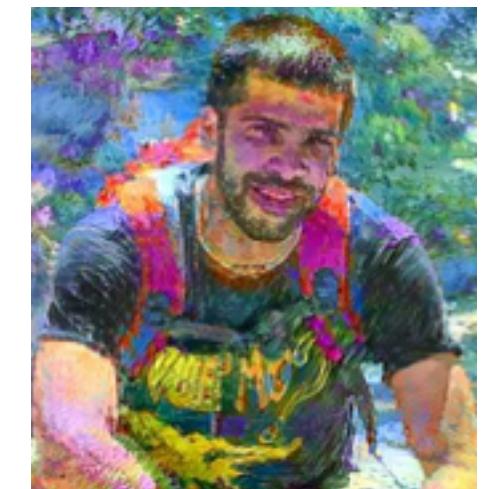


Generative models for (many-body) wavefunction reconstruction

Roger G. Melko



Juan Carrasquilla



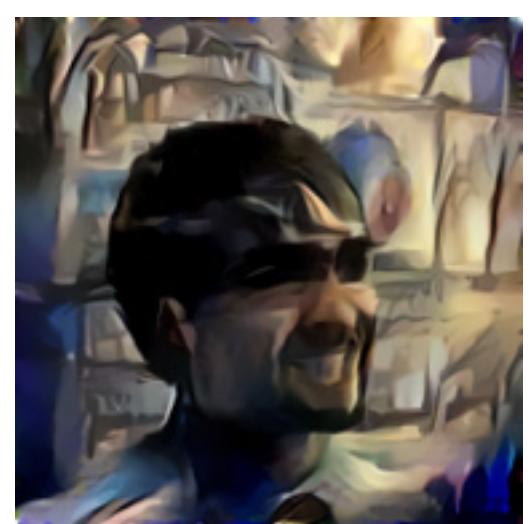
Giacomo Torlai



Bohdan Kulchytskyy



Anna Golubeva



Matt Beach

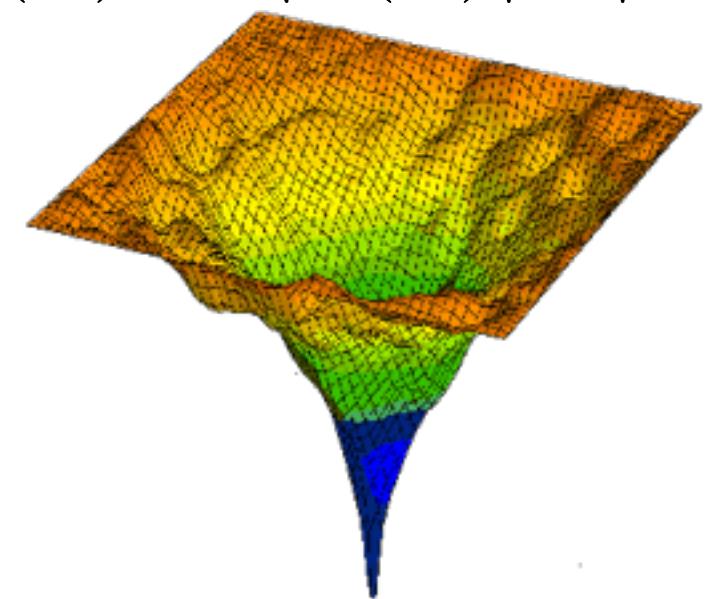
• • •

The landscape of many-body complexity

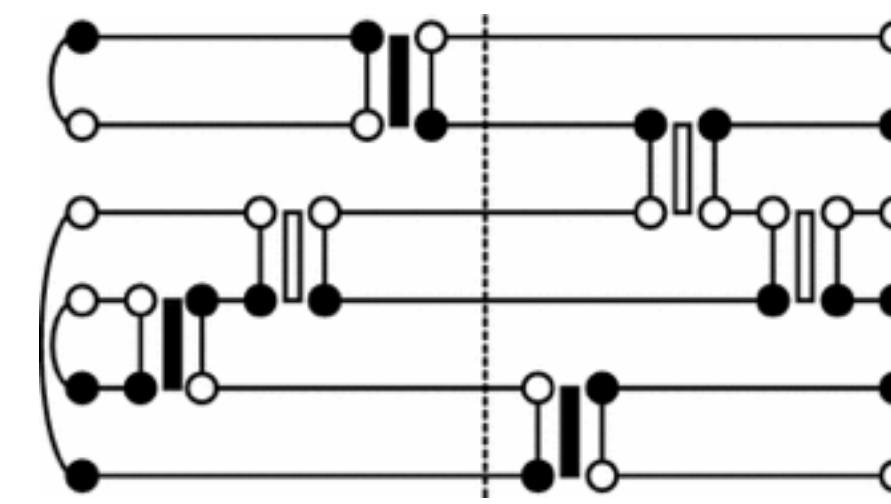
- Many body systems are typically “solved” or “simulated”. E.g. given a **Hamiltonian**, we may seek to find a ground state, or its **correlation functions**

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

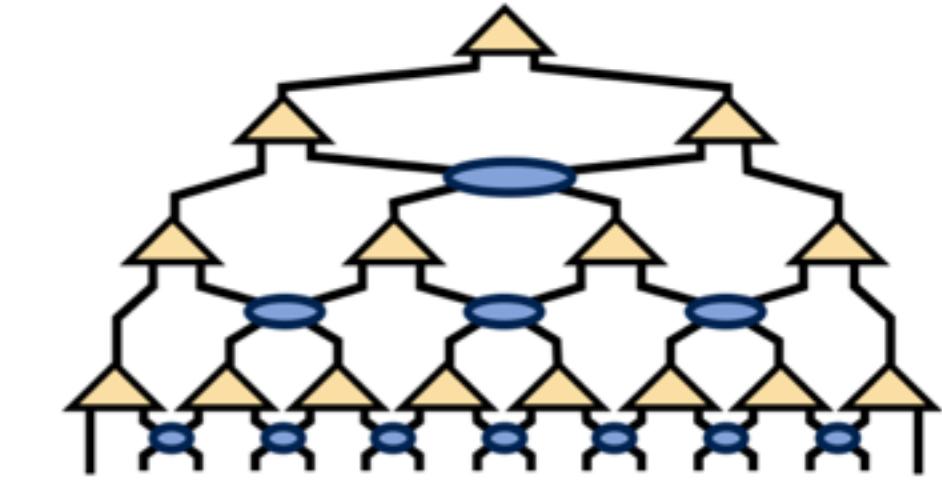
$$\epsilon(\alpha) = \langle \psi(\alpha) | H | \psi(\alpha) \rangle$$



ergodicity



sign problem



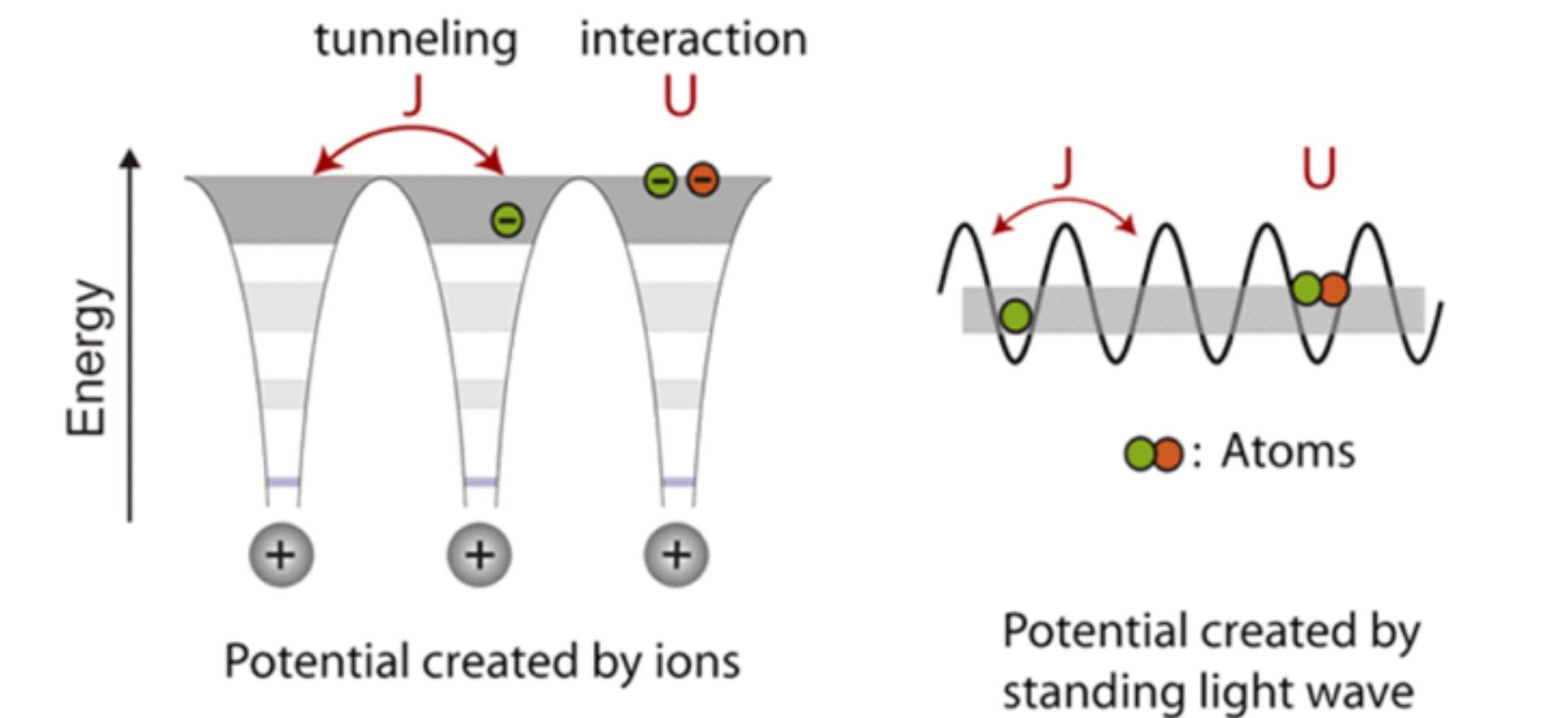
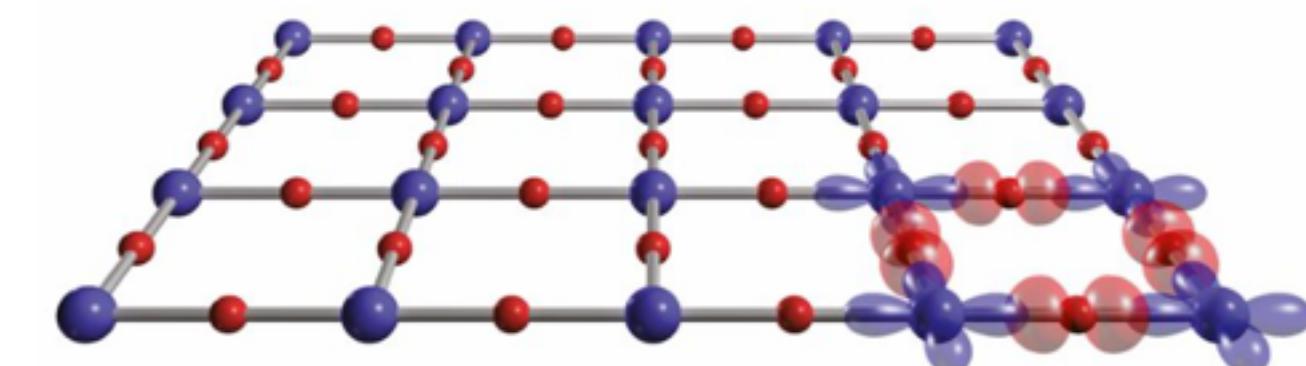
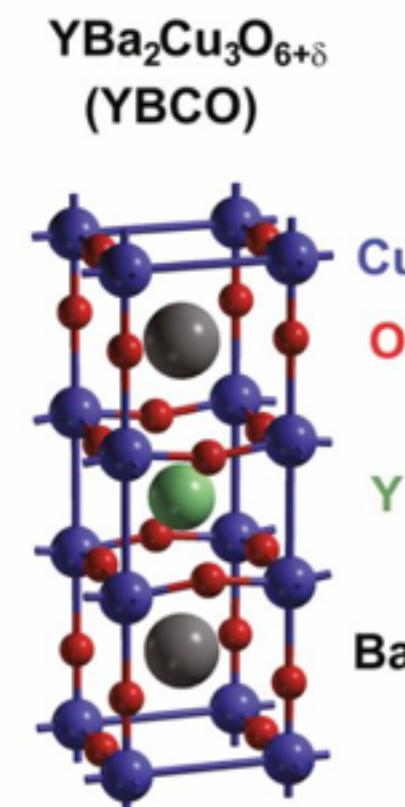
entanglement scaling

