

Variational Tensor Network Simulation of Gaussian Boson Sampling and Beyond

Jonas Vinther

*Department of Computer Science, University of Copenhagen, DK and
Niels Bohr Institute, University of Copenhagen, DK**

Michael J. Kastoryano

*Department of Computer Science, University of Copenhagen, DK
(Dated: March 12, 2025)*

The continuous variable quantum computing platform constitutes a promising candidate for realizing quantum advantage, as exemplified in Gaussian Boson Sampling. While noise in the experiments makes the computation attainable for classical simulations, it has been suggested that the addition of non-linear elements to the experiment will help retain the quantum advantage. We propose a classical simulation tool for general continuous variable sampling problems, including Gaussian Boson Sampling and beyond. We reformulate the sampling problem as that of finding the ground state of a simple few-body Hamiltonian. This allows us to employ powerful variational methods based on tensor networks and to read off the simulation error directly from the expectation value of the Hamiltonian. We validate our method by simulating Gaussian Boson Sampling, where we achieve results comparable to the state of the art. We also consider a non-Gaussian sampling problem, for which we develop novel local basis optimization techniques based on a non-linear parameterization of the implicit basis, resulting in high effective cutoffs with diminished computational overhead.

I. INTRODUCTION

Continuous variable (CV) quantum systems constitute a conceptually different framework for quantum computation as compared to discrete, qudit based architectures [1][2][3]. An essential milestone that has been pursued over the last decade, is the proof of a quantum advantage [4]. For CV systems, one of the candidates for such an experiment is the Gaussian Boson Sampling (GBS) problem [5][6]. Standard GBS is based on squeezed vacuum states of light that passes through rotation gates and beam-splitters and are subject to either threshold or Photon Number Resolved (PNR) detection at the end of the circuit [7][8]. Calculating a probability for obtaining one of the many combinations of photon number measurements is an exponentially hard computational problem in the number of excitations [9][10][11].

Previous attempts at handling this curse of dimensionality in simulating boson sampling, includes the approximate algorithm of Ref. [12] which is based on Metropolis independence sampling. The result was subsequently improved in Ref. [13] where a faster, exact algorithm was developed. The same authors further improved the runtime in Ref. [14]. In general, the computational complexity for bosonic computations can be posed as an exponential scaling in terms of a measure of non-Gaussianity [15][16]. For GBS, the measurements constitute the non-Gaussian element. Phase-space based methods are useful for validating experiments [17].

The state of the art in classical simulations of experimental GBS is based on tensor network methods and relies on the fact that the reduced density matrix of a

Gaussian state can be diagonalized exactly [18]. We explore a more straightforward approach, which can be further generalized to a larger set of multi-modal states, beyond Gaussian States, with application in e.g. non-linear boson sampling [19]. Extending the input of a GBS protocol to non-Gaussian states does not increase the scaling of the complexity [20]. It does, however, increase the resilience to noise in achieving a quantum advantage [21]. Our method can be used as an approximate classical simulation of such proposals, with which to validate the quantum computation

In this work, we construct a Hamiltonian for which the resulting quantum state of the experiment is the unique ground state with energy zero. By writing the Hamiltonian in the Fock basis and solving the resulting variational problem with e.g. DMRG [22], we obtain an approximate solution in polynomial time. By construction, the variational energy is a direct measure of the simulation error, hence accuracy can be assessed on the fly.

One of the drawbacks of our approach, is that the discarding of high photon number Fock basis elements induces an error on the low photon number probabilities. Therefore, we make use of Local Basis Optimization [23] (LBO), in order to achieve a higher cutoff at reduced computational complexity. For the case of GBS, we derive the exact optimal local basis, which additionally yields access to bounds on the truncation error.

For the general case it is not possible to derive the optimal local basis and we examine a complimentary approach to standard LBO techniques, that we denote parameterized LBO. Simply stated, we allow our physical basis to have variational parameters which are optimized simultaneously with the variational state. Crucially, the local basis transformation is implemented analytically in the Hamiltonian, such that all computations are done in the optimized basis. Similar ideas can be found in

* Contact author: jonas.vinther@nbi.ku.dk

ref. [24] [25]. We show that a better approximation of the ground state solution can be found efficiently using orders of magnitude fewer basis elements (i.e. a substantially smaller cutoff).

The article is structured as follows. In section II the main methodology of the paper is outlined along with an application to Gaussian Boson Sampling. In section III, local basis optimization is introduced and exactly solved in the case of GBS. Parameterized LBO is subsequently introduced and an optimization strategy in terms of tensor network methods is formulated. These optimization techniques are applied to a non-Gaussian sampling problem. Results are presented in sec. IV.

II. THEORY

In this section, we present the main method of the paper and subsequently relate it to the relevant physical setting. For an illustrative overview, see figure 1. For further details on the Boson sampling problem see [26] [27]. We set $\hbar = 1$ throughout.

A. Main Methodology

Let us start by considering a noiseless quantum photonics experiment. A typical Boson sampling problem consists of evolving a known input state through a set of CV-gates such as beamsplitters and squeezing and measuring the photon number for each output mode. If we denote the pure state right before the PNR measurements by $|\psi\rangle$ then the probability of a specific outcome of the PNR measurements will follow the distribution:

$$P(\mathbf{n}) := P(n_1, n_2, \dots, n_N) = \left| \left(\bigotimes_{k=1}^N \langle n_k | \right) |\psi\rangle \right|^2 =: |\langle \mathbf{n} | \psi \rangle|^2 \quad (1)$$

where the vector \mathbf{n} refers to the number of photons in each mode and $|n_k\rangle$ refer to the n -th simple quantum harmonic oscillator state in the k -th mode.

When both the input state and the CV-gates are Gaussian, $|\psi\rangle$ will be a Gaussian state and the sampling problem is referred to as Gaussian Boson Sampling (sec. II C 1). The Gaussian state is completely characterized by a $2N \times 2N$ covariance matrix and a $2N$ -dimensional displacement vector [28]. However, even then, calculating an overlap of the form $|\langle \mathbf{n} | \psi \rangle|^2$, scales exponentially in $\sum_k n_k$ [9].

Instead, one can recover all probabilities at once by treating $\langle \mathbf{n} | \psi \rangle$ as the solution to a variational problem for which one can employ tensor network methods, such as DMRG, to achieve an approximate solution in polynomial time. This formulation also allows us to treat non-Gaussian boson sampling problems.

B. Variational Formulation

We consider the noiseless case here. A discussion of the impact of noise is provided in section II F. Suppose we want to sample from a few body input state $|\psi_{\text{in}}\rangle$ that has been subject to the unitary gates of a circuit $\{U_i\}$.

$$|\psi\rangle = \prod_{i \in \text{circuit}} U_i^\dagger |\psi_{\text{in}}\rangle = U^\dagger |\psi_{\text{in}}\rangle \quad (2)$$

We will then make the crucial assumption that the initial state can be prepared from the vacuum, through a set of unitary gates $\prod_{i \in \text{state prep}} V_i = V$ and possibly PNR measurements of auxiliary modes. This includes, at the least, all Gaussian states, any non-Gaussian state that can be prepared from a Gaussian state and Fock states if formulated as in scattershot boson sampling [29].

The method relies on the following simple observation:

$$0 = \sum_n \hat{a}_n^\dagger \hat{a}_n |\mathbf{0}\rangle = \sum_n \hat{A}_n^\dagger \hat{A}_n |\psi\rangle = \mathcal{H} |\psi\rangle \quad (3)$$

where

$$\hat{A}_n = U^\dagger V^\dagger \hat{a}_n V U \quad |\psi\rangle = U^\dagger V^\dagger |\mathbf{0}\rangle \quad (4)$$

The state $|\psi\rangle$ will be the unique ground state of the positive semidefinite hamiltonian \mathcal{H} with eigenvalue zero. The uniqueness comes from the preservation of the spectrum by the similarity transformations. Thus, \mathcal{H} will also have a harmonic spectrum which is again evident from the conservation of the canonical commutation relations:

$$[\hat{A}_i, \hat{A}_j^\dagger] = \delta_{ij} \quad (5)$$

Therefore, by obtaining the ground state of the matrix representation of the Hamiltonian in the Fock-basis $\langle \mathbf{m} | \mathcal{H} | \mathbf{n} \rangle$ one has acquired $\langle \mathbf{n} | \psi \rangle$ for all PNR detection possibilities \mathbf{n} below some cutoff of the Fock-space and thus complete access to the sampled photon number distribution $|\langle \mathbf{n} | \psi \rangle|^2$. Furthermore, the variational energy $\langle \mathcal{H} \rangle$ is directly related to the simulation error, as clarified in sec. II E.

It remains to derive the Hamiltonian for experiments of interest and to answer the question of how to simulate real experiments in the presence of noise. This is done in the following sections.

C. Deriving the Hamiltonian

1. Gaussian Boson Sampling

For Gaussian Boson sampling the vacuum is subject to single mode squeezing and a linear interferometer before being measured by photon number resolved detectors. For the following it will be natural to first rotate the \hat{a}

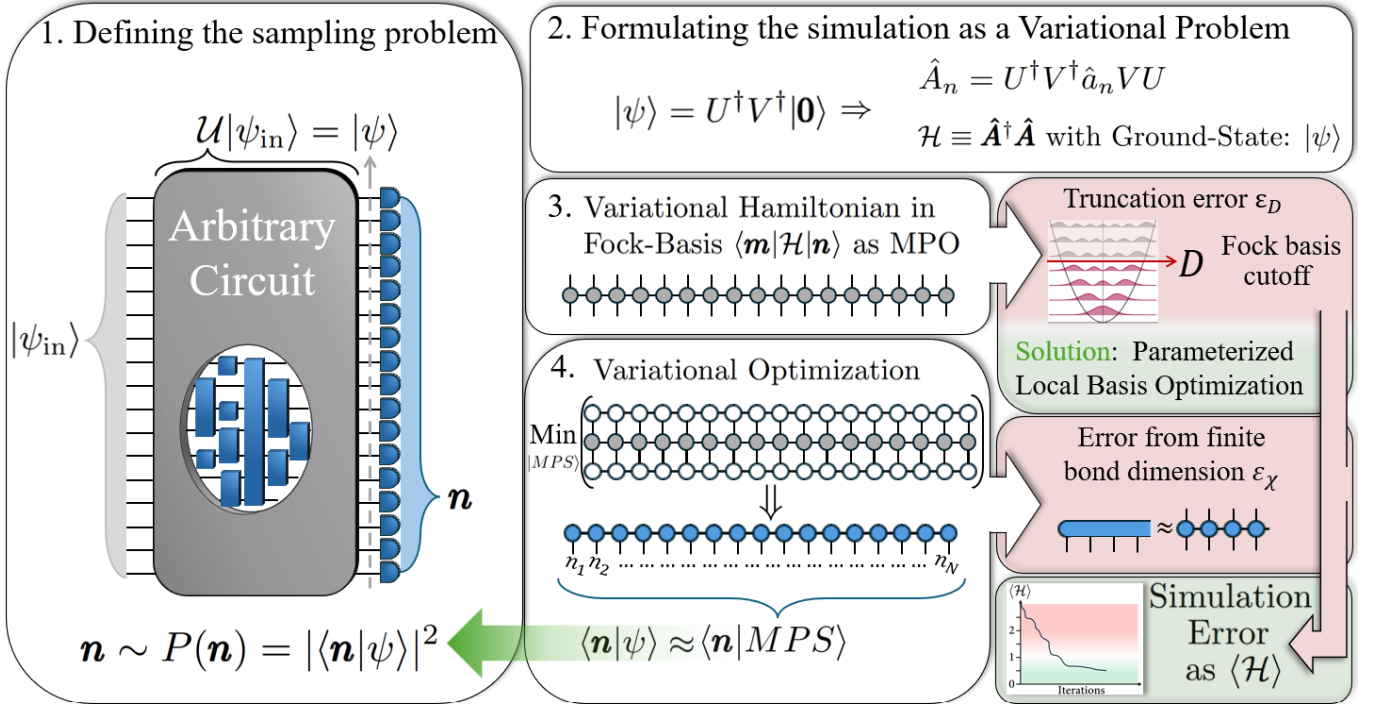


FIG. 1: Overview of our variational approach to solving Bosonic sampling problems. A Hamiltonian \mathcal{H} can be constructed, for which the target state $|\psi\rangle$ is the unique ground state with energy zero. Once the Hamiltonian is written in the Fock Basis, standard variational methods can be used to obtain an approximation to the sampling probabilities $|\langle \mathbf{n} | \psi \rangle|^2$. Furthermore, two approximation errors are identified, the magnitude of which can be assessed simply by observing how close the variational energy tend towards the true ground state energy; 0. In order to mitigate the error due to the Fock basis cutoff, we study the application of parameterized Local Basis Optimization techniques and non-linear optimization methods, the results of which are outlined in sec. IV.

and \hat{a}^\dagger into the \hat{X} and \hat{P} quadrature operators:

$$\begin{pmatrix} \hat{a}_k \\ \hat{a}_k^\dagger \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & i \\ 1 & -i \end{pmatrix} \begin{pmatrix} \hat{X}_k \\ \hat{P}_k \end{pmatrix} \quad (6)$$

$$\mathcal{H}_{|0\rangle} = \sum_k \hat{a}_k^\dagger \hat{a}_k = \frac{1}{2} \begin{pmatrix} \hat{X} \\ \hat{P} \end{pmatrix}^T \begin{pmatrix} \hat{X} \\ \hat{P} \end{pmatrix} - \frac{N}{2} \quad (7)$$

The squeezing operation simply rescales the quadrature operators in an inverse fashion to what is performed on the state:

$$\mathcal{H}_{V^\dagger|0\rangle} = \sum_k V^\dagger \hat{a}_k^\dagger \hat{a}_k V = \frac{1}{2} \begin{pmatrix} \hat{X} \\ \hat{P} \end{pmatrix}^T D^{-2} \begin{pmatrix} \hat{X} \\ \hat{P} \end{pmatrix} - \frac{N}{2} \quad (8)$$

Where D is a diagonal matrix: $D = \text{diag}(e^{-r_1} \dots e^{-r_N} e^{r_1} \dots e^{r_N})$ with $\{r_i\} \in \mathbb{R}$ specifying the squeezing amount of each mode in the Gaussian state. The transformation under the passive Gaussian transformation rotates the quadrature operators by the orthogonal matrix O :

$$\mathcal{H} = \sum_k U^\dagger V^\dagger \hat{a}_k^\dagger \hat{a}_k V U = \frac{1}{2} \begin{pmatrix} \hat{X} \\ \hat{P} \end{pmatrix}^T O^T D^{-2} O \begin{pmatrix} \hat{X} \\ \hat{P} \end{pmatrix} - \frac{N}{2} \quad (9)$$

We can express the Hamiltonian in terms of the covariance matrix of the Gaussian state $V = \frac{1}{2} O D^2 O^T$ [30] namely:

$$\mathcal{H} = \begin{pmatrix} \hat{X} \\ \hat{P} \end{pmatrix}^T \frac{V^{-1}}{4} \begin{pmatrix} \hat{X} \\ \hat{P} \end{pmatrix} - \frac{N}{2} \quad (10)$$

The ground state of such an n -dimensional anisotropic oscillator in a uniform magnetic field [31] is exactly the Gaussian state in question. Displacements are easily accounted for in the simulation (they constitute single mode gates) and are therefore neglected here.

