

Quantum Fourier Generative Models Trainable at Large Scale

Cenk Tüysüz¹, Oleksandr Kyriienko² and Michele Grossi¹

¹ European Organisation for Nuclear Research (CERN), 1211 Geneva, Switzerland

² School of Mathematical and Physical Sciences, University of Sheffield, Sheffield S10 2TN, United Kingdom

Quantum generative models provide sampling access to probability distributions that can be hard to represent and sample classically, yet training them at scale is limited by vanishing gradients and shot noise. In this work, we propose an algorithmic framework for building differentiable quantum generative models (DQGMs) [1] based on the *train-on-classical deploy-on-quantum* approach (see Fig. 1 for a visual summary). Unlike prior work that relied on the maximum mean discrepancy (MMD) loss [2, 3], our approach goes beyond matching low-order spectral moments, while enabling efficient classical training. Once the model is trained, we use inverse quantum Fourier transforms to map it into a separate sampling circuit in the computational basis. We validate the efficiency of the framework by evaluating the loss of models with more than 1000 qubits in under one second on a single GPU. We show that univariate and bivariate targets with multi-modal, oscillatory, and heavy-tailed structure can be learned to low total variation distance, while instantaneous quantum polynomial (IQP) models trained with the MMD loss perform poorly despite dedicated hyperparameter optimization. Compared to classical normalizing flow and diffusion baselines, our approach avoids oversmoothing and preserves the multi-modal structure of the target. For a financial example, we train models representing stochastic variables of daily log-returns across four assets, capturing the cross-asset correlations and outperforming MMD-based training. Finally, we deploy the trained models on superconducting quantum devices and sample the learned distributions at approximately $300\mu s$ per sample. Our work shows that quantum generative models within the train-on-classical deploy-on-quantum approach can provide both high-quality distribution matching at scale and the fast sampling access needed for inference. This extended abstract summarizes the preprint Ref. [4], which contains the theoretical analysis of the classical estimator and the complete set of benchmarks.

Introduction. Generative modelling is based on representing, learning, and sampling complex multivariate probability distributions [5–8]. Quantum generative modelling takes advantage of the intrinsic probabilistic nature of quantum systems and the ability to sample through projective measurements [9–11]. The ex-

