

Benchmarking Quantum Generative Learning: A Study on Scalability and Noise Resilience using QUARK

Florian J. Kiwit,^{1,2,*} Maximilian A. Wolf,^{1,2} Marwa Marso,^{1,2,3} Philipp Ross,¹ Jeanette M. Lorenz,^{2,4} Carlos A. Riofrío,¹ and Andre Luckow^{1,2}

¹*BMW Group, Munich, Germany*

²*Ludwig Maximilian Universität, Munich, Germany*

³*Technical University, Munich, Germany*

⁴*Fraunhofer Institute for Cognitive Systems IKS, Munich, Germany*

Quantum computing promises a disruptive impact on machine learning algorithms, taking advantage of the exponentially large Hilbert space available. However, it is not clear how to scale quantum machine learning (QML) to industrial-level applications. This paper investigates the scalability and noise resilience of quantum generative learning applications. We consider the training performance in the presence of statistical noise due to finite-shot noise statistics and quantum noise due to decoherence to analyze the scalability of QML methods. We employ rigorous benchmarking techniques to track progress and identify challenges in scaling QML algorithms, and show how characterization of QML systems can be accelerated, simplified, and made reproducible when the QUARK framework is used. We show that QGANs are not as affected by the curse of dimensionality as QCBMs and to which extent QCBMs are resilient to noise.

Keywords: Quantum Computing; Machine Learning; Noise Resilience; Generative Modeling, Benchmark Framework

I. INTRODUCTION

Systematic evaluation of quantum processors and algorithms through benchmarking offers valuable insights into the current capabilities and future potential of available quantum processing units [1–9]. However, benchmarking quantum computing is far from a straightforward task. The field is characterized by a diversity of technologies [10], each with unique requirements for precise and meaningful assessment. As a result, current benchmarks often focus on specific aspects of the technology, which can sometimes lead to an incomplete picture of the end-to-end performance of quantum computing.

The Quantum computing Application benchmark (QUARK) framework [8] was explicitly developed for challenges of application-oriented quantum computing. QUARK’s benchmarking approach ensures a comprehensive evaluation, covering the entire benchmarking pipeline from hardware to algorithmic design for the problems under investigation. Its versatility and modular implementation are central to QUARK, allowing for component expansion and customization. Additionally, it hosts benchmarks from the domain of optimization [8] and machine learning [11].

In quantum machine learning (QML), scaling algorithms and maintaining performance amidst noise are crucial for practical applications, particularly in industries reliant on generative models. This work shows an extension that enables us to include noisy simulations for QML applications. It is important to understand the limitations and track the development of current quantum hardware and algorithms over time. Concretely, we

present a comprehensive study of the scalability of QML models, evaluating their intrinsic robustness against statistical and quantum noise. We aim to bridge the gap between theoretical QML advancements and their practical implementation in real-world scenarios while using the QUARK framework to accelerate and standardize performance assessment.

The paper is organized as follows: Section II briefly introduces the basic concepts of quantum generative learning used in this paper. In Section III A, we present a description of the components of the QUARK framework. In Section IV A, we discuss the influence of statistical noise on the scalability of QML models. In Section IV B, we look at the effects of quantum noise and hardware defects on the training of QML models. And finally, in Section V, we summarize our findings.

II. QUANTUM GENERATIVE LEARNING

Generative modeling is a growing area of interest across all industries. Applications include anomaly detection, text & image generation, or speech & video synthesis. Ultimately, the objective of training a generative model is to express the underlying distribution of a dataset by a machine learning model. In QML, this model is represented by a parameterized quantum circuit (PQC) [12]. During the training of a quantum generative model, the probability amplitudes of the quantum state vector generated by a PQC are fitted to the probability distribution of the dataset; see, for example, reference [13]. We will refer to the probability mass function (PMF) of the dataset as p and to that of the state generated by the PQC as q . The absolute square values of the state vector give the PMF of the PQC.

* florian.kiwit@bmw.de

Two popular training routines are the *quantum circuit Born machine* (QCBM) [14] and the *quantum generative adversarial network* (QGAN) [15], see [16] for a detailed review of both methods. The QCBM tries to minimize the Kullback-Leibler (KL) divergence, a well-known statistical distance, between p and q by adapting the model parameters via the covariance matrix adaption evolutionary strategy (CMA-ES) [17], a gradient-free optimizer.