2. Beyond GBS: fully connected CZ-gate.

GBS poses a benchmark problem; a setting for which we can compare our method to other exact methods. In order to highlight the flexibility of our approach and to showcase our novel methods for solving the variational problem, we also consider a (artificial) highly non-Gaussian circuit that would be not so straight forward by other means. That is, a Gaussian circuit on which a completely connected phase gate is applied in between the Gaussian unitary and the PNR measurements. The

reason for the full connectivity is to ensure that we can't just solve for the Gaussian state and apply unitary gates to our representation of the state (this would for example be a viable approach if we only applied, the otherwise non-Gaussian, cubic phase gate to each output mode). Specifically we consider:

$$|\psi\rangle = \exp\left(-i\kappa \prod_k^N \hat{X}_k\right) |\psi_{\text{Gaussian}}\rangle \quad (11)$$

which leaves \hat{X} invariant but transforms \hat{P} as:

$$\hat{P}_i \rightarrow \hat{P}_i + \kappa \prod_{n \neq i} \hat{X}_n \quad (12)$$

such that if we define a "non-local" vector of operators:

$$(\hat{\xi})_i = \kappa \prod_{n \neq i} \hat{X}_n \quad (13)$$

The variational Hamiltonian for this example, is simply:

$$\mathcal{H} = \left(\hat{\mathbf{X}} - \boldsymbol{\mu}\right)^T \frac{V^{-1}}{4} \left(\hat{\mathbf{X}} - \boldsymbol{\mu}\right) - \frac{N}{2} \quad (14)$$

The presence of $\hat{\xi}$ makes the Hamiltonian non-Gaussian when we consider a circuit with $N > 2$.

D. Matrix Product States

In order to numerically treat many body quantum systems, it is necessary to restrict the state representation to only a part of the Hilbert space. Matrix Product States (MPSs) provide a succinct description of the physically relevant states, with direct access to truncation of the entanglement spectrum between the constituent Hilbert spaces of the N-body Hilbert space [32][33][34].

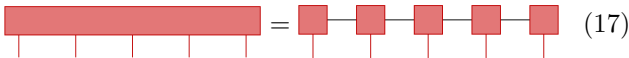
The Matrix Product State ansatz is a statement about the tensor $c_{n_1, n_2, \dots}$ containing the coefficients of the pure state in a given basis:

$$|\psi\rangle = \sum_{\mathbf{n}} c_{\mathbf{n}} |n_1\rangle \otimes |n_2\rangle \otimes \dots \otimes |n_N\rangle \quad (15)$$

Namely, that it can be written as a product of matrices:

$$c_{\mathbf{n}} = (\mathbf{M}_{n_1}^{(1)})^T M_{n_2}^{(2)} \dots M_{n_{N-1}}^{(N-1)} \mathbf{M}_{n_N}^{(N)} \quad (16)$$

Here $\mathbf{M}^{(1)}$ and $\mathbf{M}^{(N)}$ are vectors because we choose open boundary conditions for algorithmic reasons. Diagrammatically, eq. 16 looks like (for $N = 5$):



$$\text{[Diagrammatic representation of eq. 16]} \quad (17)$$

where the legs denote the indices of the corresponding tensor and connected legs denote summation over that index. The matrices $\{M_{n_i}^{(i)}\}$ are of dimension $\chi \times \chi$

where the *bond dimension* χ reflects the entanglement across the specific cut of the chain. Systems with extensive entanglement require a large bond dimension to be represented faithfully as an MPS, while low entanglement states can be efficiently described with small bond dimension. Specifically, whether a state permits an efficient MPS representation can be diagnosed from the scaling of its Rényi entropy [35][36].

Representation of the state as an MPS, allows for powerful variational methods, such as the DMRG-algorithm [22], which is able to optimize the full state by optimizing the single-site matrices one by one in a sequential manner [37]. Our parameterized local basis optimization approach (sec. III B) builds upon the MPS structure.

E. Simulation Errors

In order to represent the bosonic operators numerically, we have to introduce a cutoff of the infinite dimensional space. This cutoff together with the truncation of the ansatz space to MPS with finite bond-dimension:

Fock basis cutoff: D , MPS bond dimension: χ

constitute the two hyper-parameters of the variational problem, and together fully control the simulation error.

One of the key properties of our variational method is that the ground state energy is zero in the continuum. Therefore, any truncation or approximation error will increase the variational energy $\langle \mathcal{H} \rangle$ above zero. In other words:

$$\text{"Magnitude of Simulation Error"} \equiv \langle \mathcal{H} \rangle \geq 0 \quad (18)$$

A concrete manifestation of this relation can be found by assuming that $|\psi\rangle$ remains the exact ground state of \mathcal{H} after truncation and by exploiting the unit-sized harmonic spectrum of \mathcal{H} . Then, in the worst case:

$$F \geq 1 - \langle \mathcal{H} \rangle \quad (19)$$

where F is the fidelity, i.e. the absolute square of the overlap between the variational state and $|\psi\rangle$. The confirmation of such a bound is shown in sec. IV (fig. 4c).

1. Errors from the cutoff D

Truncating the infinite-dimensional operators introduces an error in the ground state eigenvalue problem. Let P be the projection

$$P = \bigotimes_i^N \left(\sum_{n_i=0}^{D-1} |n_i\rangle \langle n_i| \right) \quad (20)$$

and $Q = \mathbb{1} - P$ the complement. Then it becomes apparent that the eigendecomposition of the naively truncated

problem does not in general completely align with the eigendecomposition of the original problem. Namely, we are interested in $P|\psi\rangle$ where

$$\mathcal{H}|\psi\rangle = 0 \quad (21)$$

However, from insertion of the identity we see that:

$$P\mathcal{H}P|\psi\rangle = -P\mathcal{H}Q|\psi\rangle \quad (22)$$

So unless $|\psi\rangle$ is completely contained in the projected space (i.e. $Q|\psi\rangle = 0$) the ground state of $P\mathcal{H}P$ will be misaligned with $P|\psi\rangle$.

The exact effective Hamiltonian, for which the ground state will align with $P|\psi\rangle$, is found by also applying Q from the left on eq. 22 and solving for $P|\psi\rangle$ [38]:

$$(P\mathcal{H}P - P\mathcal{H}Q(Q\mathcal{H}Q)^{-1}Q\mathcal{H}P)P|\psi\rangle = 0 \quad (23)$$

In general, the exact effective Hamiltonian is not available, and one would resort to perturbation theory or targeted unitary transformations to achieve an approximate effective description [39]. In this work we simply consider $P\mathcal{H}P$, incurring the error:

$$\varepsilon_D := |||\psi\rangle - P|\psi\rangle||^2 = 1 - \langle\psi|P|\psi\rangle \quad (24)$$

ε_D captures the rotation induced on the ground state when going from the full description to the naively truncated description. For GBS ε_D can be directly assessed as shown in sec. III A.

2. Assessing the error from finite bond dimension

Any quantum state can be cast as an MPS. However, in general, the bond dimension will have to grow exponentially with system size. If instead the bond dimension is capped at a constant value, an approximation error is introduced.

In the case of GBS the compression of the Gaussian state to its MPS form is based on sequential singular value decompositions [40][18]. Therefore, we denote the compressed state $|SVD\rangle$. The singular values themselves inform about the error introduced by a finite bond dimension χ , which we denote ε_χ (see SM sec. 3 for a formal description [41]). Specifically, they provide an upper bound on ε_χ , while mitigating the complications of the infinite dimensional Hilbert space, and therefore we use these as a surrogate [42][43][44]:

$$\varepsilon_\chi \sim \sum_{n=1}^{N-1} \left(1 - \sum_{k \leq \chi_n} (\sigma_k^{(n)})^2 \right) \quad (25)$$

$\sigma^{(n)}$ are the singular values and are equal to the Schmidt coefficients when bipartitioning the state into $\{1, \dots, n\}$ and $\{n+1, \dots, N\}$.

Outside GBS, it is impractical to obtain the compressed state through sequential SVDs of the full state.

Instead, the MPS approximation $|MPS\rangle$ to the ground state $|GS\rangle$ of the Hamiltonian $P\mathcal{H}P$ is obtained through variational methods. The error induced by imposing an MPS ansatz on the variational state $|MPS\rangle$, can generally be estimated by its variance wrt. the Hamiltonian [45]:

$$\sigma_{\mathcal{H}} = \sqrt{\langle(P\mathcal{H}P)^2\rangle - \langle P\mathcal{H}P\rangle^2} \quad (26)$$

The compression error $\varepsilon'_\chi \sim \sigma_{\mathcal{H}}$ is the degree to which $|MPS\rangle$ remains an eigenstate of the Hamiltonian. Schematically, the relation between the relevant states in terms of the errors introduced is:

$$\begin{array}{ccc} & & |GS\rangle \xrightarrow{\varepsilon'_\chi} |MPS\rangle \\ & \nearrow \varepsilon_D & \\ |\psi\rangle & & \\ & \searrow \varepsilon_\chi & \\ & & |SVD\rangle \end{array} \quad (27)$$

F. Sampling in Presence of Noise

For experimental GBS, the proper description will be in terms of a mixed state and it is not immediately clear how to apply our method which derived the variational Hamiltonian from a pure state. However, for mixed state GBS, it is possible to divide the sampling problem into two parts [46][18]. Namely, a pure state GBS problem followed by a classical sampling problem, which can be done efficiently. Specifically, one decomposes the mixed state covariance matrix into two using semi-definite programming:

$$V = Q + C \quad (28)$$

Q constitutes a pure state covariance matrix and contains all the entanglement information and the exponentially hard computational complexity. C is the covariance matrix of a classical zero-mean Gaussian distribution of random displacements, which encapsulates all the noise of the experiment. After obtaining the pure state defined by Q , random displacement operations are applied onto this state and a PNR measurement is sampled from the resulting state. These samples will follow the distribution defined by the noisy GBS problem. The computationally hard part of this problem is obtaining the pure state amplitudes $\langle\mathbf{n}|\psi\rangle$, which can be done with our method.

In generality, noise in the form of photon loss, present throughout the circuit, can be commutated to the end of the circuit [47][48]. It can then be accounted for, by obtaining the pure state $\langle\mathbf{n}|\psi\rangle$ right up until the photon loss is applied. The photon loss constitutes entangling the pure state with auxiliary degrees of freedom through beam-splitters and tracing out the auxiliary degrees of freedom, thereby obtaining the density matrix describing the noisy sampling problem.

III. METHODS

In this section, we will introduce Local Basis Optimization, which is solvable in the Gaussian case. For non-Gaussian sampling problems we introduce the parameterized LBO method and an accompanying optimization strategy.

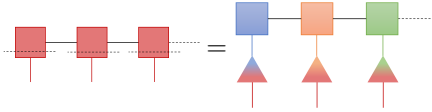
A. Optimal Local Basis and Cutoff Error Bounds

It is of central importance to reduce the error associated to the truncation of the bosonic operators. To this end Local Basis Optimization seeks to reduce the physical dimension D of the local Hilbert space by introducing a new local basis of smaller dimension $d \ll D$ [49]. The optimal local basis depends on the state that is being considered and is given by the eigenvectors of the reduced density matrix:

$$\rho_i = \text{Tr}_{k \neq i}(|\psi\rangle\langle\psi|) \quad (29)$$

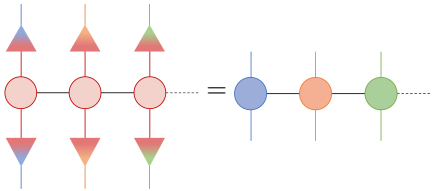
The error induced by decreasing the physical dimension is related to the eigenvalues of the reduced density matrix and is minimized by keeping the eigenvectors corresponding to the largest eigenvalues.

LBO can also be viewed as a compression of the physical index by SVDs:



$$(30)$$

where each color denotes their own basis and the red color denotes the basis comprised of photon number eigenstates. The Hamiltonian can be cast in the optimal local basis:



$$(31)$$

thus reducing the numerical effort for subsequent calculations.

Usually the target state is not known beforehand and LBO has to be carried out simultaneously with the variational optimization in a self-consistent manner [23][50]. However, in our setting, the target state is known and specifically for Gaussian states, it is possible to derive the optimal local basis analytically.

