

QUANTUM STATES WITH NEURAL NETWORKS: REPRESENTATIONS AND TOMOGRAPHY

Cours et Séminaire Collège de France 23 mai 2023 `Réseaux de neurones, apprentissage et physique quantique'

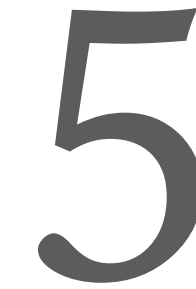
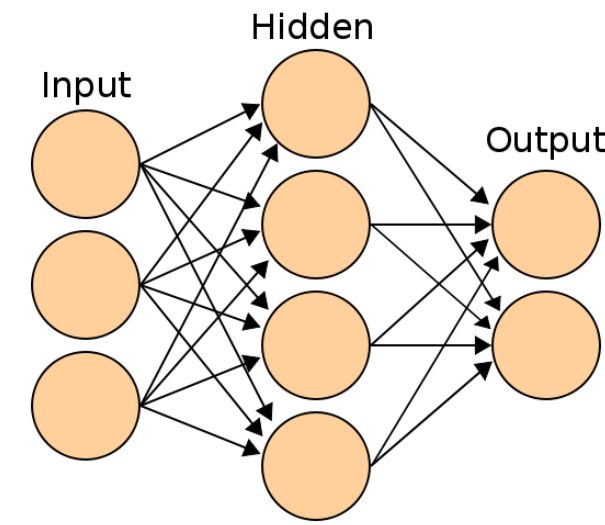
Juan Felipe Carrasquilla Álvarez
Vector Institute



Two examples of representing
quantum states with ML: **Computer
vision and natural language processing**

Computer vision

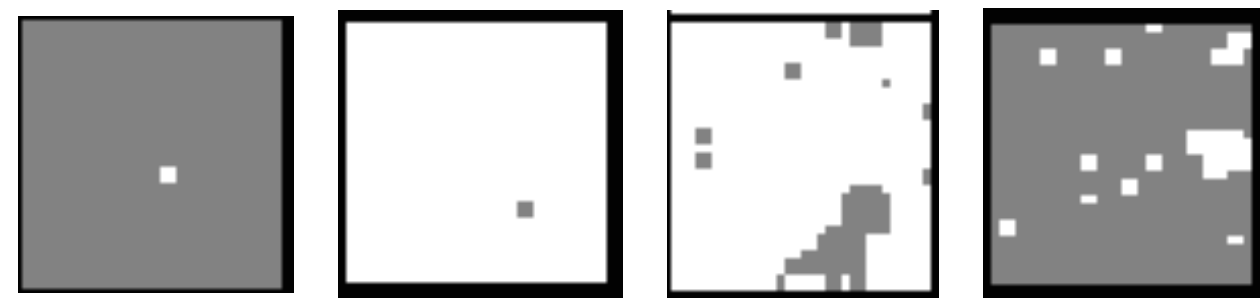
Handwritten digit classification (mnist)



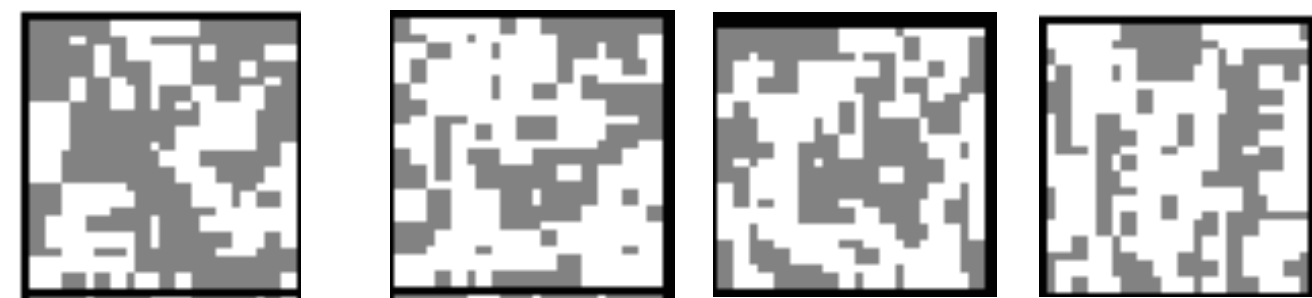
CV community has developed powerful supervised learning algorithms based on neural networks

= 5 + fluctuations

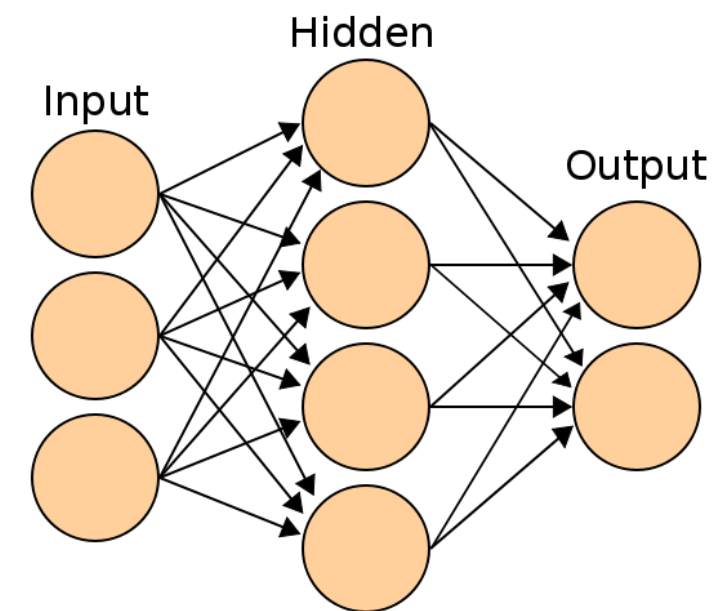
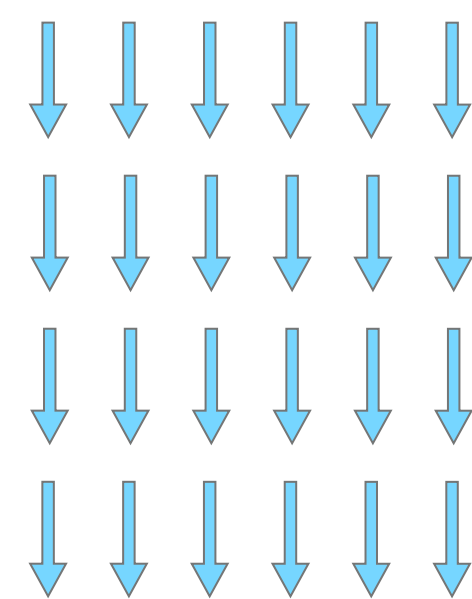
FM phase



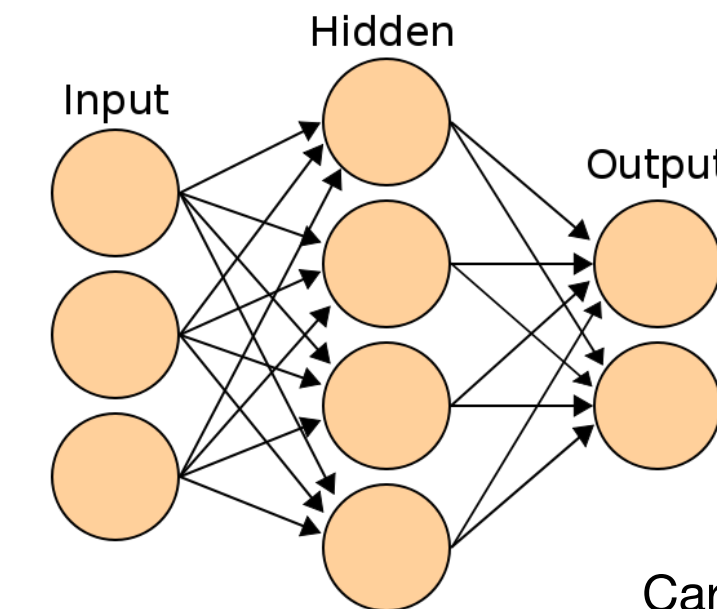
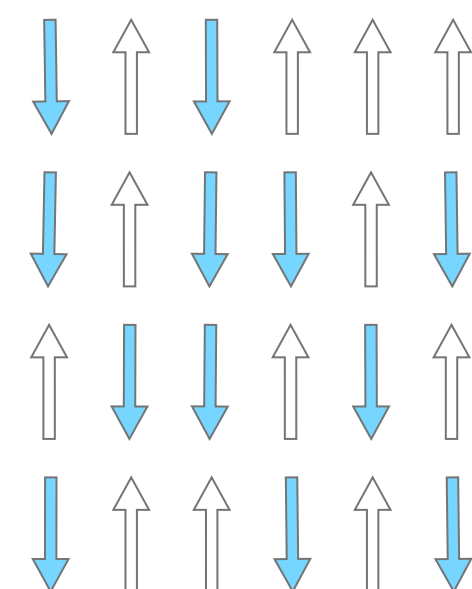
High T phase



grey = spin up
white = spin down



FM (0)



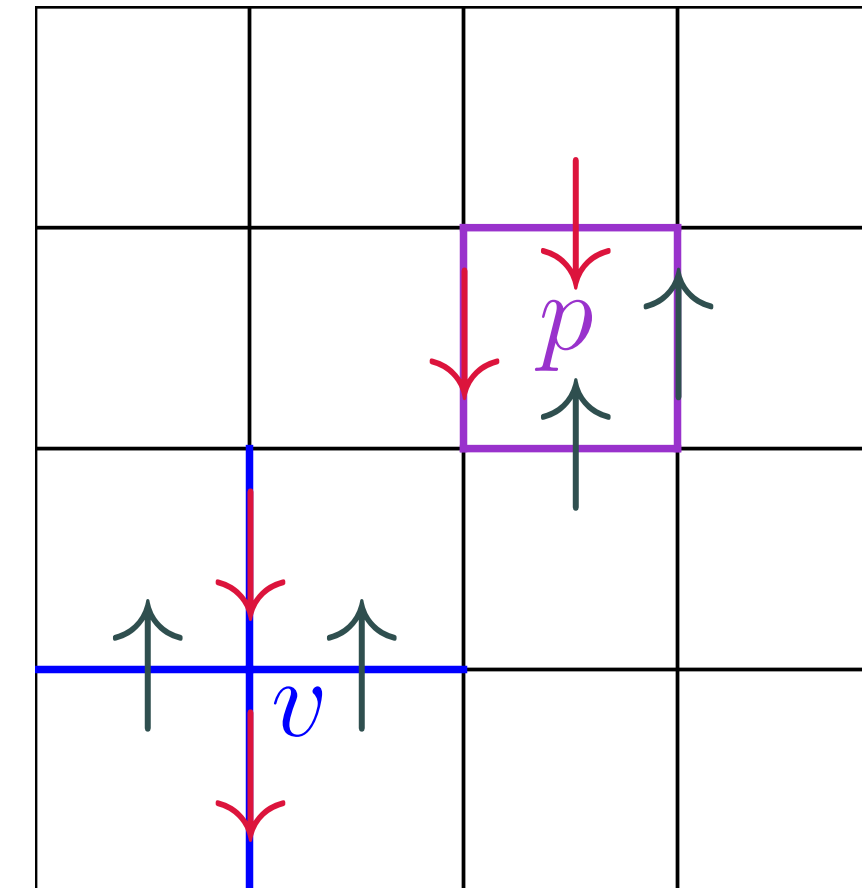
PM (1)

Wegner's Ising gauge theory

$$H = -J \sum_p \prod_{i \in p} \sigma_i^z$$

F.J. Wegner, J. Math. Phys. 12 (1971) 2259
(Kogut Rev. Mod. Phys. 51, 659 (1979))

The ground state is a highly degenerate manifold with short-range spin-spin correlations. Ground states form a soup of strings of aligned spins



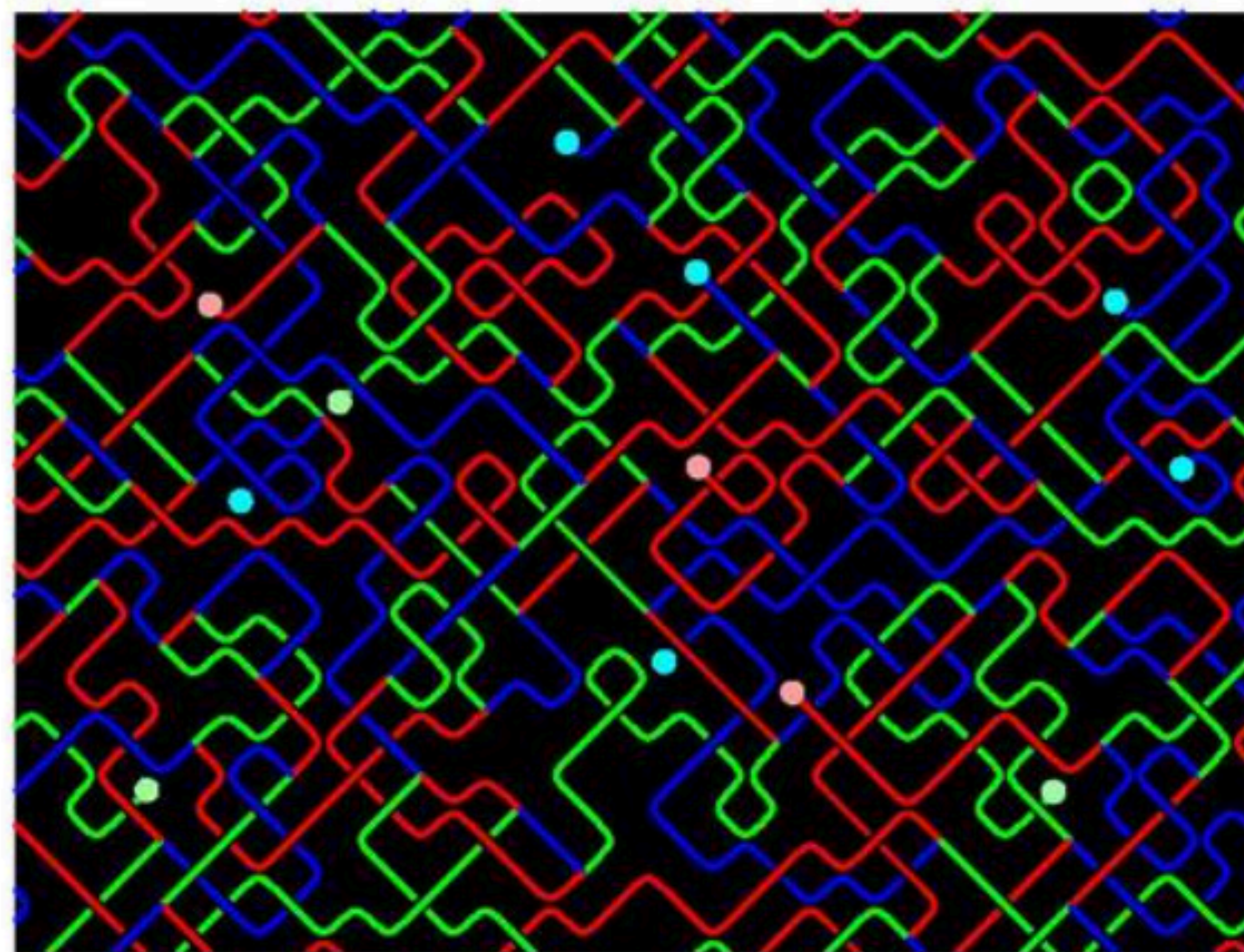
The grandmother of most lattice models for topological quantum computation



Castelnovo and Chamon Phys. Rev. B 76, 174416 (2007)

Wegner's Ising gauge theory

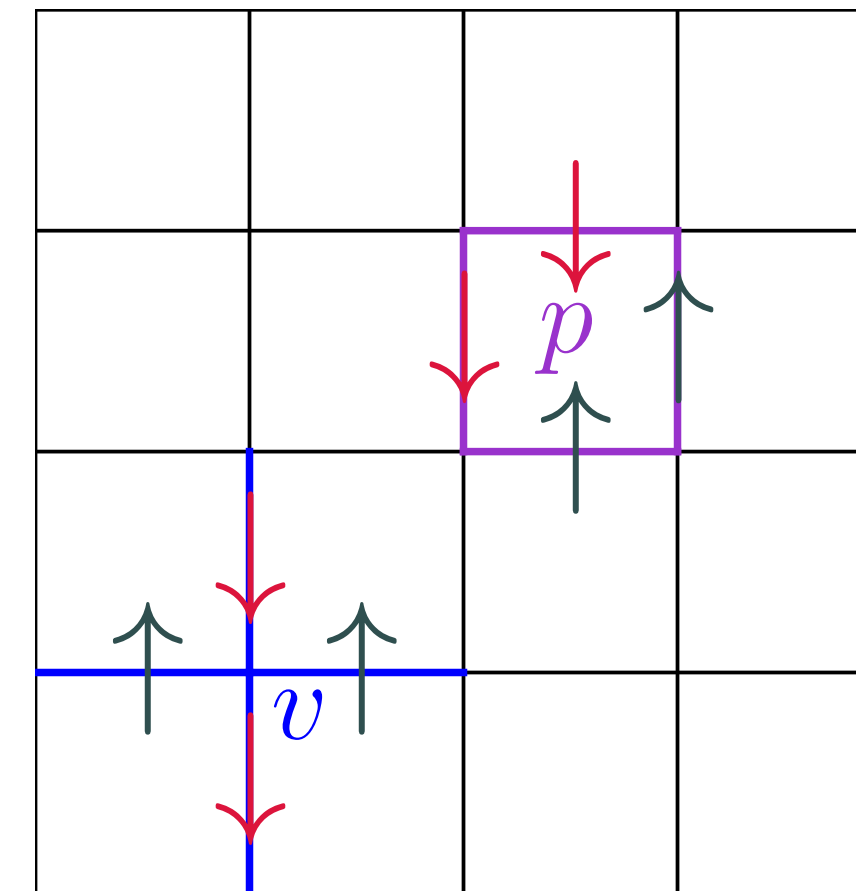
$$H = -J \sum_p \prod_{i \in p} \sigma_i^z$$



A picture of our vacua

A string-net theory of light and electrons

Levin-Wen cond-mat/0404617

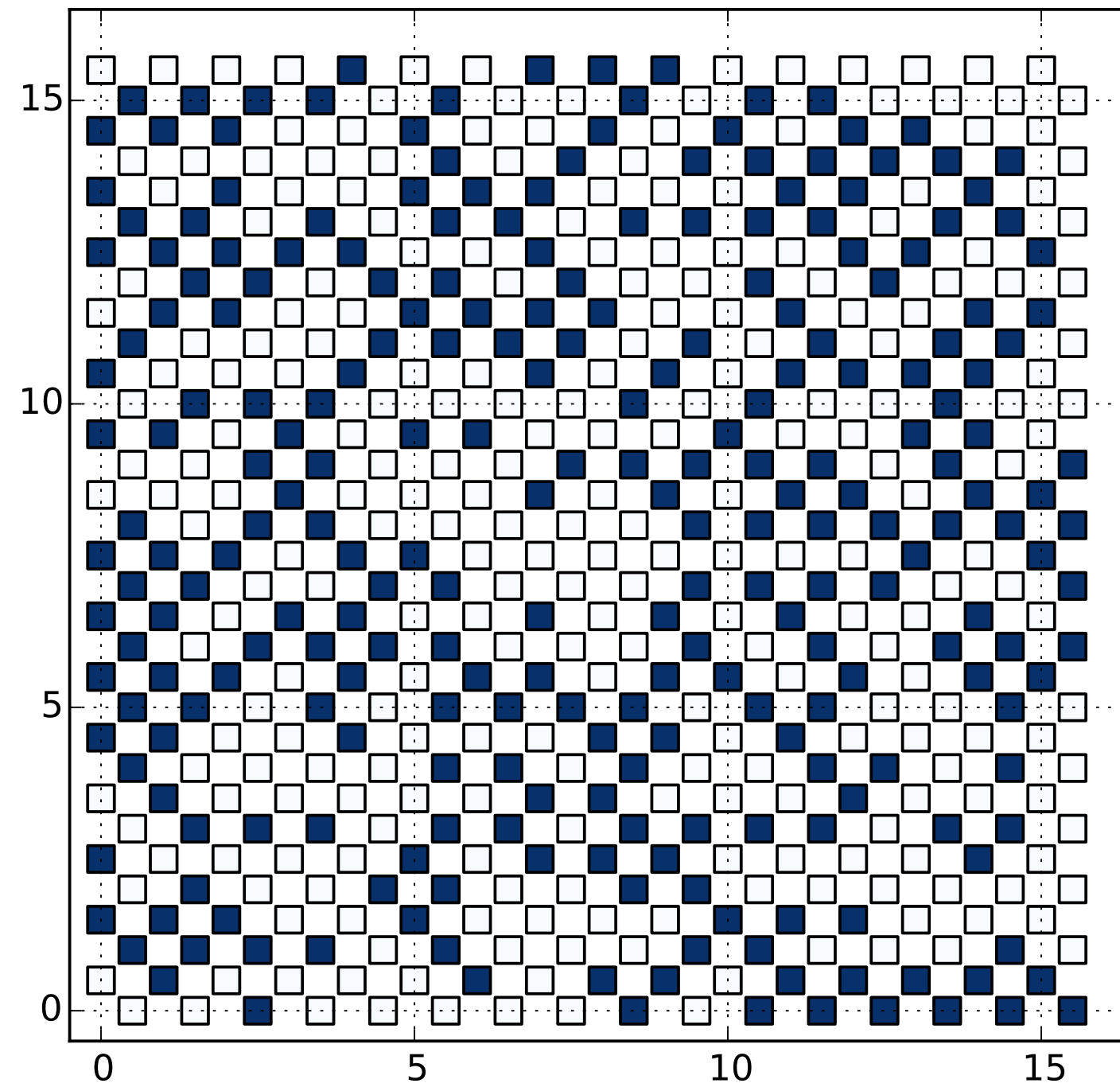
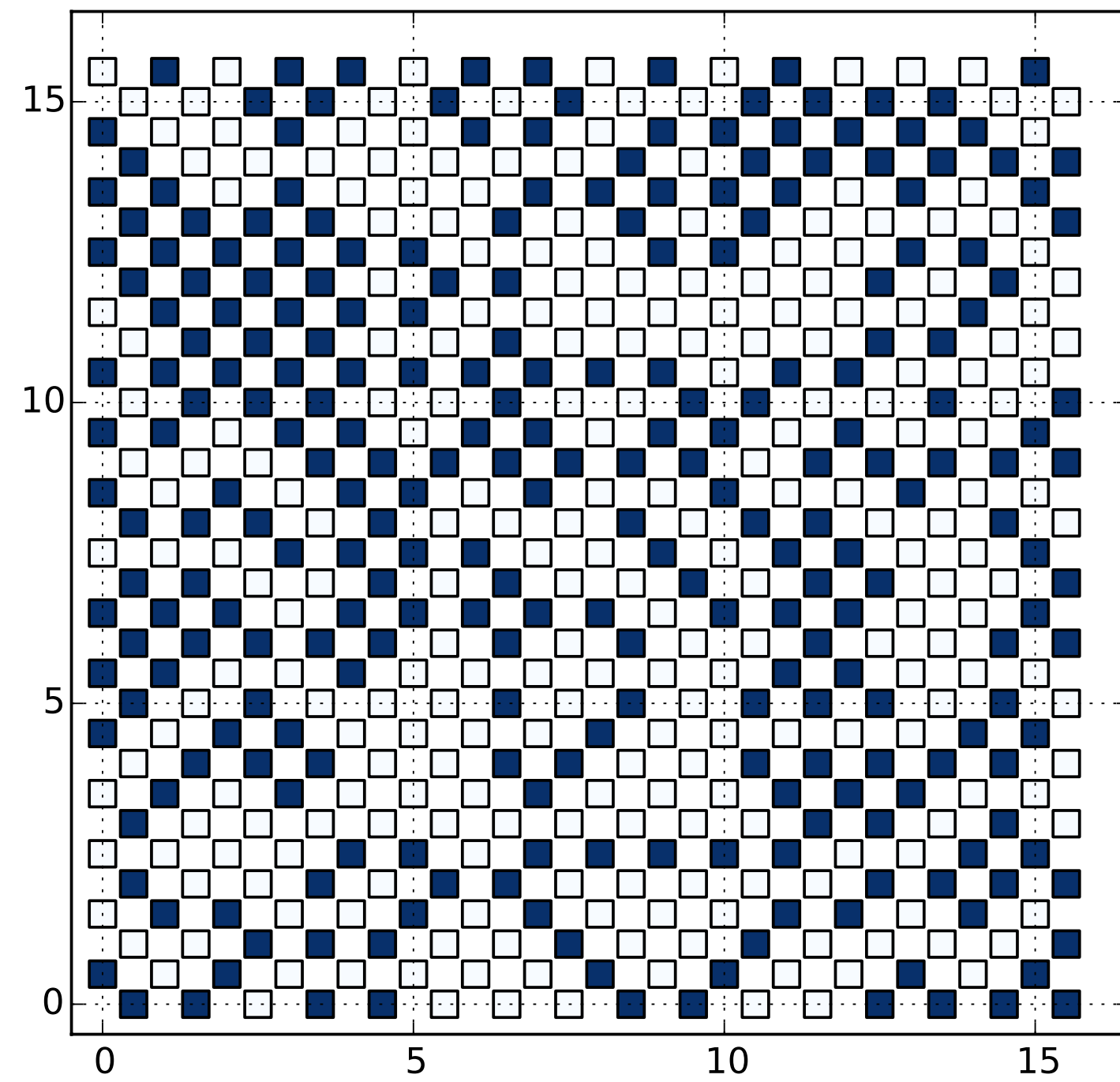


