

Reconstructing Quantum States With Generative Models

Roger Melko



UNIVERSITY OF
WATERLOO

PI QUANTUM
INTELLIGENCE
LAB



 Canada
Research
Chairs
Chaires
de recherche
du Canada


IQC Institute for
Quantum
Computing

 VECTOR
INSTITUTE

 **CREATIVE**
DESTRUCTION LAB

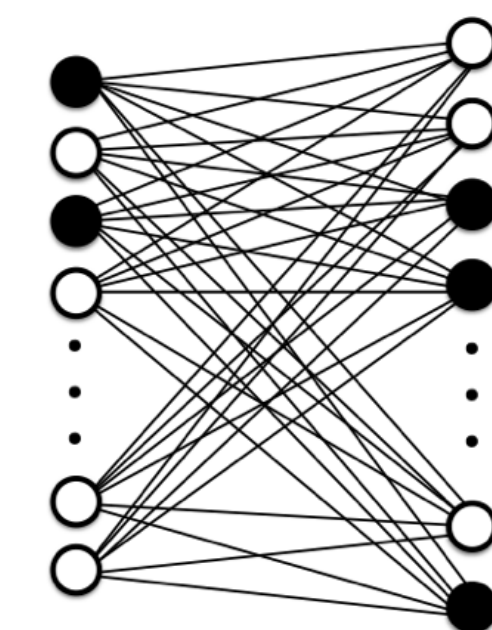
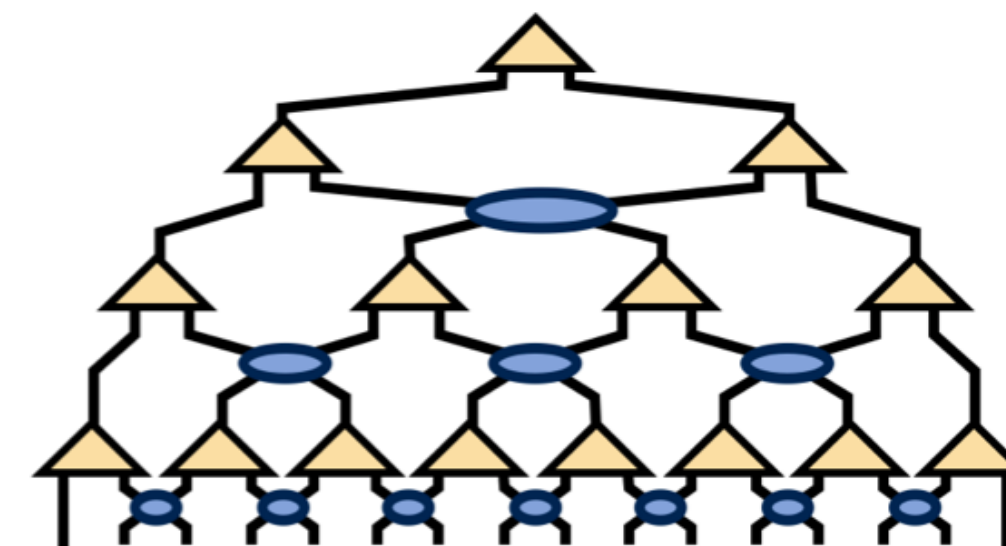
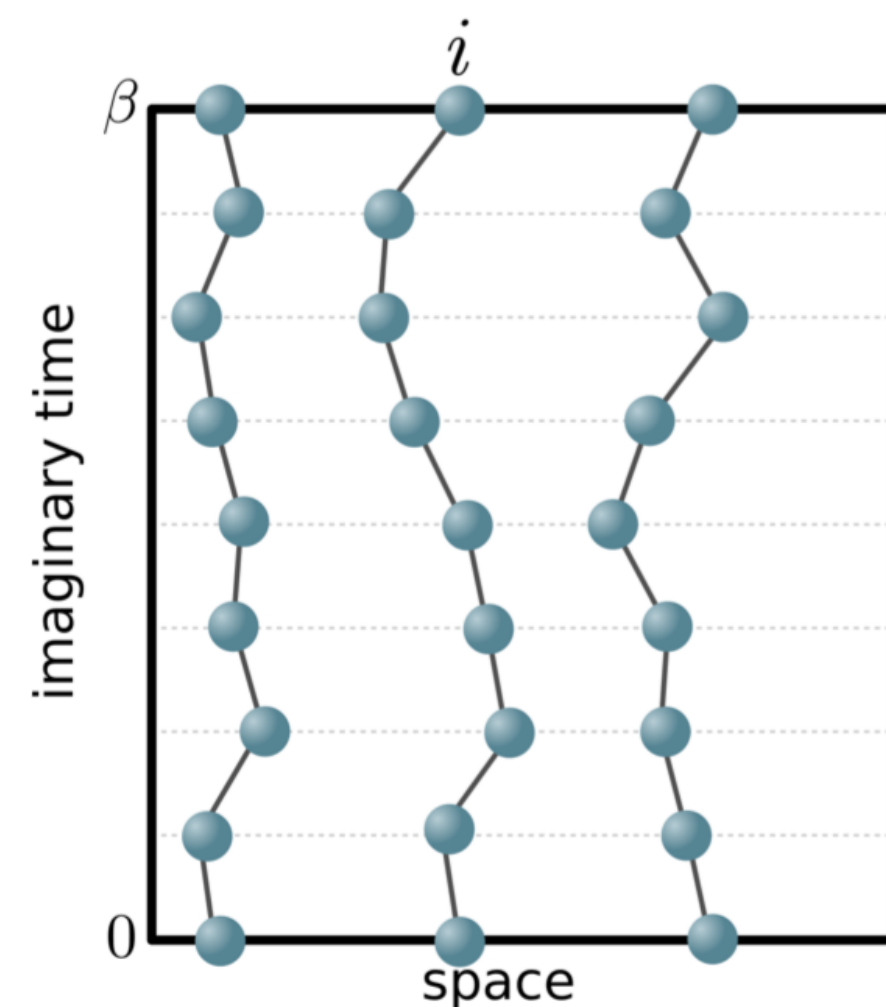
The need for quantum simulators

- What makes “solving” or “simulating” a many-body system hard?
- E.g. given a ***Hamiltonian***, we may seek to find a ground state, & calculate *correlation functions, elementary excitations, topological defects, etc.*

$$H = -t \sum_{\langle ij \rangle, \sigma} (c_{i\sigma}^\dagger c_{j\sigma} + c_{i\sigma} c_{j\sigma}^\dagger) + U \sum_i n_{i\uparrow} n_{i\downarrow}$$

$$H = J_1 \sum_{\langle ij \rangle} \mathbf{S}_i \cdot \mathbf{S}_j + J_2 \sum_{\langle\langle ij \rangle\rangle} \mathbf{S}_i \cdot \mathbf{S}_j$$

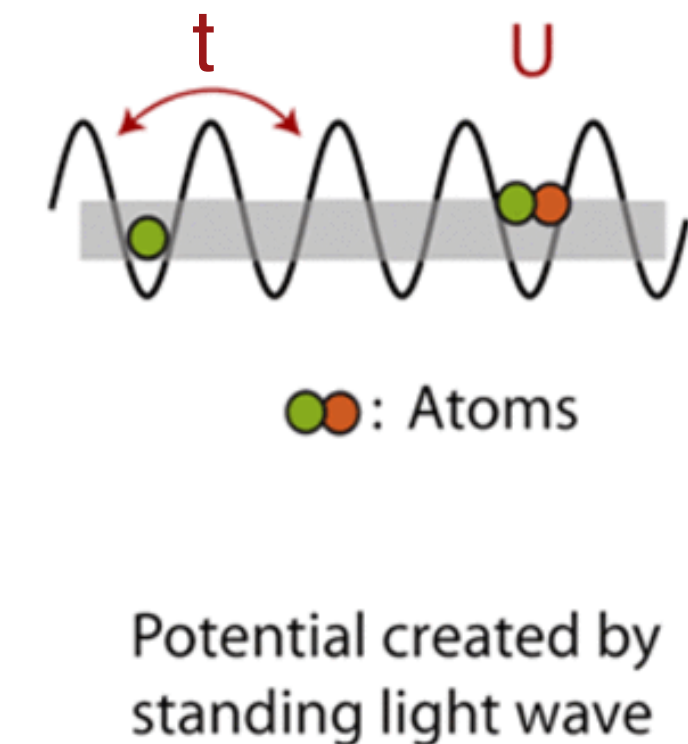
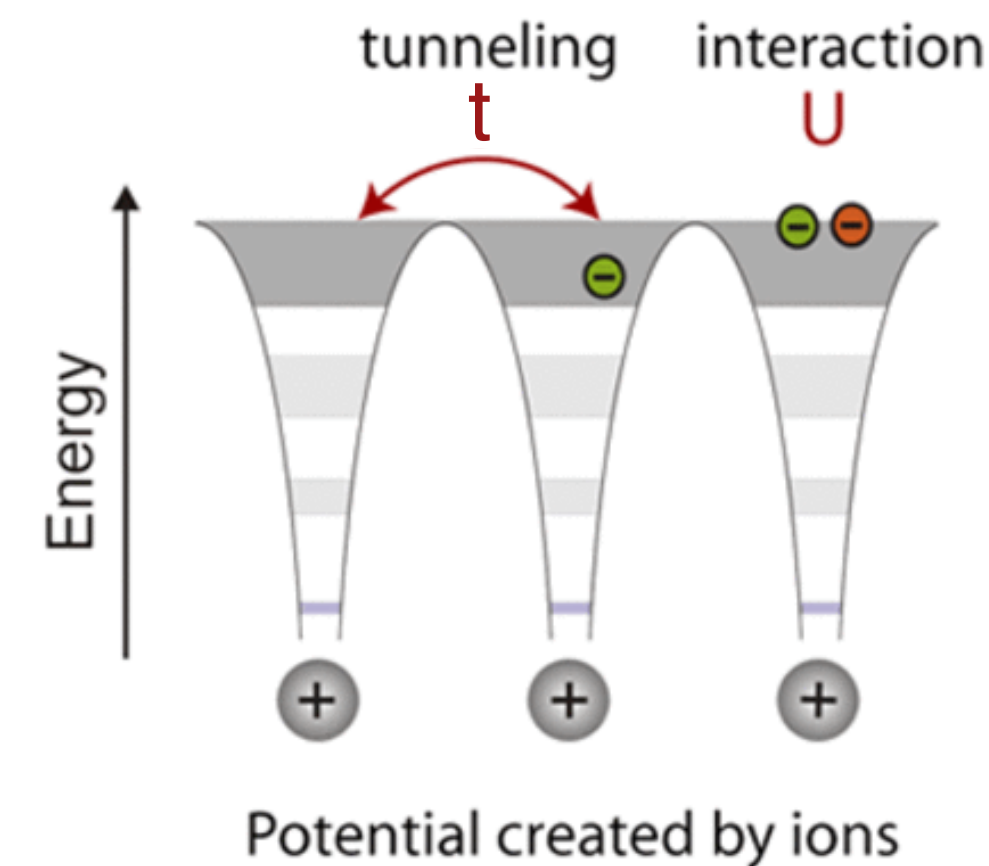
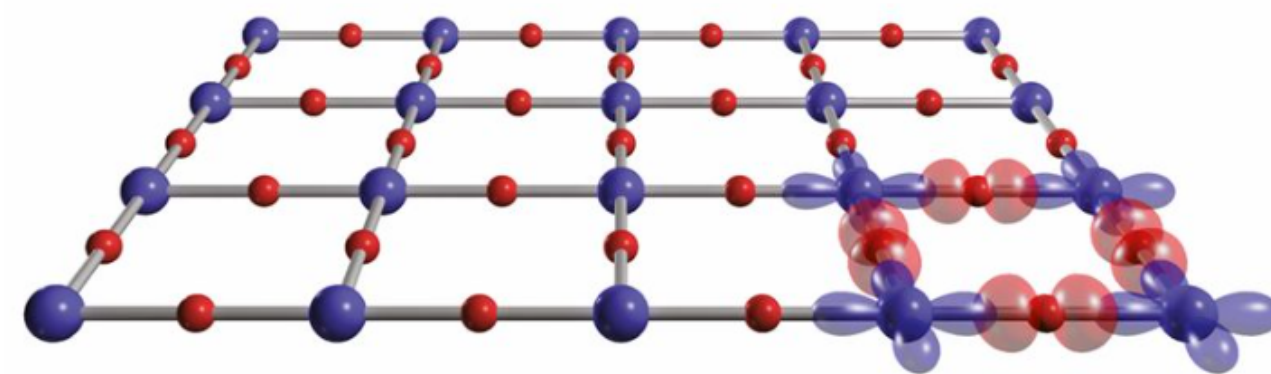
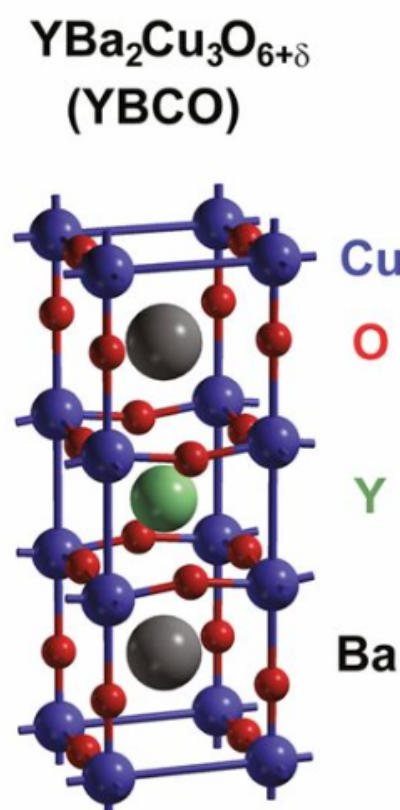
$$T = \begin{pmatrix} \alpha_1 & \beta_2 & & & & 0 \\ \beta_2 & \alpha_2 & \beta_3 & & & \\ & \beta_3 & \alpha_3 & \ddots & & \\ & & \ddots & \ddots & \beta_{m-1} & \\ & & & \beta_{m-1} & \alpha_{m-1} & \beta_m \\ 0 & & & & \beta_m & \alpha_m \end{pmatrix}$$



$$\epsilon(\lambda) = \langle \psi_\lambda | H | \psi_\lambda \rangle$$

The need for quantum simulators

- With the advent of experimental simulators, measurements of a quantum state turn this into a ***data-driven*** learning problem.



- Here, numerical methods do not need to “solve” the Hamiltonian. However, they can be used to probe the experiment and (ultimately) reconstruct the underlying quantum state

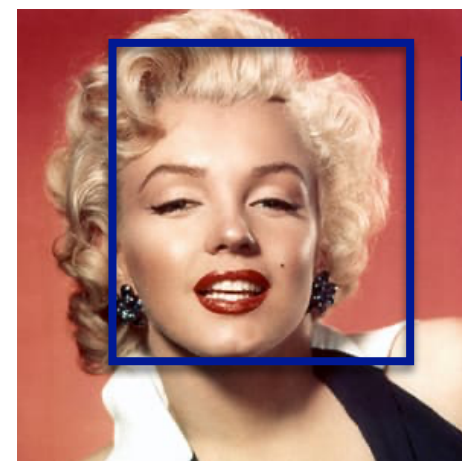
Supervised versus unsupervised learning

- Supervised learning can classify features in the data.

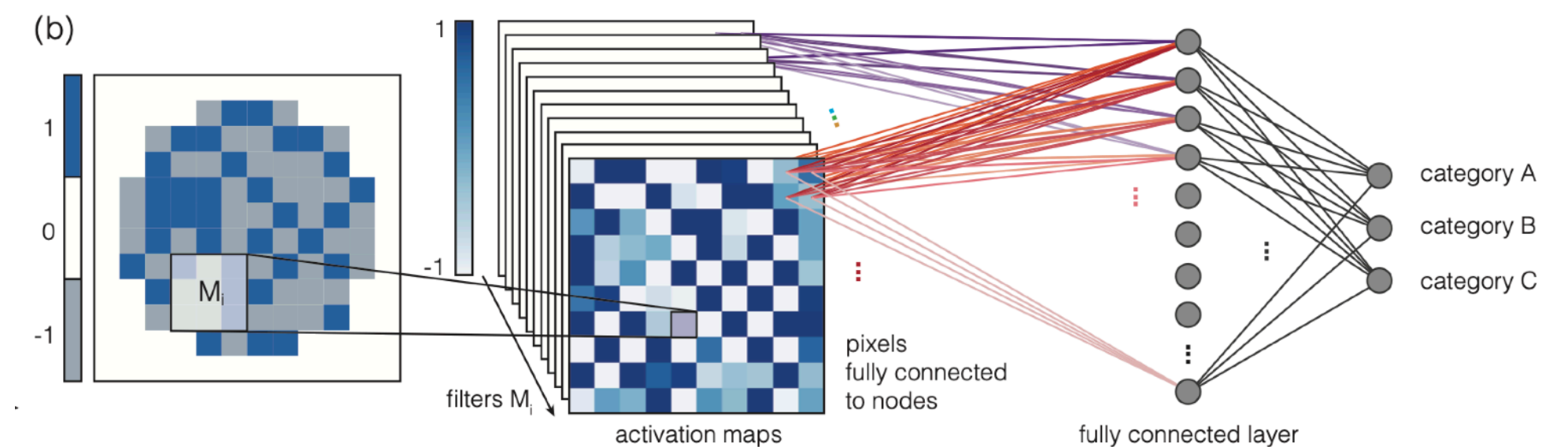
albert einstein



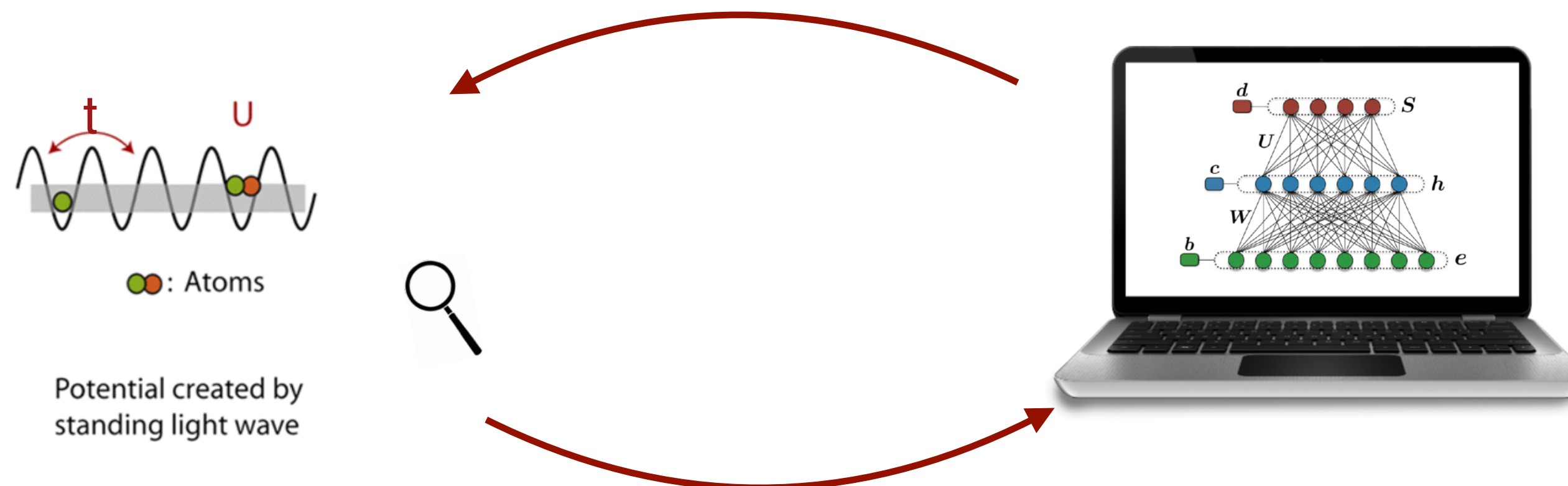
marilyn monroe



Bohrdt et al., Nature Physics 15, 921 (2019)