1. Optimal Local Basis for Gaussian States

The reduced density matrix of a Gaussian state is found by integrating out all the modes k different from

i and is itself a Gaussian state. The Gaussian density matrix is diagonalized as [28]:

$$\rho_i = \sum_{m_i=0}^{\infty} \frac{\bar{n}_i^{m_i}}{(\bar{n}_i + 1)^{m_i+1}} \hat{S}_i^\dagger |m_i\rangle \langle m_i| \hat{S}_i \quad (32)$$

with $\bar{n}_i = \langle \hat{n}_i \rangle_G$ being the mean number of photons and $|m_i\rangle$ the photon number eigenstates. The Hamiltonian is rotated to the basis spanned by $\hat{S}_i^\dagger |m_i\rangle$ by transforming under \hat{S}_i :

$$\mathcal{H} \rightarrow \hat{S}_i \mathcal{H} \hat{S}_i^\dagger \quad (33)$$

The action of the quadratic unitary \hat{S}_i on the quadrature operators is:

$$\hat{S}_i \begin{pmatrix} \hat{X}_i \\ \hat{P}_i \end{pmatrix} \hat{S}_i^\dagger = S_i \begin{pmatrix} \hat{X}_i \\ \hat{P}_i \end{pmatrix} \quad (34)$$

Where the matrix representation S_i is found by Williamson decomposition [51] of the covariance matrix of the reduced Gaussian density matrix V_i [52]:

$$V_i = S_i \begin{pmatrix} \hbar(\bar{n}_i + \frac{1}{2}) & 0 \\ 0 & \hbar(\bar{n}_i + \frac{1}{2}) \end{pmatrix} S_i^T \quad (35)$$

Notice from eq. 10 and eq. 35 that eq. 33 has the effect of diagonalizing the submatrix for mode i . This procedure is carried out for each mode.

Solving the variational problem in the optimal local basis $\prod_i \hat{S}_i^\dagger |\mathbf{m}\rangle$ does not grant immediate access to the wanted Fock state overlaps $\langle \mathbf{n} | \psi \rangle$. However, the fact that the basis transformation was local, means that it is a simple matter to apply the matrix representation of the inverse transformation:

$$\langle n_i | \psi \rangle = \sum_{m_i}^{d_i} \langle n_i | \hat{S}_i^\dagger | m_i \rangle \langle m_i | \psi_{\text{LBO}} \rangle \quad (36)$$

where $\langle m_i | \psi_{\text{LBO}} \rangle$ is to be understood as the result obtained when solving the variational problem defined by the local basis optimized Hamiltonian (eq. 33). Note, that while the representation has a cutoff d_i we can calculate the overlap for any $\langle n_i |$. This results in the notion of an effective cutoff for which the matrix $V \in \mathbb{C}^{D \times d}$ with $V_{nm} = \langle n_i | \hat{S}_i^\dagger | m_i \rangle$ is an isometry (up to machine precision):

$$\text{"Effective cutoff"} = \text{minimal } D \text{ s.t. } V^\dagger V = \mathbb{1}_{d \times d} \quad (37)$$

2. Truncation Error from Singular Values

Besides obtaining the optimal local basis from the unitaries in eq. 32, having access to all the singular values

$$(\sigma_k^{(i)})^2 \equiv \frac{\bar{n}_i^k}{(\bar{n}_i + 1)^{k+1}} \quad (38)$$

is also beneficial. Similar to eq. 25, there is an inequality associated to the SVD compression of eq. 30 and it is a simple matter to prove (see SM sec. 4 [41]) that the spectral weight left after projection is bounded by the sum of the remaining singular values:

$$N\varepsilon \geq \underbrace{1 - \langle \psi | P | \psi \rangle}_{\varepsilon_D} \geq \varepsilon \quad (39)$$

where $0 \leq \varepsilon \leq 1$ is defined as:

$$\varepsilon := 1 - \frac{1}{N} \sum_{i=1}^N \sum_{k \leq d_i} \left(\sigma_k^{(i)} \right)^2 \quad (40)$$

Empirically we find that the upper limit saturates as ε approaches zero (see figure 2).

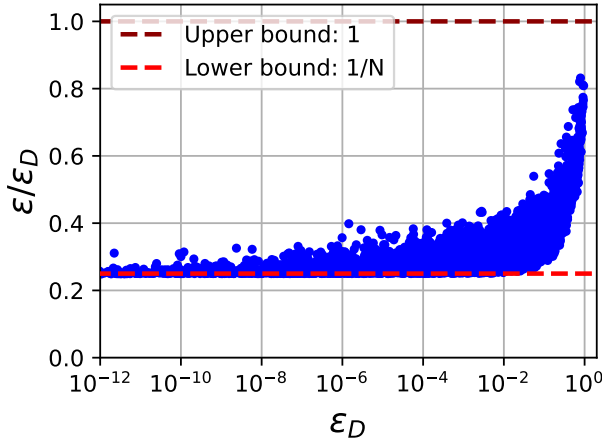


FIG. 2: This shows the bound as: $\frac{1}{N} \leq \frac{\varepsilon}{1 - \langle \psi | P | \psi \rangle} \leq 1$ with $N = 4$. The data suggests that for low enough truncation errors, the spectral weight can be assessed by: $\varepsilon_D \approx N\varepsilon$

B. Parameterized Local Basis Optimization

Since we are unable to derive the equivalent of eq. 32 for a generic non-Gaussian state, we have to resort to approximate methods for the local basis optimization. Standard techniques [24] consists of variationally optimizing d orthonormal linear combinations of the $D \gg d$ lowest lying photon number states, which then serves as the basis in which the MPS is optimized. The local basis can be improved by simultaneously shifting the photon number basis by the expectation value of \hat{X}_i during the optimization. The displacement is most aptly implemented directly in the matrix representation of the operators. These three steps are iterated upon until convergence and the next site is optimized in a similar fashion.

Incorporating more general local transformations other than displacement will enhance the representational power of the basis. By utilizing non-linear optimization methods, the associated basis parameters can be learned together with the basis, circumventing the need for sequential optimizations. We study the effects of this approach without the $(D \times d)$ dimensional linear basis transformation, however, the two methods can be combined, as we show in the supplemental material sec. 6 [41].

In choosing a parameterization for the local basis transformation, it is important to consider how the inverse transformation is obtained (eq. 36). To this end, we consider the operations included in the python module StrawberryFields [53][54]. Through the use of this library, we can obtain the matrix representation of the transformation from the local basis to the original Fock basis.

There are multiple ways of composing a local basis unitary from the single site gates. The choice will affect the implementability and the size of the space the variational ansatz exists in. We choose a simple composition which make use of each gate once to become the following parameterized unitary:

$$\mathcal{U}^\dagger = K^\dagger(\kappa) P_3^\dagger(\gamma) P_2^\dagger(s) R^\dagger(\theta) S^\dagger(re^{i\phi}) D^\dagger(\underbrace{\alpha_x + i\alpha_p}_{\alpha}) \quad (41)$$

The gates are from left to right the Kerr, Cubic phase, quadratic phase, rotation, squeezing and displacement gate. The additional variational parameters are $\{\alpha_x, \alpha_p, r, \phi, \theta, s, \gamma, \kappa\}$, all of which are $\in \mathbb{R}$. Each mode has its own set of tuneable basis parameters. The action of \mathcal{U} on the constituents of the Hamiltonian is:

$$\begin{aligned} \mathcal{U}^\dagger \hat{a} \mathcal{U} &= \hat{a} e^{i\kappa(2\hat{n}-1)} \left(\left(1 + i\frac{s}{2} \right) e^{i\theta} \cosh r + i\frac{s}{2} e^{i(\phi-\theta)} \sinh r \right) \\ &\quad - e^{-i\kappa(2\hat{n}-1)} \hat{a}^\dagger \left(\left(1 - i\frac{s}{2} \right) e^{i(\phi-\theta)} \sinh r - i\frac{s}{2} e^{i\theta} \cosh r \right) \\ &\quad + i\frac{\gamma}{2} (e^{i\theta} \cosh r + e^{i(\phi-\theta)} \sinh r) \\ &\quad \times (2\hat{n} + 1 + \hat{a} \hat{a} e^{i4\kappa(\hat{n}-1)} + e^{-i4\kappa(\hat{n}-1)} \hat{a}^\dagger \hat{a}^\dagger) + \alpha \end{aligned} \quad (42)$$

The matrix representation of the transformed operator is easily obtained, as outlined in the SM sec. 5 [41]. For reasons discussed in sec. V we adopt an optimization strategy inspired by single site DMRG. We only optimize the variational parameters related to a single site before moving on to the next, thereby optimizing the full state and the local basis in a sweeping manner, as shown in fig. 3.

IV. RESULTS

A. Gaussian Boson Sampling

The problem of Gaussian Boson sampling will now be treated with our variational approach. The problem consists of finding the ground state of eq. 10 in the optimal

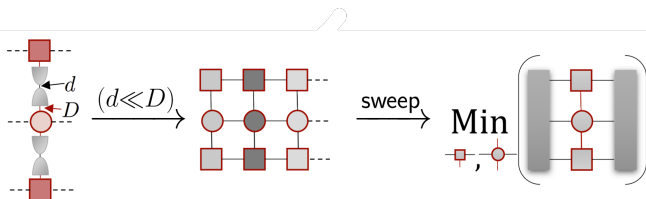


FIG. 3: Schematic of our variational algorithm that utilizes parameterized local basis optimization. It works much like DMRG, in that it defines an effective Hamiltonian for each individual site and performs an optimization for that site. The difference is that the MPO also contains variational parameters which are optimized using Gradient Descent algorithms due to the non-linearity introduced by eq. 42

local basis described in sec. III A. The variational optimization algorithm is chosen to be single site DMRG as implemented in the python library Quimb [55].

GBS is well suited as a benchmarking problem, since the pure state amplitudes can be calculated by other means and compared against. For small system sizes (up to $D^N \sim 10^6$) the python library "thewalrus" is utilized [56]. For larger system sizes, one can compare to state of the art classical simulations of GBS, also based on Tensor Network methods, such as [18].

As noted in sec. II E, there exists two simulation errors, originating from the cutoffs in the two hyper-parameters D (physical dimension) and χ (bond dimension). In order to validate this statement and more importantly, the statement of eq. 18, namely that all simulation errors are captured by the variational energy, we show the state overlaps as a function of the associated error amplitudes in figure 4. A schematic overview of the relevant states is found in eq. 27 and for the numerics, each state is to be understood as a normalized version of its projection wrt. P (eq. 20). Therefore, the y-axis of figure 4 shows to what degree the finite dimensional vectors align with each other.

The simulation error (colored blue in fig. 4), is directly correlated with the variational energy and how much of this is due to the finite bond dimension can be estimated by the variance of the variational energy (fig. 4c). How ε_D and ε_χ individually contribute to the simulation error is shown in fig. 4a and 4b with the insets establishing the correctness of these error metrics. The variationally obtained state is generally worse than the compressed state in simulating $|\psi\rangle$ due to finite ε_D . This discrepancy can be mitigated by working towards approximating the effective Hamiltonian, eq. 23.

We also compare our method to experiment, namely the Borealis M72 experiment [57]. This is done by using the procedure, described in sec. II F, to obtain a pure state covariance matrix from the noisy experiment. Using our method with $D = 6$ and $\chi = 240$, we obtain the amplitude for all overlaps between the variational state and the associated Fock states. By focusing on only the

TABLE I: Photon Number Expectations [58]

$$\left(\sum_j \langle \hat{n}_j \rangle \pm \sqrt{\sum_j \langle \hat{n}_j^2 \rangle - \langle \hat{n}_j \rangle^2} \right)$$

Experiment	Williamson Decomposition	Semi-Definite Programming	Local Basis Optimization
BorealisM16	2.016 ± 1.614	0.5487 ± 0.807	0.470 ± 0.696
BorealisM72	7.072 ± 2.895	1.749 ± 1.388	1.601 ± 1.280
BorealisM216High	36.569 ± 6.759	6.541 ± 2.686	6.094 ± 2.504
BorealisM288	46.930 ± 7.585	10.687 ± 3.422	10.106 ± 3.235

The total photon number distribution for different pure states related to the experiments of the first column.

The smaller the numbers, the easier it is to represent the state. The second column refers to the pure states obtained from Williamson Decomposition of the mixed state covariance matrices of the noisy experiments [46].

The second column extracts the pure state with minimal expected number of excitations, through the use of Semi-Definite Programming (SDP) [18]. This results in only a slight improvement when applying LBO (sec. III A) on the SDP-recovered pure state as shown in the last column. Nonetheless, an increase in the effective cutoff (eq. 37) of a factor of $D \sim 3d$ is awarded for free, with the optimal local basis approach.

two-excitation Fock state amplitudes, we can calculate the true target distribution analytically, as implemented in e.g. thewalrus. The comparison shows agreement as seen in fig. 5.

In order to investigate the benefit of the optimal local basis for GBS problems, we consider the total photon number distribution of the Borealis experiments in three different cases, as shown in table I

B. Beyond GBS with parameterized LBO

Having established the validity of our variational method and usage of the variational energy as a measure of simulation error, we will now investigate the applicability of our pLBO approach (described in section III B) by applying it to the non-Gaussian sampling problem defined in sec. II C 2.