Conversely, the QGAN follows the architecture of a classical GAN [18]. A GAN operates on the principle of adversarial training in a minimax 2-player game, employing two neural networks, the generator and the discriminator. The generator creates synthetic data instances, while the discriminator evaluates these generated samples alongside real ones. The two networks engage in a continual feedback loop, with the generator striving to improve its output and the discriminator refining its ability to differentiate between real and fake samples. Different architectures are proposed for QGANs, but a typical approach is to replace the classical generator with a PQC. The model parameters of the PQC are updated by gradient descent, and the parameter-shift rule determines the gradients, for example, reference [19]. The classical discriminator of the QGAN is optimized with the ADAM [20] optimizer, and the gradients are determined via backpropagation.

III. BENCHMARKING QUANTUM COMPUTING

In its most general form, benchmarking is the process of comparing the performance of systems by a set of measurements. The workloads used are referred to as benchmarks [21], and the metrics are the criteria to compare the performance. In quantum computing, these metrics should characterize scale, quality, and speed [22]. Depending on the scope of attributes a benchmark assesses, they can be attributed to three categories [23]. (1) Physical Benchmarks focus on the basic physical properties of quantum hardware, such as the number of qubits and quantum gates. Examples are T1 and T2 relaxation times, gate fidelity, and readout fidelity; for example, see reference [24] for an extensive review. (2) Aggregated benchmarks assess the performance over a large set of device attributes. Prominent examples are quantum volume [25] and circuit layer operations per second [26]. (3) Application-oriented benchmarks test the performance of quantum computers in real-world scenarios. They simulate specific computational tasks that quantum computers are expected to perform, e.g., optimization, machine learning and chemistry algorithms. These benchmarks are particularly important for demonstrating the practical utility and efficiency of quantum systems in solving complex, real-world problems and are a key indicator of progress toward quantum advantage. Examples include QPack [27], QED-C [6, 7, 28] and Q-Score [29].

A. The QUARK Framework

The QUARK framework [30] orchestrates application-oriented quantum benchmarks in a well-defined, standardized, reproducible and verifiable way. It remains vendor-neutral to ensure unbiased application across different quantum computing platforms. In the following, we will describe the implementation of the generative modeling application in QUARK.

A benchmarking instance is defined by configuring six distinct modules (see Figure 1): In the (1) generative modeling class, global properties, such as the number of qubits n , are defined. Subsequently, a (2) continuous or discrete dataset is selected. The discrete datasets are characterized by a constraint on the bit string of length n , while the continuous datasets include both low-dimensional synthetic and real data. The continuous datasets are passed to a (3) transformation. This ensures the data is in a standard and normalized form. The MinMax transformation maps the marginal distributions to the interval $[0, 1]$. An alternative is the probability integral transformation (PIT), which makes the marginal distributions uniformly distributed. After applying the transformation, the data is mapped to a discrete probability distribution with 2^n bins. Next, the architecture of the PQC is selected in the (4) circuit module and mapped to the (5) Qiskit [31] or PennyLane [32] SDK. The library-agnostic definition of the PQC enables comparative studies of quantum simulators from different vendors. Additionally, many variational ansatzes can be extended as the application requires. For example, the *copula circuit* [33], which naturally respects the properties of data transformed via the PIT, can only learn a probability distribution whose cumulative marginals are uniformly distributed. In the last step, a pre-trained PQC is loaded for (6) inference or a training routine is configured. Training routines include the QCBM and QGAN, discussed in the previous section, but can be extended to other QML models. After defining the benchmarking instance, QUARK orchestrates the execution, data collection and visualization of the benchmark. For a detailed report on the QUARK framework, see References [8, 11].

Furthermore, both noisy and noise-free simulators are available. Depending on the individual requirements, one can configure the backend of the circuit to include multiple sources of errors. This can be done in two ways: (1) By specifying different error sources, like readout and depolarizing errors or amplitude and phase damping and chip-agnostic parameters, e.g., the coupling map, which represents the connectivity of the qubits; the basis gates, i.e., the gates that can be used on the backend. (2) We also provide an implementation of Qiskit FakeBackends, which are a predefined snapshot of the error rates, coupling map and basis gates [34]. Additionally, the latest calibration data of IBM's quantum processors can be accessed using `qiskit_ibm_runtime`. This feature enables users to understand how quantum noise might influence

