Introduction to quantum Monte Carlo methods

Claudia Filippi and Ravindra Shinde
MESA+ Institute for Nanotechnology, Universiteit Twente, The Netherlands

Anthony Scemama and Vijay Gopal Chilkuri
Laboratoire de Chimie et Physique Quantique, CNRS, Toulouse, France

Luchon School Tutorials in Theoretical Chemistry, Jan 25-Feb 8, 2021

Targeting Real Chemical Accuracy at the Exascale project has received funding from the European Union
Horizoon 2020 research and innovation programme under Grant Agreement **No. 952165.**

$\boxed{\text{Monte Carlo methods}}$

Approaches which make repeated use of random numbers:

- to simulate truly stochastic events
- to solve $\boxed{\text{deterministic problems}}$ using probabilities

Very important class of methods in statistical mechanics

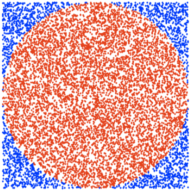
→ Sampling Boltzmann distribution

Computation of averages (integrals in many dimensions)

For quantum mechanical simulations → $\boxed{\text{Quantum Monte Carlo}}$

A simple example of a Monte Carlo simulation

Basic idea of Monte Carlo through the “dartboard method”



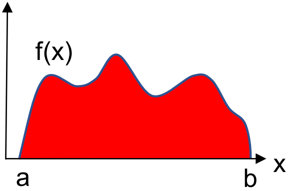
→ Throw darts, compute $\mathrm{A}_{\mathrm{cicle}}$, compute π

Throw darts which land randomly within the square

$$\frac{\#\text{ hits inside circle}}{\#\text{ hits inside the square}} \underset{\substack{\uparrow \\ \text{many, many hits}}}{=} \frac{A_{\mathrm{circle}}}{A_{\mathrm{square}}} = \frac{\pi}{4}$$

Monte Carlo integration

We want to compute the integral of f(x) in the interval $[a, b]$



$$
\begin{aligned}
I &= \int_a^b f(x)\,\mathrm{d}x = (b-a)\int_a^b f(x)\frac{1}{b-a}\,\mathrm{d}x \\
&= (b-a)\langle f\rangle_{[a,b]}
\end{aligned}
$$

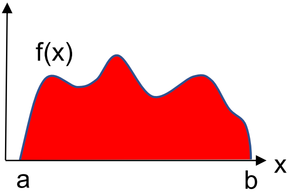
where $\langle f\rangle_{[a,b]}$ is the average of the function in the range $[a, b]$

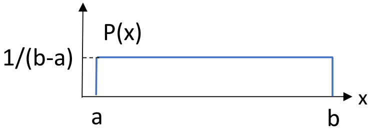
Monte Carlo integration

(2)

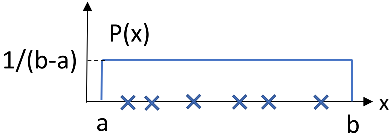
$$\langle f\rangle_{[a,b]} = \int_a^b f(x)\frac{1}{b-a}\,\mathrm{d}x$$

$$= \int_a^b f(x)P(x)\,\mathrm{d}x$$





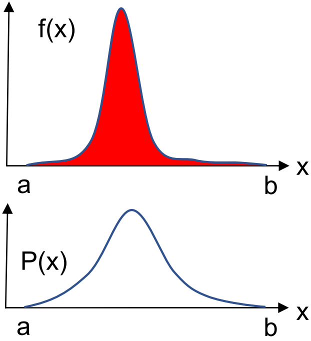
Draw M random numbers distributed unformely in $[a,b]$



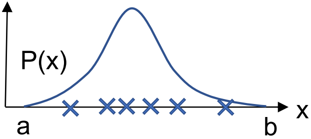
$$\longrightarrow \quad \boxed{\langle f\rangle_{[a,b]} \approx \frac{1}{M}\sum_{i=1}^{M} f(x_i)}$$

A less uniform function

$$
\begin{aligned}
I &= \int_a^b f(x)\,\mathrm{d}x \\
&= \int_a^b \frac{f(x)}{P(x)} P(x)\,\mathrm{d}x
\end{aligned}
$$



Draw M random numbers distributed as $P(x)$



$$
\longrightarrow \quad \boxed{\langle f\rangle_{[a,b]} \approx \frac{1}{M}\sum_{i=1}^{M} f(x_i)}
$$

Monte Carlo integration in a nutshell

We want to compute $\boxed{\langle A \rangle = \int_a^b A(x)P(x)}$

with $\boxed{P(x) \geq 0 \;\text{ and }\; \int_a^b P(x) = 1}$ ← a probability density!

Monte Carlo → Sample $\{x_1, \ldots, x_M\}$ from $P(x)$

$$\text{Estimate} \quad \langle A \rangle \approx \frac{1}{M} \sum_{i=1}^{M} A(x_i)$$

Statistical physics: $P(x) = \dfrac{e^{-\beta E(x)}}{Z}$, the Boltzman distribution

Quantum chemical simulations

- Density functional theory methods

  Large systems but approximate exchange/correlation

- Quantum chemistry post-Hartree-Fock methods

  Accurate on small-medium systems

  → Jungle of approaches: CI, MCSCF, CC, CASPT2 . . .

- Quantum Monte Carlo techniques

  Stochastic solution of the Schrödinger equation

  Accurate correlated calculations for medium-large systems

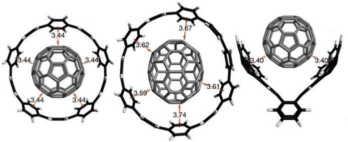
Some general words about quantum Monte Carlo methods

Stochastically solve interacting Schrödinger equation

Why (real-space) quantum Monte Carlo?

- Favorable scaling → Energy is $\boxed{O(N^4)}$
- Flexibility in choice of functional form of wave function
- Easy parallelization
- Among most accurate calculations for medium-large systems

  Routinely, molecules of up to 100 (mainly 1st/2nd-row) atoms



upto $C_{136}H_{44}$ (Alfé 2017)

A different way of writing the expectation values

Consider the expectation value of the Hamiltonian on Ψ

$$\begin{aligned} E_V &= \frac{\langle\Psi|\mathcal{H}|\Psi\rangle}{\langle\Psi|\Psi\rangle} = \frac{\int \mathrm{d}\mathbf{R}\,\Psi^*(\mathbf{R})\mathcal{H}\Psi(\mathbf{R})}{\int \mathrm{d}\mathbf{R}\,\Psi^*(\mathbf{R})\Psi(\mathbf{R})} \geq E_0 \\ &= \int \mathrm{d}\mathbf{R}\,\frac{\mathcal{H}\Psi(\mathbf{R})}{\Psi(\mathbf{R})}\,\underset{\big\downarrow}{\boxed{\frac{|\Psi(\mathbf{R})|^2}{\int \mathrm{d}\mathbf{R}|\Psi(\mathbf{R})|^2}}} \\ &= \int \mathrm{d}\mathbf{R}\,E_{\mathrm{L}}(\mathbf{R})\,P(\mathbf{R}) = \langle E_{\mathrm{L}}(\mathbf{R})\rangle_P \end{aligned}$$