The grandmother of most lattice models for topological quantum computation

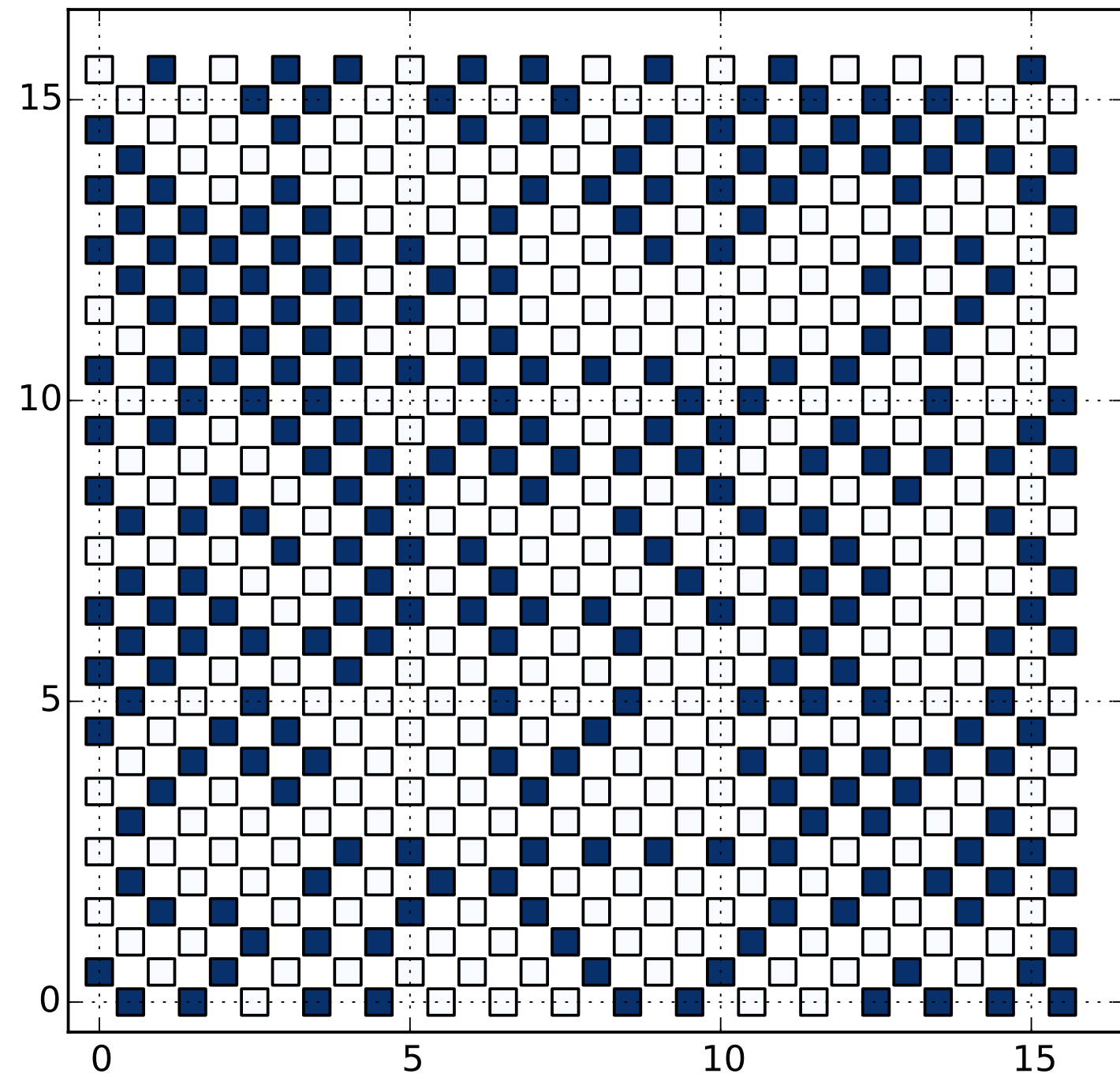


Castelnovo and Chamon Phys. Rev. B 76, 174416 (2007)

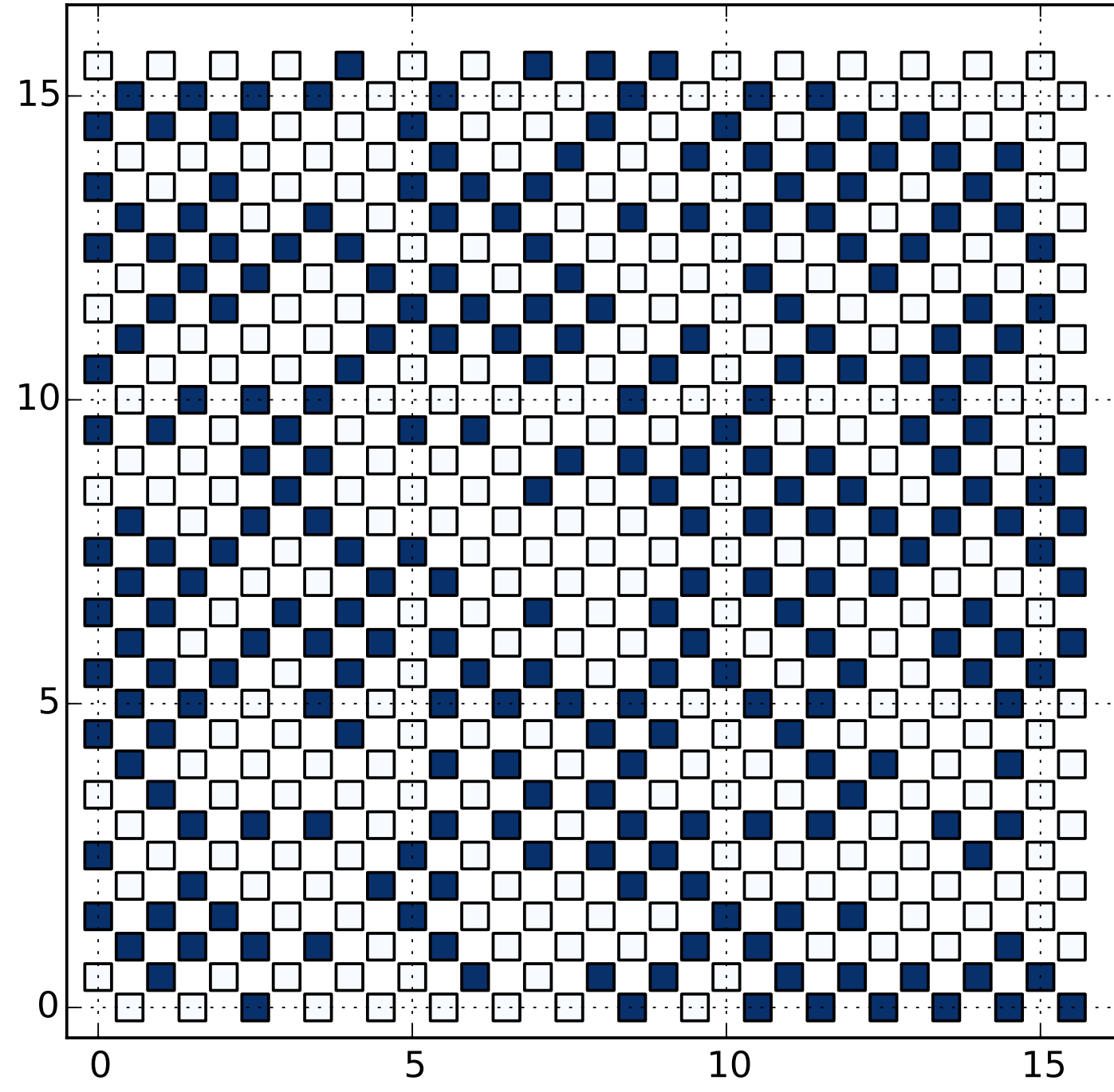
Two configurations



For two configurations



Ground state



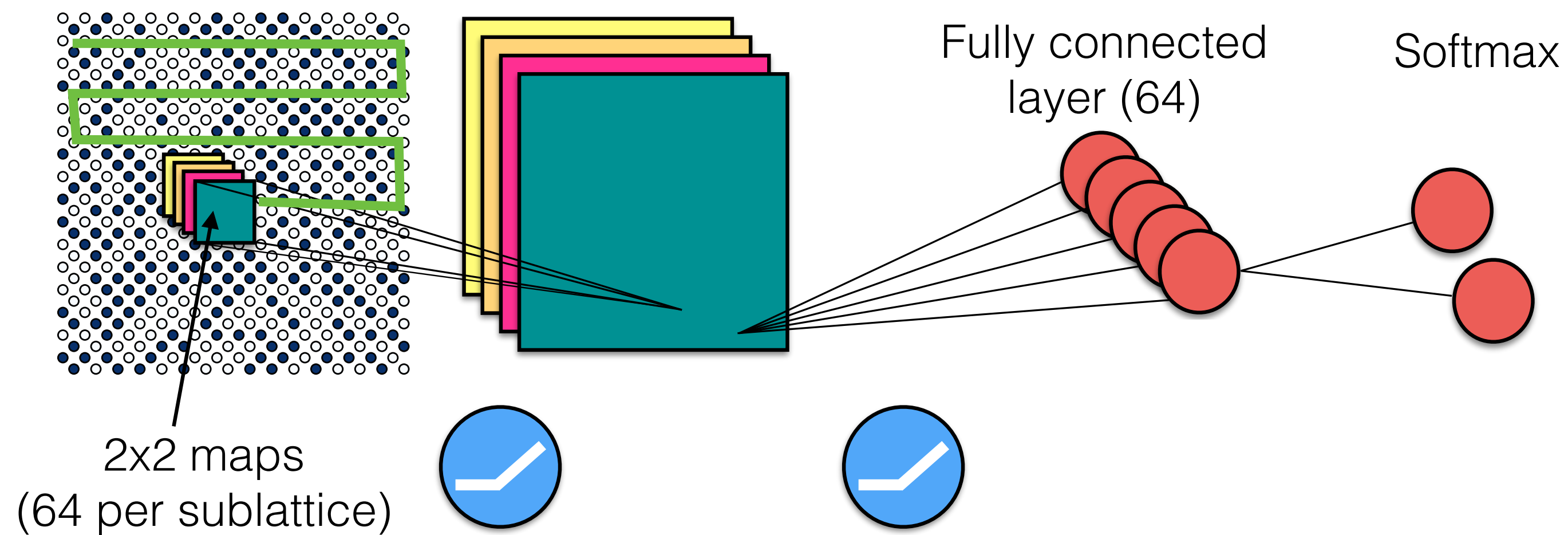
high-temperature state

Feedforward NN are difficult to apply to this problem and lead to 50% accuracy

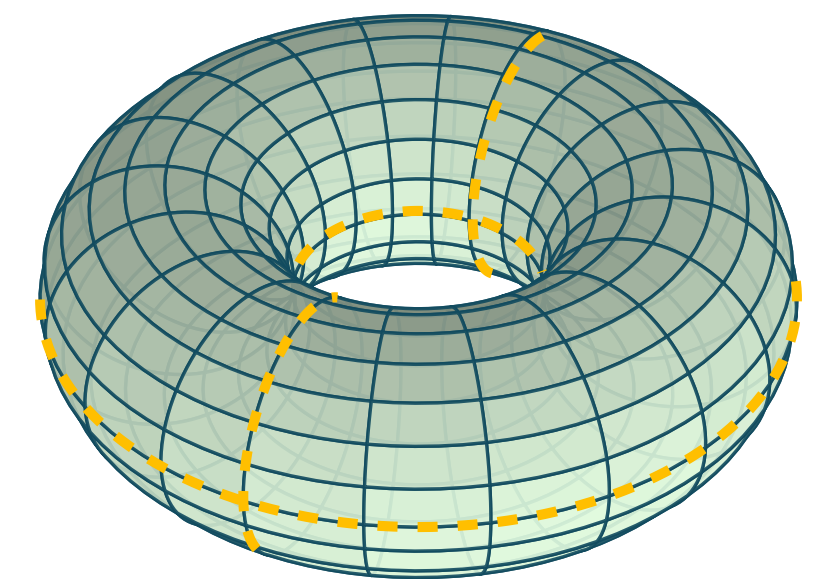
Ising gauge theory

F.J. Wegner, J. Math. Phys. 12 (1971) 2259

$$H = -J \sum_p \prod_{i \in p} \sigma_i^z$$



*99% accuracy
*easy to train**

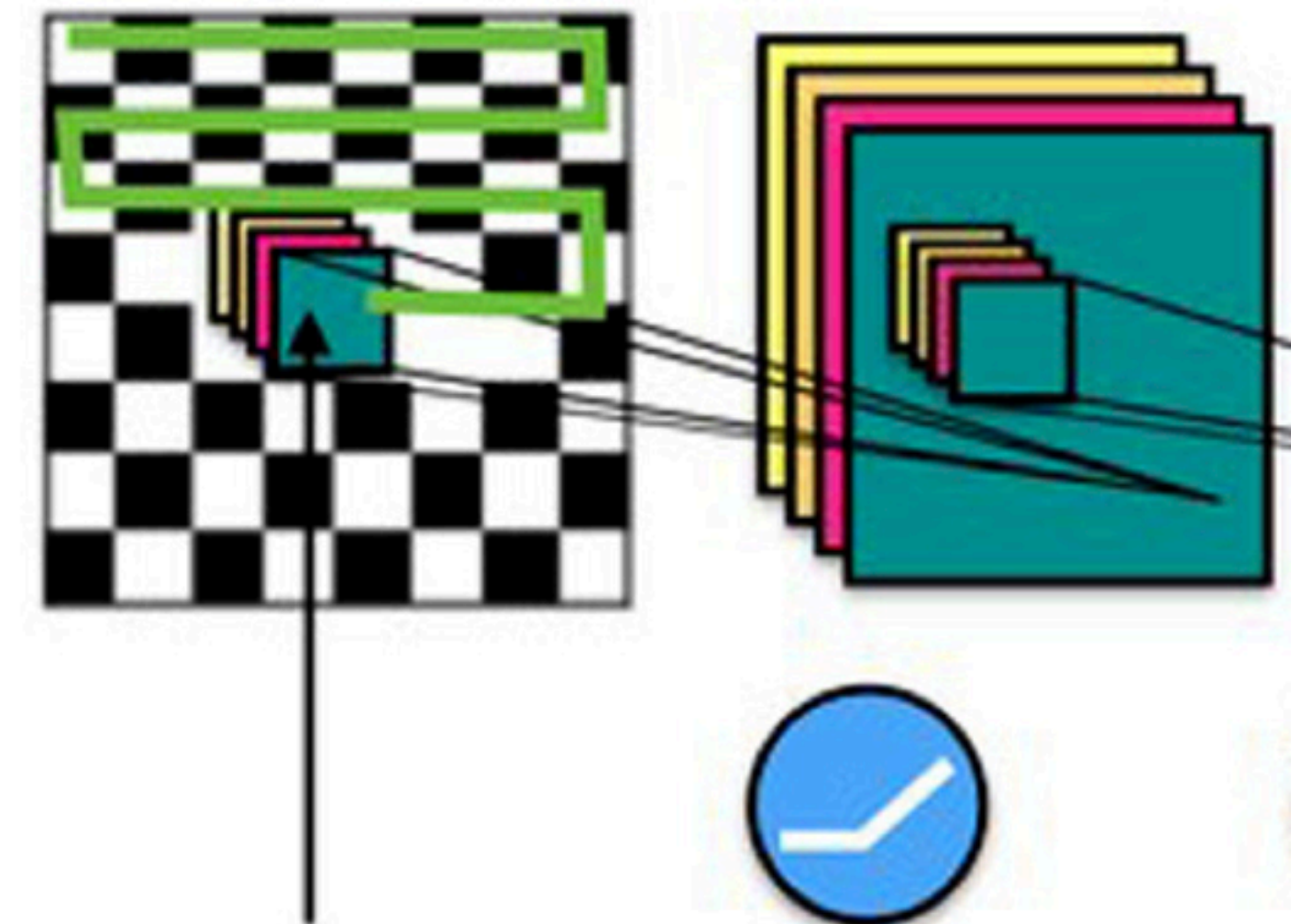


Convolutional neural network

Convolutional neural networks

- CNNs use a mathematical operation called “convolution” to process information for data that have a natural gridlike topology
- The trainable kernel $K_{i,l,m_y,m_x}^{(q)}$ (layer q , input/output channels l,i spatial shifts along y and x directions, F is a non-linearity)

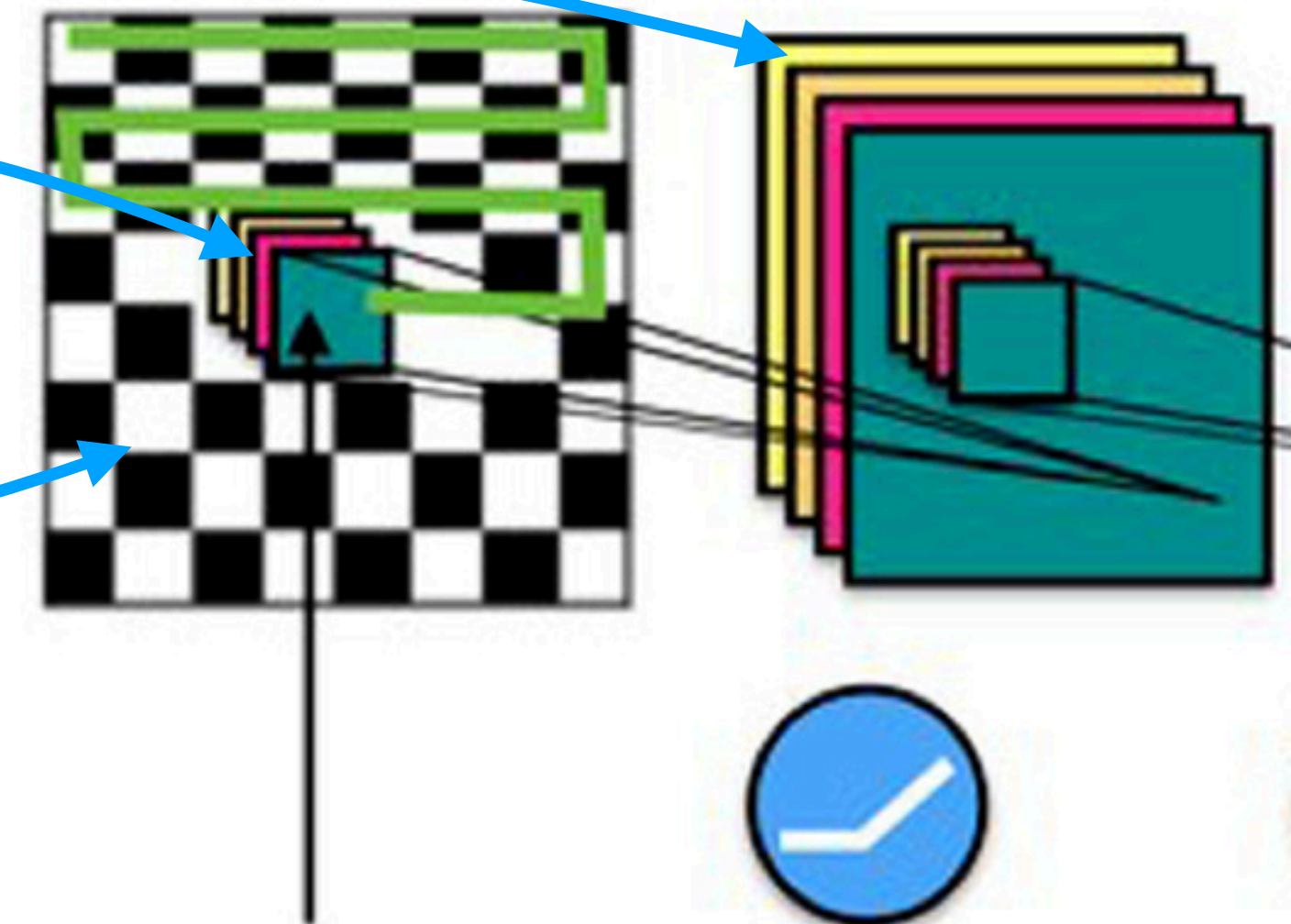
$$h_{i,j,k}^{(q)} = F \left(\sum_{l,m_y,m_x} h_{l,j+m_y,k+m_x}^{(q-1)} K_{i,l,m_y,m_x}^{(q)} \right)$$
$$:= F (K^{(q)} * h^{(q-1)}),$$



