

Navigating the trainability–expressibility trade-off in quantum machine learning

Peter Röseler^{*1,2}, Dennis Willsch^{1,3}, and Kristel Michielsen^{1,4,2}

¹Jülich Supercomputing Centre, Forschungszentrum Jülich, 52425 Jülich, Germany

²Universität zu Köln, 50923 Köln, Germany

³FH Aachen University of Applied Sciences, 52066 Aachen, Germany

⁴AIDAS, 52425 Jülich, Germany

1 Motivation

For variational quantum machine learning algorithms, expressivity alone is not sufficient. In such algorithms, the choice of the parameterized quantum circuit (PQC), often called the ansatz, is a primary performance bottleneck. An ansatz that is too restricted may not be able to exploit quantum features (low expressibility), while an overly flexible circuit can induce exponentially vanishing gradient landscapes (low trainability), which we refer to as barren plateaus [1].

An important open question is whether there exist both expressible and trainable quantum circuits [2]. Recent works suggest that highly expressive families can be prone to barren plateaus [2], while circuit classes that avoid or soften barren plateaus may do so by exploiting structures that correlate with efficient classical simulation [3–5].

While expressibility admits quantitative metrics [6, 7], quantifying trainability in an informative and broadly applicable way remains less settled [8–10]. Barren plateau theory provides diagnostic and mitigation tools, however, many of these approaches rely on specific circuit structure, initial-state requirements, or problem restrictions, and can be costly to apply at scale [11]. This makes systematic design of circuits that are both expressive and trainable difficult.

We introduce a *finite-sample, dimension-free* concentration bound for estimating gradient variance for arbitrary PQCs, initialization distributions, and cost functions. By using this as a selection criterion, trainability becomes an inductive bias in ansatz search. Our framework combines trainability and expressibility objectives to identify candidate circuits with more favorable expressibility–trainability trade-offs than commonly used ansätze. We further apply the framework to quantum neural networks (QNNs), where the discovered circuits achieve improved effective dimension at reduced parameter count. Further details are provided in Ref. [11].

2 Results

Highly expressible circuits often suffer from barren plateaus [2], rendering optimization ineffective despite good coverage. We want to diagnose whether a given parameterized circuit $U(\boldsymbol{\theta})$ acting on a fiducial state $\rho(\boldsymbol{\theta}) := U(\boldsymbol{\theta})\rho U^\dagger(\boldsymbol{\theta})$ and cost function $C(\boldsymbol{\theta}) := \text{Tr}[\rho(\boldsymbol{\theta})O]$ (where O is a Hermitian observable) is prone to exhibit barren plateaus. Following [8, 10, 12, 13] we define the barren plateau as

$$\text{Var}_{\boldsymbol{\theta}}[\partial_{\mu}C(\boldsymbol{\theta})] \leq F(n), \quad F(n) \in \mathcal{O}\left(\frac{1}{c^n}\right), \quad c > 1. \quad (1)$$

