

Analyzing the barren plateau phenomenon in training quantum neural networks with the ZX-calculus

Chen Zhao^{1,2} and Xiao-Shan Gao^{1,2}

¹Academy of Mathematics and Systems Science, Chinese Academy of Sciences

²University of Chinese Academy of Sciences

In this paper, we propose a general scheme to analyze the gradient vanishing phenomenon, also known as the barren plateau phenomenon, in training quantum neural networks with the ZX-calculus. More precisely, we extend the barren plateaus theorem from unitary 2-design circuits to any parameterized quantum circuits under certain reasonable assumptions. The main technical contribution of this paper is representing certain integrations as ZX-diagrams and computing them with the ZX-calculus. The method is used to analyze four concrete quantum neural networks with different structures. It is shown that, for the hardware efficient ansatz and the MPS-inspired ansatz, there exist barren plateaus, while for the QCNN ansatz and the tree tensor network ansatz, there exists no barren plateau.

1 Introduction

In recent years, hybrid quantum-classical algorithms are widely used in quantum chemistry [1–4], combinatorial optimization [5, 6], and quantum machine learning [7–12]. In these hybrid quantum-classical algorithms, the goal is usually training parameterized quantum circuits (PQCs) with classical optimizers. The PQC will be applied to an initial state and then the state will be measured on a quantum device. The classical optimizer will update the parameters of the PQC according to the measurement results. As the PQC can be run on noisy intermediate-scale quantum (NISQ [13]) devices, these algorithms are regarded as near-term practical quantum algorithms with potential quantum advantages.

There exist many methods to train PQCs. Some of these are gradient-based [14–17] and some are not [17, 18]. In quantum machine learning, gradient-based methods are widely used. When using gradient-based methods to train PQCs, one may suffer from the barren plateau (BP) phenomenon which was first studied in [19]. The BP phenomenon is that the gradient of parameters of the PQC will vanish exponentially in terms of the system size. It was proved that if the PQCs form unitary 2-designs, then the BP phenomenon exists [19]. This result has been extended to the case when the PQCs form approximately 2-designs in [20]. The BP phenomenon in PQCs of various structures has been proposed. For PQCs with a brick-like structure, if the PQC has locally 2-design, then the existence of BPs depends on the depth of the circuit and the cost-function [21]. Let n be the number of qubits of the PQC. For $\text{poly}(n)$ -depth PQCs with a brick-like form, there always exist BPs. For $\log(n)$ -depth PQCs with a brick-like form, if the cost-function is global, there exist BPs. Otherwise, there exists no BP when the cost-function is local. Too much entanglement will induce BPs [22, 23]. The BP phenomenon in dissipative quantum neural networks has been studied in [24]. And the noise from quantum hardware also causes BPs, which are called noise-induced BPs [25]. Several methods to avoid BPs have been proposed [20, 23, 26, 27].

Chen Zhao: zhaochen17@mails.ucas.ac.cn
 Xiao-Shan Gao: xgao@mmrc.iss.ac.cn

The above results about the BP phenomenon are obtained under certain assumptions of unitary 2-design and it is still difficult to analyze the BP phenomenon for PQCs besides those containing t -design parts. In this paper, we develop a general scheme to analyze whether there exist BP phenomena when training a concrete PQC. We focus on BP phenomena induced by the structure of PQCs and noise-induced BPs are not considered in this paper. The most important tool used in this paper is the *ZX-calculus*, a graphical language for describing and reasoning about quantum processes. The ZX-calculus was developed by Coecke and Duncan in [28, 29], which has various applications including quantum circuit synthesis [30–33], measurement-based quantum computing [34, 35], quantum error correction [36, 37], condensed matter physics [38], quantum machine learning [39], and quantum natural language processing [40]. In the ZX-calculus, the objects under consideration are ZX-diagrams, which consist of two kinds of tensors: Z-spiders and X-spiders. A ZX-diagram can be rewritten with ZX-calculus rules. Moreover, every quantum circuit can be converted into a ZX-diagram.

Let $\vec{\theta} = (\theta_1, \dots, \theta_m)$ be a set of parameters. To analyze the gradient of a PQC $U(\vec{\theta})$ with respect to a Hamiltonian H , we need to estimate the following expectation and the variance

$$\mathbf{E} \left[\frac{\partial \langle H \rangle}{\partial \theta_j} \right], \mathbf{Var} \left[\frac{\partial \langle H \rangle}{\partial \theta_j} \right] \quad (1)$$

where $\langle H \rangle$ is defined in Eq. (3). It will be shown that the expectation in Eq. (1) is always zero. The PQC is said to have *barren plateaus* if the variance in Eq. (1) vanishes exponentially in terms of the size of the PQC. The PQC is said to have *no barren plateau* or *trainable* if the variance in Eq. (1) vanishes polynomially in terms of the size of the PQC.

To estimate the expectation and variance in Eq. (1), we first represent them as ZX-diagrams. Since the expectation and the variance are integrations, the main technical contribution of this paper is representing these integrations as ZX-diagrams and computing them with the ZX-calculus when the PQC satisfies Assumption 1. More precisely, with the rewriting rules in the ZX-calculus, we prove that Eq. (1) is equal to the contraction of a tensor network with a similar structure as the PQC. Hence, the existence of BPs is totally characterized by the scaling property of the tensor network.

We use these techniques to analyze whether there exist BP phenomena in the hardware-efficient ansatz [2], the QCNN ansatz [41], the tree tensor network ansatz [42], and the MPS-inspired ansatz [43]. We show that there exist BPs in hardware-efficient ansatz and MPS-inspired ansatz, and there exists no BP in the QCNN ansatz and the tree tensor network ansatz.

This paper is organized as follows. A brief introduction to the PQC, the BP phenomenon, and the ZX-calculus will be given in Section 2. We will prove the main result that characterizes Eq. (1) in Section 3. And the analysis of four concrete PQCs is given in Section 4.

2 Preliminary

2.1 Hybrid quantum-classical algorithms

In a hybrid quantum-classical algorithm, there will be an ansatz, which is a PQC of the form

$$U(\vec{\theta}) = \prod_{j=1}^M [U_j(\theta_j) \cdot V_j]. \quad (2)$$

In (2), $U_j(\theta_j)$, $j = 1, \dots, M$ are parameterized gates, such as the rotational gates R_X, R_Y, R_Z ; and V_j are non-parameterized gates, such as the Hadamard gate H and the CNOT gate. The PQC will be applied to an initial state ρ_0 and then the state will be measured. The above procedure, which is the quantum part of the algorithm, will be run on quantum processors. Meanwhile, there will be a classical part that consists of classical processors to optimize the parameters of the PQC in the quantum part. A cost-function $L(\vec{\theta})$ will be estimated in the classical part based on the

measurement results. Usually, the expectation

$$\langle H \rangle = \text{Tr} \left(\rho_0 U^\dagger(\vec{\theta}) H U(\vec{\theta}) \right) \quad (3)$$

of a given Hamiltonian H will be regarded as the cost-function in many tasks.

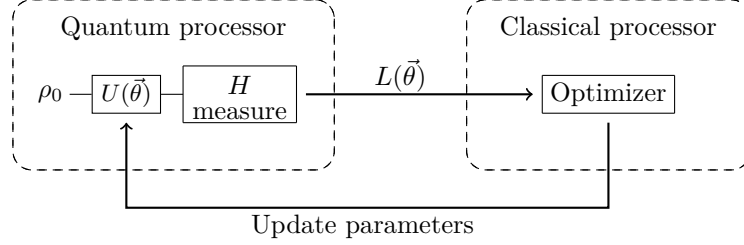


Figure 1: The hybrid quantum-classical algorithm

As demonstrated in Figure 1, the quantum part runs the PQC and obtains the measurement results and the classical part estimates the cost-function and updates the parameters. After several iterations, the cost-function may converge and be optimized. Then the training will be stopped. This is the main idea of the hybrid quantum-classical algorithm.

2.2 Barren plateau phenomenon

When the parameterized gates are of the form

$$U_j(\theta_j) = e^{-i\frac{\theta}{2}H_j},$$

where H_j satisfies $H_j^2 = I$, the gradient $\frac{\partial \langle H \rangle}{\partial \theta_j}$ can be estimated by the parameter shifting rule without changing the structure of the PQC [14]. Once we obtain the gradient, we can use gradient-based optimization methods, such as gradient descent, to optimize the parameters.

Ideally, if the gradient does not vanish too fast as the size of the PQC grows, then the gradient could be estimated efficiently and the PQC could be trained easily. However, the BP phenomenon tells us that in many cases, the gradient vanishes exponentially as the system size grows up. When this happens, the PQC is difficult to be trained. The first rigorous proof of the BP phenomenon is shown below.

Theorem 1 ([19]). *Consider a PQC $U(\vec{\theta}) = V(\theta_M, \dots, \theta_{j+1})U(\theta_j)W(\theta_{j-1}, \dots, \theta_1)$ and a Hamiltonian H . The expectation of gradient is 0 if V and W are 1-design. And the variance of gradient $\text{Var} \left[\frac{\partial \langle H \rangle}{\partial \theta_j} \right]$ vanishes exponentially in terms of the number of qubits if V or W is 2-design.*

Hence, when designing the ansatz PQC for a hybrid quantum-classical algorithm, we should analyze whether there exist BP phenomena in it to ensure that it is trainable.

2.3 The ZX-calculus

We provide a brief introduction to the ZX-calculus. For more details, please refer to [44, 45].

In the ZX-calculus, quantum states and their transformations are represented as ZX-diagrams which consist of two kinds of tensors: *Z-spiders* and *X-spiders*. A Z-spider is denoted as a green node, and an X-spider is denoted as a red node. They can be written explicitly in the Dirac

notation as follows.

$$\begin{aligned}
 m \left\{ \begin{array}{c} \vdots \\ \vdots \\ \vdots \end{array} \right\} \theta \left\{ \begin{array}{c} \vdots \\ \vdots \\ \vdots \end{array} \right\} n &:= \underbrace{|0 \dots 0\rangle}_n \underbrace{\langle 0 \dots 0|}_m + e^{i\theta} \underbrace{|1 \dots 1\rangle}_n \underbrace{\langle 1 \dots 1|}_m \\
 m \left\{ \begin{array}{c} \vdots \\ \vdots \\ \vdots \end{array} \right\} \theta \left\{ \begin{array}{c} \vdots \\ \vdots \\ \vdots \end{array} \right\} n &:= \underbrace{|+\dots+\rangle}_n \underbrace{\langle +\dots+|}_m + e^{i\theta} \underbrace{|-\dots-\rangle}_n \underbrace{\langle -\dots-|}_m
 \end{aligned} \tag{4}$$

For a spider, the edges on the left-hand side are called *input* and the edges on the right-hand side are called *output*. The angle θ is called the *phase* of the spider. For simplicity, we will omit the phase when it is zero. Spiders can be connected with wires. Hence, ZX-diagrams can be regarded as tensor networks generated with Z-spiders and X-spiders. For example, we can use ZX-diagrams to represent the following quantum states and quantum gates.

$$\begin{aligned}
 \text{○} &= |0\rangle + |1\rangle = \sqrt{2}|+\rangle & \text{■} &:= \text{○} \begin{array}{c} \frac{\pi}{2} \\ \frac{\pi}{2} \\ \frac{\pi}{2} \end{array} = \text{□} H \\
 \text{●} &= |+\rangle + |-\rangle = \sqrt{2}|0\rangle & \text{○} &= \frac{1}{\sqrt{2}} \begin{array}{c} \text{○} \\ \text{●} \end{array} \\
 \text{○} \theta &= \text{□} R_Z(\theta) & \text{○} &= \frac{1}{\sqrt{2}} \begin{array}{c} \text{○} \\ \text{■} \\ \text{○} \end{array} \\
 \text{●} \theta &= \text{□} R_X(\theta) & \text{○} &= \text{□} Z
 \end{aligned} \tag{5}$$

Here we introduce a new notation, the yellow box, to represent the Hadamard gate

$$H = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}.$$

Since the gates set $\{R_Z, R_X, H, \text{CNOT}\}$ is universal for quantum computing, in principle, one can convert every quantum circuit to a ZX-diagram with the equations in Eq. (5).

Moreover, the ZX-calculus is a powerful tool for reasoning. There are several rewriting rules in the ZX-calculus with which one can rewrite a ZX-diagram to another equivalent form. Figure 2 gives some basic rewriting rules¹ in the ZX-calculus. Here, two ZX-diagrams A, B are said to be equivalent if and only if there exists a non-zero constant $c \in \mathbb{C}$, such that $[A] = c \cdot [B]$, where $[A]$ is the matrix corresponding to the ZX-diagram A .

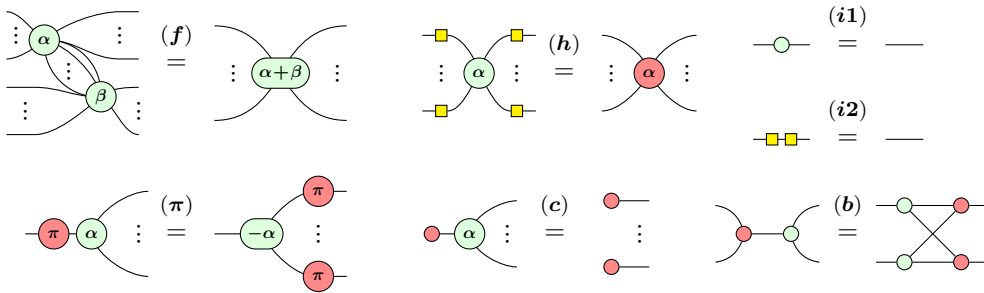


Figure 2: Some basic rewriting rules in the ZX-calculus. Here, “...” means 0 or more (figure from [30]).

Note that the ZX-calculus is *universal*. It means that any linear transformations can be represented as ZX-diagrams. Moreover, the rules in Figure 2 are *complete* for the stabilizer quantum mechanics where phases can only be multiples of $\frac{\pi}{2}$ [46, 47]. That is, if two ZX-diagrams are equivalent, then there exists a set of rewriting rules in Figure 2 that rewrites one into another.

¹(**f**) is for the “fusion” rule; (**h**) is for the “Hadamard color changing” rule; (**i1**) and (**i2**) are “identity” rules; (**pi**) is for the “ π -copy” rule; (**c**) is for the “copy” rule; (**b**) is for the “bi-algebra” rule.

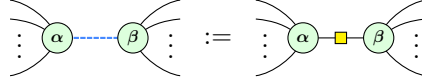
There are also completeness results for the Clifford+T quantum mechanics, where phases can be multiples of $\frac{\pi}{4}$, and for arbitrary ZX-diagrams [48–52].

In this paper, we will focus on a canonical form of the ZX-diagram, the *graph-like ZX-diagram* which is defined in [30].

Definition 1 ([30]). *A ZX-diagram is called graph-like if*

1. All spiders are Z-spiders.
2. Z-spiders are only connected via Hadamard edges.
3. There exist no parallel Hadamard edges or self-loops.
4. Every input or output is connected to a Z-spider and every Z-spider is connected to at most one input or output.

Two spiders being connected via a Hadamard edge means that they are connected with a Hadamard box. Alternatively, we will also use the dashed blue edge to represent a Hadamard edge.



All X-spiders can be rewritten to Z-spiders by using the rule (**h**) in Figure 2. Connected Hadamard boxes can be canceled with the rule (**i2**) and normal edges can be canceled with the rule (**f**). Furthermore, parallel Hadamard edges and self-loops can be canceled with rules² in Figure 3. Hence, every ZX-diagram is equivalent to a graph-like ZX-diagram [30].



Figure 3: Rules for canceling parallel edges and self-loops [30].

3 Analyzing the BP phenomenon with the ZX-calculus

In this section, we will show how to analyze the BP phenomenon with the ZX-calculus. More precisely, we will show how to estimate the expectation and the variance of the gradient of the cost function of a PQC with respect to a Hamiltonian with the ZX-calculus. The main technique we used is to compute integration over unitarians with the ZX-calculus.

Scalars are ignored in the rules in Section 2.3. However, to consider the BP phenomenon, the scalar is necessary. Hence, by using the definition of the Z-spider and the X-spider in Eq. 4, we can obtain the precise rules with scalars in Figure 4.

In this paper, we consider PQCs under the following assumptions.

Assumption 1. *The PQC $U(\vec{\theta})$ satisfies*

1. Each gate in U is one of $\{R_X, R_Z, H, \text{CNOT}\}$.
2. The parameters in $\vec{\theta} = (\theta_1, \dots, \theta_m)$ are independent uniform random variables in the interval $[-\pi, \pi]$.

²(**hopf**) is for the ‘‘Hopf’’ rule; (**sl**) is for the ‘‘self-loop canceling’’ rule; (**hsl**) is for the ‘‘Hadamard self-loop canceling’’ rule.

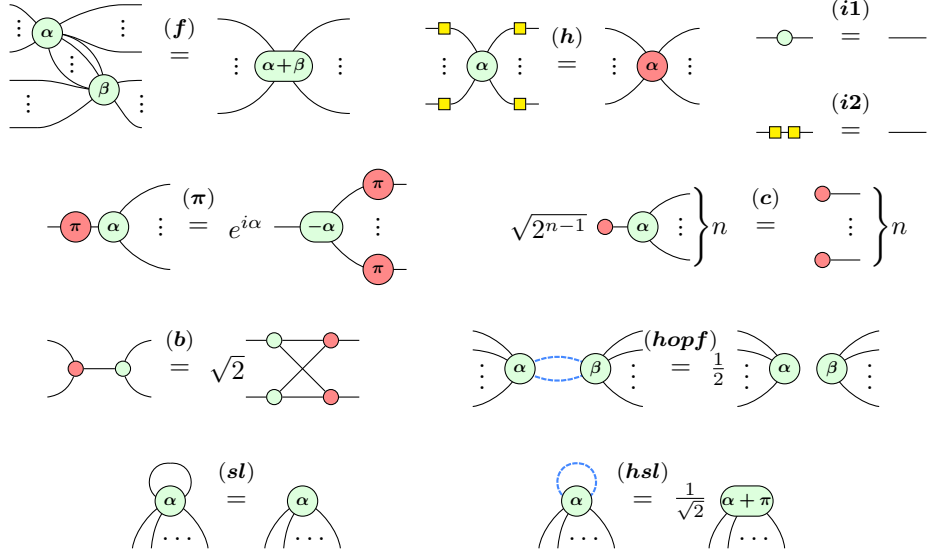


Figure 4: Rewriting rules with scalars.

We remark that in the case that a quantum circuit contains gates not satisfying Assumption 1, if one can represent these gates as composition of gates which satisfy Assumption 1, then the results of this paper still hold. For example, we will first represent R_Y using R_Z and R_X in Section 4.2.

3.1 Representing gradients as ZX-diagrams

Consider a PQC $U(\vec{\theta})$ of n -qubits and a Hamiltonian H . We assume that the input state is pure. Without loss of generality, we can also assume that we apply this PQC to an initial state $|0\rangle$. Then the expectation of H can be expressed as

$$\langle H \rangle = \langle 0 | U^\dagger(\vec{\theta}) H U(\vec{\theta}) | 0 \rangle. \quad (6)$$

As shown in Section 2.3, we can convert the PQC $U(\vec{\theta})$ to a parameterized graph-like ZX-diagram $G_U(\vec{\theta})$ with Eq. (5). Suppose that

$$U(\vec{\theta}) = c \cdot [G_U(\vec{\theta})]$$

for a constant c , then $\langle H \rangle$ can also be expressed as a ZX-diagram as demonstrated in the following equation. Here, $U(\vec{\theta})$ is under the Assumption 1.

$$\langle H \rangle = \frac{|c|^2}{2^n} \cdot \left(G_U(\vec{\theta}) \right) H \left(G_{U^\dagger}(-\vec{\theta}) \right) \quad (7)$$

We remark that in the general case, the input state can be a mixed state ρ . Because the ZX-calculus is universal, we can represent ρ as a ZX-diagram D_ρ . Then by replacing the X-spiders representing zero states on the left- and right-hand sides in Eq. (7) with D_ρ , we can still obtain a ZX-diagram representing the expectation $\langle H \rangle = \text{Tr}(\rho U^\dagger H U)$. And results in this paper still hold in this case.

If we expand the spider by the definition of the Z-spider, we can prove that the gradient $\frac{\partial \langle H \rangle}{\partial \theta_j}$ can be represented as a ZX-diagram.