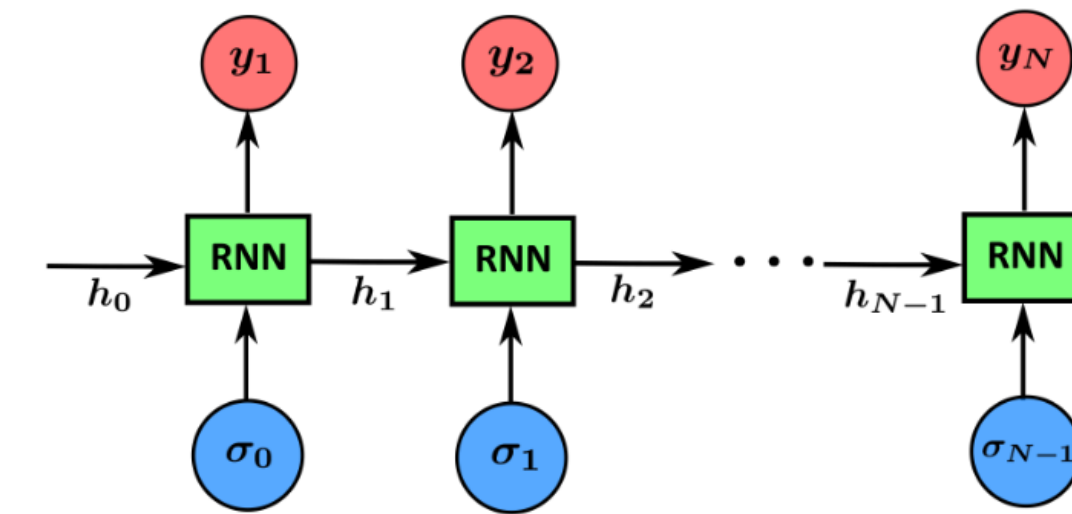
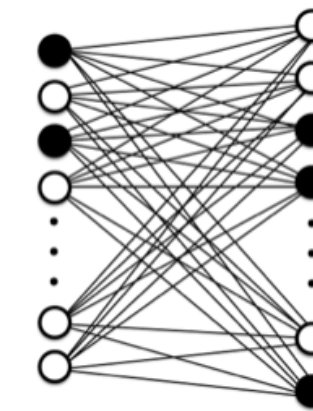
- Unsupervised learning deals with un-labelled data. We want to use it to characterize and validate the state, and give feedback to experimental design



Outline:

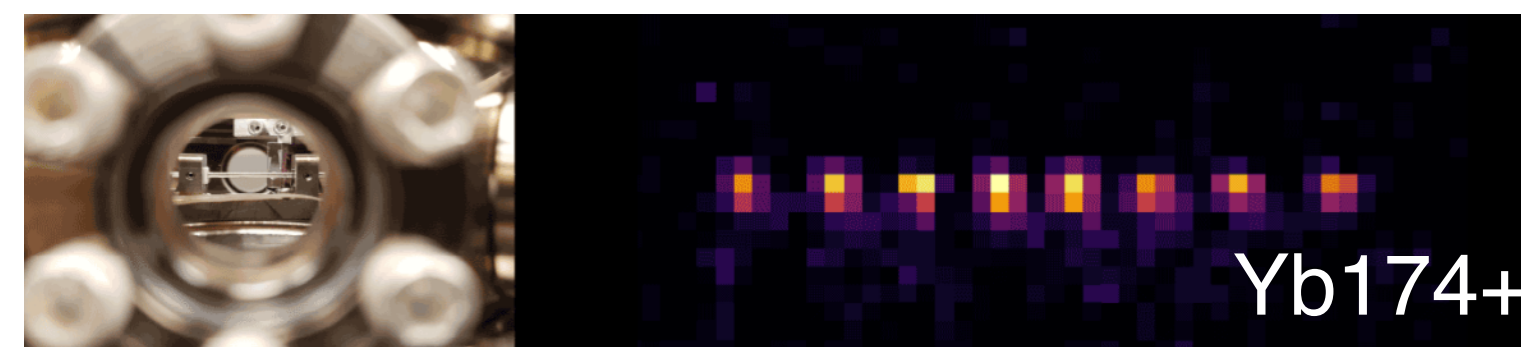
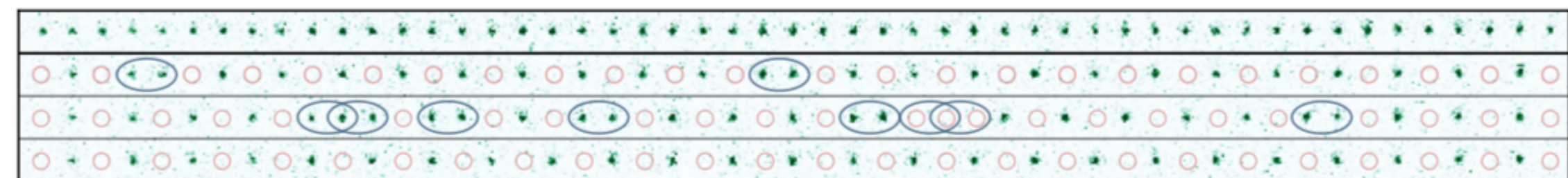
- **An introduction to generative models**

- Restricted Boltzmann machines
- Recurrent Neural Networks



- **Applications in near-term experiments**

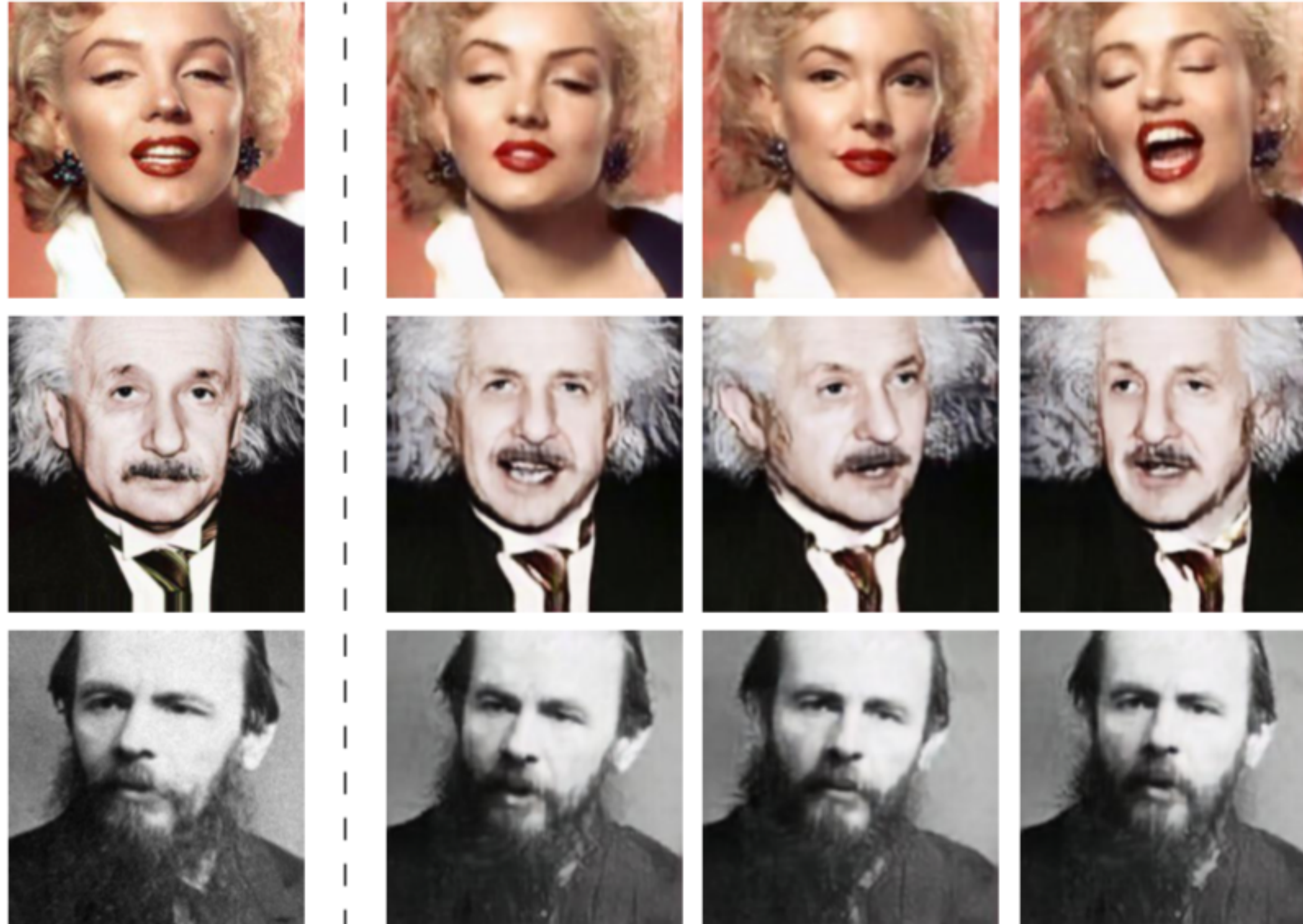
- Neutral Rydberg atoms
- Trapped ions



R. Islam Lab,
IQC Waterloo

Generative models

VoxCeleb2

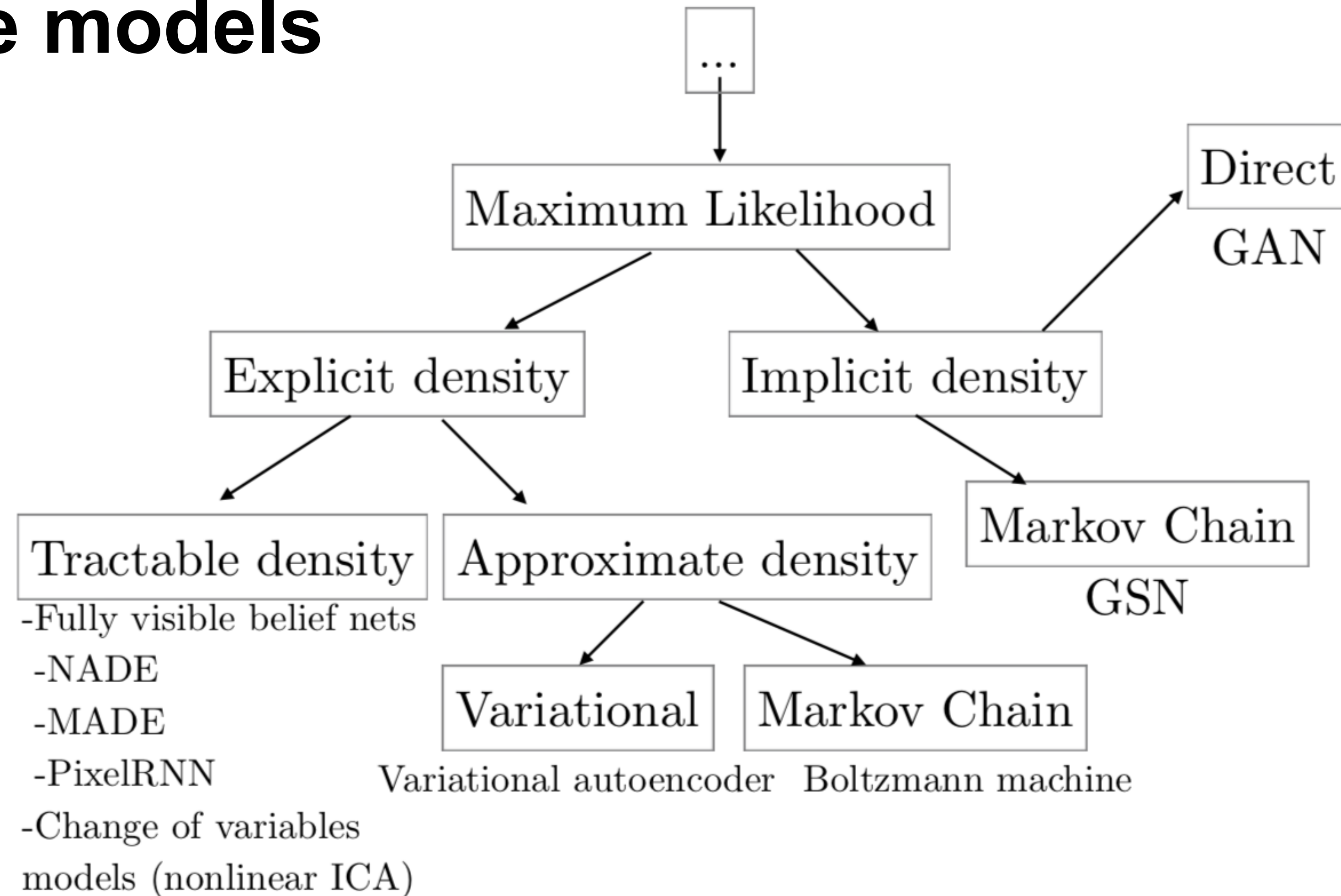


Source

Generated images

arXiv:1905.08233

Generative models

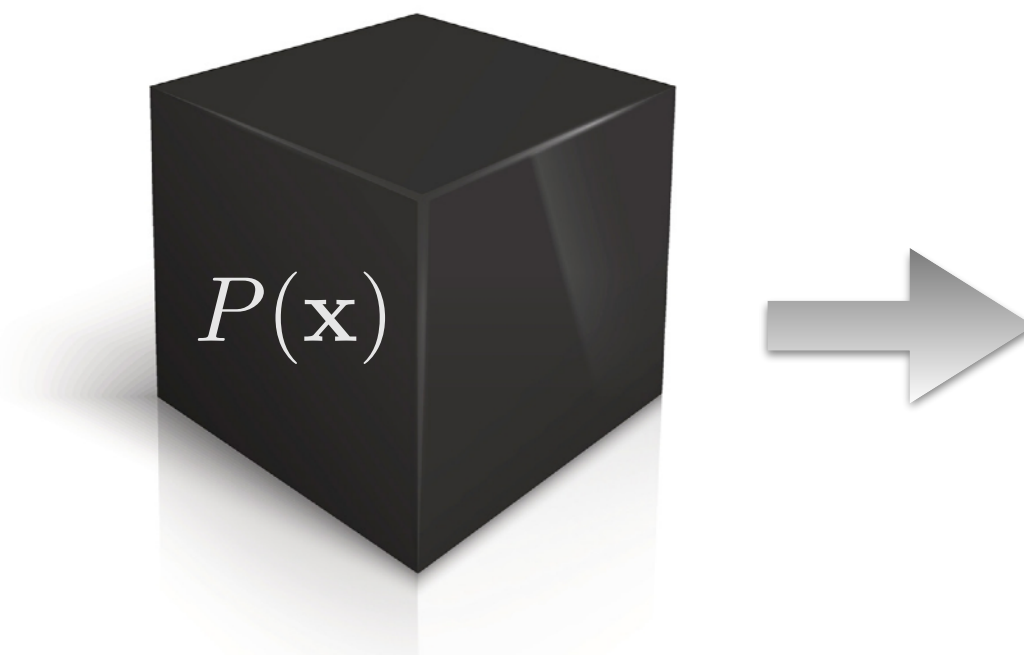


Generative Adversarial Networks
Ian Goodfellow
[arXiv:1701.00160](https://arxiv.org/abs/1701.00160)

What physicists want to know about advances in generative modeling
Michael Albergo
<https://machine19.github.io>

Generative models

Imagine that you have a large number of measurements, samples, or configurations, drawn from some unknown probability distribution



$$\mathbf{x}_1 = (1, 0, 0, 1, 1, 1, 0, 0, 0, 0, \dots, 1)$$

$$\mathbf{x}_2 = (1, 1, 1, 0, 1, 1, 0, 1, 1, 1, \dots, 1)$$

$$\mathbf{x}_3 = (0, 1, 1, 0, 0, 1, 0, 1, 0, 1, \dots, 0)$$

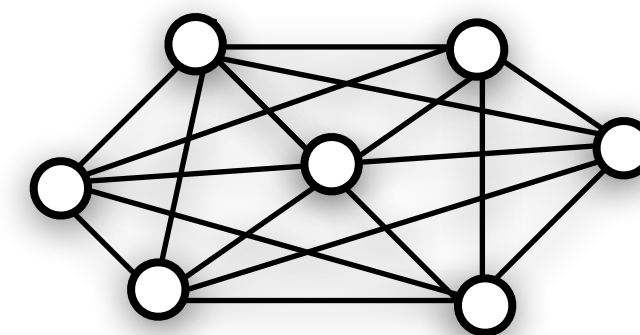
...

$$\mathbf{x}_{10,000} = (0, 0, 1, 1, 1, 1, 0, 1, 0, 0, \dots, 0)$$

$$||\mathcal{D}|| = 10,000$$

Goal: to learn from the data the best possible approximation, $p(\mathbf{x}) \approx P(\mathbf{x})$ that generalizes well to data not contained in the training set

$$p(x) \approx \frac{1}{||\mathcal{D}||} \sum_{\mathbf{x}_k \in \mathcal{D}} \delta_{\mathbf{x}, \mathbf{x}_k}$$

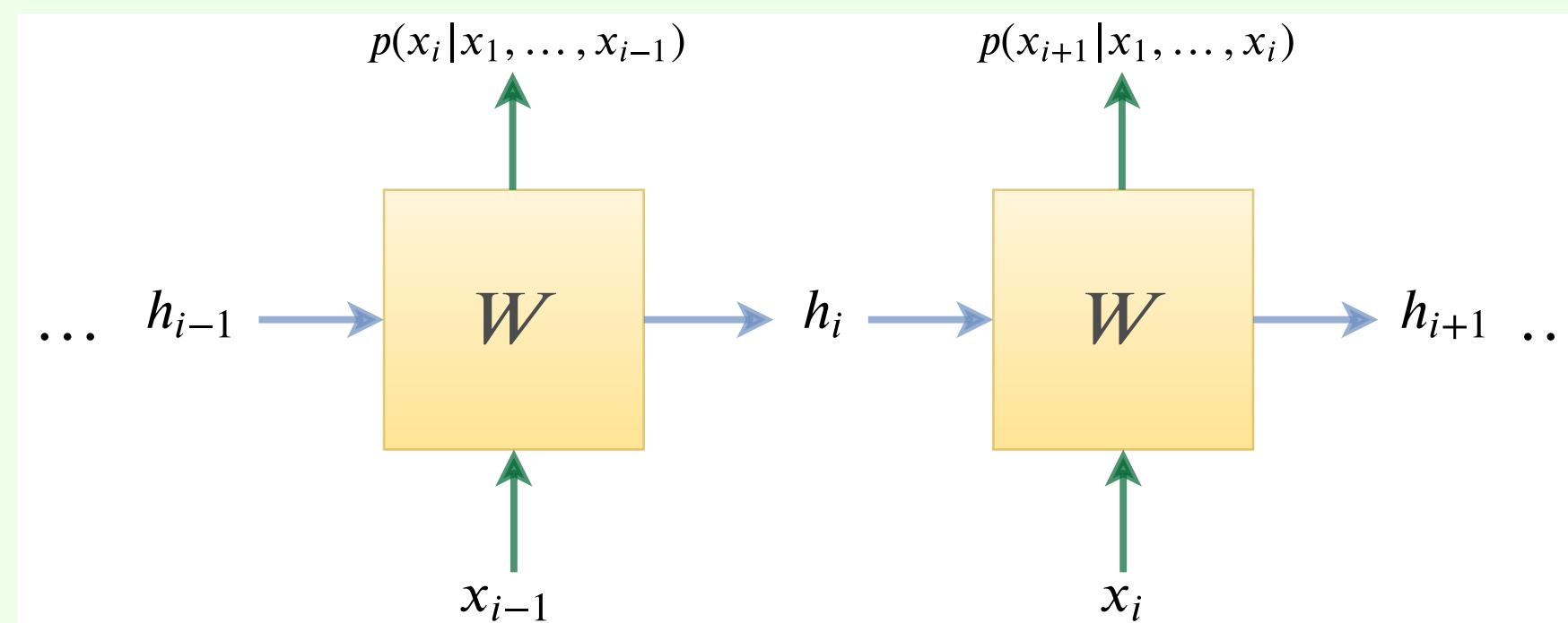


p_λ

a 'model'

