Seminar: "Solving the quantum many-body problem with artificial neural networks"

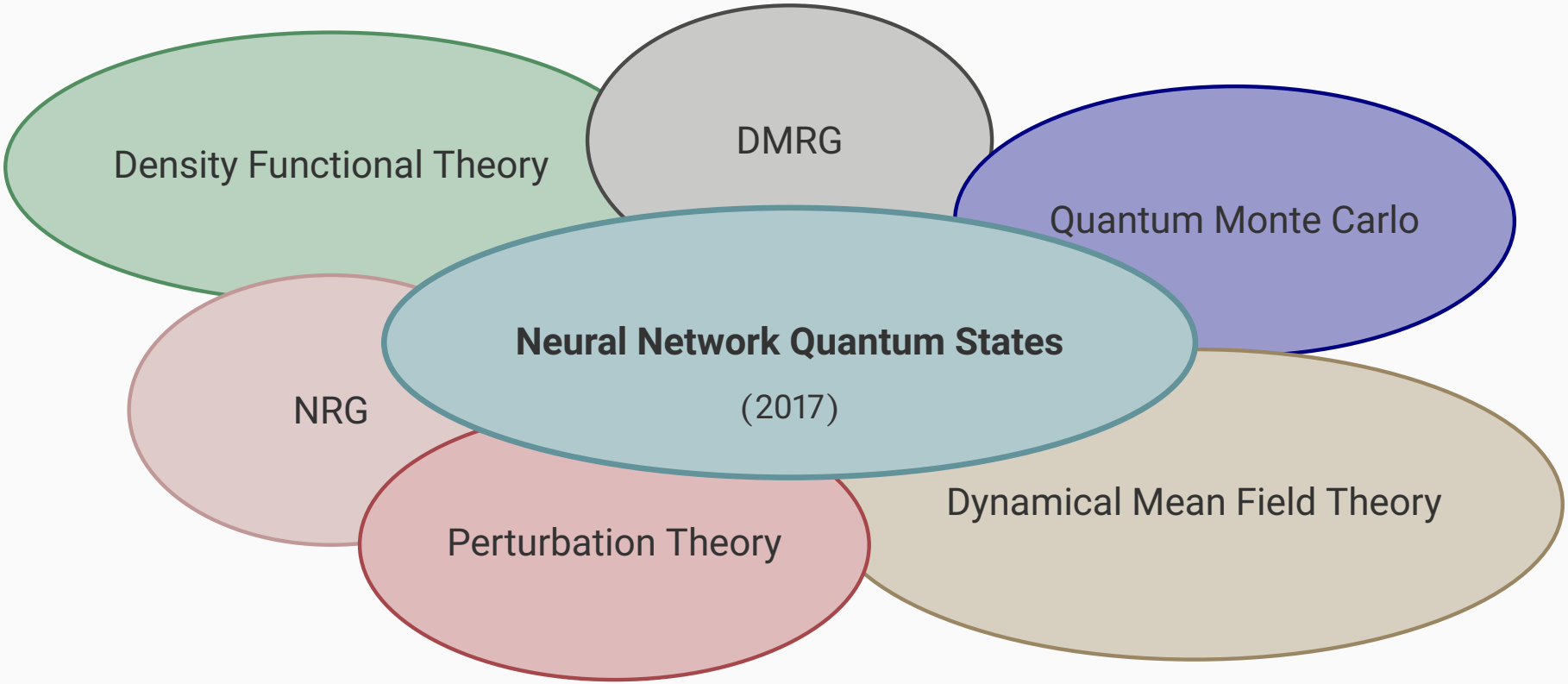
Ana Luiza R. F. Ferrari

Supervised by: Chae-Yeun Park

June 3, 2020

The many body problem

- The Hilbert space of a many-body problem grows exponentially with size.
- Analytic calculations of the wavefunction are essentially impossible.
- Approximations are used to describe these systems.



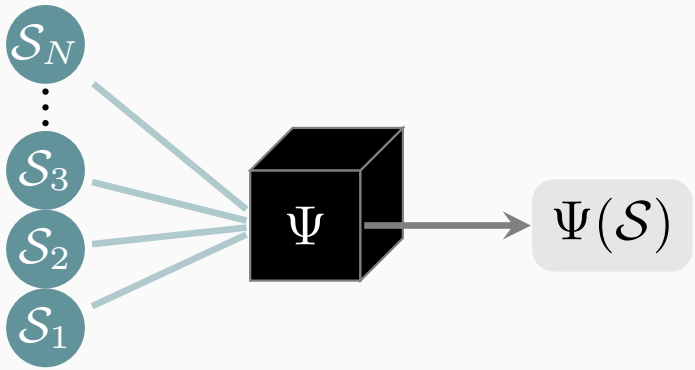
MANY-BODY PHYSICS

Solving the quantum many-body problem with artificial neural networks

Giuseppe Carleo[1]* and Matthias Troyer[1,2]

The challenge posed by the many-body problem in quantum physics originates from the difficulty of describing the nontrivial correlations encoded in the exponential complexity of the many-body wave function. Here we demonstrate that systematic machine learning of the wave function can reduce this complexity to a tractable computational form for some notable cases of physical interest. We introduce a variational representation of quantum states based on artificial neural networks with a variable number of hidden neurons. A reinforcement-learning scheme we demonstrate is capable of both finding the ground state and describing the unitary time evolution of complex interacting quantum systems. Our approach achieves high accuracy in describing prototypical interacting spins models in one and two dimensions.

