

IQP Born Machines under Data-dependent & Agnostic Initializations

Sacha Lerch

Joseph Bowles

Ricard Puig

Erik Armengol

Zoë Holmes

Supanut Thanasilp

Abstract. Instantaneous quantum polynomial-time (IQP) circuits provide a natural framework for quantum generative modeling and are believed to be classically hard to sample from in certain regimes. We study the optimization landscape of IQP-based quantum circuit Born machines trained with Maximum Mean Discrepancy (MMD) losses constructed from low-body Pauli-Z correlators. We prove that random full-range initializations exhibit barren plateaus, while identity and unbiased data-agnostic initializations admit non-vanishing loss variance lower bounds. We further introduce a data-dependent initialization that better aligns with the target distribution, yielding provable gradients and faster convergence in practice, including experiments on 150-qubit genomic datasets.

This is the extended abstract of Ref. [1].

1 Introduction

Quantum machine learning seeks to leverage quantum-mechanical effects to recognize and generate complex patterns that may be difficult for classical models to capture. Among the various learning tasks, generative modeling has emerged as one of the most promising candidates for demonstrating a practical quantum advantage. The central idea is simple: quantum states naturally define probability distributions, and measurements provide a direct way to sample from complex, potentially non-classical patterns. A widely studied framework for quantum generative modeling is the Quantum Circuit Born Machine (QCBM), where a parameterized quantum circuit is measured in the computational basis to generate samples according to the Born rule [2–6].

However, scalability of QCBMs, and more broadly variational quantum models, remains a central challenge. A particularly well-known obstacle are barren plateaus, also referred to as exponential concentration, in which the optimization landscape becomes exponentially flat as the system size grows [7, 8]. In this regime, random initialization lands in flat regions with overwhelming probability, leading to vanishing gradients and practical untrainability. Importantly, trainable regions with substantial gradients may still exist, but locating them becomes increasingly difficult at scale. This has motivated the study of alternative initialization strategies as a potential route to mitigating barren plateaus [9–15].

In parallel, Instantaneous Quantum Polynomial-time (IQP) based QCBMs have become increasingly attractive. Sampling from certain instances of IQP-QCBMs is believed to be classically hard under standard complexity-theoretic assumptions [16–19], while key quantities required for training can be estimated efficiently using classical computation. Together, these features motivate a promising “classically-train, quantum-deploy” paradigm, in which models are trained using classical resources and subsequently deployed on quantum hardware [4, 20–23]. Recent work has demonstrated that IQP-QCBMs can be trained at scales exceeding a thousand qubits [20], suggesting that IQP-QCBMs are a compelling platform for studying the scalability of quantum generative models and eventually achieving a quantum advantage.

Crucially, these recent empirical findings also suggest that IQP-QCBMs exhibit barren plateaus under standard random initialization [20]. Nevertheless, these models have been shown to remain trainable at surprisingly large scales when equipped with suitable warm-start strategies. This raises a central machine-learning question: *when does an initialization place the model in a trainable region of the loss landscape?*

Our contribution: We provide theoretical and numerical analysis of trainability for IQP-QCBMs under three families of initialization: (i) uniformly random initializations, (ii) small-angle data-agnostic initializations, and (iii) data-dependent initialization strategies. In particular, we rigorously show the emergence of barren plateaus over the whole loss landscape with all-to-all IQP topology, excluding the practicality of the full angle range initialization. For small-angle range initialization strategies, we explicitly connect the loss variance around a small patch to the curvature of the loss landscape at the initialization center. Consequently, we show that agnostic initialization strategies and data-dependent strategy under reasonable physical assumptions of a target distribution results in patches with substantial gradients, exhibiting initial trainability guarantees. More broadly, these variance lower bounds apply to a general class of non-linear losses, providing a versatile toolset for analyzing warm-starts across quantum machine learning.

Going beyond the variance analysis, we provide empirical evidence with large scale 150 qubit simulation that data-dependent initialization strategies consistently converge faster than an agnostic unbiased strategy where the initial guess corresponds to the uniform distribution over all possible bit strings, and have better performance over the (popular) identity initialization strategy. This analysis clarifies when and why data-dependent warm starts can outperform data-agnostic ones. A summary of our results is provided in Fig. 1 and Section 3.

