Statistical complexity of quantum circuits

Kaifeng Bu [1,*] Dax Enshan Koh [2,†] Lu Li, [3,4,‡] Qingxian Luo, [4,5,§] and Yaobo Zhang [6,7,‖]

[1] *Department of Physics, Harvard University, Cambridge, Massachusetts 02138, USA*
[2] *Institute of High Performance Computing, Agency for Science, Technology and Research (A*STAR), 1 Fusionopolis Way, #16-16 Connexis, Singapore 138632, Singapore*
[3] *Department of Mathematics, Zhejiang Sci-Tech University, Hangzhou, Zhejiang 310018, China*
[4] *School of Mathematical Sciences, Zhejiang University, Hangzhou, Zhejiang 310027, China*
[5] *Center for Data Science, Zhejiang University, Hangzhou, Zhejiang 310027, China*
[6] *Zhejiang Institute of Modern Physics, Zhejiang University, Hangzhou, Zhejiang 310027, China*
[7] *Department of Physics, Zhejiang University, Hangzhou, Zhejiang 310027, China*

(Received 12 December 2021; accepted 11 May 2022; published 17 June 2022)

In theoretical machine learning, the statistical complexity is a notion that measures the richness of a hypothesis space. In this work, we apply a particular measure of statistical complexity, namely, the Rademacher complexity, to the quantum circuit model in quantum computation and study how the statistical complexity depends on various quantum circuit parameters. In particular, we investigate the dependence of the statistical complexity on the resources, depth, width, and the number of input and output registers of a quantum circuit. To study how the statistical complexity scales with resources in the circuit, we introduce a magic resource measure based on the (p, q) group norm, which quantifies the amount of magic resource in the quantum channels associated with the circuit. These dependencies are investigated in the following two settings: (i) where the entire quantum circuit is treated as a single quantum channel, and (ii) where each layer of the quantum circuit is treated as a separate quantum channel. The bounds we obtain can be used to constrain the capacity of quantum neural networks in terms of their depths and widths as well as the resources in the network.

DOI: 10.1103/PhysRevA.105.062431

I. INTRODUCTION

Owing to its ability to recognize and analyze patterns in data and use them to make predictions, deep learning—a subfield of machine learning—has made a profound impact on the computing industry [1–3] and has found applications in a myriad of fields, including natural language processing [4–6], drug design [7,8], fraud detection [9,10], medical image analysis [11,12], self-driving cars [13,14], handwriting recognition [15,16], and computer vision [17,18]. A central object in many deep learning models is the neural network, an interconnected collection of nodes that can learn from data and model relationships between them [19]. Different neural networks differ in terms of their ability to learn from data, and understanding this difference is a key problem in theoretical machine learning. This ability of neural networks has been quantified by various statistical complexity measures, including the Vapnik-Chervonenkis (VC) dimension [20,21], the metric entropy [22], the Gaussian complexity [23], and the Rademacher complexity [23]. The dependence of these measures on various structure parameters of the neural network, such as its depth and width and the number of parameters in the neural network, has been studied in a number of papers [24–29].

In addition to the progress in deep learning, the last decade also saw rapid developments in quantum computing [30]. With the development of noisy intermediate-scale quantum (NISQ) hardware [31] as well as near-term quantum algorithms like the variational quantum eigensolver (VQE) [32] and the quantum approximate optimization algorithm (QAOA) [33,34], there are expectations that quantum computers are poised to revolutionize computation by speeding up the solutions of certain practical computational problems [35]. Major experimental milestones in this direction include the recent demonstrations of quantum computational supremacy [36,37] (also called quantum advantage [38]), defined to be an event in which a quantum computer empirically solves a computational problem deemed intractable for classical computers, independent of the practical value of the problem [39–42].

At the intersection of deep learning and quantum computing is the field of quantum deep learning, which has the quantum neural network—the quantum generalization of the classical neural network—as one of its central objects [43–48]. Quantum deep learning has been explored as an application of quantum machine learning, which has gained significant interest of late [49–53]. Compared with the classical neural networks, however, considerably less is known about quantum neural networks and characterizations of their

*kfbu@fas.harvard.edu
†dax_koh@ihpc.a-star.edu.sg
‡lilu93@zju.edu.cn
§luoqingxian@zju.edu.cn
‖yaobozhang@zju.edu.cn

statistical complexities. For example, the following question has hitherto remained largely unaddressed: how does the statistical complexity of quantum neural networks depend on the structure parameters of the quantum circuit underlying it as well as the amount of certain resources it contains?

In this paper, we address the above gap by characterizing the statistical complexity of quantum circuits in terms of their Rademacher complexity. To characterize the dependence of Rademacher complexity on resources in the framework of quantum resource theories [54,55], we introduce a magic resource measure [56] for quantum channels based on the (p, q) group norm. We consider the Rademacher complexity of quantum circuits in two different settings. First, we consider the case where the entire quantum circuit is treated as a single quantum channel independent of its depth or width. In this case, we find a bound for the statistical complexity that depends on a magic resource measure as well as the number of input and output qubits. Second, we consider the case where each layer of the quantum circuit is treated as a separate quantum channel. In this case, we find a bound for the statistical complexity that depends not only on the magic resource measure but also on the depth and width of the quantum circuit.

II. MAIN RESULTS

Consider m independent samples $S = (\vec{x}_1, \ldots, \vec{x}_m)$, where each $\vec{x}_i$ is encoded as a quantum state $|\psi(\vec{x}_i)\rangle$. After a quantum circuit C (e.g., C could be an instance of a variational quantum circuit or a quantum neural network) is applied to the quantum state $|\psi(\vec{x}_i)\rangle$ and a (Hermitian) observable H is measured on the output, the expected measurement outcome is given by

$$f_C(\vec{x}_i) = \mathrm{Tr}[C(|\psi(\vec{x}_i)\rangle\langle\psi(\vec{x}_i)|)H]. \tag{1}$$

In this way, each quantum circuit C defines a real-valued function f_C. Let $\mathcal{F} \circ \mathcal{C} := \{f_C : C \in \mathcal{C}\}$ denote the function class defined by the set of quantum circuits $\mathcal{C}$.

Consider the hypothesis space $\mathcal{H} = \mathcal{F} \circ \mathcal{C}$, where $\mathcal{C}$ is a given set of quantum circuits. Given m independent samples $\{(\vec{x}_i, y_i)\}_{i=1}^m$, where each $(\vec{x}_i, y_i)$ is taken to be i.i.d. from some unknown probability distribution D on some $\mathcal{X} \times \mathcal{Y}$, let us consider a loss function $l : \mathcal{Y} \times \mathcal{Y} \to \mathbb{R}$. The goal of the learning task is to find some function in the hypothesis space that minimizes the expected error:

$$L(f) = \mathbb{E}_{(\vec{x},y)\sim D}\, l(f(\vec{x}), y). \tag{2}$$

As we have access to only the m independent samples $\{(\vec{x}_i, y_i)\}_{i=1}^m$, one strategy is to find some function in hypothesis space to minimize the empirical error:

$$\hat{L}(f) = \frac{1}{m}\sum_{i=1}^{m} l(f(\vec{x}_i), y_i). \tag{3}$$

The difference between the empirical and expected error is called the *generalization error*, which determines the performance of the hypothesis function f on the unseen data drawn from the unknown probability distribution.

The Rademacher complexity is a measure of the richness of a hypothesis space and can be used to provide bounds on the generalization error associated with learning from training data [23,57]. Let us consider the Rademacher complexity of $\mathcal{F} \circ \mathcal{C}$ on m independent samples $S = \{\vec{x}_1, \ldots, \vec{x}_m\}$, defined as

$$R_S(\mathcal{F} \circ \mathcal{C}) = \mathbb{E}_{\vec{\epsilon}}\, \frac{1}{m} \sup_{C\in\mathcal{C}} \left|\sum_i \epsilon_i f_C(\vec{x}_i)\right|, \tag{4}$$

where each ϵ_i in the expectation above is a Rademacher random variable, which takes the values ± 1 with equal probability 1/2. Here, we use the Rademacher complexity as a measure of the statistical complexity of the hypothesis space $\mathcal{F} \circ \mathcal{C}$.

In this work, we also consider the connection between the Rademacher complexity and the magic resource, where the latter is a well-known resource that has been used to quantify quantum computational advantage. Let $\mathcal{P}_n$ denote the set of n-qubit Hermitian Pauli operators, i.e., operators that can be written as the n-fold tensor product of the single-qubit Pauli operators $\{I, X, Y, Z\}$ with sign $\pm$. The *Clifford unitaries* U on n qubits are defined as the unitaries that map Pauli operators to Pauli operators; that is, $UPU^\dagger \in \mathcal{P}_n$ for any $P \in \mathcal{P}_n$. An important result about the Clifford unitaries is the *Gottesman-Knill theorem*, which states that quantum circuits with Clifford unitaries acting on $|0\rangle^{\otimes n}$ and with measurements in Pauli Z basis can be simulated efficiently on a classical computer [58]. Hence, such circuits cannot yield a quantum advantage; in other words, to obtain a quantum speedup, we need unitary operators that are not Clifford. To this end, the magic property has been proposed as a way to characterize how far away a quantum gate is from the set of Clifford unitaries [56,59–62].

A. Rademacher complexity of quantum channels

1. *Rademacher complexity of arbitrary quantum channel*

Given a quantum channel $\Phi : \mathcal{L}((\mathbb{C}^2)^{\otimes n_1}) \to \mathcal{L}((\mathbb{C}^2)^{\otimes n_2})$ from n_1 qubits to n_2 qubits, we define the $4^{n_2} \times 4^{n_1}$ *representation matrix* M^Φ of Φ to be the matrix whose entries are given by

$$M^\Phi_{\vec{z}\vec{x}} = \frac{1}{2^{n_2}}\mathrm{Tr}[P_{\vec{z}}\Phi(P_{\vec{x}})], \tag{5}$$

where $\vec{x} \in \{0, 1, 2, 3\}^{n_1}$, $\vec{z} \in \{0, 1, 2, 3\}^{n_2}$, and $P_{\vec{x}}, P_{\vec{z}}$ are the corresponding Pauli operators. For any Hermitian operator P, the *representation vector* $\vec{\alpha}^P$ of P is defined as

$$\alpha^P_{\vec{z}} = \frac{1}{2^n}\mathrm{Tr}[P_{\vec{z}}P]. \tag{6}$$

For any $N_1 \times N_2$ matrix M, which can be treated as a column of N_1 row vectors, the (p, q) *group norm* of M, where $0 < p, q \leqslant \infty$, is defined as $\|M\|_{p,q} = (\frac{1}{N_1}\sum_i \|M_i\|_p^q)^{1/q}$, where the l_p norm of the ith row vector $\|M_i\|_p$ is defined as $\|M_i\|_p = (\sum_{j=1}^{N_2} |M_{ij}|^p)^{1/p}$. Of interest to us is the (p, q) group norm of the representation matrix of quantum channels. As we shall show in Appendix A, the (p, q) group norm of the representation matrix of quantum gates can be used as a resource measure to quantify the amount of magic resource in the quantum gates.

Here, we treat the entire quantum circuit as a single quantum channel. Let us define $\mathcal{C}^{n_0,n_1}_{\|\cdot\|_{p,q}\leqslant\mu}$ to be the set of quantum circuits from n_0 qubits to n_1 qubits that have a (p, q) group norm bounded by μ.

Theorem 1. Given the set of quantum circuits C from n_0 qubits to n_1 qubits with bounded (p, q) norm $\|\cdot\|_{p,q}$, the Rademacher complexity of $\mathcal{F}\circ\mathcal{C}^{n_0,n_1}_{\|\cdot\|_{p,q}\leqslant\mu}$ on m independent samples $S = \{\vec{x}_1, \dots, \vec{x}_m\}$ is bounded as follows:

(1) For $1 \leqslant p \leqslant 2$, we have

$$R_S\big(\mathcal{F}\circ\mathcal{C}^{n_0,n_1}_{\|\cdot\|_{p,q}\leqslant\mu}\big) \leqslant \mu 4^{n_1\max\{\frac{1}{p^*},\frac{1}{q}\}}\frac{\sqrt{\min\{p^*, 8n_0\}}}{\sqrt{m}}K_p(S, H). \tag{7}$$

(2) For $2 < p < \infty$, we have

$$R_S\big(\mathcal{F}\circ\mathcal{C}^{n_0,n_1}_{\|\cdot\|_{p,q}\leqslant\mu}\big) \leqslant \mu 4^{n_1\max\{\frac{1}{p^*},\frac{1}{q}\}}\frac{\sqrt{p^*}}{m^{1/p}}K_p(S, H), \tag{8}$$

where p^* is the Hölder conjugate of p, i.e., $\frac{1}{p}+\frac{1}{p^*}=1$:

$$K_p(S, H) = \|\vec{\alpha}\|_p \max_i \|\vec{f}_I(\vec{x}_i)\|_{p^*}; \tag{9}$$

and $\vec{\alpha}$ and $\vec{f}_I(\vec{x}_i)$ are the representation vectors of H and $|\psi(x_i)\rangle\langle\psi(x_i)|$ in the Pauli basis, respectively.

This result provides an upper bound on the Rademacher complexity of quantum circuits that depends on the amount of magic resource and the number of input and output qubits. (See Appendix A for a proof of Theorem 1.)

2. *Rademacher complexity of unital quantum channels*

We now consider the special case where the quantum channel Φ is unital, i.e., $\Phi(\mathbb{I}) = \mathbb{I}$. In this case, the representation matrix M^Φ has the following form:

$$M^\Phi = \begin{bmatrix} 1 & \vec{0}^T \\ \vec{0} & \hat{M}^\Phi \end{bmatrix}.$$

We shall define the *modified representation matrix* $\hat{M}^\Phi$ to be the bottom-right $(4^{n_2}-1)\times(4^{n_2}-1)$ submatrix of M^Φ. Next, note that the representation vector of a Hermitian operator P can be written as $\vec{\alpha}^P = (\alpha_0, \hat{\vec{\alpha}}^P)$. We shall call $\hat{\vec{\alpha}}^P$ the *modified representation vector* of the operator P.

For a unital channel Φ, we shall denote the (p, q) group norm of the modified representation matrix $\hat{M}^\Phi$ as $\|\hat{M}^\Phi\|_{p,q}$. Note that the (p, q) group norm of the modified representation matrix of unital quantum channels can be regarded as a magic resource measure (see Appendix B).

Similarly, let us define $\mathcal{C}^{n_0,n_1}_{\|\hat{\cdot}\|_{p,q}\leqslant\mu}$ to be the set of unital quantum circuits C from n_0 qubits to n_1 qubits with bounded norm $\|\hat{\cdot}\|_{p,q}$.

Theorem 2. Let H be a traceless observable. Given a set of unital quantum circuits from n_0 qubits to n_1 qubits with bounded norm $\|\hat{\cdot}\|_{p,q}$, the Rademacher complexity of $\mathcal{F}\circ\mathcal{C}^{n_0,n_1}_{\|\hat{\cdot}\|_{p,q}\leqslant\mu}$ on m samples $S = \{\vec{x}_1, \dots, \vec{x}_m\}$ is bounded as follows:

(1) For $1 \leqslant p \leqslant 2$, we have

$$R_S\big(\mathcal{F}\circ\mathcal{C}^{n_0,n_1}_{\|\hat{\cdot}\|_{p,q}\leqslant\mu}\big) \leqslant \mu N_1^{\max\{\frac{1}{p^*},\frac{1}{q}\}}\frac{\sqrt{\min\{p^*, 8n_0\}}}{\sqrt{m}}\hat{K}_p(S, H). \tag{10}$$

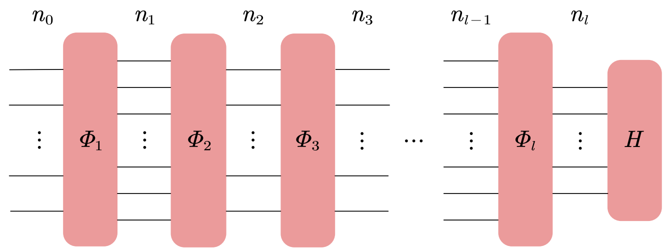

FIG. 1. Circuit diagram of a depth-l quantum circuit.

(2) For $2 < p < \infty$, we have

$$R_S\big(\mathcal{F}\circ\mathcal{C}^{n_0,n_1}_{\|\hat{\cdot}\|_{p,q}\leqslant\mu}\big) \leqslant \mu N_1^{\max\{\frac{1}{p^*},\frac{1}{q}\}}\frac{\sqrt{p^*}}{m^{1/p}}\hat{K}_p(S, H), \tag{11}$$

where $N_1 = 4^{n_1} - 1$,

$$\hat{K}_p(S, H) = \|\hat{\vec{\alpha}}\|_p \max_i \|\hat{\vec{f}}_I(\vec{x}_i)\|_{p^*}, \tag{12}$$

and $\hat{\vec{\alpha}}$ and $\hat{\vec{f}}_I(\vec{x}_i)$ are the modified representation vector of H and $|\psi(x_i)\rangle\langle\psi(x_i)|$ in the Pauli basis, respectively.

The proof of this theorem is presented in Appendix B.

B. Rademacher complexity of depth-l quantum circuits

In this section, we take the depth and width of the quantum circuits involved into account by considering the layer structure of the circuits. Consider a depth-l quantum circuit $C_l = \Phi_l \circ \Phi_{l-1} \circ \cdots \circ \Phi_1$, where the ith layer $\Phi_i : \mathcal{L}((\mathbb{C}^2)^{\otimes n_{i-1}}) \to \mathcal{L}((\mathbb{C}^2)^{\otimes n_i})$ (see Fig. 1 for a circuit diagram). We shall denote the quantum circuit as $\vec{C}_l = (\Phi_l, \Phi_{l-1}, \dots, \Phi_1)$ and the set of quantum circuits with fixed depth l and width vector $\vec{n} = (n_l, \dots, n_1, n_0)$ as

$$\begin{aligned}\mathcal{C}^{l,\vec{n}} = \{\vec{C}_l | \vec{C}_l = (\Phi_l, \Phi_{l-1}, \dots, \Phi_1),\\ \Phi_i : \mathcal{L}((\mathbb{C}^2)^{\otimes n_{i-1}}) \to \mathcal{L}((\mathbb{C}^2)^{\otimes n_i})\}.\end{aligned} \tag{13}$$

Next, let us define the resource measure for a depth-l quantum circuit $\vec{C}_l$ as follows:

$$\nu_{p,q}(\vec{C}_l) = \frac{1}{l}\sum_{i=1}^{l}\|M^{\Phi_i}\|_{p,q}, \tag{14}$$

which represents the average amount of magic resource over the layers of the quantum circuit. Let us denote $\mathcal{C}^{l,\vec{n}}_{\nu_{p,q}\leqslant\nu}$ to be the set of quantum circuits with bounded resource $\nu_{p,q} \leqslant \nu$, fixed depth l, and width vector $\vec{n}$ (see Fig. 2). Then we have the following results:

Theorem 3. Given the set of depth-l quantum circuits with bounded resource $\nu_{p,q} \leqslant \nu$, the Rademacher complexity on m independent samples $S = \{\vec{x}_1, \dots, \vec{x}_m\}$ is bounded as follows:

(1) For $1 \leqslant p \leqslant 2$, we have

$$R_S\big(\mathcal{F}\circ\mathcal{C}^{l,\vec{n}}_{\nu_{p,q}\leqslant\nu}\big) \leqslant \nu^l 4^{\|\vec{n}\|_1\max\{\frac{1}{p^*},\frac{1}{q}\}}\frac{\sqrt{\min\{p^*, 8n_0\}}}{\sqrt{m}}K_p(S, H). \tag{15}$$

(2) For $2 < p < \infty$, we have

$$R_S\big(\mathcal{F}\circ\mathcal{C}^{l,\vec{n}}_{\nu_{p,q}\leqslant\nu}\big) \leqslant \nu^l 4^{\|\vec{n}\|_1\max\{\frac{1}{p^*},\frac{1}{q}\}}\frac{\sqrt{p^*}}{m^{1/p}}K_p(S, H), \tag{16}$$

where $K_p(S, H)$ is defined by Eq. (9), and $\|\vec{n}\|_1 = \sum_{i=1}^{l} n_i$.

