

Probabilistic modeling over permutations using quantum computers

Vasilis Belis¹, Giulio Crognaletti^{2,3}, Matteo Argenton^{3,4}, Michele Grossi³, and Maria Schuld¹

¹Xanadu Quantum Technologies, Toronto, ON, M5G 2C8, Canada

²INFN Trieste & University of Trieste, Trieste, Italy

³European Organization for Nuclear Research (CERN), CH-1211 Geneva, Switzerland

⁴INFN Ferrara & University of Ferrara, Ferrara, Italy

Abstract

Quantum computers provide a super-exponential speedup for performing a Fourier transform over the symmetric group, an ability for which practical use cases have remained elusive so far. In this work, we leverage this ability to unlock spectral methods for machine learning over permutation-structured data, which appear in applications such as multi-object tracking and recommendation systems. It has been shown previously that a powerful way of building probabilistic models over permutations is to use the framework of non-Abelian harmonic analysis, as the model’s group Fourier spectrum captures the interaction complexity: “low frequencies” correspond to low order correlations, and “high frequencies” to more complex ones. This can be used to construct a Markov chain model driven by alternating steps of *diffusion* (a group-equivariant convolution) and *conditioning* (a Bayesian update). However, this approach is computationally challenging and hence limited to simple approximations. Here we construct a quantum algorithm that encodes the exact probabilistic model—a classically intractable object—into the amplitudes of a quantum state by making use of the Quantum Fourier Transform (QFT) over the symmetric group. We discuss the scaling, limitations, and practical use of such an approach, which we view as a step towards useful applications of non-Abelian QFTs.

Introduction

Probabilistic modeling over permutations is central to identity management in multi-object tracking and preference learning in recommender systems. These problems are naturally formulated within non-commutative harmonic analysis, where the statistical structure of the data is tied to the symmetric group $\mathbb{S}_n$ [1], the group of all orderings of n objects. The framework’s power lies in the statistical interpretation of the group *Fourier spectrum*: low-frequency components encode single-object statistics (first-order marginals) and simple pairwise relationships, high-frequency ones encode multi-body dependencies [1]. This group Fourier-theoretic angle enables building probabilistic models with favorable learning properties, as pioneered by Kondor [2, 3, 4, 5] and Huang [6, 7]. Classically, however, working with permutations forces strong *band-limiting* to make the computation tractable—keeping only low frequencies, hence capturing only low-order correlations. Even then, modeling up to k -th order marginals costs classical memory and time $\mathcal{O}(n^{2k})$, and toggling between Fourier and direct space via the classical Fast Fourier Transform (FFT) over $\mathbb{S}_n$ in general scales super-exponentially as $\mathcal{O}(n!n^2)$ [8].

Both issues are elegantly resolved on a quantum computer. A quantum state can store and manipulate the $n!$ -dimensional model and its spectrum, and the QFT over $\mathbb{S}_n$ [9, 10, 11] moves between the direct and Fourier space representations in $\mathcal{O}(n^3 \log n)$ [11]. Following a recurrent (Markovian) framework, we formulate a quantum algorithm that—under assumptions we argue as realistic in learning—prepares a state encoding a probabilistic model over $\mathbb{S}_n$ without band-limiting. We encode permutations in $\mathcal{O}(n \log n)$ qubits and implement *diffusion* and *conditioning*, the two operations driving the belief-state evolution (Fig. 1), efficiently.

Results

The Group Fourier Transform (GFT) of $h : G \rightarrow \mathbb{C}$ has matrix-valued coefficients $\hat{h}_\rho = \sum_{x \in G} h(x) \rho(x)$, where ρ is an irreducible representation (irrep) of dimension d_ρ . In $\mathbb{S}_n$, irreps are indexed by partitions of n , denoted $\lambda \vdash n$. If h is a distribution, $\hat{h}_\lambda = \mathbb{E}_{\sigma \sim h}[\rho_\lambda(\sigma)]$ measures how much of the structure encoded by ρ_λ is preserved [1]. We encode the $n!$ permutations with the *Lehmer code*,