2 Framework

Quantum generative model aims to learn a target distribution through the principle of quantum mechanics. Notably, QCBMs leverage the Born rule and parametrizes a quantum state as $|\psi(\theta)\rangle = U(\theta)|\psi_0\rangle$ with a parametrized circuit $U(\theta)$ and some initial state $|\psi_0\rangle$, leading to the model distribution (in the computational basis $\{|z\rangle\}_z$) of the form $p_\theta(z) = |\langle z|\psi(\theta)\rangle|^2$. In addition, an IQP-QCBM takes a circuit architecture as

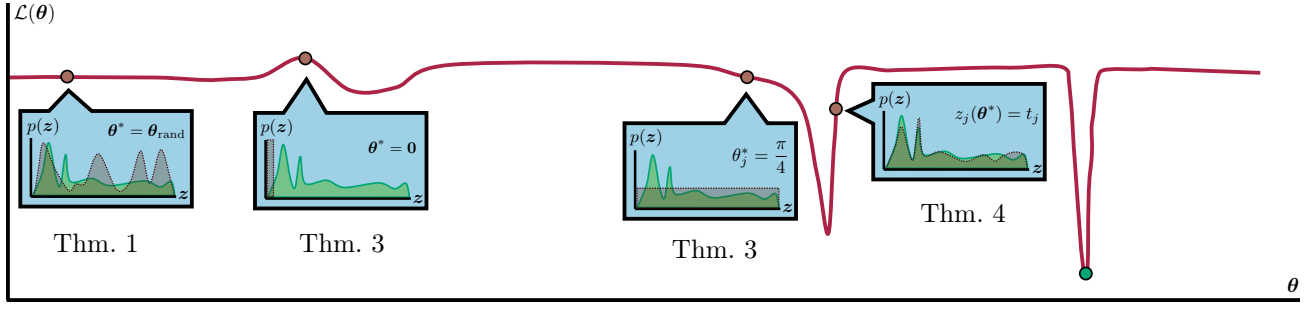


Figure 1: **Illustration of our main results.** Schematic MMD loss landscape $\mathcal{L}(\theta)$ for an IQP generative model. The insets show the corresponding computational-basis distributions at representative initialization points: the model distribution $p_\theta()$ is shown in gray, and the target distribution $p_{\text{data}}()$ in green. Identity initialization ($\theta^* = \mathbf{0}$) yields a highly biased distribution, $p_{\theta^*}(\mathbf{0}) = 1$, which corresponds to a local maximum of the MMD. An unbiased data-agnostic initialization ($\theta_j^* = \pi/4$) yields a uniform distribution over bit strings and usually corresponds to a region with gradient. A data-dependent initialization yields an initial distribution that is better aligned with the target. The green point indicates schematically the minimum corresponding to $p_{\text{data}}()$.

$U(\theta) = \exp\left(i \sum_{j=1}^n \theta_j X_j + i \sum_{(j,k) \in E} \theta_{jk} X_j X_k\right)$ where $\theta = \{\theta_i, \theta_{jk}\}$ are parameters to be optimized, X_j is a Pauli- X operator on the qubit j , and $E \subseteq \{(j,k) : j < k\}$ denotes the set of edges defining the interaction graph (i.e., the qubit topology).

The ultimate goal is to train QCBMs such that the model distribution well approximates the target distribution $p_{\theta_{\text{opt}}}(\mathbf{z}) \approx p_{\text{data}}(\mathbf{z})$ through minimizing the loss function. More specifically, we are interested in Maximum Mean Discrepancy (MMD) loss which acts as *correlator* matching between two distributions

$$\mathcal{L}(\theta) = \sum_{A \subseteq [n]} w_A (z_A(\theta) - t_A)^2, \quad (1)$$

where A represents a subset of qubits i.e., $A \subseteq [n]$ with $[n] = \{1, 2, \dots, n\}$, $z_A(\theta) = \langle \psi(\theta) | Z_A | \psi(\theta) \rangle$ is a *model correlator* associated with this subset expressed as the expectation value of the multi-qubit Pauli-Z operator $Z_A = \bigotimes_{j \in A} Z_j$, similarly $t_A = \mathbb{E}_{\mathbf{y} \sim p_{\text{data}}}[\langle \mathbf{y} | Z_A | \mathbf{y} \rangle]$ is a target correlator, and w_A are some weights in the sum.

Exponential concentration: Variational quantum algorithms exhibit a range of scalability challenges that also apply to generative models. Training can be hindered by limited circuit expressivity, local minima, or an unfavorable optimization landscape. Among these issues, exponential concentration of the loss landscape—equivalently, the barren plateau phenomenon [7, 8]—is the most thoroughly studied, including in the setting of the quantum generative modeling [24–26].