Theorem 2. *The gradient can be represented as the following equation.*

$$\frac{2^n}{|c|^2} \cdot \frac{\partial \langle H \rangle}{\partial \theta_j} = \frac{\partial}{\partial \theta_j}$$

Proof. We expand the corresponding Z-spiders according to the definition. There will be four terms on the left-hand side. Two of these terms are constants, and thus they will become 0 after taking the derivative. According to the definition of the X-spider, we can obtain the ZX-diagram on the right-hand side. The complete proof is given in Appendix B. \square

This theorem also gives a graphical proof of the parameter-shift rule in [14]. We will demonstrate it with the following example. Consider the following ansatz. We first can convert it to an equivalent ZX-diagram.

$$\begin{aligned} &|0\rangle \text{---} R_X(\theta_1) \text{---} \bullet \text{---} R_Z(\theta_3) \text{---} \\ &|0\rangle \text{---} R_X(\theta_2) \text{---} \oplus \text{---} R_Z(\theta_4) \text{---} \end{aligned} = \frac{1}{\sqrt{2}} \begin{matrix} \bullet & \theta_1 & \bullet & \theta_3 \\ \bullet & \theta_2 & \bullet & \theta_4 \end{matrix} \begin{matrix} (f), (h) \\ \bullet & \theta_1 & \bullet & \theta_3 \\ \bullet & \theta_2 & \bullet & \theta_4 \end{matrix} = \frac{1}{\sqrt{2}} \begin{matrix} \theta_1 & \theta_3 \\ \theta_2 & \theta_4 \end{matrix}$$

And the expectation $\langle H \rangle$ of a Hamiltonian H can be represented as the following ZX-diagram.

$$\langle H \rangle = \begin{matrix} |0\rangle \text{---} R_X(\theta_1) \text{---} \bullet \text{---} R_Z(\theta_3) \text{---} H \text{---} R_Z(-\theta_3) \text{---} \bullet \text{---} R_X(-\theta_1) \text{---} \langle 0| \\ |0\rangle \text{---} R_X(\theta_2) \text{---} \oplus \text{---} R_Z(\theta_4) \text{---} H \text{---} R_Z(-\theta_4) \text{---} \oplus \text{---} R_X(-\theta_2) \text{---} \langle 0| \end{matrix} = \frac{1}{\sqrt{2}} \begin{matrix} \theta_1 & \theta_3 \\ \theta_1 & \theta_4 \end{matrix} H \begin{matrix} -\theta_3 & -\theta_1 \\ -\theta_4 & -\theta_2 \end{matrix} \frac{1}{\sqrt{2}}$$

Then by Theorem 2, we can obtain the gradient directly.

$$\frac{\partial \langle H \rangle}{\partial \theta_1} = \frac{1}{\sqrt{2}} \begin{matrix} \theta_1 + \frac{\pi}{2} & \theta_3 \\ \theta_2 & \theta_4 \end{matrix} H \begin{matrix} -\theta_3 & -\theta_1 - \frac{\pi}{2} \\ -\theta_4 & -\theta_2 \end{matrix} \frac{1}{\sqrt{2}}$$

Then we can use the definition of the X-spider to obtain the parameter-shift rule as shown below.

$$\begin{aligned}
& \frac{1}{\sqrt{2}} \begin{array}{c} \textcircled{\pi} \\ \textcircled{\theta_1 + \frac{\pi}{2}} \quad \textcircled{\theta_3} \\ \textcircled{\theta_2} \quad \textcircled{\theta_4} \\ \textcircled{-\theta_3} \quad \textcircled{-\theta_1 - \frac{\pi}{2}} \\ \textcircled{-\theta_4} \quad \textcircled{-\theta_2} \end{array} \frac{1}{\sqrt{2}} \\
&= \frac{1}{\sqrt{2}} \begin{array}{c} \boxed{|+\rangle \langle +| - |-\rangle \langle -|} \\ \textcircled{\theta_1 + \frac{\pi}{2}} \quad \textcircled{\theta_3} \\ \textcircled{\theta_2} \quad \textcircled{\theta_4} \\ \textcircled{-\theta_3} \quad \textcircled{-\theta_1 - \frac{\pi}{2}} \\ \textcircled{-\theta_4} \quad \textcircled{-\theta_2} \end{array} \frac{1}{\sqrt{2}} \\
&= \frac{1}{\sqrt{2}} \begin{array}{c} \textcircled{\theta_1 + \frac{\pi}{2}} \quad \textcircled{\theta_3} \\ \textcircled{\theta_2} \quad \textcircled{\theta_4} \\ \textcircled{-\theta_3} \quad \textcircled{-\theta_1 - \frac{\pi}{2}} \\ \textcircled{-\theta_4} \quad \textcircled{-\theta_2} \end{array} \frac{1}{\sqrt{2}} - \frac{1}{\sqrt{2}} \begin{array}{c} \textcircled{\theta_1 - \frac{\pi}{2}} \quad \textcircled{\theta_3} \\ \textcircled{\theta_2} \quad \textcircled{\theta_4} \\ \textcircled{-\theta_3} \quad \textcircled{-\theta_1 + \frac{\pi}{2}} \\ \textcircled{-\theta_4} \quad \textcircled{-\theta_2} \end{array} \frac{1}{\sqrt{2}} \\
&\stackrel{(f)}{=} \frac{1}{2} \left\{ \begin{array}{c} \textcircled{\theta_1 + \frac{\pi}{2}} \quad \textcircled{\theta_3} \\ \textcircled{\theta_2} \quad \textcircled{\theta_4} \\ \textcircled{-\theta_3} \quad \textcircled{-\theta_1 - \frac{\pi}{2}} \\ \textcircled{-\theta_4} \quad \textcircled{-\theta_2} \end{array} \frac{1}{\sqrt{2}} - \begin{array}{c} \textcircled{\theta_1 - \frac{\pi}{2}} \quad \textcircled{\theta_3} \\ \textcircled{\theta_2} \quad \textcircled{\theta_4} \\ \textcircled{-\theta_3} \quad \textcircled{-\theta_1 + \frac{\pi}{2}} \\ \textcircled{-\theta_4} \quad \textcircled{-\theta_2} \end{array} \frac{1}{\sqrt{2}} \right\} \\
&= \frac{1}{2} (\langle H \rangle_{\theta_1, +} - \langle H \rangle_{\theta_1, -})
\end{aligned}$$

Here $\langle H \rangle_{\theta_1, \pm}$ means replacing the parameter θ_1 with $\theta_1 \pm \frac{\pi}{2}$ in the ansatz.

To analyze the BP phenomenon, we need to compute the expectation

$$\mathbf{E}\left(\frac{\partial \langle H \rangle}{\partial \theta_j}\right) = \int_{\vec{\theta}} p(\vec{\theta}) \frac{\partial \langle H \rangle}{\partial \theta_j} d\vec{\theta}, \quad (8)$$

and the variance

$$\mathbf{Var}\left(\frac{\partial \langle H \rangle}{\partial \theta_j}\right) = \mathbf{E}\left(\left|\frac{\partial \langle H \rangle}{\partial \theta_j}\right|^2\right) - \left(\mathbf{E}\left(\frac{\partial \langle H \rangle}{\partial \theta_j}\right)\right)^2 = \int_{\vec{\theta}} p(\vec{\theta}) \left|\frac{\partial \langle H \rangle}{\partial \theta_j}\right|^2 d\vec{\theta} - \left(\mathbf{E}\left(\frac{\partial \langle H \rangle}{\partial \theta_j}\right)\right)^2, \quad (9)$$

for $j = 1, \dots, m$. Here $p(\vec{\theta})$ is the probability of the parameters $\vec{\theta}$.

By Assumption 1, Eq. (8) and Eq. (9) can be written as

$$\mathbf{E}\left(\frac{\partial \langle H \rangle}{\partial \theta_j}\right) = \frac{1}{(2\pi)^m} \int_{\theta_1} \dots \int_{\theta_m} \frac{\partial \langle H \rangle}{\partial \theta_j} d\theta_1 \dots d\theta_m, \quad (10)$$

and

$$\mathbf{Var}\left(\frac{\partial \langle H \rangle}{\partial \theta_j}\right) = \frac{1}{(2\pi)^m} \int_{\theta_1} \dots \int_{\theta_m} \left|\frac{\partial \langle H \rangle}{\partial \theta_j}\right|^2 d\theta_1 \dots d\theta_m - \left(\mathbf{E}\left(\frac{\partial \langle H \rangle}{\partial \theta_j}\right)\right)^2, \quad (11)$$

for $j = 1, \dots, m$.

We will compute the expectation and variance of the gradients in the next two sections.

3.2 The expectation of gradients

In this section, we will compute the expectation in Eq. (10). As shown in Theorem 2, the integration

$$\frac{1}{2\pi} \int_{\theta_k} \frac{\partial \langle H \rangle}{\partial \theta_j} d\theta_k,$$

for $k = 1, \dots, m$, is also an integration of a ZX-diagram over its parameter θ_k . With the following lemma, the integration can be represented as a ZX-diagram again.

Lemma 1. *The following equation holds.*

$$\frac{1}{2\pi} \int_{\alpha} m \left\{ \begin{array}{c} \text{Z-spider with } \alpha \\ \vdots \\ \text{Z-spider with } \alpha \end{array} \right\} n \dots n \left\{ \begin{array}{c} \text{Z-spider with } -\alpha \\ \vdots \\ \text{Z-spider with } -\alpha \end{array} \right\} m \, d\alpha = m \left\{ \begin{array}{c} \text{Z-spider with } \alpha \\ \vdots \\ \text{Z-spider with } \alpha \end{array} \right\} n \dots n \left\{ \begin{array}{c} \text{Z-spider with } \alpha \\ \vdots \\ \text{Z-spider with } \alpha \end{array} \right\} m$$

Proof. Expand the Z-spiders according to its definition. There will be four terms on the left-hand side. By using

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} e^{i\alpha} d\alpha = 0,$$

only two terms left. We will use the definition of the Z-spider again to obtain the ZX-diagram on the right-hand side. The complete proof is given in Appendix C. \square

With this lemma, we give a graphical proof of the following theorem, which is also proved in [20].

Theorem 3. *Under Assumption 1, the integration*

$$\frac{1}{2\pi} \int_{\theta_j} \frac{\partial \langle H \rangle}{\partial \theta_j} d\theta_j = 0.$$

Proof. Using the relation in Lemma 1 on the ZX-diagram in Theorem 2, we have

$$\begin{aligned} \frac{2^n}{|c|^2} \cdot \frac{1}{2\pi} \int_{\theta_j} \frac{\partial \langle H \rangle}{\partial \theta_j} d\theta_j &= \text{ZX-diagram with } H \text{ and } \pi \text{ spider} \\ &= \text{ZX-diagram with } H \text{ and } \pi \text{ spider} \\ &= \text{ZX-diagram with } H \text{ and } \frac{1}{2}\pi \text{ spider} \\ &= 0. \end{aligned}$$

\square

As a corollary, the expectation of the gradient in Eq. (10) is zero.

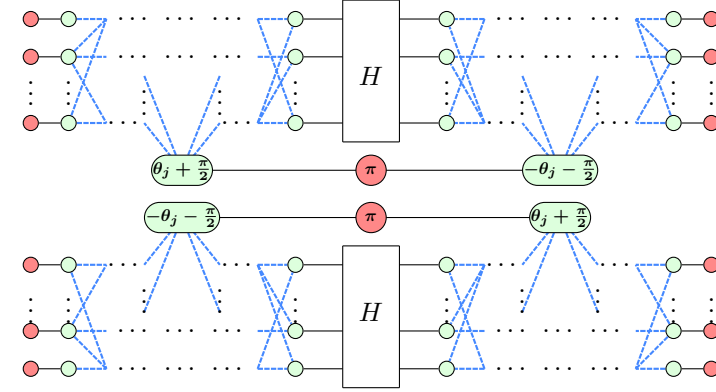
Corollary 1. *Under Assumption 1, the expectation*

$$\mathbf{E}\left(\frac{\partial \langle H \rangle}{\partial \theta_j}\right) = \frac{1}{(2\pi)^m} \int_{\theta_1} \dots \int_{\theta_m} \frac{\partial \langle H \rangle}{\partial \theta_j} d\theta_1 \dots d\theta_m = 0, \text{ for } j = 1, \dots, m.$$

3.3 The variance of gradients

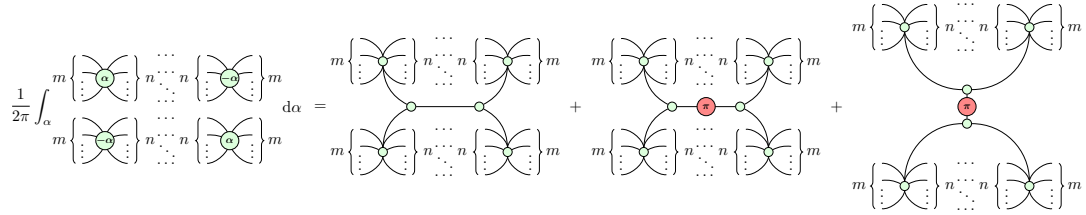
In this section, we will compute the variance in Eq. (9).

Because the gradient $\frac{\partial \langle H \rangle}{\partial \theta_j}$ is a real number and the expectation is 0, by Eq. (9), the variance is the expectation of $\left(\frac{\partial \langle H \rangle}{\partial \theta_j}\right)^2$, which can be represented as follows by Theorem 2.

$$\frac{4^n}{|c|^2} \cdot \left[\frac{\partial \langle H \rangle}{\partial \theta_j} \right]^2 =$$

(12)

Similar to Lemma 1, we can prove the following lemma.

Lemma 2. *The following equation holds.*

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} e^{i\alpha} d\alpha =$$


Proof. We expand all Z-spiders and there will be 16 terms on the left-hand side. Again, using the relation

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} e^{i\alpha} d\alpha = 0,$$

6 terms was left. We can use the definition of the Z-spider again to obtain the right-hand side. The complete proof is given in Appendix D. \square

There exist three terms after integration. Hence, computing the variance of gradients is much more complicated than computing the expectation. We denote the three ZX-diagrams in Lemma 2 as

$$T_1 = \text{Z-spider with two green circles}, \quad T_2 = \text{Z-spider with a red circle in the middle}, \quad T_3 = \text{Z-spider with a red circle on the top wire}.$$
(13)

And we introduce a new notation

$$V_U^{a_1, \dots, a_m}, \quad a_j \in \{T_1, T_2, T_3\},$$

to represent the following ZX-diagram.

$$V_U^{a_1, \dots, a_m} = \text{Diagram} \quad (14)$$

Here, $U(\theta_1, \dots, \theta_m)$ is a PQC with m parameters and G_U is the graph-like ZX-diagram corresponding to U . With this notation, we have the following theorem.

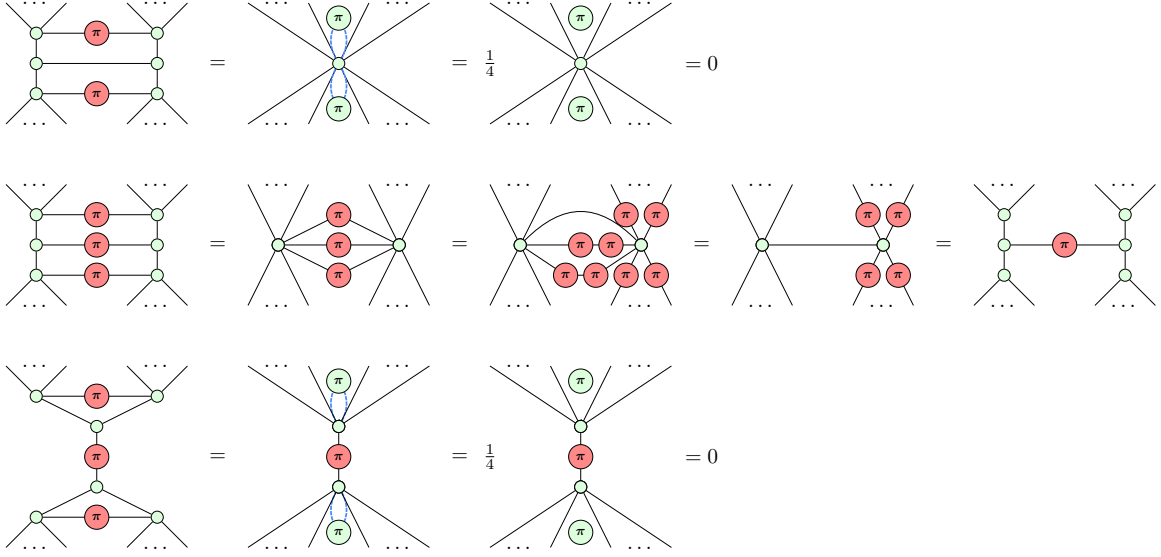
Theorem 4. *Under Assumption 1, the following equation holds.*

$$\text{Var} \left(\frac{\partial \langle H \rangle}{\partial \theta_j} \right) = \frac{|c|^2}{4^n} \cdot \sum_{a_k \in \{T_1, T_2, T_3\}, k \neq j} V_U^{a_1, \dots, a_{j-1}, T_2, a_{j+1}, \dots, a_m}.$$

Proof. By Eq. (12) and Lemma 2, we have the following equation.

$$\frac{4^n}{|c|^2} \cdot \frac{1}{2\pi} \int_{\theta_j} \left[\frac{\partial \langle H \rangle}{\partial \theta_j} \right]^2 d\theta_j = \sum_{a_j \in \{T_1, T_2, T_3\}} \text{Diagram}$$

And by the following equations,



we have

$$\frac{4^n}{|c|^2} \cdot \frac{1}{2\pi} \int_{\theta_j} \left[\frac{\partial \langle H \rangle}{\partial \theta_j} \right]^2 d\theta_j =$$

Then, by the definition of $V_U^{a_1, \dots, a_m}$, we obtain

$$\text{Var} \left(\frac{\partial \langle H \rangle}{\partial \theta_j} \right) = \frac{|c|^2}{4^n} \cdot \sum_{a_k \in \{T_1, T_2, T_3\}, k \neq j} V_U^{a_1, \dots, a_{j-1}, T_2, a_{j+1}, \dots, a_m}.$$

□

Hence, to compute the variance, we need to sum over 3^{m-1} terms of the tensor

$$V_U^{a_1, \dots, a_{j-1}, T_2, a_{j+1}, \dots, a_m}.$$

It seems inaccessible when m is large. But in many cases, we have simple ways to compute this sum.

Recall that we have converted quantum circuits to graph-like ZX-diagrams. Hence, spiders are connected with Hadamard edges. Let us consider two spiders W_j, W_k corresponding to the parameters θ_j, θ_k in G_U . Suppose that W_j and W_k are connected with a Hadamard edge. Then by the following lemma, the Hadamard edge can be removed after integration over θ_j and θ_k .

the input state, while the tensor $\tilde{H}_{c_1, \dots, c_m}$ is related to the Hamiltonian H . And P_2 is a projection that has only one non-zero entry. That is

$$P_2(2, 2, \dots, 2) = 1.$$

Also, note that there is a scalar 2 for each internal copy tensor. This scalar comes from the following equation.

$$\text{Diagrammatic equation (15)} \quad (15)$$

In Appendix A, a simple example is given to illustrate the techniques introduced in this paper.

In conclusion, computing the variance of gradients is reduced to contracting a tensor network corresponding to the circuit. In the next section, we will use these techniques to analyze the BP phenomenon for several for concrete PQCs.

4 Analyzing the BP phenomenon for four PQCs

In this section, we will analyze the BP phenomenon for four PQCs with the techniques introduced in Section 3.

4.1 Hardware-efficient ansatz

Consider a hardware-efficient ansatz [2] PQC of the following form.

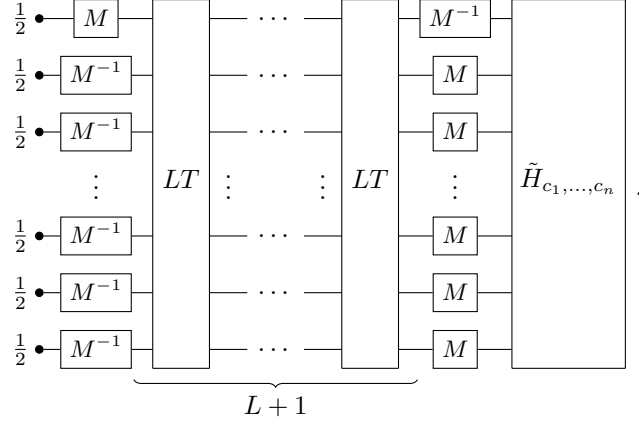
$$\text{Quantum circuit diagram (16)} \quad (16)$$

The first step is converting it to a graph-like ZX-diagram. Suppose the circuit is of n -qubits. By the conversion rules in Eq. (5) and rewriting rules in Figure 4, we obtain a graph-like ZX-diagram where most spiders are parameterized.

$$\text{Graph-like ZX-diagram} \quad \times (L + 1)$$

Here, the Z-spider with “...” represents a Z-spider with a parameter.

And the whole tensor network in Eq. (17) is



Hence, the variance will be

$$\text{Var} \left(\frac{\partial \langle H \rangle}{\partial \theta_j} \right) = \text{Diagram} \quad (18)$$

The diagram shows a tensor network for the variance. It has L_1 layers of LT on the left, followed by a central layer with boxes M, P_2, and M^{-1} on the top wire, and M^{-1} on the other wires. This central layer is labeled 'corresponding to \theta_j'. This is followed by L_2 layers of LT on the right, and finally a box labeled \tilde{H}_{c_1, \dots, c_n}.

We can prove that only two eigenvalues of the matrix LT are 1 and the norms of other eigenvalues are less than 1 (for the complete proof, please refer to Appendix G). Moreover, the eigenspace corresponding to the eigenvalue 1 is generated with two vectors

$$E_1 = \text{span}\{v_{1,2} \otimes \dots \otimes v_{1,2}, v_{1,3} \otimes \dots \otimes v_{1,3}\}, \quad v_{1,2} = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \quad v_{1,3} = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}. \quad (19)$$

Hence, LT^d will converge to P_{E_1} , the projection to the eigenspace E_1 , exponentially, as $d \rightarrow \infty$.

If we replace LT^{L_1} and LT^{L_2} with the projection P_{E_1} , then the Eq. (18) will become

$$\lim_{L_1, L_2 \rightarrow \infty} \text{Var} \left(\frac{\partial \langle H \rangle}{\partial \theta_j} \right) = \text{Diagram} \quad (20)$$

The diagram shows the limit of the variance tensor network. The L_1 and L_2 layers of LT are replaced by projection boxes labeled P_{E_1}. The central layer with M, P_2, and M^{-1} remains, labeled 'corresponding to \theta_j'. The output is \tilde{H}_{c_1, \dots, c_n}.

This term is

$$4 \cdot \frac{1}{4^n} \text{Tr}(H^2), \quad (21)$$

which is exponentially small. That is, the number of qubits n determines an exponential small limitation of the variance in Eq. (21) when the number of layers $L \rightarrow \infty$. And the number of

layers L determines how the variance will be close to the limitation. More precisely, the variance $\text{Var}\left(\frac{\partial\langle H\rangle}{\partial\theta}\right)$ is exponentially (in L) close to the exponential small (in n) value Eq. (21). Thus, there exist BPs in the hardware-efficient ansatz.

Theorem 5. *The variance of gradients in the hardware-efficient ansatz defined in Eq. (16) vanishes exponentially as the number of qubits n and the number of layers L grow up.*

Note that the above analysis can be generalized to any hardware-efficient ansätze if the entangler connects all of the qubits.

