No-go theorem of hidden variable theory and quantum machine learning

Xun Gao
JILA and Department of Physics at CU Boulder

FOCS Quantum Learning Workshop
Oct 27, 2024

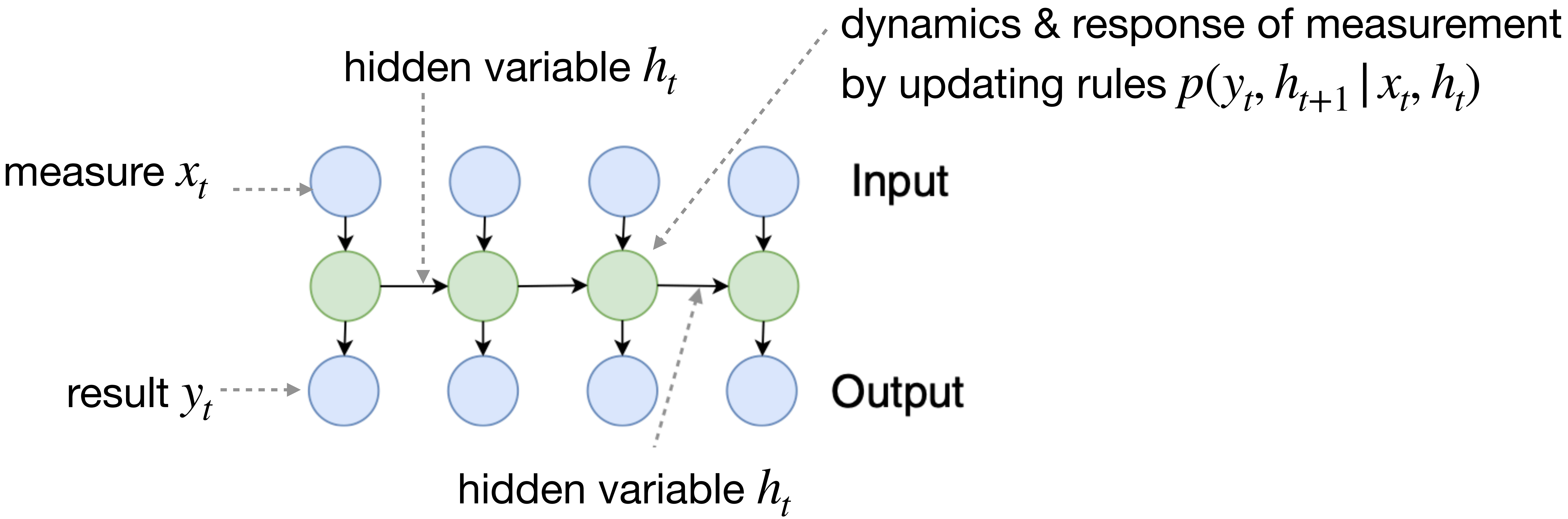
Outline

- Quantum contextuality and no-go theorem of hidden variable theory
- Applications of quantum contextuality:
  - Expressive power separation between quantum and classical neural networks
  - Performance on real-world data
- Outlook
  - Solid foundations? Sheaf cohomology?
  - Relation with Non-negative matrix factorization and communication complexity
  - Experimental challenge and other approaches

Quantum Contextuality & No-go theorem of hidden variable theory

Hidden variable theory

- What are hidden variable models? Just hidden Markov models



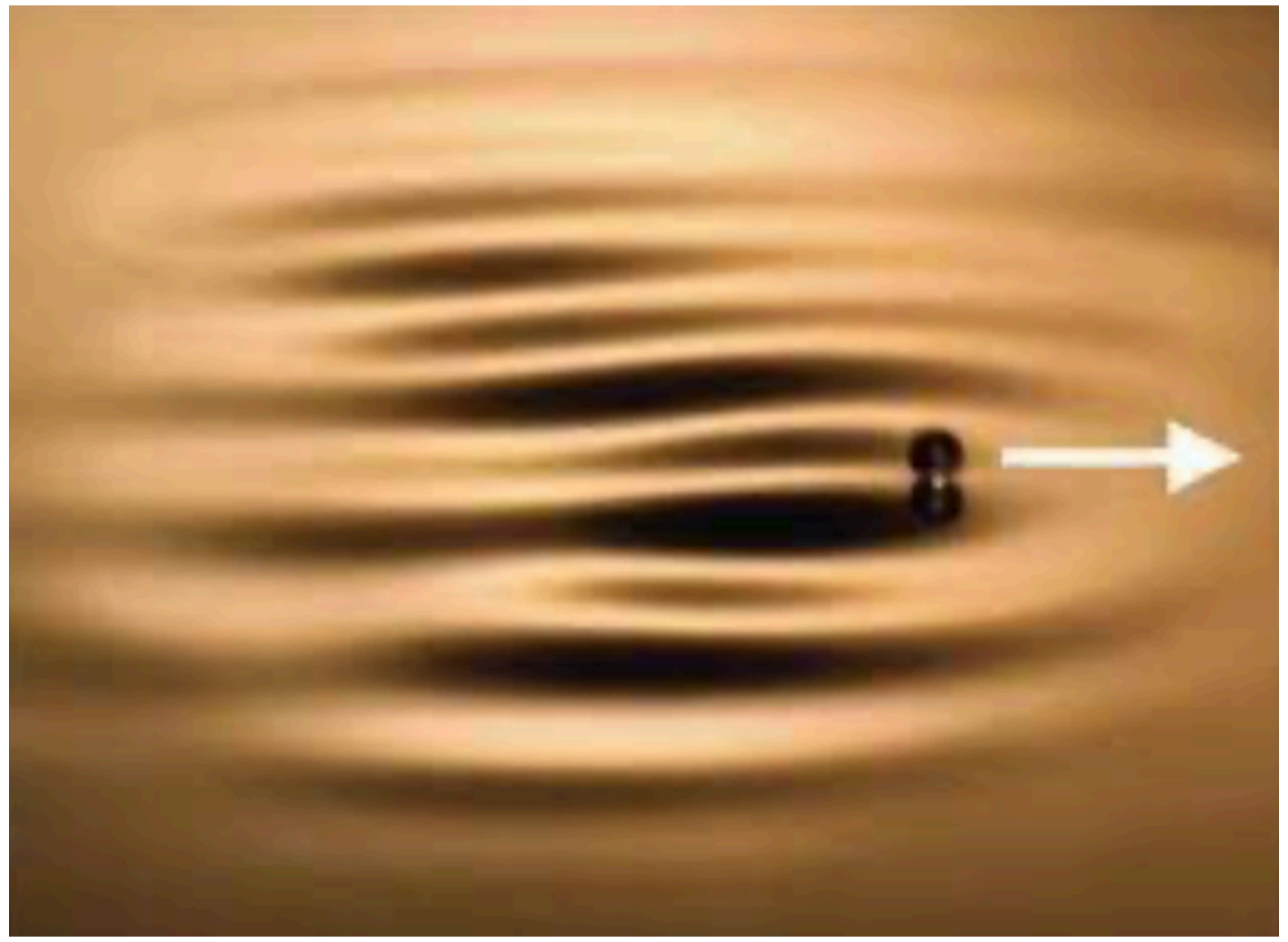
time direction

$$p(\cdots y_t \cdots | \cdots x_t \cdots) = \sum_{\cdots h_t \cdots} \cdots p(y_{t-1}, h_t | x_{t-1}, h_{t-1}) \cdot p(y_t, h_{t+1} | x_t, h_t) \cdots$$

Hidden variable theory

- Quantum mechanics described by hidden variable models?

  We don't know. Bohm's mechanics (hydrodynamics-like equation, however, non-local, contextual)
  Even more extremely, h_t is the full description of the quantum state or the whole history



- No-go theorem of hidden variable theory?

  Need further constraints
  e.g., locality, non-contextuality, bounded memory ($\dim h_t$ is limited)

Non-contextual hidden variable theory

Bell-Kochen-Specker theorem

- non-contextual hidden variable theory

I. contexts (commuting sets of observables):
$\{A, B, C\}$ and $\{A, L, M\}$ (B, C not commute with L, M)

II. well-defined joint measurement results:
$p(\cdots | A, B, C)$ and $p(\cdots | A, L, M)$ (data from experiments)

III. non-contextual condition (how to glue the data together):
the marginal conditional distribution for A are the same, given by $p(\cdot | A)$

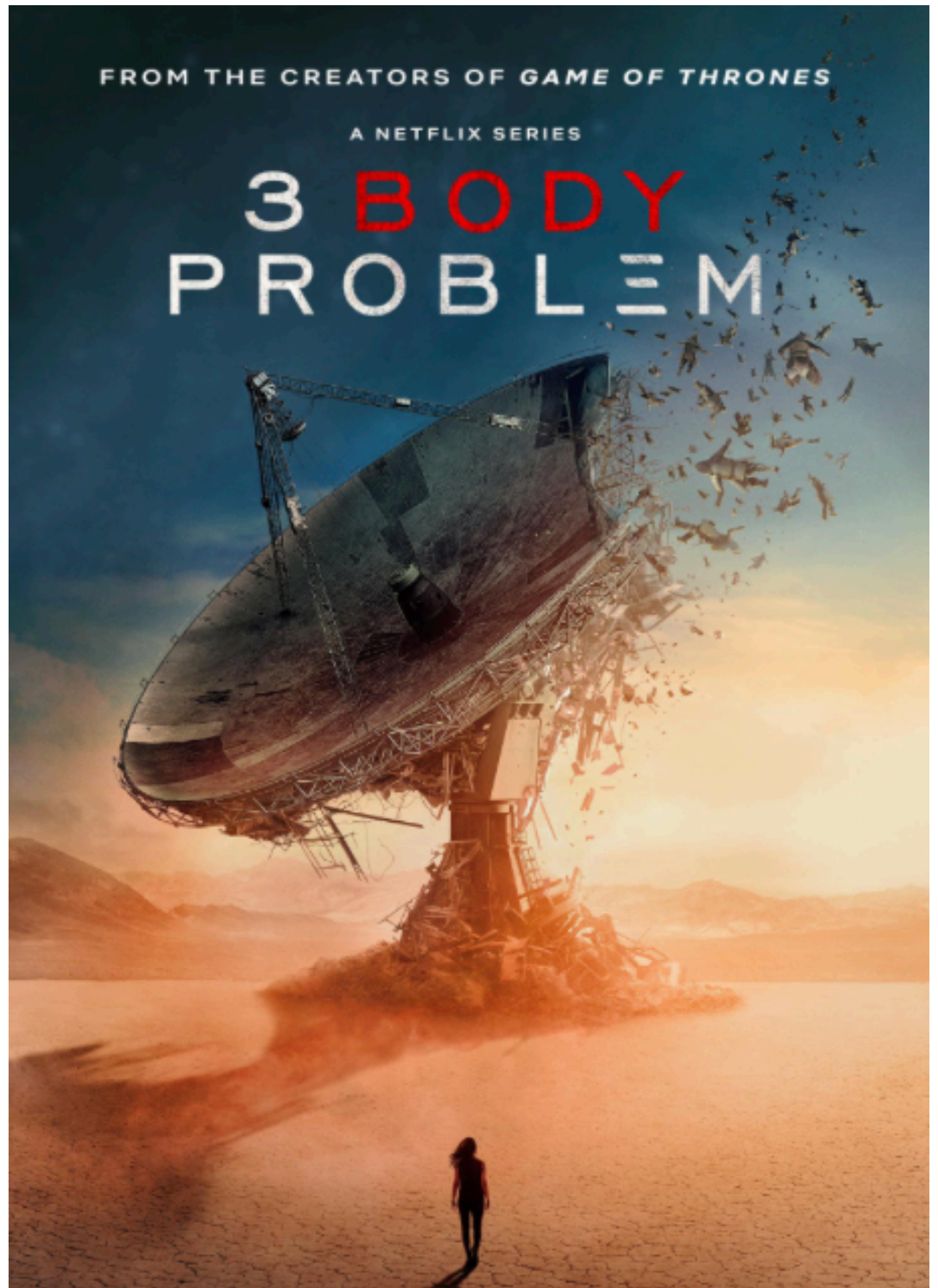
IV. non-contextual hidden variable models:
a global joint distribution $p(\cdots\cdots | A, B, C, L, M)$

Non-contextual hidden variable theory

Bell-Kochen-Specker theorem

- non-contextual hidden variable theory

Why reasonable?
Measurement does not rely on the context.
Imagining measure A first. Nature is not conspiring
(what if we measure A in the final?)



Nature is not "intentionally manipulating" the experiment

"physics does not exist" — Ye Wenjie (a character in Three-Body Problems,
a recent Sci-Fi show in Netflix, physics experiments are manipulated by alien civilization to prevent human-being to develop science)

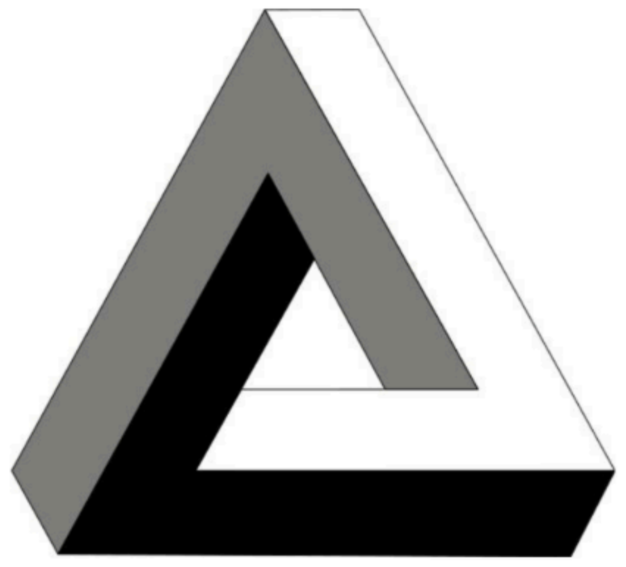
Cavalcanti E G. Classical causal models for Bell and Kochen-Specker inequality violations require fine-tuning. Physical Review X

Non-contextual hidden variable theory

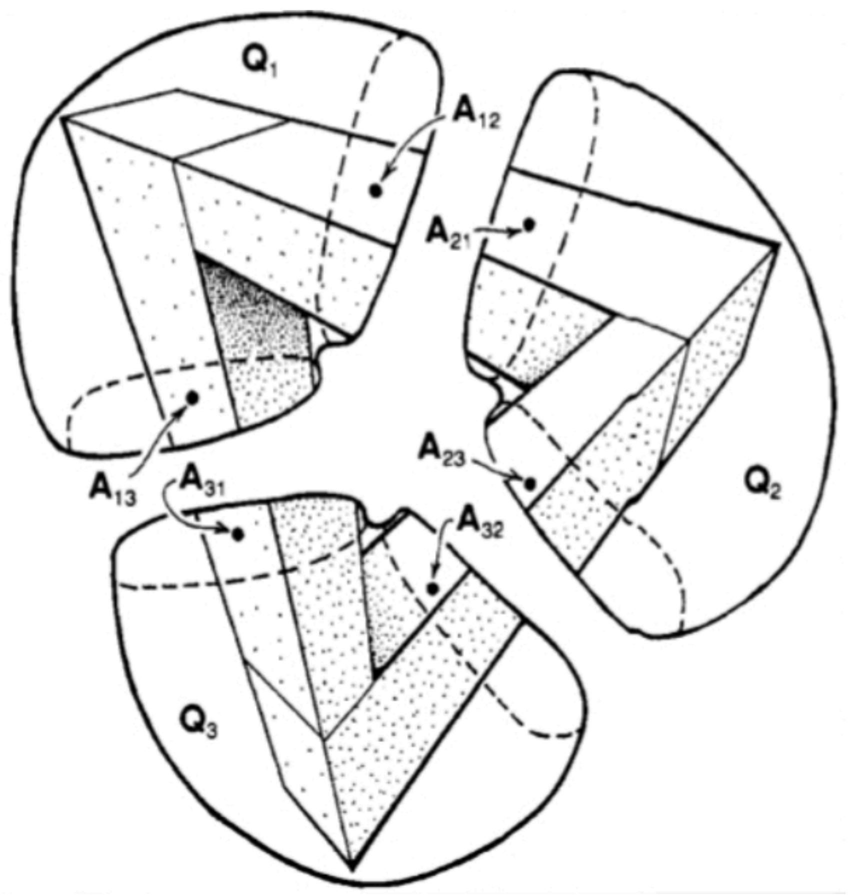
Bell-Kochen-Specker theorem

- Bell-Kochen-Specker theorem

1954
ROGER PENROSE



Non-contextual hidden variable models contradicts with the prediction of quantum mechanics



