NYSDS

October 23

2020

Many-Body Quantum Systems as a Challenge for Machine Learning

Giuseppe Carleo

Computational Quantum Science Lab
Institute of Physics
EPFL - Lausanne - Switzerland

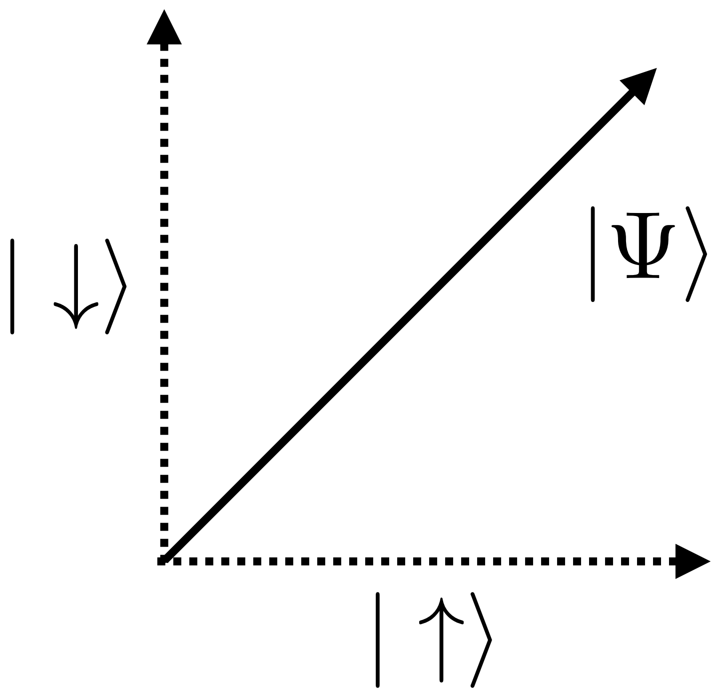
EPFL

A Major Problem In Quantum Physics

The State of a Two-Level System

The state of a
spin/qubit/etc is
a complex-valued
vector
(the wave function)

$$|\Psi\rangle = c_{\uparrow}|\uparrow\rangle + c_{\downarrow}|\downarrow\rangle$$



Probability of Observing a Given State

$$P(\uparrow) = |c_{\uparrow}|^2$$

$$P(\downarrow) = |c_{\downarrow}|^2$$

A quantum spin can be found
in either up or down state
with a given probability

The State of a Many Body System

The wave function is a vector
in a huge (2^N)
Complex Vector Space

$$|\Psi\rangle = c_{\uparrow\uparrow\uparrow\ldots}|\uparrow\uparrow\uparrow\ldots\rangle + c_{\downarrow\uparrow\uparrow\ldots}|\downarrow\uparrow\uparrow\ldots\rangle + \ldots c_{\downarrow\downarrow\downarrow\ldots}|\downarrow\downarrow\downarrow\ldots\rangle$$

Complex-Valued Coefficients

Schroedinger's Equation

$$\mathcal{H}\,|\Psi\rangle = E\,|\Psi\rangle$$

Eigenvalue Problem
for given Hamiltonian
(sparse matrix)

Quantum Many-Body Problem

[3000 BCE]

Papyrus

10 Qubits

[1455]

Book

15 Qubits

[1973]

IBM 3340

23 Qubits

[1993]

IBM 3390

35 Qubits

[2002]

Earth Simulator

46 Qubits

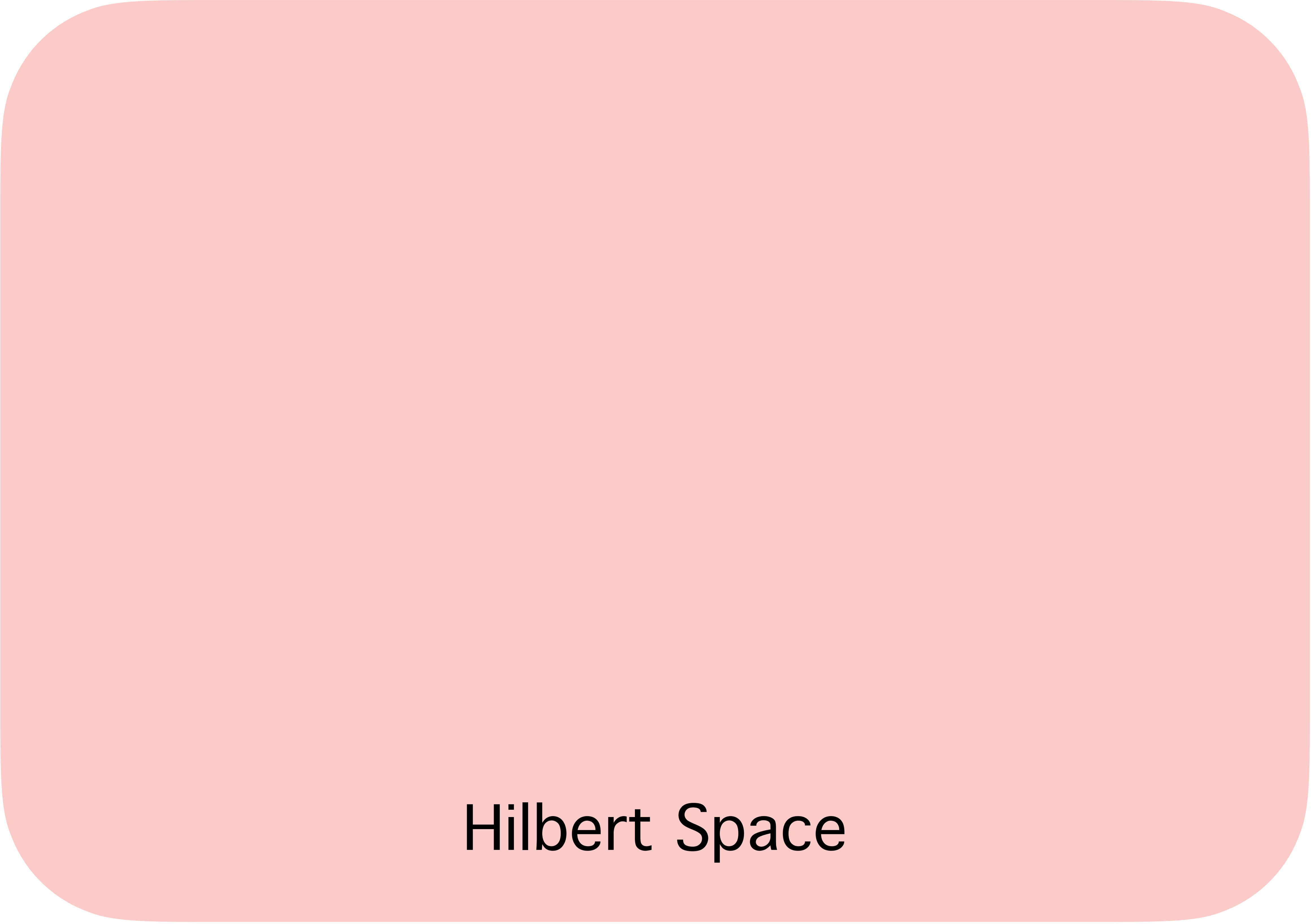
[2019]

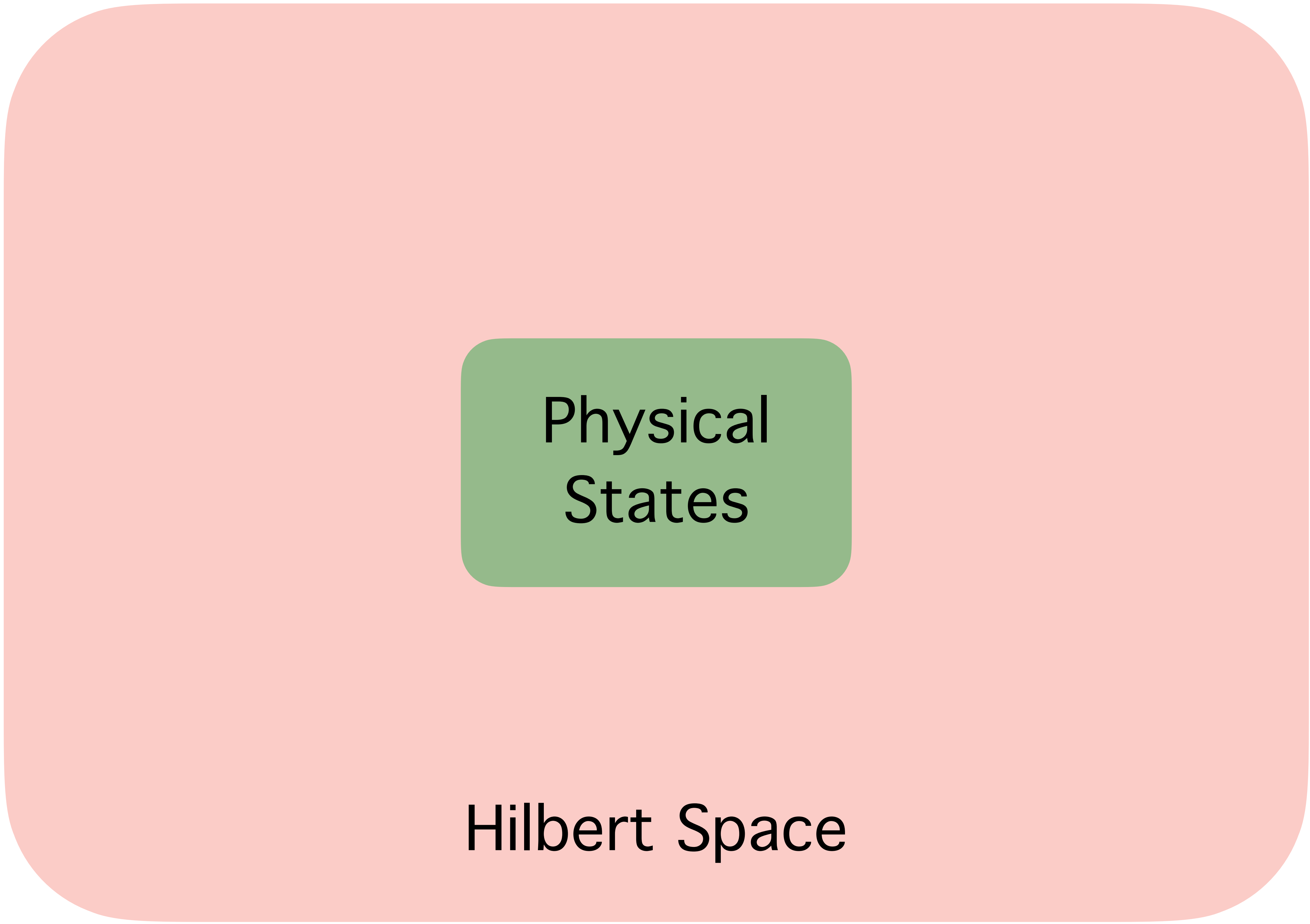
Summit

54 Qubits

Is this complexity "truly" exponential for physical systems?

Corners of Hilbert space


Hilbert Space


Physical States
Hilbert Space

Our Best Hope: Variational Formulation

$$E(\mathbf{W}) = \frac{\langle \Psi(\mathbf{W}) | \mathcal{H} | \Psi(\mathbf{W}) \rangle}{\langle \Psi(\mathbf{W}) | \Psi(\mathbf{W}) \rangle} \geq E_0$$

Rayleigh Quotient

Exact Ground-State Energy

Two ML-Inspired Approaches In This Talk

Classical Variational States	Quantum Variational States
Require a CPU/GPU/TPU…	***Require a QPU…***

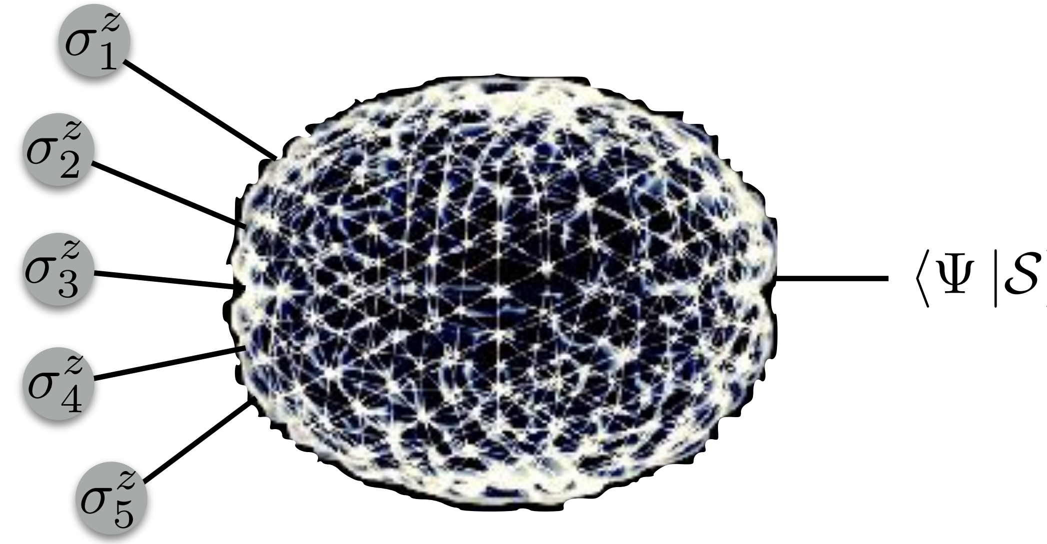
Classical Variational States

Issue

Can We Have A Sufficiently Flexible And Computationally Efficient Classical Ansatz?

