Title: Machine learning ground-state energies and many-body wave function

Speakers: Sebastiano Pilati

Collection: Machine Learning for Quantum Design

Date: July 10, 2019 - 2:45 PM

URL: http://pirsa.org/19070011

Abstract: In the first part of this presentation, I will present supervised machine-learning studies of the low-lying energy levels of disordered quantum systems. We address single-particle continuous-space models that describe cold-atoms in speckle disorder, and also 1D quantum Ising glasses. Our results show that a sufficiently deep feed-forward neural network (NN) can be trained to accurately predict low-lying energy levels. Considering the long-term prospect of using cold-atoms quantum simulator to train neural networks to solve computationally intractable problems, we consider the effect of random noise in the training data, finding that the NN model is remarkably resilient. We explore the use of convolutional NN to build scalable models and to accelerate the training process via transfer learning.

In the second part, I will discuss how generative stochastic NN, specifically, restricted and unrestricted Boltzmann machines, can be used as variational Ansatz for the ground-state many-body wave functions. In particular, we show how to employ them to boost the efficiency of projective quantum Monte Carlo (QMC) simulations, and how to automatically train them within the projective QMC simulation itself.

SP, P. Pieri, Scientific Reports 9, 5613 (2019)
E. M. Inack, G. Santoro, L. Dellâ€™Anna, SP, Physical Review B 98, 235145 (2018)

Machine learning ground-state energies and many-body wave functions

Sebastiano Pilati (University of Camerino)

UNICAM
Università di Camerino
1336

Perimeter Institute, July 2019

In collaboration with:
Pierbiagio Pieri, Nicola Saraceni, Simone Cantori (Uni. Camerino)
Estelle Maeva Inack (Perimeter Institute)
Giuseppe Santoro, Giuliano Giudici, Tommaso Parolini (SISSA)
Luca Dell'Anna (Uni. Padova)

Computing resources from:
CINECA SCAI
SuperComputing Applications and Innovation

First part:

$$\hat{H} \longrightarrow E_0 = \langle \psi_0 | \hat{H} | \psi_0 \rangle$$

Bypass quantum many-body computation (e.g, molecular dynamics)
Perspective: consider the use of cold-atom quantum simulators to train neural networks

Second part:

$$\hat{H} \longrightarrow |\psi_0\rangle$$

Build accurate ground-state wf using Boltzmann machine to boost projective QMC simulations
Applications in adiabatic quantum optimization (quantum annealing)

First part:

$$\hat{H} \longrightarrow E_0 = \langle \psi_0 | \hat{H} | \psi_0 \rangle$$

Bypass quantum many-body computation (e.g, molecular dynamics)
Perspective: consider the use of cold-atom quantum simulators to train neural networks

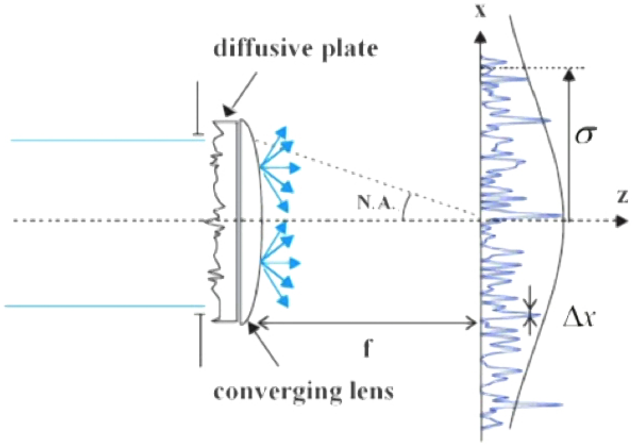
Second part:

$$\hat{H} \longrightarrow |\psi_0\rangle$$

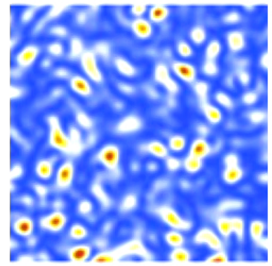
Build accurate ground-state wf using Boltzmann machine to boost projective QMC simulations
Applications in adiabatic quantum optimization (quantum annealing)

Disorder: optical speckle patterns

EXP @ LENS, Palaiseau

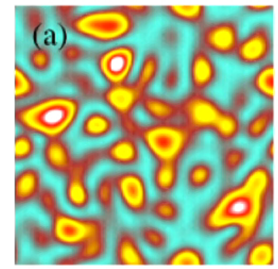


experiment @ LENS



Semeghini et al. Nat. Phys. 2015

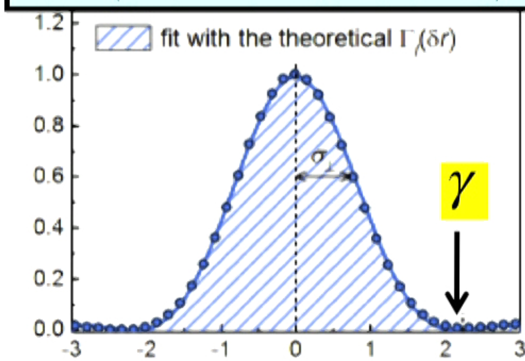
num. simulation



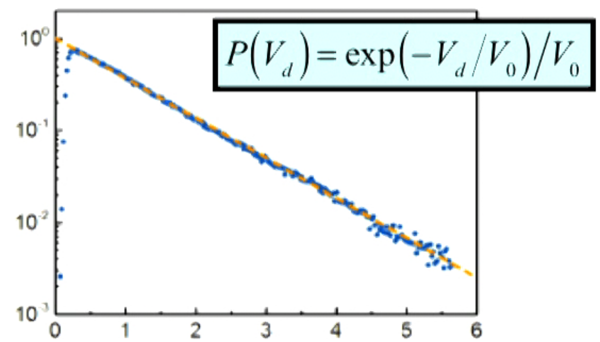
SP, Fratini PRA (2015)

Spatial correlations

$$\Gamma(x) \equiv \left\langle \left(V_d(x'+x) - \bar{V}_d\right)\left(V_d(x') - \bar{V}_d\right)\right\rangle / \bar{V}_d^2$$

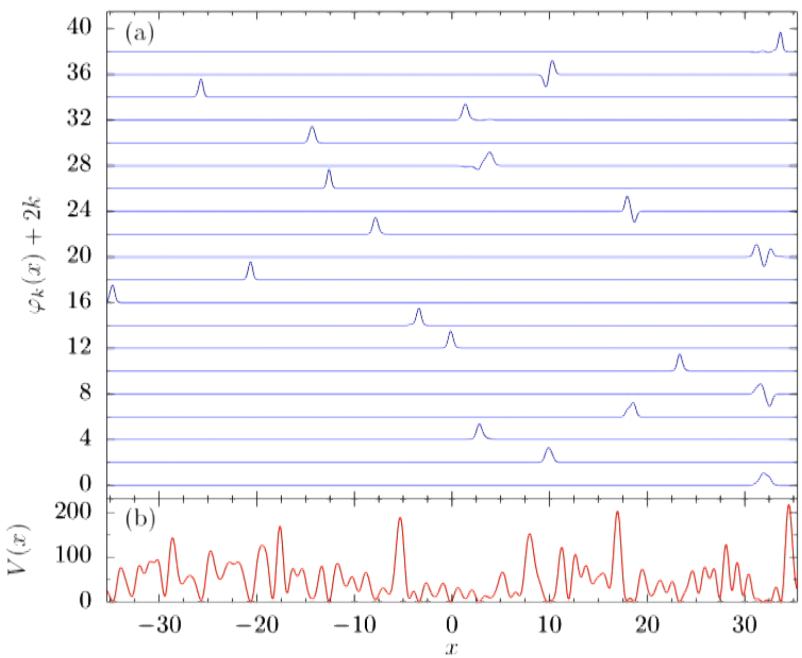


Probability distribution of intensities

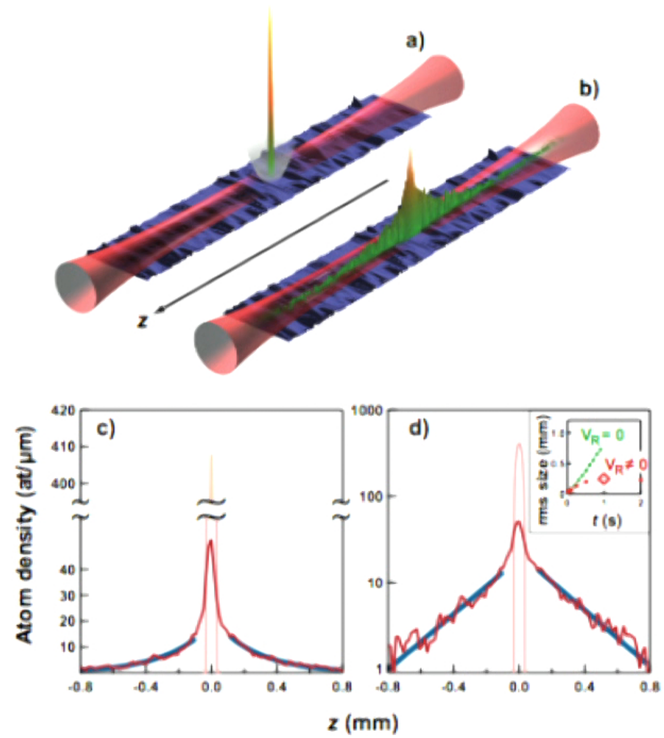


Anderson localization in 1D speckle optical potentials

Low-energy single-particle states



Exp @ Institut d'Optique, LENS (Florence)



