University of Zurich[UZH]

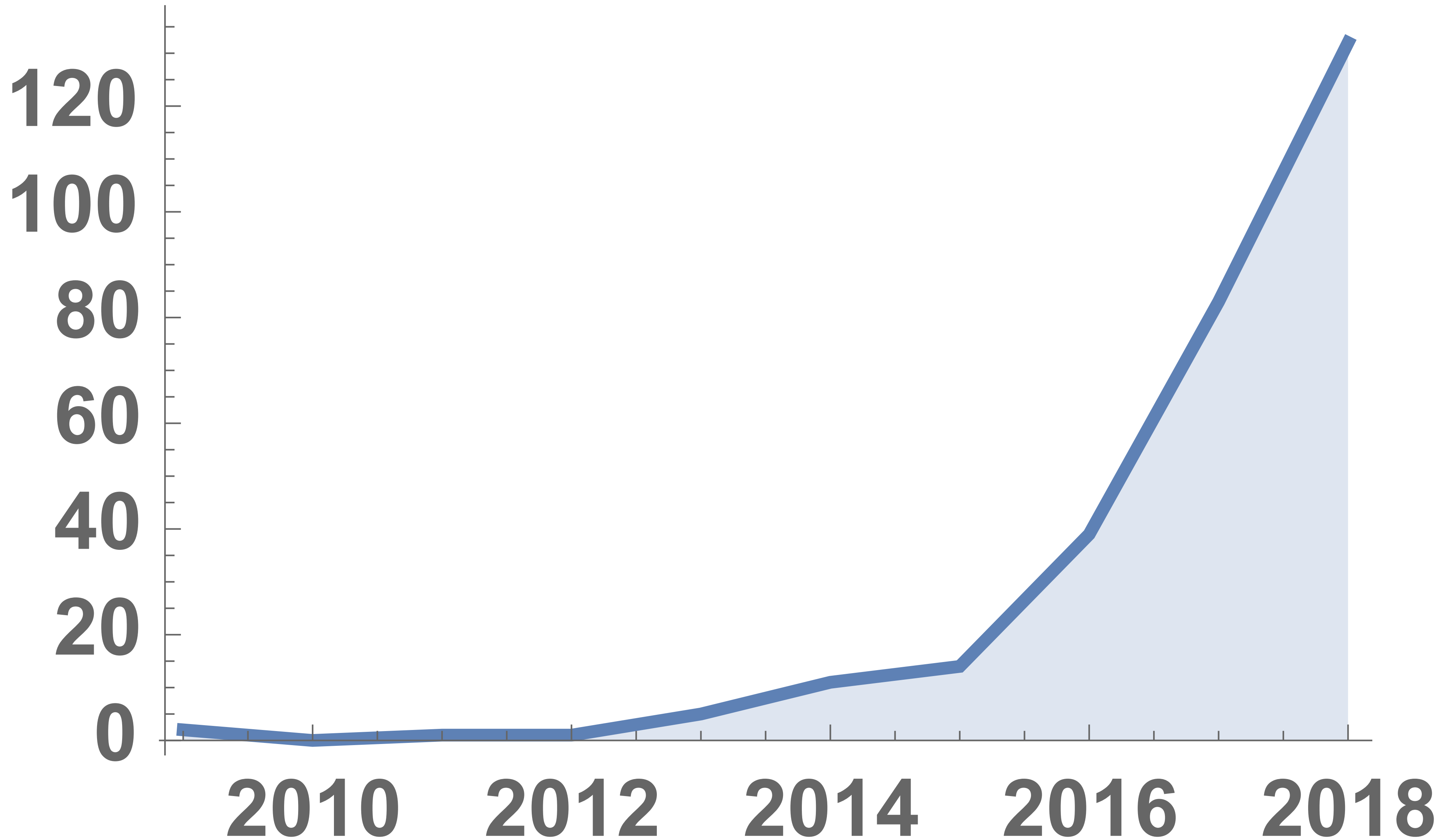
Titus Neupert

KITP, Feb 14 2019

some

Applications of Neural Networks **in Condensed Matter Physics**

Number of cond-mat papers with
“machine learning” in the abstract



University of Zurich[UZH]

PART I Phase classification

PRB 95, 245134 (2017)
Phys. Rev. B 98, 174202 (2018)
arXiv:1812.05625

PART II Neural networks as variational wave functions

PRL 121, 167204 (2018)

PART III Quantum machine learning

preliminary

Kenny Choo
(U Zurich)

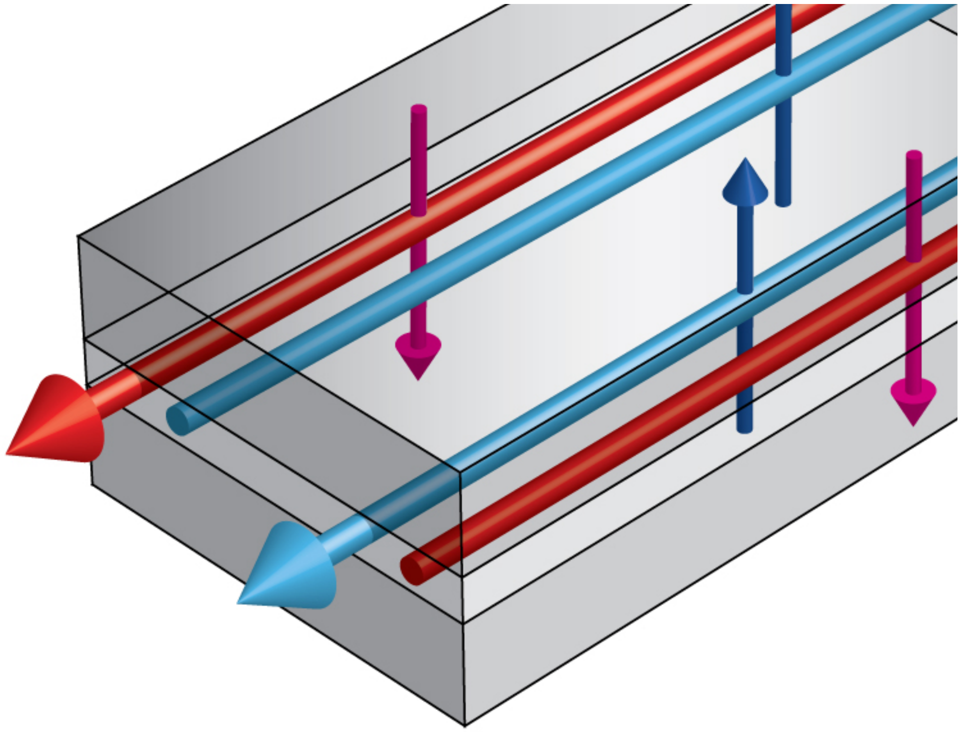
Frank Schindler
(U Zurich)

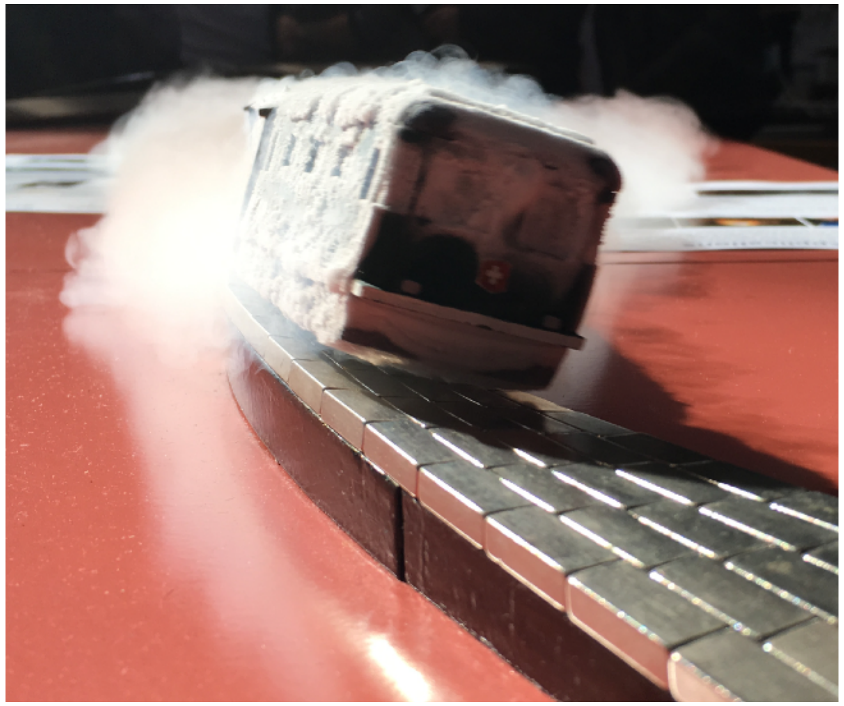
Giuseppe Carleo (Flatiron)
Nicolas Regnault (ENS Paris)
Johan Chang (UZH)
Pascal M. Vecsei (UZH)
Ruben Beynon (UZH)

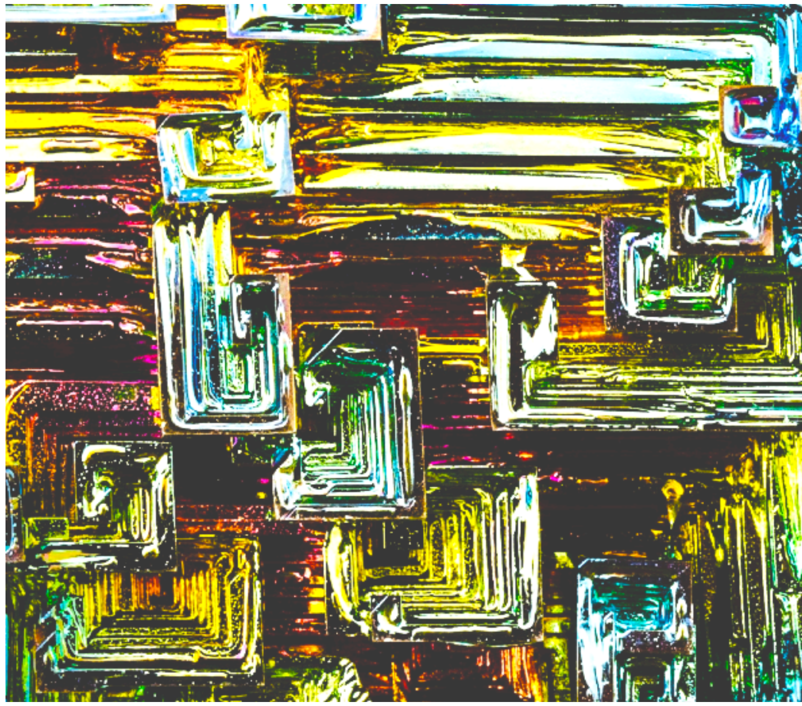
Elmer V. H. Doggen (KIT)
Konstantin S. Tikhonov (KIT)
Alexander D. Mirlin (KIT)
Dmitry G. Polyakov (KIT)
Igor V. Gornyi (KIT)

University of Zurich[UZH]

PART I Phase classification







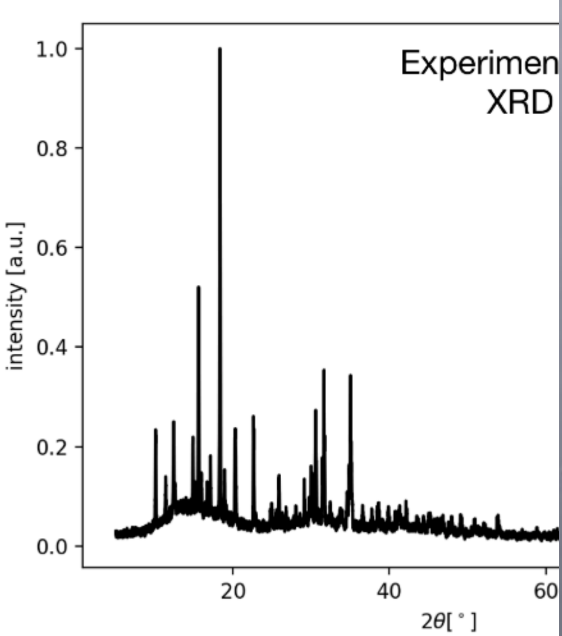
Condensed matter physics is a classification problem

University of Zurich[UZH]

Phase classification (fully supervised)

[Vecsei et al., arXiv:1812.05625]

Example: find **crystal structure** (space group/crystal system classification) from X-ray diffraction (XRD) patterns



classification in one of 230 space groups

train with theore
data (~100 000

University of Zurich[UZH]

Phase classification (fully supervised)

[Vecsei et al., arXiv:1812.05625]

Results:

quite messy experimental data on natural crystals

	Crystal systems		Space groups	
	Test set	RRUFF	Test set	RRUFF
Convolutional	85%	56%	76%	42%
Dense	73%	70%	57%	54%

If network can be uncertain about ~50% of the cases

80%

University of Zurich[UZH]

Phase classification in unknown phase diagram

Objective:

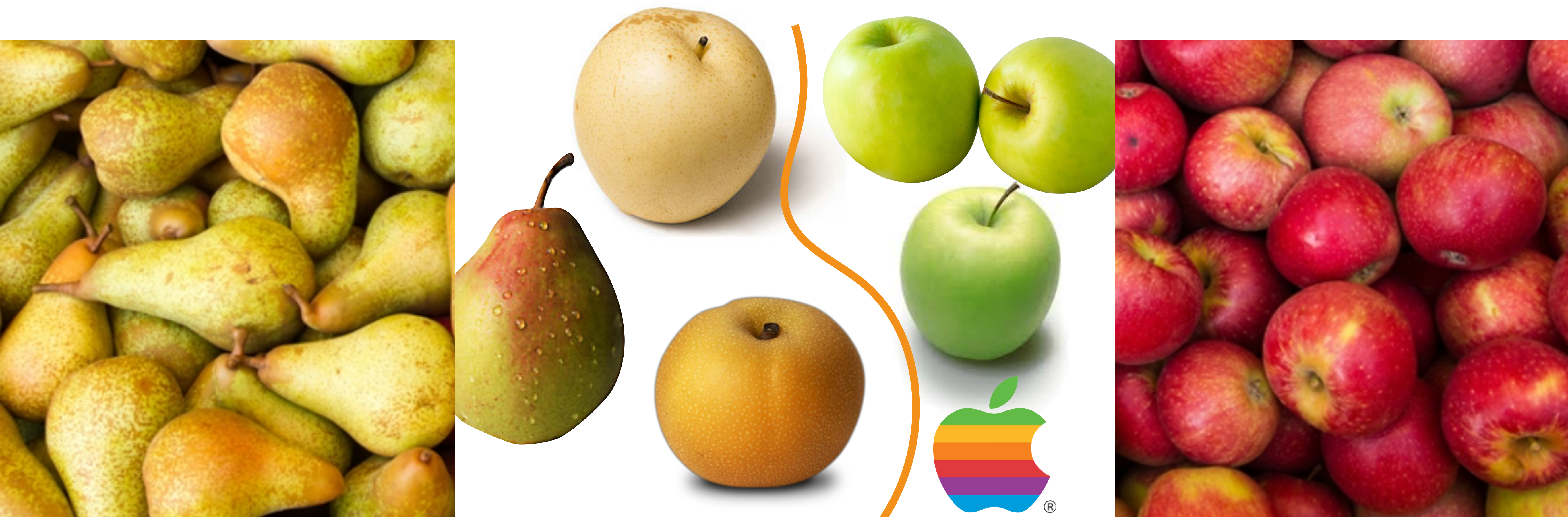
Classification of phases of matter using correlation functions

