Title: Recurrent neural networks for many-body physics

Speakers: Juan Felipe Carrasquilla Ã•lvarez

Collection: Quantum Matter Workshop

Date: November 14, 2022 - 11:30 AM

URL: https://pirsa.org/22110076

Abstract: I will discuss our recent work on the use of autoregressive neural networks for many-body physics. In particular, I will discuss two approaches to represent quantum states using these models and their applications to the reconstruction of quantum states, the simulation of real-time dynamics of open quantum systems, and the approximation of ground states of many-body systems displaying long-range order, frustration, and topological order. Finally, I will discuss how annealing in these systems can be used for combinatorial optimization where we observe solutions to problems that are orders of magnitude more accurate than simulated and simulated quantum annealing.

Recurrent neural networks for many-body physics

Juan Felipe Carrasquilla Álvarez
Vector Institute
@carrasqu

Perimeter Institute, quantum matter workshop
Nov. 14th, 2022

VECTOR INSTITUTE | INSTITUT VECTEUR · UNIVERSITY OF WATERLOO · UNIVERSITY OF TORONTO · CIFAR · NSERC CRSNG · compute canada | calcul canada

ML and physics

- Researchers in particle physics, high energy physics, condensed matter, quantum information, statistical physics, and atomic, molecular, and optical physics communities are exploring ideas related to machine learning the its implications in these fields.

Machine learning and the physical sciences. Giuseppe Carleo, Ignacio Cirac, Kyle Cranmer, Laurent Daudet, Maria Schuld, Naftali Tishby, Leslie Vogt-Maranto, and Lenka Zdeborová

Rev. Mod. Phys. **91**, 045002 (2019)

ML and quantum science research

- Machine learning in simulations of strongly correlated fermions (Chemistry, DFT, QMC)
- Machine learning phases of matter in simulations and experimental data
- Discovery of physics phenomena and laws from data
- Neural-network representations of quantum states
- Renormalization group
- Applications in quantum information, quantum control, quantum error correction, and quantum computation and more ...

Machine learning for quantum matter. **Juan Carrasquilla**. Advances in Physics: X, 5:1 (2020)

This style of research is shaped in part by the **commonalities** among of the problems that the ML and physics disciplines address.

Dimensionality of quantum many-body systems / machine learning models

- Generic specification of a quantum state requires resources exponentially large in the number of degrees of freedom N

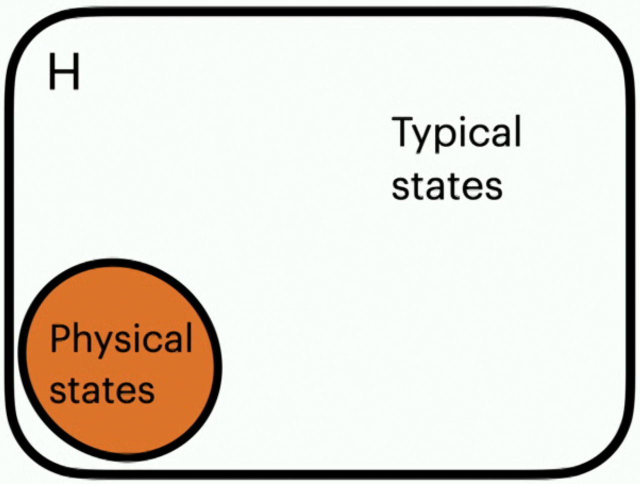
  $|\Psi\rangle$ vector with 2^N
- Today's best supercomputers can solve the wave equation **exactly** for systems with a maximum of ~45 spins.
- $2^N \sim 3.5 \times 10^{13}$

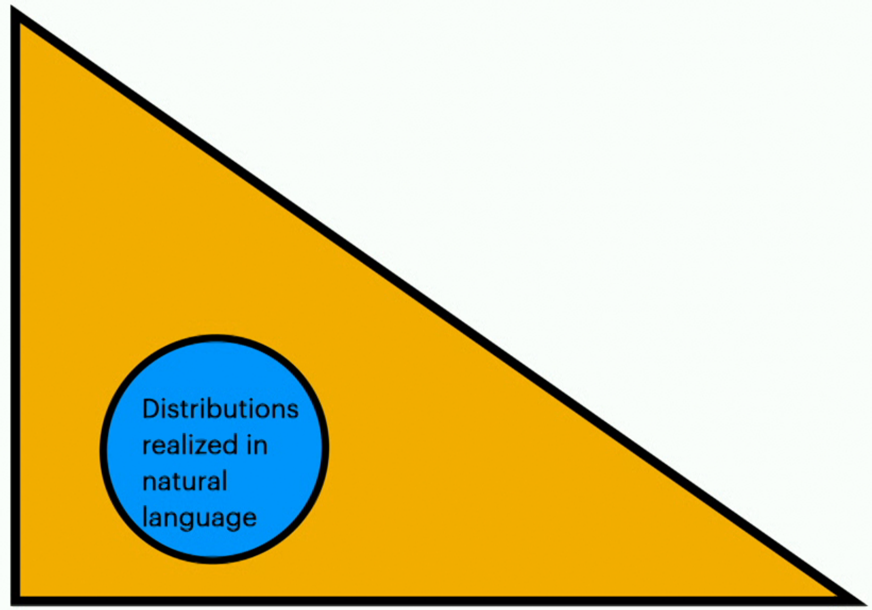
- Language models live in very high dimensional spaces too (example from “Attention is all you need”, 2017)

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

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

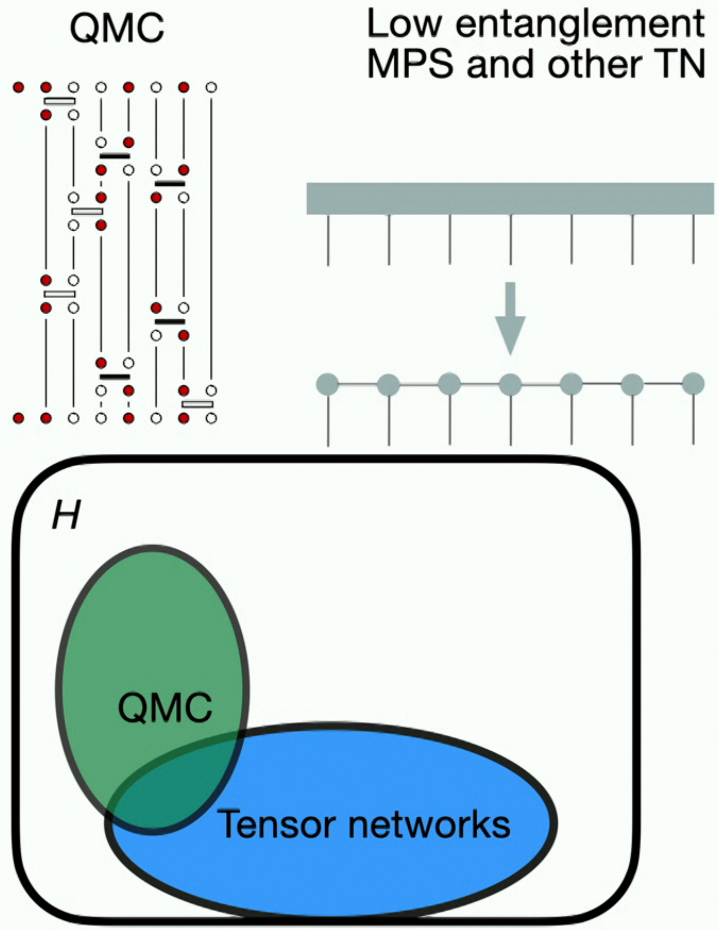
Many quantum states realized in nature live in a lower dimensional subspace. Similarly, most paragraphs are noise — probability distributions over the sentences our brain understands live in low-dimensional subspace.





