Arbitrary Polynomial Separations in Trainable Quantum Machine Learning

Eric R. Anschuetz

eans@caltech.edu

November 25, 2024

arXiv:2402.08606 [quant-ph]

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

How can one efficiently model sequential data?

How can one efficiently model sequential data?

- Translation. . .

How can one efficiently model sequential data?

- Translation. . .
- Time series modeling. . .

Sequence Modeling

How can one efficiently model sequential data?

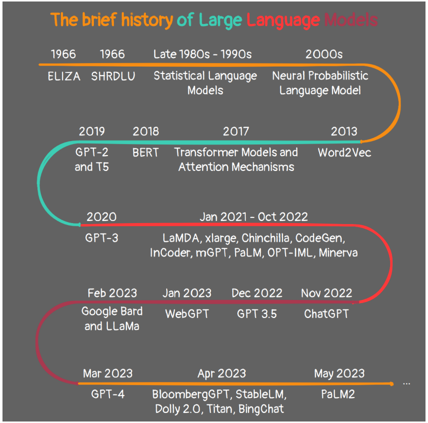
- Translation...
- Time series modeling...
- Chatbots!

Write a limerick about the status of ChatGPT.

ChatGPT is surely the best
But its servers are put to the test
With so many users chatting
It's no wonder they're lagging
But they'll fix it soon, no need to fret!

Large Language Models

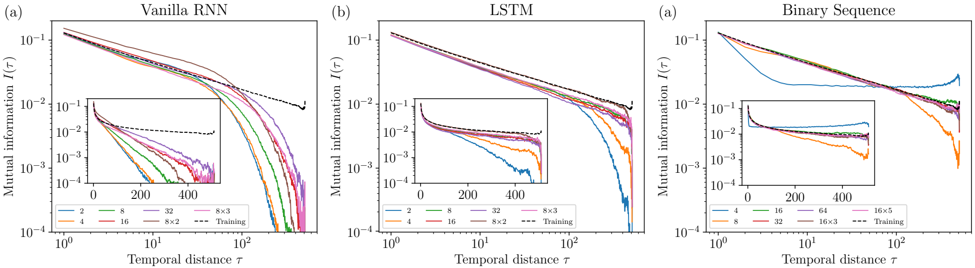
- Recent breakthroughs using *large language models*



A. Norouzi, Level Up Coding (2023)

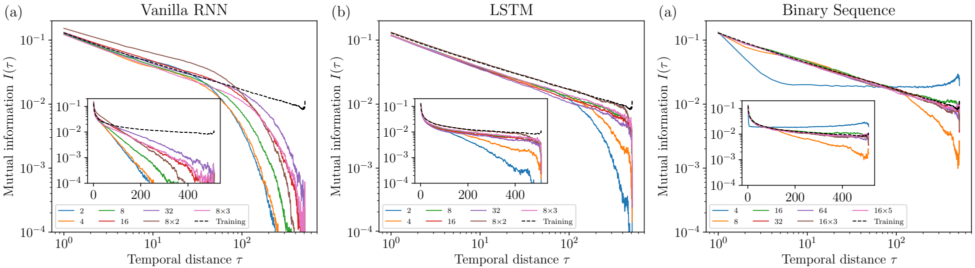
Large Language Models

- Driving factor: efficient representability of *long-range correlations*



H. Shen, arXiv:1905.04271 [cs.LG]

- Driving factor: efficient representability of *long-range correlations*



Can quantum models represent certain long-range correlations efficiently?

H. Shen, arXiv:1905.04271 [cs.LG]

Outline

(The Problem With) Quantum Neural Networks

How Can Quantum Contextuality Help?

