# Neural Network Quantum States for Atomic Nuclei

**ECT* online program**
ADVANCES IN MANY-BODY THEORIES: FROM FIRST PRINCIPLE METHODS TO QUANTUM COMPUTING AND MACHINE LEARNING

November 3, 2020

Alessandro Lovato

Argonne NATIONAL LABORATORY

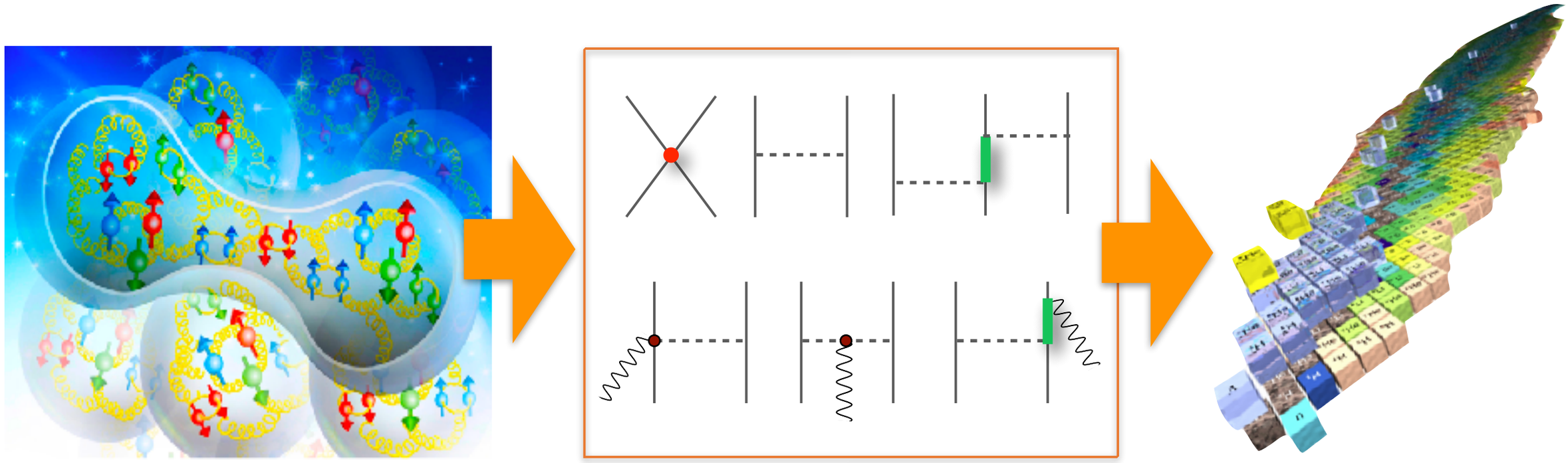
INFN Istituto Nazionale di Fisica Nucleare

In collaboration with:
C. Adams, P. Balaprakash, G. Carleo, K. Raghavan, N. Rocco

# The microscopic model of nuclear theory

In the low-energy regime, **quark and gluons are confined within hadrons** and the relevant degrees of freedoms are protons, neutrons, and pions

**Effective field theories** are the link between QCD and nuclear observables. They exploit the separation between the “hard” (nucleon mass) and “soft” (exchanged momentum) scales



Nucleons can treated as point-like particles interacting through the Hamiltonian

$$H = \sum_i \frac{\mathbf{p}_i^2}{2m} + \sum_{i<j} v_{ij} + \sum_{i<j<k} V_{ijk} + \dots$$

# “Conventional” continuum nuclear quantum Monte Carlo methods



$$H|\Psi_n\rangle = E_n|\Psi_n\rangle$$

$$J_{mn} = \langle\Psi_m|J|\Psi_n\rangle$$

# Variational Monte Carlo

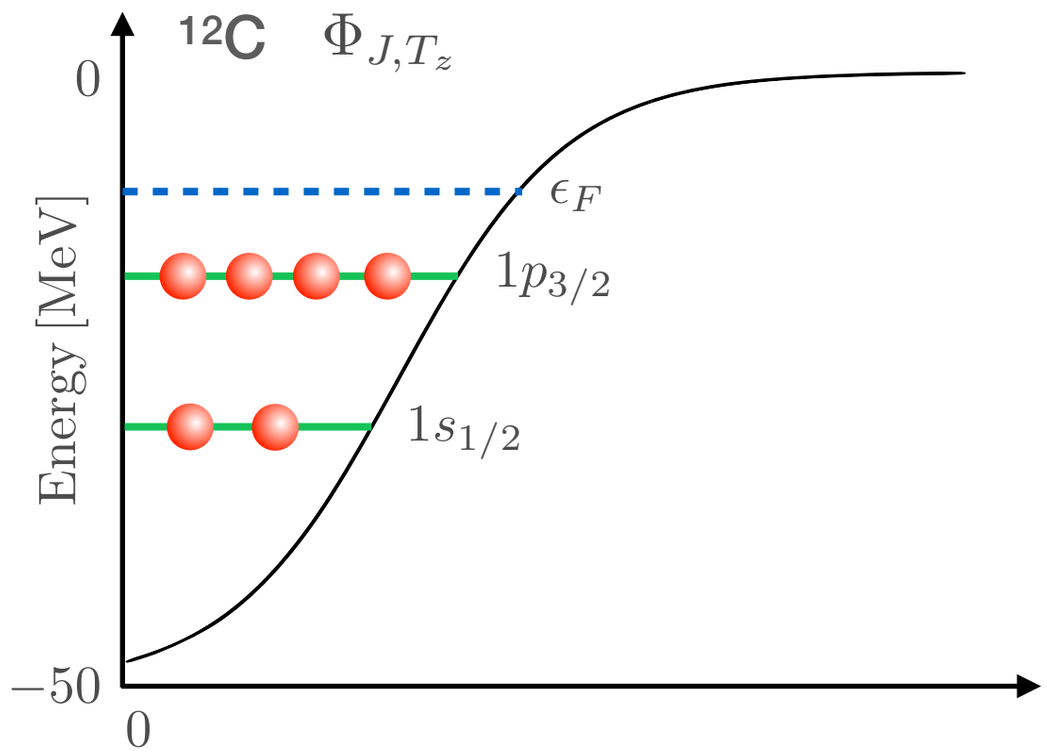
In variational Monte Carlo, one **assumes a suitable form for the trial wave function**

$$|\Psi_T\rangle = \left(1 + \sum_{ijk} F_{ijk}\right)\left(\mathcal{S}\prod_{i<j} F_{ij}\right)|\Phi_{J,T_z}\rangle$$

The best variational parameters are found by **optimizing the variational energy**

$$E_T = \langle\Psi_T|H|\Psi_T\rangle \geq E_0$$

The long-range antisymmetric is typically a Slater determinant of single-particle orbitals



# Variational Monte Carlo

In variational Monte Carlo, one **assumes a suitable form for the trial wave function**

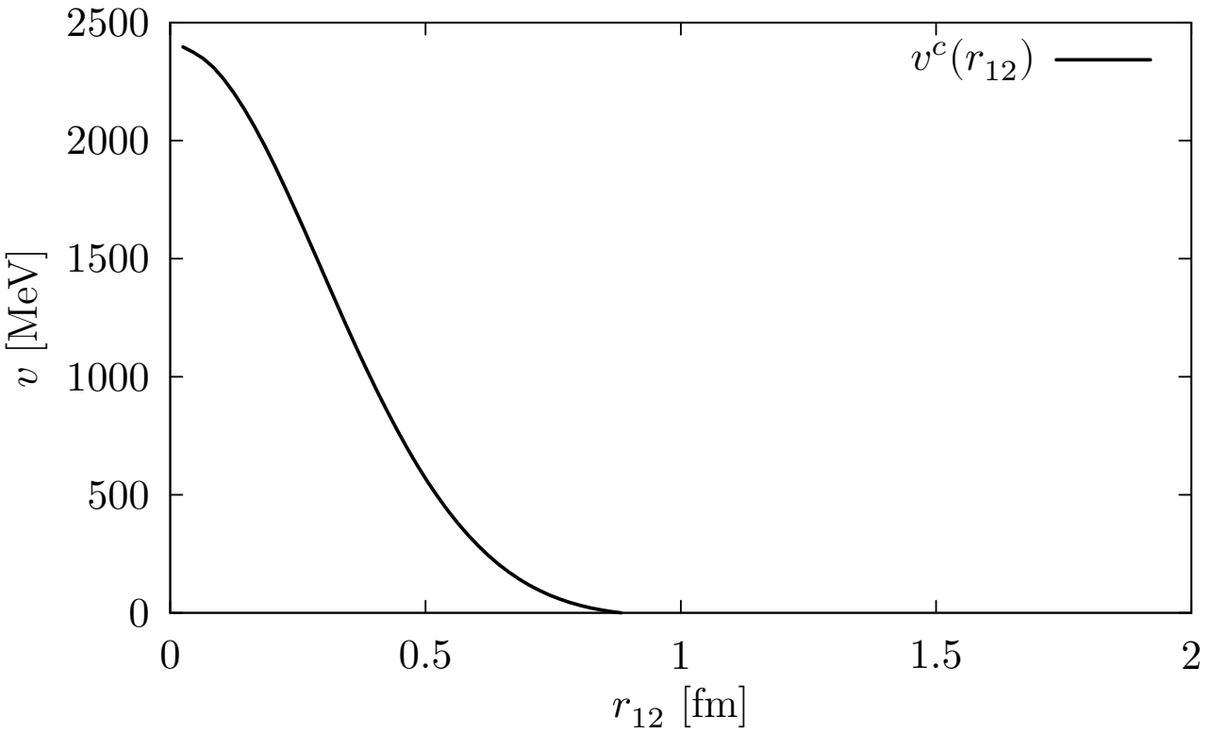
$$|\Psi_T\rangle = \left(1 + \sum_{ijk} F_{ijk}\right)\left(\mathcal{S}\prod_{i<j} F_{ij}\right)|\Phi_{J,T_z}\rangle$$

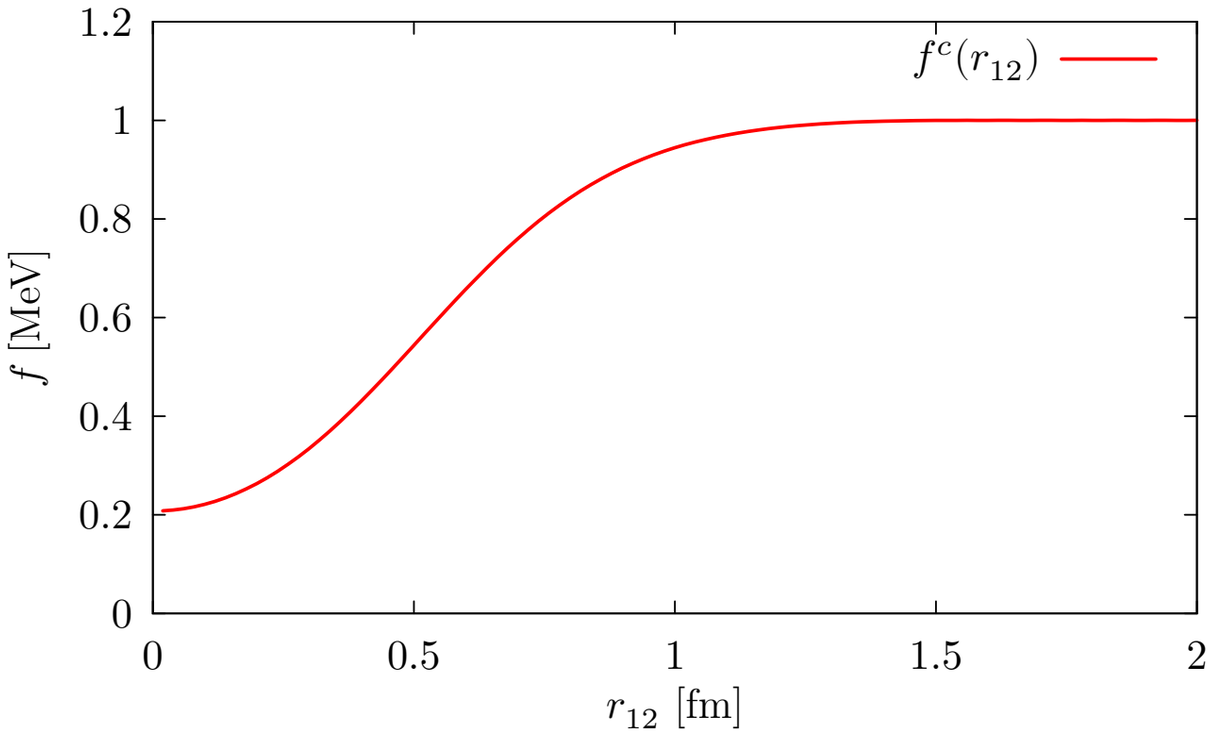
The best variational parameters are found by **optimizing the variational energy**

$$E_T = \langle\Psi_T|H|\Psi_T\rangle \geq E_0$$

The correlation operator reflects the spin-isospin dependence of the nuclear interaction

$$F_{ij} \simeq \sum_p f^p(r_{ij})O^p_{ij} \qquad F_{ijk} = \sum_x \epsilon_x V^x_{ijk}(\tilde{r}_{ij}, \tilde{r}_{ik}, \tilde{r}_{jk})$$