WE EXPLOIT THESE STRUCTURES IN COMPUTATIONAL PHYSICS

- Amount of information in states seen in nature is usually 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: Exploit the fact that some important quantum states realized in nature have little entanglement
- Both techniques have led to profound implications to our understanding of quantum systems.

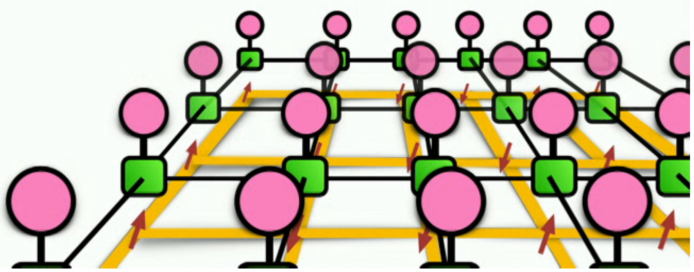


ENTER MACHINE LEARNING

- Similarly, ML techniques exploit the structure present in natural datasets e.g. images, natural language.
- Commonalities with physical systems include correlation functions and symmetries (translations, rotations, reflections, and more).
- Can we exploit the insight that both quantum and ML problems have some shared structure?
- Are these commonalities relevant/important for computational physics beyond mere resemblance?
- We have been exploring these questions for the past few years.

Central ingredient behind these two examples: Representations of classical and quantum states using recurrent neural networks

- **Carrasquilla**, Torlai, Melko, Aolita. Nature Machine Intelligence 1, 200 (2019)
- Dian Wu, Lei Wang, Pan Zhang. Solving Statistical Mechanics Using Variational Autoregressive Networks. Phys. Rev. Lett. 122, 080602 (2019)
- Hibat-Allah, Ganahl, Hayward, Melko, **Carrasquilla**. Phys. Rev. Research **2**, 023358 (2020) arXiv:2002.02973
- Hibat-Allah, Inack, Wiersema, Melko, **Carrasquilla**. Variational neural annealing. Nature Machine Intelligence volume **3**, 952–961 (2021) arXiv:2101.10154
- And more…



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

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

- This idea can be extended to quantum states.

Autoregressive models and their tractability advantages

✓ Computing $P(\boldsymbol{\sigma})$ or the amplitude $\Psi(\boldsymbol{\sigma})$ 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

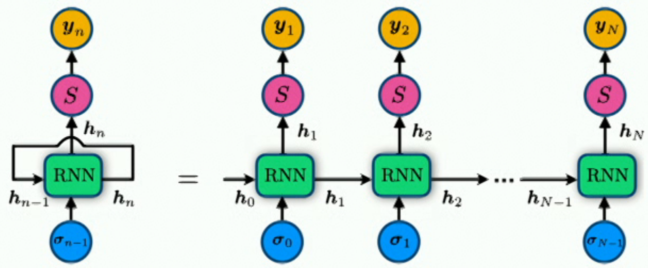
✓ Easy to encode product states and mean-field theories (e.g. Gutzwiller)

✓ We can impose some symmetries and conservation laws

✓ Advantages remain true for models quantum states $\Psi(\boldsymbol{\sigma})$ in any dimension.

RECURRENT NEURAL NETWORKS (RNN)

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

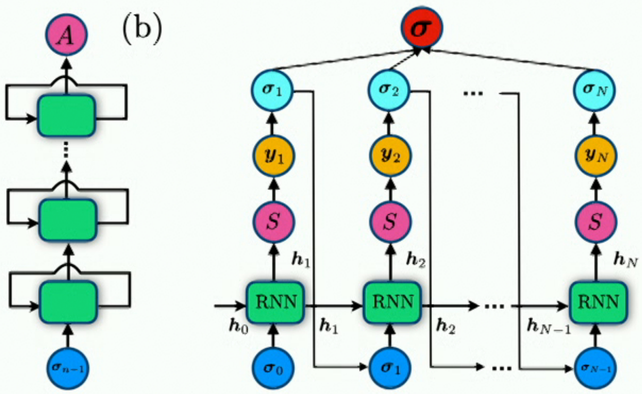


$$\boldsymbol{h}_n = f\left(W[\boldsymbol{h}_{n-1};\boldsymbol{\sigma}_{n-1}] + \boldsymbol{b}\right)$$

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

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

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



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

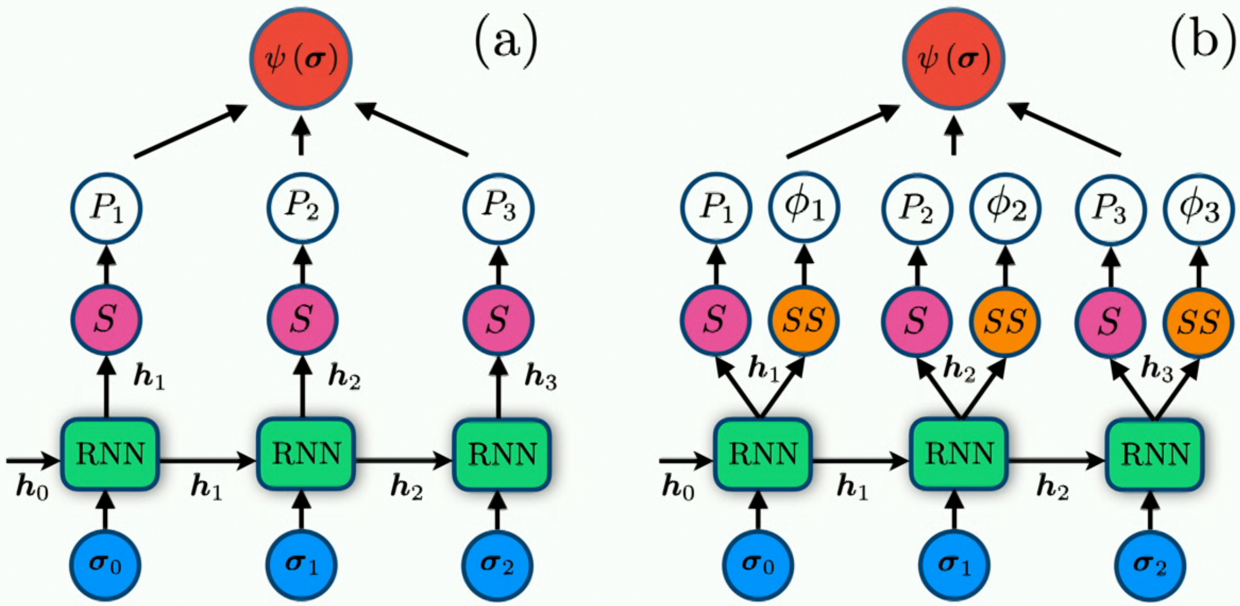
Sampling:

- 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)

RNNs wave functions



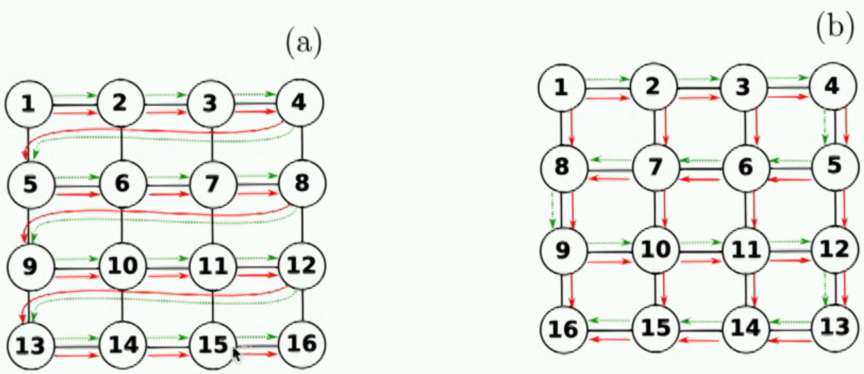
$$|\Psi\rangle = \sum_{\boldsymbol{\sigma}} \psi(\boldsymbol{\sigma})\,|\boldsymbol{\sigma}\rangle = \sum_{\boldsymbol{\sigma}} \sqrt{P(\boldsymbol{\sigma})}\,|\boldsymbol{\sigma}\rangle\,. \qquad |\Psi\rangle = \sum_{\boldsymbol{\sigma}} e^{\mathrm{i}\phi(\boldsymbol{\sigma})}\sqrt{P(\boldsymbol{\sigma})}\,|\boldsymbol{\sigma}\rangle\,.$$