J. Billy et al., Nature 453, 891 (2008)

One atom in a 1D speckle potential: NUMERICAL SOLUTION

$$\left[-\frac{\hbar^2}{2m}\frac{\mathrm{d}^2}{\mathrm{d}x^2}+V_{\mathrm{d}}(x)\right]\varphi_i(x)=E_i\varphi_i(x)$$

x_1 … x_{j-1} x_j x_{j+1} … x_N

Finite difference method, 3 point formula:

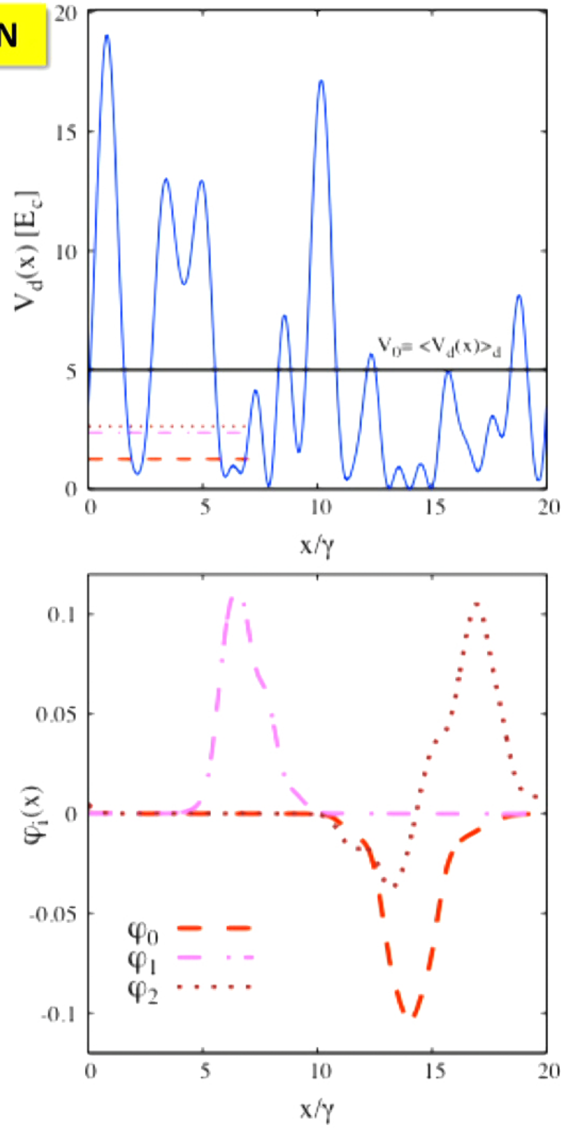
$$\frac{\mathrm{d}^2}{\mathrm{d}x^2}\cong\frac{1}{\delta x^2}\left[-\varphi(x_{j-1})+2\varphi(x_j)-\varphi(x_{j+1})\right]$$

Hamiltonian matrix:

$$H=\begin{pmatrix} \frac{\hbar^2}{m\delta x^2}+V_{\mathrm{d}}(x_1) & -\frac{\hbar^2}{2m\delta x^2} & \cdots & 0 \\ -\frac{\hbar^2}{2m\delta x^2} & \frac{\hbar^2}{m\delta x^2}+V_{\mathrm{d}}(x_2) & \cdots & \cdots \\ \cdots & \cdots & \cdots & -\frac{\hbar^2}{2m\delta x^2} \\ 0 & \cdots & -\frac{\hbar^2}{2m\delta x^2} & \frac{\hbar^2}{m\delta x^2}+V_{\mathrm{d}}(x_N) \end{pmatrix}$$

Matrix eigenvalue problem: $H\varphi_i=E_i\varphi_i$

NOTE: we actually use an 11-point formula

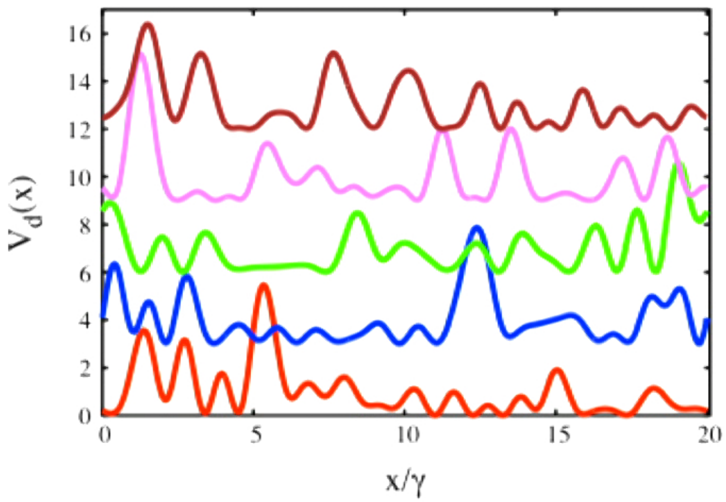


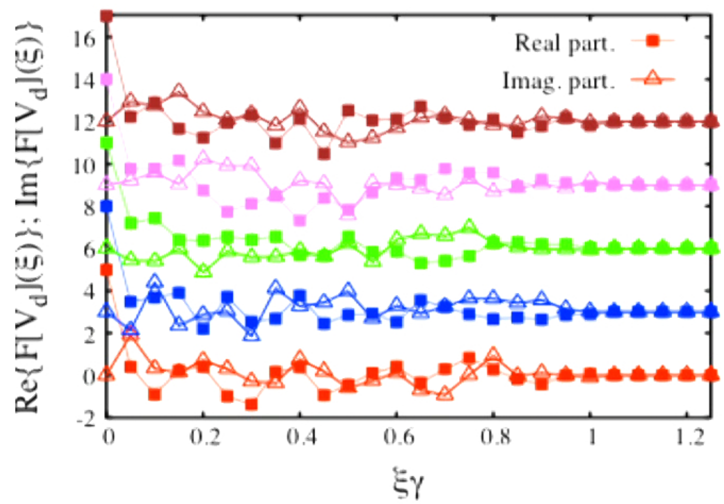
Representation: choice of features

We use knowledge about structure of speckle potential to find compact representation

Fourier transform of the optical speckle field has finite support: -1/Υ < ξ < 1/Υ
γ = Spatial correlation length of disorder

Only **42** nonzero Fourier components for *L=20γ*!

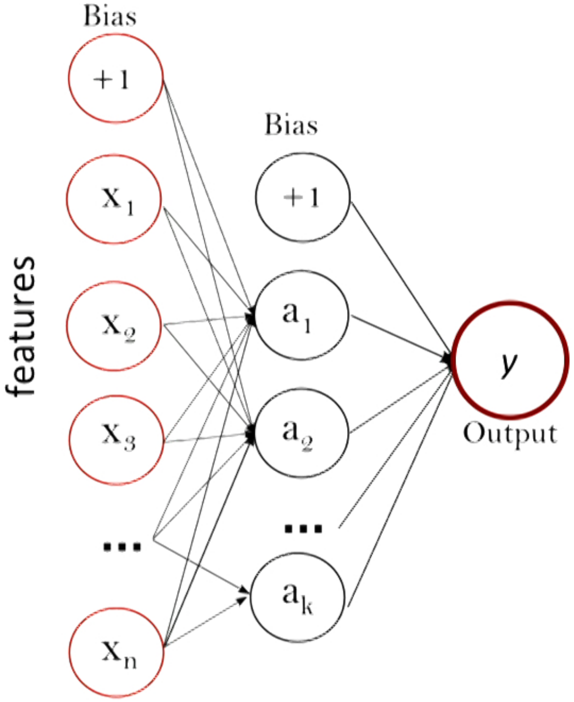


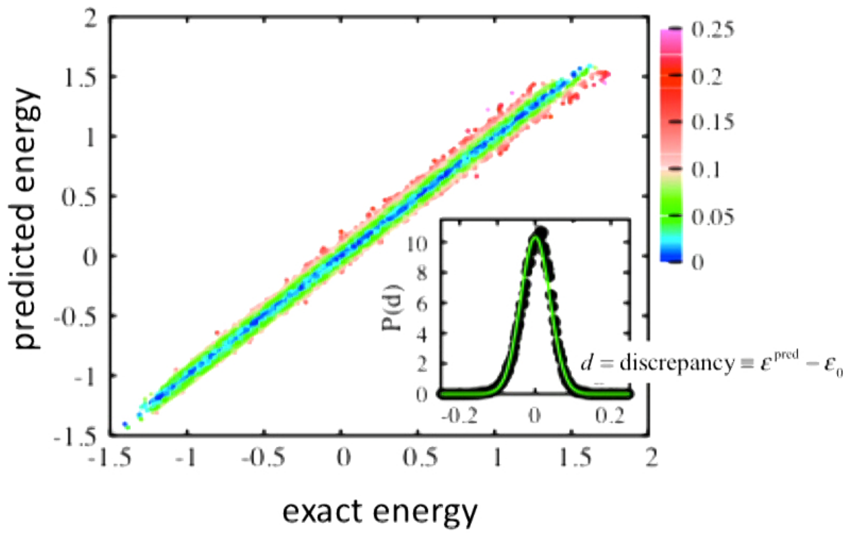


Predicting ground-state energy with deep NN

Fully-connected multi-layer NN
\# of hidden layers $N_l=3$
\# of neuron per layer $N_n = 150$
\# of instances in the training set $N_t = 80000$