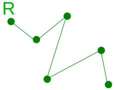
$P(\mathbf{R})$ is a probability density and $E_{\mathrm{L}}(\mathbf{R}) = \dfrac{\mathcal{H}\Psi(\mathbf{R})}{\Psi(\mathbf{R})}$ the local energy

Variational Monte Carlo: a random walk of the electrons

Use Monte Carlo integration to compute expectation values

- ▷ Sample $\mathbf{R}$ from $P(\mathbf{R})$ using Metropolis algorithm
- ▷ Average local energy $E_{\mathrm{L}}(\mathbf{R}) = \frac{\mathcal{H}\Psi(\mathbf{R})}{\Psi(\mathbf{R})}$ to obtain E_V as

$$\boxed{E_V = \langle E_{\mathrm{L}}(\mathbf{R})\rangle_P \approx \frac{1}{M}\sum_{i=1}^{M} E_{\mathrm{L}}(\mathbf{R}_i)}$$

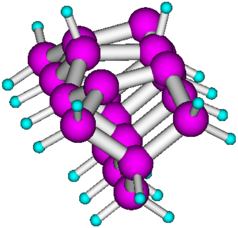


Random walk in 3N dimensions, $\mathbf{R} = (\mathbf{r}_1, \ldots, \mathbf{r}_N)$

Just a $\boxed{\text{trick}}$ to evaluate integrals in many dimensions

Is it really "just" a trick?

$Si_{21}H_{22}$



Number of electrons $4 \times 21 + 22 = 106$

Number of dimensions $3 \times 106 = \boxed{318}$

Integral on a grid with 10 points/dimension → 10^{318} points!

MC is a powerful trick ⇒ Freedom in form of the wave function Ψ

Monte Carlo integration

We want to compute an integral

$$E_V = \int \mathrm{d}\mathbf{R} E_{\mathrm{L}}(\mathbf{R}) P(\mathbf{R})$$

We sample $P(\mathbf{R}) \rightarrow$ $\boxed{E_V = \langle E_{\mathrm{L}}(\mathbf{R}) \rangle_P \approx \frac{1}{M} \sum_{i=1}^{M} E_{\mathrm{L}}(\mathbf{R}_i)}$

- Does the trick always work?
- How efficient is it?

The Central Limit Theorem

Probability density P and function f with finite mean and variance

$$\boxed{\mu} = \int \mathrm{d}x\, f(x)P(x) \qquad \boxed{\sigma^2} = \int \mathrm{d}x\, (f(x)^2 - \mu)P(x)$$

Sample M independent random variables $x_1, \ldots, x_M$ from $P(x)$

Define $$\boxed{F_M = \frac{1}{M}\sum_{i=1}^{M} f(x_i)}$$

As M increases, F_M is normally distributed as $\frac{1}{\sqrt{2\pi\sigma}} e^{-(x-\mu)^2/2\sigma_M^2}$

with a mean $\boxed{\mu}$ and variance $\boxed{\sigma_M^2 = \sigma^2/\text{M}}$

→ **Irrespective** of the original probability density function

Monte Carlo versus deterministic integration

Integration error ϵ using M integration/Monte Carlo points

- Monte Carlo methods

  $\epsilon \propto \dfrac{1}{\sqrt{M}}$ independent on dimension !

  It follows from Central Limit Theorem

  $\rightarrow$ width of Gaussian decreases as $\dfrac{\sigma}{\sqrt{M}}$ for finite variance

- Deterministic integration methods

  1-dim Simpson rule: $\epsilon \propto \dfrac{1}{M^4}$

  d-dim Simpson rule: $\epsilon \propto \dfrac{1}{M^{4/d}}$

Scaling with number of electrons

Roughly, Monte Carlo integration advantageous if $d > 8$

… for many-body wave functions $d = 3N_{\mathrm{elec}}$!

– Simpson rule integration ($\mathrm{M}_{\mathrm{int}}$ integration points)

$$\epsilon = \frac{c}{M_{\mathrm{int}}^{4/d}} = \frac{c}{M_{\mathrm{int}}^{4/3N_{\mathrm{elec}}}} \quad \Rightarrow \quad M_{\mathrm{int}} = \left(\frac{c}{\epsilon}\right)^{3N_{\mathrm{elec}}/4} \quad \boxed{\text{Exponential}}$$

– Monte Carlo integration (M_{MC} Monte Carlo samples)

$$\epsilon = \frac{\sigma}{\sqrt{M_{\mathrm{MC}}}} = c\sqrt{\frac{N_{\mathrm{elec}}}{M_{\mathrm{MC}}}} \quad \Rightarrow \quad M_{\mathrm{MC}} = \left(\frac{c}{\epsilon}\right)^2 N_{\mathrm{elec}} \quad \boxed{\text{Linear}}$$

Summary of variational Monte Carlo

Expectation value of the Hamiltonian on Ψ

$$E_V = \frac{\langle\Psi|\mathcal{H}|\Psi\rangle}{\langle\Psi|\Psi\rangle} = \int \mathrm{d}\mathbf{R}\, \frac{\mathcal{H}\Psi(\mathbf{R})}{\Psi(\mathbf{R})} \frac{|\Psi(\mathbf{R})|^2}{\int \mathrm{d}\mathbf{R}|\Psi(\mathbf{R})|^2} = \int \mathrm{d}\mathbf{R}\, E_\mathrm{L}(\mathbf{R})\, P(\mathbf{R})$$

$$\begin{aligned} E_V &= \int \mathrm{d}\mathbf{R}\, E_\mathrm{L}(\mathbf{R})\, P(\mathbf{R}) \\ \sigma^2 &= \int \mathrm{d}\mathbf{R} (E_\mathrm{L}(\mathbf{R}) - E_V)^2 P(\mathbf{R}) \end{aligned}$$

Estimate E_V and σ from M independent samples as

$$\begin{aligned} \bar{E}_V &= \frac{1}{M} \sum_{i=1}^{M} E_\mathrm{L}(\mathbf{R}_i) \\ \bar{\sigma}^2 &= \frac{1}{M-1} \sum_{i=1}^{M} (E_\mathrm{L}(\mathbf{R}_i) - \bar{E}_V)^2 \end{aligned}$$

Are there any conditions on many-body Ψ to be used in VMC?