Mohamed Hibat-Allah, Martin Ganahl, Lauren E. Hayward, Roger G. Melko, Juan Carrasquilla
Recurrent Neural Network Wave Functions. Phys. Rev. Research 2, 023358 (2020)

Two dimensional RNNs

- Replace the single hidden state that is passed from one site to another by two hidden states, with each one corresponding to the state of a neighbouring site (vertical and horizontal) — 2D geometry of the problem

- $$\boldsymbol{h}_{i,j} = f\left(W^{(h)}[\boldsymbol{h}_{i-1,j}; \boldsymbol{\sigma}_{i-1,j}] + W^{(v)}[\boldsymbol{h}_{i,j-1}; \boldsymbol{\sigma}_{i,j-1}] + \boldsymbol{b}\right)$$

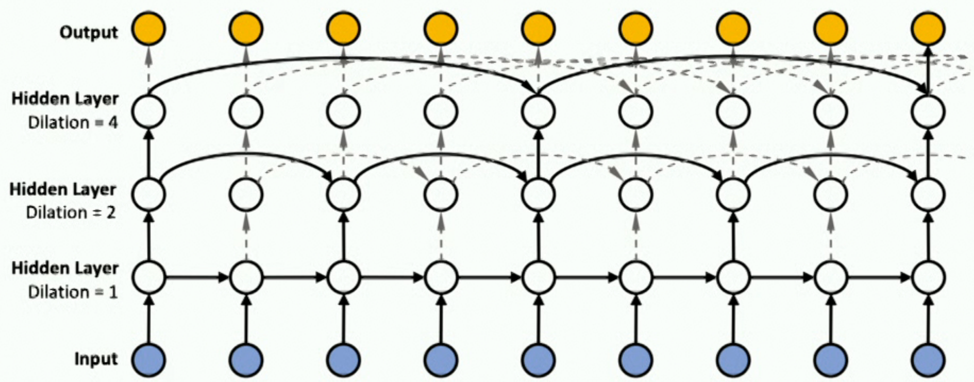


No tractability issues, can be defined in any spatial dimension

Dilated RNNs

- To account for slowly decaying correlations, one can use dilated RNNs. Different layers capture correlations at different scales. Similar in spirit to MERA

$$\boldsymbol{h}_n^{(l)} = F(W_n^{(l)}[\boldsymbol{h}^{(l)}_{\max(0,n-2^{l-1})}; \boldsymbol{h}_n^{(l-1)}] + \boldsymbol{b}_n^{(l)})$$



Shiyu Chang, Yang Zhang, Wei Han, et al. Dilated Recurrent Neural Networks. arXiv:1710.02224

Mohamed Hibat-Allah, Estelle M. Inack, Roeland Wiersema, Roger G. Melko & Juan Carrasquilla *Nature Machine Intelligence* (2021)

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

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

nature.com > nature > letters > article

nature
International journal of science

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

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

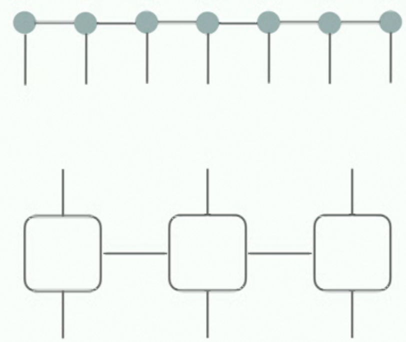
Phys. Rev. X **8**, 031022 – Published 24 July 2018

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

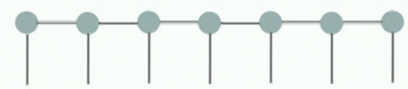
[2] **A scalable maximum likelihood method for quantum state tomography** T Baumgratz1, A Nüßeler, M Cramer and M B Plenio. New Journal of Physics, Volume 15, December 2013



HOW TO MAKE QST EFFICIENT?

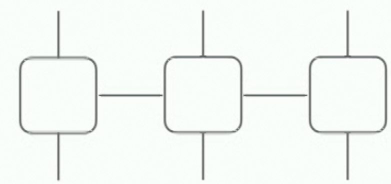
➤ Introduce a parametrization of the quantum state with good scaling if non-trivial structural information on the quantum systems under consideration is utilized: MPS[1] and MPO[2] tomography

[1] **Efficient quantum state tomography**. Marcus Cramer, Martin B. Plenio, Steven T. Flammia, Rolando Somma, David Gross, Stephen D.



Works very well, best suited to 1D systems

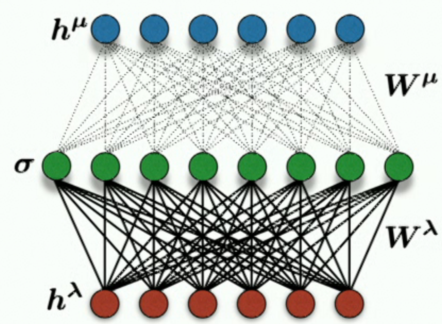
[2] **A scalable maximum likelihood method for quantum state tomography** T Baumgratz1, A Nüßeler, M Cramer and M B Plenio. New Journal of Physics, Volume 15, December 2013



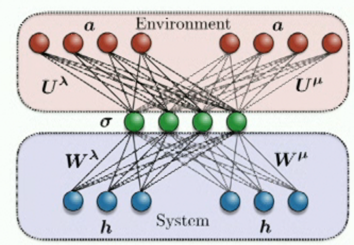
HOW TO MAKE QST EFFICIENT?

➤ 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?

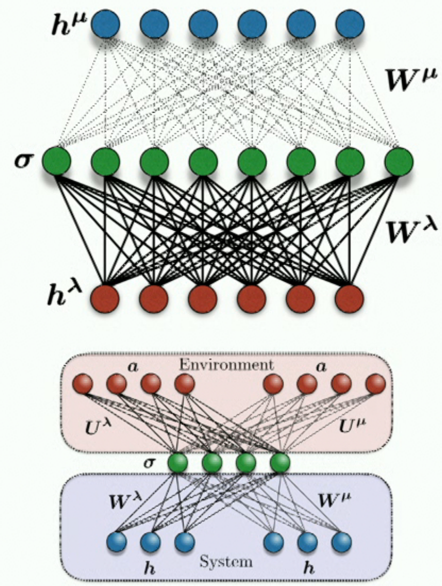
- Introduce a parametrization of the quantum state with good scaling if non-trivial structural information on the quantum systems under consideration is utilized: Restricted Boltzmann machines both for pure[3] states and mixed states[4]

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

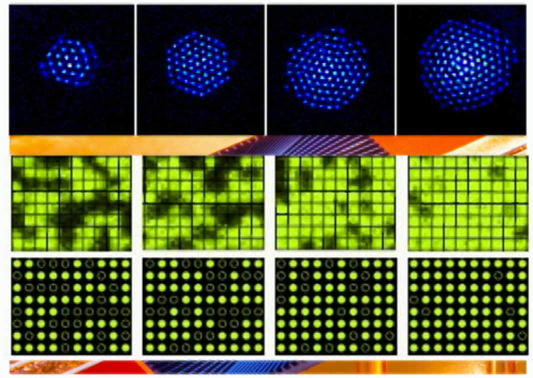
- ~~Parametrize the quantum state~~ We parametrize the measurement statistics of an informationally complete positive operator valued measure (POVM). We use recurrent neural networks and transformers.
- Use this idea to learn states from synthetic measurements mimicking experimental data as well as data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

