Notes on Gaussian states and their representation as neural networks
Ph. D. lectures on quantum machine learning (Trento)

Claudio Conti[1, 2, 3, *]

[1]Institute for Complex Systems, National Research Council (ISC-CNR), Via dei Taurini 19, 00185 Rome, Italy
[2]Department of Physics, University Sapienza, P.le Aldo Moro 5, 00185 Rome, Italy
[3]Research Center Enrico Fermi (CREF), Via Panisperna 89a, 00184 Rome, Italy [†]

(Dated: April 12, 2022)

I. PHASE SPACE REPRESENTATION

II. INTRODUCTION

In the phase space, a state is represented by a function of several variables, which we encode in a real vector $\mathbf{x}$. For a n-body system, $\mathbf{x}$ is a vector of $N = 2\,n$ real variables.

Specifically, we consider the characteristic function $\chi(\mathbf{x})$.

For Gaussian states [1], in symmetric ordering,

$$\chi(\mathbf{x}) = \exp\left(-\frac{1}{4}\mathbf{x}\cdot\boldsymbol{g}\cdot\mathbf{x} + \imath\,\mathbf{x}\cdot\boldsymbol{d}\right). \tag{1}$$

with $\boldsymbol{g}$ the $N\times N$ covariance matrix, and $\boldsymbol{d}$ the displacement vector. Gaussian states include vacuum, thermal, coherent, and squeezed states.

$\chi(\mathbf{x})$ is a complex functions with real and imaginary part, i.e.,

$$\chi(\mathbf{x}) = \chi_R(\mathbf{x}) + \imath\chi_I(\mathbf{x}), \tag{2}$$

As neural network models typically deal with real-valued quantities, in the following we will represent the characteristic functions by models with two real outputs corresponding to χ_R and χ_I. Any many-body state is represented by a couple of real functions $\chi_{R,I}$ of N real variables. A noteworthy link between the phase space representation and neural networks, is that neural networks can approximate arbitrary functions, hence they may approximate characteristic functions. As an outcome, the resulting model includes linear transformations representing gates, as displacements or interferometers, followed by a nonlinear activation, which corresponds to computing the characteristic functions (e.g, a Gaussian function); this maps the quantum state in a conventional neural network architecture that can be trained by well-known algorithms.

III. THE CHARACTERISTIC FUNCTION AND OPERATOR ORDERING

We introduce the complex vector $\boldsymbol{z}$ with components

$$\boldsymbol{z} = (z_0, z_0, \ldots, z_{n-1}, z_0^*, z_1^*, \ldots, z_{n-1}^*), \tag{3}$$

which includes the complex z_j components and their conjugate z_j^*. Given the density matrix ρ, the characteristic function is [2]

$$\chi(\boldsymbol{z}, \mathcal{P}) = \mathrm{Tr}\left\{\rho\exp\left[\sum_k\left(z_k\hat{a}_k^\dagger - z_k^*\hat{a}_k\right)\right]\right\}\exp\left(\sum_k\frac{\mathcal{P}}{2}z_k z_k^*\right) \tag{4}$$

In Eq. (4) the variable $\mathcal{P} = 0, 1, -1$, refers to the adopted operator ordering.

The characteristic function depends on $\boldsymbol{z}$ and the ordering index $\mathcal{P}$. $\mathcal{P}$ is important when determining the expected value of operators as derivatives of χ.

---

[*] claudio.conti@uniroma1.it

[†] https://www.complexlight.org

The expected value of an observable $\hat{O}$ is

$$\langle \hat{O} \rangle = \mathrm{Tr}\left(\rho \hat{O}\right). \tag{5}$$

The expected value of combinations of the annihilation and creation operators are

$$\langle \left(\hat{a}_j^\dagger\right)^m (\hat{a}_k)^n \rangle_{\mathcal{P}} = \left(\frac{\partial}{\partial z_j}\right)^m \left(-\frac{\partial}{\partial z_k^*}\right)^n \chi(\boldsymbol{z}, \mathcal{P})\Big|_{\boldsymbol{z}=0} \tag{6}$$

One realizes that the exponential term in (4) weighted by $\mathcal{P}$ changes the derivatives, and we have different expected values when varying $\mathcal{P}$.

The simplest one is the mean value of the field operator

$$\langle \hat{a}_k \rangle = -\frac{\partial \chi}{\partial z_k^*}, \tag{7}$$

which does not depend on the ordering index $\mathcal{P}$, as the mean value of

$$\langle \hat{a}_k^\dagger \rangle = \frac{\partial \chi}{\partial z_k}. \tag{8}$$

On the other hand for the mean particle number of the mode k, $\hat{n}_k = \hat{a}_k^\dagger \hat{a}_k$, we have

$$\langle \hat{a}_k^\dagger \hat{a}_k \rangle_{\mathcal{P}} = -\frac{\partial^2 \chi(\boldsymbol{z}, \mathcal{P})}{\partial z_k \partial z_k^*}\Big|_{\boldsymbol{z}=0} \tag{9}$$

For the symmetric ordering $\mathcal{P} = 0$, one has

$$\langle \hat{a}_k^\dagger \hat{a}_k \rangle_{\mathcal{P}=0} = \langle \hat{a}_k^\dagger \hat{a}_k + \hat{a}_k \hat{a}_k^\dagger \rangle; \tag{10}$$

for the normal ordering $\mathcal{P} = 1$, one has

$$\langle \hat{a}_k^\dagger \hat{a}_k \rangle_{\mathcal{P}=1} = \langle \hat{a}_k^\dagger \hat{a}_k \rangle, \tag{11}$$

finally for antinormal ordering $\mathcal{P} = -1$,

$$\langle \hat{a}_k^\dagger \hat{a}_k \rangle_{\mathcal{P}=-1} = \hat{\langle} a_k a_k^\dagger \rangle \tag{12}$$

In general, we have

$$\langle \hat{a}_k^\dagger \hat{a}_k \rangle_{\mathcal{P}} = \langle \hat{a}_k^\dagger \hat{a}_k \rangle + \frac{1}{2}(1 - \mathcal{P}), \tag{13}$$

which will be adopted below to write the neural network layer that returns the mean particle number. When not explicitly stated, we refer to symmetric ordering $\mathcal{P} = 0$.

IV. THE CHARACTERISTIC FUNCTION IN TERMS OF REAL VARIABLES

For use with the machine learning application programming interfaces as `TensorFlow`, it is convenient to use real variables. We consider the quadrature operators $\hat{x}_k$ and $\hat{p}_k$

$$\hat{a}_k = \frac{\hat{q}_k + \imath \hat{p}_k^\dagger}{\sqrt{2}}. \tag{14}$$

The quadrature operators can be cast in operator vectors $\hat{\boldsymbol{q}}$ and $\hat{\boldsymbol{p}}$ with dimension $n \times 1$,

$$\hat{\boldsymbol{q}} = \begin{pmatrix} \hat{q}_0 \\ \hat{q}_1 \\ \vdots \\ \hat{q}_{n-1} \end{pmatrix} \qquad \hat{\boldsymbol{p}} = \begin{pmatrix} \hat{p}_0 \\ \hat{p}_1 \\ \vdots \\ \hat{p}_{n-1} \end{pmatrix}. \tag{15}$$

In addition, we consider the real quantities

$$q_k = \frac{z_k - z_k^*}{\sqrt{2}\imath} \qquad p_k = \frac{z_k + z_k^*}{\sqrt{2}}, \tag{16}$$

and we collect them in two real vectors $\boldsymbol{q}$ and $\boldsymbol{p}$ with dimension $1 \times n$

$$\begin{aligned} \boldsymbol{q} &= \begin{pmatrix} q_0 & q_1 & \cdots & q_{n-1} \end{pmatrix} \\ \boldsymbol{p} &= \begin{pmatrix} p_0 & p_1 & \cdots & p_{n-1} \end{pmatrix}. \end{aligned} \tag{17}$$

Correspondingly, we have

$$\chi = \chi(\boldsymbol{q}, \boldsymbol{p}, \mathcal{P} = 0) = \text{Tr}[\rho \exp(\imath\, \boldsymbol{q} \cdot \hat{\boldsymbol{q}} + \imath\, \boldsymbol{p} \cdot \hat{\boldsymbol{p}})] \tag{18}$$

As $\hat{\boldsymbol{q}}$ and $\hat{\boldsymbol{p}}$ are column vectors, and $\boldsymbol{q}$ and $\boldsymbol{p}$ are row vectors, we omit the dot product symbol and write

$$\chi(\boldsymbol{q}, \boldsymbol{p}) = \text{Tr}[\rho \exp(\imath\, \boldsymbol{q}\, \hat{\boldsymbol{q}} + \imath\, \boldsymbol{p}\, \hat{\boldsymbol{p}})]\,. \tag{19}$$

The notation can be simplified by defining the $1 \times N$ real row vector

$$\mathbf{x} = \begin{pmatrix} q_0 & p_0 & q_1 & p_2 & \cdots q_n & p_n \end{pmatrix} \tag{20}$$

and we have

$$\chi = \chi(\mathbf{x}) = \text{Tr}\left[\rho \exp\left(\imath\, \mathbf{x}\, \hat{\mathbf{R}}\right)\right] = \chi_R(\mathbf{x}) + \imath \chi_I(\mathbf{x}) \tag{21}$$

being

$$\mathbf{x}\, \hat{\mathbf{R}} = q_0 \hat{q}_0 + p_0 \hat{p}_0 + q_1 \hat{q}_1 + \ldots + q_{n-1} \hat{q}_{n-1} + p_{n-1} \hat{p}_{n-1} = x_0 \hat{R}_0 + x_1 \hat{R}_1 + \ldots + x_{N-1} \hat{R}_{N-1} \tag{22}$$

with the column vector $N \times 1$ of operators

$$\hat{\mathbf{R}} = \begin{pmatrix} \hat{q}_0 \\ \hat{p}_0 \\ \hat{q}_1 \\ \hat{p}_1 \\ \vdots \\ \hat{q}_{n-1} \\ \hat{p}_{n-1} \end{pmatrix}. \tag{23}$$

Following the canonical commutation relation of the quadratures, in units with $\hbar = 1$,

$$\begin{aligned} [\hat{q}_j, \hat{q}_k] &= 0 \\ [\hat{p}_j, \hat{p}_k] &= 0 \\ [\hat{q}_j, \hat{p}_k] &= i \delta_{jk} \end{aligned} \tag{24}$$

we have

$$[\hat{R}_p, \hat{R}_q] = \imath J_{pq}, \tag{25}$$

where we introduced the $N \times N$ symplectic matrix $\mathbf{J}$

$$\mathbf{J} = \bigoplus_j \mathbf{J}_1 = \begin{pmatrix} 0 & 1 & 0 & 0 & \cdots & 0 & 0 \\ -1 & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & 1 & \cdots & 0 & 0 \\ 0 & 0 & -1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 0 & 1 \\ 0 & 0 & 0 & 0 & \cdots & -1 & 0 \end{pmatrix}, \tag{26}$$

being $\mathbf{J}_1 = \left(\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right)$ [1].In Eq. (26), the symbol $\bigoplus$ is the direct sum for block matrices with j running in $0, 1, \dots n-1$. The matrix $\mathbf{J}$ is such that $\mathbf{J}^{-1} = \mathbf{J}^\top$, and $\mathbf{J}^2 = -\mathbf{I}_N$ and $\mathbf{J}^\top \mathbf{J} = \mathbf{1}_N$ is the identity $N \times N$ matrix.

The expected value of $\hat{\boldsymbol{R}}$ is determined by the derivatives of the characteristic function. By using the definition of $\chi(\mathbf{x})$

$$\chi(\mathbf{x}) = \mathrm{Tr}\left(\rho e^{\imath\, \mathbf{x}\, \hat{\mathbf{R}}}\right), \tag{27}$$

one has

$$\langle \hat{R}_j \rangle = \mathrm{Tr}[\rho \hat{R}_j] = -\imath \left.\frac{\partial \chi}{\partial x_j}\right|_{\mathbf{x}=0} = \left.\frac{\partial \chi_I}{\partial x_j}\right|_{\mathbf{x}=0} \tag{28}$$

or

$$\langle \hat{\mathbf{R}} \rangle = \mathrm{Tr}[\rho \hat{\mathbf{R}}] = -\imath \left.\nabla \chi\right|_{\mathbf{x}=0} = \left.\nabla \chi_I\right|_{\mathbf{x}=0} . \tag{29}$$

As the components of $\hat{\mathbf{R}}$ are self-adjoint, their quantum expected is real, and it is given by the derivative of the imaginary part χ_I at $\mathbf{x} = 0$.