Neural-Network Quantum States

Carleo, and Troyer
Science 355, 602 (2017)



$c_{\sigma_1^z, \sigma_2^z \ldots \sigma_5^z}$

Many-Body Amplitudes

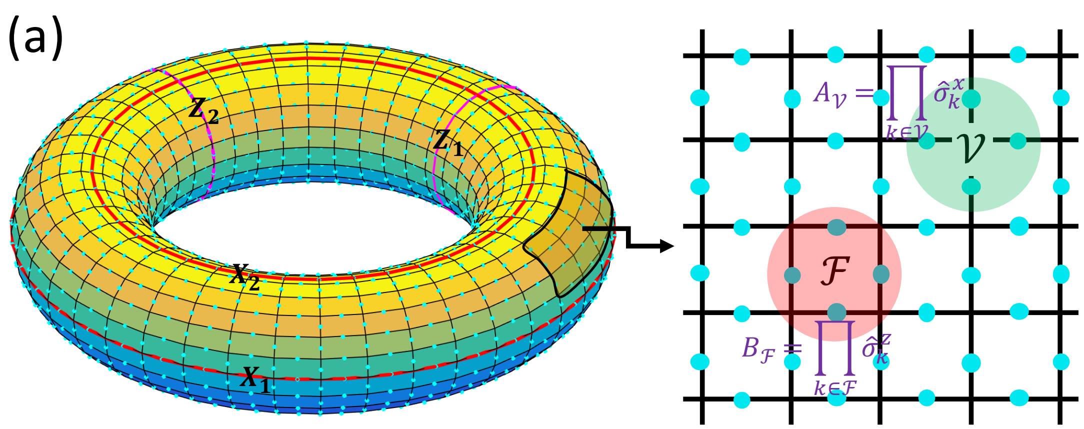
$$\Psi(s_1 \ldots s_N) = g^{(L)} \circ W^{(L)} \ldots g^{(2)} \circ W^{(2)} g^{(1)} \circ W^{(1)} \mathbf{s}$$

Quantum Numbers

Nonlinear Activation Functions Applied Component-Wise

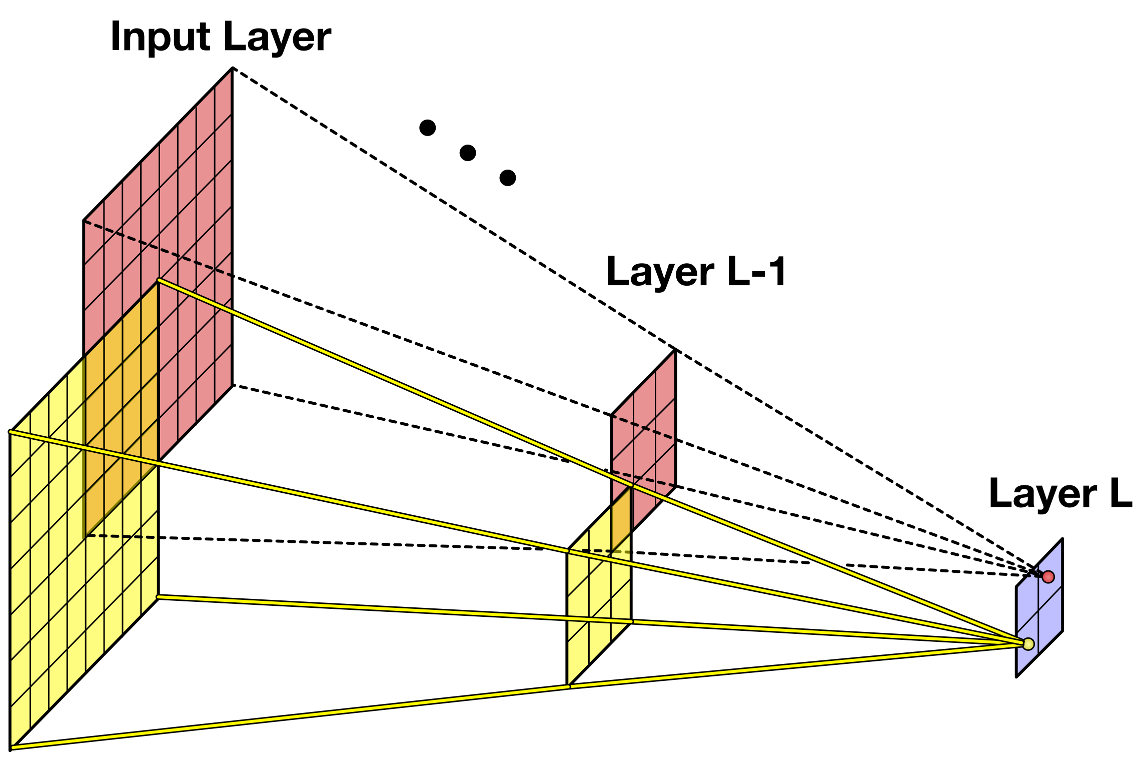
Variational Parameters

Some Properties



Compact Representations of Almost all known **efficient** Variational States (Jastrow, Laughlin, MPS, etc)

For a review: see Carleo et al.
Reviews of Modern Physics *91, 045002 (2020)*



Efficient Volume-Law States With Shallow or (better) Deep Convolutional Networks

Levine, Sharir, Cohen, and Shashua
Physical Review Letters *122, 065301 (2019)*

Ground-State Search

Variational Sampling

Expectation Minimization

$$\mathcal{L}(\mathbf{W}) = \frac{\sum_s |\Psi(s,W)|^2 E_{\rm loc}(s,W)}{\sum_s |\Psi(s,W)|^2}$$

McMillan
Phys. Rev. 138, A442 (1965)

`Loss Function`

`No “Dataset” here,` `closer in spirit to reinforcement learning`

Energy Gradient

$$\nabla_k E = 2\left(\langle \mathcal{O}_k^{*} E_{\mathrm{loc}}\rangle - \langle \mathcal{O}_k^{*}\rangle\langle E_{\mathrm{loc}}\rangle\right)$$

$$E_{\mathrm{loc}}(\sigma_1 \ldots \sigma_N) = \sum_{\sigma'_1,\ldots\sigma'_N} \frac{\Psi(\sigma'_1 \ldots \sigma'_N)}{\Psi(\sigma_1,\ldots\sigma_N)} \mathcal{H}_{\sigma,\sigma'}$$

$$\mathcal{O}_k(\sigma_1 \ldots \sigma_N) = \frac{1}{\Psi(\sigma_1 \ldots \sigma_N)} \frac{\partial \Psi(\sigma_1 \ldots \sigma_N)}{\partial p_k}$$

$$\langle \ldots \rangle = \frac{\sum_{\sigma_1 \sigma_2 \ldots \sigma_N} \left|\Psi(\sigma_1, \sigma_2 \ldots \sigma_N)\right|^2 \ldots}{\sum_{\sigma_1 \sigma_2 \ldots \sigma_N} \left|\Psi(\sigma_1, \sigma_2 \ldots \sigma_N)\right|^2}$$

Computationally Tractable States

“Computationally Tractable” States

Van den Nest
arXiv:0911.1624 (2009)

Definition 1 *An n-qubit state $|\psi\rangle$ is called ‘computationally tractable’ (CT) if the following conditions hold:*

(a) it is possible to sample in poly(n) time with classical means from the probability distribution $Prob(x) = |\langle x|\psi\rangle|^2$ on the set of n-bit strings x, and

(b) upon input of any bit string x, the coefficient $\langle x|\psi\rangle$ can be computed in poly(n) time on a classical computer.

Theorem 3 *Let $|\psi\rangle$ and $|\varphi\rangle$ be CT n-qubit states and let A be an efficiently computable sparse (not necessarily unitary) n-qubit operation with $\|A\| \leq 1$. Then there exists an efficient classical algorithm to approximate $\langle\varphi|A|\psi\rangle$ with polynomial accuracy.*

Corollary 1 *Let $|\psi\rangle$ be an n-qubit CT state and let O be a d-local observable with $d = O(\log n)$ and $\|O\| \leq 1$. Then there exists an efficient classical algorithm to estimate $\langle\psi|O|\psi\rangle$ with polynomial accuracy.*

Examples

Matrix Product States Are Computationally Tractable

PEPS are **not** Computationally Tractable

Generic neural deep quantum states are **not** computationally Tractable

NQS: MCMC Breaks Tractability

$$\langle \dots \rangle = \frac{\sum_{\sigma_1 \sigma_2 \dots \sigma_N} \left|\Psi(\sigma_1, \sigma_2 \dots \sigma_N)\right|^2 \dots}{\sum_{\sigma_1 \sigma_2 \dots \sigma_N} \left|\Psi(\sigma_1, \sigma_2 \dots \sigma_N)\right|^2}$$

Historically, MCMC is used to sample from this probability

For deep networks: can Be Inefficient Both Theoretically (Correlation Times) and Computationally (GPU)

Autoregressive Quantum States

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

$$\Psi(s_1, \ldots, s_N) = \prod_{i=1}^{N} \psi_i(s_i | s_{i-1}, \ldots, s_1)$$

$$\sum_{s'} |\psi_i(s' | s_{i-1}, \ldots, s_1)|^2 = 1$$

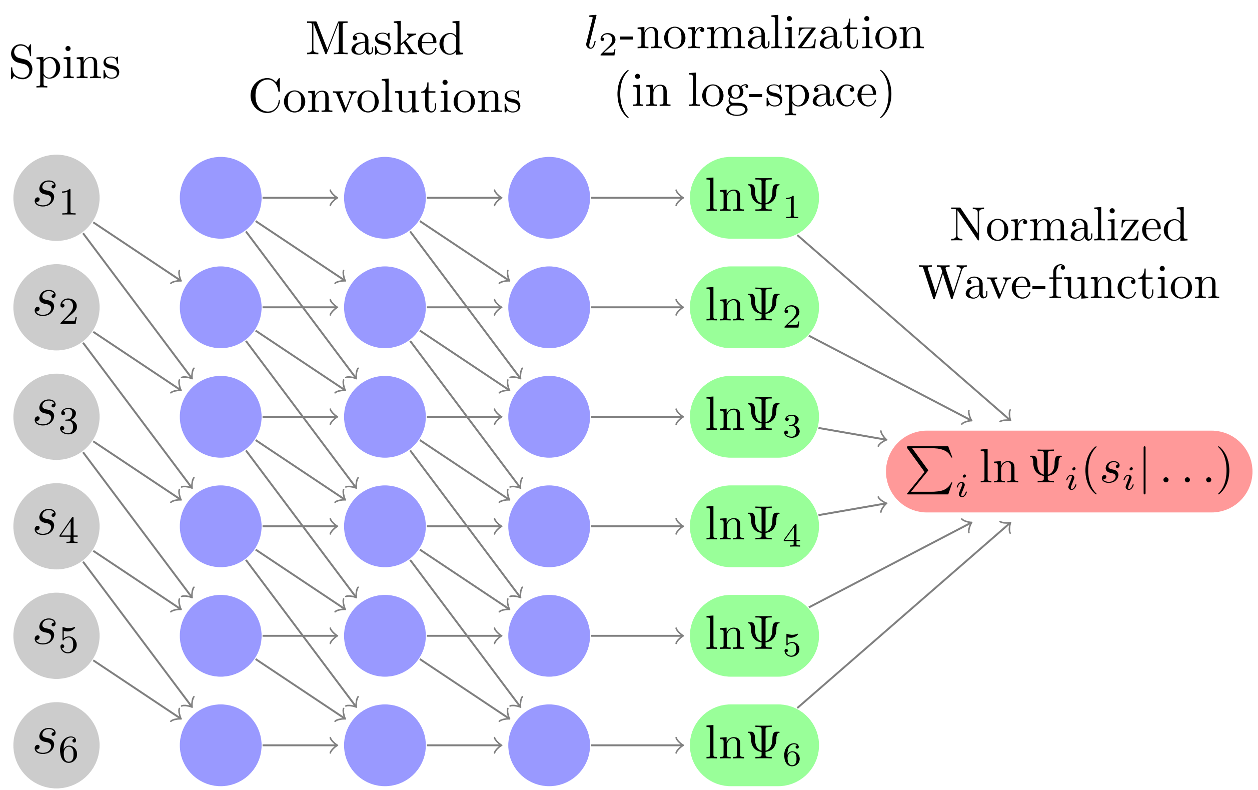
Normalized "Conditionals"

