

Transform-Graph Quantum Generative Models with Scalable Parseval Training

Aaron Mark Thomas^{1,2}, Sharu Theresa Jose¹

¹Department of Computer Science, University of Birmingham, United Kingdom

²QunaSys Europe, Copenhagen, Denmark

Motivation. Quantum Circuit Born Machines (QCBMs) form a natural generative model, whereby a parameterised quantum circuit prepares state $|\psi_\theta\rangle$ and samples random samples from the distribution $p_\theta(x) = |\langle x|\psi_\theta\rangle|^2$. Generic Born machines, however, do not provide tractable likelihoods, so scalable training commonly uses implicit objectives such as maximum mean discrepancy (MMD) [1, 2]. For bitstring distributions, common Hamming/RBF MMD kernels impose a fixed Walsh-Hadamard spectral weighting and can underweight high-Hamming-weight components, which are often required for multimodal or strongly correlated targets. Fourier-DQGM methods address this by using a train-on-classical, deploy-on-quantum paradigm: a likelihood-like objective is made tractable by a Parseval hidden-register estimator, and sampling is performed by a transform readout circuit [3]. Yet the original construction fixes a Fourier readout, a visible-hidden bipartite interaction graph, and a fixed finite- K Monte Carlo estimator. We generalise this into

$$\mathcal{M}(T, G, K), \quad T = \text{readout transform}, \quad G = \text{tractable factor graph}, \quad K = \text{hidden-sample budget}. \quad (1)$$

The Fourier-DQGM is recovered as $T = \text{QFT}$, $G = G_{\text{bipartite}}$, and fixed uniform K .

Method. Let n denote the number of visible/data qubits and m the number of hidden qubits. For a transform column $|\tilde{x}_T\rangle = T|x\rangle = \sum_t \chi_T(t; x)|t\rangle$, we train the visible all-zero marginal with hidden qubits traced out. Our first contribution is a transform graph Parseval estimator that rewrites the training marginal as a hidden-register expectation:

$$P_{\theta, T, G}(0^n | x) = \mathbb{E}_{\omega \sim \text{Unif}(\{0,1\}^m)} |g_{\theta, T, G}(\omega; x)|^2, \quad g_{\theta, T, G}(\omega; x) = \sum_{\beta \in \mathcal{B}_T(x)} c_\beta(x) g_{\theta, \beta, G}(\omega; x), \quad (2)$$

Here $\mathcal{B}_T(x)$ is the set of simple components, or “branches”, in a linear decomposition of the transform column. Branch amplitudes are added before taking the modulus square, preserving interference between components. Hartley gives two Fourier-like phase components, Chebyshev gives two phase components plus a constant-row correction, and Haar gives a sparse average-plus-wavelet sum with $n+1$ terms. The original model computes g by product factorisation. We replace the original width-zero product factorisation by bounded-width visible-visible tensor contractions. If the conditioned visible graph has local dimension q and treewidth/pathwidth w , the estimator scales as

$$\text{cost}(\hat{P}_{\theta, T, G}) = O(K B_T n q^{w+1}), \quad \mathbb{E}[-\log \hat{\mu}_K] = -\log \mu + \frac{\rho}{2K} + O(K^{-2}), \quad \rho = \frac{\text{Var}(|g|^2)}{\mu^2}, \quad (3)$$

where $\mu = P_{\theta, T, G}(0^n | x)$ and $\hat{\mu}_K$ is its K -sample Monte Carlo estimate. Thus bounded graph structure controls expressivity, while K controls the leading finite-sample log-likelihood bias.

Preliminary results. We validate the three design choices of $\mathcal{M}(T, G, K)$. *Interaction graph:* a two-layer visible chain reduces Ising-chain KLD from 0.0634 to 0.0063 and high-order Walsh-Hadamard error from 4.16 to 1.35. A two-feature ladder is most useful for cross-feature scale coupling, improving KLD from 0.796 to 0.730 and reducing estimator relative variance from 9.09 to 3.88. *Finite- K estimation:* exact diagnostics show observed log-bias correlates 0.91–0.95 with $\rho/(2K)$. At $m = 12$, fixed- K training degrades on rare-event targets, while adaptive- K and bias-corrected log-loss recover performance. *Transform basis:* transform-specific Parseval estimators match exact statevector likelihoods to machine precision. At $n = 8, m = 4, K = 1000$, Chebyshev has the best estimator conditioning ($\rho = 2.37$ versus Fourier 3.29) and wins endpoint/smooth bounded targets; Haar remains best on step and truncated-heavy-tail targets but loses some representation-space advantage on sparse rare-event targets.