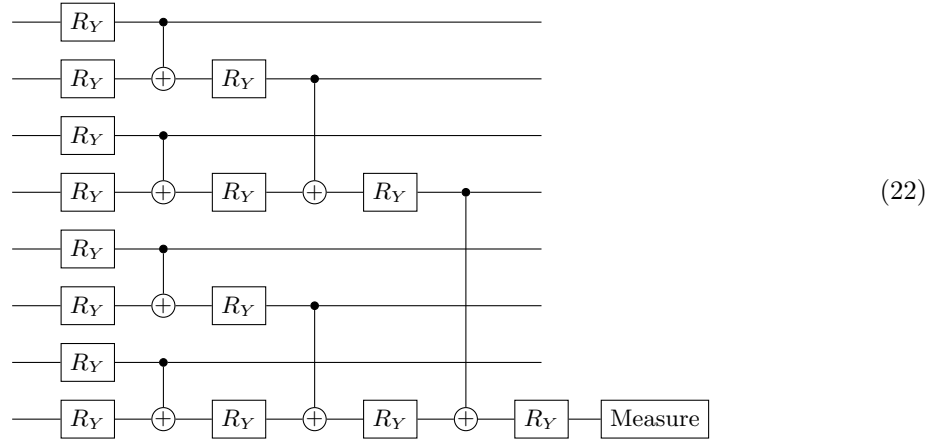
4.2 Tree tensor network ansatz

The tree tensor network is a special kind of tensor network with tree structures. The quantum analog of the tree tensor network was developed in [42]. In [53], it was proved that the sum of the variance

$$\sum_{j=1}^n \text{Var}\left(\frac{\partial\langle H\rangle}{\partial\theta_j}\right)$$

will not vanish exponentially. In this section, we will prove that not only the sum of the variance but also the variance of each parameter vanishes polynomially.

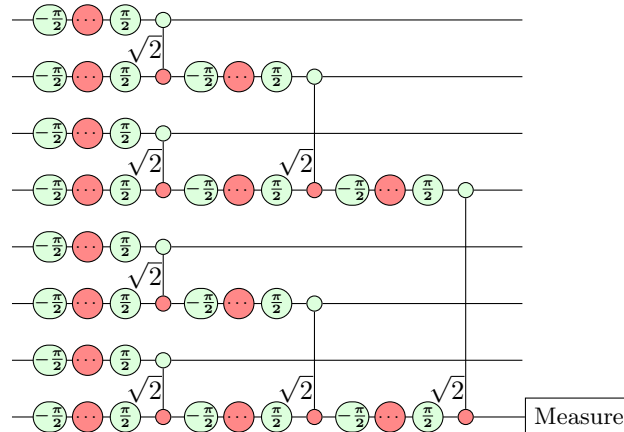
Consider the tree tensor network ansatz with n -qubit of the following form.



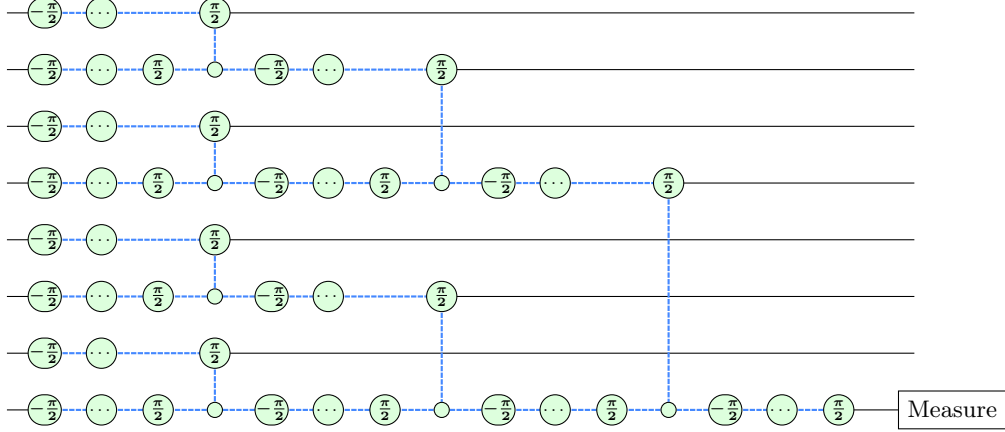
To analyze the BP phenomenon of this ansatz, we first use the gate decomposition

$$R_Y(\theta) = R_Z\left(\frac{\pi}{2}\right)R_X(\theta)R_Z\left(-\frac{\pi}{2}\right)$$

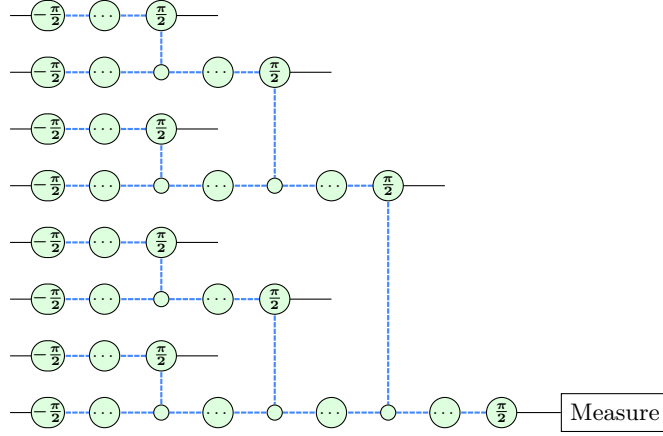
to convert the PQC to a ZX-diagram as follows.



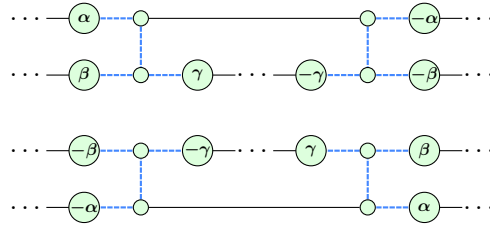
The X-spiders with phase “...” are spiders with parameters. And the ZX-diagram can be rewritten to a graph-like ZX-diagram as follows.



By using the rewriting rule (1c) in [30], we can remove the spiders with phases $\pm \frac{\pi}{2}$.



Now we are going to construct a tensor network from this graph-like ZX-diagram to represent the variance. By Eq. (12), the building block of the variance $\frac{\partial \langle H \rangle^2}{\partial \theta_j}$ is



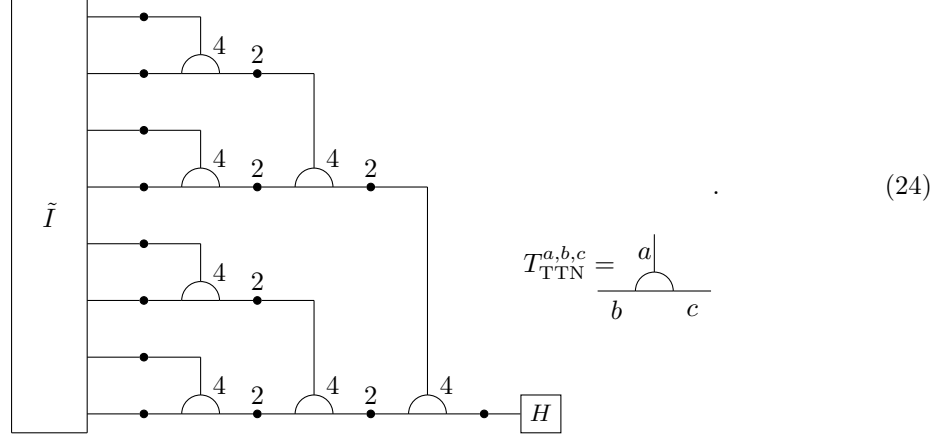
We can prove that (for the complete proof, please refer to Appendix G), after integration over the parameters α, β, γ , the building block will become

$$\frac{1}{(2\pi)^3} \int_{\alpha, \beta, \gamma} d\alpha d\beta d\gamma = \sum_{a, b, c \in T_1, T_2, T_3} T_{\text{TTN}}^{a, b, c} \quad (23)$$

Here, T_{TTN} is a $3 \times 3 \times 3$ tensor defined as follows.

$$T_{\text{TTN}}[1, \cdot, \cdot] = \frac{1}{16} \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}, \quad T_{\text{TTN}}[2, \cdot, \cdot] = \frac{1}{16} \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & 0 \\ -1 & 0 & 1 \end{pmatrix}, \quad T_{\text{TTN}}[3, \cdot, \cdot] = \frac{1}{16} \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}.$$

Hence, the variance of $\frac{\partial \langle H \rangle}{\partial \theta_j}$ can be obtained by replacing one of the copy tensors with the projection P_2 in the following tensor network.



Now let us analyze the scaling property of this tensor network. Since the Hamiltonian H is a 1-qubit Hermitian operator, it can be expressed as

$$H = k_0 I + k_1 X + k_2 Y + k_3 Z, \quad k_j \in \mathbb{R}.$$

Then

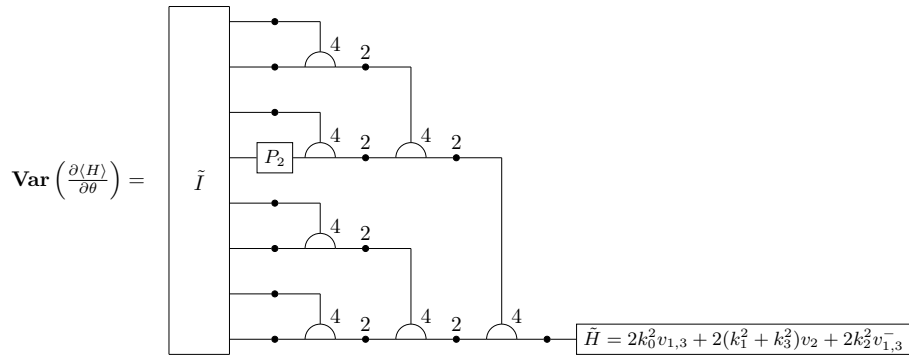
$$\tilde{H} = 2k_0^2 \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} + 2(k_1^2 + k_3^2) \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} + 2k_2^2 \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}.$$

We denote

$$v_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \quad v_{1,3} = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \quad v_{1,3}^- = \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}.$$

Note that the building block of this tensor network is $8 \cdot T_{\text{TTN}}$. By the definition of T_{TTN} , we have

With the above equations, we can compute the variance simply. For example, consider the following variance.

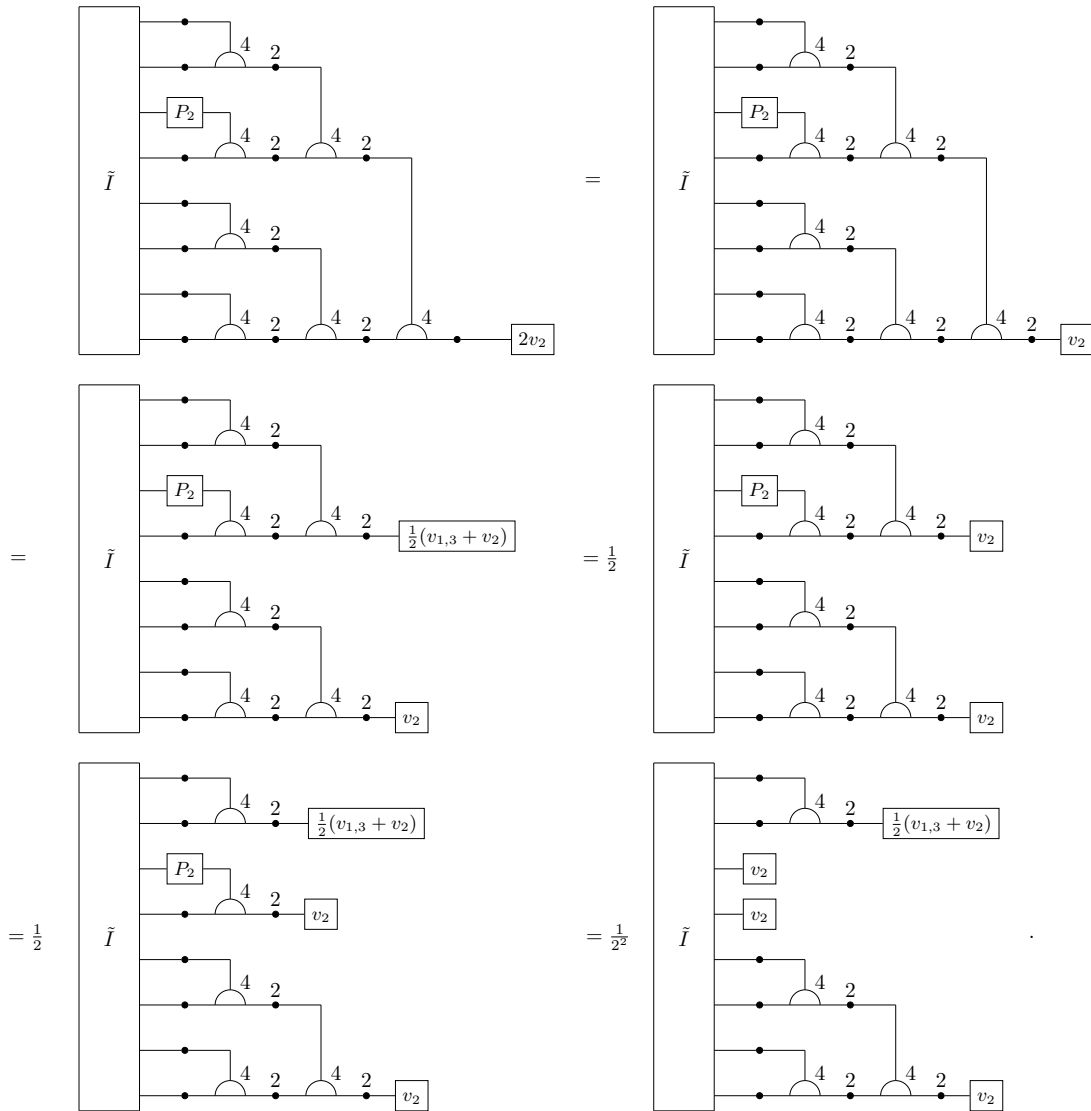


It is a linear function of \tilde{H} . Hence, we can analyze each term of \tilde{H} individually.

Since $P_2 v_{1,3} = 0$, the first term $2k_0^2 v_{1,3}$ in \tilde{H} will become 0.

Now let us consider the second term $2(k_1^2 + k_3^2) v_2$. With Eq. (25), we can expand the variance

as follows.

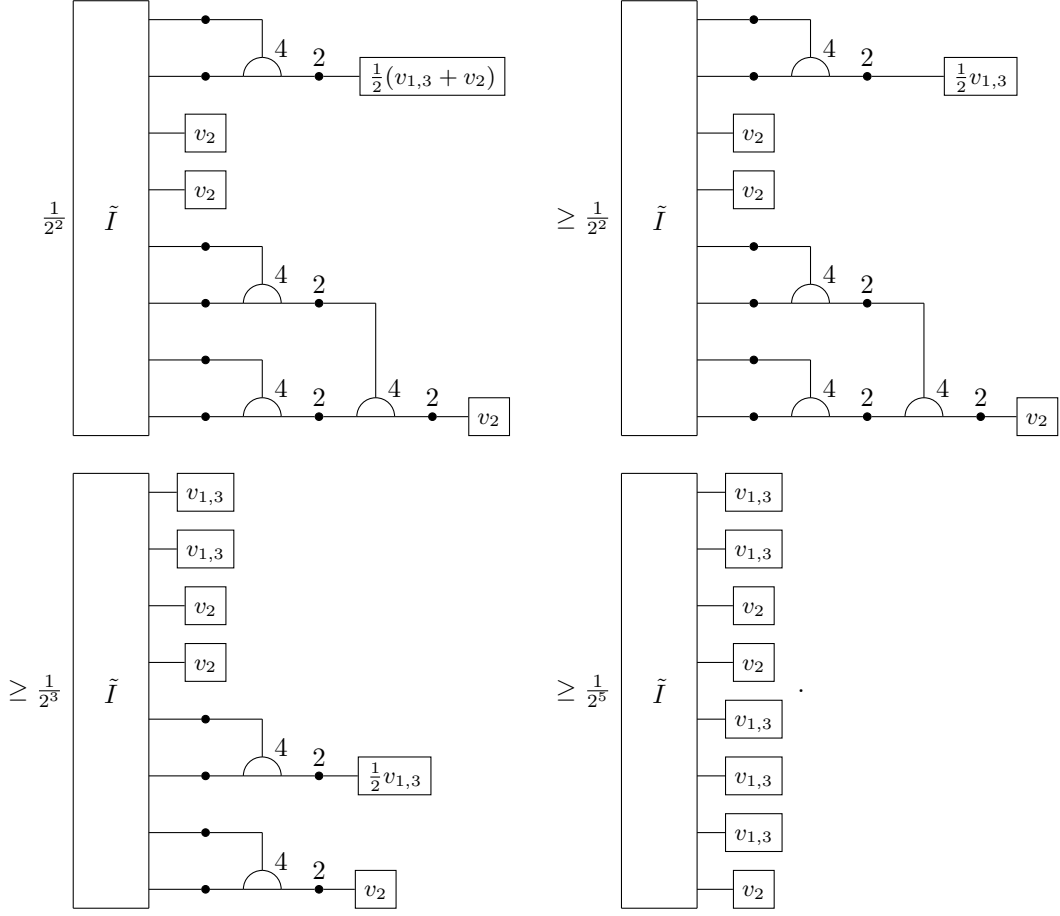


Expanding it recursively, the variance can be represented as a summation of terms of the following form,

$$\tilde{I}(u_1, \dots, u_n) = \tilde{I} \begin{matrix} \text{---} u_1 \\ \text{---} u_2 \\ \vdots \\ \text{---} u_n \end{matrix}, \text{ where } u_j \in \{v_2, v_{1,3}\}. \quad (26)$$

And by the definition of \tilde{I} , each of the term $\tilde{I}(u_1, \dots, u_n) \geq 0$. Hence, we can obtain a lower

bound,



Similarly, we have a lower bound for the term $v_{1,3}^-$.

For the general case of n -qubit, we can prove that it has a lower bound.

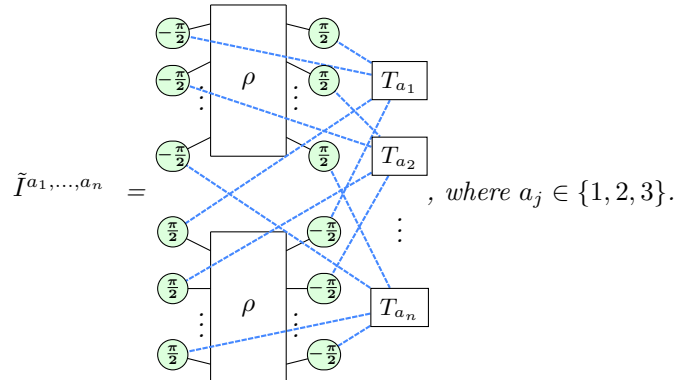
Theorem 6. For tree tensor network ansatz shown in (22), if

$$H = k_0 I + k_1 X + k_2 Y + k_3 Z,$$

then we have

$$\text{Var} \left(\frac{\partial \langle H \rangle}{\partial \theta} \right) \geq \frac{k_1^2 + k_3^2}{n^2} \tilde{I}(u_1, \dots, u_n) + \frac{k_2^2}{n^2} \tilde{I}(w_1, \dots, w_n), \quad (27)$$

for some $u_j, w_j \in \{v_{1,3}, v_2, v_{1,3}^-\}$. Here $\tilde{I}(u_1, \dots, u_n)$ is defined in Eq. (26). And \tilde{I} is a 3^n dimensional tensor which only depends on the input state. If the input state is ρ , then \tilde{I} is defined as follows.



Proof. See Appendix G.2. □

Note that $\tilde{I}(u_1, \dots, u_n)$ only depends on the input state. If

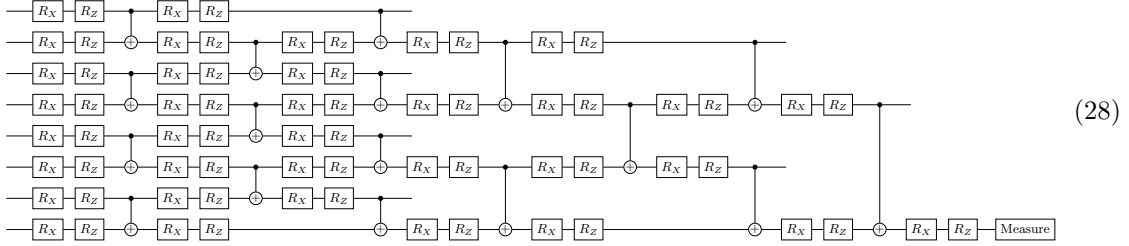
$$\tilde{I}(u_1, \dots, u_n) \in \Omega\left(\frac{1}{\text{poly}(n)}\right) \text{ or } \tilde{I}(w_1, \dots, w_n) \in \Omega\left(\frac{1}{\text{poly}(n)}\right),$$

there exist no BP in the tree tensor network ansatz.

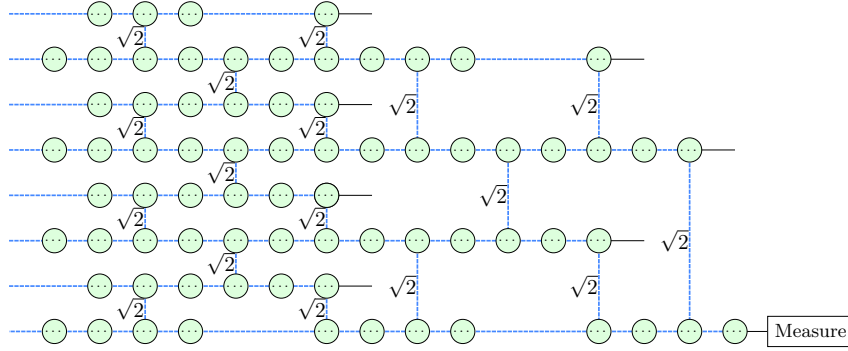
4.3 QCNN

QCNN was developed in [41]. It was proved that there exists no BP in the QCNN ansatz if the subblocks form unitary 2-design [54]. In this section, we will use the ZX-calculus to analyze the BP phenomenon in a QCNN ansatz without the assumption of unitary 2-design.

