The State Preparation of Multivariate Normal Distributions using Tree Tensor Network

Hidetaka Manabe[1] and Yuichi Sano[2]

[1]Graduate School of Engineering Science, Osaka University
[2]Department of Nuclear Engineering, Kyoto University

The quantum state preparation of probability distributions is an important subroutine for many quantum algorithms. When embedding D-dimensional multivariate probability distributions by discretizing each dimension into 2^n points, we need a state preparation circuit comprising a total of nD qubits, which is often difficult to compile. In this study, we propose a scalable method to generate state preparation circuits for D-dimensional multivariate normal distributions, utilizing tree tensor networks (TTN). We establish theoretical guarantees that multivariate normal distributions with 1D correlation structures can be efficiently represented using TTN. Based on these analyses, we propose a compilation method that uses automatic structural optimization to find the most efficient network structure and compact circuit. We apply our method to state preparation circuits for various high-dimensional random multivariate normal distributions. The numerical results suggest that our method can dramatically reduce the circuit depth and CNOT count while maintaining fidelity compared to existing approaches.

1 Introduction

Amplitude embedding of classical data is a key subroutine in various quantum algorithms such as quantum-linear-solver algorithms [17], Monte-Carlo methods [34], and quantum machine learning [3, 32]. In quantum state preparation, we encode 2^n classical data $f(x)$ into an n-qubit quantum state $|f(x)\rangle = U(x)|0\rangle$. Generally, achieving this requires $O(2^n)$ quantum operations [37, 42], which can become a critical bottleneck in the execution of algorithms. Thus, proposing efficient state preparation methods is essential for the practical implementation of quantum computer algorithms.

There are many studies on methods of quantum state preparation [1, 8, 27, 29, 30, 35, 38, 44, 45, 61, 62]. For the Fault-Tolerant Quantum Computer (FTQC), it is possible to utilize Quantum Signal Processing-based [30] or oracle-based algorithms, where error bounds and the scaling of the gate counts can be rigorously proven. However, in the case of NISQ (Noisy Intermediate-Scale Quantum devices) and early-FTQC, there are strict limitations on executable circuit depth, connectivity, and the number of qubits. Therefore, it becomes essential to use classical optimizations to drastically reduce the circuit costs.

One of the efficient compilation methods for state preparation circuits is the Fourier series loader (FSL) method [35]. Suppose we want to encode a D-dimensional function $f(\boldsymbol{x})$, discretizing each dimension into 2^n points. Implementing this would require a state preparation circuit with a nD-qubit circuits and a depth of $O(2^{nD})$. The FSL method utilizes the exponential decay of the Fourier coefficients of smooth functions, allowing for

state preparation in a depth of $O(2^{mD})$ with $m < n$ using Fourier series approximation. Nevertheless, as D increases, the circuit depth increases exponentially, making this method unsuitable for high-dimensional functions.

In this study, we propose a new method to compile the state preparation circuit for multivariate functions, particularly for multivariate normal distributions, which are widely used for many practical quantum algorithms in various settings [9, 10, 24, 25, 28, 36, 53, 58]. We utilize a tensor network-based method [2, 11, 12, 15, 19–21, 23, 26, 31, 43, 48, 49, 60], especially the Tree Tensor Network (TTN) ansatz [52], which is often used to represent quantum states with a 1D structure. We prove that a multivariate normal distribution can be efficiently represented as a TTN when there is a 1D structure in the correlations between variables. We demonstrate that by combining the FSL method and Tensor Cross Interpolation (TCI) [41, 50], it is possible to scalably prepare quantum states for multivariate normal distributions with 1D structures, independent of the dimension D and discretization parameter n.

Additionally, we propose a new approach that combines circuit compilation with automatic structural optimization [18]. Automatic structural optimization is a method to optimize network structure in TTN to represent the state more efficiently by reconnecting the legs of the local tensor throughout the process. This allows for the appropriate detection of structures in multivariate normal distributions and finding more compact circuits, even when the correlations between variables are unknown beforehand or when there is no 1D structure.

We have demonstrated our method to compile state preparation circuits for several high-dimensional multivariate normal distributions. Our method can scalably generate state preparation circuits for multivariate normal distributions with 1D structures. We also numerically calculate the relationship between the fidelity of the circuit generated and the bond dimension, confirming the theoretical results. Furthermore, we investigated the impact of the automatic structural optimization algorithms on the performance of circuit compilation. Our numerical results show that the algorithm appropriately detects the 1D structure of a multivariate normal distribution and reconfigures the network with high success probability, even for large D. Finally, we show that we can reduce the circuit depth and CNOT count by factors of 10 to 1000 compared to existing methods for random multivariate normal distributions.

This paper is organized as follows. We describe tensor network techniques used in this study in Sec. 2. We explain the proposed state preparation method in Sec. 3. In Sec. 4, we present the numerical results of our methods and discuss their performance. We conclude the paper by summarizing our contribution in Sec. 5.

2 Tensor network methods

In this section, we briefly explain the tensor network methods used in this study. For a more detailed discussion on tensor networks, please refer to the reference [7].

2.1 Matrix product states and tree tensor networks

Tensor networks are a framework for representing a large tensor in the form of a network of smaller tensors. Specifically in quantum information theory, they are used to represent the wave function of quantum many-body systems. Matrix Product States (MPS), one

type of tensor network ansatz, can be described as follows:

$$|\psi_{\mathrm{mps}}\rangle = \sum_{\{s_i\}} \sum_{\{\alpha_i\}} \left[A^{[1]s_1}_{\alpha_1} A^{[2]s_2}_{\alpha_1\alpha_2} \cdots A^{[n]s_n}_{\alpha_{n-1}} \right] |s_1 s_2 \cdots s_n\rangle \,, \tag{1}$$

where $\{s_i\}$ and $\{\alpha_i\}$ are referred to as physical and virtual indices, respectively. The dimension of the virtual indices is referred to as *bond dimension*.

The ability to represent a multi-dimensional array as an MPS can be evaluated using its singular values. Suppose $\psi(s_1, \ldots, s_i; s_{i+1}, \ldots, s_n)$ be the unfolding matrix between site $(1, \ldots, i)$ and $(i+1, \ldots, n)$. Consider the singular value decomposition (SVD) of the matrix in the form

$$\psi(s_1, \ldots, s_i; s_{i+1}, \ldots, s_n) = U_i \Sigma_i V_i \tag{2}$$

and its singular values $\{\sigma_j^{(i)}\}$ sorted in descending order. Note that the summation of the square of the singular values is 1 for a wave function $|\psi\rangle$. Then, there exists an MPS approximation $\tilde{\psi}$ with bond dimension r whose truncation error, in the sense of the Frobenius norm, is upper bounded by its singular values [39, 41]:

$$\left\| \psi(\boldsymbol{s}) - \tilde{\psi}(\boldsymbol{s}) \right\|_F \leq \sqrt{\sum_{i=1}^{n-1} \epsilon_i^2}, \tag{3}$$

$$\epsilon_i^2 = 1 - \sum_{j=0}^{r-1} {\sigma_j^{(i)}}^2. \tag{4}$$

In other words, a multi-dimensional array is well-represented by an MPS if it has a small number of dominant singular values for each unfolding matrix.

The discussion above can be generalized to Tree Tensor Networks (TTN) [52]. A TTN is a type of tensor network ansatz that does not have closed loops in the diagrammatic representation. The inner edge divides the sites into two disjoint sets, which we call the *bipartition* of the system. Similarly to the case with MPS, the error upper bound in the best TTN approximation, in terms of the Frobenius norm, depends on the distribution of the singular values of the unfolding matrix across each bipartition induced by the edges in the network.

2.2 Functional matrix product states and tree tensor network

For the discussion later, we introduce the functional version of MPS, known as functional MPS (functional Tensor-Train decomposition) [4, 13, 40, 47]. For $\psi \in L^2(\mathbb{R}^n)$, we consider the following decomposition:

$$\psi(x_1, \ldots, x_n) = \sum_{\{\alpha_i\}} \gamma_1(\alpha_0, x_1, \alpha_1)\gamma_2(\alpha_1, x_2, \alpha_2) \cdots \gamma_n(\alpha_{n-1}, x_n, \alpha_n). \tag{5}$$

To consider the error introduced by this approximation, we use an analogue of the singular value decomposition: the Schmidt decomposition. The Schmidt decomposition of the function ψ has the form

$$\psi(x_1, \ldots, x_i, x_{i+1}, \ldots, x_n) = \sum_k \sqrt{\lambda_k^{(i)}} \psi_k^{(1)}(x_1, \ldots, x_i) \psi_k^{(2)}(x_{i+1}, \ldots, x_n), \tag{6}$$

where $\psi_k^{(1)}$ and $\psi_k^{(2)}$ form orthogonal basis for each subsystem, and $\{\sqrt{\lambda_k^{(i)}}\}$ are called *Schmidt coefficients* and sorted in descending order. Then, as in the discrete version, there

exists a functional MPS approximation $\tilde{\psi}$ with bond dimension r whose truncation error, in the sense of the L^2 norm, is upper bounded by its Schmidt coefficients:

$$\|\psi - \tilde{\psi}\|_{L^2(\mathbb{R}^n)} \leq \sqrt{\sum_{i=1}^{n-1} \left(\sum_{k=r}^{\infty} \lambda_k^{(i)} \right)}. \tag{7}$$

These formulations for the functional MPS can similarly be extended to the functional TTN.

Consider tensorizing a function evaluated at points on a grid. When discretized on a sufficiently dense and large grid, the Schmidt decomposition (5) can be interpreted as the singular value decomposition (2), and the truncation error of the functional MPS (6) as that of discrete MPS (3). Therefore, if a function can be accurately approximated by a functional MPS, we can expect that its discretized multi-index tensor can similarly be approximated by a discrete MPS with the same bond dimension. In the following discussion, we consider functional MPS and functional TTN when evaluating approximation error.

2.3 Tensor cross interpolation

We can construct MPS or TTN from the original multi-index array by sequentially applying SVD. However, this method requires loading the unfolding matrix of the original tensor into memory, which becomes impractical when the tensor to be approximated itself is very large.