Supervised learning:

Training deep in the phase

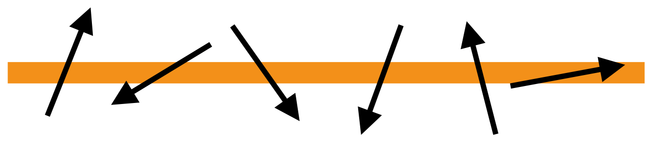
Determine phase boundary:

Apply to states for which classification is less clear



Toy problem: Many-body localization

Standard model of MBL: spin-1/2 Heisenberg chain, open boundary conditions



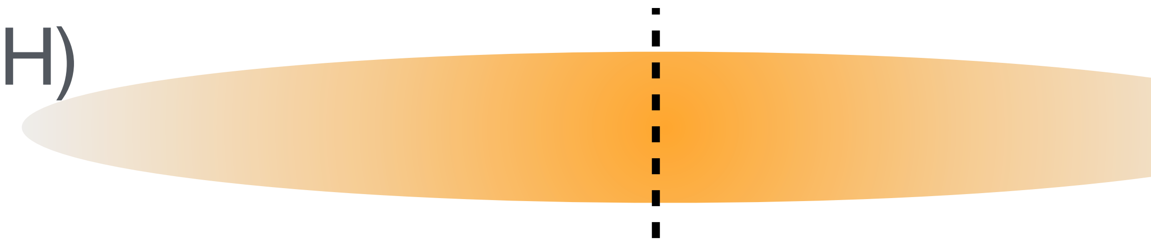
$$H = J \sum_{\mathrm{r}=1}^{N-1} \boldsymbol{S}_{\mathrm{r}} \cdot \boldsymbol{S}_{\mathrm{r}+1} + \sum_{\mathrm{r}=1}^{N} h_{\mathrm{r}} S^z_{\mathrm{r}}$$

$$J = 1$$

$$h_{\mathrm{r}} \in [-\bar{h}, \bar{h}]$$

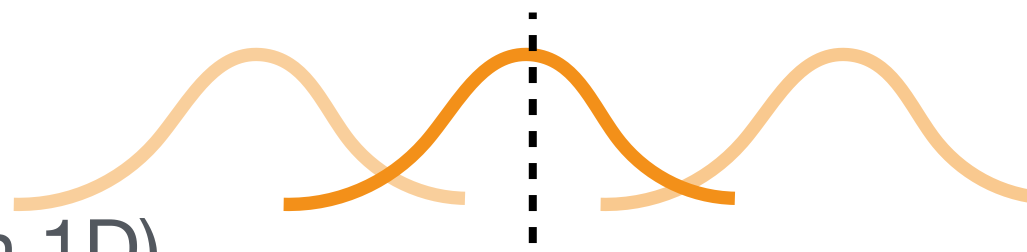
$\bar{h} \ll 1$ **thermalizing** regime (obeys ETH)

volume law entanglement



$\bar{h} \gg 1$ **many-body localized** regime

area law entanglement (constant in 1D)



regimes defined for states at finite energy density (not ground state)

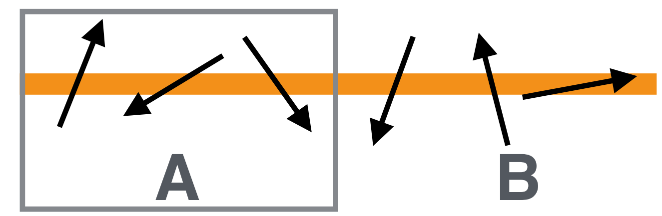
Solve with ED: N = …12, 14, 16, 18 site chain; use U(1) symmetry

University of Zurich[UZH]

Conventional classification methods

based on energy level spectrum or entanglement entropy/spectrum

$$\rho_A = \mathrm{Tr}_B \left|\Psi\right\rangle \left\langle\Psi\right| \equiv e^{-H_\mathrm{e}}$$



crude

i) **Schmidt gap**: $\lambda_1(\rho_A) - \lambda_2(\rho_A) \rightarrow 1$ for MBL (nearly pure)

needs finite size scaling

ii) Volume vs. area law **scaling** of $S(N$ …

phase transition does not correspond to maximum

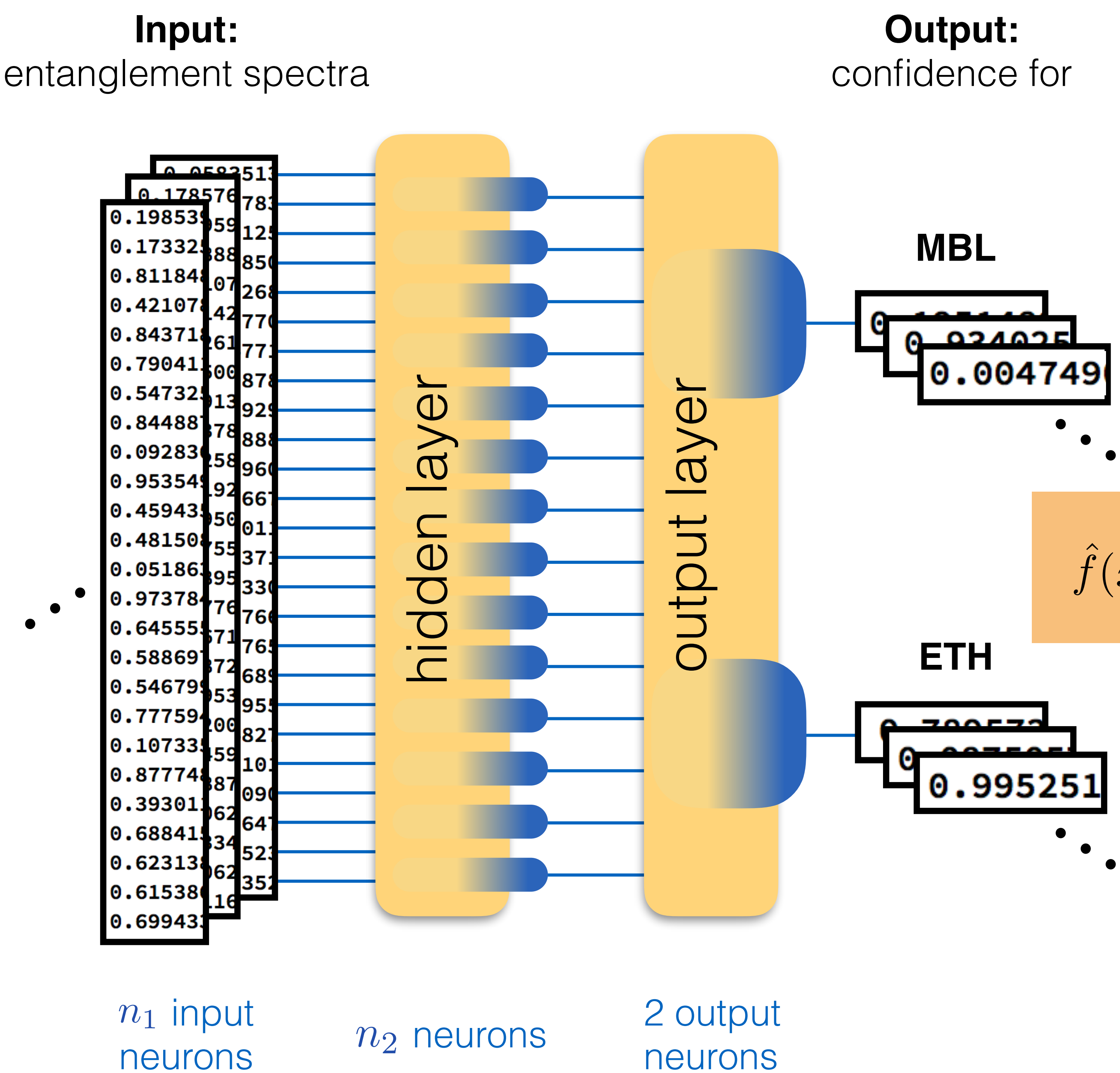
iii) **Standard deviation** of $S(N_A)$ … es

large near the transition where both MBL and ETH like states coexist

iii) **Level statistics** of either the entanglement spectrum or the energy spectrum follow distinct statistical distributions in each regime

needs large systems

Plain vanilla neural network



$$\mathrm{ReLu}_i(x) = x_i\,\theta(x_i)$$

$$\mathrm{Softmax}_i(x) = \frac{e^{-x_i}}{\sum_j e^{-x_j}}$$

Binary classification

$$\hat{f}(x) = \mathrm{Softmax}[V^{(3,2)}\,\mathrm{ReLu}(V^{(2,1)}\,x + a^{(2)}) + a^{(3)}]$$

Simplicity:

- plain vanilla
- 1 hidden layer
- activation functions ReLu and Softmax

University of Zurich[UZH]

Cost function and regularization

$$\text{Cost}(\hat{f}, f) = -\sum_{x \in \text{TD}} \sum_{i}^{2} f_i(x) \log \hat{f}_i(x) \; - \delta \sum_{x \in \text{TR}} \sum_{i}^{2} \hat{f}_i(x) \log \hat{f}_i(x) + \mu |V|^2$$

labelled training data

classifying output neurons

Cross entropy

Confidence optimization:

favors unlabelled data near phase transition to be classified confidently

random subset of spectra near transition

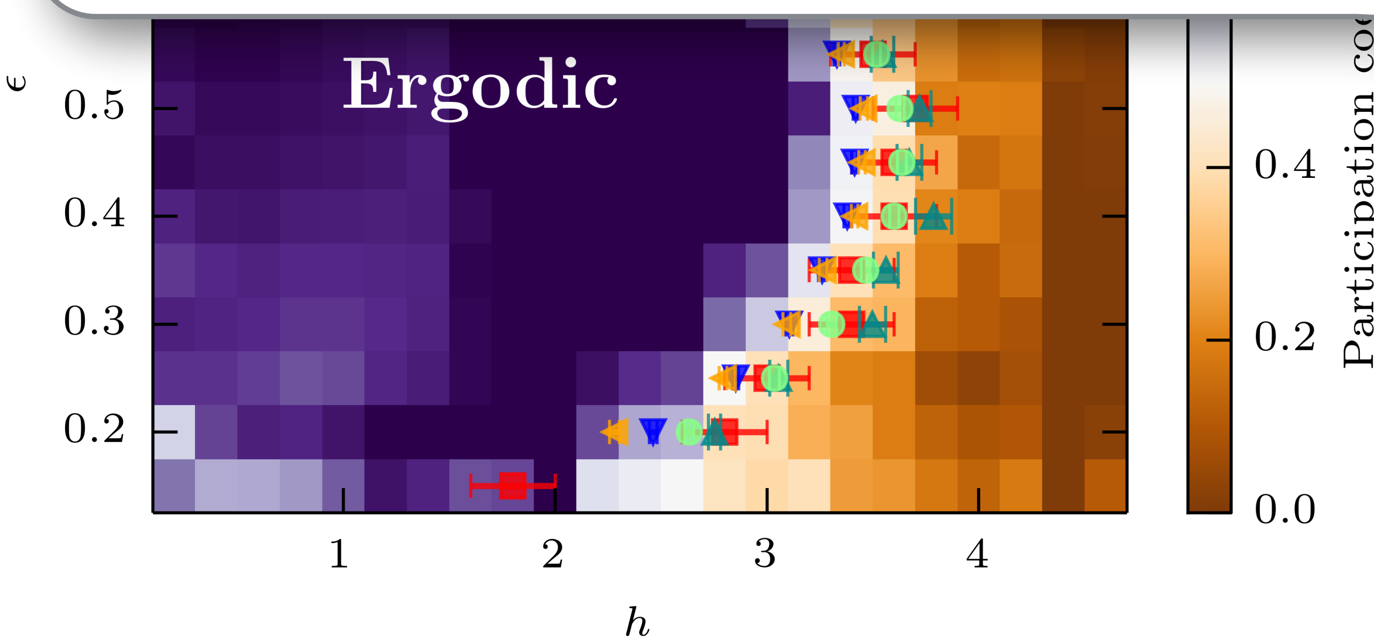
Weight decay:

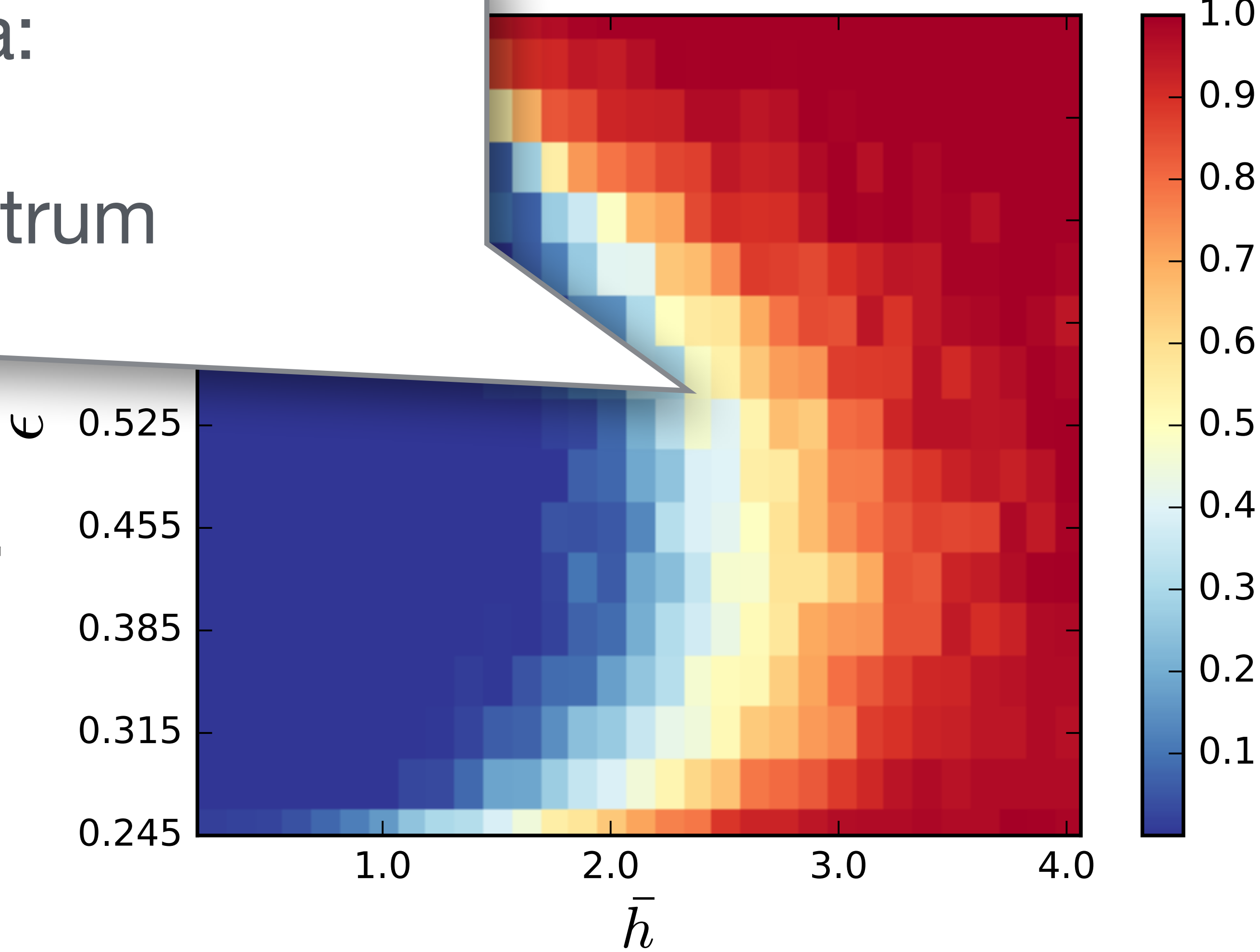
favors having only few nonzero weights/using as few neurons as possible

