Variational quantum algorithm for Gaussian discrete solitons and their boson sampling

Claudio Conti*

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

(Received 24 August 2021; revised 23 March 2022; accepted 13 July 2022; published 22 July 2022)

In the context of quantum information, highly nonlinear regimes, such as those supporting solitons, are marginally investigated. We miss general methods for quantum solitons, although they can act as entanglement generators or as self-organized quantum processors. We develop a computational approach that uses a neural network as a variational ansatz for quantum solitons in an array of waveguides. By training the resulting phase space quantum machine-slearning model, we find different soliton solutions, varying the number of particles and interaction strength. We consider Gaussian states that enable measuring the degree of entanglement and sampling the probability distribution of many-particle events. We also determine the probability of generating particle pairs and unveil that soliton bound states emit correlated pairs. These results may have a role in boson sampling with nonlinear systems and in quantum processors for entangled nonlinear waves.

DOI: 10.1103/PhysRevA.106.013518

I. INTRODUCTION

A soliton is a nonperturbative solution of a classical nonlinear wave equation; it may describe mean-field states of atoms (as in Bose-Einstein condensation) or photons (as in nonlinear optics) [1]. From a quantum mechanical perspective, a soliton may correspond to a coherent state; however, the nonlinearity may induce squeezing or non-Gaussianity [2]. The quantum properties of solitons inspired experimental investigations, as quantum nondemolition, squeezing [3–6], and photon bound states [7]. Authors have reported on theoretical studies on soliton quantum features, such as evaporation and breathing [8–13]. Following these investigations, one can argue that solitons are unique tools for fundamental many-body phenomena and sources of highly nonclassical states.

However, solitons are overlooked in quantum technologies, as they require specialized theoretical methods [14–18] that do not appear compatible with modern quantum information. On the other hand, quantum optical protocols mainly employ linear circuits, and we ask if nonlinear waves may have a role in quantum computing [19–23] or—-reciprocally—if quantum processors can generate quantum solitons. An intriguing opportunity arises from the fact that linear circuits for the quantum advantage in boson sampling (BS) [24–29] adopt the same devices used to observe discrete solitons [1,30,31]. BS in a linear optical circuit is intermediate between a many-body system not performing computation and a universal quantum computer [32]. To make a conceptual step, we can consider nonlinear phenomena in BS. A soliton is a self-organized nonlinear channel. Is the probability of specific events affected by these nonlinear channels? Can we do BS with solitons?

To merge solitons and quantum information, we need methodologies for quantum protocols on many-body states bounded by nonlinearity. Recently, neural networks (NNs) have been introduced to study many-body models and quantum circuits [33]. Quantum machine learning (QML), i.e., the development and training of parameterized quantum circuits [34–38], is a technique to approximate high-dimensional ground states [39]. In QML, one considers quantum circuits depending on trainable parameters, such as rotation angles for qubits or degree of squeezing for continuous variables [40]. Applications include quantum classifiers [41,42], or variational quantum algorithms [43,44], for optimization and quantum simulations.

Here we develop a quantum variational ansatz in the phase space for quantum soliton states. The ansatz corresponds to a trainable BS quantum processor [45,46] and can synthesize various soliton solutions. The trained NN is used to compute their quantum features, showing that soliton bound states exhibit entanglement. Also, the quantum variational algorithm is used to sample the probability distribution at specific patterns of output particles (i.e., BS) that reveal the generation of correlated bosons. The approach mixing phase space methods and machine learning turns out to be a remarkable tool for studying quantum solitons, and to predict and quantify effects not accessible with conventional techniques.

This paper is organized as follows. In Secs. II and III, we review the basics of phase space representations, characteristic functions, and outline the link with conventional NN models. In Secs. IV to X, we show how a Gaussian state can be reinterpreted as a stack of linear layers with an Gaussian activation function; we discuss how to map different quantum operators to layers in a NN. Then, in Secs. XI and XII, we represent the Gaussian ansatz for a quantum soliton as a quantum processor with trainable parameters and compute different soliton solutions. Finally, in Secs. XII and XIV, we

*claudio.conti@uniroma1.it
†https://www.complexlight.org.

use the trained neural network for various quantum features, as entanglement and BS. Conclusions are drawn in Sec. XV.

II. QUANTUM PHASE SPACE AND NEURAL NETWORKS

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 $N = 2n$ real variables. Among the many representations [47], we consider the characteristic function $\chi(\mathbf{x})$.

For Gaussian states [48], in symmetric ordering,

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

with $\boldsymbol{g}$ the $N\times N$ covariance matrixx and $\boldsymbol{d}$ the $N\times 1$ displacement vector. $\chi(\mathbf{x})$ is a complex function $\chi(\mathbf{x}) = \chi_R(\mathbf{x}) + i\chi_I(\mathbf{x})$ with real part χ_R and imaginary part χ_I. Any many-body state is represented by a couple of real functions $\chi_{R,I}$ of N real variables. As NN 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. The resulting NN 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 NN architecture that can be trained by well-known algorithms.

III. THE CHARACTERISTIC FUNCTION

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

$$\boldsymbol{z} = (z_0, z_0, \ldots, z_{n-1}, z_0^*, z_1^*, \ldots, z_{n-1}^*), \tag{2}$$

which include the complex z_j components and their conjugates z_j^*. Given the density matrix ρ, the characteristic function is [47]

$$\chi(\boldsymbol{z},\mathcal{P}) = \mathrm{Tr}\left\{\rho e^{\sum_k (z_k\hat{a}_k^\dagger - z_k^*\hat{a}_k)}\right\} e^{\sum_k \frac{\mathcal{P}}{2} z_k z_k^*}. \tag{3}$$

In Eq. (3), the variable $\mathcal{P}\in\{0,1,-1\}$ 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 values as derivatives of χ.

For an observable $\hat{O}$, we have

$$\langle\hat{O}\rangle = \mathrm{Tr}(\rho\hat{O}). \tag{4}$$

The expected value of a combination of the annihilation and creation operators is

$$\langle(\hat{a}_j^\dagger)^m(\hat{a}_k)^n\rangle_{\mathcal{P}} = \left(\frac{\partial}{\partial z_j}\right)^m\left(-\frac{\partial}{\partial z_k^*}\right)^n \chi(\boldsymbol{z},\mathcal{P})\Big|_{z=0}, \tag{5}$$

with m and n integers.

The mean value of the field operator

$$\langle\hat{a}_k\rangle = -\frac{\partial\chi}{\partial z_k^*} \tag{6}$$

does not depend on the ordering index $\mathcal{P}$, as also

$$\langle\hat{a}_k^\dagger\rangle = \frac{\partial\chi}{\partial z_k}. \tag{7}$$

The mean particle number of mode k, $\hat{n}_k = \hat{a}_k^\dagger\hat{a}_k$ is

$$\langle\hat{a}_k^\dagger\hat{a}_k\rangle_{\mathcal{P}} = -\frac{\partial^2\chi(\boldsymbol{z},\mathcal{P})}{\partial z_k\partial z_k^*}\Big|_{z=0}. \tag{8}$$

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

$$\langle\hat{a}_k^\dagger\hat{a}_k\rangle_{\mathcal{P}=0} = \langle\hat{a}_k^\dagger\hat{a}_k + \hat{a}_k\hat{a}_k^\dagger\rangle; \tag{9}$$

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

$$\langle\hat{a}_k^\dagger\hat{a}_k\rangle_{\mathcal{P}=1} = \langle\hat{a}_k^\dagger\hat{a}_k\rangle, \tag{10}$$

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

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

In general,

$$\langle\hat{a}_k^\dagger\hat{a}_k\rangle_{\mathcal{P}} = \langle\hat{a}_k^\dagger\hat{a}_k\rangle + \tfrac{1}{2}(1-\mathcal{P}), \tag{12}$$

which is adopted below for the NN layer that returns the mean particle number. When not explicitly stated, we refer to symmetric ordering $\mathcal{P}=0$.

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 $\hat{p}_k$ defined by

$$\hat{a}_k = \frac{\hat{q}_k + i\hat{p}_k^\dagger}{\sqrt{2}}. \tag{13}$$

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

$$\hat{\boldsymbol{q}} = \begin{pmatrix}\hat{q}_0\\ \hat{q}_1\\ \vdots\\ \hat{q}_{n-1}\end{pmatrix}, \quad \hat{\boldsymbol{p}} = \begin{pmatrix}\hat{p}_0\\ \hat{p}_1\\ \vdots\\ \hat{p}_{n-1}\end{pmatrix}. \tag{14}$$

In addition, we consider the real quantities

$$q_k = \frac{z_k - z_k^*}{\sqrt{2}i}, \quad p_k = \frac{z_k + z_k^*}{\sqrt{2}}, \tag{15}$$

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

$$\begin{aligned}\boldsymbol{q} &= (q_0 \quad q_1 \quad \cdots \quad q_{n-1}),\\ \boldsymbol{p} &= (p_0 \quad p_1 \quad \cdots \quad p_{n-1}).\end{aligned} \tag{16}$$

Correspondingly, we have

$$\chi = \chi(\boldsymbol{q},\boldsymbol{p},\mathcal{P}=0) = \mathrm{Tr}[\rho\exp(i\,\boldsymbol{q}\cdot\hat{\boldsymbol{q}} + i\,\boldsymbol{p}\cdot\hat{\boldsymbol{p}})]. \tag{17}$$

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

$$\chi(\boldsymbol{q},\boldsymbol{p}) = \mathrm{Tr}[\rho\exp(i\,\boldsymbol{q}\,\hat{\boldsymbol{q}} + i\,\boldsymbol{p}\,\hat{\boldsymbol{p}})]. \tag{18}$$

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

$$\mathbf{x} = (q_0 \quad p_0 \quad q_1 \quad p_2 \quad \cdots q_n \quad p_n), \tag{19}$$

and we have

$$\chi = \chi(\mathbf{x}) = \mathrm{Tr}[\rho\exp(i\,\mathbf{x}\,\hat{\mathbf{R}})] = \chi_R(\mathbf{x}) + i\chi_I(\mathbf{x}) \tag{20}$$

being

$$\begin{aligned}\mathbf{x}\,\hat{\mathbf{R}} &= q_0\hat{q}_0 + p_0\hat{p}_0 + q_1\hat{q}_1 + \cdots + q_{n-1}\hat{q}_{n-1} + p_{n-1}\hat{p}_{n-1}\\ &= x_0\hat{R}_0 + x_1\hat{R}_1 + \cdots + x_{N-1}\hat{R}_{N-1},\end{aligned} \tag{21}$$

with the $N \times 1$ column operator vector:

$$\hat{\mathbf{R}} = \begin{pmatrix} \hat{q}_0 \\ \hat{p}_0 \\ \hat{q}_1 \\ \hat{p}_1 \\ \vdots \\ \hat{q}_{n-1} \\ \hat{p}_{n-1} \end{pmatrix}. \tag{22}$$

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

$$\begin{aligned}[\hat{q}_j, \hat{q}_k] &= 0,\\ [\hat{p}_j, \hat{p}_k] &= 0,\\ [\hat{q}_j, \hat{p}_k] &= i\delta_{jk},\end{aligned} \tag{23}$$

we have

$$[\hat{R}_p, \hat{R}_q] = iJ_{pq}, \tag{24}$$

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

$$\mathbf{J} = \bigoplus_j \mathbf{J}_1 = \begin{pmatrix} 0 & 1 & 0 & 0 & \cdots & 0 & 0 \\ -1 & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & 1 & \cdots & 0 & 0 \\ 0 & 0 & -1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 0 & 1 \\ 0 & 0 & 0 & 0 & \cdots & -1 & 0 \end{pmatrix}, \tag{25}$$

being $\mathbf{J}_1 = \left(\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right)$ [48]. In Eq. (25), the symbol $\bigoplus$ is the direct sum for block matrices with j running in $0, 1, \ldots n-1$. The matrix $\mathbf{J}$ is such that $\mathbf{J}^{-1} = \mathbf{J}^\top$, $\mathbf{J}^2 = -\mathbf{1}_N$, and $\mathbf{J}^\top\mathbf{J} = \mathbf{1}_N$ is the $N \times N$ identity matrix.