Instead, we can use the tensor cross interpolation (TCI) algorithm [41], a method for directly constructing MPS and TTN from the elements of the multi-dimensional tensor. Let $s_i \in \mathbb{K}_i = \{1, \ldots, d\}$. We define the following multi-indices:

$$s_{\leq i} = s_1 s_2 \cdots s_i \in \mathbb{K}_1 \times \mathbb{K}_2 \times \cdots \times \mathbb{K}_i, \quad s_{>i} = s_{i+1} \cdots s_n \in \mathbb{K}_{i+1} \times \cdots \times \mathbb{K}_n. \tag{8}$$

We interpolate the multi-index tensor as:

$$\psi(s_1, \ldots, s_n) \simeq \psi(s_1, \mathcal{J}_{>1})[\psi(\mathcal{J}_{\leq 1}, \mathcal{J}_{>1})]^{-1} \psi(\mathcal{J}_{\leq 1}, s_2, \mathcal{J}_{>2})[\psi(\mathcal{J}_{\leq 2}, \mathcal{J}_{>2})]^{-1} \tag{9}$$

$$\cdots \psi(\mathcal{J}_{\leq n-1}, s_n), \tag{10}$$

where $\mathcal{J}_{\leq i}$ and $\mathcal{J}_{>i}$ are sets of multi-indices satisfying the following condition:

$$\mathcal{J}_{\leq i+1} \subset \mathcal{J}_{\leq i} \times \mathbb{K}_{i+1}, \quad \mathcal{J}_{>i} \subset \mathbb{K}_{i+1} \times \mathcal{J}_{>i+1}. \tag{11}$$

If the size of each $\mathcal{J}_{\leq i}, \mathcal{J}_{>i}$ is limited by χ, the original tensor is interpolated as an MPS with bond dimension χ. The choice of multi-indices defines the accuracy of the approximation. In this study, we use the ALS maxvol algorithm to find indices to use. For the details of the algorithm, please refer to the reference [41].

2.4 Canonical form of tree tensor network

For tree tensor networks, an efficient truncation method using the *canonical form* is well-known [51]. The diagram notation of the canonical form is illustrated in Fig. 1(a). We impose the isometric constraint on each tensor, except for the order-2 tensor known as the *top tensor*. We can efficiently transform any TTN into canonical form by repeatedly applying SVD to each tensor. When sweeping the position of the top tensor in canonical form, we can truncate small singular values and reduce bond dimension, which achieves a near-optimal approximation in terms of the Frobenius norm.

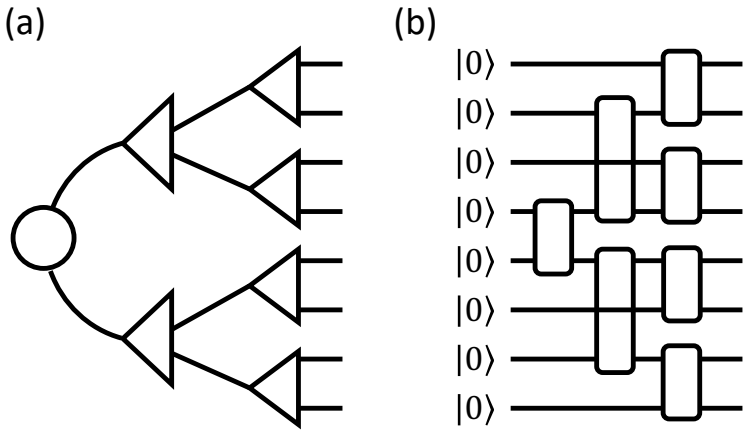

Figure 1: (a) The canonical form of the TTN. The top tensor is illustrated as the circle node and isometries as the triangle nodes. (b) The corresponding quantum circuit.

After obtaining the canonical form for the 1D tensor network states, we can easily construct the state preparation circuit as illustrated in Fig. 1(b). The quantum circuit representation of the top tensor is achieved through the brute-force state preparation method. The other tensors are isometries and can be converted into circuits using unitary synthesis [22]. Even if the bond dimension is not a power of two, it can be represented as a quantum circuit by suitably padding tensors to become a power of two.

3 Methods

Our goal is to compile a quantum circuit to generate a quantized D-dimensional multivariate normal distribution by discretizing each dimension into 2^n points. In this study, we consider the following probability density function with mean zero for simplicity:

$$p(\boldsymbol{x}) = \frac{1}{\sqrt{(2\pi)^D \det(\Sigma)}} \exp\left(-\frac{1}{2}\boldsymbol{x}^\top \Sigma^{-1} \boldsymbol{x}\right), \tag{12}$$

where $\boldsymbol{x}$ is D-dimensional vector and Σ is the covariance matrix. Then we specify the parameter a such that $[-a/2, a/2]^D$ contains effectively all of the probability density. We use the binary encoding to discretize each dimension, such that

$$x_i = -\frac{a}{2} + \frac{a}{2^n} b_i, \quad 0 \leq b_i \leq 2^n - 1, \tag{13}$$

then our target state is expressed as follows:

$$\sum_{\boldsymbol{b}} f_{\boldsymbol{b}} |\boldsymbol{b}\rangle \equiv \sum_{\boldsymbol{b}} \left(\frac{a}{2^n}\right)^{D/2} \sqrt{p(\boldsymbol{x})} |\boldsymbol{b}\rangle . \tag{14}$$

3.1 Fourier series approximation

To scale for n, we adopt the technique of using discrete Fourier transform as proposed in [35]. The discrete Fourier transform over $\mathbb{Z}_{2^n}^D$ is expressed as follows:

$$\hat{f}_{\boldsymbol{k}} = \frac{1}{\sqrt{2^{nD}}} \sum_{b_1=0}^{2^n-1} \cdots \sum_{b_D=0}^{2^n-1} f_{\boldsymbol{b}} \exp\left(-2\pi i \frac{b_1 k_1 + \cdots + b_D k_D}{2^n}\right). \tag{15}$$

Then we approximate $f_{\boldsymbol{b}}$ by truncating Fourier coefficients as follows:

$$\tilde{f}_{\boldsymbol{b}} \simeq \frac{1}{\sqrt{2^{nD}}} \sum_{k_1=-M/2}^{M/2-1} \cdots \sum_{k_D=-M/2}^{M/2-1} \hat{f}_{\boldsymbol{k}} \exp\left(2\pi i \frac{b_1 k_1 + \cdots + b_D k_D}{2^n}\right). \tag{16}$$

Here, $M := 2^m$ represents the number of Fourier coefficients to be kept for each dimension. It is shown that, if $f(\cdot)$ is a sufficiently smooth function, the required number of m is smaller than n and, moreover, does not depend on n [35]. In practice, $\hat{f}_{\boldsymbol{k}}$ is calculated as follows:

$$\hat{f}_{\boldsymbol{k}} \propto \sum_{\boldsymbol{b}} f_{\boldsymbol{b}} \exp\left(-2\pi i \frac{x_1 k_1 + \cdots + x_D k_D}{a}\right) \exp\left(-2\pi i \frac{k_1 + \cdots + k_D}{2}\right) \tag{17}$$

$$\propto (-1)^{|k_1+\cdots+k_D|} \int_{-\frac{a}{2}}^{\frac{a}{2}} \cdots \int_{-\frac{a}{2}}^{\frac{a}{2}} \sqrt{p(\boldsymbol{x})} \exp\left(-2\pi i \frac{\boldsymbol{x}\boldsymbol{k}}{a}\right) d\boldsymbol{x} \tag{18}$$

for a sufficiently fine discretization. Moreover, if a is taken to be sufficiently large, the following approximation also holds:

$$\hat{f}_{\boldsymbol{k}} \propto (-1)^{|k_1+\cdots+k_D|} \exp\left(-(2\pi/l)^2 \boldsymbol{k}^T \Sigma \boldsymbol{k}\right) \tag{19}$$

with an appropriate normalization term. Equation (19) shows that the value of $\hat{f}_{\boldsymbol{k}}$ decreases exponentially as the wavenumber $\boldsymbol{k}$ increases, indicating that retaining only a very small number of Fourier coefficients is sufficient to represent $f_{\boldsymbol{b}}$.

We first prepare the state that encodes the Fourier coefficients:

$$\sum_{k_1=-M/2}^{M/2-1} \cdots \sum_{k_D=-M/2}^{M/2-1} \hat{f}_{\boldsymbol{k}} |\boldsymbol{k}\rangle, \tag{20}$$

then we apply Inverse Quantum Fourier Transformation (QFT) to obtain the target state $\sum_{\boldsymbol{b}} \tilde{f}_{\boldsymbol{b}} |b\rangle$. The circuit cost for preparing (20) is much smaller than that for $\sum_{\boldsymbol{b}} f_{\boldsymbol{b}} |b\rangle$, and Inverse QFT can be implemented efficiently in quantum computers with circuit depth n. Therefore, we can generate the desired state with a small cost that does not scale exponentially with the discretization size n.

3.2 Representing Fourier coefficients using tree tensor network

We now want to encode the truncated Fourier coefficients $\hat{f}_{\boldsymbol{k}}$ into the quantum circuit. In the case of D-dimensional probability distributions, however, the number of truncated Fourier coefficients is 2^{mD}, which can be a very large number.

Therefore, we consider representing the Fourier coefficients with an MPS:

$$\sum_{\boldsymbol{k}} \hat{f}_{\boldsymbol{k}} |\boldsymbol{k}\rangle = \sum_{\{k_i\}} \sum_{\{\alpha_i\}} \left[A^{[1]k_1}_{\alpha_1} A^{[2]k_2}_{\alpha_1\alpha_2} \cdots A^{[D]k_D}_{\alpha_{D-1}}\right] |k_1 k_2 \cdots k_D\rangle, \tag{21}$$

or a TTN. The memory footprint required grows only polynomially with D.

3.2.1 Schmidt coefficients of multivariate normal distribution

As discussed above, the accuracy achievable in the MPS or TTN representation depends on its singular values or Schmidt coefficients. The Schmidt coefficients for a multivariate normal distribution are analytically known [6].

