Enhancing Generative Models via Quantum Correlations

Eric R. Anschuetz

Institute for Quantum Information and Matter, Caltech

Work with Xun Gao, Shengtao Wang, J. Ignacio Cirac, and Mikhail D. Lukin

Phys. Rev. X **12**, 021037

(and a bit of PRX Quantum **4**, 020338)

Quantum Computing

- Quantum mechanical systems of size n described by state vector of dimension $\exp(n)$, classical systems $\sim n$
- Simulating dynamics of generic quantum systems probably difficult for traditional computers

Multireference Nature of Chemistry: The Coupled-Cluster View

Dmitry I. Lyakh,* Monika Musiał, Victor F. Lotrich, and Rodney J. Bartlett

Quantum Theory Project, University of Florida, Gainesville, Florida 32611, United States

Møller–Plesset perturbation theory: from small molecule methods to methods for thousands of atoms

Dieter Cremer*

Quantum Computing

- Quantum system dynamics can also solve (certain) hard "classical" problems!
- Dynamics can be digitized: *quantum computation*

Algorithms for Quantum Computation: Discrete Logarithms and Factoring

Peter W. Shor

A fast quantum mechanical algorithm for database search

Lov K. Grover

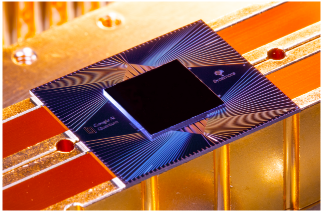
PRL **103,** 150502 (2009) PHYSICAL REVIEW LETTERS week ending 9 OCTOBER 2009

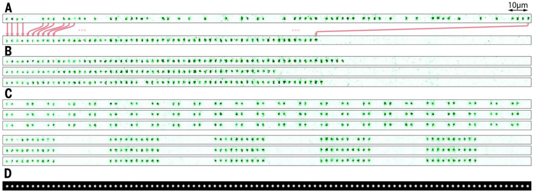
Quantum Algorithm for Linear Systems of Equations

Aram W. Harrow,[1] Avinatan Hassidim,[2] and Seth Lloyd[3]

Quantum Computing Today

- Today: have quantum devices sampling from complex distributions[1]





- One day: quantum generative models?

[1] F. Arute et al., Nature **574**, 505; M. Endres et al., Science **354**, 1024.

Quantum Machine Learning (QML)

- Many quantum generative models have been shown to be more expressive than their classical counterparts:
  - Quantum Boltzmann machines vs. classical restricted Boltzmann machines[2]...
  - Quantum GANs vs. GANs[3]...
  - ...etc.
- Proof strategy?

[2] N. Wiebe et al., arXiv:1902.05162 [quant-ph].
[3] S. Lloyd and C. Weedbrook, Phys. Rev. Lett. **121**, 040502.

Quantum Machine Learning (QML)

- Typical proof of separation:
  1. QML models can factor large numbers

Quantum Machine Learning (QML)

- Typical proof of separation:
    1. QML models can factor large numbers
    2. Factoring large numbers is probably hard using traditional computers

Quantum Machine Learning (QML)

- Typical proof of separation:
  1. QML models can factor large numbers
  2. Factoring large numbers is probably hard using traditional computers
  3. QED :)

Quantum Machine Learning (QML)

- Typical proof of separation:
  1. QML models can factor large numbers
  2. Factoring large numbers is probably hard using traditional computers
  3. QED :)
- What data should we use QML models for?

Quantum Machine Learning (QML)

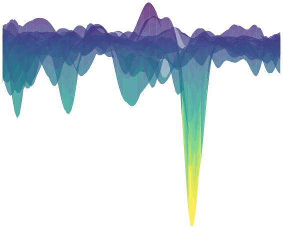
- Typical proof of separation:
  1. QML models can factor large numbers
  2. Factoring large numbers is probably hard using traditional computers
  3. QED :)
- What data should we use QML models for? ¯_(ツ)_/¯

Quantum Machine Learning (QML)

- Goal: concrete, constructive proofs of expressivity separations between traditional and quantum(-inspired) generative models
- Interpretability gives:
  - Intuition where separation holds
  - Intuition how to construct better ML models
  - Intuition on trainability...

