

Towards robust quantum advantage in sequence learning with a single reliable gate

Eric R. Anschuetz and Jonathan Z. Lu

Motivation. The study and construction of quantum algorithms that can efficiently solve problems believed or known to be hard for any classical computer is a cornerstone of quantum information science. The search for quantum advantage is, however, obstructed by significant challenges. First, the most commonly utilized notion of quantum advantage—a quantum algorithm solving a (somewhat useful computational) classical problem exponentially faster than the best (known) classical algorithm—appears to be exceedingly rare. Indeed, aside from Shor’s algorithm [1] and its Hidden Subgroup generalizations [2, 3], as well as certain quantum simulation tasks [4], few convincing exponential quantum speedups are known for classical-input problems. In an attempt to circumnavigate the apparent intractability of finding broader classes of exponential speedups, several modifications to the notion of quantum advantage have been proposed in recent years. These relaxed notions include allowing the input to the problem to be a quantum state [5], looking for a polynomial quantum speedup [6, 7] (typically super-quadratic, since quadratic speedups are relatively generic due to Grover’s algorithm [8]), and measuring advantage in terms of other resources, such as in space [9] or black-box query count [10, 11]. Investigation into each of these relaxations has yielded substantial new insight into the power and limits of quantum computation. At the same time, however, even in a broader setting of quantum advantage, we encounter a second issue: even if a problem makes a convincing case for quantum advantage, the requirements to reliably implement the corresponding quantum algorithm generally lie substantially beyond the capabilities of present-day quantum devices. For example, Shor’s algorithm, despite significant and fruitful circuit optimization efforts in recent years [12–14], requires universal fault-tolerant quantum computation (FTQC) and relatively deep circuits to reliably factor integers on noisy quantum devices. In the limiting case in which *no* fault-tolerant computation is possible on a given device, demonstrations of quantum advantage for computational problems appears unlikely [15–18], and the focus of quantum-classical separations in this setting has centered upon sampling from various distributions of random quantum circuits [19–24]. Yet in recent years, rapid advancements toward an era of early fault-tolerant quantum devices [25–28] suggests that even though universal FTQC on deep circuits remains a significant experimental challenge, quantum algorithms which can be implemented with non-zero but relatively lightweight fault tolerance requirements are within reach. This current state of quantum devices therefore presents a compelling motivation for the design of quantum algorithms that do not require reliable resources as powerful as those of, e.g., factoring, and yet provide some notion of quantum advantage on an interesting computational problem rather than a circuit sampling task. Such a goal forms the central guiding question of this work:

Can we find instances of quantum advantage (as defined in the broader setting above) on classical input-output computational problems, whose fault-tolerance requirements are far from universal FTQC?

Contributions. We propose an affirmative answer to the above question specifically in the setting of *sequence learning* in the encoder-decoder model. More precisely, a sequence learning problem in the encoder-decoder model is specified by a distribution $\mathcal{R}$ over $\mathbb{Z}_2^N \times \mathbb{Z}_2^{N'}$ and a function $f : \mathbb{Z}_2^N \times \mathbb{Z}_2^{N'} \rightarrow \mathbb{Z}_2^m$. The encoder and decoder, which may be bounded in computational power and/or memory, are respectively given $(\mathbf{x}, \mathbf{y}) \sim \mathcal{R}$, and the encoder may pass a n -bit message to the decoder. Then, the decoder must predict $f(\mathbf{x}, \mathbf{y})$ solely from $\mathbf{y}$ and the encoder’s message. Furthermore, if we assume that the encoder and decoder are bounded in space, e.g. both algorithms have $O(n)$ bits of memory and $N, N' \gg n$, then the model is taken to be in the *streaming* setting: one new bit of $\mathbf{x}$ and $\mathbf{y}$ are respectively sent per unit time to the encoder and decoder, who can then update their internal memory accordingly. This setting encompasses a broad landscape of problems and algorithmic frameworks whose utility has surged in recent years from the advent of large language models. Indeed, a quintessential example of an encoder-decoder model is a *transformer*, wherein an encoder receives a streamed long sequence of words $\mathbf{x}$ and processes the sequence into a compact latent representation that is then given to the decoder. At a later time, the decoder receives a *query* $\mathbf{y}$ about the original sequence, e.g. to recall some small chunk of the sequence which is not known ahead of time to the encoder, and the decoder must answer the query using only the compact representation of the sequence given by the encoder. Central to the performance of the model is whether this compact representation is sufficiently expressive for the decoder to accurately answer a query (with high probability over the input/query distributions). Intuitively, unless the representation has roughly the same size as the input sequence $\mathbf{x}$, a classical model should fail to answer certain queries because they depend on parts of the input that were “projected out” in the compactified representation. On the other hand, a quantum state, being much more high-dimensional, could plausibly encode the input sequence succinctly enough so that the answer to a typical given query could be answered with high accuracy by measuring the correct observable. This intuition does not universally apply on arbitrary tasks due to the non-existence of quantum random access codes [29, 30], but with carefully structured distributional tasks such a quantum compactification remains plausible even with highly structured states that require little fault tolerance for reliable preparation.

