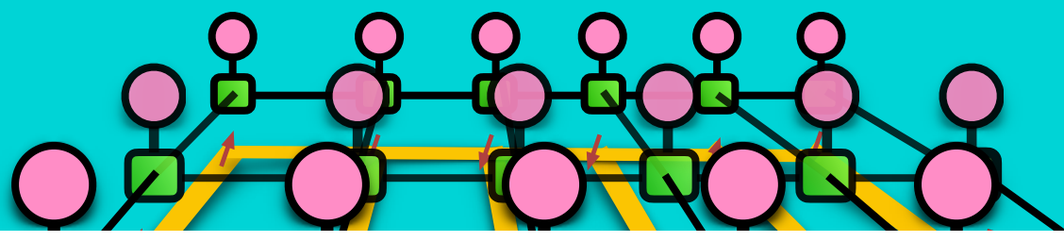
NEURAL AUTOREGRESSIVE MODELS FOR MANY-BODY PHYSICS

Juan Felipe Carrasquilla Álvarez
Vector Institute

ECT* Trento Workshop "Machine Learning for High Energy Physics, on and off the Lattice"
Sept. 27th 2021



VECTOR INSTITUTE | INSTITUT VECTEUR

UNIVERSITY OF WATERLOO

CIFAR

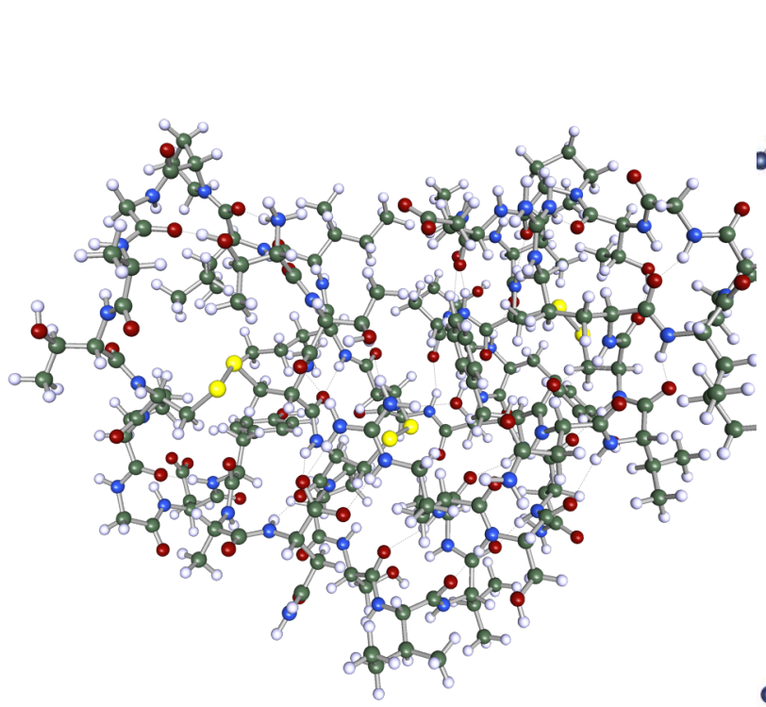
NSERC CRSNG

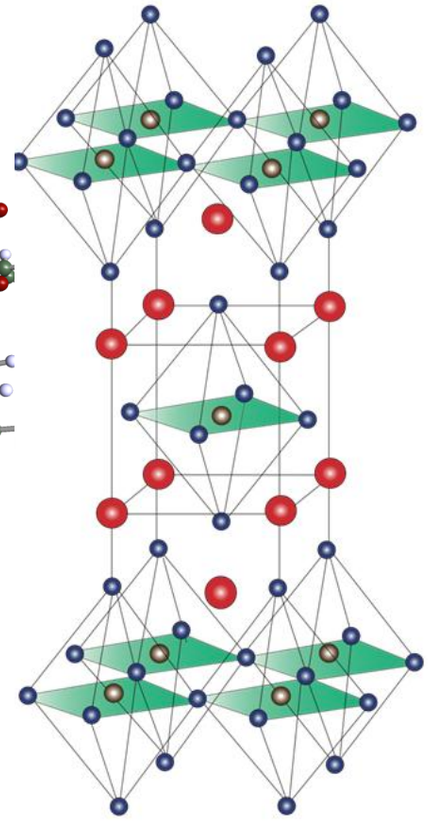
WHY USE MACHINE LEARNING FOR QUANTUM PHYSICS? COMMON STRUCTURE

THE PROBLEM OF HIGH DIMENSIONALITY IN QUANTUM MECHANICS

- Generic specification of a quantum state requires resources exponentially large in the number of degrees of freedom N
- Today's best supercomputers can solve the wave equation **exactly** for systems with a maximum of ~45 spins.
- Yet, technologically relevant problems in chemistry, condensed matter physics, and quantum computing are much larger than 45.
- We have to exploit structure of the problem

$|\Psi\rangle$ vector with 2^N



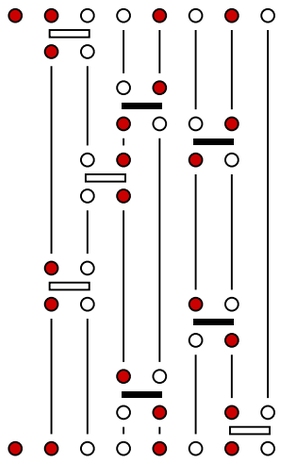


THE PROBLEM OF HIGH DIMENSIONALITY IN QUANTUM MECHANICS

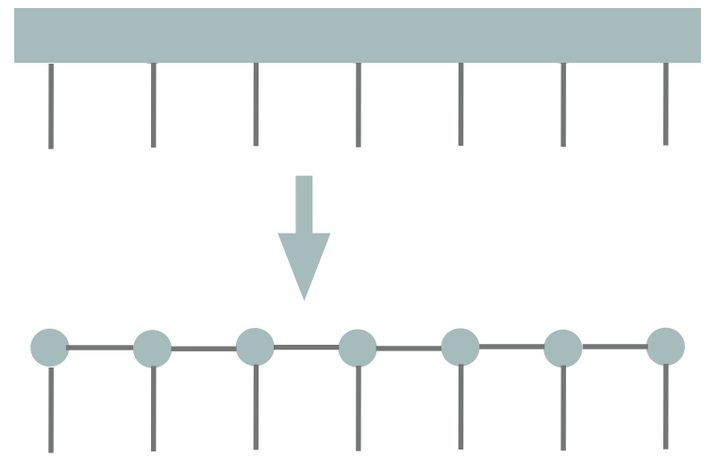
- Amount of information smaller than the maximum capacity — problems have structure and we exploit it
- Quantum Monte Carlo: stochastic exploration of most important regions of the gigantic state space.
- Tensor Networks: quantum states realized in nature have little entanglement
- Both techniques have led to profound implications to our understanding of quantum many-body physics

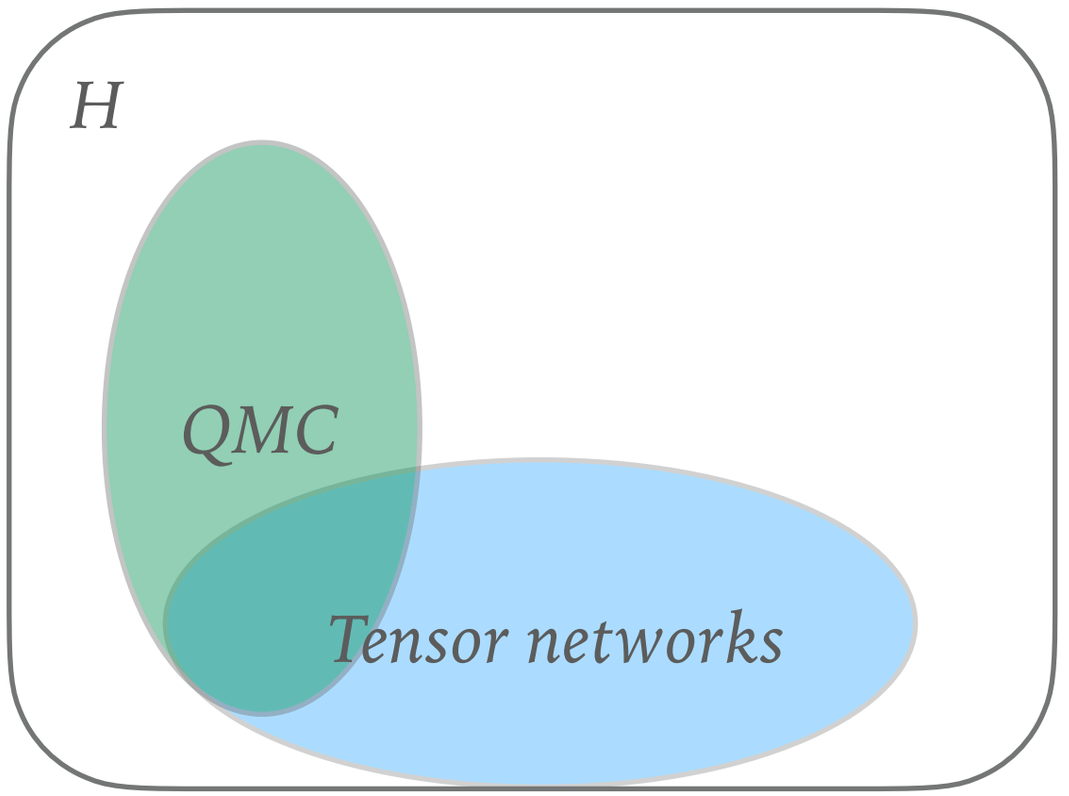
$|\Psi\rangle$ vector with 2^N

QMC



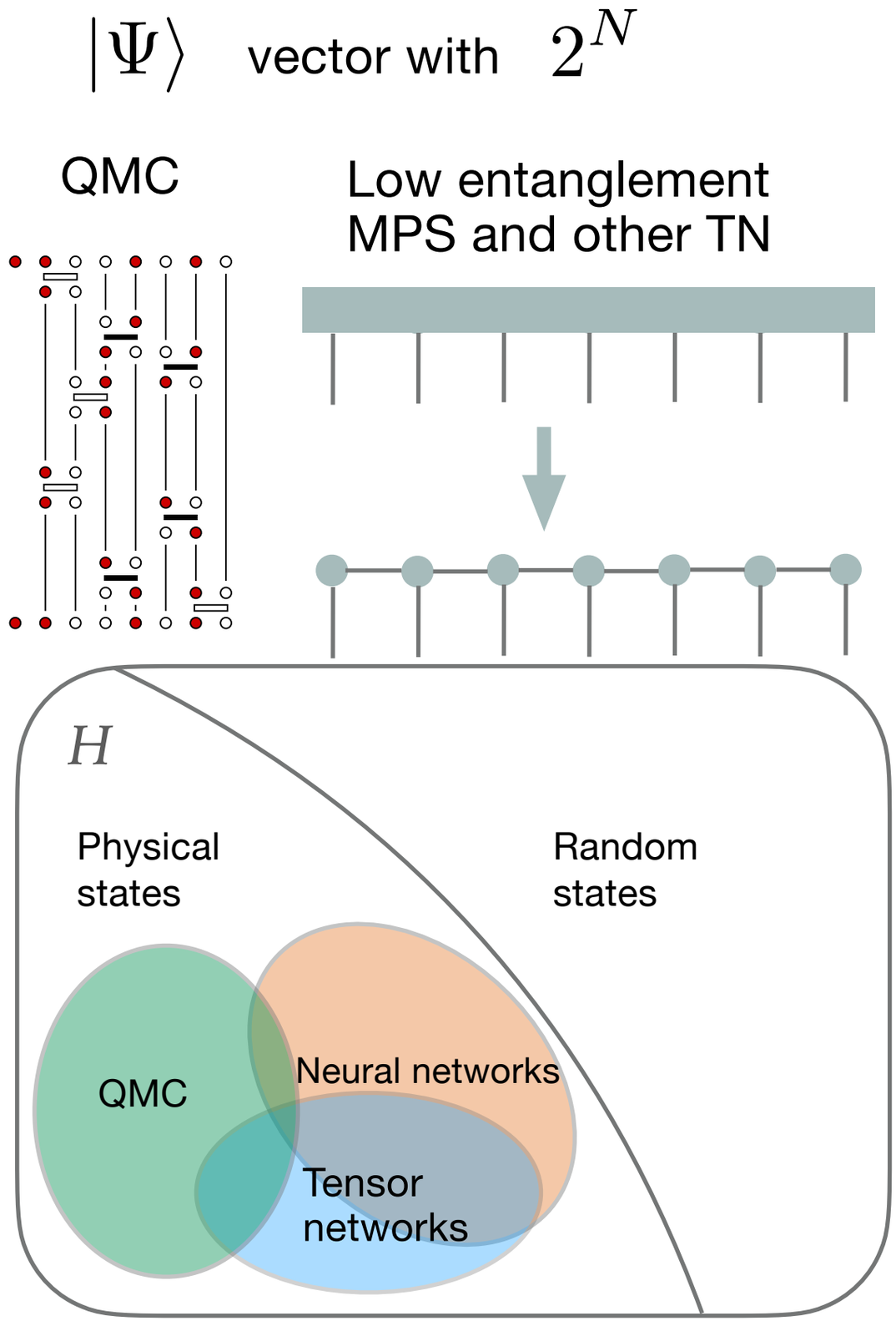
Low entanglement MPS and other TN





THERE IS STILL HOPE FOR CLASSICAL ALGORITHMS

- Machine learning (ML): ML community deals with highly structured problems arising in natural datasets.
- Insight: both quantum and ML problems have a lot of shared structure and symmetry.
- What are these commonalities and are they important beyond mere resemblance?



DIMENSIONALITY OF QUANTUM SYSTEMS VS NEURAL MACHINE TRANSLATION

$|\Psi\rangle$ vector with 2^N

- Today's best supercomputers can solve the wave equation **exactly** for systems with a maximum of ~45 spins.

$$2^N \sim 3.5 \times 10^{13}$$

- Language translation models live in very high dimensional spaces too (example from “Attention is all you need”)

$$\text{Vocab. Size}^{\text{Max length of sentence}}$$

$$8000^{100} \sim 2.03 \times 10^{390}$$

Very large state space

2. COMMONALITY IN SOME THE MATHEMATICAL OBJECTS

- In unsupervised learning researchers are interested, e.g. in understanding the underlying probability of a dataset. For instance images of handwritten digits

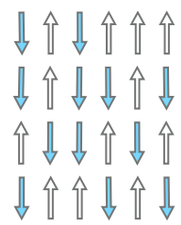
- What is the probability P(4) or what is the P(8)

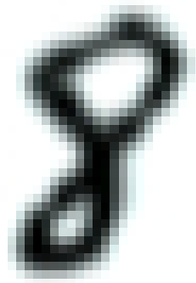
IN ML PEOPLE STUDY P(IMAGE), IN STAT MECH ?