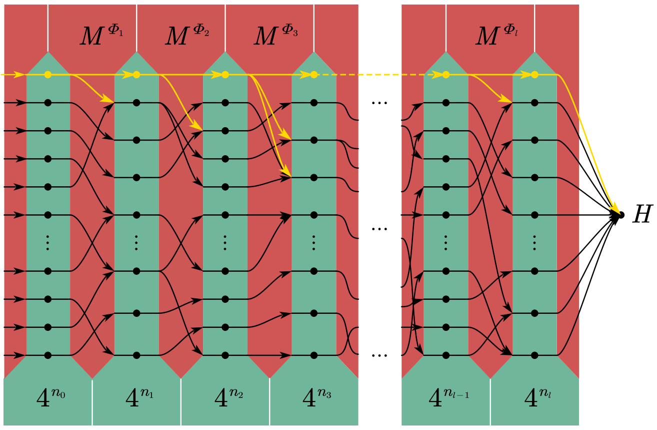

FIG. 2. Diagram illustrating the layer structure of the representation matrix of a depth-l quantum circuit.

This theorem tells us how the Rademacher complexity depends on the depth, width, and the amount of magic resource in the quantum circuits. Note that we can choose suitable p, q to reduce the exponential dependence on the width vector to polynomial dependence, for example, by taking $p^*=q=\Omega(\|\vec{n}\|_1/\log_2\|\vec{n}\|_1)$ or $p^*=q=\infty$. The proof of Theorem 3 is presented in Appendix C.

If the quantum channel in the quantum circuit is unital (for example, a unitary quantum channel), then we modify the resource measure as follows (see Fig. 3):

$$\hat{\nu}_{p,q}(\vec{C}_l)=\frac{1}{l}\sum_{i=1}^{l}\|\hat{M}^{\Phi_i}\|_{p,q}. \tag{17}$$

We are now ready to state our next result.

Theorem 4. Let H be a traceless observable. Given the set of depth-l quantum circuits with bounded resource $\hat{\nu}_{p,q}\leqslant\nu$, the Rademacher complexity of $\mathcal{F}\circ\mathcal{C}^{l,\vec{n}}_{\hat{\nu}_{p,q}\leqslant\nu}$ on m independent samples $S=\{\vec{x}_1,\dots,\vec{x}_m\}$ satisfies the following bounds:

(1) For $1\leqslant p\leqslant 2$, we have

$$R_S\big(\mathcal{F}\circ\mathcal{C}^{l,\vec{n}}_{\hat{\nu}_{p,q}\leqslant\nu}\big)\leqslant\nu^l\Pi_{i=1}^{l}N_i^{\max\{\frac{1}{p^*},\frac{1}{q}\}}\frac{\sqrt{\min\{p^*,8n_0\}}}{\sqrt{m}}\hat{K}_p(S,H). \tag{18}$$

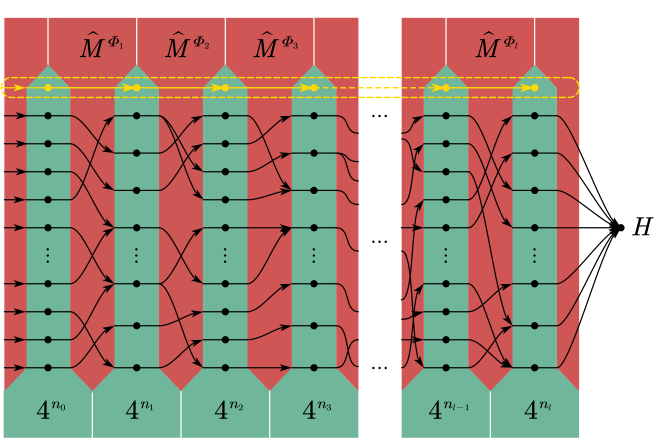

FIG. 3. Diagram illustrating the layer structure of the representation matrix of a depth-l unital quantum circuit.

(2) For $2<p<\infty$, we have

$$R_S\big(\mathcal{F}\circ\mathcal{C}^{l,\vec{n}}_{\hat{\nu}_{p,q}\leqslant\nu}\big)\leqslant\nu^l\Pi_{i=1}^{l}N_i^{\max\{\frac{1}{p^*},\frac{1}{q}\}}\frac{\sqrt{p^*}}{m^{1/p}}\hat{K}_p(S,H), \tag{19}$$

where $N_i=4^{n_i}-1$ for any $1\leqslant i\leqslant l$ and $\hat{K}_p(S,H)$ is defined by Eq. (12).

The proof of this theorem is presented in Appendix D.

Remark 5. While we based our resource measure in this paper on the arithmetic mean, we could have alternatively defined a resource measure of the quantum circuit $\vec{C}_l=(\Phi_l,\Phi_{l-1},\dots,\Phi_1)$ that is based on the geometric mean, viz.

$$\mu_{p,q}(\vec{C}_l)=\prod_{i=1}^{l}\|M^{\Phi_i}\|_{p,q}, \tag{20}$$

which is the geometric mean of the resource over the layers of the quantum circuit. By the arithmetic mean–geometric mean inequality, it is easy to see that

$$\nu_{p,q}(\vec{C}_l)\geqslant\mu_{p,q}(\vec{C}_l)^{1/l}. \tag{21}$$

Also, we could define the path norm as a resource measure as follows:

$$\gamma_{p,q}(\vec{C}_l)=\left(\frac{1}{4^{n_l}}\sum_{\vec{x}}\gamma_p^{(\vec{x})}(C_l)^q\right)^{1/q}, \tag{22}$$

where

$$\gamma_p^{(\vec{x})}(\vec{C}_l)=\left(\sum_{\substack{v_0\to v_1\to\cdots\to v_{out},\\ v_{out}=\vec{x}}}\left|M^{\Phi_l}_{\vec{x}v_{l-1}}M^{\Phi_{l-1}}_{v_{l-1}v_{l-2}}\cdots M^{\Phi_1}_{v_1v_0}\right|^p\right)^{1/p}. \tag{23}$$

The modified version of these resource measures for quantum circuits can also be similarly defined. We present similar results on the Rademacher complexity of quantum circuits based on these resource measures in Appendixes C and D.

Remark 6. Note that, for any given quantum channel Ψ, there could be many different ways to realize it by quantum circuits of the same depth l and width vector $\vec{n}$, i.e., there could be multiple circuits $\vec{C}_l=(\Phi_l,\dots,\Phi_1)$ for which $\Psi=\Phi_l\circ\cdots\circ\Phi_1$. Furthermore, note that resource measures such as $\nu_{p,q}$ depend on the realization of the channel. Hence, if we would like to define a resource measure for quantum channels Ψ that is independent of their quantum circuit realization, it would be necessary to adopt a definition like the one below:

$$\nu^{l,\vec{n}}_{p,q}(\Psi):=\min\{\nu_{p,q}(\vec{C}_l):\vec{C}_l\in\mathcal{C}^{l,\vec{n}},\Psi=C_l\}, \tag{24}$$

which quantifies the minimum amount of resources necessary to realize the target channel over all quantum circuits with a given depth and width. The quantities $\mu^{l,\vec{n}}_{p,q}$ and $\gamma^{l,\vec{n}}_{p,q}$ may also be defined analogously. These resource measures may be of independent interest in resource theory.

C. Application of Rademacher complexity in quantifying generalization error

Let us consider the gap between the empirical error $\hat{L}(f)$ defined in (3) and the true error $L(f)$ defined in (2) with an l_1-norm loss function l, where $l(x,y)=|x-y|$, which quantifies the performance of the hypothesis function f on the unseen data drawn from the unknown probability distribution.

Such a gap is also called the generalization error. The smaller the gap is, the better we learn the feature of the whole data space from the finite samples.

Proposition 7. Given a set of unital quantum circuits from n_0 qubits to n_1 qubits with bounded norm $\|\hat{\cdot}\|_{1,\infty}$, if an observable H with eigenvalues ± 1 at the end of the circuit is measured, then for any $f \in \mathcal{F} \circ \mathcal{C}^{n_0,n_1}_{\|\hat{\cdot}\|_{1,\infty}\leqslant\mu}$, the generalization error is bounded above by

$$|L(f) - \hat{L}(f)| \leqslant 4\mu \frac{\sqrt{8n_0}}{\sqrt{m}} \hat{K}_1(S,H) + 6\sqrt{\frac{\log_2(2/\delta)}{2m}}, \qquad (25)$$

with probability at least $1-\delta$, where $\hat{K}_1(S,H)$ is defined in Theorem 2.

We present a proof of the above theorem in Appendix E.

III. CONCLUSION

In this work, we studied the Rademacher complexity of quantum circuits. First, we introduced the (p,q) group norm to define the magic resource measure for quantum channels and for quantum circuits with a layered structure. Second, we proved that the Rademacher complexity of quantum circuits is bounded by its depth and width as well as its amount of magic resource, where the dependence on the width is determined by the choice of (p,q). These results reveal the dependence of statistical complexity on the resources and structure parameters (such as depth and width) of the quantum circuit.

While our results are stated in terms of the Rademacher complexity, there are other prominent choices of measures of statistical complexity, such as the VC dimension and metric entropy, that could be used. Due to the close relationship between the Rademacher complexity and the VC dimension and the metric entropy [63–66], it is straightforward to extend our results to obtain bounds on these complexity measures of quantum circuits. Another measure that has recently gained prominence is the topological entropy, a concept from dynamic systems that has recently been used to measure the complexity of classical neural networks [67]. We leave for future work the problem of generalizing the results about Rademacher complexity to topological entropy. Finally, we note that, while our results are based on expressing each quantum channel in the Pauli basis, there are also other choices of bases, or more generally frames, that can be used to express quantum channels, a notable example being the phase-space point operator basis [68,69]. How do our results generalize to the case where the basis is chosen arbitrarily? We leave this question for further work.

ACKNOWLEDGMENTS

K.B. thanks Arthur Jaffe and Zhengwei Liu for the help and support during the breakout of the COVID -19 pandemic. K.B. acknowledges the support of ARO Grants W911NF-19-1-0302 and W911NF-20-1-0082, and the support from Yau Mathematical Science Center at Tsinghua University during the visit. D.E.K. acknowledges funding support from the National Research Foundation, Singapore, through Grant NRF2021-QEP2-02-P03.

APPENDIX A: SINGLE QUANTUM CHANNELS

1. (p,q) group norm of the representation matrix of a single quantum channel

For any $N_1 \times N_2$ real-valued matrix M, which can be written as a column

$$\begin{pmatrix} M_1 \\ M_2 \\ \cdots \\ M_{N_1} \end{pmatrix} \qquad (A1)$$

of N_1 rows, we define the (p,q) *group norm*, with $0 < p,\, q \leqslant \infty$, as follows:

$$\|M\|_{p,q} = \left(\frac{1}{N_1} \sum_i \|M_i\|_p^q \right)^{1/q}, \qquad (A2)$$

where the l_p norm of the ith row vector M_i is

$$\|M_i\|_p = \left(\sum_{j=1}^{N_2} |M_{ij}|^p \right)^{1/p}. \qquad (A3)$$

The (p,q) group norm satisfies the following multiplicative property.

Lemma 8. Given two matrices M_1 and M_2, it holds that

$$\|M_1 \otimes M_2\|_{p,q} = \|M_1\|_{p,q} \|M_2\|_{p,q}. \qquad (A4)$$

Proof. This follows directly from the fact that $[M_1 \otimes M_2]_{\vec{x}_1\vec{x}_2,\vec{y}_1\vec{y}_2} = [M_1]_{\vec{x}_1\vec{y}_1}[M_2]_{\vec{x}_2\vec{y}_2}$. ■

Let $P_0 = \mathbb{I}$, $P_1 = X$, $P_2 = Y$, and $P_3 = Z$ be the single-qubit Pauli matrices. The n-qubit Pauli matrices $P_{\vec{z}}$ are defined as $P_{\vec{z}} = P_{z_1} \otimes P_{z_2} \otimes \cdots \otimes P_{z_n}$ for any vector $\vec{z} \in \{0,1,2,3\}^n$. Given a quantum channel $\Phi : \mathcal{L}((\mathbb{C}^2)^{\otimes n_1}) \to \mathcal{L}((\mathbb{C}^2)^{\otimes n_2})$ from n_1 qubits to n_2 qubits, we define the $4^{n_2} \times 4^{n_1}$ representation matrix M^{Φ} in the Pauli basis by its matrix elements as follows:

$$M^{\Phi}_{\vec{z}\vec{x}} = \frac{1}{2^{n_2}} \mathrm{Tr}[P_{\vec{z}} \Phi(P_{\vec{x}})], \qquad (A5)$$

where $\vec{x} \in \{0,1,2,3\}^{n_1}$, $\vec{z} \in \{0,1,2,3\}^{n_2}$, and $P_{\vec{x}}$ and $P_{\vec{z}}$ are the corresponding Pauli operators. From the definition of M^{Φ}, it is easy to see that the representation matrix of quantum channels in the Pauli basis satisfies the following properties:

Lemma 9. Given two quantum channels $\Phi_1 : \mathcal{L}((\mathbb{C}^2)^{\otimes n_1}) \to \mathcal{L}((\mathbb{C}^2)^{\otimes n_2})$ and $\Phi_2 : \mathcal{L}((\mathbb{C}^2)^{\otimes n_3}) \to \mathcal{L}((\mathbb{C}^2)^{\otimes n_4})$, we have

$$M^{\Phi_2 \circ \Phi_1} = M^{\Phi_2} M^{\Phi_1}, \qquad (A6)$$

$$M^{\Phi_2 \otimes \Phi_1} = M^{\Phi_2} \otimes M^{\Phi_1}, \qquad (A7)$$

$$M^{\lambda\Phi_1 + \mu\Phi_2} = \lambda M^{\Phi_1} + \mu M^{\Phi_2} \;\forall\, \lambda, \mu \in \mathbb{R}. \qquad (A8)$$

Proof. Based on definition of the representation matrix M^{Φ}, we have

$$\Phi_1(P_{\vec{x}}) = \sum_{\vec{y}} M^{\Phi_1}_{\vec{y}\vec{x}} P_{\vec{y}}.$$

Therefore, it follows that

$$M^{\Phi_2\circ\Phi_1}_{\vec{z}\vec{x}} = \frac{1}{2^n}\mathrm{Tr}[P_{\vec{z}}\Phi_2\circ\Phi_1(P_{\vec{x}})]$$
$$= \frac{1}{2^{n_4}}\mathrm{Tr}\left[P_{\vec{z}}\sum_{\vec{y}} M^{\Phi_1}_{\vec{y}\vec{x}}\Phi_2(P_{\vec{y}})\right] = \sum_{\vec{y}} M^{\Phi_2}_{\vec{z}\vec{y}}M^{\Phi_1}_{\vec{y}\vec{x}}.$$

Hence,

$$M^{\Phi_2\circ\Phi_1} = M^{\Phi_2}M^{\Phi_1}.$$

The other two identities

$$M^{\Phi_2\otimes\Phi_1} = M^{\Phi_2}\otimes M^{\Phi_1},$$
$$M^{\lambda\Phi_1+\mu\Phi_2} = \lambda M^{\Phi_1} + \mu M^{\Phi_2},$$

follow directly from the definition of the representation matrix. ■

Next, we consider the (p,q) group norm of the representation matrix for quantum channels, for the case when the channel is unitary.

Lemma 10. For a Clifford unitary U, M^U is a permutation matrix, up to a $\pm$ sign. That is, $M^U_{\vec{x}\vec{y}} = \pm\delta_{\vec{x},\pi(\vec{y})}$, where π is a permutation over the set $\{0,1,2,3\}^n$.

Proof. This comes directly from the definition of the Clifford unitaries, which map Pauli operators to Pauli operators. ■

Lemma 11. The (p,q) group norm is invariant under the left or right multiplication of M^U, if U is a Clifford unitary. That is, for any matrix A, we have

$$\|M^UA\|_{p,q} = \|AM^U\|_{p,q} = \|A\|_{p,q}. \quad \text{(A9)}$$

Proof. Based on Lemma 10, M^U is a permutation matrix up to some $\pm$ sign. Hence, AM^U is just a permutation of the columns of A with some $\pm$ sign. Thus, for the ith row vector, we have $\|(AM^U)_i\|_p = \|A_i\|_p$. Therefore, $\|AM^U\|_{p,q} = \|A\|_{p,q}$.

Similarly, M^UA is just a permutation of the columns of A with some $\pm$ sign. Thus, for the ith row vector $\|(M^UA)_i\|_p = \|A_{\pi(i)}\|_p$, where π is a permutation. Then $\sum_i\|(AM^U)_i\|_p = \sum_i\|A_{\pi(i)}\|_p = \sum_i\|A_i\|_p$, i.e., $\|AM^U\|_{p,q} = \|A\|_{p,q}$. ■

Lemma 12. Given a unitary channel U, we have the following result:

(1) For $0<p<2$, we have $\|M^U\|_{p,q}\geqslant 1$, $\|M^U\|_{p,q}=1$ iff U is a Clifford unitary.

(2) For $p>2$, $0<q<\infty$, we have $\|M^U\|_{p,q}\leqslant 1$, $\|M^U\|_{p,q}=1$ iff U is a Clifford unitary.

(3) For $p=2$, $q>0$ or $p>2, q=\infty$, we have $\|M^U\|_{p,q}=1$ for any unitary U.

Proof. First, for any unitary U, it is easy to see that M^U is an orthogonal matrix. Therefore, $\|M^U_{\vec{x}}\|_2 = 1$ for any $\vec{x}$ and $M^U_{\vec{0}} = (1,0,\dots,0)$. Therefore, we have the statement in (3).

