \mathbf{X}^{-1} \mathbf{C}_1^T \\ -\mathbf{C}_2 (\mathbf{X}^T)^{-1} \mathbf{B}_1^T \mathbf{D}_2 - \mathbf{C}_2^T \mathbf{D}_1 (\mathbf{X}^T)^{-1} \mathbf{C}_2 \\ 0_M \\ 0_M \end{bmatrix} = \begin{bmatrix} \mathbf{A}_{\mathbf{U}_f} \\ 0_{2M} \end{bmatrix}.
\]
\[ \text{(G38)} \]

So that finally, we can write:
\[
\langle \mathbf{\beta}^* | \mathcal{G}_1 \mathcal{G}_2 | \mathbf{\beta} \rangle = \exp \left( -\frac{1}{2} \left[ ||\mathbf{\beta}||^2 + ||\mathbf{\beta}'||^2 \right] \right) c_{\mathbf{U}_f} \exp \left( b_{\mathbf{U}_f}^T \mathbf{\nu} + \frac{1}{2} \mathbf{\nu}^T \mathbf{A}_{\mathbf{U}_f} \mathbf{\nu} \right).
\]
\[ \text{(G39)} \]

\[ c_{\mathbf{U}_f} \] is what we want and it contains the global phase of the composite Gaussian operator.