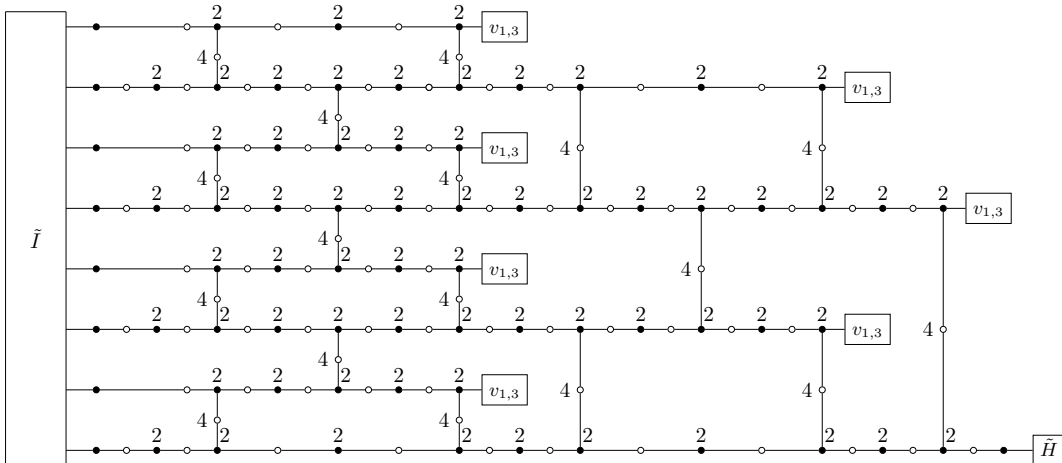
Consider a QCNN ansatz as follows.



It can be represented as the following ZX-diagram.



where the Z-spiders with “...” are parameterized. Note that this is a graph-like ZX-diagram whose spiders are all parameterized. Hence, by using Lemma 3, the variance can be obtained by replacing one of the copy tensors with the projection P_2 in the following tensor network.



Now we are going to analyze the scaling property of this tensor network. If we denote

$$T_{\text{QCNN}} = \begin{array}{c} \text{---} \bullet \text{---} \bullet \text{---} \\ | \quad | \\ \text{---} \bullet \text{---} \bullet \text{---} \end{array},$$

then we have the following equations

$$\begin{array}{c} \text{---} \bullet \text{---} \bullet \text{---} \\ | \quad | \\ \text{---} \bullet \text{---} \bullet \text{---} \end{array} \begin{array}{c} \boxed{v_{1,3}} \\ \boxed{v_{1,3}} \end{array} = \begin{array}{c} \boxed{v_{1,3}} \\ \boxed{v_{1,3}} \end{array}, \quad \begin{array}{c} \text{---} \bullet \text{---} \bullet \text{---} \\ | \quad | \\ \text{---} \bullet \text{---} \bullet \text{---} \end{array} \begin{array}{c} \boxed{v_{1,3}} \\ \boxed{v_2} \end{array} = \begin{array}{c} \boxed{\frac{1}{2}v_{1,3}^- + \frac{1}{2}v_2} \\ \boxed{\frac{1}{4}v_{1,3} + \frac{3}{4}v_2} \end{array}, \quad \begin{array}{c} \text{---} \bullet \text{---} \bullet \text{---} \\ | \quad | \\ \text{---} \bullet \text{---} \bullet \text{---} \end{array} \begin{array}{c} \boxed{v_{1,3}} \\ \boxed{v_{1,3}^-} \end{array} = \begin{array}{c} \boxed{v_{1,3}} \\ \boxed{\frac{1}{2}v_{1,3}^- + \frac{1}{2}v_2} \end{array}. \quad (29)$$

By using Eq. (29), we can expand the variance as a sum of terms of the following form.

$$\tilde{I}(u_1, \dots, u_n) = \begin{array}{c} \boxed{\tilde{I}} \\ | \quad | \quad | \\ \boxed{u_1} \\ \boxed{u_2} \\ \vdots \\ \boxed{u_n} \end{array}, \quad \text{where } u_j \in \{v_2, v_{1,3}, v_{1,3}^-\}.$$

Each of these terms is non-negative. Hence, similar to the analysis of tree tensor network ansatz, we can prove that the variance of gradients in the QCNN ansatz has a lower bound, since the QCNN ansatz is of $O(\log(n))$ -depth.

Theorem 7. *For the QCNN ansatz shown in (28), if*

$$H = k_0 I + k_1 X + k_2 Y + k_3 Z,$$

then we have

$$\text{Var} \left(\frac{\partial \langle H \rangle}{\partial \theta} \right) \geq \frac{k_2^2 + k_3^2}{n^9} \tilde{I}(u_1, \dots, u_n) + \frac{k_1^2}{n^9} \tilde{I}(w_1, \dots, w_n), \quad (30)$$

for some $u_j, w_j \in \{v_{1,3}, v_2, v_{1,3}^-\}$. Here \tilde{I} is a 3^n dimensional tensor which only depends on the input state ρ .

$$\tilde{I}^{a_1, \dots, a_n} = \begin{array}{c} \boxed{\rho} \\ \vdots \\ \vdots \\ \boxed{\rho} \end{array} \begin{array}{c} \vdots \\ \vdots \\ \vdots \\ \vdots \end{array} \begin{array}{c} \boxed{T_{a_1}} \\ \boxed{T_{a_2}} \\ \vdots \\ \boxed{T_{a_n}} \end{array}, \quad \text{where } a_j \in \{1, 2, 3\}.$$

Proof. See Appendix G.3. □

Hence, if provided

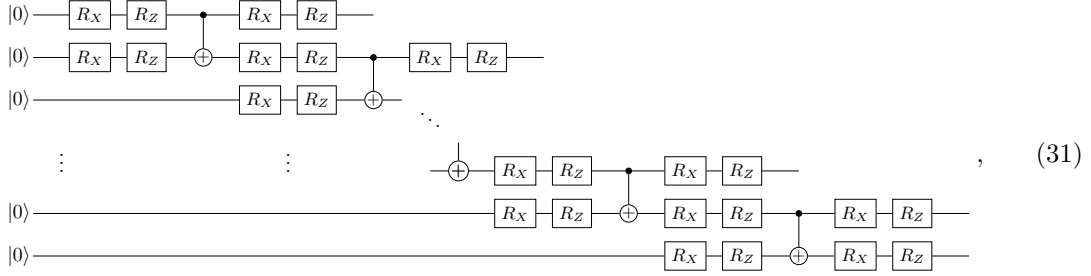
$$\tilde{I}(u_1, \dots, u_n) \in \Omega\left(\frac{1}{\text{poly}(n)}\right) \text{ or } \tilde{I}(w_1, \dots, w_n) \in \Omega\left(\frac{1}{\text{poly}(n)}\right),$$

then there exists no BP in the QCNN ansatz.

4.4 MPS-inspired ansatz

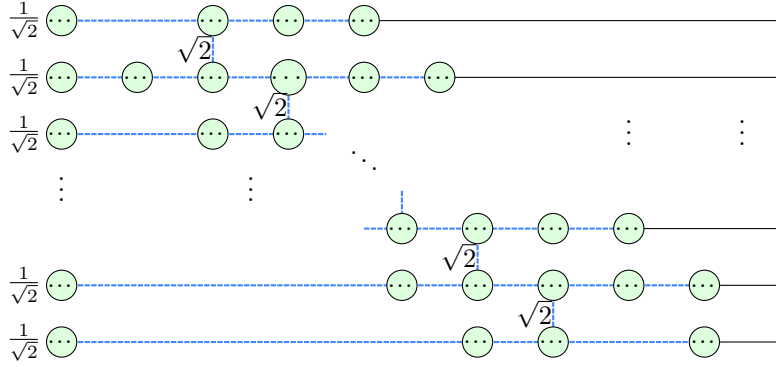
The matrix product state (MPS) is a special structure of tensor networks, which is widely used in quantum physics and machine learning [55, 56]. There are also PQCs with a similar structure as MPS, and we call it MPS-inspired ansatz. It has been shown that MPS-inspired ansatz can be implemented efficiently in quantum computers with a small number of qubits [43]. We will analyze the BP phenomenon in MPS-inspired ansatz in this section.

Let us consider the following MPS-inspired ansatz

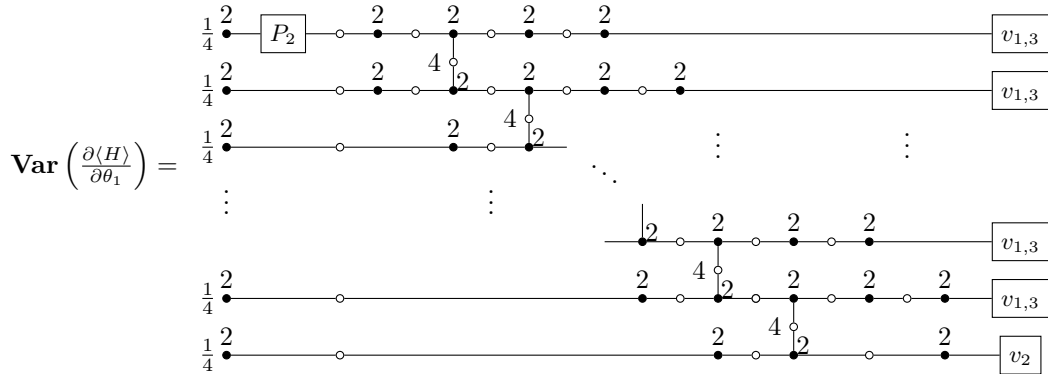


and the Hamiltonian $H = I \otimes I \cdots \otimes I \otimes X$. We will prove that the variance $\mathbf{Var} \left(\frac{\partial \langle H \rangle}{\partial \theta_1} \right)$ is exponentially small. Here θ_1 is the parameter of the first R_X gate applying on the first qubit.

Firstly, we convert the PQC into a ZX-diagram as follows.



This is a graph-like ZX-diagram whose spiders are all parameterized. We can use Lemma 3 to represent the variance as the following tensor network.



By using

$$2Mv_2 = \frac{1}{2} (v_2 + v_{1,3}^-), \quad 2Mv_{1,3} = v_{1,3},$$

and

$$\begin{array}{c} \text{---} \bullet \text{---} \boxed{v_{1,3}} \\ \text{---} \bullet \text{---} \boxed{v_2} \end{array} = \frac{1}{2} \left\{ \begin{array}{c} \boxed{v_{1,3}} \\ \boxed{v_{1,3}^-} \end{array} + \begin{array}{c} \boxed{v_{1,3}^-} \\ \boxed{v_2} \end{array} \right\}, \quad \begin{array}{c} \text{---} \bullet \text{---} \boxed{v_{1,3}} \\ \text{---} \bullet \text{---} \boxed{v_{1,3}^-} \end{array} = \begin{array}{c} \boxed{v_{1,3}^-} \\ \boxed{v_2} \end{array}, \quad \begin{array}{c} \text{---} \bullet \text{---} \boxed{v_{1,3}} \\ \text{---} \bullet \text{---} \boxed{v_{1,3}} \end{array} = \begin{array}{c} \boxed{v_{1,3}} \\ \boxed{v_{1,3}} \end{array},$$

we can simplify the variance as

$$\text{Var} \left(\frac{\partial \langle H \rangle}{\partial \theta_1} \right) = \begin{array}{c} \frac{1}{4} \begin{array}{c} \bullet \\ \text{---} \end{array} \begin{array}{c} 2 \\ \bullet \\ \text{---} \end{array} \boxed{P_2} \begin{array}{c} \circ \\ \text{---} \end{array} \begin{array}{c} 2 \\ \bullet \\ \text{---} \end{array} \boxed{\frac{3}{4} v_{1,3}^-} \\ \frac{1}{4} \begin{array}{c} \bullet \\ \text{---} \end{array} \begin{array}{c} 2 \\ \bullet \\ \text{---} \end{array} \begin{array}{c} \circ \\ \text{---} \end{array} \begin{array}{c} 2 \\ \bullet \\ \text{---} \end{array} \boxed{v_2} \\ \frac{1}{4} \begin{array}{c} \bullet \\ \text{---} \end{array} \begin{array}{c} 2 \\ \bullet \\ \text{---} \end{array} \begin{array}{c} \circ \\ \text{---} \end{array} \begin{array}{c} 2 \\ \bullet \\ \text{---} \end{array} \boxed{v_2} \\ \vdots \\ \frac{1}{4} \begin{array}{c} \bullet \\ \text{---} \end{array} \begin{array}{c} 2 \\ \bullet \\ \text{---} \end{array} \begin{array}{c} \circ \\ \text{---} \end{array} \begin{array}{c} 2 \\ \bullet \\ \text{---} \end{array} \boxed{v_2} \\ \frac{1}{4} \begin{array}{c} \bullet \\ \text{---} \end{array} \begin{array}{c} 2 \\ \bullet \\ \text{---} \end{array} \begin{array}{c} \circ \\ \text{---} \end{array} \begin{array}{c} 2 \\ \bullet \\ \text{---} \end{array} \boxed{v_2} \end{array} = \frac{1}{2^{2n-1}} \quad (32)$$

It is exponential in the number of qubits n . Hence, there exist BPs in the MPS-inspired ansatz.

5 Discussion

We developed powerful techniques to analyze the BP phenomenon for quantum neural networks training with the ZX-calculus. The quantum neural networks under consideration are PQCs under certain reasonable assumptions and the cost function is the expectation $\langle H \rangle$ of the PQC with respect to a given Hamiltonian H . The basic idea of the method is to represent the PQC, the cost function $\langle H \rangle$, and the gradients $\frac{\partial \langle H \rangle}{\partial \theta_j}$ as ZX-diagrams. And then computing the expectation and the variance of the gradient of $\langle H \rangle$ becomes computing the integration of certain ZX-diagrams. We show that these integrations are sums of ZX-diagrams that can be computed explicitly in many cases. As future works, it would be desirable to use the completeness of ZX-calculus to represent these sums by ZX-diagrams, or more generally, diagrams in other complete graphical calculi.

In principle, these techniques can be used to any given ansatz under Assumption 1. We remark that these techniques can be used to analyze the BP phenomenon for PQCs which contain t -design sub-blocks, for example, the PQCs considered in [21, 54]. Because the t -design sub-blocks can be replaced with concrete t -design PQCs and then the techniques proposed in this paper can be applied. Techniques introduced in this paper can be used in more cases including circuits with global cost functions and circuits of any depth. To analyze a PQC with a global cost function, one can represent the global Hamiltonian as a ZX-diagram and then the method introduced in this paper can be used. As shown in Section 4.1, the BP phenomenon for the hardware-efficient ansatz has been analyzed when the depth is $O(\text{poly}(n))$. In conclusion, we extend the BP theorem from unitary 2-design circuits to any parameterized quantum circuits under Assumption 1.

Using the techniques proposed in this paper, we analyzed four kinds of ansätze, including the hardware-efficient ansatz, the tree tensor network ansatz, the QCNN ansatz, and the MPS-inspired ansatz. It is shown that there exist BPs in the hardware-efficient-ansatz and the MPS-inspired ansatz, while there exists no BP in the tree tensor network ansatz and the QCNN ansatz.

Acknowledgment. This work is partially supported by a NSFC grant No.11688101 and by a NKRDG grant No.2018YFA0306702. We want to thank the anonymous referees for their valuable suggestions.

References

- [1] Alberto Peruzzo, Jarrod McClean, Peter Shadbolt, Man-Hong Yung, Xiao-Qi Zhou, Peter J. Love, Alán Aspuru-Guzik, and Jeremy L. O’Brien. A variational eigenvalue solver on a photonic quantum processor. *Nature Communications*, 5(1):4213, Jul 2014. ISSN 2041-1723. DOI: 10.1038/ncomms5213.
- [2] Abhinav Kandala, Antonio Mezzacapo, Kristan Temme, Maika Takita, Markus Brink, Jerry M. Chow, and Jay M. Gambetta. Hardware-efficient variational quantum eigensolver for

- small molecules and quantum magnets. *Nature*, 549(7671):242–246, Sep 2017. ISSN 1476-4687. DOI: [10.1038/nature23879](https://doi.org/10.1038/nature23879).
- [3] Yudong Cao, Jonathan Romero, Jonathan P. Olson, Matthias Degroote, Peter D. Johnson, Mária Kieferová, Ian D. Kivlichan, Tim Menke, Borja Peropadre, Nicolas P. D. Sawaya, Sukin Sim, Libor Veis, and Alán Aspuru-Guzik. Quantum chemistry in the age of quantum computing. *Chemical Reviews*, 119(19):10856–10915, Oct 2019. ISSN 0009-2665. DOI: [10.1021/acs.chemrev.8b00803](https://doi.org/10.1021/acs.chemrev.8b00803).
 - [4] Bela Bauer, Sergey Bravyi, Mario Motta, and Garnet Kin-Lic Chan. Quantum algorithms for quantum chemistry and quantum materials science. *Chemical Reviews*, 120(22):12685–12717, Nov 2020. ISSN 0009-2665. DOI: [10.1021/acs.chemrev.9b00829](https://doi.org/10.1021/acs.chemrev.9b00829).
 - [5] Edward Farhi, Jeffrey Goldstone, and Sam Gutmann. A quantum approximate optimization algorithm. *arXiv preprint arXiv:1411.4028*, 2014.
 - [6] Leo Zhou, Sheng-Tao Wang, Soonwon Choi, Hannes Pichler, and Mikhail D. Lukin. Quantum approximate optimization algorithm: Performance, mechanism, and implementation on near-term devices. *Phys. Rev. X*, 10:021067, Jun 2020. DOI: [10.1103/PhysRevX.10.021067](https://doi.org/10.1103/PhysRevX.10.021067).
 - [7] Jin-Guo Liu and Lei Wang. Differentiable learning of quantum circuit born machines. *Phys. Rev. A*, 98:062324, Dec 2018. DOI: [10.1103/PhysRevA.98.062324](https://doi.org/10.1103/PhysRevA.98.062324).
 - [8] Seth Lloyd and Christian Weedbrook. Quantum generative adversarial learning. *Phys. Rev. Lett.*, 121:040502, Jul 2018. DOI: [10.1103/PhysRevLett.121.040502](https://doi.org/10.1103/PhysRevLett.121.040502).
 - [9] Vojtěch Havlíček, Antonio D. Córcoles, Kristan Temme, Aram W. Harrow, Abhinav Kandala, Jerry M. Chow, and Jay M. Gambetta. Supervised learning with quantum-enhanced feature spaces. *Nature*, 567(7747):209–212, Mar 2019. ISSN 1476-4687. DOI: [10.1038/s41586-019-0980-2](https://doi.org/10.1038/s41586-019-0980-2).
 - [10] Maria Schuld, Alex Bocharov, Krysta M. Svore, and Nathan Wiebe. Circuit-centric quantum classifiers. *Phys. Rev. A*, 101:032308, Mar 2020. DOI: [10.1103/PhysRevA.101.032308](https://doi.org/10.1103/PhysRevA.101.032308).
 - [11] Marcello Benedetti, Erika Lloyd, Stefan Sack, and Mattia Fiorentini. Parameterized quantum circuits as machine learning models. *Quantum Science and Technology*, 4(4):043001, nov 2019. DOI: [10.1088/2058-9565/ab4eb5](https://doi.org/10.1088/2058-9565/ab4eb5).
 - [12] Chen Zhao and Xiao-Shan Gao. QDNN: DNN with quantum neural network layers. *arXiv preprint arXiv:1912.12660*, 2019.
 - [13] John Preskill. Quantum Computing in the NISQ era and beyond. *Quantum*, 2:79, August 2018. ISSN 2521-327X. DOI: [10.22331/q-2018-08-06-79](https://doi.org/10.22331/q-2018-08-06-79).
 - [14] Maria Schuld, Ville Bergholm, Christian Gogolin, Josh Izaac, and Nathan Killoran. Evaluating analytic gradients on quantum hardware. *Phys. Rev. A*, 99:032331, Mar 2019. DOI: [10.1103/PhysRevA.99.032331](https://doi.org/10.1103/PhysRevA.99.032331).
 - [15] Andrea Mari, Thomas R. Bromley, and Nathan Killoran. Estimating the gradient and higher-order derivatives on quantum hardware. *Phys. Rev. A*, 103:012405, Jan 2021. DOI: [10.1103/PhysRevA.103.012405](https://doi.org/10.1103/PhysRevA.103.012405).
 - [16] James Stokes, Josh Izaac, Nathan Killoran, and Giuseppe Carleo. Quantum Natural Gradient. *Quantum*, 4:269, May 2020. ISSN 2521-327X. DOI: [10.22331/q-2020-05-25-269](https://doi.org/10.22331/q-2020-05-25-269).
 - [17] Jonas M. Kübler, Andrew Arrasmith, Lukasz Cincio, and Patrick J. Coles. An Adaptive Optimizer for Measurement-Frugal Variational Algorithms. *Quantum*, 4:263, May 2020. ISSN 2521-327X. DOI: [10.22331/q-2020-05-11-263](https://doi.org/10.22331/q-2020-05-11-263).
 - [18] Ken M. Nakanishi, Keisuke Fujii, and Syngae Todo. Sequential minimal optimization for quantum-classical hybrid algorithms. *Phys. Rev. Research*, 2:043158, Oct 2020. DOI: [10.1103/PhysRevResearch.2.043158](https://doi.org/10.1103/PhysRevResearch.2.043158).
 - [19] Jarrod R. McClean, Sergio Boixo, Vadim N. Smelyanskiy, Ryan Babbush, and Hartmut Neven. Barren plateaus in quantum neural network training landscapes. *Nature Communications*, 9(1):4812, Nov 2018. ISSN 2041-1723. DOI: [10.1038/s41467-018-07090-4](https://doi.org/10.1038/s41467-018-07090-4).
 - [20] Zoë Holmes, Kunal Sharma, M. Cerezo, and Patrick J. Coles. Connecting ansatz expressibility to gradient magnitudes and barren plateaus, 2021.
 - [21] Marco Cerezo, Akira Sone, Tyler Volkoff, Lukasz Cincio, and Patrick J. Coles. Cost function dependent barren plateaus in shallow parametrized quantum circuits. *Nature Communications*, 12(1):1791, Mar 2021. ISSN 2041-1723. DOI: [10.1038/s41467-021-21728-w](https://doi.org/10.1038/s41467-021-21728-w).
 - [22] Carlos Ortiz Marrero, Mária Kieferová, and Nathan Wiebe. Entanglement induced barren plateaus. *arXiv preprint arXiv:2010.15968*, 2020.