A landscape is said to exponentially concentrate if the variance of the loss over parameters vanishes exponentially with the system size n , i.e., $\text{Var}_\theta[\mathcal{L}(\theta)] \in \mathcal{O}(\exp(-n))$. Crucially, the presence of barren plateaus over the landscape does not necessarily imply that a quantum model is untrainable. Even within a predominantly flat landscape, there may exist small regions with substantial gradients that contain favorable local minima. This observation has motivated a range of analytical and numerical studies exploring initialization strategies to bypass barren plateaus [9–15].

3 Main Results

We investigate how different initialization strategies influence variance scaling and training for the MMD loss with IQP circuit architectures. These initialization strategies are characterized by the patch centre point θ^* and how big the patch is r , including 1. full range random initialization ($\theta^* = \mathbf{0}$ and $r = \pi/2$), 2. data-agnostic small-angle identity initialization where the initial parameters are set such that the circuit becomes identity ($\theta^* = \mathbf{0}$ and $r \ll 1$), 3. data-agnostic small-angle unbiased initialization where the initial model distribution corresponds to the uniform distribution over all possible bit-strings ($\theta_j^* = \pi/4$, $\theta_{jk}^* = 0$ and $r \ll 1$), and finally 4. data-dependent initialization where the parameters are initialized based on the training data (or empirical target statistics) through marginal matching, focusing on a simplified version of the initialization used in Ref. [20].

Untrainability under random initialization. We rigorously relate the topology of the IQP architecture to the flatness of the loss landscape for full-range random initialization. For all-to-all architectures, exponential concentration emerges, thereby ruling out this strategy at scale. This is stated more formally in *Theorem 1 of the manuscript*. However, for less connected models, e.g., those with constant degree connectivity, exponential concentration can be avoided (see *Appendix E2*).

Development of a practical diagnostic framework for analyzing alternative small-angle initialization strategies in quantum machine learning. We establish a general connection between the variance of the loss over a local parameter patch and the curvature of the loss landscape at the initialization point for non-linear loss functions. This leads to a practical diagnostic tool to determine whether a given initialization strategy can avoid barren

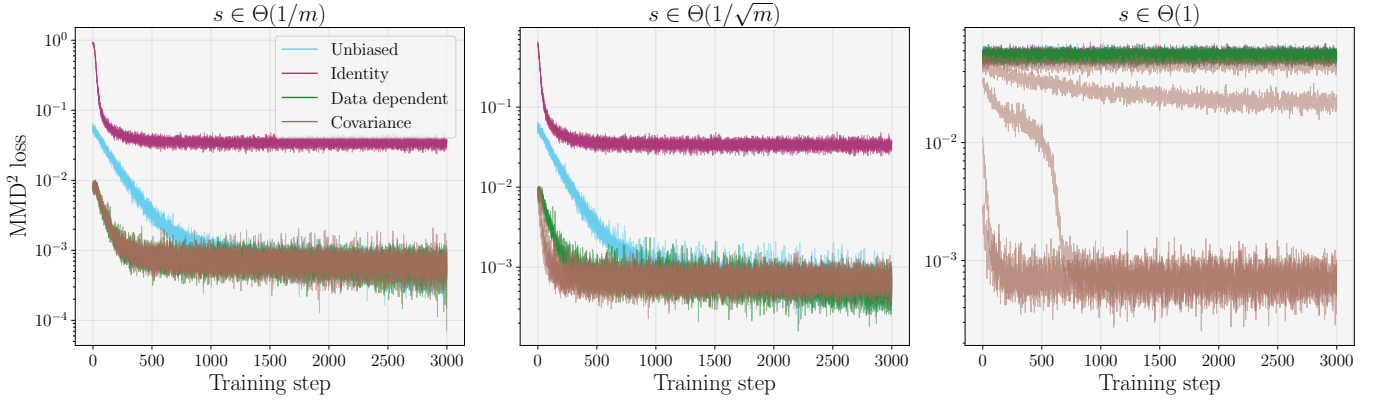


Figure 2: **Training of the low-body MMD estimator for different initializations.** Training loss $\mathcal{L}(\theta)$ as a function of gradient-descent iterations for a $n = 150$ qubit model trained on the genomic dataset. Each curve corresponds to one of the four initialization strategies: identity (red), unbiased (blue), marginal matching (green – the data-dependent strategy considered in the theoretical analysis of this work), and the covariance-based data-dependent initialization (light green – the data-dependent strategy in Ref. [20]). The three panels correspond to different initialization scales s (i.e., $\theta_j \sim \theta_j^* + \text{Unif}[-\frac{\pi}{2}s, \frac{\pi}{2}s]$): $s = 1/m$, $s = 1/\sqrt{m}$, and $s = 1$, where m denotes the total number of parameters. Each model is trained five times, with most initializations leading to similar overlapping loss curves.