Mixed states with structure, good scaling

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

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



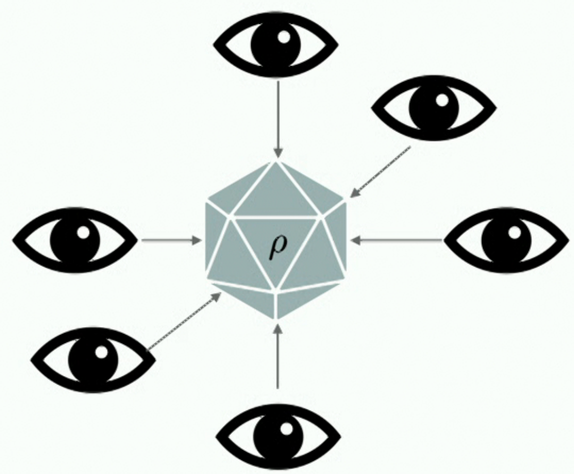
$$\text{dataset} = \begin{bmatrix} 0,0,2,\ldots 3,0 \\ 1,1,0,\ldots 0,2 \\ 0,2,1,\ldots 0,3 \\ 1,1,3,\ldots 3,0 \\ \vdots \\ 3,2,0,\ldots 1,0 \end{bmatrix} \rightarrow \rho^* \rightarrow F(\rho,\rho^*) = \mathrm{Tr}\left[\sqrt{\sqrt{\rho}\rho^*\sqrt{\rho}}\right]^2$$

MEASUREMENTS: POSITIVE OPERATOR-VALUED MEASURE (POVM)

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

INFORMATIONALLY COMPLETE MEASUREMENTS

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



$$\boldsymbol{M} = \left\{M^{(a_1)} \otimes M^{(a_2)} \otimes \ldots M^{(a_N)}\right\}_{a_1, \ldots a_N}$$

INVERTING BORN RULE

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

INFORMATIONALLY COMPLETENESS —>THIS RELATION CAN BE INVERTED

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

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

- 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(a)
- Tomography
- Real-time dynamics of close and open systems
- Measurements

INVERTING BORN RULE

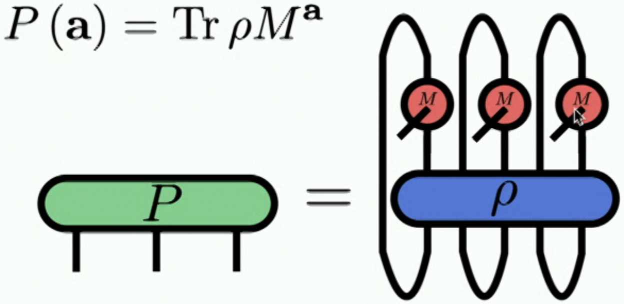
BORN RULE $P(\mathbf{a}) = \mathrm{Tr}\,\rho M^{\mathbf{a}}$

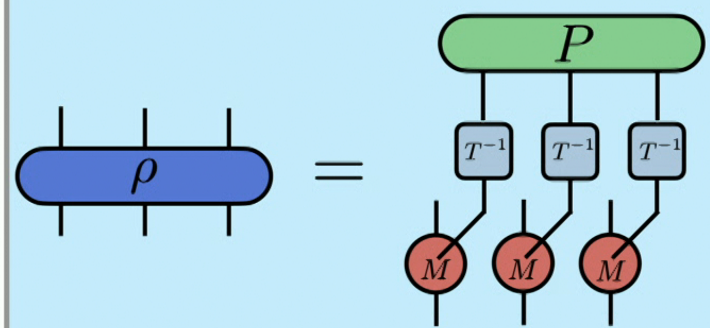
INFORMATIONALLY COMPLETENESS —>THIS RELATION CAN BE INVERTED

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

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

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





Insight: parametrize statistics of measurements and invert

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

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

$$\rho_{\mathrm{model}} = \sum_{a,a'} T^{-1}_{a,a'}\, P_{\mathrm{model}}(a')\, M^{(a)}$$

Create an autoregressive model of P(a)

Autoregressive models (RNNs and transformer)

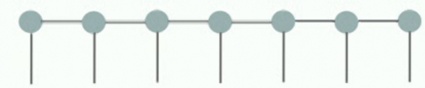
1. Allow for exact sampling
2. Tractable density $P_{\mathrm{model}}(\mathbf{a})$
3. Use maximum likelihood estimation to infer the parameters of the model so that the distribution explains the dataset of measurements.

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

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

RESULTS ON SYNTHETIC DATASETS FOR GHZ STATES

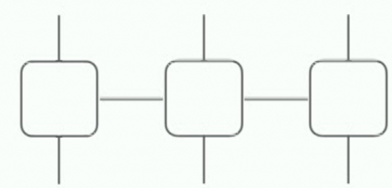
GHZ STATE $\quad |\Psi_0\rangle \equiv \alpha\,|0\rangle^{\otimes N} + \beta\,|1\rangle^{\otimes N}$



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

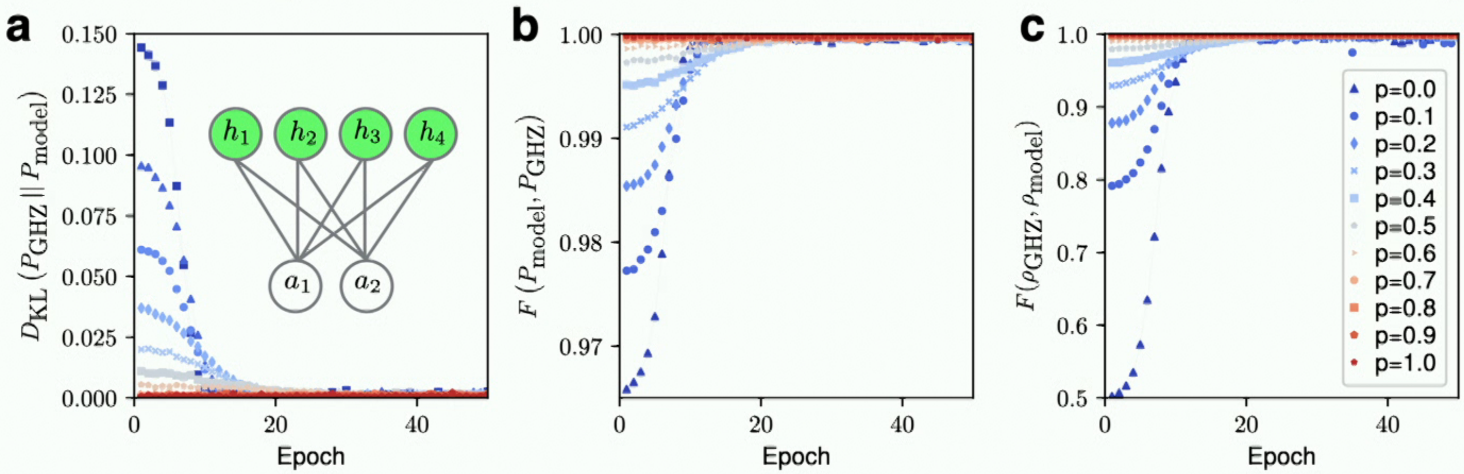
GHZ with local depolarization: a model of a decohering qubit where with probability 1 – p each qubit remains intact, while with probability p an “error” occurs.

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



Learning 2 qubit locally depolarized GHZ

$$KL\left(P_{\text{model}}|P\right) = -\sum_a P(a)\log\frac{P_{model}(a)}{P(a)} \qquad F_{\text{Classical}} = \sum_a \sqrt{P(a)P_{model}(a)} \qquad F\left(\rho,\sigma\right) = \text{Tr}\left[\sqrt{\sqrt{\rho}\sigma\sqrt{\rho}}\right]^2$$



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.