Within VMC, we can use any "computable" wave function if

▷ Continuous, normalizable, proper symmetry

▷ Finite variance

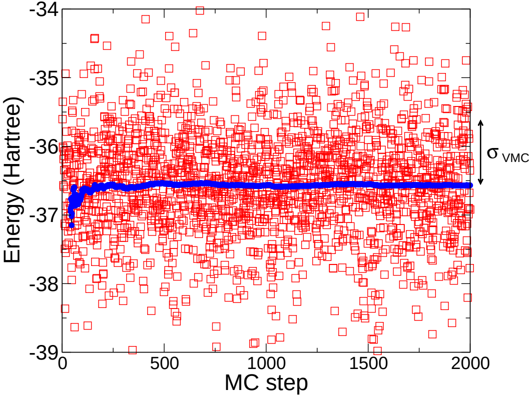
$$\sigma^2 = \frac{\langle \Psi | (\mathcal{H} - E_V)^2 | \Psi \rangle}{\langle \Psi | \Psi \rangle} = \langle (E_{\mathrm{L}}(\mathbf{R}) - E_V)^2 \rangle_P$$

since the Monte Carlo error goes as $\boxed{\mathrm{err}(E_V) \sim \frac{\sigma}{\sqrt{M}}}$

Zero variance principle: if $\Psi \to \Psi_0$, $E_{\mathrm{L}}(\mathbf{R})$ does not fluctuate

Typical VMC run

Example: Local energy and average energy of acetone (C_3H_6O)



$$E_{\mathrm{VMC}} = \langle E_{\mathrm{L}}(\mathbf{R})\rangle_P = -36.542 \pm 0.001 \text{ Hartree } (40\times 20000 \text{ steps})$$

$$\sigma_{\mathrm{VMC}} = \langle (E_{\mathrm{L}}(\mathbf{R}) - E_{\mathrm{VMC}})^2\rangle_P = 0.90 \text{ Hartree}$$

Variational Monte Carlo: To do list

– Method to $\boxed{\text{sample}}$ distribution function $\rho(\mathbf{R}) = \frac{|\Psi(\mathbf{R})|^2}{\int \mathrm{d}\mathbf{R}|\Psi(\mathbf{R})|^2}$

→ Obtain a set of $\{\mathbf{R}_1, \mathbf{R}_2, \ldots, \mathbf{R}_M\}$ distributed as $\rho(\mathbf{R})$

How? As in classical Monte Carlo with Metropolis algorithm!

– Build the wave function $\Psi(\mathbf{R})$. Which $\boxed{\text{functional form}}$?

Here, we spend most of our time, open topic of research

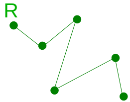
– Compute expectation values $\boxed{\frac{\langle\Psi|\mathcal{O}|\Psi\rangle}{\langle\Psi|\Psi\rangle}}$

Reformulate them to reduce fluctuations, open topic of research

How do we sample $\rho(\mathbf{R})$?

Generate a Markov chain

$$\dots \xrightarrow{M} \mathbf{R} \xrightarrow{M} \mathbf{R}' \xrightarrow{M} \mathbf{R}'' \xrightarrow{M} \dots$$



Construct $M(\mathbf{R}_{\mathrm{f}}|\mathbf{R}_{\mathrm{i}})$ as probability for transition $\mathbf{R}_{\mathrm{i}} \to \mathbf{R}_{\mathrm{f}}$ so that

- $M(\mathbf{R}_{\mathrm{f}}|\mathbf{R}_{\mathrm{i}}) \geq 0$ and $\int d\mathbf{R}_{\mathrm{f}} M(\mathbf{R}_{\mathrm{f}}|\mathbf{R}_{\mathrm{i}}) = 1$ (stochastic)
- If we start from an arbitrary distribution P_{init}, we evolve to P

→ Impose stationarity condition

Constructing M

To sample P, use M which satisfies stationarity condition:

$$\int \mathrm{d}\mathbf{R}_{\mathrm{i}}\, M(\mathbf{R}_{\mathrm{f}}|\mathbf{R}_{\mathrm{i}})\; P(\mathbf{R}_{\mathrm{i}}) = P(\mathbf{R}_{\mathrm{f}}) \quad \forall\, \mathbf{R}_{\mathrm{f}}$$

▷ Stationarity condition

$\Rightarrow$ $\boxed{\text{If we start with } P\text{, we continue to sample } P}$

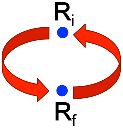
▷ Stationarity condition + stochastic property of P + ergodicity

$\Rightarrow$ $\boxed{\text{Any initial distribution will evolve to } P}$

More stringent condition

In practice, we impose detailed balance condition

$$\boxed{M(\mathbf{R}_\mathrm{f}|\mathbf{R}_\mathrm{i})\ P(\mathbf{R}_\mathrm{i}) = M(\mathbf{R}_\mathrm{i}|\mathbf{R}_\mathrm{f})\ P(\mathbf{R}_\mathrm{f})}$$



Stationarity condition can be obtained by summing over $\mathbf{R}_\mathrm{i}$

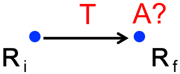
$$\int \mathrm{d}\mathbf{R}_\mathrm{i} M(\mathbf{R}_\mathrm{f}|\mathbf{R}_\mathrm{i})\ P(\mathbf{R}_\mathrm{i}) = \underbrace{\int \mathrm{d}\mathbf{R}_\mathrm{i} M(\mathbf{R}_\mathrm{i}|\mathbf{R}_\mathrm{f})}_{1}\ P(\mathbf{R}_\mathrm{f}) = P(\mathbf{R}_\mathrm{f})$$

Detailed balance is a sufficient but not necessary condition

How do we construct the transition matrix P in practice?

Metropolis method → Write P as proposal T × acceptance A

$$\boxed{M(\mathbf{R}_\mathrm{f}|\mathbf{R}_\mathrm{i}) = A(\mathbf{R}_\mathrm{f}|\mathbf{R}_\mathrm{i})\ T(\mathbf{R}_\mathrm{f}|\mathbf{R}_\mathrm{i})}$$



Let us rewrite the detailed balance condition

$$M(\mathbf{R}_\mathrm{f}|\mathbf{R}_\mathrm{i})\ P(\mathbf{R}_\mathrm{i}) = M(\mathbf{R}_\mathrm{i}|\mathbf{R}_\mathrm{f})\ P(\mathbf{R}_\mathrm{f})$$

$$A(\mathbf{R}_\mathrm{f}|\mathbf{R}_\mathrm{i})\ T(\mathbf{R}_\mathrm{f}|\mathbf{R}_\mathrm{i})\ P(\mathbf{R}_\mathrm{i}) = A(\mathbf{R}_\mathrm{i}|\mathbf{R}_\mathrm{f})\ T(\mathbf{R}_\mathrm{i}|\mathbf{R}_\mathrm{f})\ P(\mathbf{R}_\mathrm{f})$$

$$\Rightarrow \quad \frac{A(\mathbf{R}_\mathrm{f}|\mathbf{R}_\mathrm{i})}{A(\mathbf{R}_\mathrm{i}|\mathbf{R}_\mathrm{f})} = \frac{T(\mathbf{R}_\mathrm{i}|\mathbf{R}_\mathrm{f})\ P(\mathbf{R}_\mathrm{f})}{T(\mathbf{R}_\mathrm{f}|\mathbf{R}_\mathrm{i})\ P(\mathbf{R}_\mathrm{i})}$$

Choice of acceptance matrix A

Original choice by Metropolis *et al.* maximizes the acceptance

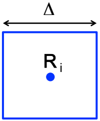
$$\boxed{A(\mathbf{R}_\mathrm{f}|\mathbf{R}_\mathrm{i}) = \min\left\{1, \frac{T(\mathbf{R}_\mathrm{i}|\mathbf{R}_\mathrm{f})\ P(\mathbf{R}_\mathrm{f})}{T(\mathbf{R}_\mathrm{f}|\mathbf{R}_\mathrm{i})\ P(\mathbf{R}_\mathrm{i})}\right\}}$$

Note: $P(\mathbf{R})$ does not have to be normalized

→ For complicated Ψ we do not know the normalization!