Quantum Machine Learning (QML)

- Goal: concrete, constructive proofs of expressivity separations between traditional and quantum(-inspired) generative models
- Interpretability gives:
  - Intuition where separation holds
  - Intuition how to construct better ML models
  - Intuition on trainability...
- (...more on this later)



We show...

1. There exist distributions easy for matrix product states to sample from, but difficult for certain Bayesian networks (superpolynomial memory separation)
2. This separation is due to (simulating) a type of quantum mechanical correlation (quantum contextuality)
3. Extensions to neural networks as well!

Outline

Outline

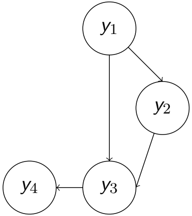
Bayesian Networks

- “Machine learning before machine learning was cool”
- Directed acyclic graphical models that encode structure in underlying probability distribution
- Depending on graph structure, fewer parameters needed to describe model
- Depending on graph structure, efficiently trainable

- Example ($y_i \in \{0, \ldots, d-1\}$):

$$p\left(y_1, y_2, y_3, y_4\right) = p_1\left(y_1\right) p_2\left(y_2 \mid y_1\right) p_3\left(y_3 \mid y_1, y_2\right) p_4\left(y_4 \mid y_3\right) \tag{1}$$

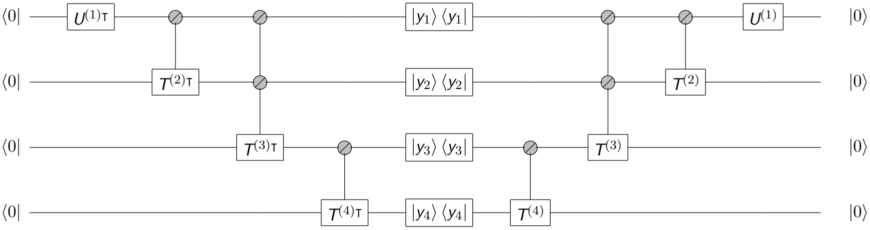
- Corresponding graphical model:



Bayesian Tensor Networks

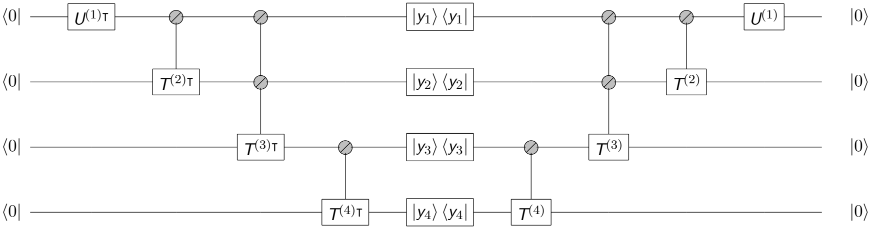
- Rewrite as tensor network with nonnegative entries:

$p(y_1, y_2, y_3, y_4) =$



- Rewrite as tensor network with nonnegative entries:

$p(y_1, y_2, y_3, y_4) =$

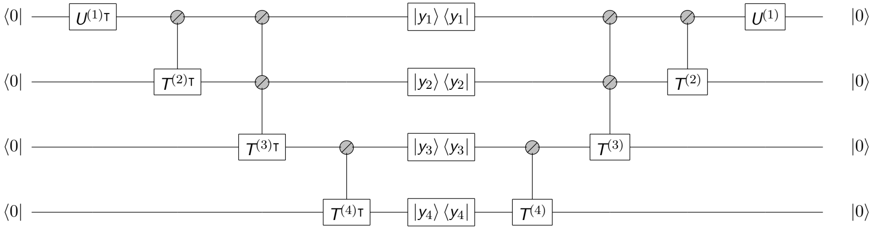


- $\langle 0| = (1, 0, \ldots, 0),\ |0\rangle = \langle 0|^\intercal$

Bayesian Tensor Networks

- Rewrite as tensor network with nonnegative entries:

$p\,(y_1, y_2, y_3, y_4) =$

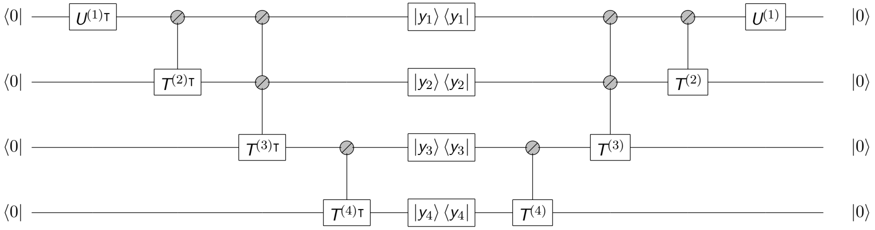


- $\langle 0| = (1, 0, \ldots, 0),\ |0\rangle = \langle 0|^\intercal$
- $|a\rangle\langle a| = \big(1 \text{ at } (a, a)\big)$ ($d \times d$ projector)

Bayesian Tensor Networks

- Rewrite as tensor network with nonnegative entries:

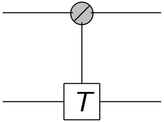
$p\left(y_1, y_2, y_3, y_4\right) =$



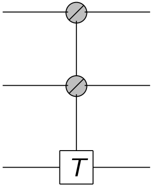
- $\langle 0| = (1, 0, \ldots, 0)$, $|0\rangle = \langle 0|^\intercal$
- $|a\rangle\langle a| = \big(1 \text{ at } (a, a)\big)$ ($d \times d$ projector)
- $(a, 0)$ entry of $U^{(1)}$: $\sqrt{p_1\left(a\right)}$ ($d \times d$ orthogonal)

Bayesian Tensor Networks

- *Uniformly controlled gates*



$$= \sum_{a=0}^{d-1} |a\rangle \langle a| \otimes T_a$$

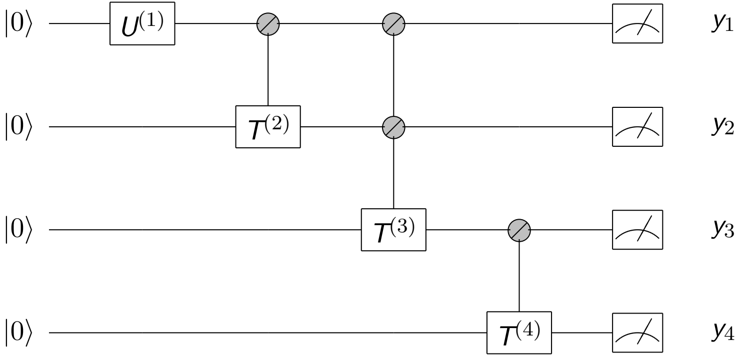


$$= \sum_{a,b=0}^{d-1} |a\rangle \langle a| \otimes |b\rangle \langle b| \otimes T_{a,b}$$

- $(c, 0)$ entry of $T^{(i)}_{a,b}$: $\sqrt{p_i\,(c \mid a, b)}$ ($d \times d$ orthogonal)

Bayesian Quantum Circuits[4]

- Throughout talk: will write in "quantum circuit notation"



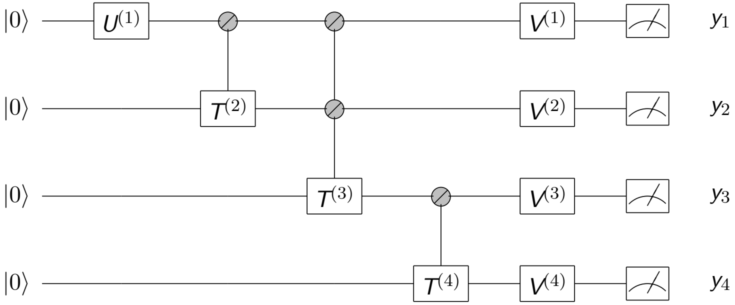
[4] G. H. Low et al., Phys. Rev. A **89**, 062315.