The expected value of $\hat{\boldsymbol{R}}$ is determined by the derivatives of the characteristic function,

$$\chi(\mathbf{x}) = \mathrm{Tr}(\rho e^{i\,\mathbf{x}\,\hat{\mathbf{R}}}). \tag{26}$$

One has

$$\langle \hat{R}_j \rangle = \mathrm{Tr}[\rho \hat{R}_j] = -i\,\frac{\partial \chi}{\partial x_j}\bigg|_{\mathbf{x}=0} = \frac{\partial \chi_I}{\partial x_j}\bigg|_{\mathbf{x}=0} \tag{27}$$

or

$$\langle \hat{\mathbf{R}} \rangle = \mathrm{Tr}[\rho \hat{\mathbf{R}}] = -i\,\boldsymbol{\nabla}\chi|_{\mathbf{x}=0} = \boldsymbol{\nabla}\chi_I|_{\mathbf{x}=0}. \tag{28}$$

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

$$\begin{aligned}&-\frac{\partial^2}{\partial z_j \partial z_j^*}\chi\\ &= -\left(\frac{\partial}{\partial q_j}\frac{\partial q_j}{\partial z_j} + \frac{\partial}{\partial p_j}\frac{\partial p_j}{\partial z_j}\right)\left(\frac{\partial}{\partial q_j}\frac{\partial q_j}{\partial z_j^*} + \frac{\partial}{\partial p_j}\frac{\partial p_j}{\partial z_j^*}\right)\chi\\ &= -\frac{1}{2}\left(\frac{\partial^2}{\partial q_j^2} + \frac{\partial^2}{\partial p_j^2}\right)\chi,\end{aligned} \tag{29}$$

after using Eqs. (15). Thus, we evaluate the expected number of photons by the second derivatives of χ_R,

$$\langle \hat{a}_j^\dagger \hat{a}_j \rangle_{\mathcal{P}=0} = -\frac{1}{2}\left(\frac{\partial^2}{\partial x_{2j}^2} + \frac{\partial^2}{\partial x_{2j+1}^2}\right)\chi_R\bigg|_{\mathbf{x}=0}, \tag{30}$$

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

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

$$\langle \hat{\mathcal{N}} \rangle_{\mathcal{P}=0} = -\tfrac{1}{2}\nabla^2\chi_R\big|_{\mathbf{x}=0}, \tag{31}$$

the Laplacian being computed with respect to all the components of $\mathbf{x}$. Equation (31) refers to symmetric ordering with $\mathcal{P} = 0$. For normal ordering, we use Eq. (12) as

$$\langle \hat{a}_k^\dagger \hat{a}_k \rangle = \langle \hat{a}_k^\dagger \hat{a}_k \rangle_{\mathcal{P}=1} = \langle \hat{a}_k^\dagger \hat{a}_k \rangle_{\mathcal{P}=0} - \tfrac{1}{2}, \tag{32}$$

and we get

$$\begin{aligned}\langle \hat{\mathcal{N}} \rangle &= \langle \hat{\mathcal{N}} \rangle_{\mathcal{P}=1} = \langle \hat{\mathcal{N}} \rangle_{\mathcal{P}=0} - \frac{n}{2}\\ &= -\frac{1}{2}\nabla^2\chi_R\bigg|_{\mathbf{x}=0} - \frac{n}{2}\\ &= -\frac{1}{2}(\nabla^2 + n)\chi|_{\mathbf{x}=0}.\end{aligned} \tag{33}$$

IV. GAUSSIAN STATES

For a Gaussian state,

$$\chi(\mathbf{x}) = \exp\left(-\tfrac{1}{4}\mathbf{x}\boldsymbol{g}\mathbf{x}^\top + i\,\mathbf{x}\boldsymbol{d}\right), \tag{34}$$

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,

$$\begin{aligned}\chi_R(\mathbf{x}) &= e^{-\frac{\mathbf{x}\boldsymbol{g}\mathbf{x}^\top}{4}}\cos(\mathbf{x}\boldsymbol{d}),\\ \chi_I(\mathbf{x}) &= e^{-\frac{\mathbf{x}\boldsymbol{g}\mathbf{x}^\top}{4}}\sin(\mathbf{x}\boldsymbol{d}).\end{aligned} \tag{35}$$

From Eq. (28), we have for $\boldsymbol{d}$, with $q = 0, 1, \ldots, N-1$:

$$\langle \hat{R}_q \rangle = \mathrm{Tr}[\hat{\rho}\hat{R}_q] = d_q = \frac{\partial \chi_I}{\partial x_q}\bigg|_{\mathbf{x}=0} \tag{36}$$

or

$$\langle \mathbf{R} \rangle = \boldsymbol{d} = \boldsymbol{\nabla}\chi_I|_{\mathbf{x}=0}. \tag{37}$$

Hence, the displacement d_q is the first moment of the Gaussian χ. The derivatives of the characteristic function can be expressed by the displacement and covariance matrix d_q and g_{pq}. This speeds up computing the derivatives of χ and the observable quantities.

For the particle number, we have

$$\begin{aligned}\langle \hat{a}_j^\dagger \hat{a}_j \rangle_{\mathcal{P}=0} &= -\frac{1}{2}\left(\frac{\partial^2}{\partial x_{2j}^2} + \frac{\partial^2}{\partial x_{2j+1}^2}\right)\bigg|_{\mathbf{x}=0}\chi_R\\ &= \frac{g_{2j,2j} + g_{2j+1,2j+1}}{4} + \frac{d_{2j}^2 + d_{2j+1}^2}{2}\end{aligned} \tag{38}$$

and

$$\langle \hat{a}_j^\dagger \hat{a}_j \rangle = -\frac{1}{2}\left(\frac{\partial^2}{\partial x_{2j}^2} + \frac{\partial^2}{\partial x_{2j+1}^2}\right)\chi_R - \frac{1}{2}$$
$$= \frac{g_{2j,2j} + g_{2j+1,2j+1}}{4} + \frac{d_{2j}^2 + d_{2j+1}^2}{2} - \frac{1}{2}, \quad (39)$$

while the second derivatives of χ_I at $\mathbf{x} = 0$ are vanishing, and we used Eq. (12). The covariance matrix $\boldsymbol{g}$ is determined by the second moments:

$$g_{pq} = \langle \{\hat{R}_p, \hat{R}_q\} \rangle - 2\langle \hat{R}_p \rangle \langle \hat{R}_q \rangle, \quad (40)$$

with

$$\{\hat{R}_p, \hat{R}_q\} = \hat{R}_p \hat{R}_q + \hat{R}_q \hat{R}_p. \quad (41)$$

We also have after Eq. (24):

$$\begin{aligned} g_{pq} &= \langle \hat{R}_p \hat{R}_q \rangle + \langle \hat{R}_q \hat{R}_p \rangle - 2\langle R_p \rangle \langle R_q \rangle \\ &= \langle \hat{R}_p \hat{R}_q \rangle + \langle \hat{R}_q \hat{R}_p \rangle - 2\langle R_p \rangle \langle R_q \rangle \\ &= 2\langle \hat{R}_p \hat{R}_q \rangle - iJ_{pq} - 2\langle R_p \rangle \langle R_q \rangle \\ &= 2\langle (\hat{R}_p - d_p)(\hat{R}_q - d_q) \rangle - iJ_{pq}. \end{aligned} \quad (42)$$

A. Vacuum state

Vacuum states are Gaussian states with $\boldsymbol{d} = \mathbf{0}$ and $\boldsymbol{g} = \mathbf{1}_N$. From Eq. (39), we get

$$\langle \hat{a}_j^\dagger \hat{a}_j \rangle_{\mathcal{P}=0} = \tfrac{1}{2}. \quad (43)$$

Correspondingly,

$$\langle \hat{a}_j^\dagger \hat{a}_j \rangle = \langle \hat{a}_j^\dagger \hat{a}_j \rangle_{\mathcal{P}=0} - \tfrac{1}{2} = 0. \quad (44)$$

We have $\langle \hat{\mathbf{R}} \rangle = 0$, and for the average photon number after Eq. (39):

$$\langle \mathcal{N} \rangle = \sum_{j=0}^{n-1} \langle \hat{a}_j^\dagger \hat{a}_j \rangle = 0. \quad (45)$$

B. Coherent state

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

$$d_0 = \sqrt{2}\mathrm{Re}(\alpha) = \frac{\alpha + \alpha^*}{\sqrt{2}},$$
$$d_1 = \sqrt{2}\mathrm{Im}(\alpha) = \frac{\alpha - \alpha^*}{i\sqrt{2}}. \quad (46)$$

From Eq. (39), we have ($j = 0$):

$$\langle \hat{a}_j^\dagger \hat{a}_j \rangle = \langle \hat{a}_j^\dagger \hat{a}_j \rangle_{\mathcal{P}=0} - \frac{1}{2} = |\alpha|^2 = \frac{d_0^2 + d_1^2}{2}. \quad (47)$$

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

$$\langle \hat{a}_j^\dagger \hat{a}_j \rangle = \frac{d_{2j}^2 + d_{2j+1}^2}{2} \quad (48)$$

and

$$\langle \mathcal{N} \rangle = \sum_{j=0}^{n-1} \langle \hat{a}_j^\dagger \hat{a}_j \rangle = \sum_{q=0}^{N-1} \frac{d_q^2}{2}. \quad (49)$$

C. Covariance matrix

Given χ, one can compute the covariance matrix and the displacement operator by derivation. $\boldsymbol{d}$ can be computed by the first derivatives, as in Eq. (28):

$$d_q = \langle \hat{R}_q \rangle = \mathrm{Tr}[\hat{\rho}\hat{R}_q] = \left.\frac{\partial \chi_I}{\partial x_q}\right|_{\mathbf{x}=0}. \quad (50)$$

Writing the characteristic function in terms of the components of $\mathbf{x}$,

$$\chi(\mathbf{x}) = \exp\left(-\frac{1}{4}\sum_{pq} g_{pq} x_p x_q + i\sum_p x_p d_p\right), \quad (51)$$

with $p, q = 0, 1, ..., N-1$, we have for the first derivative, by using the symmetry $g_{pq} = g_{qp}$,

$$\frac{\partial \chi}{\partial x_m} = \left(-\frac{1}{2}\sum_p g_{mp} x_p + i d_m\right)\chi(\mathbf{x}), \quad (52)$$

with $m = 0, 1, ..., N-1$. Evaluating Eq. (52) at $\mathbf{x} = 0$, we obtain as above, being $\chi(\mathbf{0}) = 1$:

$$\left.\frac{\partial \chi}{\partial x_m}\right|_{\mathbf{x}=0} = i\left.\frac{\partial \chi_I}{\partial x_m}\right|_{\mathbf{x}=0} = i d_m. \quad (53)$$

For the second derivatives, we have

$$\left.\frac{\partial^2 \chi_R}{\partial x_m \partial x_n}\right|_{\mathbf{x}=0} = -\frac{1}{2} g_{mn} - d_m d_n,$$
$$\left.\frac{\partial^2 \chi_I}{\partial x_m \partial x_n}\right|_{\mathbf{x}=0} = 0. \quad (54)$$

The covariance matrix is given by

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

Equation (55) is helpful to compute the covariance matrix after the model is trained to determine various features as, e.g., the degree of entanglement.

V. GATES AS LINEAR LAYERS

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

$$\hat{\tilde{\boldsymbol{a}}} = \hat{U}^\dagger \hat{\boldsymbol{a}} \hat{U}, \quad (56)$$

in terms of the density matrix, the transformation reads [47]

$$\tilde{\rho} = \hat{U}\rho\,\hat{U}^\dagger. \quad (57)$$

We start considering a transformation such that [47]

$$\hat{\tilde{\boldsymbol{a}}} = \mathbf{U}\hat{\boldsymbol{a}}, \quad (58)$$

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

$$\hat{\tilde{\boldsymbol{a}}} = \mathbf{U}\hat{\boldsymbol{a}} + \mathbf{W}\hat{\boldsymbol{a}}^\dagger. \quad (59)$$

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

The transformation in terms of the canonical variables read [48]

$$\tilde{\hat{\mathbf{R}}} = \hat{U}^{\dagger}\hat{\mathbf{R}}\hat{U} = \mathbf{M}\,\hat{\mathbf{R}} + \boldsymbol{d}'. \tag{60}$$

Linear transformations transform Gaussian states into Gaussian states [48]. A Gaussian state with covariance matrix $\boldsymbol{g}$ and displacement vector $\boldsymbol{d}$ turns into a new Gaussian state with covariance matrix

$$\tilde{\boldsymbol{g}} = \mathbf{M}\boldsymbol{g}\mathbf{M}^{\top} \tag{61}$$

and displacement vector

$$\tilde{\boldsymbol{d}} = \mathbf{M}\boldsymbol{d} + \boldsymbol{d}'. \tag{62}$$

If the transformation due to $\mathbf{U}$ is unitary, the matrix $\mathbf{M}$ is symplectic, and satisfies

$$\mathbf{M}\mathbf{J}\mathbf{M}^{\top} = \mathbf{J}, \tag{63}$$

with $\mathbf{J}$ given in Eq. (25). The inverse of $\mathbf{M}$ is found as

$$\mathbf{M}^{-1} = \mathbf{J}\mathbf{M}^{\top}\mathbf{J}^{\top}. \tag{64}$$

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

$$\hat{\boldsymbol{a}} = \begin{pmatrix} \hat{a}_0 \\ \hat{a}_1 \\ \vdots \\ \hat{a}_{n-1} \end{pmatrix}, \tag{65}$$

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

$$\hat{\boldsymbol{q}} = \begin{pmatrix} \hat{q}_0 \\ \hat{q}_1 \\ \vdots \\ \hat{q}_{n-1} \end{pmatrix}, \quad \hat{\boldsymbol{p}} = \begin{pmatrix} \hat{p}_0 \\ \hat{p}_1 \\ \vdots \\ \hat{p}_{n-1} \end{pmatrix}. \tag{66}$$

We build the $\hat{\mathbf{R}}$ vector in Eq. (22) by using auxiliary rectangular matrices $\mathbf{R}_{\mathrm{q}}$ and $\mathbf{R}_{\mathrm{p}}$ as follows:

$$\hat{\mathbf{R}} = \mathbf{R}_{\mathrm{q}}\hat{\boldsymbol{q}} + \mathbf{R}_{\mathrm{p}}\hat{\boldsymbol{p}}. \tag{67}$$

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

For $N = 4$, we have

$$\mathbf{R}_{\mathrm{q}} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1 \\ 0 & 0 \end{pmatrix} \tag{68}$$

and

$$\mathbf{R}_{\mathrm{p}} = \begin{pmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{pmatrix}, \tag{69}$$

such that

$$\mathbf{R}_{\mathrm{q}}\hat{\mathbf{q}} = \begin{pmatrix} \hat{q}_0 \\ 0 \\ \hat{q}_1 \\ 0 \end{pmatrix} \tag{70}$$

and

$$\mathbf{R}_{\mathrm{p}}\hat{\mathbf{p}} = \begin{pmatrix} 0 \\ \hat{p}_0 \\ 0 \\ \hat{p}_1 \end{pmatrix}, \tag{71}$$

and hence

$$\hat{\mathbf{R}} = \mathbf{R}_{\mathrm{q}}\hat{\mathbf{q}} + \mathbf{R}_{\mathrm{p}}\hat{\boldsymbol{p}} = \begin{pmatrix} \hat{q}_0 \\ \hat{p}_0 \\ \hat{q}_1 \\ \hat{p}_1, \end{pmatrix}, \tag{72}$$

being

$$\hat{\boldsymbol{q}} = \begin{pmatrix} \hat{q}_0 \\ \hat{q}_1 \end{pmatrix} \tag{73}$$

and

$$\hat{\boldsymbol{p}} = \begin{pmatrix} \hat{p}_0 \\ \hat{p}_1 \end{pmatrix}. \tag{74}$$