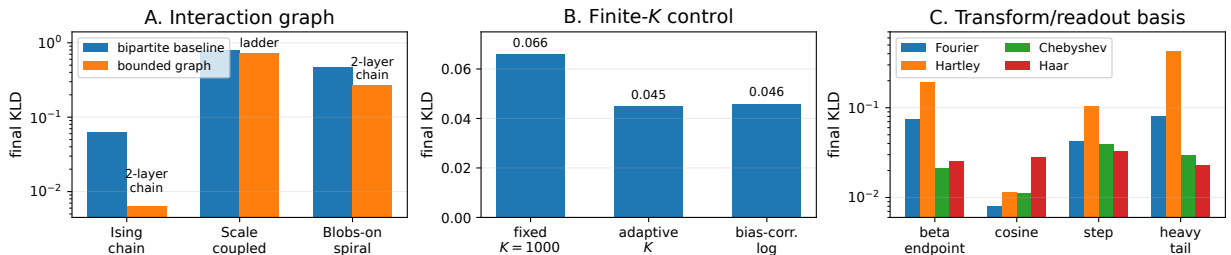


Figure 1: Summary of the three design axes in $\mathcal{M}(T, G, K)$. Left: bounded-width visible contractions improve KLD over the bipartite baseline. Middle: finite- K control improves a rare-event target at $m = 12$. Right: Fourier, Hartley, Chebyshev, and Haar readouts show target-dependent performance.

Contribution and outlook.

Our results support a unified view of transform-graph quantum generative models: Fourier product-factorised Parseval training is one member of a broader class with controllable transform basis, interaction graph, and estimator budget. We provide a classically trainable, quantum-deployable model family with explicit tractability conditions. Ongoing work combines Chebyshev readouts with two-layer chains and adaptive- K , and Haar readouts with bounded-width graphs for sparse multiscale targets. Assessing time and memory scaling as well as performance under varying shot noise budgets.

References

- [1] J.-G. Liu and L. Wang, “Differentiable learning of quantum circuit Born machines,” *Physical Review A*, vol. 98, p. 062324, 2018.
- [2] 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, p. 60, 2020.
- [3] C. Tüysüz, O. Kyriienko, and M. Grossi, “Quantum Fourier Generative Models Trainable at Large Scale,” arXiv:2606.28483, 2026.
- [4] S. Bravyi, D. Gosset, D. Grier, and L. Schaeffer, “Classical algorithms for Forrelation,” arXiv:2102.06963, 2021.
- [5] I. L. Markov and Y. Shi, “Simulating quantum computation by contracting tensor networks,” *SIAM Journal on Computing*, vol. 38, no. 3, pp. 963–981, 2008.
- [6] M. Van den Nest, “Simulating quantum computers with probabilistic methods,” *Quantum Information & Computation*, vol. 11, no. 9–10, pp. 784–812, 2011.
- [7] H.-Y. Wu, V. E. Elfving, and O. Kyriienko, “Multidimensional quantum generative modeling by quantum Hartley transform,” *Advanced Quantum Technologies*, vol. 8, no. 3, p. 2400337, 2025.
- [8] C. A. Williams, A. E. Paine, H.-Y. Wu, V. E. Elfving, and O. Kyriienko, “Quantum Chebyshev transform: mapping, embedding, learning and sampling distributions,” arXiv:2306.17026, 2023.
- [9] A. Fijany and C. P. Williams, “Quantum wavelet transforms: Fast algorithms and complete circuits,” in *NASA International Conference on Quantum Computing and Quantum Communications*, pp. 10–33, Springer, 1998.