Numerical Simulations

Outline

(The Problem With) Quantum Neural Networks

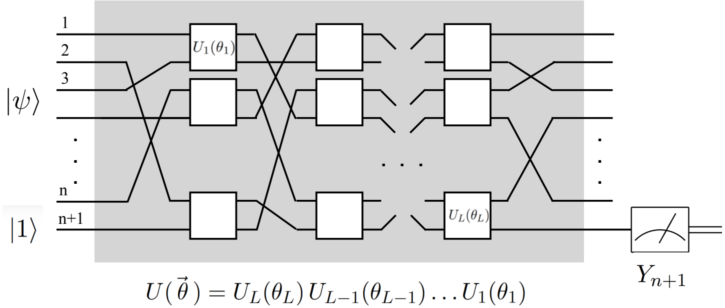
How Can Quantum Contextuality Help?

Numerical Simulations

Quantum Machine Learning (QML)

- *Quantum neural networks*
- Broadly: parameterize quantum circuit $U(\boldsymbol{\theta})$, optimize loss function

$$f(\boldsymbol{\theta}) = \mathrm{Tr}\left(U(\boldsymbol{\theta})\,\rho U(\boldsymbol{\theta})^{\dagger}\,O\right)$$



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

Quantum Neural Networks

Powerful?	Trainable?

Quantum Neural Networks

Powerful?	**Trainable?**
Yes![1]	

[1] Y. Liu et al., Nat. Phys. **17**, 1013 (2021); E. Gil-Fuster et al., arXiv:2406.07072 [quant-ph].

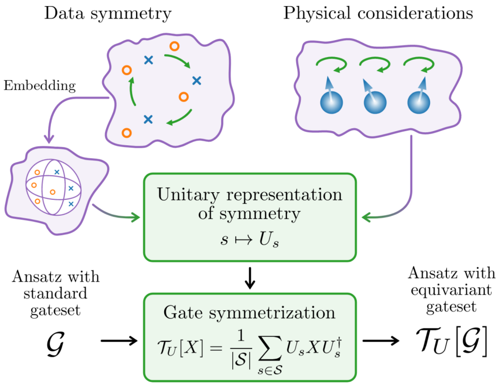
Quantum Neural Networks

Powerful?	**Trainable?**
Yes![1]	...No :(

[1] Y. Liu et al., Nat. Phys. **17**, 1013 (2021); E. Gil-Fuster et al., arXiv:2406.07072 [quant-ph].

Sidestepping Untrainability

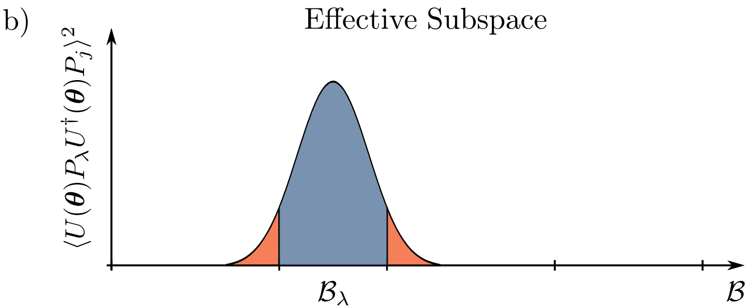
- Generally: *exploring an exponentially large Hilbert space is hard*
- What about *polynomially-sized subspaces*?



J. J. Meyer et al., PRX Quantum **4**, 010328 (2023)

Sidestepping Untrainability

- Unfortunately, quantum advantage less obvious due to constrained Hilbert space



- Few cases where there *is* an advantage seem very specific

M. Cerezo et al., arXiv:2312.09121 [quant-ph]

Wishlist

Can we balance:

- Efficient trainability?
- “Large” quantum-classical separation?
- Physical intuition?
- Constructive proofs?

Wishlist

Can we balance:

- Efficient trainability?
- "Large" quantum-classical separation?
- Physical intuition?
- Constructive proofs?

Yes!