(1) For $0<p<2$, we have $\|M^U_{\vec{x}}\|_p \geqslant \|M^U_{\vec{x}}\|_2 = 1$ for any $\vec{x}$. Therefore, $\|M^U\|_{p,q}\geqslant 1$. Besides, $\|M^U\|_{p,q}=1$ iff $\|M^U_{\vec{x}}\|_p = 1$ for any $\vec{x}$ iff every row vector $M^U_{\vec{x}}$ has only one nonzero element, which could only be ± 1, iff U is a Clifford unitary.

(2) For $p>2, 0<q<\infty$, we have $\|M^U_{\vec{x}}\|_p \leqslant \|M^U_{\vec{x}}\|_2 = 1$ for any $\vec{x}$. Therefore, $\|M^U\|_{p,q}\leqslant 1$. Besides, $\|M^U\|_{p,q}=1$ iff $\|M^U_{\vec{x}}\|_p = 1$ for any $\vec{x}$ iff every row vector $M^U_{\vec{x}}$ has only one nonzero element, which could only be ± 1, iff U is Clifford. ■

Based on the above facts, it is easy to see that the (p,q) norm of the representation matrix can be regarded as some magic resource measure of quantum gates.

Proposition 13. Given a unitary channel U, the (p,q) norm can be regarded as a resource measure satisfying the following properties:

(1) (Faithfulness) For $0<p<2$, we have $\|M^U\|_{p,q}\geqslant 1$, $\|M^U\|_{p,q}=1$ iff U is a Clifford unitary.

(1′) (Faithfulness) For $p>2$, $0<q<\infty$, we have $\|M^U\|_{p,q}\leqslant 1$, $\|M^U\|_{p,q}=1$ iff U is a Clifford unitary.

(2) (Invariance under Clifford unitaries) $\|M^{U_1\circ U\circ U_2}\|_{p,q} = \|M^U\|_{p,q}$ for any Clifford unitaries U_1 and U_2.

(3) (Multiplicity under tensor product) $\|M^{U_1\otimes U_2}\|_{p,q} = \|M^{U_1}\|_{p,q}\|M^{U_2}\|_{p,q}$.

(4) (Convexity) For $p\geqslant 1, q\geqslant 1$, we have $\|M^{\lambda U_1+(1-\lambda)U_2}\|_{p,q}\leqslant \lambda\|M^{U_1}\|_{p,q} + (1-\lambda)\|M^{U_2}\|_{p,q}$ for $\lambda\in[0,1]$.

Proof. (1) and (1′) come from Lemma 12 directly.

(2)

$$\|M^{U_1\circ U\circ U_2}\|_{p,q} = \|M^{U_1}M^UM^{U_2}\|_{p,q} = \|M^U\|_{p,q}, \quad \text{(A10)}$$

where the first equality comes from Lemma 9 and the second equality comes from Lemma 12.

(3)

$$\|M^{U_1\otimes U_2}\|_{p,q} = \|M^{U_1}\otimes M^{U_2}\|_{p,q} = \|M^{U_1}\|_{p,q}\|M^{U_2}\|_{p,q}, \quad \text{(A11)}$$

where the first equality comes from Lemma 9 and the second equality comes from Lemma 8.

(4) comes directly from the convexity of l_p and l_q norm for $p\geqslant 1$, $q\geqslant 1$. ■

2. Bounds on the Rademacher complexity of quantum channels

Let p^* denote the Hölder conjugate of p, i.e., $\frac{1}{p}+\frac{1}{p^*}=1$.

Lemma 14. For any $N_1\times N_2$ real-valued matrix M, and any vector $\vec{v}\in\mathbb{R}^{N_2}$, we have

$$\|M\vec{v}\|_{p^*}\leqslant N_1^{\max\{\frac{1}{p^*},\frac{1}{q}\}}\|M\|_{p,q}\|\vec{v}\|_{p^*}. \quad \text{(A12)}$$

Proof. First, let us prove the following inequality:

$$\|M\vec{v}\|_{p^*}\leqslant N_1^{\frac{1}{p^*}}\|M\|_{p,p^*}\|\vec{v}\|_{p^*}.$$

This inequality holds because

$$\|M\vec{v}\|_{p^*}^{p^*} = \sum_i (M_i\vec{v})^{p^*}\leqslant\sum_i\|M_i\|_p^{p^*}\|\vec{v}\|_{p^*}^{p^*} = N_1^{\frac{1}{p^*}}\|M\|_{p,p^*}\|\vec{v}\|_{p^*}.$$

If $q>p^*$, then $\max\{\frac{1}{p^*},\frac{1}{q}\} = \frac{1}{p^*}$ and $\|M\|_{p,q}\geqslant\|M\|_{p,p^*}$. Hence the inequality Eq. (A12) reduces to

$$\|M\vec{v}\|_{p^*}\leqslant N_1^{\frac{1}{p^*}}\|M\|_{p,p^*}\|\vec{v}\|_{p^*}.$$

If $q<p^*$, then $\max\{\frac{1}{p^*},\frac{1}{q}\} = \frac{1}{q}$ and $N_1^{1/q}\|M\|_{p,q}\geqslant N_1^{1/p^*}\|M\|_{p,p^*}$. Hence the inequality Eq. (A12) reduces to

$$\|M\vec{v}\|_{p^*}\leqslant N_1^{\frac{1}{p^*}}\|M\|_{p,p^*}\|\vec{v}\|_{p^*}.$$

■

Lemma 15. For any $1 \leqslant p \leqslant 2$

$$\mathbb{E}_{\vec{\epsilon}} \frac{1}{m} \left\| \sum_{i=1}^{m} \epsilon_i \vec{v}_i \right\|_{p^*} \leqslant \frac{\sqrt{\min\{p^*, 8n_0\}}}{\sqrt{m}} \max_i \|\vec{v}_i\|_{p^*}. \quad \text{(A13)}$$

For $2 < p < \infty$, we have

$$\mathbb{E}_{\vec{\epsilon}} \frac{1}{m} \left\| \sum_{i=1}^{m} \epsilon_i \vec{v}_i \right\|_{p^*} \leqslant \frac{\sqrt{p^*}}{m^{1/p*}} \max_i \|\vec{v}_i\|_{p^*}, \quad \text{(A14)}$$

where $\vec{v}_i \in \mathbb{R}^N$.

Proof. The proof is similar to that of Lemma 15 in Ref. [25]. If $1 \leqslant p \leqslant \frac{2\log_2(N)}{2\log_2(N)-1}$, then $2\log_2(N) \leqslant p^*$. Thence,

$$\begin{aligned}
\mathbb{E}_{\vec{\epsilon}} \frac{1}{m} \left\| \sum_{i=1}^{m} \epsilon_i \vec{v}_i \right\|_{p^*} &\leqslant N^{\frac{1}{p^*}} \mathbb{E}_{\vec{\epsilon}} \frac{1}{m} \mathrm{a} \left\| \sum_{i=1}^{m} \epsilon_i \vec{v}_i \right\|_{\infty} \\
&\leqslant N^{\frac{1}{2\log_2(N)}} \mathbb{E}_{\vec{\epsilon}} \frac{1}{m} \left\| \sum_{i=1}^{m} \epsilon_i \vec{v}_i \right\|_{\infty} \\
&\leqslant \sqrt{2} \mathbb{E}_{\vec{\epsilon}} \frac{1}{m} \left\| \sum_{i=1}^{m} \epsilon_i \vec{v}_i \right\|_{\infty} \\
&= \sqrt{2} \mathbb{E}_{\vec{\epsilon}} \frac{1}{m} \max_j \left| \sum_{i=1}^{m} \epsilon_i v_i(j) \right| \\
&\leqslant \sqrt{2} \frac{\sqrt{2\log_2(N)}}{m} \max_j \|(v_i(j))_i\|_2 \\
&\leqslant \sqrt{2} \frac{\sqrt{2\log_2(N)}}{\sqrt{m}} \max_i \|\vec{v}_i\|_{\infty} \\
&\leqslant \sqrt{2} \frac{\sqrt{2\log_2(N)}}{\sqrt{m}} \max_i \|\vec{v}_i\|_{p^*}.
\end{aligned}$$

If $\frac{2\log_2(N)}{2\log_2(N)-1} < p < \infty$, then by the Khintchine-Kahane inequality, we have

$$\begin{aligned}
\mathbb{E}_{\vec{\epsilon}} \frac{1}{m} \left\| \sum_{i=1}^{m} \epsilon_i \vec{v}_i \right\|_{p^*} &\leqslant \frac{1}{m} \left(\sum_j \mathbb{E}_{\vec{\epsilon}} \left| \sum_i \epsilon_i v_i(j) \right|^{p*} \right)^{\frac{1}{p^*}} \\
&\leqslant \frac{\sqrt{p^*}}{m} \left(\sum_j \|(v_i(j))_i\|_2^{p*} \right)^{\frac{1}{p^*}},
\end{aligned}$$

where

$$\left(\sum_{\vec{z}} \|(v_i(j))_i\|_2^{p*} \right)^{\frac{1}{p^*}} \leqslant \begin{cases} m^{1/2} \max_i \|\vec{v}_i\|_{p^*}, & p^* \geqslant 2 \\ m^{1/p^*} \max_i \|\vec{v}_i\|_{p^*}, & p^* < 2, \end{cases} \quad \text{(A15)}$$

and the first inequality comes from the Minkowski inequality and the second inequality from the fact that

$$(x+y)^{p^*/2} \leqslant x^{p^*/2} + y^{p^*/2},$$

for $p^*/2 < 1$. Therefore

$$\mathbb{E}_{\vec{\epsilon}} \frac{1}{m} \left\| \sum_{i=1}^{m} \epsilon_i \vec{v}_i \right\|_{p^*} \leqslant \begin{cases} \frac{\sqrt{p^*}}{m^{1/2}} \max_i \|\vec{v}_i\|_{p^*}, & p^* \geqslant 2 \\ \frac{\sqrt{p^*}}{m^{1/p}} \max_i \|\vec{v}_i\|_{p^*}, & p^* < 2. \end{cases} \quad \text{(A16)}$$

■

Lemma 16. (Massart lemma [70]) Given a finite set $A \subset \mathbb{R}^m$, we have

$$R(A) \leqslant \max_{\vec{v}\in A} \|\vec{v} - \bar{\vec{v}}\|_2 \frac{\sqrt{2\log_2|A|}}{m}, \quad \text{(A17)}$$

where $\bar{\vec{v}} = \frac{1}{|A|} \sum_{\vec{v}\in A} \vec{v}$.

Theorem 17. (Restatement of Theorem 1) Given a set of quantum circuits Φ from n_0 qubits to n_1 qubits with bounded (p,q) norm $\|\cdot\|_{p,q}$, the Rademacher complexity on m samples $S = \{\vec{x}_1, \dots, \vec{x}_m\}$ satisfies the following bounds:

(1) For $1 \leqslant p \leqslant 2$, we have

$$\begin{aligned}
&R_S(\mathcal{F} \circ \mathcal{C}_{\|\cdot\|_{p,q}\leqslant\mu}) \\
&\quad \leqslant \mu 4^{n_1 \max\{\frac{1}{p^*}, \frac{1}{q}\}} \frac{\sqrt{\min\{p^*, 8n_0\}}}{\sqrt{m}} \|\vec{\alpha}\|_p \max_i \|\vec{f}_I(\vec{x}_i)\|_{p^*}.
\end{aligned} \quad \text{(A18)}$$

(2) For $2 < p < \infty$, we have

$$\begin{aligned}
&R_S(\mathcal{F} \circ \mathcal{C}_{\|\cdot\|_{p,q}\leqslant\mu}) \\
&\quad \leqslant \mu 4^{n_1 \max\{\frac{1}{p^*}, \frac{1}{q}\}} \frac{\sqrt{p^*}}{m^{1/p}} \|\vec{\alpha}\|_p \max_i \|\vec{f}_I(\vec{x}_i)\|_{p^*}.
\end{aligned} \quad \text{(A19)}$$

Proof. First, we compute

$$\begin{aligned}
&R_S(\mathcal{F} \circ \mathcal{C}_{\|\cdot\|_{p,q}\leqslant\mu}) \\
&= \mathbb{E}_{\vec{\epsilon}} \frac{1}{m} \sup_{\Phi\in\mathcal{C}_{\|\cdot\|_{p,q}\leqslant\mu}} \left| \sum_{i=1}^{m} \epsilon_i \vec{\alpha} \vec{f}_\Phi(\vec{x}_i) \right| \\
&\leqslant \mathbb{E}_{\vec{\epsilon}} \frac{1}{m} \sup_{\Phi} \frac{\mu}{\|M^\Phi\|_{p,q}} \left| \sum_{i=1}^{m} \epsilon_i \vec{\alpha} \vec{f}_\Phi(\vec{x}_i) \right| \\
&= \mu \mathbb{E}_{\vec{\epsilon}} \frac{1}{m} \sup_{\Phi} \frac{1}{\|M^\Phi\|_{p,q}} \left| \sum_{i=1}^{m} \epsilon_i \vec{\alpha} \vec{f}_\Phi(\vec{x}_i) \right| \\
&\leqslant \mu \|\vec{\alpha}\|_p \mathbb{E}_{\vec{\epsilon}} \frac{1}{m} \sup_{\Phi} \frac{1}{\|M^\Phi\|_{p,q}} \left\| \sum_{i=1}^{m} \epsilon_i \vec{f}_\Phi(\vec{x}_i) \right\|_{p^*} \\
&\leqslant \mu \|\vec{\alpha}\|_p \mathbb{E}_{\vec{\epsilon}} \frac{1}{m} \sup_{\Phi} \frac{1}{\|M^\Phi\|_{p,q}} \left\| \sum_{i=1}^{m} \epsilon_i M^\Phi \vec{f}_I(\vec{x}_i) \right\|_{p^*} \\
&\leqslant \mu \|\vec{\alpha}\|_p N_1^{\max\{\frac{1}{p^*}, \frac{1}{q}\}} \mathbb{E}_{\vec{\epsilon}} \frac{1}{m} \left\| \sum_{i=1}^{m} \epsilon_i \vec{f}_I(\vec{x}_i) \right\|_{p^*},
\end{aligned}$$

where the third inequality follows from Lemma 14. Using Lemma 15, we get the results of this theorem. ■

APPENDIX B: SINGLE UNITAL QUANTUM CHANNEL

1. (p, q) group norm of the modified representation matrix of unital channels

If a quantum channel Φ is unital, i.e., $\Phi(\mathbb{I}) = \mathbb{I}$, then the representation matrix M^Φ has the following form:

$$M^\Phi = \begin{bmatrix} 1 & \vec{0}^T \\ \vec{0} & \hat{M}^\Phi \end{bmatrix}. \quad \text{(B1)}$$

We call $\hat{M}^\Phi$ the *modified representation matrix* of Φ. (Note that a unitary channel is a special case of a unital channel.)

For a unital channel Φ, we define the (p,q) *group norm* of the modified representation matrix $\hat{M}^{\Phi}$ as follows:

$$\begin{aligned}\|\hat{M}^{\Phi}\|_{p,q} &= \left[\frac{1}{N}\sum_{\vec{x}\neq\vec{0}}\left(\sum_{\vec{y}\neq\vec{0}}|M_{\vec{x},\vec{y}}|^p\right)^{\frac{q}{p}}\right]^{\frac{1}{q}}\\ &= \left(\frac{1}{N}\sum_{\vec{x}\neq\vec{0}}\left\|M^{\Phi}_{\vec{x}}\right\|_p^q\right)^{\frac{1}{q}},\end{aligned} \tag{B2}$$

where $N=4^n-1$.

We now state and prove the following properties of the (p,q) norm of the representation matrix $\hat{M}^U$, where U is a unitary channel.

Proposition 18. For any unitary channel U, we have the following relationships between $\|M^U\|_{p,q}$ and $\|\hat{M}^U\|_{p,q}$,

(1) For $0<p<2$, $0<q<\infty$, we have

$$\|M^U\|_{p,q}\leqslant\|\hat{M}^U\|_{p,q}, \tag{B3}$$

with equality iff U is a Clifford unitary.

(2) For $0<p<2$, $q=\infty$, we have

$$\|M^U\|_{p,\infty}=\|\hat{M}^U\|_{p,\infty}, \tag{B4}$$

for any unitary U.

(3) For $p>2$, $q>0$, we have

$$\|M^U\|_{p,q}\geqslant\|\hat{M}^U\|_{p,q}, \tag{B5}$$

with equality iff U is a Clifford unitary

(4) For $p=2$, $q>0$

$$\|M^U\|_{p,q}=\|\hat{M}^U\|_{p,q}=1. \tag{B6}$$

Proof. (4) is obvious, as M^U and $\hat{M}^U$ are orthogonal matrices.

For $0<p<2$, $q>0$ we have $\|M^U_{\vec{x}}\|_p\geqslant\|M^U_{\vec{0}}\|_p=1$ for any $\vec{x}\neq\vec{0}$. Therefore, $\|M^U\|_{p,q}\leqslant\|\hat{M}^U\|_{p,q}$ for $0<q<\infty$ and $\|M^U\|_{p,q}=\|\hat{M}^U\|_{p,q}$ for $q=\infty$. Hence, we get (2). Next, for $0<q<\infty$, $\|M^U\|_{p,q}=\|\hat{M}^U\|_{p,q}$ iff $\|M^U_{\vec{x}}\|_p=1$ for any $\vec{x}\neq\vec{0}$ iff every row vector $M^U_{\vec{x}}$ has only one nonzero element, which could only be ±1, iff U is a Clifford unitary. Hence, we get (1).

For $p>2$, $0<q\leqslant\infty$, we have $\|M^U_{\vec{x}}\|_p\leqslant\|M^U_{\vec{0}}\|_p=1$ for any $\vec{x}$. Therefore, $\|M^U\|_{p,q}\geqslant\|\hat{M}^U\|_{p,q}$. Besides, $\|M^U\|_{p,q}=\|\hat{M}^U\|_{p,q}$ iff $\|M^U_{\vec{x}}\|_p=1$ for any $\vec{x}$ iff every row vector $M^U_{\vec{x}}$ has only one nonzero element, which could only be ±1, iff U is a Clifford unitary. Therefore, we get (3). ■

A direct consequence of the above proposition is the following corollary:

Corollary 19. Given a unitary channel U, the (p,q) group norm of the modified representation matrix $\hat{M}^U$ can be regarded as a resource measure which satisfies the following properties:

(1) (Faithfulness) For $0<p<2$, $q>0$ we have $\|\hat{M}^U\|_{p,q}\geqslant1$, $\|\hat{M}^U\|_{p,q}=1$ iff U is a Clifford unitary.

(1′) (Faithfulness) For $p>2$, $q>0$, we have $\|\hat{M}^U\|_{p,q}\leqslant1$, $\|\hat{M}^U\|_{p,q}=1$ iff U is a Clifford unitary.

(2) (Invariance under a Clifford unitary) $\|\hat{M}^{U_1\circ U\circ U_2}\|_{p,q}=\|\hat{M}^U\|_{p,q}$ for any Clifford unitary U_1 and U_2.