Generative models

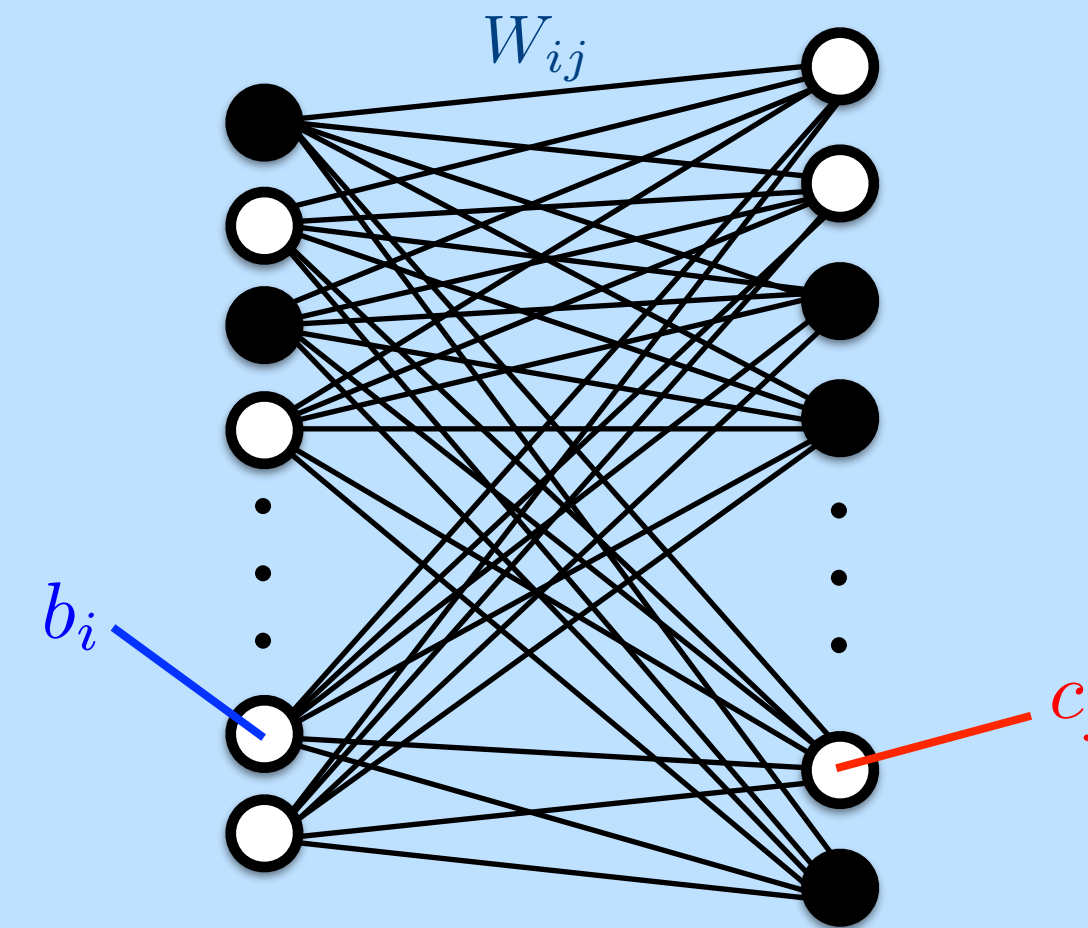
Recurrent Neural Network



$$p_{\lambda}(x_1, \dots, x_N) = \prod_{i=1}^N p(x_i | x_1, \dots, x_{i-1})$$

$$= p(x_1)p(x_2|x_1)p(x_3|x_1, x_2) \dots$$

Restricted Boltzmann Machine



$$x_i = 0, 1 \quad h_j = 0, 1$$

$$E_{\lambda}(\mathbf{x}, \mathbf{h}) = - \sum_{ij} W_{ij} x_i h_j - \sum_i b_i x_i - \sum_j c_j h_j$$

$$\lambda = \{W, b, c\} \quad \text{model parameters}$$

$$p_{\lambda} = \frac{1}{Z_{\lambda}} e^{-E_{\lambda}(\mathbf{x}, \mathbf{h})}$$

$$p_{\lambda}(\mathbf{x}) = \sum_{\mathbf{h}} p_{\lambda}(\mathbf{x}, \mathbf{h})$$

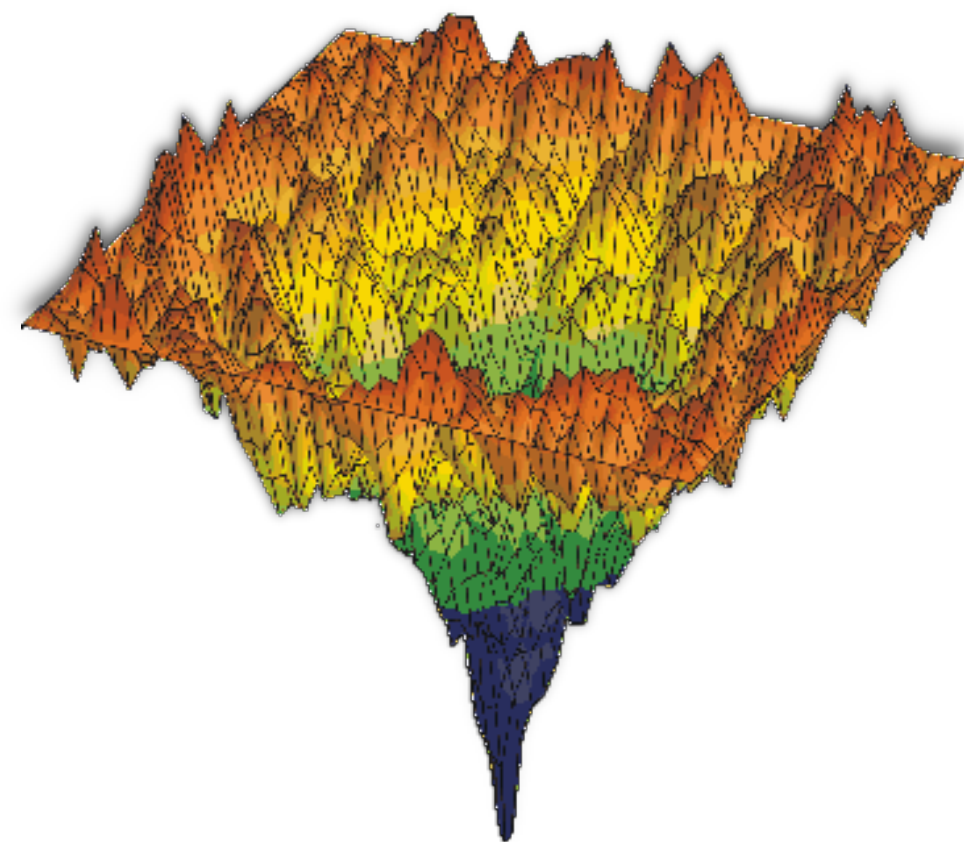
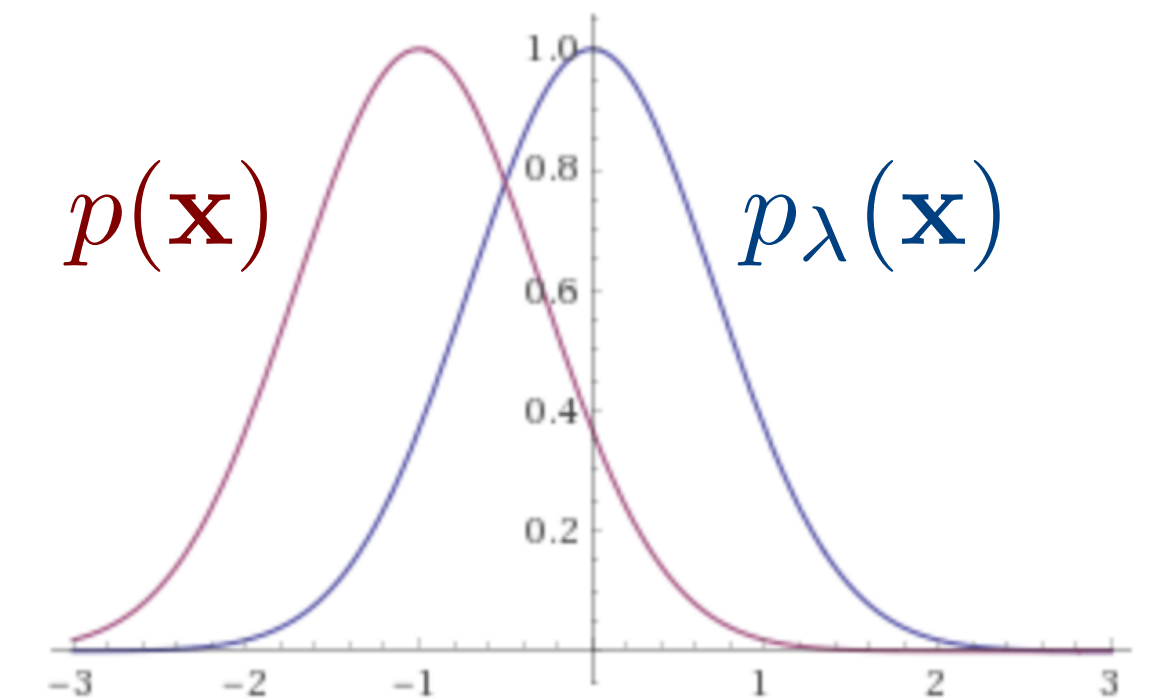
Training an RBM

Training means tuning the machine parameters to minimize the difference between the marginal distribution $p_\lambda(\mathbf{x}) = \sum_{\mathbf{h}} p_\lambda(\mathbf{x}, \mathbf{h})$ and the (unknown) physical “target” distribution

Define an optimization problem: minimize the **Kullback-Leibler divergence**

$$\text{KL}(p||p_\lambda) = \sum_{\mathbf{x}} p(\mathbf{x}) \log \frac{p(\mathbf{x})}{p_\lambda(\mathbf{x})} \geq 0$$

Equivalent to maximizing the “log-likelihood” $\mathcal{L} = \langle \log p_\lambda(\mathbf{x}) \rangle_p$



The optimization landscape is thus defined - find minima using gradient descent

$$\lambda' = \lambda - \eta \nabla \mathcal{L} \quad \lambda = \{W, b, c\}$$

The Generative Step

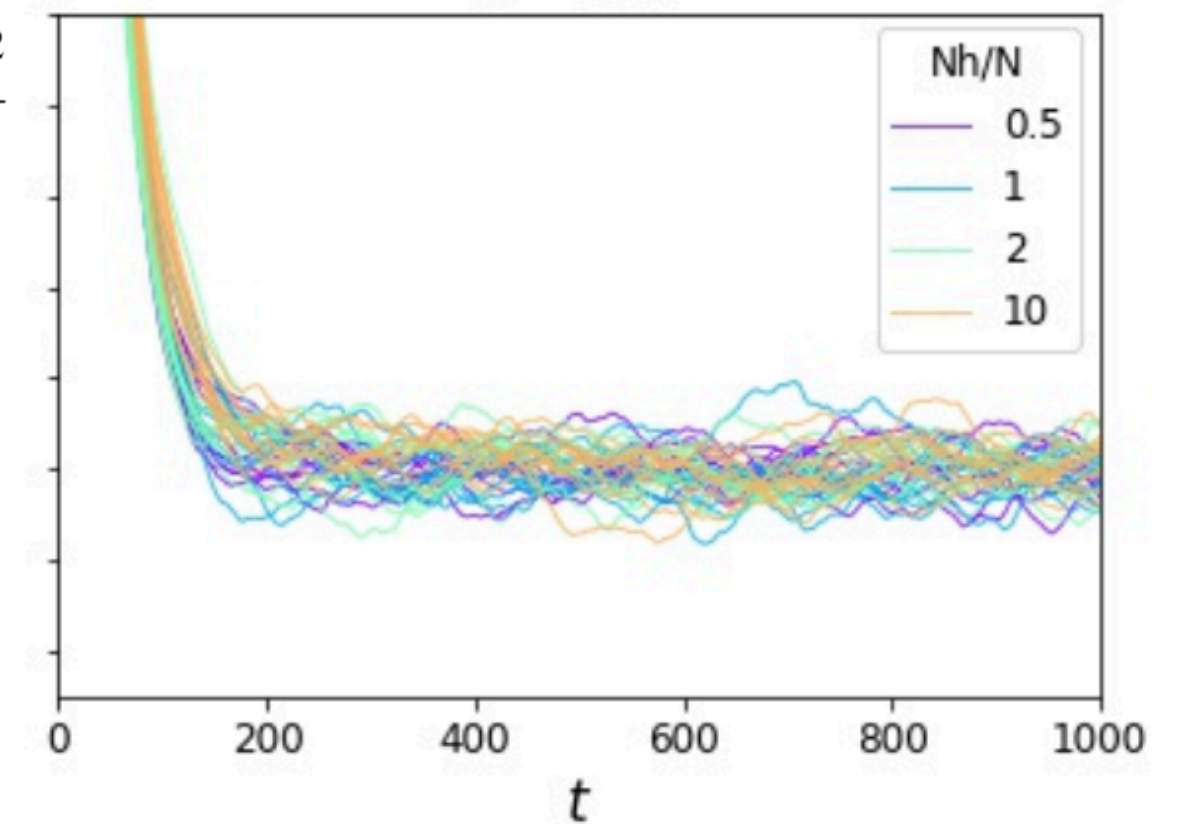
After the training is complete, and the weights and biases are stationary, one can sample new configurations



- Markov chain sampling for RBMs, etc.
- Tractable sampling for RNNs, other autoregressive models

compare to Sharir, Levine, Wies, Carleo, Shashua, Phys. Rev. Lett. 124, 020503 (2020)

$$A_Q(t) = \frac{\langle Q(i+t)Q(i) \rangle - \langle Q \rangle^2}{\langle Q^2 \rangle - \langle Q \rangle^2}$$



These samples can be used to probe physical estimators, like the energy, magnetic state, or the entanglement. These can validate the hardware against expectations, or be something inaccessible in experiment.

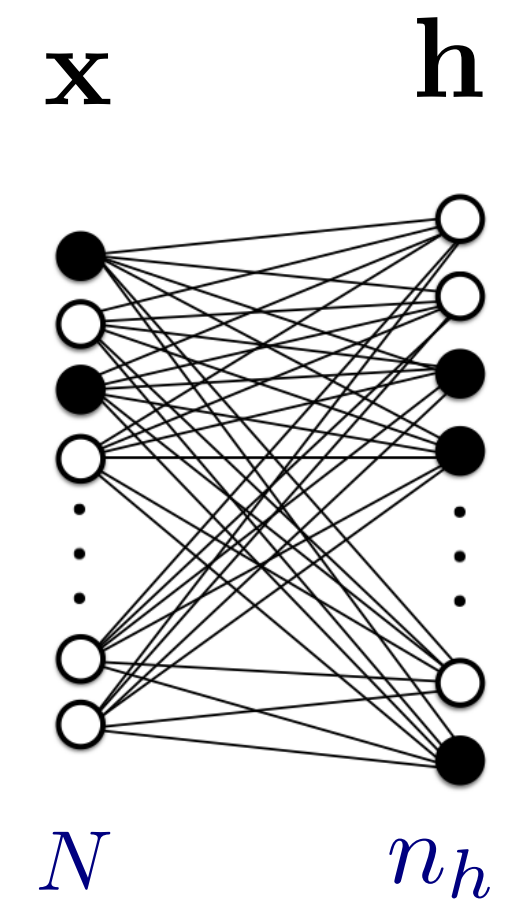
$$\langle Q \rangle = \frac{1}{M} \sum_x Q_x$$

e.g. sampling a RBM

After the training is complete, one can use Gibbs sampling to produce new configurations

$$\langle \mathcal{O} \rangle_{\text{joint}} = \sum_x \mathcal{O}_x \sum_h p_\lambda(x, h) \approx \langle \mathcal{O} \rangle_{\text{physical}}$$