From the previous expressions, one has

$$\begin{aligned} \hat{\boldsymbol{q}} &= \mathbf{R}_{\mathrm{q}}^{\top}\hat{\mathbf{R}}, \\ \hat{\boldsymbol{p}} &= \mathbf{R}_{\mathrm{p}}^{\top}\hat{\mathbf{R}}, \end{aligned} \tag{75}$$

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

$$\begin{aligned} \mathbf{R}_{\mathrm{q}}^{\top}\mathbf{R}_{\mathrm{q}} &= \mathbf{1}_n, \\ \mathbf{R}_{\mathrm{q}}^{\top}\mathbf{R}_{\mathrm{p}} &= \mathbf{0}_n, \\ \mathbf{R}_{\mathrm{p}}^{\top}\mathbf{R}_{\mathrm{p}} &= \mathbf{1}_n, \\ \mathbf{R}_{\mathrm{p}}^{\top}\mathbf{R}_{\mathrm{q}} &= \mathbf{0}_n, \\ \mathbf{R}_{\mathrm{q}}\mathbf{R}_{\mathrm{q}}^{\top} + \mathbf{R}_{\mathrm{p}}\mathbf{R}_{\mathrm{p}}^{\top} &= \mathbf{1}_N, \\ \mathbf{J}\mathbf{R}_{\mathrm{q}} &= -\mathbf{R}_{\mathrm{p}}, \\ \mathbf{J}\mathbf{R}_{\mathrm{p}} &= \mathbf{R}_{\mathrm{q}}, \\ \mathbf{R}_{\mathrm{q}}\mathbf{R}_{\mathrm{p}}^{\top} - \mathbf{R}_{\mathrm{p}}\mathbf{R}_{\mathrm{q}}^{\top} &= \mathbf{J}. \end{aligned} \tag{76}$$

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

$$\hat{\boldsymbol{a}} = \frac{\hat{\boldsymbol{q}} + i\hat{\boldsymbol{p}}}{\sqrt{2}} = \frac{\mathbf{R}_{\mathrm{q}}^{\top} + i\mathbf{R}_{\mathrm{p}}^{\top}}{\sqrt{2}}\hat{\mathbf{R}}. \tag{77}$$

Letting $\mathbf{U} = \mathbf{U}_R + i\mathbf{U}_I$ with real part $\mathbf{U}_R$ and imaginary part $\mathbf{U}_I$; and $\mathbf{U}^* = \mathbf{U}_R - i\mathbf{U}_I$. Correspondingly,

$$\begin{aligned} \tilde{\hat{\boldsymbol{q}}} &= \mathbf{U}_R\hat{\boldsymbol{q}} - \mathbf{U}_I\hat{\boldsymbol{p}}, \\ \tilde{\hat{\boldsymbol{p}}} &= \mathbf{U}_I\hat{\boldsymbol{q}} + \mathbf{U}_R\hat{\boldsymbol{p}}. \end{aligned} \tag{78}$$

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

$$\tilde{\hat{\mathbf{R}}} = \mathbf{R}_{\mathrm{q}}\tilde{\hat{\boldsymbol{q}}} + \mathbf{R}_{\mathrm{p}}\tilde{\hat{\boldsymbol{p}}} = \mathbf{M}\hat{\mathbf{R}} = (\mathbf{M}_1 + \mathbf{M}_2)\hat{\mathbf{R}}, \tag{79}$$

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

$$\begin{aligned} \mathbf{M} &= \mathbf{M}_1 + \mathbf{M}_2, \\ \mathbf{M}_1 &= \mathbf{R}_{\mathrm{q}}\mathbf{U}_R\mathbf{R}_{\mathrm{q}}^{\top} + \mathbf{R}_{\mathrm{p}}\mathbf{U}_R\mathbf{R}_{\mathrm{p}}^{\top}, \\ \mathbf{M}_2 &= \mathbf{R}_{\mathrm{p}}\mathbf{U}_I\mathbf{R}_{\mathrm{q}}^{\top} - \mathbf{R}_{\mathrm{q}}\mathbf{U}_I\mathbf{R}_{\mathrm{p}}^{\top}. \end{aligned} \tag{80}$$

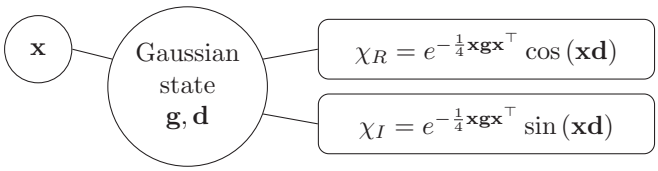

FIG. 1. Double-head model for a layer representing a Gaussian.

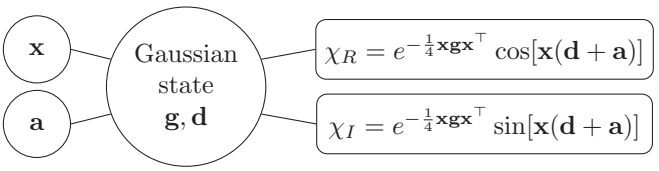

FIG. 2. Double-head model representing a Gaussian state with a bias.

Equations (80) are used in defining linear gates starting from their matrix **U**.

VI. GAUSSIAN STATES AS NEURAL NETWORKS

NNs are mathematical tools to approximate unknown functions of many variables. NNs depend on many parameters that can be fine-tuned to improve fitting a target model. Gaussian states correspond to NNs with a Gaussian activation function, which is often adopted, for example, in *kernel machines* [40]. We implement a Gaussian density matrix by a multihead 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 parts χ_R and χ_I.

Figure 1 shows the NN model for a Gaussian state with each layer having one input **x** and two outputs (the heads) χ_R and χ_I. The input is the state row vector **x**, and is the input layer. The outputs χ_R and χ_I are the output layers. The inner layer denoted Gaussian state computes the Gaussian characteristic function. The parameters of this layer, namel, the covariance matrix $\boldsymbol{g}$ and the displacement $\boldsymbol{d}$ are also indicated.

For a Gaussian state with covariance matrix **g** and displacement vector **d**, we have

$$\chi_R(\mathbf{x}) = \exp\left(-\tfrac{1}{4}\mathbf{x}\mathbf{g}\mathbf{x}^\top\right)\cos(\mathbf{x}\boldsymbol{d}),$$
$$\chi_I(\mathbf{x}) = \exp\left(-\tfrac{1}{4}\mathbf{x}\mathbf{g}\mathbf{x}^\top\right)\sin(\mathbf{x}\boldsymbol{d}). \tag{81}$$

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

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

$$\chi(x) = e^{-\frac{1}{4}\mathbf{x}\mathbf{g}\mathbf{x}^T} e^{i(\boldsymbol{d}+\boldsymbol{a})}. \tag{82}$$

Here $\boldsymbol{a}$ is a bias in the displacement, which will be useful for cascading layers as detailed below. Equation (82) corresponds to the two heads of the model, returning

$$\chi_R(\mathbf{x}) = \exp\left(-\tfrac{1}{4}\mathbf{x}\mathbf{g}\mathbf{x}^T\right)\cos[\mathbf{x}(\boldsymbol{d}+\mathbf{a})],$$
$$\chi_I(\mathbf{x}) = \exp\left(-\tfrac{1}{4}\mathbf{x}\mathbf{g}\mathbf{x}^T\right)\sin[\mathbf{x}(\boldsymbol{d}+\mathbf{a})]. \tag{83}$$

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

VII. PULLBACK

We set up our NN representation of the density matrix as a layered sequence of gates. Each gate is a unitary operator that acts on the density matrix and transforms the latter into a new state. To detail how it works, we start considering the transformation of the characteristic function under unitary operations.

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

$$\hat{\tilde{\boldsymbol{a}}} = \hat{U}^\dagger \hat{\boldsymbol{a}} \hat{U},$$
$$\tilde{\rho} = \hat{U}\rho\,\hat{U}^\dagger. \tag{84}$$

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

$$\tilde{\chi} = \mathrm{Tr}[\tilde{\rho}e^{i\mathbf{x}\hat{\mathbf{R}}}] = \mathrm{Tr}[\hat{U}\rho\hat{U}^\dagger e^{i\mathbf{x}\hat{\mathbf{R}}}] = \mathrm{Tr}[\rho\hat{U}^\dagger e^{i\mathbf{x}\hat{\mathbf{R}}}\hat{U}]$$
$$= \mathrm{Tr}[\rho e^{i\mathbf{x}\hat{U}^\dagger\hat{\mathbf{R}}\hat{U}}] = \mathrm{Tr}[\rho e^{i\mathbf{x}\hat{\tilde{\mathbf{R}}}}]. \tag{85}$$

Recalling Eq. (60),

$$\hat{\tilde{\mathbf{R}}} = \mathbf{M}\mathbf{R} + \boldsymbol{d}', \tag{86}$$

we remark that the linear transformation is determined by the symplectic matrix **M** and the displacement $\boldsymbol{d}'$. We write $\tilde{\chi}$ in terms of **M** and $\boldsymbol{d}'$, by using Eq. (86):

$$\tilde{\chi} = \mathrm{Tr}[\rho e^{i\mathbf{x}\mathbf{M}\hat{\mathbf{R}}}]e^{i\mathbf{x}\boldsymbol{d}'}. \tag{87}$$

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

$$\chi = \mathrm{Tr}[\rho e^{i\mathbf{x}\hat{\mathbf{R}}}]. \tag{88}$$

Thus, we have

$$\tilde{\chi}(\mathbf{x}) = \chi(\mathbf{x}\mathbf{M})e^{i\mathbf{x}\boldsymbol{d}'} = \chi(\mathbf{x}\mathbf{M})e^{i(\mathbf{x}\mathbf{M})(\mathbf{M}^{-1}\boldsymbol{d}')}. \tag{89}$$

From Eq. (89), 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:

$$\tilde{\chi}(\mathbf{x}) = \chi(\mathbf{y})e^{i\mathbf{y}\boldsymbol{a}}. \tag{90}$$

We give a graphical representation of the transformations in Eq. (90). We start considering a characteristic function with input vector **x** and multi-headed outputs χ_R and χ_I, as in Fig. 3. Then we generalize the model by introducing the bias vector $\boldsymbol{a}$ as in the previous section and shown in Fig. 4. Figure 5 represents the transformation in Eq. (90).

We introduce a new layer, which we call the linear layer, having as parameters the symplectic matrix **M** and the displacement $\boldsymbol{d}'$. The linear layer is shown in Fig. 6, and it has two inputs **x** and $\boldsymbol{a}$, and two outputs: $\mathbf{y} = \mathbf{x}\mathbf{M}$, a row vector with the same size of **x**, and a new displacement $\boldsymbol{b} = \mathbf{M}^{-1}(\boldsymbol{d}' + \boldsymbol{a})$, a column vector with the size of $\boldsymbol{d}$. The

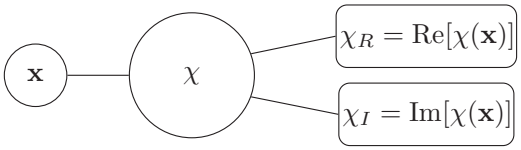

FIG. 3. Graphical representation of the model for the characteristic function.

linear layer will enable us to represent many different linear transformations as squeezing or Glauber operators.

By using the linear layer, we represent the transformation by cascading two layers. The resulting model is Fig. 7. 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. 8. This representation will be useful when considering multiple transformations.

For a Gaussian state [see Eq. (34)], the transformed characteristic function reads

$$\tilde{\chi}(\mathbf{x}) = e^{-\frac{1}{4}\mathbf{x}\mathbf{M}\mathbf{g}\mathbf{M}^\top\mathbf{x}^\top} e^{i\mathbf{x}\mathbf{M}(\boldsymbol{d}+\mathbf{M}^{-1}\boldsymbol{d}')}. \tag{91}$$

Equation (91) is graphically represented in Fig. 9 by using the Gaussian layer in Fig. 2. This representation is equivalent to Fig. 10, by the linear layer, whose action is detailed in Fig. 7. As shown in Fig. 10 linear transformation on the density matrix is actually a transformation of the variable $\mathbf{x}$. This implies that the linear transformation can be represented 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 Fig. 10.

VIII. PULLBACK CASCADING

The pullback 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 Fig. 11. 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

$$\hat{U}_1\rho\hat{U}_1^\dagger \tag{92}$$

and the new canonical vector is

$$\hat{R}_1 = \mathbf{M}_1\hat{R} + \boldsymbol{d}_1. \tag{93}$$

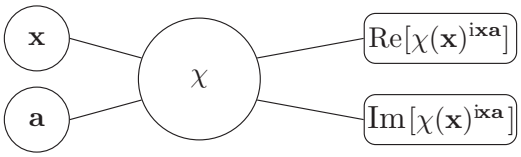

FIG. 4. Graphical representation of the generalized model for the characteristic function with bias $\boldsymbol{a}$, which is adopted in the linear transformations.

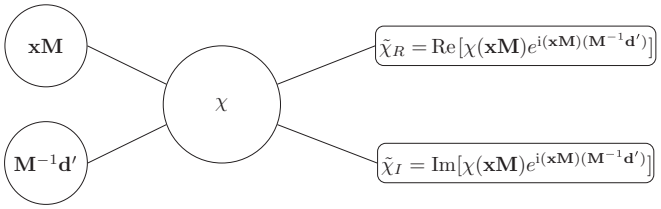

FIG. 5. Graphical representation of the transformed characteristic function in terms of the original characteristic function with modified input and bias.