# Green's function Monte Carlo

GFMC **overcomes the limitations of the variational wave-function** by using an imaginary-time  projection technique

Any trial wave function can be expanded in the complete set of eigenstates of the the Hamiltonian according to

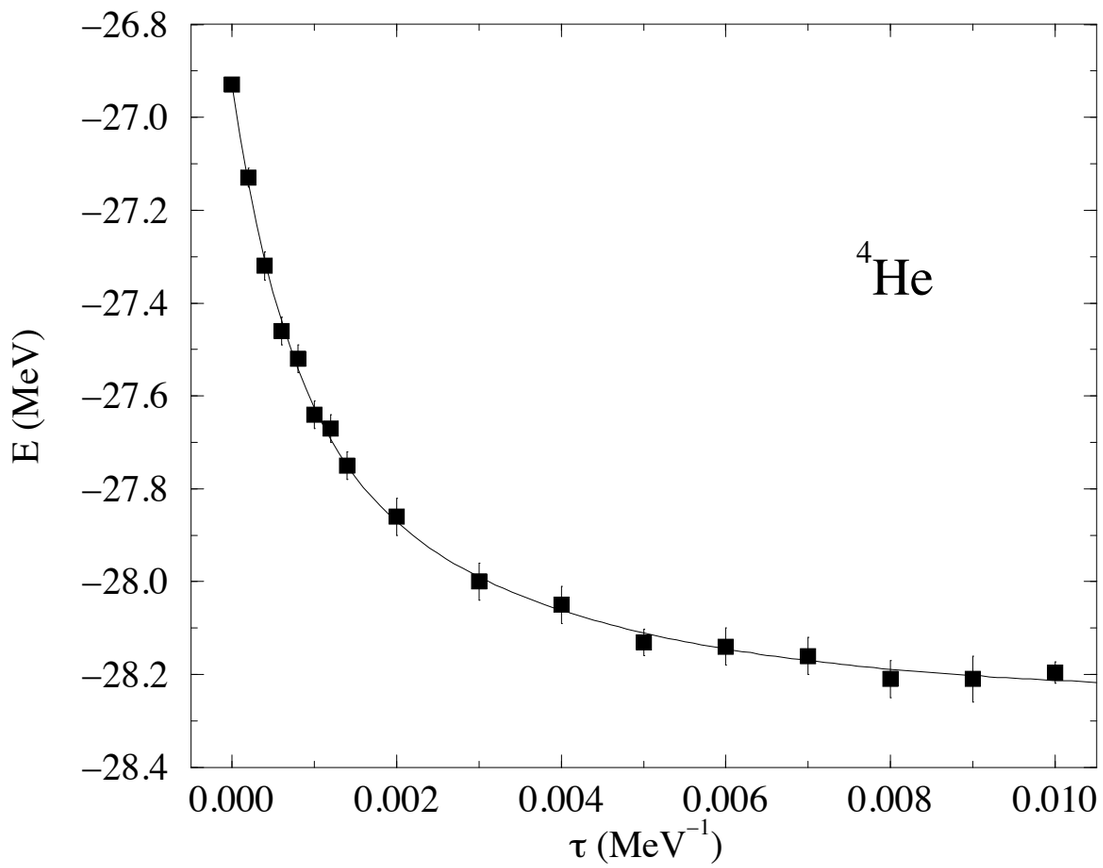
$$|\Psi_T\rangle = \sum_n c_n |\Psi_n\rangle \qquad\qquad H|\Psi_n\rangle = E_n|\Psi_n\rangle$$

GFMC **projects out the exact lowest-energy state**, provided the trial wave function it is not orthogonal to the ground state.

$$\lim_{\tau\to\infty} e^{-(H-E_0)\tau}|\Psi_T\rangle =$$

$$\lim_{\tau\to\infty} \sum_n c_n\, e^{-(E_n-E_0)\tau}|\Psi_n\rangle = c_0|\Psi_0\rangle$$

B. Pudliner et al., PRC **56**, 1720 (1997)



# Green's function Monte Carlo

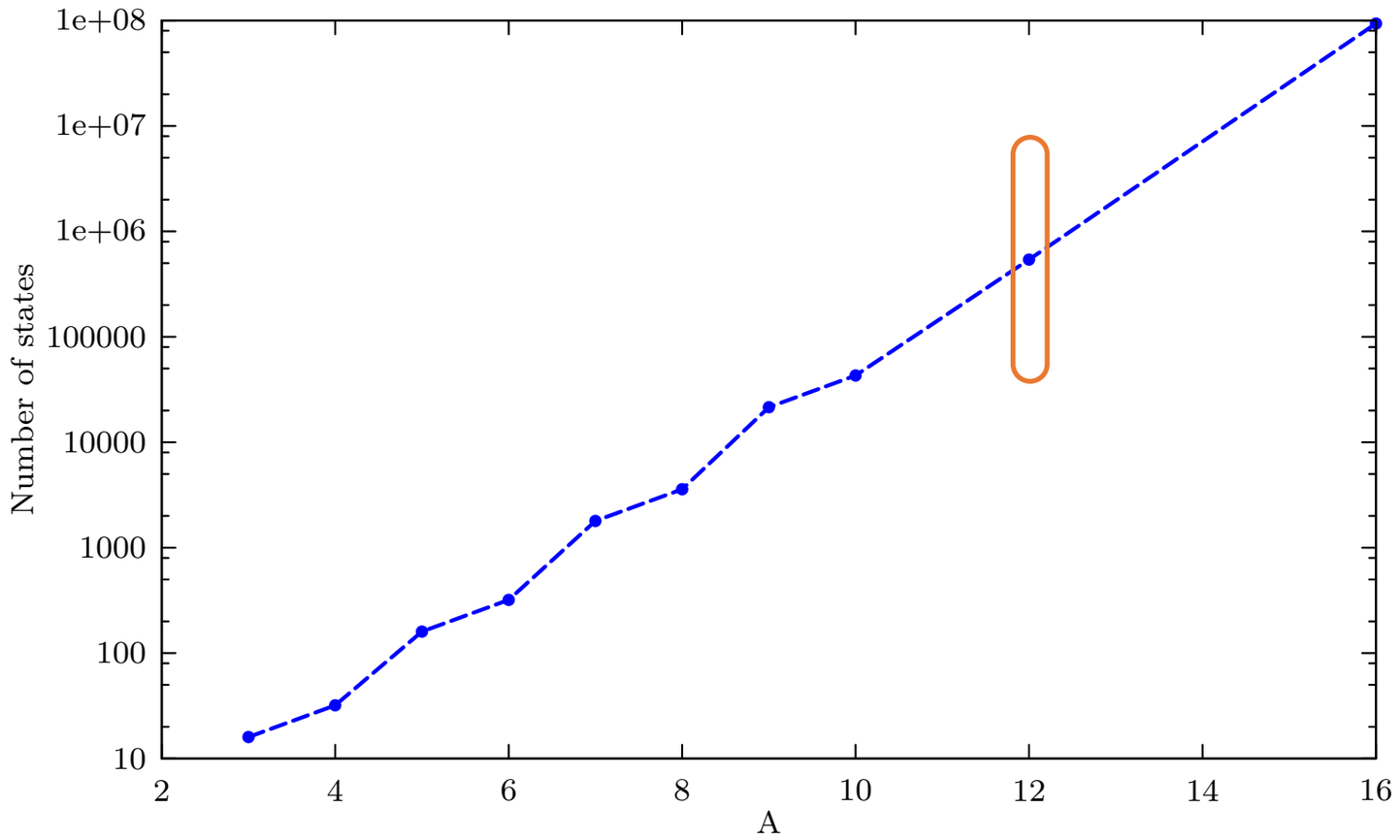
In the GFMC, **a sum over all the many-body spin-isospin states** is performed

$$\sum_{SS'}\langle S'|e^{-[V-E_0]\delta\tau}|S\rangle \simeq \sum_{SS'}\langle S'|\prod_{i<j}e^{-V_{ij}\delta\tau}|S\rangle e^{E_0\delta\tau}$$

Many-body spin-isospin states are utilized → unfavorable exponential scaling

The size of the spin state alone grows like $2^A$

$$|S\rangle = \begin{pmatrix} s_{\uparrow\uparrow\uparrow} \\ s_{\uparrow\uparrow\downarrow} \\ s_{\uparrow\downarrow\uparrow} \\ s_{\uparrow\downarrow\downarrow} \\ s_{\downarrow\uparrow\uparrow} \\ s_{\downarrow\uparrow\downarrow} \\ s_{\downarrow\downarrow\uparrow} \\ s_{\downarrow\downarrow\downarrow} \end{pmatrix}$$




GFMC is extremely accurate but **limited to A≲12 nuclei** and small (A ≤14) neutron systems

# Auxiliary-field diffusion Monte Carlo

The AFDMC method uses a spin-isospin basis given by the outer product of single-nucleon spinors

$$|S\rangle = |s_1\rangle \otimes |s_2\rangle \cdots \otimes |s_A\rangle \quad \longleftrightarrow \quad |s\rangle = s_{p\uparrow}|p\uparrow\rangle + s_{p\downarrow}|p\downarrow\rangle + s_{n\uparrow}|n\uparrow\rangle + s_{n\downarrow}|n\downarrow\rangle$$

In purely neutron systems $\tau_{ij} = 1$ and the spin-dependent part of the potential reads

$$V_{SD} = \frac{1}{2}\sum_{i\alpha j\beta} A_{i\alpha,j\beta}\sigma_i^\alpha \sigma_j^\beta = \frac{1}{2}\sum_{n=1}^{3A} O_n^2 \lambda_n \,,$$

To preserve the single-particle representation, the short-time propagator is linearized utilizing the **Hubbard-Stratonovich transformation**

$$e^{-\lambda O^2 \delta\tau/2} = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} dx e^{-x^2/2} e^{x\sqrt{-\lambda\delta\tau}\,O}$$

At each imaginary-time step the single particle spinors are rotated

$$e^{-V\delta\tau}[|s_1\rangle \otimes |s_2\rangle \cdots \otimes |s_A\rangle] = |s_1'\rangle \otimes |s_2'\rangle \cdots \otimes |s_A'\rangle$$

AFDMC **scales polynomially with the number of nucleons** but is limited to simplified potentials

K. E. Schmidt et al., PLB **446**, 99 (1999)

# Exponential complexity

An exponential in the particle number operations is required to evaluate the correlation operator

$$\left(\mathcal{S}\prod_{i<j}F_{ij}\right)|\Phi\rangle$$

Consider simple spin-pair correlation (no tensor)

$$F_{ij} = f^c(r_{ij}) + f^\sigma(r_{ij})\sigma_{ij} \quad \longleftrightarrow \quad \sigma_{ij} = 2P^\sigma_{ij} - 1$$

Let us apply the correlation operator to a three-body spin state

$$(f^c_{12} + f^\sigma_{12}\sigma_{12})(f^c_{13} + f^\sigma_{13}\sigma_{13})(f^c_{23} + f^\sigma_{23}\sigma_{23})|s_1, s_2, s_3\rangle$$

# Exponential complexity

An exponential in the particle number operations is required to evaluate the correlation operator

$$\left(\mathcal{S}\prod_{i<j}F_{ij}\right)|\Phi\rangle$$

Consider simple spin-pair correlation (no tensor)

$$F_{ij}=f^c(r_{ij})+f^\sigma(r_{ij})\sigma_{ij} \quad \longleftrightarrow \quad \sigma_{ij}=2P^\sigma_{ij}-1$$