REVIEWS OF MODERN PHYSICS

Recent | Accepted | Authors | Referees | Search | Press | About | Editorial Tea

Kochen-Specker contextuality

Costantino Budroni, Adán Cabello, Otfried Gühne, Matthias Kleinmann, and Jan-Åke Larsson
Rev. Mod. Phys. **94**, 045007 – Published 19 December 2022

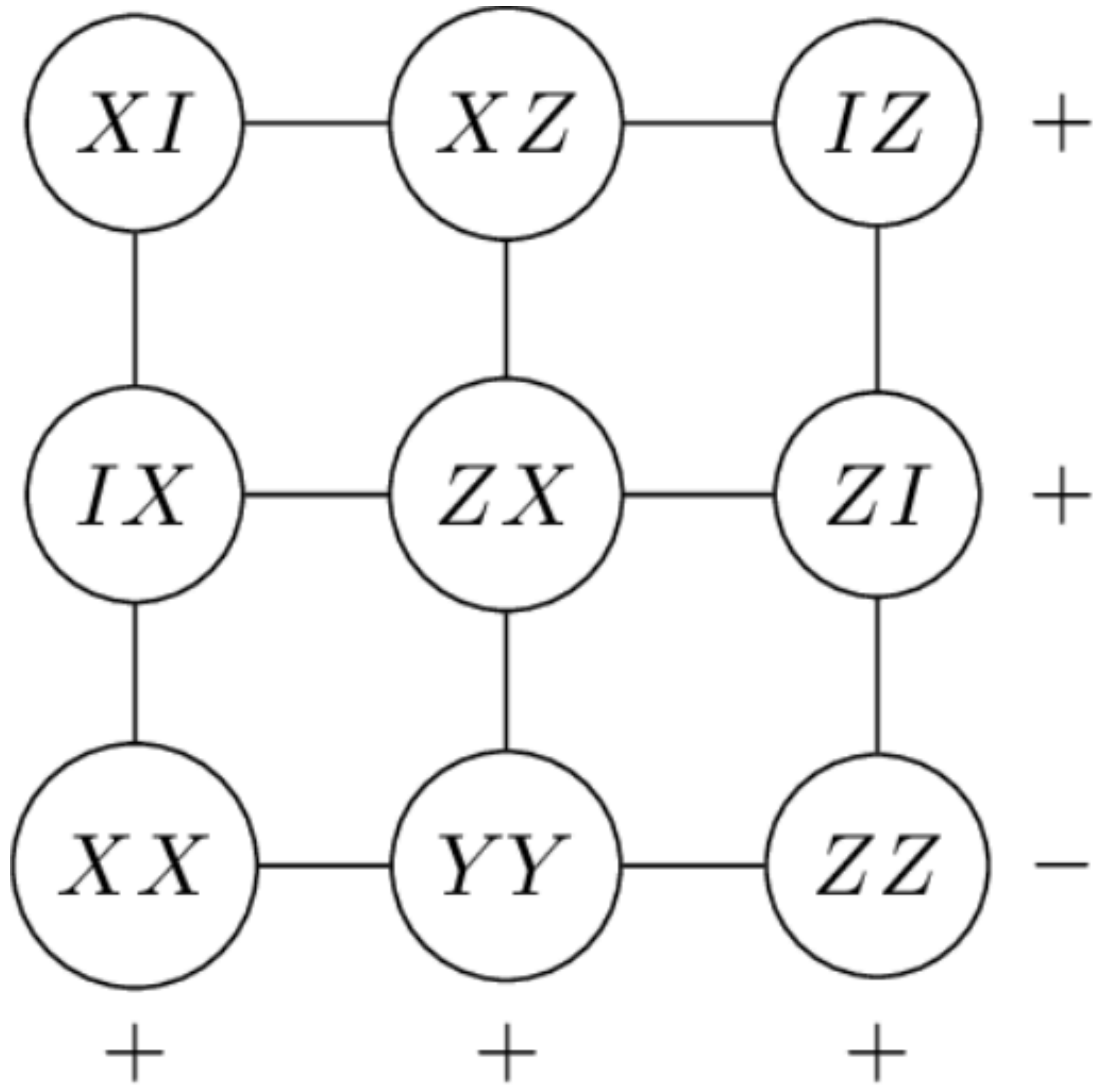
Non-contextual hidden variable theory

Bell-Kochen-Specker theorem

- Mermin-Pere's magic square

suppose a distribution over hidden variable λ
$\lambda(O)$: measurement result of O on the state λ

O is an observable, $\lambda(O)$ is the outcome when measure O on the state described by λ
writing $\lambda(O)$ means condition **III**



Contradiction!

product of row　　product of column

$$\prod_i \left(\prod_j \lambda(O_{ij}) \right) \cdot \prod_j \left(\prod_i \lambda(O_{ij}) \right)$$

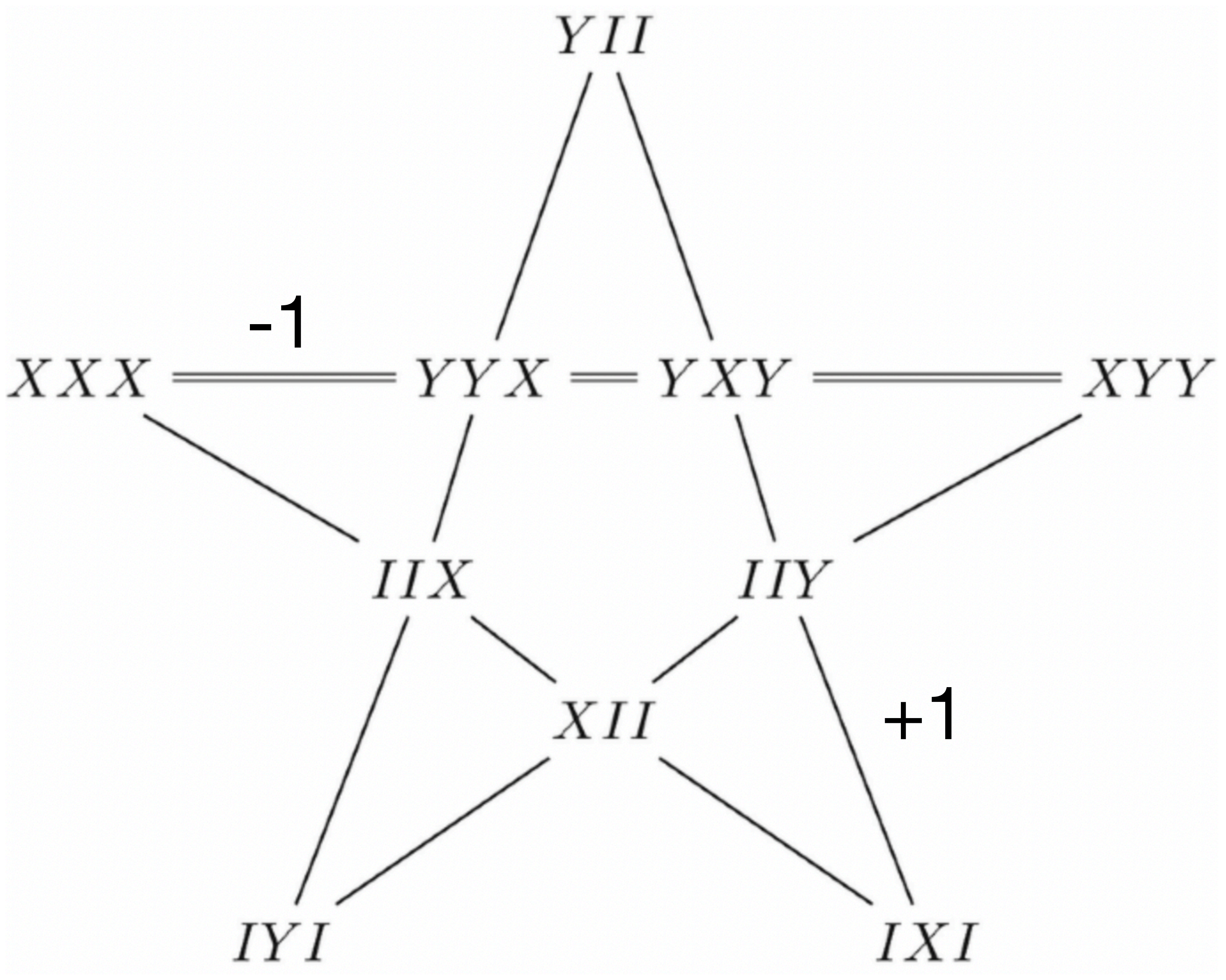
$$(+1)^5 \cdot (-1)^1 = -1$$

$$\prod_i \prod_j \lambda(O_{ij})^2 = 1$$

Non-contextual hidden variable theory

Bell-Kochen-Specker theorem

- Mermin's pentagon



REVIEWS OF MODERN PHYSICS

Recent Accepted Authors Referees Search Press About Editorial Team

Hidden variables and the two theorems of John Bell

N. David Mermin

Rev. Mod. Phys. **65**, 803 – Published 1 July 1993; Errata Rev. Mod. Phys. **85**, 919 (2013); Rev. Mod. Phys. **88**, 039902 (2016); Rev. Mod. Phys. **89**, 049901 (2017)

Both are **state-independent contextuality**

Letter | Published: 23 July 2009

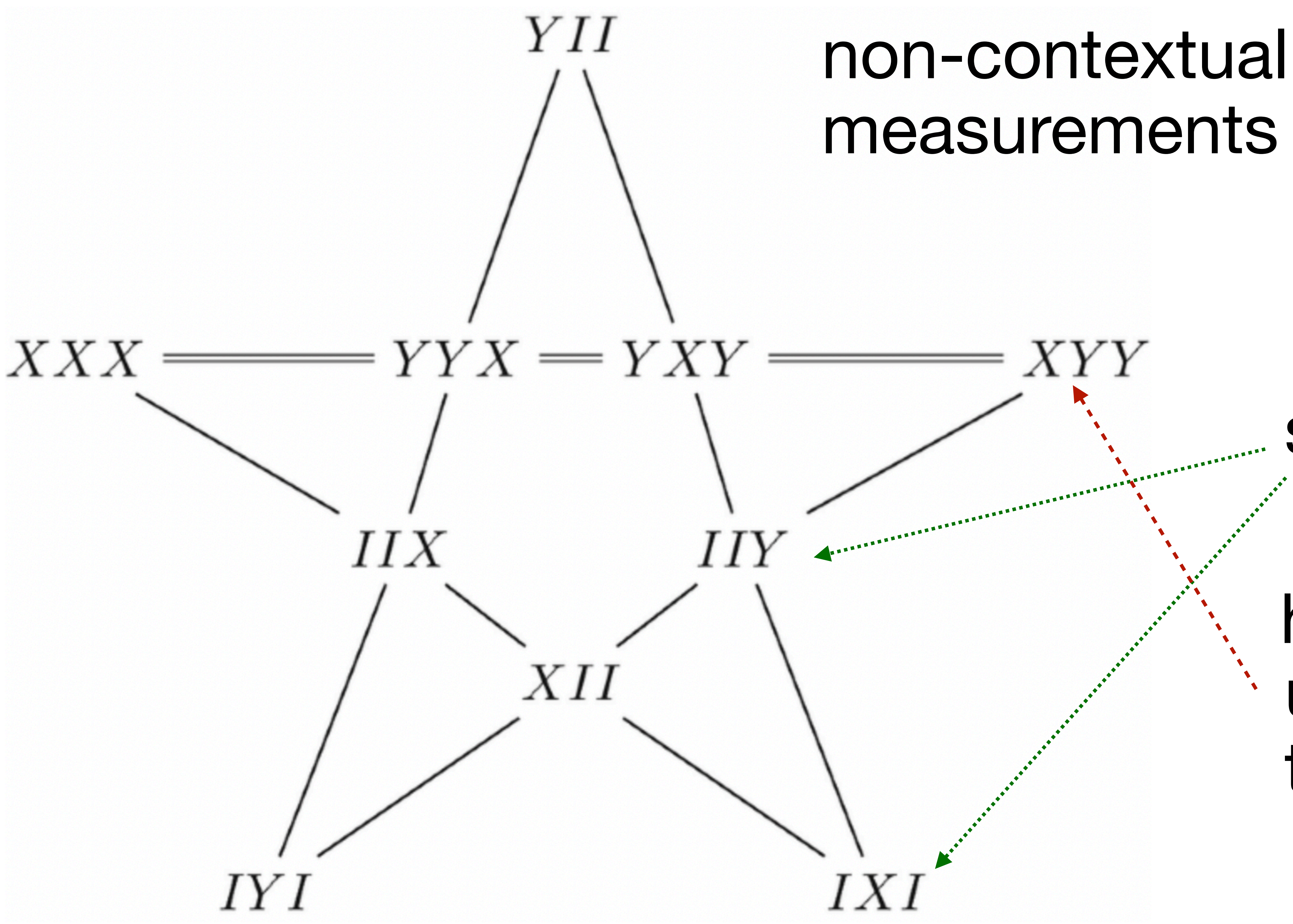
State-independent experimental test of quantum contextuality

G. Kirchmair, F. Zähringer, R. Gerritsma, M. Kleinmann, O. Gühne, A. Cabello, R. Blatt & C. F. Roos

Nature **460**, 494–497 (2009) | Cite this article

From contextuality to nonlocality

- Bell theorem on GHZ state and Mermin pentagon



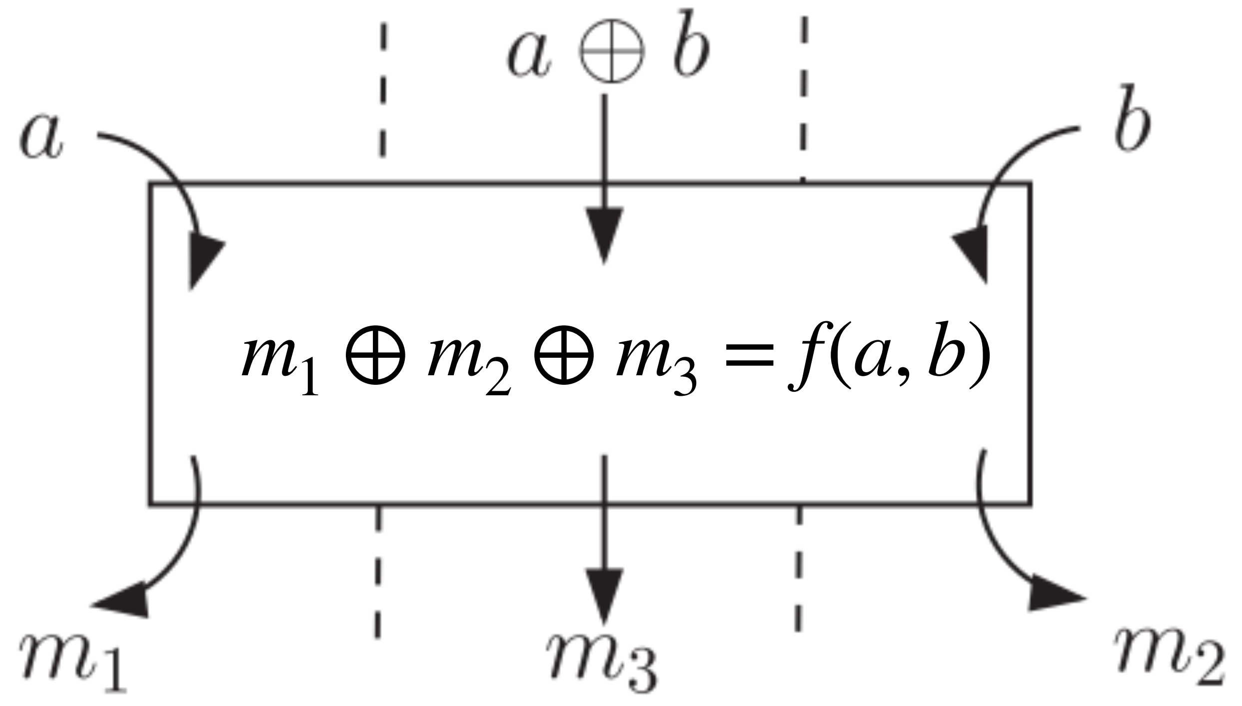
non-contextuality can be replaced by locality:
measurements on different space-like location cannot influence each other

single qubit Pauli measurement is local

how about three qubit Pauli? non-local measurement!
use GHZ state to fix them (their common eigenstate);
then no need to measure them

non-locality also **state-dependent contextuality**

Revisit Bell theorem on GHZ state



non-local game:
referee send $a, a \oplus b, b$ to player 1,2,3 respectively
player 1,2,3 return m_1, m_2, m_3
they win if $m_1 \oplus m_2 \oplus m_3 = f(a,b)$