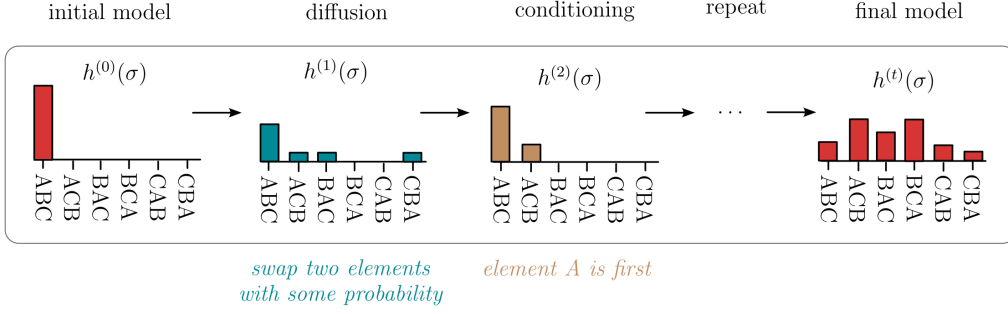


Figure 1: **The probabilistic modeling process.** The belief state over permutations of n objects evolves as $h^{(t)}(\sigma)$, initialized ($t = 0$) from a canonical assignment (shown for $n = 3$ as $ABC \rightarrow [1, 2, 3]$). It alternates *diffusion*, which spreads probability mass to model uncertainty, and *conditioning*, a Bayesian update from partial observations (the training data), repeated per observation.

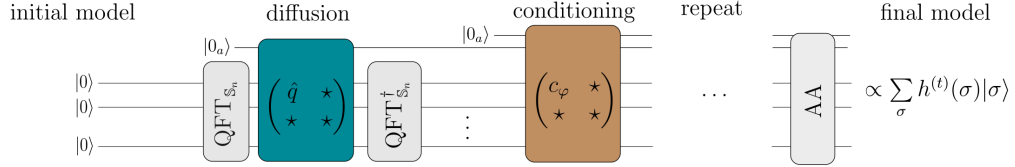


Figure 2: **Our quantum implementation:** The probability distribution is encoded in the amplitudes of the quantum state $|\psi^{(t)}\rangle \propto \sum_{\sigma} h^{(t)}(\sigma) |\sigma\rangle$. We use Lehmer’s code to map permutations $\sigma \in \mathbb{S}_n$ to $n!$ integers, which are then encoded to binary and represented by computational basis states $|\sigma\rangle$. We implement diffusion in Fourier space by block-encoding an operator $\hat{q}$ (see **Diffusion**) and the conditioning step by block-encoding the operator c_{φ} (see **Conditioning**) in direct space. The Quantum Fourier Transform (QFT) over $\mathbb{S}_n$ efficiently toggles between Fourier and direct space. Each block-encoding requires a new ancilla. At the end, Amplitude Amplification boosts the probability of successfully preparing the full posterior probability distribution.

using $\lceil \log_2(n!) \rceil \in \mathcal{O}(n \log n)$ qubits [9, 10, 11]. This factorial-number system enables us to represent permutations by computational basis states $|\sigma\rangle$. The $\mathbb{S}_n$ -QFT is defined as the unitary $\mathcal{F} = \sum_{\sigma, \lambda, i, j} \sqrt{d_{\lambda}/n!} [\rho_{\lambda}(\sigma)]_{ij} |\lambda i j\rangle \langle \sigma|$, which changes the basis from the computational one to the Fourier basis $|\lambda i j\rangle$. The classical FFT scales as $\mathcal{O}(n! n^2)$ [8], whereas the best known quantum implementation runs in $\mathcal{O}(n^3 \log n)$ [11].

Quantum Model. The model is a Markov chain of two alternating operations. *Diffusion* is a group-equivariant convolution (a random walk on the Cayley graph of $\mathbb{S}_n$ generated by transpositions), used to model uncertainty in object-position associations. *Conditioning* is a Bayesian update on new classical data that breaks equivariance. Diffusion is diagonal in Fourier space, conditioning is instead diagonal in direct space, and the QFT toggles between them efficiently. Classically, inference is bottlenecked by memory and the $\mathbb{S}_n$ -FFT, forcing band-limited approximations. We define the quantum model $|\psi^{(t)}\rangle = \frac{1}{\sqrt{N^{(t)}}} \sum_{\sigma} h^{(t)}(\sigma) |\sigma\rangle$ with $N^{(t)} = \|h^{(t)}\|_2^2$, and a natural initial state is the identity delta-peak $|\psi^{(0)}\rangle = |e\rangle \equiv |0\rangle$.