University of Zurich[UZH]

Results: **Disorder-averaged phase diagram**

Robust against choice of input data:
a) entanglement spectrum
b) differences in entanglement spectrum





Volume law coefficient of the entanglement entropy

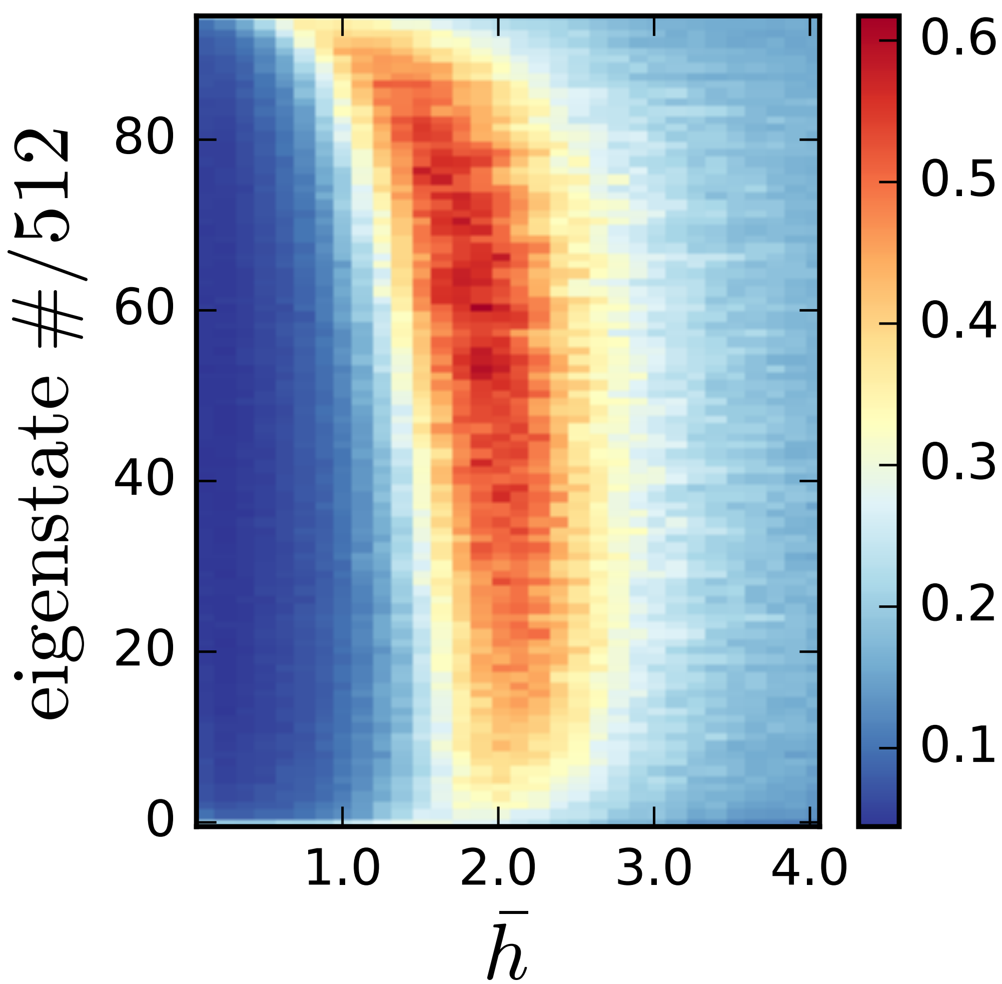
Confidence for MBL averaged over disorder realization and eigenstates

- fewer disorder realizations (40)
- smaller system (N=16)

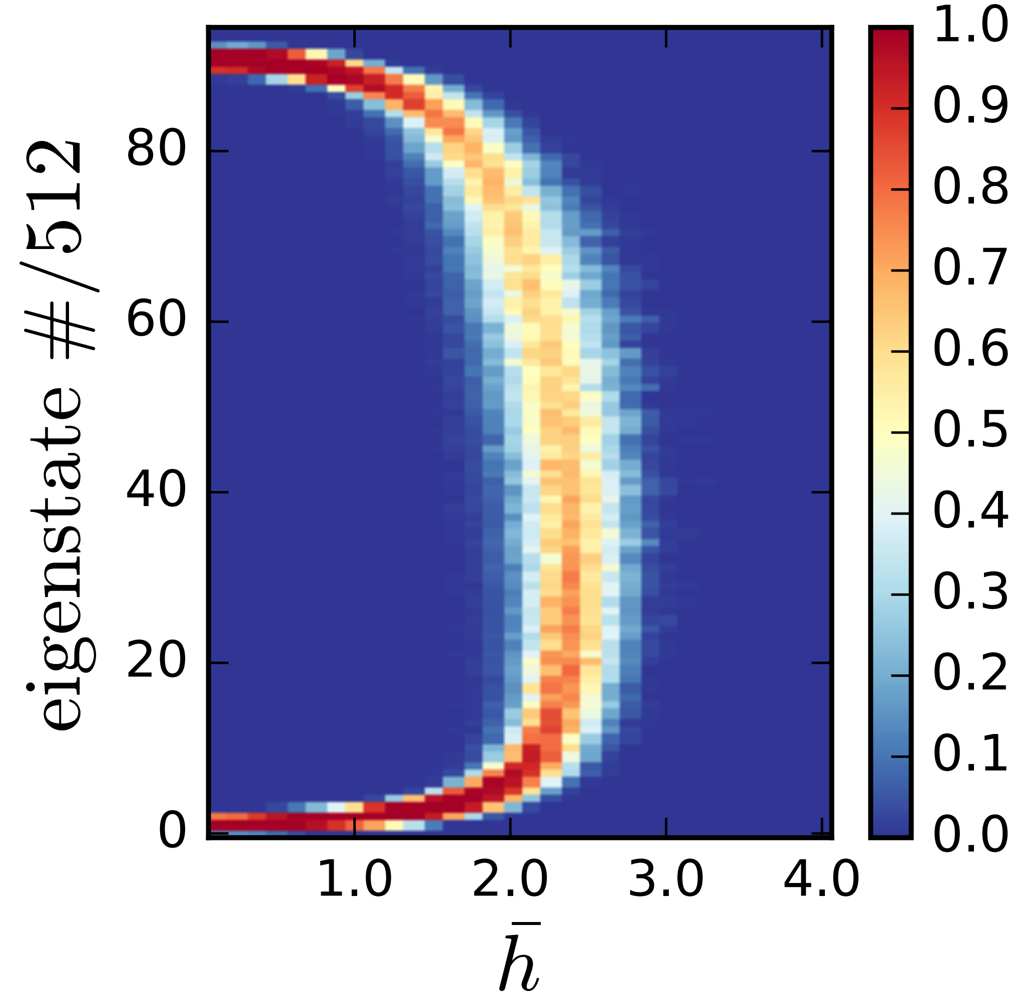
[Luitz et al., PRB 2015]

University of Zurich[UZH]

Results: Transition in single disorder realization



standard deviation of entanglement entropy over 512 consecutive eigenstates

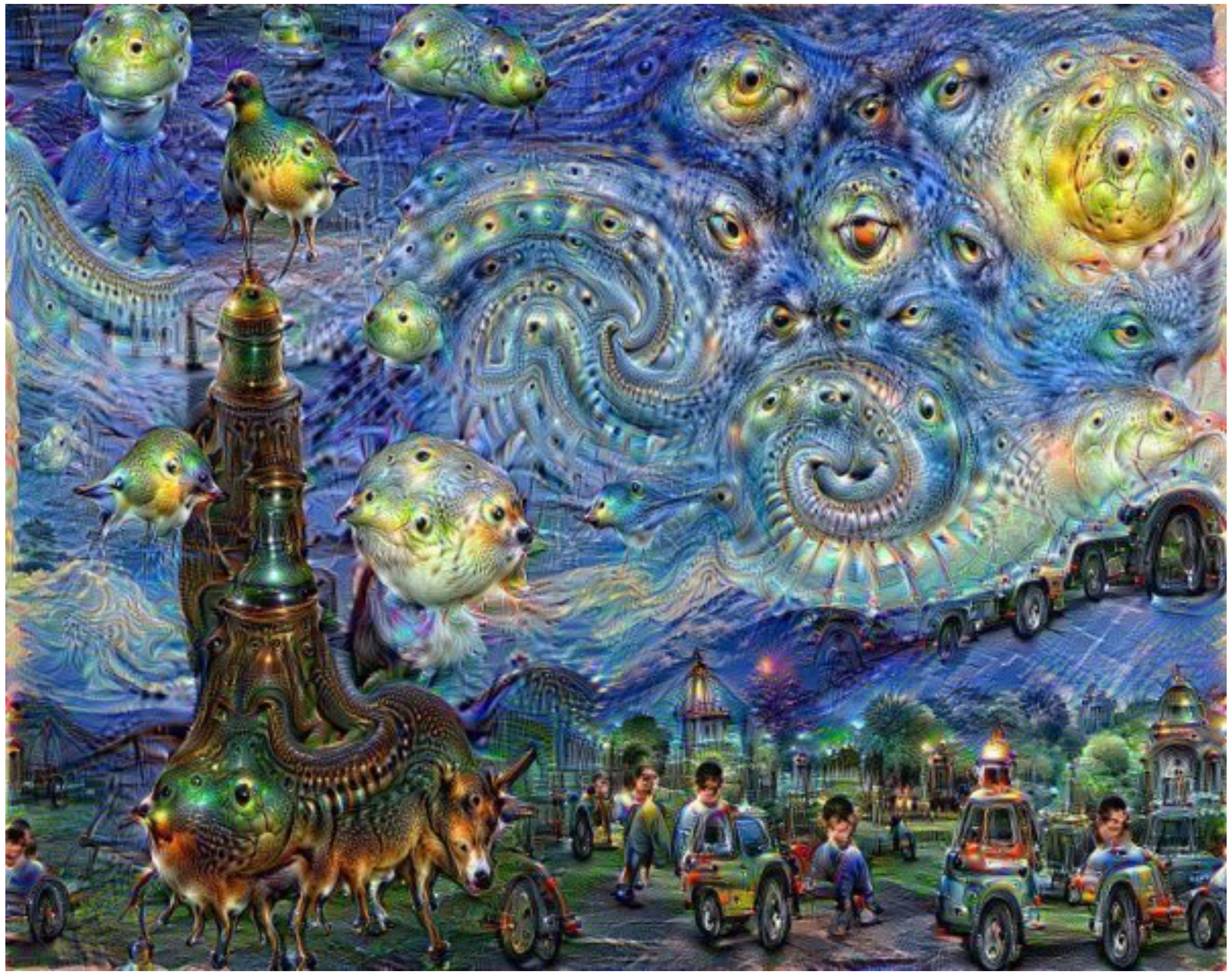


fraction of uncertainly classified states (out of 512 consecutive states)

output >0.9 taken as certain

University of Zurich[UZH]

Dreaming: **What the network has learned**

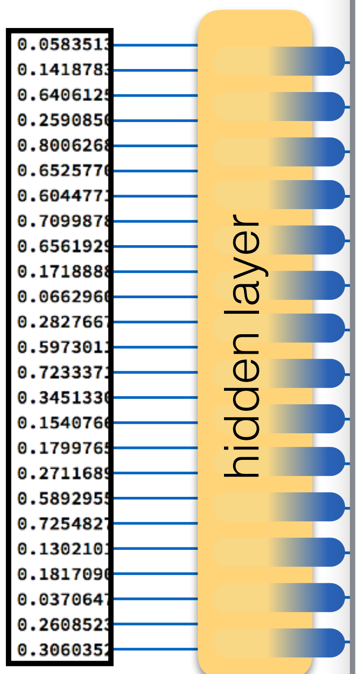


University of Zurich[UZH]