Convolutional neural networks

- CNNs use a mathematical operation called “convolution” to process information for data that have a natural gridlike topology
- The trainable kernel $K_{i,l,m_y,m_x}^{(q)}$ (layer q , input/output channels l,i , spatial shifts along y and x directions, F is a non-linearity)

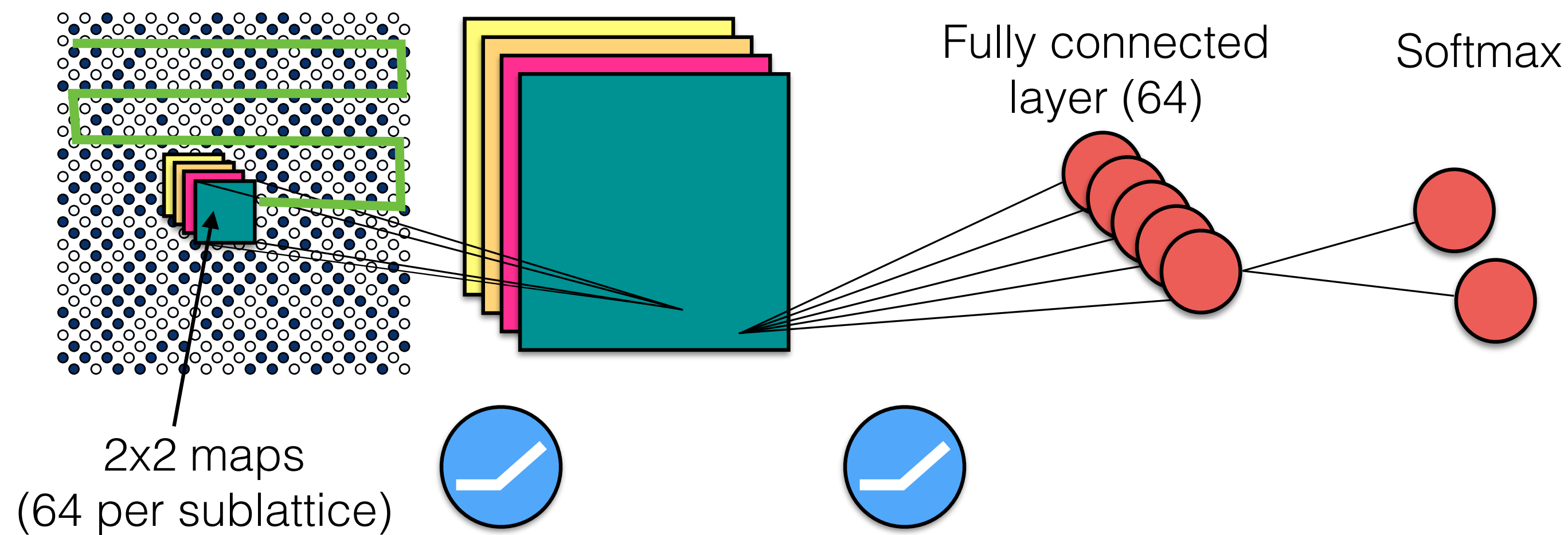
$$h_{i,j,k}^{(q)} = F \left(\sum_{l,m_y,m_x} h_{l,j+m_y,k+m_x}^{(q-1)} K_{i,l,m_y,m_x}^{(q)} \right)$$
$$:= F (K^{(q)} * h^{(q-1)}),$$



Ising gauge theory

F.J. Wegner, J. Math. Phys. 12 (1971) 2259

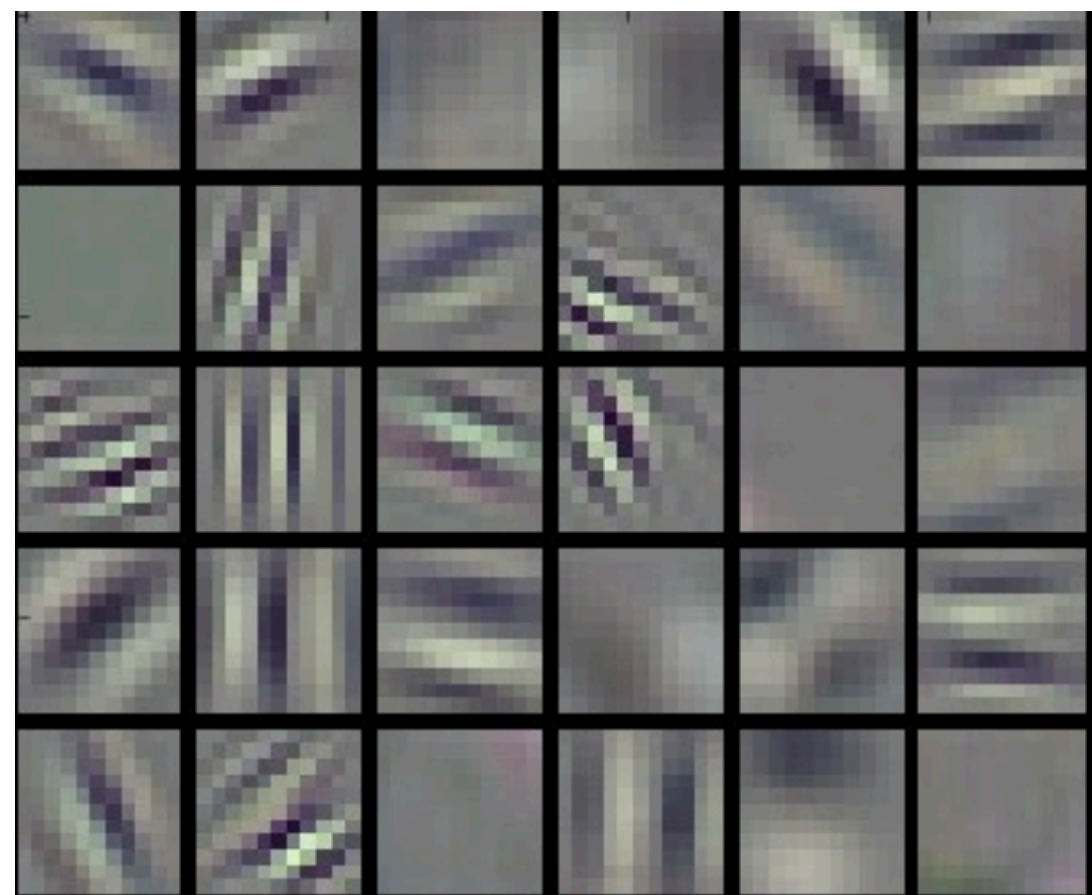
$$H = -J \sum_p \prod_{i \in p} \sigma_i^z$$



The picture we draw for what the CNN is using to distinguish the phases is that of the detection of satisfied local constraints. In few words, the neural network figures out a proxy for the energy and uses it to classify states

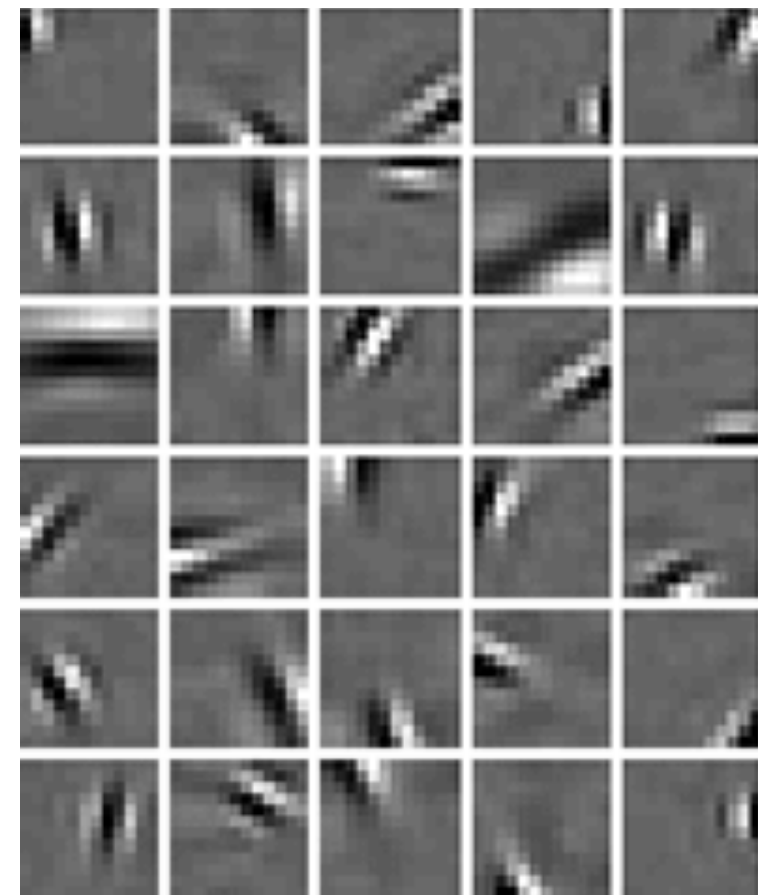
Connection with computer vision

- Convolutional neural networks revolutionized computer vision — beat humans at classifying images since 2015

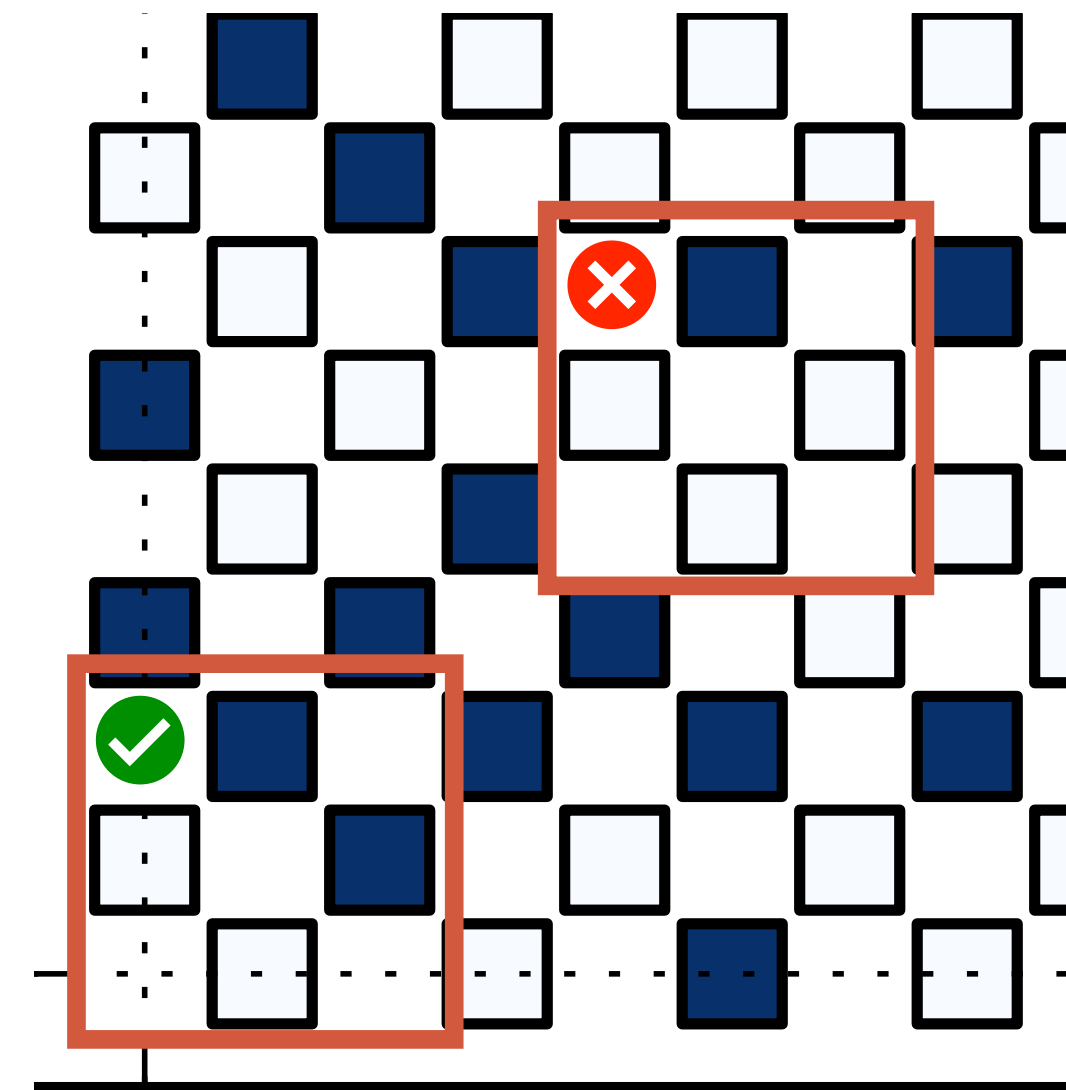


Filters used by the 1st convolutional layer

<https://cs231n.github.io/understanding-cnn/>



Gabor filters, believed to exist in visual processing in the brain (Olshausen & Field, 1997)



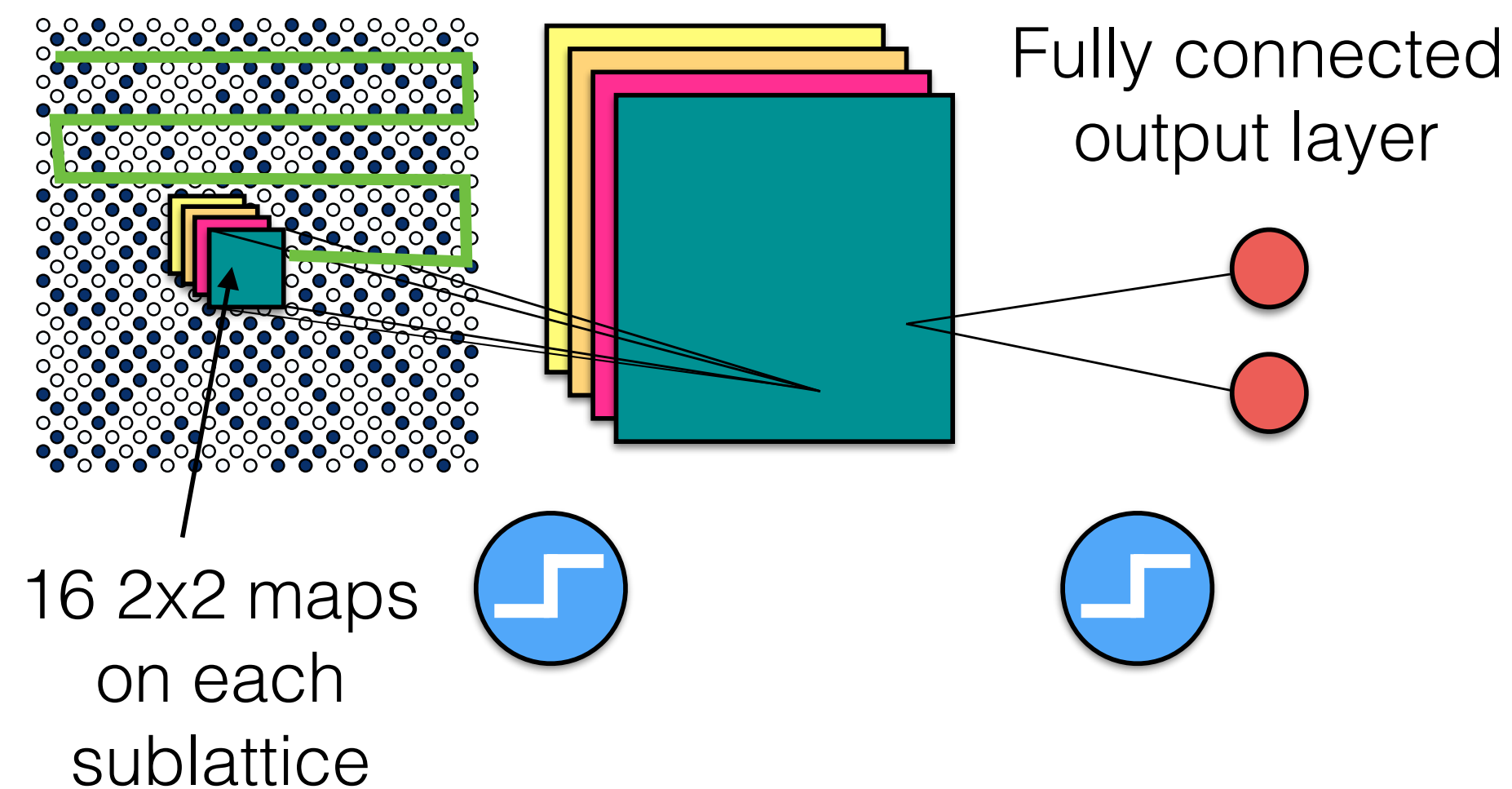
In our examples, they see magnetic monopoles or in spin ice, they see the ice rules

$$H = -J \sum_p \prod_{i \in p} \sigma_i^z$$

Analytical understanding: What does the CNN use to make predictions?

- Based on this observation we derived the weights of a streamlined convolutional network analytically designed to work well.

$$O_{\text{cold}}(\sigma_1, \dots, \sigma_N) \propto \lim_{\beta \rightarrow \infty} \exp \beta J \sum_p \prod_{i \in p} \sigma_i^z$$



100% accuracy on test sets

Analytical model for the Ising gauge theory

Convolutional layer

f	s=A	s=B	f	s=A	s=B
1	$\begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}$	$\begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}$	9	$\begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}$	$\begin{pmatrix} 1 & -1 \\ 0 & 0 \end{pmatrix}$
2	$\begin{pmatrix} -1 & 0 \\ -1 & 0 \end{pmatrix}$	$\begin{pmatrix} -1 & -1 \\ 0 & 0 \end{pmatrix}$	10	$\begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}$	$\begin{pmatrix} -1 & 1 \\ 0 & 0 \end{pmatrix}$
3	$\begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}$	$\begin{pmatrix} -1 & -1 \\ 0 & 0 \end{pmatrix}$	11	$\begin{pmatrix} 1 & 0 \\ -1 & 0 \end{pmatrix}$	$\begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}$
4	$\begin{pmatrix} 1 & 0 \\ -1 & 0 \end{pmatrix}$	$\begin{pmatrix} -1 & 1 \\ 0 & 0 \end{pmatrix}$	12	$\begin{pmatrix} -1 & 0 \\ 1 & 0 \end{pmatrix}$	$\begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}$
5	$\begin{pmatrix} -1 & 0 \\ -1 & 0 \end{pmatrix}$	$\begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}$	13	$\begin{pmatrix} -1 & 0 \\ -1 & 0 \end{pmatrix}$	$\begin{pmatrix} -1 & 1 \\ 0 & 0 \end{pmatrix}$
6	$\begin{pmatrix} -1 & 0 \\ 1 & 0 \end{pmatrix}$	$\begin{pmatrix} 1 & -1 \\ 0 & 0 \end{pmatrix}$	14	$\begin{pmatrix} -1 & 0 \\ -1 & 0 \end{pmatrix}$	$\begin{pmatrix} 1 & -1 \\ 0 & 0 \end{pmatrix}$
7	$\begin{pmatrix} 1 & 0 \\ -1 & 0 \end{pmatrix}$	$\begin{pmatrix} 1 & -1 \\ 0 & 0 \end{pmatrix}$	15	$\begin{pmatrix} -1 & 0 \\ 1 & 0 \end{pmatrix}$	$\begin{pmatrix} -1 & -1 \\ 0 & 0 \end{pmatrix}$
8	$\begin{pmatrix} -1 & 0 \\ 1 & 0 \end{pmatrix}$	$\begin{pmatrix} -1 & 1 \\ 0 & 0 \end{pmatrix}$	16	$\begin{pmatrix} -1 & 0 \\ -1 & 0 \end{pmatrix}$	$\begin{pmatrix} -1 & -1 \\ 0 & 0 \end{pmatrix}$

$$b_c = -(2 + \epsilon) \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}$$

fully-connected layer

$$W_o = \begin{pmatrix} \underbrace{1 \dots 1}_{8L^2 \text{ terms}} & \underbrace{-L^2 \dots -L^2}_{8L^2 \text{ terms}} \\ -1 \dots -1 & L^2 \dots L^2 \end{pmatrix}, \quad \text{and } b_o = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

The purpose of the filters is to individually process each plaquette in the spin configuration and determine whether its energetic constraints are satisfied or not. Basically the Conv. layer encodes the Hamiltonian