- Boltzmann distribution

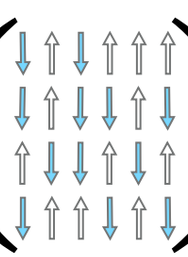
$$P(E) = \frac{e^{-E/k_B T}}{Z} \qquad E = -J \sum_{\langle i,j \rangle} \sigma_i \sigma_j$$

$$\mathrm{P}\left(\right) \longleftrightarrow \mathrm{P}\left(\right)$$





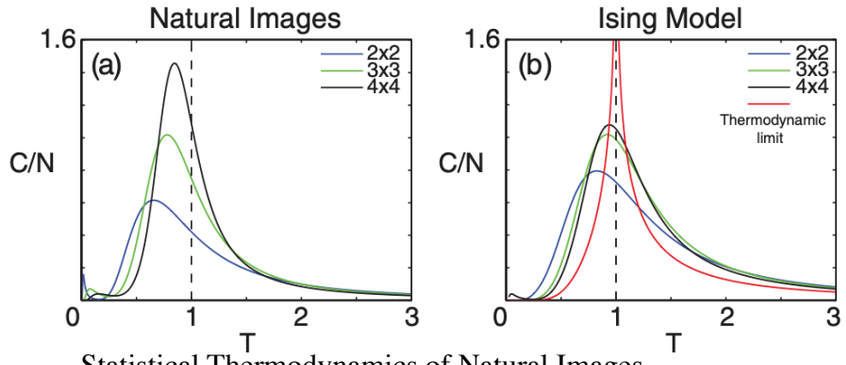
- Wavefunction $\Psi\left(\right)$



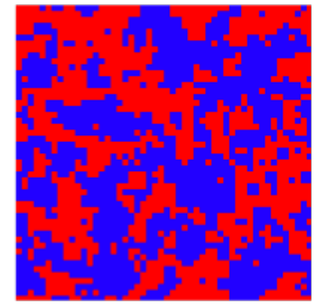
ML and statistical (and quantum) physics are interested in similar high dimensional distributions and wavefunctions

CORRELATIONS AND SYMMETRIES WITH STRIKINGLY SIMILAR STRUCTURE

- Critical correlations:
- Natural language and natural images
- Music
- Flocks of animals

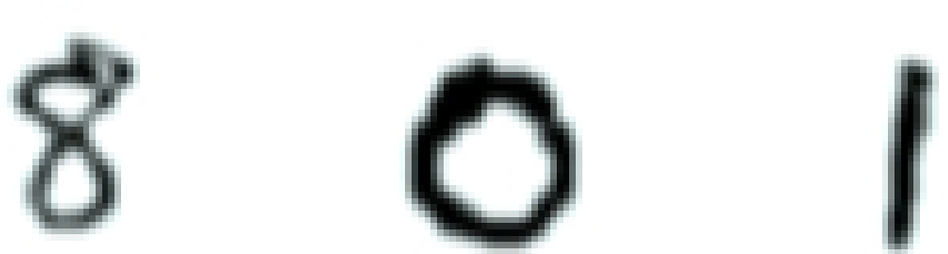


Statistical Thermodynamics of Natural Images
PRL 110, 018701 (2013)

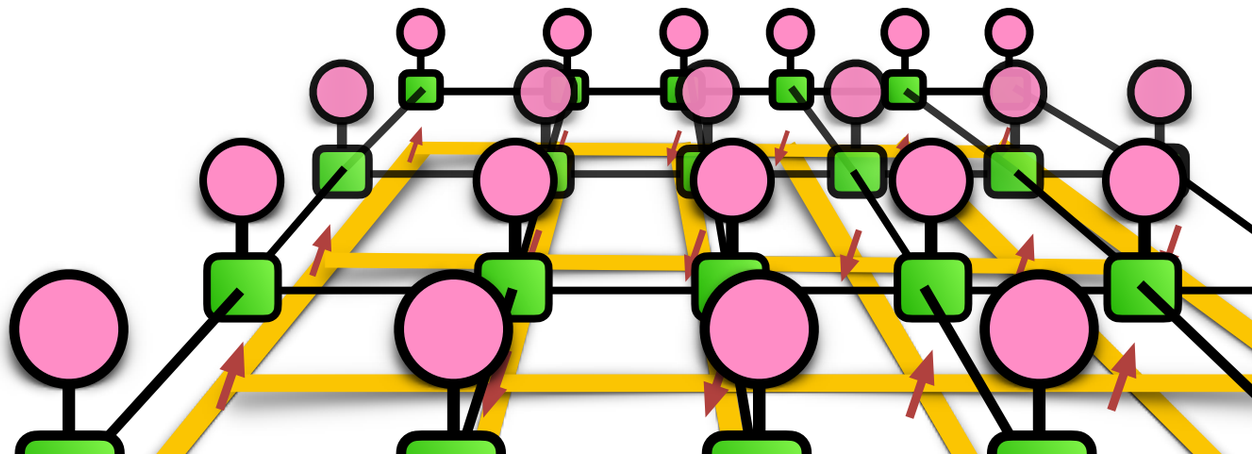


Scale-free correlations in starling flocks. PNAS 107 (26) 11865-11870

- All exhibit power-law decaying correlations identical to a (classical or quantum) at a critical point
- Translational, rotational, reflection, and other symmetries— rich



WHAT ARE AUTOREGRESSIVE MODELS?



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 $P(\boldsymbol{\sigma}) = P(\sigma_1, \sigma_2, \ldots, \sigma_N)$,

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

- To specify P in a tabular form requires exponential resources

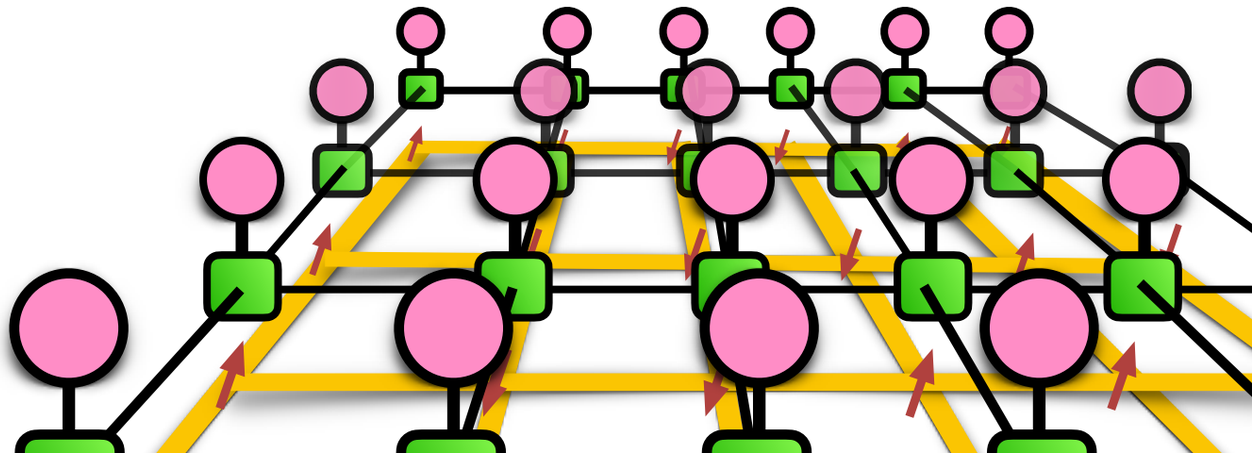
- To alleviate this exponential issue: parametrize the conditionals

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

PROBABILISTIC AUTOREGRESSIVE MODELS AND WAVE FUNCTIONS

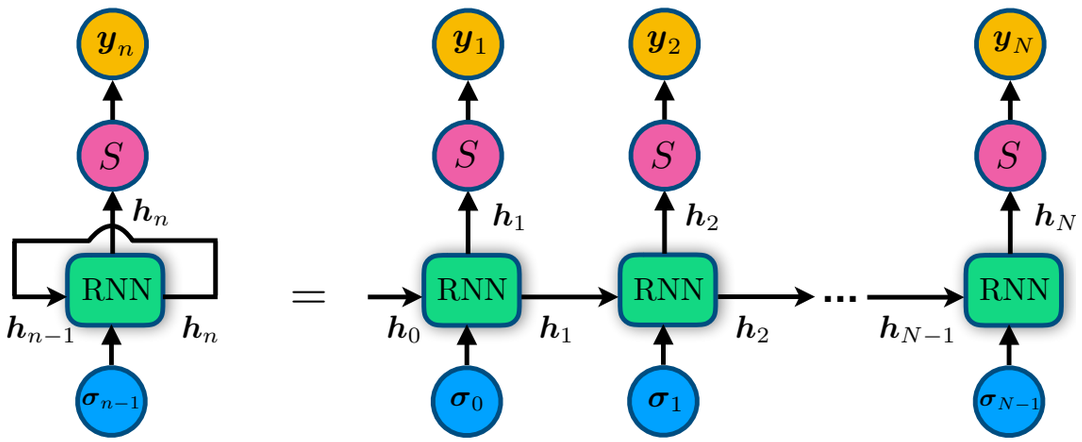
- ✓ Can be exactly sampled easily
- ✓ Computing the probability of a configuration $P(\boldsymbol{\sigma}) = P(\sigma_1, \sigma_2, \ldots, \sigma_N)$ is easy
- ✓ Can be defined in any spatial dimension
- ✓ Easy to encode mean-field theories (e.g. Gutzwiller mean-field theory)
- ✓ We can impose symmetry and other inductive biases useful for physical problems
- ✓ These properties remain true for autoregressive models of the quantum state

A CANONICAL EXAMPLE: THE RECURRENT NEURAL NETWORK



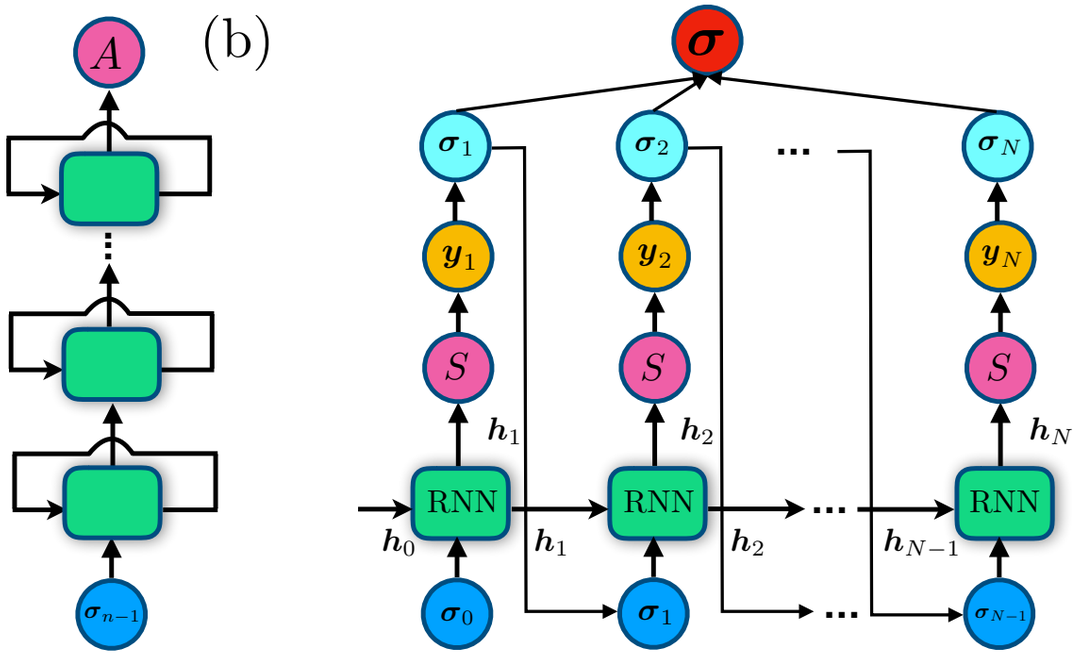
RECURRENT NEURAL NETWORKS (RNN)

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



$$\boldsymbol{h}_n = f\left(W[\boldsymbol{h}_{n-1}; \boldsymbol{\sigma}_{n-1}] + \boldsymbol{b}\right)$$

$$\boldsymbol{y}_n \equiv S\left(U\boldsymbol{h}_n + \mathbf{c}\right) \quad \text{S = Softmax}$$



$$P\left(\sigma_n | \sigma_{n-1}, \ldots, \sigma_1\right) = \boldsymbol{y}_n \cdot \boldsymbol{\sigma}_n$$

Sampling:

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

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

RNNs are universal function approximators. Schäfer and Zimmermann (2006)

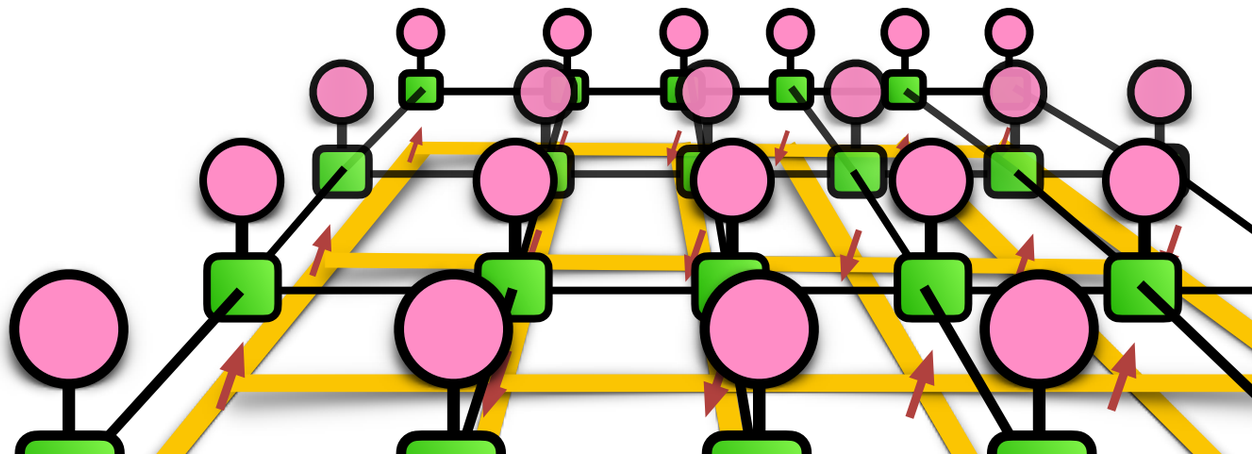
BUT THERE ARE MORE AND LIST IS LONG

- Transformers
- Neural autoregressive density estimators
- Autoregressive flows
- PixelRNN
- PixelCNN
- Wavenet
- ….

TODAY'S TALK:

- Probabilistic formulation of the quantum state
- Learning quantum states and their quantum dynamics
- Recurrent neural network wavefunctions
- Variational neural annealing

PROBABILISTIC SIMULATION OF QUANTUM MECHANICS



FEYNMAN 1981:

Simulating Physics with Computers

Richard P. Feynman

5. CAN QUANTUM SYSTEMS BE PROBABILISTICALLY SIMULATED BY A CLASSICAL COMPUTER?

Now the next question that I would like to bring up is, of course, the interesting one, i.e., Can a quantum system be probabilistically simulated by a classical (probabilistic, I'd assume) universal computer? In other words, a computer which will give the same probabilities as the quantum system does. If you take the computer to be the classical kind I've described so far, (not the quantum kind described in the last section) and there're no changes in any laws, and there's no hocus-pocus, the answer is certainly, No! This is called the hidden-variable problem: it is impossible to represent the results of quantum mechanics with a classical universal device. To learn a little bit about it, I say let us try to put the quantum equations in a form as close as

Feynman concludes:

the great difficulty. The only difference between a probabilistic classical world and the equations of the quantum world is that somehow or other it appears as if the probabilities would have to go negative, and that we do not know, as far as I know, how to simulate. Okay, that's the fundamental problem. I don't know the answer to it, but I wanted to explain that if I try my best to make the equations look as near as possible to what would be imitable by a classical probabilistic computer, I get into trouble.

Motivated the field of quantum computing

This is all still true today and is **fundamentally** linked to the notion of quantum speed-up in quantum computing.

WE CAN'T SIMULATE QUANTUM DYNAMICS EXACTLY USING PROBABILITY, BUT CAN WE DO IT APPROXIMATELY WITH ML MODELS?

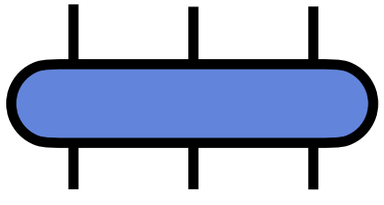
PROBABILISTIC REPRESENTATION OF QUANTUM MECHANICS

HOW IS A QUANTUM STATE TRADITIONALLY DESCRIBED?