Arbitrary Polynomial Separations in Trainable Quantum Machine Learning

Eric R. Anschuetz*[1] and Xun Gao[2]

Outline

Sequence Modeling: Translation Tasks

- *Translation task*: sample from $p(\boldsymbol{y} \mid \boldsymbol{x})$ to finite relative entropy
- Example:

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

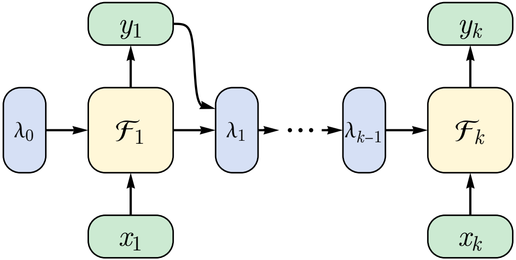
Sequence Modeling: Translation Tasks

- *Translation task*: sample from $p(\boldsymbol{y} \mid \boldsymbol{x})$ to finite relative entropy
- Example:

$$\text{Me llamo Eric.} \quad \longrightarrow \quad \begin{array}{l} \text{My name is Eric.} \\[2em] \text{I call myself Eric.} \end{array}$$

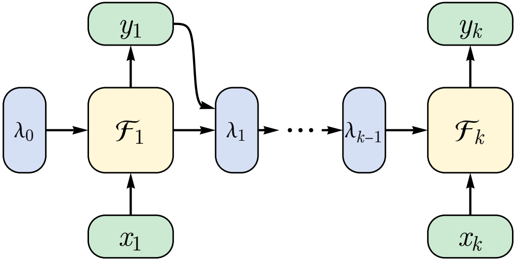
Autoregressive Models for Sequence Modeling

Autoregressive sequence model: RNNs, LSTMs, Transformer decoders, ...



Autoregressive Models for Sequence Modeling

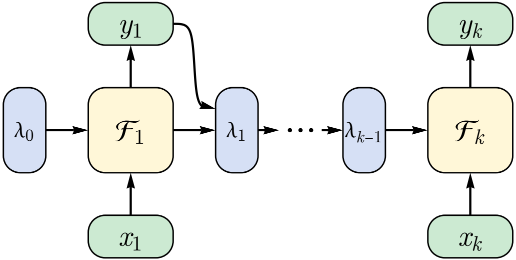
Autoregressive sequence model: RNNs, LSTMs, Transformer decoders, ...



- Modeling of long-time correlations via "memory" $\boldsymbol{\lambda}_i$

Autoregressive Models for Sequence Modeling

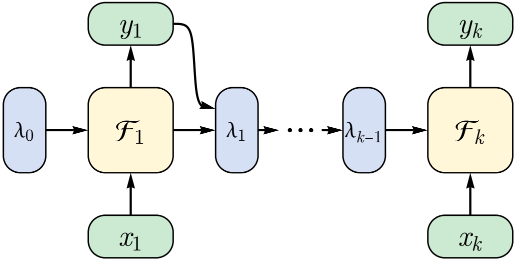
Autoregressive sequence model: RNNs, LSTMs, Transformer decoders, ...



- Modeling of long-time correlations via "memory" $\boldsymbol{\lambda}_i$
- Measure of "separation": classical vs. quantum memory needed to represent data

Autoregressive Models for Sequence Modeling

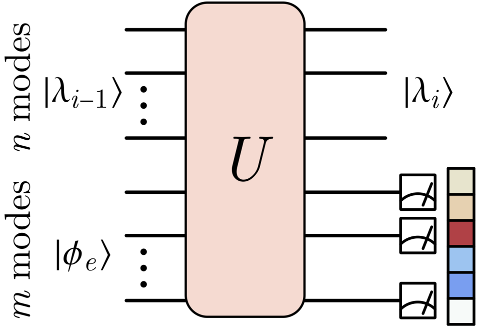
Autoregressive sequence model: RNNs, LSTMs, Transformer decoders, ...