First, we discuss the Schmidt decomposition of the two-dimensional normal distribution. The Schmidt decomposition of the wave function $\psi(x_1, x_2)$ has the form

$$\psi(x_1, x_2) = \sum_{k=0}^{\infty} \sqrt{\lambda_k} \psi_k^{(1)}(x_1) \psi_k^{(2)}(x_2). \tag{22}$$

Suppose the quantum state $\psi(x_1, x_2)$ is given by the square root of the density function of the normal distribution with correlation coefficient ρ:

$$\psi(x_1, x_2) = \sqrt{p(x_1, x_2)}, \tag{23}$$

$$p(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)}\left(\frac{(x_1-m_1)^2}{\sigma_1^2} - 2\rho\frac{(x_1-m_1)(x_2-m_2)}{\sigma_1\sigma_2} + \frac{(x_2-m_2)^2}{\sigma_2^2}\right)\right), \tag{24}$$

where m_1, m_2 are the mean values and σ_1^2, σ_2^2 are the variance of each variable. The Schmidt decomposition of this wave function [5] is given by:

$$\psi_k^{(1)}(x_1) = C_k^{(1)} H_k\left(\frac{(x_1-m_1)}{\sigma_1}\sqrt{\frac{K}{2}}\right) \exp\left(-K\frac{(x_1-m_1)^2}{4\sigma_1^2}\right), \tag{25}$$

$$\psi_k^{(2)}(x_2) = C_k^{(2)} H_k\left(\frac{(x_2-m_2)}{\sigma_2}\sqrt{\frac{K}{2}}\right) \exp\left(-K\frac{(x_2-m_2)^2}{4\sigma_2^2}\right), \tag{26}$$

where $C_k^{(1)}, C_k^{(2)}$ are the normalization factors and K is calculated by

$$K = \frac{1}{\sqrt{1-\rho^2}}. \tag{27}$$

The squares of Schmidt coefficients $\{\lambda_k\}$ form a geometric series with an initial value

$$\lambda_0 = \frac{2}{K+1} \tag{28}$$

and common ratio:

$$q = \frac{K-1}{K+1}. \tag{29}$$

An approximation error of an MPS of bond dimension r is calculated as

$$\|\psi - \psi_{\text{mps}}\|^2_{L^2(\mathbb{R}^2)} \leq \sum_{k=r}^{\infty} \lambda_k \tag{30}$$

$$= \lambda_0 \sum_{k=r}^{\infty} q^k \tag{31}$$

$$= \lambda_0 \frac{q^r}{1-q} = q^r. \tag{32}$$

Then, if we want to suppress the truncation error to ϵ, we need the bond dimension r that satisfies:

$$q^r \leq \epsilon^2 \tag{33}$$

$$\iff r \geq \frac{2}{\log\frac{1}{q}} \log\frac{1}{\epsilon}. \tag{34}$$

In other words, the required bond dimension scales as $O(\log(1/\epsilon))$, which is an efficient approximation.

The distribution of the Schmidt coefficients of the multivariate normal distribution is also known and calculated only from the covariance matrix [6]. For simplicity, we consider the multivariate normal distribution with zero mean. The Schmidt decomposition of the multivariate function for subsystems $x_1, \ldots, x_p$ and $x_{p+1}, \ldots, x_D$, assuming that $2p \leq D$, is defined as follows:

$$\psi(x_1, \ldots, x_p, x_{p+1}, \ldots, x_D) = \sum_k \sqrt{\lambda_k} \psi_k^{(1)}(x_1, \ldots, x_p) \psi_k^{(2)}(x_{p+1}, \ldots, x_D). \tag{35}$$

According to the canonical-correlation analysis [46], linear transformations can be found for variables $\xi_1, \ldots \xi_p, \xi_{p+1}, \ldots, \xi_D$:

$$\xi_i = \sum_{j=1}^{p} l_{ij} x_j, \qquad 1 \leq i \leq p \tag{36}$$

$$\xi_i = \sum_{j=p+1}^{D} m_{ij} x_j, \quad p+1 \leq i \leq D \tag{37}$$

that satisfy the following properties: (1) All ξ have unit variance and zero mean and (2) the correlation between any ξ is zero except for the pairs $(\xi_1, \xi_{p+1}), (\xi_2, \xi_{p+2}), \ldots, (\xi_p, \xi_{2p})$. The coefficients $\boldsymbol{l}, \boldsymbol{m}$ and the canonical correlation coefficients $\{\rho_i\}$ between these pairs are the solutions to the following generalized eigenvalue problem:

$$\begin{pmatrix} O & \Sigma_{12} \\ \Sigma_{21} & O \end{pmatrix} \begin{pmatrix} \boldsymbol{l} \\ \boldsymbol{m} \end{pmatrix} = \rho \begin{pmatrix} \Sigma_{11} & O \\ O & \Sigma_{22} \end{pmatrix} \begin{pmatrix} \boldsymbol{l} \\ \boldsymbol{m} \end{pmatrix}, \tag{38}$$

where $\Sigma_{11}, \Sigma_{12}, \Sigma_{21}, \Sigma_{22}$ are the submatrices of the covariance matrix:

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12} & \Sigma_{22} \end{bmatrix}. \tag{39}$$

$D - 2p$ solutions are zero and the others arise in pairs as $\pm\rho_1, \pm\rho_2, \ldots, \pm\rho_p$. We choose the positive ones. Then, the target function is decomposed as the tensor product of the probability density function of the two-dimensional normal distribution $(\xi_1, \xi_{p+1}), \ldots, (\xi_p, \xi_{2p})$ with correlation $\rho_1, \ldots, \rho_p$, and one-dimensional normal distribution $\xi_{2p+1}, \ldots, \xi_D$:

$$\psi(x_1, \ldots, x_p, x_{p+1}, \ldots, x_D) = \prod_{i=1}^{p} \left(\sum_{k_i} \sqrt{\lambda_{k_i}^{(i)}} \psi_{k_i}^{(1)}(\xi_i) \psi_{k_i}^{(2)}(\xi_{p+i}) \right) \cdot \sqrt{p(\xi_{2p+1})} \cdots \sqrt{p(\xi_D)}, \tag{40}$$

where $\{\lambda^{(i)}\}$ is the Schmidt coefficients for the pair (ξ_i, ξ_{p+i}). The Schmidt decomposition of the multivariate normal distribution is given by defining mode $\boldsymbol{k} = (k_1, \ldots, k_p)$ and the corresponding orthogonal basis as

$$\psi_{\boldsymbol{k}}^{(1)}(x_1, \ldots, x_p) = \prod_{i=1}^{p} \psi_{k_i}^{(1)}(\xi_i), \tag{41}$$

$$\psi_{\boldsymbol{k}}^{(2)}(x_{p+1}, \ldots, x_D) = \left(\prod_{i=1}^{p} \psi_{k_i}^{(2)}(\xi_{p+i}) \right) \cdot \sqrt{p(\xi_{2p+1})} \cdots \sqrt{p(\xi_D)}, \tag{42}$$

and the Schmidt coefficients are the product of each Schmidt coefficient of the pairs:

$$\sqrt{\lambda_{\boldsymbol{k}}} = \prod_{i=1}^{p} \sqrt{\lambda_{k_i}^{(i)}}. \tag{43}$$

Each $\{\lambda^{(i)}\}$ can be calculated from ρ_i and equation (27)-(29). Therefore, by solving the generalized eigenvalue problem (38), we can enumerate all the Schmidt coefficients of the multivariate normal distribution between any bipartition of the system and calculate the required bond dimensions for the tensor network approximation. The computational cost of calculating the above values only depends polynomially on the dimension D, not on the discretization size n.

3.2.2 Schmidt coefficients of a multivariate normal distribution with 1D covariance structure

The discussion above can be applied to multivariate normal distributions with any covariance matrix and bipartition. However, if we restrict to the 1D covariance structure, we can guarantee the existence of efficient MPS or TTN representations of the target state. Note that a priori rank bounds for approximations in the MPS representation for the multivariate normal distribution with weak correlations are studied in [47], This work differs in that we directly use the information of the Schmidt coefficients for proof.

Theorem 3.1. *Let $p(\boldsymbol{x})$ be a probability density function of a D-dimensional multivariate normal distribution with zero mean and covariance matrix Σ. Assume that, for any bipartition of variables induced by an inner edge of the tree tensor network,* $\operatorname{rank}\Sigma_{12} \leq l$ *where l is constant and the canonical correlations are less than 1. For every ϵ there exists a TTN approximation $\tilde{\psi}$ of $\psi(\boldsymbol{x}) = \sqrt{p(\boldsymbol{x})}$ with*

$$\|\psi(\boldsymbol{x}) - \tilde{\psi}(\boldsymbol{x})\|_{L^2(\mathbb{R}^D)} \leq \epsilon \tag{44}$$

whose bond dimension χ is upper bounded by

$$\chi \leq \left(C \cdot \log \frac{l\sqrt{D}}{\epsilon} \right)^l \tag{45}$$

for some constant C.

This theorem shows that the required bond dimension scales only poly-logarithmically in $1/\epsilon$, which indicates this kind of multivariate normal distribution can be efficiently represented as a TTN.

Proof. From (7), if the sum of the squares of the Schmidt coefficients for all bipartitions is less than ϵ^2/D, then an MPS approximation that satisfies (44) exists.

Let $q = \max_i q_i$, where q_i is the common ratio of the squared Schmidt coefficients for the pair (ξ_i, ξ_{p+i}). Consider keeping the r^l-th largest Schmidt coefficients and truncating all other small values. The truncation error is upper-bounded as follows:

$$\sum_{k=r^l}^{\infty} \lambda_k \leq 1 - (1-q^r)^l \tag{46}$$

$$\leq 1 - (1 - l \cdot q^r) = l \cdot q^r, \tag{47}$$

where we use the Weierstrass product inequality and $q^r < 1$. If we want to suppress the error by ϵ^2/D, we have to satisfy the following inequality:

$$l \cdot q^r \leq \epsilon^2/D. \tag{48}$$

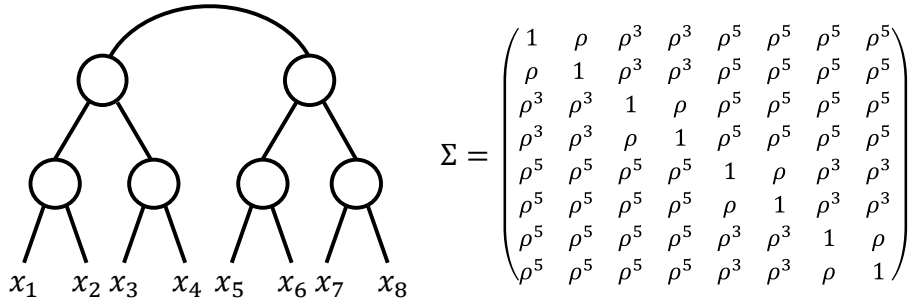

Figure 2: An example of a multivariate normal distribution with a tree-like structure. The covariance decays exponentially with the distance on the tree between variables.

By solving this we get

$$r \geq \frac{2}{\log \frac{1}{q}} \log \frac{l\sqrt{D}}{\epsilon}. \tag{49}$$

□

Example 3.1. The multivariate normal distribution with uniform correlation coefficients. Consider the following covariance matrix:

$$\Sigma = \begin{pmatrix} 1 & \rho & \rho & \cdots & \rho \\ \rho & 1 & \rho & \cdots & \rho \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \rho & \cdots & 1 \end{pmatrix}, \tag{50}$$

with $-1 < \rho < 1$. Then any bipartition of the system gives rank-one submatrix Σ_{12}. Therefore, this probability distribution can be efficiently represented as an MPS with any index ordering.

Example 3.2. The multivariate normal distribution with 1D correlations. Consider the following covariance matrix:

$$\Sigma = \begin{pmatrix} 1 & \rho & \rho^2 & \cdots & \rho^D \\ \rho & 1 & \rho & \cdots & \rho^{D-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^D & \rho^{D-1} & \rho^{D-2} & \cdots & 1 \end{pmatrix}, \tag{51}$$

with $-1 < \rho < 1$. Then, a bipartition of subsystems $x_1, \ldots, x_i$ and $x_{i+1}, \ldots, x_D$ gives a rank-one submatrix Σ_{12}. Therefore, this probability distribution can be efficiently represented as an MPS with the same ordering.