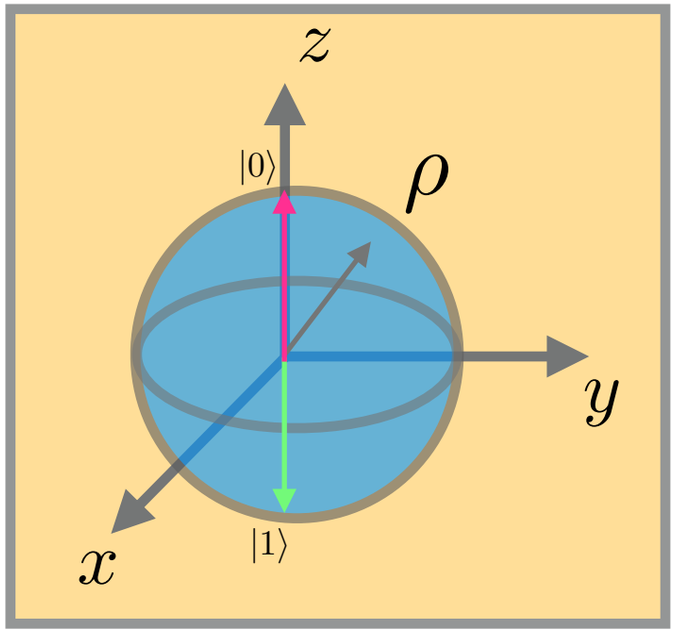
- Traditionally, a quantum state is represented through a a **density matrix** describes the statistical state of a system in quantum mechanics. Everything we can possibly know about a quantum system is encoded in the density matrix.

- For one qubit, all the states live in the Bloch sphere.

- But there are alternative representations. Are these useful?



$$\rho$$



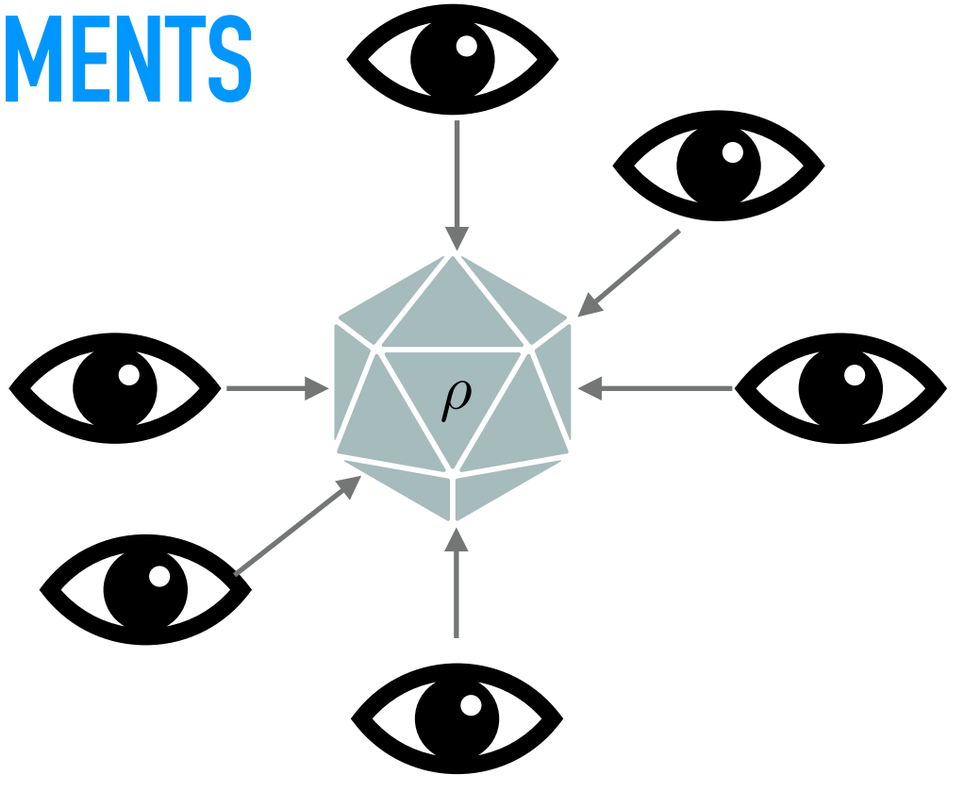
HOW TO REPRESENT A QUANTUM STATE WITH ONLY PROBABILITY?

MEASUREMENTS: POSITIVE OPERATOR-VALUED MEASURE (POVM)

- Born Rule $P(\mathbf{a}) = \mathrm{Tr}\,\rho M^{\mathbf{a}}$ quantum theory ⇔ experiment
- POVM elements $\mathbf{M} = \{M^{(a)} \mid a \in \{1, ..., m\}\}$
- Positive semidefinite operators $\sum_i M^{(a)} = \mathbb{1}$

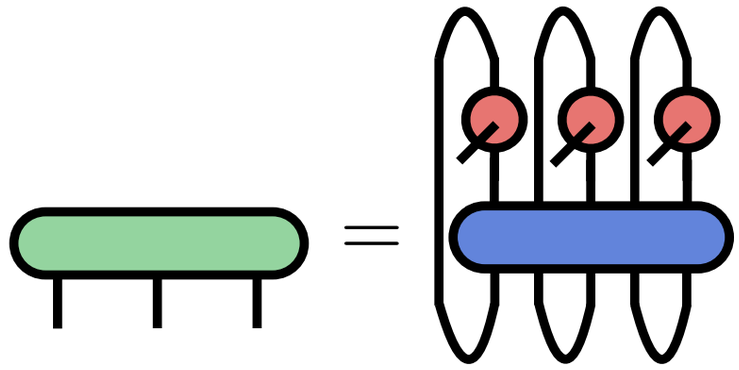
INFORMATIONALLY COMPLETE MEASUREMENTS

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



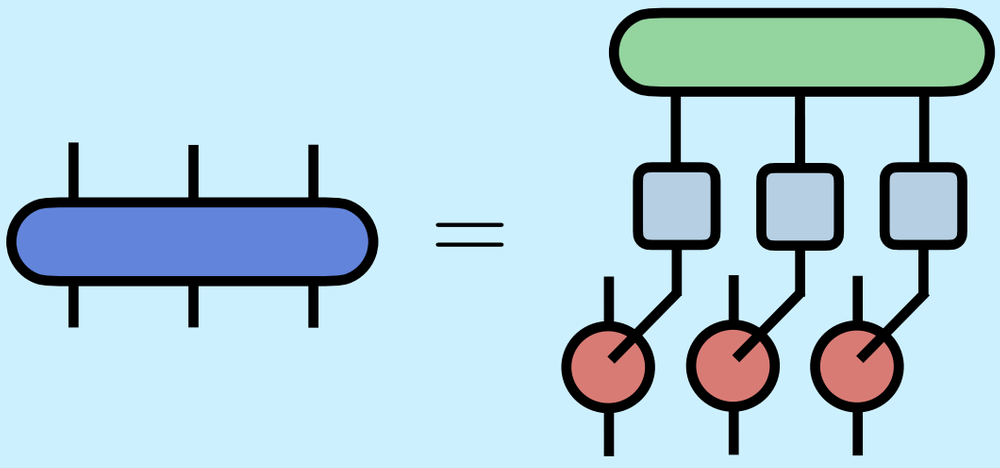
GRAPHICAL NOTATION AND INVERSE

BORN RULE $P(\mathbf{a}) = \mathrm{Tr}\,\rho M^{\mathbf{a}}$



INFORMATIONALLY COMPLETENESS ——>THIS RELATION CAN BE INVERTED

$$\rho = \sum_{a,a'} T^{-1}_{a,a'} P(a') M^{(a)}$$



- Unitary evolution
- Schrodinger equation
- Linblad equation
- Measurements
- Etc

$$i\frac{\partial \rho}{\partial t} = [\mathcal{H}, \rho] \quad \text{BORN RULE} \quad i\frac{\partial P(\boldsymbol{a}'', t)}{\partial t} = \sum_{\boldsymbol{a},\boldsymbol{a}'} \mathrm{Tr}\left(\left[\mathcal{H}, M^{(\boldsymbol{a})}\right] M^{(\boldsymbol{a}'')}\right) T^{-1}_{\boldsymbol{a},\boldsymbol{a}'} P(\boldsymbol{a}', t)$$

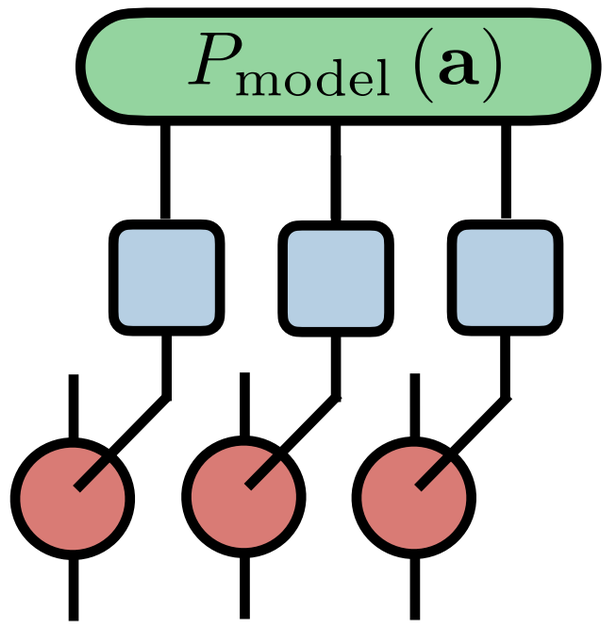
INSIGHT: PARAMETRIZE STATISTICS OF MEASUREMENTS AND INVERT

Instead of parameterizing ρ, create an autoregressive model of $P(a)$

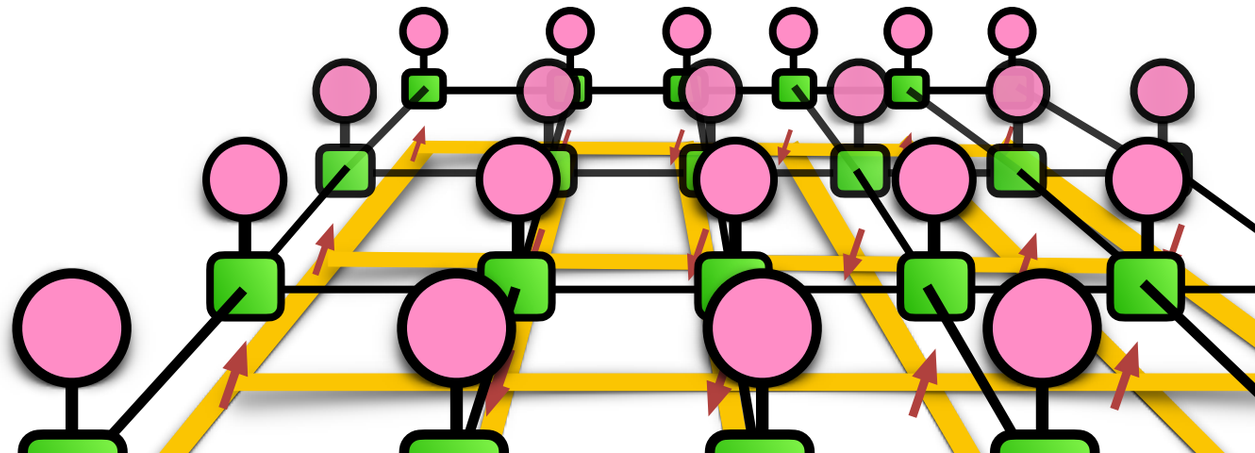
$$P(\mathbf{a}) = \mathrm{Tr}\, \rho M^{\mathbf{a}}$$

Autoregressive models (RNNs and transformer)

$\rho_{\mathrm{model}} =$



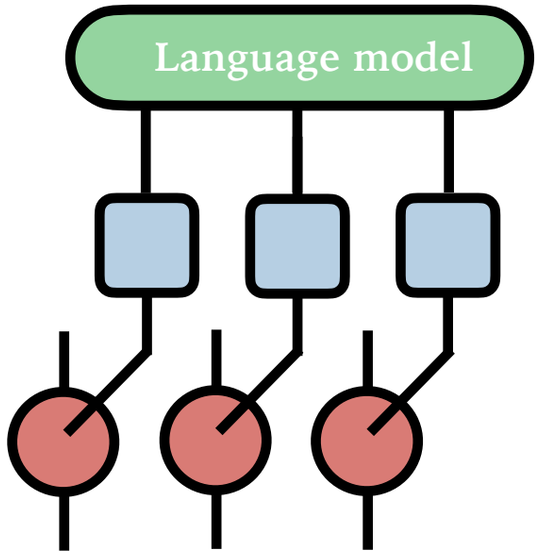
EXAMPLE: LEARN A QUANTUM STATE FROM MEASUREMENTS



NEED TO GO BEYOND STANDARD QUANTUM STATE TOMOGRAPHY

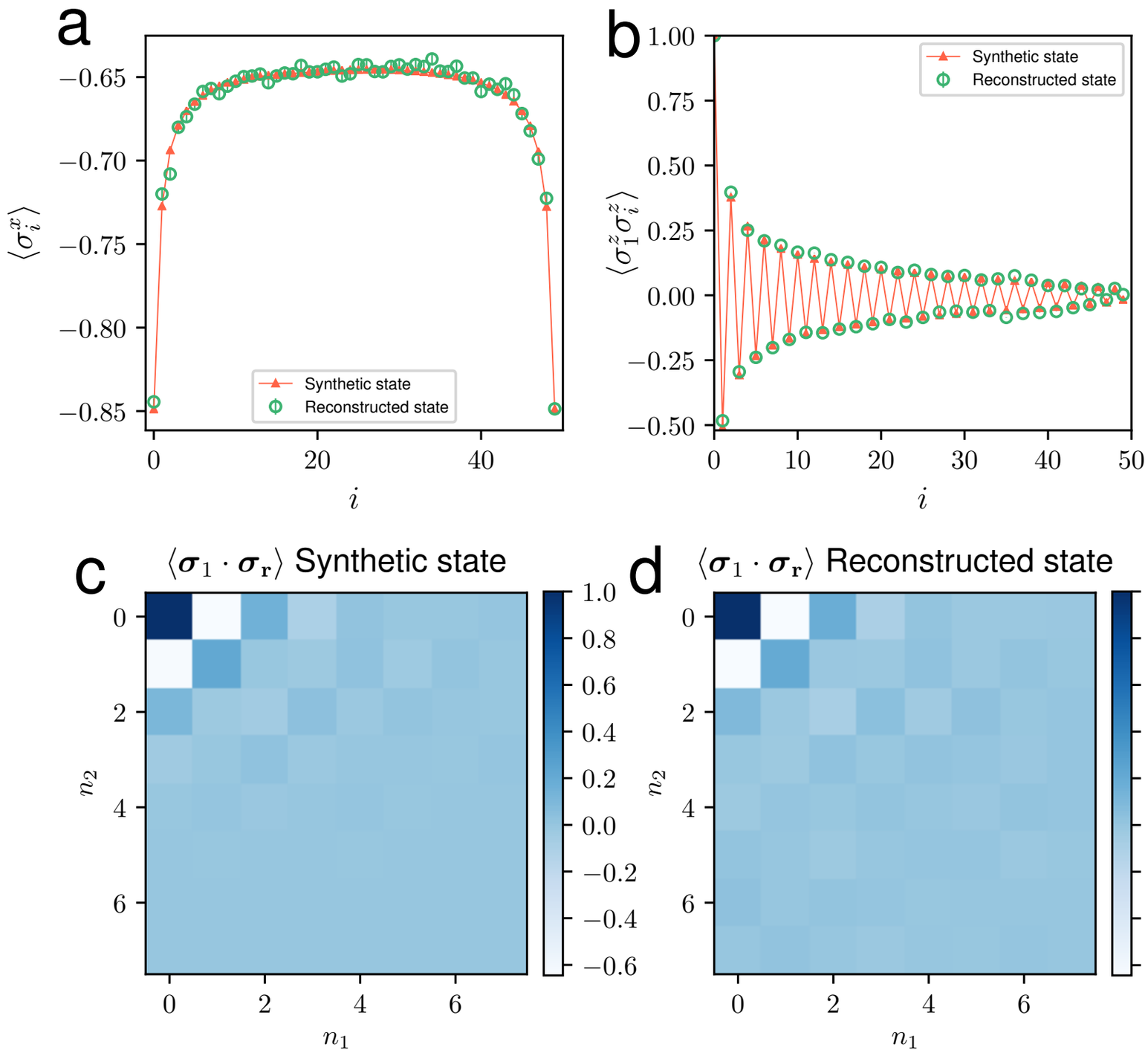
- Prepare an unknown quantum state
- Apply a measurement that probes enough information about the quantum state
- Repeat and collect the statistics of the measurement
- Infer a reconstruction of the state consistent with the measurement outcomes