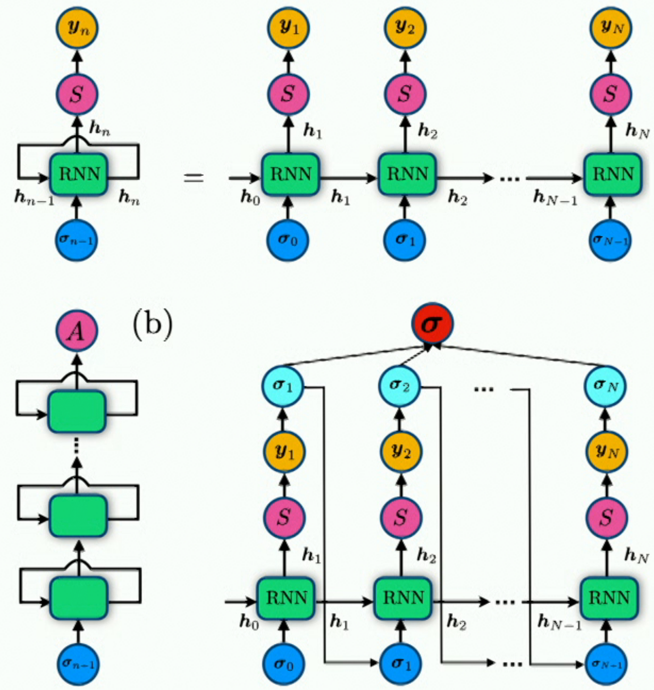
Carrasquilla, Torlai, Melko, Aolita. Nature Machine Intelligence 1, 200 (2019)

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)

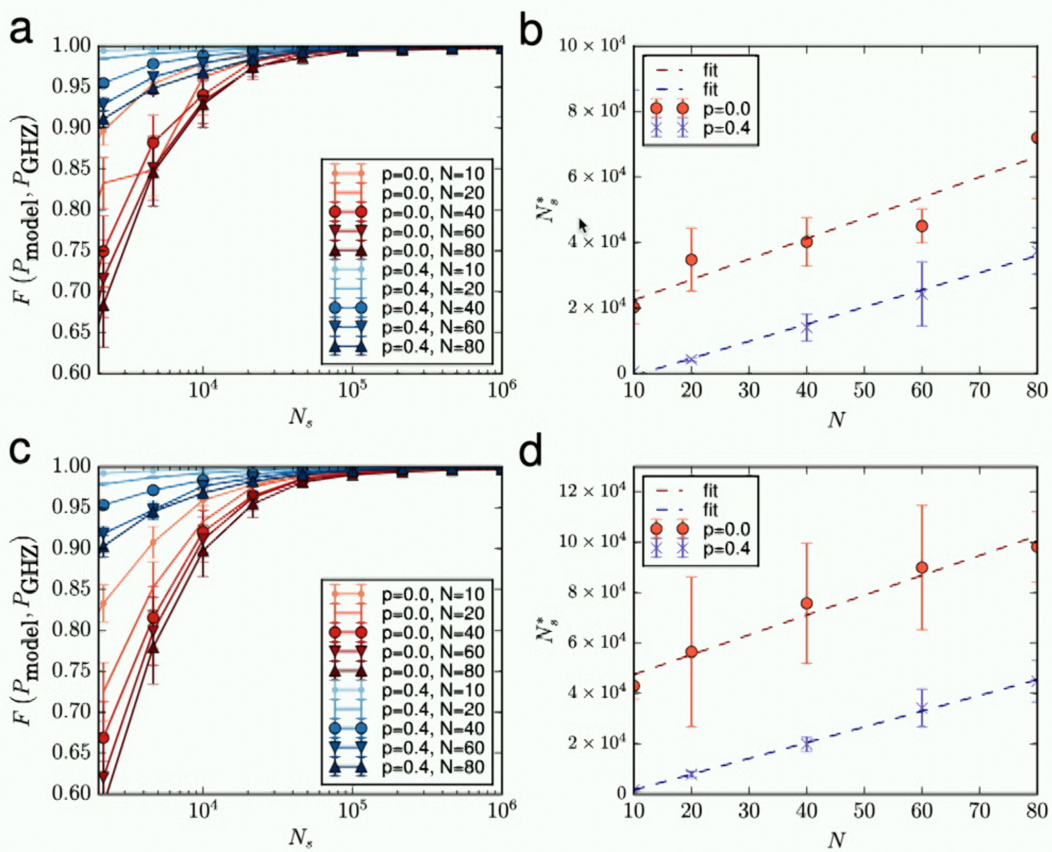
- System sizes beyond the reach of traditional quantum state tomography.
- Numerically: GHZ states require **linear** number of samples to reach fixed classical fidelity



- 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

Yoav Levine, Or Sharir, Nadav Cohen, and Amnon Shashua
Phys. Rev. Lett. **122**, 065301 – Published 12 February 2019

Numerical investigation of the sample complexity of learning



Tetrahedral POVM

Pauli measurement

Favorable scaling. Carrasquilla, Torlai, Melko, Aolita. Nature Machine Intelligence 1, 200 (2019) arXiv: 1810.10584

EXPERIMENTAL DEMONSTRATION

GHZ state with 3 qubits

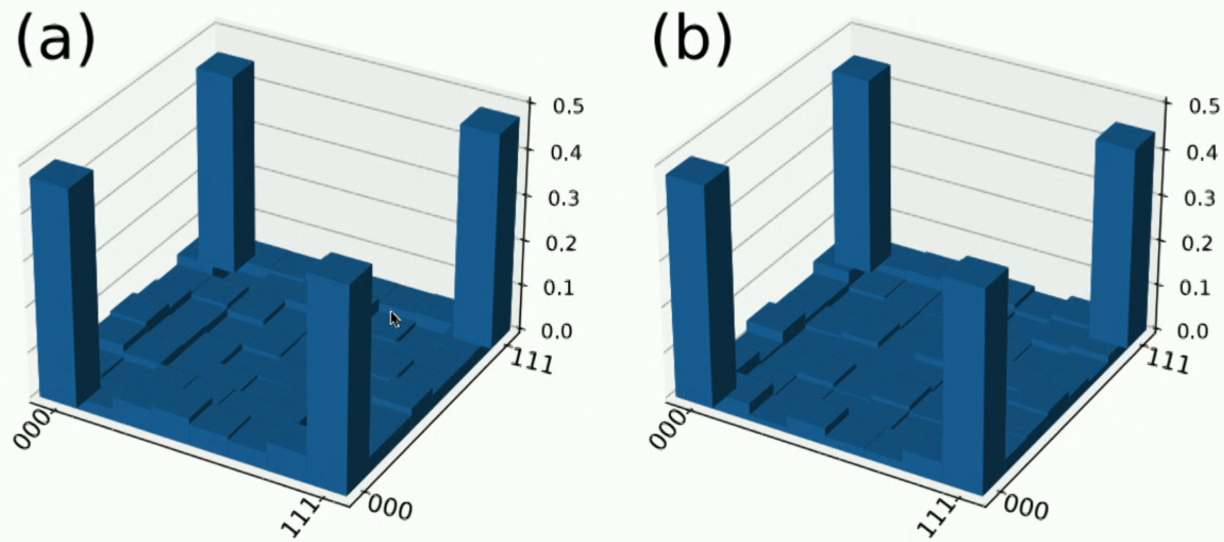
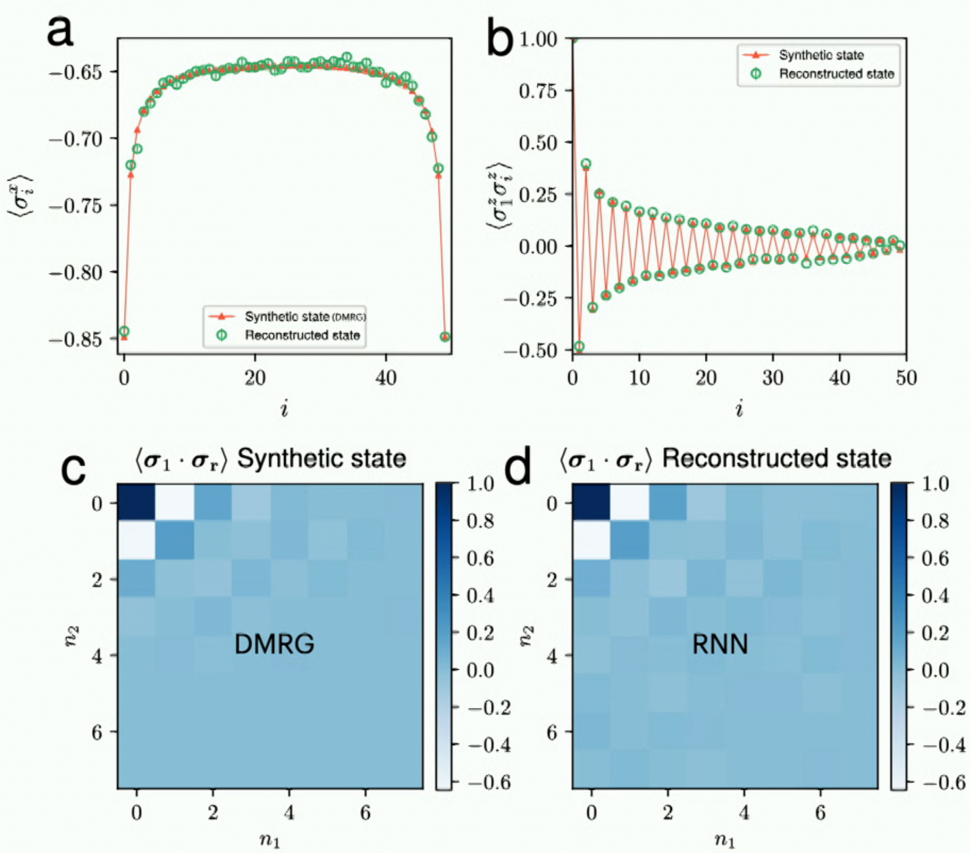