Example 3.3. The multivariate normal distribution with tree-like correlations. Consider the elements of the covariance matrix are given by its distance:

$$\Sigma_{ij} = \exp\left(-\frac{\text{distance}(i,j)}{\sigma}\right), \tag{52}$$

where the distance between site i and j is given as the length of the path on the tree as shown in Fig 2. Then, the bipartition of the system induced by each edge of the tree gives rank-one submatrix Σ_{12} and this probability distribution can be efficiently represented as a TTN. This generalizes the above two examples.

In addition to the case where the rank of the submatrix is upper-bounded, we can consider the following situation where correlation coefficients decay exponentially. In this case, however, the required bond dimension for approximation scales polynomially in $1/\epsilon$ instead of $\log(1/\epsilon)$.

Theorem 3.2. *Let $p(\boldsymbol{x})$ be a probability density function of a D-dimensional multivariate normal distribution with zero mean and covariance matrix Σ. Suppose the canonical correlations are less than 1 and decay exponentially:*

$$\rho_j \leq \alpha e^{-\theta \cdot j} \quad \alpha, \theta > 0 \tag{53}$$

for any bipartition of variables induced by an inner edge of the tree tensor network. For every ϵ there exists a TTN approximation $\tilde{\psi}$ of $\psi(\boldsymbol{x}) = \sqrt{p(\boldsymbol{x})}$ with

$$\|\psi(\boldsymbol{x}) - \tilde{\psi}(\boldsymbol{x})\|_{L^2(\mathbb{R}^D)} \leq \epsilon \tag{54}$$

whose bond dimension is upper-bounded by

$$\chi \leq \left(\frac{C_2\sqrt{D}}{\epsilon}\right)^{C_1 \log\left(C_3 \log\left(\frac{2C_1 \log(\frac{C_2\sqrt{D}}{\epsilon})\sqrt{D}}{\epsilon}\right)\right)} \tag{55}$$

for some constants C_1, C_2, C_3.

This theorem indicates that the necessary bond dimension scales as $\chi = O(\mathrm{poly}(\frac{C_2\sqrt{D}}{\epsilon}))$ if we ignore the log term in the exponent. The efficiency of the approximation is worse compared to the case where the rank of the covariance submatrix is strictly upper-bounded.

Proof. By (29),

$$q_j = \frac{K_j - 1}{K_j + 1} = \frac{\left(1 - \sqrt{1 - \rho_j^2}\right)^2}{\rho_j^2} < \rho_j \quad (0 < \rho_j < 1). \tag{56}$$

Therefore, when ρ_j decays exponentially, the common ratio q_j also decays exponentially to j.

Consider the two-step approximation: First, we retain only the largest singular value for pairs (ξ_i, ξ_{p+i}) with $i > l$. Then the approximation error by introducing this is calculated as

$$\begin{aligned}
\|\psi(\boldsymbol{x}) - \overline{\psi}(\boldsymbol{x})\|^2_{L^2(\mathbb{R}^D)} &= 1 - (1 - q_{l+1})(1 - q_{l+2}) \cdots (1 - q_p) && (57)\\
&\leq 1 - (1 - \alpha e^{-\theta(l+1)})(1 - \alpha e^{-\theta(l+2)}) \cdots (1 - \alpha e^{-\theta p}) && (58)\\
&\leq \sum_{i=l+1}^{p} \alpha e^{-\theta i} \leq \alpha \frac{e^{-\theta l}}{1 - e^{-\theta}}. && (59)
\end{aligned}$$

By solving the following inequality:

$$\alpha \frac{e^{-\theta l}}{1 - e^{-\theta}} \leq \frac{\epsilon^2}{4D}, \tag{60}$$

we get $l = C_1 \log(C_2\sqrt{D}/\epsilon)$ for constant numbers C_1, C_2.

Next, we consider approximating $\overline{\psi}(\boldsymbol{x})$ with finite Schmidt coefficients. Now the covariance matrix for the multivariate normal distribution $\overline{\psi}(\boldsymbol{x})$ only have l non-zero correlation

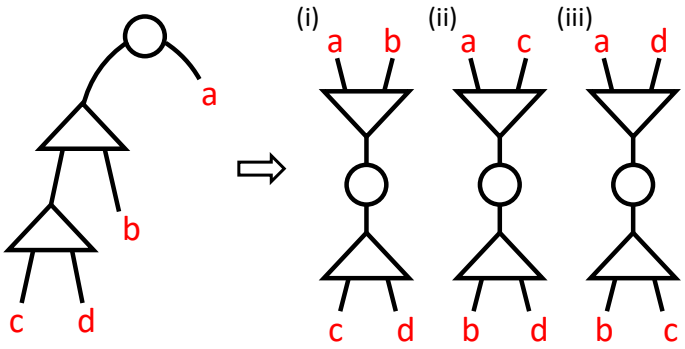

Figure 3: The process of automatic structural optimization of the TTN. When sweeping the top tensor, there are three patterns for choosing indices. By selecting the pattern that results in the lowest entanglement entropy, calculated from the singular values contained in the top tensor, the structure of TTN can be automatically optimized to the most suitable configuration.

coefficients. According to theorem 3.1, we can approximate it with error $\epsilon/(2\sqrt{D})$ by using r^l Schmidt coefficients and $r = C_3 \log(2l\sqrt{D}/\epsilon)$ for constant number C_3. Then we get

$$\|\psi(\boldsymbol{x}) - \tilde{\psi}(\boldsymbol{x})\|_{L^2(\mathbb{R}^D)} \leq \|\psi(\boldsymbol{x}) - \overline{\psi}(\boldsymbol{x})\|_{L^2(\mathbb{R}^D)} + \|\overline{\psi}(\boldsymbol{x}) - \tilde{\psi}(\boldsymbol{x})\|_{L^2(\mathbb{R}^D)} \tag{61}$$

$$\leq \epsilon/\sqrt{D} \tag{62}$$

Therefore, the necessary bond dimension to approximate the entire quantum state as TTN within precision ϵ scales as

$$r^l = C_3 \log \left(\frac{2l\sqrt{D}}{\epsilon} \right)^l \tag{63}$$

$$= C_3 \log \left(\frac{2C_1 \log\left(\frac{C_2\sqrt{D}}{\epsilon}\right)\sqrt{D}}{\epsilon} \right)^{C_1 \log\left(\frac{C_2\sqrt{D}}{\epsilon}\right)} \tag{64}$$

$$= \left(\frac{C_2\sqrt{D}}{\epsilon} \right)^{C_1 \log\left(C_3 \log\left(\frac{2C_1 \log(\frac{C_2\sqrt{D}}{\epsilon})\sqrt{D}}{\epsilon} \right) \right)} \tag{65}$$

□

3.3 Automatic structural optimization of the TTN

As discussed above, the required bond dimension for achieving enough accuracy can be estimated by the distribution of Schmidt coefficients among variables that are bipartitioned by the network. Therefore, whether a TTN can efficiently represent a multivariate normal distribution depends strongly on the tensor network structure. Even if a multivariate normal distribution has a one-dimensional correlation structure as in Th. 3.1 and 3.2, an efficient approximation is no longer guaranteed if an appropriate TTN corresponding to that structure is not used. In the case of a normal distribution, the optimal network structure can be determined by enumerating all network configurations and estimating their fidelity from the correlation coefficients. However, for large D, it becomes increasingly impractical, as there are $O(D!)$ ways to arrange variables, even when restricted to MPS.

Instead, in this study, we employ automatic structural optimization [18] to optimize network configuration. Fig. 3 shows the outline of the algorithm. Initially, we execute

TCI using an initial network structure (e.g., MPS) to represent the multivariate normal distribution. Then, we sweep the top tensor in the network using the canonical form. During this process, we contract the three tensors in Fig. 3 and obtain T_{abcd}. We consider three ways of singular value decomposition:

$$T_{abcd} = U_{abe} s_e V_{ecd} \tag{66}$$
$$T_{abcd} = U_{ace} s'_e V_{ebd} \tag{67}$$
$$T_{abcd} = U_{ade} s''_e V_{ebc} \tag{68}$$

and compute the entanglement entropy as follows:

$$EE^{(ab|cd)} = -\sum_e s_e^2 \log s_e^2 \tag{69}$$
$$EE^{(ac|bd)} = -\sum_e s_e'^2 \log s_e'^2 \tag{70}$$
$$EE^{(ad|bc)} = -\sum_e s_e''^2 \log s_e''^2 \tag{71}$$

where we assumed that TTN is normalized. Among them, we choose the structure with the smallest entanglement entropy. By repeating this process, the entire TTN network is automatically optimized to reflect the entanglement structure of the system.

Entanglement entropy takes small values when the singular values decay rapidly. As discussed previously, the faster the correlation coefficients decay, the more rapidly the Schmidt coefficients decrease. Thus, in this context, entanglement entropy indicates the strength of correlation when dividing variables into two groups. The process of automatic structural optimization can be interpreted as assuming a tree-like correlation structure in the probability distribution and searching the most suitable network configuration where the correlation between variables is minimized.

We first execute TCI using an initial tree tensor network with a sufficiently large bond dimension enough to accurately represent the multivariate normal distribution. Then, we employ structural optimization to refine the network structure and reduce the bond dimension through truncation using the canonical form. If the accuracy of the state obtained is insufficient, we can rerun TCI using the optimized network structure to achieve higher accuracy.

3.4 Optimizing the quantum circuit by tree tensor network

Up to this point, we have discussed how to represent Fourier coefficients using tensor networks efficiently. After encoding the Fourier coefficients as a quantum circuit, we can get the discretized probability distribution by applying inverse QFT with a depth of n.

We can further compile the entire state preparation circuit. When expressing Fourier coefficients with an MPS or TTN, a bond dimension 2^m is inevitably required for a physical index. This bond dimension can be reduced by utilizing the fact that the inverse QFT circuit for each dimension can be represented as a TTN, as illustrated in Fig 4. Here, the black node in TTN represents a copy tensor δ_{ijk} with bond dimension 2^m and A_j is a $2^k \times 2$ tensor whose elements are given as

$$(A_j)_l^k = e^{2\pi i k l / 2^{j+1}}. \tag{72}$$

We then apply bond truncation algorithms using the canonical form. While this approach loses the advantage of executing inverse QFT with a circuit depth of n, it allows for the potential creation of a more compact circuit. Additionally, as the size of each tensor decreases, the classical resources required for the circuit synthesis of each tensor are reduced.