Let us apply the correlation operator to a three-body spin state

$$\begin{aligned}&(f^c_{12}+f^\sigma_{12}\sigma_{12})(f^c_{13}+f^\sigma_{13}\sigma_{13})(f^c_{23}+f^\sigma_{23}\sigma_{23})|s_1,s_2,s_3\rangle\\ =&(f^c_{12}+f^\sigma_{12}\sigma_{12})(f^c_{13}+f^\sigma_{13}\sigma_{13})[(f^c_{23}-f^\sigma_{23})|s_1,s_2,s_3\rangle+2f^\sigma_{23}|s_1,s_3,s_2\rangle]\end{aligned}$$

# Exponential complexity

An exponential in the particle number operations is required to evaluate the correlation operator

$$\left(\mathcal{S}\prod_{i<j}F_{ij}\right)|\Phi\rangle$$

Consider simple spin-pair correlation (no tensor)

$$F_{ij}=f^c(r_{ij})+f^\sigma(r_{ij})\sigma_{ij} \quad \longleftrightarrow \quad \sigma_{ij}=2P^\sigma_{ij}-1$$

Let us apply the correlation operator to a three-body spin state

$$\begin{aligned}
&(f^c_{12}+f^\sigma_{12}\sigma_{12})(f^c_{13}+f^\sigma_{13}\sigma_{13})(f^c_{23}+f^\sigma_{23}\sigma_{23})|s_1,s_2,s_3\rangle \\
=&(f^c_{12}+f^\sigma_{12}\sigma_{12})(f^c_{13}+f^\sigma_{13}\sigma_{13})[(f^c_{23}-f^\sigma_{23})|s_1,s_2,s_3\rangle+2f^\sigma_{23}|s_1,s_3,s_2\rangle] \\
=&(f^c_{12}+f^\sigma_{12}\sigma_{12})[(f^c_{13}-f^\sigma_{13})(f^c_{23}-f^\sigma_{23})|s_1,s_2,s_3\rangle+2(f^c_{13}-f^\sigma_{13})f^\sigma_{23}|s_1,s_3,s_2\rangle \\
&+2f^\sigma_{13}(f^c_{23}-f^\sigma_{23})|s_3,s_2,s_1\rangle+4f^\sigma_{13}f^\sigma_{23}|s_2,s_3,s_1\rangle]
\end{aligned}$$

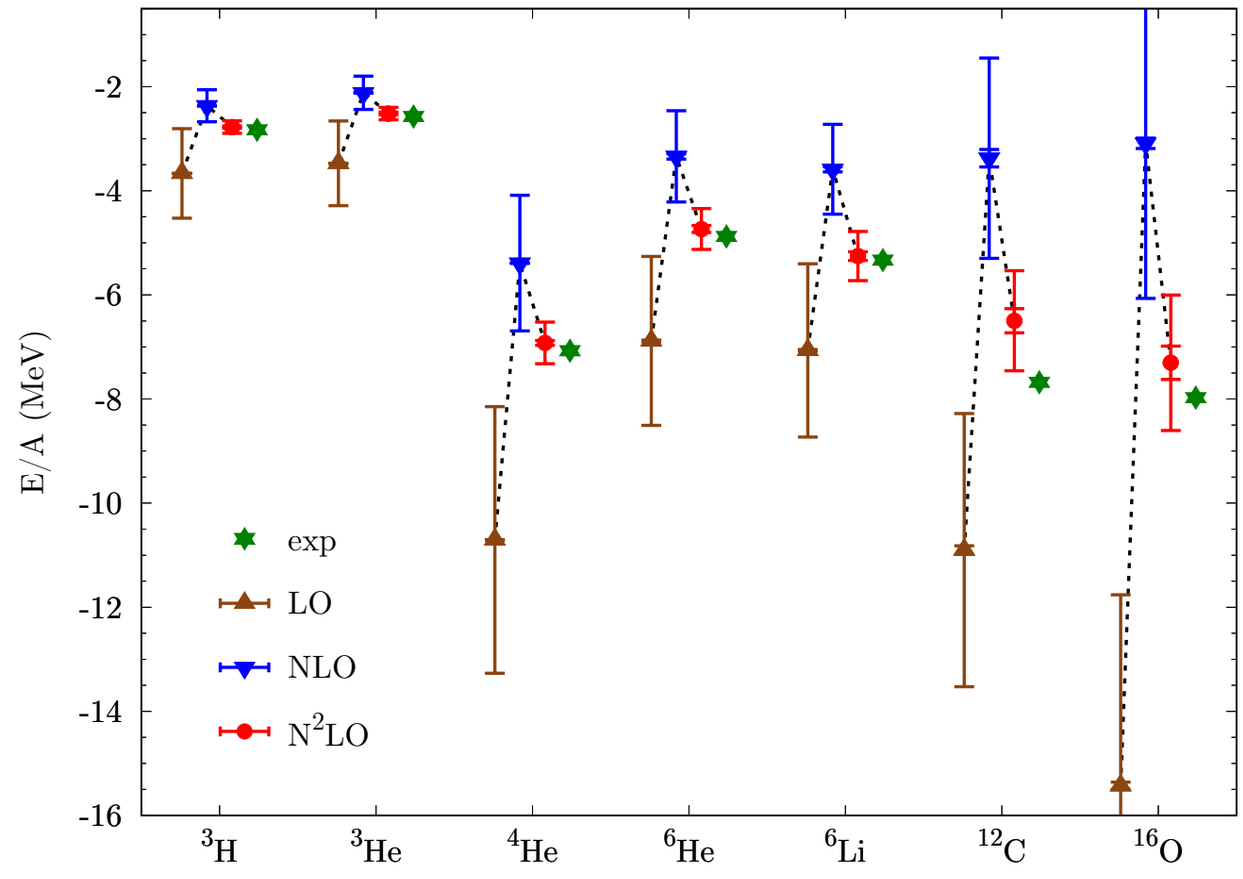
Sampling the spin-isospin state and evaluating the trial wave function's amplitude for that sampled state **still requires a number of operations exponential in the particle number**

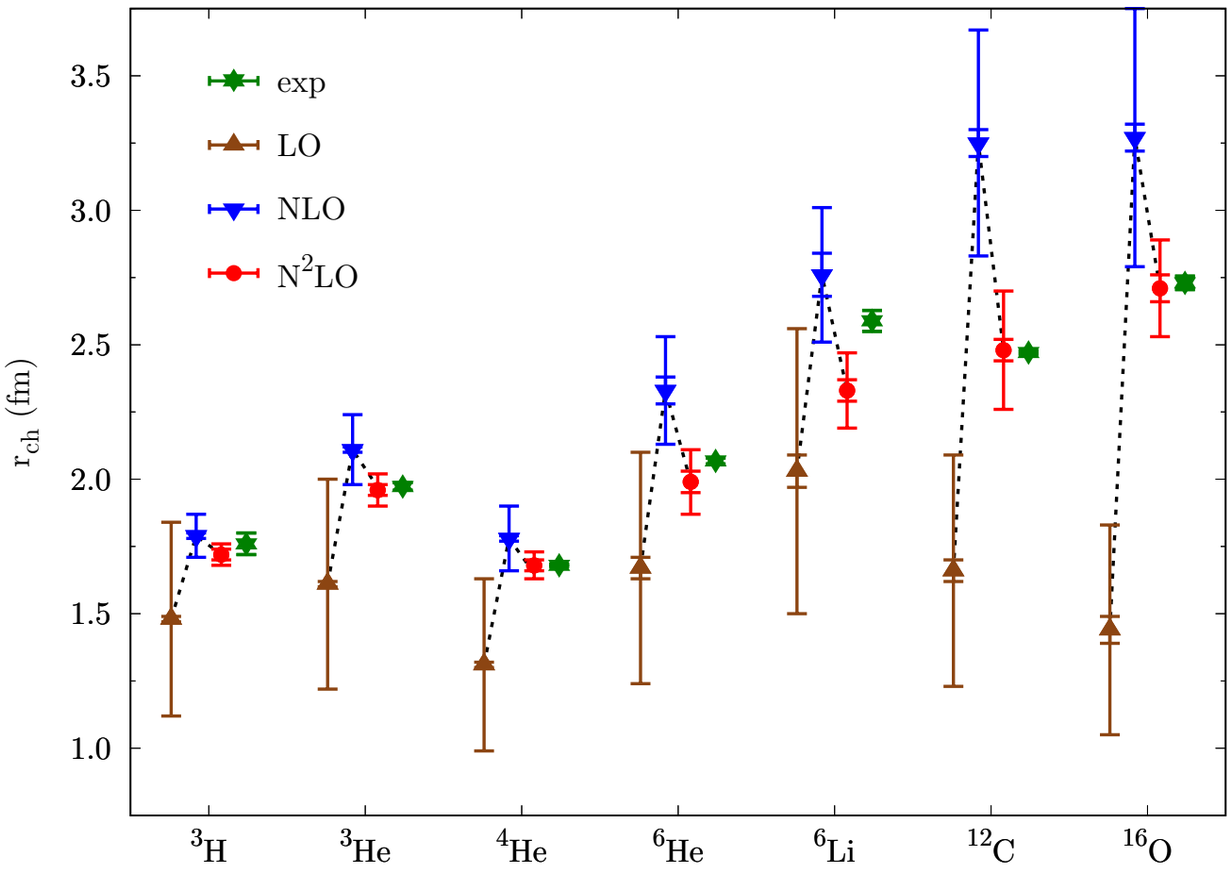
# Exponential complexity

Recent AFDMC calculations employ a **linear approximation** to spin-isospin dependent correlations. Simple but it **violates the factorization theorem**.

S. Gandolfi, AL et al., PRC **90**, 061306 (2014)

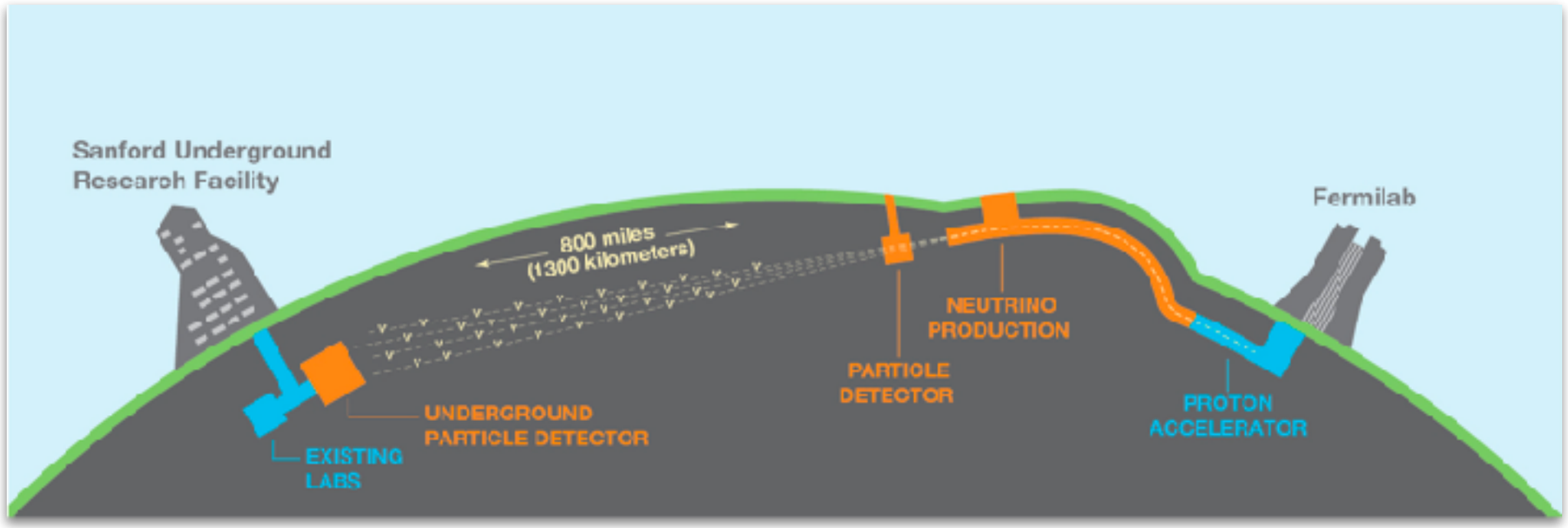
$$\left(\mathcal{S}\prod_{i<j}F_{ij}\right)|\Phi_{J,T_z}\rangle \longrightarrow \left(\prod_{i<j}f_{ij}^{c}\right)\left(1+\sum_{i<j}U_{ij}\right)|\Phi_{J,T_z}\rangle$$



