Learning Solutions to the Schrödinger equation with Neural-Network Quantum States

Giuseppe Carleo

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

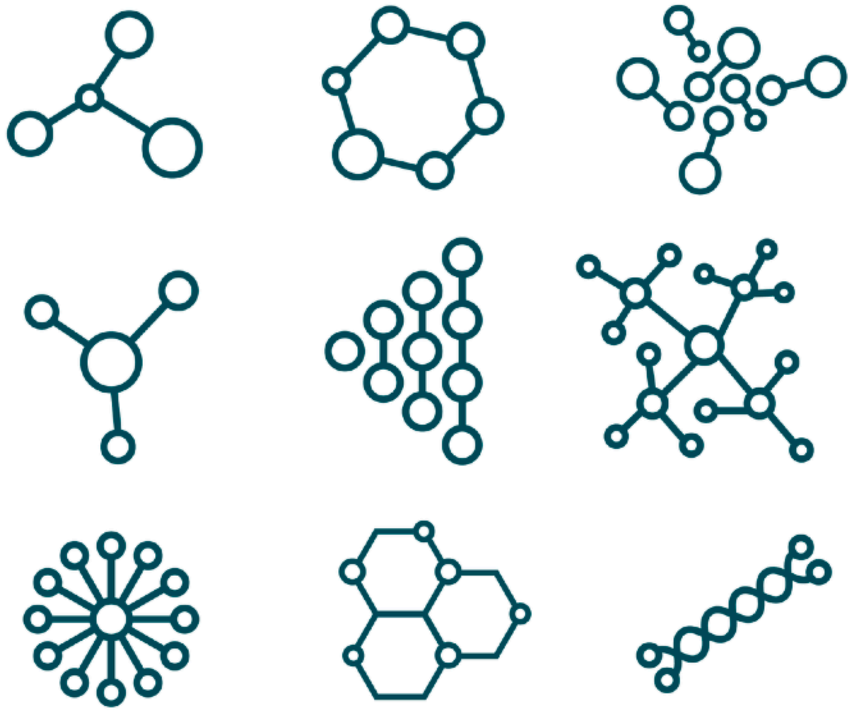
Computational Quantum Science Lab.

EPFL

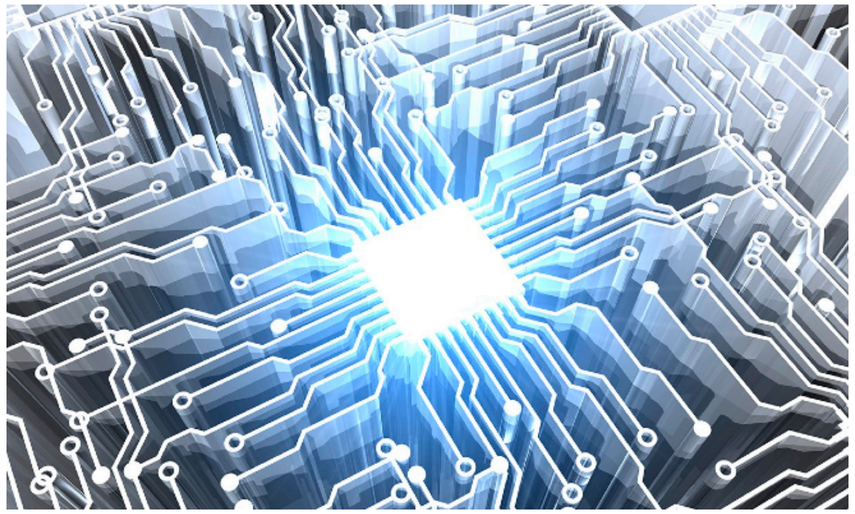
01.

The Quantum Many-Body Problem.

01.1 - Interacting Quantum Matter



E.g.
Interacting Particles in Chemistry, Material Science, Atomic Physics, Nuclear Physics...

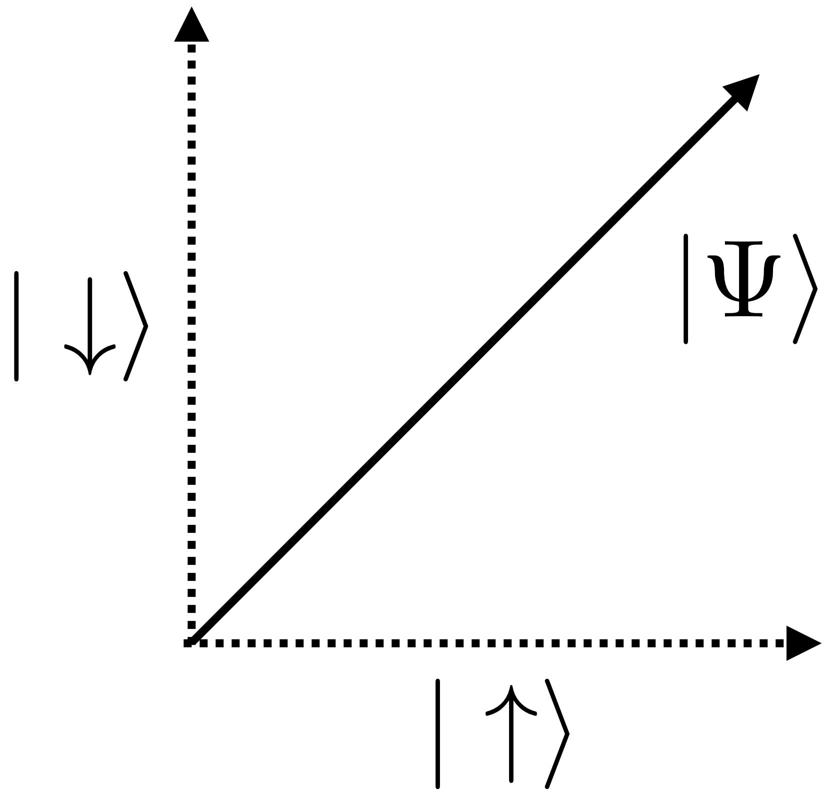


E.g.
Harnessing Entanglement in Quantum Computers, Quantum Simulators...

01.2 - Refresher: Quantum States

The state of a quantum spin is a complex-valued **vector**

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



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

01.3 - The Many-Body Wave Function

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

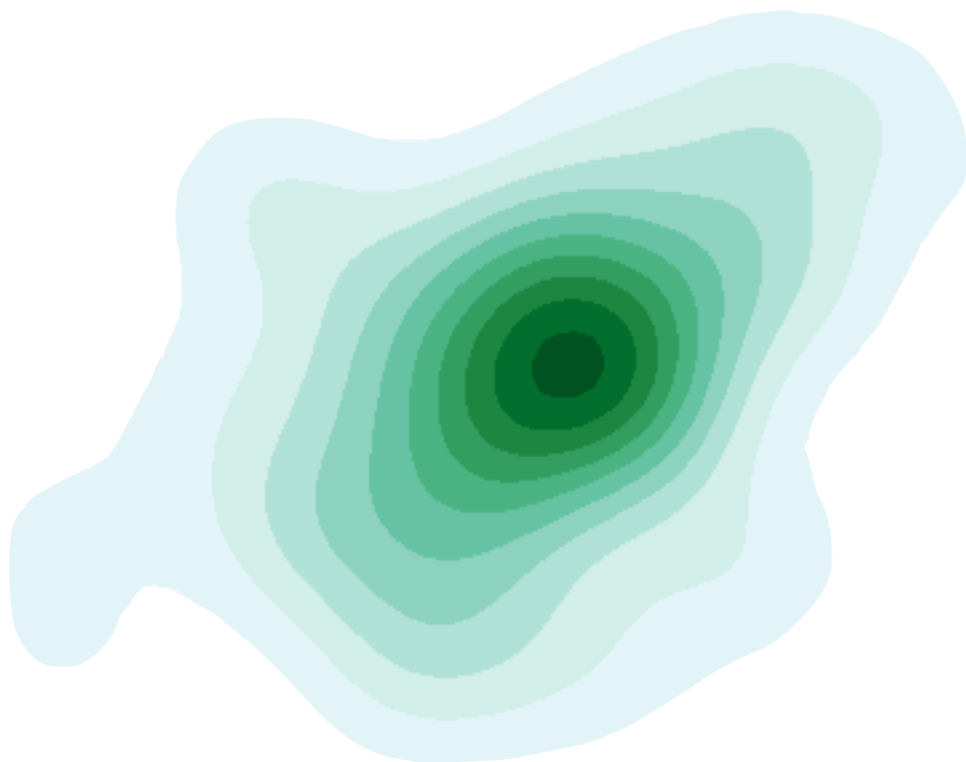
Complex-Valued Coefficients

The Wave Function is a Vector
in a Huge (2^N)
Space

The state of N
quantum particles
is a high-dimensional
"monster"

"In general the many-electron wave-function for a system of many electrons is not a legitimate scientific concept"

W. Kohn, Nobel Lecture



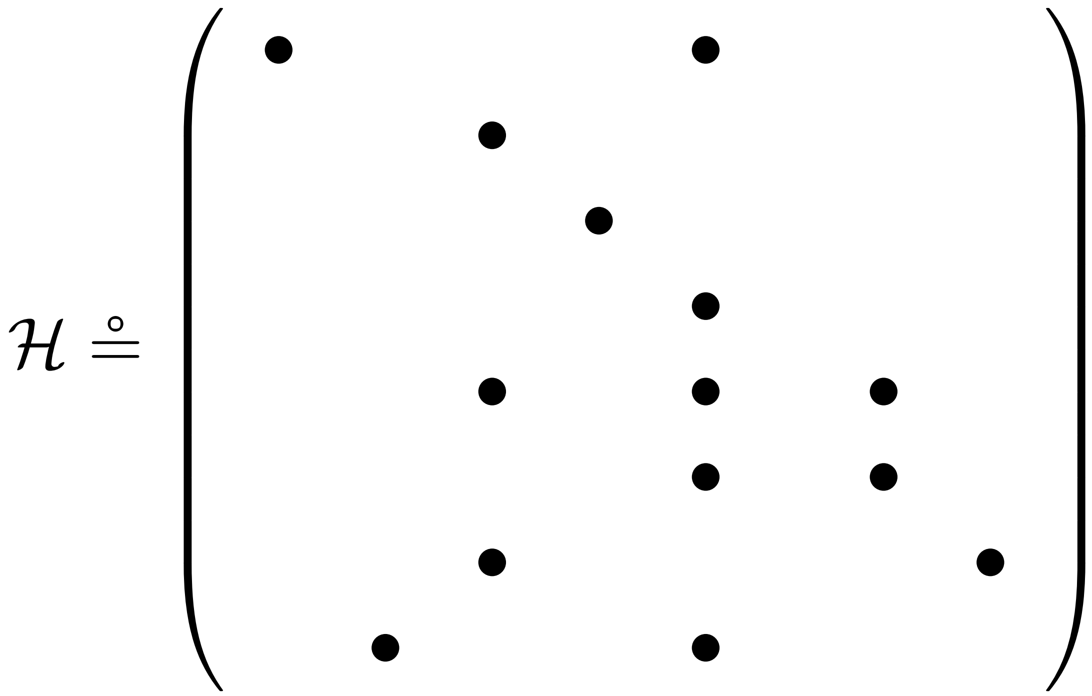
01.4 - Time-Independent Schrödinger Equation

$$\mathcal{H}|\Psi_n\rangle = E_n|\Psi_n\rangle$$

Eigenvalue Problem
for given Hamiltonian

Hamiltonian

Eigenstates

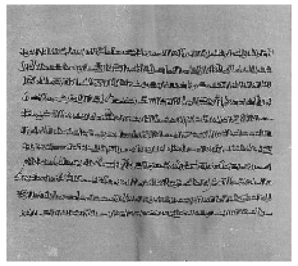


“Row-Sparse”
Matrix for
Physical
Interactions

01.5 - Exact Solutions Limited to Small Systems

[3000 BCE]

Papyrus



10 Qubits

[1455]

Book


ARTAMENE
OV
LE GRAND
CYRVS
DEDIE'
A MADAME LA DVCHESSE
DE LONGVEVILLE.
TROISIESME PARTIE.
A PARIS,
M. DC. LIV.