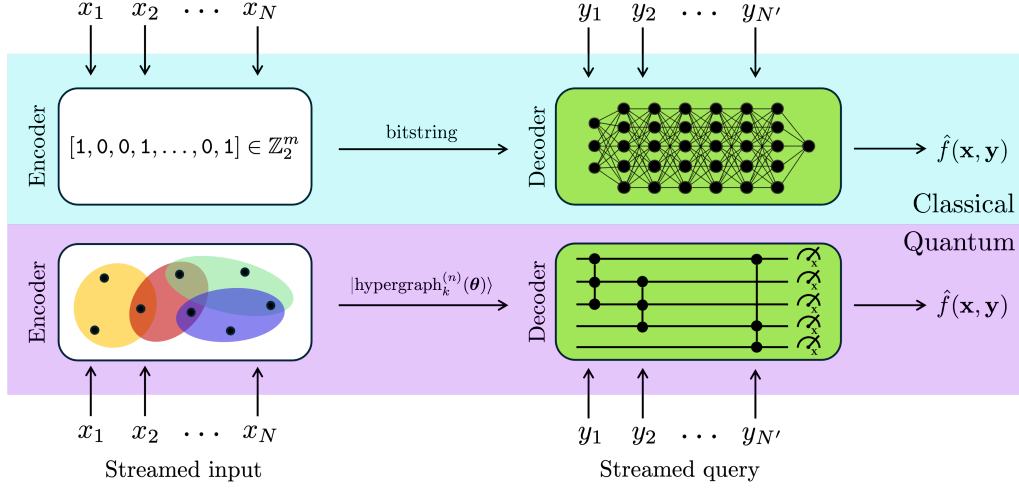


FIG. 1. Comparison of a classical and a hypergraph quantum encoder-decoder model.

To that end, we construct in this work a quantum encoder-decoder model, parameterized by some integer $k \geq 2$. The encoder maintains a n -qubit quantum state $|\psi\rangle$; as each element x_i of the input sequence $\mathbf{x} \in \mathbb{Z}_2^N$ is streamed in, it applies a corresponding gate to the state. After the input stream terminates, the encoder passes $|\psi\rangle$ to the decoder, which measures an observable and outputs the result. We prove that our model satisfies the following properties.

- (1) The model can be trained in time $n^{O(k)}$ classically, using an efficient algebraic representation of the structured family of states our model prepares. Only during inference is a quantum device necessary. The removal of quantum samples during training substantially improves efficiency, as the generation of quantum samples during long training processes is often prohibitively expensive.
- (2) The entire quantum computation uses only a *single* k -qubit quantum gate $\mathbf{U}^{(k)}$ (plus initialization to $|+\rangle^n$ and a final measurement in the $\mathbf{X}$ basis). Thus, relatively minimal fault tolerance is necessary for reliable implementation of the model, e.g. via a quantum code which transversally and addressably implements $\mathbf{U}^{(k)}$.
- (3) There exists explicit distributional sequence learning problems solvable with only n qubits in this model, whereas we provide evidence that any classical model requires representations of size $\Omega(n^k)$.