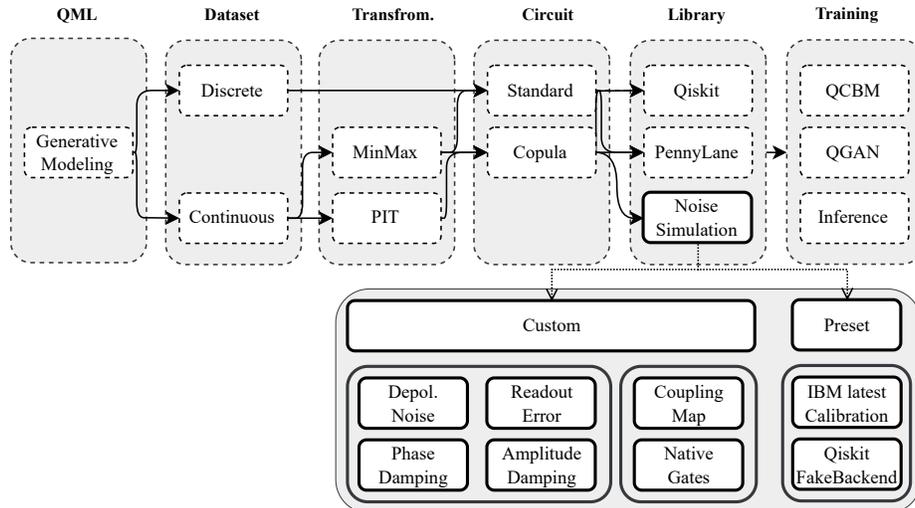


FIG. 1. Illustration of the components of the QUARK framework for quantum generative modeling with a detailed depiction of the noise module. We start at the left by defining the application, in this case, the training of a QML model. Then, the user defines the dataset for training, followed by necessary data transformations, before choosing the PQC ansatz to be used as a model. As explained in the main text, QUARK offers great flexibility regarding quantum libraries for implementing the circuit ansatz and training via QCBM or QGAN methods. At the bottom, we show the structure of the new Noise Simulation module. This module is designed to offer both predefined noise configurations and the flexibility to create entirely custom noise profiles and backend specifications. This allows for more accurate assessments of the robustness and performance of quantum algorithms under various noise conditions. The new modules are indicated by bold solid lines in the figure. More details are given in the main text.

their selected applications and compare them to an ideal environment or a real QPU.

IV. METHODOLOGY AND RESULTS

In this Section, we use the QUARK framework to study the effects of statistical and quantum noise in QML applications. First, we compare the training routines of the QCBM and the QGAN. Next, we study the effects of quantum noise in the training of QCBMs.

A. Scalability of Quantum Generative Models under the Influence of Statistical Noise

The training of the QCBM relies on estimating the full PMF q generated by the PQC. As we scale the number of qubits, n , the number of bins of the PMF scales exponentially. In the following, we will describe how the number of circuit executions, n_{shots} , scales, to keep the statistical error of the PMF, $|\alpha_i|^2$, bounded, where α_i are the probability amplitudes of the quantum state vector generated by a PQC. The probability of frequency n_i in bin i is given by the binomial distribution $B_{n_{\text{shots}}, p_i}(n_i)$, with $\sum n_i = n_{\text{shots}}$ and $p_i = n_i/n_{\text{shots}}$. Furthermore, p_i is the estimator of $|\alpha_i|^2$. On average, if $|\alpha_i|^2$ decays exponentially with the number of qubits n , i.e. $|\alpha_i|^2 \propto 1/n^2$, then the variance is proportional to $\sigma^2(p_i) \propto 2^n/n_{\text{shots}}$. Therefore, the number of circuit executions needs to scale

exponentially with the number of qubits to keep the statistical error on each estimator of $|\alpha_i|^2$ bounded. For the QCBM, the number of circuit executions of one epoch is given by $n_{\text{shots}} \cdot \lambda$. The population size λ is a hyperparameter of the CMA-ES optimizer and refers to the number of model parameters evaluated at any given iteration.

Unlike the QCBM, the number of circuit executions of the QGAN needed to update the model parameters does not increase exponentially with the number of qubits. The total number of circuit executions per epoch is given by $(2 \cdot n_{\text{parameters}} + 1) \cdot n_{\text{samples}}$. One circuit execution with the model parameters is needed to generate synthetic samples to update the discriminator. In the backward pass, two additional circuit executions per model parameter are necessary to determine the gradients with the parameter-shift rule.