Formulating the ML problem



Goal

Encode the ground state wave-function $\Psi(\mathcal{S})$ of a many-body system with $\mathcal{S} = (\mathcal{S}_1, \ldots, \mathcal{S}_N)$ in a neural network.

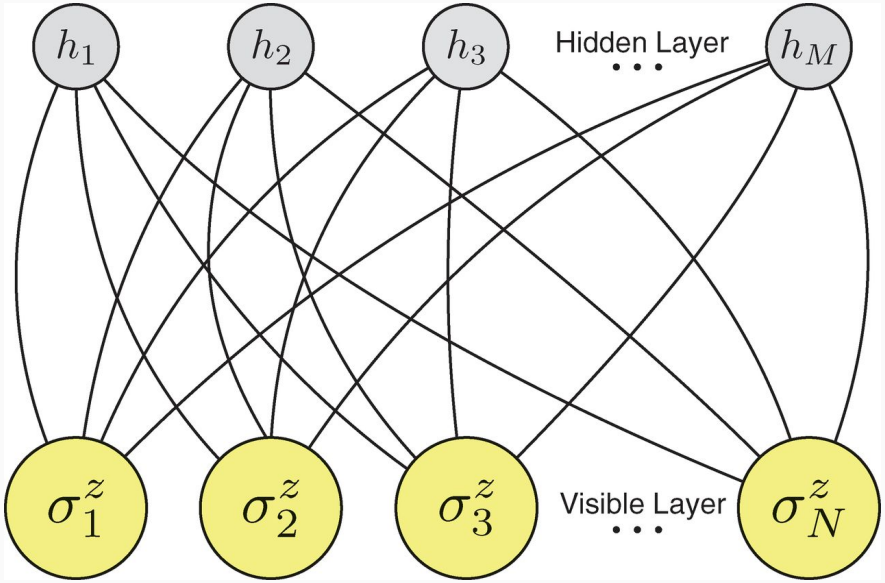
Machine learning problems can be formulated in two parts:

1. Machine: We encode a high dimensional function $\mathbf{F}_\lambda(\mathbf{x})$ in a neural network, which depends on the input *x* and on the network parameters λ.
2. Learning: We look for a set of optimal parameters λ^* which minimizes a cost function $\mathbf{C}_\lambda$ given a data set.

Architecture Restricted Boltzmann Machine

- N visible nodes: $x = \{\sigma_1^z, \ldots, \sigma_N^z\}$;
- One hidden layer with the auxiliary spin variables $h = \{h_1, \ldots, h_M\}$;
- The network parameters are complex $\Rightarrow$ **they describe the phase and amplitude of the wave function.**

$$\Psi_\lambda(x, \lambda) = \sum_{\{h_i\}} e^{\sum_j a_j \sigma_j^z + \sum_i b_i h_i + \sum_{i,j} W_{i,j} h_i \sigma_j^z}$$



- Quality: systematically improved by increasing the number of hidden variables M.;
- The hidden layer links all sites $\Rightarrow$ **non local correlations**.
- **RBMs have no intralayer interaction:** the hidden variables can be traced out

$$\Psi_\lambda(x,\lambda) = \sum_{\{h_i\}} e^{\sum_j a_j\sigma_j^z + \sum_i b_i h_i + \sum_{i,j} W_{i,j} h_i \sigma_j^z}$$
$$= e^{\sum_j a_j \sigma_j^z} \times \prod_{i=1}^{M} F_i(S),$$

where

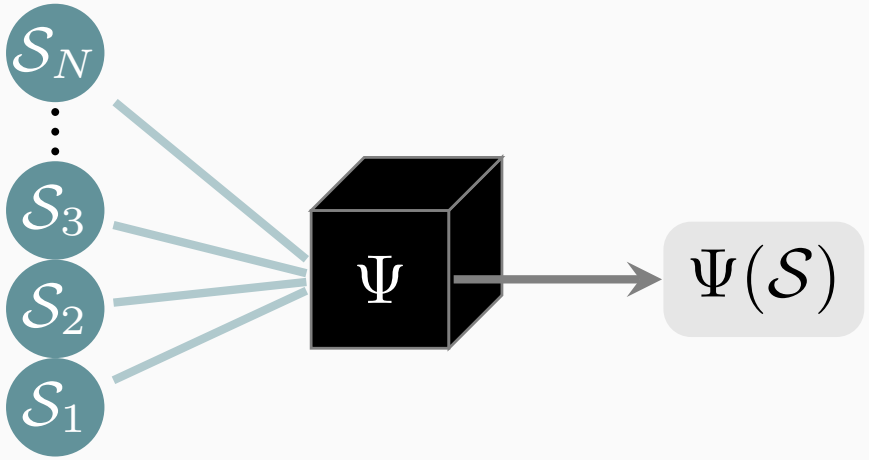
$$F_i(S) = 2\cosh\left(b_i + \sum_j W_{ij}\sigma_j^z\right)$$

Generative Modelling vs. NQS

Rajat's presentation

	Generative modelling	NQS
Encoded function	$\mathcal{P}_{\mathcal{W}}(x)$	$\Psi_{\mathcal{W}}(x)$
Cost function	Kullback-Leibler divergence	Ground state energy functional
RBM	$\mathcal{W} \in \mathbb{R}$	$\mathcal{W} \in \mathbb{C}$
Minimization method	Gradient descent	Stochastic reconfiguration

Samples and learning method



- The many-body state is unknown **(it is what we want to determine)**.
- We don't have samples of the exact wave function Ψ.