Seemingly, we have for the mean value of the number of particles in the mode j

$$\langle \hat{a}_j^\dagger \hat{a}_j \rangle_{\mathcal{P}=0} = -\left.\frac{\partial^2}{\partial z_j \partial z_j^*} \chi\right|_{\mathbf{x}=0} . \tag{30}$$

Expressing χ in terms of the real components $\boldsymbol{q}$ and $\boldsymbol{p}$, we have for the derivatives

$$\begin{gathered} -\frac{\partial^2}{\partial z_j \partial z_j^*}\chi = \\ -\left(\frac{\partial}{\partial q_j}\frac{\partial q_j}{\partial z_j} + \frac{\partial}{\partial p_j}\frac{\partial p_j}{\partial z_j}\right)\left(\frac{\partial}{\partial q_j}\frac{\partial q_j}{\partial z_j^*} + \frac{\partial}{\partial p_j}\frac{\partial p_j}{\partial z_j^*}\right)\chi = -\frac{1}{2}\left(\frac{\partial^2}{\partial q_j^2} + \frac{\partial^2}{\partial p_j^2}\right)\chi, \end{gathered} \tag{31}$$

after using (16).

Hence we may evaluate the expected number of photons by the second derivatives of χ_R, because its imaginary part is vanishing

$$\langle \hat{a}_j^\dagger \hat{a}_j \rangle_{\mathcal{P}=0} = -\frac{1}{2}\left.\left(\frac{\partial^2}{\partial x_{2j}^2} + \frac{\partial^2}{\partial x_{2j+1}^2}\right)\chi_R\right|_{\mathbf{x}=0} \tag{32}$$

with $j = 0, 1, 2, \dots, n-1$.

For the expected value of total number of particles $\mathcal{N}$, we hence have

$$\langle \hat{\mathcal{N}} \rangle_{\mathcal{P}=0} = -\frac{1}{2}\left.\nabla^2 \chi_R\right|_{\mathbf{x}=0} \tag{33}$$

the Laplacian being computed with respect to all the components of $\mathbf{x}$.

Equation (33) refer to symmetric ordering with $\mathcal{P} = 0$, for normal ordering, we use Eq. (13) as

$$\langle \hat{a}_k^\dagger \hat{a}_k \rangle = \langle \hat{a}_k^\dagger \hat{a}_k \rangle_{\mathcal{P}=1} = \langle \hat{a}_k^\dagger \hat{a}_k \rangle_{\mathcal{P}=0} - \frac{1}{2} \tag{34}$$

in (33) and we get

$$\langle \mathcal{N} \rangle = \langle \mathcal{N} \rangle_{\mathcal{P}=1} = \langle \mathcal{N} \rangle_{\mathcal{P}=0} - \frac{n}{2} = -\frac{1}{2}\left.\nabla^2 \chi_R\right|_{\mathbf{x}=0} - \frac{n}{2} = -\frac{1}{2}\left.\left(\nabla^2 + n\right)\chi\right|_{\mathbf{x}=0} \tag{35}$$

V. GAUSSIAN STATES

For a Gaussian state, the characteristic function in symmetric ordering is

$$\chi(\mathbf{x}) = \exp(-\frac{1}{4}\mathbf{x}\,\boldsymbol{g}\,\mathbf{x}^\top + \imath\,\mathbf{x}\,\boldsymbol{d}) \tag{36}$$

with $\boldsymbol{g}$ the $N\times N$ covariance matrix, and $\boldsymbol{d}$ the $N\times 1$ displacement column vector. Correspondingly, being $\mathbf{x}$ a row vector,

$$\begin{aligned}\chi_R(\mathbf{x}) &= e^{-\frac{\mathbf{x}\boldsymbol{g}\mathbf{x}^\top}{4}}\cos(\mathbf{x}\boldsymbol{d})\\ \chi_I(\mathbf{x}) &= e^{-\frac{\mathbf{x}\boldsymbol{g}\mathbf{x}^\top}{4}}\sin(\mathbf{x}\boldsymbol{d})\end{aligned} \tag{37}$$

From Eq. (29), we have for $\boldsymbol{d}$, with $q = 0, 1, \ldots, N-1$.

$$\langle\hat{R}_q\rangle = \mathrm{Tr}[\hat{\rho}\hat{R}_q] = d_q = \left.\frac{\partial\chi_I}{\partial x_q}\right|_{\mathbf{x}=0} \tag{38}$$

or

$$\langle\mathbf{R}\rangle = \boldsymbol{d} = \left.\nabla\chi_I\right|_{\mathbf{x}=0} \tag{39}$$

Hence, the displacement d_q is the first moment of the Gaussian χ.

In general, the derivatives of the characteristic function can be expressed by the displacement and covariance matrix d_q and g_{pq}. This shows that a Gaussian states is determined by the displacements and the elements g_{pq}.

Also, we have an advantage in speeding up computing the derivatives of χ, and the observable quantities.

Indeed, for the particle number, we have

$$\begin{aligned}\langle\hat{a}_j^\dagger\hat{a}_j\rangle_{\mathcal{P}=0} &= -\frac{1}{2}\left.\left(\frac{\partial^2}{\partial x_{2j}^2} + \frac{\partial^2}{\partial x_{2j+1}^2}\right)\right|_{\mathbf{x}=0}\chi_R\\ &= \frac{g_{2j,2j} + g_{2j+1,2j+1}}{4} + \frac{d_{2j}^2 + d_{2j+1}^2}{2}\end{aligned} \tag{40}$$

and

$$\begin{aligned}\langle\hat{a}_j^\dagger\hat{a}_j\rangle &= -\frac{1}{2}\left(\frac{\partial^2}{\partial x_{2j}^2} + \frac{\partial^2}{\partial x_{2j+1}^2}\right)\chi_R - \frac{1}{2}\\ &= \frac{g_{2j,2j} + g_{2j+1,2j+1}}{4} + \frac{d_{2j}^2 + d_{2j+1}^2}{2} - \frac{1}{2}\end{aligned} \tag{41}$$

while the second derivatives of χ_I at $\mathbf{x} = 0$ are vanishing, and we used (13).

The covariance matrix $\boldsymbol{g}$ is determined by the second moments

$$g_{pq} = \langle\{\hat{R}_p, \hat{R}_q\}\rangle - 2\langle\hat{R}_p\rangle\langle\hat{R}_q\rangle \tag{42}$$

with

$$\{\hat{R}_p, \hat{R}_q\} = \hat{R}_p\hat{R}_q + \hat{R}_q\hat{R}_p. \tag{43}$$

Eq. (42) shows that $\boldsymbol{g}$ is a real symmetric matrix as obtained by expected values of self-adjoint operators.

We also have after Eq. (25)

$$\begin{aligned}g_{pq} &= \langle\hat{R}_p\hat{R}_q\rangle + \langle\hat{R}_q\hat{R}_p\rangle - 2\langle R_p\rangle\langle R_q\rangle\\ &= \langle\hat{R}_p\hat{R}_q\rangle + \langle\hat{R}_q\hat{R}_p\rangle - 2\langle R_p\rangle\langle R_q\rangle\\ &= 2\langle\hat{R}_p\hat{R}_q\rangle - \imath J_{pq} - 2\langle R_p\rangle\langle R_q\rangle =\\ &= 2\langle\left(\hat{R}_p - d_p\right)\left(\hat{R}_q - d_q\right)\rangle - \imath J_{pq}\end{aligned} \tag{44}$$

A. Vacuum state

Vacuum states have $\boldsymbol{d} = \mathbf{0}_{N\times 1}$ and $\boldsymbol{g} = \mathbf{1}_{N\times N}$ the $N \times N$ identity matrix. Using these expression in (41), we get

$$\langle \hat{a}_j^\dagger \hat{a}_j \rangle_{\mathcal{P}=0} = \frac{1}{2}\ . \tag{45}$$

Correspondingly,

$$\langle \hat{a}_j^\dagger \hat{a}_j \rangle = \langle \hat{a}_j^\dagger \hat{a}_j \rangle_{\mathcal{P}=0} - \frac{1}{2} = 0. \tag{46}$$

We have $\langle \hat{\mathbf{R}} \rangle = 0$ and for the average photon number after Eq. (41)

$$\langle \mathcal{N} \rangle = \sum_{j=0}^{n-1} \langle \hat{a}_j^\dagger \hat{a}_j \rangle = 0 \tag{47}$$

B. Coherent state

A coherent state has a non-vanishing displacement vector $\boldsymbol{d} \neq \mathbf{0}$ and $\boldsymbol{g} = \mathbf{1}_N$. If we consider a single mode $|\alpha\rangle$, with $n = 1$ and $N = 2$,

$$\begin{aligned} d_0 &= \sqrt{2}\Re\left(\alpha\right) = \frac{\alpha + \alpha^*}{\sqrt{2}} \\ d_1 &= \sqrt{2}\Im\left(\alpha\right) = \frac{\alpha - \alpha^*}{\imath\sqrt{2}} \end{aligned} \tag{48}$$

From (41), we have ($j = 0$)

$$\langle \hat{a}_j^\dagger \hat{a}_j \rangle = \langle \hat{a}_j^\dagger \hat{a}_j \rangle_{\mathcal{P}=0} - \frac{1}{2} = |\alpha|^2 = \frac{d_0^2 + d_1^2}{2}. \tag{49}$$

For n distinct modes, $j = 0, 1, \ldots, n-1$ we have

$$\langle \hat{a}_j^\dagger \hat{a}_j \rangle = \frac{d_{2j}^2 + d_{2j+1}^2}{2}, \tag{50}$$

and

$$\langle \mathcal{N} \rangle = \sum_{j=0}^{n-1} \langle \hat{a}_j^\dagger \hat{a}_j \rangle = \sum_{q=0}^{N-1} \frac{d_q^2}{2}. \tag{51}$$

VI. COVARIANCE MATRIX IN TERMS OF THE DERIVATIVES OF χ

Given a generic χ, one can compute the covariance matrix and the displacement operator by derivation. The covariance matrix is defined by (42), here reported as

$$\begin{aligned} g_{pq} =& \\ &\langle \{\hat{R}_p, \hat{R}_q\} \rangle - 2\langle \hat{R}_p \rangle \langle \hat{R}_q \rangle = \\ &= \langle \hat{R}_p \hat{R}_q \rangle + \langle \hat{R}_q \hat{R}_p \rangle - 2\langle R_p \rangle \langle R_q \rangle \\ &2\mathrm{Tr}[\rho(\hat{R}_p - d_p)(\hat{R}_q - d_q)] - \imath J_{pq} \end{aligned} \tag{52}$$

and seemingly the displacement vector

$$d_p = \langle \hat{R}_p \rangle\ . \tag{53}$$

$\boldsymbol{d}$ can be computed by the first derivatives, as in Eq. (29) here written as

$$d_q = \langle \hat{R}_q \rangle = \mathrm{Tr}[\hat{\rho}\hat{R}_q] = \left.\frac{\partial \chi_I}{\partial x_q}\right|_{\mathbf{x}=0} . \tag{54}$$

Seemingly, the covariance matrix is given by

$$g_{pq} = -2 \left.\frac{\partial^2 \chi_R}{\partial x_p \partial x_q}\right|_{\mathbf{x}=0} - 2 d_p d_q = -2 \frac{\partial^2 \chi_R}{\partial x_p \partial x_q} - 2 \left.\frac{\partial \chi_I}{\partial x_p}\frac{\partial \chi_I}{\partial x_q}\right|_{\mathbf{x}=0} \tag{55}$$

These expressions, proven below, are not limited to Gaussian states. They can be used to compute $\boldsymbol{g}$ and $\boldsymbol{d}$ from a NN model. The special case of a Gaussian state is reported in section (VI B).