→ $P(\mathbf{R}) = |\Psi(\mathbf{R})|^2$

Original Metropolis method



Symmetric $T(\mathbf{R}_\mathrm{f}|\mathbf{R}_\mathrm{i}) = 1/\Delta^{3N} \ \Rightarrow\ A(\mathbf{R}_\mathrm{f}|\mathbf{R}_\mathrm{i}) = \min\left\{1, \frac{P(\mathbf{R}_\mathrm{f})}{P(\mathbf{R}_\mathrm{i})}\right\}$

Better choices of proposal matrix T

Sequential correlation $\Rightarrow$ $M_{\mathrm{eff}} < M$ independent observations

$\boxed{M_{\mathrm{eff}} = \dfrac{M}{T_{\mathrm{corr}}}}$ with T_{corr} autocorrelation time of desired observable

Aim is to achieve fast evolution and reduce correlation times

Use **freedom** in choice of T: For example, use available trial Ψ

$$T(\mathbf{R}_{\mathrm{f}}|\mathbf{R}_{\mathrm{i}}) = \mathcal{N} \exp\left[-\frac{(\mathbf{R}_{\mathrm{f}} - \mathbf{R}_{\mathrm{i}} - \mathbf{V}(\mathbf{R}_{\mathrm{i}})\tau)^2}{2\tau}\right] \quad \text{with} \quad \mathbf{V}(\mathbf{R}_{\mathrm{i}}) = \frac{\nabla\Psi(\mathbf{R}_{\mathrm{i}})}{\Psi(\mathbf{R}_{\mathrm{i}})}$$

Acceptance and T_{corr} for the total energy E_V

Example: All-electron Be atom with simple wave function

Simple Metropolis

Δ	T_{corr}	$\bar{A}$
1.00	41	0.17
0.75	21	0.28
0.50	17	0.46
0.20	45	0.75

Drift-diffusion transition

τ	T_{corr}	$\bar{A}$
0.100	13	0.42
0.050	7	0.66
0.020	8	0.87
0.010	14	0.94

Generalized Metropolis algorithm

1. Choose distribution $P(\mathbf{R})$ and proposal matrix $T(\mathbf{R}_\mathrm{f}|\mathbf{R}_\mathrm{i})$
2. Initialize the configuration $\mathbf{R}_i$
3. Advance the configuration from $\mathbf{R}_i$ to $\mathbf{R}'$
    - a) Sample $\mathbf{R}'$ from $T(\mathbf{R}'|\mathbf{R}_\mathrm{i})$.
    - b) Calculate the ratio $p = \dfrac{T(\mathbf{R}_\mathrm{i}|\mathbf{R}')}{T(\mathbf{R}'|\mathbf{R}_\mathrm{i})} \dfrac{P(\mathbf{R}')}{P(\mathbf{R}_\mathrm{i})}$
    - c) Accept or reject with probability p

      Pick a uniformly distributed random number $\chi \in [0, 1]$

      if $\chi < p$, move accepted → set $\mathbf{R}_\mathrm{f} = \mathbf{R}'$

      if $\chi > p$, move rejected → set $\mathbf{R}_\mathrm{f} = \mathbf{R}$
4. Throw away first κ configurations of equilibration time
5. Collect the averages

Variational Monte Carlo → Freedom in choice of Ψ

Monte Carlo integration allows the use of complex and accurate Ψ

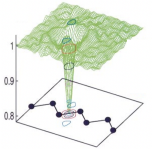
$\Rightarrow$ More compact representation of Ψ than in quantum chemistry

$\Rightarrow$ Beyond $c_0 D_{\mathrm{HF}} + c_1 D_1 + c_2 D_2 + \ldots$ millions of determinants

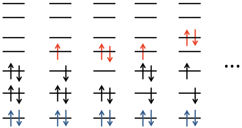
Jastrow-Slater wave function

Commonly employed compact Jastrow-Slater wave functions

$$\Psi(\mathbf{r}_1,\ldots,\mathbf{r}_N) = \mathcal{J}(\mathbf{r}_1,\ldots,\mathbf{r}_N) \times \sum_i c_i\, D_i(\mathbf{r}_1,\ldots,\mathbf{r}_N)$$



×



$\boxed{\mathcal{J}} \longrightarrow$ Jastrow correlation factor

- Explicit dependence on electron-electron distances $\boxed{r_{ij}}$

$\boxed{\sum c_i\, D_i} \longrightarrow$ Determinants of single-particle orbitals

- $\boxed{\text{Few}}$ and not millions of determinants

Jastrow factor and divergences in the potential

At interparticle coalescence points, the potential diverges as

$-\dfrac{Z}{r_{i\alpha}}$ for the electron-nucleus potential

$\dfrac{1}{r_{ij}}$ for the electron-electron potential

Local energy $\boxed{\dfrac{\mathcal{H}\Psi}{\Psi} = -\dfrac{1}{2}\sum_i \dfrac{\nabla_i^2\Psi}{\Psi} + \mathcal{V}}$ must be $\boxed{\text{finite}}$

$\Rightarrow$ Kinetic energy must have opposite divergence to the potential $\mathcal{V}$

Divergence in potential and Kato's cusp conditions

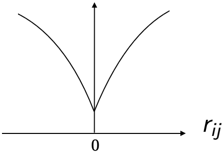
Finite local energy as $r_{ij} \to 0 \Rightarrow \Psi$ must satisfy:

$$\boxed{\left.\frac{\partial\Psi}{\partial r_{ij}}\right|_{r_{ij}=0} = \mu_{ij} q_i\, q_j \Psi(r_{ij}=0)}$$

Electron-electron cusps imposed through the Jastrow factor

Example: Simple Jastrow factor → $\mathcal{J}(r_{ij}) = \prod_{i<j} \exp\left\{ b_0 \frac{r_{ij}}{1+b\,r_{ij}} \right\}$

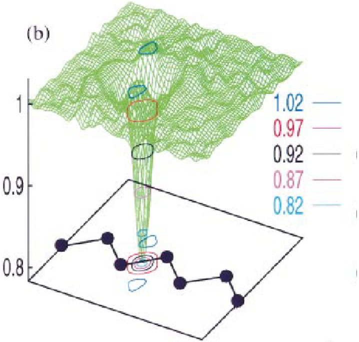
Imposes cusp conditions
+
keeps electrons apart



