AB INITIO CALCULATIONS OF MEDIUM-MASS NUCLEI WITH NEURAL QUANTUM STATES

JANE KIM
Argonne National Laboratory

Challenges in effective field theory descriptions of nuclei
Hirschegg, Austria
21 January 2026

OUTLINE

Neural quantum states (NQS) & conventional variational wave functions

Various interactions explored in this work

Preliminary results on medium-mass nuclei

Future directions

VARIATIONAL MONTE CARLO

Variational principle: $\langle E \rangle \equiv \frac{\langle \Psi | \hat{H} | \Psi \rangle}{\langle \Psi | \Psi \rangle} \geq E_0$

Monte Carlo integration:

$$\langle E \rangle = \frac{\int dX \;\; \Psi^*(X) \hat{H} \Psi(X)}{\int dX \;\; \Psi^*(X) \Psi(X)} \approx \frac{1}{N_s} \sum_{n=1}^{N_s} \frac{1}{\Psi(X_n)} \hat{H} \Psi(X_n), \quad \text{where } X_n \sim |\Psi(X)|^2$$

Simulate A nucleons with spatial, spin, and isospin degrees of freedom

$$X = \{\mathbf{x}_i\}_{i=1}^{A}$$
$$\mathbf{x}_i = (\mathbf{r}_i, s_i^z, t_i^z)$$
$$i = 1, 2, \ldots, A$$

CONVENTIONAL ANSÄTZE FOR NUCLEI

$$|\Psi\rangle = \hat{F}\,|\Phi\rangle$$

symmetric
short-range correlations

antisymmetric
long-range behavior, quantum numbers
e.g. Slater determinants, Pfaffians

e.g. $\hat{F} = \left(\hat{S}\sum_{i<j}\hat{F}_{ij}\right)\left(\hat{S}\sum_{i<j<k}(1+\hat{F}_{ijk})\right)$

$4^A \rightarrow \binom{A}{Z}\binom{A}{N_\uparrow} \sim \frac{4^A}{A}$ elements!

Evaluating the wave function involves storing and manipulating a very long vector of spin-isospin amplitudes

VMC can handle “hard” interactions, but we pay the price elsewhere

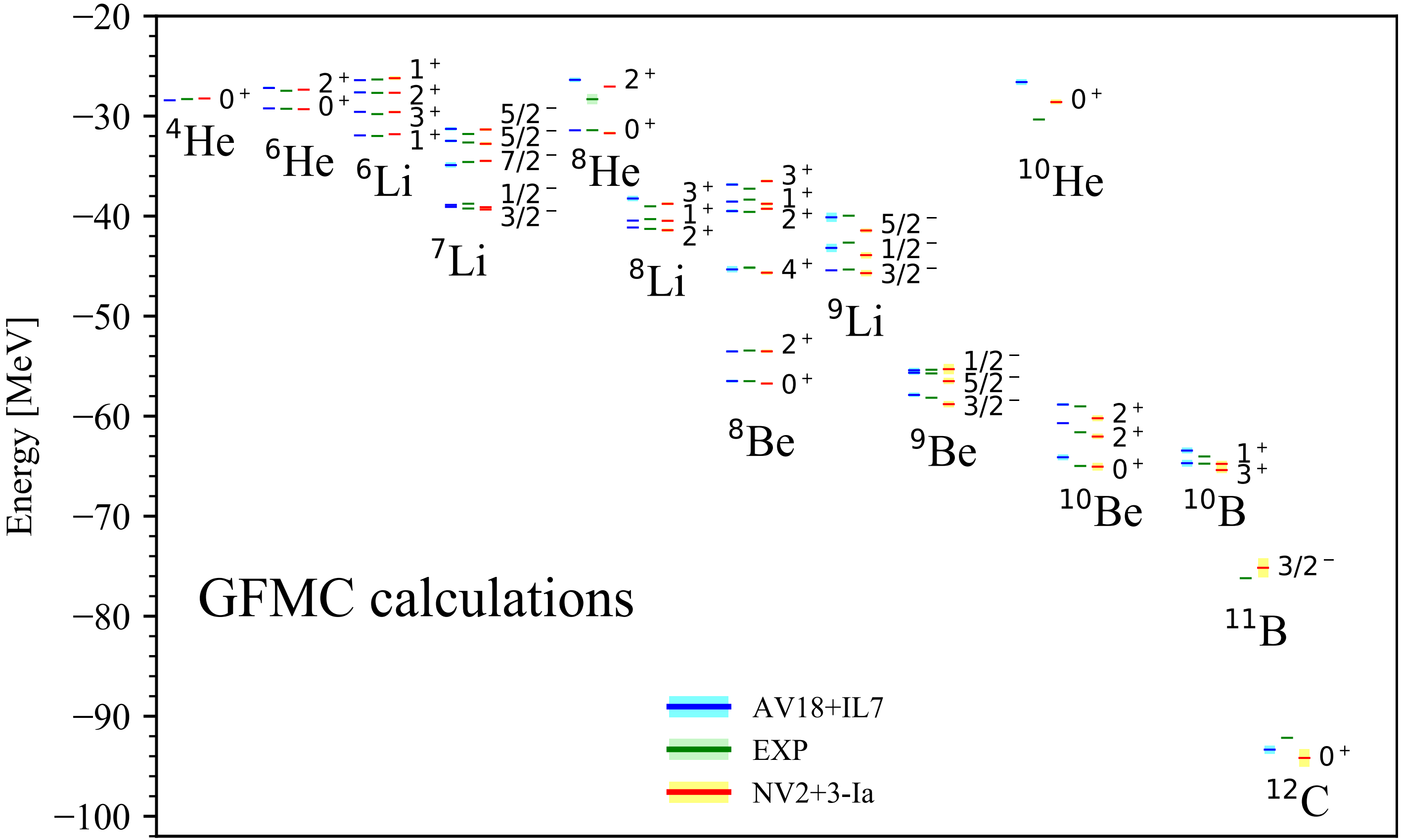
Always fine tuned with Green’s function Monte Carlo (GFMC)

ACCURACY OF VMC + GFMC

Percent-level accuracy for energies and radii

Local phenomenological and chiral interactions used here

Typically limited to $A \lesssim 12$ due to exponential scaling



M. Piarulli et al., PRL ***120****, 052503 (2018)*

G. Carleo and M. Troyer, Science 355, 6325 (2017).

NEURAL QUANTUM STATES

$$\Psi(X) = \underbrace{F(X)}_{\text{symmetric}}\underbrace{\Phi(X)}_{\text{antisymmetric}}$$

Similar idea as before, but parameterize by many feedforward neural networks…

very flexible mappings between two spaces
fast to evaluate
easy to differentiate

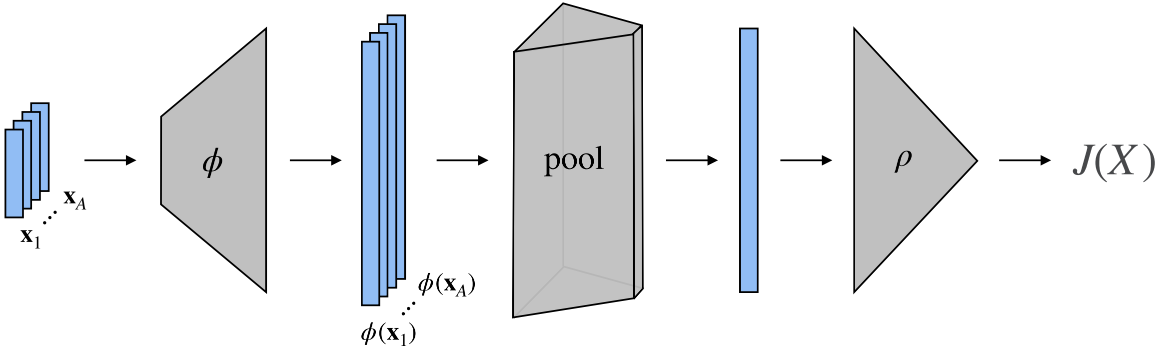
NEURAL QUANTUM STATES

$$\Psi(X) = F(X)\Phi(X)$$

symmetric

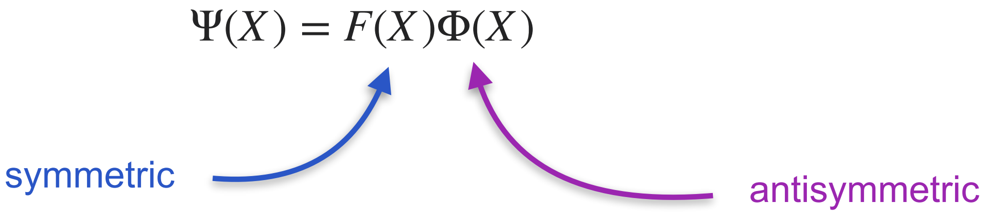
antisymmetric

Symmetric part: $F(X) = e^{J(X)}$ (positive definite)



Zaheer et al., arXiv:1703.06114 (2017)

NEURAL QUANTUM STATES



$$\Psi(X) = F(X)\Phi(X)$$

symmetric

antisymmetric

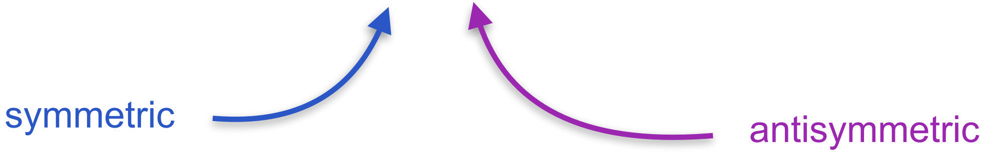
Antisymmetric part: often based on a Slater determinant

$$\Phi(X) = \det \begin{bmatrix} \phi_1(\mathbf{x}_1) & \phi_1(\mathbf{x}_2) & \cdots & \phi_1(\mathbf{x}_A) \\ \phi_2(\mathbf{x}_1) & \phi_2(\mathbf{x}_2) & \cdots & \phi_2(\mathbf{x}_A) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_A(\mathbf{x}_1) & \phi_A(\mathbf{x}_2) & \cdots & \phi_A(\mathbf{x}_A) \end{bmatrix}$$

A independent neural networks
or one with A different outputs

NEURAL QUANTUM STATES

$$\Psi(X) = F(X)\Phi(X)$$



Antisymmetric part: we use one based on a **Pfaffian**

$$\Phi(X) = \det \begin{bmatrix} \phi_1(\mathbf{x}_1) & \phi_1(\mathbf{x}_2) & \cdots & \phi_1(\mathbf{x}_A) \\ \phi_2(\mathbf{x}_1) & \phi_2(\mathbf{x}_2) & \cdots & \phi_2(\mathbf{x}_A) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_A(\mathbf{x}_1) & \phi_A(\mathbf{x}_2) & \cdots & \phi_A(\mathbf{x}_A) \end{bmatrix}$$

A independent neural networks or one with A different outputs

$$\Phi(X) = \mathrm{pf} \begin{bmatrix} 0 & \phi(\mathbf{x}_1, \mathbf{x}_2) & \cdots & \phi(\mathbf{x}_1, \mathbf{x}_A) \\ -\phi(\mathbf{x}_1, \mathbf{x}_2) & 0 & \cdots & \phi(\mathbf{x}_2, \mathbf{x}_A) \\ \vdots & \vdots & \ddots & \vdots \\ -\phi(\mathbf{x}_1, \mathbf{x}_A) & -\phi(\mathbf{x}_2, \mathbf{x}_A) & \cdots & 0 \end{bmatrix}$$

One neural network parameterizing the pairing orbital

JK et al., Commun. Phys. ***7****, 148 (2024).*

UNITARITY

First designed neural Pfaffian while studying the **unitary** Fermi gas

Divergent two-body scattering length $|a| \rightarrow \infty$ produces two-body state exactly at threshold

No intrinsic interaction length scale set by potential!

Ideal testbed for wave function flexibility, optimization, numerical stability

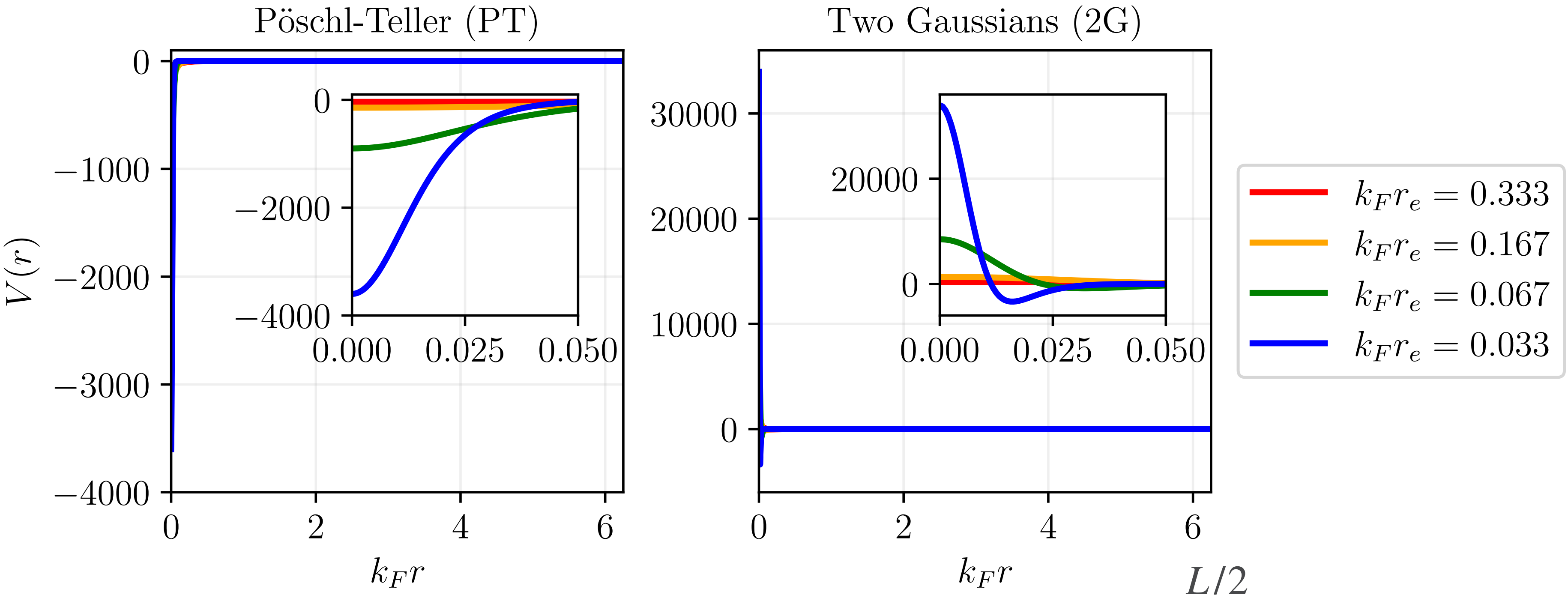
Low-energy NN s-wave interactions are near unitarity

JK et al., Commun. Phys. ***7****, 148 (2024).*

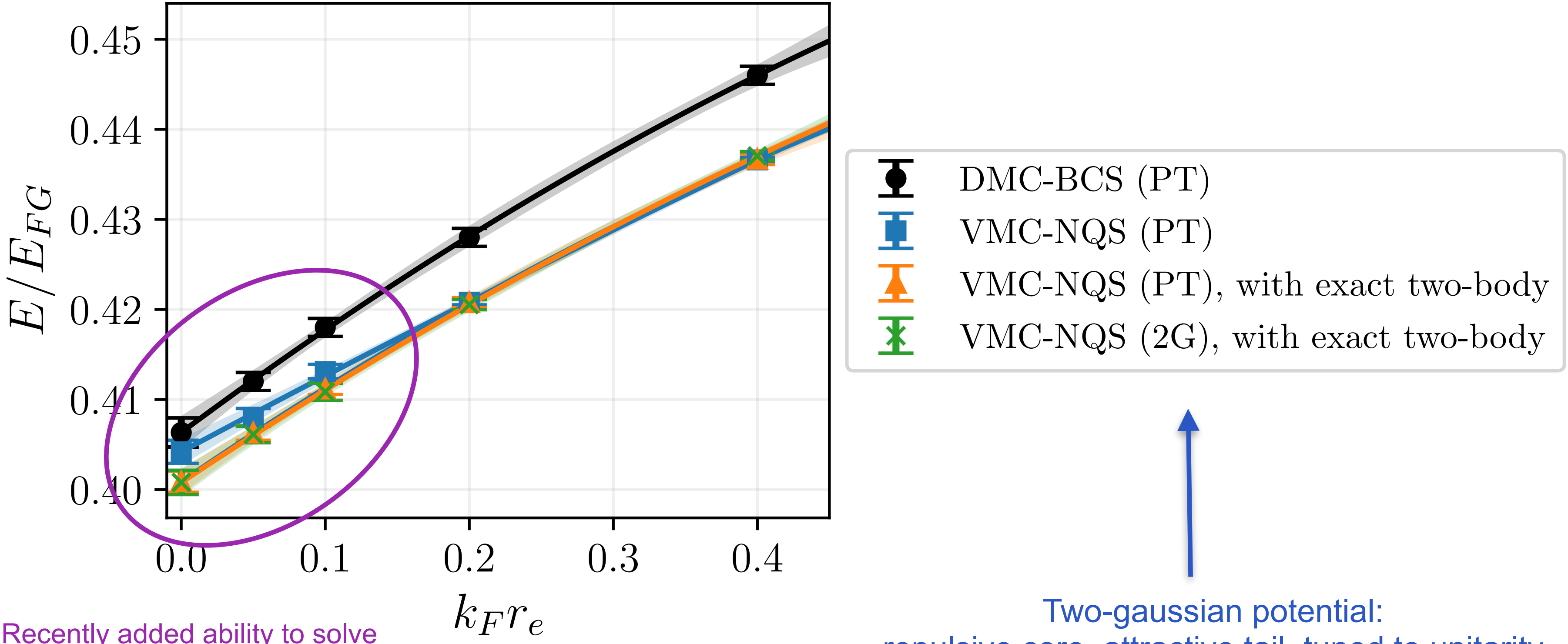
G. Pescia, JK, et al., PRB ***110****, 035108 (2024).*

UNITARITY

Two wildly different potentials tuned to unitarity should give identical results!



SHORT-RANGE CORRELATIONS



Recently added ability to solve two-body problem on-the-fly to assist short-range stability

Two-gaussian potential:
repulsive core, attractive tail, tuned to unitarity

NEURAL PFAFFIAN

Slater determinant is a special case of a Pfaffian → can capture different phases with one ansatz

Spin-isospin dependent pairing orbitals without explicit coding

Number of trainable parameters is **independent of system size**

Odd and even systems share a pairing orbital

No reliance on shell closures or explicit many-body basis, works directly in continuous space

Scales the same as a determinant

Not discussed: the graph neural network we use to build the bulk of our correlations (see refs)

JK et al., Commun. Phys. ***7****, 148 (2024).*

G. Pescia, JK, et al., PRB ***110****, 035108 (2024).*

UNITARY FERMI GAS (N=66)

Fixed-node diffusion Monte Carlo:

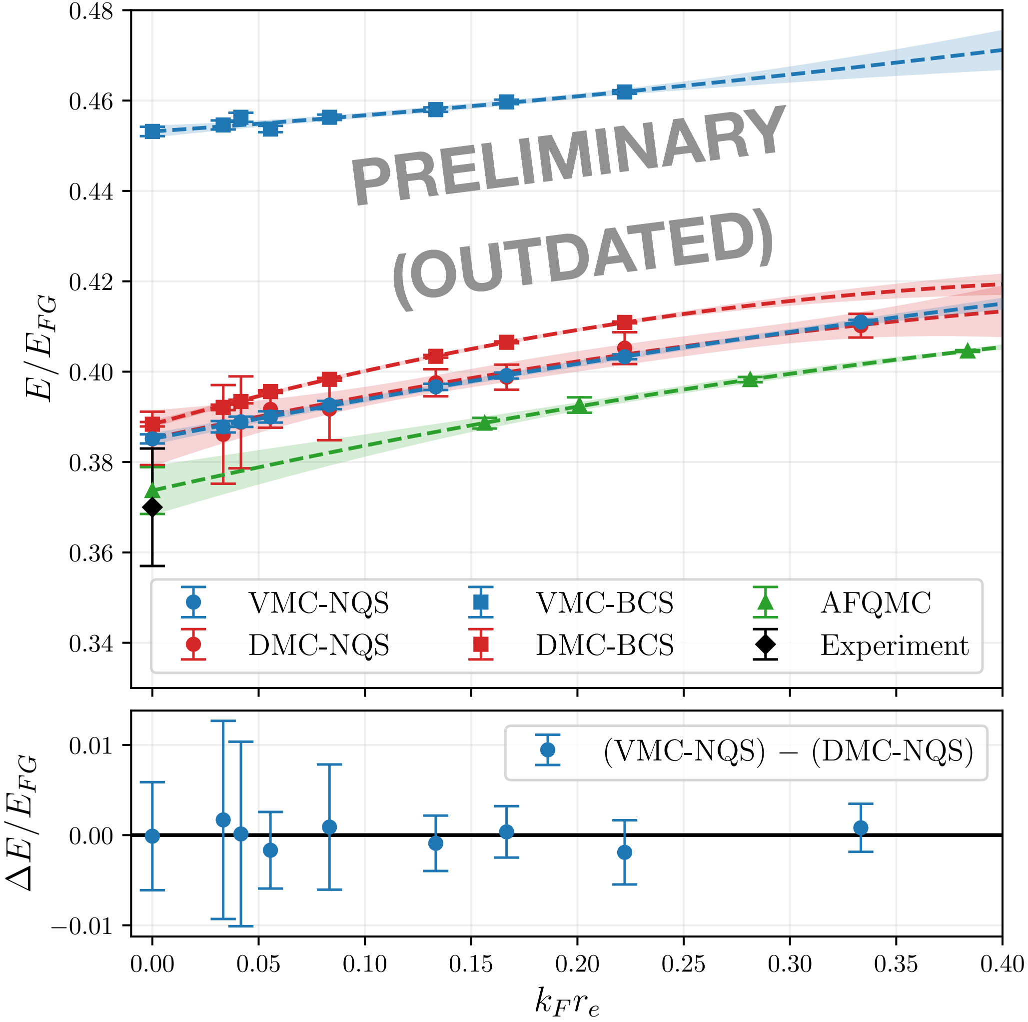
$$\lim_{\tau\to\infty}\left(\sum_i c_i e^{-\tau(E_i-E_T)}|\Psi_i\rangle\right)\propto|\Psi_0\rangle$$

In principle, gives the exact ground state

In practice, need the fixed-node approximation to control the fermionic sign problem

$$E_{VMC}\gtrsim E_{DMC}\gtrsim E_0$$

Newer calculations with even better NQS currently running



UNITARY LIMIT → PHYSICAL POINT

Unitary Fermi gas:

- Maximally nonperturbative and scale free
- Correlations everywhere, pairing
- Numerical unstable, high-variance local energies
- Good stress test of numerics, optimization, and wave function flexibility

Nuclear systems:

- Strong coupling between spin, isospin, spatial degrees of freedom
- Many operators, many competing channels
- Three-body forces

HAMILTONIANS

Key questions:

- Which aspects of nuclear structure are governed by low-energy physics?
- Which require resolving pion dynamics and short-distance details?
- How sensitive are bulk observables to physics above the pion scale?
- How much of nuclear binding is explained by near-unitary NN physics?