Bayesian Quantum Circuits

- Generally, a quantum circuit is a Bayesian quantum circuit iff:
  1. It is composed of single-node unitaries and uniformly controlled gates, where there is one target node for each gate (directed graph)
  2. There are no unitaries after control units on a node (acyclic)

Basis-Enhanced Bayesian Networks

- "Minimal quantum extension" of Bayesian quantum circuits
- Example:



- $V^{(i)}$ unitary, *only acting on a single node*

Basis-Enhanced Bayesian Quantum Circuits

- Generally, a quantum circuit is a basis-enhanced Bayesian quantum circuit iff:
  1. It is composed of single-node unitaries and uniformly controlled gates, where there is one target node for each gate (directed graph)
  2. There are no target unitaries after control units on a node (acyclic) (i.e., local noncomputational basis measurements are allowed)

Outline

Hidden Markov Models for Translation

- Given input sequence $\boldsymbol{x}$, give any correct translation $\boldsymbol{y}$

Hidden Markov Models for Translation

- Given input sequence $\boldsymbol{x}$, give any correct translation $\boldsymbol{y}$
- Example:

$$\text{Me llamo Eric.} \quad \longrightarrow \quad \begin{matrix} \text{My name is Eric.} \\ \\ \text{I call myself Eric.} \end{matrix}$$

Hidden Markov Models for Translation

- Given input sequence $\boldsymbol{x}$, give any correct translation $\boldsymbol{y}$
- Example:

$$\text{Me llamo Eric.} \quad \longrightarrow \quad \begin{array}{l} \text{My name is Eric.} \\ \\ \text{I call myself Eric.} \end{array}$$

Hidden Markov Models for Translation

- Given input sequence $\boldsymbol{x}$, give any correct translation $\boldsymbol{y}$
- Example:

$$\text{Me llamo Eric.} \longrightarrow \begin{array}{l} \text{My name is Eric.} \\ \\ \text{I call myself Eric.} \end{array}$$

- Want "finite KL divergence" with uniform distribution over translations:

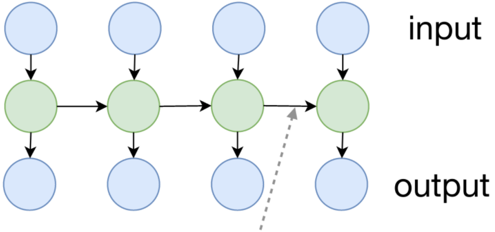
$$p_{\text{model}}(\boldsymbol{y} \mid \boldsymbol{x}) \neq 0 \implies p(\boldsymbol{y} \mid \boldsymbol{x}) \neq 0 \tag{2}$$

Hidden Markov Models (HMMs)

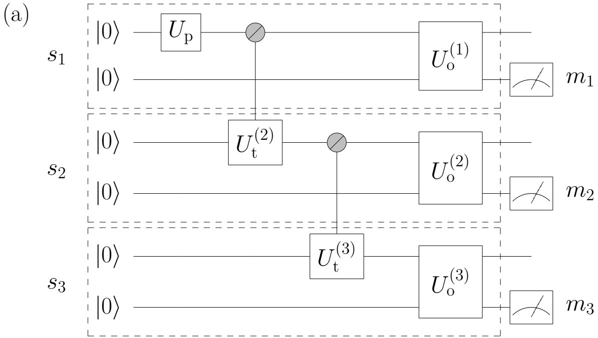
- *HMM*: assume latent variable λ,

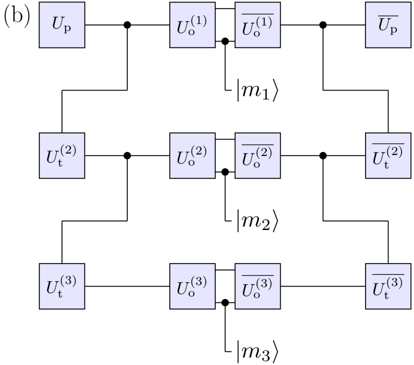
$$p\left(\lambda_t \mid \boldsymbol{x}, \boldsymbol{y}, \hat{\boldsymbol{\lambda}}_t\right) = p\left(\lambda_t \mid x_t, \lambda_{t-1}\right), \tag{3}$$

$$p\left(y_t \mid \boldsymbol{x}, \hat{\boldsymbol{y}}_t, \boldsymbol{\lambda}\right) = p\left(y_t \mid \lambda_t\right) \tag{4}$$



Basis-Enhanced HMM





- Note: basis-enhanced HMM has an efficient MPS representation!