These Are Computationally Tractable

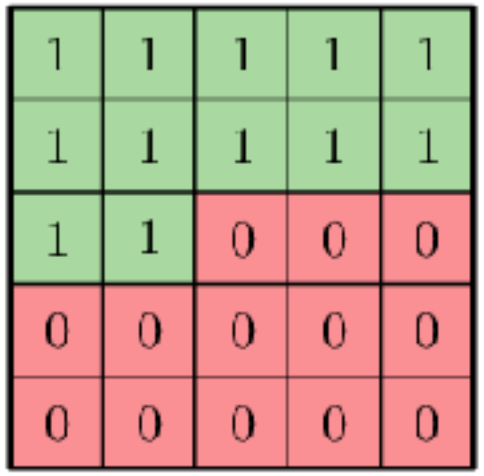
(a) Exact Sampling

(b) Computing Normalized Amplitudes is Efficient

Masked deep networks



Masked Fully Connected Network



Masked Convolutions

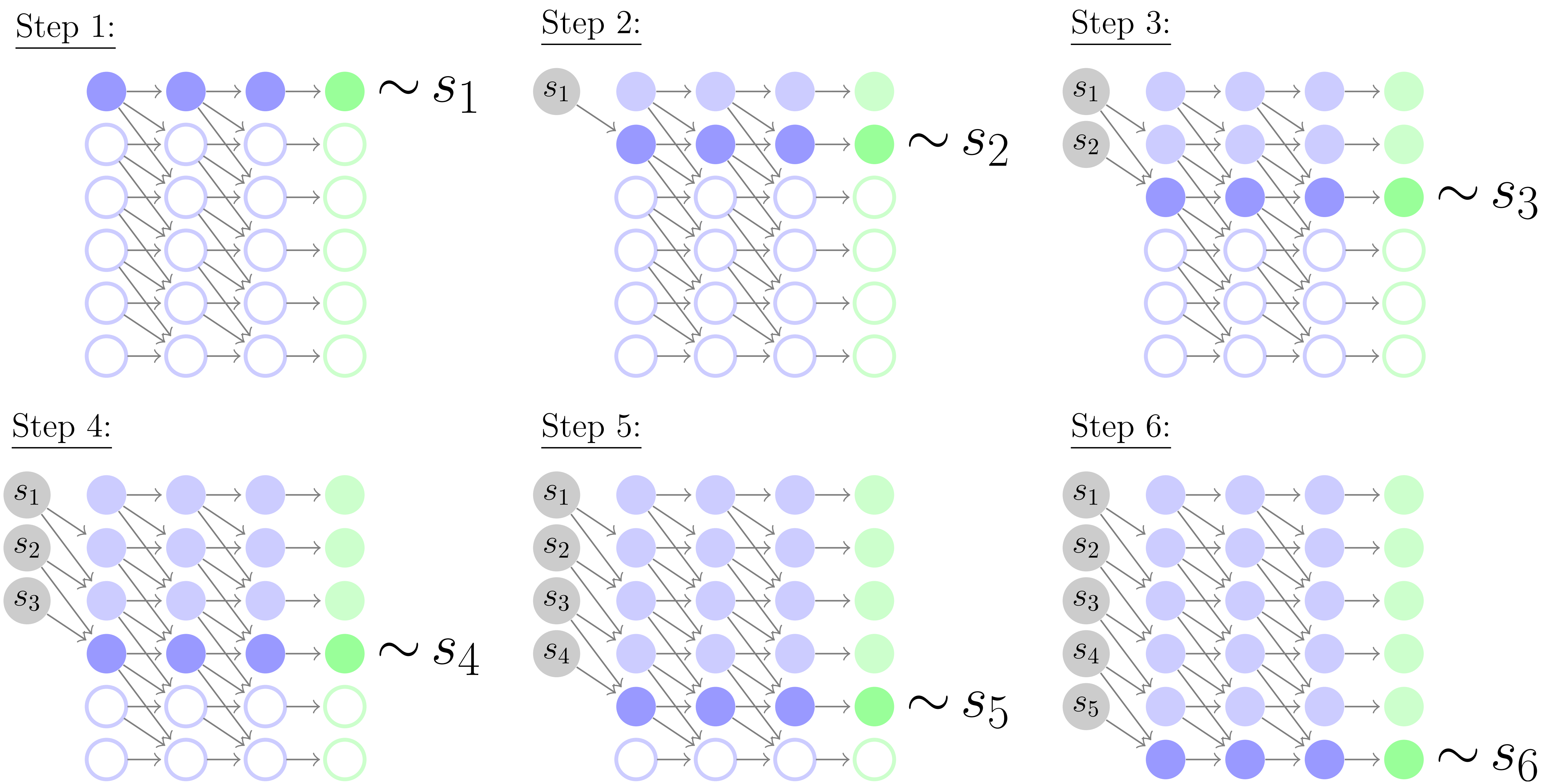
PixelCNN

Van den Oord et al. arXiv:1606.05328 (2016)

Salimans et al. arXiv:1701.05517 (2017)

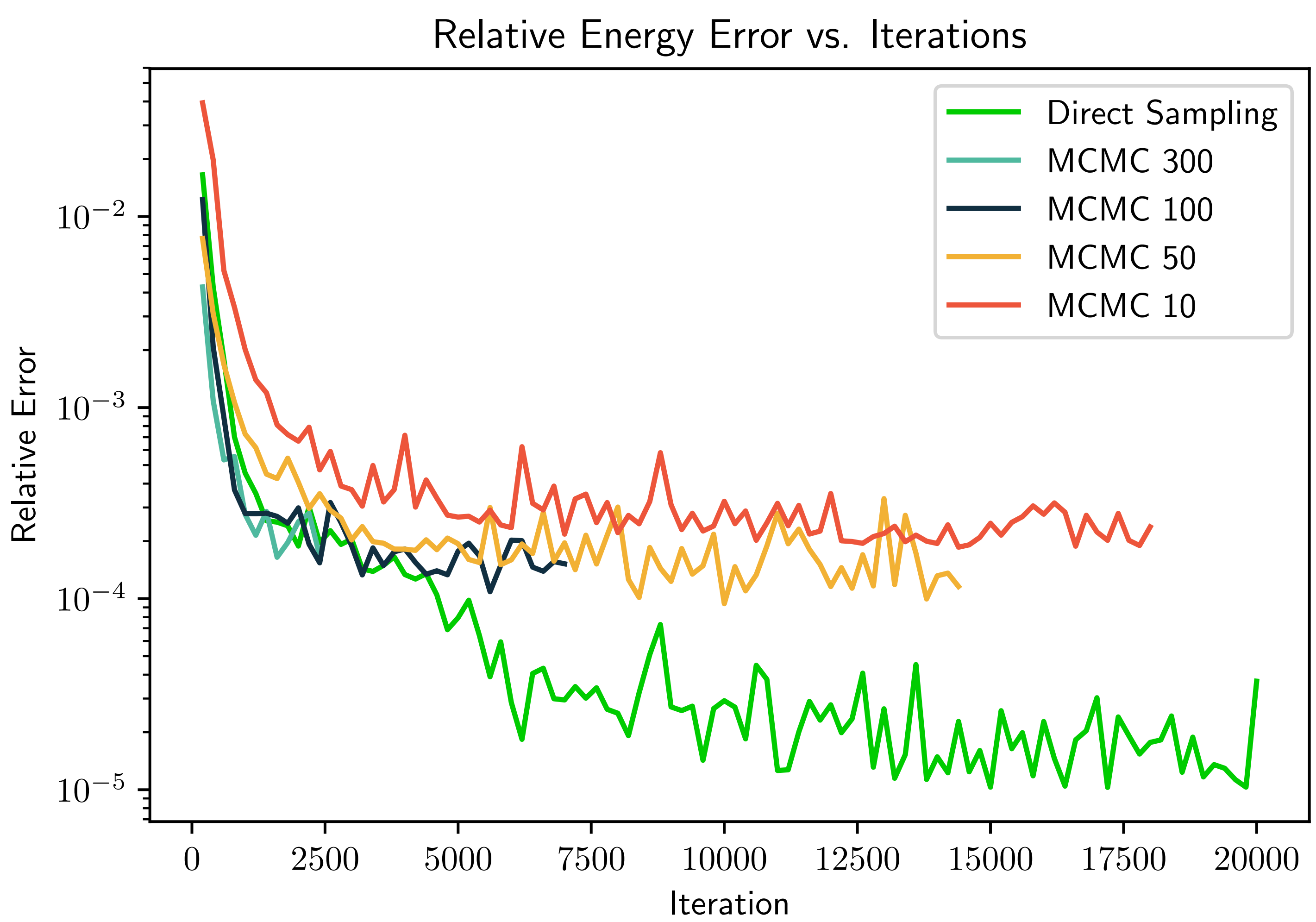
Ramachandran et al. arXiv:1704.06001 (2017)

Exact Sampling



Removing The Sampling Bottleneck Pays Off

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

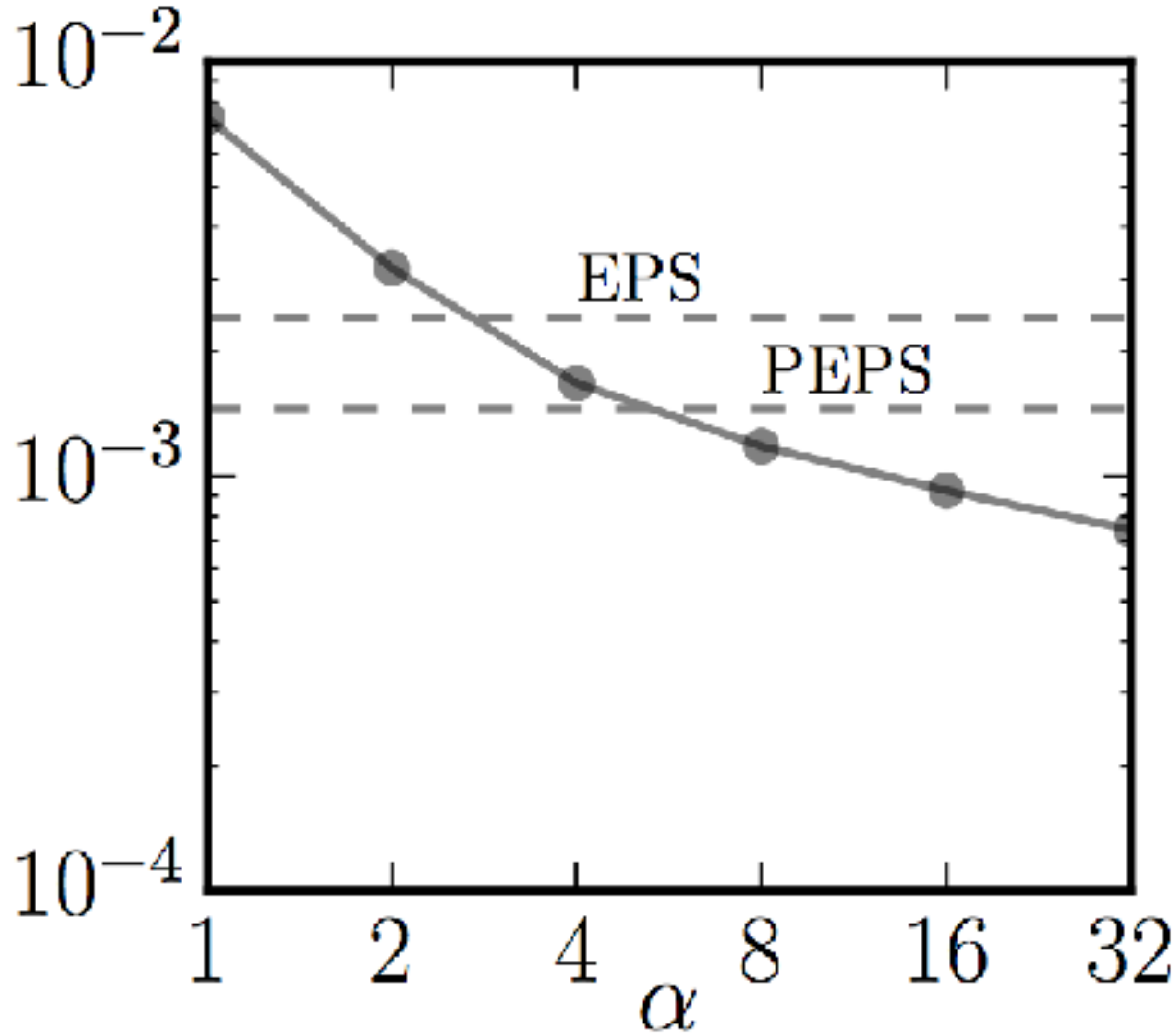


21x21 Transverse-Field Ising in 2d

Depth 20
About 1 Million Parameters

Heisenberg Model

Carleo, and Troyer
Science 355, 602 (2017)