The effect of the Jastrow factor

Pair correlation function for ↑↓ electrons in the (110) plane of Si

$g_{\uparrow\downarrow}(\mathbf{r}, \mathbf{r}')$ with one electron is at the bond center



Hood *et al.* Phys. Rev. Lett. **78**, 3350 (1997)

Why should $\Psi_{\mathrm{QMC}} = \mathcal{J} D$ work?

$$\begin{tikzcd}
\text{Full wave-function} \arrow[r] & \text{Factorized wave-function} \\
\Psi \arrow[d] & \mathcal{J}\Phi \arrow[d] \\
\text{Full Hamiltonian} \arrow[r] & \text{Effective Hamiltonian} \\
\mathcal{H} & \mathcal{H}_{\mathrm{eff}} \\
\boxed{\mathcal{H}\Psi = \mathrm{E}\Psi} \arrow[r] & \mathcal{H}\mathcal{J}\Phi = \mathrm{E}\mathcal{J}\Phi \rightarrow \dfrac{\mathcal{H}\mathcal{J}}{\mathcal{J}}\Phi = \mathrm{E}\Phi \\
 & \boxed{\mathcal{H}_{\mathrm{eff}}\Phi = \mathrm{E}\Phi}
\end{tikzcd}$$

$\mathcal{H}_{\mathrm{eff}}$ weaker Hamiltonian than $\mathcal{H}$

$\Rightarrow$ $\Phi \approx$ non-interacting wave function D

$\Rightarrow$ Quantum Monte Carlo wave function $\Psi = \mathcal{J} D$

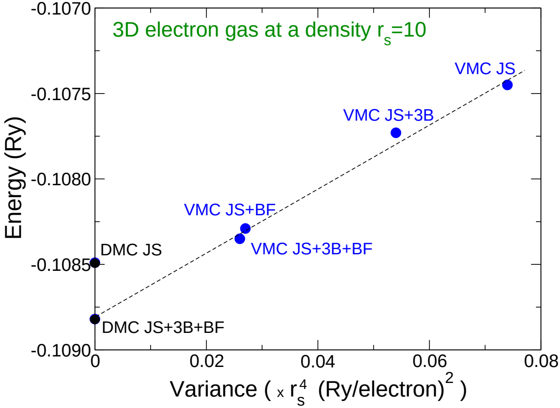
$\boxed{\text{Beyond VMC?}}$

Removing or reducing wave function bias?

⇒ Projection Monte Carlo methods

Why going beyond VMC?

Dependence of VMC from wave function Ψ



Kwon, Ceperley, Martin, Phys. Rev. B **58**, 6800 (1998)

Why going beyond VMC?

What goes in, comes out! Can we remove wave function bias?

Projector (diffusion) Monte Carlo method

- ▷ Construct an operator which inverts spectrum of $\mathcal{H}$

  Diffusion Monte Carlo $\rightarrow$ $e^{-\tau(\mathcal{H}-E_{\mathrm{ref}})}$

- ▷ Use it to stochastically project the ground state of $\mathcal{H}$

Diffusion Monte Carlo

Consider initial guess $\Psi^{(0)}$ and repeatedly apply projection operator

$$\boxed{\Psi^{(n)} = e^{-\tau(\mathcal{H}-E_{\mathrm{ref}})}\Psi^{(n-1)}}$$

Expand $\Psi^{(0)}$ on the eigenstates Ψ_i with energies E_i of $\mathcal{H}$

$$\Psi^{(n)} = e^{-n\tau(\mathcal{H}-E_{\mathrm{ref}})}\Psi^{(0)} = \sum_i \Psi_i \left\langle \Psi_i \middle| \Psi^{(0)} \right\rangle e^{-n\tau(E_i-E_{\mathrm{ref}})}$$

and obtain in the limit of $n \to \infty$

$$\lim_{n\to\infty} \Psi^{(n)} = \Psi_0 \left\langle \Psi_0 \middle| \Psi^{(0)} \right\rangle e^{-n\tau(E_0-E_{\mathrm{ref}})}$$

If we choose $E_{\mathrm{ref}} \approx E_0$, we obtain $\boxed{\lim_{n\to\infty} \Psi^{(n)} = \Psi_0}$

How do we perform the projection?

Rewrite projection equation in integral form

$$\boxed{\Psi(\mathbf{R}', t+\tau) = \int \mathrm{d}\mathbf{R}\, G(\mathbf{R}', \mathbf{R}, \tau)\Psi(\mathbf{R}, t)}$$

where $G(\mathbf{R}', \mathbf{R}, \tau) = \langle \mathbf{R}' | e^{-\tau(\mathcal{H}-E_{\mathrm{ref}})} | \mathbf{R} \rangle$

- ▷ Can we sample the wave function?

  For the moment, assume we are dealing with $\boxed{\text{bosons}}$, so $\Psi > 0$

- ▷ Can we interpret $G(\mathbf{R}', \mathbf{R}, \tau)$ as a transition probability?

  If yes, we can perform this integral by Monte Carlo integration

VMC and DMC as power methods

VMC Distribution function is given $\boxed{P(\mathbf{R}) = \frac{|\Psi(\mathbf{R})|^2}{\int \mathrm{d}\mathbf{R}|\Psi(\mathbf{R})|^2}}$

Construct M which satisfies stationarity condition so that

$$\lim_{n\to\infty} \int \mathrm{d}\mathbf{R}_n \cdots \mathrm{d}\mathbf{R}_1 M(\mathbf{R},\mathbf{R}_n) \cdots M(\mathbf{R}_3,\mathbf{R}_2) M(\mathbf{R}_2,\mathbf{R}_1) P_{\mathrm{init}}(\mathbf{R}_1) = P(\mathbf{R})$$

DMC Opposite procedure!