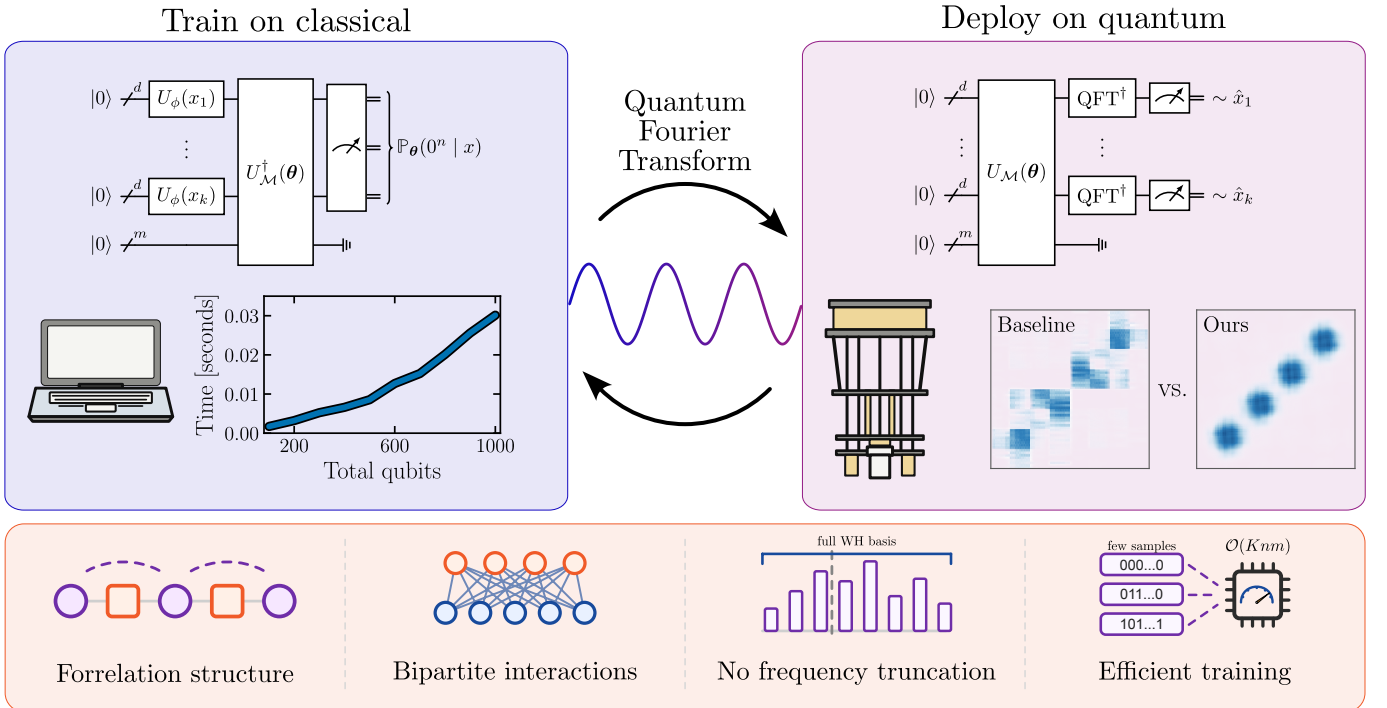


Fig. 1: **The *train-on-classical deploy-on-quantum* workflow.** The model is trained classically (left), and inverse quantum Fourier transforms map it into a sampling circuit executed on quantum hardware (right).

ponential sampling advantage for certain families of quantum circuits [12–15] motivated the development of trainable quantum generative models (QGMs) that can learn efficiently while improving the scaling for sample generation [16].

Training and inference of quantum generative models depend on their architecture and loss [17]. The trade-off between the expressivity and trainability of variational quantum circuits [18, 19] often limits training to smaller system sizes due to vanishing gradients [20, 21] and local minima [22, 23], given that efficient training implies low shot noise when training on quantum devices. However, this trade-off is alleviated if gradients can be calculated classically, at large scale and the required precision, avoiding costly parameter shift rules [24–26]. Here, generative modelling offers a distinct setting, as proposed for QGMs based on IQP circuits [27]. For this architecture, training can be done on a classically tractable loss, while inference can be computationally hard and is off-loaded to quantum devices. The separation of training and inference stages hence led to the emergence of QGMs based on the *train-on-classical deploy-on-quantum* approach [2, 3], which has also been extended to fermionic linear optics circuits [28], shallow circuits [16], and boson-sampling architectures [29–31].

For IQP-based quantum circuit Born machines, the classically tractable training objective is an MMD loss, computed from expectation values of diagonal observables [32, 33]. Effectively, this relies on matching the lowest moments of the target distribution in a spectral domain [34], where in a Walsh-Hadamard representation the kernel puts large weight on low-order correlations [35, 36]. When the target has strong correlations across features, the relevant structure can lie in high-order components that the kernel weights weakly, so these correlations contribute little to the loss and its gradient, and MMD training can miss them even when the model class can represent the target (we give a concrete example below). The underlying issue is not the use of spectral information itself [34, 37], but the fixed spectral weighting imposed by a prescribed MMD kernel. To obtain high-quality matching of QGMs to the target distribution, we need qualitatively new approaches to model architecture building and training objectives.