For f is XOR, winning prop:
Classical at most 75%
Quantum 1

a, b	
0,0	$X \otimes X \otimes X \vert \text{GHZ}\rangle = + \vert \text{GHZ}\rangle$
0,1	$X \otimes Y \otimes Y \vert \text{GHZ}\rangle = - \vert \text{GHZ}\rangle$
1,1	$Y \otimes X \otimes Y \vert \text{GHZ}\rangle = - \vert \text{GHZ}\rangle$
1,0	$Y \otimes Y \otimes X \vert \text{GHZ}\rangle = - \vert \text{GHZ}\rangle$

$a, a \oplus b, b$ $= 0$ to measure X
$= 1$ to measure Y

measurement result is $(-1)^{m_i}$

always $(-1)^{m_1+m_2+m_3} = (-1)^{\text{OR}(a,b)}$

just a different way to interpret "Bell's theorem without inequalities"

Extending to measurement-based quantum computing

PHYSICAL REVIEW LETTERS

Computational Power of Correlations

Janet Anders* and Dan E. Browne†

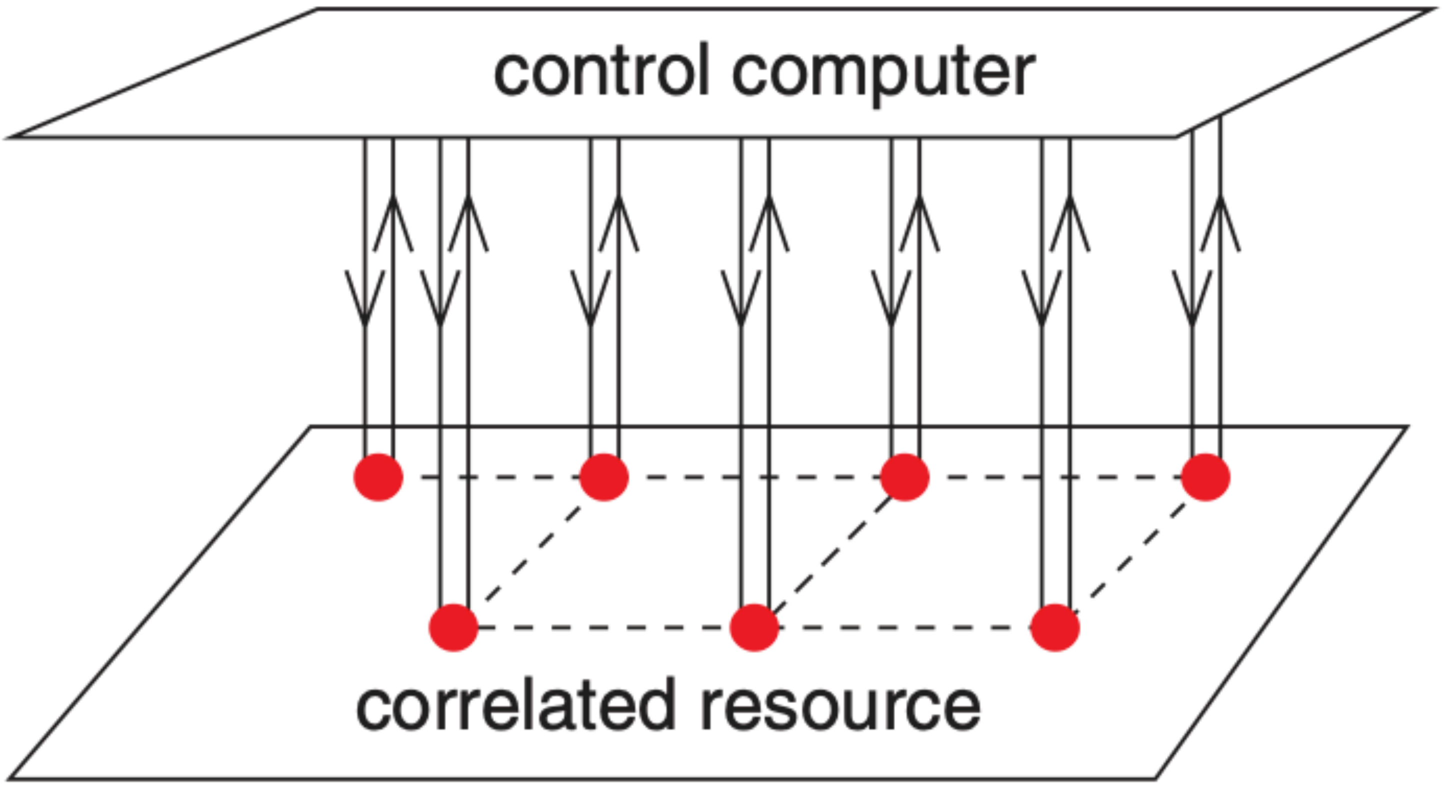
generalize →

PHYSICAL REVIEW A **88**, 022322 (2013)

Contextuality in measurement-based quantum computation

Robert Raussendorf*
Department of Physics and Astronomy, University of British Columbia, Vancouver, British Columbia V6T 1Z1, Canada
(Received 1 May 2013; revised manuscript received 11 July 2013; published 19 August 2013)

linear computation



either classical or quantum
(either entangled or not)

Deterministic computation of non-linear function on $\mathbb{Z}_2$ implies no non-contextual hidden variable theory to explain all the measurement results during the computation

Contextuality is the resource to compute non-linear function in the MBQC model

to use the computational power of the whole system to detect the property of the resource

Nonlinearity (e.g. deviation from linear test? Fourier transformation?) <==> metric of contextuality ?

A Quantum Neural Network Enhanced by Contextuality

High level idea of Quantum Contextuality

(from Wikipedia) the measurement result (assumed pre-existing) of a quantum observable is dependent upon which other commuting observables are within the same measurement set.

Video game: Monument Valley
inspired from M.C.Esher's "Waterfall"

More generally, locally "consistent", globally "inconsistent"

$$p(\cdots y_t \cdots \mid \cdots x_t \cdots) = \sum_{\cdots h_t \cdots} \cdots p(y_{t-1}, h_t \mid x_{t-1}, h_{t-1}) \cdot p(y_t, h_{t+1} \mid x_t, h_t) \cdots$$

In order to predict measurement results correctly, contextuality requires more memory to memorize the "context"

similar to **linguistic contextuality** in language problems

Quantum vs. Linguistic Contextuality

- Analogy

quantum contextuality	to predict measurement results need to memorize the “context”	constraints in a context
linguistic contextuality	the meaning of a word depends on the context	grammar, fixed phrases, etc.

- Sheaf-cohomology

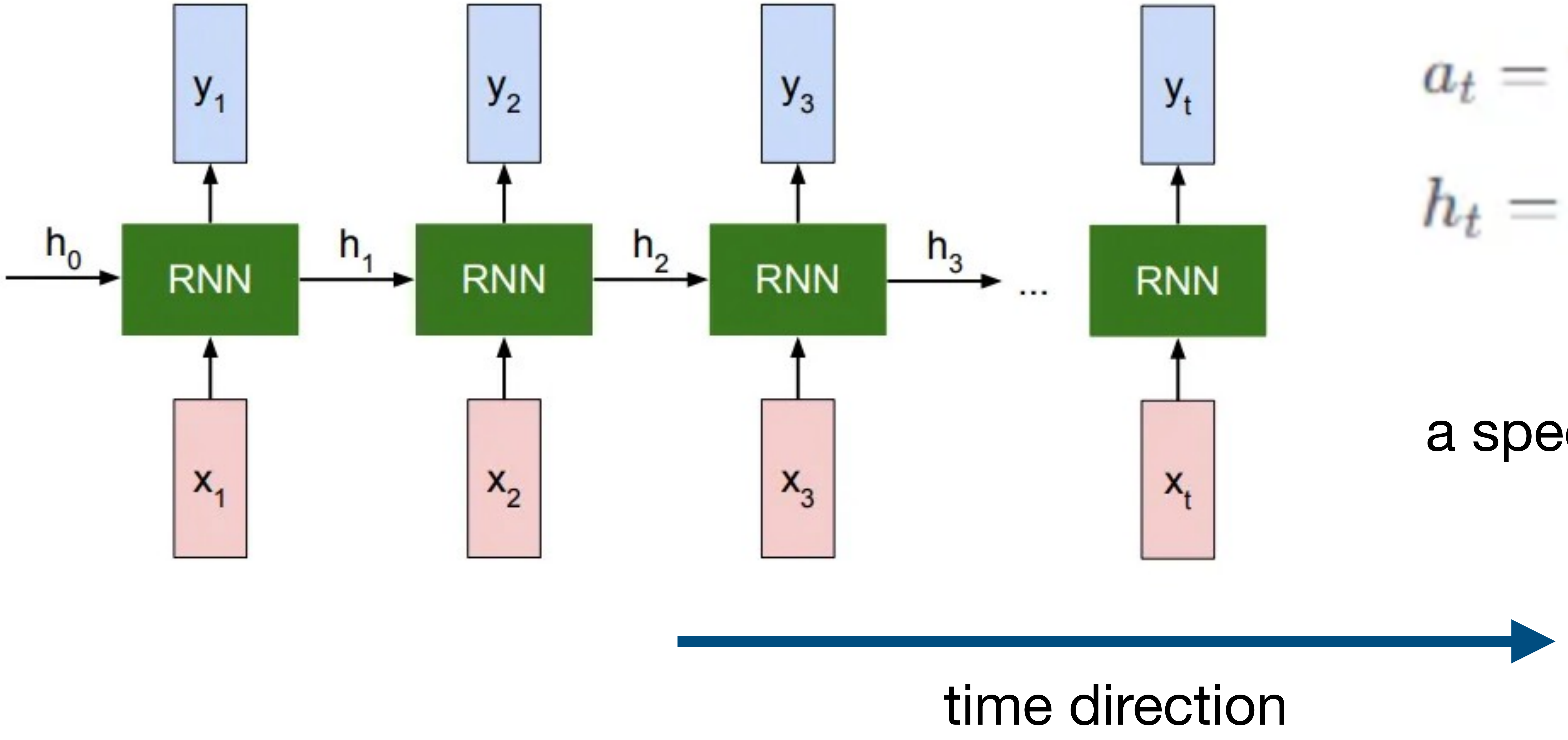
Quantum Contextuality: Abramsky, Samson, and Adam Brandenburger. "The sheaf-theoretic structure of non-locality and contextuality." *New Journal of Physics* (2011)

Natural Language: Lo, Kin Ian, Mehrnoosh Sadrzadeh, and Shane Mansfield. "Developments in Sheaf-Theoretic Models of Natural Language Ambiguities." *arXiv:2402.04505* (2024).

Recurrent Neural Networks

Deterministic HHM with continuous variables

- Recurrent Neural Networks (sequencial models, translation-invariant on time)

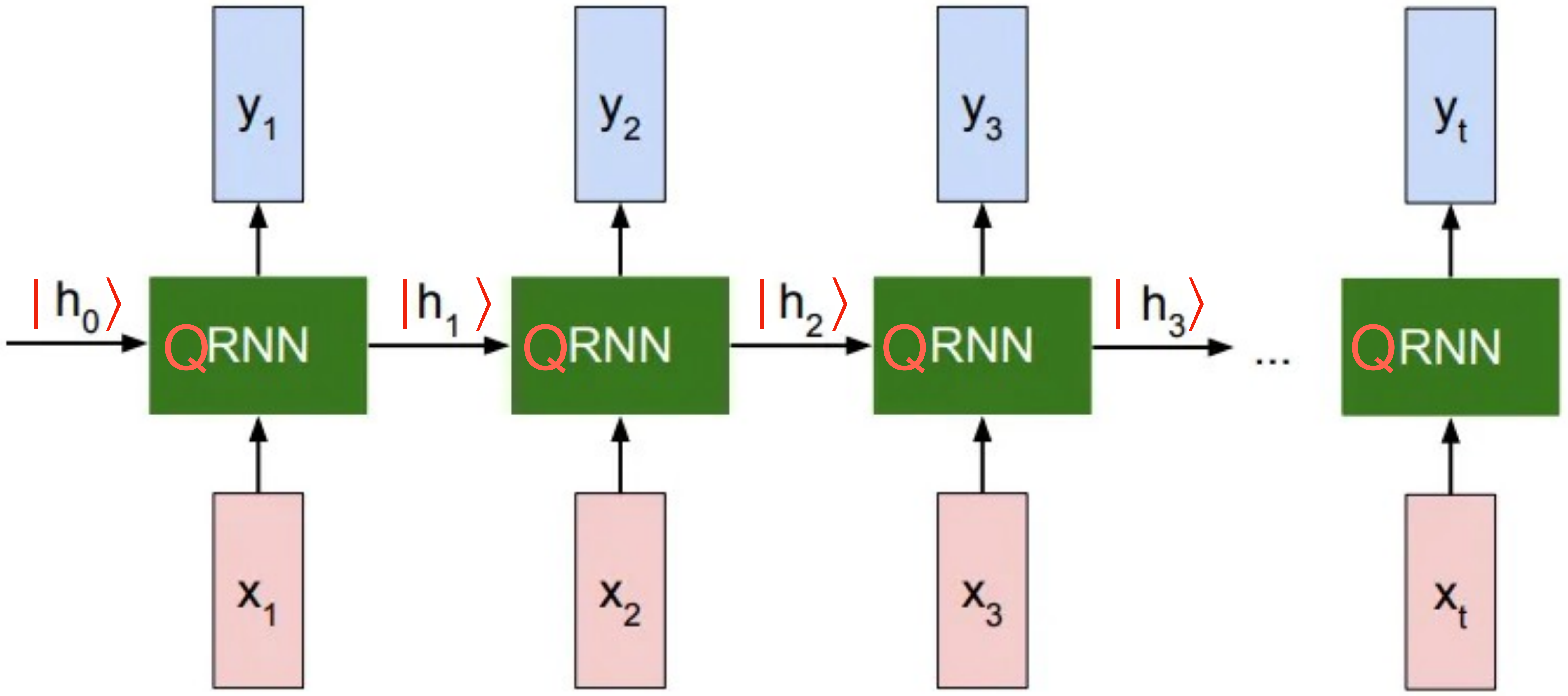


$$a_t = W_{hh} \cdot h_{t-1} + W_{hx} \cdot x_t + b_h$$

$$h_t = \tanh(a_t)$$

a special case of hidden Markov models

The Quantum Neural Networks



y: measurement result

QRNN: gaussian unitary

$x = (x_1, x_2, x_3, x_4, \cdots)$: choose the measurement $x_1\hat{x}_1 + x_2\hat{p}_1 + x_3\hat{x}_2 + x_4\hat{p}_2 + \cdots \mod 2\pi$