15 Qubits

[1973]

IBM 3340

23 Qubits

[1993]

IBM 3390

35 Qubits

[2002]

Earth Simulator

46 Qubits

[2019]

Summit

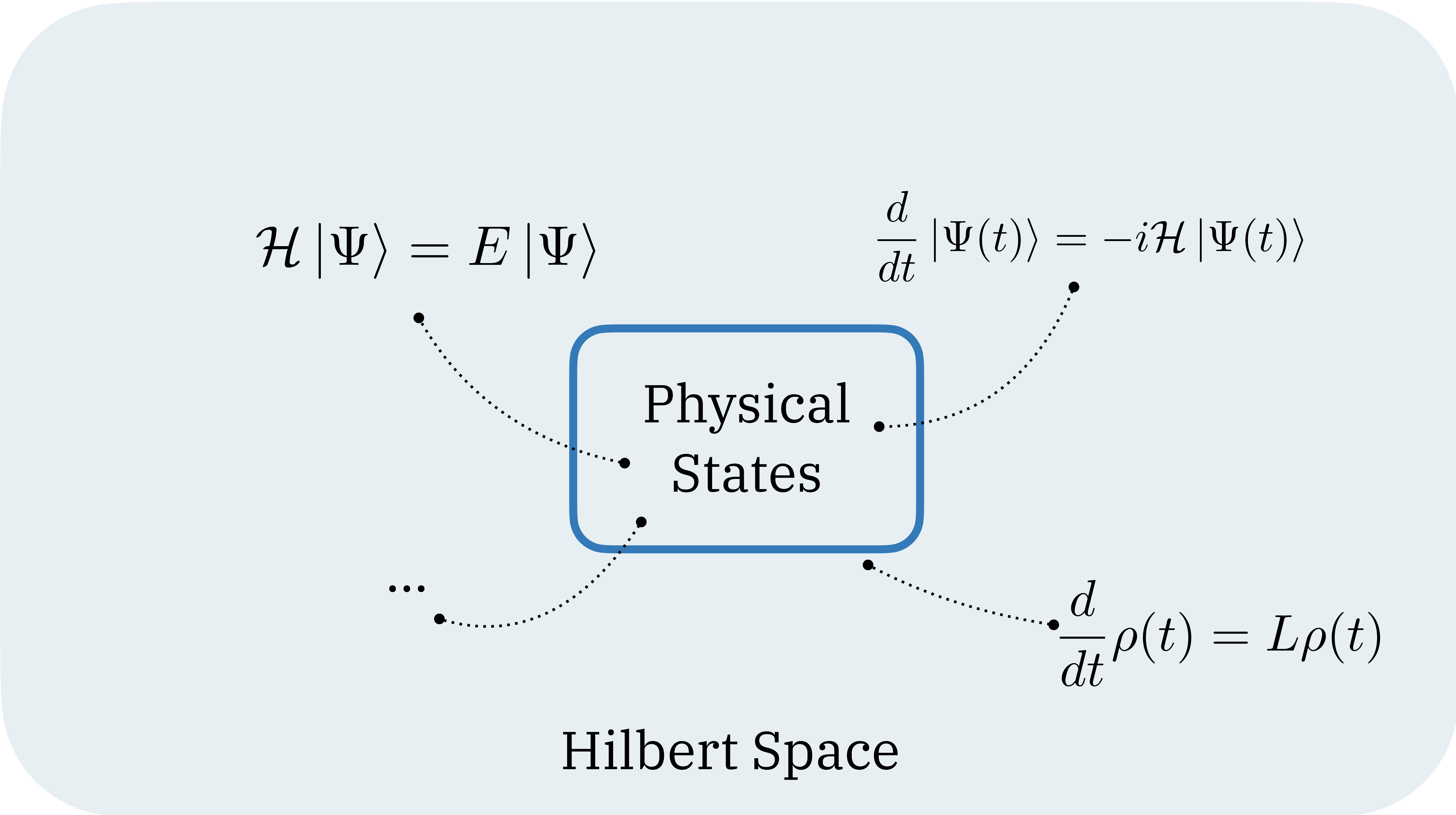
54 Qubits

Time →

02.

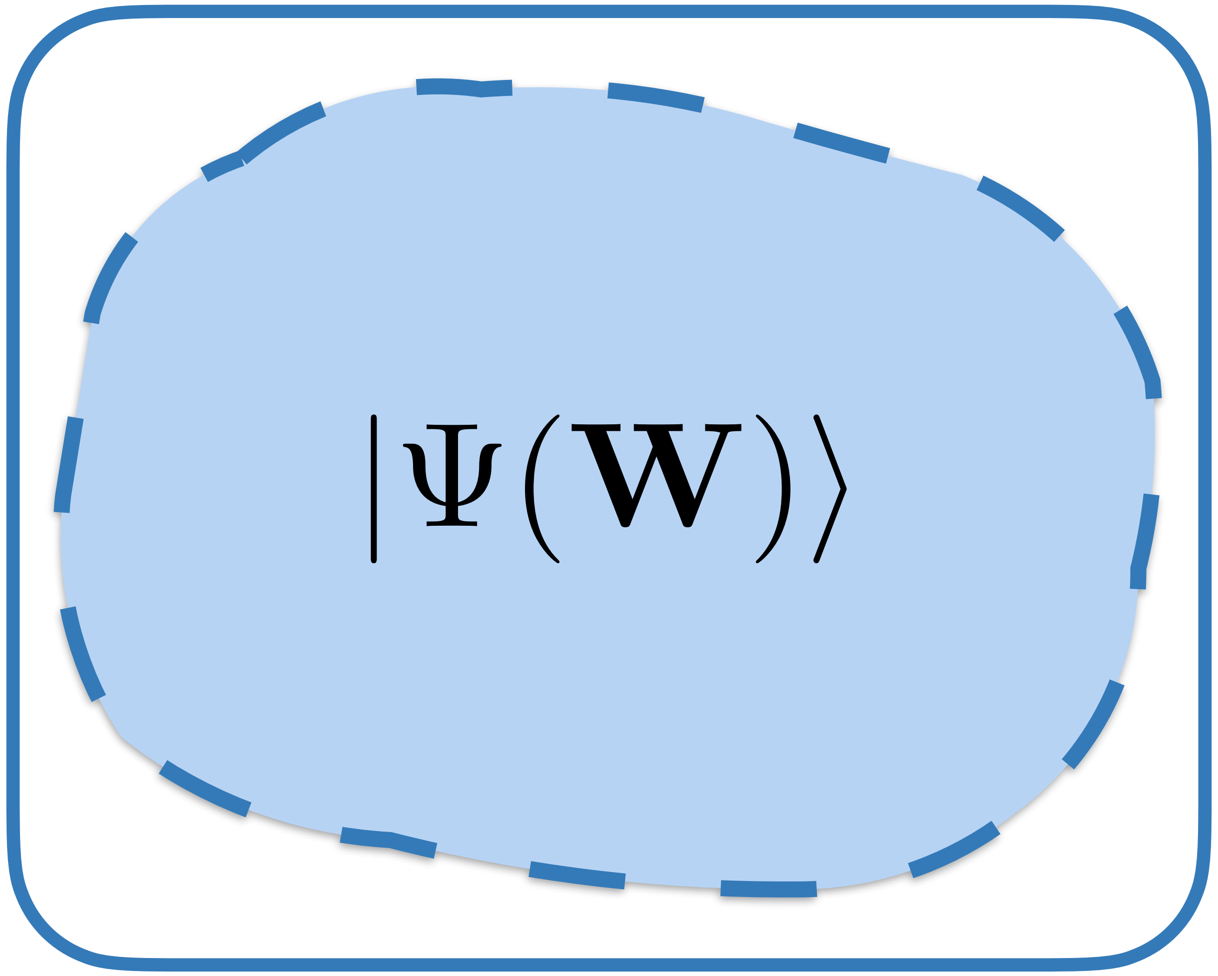
Variational Representations.

02.1 - Corners of the Hilbert space



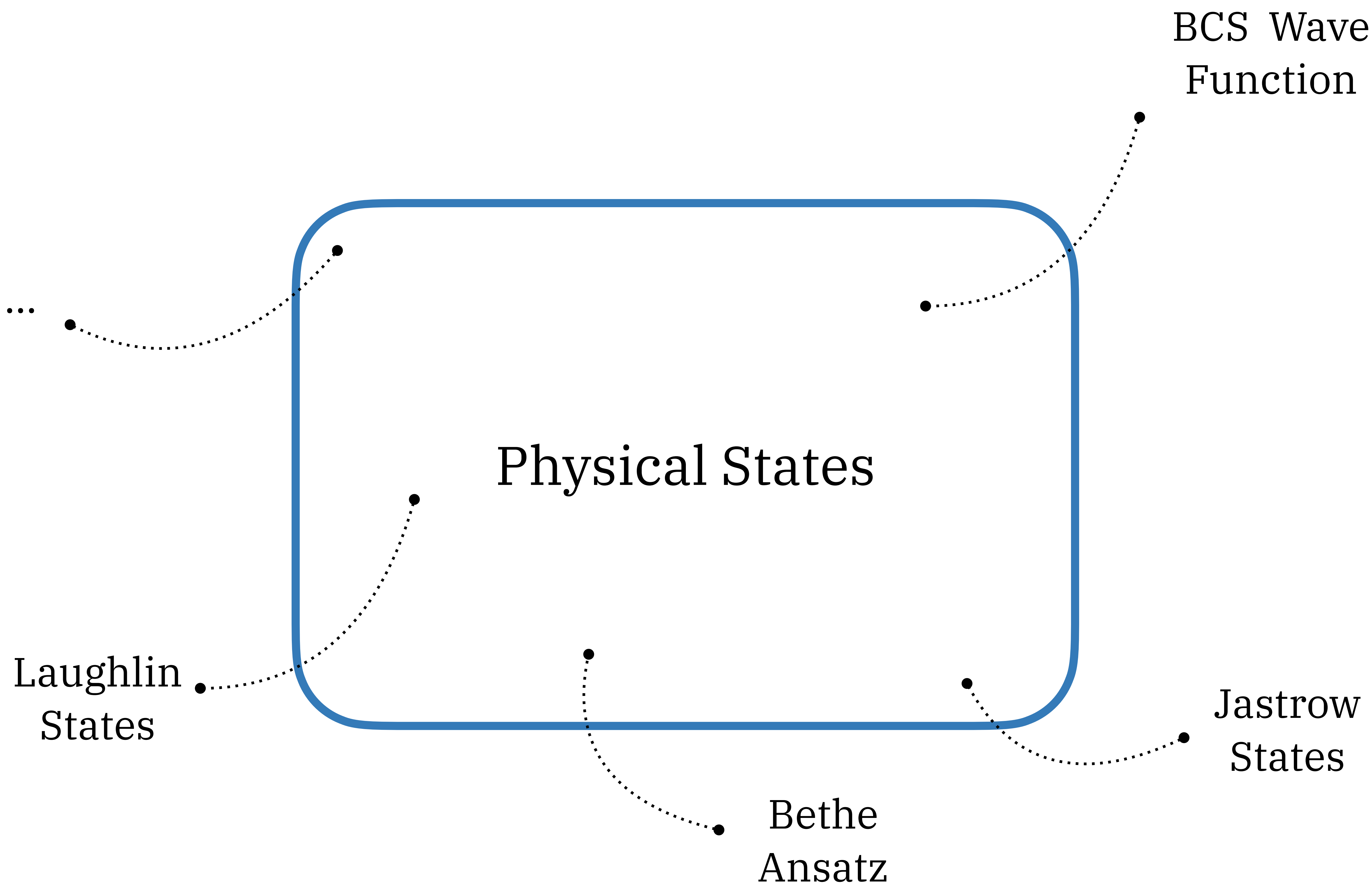
O2.2 - Variational Representations

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



$$\langle Z_1 Z_2 \ldots Z_n|\Psi(W)\rangle = \Psi(Z_1, Z_2 \ldots Z_N; W) = c_{Z_1,Z_2,\ldots Z_N}(W)$$

02.3 - Physics-Inspired Representations

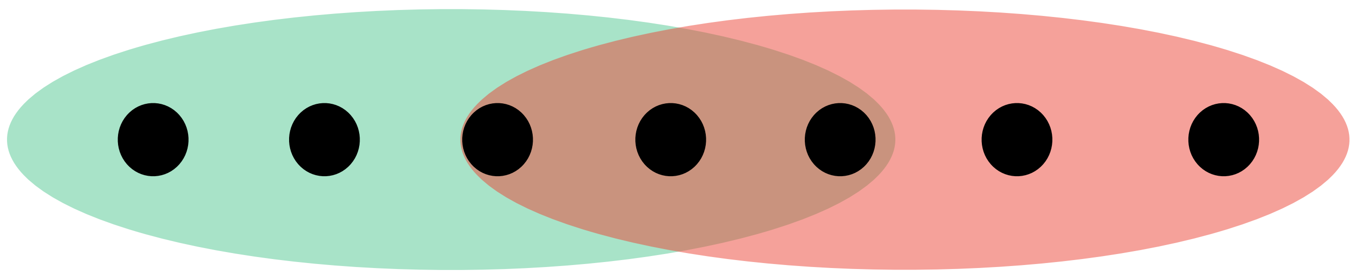


O2.4 - General Purpose: Matrix Product States

$$\langle Z_1 Z_2 \ldots Z_n | \Psi(W) \rangle = \mathrm{Tr}\left[M(Z_1; W) M(Z_2; W) \ldots M(Z_n; W) \right]$$

