Bulletin of Networking, Computing, Systems, and Software – www.bncss.org, ISSN 2186-5140
Volume 11, Number 1, pages 7–11, January 2022

Learnability and expressibility of variational quantum circuit: VC and PAC-Bayesian theory

1st Chih-Chieh Chen
Grid Inc.
Tokyo, Japan
chen.chih.chieh@gridsolar.jp

2nd Masaya Watabe
Engineering Department
The University of Electro-Communications
Tokyo, Japan

3rd Kodai Shiba
Grid Inc.
Tokyo, Japan

4th Masaru Sogabe
Grid Inc.
Tokyo, Japan

5th Katsuyoshi Sakamoto
Engineering Department
The University of Electro-Communications
Tokyo, Japan

6th Tomah Sogabe
Engineering Department
The University of Electro-Communications
Tokyo, Japan
sogabe@uec.ac.jp

***Abstract*—Variational quantum circuit algorithm is suggested as a machine learning paradigm to utilize the power of Noisy Intermediate-Scale Quantum hardware. Recent efforts lead to better theoretical understanding of the expressibility and generalization ability of quantum circuits. In this work, we compare our previous results on the quantum circuit learnability based on VC theory to some preliminary results based on PAC-Bayesian method. Some recent developments based on other generalization bounds due to other research groups are also reviewed and discussed.**

***Index Terms*—Quantum circuit, PAC learning, VC theory, PAC-Bayesian**

I. Introduction

A. Background

Supervised machine learning [1] is a possible direction to harvest the advantage of Noisy Intermediate-Scale Quantum (NISQ) [2], [3] hardware. Applying variational quantum circuit [4], [5] for hybrid quantum-classical supervised machine learning leads to the quantum circuit learning (QCL) method (or quantum neural network) [6]–[10]. There are recent efforts toward the understanding of expressitivity and model complexity of QCL [11]–[15]. The learnability of machines can be understood with the Probably Approximately Correct (PAC) model proposed by Valiant [16]–[19]. If a learning machine has too much model complexity, it is possible to overfit the data and the generalization ability would be bad. For classiciation tasks, Vapnik–Chervonenkis (VC) theory [20], [21] can be used to establish the generalization ability by using the VC dimension [22]. For regression tasks, the pseudo-dimension [23] and the fat-shattering dimension [24], [25] could be used. PAC-Bayesian method [26]–[28] provides an alternative way to get uniform PAC bounds, which is usually tighter than that of VC theory [1]. Recent large-scale comparative experiments of various generalization measures suggests that PAC-Bayesian bound is useful [29].

B. Our results

Our recent study gives a VC dimesnion upper bound in terms of circuit width and circuit depth, and provides a recipe to control the model complexity of QCL [30]–[32]. In this report, we present a summary and some further discussions and investigations related to the expressibility and learnability. In particular, we provide a PAC-Bayesian bound. The PAC-Bayesian generalization bound is then compared to the VC generalization bound in the case of QCL.

C. Related works

Many studies on the expressibility of quantum circuit have been published [11]–[15]. Previous learnability results for quantum machine learning are based on fat-shattering dimension for regression [33], pseudo-dimension for regression [34], or quantum sample complexity for coherent quantum states samples [35]. Recent researches on the quantum circuit learnability lead to various generalization bounds. Abbas et al. [36] give input-dependent results by using Fisher information. Huang et al. [37] also provide input-dependent result based on a different measure. Caro et al. give generalization bound by using Rademacher complexity for Lipschitz loss functions [38]. Their result is asymptotically equivalent to our result and they also include more general encoding scheme. Du et al. [39] give generalization bound for qudit systems by using covering number bound [40] for Lipschitz loss functions. Bu et al. [41] give Rademacher complexity bound in terms of $L_{p,q}$ matrix norm. Another VC-dimension upper bound, which is different from our result, is recently proposed in Ref. [42]. Their bound is in terms of the property of output measurable operator.