Advantage in Basis-Enhanced Hidden Markov Models

Theorem

There exists a family of basis-enhanced HMMs with M states per time step that, to be approximated to finite KL divergence by a classical hidden Markov model, requires $M^{(\log(M))}$ hidden states per time step.

- There exist distributions with efficient MPS representations (bond dimension M^2) but no efficient HMM representation!

Outline

Quantum Contextuality

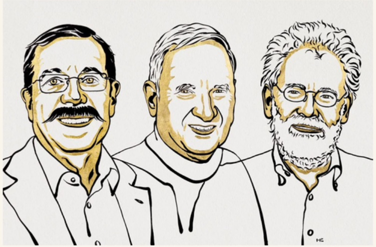
- Properties of a quantum mechanical system have no definite value
- Nobel Prize in Physics awarded for experimental demonstration of this just last year!

Nobel Prize in Physics

The 2022 physics laureates

The Nobel Prize in Physics 2022 was awarded to Alain Aspect, John F. Clauser and Anton Zeilinger “for experiments with entangled photons, establishing the violation of Bell inequalities and pioneering quantum information science”.

Their results have cleared the way for new technology based upon quantum information.



Ill. Niklas Elmehed © Nobel Prize Outreach

Example of Quantum Contextuality

- Quantum mechanics: can construct "quantum variables" (q-numbers) such that:

$$\begin{array}{ccccccc}
q_{11} & \times & q_{12} & \times & q_{13} & = & +1 \\
\times & & \times & & \times & & \\
q_{21} & \times & q_{22} & \times & q_{23} & = & +1 \\
\times & & \times & & \times & & \\
q_{31} & \times & q_{32} & \times & q_{33} & = & +1 \\
\shortparallel & & \shortparallel & & \shortparallel & & \\
+1 & & +1 & & -1 & &
\end{array}$$

- Can classical variable assignments do this?

Example of Quantum Contextuality

- Classical attempt:

$$
\begin{array}{ccccccc}
1 & \times & 1 & \times & 1 & = & +1 \\
\times & & \times & & \times & & \\
1 & \times & 1 & \times & 1 & = & +1 \\
\times & & \times & & \times & & \\
1 & \times & 1 & \times & 1 & = & +1 \\
\shortparallel & & \shortparallel & & \shortparallel & & \\
+1 & & +1 & & \color{red}{+1} & &
\end{array}
$$

Example of Quantum Contextuality

- Classical attempt:

$$\begin{array}{ccccccc}
1 & \times & 1 & \times & 1 & = & +1 \\
\times & & \times & & \times & & \\
1 & \times & 1 & \times & 1 & = & +1 \\
\times & & \times & & \times & & \\
1 & \times & 1 & \times & \color{red}{-1} & = & \color{red}{-1} \\
\shortparallel & & \shortparallel & & \shortparallel & & \\
+1 & & +1 & & -1 & &
\end{array}$$

Example of Quantum Contextuality

- Classical attempt:

$$\begin{array}{ccccccc}
1 & \times & 1 & \times & 1 & = & +1 \\
\times & & \times & & \times & & \\
1 & \times & 1 & \times & 1 & = & +1 \\
\times & & \times & & \times & & \\
1 & \times & \color{red}{-1} & \times & \color{red}{-1} & = & +1 \\
\shortparallel & & \shortparallel & & \shortparallel & & \\
+1 & & \color{red}{-1} & & -1 & &
\end{array}$$

Example of Quantum Contextuality

- Classical attempt:

$$\begin{array}{ccccccc}
1 & \times & 1 & \times & 1 & = & +1 \\
\times & & \times & & \times & & \\
1 & \times & 1 & \times & 1 & = & +1 \\
\times & & \times & & \times & & \\
1 & \times & \color{red}{-1} & \times & \color{red}{-1} & = & +1 \\
\shortparallel & & \shortparallel & & \shortparallel & & \\
+1 & & \color{red}{-1} & & -1 & &
\end{array}$$

