Variational Monte Carlo and Machine Learning

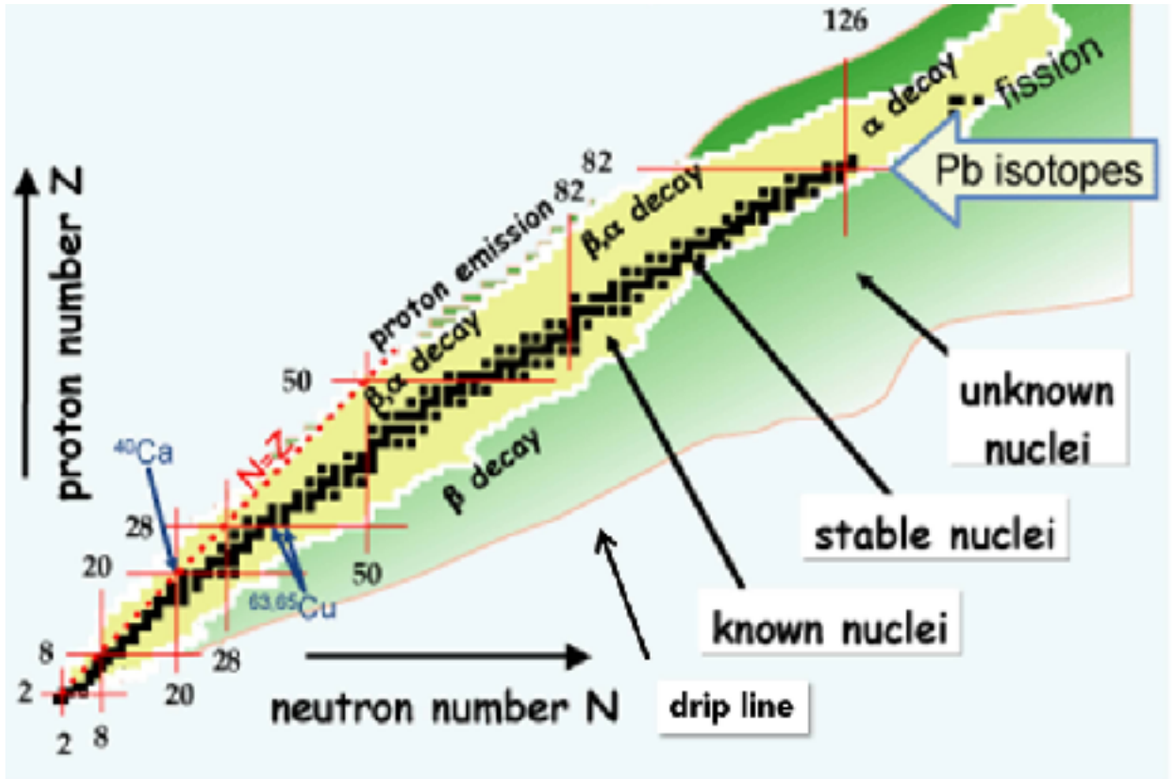
Artificial Intelligence for Nuclear Physics Winter School
January 12, 2021

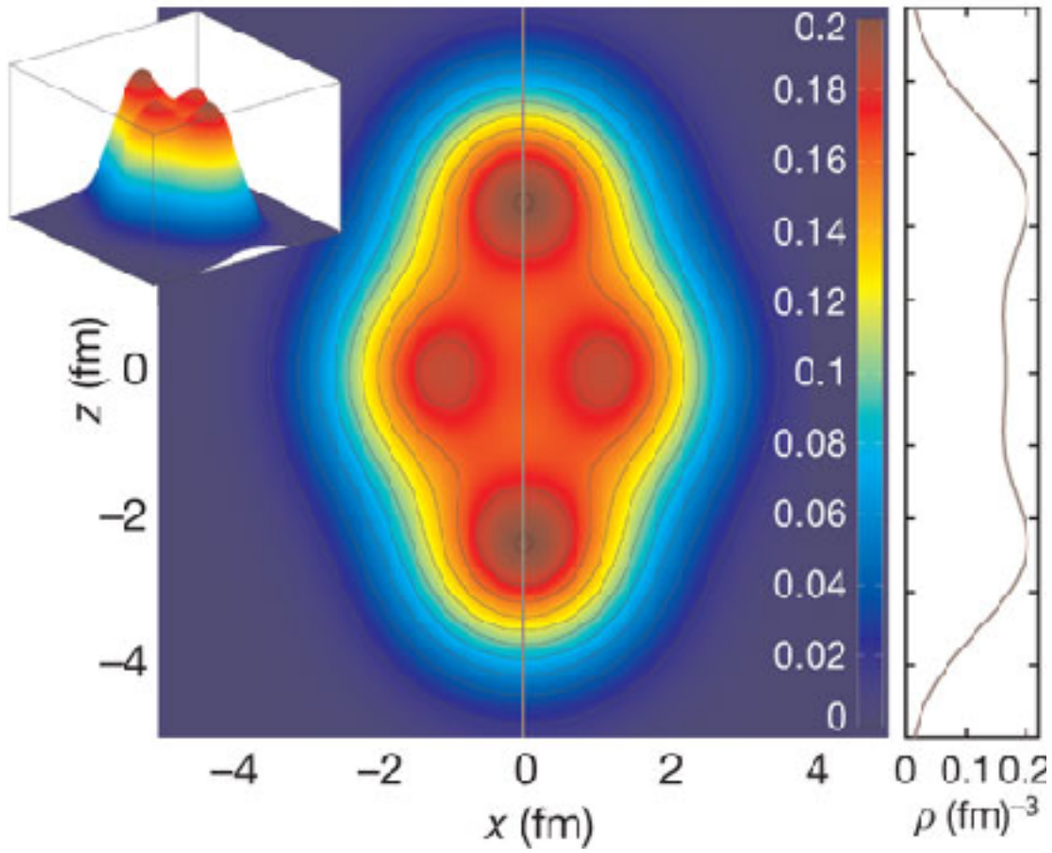
Corey Adams & Alessandro Lovato

Argonne NATIONAL LABORATORY

Ab-initio nuclear theory

- Atomic nuclei are strongly interacting many-body systems exhibiting self emerging properties including: shell structure, pairing and superfluidity, deformation, and clustering.