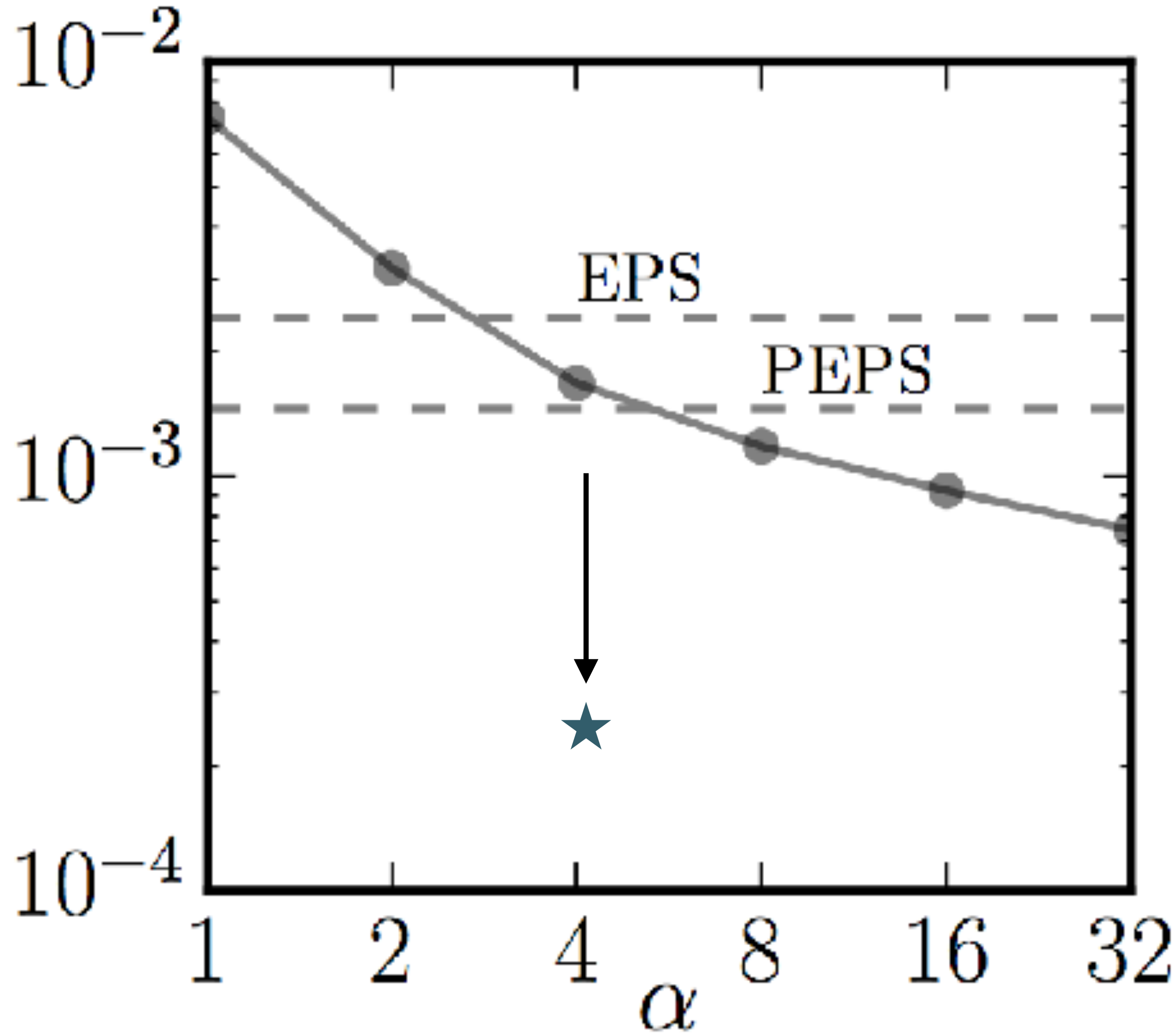


10 by 10 cluster

```
Early (2016) Results With
Shallow (RBM) Network
```

Heisenberg Model

Choo, Neupert, and Carleo
Phys. Rev. B 100, 125124 (2019)