II. Methods

A. Variational quantum circuit learning

In supervised binary classification learning problem, there is an unknown target function $f : X \mapsto Y = \{-1, 1\}$, and we are given some training dataset $\{(\vec{x}_i, y_i = f(\vec{x}_i))|\vec{x}_i \in X\}$ where $\vec{x}_i$ is drawn from some unknown distribution $\vec{x}_i \sim P(\vec{x}_i)$. (In the setting that the training data is noisy, the dataset is drawn from some unknown joint distribution $(\vec{x}_i, y_i) \sim P(\vec{x}, y)$.) The goal of learning is to obtain a hypothesis $h : X \mapsto Y$ such that the prediction error (out-of-sample error) $E_{out}(h) =$

$\mathbb{P}_{\vec{x}\sim P(\vec{x})}[h(\vec{x}) \neq f(\vec{x})]$ (or $E_{out}(h) = \mathbb{P}_{(\vec{x},y)\sim P(\vec{x},y)}[h(\vec{x}) \neq y]$ for the noisy case) is small.

In QCL [6], [7] we use quantum circuit as the hypothesis set. The number of qubits n is also called circuit width. The circuit starts with an input layer where the input data $\vec{x}$ is encoded by $|\psi_{in}(\vec{x})\rangle = \otimes_{i=0}^{n-1} U(\theta_{i,in}(\vec{x}))|0\rangle$. After the input layer, the trainable part $U_\theta = (U_{loc}^{(L)}(\theta_L)U_{ent})...(U_{loc}^{(1)}(\theta_1)U_{ent})U_{loc}^{(0)}(\theta_0)$ consists of alternating layers of trainable single qubit rotations $U_{loc}^{(j)}(\theta_j) = \otimes_{i=0}^{n-1} U(\theta_{i,j})$ and fixed (not trainable) entangling layer. The local rotation could include any subset of $\{R_X, R_Y, R_Z\}$. The number of layers is a hyper-parameter, also known as circuit depth L. The measurement result is used to compute the Z expectation values of i-th qubit

$$\langle Z_i(\theta, \vec{x})\rangle \qquad = \mathrm{Tr}(\rho_{in}(\vec{x})U_\theta^\dagger Z_i U_\theta), \tag{1}$$

where $\rho_{in} = |\psi_{in}(\vec{x})\rangle\langle\psi_{in}(\vec{x})|$. The expectation value is then thresholded by some value t for the binary classification, leading to a hypothesis set $H = \{\mathrm{sgn}(\langle Z_i(\theta, \vec{x})\rangle - t)\}$. Multi-class classification is done by the usual softmax method. Since we are interested in NISQ applications, we use the Hardware Efficient Ansatz (HEA) where the entangler consists of 2-qubit quantum gate acting on nearest neighbor qubits for some lattice. The learning circuit studied in this work is depicted in , where the entangling lattice is an 1D ring with periodic boundary condition, and the entangling gate is the controlled-Z (CZ) gate. The local unitaries are $\{R_Y, R_Z\}$.

B. Statistical learning theory: VC theory

We use the definition that the generalization error is $E_{out}(h) - E_{in}(h)$ for a hypothesis $h \in H$, where $E_{out}(h) = \mathbb{E}_{(\vec{x},y)\sim P(\vec{x},y)}[\mathbb{1}[h(\vec{x}) \neq y]]$ is the out-of-sample error (prediction error) and $E_{in}(h) = \frac{1}{N}\sum_{i=1}^{N} \mathbb{1}[h(\vec{x}_i) \neq y_i]$ is the in-sample-error (training error). $\mathbb{E}$ is the expectation taken over the unknown input distribution. $\mathbb{1}$ is the indicator function. The VC generalization error bound is [20], [43]

$$E_{out}(h) - E_{in}(h) \leq \sqrt{\frac{8}{N}\ln(\frac{4m_H(2N)}{\delta})}, \tag{2}$$

with probability $\geq 1-\delta$. N is the sample size. δ is the confidence interval. The quantity $m_H(N)$ is a function which could be bounded by $m_H(N) \leq N^{d_{VC}} + 1$ for finite VC-dimension d_{VC}. VC dimension is the maximum number of points that can be shattered by the hypothesis set H. In general, d_{VC} could be infinite for an uncountable hypothesis set. If d_{VC} is finite, then the generalization ability is guaranteed by the VC bound. There are several advantages of VC theory [21]: (1) VC bound is independent of the input distribution. (2) VC bound is non-asymptotic, and it can be applied when the size of training data set is small. (3) VC dimension is a property of the hypothesis set, and is independent of individual hypothesis and independent of the algorithm being used.