SP, Pieri, Scientific Reports (2019)
Related work: Mills, Spanner, Tamblyn, PRA (2017)





Measure of accuracy: coefficient of determination R^2

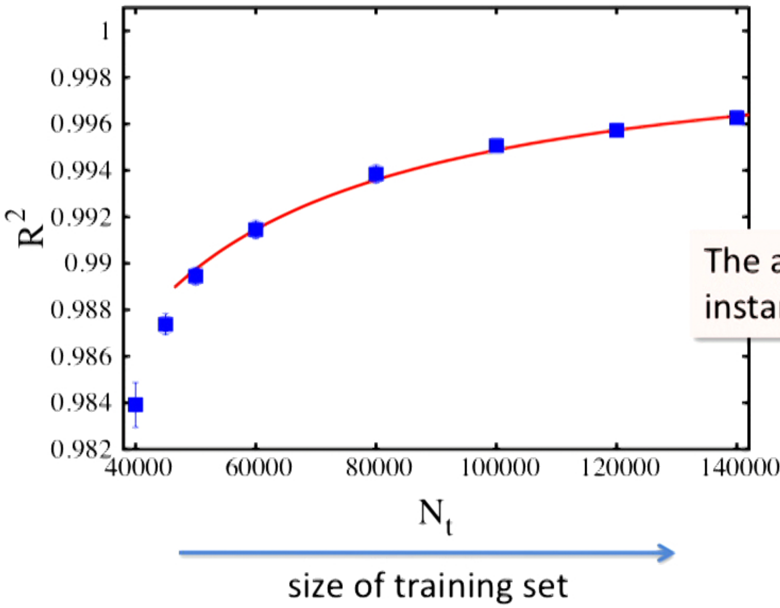
$$R^2 = 1 - \frac{\sum_{i=1}^{N_{\text{test}}} \left(y_i - F(\mathbf{x}_i) \right)^2}{\sum_{i=1}^{N_{\text{test}}} \left(y_i - \bar{y} \right)^2}$$

where: $\bar{y} = \frac{1}{N_{\text{test}}} \sum_{i=1}^{N_{\text{test}}} y_i$

$R^2 = 1$ corresponds to perfect predictions

$R^2 = 0$ corresponds to constant function equal to average

R^2 can be negative

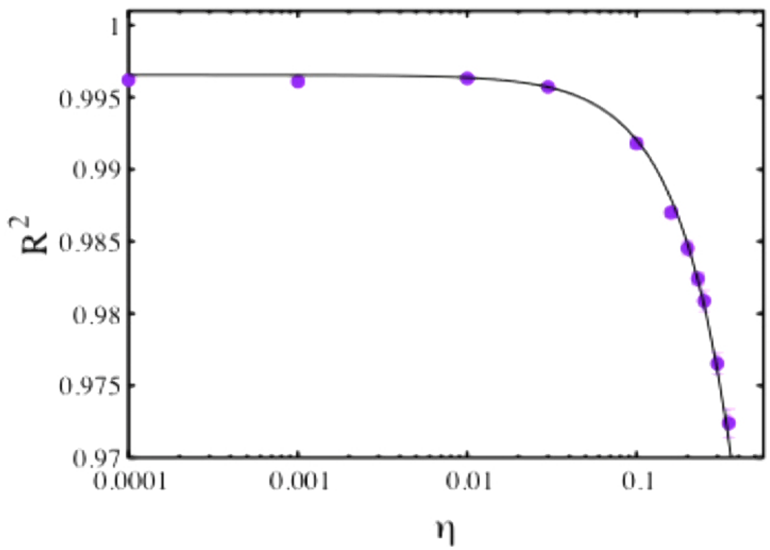


The accuracy improves with the number of instances in the training set

Cold-atom quantum simulators

Q: could we use QS to train NN to solve computationally intractable problem?

ANALYSIS OF NOISE SENSITIVITY: Training on synthetic data with added noise

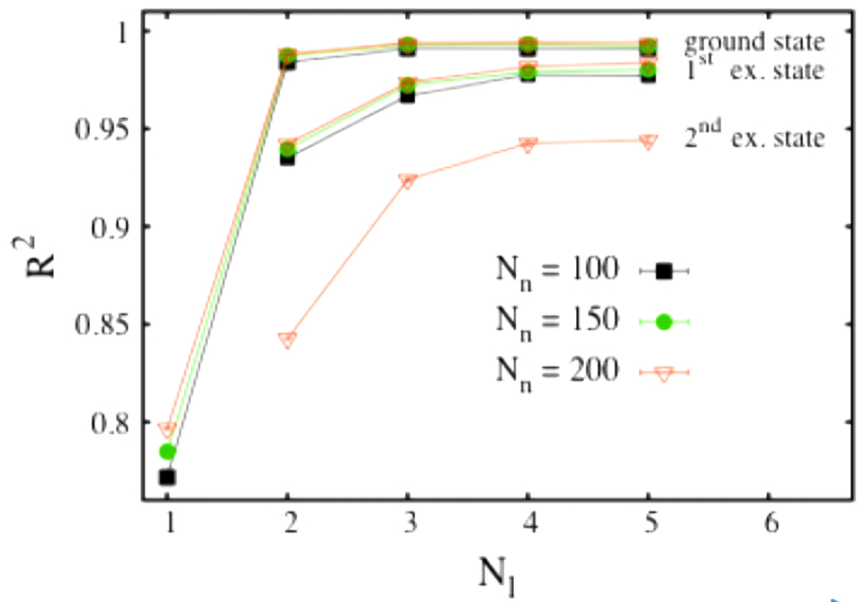


added noise

The NN is remarkably resilient, it can filter signal from noise

Gaussian noise with stand. dev. proportional to the data stand. dev. times η

Excited state energies

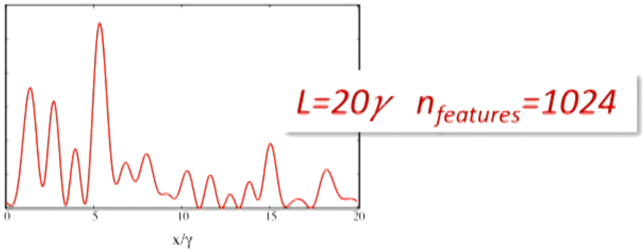


# of hidden layers

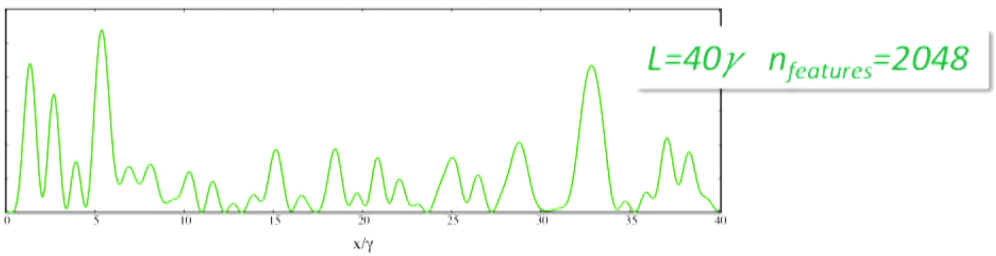
➔ it is more difficult to learn excited-state energies

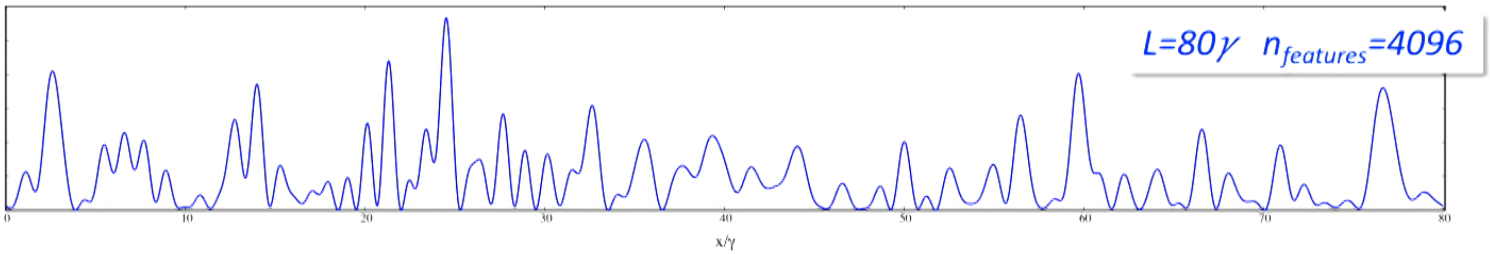
Scalabale neural networks

- Adaptive neural network that can address different system sizes
- Heterogeneous training
- Extrapolation: making prediction for larger system sizes