Matrices
DxD

S. White
Phys. Rev. Lett. 69, 2863 (1992)



Simple Algebra

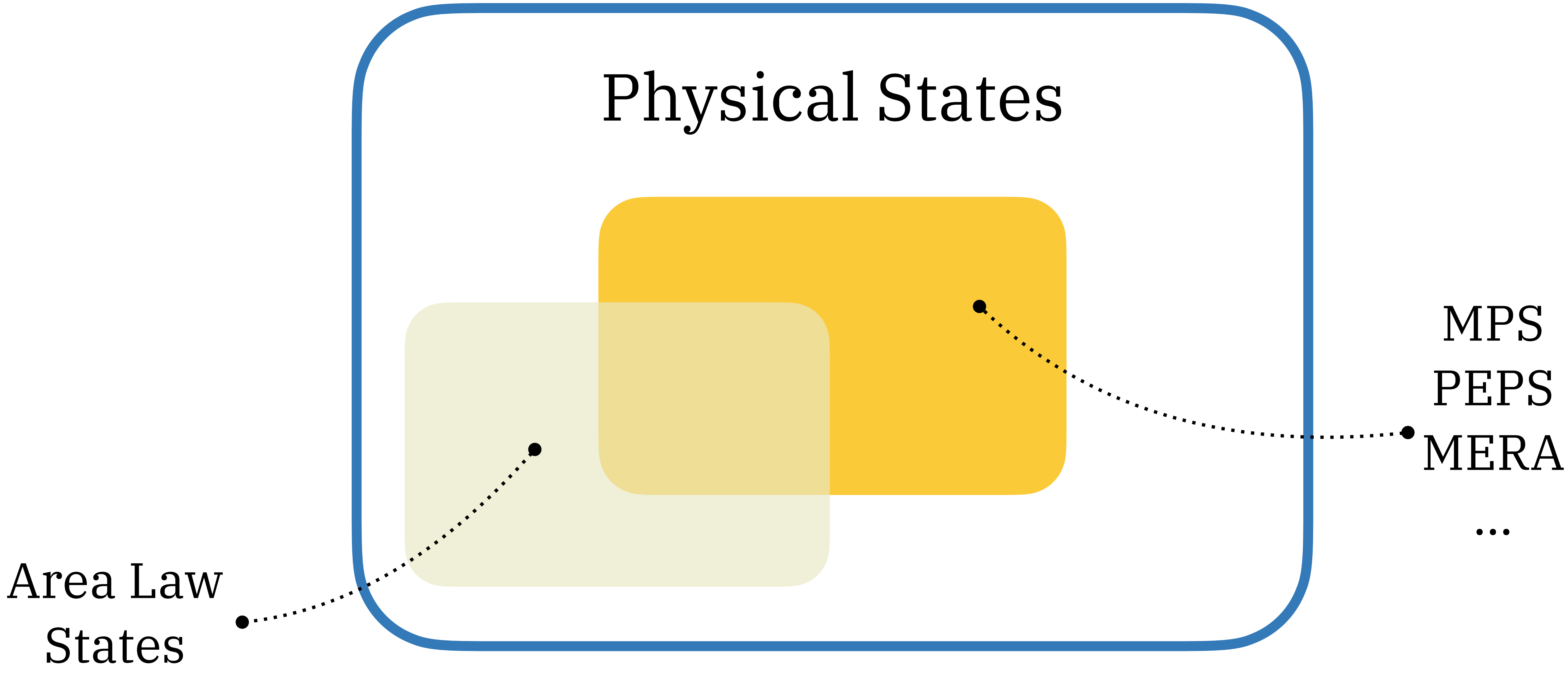
Efficient Compression
of Wave-Function

“Polynomial”
complexity

Low Entanglement

Many-Body State
Specified by Small Set
of Local Quantities

02.5 - Tensor Networks Representations

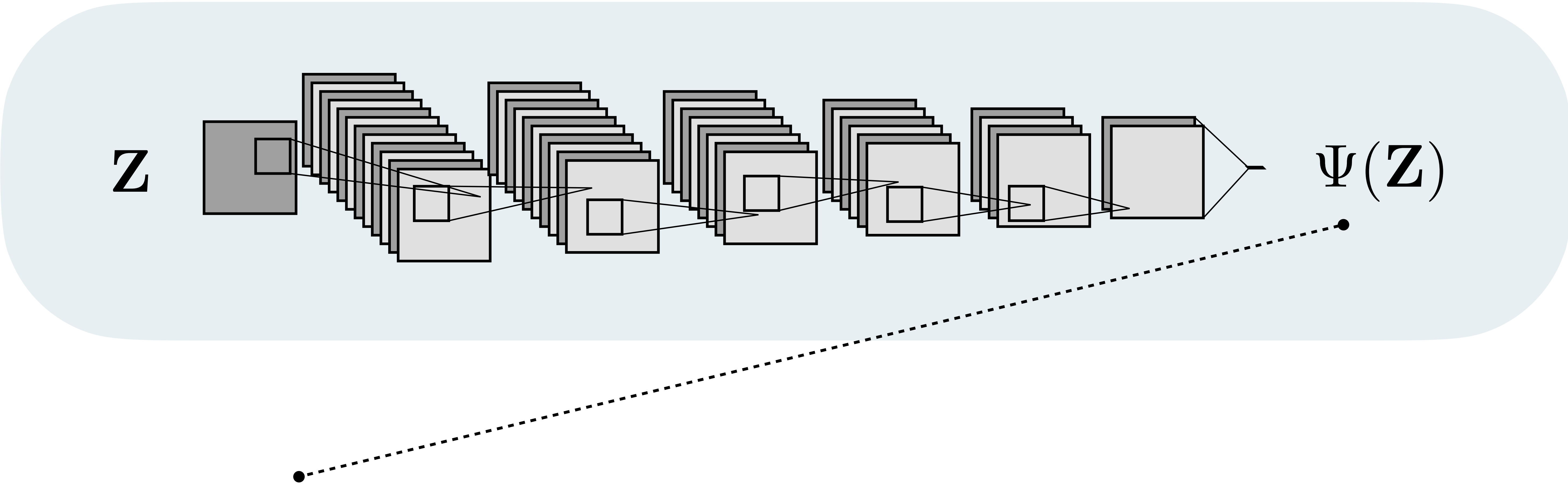


03.

Neural-Network Representations.

03.1 - Neural Quantum States

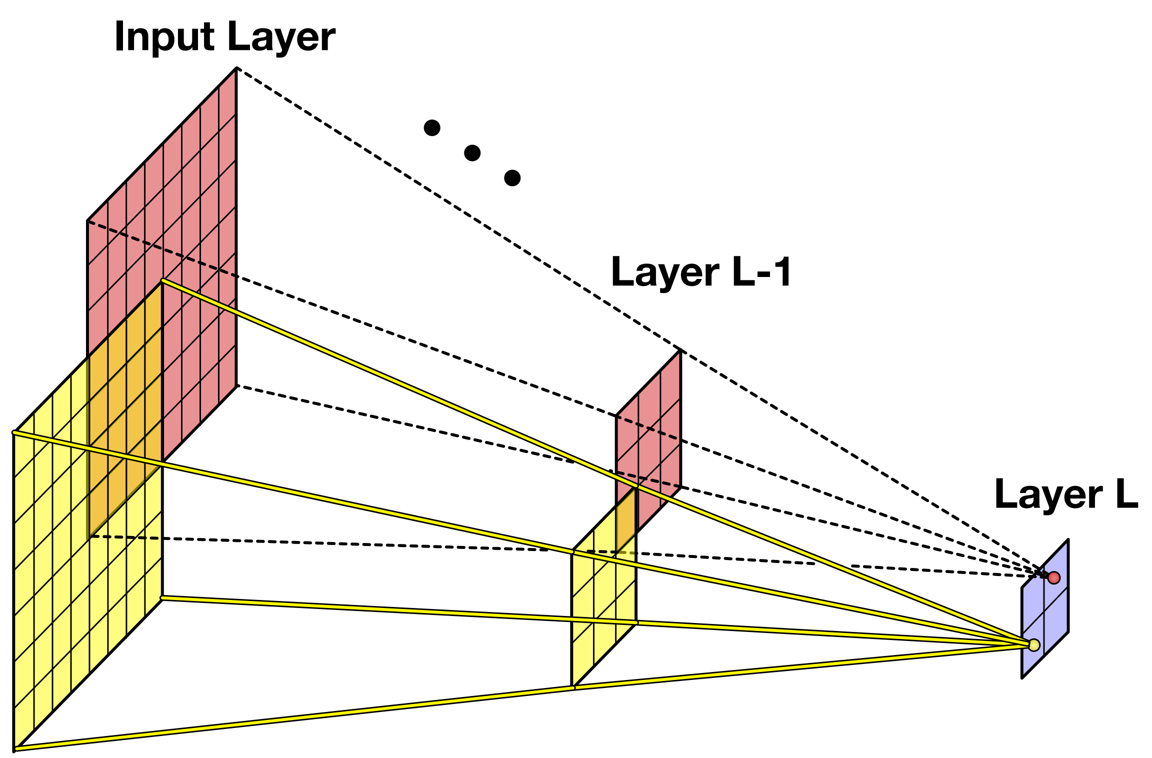
Carleo, and Troyer
Science 355, 602 (2017)



$$\langle Z_1 Z_2 \ldots Z_N | \Psi \rangle = g^{(L)} \circ W^{(L)} \ldots g^{(2)} \circ W^{(2)} g^{(1)} \circ W^{(1)} \mathbf{Z}$$

03.2 - Representation and Entanglement Properties

$$\langle \mathbf{Z} | \Psi \rangle = \sum_{q=0}^{2n} \Phi_q \left(\sum_{p=1}^{n} \phi_{q,p}(Z_p) \right)$$



Universal Approximation Theorems

Kolmogorov and Arnold (1956)

Cybenko (1989)

Volume-Law States

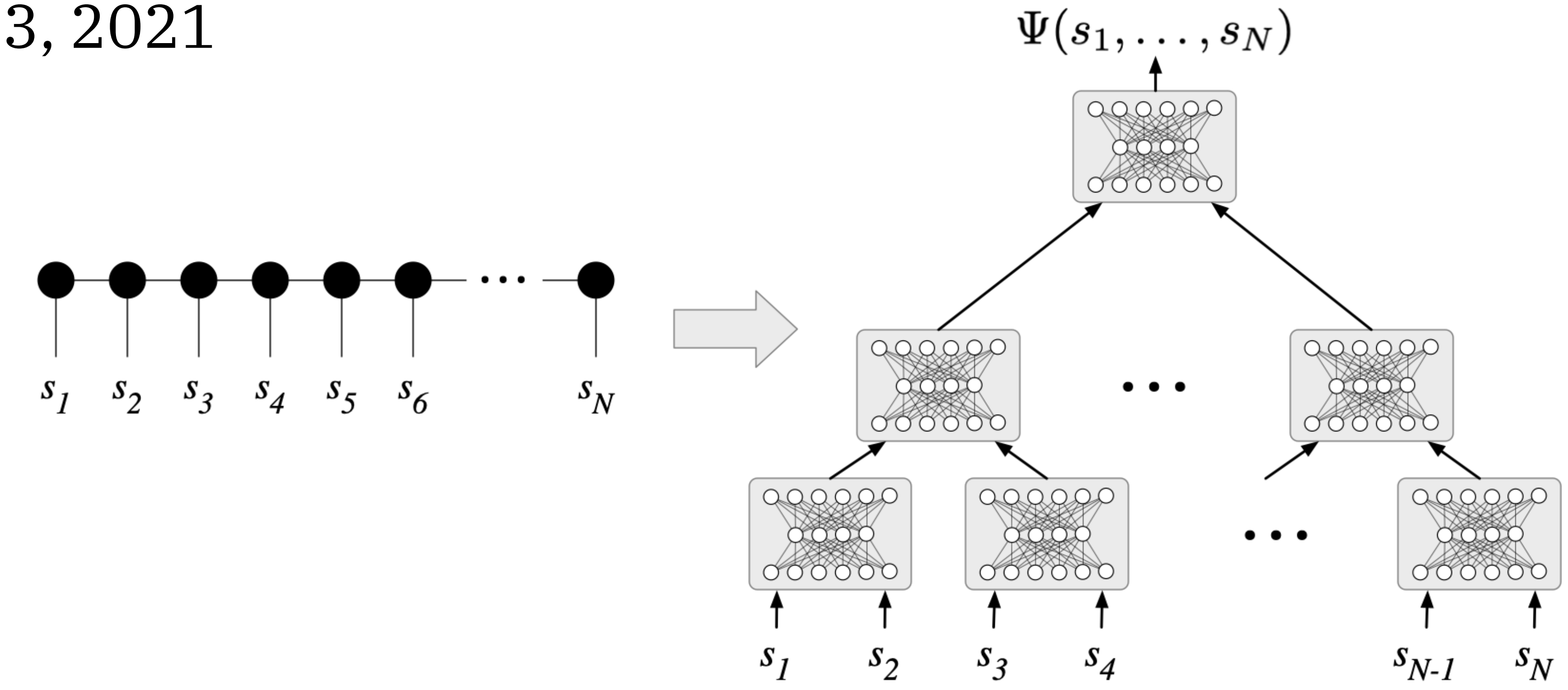
Deng, Li, and Das Sarma PRX *7, 021021 (2017)*

....

Levine, Sharir, Cohen, and Shashua PRL *122, 065301 (2019)*

03.3 - Neural-Tensor Contractions

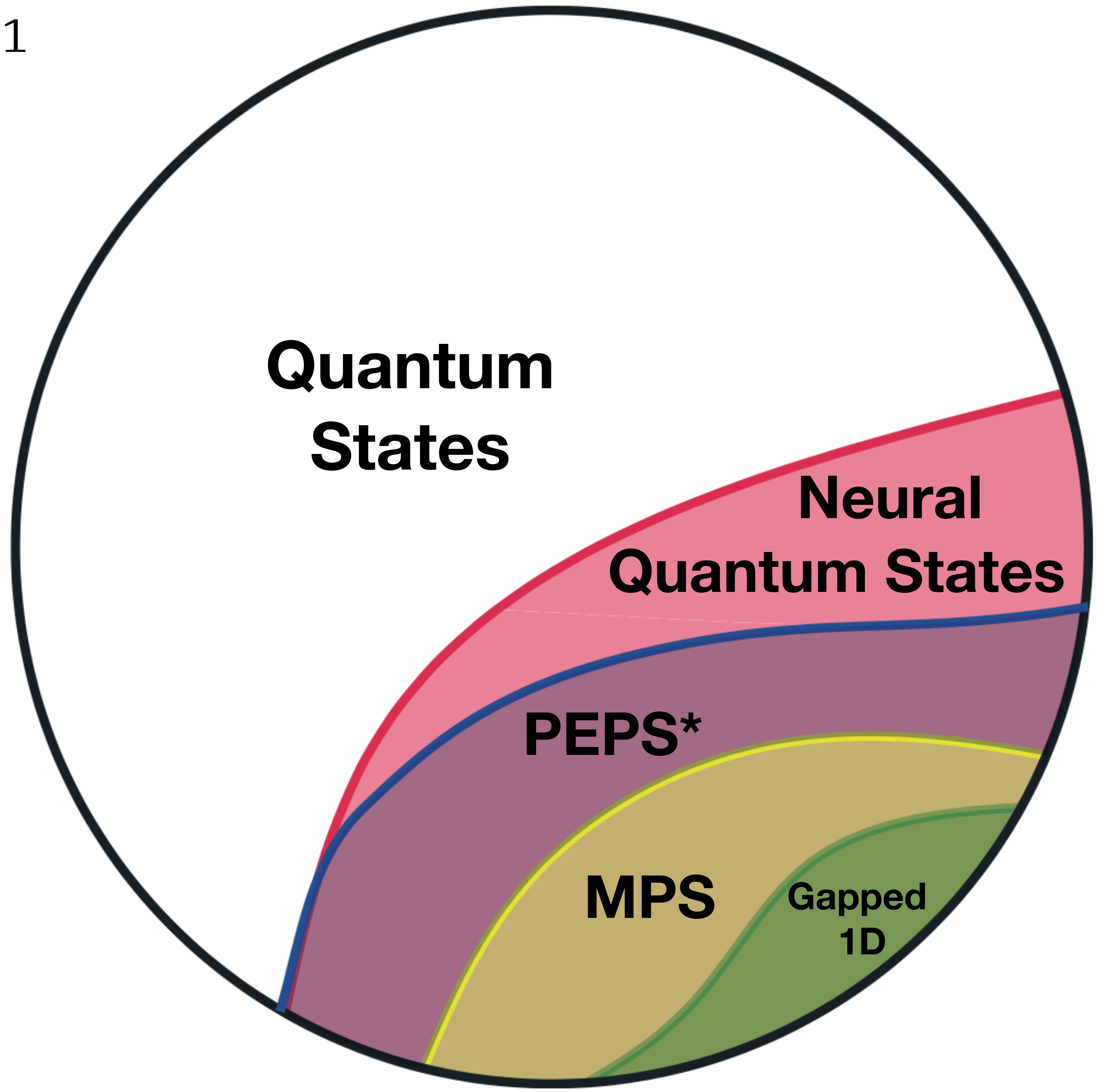
Sharir, Shashua, and Carleo
arXiv:2103.10293, 2021



Corollary 1 *For any tensor network quantum state with a contraction scheme of run-time k, and at most b bits of precision in computations and parameters, there exists a neural network that approximate it with a maximal error of ϵ and of run-time (number of edges)* $O\left(k + \ln^2\left(\frac{kb}{\epsilon}\right) + \ln\left(\frac{1}{\epsilon}\right)\sqrt{\frac{1}{\epsilon}}\right)$.