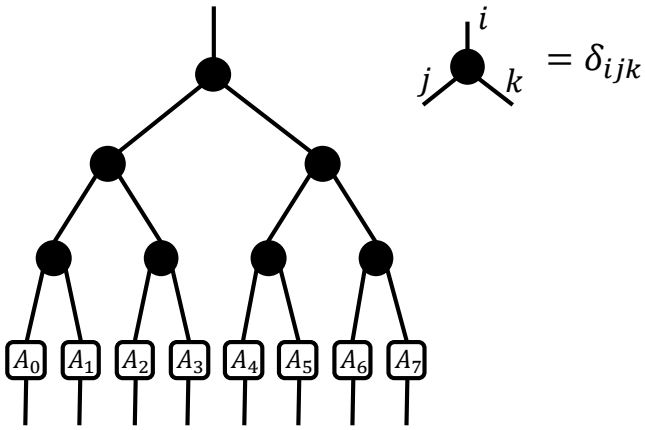

Figure 4: The tree tensor network representing inverse QFT. The black node represents a copy tensor. The definition of tensors $\{A_j\}_{j=0}^{n}$ is given in the main text. The number of qubits for each dimension is set as $n = 8$ in this figure.

4 Results

Here we present our simulation results for the compilation method of the state preparation circuit of multivariate normal distribution described in Sec. 3. We set the parameters as $a = 20$ ($x \in [-a/2, a/2]^D$), $n = 8$ (discretizing each dimension into 2^n points) and $m = 5$ (keep 2^m Fourier coefficients for each dimension) to discretize the probability density function with sufficient accuracy. We first use TCI to express the Fourier coefficients (21) as an MPS with bond dimension χ', where χ' is set sufficiently large to represent the target Fourier coefficients. In this study, we set $\chi' = 64$.

The fidelity between the original and approximated states of the i-th step of truncation can be calculated using the singular values as follows:

$$f^{(i)} = \sum_{j=0}^{\chi-1} {\sigma_j^{(i)}}^2. \tag{73}$$

Note that we normalize the quantum state after each truncation step so that the summation of the squares of singular values is always one. After completing the truncation process, the fidelity of the final state is predicted as the product of the fidelity at each step:

$$f = \prod_i f^{(i)} \tag{74}$$

This approach serves as a good estimate. Subsequently, the Frobenius norm is restored using the following equation:

$$\| \, |\psi\rangle - |\tilde{\psi}\rangle \, \|_F = \sqrt{2(1 - \sqrt{f})} \tag{75}$$

The tensor-network algorithms are implemented using the Python library Quimb [14]. Each data point on the plots is obtained by averaging 100 random instances.

4.1 Multivariate normal distribution with low-rank structure

To numerically test the theorem 3.1, we represented a multivariate normal distribution with a low-rank covariance matrix using MPS. We utilized a simple covariance matrix with a

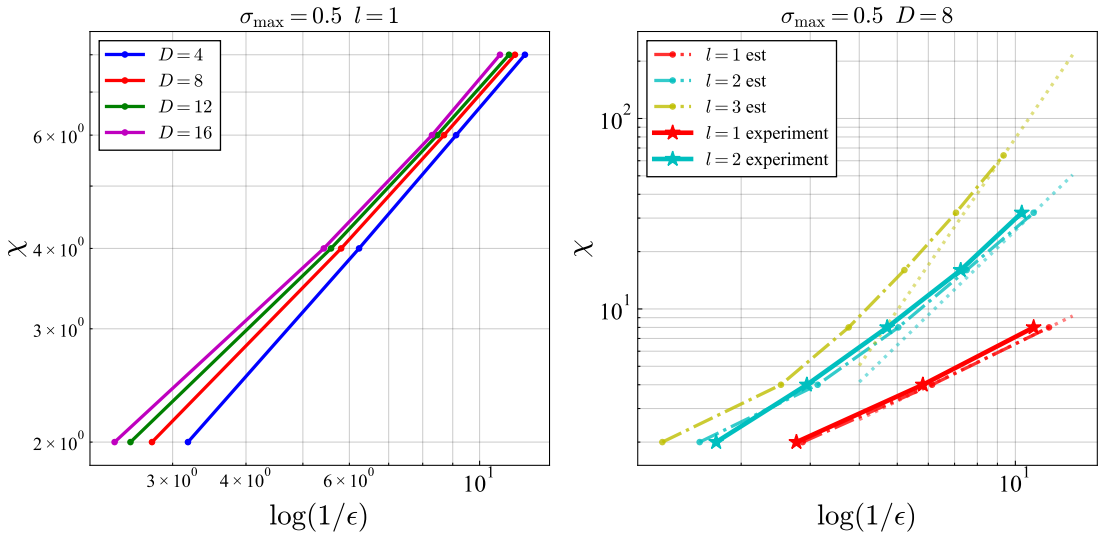

Figure 5: The bond dimension of the MPS is plotted as a function of the accuracy $\log(1/\epsilon)$ (Left) for various dimensions of the multivariate normal distribution and (right) for various ranks of the submatrix of the covariance matrix. "est" represents the estimation value from the distribution of the Schmidt coefficients and "experiment" represents the actual value from the numerical experiment. The dotted lines represent the predicted polynomial growth rate in $\log(1/\epsilon)$.

low-rank structure similar to the Example 3.2, with σ uniformly sampled from $[0, \sigma_{\max}]$ and $\sigma_{\max} = 0.5$. We consider a set of l multivariate normal distributions. We partition the variables into non-overlapping subgroups $\{x_i^{(j)}\}$ as follows: for $1 \leq j \leq l$, let

$$x_1^{(j)} \in \{x_1, x_2, \ldots, x_l\}, \quad x_2^{(j)} \in \{x_{l+1}, \ldots, x_{2l}\}, \quad \cdots . \tag{76}$$

Each subgroup follows a multivariate normal distribution with a 1D structure. For the probability density function constructed in this manner, the condition on the rank of the covariance matrix $\mathrm{rank}(\Sigma_{12}) \leq l$ in Theorem 3.1 is satisfied.

In Fig. 5, we illustrate the relationship between the accuracy obtained from MPS approximation and the corresponding bond dimension. In the left figure, the rank l is fixed at 1 and the dimension D is varied. As can be seen from the figure, the required bond dimension increases linearly with $\log(1/\epsilon)$. Moreover, the required bond dimension grows only gradually as D increases, suggesting that this approach scales effectively with dimension.

In the right plot, the dimension D is fixed, and the rank l is varied. The dash-dot line uses the distribution of Schmidt coefficients to directly estimate the achievable accuracy from the bond dimension, while the solid line calculates accuracy by obtaining the actual MPS. The results confirm the Theorem 3.1 that the required bond dimension increases with $\log(1/\epsilon)^l$. Furthermore, the accuracy obtained from the actual MPS is found to be very close to that predicted from the Schmidt coefficients, confirming the effectiveness of this bound in estimating the necessary bond dimension given the target accuracy.

4.2 Multivariate normal distribution with exponentially decaying correlation coefficients

Next, we represent a multivariate normal distribution in which canonical correlations decay exponentially. To construct such a distribution, we consider the following setup: the sites

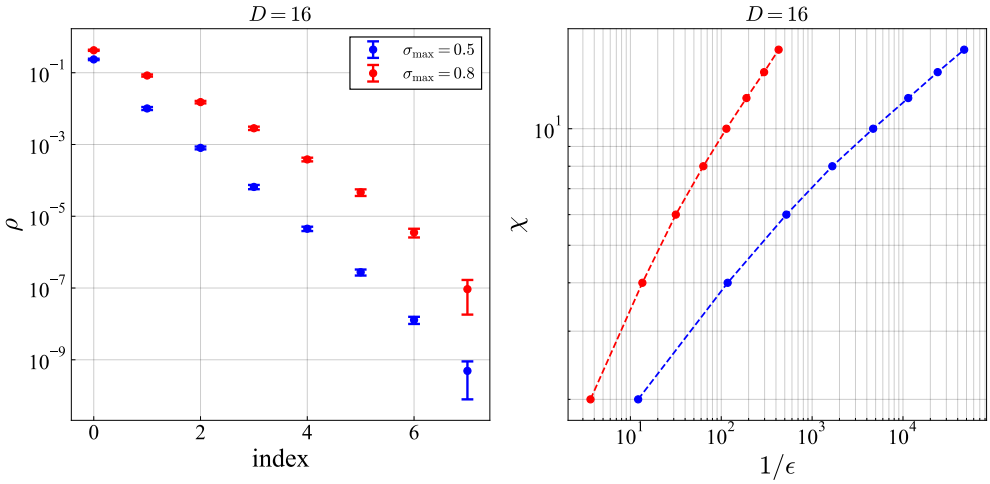

Figure 6: (left) The distribution of the canonical correlation coefficients when the system with $D = 16$ is divided in half. (right) The bond dimension of the MPS is plotted as a function of the accuracy $1/\epsilon$. Different maximum values of covariance $\sigma_{\max}$ are plotted in different colors.

are aligned in 1D and the covariance between site i and j is defined as

$$\mathrm{cov}_{ij} = (\sqrt{\sigma_i \sigma_j})^{|i-j|}, \tag{77}$$

where σ_i is uniformly sampled from $[0, \sigma_{\max}]$. We numerically verified that the correlations decay exponentially with the distance between sites. Fig. 6 (left) shows the distribution of canonical correlations between sites 1 to 8 and 9 to 16 for the multivariate normal distribution with dimension $D = 16$. As can be seen from the figure, the canonical correlations decay exponentially with the index number, satisfying the condition $\rho_j \leq \alpha e^{-\theta j}$ in Theorem 3.2.

Fig. 6 (right) shows the relationship between accuracy and the required bond dimension for approximating the above multivariate normal distribution with MPS. Note that the x-axis represents $1/\epsilon$, not $\log(1/\epsilon)$. As can be seen, polynomial growth in $1/\epsilon$ is confirmed, which is consistent with Theorem 3.2. These numerical results demonstrate that when correlations decay exponentially in 1D, it is possible to efficiently represent them using MPS regardless of the dimension D, although the efficiency significantly decreases compared to cases when the rank is strictly bounded.

4.3 Structual optimization with 1D tree structure

We also evaluated the performance of the structural optimization introduced in Sec. 3.3. We consider a multivariate normal distribution with a tree-like structure as presented in the Example 3.3. The initial structure of a tree is generated randomly. We approximate the distribution using a standard MPS using TCI without considering the covariance structure. Then, we employed automatic structural optimization to modify the tensor network, as depicted in Fig. 7 (a), and verify whether it could restore the original tree structure.