- word2vec such that x (y) encode the original (translated) word; the gaussian unitary has training parameters
- If measure $x_1\hat{x}_1 + x_2\hat{p}_1 + x_3\hat{x}_2 + x_4\hat{p}_2 + \cdots$ (homodyne measurement), Gaussian optics (linear optics); there is non-contextual hidden variable theory: $\rho \leftrightarrow W_\rho(x_1, p_1, x_2, p_2, \cdots)$ (Wigner function)
- If measure $(\hat{x}_1\hat{p}_1 + \hat{p}_1\hat{x}_1) \otimes (\hat{x}_2\hat{p}_2 + \hat{p}_2\hat{x}_2) \otimes \cdots$, Gaussian BosonSampling

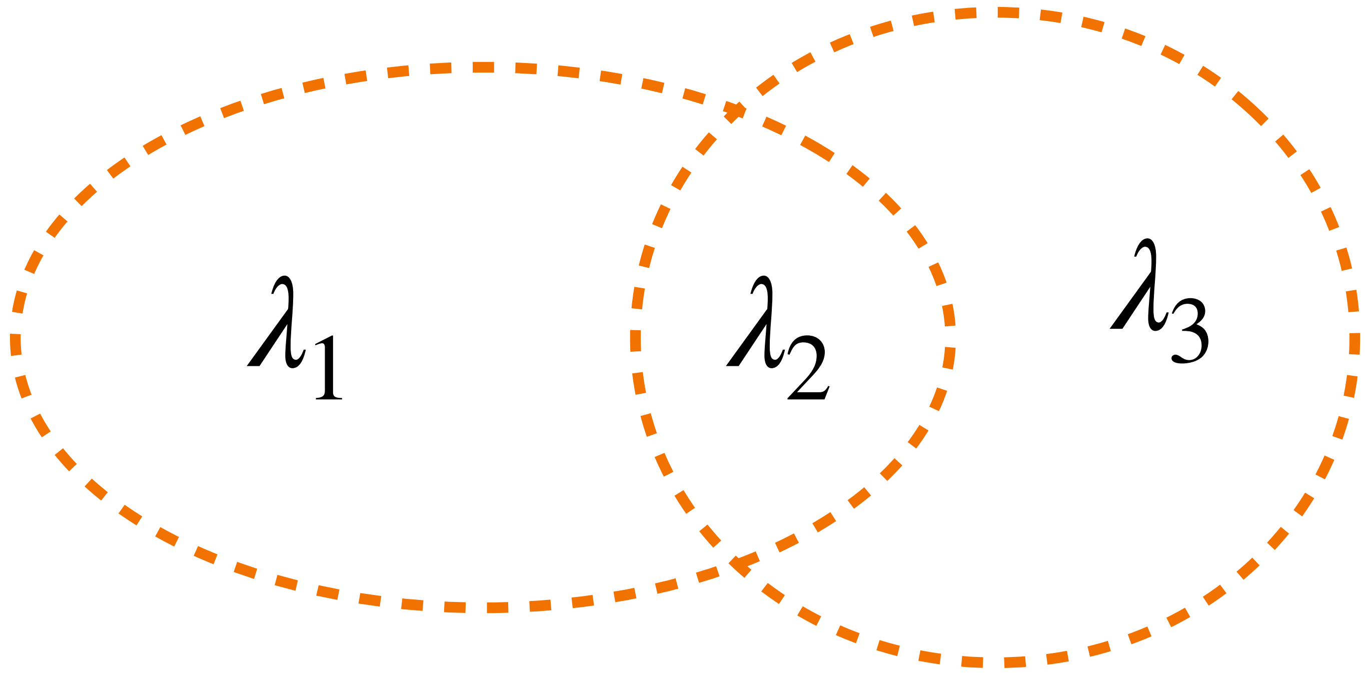
The Quantum Neural Networks

Theoretical results: there exists $p(y_1y_2\cdots|x_1x_2\cdots)$ to approximate, such that

1. quantum model: N hidden neurons (bosonic modes);
2. any classical models: at least $\propto N^2$ hidden neurons (can be extended to $\propto n^k$ but non-gaussian unitary)

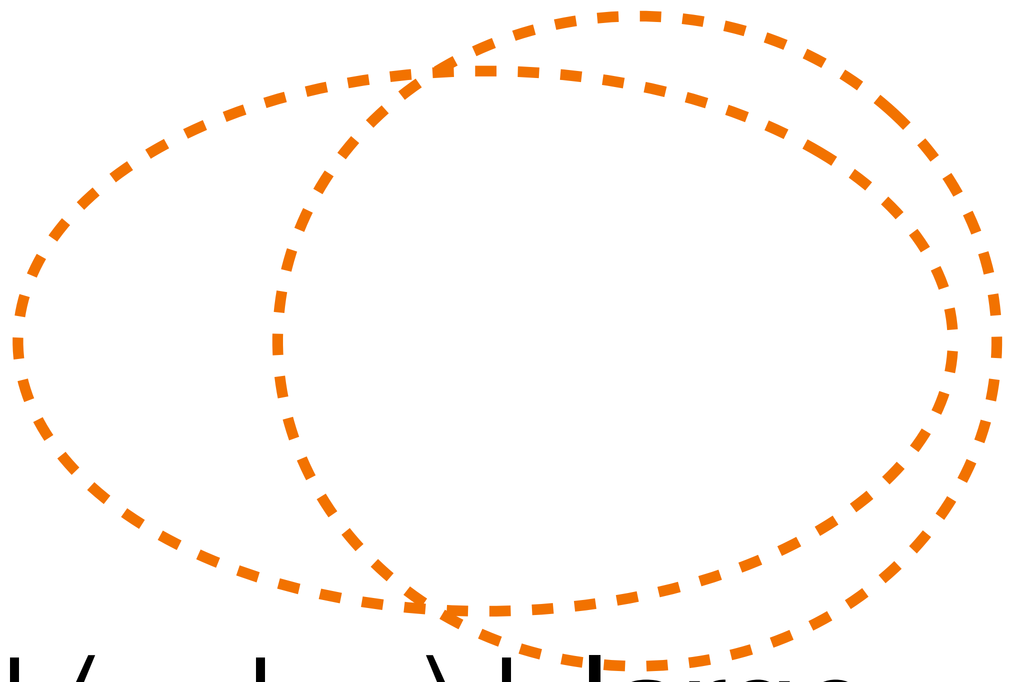
In HVM, a quantum state ⇔ a distribution over hidden variables

Naively, large overlap of 2 states (non-zero inner product) ⇒
the distributions have large overlap
(e.g., Gaussian states by Wigner function rep.)



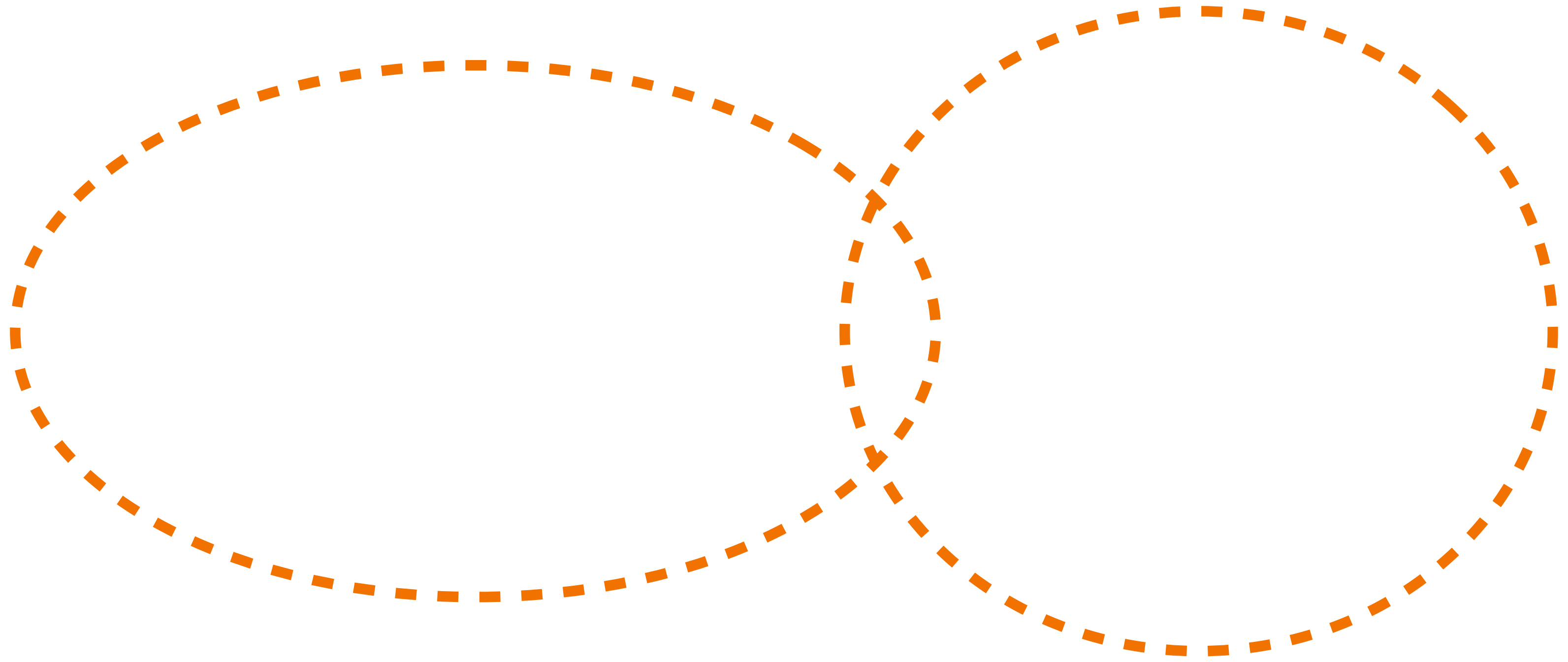
$|\psi_1\rangle$: distribution over λ_1, λ_2

$|\psi_2\rangle$: distribution over λ_2, λ_3



$|\langle\psi_1|\psi_2\rangle|$ large

$|\langle\psi_1|\psi_2\rangle|$ small



hidden variables ~
dim of the Hilbert space

Sketch of the proof

If two states are orthogonal $\Rightarrow$ no common hidden variable λ
Proof: $\lambda(O)$ gives the same result; but there is O to fully distinguish between them deterministically

What if states are not orthogonal?

$$|\psi_1\rangle = |00\rangle$$

$$|\psi_2\rangle = |++\rangle$$

$$|\psi_3\rangle = \mathrm{CZ}|++\rangle$$

$Z\otimes I$	$I\otimes Z$	$Z\otimes Z$
$I\otimes X$	$X\otimes I$	$X\otimes X$
$Z\otimes X$	$X\otimes Z$	$Y\otimes Y$

assume there is a common λ, measure YY

$\Downarrow$

non-zero prob, get $\lambda \to \lambda'$

$\Downarrow$

measure ZZ on λ', non-zero prob with the same results

measure YY, non-zero prob for all 3 states

$\Downarrow$

get $YY = 1$, and states $\frac{1+YY}{2}|\psi_{1,2}\rangle, |\psi_3\rangle$

$\Downarrow$

measure ZZ to fully distinguish $\frac{1+YY}{2}|\psi_{1,2}\rangle$

Pusey M F, Barrett J, Rudolph T. On the reality of the quantum state[J]. Nature Physics, 2012, 8(6): 475-478.

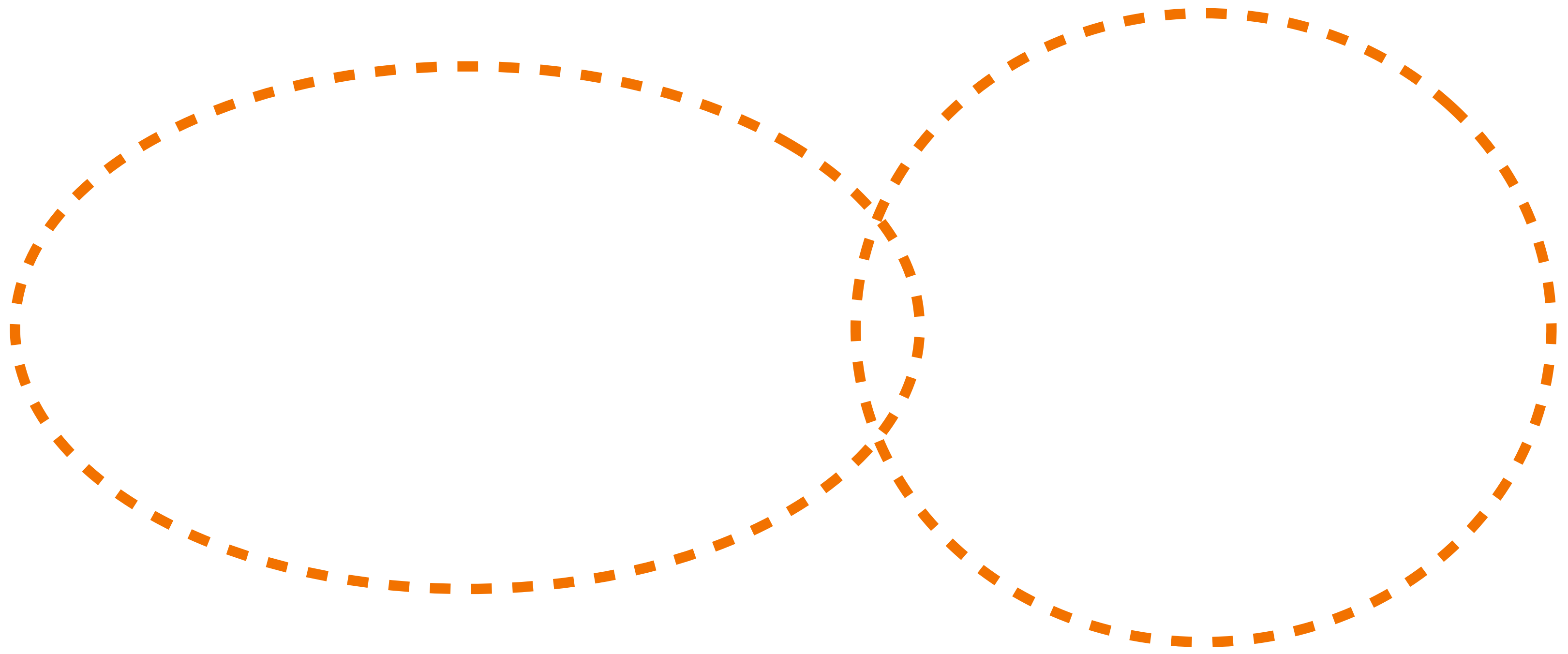
Karanjai A, Wallman J J, Bartlett S D. Contextuality bounds the efficiency of classical simulation of quantum processes[J]. arXiv preprint arXiv:1802.07744, 2018.

Sketch of the proof

The “density” of such kind of triples are very large:
for any m states involved, we can find at least one such triples; but $m \ll$ # states

handwavingly:

even $|\langle\psi_1|\psi_2\rangle|$ large



hidden variables ~ # quantum states
>> dim of Hilbert space

XG, Anschuetz, E. R., Wang, S. T., Cirac, J. I., & Lukin, M. D. Enhancing generative models via quantum correlations. PRX (2022).