Dreaming: What the network has learned

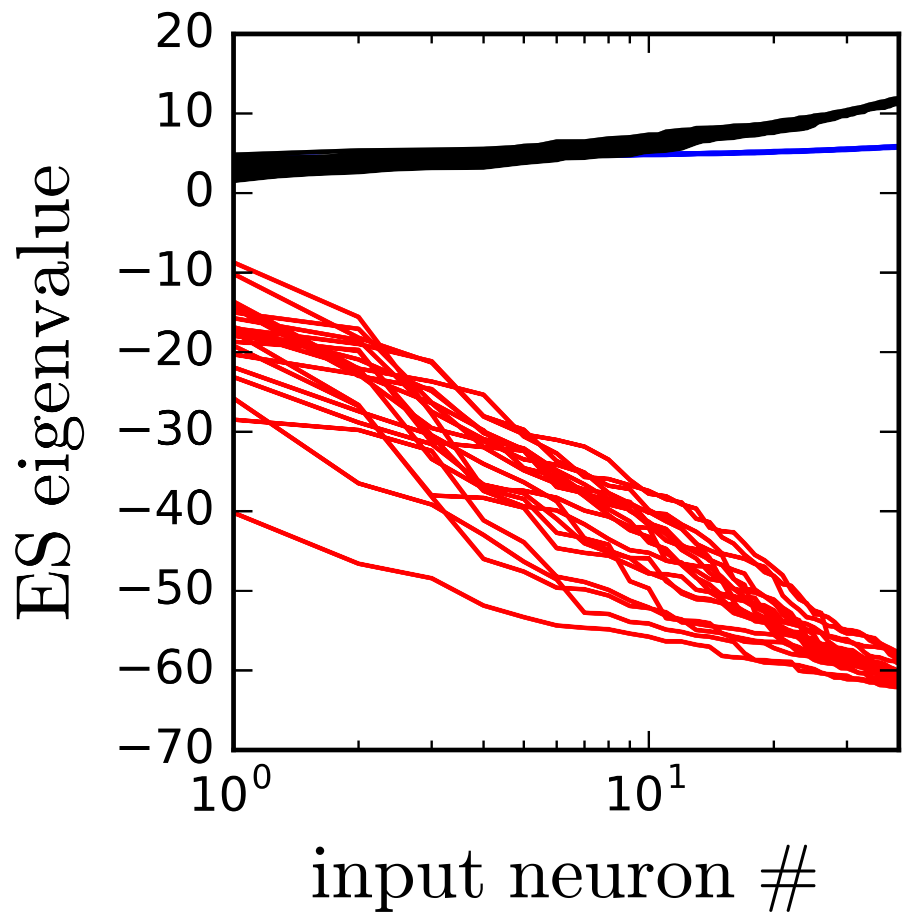
Creation by

random input

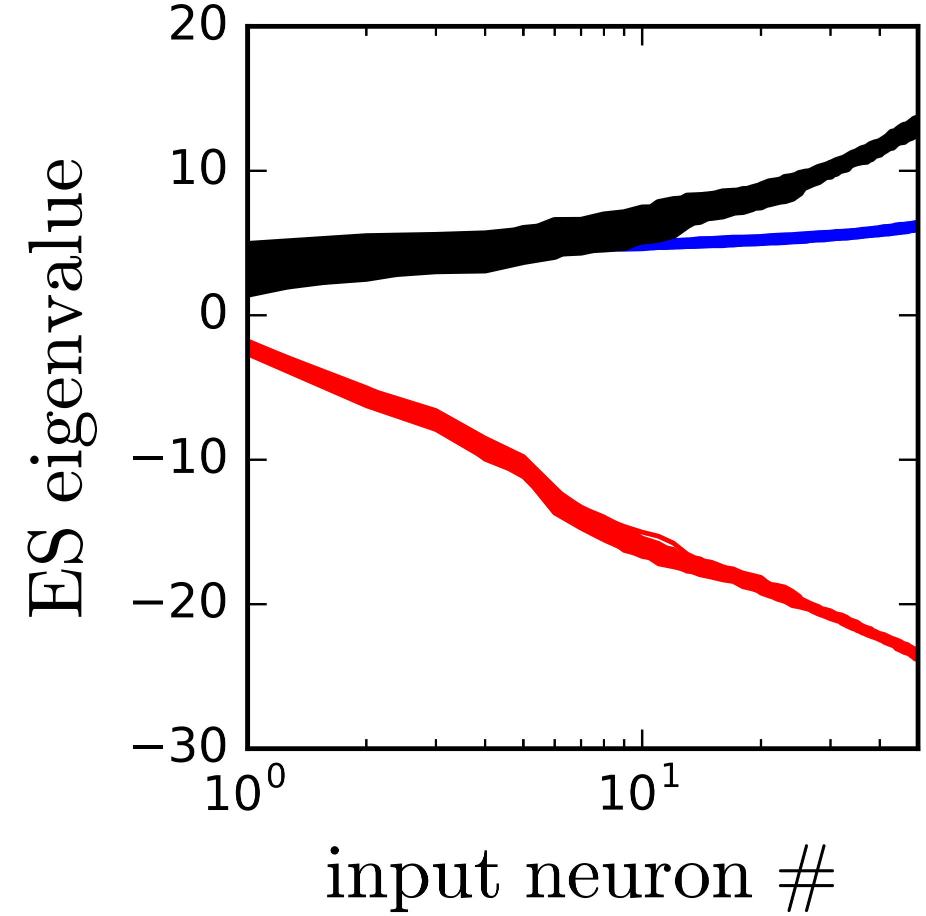


- modify inpu
  reached
- input is ins
- repeat with

- initial input



actual entanglement spectra



dreamed entanglement spectra

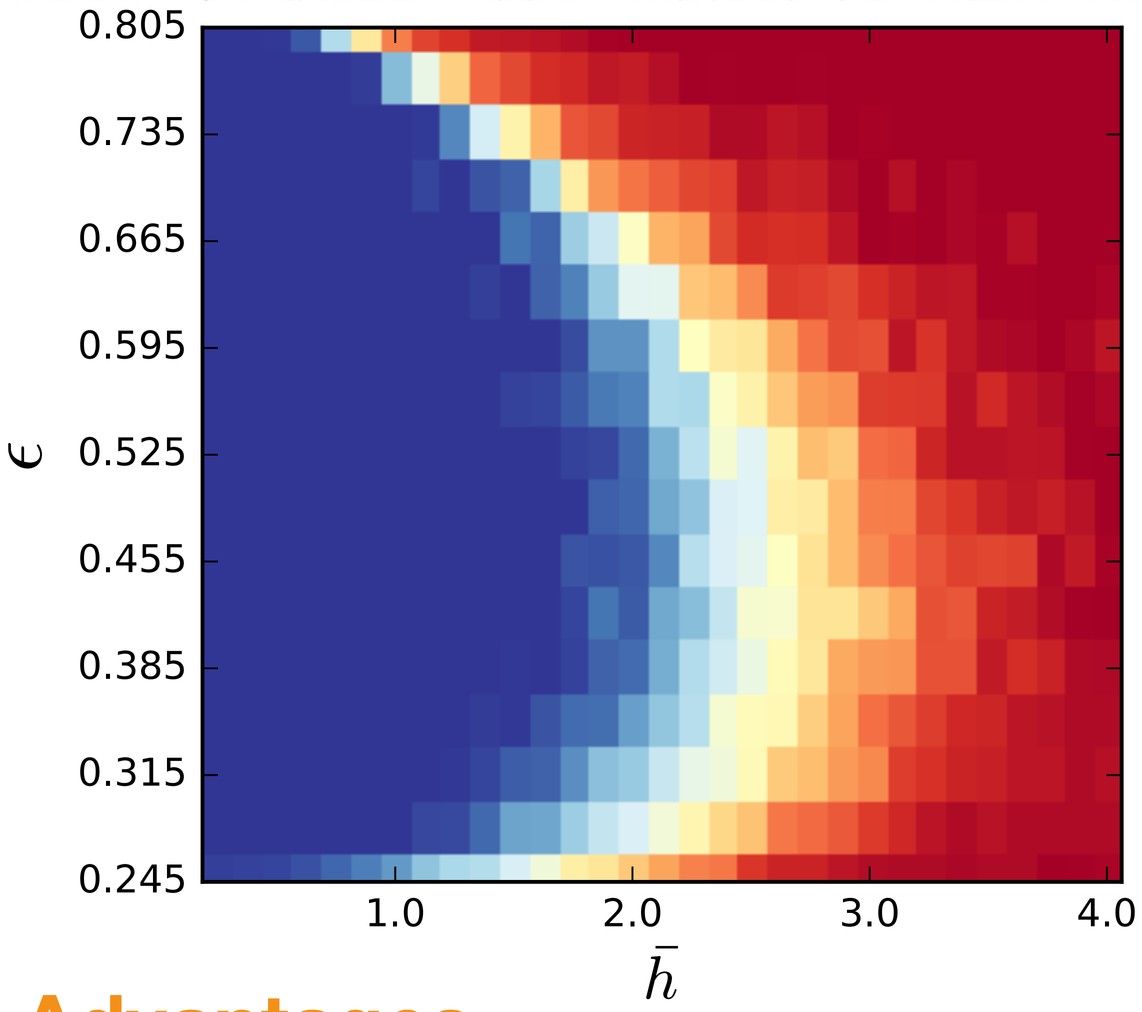
Shape reproduced, magnitude not

Reproduces power-law form of entanglement spectra

[Serbyn et al., PRL 2016]

Summary Part I

- great performance, comparable to established (physical) methods
- works with **less data** than physical quantities
- **simple and natural** choice of network and cost function; no tweaking; **confidence optimization**
- blueprint for other phase classification applications using NNs



Problems

- quantitative correctness not guaranteed
- discovery of new phases
- interpretability

Advantages

- simple and performant
- no physical insight about phase characteristics needed

University of Zurich[UZH]

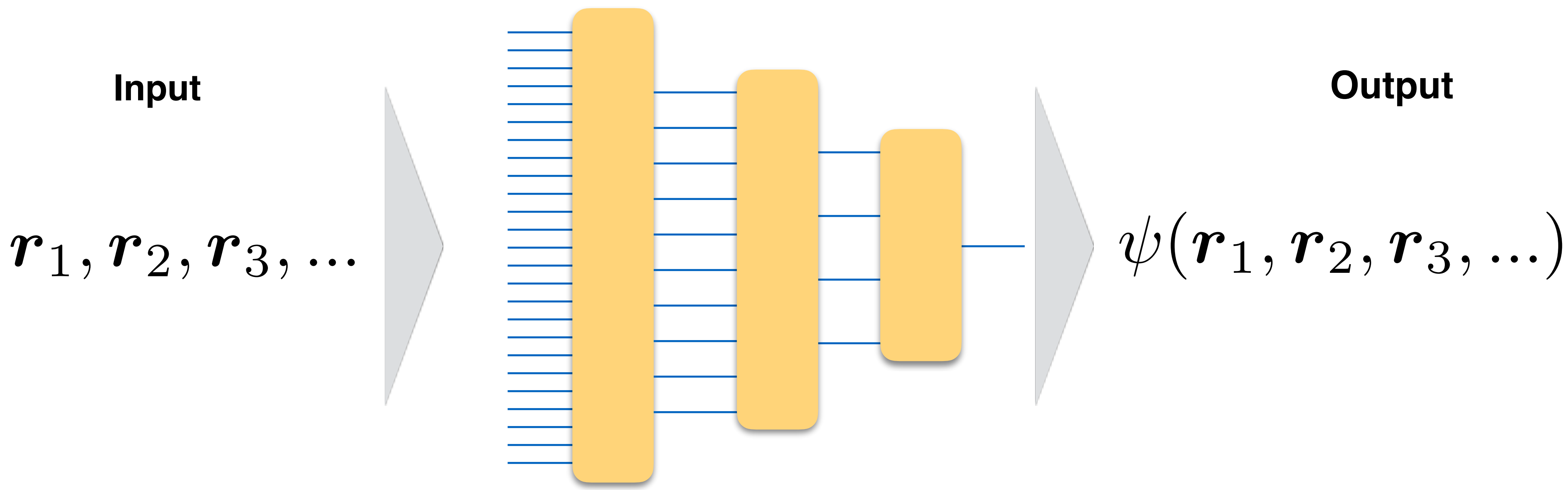
PART II

Neural networks as variational wave functions for quantum many-body problems

University of Zurich[UZH]

Explore utility of neural networks as variational wave functions

[G. Carleo and M. Troyer, Science **355** (2017)]



Network represents one (compressed) many-body quantum state

Determine eigenstates of a given Hamiltonian variationally

Promise: also works for long-range entangled states (topologically ordered, Chern insulators, chiral p-wave superconductors, …)

[D. L. Deng et al., Phys. Rev. X, Phys. Rev. X 7, 021021]

University of Zurich[UZH]

2 problems

1) Compute simultaneous eigenstates of (non-local) symmetries and of the Hamiltonian

- dispersion of excitations
- target specific excited states

2) Compute (at least low-lying) excited states

- gaps
- (topological) degeneracies

Goal: a method that would be ready to compete with ED and DMRG for **generic** problems

Network architecture

Random Boltzmann machine (one hidden layer)

$$\Psi(\boldsymbol{\sigma}) = \sum_{\boldsymbol{h}} e^{\sum_j a_j \sigma_j + \sum_i b_i h_j + \sum_{ij} h_i W_{ij} \sigma_j}$$

hidden spins

complex weights/biases

$$\log(\Psi(\boldsymbol{\sigma})) = \sum_j a_j \sigma_j + \sum_i \log\left[\cosh\left(b_i + \sum_j W_{ij}\sigma_j\right)\right]$$

no direct probabilistic interpretation due to complex numbers

Feed forward neural network with one hidden layer and log(cosh) activation function

$$\log\left(\Psi(\boldsymbol{\sigma})\right) = b + \sum_i w_i \log\left[\cosh\left(b_i + \sum_j W_{ij}\sigma_j\right)\right]$$

University of Zurich[UZH]

Problem 1) **Symmetries**

How to implement nonlocal symmetries? $\Psi(\boldsymbol{\sigma}) \longleftrightarrow \Psi(\boldsymbol{\sigma}')$

Linear operators in Hilbert space, but RBM is a nonlinear function
No natural way to extend action to hidden spins

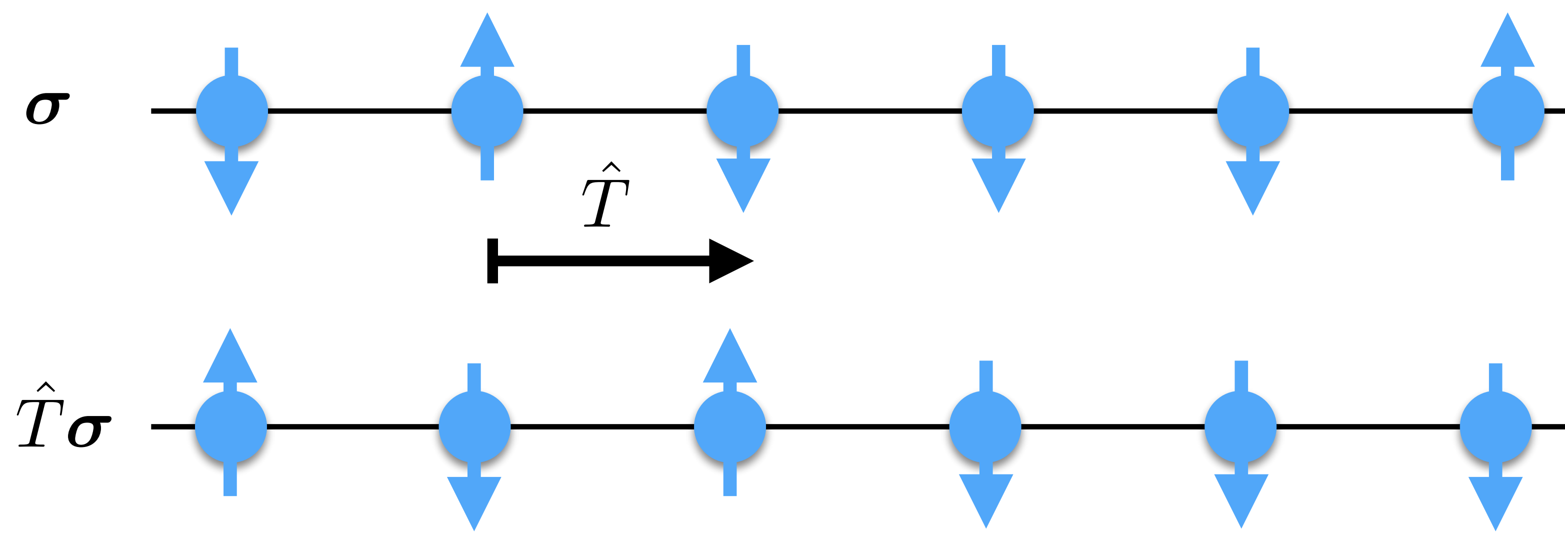
Example: **Translation symmetry** (by lattice spacing)

Eigenstate satisfies:

$$\hat{T}\,|\Psi\rangle = e^{ik}\,|\Psi\rangle$$

$$\implies \langle\boldsymbol{\sigma}|\,\hat{T}\,|\Psi\rangle = e^{ik}\,\langle\boldsymbol{\sigma}|\Psi\rangle$$

$$\implies \Psi(\hat{T}^{-1}\boldsymbol{\sigma}) = e^{ik}\Psi(\boldsymbol{\sigma}),$$



$$\boxed{\log\Psi(\hat{T}\boldsymbol{\sigma}) = ik + \log\Psi(\boldsymbol{\sigma})}$$

nonlinear constraint on weights and biases

Solution: only evaluate network in canonical configurations and compute others

$$\boldsymbol{\sigma} = (1,0,1,1,0,0) \rightarrow (0,0,1,0,1,1) = \hat{T}^2\boldsymbol{\sigma} = \boldsymbol{\sigma}_{\text{canonical}}$$