^{*}Corresponding author: p.roeseler@fz-juelich.de

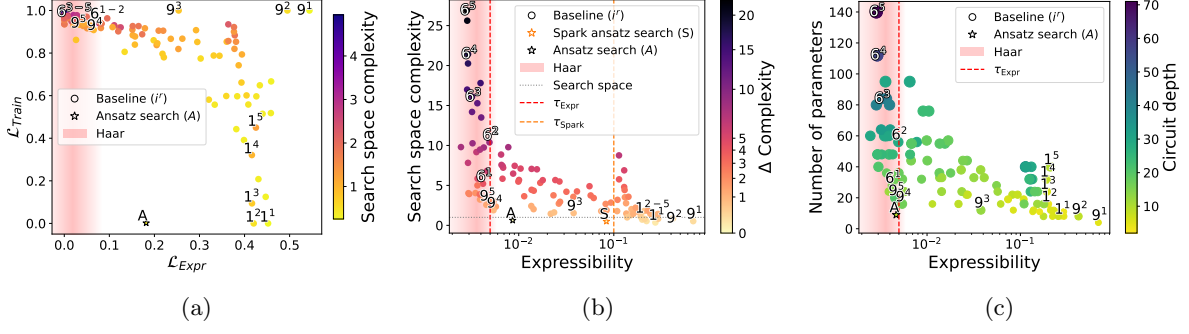


Figure 1: Metric-guided ansatz optimization. (a) Trade-off between expressibility loss $\mathcal{L}_{\text{Expr}}$ and trainability loss $\mathcal{L}_{\text{Train}}$. (b) Hardware-constrained expressibility search on the 5-qubit IQM Spark device with native gates and star connectivity; marker color encodes the Δ -complexity, i.e., the fold increase in the allowed architecture search space, and the dashed vertical lines indicate the expressibility thresholds τ_{Expr} and τ_{Spark} . (c) Expressibility versus number of parameters; marker size is proportional to the number of gates. Markers denote benchmark ansätze i^r from Ref. [6], searched circuits (A), and the hardware-constrained IQM Spark ansatz (S), with r the number of repetitions. The shaded *Haar* band shows a $\pm 3\sigma$ interval from 5,000 Haar-random state pairs.

However, since this definition is asymptotic in the number of qubits n , it is primarily suited to statements about scaling. We therefore adopt an *operational* notion of barren plateaus: for fixed n we declare a circuit effectively non-trainable if the typical gradient fluctuations fall below a threshold $\tau_{\text{BP}} \in \mathbb{R}^+$, i.e., $\text{Var}_{\boldsymbol{\theta}}[\partial_{\mu} C(\boldsymbol{\theta})] \leq \tau_{\text{BP}}$. This raises two practical questions: (i) how to set τ_{BP} , and (ii) how to estimate $\text{Var}_{\boldsymbol{\theta}}[\partial_{\mu} C(\boldsymbol{\theta})]$ from finitely many samples.

To address (i), we assume that the algorithm employing the PQC is executed on noisy intermediate-scale quantum (NISQ) devices. If the gradient fluctuations are smaller than the estimation noise floor $\text{Pr}_{\text{PQC}}(\text{err})$, optimization can become ineffective because updates are dominated by noise rather than signal [14, 15]. Motivated by signal-to-noise analyses of gradient-based learning in stochastic optimization [16, 17], we therefore deem a PQC trainable if its gradient-to-noise ratio exceeds the threshold τ_{BP} :

$$\frac{\text{Var}_{\boldsymbol{\theta}}[\partial_{\mu} C(\boldsymbol{\theta})]}{\text{Pr}_{\text{PQC}}(\text{err})} > \tau_{\text{BP}}. \quad (2)$$

We emphasize that noise does not necessarily imply poor optimization [18, 19].

To address (ii), we estimate gradient fluctuations from finitely many samples using a concentration argument based on the empirical Bernstein bounds of Maurer and Pontil [20, Thm. 7]. Since this requires uniformly bounded samples, we clip the sampled gradients, a standard technique in machine learning [21–25]. The resulting bound (Theorem 2.1) controls the deviation between the empirical and true standard deviation of the gradient distribution, while Lemma 1 of Ref. [11] relates this certificate to variance control in the small-variance regime relevant for barren plateaus.

Theorem 2.1 *Let $X_1, \dots, X_m$ be i.i.d. bounded gradient samples with $X_i \in [L, U]$, $L < U$, and $R := U - L$. For $m \geq 3$, let s_m^2 denote their unbiased sample variance. Then, for all $\delta > 0$:*

$$\Pr \left(\left| \sqrt{s_m^2} - \sqrt{\mathbb{E}[s_m^2]} \right| \leq \sqrt{\frac{2R^2 \ln(2/\delta)}{m-1}} \right) \geq 1 - \delta. \quad (3)$$

We combine this trainability objective with a modified expressibility metric based on Ref. [6] and use Bayesian optimization as a sample-efficient black-box search over circuit architectures [11]. The search identifies ansätze in an intermediate regime of the expressibility–trainability