Eric Anschuetz, Hongye Hu, Jinlong Huang, **XG**. Interpretable Quantum Advantage in Neural Sequence Learning, PRX Quantum (2023)

Anschuetz, E. R., **XG**. Arbitrary Polynomial Separations in Trainable Quantum Machine Learning. arXiv:2402.08606 (2024)

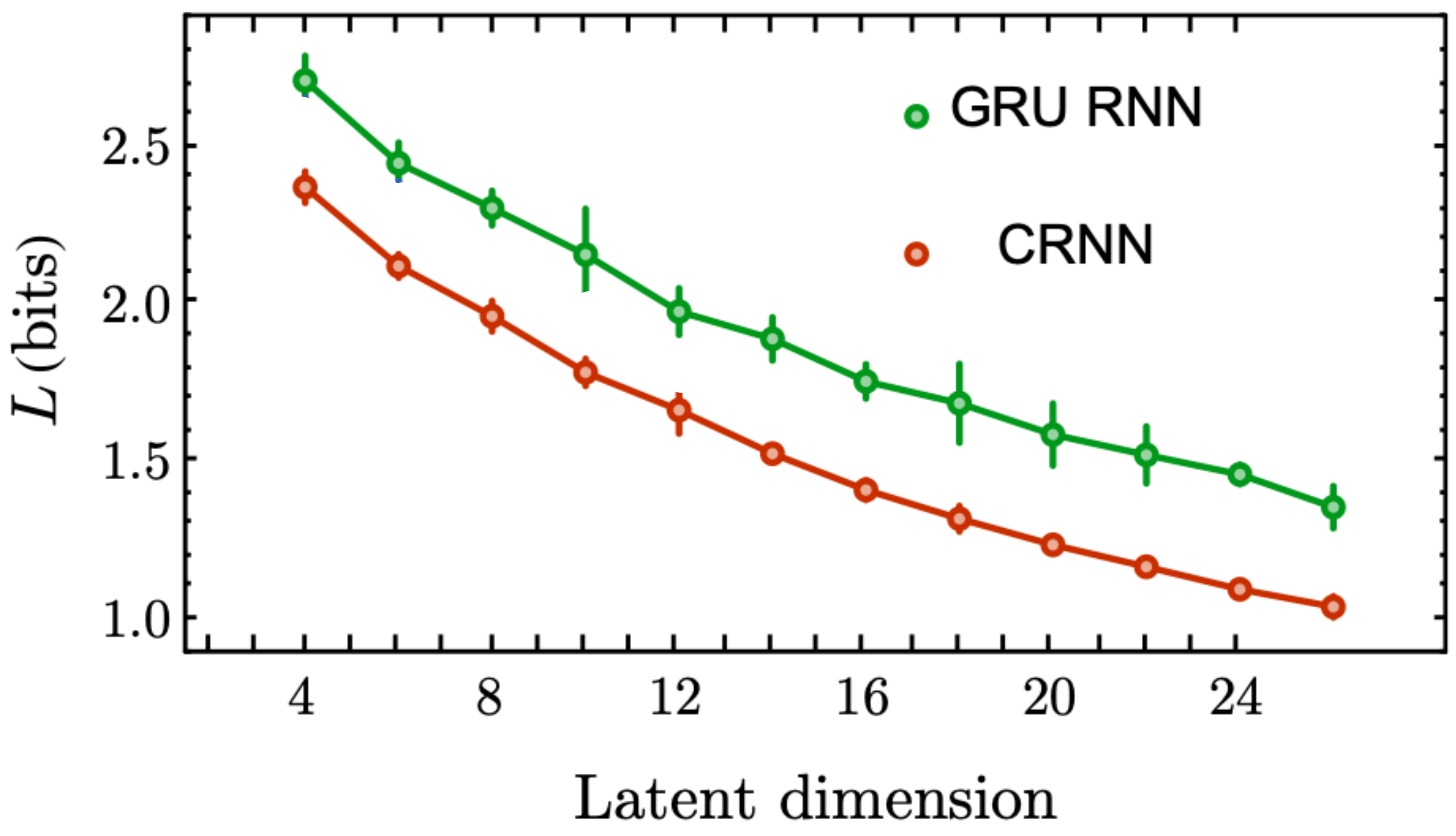
Real-world data

Spanish-English translation

Numerical results: Spanish-to-English translation

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."

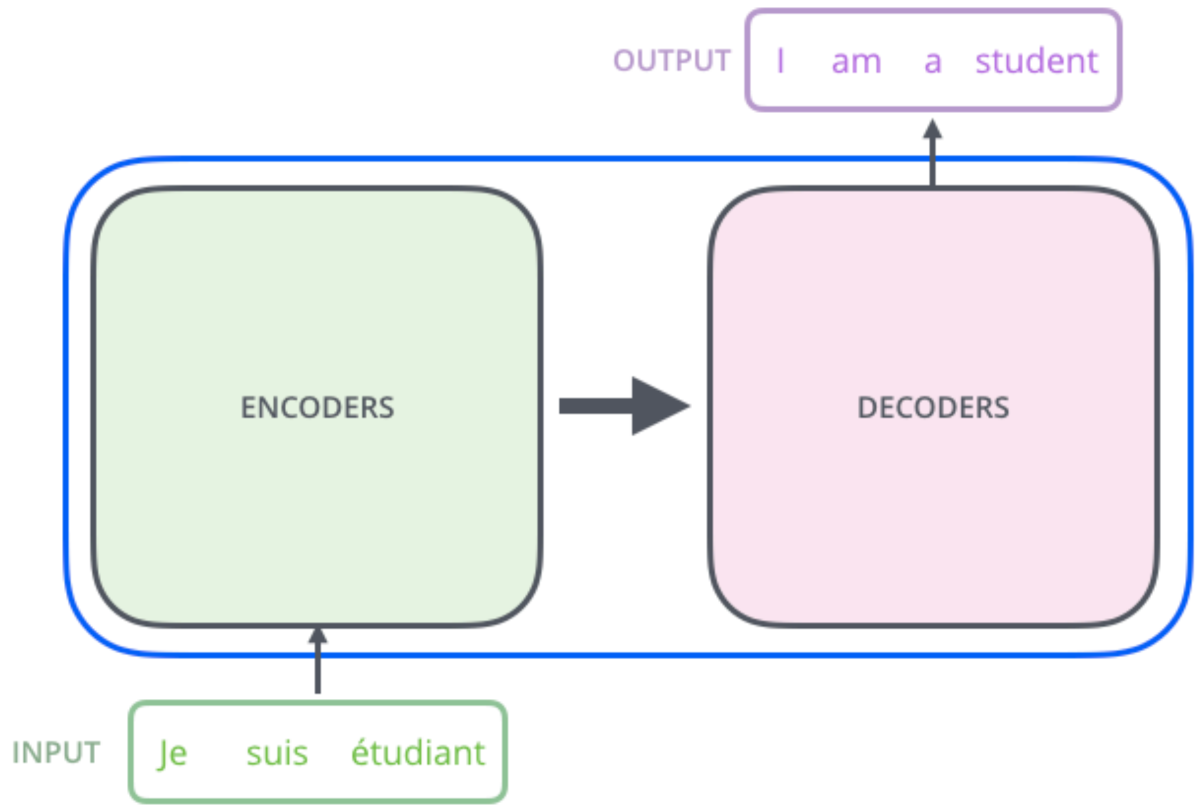
Input	"Admití que estaba equivocada."
Truth	"I admitted that I was wrong."
CRNN	"I was wrong to say that."
GRU	"They had a thing to be true."

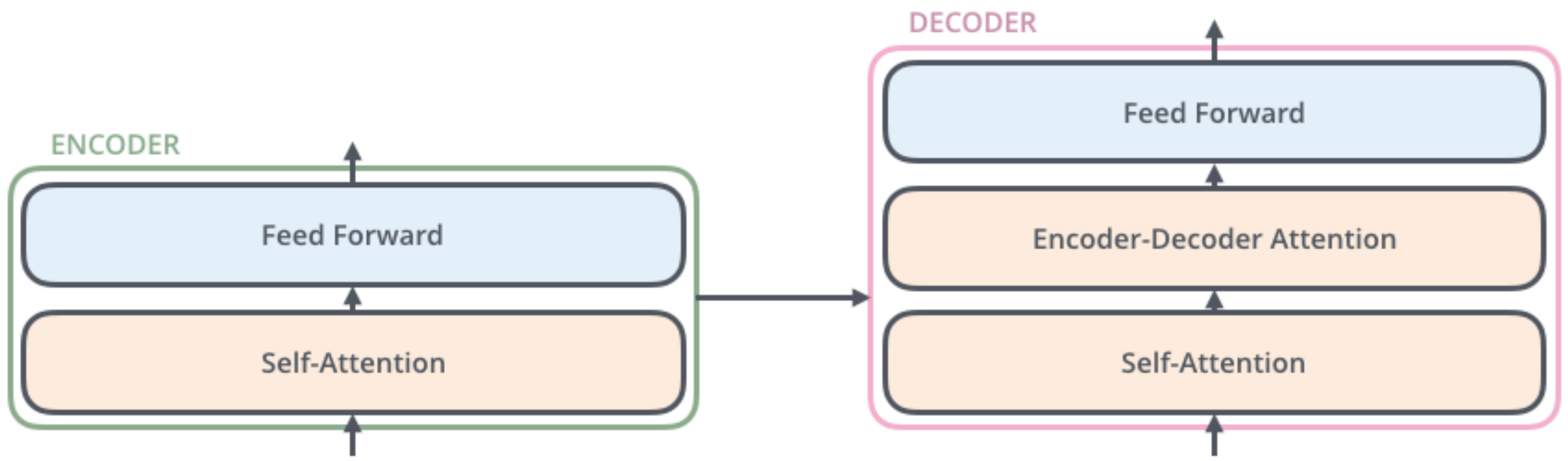


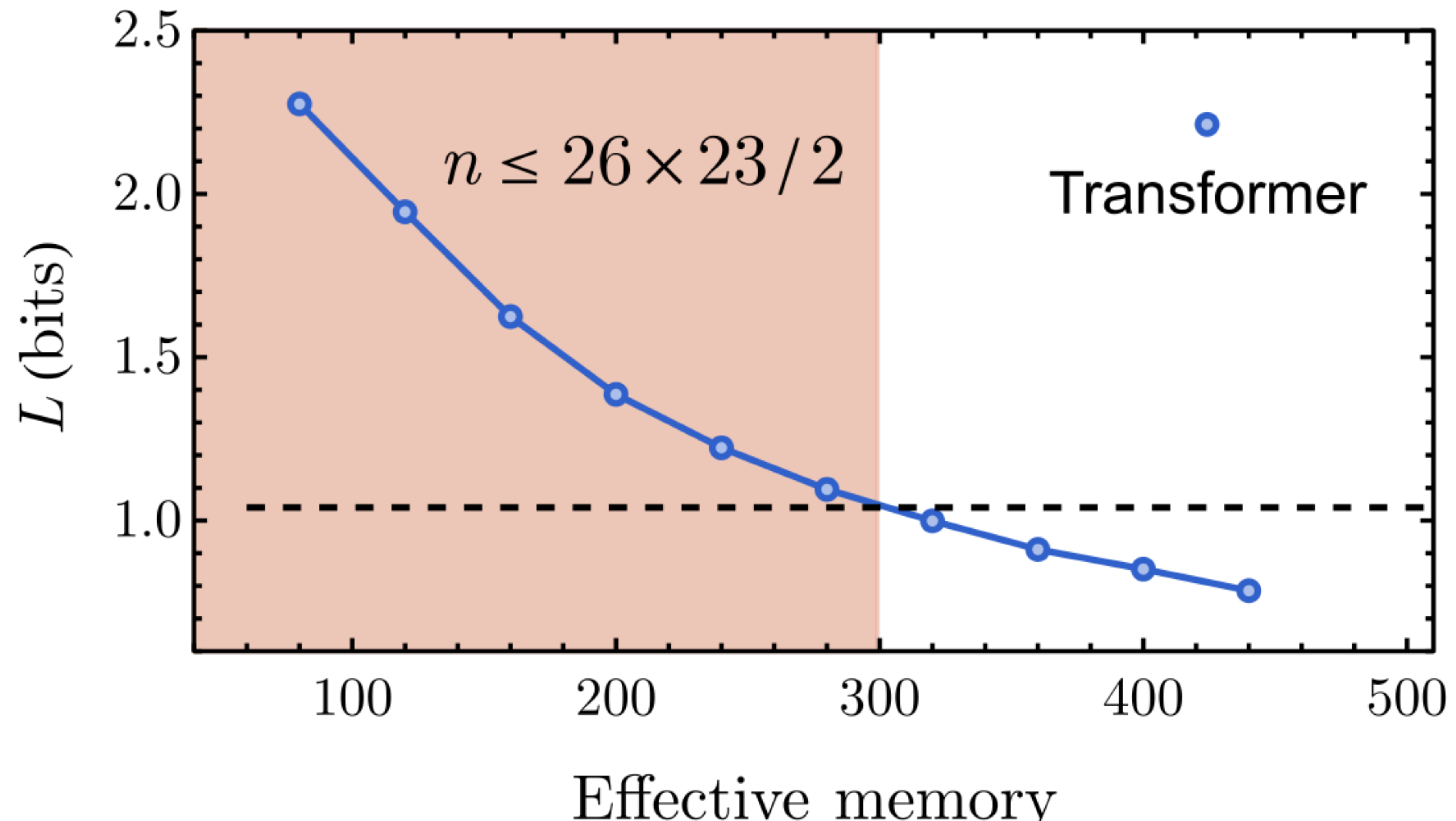
CRNN: Contextual Recurrent Neural Network which is the quantum model
GRU (gated-recurrent-unit): variation of LSTM (basically the best RNN architecture)
here we restrict both models with **just 26 neurons** s.t. we can simulate CRNN

Compared with Transformer

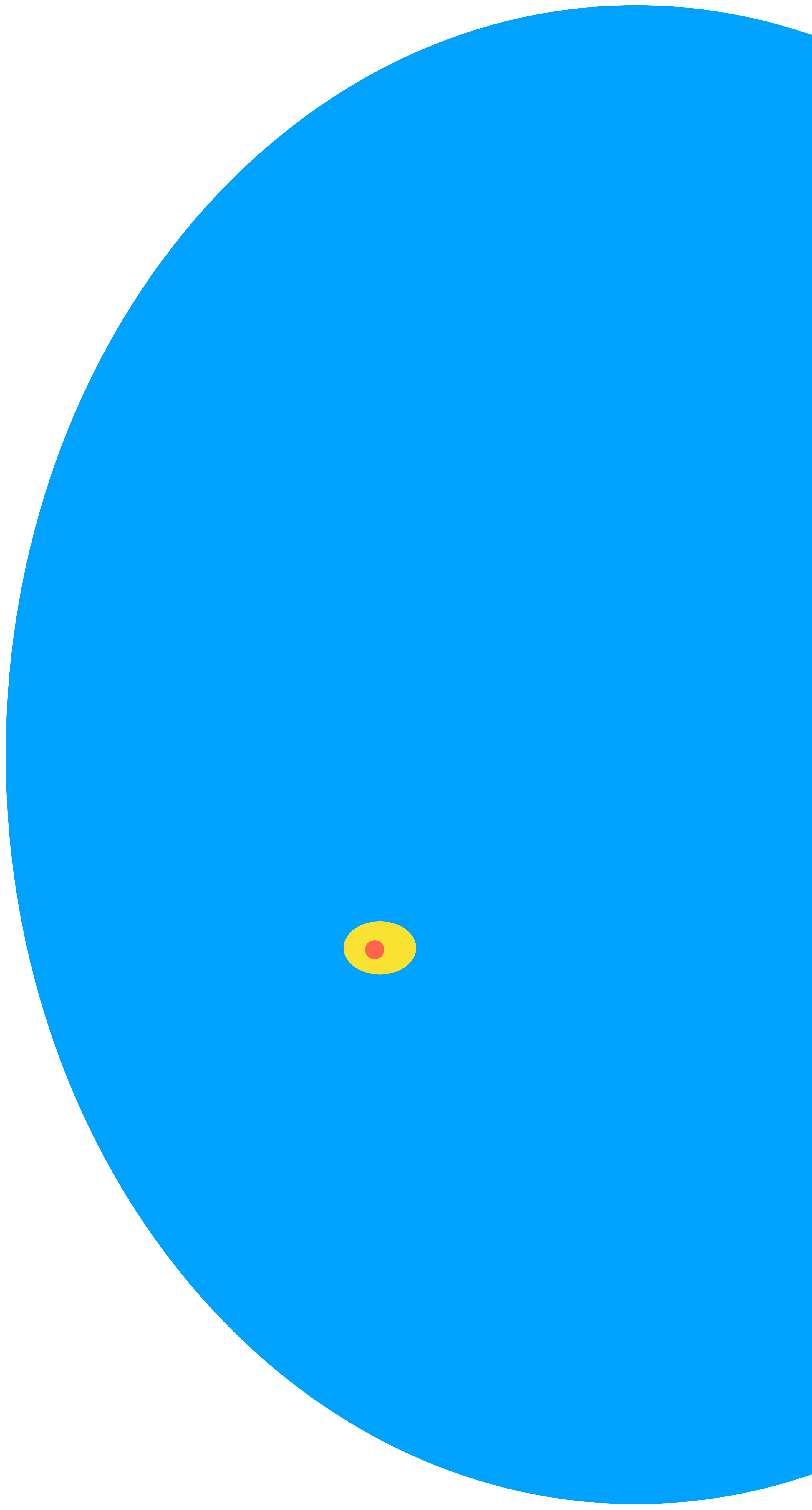
- Transformers (building block of Large Language Model)