This work:

- Use the **simplest LO pionless EFT Hamiltonians**
- Treat them as a probe of universality, not a convergent microscopic theory

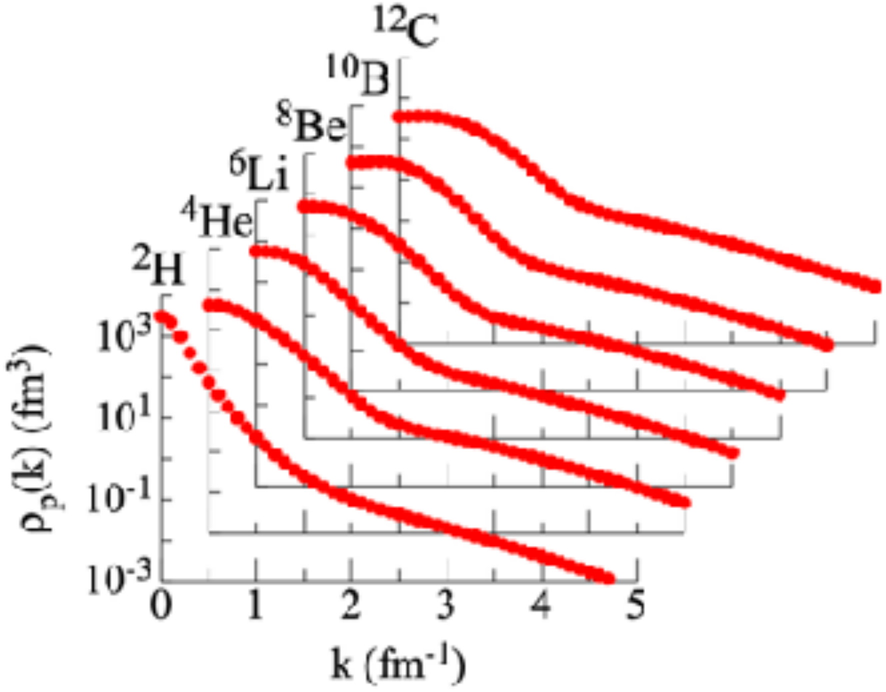
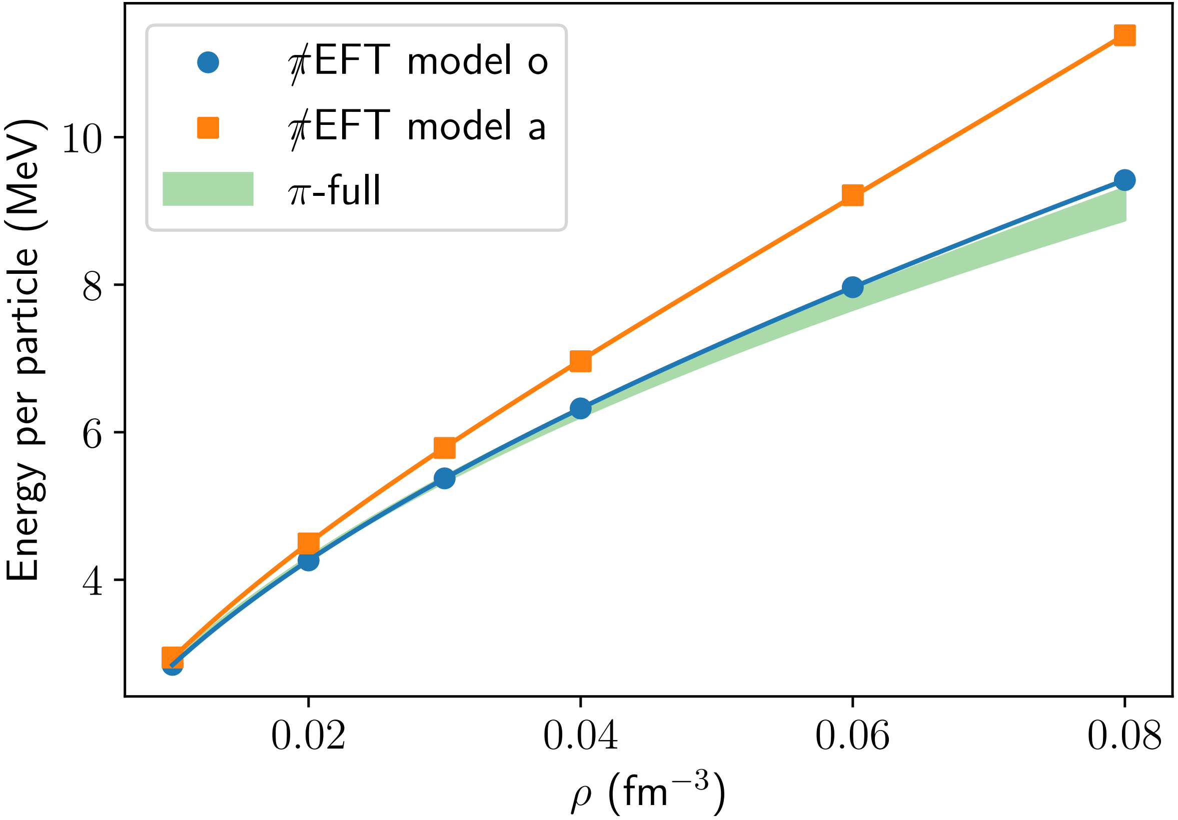


Figure: R. B. Wiringa et al., PRC **89**, 024305 (2014)

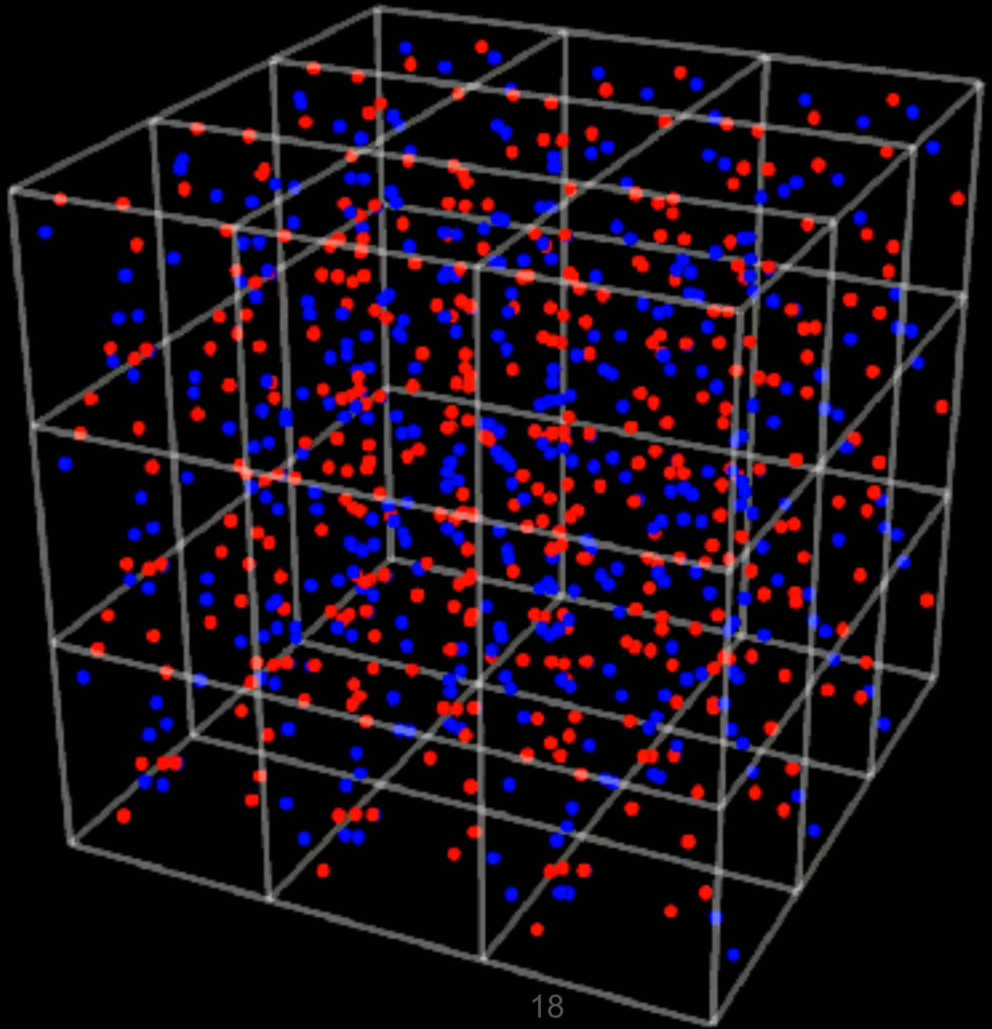
NUCLEAR MATTER



This simple LO pionless EFT Hamiltonian compared to pionfull Hamiltonians at low densities

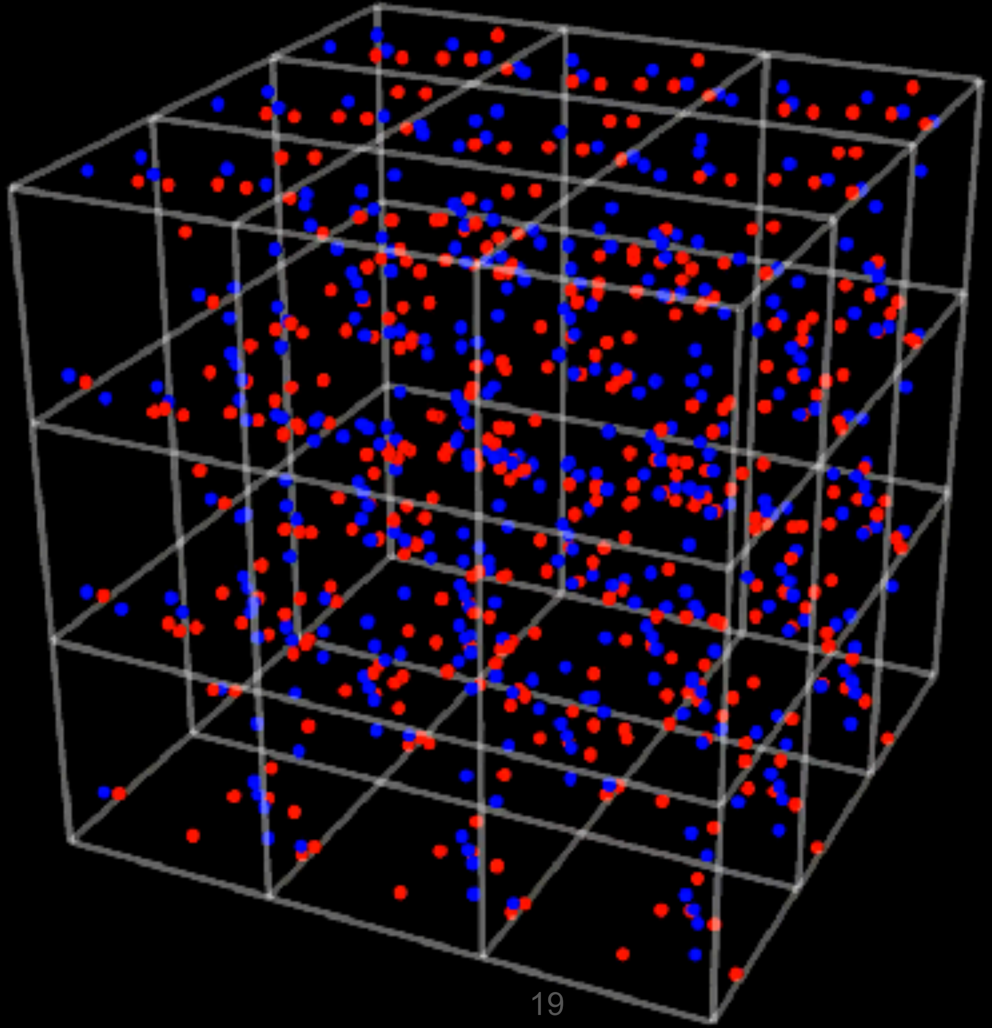
B. Fore, JK, et al., PRR 5, 033062 (2023).

Symmetric Nuclear Matter



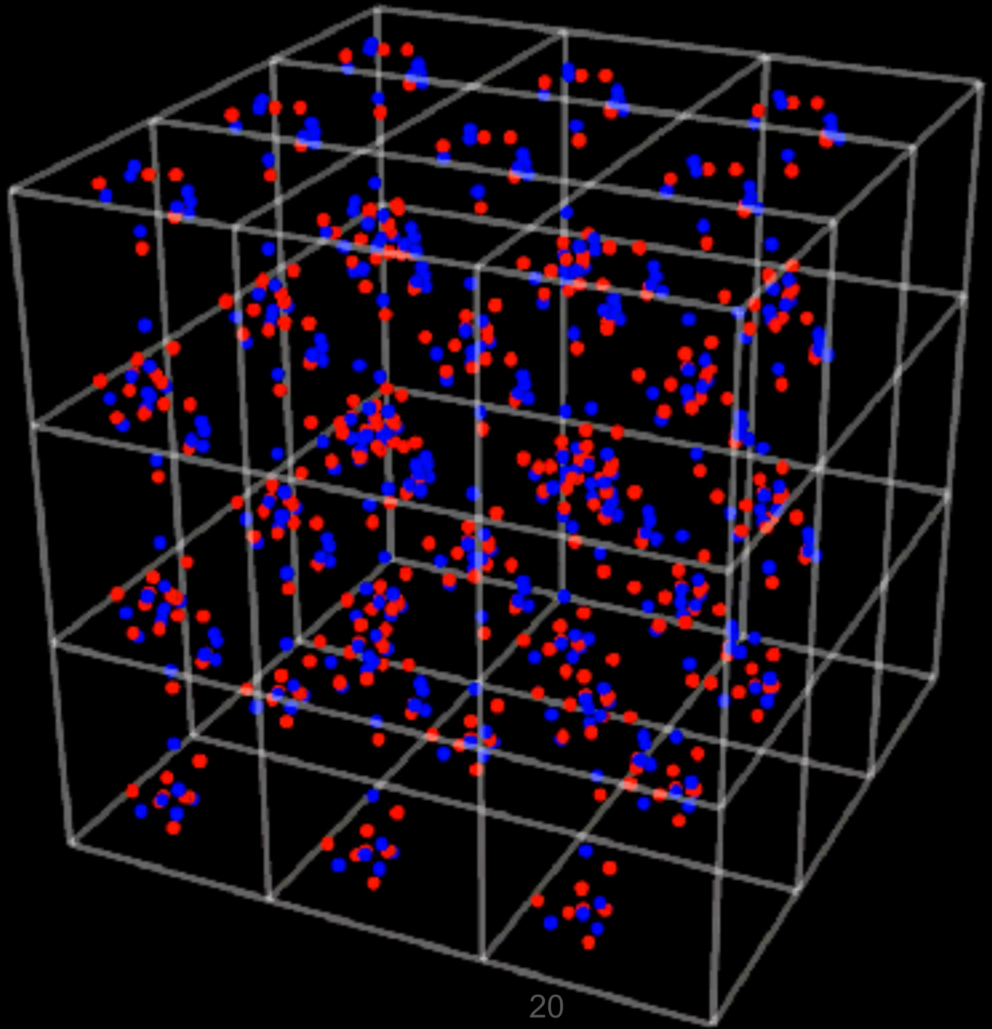
$n_B = 0.08\ \mathrm{fm}^{-3}$

Symmetric Nuclear Matter



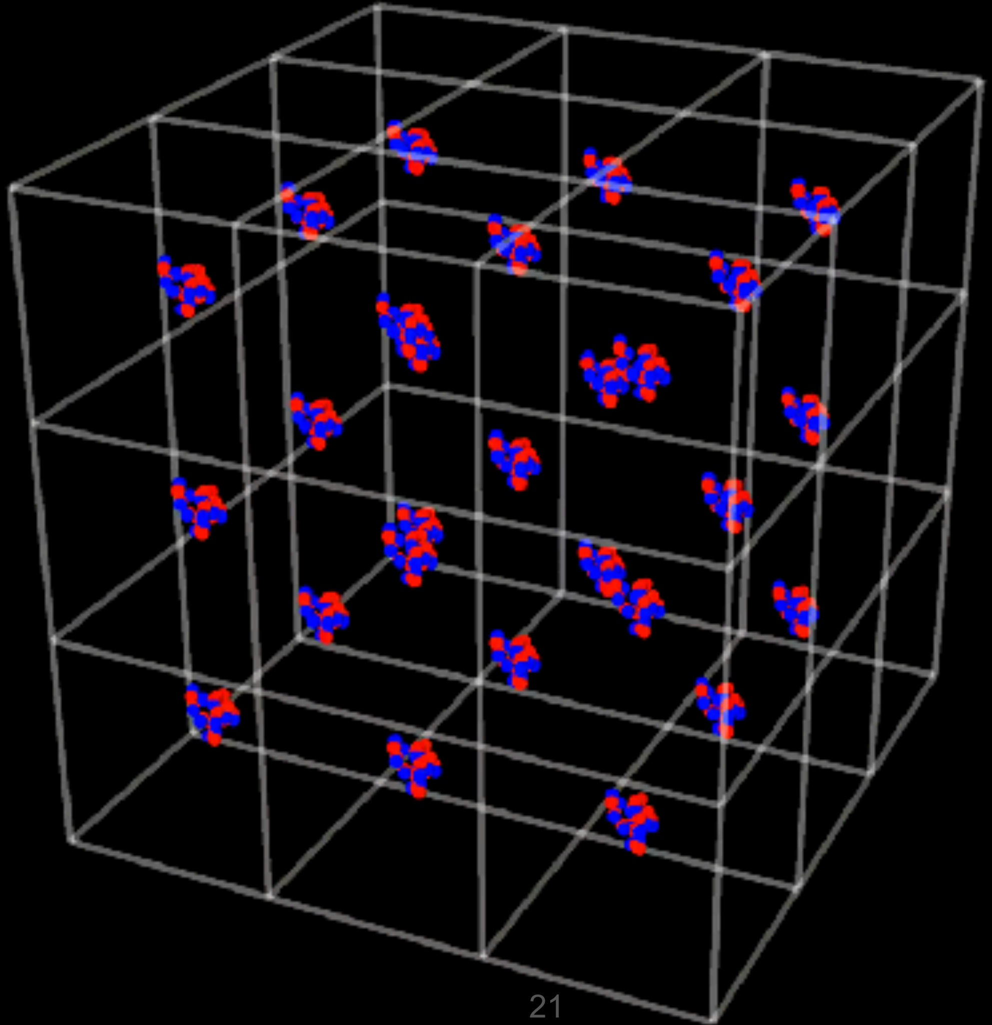
$n_B = 0.04 \ \mathrm{fm}^{-3}$

Symmetric
Nuclear
Matter



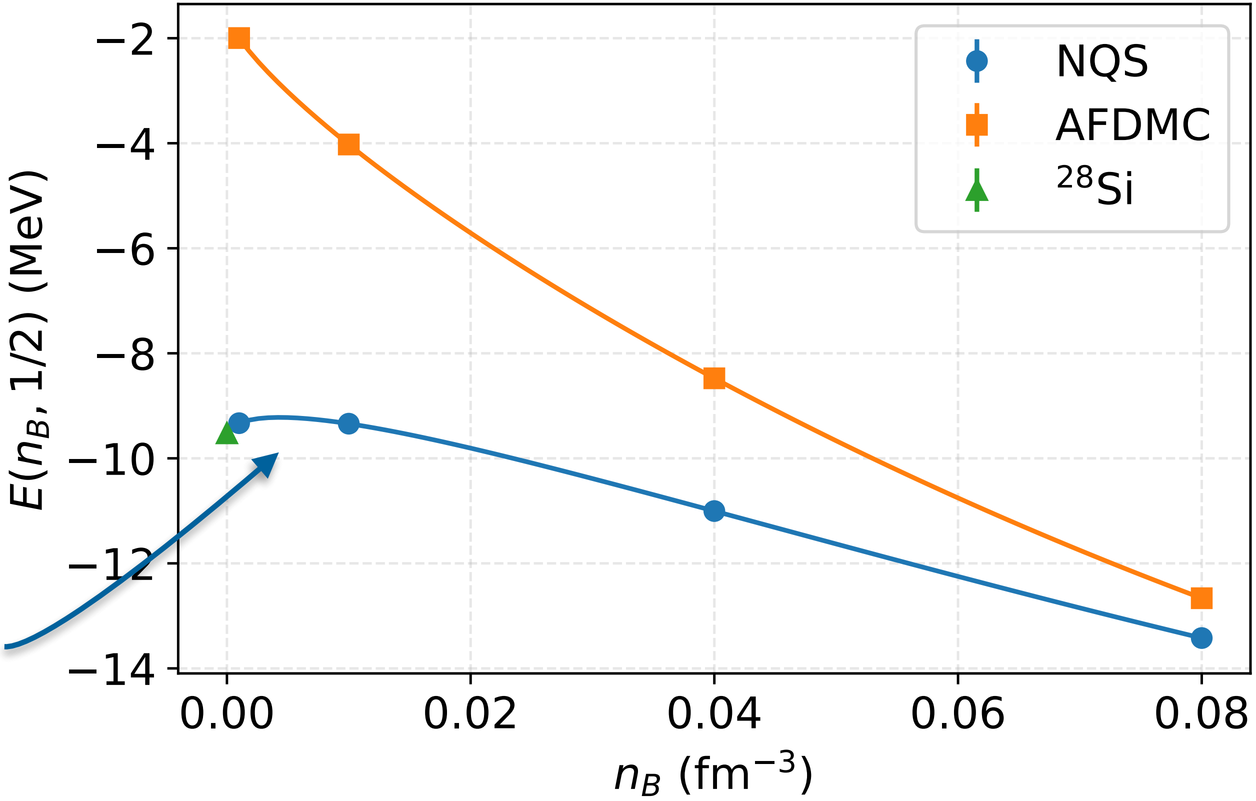
$n_B = 0.01\ \mathrm{fm}^{-3}$

Symmetric
Nuclear
Matter



$n_B = 0.001\ \mathrm{fm}^{-3}$

SYMMETRIC NUCLEAR MATTER



Clustering at
low densities?