(3) (Convexity) For $p\geqslant1$, we have $\|\hat{M}^{\lambda U_1+(1-\lambda)U_2}\|_{p,q}\leqslant\lambda\|\hat{M}^{U_1}\|_{p,q}+(1-\lambda)\|\hat{M}^{U_2}\|_{p,q}$.

Proposition 20. Let U_1 and U_2 be unitary channels.

(1) For $0<p<2$, $0<q<\infty$, we have

$$\|\hat{M}^{U_1}\otimes\hat{M}^{U_2}\|_{p,q}\geqslant\|\hat{M}^{U_1\otimes U_2}\|_{p,q}, \tag{B7}$$

with equality iff U_1 and U_2 are Clifford unitaries.

(2) For $0<p<2$, $q=\infty$, we have

$$\|\hat{M}^{U_1}\otimes\hat{M}^{U_2}\|_{p,\infty}=\|\hat{M}^{U_1\otimes U_2}\|_{p,\infty}, \tag{B8}$$

for any unitaries U_1 and U_2.

(3) For $p>2$, $0<q\leqslant\infty$, we have

$$\|\hat{M}^{U_1}\otimes\hat{M}^{U_2}\|_{p,q}\leqslant\|\hat{M}^{U_1\otimes U_2}\|_{p,q}. \tag{B9}$$

For $p>2$, $0<q<\infty$, "=" holds iff U_1 and U_2 are Clifford unitaries.

(4) For $p=2$, $0<q\leqslant\infty$, we have

$$\|\hat{M}^{U_1}\otimes\hat{M}^{U_2}\|_{2,q}=\|\hat{M}^{U_1\otimes U_2}\|_{2,q}=1. \tag{B10}$$

Proof. (3) is obvious as both $\hat{M}^{U_1}\otimes\hat{M}^{U_2}$ and $\hat{M}^{U_1\otimes U_2}$ are orthogonal matrices.

Using the property

$$\|\hat{M}^{U_1}\otimes\hat{M}^{U_2}\|_{p,q}=\|\hat{M}^{U_1}\|_{p,q}\|\hat{M}^{U_2}\|_{p,q},$$

we find that for $0<q<\infty$,

$$\begin{aligned}&\|\hat{M}^{U_1}\otimes\hat{M}^{U_2}\|^q_{p,q}\\ &=\left(\frac{1}{N_1}\sum_{\vec{x}_1\neq\vec{0}}\left\|M^{U_1}_{\vec{x}_1}\right\|_p^q\right)\left(\frac{1}{N_2}\sum_{\vec{x}_2\neq\vec{0}}\left\|M^{U_2}_{\vec{x}_2}\right\|_p^q\right),\\ &\|\hat{M}^{U_1\otimes U_2}\|^q_{p,q}\\ &=\left(\frac{1}{N_1N_2+N_1+N_2}\right)\sum_{(\vec{x}_1,\vec{x}_2)\neq(\vec{0},\vec{0})}\left\|M^{U_1}_{\vec{x}_1}\right\|_p^q\left\|M^{U_2}_{\vec{x}_2}\right\|_p^q\\ &=\left(\frac{1}{N_1N_2+N_1+N_2}\right)\left(\sum_{\vec{x}_1\neq\vec{0},\vec{x}_2\neq\vec{0}}\left\|M^{U_1}_{\vec{x}_1}\right\|_p^q\left\|M^{U_2}_{\vec{x}_2}\right\|_p^q\right.\\ &\quad\left.+\sum_{\vec{x}_1\neq\vec{0},\vec{x}_2=\vec{0}}\left\|M^{U_1}_{\vec{x}_1}\right\|_p^q\left\|M^{U_2}_{\vec{x}_2}\right\|_p^q+\sum_{\vec{x}_1=\vec{0},\vec{x}_2\neq\vec{0}}\left\|M^{U_1}_{\vec{x}_1}\right\|_p^q\left\|M^{U_2}_{\vec{x}_2}\right\|_p^q\right)\\ &=\left(\frac{1}{N_1N_2+N_1+N_2}\right)\left(\sum_{\vec{x}_1\neq\vec{0},\vec{x}_2\neq\vec{0}}\left\|M^{U_1}_{\vec{x}_1}\right\|_p^q\left\|M^{U_2}_{\vec{x}_2}\right\|_p^q\right.\\ &\quad\left.+\sum_{\vec{x}_1\neq\vec{0}}\left\|M^{U_1}_{\vec{x}_1}\right\|_p^q+\sum_{\vec{x}_2\neq\vec{0}}\left\|M^{U_2}_{\vec{x}_2}\right\|_p^q\right),\end{aligned}$$

where $N_1=4^{n_1}-1$, $N_2=4^{n_2}-1$. Hence to compare $\|\hat{M}^{U_1}\otimes\hat{M}^{U_2}\|_{p,q}$ and $\|\hat{M}^{U_1\otimes U_2}\|_{p,q}$, we need only to compare

$$\left(\frac{1}{N_1}\sum_{\vec{x}_1\neq\vec{0}}\left\|M^{U_1}_{\vec{x}_1}\right\|_p^q\right)\left(\frac{1}{N_2}\sum_{\vec{x}_2\neq\vec{0}}\left\|M^{U_2}_{\vec{x}_2}\right\|_p^q\right),$$

and

$$\frac{1}{N_1+N_2}\left(\sum_{\vec{x}_1\neq\vec{0}}\|M^{U_1}_{\vec{x}_1}\|^q_p+\sum_{\vec{x}_2\neq\vec{0}}\|M^{U_2}_{\vec{x}_2}\|^q_p\right).$$

To this end, let us consider a simple inequality first. It is easy to verify the following two inequalities:

(1) For $a, b \geqslant 1$, we have

$$ab \geqslant \frac{N_1a+N_2b}{N_1+N_2}. \tag{B11}$$

Moreover, equality holds iff $a=b=1$.

(2) For $0<a, b\leqslant 1$, we have

$$ab \leqslant \frac{N_1a+N_2b}{N_1+N_2}. \tag{B12}$$

Moreover, equality holds iff $a=b=1$.

Thus, for $0<p<2$, $0<q<\infty$, we have $\|M^U_{\vec{x}}\|^q_p\geqslant 1$ for any $\vec{x}\neq\vec{0}$, and $\|M^U_{\vec{x}}\|^q_p=1$ for all $\vec{x}\neq 0$ iff U is Clifford. Let

$$a=\frac{1}{N_1}\sum_{\vec{x}_1\neq\vec{0}}\|M^{U_1}_{\vec{x}_1}\|^q_p,\quad b=\frac{1}{N_2}\sum_{\vec{x}_2\neq\vec{0}}\|M^{U_2}_{\vec{x}_2}\|^q_p.$$

Then by the first inequality (1), we have

$$\left(\frac{1}{N_1}\sum_{\vec{x}_1\neq\vec{0}}\|M^{U_1}_{\vec{x}_1}\|^q_p\right)\left(\frac{1}{N_2}\sum_{\vec{x}_2\neq\vec{0}}\|M^{U_2}_{\vec{x}_2}\|^q_p\right)$$
$$\geqslant\frac{1}{N_1+N_2}\left(\sum_{\vec{x}_1\neq\vec{0}}\|M^{U_1}_{\vec{x}_1}\|^q_p+\sum_{\vec{x}_2\neq\vec{0}}\|M^{U_2}_{\vec{x}_2}\|^q_p\right).$$

Therefore, for $0<p<2$, $0<q<\infty$, we have

$$\|\hat{M}^{U_1}\otimes\hat{M}^{U_2}\|_{p,q}\geqslant\|\hat{M}^{U_1\otimes U_2}\|_{p,q},$$

where equality holds iff U_1 and U_2 are Clifford unitaries.

Similarly, for $p>2$, $0<q<\infty$, we have $\|M^U_{\vec{x}}\|^q_p\leqslant 1$ for any $\vec{x}\neq\vec{0}$, and $\|M^U_{\vec{x}}\|^q_p=1$ for all $\vec{x}\neq 0$ iff U is Clifford.

Let

$$a=\frac{1}{N_1}\sum_{\vec{x}_1\neq\vec{0}}\|M^{U_1}_{\vec{x}_1}\|^q_p,\quad b=\frac{1}{N_2}\sum_{\vec{x}_2\neq\vec{0}}\|M^{U_2}_{\vec{x}_2}\|^q_p.$$

Then by the second inequality (2), we have

$$\left(\frac{1}{N_1}\sum_{\vec{x}_1\neq\vec{0}}\|M^{U_1}_{\vec{x}_1}\|^q_p\right)\left(\frac{1}{N_2}\sum_{\vec{x}_2\neq\vec{0}}\|M^{U_2}_{\vec{x}_2}\|^q_p\right)$$
$$\leqslant\frac{1}{N_1+N_2}\left(\sum_{\vec{x}_1\neq\vec{0}}\|M^{U_1}_{\vec{x}_1}\|^q_p+\sum_{\vec{x}_2\neq\vec{0}}\|M^{U_2}_{\vec{x}_2}\|^q_p\right).$$

Therefore, for $p>2$, $0<q<\infty$, we have

$$\|\hat{M}^{U_1}\otimes\hat{M}^{U_2}\|_{p,q}\leqslant\|\hat{M}^{U_1\otimes U_2}\|_{p,q}.$$

Moreover, equality holds iff U_1 and U_2 are Clifford unitaries.

Now, let us consider the case where $q=\infty$. For $q=\infty$, we have

$$\|\hat{M}^{U_1}\otimes\hat{M}^{U_2}\|_{p,\infty}=\max_{\vec{x}_1\neq\vec{0}}\|M^{U_1}_{\vec{x}_1}\|_p\max_{\vec{x}_2\neq\vec{0}}\|M^{U_2}_{\vec{x}_2}\|_p,$$

and

$$\begin{aligned}\|\hat{M}^{U_1\otimes U_2}\|_{p,\infty}&=\max_{(\vec{x}_1,\vec{x}_2)\neq(\vec{0},\vec{0})}\|M^{U_1}_{\vec{x}_1}\|_p\|M^{U_2}_{\vec{x}_2}\|_p\\&=\max\left\{\max_{\vec{x}_1\neq\vec{0}}\|M^{U_1}_{\vec{x}_1}\|_p\max_{\vec{x}_2\neq\vec{0}}\|M^{U_2}_{\vec{x}_2}\|_p,\right.\\&\qquad\left.\max_{\vec{x}_1\neq\vec{0}}\|M^{U_1}_{\vec{x}_1}\|_p,\max_{\vec{x}_2\neq\vec{0}}\|M^{U_2}_{\vec{x}_2}\|_p\right\}.\end{aligned}$$

Hence, for $0<p<2$, we have $\|M^U_{\vec{x}}\|_p\geqslant 1$ for any $\vec{x}\neq\vec{0}$; therefore,

$$\begin{aligned}&\max_{\vec{x}_1\neq\vec{0}}\|M^{U_1}_{\vec{x}_1}\|_p\max_{\vec{x}_2\neq\vec{0}}\|M^{U_2}_{\vec{x}_2}\|_p\\&\quad\geqslant\max\left\{\max_{\vec{x}_1\neq\vec{0}}\|M^{U_1}_{\vec{x}_1}\|_p,\max_{\vec{x}_2\neq\vec{0}}\|M^{U_2}_{\vec{x}_2}\|_p\right\}.\end{aligned}$$

That is,

$$\|\hat{M}^{U_1}\otimes\hat{M}^{U_2}\|_{p,\infty}=\|\hat{M}^{U_1\otimes U_2}\|_{p,\infty}.$$

For $p>2$, we have $\|M^U_{\vec{x}}\|_p\leqslant 1$ for any $\vec{x}\neq\vec{0}$; therefore,

$$\begin{aligned}&\max_{\vec{x}_1\neq\vec{0}}\|M^{U_1}_{\vec{x}_1}\|_p\max_{\vec{x}_2\neq\vec{0}}\|M^{U_2}_{\vec{x}_2}\|_p\\&\quad\leqslant\max\left\{\max_{\vec{x}_1\neq\vec{0}}\|M^{U_1}_{\vec{x}_1}\|_p,\max_{\vec{x}_2\neq\vec{0}}\|M^{U_2}_{\vec{x}_2}\|_p\right\}.\end{aligned}$$

■

2. Rademacher complexity of single unital quantum circuit

In this section, we assume for simplicity that the observable H is traceless, which implies that $\alpha_{\vec{0}}=0$.

Theorem 21. (Restatement of Theorem 2) Given the set of unital quantum circuits Φ from n_0 qubits to n_1 qubits with bounded (p,q) norm of the modified representation matrix, the Rademacher complexity on m samples $S=\{\vec{x}_1,\ldots,\vec{x}_m\}$ satisfies the following bounds:

(1) For $1\leqslant p\leqslant 2$, we have

$$\begin{aligned}R_S\big(\mathcal{F}\circ\mathcal{C}^{n_0,n_1}_{\|\hat{\cdot}\|_{p,q}\leqslant\mu}\big)&\leqslant\mu N_1^{\max\{\frac{1}{p^*},\frac{1}{q}\}}\frac{\sqrt{\min\{p^*,8n_0\}}}{\sqrt{m}}\\&\quad\times\|\hat{\vec{\alpha}}\|_p\max_i\|\hat{\vec{f}}_I(\vec{x}_i)\|_{p^*},\end{aligned}\tag{B13}$$

where $N_1=4^{n_1}-1$.

(2) For $2<p<\infty$, we have

$$R_S\big(\mathcal{F}\circ\mathcal{C}^{n_0,n_1}_{\|\hat{\cdot}\|\leqslant\mu}\big)\leqslant\mu N_1^{\max\{\frac{1}{p^*},\frac{1}{q}\}}\frac{\sqrt{p^*}}{m^{1/p}}\|\hat{\vec{\alpha}}\|_p\max_i\|\hat{\vec{f}}_I(\vec{x}_i)\|_{p^*}. \tag{B14}$$

Proof. Since $\alpha_{\vec{x}}=0$, it follows that $\vec{\alpha}=(0,\hat{\vec{\alpha}})$, where $\hat{\vec{\alpha}}\in\mathbb{R}^{N_1}$. Hence,

$$\begin{aligned}f_\Phi(\vec{x})&=\mathrm{Tr}[\Phi(|\psi(\vec{x})\rangle\langle\psi(\vec{x})|)H]\\&=\sum_{\vec{z}\neq\vec{0}}\alpha_{\vec{z}}\mathrm{Tr}[\Phi(|\psi(\vec{x}_i)\rangle\langle\psi(\vec{x}_i)|)P_{\vec{z}}]\\&=\hat{\vec{\alpha}}\hat{\vec{f}}_\Phi(\vec{x}),\end{aligned}$$

where $\hat{\vec{f}}_\Phi(\vec{x}) = [\hat{f}^{\vec{z}}_\Phi(\vec{x})]_{\vec{z}\neq\vec{0}} \in \mathbb{R}^{N_1}$ and $N_1 = 4^{n_1} - 1$. Similarly, for unital quantum channels Φ, we have

$$\hat{\vec{f}}_\Phi(\vec{x}) = \hat{M}^\Phi \hat{\vec{f}}_I(\vec{x}). \tag{B15}$$

Therefore, we have

$$\begin{aligned}
&R_S\big(\mathcal{F}\circ\mathcal{C}^{n_0,n_1}_{\|\cdot\|\leqslant\mu}\big)\\
&= \mathbb{E}_{\vec{\epsilon}}\frac{1}{m}\sup_{\Phi\in\mathcal{C}^{n_0,n_1}_{\|\cdot\|\leqslant\mu}}\left|\sum_{i=1}^m \epsilon_i\hat{\vec{\alpha}}\hat{\vec{f}}_\Phi(\vec{x}_i)\right|\\
&\leqslant \mathbb{E}_{\vec{\epsilon}}\frac{1}{m}\sup_{\Phi}\frac{\mu}{\|\hat{M}^\Phi\|_{p,q}}\left|\sum_{i=1}^m \epsilon_i\hat{\vec{\alpha}}\hat{\vec{f}}_\Phi(\vec{x}_i)\right|\\
&= \mu\mathbb{E}_{\vec{\epsilon}}\frac{1}{m}\sup_{\Phi}\frac{1}{\|\hat{M}^\Phi\|_{p,q}}\left|\sum_{i=1}^m \epsilon_i\hat{\vec{\alpha}}\hat{\vec{f}}_\Phi(\vec{x}_i)\right|\\
&\leqslant \mu\big\|\hat{\vec{\alpha}}\big\|_p\mathbb{E}_{\vec{\epsilon}}\frac{1}{m}\sup_{\Phi}\frac{1}{\|\hat{M}^\Phi\|_{p,q}}\left\|\sum_{i=1}^m \epsilon_i\hat{\vec{f}}_\Phi(\vec{x}_i)\right\|_{p^*}\\
&\leqslant \mu\big\|\hat{\vec{\alpha}}\big\|_p\mathbb{E}_{\vec{\epsilon}}\frac{1}{m}\sup_{\Phi}\frac{1}{\|\hat{M}^\Phi\|_{p,q}}\left\|\sum_{i=1}^m \epsilon_i\hat{M}^\Phi\hat{\vec{f}}_I(\vec{x}_i)\right\|_{p^*}\\
&\leqslant \mu\big\|\hat{\vec{\alpha}}\big\|_p N_1^{\max\{\frac{1}{p^*},\frac{1}{q}\}}\mathbb{E}_{\vec{\epsilon}}\frac{1}{m}\left\|\sum_{i=1}^m \epsilon_i\hat{\vec{f}}_I(\vec{x}_i)\right\|_{p^*},
\end{aligned}$$

where the third inequality come from the Lemma 14. Using Lemma 15, we get the results of the theorem. ■

APPENDIX C: DEEP QUANTUM CIRCUITS

Consider a depth-l quantum circuit, where each layer of the quantum circuit is a treated as a quantum channel. We denote the depth-l quantum circuit as $\vec{C}_l$ as follows

$$\vec{C}_l = (\Phi_l, \Phi_{l-1}, \dots, \Phi_1), \tag{C1}$$

where the ith layer $\Phi_i : \mathcal{L}((\mathbb{C}^2)^{\otimes n_{i-1}}) \to \mathcal{L}((\mathbb{C}^2)^{\otimes n_i})$ (see Fig. 1).

Let us define $\mathcal{C}^{l,\vec{n}}$ with the width vector $\vec{n} = (n_l, \dots, n_0)$ to be the set of all depth-l quantum circuits $\vec{C}_l = (\Phi_l, \Phi_{l-1}, \dots, \Phi_1)$, where the ith layer $\Phi_i : \mathcal{L}((\mathbb{C}^2)^{\otimes n_{i-1}}) \to \mathcal{L}((\mathbb{C}^2)^{\otimes n_i})$. In this section, we introduce three resource measures to quantify the amount of magic resource in quantum circuits by making use of the (p, q) group norm.