A. *Proof of Eq. (54) and (55) for a generic state

Proof. We consider the definition of the characteristic function in Eq. (27) here reported

$$\chi(\mathbf{x}) = \mathrm{Tr}\left(\rho e^{\imath\, \mathbf{x}\hat{\mathbf{R}}}\right), \tag{56}$$

We first consider the first moment, by deriving Eq. 56 w.r.t. x_q we have

$$\frac{\partial \chi}{\partial x_q} = \mathrm{Tr}\left(\imath \hat{R}_q \rho e^{\imath\, \mathbf{x}\hat{\mathbf{R}}}\right), \tag{57}$$

which gives at $\mathbf{x} = 0$

$$\left.\frac{\partial \chi}{\partial x_q}\right|_{\mathbf{x}=0} = \imath \mathrm{Tr}\left(\hat{R}_q \rho\right) = \imath \langle \hat{R}_q \rangle, \tag{58}$$

as $\chi = \chi_R + \imath \chi$ and $\langle \hat{R}_q \rangle$ is real because $\hat{R}_q$ is self-adjoint, we obtain the following

$$\begin{aligned} \langle \hat{R}_q \rangle &= -\imath \left.\frac{\partial \chi}{\partial x_q}\right|_{\mathbf{x}=0} = \left.\frac{\partial \chi_I}{\partial x_q}\right|_{\mathbf{x}=0} \\ \left.\frac{\partial \chi_R}{\partial x_q}\right|_{\mathbf{x}=0} &= 0 \,. \end{aligned} \tag{59}$$

For the second derivatives, we write

$$\begin{aligned} \langle \hat{\mathrm{R}}_p \hat{\mathrm{R}}_q \rangle = \mathrm{Tr}\left(\rho \hat{\mathrm{R}}_p \hat{\mathrm{R}}_q\right) &= \mathrm{Tr}\left.\left[\rho \frac{\partial e^{\imath\, x_p \hat{\mathrm{R}}_p}}{\partial(\imath x_p)} \frac{\partial e^{\imath\, x_q \hat{\mathrm{R}}_q}}{\partial(\imath x_q)}\right]\right|_{\mathbf{x}=0} = \\ &- \frac{\partial^2}{\partial x_p \partial x_q} \mathrm{Tr}\left.\left(\rho e^{\imath\, x_p \hat{\mathrm{R}}_p} e^{\imath\, x_q \hat{\mathrm{R}}_q}\right)\right|_{\mathbf{x}=0} \end{aligned} \tag{60}$$

Then we have

$$e^{\imath\, x_p \hat{\mathrm{R}}_p} e^{\imath\, x_q \hat{\mathrm{R}}_q} = e^{\imath\, x_q \hat{\mathrm{R}}_q + \imath\, x_q \hat{\mathrm{R}}_q} e^{-\frac{1}{2}\left[\hat{\mathrm{R}}_p, \hat{\mathrm{R}}_q\right] x_p x_q} = e^{\imath\, x_q \hat{\mathrm{R}}_q + \imath\, x_q \hat{\mathrm{R}}_q} e^{-\frac{\imath}{2} J_{pq} x_p x_q}, \tag{61}$$

which used in (60)

$$\begin{aligned} \langle \hat{\mathrm{R}}_p \hat{\mathrm{R}}_q \rangle = &- \frac{\partial^2}{\partial x_p \partial x_q} \mathrm{Tr}\left.\left(\rho e^{\imath\, x_p \hat{\mathrm{R}}_p} e^{\imath\, x_q \hat{\mathrm{R}}_q}\right)\right|_{\mathbf{x}=0} = \\ &- \frac{\partial^2}{\partial x_p \partial x_q} \mathrm{Tr}\left.\left(\rho e^{\imath\, x_q \hat{\mathrm{R}}_q + \imath\, x_q \hat{\mathrm{R}}_q} e^{-\frac{\imath}{2} J_{pq} x_p x_q}\right)\right|_{\mathbf{x}=0} = \\ &- \frac{\partial^2}{\partial x_p \partial x_q} \mathrm{Tr}\left.\left(\rho e^{\imath\, \mathbf{x}\hat{\boldsymbol{R}}} e^{-\frac{\imath}{2} J_{pq} x_p x_q}\right)\right|_{\mathbf{x}=0} = \\ &- \frac{\partial^2}{\partial x_p \partial x_q} \left.\left[\chi(\mathbf{x}) e^{-\frac{\imath}{2} J_{pq} x_p x_q}\right]\right|_{\mathbf{x}=0} = \\ &- \left.\frac{\partial^2}{\partial x_p \partial x_q} \chi(\mathbf{x})\right|_{\mathbf{x}=0} + \frac{\imath}{2} J_{pq} \end{aligned} \tag{62}$$

Using the previous result in (see Eq. 52)

$$g_{pq} = \langle \hat{R}_p \hat{R}_q \rangle + \langle \hat{R}_q \hat{R}_p \rangle - 2\langle R_p \rangle \langle R_q \rangle \tag{63}$$

we have ($J_{pq} = -J_{qp}$)

$$g_{pq} = -2 \left. \frac{\partial^2}{\partial x_p \partial x_q} \chi(\mathbf{x}) \right|_{\mathbf{x}=0} - 2 d_p d_q, \tag{64}$$

and taking into account that g_{pq} and d_p are real, we have the proof. □

B. *Proof of Eq. (54) and (55) for a Gaussian state

Proof. By writing the Gaussian characteristic function in terms of the components of $\mathbf{x}$

$$\chi(\mathbf{x}) = \exp\left(-\frac{1}{4} \sum_{pq} g_{pq} x_p x_q + \imath \sum_p x_p d_p \right), \tag{65}$$

with $p, q = 0, 1, ..., N-1$.

We have for the first derivative, by using the symmetry $g_{pq} = g_{qp}$,

$$\frac{\partial \chi}{\partial x_m} = \left(-\frac{1}{2} \sum_p g_{mp} x_p + \imath d_m \right) \chi(\mathbf{x}) \tag{66}$$

with $m = 0, 1, ..., N-1$.

Evaluating Eq. (66) at $\mathbf{x} = 0$, we obtain as above, being $\chi(\mathbf{0}) = 1$,

$$\left. \frac{\partial \chi}{\partial x_m} \right|_{\mathbf{x}=0} = \imath \left. \frac{\partial \chi_I}{\partial x_m} \right|_{\mathbf{x}=0} = \imath d_m \tag{67}$$

For the second derivative of Eq. (65) we have

$$\begin{aligned} &\frac{\partial^2 \chi}{\partial x_m \partial x_n} = \\ &-\frac{1}{2} g_{mn} \chi(\mathbf{x}) + \left(-\frac{1}{2} \sum_p g_{mp} x_p + \imath d_m \right) \left(-\frac{1}{2} \sum_p g_{np} x_p + \imath d_n \right) \chi(\mathbf{x}) \end{aligned} \tag{68}$$

with $n = 0, 1, ..., N-1$. Evaluating Eq. (68) at $\mathbf{x} = 0$ we have

$$\begin{aligned} \left. \frac{\partial^2 \chi_R}{\partial x_m \partial x_n} \right|_{\mathbf{x}=0} &= -\frac{1}{2} g_{mn} - d_m d_n \\ \left. \frac{\partial^2 \chi_I}{\partial x_m \partial x_n} \right|_{\mathbf{x}=0} &= 0 \end{aligned} \tag{69}$$

and using Eq. (67) in the first of Eq. (69) we have the proof. □

VII. LINEAR TRANSFORMATIONS

Gates represented by unitary operations are of relevance in many applications. We are interested to those gates such that the new annihilation operators are expressed as a linear combination of the input operators. If the new operator is

$$\tilde{\hat{\boldsymbol{a}}} = \hat{U}^\dagger \hat{\boldsymbol{a}} \hat{U}\ , \tag{70}$$

in term of the density matrix, the transformation reads [2]

$$\tilde{\rho} = \hat{U} \rho \hat{U}^\dagger. \tag{71}$$

We start considering transformation such that

$$\hat{\tilde{\boldsymbol{a}}} = \mathbf{U}\hat{\boldsymbol{a}}. \tag{72}$$

with $\mathbf{U}$ a $n \times n$ complex matrix, with $\hat{\boldsymbol{a}}$ and $\hat{\tilde{\boldsymbol{a}}}$ column vectors of operators with dimension $n \times 1$. In a later section, we will consider the more general case with

$$\hat{\tilde{\boldsymbol{a}}} = \mathbf{U}\hat{\boldsymbol{a}} + \mathbf{W}\hat{\boldsymbol{a}}^\dagger. \tag{73}$$

For the moment, we have $\mathbf{W} = 0$.

The transformation in terms of the canonical variables read [1]

$$\hat{\tilde{\mathbf{R}}} = \hat{U}^\dagger \hat{\mathbf{R}} \hat{U} = \mathbf{M}\,\hat{\mathbf{R}} + \boldsymbol{d}' \tag{74}$$

Linear transformations transform Gaussian states into Gaussian states [1]. A Gaussian state with covariance matrix $\boldsymbol{g}$ and displacement vector $\boldsymbol{d}$, turns into a new Gaussian state with covariance matrix

$$\tilde{\boldsymbol{g}} = \mathbf{M}\,\boldsymbol{g}\,\mathbf{M}^\top \tag{75}$$

and displacement vector

$$\tilde{\boldsymbol{d}} = \mathbf{M}\,\boldsymbol{d} + \boldsymbol{d}'. \tag{76}$$

If the transformation due to $\mathbf{U}$ is unitary, the matrix $\mathbf{M}$ is symplectic, that is the following relation is satisfied

$$\mathbf{M}\mathbf{J}\mathbf{M}^\top = \mathbf{J} \tag{77}$$

with $\mathbf{J}$ given in Eq. (26).

Equivalently, the inverse of $\mathbf{M}$ is found as

$$\mathbf{M}^{-1} = \mathbf{J}\mathbf{M}^\top\mathbf{J}^\top \tag{78}$$

A. Proof of Eqs. (75) and (76)

Proof. We consider the definition of the characteristic function in the

$$\chi = \mathrm{Tr}\left(\rho e^{\imath \mathbf{x}\hat{\boldsymbol{R}}}\right) \tag{79}$$

The density matrix under the action of a unitary operator $\hat{U}$, transforms as

$$\tilde{\rho} = \hat{U}\rho\hat{U}^\dagger, \tag{80}$$

so that the corresponding characteristic function reads

$$\tilde{\chi} = \mathrm{Tr}\left(\tilde{\rho} e^{\imath \mathbf{x}\hat{\boldsymbol{R}}}\right) = \mathrm{Tr}\left(\hat{U}\rho\hat{U}^\dagger \rho e^{\imath \mathbf{x}\hat{\boldsymbol{R}}}\right) = \mathrm{Tr}\left(\rho\hat{U}^\dagger e^{\imath \mathbf{x}\hat{\boldsymbol{R}}}\hat{U}\right) = \mathrm{Tr}\left(\rho e^{\imath \hat{\mathbf{x}} U^\dagger \hat{\boldsymbol{R}}\hat{U}}\right) \tag{81}$$

where we use the cyclic property of the trace and

$$\hat{U}^\dagger f(\hat{C})\hat{U} = f(\hat{U}^\dagger \hat{C}\hat{U}) \tag{82}$$

as $\hat{U}$ is unitary [2]. Using Eq. (74)

$$\tilde{\chi} = \mathrm{Tr}\left(\tilde{\rho} e^{\imath \mathbf{x}\mathbf{M}\hat{\boldsymbol{R}} + \imath \mathbf{x}\boldsymbol{d}'}\right) = \chi(\mathbf{x}\mathbf{M}) e^{\imath \mathbf{x}\boldsymbol{d}'} \tag{83}$$

For a Gaussian χ as in Eq. (36), this transformation reads

$$\tilde{\chi}(\mathbf{x}) = \exp\left[-\frac{1}{4}\mathbf{x}\mathbf{M}\;\boldsymbol{g}\,\mathbf{M}^\top\mathbf{x}^\top + \imath\,\mathbf{x}\left(\mathbf{M}\boldsymbol{d} + \boldsymbol{d}'\right)\right] \tag{84}$$

which gives Eqs. (75) and (76). □

VIII. THE U AND M MATRICES

It is instructive to deepen the link the matrix $\mathbf{U}$ with the matrix $\mathbf{M}$ for later use. We consider the $n \times 1$ vector of the annihilation operators $\hat{\boldsymbol{a}}$,

$$\hat{\boldsymbol{a}} = \begin{pmatrix} \hat{a}_0 \\ \hat{a}_1 \\ \vdots \\ \hat{a}_{n-1} \end{pmatrix} \tag{85}$$

and the corresponding $n \times 1$ vectors of positions $\hat{\boldsymbol{q}}$ and momenta vector $\hat{\boldsymbol{p}}$,

$$\hat{\boldsymbol{q}} = \begin{pmatrix} \hat{q}_0 \\ \hat{q}_1 \\ \vdots \\ \hat{q}_{n-1} \end{pmatrix} \qquad \hat{\boldsymbol{p}} = \begin{pmatrix} \hat{p}_0 \\ \hat{p}_1 \\ \vdots \\ \hat{p}_{n-1} \end{pmatrix}. \tag{86}$$

We build the $\hat{\mathbf{R}}$ vector in Eq. (23) by using auxiliary rectangular matrices $\mathbf{R}_\mathrm{q}$ and $\mathbf{R}_\mathrm{p}$ as follows

$$\hat{\mathbf{R}} = \mathbf{R}_\mathrm{q}\hat{\boldsymbol{q}} + \mathbf{R}_\mathrm{p}\hat{\boldsymbol{p}}. \tag{87}$$

$\mathbf{R}_\mathrm{q}$ and $\mathbf{R}_\mathrm{p}$ are matrices with size $N \times n$.