Supervised learning is not possible here!

Consistent reinforcement learning (Variational Monte Carlo)

Cost function Variational principle

- The variational principle gives us an upper bound for the ground state energy

$$E[\Psi] = \langle \Psi | \mathcal{H} | \Psi \rangle \geq E_0$$

- It is possible to find the ground state by minimizing the energy functional.

$$\Psi_0 = \text{argmin}_\Psi E[\Psi] \xrightarrow{\text{Ansatz: } |\Psi_\lambda\rangle} \Psi_0 = \text{argmin}_\lambda E[\Psi_\lambda]$$

Optimization problem: the energy functional is the cost function.

Variational Quantum Monte Carlo Expectation Values

- We can "prepare" expectation values for a **Monte Carlo** application in the following way

$$\langle \mathcal{O} \rangle = \frac{\langle \Psi | \mathcal{O} | \Psi \rangle}{\langle \Psi | \Psi \rangle} \xrightarrow{\text{Completeness}} \langle \mathcal{O} \rangle = \frac{\sum_x \langle \Psi | x \rangle \langle x | \mathcal{O} | \Psi \rangle}{\sum_x \langle \Psi | x \rangle \langle x | \Psi \rangle}$$

$$\langle \mathcal{O} \rangle = \frac{\sum_x |\Psi(x)|^2 \mathcal{O}_L(x)}{\sum_x |\Psi(x)|^2}$$

$$\mathcal{O}_L(x) = \sum_{xx'} \mathcal{O}_{xx'} \frac{\Psi(x')}{\Psi(x)}$$

- Thus, we identify a probability distribution for the operator $\mathcal{O}_L$

$$\langle \mathcal{O} \rangle = \sum_x P(x) \mathcal{O}_L(x) \quad , \quad P(x) = \frac{|\Psi(x)|^2}{\sum_x |\Psi(x)|^2}$$

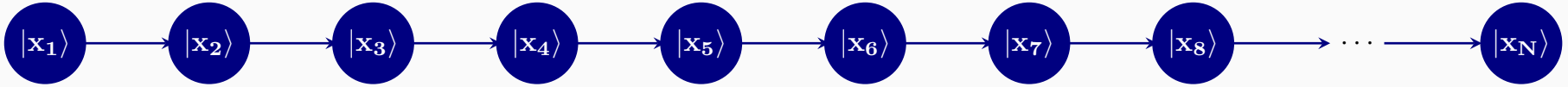
- Now, we can use the Metropolis Hastings algorithm to sample configurations from $P(x)$ and to calculate the expectation values.

Expectation Values

$$\langle \mathcal{O} \rangle = \frac{\langle \Psi | \mathcal{O} | \Psi \rangle}{\langle \Psi | \Psi \rangle} \Longrightarrow \mathbb{E}[\mathcal{O}_L] = \frac{1}{N_s} \sum_{k=1}^{N_s} \mathcal{O}_L(x_k)$$

- The modified operators $\mathcal{O}_L$ are **Local operators**.
- In particular, we will use the local energy

$$\mathcal{H}_L(x) = \sum_{x'} \mathcal{H}_{xx'} \frac{\Psi(x')}{\Psi(x)}$$

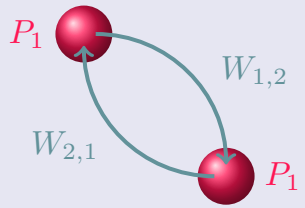


Given a configuration $|x_j\rangle$, construct a suggestion for the next configuration $|x_i\rangle$ via some change of $|x_j\rangle$ (e.g. spin flip)

Calculate the ratio of probabilities $\mathcal{P} = \frac{p(x_i)}{p(x_j)} = \frac{|\Psi(x_i)|^2}{|\Psi(x_j)|^2}$

`if` $P > 1$: $|x_j\rangle \rightarrow |x_i\rangle$
`else`: $|x_j\rangle \rightarrow |x_i\rangle$ with probability P