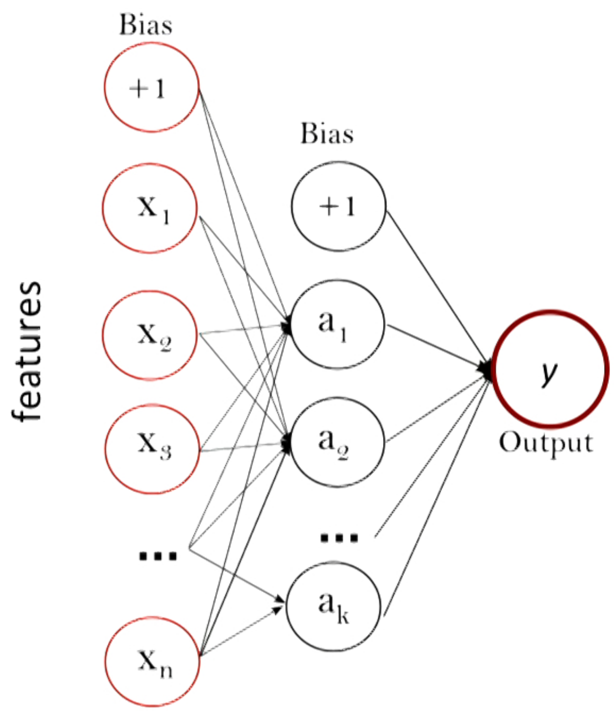
Features: potential values of a spatial grid



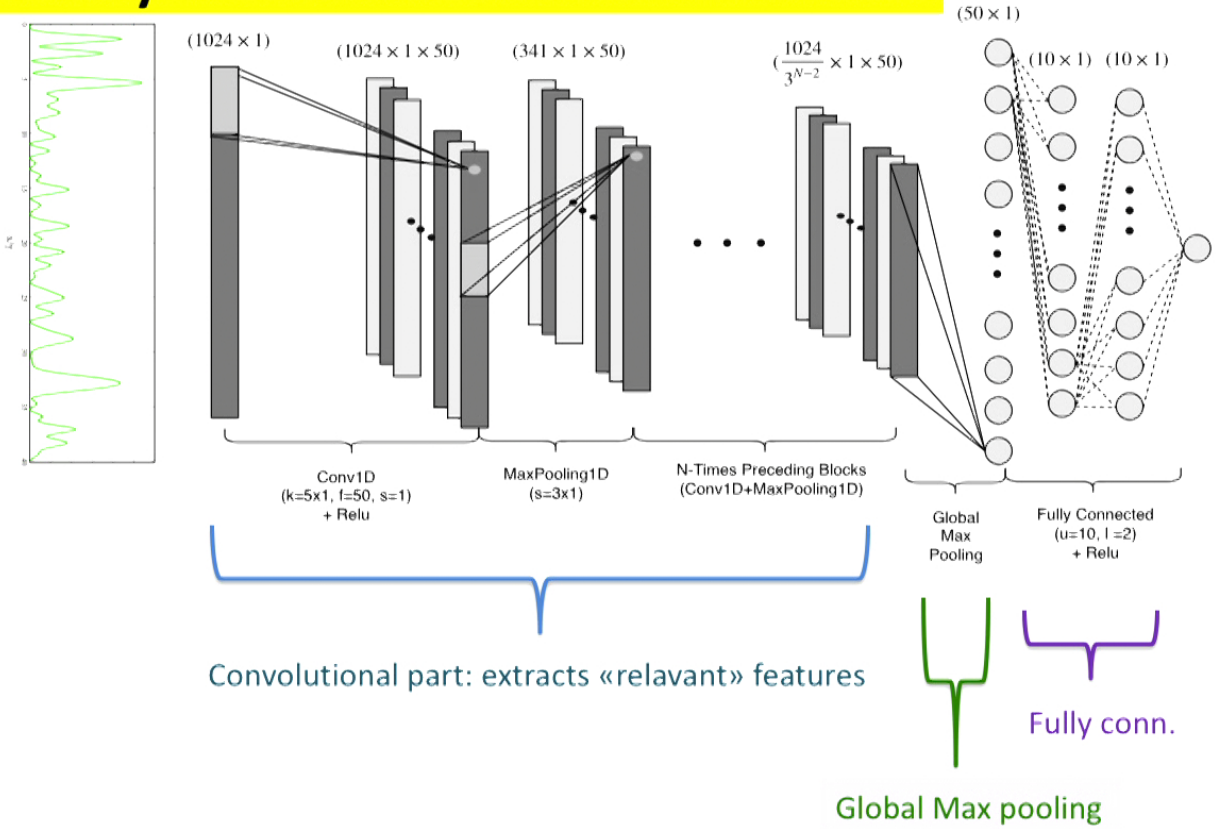


Large number of features: problematic for fully-connected neural networks

The number of parameters scales as: $n_{features} \times n_{hidden\ units}$



Scalability: convolutional neural networks



Extensive conv. NN: see *Ryczko, Strubbe, Tamblyn, Chemical Science (2019)*

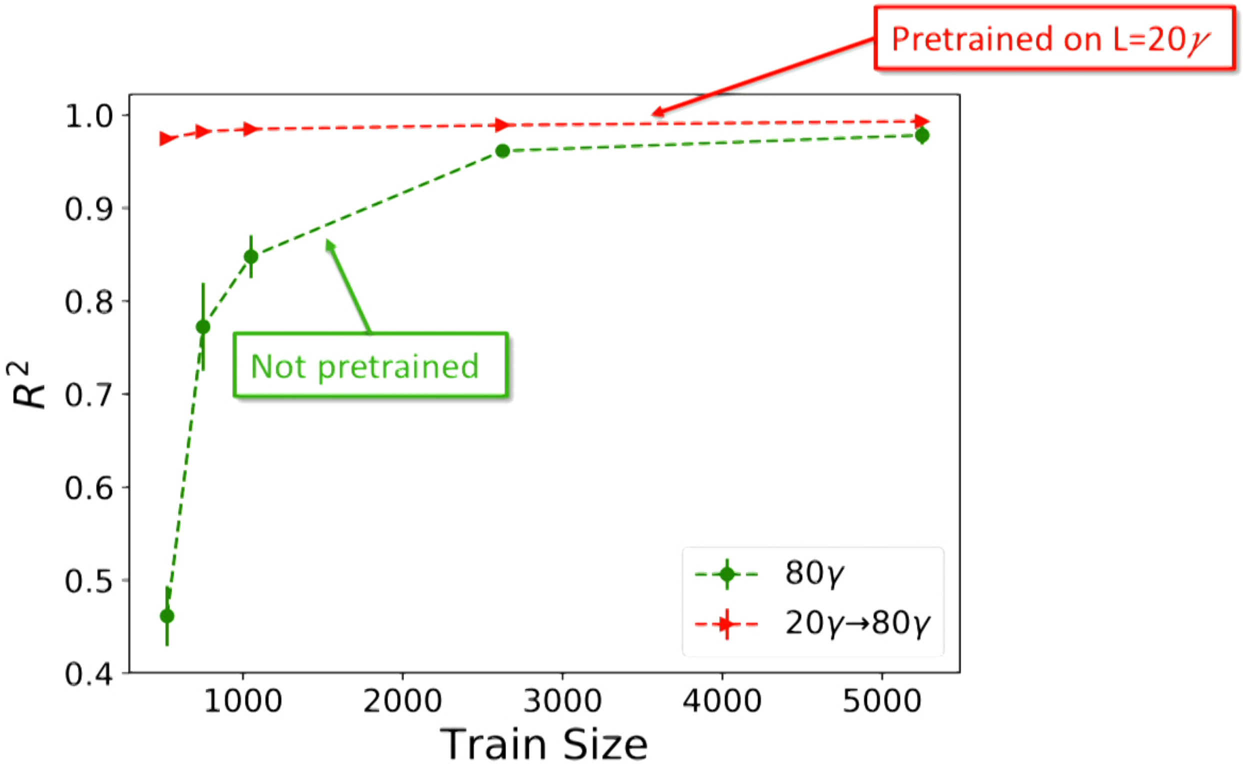
$$E_{TOT} \approx E_1 + E_2 + E_3 + \ldots$$

Transfer learning

If you have a small training set, you can use a NN that was pretrained on a similar task.