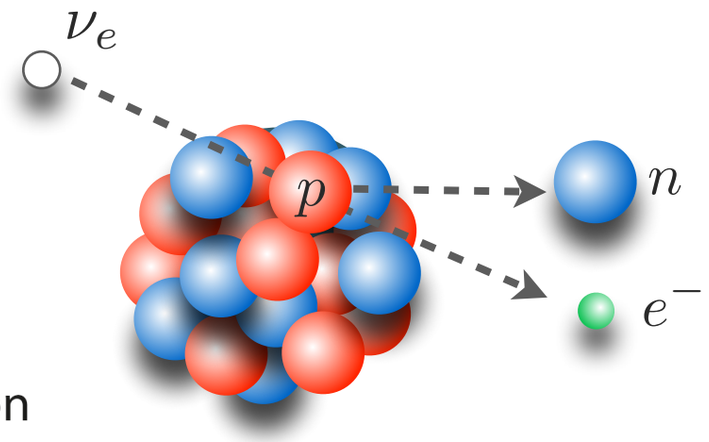

D. Lonardoni et al., PRL **120**, 122502 (2018)

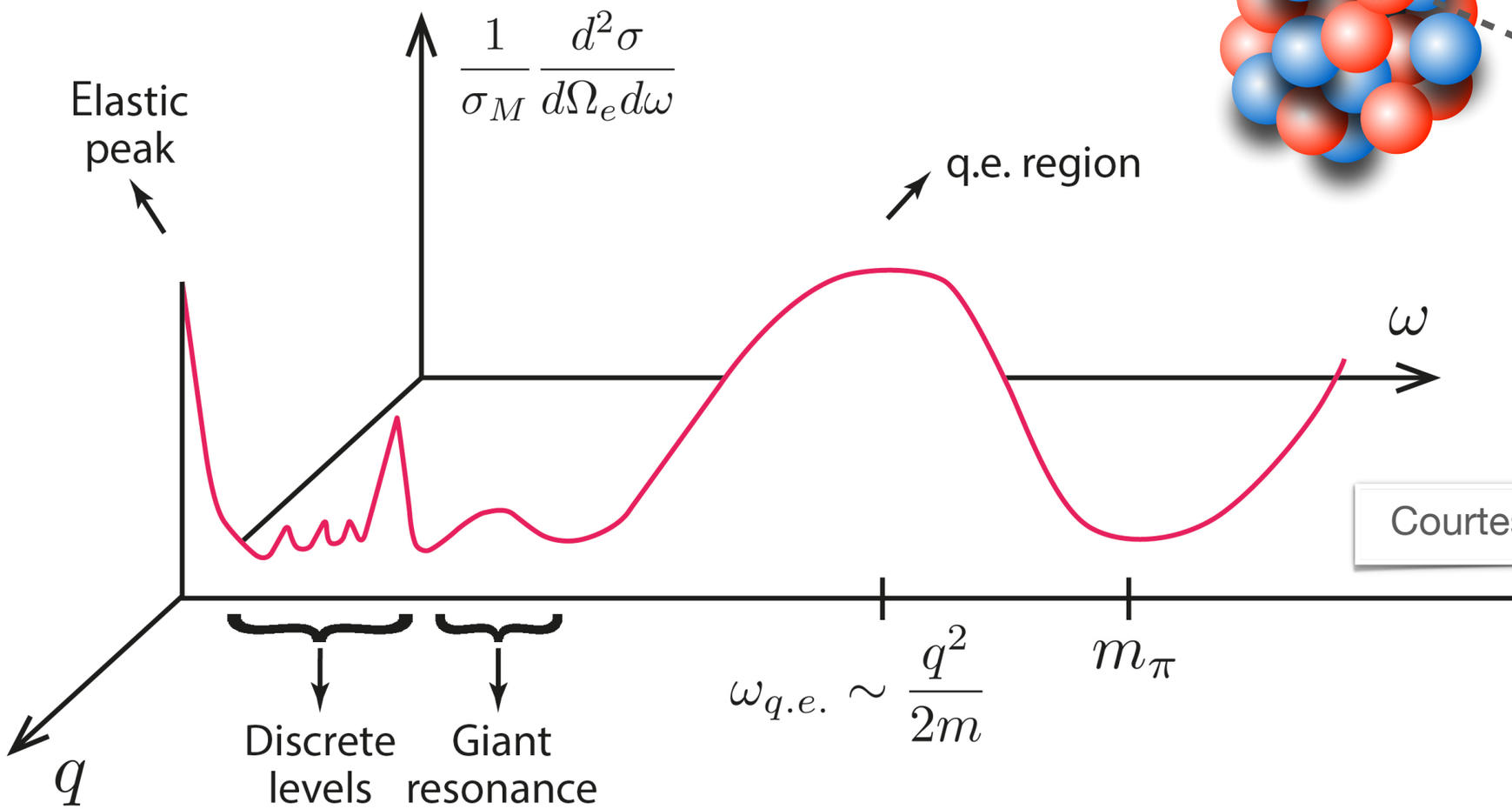
# Lepton-nucleus scattering with ML methods



# Lepton-nucleus scattering

The inclusive cross section is characterized by a variety of reaction mechanisms




Courtesy of Saori Pastore

The response functions contain all information on the structure and dynamics of the target

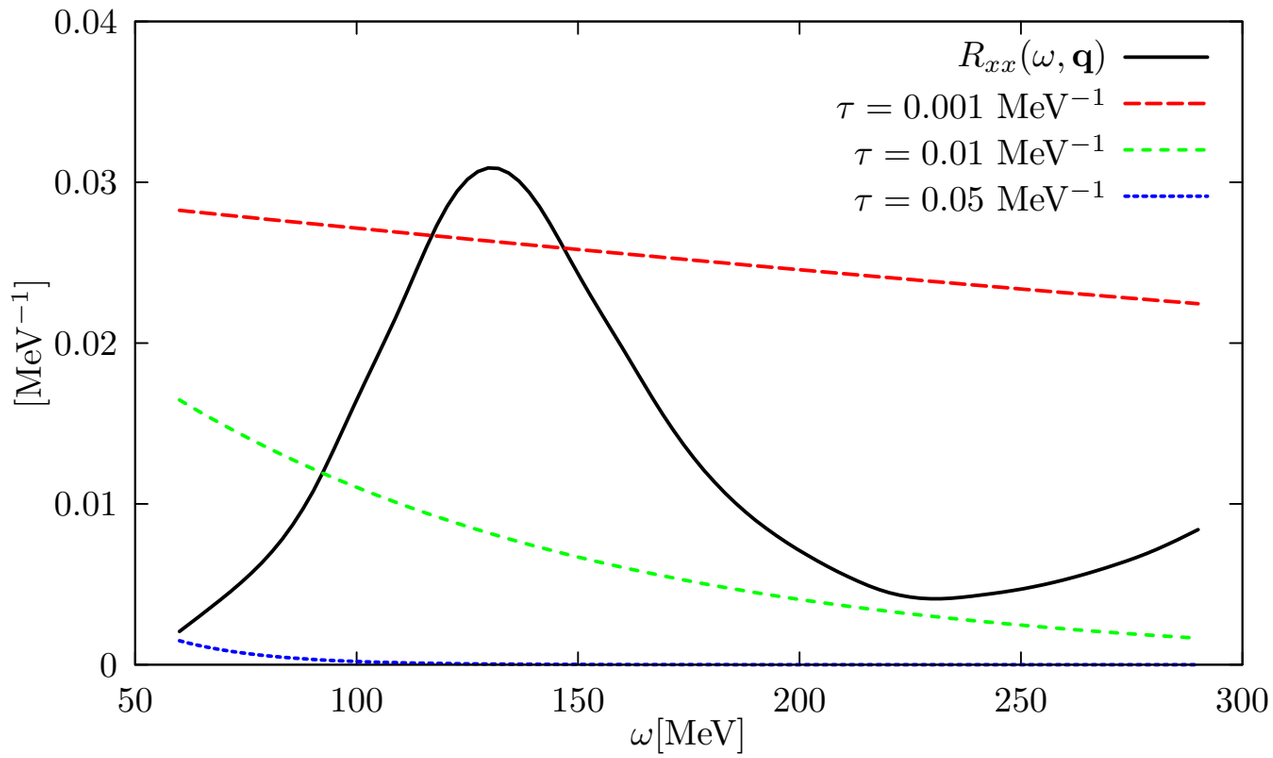
$$R_{\alpha\beta}(\omega, \mathbf{q}) = \sum_f \langle \Psi_0 | J^\dagger_\alpha(\mathbf{q}) | \Psi_f \rangle \langle \Psi_f | J_\beta(\mathbf{q}) | \Psi_0 \rangle \delta(\omega - E_f + E_0)$$

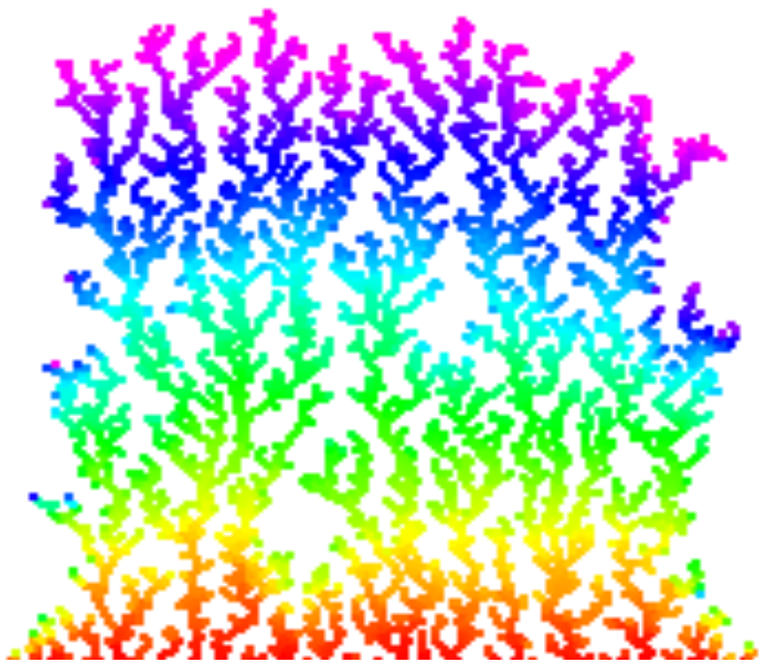
# Lepton-nucleus scattering

On type of integral transform is the **Laplace transform**

$$E_{\alpha\beta}(\tau, \mathbf{q}) \equiv \int d\omega e^{-\omega\tau} R_{\alpha\beta}(\omega, \mathbf{q})$$

At finite imaginary time the contributions from large energy transfer are quickly suppressed



The system is first heated up by the transition operator.
Its cooling determines the **Euclidean response** of the system

$$\sum_f |\Psi_f\rangle\langle\Psi_f|$$

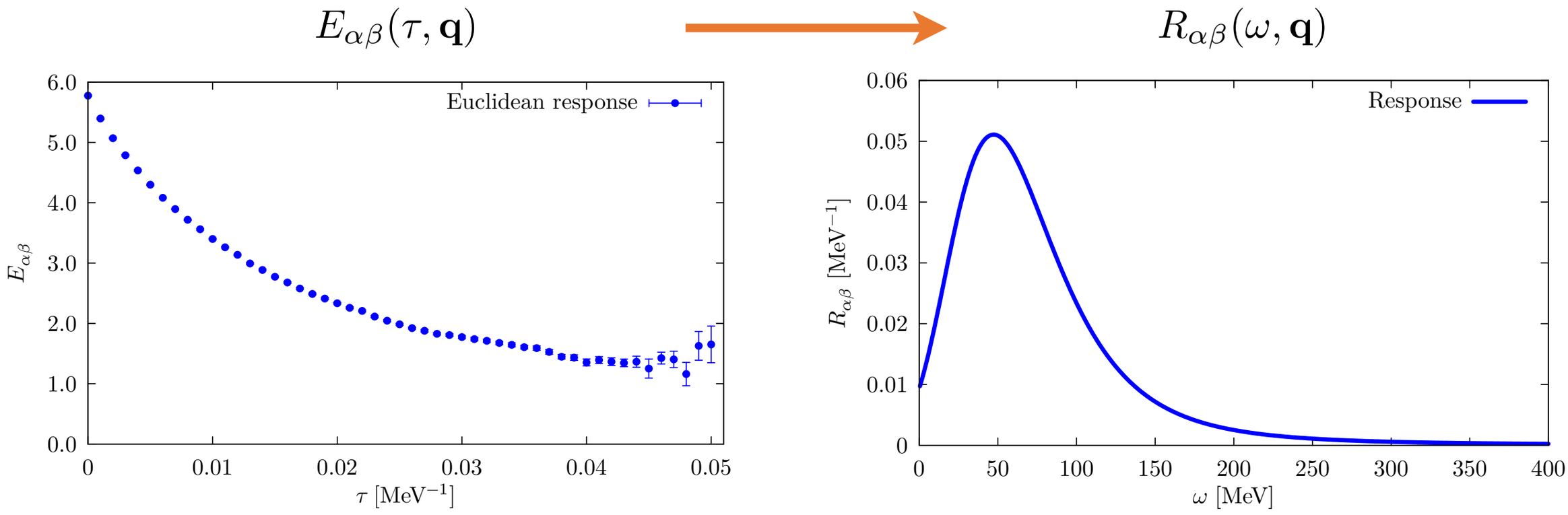
$$E_{\alpha\beta}(\tau, \mathbf{q}) = \langle\Psi_0|J^\dagger_\alpha(\mathbf{q}) e^{-(H-E_0)\tau} J_\beta(\mathbf{q})|\Psi_0\rangle$$

Analogous techniques are used in Lattice QCD and condensed matter Physics

# Lepton-nucleus scattering

The Euclidean response formalism allows one to extract dynamical properties of the system from ground-state calculations

**Inverting the Euclidean response is an ill posed problem**: any set of observations is limited and noisy and the situation is even worse since the kernel is a smoothing operator.