The parametric local basis transformations are implemented directly in the matrix representation of eq. 42. Consequently, we would like direct access to the basis parameters in the Hamiltonian, so that they can be updated at each iteration of the optimization algorithm. The most efficient way of doing so, is to have an algebraic construction of the MPO so that the MPO cores can quickly be generated with the new basis-parameters. The problem of writing down an MPO can be solved with Finite State Machines [59][60][61]. However, in the SM sec. 7 [41], we

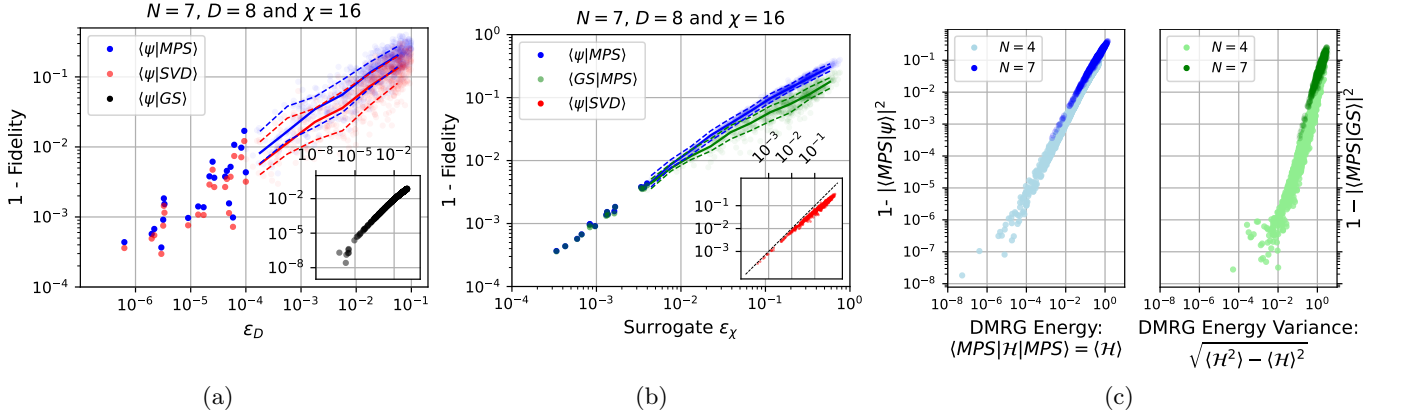


FIG. 4: Each point is a GBS simulation of a random pure state covariance matrix in the optimal local basis and with zero means. **(a)&(b)** "Fidelity" between two states $|\psi\rangle$ and $|\phi\rangle$ refers to $\frac{|\langle\phi|P|\psi\rangle|^2}{\langle P\rangle_\phi\langle P\rangle_\psi}$, i.e. how much they align. The full line is the mean of a bin of points while the dashed are the $\pm 31.4\%$ percentile (corresponding to \pm one standard deviation for Gaussian distributions). "Surrogate ϵ_χ " is the upper bound on ϵ_χ , eq. 25. The fact that the normalized overlap between $|\psi\rangle$ and $|SVD\rangle$ is well correlated and bounded by this metric is an indicator that this is in fact a good surrogate for ϵ_χ . **(c)** In all cases $D=8$ and $\chi=16$. Clearly, the simulation errors can be estimated by the different moments of the variational Hamiltonian as explained in sec. II E.

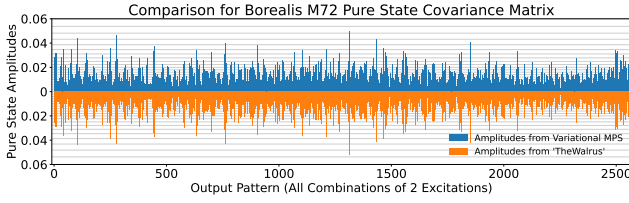


FIG. 5: Direct comparison of pure state amplitudes obtained through our variational approach to the true distribution for the Borealis M72 experiment [57], as defined by the Q in eq. 28. The variational energy in this case is $\langle H \rangle \approx 0.159$ while the normalized overlap with the compressed true state is: $|\langle SVD|MPS \rangle|^2 \approx 0.977$

present our manual approach for constructing the relevant MPOs for the non-Gaussian Hamiltonian eq. 14

As a simple initial example, we consider the fully connected CZ-gate acting on the vacuum, such that the covariance matrix is the identity matrix and the displacements are zero. In this case, the only parameter defining the sampling problem is κ , with $\kappa=0$ being the trivial problem of sampling the vacuum. When κ is raised, it will be increasingly more difficult to simulate the sampling problem, as captured by deviations in the variational energy $\langle H \rangle$ from zero (sec. II E). In figure 6a we show the variational energy for the pLBO approach compared to doing no basis optimization. The pLBO improves the simulation because it is able to diminish ϵ_D by achieving a high effective cutoff. ϵ_χ is practically negligible in this example since increasing χ had no effect on the variational energy. However, the Fock-basis energy is

for the dense ground state ($|GS\rangle$).

If we apply the global CZ-gate on GBS problems instead, we achieve the results of figures 6b and 6c. Each point is a randomly generated pure state covariance matrix, post-selected to not deviate too far from the identity matrix, which results in the GBS problem being easily simulated. However, raising κ increases the complexity. Still, pLBO is better able to simulate the sampling problem with minimal overhead. The reason is again the increased cutoff as seen in fig. 6c. Interestingly the variance of the Hamiltonian and thus ϵ'_χ is affected by the application of the single site gates of pLBO. While single site gates cannot change the entanglement structure, they will affect the truncation error ϵ_D which in turn affect the ground state $|GS(\epsilon_D)\rangle$ and its entanglement structure.

To prove the scalability of our approach, we apply our method to a likewise sampling problem with $N=71$, as shown in fig. 7. Similar results are found in this moderately large N regime and indicates the usefulness of LBO techniques for bosonic variational problems.

V. DISCUSSION

1. Gaussian Boson Sampling

State of the art simulation of GBS [18] compresses the Gaussian state to MPS form through the use of sequential singular value decompositions [40]. While such a single-shot compression technique is quasi-optimal (eq. A.16), it is generally the case, that the MPS obtained from iterative methods, such as DMRG, will better approximate the optimal MPS representation [62][63]. Therefore, if

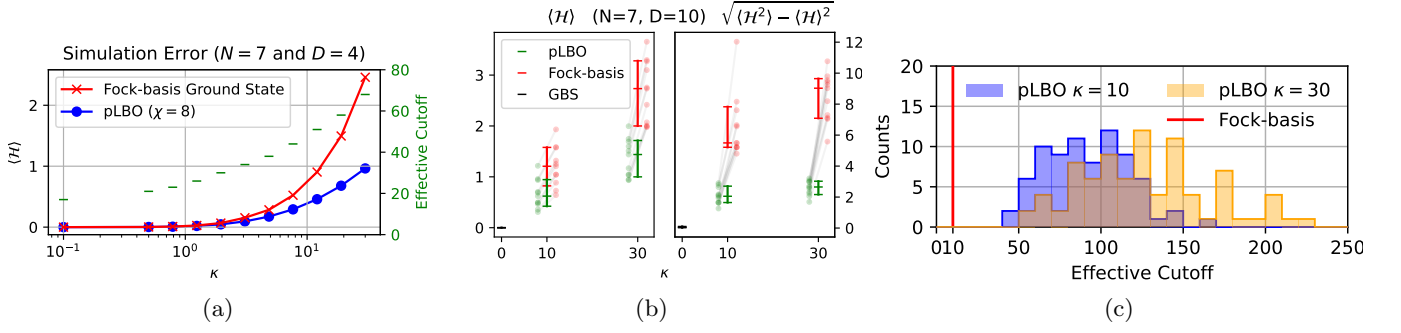


FIG. 6: **(a)** The variational energy as a function of the “strength” (κ) of the non-Gaussian circuit for the Fock basis and the pLBO approach. The adaptability of the parameterized local basis increases the range of applicability of our variational approach for sampling problems. Increasing χ in this case does not lower $\langle \mathcal{H} \rangle$, however, it still lowers $\sigma_{\mathcal{H}}$, meaning ε_D is the dominant error. **(b)** Simulation of eleven GBS problems with the non-Gaussian global CZ-gate applied, for increasing κ . The horizontal lines indicate the median and the 50% \pm 34.1%th quantile. The grey lines group points that share the same GBS covariance matrix. With this we clearly see that pLBO always betters the simulation and increasingly so for more complex simulations. **(c)** Distribution of the effective cutoffs (eq. 37) for the pLBO simulations of figure b. Recall that all computations for pLBO is done at a numerical cutoff of $D = 10$, however, an increase of an order of magnitude is achieved in the corresponding cutoff in the Fock basis. Furthermore, the increase is not simply due to a large displacement or squeezing, but a learned consortium of the basis parameters, as we show in the supplemental material figure 9 [41]

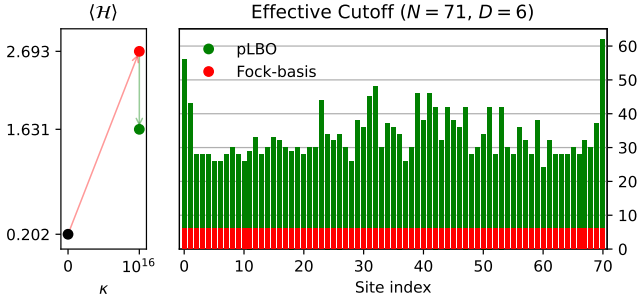


FIG. 7: Application of our pLBO approach to a sampling problem with $N = 71$ modes, showcasing its scalability. For GBS ($\kappa = 0$) $\chi = 240$ whereas for the non Gaussian case the optimization converged already at $\chi = 64$. The reason for the exponentially large κ in this case is due to the parameterization of the global CZ-gate not accounting for the number of modes:

$$\exp(-i\kappa \prod_n x_n) = \exp(-i \prod_n \sqrt[\kappa]{\kappa} x_n).$$

the error ε_D is eliminated, e.g. by deriving the exact effective Hamiltonian (eq. 23), our method provides a way to obtain the optimal MPS representation of general continuous variable sampling problems. While efforts towards obtaining the effective Hamiltonian could help alleviate the ε_D error another approach consists of utilizing the symmetries present in the sampling problems [48], thereby effectively increasing the number of basis states [64].

2. Optimization strategies and gradient based optimization algorithms

By framing the sampling problem as a variational problem we can leverage all the techniques developed for solving such optimization problems. While we will not provide an extensive overview of such methods, we will here touch on our experiences, especially wrt. the parameterized Local Basis Optimization.

Since we are unable to employ linear methods (due to eq. 42), we opt for gradient based optimization methods, where the gradients are obtained by auto differentiation. If we were to optimize the whole state at once (akin to what is done in [65]), the memory consumption of the gradient calculation would become the limiting factor. Instead, each site is optimized sequentially akin to single site DMRG. Seemingly, this has a regularizing effect and yields consistently equal or better minima than optimizing all variational parameters at once. However, further studies would have to distinguish whether this is dependent on the specific optimization algorithm.

As mentioned earlier, we have omitted the $(D \times d)$ linear transformation which would possibly diminish the critical error ε_D . We do this to focus the study on the variational parameters that are implemented algebraically. Still, in the supplemental material [41] sec. 6 we apply a single round of linear basis transformation optimization to the results of fig. 7. The gain is larger for the Fock-basis simulation since the pLBO already has enhanced the basis. However, we expect that with a simultaneous optimization of the two methods, they will be better suited to leverage each other and an even higher effective cutoff may be reached. This remains for future

studies to verify.

Regarding optimization algorithms, the division of the optimization into sub-problems is the setting for Stochastic Gradient Descent. Within this group of algorithms, ADAM [66] is a popular choice and we also found this to perform better than non-linear Conjugate Methods in reaching better minima, but at the cost of more iterations. L-BFGS [67] was also found to converge fast, but only to local minima, from which SGD could be employed to converge towards a better minimum. We believe this "initially fast then slow" behavior of L-BFGS is related to the considerations necessary when implementing DMRG [68]. Namely, that the immediate site should not be fully optimized within a sweep, since the effective Hamiltonian is only tentative as the other sites have yet to converge. This is in conflict with the line-search employed during a step of L-BFGS.

Using DMRG, the bond dimension χ is iteratively increased from $\chi = \mathcal{O}(1)$ to the wanted final bond dimension [69]. Doing this increases the speed of convergence, presumably due to a smaller need for large Krylov spaces, and it could be interesting to also implement the subspace expansion method in our optimization procedure [70][71]. However, at least for the present non-Gaussian sampling problem, we found that running pLBO at a small bond dimension ($\chi = \mathcal{O}(D)$) yielded equally good local basis as running it for the maximum bond dimension. Thus, we propose the following optimization procedure that best incorporates the strengths of both DMRG and the pLBO approach:

- Use DMRG in the Fock basis to produce a good initial guess for pLBO by sweeping until $\chi = \mathcal{O}(1) \rightarrow \chi = \mathcal{O}(D)$
- Optimize the local basis together with the variational state using Stochastic Gradient Descent.
- Use DMRG in the obtained optimal local basis to converge towards the optimal variational state with the maximal achievable bond dimension.

In fact, all results for the non-Gaussian example, except those shown in fig. 6a, were produced with this procedure.

3. Further direction of studies for pLBO

The parameterization of the local basis transformation eq. 41 is only one choice of many. The only overhead exists in the implementation of the matrix representation (see SM sec. 5 [41]). Therefore, it could be worthwhile to better understand the variational class spanned by the different parameterizations, in order to pick the optimal one.

In order to truly leverage the non-linear capabilities, we will now speculate in methods that would be able to also reduce the computational bond dimension, akin to

how the pLBO is able to reduce the computational physical dimension. One way of achieving this is to consider a parameterized classically simulatable circuit, e.g. a low-depth application of two-site gates. The transformation of the Hamiltonian under such a circuit would produce a collection of new terms with the coefficients depending non-linearly on the introduced parameters. Using Finite State Machines, one would be able to formulate the MPO in such a way, that the tuneable parameters are easily accessed and can thus be iteratively optimized using gradient based methods. The result is a ground state optimization algorithm that simultaneously learns an MPS representation, but which necessitates a smaller bond-physical-dimension than in the original formulation, together with the parameters for a small depth circuit that can be applied to retrieve the state in the original representation. Similar ideas of divide-and-conquer are utilized in simulating qubit-based quantum circuits in the form of a joint Schrödinger-Heisenberg picture approach [72][73][74]. Lastly, we emphasize that our simulation tool can be straightforwardly applied to the setup in [21], which may help determine the noise threshold for efficient classical approximate simulations in non-linear Boson Sampling.