B. Fore, JK, et al. Commun. Phys. 8, 108 (2025).

COMPOSITION OF NEUTRON STAR CRUSTS

Assume rates of beta decay and inverse beta decay are equal:

$$n \rightarrow p + e^- + \bar{\nu}_e$$

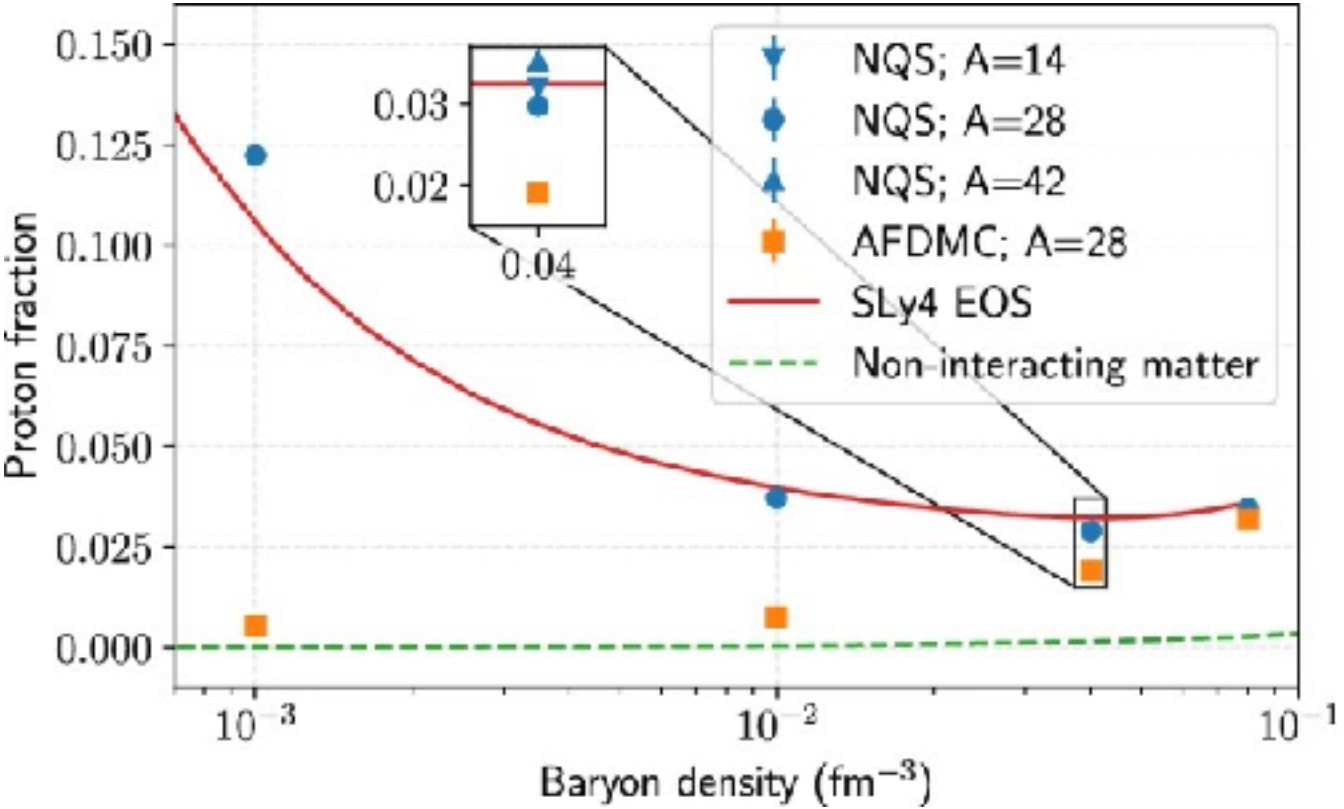
$$p + e^- \rightarrow n + \nu_e$$

$$\mu_n = \mu_p + \mu_e$$

Assume charge neutrality:

$$n_p = n_e$$

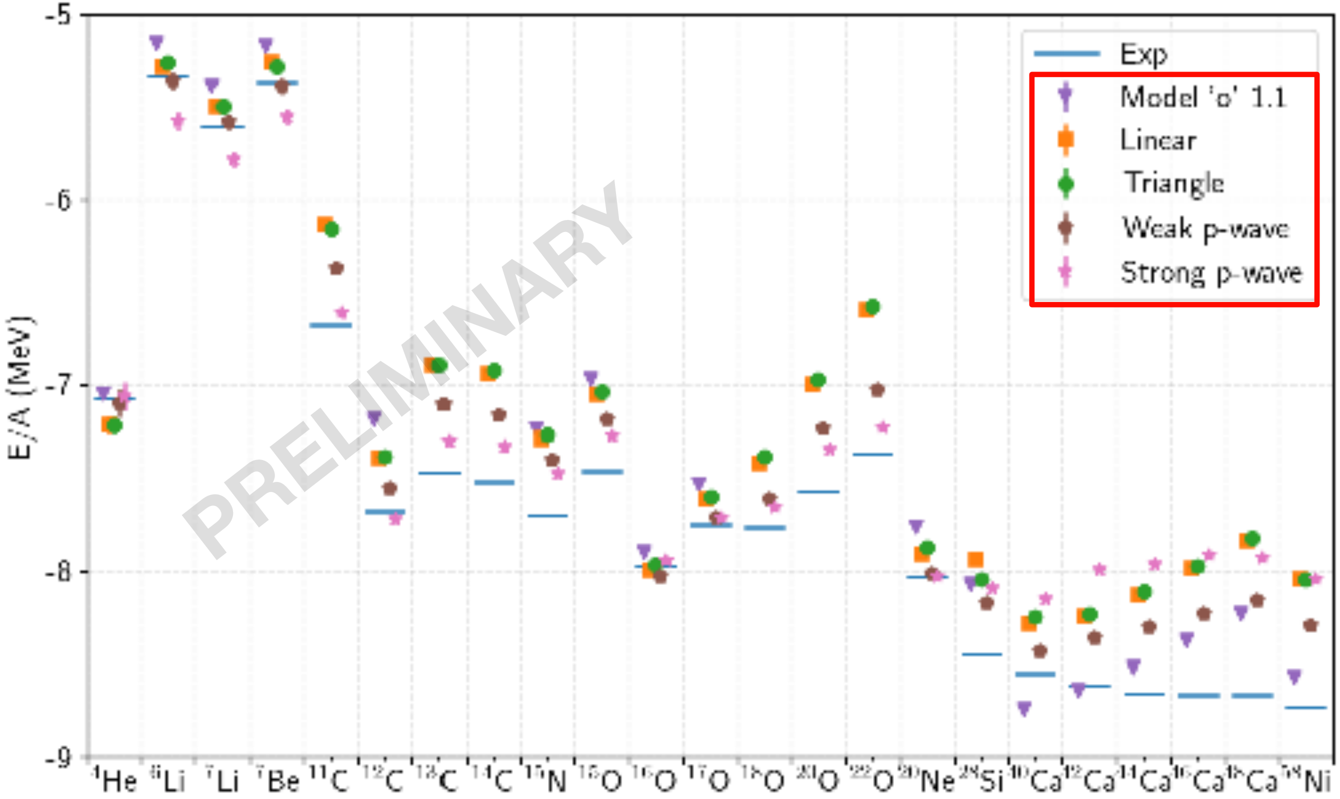
NQS agrees better with phenomenological Skryme models than AFDMC



B. Fore, JK, et al., Commun. Phys. **8**, 108 (2025).

B. Fore, JK, et al., in prep.

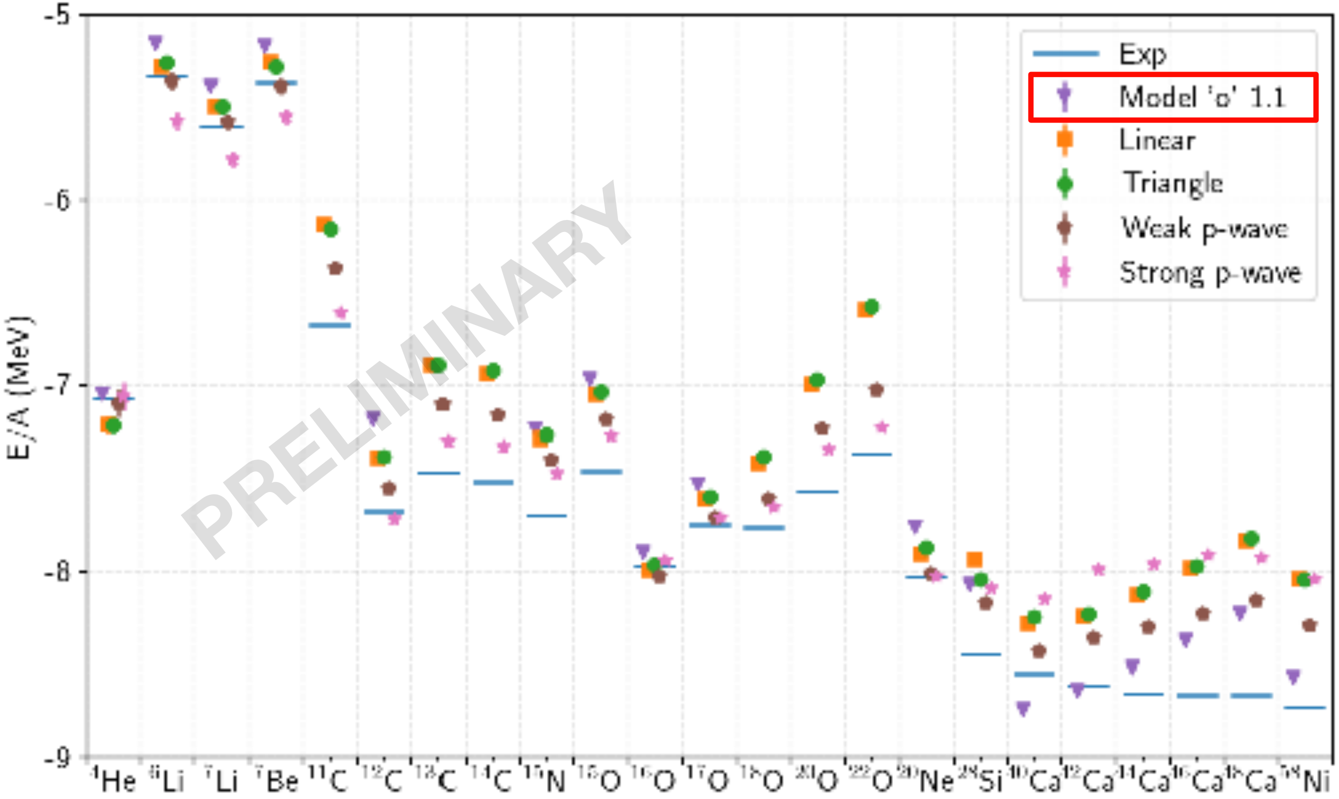
RESULTS: MEDIUM-MASS NUCLEI



Different LO pionless EFT Hamiltonians

B. Fore, JK, et al., in prep.

RESULTS: MEDIUM-MASS NUCLEI



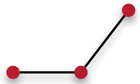
Our baseline
(same interaction we used for dilute nuclear matter)

R. Schiavilla et al.,
PRC 103, 054003 (2021).

LECs: C_{01}, C_{10}, c_E

Gaussian regulators:

$R_0 \approx 1.55$ fm
$R_1 \approx 1.83$ fm
$R_3 = 1.1$ fm



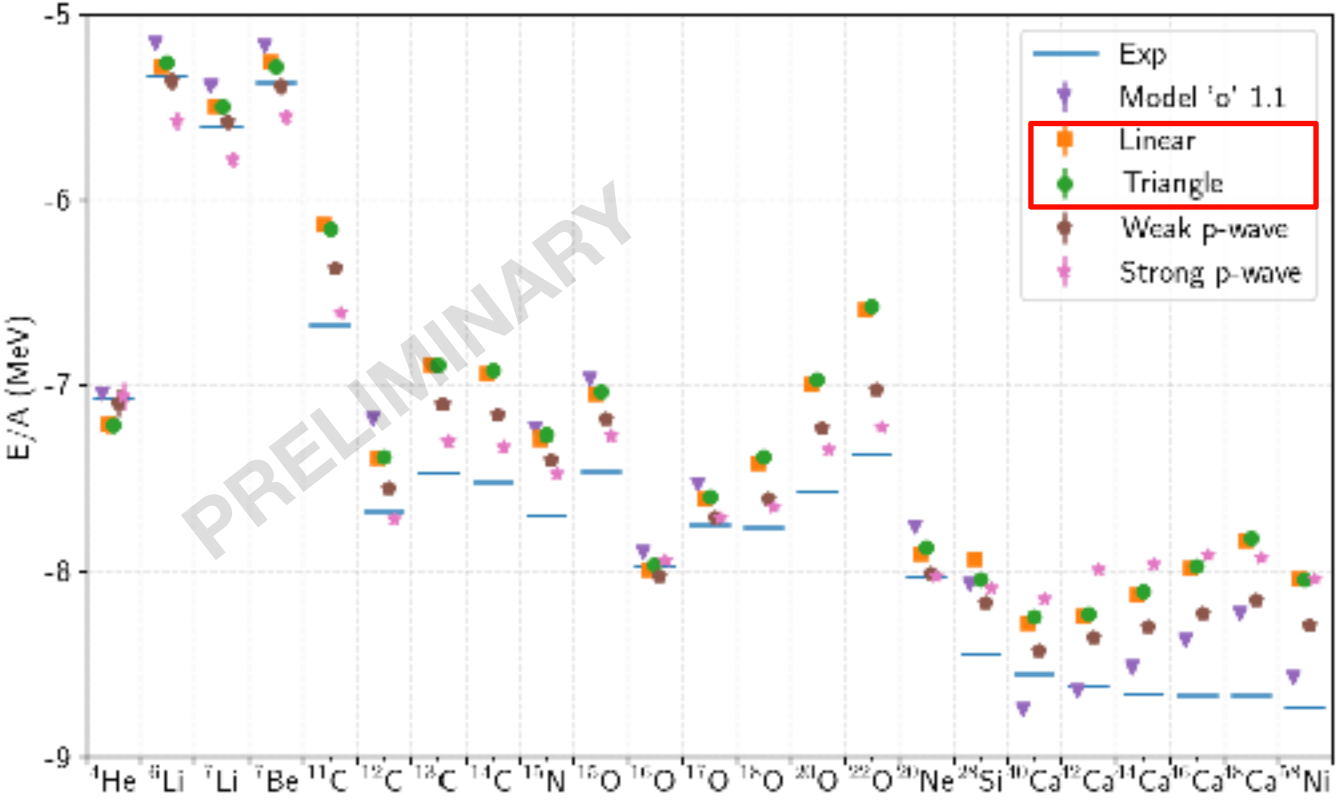
Three-body contact involves **two** pair distances

Fit to:

- np scattering lengths and effective ranges in $S/T = 0/1,\ 1/0$
- triton binding energy

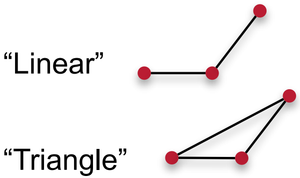
B. Fore, JK, et al., in prep.

RESULTS: MEDIUM-MASS NUCLEI



Add CD and CA terms at LO to the 2-body interaction for $S/T = 0/1$

Fit different 3-body interactions by hand using NQS calculations of He-4 and O-16



Gaussian regulators:

$R_0 = 1.537$ fm

$R_1 = 1.813$ fm

Two additional LECs: C_{CD}, C_{CA}

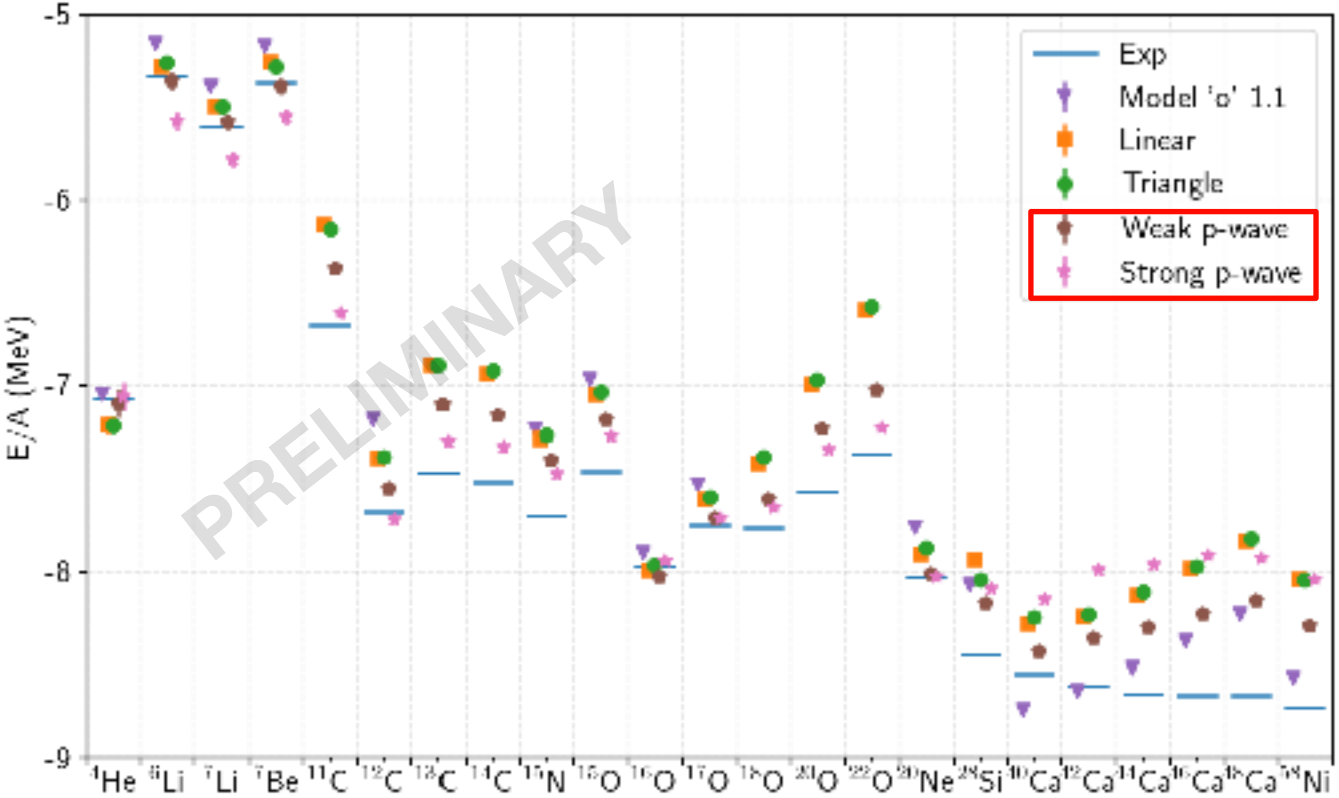
All two-body LECs fit to low-energy phase shifts up to 5 MeV

np, pp - Nijmegen

nn - Argonne $v18$

B. Fore, JK, et al., in prep.

RESULTS: MEDIUM-MASS NUCLEI



Add S/T = 0/0, 1/1 to model “o”

Fit different 3-body interactions by hand using NQS calculations of He-4 and O-16

M. Gattobigio et al.,
PRC 100, 034004 (2019).

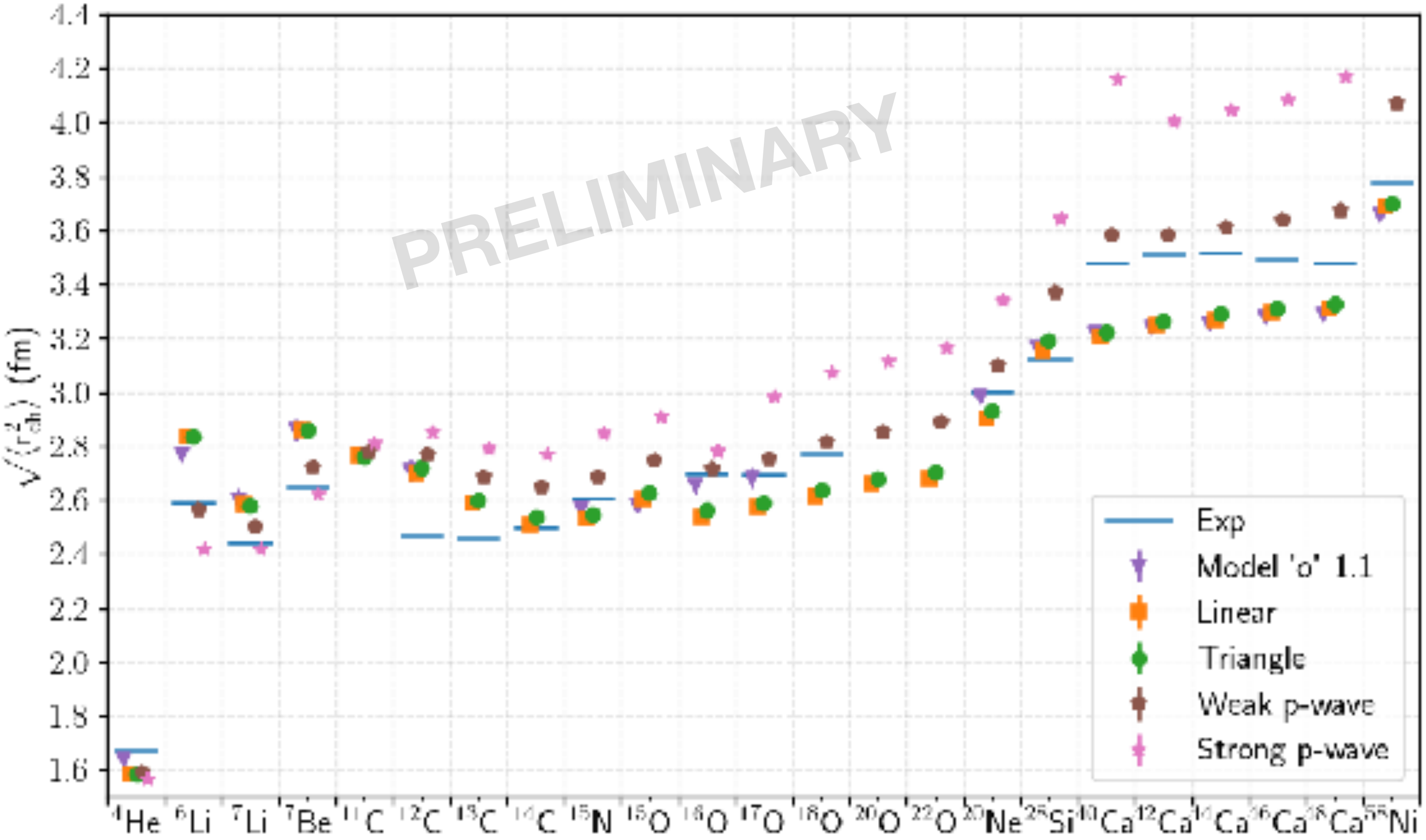
Fit to:

- s-wave scattering lengths and effective ranges
- triton binding energy
- C_{00} and C_{11} tuned to AV14 phase shifts for p-wave sector

C_{00} fixed, smaller C_{11} used for “weak” and larger for “strong”

B. Fore, JK, et al., in prep.

RESULTS: MEDIUM-MASS NUCLEI



FUTURE DIRECTIONS

R. Curry, et al., arXiv:2510.15860 (2025).

Uncertainty quantification:

All error bars reported here are statistical errors not model uncertainties

Reduce uncertainties from many-body method as much as possible, push all uncertainties to Hamiltonian

Construct reduced order models using trained NQS as basis states

Ni-58 calculations required $\sim$9,000 GPU hours → small generalized eigenvalue problem

Excited states:

D. Pfau et al., Science 385, 6711 (2024).
Hendry et al., arXiv:2507.10287 (2025).
Kahn et al., arXiv:2507.08930 (2025).

Newly developed Grassmann VMC developed in the NQS community

Evolve a subspace of NQS as one variational object

Plus everything that has already been done with conventional variational states…

CONCLUSION

NQS provide a flexible framework for describing strongly correlated quantum systems, applicable to both **finite nuclei and infinite matter**, and capable of capturing **emergence of different phases.**

High-accuracy energies and radii can be achieved for $A \leq 58$, far **beyond the reach of conventional VMC.**

Initial calculations with extremely simple pionless EFT Hamiltonians suggest that aspects of **low-energy physics persist in medium-mass nuclei**, but rigorous uncertainty quantification is imperative.

Thank you!

VNIVERSITAT DE VALÈNCIA	Alessandro Lovato
Argonne NATIONAL LABORATORY	Bryce Fore Anthony Tropiano
Los Alamos NATIONAL LABORATORY	Stefano Gandolfi
UNIVERSITY OF OSLO	Morten Hjorth-Jensen
EPFL	Giuseppe Carleo Gabriel Pescia
ETH zürich	Jannes Nys

PERIODIC BOUNDARY CONDITIONS

Minimum-image convention

For nuclear matter, we also sum the interaction over nearby boxes

Periodic separation vectors:

$$\mathbf{r}_{ij} = \mathbf{r}_i - \mathbf{r}_j \quad \mapsto \quad \tilde{\mathbf{r}}_{ij} = \left(\cos\left(\frac{2\pi}{L}\mathbf{r}_{ij}\right),\ \sin\left(\frac{2\pi}{L}\mathbf{r}_{ij}\right) \right)$$

Periodic distances:

$$r_{ij} = \|\mathbf{r}_{ij}\| \quad \mapsto \quad \tilde{r}_{ij} = \left\| \sin\left(\frac{\pi}{L}\mathbf{r}_{ij}\right) \right\|$$

PARITY AND TIME-REVERSAL

For the unpolarized system:

$$\Psi^{P}(R,S) = \Psi(R,S) + \Psi(-R,S)$$

$$\Psi^{PT}(R,S) = \Psi^{P}(R,S) + (-1)^{N/2}\Psi^{P}(R,-S)$$

Not necessary, but can help the NQS train with fewer iterations

Symmetries can be added or removed at will

GRASSMANN VMC

In August 2024, Pfau et al. showed that using a determinant of many-body states could be used to find the low-lying states of atoms and molecules.

In July 2025, two articles simultaneously formalized the idea in terms of Grassmannian geometry. Benchmarked on 2D Heisenberg model and transverse field Ising model.