Figure 3. Benchmarking AQT (a) to MLE tomography offered by IBM's Qiskit library (b) for a noisy 3-qubit GHZ state data generated on the IBMQ_OURENSE quantum computer. Each bar represents the absolute value of a density matrix (DM) element.

Peter Cha, Paul Ginsparg, Felix Wu, Juan Carrasquilla, Peter L. McMahon, Eun-Ah Kim. Mach. Learn.: Sci. Technol. 3 01LT01 (2022)

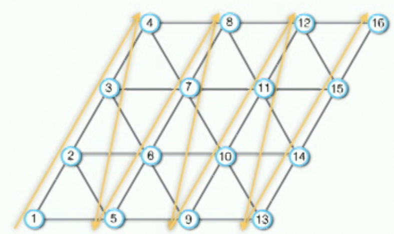
Learning Ground states of local hamiltonians from data



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

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

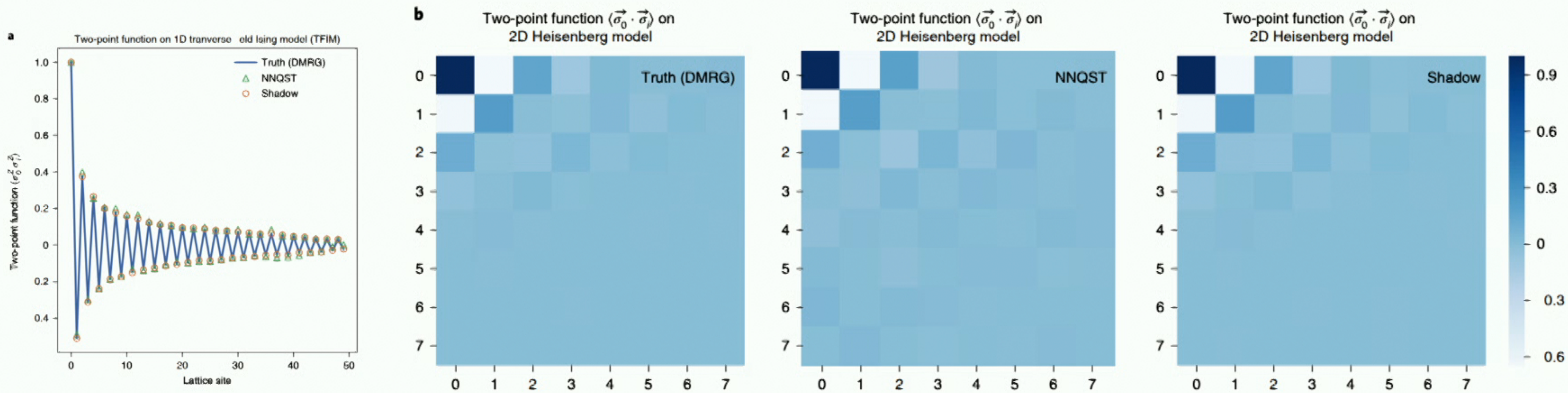
$$H = J\sum_{i,j}\boldsymbol{\sigma}_i.\boldsymbol{\sigma}_j$$



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

Connection to classical shadows

- The data sets we collected to train our model were eventually theoretically analyzed in detail — classical shadows.



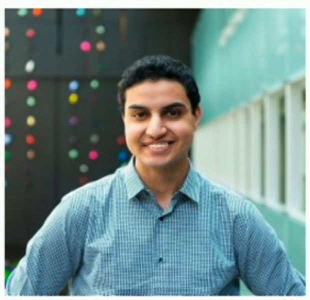
- "many properties of a quantum state can be accurately and efficiently predicted with a rigorous performance guarantee" In part explains the success of our approach. We are currently studying what generalization in ML model implications.

Hsin-Yuan Huang, Richard Kueng & John Preskill .Nature Physics. 16, 1050–1057 (2020)

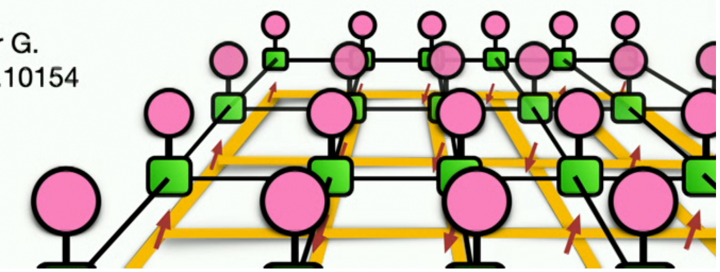
Extensions of these ideas

- 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 arXiv preprint arXiv:2006.02424
- 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)

Example 2: Variational neural annealing



Mohamed Hibat-Allah, Estelle M. Inack, Roeland Wiersema, Roger G. Melko, Juan Carrasquilla. Variational neural annealing. arXiv:2101.10154 Nature Machine Intelligence volume **3**, 952–961 (2021)



Annealing and simulated annealing

- Origin of this observation: many of these computationally hard but technologically relevant problems can be formulated as finding the ground state of an Ising Hamiltonian*: $H_{\text{target}} = -\sum_{i<j} J_{ij}\,\sigma_i\sigma_j - \sum_{i=1}^{N} h_i\sigma_i$

- Already in 1983: design of computer chips—component placement, wiring.