1. Multiplication (p, q) depth-norm

In this section, let us define the multiplication (p, q) depth-norm for depth-l quantum circuits $\vec{C}_l = (\Phi_l, \Phi_{l-1}, \dots, \Phi_1)$ as follows:

$$\mu_{p,q}(\vec{C}_l) = \prod_{i=1}^l \|M^{\Phi_i}\|_{p,q}. \tag{C2}$$

Proposition 22. The multiplication (p, q) depth-norm satisfies the following properties:

(1) Given a depth-l quantum circuit $\vec{C}_l$ and a depth-m quantum circuit $\vec{C}_m$, we have

$$\mu_{p,q}(\vec{C}_l\circ\vec{C}_m) = \mu_{p,q}(\vec{C}_l)\mu_{p,q}(\vec{C}_m), \tag{C3}$$

where $\vec{C}_l\circ\vec{C}_m := (\vec{C}_l, \vec{C}_m)$.

(2) Given two depth-l quantum circuits C_l and C'_l, we have

$$\mu_{p,q}(\vec{C}_l\otimes\vec{C}'_l) = \mu_{p,q}(\vec{C}_l)\mu_{p,q}(\vec{C}'_l), \tag{C4}$$

where $\vec{C}_l\otimes\vec{C}'_l := (\Phi_l\otimes\Phi'_l, \dots, \Phi_1\otimes\Phi'_1)$ for $\vec{C}_l = (\Phi_l, \dots, \Phi_1), \vec{C}'_l = (\Phi'_l, \dots, \Phi'_1)$.

Proof. These two properties follow directly from the definition of $\mu_{p,q}$. ■

Note that, for the depth-l quantum circuit $\vec{C}_l$, where each layer contains only unitary gates, i.e., $\vec{C}_l = (U_l, U_{l-1}, \dots, U_1)$, $\mu_{p,q}$ can be viewed as a magic resource measure.

Lemma 23. Given a depth-l quantum circuit $\vec{C}_l = (U_l, U_{l-1}, \dots, U_1)$, we have the following:

(1) (Faithfulness) For $0 < p < 2$, $q > 0$, it holds that $\mu_{p,q}(\vec{C}_l) \geqslant 1$, and $\mu_{p,q}(\vec{C}_l) = 1$ iff $\vec{C}_l$ is a Clifford circuit, i.e., each U_i is a Clifford unitary.

(1′) (Faithfulness) For $p > 2$, $0 < q < \infty$, it holds that $\mu_{p,q}(\vec{C}_l) \leqslant 1$, and $\mu_{p,q}(\vec{C}_l) = 1$ iff $\vec{C}_l$ is a Clifford circuit.

(2) (Invariance under Clifford circuits) For $p > 0$, $q > 0$, we have $\mu_{p,q}(\vec{C}_1\circ\vec{C}_l\circ\vec{C}_2) = \mu_{p,q}(\vec{C}_l)$ if $\vec{C}_1$, $\vec{C}_2$ are Clifford circuits.

Proof. This lemma follows directly from Lemma 12 and Proposition 22. ■

Next, let us denote the set of depth-l quantum circuits $\vec{C}_l$ with bounded depth-norm $\mu_{p,q}$ as $\mathcal{C}^{l,\vec{n}}_{\mu_{p,q}\leqslant\mu}$, that is,

$$\mathcal{C}^{l,\vec{n}}_{\mu_{p,q}\leqslant\mu} := \{\vec{C}_l \in \mathcal{C}^{l,\vec{n}} : \mu_{p,q}(\vec{C}_l) \leqslant \mu\}. \tag{C5}$$

Lemma 24. Given the set of depth-l quantum circuits $\mathcal{C}^{l,\vec{n}}$ and the set of depth-l quantum circuits $\mathcal{C}^{l,\vec{n}'}$, where $\vec{n} = (\vec{n}', n_l)$ and $\vec{n}' = (\vec{n}'', n_{l-1})$, then $\forall\,\vec{\epsilon} \in \{\pm 1\}^m$, we have

$$\begin{aligned}
&\sup_{\vec{C}_l\in\mathcal{C}^{l,\vec{n}}}\frac{1}{\mu_{p,q}(\vec{C}_l)}\left\|\sum_{i=1}^m \epsilon_i\vec{f}_{C_l}(\vec{x}_i)\right\|_{p^*}\\
&\leqslant 4_l^{n_l\max\{\frac{1}{p^*},\frac{1}{q}\}}\sup_{\vec{C}_{l-1}\in\mathcal{C}^{l,\vec{n}}}\frac{1}{\mu_{p,q}(\vec{C}_{l-1})}\left\|\sum_{i=1}^m \epsilon_i\vec{f}_{C_{l-1}}(\vec{x}_i)\right\|_{p^*}.
\end{aligned}\tag{C6}$$

Thus,

$$\begin{aligned}
&\sup_{\vec{C}_l\in\mathcal{C}^{l,\vec{n}}}\frac{1}{\mu_{p,q}(\vec{C}_l)}\left\|\sum_{i=1}^m \epsilon_i f_{C_l}(\vec{x}_i)\right\|_{p^*}\\
&\leqslant \prod_{i=1}^l 4^{n_i\max\{\frac{1}{p^*},\frac{1}{q}\}}\left\|\sum_{i=1}^m \epsilon_i f_I(\vec{x}_i)\right\|_{p^*}.
\end{aligned}\tag{C7}$$

Proof. The lemma follows from

$$\begin{aligned}
&\sup_{\vec{C}_l\in\mathcal{C}^{l,\vec{n}}}\frac{1}{\mu_{p,q}(\vec{C}_l)}\left\|\sum_{i=1}^m \epsilon_i\vec{f}_{C_l}(\vec{x}_i)\right\|_{p^*}\\
&= \sup_{\vec{C}_{l-1}\in\mathcal{C}^{l-1,\vec{n}'}}\frac{1}{\mu_{p,q}(\vec{C}_{l-1})\|M^{\Phi_l}\|_{p,q}}\left\|\sum_{i=1}^m \epsilon_i M^{\Phi_l}\vec{f}_{C_{l-1}}(\vec{x}_i)\right\|_{p^*}\\
&\leqslant 4^{n_l\max\{\frac{1}{p^*},\frac{1}{q}\}}\sup_{\vec{C}_{l-1}\in\mathcal{C}^{l-1,\vec{n}'}}\frac{1}{\mu_{p,q}(\vec{C}_{l-1})}\left\|\sum_{i=1}^m \epsilon_i\vec{f}_{C_{l-1}}(\vec{x}_i)\right\|_{p^*},
\end{aligned}$$

where the inequality follows from Lemma 14. ■

Theorem 25. Given the set of depth-l quantum circuits with bounded depth-norm $\mu_{p,q}$, the Rademacher complexity on m samples $S=\{\vec{x}_1,\dots,\vec{x}_m\}$ satisfies the following bounds:

(1) For $1\leqslant p\leqslant 2$, we have

$$R_S\big(\mathcal{F}\circ\mathcal{C}^{l,\vec{n}}_{\mu_{p,q}\leqslant\mu}\big)\leqslant \mu 4^{(\sum_{i=1}^l n_i)\max\{\frac{1}{p^*},\frac{1}{q}\}}\frac{\sqrt{\min\{p^*,8n_0\}}}{\sqrt{m}}\times\|\vec{\alpha}\|_p\max_i\|\vec{f}_I(\vec{x}_i)\|_{p^*}. \tag{C8}$$

(2) For $2<p<\infty$, we have

$$R_S\big(\mathcal{F}\circ\mathcal{C}^{l,\vec{n}}_{\mu_{p,q}\leqslant\mu}\big)\leqslant \mu 4^{(\sum_{i=1}^l n_i)\max\{\frac{1}{p^*},\frac{1}{q}\}}\times\frac{\sqrt{p^*}}{m^{1/p}}\|\vec{\alpha}\|_p\max_i\|\vec{f}_I(\vec{x}_i)\|_{p^*}. \tag{C9}$$

Proof. These bounds follow from

$$\begin{aligned}
&R_S\big(\mathcal{F}\circ\mathcal{C}^{l,\vec{n}}_{\mu_{p,q}\leqslant\mu}\big)\\
&=\mathbb{E}_{\vec{\epsilon}}\frac{1}{m}\sup_{\vec{C}_l\in\mathcal{C}^{l,\vec{n}}_{\mu_{p,q}\leqslant\mu}}\left|\sum_{i=1}^m\epsilon_i\vec{\alpha}\vec{f}_{C_l}(\vec{x}_i)\right|\\
&\leqslant\mathbb{E}_{\vec{\epsilon}}\frac{1}{m}\sup_{\vec{C}_l\in\mathcal{C}^{l,\vec{n}}}\frac{\mu}{\mu_{p,q}(\vec{C}_l)}\left|\sum_{i=1}^m\epsilon_i\vec{\alpha}\vec{f}_{C_l}(\vec{x}_i)\right|\\
&=\mu\mathbb{E}_{\vec{\epsilon}}\frac{1}{m}\sup_{\vec{C}_l\in\mathcal{C}^{l,\vec{n}}}\frac{1}{\mu_{p,q}(\vec{C}_l)}\left|\sum_{i=1}^m\epsilon_i\vec{\alpha}\vec{f}_{C_l}(\vec{x}_i)\right|\\
&\leqslant\mu\|\vec{\alpha}\|_p\mathbb{E}_{\vec{\epsilon}}\frac{1}{m}\sup_{C_l\in\mathcal{C}^{l,\vec{n}}}\frac{1}{\mu_{p,q}(\vec{C}_l)}\left\|\sum_{i=1}^m\epsilon_i\vec{f}_{C_l}(\vec{x}_i)\right\|_{p^*}\\
&\leqslant\mu\|\vec{\alpha}\|_p\prod_{i=1}^l 4^{n_i\max\{\frac{1}{p^*},\frac{1}{q}\}}\mathbb{E}_{\vec{\epsilon}}\frac{1}{m}\left\|\sum_{i=1}^m\epsilon_i\vec{f}_I(\vec{x}_i)\right\|_{p^*},
\end{aligned}$$

where the second inequality comes from Lemma 24. Using Lemma 15, we obtain the results of the theorem. ■

To get rid of the exponential dependence on the width of the quantum neural network, we need to take $p^*\geqslant\sum_{i=1}^{l-1}n_i$, $q\geqslant\sum_{i=1}^{l-1}n_i$. For example, we could take $p=1,q=\infty$.

Proposition 26. Given the set of depth-l quantum circuits with bounded $\mu_{1,\infty}$ norm, the Rademacher complexity on m samples $S=\{\vec{x}_1,\dots,\vec{x}_m\}$ satisfies the following bounds:

$$R_S\big(\mathcal{F}\circ\mathcal{C}^{l,\vec{n}}_{\mu_{1,\infty}\leqslant\mu}\big)\leqslant\mu\frac{\sqrt{8n_0}}{\sqrt{m}}\|\vec{\alpha}\|_1\max_i\big\|\vec{f}_I(\vec{x}_i)\big\|_\infty. \tag{C10}$$

We denote the set of depth-l variational quantum circuits with parameters $\vec{\theta}$ and fixed structure $\mathcal{A}$ by $\mathcal{C}^{l,n}_{\mathcal{A},\vec{\theta}}$. By "fixed structure," we mean that the position of each parametrized gate is fixed. Then, for the ith layer of the variational quantum circuit, denoted as $\Phi_i(\vec{\theta})$, let us define $\mu_i:=\sup_{\vec{\theta}_i}\|M^{\Phi_i(\vec{\theta})}\|_{1,\infty}$. Therefore, for any such depth-l variational quantum circuit with a fixed structure, we have $\mu_{p,q}\leqslant\prod_i\mu_i$. It follows that the Rademacher complexity of the class of depth-l variational quantum circuits with fixed structure is bounded by

$$R_S\big(\mathcal{F}\circ\mathcal{C}^{l,n}_{\mathcal{A},\vec{\theta}}\big)\leqslant\prod_i\mu_i\frac{\sqrt{8n_0}}{\sqrt{m}}\|\vec{\alpha}\|_1\max_i\|\vec{f}_I(\vec{x}_i)\|_\infty. \tag{C11}$$

2. Summation (p, q) depth-norm

In this section, let us define the summation (p,q) depth-norm for depth-l quantum circuits $\vec{C}_l=(\Phi_l,\dots,\Phi_1)$ as follows:

$$\nu^{(r)}_{p,q}(\vec{C}_l)=\left(\frac{1}{l}\sum_{i=1}^l\|M^{\Phi_i}\|^r_{p,q}\right)^{\frac{1}{r}}, \tag{C12}$$

for any $r>0$. For example, if we take $r=1$, then

$$\nu_{p,q}(\vec{C}_l)=\frac{1}{l}\sum_{i=1}^l\|M^{\Phi_i}\|_{p,q}, \tag{C13}$$

which is the average value of the amount of resources in each layer of the quantum circuit.

Proposition 27. The summation (p,q) depth-norm satisfy the following properties:

(1) Given a depth-l quantum circuit $\vec{C}_l$ and a depth-m quantum circuit $\vec{C}_m$, we have

$$(l+m)\big[\nu^{(r)}_{p,q}(\vec{C}_l\circ\vec{C}_m)\big]^r=l\big[\nu^{(r)}_{p,q}(\vec{C}_l)\big]^r+m\big[\nu^{(r)}_{p,q}(\vec{C}_m)\big]^r. \tag{C14}$$

(2) Given two depth-l quantum circuits C_l and C'_l, we have

$$\nu^{(r)}_{p,q}(\vec{C}_l\otimes\vec{C}'_l)\leqslant\nu^{(s)}_{p,q}(\vec{C}_l)\nu^{(t)}_{p,q}(\vec{C}'_l), \tag{C15}$$

where $r,s,t>0$ and $\frac{1}{s}+\frac{1}{t}=\frac{1}{r}$.

Proof. (1) follows directly from the definition of $\nu^{(r)}_{p,q}$, and (2) follows directly from Hölder's inequality. ■

Similarly, for the depth-l quantum circuit $\vec{C}_l$, where each layer only contains unitary gates, i.e., $\vec{C}_l=(U_l,U_{l-1},\dots,U_1)$, $\nu^r_{p,q}$ can be viewed as a magic resource measure.

Lemma 28. Given a depth-l quantum circuits $\vec{C}_l=(U_l,U_{l-1},\dots,U_1)$, we have

(1) (Faithfulness) For $0<p<2$, $q>0$, $r>0$: $\nu^{(r)}_{p,q}(\vec{C}_l)\geqslant 1$, and $\nu^{(r)}_{p,q}(\vec{C}_l)=1$ iff $\vec{C}_l$ is a Clifford circuit.

(1′) (Faithfulness) For $p>2$, $0<q<\infty$, $r>0$: $\nu^{(r)}_{p,q}(\vec{C}_l)\leqslant 1$, and $\nu^{(r)}_{p,q}(\vec{C}_l)=1$ iff $\vec{C}_l$ is a Clifford circuit.

(2) (Nonincreasing under Clifford circuits) For $0<p<2$, $q>0$, $r>0$, we have $\nu^{(r)}_{p,q}(\vec{C}_1\circ\vec{C}_l\circ\vec{C}_2)\leqslant\nu^{(r)}_{p,q}(\vec{C}_l)$ if $\vec{C}_1,\vec{C}_2$ are Clifford circuits.

(2′) (Nondecreasing under Clifford circuits) For $p>2$, $0<q<\infty$, $r>0$, we have $\nu^{(r)}_{p,q}(\vec{C}_1\circ\vec{C}_l\circ\vec{C}_2)\geqslant\nu^{(r)}_{p,q}(\vec{C}_l)$ if $\vec{C}_1,\vec{C}_2$ are Clifford circuits.

Proof. This lemma comes directly from Propositions 12 and 27. ■

Note that for $p>2$, $0<q<\infty$, we can define the resource measure as $1-\nu^{(r)}_{p,q}$, in which case the resource measure also satisfies the properties of faithfulness and nonincreasingness under Clifford circuits.

Let us define the set of depth-l quantum circuits with bounded $\nu^{(r)}_{p,q}$ norm by $\mathcal{C}^{l,\vec{n}}_{\nu^{(r)}_{p,q}\leqslant\nu}$. It is easy to see the following relationship $\nu^{(r)}_{p,q}$ and $\mu_{p,q}$

$$\nu^{(r)}_{p,q}(\vec{C}_l)\geqslant\mu_{p,q}(\vec{C}_l)^{1/l}, \tag{C16}$$

which follows directly from the arithmetic mean-geometric mean inequality. Hence we have

$$\mathcal{C}^{l,\vec{n}}_{\nu^{(r)}_{p,q}\leqslant\nu} \subseteq \mathcal{C}^{l,\vec{n}}_{\mu_{p,q}\leqslant\nu^l}. \tag{C17}$$

This allows us to obtain the following result on the Rademacher complexity of quantum circuits with bounded $\nu^r_{p,q}$ norm directly from Theorem 25.

Theorem 29. (Restatement of Theorem 3) Given the set of depth-l quantum circuits with bounded $\nu^{(r)}_{p,q}$, the Rademacher complexity on m independent samples $S = \{\vec{x}_1, \ldots, \vec{x}_m\}$ satisfies the following bounds:

(1) For $1 \leqslant p \leqslant 2$, we have

$$R_S\big(\mathcal{F}\circ\mathcal{C}^{l,\vec{n}}_{\nu^{(r)}_{p,q}\leqslant\nu}\big) \leqslant \nu^l 4^{(\sum_{i=1}^{l} n_i)\max\{\frac{1}{p^*},\frac{1}{q}\}}\frac{\sqrt{\min\{p^*, 8n_0\}}}{\sqrt{m}} \times \|\vec{\alpha}\|_p \max_i \|\vec{f}_I(\vec{x}_i)\|_{p^*}. \tag{C18}$$

(2) For $2 < p < \infty$, we have

$$R_S\big(\mathcal{F}\circ\mathcal{C}^{l,\vec{n}}_{\nu^{(r)}_{p,q}\leqslant\nu}\big) \leqslant \nu^l 4^{(\sum_{i=1}^{l} n_i)\max\{\frac{1}{p^*},\frac{1}{q}\}} \times \frac{\sqrt{p^*}}{m^{1/p}}\|\vec{\alpha}\|_p \max_i \|\vec{f}_I(\vec{x}_i)\|_{p^*}. \tag{C19}$$

To get rid of the exponential dependence on the width of quantum neural networks, we need to take $p^* \geqslant \sum_{i=1}^{l-1} n_i$ and $q \geqslant \sum_{i=1}^{l-1} n_i$. For example, we could take $p = 1, q = \infty$.