$\rho^* =$



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

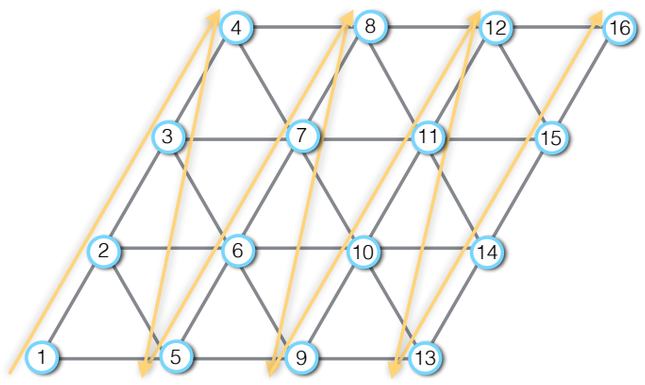
LEARNING GROUND STATES OF LOCAL HAMILTONIANS FROM DATA (RNN)



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

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

$$H = J \sum_{i,j} \boldsymbol{\sigma}_i . \boldsymbol{\sigma}_j$$



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

EXPERIMENTAL DEMONSTRATION

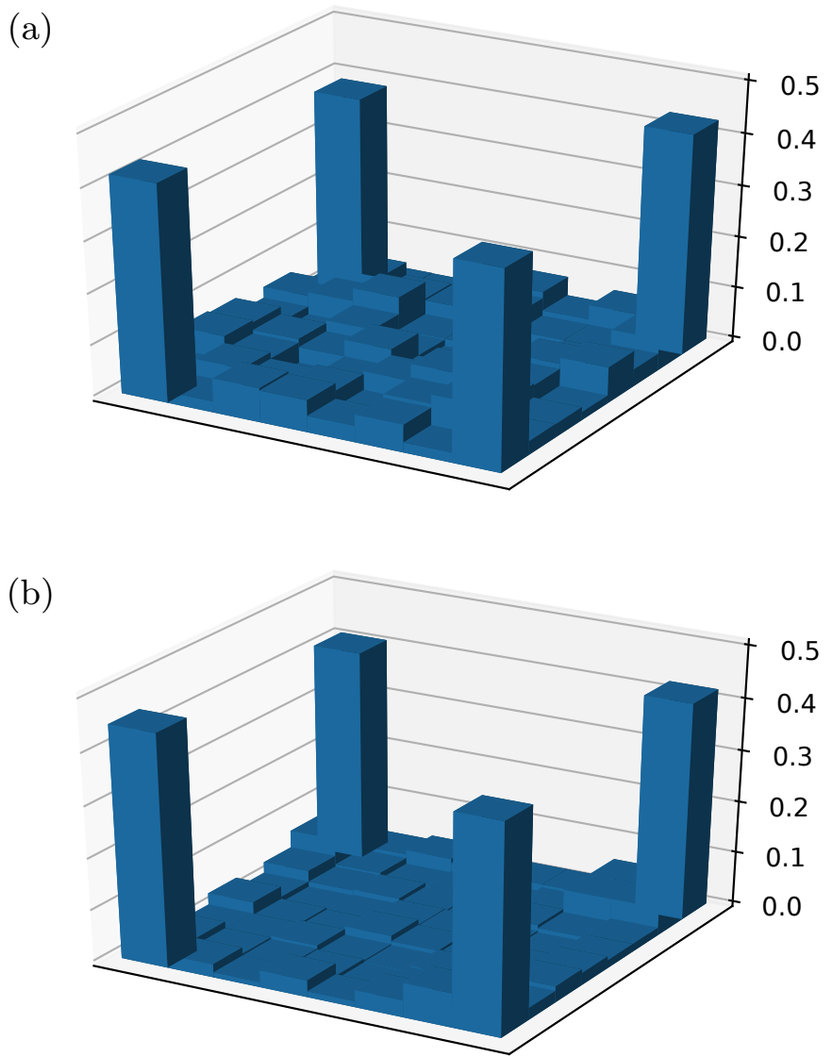
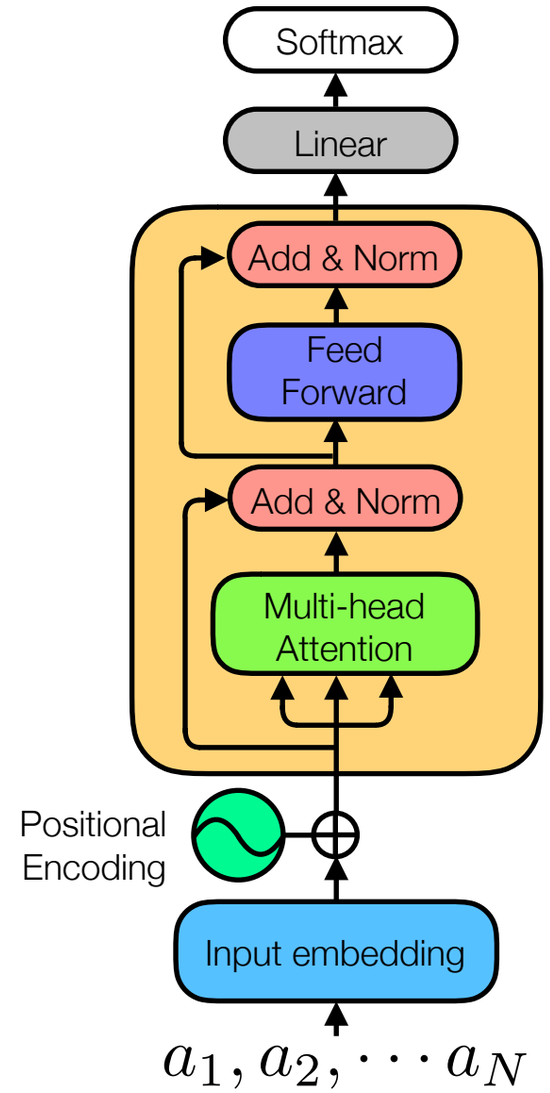


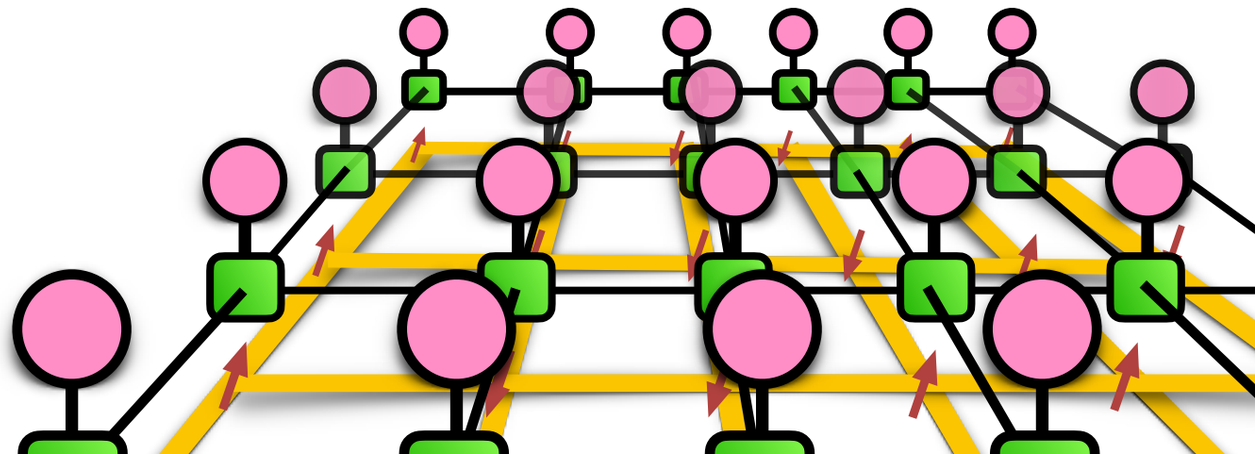
FIG. 5. Benchmarking AQT (a) to MLE tomography offered by IBM's Qiskit library (b) for a noisy 3-qubit QHZ state data generated on the IBMQ_OURENSE quantum computer. Each bar represents the absolute value of a density matrix (DM) element.

GHZ state with 3 qubits



Peter Cha, Paul Ginsparg, Felix Wu, Juan Carrasquilla, Peter L. McMahon, Eun-Ah Kim. arXiv:2006.12469 (2020)

SIMULATION OF QUANTUM CIRCUITS WITH AUTOREGRESSIVE MODELS



CIRCUITS IN THE PROBABILISTIC LANGUAGE

Quantum circuits and quantum computing

$$\rho_U = U \rho U^\dagger$$

Unitary matrices U

BORN RULE

Evolution of probability is **almost** classical (e.g. similar to MCMC)

$$P_U(\mathbf{a}'') = \sum_{\mathbf{a}'} P(\mathbf{a}') O_{\mathbf{a}',\mathbf{a}''}$$

Probabilistic gates: **Somewhat** (or quasi-) stochastic matrices

$$O_{\mathbf{a}'\mathbf{a}''} = \sum_{\mathbf{a}} \mathrm{Tr}(U M^{(\mathbf{a})} U^\dagger M^{(\mathbf{a}'')}) T^{-1}_{\mathbf{a},\mathbf{a}'}$$

APPLY ONE GATE

$$P_U(\mathbf{a}'') = \sum_{\mathbf{a}'} P(\mathbf{a}')O_{\mathbf{a}',\mathbf{a}''}$$

Take an initial distribution

Multiply it by a somewhat stochastic matrix

Results in an evolved distribution Pu(a)

APPLY ONE GATE

$$P_U(\mathbf{a}'') = \sum_{\mathbf{a}'} P(\mathbf{a}') O_{\mathbf{a}',\mathbf{a}''}$$

Introduce a model $P_\theta(\mathbf{a})$

Compute a divergence between model and evolved $P_U(\mathbf{a})$ through sampling

Minimize the distance

$$D_{\mathrm{KL}}\left(P_U || P_\theta\right) = -\sum_{\mathbf{a}} P_U\left(\mathbf{a}\right) \ln \frac{P_\theta\left(\mathbf{a}\right)}{P_U\left(\mathbf{a}\right)}$$

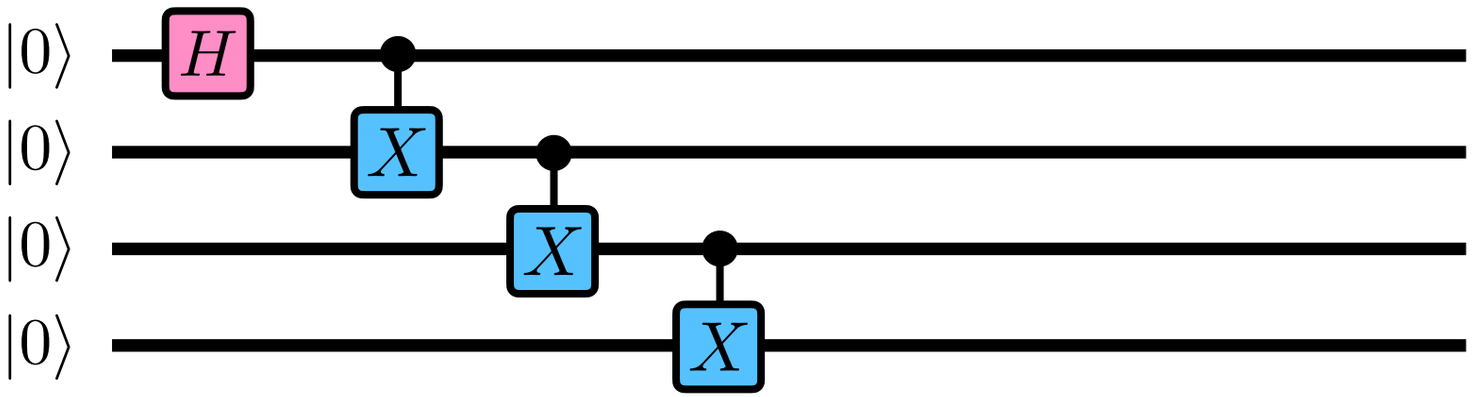
$$D_{\mathrm{KL}}\left(P_U || P_\theta\right) = H(P_U, P_\theta) - H(P_U)$$

$$H(P_U, P_\theta) = -\sum_{\mathbf{a}} P_U\left(\mathbf{a}\right) \ln P_\theta\left(\mathbf{a}\right) = -\sum_{\mathbf{a},\mathbf{a}'} P\left(\mathbf{a}'\right) O_{\mathbf{a}\mathbf{a}'} \ln P_\theta\left(\mathbf{a}\right)$$