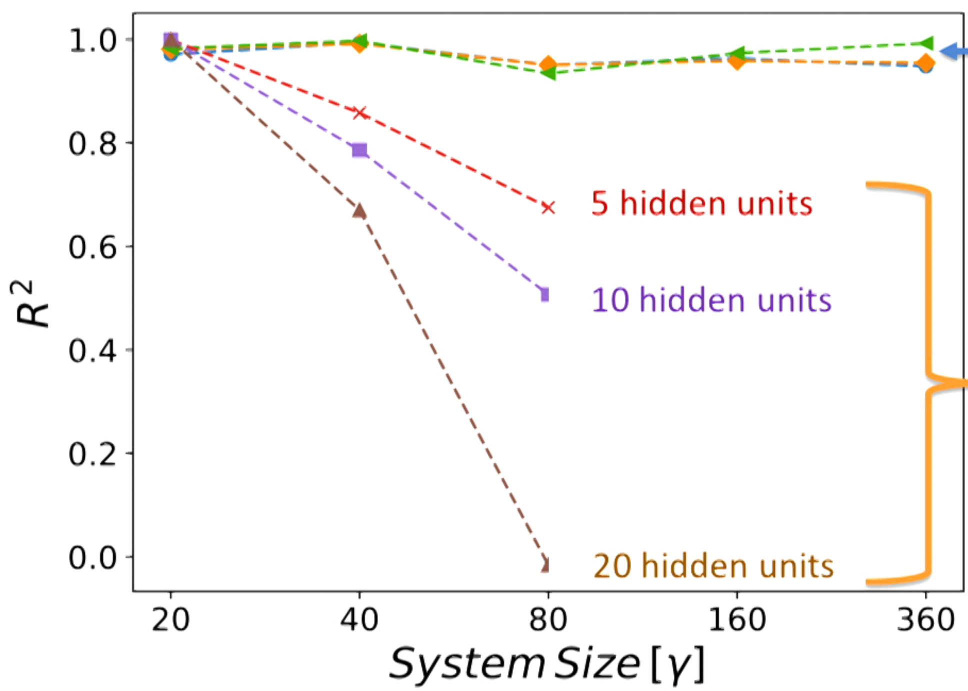
EXAMPLES: Oxford VGG Model, Google Inception Model, Microsoft ResNet Model

We specialize on L=80γ a NN pretrained on L=20γ



Extrapolation: making predictions for larger systems

Convolutional neural network with global max pooling layer



Heterogeneous training:
140K for L=20 and 5K for L=40

Homogeneous training
140K for L=20

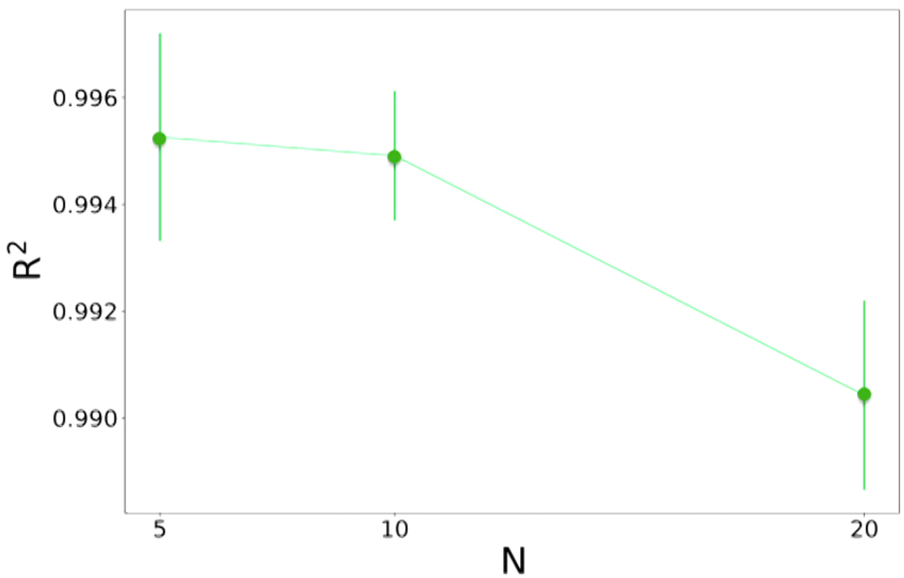
Disordered quantum Ising model in 1D

$$\hat{H} = -\sum_{\langle ij \rangle} J_{ij} \sigma_i^z \sigma_j^z - \Gamma \sum_i \sigma_i^x \qquad \sigma_i^z, \sigma_i^x \equiv \text{Pauli matrices}$$

$$J_{ij} = \mathrm{ran}\left(\left[-1,1\right]\right) \qquad \Gamma = 0.5$$

Features: couplings J_{ij}

Training set size: 30000



5 Conv. Layers, 20 filters (dim.=2)
+2 fully conn. (10 units)

WORK IN PROGRESS
Now gathering QMC data

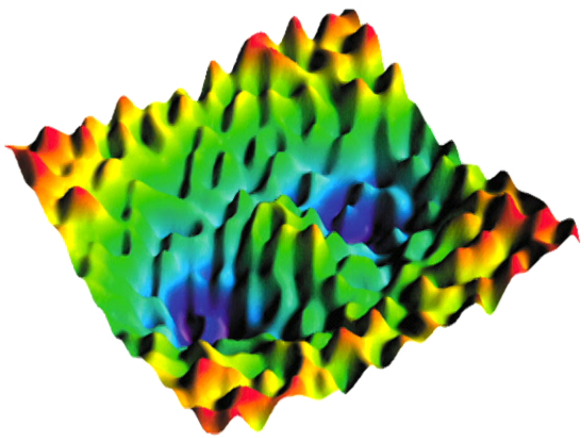
Second part:

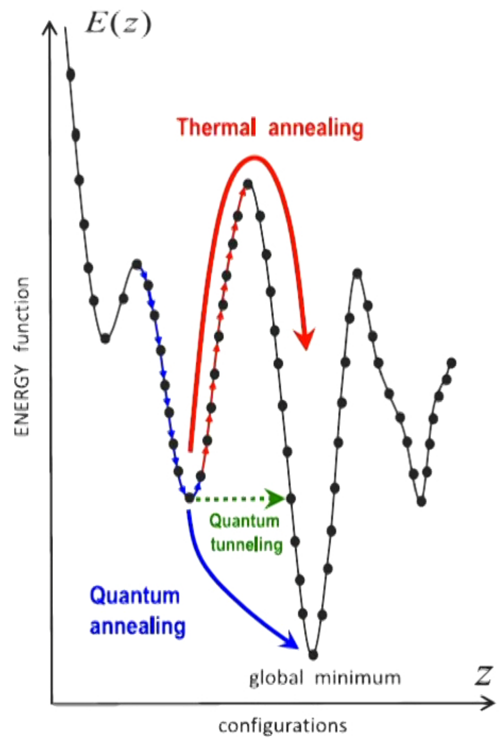
Boltzmann machines for projective QMC simulations

Two families of quantum MC algorithms:

- Path-Integral MC: classical MC on a D+1-dimensional system
- Projective QMC: stochastic simulation of Schrödinger eq. in imaginary time

Solving hard optimization problems: simulated (Classical) Annealing vs Quantum Annealing



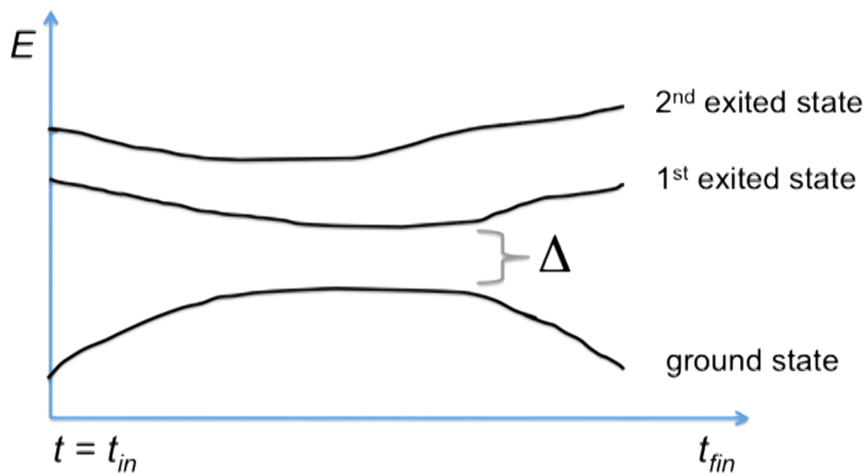


Adiabatic quantum computing: Quadratic Unconstrained Binary Optimization

$$H = -\sum_{ij} J_{ij}\sigma_i^z\sigma_j^z - \sum_i h_i\sigma_i^z - \Gamma(t)\sum_i \sigma_i^x$$

Ising glass:
Finding the ground state is a hard problem

Transverse field:
Annealed from $\Gamma >> J_{ij}$ to $\Gamma = 0$



How slow?
Adiabatic theorem:

$$t_{fin} >> \frac{\alpha}{\Delta^2}$$

$\Delta \equiv$ smallest gap

→→→→→→→→→→→

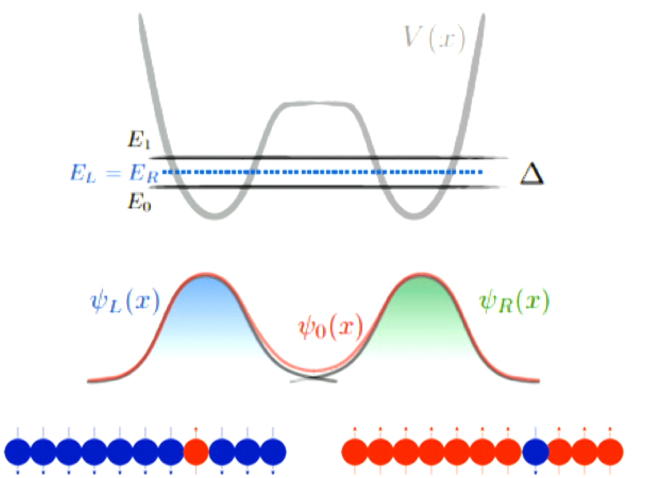
↑↑↓↑↓↓↑↑↓↑↑↓↑↓↓

QMC tunneling time: finite-temperature PIMC

Isakov, Mazzola, Smelyanskiy, Jiang, Boixo, Neven, Troyer, PRL (2016)
Mazzola, Smelyanskiy, Troyer, PRB (2017)