Fig. 7 (b) displays the relationship between the dimension D of the multivariate normal distribution and the success rate of structural optimization, i.e. the probability of completely restoring the original structure. Here, χ represents the maximum bond dimension of the tensor network during the process. We see that even for high-dimensional distributions like $D = 16$, automatic structural optimization can replicate the original structure with

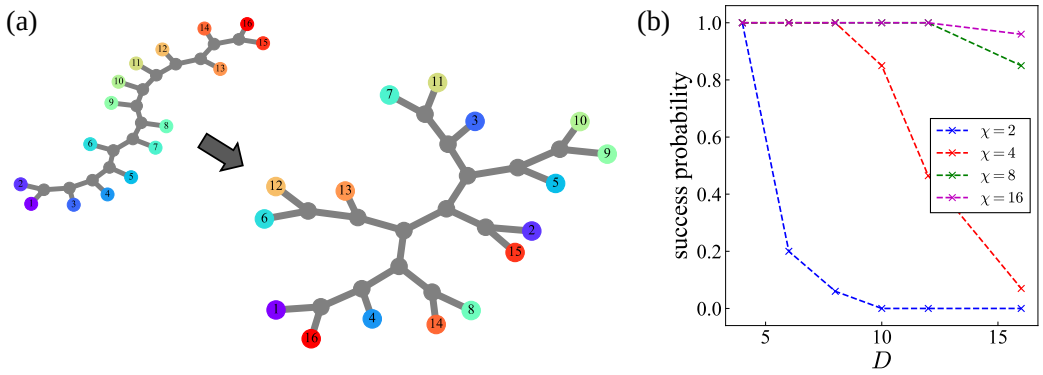

Figure 7: (a) Initially, the distribution is represented as an MPS, and after automatic structural optimization, the tree-like structure is recovered. (b) The success probability of the structural optimization is plotted as a function of the dimension D. Different bond dimensions in the tensor network are represented in different colors.

high probability. Additionally, increasing the bond dimension provides more accurate information about the bipartite entanglement entropy of the system, allowing a higher success rate. These numerical results indicate that despite being a heuristic algorithm, structural optimization is extremely powerful in capturing the 1D structure of distributions.

4.4 State preparation of random normal distribution

Now we move on to the compilation of the state preparation circuits for the multidimensional normal distribution without any 1D or tree-like structure.

Before discussing the results, we outline how the generated circuits are evaluated. We assume all-to-all connectivity. The main metrics adopted here are the number of CNOT gates and the depth of the circuit, both of which are very important, especially for NISQ computers. Regarding the number of CNOT gates, it is known that an isometry with a p-qubit input and a q-qubit output in the tensor network can be implemented using $O(2^{p+q})$ CNOT gates [22]. For simplicity and comparison purposes, we treat this number as 2^{p+q}, essentially equating it to the memory size required by the tensor network. Circuit depth is the necessary depth for implementing all two-qubit gates and is calculated by identifying the longest path from the root to the leaf within the TTN. If there is an isometry of size 2^{p+q} along a path, then 2^{p+q} is added to the circuit depth. Using these metrics, we estimate the cost of the circuits. Actual circuit synthesis on quantum hardware is beyond the scope of this work and is left for future investigation.

We consider preparing the quantum state of a 4-dimensional multivariate normal distribution with a randomly generated covariance matrix. Each element of the covariance matrix is generated from the uniform random distribution $[-\sigma_{\max}, \sigma_{\max}]$, except for the diagonal elements, which are all 1.0 for regularization.

In Fig. 8, we present the circuit depth, CNOT count, and infidelity of the encoding circuit generated by our method. For comparison, we also plot the costs using the FSL method without tensor network approximations, which serves as a baseline [35]. As shown in the figure, our approach compiles circuits that achieve high accuracy while using 10 to 1000 times fewer resources than the baseline. Additionally, by using a lower bond dimension, it is possible to perform state preparation at a lower cost, albeit at the expense of accuracy. This is particularly useful for NISQ algorithms, where the required precision is lower, allowing the use of circuits that require very few resources. Notably, at $\sigma_{\max} = 0.2$,

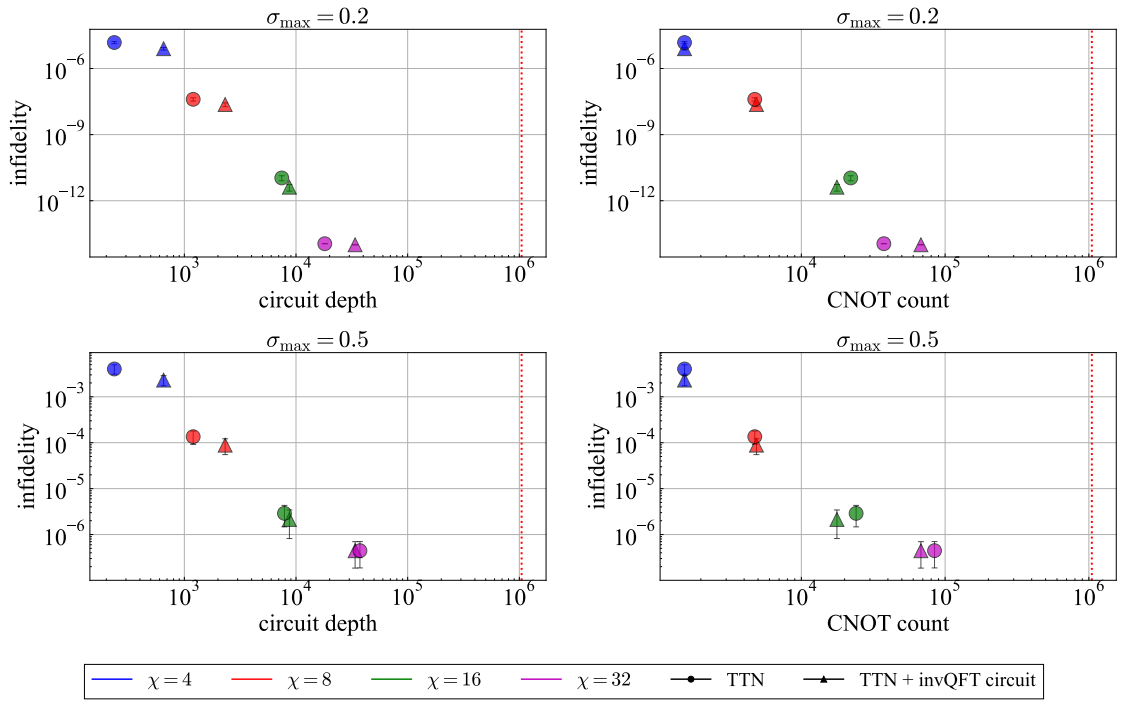

Figure 8: The circuit depth and CNOT count of the encoding circuit, and the infidelity of the quantum state generated by the encoding circuit. The parameters are set as $\sigma_{\max} = 0.2, 0.5$. The circle points show the result by representing inverse QFT as a TTN, discussed in Sec. 3.4, and triangle points show the result by using standard inverse QFT circuit. The red line implies the result without MPS approximation.

we can further reduce the circuit depth and CNOT count that optimize the entire circuit through a TTN representation, as discussed in Sec. 3.4.

On the other hand, increasing $\sigma_{\max}$ leads to a significant reduction in accuracy. $\sigma_{\max}$ represents the strength of the correlations between variables. Stronger correlations result in larger Schmidt coefficients, making tensor network approximations more challenging.

Finally, we discuss the effect of network structure and automatic structural optimization. Fig. 9 shows the infidelity of the state preparation circuit by varying the strategy for determining network structure. We get the optimal network structure by enumerating all possible candidates and estimating the fidelity. It is found that using tensor network structural optimization achieves accuracy close to that of optimal network structure, compared to when optimization is not employed. This highlights the significant influence of network structure on the performance of tensor network-based quantum circuit compilation and underscores the potential of structural optimization techniques in finding the near-optimal ordering, even when there are no 1D or tree-like structures.

5 Discussion

In this study, we proposed a scalable circuit compilation method for multivariate normal distributions using TTN. We first considered multivariate normal distributions with 1D correlation structures. We examined two cases: the rank of the covariance matrix is bounded, and the correlation coefficients decay exponentially with the distance between variables. We theoretically guaranteed that these probability distributions can be efficiently represented using TTN.

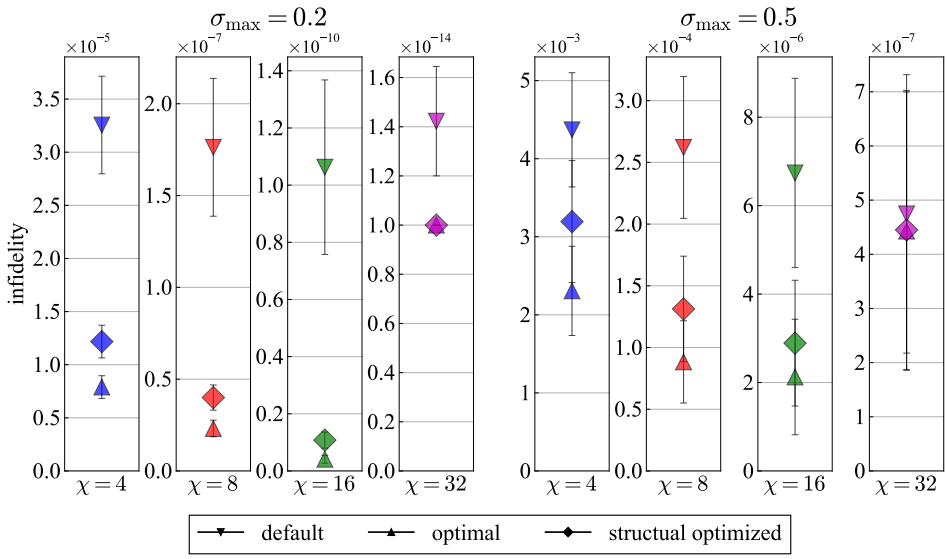

Figure 9: The infidelity of the encoding circuit with different strategies for determining network structure. The parameters are set as $\sigma_{\max} = 0.2, 0.5$. The lower triangle points show the result without any optimization of the ordering. The upper triangle points show the result with optimal network structure. The diamond points show the result with the automatic structural optimization method discussed in Sec. 3.3.

Additionally, we introduced automatic structural optimization into our circuit compilation method. This optimization rearranges the structure of tensor networks based on the bipartite entanglement entropy of the system. By applying it to multivariate probability distributions, we found that it searches for structures that appropriately reflect the 1D correlation structure of the distribution.

We combined these techniques with the FSL method, TCI, and truncation method using the canonical form. We numerically demonstrated that we can scalably compile the state preparation circuit for normal multivariate distributions with 1D structures. Moreover, even for random normal distributions without a specific structure, we confirmed numerically that our approach could significantly reduce the circuit depth and the number of CNOT gates compared to existing methods.