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

$$\hat{U}_2\hat{U}_1\rho\hat{U}_1^\dagger\hat{U}_2^\dagger, \tag{94}$$

and the final vector of observables is

$$\hat{\mathbf{R}}_2 = \mathbf{M}_2\hat{\mathbf{R}}_1 + \boldsymbol{d}_2 = \mathbf{M}_2\mathbf{M}_1\hat{\mathbf{R}} + \boldsymbol{d}_2 + \mathbf{M}_2\boldsymbol{d}_1. \tag{95}$$

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

$$\begin{aligned}\mathbf{M} &= \mathbf{M}_2\mathbf{M}_1,\\ \boldsymbol{d}' &= \mathbf{M}_2\boldsymbol{d}_1 + \boldsymbol{d}_2.\end{aligned} \tag{96}$$

We have the following (Fig. 12): *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. 12, the output of the linear layer corresponding to $\hat{U}_2$ is

$$\begin{aligned}\mathbf{x}_2 &= \mathbf{x}\mathbf{M}_2,\\ \boldsymbol{a}_2 &= \mathbf{M}_2^{-1}(\boldsymbol{d}_2 + \boldsymbol{a}).\end{aligned} \tag{97}$$

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

$$\begin{aligned}\mathbf{x}_1 &= \mathbf{x}_2\mathbf{M}_1 = \mathbf{x}\mathbf{M}_2\mathbf{M}_1,\\ \boldsymbol{a}_1 &= \mathbf{M}_1^{-1}(\boldsymbol{d}_1 + \boldsymbol{a}_2) = \mathbf{M}_1^{-1}\big(\boldsymbol{d}_1 + \mathbf{M}_2^{-1}\boldsymbol{d}_2 + \mathbf{M}_2^{-1}\boldsymbol{a}\big).\end{aligned} \tag{98}$$

For the layer with parameters $(\mathbf{M}, \boldsymbol{d}')$, we have, see Eqs. (96):

$$\begin{aligned}\mathbf{y} &= \mathbf{x}\mathbf{M} = \mathbf{x}\mathbf{M}_2\mathbf{M}_1,\\ \boldsymbol{b} &= \mathbf{M}^{-1}(\boldsymbol{d}' + \boldsymbol{a})\\ &= \mathbf{M}_1^{-1}\mathbf{M}_2^{-1}(\mathbf{M}_2\boldsymbol{d}_1 + \boldsymbol{d}_2 + \boldsymbol{a})\\ &= \mathbf{M}_1^{-1}\big(\boldsymbol{d}_1 + \mathbf{M}_2^{-1}\boldsymbol{d}_2 + \mathbf{M}_2^{-1}\boldsymbol{a}\big).\end{aligned} \tag{99}$$

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

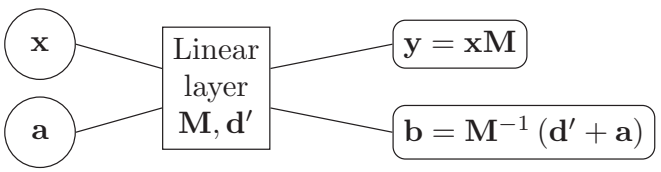

FIG. 6. Linear layer to transform the input variables to the χ layer.

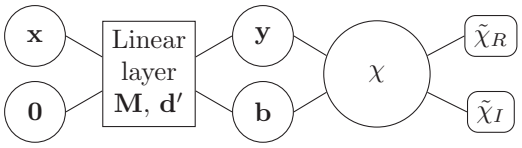

FIG. 7. Graphical representation of a linear transformation as a pullback of a linear layer from the original characteristic function. The resulting model corresponds the transformed characteristic function in Fig. 5. Note that here $\boldsymbol{a} = 0$.

By extending the previous argument to three or more transformations, one realizes that the cascade of an arbitrary number of transformations corresponds to a cascade of pullbacks in reverse order. For M transformations, reverse order means that one first makes a pullback of operator 1, then operator 2, etc. In the flow of data in the network, the input $\mathbf{x}$ first enters operator M, then $M-1$ and so forth to passing through the linear layer corresponding to the transformation 1. The case $M = 3$ is shown as an example in Fig. 13.

IX. 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 [47]

$$\hat{\mathcal{D}}(\alpha) = \exp(\alpha^* \hat{a} - \alpha \hat{a}^\dagger). \tag{100}$$

A coherent state is obtained by displacing the vacuum state:

$$|\alpha\rangle = \hat{\mathcal{D}}(\alpha)|0\rangle. \tag{101}$$

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

$$\tilde{\hat{a}} = \hat{U}^\dagger \hat{a} \hat{U} = \hat{a} + \alpha, \tag{102}$$

which implies for the canonical vector

$$\tilde{\hat{\mathbf{R}}} = \hat{U}^\dagger \hat{\mathbf{R}} \hat{U} = \begin{pmatrix} \hat{x} \\ \hat{p} \end{pmatrix} + \begin{pmatrix} d_0 \\ d_1 \end{pmatrix}, \tag{103}$$

with

$$d_0 = \sqrt{2}\mathrm{Re}(\alpha),$$
$$d_1 = \sqrt{2}\mathrm{Im}(\alpha). \tag{104}$$

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

$$\tilde{\hat{\boldsymbol{a}}} = \hat{U}^\dagger \hat{\boldsymbol{a}} \hat{U} = \hat{\boldsymbol{a}} + \boldsymbol{\alpha}, \tag{105}$$

which implies for the canonical vector

$$\tilde{\hat{\mathbf{R}}} = \hat{U}^\dagger \hat{\mathbf{R}} \hat{U} = \hat{\mathbf{R}} + \boldsymbol{d}, \tag{106}$$

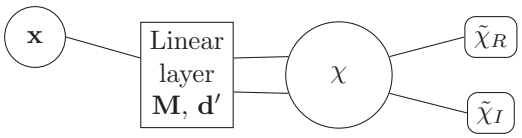

FIG. 8. Simplified model of Fig. 7, omitting the internal variables $\mathbf{y}$ and $\boldsymbol{b}$ and the zero input bias $\boldsymbol{a}$.

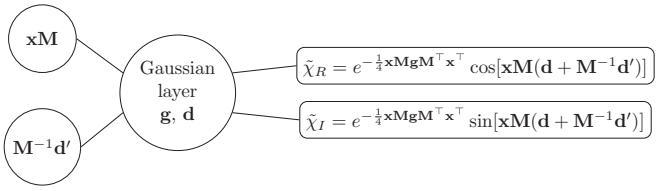

FIG. 9. Single-layer multiheaded model for a linear transformation of a Gaussian state.

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

$$d_{2j} = \sqrt{2}\mathrm{Re}(\alpha_j),$$
$$d_{2j+1} = \sqrt{2}\mathrm{Im}(\alpha_j) \tag{107}$$

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

To create a NN model that represents a coherent state with a given displacement vector $\boldsymbol{d}_{\text{target}}$, 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 $\boldsymbol{d}_{\text{target}}$. No bias is needed (see Fig. 14 with $\boldsymbol{a} = \mathbf{0}$).

X. THE SQUEEZING LAYER

Different nonclassical states are generated from vacuum by unitary operators resulting into the following linear relation:

$$\tilde{\hat{\boldsymbol{a}}} = \mathbf{U}\hat{\boldsymbol{a}} + \mathbf{W}\hat{\boldsymbol{a}}^\dagger, \tag{108}$$

which generalizes Eq. (58). Using the matrices $\mathbf{U}$ and $\mathbf{W}$, one obtains the corresponding symplectic matrix $\mathbf{M}$.

Let

$$\mathbf{U} = \mathbf{U}_R + i\mathbf{U}_I \tag{109}$$

and

$$\mathbf{W} = \mathbf{W}_R + i\mathbf{W}_I. \tag{110}$$

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

$$\tilde{\hat{\boldsymbol{a}}} = \frac{\tilde{\hat{\boldsymbol{q}}} + i\tilde{\hat{\boldsymbol{p}}}}{\sqrt{2}} = \mathbf{U}\hat{\boldsymbol{a}} + \mathbf{W}\hat{\boldsymbol{a}}^\dagger, \tag{111}$$

$$\tilde{\hat{\boldsymbol{a}}}^\dagger = \frac{\tilde{\hat{\boldsymbol{q}}} - i\tilde{\hat{\boldsymbol{p}}}}{\sqrt{2}} = \mathbf{U}^*\hat{\boldsymbol{a}}^\dagger + \mathbf{W}^*\hat{\boldsymbol{a}}. \tag{112}$$

We have

$$\tilde{\hat{\boldsymbol{q}}} = (\mathbf{U}_R + \mathbf{W}_R)\hat{\boldsymbol{q}} + (-\mathbf{U}_I + \mathbf{W}_I)\hat{\boldsymbol{p}},$$
$$\tilde{\hat{\boldsymbol{p}}} = (\mathbf{U}_I + \mathbf{W}_I)\hat{\boldsymbol{q}} + (\mathbf{U}_R - \mathbf{W}_R)\hat{\boldsymbol{p}}, \tag{113}$$

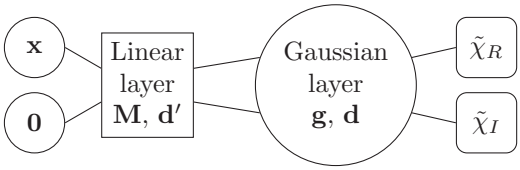

FIG. 10. Two-layer multiheaded model equivalent to Fig. 9.

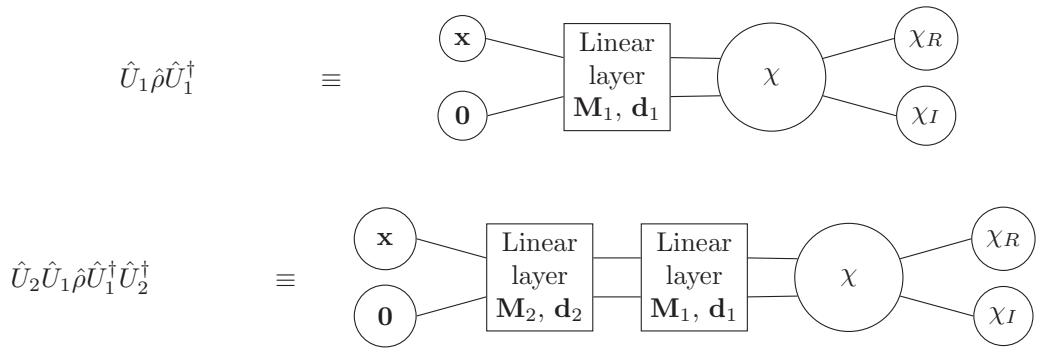

FIG. 11. Linear transformations and pullbacks. A single transformation with unitary operator $\hat{U}_1$ corresponds to a single pullback. A double transformation with unitary operators $\hat{U}_1$ and $\hat{U}_2$ corresponds to a double pullback. Note that the flow of data from $\mathbf{x}$ to the output through the network is in reverse order with respect to the two transformations.

which generalize Eqs. (78). Following the same arguments for Eqs. (80), we have for the symplectic matrix

$$\begin{aligned}\mathbf{M} &= \mathbf{M}_1 + \mathbf{M}_2,\\ \mathbf{M}_1 &= \mathbf{R}_{\rm q}(\mathbf{U}_R + \mathbf{W}_R)\mathbf{R}_{\rm q}^\top + \mathbf{R}_{\rm p}(\mathbf{U}_R - \mathbf{W}_R)\mathbf{R}_{\rm p}^\top,\\ \mathbf{M}_2 &= \mathbf{R}_{\rm p}(\mathbf{U}_I + \mathbf{W}_I)\mathbf{R}_{\rm q}^\top + \mathbf{R}_{\rm q}(-\mathbf{U}_I + \mathbf{W}_I)\mathbf{R}_{\rm p}^\top. \end{aligned} \quad (114)$$

Equation (114) is used for programming layers representing squeezing operators.

A. Single-mode squeezed state

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

$$\hat{\mathbf{R}} = \begin{pmatrix}\hat{x}\\ \hat{p}\end{pmatrix}. \quad (115)$$

Using the squeezing operator with parameters r and θ [47],

$$\tilde{\hat{a}} = \hat{S}^\dagger\hat{a}\hat{S} = \cosh(r)\hat{a} - e^{i\theta}\sinh(r)\hat{a}^\dagger, \quad (116)$$

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

$$\begin{aligned} U_R &= \cosh(r),\\ U_I &= 0,\\ W_R &= -\cos(\theta)\sinh(r),\\ W_I &= -\sin(\theta)\sinh(r), \end{aligned} \quad (117)$$

and, by Eqs. (114), we find the symplectic operator for the squeezing:

$$\mathbf{M}_s(r,\theta) = \begin{pmatrix}\cosh(r) - \cos(\theta)\sinh(r) & -\sin(\theta)\sinh(r)\\ -\sin(\theta)\sinh(r) & \cosh(r) + \cos(\theta)\sinh(r)\end{pmatrix}. \quad (118)$$

For $\theta = 0$, i.e., for a real squeezing parameter, we have

$$\mathbf{M}_s(r,0) = \begin{pmatrix}\exp(-r) & 0\\ 0 & \exp(r)\end{pmatrix}. \quad (119)$$

B. Multimode squeezed vacuum model

We are interested to multimode systems, so we consider an $n-$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. (118) for the mode to be squeezed denoted by an index $n_{\rm squeezed}$ in the range 0 to $n-1$ (see Fig. 15). For example, for $n = 4$, and $n_{\rm squeezed} = 0$, we have

$$\mathbf{M} = \begin{pmatrix} M_{s,11} & M_{s,12} & 0 & 0 & 0 & 0 & 0 & 0\\ M_{s,21} & M_{s,22} & 0 & 0 & 0 & 0 & 0 & 0\\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0\\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0\\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}. \quad (120)$$

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

$$\mathbf{M} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0\\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & M_{s,11} & M_{s,12} & 0 & 0\\ 0 & 0 & 0 & 0 & M_{s,21} & M_{s,22} & 0 & 0\\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0\\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}. \quad (121)$$

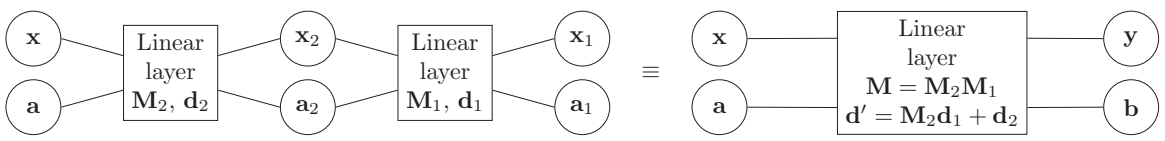

FIG. 12. Equivalence of two cascaded linear layers with a single linear layer.