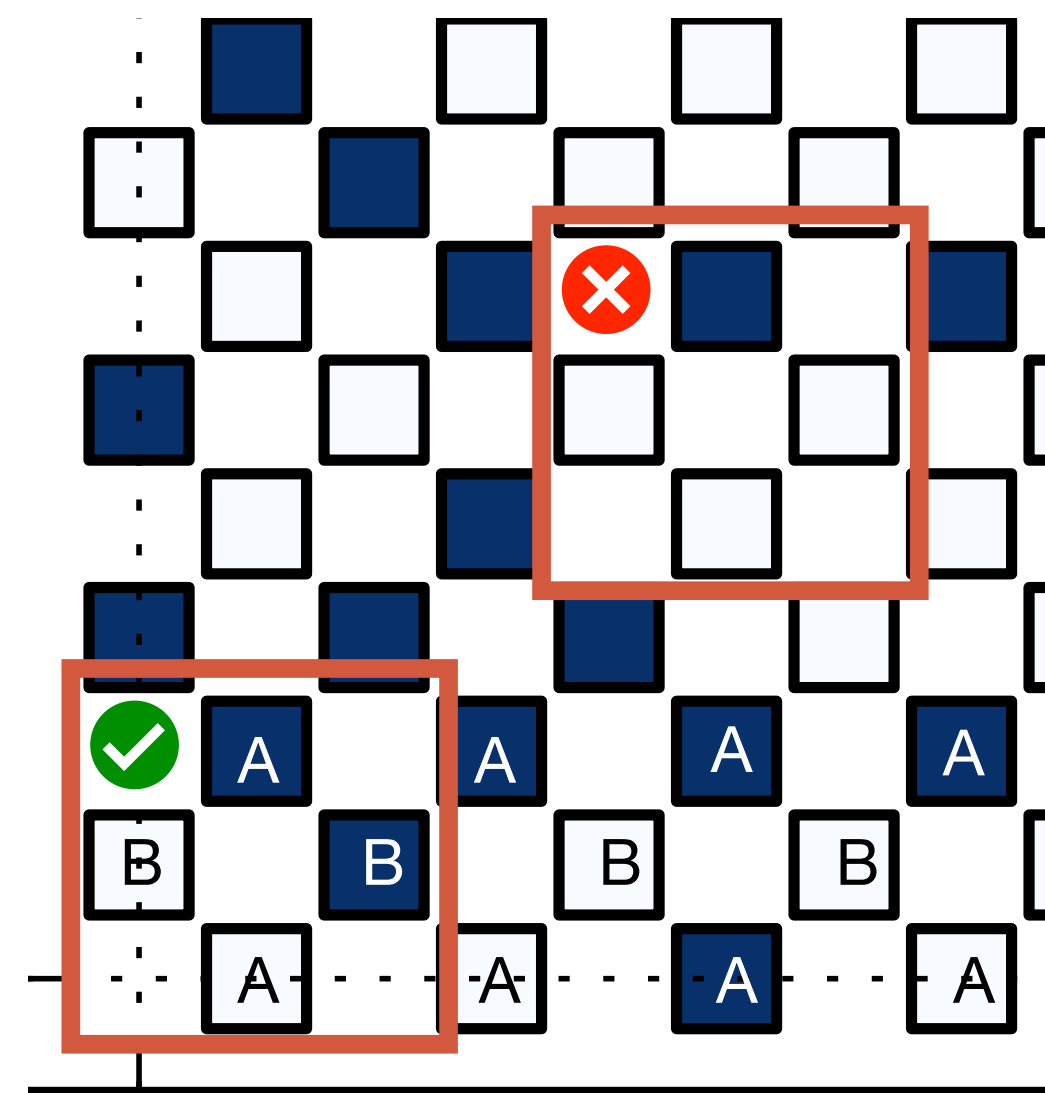
Analytical model for the Ising gauge theory

Convolutional layer

f	s=A	s=B	f	s=A	s=B
1	$\begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}$	$\begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}$	9	$\begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}$	$\begin{pmatrix} 1 & -1 \\ 0 & 0 \end{pmatrix}$
2	$\begin{pmatrix} -1 & 0 \\ -1 & 0 \end{pmatrix}$	$\begin{pmatrix} -1 & -1 \\ 0 & 0 \end{pmatrix}$	10	$\begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}$	$\begin{pmatrix} -1 & 1 \\ 0 & 0 \end{pmatrix}$
3	$\begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}$	$\begin{pmatrix} -1 & -1 \\ 0 & 0 \end{pmatrix}$	11	$\begin{pmatrix} 1 & 0 \\ -1 & 0 \end{pmatrix}$	$\begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}$
4	$\begin{pmatrix} 1 & 0 \\ -1 & 0 \end{pmatrix}$	$\begin{pmatrix} -1 & 1 \\ 0 & 0 \end{pmatrix}$	12	$\begin{pmatrix} -1 & 0 \\ 1 & 0 \end{pmatrix}$	$\begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}$
5	$\begin{pmatrix} -1 & 0 \\ -1 & 0 \end{pmatrix}$	$\begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}$	13	$\begin{pmatrix} -1 & 0 \\ -1 & 0 \end{pmatrix}$	$\begin{pmatrix} -1 & 1 \\ 0 & 0 \end{pmatrix}$
6	$\begin{pmatrix} -1 & 0 \\ 1 & 0 \end{pmatrix}$	$\begin{pmatrix} 1 & -1 \\ 0 & 0 \end{pmatrix}$	14	$\begin{pmatrix} -1 & 0 \\ -1 & 0 \end{pmatrix}$	$\begin{pmatrix} 1 & -1 \\ 0 & 0 \end{pmatrix}$
7	$\begin{pmatrix} 1 & 0 \\ -1 & 0 \end{pmatrix}$	$\begin{pmatrix} 1 & -1 \\ 0 & 0 \end{pmatrix}$	15	$\begin{pmatrix} -1 & 0 \\ 1 & 0 \end{pmatrix}$	$\begin{pmatrix} -1 & -1 \\ 0 & 0 \end{pmatrix}$
8	$\begin{pmatrix} -1 & 0 \\ 1 & 0 \end{pmatrix}$	$\begin{pmatrix} -1 & 1 \\ 0 & 0 \end{pmatrix}$	16	$\begin{pmatrix} -1 & 0 \\ -1 & 0 \end{pmatrix}$	$\begin{pmatrix} -1 & -1 \\ 0 & 0 \end{pmatrix}$

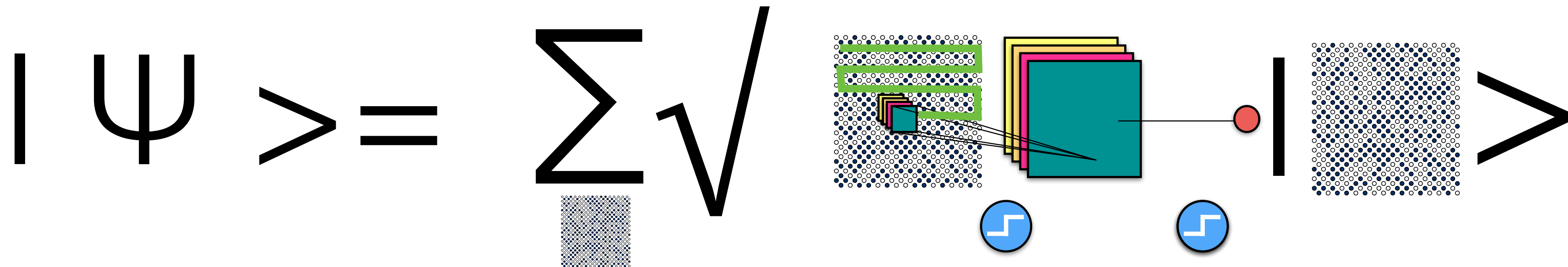
$$b_c = -(2 + \epsilon) \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}$$

fully-connected layer



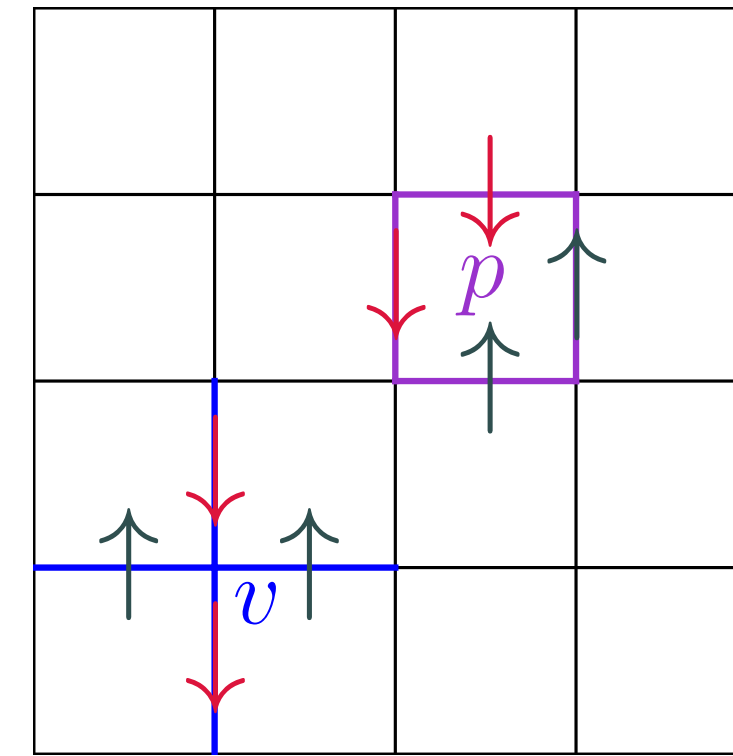
The purpose of the filters is to individually process each plaquette in the spin configuration and determine whether its energetic constraints are satisfied or not. The final layer counts good plaquettes

Connection to quantum states?



Ground state of the toric code

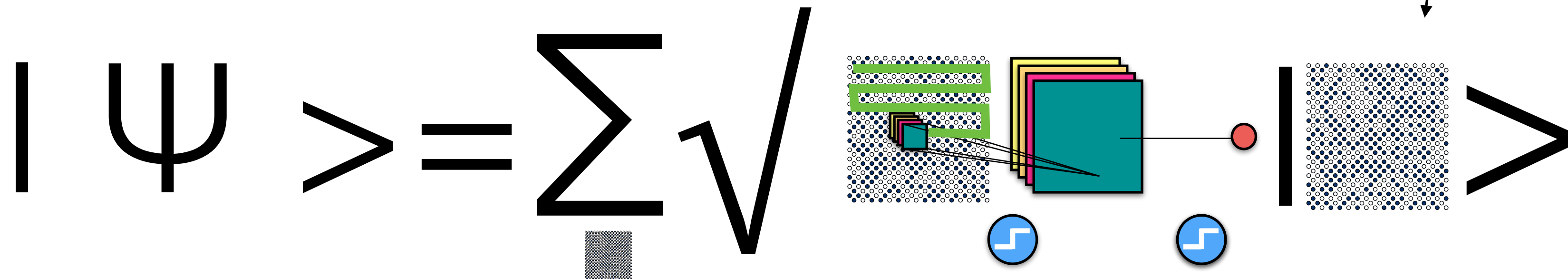
$$H = -J_p \sum_p \prod_{i \in p} \sigma_i^z - J_v \sum_v \prod_{i \in v} \sigma_i^x$$



$$|\Psi_{\text{TC}}\rangle \propto \lim_{\beta \rightarrow \infty} \sum_{\sigma_1, \dots, \sigma_N} e^{\frac{\beta}{2} J \sum_p \prod_{i \in p} \sigma_i^z} |\sigma_1, \dots, \sigma_N\rangle$$

PEPS : F. Verstraete, M. M. Wolf, D. Perez-Garcia, J. I. Cirac Phys. Rev. Lett. 96, 220601 (2006).

$$O_{\text{cold}}(\sigma_1, \dots, \sigma_N) \propto \lim_{\beta \rightarrow \infty} \exp \beta J \sum_p \prod_{i \in p} \sigma_i^z$$



J. Carrasquilla and R. G. Melko. Nature Physics 13, 431–434 (2017)

Dong-Ling Deng et al Phys. Rev. X 7, 021021 (2017)

Jing Chen, Song Cheng, Haidong Xie, Lei Wang, Tao Xiang arXiv:1701.04831 RBMs

Natural Language Processing

Natural language processing

- Branch of computer science, linguistics, and machine learning concerned with giving computers the ability to process text and spoken words in a similar way humans do it.
- Machine translation
- Speech recognition
- Sentiment analysis
- Automatic summarization of text
- Text to image/video generation

ChatGPT

- ChatGPT is a machine learning model which interacts in a conversational way
- Dialogue format makes it possible for ChatGPT to answer followup questions, admit its mistakes, challenge incorrect premises, and reject inappropriate requests.

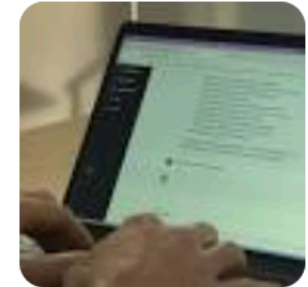


CBC

[Can the new AI tool ChatGPT replace human work? Judge for yourself](#)

A new artificial intelligence tool using natural language processing has captured the public's imagination, amassing more than a million...

1 day ago

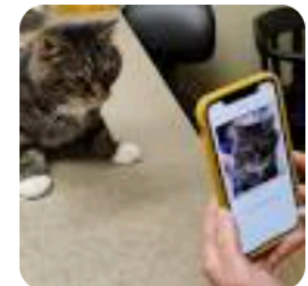


Global News

[ChatGPT: Everything to know about the viral, 'groundbreaking' AI bot - National | Globalnews.ca](#)

Users can ask the AI to write essays, poems or scripts, or even translate or summarize text. ChatGPT can also answer questions on a wide...

21 hours ago



CNET

[The 5 Best Uses \(So Far\) for ChatGPT's AI Chatbot](#)

The new AI tool ChatGPT has inspired excitement and worry with its ability to instantly answer complex questions. In the days after its...

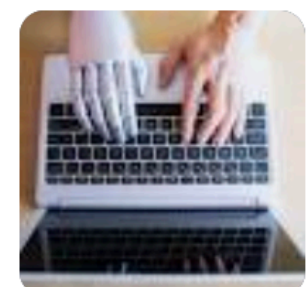
1 day ago



The Toronto Star

[I wrote a story about ChatGPT's AI. Then I dared it to write a better one](#)

Like a million or so other artificial intelligence dumb-dumbs who discovered ChatGPT this week, I couldn't wait to show off the capabilities...



Is Mona Lisa smiling? ChatGPT's take



is mona lisa smiling?



The smile of the Mona Lisa has been a subject of much discussion and interpretation over the years. Leonardo da Vinci's famous portrait depicts a slight smile, which has often been described as enigmatic or mysterious. It's not an overt, obvious smile, but rather a subtle expression that has captivated viewers and sparked numerous theories and debates.

The perception of the Mona Lisa's smile can vary from person to person, and different individuals may interpret it differently. Some see it as a gentle, content smile, while others perceive it as somewhat ambiguous, with a touch of sadness or even a hint of irony.

Ultimately, the interpretation of the smile is subjective and open to personal perception.



ChatGPT: is Mona Lisa smiling?



I'll do my best to create an ASCII art representation of the Mona Lisa, however, please keep in mind that it may not be a very detailed or realistic representation. Here is my attempt:



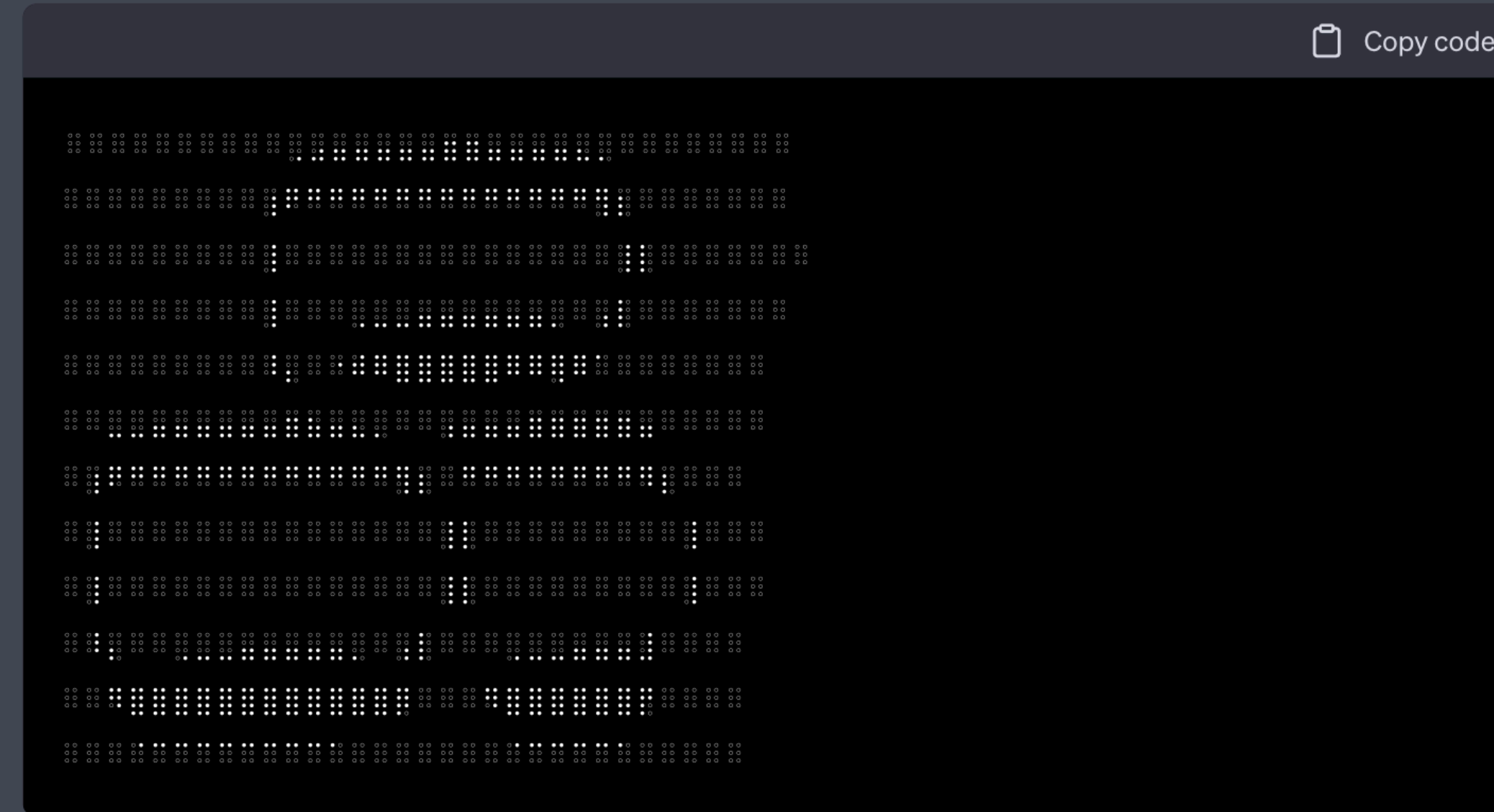
Please note that this is a very simple and minimal representation of Monalisa, as ASCII art is not my primary function. It may not be the best representation of Monalisa.



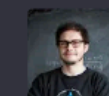
create an ascii representation of mona lisa



Certainly! Here's an ASCII representation of the Mona Lisa:



Please note that ASCII art may not capture the full complexity and detail of the original artwork, but I hope you enjoy this representation!



I am enjoying it as much as I did the first time I asked you to do it

How do these models work?