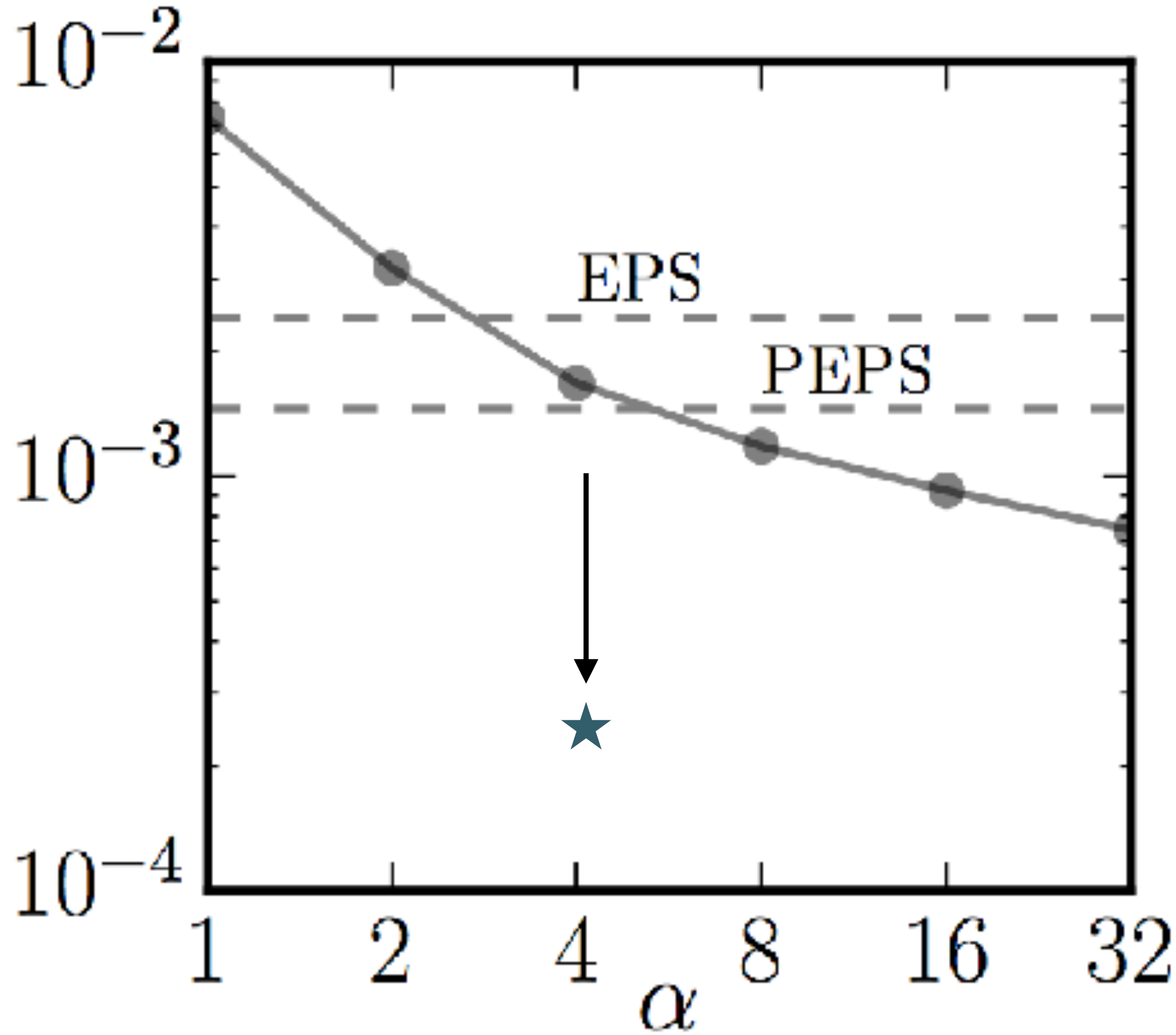


10 by 10 cluster

(Mildly) deep network further improves

Heisenberg Model

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



10 by 10 cluster

(Actual) deep network further improves

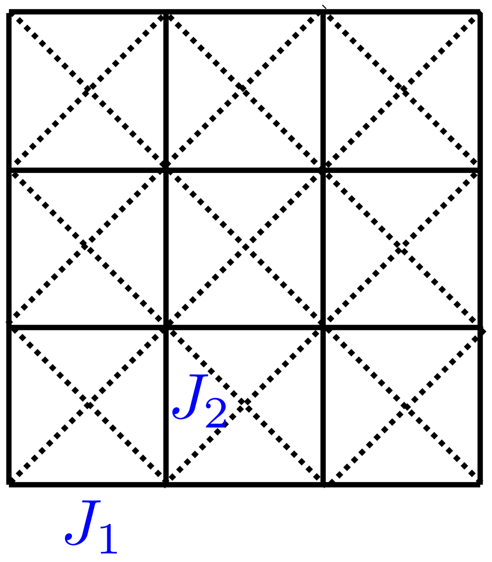
~4x10^-5

A Challenging Benchmark

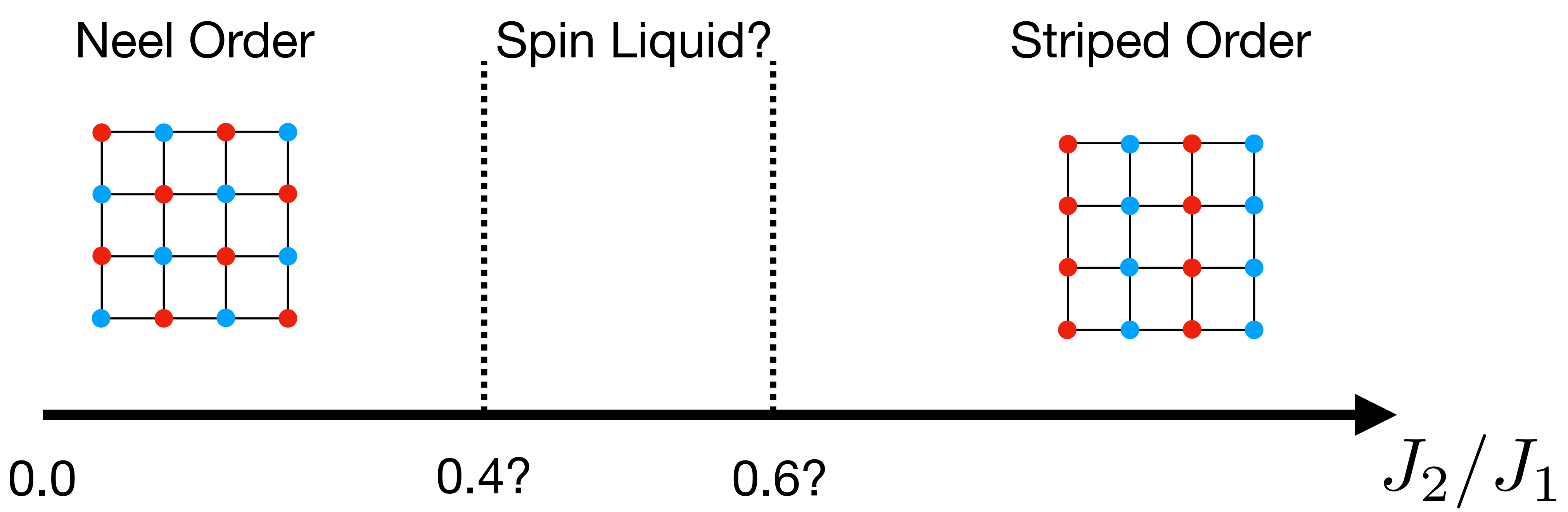
Frustrated 2D Spins

J1-J2 Model

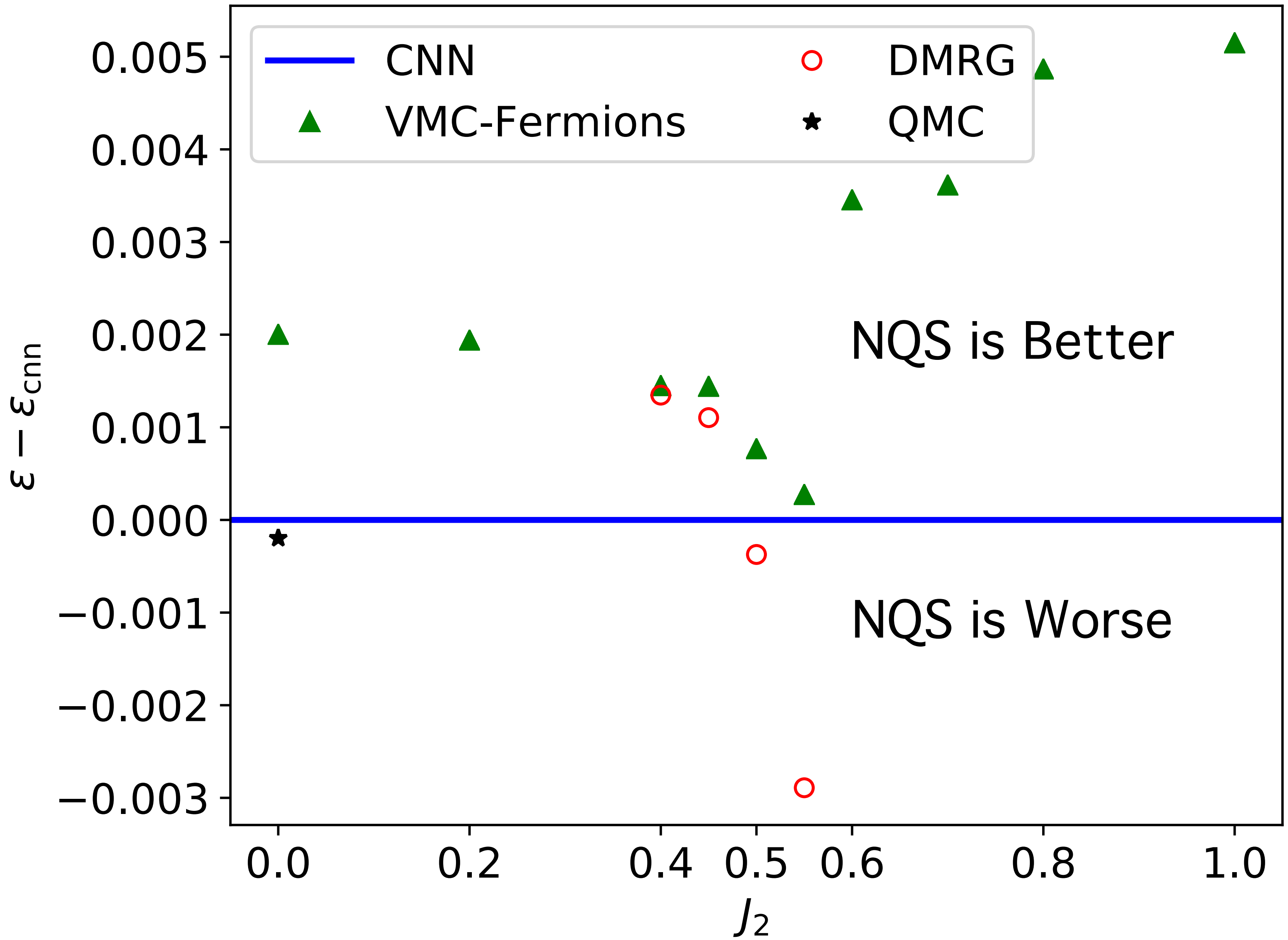
$$\hat{H} = J_1 \sum_{\langle ij \rangle} \hat{\boldsymbol{S}}_i \cdot \hat{\boldsymbol{S}}_j + J_2 \sum_{\langle\langle ij \rangle\rangle} \hat{\boldsymbol{S}}_i \cdot \hat{\boldsymbol{S}}_j$$



Phase Diagram



J1-J2 Model Square Lattice

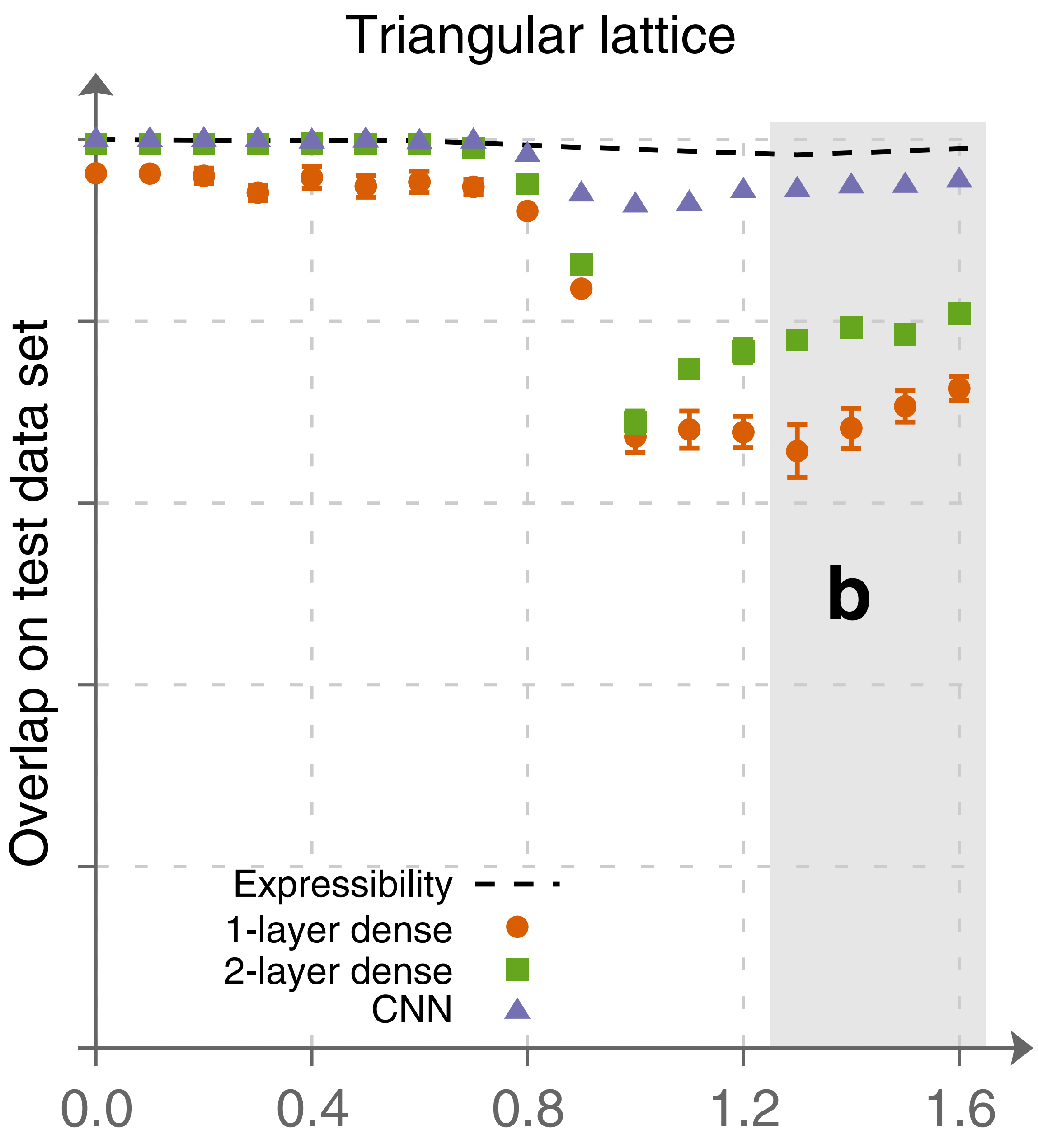


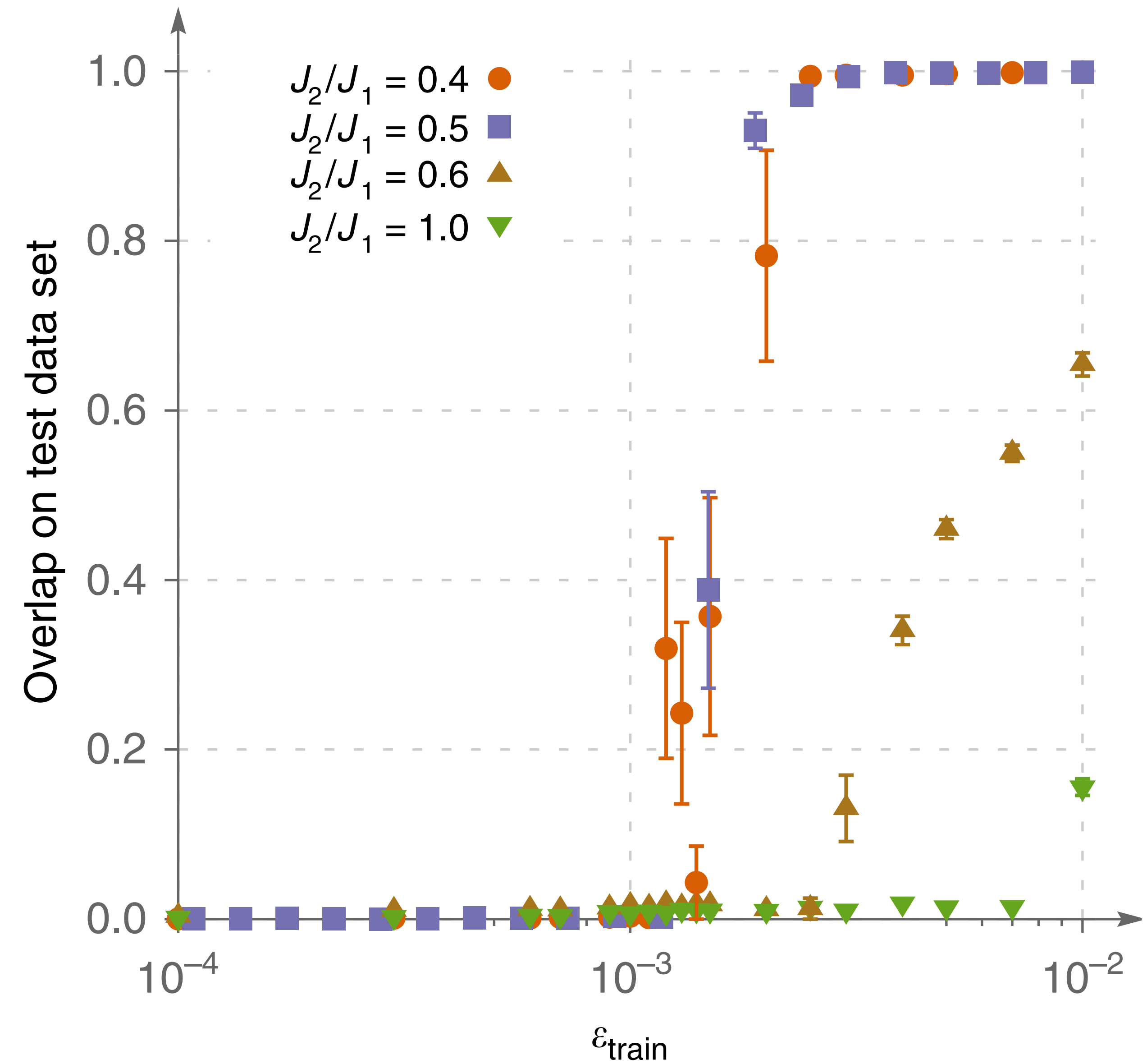
10 by 10 cluster

Choo, Neupert, and Carleo
Phys. Rev. B 100, 125124 (2019)

```
Tangible Improvements on Open
     Challenging Models
```

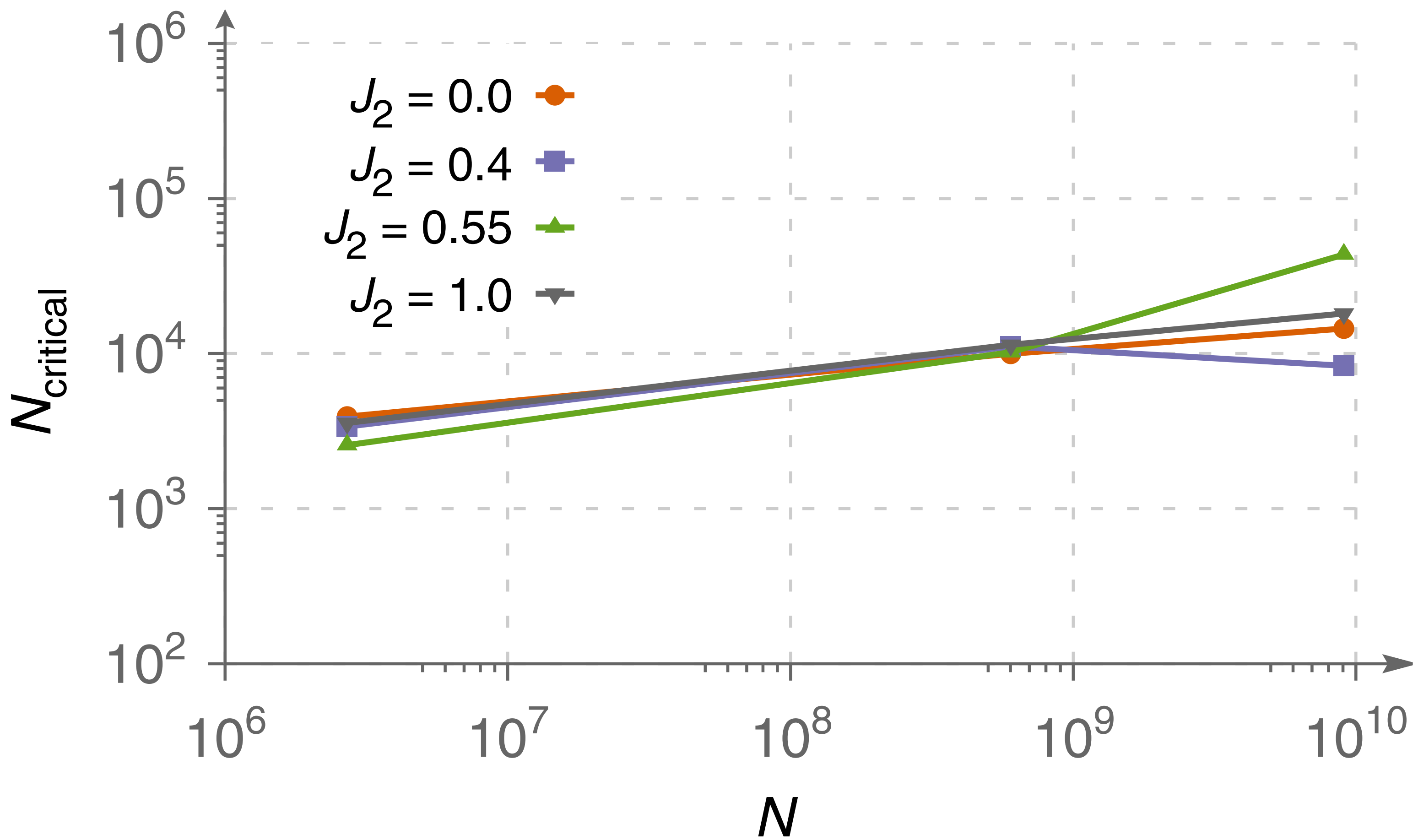
Origin of the challenge: generalization for sign structure