The matrix M is given → $\boxed{M \equiv G = \langle \mathbf{R}' | e^{-\tau(\mathcal{H}-E_{\mathrm{ref}})} | \mathbf{R} \rangle}$

We do not know $\boxed{P}$!

$$\lim_{n\to\infty} \int \mathrm{d}\mathbf{R}_n \cdots \mathrm{d}\mathbf{R}_1 G(\mathbf{R},\mathbf{R}_n) \cdots G(\mathbf{R}_3,\mathbf{R}_2) G(\mathbf{R}_2,\mathbf{R}_1) P_{\mathrm{init}}(\mathbf{R}_1) = \Psi_0(\mathbf{R})$$

In either case, we want to find the dominant eigenvector of M

What can we say about the Green's function?

$$G(\mathbf{R}', \mathbf{R}, \tau) = \langle \mathbf{R}' | e^{-\tau(\mathcal{H}-E_{\mathrm{ref}})} | \mathbf{R} \rangle$$

$G(\mathbf{R}', \mathbf{R}, \tau)$ satisfies the imaginary-time Schrödinger equation

$$\boxed{(\mathcal{H} - E_{\mathrm{ref}}) G(\mathbf{R}, \mathbf{R}_0, t) = -\frac{\partial G(\mathbf{R}, \mathbf{R}_0, t)}{\partial t}}$$

with $G(\mathbf{R}', \mathbf{R}, 0) = \delta(\mathbf{R}' - \mathbf{R})$

Evolution equation of the probability distribution

We can understand the behavior of G which satisfies

$$\boxed{(\mathcal{H} - E_{\mathrm{ref}})G(\mathbf{R}, \mathbf{R}_0, t) = -\frac{\partial G(\mathbf{R}, \mathbf{R}_0, t)}{\partial t}}$$

to understand evolution of the distribution Ψ

$$\Psi(\mathbf{R}, t) = \int \mathrm{d}\mathbf{R}_0 \, G(\mathbf{R}, \mathbf{R}_0, t)\Psi^{(0)}(\mathbf{R}_0)$$

which satisfies the imaginary-time Schrödinger equation

$$\boxed{(\mathcal{H} - E_{\mathrm{ref}})\Psi(\mathbf{R}, t) = -\frac{\partial \Psi(\mathbf{R}, t)}{\partial t}}$$

$$\boxed{\text{Can we interpret } G(\mathbf{R}', \mathbf{R}, \tau) \text{ as a transition probability?}} \tag{1}$$

$$\boxed{\mathcal{H} = \mathcal{T}}$$

Imaginary-time Schrödinger equation is a diffusion equation

$$-\frac{1}{2}\nabla^2 G(\mathbf{R}, \mathbf{R}_0, t) = -\frac{\partial G(\mathbf{R}, \mathbf{R}_0, t)}{\partial t}$$

The Green's function is given by a Gaussian

$$G(\mathbf{R}', \mathbf{R}, \tau) = (2\pi\tau)^{-3N/2} \exp\left[-\frac{(\mathbf{R}' - \mathbf{R})^2}{2\tau}\right]$$

$$\boxed{\text{Positive and can be sampled}}$$

$$\boxed{\text{Can we interpret } G(\mathbf{R}', \mathbf{R}, \tau) \text{ as a transition probability?}} \qquad (2)$$

$$\boxed{\mathcal{H} = \mathcal{V}}$$

$$(\mathcal{V}(\mathbf{R}) - E_{\mathrm{ref}})G(\mathbf{R}, \mathbf{R}_0, t) = -\frac{\partial G(\mathbf{R}, \mathbf{R}_0, t)}{\partial t} \,,$$

The Green's function is given by

$$G(\mathbf{R}', \mathbf{R}, \tau) = \exp\left[-\tau \left(\mathcal{V}(\mathbf{R}) - E_{\mathrm{ref}}\right)\right] \delta(\mathbf{R} - \mathbf{R}'),$$

$\boxed{\text{Positive}}$ but does not preserve the normalization

It is a factor by which we multiply the distribution $\Psi(\mathbf{R}, t)$

$\boxed{\mathcal{H} = \mathcal{T} + \mathcal{V} \text{ and a combination of diffusion and branching}}$

Let us combine previous results

$$\boxed{G(\mathbf{R}', \mathbf{R}, \tau) \approx (2\pi\tau)^{-3N/2} \exp\left[-\frac{(\mathbf{R}' - \mathbf{R})^2}{2\tau}\right] \exp\left[-\tau\left(\mathcal{V}(\mathbf{R}) - E_{\mathrm{T}}\right)\right]}$$

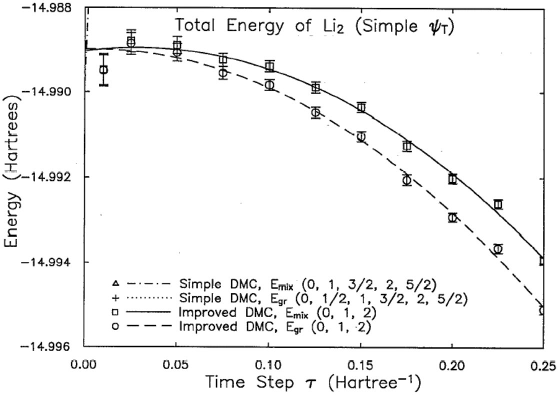
Diffusion + branching factor leading to survival/death/cloning

Why? Trotter's theorem → $\boxed{e^{(A+B)\tau} = e^{A\tau} e^{B\tau} + \mathcal{O}(\tau^2)}$

→ Green's function in the $\boxed{\text{short-time approximation}}$ to $\mathcal{O}(\tau^2)$

Time-step extrapolation

Example: Energy of Li_2 versus time-step τ



Umrigar, Nightingale, Runge, J. Chem. Phys. **94**, 2865 (1993)

Diffusion Monte Carlo as a branching random walk

The basic DMC algorithm is rather simple:

1. Sample $\Psi^{(0)}(\mathbf{R})$ with the Metropolis algorithm

   Generate M_0 walkers $\mathbf{R}_1, \ldots, \mathbf{R}_{M_0}$ (zeroth generation)

2. Diffuse each walker as $\boxed{\mathbf{R}' = \mathbf{R} + \xi}$

   where ξ is sampled from $g(\xi) = (2\pi\tau)^{-3N/2} \exp\left(-\xi^2/2\tau\right)$

3. For each walker, compute the factor

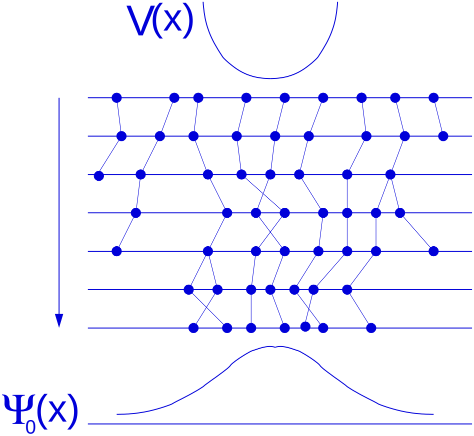
$$\boxed{p = \exp\left[-\tau(\mathcal{V}(\mathbf{R}) - E_{\mathrm{ref}})\right]}$$

   p is the probability to survive/proliferate/die

4. Adjust E_{ref} so that population fluctuates around target M_0

→ After many iterations, walkers distributed as $\Psi_0(\mathbf{R})$

Diffusion and branching in a harmonic potential



Walkers proliferate/die where potential is lower/higher than E_{ref}

Problems with simple algorithm

The simple algorithm is inefficient and unstable

- Potential can vary a lot and be unbounded

  e.g. electron-nucleus interaction → Exploding population

- Branching factor grows with system size

Importance sampling

Start from integral equation

$$\Psi(\mathbf{R}', t+\tau) = \int d\mathbf{R}\, G(\mathbf{R}', \mathbf{R}, \tau)\Psi(\mathbf{R}, t)$$