Grassmannian $\mathbf{Gr}_M(\mathscr{H})$ = set of all M-dimensional linear subspaces of a Hilbert space $\mathscr{H}$

- Each “point” is a whole subspace
- Smooth differentiable manifold, natural metric, symplectic structure
- Many representations to exploit

D. Pfau et al., Science 385, 6711 (2024). Hendry et al., arXiv:2507.10287 (2025). Kahn et al., arXiv:2507.08930 (2025).

GRASSMANN VMC

Ordered basis $\Psi = (\psi_1, \psi_2, \ldots, \psi_M)$

Matrix representation $|\Psi]] = [\,|\psi_1\rangle, \,|\psi_2\rangle, \,\ldots, \,|\psi_M\rangle]^T$

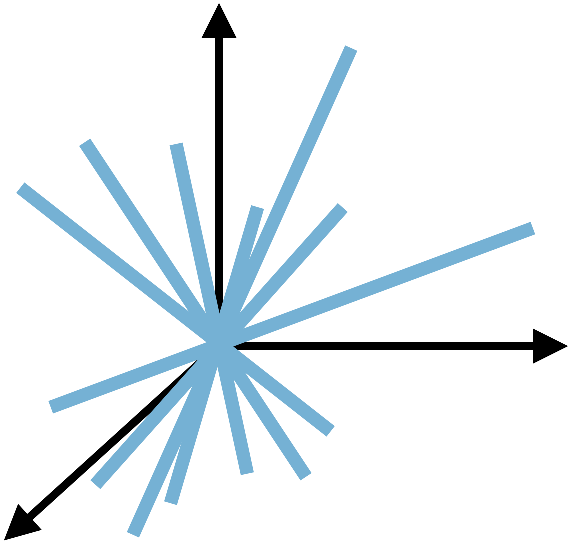
Wedge-product representation $|\Psi\rangle\rangle = \hat{\mathscr{A}}\,|\Psi]] \quad \Longrightarrow \quad \Psi(X) = \langle\langle X\,|\,\Psi\rangle\rangle \propto \det[\psi_i(x_j)]$

Operator expectation value becomes a matrix:

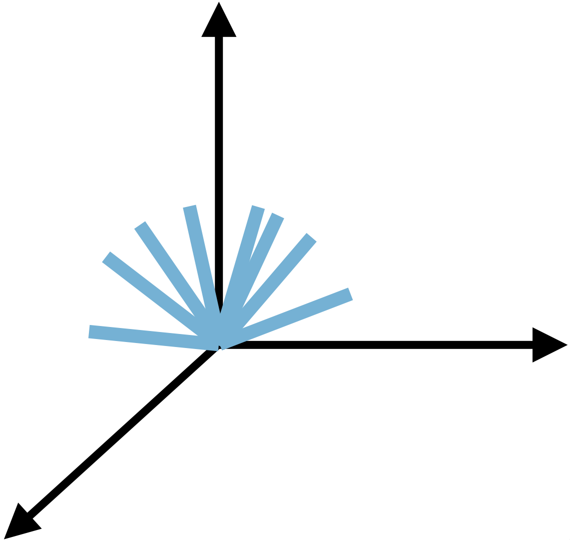
$$\frac{\langle \psi\,|\,\hat{H}\,|\,\psi\rangle}{\langle \psi\,|\,\psi\rangle} \quad \Longrightarrow \quad [[\Psi\,|\,\Psi]]^{-1}[[\Psi\,|\,\hat{\mathbb{H}}\,|\,\Psi]]$$

D. Pfau et al., Science 385, 6711 (2024). Hendry et al., arXiv:2507.10287 (2025). Kahn et al., arXiv:2507.08930 (2025).

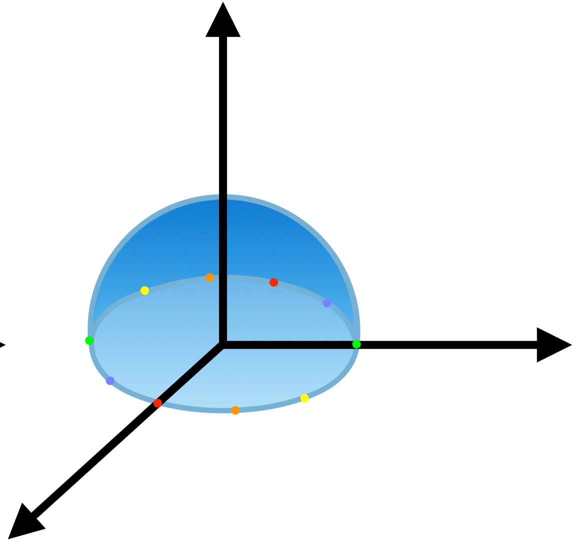
EXAMPLE: $\mathbf{Gr}_1(\mathbb{R}^3)$



1d subspaces



Normalize and select $z > 0$



Real projective plane RP^2

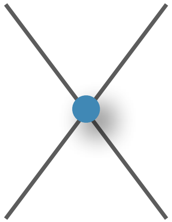
PIONLESS EFT (INSPIRED) HAMILTONIAN

A low-energy EFT of QCD with only nucleons as degrees of freedom

Appropriate when momenta are well below the pion mass, or for nuclear matter up to $\sim n_0/2$

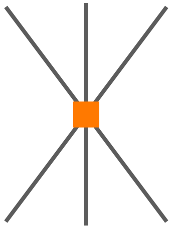
$$\hat{H}_{LO} = -\sum_i \frac{\nabla_i^2}{2m_i} + \sum_{i<j} v_{ij} + \sum_{i<j<k} V_{ijk}$$

- NN potential: fit to np scattering lengths and effective radii and the deuteron binding energy
- 3NF adjusted to reproduce the ^{3}H binding energy.



$$v_{ij}^{\mathrm{CI}} = \sum_{p=1}^{4} v^p(r_{ij}) O_{ij}^p \, ,$$

$$O_{ij}^{p=1,4} = (1, \tau_{ij}, \sigma_{ij}, \sigma_{ij}\tau_{ij})$$



$$V_{ijk} = \tilde{c}_E \sum_{\mathrm{cyc}} e^{-(r_{ij}^2 + r_{jk}^2)/R_3^2}$$

R. Schiavilla, et al., PRC ***103****, 054003(2021)*

VARIATIONAL MONTE CARLO

Based on the variational principle:

$$E(\theta) = \frac{\langle \Psi_\theta | \hat{H} | \Psi_\theta \rangle}{\langle \Psi_\theta | \Psi_\theta \rangle} \quad \Longrightarrow \quad \min_\theta E(\theta) \geq E_0$$

Use Monte Carlo integration to handle the high-dimensional integrals:

$$\frac{\langle X | \Psi_\theta \rangle}{\langle X | \Psi_\theta \rangle}$$

$P(X)$ $E_\theta(X)$

$$\frac{\langle \Psi_\theta | \hat{H} | \Psi_\theta \rangle}{\langle \Psi_\theta | \Psi_\theta \rangle} = \frac{\int dX \langle \Psi_\theta | X \rangle \langle X | \hat{H} | \Psi_\theta \rangle}{\langle \Psi_\theta | \Psi_\theta \rangle} = \frac{\int dX \, |\Psi_\theta(X)|^2 \frac{1}{\Psi_\theta(X)} \hat{H} \Psi_\theta(X)}{\langle \Psi_\theta | \Psi_\theta \rangle}$$

$$= \int dX P(X) E_\theta(X) \approx \mathbb{E}_{X \sim P(X)} E_\theta(X)$$

OPTIMIZATION

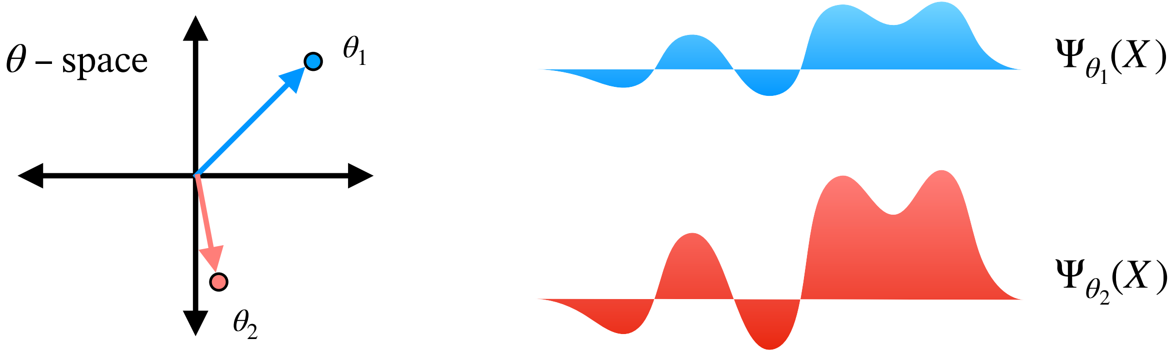
Take the gradient of the energy w.r.t. variational parameters:

$$\nabla_\theta E(\theta) = 2\left(\frac{\langle \Psi_\theta | \hat{H} | \nabla_\theta \Psi_\theta \rangle}{\langle \Psi_\theta | \Psi_\theta \rangle} - E(\theta) \frac{\langle \Psi_\theta | \nabla_\theta \Psi_\theta \rangle}{\langle \Psi_\theta | \Psi_\theta \rangle} \right)$$

Update the parameters using SGD: $\theta \mapsto \theta - \eta \nabla_\theta E(\theta)$

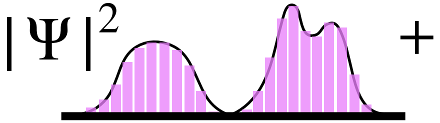
First-order optimization is fine for small systems, but when our NQS are compact, we can do better…

STOCHASTIC RECONFIGURATION



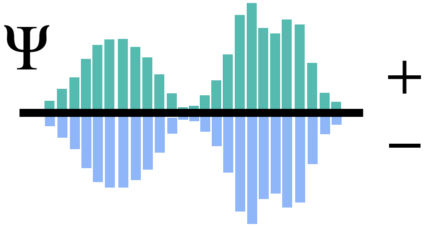
Precondition the gradient with the **quantum geometric tensor**: $\theta \mapsto \theta - \eta S^{-1} \nabla_\theta E(\theta)$

$$S_{ij} = \frac{\langle \partial_i \Psi_\theta | \partial_j \Psi_\theta \rangle}{\langle \Psi_\theta | \Psi_\theta \rangle} - \frac{\langle \partial_i \Psi_\theta | \Psi_\theta \rangle \langle \Psi_\theta | \partial_j \Psi_\theta \rangle}{\langle \Psi_\theta | \Psi_\theta \rangle}$$



Neural Quantum States

- Use artificial neural networks to write a flexible trial wave function



Variational Monte Carlo

- Optimize explicit trial wave function
- Accuracy limited by quality of trial
- No sign problem!

Based on variational principle

$$\frac{\langle \Psi_\theta | \hat{H} | \Psi_\theta \rangle}{\langle \Psi_\theta | \Psi_\theta \rangle} \geq E_0$$

Can compute any $\langle \hat{O} \rangle$ even when $[\hat{O}, \hat{H}] \neq 0$

Diffusion Monte Carlo

- Wave function is a distribution of configurations
- Variational (strict upper bound)
- Sign problem controlled by fixed-node approximation

Auxiliary-Field Quantum Monte Carlo

- Wave function is a distribution of mean-fields
- Sign problem controlled by constrained-path approximation
- Not variational when constrained

Based on imaginary-time projection

$$\lim_{\tau \to \infty} e^{-\hat{H}\tau/\hbar}\, \Psi(0) \propto \Psi_0$$

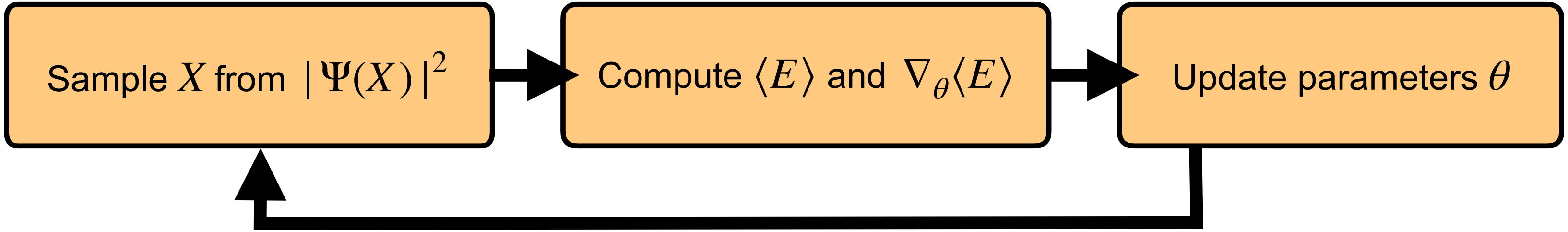
Biased estimates of $\langle \hat{O} \rangle$ when $[\hat{O}, \hat{H}] \neq 0$

NEURAL QUANTUM STATES

Variational wave functions based on artificial neural networks

Ab initio: Solve the Schrödinger equation starting from assumed Hamiltonian $\hat{H}$

What is the **input** to the network?	Many-body configurations $X = \{\mathbf{x}_i\}_{i=1}^{N}$
What is the **output** of the network?	The amplitude $\Psi(X)$
How do you **train** the network?	Gradient descent (or stochastic reconfiguration)

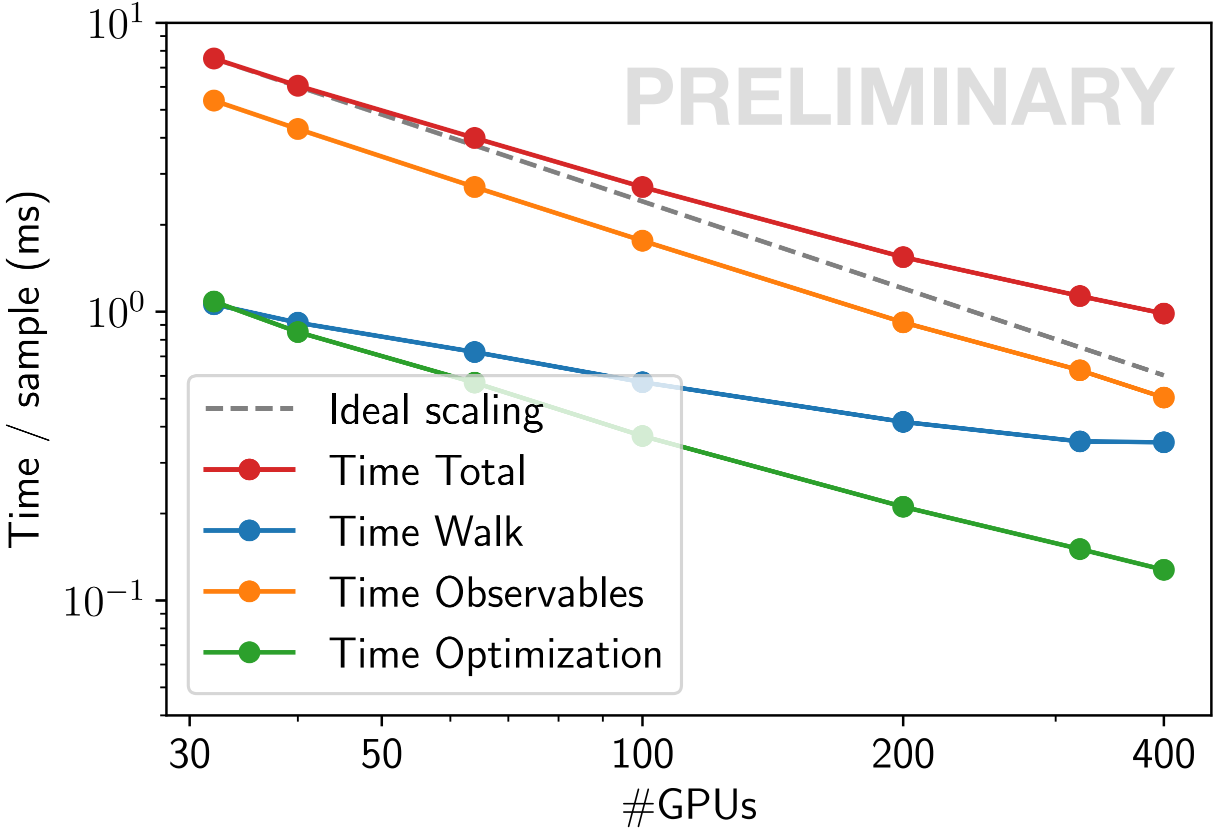

Sample X from $|\Psi(X)|^2$

Compute $\langle E \rangle$ and $\nabla_\theta \langle E \rangle$

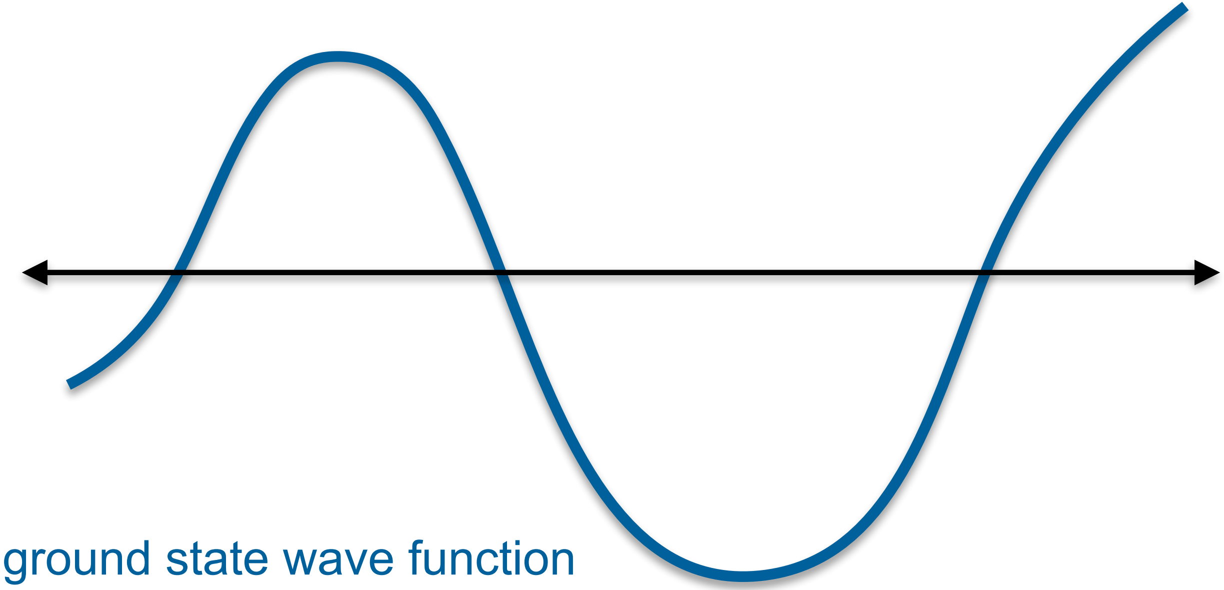
Update parameters θ

G. Carleo and M. Troyer, Science 355, 6325 (2017).

SCALING

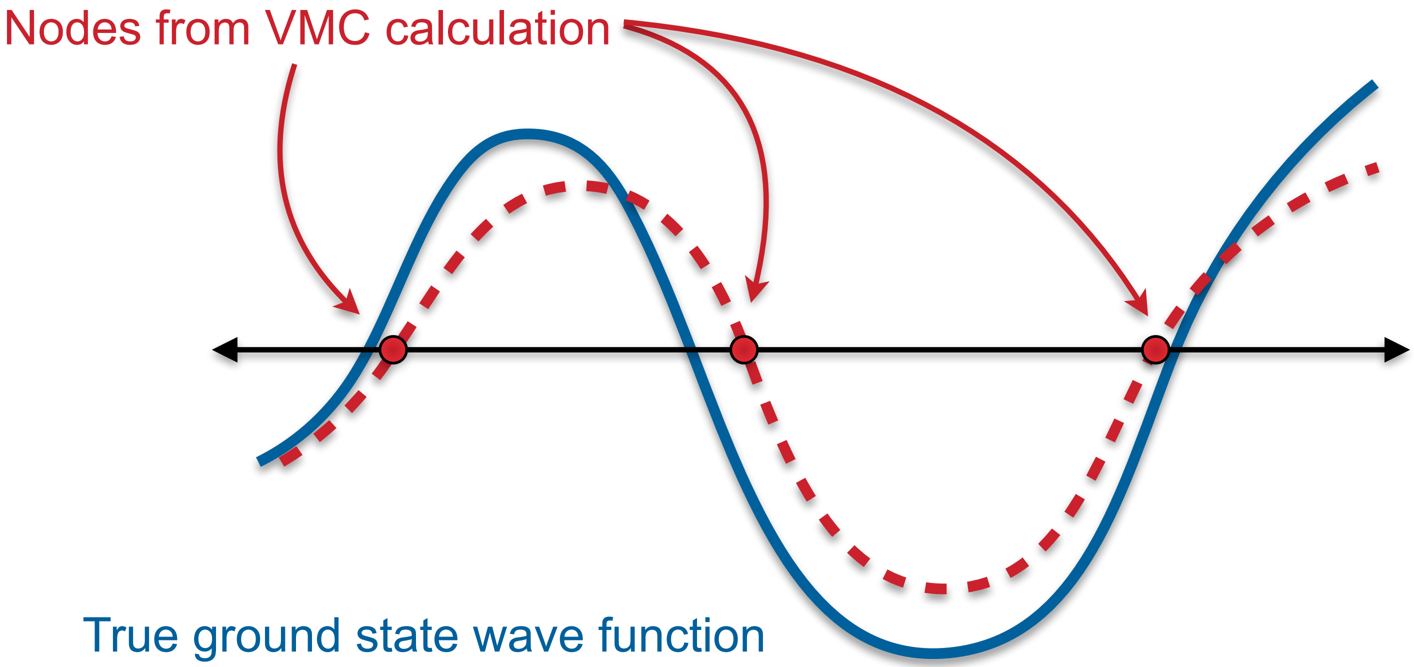

PRELIMINARY
10^1
10^0
10^{-1}
Time / sample (ms)
30
50
100
200
400
#GPUs
Ideal scaling
Time Total
Time Walk
Time Observables
Time Optimization