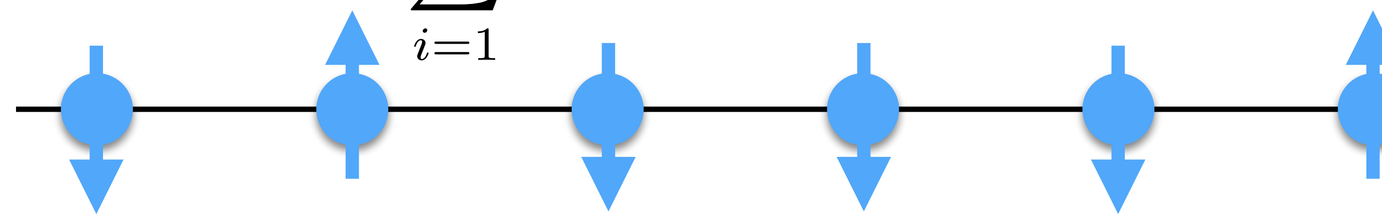
$$\boxed{\log\Psi(\boldsymbol{\sigma}) = 2ik + \log\Psi_N(\boldsymbol{\sigma}_{\text{canonical}})}$$

University of Zurich[UZH]

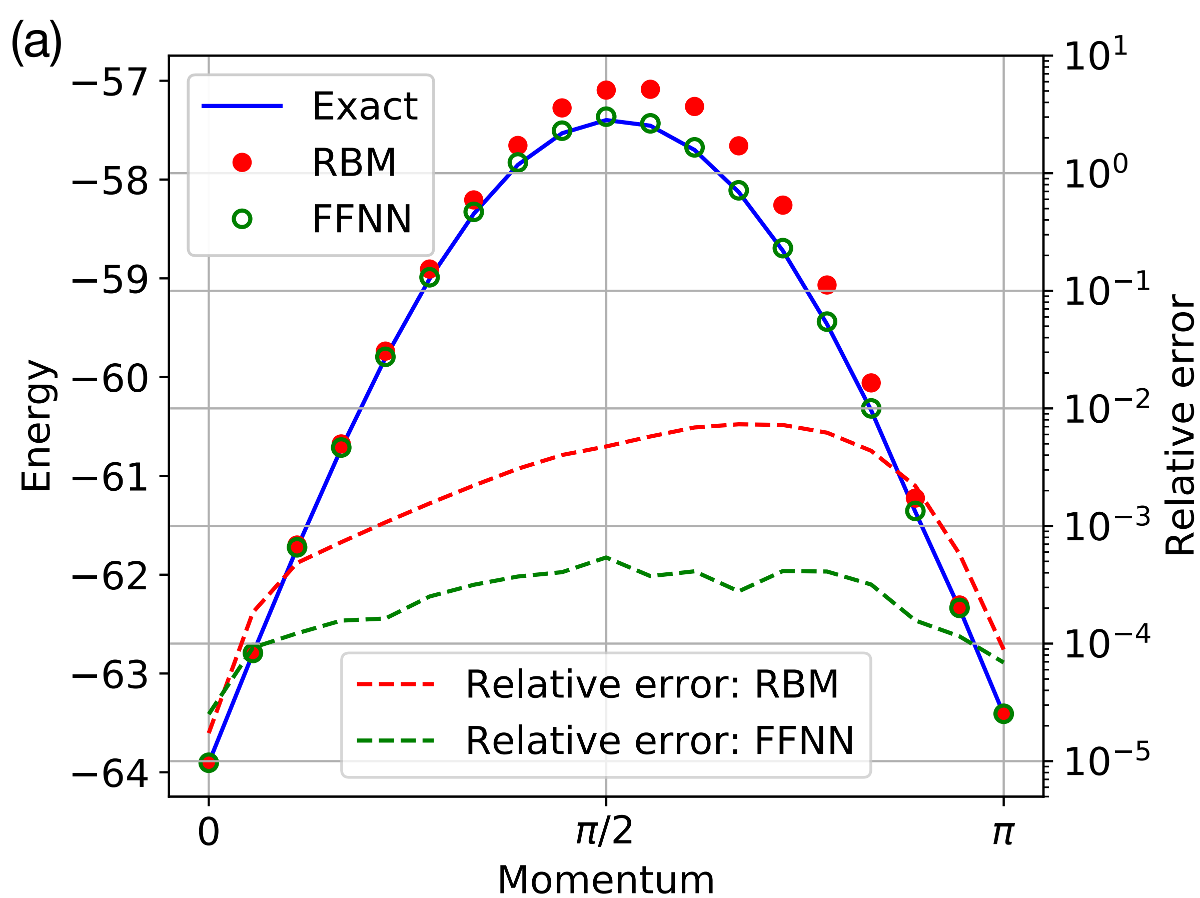
Results 1) **Symmetries**

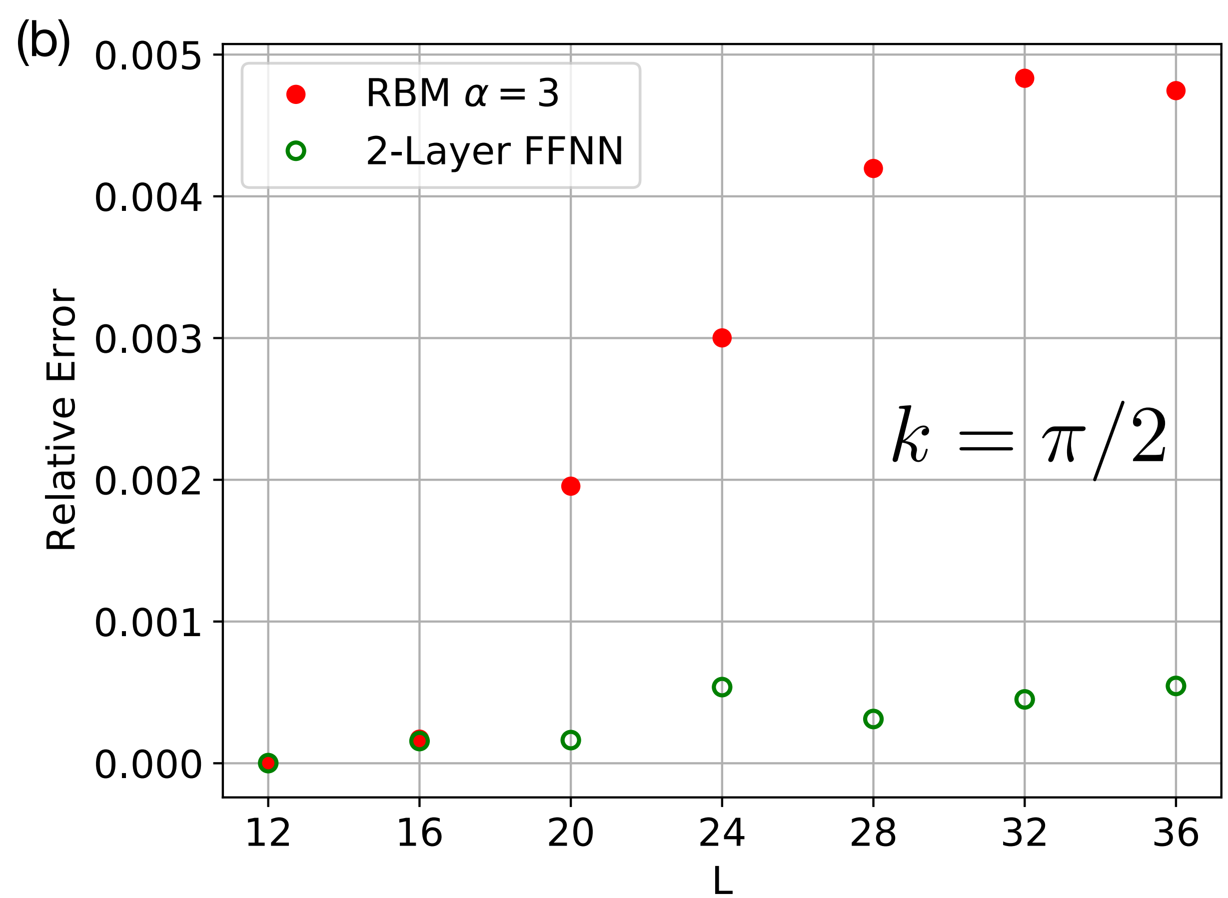
Spin-1/2 Heisenberg antiferromagnet

$$\hat{H} = 4 \sum_{i=1}^{L} \hat{\boldsymbol{S}}_i \cdot \hat{\boldsymbol{S}}_{i+1}$$



PBC, 36 sites, 72 hidden units





~4000 network parameters

vs.

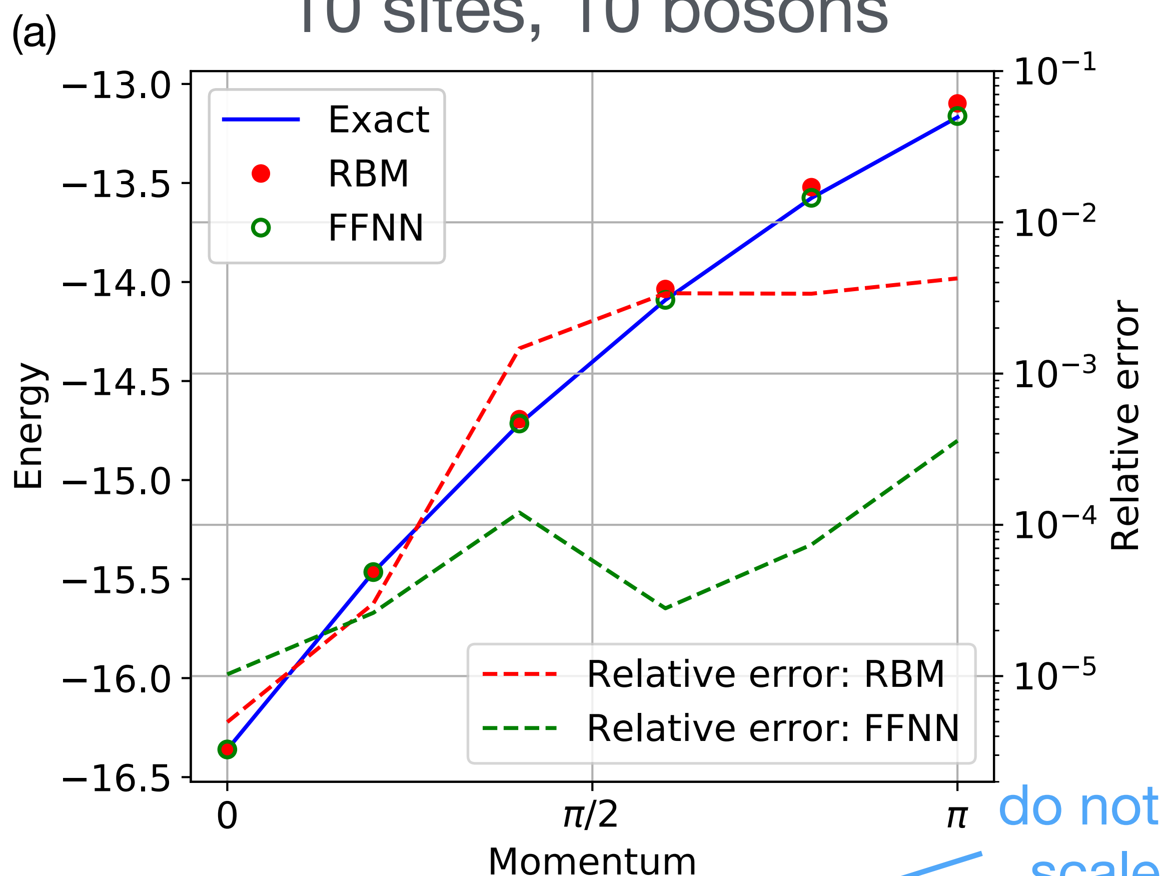
3x10⁹ parameters in ED wave function

University of Zurich[UZH]

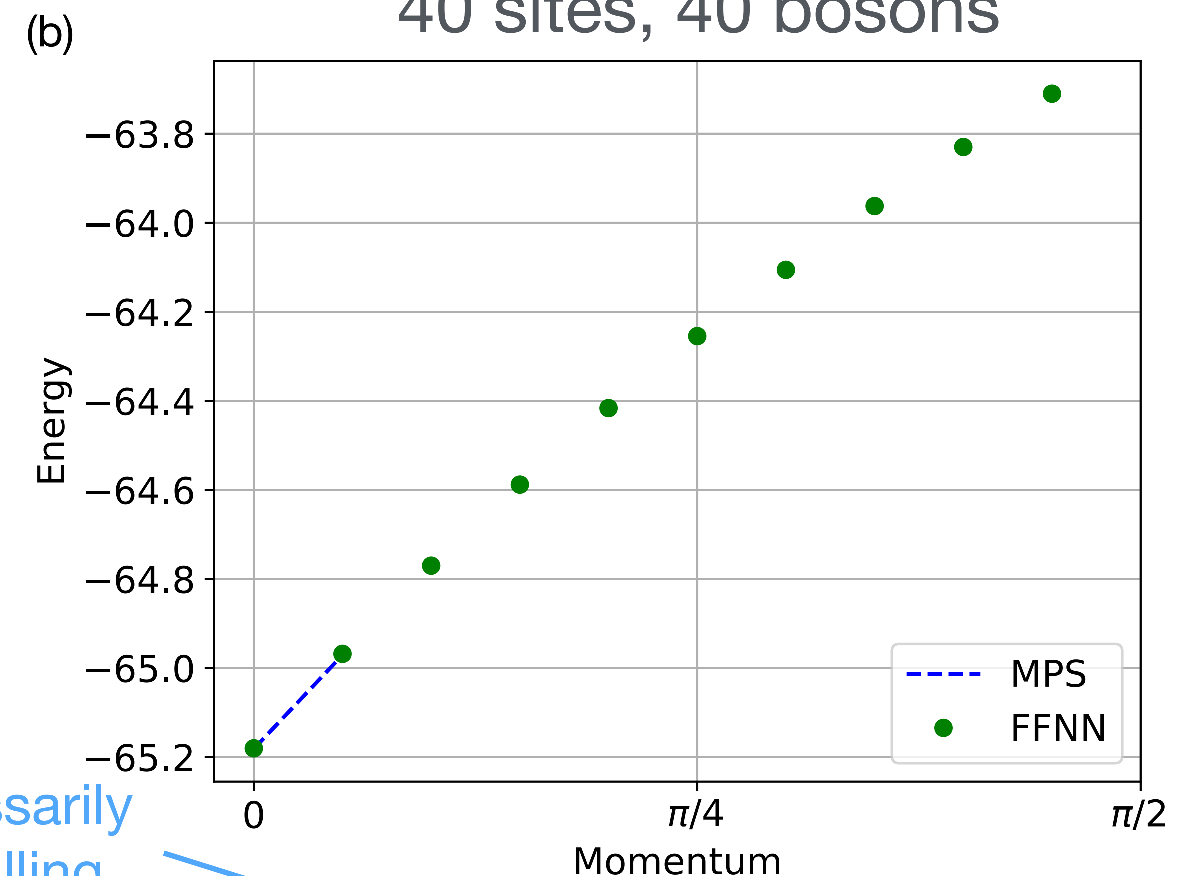
Results 1) **Symmetries**

Bose-Hubbard chain $$\hat{H} = -t \sum_{i=1}^{L} (\hat{c}_i^\dagger \hat{c}_{i+1} + \text{h.c.}) + \frac{U}{2} \sum_{i=1}^{L} \hat{n}_i (\hat{n}_i - 1)$$ U=1