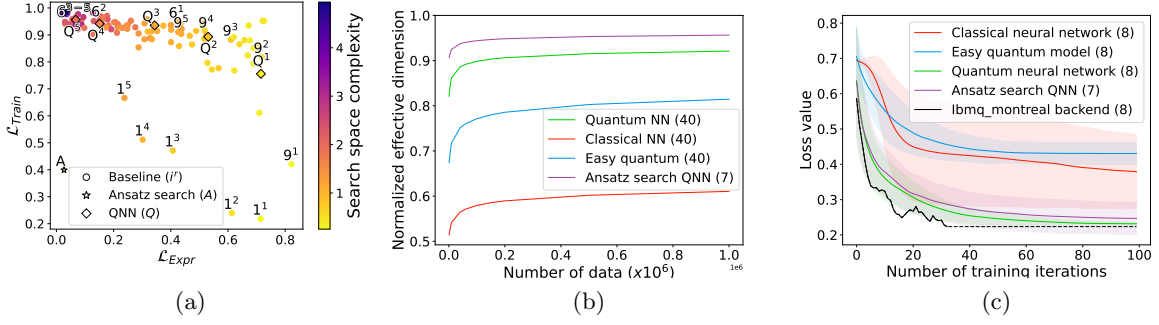


Figure 2: QNN case study following the setup of [27]. (a) Search outcomes in the expressibility–trainability plane for the considered search space. (b) Estimated effective dimension for the reference QNN architecture from [27] and for the discovered PQC. (c) Training performance on the dataset of [27], using the same optimizer and protocol, comparing the reference architecture to our discovered PQC.

plane, where they satisfy moderate thresholds for both metrics (Fig. 1a). For commonly used circuits [6], the region of simultaneously high expressibility and high trainability appears largely inaccessible, consistent with theoretical links between expressibility and gradient concentration [1, 2].

To test whether the same search principle remains useful under hardware constraints, we next consider the simpler but important objective of finding compact circuits with high expressibility. On the 5-qubit IQM Spark device [26], we restrict the search to the native gate set and star connectivity, and identify more expressive hardware-compatible circuits within the considered search space (Fig. 1b). When these hardware constraints are removed, the same procedure finds circuits with expressibility comparable to the best benchmark ansätze of Ref. [6], while using fewer parameters and lower circuit complexity (Fig. 1c).

Beyond these architecture-level benchmarks, we ask whether the searched circuits also lead to improvements in a quantum-machine-learning task. We therefore apply the search to the QNN protocol of Ref. [27]. As in the metric-level study, the discovered ansatz lies in a sparsely populated region of the expressibility–trainability plane (Fig. 2a). Despite its low parameter count, it achieves an improved effective dimension compared with the reference architecture (Fig. 2b). This increased functional capacity, however, does not translate into improved classification performance in the considered training experiment (Fig. 2c), possibly because the benchmark task is too simple for the additional capacity to yield a measurable advantage.

3 Discussion

Our results suggest that trainability can be used as an explicit design criterion for PQCs. By combining a dimension-free trainability certificate with an expressibility objective, the framework identifies ansätze with favorable metric-level properties, including circuits that remain expressive while avoiding strongly suppressed gradient fluctuations. This does not replace analytical circuit design, but provides a complementary tool for exploring architecture space and generating candidate circuits whose structural features can later be studied analytically.

Several limitations remain. The scaling of the discovered QNN ansatz and the structural origin of its favorable properties remain open questions. In this work, we did not explicitly optimize for indicators of classical hardness, but instead relied on the observed relation between expressibility and entanglement [11]. Future work could incorporate such indicators more directly (e.g., magic [28]) and include further desirable properties in the ansatz-design objective. More generally, the same framework may be useful for other variational quantum algorithms [11].