Example: fermionic Hubbard model

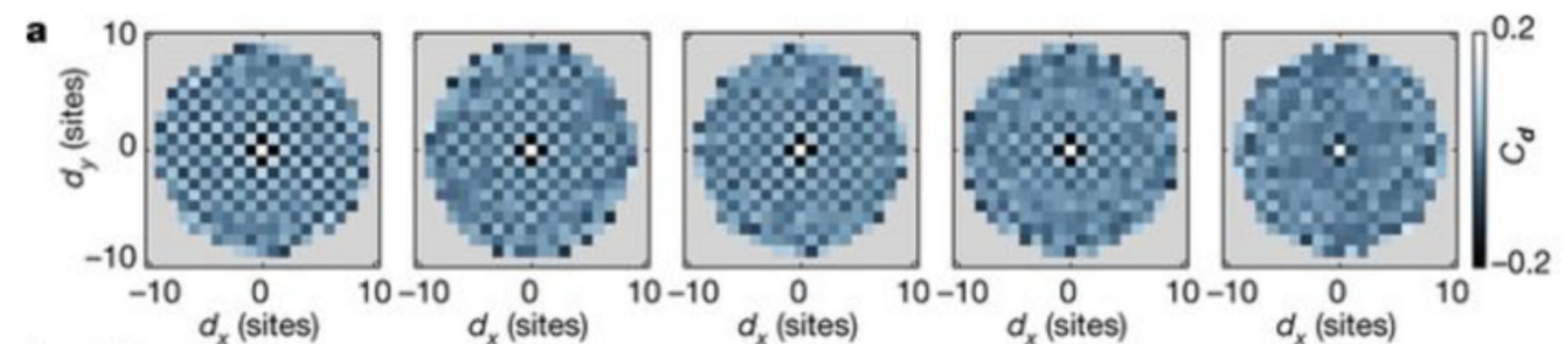
$$\hat{H} = -t \sum_{\langle \mathbf{j}, \mathbf{l} \rangle \sigma} \left(c_{\mathbf{j}\sigma}^\dagger c_{\mathbf{l}\sigma} + c_{\mathbf{l}\sigma}^\dagger c_{\mathbf{j}\sigma} \right) + U \sum_{\mathbf{j}} n_{\mathbf{j}\uparrow} n_{\mathbf{j}\downarrow} - \mu \sum_{\mathbf{j}} (n_{\mathbf{j}\uparrow} + n_{\mathbf{j}\downarrow})$$

The landscape of many-body complexity

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



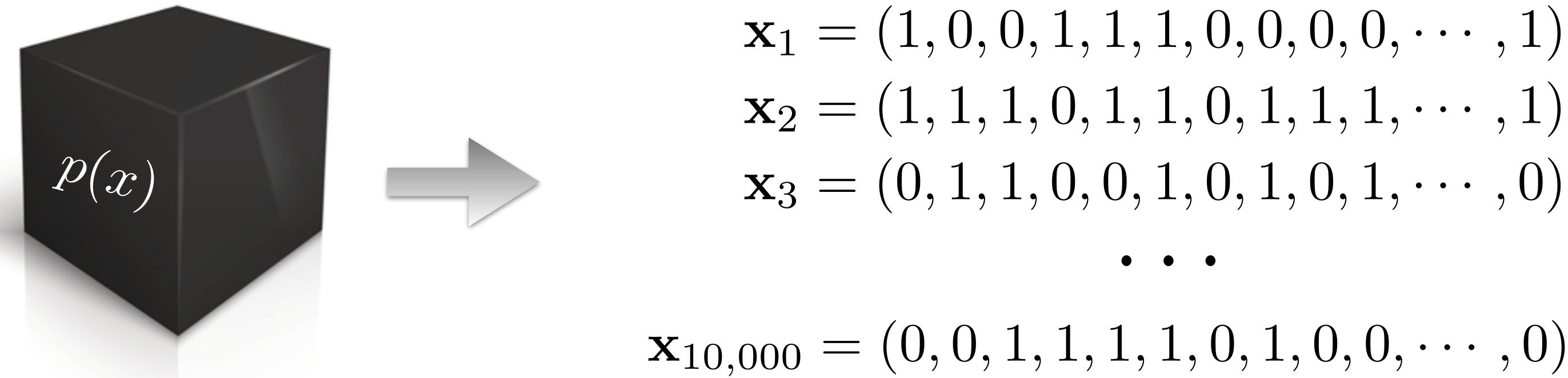
E.g. Mazurenko et. al. Nature 545, 462 (2017)



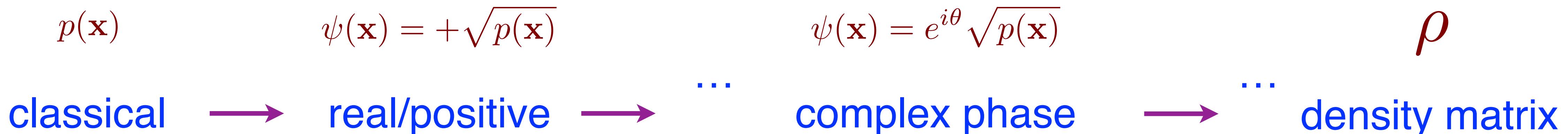
Q: how does our notion of complexity transfer?

Learning (quantum) states from data

- Target projective measurements on a large number of qubits



- Systematically increase the complexity of the learning problem

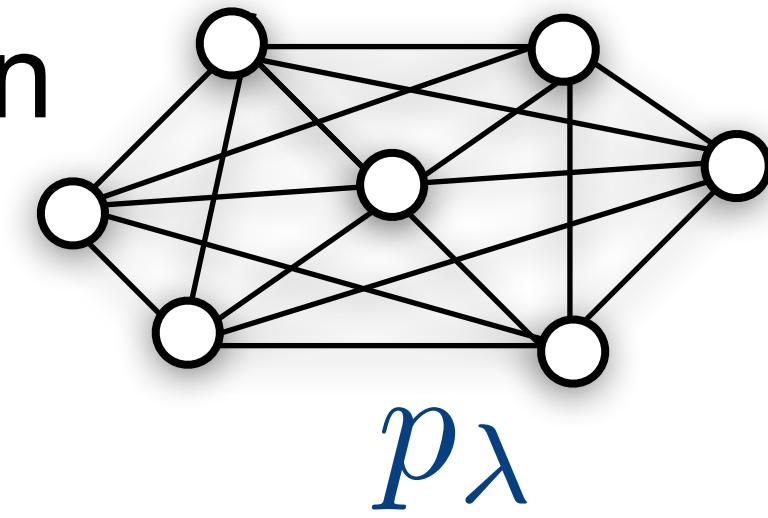


Strategy: probabilistic graphical models

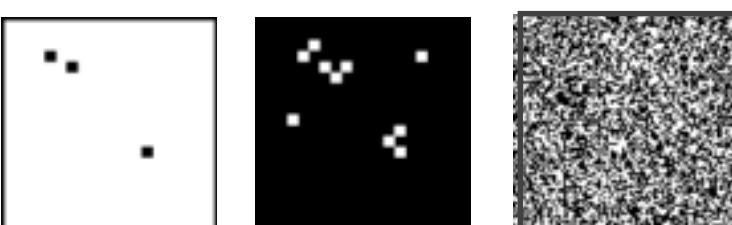
- Goal: take the data and *learn* the best possible approximation

$$p_\lambda(\mathbf{x}) \approx p(\mathbf{x})$$

- Parameterize with a graphical representation of a probability distribution



- Can be used to generate new data samples

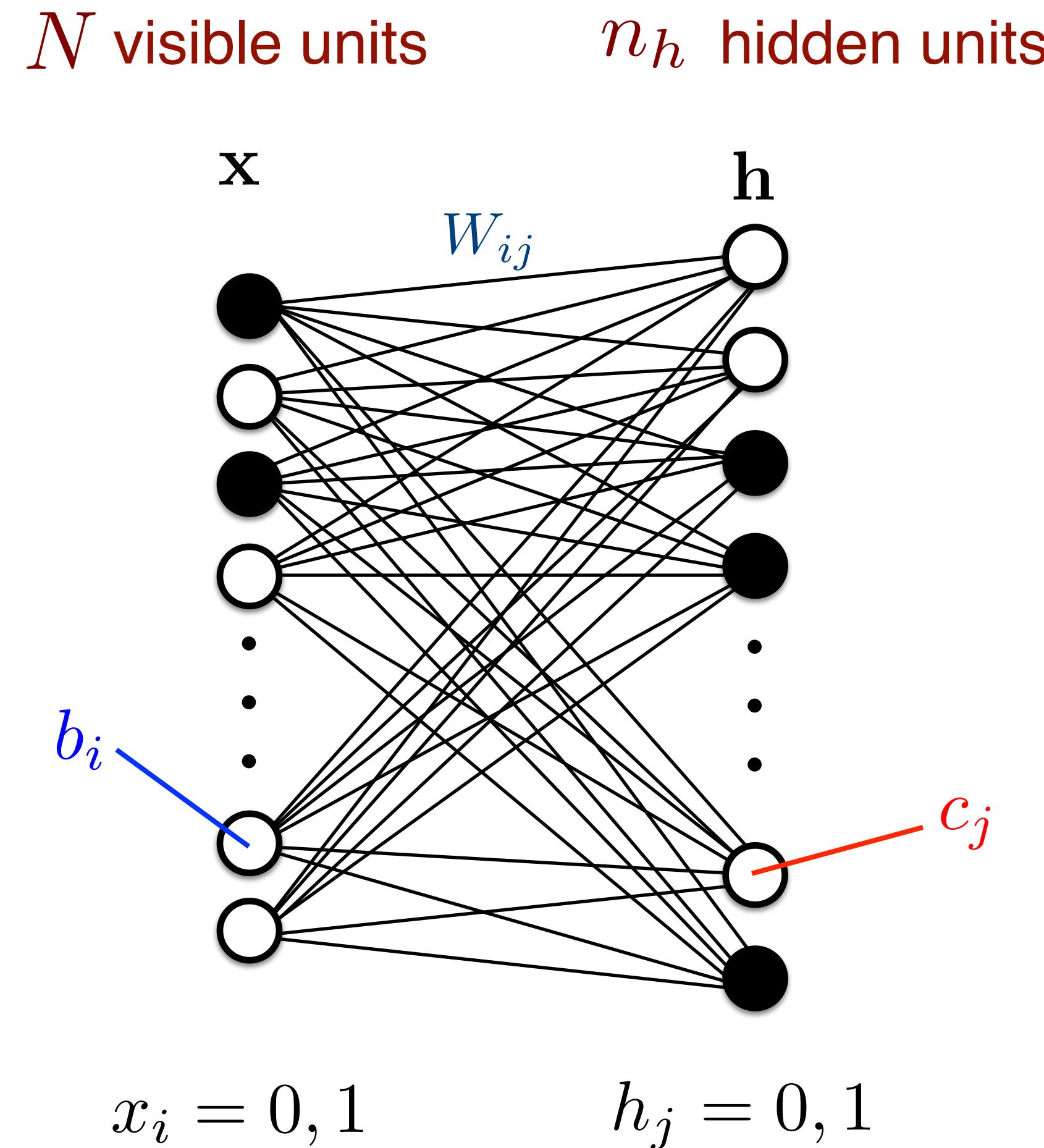


- These can be used e.g. to calculate new estimators

$$\langle \mathcal{O} \rangle = \frac{1}{N_{\text{MCS}}} \sum_{\mathbf{x}} \mathcal{O}_{\mathbf{x}}$$

Restricted Boltzmann Machine

Smolensky, Hinton, Salakhutdinov, Bengio



RBMs are “energy-based” models:

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

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

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

marginal probability distribution

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

Training the RBM

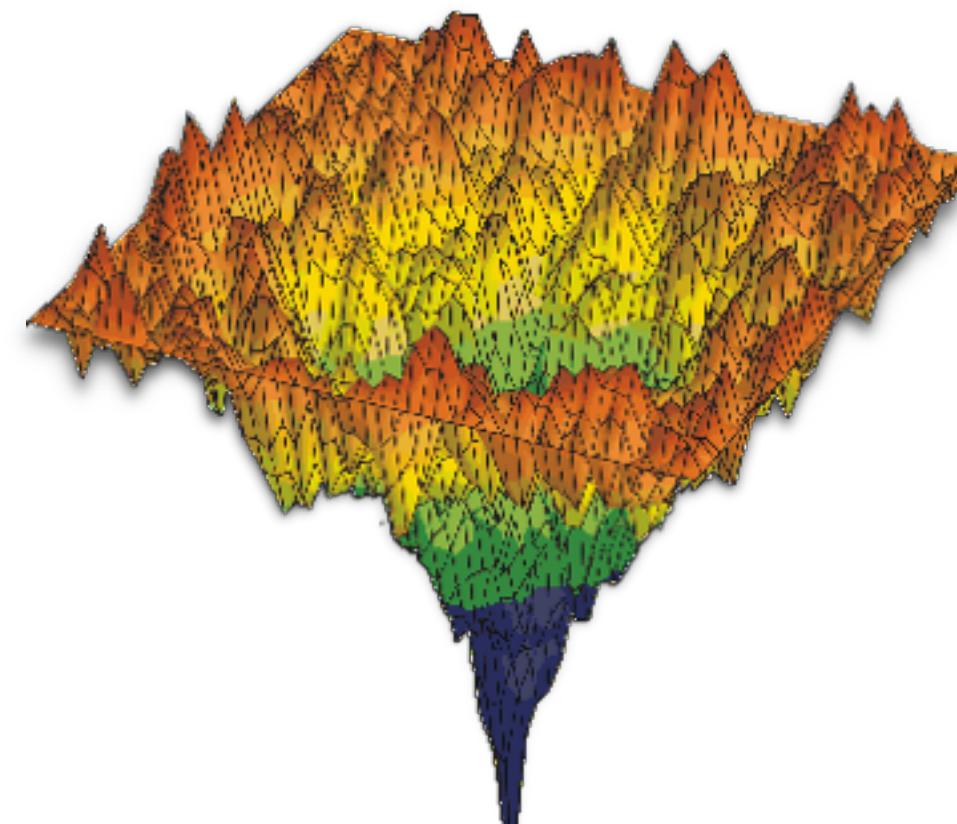
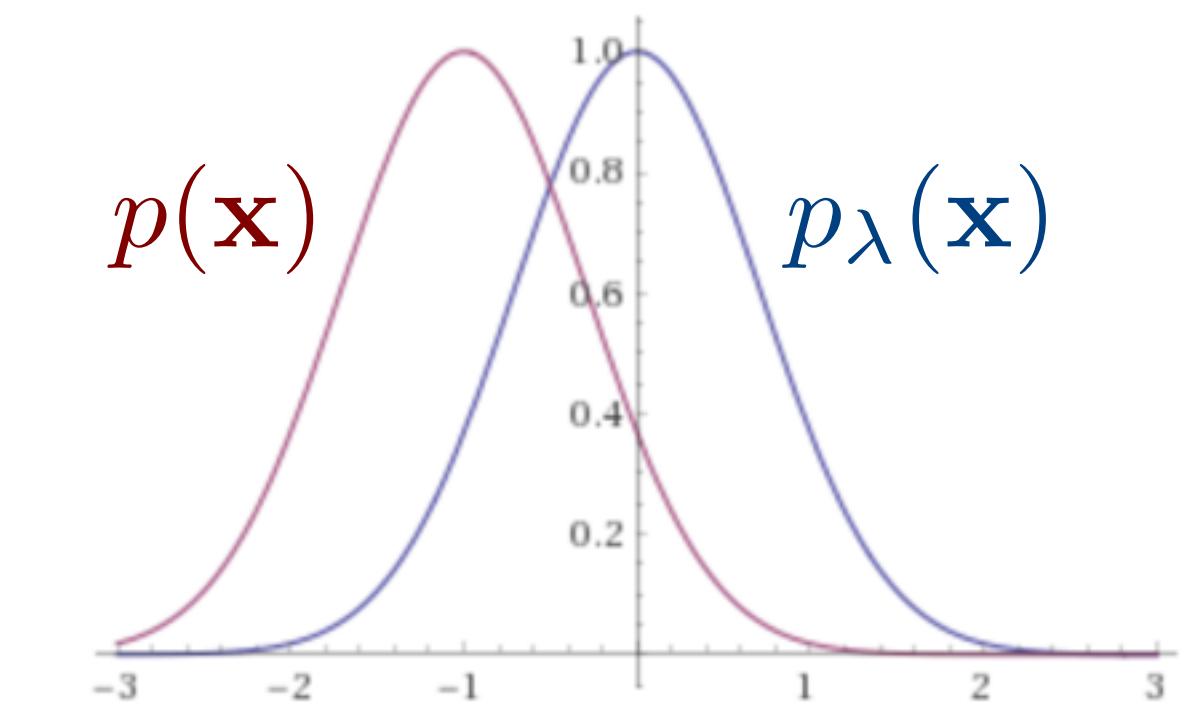
G. Hinton, Neural computation 14, 1771 (2002)

Training means tuning the machine parameters to minimize the difference between the marginal distribution $p_\lambda(\mathbf{x}) = \sum_h p_\lambda(\mathbf{x}, \mathbf{h})$ 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

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

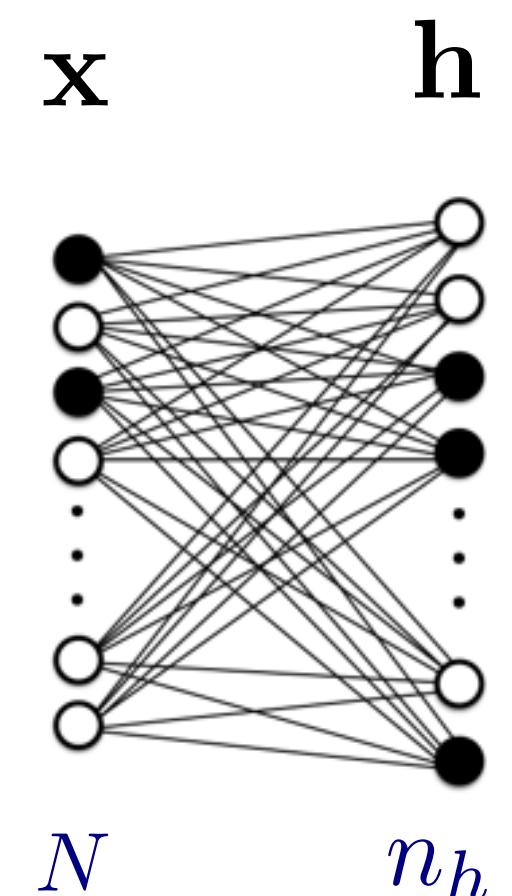
Sampling the RBM

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

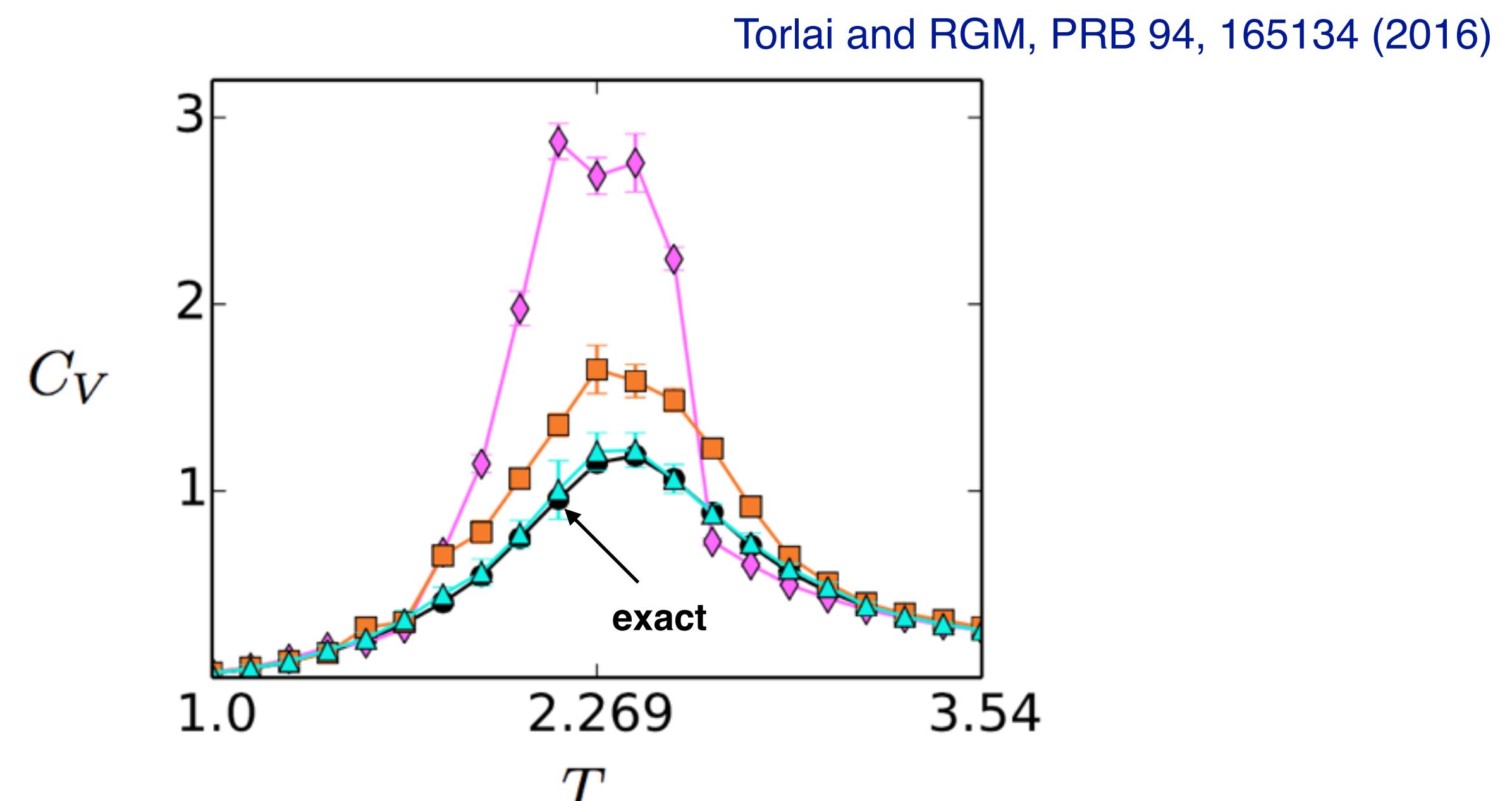
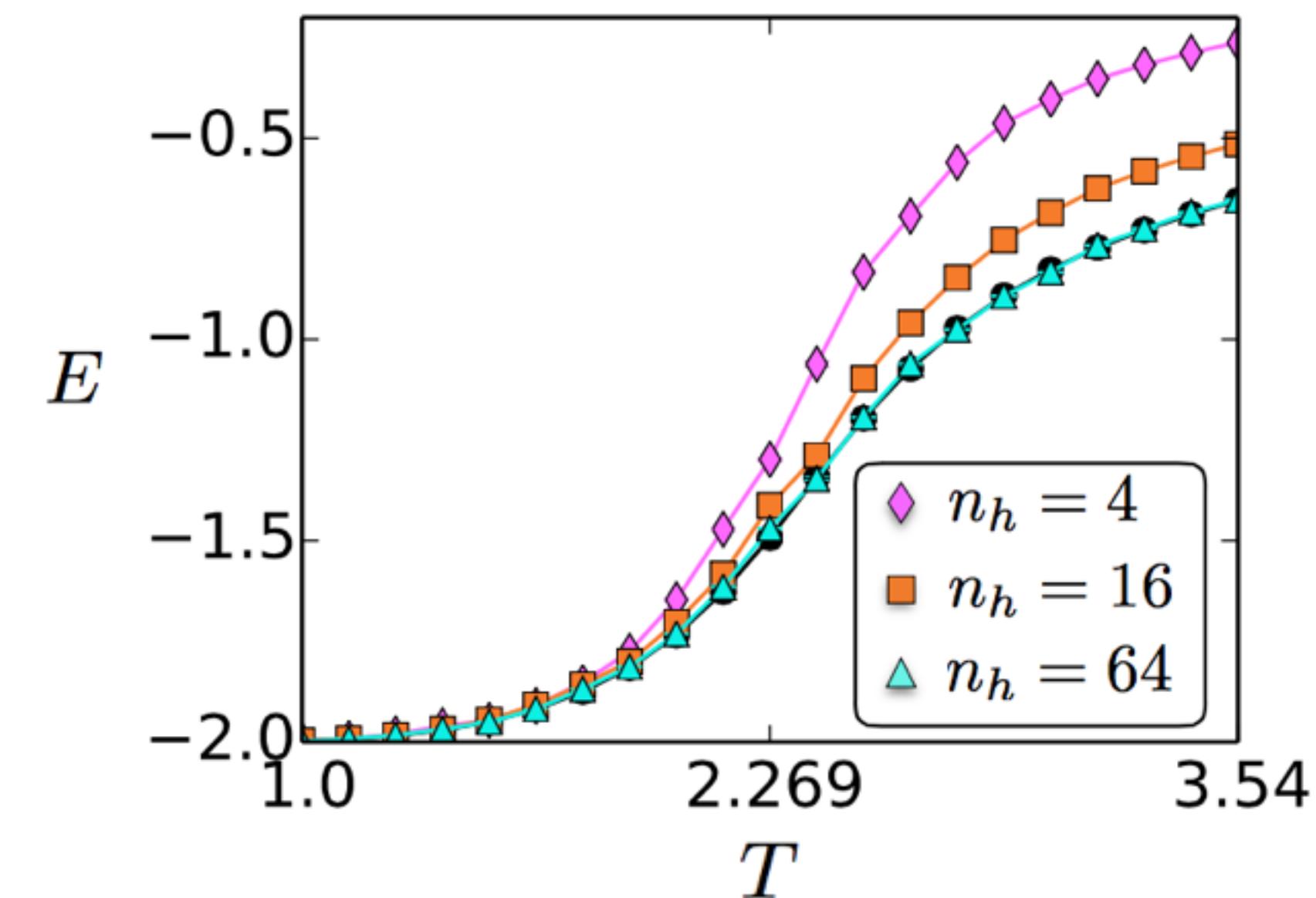
$$\langle \mathcal{O} \rangle_{\text{joint}} = \frac{1}{Z} \sum_x \mathcal{O}_x \left[\sum_h p_\lambda(x, h) \right] \approx \langle \mathcal{O} \rangle_{\text{physical}}$$