Whether we use a classical or a quantum model, the total computational time of the training and inference scales as $n^{O(k)}$. The separation is instead in terms of *memory* used by the models, or equivalently in terms of the *communication complexity* between encoder and decoder. This separation may nevertheless yield practical utility if we interpret the sequence learning problem to be an indefinite process over time; this spacetime-advantage-per-token can then be seen as a *power* advantage during inference, assuming that classical models access substantial portions of their memory for each new token. On the choice of k , we note that while larger k increases the potential quantum-classical gap, it also results in a more complicated gate, thereby becoming more challenging for fault-tolerant implementation. k should therefore be as large as possible, subject to the fault-tolerant capabilities of the corresponding quantum device.

Overview. A k -uniform hypergraph on n nodes is a hypergraph in which every hyperedge links k nodes (when $k = 2$, this definition coincides with a standard graph). The states our model prepares are k -uniform hypergraph state on n qubits, specified by a bitstring $\mathbf{g} \in \mathbb{Z}_2^{\binom{n}{k}}$ whose entries are associated with subsets S of $[n] = \{0, \dots, n-1\}$ of size k known as hyperedges. In this convention,

$$|\mathbf{g}\rangle := \prod_{S \subseteq [n]: |S|=k} (\mathbf{C}^{k-1} \mathbf{Z})_S^{\mathbf{g}_S} |+\rangle^{\otimes n}. \quad (1)$$

Circuits which produce this state from $|0^n\rangle$ are special cases of instantaneous quantum poly-time (IQP) circuits, which have been extensively studied in the context of quantum advantage in sampling [25, 31, 32]. Informally, the quantum encoder on each sequence elements x_i computes a corresponding $\mathbf{C}^{k-1} \mathbf{Z}$ gate to apply, preparing a hypergraph state based on the input. Based on the query, the decoder computes an observable—which can be implemented via only $\mathbf{C}^{k-1} \mathbf{Z}$ gates, a full $\mathbf{X}$ -basis measurement, and efficient classical post-processing—to measure as its prediction. By comparison, a corresponding classical model would instead prepare a bitstring that it modifies according to the input sequence, and the decoder would then compute a function of this bitstring and the query for its output (see Fig. 1).

The advantage of hypergraph states is that their circuit complexity can be made arbitrarily large by increasing k , yet they require essentially only *a single gate* to prepare. In addition, they have a succinct algebraic representation via the hypergraph, and thus only have $\text{poly}(n)$ effective dimension rather than 2^n . This low effective dimension enables efficient training via gradient descent over random initializations as compared to a generic quantum model [33]. More precisely, we train the model classically by tracking the hypergraph itself, which uses space $O(n^k)$. Since direct training of a discrete model is challenging, we introduce a relax-and-round trick: during training, we continuously parameterize gates to be of the form $\exp(-i\theta(\mathbf{I} - \mathbf{Z})^{\otimes k})$ (the hypergraph representation becomes a vector of real numbers instead of bits), and round at the end so that all gates become either $\mathbf{I}$ or $\mathbf{C}^{k-1}\mathbf{Z}$. These techniques underpin properties (1) and (2) from above.