- *No classical assignment possible!*

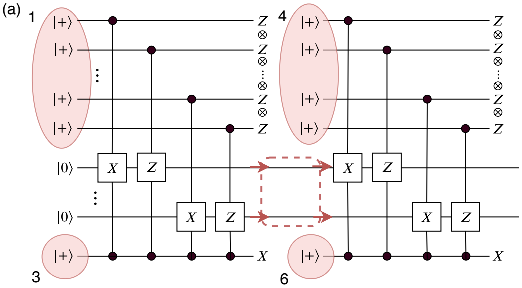
Example of Quantum Contextuality

$$
\begin{array}{ccccccc}
1 & \times & 1 & \times & 1 & = & +1 \\
\times & & \times & & \times & & \\
1 & \times & 1 & \times & 1 & = & +1 \\
\times & & \times & & \times & & \\
1 & \times & 1 & \times & q_{33} & = & +1 \\
\shortparallel & & \shortparallel & & \shortparallel & & \\
+1 & & +1 & & -1 & &
\end{array}
$$

- Value of q_{33} depends on if accessed with variables in row or in column
- Quantum mechanics allows *context-dependent* values for variables with no memory overhead!

Sketch of Basis-Enhanced Bayesian Advantage Proof

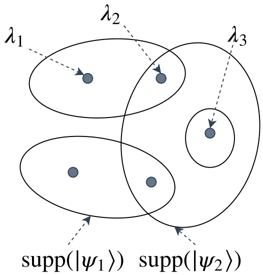
- Idea: construct translation task where:
  - $\boldsymbol{x}$ describes q-numbers
  - $\boldsymbol{y}$ are values of q-numbers when measured sequentially in a quantum mechanical system of size V
- Basis-enhanced HMM (MPS) with bond dimension $\sim M$ can simulate these measurements when $V \sim \log(M)$ ("phase estimation")



Sketch of Basis-Enhanced Bayesian Advantage Proof

- Classically: require a hidden state in HMM to represent any given context
- Required memory H using classical HMM:

$$H \geq \# \text{ of contexts} \geq \frac{\# \text{ inputs}}{\# \text{ number of inputs per context}} \tag{5}$$



Sketch of Basis-Enhanced Bayesian Advantage Proof

- $\exists$ a quantum system of size V with q-numbers such that:

$$\#\text{ inputs} \sim 2^{V^2/2} \tag{6}$$

$$\#\text{ number of inputs per context} \sim 2^{V^2/4} \tag{7}$$

- $\implies H \gtrsim 2^{V^2/4} \gtrsim M^{\log(M)}$ $\square$

Sketch of Basis-Enhanced Bayesian Advantage Proof

- $\exists$ a quantum system of size V with q-numbers such that:

$$\# \text{ inputs} \sim 2^{V^2/2} \tag{6}$$

$$\# \text{ number of inputs per context} \sim 2^{V^2/4} \tag{7}$$

- $\implies H \gtrsim 2^{V^2/4} \gtrsim M^{\log(M)}$ $\square$
- We show:
  - Consider inputs describing the sequential measurement of V commuting V-qubit Pauli operators ($\sim 2^{V^2/2}$)
  - Show every $\sim 2^{V^2/4}$ have at least 9 comprising a Mermin–Peres magic square

Takeaways

- Quantum(-inspired) models can efficiently store long-range correlations through quantum contexts
- Search for practical separations in data with long-range correlations
- Is this true for other quantum(-inspired) models?

Takeaways

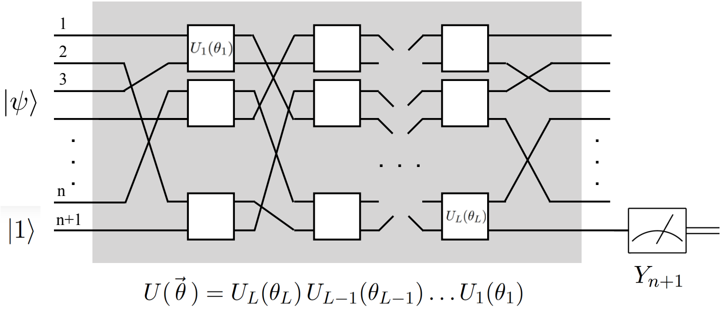
- Quantum(-inspired) models can efficiently store long-range correlations through quantum contexts
- Search for practical separations in data with long-range correlations
- Is this true for other quantum(-inspired) models?
- (Yes)