C. PAC-Bayesian analysis

PAC-Bayesian approach was proposed by McAllester [26]–[28]. We use a simple version of PAC-Bayesian bound with only prior probability distribution [26]

$$E_{out}(U) - E_{in}(U) \tag{3}$$
$$\leq \sqrt{\frac{1}{2N}[\ln(\frac{1}{P(U)} + \ln(\frac{1}{\delta})) + 2\ln(N)]} + \frac{1}{N}. \tag{4}$$

$U \subseteq H$ is a subset of the hypothesis set H. $E_{out}(U) = \mathbb{E}_{h\in U}E_{out}(h)$ and $E_{in}(U) = \mathbb{E}_{h\in U}E_{in}(h)$ are expectation value of E_{out} and E_{in} for random hypothesis drawn from U according to the prior distribution. This bound depends on the subset U and the prior probability. The full version of PAC-Bayesian bound also depends on a posterior probability distribution over the hypothesis set [28], but we use this version for its simplicity.

D. Tensor networks: algorithmic light cone

Due to the locality and unitarity of 2-qubit entangling gate used in HEA, the light cone limitation in the tensor network can be applied [44]. The tensor contraction outside of the light cone gives identities, hence only the qubits covered by the light cone contribute to the hypothesis set. This limitation leads to the limitation to the degree of trignometric polynomials and hence limits the VC dimension [30].

III. Results and Discussions

To use the PAC-Bayesian bound, consider the subset U_ϵ which is ϵ close to an arbitrary hypothesis h with some distance measure. The setting is depicted in Fig. 1. The

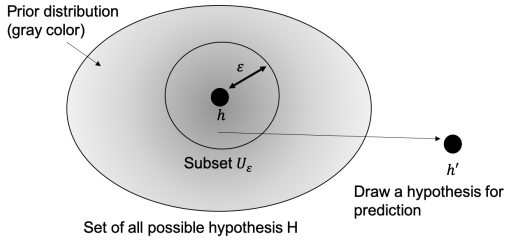

Fig. 1. PAC-Bayesian setting.

distance measure that we use is the quantum state fidelity. The probability density over fidelity $F = |\langle\psi|\phi\rangle|^2$ for random states $|\psi\rangle$ and $|\phi\rangle$ drawn from n-qubit Haar distribution [45] is given by

$$p(F) = (2^n - 1)F^{2^n-2}, \tag{5}$$

where n is the number of qubits (circuit width). We calculate the cumulative distribution

$$P(U_\epsilon) = P_{Harr}(\epsilon) = \mathbb{P}[F > (1-\epsilon)] = \epsilon^{2^n-1}. \tag{6}$$

See Fig. 2 for illustration. This provides the PAC-Bayesian model complexity

$$\ln(\frac{1}{P(U_\epsilon)}) = (2^n - 1)\ln(\frac{1}{\epsilon}) \tag{7}$$

To make the comparisons between VC bound [30] and PAC-Bayesian bound more readable, we make following two simplification: (1). Assuming $d = n$, which means the input feature space dimension is equal to the circuit width. (2). We assume that 1D HEA PBC is used, and a single qubit output measurement result $\langle Z_0 \rangle$ is used for prediction. With these two simplifications, the effective circuit width covered by the light cone is given by $n' = \min\{n, 2L + 1\}$. The PAC-Bayesian complexity is then upper bounded by

$$\ln(\frac{1}{P(U_\epsilon)}) \leq (2^{n'} - 1) \ln(\frac{1}{\epsilon}). \tag{8}$$

The VC dimension is upper bounded by

$$d_{VC} \leq 3^{2n'}. \tag{9}$$

Both VC and PAC-Bayesian complexity scales exponentially with respect to n'. To ease the computation of VC bound, we use the approximation $\ln((2N)^{d_{VC}} + 1) \approx d_{VC} \ln(2N)$. The comparisons of VC bound vs PAC-Bayesian bound scaling with respect to sample size, circuit depth, and circuit width, are plotted in Fig. 3. The parameters used are $\delta = 0.3$ and $\epsilon = 0.01$.