03.4 - Representability Diagram

Sharir, Shashua, and Carleo
arXiv:2103.10293, 2021



04.

Learning the Ground State.

04.1 - 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

Expectation Minimization

$$E(\mathbf{W}) = \frac{\sum_Z |\Psi(Z;W)|^2 E_{\mathrm{loc}}(Z;W)}{\sum_Z |\Psi(Z;W)|^2}$$

McMillan, Phys. Rev. 138, *A442 (1965)*

04.2 - Energy Gradients

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

$$E_{\text{loc}}(Z;W) = \sum_{Z'} \frac{\Psi(Z';W)}{\Psi(Z;W)}\langle Z|\mathcal{H}|Z'\rangle$$

$$\mathcal{O}_k(Z;W) = \frac{1}{\Psi(Z;W)}\frac{\partial\Psi(Z;W)}{\partial W_k}$$

$$\langle F\rangle = \frac{\sum_Z |\Psi(Z;W)|^2 F(Z)}{\sum_Z |\Psi(Z;W)|^2}$$

04.3 - Natural Gradients

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

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

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

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

Quantum Geometric Tensor or Quantum Fisher Information

Equivalent to Imaginary-Time Evolution (Power Method)in Variational Manifold

04.4 - Variational Learning Algorithm

1. Sample $Z^{(1)} \dots Z^{(M)}$ from $P(Z;W) = \frac{|\Psi(Z;W)|^2}{\sum_{Z'} |\Psi(Z';W)|^2}$

2. Estimate Expectation Values and Gradient $\langle F \rangle \simeq \frac{1}{M} \sum_{i}^{M} F(Z^{(i)})$

3. Estimate Quantum Fisher

4. Update Parameters $W' = W - \eta G$

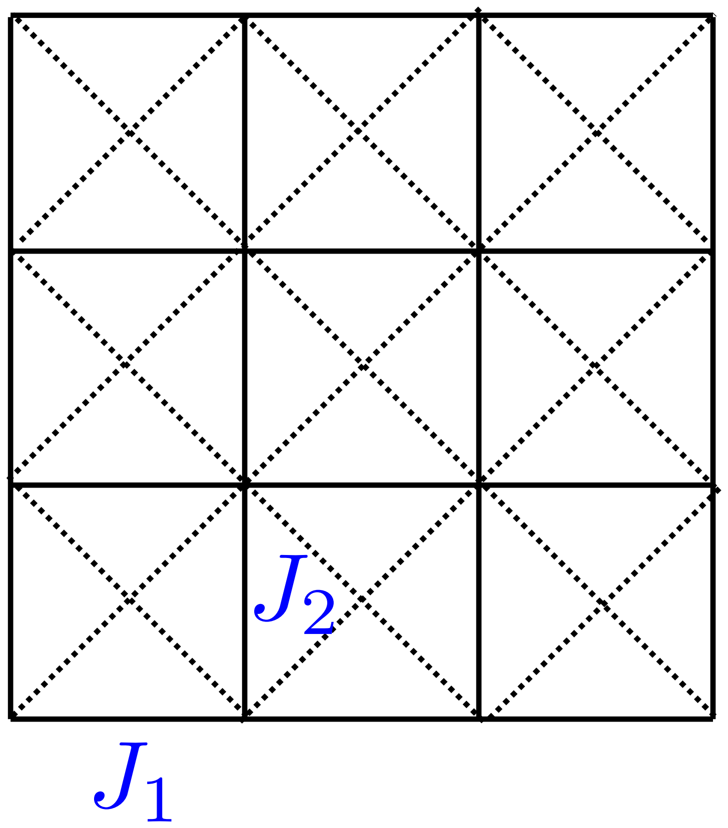
05.

Example Applications.

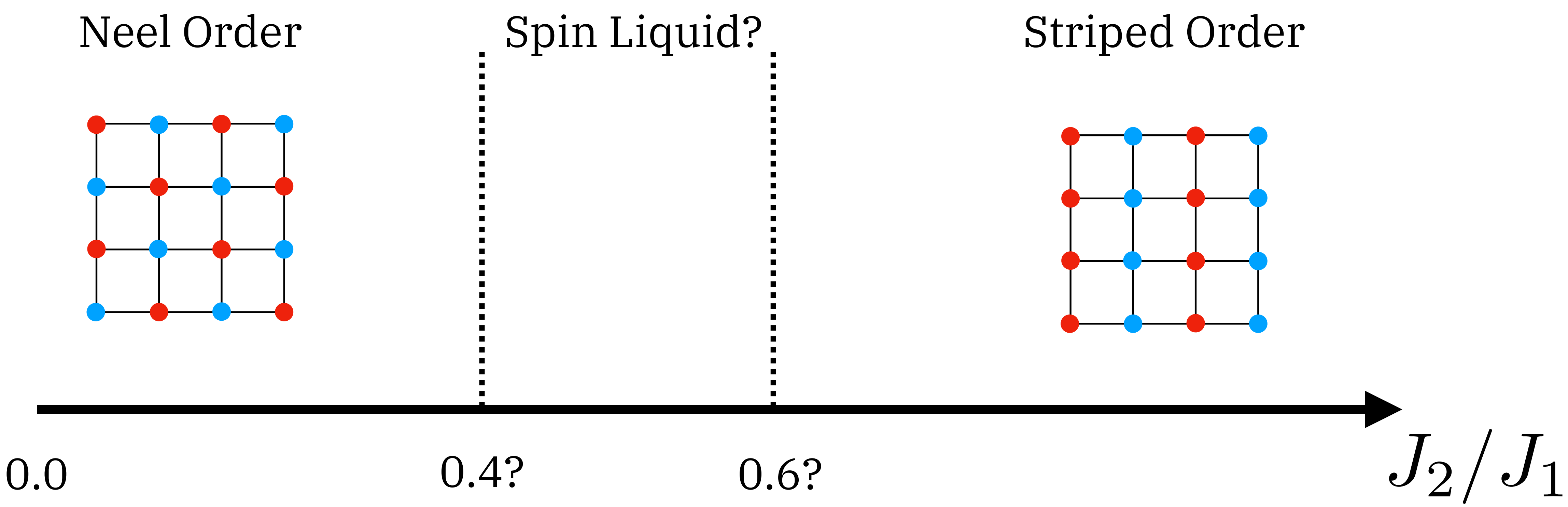
O5.1 - Frustrated Spins

J1-J2 Model

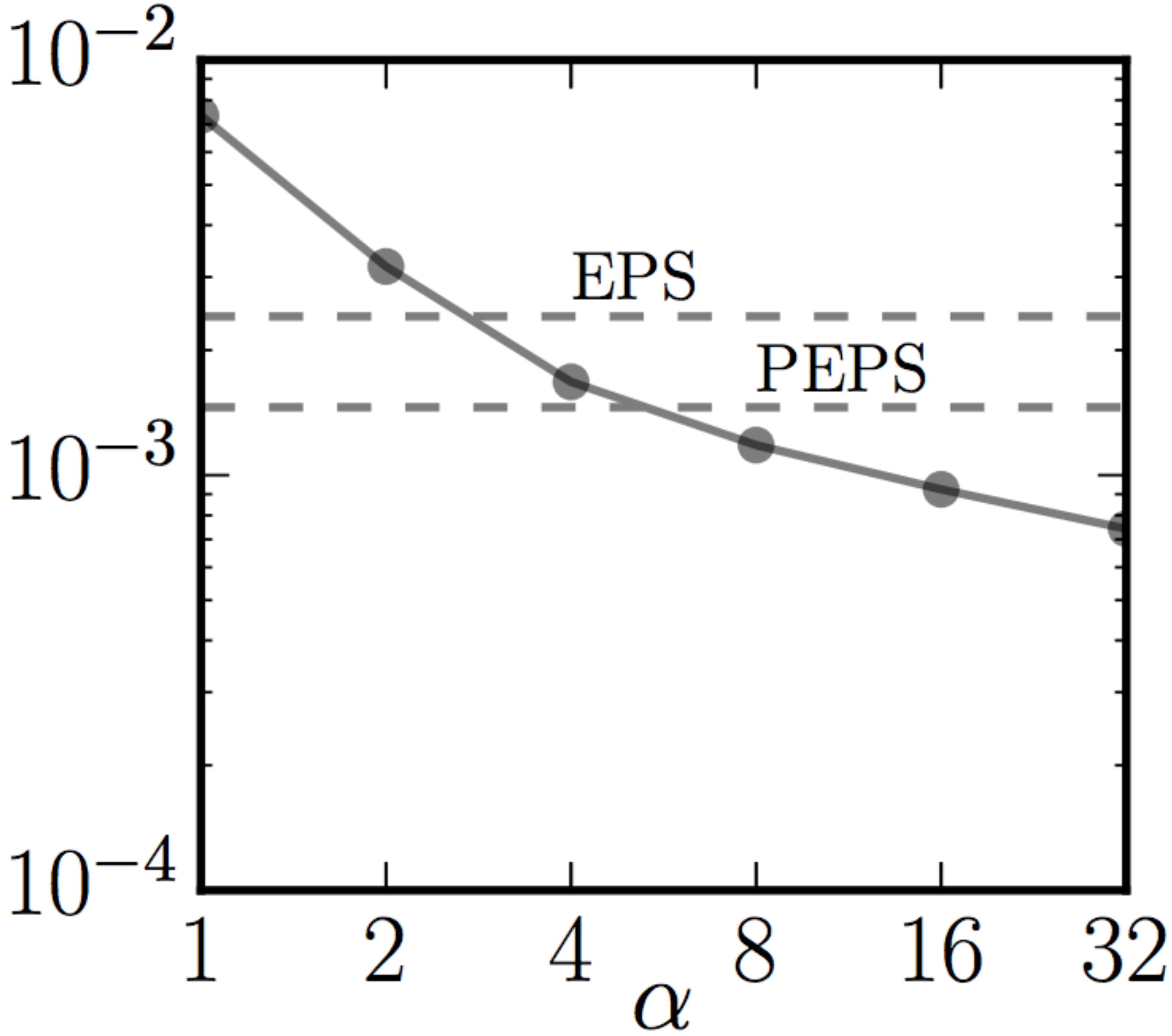
$$\hat{H} = J_1 \sum_{\langle ij \rangle} \hat{\boldsymbol{S}}_i \cdot \hat{\boldsymbol{S}}_j + J_2 \sum_{\langle\langle ij \rangle\rangle} \hat{\boldsymbol{S}}_i \cdot \hat{\boldsymbol{S}}_j$$