Experimental Design: To showcase the different scaling behaviors of QCBMs and QGANs, we track the KL divergence as a function of the cumulative number of circuit executions, as depicted in Figure 2. We fitted the copula circuit with 12 qubits and a depth of one to a dataset resembling the shape of the letter X, using the quantum noise-free Qiskit AerSimulator. For the QCBM, we use a population size of $\lambda = 5$ and train the models with $4 \cdot 10^5$ and $1 \cdot 10^6$ circuit executions to determine the PMF generated by the PQC. For the QGAN, we use a batch size of 20 and alternately update the generator and discriminator on each mini-batch. While training the generative models with a shot-based simulator, we report

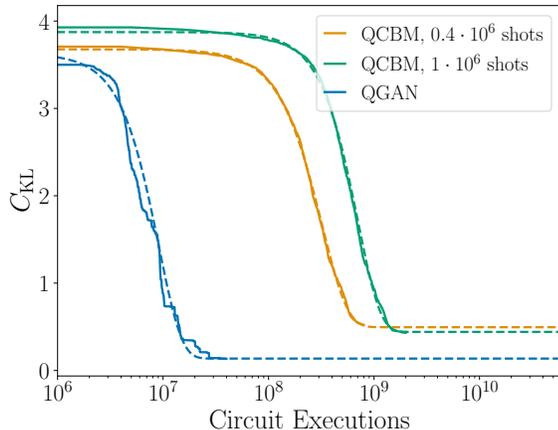


FIG. 2. KL divergence, C_{KL} , as a function of the cumulative number of circuit executions for the QCBM and the QGAN. The dashed lines indicate the stretched exponential function $f(x) = \alpha \cdot \exp(-\beta x^\gamma) + C_{\text{KL}}^{\text{conv}}$ fitted to the loss curves. For the training of the QCBM we show results of experiments, where we used $0.4 \cdot 10^6$ and $1 \cdot 10^6$ circuit executions to determine the PMF generated by the PQC. The QGAN converges with more than an order of magnitude fewer circuit executions than the QCBM to a lower limit. Each model was trained ten times and mean values and standard error on the mean $\sigma/\sqrt{10}$ are depicted as the solid lines and shaded areas, respectively.

the KL divergence between the PMF¹ of the PQC and the target distribution. To compare the performance of the trained models, we fit a stretched exponential function, $f(x) = \alpha \cdot \exp(-\beta x^\gamma) + C_{\text{KL}}^{\text{conv}}$, to the loss curves and report the limit of the KL divergence $C_{\text{KL}}^{\text{conv}}$.

Discussion: For the QCBM, increasing the number of circuit executions to estimate the PMF of the state generated by the PQC from $4 \cdot 10^5$ to $1 \cdot 10^6$ leads to a slight decrease of the KL divergence at convergence from $C_{\text{KL}}^{\text{conv}} = 0.49 \pm 0.13$ to $C_{\text{KL}}^{\text{conv}} = 0.44 \pm 0.12$. The QGAN achieves faster convergence than the QCBM, requiring more than one order of magnitude fewer circuit executions. In addition to fewer circuit executions, the KL divergence of the QGAN converges to a lower limit with a value $C_{\text{KL}}^{\text{conv}} = 0.14 \pm 0.01$. To match the limit of the KL divergence of the QGAN with the QCBM, we would need to increase the circuit executions further, and the separation with respect to the circuit executions would become even more dominant.

Limitations: The QCBM was trained with the gradient-free optimizer CMA-ES. However, a gradient-based training of the QCBM [35] was not investigated and might lead to faster convergence despite needing more circuit evaluations per iteration to estimate the gradi-

ents. Furthermore, we used only one dataset; exploring how the findings generalize to different datasets is an interesting path for future research.

B. Noisy training of QCBMs

In the noisy intermediate-scale quantum (NISQ) era, the performance of QML algorithms is impacted by the presence of quantum noise [38, 39]. This noise stems from various sources, such as decoherence, imperfect gate operations, and environmental interference, and leads to a decreased fidelity of the quantum state generated by a PQC; for a detailed discussion of how noise can influence the training of PQCs, see [40].

Experimental design: We investigate the robustness of QCBMs against quantum noise to characterize the limits of current quantum hardware. To this aim, we train the QCBM and investigate the KL divergence at convergence under varying noise conditions. We fit the copula circuit with 6, 8, 10, and 12 qubits to a dataset that resembles the shape of the letter X. We execute the circuit $1 \cdot 10^4$ times to estimate the PMF. The gate set used for the copula circuit is native to IonQ Harmony. The number of gates of the circuit is reported in Table I. We vary the probability of (a) readout and (b) two-qubit depolarizing errors, as depicted in Figure 3. Readout errors are represented by a bit-flip channel, which means that with a probability p_{10} (p_{01}), the prepared input state $|1\rangle$ ($|0\rangle$) yields the measurement outcome 0 (1). In our experiments, we set $p_{10} = p_{01}$ [39]. Two-qubit depolarizing errors are characterized by the error rate, p_{depol} , that a two-qubit gate creates the fully mixed state instead of the desired output state of the operation.