Detailed Balance



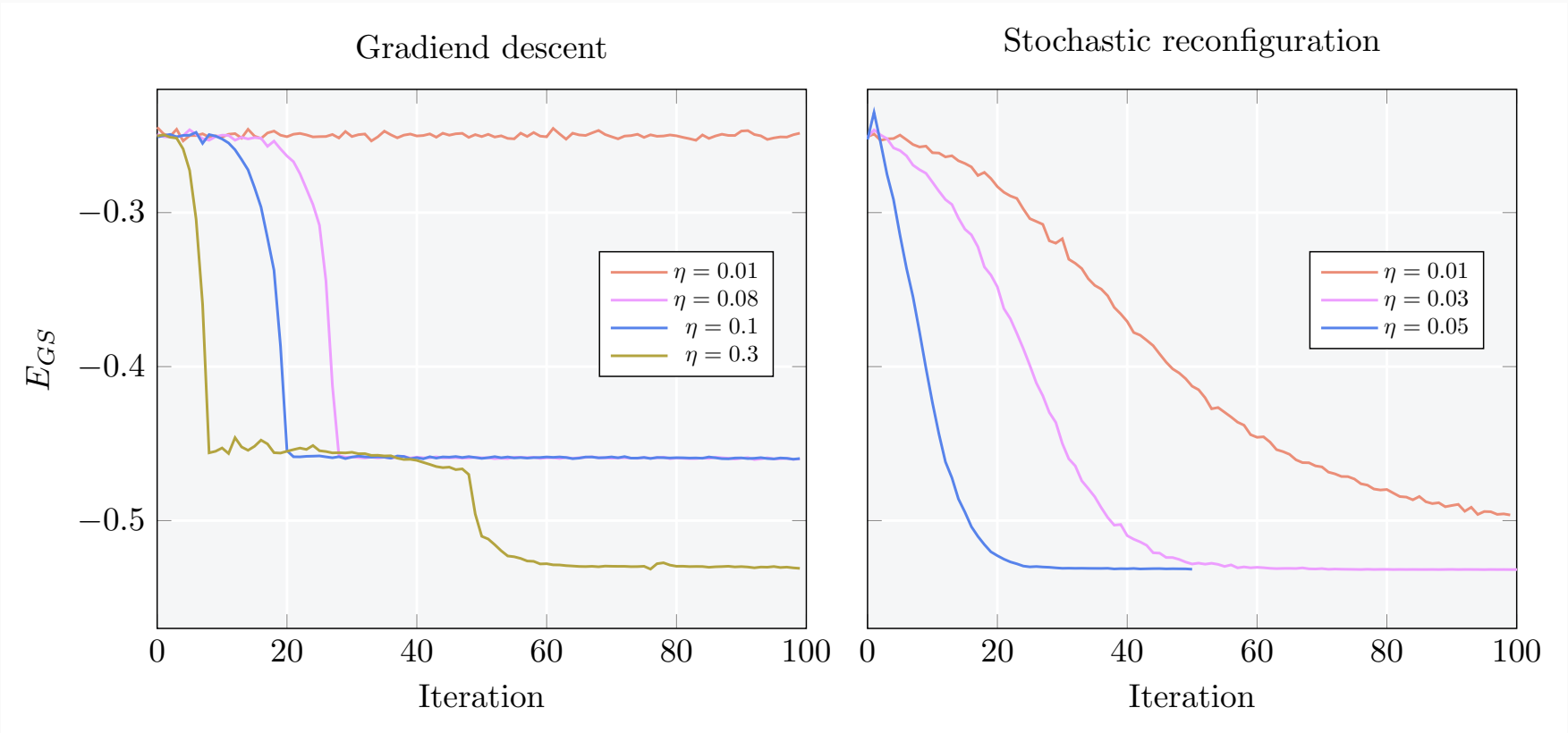
Detailed Balance

Our transition probability $\mathcal{P}$ fulfils detailed balance such that **in equilibrium** we are sampling from

$$P(x) = \frac{|\Psi(x)|^2}{\sum_x |\Psi(x)|^2}$$

Stochastic reconfiguration vs Gradient descent[1]

- Simulations for the transverse field Ising model.
- **Gradient descent:** the updates get stuck oscillating between wells in the energy landscape.
- A new minimization method is required.



[1] available code by: Emily Davis, Kevin Fischer, and Alex Hristov

- The idea is to update the parameters λ in a way that mimics a time evolution

$$|\Psi_\epsilon\rangle = e^{-\epsilon\mathcal{H}}|\Psi_{0,\lambda}\rangle \approx (1-\epsilon\mathcal{H})|\Psi_{0,\lambda}\rangle$$

For small ϵ

$$|\Psi_\epsilon\rangle \approx |\Psi_{\lambda+\delta\lambda}\rangle$$

Are we closer to the ground state?

- $\mathcal{H} = E_0|E_0\rangle\langle E_0| + E_1|E_1\rangle\langle E_1| + \ldots$, such that $E_0 < E_1 < E_2 \ldots$
- After N steps:

$$\begin{aligned}|\Psi_{N\epsilon}\rangle &= e^{-\epsilon N\mathcal{H}}|\Psi_{0\lambda}\rangle \\ &= \alpha_0 e^{-\epsilon N E_0}|E_0\rangle + \alpha_1 e^{-\epsilon N E_1}|E_1\rangle + \ldots \\ &\to |E_0\rangle \quad \text{YES!}\end{aligned}$$

- By investigating how the parameters change, we find

$$\lambda \to \lambda - \eta S^{-1}\nabla_\lambda\langle\mathcal{H}\rangle$$

Fisher Matrix

- Differentiates between SR and plain GD;
- Takes the curvature of the quantum space into account;

$$S_{\alpha,\beta} = \langle \mathcal{O}_\alpha^\dagger \mathcal{O}_\beta \rangle - \langle \mathcal{O}_\alpha^\dagger \rangle \langle \mathcal{O}_\beta \rangle$$

$$\mathcal{O}_{\omega_{ij}} |x\rangle = \frac{\partial \log(\Psi_\lambda(x))}{\partial \omega_{ij}} |x\rangle$$