- [23] Taylor L. Patti, Khadijeh Najafi, Xun Gao, and Susanne F. Yelin. Entanglement devised barren plateau mitigation, 2020.
- [24] Kunal Sharma, M. Cerezo, Lukasz Cincio, and Patrick J. Coles. Trainability of dissipative perceptron-based quantum neural networks, 2020.
- [25] Samson Wang, Enrico Fontana, Marco Cerezo, Kunal Sharma, Akira Sone, Lukasz Cincio, and Patrick J Coles. Noise-induced barren plateaus in variational quantum algorithms. *arXiv preprint arXiv:2007.14384*, 2020.
- [26] Edward Grant, Leonard Wossnig, Mateusz Ostaszewski, and Marcello Benedetti. An initialization strategy for addressing barren plateaus in parametrized quantum circuits. *Quantum*, 3:214, December 2019. ISSN 2521-327X. DOI: [10.22331/q-2019-12-09-214](https://doi.org/10.22331/q-2019-12-09-214).
- [27] Tyler Volkoff and Patrick J Coles. Large gradients via correlation in random parameterized quantum circuits. *Quantum Science and Technology*, 6(2):025008, jan 2021. DOI: [10.1088/2058-9565/abd891](https://doi.org/10.1088/2058-9565/abd891).
- [28] Bob Coecke and Ross Duncan. Interacting quantum observables. In *Proceedings of the 37th International Colloquium on Automata, Languages and Programming (ICALP)*, Lecture Notes in Computer Science, 2008. DOI: [10.1007/978-3-540-70583-3_25](https://doi.org/10.1007/978-3-540-70583-3_25).
- [29] Bob Coecke and Ross Duncan. Interacting quantum observables: categorical algebra and diagrammatics. *New Journal of Physics*, 13:043016, 2011. DOI: [10.1088/1367-2630/13/4/043016](https://doi.org/10.1088/1367-2630/13/4/043016).
- [30] Ross Duncan, Aleks Kissinger, Simon Pedrix, and John van de Wetering. Graph-theoretic Simplification of Quantum Circuits with the ZX-calculus. *Quantum*, 4:279, 6 2020. ISSN 2521-327X. DOI: [10.22331/q-2020-06-04-279](https://doi.org/10.22331/q-2020-06-04-279).
- [31] Aleks Kissinger and John van de Wetering. Reducing T-count with the ZX-calculus. *Physical Review A*, 102:022406, 8 2020. DOI: [10.1103/PhysRevA.102.022406](https://doi.org/10.1103/PhysRevA.102.022406).
- [32] Alexander Cowtan, Silas Dilkes, Ross Duncan, Will Simmons, and Seyon Sivarajah. Phase Gadget Synthesis for Shallow Circuits. In Bob Coecke and Matthew Leifer, editors, Proceedings 16th International Conference on *Quantum Physics and Logic*, Chapman University, Orange, CA, USA., 10-14 June 2019, volume 318 of *Electronic Proceedings in Theoretical Computer Science*, pages 213–228. Open Publishing Association, 2020. DOI: [10.4204/EPTCS.318.13](https://doi.org/10.4204/EPTCS.318.13).
- [33] Michael Hanks, Marta P. Estarellas, William J. Munro, and Kae Nemoto. Effective Compression of Quantum Braided Circuits Aided by ZX-Calculus. *Physical Review X*, 10:041030, 2020. DOI: [10.1103/PhysRevX.10.041030](https://doi.org/10.1103/PhysRevX.10.041030).
- [34] Ross Duncan. A graphical approach to measurement-based quantum computing. In Mehrnoosh Sadrzadeh Chris Heunen and Edward Grefenstette, editors, *Quantum Physics and Linguistics: A Compositional, Diagrammatic Discourse*. 2013. ISBN 9780199646296. DOI: [10.1093/acprof:oso/9780199646296.001.0001](https://doi.org/10.1093/acprof:oso/9780199646296.001.0001).
- [35] Miriam Backens, Hector Miller-Bakewell, Giovanni de Felice, Leo Lobski, and John van de Wetering. There and back again: A circuit extraction tale. *arXiv preprint arXiv:2003.01664*, 2020.
- [36] Nicholas Chancellor, Aleks Kissinger, Joschka Roffe, Stefan Zohren, and Dominic Horsman. Graphical Structures for Design and Verification of Quantum Error Correction. *arXiv preprint arXiv:1611.08012*, 2016.
- [37] Niel de Beaudrap and Dominic Horsman. The ZX calculus is a language for surface code lattice surgery. *Quantum*, 4:218, January 2020. ISSN 2521-327X. DOI: [10.22331/q-2020-01-09-218](https://doi.org/10.22331/q-2020-01-09-218).
- [38] Richard D. P. East, John van de Wetering, Nicholas Chancellor, and Adolfo G. Grushin. AKLT-states as ZX-diagrams: diagrammatic reasoning for quantum states. *arXiv preprint arXiv:2012.01219*, 2020.
- [39] Richie Yeung. Diagrammatic design and study of ansätze for quantum machine learning, 2020.
- [40] Bob Coecke, Giovanni de Felice, Konstantinos Meichanetzidis, and Alexis Toumi. Foundations for Near-Term Quantum Natural Language Processing. *arXiv preprint arXiv:2012.03755*, 2020.
- [41] Iris Cong, Soonwon Choi, and Mikhail D. Lukin. Quantum convolutional neural networks. *Nature Physics*, 15(12):1273–1278, Dec 2019. ISSN 1745-2481. DOI: [10.1038/s41567-019-0648-8](https://doi.org/10.1038/s41567-019-0648-8).

- [42] Edward Grant, Marcello Benedetti, Shuxiang Cao, Andrew Hallam, Joshua Lockhart, Vid Stojevic, Andrew G. Green, and Simone Severini. Hierarchical quantum classifiers. *npj Quantum Information*, 4(1):65, Dec 2018. ISSN 2056-6387. DOI: [10.1038/s41534-018-0116-9](https://doi.org/10.1038/s41534-018-0116-9).
- [43] Jin-Guo Liu, Yi-Hong Zhang, Yuan Wan, and Lei Wang. Variational quantum eigensolver with fewer qubits. *Phys. Rev. Research*, 1:023025, Sep 2019. DOI: [10.1103/PhysRevResearch.1.023025](https://doi.org/10.1103/PhysRevResearch.1.023025).
- [44] Bob Coecke and Aleks Kissinger. *Picturing Quantum Processes*. Cambridge University Press, 2017. DOI: [10.1007/978-3-319-91376-6](https://doi.org/10.1007/978-3-319-91376-6).
- [45] John van de Wetering. ZX-calculus for the working quantum computer scientist. *arXiv preprint arXiv:2012.13966*, 2020.
- [46] Miriam Backens. The ZX-calculus is complete for stabilizer quantum mechanics. *New Journal of Physics*, 16(9):093021, 2014. DOI: [10.1088/1367-2630/16/9/093021](https://doi.org/10.1088/1367-2630/16/9/093021).
- [47] Miriam Backens. Making the stabilizer ZX-calculus complete for scalars. In Chris Heunen, Peter Selinger, and Jamie Vicary, editors, *Proceedings of the 12th International Workshop on Quantum Physics and Logic (QPL 2015)*, volume 195 of *Electronic Proceedings in Theoretical Computer Science*, pages 17–32, 2015. DOI: [10.4204/EPTCS.195.2](https://doi.org/10.4204/EPTCS.195.2).
- [48] Emmanuel Jeandel, Simon Perdrix, and Renaud Vilmart. A Complete Axiomatisation of the ZX-Calculus for Clifford+T Quantum Mechanics. In *Proceedings of the 33rd Annual ACM/IEEE Symposium on Logic in Computer Science, LICS '18*, pages 559–568, New York, NY, USA, 2018. ACM. ISBN 978-1-4503-5583-4. DOI: [10.1145/3209108.3209131](https://doi.org/10.1145/3209108.3209131).
- [49] Emmanuel Jeandel, Simon Perdrix, and Renaud Vilmart. Diagrammatic Reasoning Beyond Clifford+T Quantum Mechanics. In *Proceedings of the 33rd Annual ACM/IEEE Symposium on Logic in Computer Science, LICS '18*, pages 569–578, New York, NY, USA, 2018. ACM. ISBN 978-1-4503-5583-4. DOI: [10.1145/3209108.3209139](https://doi.org/10.1145/3209108.3209139).
- [50] Amar Hadzihasanovic, Kang Feng Ng, and Quanlong Wang. Two complete axiomatisations of pure-state qubit quantum computing. In *Proceedings of the 33rd Annual ACM/IEEE Symposium on Logic in Computer Science, LICS '18*, page 502–511, New York, NY, USA, 2018. Association for Computing Machinery. ISBN 9781450355834. DOI: [10.1145/3209108.3209128](https://doi.org/10.1145/3209108.3209128).
- [51] Quanlong Wang. *Completeness of the ZX-calculus*. PhD thesis, University of Oxford, 2018.
- [52] Emmanuel Jeandel, Simon Perdrix, and Renaud Vilmart. Completeness of the zx-calculus. *Logical Methods in Computer Science*, 6 2020. DOI: [10.23638/LMCS-16\(2:11\)2020](https://doi.org/10.23638/LMCS-16(2:11)2020).
- [53] Kaining Zhang, Min-Hsiu Hsieh, Liu Liu, and Dacheng Tao. Toward trainability of quantum neural networks. *arXiv preprint arXiv:2011.06258*, 2020.
- [54] Arthur Pesah, M Cerezo, Samson Wang, Tyler Volkoff, Andrew T Sornborger, and Patrick J Coles. Absence of barren plateaus in quantum convolutional neural networks. *arXiv preprint arXiv:2011.02966*, 2020.
- [55] F. Verstraete, V. Murg, and J.I. Cirac. Matrix product states, projected entangled pair states, and variational renormalization group methods for quantum spin systems. *Advances in Physics*, 57(2):143–224, 2008. DOI: [10.1080/14789940801912366](https://doi.org/10.1080/14789940801912366).
- [56] Zhao-Yu Han, Jun Wang, Heng Fan, Lei Wang, and Pan Zhang. Unsupervised generative modeling using matrix product states. *Phys. Rev. X*, 8:031012, Jul 2018. DOI: [10.1103/PhysRevX.8.031012](https://doi.org/10.1103/PhysRevX.8.031012).

A A simple example

In this section, we will use a simple example considered in Section 3.1 to illustrate the techniques proposed in this paper. Recall that the PQC considered is

$$\begin{array}{c}
 |0\rangle \text{---} \boxed{R_X(\theta_1)} \text{---} \bullet \text{---} \boxed{R_Z(\theta_3)} \text{---} \\
 |0\rangle \text{---} \boxed{R_X(\theta_2)} \text{---} \oplus \text{---} \boxed{R_Z(\theta_4)} \text{---}
 \end{array}
 =
 \begin{array}{c}
 \frac{1}{\sqrt{2}} \text{---} \text{red circle } \theta_1 \text{---} \text{green circle } \theta_3 \text{---} \\
 \frac{1}{\sqrt{2}} \text{---} \text{red circle } \theta_2 \text{---} \text{red circle } \theta_4 \text{---}
 \end{array}
 (\mathbf{f}), (\mathbf{h})
 =
 \begin{array}{c}
 \frac{1}{\sqrt{2}} \text{---} \text{green circle } \theta_1 \text{---} \text{green circle } \theta_3 \text{---} \\
 \frac{1}{\sqrt{2}} \text{---} \text{green circle } \theta_2 \text{---} \text{green circle } \theta_4 \text{---}
 \end{array}$$

and the gradient with respect to θ_1 can be represented as

$$\frac{\partial \langle H \rangle}{\partial \theta_1} = \frac{1}{\sqrt{2}} \begin{array}{c} \pi \\ \theta_1 + \frac{\pi}{2} \quad \theta_3 \\ \theta_2 \quad \theta_4 \\ H \\ -\theta_3 \quad -\theta_1 - \frac{\pi}{2} \\ \theta_4 \quad \theta_2 \\ \frac{1}{\sqrt{2}} \end{array}.$$

By Lemma 1, the expectation of this gradient is zero. The variance of the gradient is the following integration.

$$\mathbf{Var} \left(\frac{\partial \langle H \rangle}{\partial \theta_1} \right) = \frac{1}{16(2\pi)^4} \int_{\theta_1} \int_{\theta_2} \int_{\theta_3} \int_{\theta_4} \begin{array}{c} \pi \\ \theta_1 + \frac{\pi}{2} \quad \theta_3 \\ \theta_2 \quad \theta_4 \\ H \\ -\theta_3 \quad -\theta_1 - \frac{\pi}{2} \\ \theta_4 \quad \theta_2 \\ H \\ -\theta_1 - \frac{\pi}{2} \quad -\theta_3 \\ \theta_3 \quad \theta_1 + \frac{\pi}{2} \\ \pi \end{array} d\theta_1 d\theta_2 d\theta_3 d\theta_4.$$

Now we choose the Hamiltonian $H = X \otimes X$. Then by Lemma 2 and Theorem 4, we have

$$16 \cdot \mathbf{Var} \left(\frac{\partial \langle H \rangle}{\partial \theta_1} \right) = \sum_{a_1, a_2, a_3, a_4 \in \{T_1, T_2, T_3\}} \begin{array}{c} \pi \\ \pi \\ \pi \\ a_1 \quad a_2 \quad a_3 \quad a_4 \\ \pi \\ \pi \\ \pi \\ \pi \end{array} = \sum_{a_2, a_3, a_4 \in \{T_1, T_2, T_3\}} \begin{array}{c} \pi \\ \pi \\ \pi \\ a_2 \quad a_3 \quad a_4 \\ \pi \\ \pi \\ \pi \end{array}.$$

Using Lemma 3 recursively, we can break the large ZX-diagram into small parts as follows.

$$\begin{aligned} 16 \cdot \mathbf{Var} \left(\frac{\partial \langle H \rangle}{\partial \theta_1} \right) &= \sum_{a_2, a_3, a_4 \in \{T_1, T_2, T_3\}} \begin{array}{c} \pi \\ \pi \\ \pi \\ a_2 \quad a_3 \quad a_4 \\ \pi \\ \pi \\ \pi \end{array} = \sum_{a_2, a_3, a_4 \in \{T_1, T_2, T_3\}} M_{T_2, a_3} \cdot \begin{array}{c} \pi \\ \pi \\ \pi \\ a_2 \quad a_3 \quad a_4 \\ \pi \\ \pi \\ \pi \end{array} \\ &= \sum_{a_2, a_3, a_4 \in \{T_1, T_2, T_3\}} M_{T_2, a_3} \cdot M_{a_2, a_3} \cdot \begin{array}{c} \pi \\ \pi \\ \pi \\ a_2 \quad a_3 \quad a_4 \\ \pi \\ \pi \\ \pi \end{array} = \sum_{a_2, a_3, a_4 \in \{T_1, T_2, T_3\}} M_{T_2, a_3} \cdot M_{a_2, a_3} \cdot M_{a_2, a_4} \cdot \begin{array}{c} \pi \\ \pi \\ \pi \\ a_2 \quad a_3 \quad a_4 \\ \pi \\ \pi \\ \pi \end{array} \\ &= \sum_{a_2, a_3, a_4 \in \{T_1, T_2, T_3\}} M_{T_2, a_3} \cdot M_{a_2, a_3} \cdot M_{a_2, a_4} \cdot \begin{array}{c} \pi \\ \pi \\ \pi \\ a_2 \quad a_3 \quad a_4 \\ \pi \\ \pi \\ \pi \end{array}. \end{aligned}$$

We denote

$$\begin{aligned} \tilde{I}_{a_1, a_2} &= \begin{array}{c} \circ \quad \circ \\ \diagdown \quad \diagup \\ a_1 \\ \diagup \quad \diagdown \\ \circ \quad \circ \end{array} \cdot \begin{array}{c} \circ \quad \circ \\ \diagdown \quad \diagup \\ a_2 \\ \diagup \quad \diagdown \\ \circ \quad \circ \end{array}, \\ \tilde{H}_{a_3, a_4} &= \begin{array}{c} \circ \quad \circ \\ \diagdown \quad \diagup \\ \pi \\ \diagup \quad \diagdown \\ a_3 \\ \diagdown \quad \diagup \\ \pi \\ \diagup \quad \diagdown \\ \circ \quad \circ \end{array} \cdot \begin{array}{c} \circ \quad \circ \\ \diagdown \quad \diagup \\ \pi \\ \diagup \quad \diagdown \\ a_4 \\ \diagdown \quad \diagup \\ \pi \\ \diagup \quad \diagdown \\ \circ \quad \circ \end{array}. \end{aligned}$$

Then the variance can be represented as the following tensor network.

$$16 \cdot \mathbf{Var} \left(\frac{\partial \langle H \rangle}{\partial \theta_1} \right) = \sum_{a_2, a_3, a_4 \in \{T_1, T_2, T_3\}} M_{T_2, a_3} \cdot M_{a_2, a_3} \cdot M_{a_2, a_4} \cdot \tilde{I}_{T_2, a_2} \cdot \tilde{H}_{a_3, a_4} = \begin{array}{c} \boxed{\tilde{I}} \quad \boxed{P_2} \\ \diagdown \quad \diagup \\ \bullet \quad \bullet \\ \diagup \quad \diagdown \\ \bullet \quad \bullet \\ \boxed{\tilde{H}} \end{array}.$$

By contracting this tensor network, we finally obtain

$$\mathbf{Var} \left(\frac{\partial \langle H \rangle}{\partial \theta_1} \right) = \frac{1}{16} \cdot \frac{3}{4}.$$

B Proof of theorem 2

Theorem 2. *The gradient can be represented as the following equation.*

$$\begin{aligned} \frac{2^n}{|c|^2} \cdot \frac{\partial \langle H \rangle}{\partial \theta_j} &= \frac{\partial}{\partial \theta_j} \begin{array}{c} \circ \quad \circ \quad \dots \quad \circ \quad \circ \\ \diagdown \quad \diagup \quad \dots \quad \diagdown \quad \diagup \\ \vdots \quad \vdots \quad \dots \quad \vdots \quad \vdots \\ \theta_j \quad \dots \quad \theta_j \\ \diagup \quad \diagdown \quad \dots \quad \diagup \quad \diagdown \\ \circ \quad \circ \quad \dots \quad \circ \quad \circ \end{array} \begin{array}{c} \boxed{H} \\ \vdots \\ \boxed{H} \end{array} \begin{array}{c} \circ \quad \circ \quad \dots \quad \circ \quad \circ \\ \diagdown \quad \diagup \quad \dots \quad \diagdown \quad \diagup \\ \vdots \quad \vdots \quad \dots \quad \vdots \quad \vdots \\ -\theta_j \quad \dots \quad -\theta_j \\ \diagup \quad \diagdown \quad \dots \quad \diagup \quad \diagdown \\ \circ \quad \circ \quad \dots \quad \circ \quad \circ \end{array} \\ &= \begin{array}{c} \circ \quad \circ \quad \dots \quad \circ \quad \circ \\ \diagdown \quad \diagup \quad \dots \quad \diagdown \quad \diagup \\ \vdots \quad \vdots \quad \dots \quad \vdots \quad \vdots \\ \theta_j + \frac{\pi}{2} \quad \dots \quad -\theta_j - \frac{\pi}{2} \\ \diagup \quad \diagdown \quad \dots \quad \diagup \quad \diagdown \\ \circ \quad \circ \quad \dots \quad \circ \quad \circ \end{array} \begin{array}{c} \boxed{H} \\ \vdots \\ \boxed{H} \end{array} \begin{array}{c} \circ \quad \circ \quad \dots \quad \circ \quad \circ \\ \diagdown \quad \diagup \quad \dots \quad \diagdown \quad \diagup \\ \vdots \quad \vdots \quad \dots \quad \vdots \quad \vdots \\ \pi \\ \diagup \quad \diagdown \quad \dots \quad \diagup \quad \diagdown \\ \circ \quad \circ \quad \dots \quad \circ \quad \circ \end{array} \end{aligned}$$

Proof. Consider the spiders corresponding to θ_j . We expand the spiders as follows.

$$m \left\{ \begin{array}{c} \circ \quad \circ \\ \diagdown \quad \diagup \\ \theta_j \\ \diagup \quad \diagdown \\ \circ \quad \circ \end{array} \right\} n \dots n \left\{ \begin{array}{c} \circ \quad \circ \\ \diagdown \quad \diagup \\ -\theta_j \\ \diagup \quad \diagdown \\ \circ \quad \circ \end{array} \right\} m = \begin{array}{l} |0\rangle^{\otimes n} \langle 0|^{\otimes m} \dots |0\rangle^{\otimes m} \langle 0|^{\otimes n} + e^{i\theta_j} |1\rangle^{\otimes n} \langle 1|^{\otimes m} \dots |0\rangle^{\otimes m} \langle 0|^{\otimes n} \\ + e^{-i\theta_j} |0\rangle^{\otimes n} \langle 0|^{\otimes m} \dots |1\rangle^{\otimes m} \langle 1|^{\otimes n} + |1\rangle^{\otimes n} \langle 1|^{\otimes m} \dots |1\rangle^{\otimes m} \langle 1|^{\otimes n} \end{array}.$$