Numerical results. We begin with univariate benchmark distributions, where the learned density can be compared to the target on a high-resolution grid. Fig. 2a shows four targets defined on a 2^{16} -point discretization of $x \in [0, 1]$, chosen to probe localization, separated modes, oscillatory structure, and asymmetric heavy-

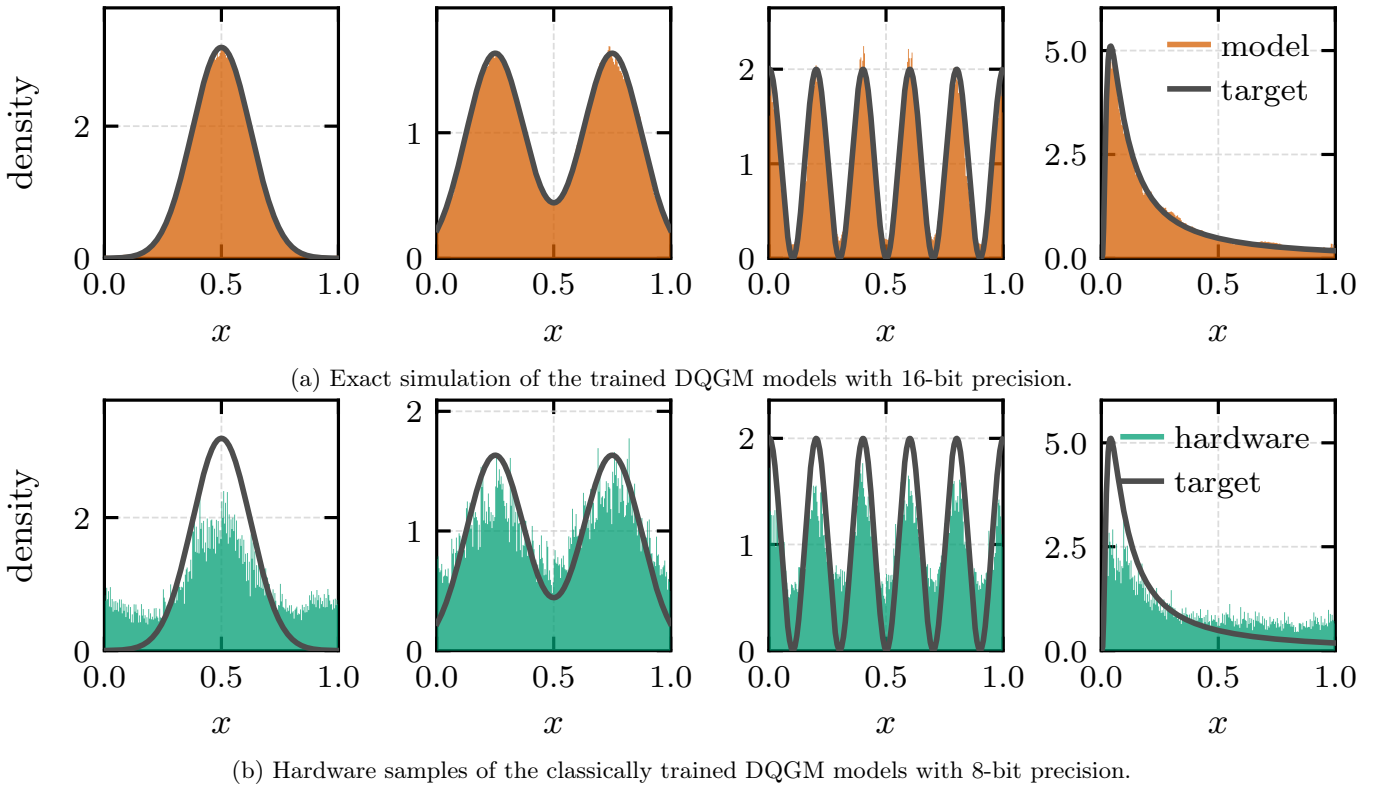


Fig. 2: **Univariate benchmarks in exact simulation and on quantum hardware.** Each panel compares the target density (grey line) with the model (histogram) for a localized Gaussian, a two-component Gaussian mixture, an oscillatory cosine density, and a truncated Lévy distribution.