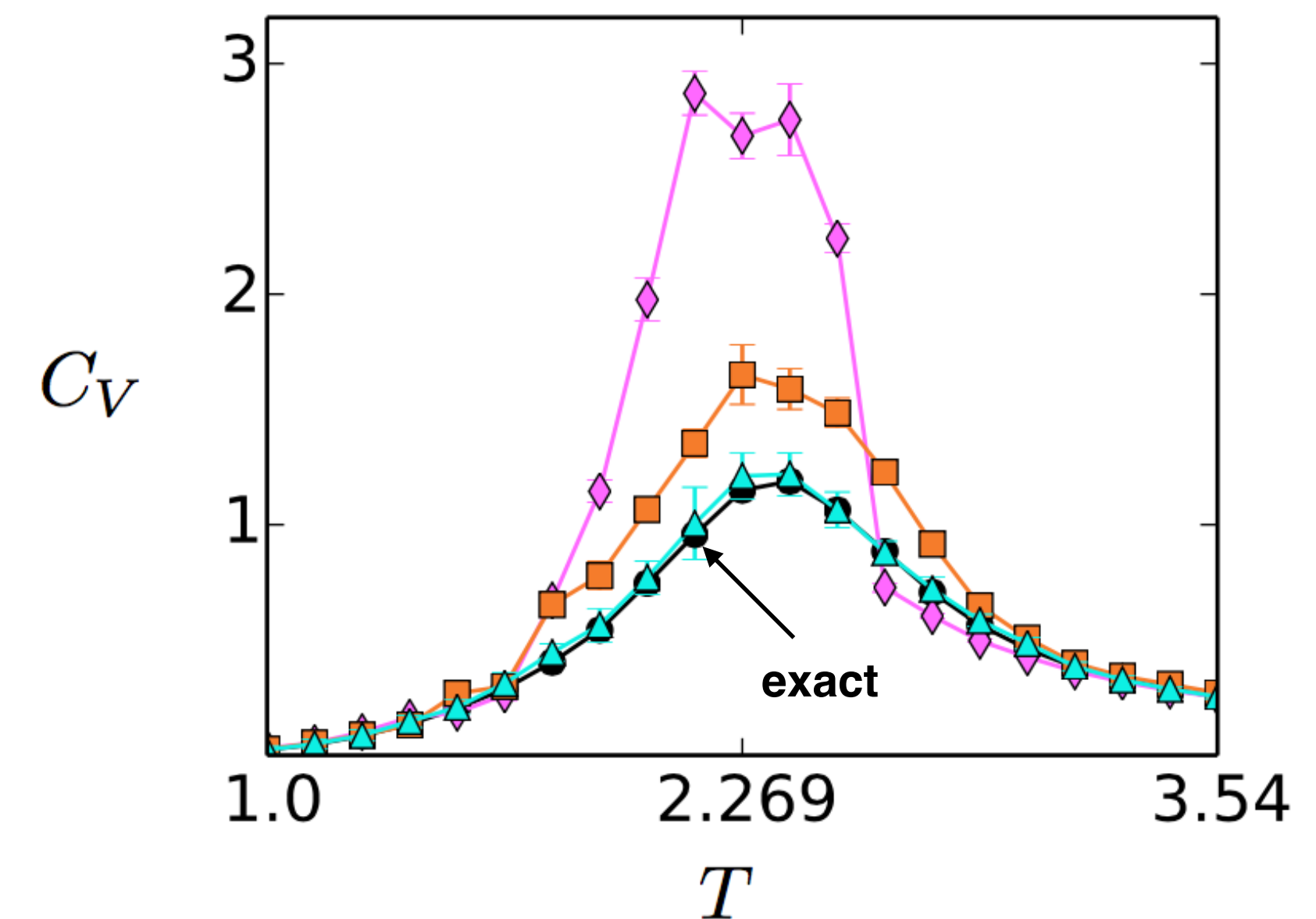
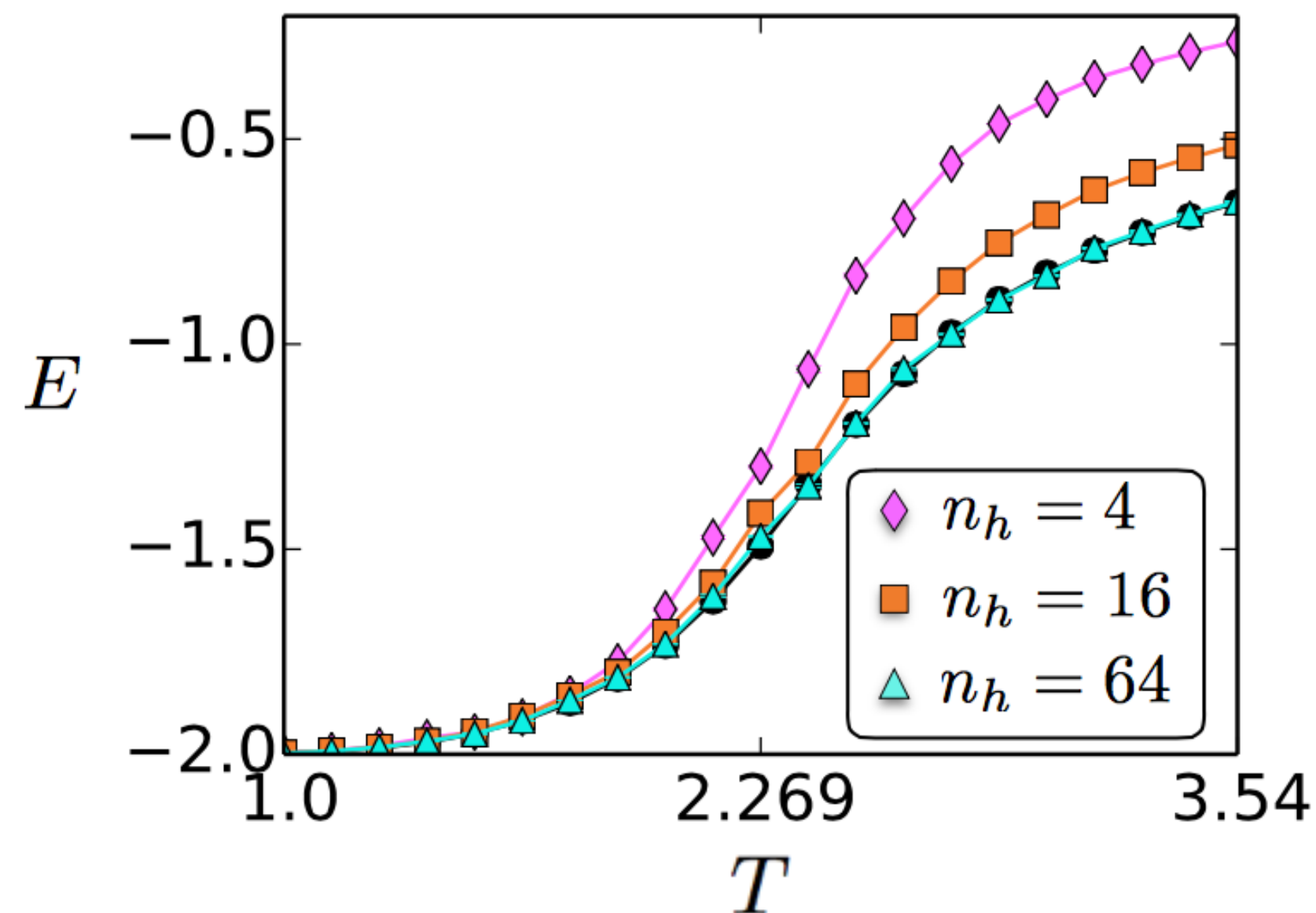
$p_\lambda(x) \rightarrow p(x)$



restrict observables $\mathcal{O} = \mathcal{O}_x$

Classical 2D
Ising model

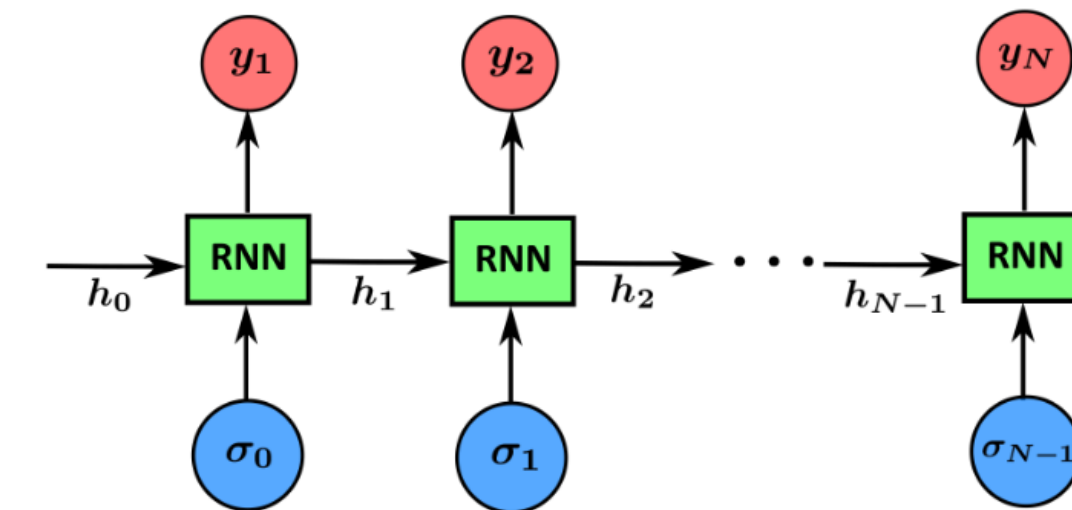
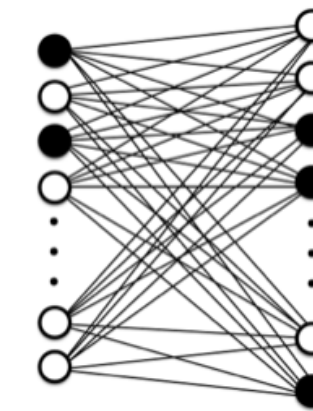
$N = 64$



Outline:

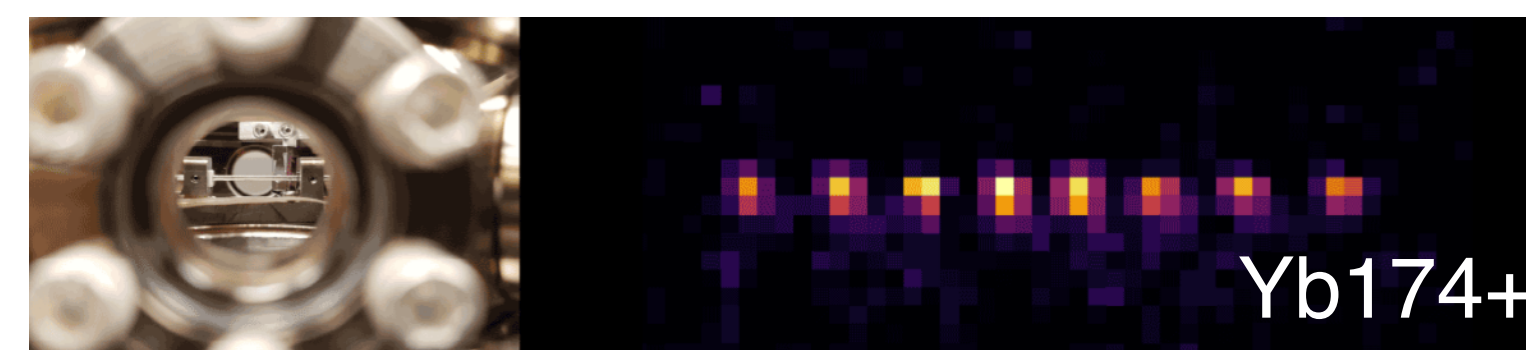
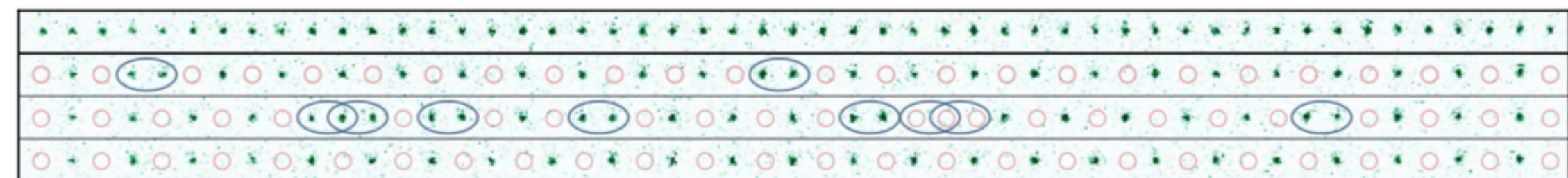
- **An introduction to generative models**

- Restricted Boltzmann machines
- Recurrent Neural Networks



- **Applications in near-term experiments**

- Neutral Rydberg atoms
- Trapped ions

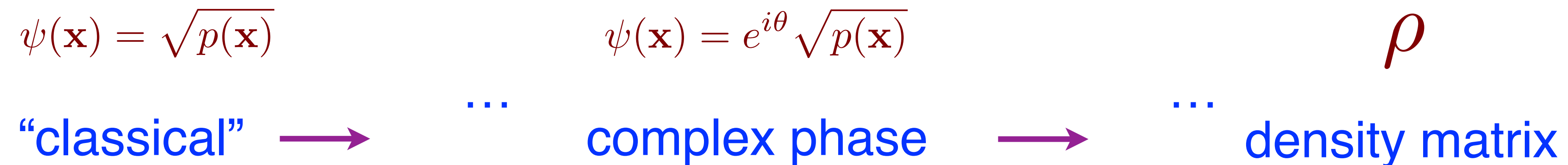


R. Islam Lab,
IQC Waterloo

Learning quantum states

Torlai, Mazzola, Carrasquilla, Troyer, RGM, Carleo,
Nature Physics 14, 447-450 (2018)

- Explicit density models can represent pure & mixed quantum states

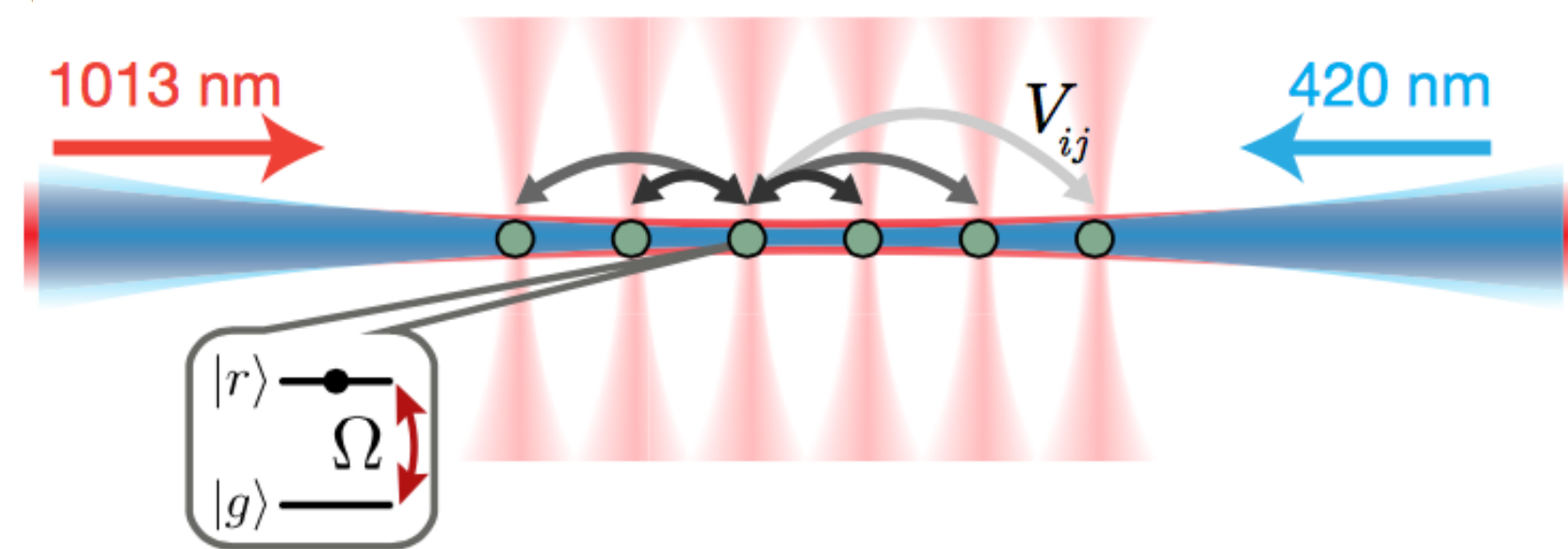
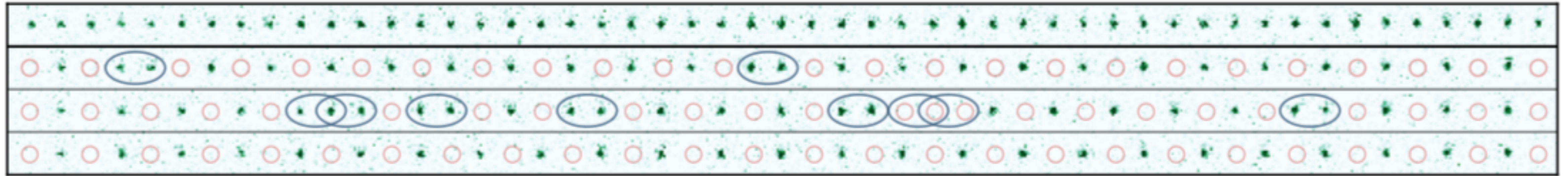


- Let’s focus first on the class of groundstate wavefunction where in some “convenient” basis coefficients are all positive and real

$$H = \begin{pmatrix} d & -|o| & -|o| & -|o| & -|o| \\ -|o| & d & -|o| & -|o| & -|o| \\ -|o| & -|o| & d & -|o| & -|o| \\ -|o| & -|o| & -|o| & d & -|o| \\ -|o| & -|o| & -|o| & -|o| & d \end{pmatrix} \quad \text{“stoquastic” Hamiltonians} \quad \psi_{\lambda}(\mathbf{x}) = \sqrt{p_{\lambda}(\mathbf{x})}$$

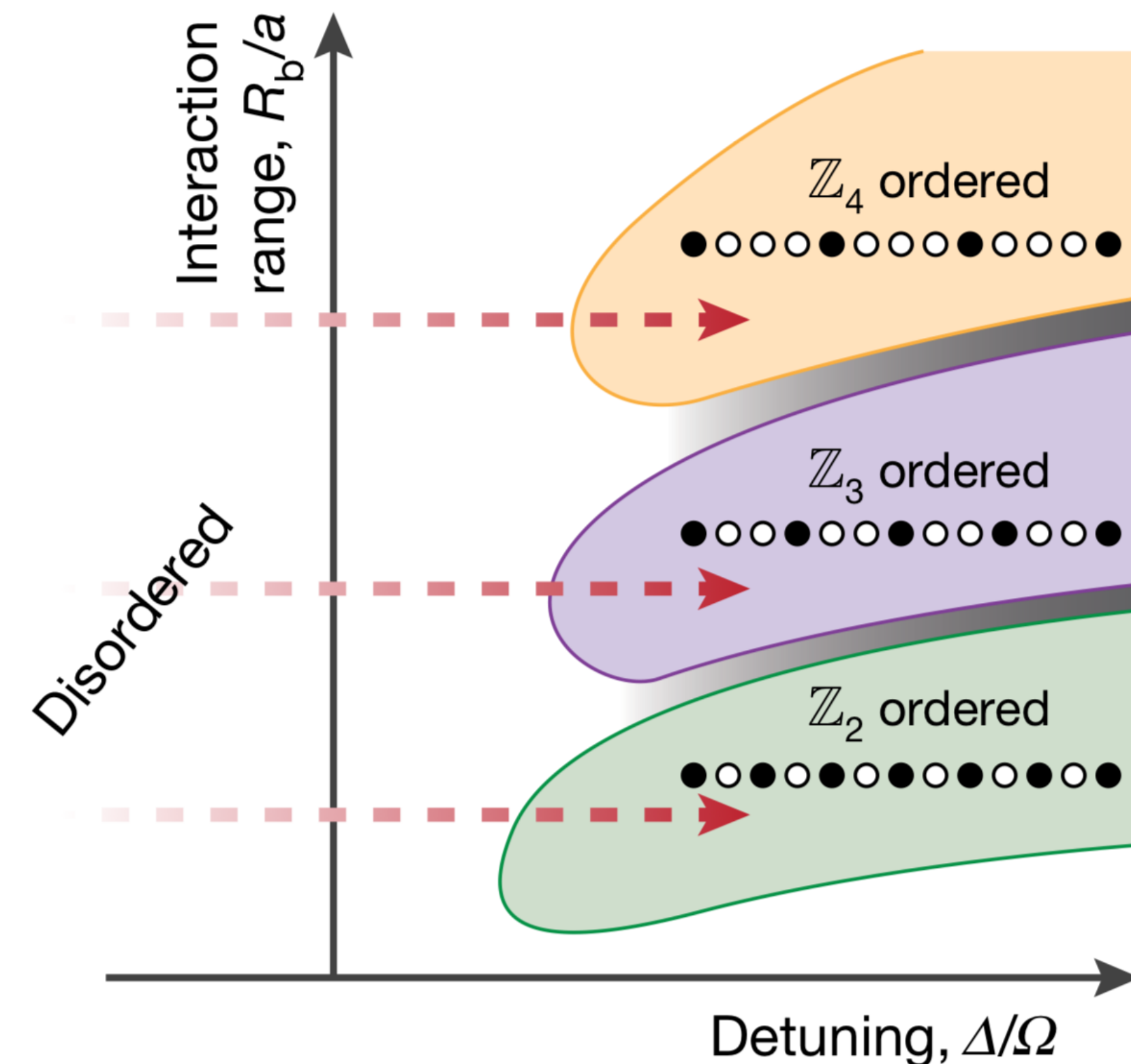
Experiments: Rydberg atoms

Bernien, Schwartz, Keesling, Levine, Omran, Pichler, Choi, Zibrov, Endres, Greiner, Vuletić, Lukin Nature 551, 579 (2017)



$$H = \Omega \sum_i \sigma_i^x - \Delta \sum_i n_i + \sum_{i < j} V_{ij} n_i n_j$$

$$\sigma^x = |g\rangle\langle r| + |r\rangle\langle g| \quad n = |r\rangle\langle r|$$