Multivariate normal distributions play an important role in various practical quantum algorithms. An important example is quantum algorithms for quantum field theory [24, 25]. In lattice ϕ^4 theory, the vacuum state $|\mathrm{vac}\rangle$ is represented as a multivariate normal distribution in the field variables $\phi(\boldsymbol{x})$, with the covariance matrix defined as

$$G(\boldsymbol{x} - \boldsymbol{y}) = \langle \mathrm{vac}|\phi(\boldsymbol{x})\phi(\boldsymbol{y})|\mathrm{vac}\rangle \,. \tag{78}$$

In the free theory, the covariance matrix $G(\boldsymbol{x} - \boldsymbol{y})$ decays exponentially with the distance $|\boldsymbol{x} - \boldsymbol{y}|$. Therefore, when the lattice is defined in one spatial dimension, the vacuum state can be well approximated using MPS, allowing efficient compilation via our proposed methods.

Another important application lies in quantum algorithms for finance. In finance, the relationship between the risks and returns of multiple assets is modeled with a multivariate normal distribution; therefore, preparing a multivariate normal quantum state is essential for quantum Monte Carlo integration applied to derivative pricing [54, 55] and risk analysis [33, 58, 59]. These financial datasets are not guaranteed to exhibit exact tree-like correlation structures. However, as demonstrated in our numerical experiments,

our heuristic optimization algorithm may help reduce the cost of preparing such quantum states.

One of the advantages of our method is the ability to calculate achievable fidelity in advance. There are a lot of methods for circuit compiling for NISQ algorithms [2, 11, 12, 15, 19–21, 23, 26, 31, 38, 43, 48, 49, 60, 61], which also produce low-cost circuits at the expense of fidelity. However, many algorithms are based on variational optimization and there are often no theoretical guarantees or estimates regarding the accuracy of their approximations. Our method does not depend on variational optimization and can estimate the achievable fidelity from the covariance matrix and bond dimension in the network. Although this study was limited to multivariate normal distributions with theoretical guarantees, it can be extended to other multivariate probability distributions and multivariable functions [56].

Another advantage of our method is its hardware compatibility. While our numerical simulations assume all-to-all connectivity, the hierarchical structure of the TTN allows each local tensor to be encoded into a local quantum circuit and connected using SWAP gates in a layered manner. If the bond dimension is sufficiently small, the cost of isometry synthesis can also be kept low. Therefore, the proposed method is expected to be practical even on quantum hardware with limited qubit connectivity. We left fully synthesizing circuits and testing them on real hardware for future work.

From the perspective of tensor network algorithms, this research can also be considered a new application of automatic structural optimization. Automatic structural optimization, developed in the field of many-body physics, has been used to detect the entanglement structure of systems. In [16], the techniques were applied to the field of machine learning, utilizing mutual information to efficiently represent datasets. This work demonstrates that these techniques can also be applied to the field of quantum circuit optimization, underscoring the potential applicability across various fields.

Future directions could involve the use of other tensor network algorithms, especially those designed for two-dimensional systems, for the compilation of state preparation circuits. Moreover, since automatic structural optimization remains a heuristic algorithm, there is a potential to develop theoretically guaranteed or more sophisticated structural optimization algorithms.

An open-source implementation of the method for detecting the correlation structure of multivariate normal distributions is available and has been described in a subsequent paper [57].

Acknowledgement

H. Manabe is supported by JSTCOI-NEXT program Grant Numbers JPMJPF2014. Y. Sano is supported by JSPS KAKENHI Grant Numbers JP23KJ1178. We also thank to the Quantum Computing for Quantum Field Theory summer school 2023 for providing us with the opportunity for collaboration.

References

[1] Christian W. Bauer, Plato Deliyannis, Marat Freytsis, and Benjamin Nachman. Practical considerations for the preparation of multivariate Gaussian states on quantum computers. *arXiv preprint arXiv:2109.10918*, 2021. DOI: 10.48550/arXiv.2109.10918.

[2] Matan Ben-Dov, David Shnaiderov, Adi Makmal, and Emanuele G. Dalla Torre. Approximate encoding of quantum states using shallow circuits. *npj Quantum Information*, 10(1):1–8, 2024. ISSN 2056-6387. DOI: 10.1038/s41534-024-00858-1.

[3] Jacob Biamonte, Peter Wittek, Nicola Pancotti, Patrick Rebentrost, Nathan Wiebe, and Seth Lloyd. Quantum machine learning. *Nature*, 549(7671):195–202, 2017. ISSN 1476-4687. DOI: 10.1038/nature23474.

[4] Daniele Bigoni, Allan P. Engsig-Karup, and Youssef M. Marzouk. Spectral tensor-train decomposition. *SIAM Journal on Scientific Computing*, 38(4):A2405–A2439, 2016. ISSN 1064-8275, 1095-7197. DOI: 10.1137/15M1036919.

[5] A. Yu. Bogdanov, Yu. I. Bogdanov, and K. A. Valiev. Schmidt information and entanglement of quantum systems. *Moscow University Computational Mathematics and Cybernetics*, 31(1):33–42, 2007. ISSN 1934-8428. DOI: 10.3103/S0278641907010074.

[6] Yu I. Bogdanov, N. A. Bogdanova, D. V. Fastovets, and V. F. Luckichev. Schmidt decomposition and multivariate statistical analysis. In *International Conference on Micro- and Nano-Electronics 2016*, volume 10224, pages 706–715. SPIE, 2016. DOI: 10.1117/12.2266891.

[7] Jacob C. Bridgeman and Christopher T. Chubb. Hand-waving and Interpretive Dance: An Introductory Course on Tensor Networks. *Journal of Physics A: Mathematical and Theoretical*, 50(22):223001, 2017. ISSN 1751-8113, 1751-8121. DOI: 10.1088/1751-8121/aa6dc3.

[8] Almudena Carrera Vazquez and Stefan Woerner. Efficient State Preparation for Quantum Amplitude Estimation. *Physical Review Applied*, 15(3):034027, 2021. ISSN 2331-7019. DOI: 10.1103/PhysRevApplied.15.034027.

[9] Anirban Narayan Chowdhury and Rolando D. Somma. Quantum algorithms for Gibbs sampling and hitting-time estimation. *Quantum Info. Comput.*, 17(1-2):41–64, 2017. ISSN 1533-7146. DOI: 10.26421/QIC17.1-2-3.

[10] Daniel J. Egger, Ricardo García Gutiérrez, Jordi Cahué Mestre, and Stefan Woerner. Credit Risk Analysis Using Quantum Computers. *IEEE Transactions on Computers*, 70(12):2136–2145, 2021. ISSN 1557-9956. DOI: 10.1109/TC.2020.3038063.

[11] Paula García-Molina, Javier Rodríguez-Mediavilla, and Juan José García-Ripoll. Quantum Fourier analysis for multivariate functions and applications to a class of Schrödinger-type partial differential equations. *Physical Review A*, 105(1):012433, 2022. DOI: 10.1103/PhysRevA.105.012433.

[12] Javier Gonzalez-Conde, Thomas W. Watts, Pablo Rodriguez-Grasa, and Mikel Sanz. Efficient quantum amplitude encoding of polynomial functions. *Quantum*, 8:1297, 2024. ISSN 2521-327X. DOI: 10.22331/q-2024-03-21-1297.

[13] Alex Gorodetsky, Sertac Karaman, and Youssef Marzouk. A continuous analogue of the tensor-train decomposition. *Computer Methods in Applied Mechanics and Engineering*, 347:59–84, 2019. ISSN 0045-7825. DOI: 10.1016/j.cma.2018.12.015.

[14] Johnnie Gray. Quimb: A python package for quantum information and many-body calculations. *Journal of Open Source Software*, 3(29):819, 2018. ISSN 2475-9066. DOI: 10.21105/joss.00819.

[15] Prithvi Gundlapalli and Junyi Lee. Deterministic and Entanglement-Efficient Preparation of Amplitude-Encoded Quantum Registers. *Physical Review Applied*, 18(2): 024013, 2022. ISSN 2331-7019. DOI: 10.1103/PhysRevApplied.18.024013.

[16] Kenji Harada, Tsuyoshi Okubo, and Naoki Kawashima. Tensor tree learns hidden relational structures in data to construct generative models. *Machine Learning: Science and Technology*, 6(2):025002, 2025. ISSN 2632-2153. DOI: 10.1088/2632-2153/adc2c7.

[17] Aram W. Harrow, Avinatan Hassidim, and Seth Lloyd. Quantum Algorithm for Linear Systems of Equations. *Physical Review Letters*, 103(15):150502, 2009. ISSN 0031-9007, 1079-7114. DOI: 10.1103/PhysRevLett.103.150502.

[18] Toshiya Hikihara, Hiroshi Ueda, Kouichi Okunishi, Kenji Harada, and Tomotoshi Nishino. Automatic structural optimization of tree tensor networks. *Physical Review Research*, 5(1):013031, 2023. ISSN 2643-1564. DOI: 10.1103/PhysRevResearch.5.013031.

[19] Adam Holmes and A. Y. Matsuura. Efficient Quantum Circuits for Accurate State Preparation of Smooth, Differentiable Functions. In *2020 IEEE International Conference on Quantum Computing and Engineering (QCE)*, pages 169–179. IEEE Computer Society, 2020. ISBN 978-1-72818-969-7. DOI: 10.1109/QCE49297.2020.00030.

[20] Jason Iaconis and Sonika Johri. Tensor Network Based Efficient Quantum Data Loading of Images. *arXiv preprint arXiv:2310.05897*, 2023. DOI: 10.48550/arXiv.2310.05897.

[21] Jason Iaconis, Sonika Johri, and Elton Yechao Zhu. Quantum state preparation of normal distributions using matrix product states. *npj Quantum Information*, 10(1): 1–11, 2024. ISSN 2056-6387. DOI: 10.1038/s41534-024-00805-0.

[22] Raban Iten, Roger Colbeck, Ivan Kukuljan, Jonathan Home, and Matthias Christandl. Quantum Circuits for Isometries. *Physical Review A*, 93(3):032318, 2016. ISSN 2469-9926, 2469-9934. DOI: 10.1103/PhysRevA.93.032318.

[23] Bernhard Jobst, Kevin Shen, Carlos A. Riofrío, Elvira Shishenina, and Frank Pollmann. Efficient MPS representations and quantum circuits from the Fourier modes of classical image data. *Quantum*, 8:1544, 2024. DOI: 10.22331/q-2024-12-03-1544.

[24] Stephen P. Jordan, Keith S. M. Lee, and John Preskill. Quantum Algorithms for Quantum Field Theories. *Science*, 336(6085):1130–1133, 2012. ISSN 0036-8075, 1095-9203. DOI: 10.1126/science.1217069.