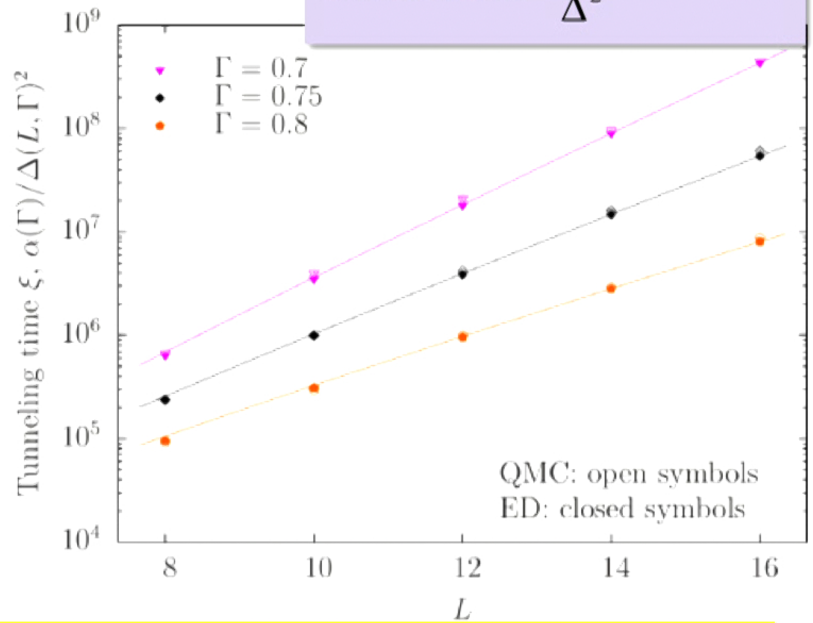
Ferromagnetic quantum Ising chain *Γ/J<1*

energy gap: $\Delta \propto \exp(-\alpha L)$



Open symbols: PIMC tunneling time

Closed symbols: $\dfrac{\alpha}{\Delta^2}$

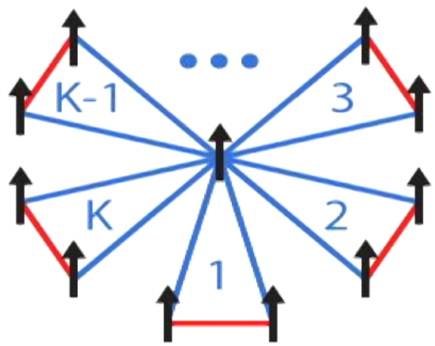


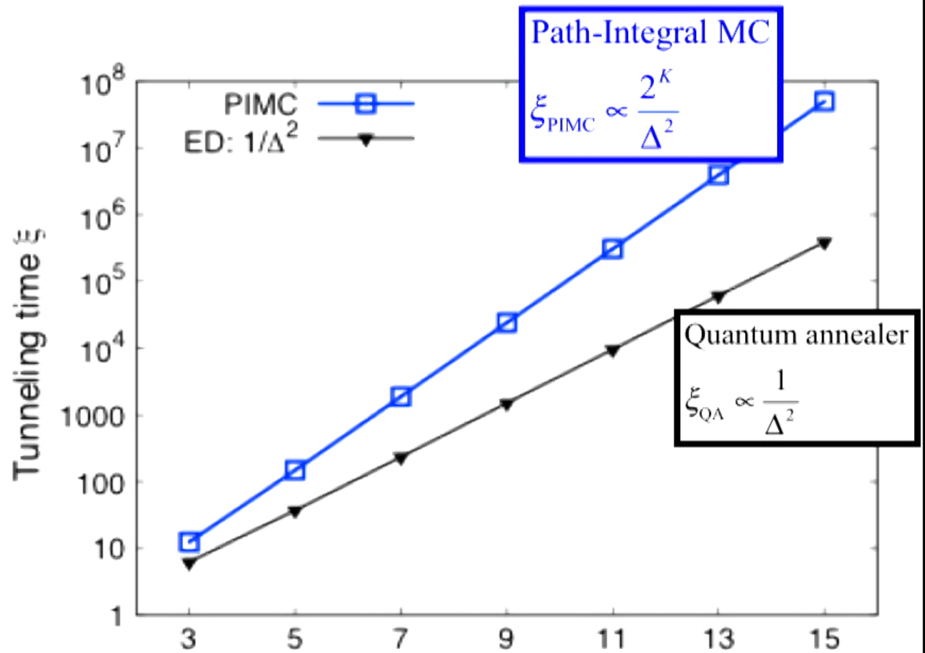
The PIMC algorithm efficiently simulates incoherent quantum tunneling.
Is this general?

Note: PIMC with open-boundary condition in imaginary time it scales as $1/\Delta$

Shamrock: a model of frustrated rings

E. Andriyash and M. H. Amin (D-Wave Systems Inc.), *"Can quantum Monte Carlo simulate quantum annealing?"*, arXiv:1703.09277, 2017





Recently realized @Google
Kafri, D., Quintana, C., Chen, Y., Martinis, J., & Neven, H. Progress Towards Quantum Annealer v2. 0, Bulletin of the American Physical Society (2018).

➢ Path-integral slows down due to "topological" obstruction, **slower than Quantum annealer!**

Projective Monte Carlo for Quantum Ising models

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

$$\psi(\mathbf{S},\tau) = \exp(-\tau H)\psi(\mathbf{S},0) \underset{\tau\to\infty}{\approx} \psi_0(\mathbf{S},0)$$ Schrödinger eq. in imaginary time

$$\psi(\mathbf{S},\tau+\Delta\tau) = \sum_{\mathbf{S}'} G(\mathbf{S}',\mathbf{S},\Delta\tau)\psi(\mathbf{S}',\tau)$$ defines a Markov process

$G(\mathbf{S}',\mathbf{S},\Delta\tau) \geq 0 \Rightarrow$ no sign problem (stoquastic Hamiltonian)

$\sum_{\mathbf{S}'} G(\mathbf{S}',\mathbf{S},\Delta\tau) \neq 1 \Rightarrow$ not a standard Markov process $\Rightarrow$ kill or clone random walkers

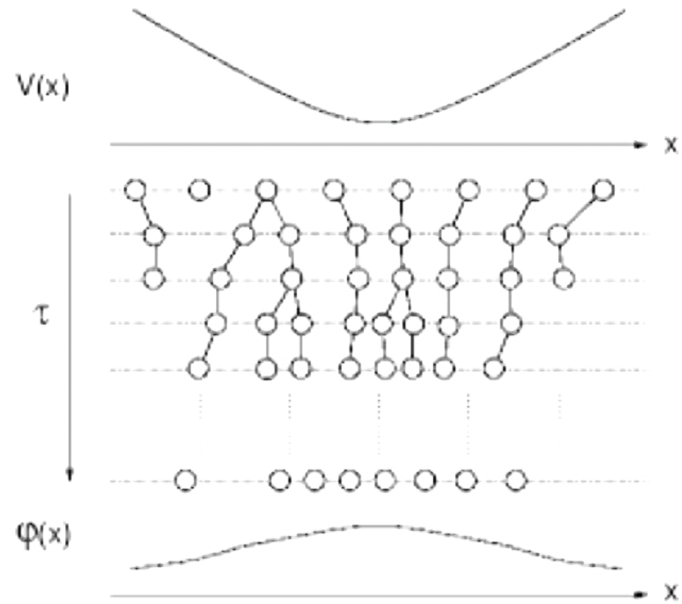
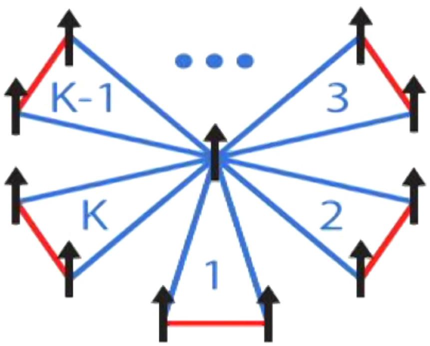
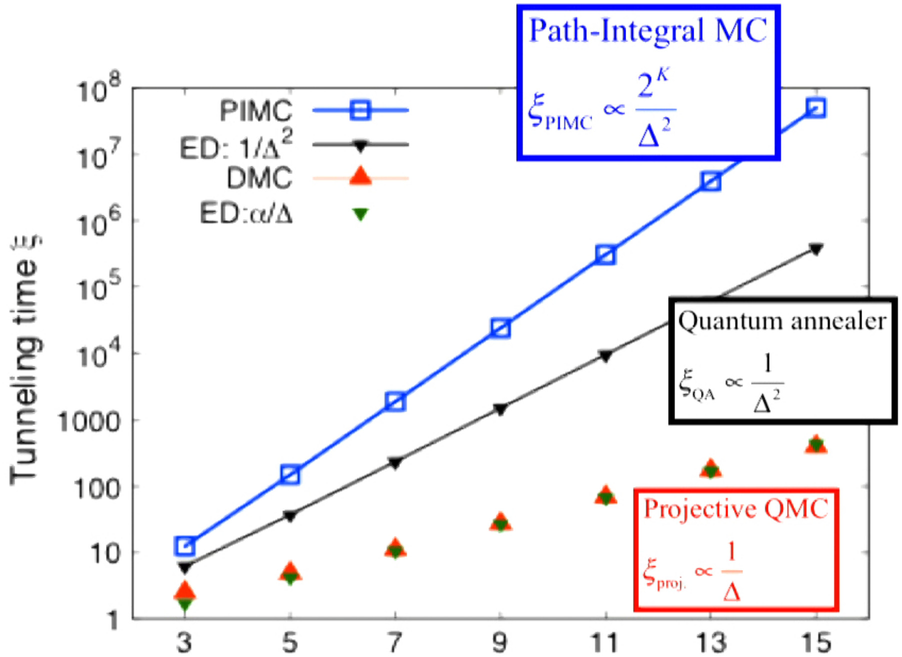