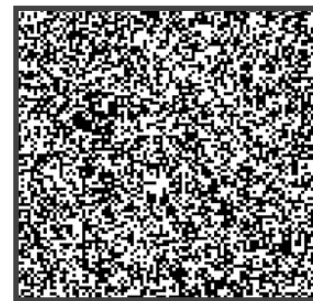
Learning (real +ve) wavefunctions

Torlai, Mazzola, Carrasquilla, Troyer, RGM, Carleo,
Nature Physics 14, 447-450 (2018)

In the case where the wavefunction is
real and positive in a certain basis

$$\psi_{\lambda}(\mathbf{x}) \propto \sqrt{p_{\lambda}(\mathbf{x})}$$

Train with samples in the σ^z basis



and after calculate estimators from samples
produced on the trained machine

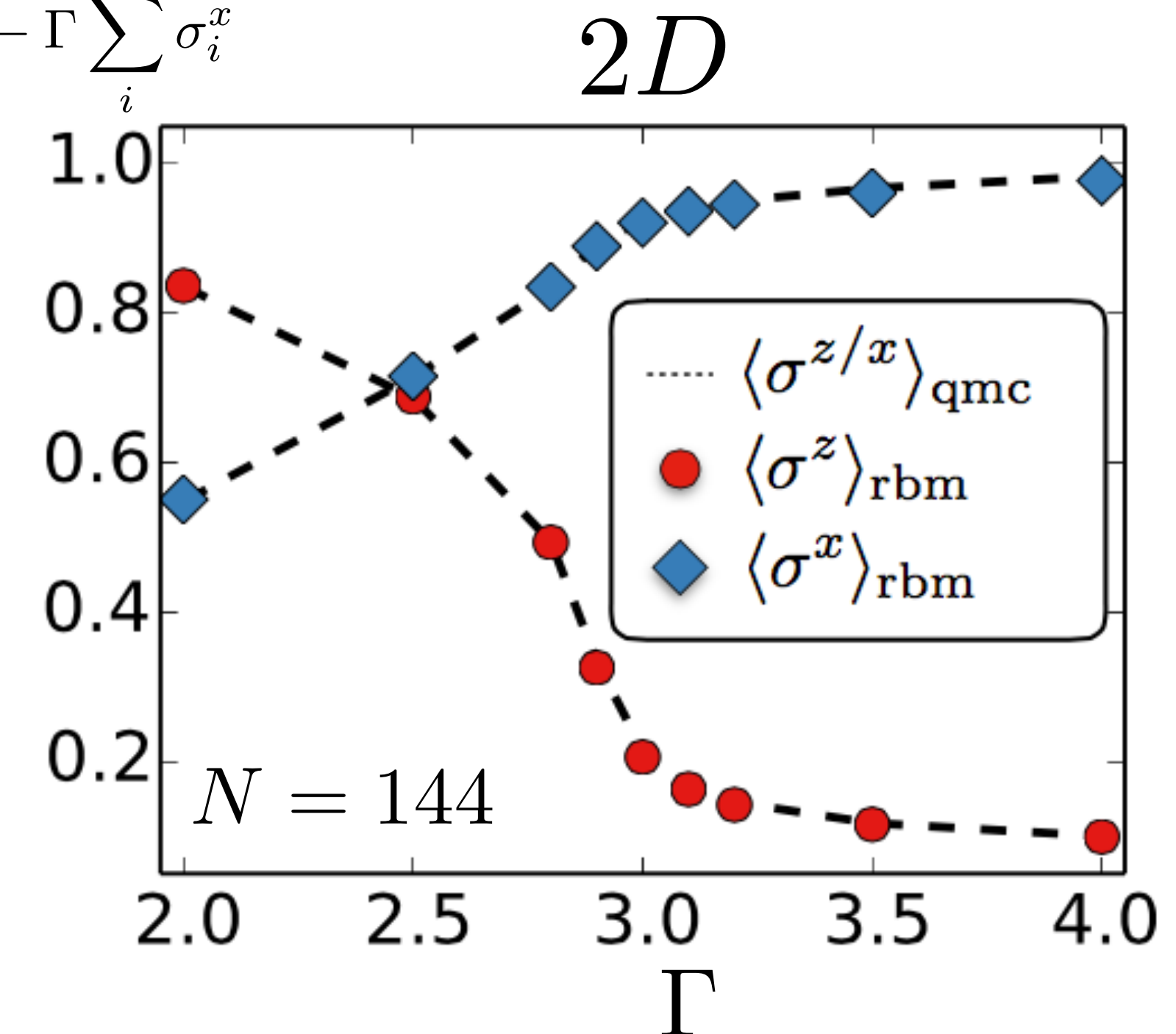
$$\langle \mathcal{O}^D \rangle = \sum_{\mathbf{x}} p_{\lambda}(\mathbf{x}) \mathcal{O}_{\mathbf{x}}$$

$$\langle \mathcal{O}^{OD} \rangle = \sum_{\mathbf{x}\mathbf{x}'} \sqrt{p_{\lambda}(\mathbf{x})} \sqrt{p_{\lambda}(\mathbf{x}')} \mathcal{O}_{\mathbf{x}\mathbf{x}'}$$

$$= \sum_{\mathbf{x}} p_{\lambda}(\mathbf{x}) \sum_{\mathbf{x}'} \frac{\sqrt{p_{\lambda}(\mathbf{x}')}}{\sqrt{p_{\lambda}(\mathbf{x})}} \mathcal{O}_{\mathbf{x}\mathbf{x}'}$$

“local” estimator

$$H = - \sum_{\langle ij \rangle} \sigma_i^z \sigma_j^z - \Gamma \sum_i \sigma_i^x$$



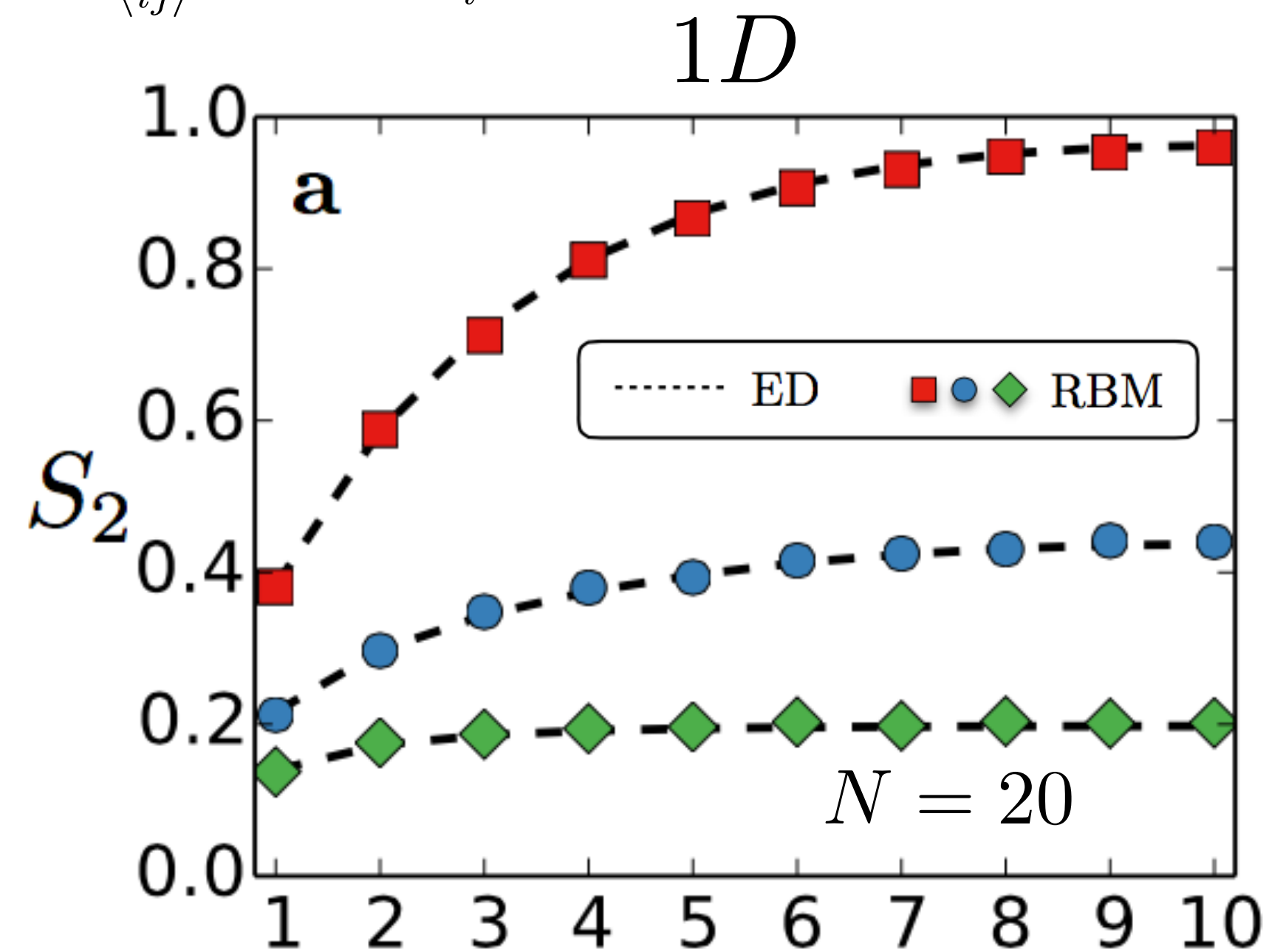
Learning (real +ve) wavefunctions

Torlai, Mazzola, Carrasquilla, Troyer, RGM, Carleo,
Nature Physics 14, 447-450 (2018)

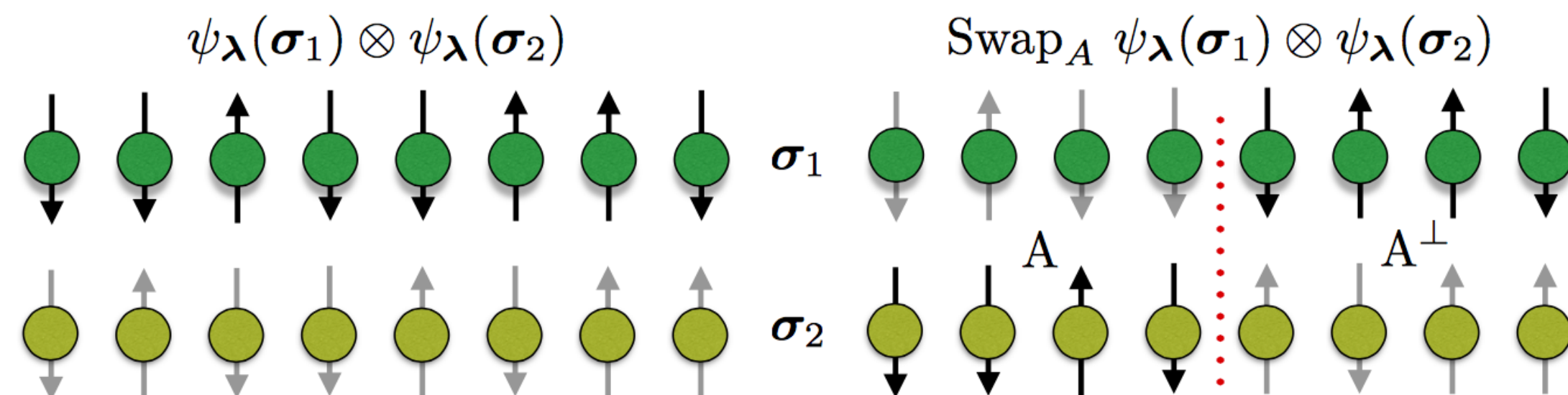
Entanglement entropy:

$$S_2(\rho_A) = -\log [\text{Tr}(\rho_A^2)]$$

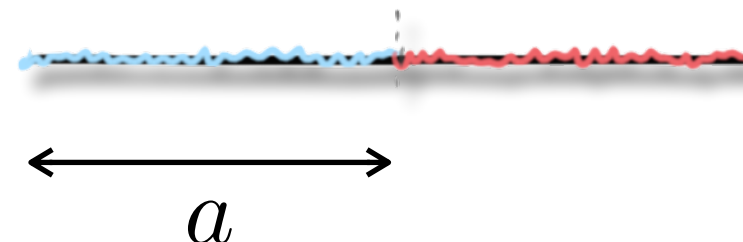
$$H = -\sum_{\langle ij \rangle} \sigma_i^z \sigma_j^z - \Gamma \sum_i \sigma_i^x$$



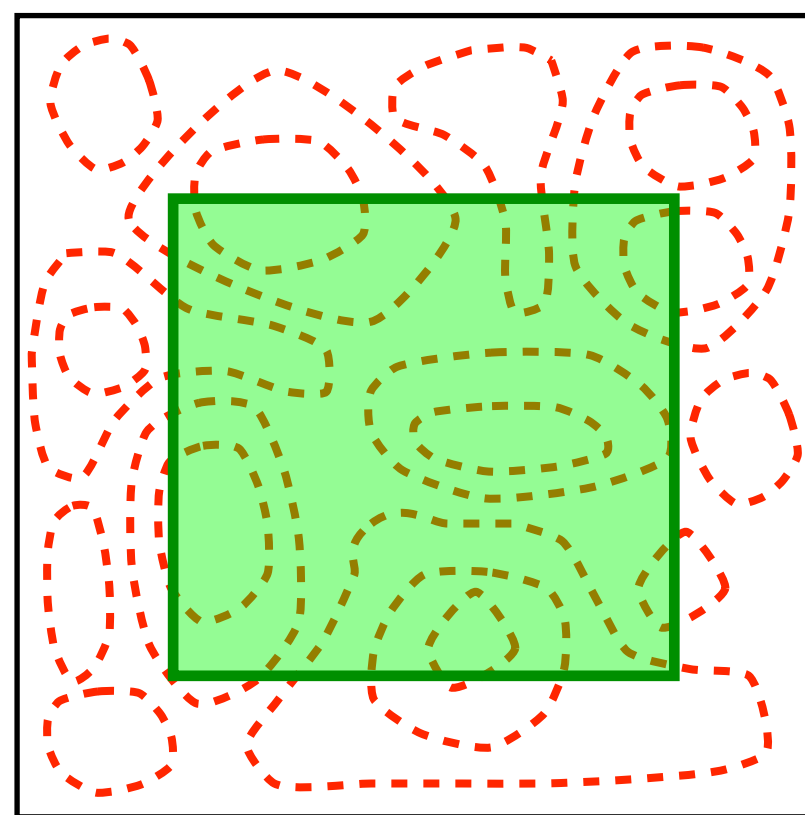
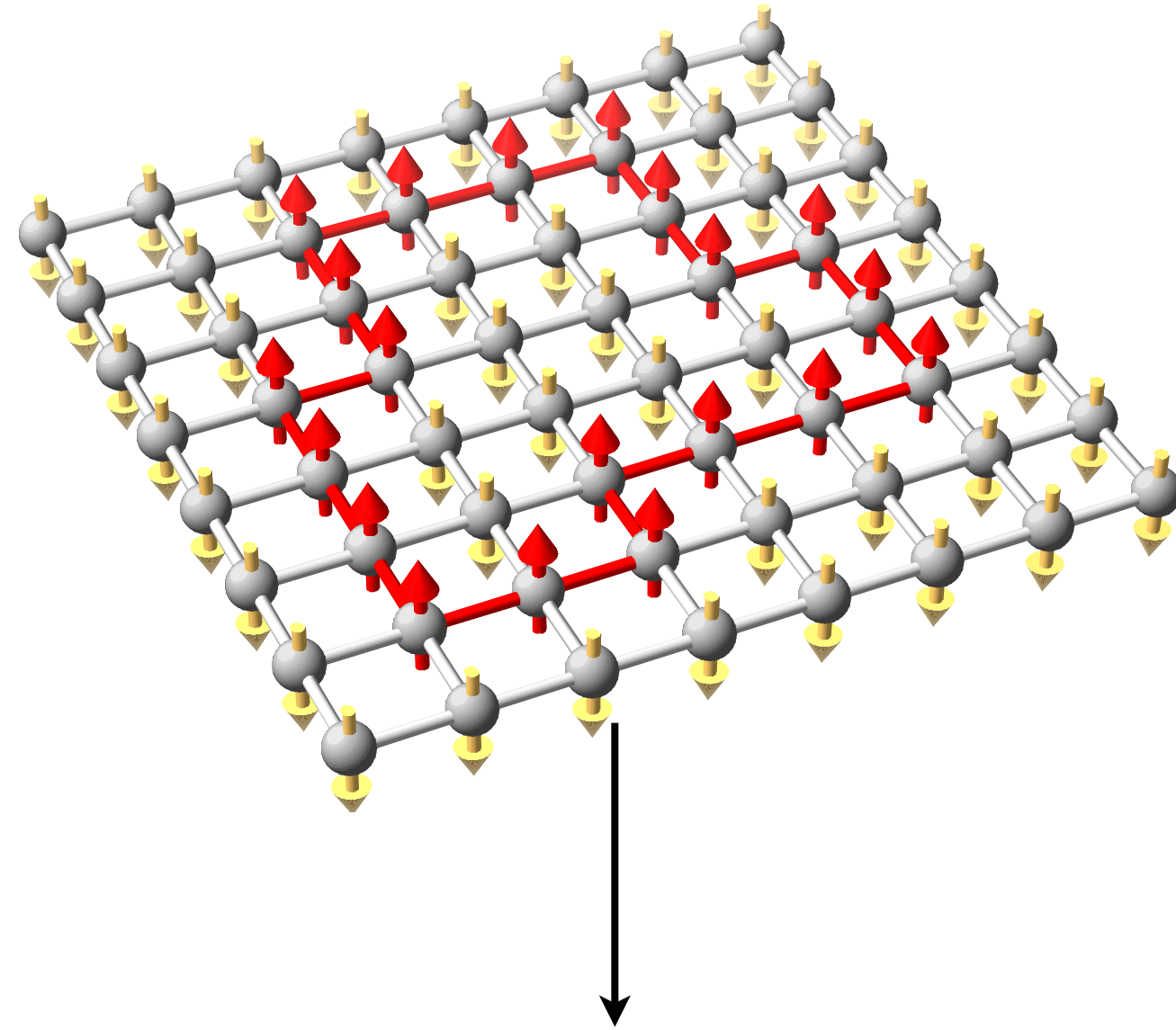
$$\psi_\lambda(\mathbf{x}) \propto \sqrt{p_\lambda(\mathbf{x})}$$



The expectation of the swap operator can be measured in the replicated (trained) RBM with a simple, universal algorithm that scales well.