Phase Diagram



05.2 - Heisenberg Limit - Shallow Net

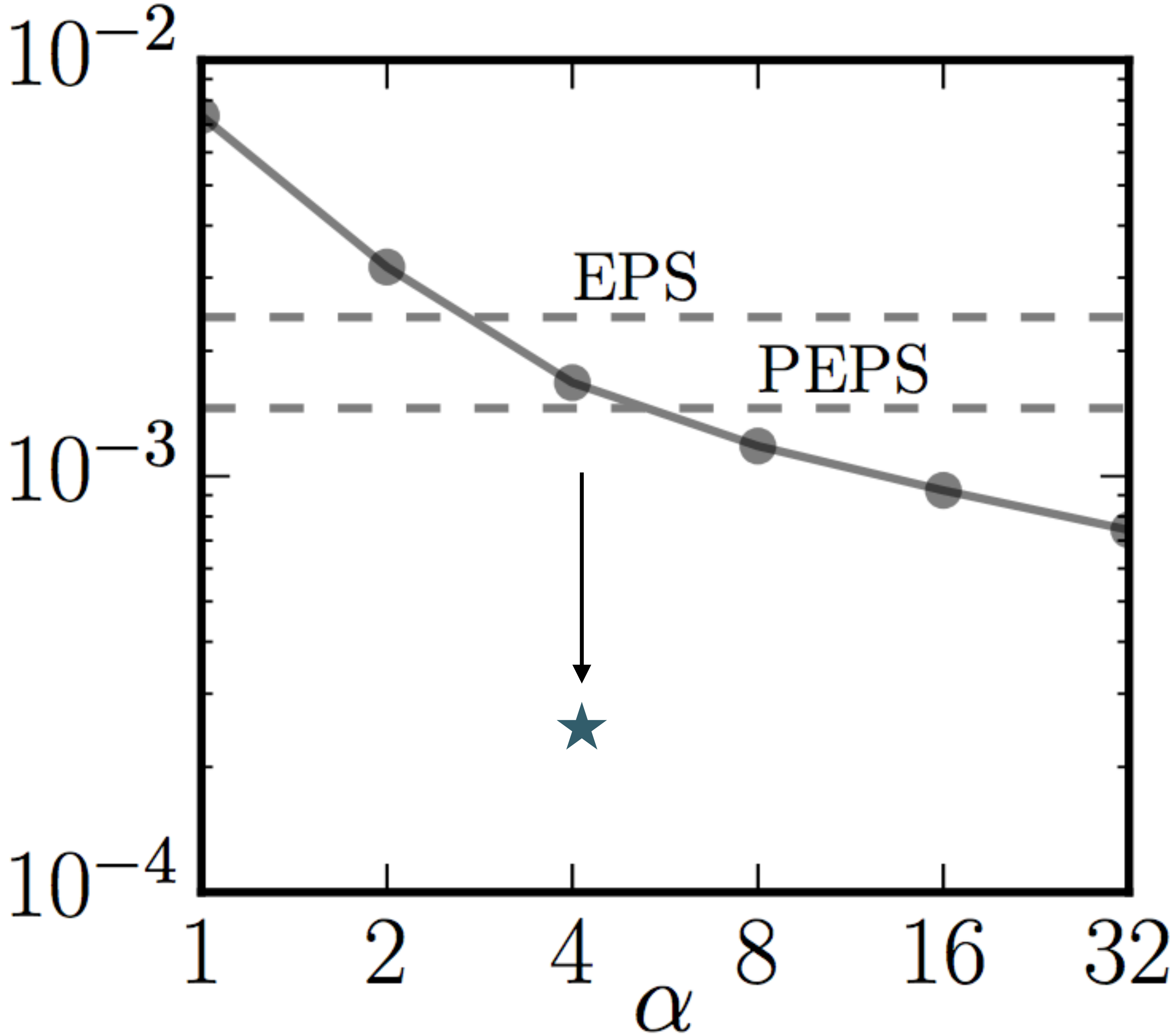


10 by 10 cluster

Early (2016) Results With Shallow (RBM) Network

Carleo, and Troyer
Science 355, 602 (2017)

05.3 - Heisenberg Limit - Deeper Net

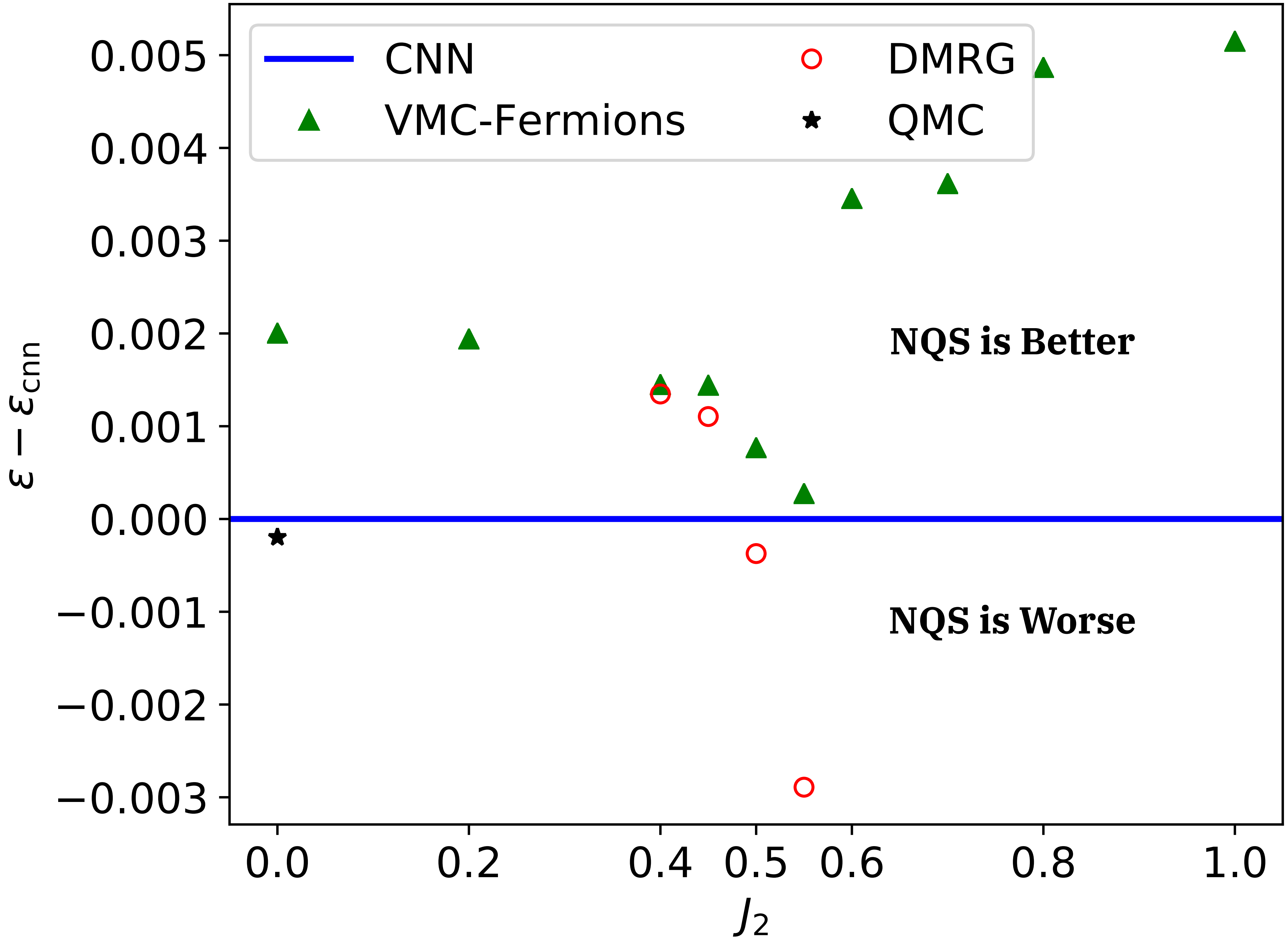


10 by 10 cluster

(Mildly) deep CNN further improves

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

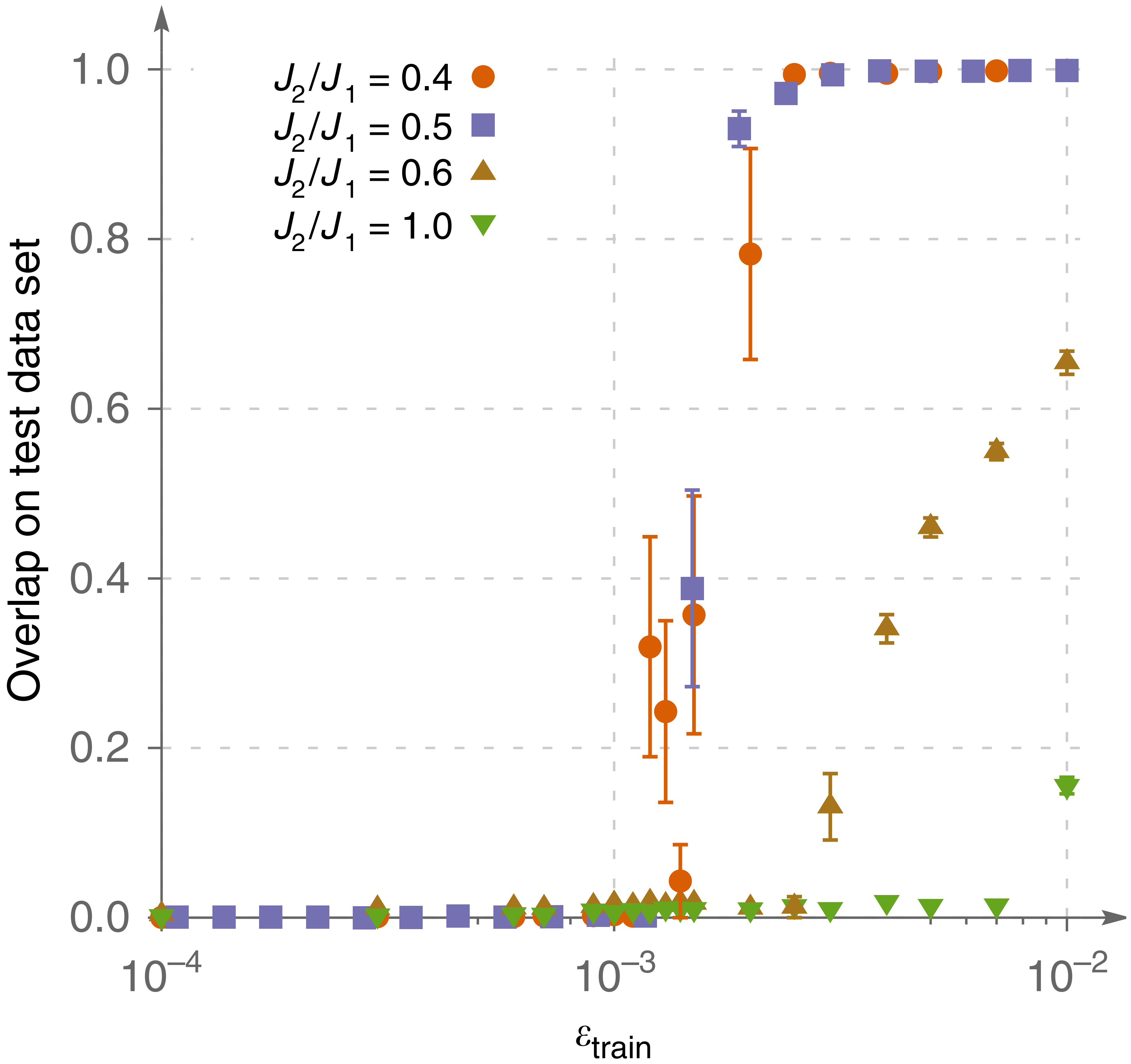
05.4 - Frustrated Case: Accuracy Diagram



10 by 10 cluster

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

O5.5 - Origin of the Challenge?



Frustrated Phases Have Large Sample Complexity

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

05.6 - Continuous Improvements...

TABLE II. Comparison of ground-state energy for the 10×10 lattice at $J_2 = 0.5$ among different wave functions. The wave functions in bold font use neural networks. In Ref. [18], p-th order Lanczos steps are applied to the VMC wave function.

Energy per site	Wave function	Reference
−0.494757(12)	**Neural quantum state**	65
−0.49516(1)	**CNN**	60
−0.49521(1)	VMC(p=0)	18
−0.495530	DMRG	22
−0.49575(3)	**RBM-fermionic w.f.**	63
−0.497549(2)	VMC(p=2)	18
−0.497629(1)	**RBM+PP**	present study

Nomura, and Imada
arXiv:2005.14142 (2020)

05.7 - Fermions: Back to the Spin Problem

Map Fermions to Spins

Choo, Mezzacapo, and Carleo
Nature Comm. 11, 2368 (2020)

Jordan-Wigner Mapping

Pro: Simple Mapping

Con: N-Body, non-local Spin Operators

Bravyi-Kitaev Mapping

Pro: log(N)-Body, quasi-local Spin Operators

Con: More Involved Mapping

O5.8 - Jordan-Wigner Mapping

$$c_j \to \left(\prod_{i=0}^{j-1} \sigma_i^z\right) \sigma_j^-$$

$$c_j^\dagger \to \left(\prod_{i=0}^{j-1} \sigma_i^z\right) \sigma_j^+$$

Jordan Wigner “strings” take into account exchange symmetry

$$H_q = \sum_{j=1}^{r} h_j \boldsymbol{\sigma}_j$$

Spin Hamiltonian is a sum of product of Pauli matrices

05.9 - Bravyi-Kitaev Mapping

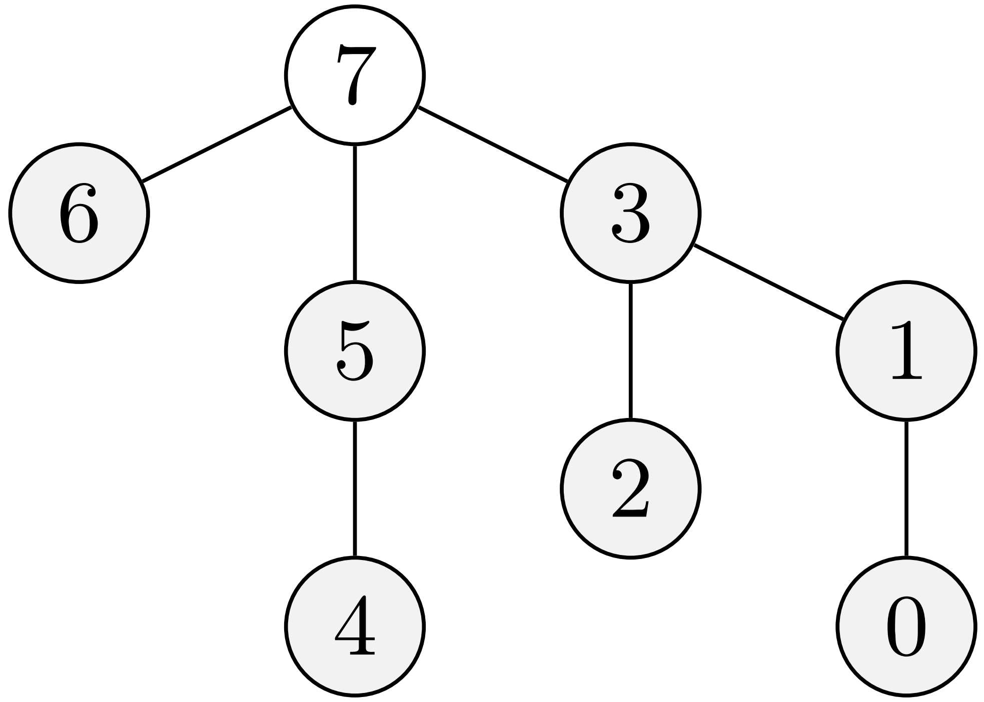
$$c_j \to \frac{1}{2}\left(A_j \sigma_j^x B_j + i A_j \sigma_j^x B_j\right)$$