For $N = 4$, we have

$$\mathbf{R}_\mathrm{q} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix} \tag{88}$$

and

$$\mathbf{R}_\mathrm{p} = \begin{pmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{pmatrix}, \tag{89}$$

such that

$$\mathbf{R}_\mathrm{q}\hat{\mathbf{q}} = \begin{pmatrix} \hat{q}_0 \\ 0 \\ \hat{q}_1 \\ 0 \end{pmatrix}, \tag{90}$$

and

$$\mathbf{R}_\mathrm{p}\hat{\mathbf{p}} = \begin{pmatrix} 0 \\ \hat{p}_0 \\ 0 \\ \hat{p}_1 \end{pmatrix}, \tag{91}$$

and hence

$$\hat{\mathbf{R}} = \mathbf{R}_\mathrm{q}\hat{\mathbf{q}} + \mathbf{R}_\mathrm{p}\hat{\boldsymbol{p}} = \begin{pmatrix} \hat{q}_0 \\ \hat{p}_0 \\ \hat{q}_1 \\ \hat{p}_1, \end{pmatrix}. \tag{92}$$

being

$$\hat{\boldsymbol{q}} = \begin{pmatrix} \hat{q}_0 \\ \hat{q}_1 \end{pmatrix}, \tag{93}$$

and

$$\hat{\boldsymbol{p}} = \begin{pmatrix} \hat{p}_0 \\ \hat{p}_1 \end{pmatrix}. \tag{94}$$

From the previous expressions, one has

$$\begin{aligned} \hat{\boldsymbol{q}} &= \mathbf{R}_{\mathrm{q}}^{\top}\hat{\mathbf{R}} \\ \hat{\boldsymbol{p}} &= \mathbf{R}_{\mathrm{p}}^{\top}\hat{\mathbf{R}} \end{aligned} \tag{95}$$

and the rectangular matrices $\mathbf{R}_{\mathrm{q}}$ and $\mathbf{R}_{\mathrm{p}}$ satisfy

$$\begin{aligned} &\mathbf{R}_{\mathrm{q}}^{\top}\mathbf{R}_{\mathrm{q}} = \mathbf{1}_n \\ &\mathbf{R}_{\mathrm{q}}^{\top}\mathbf{R}_{\mathrm{p}} = \mathbf{0}_n \\ &\mathbf{R}_{\mathrm{p}}^{\top}\mathbf{R}_{\mathrm{p}} = \mathbf{1}_n \\ &\mathbf{R}_{\mathrm{p}}^{\top}\mathbf{R}_{\mathrm{q}} = \mathbf{0}_n \\ &\mathbf{R}_{\mathrm{q}}\mathbf{R}_{\mathrm{q}}^{\top} + \mathbf{R}_{\mathrm{p}}\mathbf{R}_{\mathrm{p}}^{\top} = \mathbf{1}_N \\ &\mathbf{J}\mathbf{R}_{\mathrm{q}} = -\mathbf{R}_{\mathrm{p}} \\ &\mathbf{J}\mathbf{R}_{\mathrm{p}} = \mathbf{R}_{\mathrm{q}} \\ &\mathbf{R}_{\mathrm{q}}\mathbf{R}_{\mathrm{p}}^{\top} - \mathbf{R}_{\mathrm{p}}\mathbf{R}_{\mathrm{q}}^{\top} = \mathbf{J} \end{aligned} \tag{96}$$

We recall that, for $N = 4$ and $n = 2$ qbits, we have

$$\mathbf{J} = \bigoplus_{j=1}^{n} \mathbf{J}_1 = \begin{pmatrix} 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \end{pmatrix}. \tag{97}$$

By the matrices $\mathbf{R}_q$ and $\mathbf{R}_p$, we can write the $\hat{\boldsymbol{a}}$ vector as follows

$$\hat{\boldsymbol{a}} = \frac{\hat{\boldsymbol{q}} + \imath\hat{\boldsymbol{p}}}{\sqrt{2}} = \frac{\mathbf{R}_{\mathrm{q}}^{\top} + \imath\mathbf{R}_{\mathrm{p}}^{\top}}{\sqrt{2}}\hat{\mathbf{R}}. \tag{98}$$

For a linear transformation $\hat{U}$ from the vector $\hat{\boldsymbol{a}}$ to the new $\hat{\tilde{\boldsymbol{a}}}$,

$$\begin{aligned} \hat{\tilde{\boldsymbol{a}}} &= \mathbf{U}\hat{\boldsymbol{a}} \\ \hat{\tilde{\boldsymbol{a}}}^{\dagger} &= \mathbf{U}^{*}\hat{\boldsymbol{a}}^{\dagger} \end{aligned} \tag{99}$$

with $\mathbf{U} = \mathbf{U}_R + \imath\mathbf{U}_I$ a complex matrix with real part $\mathbf{U}_R$ and imaginary part $\mathbf{U}_I$; and $\mathbf{U}^* = \mathbf{U}_R - \imath\mathbf{U}_I$. Correspondingly

$$\begin{aligned} \hat{\tilde{\boldsymbol{q}}} &= \mathbf{U}_R\hat{\boldsymbol{q}} - \mathbf{U}_I\hat{\boldsymbol{p}} \\ \hat{\tilde{\boldsymbol{p}}} &= \mathbf{U}_I\hat{\boldsymbol{q}} + \mathbf{U}_R\hat{\boldsymbol{p}} \end{aligned} \tag{100}$$

As $\hat{\mathbf{R}}$ transforms in $\hat{\tilde{\mathbf{R}}}$, we have

$$\hat{\tilde{\mathbf{R}}} = \mathbf{R}_{\mathrm{q}}\hat{\tilde{\boldsymbol{q}}} + \mathbf{R}_{\mathrm{p}}\hat{\tilde{\boldsymbol{p}}} = \mathbf{M}\hat{\mathbf{R}} = (\mathbf{M}_1 + \mathbf{M}_2)\,\hat{\mathbf{R}} \tag{101}$$

with $\mathbf{M}_1$, $\mathbf{M}_2$, and $\mathbf{M}$ real matrices such that

$$\begin{aligned} \mathbf{M} &= \mathbf{M}_1 + \mathbf{M}_2 \\ \mathbf{M}_1 &= \mathbf{R}_{\mathrm{q}}\mathbf{U}_R\mathbf{R}_{\mathrm{q}}^{\top} + \mathbf{R}_{\mathrm{p}}\mathbf{U}_R\mathbf{R}_{\mathrm{p}}^{\top} \\ \mathbf{M}_2 &= \mathbf{R}_{\mathrm{p}}\mathbf{U}_I\mathbf{R}_{\mathrm{q}}^{\top} - \mathbf{R}_{\mathrm{q}}\mathbf{U}_I\mathbf{R}_{\mathrm{p}}^{\top} \end{aligned} \tag{102}$$

The matrix $\mathbf{M}$ is symplectic if $\mathbf{M}\mathbf{J}\mathbf{M}^{\top} = \mathbf{J}$, and the matrix $\mathbf{U}$ is unitary if $\mathbf{U}^{\dagger}\mathbf{U} = (\mathbf{U}^*)^{\top}\mathbf{U} = \mathbf{1}_n$.

For an arbitraty matrix $\mathbf{U}$, the matrix $\mathbf{M}$ is not in general a symplectic matrix, hence does not represent a linear transformation. We show in the following that if $\hat{U}$ is unitary, the matrix $\mathbf{M}$ is symplectic.

Proposition — If $\hat{U}$ is a unitary operator, i.e., $\hat{U}^{\dagger}\hat{U} = 1$, the matrix $\mathbf{U}$ is unitary.

Proof. If $\hat{U}$ is a unitary trasformation, the expected values must be invariant, and we have

$$\begin{aligned}\langle\hat{\tilde{\boldsymbol{a}}}^\dagger\hat{\tilde{\boldsymbol{a}}}\rangle = \mathrm{Tr}\left(\tilde{\rho}\hat{\tilde{\boldsymbol{a}}}^\dagger\hat{\tilde{\boldsymbol{a}}}\right) = \mathrm{Tr}\left(\hat{U}\rho\hat{U}^\dagger\hat{\tilde{\boldsymbol{a}}}^\dagger\hat{\tilde{\boldsymbol{a}}}\right)\\ =&\mathrm{Tr}\left(\rho\hat{U}^\dagger\hat{\tilde{\boldsymbol{a}}}^\dagger\hat{\tilde{\boldsymbol{a}}}\hat{U}\right)\\ =&\mathrm{Tr}\left(\rho\hat{U}^\dagger\hat{\tilde{\boldsymbol{a}}}^\dagger\hat{U}^\dagger\hat{U}\hat{\tilde{\boldsymbol{a}}}\hat{U}\right)\\ =&\mathrm{Tr}\left(\rho\boldsymbol{a}^\dagger\mathbf{U}^\dagger\mathbf{U}\boldsymbol{a}\right)\end{aligned} \tag{103}$$

where we used the cyclic property for the trace and the unitarity for $\hat{U}$. As we must have

$$\langle\hat{\tilde{\boldsymbol{a}}}^\dagger\hat{\tilde{\boldsymbol{a}}}\rangle = \langle\hat{\boldsymbol{a}}^\dagger\hat{\boldsymbol{a}}\rangle = \mathrm{Tr}\left(\rho\boldsymbol{a}^\dagger\boldsymbol{a}\right) \tag{104}$$

we find

$$\mathbf{U}^\dagger\mathbf{U} = \mathbf{1}_n, \tag{105}$$

which is the proof. □

Proposition — If U is unitary, M is symplectic

Proof. We have after Eq. (102)

$$\mathbf{M} = \mathbf{R}_\mathrm{q}\mathbf{U}_R\mathbf{R}_\mathrm{q}^\top + \mathbf{R}_\mathrm{p}\mathbf{U}_R\mathbf{R}_\mathrm{p}^\top + \mathbf{R}_\mathrm{p}\mathbf{U}_I\mathbf{R}_\mathrm{q}^\top - \mathbf{R}_\mathrm{q}\mathbf{U}_I\mathbf{R}_\mathrm{p}^\top \tag{106}$$

and using Eqs. (96)

$$\mathbf{M}^\top = \mathbf{R}_\mathrm{q}\mathbf{U}_R^\top\mathbf{R}_\mathrm{q}^\top + \mathbf{R}_\mathrm{p}\mathbf{U}_R^\top\mathbf{R}_\mathrm{p}^\top + \mathbf{R}_\mathrm{q}\mathbf{U}_I^\top\mathbf{R}_\mathrm{p}^\top - \mathbf{R}_\mathrm{p}\mathbf{U}_I^\top\mathbf{R}_\mathrm{q}^\top \tag{107}$$

$$\mathbf{J}\mathbf{M}^\top = -\mathbf{R}_\mathrm{p}\mathbf{U}_R^\top\mathbf{R}_\mathrm{q}^\top + \mathbf{R}_\mathrm{q}\mathbf{U}_R^\top\mathbf{R}_\mathrm{p}^\top - \mathbf{R}_\mathrm{p}\mathbf{U}_I^\top\mathbf{R}_\mathrm{p}^\top - \mathbf{R}_\mathrm{q}\mathbf{U}_I^\top\mathbf{R}_\mathrm{q}^\top \tag{108}$$

which give after some algebra

$$\begin{aligned}&\mathbf{M}\mathbf{J}\mathbf{M}^\top =\\ &\mathbf{R}_\mathrm{q}\left(\mathbf{U}_R\mathbf{U}_R^\top + \mathbf{U}_I\mathbf{U}_I^\top\right)\mathbf{R}_\mathrm{p}^\top - \mathbf{R}_\mathrm{p}\left(\mathbf{U}_R\mathbf{U}_R^\top + \mathbf{U}_I\mathbf{U}_I^\top\right)\mathbf{R}_\mathrm{q}^\top +\\ &-\mathbf{R}_\mathrm{q}\left(\mathbf{U}_R\mathbf{U}_I^\top - \mathbf{U}_I\mathbf{U}_R^\top\right)\mathbf{R}_\mathrm{q}^\top - \mathbf{R}_\mathrm{p}\left(\mathbf{U}_R\mathbf{U}_I^\top - \mathbf{U}_I\mathbf{U}_R^\top\right)\mathbf{R}_\mathrm{p}^\top\end{aligned} \tag{109}$$

As $\mathbf{U}$ is a unitary matrix, one has $\mathbf{U}^\dagger\mathbf{U} = (\mathbf{U}_R - \imath\mathbf{U}_I)^\top(\mathbf{U}_R + \imath\mathbf{U}_I) = \mathbf{1}_n$, that is

$$\begin{aligned}\mathbf{U}_R\mathbf{U}_R^\top + \mathbf{U}_I\mathbf{U}_I^\top = \mathbf{1}_n\\ \mathbf{U}_R\mathbf{U}_I^\top - \mathbf{U}_I\mathbf{U}_R^\top = \mathbf{0}_n,\end{aligned} \tag{110}$$

and after Eq. (109) we have

$$\mathbf{M}\mathbf{J}\mathbf{M}^\top = \mathbf{R}_\mathrm{q}\mathbf{R}_\mathrm{p}^\top - \mathbf{R}_\mathrm{p}\mathbf{R}_\mathrm{q}^\top = \mathbf{J} \tag{111}$$

where we used the last of Eqs. (96). □

IX. GAUSSIAN STATES AS A NEURAL NETWORK

Neural networks are mathematical tools to approximate unknown functions of many variables. NN depend on many parameters that can be fine tuned to improve fitting a target model. Gaussian states correspond to NN with a Gaussian activation function, which is often adopted, for example, in *Kernel machines* [3]. We implement a Gaussian density matrix by a "multi-head model", that is a model, which has a computational backbone that processes the inputs and multiple output layers to return different quantities. We start considering as "heads" the real and imaginary part χ_R and χ_I of χ as outputs.