Westerhout, Astrakhantsev, Tikhonov, Katsnelson, Bagrov
Nature Comm. 11, 1593 (2020)

Main Open Issue: Sample Complexity For Sign Structure

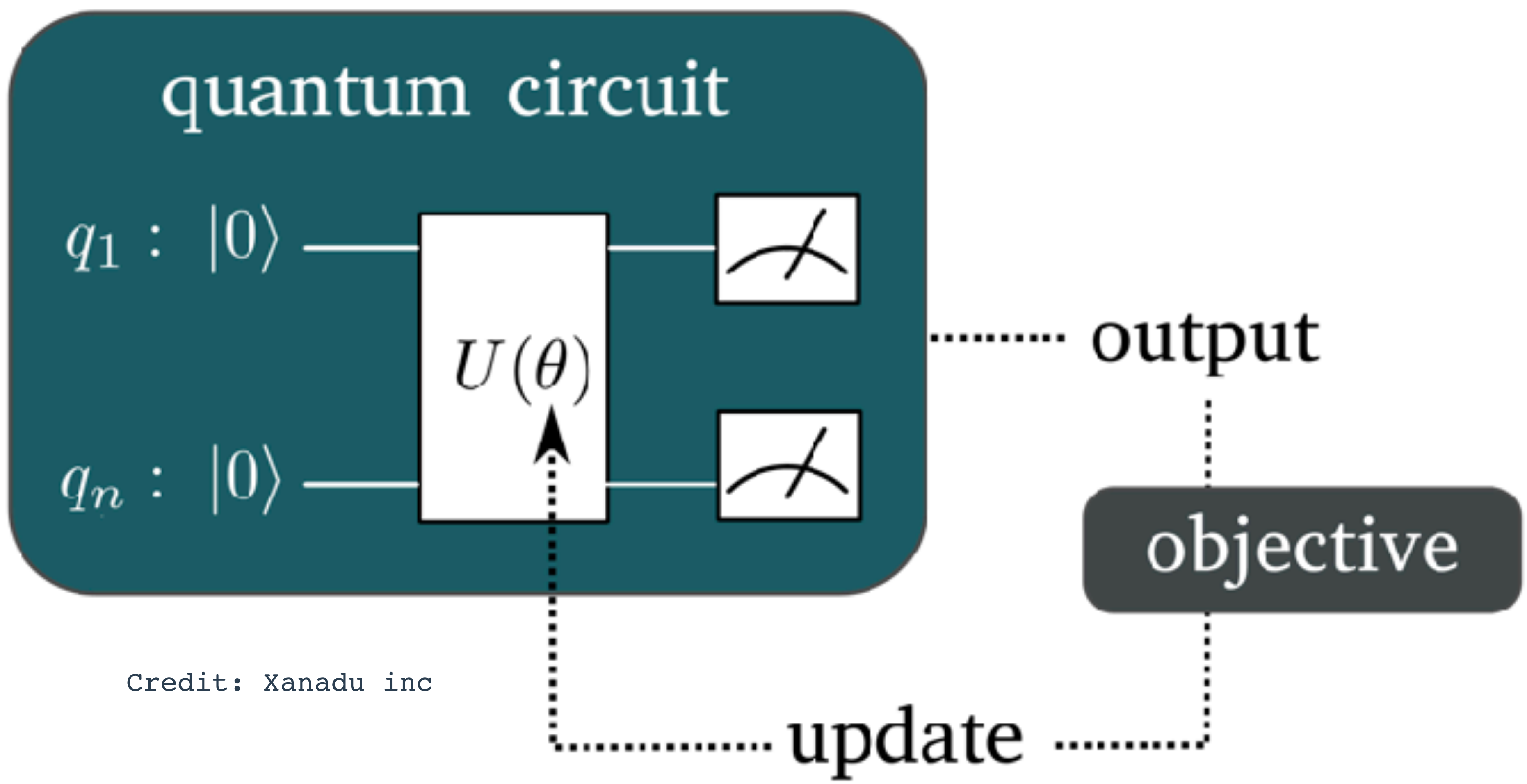


N_s=20 N_s=36

Westerhout, Astrakhantsev, Tikhonov, Katsnelson, Bagrov
Nature Comm. 11, 1593 (2020)

Quantum Variational States

Variational Quantum Eigensolvers



Credit: Xanadu inc

$$E(\theta) = \langle \Psi_0 | U(\theta)^\dagger \mathcal{H} U(\theta) | \Psi_0 \rangle$$

$$\mathcal{H} = \sum_k c_k \sigma_1^{(x,y,z)} \ldots \sigma_N^{(x,y,z)}$$

Energy is Estimated Stochastically on Samples From Measurements

Strong Connection With Stochastic Classical Counterparts

Stokes, Izaac, Killoran, and Carleo
Quantum 4, 269 (2020)

Higher-Order Optimizer for Quantum Machine Learning

Quantum Variational Imaginary-Time Evolution

Quantum Variational Real-Time Evolution

Stochastic Estimates of the Quantum Geometric Tensor

Stochastic Reconfiguration

Time-Dependent Variational Monte Carlo

Natural Gradient

Stochastic Reconfiguration (Natural Gradient)

Sandro Sorella et al.
Physical Review Letters
80, 4558 (1998)

Shun-Ichi Amari
Journal Neural Computation
10, 251 (1998)

$$\sum_{k'} S_{k,k'} \Delta p_{k'} = -G_k$$

"Second-Order" Method

$$S_{k,k'} = \langle \mathcal{O}_k^\star \mathcal{O}_{k'} \rangle - \langle \mathcal{O}_k^\star \rangle \langle \mathcal{O}_{k'} \rangle$$

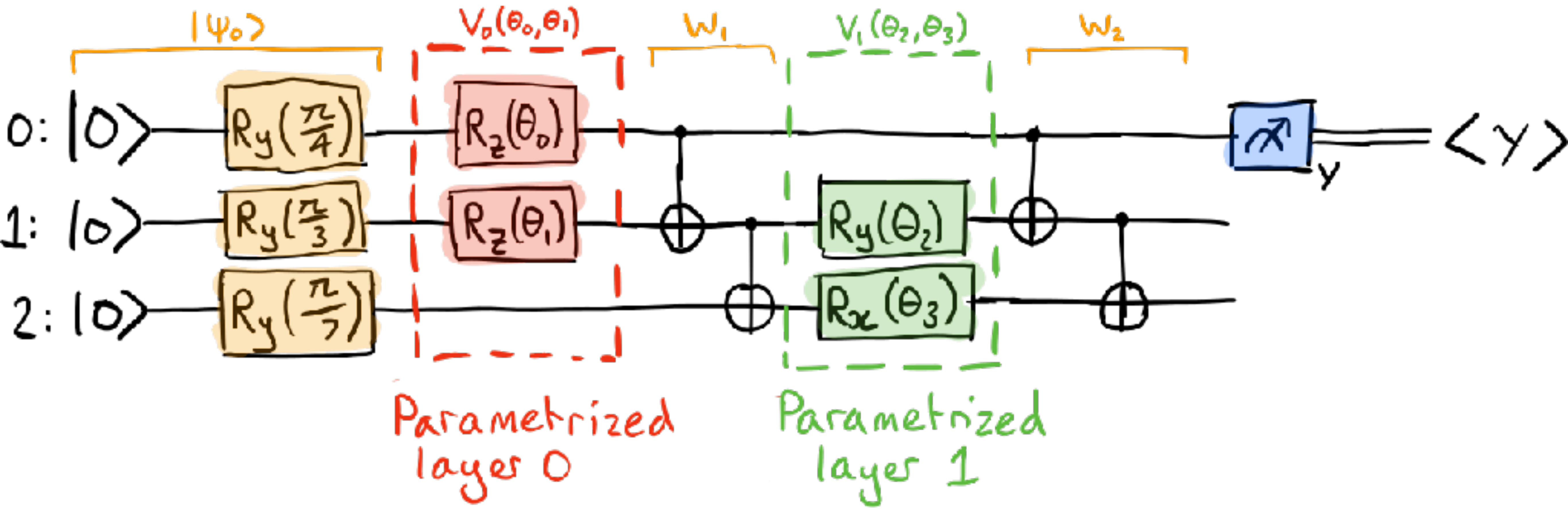
Quantum Geometric Tensor or Quantum Fisher Information

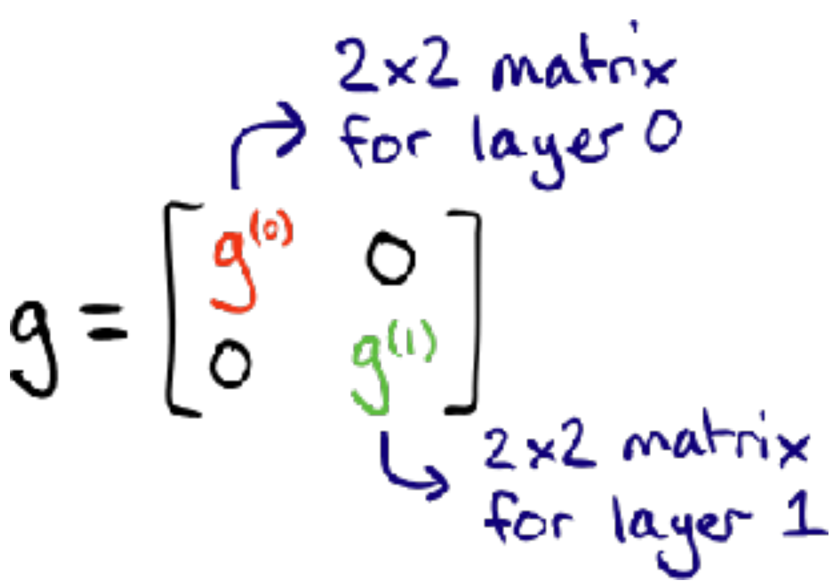
Linear System to be Solved at each iteration of the optimizer

Sparse solvers can be used when dealing with large number of parameters

Quantum Natural Gradient

Stokes, Izaac, Killoran, and Carleo
Quantum 4, 269 (2020)