RESULTS: STATE PREPARATION FOR SIMPLE STATES

GHZ state

a

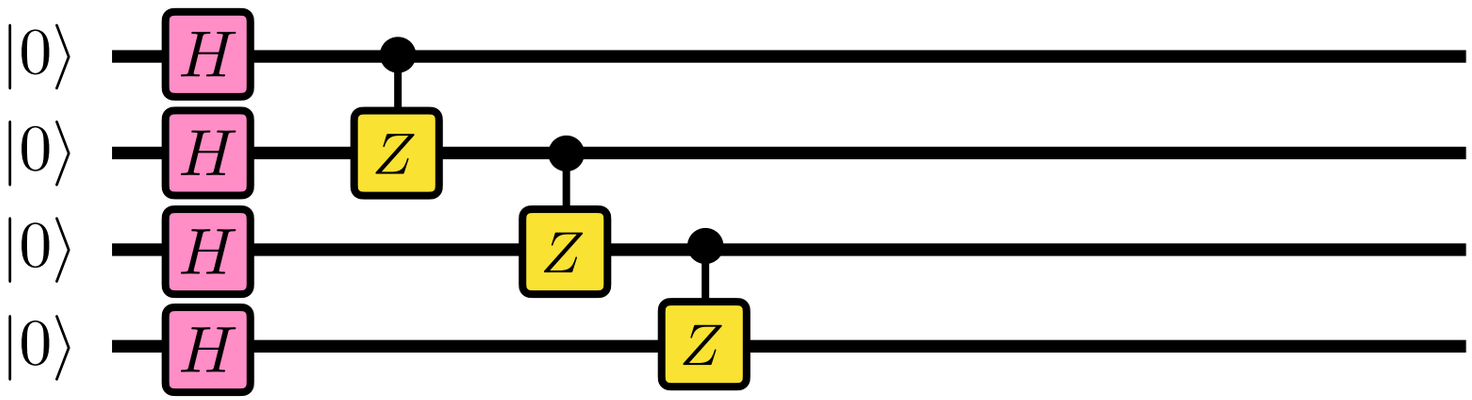


$$\mathrm{CNOT} = cX = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

$$H_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$$

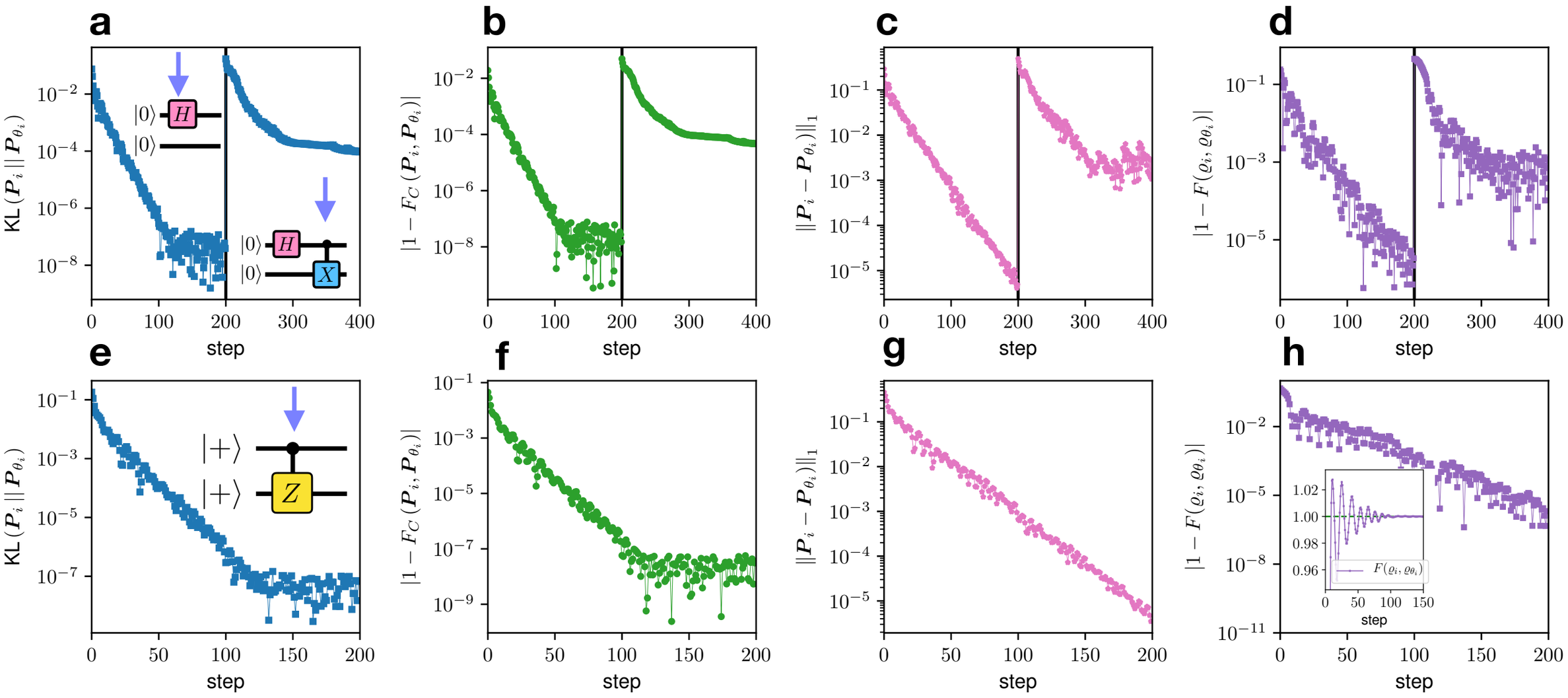
Linear graph state

b



$$cZ = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix}$$

TRAINING DYNAMICS OF THE BELL/GRAPH STATE PREPARATION



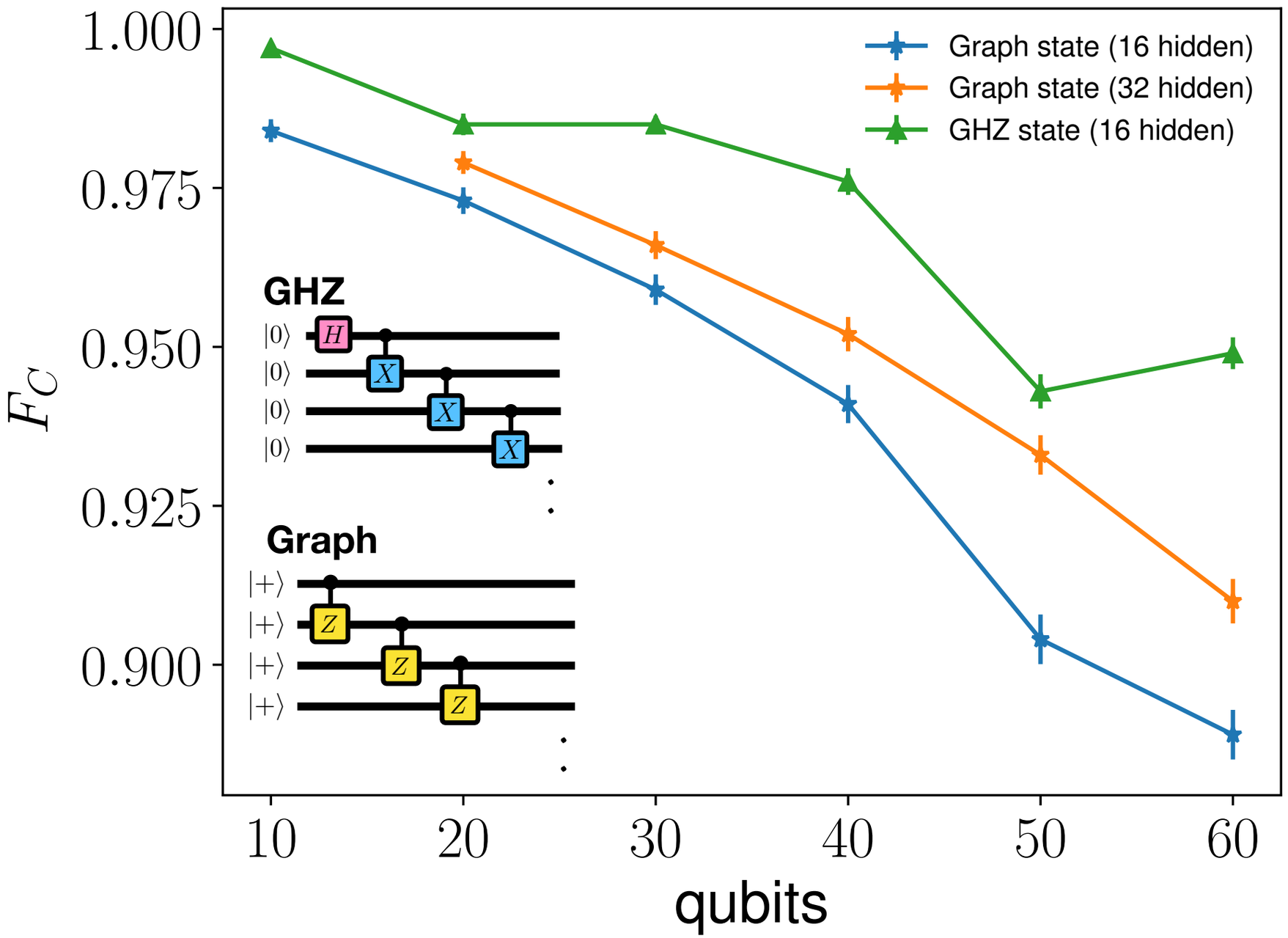
Based on a small transformer

Juan Carrasquilla, Di Luo, Felipe Pérez, Ashley Milsted, Bryan K. Clark, Maksims Volkovs, and Leandro Aolita
Phys. Rev. A 104, 032610

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

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

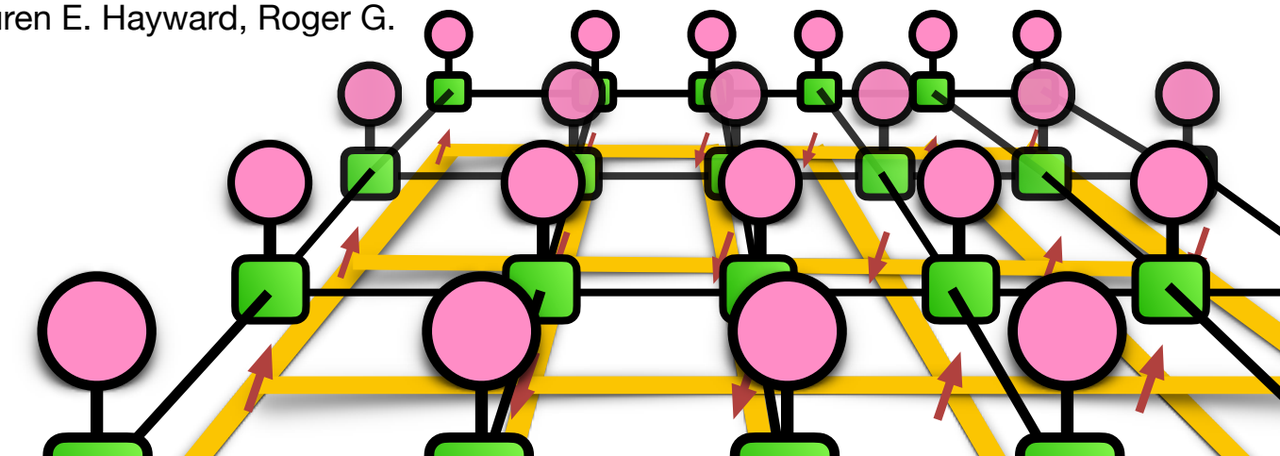
SCALING TO A LARGER NUMBER OF QUBITS



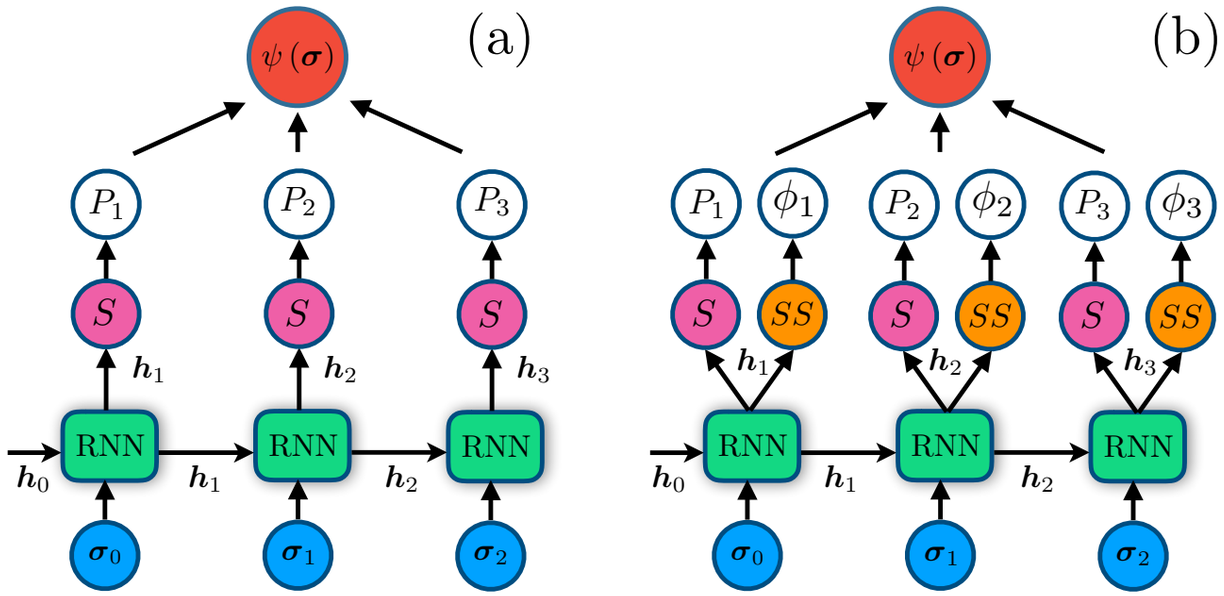
arXiv:1912.11052

RECURRENT NEURAL NETWORK WAVEFUNCTIONS

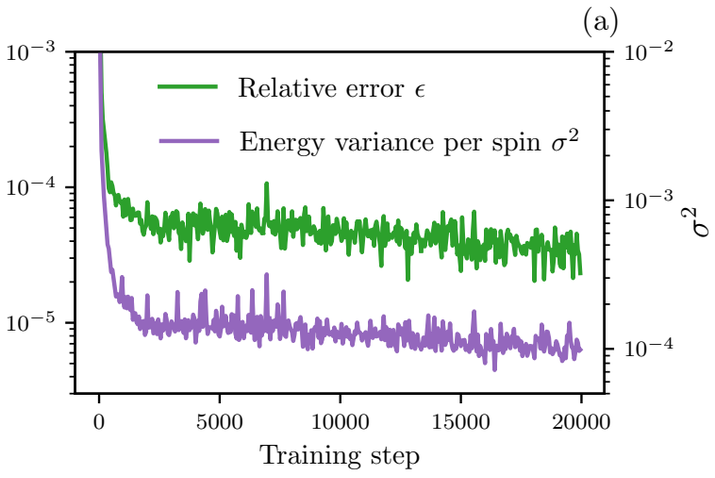
Mohamed Hibat-Allah, Martin Ganahl, Lauren E. Hayward, Roger G. Melko, and Juan Carrasquilla
Phys. Rev. Research **2**, 023358 (2020)



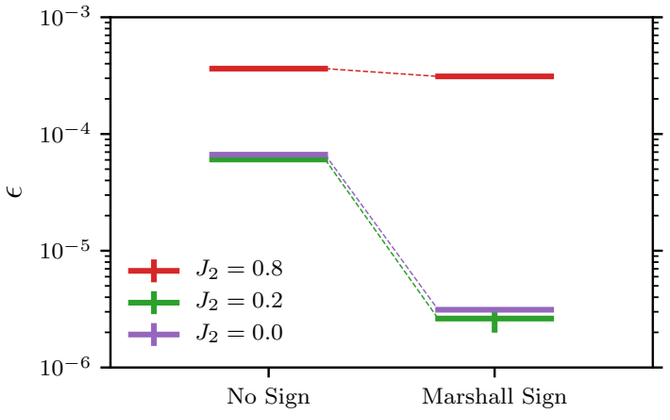
RECURRENT NEURAL NETWORK WAVEFUNCTIONS



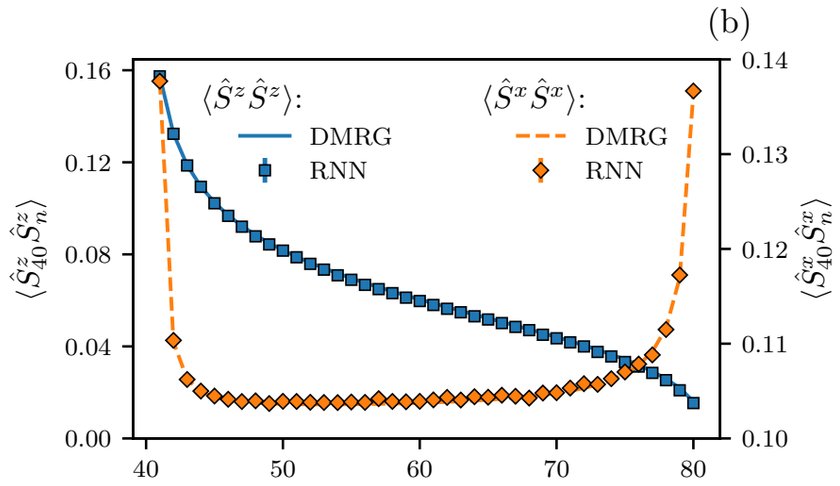
$$|\Psi\rangle = \sum_{\boldsymbol{\sigma}} \psi(\boldsymbol{\sigma}) |\boldsymbol{\sigma}\rangle = \sum_{\boldsymbol{\sigma}} \sqrt{P(\boldsymbol{\sigma})} |\boldsymbol{\sigma}\rangle \qquad |\Psi\rangle = \sum_{\boldsymbol{\sigma}} \exp(\mathrm{i}\phi(\boldsymbol{\sigma})) \sqrt{P(\boldsymbol{\sigma})} |\boldsymbol{\sigma}\rangle$$