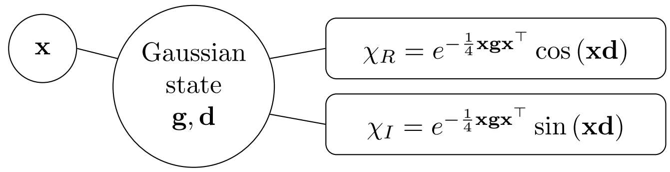

FIG. 1. Double-head model for a layer representing a Gaussian.

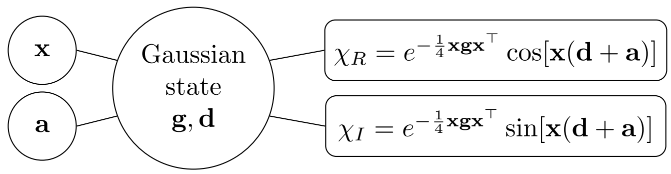

FIG. 2. Double-head model representing a Gaussian state with a bias.

X. THE SIMPLEST NEURAL NETWORK FOR A GAUSSIAN STATE

Figure 1 shows the NN model for a Gaussian state with each layer having one input $\mathbf{x}$ and two outputs (the "heads") χ_R and χ_I. The input is the state row vector $\mathbf{x}$, and is the input layer. The outputs χ_R and χ_I are the output layers.

The inner layer denoted "Gaussian state," compute the Gaussian characteristic function. The parameters of this layer, namely the covariance matrix $\boldsymbol{g}$ and the displacement $\boldsymbol{d}$ are also indicated.

For a Gaussian state with covariance matrix $\mathbf{g}$ and displacement vector $\mathbf{d}$ we have

$$\chi(\mathbf{x}) = \chi_R(\mathbf{x}) + \imath\chi_I(\mathbf{x}) \tag{112}$$

with the real and imaginary part

$$\begin{aligned}\chi_R(\mathbf{x}) &= \exp(-\tfrac{1}{4}\mathbf{x}\mathbf{g}\mathbf{x}^\top)\cos(\mathbf{x}\boldsymbol{d})\\ \chi_I(\mathbf{x}) &= \exp(-\tfrac{1}{4}\mathbf{x}\mathbf{g}\mathbf{x}^\top)\sin(\mathbf{x}\boldsymbol{d}).\end{aligned} \tag{113}$$

However, it is convenient to consider a more general model that includes a further input vector.

XI. GAUSSIAN NEURAL NETWORK WITH BIAS INPUT

For later convenience, we generalize the Gaussian layer to include a further bias input $\mathbf{a}$ with the same dimensions of $\boldsymbol{d}$, in addition to $\mathbf{x}$, such that the output is

$$\chi(x) = e^{-\frac{1}{4}\mathbf{x}\mathbf{g}\mathbf{x}^T} e^{\imath(\boldsymbol{d}+\boldsymbol{a})}, \tag{114}$$

Here $\boldsymbol{a}$ is a bias in the displacement, which will be useful for cascading layers as detailed below.

Equation (114) corresponds to the two heads of the model returning

$$\begin{aligned}\chi_R(\mathbf{x}) &= \exp(-\tfrac{1}{4}\mathbf{x}\mathbf{g}\mathbf{x}^T)\cos[\mathbf{x}(\boldsymbol{d}+\mathbf{a})]\\ \chi_I(\mathbf{x}) &= \exp(-\tfrac{1}{4}\mathbf{x}\mathbf{g}\mathbf{x}^T)\sin[\mathbf{x}(\boldsymbol{d}+\mathbf{a})].\end{aligned} \tag{115}$$

Figure 2 shows a graphical representation of the generalized Gaussian NN with bias.

We define a new layer, which takes as parameters the covariance matrix $\boldsymbol{g}$ and the displacement vector $\boldsymbol{d}$.

XII. THE VACUUM LAYER

The vacuum state is a Gaussian state with unitary covariance matrix $\boldsymbol{g} = \mathbf{1}_N$ and zero displacement (see Fig. 3). Also for the vacuum, as particular case of Gaussian state, we may consider the presence of a bias vector.

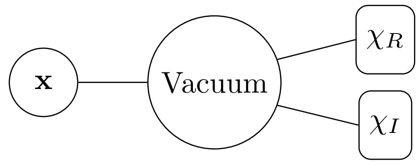

FIG. 3. The model for vacuum state.

XIII. PULLBACK

We setup our neural network representation of the density matrix as a layered sequence of gates. Each gate is a unitary operator that acts on the density matrix, and transform the latter into a new state. To detail how it works, we start considering the transformation of the characteristic functions under unitary operations.

We consider a linear transformation by the operator $\hat{U}$. The operators and the density matrix changes as follows

$$\begin{aligned} \hat{\tilde{\boldsymbol{a}}} &= \hat{U}^\dagger \hat{\boldsymbol{a}} \hat{U} \\ \tilde{\rho} &= \hat{U} \rho \hat{U}^\dagger . \end{aligned} \tag{116}$$

Correspondingly, for the transformed characteristic functions $\tilde{\chi}$, we have

$$\begin{aligned} \tilde{\chi} &= \mathrm{Tr}\left[\tilde{\rho} e^{\imath \mathbf{x} \hat{\mathbf{R}}}\right] = \mathrm{Tr}\left[\hat{U}\rho\hat{U}^\dagger e^{\imath \mathbf{x} \hat{\mathbf{R}}}\right] = \mathrm{Tr}\left[\rho\hat{U}^\dagger e^{\imath \mathbf{x} \hat{\mathbf{R}}}\hat{U}\right] \\ &= \mathrm{Tr}\left[\rho e^{\imath \mathbf{x} \hat{U}^\dagger \hat{\mathbf{R}} \hat{U}}\right] = \mathrm{Tr}\left[\rho e^{\imath \mathbf{x} \hat{\tilde{\mathbf{R}}}}\right] . \end{aligned} \tag{117}$$

Recalling Eq. (74)

$$\hat{\tilde{\mathbf{R}}} = \mathbf{M}\mathbf{R} + \boldsymbol{d}', \tag{118}$$

we remark that the linear transformation is determined by the symplectic matrix $\mathbf{M}$ and the displacement $\boldsymbol{d}'$. We write $\tilde{\chi}$ in terms of $\mathbf{M}$ and $\boldsymbol{d}'$, by using Eq. (118),

$$\tilde{\chi} = \mathrm{Tr}\left[\rho e^{\imath \mathbf{x} \mathbf{M} \hat{\mathbf{R}}}\right] e^{\imath \mathbf{x} \boldsymbol{d}'} . \tag{119}$$

On the other hand, the expression for $\chi(\mathbf{x})$ is the following

$$\chi = \mathrm{Tr}\left[\rho e^{\imath \mathbf{x} \hat{\mathbf{R}}}\right] . \tag{120}$$

Thus we have

$$\tilde{\chi}(\mathbf{x}) = \chi(\mathbf{x}\mathbf{M}) e^{\imath \mathbf{x} \boldsymbol{d}'} = \chi(\mathbf{x}\mathbf{M}) e^{\imath (\mathbf{x}\mathbf{M})(\mathbf{M}^{-1}\boldsymbol{d}')} . \tag{121}$$

From Eq. (121), we see that the modified characteristic function depends on the modified input vector $\mathbf{x}\mathbf{M}$ and has a modified displacement $\mathbf{M}^{-1}\boldsymbol{d}'$.

Introducing the new input vector $\mathbf{y} = \mathbf{x}\mathbf{M}$ and the bias $\boldsymbol{a} = \mathbf{M}^{-1}\boldsymbol{d}'$, we can express the transformed characteristic function $\tilde{\chi}$ in terms of the original χ as follows

$$\tilde{\chi}(\mathbf{x}) = \chi(\mathbf{y}) e^{\imath \mathbf{y} \boldsymbol{a}} . \tag{122}$$

XIV. THE PULLBACK LAYER

We give a graphical representation of the transformations in Eq. (122). We start considering a characteristic function with input vector $\mathbf{x}$ and multi-headed outputs χ_R and χ_I, as in figure 4. Then we generalize the model by introducing the bias vector $\boldsymbol{a}$ as in the previous chapter and shown in 5. Fig. 6 represents the transformation in equation (122).

We introduce a new layer, which we call "linear layer", having as parameters the symplectic matrix $\mathbf{M}$ and the displacement $\boldsymbol{d}'$. The linear layer is shown in Fig. 7, and it has two inputs $\mathbf{x}$ and $\boldsymbol{a}$, and two outputs: $\mathbf{y} = \mathbf{x}\mathbf{M}$, a row vector with the same size of $\mathbf{x}$, and a new displacement $\boldsymbol{b} = \mathbf{M}^{-1}(\boldsymbol{d}' + \boldsymbol{a})$ a column vector with the size of

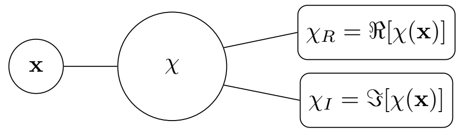

FIG. 4. Graphical rapresentation of the model for the characteristic function.

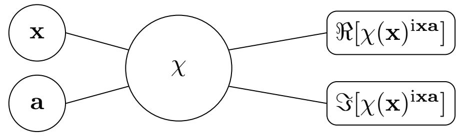

FIG. 5. Graphical rapresentation of the generalized model for the characteristic function with bias $\boldsymbol{a}$, which is adopted in the linear transformations.

$\boldsymbol{d}$. The linear layer will enable to represent many different linear transformations, as squeezing operators or Glauber operators. The bias is needed to cascade multiple transformations, as detailed below.

By using the linear layer, we represent the transformation by cascading two layers. The resulting model is Fig. 8. The model can be described as the pullback of the linear layer from the characteristic function layer. The matrix $\mathbf{M}$ first acts on the input $\mathbf{x}$, and then the function χ is evaluated on the vector $\mathbf{y}$. A simplified diagram is given in Fig. 9. This representation will be useful when considering multiple transformation on states.

XV. PULLBACK OF GAUSSIAN STATES

For a Gaussian state (see Eq. (36)) the trasformed characteristic function reads

$$\tilde{\chi}(\mathbf{x}) = e^{-\frac{1}{4}\mathbf{x}\mathbf{M}\boldsymbol{g}\mathbf{M}^\top\mathbf{x}^\top} e^{\imath\mathbf{x}\mathbf{M}(\boldsymbol{d}+\mathbf{M}^{-1}\boldsymbol{d}')}. \tag{123}$$

Equation (123) is graphically represented in Fig. 10, by using the Gaussian layer in in Fig. 2. This representation is equivalent to Fig. 11, by the linear layer, whose action is detailed in Fig. 8. As shown in Fig. 11 linear transformation on the density matrix is actually a transformation of the variable $\mathbf{x}$. This implies that the linear transformation can be represent first by a linear gate, followed by the Gaussian gate. This may be described as “pulling back” the linear operation before the Gaussian gate. By using these two models, the linear transformation of the Gaussian state is a cascade of a linear pullback layer and a Gaussian state layer, as shown in figure 11.

XVI. PULLBACK CASCADING

The pullback approach is helpful when being in the presence of multiple transformations. A sequence of linear transforms is equivalent to a sequence of pullbacks in reverse order as sketched in figure 12. For example, we consider a system originally described by the density matrix ρ and canonical observables $\hat{\mathbf{R}}$. First, the system is subject to a linear transformation with operator $\hat{U}_1$, such that the density matrix becomes

$$\hat{U}_1\rho\hat{U}_1^\dagger \tag{124}$$

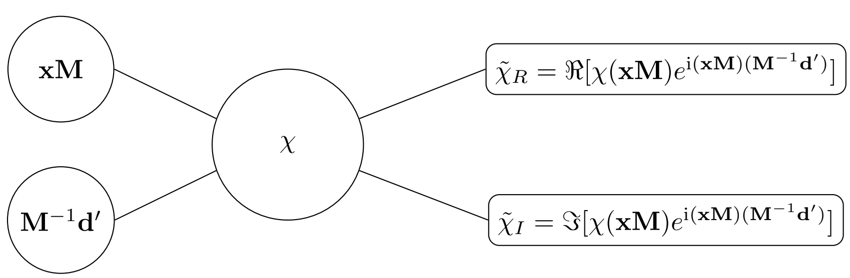

FIG. 6. Graphical representation of the transformed characteristic function in terms of the original characteristic function with modified input and bias.

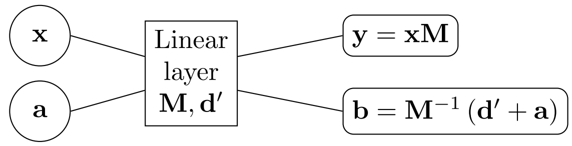

FIG. 7. Linear layer to transform the input variables to the χ layer.