Seems that it requires $n^2/2$ hidden neurons?
perhaps coincidence



No-go theorem of general hidden variable theory

Go-beyond non-contextual assumption

- no need to assume locality, non-contextuality, no fine-tune, etc.

  only need to assume the “size” (cardinality, bond dimension, dimension, # neurons, #bits) of hidden Markov model is bounded

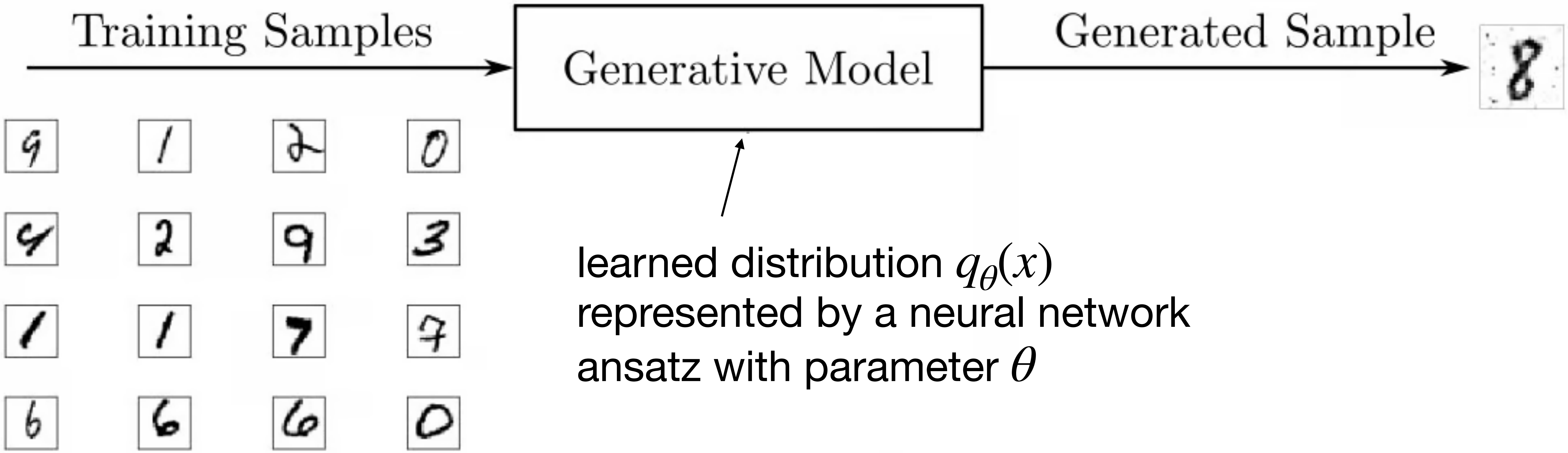
XG, Anschuetz, E. R., Wang, S. T., Cirac, J. I., & Lukin, M. D. Enhancing generative models via quantum correlations. PRX (2022).

Eric Anschuetz, Hongye Hu, Jinlong Huang, **XG**. PRX Quantum (2023)

Anschuetz, E. R., **XG**. Arbitrary Polynomial Separations in Trainable Quantum Machine Learning. arXiv:2402.08606 (2024)

See also the discussion from space complexity point of view:
Karanjai A, Wallman J J, Bartlett S D. Contextuality bounds the efficiency of classical simulation of quantum processes[J]. arXiv preprint arXiv:1802.07744, 2018.

Why expressive power?



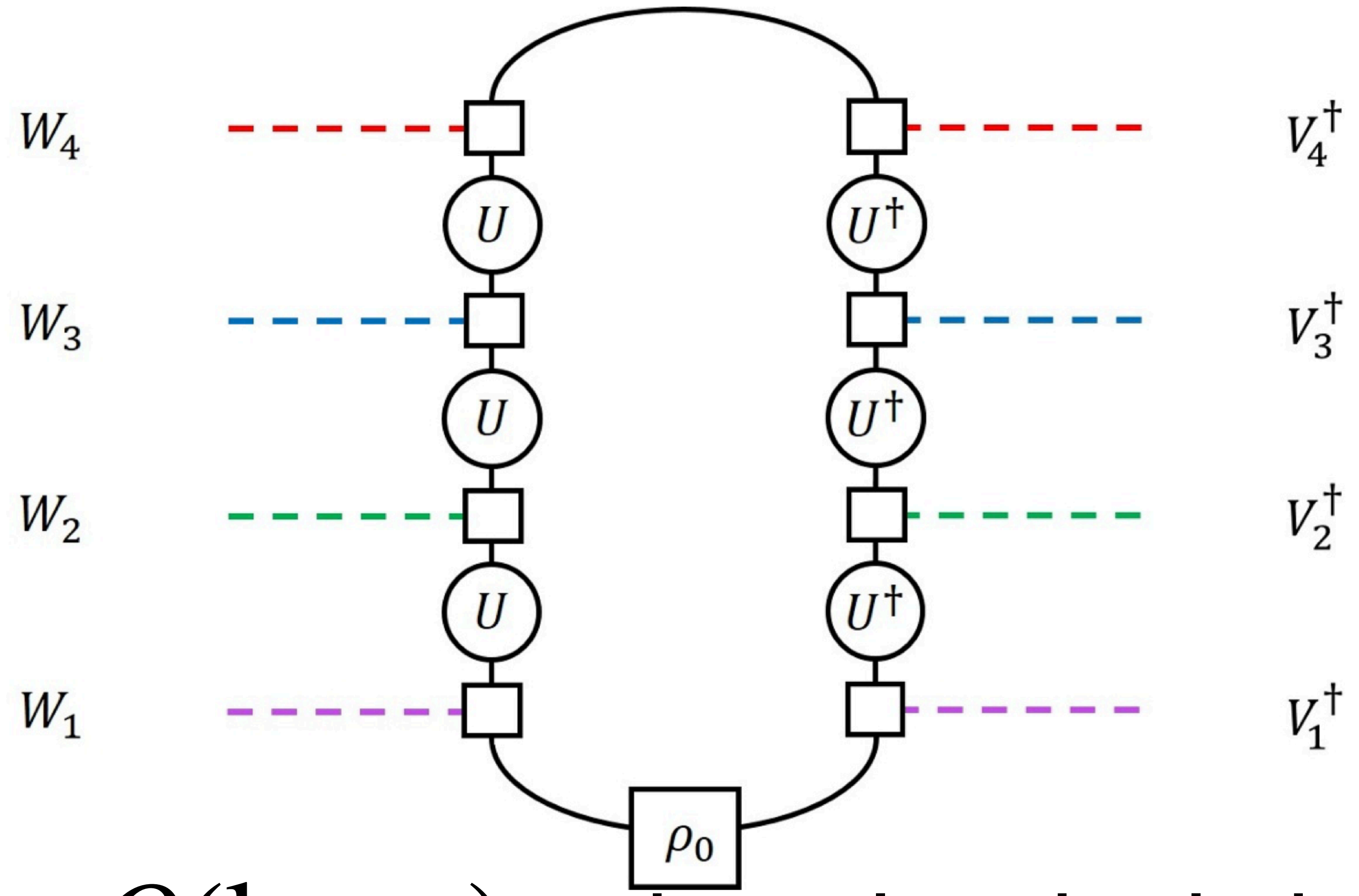
learned distribution $q_\theta(x)$
represented by a neural network
ansatz with parameter θ

Intuitively, smaller size of θ, less number of samples

Time and sample complexity of training?

These works are only focusing on expressive power, the training part is not very clear in detail

- Barren plateau: highly likely to avoid (numerics and Lie algebra structure)
- No back-propagation:
  - translational invariant and shallow in each step



- Gaussian unitary has at most $O(n^2)$ parameters:
  - fully determined by 2-point correlation function), perhaps $O(\log n)$ using classical shadow
  - may need classical shadow tomography for super-density operator

Huang H Y, Kueng R, Preskill J. Predicting many properties of a quantum system from very few measurements[J]. Nature Physics, 2020, 16(10): 1050-1057.

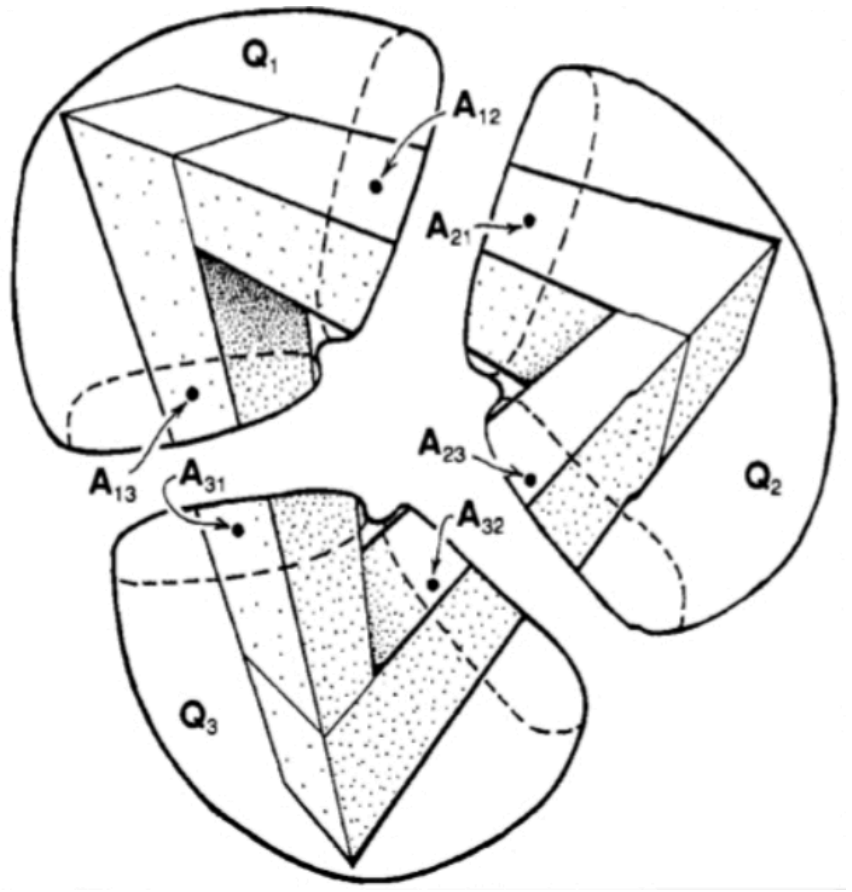
Cotler J, Jian C M, Qi X L, et al. Superdensity operators for spacetime quantum mechanics[J]. Journal of High Energy Physics, 2018, 2018(9): 1-57.

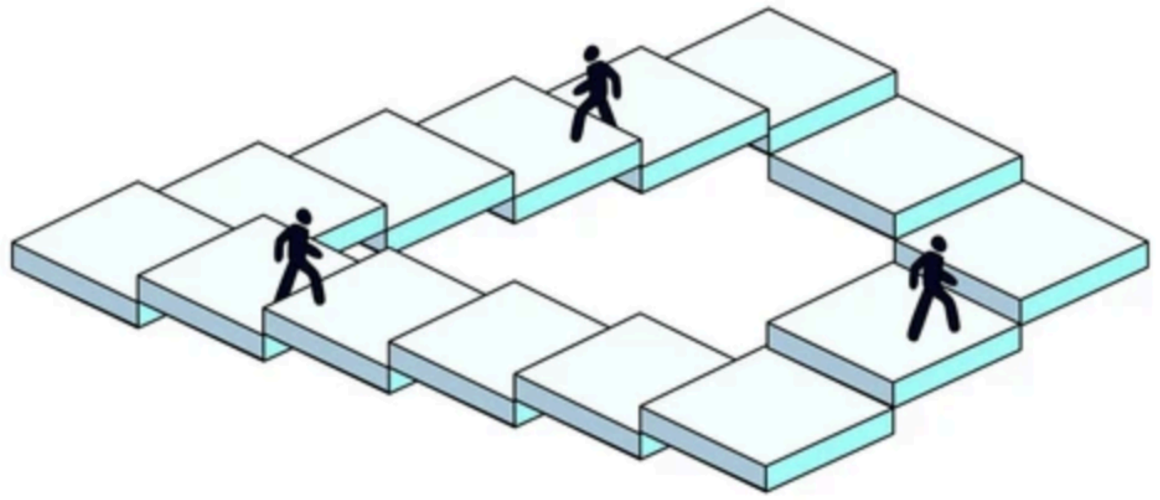
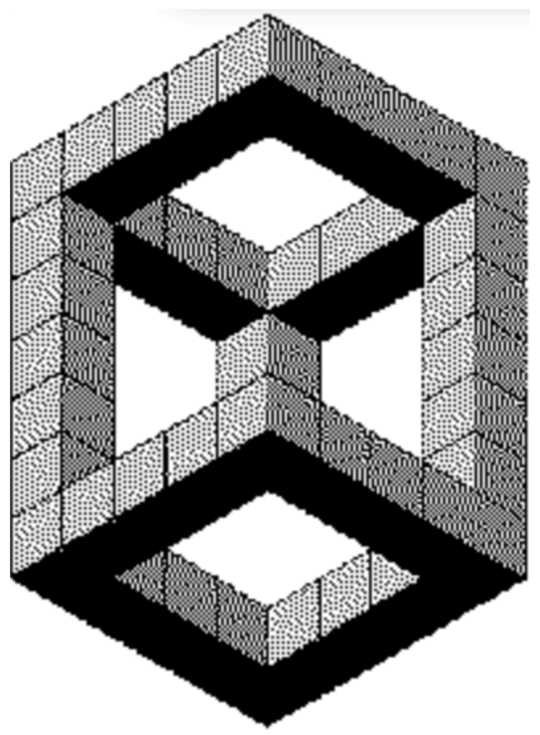
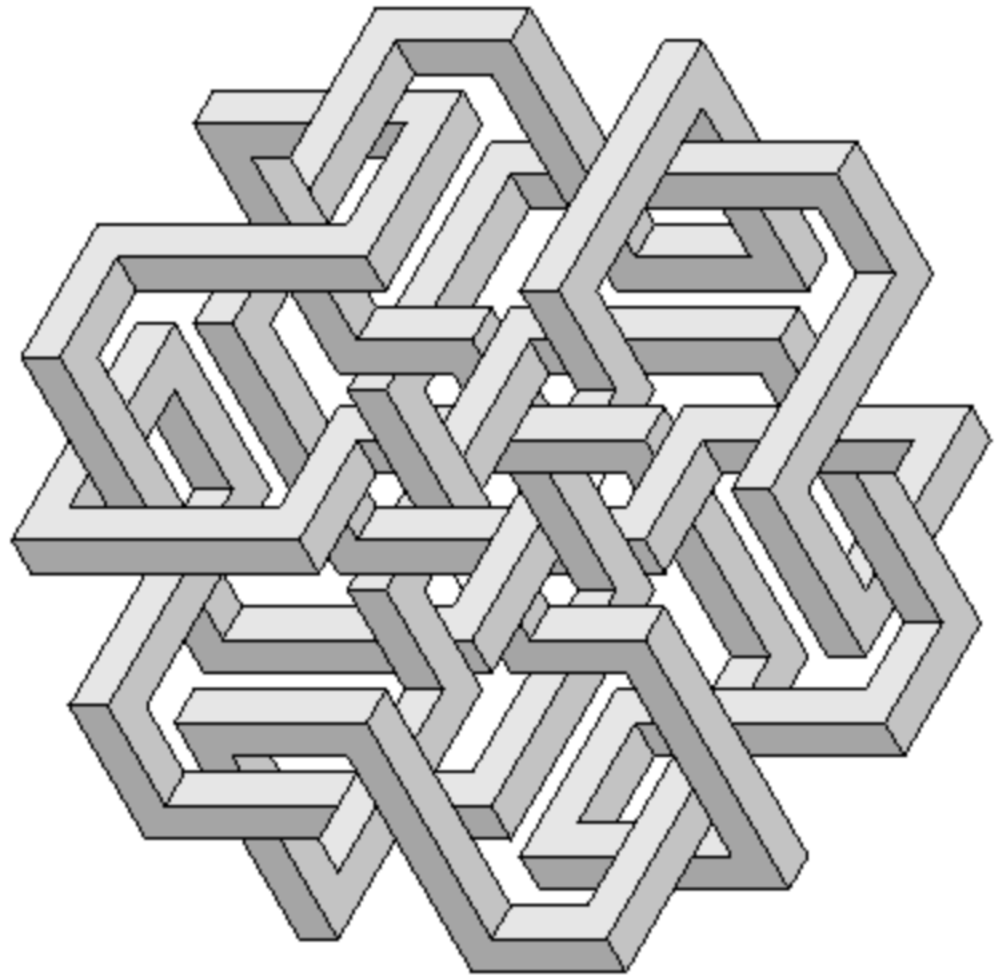
Outlook