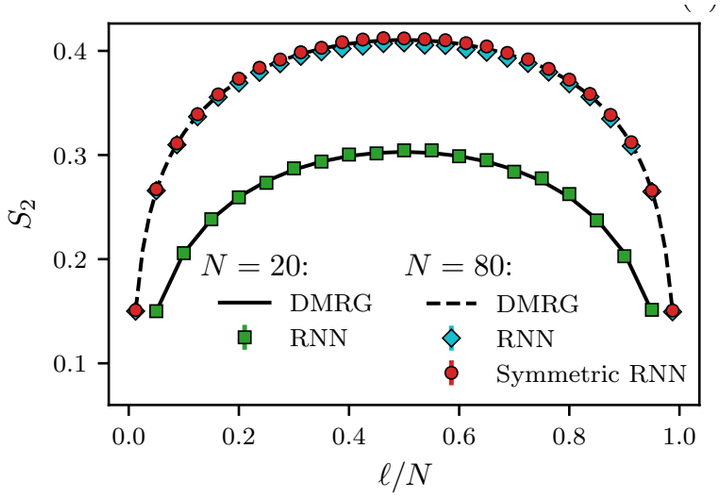
TFIM 1-D N=1000



J1-J2 model in 1-D



TFIM 1-D N=80

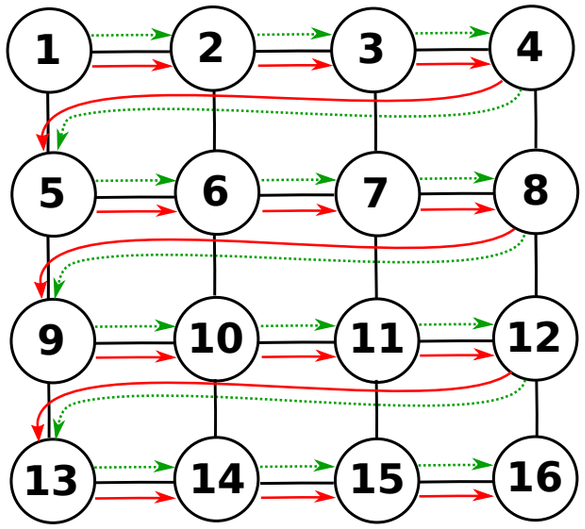


TFIM 1-D N=80

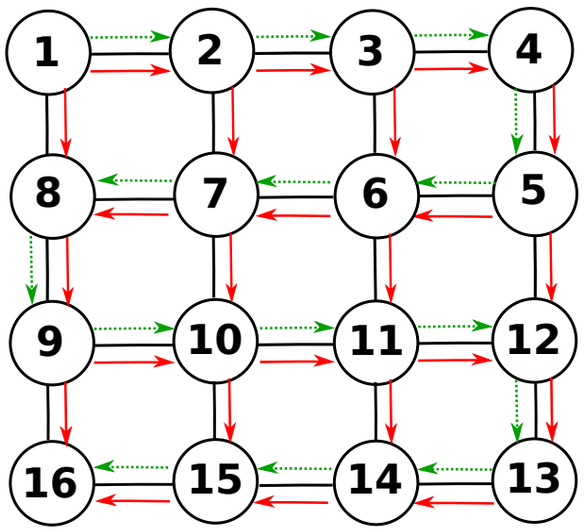
Symmetries: Spin inversion, mirror reflection, Sz. Sign: different Marshall signs for the J1-J2 model can be encoded

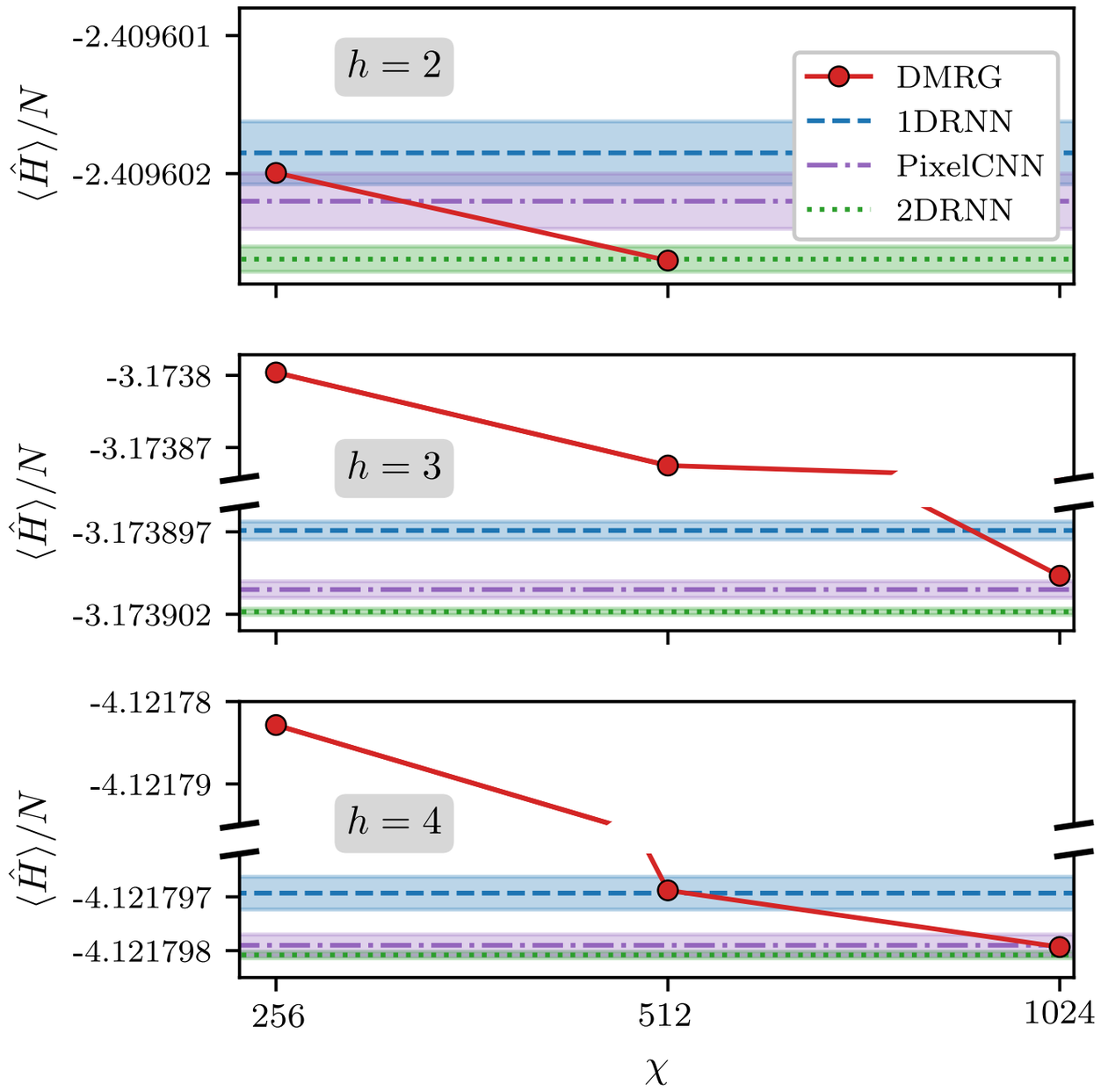
Mohamed Hibat-Allah, Martin Ganahl, Lauren E. Hayward, Roger G. Melko, and Juan Carrasquilla Phys. Rev. Research **2**, 023358 (2020)

RECURRENT NEURAL NETWORK WAVEFUNCTIONS IN 2 DIMENSIONS



(b)



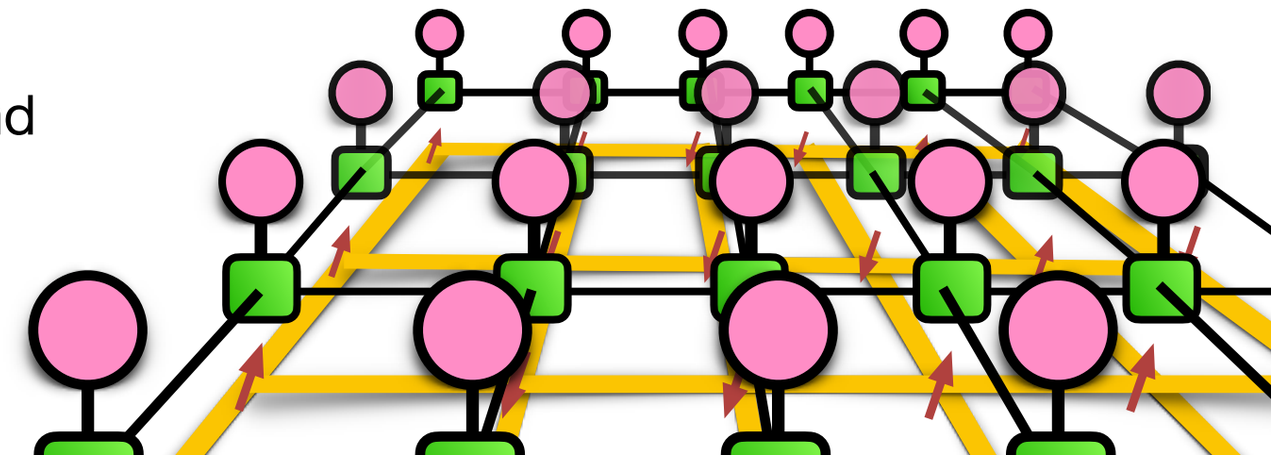


Very accurate and compact: orders of magnitude less parameters than DMRG, pixel CNN. Unlike PEPS, the 2d RNN is tractable.

Mohamed Hibat-Allah, Martin Ganahl, Lauren E. Hayward, Roger G. Melko, and Juan Carrasquilla. Phys. Rev. Research **2**, 023358 (2020)

VARIATIONAL NEURAL ANNEALING

Mohamed Hibat-Allah, Estelle M. Inack, Roeland Wiersema, Roger G. Melko, Juan Carrasquilla. Variational neural annealing. arXiv:2101.10154



Combinatorial optimization

- Many important challenges in science and technology can be cast as optimization problems.

- Areas include artificial intelligence, machine learning, auction theory, software engineering, applied mathematics and theoretical computer science

- Traveling salesman problem, nurse scheduling problem, Vehicle routing problem, factoring, chip placement, etc

Combinatorial optimization by simulated annealing

- Deep connection between materials science and optimization. S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, “Optimization by simulated annealing,” Science 220, 671–680 (1983).

- Annealing in materials and metallurgy: a crystalline solid is heated, kept at high temperature for a while, and then slowly cooled down to its lowest energy state— enables **control** of the properties of the material.

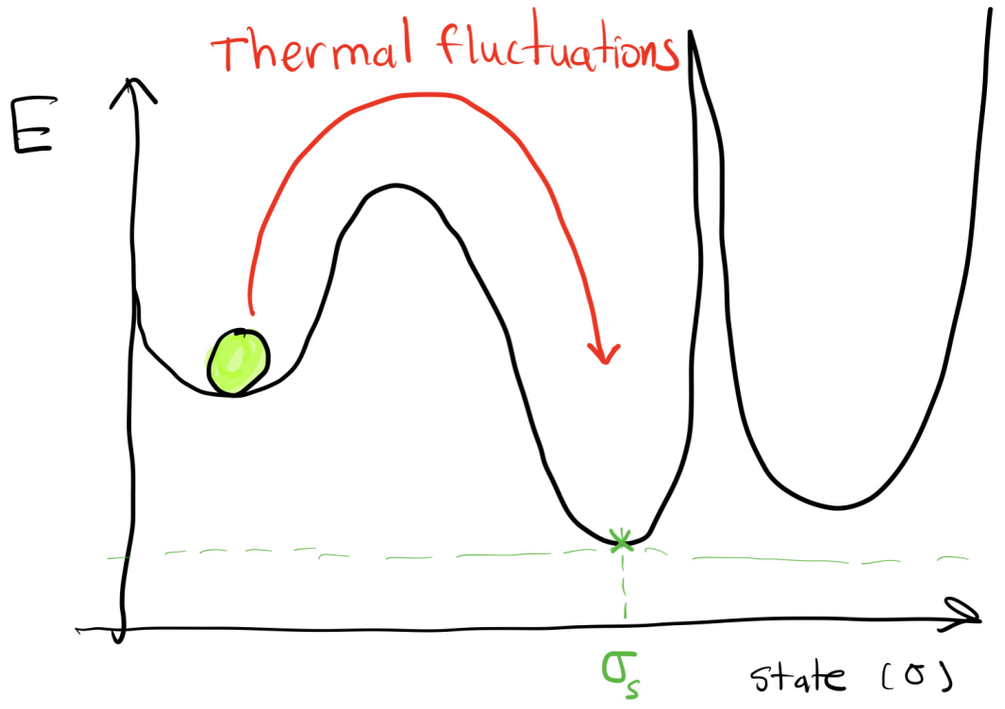
- “We show how the Metropolis algorithm for approximate numerical simulation of many-body system at a finite temperature provides a natural tool for bringing techniques of statistical mechanics to bear on optimization”.

Combinatorial optimization by simulated annealing

- SA mirrors the analogous annealing process in materials science and metallurgy

- The SA algorithm explores an optimization problem's energy landscape via a gradual decrease in thermal fluctuations generated by the Metropolis-Hastings algorithm.

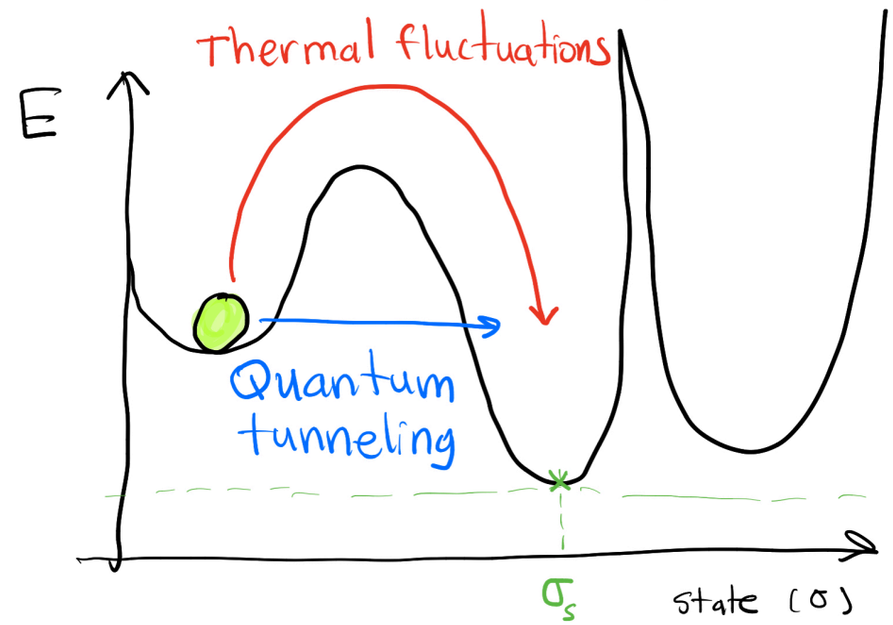
- Temperature is reduced slowly according to some user-defined schedule.