Multiply each side by trial Ψ_{T} and define $\boxed{\Pi(\mathbf{R}, t) = \Psi_{\mathrm{T}}(\mathbf{R})\Psi(\mathbf{R}, t)}$

$$\Pi(\mathbf{R}', t+\tau) = \int d\mathbf{R}\, \tilde{G}(\mathbf{R}', \mathbf{R}, \tau)\Pi(\mathbf{R}, t)$$

where the importance sampled Green's function is

$$\boxed{\tilde{G}(\mathbf{R}', \mathbf{R}, \tau) = \Psi_{\mathrm{T}}(\mathbf{R}')\langle \mathbf{R}' | e^{-\tau(\mathcal{H}-E_{\mathrm{ref}})} | \mathbf{R}\rangle / \Psi_{\mathrm{T}}(\mathbf{R})}$$

We obtain $\boxed{\lim_{n\to\infty} \Pi(\mathbf{R}) = \Psi_{\mathrm{T}}(\mathbf{R})\Psi_0(\mathbf{R})}$

Importance sampled Green's function

The importance sampled $\tilde{G}(\mathbf{R}, \mathbf{R}_0, \tau)$ satisfies

$$\boxed{-\frac{1}{2}\nabla^2 \tilde{G} + \nabla \cdot [\tilde{G}\,\mathbf{V}(\mathbf{R})] + [E_{\mathrm{L}}(\mathbf{R}) - E_{\mathrm{ref}}]\,\tilde{G} = -\frac{\partial \tilde{G}}{\partial \tau}}$$

with quantum velocity $\mathbf{V}(\mathbf{R}) = \dfrac{\nabla \Psi_{\mathrm{T}}(\mathbf{R})}{\Psi_{\mathrm{T}}(\mathbf{R})}$ and $E_{\mathrm{L}}(\mathbf{R}) = \dfrac{\mathcal{H}\Psi_{\mathrm{T}}(\mathbf{R})}{\Psi_{\mathrm{T}}(\mathbf{R})}$

We now have $\boxed{\text{drift}}$ in addition to diffusion and branching terms

Trotter's theorem $\Rightarrow$ Consider them separately for small enough τ

The drift-diffusion-branching Green's function

Drift-diffusion-branching short-time Green's function is

$$\tilde{G}(\mathbf{R}',\mathbf{R},\tau) = (2\pi\tau)^{-3N/2}\,\exp\left[-\frac{(\mathbf{R}'-\mathbf{R}-\tau\mathbf{V}(\mathbf{R}))^2}{2\tau}\right]\times$$
$$\times\,\exp\left\{-\tau\left(E_{\mathrm{L}}(\mathbf{R})-E_{\mathrm{ref}}\right)\right\}$$

What is new in the drift-diffusion-branching expression?

- $\mathbf{V}(\mathbf{R})$ pushes walkers where Ψ is large
- $E_{\mathrm{L}}(\mathbf{R})$ is better behaved than the potential $\mathcal{V}(\mathbf{R})$

  Cusp conditions $\Rightarrow$ No divergences when particles approach

  As $\Psi_{\mathrm{T}} \to \Psi_0$, $E_{\mathrm{L}} \to E_0$ and branching factor is smaller

Basic DMC algorithm with importance sampling

1. Sample initial walkers from $|\Psi_{\mathrm{T}}(\mathbf{R})|^2$
2. Drift and diffuse the walkers as $\mathbf{R}' = \mathbf{R} + \tau \mathbf{V}(\mathbf{R}) + \xi$
   where ξ is sampled from $g(\xi) = (2\pi\tau)^{-3N/2} \exp\left(-\xi^2/2\tau\right)$
3. Branching step as in the simple algorithm but with the factor

$$p = \exp\left\{-\tau[(E_{\mathrm{L}}(\mathbf{R}) + E_{\mathrm{L}}(\mathbf{R}'))/2 - E_{\mathrm{ref}}]\right\}$$

4. Adjust the trial energy to keep the population stable

→ After many iterations, walkers distributed as $\Psi_{\mathrm{T}}(\mathbf{R})\Psi_0(\mathbf{R})$

Electrons are fermions!

We assumed that $\Psi_0 > 0$ and that we are dealing with bosons

Fermions → Ψ is antisymmetric and changes sign!

Fermion Sign Problem

All fermion QMC methods suffer from sign problems

These sign problems look different but have the same “flavour”

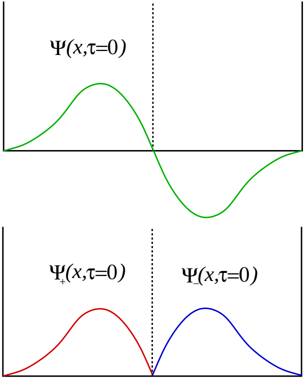
Arise when you treat something non-positive as probability density

The DMC Sign Problem

How can we impose antisymmetry in simple DMC method?

Idea Evolve separate positive and negative populations of walkers

Simple 1D example: Antisymmetric wave function $\Psi(x, \tau = 0)$



Rewrite $\Psi(x, \tau = 0)$ as

$$\Psi = \Psi_+ - \Psi_-$$

where

$$\begin{aligned} \Psi_+ &= \frac{1}{2}(|\Psi| + \Psi) \\ \Psi_- &= \frac{1}{2}(|\Psi| - \Psi) \end{aligned}$$

Particle in a box and the fermionic problem

The imaginary-time Schrödinger equation

$$\boxed{\mathcal{H}\Psi = -\frac{\partial\Psi}{\partial t}}$$

is linear, so solving it with the initial condition

$$\Psi(x, t=0) \;\; = \;\; \Psi_+(x, t=0) - \Psi_-(x, t=0)$$

is equivalent to solving

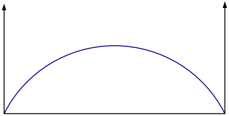
$$\boxed{\mathcal{H}\Psi_+ = -\frac{\partial\Psi_+}{\partial t}} \qquad \text{and} \qquad \boxed{\mathcal{H}\Psi_- = -\frac{\partial\Psi_-}{\partial t}}$$

separately and subtracting one solution from the other

Particle in a box and the fermionic problem (2)

- ▷ Since $E_0^s < E_0^a$, both Ψ_+ and Ψ_- evolve to Ψ_0^s

$\Psi_\pm \longrightarrow$



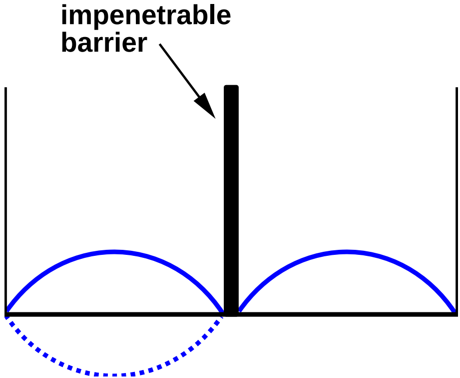
- ▷ Antisymmetric component exponentially harder to extract

$$\frac{|\Psi_+ - \Psi_-|}{|\Psi_+ + \Psi_-|} \propto \frac{e^{-E_0^a t}}{e^{-E_0^s t}} \quad \text{as} \quad t \to \infty$$

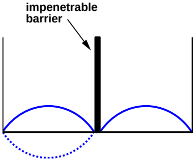
The Fixed-Node Approximation

Problem Small antisymmetric part swamped by random errors

Solution Fix the nodes! (If you don't know them, guess them)



Fixed-node algorithm in simple DMC



How do we impose this additional boundary condition?