Image from: J. Thijssen, Computational Physics, Cambridge University Press

Shamrock: a model of frustrated rings

E. Andriyash and M. H. Amin (D-Wave Systems Inc.), *"Can quantum Monte Carlo simulate quantum annealing?"*, arXiv:1703.09277, 2017





Recently realized @Google
Kafri, D., Quintana, C., Chen, Y., Martinis, J., & Neven, H. Progress Towards Quantum Annealer v2. 0, Bulletin of the American Physical Society (2018).

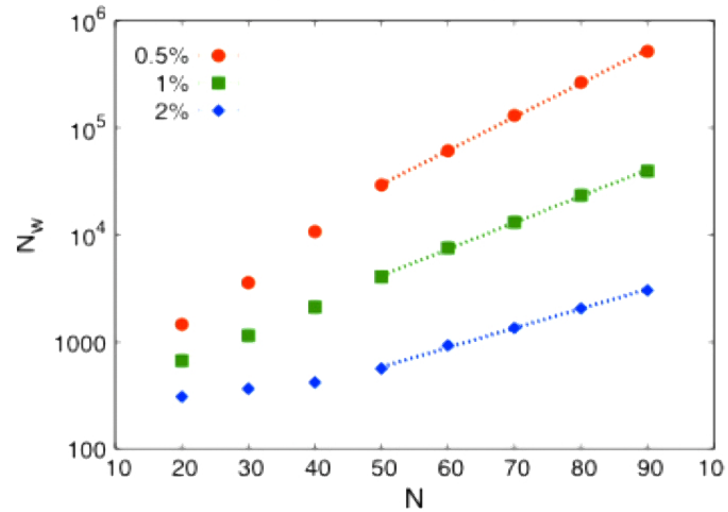
- Path-integral slows down due to "topological" obstruction, **slower than Quantum annealer!**
- Projective QMC like $1/\Delta$ (i.e., "faster" than QA)

E. M. Inack, G. Giudici,T. Parolini, G.E. Santoro, SP, PRA (2018)

Computational cost of projective QMC simulations

Notice: any diagonal Hamiltonian is stoquastic (sign-problem free). Finding its ground state encompasses hard classical optimization problems such as k-SAT or MAX-CUT.
Bravyi, Quant. Inf. Comp., Vol. 15, No. 13/14, pp. 1122-1140 (2015)

# of walkers required to keep relative err. fixed



System size

Exponentially growing computational cost, even without sign problem

Note: here we use "simple" PQMC algorithm: no guiding wave function.

IMPORTANCE SAMPLING

Introduce guiding wave function $\equiv \psi_{\mathrm{G}}(\mathbf{x})$

Modified master eq.: $\Psi(\mathbf{x},\tau+\Delta\tau)\psi_{\mathrm{G}}(\mathbf{x}) = \sum_{\mathbf{x}'} \tilde{G}(\mathbf{x},\mathbf{x}',\Delta\tau)\Psi(\mathbf{x}',\tau)\psi_{\mathrm{G}}(\mathbf{x}')$

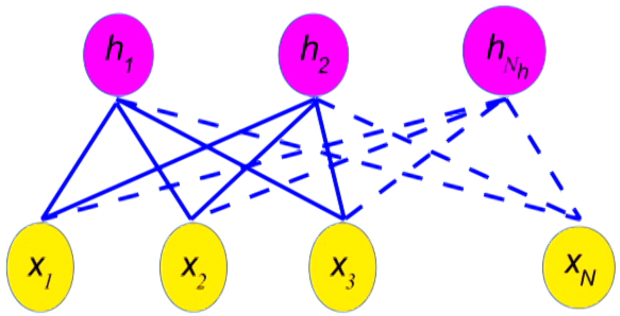
Modified Green's function: $\tilde{G}(\mathbf{x},\mathbf{x}',\Delta\tau) = \langle \mathbf{x} | \exp(-\Delta\tau \hat{H} - E_{\mathrm{REF}}) | \mathbf{x}' \rangle \dfrac{\psi_{\mathrm{G}}(\mathbf{x})}{\psi_{\mathrm{G}}(\mathbf{x}')}$

The guiding wf reduces computational cost and statistical fluctuations

Here, we adopt neural network states.

Restricted Boltzmann machines

Carleo, Troyer, Science 2017



$$\psi(\mathbf{x}) = \prod_j \exp(a_j x_j) \prod_i 2\cosh\left(b_i + \sum_j w_{ij} x_j\right)$$

$\propto N \times N_h$ variational parameters

Hidden spins integrated out

Unrestricted Boltzmann machine

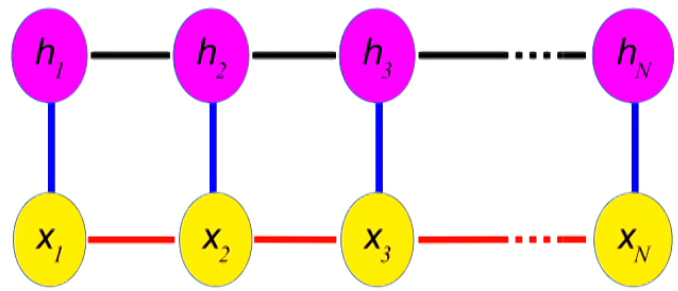
alias shadow wave-function

Reatto, Masserini, PRB 1988

Vitiello, Runge, Kalos PRL 1988

≈ variational imaginary-time Ansatz with P=1

Beach, Melko, Grover, Hsieh 2019



$$\psi(\mathbf{x}) = \sum_{\mathbf{h}} \phi(\mathbf{x}, \mathbf{h})$$

$$\phi(\mathbf{x}, \mathbf{h}) = \exp\left(-k_1 \sum_i x_i x_{i+1}\right) \exp\left(-k_2 \sum_i h_i h_{i+1}\right) \exp\left(-k_3 \sum_i x_i h_i\right)$$

$k_1, k_2, k_3 = 3$ variational parameters

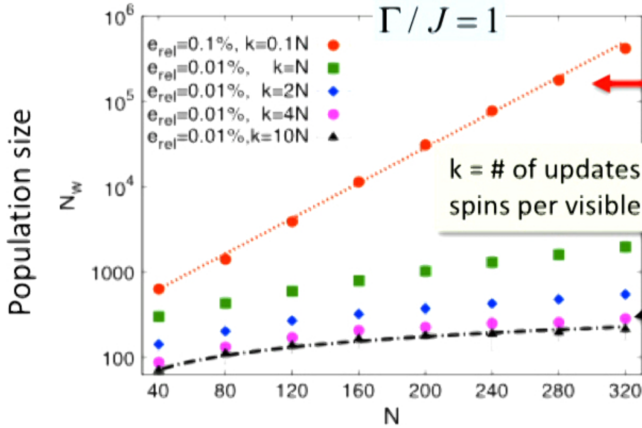
Need to sample hidden spins

Inack, Dell'Anna, Santoro, SP, PRB 2018

Computational complexity of PQMC guided by unrestricted BM

Inack, Dell'Anna, Santoro, SP, PRB 2018

- Needs combined sampling of both visible and hidden spins
- Correlations among hidden-spin configurations affect systematic errors