ACKNOWLEDGMENTS

The authors would like to thank Rafael E. Barfknecht for his contributions in initializing the project. We further thank Andreas Michelsen, Jonas Neergaard-Nielsen and Karsten Flensberg for useful discussions. We acknowledge support from the AWS Center for Quantum Computing, the Carlsberg foundation and from the Novo Nordisk Foundation, Grant number NNF22SA0081175, Quantum Computing Programme.

The data and the code used to generate the results are accessible on GitHub at: github.com/Vinther901/Variational-Tensor-Networks-Boson-Sampling [75]

Supplemental Material

1. Effect of truncation on the spectrum of \mathcal{H}

Equation 3 and 5 results in: $\text{spec}(\mathcal{H}) \in \{0, 1, 2, \dots\} = \mathbb{N}_0$. Therefore, the distribution of spectral weight for the variational state $|MPS\rangle$ can be alluded to by:

$$\langle \mathcal{H} \rangle_{|MPS\rangle} = 0 \cdot |c_0|^2 + \sum_{m=1}^{\infty} m |c_m|^2 \quad (\text{A.1})$$

where $\sum_{m=0} |c_m|^2 = 1$ and $|c_0|^2$ is the object of interest. The $|c_m|^2$ is the total amplitude for all degenerate eigenvectors corresponding to the eigenvalue m and, in principle, the $|c_m|^2$'s could be fitted by also considering higher moments $\langle \mathcal{H}^n \rangle$.

One could be concerned that once the projections are introduced and thus the spectrum altered, this will no longer be true. However, it can be shown (fig. 8), that introducing the projections can only increase the eigenvalues [76]. Furthermore, when they do increase, it is because the truncation P introduced an error. Therefore, this metric remains trustworthy and a variational energy $\langle \mathcal{H} \rangle_{P|MP S} < 1$ determines an acceptable simulation, depending on the wanted degree of accuracy. A larger variational energy could still result in an acceptable simulation, however, it would be harder to recognize.

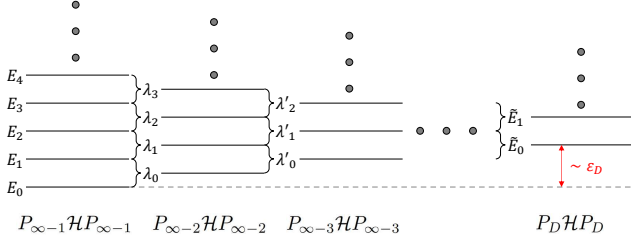
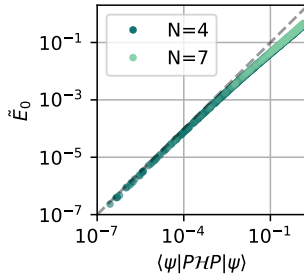


FIG. 8: Outline of how the energies of the truncated Hamiltonian is related to the original spectrum. $P_{\infty-1}$ is a projection onto a large enough Hilbert space such that the relevant part of the spectrum remain unaffected. The result of Cauchy's interlace theorem [77] [76] is that $\lambda_0 \leq \lambda'_0 \leq \lambda_1 \leq \lambda'_1 \leq \dots$ as exemplified by the brackets in the figure. The vertical lines are just one arbitrary manifestation and should not be taken too seriously (except for the E_i 's which truly are harmonic).

To first order in perturbation theory $\tilde{E}_0 \lesssim \langle \psi | P \mathcal{H} P | \psi \rangle$ (with $E_0 = 0$ and the inequality from numerical observation). Indeed, the ground state energy forms a very nice line as a function of $\langle \mathcal{H} \rangle_{P|\psi}$



2. Continuous Variable construction of \mathcal{H}

Suppose we want to sample from a few-body wavefunction of the form:

$$\psi(\mathbf{x}) \propto e^{-\varphi(\mathbf{x})} \quad (\text{A.2})$$

We can then define a generalized ladder operator:

$$\hat{A}_k = \sum_l M_{kl}^{-1/2} \left(\hat{P}_l - i \frac{\partial \varphi(\mathbf{x})}{\partial x_l} \Big|_{\mathbf{x}=\hat{\mathbf{X}}} \right) \quad (\text{A.3})$$

or in vector notation:

$$\hat{\mathbf{A}} = M^{-1/2} (\hat{\mathbf{P}} - i \nabla \varphi(\hat{\mathbf{X}})), \quad (\text{A.4})$$

where $\hat{\mathbf{X}}, \hat{\mathbf{P}}$ are the canonical position and momentum operators. The positive semi-definite matrix M is the mass-tensor. It is complimentary to the wavefunction and not given from experiment. The generalized ladder operators eq. A.4 are related to those found in supersymmetric quantum mechanics [78]

If we set the mass matrix to be the identity:

$$\hat{A}_k = \hat{P}_k - i \frac{\partial \varphi(\mathbf{x})}{\partial x_k} \Big|_{\mathbf{x}=\hat{\mathbf{X}}} \quad (\text{A.5})$$

then it is still true that $\hat{\mathbf{A}}|\psi\rangle = 0$ and that $\hat{\mathbf{A}}^\dagger \hat{\mathbf{A}}$ is positive semi-definite. However, with the introduction of a cutoff: $\hat{P} = \sum_k^{D-1} |\mathbf{k}\rangle \langle \mathbf{k}|$ we generally have, $\hat{P} \hat{A}_n \hat{P} |\psi\rangle = |\varphi_n\rangle$ where $|\varphi_n\rangle = 0$, only when no truncation error is made from the projection. Notice we still have positive semi-definiteness:

$$\langle \hat{\mathbf{A}}^\dagger \hat{\mathbf{A}} \rangle_{P\psi} = \sum_n \langle \varphi_n | \varphi_n \rangle \quad (\text{A.6})$$

with each $\langle \varphi_n | \varphi_n \rangle \geq 0$. We can then consider the introduction of an arbitrary symmetric and positive-definite matrix, denoted the mass matrix M :

$$\langle \hat{\mathbf{A}}^\dagger \hat{\mathbf{A}} \rangle_{P\psi} \rightarrow \langle \hat{\mathbf{A}}^\dagger M^{-1} \hat{\mathbf{A}} \rangle_{P\psi} \quad (\text{A.7})$$

such that

$$\langle \hat{\mathbf{A}}^\dagger M^{-1} \hat{\mathbf{A}} \rangle_{P\psi} = \sum_{nml} \langle \varphi_n | M_{nl}^{-1/2} M_{lm}^{-1/2} | \varphi_m \rangle = \sum_l \langle \tilde{\varphi}_l | \tilde{\varphi}_l \rangle \quad (\text{A.8})$$

which means that positive semi-definiteness is still obeyed but also that, once a truncation P is introduced, the simulation can be affected by the choice of M . Furthermore, the choice of M also has an effect on the commutation relations:

$$[\hat{A}_n, \hat{A}_m] = 0 \quad (\text{A.9})$$

$$[\hat{A}_n, \hat{A}_m^\dagger] = 2M_{ni}^{-1/2} \text{Re} \left(\frac{\partial^2 \varphi(\hat{\mathbf{X}})}{\partial x_i \partial x_j} \right) M_{jm}^{-1/2} \quad (\text{A.10})$$

such that if $\text{Re}(\varphi(\mathbf{x}))$ is at most quadratic, M can be chosen such that $[\hat{\mathbf{A}}, \hat{\mathbf{A}}^\dagger] = \mathbb{1}$. In the case of GBS and with this choice of M , a construction of \mathcal{H} from eq. A.4 yields exactly eq. 10.

Empirically (at least for GBS) the M that make the generalized ladder operators obey the canonical commutation relations also improves the simulation, as opposed to choosing M to be the identity. In general however, numerics indicate that even though it provides an improvement, it will not be the optimal choice. The optimal choice remains elusive and it also seems to dependent on the specific choice of D .

3. Compression to MPS form

Imposing an MPS form on a state, with maximum bond dimension χ , usually incurs an approximation error. In order to define this error, first note that MPSS

of a given bond dimension constitute a smooth manifold [79]

$$\mathcal{M}_\chi = \{|\Phi\rangle : \text{MPS-rank}(|\Phi\rangle) = \chi\} \quad (\text{A.11})$$

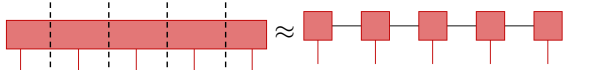
$\text{MPS-rank}(|\Phi\rangle) = \chi$ means that $|\Phi\rangle$ can be decomposed as eq. 15 and 16 with the predefined bond dimension structure χ . By compression of a generic state $|\psi\rangle$ to its MPS form $|\psi_{\text{MPS}}\rangle$, we mean a retraction from the point generally outside the manifold onto the manifold [37].

$$|\psi_{\text{MPS}}\rangle = \underset{|\Phi\rangle \in \mathcal{M}_\chi}{\text{argmin}} |||\psi\rangle - |\Phi\rangle||^2 \quad (\text{A.12})$$

There exist various approximations to $|\psi_{\text{MPS}}\rangle$ [37][80]. The one used in this paper relies on sequential singular value decompositions [40][18].

$$|\psi_{\text{MPS}}\rangle \approx |S\tilde{V}D\rangle, \quad |S\tilde{V}D\rangle = \frac{|S\tilde{V}D\rangle}{\sqrt{\langle S\tilde{V}D|S\tilde{V}D\rangle}} \quad (\text{A.13})$$

hence the name "SVD" for the compressed state. The tilde is to indicate that the state generally will not be normalized and a subsequent renormalization is necessary. Diagrammatically, if we denote a singular value decomposition by a dashed line, we may write:



$$(\text{A.14})$$

where the approximation is due to discarding all but the χ largest singular values. Note that the SVDs cannot be carried out simultaneously and it is this arbitrariness that demotes it from being an optimal rank χ compression.

Similarly to ε_D we define:

$$\varepsilon_\chi := |||\psi\rangle - |S\tilde{V}D\rangle||^2 \quad (\text{A.15})$$

which obeys the following bound [42][43][44]:

$$\varepsilon_\chi \leq \sum_{n=1}^{N-1} \sum_{k>\chi_i} \left(\sigma_k^{(n)}\right)^2 \leq (N-1) |||\psi\rangle - |\psi_{\text{MPS}}\rangle||^2 \quad (\text{A.16})$$

where $\sigma^{(n)}$ are the Schmidt coefficients when bipartitioning the state into $\{1, \dots, n\}$ and $\{n+1, \dots, N\}$. The fact that the compression by sequential SVDs is only quasi-optimal, is made manifest by the second inequality. The first inequality can also be found in the physics literature [36][80][81][82]. In practice, we circumvent the complications of the infinite dimensional Hilbert space in the definition of ε_χ , by using the sum of squared singular values as a surrogate, eq. 25.

4. Proof of ε_D bound (eq. 39)

Similar to eq. A.16, the subsequent application of SVDs in the LBO procedure has its own quasi-optimality

condition associated to it:

$$|||\psi\rangle - |\psi_{\text{LBO}}\rangle||^2 \leq \sum_{i=1}^N \sum_{k>d_i} \left(\sigma_k^{(i)}\right)^2 \leq N |||\psi\rangle - |\psi_{\text{opt}}\rangle||^2 \quad (\text{A.17})$$

where, now, $\sigma_k^{(i)}$ are the eigenvalues of eq. 29, $|\psi_{\text{LBO}}\rangle$ is the physical state after truncation of the physical dimension in the optimal local basis and $|\psi_{\text{opt}}\rangle$ is the associated best rank $\{r_i\}$ approximation of the state $|\psi\rangle$.

However, the LBO based on the diagonalization of the reduced density matrices of the exact target state, is the optimal one [49], which means that:

$$|\psi_{\text{LBO}}\rangle = P|\psi\rangle = |\psi_{\text{opt}}\rangle \quad (\text{A.18})$$

and the quasi-optimality condition eq. A.17 becomes eq. 39 by the two observations:

$$|||\psi\rangle - P|\psi\rangle||^2 = 1 - \langle\psi|P|\psi\rangle = \varepsilon_D \quad (\text{A.19})$$

$$\sum_{k=0}^{\infty} \left(\sigma_k^{(i)}\right)^2 = 1 \quad (\text{A.20})$$

One way of seeing A.18 is to note that what made eq. A.13 and eq. A.14 approximate, is that each sequential SVD is affected by the one before it. When considering eq. 30 this is no longer the case, each SVD can be performed in unison and the compression will be the optimal one [49].

5. Numerical implementation of the transformed bosonic operators eq. 42

The inability of a matrix representation of \hat{X} and \hat{P} to obey the canonical commutation relation is easily seen by taking the trace:

$$\text{Tr}([X, P]) = 0 \neq i\hbar \text{Tr}(\mathbb{1}) \quad (\text{A.21})$$

In the SQHO basis the "error" will accumulate near the point of truncation:

$$[X, P] \doteq i\hbar \begin{pmatrix} 1 & 0 & & \\ 0 & \ddots & \ddots & \\ & \ddots & 1 & 0 \\ & & 0 & 1-D \end{pmatrix} \quad (\text{A.22})$$

Therefore, if we rely on matrix multiplication to construct the matrix representation of eq. 42 we should be wary of this defect.