Taking the partial derivative of θ_j on the two sides, we obtain

$$\begin{aligned} \frac{\partial}{\partial \theta_j} m \left\{ \begin{array}{c} \circ \quad \circ \\ \diagdown \quad \diagup \\ \theta_j \\ \diagup \quad \diagdown \\ \circ \quad \circ \end{array} \right\} n \dots n \left\{ \begin{array}{c} \circ \quad \circ \\ \diagdown \quad \diagup \\ -\theta_j \\ \diagup \quad \diagdown \\ \circ \quad \circ \end{array} \right\} m &= i e^{i\theta_j} |1\rangle^{\otimes n} \langle 1|^{\otimes m} \dots |0\rangle^{\otimes n} \langle 0|^{\otimes m} - i e^{-i\theta_j} |0\rangle^{\otimes n} \langle 0|^{\otimes m} \dots |1\rangle^{\otimes n} \langle 1|^{\otimes m} \\ &= e^{i(\theta_j + \frac{\pi}{2})} |1\rangle^{\otimes n} \langle 1|^{\otimes m} \dots |0\rangle^{\otimes n} \langle 0|^{\otimes m} + e^{-i(\theta_j + \frac{\pi}{2})} |0\rangle^{\otimes n} \langle 0|^{\otimes m} \dots |1\rangle^{\otimes n} \langle 1|^{\otimes m} \\ &= m \left\{ \begin{array}{c} \circ \quad \circ \\ \diagdown \quad \diagup \\ \theta_j + \frac{\pi}{2} \\ \diagup \quad \diagdown \\ \circ \quad \circ \end{array} \right\} n \dots n \left\{ \begin{array}{c} \circ \quad \circ \\ \diagdown \quad \diagup \\ -\theta_j - \frac{\pi}{2} \\ \diagup \quad \diagdown \\ \circ \quad \circ \end{array} \right\} m \end{aligned}$$

□

C Proof of lemma 1

Lemma 1. *The following equation holds.*

□

E Proof of lemma 3

Lemma 3. *The following equation holds.*

$$\frac{1}{(2\pi)^2} \int_{\theta_j} \int_{\theta_k} \begin{array}{c} \text{---} \text{---} \text{---} \text{---} \text{---} \\ \text{---} \text{---} \text{---} \text{---} \text{---} \\ \text{---} \text{---} \text{---} \text{---} \text{---} \\ \text{---} \text{---} \text{---} \text{---} \text{---} \\ \text{---} \text{---} \text{---} \text{---} \text{---} \end{array} \begin{array}{c} \text{---} \text{---} \text{---} \text{---} \text{---} \\ \text{---} \text{---} \text{---} \text{---} \text{---} \\ \text{---} \text{---} \text{---} \text{---} \text{---} \\ \text{---} \text{---} \text{---} \text{---} \text{---} \\ \text{---} \text{---} \text{---} \text{---} \text{---} \end{array} d\theta_j d\theta_k = \sum_{a_j, a_k \in \{T_1, T_2, T_3\}} M_{a_j, a_k} \begin{array}{c} \text{---} \text{---} \text{---} \text{---} \text{---} \\ \text{---} \text{---} \text{---} \text{---} \text{---} \\ \text{---} \text{---} \text{---} \text{---} \text{---} \\ \text{---} \text{---} \text{---} \text{---} \text{---} \\ \text{---} \text{---} \text{---} \text{---} \text{---} \end{array}$$

Here M_{a_j, a_k} is the element on the a_j -th row and a_k -th column in the following 3×3 matrix

$$M = \frac{1}{4} \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & -1 \\ 1 & -1 & 1 \end{pmatrix}$$

Proof. By Lemma 2, we have

$$\frac{1}{(2\pi)^2} \int_{\theta_j} \int_{\theta_k} \begin{array}{c} \text{---} \text{---} \text{---} \text{---} \text{---} \\ \text{---} \text{---} \text{---} \text{---} \text{---} \\ \text{---} \text{---} \text{---} \text{---} \text{---} \\ \text{---} \text{---} \text{---} \text{---} \text{---} \\ \text{---} \text{---} \text{---} \text{---} \text{---} \end{array} \begin{array}{c} \text{---} \text{---} \text{---} \text{---} \text{---} \\ \text{---} \text{---} \text{---} \text{---} \text{---} \\ \text{---} \text{---} \text{---} \text{---} \text{---} \\ \text{---} \text{---} \text{---} \text{---} \text{---} \\ \text{---} \text{---} \text{---} \text{---} \text{---} \end{array} d\theta_j d\theta_k = \sum_{a_j, a_k \in \{T_1, T_2, T_3\}} \begin{array}{c} \text{---} \text{---} \text{---} \text{---} \text{---} \\ \text{---} \text{---} \text{---} \text{---} \text{---} \\ \text{---} \text{---} \text{---} \text{---} \text{---} \\ \text{---} \text{---} \text{---} \text{---} \text{---} \\ \text{---} \text{---} \text{---} \text{---} \text{---} \end{array}$$

Hence, it is sufficient to prove

$$\begin{array}{c} \text{---} \text{---} \text{---} \text{---} \text{---} \\ \text{---} \text{---} \text{---} \text{---} \text{---} \\ \text{---} \text{---} \text{---} \text{---} \text{---} \\ \text{---} \text{---} \text{---} \text{---} \text{---} \\ \text{---} \text{---} \text{---} \text{---} \text{---} \end{array} = M_{a_j, a_k} \begin{array}{c} \text{---} \text{---} \text{---} \text{---} \text{---} \\ \text{---} \text{---} \text{---} \text{---} \text{---} \\ \text{---} \text{---} \text{---} \text{---} \text{---} \\ \text{---} \text{---} \text{---} \text{---} \text{---} \\ \text{---} \text{---} \text{---} \text{---} \text{---} \end{array}$$

for $a_j, a_k \in \{T_1, T_2, T_3\}$.

For $a_j = a_k = T_1$, we have

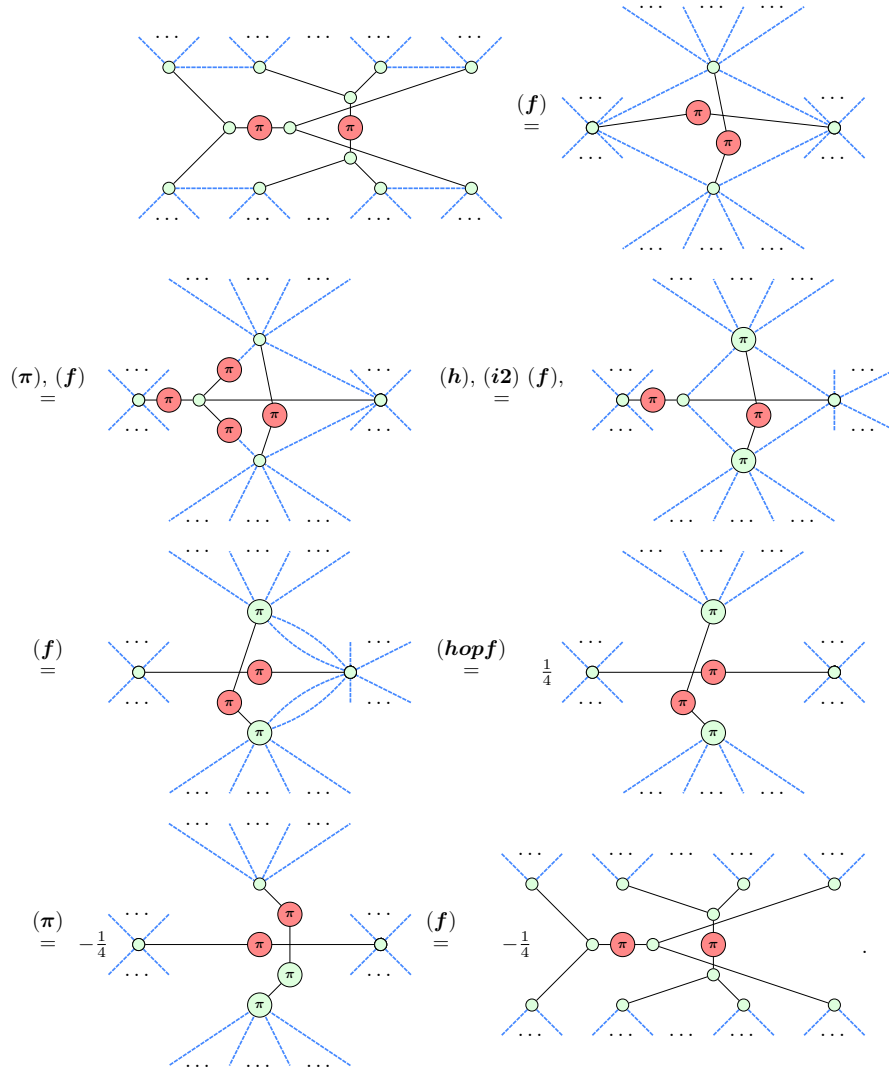
$$\begin{array}{c} \text{---} \text{---} \text{---} \text{---} \text{---} \\ \text{---} \text{---} \text{---} \text{---} \text{---} \\ \text{---} \text{---} \text{---} \text{---} \text{---} \\ \text{---} \text{---} \text{---} \text{---} \text{---} \\ \text{---} \text{---} \text{---} \text{---} \text{---} \end{array} \stackrel{(f)}{=} \begin{array}{c} \text{---} \text{---} \text{---} \text{---} \text{---} \\ \text{---} \text{---} \text{---} \text{---} \text{---} \\ \text{---} \text{---} \text{---} \text{---} \text{---} \\ \text{---} \text{---} \text{---} \text{---} \text{---} \\ \text{---} \text{---} \text{---} \text{---} \text{---} \end{array} \quad (33)$$

$$\stackrel{(hopf)}{=} \frac{1}{4} \begin{array}{c} \text{---} \text{---} \text{---} \text{---} \text{---} \\ \text{---} \text{---} \text{---} \text{---} \text{---} \\ \text{---} \text{---} \text{---} \text{---} \text{---} \\ \text{---} \text{---} \text{---} \text{---} \text{---} \\ \text{---} \text{---} \text{---} \text{---} \text{---} \end{array} \stackrel{(f)}{=} \frac{1}{4} \begin{array}{c} \text{---} \text{---} \text{---} \text{---} \text{---} \\ \text{---} \text{---} \text{---} \text{---} \text{---} \\ \text{---} \text{---} \text{---} \text{---} \text{---} \\ \text{---} \text{---} \text{---} \text{---} \text{---} \\ \text{---} \text{---} \text{---} \text{---} \text{---} \end{array}$$

For $(a_j, a_k) \notin \{(T_2, T_3), (T_3, T_2)\}$, the proof is almost the same as that of Eq. (33).

Now, let us consider the case when $(a_j, a_k) = (T_2, T_3)$. We can use the rules in Figure 4 as

follows.



□

F Proof of lemma 4

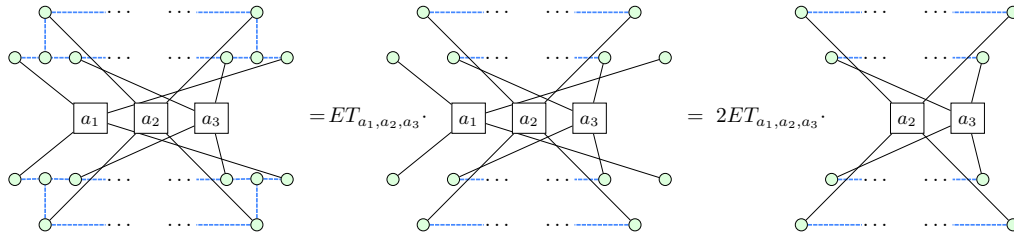
Lemma 4. *The following equation holds.*

$$\frac{1}{(2\pi)^3} \int_{\theta_1, \theta_2, \theta_3} \begin{array}{c} \theta_2 \dots \dots \dots -\theta_2 \\ \theta_1 \dots \theta_3 \dots \dots -\theta_3 \dots -\theta_1 \\ -\theta_1 \dots -\theta_3 \dots \dots \theta_3 \dots \theta_1 \\ -\theta_2 \dots \dots \dots \theta_2 \end{array} d\theta_1 d\theta_2 d\theta_3 = \sum_{a_1, a_2, a_3 \in \{T_1, T_2, T_3\}} ET_{a_1, a_2, a_3} \begin{array}{c} \dots \dots \dots \\ \dots \dots \dots \\ \dots \dots \dots \\ \dots \dots \dots \end{array} \begin{array}{c} a_1 \\ a_2 \\ a_3 \end{array}$$

Here ET_{a_1, a_2, a_3} is an element in the following $3 \times 3 \times 3$ tensor

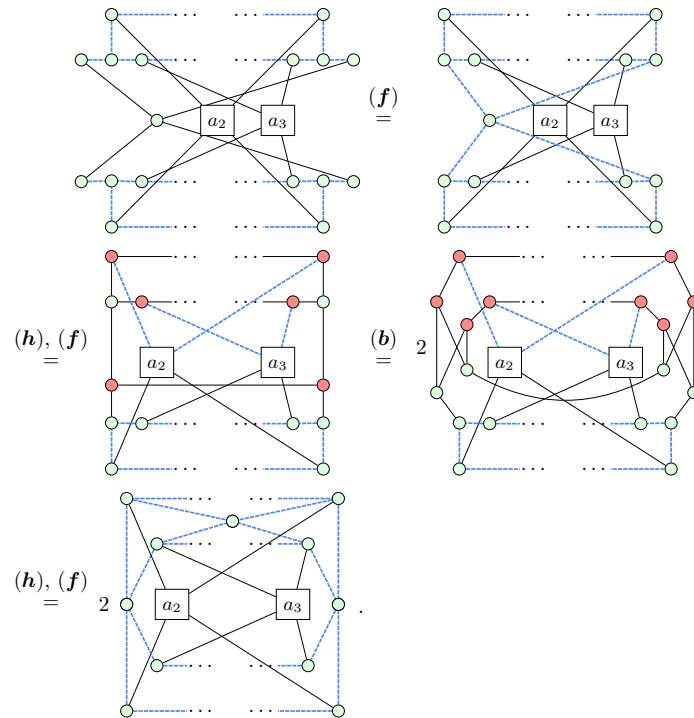
$$ET[1, \cdot, \cdot] = \frac{1}{8} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad ET[2, \cdot, \cdot] = \frac{1}{8} \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad ET[3, \cdot, \cdot] = \frac{1}{8} \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}.$$

Proof. By Lemma 2, it is sufficient to prove that

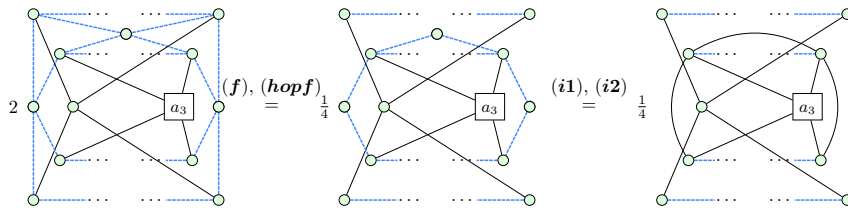


for $a_1, a_2, a_3 \in \{T_1, T_2, T_3\}$.

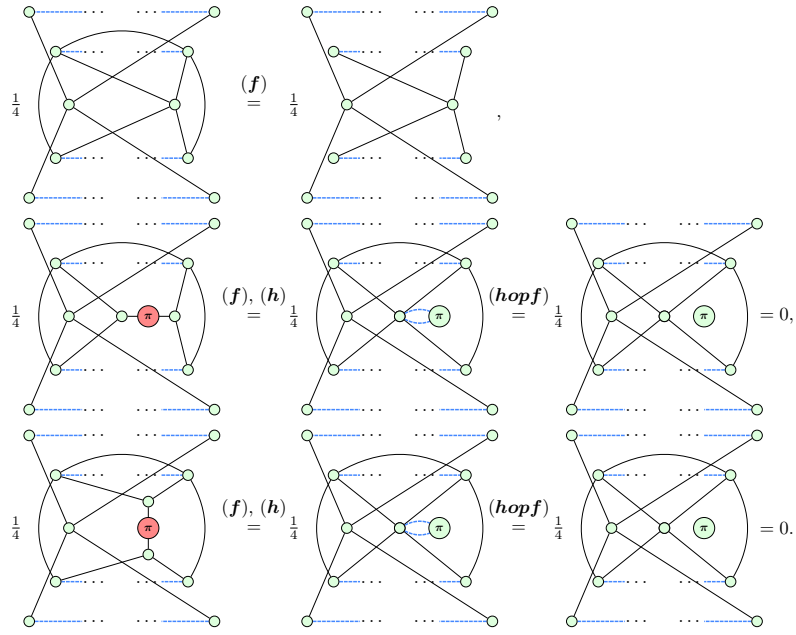
We first consider the case when $a_1 = T_1$. Since $a_1 = T_1$, we have



Now, if $a_2 = T_1$, then we have

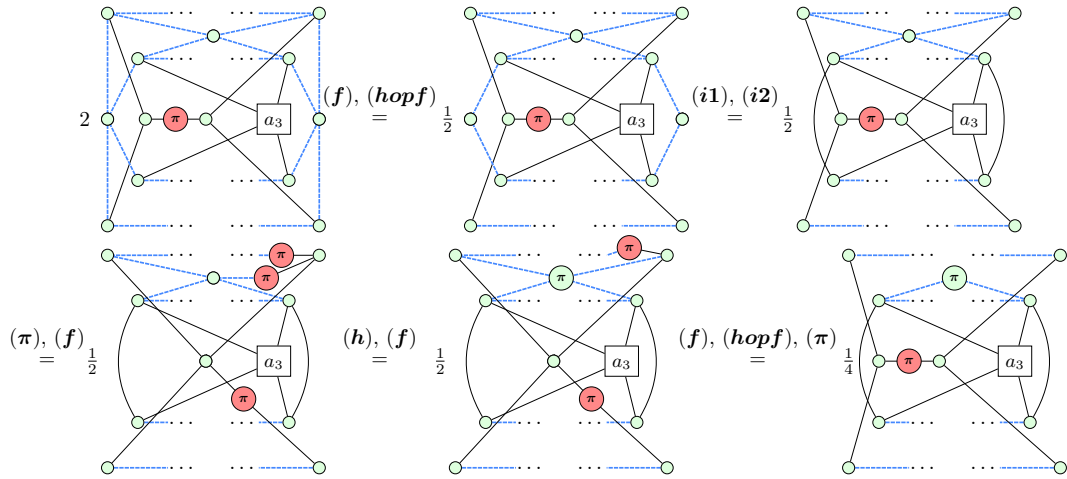


Hence,

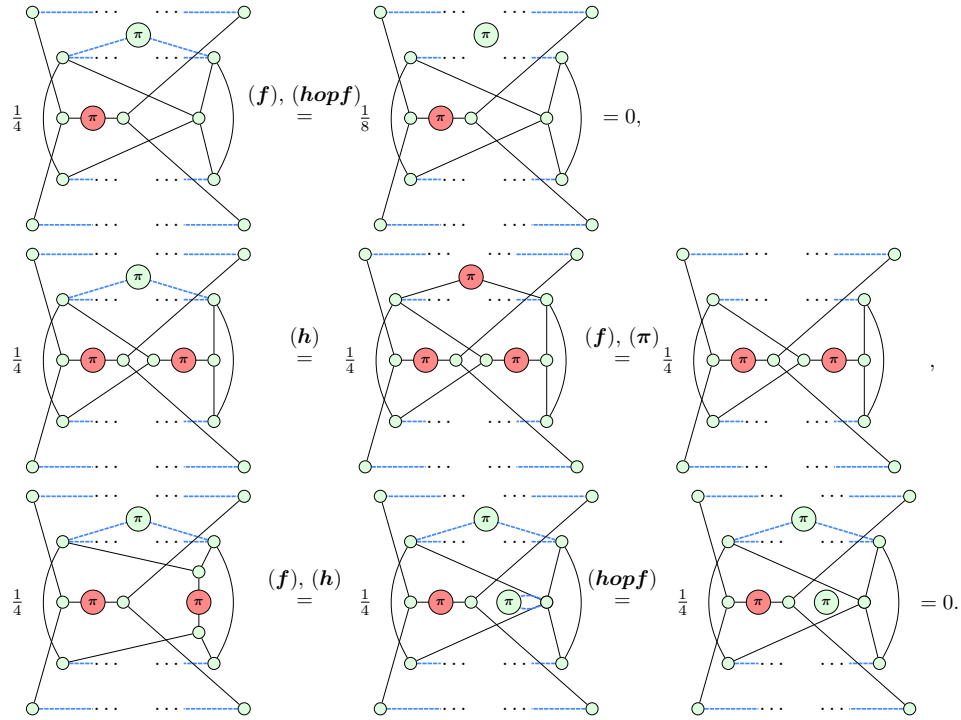


That is $ET[1, 1, \cdot] = \frac{1}{8} (1 \ 0 \ 0)$.

If $a_2 = T_2$, then

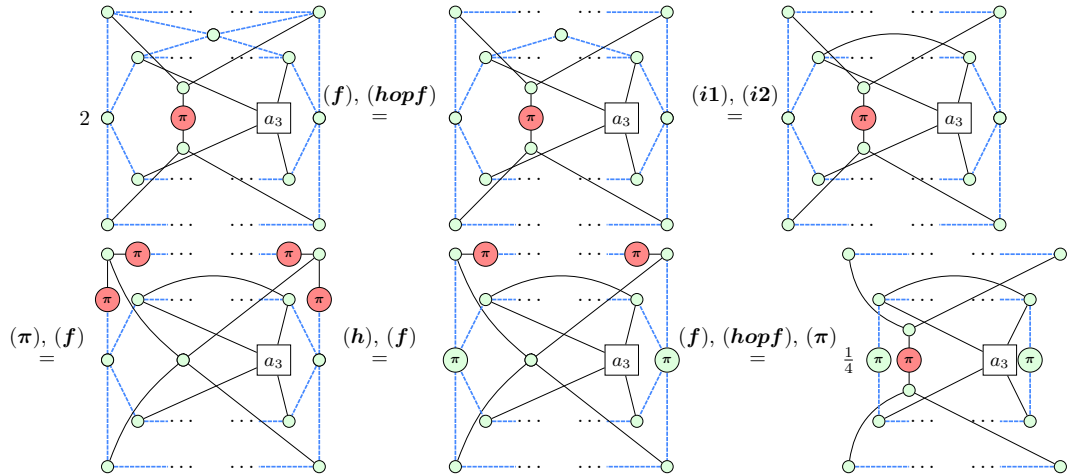


Hence,

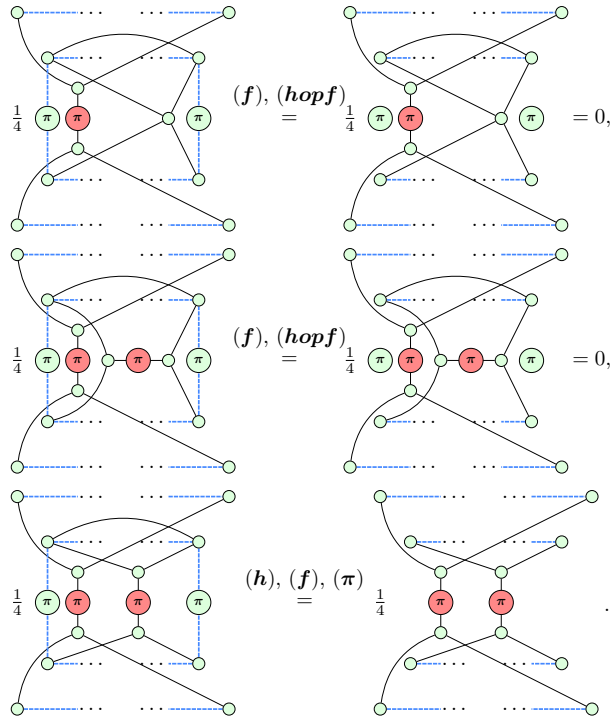


That is $ET[1, 2, \cdot] = \frac{1}{8} \begin{pmatrix} 0 & 1 & 0 \end{pmatrix}$.