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

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



Classical 2D Ising model



Learning (real +ve) wavefunctions

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

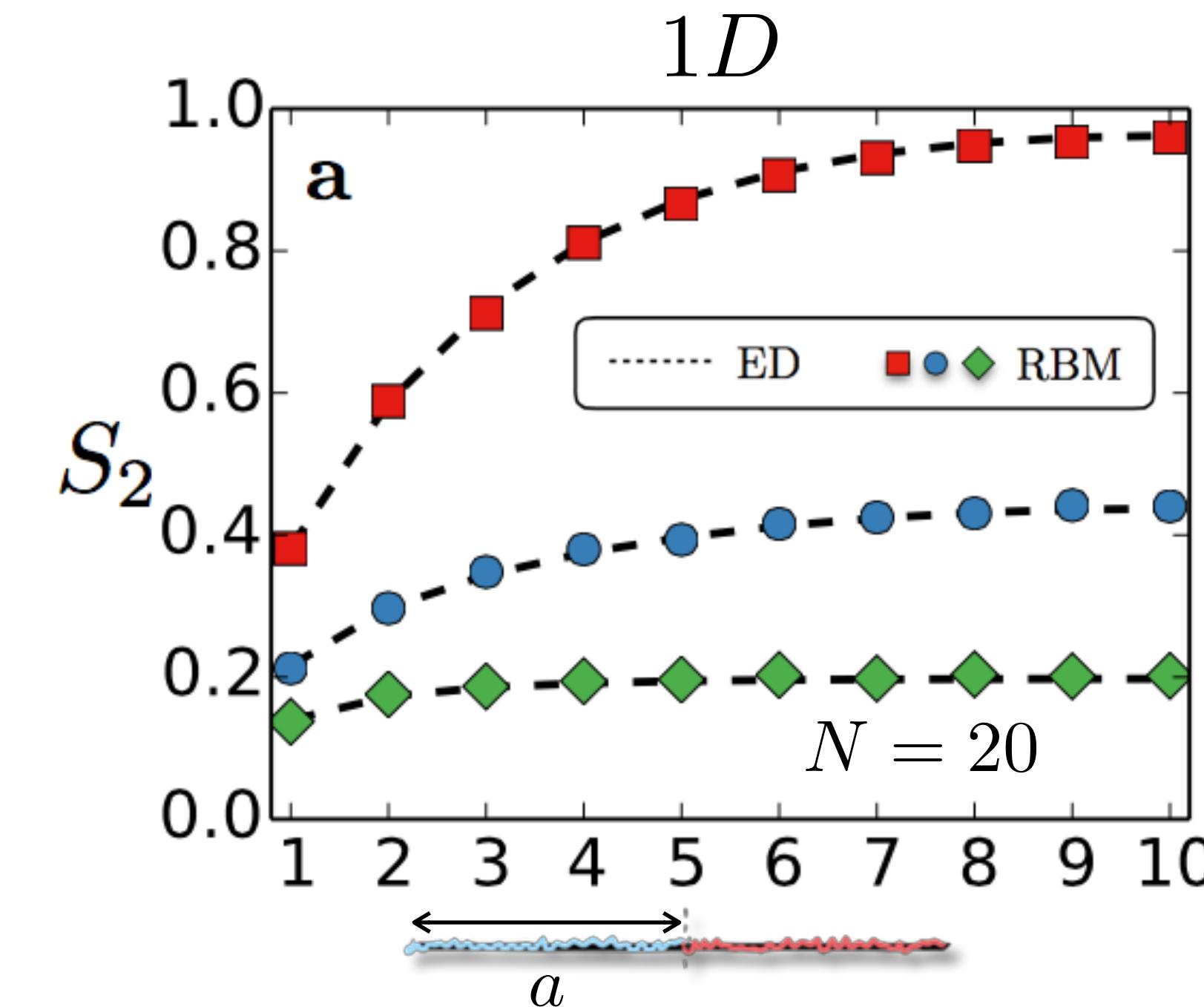
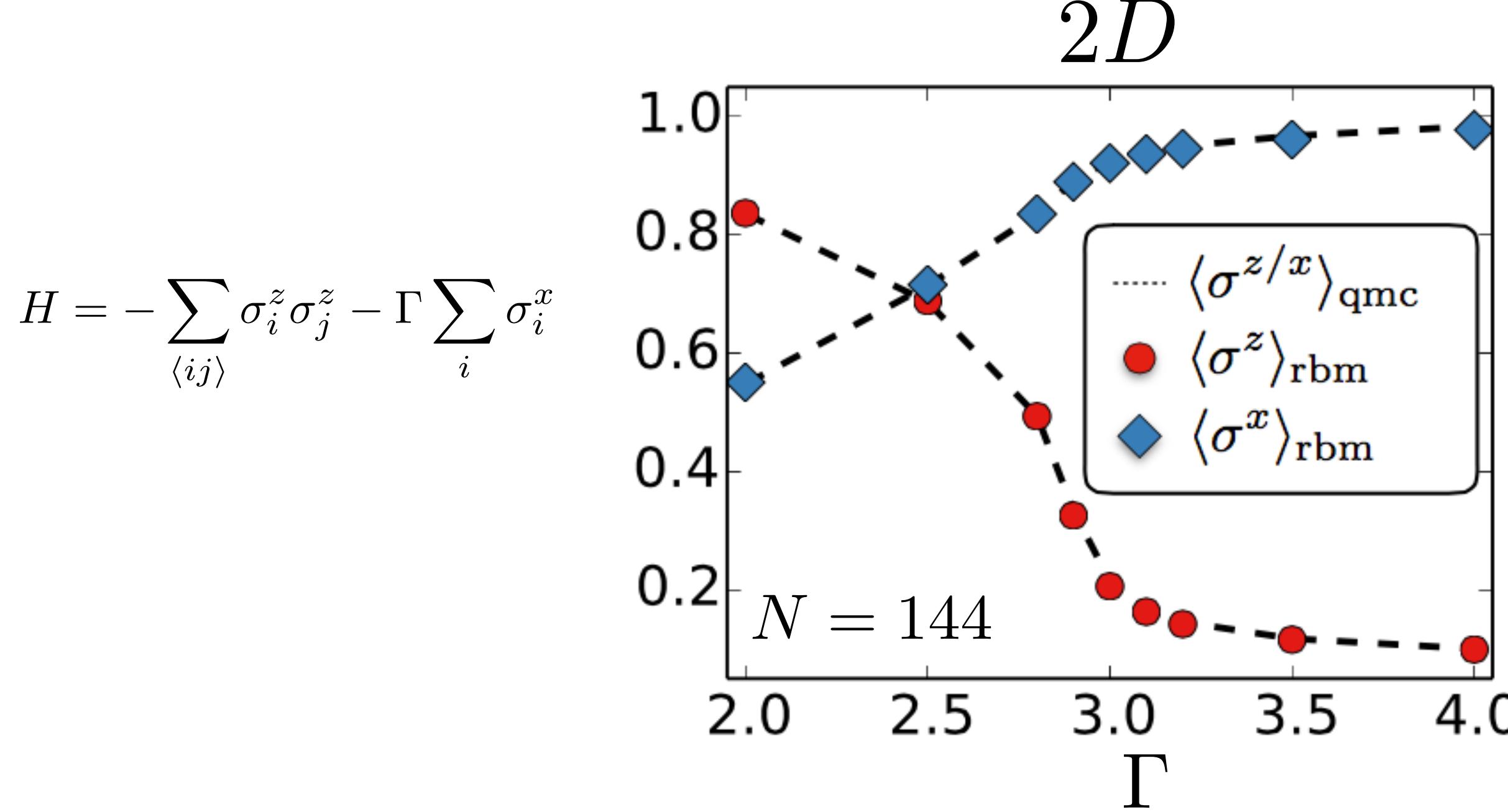
$$\psi_\lambda(\mathbf{x}) \propto \sqrt{p_\lambda(\mathbf{x})}$$

Can be trained with samples in only one basis (the computational basis)