- Modeling of long-time correlations via "memory" $\boldsymbol{\lambda}_i$
- Measure of "separation": classical vs. quantum memory needed to represent data
- Bonus: want $\mathcal{F}_i$ to be low-depth (online learning setting)

k-Hypergraph Recurrent Neural Networks

- Consider: $\mathcal{F}_i$ degree-$(k-1)$ polynomials in $\boldsymbol{\lambda}_{i-1}$
- Quantize dynamics: *k-HRNN*

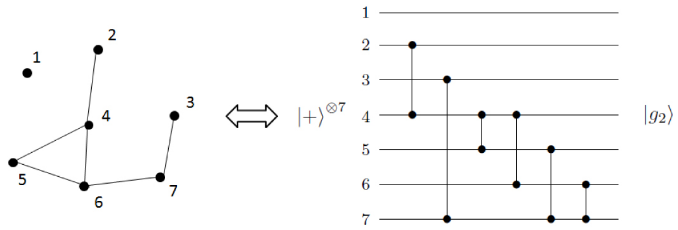


- Turns out: equivalent to sequentially measuring *hypergraph state stabilizers*

k-Uniform Hypergraph States

Graph state associated with $G = (V, E)$:

$$|\psi\rangle = \prod_{e \in E} CZ_e |+\rangle^{\otimes n}$$

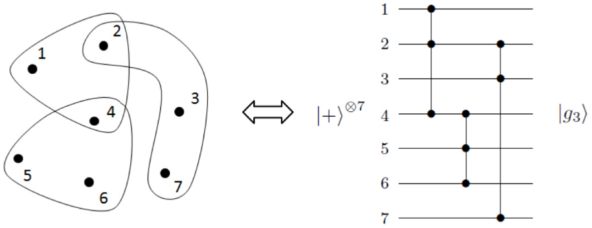


M. Rossi et al., New J. Phys. **15**, 113022 (2013)

k-Uniform Hypergraph States

Hypergraph state associated with $G = (V, E)$:

$$|\psi\rangle = \prod_{e \in E} \mathrm{C}^{k-1} Z_e \, |+\rangle^{\otimes n}$$

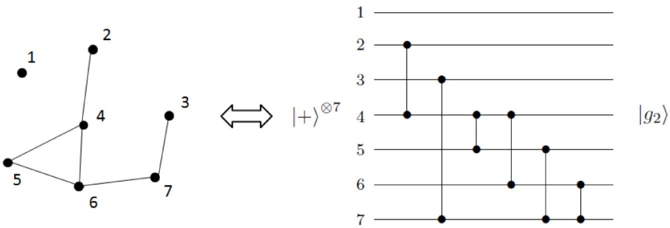


M. Rossi et al., New J. Phys. **15**, 113022 (2013)

k-Uniform Hypergraph States

Graph state stabilizers:

$$s_v = X_v \prod_{v' \in E \perp v} Z_{v'}$$

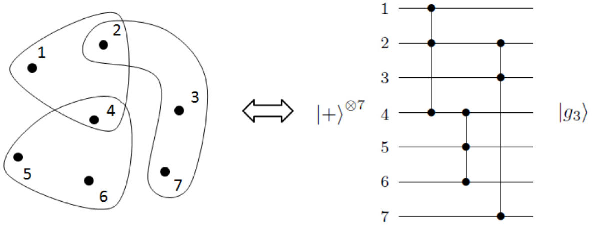


M. Rossi et al., New J. Phys. **15**, 113022 (2013)

k-Uniform Hypergraph States

Hypergraph state stabilizers:

$$s_v = X_v \prod_{v' \in E \perp v} C^{k-2} Z_{v'}$$

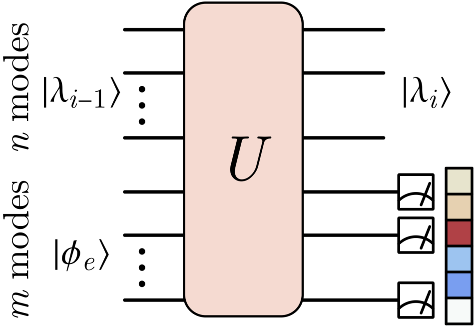


M. Rossi et al., New J. Phys. **15**, 113022 (2013)

k-Hypergraph Recurrent Neural Networks

Efficiency:

- CV: dynamics under poly (n)-sized Lie algebra $\implies$ efficient trainability![2]
- Each stabilizer measurement can be done in $\log(n)$ depth

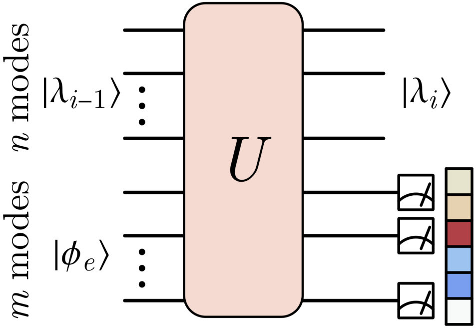


[2] **ERA**, arXiv:2408.11901 [quant-ph].