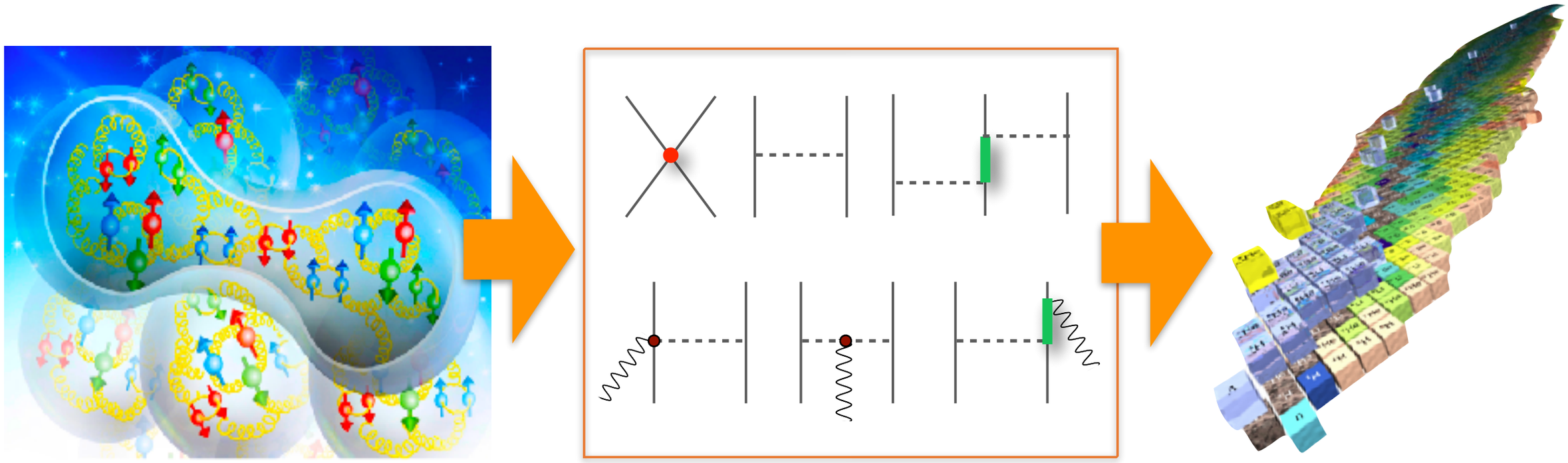


• Understanding their structure, reactions, and electroweak properties within a unified framework well-rooted in quantum chromodynamics has been a long-standing goal of nuclear physics.

From QCD to nuclear observables

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

Effective field theories are the link between quantum chromodynamics and nuclear observables.



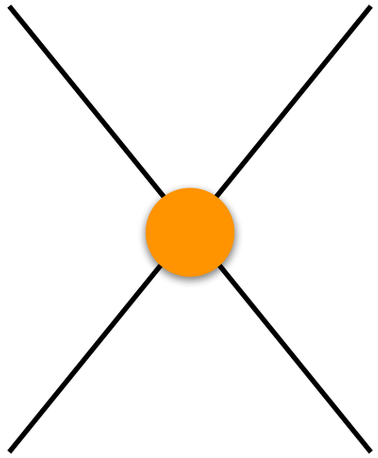
Nucleons can treated as point-like particles interacting through a non-relativistic Hamiltonian

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

A pion-less nuclear Hamiltonian

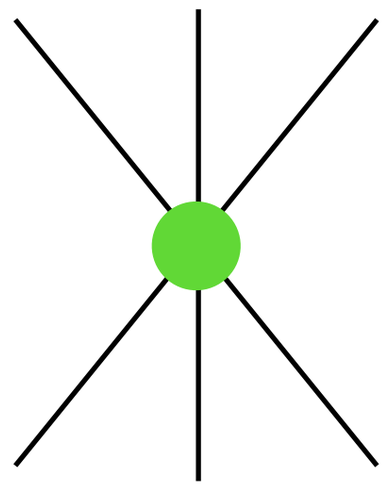
A prototypal nuclear Hamiltonian can be obtained within **pionless effective field theory**, where nucleons interact via two- and three-body contact potentials. At leading order:

$$v_{12} = C_1 v_\Lambda(r_{12}) + C_2 v_\Lambda(r_{12})\sigma_{12}$$



C_1 and C_2 are determined on nucleon-nucleon scattering data

$$V_{123} = D_0 \sum_{\text{cyc}} v_\Lambda(r_{12}) v_\Lambda(r_{13})$$



The parameter D_0 is fixed on the binding energy of ^{3}H and prevent its collapse

Despite its simplicity, solving this Hamiltonian entails most of the difficulties encountered when dealing with sophisticated chiral effective field theory potentials;

The nuclear many-body problem

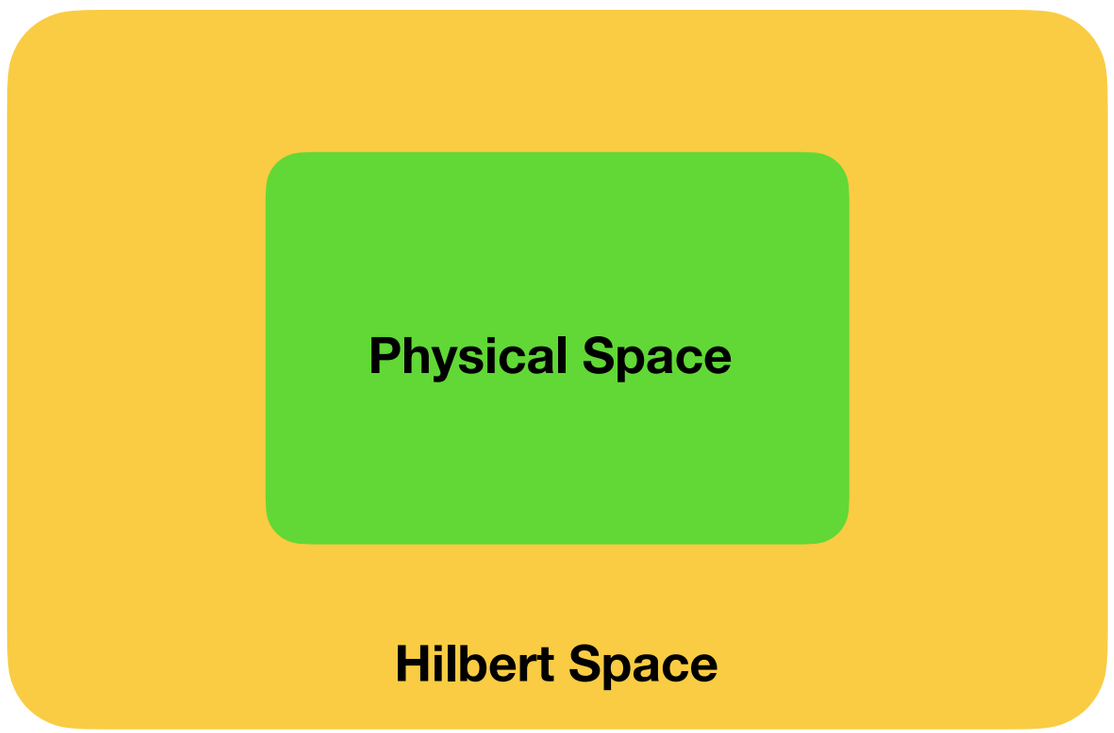
Non relativistic many body theory is aimed at solving the Schrödinger equation

$$H\Psi_n(x_1,\ldots,x_A) = E_n\Psi_n(x_1,\ldots,x_A) \quad \longleftrightarrow \quad x_i = \{\mathbf{r}_i, s_{i,z}, t_{i,z}\}$$