References

- [1] Jarrod R. McClean, Sergio Boixo, Vadim N. Smelyanskiy, Ryan Babbush, and Hartmut Neven. Barren plateaus in quantum neural network training landscapes. *Nature Communications*, 9:4812, 2018. doi: 10.1038/s41467-018-07090-4.
- [2] Zoë Holmes, Kunal Sharma, M. Cerezo, and Patrick J. Coles. Connecting ansatz expressibility to gradient magnitudes and barren plateaus. *PRX Quantum*, 3:010313, Jan 2022. doi: 10.1103/PRXQuantum.3.010313.
- [3] M. Cerezo, M. Larocca, D. García-Martín, et al. Does provable absence of barren plateaus imply classical simulability? *Nature Communications*, 16:7907, 2025. doi: 10.1038/s41467-025-63099-6.
- [4] Lorenzo Leone, Salvatore F.E. Oliviero, Lukasz Cincio, and M. Cerezo. On the practical usefulness of the Hardware Efficient Ansatz. *Quantum*, 8:1395, July 2024. ISSN 2521-327X. doi: 10.22331/q-2024-07-03-1395.
- [5] Elies Gil-Fuster, Casper Gyurik, Adrián Pérez-Salinas, and Vedran Dunjko. On the relation between trainability and dequantization of variational quantum learning models. *arXiv preprint arXiv:2406.07072*, 2025.
- [6] Sukin Sim, Peter D. Johnson, and Alán Aspuru-Guzik. Expressibility and entangling capability of parameterized quantum circuits for hybrid quantum-classical algorithms. *Adv. Quantum Technol.*, 2(12):1900070, 2019. doi: 10.1002/qute.201900070.
- [7] Yuxuan Du, Zhuozhuo Tu, Xiao Yuan, and Dacheng Tao. Efficient measure for the expressivity of variational quantum algorithms. *Physical Review Letters*, 128(8):080506, 2022. doi: 10.1103/PhysRevLett.128.080506.
- [8] John Cunningham and Jiajun Zhuang. Investigating and mitigating barren plateaus in variational quantum circuits: a survey. *Quantum Information Processing*, 24(1):48, 2025. doi: 10.1007/s11128-025-04665-1.
- [9] Martin Larocca, Supanut Thanasilp, Samson Wang, M. Cerezo, and Patrick J. Coles. Barren plateaus in variational quantum computing. *Nature Reviews Physics*, 7:174–189, 2025. doi: 10.1038/s42254-025-00813-9.
- [10] Haonan Qi, Lei Wang, Hongbin Zhu, et al. The barren plateaus of quantum neural networks: review, taxonomy and trends. *Quantum Information Processing*, 22(12):435, 2023. doi: 10.1007/s11128-023-04188-7.
- [11] Peter Röseler, Dennis Willsch, and Kristel Michielsen. How to find expressible and trainable parameterized quantum circuits? *arXiv preprint arXiv:2603.14451*, 2026.
- [12] M Cerezo and Patrick J Coles. Higher order derivatives of quantum neural networks with barren plateaus. *Quantum Science and Technology*, 6(3):035006, jun 2021. doi: 10.1088/2058-9565/abf51a.
- [13] Martin Larocca, Piotr Czarnik, Kunal Sharma, Gopikrishnan Muraleedharan, Patrick J. Coles, and M. Cerezo. Diagnosing Barren Plateaus with Tools from Quantum Optimal Control. *Quantum*, 6:824, September 2022. ISSN 2521-327X. doi: 10.22331/q-2022-09-29-824.
- [14] Samson Wang, Enrico Fontana, M. Cerezo, Kunal Sharma, Akihiko Sone, Lukasz Cincio, and Patrick J. Coles. Noise-induced barren plateaus in variational quantum algorithms. *Nature Communications*, 12:6961, 2021. doi: 10.1038/s41467-021-27045-6.