For property (3), we next construct a distributional problem, which we call the *distributed polynomial evaluation* problem parameterized by an integer k , that we believe has a quantum-classical separation in the encoder-decoder setting. The separation we conjecture is entirely in terms of one-way communication complexity from encoder to decoder, and therefore can be studied independently of the details of the encoder/decoder implementations. For each k , the task is to evaluate a degree- k homogeneous polynomial $P : \mathbb{Z}_2^n \rightarrow \mathbb{Z}_2$ on n variables $x_1, \dots, x_n$ over $\mathbb{Z}_2$ at a particular point $\mathbf{x}$, where both P and $\mathbf{x}$ are chosen from some explicit distribution; the encoder is given a description of the polynomial, sends some latent representation of the polynomial to the decoder, and the decoder is given the point $\mathbf{x}$ on which to evaluate the polynomial as well as certain classical advice about the polynomial. We show that the hypergraph quantum encoder-decoder model solves this task to perfect accuracy for *any distribution* when the memory of the model is $\Omega(n)$. While we are unable to prove a generic classical lower bound, we prove the existence of an explicit distribution wherein linear classical strategies require $\Omega(n^k)$ memory to achieve accuracy $1/2 + \Omega(1)$.

While communication complexity offers a clean theoretical abstraction for proving separations, models in practice also require efficient training which can pose a further obstruction to natural classical models, such as recurrent neural network (RNN) encoder-decoders. We find numerically, for example, that even for simple problems such as deciding if $\mathbf{x}, \mathbf{y} \in \mathbb{Z}_2^{\binom{n}{k}}$ are equal or differ in a small number of places (which by hashing strategies requires only $O(1)$ bits of communication), our quantum model trains quickly to near-perfect accuracy, while RNN-based classical models fail to train even with memory $\gg n$ and warm-start tricks. This separation suggests that in practice, hypergraph quantum models may be more broadly useful for generic sequence learning tasks beyond carefully designed problems optimized for rigorous separations.

Related work. The construction of quantum models which admit fully classical training has been explored in the context of making generative models that use IQP circuits to sample from a trained distribution [34]. Such models are prone, however, to data acquisition obstacles: assuming that the desired quantum circuit is not already known, if the learned distribution is classically hard to sample from, then either the training dataset is itself hard to generate or the learned distribution does not match the training data distribution. Whether a relaxation of polynomial sampling advantage exists in this model remains open. Related to our abstracted communication complexity separation, a result of Kerenidis and Raz [9], and independently Gavinsky, Kempe, and de Wolf [35], proves an exponential quantum-classical communication complexity separation for the *boolean hidden matching* problem. The quantum protocol, however, uses an unstructured quantum phase state whose preparation is likely challenging for early fault-tolerant devices. Similarly, recent work demonstrates an exponential space separation for quantum versus classical encoder-decoder models in sequence learning problems [36], demonstrating the significant potential for quantum advantage broadly across sequence learning tasks. As in our case, the training time in the quantum setting scales as the required classical space complexity, with separation primarily in the size of the latent data. Our model can therefore be seen as a proposal for a smoothed interpolation between fully classical and fully quantum models: the conjectured separation is polynomial, but the training time is also polynomial. Moreover, the minimal fault-tolerance requirements of hypergraph models suggest that they may serve as a useful intermediary realization of quantum-classical separations on the path towards exponential separations and universal FTQC. Finally, polynomial separations in learning models have been proposed in the continuous-variable quantum computation setting [37], giving further evidence for the promise of advantage in sequence learning. The separation in [37] is, however, in terms of achieving a finite KL divergence loss, which does not necessarily capture real-world performance. This is because a model can achieve prediction accuracy arbitrarily close to 1, and yet still have infinite KL divergence due to a difference in distributional support. Our proposal can be seen as an explicit qubit-model version of [37], using a single gate and with a conjectured separation in terms of *prediction accuracy* rather than a particular loss function.

Outlook. A key open question is the development of techniques to prove our conjectured classical lower bound. Moreover, our model captures a broad family of learning tasks which we believe to contain many problems with quantum-classical space separations. Can we find more explicit distributional sequence learning tasks with plausible or provable quantum advantage? Finally, our model is designed around k -uniform hypergraph states for their algebraic structure. Understanding if distinct families of states with similarly efficient classical representability may yield quantum-classical separations on broader families of computational and learning problems.