[25] Stephen P. Jordan, Keith S. M. Lee, and John Preskill. Quantum computation of scattering in scalar quantum field theories. *Quantum Info. Comput.*, 14(11-12):1014–1080, 2014. ISSN 1533-7146. DOI: 10.48550/arXiv.1112.4833.

[26] Raghav Jumade and Nicolas PD Sawaya. Data is often loadable in short depth: Quantum circuits from tensor networks for finance, images, fluids, and proteins. *arXiv preprint arXiv:2309.13108*, 2023. DOI: 10.48550/arXiv.2309.13108.

[27] Alexei Kitaev and William A. Webb. Wavefunction preparation and resampling using a quantum computer. *arXiv preprint arXiv:0801.0342*, 2009. DOI: 10.48550/arXiv.0801.0342.

[28] Xiangyu Li, Xiaolong Yin, Nathan Wiebe, Jaehun Chun, Gregory K. Schenter, Margaret S. Cheung, and Johannes Mülmenstädt. Potential quantum advantage for simulation of fluid dynamics. *Physical Review Research*, 7(1):013036, 2025. DOI: 10.1103/PhysRevResearch.7.013036.

[29] Gabriel Marin-Sanchez, Javier Gonzalez-Conde, and Mikel Sanz. Quantum algorithms for approximate function loading. *Physical Review Research*, 5(3):033114, 2023. ISSN 2643-1564. DOI: 10.1103/PhysRevResearch.5.033114.

[30] Sam McArdle, András Gilyén, and Mario Berta. Quantum state preparation without coherent arithmetic. *arXiv preprint arXiv:2210.14892*, 2022. DOI: 10.48550/arXiv.2210.14892.

[31] Ar A. Melnikov, A. A. Termanova, S. V. Dolgov, F. Neukart, and M. R. Perelshtein. Quantum state preparation using tensor networks. *Quantum Science and Technology*, 8(3):035027, 2023. ISSN 2058-9565. DOI: 10.1088/2058-9565/acd9e7.

[32] Kosuke Mitarai, Makoto Negoro, Masahiro Kitagawa, and Keisuke Fujii. Quantum Circuit Learning. *Physical Review A*, 98(3):032309, 2018. ISSN 2469-9926, 2469-9934. DOI: 10.1103/PhysRevA.98.032309.

[33] Koichi Miyamoto. Quantum algorithm for calculating risk contributions in a credit portfolio. *EPJ Quantum Technology*, 9(1):1–16, 2022. ISSN 2196-0763. DOI: 10.1140/epjqt/s40507-022-00153-y.

[34] Ashley Montanaro. Quantum speedup of Monte Carlo methods. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 471(2181): 20150301, 2015. ISSN 1364-5021, 1471-2946. DOI: 10.1098/rspa.2015.0301.

[35] Mudassir Moosa, Thomas W. Watts, Yiyou Chen, Abhijat Sarma, and Peter L. McMahon. Linear-depth quantum circuits for loading Fourier approximations of arbitrary functions. *Quantum Science and Technology*, 9(1):015002, 2023. ISSN 2058-9565. DOI: 10.1088/2058-9565/acfc62.

[36] Kohei Morimoto, Yusuke Takase, Kosuke Mitarai, and Keisuke Fujii. Continuous optimization by quantum adaptive distribution search. *Physical Review Research*, 6 (2):023191, 2024. DOI: 10.1103/PhysRevResearch.6.023191.

[37] Mikko Möttönen, Juha J. Vartiainen, Ville Bergholm, and Martti M. Salomaa. Transformation of quantum states using uniformly controlled rotations. *Quantum Info. Comput.*, 5(6):467–473, 2005. ISSN 1533-7146. DOI: 10.26421/QIC5.6-5.

[38] Kouhei Nakaji, Shumpei Uno, Yohichi Suzuki, Rudy Raymond, Tamiya Onodera, Tomoki Tanaka, Hiroyuki Tezuka, Naoki Mitsuda, and Naoki Yamamoto. Approximate amplitude encoding in shallow parameterized quantum circuits and its application to financial market indicators. *Physical Review Research*, 4(2):023136, 2022. DOI: 10.1103/PhysRevResearch.4.023136.

[39] I. V. Oseledets. Tensor-Train Decomposition. *SIAM Journal on Scientific Computing*, 33(5):2295–2317, 2011. ISSN 1064-8275, 1095-7197. DOI: 10.1137/090752286.

[40] I. V. Oseledets. Constructive Representation of Functions in Low-Rank Tensor Formats. *Constructive Approximation*, 37(1):1–18, 2013. ISSN 1432-0940. DOI: 10.1007/s00365-012-9175-x.

[41] Ivan Oseledets and Eugene Tyrtyshnikov. TT-cross approximation for multidimensional arrays. *Linear Algebra and its Applications*, 432(1):70–88, 2010. ISSN 0024-3795. DOI: 10.1016/j.laa.2009.07.024.

[42] Martin Plesch and Časlav Brukner. Quantum-state preparation with universal gate decompositions. *Physical Review A*, 83(3):032302, 2011. ISSN 1050-2947, 1094-1622. DOI: 10.1103/PhysRevA.83.032302.

[43] Shi-Ju Ran. Encoding of Matrix Product States into Quantum Circuits of One- and Two-Qubit Gates. *Physical Review A*, 101(3):032310, 2020. ISSN 2469-9926, 2469-9934. DOI: 10.1103/PhysRevA.101.032310.

[44] Arthur G. Rattew and Bálint Koczor. Preparing Arbitrary Continuous Functions in Quantum Registers With Logarithmic Complexity. *arXiv preprint arXiv:2205.00519*, 2022. DOI: 10.48550/arXiv.2205.00519.

[45] Arthur G. Rattew, Yue Sun, Pierre Minssen, and Marco Pistoia. The Efficient Preparation of Normal Distributions in Quantum Registers. *Quantum*, 5:609, 2021. ISSN 2521-327X. DOI: 10.22331/q-2021-12-23-609.

[46] W. D. Ray. The Advanced Theory of Statistics, Vol. 3: Design and Analysis and Time Series. *Royal Statistical Society. Journal. Series A: General*, 147(3):523, 1984. ISSN 0035-9238. DOI: 10.2307/2981604.

[47] Paul B. Rohrbach, Sergey Dolgov, Lars Grasedyck, and Robert Scheichl. Rank Bounds for Approximating Gaussian Densities in the Tensor-Train Format. *SIAM/ASA Journal on Uncertainty Quantification*, 10(3):1191–1224, 2022. ISSN 2166-2525. DOI: 10.1137/20M1314653.

[48] Manuel S. Rudolph, Jacob Miller, Danial Motlagh, Jing Chen, Atithi Acharya, and Alejandro Perdomo-Ortiz. Synergy Between Quantum Circuits and Tensor Networks: Short-cutting the Race to Practical Quantum Advantage. *arXiv preprint arXiv:2208.13673*, 2023. DOI: 10.48550/arXiv.2208.13673.

[49] Yuichi Sano and Ikko Hamamura. Quantum State Preparation for Probability Distributions with Mirror Symmetry Using Matrix Product States. *arXiv preprint arXiv:2403.16729*, 2024. DOI: 10.48550/arXiv.2403.16729.

[50] Dmitry V. Savostyanov. Quasioptimality of maximum-volume cross interpolation of tensors. *Linear Algebra and its Applications*, 458:217–244, 2014. ISSN 00243795. DOI: 10.1016/j.laa.2014.06.006.

[51] Ulrich Schollwoeck. The density-matrix renormalization group in the age of matrix product states. *Annals of Physics*, 326(1):96–192, 2011. ISSN 00034916. DOI: 10.1016/j.aop.2010.09.012.

[52] Y.-Y. Shi, L.-M. Duan, and G. Vidal. Classical simulation of quantum many-body systems with a tree tensor network. *Physical Review A*, 74(2):022320, 2006. ISSN 1050-2947, 1094-1622. DOI: 10.1103/PhysRevA.74.022320.

[53] Vladimir Skavysh, Sofia Priazhkina, Diego Guala, and Thomas R. Bromley. Quantum monte carlo for economics: Stress testing and macroeconomic deep learning. *Journal of Economic Dynamics and Control*, 153:104680, 2023. ISSN 0165-1889. DOI: 10.1016/j.jedc.2023.104680.

[54] Nikitas Stamatopoulos and William J. Zeng. Derivative Pricing using Quantum Signal Processing. *Quantum*, 8:1322, 2024. DOI: 10.22331/q-2024-04-30-1322.

[55] Queenie Sun, Nicholas Grablevsky, Huaizhang Deng, and Pooya Azadi. Quantum Computing for Multi Period Asset Allocation. *arXiv preprint arXiv:2410.11997*, 2024. DOI: 10.48550/arXiv.2410.11997.

[56] Joseph Tindall, Miles Stoudenmire, and Ryan Levy. Compressing multivariate functions with tree tensor networks. *arXiv preprint arXiv:2410.03572*, 2024. DOI: 10.48550/arXiv.2410.03572.

[57] Ryo Watanabe, Hidetaka Manabe, Toshiya Hikihara, and Hiroshi Ueda. TTNOpt: Tree tensor network package for high-rank tensor compression. *arXiv preprint arXiv:2505.05908*, 2025. DOI: 10.48550/arXiv.2505.05908.

[58] Sascha Wilkens and Joe Moorhouse. Quantum computing for financial risk measurement. *Quantum Information Processing*, 22(1):51, 2023. ISSN 1573-1332. DOI: 10.1007/s11128-022-03777-2.

[59] Stefan Woerner and Daniel J. Egger. Quantum risk analysis. *npj Quantum Information*, 5(1):1–8, 2019. ISSN 2056-6387. DOI: 10.1038/s41534-019-0130-6.

[60] Peng-Fei Zhou, Rui Hong, and Shi-Ju Ran. Automatically differentiable quantum circuit for many-qubit state preparation. *Physical Review A*, 104(4):042601, 2021. ISSN 2469-9926, 2469-9934. DOI: 10.1103/PhysRevA.104.042601.

[61] Elton Yechao Zhu, Sonika Johri, Dave Bacon, Mert Esencan, Jungsang Kim, Mark Muir, Nikhil Murgai, Jason Nguyen, Neal Pisenti, Adam Schouela, Ksenia Sosnova, and Ken Wright. Generative quantum learning of joint probability distribution functions. *Physical Review Research*, 4(4):043092, 2022. ISSN 2643-1564. DOI: 10.1103/PhysRevResearch.4.043092.

[62] Julien Zylberman and Fabrice Debbasch. Efficient Quantum State Preparation with Walsh Series. *arXiv preprint arXiv:2307.08384*, 2023. DOI: 10.48550/arXiv.2307.08384.