**Maximum-entropy techniques:**

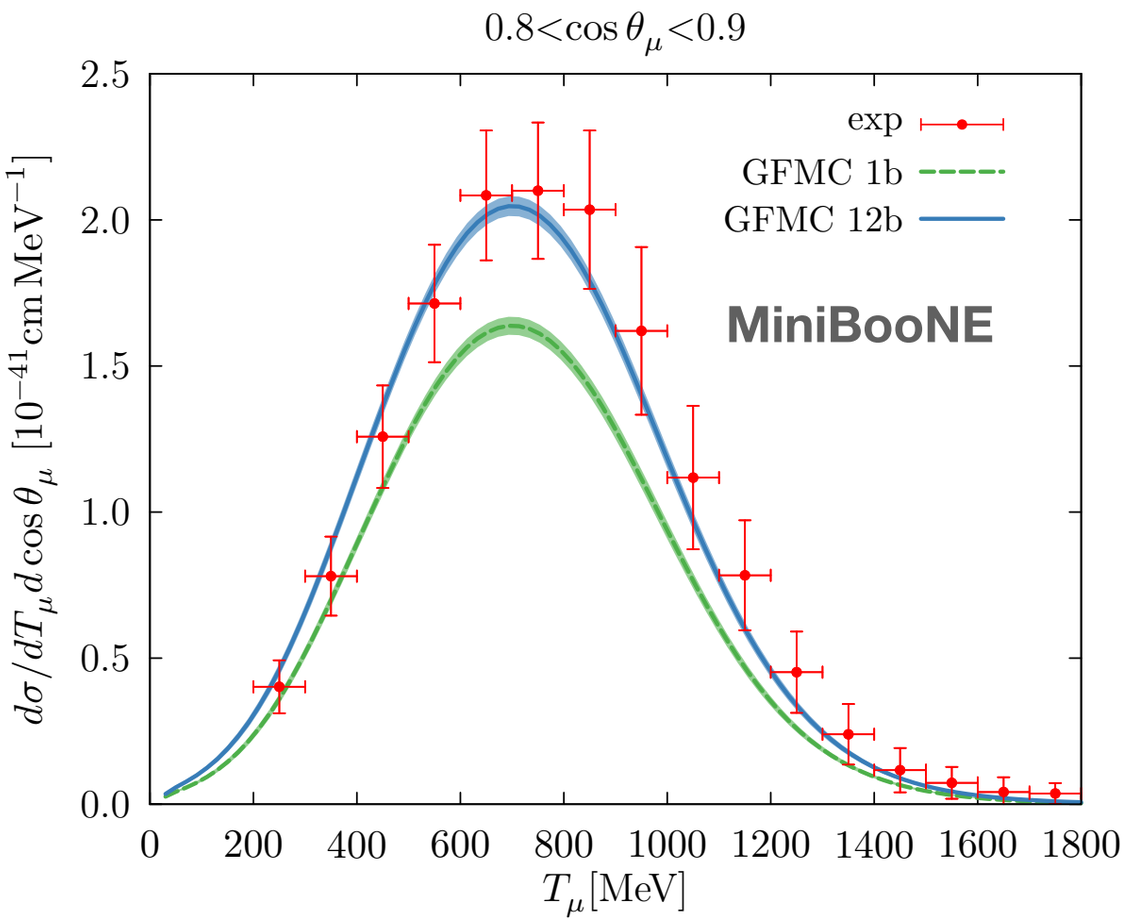
- Reliable enough for quasi-elastic responses;
- Fail to reproduce the low-energy structure of the response functions;

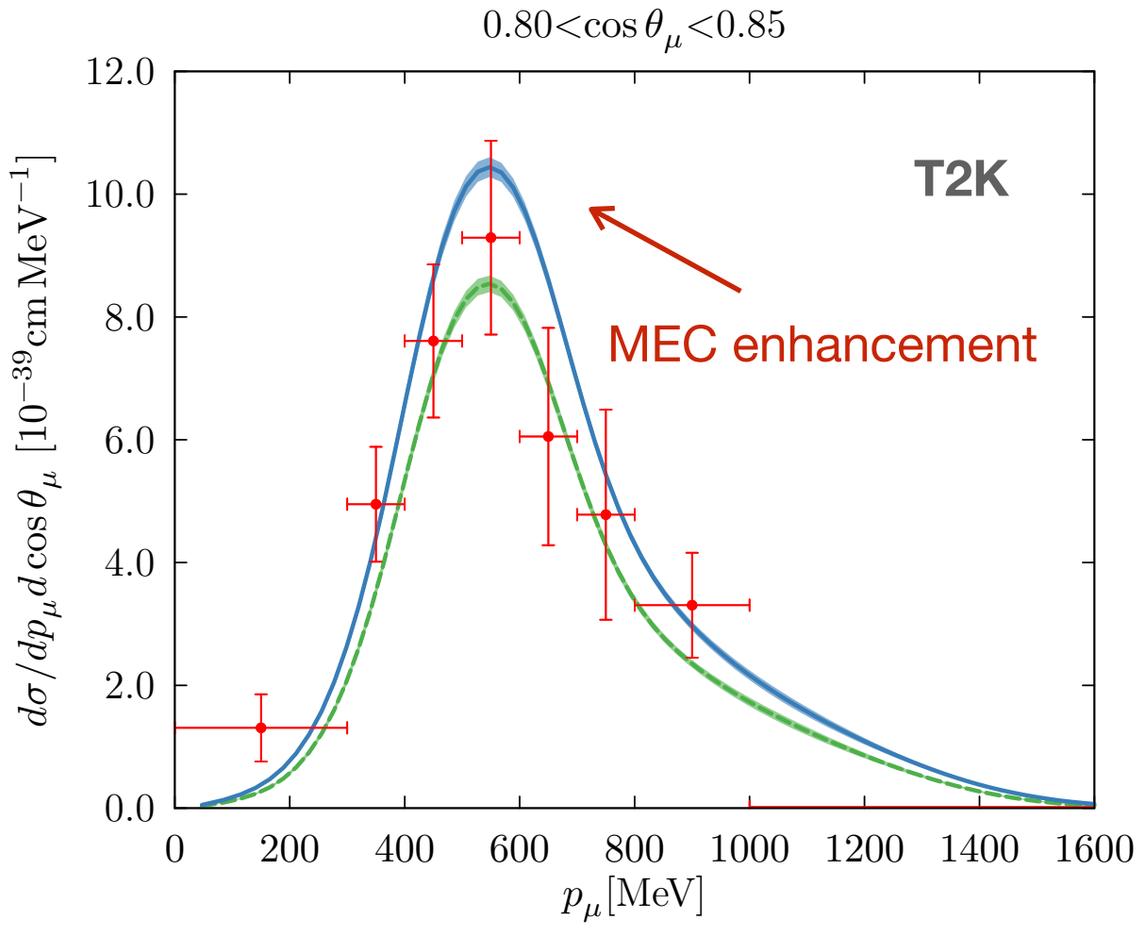
# MiniBoone cross sections

AL et al., Phys. Rev. X 10, **031068** (2020)

First microscopic calculation of neutrino-$^{12}$C cross section

$$\left\langle \frac{d\sigma}{dT_\mu\, d\cos\theta_\mu} \right\rangle = \int dE_\nu\, \phi(E_\nu)\, \frac{d\sigma(E_\nu)}{dT_\mu\, d\cos\theta_\mu}$$




Treating larger nuclei, including $^{16}$O and $^{40}$Ar:

- Higher noise level in the Euclidean responses;
- Non-negligible low-energy transitions;

→ Need to go beyond Maximum Entropy

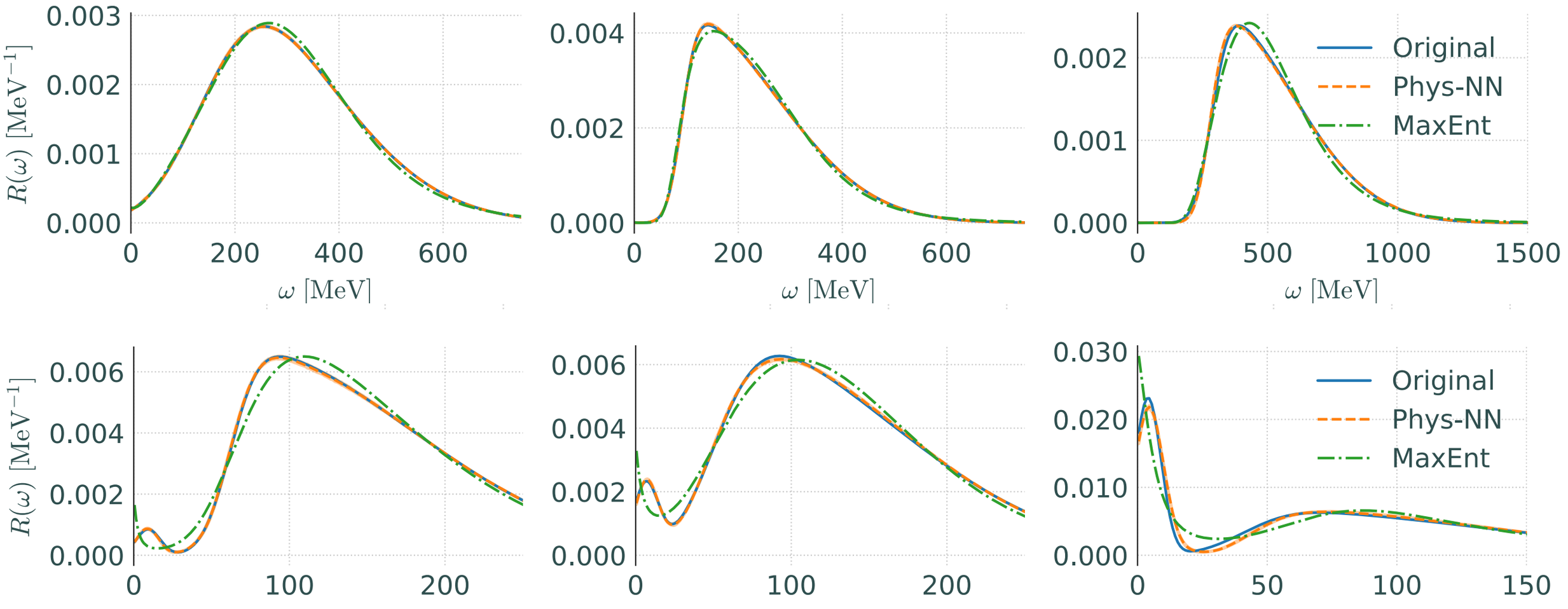
# Lepton-nucleus scattering

Several works have demonstrated that ML approaches are suitable for tackling inverse problems in condensed-matter physics