$$c_j^\dagger \to \frac{1}{2}\left(A_j \sigma_j^x B_j - i A_j \sigma_j^x B_j\right)$$

$$A_j = \left(\Pi_{k \in U(j)} \sigma_k^x\right)$$

$$B_j = \left(\Pi_{k \in P(j)} \sigma_k^z\right)$$

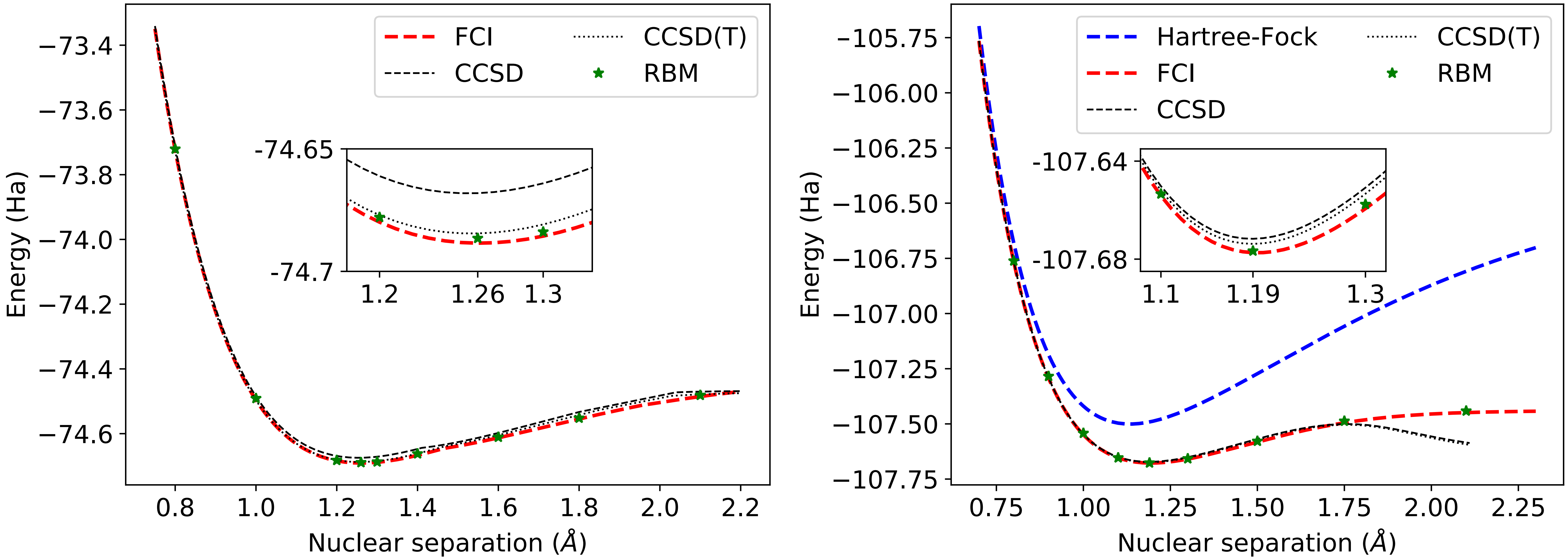
"Quasi-Local"
Operators



Compute signs using Fenwick trees instead of linear products

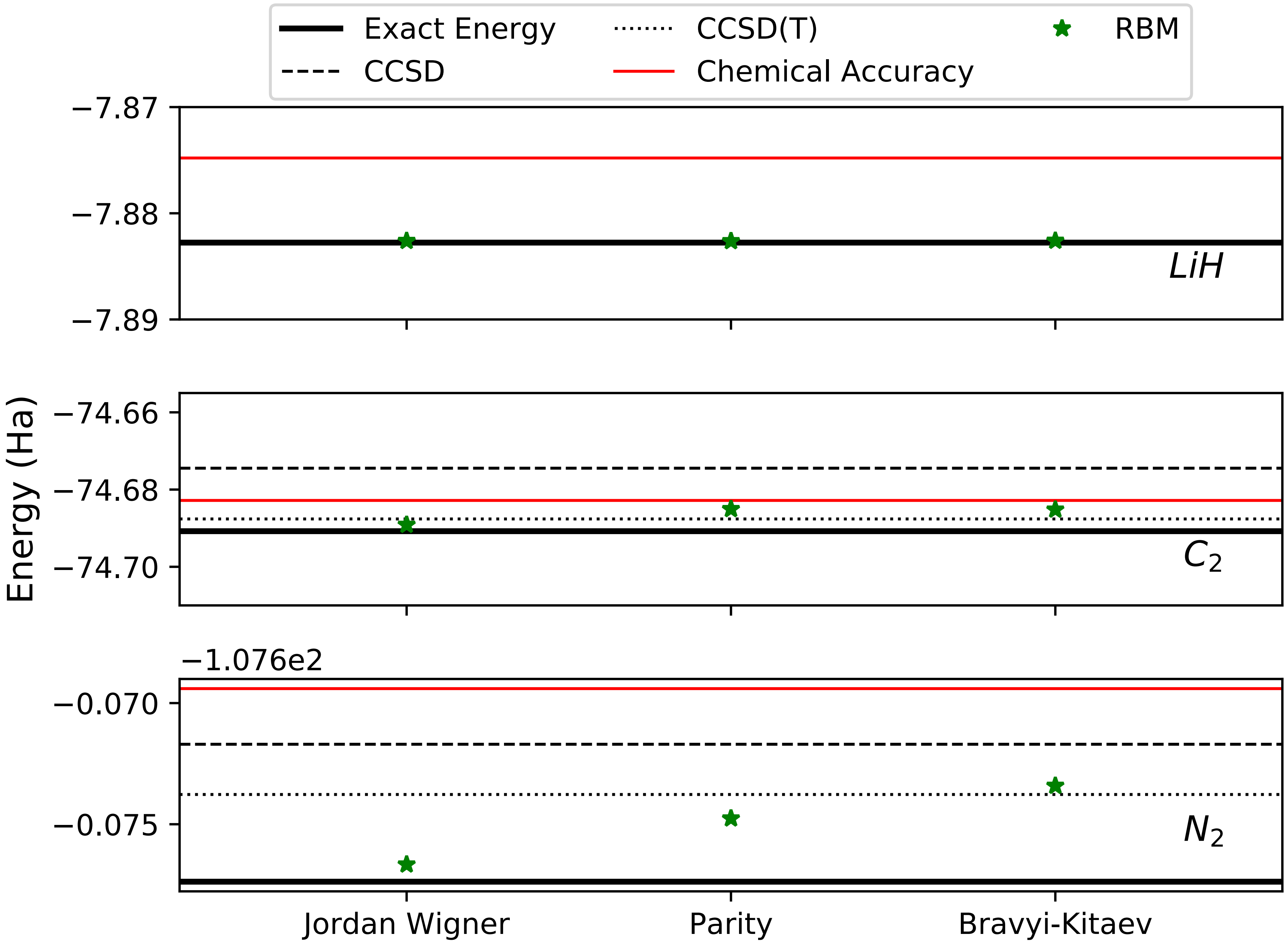
Strings have log(N) length instead of N

O5.10 - Dissociation Curves for C_2 and N_2



STO-3G Basis Set
Single-Layer Network

05.11 - Different Mappings



Ansatz is almost insensitive to the locality of the mapping

06.

Computationally Tractable States.

06.1 - Definition and Properties

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.*

Van Den Nest
arXiv:0911.1624 (2009)

06.2 - Examples

Matrix Product States Are Computationally Tractable

Jastrow, Backflow, PEPS etc states are not computationally tractable

Generic neural deep quantum states are not computationally Tractable

06.3 - Autoregressive Quantum States

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

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

Normalized
"Conditionals"

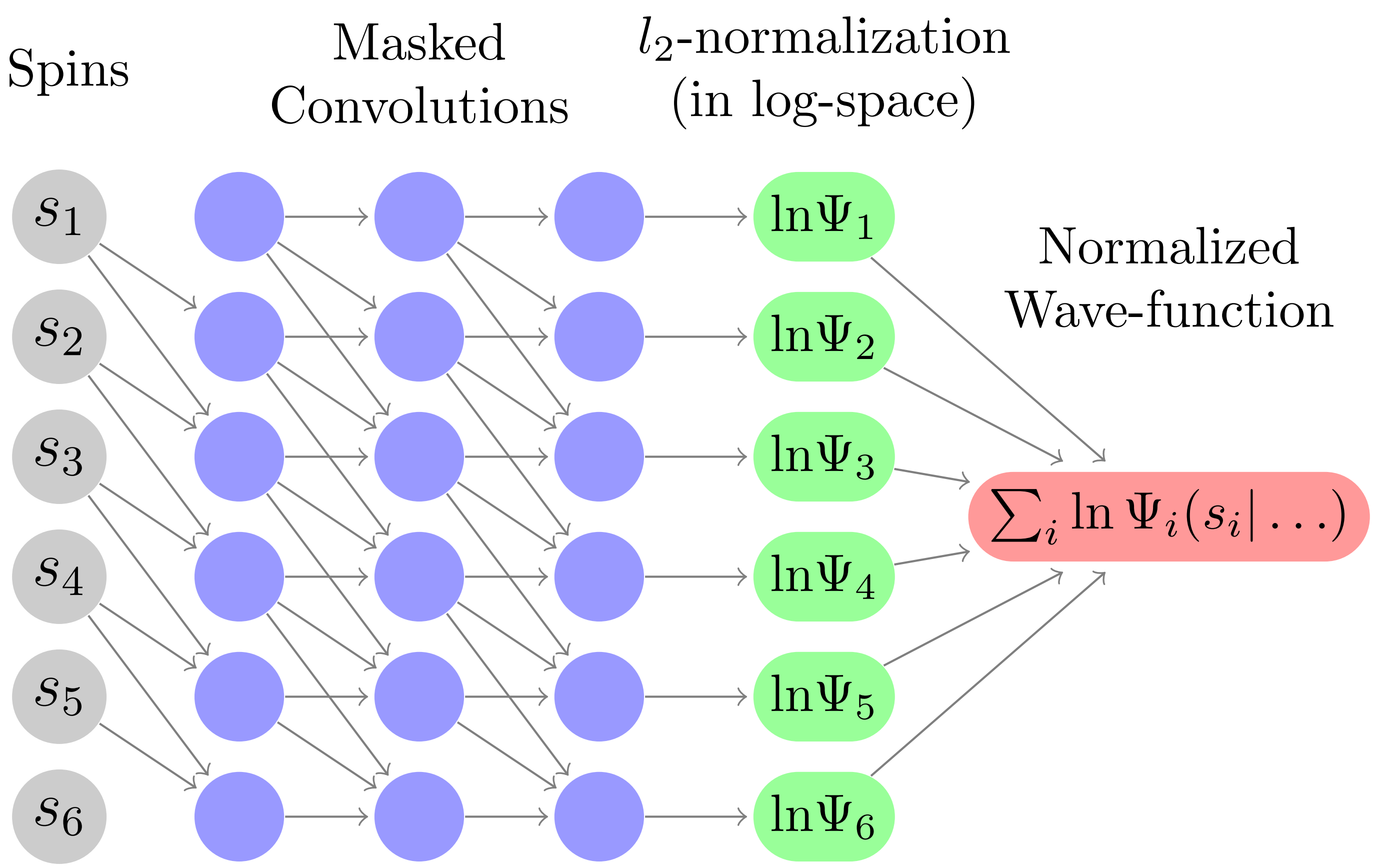
Sharir, Levine, Wies, Carleo, and Shashua
Phys. Rev. Lett. 124, 020503 (2020)