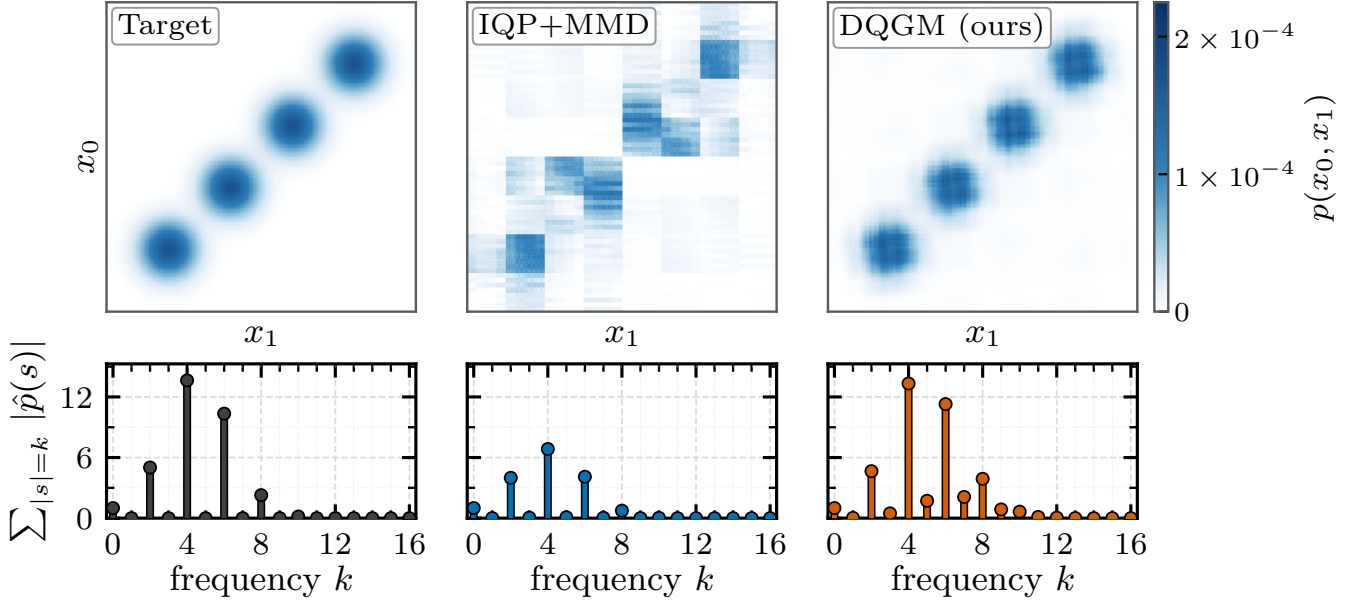


Fig. 3: **Bivariate benchmark against an IQP+MMD baseline.** The top row shows the target and learned model densities. The bottom row shows the Walsh-Hadamard spectrum $(\hat{p}(s) = \sum_{x \in \{0,1\}^n} p(x) (-1)^{s \cdot x})$ of each, aggregated by Hamming weight k .

tailed decay. All runs share the same architecture with $m = 4$ hidden qubits, initialization, hyperparameters, and the negative log-likelihood objective, and each completes in under a minute on a single GPU. While training uses our Monte Carlo estimator of the loss, the curves in Fig. 2a show the exact distribution of the trained models on the discrete support, not finite-sample histograms, and reproduce the dominant structure of all four targets. The final errors are low. The Kullback-Leibler divergence lies in $[0.0024, 0.0313]$ and the total variation distance in $[0.0206, 0.0653]$, with the largest error for the oscillatory target.