An exact solution of this equation is an **exponentially hard problem**

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

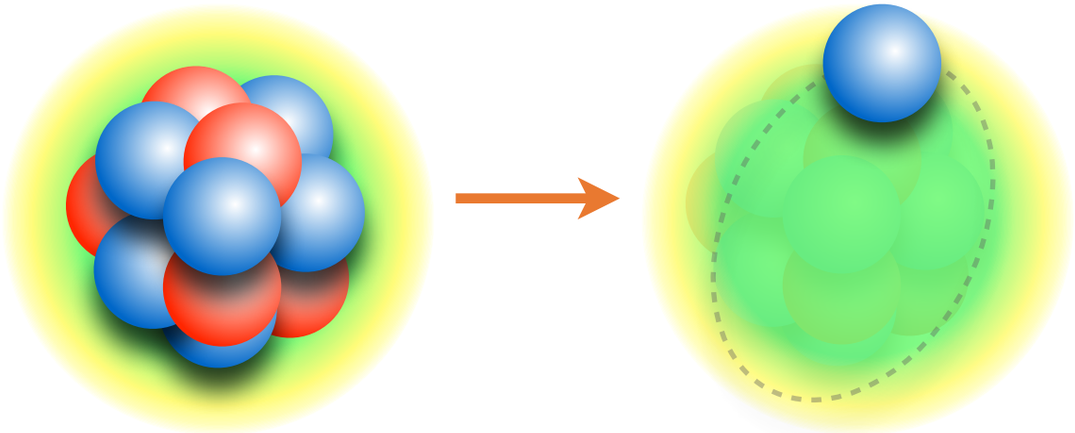
The majority of quantum states of interest for have **distinctive features and intrinsic structure**.



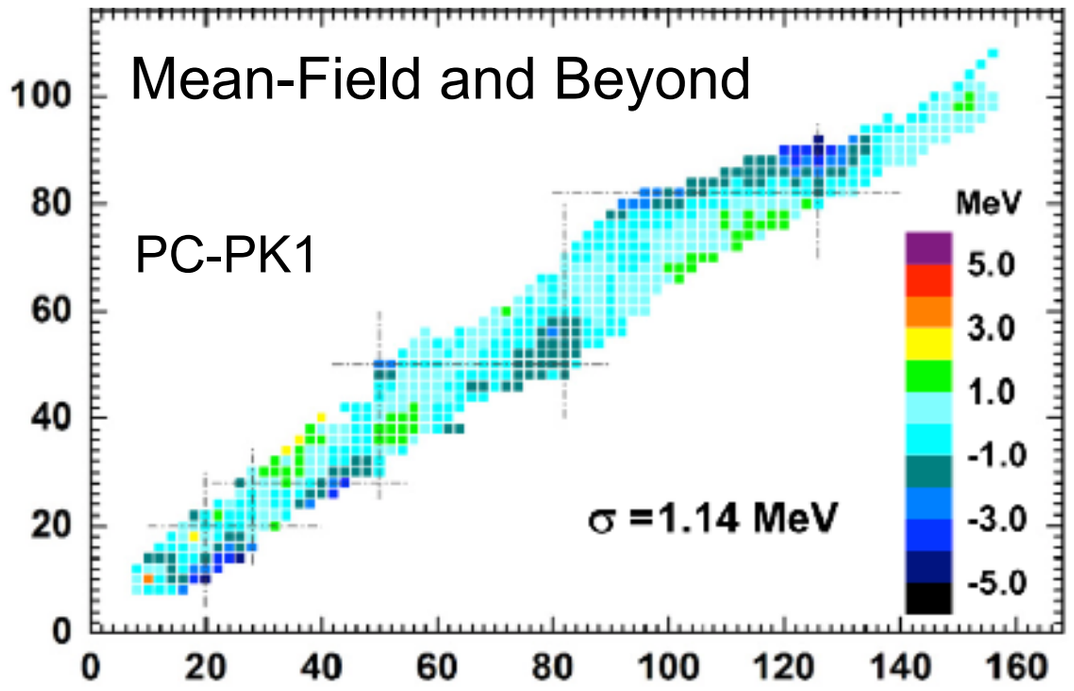
Mean-field approximations

• Mean field theory: nucleons are independent particles subject to an average nuclear potential

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



• The interaction is usually fitted on nuclear binding energies and charge radii of stable nuclei



P. W. Zhao, et al. PRC **82**, 054319 (2010)

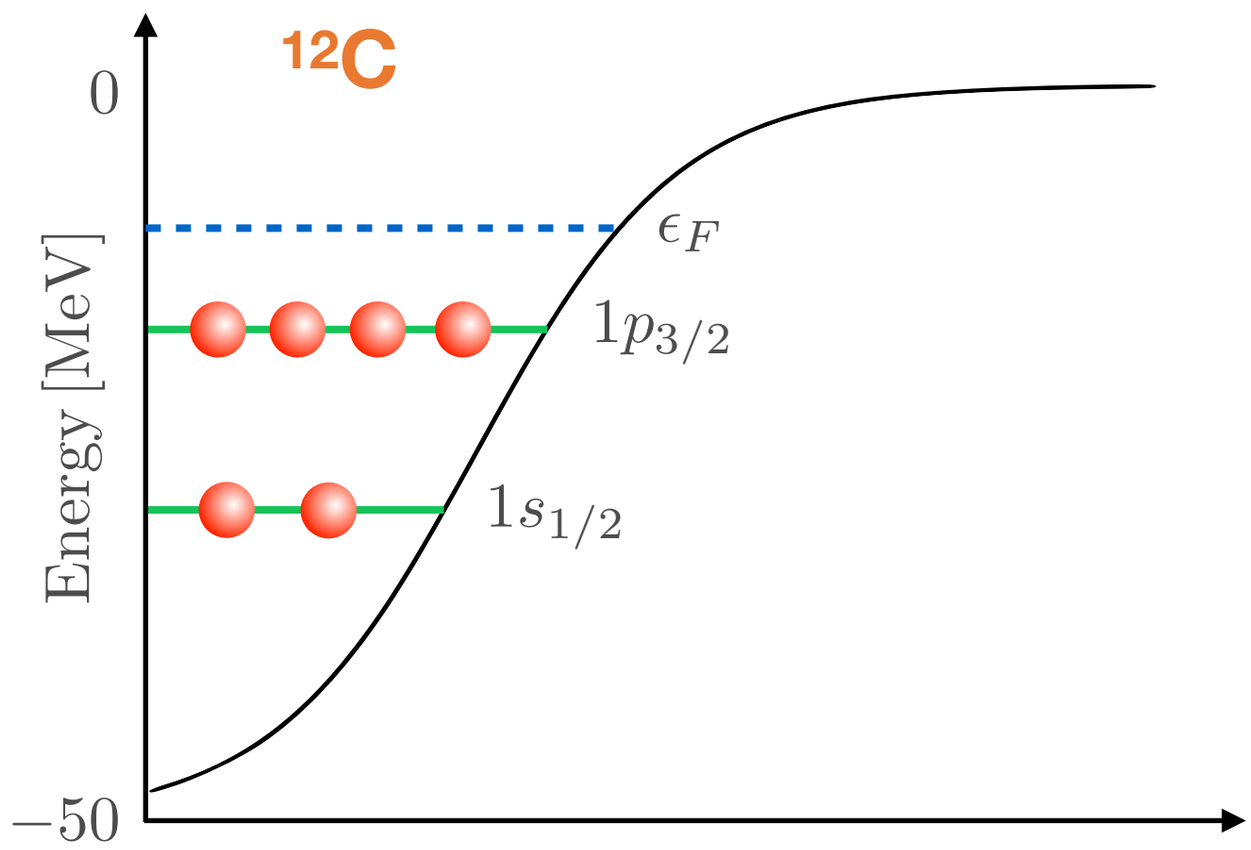
- Despite being the **tool of choice for describing large nuclei**:
  - ✳ Nucleon-nucleon scattering data and deuteron properties are ignored
  - ✳ There is no clear way to derive effective currents
  - ✳ The average procedure depends upon the (large) system of interest