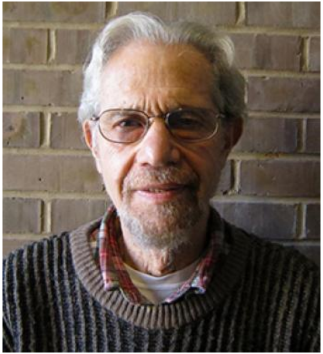
k-Hypergraph Recurrent Neural Networks

Expressivity:

- Stabilizers are extremely *contextual*



Measuring observables $\neq$ "revealing" hidden classical assignments!



S. Kochen and E. Specker, Indiana Univ. Math. J. **17**, 59 (1968)

$k = 2$ example:

$$\begin{array}{ccccccc} X_1 & \cdot & X_2 & \cdot & X_1X_2 & = & +1 \\ \cdot & & \cdot & & \cdot & & \\ X_1Z_2 & \cdot & Z_1X_2 & \cdot & Y_1Y_2 & = & +1 \\ \cdot & & \cdot & & \cdot & & \\ Z_2 & \cdot & Z_1 & \cdot & Z_1Z_2 & = & +1 \\ \shortparallel & & \shortparallel & & \shortparallel & & \\ +1 & & +1 & & -1 & & \end{array}$$

$k = 2$ example:

$$\begin{array}{ccccccc} X_1 & \cdot & X_2 & \cdot & X_1X_2 & = & +1 \\ \cdot & & \cdot & & \cdot & & \\ X_1Z_2 & \cdot & Z_1X_2 & \cdot & Y_1Y_2 & = & +1 \\ \cdot & & \cdot & & \cdot & & \\ Z_2 & \cdot & Z_1 & \cdot & Z_1Z_2 & = & +1 \\ \parallel & & \parallel & & \parallel & & \\ +1 & & +1 & & -1 & & \end{array}$$

- Can classical variable assignments do this?

Example of Quantum Contextuality

Classical attempt:

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

Example of Quantum Contextuality

Classical attempt:

$$\begin{array}{ccccccc} 1 & \cdot & 1 & \cdot & 1 & = & +1 \\ \cdot & & \cdot & & \cdot & & \\ 1 & \cdot & 1 & \cdot & 1 & = & +1 \\ \cdot & & \cdot & & \cdot & & \\ 1 & \cdot & 1 & \cdot & -1 & = & -1 \\ \shortparallel & & \shortparallel & & \shortparallel & & \\ +1 & & +1 & & -1 & & \end{array}$$

Example of Quantum Contextuality

Classical attempt:

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

Example of Quantum Contextuality

Classical attempt:

$$\begin{array}{ccccccc} 1 & \cdot & 1 & \cdot & 1 & = & +1 \\ \cdot & & \cdot & & \cdot & & \\ 1 & \cdot & 1 & \cdot & 1 & = & +1 \\ \cdot & & \cdot & & \cdot & & \\ 1 & \cdot & -1 & \cdot & -1 & = & +1 \\ \shortparallel & & \shortparallel & & \shortparallel & & \\ +1 & & -1 & & -1 & & \end{array}$$

- *No classical assignment possible!*

Example of Quantum Contextuality

$$
\begin{array}{ccccccc}
1 & \cdot & 1 & \cdot & 1 & = & +1 \\
\cdot & & \cdot & & \cdot & & \\
1 & \cdot & 1 & \cdot & 1 & = & +1 \\
\cdot & & \cdot & & \cdot & & \\
1 & \cdot & 1 & \cdot & Z_1 Z_2 & = & +1 \\
\parallel & & \parallel & & \parallel & & \\
+1 & & +1 & & -1 & &
\end{array}
$$

- "Value" of $Z_1 Z_2$ depends if measured with observables in row or in column
- Quantum mechanics allows *context-dependent* values for observable measurements
- Classical simulation: forced to memorize measurement context!