Common math between quantum & linguistic Contextuality

- Sheaf-cohomology

  The math to study something “locally consistent but globally inconsistent”

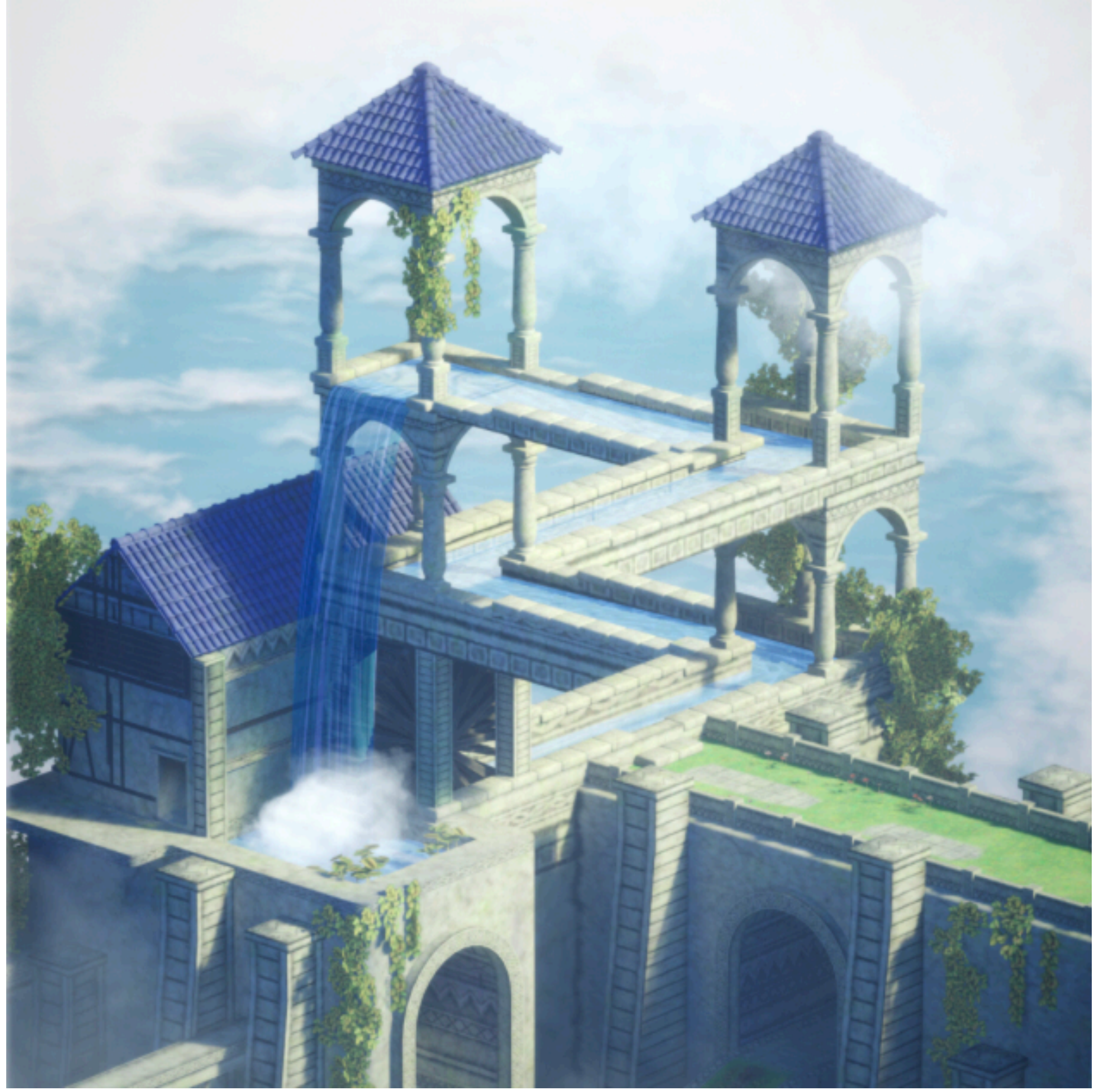


Penrose, Roger. On the cohomology of impossible figures

Cervantes V H, Dzhafarov E N. Contextuality analysis of impossible figures

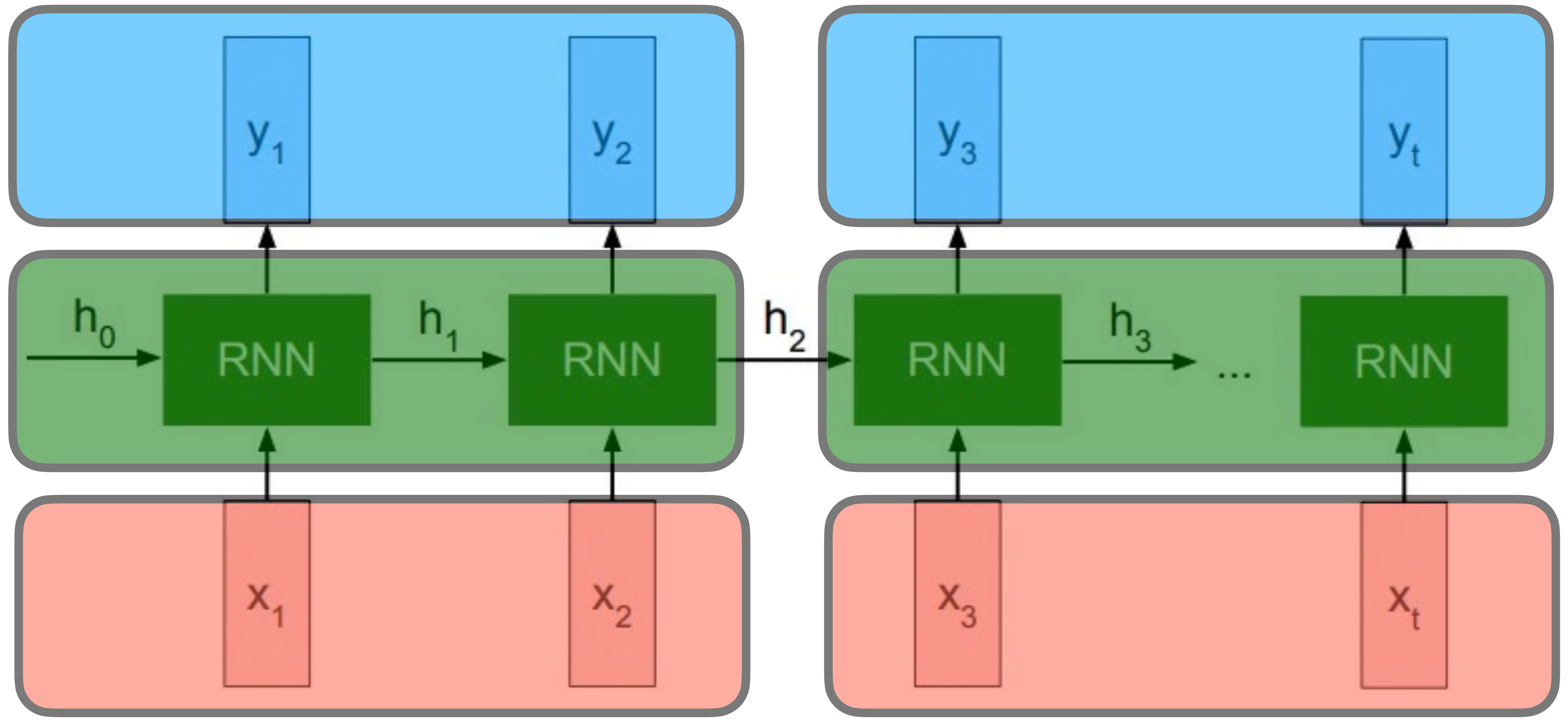
Common math between quantum & linguistic Contextuality

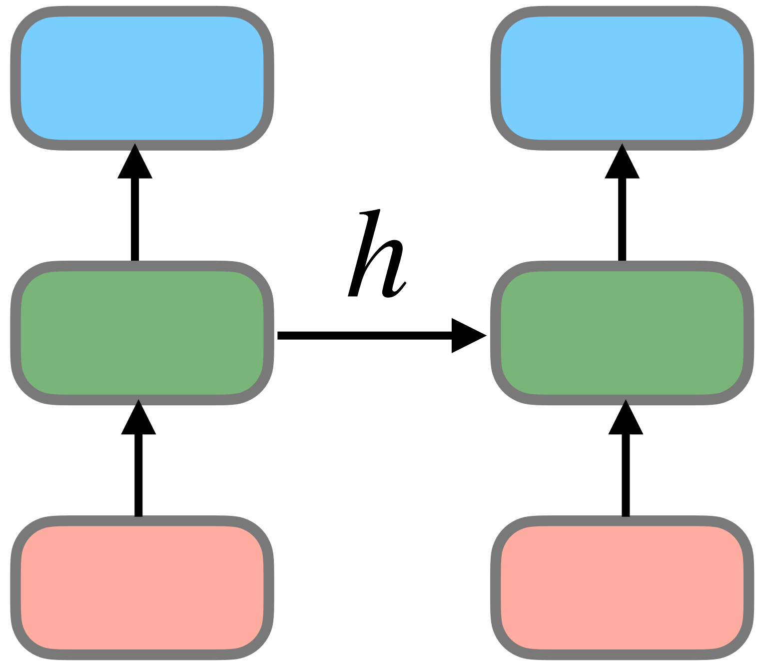


The right one is from ChatGPT and DALL-E: "Here is an illustration inspired by Escher’s “Waterfall,” depicting an impossible and surreal structure where water flows uphill and cascades down again in an endless loop.”

Relation with communication complexity

Theoretical results: there exists $p(y_1 y_2 \cdots | x_1 x_2 \cdots)$ to approximate, such that
1. quantum model: $D = \log N$ qubits;
2. any classical models: at least $\Omega(N) \propto \exp(D)$ bits hidden variables.



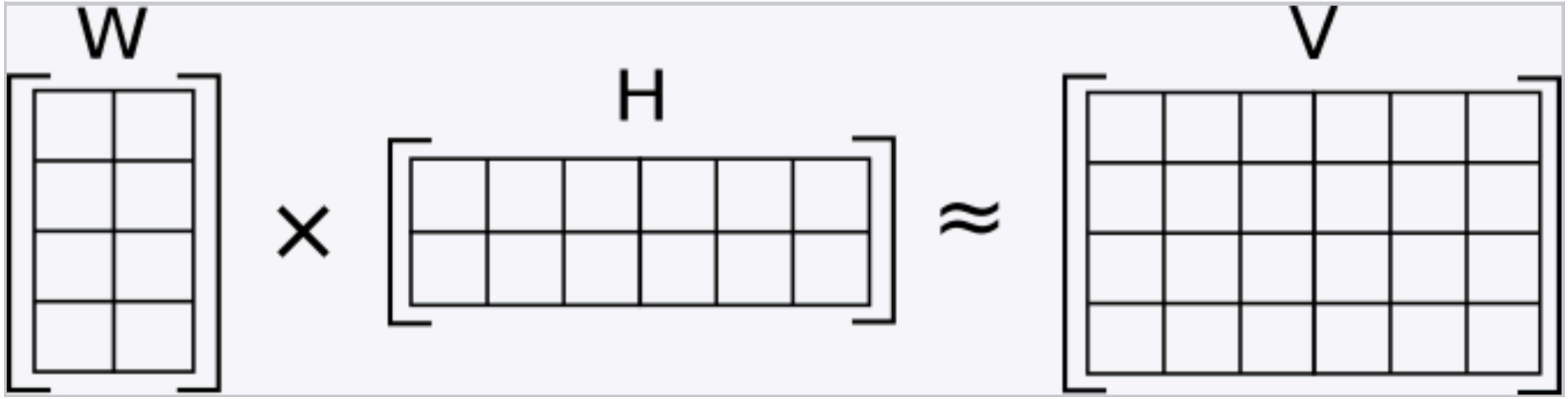


$\log N$ vs. $\Omega(N)$ separation in communication complexity
hidden neurons ⇔ one-way communication complexity

The computational complexity for each unit cell is exponentially long $\propto N$
how quantum play a role? (there is already exponential separation in expressive power but based on complexity assumption instead of unconditional proof here)

XG, Zhang, Z.Y., Duan, L. M. A quantum machine learning algorithm based on generative models. Sci.Adv. (2018).
Raz, Ran. "Exponential separation of quantum and classical communication complexity." *STOC* 1999.