Many-body wave function

Mean field methods assume the ground-state wave function to be a Slater determinant of single-particle waves functions

$$\Phi_0(x_1, \ldots, x_A) = \mathcal{A}[\phi_{n_1}(x_1) \ldots \phi_{n_A}(x_A)]$$

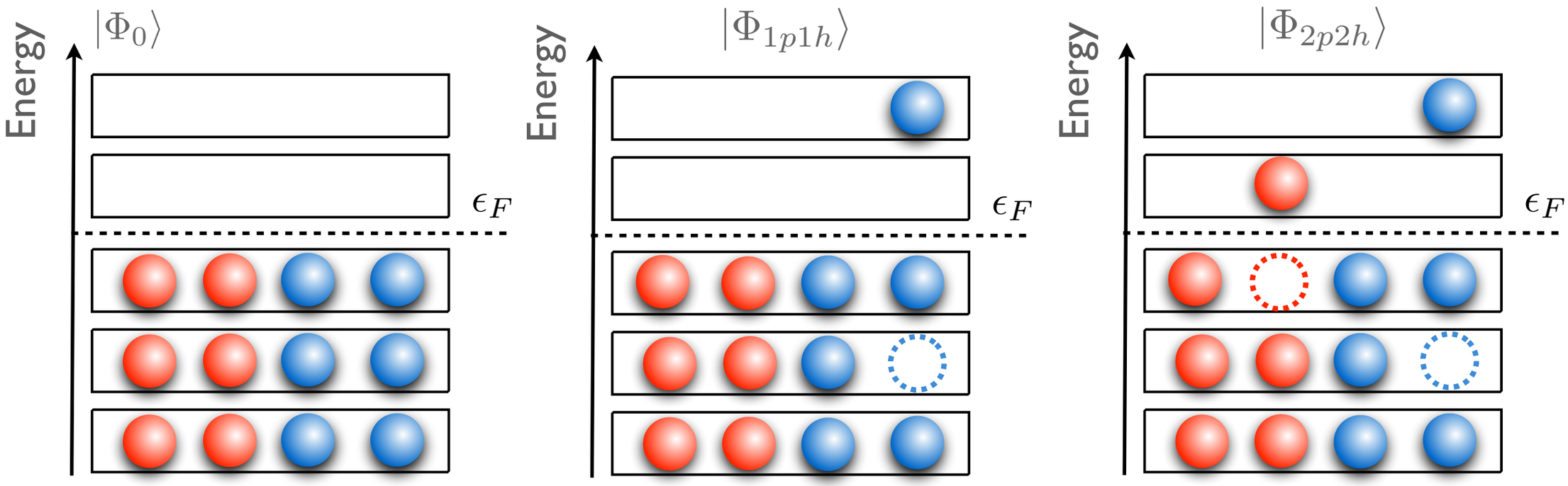
Extreme “single-particle” representations can be obtained using harmonic oscillator or Wood-Saxon single-particle potentials



Many-body wave function

Excited states are constructed removing *n* occupied states from the Slater determinant and replacing them with *n* virtual (unoccupied) states

$$\Phi_{p_i,h_i}(x_1 \dots x_A) = \mathcal{A}[\phi_{n_1}(x_1) \dots \phi_{p_i}(x_i) \dots \phi_{n_A}(x_A)]$$



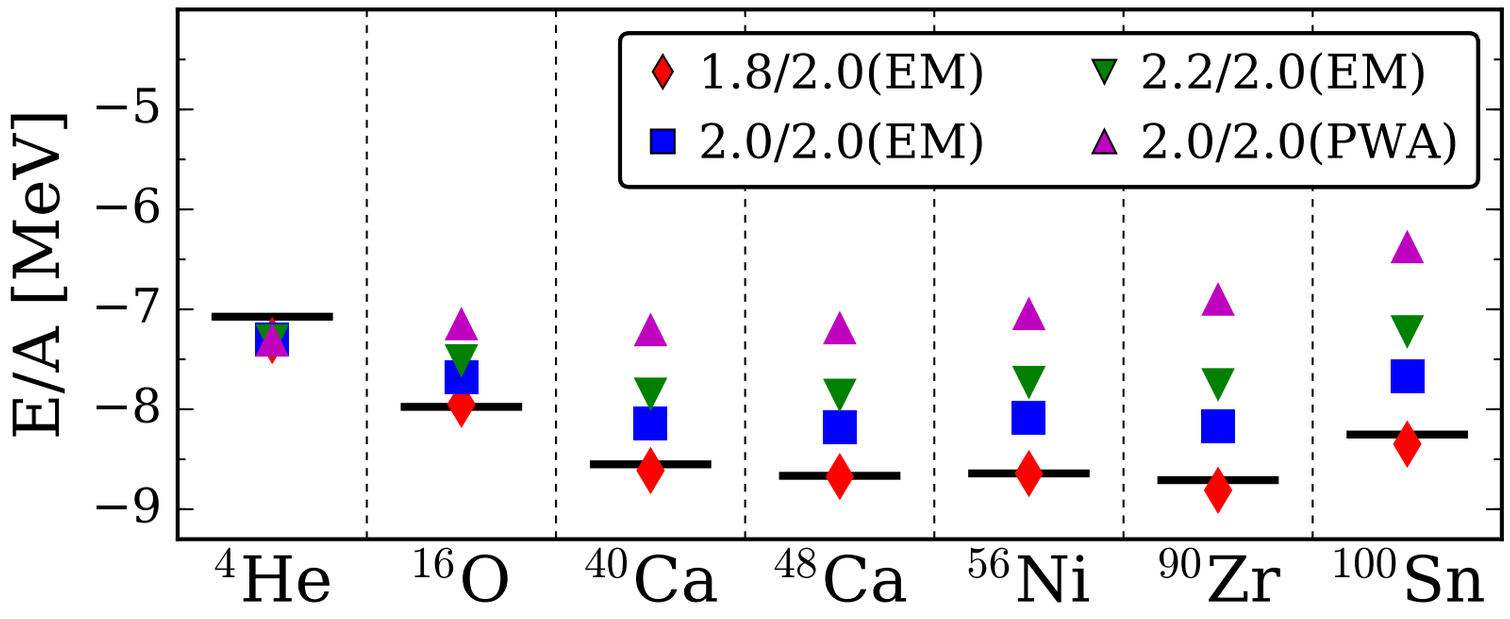
The eigenstate of the Hamiltonian is a **linear combination of n-particles n-holes states**

$$|\Psi_n\rangle = \sum c^n_{p_i,h_i} |\Phi_{p_i,h_i}\rangle \qquad H|\Psi_n\rangle = E_n|\Psi_n\rangle$$