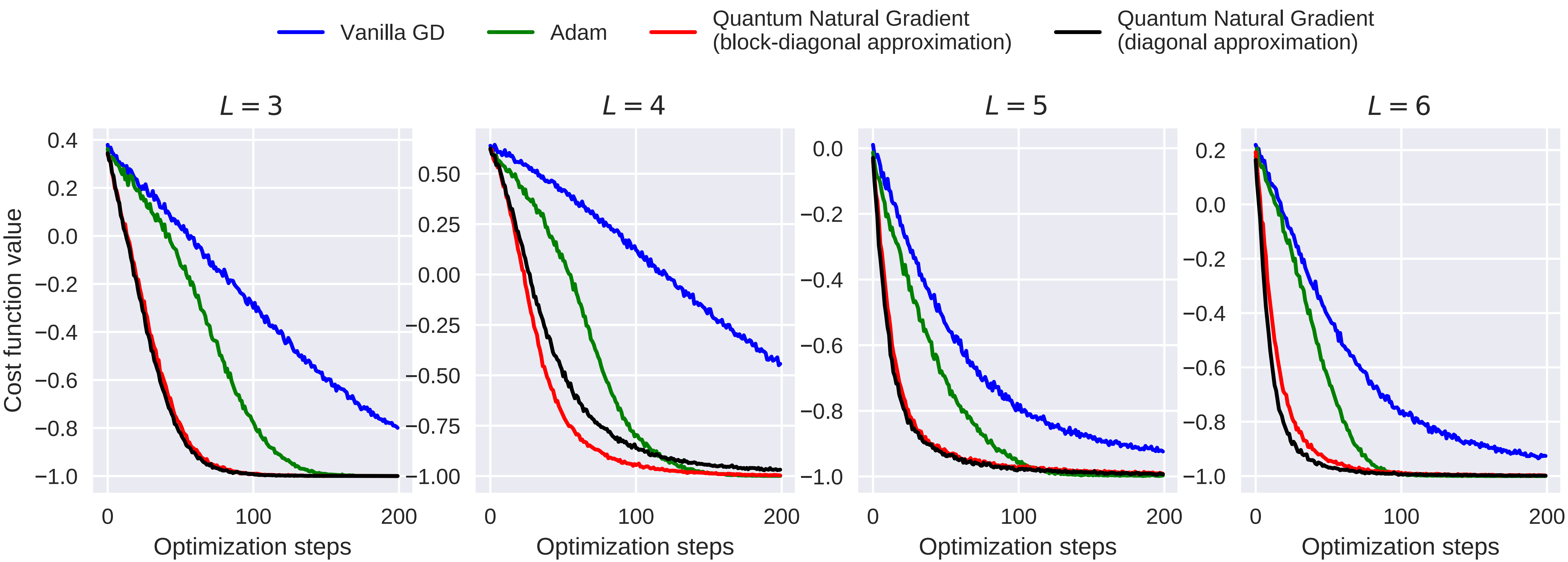




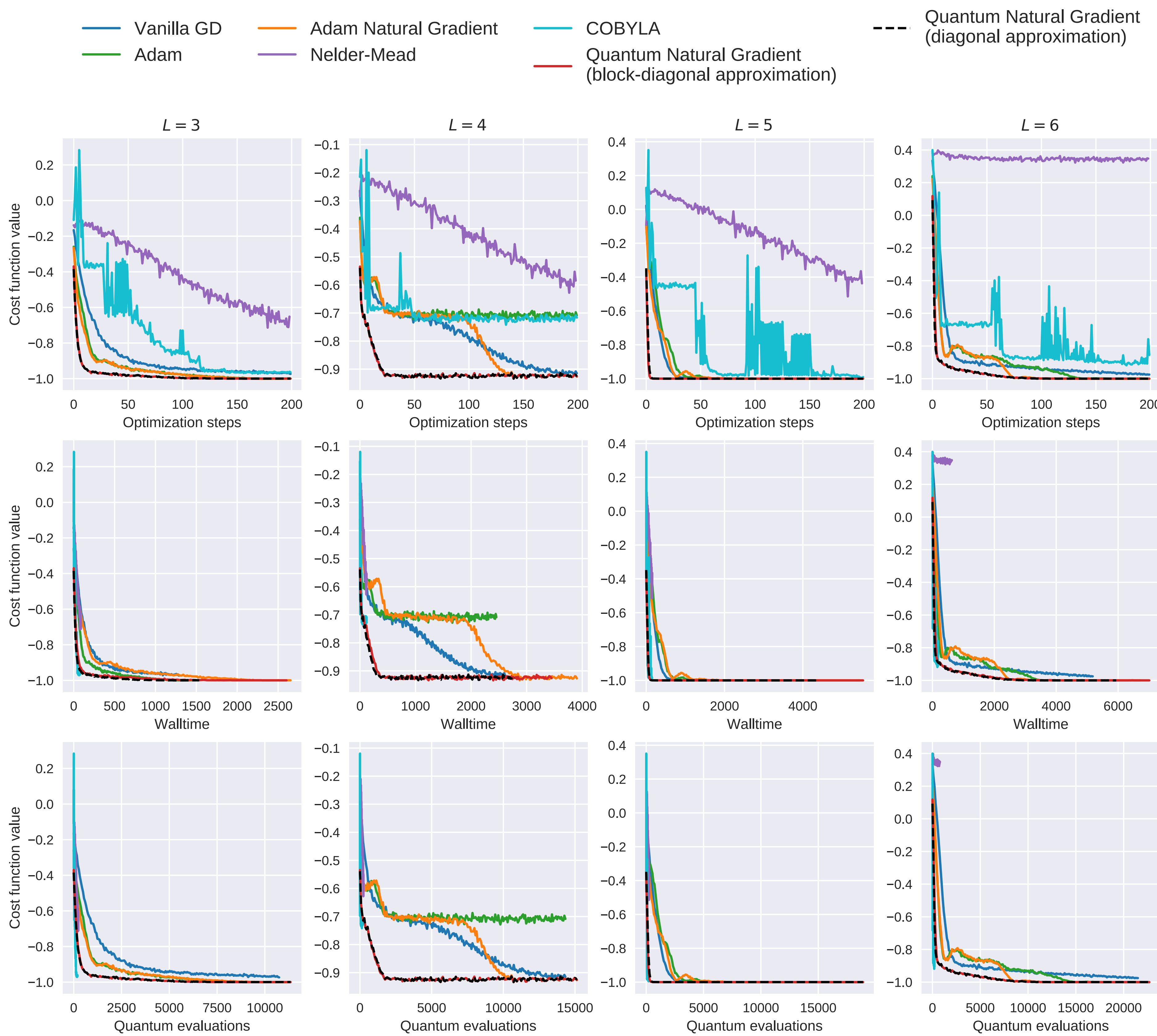
$$\theta_{t+1} = \theta_t - \eta g^{+}(\theta)\nabla L(\theta),$$

Faster Convergence

Stokes, Izaac, Killoran, and Carleo
Quantum 4, 269 (2020)



Quantum Evaluations



Software

The NetKet Project

www.netket.org

NetKet: A Machine Learning Toolkit for Many-Body Quantum Systems

Giuseppe Carleo,[1] Kenny Choo,[2] Damian Hofmann,[3] James E. T. Smith,[4] Tom Westerhout,[5] Fabien Alet,[6] Emily J. Davis,[7] Stavros Efthymiou,[8] Ivan Glasser,[8] Sheng-Hsuan Lin,[9] Marta Mauri,[1,10] Guglielmo Mazzola,[11] Christian B. Mendl,[12] Evert van Nieuwenburg,[13] Ossian O'Reilly,[14] Hugo Théveniaut,[6] Giacomo Torlai,[1] and Alexander Wietek[1]

[1] *Center for Computational Quantum Physics, Flatiron Institute, 162 5th Avenue, NY 10010, New York, USA*
[2] *Department of Physics, University of Zurich, Winterthurerstrasse 190, 8057 Zürich, Switzerland*
[3] *Max Planck Institute for the Structure and Dynamics of Matter, Luruper Chaussee 149, 22761 Hamburg, Germany*
[4] *Department of Chemistry, University of Colorado Boulder, Boulder, Colorado 80302, USA*
[5] *Institute for Molecules and Materials, Radboud University, NL-6525 AJ Nijmegen, The Netherlands*
[6] *Laboratoire de Physique Théorique, IRSAMC, Université de Toulouse, CNRS, UPS, 31062 Toulouse, France*
[7] *Department of Physics, Stanford University, Stanford, California 94305, USA*
[8] *Max-Planck-Institut für Quantenoptik, Hans-Kopfermann-Straße 1, 85748 Garching bei München, Germany*
[9] *Department of Physics, T42, Technische Universität München, James-Franck-Straße 1, 85748 Garching bei München, Germany*
[10] *Dipartimento di Fisica, Università degli Studi di Milano, via Celoria 16, I-20133 Milano, Italy*
[11] *Theoretische Physik, ETH Zürich, 8093 Zürich, Switzerland*
[12] *Technische Universität Dresden, Institute of Scientific Computing, Zellescher Weg 12-14, 01069 Dresden, Germany*
[13] *Institute for Quantum Information and Matter, California Institute of Technology, Pasadena, CA 91125, USA*
[14] *Southern California Earthquake Center, University of Southern California, 3651 Trousdale Pkwy, Los Angeles, CA 90089, USA*

SoftwareX 10, 100311 (2019)

```python
import netket as nk

# 1D Lattice
g = nk.graph.Hypercube(length=20, n_dim=1, pbc=True)

# Hilbert space of spins on the graph
hi = nk.hilbert.Spin(s=0.5, graph=g)

# Ising spin hamiltonian
ha = nk.operator.Ising(h=1.0, hilbert=hi)

# RBM Spin Machine
ma = nk.machine.RbmSpin(alpha=1, hilbert=hi)
ma.init_random_parameters(seed=1234, sigma=0.01)

# Metropolis Local Sampling
sa = nk.sampler.MetropolisLocal(machine=ma)

# Optimizer
op = nk.optimizer.Sgd(learning_rate=0.1)

# Stochastic reconfiguration
gs = nk.variational.Vmc(
    hamiltonian=ha,
    sampler=sa,
    optimizer=op,
    n_samples=1000,
    diag_shift=0.1,
    method='Sr')

gs.run(output_prefix='test', n_iter=300)
```

Release 3.0 (soon out...)

Same APIs but Almost Entire Rewriting Under the Hood

Pure Python:
Entirely Remove
C++ codebase

Numba Kernels
For Few Hotspots

Numba

Seamless
Integration with
DL Frameworks

Support for **GPU**
and **TPU**

GPU Tests on V100 Cards

Up to 100X Speedup on Gradients

Up to 30X Speedup on Sampling

Overview

Outlook/Challenges

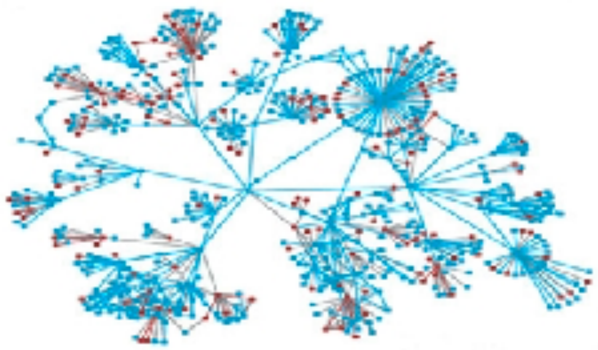
We need better
optimization strategies
and parameterizations
for signs/phases

Still "Heroic"
phase for
fermions
developments

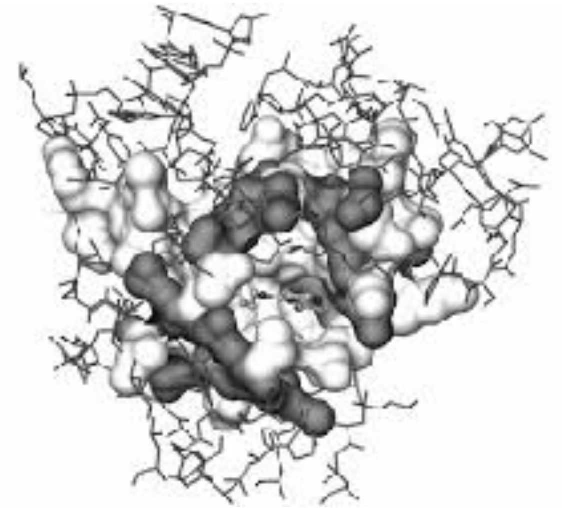
Symmetries in
networks play a
fundamental role

Machine Learning in Physics

Particle Physics



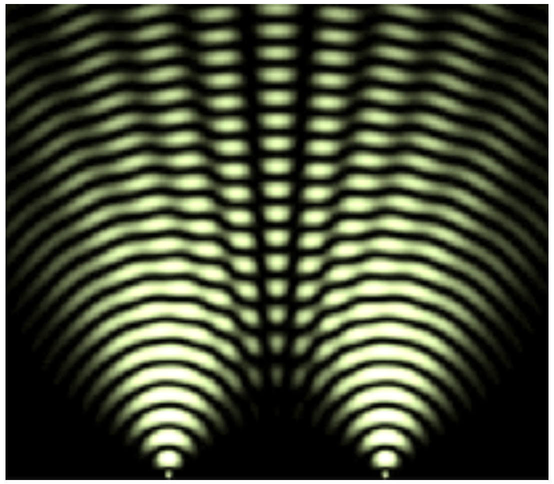
Statistical Physics



Chemistry/ Materials

Astrophysics

Carleo, Cirac, Cranmer, Daudet, Schuld, Tishby, Vogt, and Zdeborovà
Rev. Mod. Phys. 91, 045002 (2019)