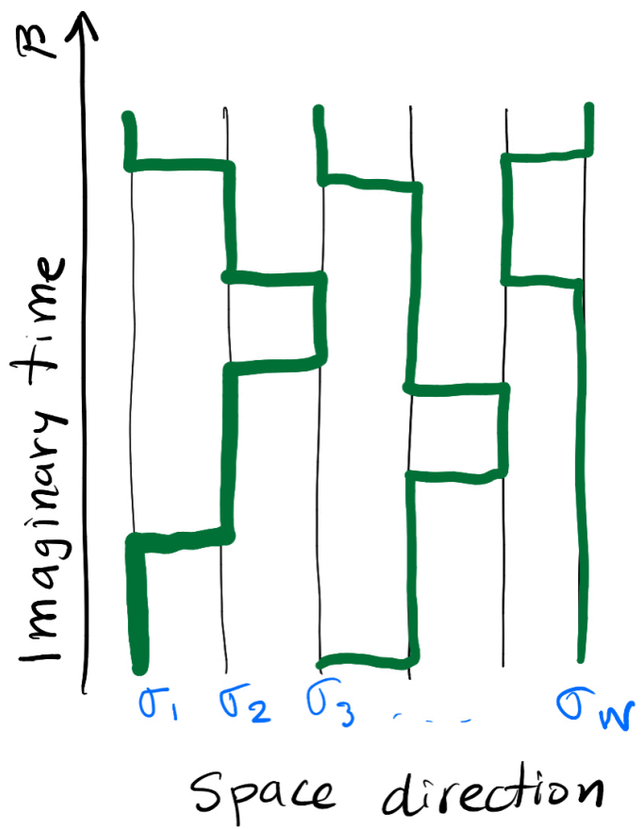


Combinatorial optimization by quantum annealing

- Quantum annealing: Solve an optimization problem using quantum effects:
  $\hat{H}(t) = \hat{H}_{\text{target}} + f(t)\hat{H}_D$

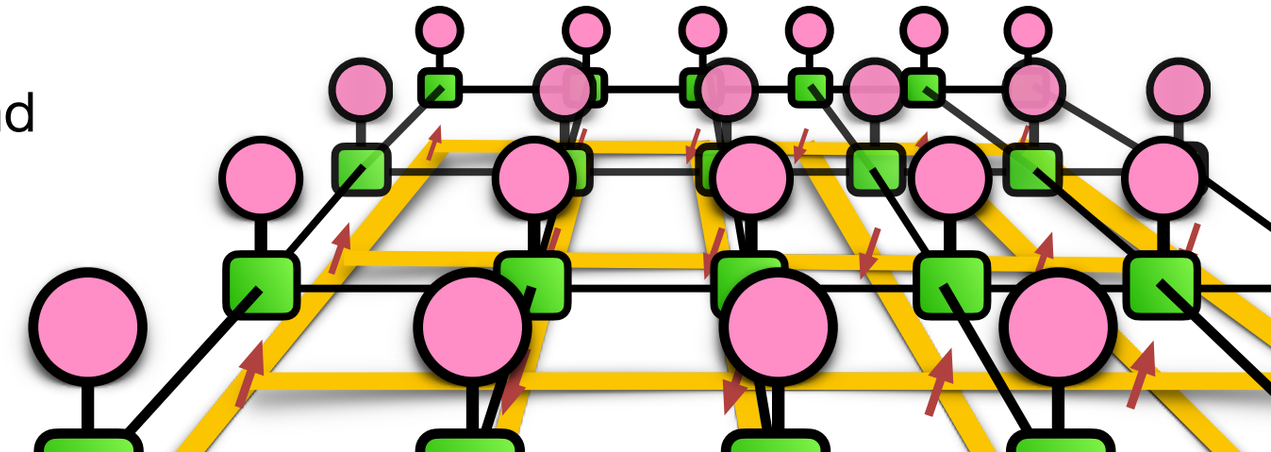
- It is possible to emulate quantum annealing using quantum Monte Carlo—simulated quantum annealing





CAN WE SIMULATE THESE OPTIMIZATION TECHNIQUES VARIATIONALLY?

Mohamed Hibat-Allah, Estelle M. Inack, Roeland Wiersema, Roger G. Melko, Juan Carrasquilla. Variational neural annealing. arXiv:2101.10154



Variational classical annealing

- How to train the model $P_{\boldsymbol{\theta}}(\boldsymbol{\sigma})$ so that it mimics the annealing of the Boltzmann distribution?

- Use variational principle and optimize model's free energy $F_{\boldsymbol{\theta}}(t) = \langle H_{\text{target}} \rangle_{\boldsymbol{\theta}} - T(t)\, S(P_{\boldsymbol{\theta}}) \geq F(t)$

- $F(t)$ is the true free energy of the system at temperature T(t).

- $S(P_{\boldsymbol{\theta}})$ is the entropy of the model $P_{\boldsymbol{\theta}}(\boldsymbol{\sigma})$

- As in SA, temperature is decreased from an initial value T_0 to 0 using a linear schedule function $T(t) = T_0(1-t)$, where $t \in [0,1]$

Mohamed Hibat-Allah, Estelle M. Inack, Roeland Wiersema, Roger G. Melko, Juan Carrasquilla. Variational neural annealing. arXiv:2101.10154.
Dian Wu, Lei Wang, Pan Zhang. Solving Statistical Mechanics Using Variational Autoregressive Networks. Phys. Rev. Lett. 122, 080602 (2019)

Variational quantum annealing

- We can extended this idea to simulated **quantum** annealing
- Promote RNN to a quantum state: $P_{\boldsymbol{\theta}}(\boldsymbol{\sigma}) \rightarrow \Psi_{\boldsymbol{\theta}}(\boldsymbol{\sigma})$
- Modify the cost function:
- $F_{\boldsymbol{\theta}}(t) = \langle H_{\text{target}} \rangle_{\boldsymbol{\theta}} - T(t)\, S(P_{\boldsymbol{\theta}}) \rightarrow E_{\theta} = \langle \Psi_{\boldsymbol{\theta}} | \hat{H}_{\text{target}} | \Psi_{\boldsymbol{\theta}} \rangle - \Gamma(t) \langle \Psi_{\boldsymbol{\theta}} | \hat{H}_{\text{driver}} | \Psi_{\boldsymbol{\theta}} \rangle$
- $\hat{H}_{\text{driver}} = \sum_i \hat{\sigma}_i^x$ typical choice in quantum annealing
- Slowly decrease quantum tunnelling $\Gamma(t) = \Gamma_0(1 - t)$, where $t \in [0,1]$

Variational neural annealing

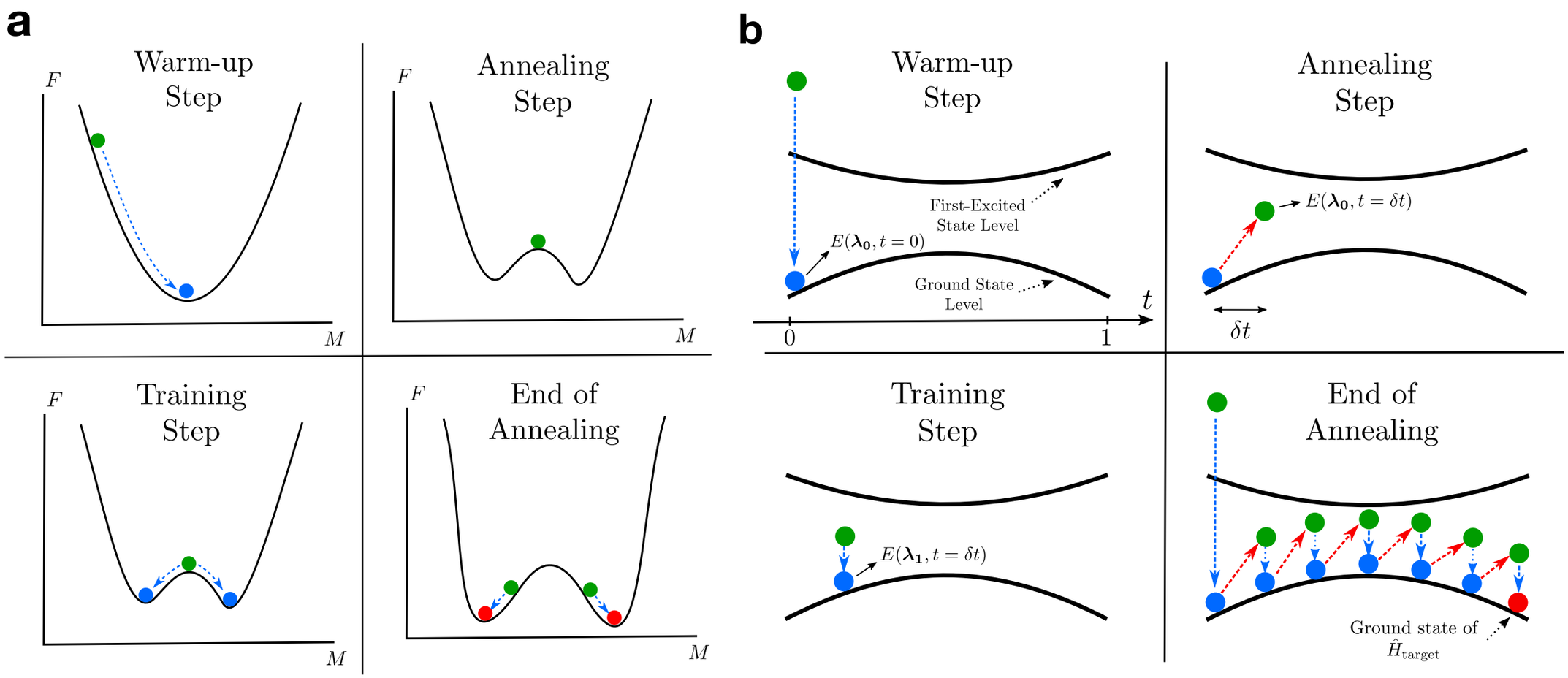


Figure 2. Variational neural annealing protocols. (a) The variational classical annealing (VCA) algorithm steps. A warm-up step brings the initialized variational state (green dot) close to the minimum of the free energy (cyan dot) at a given value of the order parameter M. This step is followed by an annealing and a training step that brings the variational state back to the new free energy minimum. Repeating the last two steps until $T(t=1)=0$ (red dots) produces approximate solutions to H_{target} if the protocol is conducted slowly enough. This schematic illustration corresponds to annealing through a continuous phase transition with an order parameter M. (b) Variational quantum annealing (VQA). VQA includes a warm-up step, followed by an annealing and a training step, which brings the variational energy (green dot) closer to the new a ground state energy (cyan dot). We loop over the previous two steps until reaching the target ground state of $\hat{H}_{\text{target}}$ (red dot) if annealing is performed slowly enough.

Ising chains

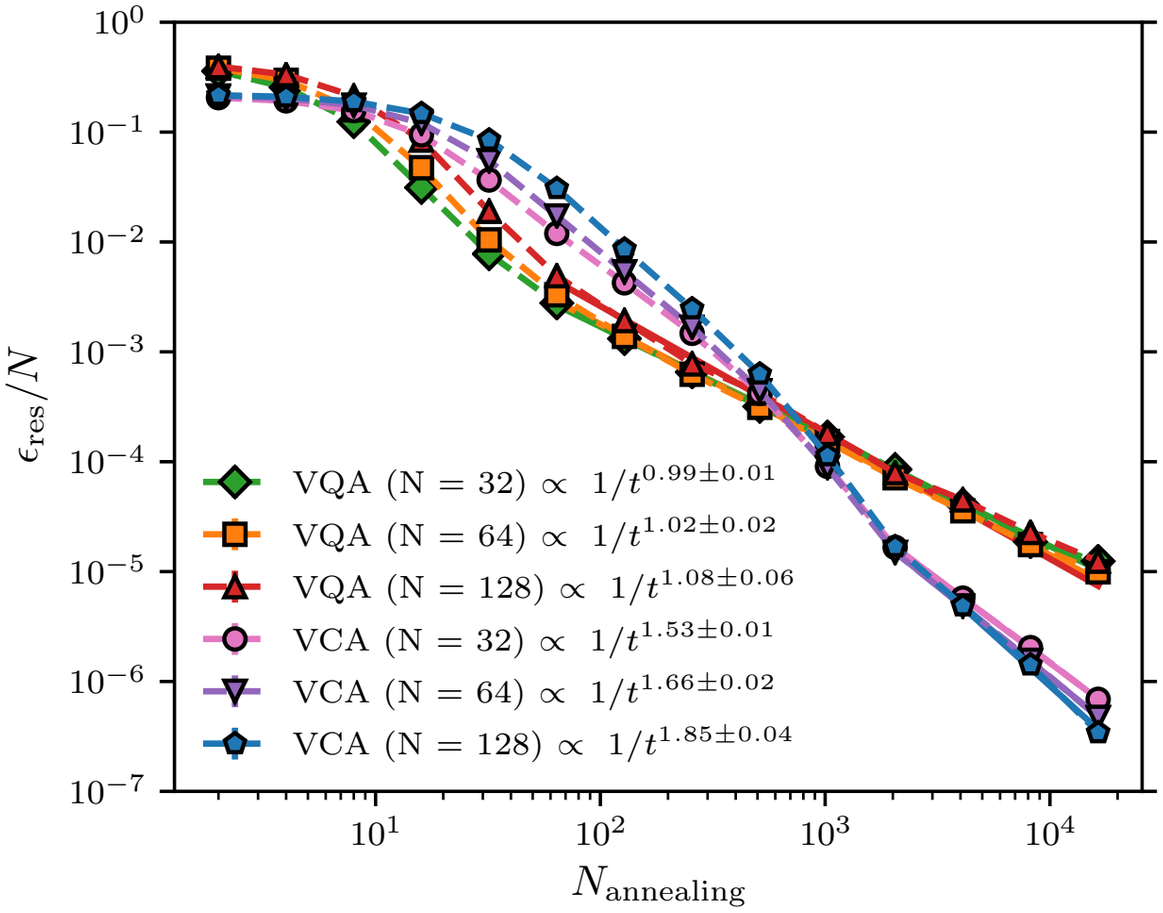


Figure 3. Variational neural annealing on a random Ising chain. Here we represent the residual energy per site $\epsilon_{\rm res}/N$ vs the number of annealing steps $N_{\rm annealing}$ for both VQA and VCA. The system sizes are $N = 32, 64, 128$. We use random positive couplings $J_{i,i+1} \in [0, 1)$ (see text for more details). The error bars represent the one s.d. statistical uncertainty calculated over different disorder realizations [28].

- $H_{\rm target} = -\sum_{i=1}^{N-1} J_{i,i+1}\sigma_i\sigma_{i+1}\,.$
- $J_{i,i+1} \in [0,1)$
- $\epsilon_{\rm res} = \left[\langle H_{\rm target}\rangle_{\rm av} - E_{\rm G}\right]_{\rm dis}$
- $\left[\ldots\right]_{\rm dis} = \exp(\langle \ln(\ldots)\rangle_{\rm av})$

Ising chains

- QA — $[\epsilon_{\text{res}}]^{\text{SA}} \sim \ln^{-2} \gamma_{SA} \tau$ * with $\gamma_{SA} = 6.5$
- SA — $[\epsilon_{\text{res}}]^{\text{QA-RT}} \sim \ln^{-\zeta} \gamma \tau$ * with $\gamma = 0.13, \zeta = 3.4$
- QA imaginary time — $1/\tau^{1.5-2.0}$ *
- VQA $1/\tau$ and VCA $1/\tau^{1.5-1.9}$
- Numerics suggest a speedup with respect to quantum annealing **and** simulated annealing, close to QA - IT

* Tommaso Zanca and Giuseppe E. Santoro, “Quantum annealing speedup over simulated annealing on random ising chains,” Phys. Rev. B 93, 224431 (2016).

* T. Caneva, R. Fazio, and G. E. Santoro, Phys. Rev. B 76, 144427 (2007).