Many-body wave function

Methods relying on single-particle basis expansions include the **no-core shell model**, the **coupled-cluster theory**, the **in-medium similarity renormalization group method**

Nuclei with up to to A=100 protons and neutrons can be described starting from the individual interactions among their constituents



T. D. Morris at al., PRL **120**, 152503 (2018)

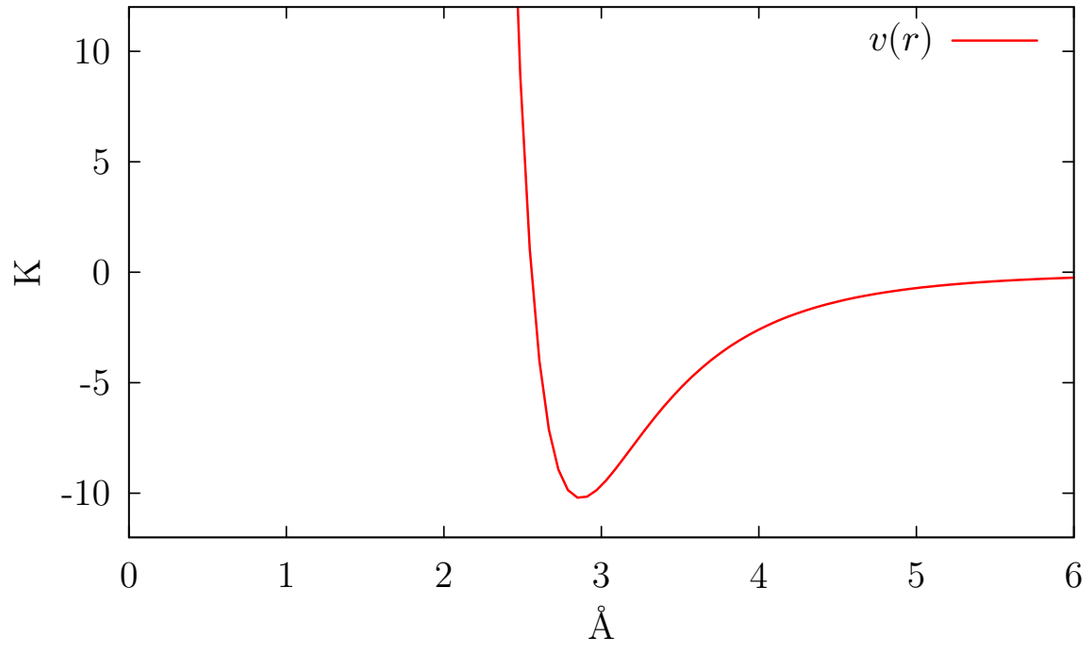
Single-particle basis expansion methods have difficulties in **modeling nuclear short-range dynamics**, relevant for electron- and neutrino-nucleus scattering;

Variational Monte Carlo

Variational Monte Carlo uses the **stochastic integration method** to evaluate the expectation value of the Hamiltonian for a chosen trial wave function

The interaction between ^{4}He atoms forming an homogeneous liquid can be parametrized by means of the two-body **Lennard-Jones potential**

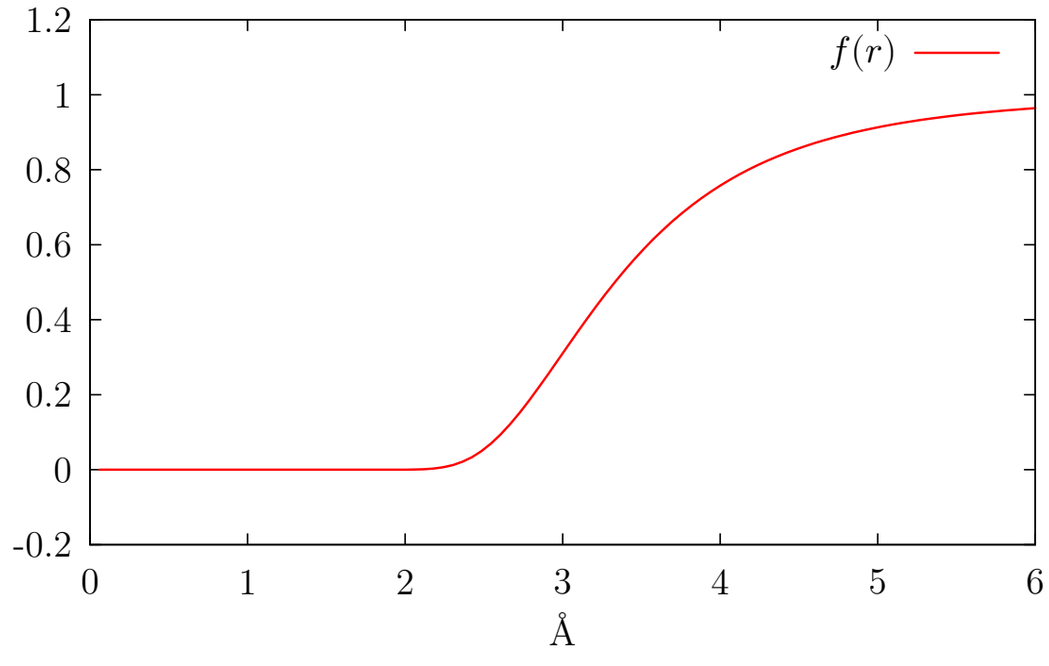
$$v(r) = 4\epsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right]$$



A reasonable trial wave function is small where the potential is repulsive and large where the potential is attractive

$$\Psi_T(R) = \prod_{i<j} f(r_{ij})$$

$$f(r) = \exp\left[-\frac{1}{2}\left(\frac{b}{r}\right)^{5}\right]$$



Variational Monte Carlo

The **variational principle** guarantees that the trial energy is greater than or equal to the ground-state energy with the same quantum numbers

$$E_T = \frac{\langle \Psi_T | H | \Psi_T \rangle}{\langle \Psi_T | \Psi_T \rangle} \geq E_0$$

The variational parameters are found **minimizing the energy**. In the liquid ^{4}He case, where the wave-function is determined by a single parameters, this amounts to finding b such that

$$\frac{\partial E_T}{\partial b} = 0$$

Computing the trial energy for a given set of variational parameters requires evaluating a **high-dimensional integral**