Next, we test whether classically trained circuits can be sampled on present quantum hardware, executing only the final sampling step on the device. Device noise limits the precision that can be sampled faithfully, so we train 8-bit versions of the models for deployment. Using the same family of univariate targets, we sample each trained model on `ibm_aachen` and report raw histograms without error mitigation. Fig. 2b shows that the hardware samples retain the dominant learned structure despite device noise. For the system sizes considered here, the main remaining bottleneck to scale our results is not classical training but accurate sampling from the trained circuit on quantum hardware.

For the final benchmark, we train both models on the same four-peak Gaussian target, matching the IQP+MMD baseline to the DQGM in hidden qubits, gate order, and parameter count. Fig. 3 shows that the IQP+MMD baseline captures part of the coarse structure but produces block-like artifacts and does not reproduce the four localized peaks as accurately as the DQGM.

We attribute this gap to two limitations of the IQP+MMD approach. First, IQP circuits are limited in expressivity [3], and matching the DQGM requires up to four-qubit gates and exact evaluation of the MMD loss, which does not scale beyond small systems. Second, the approach assumes that the Walsh-Hadamard spectrum of the target concentrates at low Hamming weight, an assumption that fails for correlated multivariate distributions such as the four-peak target. As the lower panel of Fig. 3 shows, the IQP+MMD spectrum matches the target only up to Hamming weight $k = 3$, whereas the DQGM matches all dominant frequencies. Ref. [4] details the dedicated hyperparameter optimization for the baselines and our models, and we release our implementation and data in Refs. [38, 39].

Conclusion. Our results show that classical training need not limit the quality of quantum generative models. The framework evaluates the loss of models with more than 1000 qubits in under one second on a single GPU, learns multi-modal and correlated targets that MMD-based training misses, and samples at approximately $300 \mu\text{s}$ per sample on superconducting hardware. As logical error rates of early fault-tolerant devices decrease, the same circuits could be sampled with fidelity sufficient to resolve the learned distribution, extending quantum generative modelling beyond the reach of exact classical simulation.