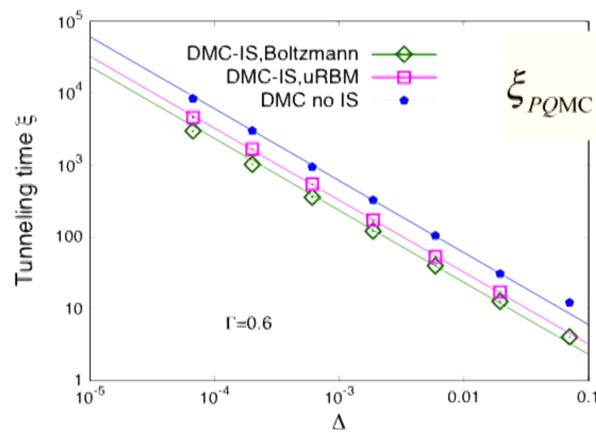
k = # of updates of hidden spins per visible spin

$$N_W \propto \exp(aN) \qquad a = 0.023(3)$$

$$N_W \propto N^b \qquad b = 0.5(1)$$

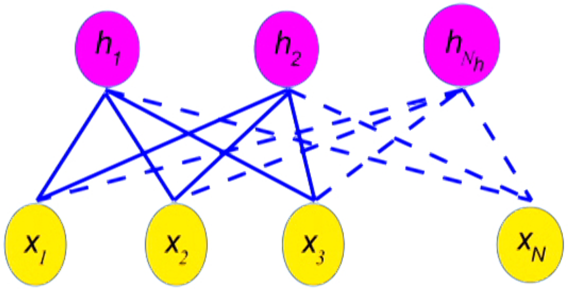
➔ polynomial computational cost

➔ scaling of tunneling time is not affected by guiding wf



$$\xi_{PQMC} \propto \frac{1}{\Delta}$$

Restricted Boltzmann machine: unsupervised learning



Quantum state tomography:
Torlai, Mazzola, Carrasquilla, Troyer, Melko, Carleo, Nat. Phys. (2018)

Marginal probability: $P_{\mathbf{W}}(\mathbf{x}) = \sum_h P_{\mathbf{W}}(\mathbf{x},\mathbf{h}) = \frac{1}{Z}\sum_h \exp\left[-H_{\mathrm{RBM}}(\mathbf{x},\mathbf{h})\right]$

Partition function: $Z = \sum_{\mathbf{x},\mathbf{h}} \exp\left[-H_{\mathrm{RBM}}(\mathbf{x},\mathbf{h})\right]$

Log-likelihood: $L(\mathbf{W}) = \sum_{k=1}^{N_{\mathrm{train}}} \ln P_{\mathbf{W}}(\mathbf{x}_k)$

Maximize log-likelihood, minimize KL divergence

Gradient ascent update rule: $W_m^{n+1} = W_m^n + \eta \frac{\partial L(\mathbf{W})}{\partial W_m}$

Gradient of log-likelihood: $\frac{\partial L(\mathbf{W})}{\partial J_{ij}} \propto \langle x_j h_i \rangle_{\mathrm{data}} - \langle x_j h_i \rangle_{\mathrm{model}}$

Performed via k-step contrastive divergence

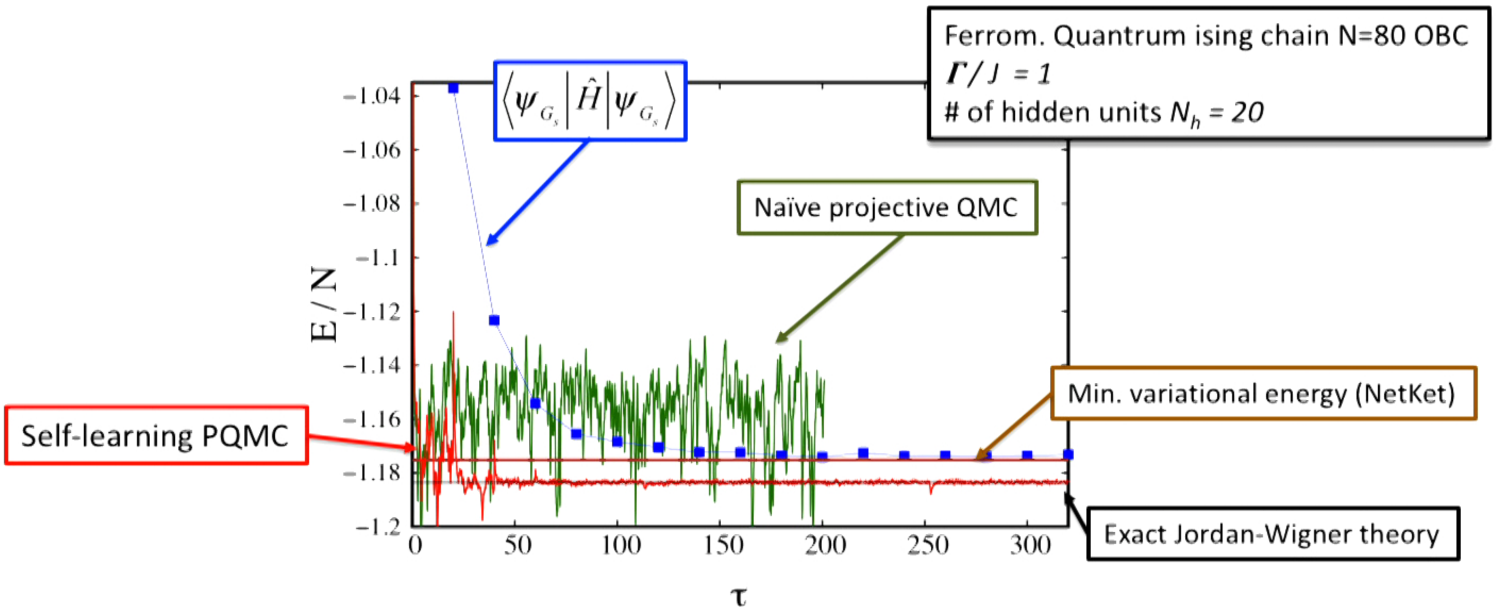
Self-learning projective QMC simulation

SP, Inack, Pieri, arXiv: arXiv:1907.00907 (2019)

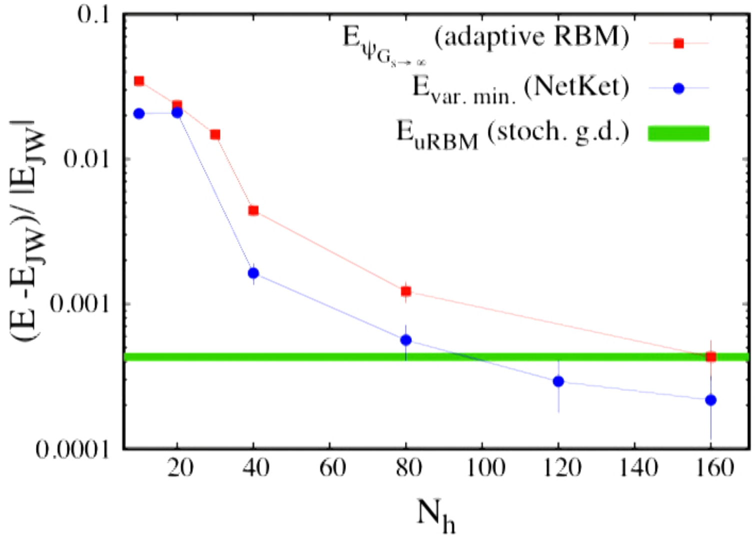
- The RBM learns the randomw-walker distribution: $P(\mathbf{x}) \propto \psi_G(\mathbf{x})\psi_0(\mathbf{x})$
- Guiding wf for the next stint: $\psi_G(\mathbf{x}) = \sqrt{P(\mathbf{x})}$ stoquastic model $\Rightarrow \psi_0(\mathbf{x}) \geq 0$

$\psi_{G_0}(\mathbf{x})\psi_0(\mathbf{x})$ $\quad \sqrt{\psi_{G_0}(\mathbf{x})\psi_0(\mathbf{x})}\psi_0(\mathbf{x})$ $\quad \ldots \quad \ldots \quad$ $\psi_{G_0}^{1/2^2}(\mathbf{x})\psi_0^{2-1/2^n}(\mathbf{x})$ $\quad \ldots$

imaginary time

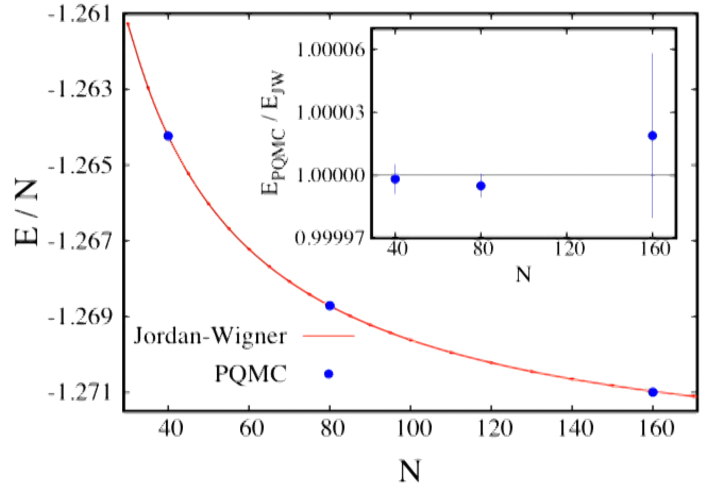


Log-likelihood maximization versus variational energy minimization (NetKet)



# of hidden spins

Ferrom. Quantrum ising chain N=80 OBC
$\Gamma / J = 1$



System size

SP, Inack, Pieri, arXiv: arXiv:1907.00907 (2019)