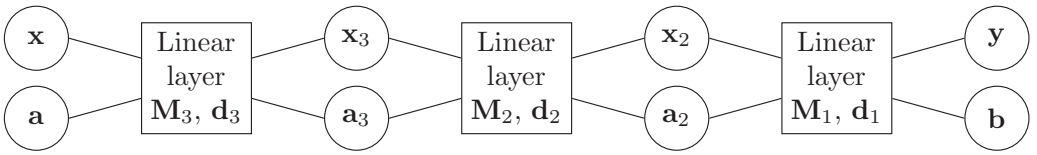

FIG. 13. Example of cascading three transformations. Note the reverse order of layers from right to left.

C. Squeezed coherent states

The squeezed operator is represented by a linear gate we can cascade with other layers, suxh as the Glauber displacement layer. Figure 16(a) 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 $\zeta = r\,e^{i\theta}$ is the complex squeezing parameter, that is,

$$|\alpha,\zeta\rangle = \hat{\mathcal{D}}(\alpha)\hat{\mathcal{S}}(\zeta)|0\rangle. \tag{122}$$

The resulting state has $\langle\hat{a}\rangle = \alpha$, and it is squeezed.

D. Squeezing the displaced vacuum

A different squeezed coherent state is obtained by squeezing a coherent state. We first pull back a displacement layer from the vacuum to create the coherent state, and then pull back a squeezing layer, as shown in Fig. 16(b). This corresponds to the following Eq. (123):

$$|\alpha\cosh(r) - \alpha^* e^{i\theta}\sinh(r),\zeta\rangle = \hat{\mathcal{S}}(\zeta)\hat{\mathcal{D}}(\alpha)|0\rangle. \tag{123}$$

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 [47], i.e.,

$$\langle\hat{a}\rangle = \alpha\cosh(r) - \alpha^* e^{i\theta}\sinh(r). \tag{124}$$

XI. COMPUTING THE HAMILTONIAN

In this paper, we are interested in the state of many-body Hamiltonians, which correspond to classical solitons. We consider the Hamiltonian

$$\hat{H} = -\sum_j [\hat{\psi}_j^\dagger\hat{\psi}_{j+1} + \hat{\psi}_j^\dagger\hat{\psi}_{j-1}] + \frac{\gamma}{2}\hat{\psi}_j^\dagger\hat{\psi}_j^\dagger\hat{\psi}_j\hat{\psi}_j = \hat{K} + \hat{V}, \tag{125}$$

with the potential energy

$$\hat{V} = \frac{\gamma}{2}\sum_j (\hat{n}_j^2 - \hat{n}_j), \tag{126}$$

being $\hat{n}_i = \hat{\psi}_i^\dagger\hat{\psi}_i$, and the kinetic energy

$$\hat{K} = \sum_{ij}\omega_{ij}\hat{\psi}_i^\dagger\hat{\psi}_j, \tag{127}$$

being $\omega_{ij} = -\delta_{i,j-1} - \delta_{i,j+1}$ with i, j in $[0, n-1]$ and with homogeneous boundary conditions, $\omega_{0,-1} = \omega_{n-1,n} = 0$.

Starting from the vacuum state, one can build the NN model of an arbitrary state by multiple pullbacks. We compute the mean value of observable quantities as $\hat{H}$ and $\hat{\mathcal{N}}$ as derivatives of χ at $\mathbf{x} = 0$. We have in symmetric ordering [47]

$$\langle\hat{K}\rangle = \sum_{jk}\omega_{jk}\left[\frac{\partial^2}{\partial\alpha_j\partial(-\alpha_k^*)}\chi(\boldsymbol{\alpha}) - \frac{\delta_{jk}}{2}\chi\right]\Bigg|_{\boldsymbol{\alpha}=0}, \tag{128}$$

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

$$\langle\hat{V}\rangle = \frac{\gamma}{2}\sum_j \frac{\partial^4\chi}{\partial\alpha_j^2\partial(-\alpha_j^*)^2} - 2\frac{\partial^2\chi}{\partial\alpha_j\partial(-\alpha_j^*)} + \frac{\chi}{2}\Bigg|_{\boldsymbol{\alpha}=0}. \tag{129}$$

These quantities can be computed by using automatic differentiation on the NN model.

Specifically,

$$\begin{aligned}\langle\hat{K}\rangle = -\frac{1}{2}\sum_{mn} &\times\left(\frac{\partial^2\chi}{\partial q_m\partial q_n} + \frac{\partial^2\chi}{\partial p_m\partial p_n} + \chi\right)\omega_{mn}^R \\ &+ \left(\frac{\partial^2\chi}{\partial q_m\partial p_n} - \frac{\partial^2\chi}{\partial p_n\partial p_q}\right)\omega_{mn}^I\Bigg|_{\mathbf{x}=0},\end{aligned} \tag{130}$$

where

$$\omega_{mn} = \omega_{mn}^R + i\omega_{mn}^I, \tag{131}$$

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

For the potential energy,

$$\langle\hat{V}\rangle = \sum_{nm} V_{nm}\langle\hat{a}_n^\dagger\hat{a}_m^\dagger\hat{a}_n\hat{a}_m\rangle, \tag{132}$$

being, for a local interaction:

$$V_{nm} = \tfrac{1}{2}\delta_{nm}. \tag{133}$$

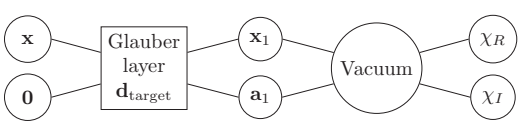

FIG. 14. Model for a coherent state obtained by pullback of a displacement operator (Glauber layer) from the vacuum.

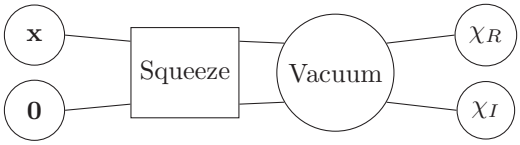

FIG. 15. A squeezed coherent state obtained by pullback with a single mode squeezing layer from the vacuum.

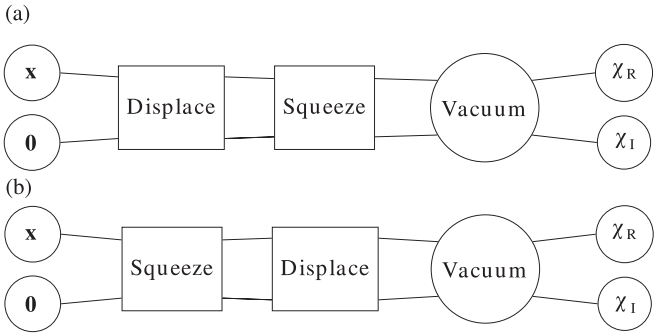

FIG. 16. Pullback model for a squeezed coherent state by displacing a squeezed vacuum (top panel). Pullback model for squeezed coherent state by squeezing a displaced vacuum (bottom panel). Here all the layers act on the same mode.

For a mode with index $j = 0, 1, \ldots, n-1$, the normal ordering product is

$$\langle \hat{n}_j^2 \rangle = \langle \hat{a}_j^\dagger \hat{a}_j \hat{a}_j^\dagger \hat{a}_j \rangle = \langle \hat{a}_j^\dagger \hat{a}_j^\dagger \hat{a}_j \hat{a}_j \rangle + \langle \hat{a}_j^\dagger \hat{a}_j \rangle, \tag{134}$$

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

$$\langle \hat{a}_j^\dagger \hat{a}_j \rangle = \langle \hat{n}_j \rangle = -\frac{\partial^2 \chi}{\partial z_j \, \partial z_j^*}\bigg|_{z=0} = -\frac{1}{2}\left(\partial_{q_j}^2 + \partial_{p_j}^2\right)\chi_R\Big|_{\mathbf{x}=0} - \frac{1}{2} \tag{135}$$

and

$$\langle \hat{a}_j^\dagger \hat{a}_j^\dagger \hat{a}_j \hat{a}_j \rangle = \left(\frac{\partial}{\partial z_j}\right)^2 \left(-\frac{\partial}{\partial z_j^*}\right)^2 \chi\Big|_{\mathbf{x}=0} - 2\left(\frac{\partial}{\partial z_j}\right)\left(-\frac{\partial}{\partial z_j^*}\right)\chi\Big|_{\mathbf{x}=0} + \frac{1}{2}. \tag{136}$$

Equation (136) can be expressed in terms of the derivatives with respect to q_j and p_j as follows:

$$\langle \hat{a}_j^\dagger \hat{a}_j^\dagger \hat{a}_j \hat{a}_j \rangle = \frac{1}{4}\left(\partial_{q_j}^2 + \partial_{p_j}^2\right)^2 \chi_R\Big|_{\mathbf{x}=0} + \left(\partial_{q_j}^2 + \partial_{p_j}^2\right)\chi_R\big|_{\mathbf{x}=0} + \frac{1}{2}, \tag{137}$$

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

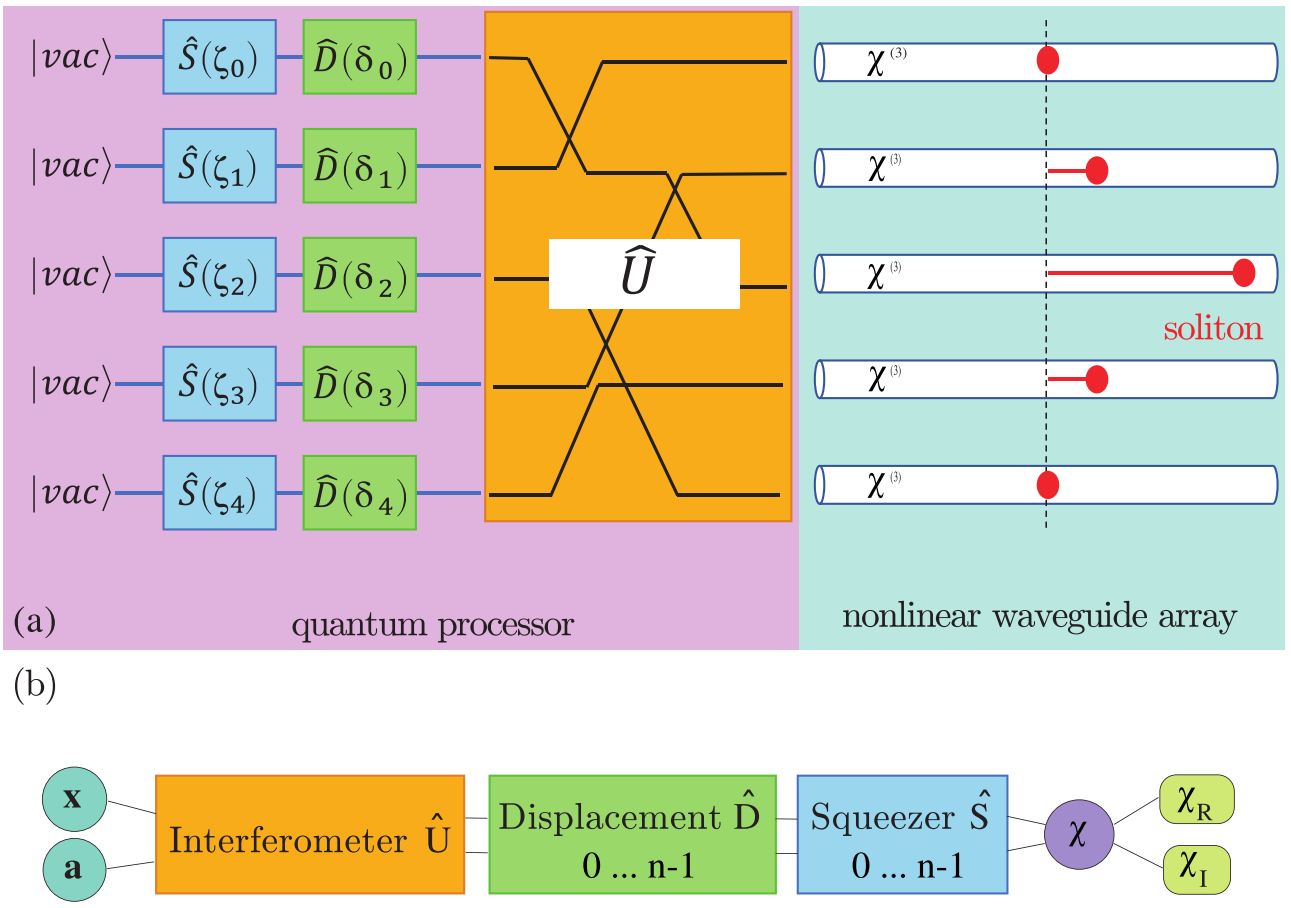

FIG. 17. (a) Quantum processor for solitons in coupled optical waveguides. n vacuum spatial modes are squeezed, displaced, and mixed by a trainable interferometer $\hat{U}$ ($n = 5$ in the figure). (b) Multiple pullbacks as a quantum soliton variational ansatz.