From eq. 42 we see that the sub/super-diagonal matrices a and a^\dagger only appears quadratically. The Hamiltonian eq. A.24 is also only quadratic in X or P such that a and a^\dagger only appears to the fourth power in any computation. Therefore, we can implement a matrix representation of eq. 42 simply by initializing a matrix representation of dimension $D+4$, construct the necessary

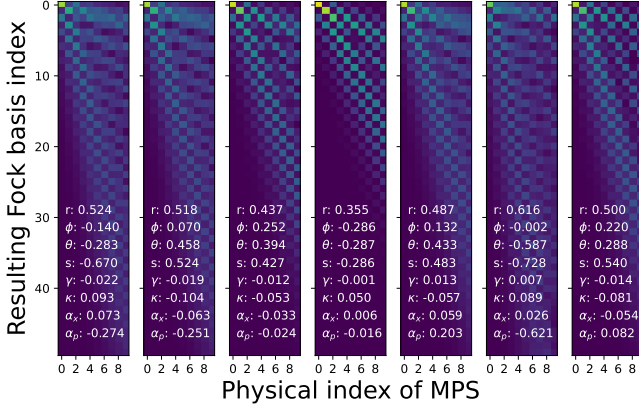


FIG. 9: Overview of the matrices $U_{nm}^{(i)}$ used to transform the variational state from the optimized local basis $|m_i\rangle$ to the Fock basis $|n_i\rangle$: $U_{nm}^{(i)} = \langle n_i | \mathcal{U}^{(i)} | m_i \rangle$. This example is taken from one of the results in fig. 6 with a large effective cutoff.

Hamiltonian terms of the MPO through matrix multiplication and then truncating the matrix representations from $D + 4 \rightarrow D$, thereby discarding the faulty entries.

6. Unifying Parameterized- and Linear Basis Optimization

Presented in figure A.23 (left) is the result of fig. 7 with the simulations of the non Gaussian experiments subjugated to a round of Local Basis Optimization similar to the approach of Ref. [24]. Specifically, starting from the MPS with physical dimension d (namely those that produced the results corresponding to the filled dots), the linear isometry eq. A.23 (right) is applied to each leg, effectively increasing the physical dimension from d to D .

The isometries are optimized similar to how the MPS is optimized in DMRG. Starting from e.g. the left, an effective Hamiltonian is defined for the left-most isometry. To optimize it, is to replace it with the lowest energy state of the effective Hamiltonian. All the isometries are optimized in a sweeping manner and after obtaining the optimal isometries the MPS is optimized with DMRG in the new d dimensional basis. The unfilled dots of fig. A.23 correspond to the results obtained after one such sweep of updating the isometries and running DMRG. Note that this is not the same procedure as applied in [24]. Furthermore, the parameterized- and standard Local Basis Optimization doesn't have to be applied separately. In fact, we suspect the optimal method would be to use L-BFGS to learn eq. A.23 simultaneously with the variational basis parameters implemented in the MPO and then use linear methods for learning the matrices of the MPS. However, this remains as a question for further studies.

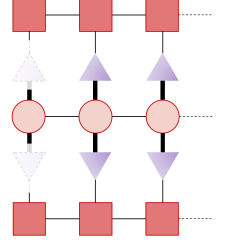
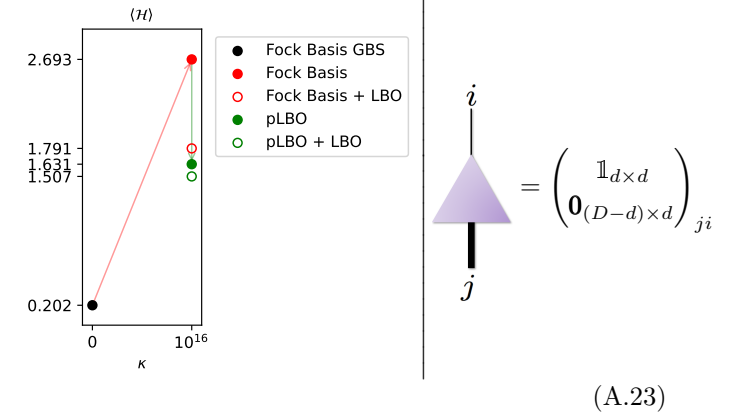


FIG. 10: Effective Hamiltonian for isometries



7. Algebraic MPO construction

The goal is to construct the MPO representation of eq. 14. Without loss of generality we can set $\boldsymbol{\mu} = \mathbf{0} = \boldsymbol{\nu}$ because it is a simple matter to introduce them later, through a redefinition of the quadrature operators (excluding those that stem from $\hat{\boldsymbol{\xi}}$); $\hat{X} \rightarrow \hat{X} - \mu$ and $\hat{P} \rightarrow \hat{P} - \nu$. While it is not immediately clear how to write up the MPO for the full Hamiltonian, we begin by writing out the Hamiltonian as:

$$\begin{aligned} \mathcal{H} = & \mathcal{H}_{\text{Gaussian}} + \kappa \sum_i \beta_{ii} \hat{P}_i \prod_{n \neq i} \hat{X}_n + \kappa^2 \sum_i b_{ii} \prod_{n \neq i} \hat{X}_n^2 \\ & + \kappa \sum_{i \neq j} \prod_{n \neq i, j} \hat{X}_n (\gamma_{ij} \hat{X}_i^2 + b_{ij} (\hat{X}_i \hat{P}_i + \hat{P}_i \hat{X}_i)) + \kappa \text{Tr} \gamma \prod_n \hat{X}_n + \kappa^2 \sum_{i > j} \beta_{ij} \hat{X}_i \hat{X}_j \prod_{n \neq i, j} \hat{X}_n^2 \end{aligned} \quad (\text{A.24})$$

The color coding is for more easily matching MPOs to Hamiltonian terms, also recall that this κ is due to eq. 11, not eq. 41. Here, $\mathcal{H}_{\text{Gaussian}}$ is eq. 10 and

$$\alpha_{ij} := \frac{1}{2}(V^{-1})_{ij} \quad \beta_{ij} := \frac{1}{2}(V^{-1})_{(N+i)(N+j)} =: 2b_{ij} \quad \gamma_{ij} := \frac{1}{2}(V^{-1})_{(i)(N+j)} \quad (\text{A.25})$$

The approach will be to write each term as an MPO and write the MPO of the full Hamiltonian as a block-diagonal MPO. This way, it is easier to derive the form of the MPO while we can still utilize the sparseness of the full MPO in the computations during optimization. We will for pedagogical reasons assume an odd number of modes, but it is straightforward to infer the MPOs for the case of an even number of modes. The index for the middle-most mode is denoted $N_m = (N + 1)/2$. All the MPOs will be of the form:

$$\text{MPO} : \left(\prod_{s=1}^{N_m-1} L_s \right) M_{N_m} \prod_{s=N_m+1}^N R_s \quad (\text{A.26})$$

where the L_s , M_{N_m} and R_s are matrices containing operators that only act in the Hilbert space corresponding to the sub-index s . The idea behind every construction is to use the L_s and R_s to collect the relevant operator strings and then collapse the vector-matrix-vector product with the right coefficients in M_{N_m} , in order to yield all the terms in the Hamiltonian. This will become clearer by considering the explicit construction of the matrices/MPO-cores in the following.

a. $\mathcal{H}_{\text{Gaussian}}$ MPO

Firstly, we define the single site operator:

$$\hat{S}_i = \frac{\alpha_{ii}}{2} \hat{X}_i^2 + \frac{\beta_{ii}}{2} \hat{P}_i^2 + \frac{\gamma_{ii}}{2} (\hat{X}_i \hat{P}_i + \hat{P}_i \hat{X}_i) - \frac{1}{2} \quad (\text{A.27})$$

such that the cores can be written as:

$$L_1 = \begin{pmatrix} \hat{X}_1 & \hat{P}_1 & \mathbb{1} & \hat{S}_1 \end{pmatrix} \quad (\text{A.28})$$

$$R_N^T = \begin{pmatrix} \hat{X}_N & \hat{P}_N & \mathbb{1} & \hat{S}_N \end{pmatrix} \quad (\text{A.29})$$

$$R_{s \neq N} = \begin{pmatrix} 0 & 0 & 0 & \hat{X}_s & 0 \\ 0 & 0 & \dots & 0 & \hat{P}_s & 0 \\ \mathbb{1} & 0 & & 0 & 0 & 0 \\ 0 & \mathbb{1} & & 0 & 0 & 0 \\ \vdots & & \ddots & \vdots & & \\ 0 & 0 & & \mathbb{1} & 0 & 0 \\ 0 & 0 & \dots & 0 & \mathbb{1} & 0 \\ \hat{X}_s \alpha_{s(s+1)} & \hat{X}_s \gamma_{s(s+1)} & & \hat{X}_s \gamma_{sN} & \hat{S}_s & \mathbb{1} \\ + \hat{P}_s \gamma_{(s+1)s} & + \hat{P}_s \beta_{s(s+1)} & & + \hat{P}_s \beta_{sN} & & \end{pmatrix} \quad (\text{A.30})$$

$$M_{N_m} = \begin{pmatrix} \alpha_{(N_m-1)(N_m+1)} & \gamma_{(N_m-1)(N_m+1)} & \alpha_{(N_m-1)N} & \gamma_{(N_m-1)N} & \alpha_{(N_m-1)N_m} \hat{X}_{N_m} + \gamma_{(N_m-1)N_m} \hat{P}_{N_m} & 0 \\ \gamma_{(N_m+1)(N_m-1)} & \beta_{(N_m-1)(N_m+1)} & \dots & \gamma_{N(N_m-1)} & \beta_{(N_m-1)N} & \gamma_{N_m(N_m-1)} \hat{X}_{N_m} + \beta_{(N_m-1)N_m} \hat{P}_{N_m} & 0 \\ & \vdots & \ddots & & \vdots & & \\ \alpha_{1(N_m+1)} & \gamma_{1(N_m+1)} & \alpha_{1N} & \gamma_{1N} & \alpha_{1N_m} \hat{X}_{N_m} + \gamma_{1N_m} \hat{P}_{N_m} & 0 \\ \gamma_{(N_m+1)1} & \beta_{1(N_m+1)} & \gamma_{N1} & \beta_{1N} & \gamma_{N_m1} \hat{X}_{N_m} + \beta_{1N_m} \hat{P}_{N_m} & 0 \\ \hat{X}_{N_m} \alpha_{N_m(N_m+1)} & \hat{X}_{N_m} \gamma_{N_m(N_m+1)} & \dots & \hat{X}_{N_m} \alpha_{N_m N} & \hat{X}_{N_m} \gamma_{N_m N} & \hat{S}_{N_m} & \mathbb{1} \\ + \hat{P}_{N_m} \gamma_{(N_m+1)N_m} & + \hat{P}_{N_m} \beta_{N_m(N_m+1)} & \dots & + \hat{P}_{N_m} \gamma_{N N_m} & + \hat{P}_{N_m} \beta_{N N_m} & & 0 \\ 0 & 0 & 0 & 0 & 0 & \mathbb{1} & 0 \end{pmatrix} \quad (\text{A.31})$$

The L_s are similar to the R_s^T but,

$$\begin{aligned} \text{with the row:} & \begin{pmatrix} \hat{X}_s \alpha_{s(s+1)} & \hat{X}_s \gamma_{s(s+1)} & \dots & \hat{X}_s \alpha_{s(s+1)} & \hat{X}_s \gamma_{sN} & \hat{S}_s & \mathbb{1} \\ + \hat{P}_s \gamma_{(s+1)s} & + \hat{P}_s \beta_{s(s+1)} & & + \hat{P}_s \gamma_{(s+1)s} & + \hat{P}_s \beta_{sN} & & \end{pmatrix} \\ \text{replaced by:} & \begin{pmatrix} \hat{X}_s \alpha_{s(s-1)} & \hat{X}_s \gamma_{s(s-1)} & \dots & \hat{X}_s \alpha_{s1} & \hat{X}_s \gamma_{s1} & \hat{S}_s & \mathbb{1} \\ + \hat{P}_s \gamma_{(s-1)s} & + \hat{P}_s \beta_{s(s-1)} & & + \hat{P}_s \gamma_{1s} & + \hat{P}_s \beta_{s1} & & \end{pmatrix} \end{aligned}$$

As an example, consider $N = 5$:

$$\begin{pmatrix} \hat{X}_1 & \hat{P}_1 & \mathbb{1} & \hat{S}_1 \end{pmatrix} \begin{pmatrix} 0 & 0 & \mathbb{1} & 0 & 0 & \alpha_{12}\hat{X}_2 + \gamma_{12}\hat{P}_2 \\ 0 & 0 & 0 & \mathbb{1} & 0 & \gamma_{21}\hat{X}_2 + \beta_{12}\hat{P}_2 \\ \hat{X}_2 & \hat{P}_2 & 0 & 0 & \mathbb{1} & \hat{S}_2 \\ 0 & 0 & 0 & 0 & 0 & \mathbb{1} \end{pmatrix} \begin{pmatrix} \alpha_{24} & \gamma_{24} & \alpha_{25} & \gamma_{25} & \alpha_{23}\hat{X}_3 + \gamma_{23}\hat{P}_3 & 0 \\ \gamma_{42} & \beta_{24} & \gamma_{52} & \beta_{25} & \gamma_{32}\hat{X}_3 + \beta_{23}\hat{P}_3 & 0 \\ \alpha_{14} & \gamma_{14} & \alpha_{15} & \gamma_{15} & \alpha_{13}\hat{X}_3 + \gamma_{13}\hat{P}_3 & 0 \\ \gamma_{41} & \beta_{14} & \gamma_{51} & \beta_{15} & \gamma_{31}\hat{X}_3 + \beta_{13}\hat{P}_3 & 0 \\ \hat{X}_3\alpha_{34} + \hat{P}_3\gamma_{43} & \hat{X}_3\gamma_{34} + \hat{P}_3\beta_{34} & \hat{X}_3\alpha_{35} + \hat{P}_3\gamma_{53} & \hat{X}_3\gamma_{35} + \hat{P}_3\beta_{35} & \hat{S}_3 & \mathbb{1} \\ 0 & 0 & 0 & 0 & \mathbb{1} & 0 \end{pmatrix} \begin{pmatrix} 0 & 0 & \hat{X}_4 & 0 \\ 0 & 0 & \hat{P}_4 & 0 \\ \mathbb{1} & 0 & 0 & 0 \\ 0 & \mathbb{1} & 0 & 0 \\ 0 & 0 & \mathbb{1} & 0 \\ \hat{X}_4\alpha_{45} + \hat{P}_4\gamma_{54} & \hat{X}_4\gamma_{45} + \hat{P}_4\beta_{45} & \hat{S}_4 & \mathbb{1} \end{pmatrix} \begin{pmatrix} \hat{X}_5 \\ \hat{P}_5 \\ \mathbb{1} \\ \hat{S}_5 \end{pmatrix}$$