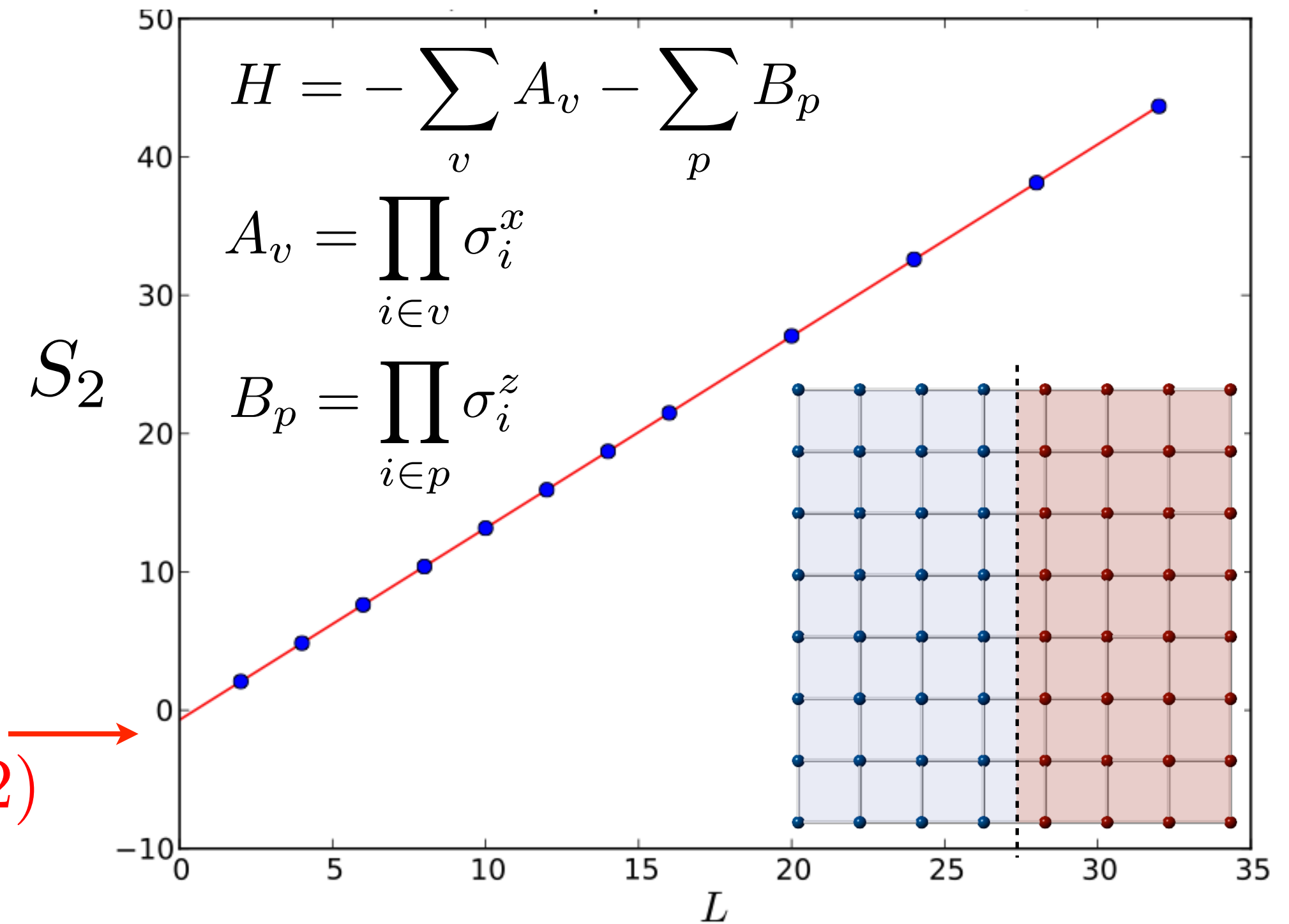
E.g. topological entanglement entropy



possible boundary configurations

$$\Omega \sim 2^{\ell-1}$$

$$\gamma = -\ln(2)$$

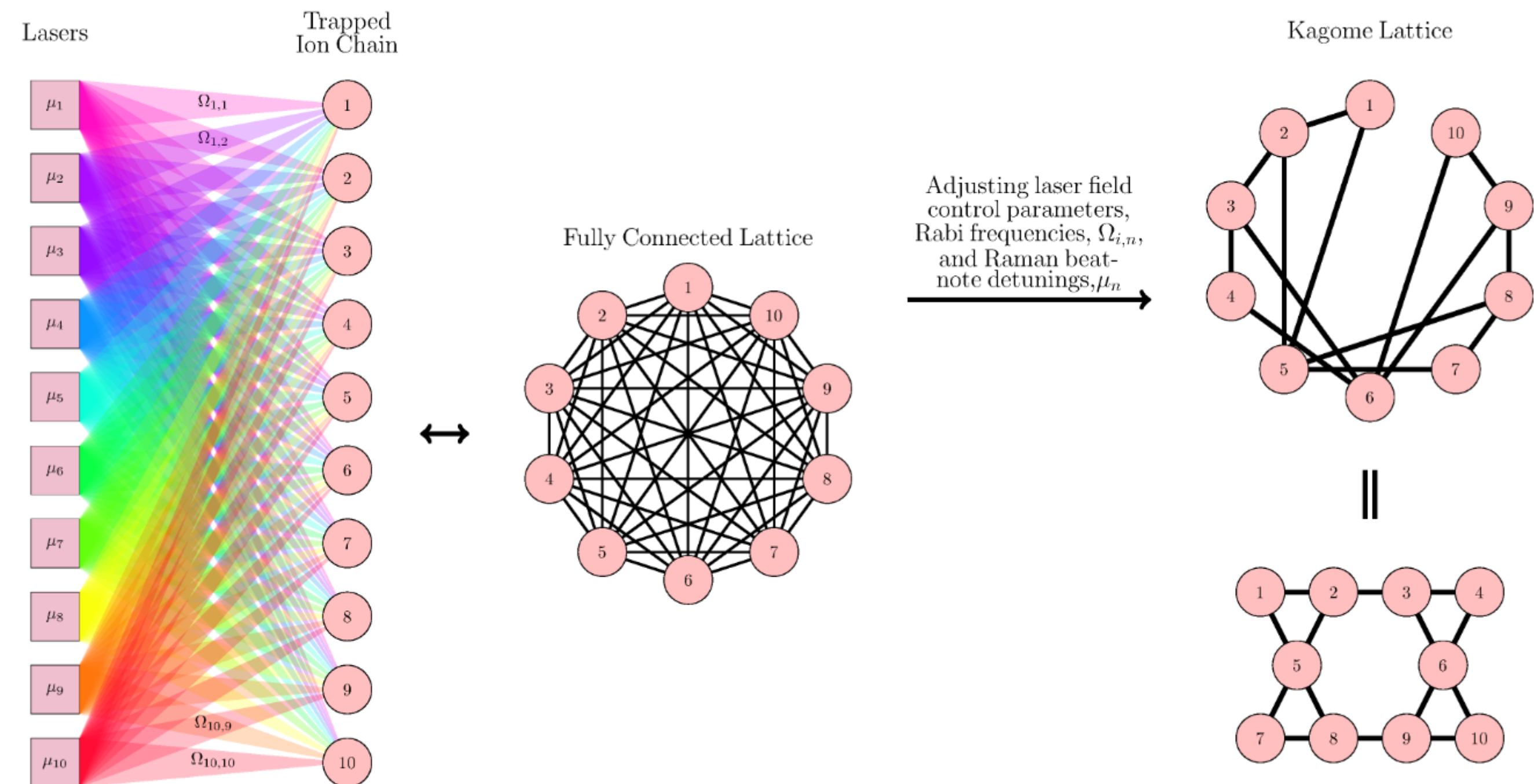
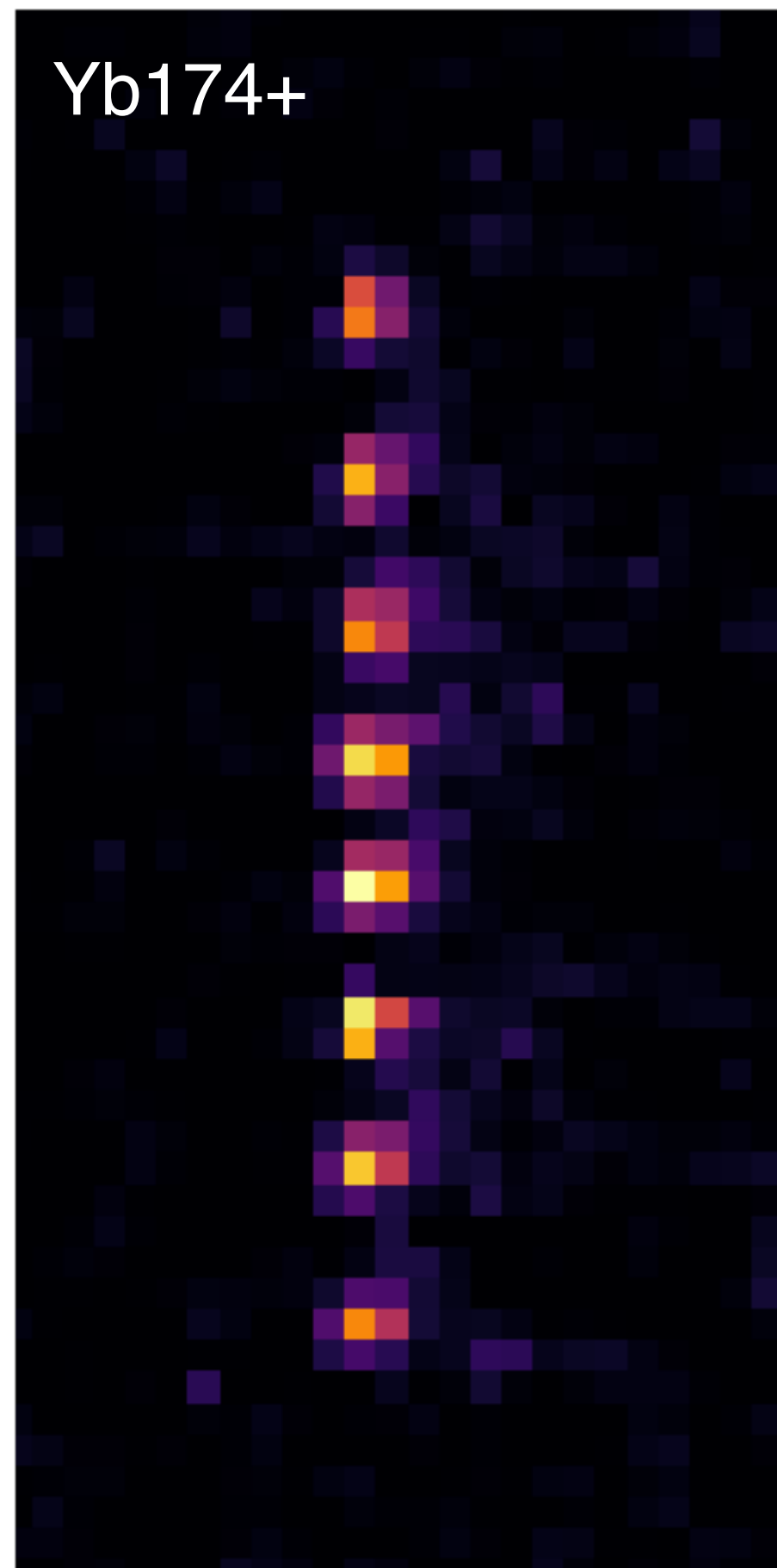


$$S_n = a\ell - \gamma$$

$$\gamma = \ln(2)$$

Trapped ions

“Machine learning design of a trapped-ion quantum spin simulator”,
Yi Hong Teoh , Marina Drygala, Roger G. Melko, and Rajibul Islam arXiv:1910.02496



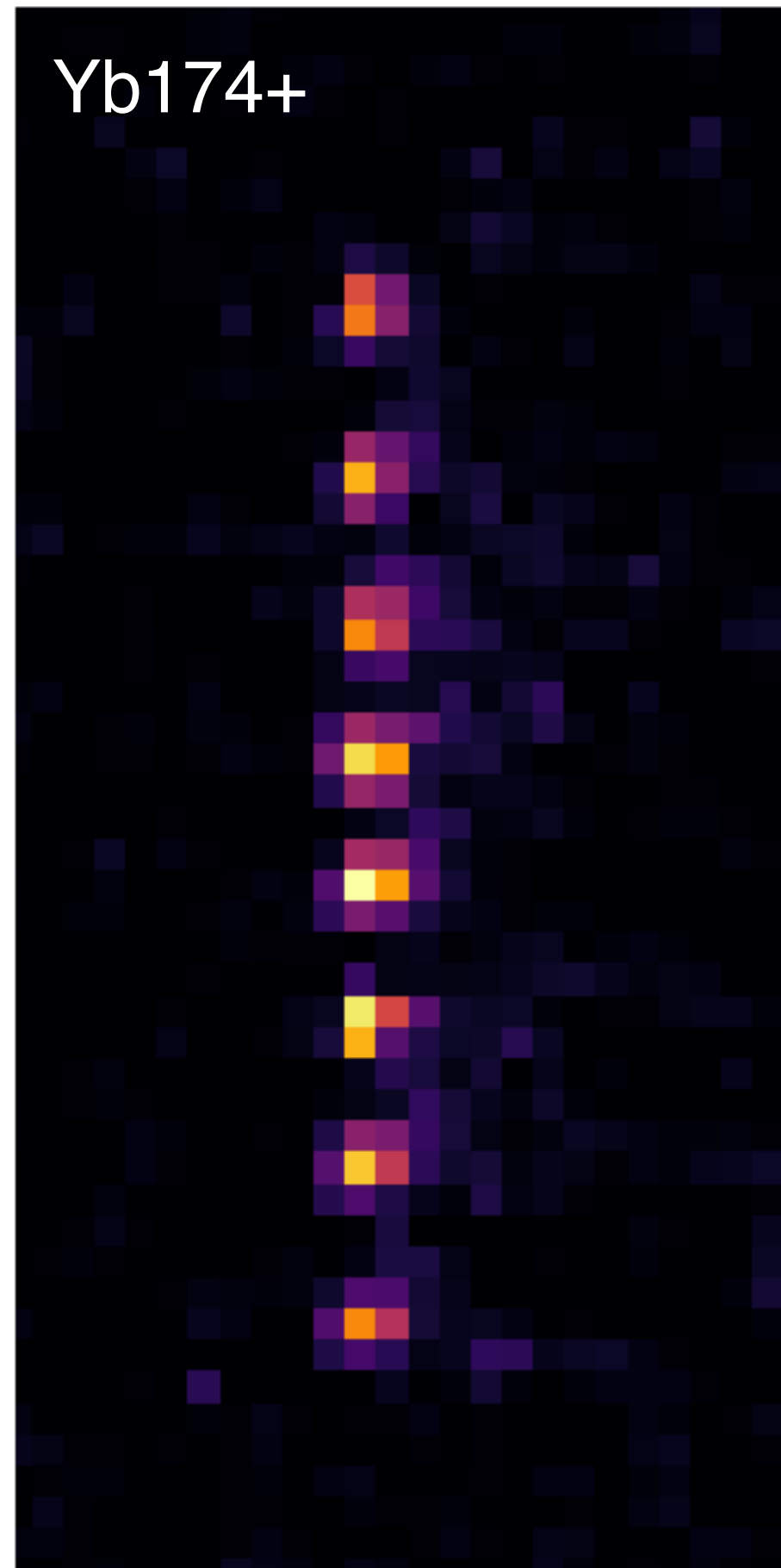
Forward problem

Inverse problem

$$J_{ij}(\Omega_{i,n}, \mu_n) = \sum_n^N \Omega_{i,n} \Omega_{j,n} \sum_m^N \frac{\eta_{i,m} \eta_{j,m} \omega_m}{\mu_n^2 - \omega_m^2}$$

Trapped ions

“Machine learning design of a trapped-ion quantum spin simulator”,
Yi Hong Teoh , Marina Drygala, Roger G. Melko, and Rajibul Islam arXiv:1910.02496



Can produce non-stoquastic Hamiltonians

$$H = \sum_{i,j} J_{ij} S_i^z S_j^z + \sum_{i,j} \tilde{J}_{ij} (S_i^x S_j^x + S_i^y S_j^y)$$

Requires a parameterization with an amplitude and a phase

$$\psi(\mathbf{x}) = e^{i\theta} \sqrt{p(\mathbf{x})}$$

Reconstruction requires measurements in more than one basis

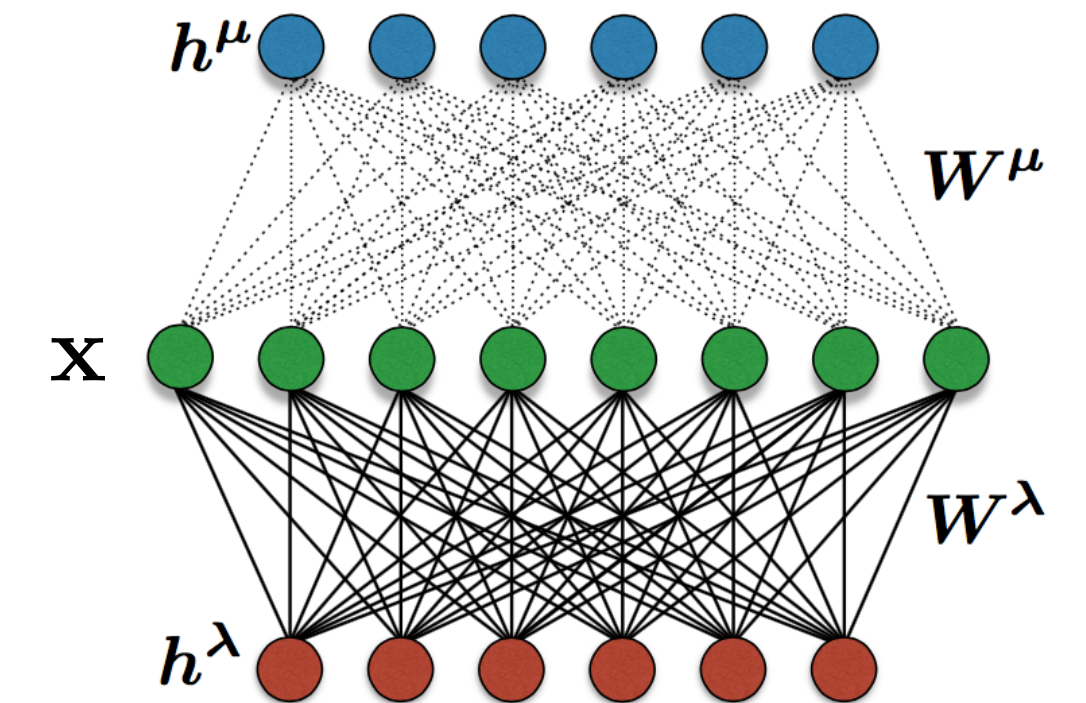
$$R_x(\theta) \quad R_y(\theta)$$

Learning complex wavefunctions

Torlai, Mazzola, Carrasquilla, Troyer, RGM, Carleo,
Nature Physics 14, 447-450 (2018)

For a more generic wavefunction with amplitude and phase, represent both with hidden units

$$\psi_{\lambda,\mu}(\mathbf{x}) \propto \sqrt{p_{\lambda}(\mathbf{x})} e^{i\phi_{\mu}(\mathbf{x})}$$



Now, different bases are needed to estimate both the amplitude and phases of the target state.

$$\mathcal{L} = \sum_b^{N_b} \sum_{\mathbf{x}_b} \log |\psi_{\lambda,\mu}(\mathbf{x}_b)|^2$$

N_b = number of bases

$\{X, X, Z, Z, \dots\}, \{Z, X, X, Z, \dots\}, \{Z, Z, X, X, \dots\},$

state rotated into basis b with the appropriate unitary

From this, calculate $\nabla_{\lambda} \mathcal{L}$ and $\nabla_{\mu} \mathcal{L}$, use stochastic gradient descent, etc.