Proposition 30. Given the set of depth-l quantum circuits with bounded $\nu^{(r)}_{1,\infty}$, the Rademacher complexity on m independent samples $S = \{\vec{x}_1, \ldots, \vec{x}_m\}$ satisfies the following bounds:

$$R_S\big(\mathcal{F}\circ\mathcal{C}^{l,\vec{n}}_{\nu^{(r)}_{1,\infty}\leqslant\nu}\big) \leqslant \nu^l \frac{\sqrt{8n_0}}{\sqrt{m}}\|\vec{\alpha}\|_p \max_i \|\vec{f}_I(\vec{x}_i)\|_{\infty}. \tag{C20}$$

3. (p, q) path norm

Let us define the (p, q) path-norm for the depth-l quantum circuits $\vec{C}_l = (\Phi_l, \Phi_{l-1}, \ldots, \Phi_1)$. First, for a fixed output $P_{\vec{z}}$, where $\vec{z} \in \{0, 1, 2, 3\}^{n_l}$, let us define

$$\gamma^{(\vec{z})}_p(\vec{C}_l) = \left(\sum_{\substack{v_0\to v_1\to\ldots\to v_{out},\\ v_{out}=\vec{z}}} \left|M^{\Phi_l}_{\vec{z}v_{l-1}}M^{\Phi_{l-1}}_{v_{l-1}v_{l-2}}\cdots M^{\Phi_1}_{v_1v_0}\right|^p\right)^{1/p}. \tag{C21}$$

Hence, we can define the (p, q) path-norm for the depth-l quantum circuits as follows:

$$\gamma_{p,q}(\vec{C}_l) = \left(\frac{1}{4^{n_l}}\sum_{\vec{z}}\gamma^{(\vec{z})}_p(\vec{C}_l)^q\right)^{1/q}. \tag{C22}$$

Proposition 31. The multiplication (p, q) path-norm satisfies the following properties:

(1) Given a depth-l quantum circuit $\vec{C}_l$ and a depth-m quantum circuit $\vec{C}_m$, we have

$$\gamma_{p,q}(\vec{C}_l\circ\vec{C}_m) \leqslant \gamma_{p,q}(\vec{C}_l)\gamma_{p,\infty}(\vec{C}_m). \tag{C23}$$

(2) Given two depth-l quantum circuits $\vec{C}_l$ and $\vec{C}'_l$, we have

$$\gamma_{p,q}(\vec{C}_l\otimes\vec{C}'_l) = \gamma_{p,q}(\vec{C}_l)\gamma_{p,q}(\vec{C}'_l). \tag{C24}$$

Proof.

(1) holds because

$$\begin{aligned}
\gamma^{(\vec{z})}_p(\vec{C}_l\circ\vec{C}_m) &= \left(\sum_{v_0\to\ldots\to v_m\to v_{m+1}\to\ldots\to v_{l+m}=\vec{z}} |M_{\vec{z}\vec{v}_{l+m-1}}|^p\ldots|M_{\vec{v}_{m+1}\vec{v}_m}|^p\ldots|M_{\vec{v}_1\vec{v}_0}|^p\right)^{1/p}\\
&= \left(\sum_{v_m\to v_{m+1}\to\ldots\to v_{l+m}=\vec{z}} |M_{\vec{z}\vec{v}_{l+m-1}}|^p\ldots|M_{\vec{v}_{m+1}\vec{v}_m}|^p\big[\gamma^{\vec{v}_m}_p(\vec{C}_m)\big]^p\right)^{1/p}\\
&\leqslant \left(\sum_{v_m\to v_{m+1}\to\ldots\to v_{l+m}=\vec{z}} |M_{\vec{z}\vec{v}_{l+m-1}}|^p\ldots|M_{\vec{v}_{m+1}\vec{v}_m}|^p\right)^{1/p} \max_{\vec{v}_m}\gamma^{(\vec{v}_m)}_p(\vec{C}_m)\\
&= \gamma^{(\vec{z})}_p(\vec{C}_l)\gamma_{p,\infty}(\vec{C}_m).
\end{aligned}$$

Therefore, we have $\gamma_{p,q}(\vec{C}_l\circ\vec{C}_m) \leqslant \gamma_{p,q}(\vec{C}_l)\gamma_{p,\infty}(\vec{C}_m)$.

(2) holds because

$$\begin{aligned}
\gamma^{(\vec{z}_1\vec{z}_2)}_p(\vec{C}_l\otimes\vec{C}'_l) &= \left(\sum_{v_0\to v_1\to v_2\to\ldots\to v_{out}, v_{out}=\vec{z}_1\vec{z}_2} \left|M^{\Phi_l\otimes\Phi'_l}_{\vec{z}v_{l-1}}M^{\Phi_{l-1}\otimes\Phi'_{l-1}}_{v_{l-1}v_{l-2}}\cdots M^{\Phi_1\otimes\Phi'_1}_{v_1v_0}\right|^p\right)^{1/p}\\
&= \gamma^{(\vec{z}_1)}_p(\vec{C}_l)\gamma^{(\vec{z}_2)}_p(\vec{C}'_l),
\end{aligned}$$

where the second equality comes from the fact that $M^{\Phi\otimes\Phi'} = M^{\Phi}\otimes M^{\Phi'}$. ■

Proposition 32. For any depth-l quantum circuit $\vec{C}_l$, we have the following relationship: For any $0 < p \leqslant 1$, $q > 0$, we have

$$\gamma_{p,q}(\vec{C}_l) \geqslant \|M^{C_l}\|_{p,q}. \tag{C25}$$

Proof. To prove this result, we only need to prove that for any $\vec{z}$, we have

$$\gamma_p^{(\vec{z})}(\vec{C}_l) \leqslant \left\|M_{\vec{z}}^{C_l}\right\|_p.$$

This is because

$$\left\|M_{\vec{z}}^{C_l}\right\|_p = \left(\sum_{\vec{v}_0}\left|\sum_{\vec{v}_1,\ldots,\vec{v}_{l-1}} M^{\Phi_l}_{\vec{x}\vec{v}_{l-1}} M^{\Phi_{l-1}}_{\vec{v}_{l-1}\vec{v}_{l-2}}\cdots M^{\Phi_1}_{\vec{v}_1\vec{v}_0}\right|^p\right)^{1/p} \leqslant \left(\sum_{\vec{v}_0}\sum_{\vec{v}_1,\ldots,\vec{v}_{l-1}}\left|M^{\Phi_l}_{\vec{x}\vec{v}_{l-1}} M^{\Phi_{l-1}}_{\vec{v}_{l-1}\vec{v}_{l-2}}\cdots M^{\Phi_1}_{\vec{v}_1\vec{v}_0}\right|^p\right)^{1/p} = \gamma_p^{(\vec{z})}(\vec{C}_l).$$

■

For a depth-l quantum circuit $\vec{C}_l$, where each layer contains only unitary gates, i.e., $\vec{C}_l = (U_l, U_{l-1}, \ldots, U_1)$, the path norm $\gamma_{p,q}$ can be viewed as a magic resource measure.

Lemma 33. Given a depth-l quantum circuit $\vec{C}_l = (U_l, U_{l-1}, \ldots, U_1)$, we have

(1) (Faithfulness) For $0 < p \leqslant 1$, $q > 0$: $\gamma_{p,q}(\vec{C}_l) \geqslant 1$, $\gamma_{p,q}(\vec{C}_l) = 1$ iff $\vec{C}_l$ is a Clifford circuit.

(2) (Invariance under Clifford circuits) $\gamma_{p,q}(\vec{C}_1 \circ \vec{C}_l \circ \vec{C}_2) = \gamma_{p,q}(\vec{C}_l)$ if $\vec{C}_1$ and $\vec{C}_2$ are Clifford circuits.

Proof. $\gamma_{p,q}(\vec{C}_l) \geqslant 1$ comes from the facts that $\gamma_{p,q}(\vec{C}_l) \geqslant \|M^{C_l}\|_{p,q}$ and $\|M^{C_l}\|_{p,q} \geqslant 1$ (by Lemma 12).

Finally, the invariance under Clifford circuits has been proved in Proposition 31. ■

Let us define the normalized representation matrix of the quantum channel in the depth-l quantum circuit C_l as follows:

$$m^{\Phi_{k+1}}_{\vec{z}\vec{x},p} = \frac{M^{\Phi_{k+1}}_{\vec{z}\vec{x}}\gamma_p^{(\vec{x})}(\vec{C}_k)}{\gamma_p^{(\vec{z})}(\vec{C}_{k+1})}. \tag{C26}$$

It is easy to see that for any row vector $m^{\Phi_{k+1}}_{\vec{z},p}$, we have

$$\left\|m^{\Phi_{k+1}}_{\vec{z},p}\right\|_p = \left(\sum_{\vec{x}}\left|m^{\Phi_{k+1}}_{\vec{z}\vec{x},p}\right|^p\right)^{1/p} = 1\ \forall\,\vec{z}. \tag{C27}$$

Besides, it is easy to verify that

$$\gamma_p^{(\vec{z})}(\vec{C}_l) m_{\vec{z}}^{\Phi_l} m^{\Phi_{l-1}}\cdots m^{\Phi_1} = M_{\vec{z}}^{\Phi_l} M^{\Phi_{l-1}}\cdots M^{\Phi_1}. \tag{C28}$$

Therefore,

$$f_{C_l}(\vec{x}) = \vec{\alpha}\vec{f}_{C_l}(\vec{x}) = \vec{\alpha}D(\gamma(\vec{C}_l))\vec{\vec{f}}_{C_l}(\vec{x}), \tag{C29}$$

where $D(\gamma(C_l)) = \mathrm{diag}(\gamma_p^{(\vec{z})})_{\vec{z}}$, $\vec{\vec{f}}_{C_l}(\vec{x}) = m^{\Phi_l}\vec{\vec{f}}_{C_l}(\vec{x})$, and $\vec{\vec{f}}_{C_1}(\vec{x}) = m^{\Phi_1}\vec{f}_I(\vec{x})$. It is easy to see that

$$\|D(C_l)m^{\Phi_l}\|_{p,q} = \left(\frac{1}{N}\sum_{\vec{z}}\gamma_p^{(\vec{z})}(C_l)^q\right)^{\frac{1}{q}}. \tag{C30}$$

Lemma 34. For any $p^* > 0$, the following statement holds for any k-depth quantum circuits $\vec{C}_k = (\Phi_k, \Phi_{k-1}, \ldots, \Phi_1)$:

$$\mathbb{E}_{\vec{\epsilon}}\frac{1}{m}\sup_{\vec{C}_k\in\mathcal{C}^{k,\vec{n}}}\left\|\sum_{i=1}^m \epsilon_i\vec{\vec{f}}_{C_k}(\vec{x}_i)\right\|_{p^*} \leqslant 4^{n_k\frac{1}{p^*}}\mathbb{E}_{\vec{\epsilon}}\frac{1}{m}\sup_{\vec{C}_{k-1}\in\mathcal{C}^{k-1,\vec{n}}}\left\|\sum_{i=1}^m \epsilon_i\vec{\vec{f}}_{C_{k-1}}(\vec{x}_i)\right\|_{p^*}. \tag{C31}$$

Proof. This lemma comes from the fact that

$$\begin{aligned}\mathbb{E}_{\vec{\epsilon}}\frac{1}{m}\sup_{\vec{C}_k\in\mathcal{C}^{k,\vec{n}}}\left\|\sum_{i=1}^m \epsilon_i\vec{\vec{f}}_{C_k}(\vec{x}_i)\right\|_{p^*} &\leqslant 4^{n_k\frac{1}{p^*}}\mathbb{E}_{\vec{\epsilon}}\frac{1}{m}\sup_{\vec{C}_k\in\mathcal{C}^{k,\vec{n}}}\left\|\sum_{i=1}^m \epsilon_i\vec{\vec{f}}_{C_k}(\vec{x}_i)\right\|_{\infty}\\ &= 4^{n_k\frac{1}{p^*}}\mathbb{E}_{\vec{\epsilon}}\frac{1}{m}\sup_{\vec{C}_{k-1}\in\mathcal{C}^{k-1,\vec{n}}}\sup_{\vec{z}}\left|\sum_{i=1}^m \epsilon_i m_{\vec{z}}^{\Phi_k}\vec{\vec{f}}_{C_{k-1}}(\vec{x}_i)\right|\\ &\leqslant 4^{n_k\frac{1}{p^*}}\mathbb{E}_{\vec{\epsilon}}\frac{1}{m}\sup_{\vec{C}_{k-1}\in\mathcal{C}^{k-1,\vec{n}}}\left\|\sum_{i=1}^m \epsilon_i\vec{\vec{f}}_{C_{k-1}}(\vec{x}_i)\right\|_{p^*}.\end{aligned}$$

■

Theorem 35. Given the set of depth-l quantum circuits with bounded path norm $\gamma_{p,q}$, the Rademacher complexity on m independent samples $S = \{\vec{x}_1, \ldots, \vec{x}_m\}$ satisfies the following bounds:

(1) For $1 \leqslant p \leqslant 2$, we have

$$R_S\big(\mathcal{F} \circ \mathcal{C}^{l,\vec{n}}_{\gamma_{p,q}(\vec{C}_l)\leqslant\gamma}\big) \leqslant \gamma 4^{n_l \max\{\frac{1}{p^*},\frac{1}{q}\}} \prod_{i=1}^{l-1} 4^{n_i \frac{1}{p^*}} \frac{\sqrt{\min\{p^*, 8n_0\}}}{\sqrt{m}} \|\vec{\alpha}\|_p \max_i \|\vec{f}_I(\vec{x}_i)\|_{p^*}. \tag{C32}$$

(2) For $2 < p < \infty$, we have

$$R_S\big(\mathcal{F} \circ \mathcal{C}^{l,\vec{n}}_{\gamma_{p,q}(\vec{C}_l)\leqslant\gamma}\big) \leqslant \gamma 4^{n_l \max\{\frac{1}{p^*},\frac{1}{q}\}} \prod_{i=1}^{l-1} 4^{n_i \frac{1}{p^*}} \frac{\sqrt{p^*}}{m^{1/p}} \|\vec{\alpha}\|_p \max_i \|\vec{f}_I(\vec{x}_i)\|_{p^*}. \tag{C33}$$

Proof. First we compute the following:

$$\begin{aligned}
R_S\big(\mathcal{F} \circ \mathcal{C}^{l,\vec{n}}_{\gamma_{p,q}(\vec{C}_l)\leqslant\gamma}\big) &= \mathbb{E}_{\vec{\epsilon}} \frac{1}{m} \sup_{\vec{C}_l \in \mathcal{C}^{l,\vec{n}}_{\gamma_{p,q}(\vec{C}_l)\leqslant\gamma}} \left| \sum_{i=1}^m \epsilon_i \vec{\alpha} \vec{f}_{C_l}(\vec{x}_i) \right| \\
&= \mathbb{E}_{\vec{\epsilon}} \frac{1}{m} \sup_{\vec{C}_l \in \mathcal{C}^{l,\vec{n}}} \frac{\gamma}{\gamma_{p,q}(\vec{C}_l)} \left| \sum_{i=1}^m \epsilon_i \vec{\alpha} D(C_l) \vec{\tilde{f}}_{C_l}(\vec{x}_i) \right| \\
&\leqslant \|\vec{\alpha}\|_p \mathbb{E}_{\vec{\epsilon}} \frac{1}{m} \sup_{\vec{C}_l \in \mathcal{C}^{l,\vec{n}}} \frac{\gamma}{\gamma_{p,q}(\vec{C}_l)} \left\| \sum_{i=1}^m \epsilon_i D(C_l) \vec{\tilde{f}}_{C_l}(\vec{x}_i) \right\|_{p^*} \\
&= \|\vec{\alpha}\|_p \mathbb{E}_{\vec{\epsilon}} \frac{1}{m} \sup_{\vec{C}_l \in \mathcal{C}^{l,\vec{n}}} \frac{\gamma}{\gamma_{p,q}(\vec{C}_l)} \left\| \sum_{i=1}^m \epsilon_i D(C_l) m^{\Phi_l} \vec{\tilde{f}}_{C_{l-1}}(\vec{x}_i) \right\|_{p^*} \\
&\leqslant \|\vec{\alpha}\|_p \mathbb{E}_{\vec{\epsilon}} \frac{1}{m} \sup_{\vec{C}_l \in \mathcal{C}^{l,\vec{n}}} \frac{\gamma}{\gamma_{p,q}(\vec{C}_l)} 4^{n_l \max\{\frac{1}{p^*},\frac{1}{q}\}} \|D(C_l) m^{\Phi_l}\|_{p,q} \left\| \sum_{i=1}^m \epsilon_i \vec{\tilde{f}}_{C_{l-1}}(\vec{x}_i) \right\|_{p^*} \\
&= \gamma \|\vec{\alpha}\|_p 4^{n_l \max\{\frac{1}{p^*},\frac{1}{q}\}} \mathbb{E}_{\vec{\epsilon}} \frac{1}{m} \sup_{\vec{C}_l \in \mathcal{C}^{l,\vec{n}}} \left\| \sum_{i=1}^m \epsilon_i \vec{\tilde{f}}_{C_{l-1}}(\vec{x}_i) \right\|_{p^*} \\
&\leqslant \gamma \|\vec{\alpha}\|_p 4_l^{n_l \max\{\frac{1}{p^*},\frac{1}{q}\}} \prod_{i=1}^{l-1} 4^{n_i \frac{1}{p^*}} \mathbb{E}_{\vec{\epsilon}} \frac{1}{m} \left\| \sum_{i=1}^m \epsilon_i \vec{\tilde{f}}_I(\vec{x}_i) \right\|_{p^*},
\end{aligned}$$

where the last inequality comes from Lemma 34. The theorem follows from this and Lemma 15. ■

APPENDIX D: DEEP UNITAL QUANTUM CIRCUITS

In this section, let us consider depth-l unital quantum circuits, where each layer of the quantum circuit is a quantum channel. Furthermore, we shall assume that the observable H is traceless. Unlike the previous section, we consider the (p, q) norm of the modified representation matrix, where $\mathbb{I}$, and only $\mathbb{I}$, is mapped to $\mathbb{I}$.

1. Modified multiplication (p, q) depth-norm

Let us define the modified multiplication (p, q) depth-norm for depth-l unital quantum circuits $\vec{C}_l = (\Phi_l, \Phi_{l-1}, \dots, \Phi_1)$ as follows:

$$\hat{\mu}_{p,q}(\vec{C}_l) = \prod_{i=1}^{l} \|\hat{M}^{\Phi_i}\|_{p,q}. \tag{D1}$$

Lemma 36. Given a depth-l quantum circuit $\vec{C}_l = (U_l, U_{l-1}, \dots, U_1)$, we have the following:

(1) (Faithfulness) For $0 < p < 2$, $q > 0$: $\hat{\mu}_{p,q}(\vec{C}_l) \geqslant 1$, and $\hat{\mu}_{p,q}(\vec{C}_l) = 1$ iff $\vec{C}_l$ is a Clifford circuit.

(1′) (Faithfulness) For $p > 2$, $q > 0$: $\hat{\mu}_{p,q}(\vec{C}_l) \leqslant 1$, and $\hat{\mu}_{p,q}(\vec{C}_l) = 1$ iff $\vec{C}_l$ is a Clifford circuit.

(2) (Invariance under Clifford circuits) For $p > 0$, $q > 0$, we have $\hat{\mu}_{p,q}(\vec{C}_1 \circ \vec{C}_l \circ \vec{C}_2) = \hat{\mu}_{p,q}(\vec{C}_l)$ if $\vec{C}_1$, $\vec{C}_2$ are Clifford circuits.