(A.32)

Doing the matrix multiplication gives all the terms of the Hamiltonian used for Gaussian Boson Sampling:

$$\mathcal{H}_{\text{Gaussian}} = \frac{1}{2} \begin{pmatrix} \hat{\mathbf{X}} & \hat{\mathbf{P}} \end{pmatrix} \begin{pmatrix} \alpha & \gamma \\ \gamma^T & \beta \end{pmatrix} \begin{pmatrix} \hat{\mathbf{X}} \\ \hat{\mathbf{P}} \end{pmatrix} - \frac{N}{2} \quad (\text{A.33})$$

b. Second and third terms as MPOs

The MPOs of the second and third terms are more easily written up as:

$$\sum_i \beta_{ii} \hat{P}_i \prod_{n \neq i} \hat{X}_n = \begin{pmatrix} \beta_{11} \hat{P}_1 & \hat{X}_1 \end{pmatrix} \left(\prod_{n=2}^{N-1} \begin{pmatrix} \hat{X}_n & 0 \\ \beta_{nn} \hat{P}_n & \hat{X}_n \end{pmatrix} \right) \begin{pmatrix} \hat{X}_N \\ \beta_{NN} \hat{P}_N \end{pmatrix} \quad (\text{A.34})$$

$$\sum_i b_{ii} \prod_{n \neq i} \hat{X}_n^2 = \begin{pmatrix} b_{11} & \hat{X}_1^2 \end{pmatrix} \left(\prod_{n=2}^{N-1} \begin{pmatrix} \hat{X}_n^2 & 0 \\ b_{nn} & \hat{X}_n^2 \end{pmatrix} \right) \begin{pmatrix} \hat{X}_N^2 \\ b_{NN} \end{pmatrix} \quad (\text{A.35})$$

c. The MPO of the fourth and fifth term

Now, we consider the explicit construction of the following MPO:

$$\sum_{i \neq j} \prod_{n \neq i, j} \hat{X}_n (\gamma_{ij} \hat{X}_i^2 + b_{ij} (\hat{X}_i \hat{P}_i + \hat{P}_i \hat{X}_i)) + \text{Tr} \gamma \prod_n \hat{X}_n = \left(\prod_{s=1}^{N_m-1} L_s \right) M_{N_m} \prod_{s=N_m+1}^N R_s \quad (\text{A.36})$$

To this end we define: $\tilde{P}_n = \hat{X}_n \hat{P}_n + \hat{P}_n \hat{X}_n$ such that the cores can be written as:

$$L_1 = \begin{pmatrix} \mathbb{1} & \hat{X}_1 & \hat{X}_1^2 & \tilde{P}_1 \end{pmatrix} \quad (\text{A.37})$$

$$R_N^T = \begin{pmatrix} \mathbb{1} & \hat{X}_N & \hat{X}_N^2 & \tilde{P}_N \end{pmatrix} \quad (\text{A.38})$$

$$R_s = \begin{pmatrix} \hat{X}_s & 0 & \dots & & & & \\ 0 & \ddots & & & & & \\ \vdots & & \hat{X}_s & 0 & 0 & 0 & \\ & & 0 & \mathbb{1} & 0 & 0 & \\ & & 0 & \hat{X}_s & 0 & 0 & \\ & & 0 & \hat{X}_s^2 & 0 & 0 & \\ & & 0 & \tilde{P}_s & 0 & 0 & \\ & & 0 & 0 & \hat{X}_s & 0 & \\ & & 0 & 0 & 0 & \hat{X}_s & \\ & & & & \ddots & & \\ & & & & & 0 & \hat{X}_s & 0 \\ \gamma_{sN} \hat{X}_s^2 + b_{sN} \tilde{P}_s & \dots & \gamma_{s(s+1)} \hat{X}_s^2 + b_{s(s+1)} \tilde{P}_s & 0 & \gamma_{(s+1)s} & b_{(s+1)s} & \dots & b_{Ns} & \hat{X}_s \end{pmatrix} \quad (\text{A.39})$$

$$M_{N_m} = \begin{pmatrix} 0 & \dots & 0 & \gamma_{N_m 1} \hat{X}_{N_m}^2 + b_{N_m 1} \tilde{P}_{N_m} & \gamma_{(N_m+1)1} \hat{X}_{N_m} & b_{(N_m+1)1} \hat{X}_{N_m} & \dots & \gamma_{N1} \hat{X}_{N_m} & b_{N1} \hat{X}_{N_m} & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & \dots & 0 & \gamma_{N_m(N_m-1)} \hat{X}_{N_m}^2 + b_{N_m(N_m-1)} \tilde{P}_{N_m} & \gamma_{(N_m+1)(N_m-1)} \hat{X}_{N_m} & b_{(N_m+1)(N_m-1)} \hat{X}_{N_m} & \dots & \gamma_{N(N_m-1)} \hat{X}_{N_m} & b_{N(N_m-1)} \hat{X}_{N_m} & 0 \\ \gamma_{N_m N} \hat{X}_{N_m}^2 + b_{N_m N} \tilde{P}_{N_m} & \dots & \gamma_{N_m(N_m+1)} \hat{X}_{N_m}^2 + b_{N_m(N_m+1)} \tilde{P}_{N_m} & \text{Tr} \gamma \hat{X}_{N_m} & \gamma_{(N_m+1)N_m} & b_{(N_m+1)N_m} & \dots & \gamma_{N(N_m)} & b_{N(N_m)} & \hat{X}_{N_m} \\ \gamma_{(N_m-1)N} \hat{X}_{N_m} & \dots & \gamma_{(N_m-1)(N_m+1)} \hat{X}_{N_m} & \gamma_{(N_m-1)N_m} & 0 & 0 & \dots & \gamma_{NN_m} & b_{NN_m} & \hat{X}_{N_m} \\ b_{(N_m-1)N} \hat{X}_{N_m} & \dots & b_{(N_m-1)(N_m+1)} \hat{X}_{N_m} & b_{(N_m-1)N_m} & 0 & \ddots & \dots & \gamma_{NN_m} & b_{NN_m} & \hat{X}_{N_m} \\ \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ \gamma_{1N} \hat{X}_{N_m} & \dots & \gamma_{1(N_m+1)} \hat{X}_{N_m} & \gamma_{1N_m} & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ b_{1N} \hat{X}_{N_m} & \dots & b_{1(N_m+1)} \hat{X}_{N_m} & b_{1N_m} & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & \dots & 0 & \hat{X}_{N_m} & \vdots & \vdots & \ddots & 0 & 0 & 0 \end{pmatrix} \quad (\text{A.40})$$

The L_s are similar to the R_s^T but,

$$\begin{aligned} \text{with the row:} & \quad \left(\gamma_{sN} \hat{X}_s^2 + b_{sN} \tilde{P}_s \quad \dots \quad \gamma_{s(s+1)} \hat{X}_s^2 + b_{s(s+1)} \tilde{P}_s \quad 0 \quad \gamma_{(s+1)s} \quad b_{(s+1)s} \quad \dots \quad b_{Ns} \quad \hat{X}_s \right) \\ \text{replaced by:} & \quad \left(\gamma_{s1} \hat{X}_s^2 + b_{s1} \tilde{P}_s \quad \dots \quad \gamma_{s(s-1)} \hat{X}_s^2 + b_{s(s-1)} \tilde{P}_s \quad 0 \quad \gamma_{(s-1)s} \quad b_{(s-1)s} \quad \dots \quad b_{1s} \quad \hat{X}_s \right) \end{aligned}$$

d. The MPO for the last term

Again, we write the MPO as:

$$\sum_{i>j} \beta_{ij} \hat{X}_i \hat{X}_j \prod_{n \neq i,j} \hat{X}_n^2 = \left(\prod_{s=1}^{N_m-1} L_s \right) M_{N_m} \prod_{s=N_m+1}^N R_s \quad (\text{A.41})$$

this time with:

$$L_1 = \begin{pmatrix} \hat{X}_1^2 & \hat{X}_1 \end{pmatrix} \quad (\text{A.42})$$

$$R_N^T = \begin{pmatrix} \hat{X}_N^2 & \hat{X}_N \end{pmatrix} \quad (\text{A.43})$$

$$R_s = \begin{pmatrix} \hat{X}_s^2 & 0 & & & \\ \hat{X}_s & 0 & & & \\ 0 & \hat{X}_s^2 & & & \\ & & \ddots & & \\ & & & \hat{X}_s^2 & \\ 0 & \beta_{(s+1)s}\hat{X}_s & \dots & \beta_{Ns}\hat{X}_s & \hat{X}_s^2 \end{pmatrix} \quad (\text{A.44})$$

$$M_{N_m} = \begin{pmatrix} 0 & \beta_{(N_m+1)N_m}\hat{X}_{N_m} & \dots & \beta_{NN_m}\hat{X}_{N_m} & \hat{X}_{N_m}^2 \\ \beta_{N_m(N_m-1)}\hat{X}_{N_m} & \beta_{(N_m+1)(N_m-1)}\hat{X}_{N_m}^2 & \dots & \beta_{N(N_m-1)}\hat{X}_{N_m}^2 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \beta_{N_m1}\hat{X}_{N_m} & \beta_{(N_m+1)1}\hat{X}_{N_m}^2 & \dots & \beta_{N1}\hat{X}_{N_m}^2 & 0 \\ \hat{X}_{N_m}^2 & 0 & \dots & 0 & 0 \end{pmatrix} \quad (\text{A.45})$$

The L_s are similar to the R_s^T but,

$$\begin{aligned} &\text{with the row:} \quad \begin{pmatrix} 0 & \beta_{(s+1)s}\hat{X}_s & \dots & \beta_{Ns}\hat{X}_s & \hat{X}_s^2 \end{pmatrix} \\ &\text{replaced by:} \quad \begin{pmatrix} 0 & \beta_{s(s-1)}\hat{X}_s & \dots & \beta_{s1}\hat{X}_s & \hat{X}_s^2 \end{pmatrix} \end{aligned}$$

-
- [1] S. L. Braunstein and P. van Loock, Quantum information with continuous variables, *Rev. Mod. Phys.* **77**, 513 (2005).
 - [2] M. Gu, C. Weedbrook, N. C. Menicucci, T. C. Ralph, and P. van Loock, Quantum computing with continuous-variable clusters, *Phys. Rev. A* **79**, 062318 (2009).
 - [3] J. E. Bourassa, R. N. Alexander, M. Vasmer, A. Patil, I. Tzitrin, T. Matsuura, D. Su, B. Q. Baragiola, S. Guha, G. Dauphinais, K. K. Sabapathy, N. C. Menicucci, and I. Dhand, Blueprint for a Scalable Photonic Fault-Tolerant Quantum Computer, *Quantum* **5**, 392 (2021).
 - [4] S. Aaronson and A. Arkhipov, The computational complexity of linear optics, *Theory of Computing* **9**, 143 (2013).
 - [5] C. S. Hamilton, R. Kruse, L. Sansoni, S. Barkhofen, C. Silberhorn, and I. Jex, Gaussian boson sampling, *Phys. Rev. Lett.* **119**, 170501 (2017).
 - [6] A. Deshpande, A. Mehta, T. Vincent, N. Quesada, M. Hinsche, M. Ioannou, L. Madsen, J. Lavoie, H. Qi, J. Eisert, D. Hangleiter, B. Felferman, and I. Dhand, Quantum computational advantage via high-dimensional gaussian boson sampling, *Science Advances* **8**, eabi7894 (2022), <https://www.science.org/doi/pdf/10.1126/sciadv.abi7894>.
 - [7] N. Quesada, J. M. Arrazola, and N. Killoran, Gaussian boson sampling using threshold detectors, *Phys. Rev. A* **98**, 062322 (2018).
 - [8] H.-S. Zhong, L.-C. Peng, Y. Li, Y. Hu, W. Li, J. Qin, D. Wu, W. Zhang, H. Li, L. Zhang, Z. Wang, L. You, X. Jiang, L. Li, N.-L. Liu, J. P. Dowling, C.-Y. Lu, and J.-W. Pan, Experimental gaussian boson sampling, *Science Bulletin* **64**, 511 (2019).
 - [9] J. F. F. Bulmer, B. A. Bell, R. S. Chadwick, A. E. Jones, D. Moise, A. Rigazzi, J. Thorbecke, U.-U. Haus, T. V. Vaerenbergh, R. B. Patel, I. A. Walmsley, and A. Laing, The boundary for quantum advantage in gaussian boson sampling, *Science Advances* **8**, eabl9236 (2022), <https://www.science.org/doi/pdf/10.1126/sciadv.abl9236>.
 - [10] N. Quesada and J. M. Arrazola, Exact simulation of gaussian boson sampling in polynomial space and exponential time, *Phys. Rev. Res.* **2**, 023005 (2020).
 - [11] B. Gupta, J. M. Arrazola, N. Quesada, and T. R. Bromley, Classical benchmarking of gaussian boson sampling on the titan supercomputer, *Quantum Information Processing* **19**, 249 (2020).
 - [12] A. Neville, C. Sparrow, R. Clifford, E. Johnston, P. M. Birchall, A. Montanaro, and A. Laing, Classical boson sampling algorithms with superior performance to near-term experiments, *Nature Physics* **13**, 1153 (2017).
 - [13] P. Clifford and R. Clifford, The classical complexity of boson sampling, in *Proceedings of the 2018 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pp. 146–155, <https://epubs.siam.org/doi/pdf/10.1137/1.9781611975031.10>.
 - [14] P. Clifford and R. Clifford, Faster classical boson sampling (2020), arXiv:2005.04214 [quant-ph].