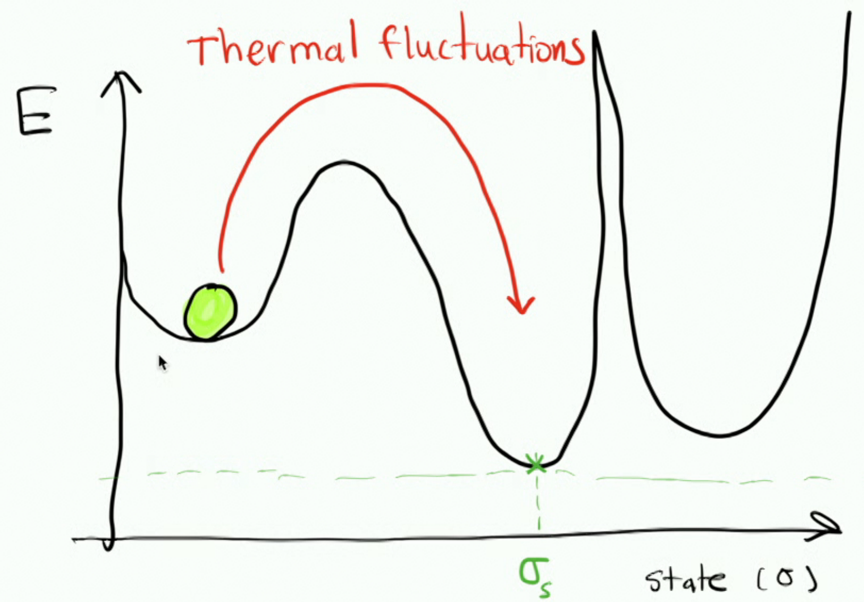
- A wide array of problems has been mapped to Ising models since.

*Andrew Lucas, "Ising formulations of many np problems," Front. Phys. 2, 5 (2014).

Combinatorial optimization by simulated annealing

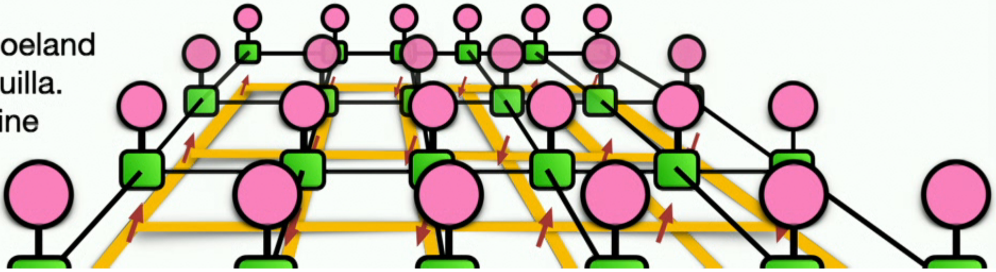
- The SA algorithm explores an optimization problem's energy landscape via a gradual decrease in thermal fluctuations generated by the Metropolis-Hastings algorithm.

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



CAN WE SIMULATE THESE OPTIMIZATION TECHNIQUES VARIATIONALLY?

Mohamed Hibat-Allah, Estelle M. Inack, Roeland Wiersema, Roger G. Melko, Juan Carrasquilla. Variational neural annealing. Nature Machine Intelligence (2021)



Variational classical annealing

- Target the **Boltzmann distribution** with a ML model $P_{\boldsymbol{\theta}}(\boldsymbol{\sigma})$.

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

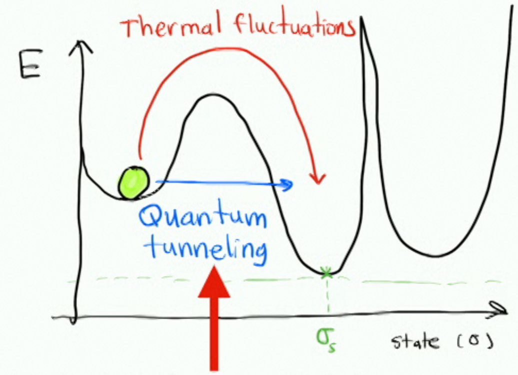
- $S(P_{\boldsymbol{\theta}})$ is the entropy of the model $P_{\boldsymbol{\theta}}(\boldsymbol{\sigma})$

- Requires samples $\boldsymbol{\sigma}^{(i)} \sim P_{\boldsymbol{\theta}}$ and fast access to $P_{\boldsymbol{\theta}}(\boldsymbol{\sigma})$ $-$ important to use autoregressive models as these models meet the two requirements.

- 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 $t \in [0,1]$, as the models are trained.

- Dian Wu, Lei Wang, Pan Zhang. Solving Statistical Mechanics Using Variational Autoregressive Networks. Phys. Rev. Lett. 122, 080602 (2019)
- Mohamed Hibat-Allah, Estelle M. Inack, Roeland Wiersema, Roger G. Melko, Juan Carrasquilla. Variational neural annealing. arXiv:2101.10154 Nature Machine Intelligence volume 3, 952–961 (2021)

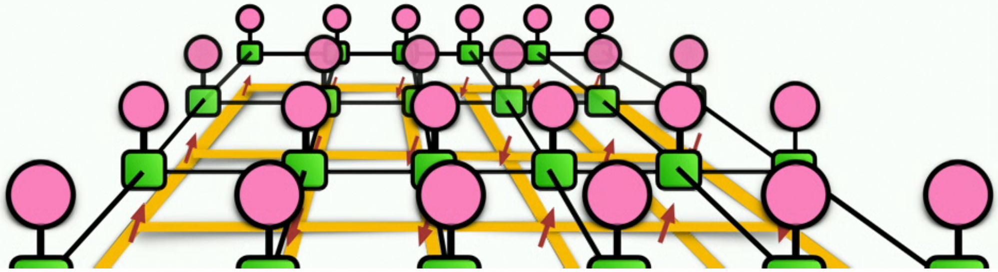
Variational quantum annealing



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

RESULTS

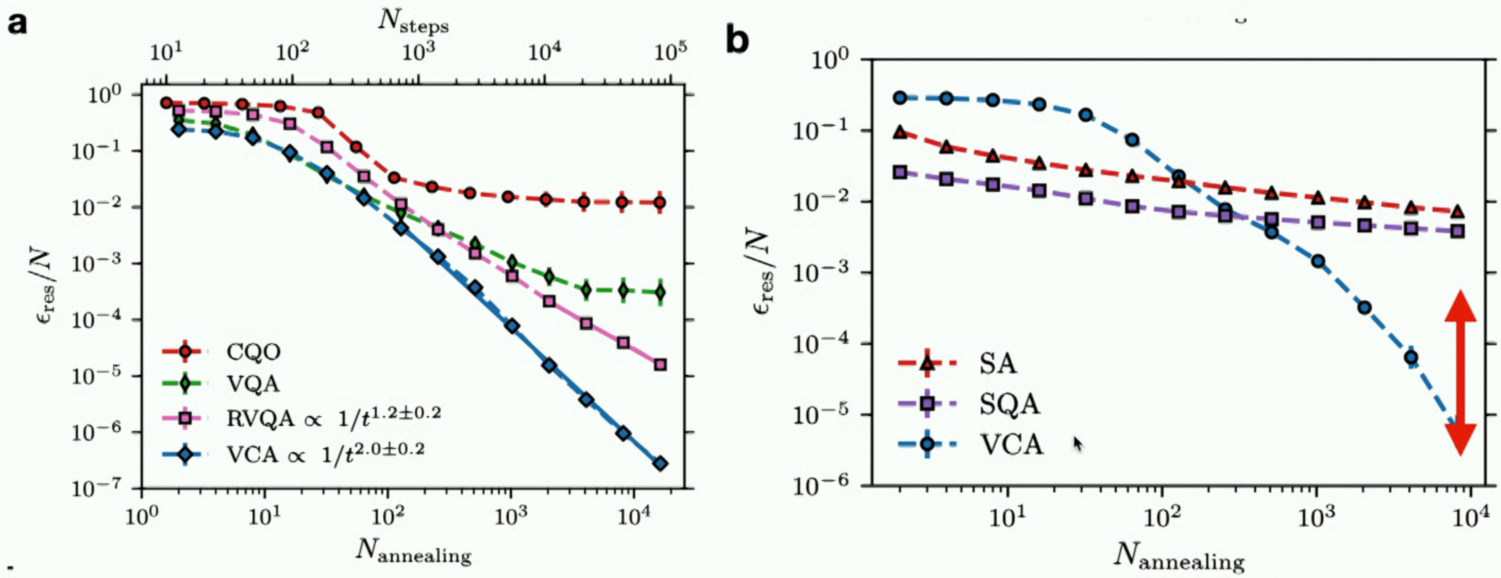
Mohamed Hibat-Allah, Estelle M. Inack, Roeland Wiersema, Roger G. Melko, Juan Carrasquilla. Variational neural annealing. Nature Machine Intelligence (2021)



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 multiple algorithms: no annealing (CQO), Variational classical annealing, Variational Quantum Annealing 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 $N_{annealing}$. 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.