-
- [1] P. W. Shor, Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer, *SIAM review* **41**, 303 (1999).
 - [2] M. Ettinger, P. Høyer, and E. Knill, The quantum query complexity of the hidden subgroup problem is polynomial, *Information Processing Letters* **91**, 43 (2004).
 - [3] A. Y. Kitaev, Quantum measurements and the abelian stabilizer problem, arXiv preprint quant-ph/9511026 (1995).
 - [4] A. J. Daley, I. Bloch, C. Kokail, S. Flannigan, N. Pearson, M. Troyer, and P. Zoller, Practical quantum advantage in quantum simulation, *Nature* **607**, 667 (2022).
 - [5] H.-Y. Huang, M. Broughton, J. Cotler, S. Chen, J. Li, M. Mohseni, H. Neven, R. Babbush, R. Kueng, J. Preskill, *et al.*, Quantum advantage in learning from experiments, *Science* **376**, 1182 (2022).
 - [6] A. Schmidhuber, R. O'Donnell, R. Kothari, and R. Babbush, Quartic quantum speedups for planted inference, *Physical Review X* **15**, 021077 (2025).
 - [7] A. Schmidhuber and A. Zlokapa, Quartic quantum speedups for community detection, arXiv preprint arXiv:2510.08494 (2025).
 - [8] L. K. Grover, A fast quantum mechanical algorithm for database search, in *Proceedings of the Twenty-Eighth Annual ACM Symposium on Theory of Computing*, STOC '96 (Association for Computing Machinery, New York, NY, USA, 1996) pp. 212–219.
 - [9] I. Kerenidis and R. Raz, The one-way communication complexity of the boolean hidden matching problem, arXiv preprint quant-ph/0607173 (2006).
 - [10] D. R. Simon, On the power of quantum computation, *SIAM journal on computing* **26**, 1474 (1997).
 - [11] S. Aaronson, Open problems related to quantum query complexity, *ACM Transactions on Quantum Computing* **2**, 1 (2021).
 - [12] M. Ekerå and J. Gärtner, A high-level comparison of state-of-the-art quantum algorithms for breaking asymmetric cryptography, arXiv preprint arXiv:2405.14381 (2024).
 - [13] C. Gidney, How to factor 2048 bit rsa integers with less than a million noisy qubits, arXiv preprint arXiv:2505.15917 (2025).
 - [14] M. Cain, Q. Xu, R. King, L. R. Picard, H. Levine, M. Endres, J. Preskill, H.-Y. Huang, and D. Bluvstein, Shor's algorithm is possible with as few as 10,000 reconfigurable atomic qubits, arXiv preprint arXiv:2603.28627 (2026).
 - [15] X. Gao and L. Duan, Efficient classical simulation of noisy quantum computation, arXiv preprint arXiv:1810.03176 (2018).
 - [16] T. Schuster, C. Yin, X. Gao, and N. Y. Yao, A polynomial-time classical algorithm for noisy quantum circuits, *Physical Review X* **15**, 041018 (2025).
 - [17] E. Fontana, M. S. Rudolph, R. Duncan, I. Rungger, and C. Cîrstoiu, Classical simulations of noisy variational quantum circuits, *npj Quantum Information* **11**, 84 (2025).
 - [18] S. Cichy, P. K. Faehrmann, L. Bittel, J. Eisert, and H. Pashayan, Classical simulation of noisy quantum circuits via locally entanglement-optimal unravelings, arXiv preprint arXiv:2508.05745 (2025).
 - [19] F. Arute, K. Arya, R. Babbush, D. Bacon, J. C. Bardin, R. Barends, R. Biswas, S. Boixo, F. G. Brandao, D. A. Buell, *et al.*, Quantum supremacy using a programmable superconducting processor, *nature* **574**, 505 (2019).