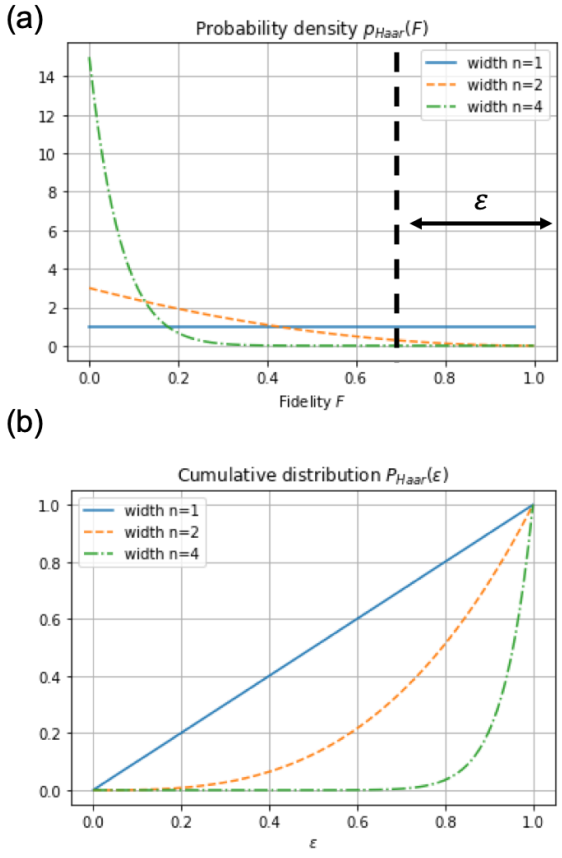

Fig. 2. Haar distribution. n is the number of qubits (circuit width) (a). Probability density of fidelity. (b). Cumulative distribution of fidelity.

The scaling behavior of PAC-Bayesian bound is similar to the VC bound as expected. The PAC-Bayesian bound is generally better, although it depends on the arbitrary parameter ϵ.

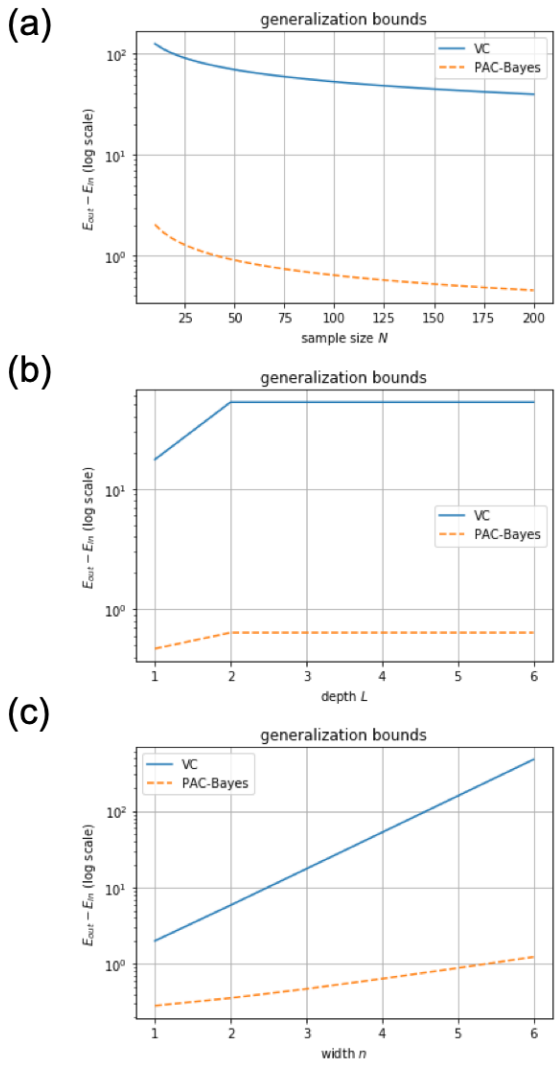

Fig. 3. Comparing the VC bound to the PAC-Bayes bound. (a). Sample size scaling. $L = 4$. $N = 100$. (b). Circuit depth scaling. $n = 4$. $L = 4$. (c). Circuit width scaling. $n = 4$. $N = 100$.