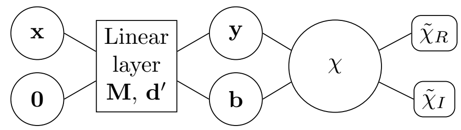

FIG. 8. Graphical representation of a linear transformation as a pullback of a linear layer from the original characteristic function. The resulting model corresponds the transformed characteristic function in Fig. 6. Note that here $\boldsymbol{a} = 0$.

and the new canonical vector is

$$\hat{R}_1 = \mathbf{M}_1 \hat{R} + \boldsymbol{d}_1. \tag{125}$$

We then consider a second transformation with unitary operator $\hat{U}_2$ and parameters $\mathbf{M}_2$ and $\boldsymbol{d}_2$, such that the final density matrix reads

$$\hat{U}_2 \hat{U}_1 \rho \hat{U}_1^\dagger \hat{U}_2^\dagger, \tag{126}$$

and the final vector of obverables is

$$\hat{\mathbf{R}}_2 = \mathbf{M}_2 \hat{\mathbf{R}}_1 + \boldsymbol{d}_2 = \mathbf{M}_2 \mathbf{M}_1 \hat{\mathbf{R}} + \boldsymbol{d}_2 + \mathbf{M}_2 \boldsymbol{d}_1. \tag{127}$$

We remark that $\mathbf{M}_1\mathbf{M}_2 \neq \mathbf{M}_2\mathbf{M}_1$, hence the order of the two tranformation is relevant as the corresponding unitary operators $\hat{U}_1$ and $\hat{U}_2$ do not commute. Hence the sequence of the two linear transformations is equivalent to a single transformation with parameters

$$\begin{aligned} \mathbf{M} &= \mathbf{M}_2 \mathbf{M}_1 \\ \boldsymbol{d}' &= \mathbf{M}_2 \boldsymbol{d}_1 + \boldsymbol{d}_2 \end{aligned} \tag{128}$$

We have the following (figure 13)

Proposition — The sequence of the two linear transfomations $\hat{U}_1$ and $\hat{U}_2$ is equivalent to the pullback of the two linear layers with parameters $(\mathbf{M}_1, \boldsymbol{d}_1)$ and $(\mathbf{M}_2, \boldsymbol{d}_2)$.

Proof. As indicated in Fig. 13 the ouput of the linear layer corresponding to $\hat{U}_2$ is

$$\begin{aligned} \mathbf{x}_2 &= \mathbf{x}\mathbf{M}_2 \\ \boldsymbol{a}_2 &= \mathbf{M}_2^{-1}\left(\boldsymbol{d}_2 + \boldsymbol{a}\right) \end{aligned} \tag{129}$$

The output of the linear layer corresponding to $\hat{U}_1$ is

$$\begin{aligned} \mathbf{x}_1 &= \mathbf{x}_2 \mathbf{M}_1 = \mathbf{x}\mathbf{M}_2\mathbf{M}_1 \\ \boldsymbol{a}_1 &= \mathbf{M}_1^{-1}\left(\boldsymbol{d}_1 + \boldsymbol{a}_2\right) = \mathbf{M}_1^{-1}\left(\boldsymbol{d}_1 + \mathbf{M}_2^{-1}\boldsymbol{d}_2 + \mathbf{M}_2^{-1}\boldsymbol{a}\right) \end{aligned} \tag{130}$$

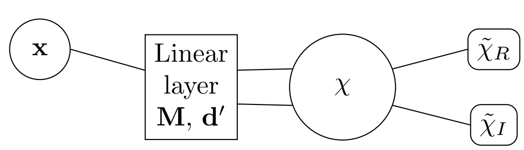

FIG. 9. Simplified model of Fig. 8, omitting the internal variables $\mathbf{y}$ and $\boldsymbol{b}$ and the zero input bias $\boldsymbol{a}$.

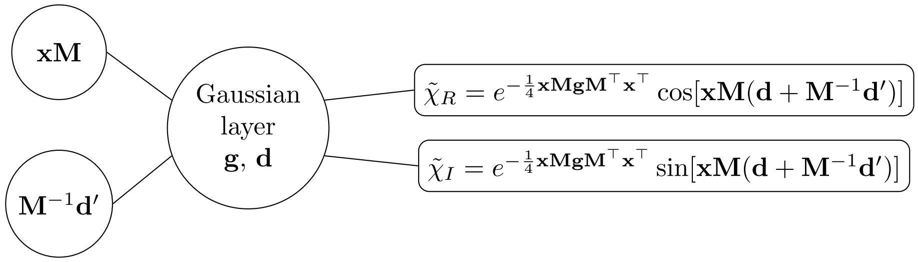

FIG. 10. Single-layer multiheaded model for a linear transformation of a Gaussian state.

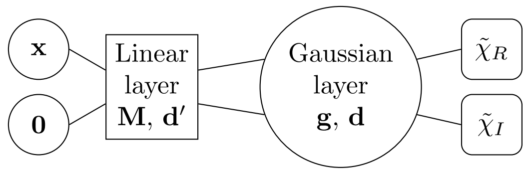

FIG. 11. Two-layer multiheaded model equivalent to Fig. 10.

For the layer with parameters $(\mathbf{M}, \boldsymbol{d}')$ we have, see (128),

$$
\begin{aligned}
\mathbf{y} &= \mathbf{x}\mathbf{M} = \mathbf{x}\mathbf{M}_2\mathbf{M}_1 \\
\boldsymbol{b} &= \mathbf{M}^{-1}\left(\boldsymbol{d}' + \boldsymbol{a}\right) \\
&= \mathbf{M}_1^{-1}\mathbf{M}_2^{-1}\left(\mathbf{M}_2\boldsymbol{d}_1 + \boldsymbol{d}_2 + \boldsymbol{a}\right) \\
&= \mathbf{M}_1^{-1}\left(\boldsymbol{d}_1 + \mathbf{M}_2^{-1}\boldsymbol{d}_2 + \mathbf{M}_2^{-1}\boldsymbol{a}\right)
\end{aligned}
\tag{131}
$$

As $\mathbf{y} = \mathbf{x}_1$ and $\boldsymbol{b} = \boldsymbol{a}_1$ we have the proof. □

By extending the previous argument to three or more transformations, one realizes that the cascade an arbitrary number of transformation corresponds to a cascade of pullbacks in reverse order. For M transformations, reverse

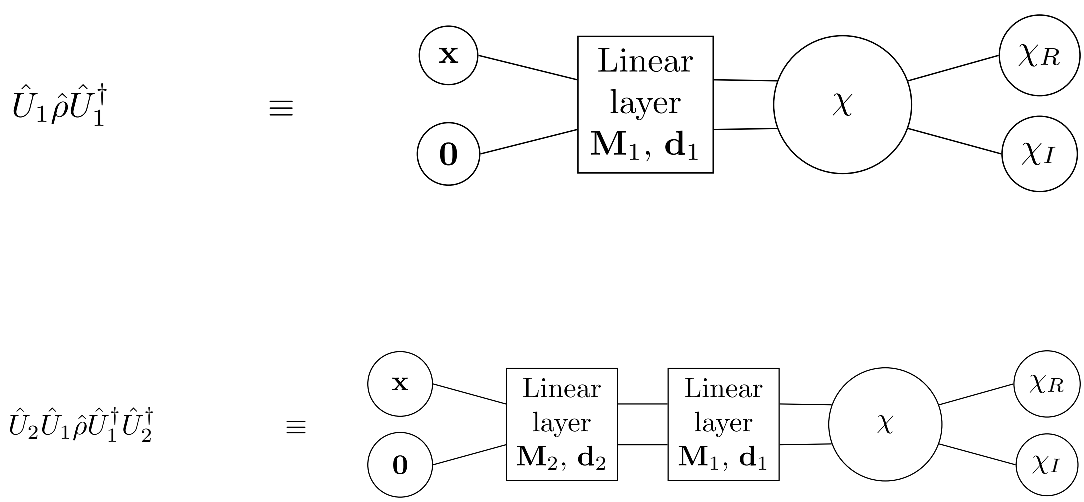

FIG. 12. Linear trasformations and pullbacks. A single transformation with unitary operator $\hat{U}_1$ corresponds to a single pullback. A double transformation with unitary operators $\hat{U}_1$ and $\hat{U}_2$ corresponds to a double pullback. Note that the flow of data from $\mathbf{x}$ to the output through the nextwork is in reverse order with respect to the two transformations.

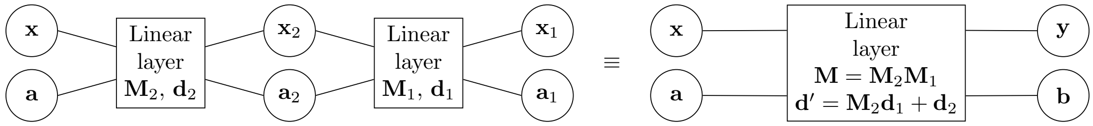

FIG. 13. Equivalence of two cascaded linear layers with a single linear layer.

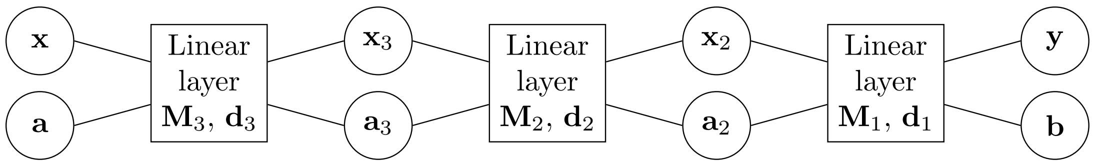

FIG. 14. Example of cascading three transformations. Note the reverse order of layer from right to left.

order means that one first make a pullback of operator 1, then operator 2, etc. In the flow of data in the network, the data $\mathbf{x}$ first enter operator M, then $M-1$ and so forth to passing throug the linear layer corresponding to the transformation 1. The case $M=3$ is shown as example in figure 14

XVII. THE GLAUBER DISPLACEMENT LAYER

Starting from the linear layer, we can define specialized layers corresponding to different unitary operators. The first we describe is the displacement operator, or Glauber operator, defined by

$$\hat{\mathcal{D}}(\alpha) = \exp\left(\alpha^* \hat{a} - \alpha \hat{a}^\dagger\right) \tag{132}$$

A coherent state is obtained by displacing the vacuum state:

$$|\alpha\rangle = \hat{\mathcal{D}}(\alpha)|0\rangle. \tag{133}$$

For a single mode, letting $\hat{U} = \hat{\mathcal{D}}(\alpha)$, we have

$$\hat{\tilde{a}} = \hat{U}^\dagger \hat{a} \hat{U} = \hat{a} + \alpha, \tag{134}$$

which implies for the canonical vector

$$\hat{\tilde{\mathbf{R}}} = \hat{U}^\dagger \hat{\mathbf{R}} \hat{U} = \begin{pmatrix} \hat{x} \\ \hat{p} \end{pmatrix} + \begin{pmatrix} d_0 \\ d_1 \end{pmatrix} \tag{135}$$

with

$$\begin{aligned} d_0 &= \sqrt{2}\Re(\alpha) \\ d_1 &= \sqrt{2}\Im(\alpha). \end{aligned} \tag{136}$$

For a many-body displacement $\hat{\mathcal{D}}(\boldsymbol{\alpha})$, wit $\boldsymbol{\alpha} = (\alpha_0, \alpha_1, ..., \alpha_n)^\top$, we have

$$\hat{\tilde{\boldsymbol{a}}} = \hat{U}^\dagger \hat{\boldsymbol{a}} \hat{U} = \hat{\boldsymbol{a}} + \boldsymbol{\alpha}, \tag{137}$$

which implies for the canonical vector

$$\hat{\tilde{\mathbf{R}}} = \hat{U}^\dagger \hat{\mathbf{R}} \hat{U} = \hat{\mathbf{R}} + \boldsymbol{d} \tag{138}$$

with $(j = 0, 1, \ldots, n-1)$

$$\begin{aligned} d_{2j} &= \sqrt{2}\Re(\alpha_j) \\ d_{2j+1} &= \sqrt{2}\Im(\alpha_j) \end{aligned} \tag{139}$$

For the Glauber layer we hence have $\mathbf{M} = \mathbf{1}_N$.

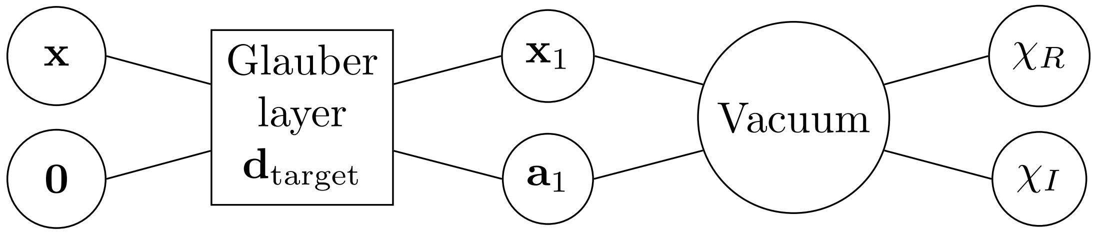

FIG. 15. Model for a coherent state obtained by pullback of a Glauber displacement operator (Glauber layer) from the vacuum.