Proof. This lemma follows directly from Proposition 19 and 22. ■

Lemma 37. Given the set of depth-l unital quantum circuit $\mathcal{C}^{l,\vec{n}}$ and the set of depth-l unital quantum circuit $\mathcal{C}^{l,\vec{n}'}$, where $\vec{n} = (\vec{n}', n_l)$ and $\vec{n}' = (\vec{n}'', n_{l-1})$, then, $\forall\, \vec{\epsilon} \in \{\pm 1\}^m$, we have

$$\begin{aligned}
&\sup_{\vec{C}_l \in \mathcal{C}^{l,\vec{n}}} \frac{1}{\hat{\mu}_{p,q}(\vec{C}_l)} \left\| \sum_{i=1}^m \epsilon_i \hat{\vec{f}}_{C_l}(\vec{x}_i) \right\|_{p^*} \\
&\quad \leqslant N_l^{\max\{\frac{1}{p^*},\frac{1}{q}\}} \sup_{\vec{C}_{l-1} \in \mathcal{C}^{l,\vec{n}}} \frac{1}{\hat{\mu}_{p,q}(\vec{C}_{l-1})} \left\| \sum_{i=1}^m \epsilon_i \hat{\vec{f}}_{C_{l-1}}(\vec{x}_i) \right\|_{p^*},
\end{aligned} \tag{D2}$$

where $N_l = 4^{n_l} - 1$. Thus,

$$\begin{aligned}
&\sup_{\vec{C}_l \in \mathcal{C}^{l,\vec{n}}} \frac{1}{\hat{\mu}_{p,q}(\vec{C}_l)} \left\| \sum_{i=1}^m \epsilon_i \hat{\vec{f}}_{C_l}(\vec{x}_i) \right\|_{p^*} \\
&\quad \leqslant \prod_{i=1}^{l} N_i^{\max\{\frac{1}{p^*},\frac{1}{q}\}} \left\| \sum_{i=1}^m \epsilon_i \hat{\vec{f}}_I(\vec{x}_i) \right\|_{p^*}.
\end{aligned} \tag{D3}$$

Proof. This is because

$$\begin{aligned}&\sup_{\vec{C}_l\in\mathcal{C}^{l,\vec{n}}}\frac{1}{\hat{\mu}_{p,q}(\vec{C}_l)}\left\|\sum_{i=1}^m\epsilon_i\hat{\vec{f}}_{C_l}(\vec{x}_i)\right\|_{p^*}\\&=\sup_{\vec{C}_{l-1}\in\mathcal{C}^{l-1,\vec{n}'}}\frac{1}{\hat{\mu}_{p,q}(\vec{C}_{l-1})\|\hat{M}^{\Phi_l}\|_{p,q}}\left\|\sum_{i=1}^m\epsilon_i\hat{M}^{\Phi_l}\hat{\vec{f}}_{C_{l-1}}(\vec{x}_i)\right\|_{p^*}\\&\leqslant N_l^{\max\{\frac{1}{p^*},\frac{1}{q}\}}\sup_{\vec{C}_{l-1}\in\mathcal{C}^{l-1,\vec{n}'}}\frac{1}{\hat{\mu}_{p,q}(\vec{C}_{l-1})}\left\|\sum_{i=1}^m\epsilon_i\hat{\vec{f}}_{C_{l-1}}(\vec{x}_i)\right\|_{p^*},\end{aligned}$$

where the inequality follows from Lemma 14. ■

Theorem 38. Given the set of depth-l unital quantum circuits with bounded depth-norm $\hat{\mu}_{p,q}$, the Rademacher complexity on m samples $S=\{\vec{x}_1,\ldots,\vec{x}_m\}$ satisfies the following bounds:

(1) For $1\leqslant p\leqslant 2$, we have

$$\begin{aligned}&R_S\big(\mathcal{F}\circ\mathcal{C}^{l,\vec{n}}_{\hat{\mu}_{p,q}\leqslant\mu}\big)\\&\leqslant\mu\prod_{i=1}^lN_i^{\max\{\frac{1}{p^*},\frac{1}{q}\}}\frac{\sqrt{\min\{p^*,8n_0\}}}{\sqrt{m}}\|\hat{\vec{\alpha}}\|_p\max_i\|\hat{\vec{f}}_I(\vec{x}_i)\|_{p^*}.\end{aligned}\tag{D4}$$

(2) For $2<p<\infty$, we have

$$R_S\big(\mathcal{F}\circ\mathcal{C}^{l,\vec{n}}_{\hat{\mu}_{p,q}\leqslant\mu}\big)\leqslant\mu\prod_{i=1}^lN_i^{\max\{\frac{1}{p^*},\frac{1}{q}\}}\frac{\sqrt{p^*}}{m^{1/p}}\|\hat{\vec{\alpha}}\|_p\max_i\|\hat{\vec{f}}_I(\vec{x}_i)\|_{p^*},\tag{D5}$$

where $N_i=4^{n_i}-1$.

Proof. The theorem follows from

$$\begin{aligned}&R_S\big(\mathcal{F}\circ\mathcal{C}^{l,\vec{n}}_{\hat{\mu}_{p,q}\leqslant\mu}\big)\\&=\mathbb{E}_{\vec{\epsilon}}\frac{1}{m}\sup_{\vec{C}_l\in\mathcal{C}^{l,\vec{n}}_{\hat{\mu}_{p,q}\leqslant\mu}}\left|\sum_{i=1}^m\epsilon_i\hat{\vec{\alpha}}\hat{\vec{f}}_{C_l}(\vec{x}_i)\right|\\&\leqslant\mathbb{E}_{\vec{\epsilon}}\frac{1}{m}\sup_{\vec{C}_l\in\mathcal{C}^{l,\vec{n}}}\frac{\mu}{\hat{\mu}_{p,q}(\vec{C}_l)}\left|\sum_{i=1}^m\epsilon_i\hat{\vec{\alpha}}\hat{\vec{f}}_{C_l}(\vec{x}_i)\right|\\&=\mu\mathbb{E}_{\vec{\epsilon}}\frac{1}{m}\sup_{\vec{C}_l\in\mathcal{C}^{l,\vec{n}}}\frac{1}{\hat{\mu}_{p,q}(\vec{C}_l)}\left|\sum_{i=1}^m\epsilon_i\hat{\vec{\alpha}}\hat{\vec{f}}_{C_l}(\vec{x}_i)\right|\\&\leqslant\mu\|\hat{\vec{\alpha}}\|_p\mathbb{E}_{\vec{\epsilon}}\frac{1}{m}\sup_{\vec{C}_l\in\mathcal{C}^{l,\vec{n}}}\frac{1}{\hat{\mu}_{p,q}(\vec{C}_l)}\left\|\sum_{i=1}^m\epsilon_i\hat{\vec{f}}_{C_l}(\vec{x}_i)\right\|_{p^*}\\&\leqslant\mu\|\hat{\vec{\alpha}}\|_p\prod_i^lN_i^{\max\{\frac{1}{p^*},\frac{1}{q}\}}\mathbb{E}_{\vec{\epsilon}}\frac{1}{m}\left\|\sum_{i=1}^m\epsilon_i\hat{\vec{f}}_I(\vec{x}_i)\right\|_{p^*},\end{aligned}$$

where the second inequality follows from Lemma 37. Using Lemma 15, we get the results of the theorem. ■

If we take $p=1$, $q=\infty$, then we have the following results directly from the previous results.

Proposition 39. Given the set of depth-l unital quantum circuits with bounded $\hat{\mu}_{1,\infty}$ norm, the Rademacher complexity on m samples $S=\{\vec{x}_1,\ldots,\vec{x}_m\}$ satisfies the following bounds:

$$R_S\big(\mathcal{F}\circ\mathcal{C}^{l,\vec{n}}_{\hat{\mu}_{1,\infty}\leqslant\mu}\big)\leqslant\mu\frac{\sqrt{8n_0}}{\sqrt{m}}\|\hat{\vec{\alpha}}\|_p\max_i\|\hat{\vec{f}}_I(\vec{x}_i)\|_\infty.\tag{D6}$$

Hence, consider a depth-l variational unitary quantum circuit on n qubits with parameters $\vec{\theta}$ and a fixed structure $\mathcal{A}$, where each layer $U_i(\vec{\theta})=\otimes_{j=1}^{k_i}U_i^{(j)}(\vec{\theta}_j)$. Then based on the properties of the (p,q) norm of the modified matrix $\hat{M}$ of unitary channels,

$$\begin{aligned}\mu_i&:=\sup_{\vec{\theta}_i}\|\hat{M}^{U_i(\vec{\theta})}\|_{1,\infty}=\sup_{\vec{\theta}}\prod_j\|\hat{M}^{U_i^{(j)}(\vec{\theta}_j)}\|_{1,\infty}\\&=\prod_j\sup_{\vec{\theta}_j}\|\hat{M}^{U_i^{(j)}(\vec{\theta}_j)}\|_{1,\infty}=\prod_j\mu_i^{(j)}.\end{aligned}\tag{D7}$$

Therefore, for any such depth-l variational quantum circuits with fixed structure, we have $\hat{\mu}_{1,\infty}\leqslant\prod_i\prod_j\mu_i^{(j)}$ and each $\mu_i^{(j)}\geqslant 1$.

Corollary 40. The Rademacher complexity of the quantum circuits class of depth-l variational quantum circuits with fixed structure is bounded by

$$R_S\big(\mathcal{F}\circ\mathcal{C}^{l,n}_{\mathcal{A},\vec{\theta}}\big)\leqslant\prod_i\prod_j\mu_i^{(j)}\frac{\sqrt{8n}}{\sqrt{m}}\|\hat{\vec{\alpha}}\|_p\max_i\|\vec{f}_I(\vec{x}_i)\|_\infty.\tag{D8}$$

2. Modified summation (p, q) depth-norm

Let us define the modified summation (p,q) depth-norm for depth-l quantum circuits as follows:

$$\hat{\nu}^r_{p,q}(\vec{C}_l)=\left(\frac{1}{l}\sum_{i=1}^l\|\hat{M}^{\Phi_i}\|^r_{p,q}\right)^{\frac{1}{r}},\tag{D9}$$

for any $r>0$.

Lemma 41. Given a depth-l quantum circuit $\vec{C}_l=(U_l,U_{l-1},\ldots,U_1)$, we have the following

(1) (Faithfulness) For $0<p<2$, $q>0$, $r>0$: $\hat{\nu}^{(r)}_{p,q}(\vec{C}_l)\geqslant 1$, and $\hat{\nu}^{(r)}_{p,q}(\vec{C}_l)=1$ iff $\vec{C}_l$ is a Clifford circuit.

(1′) (Faithfulness) For $p>2$, $q>0$, $r>0$: $\hat{\nu}^{(r)}_{p,q}(\vec{C}_l)\leqslant 1$ and $\hat{\nu}^{(r)}_{p,q}(\vec{C}_l)=1$ iff $\vec{C}_l$ is a Clifford circuit.

(2) (Nonincreasing under Clifford circuits) For $0<p<2$, $q>0$, $r>0$, we have $\hat{\nu}^{(r)}_{p,q}(\vec{C}_1\circ\vec{C}_l\circ\vec{C}_2)\leqslant\hat{\nu}^{(r)}_{p,q}(\vec{C}_l)$ if $\vec{C}_1$, $\vec{C}_2$ are Clifford circuits.

(2′) (Nondecreasing under Clifford circuits) For $p>2$, $q>0$, $r>0$, we have $\hat{\nu}^{(r)}_{p,q}(\vec{C}_1\circ\vec{C}_l\circ\vec{C}_2)\geqslant\hat{\nu}^{(r)}_{p,q}(\vec{C}_l)$ if $\vec{C}_1$, $\vec{C}_2$ are Clifford circuits.

Proof. This lemma comes directly from Lemma 19 and Proposition 27. ■

Based on the following relationship between $\hat{\nu}^{(r)}_{p,q}$ and $\hat{\mu}_{p,q}$:

$$\hat{\nu}^{(r)}_{p,q}(\vec{C}_l)\geqslant\hat{\mu}_{p,q}(\vec{C}_l)^{1/l},\tag{D10}$$

we have

$$\mathcal{C}^{l,\vec{n}}_{\hat{\nu}_{p,q}\leqslant\nu}\subseteq\mathcal{C}^{l,\vec{n}}_{\hat{\mu}_{p,q}\leqslant\nu^l}.\tag{D11}$$

We obtain the following results directly from Theorem 38.

Theorem 42. (Restatement of Theorem 4) Given the set of depth-l unital quantum circuits with bounded $\hat{\nu}^{(r)}_{p,q}$ norm, the Rademacher complexity on m samples $S = \{\vec{x}_1, \ldots, \vec{x}_m\}$ satisfies the following bounds:

(1) For $1 \leqslant p \leqslant 2$, we have

$$R_S(\mathcal{F} \circ \mathcal{C}^{l,\vec{n}}_{\hat{\nu}^r_{p,q} \leqslant \nu}) \leqslant \nu^l \prod_{i=1}^{l} N_i^{\max\{\frac{1}{p^*},\frac{1}{q}\}} \frac{\sqrt{\min\{p^*, 8n_0\}}}{\sqrt{m}} \|\hat{\vec{\alpha}}\|_p \max_i \|\vec{f}_I(\vec{x}_i)\|_{p^*}. \tag{D12}$$

(2) For $2 < p < \infty$, we have

$$R_S(\mathcal{F} \circ \mathcal{C}^{l,\vec{n}}_{\hat{\nu}^{(r)}_{p,q} \leqslant \nu}) \leqslant \nu^l \prod_{i=1}^{l} N_i^{\max\{\frac{1}{p^*},\frac{1}{q}\}} \frac{\sqrt{p^*}}{m^{1/p}} \|\hat{\vec{\alpha}}\|_p \max_i \|\vec{f}_I(\vec{x}_i)\|_{p^*}. \tag{D13}$$

If we take $p = 1, q = \infty$, then we get the following results directly from the previous results.

Proposition 43. Given the set of depth-l unital quantum circuits with bounded $\hat{\nu}^r_{1,\infty}$ norm, the Rademacher complexity on m samples $S = \{\vec{x}_1, \ldots, \vec{x}_m\}$ satisfies the following bounds:

$$R_S(\mathcal{F} \circ \mathcal{C}^{l,\vec{n}}_{\hat{\nu}^r_{1,\infty} \leqslant \nu}) \leqslant \nu^l \frac{\sqrt{8n_0}}{\sqrt{m}} \|\hat{\vec{\alpha}}\|_p \max_i \|\hat{\vec{f}}_I(\vec{x}_i)\|_\infty. \tag{D14}$$

3. Modified (p, q) path-norm

Let us define the modified (p, q) path-norm for depth-l circuits by

$$\hat{\gamma}_{p,q}(\vec{C}_l) = \left(\frac{1}{N_l}\sum_{\vec{x}\neq\vec{0}} \gamma_p^{(\vec{x})}(\vec{C}_l)^q\right)^{1/q}, \tag{D15}$$

where $N_l = 4^{n_l} - 1$.

Let us define the normalized representation matrix of quantum channels in depth-l quantum circuits $\vec{C}_l$ as follows:

$$\hat{m}^{\Phi_{k+1}}_{\vec{z}\vec{x},p} = \frac{\hat{M}^{\Phi_{k+1}}_{\vec{z}\vec{x}} \gamma_p^{(\vec{x})}(\vec{C}_k)}{\gamma_p^{(\vec{z})}(\vec{C}_{k+1})}. \tag{D16}$$

It is easy to see that for any row vector $\hat{m}^{\Phi_{k+1}}_{\vec{z},p}$, we have

$$\left\|\hat{m}^{\Phi_{k+1}}_{\vec{z},p}\right\|_p = \left(\sum_{\vec{x}} \left|\hat{m}^{\Phi_{k+1}}_{\vec{z}\vec{x},p}\right|^p\right)^{1/p} = 1 \ \forall\, \vec{z}. \tag{D17}$$

Besides, it is easy to verify that

$$\gamma_p^{(\vec{z})}(\vec{C}_l)\hat{m}^{\Phi_l}_{\vec{z}}\hat{m}^{\Phi_{l-1}}\cdots\hat{m}^{\Phi_1} = \hat{M}^{\Phi_l}_{\vec{z}}\hat{M}^{\Phi_{l-1}}\cdots\hat{M}^{\Phi_1}. \tag{D18}$$

Therefore,

$$f_{C_l}(\vec{x}) = \hat{\vec{\alpha}}\hat{\vec{f}}_{C_l}(\vec{x}) = \hat{\vec{\alpha}}\hat{D}(\gamma(\vec{C}_l))\hat{\hat{\vec{f}}}_{C_l}(\vec{x}), \tag{D19}$$

where $\hat{D}(\gamma(C_l)) = \mathrm{diag}(\gamma_p^{\vec{z}})_{\vec{z}\neq\vec{0}}$, $\hat{\hat{\vec{f}}} = m^{\Phi_l}\hat{\hat{\vec{f}}}_{C_l}(\vec{x})$, and $\hat{\hat{\vec{f}}}_{C_1}(\vec{x}) = m^{\Phi_1}\hat{\vec{f}}_I(\vec{x})$. It is easy to see that

$$\|\hat{D}(C_l)\hat{m}^{\Phi_l}\|_{p,q} = \left(\frac{1}{N_l}\sum_{\vec{z}\neq\vec{0}}\gamma_p^{(\vec{z})}(\vec{C}_l)^q\right)^{\frac{1}{q}}. \tag{D20}$$

Similarly to $\gamma_{p,q}$, $\hat{\gamma}_{p,q}$ satisfies the following property:

Proposition 44. For any depth-l unital quantum circuit $\vec{C}_l$, we have the following relationship: For any $0 < p \leqslant 1, q > 0$, we have

$$\hat{\gamma}_{p,q}(\vec{C}_l) \geqslant \|\hat{M}^{C_l}\|_{p,q}. \tag{D21}$$

Proof. The proof is similar to that of Proposition 32. ■

Lemma 45. Given a depth-l quantum circuit $\vec{C}_l = (U_l, U_{l-1}, \ldots, U_1)$, we have the following:

(1) (Faithfulness) For $0 < p \leqslant 1$, $q > 0$, $\hat{\gamma}_{p,q}(\vec{C}_l) \geqslant 1$, $\gamma_{p,q}(\vec{C}_l) = 1$ iff $\vec{C}_l$ is Clifford.

(2) (Invariance under Clifford circuits) $\hat{\gamma}_{p,q}(\vec{C}_1 \circ \vec{C}_l \circ \vec{C}_2) = \hat{\gamma}_{p,q}(\vec{C}_l)$ if $\vec{C}_1$ and $\vec{C}_2$ are Clifford circuits.