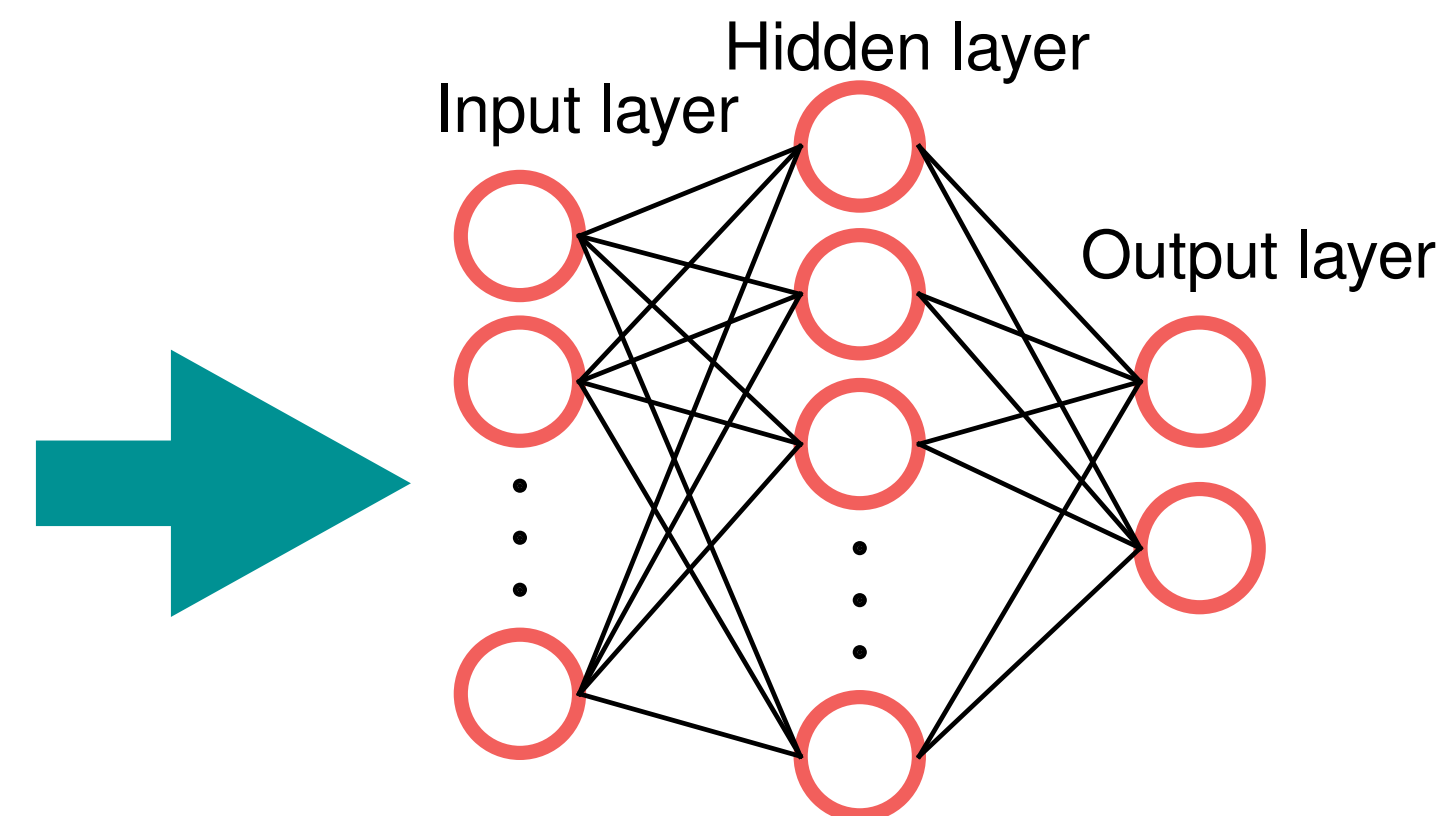
Large Language models

- Large language models are based on neural networks
- Neural networks are powerful universal function approximators and can in principle represent any function if given enough parameters
- In particular, in NLP these model objects like this:

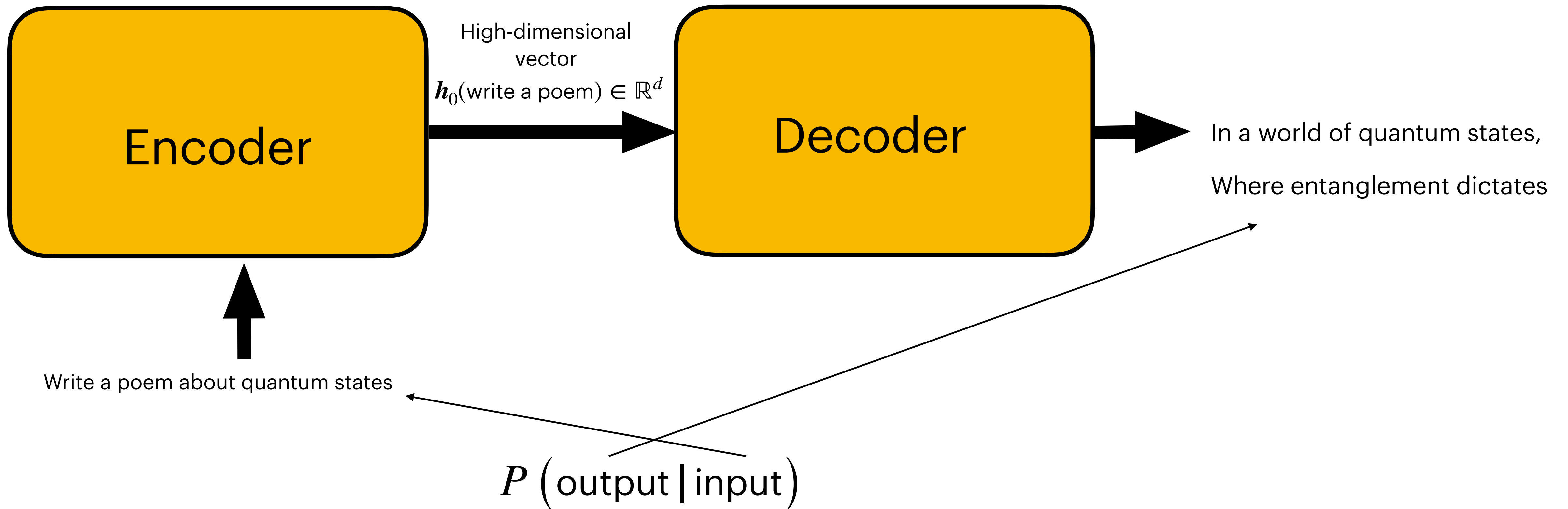
$P(\text{output text}|\text{input text})$

$P(\text{English}|\text{Spanish})$

$P(\text{Spanish})$

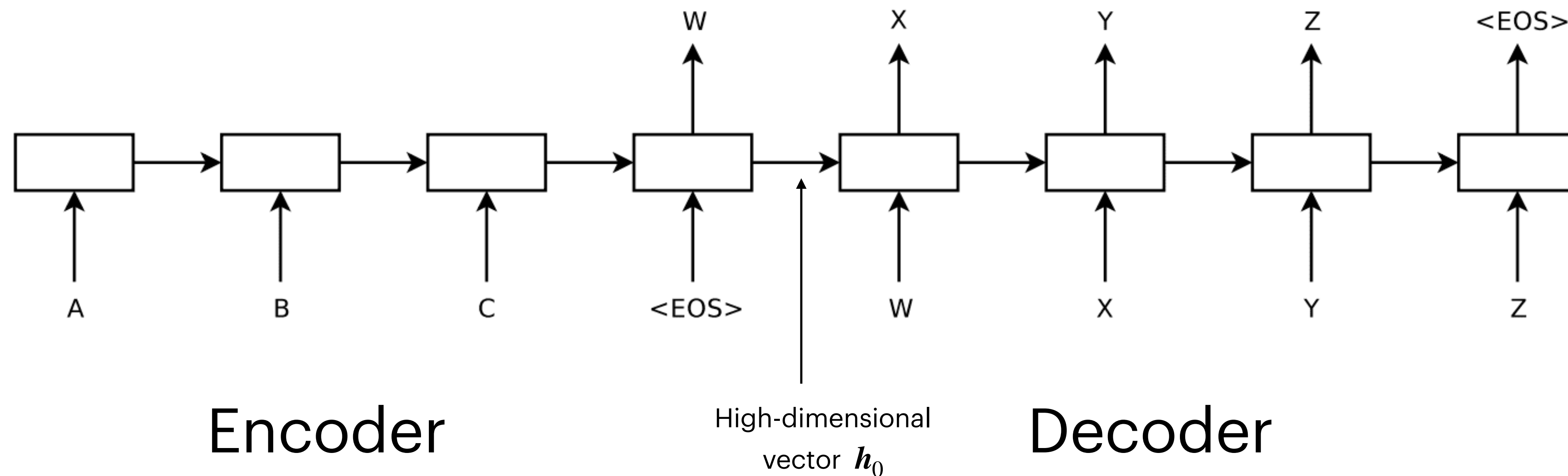


Large language models



Large Language models

Recurrent neural networks



Large Language models

Transformer

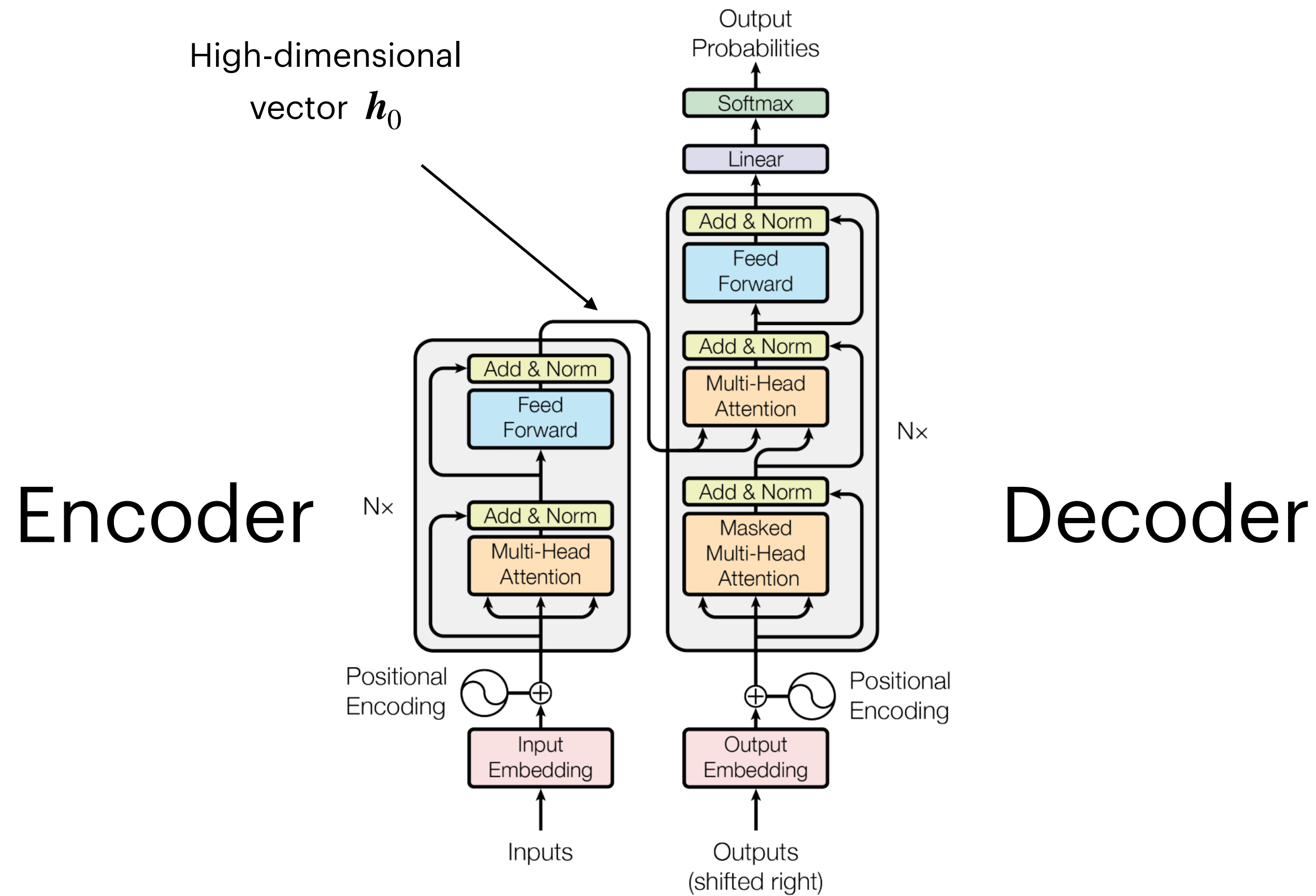
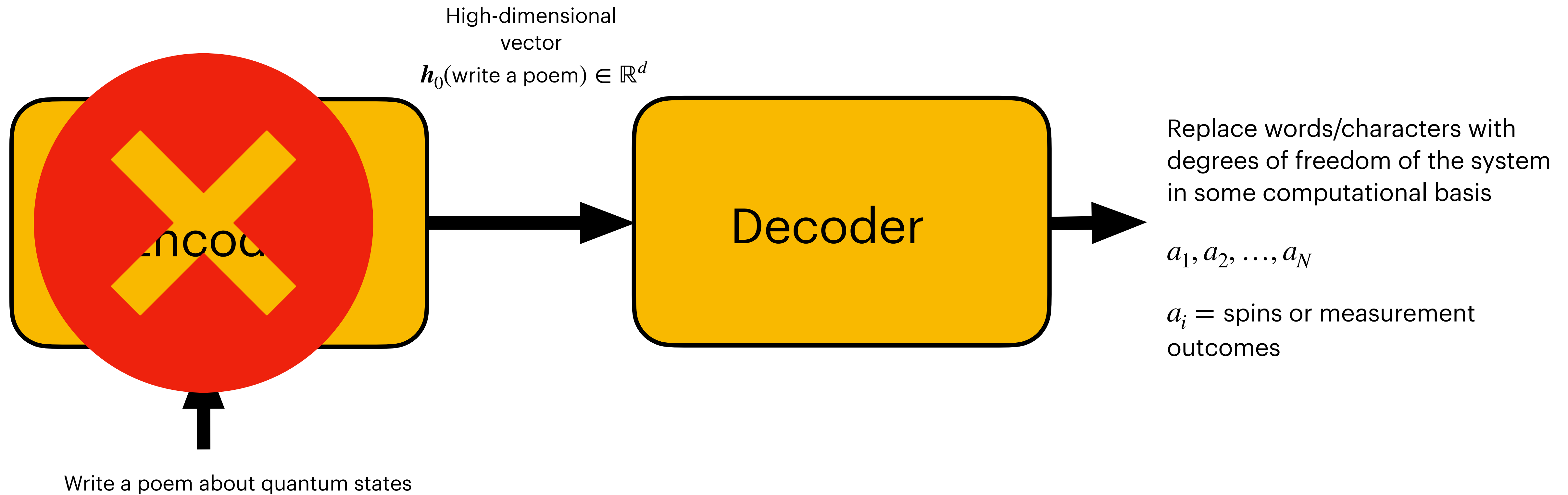


Figure 1: The Transformer - model architecture.

**How to use these models to
represent physical states?**


Representing states with language models



Input/output is given using some word dictionary = [word 1, word 2, ..., word N] \longrightarrow Input/output is given using some word dictionary = [spin up, spin down]

PROBABILISTIC AUTOREGRESSIVE MODELS

- The term *autoregressive* originates from time-series models: **observations from the previous time-steps are used to predict the value at the current time step.**
- Consider a probability distribution $P(\boldsymbol{\sigma}) = P(\sigma_1, \sigma_2, \dots, \sigma_N)$,

$$P(\sigma_1, \sigma_2, \dots, \sigma_N) = P(\sigma_1)P(\sigma_2|\sigma_1)P(\sigma_3|\sigma_1, \sigma_2) \dots P(\sigma_N|\sigma_1, \sigma_2, \dots, \sigma_{N-1})$$


- To specify P in a tabular form requires **exponential** resources
- To alleviate this exponential issue: parametrize the conditionals

$$P(\sigma_i|\sigma_{<i}) = P_\theta(\sigma_i|\sigma_{<i})$$

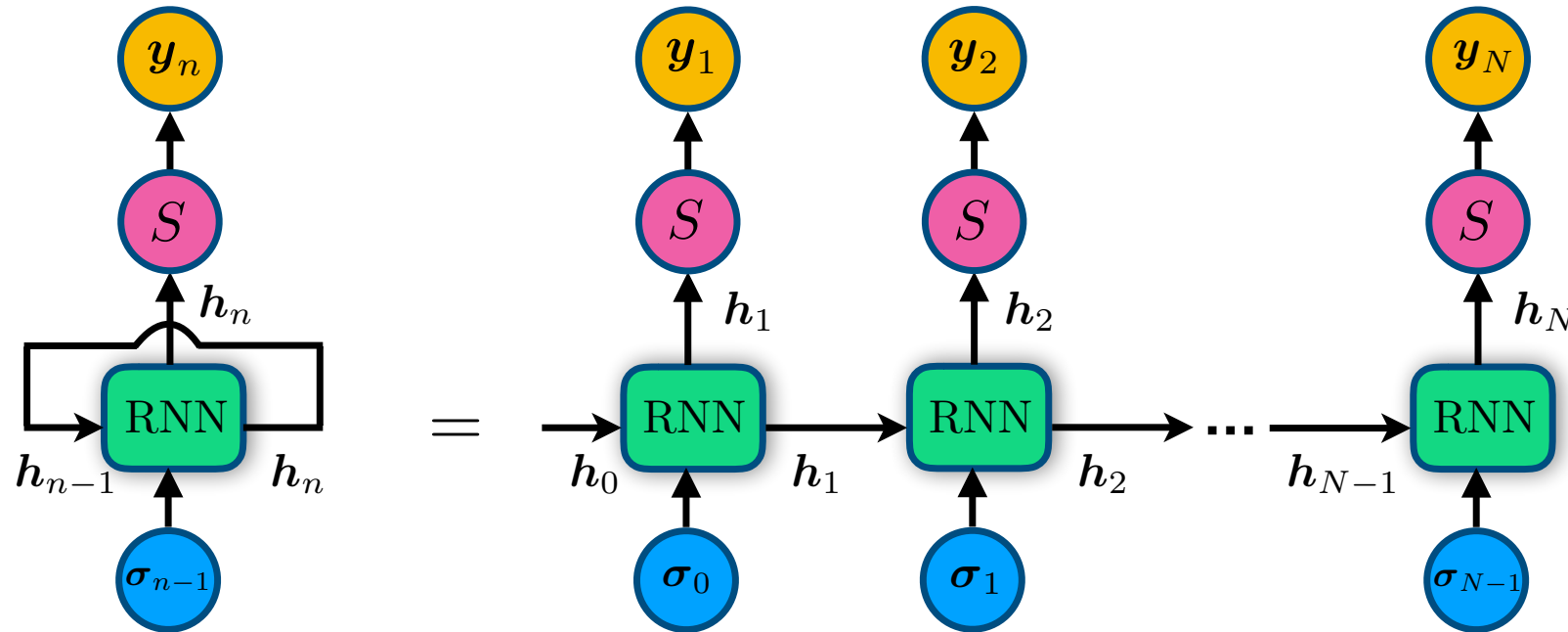
- This idea can be extended to quantum states.

Autoregressive models and their tractability advantages

- ✓ Computing $P(\boldsymbol{\sigma})$ is efficient given $\boldsymbol{\sigma}$
- ✓ Can be exactly sampled efficiently in linear time in N
- ✓ Can be easily defined in **any** spatial dimension (e.g. 2d and 3d) — no tractability issues (compared to, eg, PEPS)
- ✓ Their expressive power can be systematically improved
- ✓ Easy to encode product states and mean-field theories
- ✓ We can impose some symmetries and conservation laws
- ✓ Advantages remain true for models quantum states in any dimension.

RECURRENT NEURAL NETWORKS (RNN)

► The key building block of an RNN is a recurrent cell

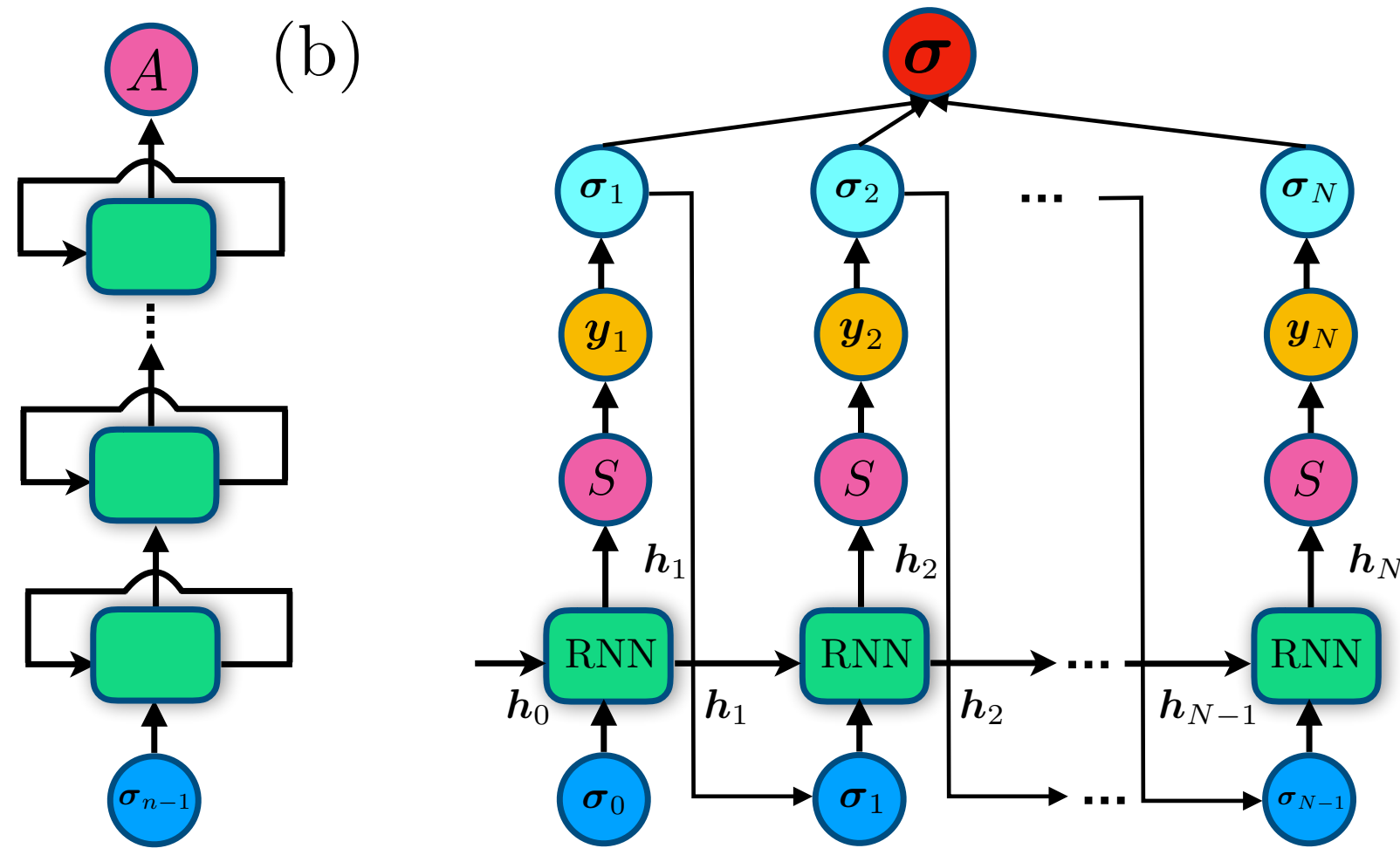


$$h_n = f(W[h_{n-1}; \sigma_{n-1}] + b)$$

$h_n \in \mathbb{R}^d$ — d measures the expressive power of the model

$$y_n \equiv S(Uh_n + c)$$

$$S(x_i) = \frac{e^{x_i}}{\sum_j e^{x_j}} \text{ — Softmax function}$$



$$y_n = P(\sigma_n | \sigma_1, \dots, \sigma_{n-1})$$

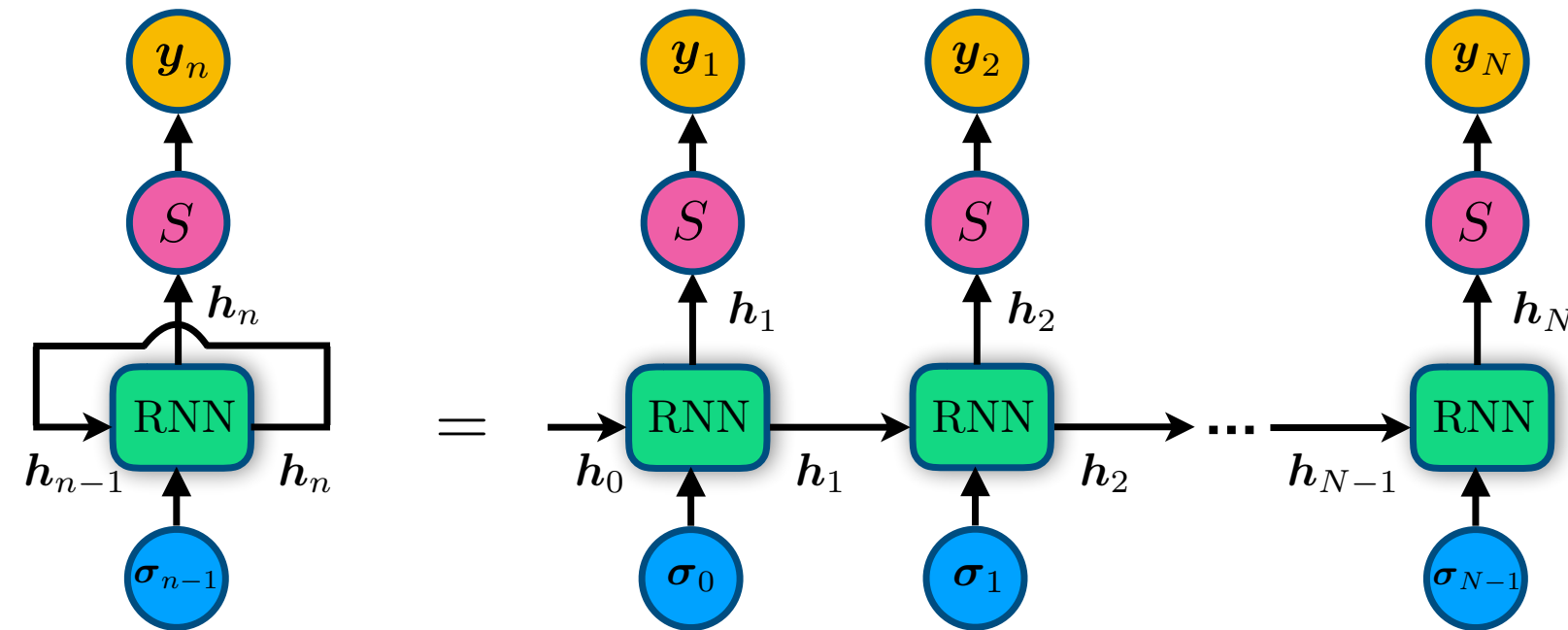
Sampling:

- Sample each conditional
- Input the sample to the next step

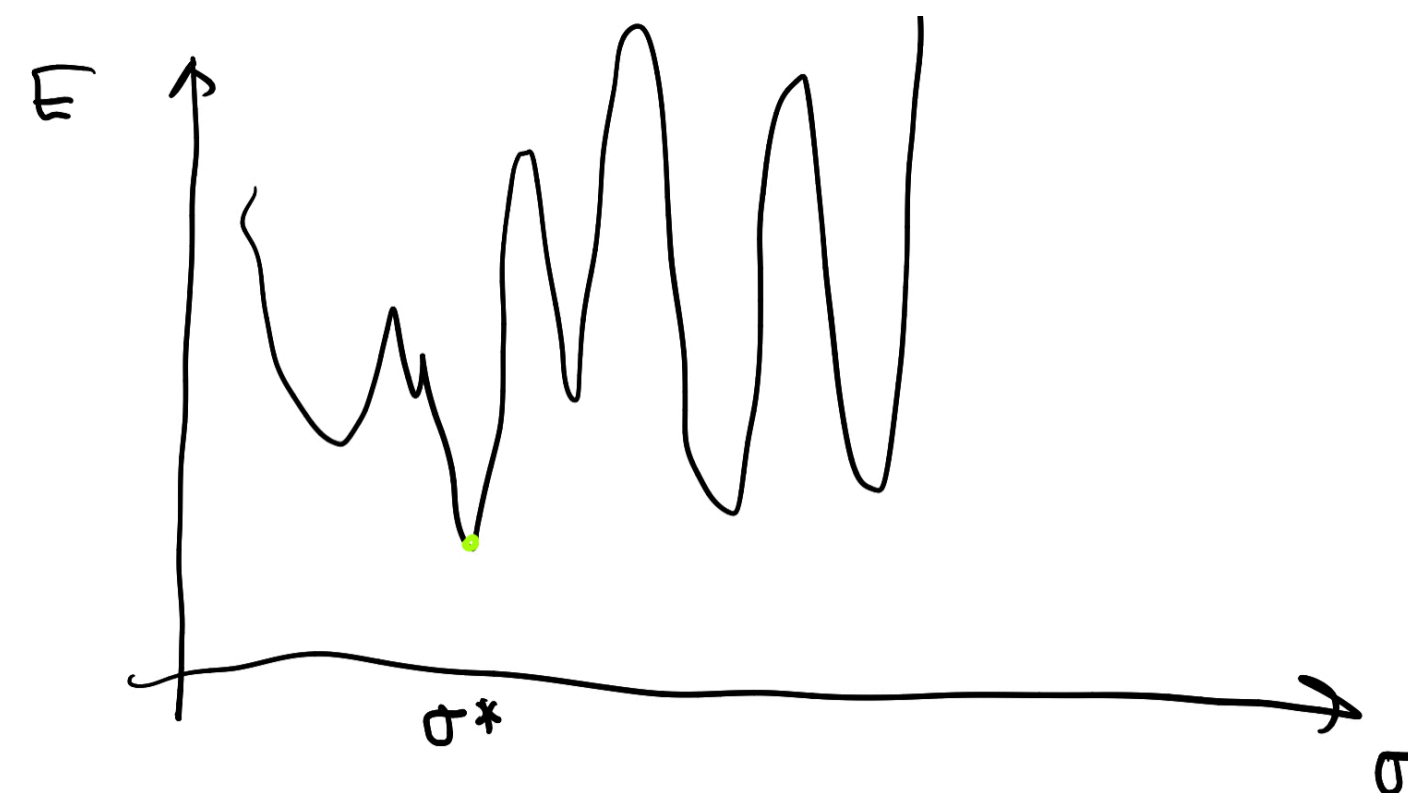
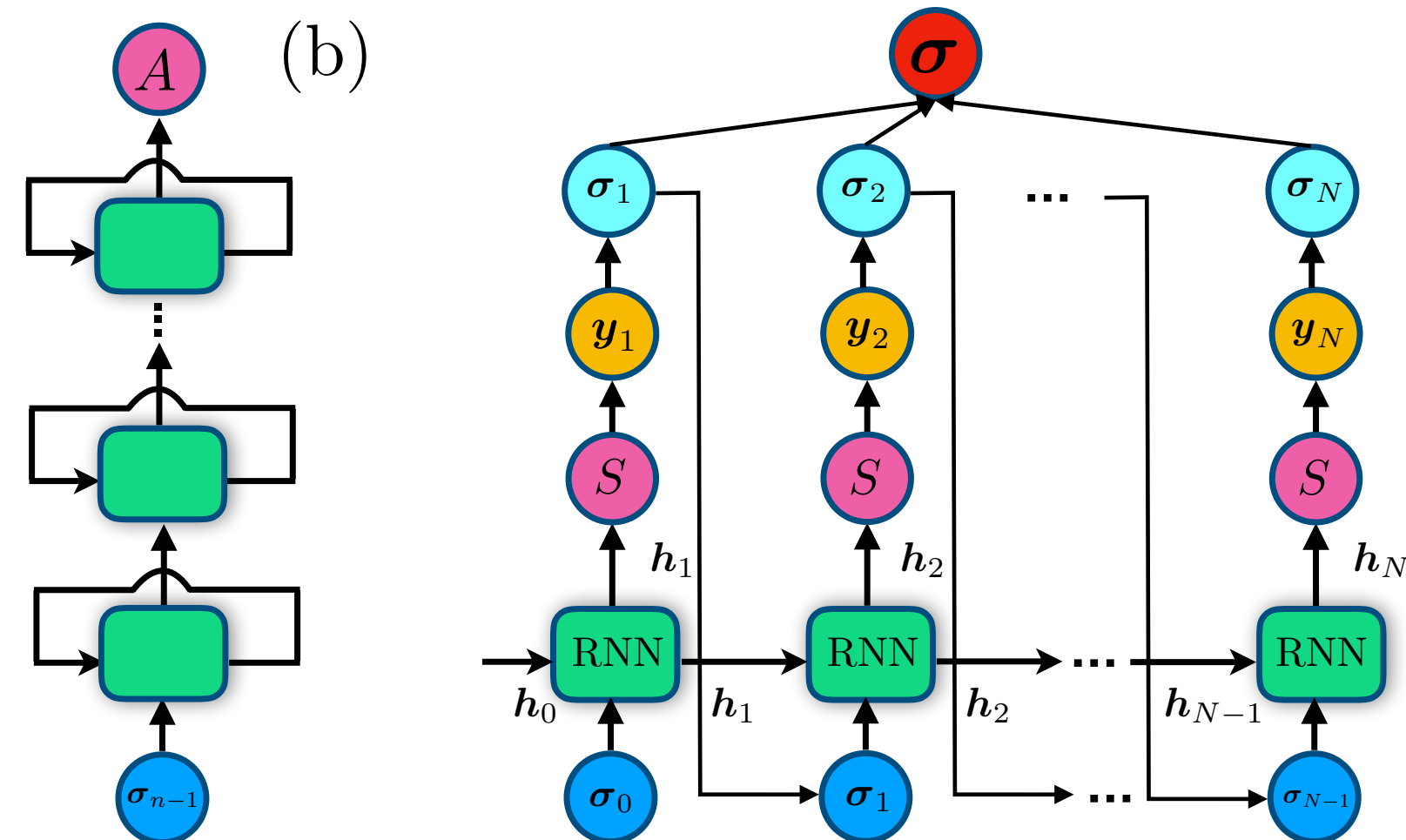
$$P(\sigma_1, \sigma_2, \dots, \sigma_N) = P(\sigma_1)P(\sigma_2|\sigma_1)P(\sigma_3|\sigma_1, \sigma_2) \dots P(\sigma_N|\sigma_1, \sigma_2, \dots, \sigma_{N-1})$$

RECURRENT NEURAL NETWORKS (RNN)

► The key building block of an RNN is a recurrent cell



- Key insight about this model choice:
- It can be sampled exactly even if it encodes a rough landscape.



$$P(\sigma_1, \sigma_2, \dots, \sigma_N) = P(\sigma_1)P(\sigma_2|\sigma_1)P(\sigma_3|\sigma_1, \sigma_2) \dots P(\sigma_N|\sigma_1, \sigma_2, \dots, \sigma_{N-1})$$

Example 1: Data-driven reconstruction of quantum states

Quantum state tomography

Quantum state tomography is the process of reconstructing the quantum state by **measurements** on the system. It “is the gold standard for verification and benchmarking of quantum devices”*

Useful for:

- Characterizing optical signals
- Diagnosing and detecting errors in state preparation, e.g. states produced by quantum computers reliably.
- Entanglement verification

* Efficient quantum state tomography. Marcus Cramer, Martin B. Plenio, Steven T. Flammia, Rolando Somma, David Gross, Stephen D. Bartlett, Olivier Landon-Cardinal, David Poulin & Yi-Kai Liu. Nature Communications volume 1, Article number: 149

Need to go beyond standard quantum state tomography


- Progress in controlling large quantum systems.
- Availability of arbitrary measurements performed with good accuracy.
- The bottleneck limiting progress in the estimation of states: **curse of dimensionality of traditional techniques.**
- **Traditional quantum state tomography scales exponentially—Can we perform reconstruction of large systems?**



Synthetic Quantum devices are growing fast

Article | [Published: 07 July 2021](#)

Quantum phases of matter on a 256-atom programmable quantum simulator

[Sepehr Ebadi](#), [Tout T. Wang](#), [Harry Levine](#), [Alexander Keesling](#), [Giulia Semeghini](#), [Ahmed Omran](#), [Dolev Bluvstein](#), [Rhine Samajdar](#), [Hannes Pichler](#), [Wen Wei Ho](#), [Soonwon Choi](#), [Subir Sachdev](#), [Markus Greiner](#), [Vladan Vuletić](#) & [Mikhail D. Lukin](#) 

[Nature](#) **595**, 227–232 (2021) | [Cite this article](#)


[nature.com](#) > [nature](#) > [letters](#) > [article](#)



Letter | [Published: 22 August 2018](#)

Observation of topological phenomena in a programmable lattice of 1,800 qubits

[Andrew D. King](#) , [Juan Carrasquilla](#), [...] [Mohammad H. Amin](#)

[Nature](#) **560**, 456–460 (2018) | [Download Citation](#) 

Science

[Current Issue](#) [First release papers](#) [Archive](#) [About](#)  [Submit manuscript](#)

[HOME](#) > [SCIENCE](#) > [VOL. 374, NO. 6572](#) > [PROBING TOPOLOGICAL SPIN LIQUIDS ON A PROGRAMMABLE QUANTUM SIMULATOR](#)

RESEARCH ARTICLE | TOPOLOGICAL MATTER



Probing topological spin liquids on a programmable quantum simulator

[G. SEMEGHINI](#) , [H. LEVINE](#) , [A. KEESLING](#) , [S. EBADI](#) , [T. T. WANG](#) , [D. BLUVSTEIN](#) , [R. VERRESEN](#) , [H. PICHLER](#) , [M. KALINOWSKI](#), [...] [M. D. LUKIN](#)

 [+7 authors](#) [Authors Info & Affiliations](#)

PHYSICAL REVIEW X

[Highlights](#) [Recent](#) [Subjects](#) [Accepted](#) [Collections](#) [Authors](#) [Referees](#) [Search](#) [Press](#) [About](#) [Staff](#)

Open Access

Quantum Chemistry Calculations on a Trapped-Ion Quantum Simulator

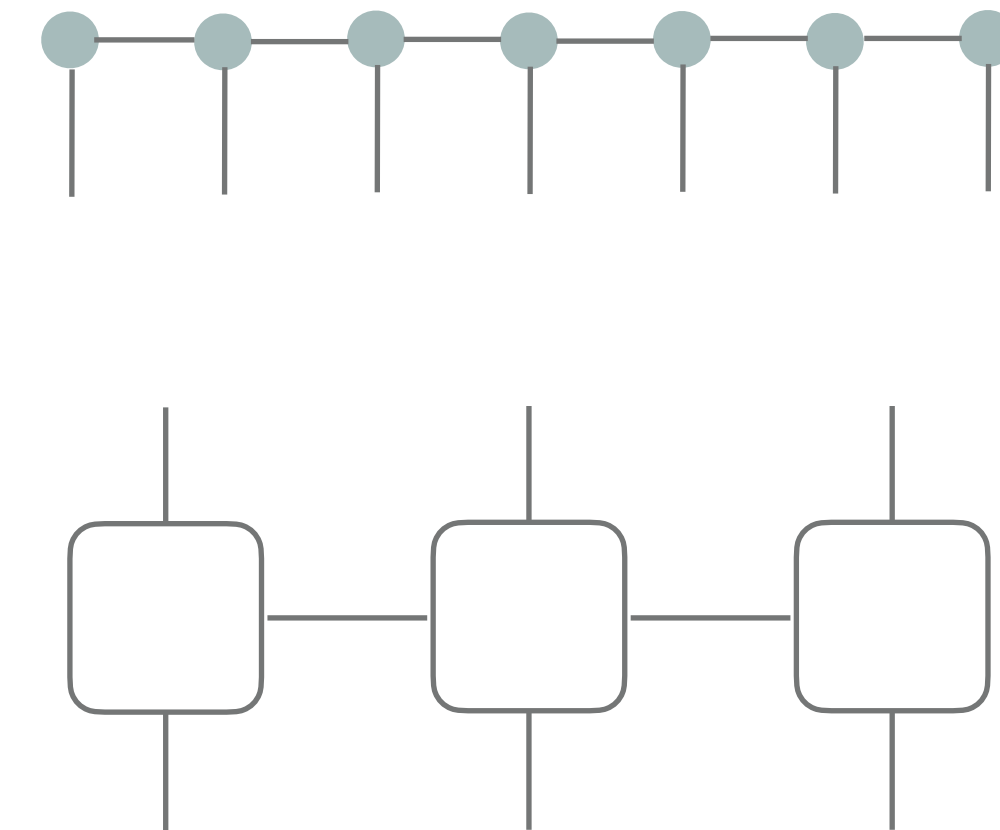
[Cornelius Hempel](#), [Christine Maier](#), [Jonathan Romero](#), [Jarrod McClean](#), [Thomas Monz](#), [Heng Shen](#), [Petar Jurcevic](#), [Ben P. Lanyon](#), [Peter Love](#), [Ryan Babbush](#), [Alán Aspuru-Guzik](#), [Rainer Blatt](#), and [Christian F. Roos](#)
Phys. Rev. X **8**, 031022 – Published 24 July 2018

HOW TO MAKE QST EFFICIENT?

- Introduce a parametrization of the quantum state with good scaling if non-trivial structural information on the quantum systems under consideration is utilized: [MPS\[1\]](#) and [MPO\[2\]](#) tomography

[1] **Efficient quantum state tomography.** Marcus Cramer, Martin B. Plenio, Steven T. Flammia, Rolando Somma, David Gross, Stephen D. Bartlett, Olivier Landon-Cardinal, David Poulin & Yi-Kai Liu. Nature Communications volume **1**, Article number: 149

[2] **A scalable maximum likelihood method for quantum state tomography** T Baumgratz¹, A Nüßeler, M Cramer and M B Plenio. New Journal of Physics, Volume 15, December 2013



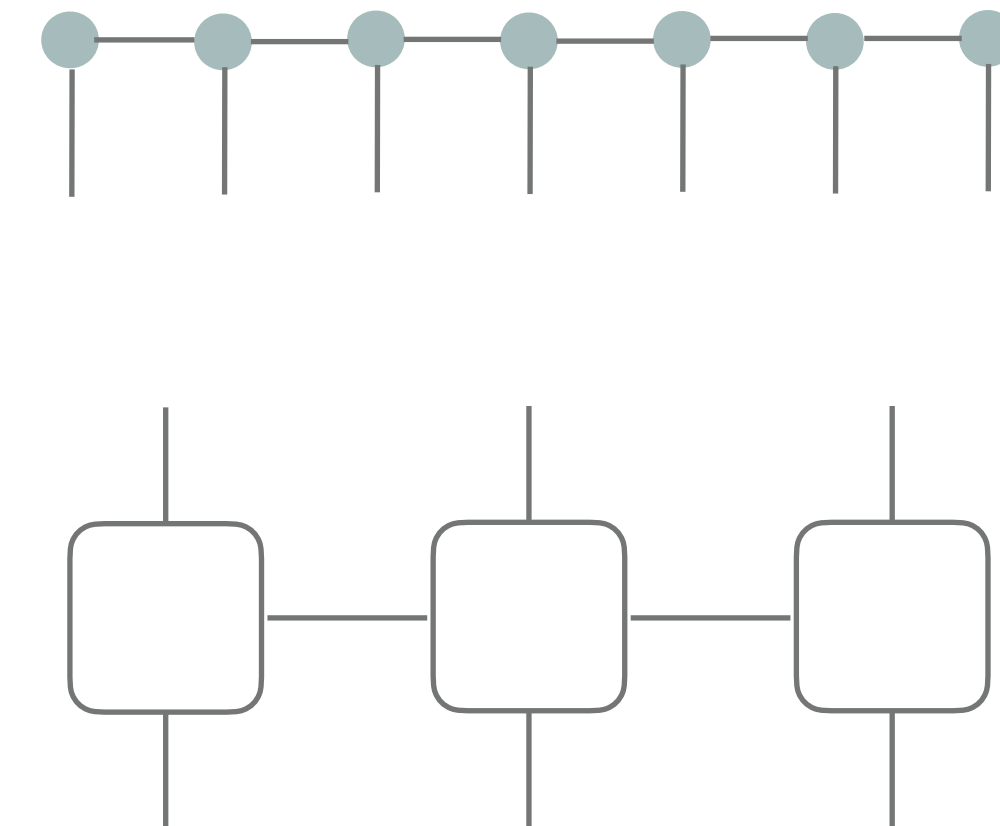
HOW TO MAKE QST EFFICIENT?