Diffusion. A random walk with diffusion kernel q evolves h by convolution, $h^{(t+1)} = q \star h^{(t)}$, an irrep-wise matrix product $\hat{h}_{\lambda}^{(t+1)} = \hat{q}_{\lambda} \cdot \hat{h}_{\lambda}^{(t)}$ in Fourier space. We take $q(\tau) = p$ for $\tau = e$, $(1-p)/\binom{n}{2}$ for transpositions $\tau \in (2, 1^{n-2})$, and 0 otherwise, with p the identity probability. This kernel is a class function, so by Schur’s lemma $\hat{q}_{\lambda} = c_{\lambda} \mathbb{I}_{\lambda}$ with $c_{\lambda} = p + (1-p)\chi_{\lambda}((2, 1^{n-2}))/d_{\lambda}$, where χ_{λ} is the character on transpositions. Diffusion thus does not mix irreps and acts as an isotropic equivariant convolution; d steps give $\hat{h}_{\lambda}^{(t+d)} = c_{\lambda}^d \hat{h}_{\lambda}^{(t)}$. We efficiently block-encode $\hat{q}^d = \bigoplus_{\lambda} c_{\lambda}^d \mathbb{I}_{\lambda}$ into a unitary D , either using Quantum Read-Only Memory (QROM) or coherent arithmetic depending on the problem size n . The following claims summarize our results on asymptotic scaling guarantees for the model.

Claim 1 (Diffusion success probability). *At $t = 0$, $p_s^{(0)} = \|\hat{q}|\hat{\psi}^{(0)}\rangle\|^2 = p^2 + \frac{2(1-p)^2}{n(n-1)}$ exactly, so standard amplitude amplification reaches success ≈ 1 in $\mathcal{O}(1/\sqrt{p_s})$ iterations. For d steps at arbitrary t : for $p > 1/2$, $p_s^{(t)} \geq (2p-1)^{2d}$ (constant in n), while for rational $p = a/b \in (0, 1)$, $p_s^{(t)} \geq (4/(b^2n^4))^d$ (inverse-polynomial in n).*

Conditioning. Conditioning applies Bayes’ rule in *direct space*, a point-wise product $h^{(t+1)}(\sigma) \propto h^{(t)}(\sigma)h(\varphi|\sigma)$. Here $\varphi : \mathbb{S}_n \rightarrow \{0, 1\}$ flags permutations consistent with an observation, and the likelihood $h(\varphi|\sigma)$ equals s if $\varphi(\sigma) = 1$ and $1-s$ otherwise, with $s \in (0.5, 1]$. This allows tuning trust in the data: $s = 1$, i.e. *hard* likelihood, encodes complete trust in data, and vice-versa for $s \rightarrow 0.5$. We block-encode the diagonal operator $c_\varphi = \sum_\sigma h(\varphi|\sigma)|\sigma\rangle\langle\sigma|$ into a unitary $C(\varphi)$ by computing $\varphi(\sigma)$ coherently, applying a controlled R_y , then uncomputing; $C(\varphi)$ depends only on φ , not on the current model. For common observations—*partial assignments* (fixing positions of $k \ll n$ objects) and *partial rankings* (fixing relative order of $k+1$ objects)—a *reorder-update* approach applies a unitary U_π mapping the relevant Lehmer indices to the string boundary so φ acts locally, then $C(\varphi)$, and finally $U_\pi^\dagger$ restores the canonical ordering. Under a hard likelihood $C(\varphi)$ becomes a projector realized by measurement, needing no extra ancillas.

Claim 2 (Reorder-update cost & conditioning success). *For an order- k constraint, U_π has depth $\mathcal{O}(kn^2 \log n)$ and computing φ at most $\mathcal{O}(k^2 \log^2 n)$. The success probability is $p_s = \|c_\varphi|\psi^{(t)}\rangle\|^2 = h(\varphi)^2 N^{(t+1)}/N^{(t)}$, with $h(\varphi)$ the total probability of consistent permutations. Well-informed models (large $h(\varphi)$) condition with high success, while vanishing success mainly afflicts less useful models.*

After applying T pairs of diffusion and conditioning, using the above results on the respective success probabilities, we use Fixed-Point Amplitude Amplification (FPAA) (Fig. 2) to ensure a coherent (deterministic) computation by boosting the total success probability close to one [12]. Furthermore, we show, without extra assumptions on the data generating process, that asymptotic scalability requires $T \in \mathcal{O}(\log n)$ in the *worst case*. Note that T is *not* the *total circuit depth*, it is the number of diffusion-conditioning pairs. Circuit depth is dominated by the polynomially deep $\mathbb{S}_n$ -QFTs.