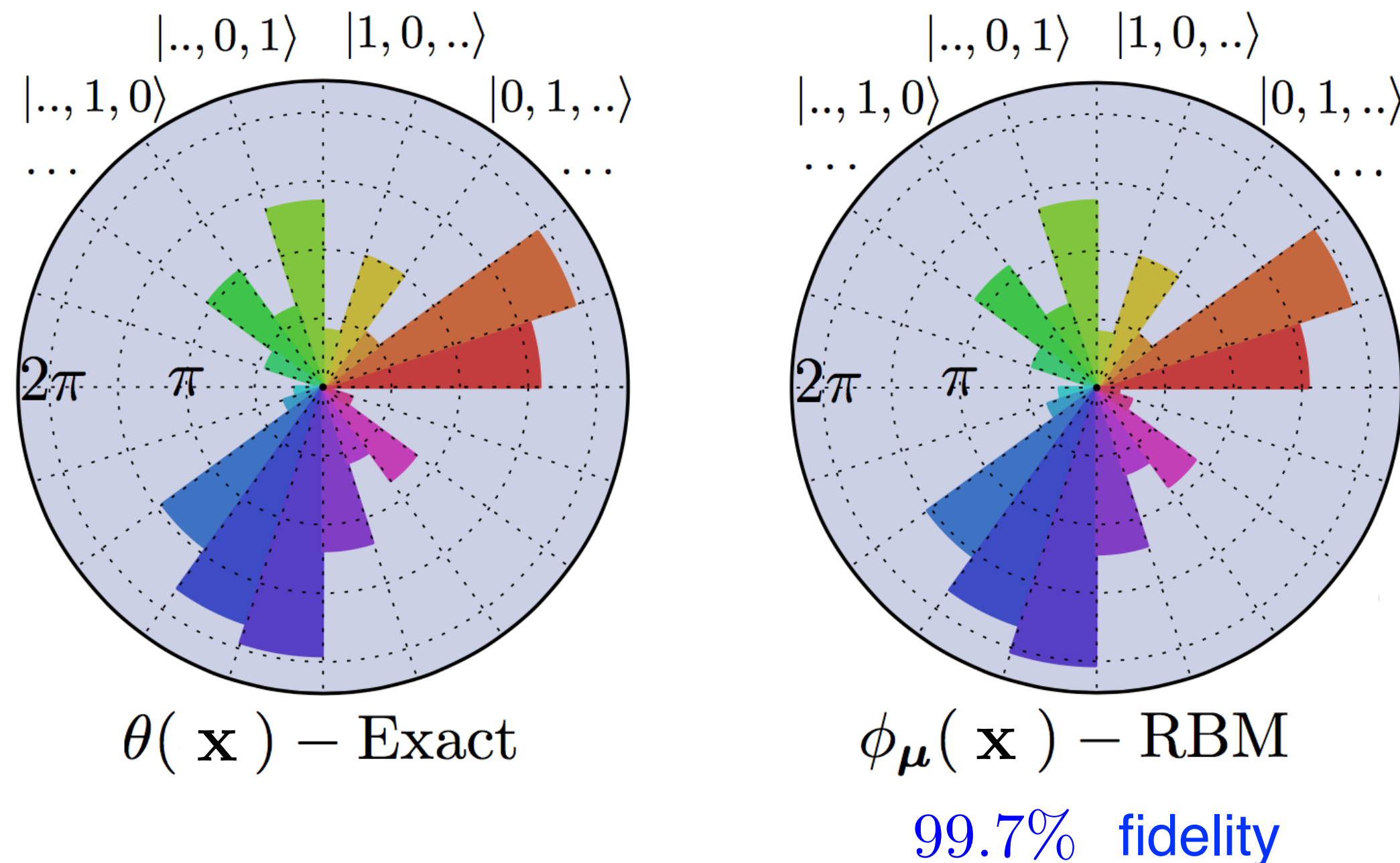
In practice, training is done in two stages: learning of amplitude first, then optimization of the phase parameters.

Learning complex wavefunctions

Torlai, Mazzola, Carrasquilla, Troyer, RGM, Carleo,
Nature Physics 14, 447-450 (2018)

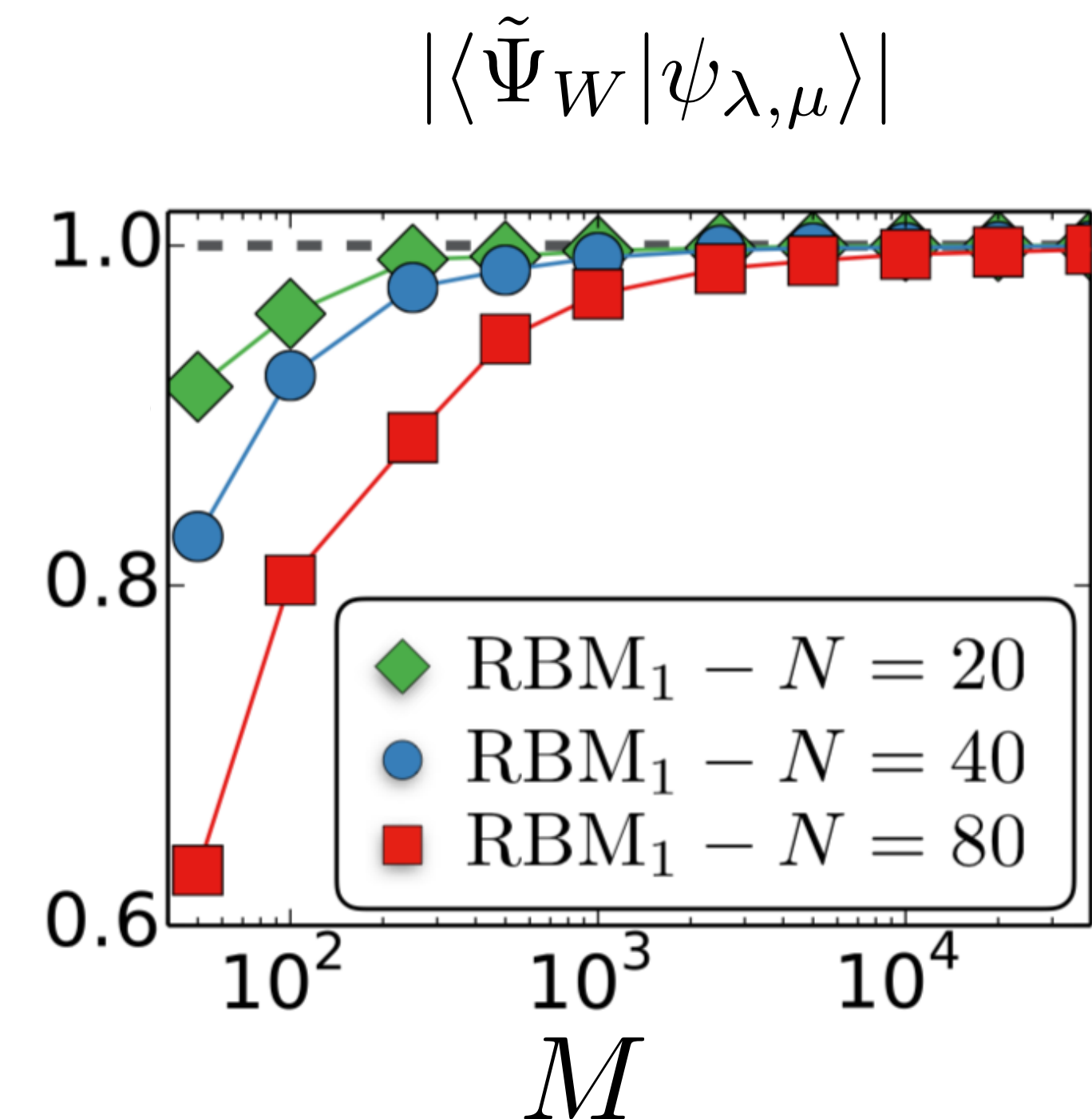
N-qubit “W” state, with local random phase shifts applied

$$|\tilde{\Psi}_W\rangle = \frac{1}{\sqrt{N}} \left(e^{i\theta_1} |100\dots\rangle + e^{i\theta_2} |010\dots\rangle + \dots + e^{i\theta_N} |0\dots 01\rangle \right)$$



$N = 20$ qubits

$M = 2 \times 10^5$ measurements



Number of bases used for training: $N_b \propto 2N$

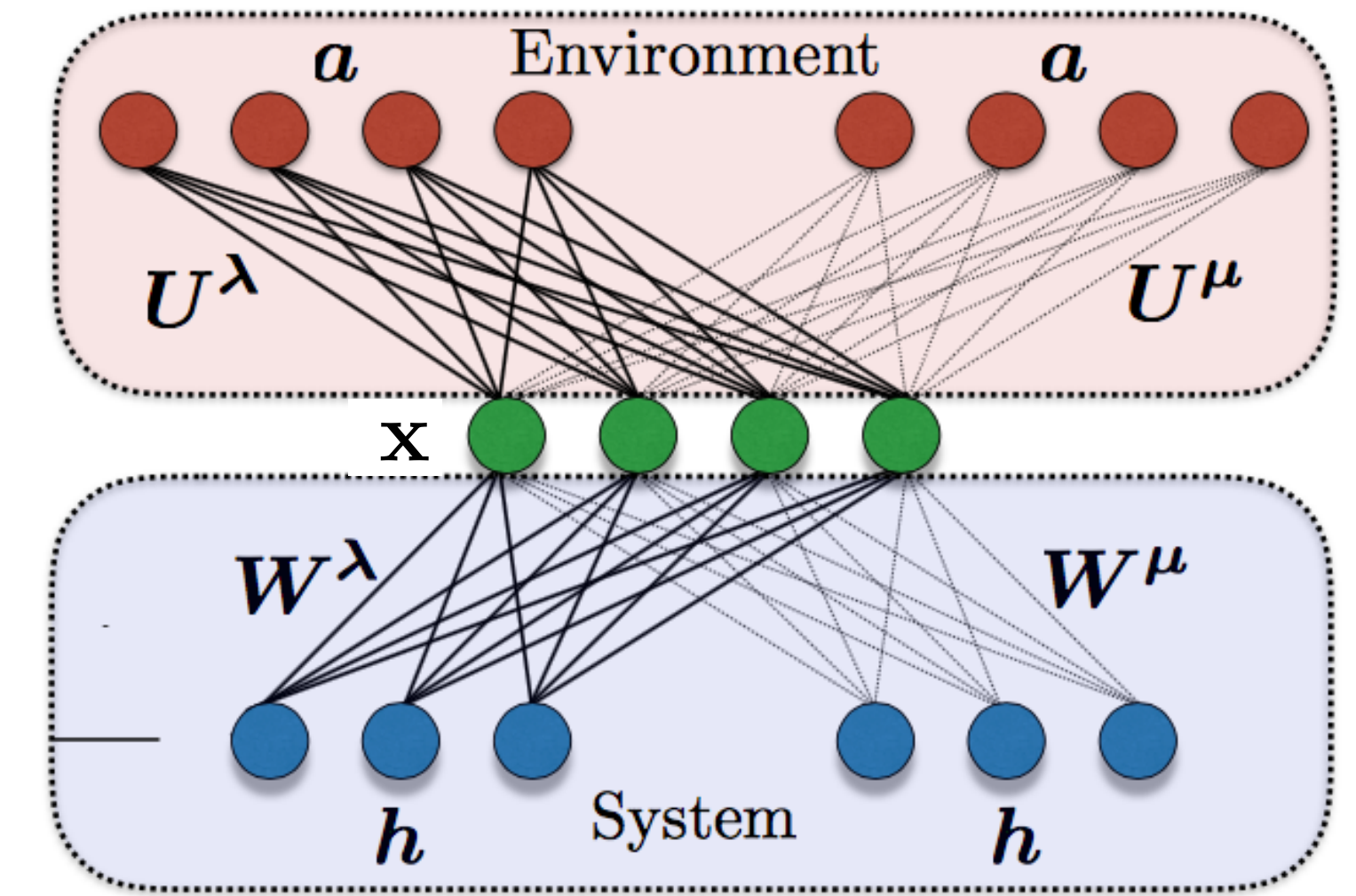
Learning density matrices

Torlai and RGM, Phys. Rev. Lett. 120, 240503 (2018)

Can extend our RBM to represent mixed states described by density matrices

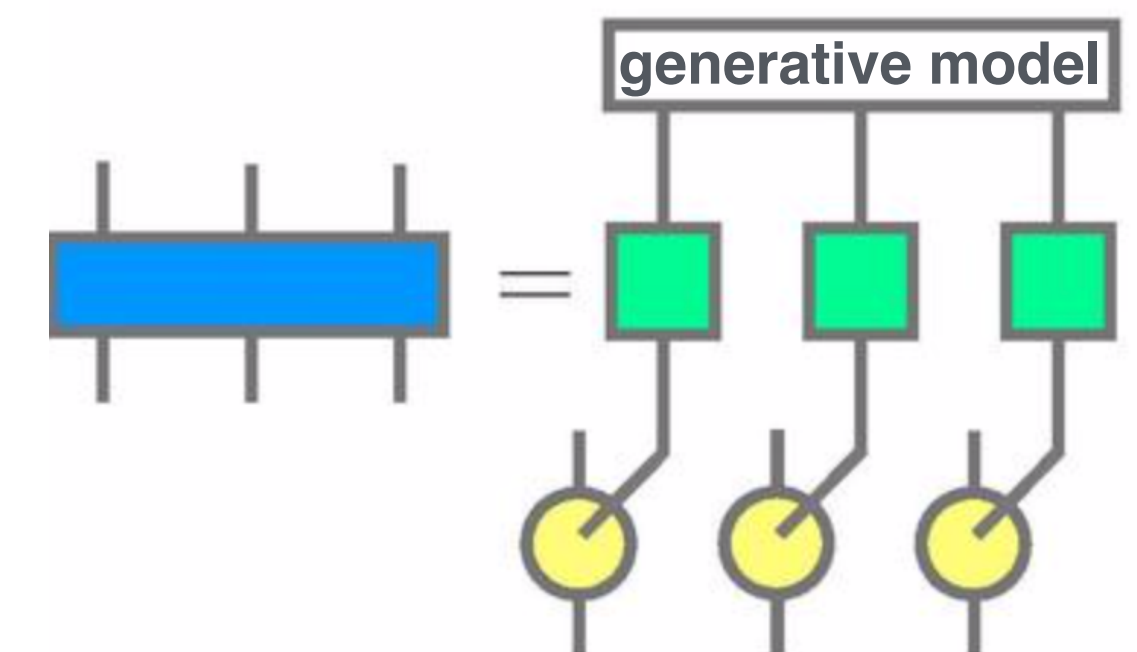
$$\psi_{\lambda,\mu}(\mathbf{x}, \mathbf{a}) \propto \sqrt{p_{\lambda}(\mathbf{x}, \mathbf{a})} e^{i\phi_{\mu}(\mathbf{x}, \mathbf{a})}$$

$$\rho_{\lambda,\mu}(\mathbf{x}, \mathbf{x}') = \sum_{\mathbf{a}} \psi_{\lambda,\mu}(\mathbf{x}, \mathbf{a}) \psi_{\lambda,\mu}^*(\mathbf{x}', \mathbf{a})$$



Can introduce a factorization of the state in terms of a probability distribution and a set of tensors - allows for inversion of the Born rule.

$$p(\mathbf{x}) = \text{Tr} \rho M^{\mathbf{x}}$$



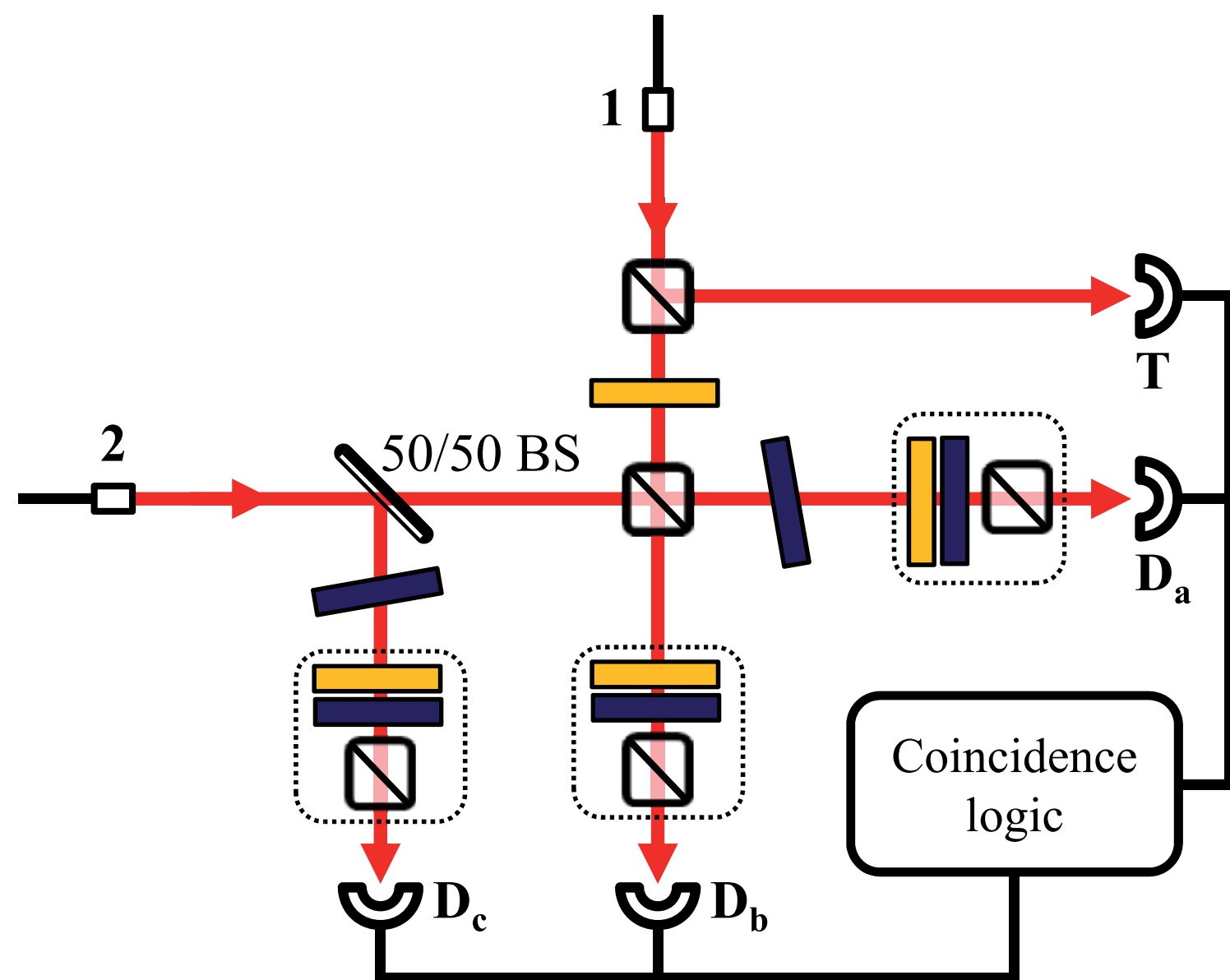
RBM, RNNs, VAEs, GANs, and more...