FIXED-NODE DIFFUSION MONTE CARLO



True ground state wave function

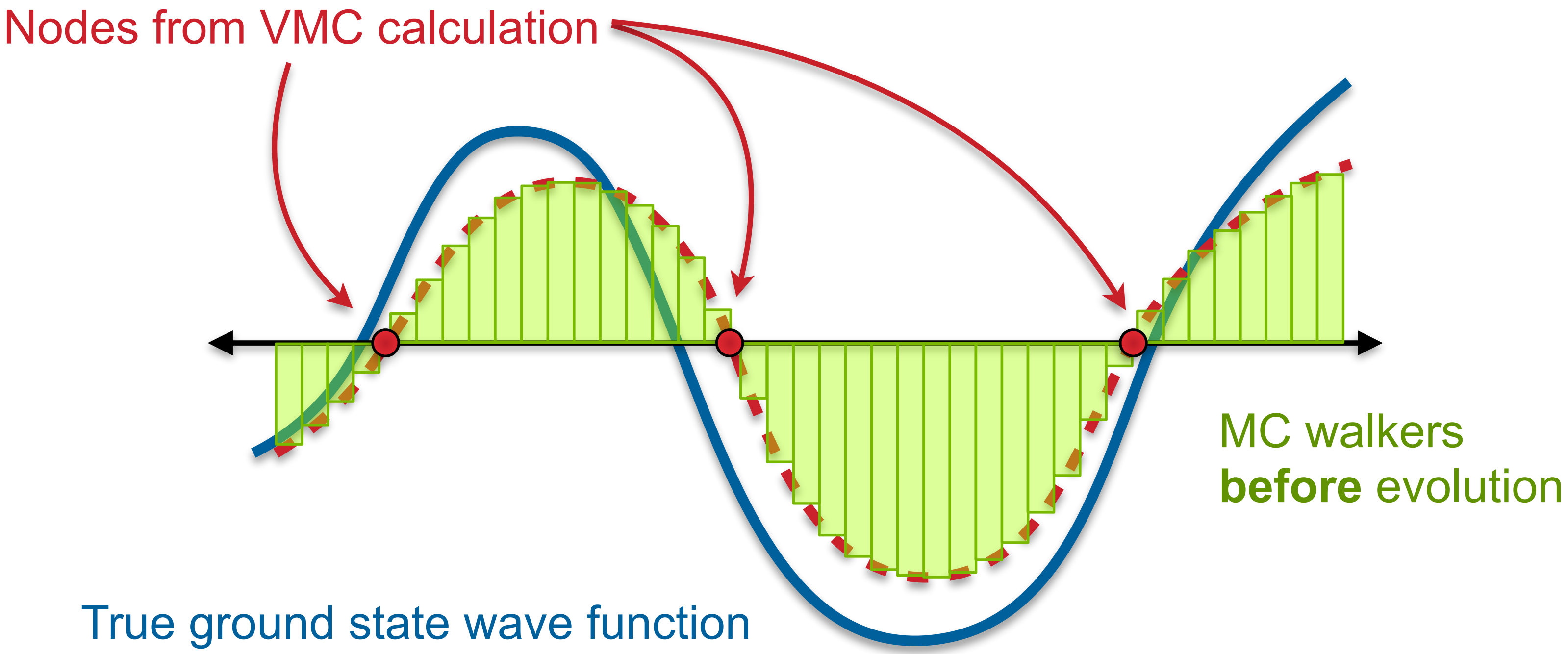
FIXED-NODE DIFFUSION MONTE CARLO



Nodes from VMC calculation

True ground state wave function

FIXED-NODE DIFFUSION MONTE CARLO

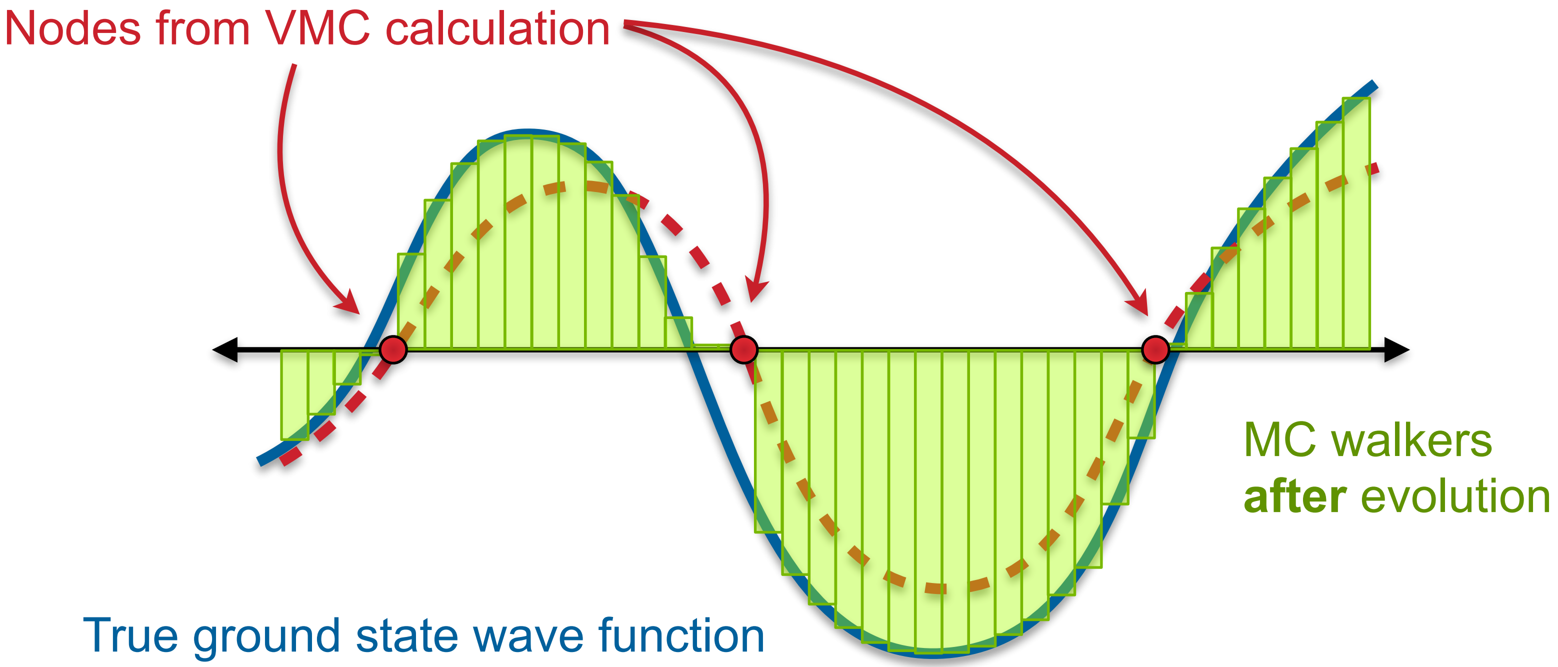


Nodes from VMC calculation

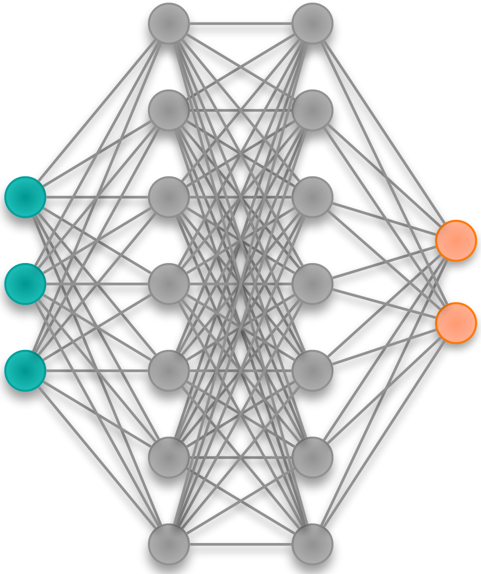
MC walkers **before** evolution

True ground state wave function

FIXED-NODE DIFFUSION MONTE CARLO



FEEDFORWARD NEURAL NETWORKS



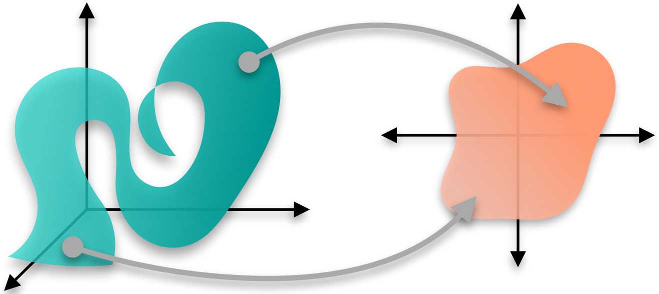
Inspired by the structure of the brain

Nodes are organized into layers, connections between neighboring layers

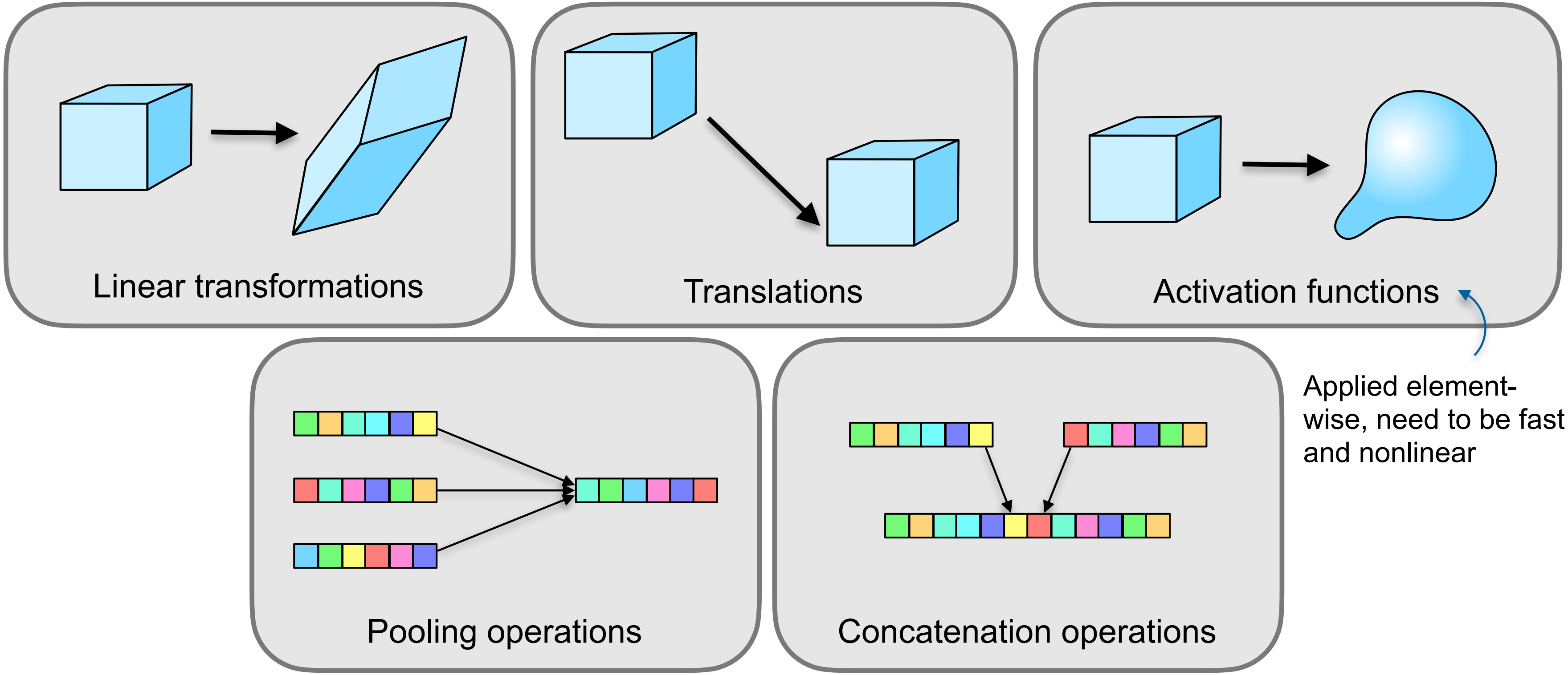
Compose $\underbrace{\text{affine transformations}}_{\text{highly optimized, trainable}}$ with $\underbrace{\text{nonlinear activation functions}}_{\text{simple, fast, fixed}}$

Universal approximation theorem: An FFNN with one hidden layer and enough hidden neurons can approximate any continuous function on a compact domain, to any desired accuracy.

Backpropagation: Method for computing the gradient of an FFNN using the chain rule.



BUILDING BLOCKS*



Linear transformations

Translations

Activation functions

Applied element-wise, need to be fast and nonlinear

Pooling operations

Concatenation operations

* for the neural networks in this talk…

BUILDING A NEURAL QUANTUM STATE

Fermionic wave functions need to be antisymmetric w.r.t. particle exchange

$$\Psi(X) = \underset{\text{symmetric}}{\underset{\nearrow}{e^{J(X)}}}\,\underset{\text{antisymmetric}}{\underset{\nwarrow}{\Phi(X)}}$$

$$X = \{\mathbf{x}_i\}_{i=1}^{N}$$
$$\mathbf{x}_i = (\mathbf{r}_i,\ s_i^z)$$

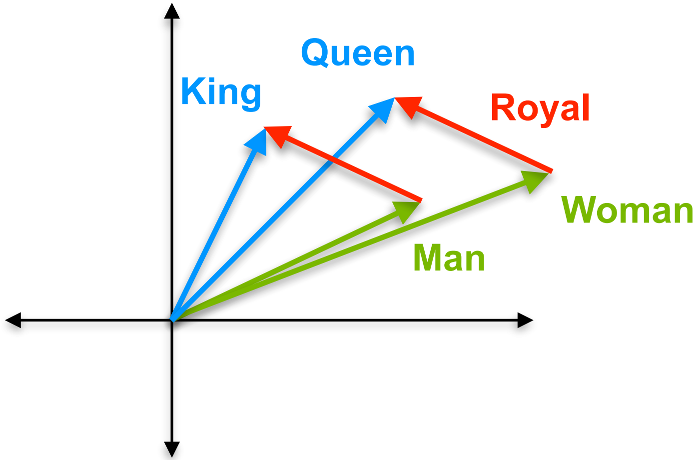
Use a **permutation-invariant** Deep Set for the symmetric part:

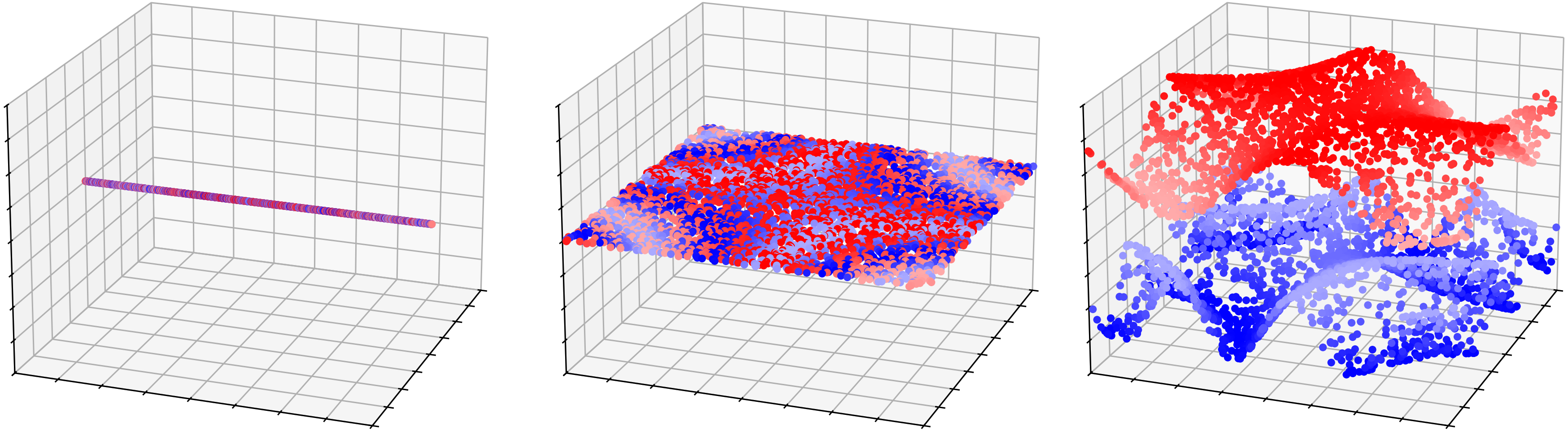
$$J(X) = \rho\left(\text{pool}\left(\{\phi(\mathbf{x}_i)\}_{i=1}^{N}\right)\right)$$

EMBEDDINGS

Learned mapping of inputs to a vector space (usually lower dimensional)

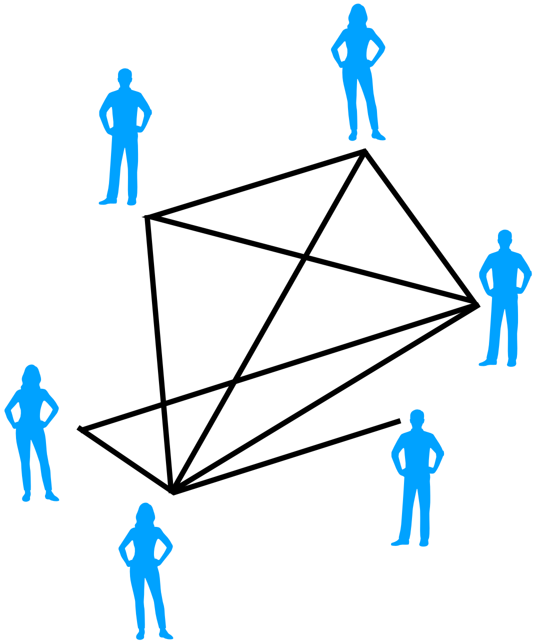
Correlations can be easier to disentangle by temporarily mapping to a higher dimensional space





GRAPH NEURAL NETWORKS

Generalize convolutional neural networks by extending local neighborhood aggregation from regular grids to arbitrary graph structures



How will my friend’s decision to attend my party be influenced by our friends?

Nodes = people (availability, location, age, job, etc…)

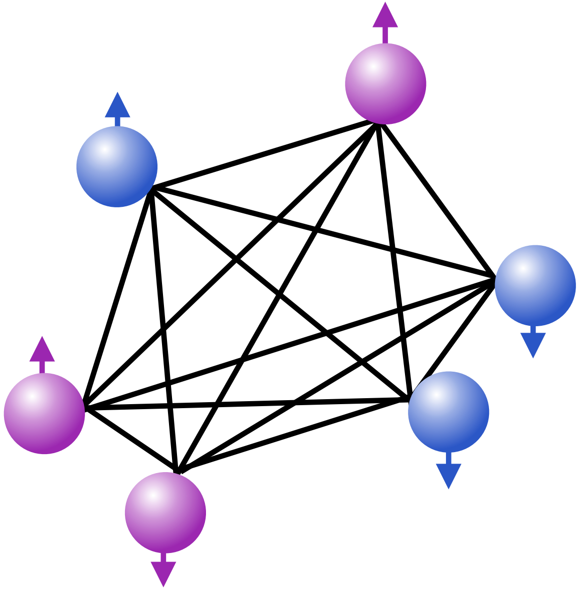
Edges = relationships (closeness, communication frequency, duration of friendship, shared interests…)

Graph neural networks preserve the structure of the graph

Permutation equivariance

GRAPH NEURAL NETWORKS

Generalize convolutional neural networks by extending local neighborhood aggregation from regular grids to arbitrary graph structures



How will a particle's state be influenced by the other particles?

Nodes = particles (spatial coordinates, spin, isospin…)

Edges = relationships (distance, separation vector, spin/isospin alignment, …)

Graph neural networks preserve the structure of the graph

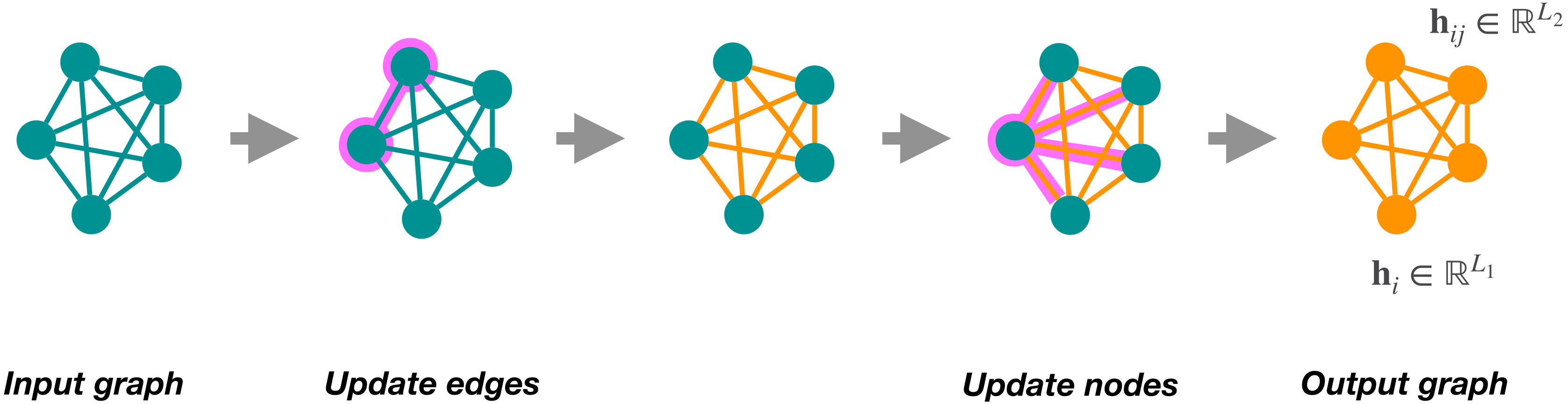
Permutation equivariance

MESSAGE-PASSING NEURAL NETWORKS

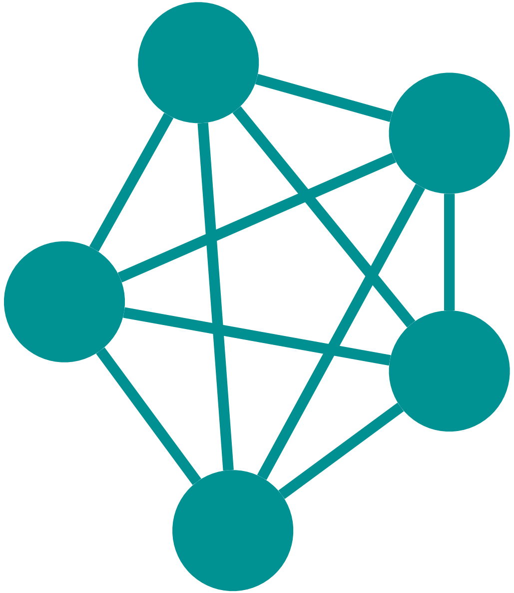
Iteratively builds local correlations by passing “messages” along edges

Used to find better representations of one- and two-body features of quantum system

New embeddings are used in place of the raw features



MESSAGE-PASSING NEURAL NETWORKS



$$\mathbf{h}_i^{(0)} = \mathbf{x}_i = (\mathbf{r}_i,\ s_i^z,\ t_i^z)$$

$$\mathbf{h}_{ij}^{(0)} = \mathbf{x}_{ij} = (\mathbf{r}_{ij},\ \|\mathbf{r}_{ij}\|,\ s_i^z,\ s_j^z,\ t_i^z,\ t_j^z)$$

for $t = 1, \ldots, T$:

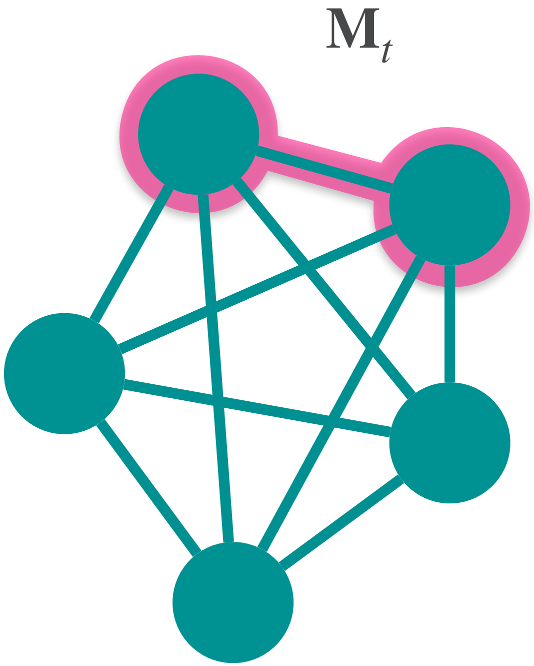
$$\mathbf{m}_{ij}^{(t)} = \mathbf{M}_t\left(\mathbf{h}_i^{(t-1)},\ \mathbf{h}_j^{(t-1)},\ \mathbf{h}_{ij}^{(t-1)}\right)$$

$$\mathbf{h}_{ij}^{(t)} = \left(\mathbf{x}_{ij},\ \mathbf{G}_t\left(\mathbf{h}_{ij}^{(t-1)}, \mathbf{m}_{ij}^{(t)}\right)\right)$$

$$\mathbf{m}_i^{(t)} = \text{pool}\left(\{\mathbf{m}_{ij}^{(t)} \,|\, j \neq i\}\right)$$

$$\mathbf{h}_i^{(t)} = \left(\mathbf{x}_i,\ \mathbf{F}_t\left(\mathbf{h}_i^{(t-1)}, \mathbf{m}_i^{(t)}\right)\right)$$

MESSAGE-PASSING NEURAL NETWORKS



$$\mathbf{h}_i^{(0)} = \mathbf{x}_i = (\mathbf{r}_i,\ s_i^z,\ t_i^z)$$

$$\mathbf{h}_{ij}^{(0)} = \mathbf{x}_{ij} = (\mathbf{r}_{ij},\ \|\mathbf{r}_{ij}\|,\ s_i^z,\ s_j^z,\ t_i^z,\ t_j^z)$$

for $t = 1, \ldots, T$:

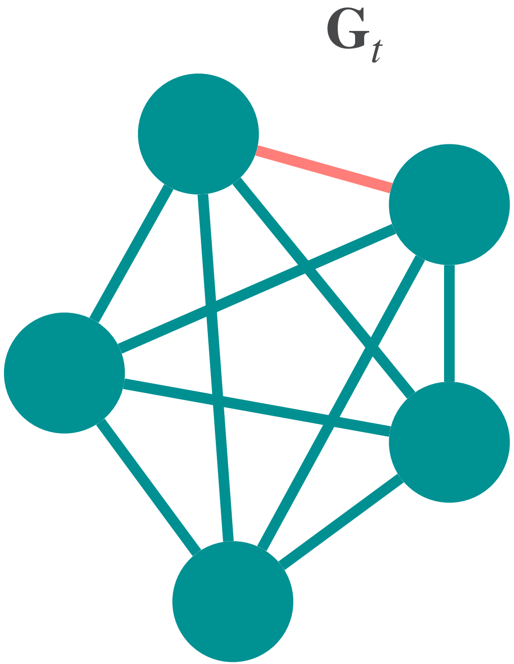
$$\mathbf{m}_{ij}^{(t)} = \mathbf{M}_t\left(\mathbf{h}_i^{(t-1)},\ \mathbf{h}_j^{(t-1)},\ \mathbf{h}_{ij}^{(t-1)}\right)$$

$$\mathbf{h}_{ij}^{(t)} = \left(\mathbf{x}_{ij},\ \mathbf{G}_t\left(\mathbf{h}_{ij}^{(t-1)}, \mathbf{m}_{ij}^{(t)}\right)\right)$$

$$\mathbf{m}_i^{(t)} = \text{pool}\left(\{\mathbf{m}_{ij}^{(t)} \,|\, j \neq i\}\right)$$

$$\mathbf{h}_i^{(t)} = \left(\mathbf{x}_i,\ \mathbf{F}_t\left(\mathbf{h}_i^{(t-1)}, \mathbf{m}_i^{(t)}\right)\right)$$

MESSAGE-PASSING NEURAL NETWORKS



$$\mathbf{h}_i^{(0)} = \mathbf{x}_i = (\mathbf{r}_i,\ s_i^z,\ t_i^z)$$

$$\mathbf{h}_{ij}^{(0)} = \mathbf{x}_{ij} = (\mathbf{r}_{ij},\ \|\mathbf{r}_{ij}\|,\ s_i^z,\ s_j^z,\ t_i^z,\ t_j^z)$$

for $t = 1, \ldots, T$:

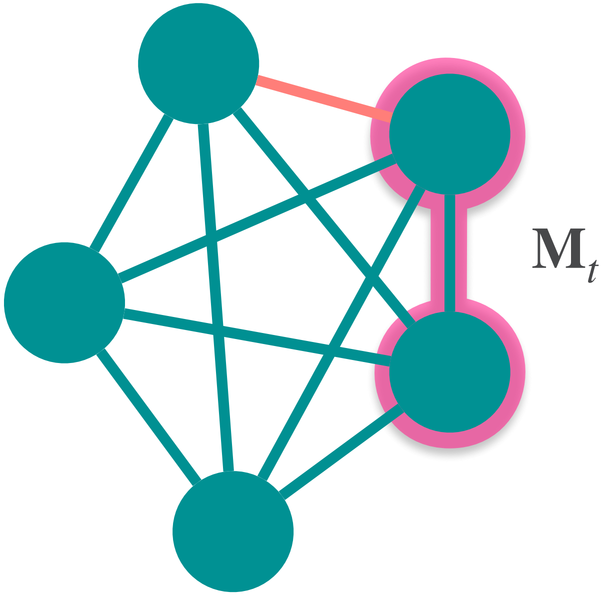
$$\mathbf{m}_{ij}^{(t)} = \mathbf{M}_t\left(\mathbf{h}_i^{(t-1)},\ \mathbf{h}_j^{(t-1)},\ \mathbf{h}_{ij}^{(t-1)}\right)$$

$$\mathbf{h}_{ij}^{(t)} = \left(\mathbf{x}_{ij},\ \mathbf{G}_t\left(\mathbf{h}_{ij}^{(t-1)}, \mathbf{m}_{ij}^{(t)}\right)\right)$$

$$\mathbf{m}_i^{(t)} = \mathrm{pool}\left(\{\mathbf{m}_{ij}^{(t)} \,|\, j \neq i\}\right)$$

$$\mathbf{h}_i^{(t)} = \left(\mathbf{x}_i,\ \mathbf{F}_t\left(\mathbf{h}_i^{(t-1)}, \mathbf{m}_i^{(t)}\right)\right)$$

MESSAGE-PASSING NEURAL NETWORKS



$$\mathbf{h}_i^{(0)} = \mathbf{x}_i = (\mathbf{r}_i,\ s_i^z,\ t_i^z)$$

$$\mathbf{h}_{ij}^{(0)} = \mathbf{x}_{ij} = (\mathbf{r}_{ij},\ \|\mathbf{r}_{ij}\|,\ s_i^z,\ s_j^z,\ t_i^z,\ t_j^z)$$

for $t = 1, \ldots, T$:

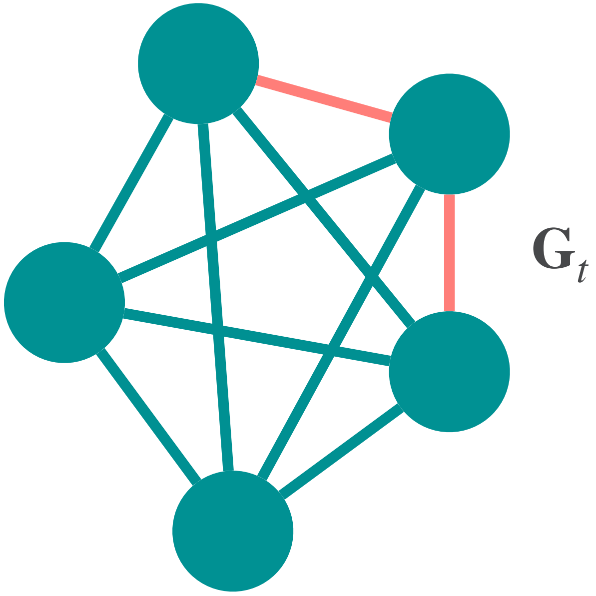
$$\mathbf{m}_{ij}^{(t)} = \mathbf{M}_t\left(\mathbf{h}_i^{(t-1)},\ \mathbf{h}_j^{(t-1)},\ \mathbf{h}_{ij}^{(t-1)}\right)$$

$$\mathbf{h}_{ij}^{(t)} = \left(\mathbf{x}_{ij},\ \mathbf{G}_t\left(\mathbf{h}_{ij}^{(t-1)}, \mathbf{m}_{ij}^{(t)}\right)\right)$$

$$\mathbf{m}_i^{(t)} = \text{pool}\left(\{\mathbf{m}_{ij}^{(t)} \mid j \neq i\}\right)$$

$$\mathbf{h}_i^{(t)} = \left(\mathbf{x}_i,\ \mathbf{F}_t\left(\mathbf{h}_i^{(t-1)}, \mathbf{m}_i^{(t)}\right)\right)$$

MESSAGE-PASSING NEURAL NETWORKS



$$\mathbf{h}_i^{(0)} = \mathbf{x}_i = (\mathbf{r}_i,\ s_i^z,\ t_i^z)$$

$$\mathbf{h}_{ij}^{(0)} = \mathbf{x}_{ij} = (\mathbf{r}_{ij},\ \|\mathbf{r}_{ij}\|,\ s_i^z,\ s_j^z,\ t_i^z,\ t_j^z)$$

for $t = 1, \ldots, T$:

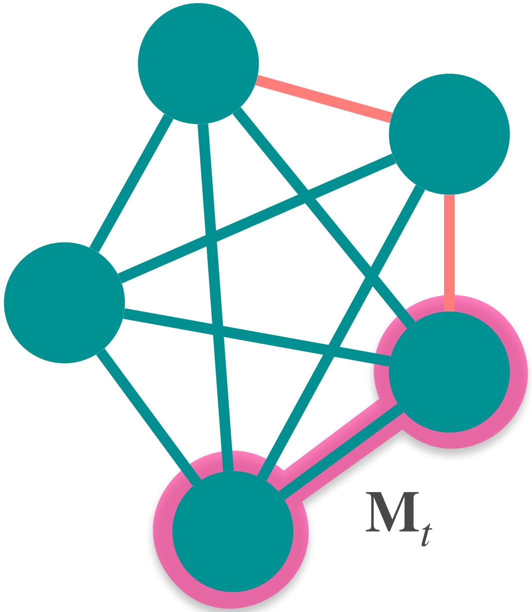
$$\mathbf{m}_{ij}^{(t)} = \mathbf{M}_t\left(\mathbf{h}_i^{(t-1)}, \mathbf{h}_j^{(t-1)}, \mathbf{h}_{ij}^{(t-1)}\right)$$

$$\mathbf{h}_{ij}^{(t)} = \left(\mathbf{x}_{ij}, \mathbf{G}_t\left(\mathbf{h}_{ij}^{(t-1)}, \mathbf{m}_{ij}^{(t)}\right)\right)$$

$$\mathbf{m}_i^{(t)} = \mathrm{pool}\left(\{\mathbf{m}_{ij}^{(t)} \,|\, j \neq i\}\right)$$

$$\mathbf{h}_i^{(t)} = \left(\mathbf{x}_i, \mathbf{F}_t\left(\mathbf{h}_i^{(t-1)}, \mathbf{m}_i^{(t)}\right)\right)$$

MESSAGE-PASSING NEURAL NETWORKS



$$\mathbf{h}_i^{(0)} = \mathbf{x}_i = (\mathbf{r}_i,\ s_i^z,\ t_i^z)$$

$$\mathbf{h}_{ij}^{(0)} = \mathbf{x}_{ij} = (\mathbf{r}_{ij},\ \|\mathbf{r}_{ij}\|,\ s_i^z,\ s_j^z,\ t_i^z,\ t_j^z)$$

for $t = 1, \ldots, T$:

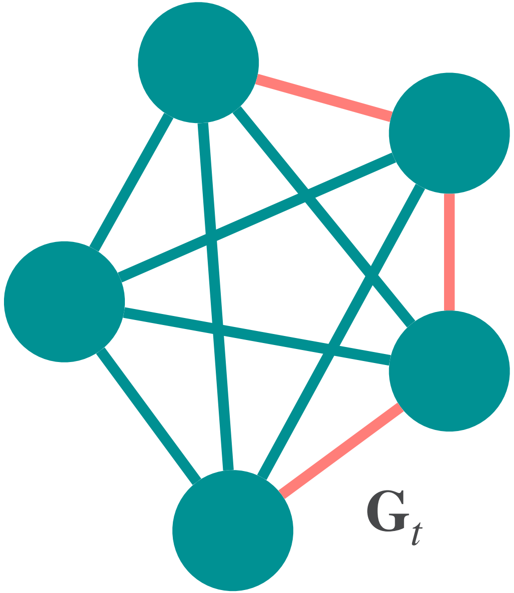
$$\mathbf{m}_{ij}^{(t)} = \mathbf{M}_t\left(\mathbf{h}_i^{(t-1)},\ \mathbf{h}_j^{(t-1)},\ \mathbf{h}_{ij}^{(t-1)}\right)$$

$$\mathbf{h}_{ij}^{(t)} = \left(\mathbf{x}_{ij},\ \mathbf{G}_t\left(\mathbf{h}_{ij}^{(t-1)}, \mathbf{m}_{ij}^{(t)}\right)\right)$$

$$\mathbf{m}_i^{(t)} = \mathrm{pool}\left(\{\mathbf{m}_{ij}^{(t)} \,|\, j \neq i\}\right)$$

$$\mathbf{h}_i^{(t)} = \left(\mathbf{x}_i,\ \mathbf{F}_t\left(\mathbf{h}_i^{(t-1)}, \mathbf{m}_i^{(t)}\right)\right)$$

MESSAGE-PASSING NEURAL NETWORKS



$$\mathbf{h}_i^{(0)} = \mathbf{x}_i = (\mathbf{r}_i,\ s_i^z,\ t_i^z)$$

$$\mathbf{h}_{ij}^{(0)} = \mathbf{x}_{ij} = (\mathbf{r}_{ij},\ \|\mathbf{r}_{ij}\|,\ s_i^z,\ s_j^z,\ t_i^z,\ t_j^z)$$

for $t = 1, \ldots, T$:

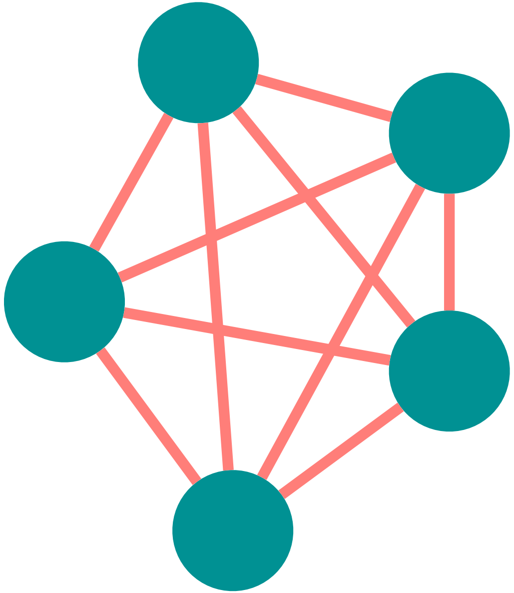
$$\mathbf{m}_{ij}^{(t)} = \mathbf{M}_t\left(\mathbf{h}_i^{(t-1)},\ \mathbf{h}_j^{(t-1)},\ \mathbf{h}_{ij}^{(t-1)}\right)$$

$$\mathbf{h}_{ij}^{(t)} = \left(\mathbf{x}_{ij},\ \mathbf{G}_t\left(\mathbf{h}_{ij}^{(t-1)}, \mathbf{m}_{ij}^{(t)}\right)\right)$$

$$\mathbf{m}_i^{(t)} = \text{pool}\left(\{\mathbf{m}_{ij}^{(t)} \,|\, j \neq i\}\right)$$

$$\mathbf{h}_i^{(t)} = \left(\mathbf{x}_i,\ \mathbf{F}_t\left(\mathbf{h}_i^{(t-1)}, \mathbf{m}_i^{(t)}\right)\right)$$

MESSAGE-PASSING NEURAL NETWORKS



$$\mathbf{h}_i^{(0)} = \mathbf{x}_i = (\mathbf{r}_i,\ s_i^z,\ t_i^z)$$

$$\mathbf{h}_{ij}^{(0)} = \mathbf{x}_{ij} = (\mathbf{r}_{ij},\ \|\mathbf{r}_{ij}\|,\ s_i^z,\ s_j^z,\ t_i^z,\ t_j^z)$$

for $t = 1, \ldots, T:$

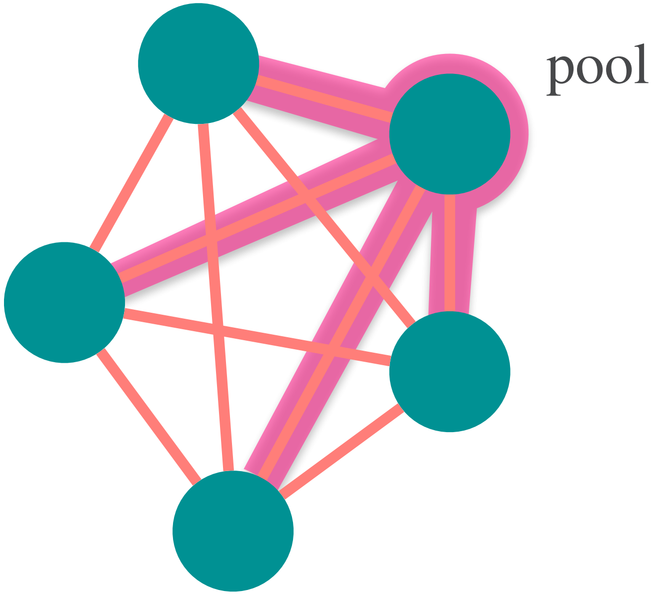
$$\mathbf{m}_{ij}^{(t)} = \mathbf{M}_t\left(\mathbf{h}_i^{(t-1)},\ \mathbf{h}_j^{(t-1)},\ \mathbf{h}_{ij}^{(t-1)}\right)$$

$$\mathbf{h}_{ij}^{(t)} = \left(\mathbf{x}_{ij},\ \mathbf{G}_t\left(\mathbf{h}_{ij}^{(t-1)}, \mathbf{m}_{ij}^{(t)}\right)\right)$$

$$\mathbf{m}_i^{(t)} = \mathrm{pool}\left(\{\mathbf{m}_{ij}^{(t)} \,|\, j \neq i\}\right)$$

$$\mathbf{h}_i^{(t)} = \left(\mathbf{x}_i,\ \mathbf{F}_t\left(\mathbf{h}_i^{(t-1)}, \mathbf{m}_i^{(t)}\right)\right)$$

MESSAGE-PASSING NEURAL NETWORKS



$$\mathbf{h}_i^{(0)} = \mathbf{x}_i = (\mathbf{r}_i,\ s_i^z,\ t_i^z)$$

$$\mathbf{h}_{ij}^{(0)} = \mathbf{x}_{ij} = (\mathbf{r}_{ij},\ \|\mathbf{r}_{ij}\|,\ s_i^z,\ s_j^z,\ t_i^z,\ t_j^z)$$

for $t = 1, \ldots, T$:

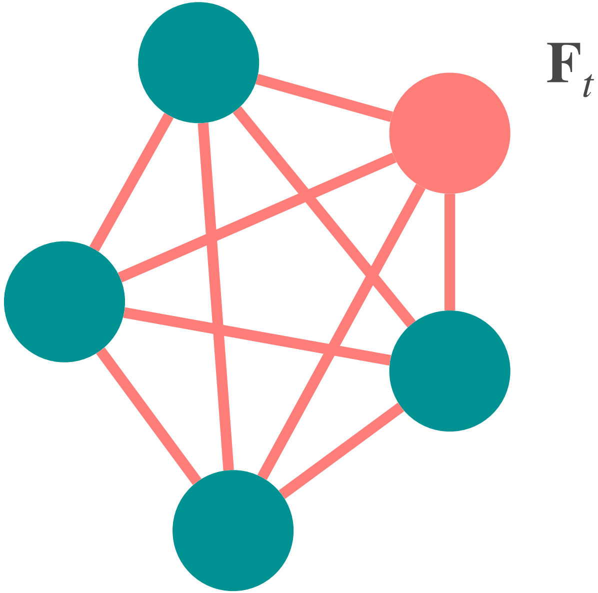
$$\mathbf{m}_{ij}^{(t)} = \mathbf{M}_t\left(\mathbf{h}_i^{(t-1)},\ \mathbf{h}_j^{(t-1)},\ \mathbf{h}_{ij}^{(t-1)}\right)$$

$$\mathbf{h}_{ij}^{(t)} = \left(\mathbf{x}_{ij},\ \mathbf{G}_t\left(\mathbf{h}_{ij}^{(t-1)}, \mathbf{m}_{ij}^{(t)}\right)\right)$$

$$\mathbf{m}_i^{(t)} = \mathrm{pool}\left(\{\mathbf{m}_{ij}^{(t)} \,|\, j \neq i\}\right)$$

$$\mathbf{h}_i^{(t)} = \left(\mathbf{x}_i,\ \mathbf{F}_t\left(\mathbf{h}_i^{(t-1)}, \mathbf{m}_i^{(t)}\right)\right)$$

MESSAGE-PASSING NEURAL NETWORKS



$$\mathbf{h}_i^{(0)} = \mathbf{x}_i = (\mathbf{r}_i,\ s_i^z,\ t_i^z)$$

$$\mathbf{h}_{ij}^{(0)} = \mathbf{x}_{ij} = (\mathbf{r}_{ij},\ \|\mathbf{r}_{ij}\|,\ s_i^z,\ s_j^z,\ t_i^z,\ t_j^z)$$

for $t = 1, \ldots, T$:

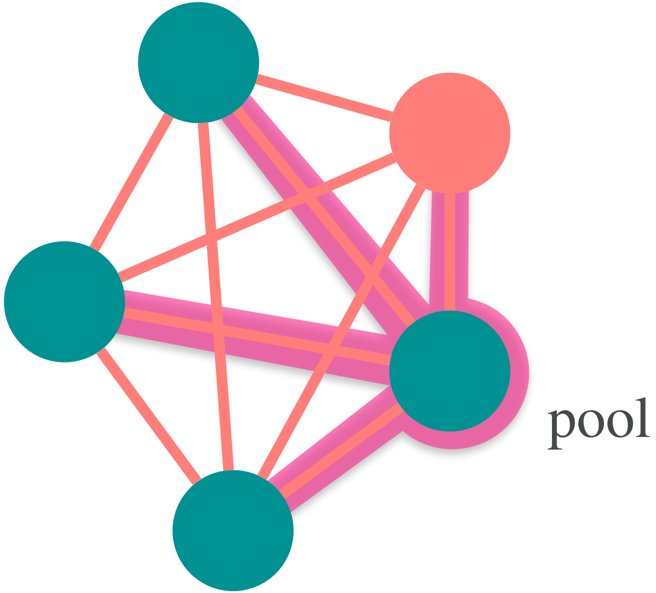
$$\mathbf{m}_{ij}^{(t)} = \mathbf{M}_t\left(\mathbf{h}_i^{(t-1)},\ \mathbf{h}_j^{(t-1)},\ \mathbf{h}_{ij}^{(t-1)}\right)$$

$$\mathbf{h}_{ij}^{(t)} = \left(\mathbf{x}_{ij},\ \mathbf{G}_t\left(\mathbf{h}_{ij}^{(t-1)}, \mathbf{m}_{ij}^{(t)}\right)\right)$$

$$\mathbf{m}_i^{(t)} = \mathrm{pool}\left(\{\mathbf{m}_{ij}^{(t)} \,|\, j \neq i\}\right)$$

$$\mathbf{h}_i^{(t)} = \left(\mathbf{x}_i,\ \mathbf{F}_t\left(\mathbf{h}_i^{(t-1)}, \mathbf{m}_i^{(t)}\right)\right)$$

MESSAGE-PASSING NEURAL NETWORKS



$$\mathbf{h}_i^{(0)} = \mathbf{x}_i = (\mathbf{r}_i,\ s_i^z,\ t_i^z)$$

$$\mathbf{h}_{ij}^{(0)} = \mathbf{x}_{ij} = (\mathbf{r}_{ij},\ \|\mathbf{r}_{ij}\|,\ s_i^z,\ s_j^z,\ t_i^z,\ t_j^z)$$

for $t = 1, \ldots, T$:

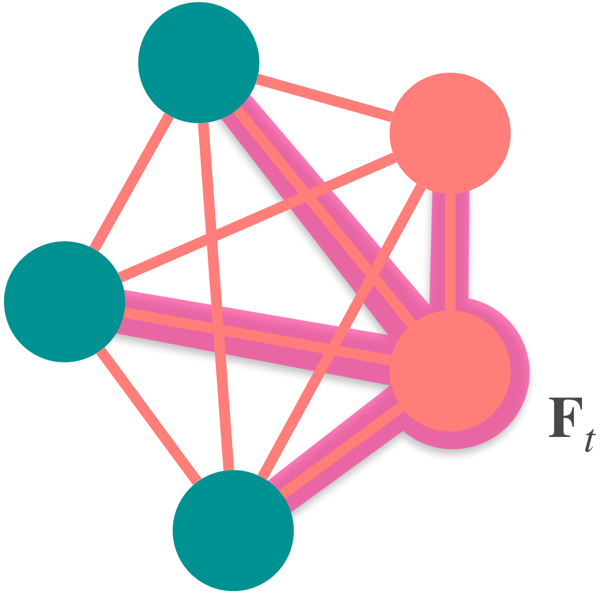
$$\mathbf{m}_{ij}^{(t)} = \mathbf{M}_t\left(\mathbf{h}_i^{(t-1)},\ \mathbf{h}_j^{(t-1)},\ \mathbf{h}_{ij}^{(t-1)}\right)$$

$$\mathbf{h}_{ij}^{(t)} = \left(\mathbf{x}_{ij},\ \mathbf{G}_t\left(\mathbf{h}_{ij}^{(t-1)}, \mathbf{m}_{ij}^{(t)}\right)\right)$$

$$\mathbf{m}_i^{(t)} = \mathrm{pool}\left(\{\mathbf{m}_{ij}^{(t)} \mid j \neq i\}\right)$$

$$\mathbf{h}_i^{(t)} = \left(\mathbf{x}_i,\ \mathbf{F}_t\left(\mathbf{h}_i^{(t-1)}, \mathbf{m}_i^{(t)}\right)\right)$$

MESSAGE-PASSING NEURAL NETWORKS



$$\mathbf{h}_i^{(0)} = \mathbf{x}_i = (\mathbf{r}_i,\ s_i^z,\ t_i^z)$$

$$\mathbf{h}_{ij}^{(0)} = \mathbf{x}_{ij} = (\mathbf{r}_{ij},\ \|\mathbf{r}_{ij}\|,\ s_i^z,\ s_j^z,\ t_i^z,\ t_j^z)$$

for $t = 1, \ldots, T$:

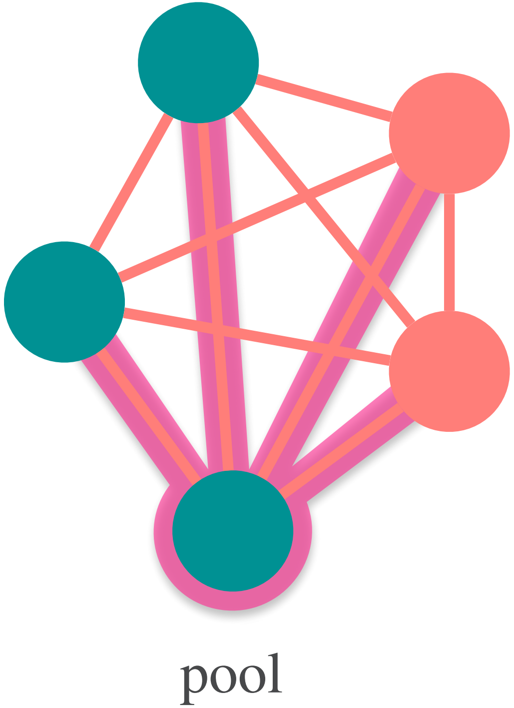
$$\mathbf{m}_{ij}^{(t)} = \mathbf{M}_t\left(\mathbf{h}_i^{(t-1)},\ \mathbf{h}_j^{(t-1)},\ \mathbf{h}_{ij}^{(t-1)}\right)$$

$$\mathbf{h}_{ij}^{(t)} = \left(\mathbf{x}_{ij},\ \mathbf{G}_t\left(\mathbf{h}_{ij}^{(t-1)}, \mathbf{m}_{ij}^{(t)}\right)\right)$$

$$\mathbf{m}_i^{(t)} = \mathrm{pool}\left(\{\mathbf{m}_{ij}^{(t)} \,|\, j \neq i\}\right)$$

$$\mathbf{h}_i^{(t)} = \left(\mathbf{x}_i,\ \mathbf{F}_t\left(\mathbf{h}_i^{(t-1)}, \mathbf{m}_i^{(t)}\right)\right)$$