BM Ansatz $\Psi_\lambda(x)$

$$\mathcal{O}_{a_i} = \sigma_i^z$$

$$\mathcal{O}_{b_i} = \tanh(\theta_i(x))$$

$$\mathcal{O}_{w_{ij}} = \sigma_i^z \tanh(\theta_j(x))$$

$$\theta = b_j + W_{ij}\sigma_j^z$$

$$\lambda \to \lambda - \eta S^{-1} \nabla_\lambda \langle \mathcal{H} \rangle$$

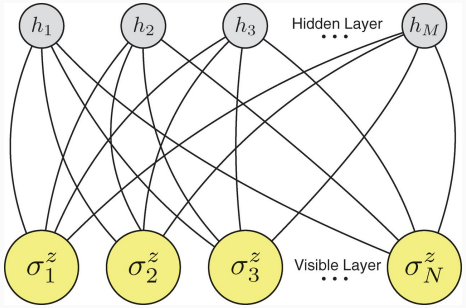
Energy gradient

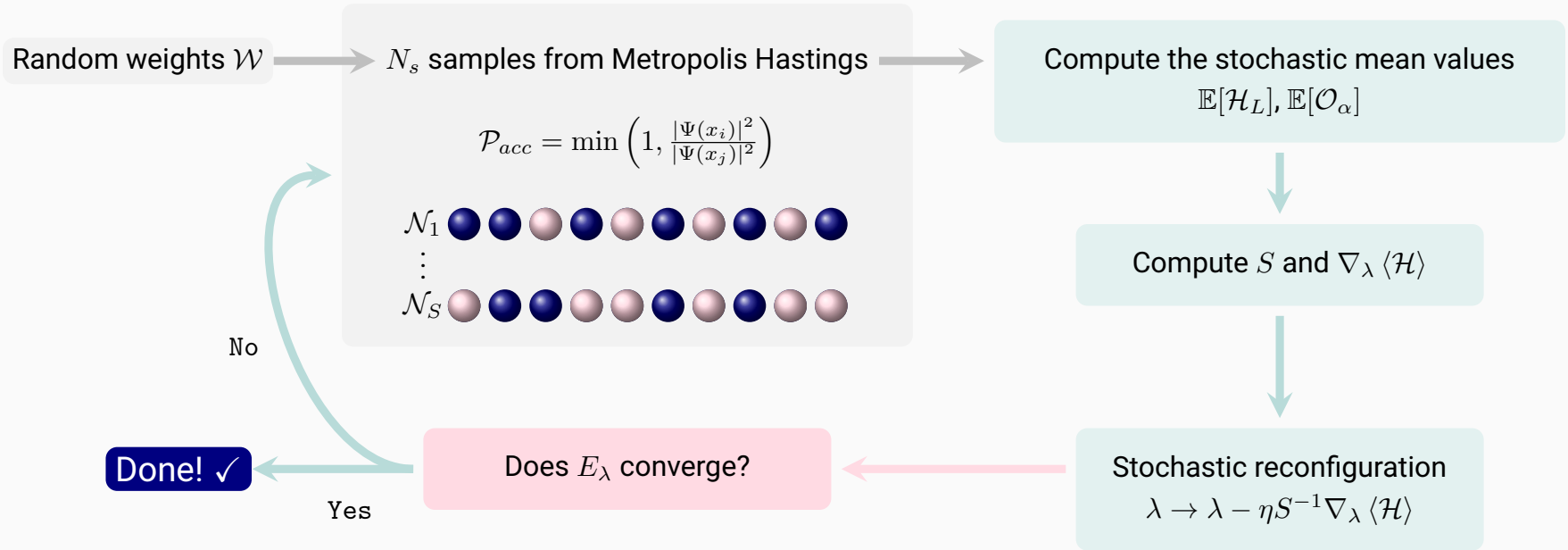
$$\partial_\alpha \langle \mathcal{H} \rangle = \mathbb{E}[\mathcal{O}_\alpha \mathcal{H}_L] - \mathbb{E}[\mathcal{O}_\alpha]\mathbb{E}[\mathcal{H}_L]$$

Learning rate

Proportional to ϵ, $\eta \approx 10^{-2}, 10^{-3}$

Overview


h_1
h_2
h_3
Hidden Layer
h_M
σ_1^z
σ_2^z
σ_3^z
Visible Layer
σ_N^z


Random weights $\mathcal{W}$
N_s samples from Metropolis Hastings
$\mathcal{P}_{acc} = \min\left(1, \frac{|\Psi(x_i)|^2}{|\Psi(x_j)|^2}\right)$
$\mathcal{N}_1$
$\mathcal{N}_S$
Compute the stochastic mean values
$\mathbb{E}[\mathcal{H}_L], \mathbb{E}[\mathcal{O}_\alpha]$
Compute S and $\nabla_\lambda \langle \mathcal{H} \rangle$
Stochastic reconfiguration
$\lambda \to \lambda - \eta S^{-1} \nabla_\lambda \langle \mathcal{H} \rangle$
Does E_λ converge?
No
Yes
Done! ✓

Models Transverse field Ising and AFM Heisenberg

- This paper analyses two model Hamiltonians

transverse-field Ising	antiferromagnetic Heisenberg
$\mathcal{H}_{TFI} = -h\sum_i \sigma_i^x - \sum_{i,j} \sigma_i^z \sigma_j^z$	$\mathcal{H}_{AFM} = \sum_{i,j} \sigma_i^x \sigma_j^x + \sigma_i^y \sigma_j^y + \sigma_i^z \sigma_j^z$

- RBM gives us the wavefunction on the computational basis.

$$\Psi_M(\mathcal{S}, \mathcal{W}) = \sum_{h_i} e^{\sum_j a_j \sigma_j^z + \sum_i b_i h_i + \sum_{ij} W_{i,j} h_i \sigma_j^z} = \langle S|\Psi_M\rangle$$

$$|\Psi_M\rangle = \sum_S \langle S|\Psi_M\rangle |S\rangle$$

which determines the whole state **(amplitude and phase)**.