However, this extra parameter could be useful for QCL, since the quantum circuit hypothesis set can only be evaluated by stochastic sampling from a quantum processor. The parameter ϵ could serve as an estimation for sampling error (due to finite number of shots) and hardware noise for quantum hardware. How to incorporate the sampling error and hardware noise into the PAC-Bayesian bound remains unknown to us.

IV. Conclusion

In this work, we provide a short discussion of various generalization error bounds for variational quantum circuit learning. We further provide a bound by using PAC-Bayesian analysis, and the performance is compared to previous result of VC dimension for simple cases. We see that the PAC-Bayesian bound has similar behavior and has better generalization performance. However, the PAC-Bayesian bound depends on an arbitrary parameter ϵ. The arbitrariness can not be

avoided in PAC-Bayesian, since the bound depends on the prior and posterior distributions over hypothesis set. Further investigation is required to understand the applicability of the PAC-Bayesian bound in the QCL setting.

REFERENCES

[1] C. M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Berlin, Heidelberg: Springer-Verlag, 2006.

[2] J. Preskill, "Quantum Computing in the NISQ era and beyond," *Quantum*, vol. 2, p. 79, Aug. 2018. [Online]. Available: https://doi.org/10.22331/q-2018-08-06-79

[3] 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, B. Burkett, Y. Chen, Z. Chen, B. Chiaro, R. Collins, W. Courtney, A. Dunsworth, E. Farhi, B. Foxen, A. Fowler, C. Gidney, M. Giustina, R. Graff, K. Guerin, S. Habegger, M. P. Harrigan, M. J. Hartmann, A. Ho, M. Hoffmann, T. Huang, T. S. Humble, S. V. Isakov, E. Jeffrey, Z. Jiang, D. Kafri, K. Kechedzhi, J. Kelly, P. V. Klimov, S. Knysh, A. Korotkov, F. Kostritsa, D. Landhuis, M. Lindmark, E. Lucero, D. Lyakh, S. Mandrà, J. R. McClean, M. McEwen, A. Megrant, X. Mi, K. Michielsen, M. Mohseni, J. Mutus, O. Naaman, M. Neeley, C. Neill, M. Y. Niu, E. Ostby, A. Petukhov, J. C. Platt, C. Quintana, E. G. Rieffel, P. Roushan, N. C. Rubin, D. Sank, K. J. Satzinger, V. Smelyanskiy, K. J. Sung, M. D. Trevithick, A. Vainsencher, B. Villalonga, T. White, Z. J. Yao, P. Yeh, A. Zalcman, H. Neven, and J. M. Martinis, "Quantum supremacy using a programmable superconducting processor," *Nature*, vol. 574, no. 7779, pp. 505–510, Oct. 2019.

[4] 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," *Nature Communications*, vol. 5, p. 4213, Jul. 2014.

[5] J. R. McClean, J. Romero, R. Babbush, and A. Aspuru-Guzik, "The theory of variational hybrid quantum-classical algorithms," *New Journal of Physics*, vol. 18, no. 2, p. 023023, feb 2016. [Online]. Available: https://doi.org/10.1088/1367-2630/18/2/023023

[6] K. Mitarai, M. Negoro, M. Kitagawa, and K. Fujii, "Quantum circuit learning," *Phys. Rev. A*, vol. 98, p. 032309, Sep 2018. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevA.98.032309

[7] V. Havlíček, A. D. Córcoles, K. Temme, A. W. Harrow, A. Kandala, J. M. Chow, and J. M. Gambetta, "Supervised learning with quantum-enhanced feature spaces," *Nature*, vol. 567, no. 7747, pp. 209–212, Mar. 2019.

[8] M. Schuld, V. Bergholm, C. Gogolin, J. Izaac, and N. Killoran, "Evaluating analytic gradients on quantum hardware," *Phys. Rev. A*, vol. 99, p. 032331, Mar 2019. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevA.99.032331

[9] E. Farhi and H. Neven, "Classification with Quantum Neural Networks on Near Term Processors," *arXiv e-prints*, p. arXiv:1802.06002, Feb. 2018.