- Introduce a parametrization of the quantum state with good scaling if non-trivial structural information on the quantum systems under consideration is utilized: [MPS\[1\]](#) and [MPO\[2\]](#) tomography

[1] **Efficient quantum state tomography.** Marcus Cramer, Martin B. Plenio, Steven T. Flammia, Rolando Somma, David Gross, Stephen D. Bartlett, Otfried Zeitner, Christof Doherty, David Reich, & Yu. Kitaev. Nat. Commun. 4, 1794 (2013)

Works very well, best suited to 1D systems

[2] **A scalable maximum likelihood method for quantum state tomography** T Baumgratz¹, A Nüßeler, M Cramer and M B Plenio. New Journal of Physics, Volume 15, December 2013

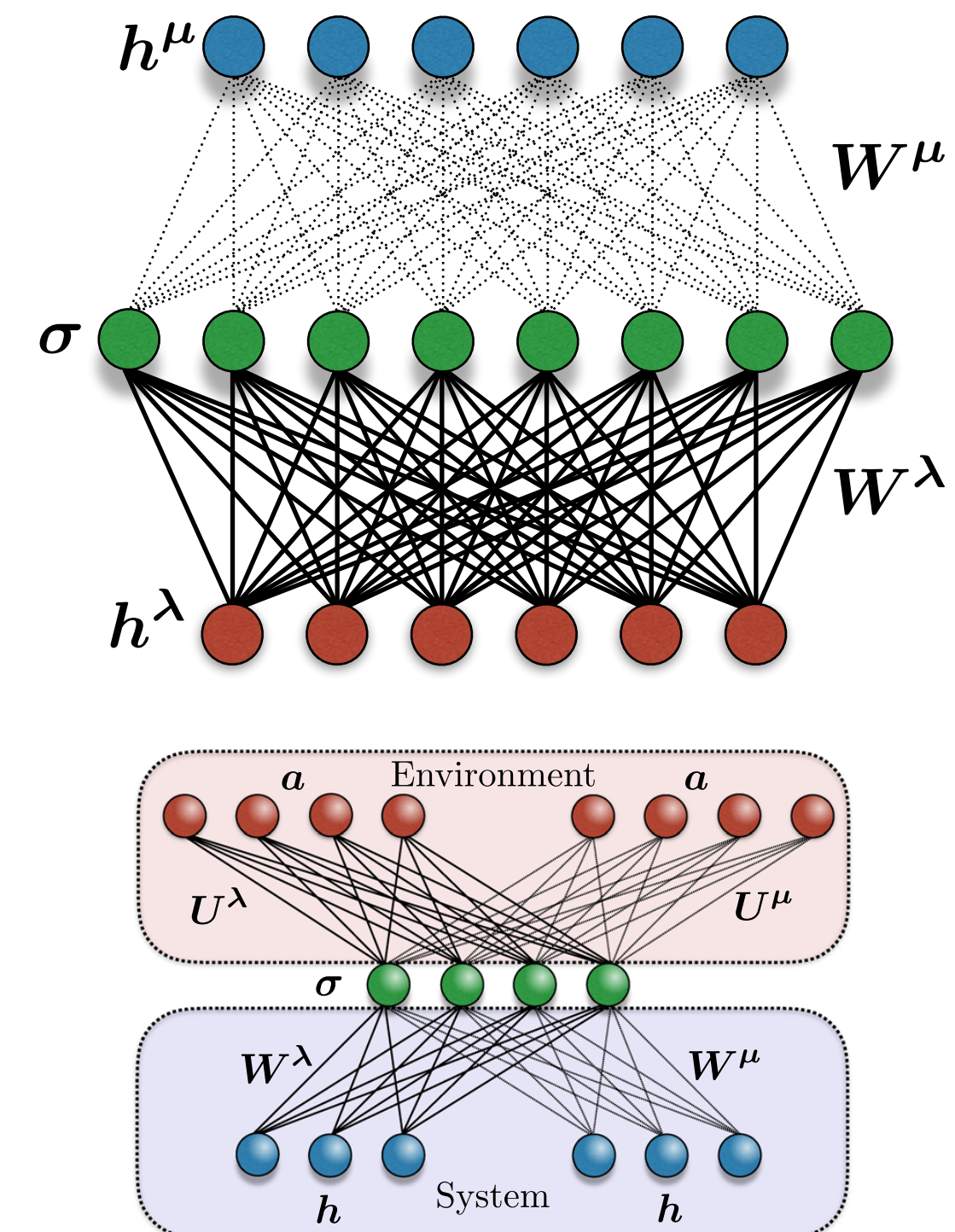


HOW TO MAKE QST EFFICIENT?

- Introduce a parametrization of the quantum state with good scaling if non-trivial structural information on the quantum systems under consideration is utilized:
Restricted Boltzmann machines both for pure[3] states and mixed states[4]

[3] **Neural-network quantum state tomography.** G. Torlai, G. Mazzola, J. Carrasquilla, M. Troyer, R. Melko, and G. Carleo, Nat. Phys. 14, 447 (2018).

[4] **Latent Space Purification via Neural Density Operators.** Giacomo Torlai and Roger G. Melko. Phys. Rev. Lett. 120, 240503 (2018)

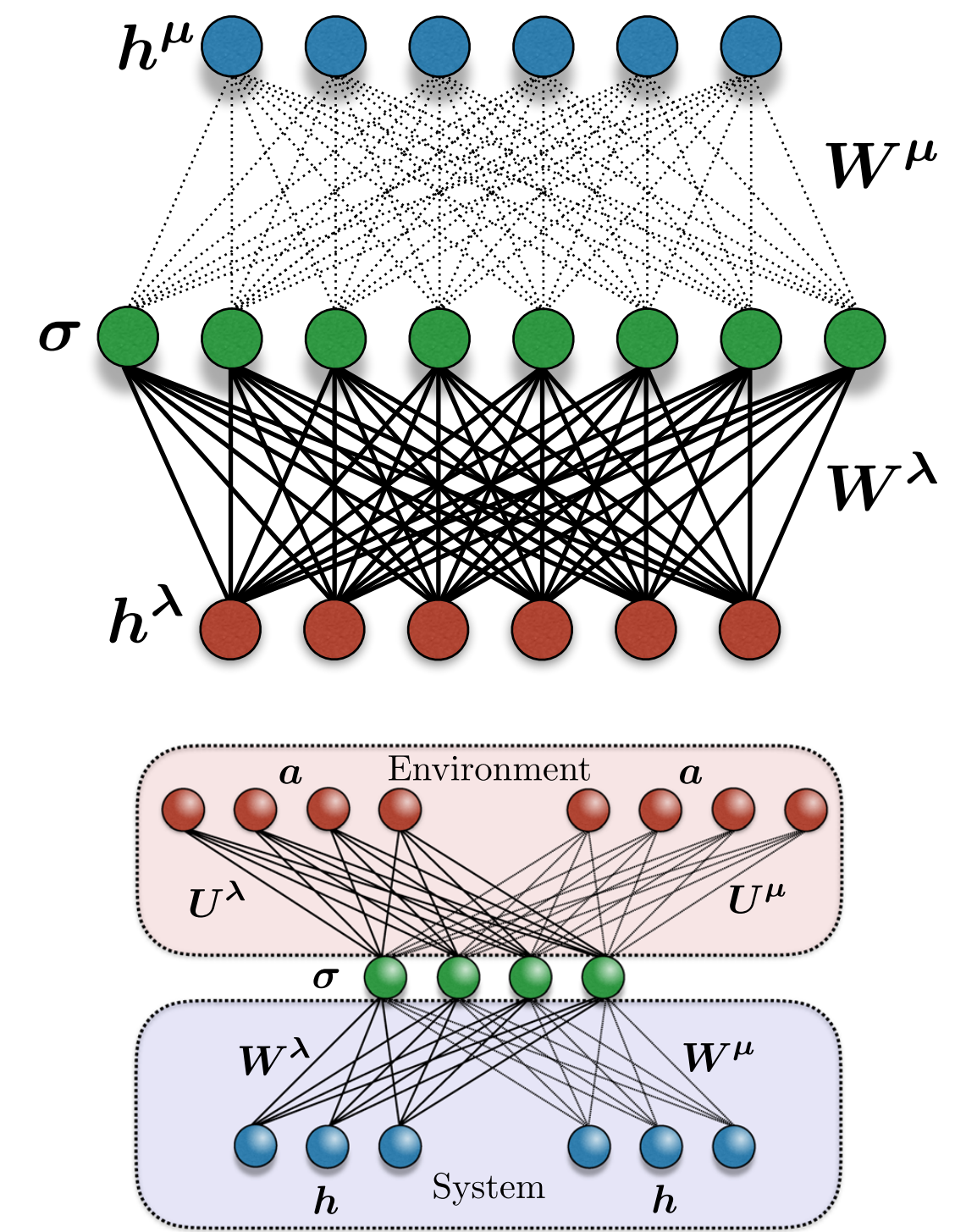


HOW TO MAKE QST EFFICIENT?

- Introduce a parametrization of the quantum state with good scaling if non-trivial structural information on the quantum systems under consideration is utilized:
Restricted Boltzmann machines both for pure[3] states and mixed states[4]

[3] Pure states with structure, high fidelity/good scaling
 Phys. 14, 447 (2018).

[4] Latent Space Purification via Neural Density Operators
 Mixed states with structure, high fidelity/bad scaling

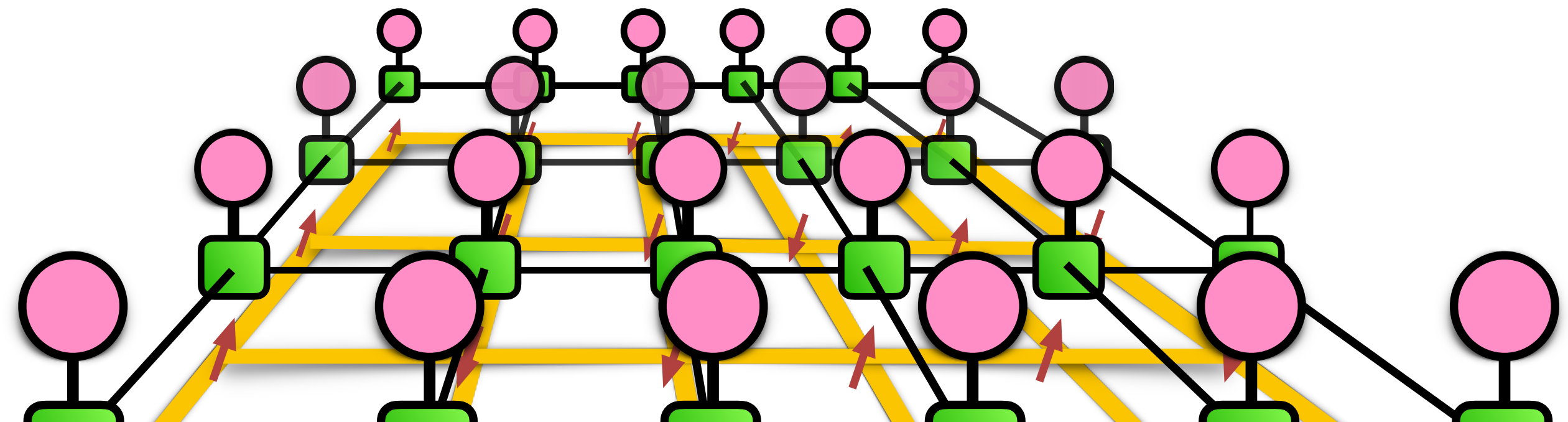


PROPOSAL

- ~~Parametrize the quantum state~~ We parametrize the measurement statistics of an informationally complete positive operator valued measure (POVM). We use recurrent neural networks and transformers.
- Use this idea to learn states from synthetic measurements mimicking experimental data as well as data from a small experiment.

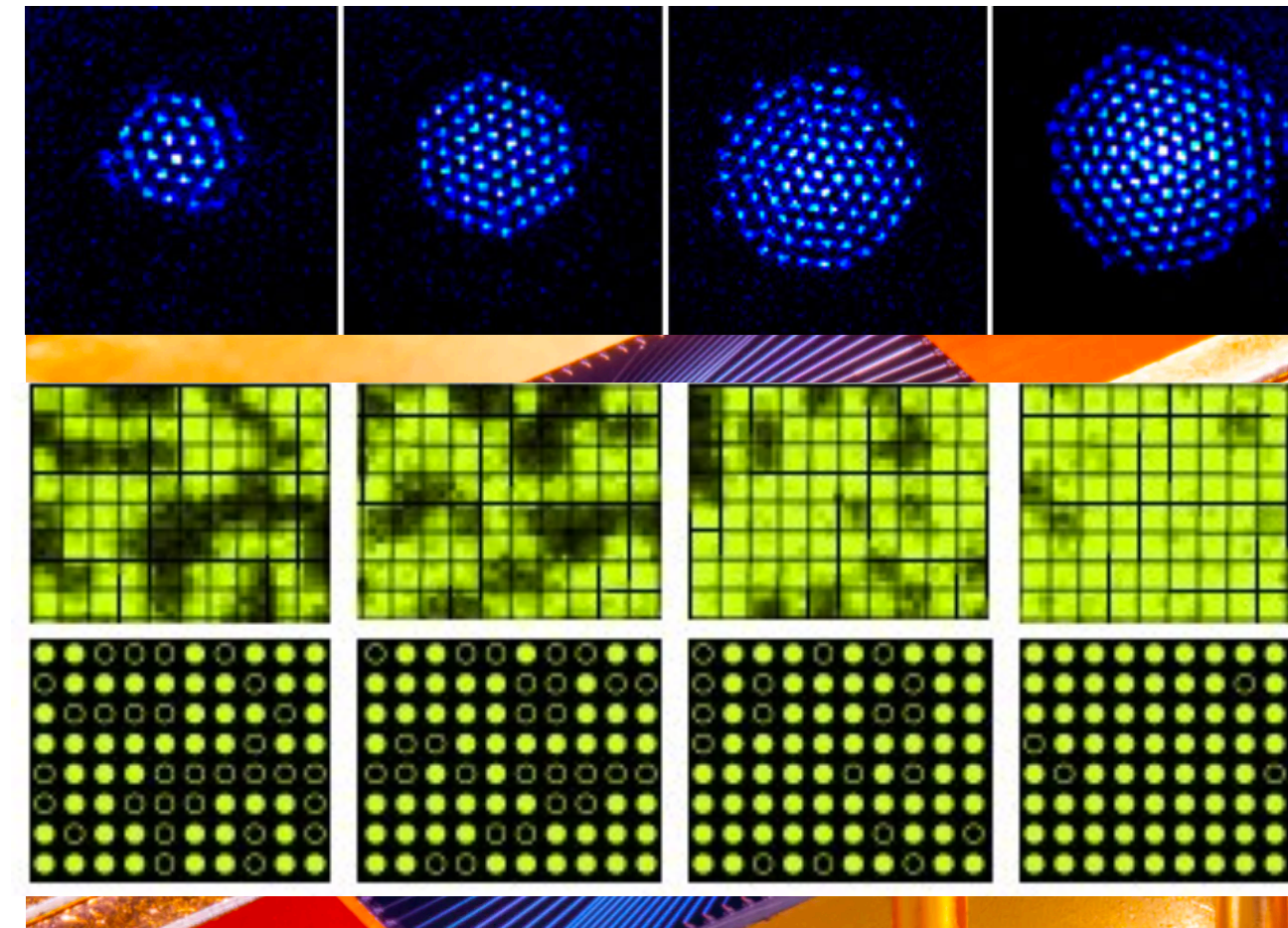
Mixed states with structure, good scaling

LEARNING QUANTUM STATES WITH RNNs



Reconstructing a quantum state

- Prepare an unknown quantum state
- Apply a measurement that probes enough information about the quantum state
- Repeat and collect the statistics of the measurement
- Infer a reconstruction of the state consistent with the measurement outcomes
- Certify the reconstruction



dataset =
$$\begin{bmatrix} 0,0,2,\dots,3,0 \\ 1,1,0,\dots,0,2 \\ 0,2,1,\dots,0,3 \\ 1,1,3,\dots,3,0 \\ \vdots \\ 3,2,0,\dots,1,0 \end{bmatrix} \longrightarrow \rho^* \longrightarrow F(\rho, \rho^*) = \text{Tr} \left[\sqrt{\sqrt{\rho} \rho^* \sqrt{\rho}} \right]^2$$

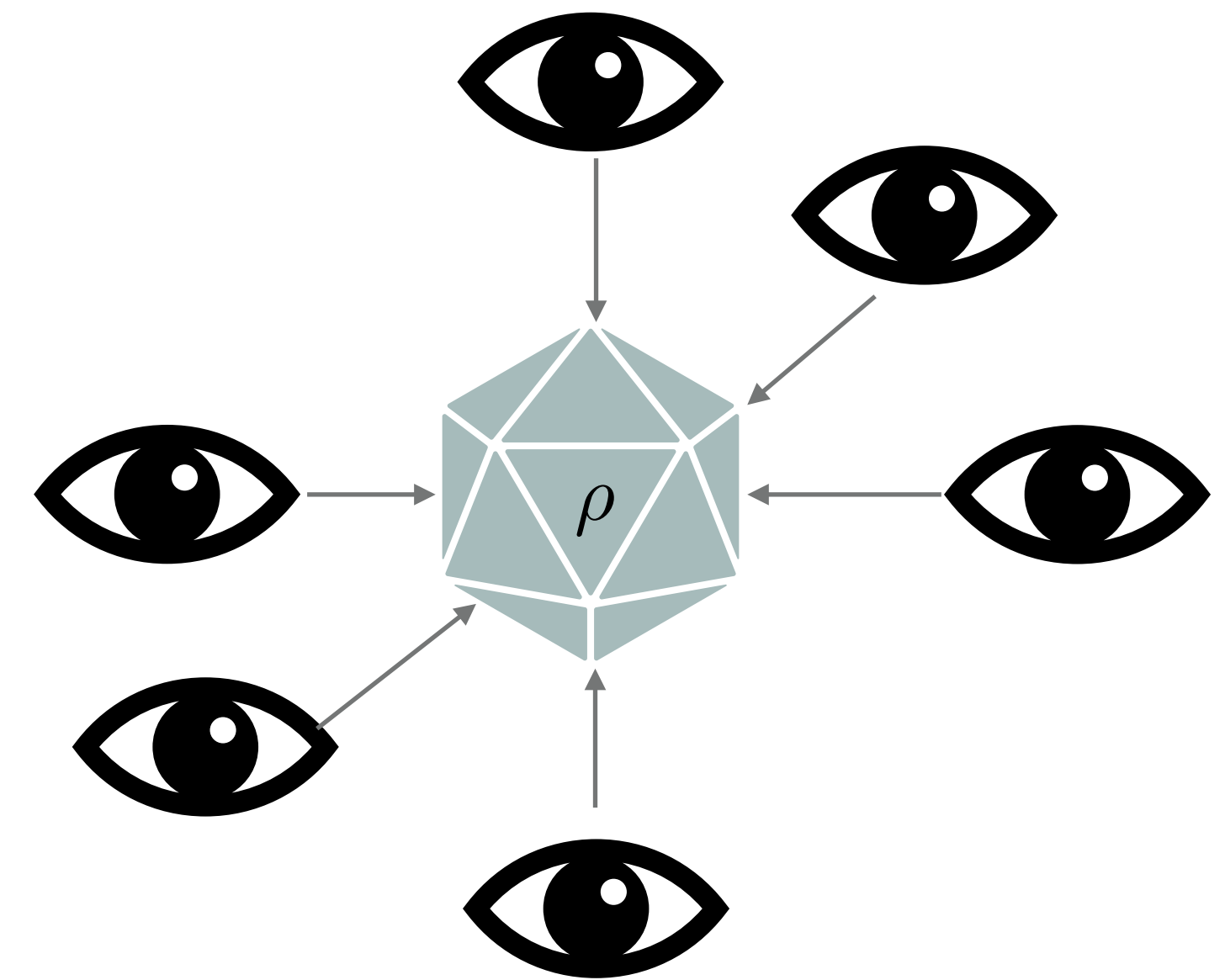
MEASUREMENTS: POSITIVE OPERATOR-VALUED MEASURE (POVM)

- ▶ POVM elements $\mathbf{M} = \{M^{(a)} \mid a \in \{1, \dots, m\}\}$
- ▶ Positive semidefinite operators $\sum_i M^{(a)} = \mathbb{1}$
- ▶ \mathbf{M} describe the experimental measurements
- ▶ Born Rule $P(a) = \text{Tr} [\rho M^{(a)}]$ quantum theory \leftrightarrow experiment

INFORMATIONALLY COMPLETE MEASUREMENTS

- The measurement statistics $P(\mathbf{a})$ contains all of the information about the state.
- Relation between ρ and distribution $P(\mathbf{a})$ can be inverted

$$\mathbf{M} = \{M^{(a_1)} \otimes M^{(a_2)} \otimes \dots \otimes M^{(a_N)}\}_{a_1, \dots, a_N}$$



INVERTING BORN RULE

BORN RULE

$$P(\mathbf{a}) = \text{Tr} \rho M^{\mathbf{a}}$$

INFORMATIONALLY COMPLETENESS \longrightarrow THIS RELATION CAN BE INVERTED

$$\rho = \sum_{a, a'} T_{a, a'}^{-1} P(a') M^{(a)}$$

$$T_{a, a'} = \text{Tr}[M^{(a)} M^{(a')}]$$

- Factorization of the state in terms of a probability distribution and a set of single-qubit tensors.
- All the entanglement and potential complexity of the state comes from the structure of the $P(\mathbf{a})$
- Tomography
- Real-time dynamics of close and open systems
- Measurements
- Etc

Insight: parametrize statistics of measurements and invert

$$P(\mathbf{a}) = \text{Tr} \rho M^{\mathbf{a}}$$

$$P_{\text{model}}(\mathbf{a}) \longrightarrow$$

$$\rho_{\text{model}} = \sum_{a,a'} T_{a,a'}^{-1} P_{\text{model}}(a') M^{(a)}$$

Create an autoregressive model of $P(\mathbf{a})$

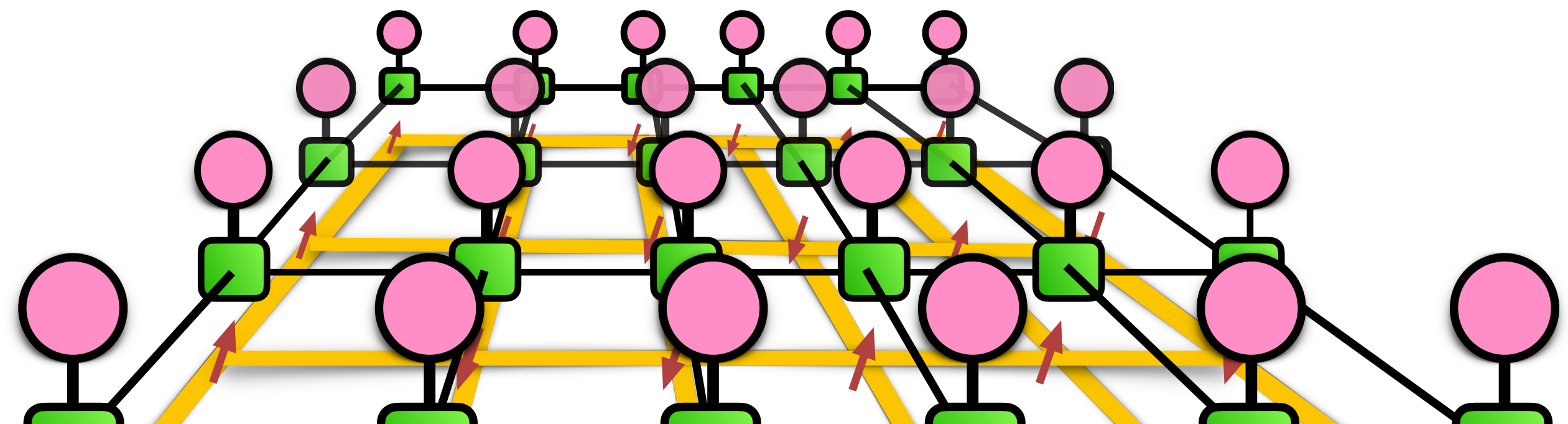
Autoregressive models (e.g., RNNs or transformer)