10 sites, 10 bosons



40 sites, 40 bosons



do not necessarily scale with filling

~900 network parameters

vs. 1’000’000 in ED state

~6500 network parameters

vs. 5x10^{22} in ED state

University of Zurich[UZH]

Problem 2) **Excited states**

Found ground state $\Psi_0(\boldsymbol{\sigma})$, want to find lowest excited state.

$$\Psi = \Phi_1 - \lambda \Phi_0$$

$$\lambda = \frac{\langle \Phi_0 | \Phi_1 \rangle}{\langle \Phi_0 | \Phi_0 \rangle} \qquad \lambda = \sum_{\boldsymbol{\sigma}} \left(\frac{\Phi_1(\boldsymbol{\sigma})}{\Phi_0(\boldsymbol{\sigma})} \right) \frac{|\Phi_0(\boldsymbol{\sigma})|^2}{\sum_{\boldsymbol{\sigma}'} |\Phi_0(\boldsymbol{\sigma}')|^2} \approx \left\langle \frac{\Phi_1(\boldsymbol{\sigma})}{\Phi_0(\boldsymbol{\sigma})} \right\rangle_M$$

$\langle \Phi_0 | \Psi \rangle = 0$ orthogonal state

1) get trial state Φ_1

2) sample ground state wave function to compute λ

3) perform stochastic reconfiguration step with $\Psi = \Phi_1 - \lambda \Phi_0$

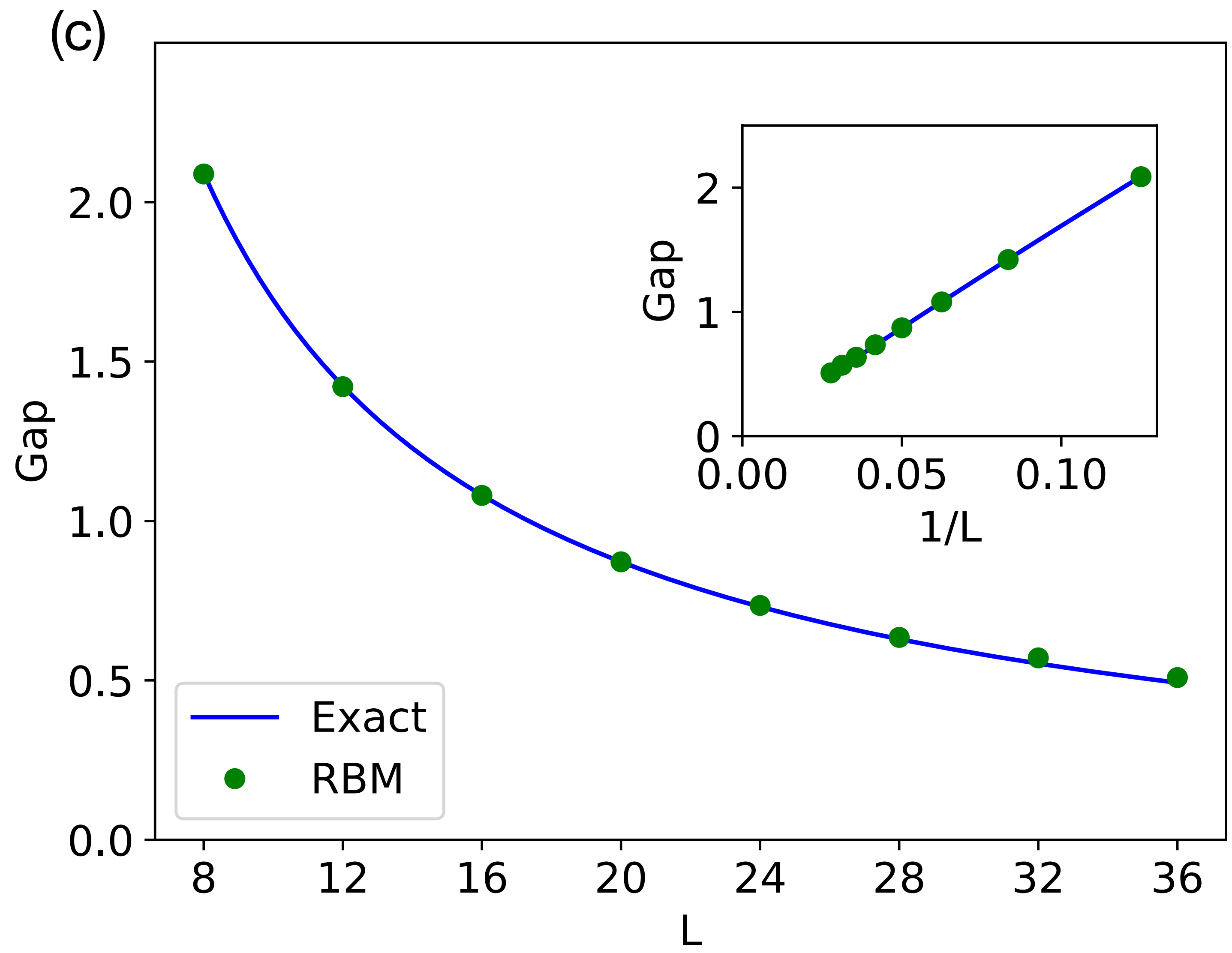
Error $\frac{\langle \Phi_0 | \Psi \rangle}{\langle \Phi_0 | \Phi_0 \rangle} \cdot \frac{\langle \Psi | \Phi_0 \rangle}{\langle \Psi | \Psi \rangle}$ (residual ground state overlap) remains below 1% for sample size of 2000.

University of Zurich[UZH]

Results 2) **Excited states**

Spin-1/2 Heisenberg antiferromagnet

$$\hat{H} = 4\sum_{i=1}^{L} \hat{\boldsymbol{S}}_i \cdot \hat{\boldsymbol{S}}_{i+1}$$

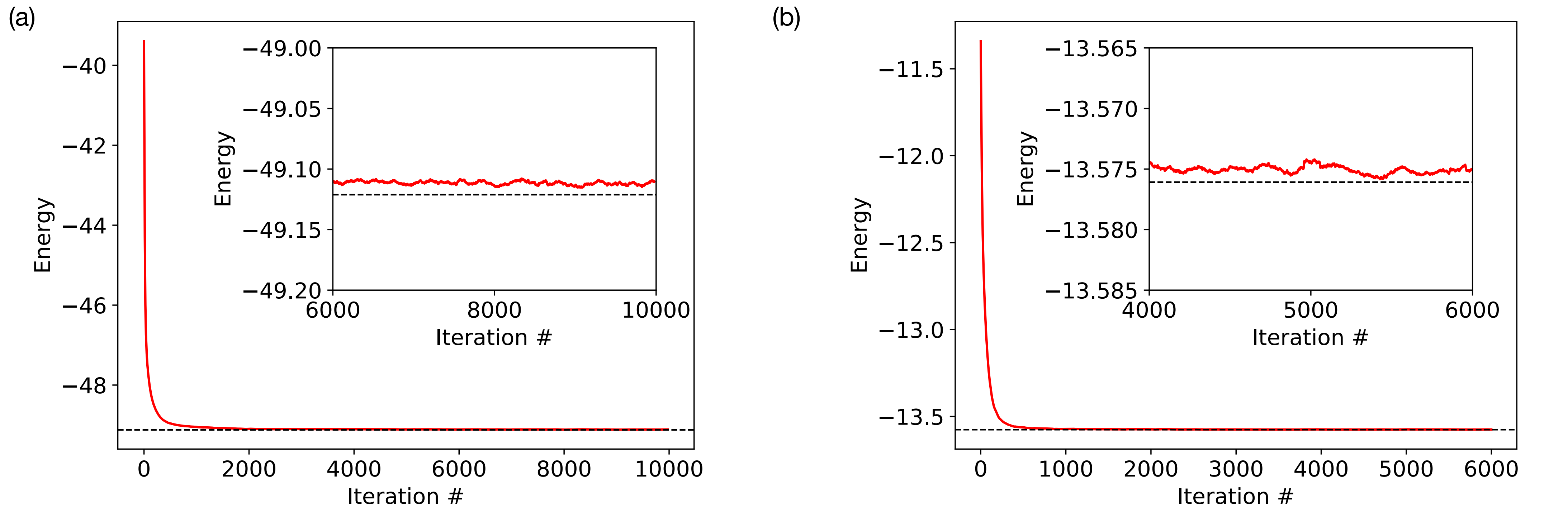


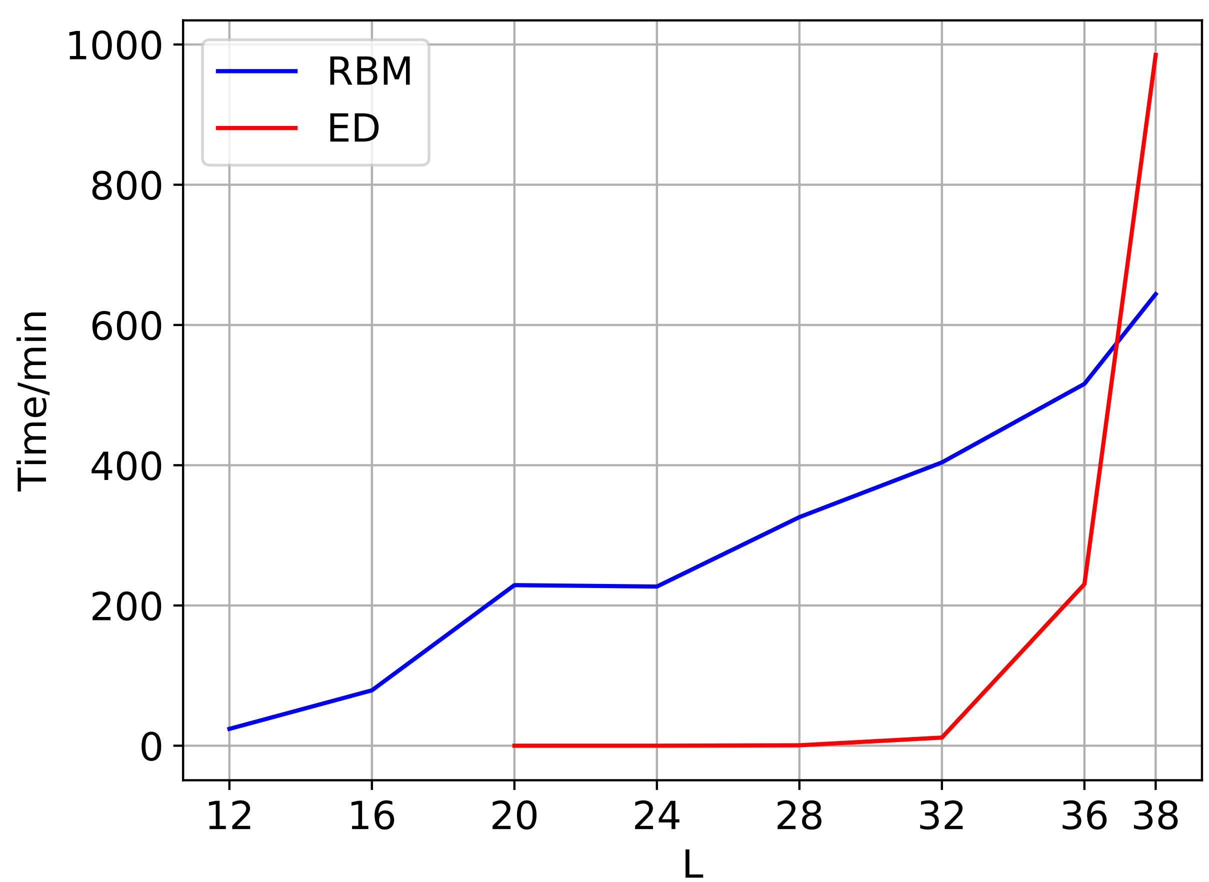
4L (2L) hidden units in ground (excited) state

University of Zurich[UZH]

Scaling and performance

Typical convergence behavior





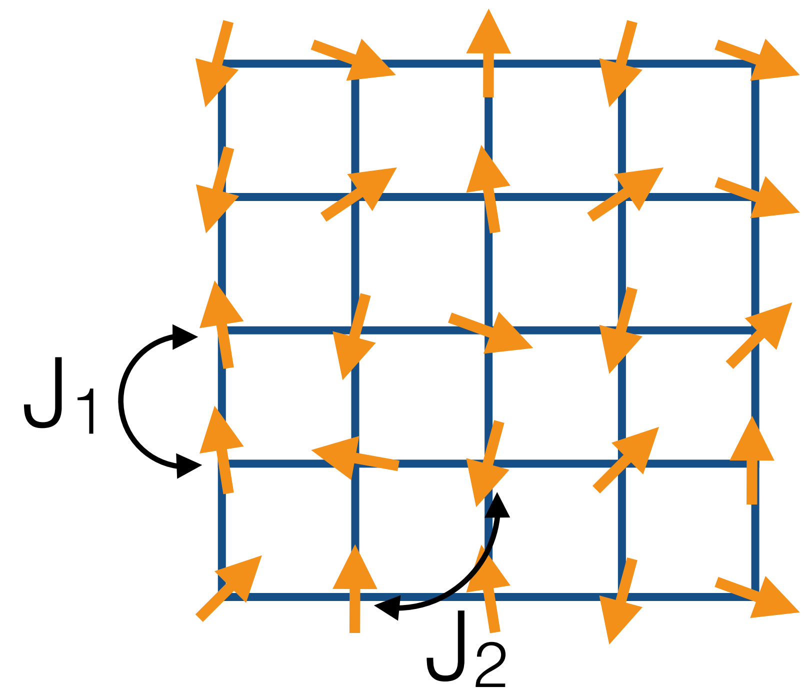
Spin-1/2 Heisenberg antiferromagnet

hard to compare computational cost
here: CPU wall-times

University of Zurich[UZH]