References

1. O. Kyriienko, A. E. Paine, and V. E. Elfving, “Protocols for trainable and differentiable quantum generative modeling,” *Physical Review Research*, vol. 6, no. 3, p. 033291, Sep. 2024.
2. E. Recio-Armengol, S. Ahmed, and J. Bowles, “Train on classical, deploy on quantum: scaling generative quantum machine learning to a thousand qubits,” 2026. [Online]. Available: <https://arxiv.org/abs/2503.02934>
3. A. Kurkin, K. Shen, S. Pielawa, H. Wang, and V. Dunjko, “Universality and kernel-adaptive training for classically trained, quantum-deployed generative models,” 2025. [Online]. Available: <https://arxiv.org/abs/2510.08476>
4. C. Tüysüz, O. Kyriienko, and M. Grossi, “Quantum fourier generative models trainable at large scale,” 2026. [Online]. Available: <https://arxiv.org/abs/2606.28483>
5. Y. Bengio, L. Yao, G. Alain, and P. Vincent, “Generalized denoising auto-encoders as generative models,” 2013. [Online]. Available: <https://arxiv.org/abs/1305.6663>
6. D. Rezende and S. Mohamed, “Variational inference with normalizing flows,” in *Proceedings of the 32nd International Conference on Machine Learning*, ser. Proceedings of Machine Learning Research, F. Bach and D. Blei, Eds., vol. 37. Lille, France: PMLR, 07–09 Jul 2015, pp. 1530–1538. [Online]. Available: <https://proceedings.mlr.press/v37/rezende15.html>
7. J. Song, C. Meng, and S. Ermon, “Denoising diffusion implicit models,” in *International Conference on Learning Representations*, 2021. [Online]. Available: <https://openreview.net/forum?id=St1giarCHLP>
8. Y. Song, J. Sohl-Dickstein, D. P. Kingma, A. Kumar, S. Ermon, and B. Poole, “Score-based generative modeling through stochastic differential equations,” in *International Conference on Learning Representations*, 2021. [Online]. Available: <https://openreview.net/forum?id=PXTIG12RRHS>
9. J.-G. Liu and L. Wang, “Differentiable learning of quantum circuit born machines,” *Phys. Rev. A*, vol. 98, p. 062324, Dec 2018. [Online]. Available: <https://link.aps.org/doi/10.1103/PhysRevA.98.062324>
10. C. Zoufal, A. Lucchi, and S. Woerner, “Quantum generative adversarial networks for learning and loading random distributions,” *npj Quantum Information*, vol. 5, no. 1, p. 103, Nov 2019. [Online]. Available: <https://doi.org/10.1038/s41534-019-0223-2>
11. B. Coyle, D. Mills, V. Danos, and E. Kashefi, “The born supremacy: quantum advantage and training of an ising born machine,” *npj Quantum Information*, vol. 6, no. 1, p. 60, Jul 2020. [Online]. Available: <https://doi.org/10.1038/s41534-020-00288-9>
12. M. J. Bremner, R. Jozsa, and D. J. Shepherd, “Classical simulation of commuting quantum computations implies collapse of the polynomial hierarchy,” *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 467, no. 2126, pp. 459–472, Aug. 2010.
13. S. Aaronson and A. Arkhipov, “The computational complexity of linear optics,” *Theory of Computing*, vol. 9, no. 4, pp. 143–252, 2013. [Online]. Available: <https://theoryofcomputing.org/articles/v009a004>
14. 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, and et al., “Quantum supremacy using a programmable superconducting processor,” *Nature*, vol. 574, no. 7779, pp. 505–510, Oct 2019. [Online]. Available: <https://doi.org/10.1038/s41586-019-1666-5>
15. D. Hangleiter and J. Eisert, “Computational advantage of quantum random sampling,” *Rev. Mod. Phys.*, vol. 95, p. 035001, Jul 2023. [Online]. Available: <https://link.aps.org/doi/10.1103/RevModPhys.95.035001>
16. H.-Y. Huang, M. Broughton, N. Eassa, H. Neven, R. Babbush, and J. R. McClean, “Generative quantum advantage for classical and quantum problems,” 2025. [Online]. Available: <https://arxiv.org/abs/2509.09033>
17. S. Herbst, I. Brandić, and A. Pérez-Salinas, “Limits of quantum generative models with classical sampling hardness,” 2025. [Online]. Available: <https://arxiv.org/abs/2512.24801>
18. Z. Holmes, K. Sharma, M. Cerezo, and P. J. Coles, “Connecting ansatz expressibility to gradient magnitudes and barren plateaus,” *PRX Quantum*, vol. 3, p. 010313, Jan 2022. [Online]. Available: <https://link.aps.org/doi/10.1103/PRXQuantum.3.010313>
19. M. Cerezo, M. Larocca, D. García-Martín, N. L. Diaz, P. Braccia, E. Fontana, M. S. Rudolph, P. Bermejo, A. Ijaz, S. Thanasilp, E. R. Anschuetz, and Z. Holmes, “Does provable absence of barren plateaus imply classical simulability?” *Nature Communications*, vol. 16, no. 1, p. 7907, Aug 2025. [Online]. Available: <https://doi.org/10.1038/s41467-025-63099-6>
20. J. R. McClean, S. Boixo, V. N. Smelyanskiy, R. Babbush, and H. Neven, “Barren plateaus in quantum neural network training landscapes,” *Nature Communications*, vol. 9, no. 1, p. 4812, Nov. 2018.
21. M. Larocca, S. Thanasilp, S. Wang, K. Sharma, J. Biamonte, P. J. Coles, L. Cincio, J. R. McClean, Z. Holmes, and M. Cerezo, “Barren plateaus in variational quantum computing,” *Nature Reviews Physics*, vol. 7, no. 4, pp. 174–189, Apr 2025.
22. X. You and X. Wu, “Exponentially many local minima in quantum neural networks,” 2021. [Online]. Available: <https://arxiv.org/abs/2110.02479>
23. E. R. Anschuetz, “Critical points in quantum generative models,” 2023. [Online]. Available: <https://arxiv.org/abs/2109.06957>
24. 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>
25. O. Kyriienko and V. E. Elfving, “Generalized quantum circuit differentiation rules,” *Phys. Rev. A*, vol. 104, p. 052417, Nov 2021. [Online]. Available: <https://link.aps.org/doi/10.1103/PhysRevA.104.052417>
26. D. Wierichs, J. Izaac, C. Wang, and C. Y.-Y. Lin, “General parameter-shift rules for quantum gradients,” *Quantum*, vol. 6, p. 677, Mar. 2022. [Online]. Available: <https://doi.org/10.22331/q-2022-03-30-677>
27. S. Kasture, O. Kyriienko, and V. E. Elfving, “Protocols for classically training quantum generative models on probability distributions,” *Phys. Rev. A*, vol. 108, p. 042406, Oct 2023. [Online]. Available: <https://link.aps.org/doi/10.1103/PhysRevA.108.042406>
28. B. Bakó, Z. Kolarovszki, and Z. Zimborás, “Fermionic born machines: Classical training of quantum generative models based on fermion sampling,” 2025. [Online]. Available: <https://arxiv.org/abs/2511.13844>