XVIII. A NEURAL NETWORK REPRESENTATION OF A COHERENT STATE

To create a neural network model that represents a coherent state with a given displacement vector $\boldsymbol{d}$, one can start from the vacuum Gaussian state with $\boldsymbol{g} = \mathbf{1}_N$ and $\boldsymbol{d} = \mathbf{0}$ and pullback a linear gate with $\mathbf{M} = \mathbf{1}_N$, and displacement $\boldsymbol{d}$. No bias is needed $\boldsymbol{a} = \mathbf{0}$.

XIX. THE GENERALIZED SYMPLECTIC OPERATOR

Different nonclassical states are generate from vacuum by unitary operator resulting into the following linear relation

$$\hat{\tilde{\boldsymbol{a}}} = \mathbf{U}\hat{\boldsymbol{a}} + \mathbf{W}\hat{\boldsymbol{a}}^\dagger, \tag{140}$$

which generalizes Eq. (99). Using the matrices $\mathbf{U}$ and $\mathbf{W}$, one can obtain the corresponding symplectic matrix $\mathbf{M}$. Let

$$\mathbf{U} = \mathbf{U}_R + \imath\mathbf{U}_I \tag{141}$$

and

$$\mathbf{W} = \mathbf{W}_R + \imath\mathbf{W}_I, \tag{142}$$

we write the $\hat{\tilde{\boldsymbol{a}}}$ as follows

$$\hat{\tilde{\boldsymbol{a}}} = \frac{\hat{\tilde{\boldsymbol{q}}} + \imath\hat{\tilde{\boldsymbol{p}}}}{\sqrt{2}} = \mathbf{U}\hat{\boldsymbol{a}} + \mathbf{W}\hat{\boldsymbol{a}}^\dagger \tag{143}$$

$$\hat{\tilde{\boldsymbol{a}}}^\dagger = \frac{\hat{\tilde{\boldsymbol{q}}} - \imath\hat{\tilde{\boldsymbol{p}}}}{\sqrt{2}} = \mathbf{U}^*\hat{\boldsymbol{a}}^\dagger + \mathbf{W}^*\hat{\boldsymbol{a}}, \tag{144}$$

we have

$$\begin{aligned}\hat{\tilde{\boldsymbol{q}}} &= (\mathbf{U}_R + \mathbf{W}_R)\hat{\boldsymbol{q}} + (-\mathbf{U}_I + \mathbf{W}_I)\hat{\boldsymbol{p}}\\ \hat{\tilde{\boldsymbol{p}}} &= (\mathbf{U}_I + \mathbf{W}_I)\hat{\boldsymbol{q}} + (\mathbf{U}_R - \mathbf{W}_R)\hat{\boldsymbol{p}}\end{aligned} \tag{145}$$

which generalize Eq. (100) Following the same arguments for and Eq. (102), we have for the symplectic matrix

$$\begin{aligned}\mathbf{M} &= \mathbf{M}_1 + \mathbf{M}_2\\ \mathbf{M}_1 &= \mathbf{R}_{\mathrm{q}}(\mathbf{U}_R + \mathbf{W}_R)\mathbf{R}_{\mathrm{q}}^\top + \mathbf{R}_{\mathrm{p}}(\mathbf{U}_R - \mathbf{W}_R)\mathbf{R}_{\mathrm{p}}^\top\\ \mathbf{M}_2 &= \mathbf{R}_{\mathrm{p}}(\mathbf{U}_I + \mathbf{W}_I)\mathbf{R}_{\mathrm{q}}^\top + \mathbf{R}_{\mathrm{q}}(-\mathbf{U}_I + \mathbf{W}_I)\mathbf{R}_{\mathrm{p}}^\top\end{aligned} \tag{146}$$

(146) is used for generating proper layers for a given linear transformation.

XX. SINGLE-MODE SQUEEZED STATE

We first consider a single model squeezed state, so that $N = 2$, and

$$\hat{\mathbf{R}} = \begin{pmatrix} \hat{x} \\ \hat{p} \end{pmatrix} \tag{147}$$

Using the squeezing operator with parameters r and θ [2]

$$\hat{\tilde{a}} = \hat{S}^\dagger \hat{a} \hat{S} = \cosh(r)\hat{a} - e^{\imath\theta}\sinh(r)\hat{a}^\dagger \tag{148}$$

we have that the matrix $\mathbf{U}_{R,I}$ and $\mathbf{W}_{R,I}$ are complex scalars as follows

$$\begin{aligned} U_R &= \cosh(r) \\ U_I &= 0 \\ W_R &= -\cos(\theta)\sinh(r) \\ W_I &= -\sin(\theta)\sinh(r) \end{aligned} \tag{149}$$

and by (146) we find the symplectic operator for the squeezing

$$\mathbf{M}_s(r,\theta) = \begin{pmatrix} \cosh(r) - \cos(\theta)\sinh(r) & -\sin(\theta)\sinh(r) \\ -\sin(\theta)\sinh(r) & \cosh(r) + \cos(\theta)\sinh(r) \end{pmatrix} \tag{150}$$

One can verify by direct calculation that $\mathbf{M}$ is symplectic, i.e. $\mathbf{M}^\top \mathbf{J}\mathbf{M} = \mathbf{J}$.

In the special case $\theta = 0$, i.e., for a real squeezing parameter, we have

$$\mathbf{M}_s(r,0) = \begin{pmatrix} \exp(-r) & 0 \\ 0 & \exp(r) \end{pmatrix} \tag{151}$$

XXI. MULTI-MODE SQUEEZED VACUUM MODEL

We are interested to multi-mode systems, so we consider a $n-$body state,and apply the squeezing operator to one mode. The single-mode squeezing operator is obtained by a linear gate with $\boldsymbol{d'} = 0$ and $\mathbf{M}$ given by Eq. (150) for the mode to be squezeed corresponding to an index $n_{squeezed}$ in the range 0 to $n-1$. For example, for $n = 4$, and $n_{\text{squeezed}} = 0$, we have

$$\mathbf{M} = \begin{pmatrix} M_{s,11} & M_{s,12} & 0 & 0 & 0 & 0 & 0 & 0 \\ M_{s,21} & M_{s,22} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \tag{152}$$

Seemingly for $n = 4$, $n_{\text{squeezed}} = 2$

$$\mathbf{M} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & M_{s,11} & M_{s,12} & 0 & 0 \\ 0 & 0 & 0 & 0 & M_{s,21} & M_{s,22} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \tag{153}$$

A model for squeezed vacuum is built by pulling back from the vacuum a linear layer as in Figure 16, as in the following

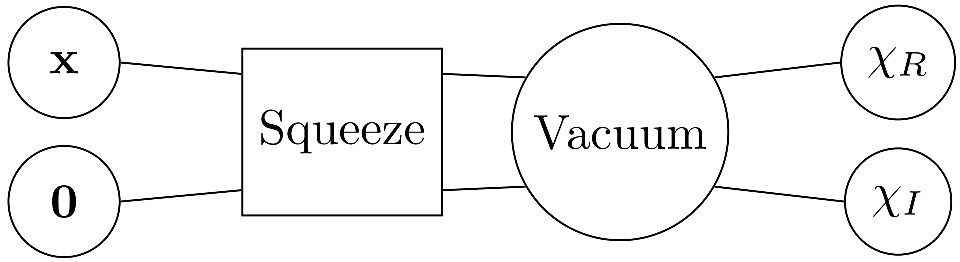

FIG. 16. A squeezed coherent state obtained by pullback with a single mode squeezing layer from the vacuum.

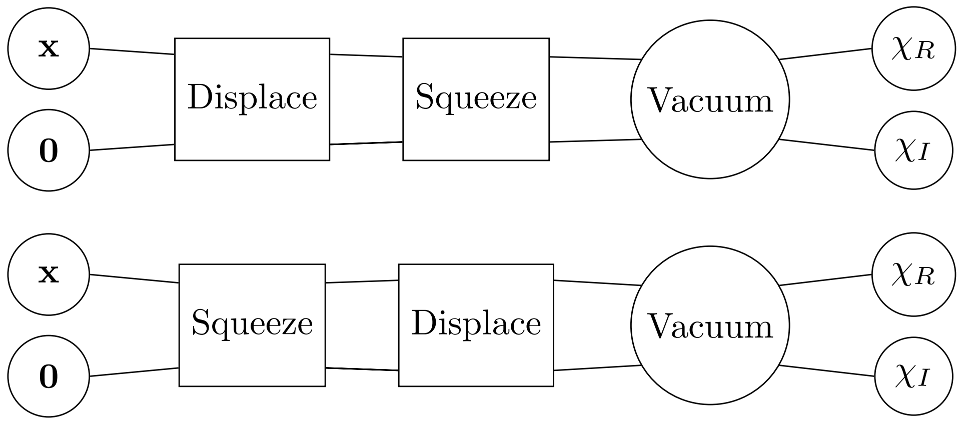

FIG. 17. Pullback model for squeezed coherent state by displacing a squeezed vacuum (top panel). Pullback model for squeezed coherent state by squeezing a displaced vacuum (bottom panel). Here all the layers act on the same mode.

XXII. SQUEEZED COHERENT STATES

A. Displacing the squeezed vacuum

The squeeze operator is represented by a linear gate, we can cascade with other layers, as the Glauber displacement layer. Figure 17a shows a model to generate a squeezed coherent state from the squeezed vacuum by a displacement layer.

The squeezed coherent states $|\alpha,\zeta\rangle$ are built by applying the displacement operator $\hat{\mathcal{D}}(\alpha)$ to a squeezed vacuum $\hat{\mathcal{S}}(\zeta)|0\rangle$, where $\zeta = r\,e^{\imath\theta}$ is the complex squeezing parameter, that is

$$|\alpha,\zeta\rangle = \hat{\mathcal{D}}(\alpha)\hat{\mathcal{S}}(\zeta)|0\rangle \tag{154}$$

The resulting state has $\langle\hat{a}\rangle = \alpha$, and it is squeezed. In the following coding example, we first define a displacement layer with a constant displacement $d_j = 3.0$ for $j = 0, 1, ..., N-1$, and then we concatenate the layers as follows.

B. Squeezing the displaced vacuum

A different squeezed coherent state is obtained by squeezing a coherent state. We first pullback a displacement layer from the vacuum to create the coherent state, and then pullback a squeezing layer, as shown in Fig. 17b. This corresponds the following equation (155)

$$|\alpha\cosh(r) - \alpha^* e^{\imath\theta}\sinh(r), \zeta\rangle = \hat{\mathcal{S}}(\zeta)\hat{\mathcal{D}}(\alpha)|0\rangle \tag{155}$$

The result is a squeezed coherent state with the same eigenvalues for the covariance matrix as above, but the mean value of the displacement is changed[2], i.e.,

$$\langle\hat{a}\rangle = \alpha\cosh(r) - \alpha^* e^{\imath\theta}\sinh(r). \tag{156}$$

XXIII. COMPUTING THE HAMILTONIAN

Given the vacuum state with characteristic function χ, one can build the NN model of an arbitrary state by multiple pullbacks (see Fig. 6 of main manuscript).

We compute the mean value of observable quantities as $\hat{H}$ and $\hat{N}$ as derivatives of the characteristic function computed in phase-space origin $\mathbf{x} = 0$.

We have in symmetric ordering [2]

$$\langle \hat{K} \rangle = \sum_{jk} \omega_{jk} \left[\frac{\partial^2}{\partial \alpha_j \partial(-\alpha_k^*)} \chi(\boldsymbol{\alpha}) - \frac{\delta_{jk}}{2} \chi \right] \Bigg|_{\boldsymbol{\alpha}=0} \tag{157}$$

where $\boldsymbol{\alpha} = \left(\alpha_0 \dots \alpha_{n-1}, \alpha_0^* \dots \alpha_{n-1}^*\right)$, and $\sqrt{2}\alpha_j = x_{2j} + \imath x_{2j+1}$, and for the interaction term

$$\langle \hat{V} \rangle = \frac{\gamma}{2} \sum_j \frac{\partial^4 \chi}{\partial \alpha_j^2 \partial(-\alpha_j^*)^2} - 2 \frac{\partial^2 \chi}{\partial \alpha_j \partial(-\alpha_j^*)} + \frac{\chi}{2} \Bigg|_{\boldsymbol{\alpha}=0} . \tag{158}$$

These quantities can be computed by using automatic differentiation on the NN model. However, as we are dealing with a Gaussian state, derivatives can be explicitly computed algebraically by the tensors $\boldsymbol{g}$ and $\boldsymbol{d}$, which speeds up computing and training the model.