Neural network ground states

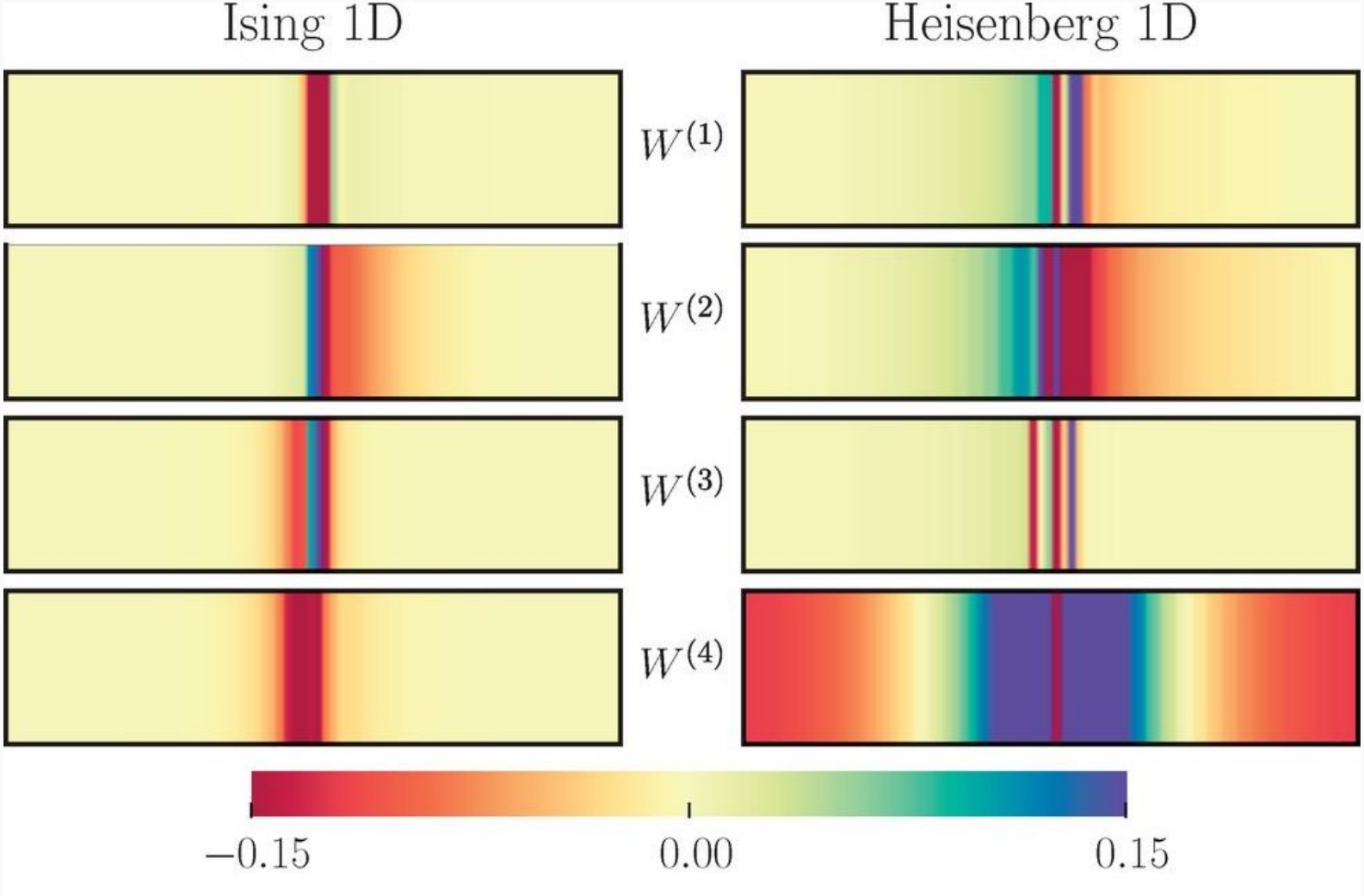


Figure: **Neural Network representation of the many body ground states.** The horizontal axis represents the 80 lattice sites. The colormap encodes the *f*th feature map $W_j^{(f)}$ for each site.