Interpretable Quantum Advantage in Neural Sequence Learning

Eric R. Anschuetz,[1,*] Hong-Ye Hu,[2,3,4] Jin-Long Huang,[2] and Xun Gao[4,†]

Outline

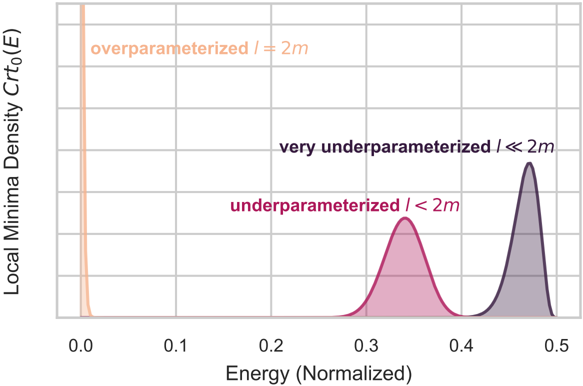
- Why does interpretability matter?
- *Quantum neural networks*[5]:



[5] E. Farhi and H. Neven, arXiv:1802.06002 [quant-ph].

Why Does Interpretability Matter?

- Generically: QNNs *untrainable*[6]
  - Expressivity $\leftrightarrow$ Trainability
- Solution: want *minimal* quantum neural network achieving a quantum advantage



[6] **ERA**, ICLR 2022; **ERA** and B. T. Kiani, Nat. Commun. **13**, 7760.

Extensions to Neural Networks

- Construct "minimal" quantum neural network exhibiting contextuality $\implies$

Theorem (Neural network expressivity separation[7], informal)

Classical neural networks of memory less than* $\frac{n(n-3)}{2}^{\dagger}$ *cannot accurately perform a certain translation task that a trainable quantum RNN of size n can perfectly perform.*

- Separation tight: $\sim n^2$-size quantum-inspired classical model can achieve this

[7] **ERA** et al., PRX Quantum **4**, 020338.

Extensions to Neural Networks

- Construct "minimal" quantum neural network exhibiting contextuality $\implies$

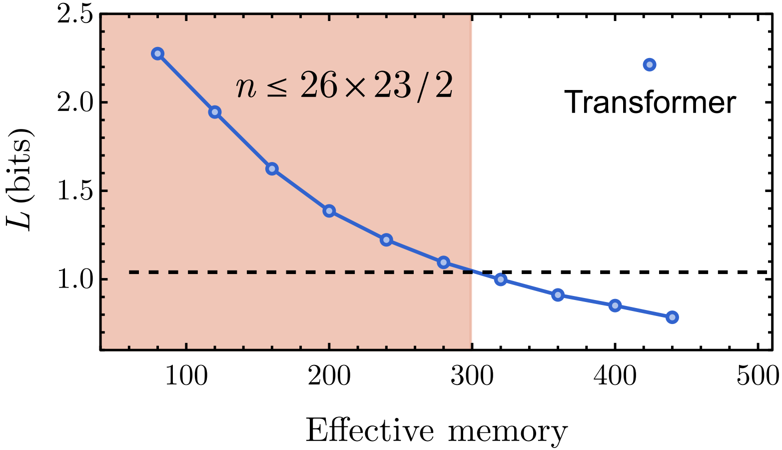
Theorem (Neural network expressivity separation[7], informal)

Classical neural networks of memory less than* $\frac{n(n-3)}{2}^{\dagger}$ *cannot accurately perform a certain translation task that a trainable quantum RNN of size n can perfectly perform.*

- Separation tight: $\sim n^2$-size quantum-inspired classical model can achieve this
- *Includes: RNNs, LSTMs, GRUs, Transformers, ...
- †In progress: $\sim n^k$ separation

[7] **ERA** et al., PRX Quantum **4**, 020338.

Simulations on Real-World Translation Tasks



Future Directions

- Ways to *a priori* evaluate data to see if amenable to quantum(-inspired) representation?
- How ubiquitous is separation on sequence data with long-range correlations?
- Do our results give a useful quantum-inspired classical neural network?
- How amenable are these architectures to experimental implementation?

Questions?

Thank you!