Proof. The proof is similar to that of Lemma 33. ■

Theorem 46. Given the set of depth-l unital quantum circuits with bounded path norm $\gamma_{p,q}$, the Rademacher complexity on m samples $S = \{\vec{x}_1, \ldots, \vec{x}_m\}$ satisfies the following bounds:

(1) For $1 \leqslant p \leqslant 2$, we have

$$R_S(\mathcal{F} \circ \mathcal{C}^{l,\vec{n}}_{\hat{\gamma}_{p,q}(C_l) \leqslant \gamma}) \leqslant \gamma N_l^{\max\{\frac{1}{p^*},\frac{1}{q}\}} \prod_{i=1}^{l-1} N_i^{\frac{1}{p^*}} \frac{\sqrt{\min\{p^*, 8n_0\}}}{\sqrt{m}} \times \|\hat{\vec{\alpha}}\|_p \max_i \|\vec{f}_I(\vec{x}_i)\|_{p^*}. \tag{D22}$$

(2) For $2 < p < \infty$, we have

$$R_S(\mathcal{F} \circ \mathcal{C}^{l,\vec{n}}_{\hat{\gamma}_{p,q}(C_l) \leqslant \gamma}) \leqslant \gamma N_l^{\max\{\frac{1}{p^*},\frac{1}{q}\}} \prod_{i=1}^{l-1} N_i^{\frac{1}{p^*}} \frac{\sqrt{p^*}}{m^{1/p}} \|\hat{\vec{\alpha}}\|_p \max_i \left\|\vec{f}_I(\vec{x}_i)\right\|_{p^*}, \tag{D23}$$

where $N_i = 4^{n_i} - 1$.

Proof. The statement in the theorem holds because

$$\begin{aligned} R_S(\mathcal{F} \circ \mathcal{C}^{l,\vec{n}}_{\hat{\gamma}_{p,q}(C_l) \leqslant \gamma}) &= \mathbb{E}_{\vec{\epsilon}} \frac{1}{m} \sup_{C_l \in \mathcal{C}^{l,\vec{n}}_{\hat{\gamma}_{p,q}(C_l)\leqslant\gamma}} \left|\sum_{i=1}^m \epsilon_i \hat{\vec{\alpha}}\hat{\vec{f}}_{C_l}(\vec{x}_i)\right| \\ &= \mathbb{E}_{\vec{\epsilon}} \frac{1}{m} \sup_{C_l \in \mathcal{C}^{l,\vec{n}}} \frac{\gamma}{\hat{\gamma}_{p,q}(C_l)} \left|\sum_{i=1}^m \epsilon_i \hat{\vec{\alpha}}\hat{D}(C_l)\hat{\hat{\vec{f}}}_{C_l}(\vec{x}_i)\right| \\ &\leqslant \|\hat{\vec{\alpha}}\|_p \mathbb{E}_{\vec{\epsilon}} \frac{1}{m} \sup_{C_l \in \mathcal{C}^{l,\vec{n}}} \frac{\gamma}{\hat{\gamma}_{p,q}(C_l)} \left\|\sum_{i=1}^m \epsilon_i D(C_l)\hat{\hat{\vec{f}}}_{C_l}(\vec{x}_i)\right\|_{p^*} \end{aligned}$$

$$
\begin{aligned}
&= \|\hat{\vec{\alpha}}\|_p \mathbb{E}_{\vec{\epsilon}} \frac{1}{m} \sup_{C_l \in \mathcal{C}^{l,\vec{n}}} \frac{\gamma}{\hat{\gamma}_{p,q}(C_l)} \left\| \sum_{i=1}^m \epsilon_i \hat{D}(C_l) \hat{m}^{\Phi_l} \vec{\hat{f}}_{C_{l-1}}(\vec{x}_i) \right\|_{p^*} \\
&\leqslant \|\hat{\vec{\alpha}}\|_p \mathbb{E}_{\vec{\epsilon}} \frac{1}{m} \sup_{C_l \in \mathcal{C}^{l,\vec{n}}} \frac{\gamma}{\hat{\gamma}_{p,q}(C_l)} (N-1)^{\max\{\frac{1}{p^*},\frac{1}{q}\}} \left\| \hat{D}(C_l) \hat{m}^{\Phi_l} \right\|_{p,q} \left\| \sum_{i=1}^m \epsilon_i \vec{\hat{f}}_{C_{l-1}}(\vec{x}_i) \right\|_{p^*} \\
&= \gamma \|\hat{\vec{\alpha}}\|_p N^{\max\{\frac{1}{p^*},\frac{1}{q}\}} \mathbb{E}_{\vec{\epsilon}} \frac{1}{m} \sup_{C_l \in \mathcal{C}^{l,\vec{n}}} \left\| \sum_{i=1}^m \epsilon_i \vec{\hat{f}}_{C_{l-1}}(\vec{x}_i) \right\|_{p^*} \\
&\leqslant \gamma \|\hat{\vec{\alpha}}\|_p N^{\max\{\frac{1}{p^*},\frac{1}{q}\}} \prod_{i=1}^{l-1} N_i^{\frac{1}{p^*}} \mathbb{E}_{\vec{\epsilon}} \frac{1}{m} \left\| \sum_{i=1}^m \epsilon_i \vec{\hat{f}}_{I}(\vec{x}_i) \right\|_{p^*},
\end{aligned}
$$

where the last inequality follows from Lemma 34. Using Lemma 15 completes the proof of the theorem. ■

APPENDIX E: PROOF OF PROPOSITION 7

Now let us prove Proposition 7, which comes directly from Theorem 2 with $p = 1$, $q = \infty$ and the following lemma:

Lemma 47 (Bartlett and Mendelson [23]). If the loss function $l(f(\vec{x}), \vec{y})$ takes values in $[0, B]$, then for any $\delta > 0$, the following statement holds for any function $f \in \mathcal{F}$ with probability at least $1 - \delta$:

$$|L(f) - \hat{L}(f)| \leqslant +2B\hat{R}_S(l_{\mathcal{F}}) + 3B\sqrt{\frac{\log_2(2/\delta)}{2m}},$$

where the function class $l_{\mathcal{F}} := \{l_f : (\vec{x}, \vec{y}) \to l(f(\vec{x}), \vec{y}) \mid f \in \mathcal{F}\}$, and $\hat{R}_S(l_{\mathcal{F}})$ is the Rademacher complexity of the function class $l_{\mathcal{F}}$ on the m given samples $S = \{(\vec{x}_i, \vec{y}_i)\}_{i=1}^m$.

[1] Y. LeCun, Y. Bengio, and G. Hinton, Deep learning, Nature (London) **521**, 436 (2015).

[2] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning* (MIT Press, Cambridge, 2016); http://www.deeplearningbook.org

[3] K. P. Murphy, *Probabilistic Machine Learning: An Introduction* (MIT Press, Cambridge, 2021).

[4] T. Young, D. Hazarika, S. Poria, and E. Cambria, Recent trends in deep learning based natural language processing, IEEE Comput. Intell. Mag. **13**, 55 (2018).

[5] L. Deng and Y. Liu, *Deep Learning in Natural Language Processing* (Springer, 2018).

[6] H. Li, Deep learning for natural language processing: Advantages and challenges, Natl. Sci. Rev. **5**, 24 (2018).

[7] Y. Jing, Y. Bian, Z. Hu, L. Wang, and X.-Q. Xie, Deep learning for drug design: An artificial intelligence paradigm for drug discovery in the big data era, AAPS J. **20**, 58 (2018).

[8] E. Gawehn, J. A. Hiss, and G. Schneider, Deep learning in drug discovery, Mol. Inf. **35**, 3 (2016).

[9] A. Roy, J. Sun, R. Mahoney, L. Alonzi, S. Adams, and P. Beling, Deep learning detecting fraud in credit card transactions, in *2018 Systems and Information Engineering Design Symposium (SIEDS)* (2018), pp. 129–134.

[10] A. Pumsirirat and L. Yan, Credit card fraud detection using deep learning based on auto-encoder and restricted Boltzmann machine, Int. J. Adv. Comput. Sci. Appl. **9**, 18, (2018).

[11] D. Shen, G. Wu, and H.-I. Suk, Deep learning in medical image analysis, Annu. Rev. Biomed. Eng. **19**, 221 (2017).

[12] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. W. M. van der Laak, B. van Ginneken, and C. I. Sánchez, A survey on deep learning in medical image analysis, Med. Image Anal. **42**, 60 (2017).

[13] S. Ramos, S. Gehrig, P. Pinggera, U. Franke, and C. Rother, Detecting unexpected obstacles for self-driving cars: Fusing deep learning and geometric modeling, in *2017 IEEE Intelligent Vehicles Symposium (IV)* (2017), pp. 1025–1032.

[14] Q. Rao and J. Frtunikj, Deep learning for self-driving cars: Chances and challenges, in *2018 IEEE/ACM 1st International Workshop on Software Engineering for AI in Autonomous Systems (SEFAIAS)* (2018), pp. 35.

[15] V. Pham, T. Bluche, C. Kermorvant, and J. Louradour, Dropout improves recurrent neural networks for handwriting recognition, in *2014 14th International Conference on Frontiers in Handwriting Recognition* (2014), pp. 285–290.

[16] M. M. Abu Ghosh and A. Y. Maghari, A comparative study on handwriting digit recognition using neural networks, in *2017 International Conference on Promising Electronic Technologies (ICPET)* (2017), pp. 77–81.

[17] A. Voulodimos, N. Doulamis, A. Doulamis, E. Protopapadakis, and D. Andina, Deep learning for computer vision: A brief review, Intell. Neuroscience **2018**, 7068349 (2018).

[18] M. A. Ponti, L. S. F. Ribeiro, T. S. Nazare, T. Bui, and J. Collomosse, Everything you wanted to know about deep learning for computer vision but were afraid to ask, in *2017 30th SIBGRAPI Conference on Graphics, Patterns and Images Tutorials (SIBGRAPI-T)* (2017), pp. 17–41.

[19] M. A. Nielsen, *Neural Networks and Deep Learning* (Determination Press, San Francisco, 2015), Vol. 2018.

[20] V. N. Vapnik and A. Ya. Chervonenkis, On the uniform convergence of relative frequencies of events to their probabilities, Theory Probab. Its Appl. (Engl. Transl.) **16**, 264 (1971).

[21] V. N. Vapnik and A. Ya. Chervonenkis, Necessary and sufficient conditions for the uniform convergence of means to their expectations, Theory Probab. Its Appl. (Engl. Transl.) **26**, 532 (1982).

[22] V. M. Tikhomirov, ε-entropy and ε-capacity of sets in functional spaces, in *Selected Works of AN Kolmogorov* (Springer, 1993), pp. 86–170.

[23] P. L. Bartlett and S. Mendelson, Rademacher and Gaussian complexities risk bounds and structural results, in *Computational Learning Theory: COLT 2001*, edited by D. Helmbold and B. Williamson, Lecture Notes in Computer Science Vol. 2111 (Springer, Berlin, Heidelberg, 2001).

[24] M. Telgarsky, Benefits of depth in neural networks, in *29th Annual Conference on Learning Theory*, Proceedings of Machine Learning Research, edited by Vitaly Feldman, Alexander Rakhlin, and Ohad Shamir (PMLR, Columbia University, New York, 2016), Vol. 49, pp. 1517–1539.

[25] B. Neyshabur, R. Tomioka, and N. Srebro, Norm-based capacity control in neural networks, in *Proceedings of The 28th Conference on Learning Theory*, Proceedings of Machine Learning Research (PMLR, Paris, France, 2015), Vol. 40, pp. 1376–1401.

[26] N. Harvey, C. Liaw, and A. Mehrabian, Nearly-tight VC-dimension bounds for piecewise linear neural networks, in *Proceedings of the 2017 Conference on Learning Theory*, Proceedings of Machine Learning Research, edited by S. Kale and O. Shamir (PMLR, Amsterdam, 2017), Vol. 65, pp. 1064–1068.

[27] P. L. Bartlett, D. J. Foster, and M. J. Telgarsky, Spectrally-normalized margin bounds for neural networks, in *Advances in Neural Information Processing Systems*, edited by I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (Curran Associates, Inc., 2017), Vol. 30, pp. 6240–6249.

[28] B. Neyshabur, S. Bhojanapalli, D. Mcallester, and N. Srebro, Exploring generalization in deep learning, in *Advances in Neural Information Processing Systems* (Curran Associates, Inc., 2017), Vol. 30, pp. 5947–5956.

[29] N. Golowich, A. Rakhlin, and O. Shamir, Size-independent sample complexity of neural networks, in *Proceedings of the 31st Conference On Learning Theory*, Proceedings of Machine Learning Research (PMLR, 2018), Vol. 75, pp. 297–299.

[30] National Academies of Sciences, Engineering, Quantum Computing: Progress and Prospects, National Academies Press (2018).

[31] J. Preskill, Quantum computing in the NISQ era and beyond, Quantum **2**, 79 (2018).

[32] A. Peruzzo, J. McClean, P. Shadbolt, M.-H. Yung, X.-Q. Zhou, P. J. Love, A. Aspuru-Guzik, and J. L. O'Brien, A variational eigenvalue solver on a photonic quantum processor, Nat. Commun. **5**, 4213 (2014).

[33] E. Farhi, J. Goldstone, and S. Gutmann, A quantum approximate optimization algorithm, arXiv:1411.4028.

[34] L. Zhou, S.-T. Wang, S. Choi, H. Pichler, and M. D. Lukin, Quantum Approximate Optimization Algorithm: Performance, Mechanism, and Implementation on Near-Term Devices, Phys. Rev. X **10**, 021067 (2020).

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

[36] F. Arute, K. Arya, R. Babbush, D. Bacon, J. C. Bardin, R. Barends, R. Biswas, S. Boixo, F. G. S. L. Brandao, D. A. Buell *et al.*, Quantum supremacy using a programmable superconducting processor, Nature (London) **574**, 505 (2019).

[37] H.-S. Zhong, H. Wang, Y.-H. Deng, M.-C. Chen, L.-C. Peng, Y.-H. Luo, J. Qin, D. Wu, X. Ding, Y. Hu *et al.*, Quantum computational advantage using photons, Science **370**, 1460 (2020).

[38] C. Palacios-Berraquero, L. Mueck, and D. M. Persaud, Instead of 'supremacy' use 'quantum advantage', Nature (London) **576**, 213 (2019).

[39] J. Preskill, Quantum computing and the entanglement frontier, arXiv:1203.5813.

[40] A. P. Lund, M. J. Bremner, and T. C. Ralph, Quantum sampling problems, Boson sampling and quantum supremacy, npj Quantum Inf. **3**, 15 (2017).

[41] A. W. Harrow and A. Montanaro, Quantum computational supremacy, Nature (London) **549**, 203 (2017).

[42] A. M. Dalzell, A. W. Harrow, D. E. Koh, and R. L. La Placa, How many qubits are needed for quantum computational supremacy? Quantum **4**, 264 (2020).

[43] E. Farhi and H. Neven, Classification with quantum neural networks on near term processors, arXiv:1802.06002.

[44] K. Beer, D. Bondarenko, T. Farrelly, T. J. Osborne, R. Salzmann, D. Scheiermann, and R. Wolf, Training deep quantum neural networks, Nat. Commun. **11**, 808 (2020).

[45] K. Sharma, M. Cerezo, L. Cincio, and P. J. Coles, Trainability of dissipative perceptron-based quantum neural networks, arXiv:2005.12458.

[46] M. Schuld, I. Sinayskiy, and F. Petruccione, The quest for a quantum neural network, Quantum Inf. Process. **13**, 2567 (2014).

[47] N. Killoran, T. R. Bromley, J. M. Arrazola, M. Schuld, N. Quesada, and S. Lloyd, Continuous-variable quantum neural networks, Phys. Rev. Research **1**, 033063 (2019).

[48] I. Cong, S. Choi, and M. D. Lukin, Quantum convolutional neural networks, Nat. Phys. **15**, 1273 (2019).

[49] S. Lloyd, M. Mohseni, and P. Rebentrost, Quantum algorithms for supervised and unsupervised machine learning, arXiv:1307.0411.

[50] P. Wittek, *Quantum Machine Learning: What Quantum Computing Means to Data Mining* (Academic Press, 2014).

[51] J. Biamonte, P. Wittek, N. Pancotti, P. Rebentrost, N. Wiebe, and S. Lloyd, Quantum machine learning, Nature (London) **549**, 195 (2017).

[52] C. Ciliberto, M. Herbster, A. D. Ialongo, M. Pontil, A. Rocchetto, S. Severini, and L. Wossnig, Quantum machine learning: a classical perspective, Proc. R. Soc. London A **474**, 20170551 (2018).

[53] V. Dunjko and H. J. Briegel, Machine learning & artificial intelligence in the quantum domain: A review of recent progress, Rep. Prog. Phys. **81**, 074001 (2018).

[54] B. Coecke, T. Fritz, and R. W. Spekkens, A mathematical theory of resources, Inf. Comput. **250**, 59 (2016).

[55] E. Chitambar and G. Gour, Quantum resource theories, Rev. Mod. Phys. **91**, 025001 (2019).

[56] M. Howard and E. Campbell, Application of a Resource Theory for Magic States to Fault-Tolerant Quantum Computing, Phys. Rev. Lett. **118**, 090501 (2017).

[57] V. Koltchinskii, Local Rademacher complexities and oracle inequalities in risk minimization, Ann. Stat. **34**, 2593 (2006).

[58] D. Gottesman, The Heisenberg representation of quantum computers, Group22: Proceedings of the XXII International Colloquium on Group Theoretical Methods in Physics, 32–43 (1999).

[59] S. Bravyi, G. Smith, and J. A. Smolin, Trading Classical and Quantum Computational Resources, Phys. Rev. X **6**, 021043 (2016).

[60] S. Bravyi, D. Browne, P. Calpin, E. Campbell, D. Gosset, and M. Howard, Simulation of quantum circuits by low-rank stabilizer decompositions, Quantum **3**, 181 (2019).

[61] K. Bu and D. E. Koh, Efficient Classical Simulation of Clifford Circuits with Nonstabilizer Input States, Phys. Rev. Lett. **123**, 170502 (2019).

[62] K. Bu and D. E. Koh, Classical simulation of quantum circuits by half Gauss sums, Commun. Math. Phys. **390**, 471 (2022).

[63] R. M. Dudley, The sizes of compact subsets of Hilbert space and continuity of Gaussian processes, J. Funct. Anal. **1**, 290 (1967).

[64] V. N. Sudakov, Gaussian random processes and measures of solid angles in a Hilbert space, Dokl. Math. **12**, 412 (1971).

[65] S. Mendelson and R. Vershynin, Entropy and the combinatorial dimension, Invent. Math. **152**, 37 (2003).

[66] M. C. Caro and I. Datta, Pseudo-dimension of quantum circuits, Quantum Mach. Intell. **2**, 14 (2020).

[67] K. Bu, Y. Zhang, and Q. Luo, Depth-width trade-offs for neural networks via topological entropy, arXiv:2010.07587.

[68] V. Veitch, S. A. H. Mousavian, D. Gottesman, and J. Emerson, The resource theory of stabilizer quantum computation, New J. Phys. **16**, 013009 (2014).

[69] C. Ferrie and J. Emerson, Framed Hilbert space: Hanging the quasi-probability pictures of quantum theory, New J. Phys. **11**, 063040 (2009).

[70] S. Shalev-Shwartz and S. Ben-David, *Understanding Machine Learning: From Theory to Algorithms* (Cambridge University Press, New York, 2014).