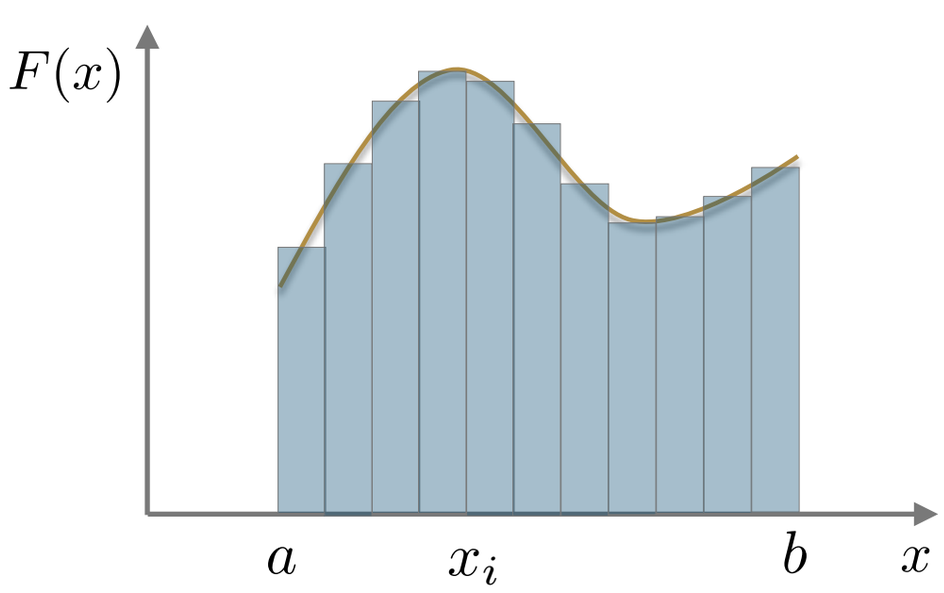
$$E_T = \frac{\int dR\, \Psi_T^*(R) H \Psi_T(R)}{\int dR\, \Psi_T^*(R) \Psi_T(R)} \qquad R \equiv \mathbf{r}_1, \ldots, \mathbf{r}_A$$

Multi-dimensional integrals

Goal: compute the D-dimensional integral

$$I(D) = \int_{a_1}^{b_1} dx_1 \ldots \int_{a_D}^{b_D} dx_D F(x_1, \ldots, x_D)$$

In the one-dimensional case, we can divide the area below to the curve into rectangles



$$\begin{cases} I(1) \simeq h \sum_i F(x_i) \\ \Delta(1) = h^2 |F'(x_i)| + O(h^3) \\ h \propto \frac{1}{N} \end{cases}$$

How many points do we need to achieve a given precision ?

$$\frac{\Delta(1)}{I(1)} = \epsilon \longrightarrow N \propto \frac{1}{\epsilon}$$

Multi-dimensional integrals

In the D-dimensional case, the following relations hold:

$$\begin{cases} I(D) \simeq h^D \sum_i F(x_i) \\ \Delta(D) = h^{D+1} |\nabla F(x_i)| + O(h^{D+2}) \\ h \propto \dfrac{1}{N^D} \end{cases} \longrightarrow \quad N \propto \frac{1}{\epsilon^D}$$

Suppose we want to compute the expectation value of the Hamiltonian for a 12-particles system with a precision 0.1

In this case we are dealing with 12 x 3 = 36-dimensional integral

$$D = 36 \longrightarrow N \propto 10^{36}$$

Clearly, numerical quadrature cannot be used to compute the Hamiltonian expectation values of many-particle systems

The central limit theorem

Suppose that the N continuum random variables $x_1, \ldots, x_N$ are drawn from the probability distribution $P(x)$ and consider the function $f(x)$. We may define a new random variable

$$S_N = \frac{1}{N} \sum_{i=1}^{N} f(x_i)$$

If the samples are statistically independent, the **central limit theorem** states that the probability distribution of S_N is Gaussian

$$P(S_N) = \frac{1}{\sqrt{2\pi\sigma_N^2}} e^{\frac{(S_N - \bar{S}_N)^2}{2\sigma_N^2}}$$

where the average and the variance of S_N are given by

$$\bar{S}_N = \int dx P(x) f(x) \qquad \sigma_N = \sqrt{\frac{1}{N}\left[\int dx P(x) f^2(x) - \bar{S}_N\right]}$$

These results **hold true for any dimensionality of the space** in which the variable x is defined

The central limit theorem

The central limit theorem provides a recipe to **evaluate multi-dimensional integrals** of the form

$$I = \int dx f(x)$$

- Since the probability density has to be positive definite, rewrite the integral as:

$$I = \int dx P(x) \frac{f(x)}{P(x)}$$

- Sample N (with N “large”) points from the probability density $P(x)$
- Average the N values of $f(x_i)$ and $f^2(x_i)$

$$I = \frac{1}{N}\sum_{i=1}^{N} f(x_i) \pm \sqrt{\frac{1}{(N-1)}\left[\frac{1}{N}\sum_{i=1}^{N} f^2(x_i) - \left(\frac{1}{N}\sum_{i=1}^{N} f(x_i)\right)^2\right]}$$

This method provides an **estimate of both the integral and of its uncertainty**

Variational Monte Carlo

The expectation value of the Hamiltonian for a system of A particles is given by

$$E_T = \frac{\langle \Psi_T | H | \Psi_T \rangle}{\langle \Psi_T | \Psi_T \rangle} = \frac{\int dR \langle \Psi_T | R \rangle \langle R | H | \Psi_T \rangle}{\int dR \langle \Psi_T | R \rangle \langle R | \Psi_T \rangle} = \frac{\int dR |\Psi_T(R)|^2 E_L(R)}{\int dR |\Psi_T(R)|^2}$$

Using the central limit theorem, we can estimate the energy expectation value as

$$\langle E_T \rangle = \frac{1}{N_s} \sum_{R_n} E_L(R_n) \quad \longleftrightarrow \quad E_L(R) \equiv \frac{H\Psi_T(R)}{\Psi_T(R)}$$

Where the walkers are sampled from

$$P(R) = \frac{|\Psi_T(R)|^2}{\int dR |\Psi_T(R)|^2}$$

The integration error can also be readily estimated by

$$\sigma_{E_T} = \sqrt{\frac{\langle E_T^2 \rangle - \langle E_T \rangle^2}{N_s - 1}} \quad \longleftrightarrow \quad \langle E_T^2 \rangle = \frac{1}{N_s} \sum_{R_n} E_L^2(R_n)$$

$M(RT)^2$ algorithm

The algorithm was first described in a paper by Metropolis, Rosenbluth, Rosenbluth, Teller and Teller $M(RT)^2$. It shares common features to the rejection techniques because:

- It involves explicitly proposing a tentative value of the variable we want to sample, which may be rejected.
- The normalization of the sampled function is irrelevant.