plateaus. Intuitively, if the curvature (i.e., the Hessian) at the center of the patch is not exponentially vanishing ($\in \mathcal{O}(1/\text{poly}(n))$), the loss exhibit at least polynomially large variance within any patch (with width $r \in \mathcal{O}(1/\text{poly}(n))$). For MMD loss and IQP-QCBMs, this is shown in [Theorem 2](#). Furthermore, we show that the curvature decomposes into two distinct components: a *mismatch-driven* term and a *model-sensitivity* term (see [Eq. \(19\) of the manuscript](#)) where the first term scales with the discrepancy between model and target correlators and the latter term provides the intrinsic dependence of the model architecture.

More broadly, our curvature-based diagnostic framework applies to a general class of non-linear loss functions beyond those considered here. As such, it provides a general tool for analyzing warm-start strategies in quantum machine learning. This general theorem is shown in [Appendix G](#).

Theoretical guarantees to avoid barren plateaus with data-agnostic and data-dependent strategies.

We show that both data-agnostic strategies yields sufficient curvature to ensure polynomially large variances, thereby evading barren plateaus as the system scales (see [Theorem 3](#)). However, their practical utility differs significantly. While identity initialization is a popular approach, the resulting model distribution at this point is a delta distribution concentrated on the all-zero bit string. Such a highly biased starting point is often detrimental to generative modeling (as shown in Fig. 2). In contrast, the unbiased initialization produces a uniform distribution over all bit strings, representing perhaps the most natural prior for a data-agnostic strategy.

On the other hand, for many real world data sets (see [Assumption 1](#) for dominating low-body target marginals and [Assumption 2](#) for no peak target around single bit-string), we prove that the data-dependent strategy guarantees inverse-polynomial local curvature, thereby ensuring non-vanishing patch variance for the MMD loss (see [Theorem 4 of the manuscript](#)). To our knowledge, this provides the first theoretical evidence that data-dependent initialization can systematically circumvent barren plateaus. Lastly, the numerical simulation of variance scaling (shown in the manuscript) well aligns with the theoretical prediction.

Empirical evidence of superior performance of the data-dependent initialization.

Large-scale numerical experiments, including simulations up to 150 qubits on genomic datasets (as shown in Fig. 2), further demonstrate that the data-dependent strategy consistently achieves faster convergence and improved model fidelity compared to data-agnostic alternatives. These results highlight an important distinction: while both agnostic and data-dependent strategies can avoid vanishing gradients, data-dependent initialization places the model in a significantly more favorable optimization region.

Overall impact Together, our results establish a principled route to scalable quantum generative modeling, showing that trainability at large scale is not merely a matter of architecture, but crucially depends on how models are initialized. This provides both theoretical insight and practical guidance, paving a way towards a quantum advantage in generative learning as well as advancing the broader goal of quantum machine learning.