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

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

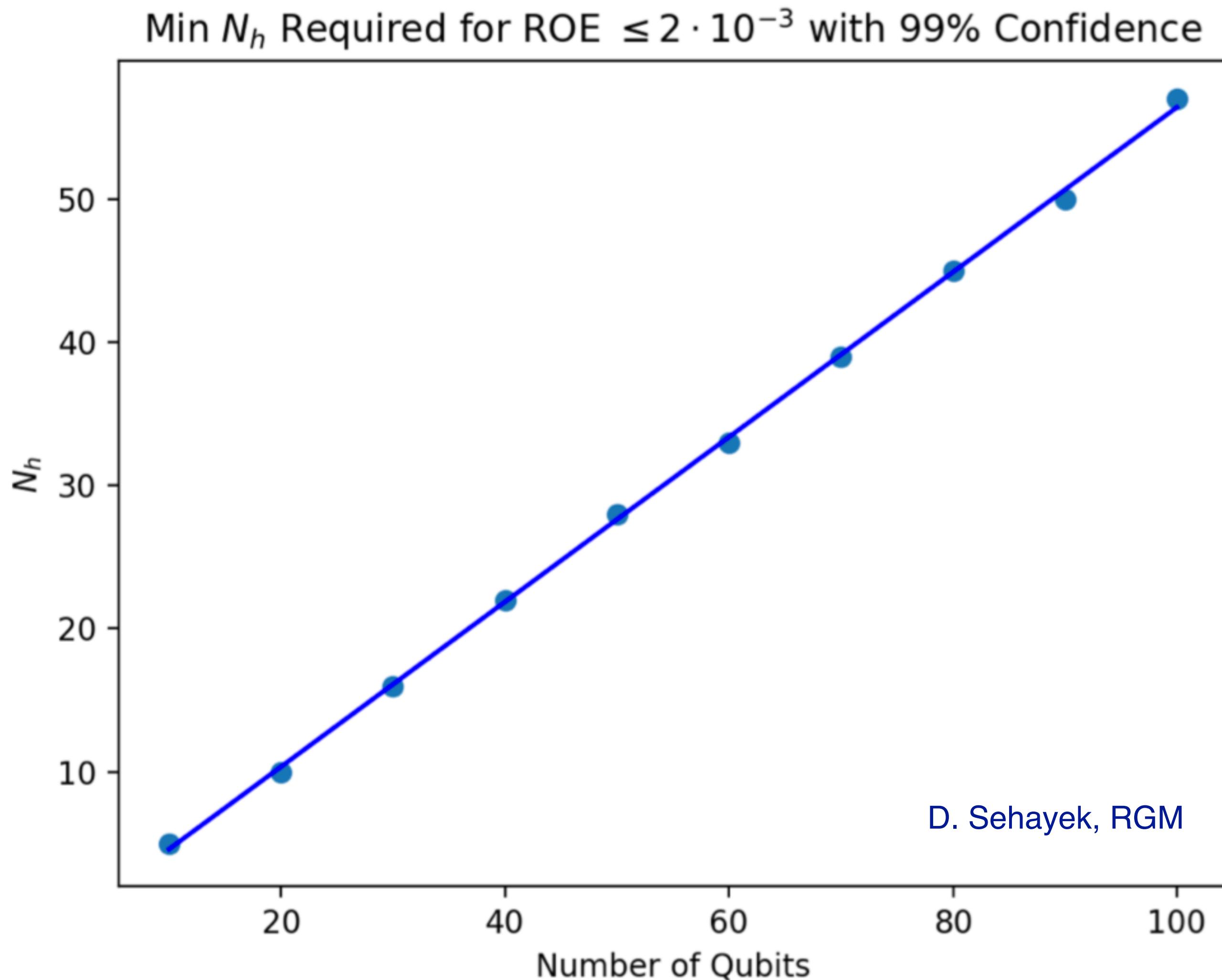
“local” estimator



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

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

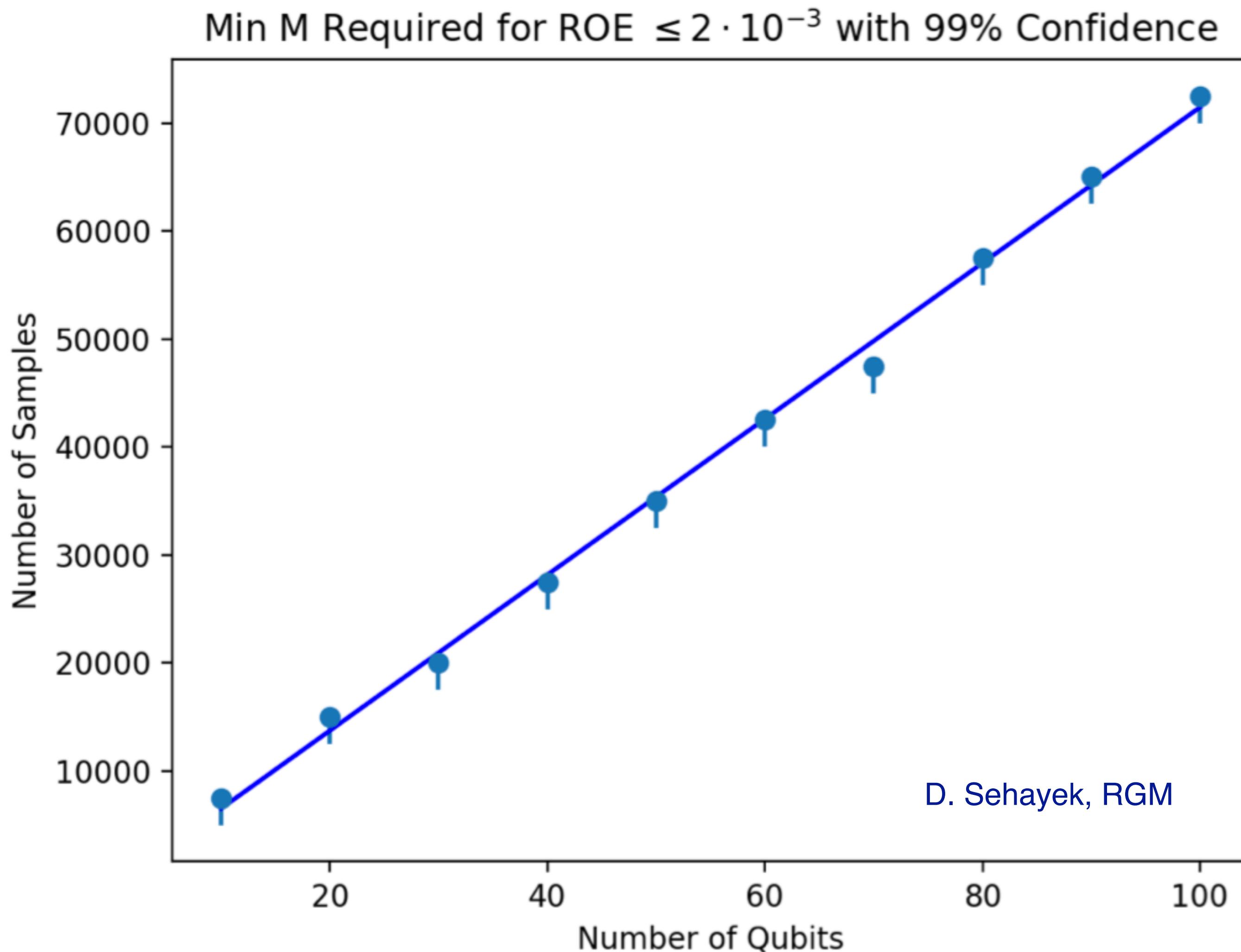
1D quantum critical point



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

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

1D quantum critical point

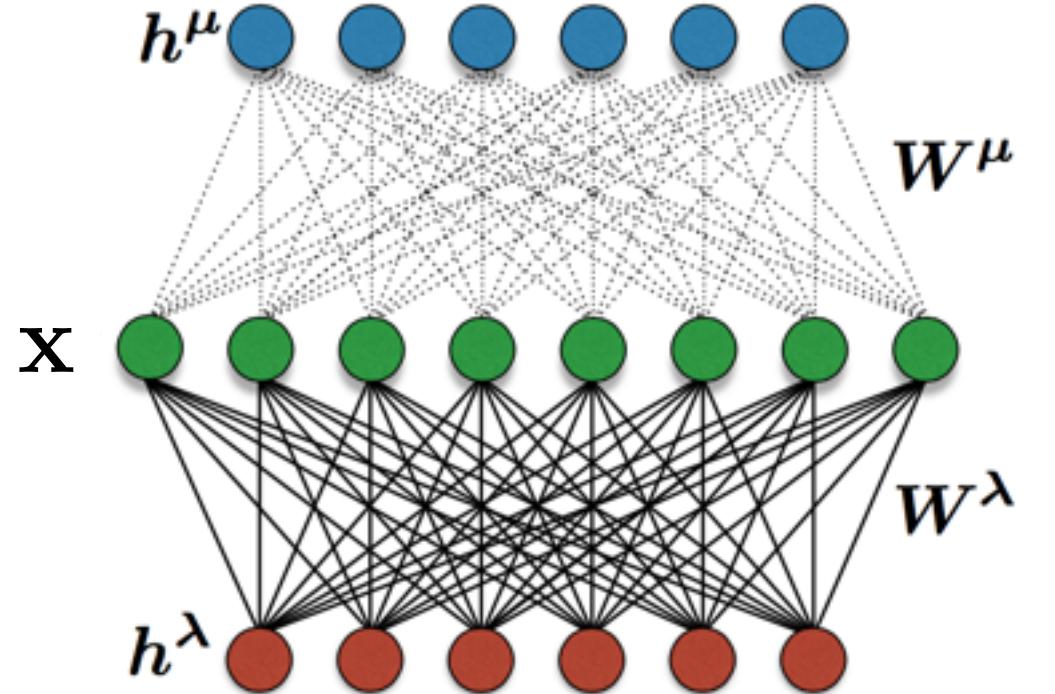


Learning complex wavefunctions

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.