Discussion: Figure 3a illustrates the influence of readout error on the KL divergence at convergence $C_{\text{KL}}^{\text{conv}}$. With increasing error rate, $C_{\text{KL}}^{\text{conv}}$ increases linearly. Even at an error rate of 0.1, the QCBM maintains robustness against readout errors, as the $C_{\text{KL}}^{\text{conv}}$ is still below the random baseline². Unsurprisingly, the two-qubit depolarizing error shows a stronger effect on $C_{\text{KL}}^{\text{conv}}$, as depicted in Figure 3b for circuits with more gates, as multiple

n_{qubits}	1-Qubit Gates	2-Qubit Gates
6	21	9
8	28	16
10	35	25
12	42	36

TABLE I. The total number of the one-qubit gates (RZ , SX , RX , H) and two-qubit gates (RXX , CX) of the compiled circuit for 6, 8, 10 and 12 qubits.

¹ Here we use the precise PMF, not the estimated PMF, to circumvent the influence of shot-noise when comparing the model performance.

² The random baseline denotes the KL divergence between the training set PMF and the uniform distribution.

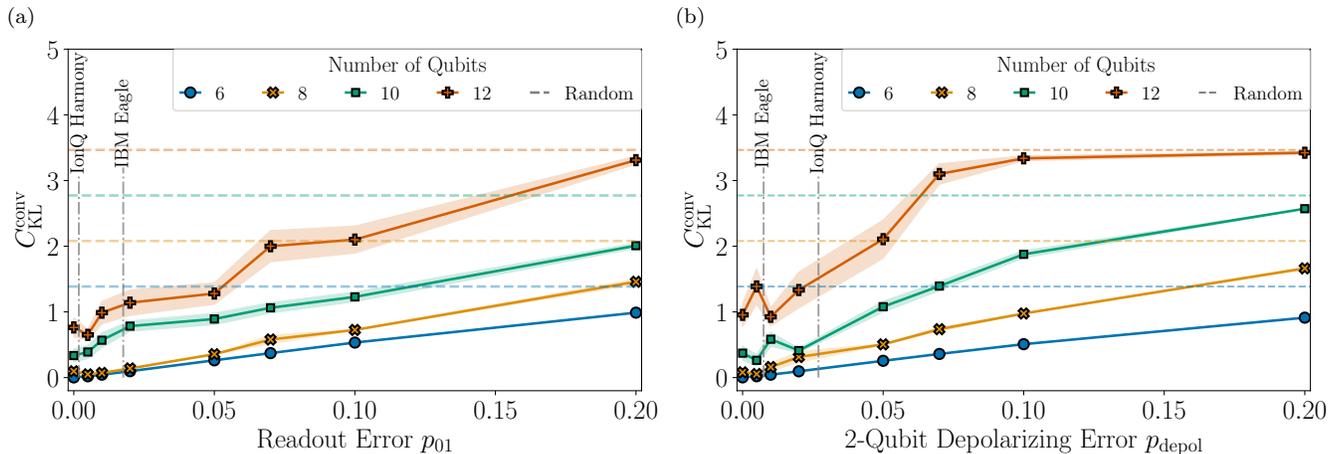


FIG. 3. Mean KL divergence after $4 \cdot 10^4$ circuit evaluations as a function of (a) the readout error and (b) the two-qubit depolarizing error for different circuit widths of a QCBM with the standard error ($\sigma/\sqrt{8}$). The vertical grey lines denote the error value of the IBM Eagle processor (Median ECR error: $7.477 \cdot 10^{-3}$ and readout error: $1 \cdot 10^{-2}$ of IBM Sherbrooke [36]) and IonQ Harmony (SPAM error: $1.8 \cdot 10^{-3}$ and two-qubit gate errors: $2.7 \cdot 10^{-2}$ [37]). The colored horizontal lines denote the KL divergence of the training data and a uniform distribution.

operations are performed per qubit, instead of measuring the state only once. The number of two-qubit gates increases with the circuit width (see Table I), so does the influence of p_{depol} on the performance of the QCBM. The performance is still below the random baseline for 6 qubits and $p_{\text{depol}} = 0.2$. For 12 qubits, however, the model performance corresponds to random guessing already for $p_{\text{depol}} = 0.1$.