 - [20] M. DeCross, R. Haghshenas, M. Liu, E. Rinaldi, J. Gray, Y. Alexeev, C. H. Baldwin, J. P. Bartolotta, M. Bohn, E. Chertkov, *et al.*, Computational power of random quantum circuits in arbitrary geometries, *Physical Review X* **15**, 021052 (2025).
 - [21] M. A. Broome, A. Fedrizzi, S. Rahimi-Keshari, J. Dove, S. Aaronson, T. C. Ralph, and A. G. White, Photonic boson sampling in a tunable circuit, *Science* **339**, 794 (2013).
 - [22] J. B. Spring, B. J. Metcalf, P. C. Humphreys, W. S. Kolthammer, X.-M. Jin, M. Barbieri, A. Datta, N. Thomas-Peter, N. K. Langford, D. Kundys, *et al.*, Boson sampling on a photonic chip, *Science* **339**, 798 (2013).
 - [23] M. Tillmann, B. Dakić, R. Heilmann, S. Nolte, A. Szameit, and P. Walther, Experimental boson sampling, *Nature photonics* **7**, 540 (2013).
 - [24] A. Crespi, R. Osellame, R. Ramponi, D. J. Brod, E. F. Galvao, N. Spagnolo, C. Vitelli, E. Maiorino, P. Mataloni, and F. Sciarrino, Integrated multimode interferometers with arbitrary designs for photonic boson sampling, *Nature photonics* **7**, 545 (2013).
 - [25] D. Bluvstein, S. J. Evered, A. A. Geim, S. H. Li, H. Zhou, T. Manovitz, S. Ebadi, M. Cain, M. Kalinowski, D. Hangleiter, *et al.*, Logical quantum processor based on reconfigurable atom arrays, *Nature* **626**, 58 (2024).
 - [26] Quantum error correction below the surface code threshold, *Nature* **638**, 920 (2025).
 - [27] S. Bravyi, D. Gosset, R. Koenig, and M. Tomamichel, Quantum advantage with noisy shallow circuits, *Nat. Phys.* **16**, 1040 (2020).
 - [28] C. Zhao, C. Duckering, A. Gu, N. Maskara, and H. Zhou, Towards ultra-high-rate quantum error correction with reconfigurable atom arrays, arXiv preprint arXiv:2604.16209 (2026).
 - [29] A. S. Holevo, Bounds for the quantity of information transmitted by a quantum communication channel, *Problemy Peredachi Informatsii* **9**, 3 (1973).
 - [30] A. Nayak, Optimal lower bounds for quantum automata and random access codes, in *40th Annual Symposium on Foundations of Computer Science (Cat. No. 99CB37039)* (IEEE, 1999) pp. 369–376.
 - [31] 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* **467**, 459 (2011).
 - [32] M. J. Bremner, A. Montanaro, and D. J. Shepherd, Average-case complexity versus approximate simulation of commuting

- quantum computations, *Physical review letters* **117**, 080501 (2016).
- [33] E. R. Anschuetz, A unified theory of quantum neural network loss landscapes, in *International Conference on Learning Representations*, edited by Y. Yue, A. Garg, N. Peng, F. Sha, and R. Yu (OpenReview, 2025).
 - [34] E. Recio-Armengol, S. Ahmed, and J. Bowles, Train on classical, deploy on quantum: scaling generative quantum machine learning to a thousand qubits, arXiv preprint arXiv:2503.02934 (2025).
 - [35] D. Gavinsky, J. Kempe, and R. de Wolf, Exponential separation of quantum and classical one-way communication complexity for a boolean function, arXiv preprint quant-ph/0607174 (2006).
 - [36] H. Zhao, A. Zlokapa, H. Neven, R. Babbush, J. Preskill, J. R. McClean, and H.-Y. Huang, Exponential quantum advantage in processing massive classical data, arXiv preprint arXiv:2604.07639 (2026).
 - [37] E. R. Anschuetz and X. Gao, Arbitrary polynomial separations in trainable quantum machine learning, *Quantum* **10**, 1976 (2026).