2D frustrated magnets: J_1-J_2 model on square lattice

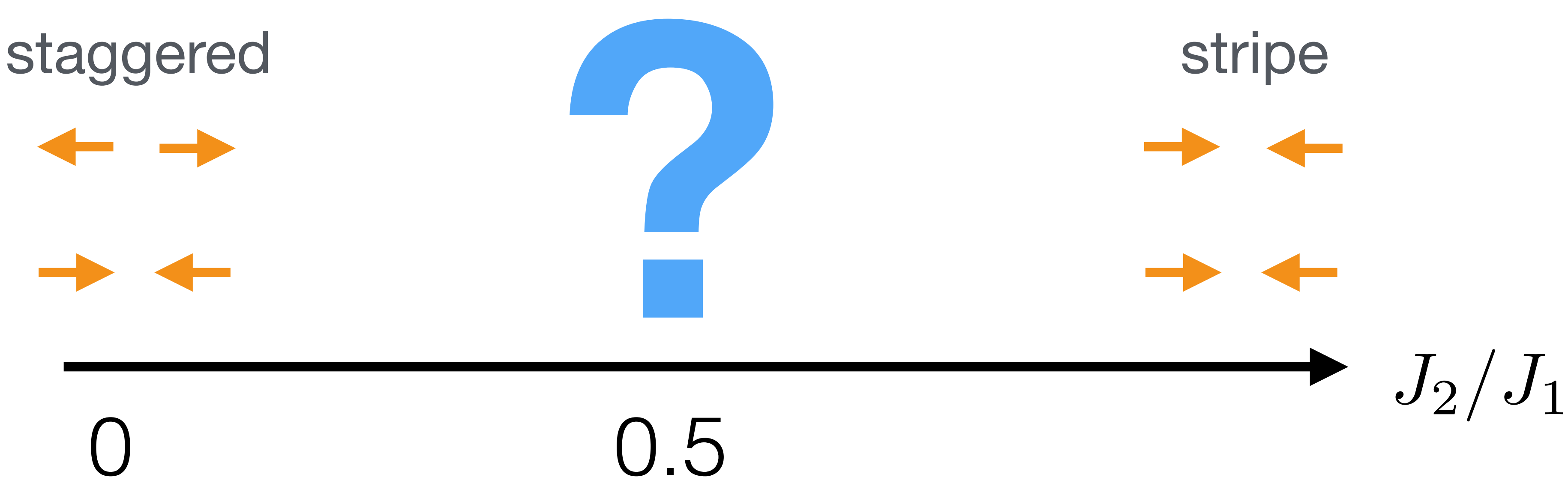
$$H = J_1 \sum_{\mathrm{NN}} \boldsymbol{S}_i \cdot \boldsymbol{S}_j + J_2 \sum_{\mathrm{NNN}} \boldsymbol{S}_i \cdot \boldsymbol{S}_j$$



Spin-1/2 Heisenberg model

highly frustrated for $J_1 \sim J_2$; sign problem

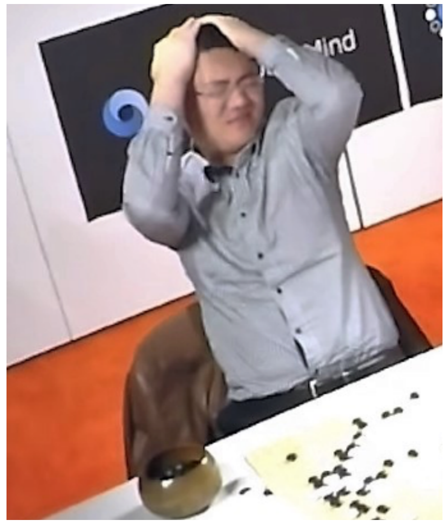
extensively studied (ED, DMRG, VMC, …)



University of Zurich[UZH]

The Go challenge

October 2015
2nd dan



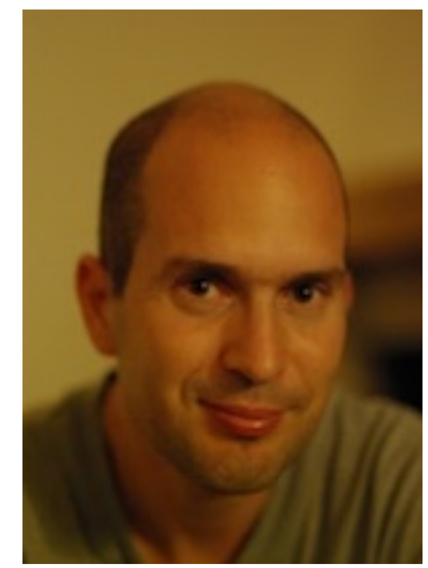
March 2016
9th dan

AlphaGo

May 2017
world champion

J_1-J_2 challenge

ED

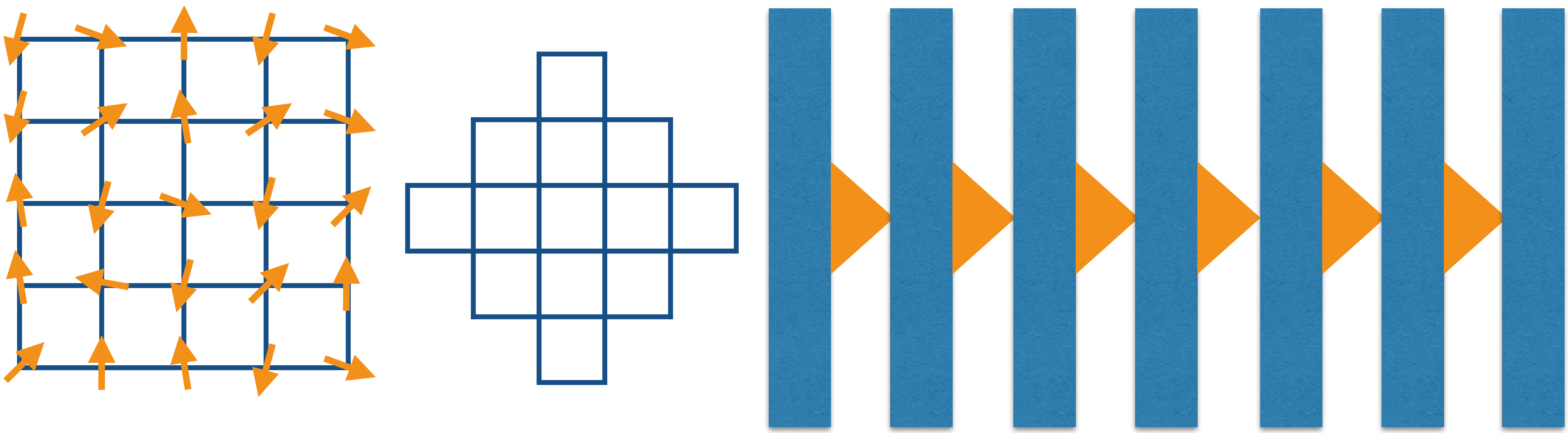


VMC

DMRG

University of Zurich[UZH]

Convolutional, complex, deep



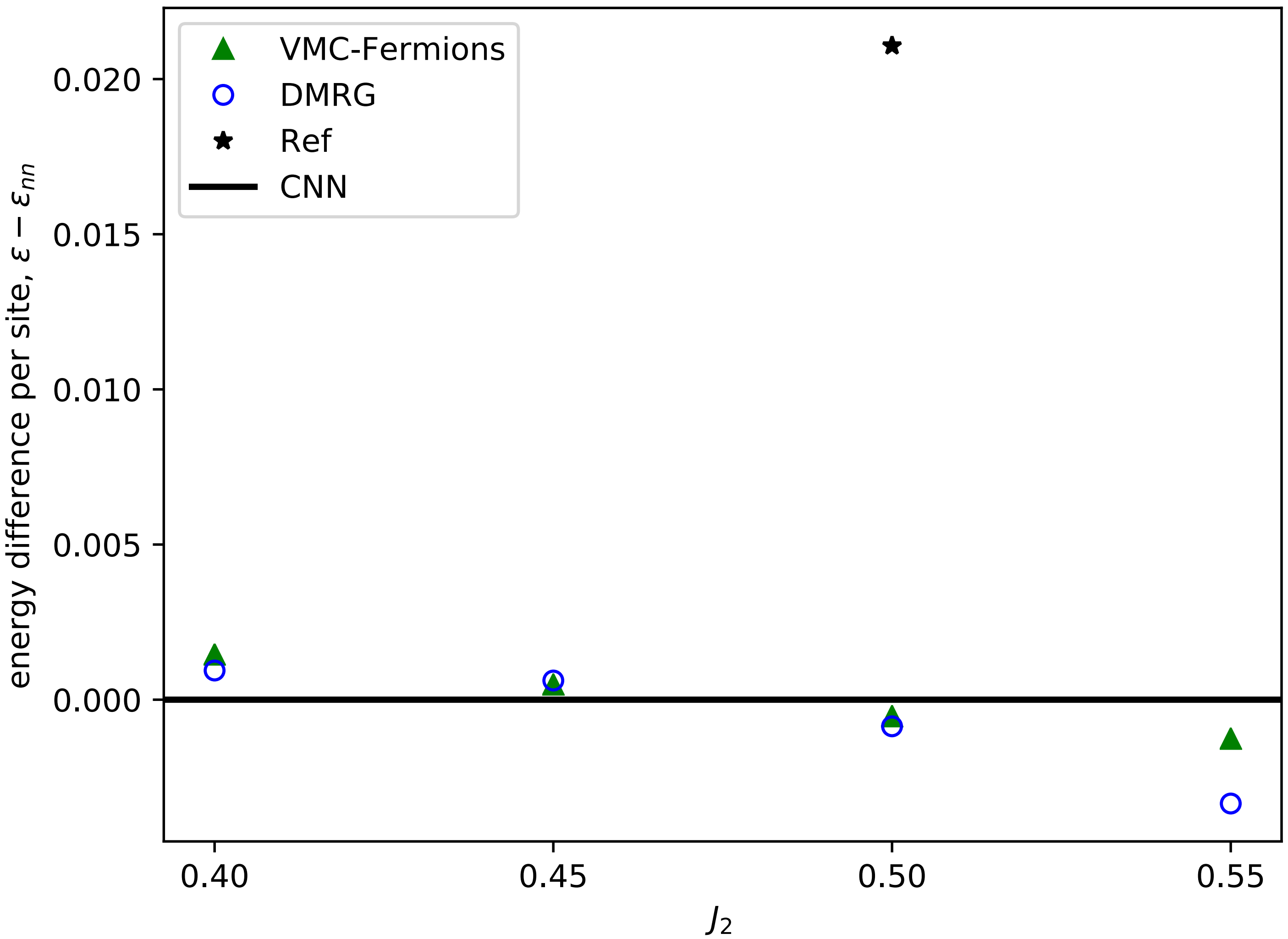
No of filters: 12 10 8 6 4 2 avg pooling

~ **3000 parameters** independent of system size

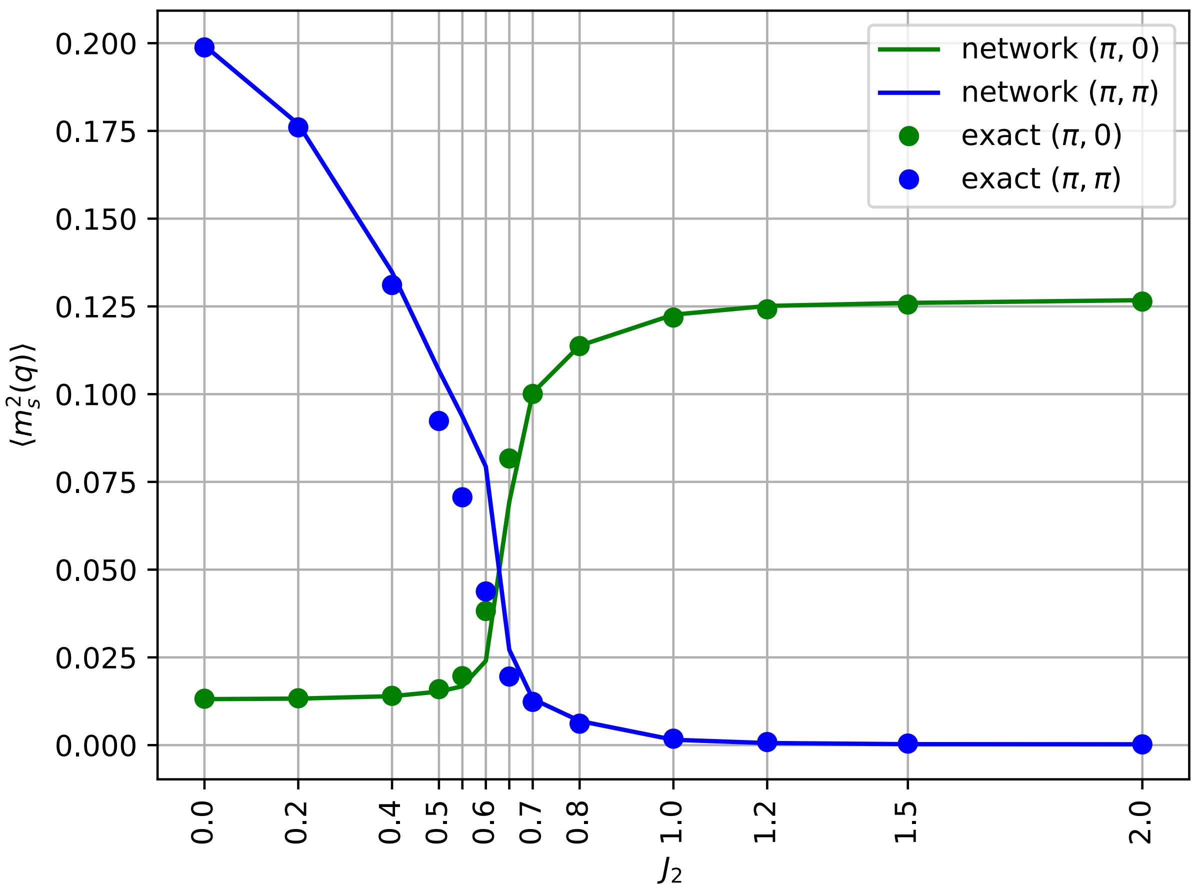
University of Zurich[UZH]

Results

ENERGIES
10x10 lattice



ORDER PARAMETER
6x6 lattice



VMC [W.-J. Hu, F. Becca, A. Parola, and S. Sorella, PRB 88, 060402 (2013)]
DMRG [S.-S. Gong, W. Zhu, D. N. Sheng, O. I. Motrunich, and M. P. A. Fisher, PRL 113, 027201 (2014)]
NN [X. Liang, W.-Y. Liu, P.-Z. Lin, G.-C. Guo, Y.-S. Zhang, and L. He, PRB 98, 104426 (2018)] (10000 parameters)

Best energies in small or large J_2/J_1
worse in the middle

University of Zurich[UZH]

Comparison for 10x10 lattice

ED
completely unbiased
2^{100} parameters, 2^{60} terrabyte
1 lightyear stack of hard disks

DMRG
entanglement bias
universal ansatz
O(million) parameters

VMC
physics-inspired, problem-specific ansatz
few parameters

Neural network states
unknown bias
universal ansatz
~3000 parameters

University of Zurich[UZH]

Summary Part II

- Neural networks: powerful class of variational quantum states
- FFNN better than RBM
- deeper, convolutional is better
- implementation of nonlocal symmetries
- access to low-lying excited states
- depending on model and regime: competitive with established techniques

University of Zurich[UZH]