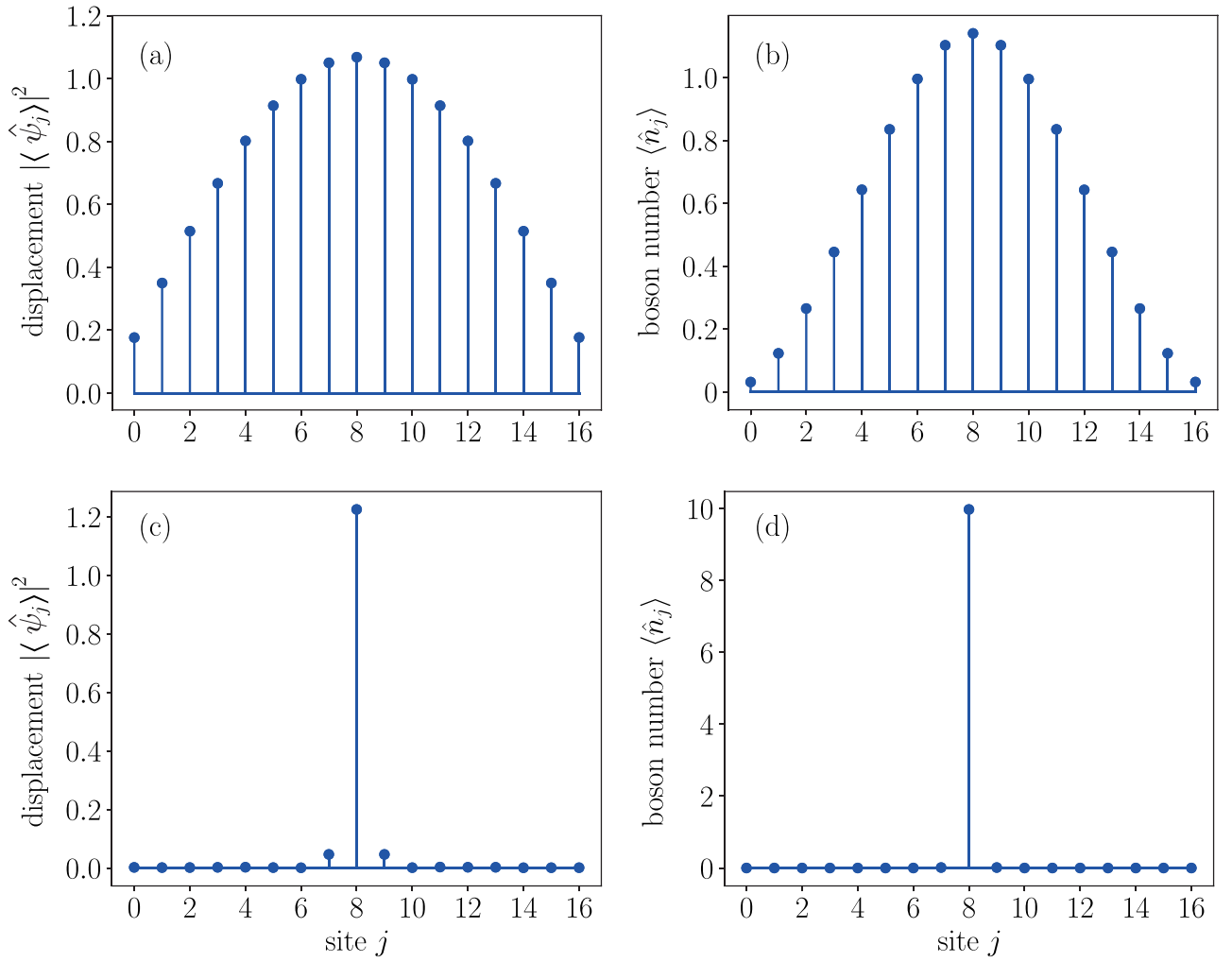

FIG. 18. (a), (b) Ground state for $N_T = 10$ and $\gamma = -0.01$: (a) displacement $|\langle\hat{\psi}_j\rangle|^2$; (b) mean boson number $\langle\hat{n}_j\rangle$; (c), (d) as in (a), (b) with $N_T = 10$ and $\gamma = -1$ ($n = 17$).

The real part χ_R enters in Eq. (137) as $\langle\hat{n}_j^2\rangle$ in Eq. (134) 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 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):

$$\begin{aligned}\frac{\partial^4\chi}{\partial x_s\partial x_p\partial x_q\partial x_r}\bigg|_{\mathbf{x}=0} &= \frac{1}{4}g_{sp}g_{qr} + \frac{1}{4}g_{sq}g_{pr} + \frac{1}{4}g_{sr}g_{pq} \\ &+ \frac{1}{2}g_{sp}d_q\,d_r + \frac{1}{2}g_{sq}d_p\,d_r + \frac{1}{2}g_{sr}d_p\,d_q \\ &+ \frac{1}{2}g_{pq}d_r\,d_s + \frac{1}{2}g_{pr}d_q\,d_s + \frac{1}{2}g_{qr}d_p\,d_s \\ &+ d_sd_pd_qd_r. \end{aligned} \tag{138}$$

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

$$\frac{\partial^4\chi}{\partial x_q^4}\bigg|_{\mathbf{x}=0} = \frac{3}{4}g_{qq}^2 + 3\,g_{qq}d_q^2 + d_q^4. \tag{139}$$

Seemingly, we have

$$\begin{aligned}\frac{\partial^4\chi}{\partial x_q^2\partial x_p^2}\bigg|_{\mathbf{x}=0} &= \frac{1}{4}g_{qq}g_{pp} + \frac{1}{2}g_{pq}^2 \\ &+ \frac{1}{2}g_{pp}d_q^2 + \frac{1}{2}g_{qq}d_p^2 + 2g_{pq}d_p\,d_q + d_p^2d_q^2. \end{aligned} \tag{140}$$

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

$$\begin{aligned}\frac{\partial^2\chi_R}{\partial x_m\partial x_n}\bigg|_{\mathbf{x}=0} &= -\frac{1}{2}g_{mn} - d_md_n, \\ \frac{\partial^2\chi_I}{\partial x_m\partial x_n}\bigg|_{\mathbf{x}=0} &= 0. \end{aligned} \tag{141}$$

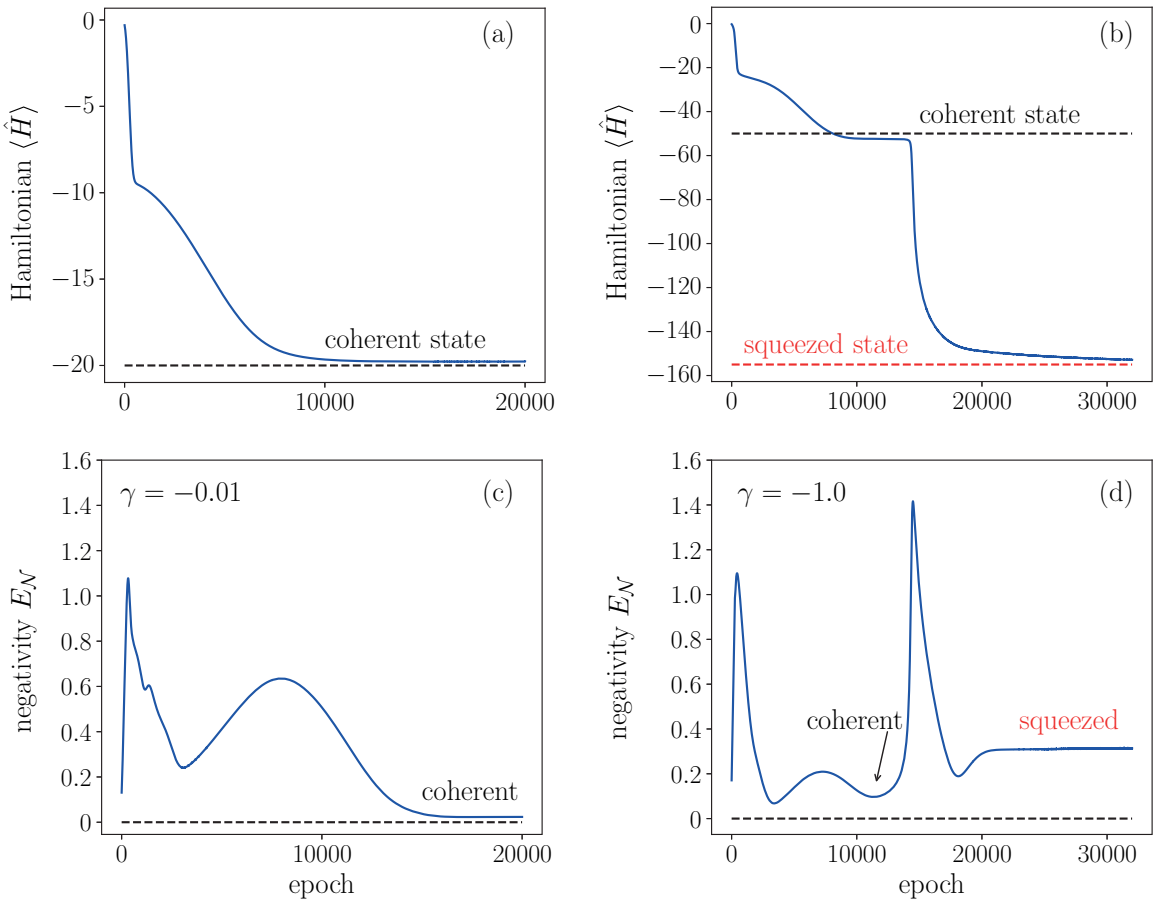

FIG. 19. Training history for Fig. 18: (a) $\langle H \rangle$ rapidly reaches the minimum for $\gamma = -0.01$ (dashed line is the coherent state with $\langle \hat{H} \rangle \simeq -2N_T$); (b) as in (a) for $\gamma = -1$. Dashed line is Eq. (143). The model explores plateaus corresponding to solitons with different degrees of entanglement; (c) logarithmic negativity for $\gamma = -0.01$, no entanglement (dashed line); (d) as in (c) for $\gamma = -1$, the ground state is entangled.

XII. QUANTUM SOLITON VARIATIONAL ANSATZ

Figure 17 shows the quantum processor to synthesize quantum lattice solitons in a discrete array with n sites. The circuit is composed by n squeezers with complex squeezing parameters ζ_j and displacement operators with complex displacements δ_j, being $j = 0, 1, \ldots, n-1$ hereafter. A unitary interferometer $\hat{U}$ mixes the modes before entering the waveguide array. The quantum processor is represented as a NN in the phase space. All the parameters of the squeezers, displacements, and interferometers are trainable and initialized as random variables. We refer to the state generated by the quantum circuit as the quantum soliton variational ansatz (QSVA). We build the NN model of an arbitrary state by multiple pullbacks from the vacuum. Figure 17(b) shows the NN model for the QSVA in Fig. 17(a).

We obtain the mean values of $\hat{H}$ and $\hat{\mathcal{N}}$ as derivatives of χ. As described above, for a Gaussian state, derivatives are obtained algebraically by $\boldsymbol{g}$ and $\boldsymbol{d}$, which speeds up training by conventional steepest descent. We use as cost function $\exp(\langle \hat{H} \rangle / n)$, ($\langle \hat{H} \rangle < 0$ at a minimum). As additional cost function, we use $(\langle \hat{\mathcal{N}} \rangle - N_T)^2$ to constraint the target mean boson number as $\langle \hat{\mathcal{N}} \rangle = N_T$.

The first challenge is training the variational quantum circuit to generate single quantum solitons. The simplest soliton is localized in one of the sites (as $\gamma < 0$). We want to understand if the trained QSVA can approximate a quantum version of the single soliton, such that one can use quantum processors to generate quantum solitons and investigate their physics.

We find that the NN furnishes quantum solitons with lower energy than the classical solution so the quantum solitons form a larger class of nonlinear waves. Starting from randomly generated weights, the soliton is found after training the NN model which minimizes the cost function, and $\langle \hat{H} \rangle$ at fixed $\langle \hat{\mathcal{N}} \rangle = N_T$.

We first consider small interaction strength $|\gamma|$, at which we expected delocalized solutions. We show in Fig. 18(a) the displacements $|\langle \psi_j \rangle|^2 = (d_{2j}^2 + d_{2j+1}^2)/2$ after the training. In Fig. 18(b), we show the mean boson number $\langle \hat{n}_j \rangle$, for $N_T = 10$ and $\gamma = -0.01$ after thousands of epochs; the profile is delocalized. Strong localization is obtained for $N_T = 10$ and $\gamma = -1$ as in Figs. 18(b) and 18(c). In Fig. 19, we report $\langle \hat{H} \rangle$ when varying the number of training epochs for $N_T = 10$. The algorithm converges after some thousands of epochs, and explores different solutions. Figure 19(a) shows the training

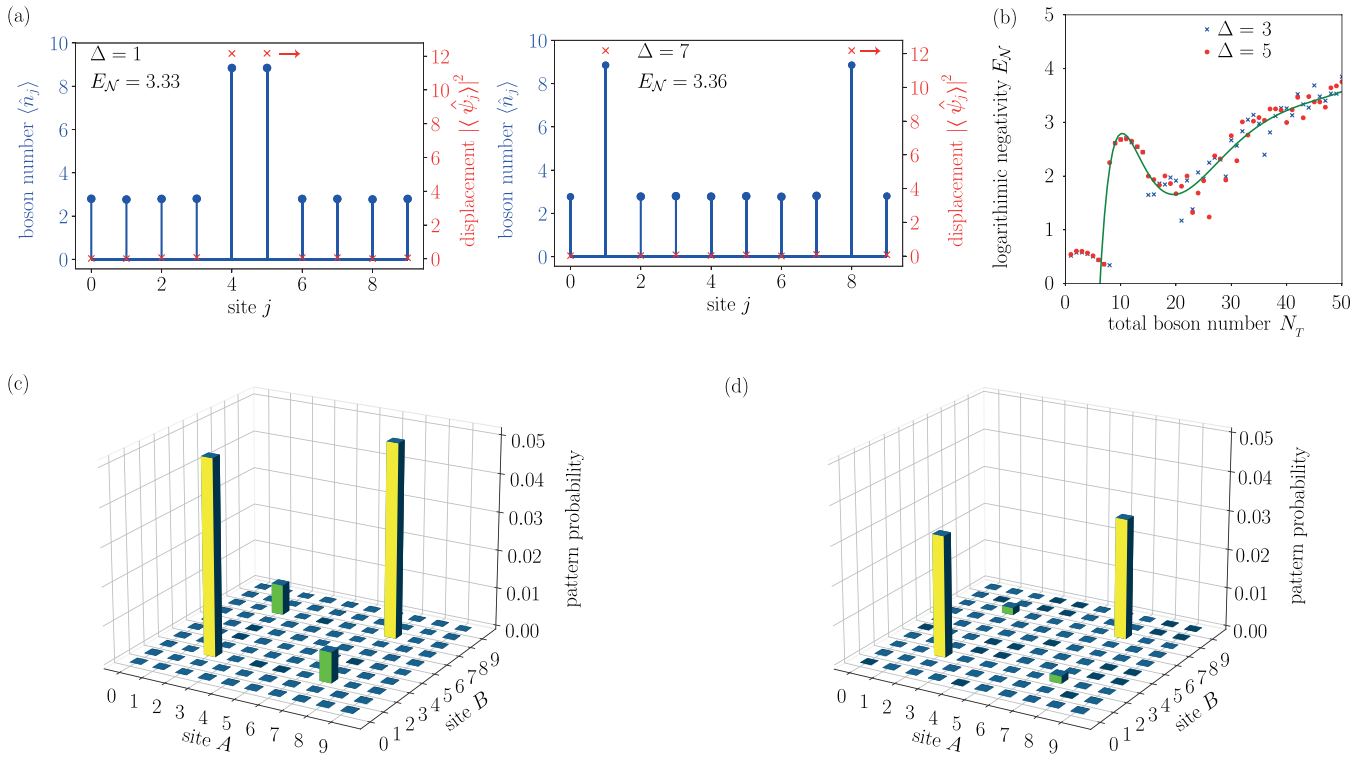

FIG. 20. (a) Boson distribution and mean displacement (right axis) for two representative bound-soliton solutions with $N_T = 40$ and $\gamma = -1.0$; (b) logarithmic negativity versus the boson number N_T for different Δ; (c) probabilities for events containing $\bar{n}_T = 2$ particles in two varying sites A and B and $N_T = 10$; (d) as in (c) with two particles per site ($\bar{n}_T = 4$). The probability is not vanishing only when the particles are at the soliton sites, denoting the generation of entangled pairs ($n = 10$).