$M(RT)^2$ algorithm has its own advantages and disadvantages:

Pros

- It can be used to sample essentially any density function regardless of analytic complexity in any number of dimensions
- It is of very great simplicity.
- It “embarrassingly parallel”

Cons

- Sampling is correct only asymptotically
- Consecutive variables produced are often very strongly correlated
- Many sampled configurations are disregarded

M(RT)^2 algorithm

Our goal is to sample the probability distribution described by

$$P(x) \equiv |\Psi_T(x)|^2$$

The M(RT)^2 algorithm is **based on the idea of random walk**. A set of random configurations (or walkers) are generated by applying the transformation.

$$P_{i+1}(x_{i+1}) = \int dx_i P_i(x_i) T(x_i \to x_{i+1})$$

Transition probability

By recursively applying the same transformation we get

$$P_n(x_n) = \int dx_1 \dots dx_{n-1} P_1(x_1) T(x_1 \to x_2) \dots T(x_{n-1} \to x_n)$$

Under some very general conditions it can be proven that

$$\lim_{n\to\infty} P_n(x_n) = P(x) \longrightarrow P(x) \text{ only depends on } T$$

M(RT)^2 algorithm

Let us impose a further condition, i.e. that the asymptotic distribution is an “equilibrium” state:

$$P(x)T(x \to y) = P(y)T(y \to x)$$

The latter is called **detailed balance condition**, because it does not hold only on average, but it tells that point by point there is no net flux of probability

We can arbitrarily split the transition probability in two terms

$$T(x \to y) = G(x \to y)A(x \to y)$$

It describes the probability of moving the walker from $x \to y$.

It tells wether the proposed move is accepted or rejected.

The detailed balance then reads

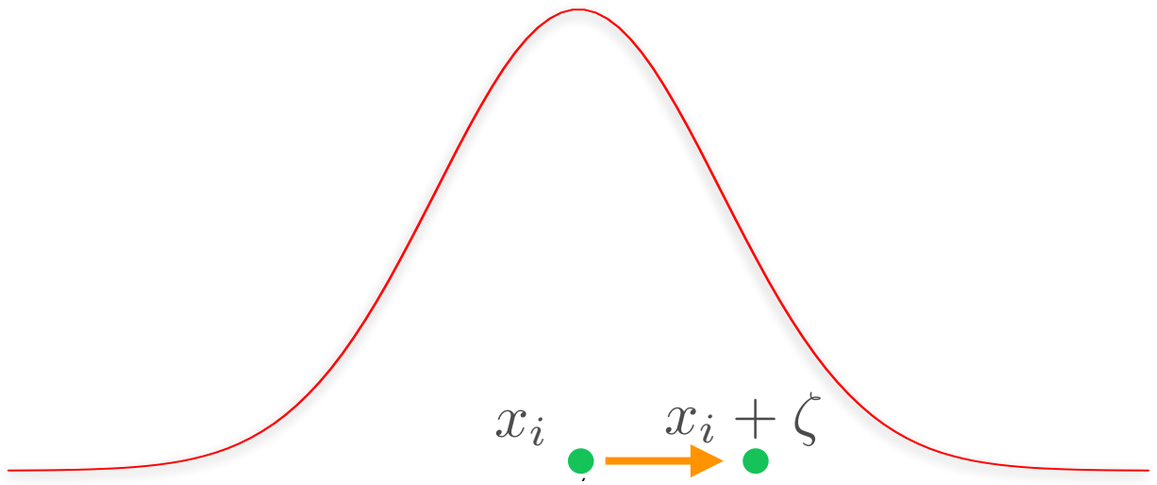
$$\frac{A(y \to x)}{A(x \to y)} = \frac{P(x)G(x \to y)}{P(y)G(y \to x)}$$

It can be easily checked that the following acceptance probability satisfies the above requirement

$$A(y \to x) = \min\left(1, \frac{P(x)G(x \to y)}{P(y)G(y \to x)}\right)$$

M(RT)^2 algorithm

A common choice for $G(x \to y)$ is a Gaussian distribution centered in zero. In this case, at each step of the propagations, the walkers are moved by $x_{i+1} = x_i + \zeta$



In the multiple-particle case, the one dimensional gaussian is replaced by a three-dimensional gaussian for each particle.

Since the probability of going from x to y is equal to the one of going from y to x, it turns out that

$$G(x \to y) = G(y \to x) \quad \longleftrightarrow \quad A(y \to x) = \min\left(1, \frac{P(x)}{P(y)}\right)$$

M(RT)^2 applied to VMC

At this point, we can describe the Metropolis algorithm for a VMC calculation

Step 0 - Start from an arbitrary distribution of walkers on the coordinate x

Step 1 - Move the walkers according to $G(x_i \rightarrow y_{i+1})$, i.e. $y_{i+1} = x_i + \zeta$

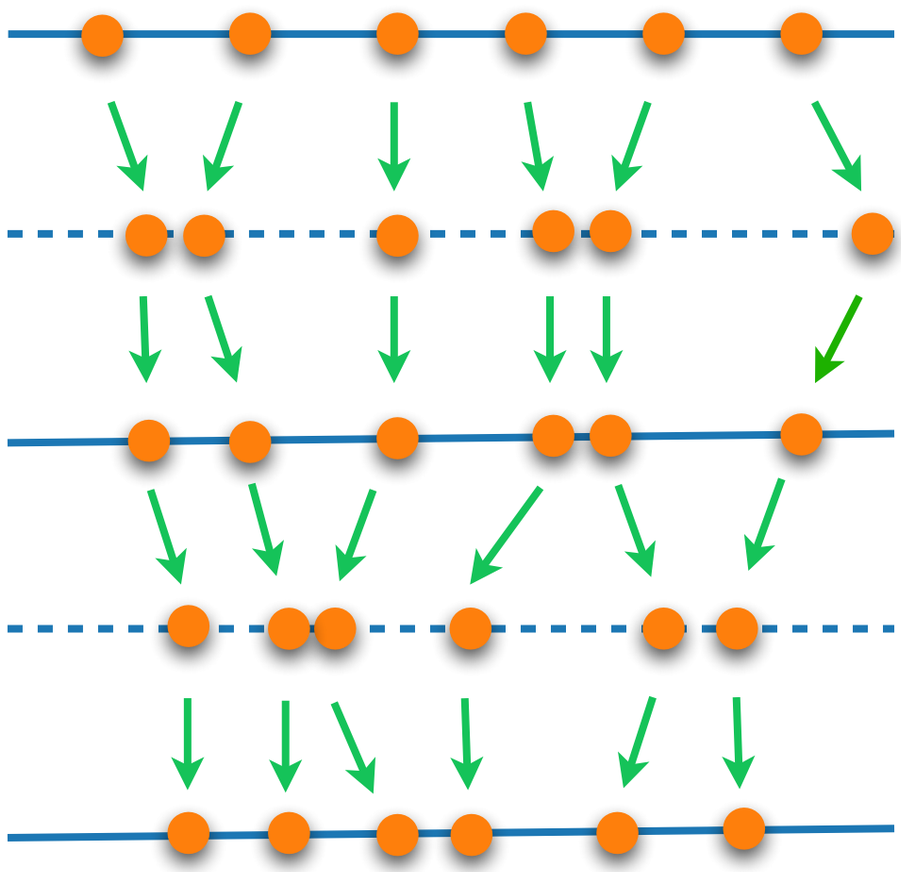
Step 2 - Compute the acceptance probability $A(x_i \rightarrow y_{i+1}) = \min\left(1, \frac{|\Psi_T(y_{i+1})|^2}{|\Psi_T(x_i)|^2}\right)$