Quantum Physics

Issue 2: Resources

Number of Measurements

The Problem

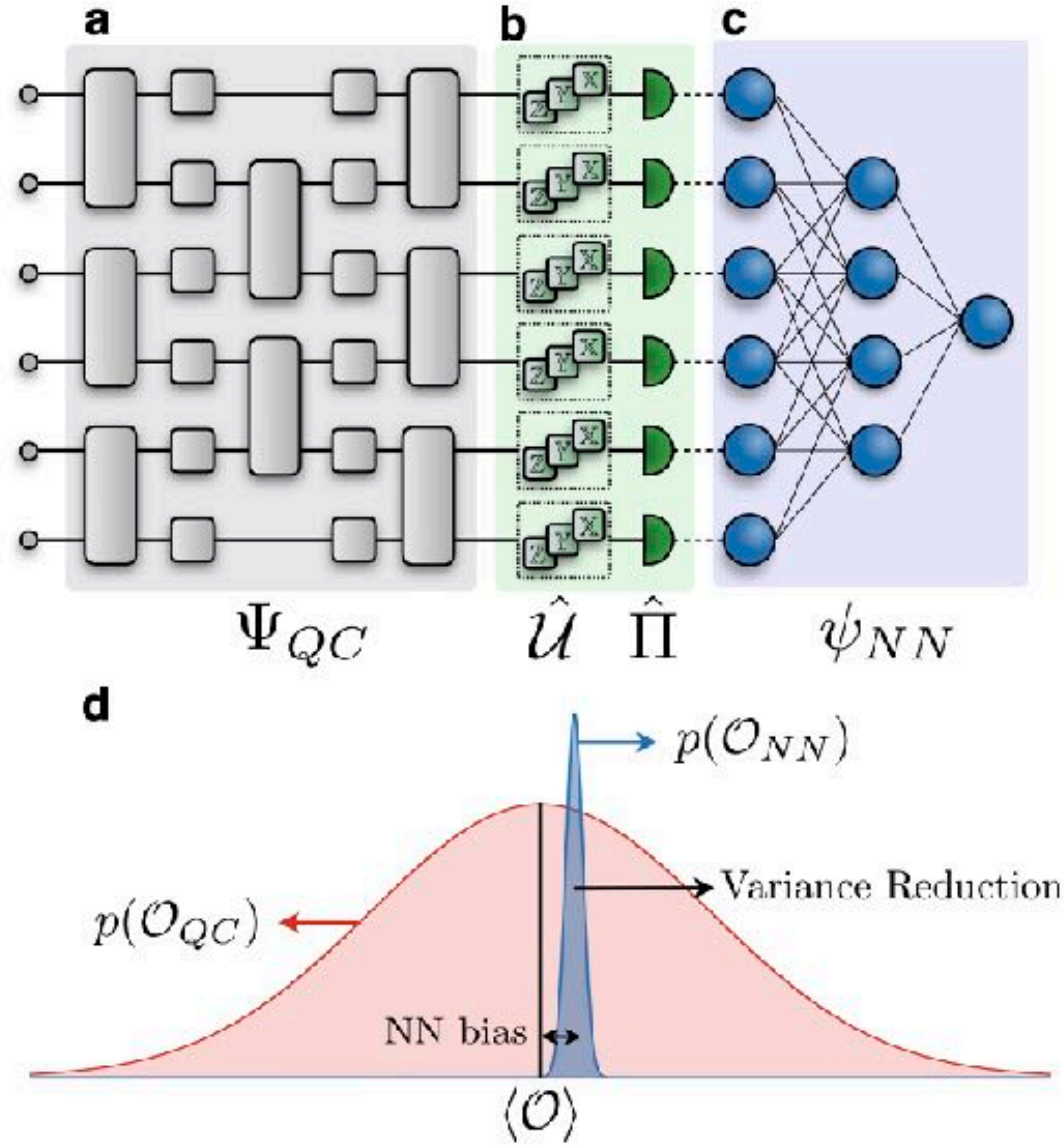
Precise Estimates of Physical Observables typically require several millions or more of measurements even on very small systems

$$\mathcal{H} = \sum_k c_k \sigma_1^{(x,y,z)} \dots \sigma_N^{(x,y,z)}$$

Example: Estimate Energy Using Measurements in Pauli Basis

Neural-Network State Parameterization

Torlai, Mazzola, Carleo, and Mezzacapo
Phys. Rev. Research 2, 022060 (2020)



Neural-Network Tomography

Torlai, Mazzola, Carrasquilla, Troyer, Melko, and Carleo
Nature Physics (2018)

Neural-Network Training: find $\mathbf{W}$ such that

$|\Psi_b(\mathbf{s}, \mathbf{W})|^2 \simeq P_b(\mathbf{s})$ in all given bases

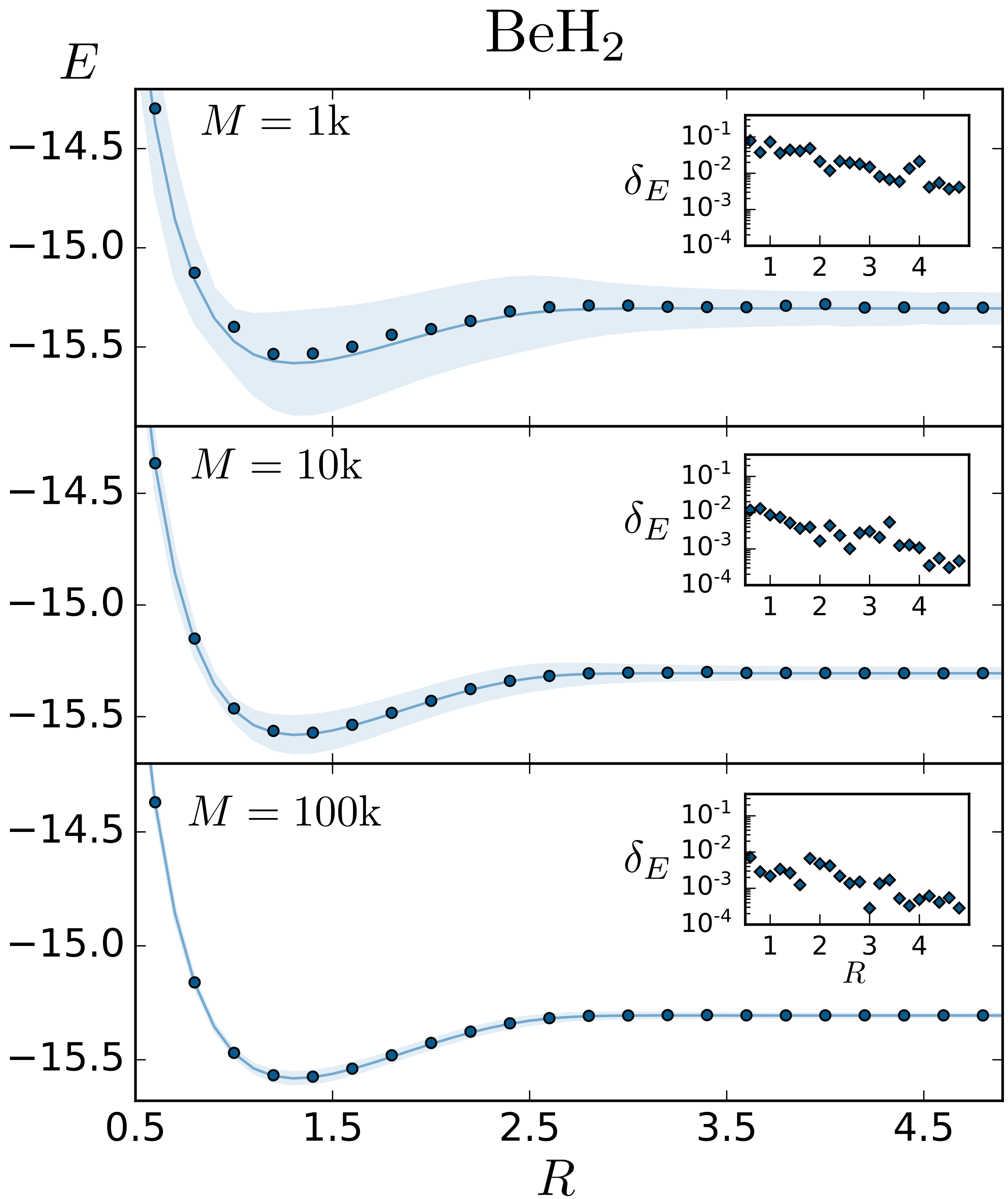
$$\Psi_b(\mathbf{s}, \mathbf{W}) = \sum_{\mathbf{s}'} \Psi(\mathbf{s}, \mathbf{W}) U^b_{s,s'}$$

Sparse Unitary Matrices

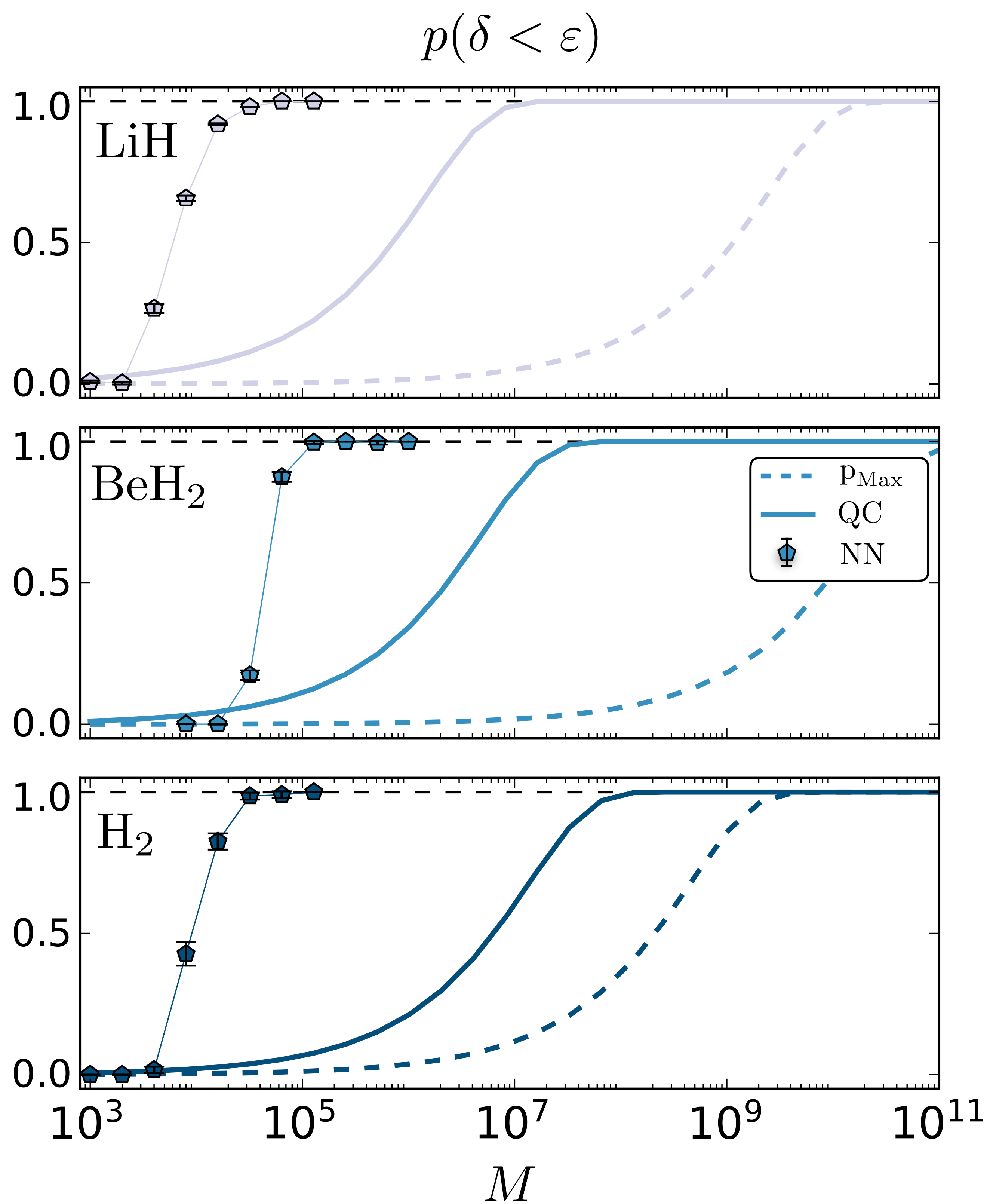
$$\mathcal{L}(\mathbf{W}) = \sum_b \sum_{\mathbf{s}} P_b(\mathbf{s}) \log \frac{P_b(\mathbf{s})}{|\Psi_b(\mathbf{s}, \mathbf{W})|^2}$$

Minimise Sum of Kullback-Leibler Divergences

Quantum Chemistry Problems



Prob. Of “Chemical Accuracy”



Reduce number of measurements by few orders of magnitude

Torlai, Mazzola, Carleo, and Mezzacapo
Phys. Rev. Research 2, 022060 (2020)