when $\gamma = -0.01$: the NN converges to the delocalized state in Figs. 18(a) and 18(b). To compare with analytical estimates, we consider the continuum limit and a nonsqueezed coherent state with a sinusoidal profile. The value of the Hamiltonian $\langle \hat{H} \rangle \simeq -2N_T$ [dashed line in Fig. 19(a)]. Figure 19(b) shows the strong interaction case $\gamma = -1$. During the training, the system settles in two plateaus corresponding to $\langle \hat{H} \rangle \cong -50$ and $\langle \hat{H} \rangle \cong -155$. These values are understood in terms of approximate solutions, in which $\langle \hat{\psi}_{j \neq A} \rangle = 0$, with $A = \lfloor \frac{n}{2} \rfloor + 1$, where the soliton is localized. If we consider the many-body state $|\boldsymbol{\alpha}\rangle = \hat{D}_A(\boldsymbol{\alpha})|\mathbf{vac}\rangle$ with $\alpha_i = \delta_{iA}\alpha$, we have

$$\langle \hat{H} \rangle \simeq \frac{\gamma}{2}\left(\sum_i \left\langle |\hat{n}_i^2| \right\rangle - \langle |\hat{n}_i| \rangle \right) = \frac{\gamma}{2} \sum_i |\alpha_i|^4 = \frac{\gamma}{2}|\alpha|^4. \tag{142}$$

One has $\langle \hat{N} \rangle = |\alpha|^2 = N_T$ and $\langle \hat{H} \rangle = \frac{\gamma}{2} N_T^2$ which gives $\langle \hat{H} \rangle \simeq -50$ (first plateau in Fig. 19, dashed line). The difference due to the neglected $\langle \hat{K} \rangle$.

The lowest energy solution occurs at a larger number of epochs [Fig. 19(b)]. The plateau can be estimated by considering a squeezed coherent state $|\boldsymbol{\alpha}, \zeta\rangle = \hat{D}_A(\boldsymbol{\alpha})\hat{S}_A(\zeta)|\mathbf{vac}\rangle$ where $\hat{S}_A$ is the squeezing operator at $j = A$ with $\zeta = re^{i\theta}$. For $|\boldsymbol{\alpha}, \zeta\rangle$, one has

$$\begin{aligned} \langle \hat{\mathcal{N}} \rangle &= \sinh(r)^2 + |\alpha|^2 = N_T, \\ \langle \hat{H} \rangle &= \tfrac{5}{8} - 2|\alpha^2| + |\alpha|^4 + (-1 + 2|\alpha|^2)\cosh(2r) \\ &\quad + \tfrac{3}{8}\cosh(4r) - |\alpha|^2 \cos(2\phi - \theta)\sinh(2r), \end{aligned} \tag{143}$$

with minimum $\langle \hat{H} \rangle \simeq -155$ ($N_T = 10$, red dashed line).

XIII. ENTANGLED SOLITONS

Having evidence that the ground state is not the classical solution, we are interested to understand if other nonclassical features are present. We consider precisely the degree of entanglement.

Entanglement in Gaussian states is estimated by the logarithmic negativity [13,49–51]. We partition the system in the soliton site, Alice A at $A = \lfloor \frac{n}{2} \rfloor + 1$, and the sites with $j \neq A$, Bob.

We retrieve the covariance matrix $\boldsymbol{g}$ from the trained NN model, and its partial transpose $\tilde{\boldsymbol{g}}$ [51], obtained by multiplying by -1 the elements of Alice momenta $\hat{p}_A$. The symplectic eigenvalues $\tilde{c}_0, \tilde{c}_1, \ldots, \tilde{c}_{n-1}$ are moduli of the eigenvalues of $\mathbf{J}^\top \tilde{\boldsymbol{g}}/2$, and logarithmic negativity is

$$E_{\mathcal{N}} = -\sum_{j=0}^{n-1} \log_2 \min\{1, 2c_j\}. \tag{144}$$

Figure 19(c) shows the $E_{\mathcal{N}}$ for $\gamma = -0.01$ during the training. Despite the system initially exploring randomly generated entangled solution, the asymptotic state is a coherent state, with vanishing entanglement (dashed line) as expected for a negligible interaction.

Figure 19(d) shows $E_{\mathcal{N}}$ for $\gamma = -1$. The local minimum of $\langle \hat{H} \rangle$, the coherent state, has a smaller entanglement than the squeezed ground state at epoch 3×10^4. Nevertheless, the asymptotic value of $E_{\mathcal{N}}$ is negligible if compared with the bounded solitons considered in the following.

Given the nonclassical features in the single-soliton ground state, we study the broader family of multiple solitons. We expect that strong nonlinear localization in different channels increases the degree of entanglement. We study two-soliton bound states [1,30,31], which can be approximated by training the network by proper loss functions. We fix the location of two sites, denoted A and $B = A + \Delta$, and we introduce as cost function $\exp(-\langle \hat{n}_A \rangle)$ to maximize the bosons at A and $\exp(\langle \hat{n}_A - \hat{n}_B \rangle)$ to generate a solution with the same boson number in B.

Figure 20 shows the coupled solitons for different peak positions ($N_T = 40$). At variance with the single soliton in Fig. 18, the boson number is not vanishing in all the sites, but $|\langle \psi_j \rangle|^2$ is not negligible only at the solitons, outlining the onset of squeezed nonlocal states.

The bound solitons are entangled. If we consider a bipartite system composed by the site in A and the rest of the array, $E_{\mathcal{N}}$ depends on the distance Δ between the two solitons (Fig. 20). The logarithmic negativity depends on the number of total bosons N_T: Figure 20(b) shows $E_{\mathcal{N}}$ versus N_T for various Δ. The entanglement is small and comparable with the single soliton for case $N_T < 10$, and growing when $N_T > 10$.

XIV. BOSON SAMPLING

As we show in the following, the entanglement is connected to correlated particles in different output channels. We compute the probability $\Pr(\bar{\mathbf{n}})$ of finding $\bar{n}_0$ bosons in mode 0, $\bar{n}_1$ in mode 1, and so forth, where $\bar{\mathbf{n}} = (\bar{n}_0, \bar{n}_1, \ldots, \bar{n}_{n-1})$ is a given particle pattern. Letting ρ be the density matrix, one has $\Pr(\bar{\mathbf{n}}) = \mathrm{Tr}[\rho |\bar{\mathbf{n}}\rangle\langle\bar{\mathbf{n}}|]$, with $|\bar{\mathbf{n}}\rangle\langle\bar{\mathbf{n}}| = \otimes_j |\bar{n}_j\rangle\langle\bar{n}_j|$. Correspondingly [52],

$$\Pr(\bar{\mathbf{n}}) = \frac{1}{\bar{\mathbf{n}}!} \prod_j \left(\frac{\partial^2}{\partial \alpha_j \partial \alpha_j^*} \right)^{\bar{n}_j} e^{\sum_j |\alpha|_j^2} Q_\rho(\boldsymbol{\alpha}) \Big|_{\boldsymbol{\alpha}=0}, \qquad (145)$$

where $\bar{\mathbf{n}}! = \bar{n}_0! \bar{n}_1! \ldots \bar{n}_{n-1}!$ and $Q_\rho = \pi^n \langle \boldsymbol{\alpha} | \rho | \boldsymbol{\alpha} \rangle$ is the Q representation of the density matrix [47,53].

We obtain $\Pr(\bar{\mathbf{n}})$ as in Ref. [54]. We consider an event with $\bar{n}_T$ particles in two sites A and B. Specifically, the events such that (i) $\bar{n}_A = \bar{n}_B = n_T/2$ for $A \neq B$ or (ii) $\bar{n}_A = \bar{n}_B = \bar{n}_T$ when $A = B$, and $\bar{n}_j = 0$ elsewhere. For $n_T = 2$, this corresponds to observing a pair of particles in a single site or two particles in two distinct sites.

Figure 20(c) shows the probability for $\bar{n}_T = 2$ varying A and B when $N_T = 10$. Such a probability is different from zero only when A or B are at a soliton position [$A = 2$ and $B = 7$ in Fig. 20(c)]. When $A = B$ (yellow bar), we have the probability of observing the two particles in one single soliton. When A and B correspond to the sites of the two solitons (green bar), we have the probability of observing a couple of particles in the two soliton sites. For all other patterns with $\bar{n}_T = 2$, there is a negligible probability of observing a pair. Hence, pairs of particles appear either in a single soliton or entangled in the two solitons. When $\bar{n}_T = 4$, we have a nonvanishing probability of observing either four bosons in a soliton or two entangled pairs in the two solitons [Fig. 20(d)]. Particles are hence observed simultaneously only when they are monitored at the two localized solitons, demonstrating quantum correlation and the nonlocality.

XV. CONCLUSION

Discrete waveguide arrays became popular for the quantum advantage in linear BS. One can argue about the role of nonlinearity, which—classically—supports discrete self-localized solitons.

For unveiling quantum effects in discrete solitons, we adopted QML. Variational quantum circuits enable the theoretical investigation of quantum nonlinear waves and their bound states. Also, their physical realization as tunable quantum processors opens the way to experiments on nonperturbative nonlinear regimes.

We considered Gaussian states, as they are expected to conform to experimental observations at high intensities, and also allowed us to rigorously obtain the degree of entanglement. Furthermore, the NN formulation is compatible with the Gaussian BS protocol in the presence of nonlinearity.

We found that solitonic states alter the observation of multiparticle events. Entangled pairs emerge from self-localized bounded solitons, which can be synthesized by trained quantum circuits.

QML and quantum variational algorithms open possibilities for the physics and the applications of quantum nonlinear waves. The methodology can be extended beyond Gaussian states and generalized to continuous systems and multidimensional arrays with arbitrary networks.

ACKNOWLEDGMENT

We acknowledge support from Horizon 2020 QuantERA grant QUOMPLEX, by National Research Council (CNR), Grant No. 731473, and PRIN PELM (No. 20177PSCKT).

APPENDIX: GRAPH AND PARAMETERS OF THE MODEL

Figure 21 shows a graphical representation of the graph of the model. We optimize the model to minimize the Hamiltonian; the resulting parameters give the circuit that produces a state with the minimum value of energy. The trainable parameters in the model, i.e., those corresponding to the n displacement layers (each layer with one 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 the model has $N^2/4 + N$ independent variables. Training is done by the Adam algorithm (`TensorFlow v2.7.0`). Gaussian BS from the trained model is done following Ref. [54].