1. Allow for exact sampling
2. Tractable density $P_{\text{model}}(\mathbf{a})$
3. Use maximum likelihood estimation to infer the parameters of the model so that the distribution explains the dataset of measurements.

$$\text{dataset} = \begin{bmatrix} 0,0,2,\dots,3,0 \\ 1,1,0,\dots,0,2 \\ 0,2,1,\dots,0,3 \\ 1,1,3,\dots,3,0 \\ \vdots \\ 3,2,0,\dots,1,0 \end{bmatrix}$$

RESULTS

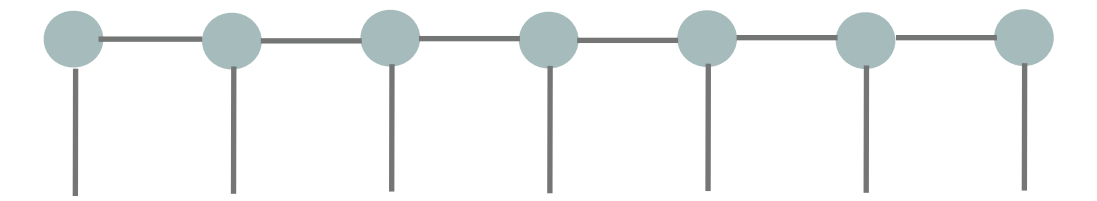
Carrasquilla, Torlai, Melko, Aolita. Nature Machine Intelligence 1, 200 (2019)



RESULTS ON SYNTHETIC DATASETS FOR GHZ STATES

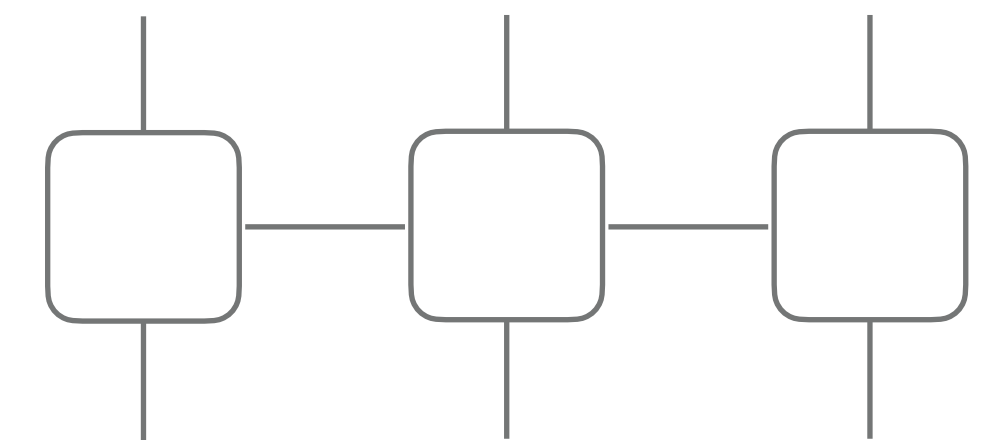
GHZ STATE $|\Psi_0\rangle \equiv \alpha |0\rangle^{\otimes N} + \beta |1\rangle^{\otimes N}$

$$\begin{aligned} \varrho_0 &:= |\Psi_0\rangle \langle \Psi_0| \\ &= |\alpha|^2 |0\rangle \langle 0|^{\otimes N} + |\beta|^2 |1\rangle \langle 1|^{\otimes N} + \left(\alpha\beta^* |0\rangle \langle 1|^{\otimes N} + \text{h.c.} \right) \end{aligned}$$



GHZ with local depolarization: a model of a decohering qubit where with probability $1 - p$ each qubit remains intact, while with probability p an “error” occurs.

$$\mathcal{E}_i \varrho_0 = (1 - p) \varrho_0 + \frac{p}{3} \left(\sigma_i^{(1)} \varrho_0 \sigma_i^{(1)} + \sigma_i^{(2)} \varrho_0 \sigma_i^{(2)} + \sigma_i^{(3)} \varrho_0 \sigma_i^{(3)} \right)$$

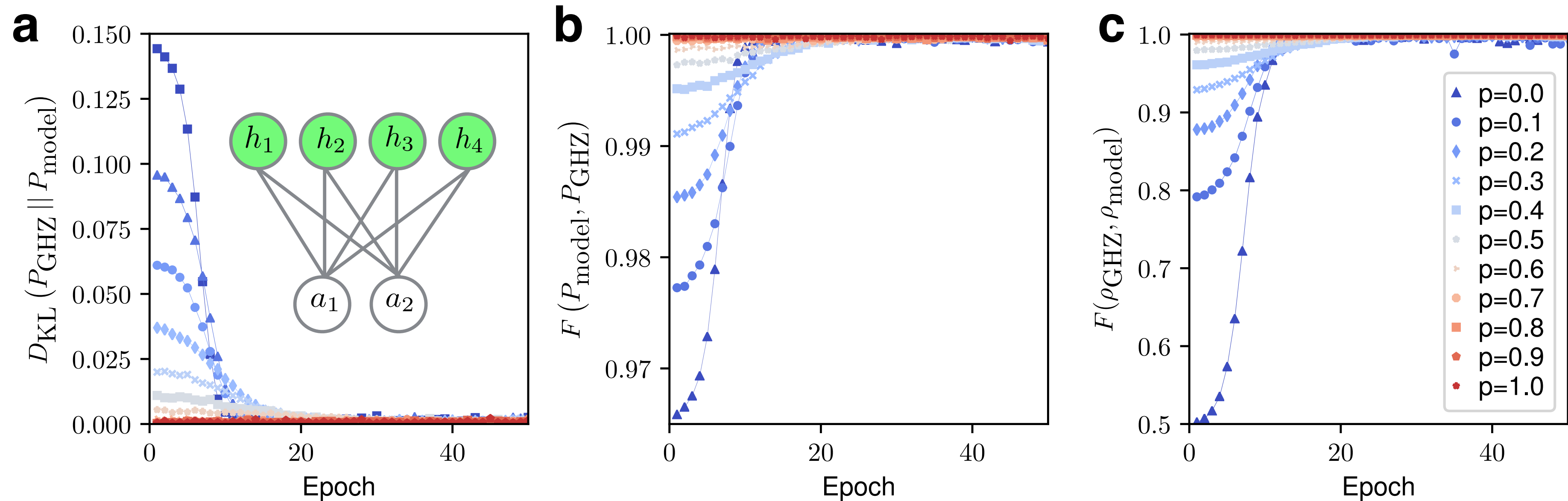


Learning 2 qubit locally depolarized GHZ

$$KL(P_{\text{model}}|P) = - \sum_a P(a) \log \frac{P_{\text{model}}(a)}{P(a)}$$

$$F_{\text{Classical}} = \sum_a \sqrt{P(a)P_{\text{model}}(a)}$$

$$F(\rho, \sigma) = \text{Tr} \left[\sqrt{\sqrt{\rho}\sigma\sqrt{\rho}} \right]^2$$



This result is obtained by parametrizing the $P(a)$ with an RBM.

We found it very difficult to scale the learning to larger system sizes with an RBM.

Larger systems: recurrent neural network model and results

GHZ results

40 qubit $p=0$

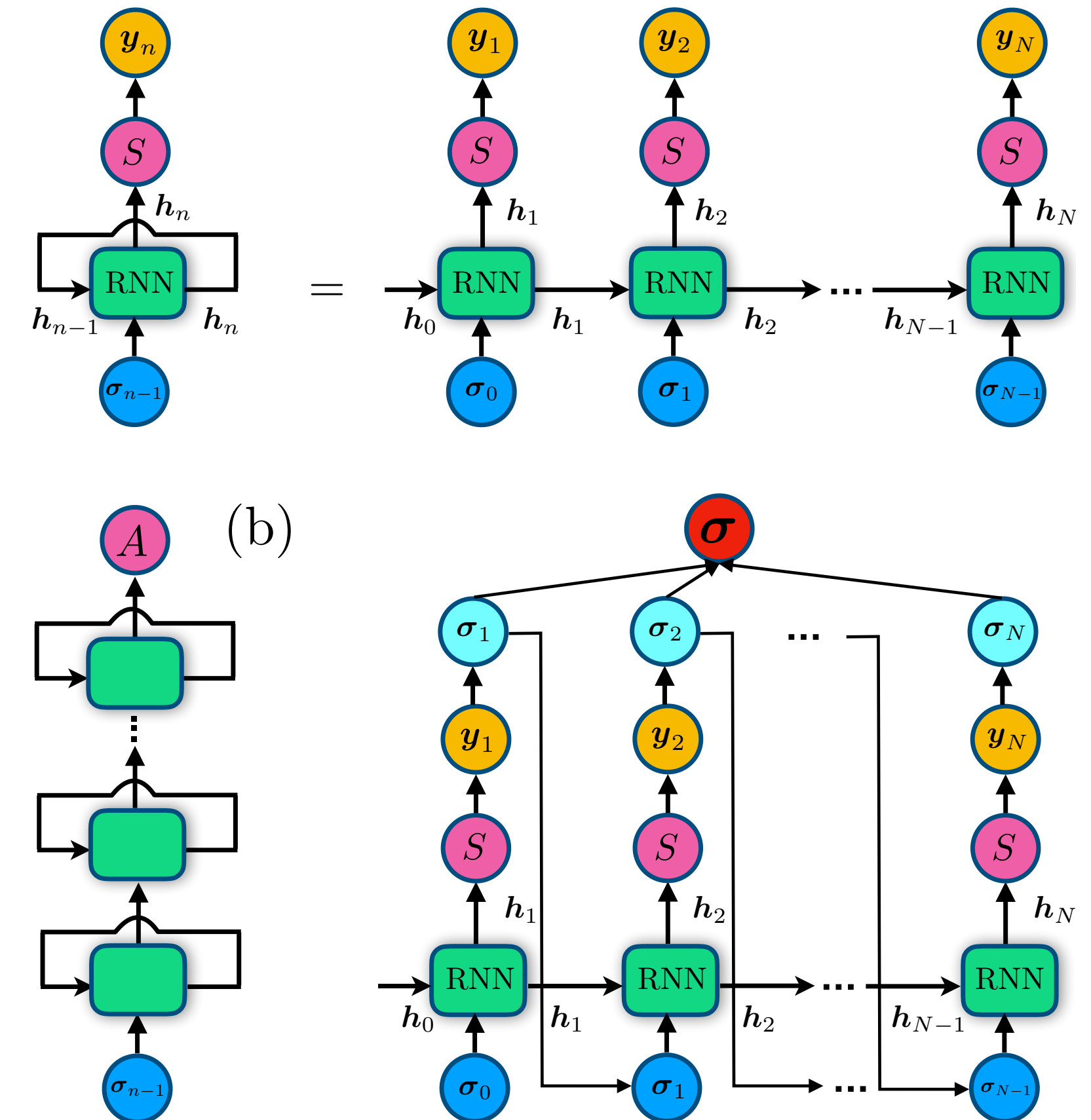
Fc = 0.9992(4)

80 qubits $p=0.01$

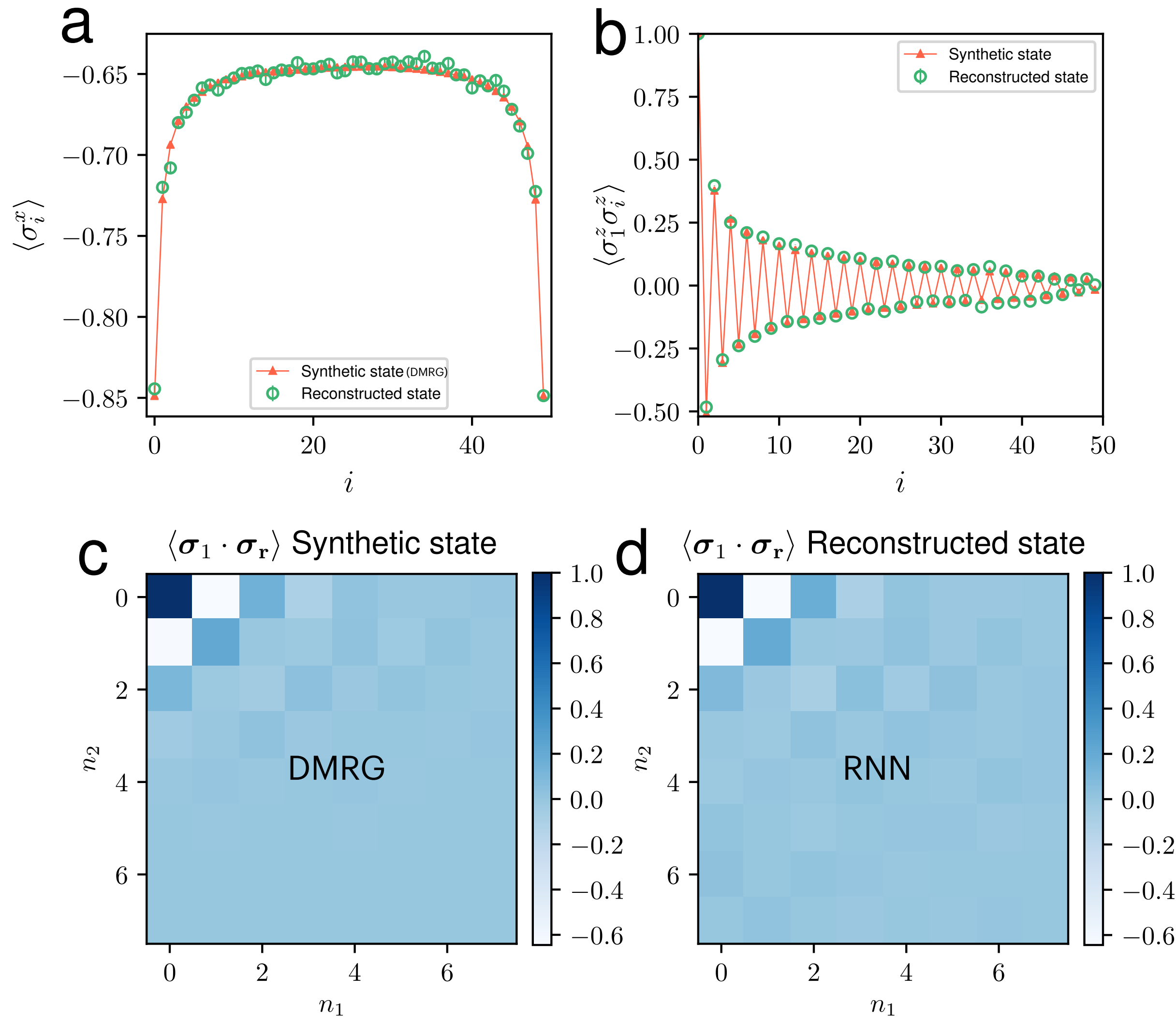
Fc = 0.9988(1)

- System sizes beyond the reach of traditional quantum state tomography.
- Numerically: GHZ states require **linear** number of samples to reach fixed classical fidelity

- Key aspect about this model choice:
- It can represent high-dimensional probability distributions compactly.
- Behaviour similar to a matrix product state if we linearize the model — matches the structure of the state we are targeting well



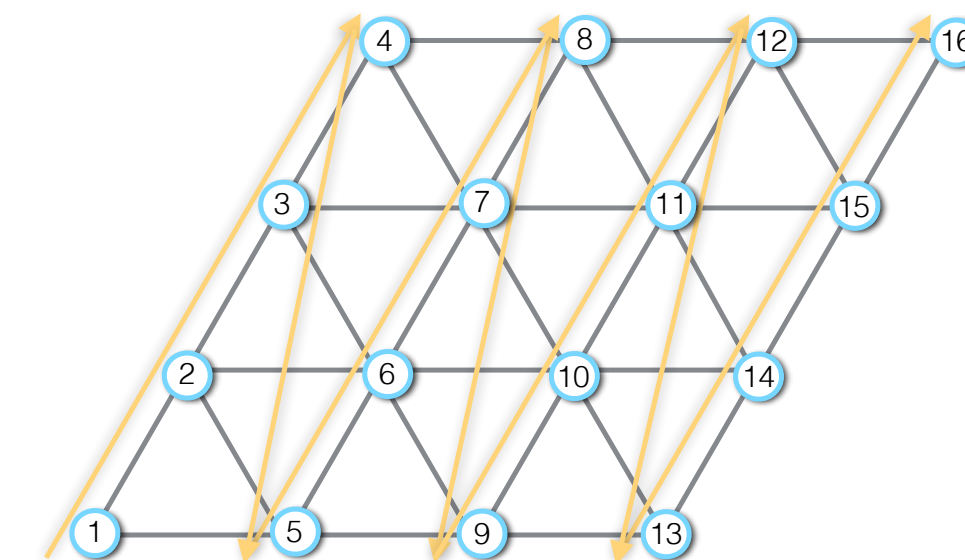
Learning Ground states of local hamiltonians from data



$$\mathcal{H} = J \sum_{ij} \sigma_i^z \sigma_j^z + h \sum_i \sigma_i^x$$

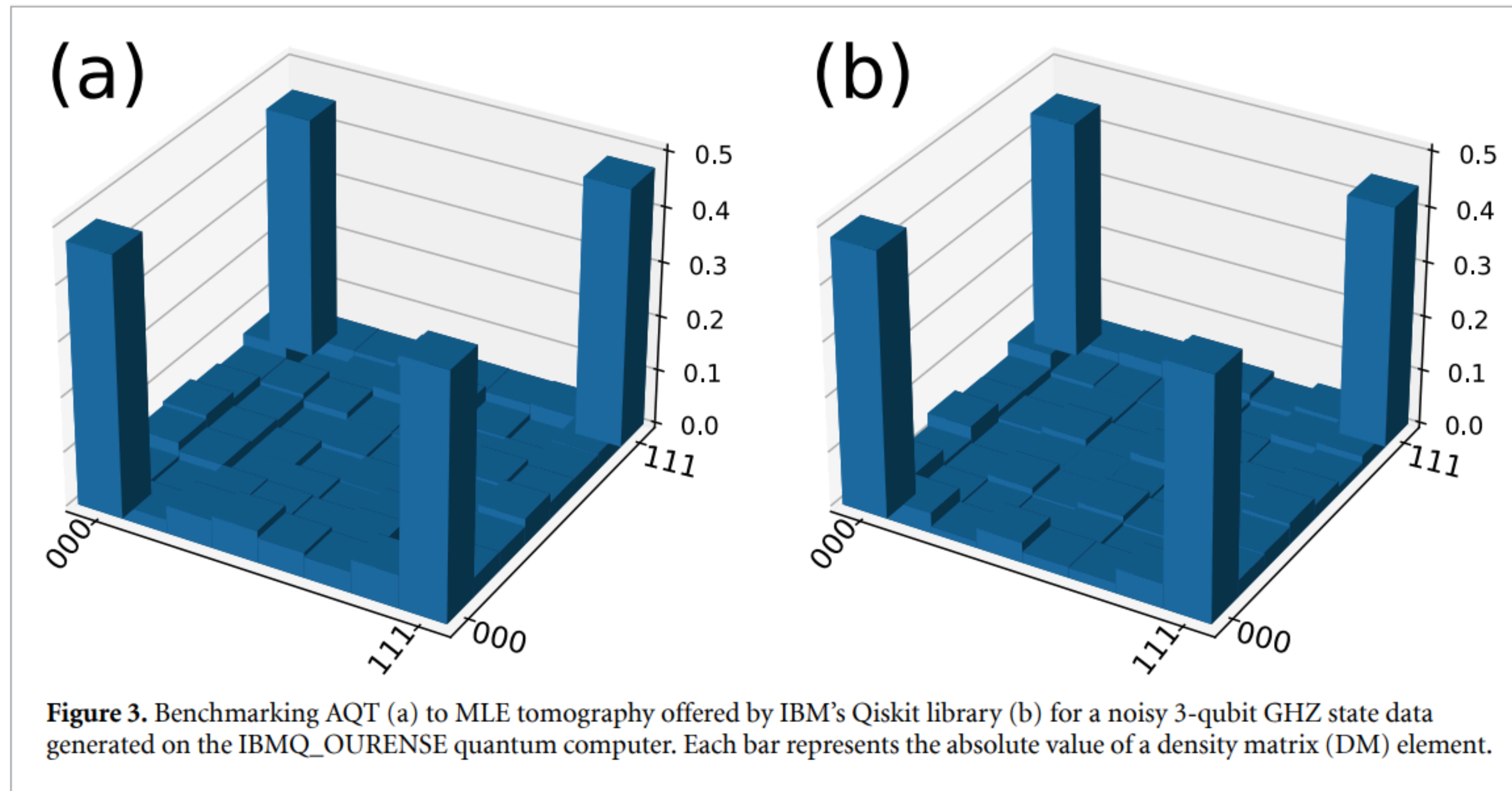
N=50 spins. P(a) is a deep (3 layer GRU) recurrent neural network language model.

$$H = J \sum_{i,j} \sigma_i \cdot \sigma_j$$



EXPERIMENTAL DEMONSTRATION

GHZ state with 3 qubits



Extensions of these ideas

- Probabilistic simulation of quantum circuits using a deep-learning architecture. Carrasquilla, Luo, Perez, et al. PHYSICAL REVIEW A **104**, 032610 (2021)
- Quantum process tomography with unsupervised learning and tensor networks. Torlai, Wood, Acharya, Carleo, Carrasquilla, Aolita. Nature Communications volume **14**, Article number: 2858 (2023)
- Autoregressive Neural Network for Simulating Open Quantum Systems via a Probabilistic Formulation. Luo, Chen, Carrasquilla, Clark. Phys. Rev. Lett. **128**, 090501 (2022)
- Time-dependent variational principle for open quantum systems with artificial neural networks. Reh, Schmitt, Gärttner. Phys. Rev. Lett. **127**, 230501 (2021)

Conclusions

- Introduced 2 examples of applications of ML ideas quantum systems.
- Ground states of the toric code.
- Quantum state reconstruction with RNNs.
- Body of recent work showcases the opportunities that machine learning techniques, ideas, and research culture can spark in the field of quantum physics.
- Now is a privileged time for quantum research — enormous opportunities arising from artificial intelligence and quantum computing, two of today's most promising computational paradigms.
- The cross-fertilization is exciting and growing steadily.