[10] M. Cerezo, A. Arrasmith, R. Babbush, S. C. Benjamin, S. Endo, K. Fujii, J. R. McClean, K. Mitarai, X. Yuan, L. Cincio, and et al., "Variational quantum algorithms," *Nature Reviews Physics*, vol. 3, no. 9, p. 625–644, Aug 2021. [Online]. Available: http://dx.doi.org/10.1038/s42254-021-00348-9

[11] S. Sim, P. D. Johnson, and A. Aspuru-Guzik, "Expressibility and entangling capability of parameterized quantum circuits for hybrid quantum-classical algorithms," *Advanced Quantum Technologies*, vol. 2, no. 12, p. 1900070, 2019. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/qute.201900070

[12] K. Nakaji and N. Yamamoto, "Expressibility of the alternating layered ansatz for quantum computation," *Quantum*, vol. 5, p. 434, Apr. 2021. [Online]. Available: https://doi.org/10.22331/q-2021-04-19-434

[13] M. Schuld, R. Sweke, and J. J. Meyer, "Effect of data encoding on the expressive power of variational quantum-machine-learning models," *Phys. Rev. A*, vol. 103, p. 032430, Mar 2021. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevA.103.032430

[14] T. Hubregtsen, J. Pichlmeier, P. Stecher, and K. Bertels, "Evaluation of Parameterized Quantum Circuits: on the relation between classification accuracy, expressibility and entangling capability," *arXiv e-prints*, p. arXiv:2003.09887, Mar. 2020.

[15] T. Goto, Q. Hoan Tran, and K. Nakajima, "Universal Approximation Property of Quantum Feature Map," *arXiv e-prints*, p. arXiv:2009.00298, Sep. 2020.

[16] L. G. Valiant, "A theory of the learnable," *Commun. ACM*, vol. 27, no. 11, p. 1134–1142, Nov. 1984. [Online]. Available: https://doi.org/10.1145/1968.1972

[17] M. Mohri, A. Rostamizadeh, and A. Talwalkar, *Foundations of Machine Learning*. The MIT Press, 2018.

[18] S. Shalev-Shwartz and S. Ben-David, *Understanding Machine Learning: From Theory to Algorithms*. USA: Cambridge University Press, 2014.

[19] M. J. Kearns and U. V. Vazirani, *An Introduction to Computational Learning Theory*. Cambridge, MA, USA: MIT Press, 1994.

[20] V. N. Vapnik and A. Y. Chervonenkis, "On the uniform convergence of relative frequencies of events to their probabilities," *Theory of Probability & Its Applications*, vol. 16, no. 2, pp. 264–280, 1971. [Online]. Available: https://doi.org/10.1137/1116025

[21] V. N. Vapnik, *The Nature of Statistical Learning Theory*. Springer-Verlag New York, Inc., 1999.

[22] A. Blumer, A. Ehrenfeucht, D. Haussler, and M. K. Warmuth, "Learnability and the vapnik-chervonenkis dimension," *J. ACM*, vol. 36, no. 4, p. 929–965, Oct. 1989. [Online]. Available: https://doi.org/10.1145/76359.76371

[23] D. Pollard, *Convergence of Stochastic Processes*. Springer, 1984.

[24] M. J. Kearns and R. E. Schapire, "Efficient distribution-free learning of probabilistic concepts," *Journal of Computer and System Sciences*, vol. 48, no. 3, pp. 464–497, 1994. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0022000005800625

[25] P. L. Bartlett, P. M. Long, and R. C. Williamson, "Fat-shattering and the learnability of real-valued functions," *Journal of Computer and System Sciences*, vol. 52, no. 3, pp. 434–452, 1996. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0022000096900331

[26] D. A. McAllester, "Some pac-bayesian theorems," in *Proceedings of the Eleventh Annual Conference on Computational Learning Theory*, ser. COLT' 98. New York, NY, USA: Association for Computing Machinery, 1998, p. 230–234. [Online]. Available: https://doi.org/10.1145/279943.279989

[27] ——, "Some pac-bayesian theorems," *Mach. Learn.*, vol. 37, no. 3, p. 355–363, Dec. 1999. [Online]. Available: https://doi.org/10.1023/A:1007618624809

[28] ——, "Pac-bayesian stochastic model selection," *Mach. Learn.*, vol. 51, no. 1, p. 5–21, Apr. 2003. [Online]. Available: https://doi.org/10.1023/A:1021840411064

[29] Y. Jiang, B. Neyshabur, H. Mobahi, D. Krishnan, and S. Bengio, "Fantastic Generalization Measures and Where to Find Them," *arXiv e-prints*, p. arXiv:1912.02178, Dec. 2019.