Limitations: Our focus on readout and two-qubit depolarizing errors provides a foundation for understanding specific noise sources, but other factors, such as phase errors and crosstalk, are not considered. Our evaluation primarily considers the noise levels of state-of-the-art trapped ion quantum computers. However, the landscape of quantum hardware is rapidly evolving, and the generalizability of our findings to other quantum platforms with potentially different error characteristics needs further exploration. Incorporating mitigation and correction strategies for quantum errors could offer additional insights into enhancing the robustness of QCBMs in noisy quantum environments.

V. CONCLUSION AND FUTURE WORK

In this work, we use and extend the QUARK benchmarking framework and illustrate its functioning with two applications that consider different aspects of noise in QML: statistical and quantum noise. The modular structure of QUARK makes it a versatile tool for a broad spectrum of research applications in QML and quantum computing in general.

Our experiments focused on the performance characterization of quantum generative models. A comparative analysis of QCBMs and QGANs revealed differences in

their efficiencies. Remarkably, QGANs achieved faster convergence with reduced computational demands. Despite the recent advancements in QPU architectures leading to reduced error rates, noise is still a limiting factor in the current NISQ era. To effectively mitigate the impact of noise in QML, strategies such as minimizing gate counts, employing QPUs with lower noise profiles, and designing circuits with inherently more resistant architectures must be pursued.

Our studies on the influence of noise on the training of QCBMs are limited to readout and two-qubit depolarizing errors. Future studies may benefit from incorporating a more complete noise model, including noise sources such as amplitude damping or crosstalk between qubit pairs. Furthermore, extending our studies to different quantum generative models, such as QGANs, would be interesting. After simulating the influence of noise on the performance of quantum generative models, conducting the experiments on quantum hardware would be a natural step. One compelling aspect to explore is investigating if a model trained with a noisy simulator has learned to resist the noise. Based on our experience, the QUARK framework should be the tool of choice for future research.

ACKNOWLEDGMENTS

PR and CAR were partly funded by the German Ministry for Education and Research (BMB+F) in the project QAI2-Q-KIS under Grant 13N15583. AL was partly funded by the Bavarian State Ministry of Economic Affairs in the project BenchQC under Grant DIK-0425/03.