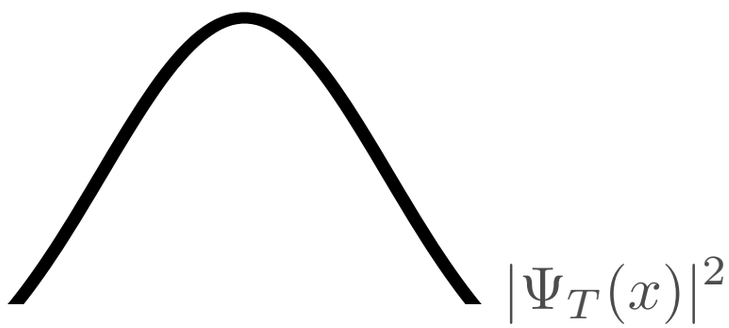
Step 3 - Accept or reject the proposed move

$$\frac{|\Psi_T(y_{i+1})|^2}{|\Psi_T(x_i)|^2} > \xi \quad \rightarrow \quad x_{i+1} = y_{i+1}$$

$$\frac{|\Psi_T(y_{i+1})|^2}{|\Psi_T(x_i)|^2} \leq \xi \quad \rightarrow \quad x_{i+1} = x_i$$

M(RT)^2 applied to VMC





- The walkers are sampled from an initial distribution
- Random Gaussian move
- Acceptance/rejection of the move
- Random Gaussian move
- Acceptance/rejection of the move
- Iterate until enough configurations are sampled

M(RT)^2 applied to VMC

A set of subsequent samplings will be correlated with each other and not correctly reflect the desired distribution

- To tame this problem, we can **pick walkers every "N_{void}" steps,** (where N_{void} is ideally larger than the correlation time of the chain)

- The correlation time can be reduced by increasing the size of the Gaussian step, but this will also increase the likelihood of rejecting the proposed move;

- A good choice for the size of the step is such that the acceptance probability is ~ 0.5

It typically takes ~1000 steps for the chain to equilibrate to the desired distribution. These samples must be disregarded (burn-in time).

Keeping these caveats into account, **we now know how to sample our walkers from**

$$P(R) = \frac{|\Psi_T(R)|^2}{\int dR |\Psi_T(R)|^2}$$

M(RT)^2 applied to VMC

Let us recall the trial energy definition:

$$E_T = \frac{\langle \Psi_T | H | \Psi_T \rangle}{\langle \Psi_T | \Psi_T \rangle} = \frac{\int dR \langle \Psi_T | R \rangle \langle R | H | \Psi_T \rangle}{\int dR \langle \Psi_T | R \rangle \langle R | \Psi_T \rangle} = \frac{\int dR |\Psi_T(R)|^2 E_L(R)}{\int dR |\Psi_T(R)|^2}$$

Using the central limit theorem, we can estimate the energy expectation value as

$$\langle E_T \rangle = \frac{1}{N_s} \sum_{R_n} E_L(R_n) \quad \longleftrightarrow \quad E_L(R) \equiv \frac{H\Psi_T(R)}{\Psi_T(R)}$$

Where the walkers are sampled from

$$P(R) = \frac{|\Psi_T(R)|^2}{\int dR |\Psi_T(R)|^2}$$

The integration error can also be readily estimated by

$$\sigma_{E_T} = \sqrt{\frac{\langle E_T^2 \rangle - \langle E_T \rangle^2}{N_s - 1}} \quad \longleftrightarrow \quad \langle E_T^2 \rangle = \frac{1}{N_s} \sum_{R_n} E_L^2(R_n)$$

Energy minimization

The trial wave function depends on a set of variational parameters

$$|\Psi_T\rangle \longrightarrow |\Psi_T(\mathbf{p})\rangle$$

It is convenient to introduce the state corresponding to the derivative of the wave function with respect to the i-th variational parameter

$$O^i|\Psi_T(\mathbf{p})\rangle \equiv \frac{\partial}{\partial p_i}|\Psi_T(\mathbf{p})\rangle$$

Assuming a real trial wave function, the energy derivative reads

$$\begin{aligned}\frac{\partial E_T}{\partial p_i} &= 2\left[\frac{\langle\Psi_T|HO^i|\Psi_T\rangle}{\langle\Psi_T|\Psi_T\rangle} - \frac{\langle\Psi_T|H|\Psi_T\rangle}{\langle\Psi_T|\Psi_T\rangle}\frac{\langle\Psi_T|O^i|\Psi_T\rangle}{\langle\Psi_T|\Psi_T\rangle}\right] \\ &= \langle E_T O^i\rangle - \langle E_T\rangle\langle O^i\rangle\end{aligned}$$

The variational parameters can be updated in the spirit of the **stochastic gradient descent** as

$$p_i^{n+1} = p_i^n - \tau\frac{\partial E_T}{\partial p_i} \longrightarrow \tau \simeq 0.001$$

The quantum Harmonic Oscillator

Let us consider the prototypal problem of a collection of A independent (decoupled) quantum Harmonic oscillators, in N dimensions

$$H = -\frac{1}{2}\sum_{i=1}^{A}\nabla_i^2 + \sum_{i=1}^{A}\frac{\mathbf{r}_i^2}{2}$$

We assume a trial wave function of the form

$$\Psi(R) = \exp\left(-\alpha\sum_{i=1}^{A}\mathbf{r}_i^2\right)$$

So that the exact ground-state wave function is recovered for $\alpha = 1/2$

$$\Psi_0(R) = \exp\left(-\frac{1}{2}\sum_{i=1}^{A}\mathbf{r}_i^2\right) \quad \longleftrightarrow \quad E_0 = A \times N \times \frac{1}{2}$$

The quantum Harmonic Oscillator

The local energy is the sum of the kinetic and potential contributions

$$E_L(R) = \frac{\langle R|H|\Psi_T\rangle}{\langle R|\Psi_T\rangle} = \frac{\langle R|T|\Psi_T\rangle}{\langle R|\Psi_T\rangle} + \frac{\langle R|V|\Psi_T\rangle}{\langle R|\Psi_T\rangle}$$

The kinetic energy involves the second derivative of the trial wave function

$$T_L(R) = -\frac{1}{2}\sum_{i=1}^{A}\frac{\nabla_i^2\Psi_T(R)}{\Psi_T(R)} = -\frac{1}{2}\sum_{i=1}^{A}\left(-2\alpha N + 4\alpha^2\mathbf{r}_i^2\right) = \sum_{i=1}^{A}\left(\alpha N - 2\alpha^2\mathbf{r}_i^2\right)$$

The potential energy is more immediate to evaluate

$$V_L(R) = \frac{1}{2}\sum_{i=1}^{A}\mathbf{r}_i^2$$

Question: What happens for $\alpha = 1/2$?

$$E_L(R) = \frac{1}{2} \times A \times N$$