[30] C.-C. Chen, M. Watabe, K. Shiba, M. Sogabe, K. Sakamoto, and T. Sogabe, "On the expressibility and overfitting of quantum circuit learning," *ACM Transactions on Quantum Computing*, vol. 2, no. 2, Jul. 2021. [Online]. Available: https://doi.org/10.1145/3466797

[31] M. Watabe, K. Shiba, C.-C. Chen, M. Sogabe, K. Sakamoto, and T. Sogabe, "Quantum circuit learning with error backpropagation algorithm and experimental implementation," *Quantum Reports*, vol. 3, no. 2, pp. 333–349, 2021. [Online]. Available: https://www.mdpi.com/2624-960X/3/2/21

[32] C.-C. Chen, M. Watabe, K. Shiba, M. Sogabe, K. Sakamoto, and T. Sogabe, "A study on the expressibility and learnability of quantum circuit learning," Tech. Rep. 5, jun 2021. [Online]. Available: http://id.nii.ac.jp/1001/00211674/

[33] S. Aaronson, "The learnability of quantum states," *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 463, no. 2088, pp. 3089–3114, 2007. [Online]. Available: https://royalsocietypublishing.org/doi/abs/10.1098/rspa.2007.0113

[34] M. C. Caro and I. Datta, "Pseudo-dimension of quantum circuits," *Quantum Machine Intelligence*, vol. 2, no. 2, Nov 2020. [Online]. Available: http://dx.doi.org/10.1007/s42484-020-00027-5

[35] S. Arunachalam and R. de Wolf, "Optimal quantum sample complexity of learning algorithms," *Journal of Machine Learning Research*, vol. 19, no. 71, pp. 1–36, 2018. [Online]. Available: http://jmlr.org/papers/v19/18-195.html

[36] A. Abbas, D. Sutter, C. Zoufal, A. Lucchi, A. Figalli, and S. Woerner, "The power of quantum neural networks," *arXiv e-prints*, p. arXiv:2011.00027, Oct. 2020.

[37] H.-Y. Huang, M. Broughton, M. Mohseni, R. Babbush, S. Boixo, H. Neven, and J. R. McClean, "Power of data in quantum machine learning," *Nature Communications*, vol. 12, p. 2631, Jan. 2021.

[38] M. C. Caro, E. Gil-Fuster, J. J. Meyer, J. Eisert, and R. Sweke, "Encoding-dependent generalization bounds for parametrized quantum circuits," *arXiv e-prints*, p. arXiv:2106.03880, Jun. 2021.

[39] Y. Du, Z. Tu, X. Yuan, and D. Tao, "An efficient measure for the expressivity of variational quantum algorithms," *arXiv e-prints*, p. arXiv:2104.09961, Apr. 2021.

[40] T. Barthel and J. Lu, "Fundamental limitations for measurements in quantum many-body systems," *Phys. Rev. Lett.*, vol. 121, p. 080406, Aug 2018. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevLett.121.080406

[41] K. Bu, D. Enshan Koh, L. Li, Q. Luo, and Y. Zhang, "On the statistical complexity of quantum circuits," *arXiv e-prints*, p. arXiv:2101.06154, Jan. 2021.

[42] C. Gyurik, D. van Vreumingen, and V. Dunjko, "Structural risk minimization for quantum linear classifiers," *arXiv e-prints*, p. arXiv:2105.05566, May 2021.

[43] Y. S. Abu-Mostafa, M. Magdon-Ismail, and H.-T. Lin, *Learning From Data*. AMLBook, 2012.

[44] U. Schollwöck, "The density-matrix renormalization group in the age of matrix product states," *Annals of Physics*, vol. 326, no. 1, pp. 96–192, 2011. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0003491610001752

[45] K. Życzkowski and H.-J. Sommers, "Average fidelity between random quantum states," *Phys. Rev. A*, vol. 71, p. 032313, Mar 2005. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevA.71.032313