H. Yoo et al., Phys. Rev. B **98**, 245101 (2018)

R. Fournier et al., Phys. Rev. Lett. **124**, 056401 (2020)

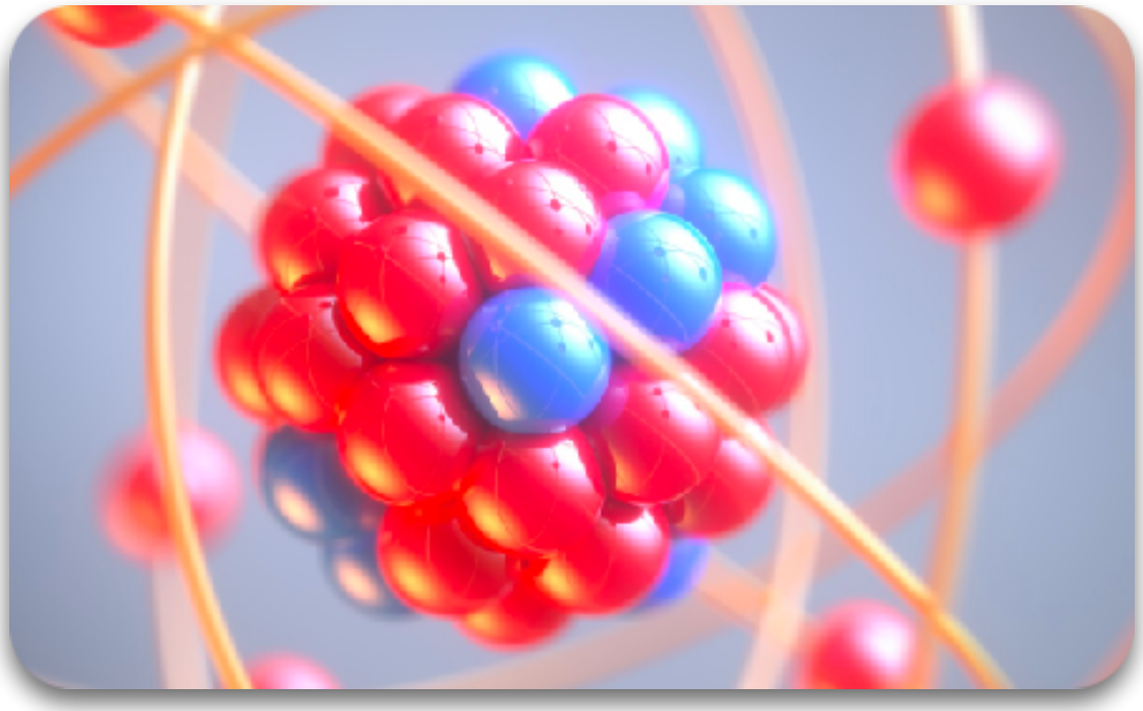
We propose a **Physics-inspired ANN** to approximate the inverse of the Laplace transform of nuclear electromagnetic response functions




**Krishnan's talk**

K. Raghavan et al., ArXiv 2010.12703

# Neural Network Quantum States for Atomic Nuclei



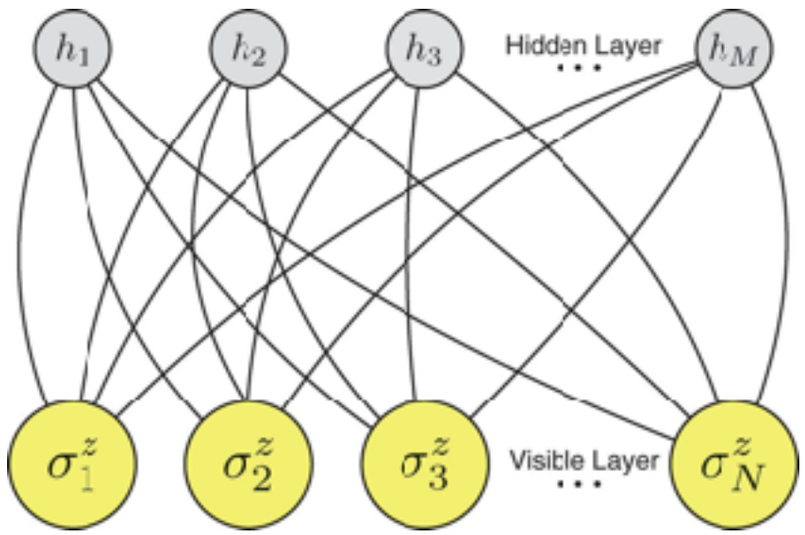
# Neural-network quantum states

- Artificial neural networks (ANNs) can compactly represent complex high-dimensional functions;

- Variational representations of **spin-systems** quantum states based on ANNs have been found to outperform conventional variational ansatz;

G. Carleo et al. Science **355**, 602 (2017)

G. Carleo et al. Nat. Commun. **9**, 532 (2018)

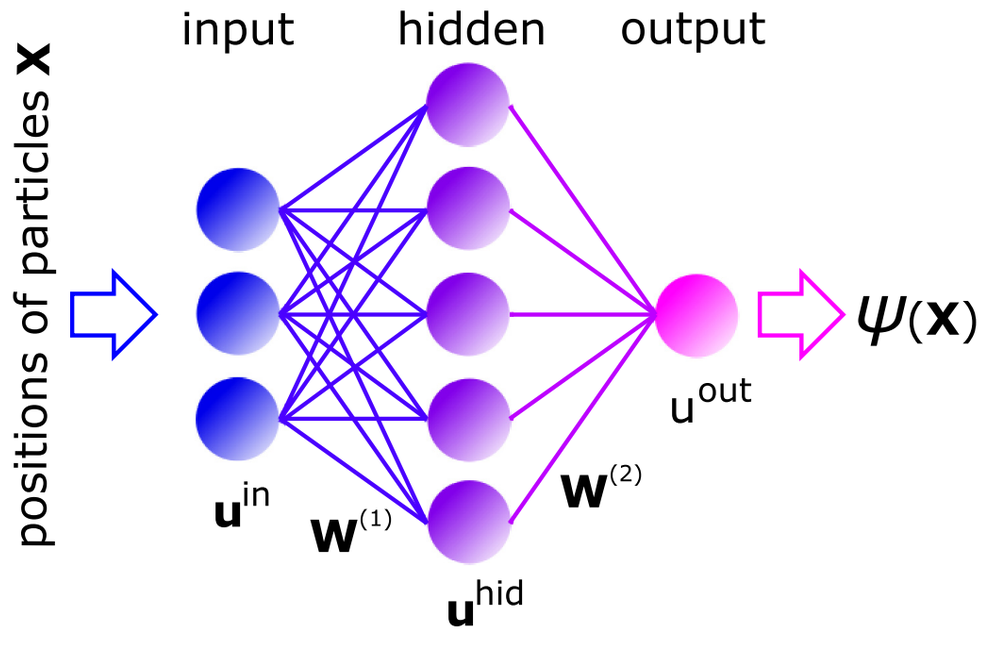



- Applications to the continuum to **few-body systems** and **quantum chemistry** problems have followed shortly thereafter;

H. Saito, J. Phys. Soc. Jpn. **87**, 074002 (2018)

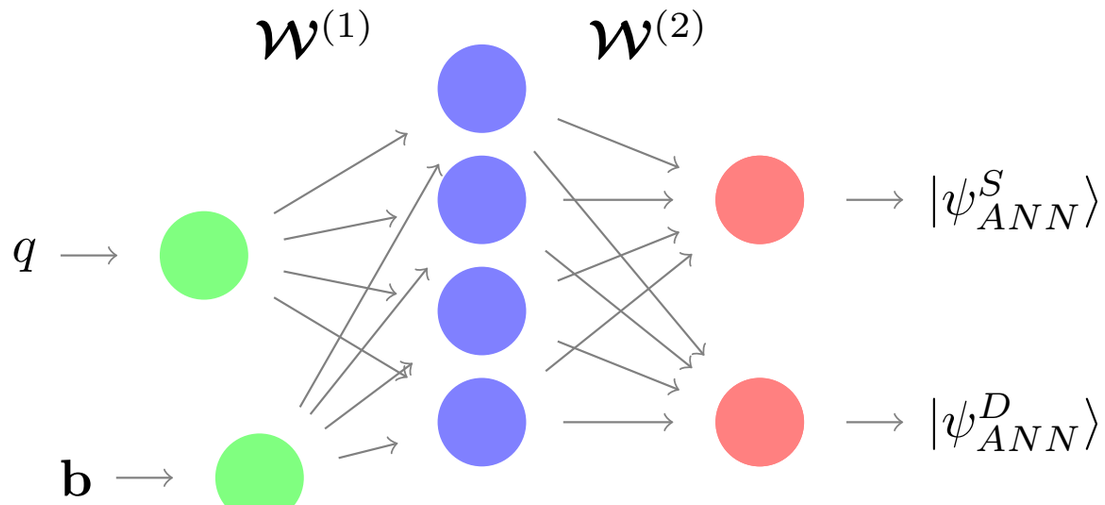
Pfau et al., arXiv:1909.02487 (2019)

Hermann et al., arXiv:1909.08423 (2019)

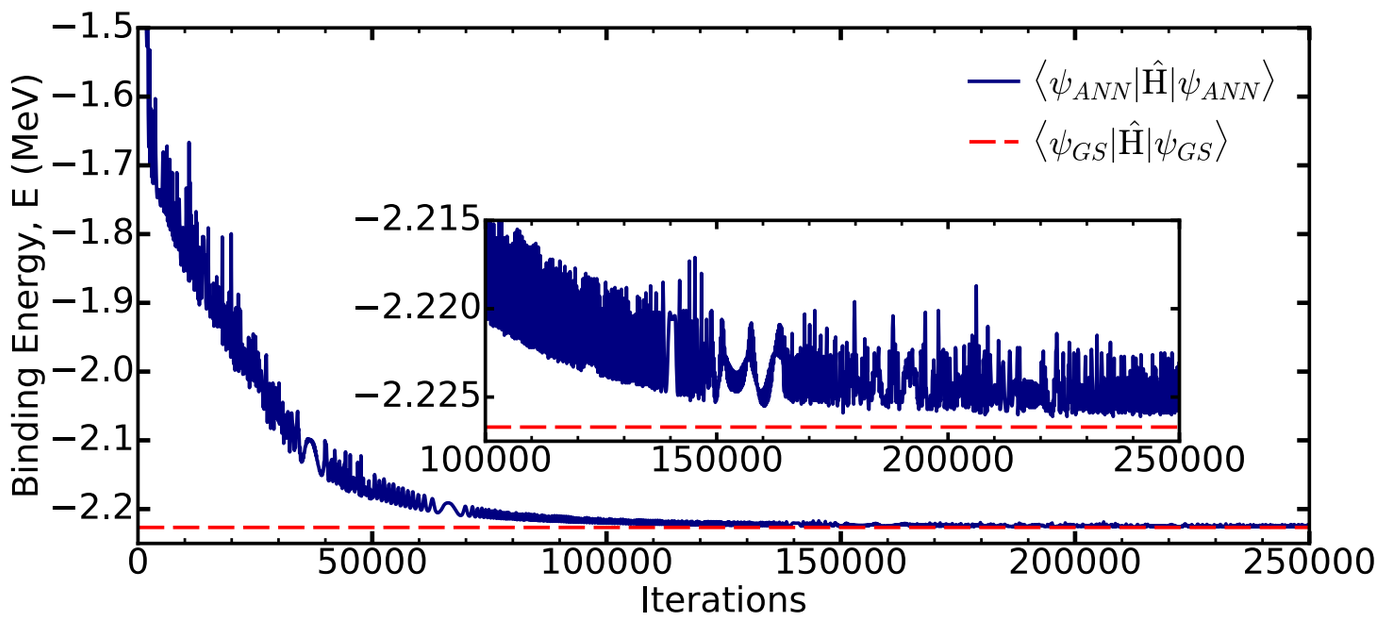


# Machine-learning the deuteron

- ANN were recently applied to solve the deuteron in momentum space using the sophisticated N3LO Entem-Machleidt chiral-EFT nucleon-nucleon force




- The parameters of the ANN are optimized minimizing the variational energy using RMSprop




**James Keeble talk**

Keeble, Rios, PLB **809**, 135743 (2020)

# Light nuclei with an ANN Jastrow ansatz

- We consider a the leading-order pionless-EFT Hamiltonian, which includes a three-body force

$$H_{LO} = -\sum_i \frac{\vec{\nabla}_i^2}{2m_N} + \sum_{i<j} \left(C_1 + C_2\, \vec{\sigma_i} \cdot \vec{\sigma_j}\right) e^{-r_{ij}^2 \Lambda^2/4} + D_0 \sum_{i<j<k} \sum_{\rm cyc} e^{-\left(r_{ik}^2 + r_{ij}^2\right)\Lambda^2/4}$$

- We introduce an ANN representation of the variational wave function of the form

$$|\Psi_V^{\rm ANN}\rangle = e^{U(\mathbf{r}_1,\ldots,\mathbf{r}_A)}|\Phi\rangle$$

- More general Ansatz than standard product of two- and three-body Jastrow terms

$$|\Psi_V^J\rangle = \prod_{i<j<k}\left(1 - \sum_{\rm cyc} u(r_{ij})u(r_{jk})\right)\prod_{i<j} f(r_{ij})|\Phi\rangle\,,$$