29. Z. Kolarovszki, B. Bakó, M. Oszmaniec, C. Oh, and Z. Zimborás, “Generative modeling with gaussian boson sampling: classically trainable bosonic born machines,” 2026. [Online]. Available: <https://arxiv.org/abs/2603.11195>
30. F. Gottlieb, R. Mezher, B. Ventura, S. Mansfield, and A. Salavrakos, “Efficient training of photonic quantum generative models,” 2026. [Online]. Available: <https://arxiv.org/abs/2603.08793>
31. A. Kurkin, U. Chabaud, Z. Kolarovszki, B. Bakó, Z. Zimborás, and V. Dunjko, “Universality of classically trainable, quantum-deployed boson-sampling generative models,” 2026. [Online]. Available: <https://arxiv.org/abs/2603.11014>
32. M. V. den Nest, “Simulating quantum computers with probabilistic methods,” 2010. [Online]. Available: <https://arxiv.org/abs/0911.1624>
33. M. S. Rudolph, S. Lerch, S. Thanasilp, O. Kiss, O. Shaya, S. Vallecorsa, M. Grossi, and Z. Holmes, “Trainability barriers and opportunities in quantum generative modeling,” *npj Quantum Information*, vol. 10, no. 1, p. 116, Nov. 2024.
34. V. Belis, J. Bowles, R. Gupta, E. Peters, and M. Schuld, “Spectral methods: crucial for machine learning, natural for quantum computers?” 2026. [Online]. Available: <https://arxiv.org/abs/2603.24654>
35. M. Herrero-Gonzalez, B. Coyle, K. McDowall, R. Grassie, S. Beentjes, A. Khamseh, and E. Kashefi, “The born ultimatum: Conditions for classical surrogation of quantum generative models with correlators,” 2025. [Online]. Available: <https://arxiv.org/abs/2511.01845>
36. K. Shen, S. Pielawa, V. Dunjko, and H. Wang, “Characterizing trainability of instantaneous quantum polynomial circuit born machines,” 2026. [Online]. Available: <https://arxiv.org/abs/2602.11042>
37. D. Wakeham and M. Schuld, “Inference, interference and invariance: How the quantum fourier transform can help to learn from data,” 2024. [Online]. Available: <https://arxiv.org/abs/2409.00172>
38. C. Tüysüz *et al.*, <https://github.com/cnktysz/DQGM>, 2026, GitHub repository.
39. —, <https://doi.org/10.6084/m9.figshare.32789952>, 2026, Public dataset repository.