Inference and Sampling. We highlight the difference to the classical case: classically the band-limited model serves only low-order-marginal inference and sampling is hard, whereas the quantum state gives *sampling access* in direct and Fourier space, attractive for generative modeling applications. As we use *amplitude encoding*, namely $\psi_\sigma^{(T)} = h^{(T)}(\sigma)/\sqrt{N^{(T)}}$, measurements on the model sample the sharpened $p(\sigma) \propto |h^{(T)}(\sigma)|^2$. Similarly, this approach allows *Fourier sampling* of an irrep label λ with probability $\propto \|\hat{\psi}_\lambda^{(t)}\|_F^2$. While this sharpening introduces distortions, we argue that they can be beneficial for certain task. For instance, we propose *maximum-a-posteriori* estimation (top- k configurations, a hard optimization task classically) by recasting it as sampling task from a sharpened distribution: block-encoding the diagonal amplitude matrix and applying a degree- m polynomial via Quantum Singular Value Transformation [13, 14] maps the state amplitude $\psi_\sigma \mapsto \psi_\sigma^m$, sharpening sampling even more, to $\propto h(\sigma)^{2m}$, thus isolating the modes of the model. One can also estimate marginal MAP for free by measuring part of the register.

Conclusions

We introduced a quantum implementation of an exact probabilistic model over permutations, based on spectral methods on $\mathbb{S}_n$ —a first step towards a classically hard, quantumly feasible, and practically useful QML algorithm.

More broadly, machine learning is a promising, largely unexplored application area for non-Abelian QFTs. We argue that building on Diaconis’ insights [1], quantum states elegantly manipulate a probabilistic model in both direct and Fourier space, with care needed because we manipulate amplitudes, not the Born-rule sampling distribution. This is an instance of a wider theme of spectral methods that can be viewed simultaneously as crucial for designing ML models and natural for quantum computers, since the QFT exposes the model’s Fourier spectrum to the full toolbox of quantum algorithms [15].

References

- [1] Persi Diaconis. Group representations in probability and statistics. *Lecture notes-monograph series*, 11:i–192, 1988.
- [2] Risi Kondor, Andrew Howard, and Tony Jebara. Multi-object tracking with representations of the symmetric group. In Marina Meila and Xiaotong Shen, editors, *Proceedings of the Eleventh International Conference on Artificial Intelligence and Statistics*, volume 2 of *Proceedings of Machine Learning Research*, pages 211–218, San Juan, Puerto Rico, 21–24 Mar 2007. PMLR.
- [3] Risi Kondor. The skew spectrum of functions on finite groups and their homogeneous spaces. *arXiv preprint arXiv:0712.4259*, 2007.
- [4] Risi Kondor. A fourier space algorithm for solving quadratic assignment problems. In *Proceedings of the twenty-first annual ACM-SIAM symposium on Discrete algorithms*, pages 1017–1028. SIAM, 2010.
- [5] Risi Kondor and Walter Dempsey. Multiresolution analysis on the symmetric group. *Advances in Neural Information Processing Systems*, 25, 2012.
- [6] Jonathan Huang, Carlos Guestrin, and Leonidas Guibas. Fourier theoretic probabilistic inference over permutations. *Journal of Machine Learning Research*, 10(37):997–1070, 2009.
- [7] Jonathan Huang, Carlos Guestrin, and Leonidas J Guibas. Efficient inference for distributions on permutations. *Advances in neural information processing systems*, 20, 2007.
- [8] David K. Maslen. The efficient computation of fourier transforms on the symmetric group. *Math. Comput.*, 67(223):1121–1147, July 1998.
- [9] Robert Beals. Quantum computation of fourier transforms over symmetric groups. In *Proceedings of the Twenty-Ninth Annual ACM Symposium on Theory of Computing*, STOC ’97, page 48–53, New York, NY, USA, 1997. Association for Computing Machinery.
- [10] Cristopher Moore, Daniel Rockmore, and Alexander Russell. Generic quantum fourier transforms. *ACM Trans. Algorithms*, 2(4):707–723, October 2006.
- [11] Yasuhito Kawano and Hiroshi Sekigawa. Quantum fourier transform over symmetric groups — improved result. *Journal of Symbolic Computation*, 75:219–243, 2016. Special issue on the conference ISSAC 2014: Symbolic computation and computer algebra.
- [12] Theodore J. Yoder, Guang Hao Low, and Isaac L. Chuang. Fixed-Point Quantum Search with an Optimal Number of Queries. *Phys. Rev. Lett.*, 113(21):210501, 2014.
- [13] András Gilyén, Yuan Su, Guang Hao Low, and Nathan Wiebe. Quantum singular value transformation and beyond: exponential improvements for quantum matrix arithmetics. In *51st Annual ACM SIGACT Symposium on Theory of Computing*, 6 2018.
- [14] John M. Martyn, Zane M. Rossi, Andrew K. Tan, and Isaac L. Chuang. Grand unification of quantum algorithms. *PRX Quantum*, 2(4), December 2021.
- [15] Vasilis Belis, Joseph Bowles, Rishabh Gupta, Evan Peters, and Maria Schuld. Spectral methods: crucial for machine learning, natural for quantum computers?, 2026.