▷ Annihilate walkers that bump into barrier (and into walls)

→ This step enforces $\boxed{\Psi = 0}$ boundary conditions

→ In each nodal pocket, evolution to ground state in pocket

Numerically $\boxed{\text{stable}}$ algorithm (no exponentially growing noise)

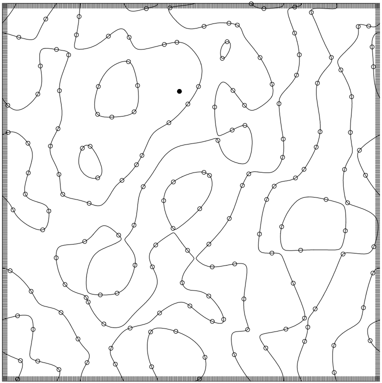
→ Solution is exact if nodes are exact

→ Best solution consistent with the assumed nodes

For many electrons, what are the nodes? A complex beast

Many-electron wave function $\Psi(\mathbf{R}) = \Psi(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N)$

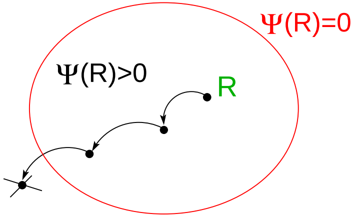
$\boxed{\text{Node}}$ → surface where $\Psi = 0$ and across which Ψ changes sign



A 2D slice through the 321-dimensional nodal surface
of a gas of 161 spin-up electrons.

Use the nodes of trial Ψ_{T} → Fixed-node approximation

Use the nodes of the best available trial Ψ_{T} wave function



Find best solution with same nodes as trial wave function Ψ_{T}

Fixed-node solution exact if the nodes of trial Ψ_{T} are exact

Easy to implement in DMC with importance sampling: $\Pi \geq 0$

Have we solved all our problems?

Results depend on the nodes of the trail wave function Ψ

Diffusion Monte Carlo as a black-box approach?

ϵ_{MAD} for atomization energy of the G1 set

DMC with optimized CAS wave functions

	DMC	CCSD(T)/aug-cc-pVQZ	
ϵ_{MAD}	1.2	2.8	kcal/mol

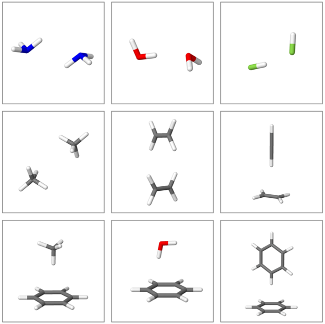
Petruzielo, Toulouse, Umrigar, JCP **136**, 124116 (2012)

With “some” effort on Ψ, we can do rather well

Diffusion Monte Carlo as a black-box approach?

Non-covalent interaction energies for 9 compounds from S22 set

DMC with B3LYP/aug-cc-PVTZ orbitals versus CCSD(T)/CBS



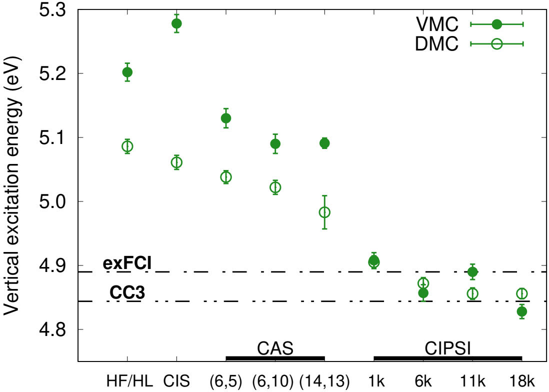
$\boxed{\Delta_{\mathrm{MAD}} = 0.058 \text{ kcal/mol}}$

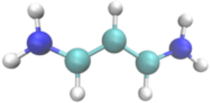
Dubecky *et al.*, JCTC **9**, 4287 (2013)

With "practically no" effort on Ψ, we can do rather well

Diffusion Monte Carlo end excitation energy

Excitation energy and wave function dependence





Cuzzocrea, Scemama, Briels, Moroni, Filippi, JCTC **16**, 4203 (2020)

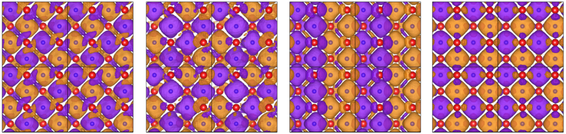
DMC is not a panacea but effort on Ψ pays off!

DMC and solid state calculations

Example: Structural/magnetic properties of superconducting FeSe

→ Accurate lattice constants, bulk moduli, and band dispersion

→ Resolving relative energetics of different magnetic ordering



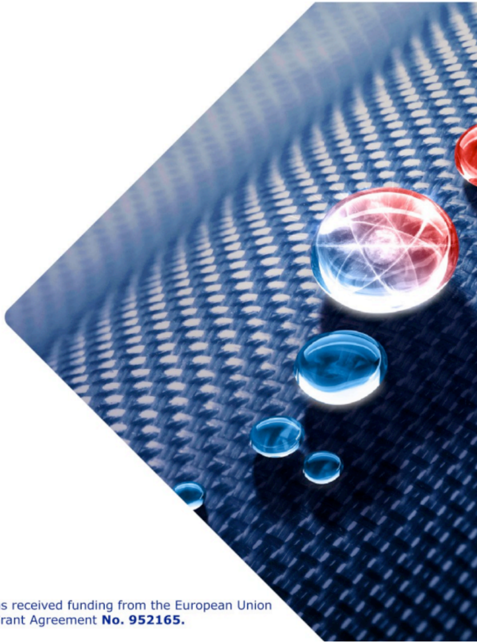
Busemeyer, Dagrada, Sorella, Casula, and Wagner PRB (2016)

Other applications of quantum Monte Carlo methods

- Electronic structure calculations
- Strongly correlated systems (Hubbard, t-J, . . .)
- Quantum spin systems (Ising, Heisenberg, XY, . . .)
- Liquid-solid helium, liquid-solid interface, droplets
- Atomic clusters
- Nuclear structure
- Lattice gauge theory

Both zero (ground state) and finite temperature

Targeting Real chemical accuracy at the EXascale



Targeting Real Chemical Accuracy at the Exascale project has received funding from the European Union Horizoon 2020 research and innovation programme under Grant Agreement **No. 952165.**