PART III Quantum machine learning

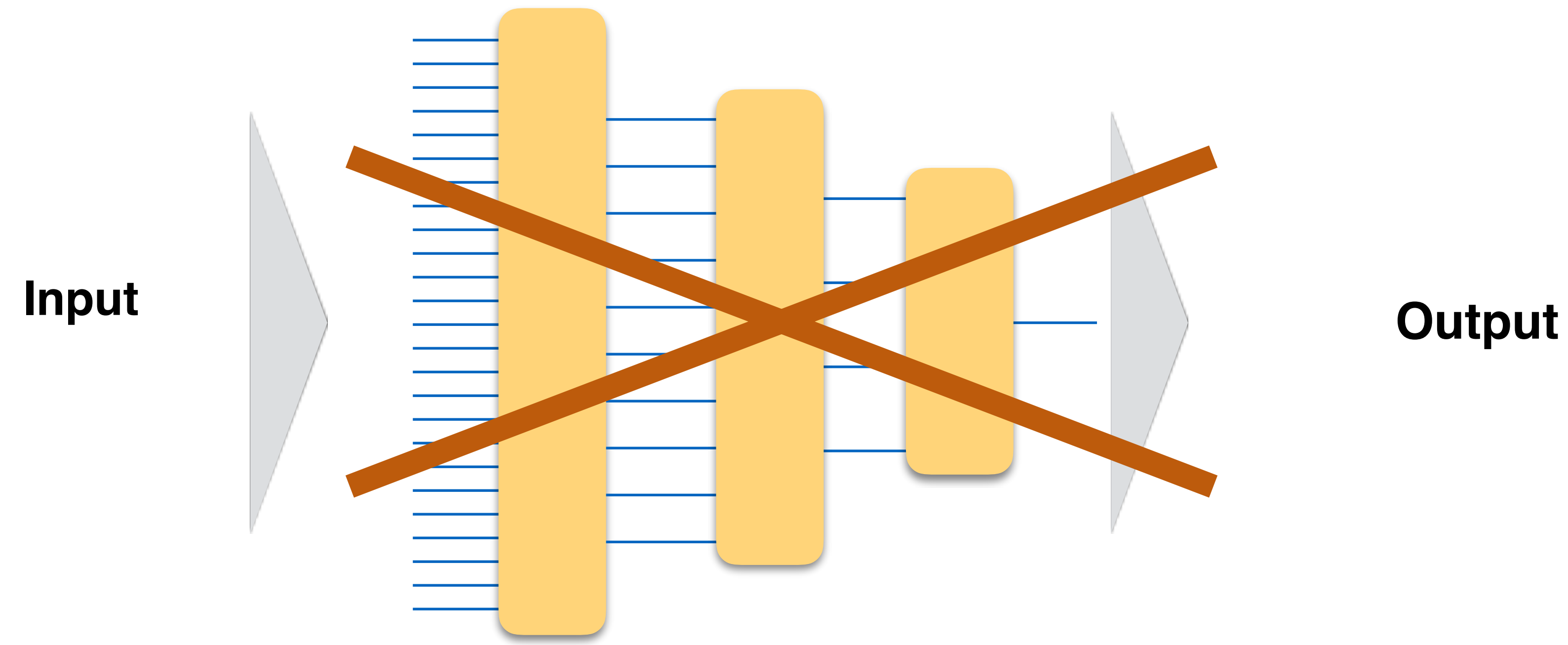
University of Zurich[UZH]

Goal: Use quantum architectures for machine learning tasks

Fundamental difference:

- neural networks are **nonlinear**
- quantum evolution is **unitary (=linear)**

Nonlinearity through **measurement step**



University of Zurich[UZH]

Network architecture

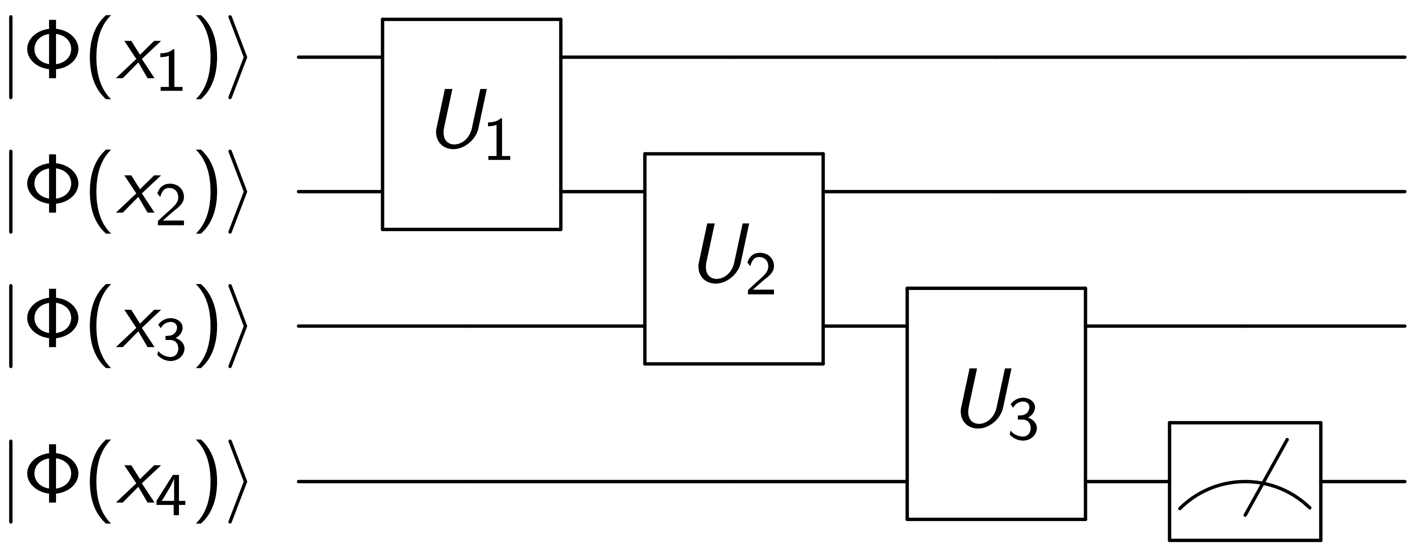
Inspired by **matrix product states**

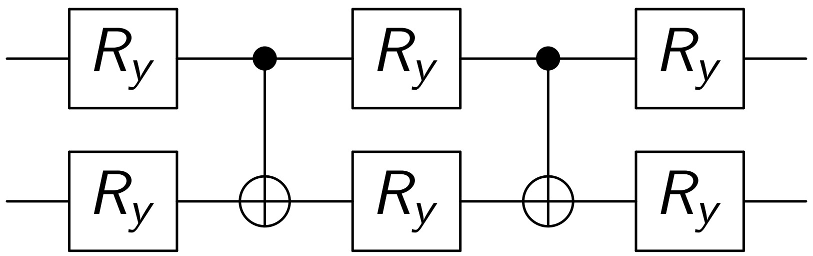
[M. Stoudenmire and D. J. Schwab, arXiv: 1605.05775.]
[I. Glasser, N. Pancotti, J I. Cirac, arXiv:1806.05964.]

Inscribe data in initial state (only real wave functions):

$$|\Phi(\boldsymbol{x})\rangle = \begin{bmatrix}\cos\left(\frac{\pi}{2}x_1\right)\\ \sin\left(\frac{\pi}{2}x_1\right)\end{bmatrix} \otimes \begin{bmatrix}\cos\left(\frac{\pi}{2}x_2\right)\\ \sin\left(\frac{\pi}{2}x_2\right)\end{bmatrix} \otimes \cdots \otimes \begin{bmatrix}\cos\left(\frac{\pi}{2}x_N\right)\\ \sin\left(\frac{\pi}{2}x_N\right)\end{bmatrix}$$

Network of successively applied **unitaries**

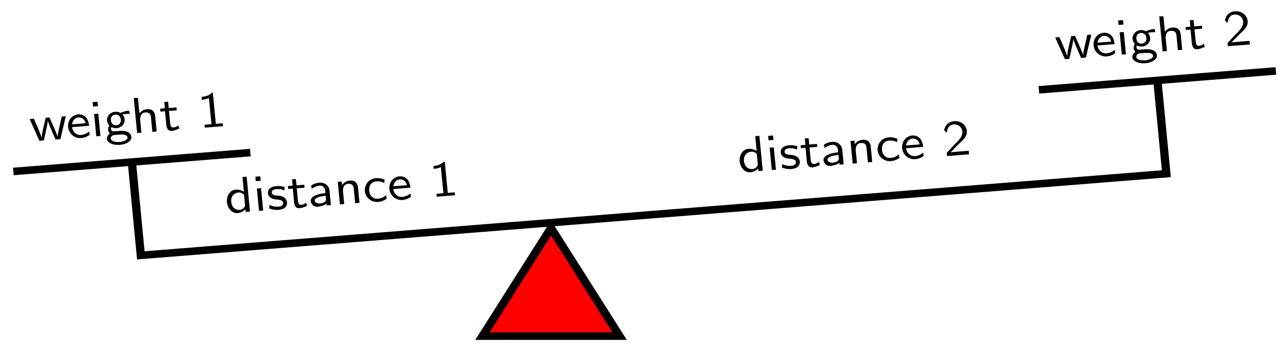




6 free parameters per unitary

University of Zurich[UZH]

Toy problem: Balance

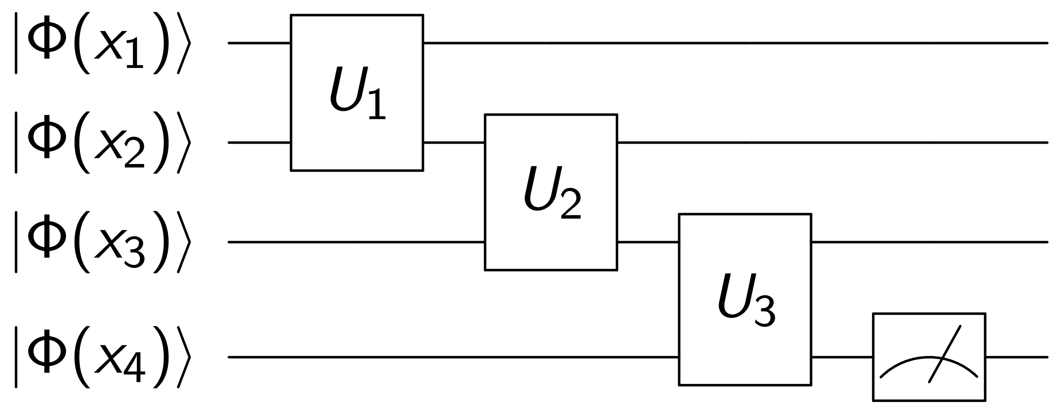


Training data:

arm length and weight of a scale

Label:

scale tips left or right



Training on classical computer

	training set	test set
accuracy	0.97	0.95

Performance of trained network on IBM Q 20 Tokyo

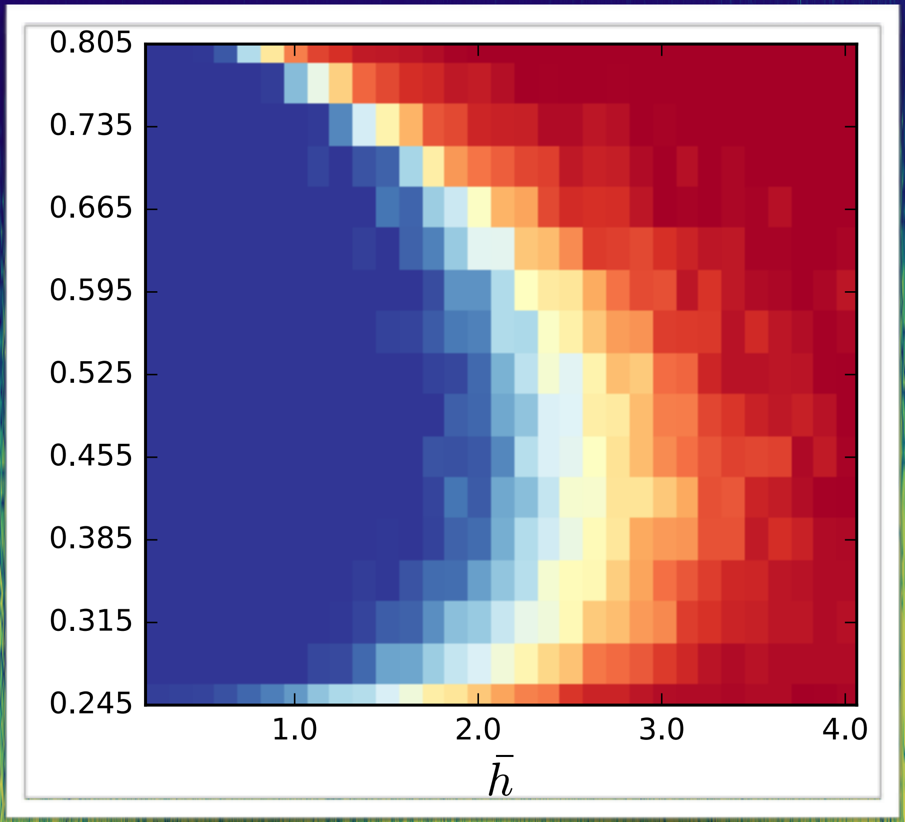
	measured on quantum computer	predicted
accuracy on test set	0.94	0.95
loss on test set	0.031	0.023

University of Zurich[UZH]

Summary

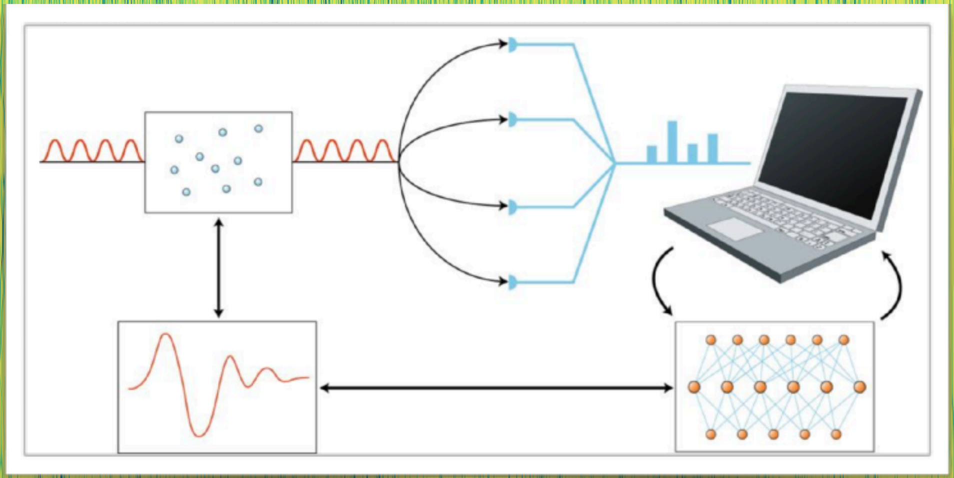
PI: Phase classification

- NN are performant aids for some tasks
- interpretability/scientific rigor biggest challenge
- performant even with small input



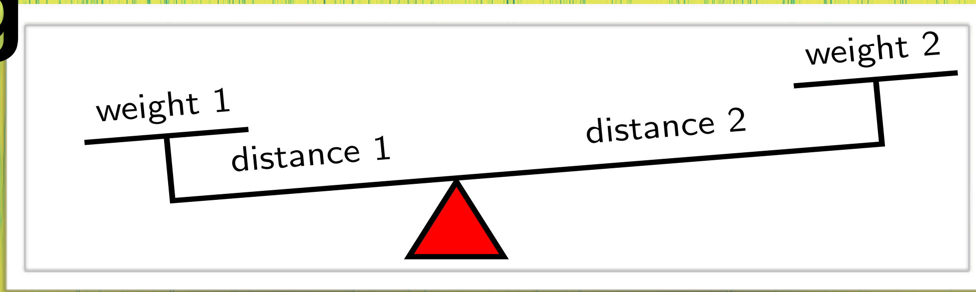
PII: Variational Wave functions

- potentially powerful new approach to many-body quantum systems
- companion tool for quantum simulators



PIII: Quantum Machine Learning

- promising short-term application for analogue quantum computers due to statistical nature
- no rigorous performance results



University of Zurich[UZH]