- [15] M. Schumann, F. K. Wilhelm, and A. Ciani. Emergence of noise-induced barren plateaus in arbitrary layered noise models. *Quantum Science and Technology*, 9(4):045019, 2024. doi: 10.1088/2058-9565/ad6285.
- [16] John Roberts and Russ Tedrake. Signal-to-noise ratio analysis of policy gradient algorithms. In D. Koller, D. Schuurmans, Y. Bengio, and L. Bottou, editors, *Advances in Neural Information Processing Systems*, volume 21. Curran Associates, Inc., 2008. URL https://proceedings.neurips.cc/paper_files/paper/2008/file/8df707a948fac1b4a0f97aa554886ec8-Paper.pdf.
- [17] Sam McCandlish, Jared Kaplan, Dario Amodei, and OpenAI Dota Team. An empirical model of large-batch training. *arXiv preprint arXiv:1812.06162*, 2018.
- [18] Samuel Smith, Erich Elsen, and Soham De. On the generalization benefit of noise in stochastic gradient descent. In Hal Daumé III and Aarti Singh, editors, *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pages 9058–9067. PMLR, 13–18 Jul 2020. URL <https://proceedings.mlr.press/v119/smith20a.html>.
- [19] Junyu Liu, Frederik Wilde, Antonio Anna Mele, Xin Jin, Liang Jiang, and Jens Eisert. Stochastic noise can be helpful for variational quantum algorithms. *Phys. Rev. A*, 111: 052441, May 2025. doi: 10.1103/PhysRevA.111.052441.
- [20] Andreas Maurer and Massimiliano Pontil. Empirical bernstein bounds and sample variance penalization. *arXiv preprint arXiv:0907.3740*, 2009.
- [21] Razvan Pascanu, Tomas Mikolov, and Yoshua Bengio. On the difficulty of training recurrent neural networks. In Sanjoy Dasgupta and David McAllester, editors, *Proceedings of the 30th International Conference on Machine Learning*, volume 28 of *Proceedings of Machine Learning Research*, pages 1310–1318, Atlanta, Georgia, USA, June 2013. PMLR. URL <https://proceedings.mlr.press/v28/pascanu13.html>.
- [22] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016. <http://www.deeplearningbook.org>.
- [23] Martin Abadi, Andy Chu, Ian Goodfellow, H. Brendan McMahan, Ilya Mironov, Kunal Talwar, and Li Zhang. Deep learning with differential privacy. CCS ’16, page 308–318, New York, NY, USA, 2016. Association for Computing Machinery. ISBN 9781450341394. doi: 10.1145/2976749.2978318.
- [24] Jingzhao Zhang, Tianxing He, Suvrit Sra, and Ali Jadbabaie. Why gradient clipping accelerates training: A theoretical justification for adaptivity. *arXiv preprint arXiv:1905.11881*, 2019.
- [25] Anastasia Koloskova, Hadrien Hendrikx, and Sebastian U Stich. Revisiting gradient clipping: Stochastic bias and tight convergence guarantees. In Andreas Krause, Emma Brunskill, Kyunghyun Cho, Barbara Engelhardt, Sivan Sabato, and Jonathan Scarlett, editors, *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pages 17343–17363. PMLR, 23–29 Jul 2023. URL <https://proceedings.mlr.press/v202/koloskova23a.html>.
- [26] Jani Rönkkö, Olli Ahonen, Ville Bergholm, Andreas Jäger, Michael Orszag, and Mikko Möttönen. On-premises superconducting quantum computer for education and research. *EPJ Quantum Technology*, 11:32, 2024. doi: 10.1140/epjqt/s40507-024-00243-z.

- [27] Amira Abbas, David Sutter, Christa Zoufal, Aurelien Lucchi, Alessio Figalli, and Stefan Woerner. The power of quantum neural networks. *Nature Computational Science*, 1:403–409, June 2021. doi: 10.1038/s43588-021-00084-1.
- [28] Lorenzo Leone, Salvatore F. E. Oliviero, and Alioscia Hamma. Stabilizer rényi entropy. *Phys. Rev. Lett.*, 128:050402, Feb 2022. doi: 10.1103/PhysRevLett.128.050402.