Specifically, we get

$$\langle \hat{K} \rangle = -\frac{1}{2} \sum_{mn} \left[\frac{\partial^2 \chi}{\partial q_m \partial q_n} + \frac{\partial^2 \chi}{\partial p_m \partial p_n} + \chi \right] \omega_{mn}^R + \left(\frac{\partial^2 \chi}{\partial q_m \partial p_n} - \frac{\partial^2 \chi}{\partial p_n \partial p_q} \right) \omega_{mn}^I \Bigg|_{\mathbf{x}=0} \tag{159}$$

where

$$\omega_{mn} = \omega_{mn}^R + \imath \omega_{mn}^I \tag{160}$$

and $q_j = x_{2j}$ and $p_j = x_{2j+1}$ with $j = 0, 1, ..., N/2 - 1$.

For the potential energy we have

$$\langle \hat{V} \rangle = \sum_{nm} V_{nm} \langle \hat{a}_n^\dagger \hat{a}_m^\dagger \hat{a}_n \hat{a}_m \rangle \ , \tag{161}$$

and for a local interaction, we have

$$V_{nm} = \frac{1}{2} \delta_{nm} \ . \tag{162}$$

We also have

$$\langle \hat{a}_j^\dagger \hat{a}_j^\dagger \hat{a}_j \hat{a}_j \rangle = \frac{1}{4} \left(\partial_{q_j}^2 + \partial_{p_j}^2 \right)^2 \chi + \left(\partial_{q_j}^2 + \partial_{p_j}^2 \right) \chi + \frac{1}{2} \chi \Big|_{\mathbf{x}=0} \tag{163}$$

Seemingly, for the average particle number we have

$$\langle \hat{N}_T \rangle = \sum_j \langle \hat{n}_j \rangle = -\frac{1}{2} \left(\partial_{q_j}^2 + \partial_{p_j}^2 \right) \tag{164}$$

XXIV. OBSERVABLES IN TERMS OF THE COVARIANCE MATRIX AND THE DISPLACEMENT

As we are dealing with Gaussian variables, we can express observables in terms of the covariance matrix and the displacement.

We have, for a mode with index $j = 0, 1, \dots, n-1$ the normal ordering product

$$\langle \hat{n}_j^2 \rangle = \langle \hat{a}_j^\dagger \hat{a}_j \hat{a}_j^\dagger \hat{a}_j \rangle = \langle \hat{a}_j^\dagger \hat{a}_j^\dagger \hat{a}_j \hat{a}_j \rangle + \langle \hat{a}_j^\dagger \hat{a}_j \rangle \tag{165}$$

where we used $[\hat{a}_j, \hat{a}_j^\dagger] = \hat{a}_j\hat{a}_j^\dagger - \hat{a}_j^\dagger\hat{a}_j = 1$.

We have

$$\langle \hat{a}_j^\dagger \hat{a}_j\rangle = \langle \hat{n}_j\rangle = -\left.\frac{\partial^2\chi}{\partial z_j\ \partial z_j^*}\right|_{\boldsymbol{z}=0} = -\frac{1}{2}\left(\partial_{q_j}^2 + \partial_{p_j}^2\right)\chi_R\Big|_{\mathbf{x}=0} - \frac{1}{2} \;. \tag{166}$$

and

$$\langle \hat{a}_j^\dagger\hat{a}_j^\dagger\hat{a}_j\hat{a}_j\rangle = \left(\frac{\partial}{\partial z_j}\right)^2\left(-\frac{\partial}{\partial z_j^*}\right)^2\chi\Bigg|_{\mathbf{x}=0} - 2\left(\frac{\partial}{\partial z_j}\right)\left(-\frac{\partial}{\partial z_j^*}\right)\chi\Bigg|_{\mathbf{x}=0} + \frac{1}{2} \;. \tag{167}$$

Eq. (167) can be expressed in terms of the derivatives w.r.t. q_j and p_j as follows

$$\langle \hat{a}_j^\dagger\hat{a}_j^\dagger\hat{a}_j\hat{a}_j\rangle = \frac{1}{4}\left(\partial_{q_j}^2 + \partial_{p_j}^2\right)^2\chi_R\Big|_{\mathbf{x}=0} + \left(\partial_{q_j}^2 + \partial_{p_j}^2\right)\chi_R\Big|_{\mathbf{x}=0} + \frac{1}{2} \tag{168}$$

where, as above, $q_j = x_{2j}$ and $p_j = x_{2j+1}$ with $j = 0, 1, ..., N/2 - 1$.

The real part χ_R enters in Eq. (168) as $\langle \hat{n}_j^2\rangle$ in Eq. (165) is real valued.

In general, we need to evaluate the fourth order derivatives $\frac{\partial^4\chi}{\partial x_s\partial x_p\partial x_q\partial x_r}$, at $\mathbf{x} = 0$. The computation of the fourth-order derivatives may be resource-demanding as it grows with N^4. For the specific case of Gaussian states, it can be simplified as the fourth order derivatives can be expressed in terms of the covariance matrix $\boldsymbol{g}$.

We have by direct derivation of Eq. ((1))

$$\begin{aligned}\left.\frac{\partial^4\chi}{\partial x_s\partial x_p\partial x_q\partial x_r}\right|_{\mathbf{x}=0} \\ &= \frac{1}{4}g_{sp}g_{qr} + \frac{1}{4}g_{sq}g_{pr} + \frac{1}{4}g_{sr}g_{pq} \\ &+ \frac{1}{2}g_{sp}d_q\, d_r + \frac{1}{2}g_{sq}d_p\, d_r + \frac{1}{2}g_{sr}d_p\, d_q \\ &+ \frac{1}{2}g_{pq}d_r\, d_s + \frac{1}{2}g_{pr}d_q\, d_s + \frac{1}{2}g_{qr}d_p\, d_s \\ &+ d_s d_p d_q d_r \;,\end{aligned} \tag{169}$$

When considering the diagonal terms we have (no implicit sum is present in the following formula)

$$\left.\frac{\partial^4\chi}{\partial x_q^4}\right|_{\mathbf{x}=0} = \frac{3}{4}g_{qq}^2 + 3\, g_{qq}d_q^2 + d_q^4 \;, \tag{170}$$

Seemingly, we have

$$\left.\frac{\partial^4\chi}{\partial x_q^2\partial x_p^2}\right|_{\mathbf{x}=0} = \frac{1}{4}g_{qq}g_{pp} + \frac{1}{2}g_{pq}^2 + \frac{1}{2}g_{pp}d_q^2 + \frac{1}{2}g_{qq}d_p^2 + 2g_{pq}d_p\, d_q + d_p^2 d_q^2 \;. \tag{171}$$

Hence, for Gaussian states, we do not need to evaluate explicitly the fourth order derivatives, but we can combine the componentes of $\boldsymbol{g}$ and $\boldsymbol{d}$.

Seemingly, for the second derivatives we have

$$\begin{aligned}\left.\frac{\partial^2\chi_R}{\partial x_m\partial x_n}\right|_{\mathbf{x}=0} &= -\frac{1}{2}g_{mn} - d_m d_n \\ \left.\frac{\partial^2\chi_I}{\partial x_m\partial x_n}\right|_{\mathbf{x}=0} &= 0\end{aligned} \tag{172}$$

XXV. GRAPH AND PARAMETERS OF THE MODEL

Summarizing, our model include layers representing trainable gates, and layers for computing the Hamiltonian. Figure 18 show a graphical representation of the graph of the model.

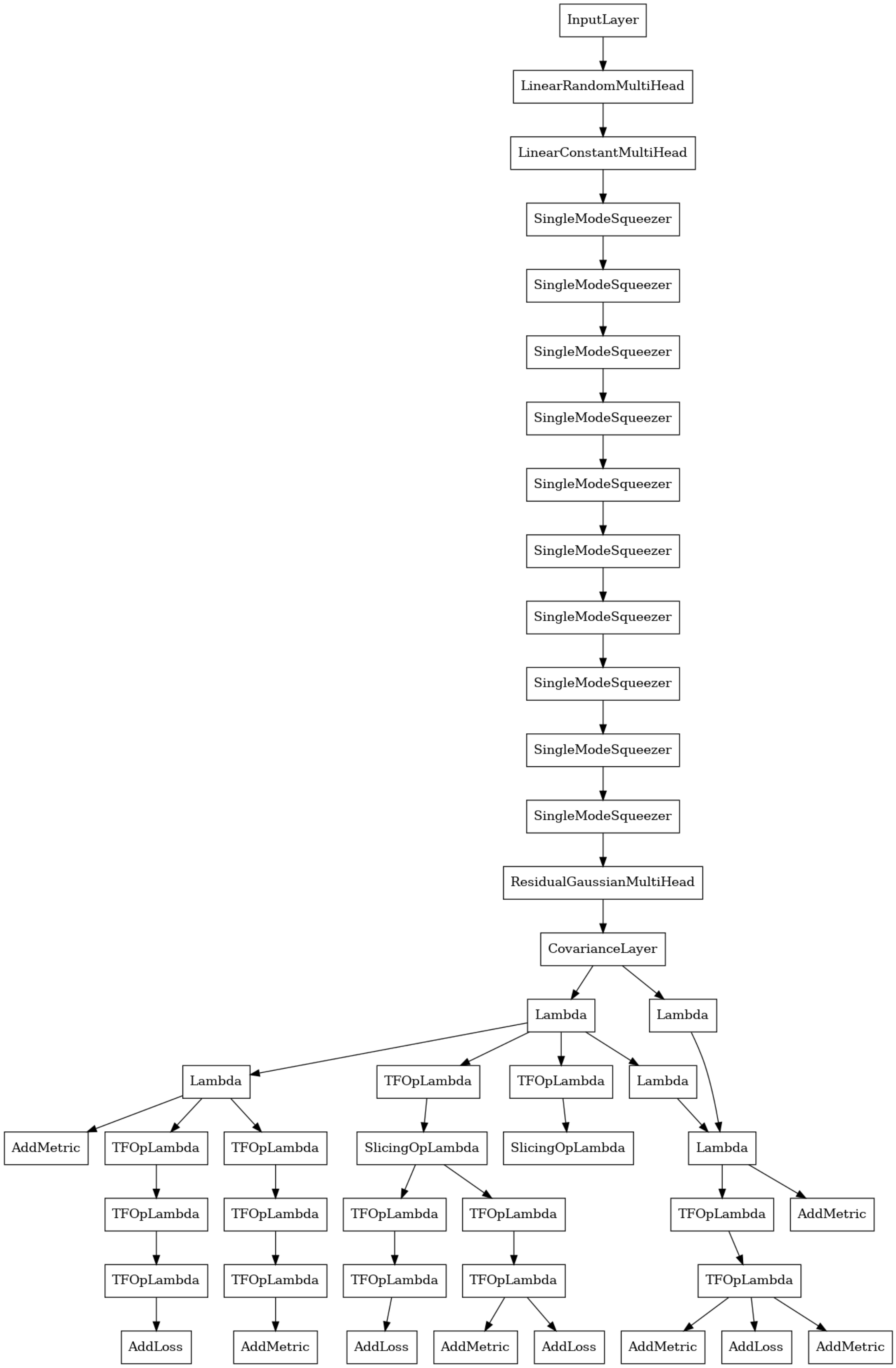

FIG. 18. Graph of a multi-head modek. One can notice the central computational core represented by the cascade of multi-head layers, including the interferometers `LinearRandomMultiHead`, displacements `LinearConstantMultiHead` and squeezing layers `SingleModeSqueezer`. Then we have the activation layer, computing the Gaussian characteristic function `ResidualGaussianMultiHead`, and the layer to determine the covariance matrix `CovarianceLayer`. After that a series of “Lambda” layers, i.e., simple computing layers for the cost functions and the various metrics during the training.

We optimize the model in order to minimize the Hamiltonian or other cost function, the resulting parameters gives the circuit that produce a state with the minimum value of energy.

The parameters we use are the trainable parameters in the model, i.e., those corresponding to the n displacement layers (each layer with 1 complex parameter), the n squeezing layers (each layer with 1 complex parameter), and the unitary matrix representing the trainable interferometer (with $N^2/4$ independent parameters). Thus overall the model has $N^2/4 + N$ independent variables.

Training is done by a the Adam algorithm (`TensorFlow v2.7.0`). Gaussian boson sampling from the trained model is done following [4].

[1] X. Wang, T. Hiroshima, A. Tomita, and M. Hayashi, Quantum information with gaussian states, Phys. Rep. **448**, 1 (2007).
[2] S. M. Barnett and P. M. Radmore, *Methods in Theoretical Quantum Optics* (Oxford University Press, New York, 1997).
[3] M. Schuld and F. Petruccione, *Supervised Learning with Quantum Computers*, edited by Springer (Springer, 2018).
[4] C. Conti, Training gaussian boson sampling by quantum machine learning, Quantum Machine Intelligence **3**, 26 (2021).