- [1] Timothy Proctor, Kenneth Rudinger, Kevin Young, Erik Nielsen, and Robin Blume-Kohout. Measuring the capabilities of quantum computers. *Nature Physics*, 18(1): 75–79, 1 2022. ISSN 1745-2481. doi:10.1038/s41567-021-01409-7. URL <https://doi.org/10.1038/s41567-021-01409-7>.
- [2] Alexander Erhard, Joel J. Wallman, Lukas Postler, Michael Meth, Roman Stricker, Esteban A. Martinez, Philipp Schindler, Thomas Monz, Joseph Emerson, and Rainer Blatt. Characterizing large-scale quantum computers via cycle benchmarking. *Nature Communications*, 10(1):5347, 11 2019. ISSN 2041-1723. doi:10.1038/s41467-019-13068-7. URL <https://doi.org/10.1038/s41467-019-13068-7>.
- [3] Robin Blume-Kohout and Kevin C. Young. A volumetric framework for quantum computer benchmarks. *Quantum*, 4:362, November 2020. ISSN 2521-327X. doi:10.22331/q-2020-11-15-362. URL <https://doi.org/10.22331/q-2020-11-15-362>.
- [4] Daniel Mills, Seyon Sivarajah, Travis L. Scholten, and Ross Duncan. Application-Motivated, Holistic Benchmarking of a Full Quantum Computing Stack. *Quantum*, 5:415, March 2021. ISSN 2521-327X. doi:10.22331/q-2021-03-22-415. URL <https://doi.org/10.22331/q-2021-03-22-415>.
- [5] Salonik Resch and Ulya R. Karpuzcu. Benchmarking quantum computers and the impact of quantum noise. *ACM Comput. Surv.*, 54(7), 07 2021. ISSN 0360-0300. doi:10.1145/3464420. URL <https://doi.org/10.1145/3464420>.
- [6] Thomas Lubinski, Sonika Johri, Paul Varosy, Jeremiah Coleman, Luning Zhao, Jason Necaie, Charles H. Baldwin, Karl Mayer, and Timothy Proctor. Application-oriented performance benchmarks for quantum computing, 2023.
- [7] Thomas Lubinski, Joshua J. Goings, Karl Mayer, Sonika Johri, Nithin Reddy, Aman Mehta, Niranjan Bhatta, Sonny Rappaport, Daniel Mills, Charles H. Baldwin, Luning Zhao, Aaron Barbosa, Smarak Maity, and Pranav S. Mundada. Quantum algorithm exploration using application-oriented performance benchmarks, 2024.
- [8] Jernej Rudi Finžgar, Philipp Ross, Leonhard Holscher, Johannes Klepsch, and Andre Luckow. QUARK: A framework for quantum computing application benchmarking. In *2022 IEEE International Conference on Quantum Computing and Engineering (QCE)*. IEEE, 9 2022. doi:10.1109/qce53715.2022.00042. URL <https://doi.org/10.1109/qce53715.2022.00042>.
- [9] Joseph Bowles, Shah Nawaz Ahmed, and Maria Schuld. Better than classical? the subtle art of benchmarking quantum machine learning models, 2024.
- [10] T. D. Ladd, F. Jelezko, R. Laflamme, Y. Nakamura, C. Monroe, and J. L. O’Brien. Quantum computers. *Nature*, 464(7285):45–53, 03 2010. ISSN 1476-4687. doi:10.1038/nature08812. URL <https://doi.org/10.1038/nature08812>.
- [11] Florian J. Kiwit, Marwa Marso, Philipp Ross, Carlos A. Riofrío, Johannes Klepsch, and Andre Luckow. Application-oriented benchmarking of quantum generative learning using quark, 2023.
- [12] Marcello Benedetti, Erika Lloyd, Stefan Sack, and Mattia Fiorentini. Parameterized quantum circuits as machine learning models. *Quantum Science and Technology*, 4(4): 043001, 11 2019. doi:10.1088/2058-9565/ab4eb5. URL <https://doi.org/10.1088/2058-9565/ab4eb5>.
- [13] Maria Schuld and Francesco Petruccione. *Machine Learning with Quantum Computers*. Quantum Science and Technology. Springer International Publishing, Cham, 2021. doi:10.1007/978-3-030-83098-4. URL <https://link.springer.com/10.1007/978-3-030-83098-4>.
- [14] Marcello Benedetti, Delfina Garcia-Pintos, Oscar Perdomo, Vicente Leyton-Ortega, Yunseong Nam, and Alejandro Perdomo-Ortiz. A generative modeling approach for benchmarking and training shallow quantum circuits. *npj Quantum Information*, 5(1):1–9, 2019.
- [15] Seth Lloyd and Christian Weedbrook. Quantum generative adversarial learning. *Phys. Rev. Lett.*, 121:040502, 07 2018. doi:10.1103/PhysRevLett.121.040502. URL <https://link.aps.org/doi/10.1103/PhysRevLett.121.040502>.
- [16] Carlos A. Riofrío, Oliver Mitevski, Caitlin Jones, Florian Krellner, Aleksandar Vučković, Joseph Doetsch, Johannes Klepsch, Thomas Ehmer, and Andre Luckow. A performance characterization of quantum generative models, 2023.
- [17] Nikolaus Hansen, Youhei Akimoto, and Petr Baudis. CMA-ES/pycma on Github. Zenodo, DOI:10.5281/zenodo.2559634, February 2019. URL <https://doi.org/10.5281/zenodo.2559634>.
- [18] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence, and K.Q. Weinberger, editors, *Advances in Neural Information Processing Systems*, volume 27. Curran Associates, Inc., 2014. URL <https://proceedings.neurips.cc/paper/2014/file/5ca3e9b122f61f8f06494c97b1afccf3-Paper.pdf>.
- [19] Maria Schuld, Ville Bergholm, Christian Gogolin, Josh Izaac, and Nathan Killoran. Evaluating analytic gradients on quantum hardware. *Phys. Rev. A*, 99:032331, 03 2019. doi:10.1103/PhysRevA.99.032331. URL <https://link.aps.org/doi/10.1103/PhysRevA.99.032331>.
- [20] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization, 2017.
- [21] R. Jain. *The art of computer systems performance analysis: techniques for experimental design, measurement, simulation, and modeling*. Wiley New York, 1991.
- [22] Mirko Amico, Helena Zhang, Petar Jurcevic, Lev S. Bishop, Paul Nation, Andrew Wack, and David C. McKay. Defining standard strategies for quantum benchmarks, 2023.
- [23] Junchao Wang, Guoping Guo, and Zheng Shan. Sok: Benchmarking the performance of a quantum computer. *Entropy*, 24(10):1467, 2022.
- [24] Jens Eisert, Dominik Hangleiter, Nathan Walk, Ingo Roth, Damian Markham, Rhea Parekh, Ulysse Chabaud, and Elham Kashefi. Quantum certification and benchmarking. *Nature Reviews Physics*, 2(7):382–390, 07 2020. ISSN 2522-5820. doi:10.1038/s42254-020-0186-4. URL <https://doi.org/10.1038/s42254-020-0186-4>.
- [25] Andrew W. Cross, Lev S. Bishop, Sarah Sheldon, Paul D.