Carrasquilla, Torlai, RGM, Aolita,
Nature Machine Intelligence 1, 155–161 (2019)

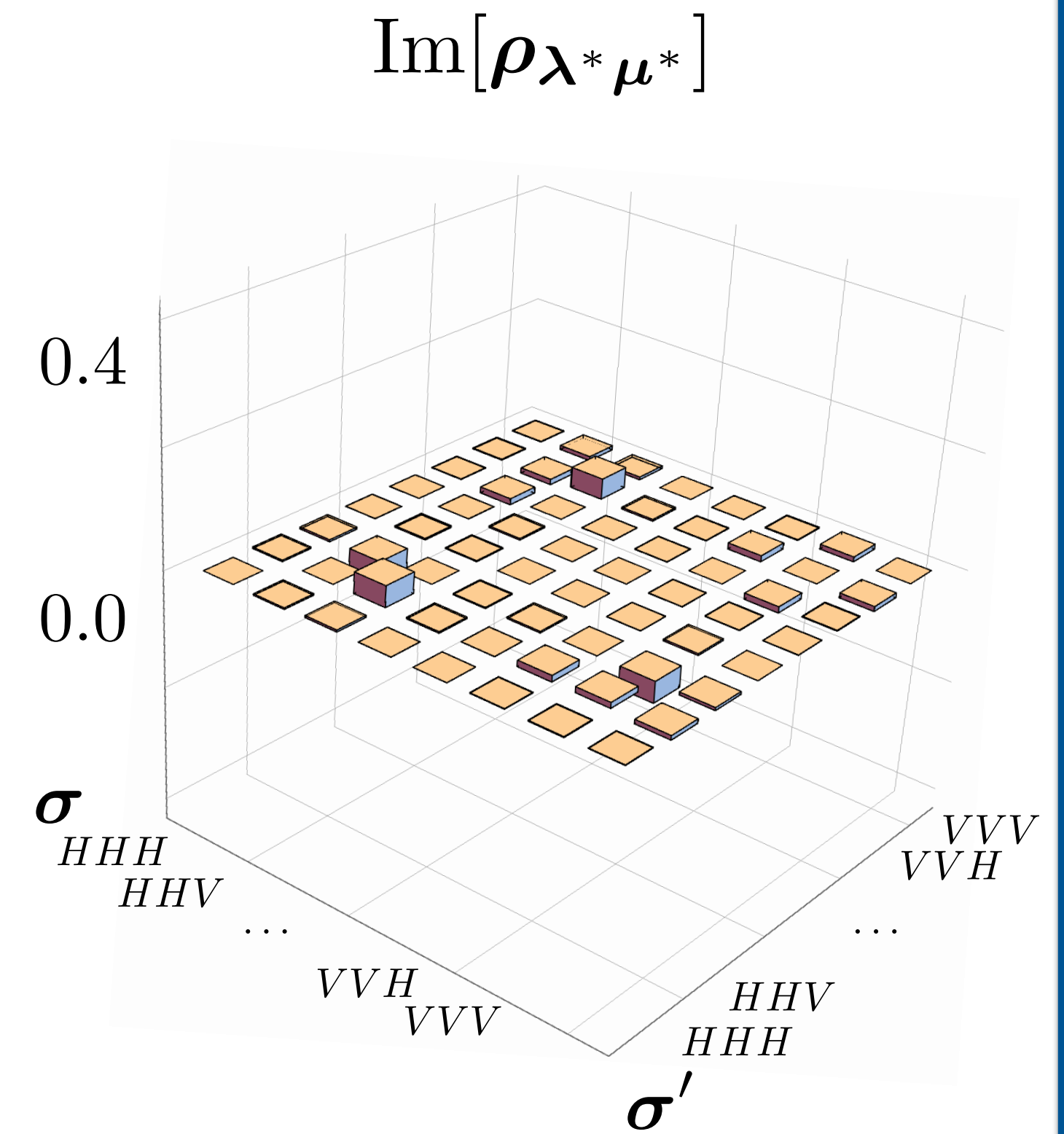
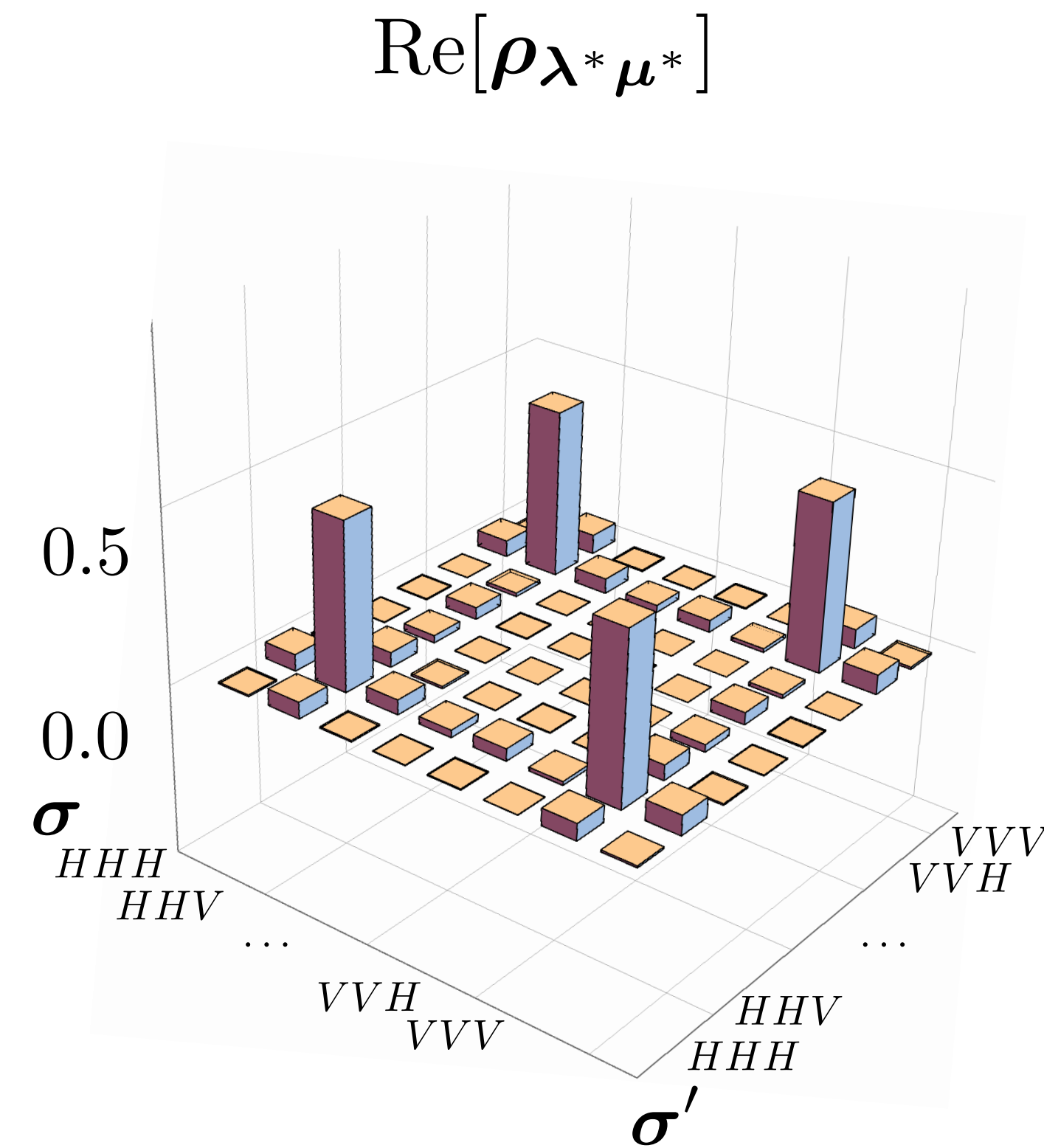
Experiment: entangled photons

G. Torlai

$$|\psi\rangle = \frac{1}{\sqrt{2}} (|HHV\rangle + |VVH\rangle)$$

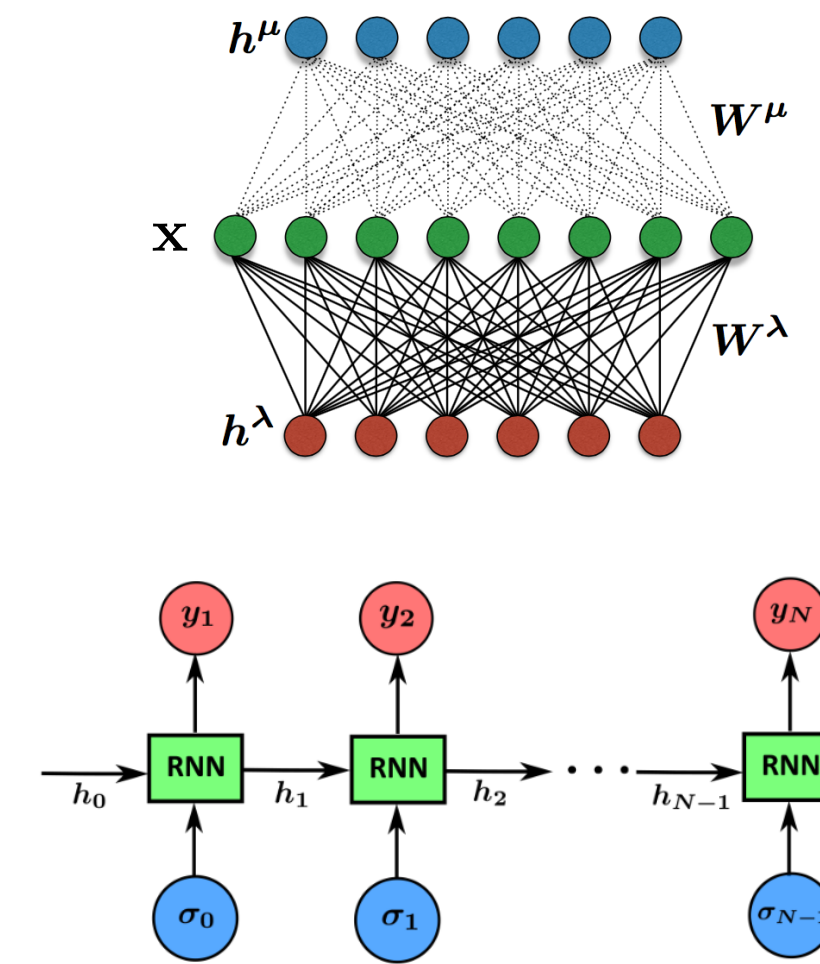


K. Resch lab, IQC Waterloo



Discussion:

- **Generative models** are a powerful tool for performing state reconstruction using projective measurement data



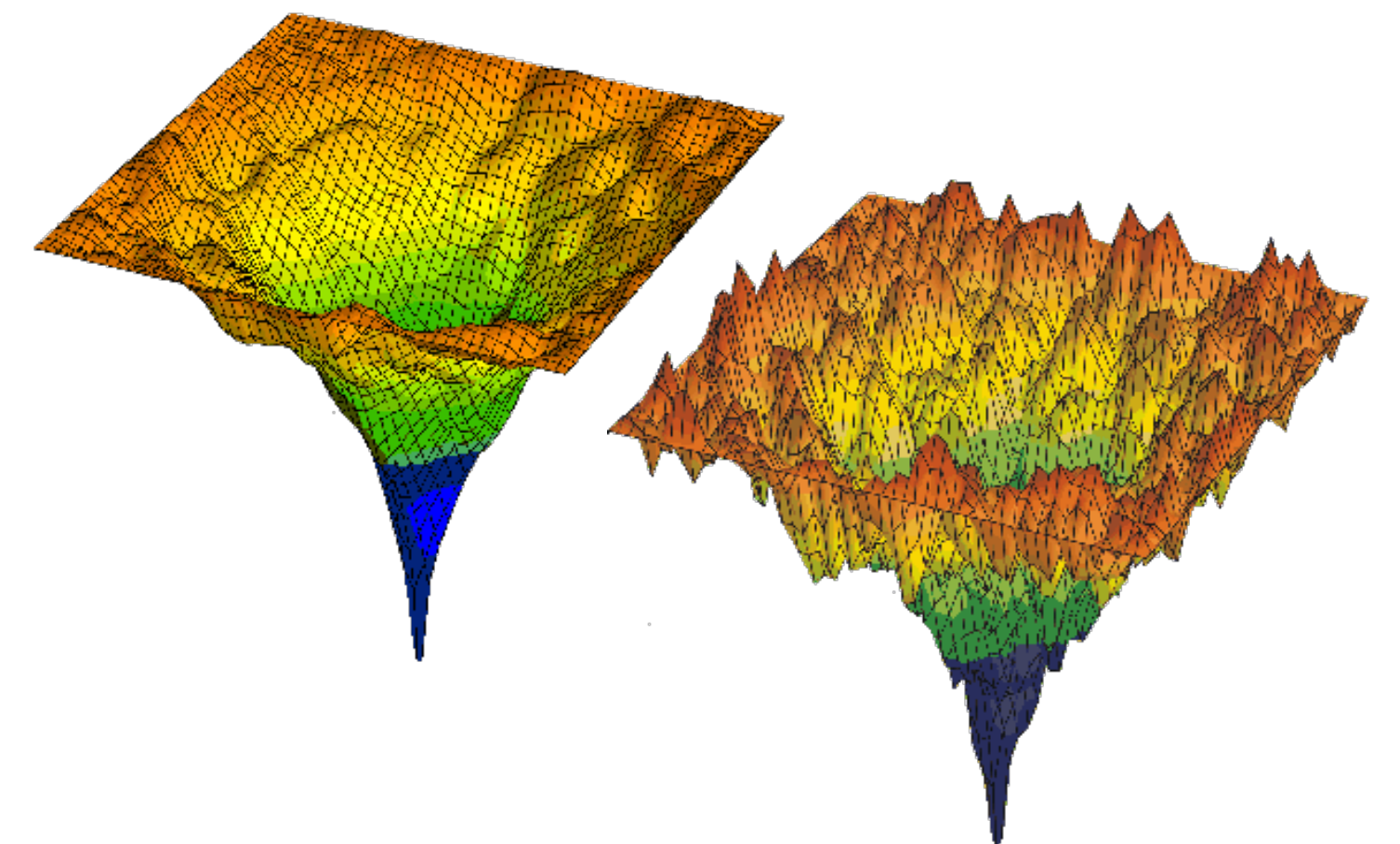
$$n_h \propto \text{poly}(N)$$

$$M \propto \text{poly}(N)$$

$$N_b \propto \text{poly}(N)$$

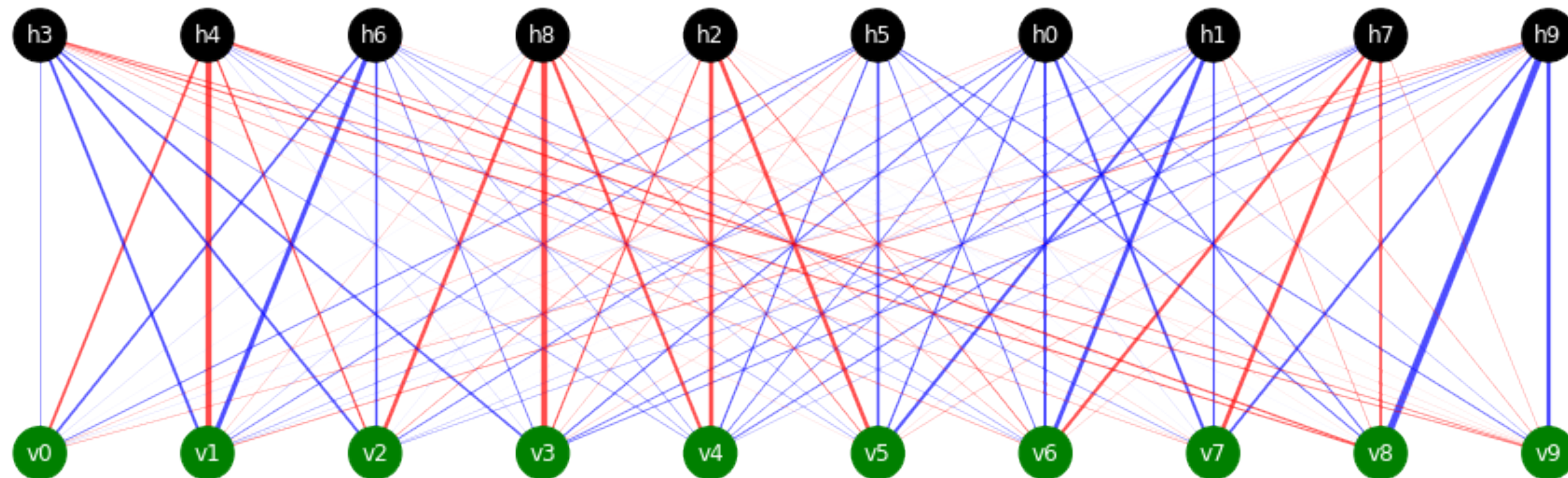
- We don't have a theory relating the “structure” of a quantum state to the **efficiency** of reconstruction

- How is learning from data **different** than optimizing parameters variationally (with knowledge of the Hamiltonian)?



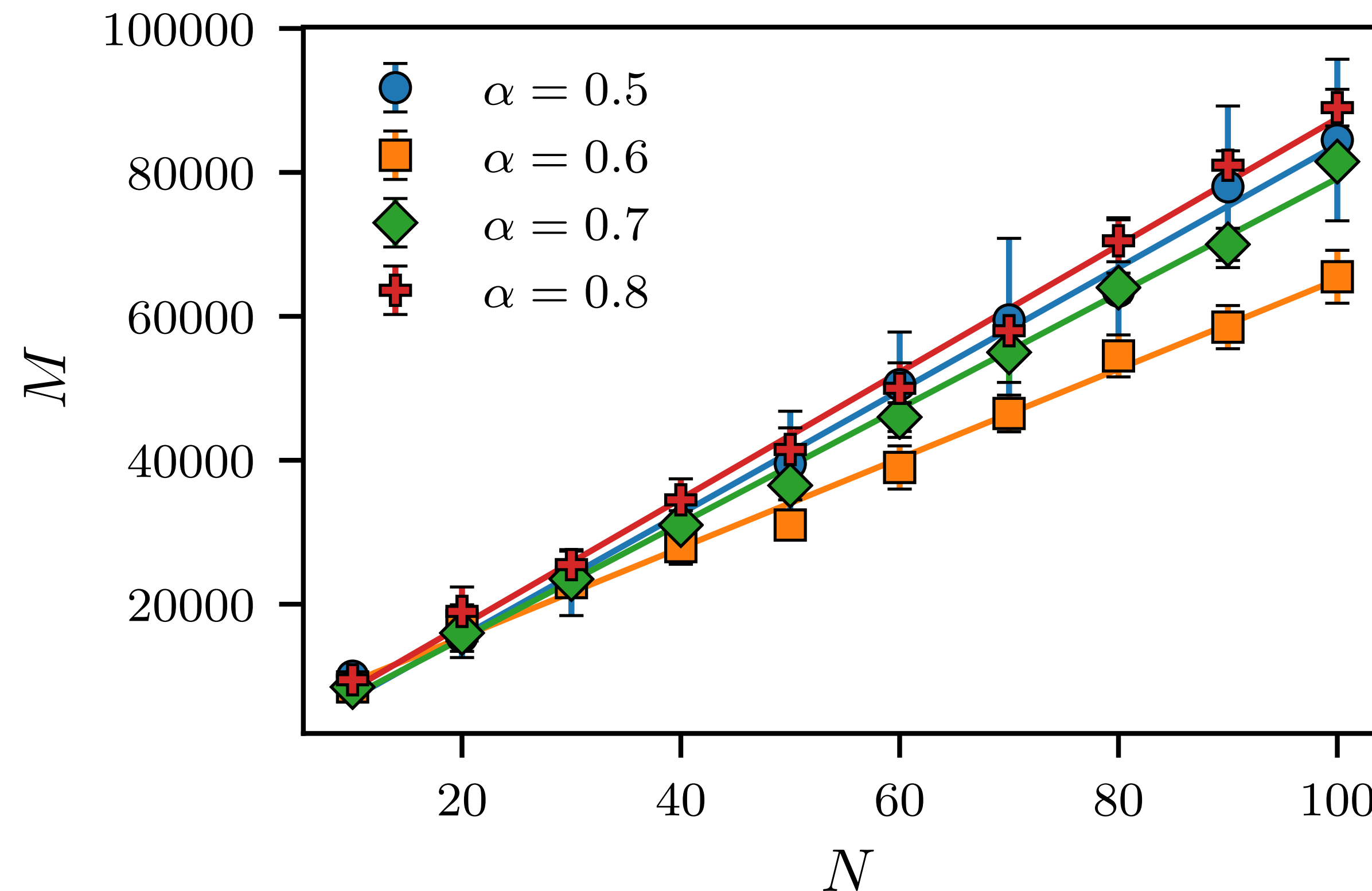
Pruning the trained weights

Sehayek, Golubeva, RGM ... unpublished



- Minimum number of *measurements* to achieve a give accuracy in energy?

$$h/J = 1$$



Aaronson :

if one is only concerned about learning a state well enough to predict the outcomes of most measurements drawn from it, the exponential cost usually associated with full state tomography is reduced to a linear scaling in N .