If $a_2 = T_3$, then



Hence.

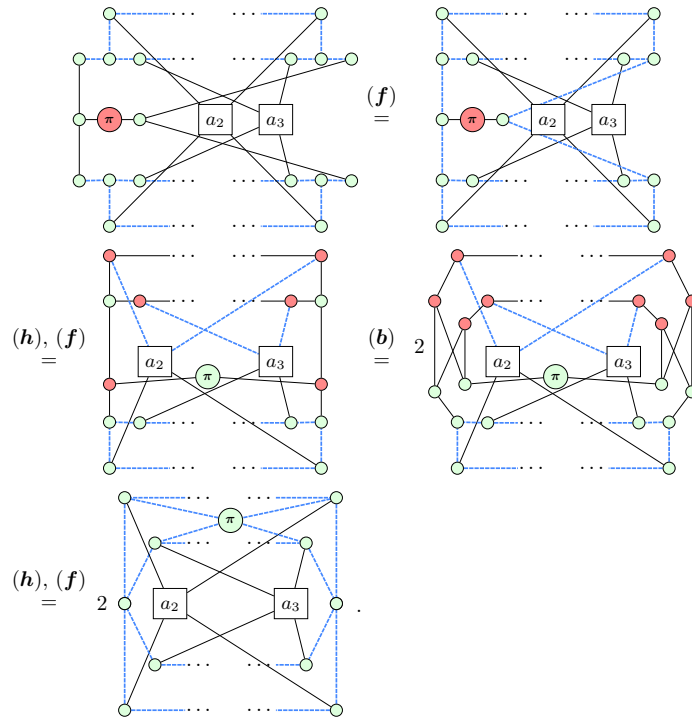


That is $ET[1, 3, \cdot] = \frac{1}{8} \begin{pmatrix} 0 & 0 & 1 \end{pmatrix}$.

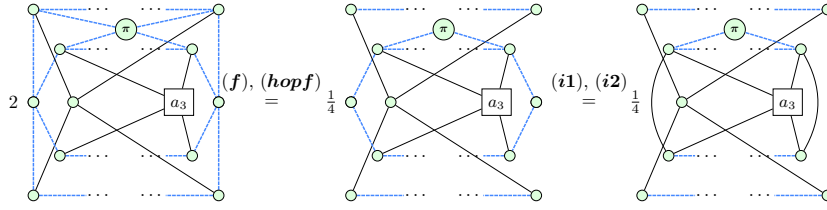
By now, we have proved that

$$ET[1, \cdot, \cdot] = \frac{1}{8} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}. \quad (34)$$

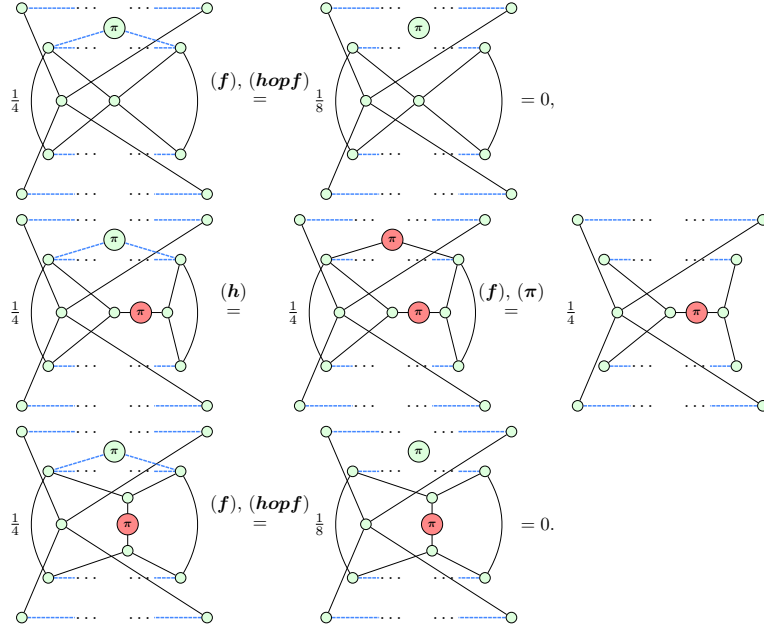
Now, let us consider the case when $a_1 = T_2$. Since $a_1 = T_2$, we have



Now, if $a_2 = T_1$, then we have

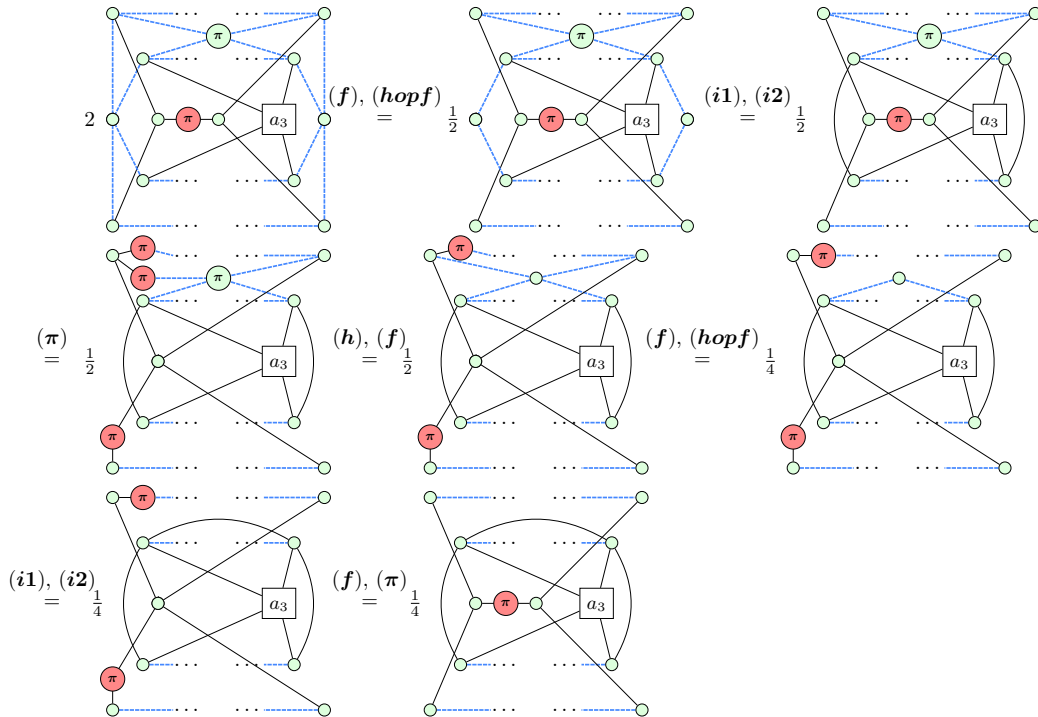


Hence,

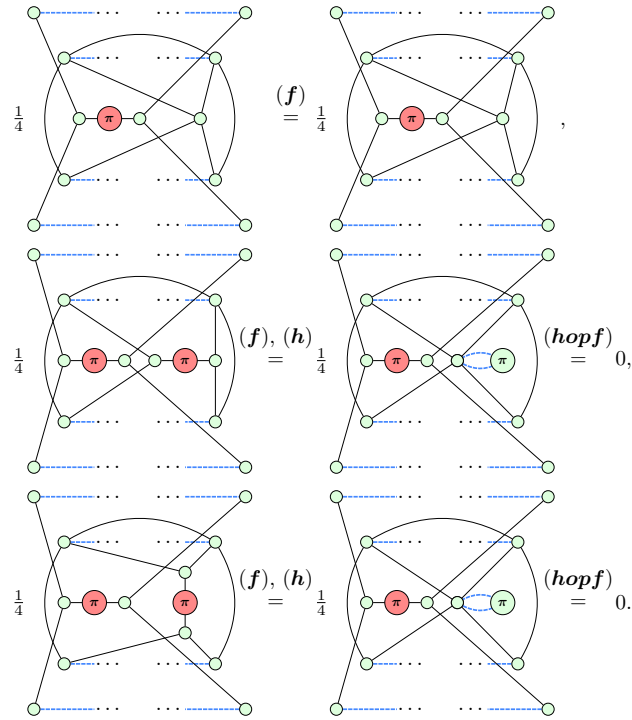


That is $ET[1, 1, \cdot] = \frac{1}{8} \begin{pmatrix} 0 & 1 & 0 \end{pmatrix}$.

If $a_2 = T_2$, then

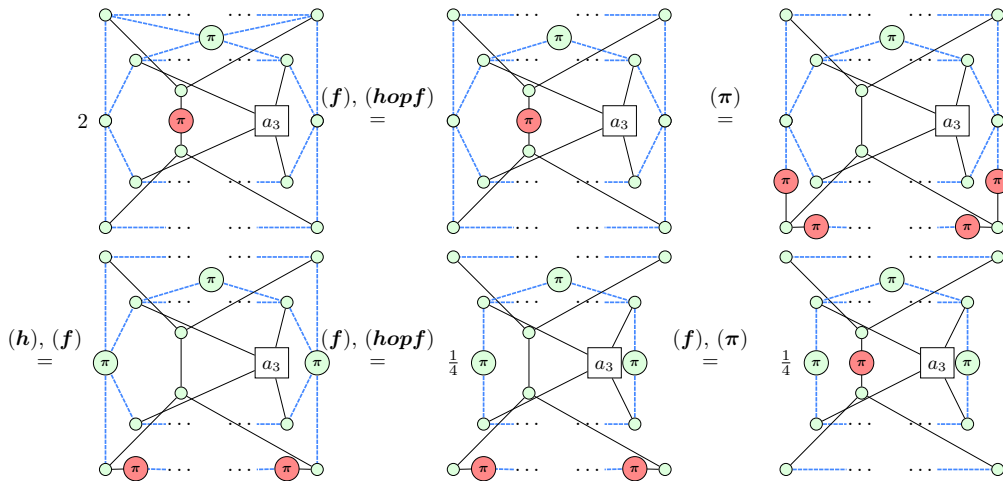


Hence,

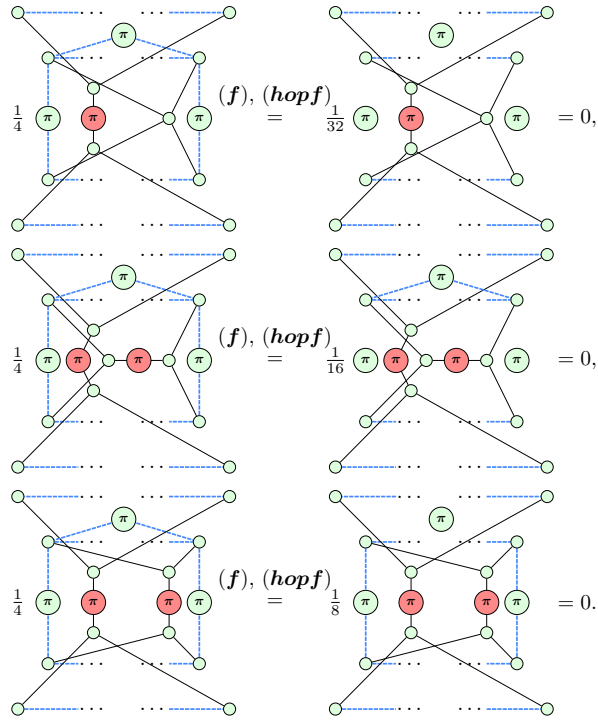


That is $ET[1, 2, \cdot] = \frac{1}{8} (1 \ 0 \ 0)$.

If $a_2 = T_3$, then



Hence.



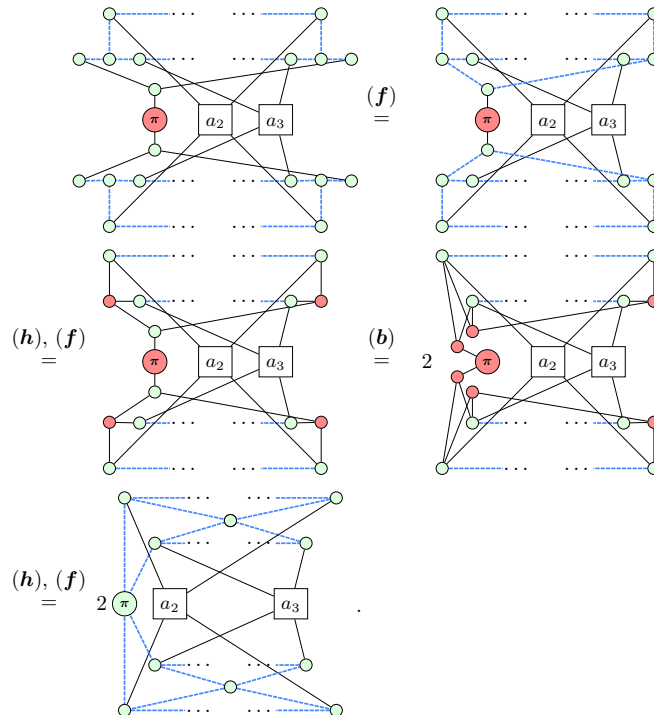
That is $ET[1, 3, \cdot] = \frac{1}{8} \begin{pmatrix} 0 & 0 & 0 \end{pmatrix}.$

By now, we have proved that

$$ET[2, \cdot, \cdot] = \frac{1}{8} \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}. \quad (35)$$

Let us consider the case when $a_1 = T_3.$

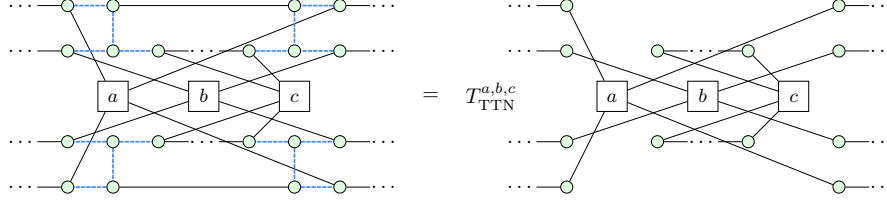
When $a_1 = T_3,$ we have



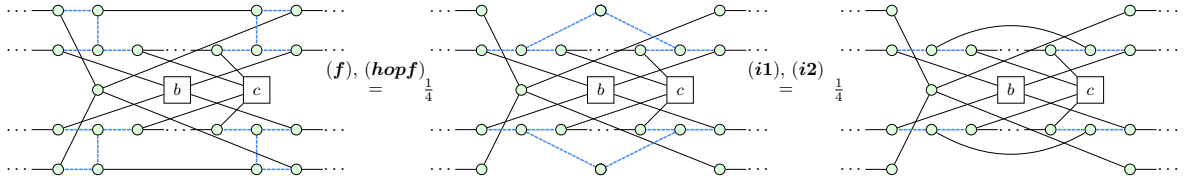
G.2 Tree tensor network ansatz

In Section 4.2, we used Eq. (23). Here we will prove this equation.

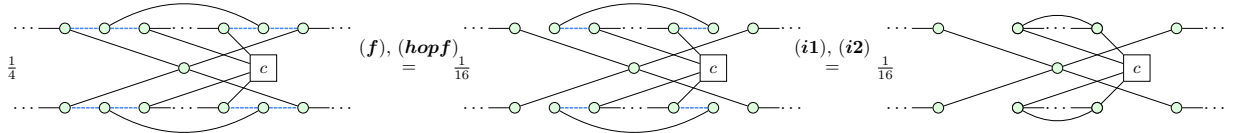
Proof of Eq. (23). By Lemma 2, it suffices to prove that



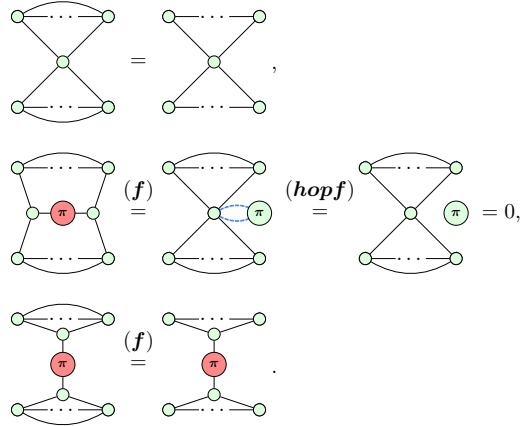
Similar to the proof of Lemma 4, we first consider the case when $a = T_1$. When $a = T_1$, we have



Now, a is disconnected with b and c . Hence, we can consider b and c separately. If $b = T_1$, then



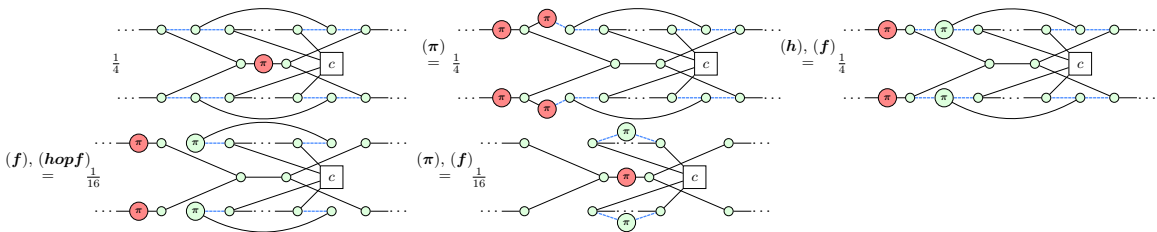
Hence,



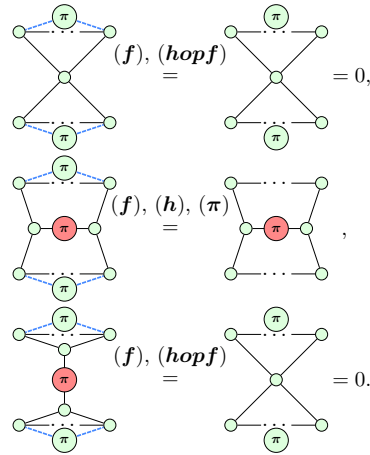
That it

$$T_{\text{TTN}}[1, 1, \cdot] = \frac{1}{16} \begin{pmatrix} 1 & 0 & 1 \end{pmatrix}.$$

If $b = T_2$, then



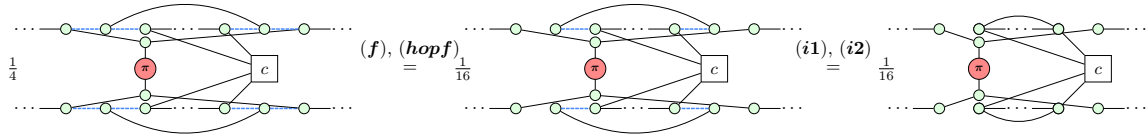
Hence,



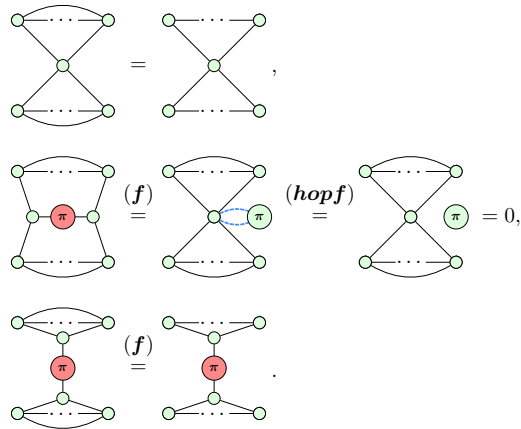
That it

$$T_{TTN}[1, 2, \cdot] = \frac{1}{16} \begin{pmatrix} 0 & 1 & 0 \end{pmatrix}.$$

If $b = T_3$, then



Hence,



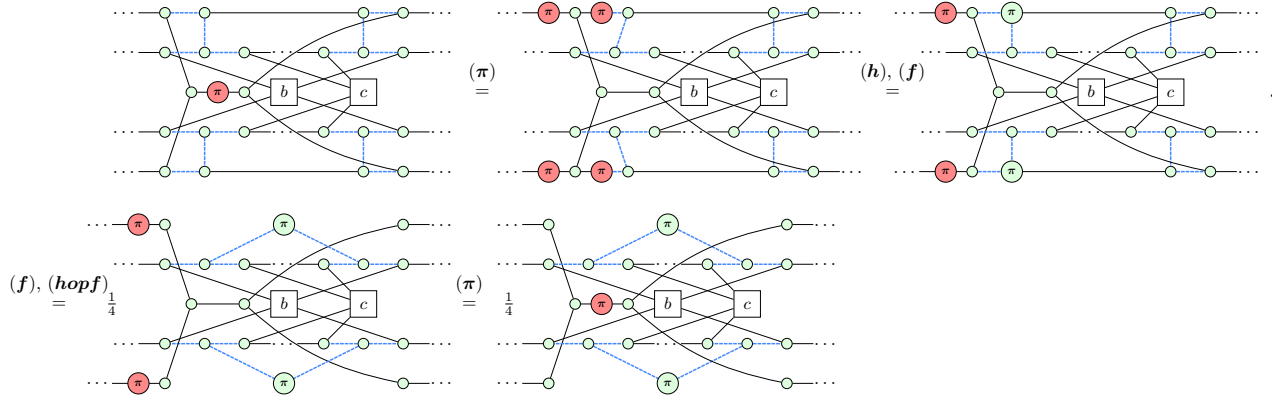
That it

$$T_{TTN}[1, 3, \cdot] = \frac{1}{16} \begin{pmatrix} 1 & 0 & 1 \end{pmatrix}.$$

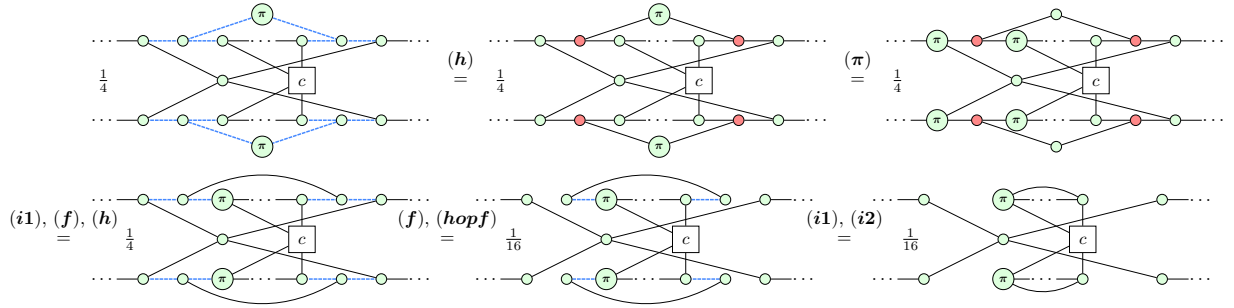
By now, we have proved that

$$T_{TTN}[1, \cdot, \cdot] = \frac{1}{16} \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}. \quad (39)$$