These Are Computationally Tractable

(a) Exact Sampling

(b) Computing Normalized Amplitudes is Efficient

06.4 - Using Masked Deep Networks



Masked Fully Connected Network

1	1	1	1	1
1	1	1	1	1
1	1	0	0	0
0	0	0	0	0
0	0	0	0	0

Masked Convolutions

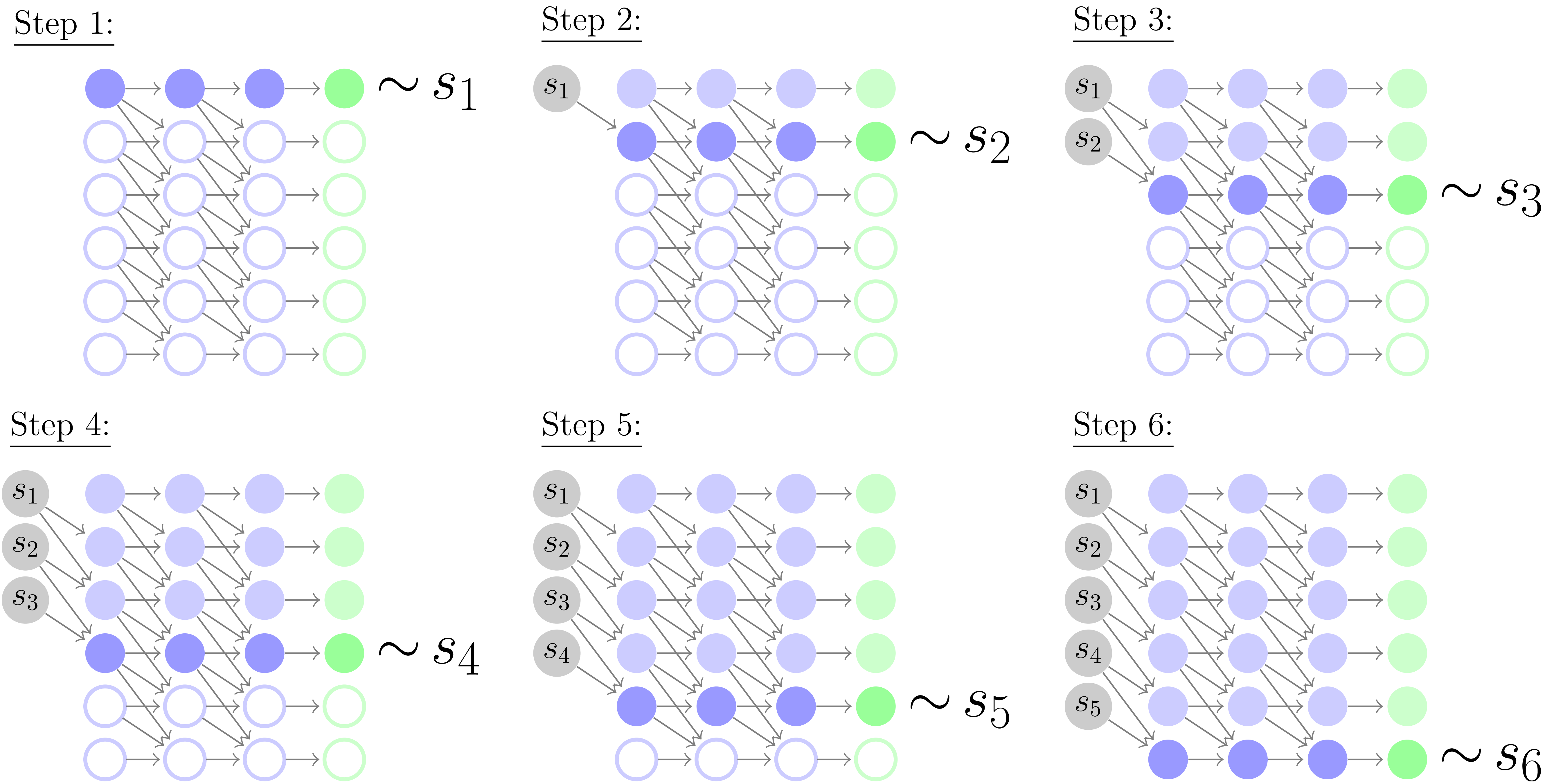
PixelCNN

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

Salimans et al.
arXiv:1701.05517 (2017)

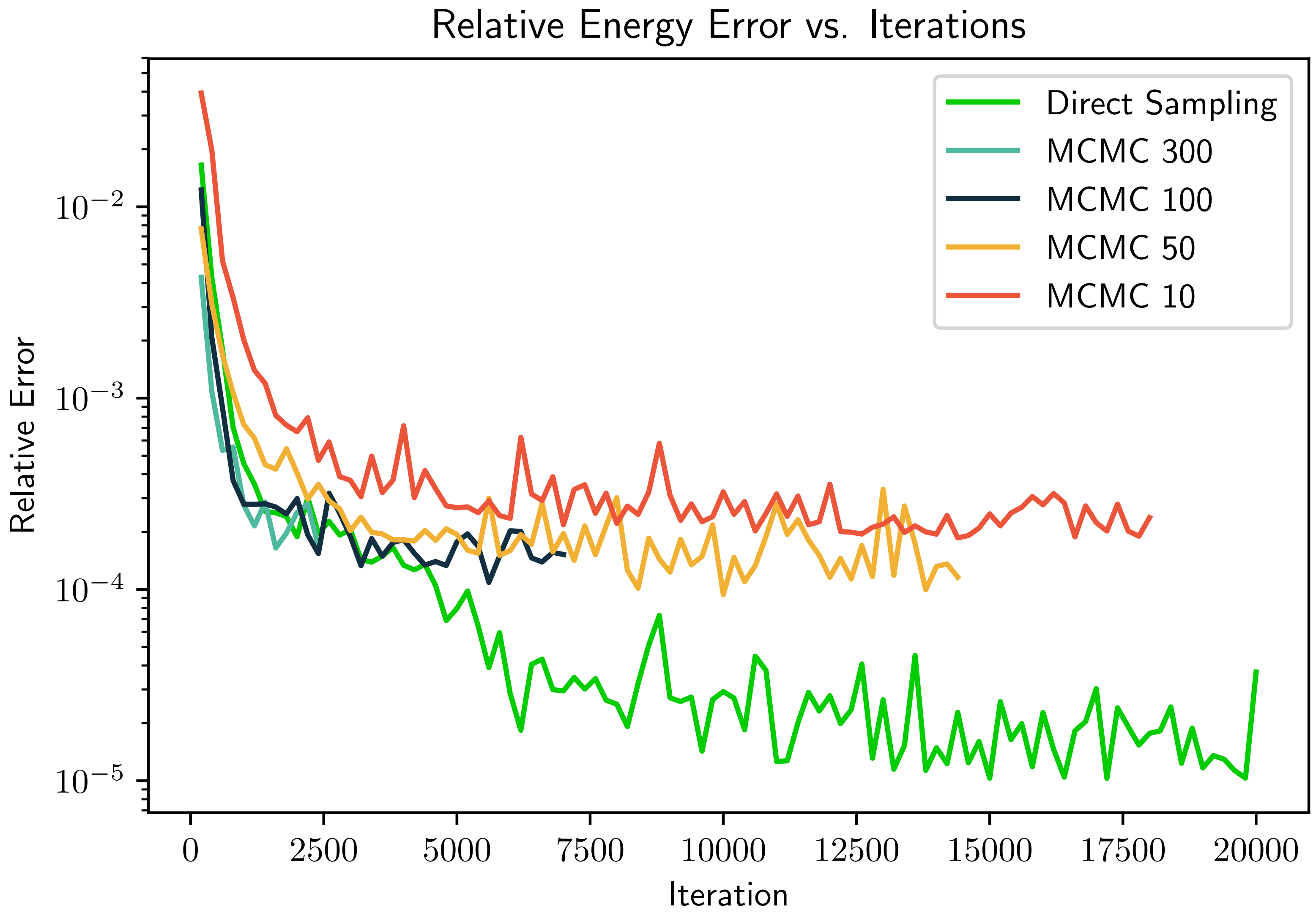
Ramachandran et al.
arXiv:1704.06001 (2017)

06.5 - Exact Sampling



O6.6 - Removing the Sampling Bottleneck Pays Off

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

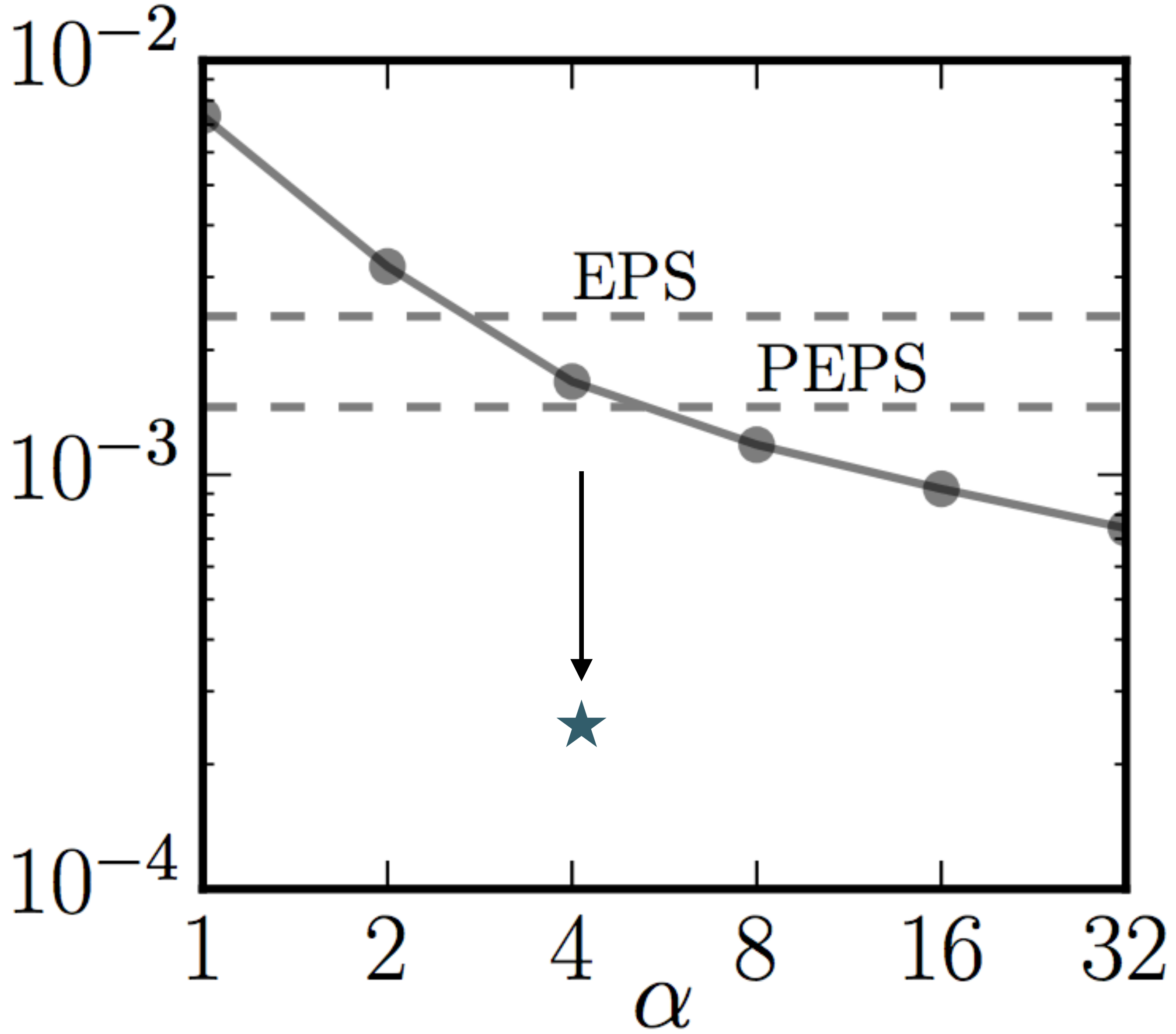


21x21
Transverse-Field
Ising in 2d

Depth 20 About
1 Million
Parameters

06.7 - Accuracy on Spins Only: Heisenberg Model

Carleo, and Troyer
Science 355, 602 (2017)



10 by 10 cluster

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

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

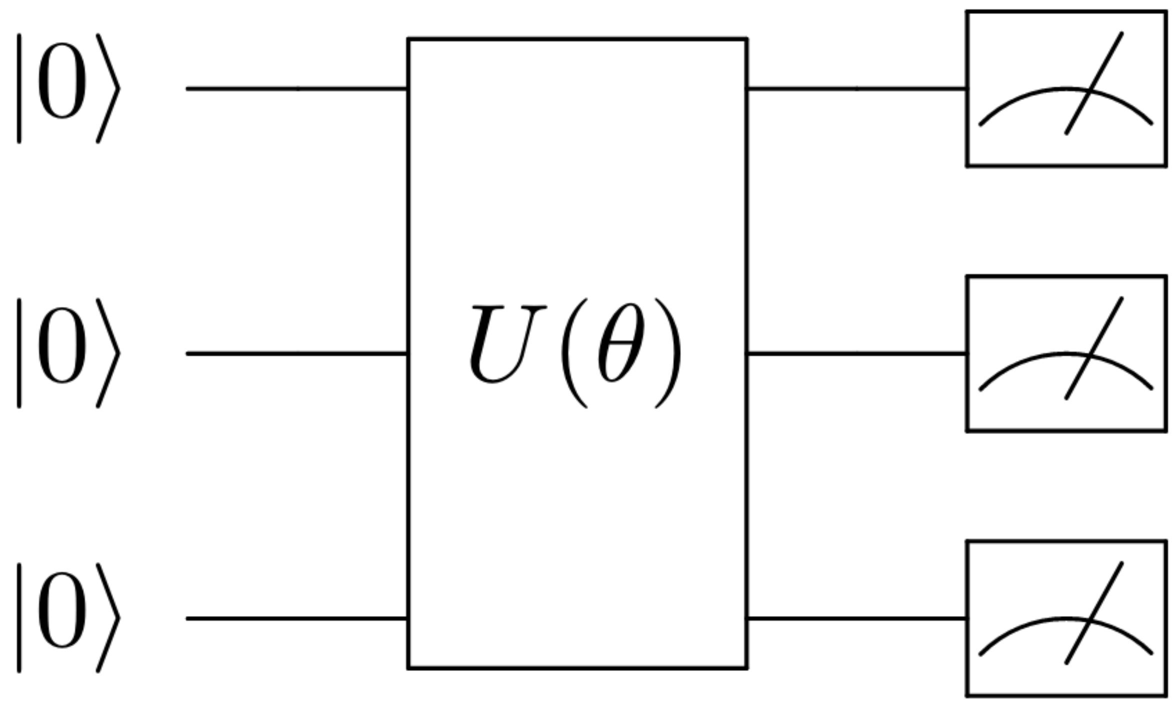
Significantly Higher Accuracy Than Shallow Networks

~4x10^-5

07.

Quantum Variational Representations.