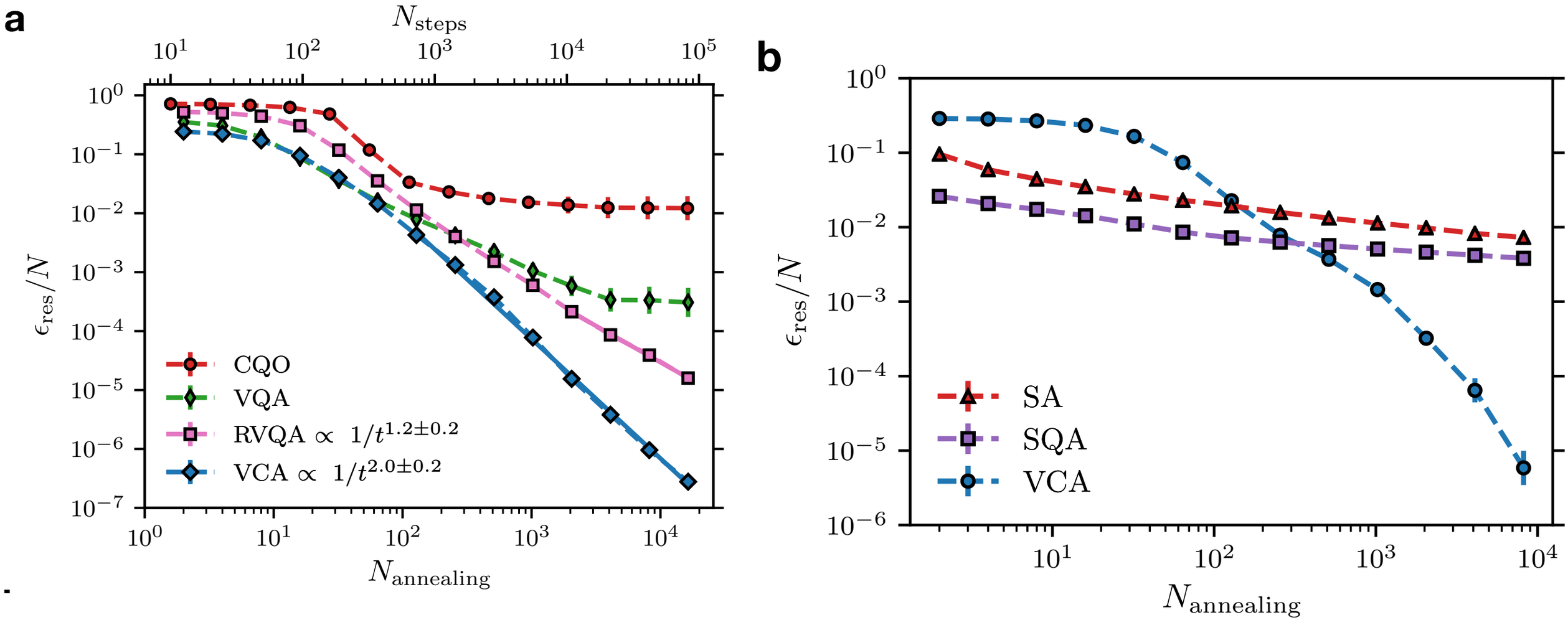
Edwards-Anderson Model

- $H_{\text{target}} = -\sum_{\langle i,j \rangle} J_{ij}\sigma_i\sigma_j$ in a 2-dimensional square lattice

- $J_{ij} \in [-1,1)$

- We test no annealing (CQO), VCA, VQA and an entropy regularized VQA: $\tilde{F}_{\boldsymbol{\theta}}(t) = \langle \hat{H}(t) \rangle_{\boldsymbol{\theta}} - T(t) S_{\text{pseudo}}(|\Psi_{\boldsymbol{\theta}}|^2)$ where $\hat{H}(t) = \hat{H}_{\text{target}} - \Gamma(t)\sum_i \hat{\sigma}_i^x$

Edwards-Anderson Model



Benchmarking the two-dimensional Edwards-Anderson spin glass. (a) A comparison between VCA, VQA, RVQA, and CQO on a 10×10 lattice by plotting the residual energy per site vs $N_{\text{annealing}}$. For CQO, we report the residual energy per site vs the number of optimization steps N_{steps}. (b) Comparison between SA, SQA with $P = 20$ trotter slices, and VCA on a 40×40 lattice. The annealing speed is the same for SA, SQA and VCA.

Variational annealing on fully connected spin glasses

- Sherrington- Kirkpatrick (SK) model provides a conceptual framework — role of disorder and frustration in widely diverse systems ranging from materials to combinatorial optimization and machine learning.

- $H_{\text{target}} = -\frac{1}{2}\sum_{i \neq j} \frac{J_{ij}}{\sqrt{N}} \sigma_i \sigma_j$

- J_{ij} is a symmetric matrix sampled from a gaussian distribution with mean 0 and variance 1.

Mohamed Hibat-Allah, Estelle M. Inack, Roeland Wiersema, Roger G. Melko, Juan Carrasquilla. Variational neural annealing. arXiv:2101.10154

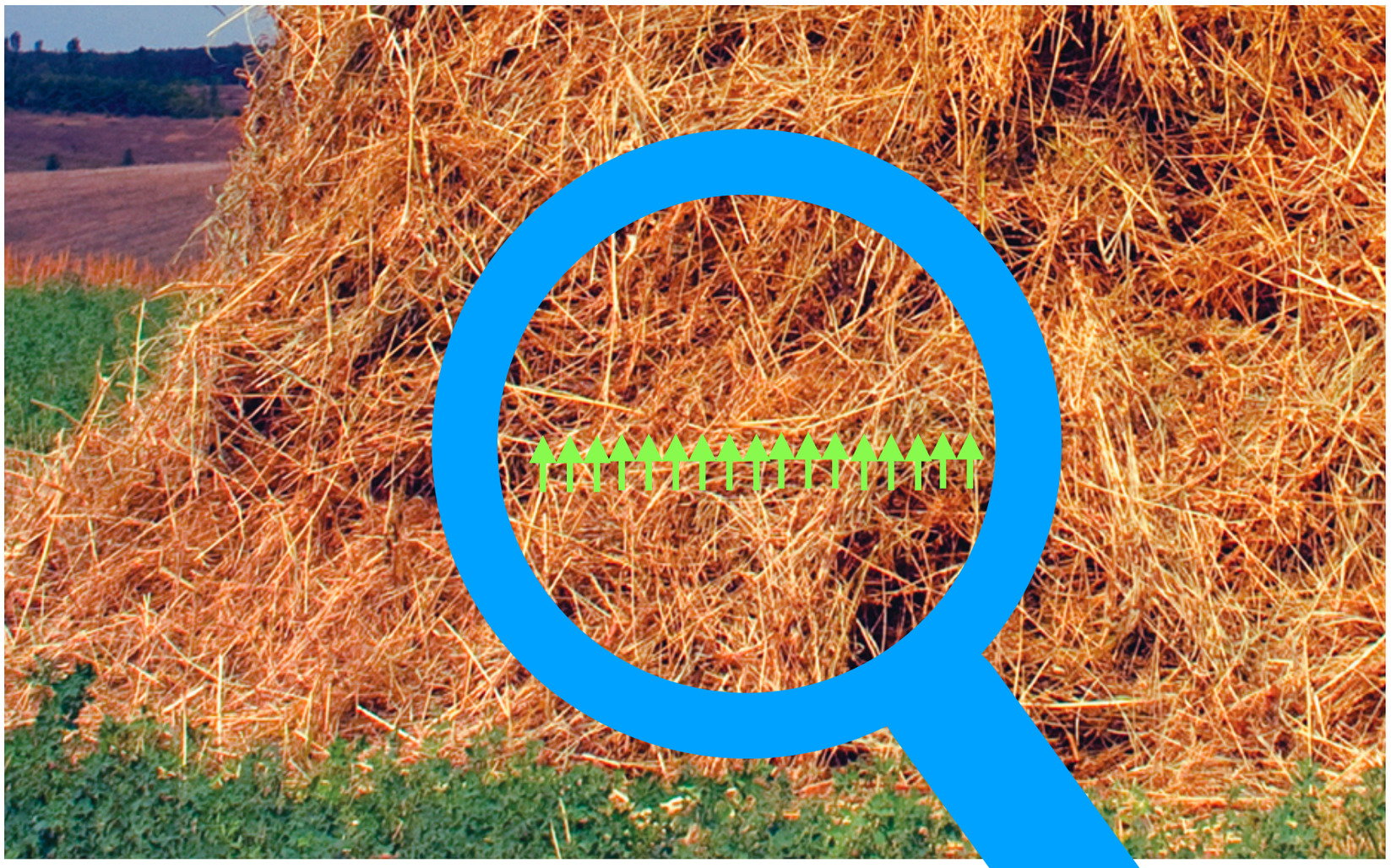
Variational annealing on fully connected spin glasses

- Wishart planted ensemble (WPE): fully connected model

Firas Hamze, Jack Raymond, Christopher A. Pattison, Katja Biswas, and Helmut G. Katzgraber, “Wishart planted ensemble: A tunably rugged pairwise ising model with a first-order phase transition,” Physical Review E 101 (2020)

Variational annealing on fully connected spin glasses

- Wishart planted ensemble (WPE)



Firas Hamze, Jack Raymond, Christopher A. Pattison, Katja Biswa… …elmut G. Katzgraber, “Wishart planted ensemble: A tunably rugged pairwise ising model with a first… …phase transition,” Physical Review E 101 (2020)

Variational annealing on fully connected spin glasses

- Wishart planted ensemble (WPE), which is a class of zero-field Ising models with a first-order phase transition and tunable algorithmic hardness.

- $H_{\text{target}} = -\frac{1}{2}\sum_{i \neq j} J_{ij}^{\alpha}\sigma_i\sigma_j$

- J_{ij}^{α} is a symmetric matrix. $J^{\alpha} = \tilde{J}^{\alpha} - \text{diag}(\tilde{J})$

- $\tilde{J}^{\alpha} = -\frac{1}{N}W_{\alpha}W_{\alpha}^{\mathsf{T}}$ (Wishart)

- W_{α} is an $N \times \lfloor \alpha N \rfloor$ random matrix satisfying $W_{\alpha}t_{\text{ferro}} = 0$ where $t_{\text{ferro}} = (+1, +1, \ldots, +1)$ is the ferromagnetic state

Firas Hamze, Jack Raymond, Christopher A. Pattison, Katja Biswas, and Helmut G. Katzgraber, “Wishart planted ensemble: A tunably rugged pairwise ising model with a first-order phase transition,” Physical Review E 101 (2020)

Variational annealing on fully connected spin glasses

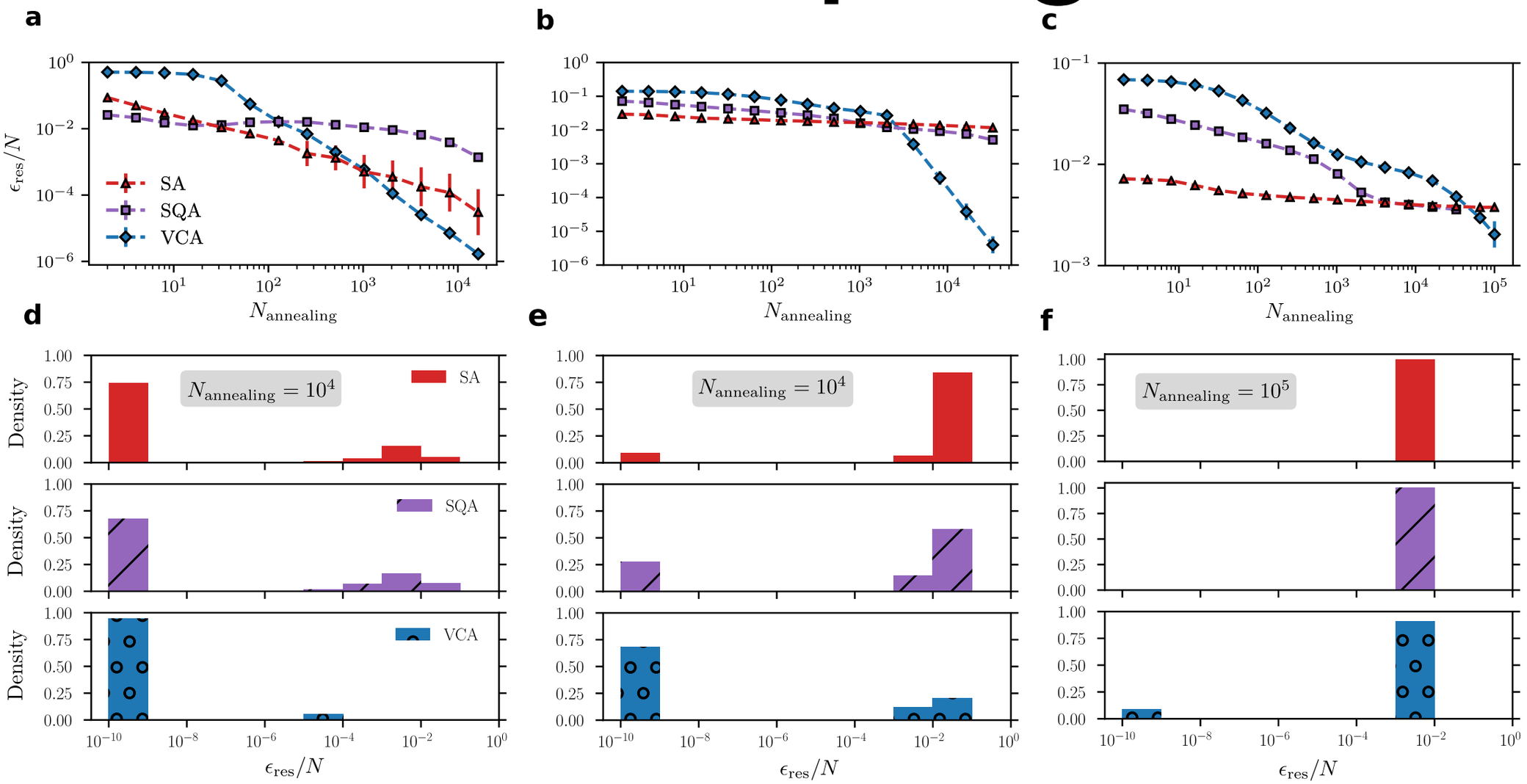


Figure 5. Benchmarking SA, SQA ($P = 100$ trotter slices) and VCA on the Sherrington-Kirkpatrick (SK) model and the Wishart planted ensemble (WPE). Panels (a),(b), and (c) display the residual energy per site as a function of $N_{\rm annealing}$. (a) The SK model with $N = 100$ spins. (b) WPE with $N = 32$ spins and $\alpha = 0.5$. (c) WPE with $N = 32$ spins and $\alpha = 0.25$. Panels (d), (e) and (f) display the residual energy histogram for each of the different techniques and models in panels (a),(b), and (c), respectively. The histograms use 25000 data points for each method. Note that we choose a minimum threshold of 10^{-10} for $\epsilon_{\rm res}/N$, which is within our numerical accuracy.

Mohamed Hibat-Allah, Estelle M. Inack, Roeland Wiersema, Roger G. Melko, Juan Carrasquilla. Variational neural annealing. arXiv:2101.10154

NEURAL AUTOREGRESSIVE MODELS FOR MANY-BODY PHYSICS

- We have explored quantum state reconstruction with RNNs (Nature Machine Intelligence, vol. 1, 155-161 (2019)) and Transformers (arXiv:2006.12469)
- Simulation of quantum circuits with transformers (arXiv:1912.11052)
- Recurrent neural network wavefunctions — extremely accurate ground states, very compact representation in 1d and 2d (Phys. Rev. Research 2, 023358 (2020))
- Variational neural annealing: produces very accurate solutions to challenging spin glass problems beyond SA and SQA (arXiv:2101.10154)
- Neural Error Mitigation of Near-Term Quantum Simulations (arXiv:2105.08086) (transformer)
- Simulation of open system dynamics (arXiv:2009.05580) (transformer)
- Transfer learning based on physical principles (arXiv:2003.02647) (RNN)