- Nation, and Jay M. Gambetta. Validating quantum computers using randomized model circuits. *Phys. Rev. A*, 100:032328, 09 2019. doi:10.1103/PhysRevA.100.032328. URL <https://link.aps.org/doi/10.1103/PhysRevA.100.032328>.
- [26] Andrew Wack, Hanhee Paik, Ali Javadi-Abhari, Petar Jurcevic, Ismael Faro, Jay M. Gambetta, and Blake R. Johnson. Quality, speed, and scale: three key attributes to measure the performance of near-term quantum computers, 2021.
- [27] Matthias Möller Koen Mesman, Zaid Al-Ars. Qpack: Quantum approximate optimization algorithms as universal benchmark for quantum computers. *arXiv*, 2022. doi:10.48550/arXiv.2103.17193. URL <https://doi.org/10.48550/arXiv.2103.17193>.
- [28] Thomas Lubinski, Carleton Coffrin, Catherine McGeoch, Pratik Sathe, Joshua Apanavicius, and David E. Bernal Neira. Optimization applications as quantum performance benchmarks, 2023.
- [29] Simon Martiel, Thomas Ayril, and Cyril Allouche. Benchmarking quantum coprocessors in an application-centric, hardware-agnostic, and scalable way. *IEEE Transactions on Quantum Engineering*, 2:1–11, 2021. ISSN 2689-1808. doi:10.1109/tqe.2021.3090207. URL <http://dx.doi.org/10.1109/TQE.2021.3090207>.
- [30] GitHub. Quark: A framework for quantum computing application benchmarking. <https://github.com/QUARK-framework/QUARK>, 2023.
- [31] Qiskit contributors. Qiskit: An open-source framework for quantum computing, 2023.
- [32] Ville Bergholm, Josh Izaac, Maria Schuld, Christian Gogolin, M Sohaib Alam, Shahnawaz Ahmed, Juan Miguel Arrazola, Carsten Blank, Alain Delgado, Soran Jahangiri, et al. Pennylane: Automatic differentiation of hybrid quantum-classical computations. *arXiv preprint arXiv:1811.04968*, 2018.
- [33] Elton Yechao Zhu, Sonika Johri, Dave Bacon, Mert Esencan, Jungsang Kim, Mark Muir, Nikhil Murgai, Jason Nguyen, Neal Pseni, Adam Schouela, Ksenia Sosnova, and Ken Wright. Generative quantum learning of joint probability distribution functions. *Phys. Rev. Res.*, 4:043092, 11 2022. doi:10.1103/PhysRevResearch.4.043092. URL <https://link.aps.org/doi/10.1103/PhysRevResearch.4.043092>.
- [34] Erick Winston and Diego Moreda. Qiskit backends: What they are and how to work with them. *Medium*, 2018. URL <https://medium.com/qiskit/qiskit-backends-what-they-are-and-how-to-work-with-them-fb66b3bd0463>.
- [35] Jin-Guo Liu and Lei Wang. Differentiable learning of quantum circuit born machines. *Phys. Rev. A*, 98, 12 2018. doi:10.1103/PhysRevA.98.062324. URL <https://link.aps.org/doi/10.1103/PhysRevA.98.062324>.
- [36] Matthias Steffen, Jerry Chow, Sarah Sheldon, and Doug McClure. IBM Research A new eagle in the poughkeepsie quantum datacenter: Ibm quantum’s most performant system yet. <https://research.ibm.com/blog/eagle-quantum-error-mitigation>, 2024.
- [37] IonQ Collaborators. IonQ Harmony. <https://ionq.com/quantum-systems/harmony>, 2024.
- [38] John Preskill. Quantum computing in the nisq era and beyond. *Quantum*, 2:79, 2018. doi:10.22331/q-2018-08-06-79. URL <https://doi.org/10.22331/q-2018-08-06-79>.
- [39] Michael A. Nielsen and Isaac L. Chuang. *Quantum Computation and Quantum Information: 10th Anniversary Edition*. Cambridge University Press, 2010. doi:10.1017/CBO9780511976667.
- [40] Marita Oliv, Andrea Matic, Thomas Messerer, and Jeanette Miriam Lorenz. Evaluating the impact of noise on the performance of the variational quantum eigensolver, 2022.