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



state rotated into basis b with the appropriate unitary

N_b = number of bases

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

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

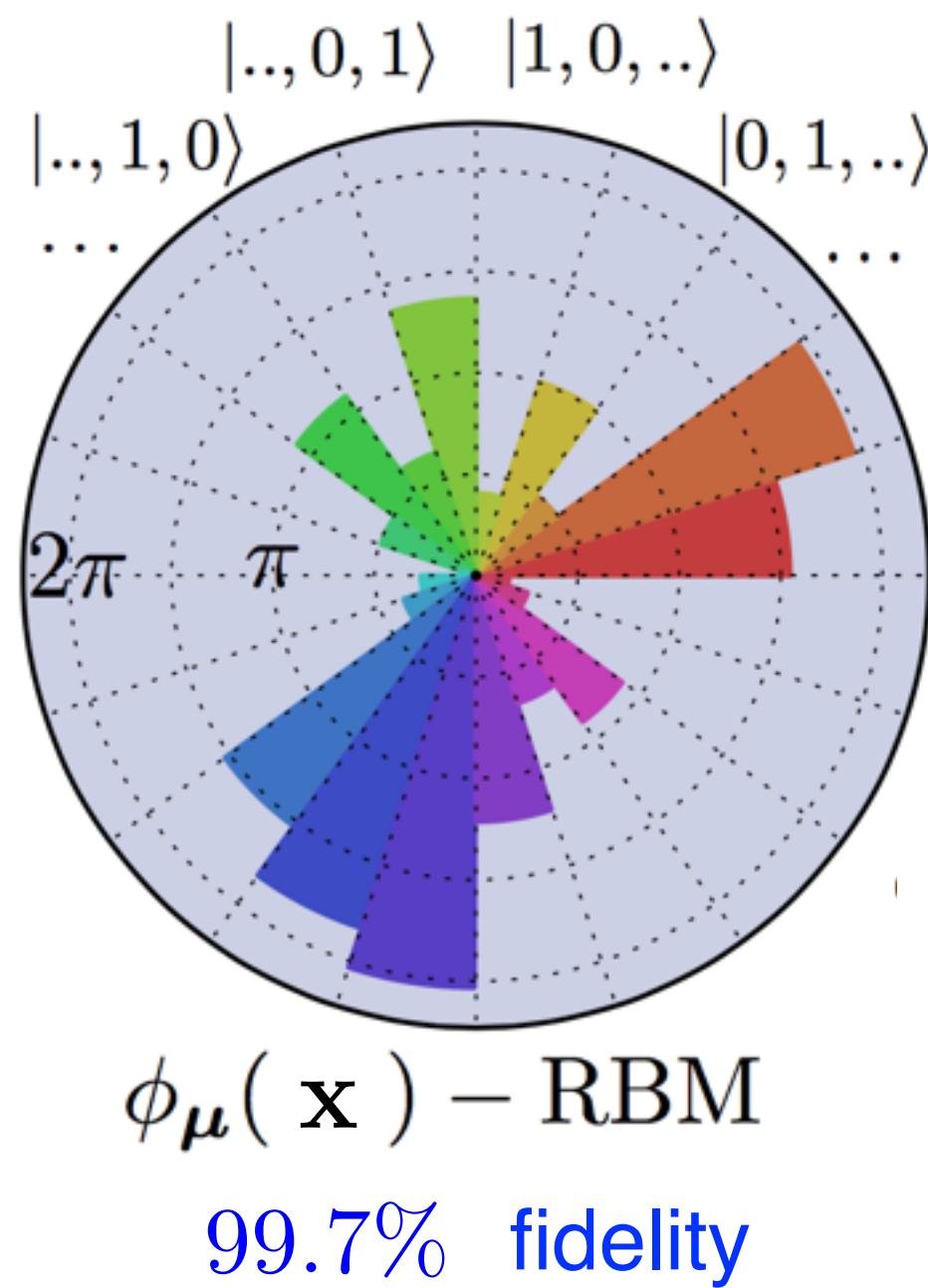
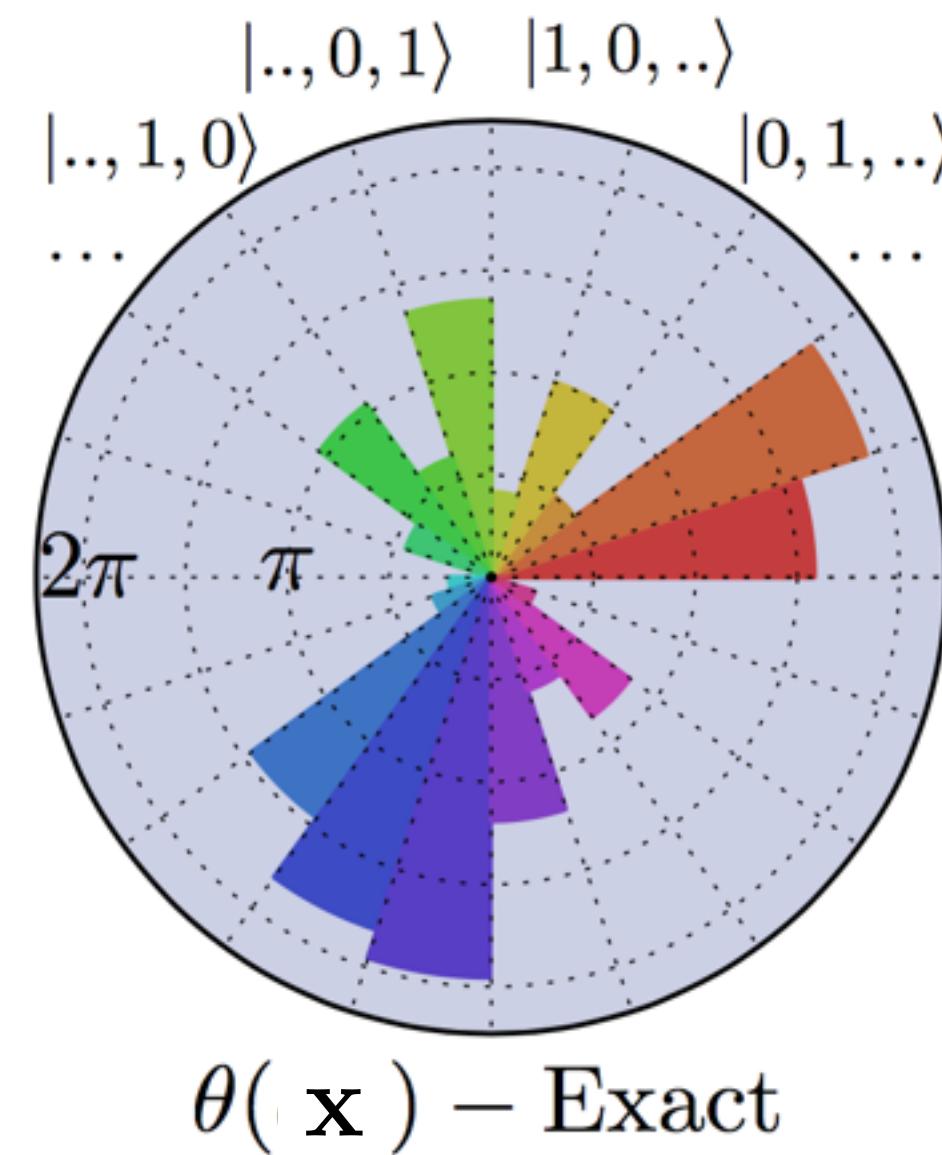
In practice, training is done in two stages: learning of amplitude first, then optimization of the phase parameters.

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

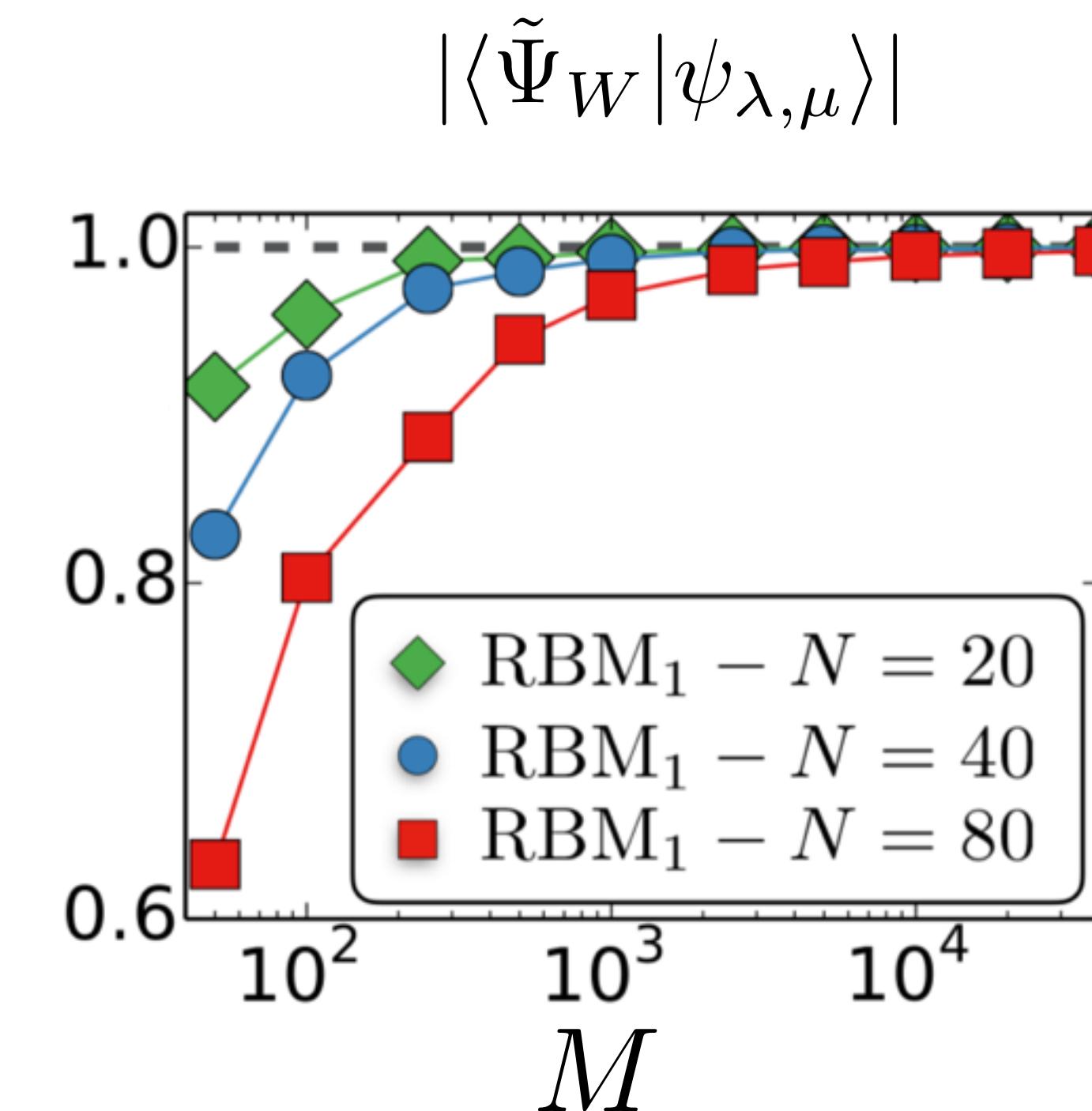
$$|\tilde{\Psi}_W\rangle = \frac{1}{\sqrt{N}} \left(e^{i\theta_1} |100\dots\rangle + e^{i\theta_2} |010\dots\rangle + \dots + e^{i\theta_N} |0\dots01\rangle \right)$$