- For the s-shell nuclei that we consider, we assume the mean-field part to only depend upon the spin and isospin degrees of freedom

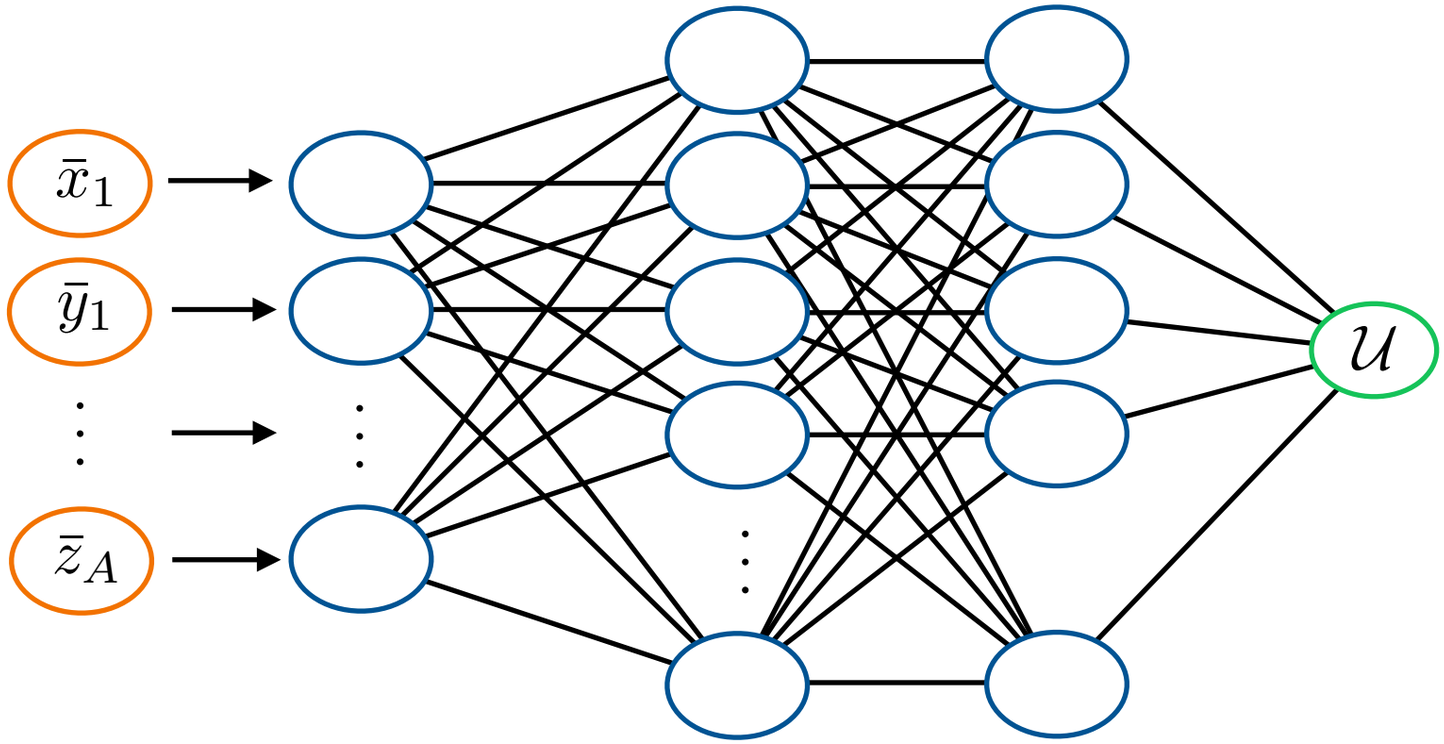
$$|\Phi_{^2{\rm H}}\rangle = \mathcal{A}|\uparrow p, \uparrow n\rangle \longrightarrow \sigma_{12}|\Phi_{^2{\rm H}}\rangle = |\Phi_{^2{\rm H}}\rangle$$

$$|\Phi_{^3{\rm He}}\rangle = \mathcal{A}|\uparrow p, \downarrow p, \uparrow n\rangle \longrightarrow \sigma_{12}|\Phi_{^3{\rm He}}\rangle = -|\Phi_{^3{\rm He}}\rangle$$

$$|\Phi_{^4{\rm He}}\rangle = \mathcal{A}|\uparrow p, \downarrow p, \uparrow n, \downarrow n, \rangle \longrightarrow \sigma_{12}|\Phi_{^4{\rm He}}\rangle = -|\Phi_{^4{\rm He}}\rangle$$

# Light nuclei with an ANN Jastrow ansatz

The correlation factor is parameterized with an ANN comprised of four fully connected layers




- The center of mass contributions to the kinetic energy are removed by $\bar{\mathbf{r}}_i = \mathbf{r}_i - \mathbf{R}_{\mathrm{CM}}$
- The kinetic energy requires the derivatives of $\mathcal{U}$. We use differentiable **softplus activation functions.**
- A Gaussian function is added to effectively confine the nucleons within a finite volume

$$\mathcal{U}(\mathbf{r}_1, \ldots, \mathbf{r}_A) \longrightarrow \mathcal{U}(\mathbf{r}_1, \ldots, \mathbf{r}_A) - \alpha \sum_i \bar{\mathbf{r}}_i^2$$

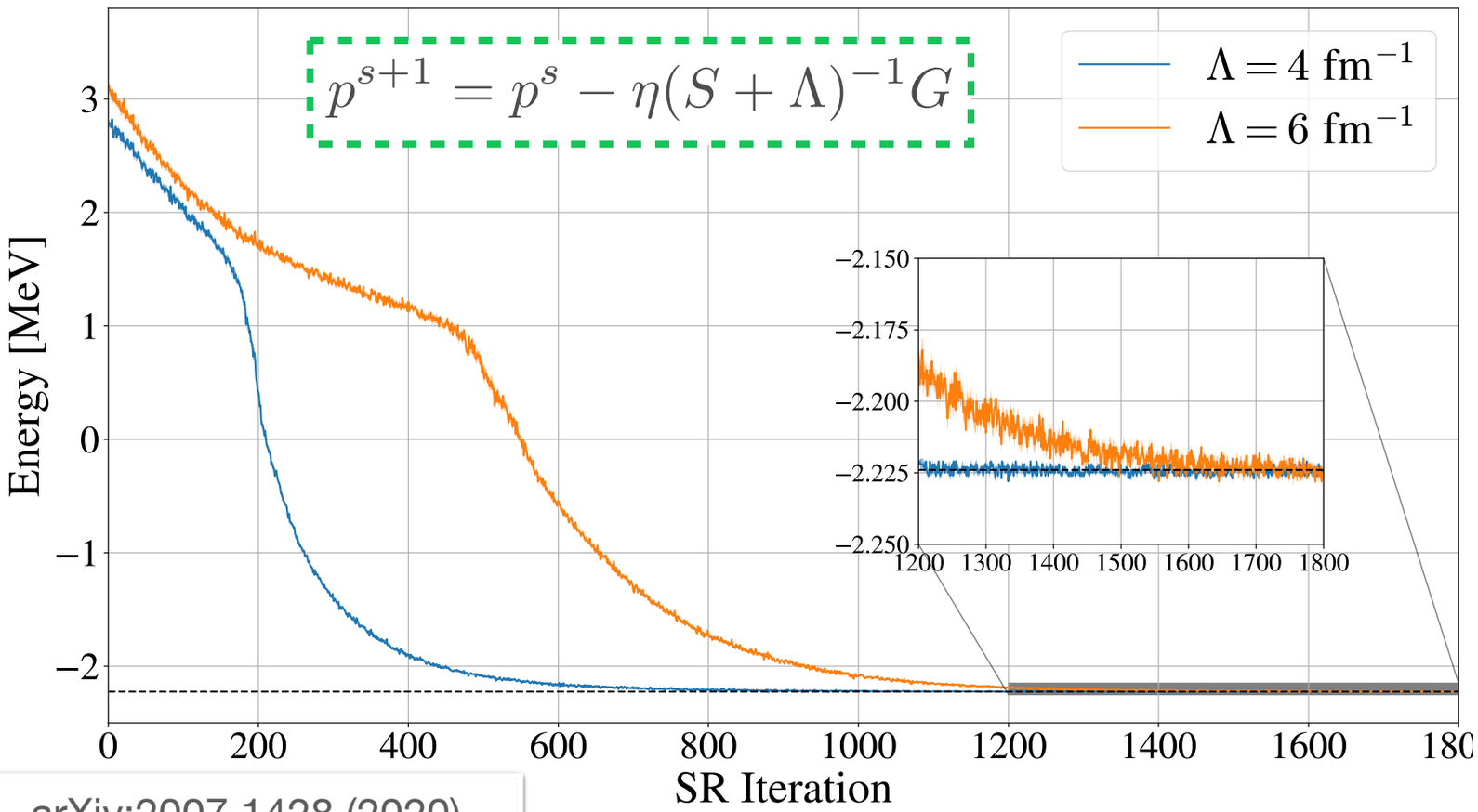
# Light nuclei with an ANN Jastrow ansatz

The optimal values of the **18304 trainable parameters** are found minimizing

$$E_V = \frac{\langle \Psi_V | H | \Psi_V \rangle}{\langle \Psi_V | \Psi_V \rangle}$$

Using a pre-conditioner based on the quantum Fisher information (analogous to the **stochastic reconfiguration method**) is more efficient than stochastic gradient descent

$$G_i = 2\left(\frac{\langle \partial_i \Psi_V | H | \Psi_V \rangle}{\langle \Psi_V | \Psi_V \rangle} - E_V \frac{\langle \partial_i \Psi_V | \Psi_V \rangle}{\langle \Psi_V | \Psi_V \rangle}\right) \quad S_{ij} = \frac{\langle \partial_i \Psi_V | \partial_j \Psi_V \rangle}{\langle \Psi_V | \Psi_V \rangle} - \frac{\langle \partial_i \Psi_V | \Psi_V \rangle \langle \Psi_V | \partial_j \Psi_V \rangle}{\langle \Psi_V | \Psi_V \rangle \langle \Psi_V | \Psi_V \rangle}$$