- [15] U. Chabaud, G. Ferrini, F. Grosshans, and D. Markham, Classical simulation of gaussian quantum circuits with non-gaussian input states, *Phys. Rev. Res.* **3**, 033018 (2021).
- [16] U. Chabaud and M. Walschaers, Resources for bosonic quantum computational advantage, *Phys. Rev. Lett.* **130**, 090602 (2023).
- [17] A. S. Dellios, M. D. Reid, and P. D. Drummond, Simulating gaussian boson sampling quantum computers, *AAPPS Bulletin* **33**, 31 (2023).
- [18] C. Oh, M. Liu, Y. Alexeev, B. Fefferman, and L. Jiang, Classical algorithm for simulating experimental gaussian boson sampling, *Nature Physics* 10.1038/s41567-024-02535-8 (2024).
- [19] N. Spagnolo, D. J. Brod, E. F. Galvão, and F. Sciarrino, Non-linear boson sampling, *npj Quantum Information* **9**, 3 (2023).
- [20] C. S. Hamilton and I. Jex, Boson sampling from non-gaussian states, *Phys. Rev. A* **109**, 052427 (2024).
- [21] G. Bressanini, H. Kwon, and M. S. Kim, Noise thresholds for classical simulability of nonlinear boson sampling, *Phys. Rev. A* **106**, 042413 (2022).
- [22] S. R. White, Density matrix formulation for quantum renormalization groups, *Phys. Rev. Lett.* **69**, 2863 (1992).
- [23] C. Zhang, E. Jeckelmann, and S. R. White, Density matrix approach to local hilbert space reduction, *Phys. Rev. Lett.* **80**, 2661 (1998).
- [24] C. Guo, A. Weichselbaum, J. von Delft, and M. Vojta, Critical and strong-coupling phases in one- and two-bath spin-boson models, *Phys. Rev. Lett.* **108**, 160401 (2012).
- [25] A. Alvermann and H. Fehske, Sparse polynomial space approach to dissipative quantum systems: Application to the sub-ohmic spin-boson model, *Phys. Rev. Lett.* **102**, 150601 (2009).
- [26] D. J. Brod, E. F. Galvão, A. Crespi, R. Osellame, N. Spagnolo, and F. Sciarrino, Photonic implementation of boson sampling: a review, *Advanced Photonics* **1**, 034001 (2019).
- [27] R. Kruse, C. S. Hamilton, L. Sansoni, S. Barkhofen, C. Silberhorn, and I. Jex, Detailed study of gaussian boson sampling, *Phys. Rev. A* **100**, 032326 (2019).
- [28] A. Serafini, *Quantum Continuous Variables: A Primer of Theoretical Methods* (CRC Press, 2017).
- [29] A. P. Lund, A. Laing, S. Rahimi-Keshari, T. Rudolph, J. L. O'Brien, and T. C. Ralph, Boson sampling from a gaussian state, *Phys. Rev. Lett.* **113**, 100502 (2014).
- [30] A. Serafini, *Quantum Continuous Variables: A Primer of Theoretical Methods* (CRC Press, 2017) pp. 87–88.
- [31] E. E. Bergmann and A. Holz, Exact solutions of an n-dimensional anisotropic oscillator in a uniform magnetic field., *Nuovo Cim. 7B: No. 2*, 265-76(11 Feb 1972). 10.1007/BF02743599 (1972).
- [32] J. Biamonte, Lectures on quantum tensor networks (2020), arXiv:1912.10049 [quant-ph].
- [33] J. C. Bridgeman and C. T. Chubb, Hand-waving and interpretive dance: an introductory course on tensor networks, *Journal of Physics A: Mathematical and Theoretical* **50**, 223001 (2017).
- [34] M. C. Bañuls, Tensor network algorithms: A route map, *Annual Review of Condensed Matter Physics* **14**, 173 (2023).
- [35] N. Schuch, M. M. Wolf, F. Verstraete, and J. I. Cirac, Entropy scaling and simulability by matrix product states, *Phys. Rev. Lett.* **100**, 030504 (2008).
- [36] F. Verstraete and J. I. Cirac, Matrix product states represent ground states faithfully, *Phys. Rev. B* **73**, 094423 (2006).
- [37] U. Schollwöck, The density-matrix renormalization group in the age of matrix product states, *Annals of Physics* **326**, 96 (2011), january 2011 Special Issue.
- [38] S. Rychkov and L. G. Vitale, Hamiltonian truncation study of the φ^4 theory in two dimensions, *Phys. Rev. D* **91**, 085011 (2015).
- [39] J. Wurtz, P. W. Claeys, and A. Polkovnikov, Variational schrieffer-wolff transformations for quantum many-body dynamics, *Phys. Rev. B* **101**, 014302 (2020).
- [40] G. Vidal, Efficient classical simulation of slightly entangled quantum computations, *Phys. Rev. Lett.* **91**, 147902 (2003).
- [41] Supplemental Material [url], which includes Refs. [76] [77] [78] [79] [37] [80] [40] [18] [42] [43] [44] [36] [81] [82] [49] [24].
- [42] I. Oseledets and E. Tyrtysnikov, Tt-cross approximation for multidimensional arrays, *Linear Algebra and its Applications* **432**, 70 (2010).
- [43] I. V. Oseledets, Tensor-train decomposition, *SIAM Journal on Scientific Computing* **33**, 2295 (2011), <https://doi.org/10.1137/090752286>.
- [44] L. De Lathauwer, B. De Moor, and J. Vandewalle, A multilinear singular value decomposition, *SIAM Journal on Matrix Analysis and Applications* **21**, 1253 (2000), <https://doi.org/10.1137/S0895479896305696>.
- [45] I. P. McCulloch, From density-matrix renormalization group to matrix product states, *Journal of Statistical Mechanics: Theory and Experiment* **2007**, P10014 (2007).
- [46] N. Quesada, R. S. Chadwick, B. A. Bell, J. M. Arrazola, T. Vincent, H. Qi, and R. García-Patrón, Quadratic speed-up for simulating gaussian boson sampling, *PRX Quantum* **3**, 010306 (2022).
- [47] M. Oszmaniec and D. J. Brod, Classical simulation of photonic linear optics with lost particles, *New Journal of Physics* **20**, 092002 (2018).
- [48] M. Liu, C. Oh, J. Liu, L. Jiang, and Y. Alexeev, Simulating lossy gaussian boson sampling with matrix-product operators, *Phys. Rev. A* **108**, 052604 (2023).
- [49] A. Weiße, H. Fehske, G. Wellein, and A. R. Bishop, Optimized phonon approach for the diagonalization of electron-phonon problems, *Phys. Rev. B* **62**, R747 (2000).
- [50] C. Brockett, F. Dorfner, L. Vidmar, F. Heidrich-Meisner, and E. Jeckelmann, Matrix-product-state method with a dynamical local basis optimization for bosonic systems out of equilibrium, *Phys. Rev. B* **92**, 241106 (2015).
- [51] J. Williamson, On the algebraic problem concerning the normal forms of linear dynamical systems, *American Journal of Mathematics* **58**, 141 (1936).
- [52] M. Houde, W. McCutcheon, and N. Quesada, Matrix decompositions in quantum optics: Takagi/autonne, bloch-messiah/euler, iwasaawa, and williamson (2024), arXiv:2403.04596 [quant-ph].
- [53] N. Killoran, J. Izaac, N. Quesada, V. Bergholm, M. Amy, and C. Weedbrook, Strawberry Fields: A Software Platform for Photonic Quantum Computing, *Quantum* **3**, 129 (2019).
- [54] T. R. Bromley, J. M. Arrazola, S. Jahangiri, J. Izaac, N. Quesada, A. D. Gran, M. Schuld, J. Swinerton, Z. Zabaneh, and N. Killoran, Applications of near-term photonic quantum computers: software and algorithms,

- Quantum Science and Technology **5**, 034010 (2020).
- [55] J. Gray, quimb: a python library for quantum information and many-body calculations, Journal of Open Source Software **3**, 819 (2018).
 - [56] B. Gupt, J. Izaac, and N. Quesada, The walrus: a library for the calculation of hafnians, hermite polynomials and gaussian boson sampling, Journal of Open Source Software **4**, 1705 (2019).
 - [57] L. S. Madsen, F. Laudenbach, M. F. Askarani, F. Rortais, T. Vincent, J. F. F. Bulmer, F. M. Miatto, L. Neuhaus, L. G. Helt, M. J. Collins, A. E. Lita, T. Gerrits, S. W. Nam, V. D. Vaidya, M. Menotti, I. Dhand, Z. Vernon, N. Quesada, and J. Lavoie, Quantum computational advantage with a programmable photonic processor, Nature **606**, 75 (2022).
 - [58] G. Vallone, G. Cariolaro, and G. Pierobon, Means and covariances of photon numbers in multimode gaussian states, Phys. Rev. A **99**, 023817 (2019).
 - [59] S. Paeckel, T. Köhler, A. Swoboda, S. Manmana, U. Schollwöck, and C. Hubig, Time-evolution methods for matrix-product states, Annals of Physics **411**, 167998 (2019).
 - [60] G. M. Crosswhite and D. Bacon, Finite automata for caching in matrix product algorithms, Phys. Rev. A **78**, 012356 (2008).
 - [61] S. Paeckel, T. Köhler, and S. R. Manmana, Automated construction of $U(1)$ -invariant matrix-product operators from graph representations, SciPost Phys. **3**, 035 (2017).
 - [62] B. N. Khoromskij, *Tensor Numerical Methods in Scientific Computing* (De Gruyter, Berlin, Boston, 2018) pp. 143–145.
 - [63] While precise studies on this topic has eluded our attention, an accessible prototypical example is given in [83], where the optimization algorithm produces a better compression than the single-shot SVD-based compression.
 - [64] T. Xiang, Implementation of symmetries, in *Density Matrix and Tensor Network Renormalization* (Cambridge University Press, 2023) p. 123–136.
 - [65] M. Rakhuba, A. Novikov, and I. Oseledets, Low-rank riemannian eigensolver for high-dimensional hamiltonians, Journal of Computational Physics **396**, 718 (2019).
 - [66] D. P. Kingma and J. Ba, Adam: A method for stochastic optimization (2017), arXiv:1412.6980 [cs.LG].
 - [67] D. C. Liu and J. Nocedal, On the limited memory bfgs method for large scale optimization, Mathematical Programming **45**, 503 (1989).
 - [68] TensorNetwork.org: Density Matrix Renormalization Group Algorithm (DMRG), author: Miles Stoudenmire, https://tensornetwork.org/mps/algorithms/dmrg/#toc_4 (2021), Accessed: 2024-10-24.
 - [69] C. Hubig, I. P. McCulloch, U. Schollwöck, and F. A. Wolf, Strictly single-site dmrg algorithm with subspace expansion, Phys. Rev. B **91**, 155115 (2015).
 - [70] S. V. Dolgov and D. V. Savostyanov, Alternating minimal energy methods for linear systems in higher dimensions, SIAM Journal on Scientific Computing **36**, A2248 (2014), <https://doi.org/10.1137/140953289>.
 - [71] S. V. Dolgov and D. V. Savostyanov, Corrected one-site density matrix renormalization group and alternating minimal energy algorithm, in *Numerical Mathematics and Advanced Applications - ENUMATH 2013*, edited by A. Abdulle, S. Deparis, D. Kressner, F. Nobile, and M. Picasso (Springer International Publishing, Cham, 2015) pp. 335–343.
 - [72] D. Kennes and C. Karrasch, Extending the range of real time density matrix renormalization group simulations, Computer Physics Communications **200**, 37 (2016).
 - [73] T. Begušić, J. Gray, and G. K.-L. Chan, Fast and converged classical simulations of evidence for the utility of quantum computing before fault tolerance, Science Advances **10**, eadk4321 (2024), <https://www.science.org/doi/pdf/10.1126/sciadv.adk4321>.
 - [74] J. Tindall, M. Fishman, E. M. Stoudenmire, and D. Sels, Efficient tensor network simulation of ibm’s eagle kicked ising experiment, PRX Quantum **5**, 010308 (2024).
 - [75] Github repository: <https://github.com/Vinther901/Variational-Tensor-Networks-Boson-Sampling>.
 - [76] R. A. Horn and C. R. Johnson, *Matrix Analysis*, p. 183 (Cambridge University Press, 1985).
 - [77] S.-G. Hwang, Cauchy’s interlace theorem for eigenvalues of hermitian matrices, The American Mathematical Monthly **111**, 157 (2004).
 - [78] E. D. Filho and M. A. C. Ribeiro, Generalized ladder operators for shape-invariant potentials, Physica Scripta **64**, 548 (2001).
 - [79] J. Haegeman, M. Mariën, T. J. Osborne, and F. Verstraete, Geometry of matrix product states: Metric, parallel transport, and curvature, Journal of Mathematical Physics **55**, 021902 (2014), https://pubs.aip.org/aip/jmp/article-pdf/doi/10.1063/1.4862851/15641654/021902_1.online.pdf.
 - [80] R.-Y. Sun, T. Shirakawa, and S. Yunoki, Improved real-space parallelizable matrix-product state compression and its application to unitary quantum dynamics simulation, Phys. Rev. B **110**, 085149 (2024).
 - [81] J. I. Cirac, D. Pérez-García, N. Schuch, and F. Verstraete, Matrix product states and projected entangled pair states: Concepts, symmetries, theorems, Rev. Mod. Phys. **93**, 045003 (2021).
 - [82] Y. Zhou, E. M. Stoudenmire, and X. Waintal, What limits the simulation of quantum computers?, Phys. Rev. X **10**, 041038 (2020).
 - [83] t3f python library: Riemannian optimization, author: Alexander Novikov, <https://t3f.readthedocs.io/en/latest/tutorials/riemannian.html> (2017), Accessed: 2024-10-24.