$N = 20$ qubits

$M = 2 \times 10^5$ measurements

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



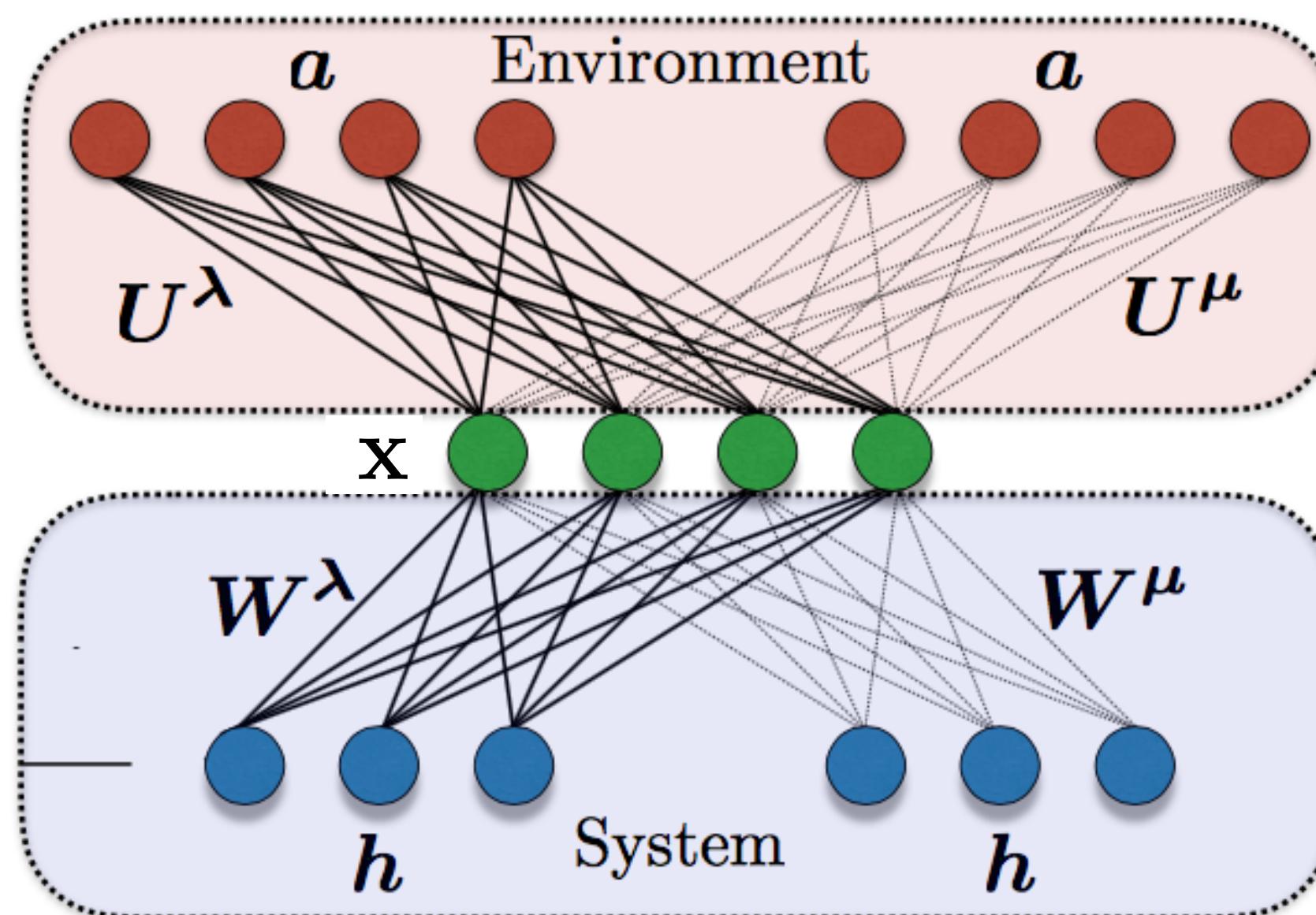
Learning density matrices 1

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

- Can extend an RBM to represent mixed state density matrices

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

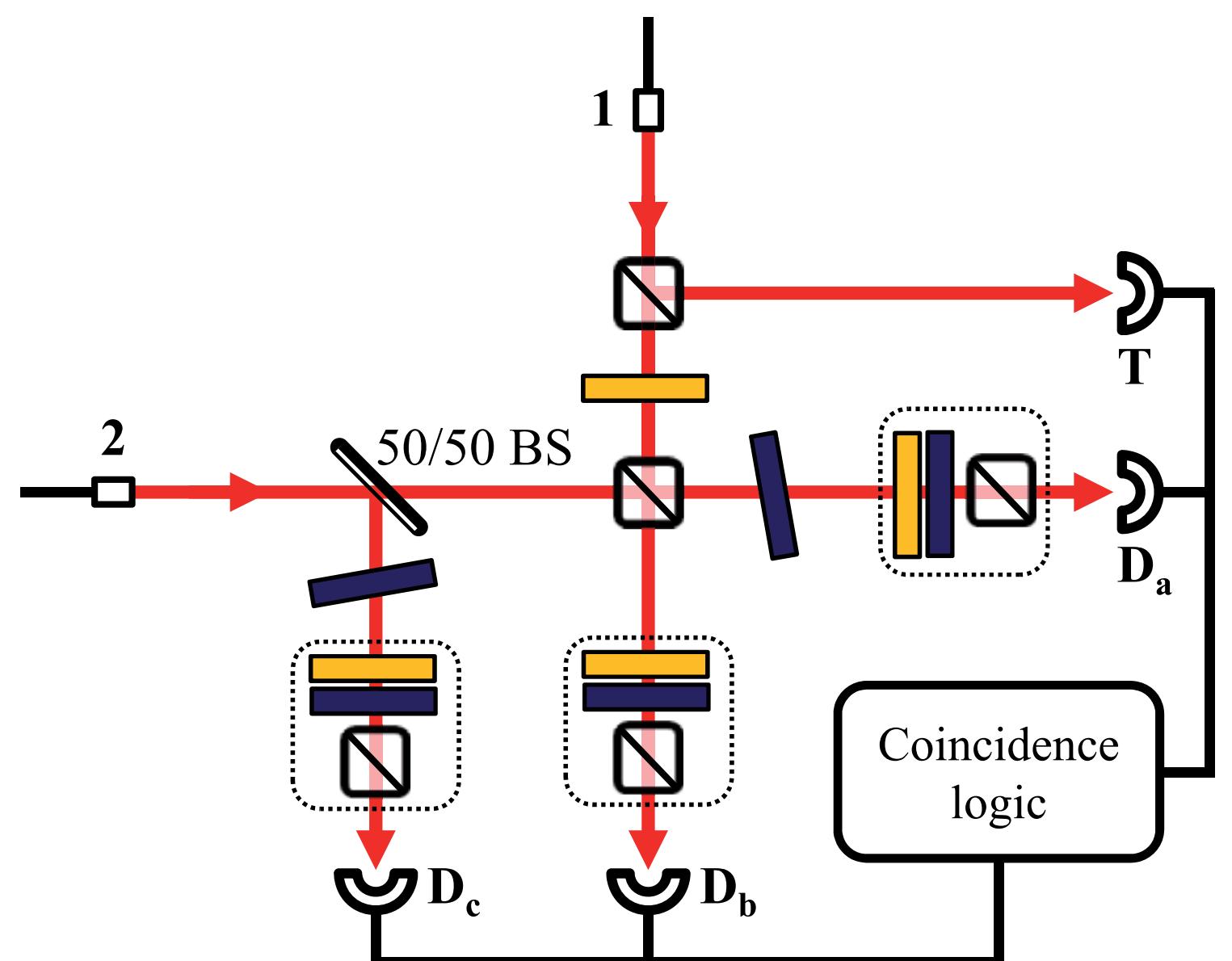
$$\rho_{\lambda,\mu}(\mathbf{x}, \mathbf{x}') = \sum_{\mathbf{a}} \psi_{\lambda,\mu}(\mathbf{x}, \mathbf{a}) \psi_{\lambda,\mu}^*(\mathbf{x}', \mathbf{a})$$



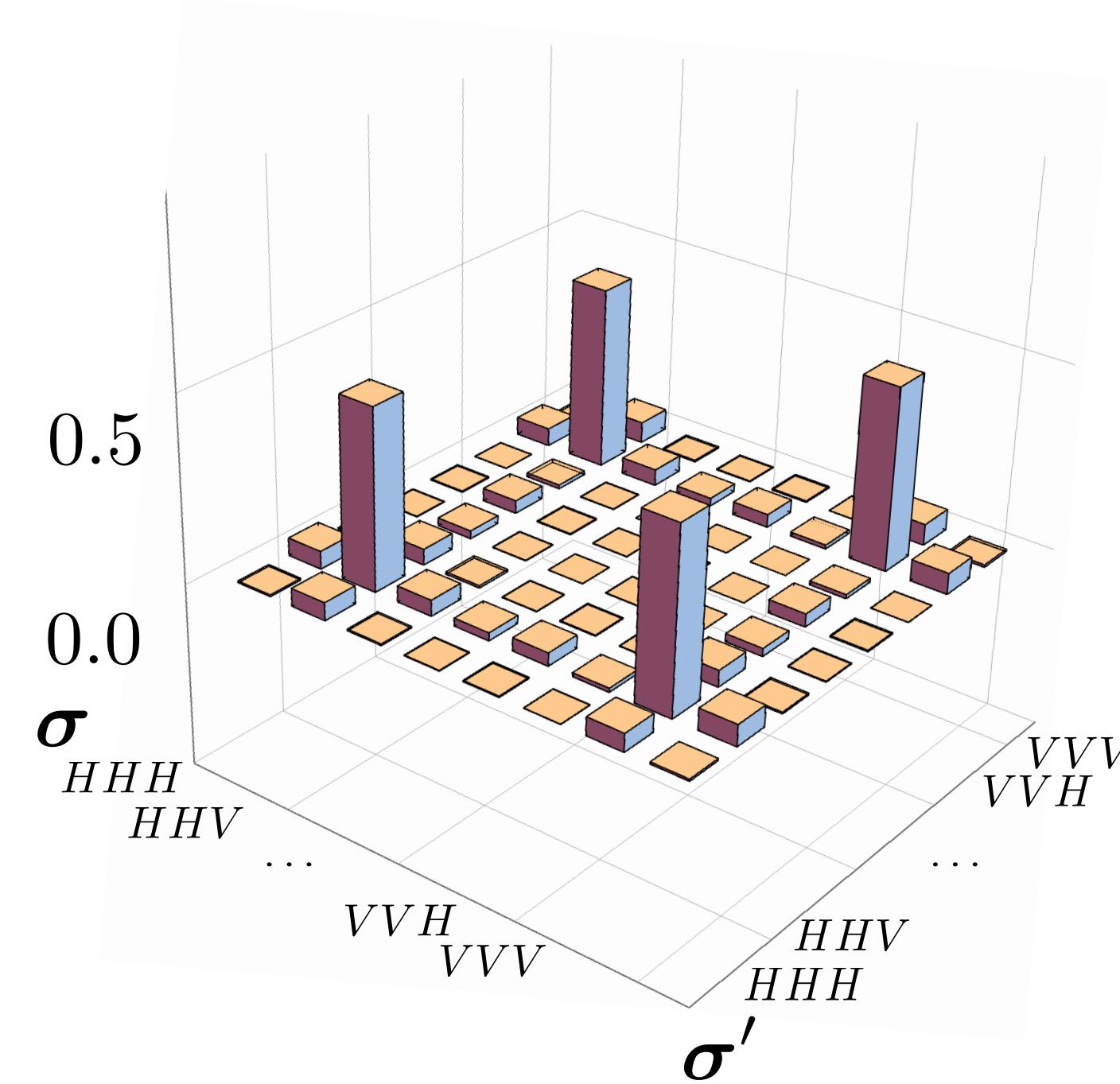
- Parameterization is physically appealing but appears to scale exponentially as a learning problem

Experiment: entangled photons

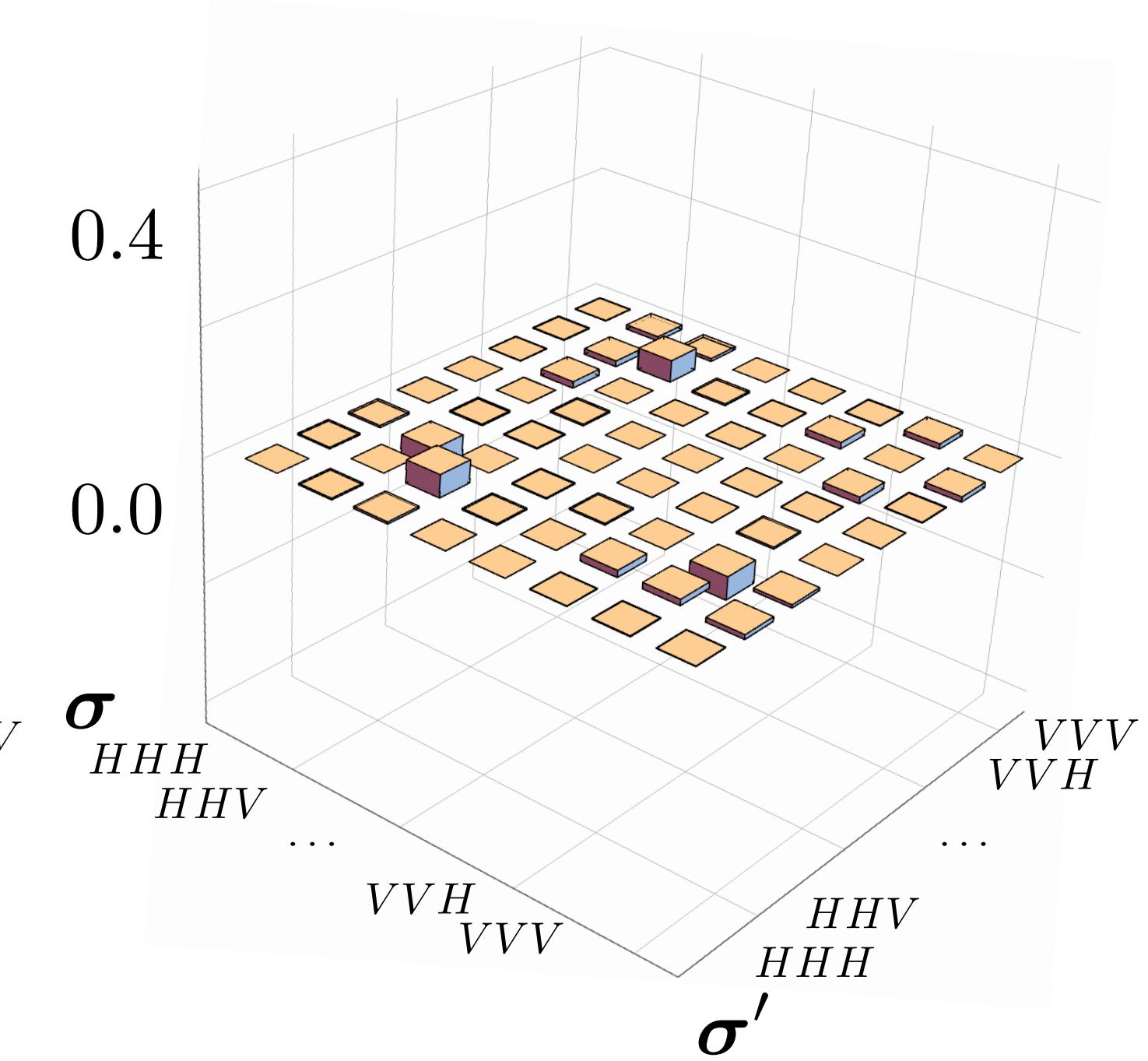
$$|\psi\rangle = \frac{1}{\sqrt{2}} (|HHV\rangle + |VVH\rangle)$$