References

- [1] Sacha Lerch, Joseph Bowles, Ricard Puig, Erik Armengol, Zoë Holmes, and Supanut Thanasilp. Iqp born machines under data-dependent and agnostic initialization strategies. *arXiv preprint arXiv:2603.14576*, 2026.
- [2] Ryan Sweke, Jean-Pierre Seifert, Dominik Hangleiter, and Jens Eisert. On the quantum versus classical learnability of discrete distributions. *Quantum*, 5:417, 2021.
- [3] Xun Gao, Eric R. Anschuetz, Sheng-Tao Wang, J. Ignacio Cirac, and Mikhail D. Lukin. Enhancing generative models via quantum correlations. *Phys. Rev. X*, 12:021037, May 2022.
- [4] Hsin-Yuan Huang, Michael Broughton, Norhan Eassa, Hartmut Neven, Ryan Babbush, and Jarrod R McClean. Generative quantum advantage for classical and quantum problems. *arXiv preprint arXiv:2509.09033*, 2025.
- [5] Mohamed Hibat-Allah, Marta Mauri, Juan Carrasquilla, and Alejandro Perdomo-Ortiz. A framework for demonstrating practical quantum advantage: comparing quantum against classical generative models. *Communications Physics*, 7(1):68, 2024.
- [6] Brian Coyle, Daniel Mills, Vincent Danos, and Elham Kashefi. The born supremacy: quantum advantage and training of an ising born machine. *npj Quantum Information*, 6(1):60, 2020.
- [7] Jarrod R McClean, Sergio Boixo, Vadim N Smelyanskiy, Ryan Babbush, and Hartmut Neven. Barren plateaus in quantum neural network training landscapes. *Nature Communications*, 9(1):1–6, 2018.
- [8] Martin Larocca, Supanut Thanasilp, Samson Wang, Kunal Sharma, Jacob Biamonte, Patrick J Coles, Lukasz Cincio, Jarrod R McClean, Zoë Holmes, and M Cerezo. Barren plateaus in variational quantum computing. *Nature Reviews Physics*, 7(4):174–189, 2025.
- [9] Ricard Puig, Marc Drudis, Supanut Thanasilp, and Zoë Holmes. Variational quantum simulation: A case study for understanding warm starts. *PRX Quantum*, 6:010317, Jan 2025.
- [10] Hela Mhiri, Ricard Puig, Sacha Lerch, Manuel S Rudolph, Thiparat Chotibut, Supanut Thanasilp, and Zoë Holmes. A unifying account of warm start guarantees for patches of quantum landscapes. *arXiv preprint arXiv:2502.07889*, 2025.
- [11] Antonio A Mele, Glen B Mbeng, Giuseppe E Santoro, Mario Collura, and Pietro Torta. Avoiding barren plateaus via transferability of smooth solutions in a Hamiltonian variational ansatz. *Physical Review A*, 106(6):L060401, 2022.
- [12] Edward Grant, Leonard Wossnig, Mateusz Ostaszewski, and Marcello Benedetti. An initialization strategy for addressing barren plateaus in parametrized quantum circuits. *Quantum*, 3:214, 2019.
- [13] Tyler Volkoff and Patrick J Coles. Large gradients via correlation in random parameterized quantum circuits. *Quantum Science and Technology*, 6(2):025008, 2021.
- [14] Kaining Zhang, Liu Liu, Min-Hsiu Hsieh, and Dacheng Tao. Escaping from the barren plateau via Gaussian initializations in deep variational quantum circuits. In *Advances in Neural Information Processing Systems*, 2022.
- [15] Chae-Yeun Park and Nathan Killoran. Hamiltonian variational ansatz without barren plateaus. *Quantum*, 8:1239, 2024.
- [16] Dan Shepherd and Michael J Bremner. Temporally unstructured quantum computation. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 465(2105):1413–1439, 2009.
- [17] Michael J Bremner, Richard Jozsa, and Dan 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*, 467(2126):459–472, 2011.
- [18] Michael J Bremner, Ashley Montanaro, and Dan J Shepherd. Average-case complexity versus approximate simulation of commuting quantum computations. *Physical review letters*, 117(8):080501, 2016.
- [19] Dominik Hangleiter and Jens Eisert. Computational advantage of quantum random sampling. *Reviews of Modern Physics*, 95(3):035001, 2023.
- [20] Erik Recio-Armengol, Shahnawaz Ahmed, and Joseph Bowles. Train on classical, deploy on quantum: scaling generative quantum machine learning to a thousand qubits. *arXiv preprint arXiv:2503.02934*, 2025.
- [21] Mario Herrero-Gonzalez, Brian Coyle, Kieran McDowall, Ross Grassie, Sjoerd Beentjes, Ava Khamseh, and Elham Kashefi. The born ultimatum: Conditions for classical surrogation of quantum generative models with correlators. *arXiv preprint arXiv:2511.01845*, 2025.
- [22] Sachin Kasture, Oleksandr Kyriienko, and Vincent E Elfving. Protocols for classically training quantum generative models on probability distributions. *Physical Review A*, 108(4):042406, 2023.
- [23] Andrii Kurkin, Kevin Shen, Susanne Pielawa, Hao Wang, and Vedran Dunjko. Universality and kernel-adaptive training for classically trained, quantum-deployed generative models. *arXiv preprint arXiv:2510.08476*, 2025.
- [24] Manuel S Rudolph, Sacha Lerch, Supanut Thanasilp, Oriel Kiss, Oxana Shaya, Sofia Vallecorsa, Michele Grossi, and Zoë Holmes. Trainability barriers and opportunities in quantum generative modeling. *npj Quantum Information*, 10(1):116, 2024.

- [25] Kevin Shen, Susanne Pielawa, Vedran Dunjko, and Hao Wang. Characterizing trainability of instantaneous quantum polynomial circuit born machines. *arXiv preprint arXiv:2602.11042*, 2026.
- [26] Olli Hirviniemi, Afrad Basheer, and Thomas Cope. Preventing barren plateaus in continuous quantum generative models. *arXiv preprint arXiv:2602.10049*, 2026.