Provable Expressivity Separation

- Input sequence $\boldsymbol{x}$: $\mathrm{O}(n)$ hypergraph state stabilizers to sequentially measure
- Output sequence $\boldsymbol{y}$: measurement outcomes consistent with quantum mechanics

Provable Expressivity Separation

- Input sequence $\boldsymbol{x}$: $\mathrm{O}(n)$ hypergraph state stabilizers to sequentially measure
- Output sequence $\boldsymbol{y}$: measurement outcomes consistent with quantum mechanics

Theorem (HRNN expressivity separation, informal)

Classical neural networks of width less than $\binom{n}{k} - 1$ cannot perform this task to any finite relative entropy.

- Alternatively: n vs. $\binom{n}{k}$, n-party communication complexity separation where the quantum parties are depth $\mathrm{O}(\log(n))$

Outline

(The Problem With) Quantum Neural Networks

How Can Quantum Contextuality Help?

Numerical Simulations

Simulations on Real-World Translation Tasks

Can view contextuality as a type of correlation present in empirical data which:

- Inhibits efficient classical representations
- Does *not* inhibit efficient quantum representations

S. Abramsky and A. Brandenburger, New J. Phys. **13**, 113036 (2011)

Simulations on Real-World Translation Tasks

Experimental test on realistic data set:

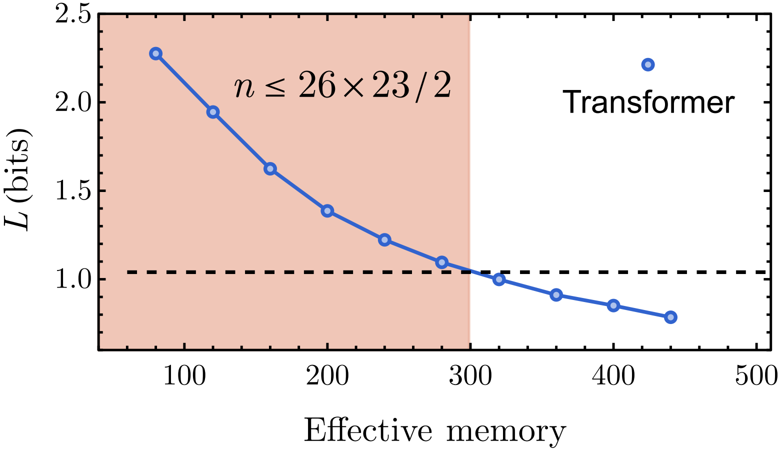
- Spanish-to-English translation
- ≈500,000 model parameters
- $k = 2$

ERA et al., PRX Quantum **4**, 020338 (2023)

Simulations on Real-World Translation Tasks

Input	"Debemos limpiar la cocina."	
Truth	"We must clean up the kitchen."	
CRNN	"We must clean the kitchen."	✓
GRU	"We have to turn the right address."	✗
Input	"Admití que estaba equivocada."	
Truth	"I admitted that I was wrong."	
CRNN	"I was wrong to say that."	~
GRU	"They had a thing to be true."	✗
Input	"¿Cual es el lugar más bonito del mundo?"	
Truth	"What's the most beautiful place in the world?"	
CRNN	"What's the world largest place?"	✗
GRU	"What's the best of is in?"	✗
Input	"La caja es pesada."	
Truth	"The box is heavy."	
CRNN	"The box is heavy."	✓
GRU	"My box is."	✗

Simulations on Real-World Translation Tasks



Future Directions

- Ways to *a priori* evaluate data to see if amenable to quantum representation?
- Do our results give a useful quantum-inspired classical model?
- How amenable are these architectures to early forms of error-correction/mitigation and experimental implementation?

Questions?

Thank you!