$\text{Re}[\rho_{\lambda^* \mu^*}]$



$\text{Im}[\rho_{\lambda^* \mu^*}]$



K. Resch lab, IQC Waterloo

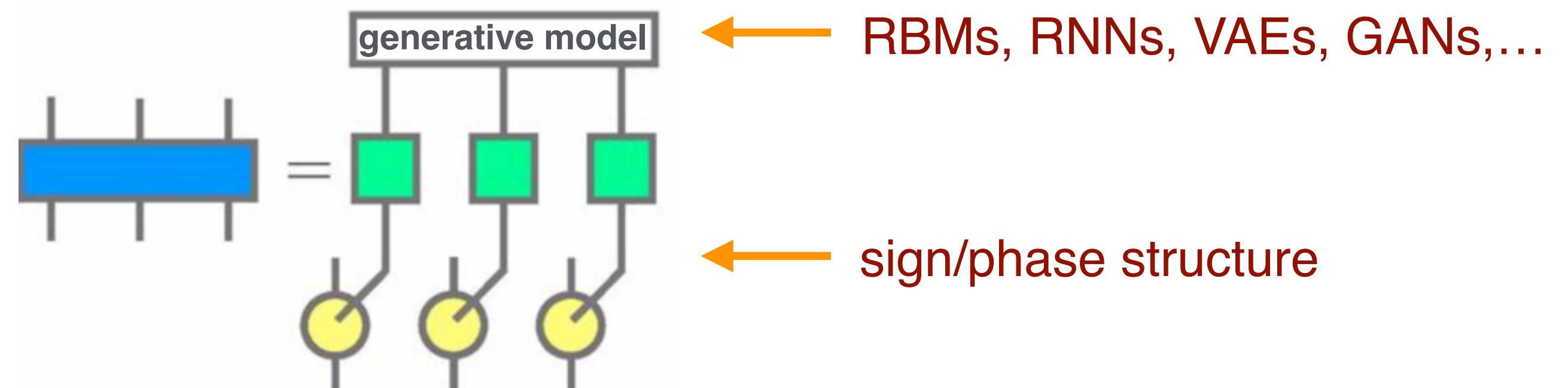
$\mathcal{F} \approx 85\%$

Learning density matrices 2

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

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

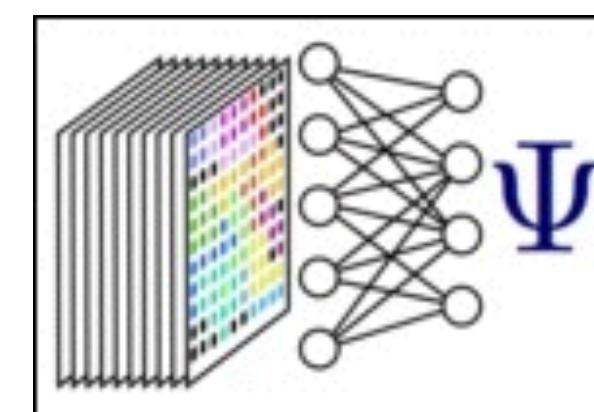
$$p(\mathbf{x}) = \text{Tr} \rho M^{\mathbf{x}}$$



Kavli Institute for
Theoretical Physics
University of California, Santa Barbara

Learning and representing quantum states with probability
Juan Carrasquilla, Vector Institute

<http://online.kitp.ucsb.edu/online/machine19/carrasquilla/>



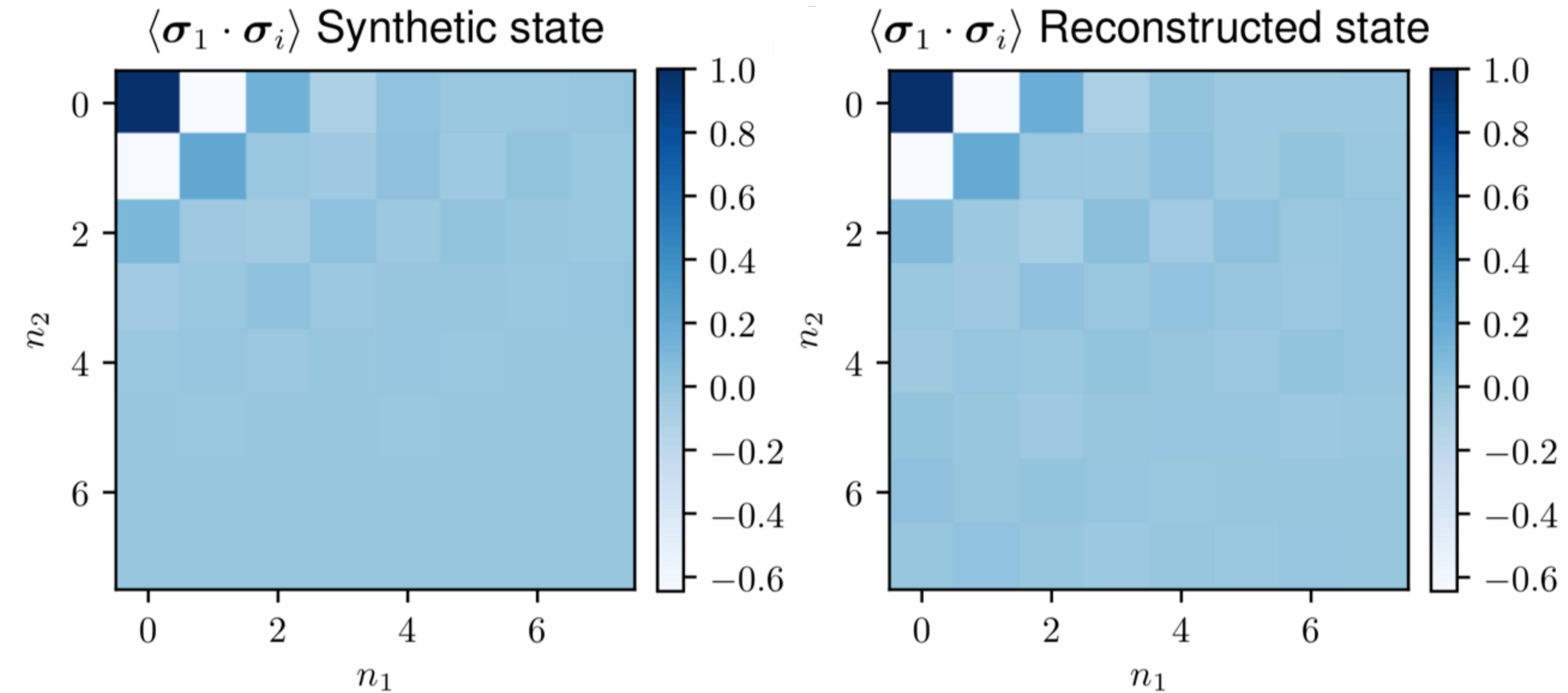
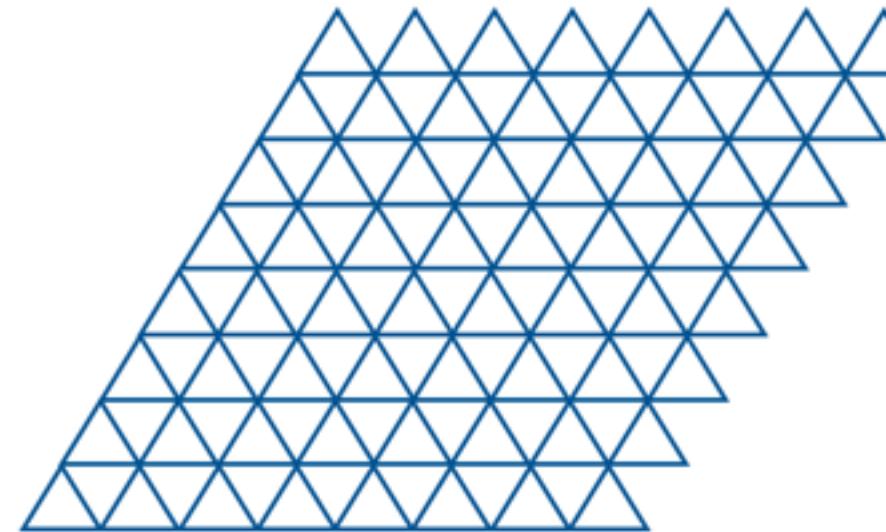
Learning density matrices 2

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

- Correlation functions for the ground state of the triangular lattice Heisenberg antiferromagnet

$$H = \sum_{\langle i,j \rangle} \mathbf{S}_i \cdot \mathbf{S}_j$$

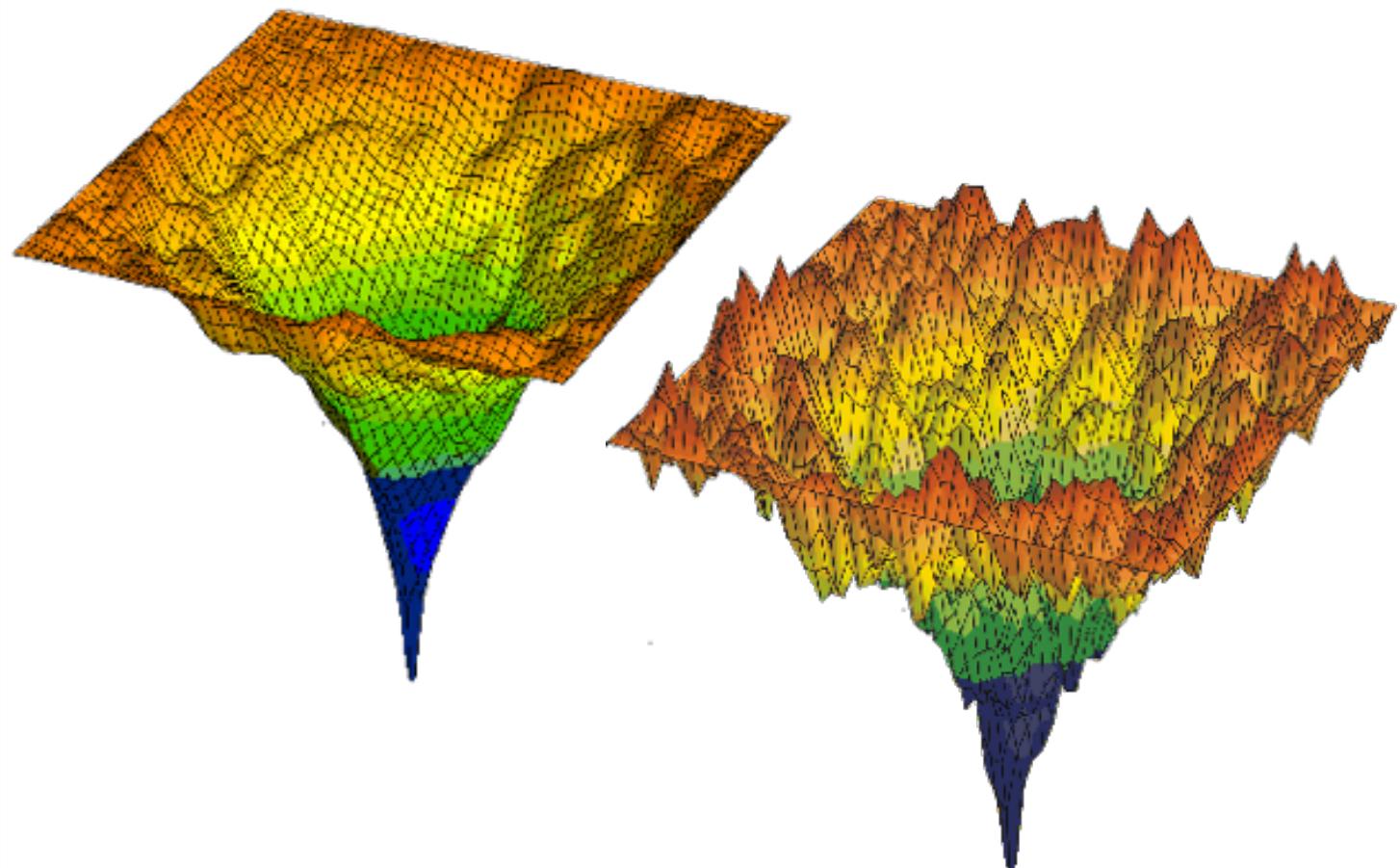
$$N = L \times L = 8 \times 8$$



Outlook:

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

$$\begin{aligned}n_h &\propto \text{poly}(N) \\M &\propto \text{poly}(N) \\N_b &\propto \text{poly}(N)\end{aligned}$$



- How is learning from data **different** than optimizing neural network parameters variationally with knowledge of the Hamiltonian?
- **Generative modelling** can potentially become very powerful when combined with current experiments