07.1 - General Setup



Parameterized Quantum Circuit

$$L(\theta)$$
$$\nabla_\theta L(\theta)$$

Stochastic Estimate of Loss Function and Gradients

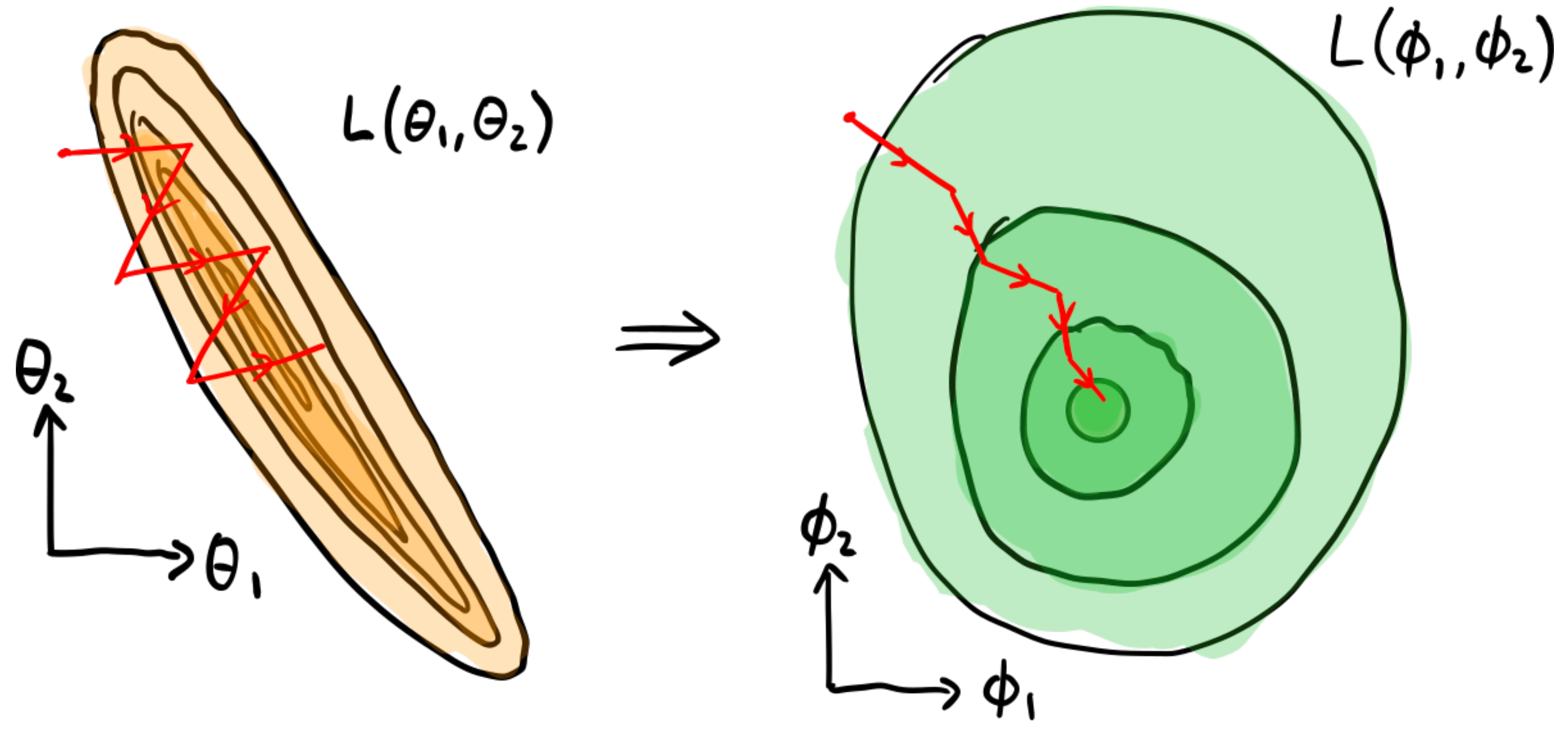
$$\theta^{(k+1)} = \theta^k - \eta \nabla_\theta L(\theta)$$

Iterative Minimization

07.2 - Quantum Natural Gradient

$$\theta^{(k+1)} = \theta^k - \eta g^{-1}(\theta^k)\nabla_\theta L(\theta)$$

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



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

$$g_{ij}(\theta) = \mathrm{Re}\left[\left\langle \frac{\partial\Psi}{\partial\theta_i}\middle|\frac{\partial\Psi}{\partial\theta_j}\right\rangle - \left\langle \frac{\partial\Psi}{\partial\theta_i}\middle|\Psi\right\rangle\left\langle\Psi\middle|\frac{\partial\Psi}{\partial\theta_j}\right\rangle\right]$$

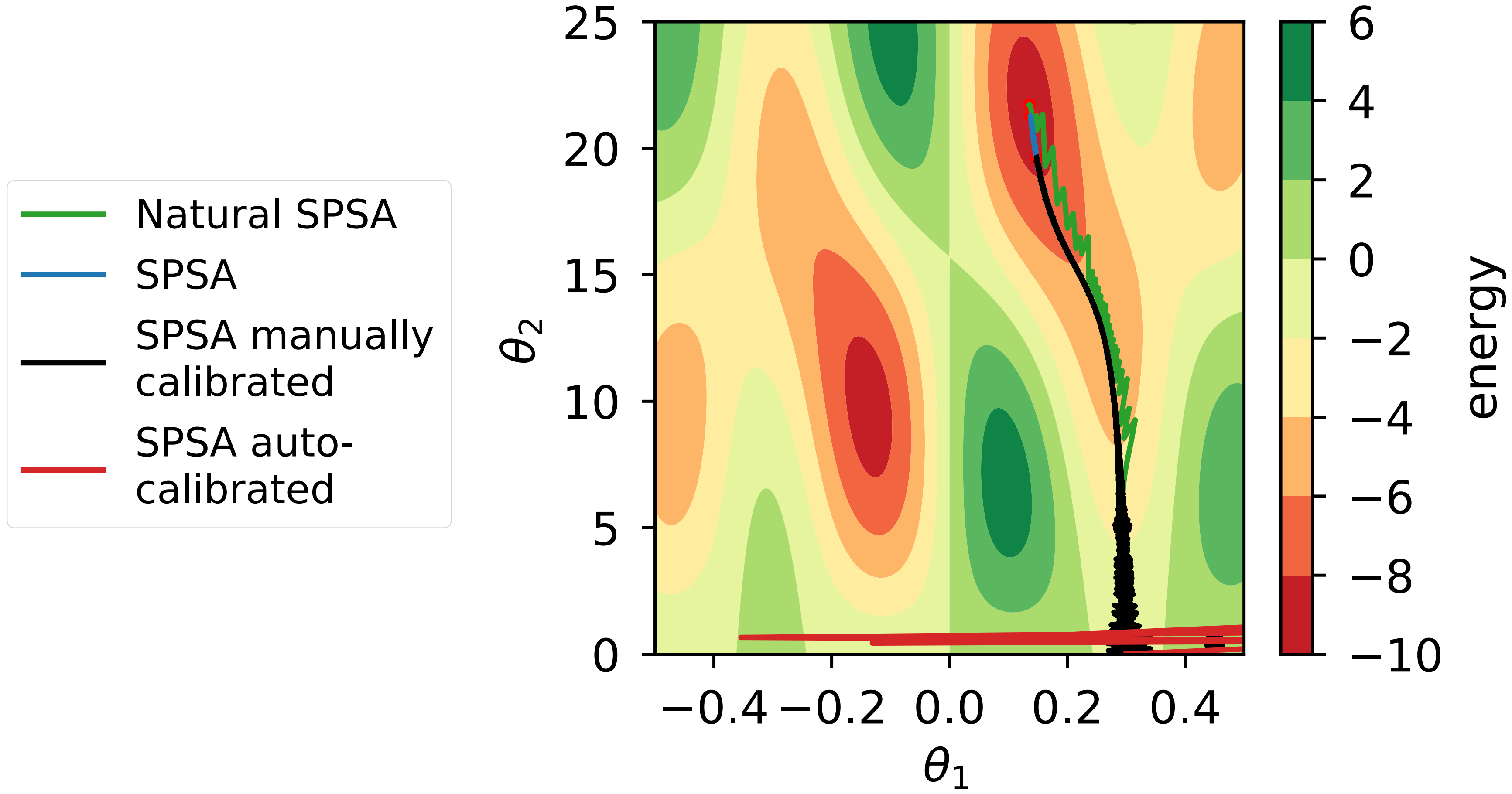
07.3 - Strong Interplay With Classical Methods

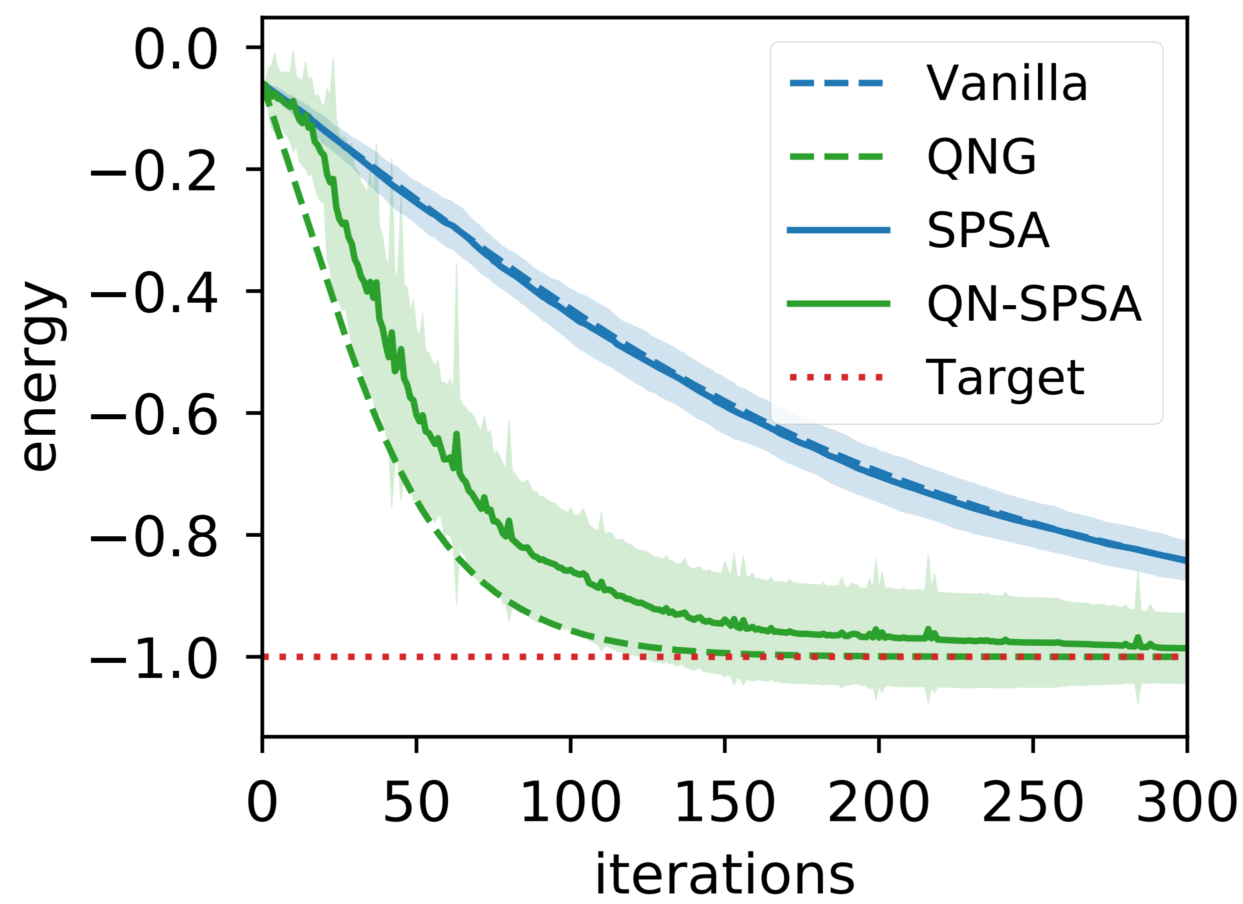
Problem	Classical Stochastic	Quantum
Variational Minimization	Variational Monte Carlo [*McMillan*, 1965]	Variational Quantum Eigensolver [*Peruzzo et al*, 2014]
Variational Imaginary Time Evolution	Stochastic Reconfiguration [*Sorella*, 1998]	[*McArdle et al*, 2019]
Variational Real Time Evolution	Time-Dependent Variational Monte Carlo [*Carleo et al*, 2012]	TDVA [*Lee and Benjamin*, 2017]
Machine Learning	Natural Gradient Descent [*Amari*, 1998]	Quantum Natural Gradient Descent [*Stokes et al*, 2020]

07.4 - Fast Quantum Natural Gradient: QN-SPSA

$$g_{ij}(\theta) = -\frac{1}{2}\frac{\partial}{\partial\theta_i}\frac{\partial}{\partial\theta_j}\left|\langle\psi_{\theta'}|\psi_\theta\rangle\right|^2\Big|_{\theta'=\theta}$$

Gacon, Zoufal, Carleo, and Woerner
arXiv:2103.09232, (2021)





07.

Outlook.

Classical Variational Simulation

Highly Entangled State

Exact Sampling [Autoregressive]

General Guiding Principle for Networks?

Efficiently Enforce Symmetries

Quantum Variational Simulation

Potentially More Expressive

“Arbitrary” Unitaries

Noise

Shallow Circuits

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)
```

Numba

Sharir, Shashua, and Carleo
arXiv:2103.10293, 2021

Choo, Mezzacapo, and Carleo
Nat. Comm. 11, 2368 (2020)

Choo, Mezzacapo, and Carleo
Nat. Comm. 11, 2368 (2020)

Stokes, Moreno, Pnevmatikakis, and Carleo
Phys. Rev. B 102, 205122 (2020)

Gacon, Zoufal, Carleo, and Woerner
arXiv:2103.09232, (2021)