MESSAGE-PASSING NEURAL NETWORKS



$$\mathbf{h}_i^{(0)} = \mathbf{x}_i = (\mathbf{r}_i,\ s_i^z,\ t_i^z)$$

$$\mathbf{h}_{ij}^{(0)} = \mathbf{x}_{ij} = (\mathbf{r}_{ij},\ \|\mathbf{r}_{ij}\|,\ s_i^z,\ s_j^z,\ t_i^z,\ t_j^z)$$

for $t = 1, \ldots, T$:

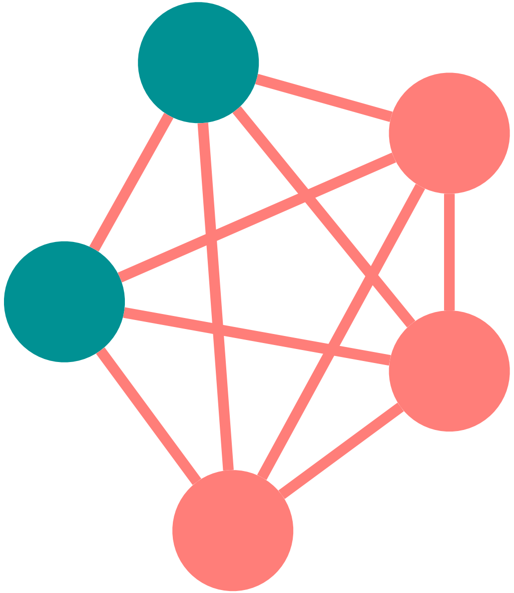
$$\mathbf{m}_{ij}^{(t)} = \mathbf{M}_t\left(\mathbf{h}_i^{(t-1)},\ \mathbf{h}_j^{(t-1)},\ \mathbf{h}_{ij}^{(t-1)}\right)$$

$$\mathbf{h}_{ij}^{(t)} = \left(\mathbf{x}_{ij},\ \mathbf{G}_t\left(\mathbf{h}_{ij}^{(t-1)}, \mathbf{m}_{ij}^{(t)}\right)\right)$$

$$\mathbf{m}_i^{(t)} = \mathrm{pool}\left(\{\mathbf{m}_{ij}^{(t)} \,|\, j \neq i\}\right)$$

$$\mathbf{h}_i^{(t)} = \left(\mathbf{x}_i,\ \mathbf{F}_t\left(\mathbf{h}_i^{(t-1)}, \mathbf{m}_i^{(t)}\right)\right)$$

MESSAGE-PASSING NEURAL NETWORKS



$\mathbf{F}_t$

$$\mathbf{h}_i^{(0)} = \mathbf{x}_i = (\mathbf{r}_i,\ s_i^z,\ t_i^z)$$

$$\mathbf{h}_{ij}^{(0)} = \mathbf{x}_{ij} = (\mathbf{r}_{ij},\ \|\mathbf{r}_{ij}\|,\ s_i^z,\ s_j^z,\ t_i^z,\ t_j^z)$$

for $t = 1, \ldots, T$:

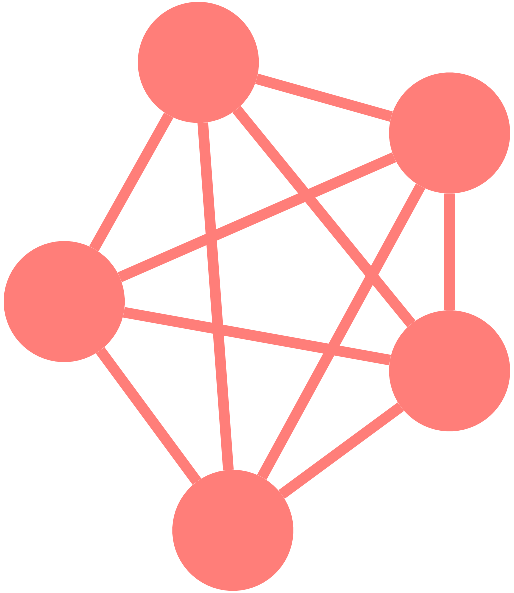
$$\mathbf{m}_{ij}^{(t)} = \mathbf{M}_t\left(\mathbf{h}_i^{(t-1)},\ \mathbf{h}_j^{(t-1)},\ \mathbf{h}_{ij}^{(t-1)}\right)$$

$$\mathbf{h}_{ij}^{(t)} = \left(\mathbf{x}_{ij},\ \mathbf{G}_t\left(\mathbf{h}_{ij}^{(t-1)}, \mathbf{m}_{ij}^{(t)}\right)\right)$$

$$\mathbf{m}_i^{(t)} = \text{pool}\left(\{\mathbf{m}_{ij}^{(t)} \,|\, j \neq i\}\right)$$

$$\mathbf{h}_i^{(t)} = \left(\mathbf{x}_i,\ \mathbf{F}_t\left(\mathbf{h}_i^{(t-1)}, \mathbf{m}_i^{(t)}\right)\right)$$

MESSAGE-PASSING NEURAL NETWORKS



$$\mathbf{h}_i^{(0)} = \mathbf{x}_i = (\mathbf{r}_i,\ s_i^z,\ t_i^z)$$

$$\mathbf{h}_{ij}^{(0)} = \mathbf{x}_{ij} = (\mathbf{r}_{ij},\ \|\mathbf{r}_{ij}\|,\ s_i^z,\ s_j^z,\ t_i^z,\ t_j^z)$$

for $t = 1, \ldots, T$:

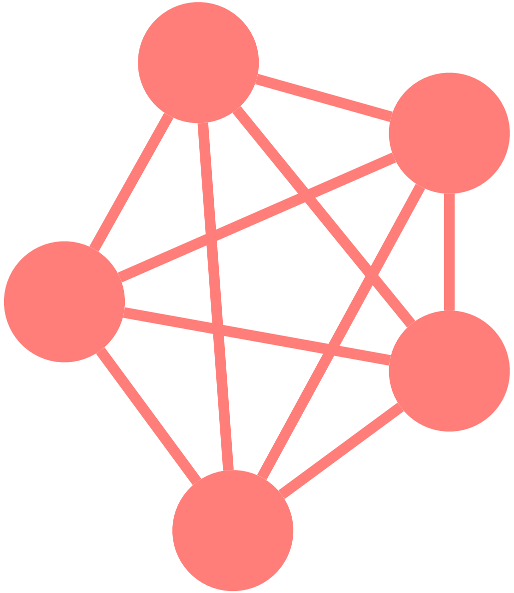
$$\mathbf{m}_{ij}^{(t)} = \mathbf{M}_t\left(\mathbf{h}_i^{(t-1)},\ \mathbf{h}_j^{(t-1)},\ \mathbf{h}_{ij}^{(t-1)}\right)$$

$$\mathbf{h}_{ij}^{(t)} = \left(\mathbf{x}_{ij},\ \mathbf{G}_t\left(\mathbf{h}_{ij}^{(t-1)}, \mathbf{m}_{ij}^{(t)}\right)\right)$$

$$\mathbf{m}_i^{(t)} = \mathrm{pool}\left(\{\mathbf{m}_{ij}^{(t)} \,|\, j \neq i\}\right)$$

$$\mathbf{h}_i^{(t)} = \left(\mathbf{x}_i,\ \mathbf{F}_t\left(\mathbf{h}_i^{(t-1)}, \mathbf{m}_i^{(t)}\right)\right)$$

MESSAGE-PASSING NEURAL NETWORKS



$$\mathbf{h}_i^{(0)} = \mathbf{x}_i = (\mathbf{r}_i,\ s_i^z,\ t_i^z)$$

$$\mathbf{h}_{ij}^{(0)} = \mathbf{x}_{ij} = (\mathbf{r}_{ij},\ \|\mathbf{r}_{ij}\|,\ s_i^z,\ s_j^z,\ t_i^z,\ t_j^z)$$

for $t = 1, \ldots, T$:

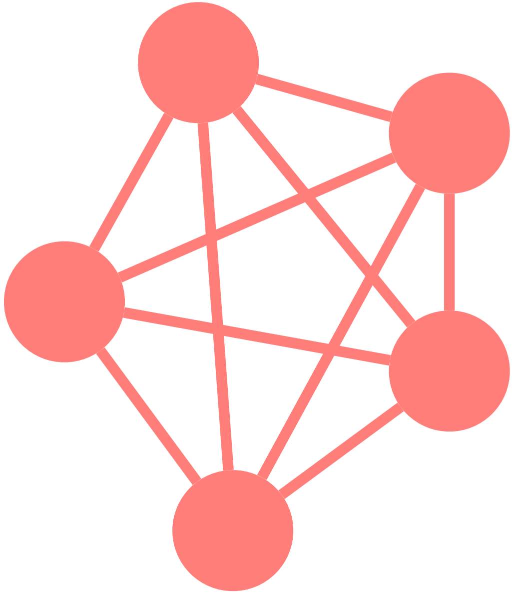
$$\mathbf{m}_{ij}^{(t)} = \mathbf{M}_t\left(\mathbf{h}_i^{(t-1)},\ \mathbf{h}_j^{(t-1)},\ \mathbf{h}_{ij}^{(t-1)}\right)$$

$$\mathbf{h}_{ij}^{(t)} = \left(\mathbf{x}_{ij}, \mathbf{G}_t\left(\mathbf{h}_{ij}^{(t-1)}, \mathbf{m}_{ij}^{(t)}\right)\right)$$

$$\mathbf{m}_i^{(t)} = \mathrm{pool}\left(\{\mathbf{m}_{ij}^{(t)} \,|\, j \neq i\}\right)$$

$$\mathbf{h}_i^{(t)} = \left(\mathbf{x}_i, \mathbf{F}_t\left(\mathbf{h}_i^{(t-1)}, \mathbf{m}_i^{(t)}\right)\right)$$

MESSAGE-PASSING NEURAL NETWORKS



$$\mathbf{h}_i^{(0)} = \mathbf{x}_i = (\mathbf{r}_i,\ s_i^z,\ t_i^z)$$

$$\mathbf{h}_{ij}^{(0)} = \mathbf{x}_{ij} = (\mathbf{r}_{ij},\ \|\mathbf{r}_{ij}\|,\ s_i^z,\ s_j^z,\ t_i^z,\ t_j^z)$$

for $t = 1, \ldots, T$:

$$\mathbf{m}_{ij}^{(t)} = \mathbf{M}_t\left(\mathbf{h}_i^{(t-1)}, \mathbf{h}_j^{(t-1)}, \mathbf{h}_{ij}^{(t-1)}\right)$$

$$\mathbf{h}_{ij}^{(t)} = \left(\underline{\mathbf{x}_{ij},\ \mathbf{G}_t\left(\mathbf{h}_{ij}^{(t-1)}, \mathbf{m}_{ij}^{(t)}\right)}\right)$$

$$\mathbf{m}_i^{(t)} = \mathrm{pool}\left(\{\mathbf{m}_{ij}^{(t)} \,|\, j \neq i\}\right)$$

$$\mathbf{h}_i^{(t)} = \left(\underline{\mathbf{x}_i,\ \mathbf{F}_t\left(\mathbf{h}_i^{(t-1)}, \mathbf{m}_i^{(t)}\right)}\right)$$

BUILDING A NEURAL QUANTUM STATE

Fermionic wave functions need to be antisymmetric w.r.t. particle exchange

$$\Psi(X) = e^{J(X)}\Phi(X)$$

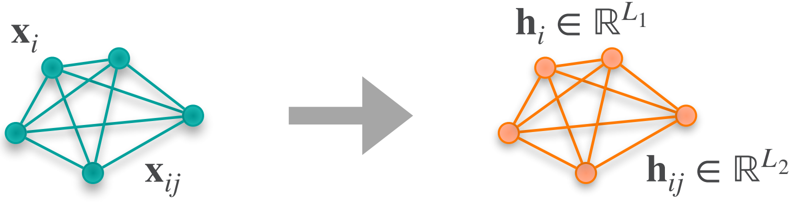
symmetric   antisymmetric

$$X = \{\mathbf{x}_i\}_{i=1}^{N}$$
$$\mathbf{x}_i = (\mathbf{r}_i,\ s_i^z)$$

Use a **permutation-invariant** Deep Set for the symmetric part:

$$J(X) = \rho\left(\text{pool}\left(\{\phi(\mathbf{x}_i)\}_{i=1}^{N}\right)\right)$$

Use a **permutation-equivariant** graph neural network to find a higher-dimensional embedding:



BUILDING A NEURAL QUANTUM STATE

Fermionic wave functions need to be antisymmetric w.r.t. particle exchange

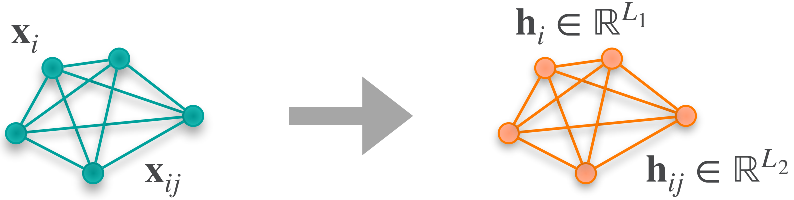
$$\Psi(X) = \underset{\text{symmetric}}{e^{J(X)}}\underset{\text{antisymmetric}}{\Phi(X)}$$

$$X = \{\mathbf{x}_i\}_{i=1}^{N}$$
$$\mathbf{x}_i = (\mathbf{r}_i,\ s_i^z)$$

Use a **permutation-invariant** Deep Set for the symmetric part:

$$J(X) = \rho\left(\mathrm{pool}\left(\{\phi(\cancelto{\mathbf{h}_i}{\mathbf{x}_i})\}_{i=1}^{N}\right)\right)$$

Use a **permutation-equivariant** graph neural network to find a higher-dimensional embedding:



ATTENTION MECHANISMS

First developed for natural language processing

Scores the relevance of each feature w.r.t. the context, then uses those scores to weight the features

One of the key elements of transformer architectures

Example: What’s the meaning of the word **bank**?

A. Vaswani et al., arXiv:1706.03762 (2017).

ATTENTION MECHANISMS

First developed for natural language processing

Scores the relevance of each feature w.r.t. the context, then uses those scores to weight the features

One of the key elements of transformer architectures

Example: What’s the meaning of the word **bank**?

They decided to meet at the **bank**.

It depends on the context!

She withdrew money from the **bank**.

He watched the sunset from the river **bank**.

A. Vaswani et al., arXiv:1706.03762 (2017).

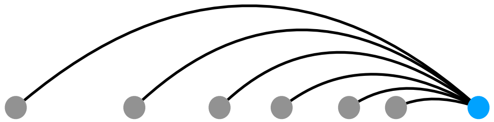
ATTENTION MECHANISMS

First developed for natural language processing

Scores the relevance of each feature w.r.t. the context, then uses those scores to weight the features

One of the key elements of transformer architectures

Example: What’s the meaning of the word **bank**?



They decided to meet at the **bank.**

It depends on the context!

She withdrew money from the **bank.**

He watched the sunset from the river **bank.**

A. Vaswani et al., arXiv:1706.03762 (2017).

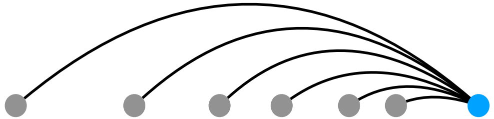
ATTENTION MECHANISMS

First developed for natural language processing

Scores the relevance of each feature w.r.t. the context, then uses those scores to weight the features

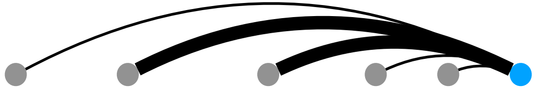
One of the key elements of transformer architectures

Example: What’s the meaning of the word **bank**?



They decided to meet at the **bank**.

It depends on the context!



She withdrew money from the **bank**.

He watched the sunset from the river **bank**.

A. Vaswani et al., arXiv:1706.03762 (2017).

ATTENTION MECHANISMS

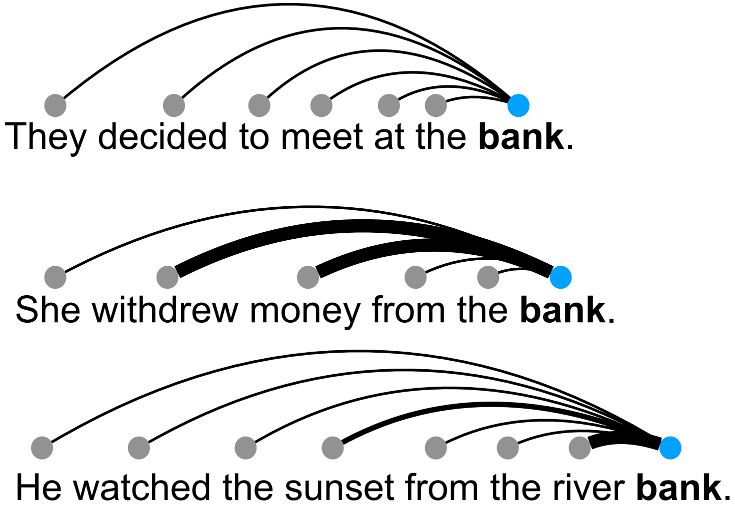
First developed for natural language processing

Scores the relevance of each feature w.r.t. the context, then uses those scores to weight the features

One of the key elements of transformer architectures

Example: What's the meaning of the word **bank**?

It depends on the context!



They decided to meet at the **bank**.

She withdrew money from the **bank**.

He watched the sunset from the river **bank**.

A. Vaswani et al., arXiv:1706.03762 (2017).

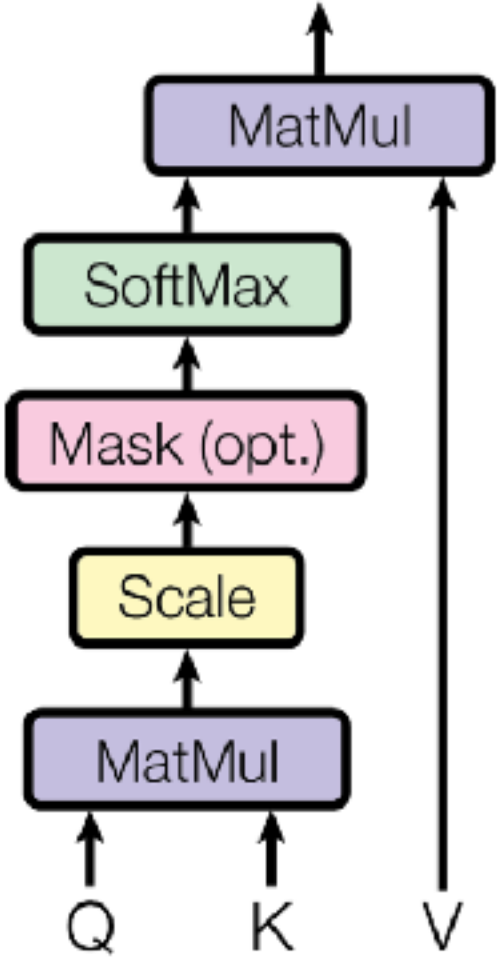
ATTENTION MECHANISMS

Query: What information am looking for? (question) $Q = XW_Q$

Key: How should I recognize the information? (label) $K = XW_K$

Value: What is the information I actually retrieve? (content) $V = XW_V$

$$\text{Attention(Q, K, V)} = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$



A. Vaswani et al., arXiv:1706.03762 (2017).

HOMOGENEOUS ELECTRON GAS

Classic benchmark system

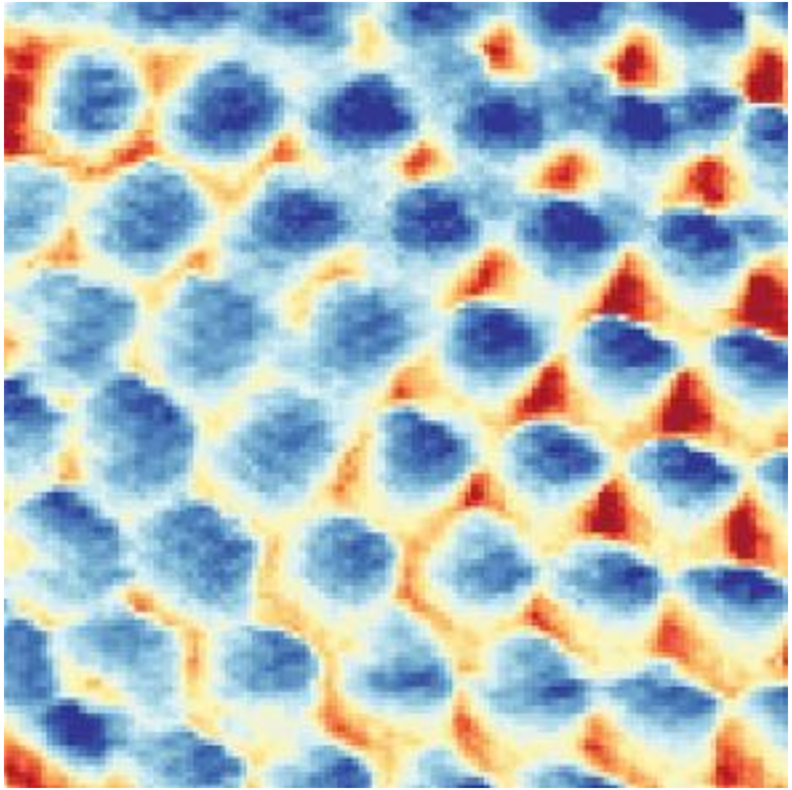
Many-body methods begin to struggle at low densities, where correlations from the long-range Coulomb interaction dominate $\Longrightarrow$ Wigner crystal

In this work, we developed a message-passing neural network with an attention mechanism

Far fewer trainable parameters than other NQS

Scales easily to larger systems

"Pays attention" to relevant features based on neighboring particles



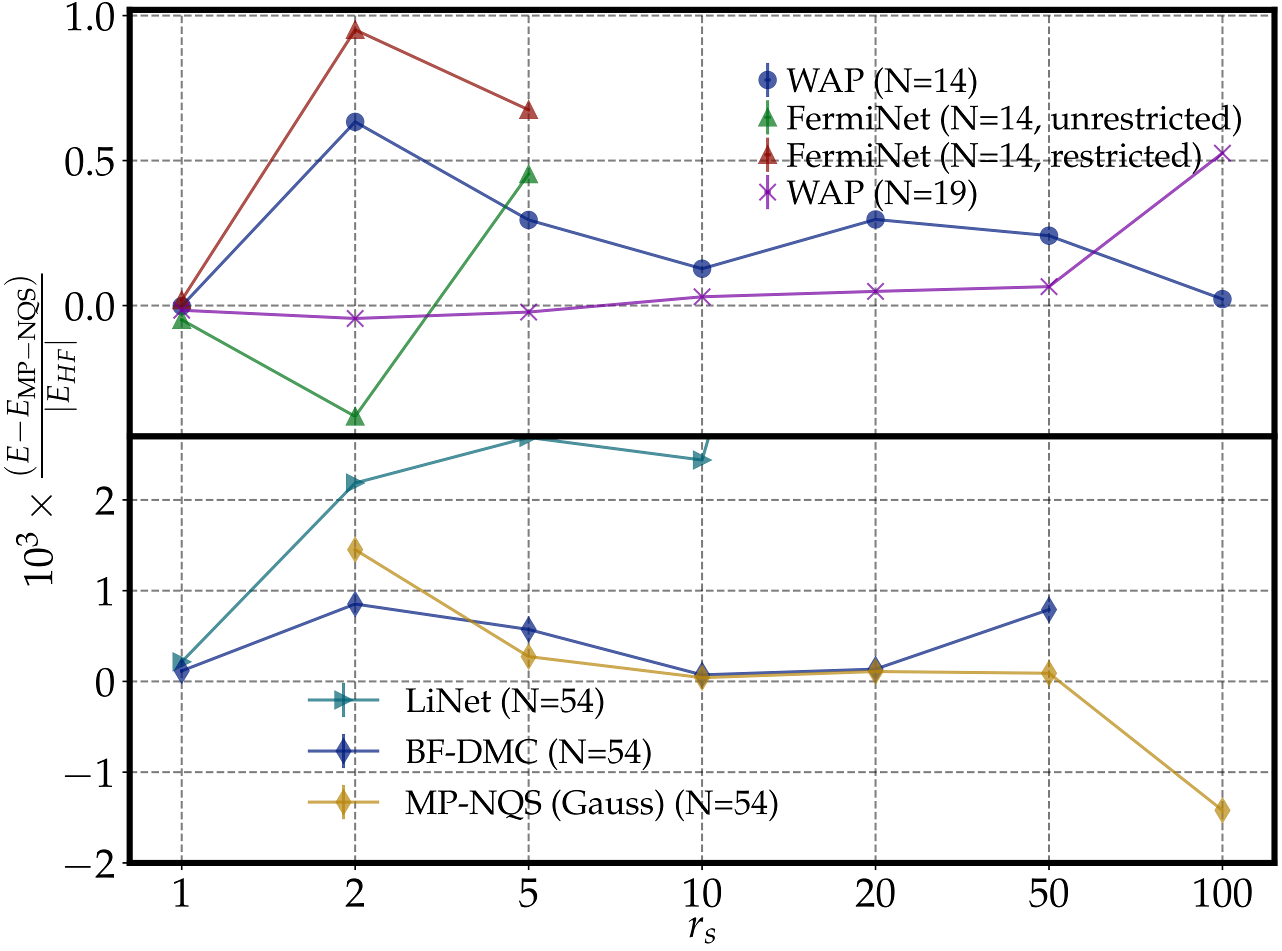
*G. Pescia, JK, et al., PRB **110**, 035108 (2024).*

Yen-Chen Tsui, Princeton University.