Neural network ground states

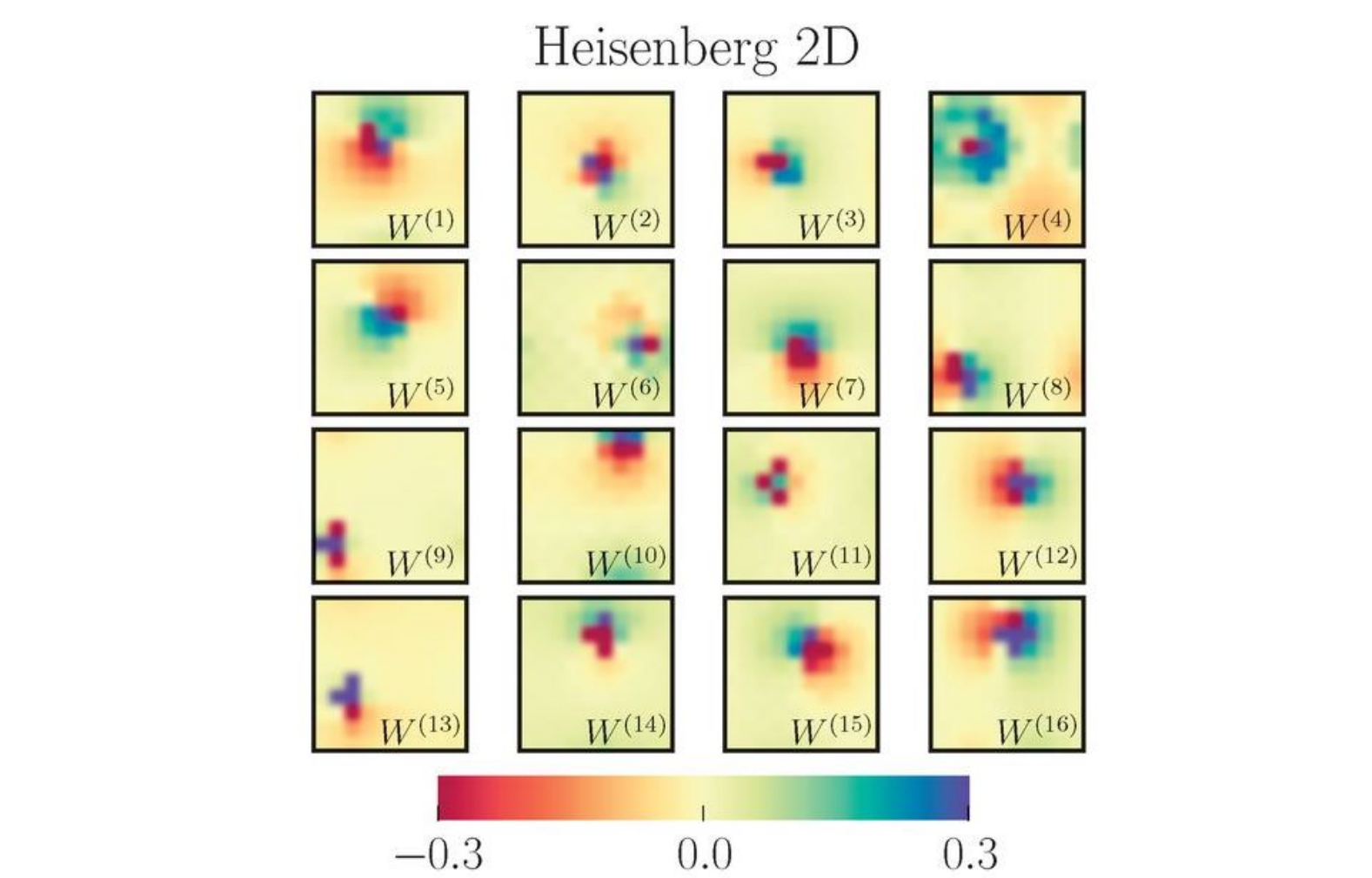


Figure: **Neural Network representation of the many body ground states.** The horizontal axis represent the 10×10 lattice sites. The colormap encodes the fth feature map $W_j^{(f)}$ for each site. The density of sites is $alpha = 16$.

Ground state energies

- By increasing the hidden variable density α, we systematically improve our results.

Controllable and arbitrary accuracy

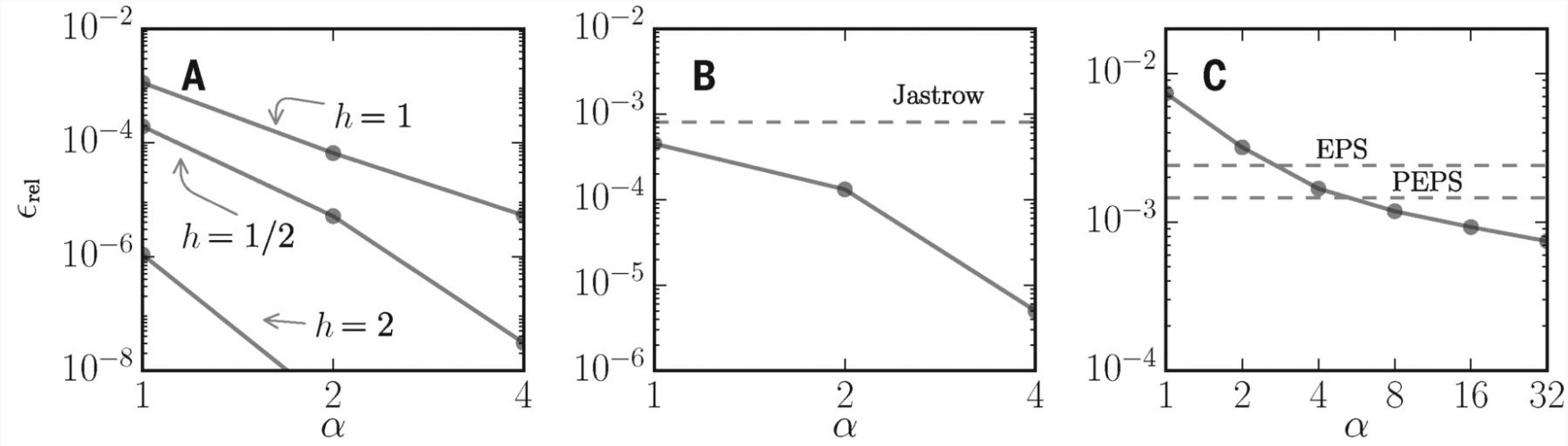


Figure: Error for the ground state energy of **A.** 1D TFI model, 80 sites. **B.** 1D AFH model, 80 sites. **C.** AFH model, 10×10 sites.