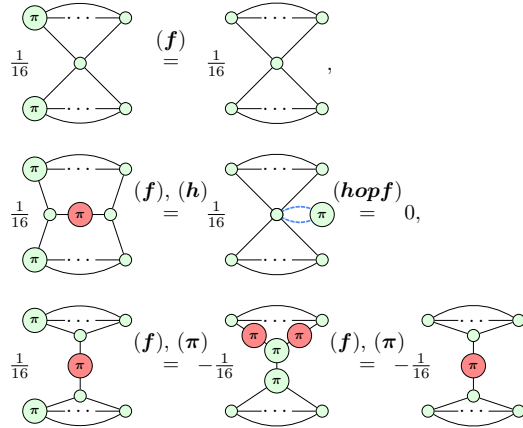
Now, let us consider the case when $a = T_2$. When $a = T_2$, we have



Now, a is disconnected with b and c . Hence, we can consider b and c individually. If $b = T_1$, then



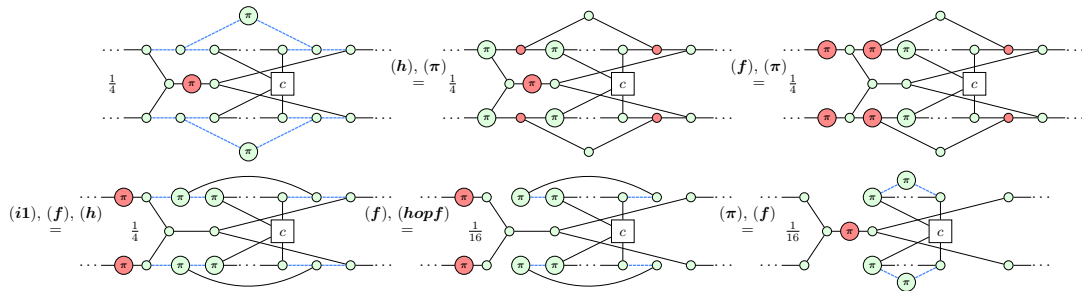
Hence,



That it

$$T_{\text{TTN}}[2, 1, \cdot] = \frac{1}{16} \begin{pmatrix} 1 & 0 & -1 \end{pmatrix}.$$

If $b = T_2$, then



Hence,

$$\begin{aligned}
 \frac{1}{16} \begin{array}{c} \pi \\ \vdots \\ \pi \end{array} & \stackrel{(f), (hopf)}{=} \frac{1}{16} \begin{array}{c} \pi \\ \vdots \\ \pi \end{array} = 0, \\
 \frac{1}{16} \begin{array}{c} \pi \\ \vdots \\ \pi \end{array} & \stackrel{(f), (\pi)}{=} \frac{1}{16} \begin{array}{c} \pi \\ \vdots \\ \pi \end{array}, \\
 \frac{1}{16} \begin{array}{c} \pi \\ \vdots \\ \pi \end{array} & \stackrel{(f), (hopf)}{=} \frac{1}{16} \begin{array}{c} \pi \\ \vdots \\ \pi \end{array} = 0.
 \end{aligned}$$

That it

$$T_{\text{TTN}}[2, 2, \cdot] = \frac{1}{16} \begin{pmatrix} 0 & 1 & 0 \end{pmatrix} .ow$$

If $b = T_3$, then

$$\begin{aligned}
 \frac{1}{4} \begin{array}{c} \pi \\ \vdots \\ \pi \end{array} & \stackrel{(h)}{=} \frac{1}{4} \begin{array}{c} \pi \\ \vdots \\ \pi \end{array} \stackrel{(\pi)}{=} \frac{1}{4} \begin{array}{c} \pi \\ \vdots \\ \pi \end{array} \\
 \stackrel{(i1), (f), (h)}{=} \frac{1}{4} \begin{array}{c} \pi \\ \vdots \\ \pi \end{array} & \stackrel{(f), (hopf)}{=} \frac{1}{16} \begin{array}{c} \pi \\ \vdots \\ \pi \end{array} \stackrel{(i1), (i2)}{=} \frac{1}{16} \begin{array}{c} \pi \\ \vdots \\ \pi \end{array} \\
 \stackrel{(f), (\pi)}{=} -\frac{1}{16} \begin{array}{c} \pi \\ \vdots \\ \pi \end{array}
 \end{aligned}$$

Hence,

$$\begin{aligned}
 -\frac{1}{16} \begin{array}{c} \pi \\ \vdots \\ \pi \end{array} & \stackrel{(f)}{=} -\frac{1}{16} \begin{array}{c} \pi \\ \vdots \\ \pi \end{array}, \\
 -\frac{1}{16} \begin{array}{c} \pi \\ \vdots \\ \pi \end{array} & \stackrel{(f), (h)}{=} -\frac{1}{16} \begin{array}{c} \pi \\ \vdots \\ \pi \end{array} \stackrel{(hopf)}{=} 0, \\
 -\frac{1}{16} \begin{array}{c} \pi \\ \vdots \\ \pi \end{array} & \stackrel{(f), (\pi)}{=} \frac{1}{16} \begin{array}{c} \pi \\ \vdots \\ \pi \end{array} \stackrel{(f), (\pi)}{=} \frac{1}{16} \begin{array}{c} \pi \\ \vdots \\ \pi \end{array}.
 \end{aligned}$$

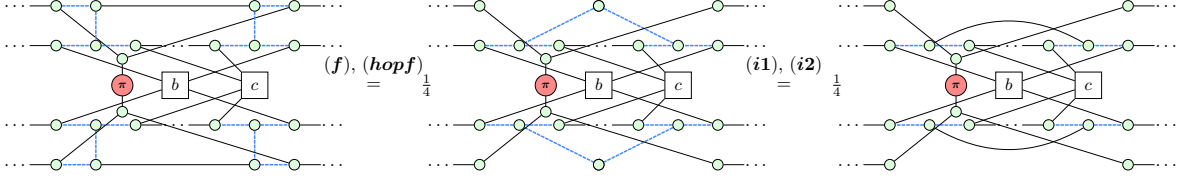
That it

$$T_{\text{TTN}}[2, 3, \cdot] = \frac{1}{16} \begin{pmatrix} -1 & 0 & 1 \end{pmatrix} .$$

By now, we have proved that

$$T_{\text{TTN}}[2, \cdot, \cdot] = \frac{1}{16} \begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & 0 \\ -1 & 0 & 1 \end{pmatrix}. \quad (40)$$

Now, let us consider the case when $a = T_3$. When $a = T_3$, we have



Now, a is disconnected with b and c . And the parts of b and c are the same as that of $a = T_1$. Hence, we have

$$T_{\text{TTN}}[3, \cdot, \cdot] = \frac{1}{16} \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}. \quad (41)$$

□

Now we are going to prove that there is a lower bound for the variance of gradients in the tree tensor network ansatz.

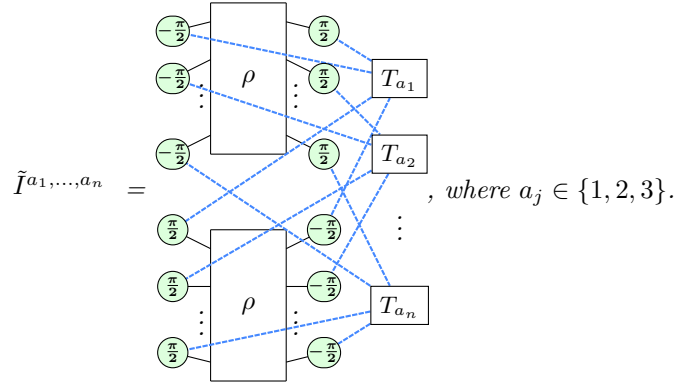
Theorem 6. For tree tensor network ansatz shown in (22), if

$$H = k_0 I + k_1 X + k_2 Y + k_3 Z,$$

then we have

$$\text{Var} \left(\frac{\partial \langle H \rangle}{\partial \theta} \right) \geq \frac{k_1^2 + k_3^2}{n^2} \tilde{I}(u_1, \dots, u_n) + \frac{k_2^2}{n^2} \tilde{I}(w_1, \dots, w_n), \quad (27)$$

for some $u_j, w_j \in \{v_{1,3}, v_2, v_{1,3}^-\}$. Here $\tilde{I}(u_1, \dots, u_n)$ is defined in Eq. (26). And \tilde{I} is a 3^n dimensional tensor which only depends on the input state. If the input state is ρ , then \tilde{I} is defined as follows.



Proof. We first prove that

$$\tilde{I}(u_1, \dots, u_n) \geq 0, \quad u_j \in \{v_2, v_{1,3}, v_{1,3}^-\},$$

where

$$v_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \quad v_{1,3} = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}, \quad v_{1,3}^- = \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix}.$$

Note that if u_1 is $v_{1,3}$ or $v_{1,3}^-$, then

$$\tilde{I}(v_{1,3}, \dots) = \begin{array}{c} \begin{array}{c} \text{---} \frac{\pi}{2} \text{---} \\ \text{---} \frac{\pi}{2} \text{---} \\ \vdots \\ \text{---} \frac{\pi}{2} \text{---} \\ \vdots \\ \text{---} \frac{\pi}{2} \text{---} \\ \vdots \\ \text{---} \frac{\pi}{2} \text{---} \\ \vdots \\ \text{---} \frac{\pi}{2} \text{---} \end{array} \begin{array}{c} \rho \\ \vdots \\ \rho \end{array} \begin{array}{c} \frac{\pi}{2} \text{---} \\ \vdots \\ \frac{\pi}{2} \text{---} \\ \vdots \\ \frac{\pi}{2} \text{---} \\ \vdots \\ \frac{\pi}{2} \text{---} \\ \vdots \\ \frac{\pi}{2} \text{---} \end{array} \begin{array}{c} T_1 + T_3 \\ \vdots \\ \dots \end{array} \end{array}, \quad \tilde{I}(v_{1,3}^-, \dots) = \begin{array}{c} \begin{array}{c} \text{---} \frac{\pi}{2} \text{---} \\ \text{---} \frac{\pi}{2} \text{---} \\ \vdots \\ \text{---} \frac{\pi}{2} \text{---} \\ \vdots \\ \text{---} \frac{\pi}{2} \text{---} \\ \vdots \\ \text{---} \frac{\pi}{2} \text{---} \\ \vdots \\ \text{---} \frac{\pi}{2} \text{---} \end{array} \begin{array}{c} \rho \\ \vdots \\ \rho \end{array} \begin{array}{c} \frac{\pi}{2} \text{---} \\ \vdots \\ \frac{\pi}{2} \text{---} \\ \vdots \\ \frac{\pi}{2} \text{---} \\ \vdots \\ \frac{\pi}{2} \text{---} \\ \vdots \\ \frac{\pi}{2} \text{---} \end{array} \begin{array}{c} T_1 - T_3 \\ \vdots \\ \dots \end{array} \end{array}.$$

By

$$\begin{aligned} T_1 + T_3 &= \begin{array}{c} \diagup \quad \diagdown \\ \text{---} \text{---} \\ \diagdown \quad \diagup \end{array} + \begin{array}{c} \text{---} \\ \text{---} \pi \text{---} \\ \text{---} \end{array} = |00\rangle\langle 00| + |10\rangle\langle 10| + |01\rangle\langle 01| + |11\rangle\langle 11|, \\ &= (|0\rangle\langle 0| + |1\rangle\langle 1|) \otimes (|0\rangle\langle 0| + |1\rangle\langle 1|) = \begin{array}{c} \text{---} \text{---} \\ \text{---} \text{---} \end{array}, \end{aligned}$$

$$\begin{aligned} T_1 - T_3 &= \begin{array}{c} \diagup \quad \diagdown \\ \text{---} \text{---} \\ \diagdown \quad \diagup \end{array} + \begin{array}{c} \text{---} \\ \text{---} \pi \text{---} \\ \text{---} \end{array} = |00\rangle\langle 00| - |10\rangle\langle 10| - |01\rangle\langle 01| + |11\rangle\langle 11| \\ &= (|0\rangle\langle 0| - |1\rangle\langle 1|) \otimes (|0\rangle\langle 0| - |1\rangle\langle 1|) = \begin{array}{c} \text{---} \pi \text{---} \\ \text{---} \pi \text{---} \end{array}, \end{aligned}$$

$$T_2 = \begin{array}{c} \diagup \quad \text{---} \pi \text{---} \quad \diagdown \\ \text{---} \quad \quad \quad \text{---} \\ \diagdown \quad \quad \quad \diagup \end{array} = \begin{array}{c} \text{---} |0\rangle\langle 1| \text{---} \\ \text{---} |0\rangle\langle 1| \text{---} \end{array} + \begin{array}{c} \text{---} |1\rangle\langle 0| \text{---} \\ \text{---} |1\rangle\langle 0| \text{---} \end{array},$$

we can expand $\tilde{I}(u_1, \dots, u_n)$ as a sum of squares. Thus, we have proved that

$$\tilde{I}(u_1, \dots, u_n) \geq 0.$$

Now let us consider the lower bound of $\mathbf{Var}\left(\frac{\partial\langle H \rangle}{\partial\theta}\right)$. By the graph-like ZX-diagram just obtained, we have $\tilde{H} = \begin{pmatrix} \tilde{H}_1 \\ \tilde{H}_2 \\ \tilde{H}_3 \end{pmatrix}$ which is defined as follows.

$$\tilde{H}_j = \begin{array}{c} \begin{array}{c} \text{---} \frac{\pi}{2} \text{---} \\ \text{---} \frac{\pi}{2} \text{---} \\ \vdots \\ \text{---} \frac{\pi}{2} \text{---} \\ \vdots \\ \text{---} \frac{\pi}{2} \text{---} \\ \vdots \\ \text{---} \frac{\pi}{2} \text{---} \end{array} \begin{array}{c} H \\ \vdots \\ H \end{array} \begin{array}{c} \frac{\pi}{2} \text{---} \\ \vdots \\ \frac{\pi}{2} \text{---} \\ \vdots \\ \frac{\pi}{2} \text{---} \\ \vdots \\ \frac{\pi}{2} \text{---} \end{array} \begin{array}{c} T_j \\ \vdots \\ T_j \end{array} \end{array}, \quad \text{for } j = 1, 2, 3.$$

Suppose that $H = k_0I + k_1X + k_2Y + k_3Z$. Then we can obtain

$$\tilde{H} = 2k_0^2v_{1,3} + 2(k_1^2 + k_3^2)v_2 + 2k_2^2v_{1,3}^-. \quad (42)$$

And by the definition of T_{TTN} , we can obtain Eq. (25). Hence, we can expand the variance as

$$\mathbf{Var}\left(\frac{\partial\langle H \rangle}{\partial\theta}\right) = \sum_{u_j \in \{v_2, v_{1,3}, v_{1,3}^-\}} a(u_1, \dots, u_n) \cdot \tilde{I}(u_1, \dots, u_n),$$

where $a(u_1, \dots, u_n)$ is a non-negative number. We will prove that there exists one choice of (u_1, \dots, u_n) such that

$$a(u_1, \dots, u_n) \in \Omega\left(\frac{1}{\text{poly}(n)}\right).$$

When we use Eq. (25) to expand the variance, the only case that will cause a coefficient < 1 is the case when we use the second equation in Eq. (25). Hence, the coefficients $a(u_1, \dots, u_n)$ only depend on the number of times that we use the second equation. And it depends on the location of θ . The worst case is that θ is the parameter of the first gate applying to the first qubit. In this case, we need to use the second equation in Eq. (25) for $2 \log(n) - 1$ times. That means we have a lower bound for the variance

$$\text{Var}\left(\frac{\partial \langle H \rangle}{\partial \theta}\right) \geq \frac{k_1^2 + k_3^2}{n^2} \tilde{I}(u_1, \dots, u_n) + \frac{k_2^2}{n^2} \tilde{I}(w_1, \dots, w_n),$$

for some $u_j, w_j \in \{v_{1,3}, v_2, v_{1,3}^-\}$.

Note that $I(u_1, \dots, u_n)$ only depends on the input state ρ . If

$$\tilde{I}(u_1, \dots, u_n) \in \Omega\left(\frac{1}{\text{poly}(n)}\right) \text{ or } \tilde{I}(w_1, \dots, w_n) \in \Omega\left(\frac{1}{\text{poly}(n)}\right),$$

then there exists no BP in the tree tensor network ansatz. \square

G.3 QCNN

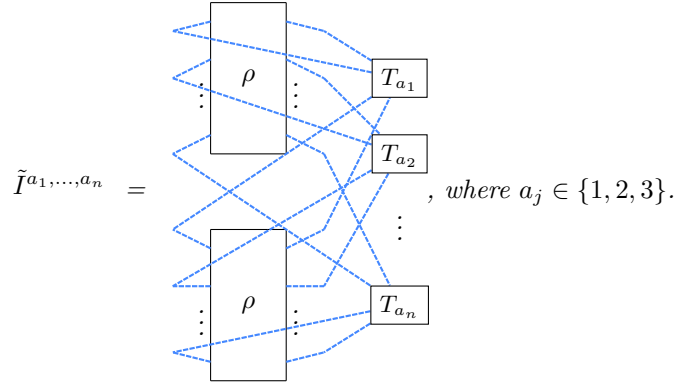
Theorem 7. For the QCNN ansatz shown in (28), if

$$H = k_0 I + k_1 X + k_2 Y + k_3 Z,$$

then we have

$$\text{Var}\left(\frac{\partial \langle H \rangle}{\partial \theta}\right) \geq \frac{k_2^2 + k_3^2}{n^9} \tilde{I}(u_1, \dots, u_n) + \frac{k_1^2}{n^9} \tilde{I}(w_1, \dots, w_n), \quad (30)$$

for some $u_j, w_j \in \{v_{1,3}, v_2, v_{1,3}^-\}$. Here \tilde{I} is a 3^n dimensional tensor which only depends on the input state ρ .



Proof. The proof is similar to that of the tree tensor network ansatz (Theorem 6), so we will only give a sketch of the proof. We also have

$$\tilde{I}(u_1, \dots, u_n) \geq 0, \quad u_j \in \{v_{1,3}, v_2, v_{1,3}^-\}.$$

From the graph-like ZX-diagram, we have

$$\tilde{H}_j = \begin{array}{c} \circ \quad \square \quad \circ \\ \quad \quad H \quad \quad \\ \quad \quad \quad \quad \quad \quad \\ \circ \quad \square \quad \circ \\ \quad \quad T_j \quad \quad \\ \quad \quad \quad \quad \quad \quad \\ \circ \quad \square \quad \circ \\ \quad \quad H \quad \quad \end{array}, \text{ for } j = 1, 2, 3.$$

Hence,

$$\tilde{H} = 2k_0^2 v_{1,3} + 2(k_1^2 + k_2^2)v_2 + 2k_3^2 v_{1,3}^-.$$

We will analyze each term of \tilde{H} in the variance.

Using Eq. (29), the term $2k_0^2 v_{1,3}$ will become 0 by $P_2 v_{1,3} = 0$. The terms $2(k_1^2 + k_2^2)v_2$ and $2k_3^2 v_{1,3}^-$ will generate terms containing v_2 after expanding using Eq. (29). And each time after generating terms containing v_2 , a coefficient $\geq \frac{1}{8}$ will be multiplied to the variance. Hence, if we want to bring v_2 to P_2 , we need to generate terms containing v_2 for l times, where l is a path from the location of v_2 to the location of P_2 .

By the structure of the QCNN ansatz, we have

$$l \geq 3 \log(n).$$

It will generate a coefficient $\geq \frac{1}{8^{3 \log(n)}}$.

Hence we have

$$\mathbf{Var} \left(\frac{\partial \langle H \rangle}{\partial \theta} \right) \geq \frac{k_2^2 + k_3^2}{n^9} \tilde{I}(u_1, \dots, u_n) + \frac{k_1^2}{n^9} \tilde{I}(w_1, \dots, w_n),$$

for some $u_j, w_j \in \{v_{1,3}, v_2, v_{1,3}^-\}$. □