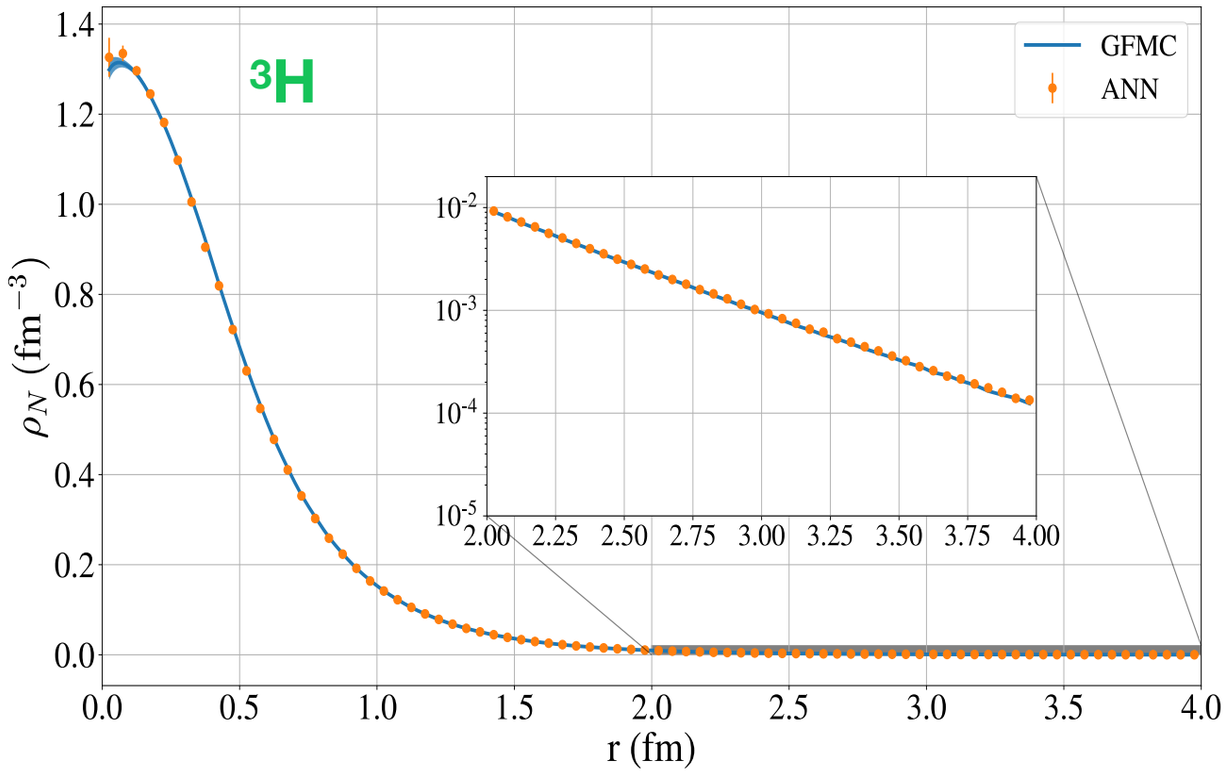

C. Adams, AL, et al., arXiv:2007.1428 (2020)

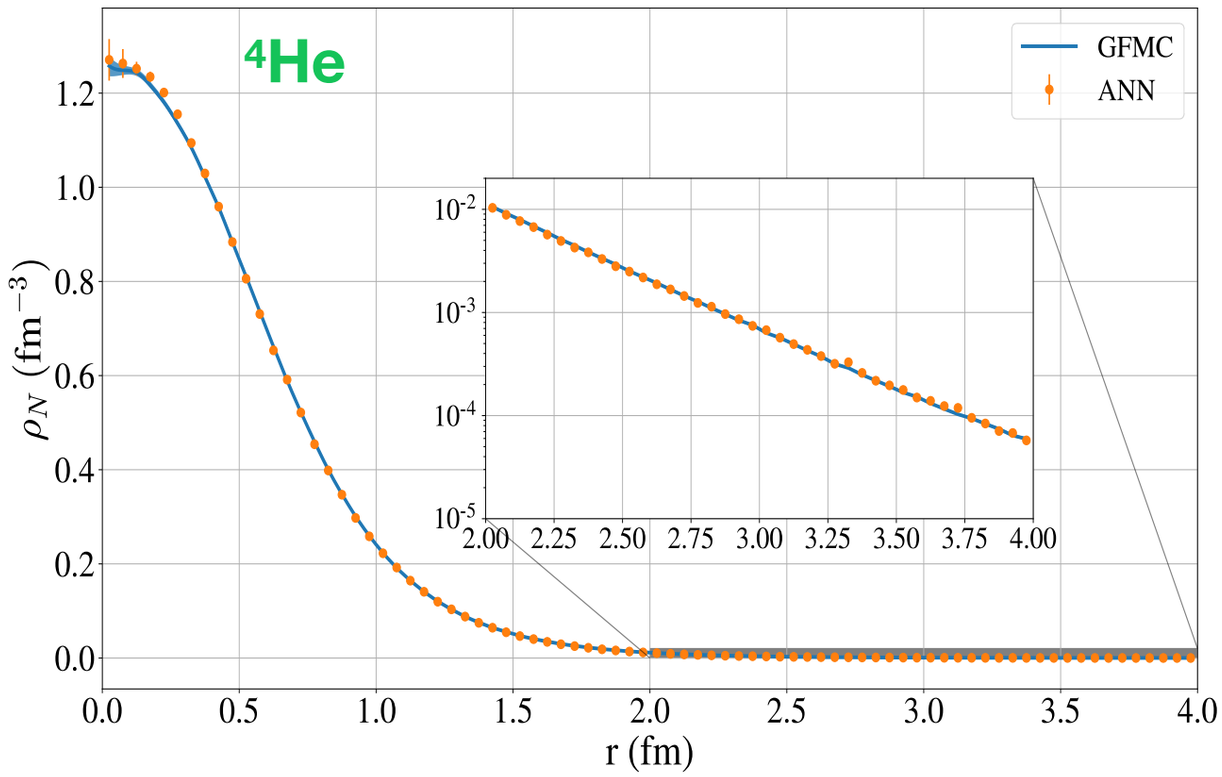
# Light nuclei with an ANN Jastrow ansatz

- The ANN ansatz outperforms standard Jastrow correlations, but lacks spin-isospin correlations

| | $\Lambda$ | VMC-ANN | VMC-JS | GFMC | $\text{GFMC}_c$ |
|---|---|---|---|---|---|
| $^2$H | 4 fm$^{-1}$ | $-2.224(1)$ | $-2.223(1)$ | $-2.224(1)$ | - |
| | 6 fm$^{-1}$ | $-2.224(4)$ | $-2.220(1)$ | $-2.225(1)$ | - |
| $^3$H | 4 fm$^{-1}$ | $-7.81(1)$ | $-7.80(1)$ | $-8.38(2)$ | $-7.82(1)$ |
| | 6 fm$^{-1}$ | $-7.79(3)$ | $-7.74(1)$ | $-8.38(2)$ | $-7.81(1)$ |
| $^4$He | 4 fm$^{-1}$ | $-22.76(3)$ | $-22.54(1)$ | $-23.62(3)$ | $-22.77(2)$ |
| | 6 fm$^{-1}$ | $-23.99(6)$ | $-23.44(2)$ | $-25.06(3)$ | $-24.10(2)$ |

- Excellent results for ground-state energies and for single-nucleon densities




C. Adams, AL, et al., arXiv:2007.1428 (2020)

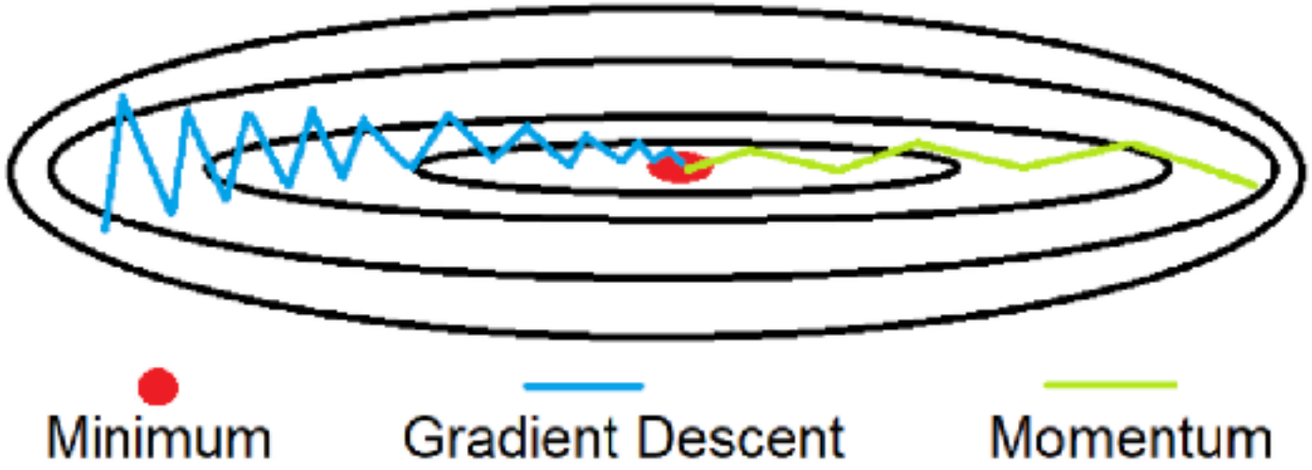
# Some perspectives

We are improving the convergence of the training combining SR with state-of-the-art optimization methods developed in ML applications;

$$G_i \longrightarrow (S+\Lambda)^{-1}_{ij} G_i$$

As a first example, we considered the "**Momentum**" method, which helps accelerate SGD in the relevant direction and dampens oscillations;

$$\begin{cases} v^s = (S^s + \Lambda)^{-1}_{ij} G^s + \gamma v^{s-1} \\ p^{s+1} = p^s - v^s \end{cases}$$




With momentum we already achieved a tenfold speedup in the convergence. We plan to test **RMSprop** and **Adam**, two of the most popular optimization methods.

# Some perspectives

Including spin-dependent correlations requires encoding the anti-symmetry of the wave function.
We are investigating the use of a **generalized back-flow ansatz**

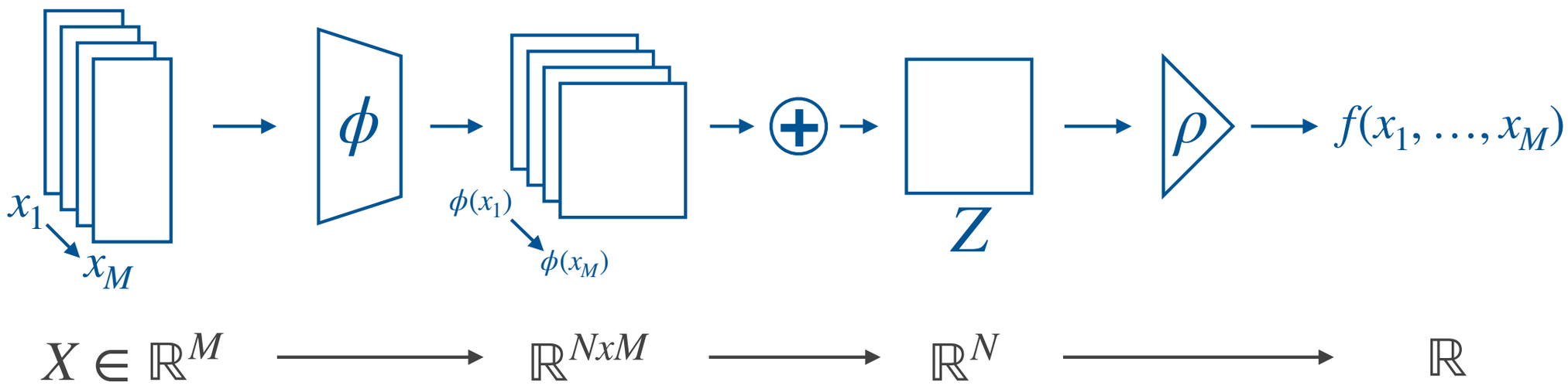
$$\hat{\psi}(X) = \det \begin{pmatrix} \phi_1(x_1; x_{j\neq 1}) & \phi_1(x_2; \{x_{j\neq 2}\}) & \dots & \phi_1(x_A; \{x_{j\neq A}\}) \\ \phi_2(x_1; x_{j\neq 1}) & \phi_2(x_2; \{x_{j\neq 2}\}) & \dots & \phi_2(x_A; \{x_{j\neq A}\}) \\ \vdots & \vdots & \dots & \vdots \\ \phi_A(x_1; x_{j\neq 1}) & \phi_A(x_2; \{x_{j\neq 2}\}) & \dots & \phi_A(x_A; \{x_{j\neq A}\}) \end{pmatrix},$$

Hermann et al., arXiv:1909.08423 (2019)

Pfau et al., arXiv:1909.02487 (2019)

- We introduced the generalized coordinate $x_i = \{\mathbf{r}_i, s_i^z, t_i^z\}$

- The functions $\phi_i(x_i; x_{j\neq i})$ must be invariant under the exchange of the order of $x_{j\neq i}$: need to Generalization of the **Deep Set architecture** to the complex case




Wagstaff et al., arXiv:1901.09006 (2019)

# Some perspectives

Our primary goal is to devise **accurate wave functions for medium-mass nuclei** otherwise inaccessible by conventional nuclear QMC methods;

Compact representations useful to both theory and experimental communities

ANN allows study **properties of excited states**, a central task and nontrivial challenge for several nuclear many-body quantum approaches;

K. Choo et al. Phys. Rev. Lett. **121**, 167204 (2018)

ANN can be extended to solve the time-dependent Schrödinger equation, similarly to the time-dependent variational Monte Carlo method. The distance to be minimized is

$$\min_{\mathbf{p}} \left\| i\frac{\partial \Psi_{\mathrm{V}}^{\mathrm{ANN}}}{\partial t} - H\Psi_{\mathrm{V}}^{\mathrm{ANN}} \right\|$$

Response function calculations; interplay with quantum-computing methods

# Summary and outlook

- We have developed a noise-resilient **ML-based protocol for inverting the Laplace transform**
  - More accurate than Maximum Entropy, especially in the low energy-transfer region;
  - Resilient to noise in the input Euclidean response, see Krishnan's talk;
  - Applicable to other integral-transform, including the Lorentz and the Gauss;

- Encouraging results with **ANN representation of nuclear variational states**
  - The ANN correlation outperforms existing two- and three-body Jastrow ansatz;
  - Using the SR method enables relatively fast training, further improved when used in combination with state-of-the-art ML optimization methods;
  - Correct behavior of single-particle densities, including the slowly decaying tail;
  - Extension to spin-isospin dependent correlation (and to larger nuclei) requires encoding permutation-invariant symmetry;
  - Access to excited states and nuclear response functions;