Relation with non-negative matrix factorization



rank=2 in this example

If **each entry** in W and H ≥ 0,
non-negative rank

Correlation/Communication complexity of generating bipartite states

Rahul Jain* Yaoyun Shi† Zhaohui Wei‡ Shengyu Zhang§

Positive Semi-Definite Rank (PSD rank):

$$V = \sum_{i}^{K} \mathrm{tr}(P_i Q_i), \text{ where } P_i, Q_i \geq 0 \text{ (positive semi-definite)}$$

contextuality may give a separation

Potential experiments

- GKP —> other non-Gaussian state

R. Booth, et al.. "Contextuality and Wigner negativity are equivalent for continuous-variable quantum measurements." PRL (2022)
J. Haferkamp, et. al. "Equivalence of contextuality and Wigner function negativity in continuous-variable quantum optics." arXiv:2112.14788 (2021).

measure $e^{i(a\hat{x}+b\hat{p})}$:

$a\hat{x} + b\hat{p} \mod 2\pi$ $\Longleftrightarrow$

- Gaussian BosonSampling, almost gaussian except measurement (photon number)

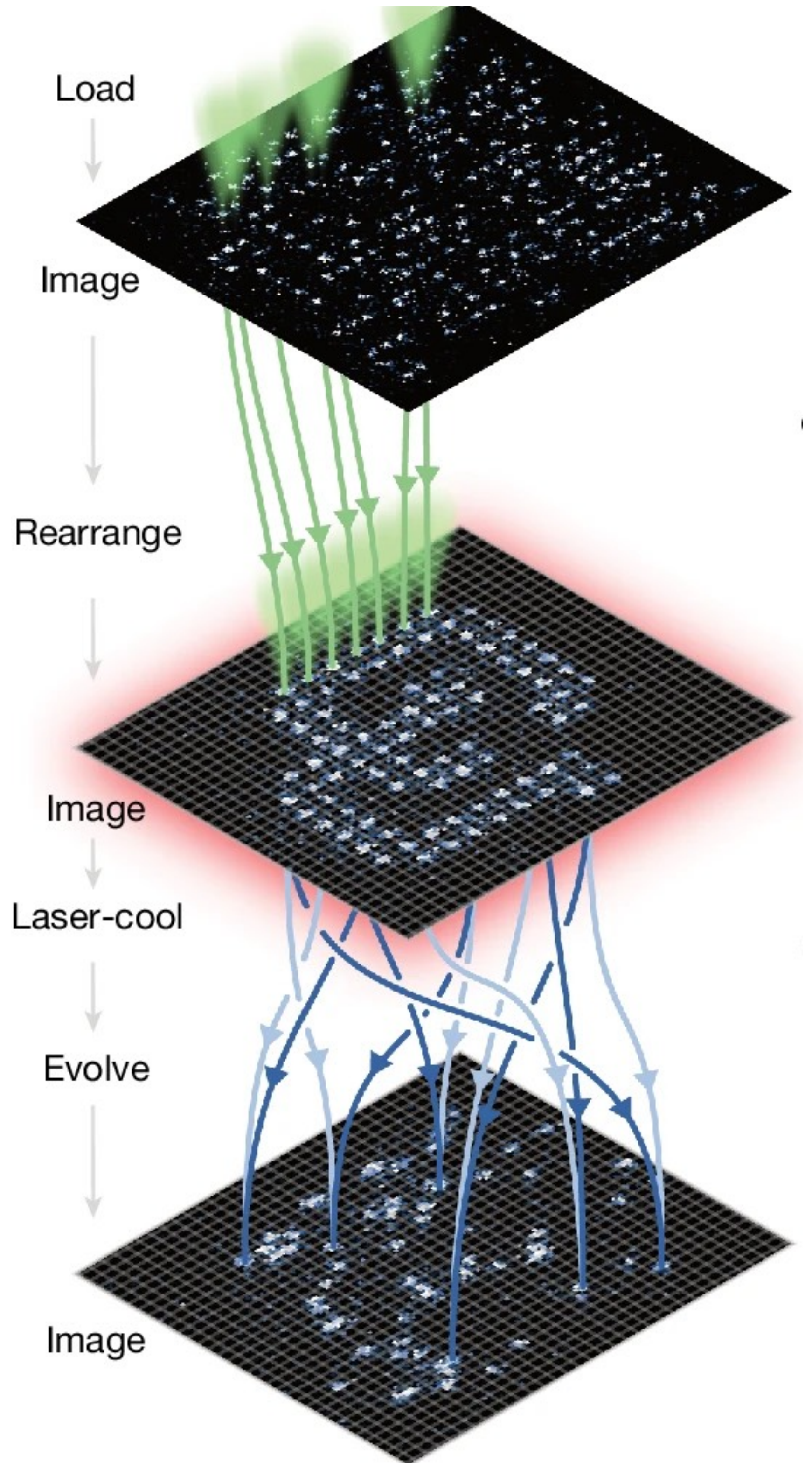
like a layer of linear transformation (Gaussian unitary)
+ nonlinear activation function (non-Gaussian measurements)

- Bose-Hubbard model in atomic system

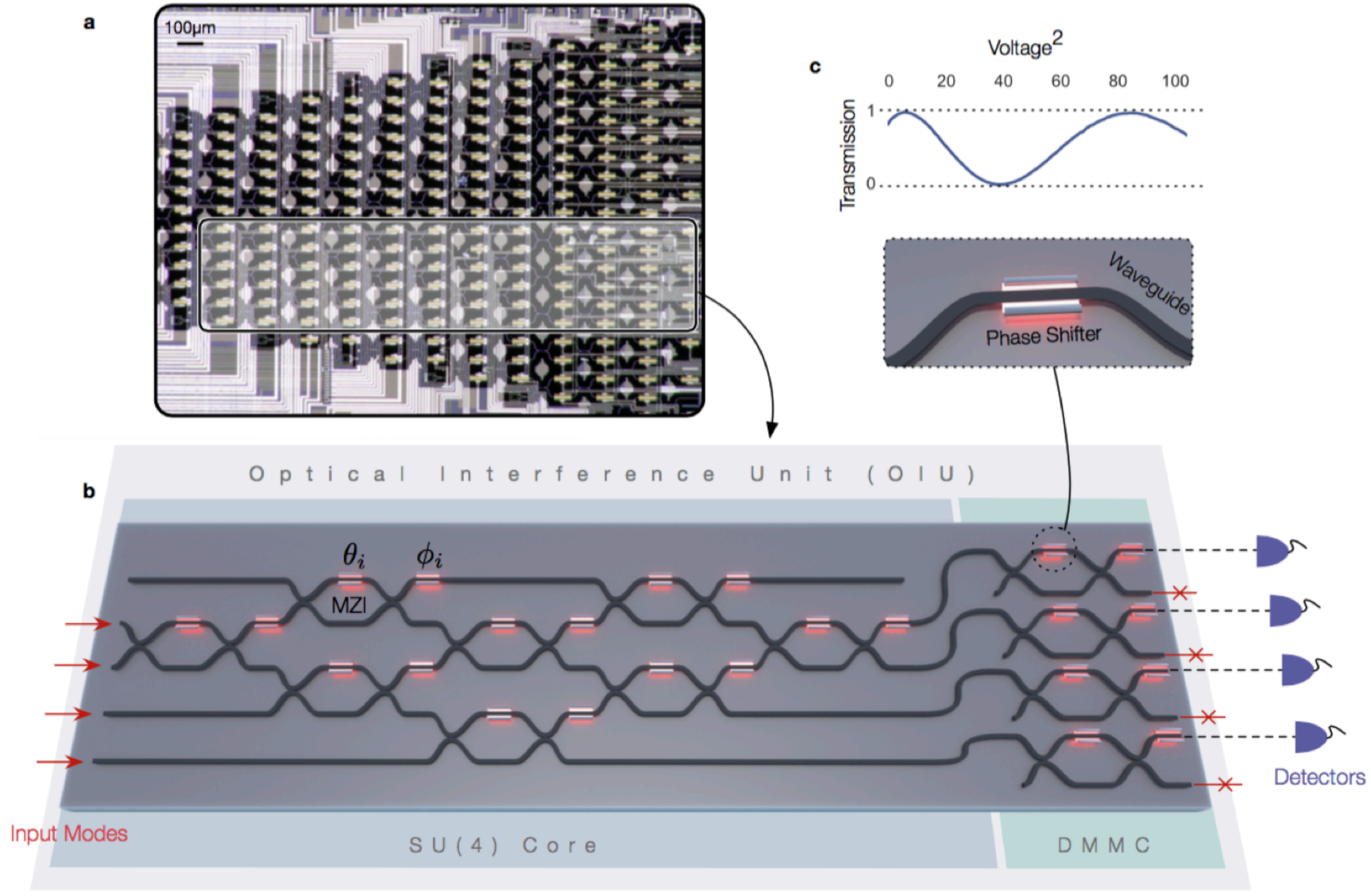
An atomic boson sampler

Aaron W. Young, Shawn Geller, William J. Eckner, Nathan Schine, Scott Glancy, Emanuel Knill & Adam M. Kaufman

Nature **629**, 311–316 (2024) | Cite this article



Photonic Neural Networks



Deep learning with coherent nanophotonic circuits

Yichen Shen, Nicholas C. Harris, Scott Skirlo, Mihika Prabhu, Tom Baehr-Jones, Michael Hochberg, Xin Sun, Shijie Zhao, Hugo Larochelle, Dirk Englund & Marin Soljačić

Nature Photonics **11**, 441–446 (2017) | Cite this article

82k Accesses | **2055** Citations | **513** Altmetric | Metrics

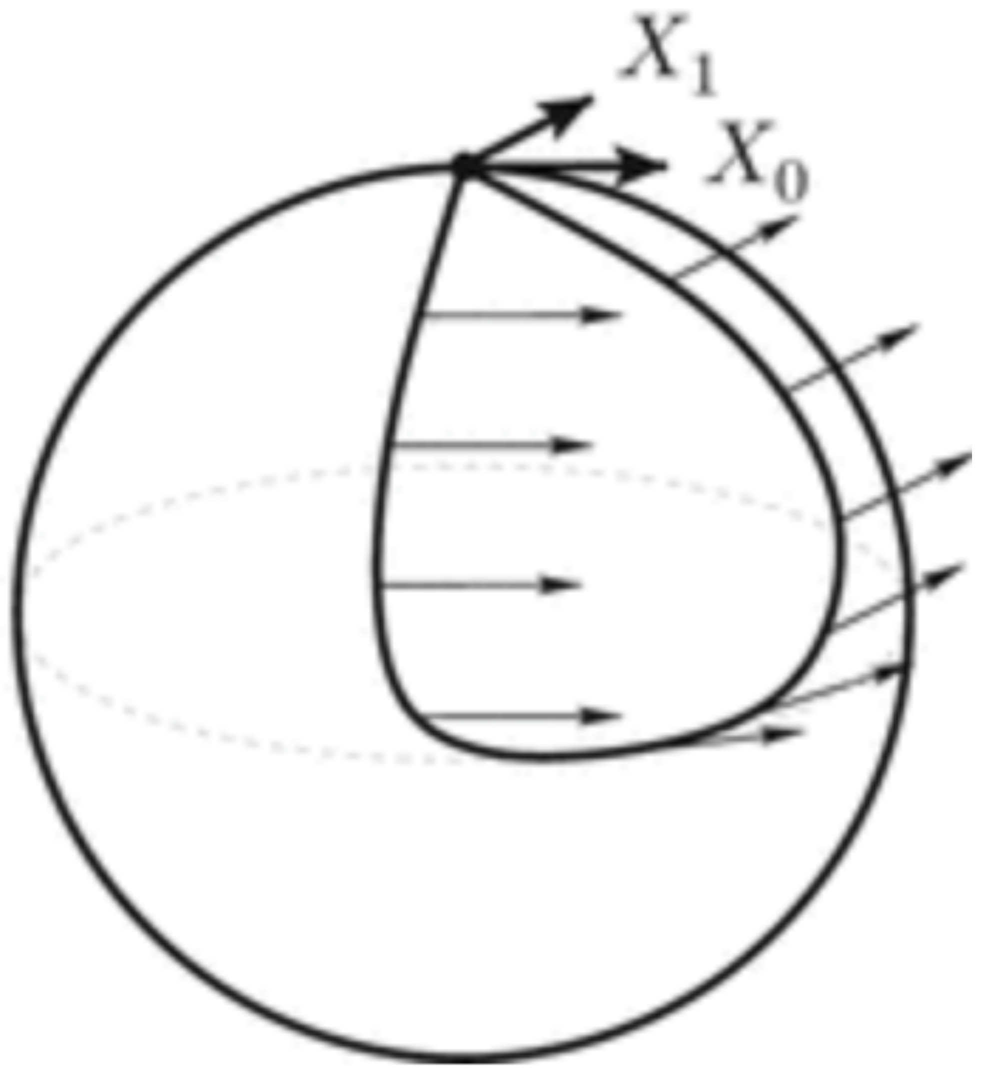
However, not easy to implement non-linear activation function (usually optical signal to electric signal then some information processing, this will destroy these two advantages)

Two advantages:
1. faster computation
2. energy-saving

Contextuality for non-linear function?

inspired from Raussendorf's result ($\mathbb{Z}_2 \rightarrow \mathbb{R}$)

Extending contextuality



Pancharatnam-Berry phase

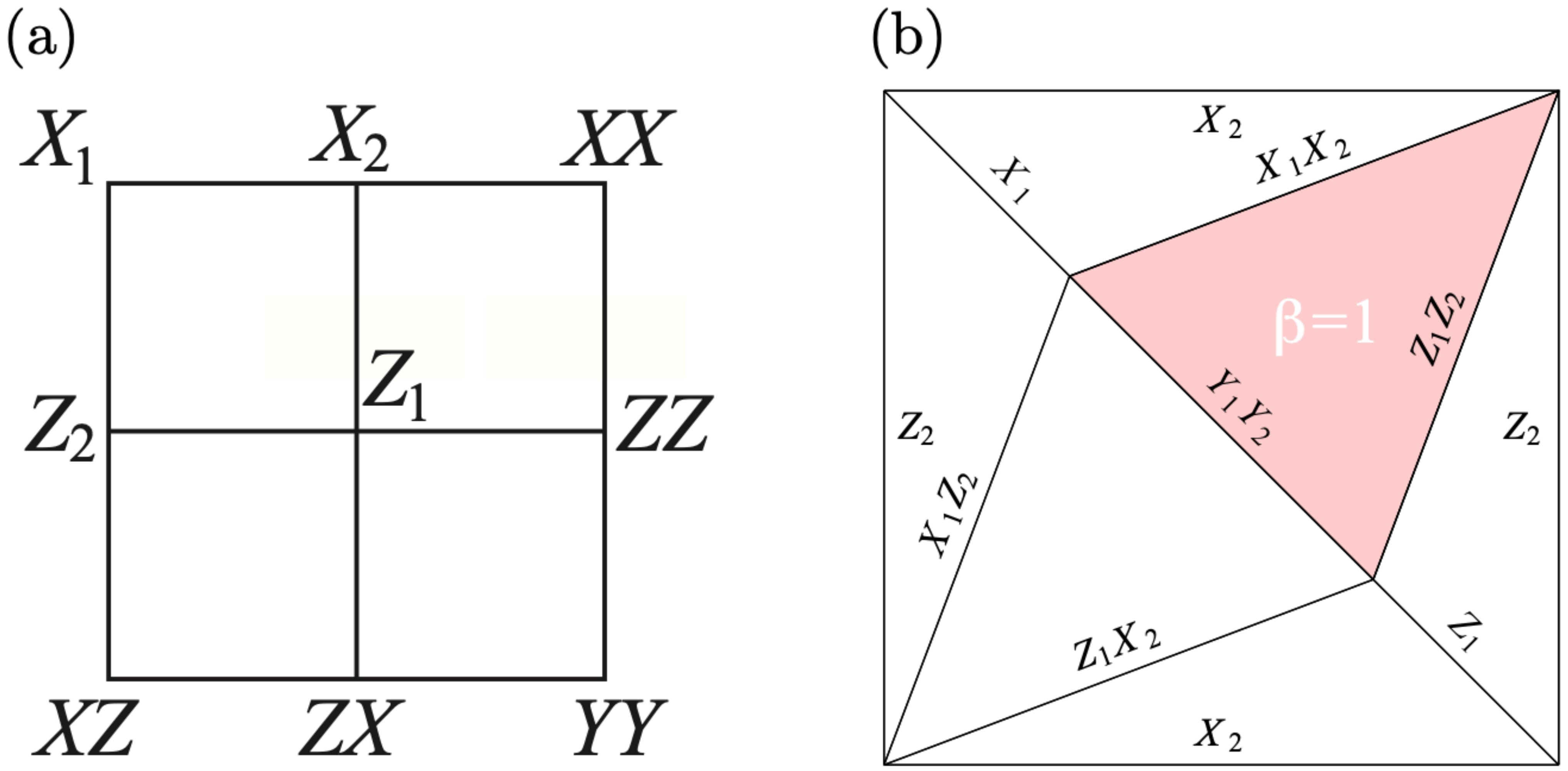
$$\phi \equiv -\mathrm{Im}\ln[\langle u_0|u_1\rangle\langle u_1|u_2\rangle\cdots\langle u_{N-1}|u_0\rangle] = -\sum_{j=0}^{N-1}\mathrm{Im}\ln\langle u_j|u_{j+1}\rangle$$

$|\tilde{u}_j\rangle = e^{-i\beta_j}|u_j\rangle$ $\quad$ β_j could be arbitrary locally

holonomy (path-dependence) vs. context-dependence

Berry Phase: inconsistency or frustration to assign phases to quantum states to observables globally although the flexibility to assign phases "locally"

Extending contextuality



Cihan Okay, Sam Roberts, Stephen D. Bartlett, and Robert Raussendorf. Topological proofs of contextuality in quantum mechanics. ArXiv:1701.01888.

Contextuality ⇔ "Chern number" $\neq 0$

Contextuality: inconsistency or frustration to assign measurement results to observables globally although the flexibility to assign measurement results "locally"

Thank you for your attention!