$\mathcal{O}(10^4)$ trainable parameters instead of $\mathcal{O}(10^{5-7})$

$\mathcal{O}(10^3)$ training iterations instead of $\mathcal{O}(10^5)$

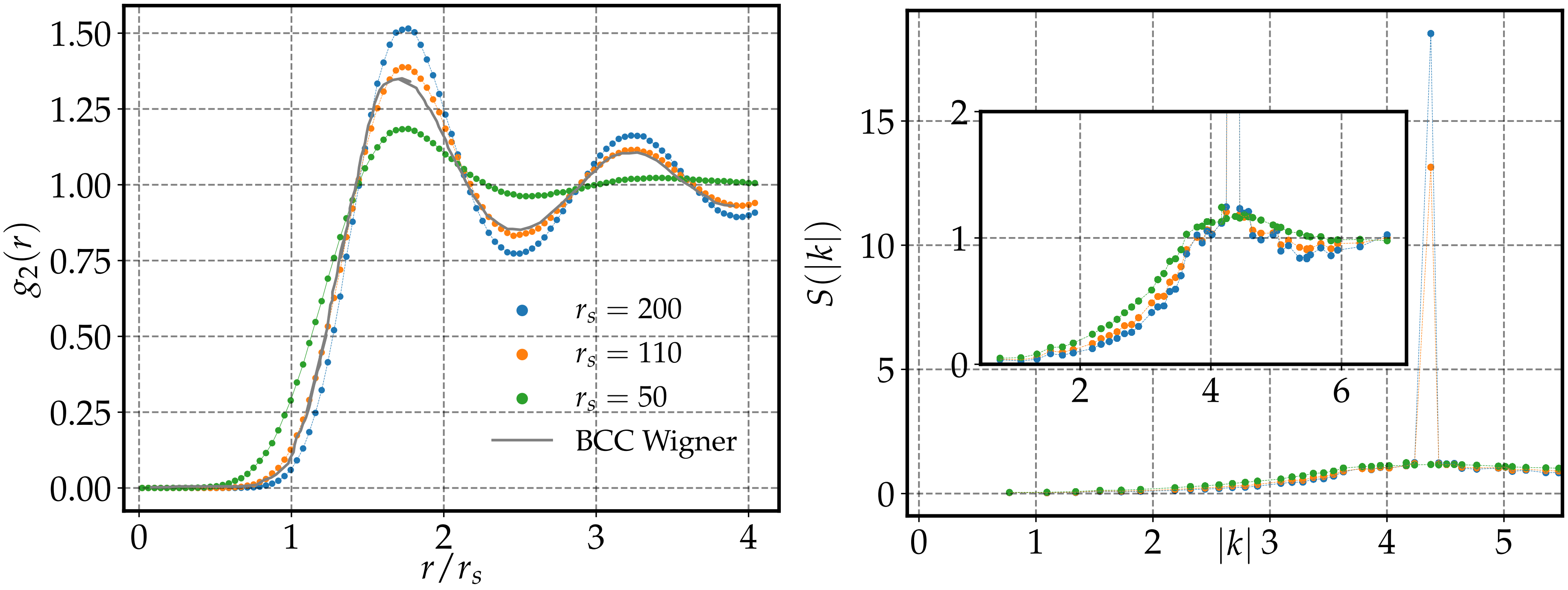
Smaller NQS can use stochastic reconfiguration (2nd order optimization)



G. Pescia, JK, et al., PRB **110**, 035108 (2024).

Efficiency allowed calculations of $N = 128$ in continuous space – largest system tackled by NQS

Evidence of Wigner crystallization appears between $r_s = 50 - 110$



G. Pescia, JK, et al., PRB **110**, 035108 (2024).

BUILDING A NEURAL QUANTUM STATE

Fermionic wave functions need to be antisymmetric w.r.t. particle exchange

$$\Psi(X) = e^{J(X)}\Phi(X)$$

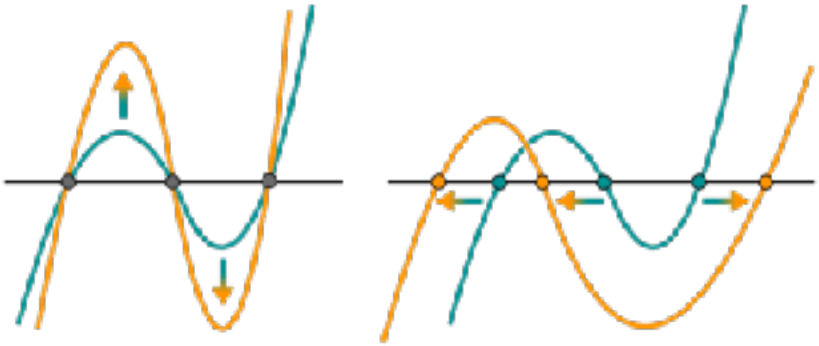
symmetric

antisymmetric

$$X = \{\mathbf{x}_i\}_{i=1}^{N}$$

$$\mathbf{x}_i = (\mathbf{r}_i,\ s_i^z)$$

The antisymmetric part is entirely responsible for the nodal structure of the wave function



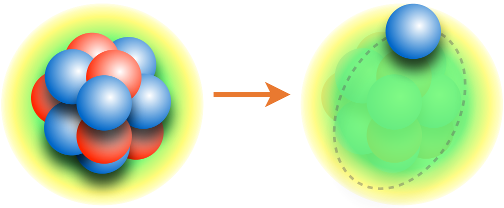
ANTISYMMETRY

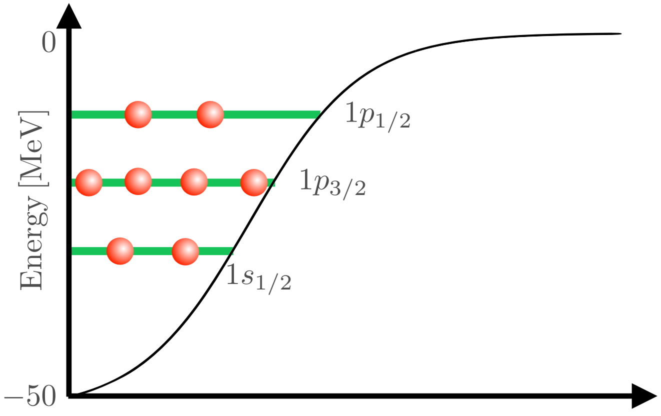
Antisymmetry under particle exchange: $\Phi(\ldots, \mathbf{x}_i, \ldots, \mathbf{x}_j, \ldots) = -\,\Phi(\ldots, \mathbf{x}_j, \ldots, \mathbf{x}_i, \ldots)$

Mean-field approximation: Particles assumed to be independent, subject to an average field

Ground state wave function is an antisymmetrized product of single-particle states

$$\left[\sum_{i<j} v_{ij} + \sum_{i<j<k} V_{ijk}\right] \mapsto \sum_{i} u_i$$





ANTISYMMETRY

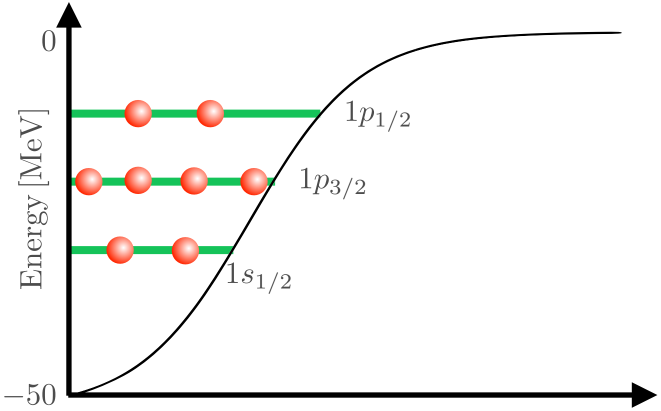
Scales as $\mathscr{O}(N!)$

$$\Phi(X) = \hat{\mathscr{A}}\left[\phi_1(\mathbf{x}_1)\cdots\phi_N(\mathbf{x}_N)\right] = \frac{1}{N!}\sum_{\sigma\in S_N}\operatorname{sgn}(\sigma)\phi_{\sigma(1)}(\mathbf{x}_1)\cdots\phi_{\sigma(N)}(\mathbf{x}_N)$$

Slater determinant: More efficient way of antisymmetrizing a product of single-particle orbitals

$$\Phi(X) = \det\begin{bmatrix}\phi_1(\mathbf{x}_1) & \phi_1(\mathbf{x}_2) & \cdots & \phi_1(\mathbf{x}_N)\\ \phi_2(\mathbf{x}_1) & \phi_2(\mathbf{x}_2) & \cdots & \phi_2(\mathbf{x}_N)\\ \vdots & \vdots & \ddots & \vdots\\ \phi_N(\mathbf{x}_1) & \phi_N(\mathbf{x}_2) & \cdots & \phi_N(\mathbf{x}_N)\end{bmatrix}$$

Scales as $\mathscr{O}(N^3)$



NEURAL SLATER DETERMINANTS

The antisymmetric part of a NQS is usually based on a Slater determinant:

$$\Phi(X) = \det \begin{bmatrix} \phi_1(\mathbf{x}_1) & \phi_1(\mathbf{x}_2) & \cdots & \phi_1(\mathbf{x}_N) \\ \phi_2(\mathbf{x}_1) & \phi_2(\mathbf{x}_2) & \cdots & \phi_2(\mathbf{x}_N) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_N(\mathbf{x}_1) & \phi_N(\mathbf{x}_2) & \cdots & \phi_N(\mathbf{x}_N) \end{bmatrix}$$

Replace with neural networks

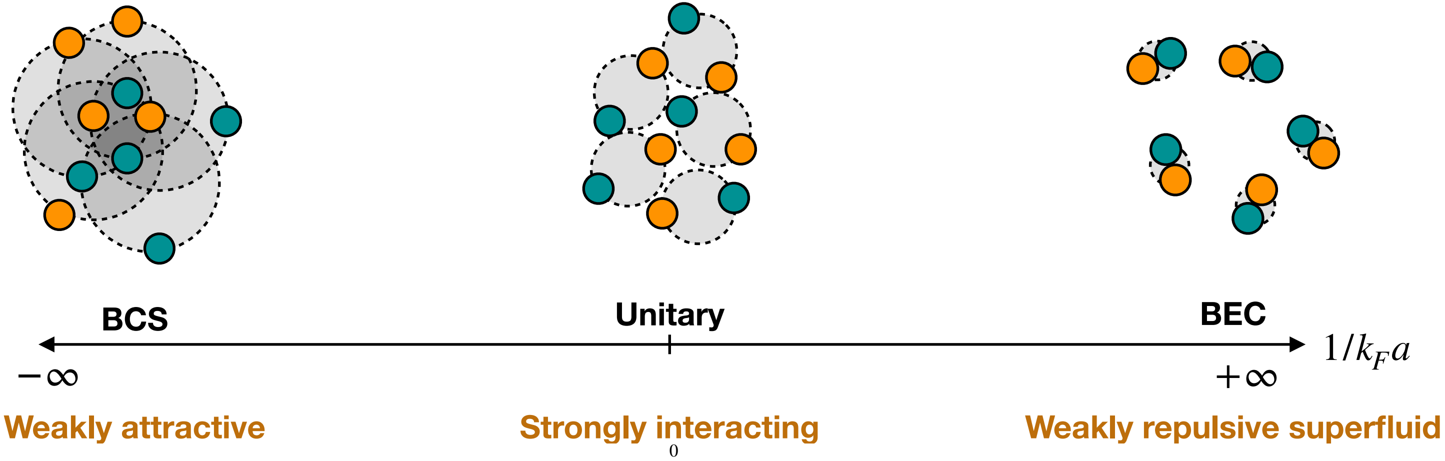
This is insufficient for strongly-correlated fermions…

ULTRACOLD FERMI GASES

Characterized by strong, attractive, short-ranged interactions

Can be experimentally measured and manipulated with high accuracy

Ideal testbed for developing a flexible NQS capable of capturing superfluidity



JK et al., Commun. Phys. ***7****, 148 (2024).*

NEURAL PFAFFIAN

Simplest and most general way to build an antisymmetrized product of pairing orbitals

$$\Phi(X) = \mathrm{pf}\begin{bmatrix} 0 & \phi(\mathbf{x}_1, \mathbf{x}_2) & \cdots & \phi(\mathbf{x}_1, \mathbf{x}_N) \\ -\phi(\mathbf{x}_1, \mathbf{x}_2) & 0 & \cdots & \phi(\mathbf{x}_2, \mathbf{x}_N) \\ \vdots & \vdots & \ddots & \vdots \\ -\phi(\mathbf{x}_1, \mathbf{x}_N) & -\phi(\mathbf{x}_2, \mathbf{x}_N) & \cdots & 0 \end{bmatrix}$$

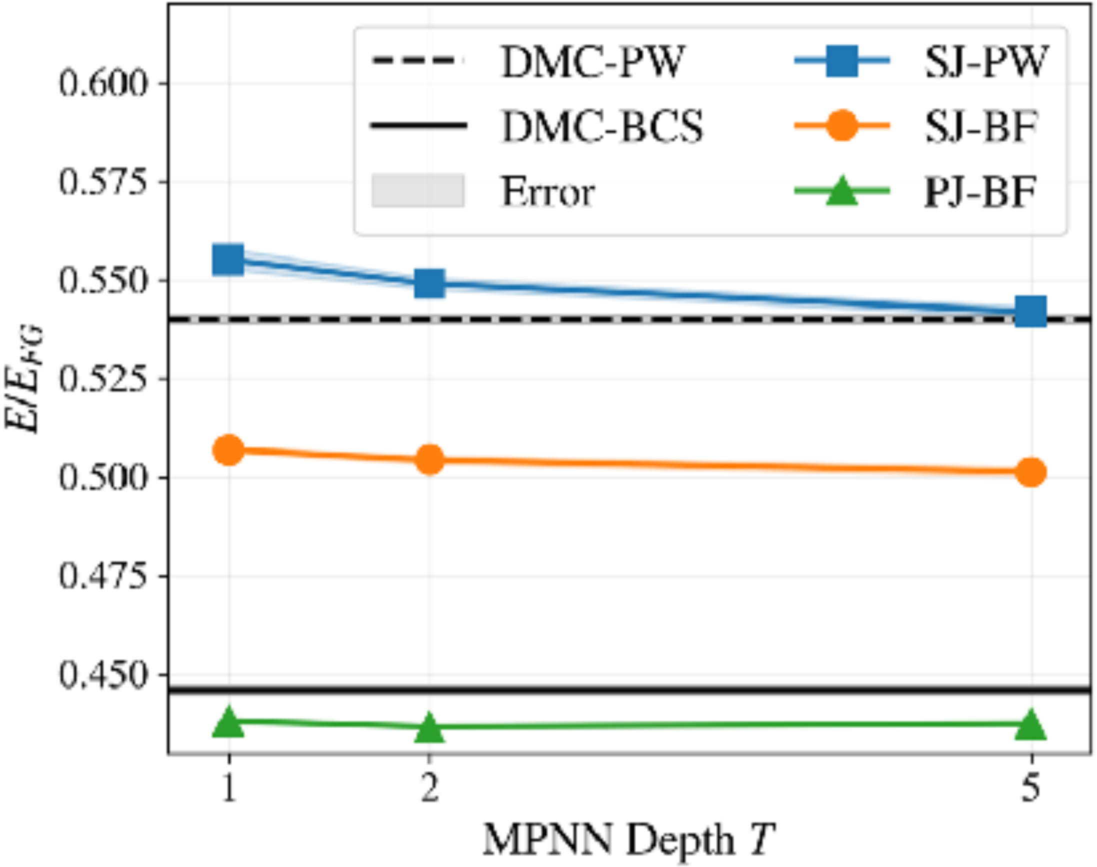
Only one neural network

$$\phi(\mathbf{x}_i, \mathbf{x}_j) = \nu(\mathbf{x}_i, \mathbf{x}_j) - \nu(\mathbf{x}_j, \mathbf{x}_i)$$

For a special choice of ϕ, this is equivalent to a Slater determinant!

JK et al., *Commun. Phys.* ***7***, *148 (2024).*

NEURAL PFAFFIAN

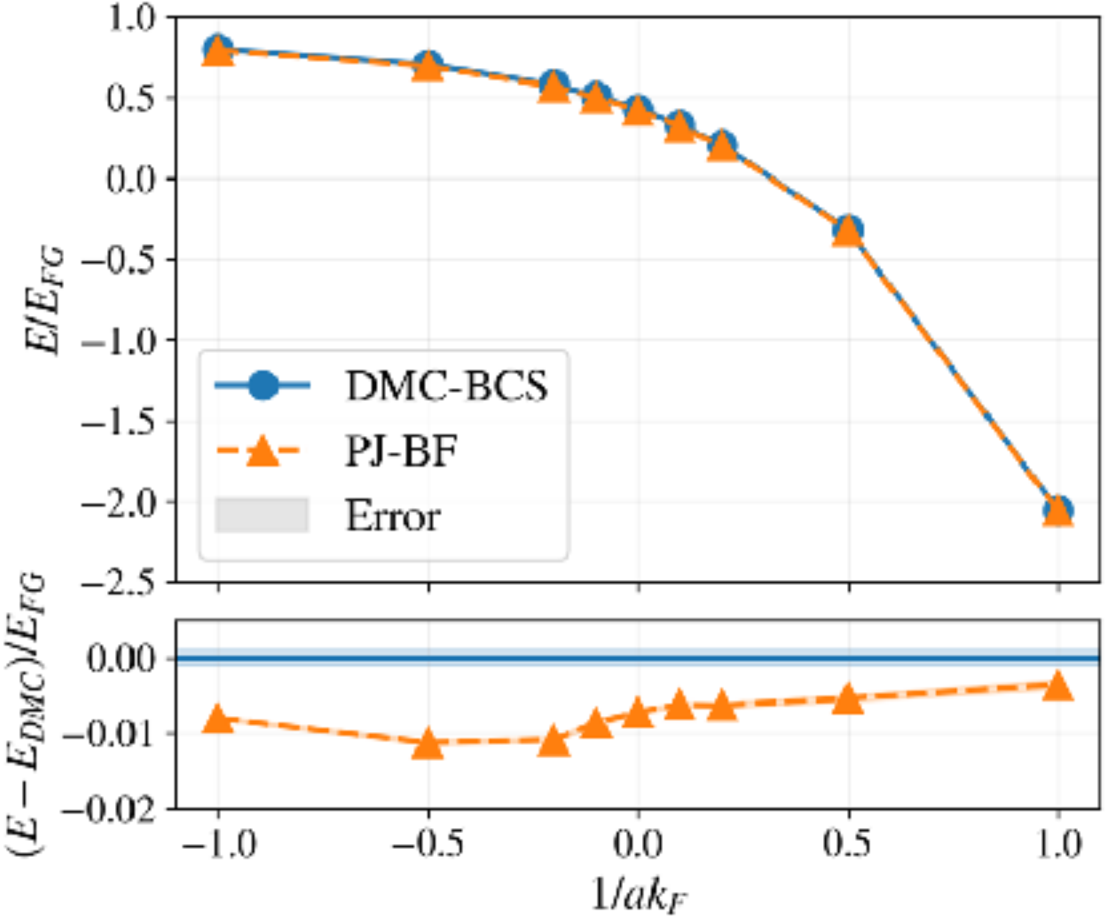


Neural Slater determinant is not sufficient for strongly paired system

Neural Pfaffian performs significantly better than diffusion Monte Carlo with BCS nodal surface

JK et al., Commun. Phys. ***7****, 148 (2024).*

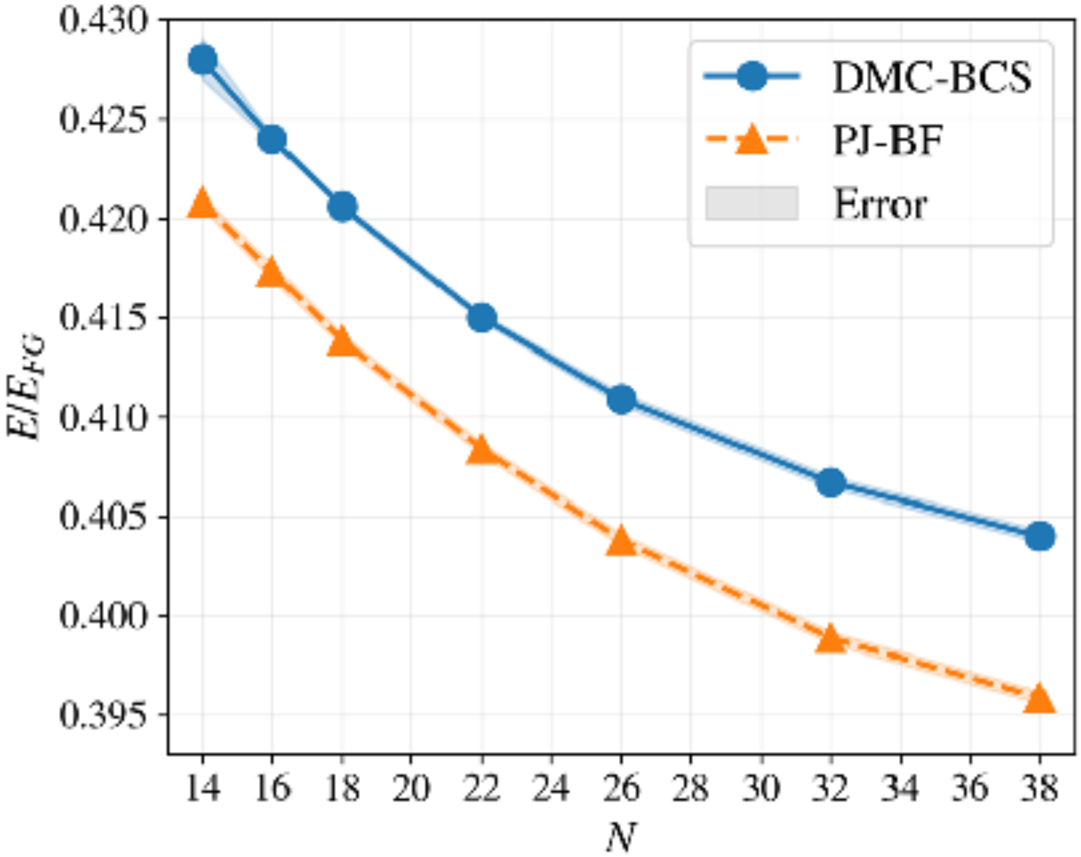
NEURAL PFAFFIAN



Results are consistent throughout the BCS-BEC crossover region

JK et al., Commun. Phys. ***7****, 148 (2024).*

NEURAL PFAFFIAN



The number of trainable parameters **does not** depend on particle number

Transfer learning can be exploited to reach larger systems faster

JK et al., Commun. Phys. ***7****, 148 (2024).*