Unitary dynamics

- NQS can be extended to the time dependent Schrödinger equation.

$$\mathcal{W} \to \mathcal{W}(t)$$

time dependent network weights

- The new cost function is

$$R(\dot{\mathcal{W}}(t)) = \text{dist}(\partial_t \Psi, -i\mathcal{H}\Psi)$$ **Dirac-Frenkel variational principle**

- Now the weights are updated according to

$$\dot{\mathcal{W}}(t) = -iS^{-1}(t)\nabla_{\mathcal{W}}\mathcal{H}(t)$$

- New sampling method $\Rightarrow$ Time-Dependent Variational Monte Carlo [2]

[2] Carleo et al. (again ☺)

Quantum quench

The system is initially in the ground state for a certain field h_i. At $t = 0$, we instantaneously change the field to a new value h_j, and let the system evolve under the new Hamiltonian.

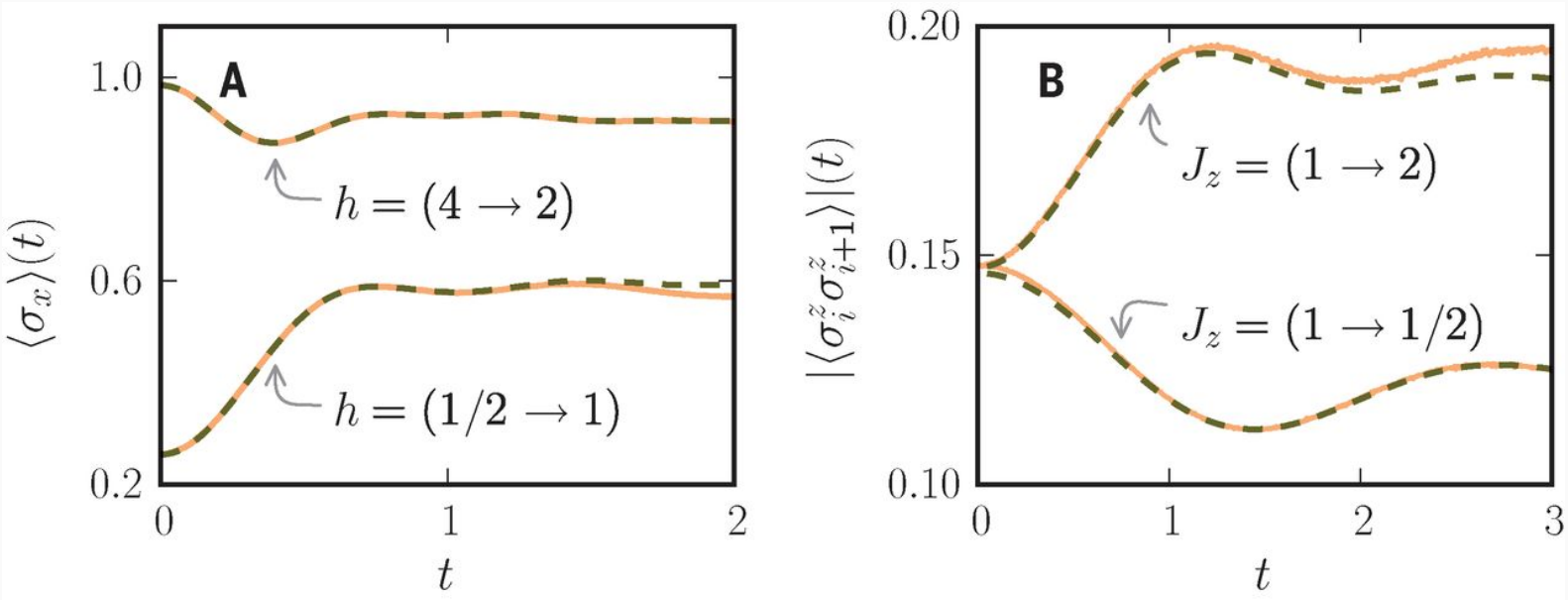


Figure: NQS results (**solid lines**) for a quantum quench in the parameters. **A.** Transverse spin polarization in the TFI model. **B.** Time dependent nearest neighbors spin correlations in the AFH model. The **dashed lines** are the exact results.

Outlook and summary

Summary

- The **high accuracy of the unitary dynamics** shows that NQS have applications beyond ground state physics.
- The method provides the best variational results reported to date for the 2D AFH.

Outlook

- Treating quantum systems other than interacting spins;
- Formal analysis of the NQS entanglement properties;
- More advanced architectures;
- Answers to challenging questions about interacting fermions.

Questions? ☺

References I

[Carleo and Troyer, 2017] Carleo, G. and Troyer, M. (2017).
Solving the quantum many-body problem with artificial neural networks.
Science, 355(6325):602–606.

[Davis et al., 2018] Davis, E., Fischer, K., and Hristov, A. (2018).
A neural network approach to the quantum many-body problem.
CS230: Deep Learning, Winter 2018, Stanford University.

[Park and Kastoryano, 2020] Park, C.-Y. and Kastoryano, M. J. (2020).
Geometry of learning neural quantum states.
Physical Review Research, 2(2).