Theory of Quantum Annealing of an Ising Spin Glass
GIUSEPPE E. SANTORO, ROMAN MARTOŇÁK, ERIO TOSATTI, AND ROBERTO CAR. SCIENCE 295 2427-2430 (2002)

Hibat-Allah, Inack, Wiersema, Melko, Carrasquilla. Variational neural annealing. Nature Machine Intelligence (2021)

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 multiple algorithms: no annealing (CQO), Variational classical annealing, Variational Quantum Annealing 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 $\hat{H}(t) = \hat{H}_{\text{target}} - \Gamma(t) \sum_i \hat{\sigma}_i^x$

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.

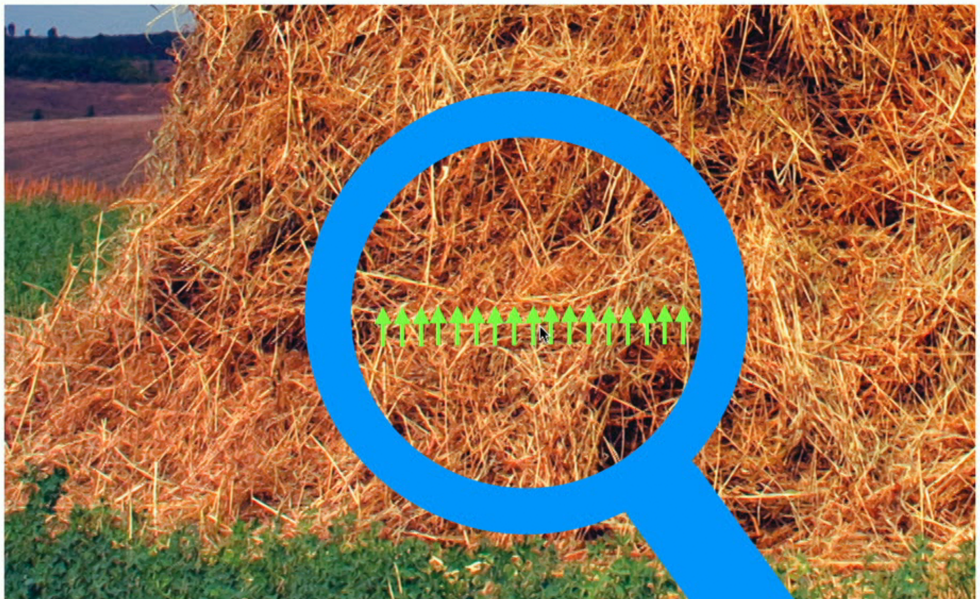
- $H_{\text{target}} = -\frac{1}{2}\sum_{i\neq j}\frac{J_{ij}}{\sqrt{N}}\sigma_i\sigma_j$

- 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)



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

Variational annealing on FC spin glasses

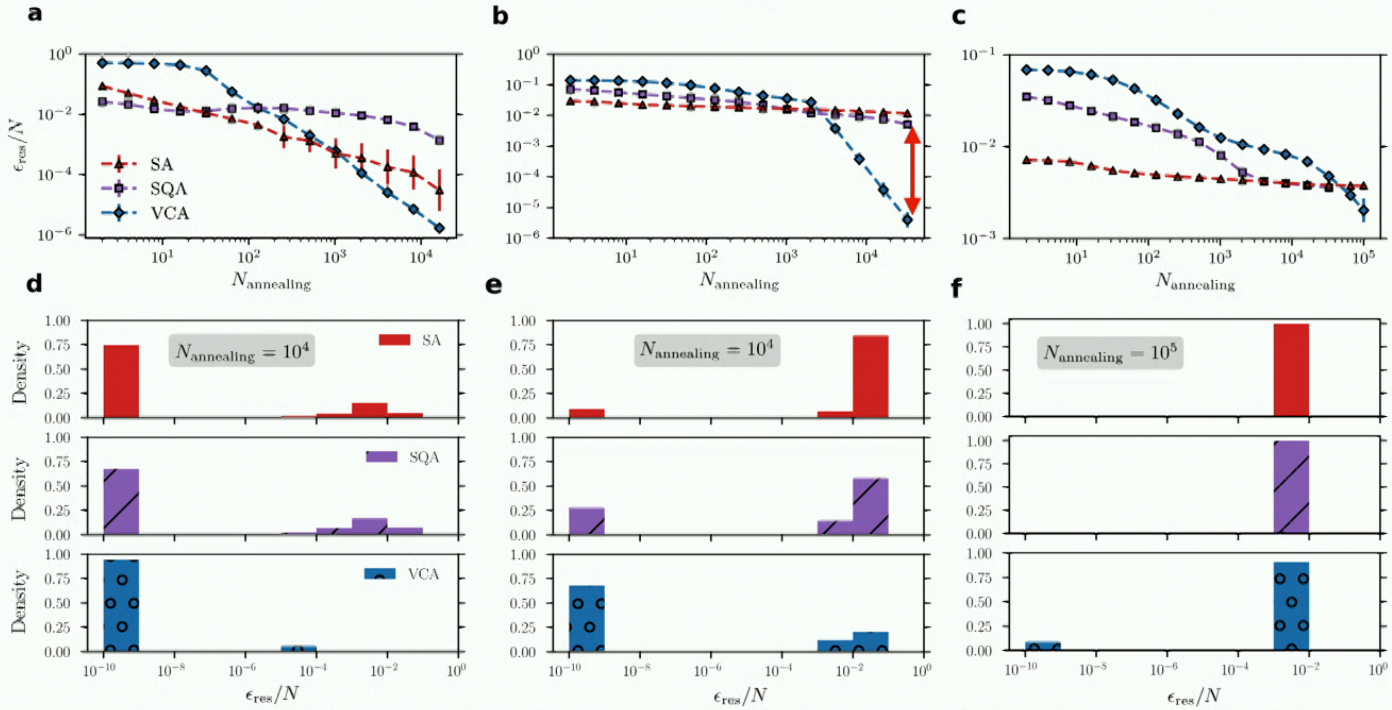


Figure 5. Benchmarking SA, SQA (P = 100 trotter slices) and VCA on the Sherrington-Kirkpatrick (SK) model and the Wishart planted ensemble (WPE). Panels (a),(b), and (c) display the residual energy per site as a function of $N_{\text{annealing}}$. (a) The SK model with $N = 100$ spins. (b) WPE with $N = 32$ 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 ϵ_{res}/N, which is within our numerical accuracy.

Hibat-Allah, Inack, Wiersema, Melko, Carrasquilla. Variational neural annealing. Nature Machine Intelligence (2021)

We use dilated RNNs

Conclusions

- Introduced 2 examples of applications of recurrent neural networks in quantum many-body physics.
- Quantum state reconstruction with RNNs.
- Introduced a variational neural annealing: Swapping **Boltzmann distribution** for a model with **efficient sampling** leads to better modelling: Spin glasses, frustrated models in QMC.
- 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.