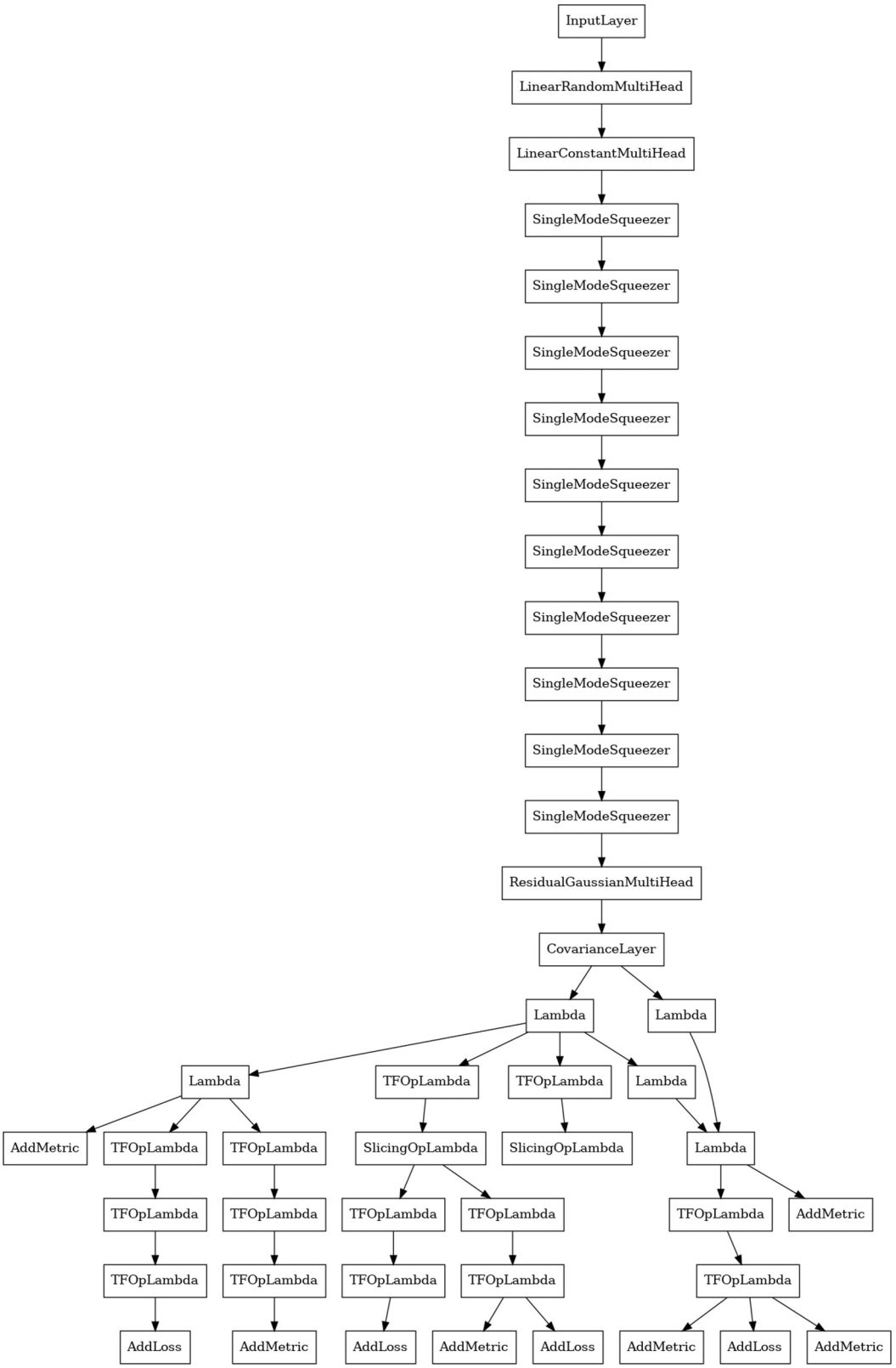

FIG. 21. Graph of the multihead model for the quantum solitons. One can notice the central computational core represented by the cascade of multihead layers, including the interferometers `LinearRandomMultiHead`, displacements `LinearConstantMultiHead`, and squeezing layers `SingleModeSqueezer`. Then 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.

[1] Y. Kivshar and G. P. Agrawal, *Optical Solitons* (Academic Press, New York, 2003)

[2] C. Conti, Fluctuations and non-Gaussianity in 3D+1 quantum nonlocal nonlinear waves: towards quantum gravity and quantum fluid technologies, arXiv:2202.10741.

[3] S. R. Friberg, S. Machida, and Y. Yamamoto, Quantum-Nondemolition Measurement of the Photon Number of an Optical Soliton, Phys. Rev. Lett. **69**, 3165 (1992).

[4] P. D. Drummond, R. M. Shelby, S. R. Friberg, and Y. Yamamoto, Quantum solitons in optical fibres, Nature (London) **365**, 307 (1993).

[5] C. Silberhorn, P. K. Lam, O. Weis, F. Konig, N. Korolova, and G. Leuchs, Generation of Continuous Variable Einstein-Podolsky-Rosen Entanglement Via the Kerr Nonlinearity in an Optical Fiber, Phys. Rev. Lett. **86**, 4267 (2001).

[6] C. X. Yu, H. A. Haus, and E. P. Ippen, Soliton squeezing at the gigahertz rate in a Sagnac loop, Opt. Lett. **26**, 669 (2001).

[7] Q.-Y. Liang, A. V. Venkatramani, S. H. Cantu, T. L. Nicholson, M. J. Gullans, A. V. Gorshkov, J. D. Thompson, C. Chin, M. D. Lukin, and V. Vuletić, Observation of three-photon bound states in a quantum nonlinear medium, Science **359**, 783 (2018).

[8] G. Leuchs, T. C. Ralph, C. Silberhorn, and N. Korolkova, Scheme for the generation of entangled solitons for quantum communication, J. Mod. Opt. **46**, 1927 (1999).

[9] R.-K. Lee, Y. Lai, and B. A. Malomed, Generation of photon-number-entangled soliton pairs through interactions, Phys. Rev. A **71**, 013816 (2005).

[10] Y. Lai and R.-K. Lee, Entangled Quantum Nonlinear Schrödinger Solitons, Phys. Rev. Lett. **103**, 013902 (2009).

[11] Leone Di Mauro Villari, D. Faccio, F. Biancalana, and C. Conti, Quantum soliton evaporation, Phys. Rev. A **98**, 043859 (2018).

[12] O. V. Marchukov, B. A. Malomed, V. Dunjko, J. Ruhl, M. Olshanii, R. G. Hulet, and V. A. Yurovsky, Quantum Fluctuations of the Center of Mass and Relative Parameters of Nonlinear Schrödinger Breathers, Phys. Rev. Lett. **125**, 050405 (2020).

[13] V. O. Martynov, V. O. Munyaev, and L. A. Smirnov, Generation of entangled states of light using discrete solitons in waveguide arrays, arXiv:2011.07662.

[14] M. J. Potasek and B. Yurke, Squeezed-light generation in a medium governed by the nonlinear Schrödinger equation, Phys. Rev. A **35**, 3974 (1987).

[15] Y. Lai and H. A. Haus, Quantum theory of solitons in optical fibers. I. Time-dependent Hartree approximation, Phys. Rev. A **40**, 844 (1989).

[16] E. M. Wright, Quantum theory of soliton propagation in an optical fiber using the Hartree approximation, Phys. Rev. A **43**, 3836 (1991).

[17] A. Scott, J. Eilbeck, and H. Gilhøj, Quantum lattice solitons, Physica D **78**, 194 (1994).

[18] D. Yao, Quantum fluctuations of optical solitons in fibers, Phys. Rev. A **52**, 4871 (1995).

[19] G. Marcucci, D. Pierangeli, and C. Conti, Theory of Neuromorphic Computing by Waves: Machine Learning by Rogue Waves, Dispersive Shocks, and Solitons, Phys. Rev. Lett. **125**, 093901 (2020).

[20] R. A. Herrera, Evaluating a neural network and a convolutional neural network for predicting soliton properties in a quantum noise environment, J. Opt. Soc. Am. B **37**, 3094 (2020).

[21] N. A. Silva, T. D. Ferreira, and A. Guerreiro, Reservoir computing with solitons, New J. Phys. **23**, 023013 (2021).

[22] J.-C. Pu, J. Li, and Y. Chen, Soliton, breather, and rogue wave solutions for solving the nonlinear Schrödinger equation using a deep learning method with physical constraints, Chin. Phys. B **30**, 060202 (2021).

[23] V. Lopez-Pastor and F. Marquardt, Self-learning machines based on Hamiltonian echo backpropagation, arXiv:2103.04992.

[24] M. Tillmann, B. Dakić, R. Heilmann, S. Nolte, A. Szameit, and P. Walther, Experimental boson sampling, Nat. Photonics **7**, 540 (2013).

[25] M. A. Broome, A. Fedrizzi, S. Rahimi-Keshari, J. Dove, S. Aaronson, T. C. Ralph, and A. G. White, Photonic boson sampling in a tunable circuit, Science **339**, 794 (2013).

[26] J. B. Spring, B. J. Metcalf, P. C. Humphreys, W. S. Kolthammer, X.-M. Jin, M. Barbieri, A. Datta, N. Thomas-Peter, N. K. Langford, D. Kundys, J. C. Gates, B. J. Smith, P. G. R. Smith, and I. A. Walmsley, Boson sampling on a photonic chip, Science **339**, 798 (2013).

[27] J. Carolan, J. D. A. Meinecke, P. J. Shadbolt, N. J. Russell, N. Ismail, K. Wörhoff, T. Rudolph, M. G. Thompson, J. L. O'Brien, J. C. F. Matthews, and A. Laing, On the experimental verification of quantum complexity in linear optics, Nat. Photonics **8**, 621 (2014).

[28] N. Spagnolo, C. Vitelli, M. Bentivegna, D. J. Brod, A. Crespi, F. Flamini, S. Giacomini, G. Milani, R. Ramponi, P. Mataloni, R. Osellame, E. F. Galvão, and F. Sciarrino, Experimental validation of photonic boson sampling, Nat. Photonics **8**, 615 (2014).

[29] H.-S. Zhong, H. Wang, Y.-H. Deng, M.-C. Chen, L.-C. Peng, Y.-H. Luo, J. Qin, D. Wu, X. Ding, Y. Hu, P. Hu, X.-Y. Yang, W.-J. Zhang, H. Li, Y. Li, X. Jiang, L. Gan, G. Yang, L. You, Z. Wang *et al.*, Quantum computational advantage using photons, Science **370**, 1460 (2020).

[30] F. Lederer, G. I. Stegeman, D. N. Christodoulides, G. Assanto, M. Segev, and Y. Silberberg, Discrete solitons in optics, Phys. Rep. **463**, 1 (2008).

[31] Y. V. Kartashov, B. A. Malomed, and L. Torner, Solitons in nonlinear lattices, Rev. Mod. Phys. **83**, 247 (2011).

[32] S. Aaronson and A. Arkhipov, The computational complexity of linear optics, Theory Comput. **9**, 143 (2013).

[33] G. Carleo, I. Cirac, K. Cranmer, L. Daudet, M. Schuld, N. Tishby, L. Vogt-Maranto, and L. Zdeborová, Machine learning and the physical sciences, Rev. Mod. Phys. **91**, 045002 (2019).

[34] M. Broughton, G. Verdon, T. McCourt, A. J. Martinez, J. H. Yoo, S. V. Isakov, P. Massey, M. Y. Niu, R. Halavati, E. Peters, M. Leib, A. Skolik, M. Streif, D. V. Dollen, J. R. McClean, S. Boixo, D. Bacon, A. K. Ho, H. Neven, and M. Mohseni, Tensorflow quantum: A software framework for quantum machine learning, arXiv:2003.02989.

[35] G. Marcucci, D. Pierangeli, P. W. H. Pinkse, M. Malik, and C. Conti, Programming multi-level quantum gates in disordered computing reservoirs via machine learning, Opt. Express **28**, 14018 (2020).

[36] F. Marquardt, Machine learning and quantum devices, SciPost Phys. Lect. Notes **29** (2021).

[37] F. Tacchino, S. Mangini, P. K. Barkoutsos, C. Macchiavello, D. Gerace, I. Tavernelli, and D. Bajoni, Variational learning for

quantum artificial neural networks, IEEE Trans. Quantum Eng. **2**, 3101110 (2021).

[38] A. Kardashin, A. Pervishko, J. Biamonte, and D. Yudin, Numerical hardware-efficient variational quantum simulation of a soliton solution, Phys. Rev. A **104**, L020402 (2021).

[39] H.-Y. Huang, R. Kueng, G. Torlai, V. V. Albert, and J. Preskill, Provably efficient machine learning for quantum many-body problems, arXiv:2106.12627.

[40] M. Schuld and F. Petruccione, *Supervised Learning with Quantum Computers* (Springer, Cham, Switzerland, 2018).

[41] P. Rebentrost, M. Mohseni, and S. Lloyd, Quantum Support Vector Machine for Big Data Classification, Phys. Rev. Lett. **113**, 130503 (2014).

[42] M. L. Wall and G. D'Aguanno, Tree-tensor-network classifiers for machine learning: From quantum inspired to quantum assisted, Phys. Rev. A **104**, 042408 (2021).

[43] F. Vicentini, A. Biella, N. Regnault, and C. Ciuti, Variational Neural-Network Ansatz for Steady States in Open Quantum Systems, Phys. Rev. Lett. **122**, 250503 (2019).

[44] M. Cerezo, A. Arrasmith, R. Babbush, S. C. Benjamin, S. Endo, K. Fujii, J. R. McClean, K. Mitarai, X. Yuan, L. Cincio, and P. J. Coles, Variational quantum algorithms, Nat. Rev. Phys. **3**, 625 (2021).

[45] J. M. Arrazola, V. Bergholm, K. BrÃ¡dler, T. R. Bromley, M. J. Collins, I. Dhand, A. Fumagalli, T. Gerrits, A. Goussev, L. G. Helt, J. Hundal, T. Isacsson, R. B. Israel, J. Izaac, S. Jahangiri, R. Janik, N. Killoran, S. P. Kumar, J. Lavoie, A. E. Lita *et al.*, Quantum circuits with many photons on a programmable nanophotonic chip, Nature (London) **591**, 54 (2021).

[46] F. Hoch, S. Piacentini, T. Giordani, Z.-N. Tian, M. Iuliano, C. Esposito, A. Camillini, G. Carvacho, F. Ceccarelli, N. Spagnolo, A. Crespi, F. Sciarrino, and R. Osellame, Boson sampling in a reconfigurable continuously-coupled 3D photonic circuit, arXiv:2106.08260.

[47] S. M. Barnett and P. M. Radmore, *Methods in Theoretical Quantum Optics* (Oxford University Press, New York, 1997).

[48] X. Wang, T. Hiroshima, A. Tomita, and M. Hayashi, Quantum information with Gaussian states, Phys. Rep. **448**, 1 (2007).

[49] G. Adesso, A. Serafini, and F. Illuminati, Extremal entanglement and mixedness in continuous variable systems, Phys. Rev. A **70**, 022318 (2004).

[50] M. B. Plenio and S. Virmani, An introduction to entanglement measures, Quantum Inf. Comput. **7**, 1 (2007).

[51] G. Vidal and R. F. Werner, Computable measure of entanglement, Phys. Rev. A **65**, 032314 (2002).

[52] R. Kruse, C. S. Hamilton, L. Sansoni, S. Barkhofen, C. Silberhorn, and I. Jex, A detailed study of Gaussian boson sampling, Phys. Rev. A **100**, 032326 (2019).

[53] C. W. Gardiner and P. Zoller, *Quantum Noise*, 3rd ed. (Springer-Verlag, Berlin, 2004).

[54] C. Conti, Training Gaussian boson sampling by quantum machine learning, Quantum Mach. Intell. **3**, 26 (2021).