Quantum Monte Carlo (QMC)

December 1, 2022
Motivation

We simulate many-body quantum systems by solving the Schrödinger equation stochastically, because there are too many degrees of freedom in anything but trivial systems.

Before you begin, you need:

- Hamiltonian (fix particle masses and their interactions)
- Particle statistics: bosons, fermions, or sometimes boltzmannons (forget indistinguishability)
  - A non-relativistic Hamiltonian has no idea if particles are bosons or fermions
- Do we want properties for the ground state, for excited states, or for a finite temperature system?
- Do you really need a first-principles result? DFT is faster than QMC.
QMC methods

- **Variational Monte Carlo (VMC)**
  - Sample particle positions from a guessed wave function and simulate the consequences in a system with a certain Hamiltonian
  - You may improve your guess (for lower energy or variance), but it’s not automatic

- **Diffusion Monte Carlo (DMC)**
  - Usually for sampling particle positions in the ground state
  - Numerically improves upon a guessed wave function

- **Path Integral Monte Carlo (PIMC)**
  - Simulate finite temperature quantum systems
  - Manifest Feynman’s path integral representation of quantum mechanics
  - Also a ground-state variant exists
1940’s  E. Fermi: Schrödinger equation can be written as a diffusion equation; Physical principles of Diffusion Monte Carlo (DMC) discovered.

1942  A. A. Frost: Evaluation of the energy of few simple molecules by picking few representative points in the configuration space; Variational Monte Carlo (VMC) is born.

1947  N. Metropolis and S. Ulam: Diffusion Monte Carlo developed into an algorithm; “...as suggested by Fermi ...”.

1950’s  R. Feynman: Path integral formulation of quantum mechanics; Physical principles of Path Integral Monte Carlo (PIMC) discovered

1962  M. H Kalos: Solved the Schrödinger equation in the form of an integral equation: Green’s function Monte Carlo (GFMC) is born.

1964  H. Conroy: Proposed, that points should be picked at random from the configuration space with probability proportional to $\Psi^2_T$, the square of the trial wave function. Application to $H_2^+$, $H^-$, $HeH^{++}$, $He$, $H_2$, and $Li$ - the first application to a fermion system!
History

1965  W. L. McMillan: Energy of liquid $^4$He; Applied the Metropolis algorithm to sample points with probability proportional to $\Psi^2_T$.


1975  J. B. Anderson: First application of DMC to electronic systems

1980  D. M. Ceperley and B. J. Alder: fixed-node DMC for electron gas

1982  P. J. Reynolds, D. M. Ceperley, B. J. Alder and W. A. Lester, Jr.: fixed-node DMC for molecules

1986  D. M. Ceperley and E. L. Pollock: PIMC calculation of the properties of liquid $^4$He at finite temperatures.

1998  N.V. Prokofev, B. V. Svistunov, and I. S. Tupitsyn: Worm algorithm

2009  G. H. Booth, A. J. W. Thom, and A. Alavi: Fermion MC as a game of life in Slater determinant space
See QMC @ google scholar. The search “Quantum Monte Carlo” gave about 14900 hits for 2022 alone. Since 2018 about 38 800 hits, and since 2010 about 272 000 hits.

Some recent topics have been

▶ Fermion QMC; Auxiliary Field Monte Carlo
▶ Neural networks, machine learning and QMC
▶ Quantum computing and QMC
Computational cost

- Diffusion Monte Carlo (DMC) $\sim 10\times$ Variational Monte Carlo (VMC)
- Atomic and molecular physics: fixed-node diffusion Monte Carlo, coupled cluster method CCSD(T)
  - $N$ electrons, fixed-node DMC as $N^4$
  - CCSD(T) scales as $N^7 \Rightarrow$ small systems
    CCSD(T) stands for Coupled Cluster Singles Doubles (Triples), indicating the level of included diagrams; triples are partially included. The method is extremely accurate, but the workload is tremendous.
  - fixed-node DMC scales as $Z^{5.5-6.5}$ with the atomic number $Z$. 
When is MC effective?

The probability that the point \((x, y)\) is inside the circle is

\[
P(\text{hit circle}) = \frac{\text{Area of a quarter unit circle}}{\text{Area of a box}} = \frac{\pi}{4}. \tag{1}
\]

Very ineffective!

\[
I = \int_0^1 dx f(x). \tag{2}
\]

Estimate

\[
I \approx I_N = \frac{1}{N} \sum_{i=1}^{N} f(x_i), \tag{3}
\]

Multidimensional integrals: The partition function of \(N\) classical atoms is

\[
Z \propto \int d\mathbf{r}_1 d\mathbf{r}_2 \ldots d\mathbf{r}_N \exp \left[ -\beta V(\mathbf{r}_1, \ldots, \mathbf{r}_N) \right]. \tag{4}
\]
When is MC effective?

The variance or the square of the standard deviation of independent samples is

\[
\sigma_I^2 \approx \frac{1}{N} \left[ \frac{1}{N} \sum_{i=1}^{N} f(x_i)^2 - \left( \frac{1}{N} \sum_{i=1}^{N} f(x_i) \right)^2 \right],
\]

so

\[
\sigma_I \sim \frac{1}{\sqrt{N}}.
\]

If one has to evaluate the integrand \(N\) times to calculate a \(d\)-dimensional integral, then for each dimension one has \(\sim N^{1/d}\) points, whose distance is \(h \sim N^{-1/d}\). The error for every integration cell of volume \(h^d\) is in the trapezoid rule \(\sim h^{d+2}\), so the total error will be \(N\) times that, \(\sim N^{-2/d}\). For high-dimensional integrals this error diminishes very slowly with increasing number of sampled points \(N\).
When is MC effective?

Why is MC effective in multidimensional integrations? Heuristically, a 3D cube and a regular grid of $N$ points, take a 2D projection on a square: you see only $N^{2/3}$ points $\Rightarrow$ holes in space, significant parts of space never get any attention.

Ways to do better:

- deterministically: quasi-Monte Carlo methods; ”quasi” because the points are not random
- completely random points: Monte Carlo
Correlated samples and MC accuracy

For correlated samples (i.e. samples not statistically independent)

\[
\text{error in mean} = \sqrt{\frac{2\tau_{\text{int}}}{N}} \sigma ,
\]

where \(\sigma\) is the variance of \(N\) independent samples. The integrated autocorrelation time \(\tau_{\text{int}}\) describes how much the samples are correlated. The bad news is that close to a critical point \(\tau_{\text{int}}\) diverges. This is known as critical slowing down, and it tends to make MC calculations slow near phase transitions.

The term ”critical slowing down” can mean either

- Decreasing rate of recovery from small perturbations. It signals approaching criticality, such as an ecological disaster. Used as an early-warning mechanism.

- An MC algorithm may sample the phase space inefficiently due to long-range correlations near criticality. As a result, the MC simulation spends long time in very similar configurations. A better MC algorithm does not suffer from critical slowing down. A typical example is the Ising model simulation, single flip vs. cluster.
Importance sampling

In the hit-miss calculation of $\pi$, it would be most efficient to minimize the miss rate. The point is, that you must be able to calculate the area enclosing the unit circle. Obviously the most efficient hit-miss method would use the unit circle itself - but the area is $\pi$ which you presumable don’t know yet. You need a square or a polygon that preferably fits tightly around the unit circle.

As a formula, the square hit-miss method is

$$\pi = \int_{\text{circle}} d^2r \ 1 = \int_{-1}^{1} \int_{-1}^{1} dx dy f(x, y),$$

(8)

where $f(x, y) = 1$, if $x^2 + y^2 < 1$, and zero elsewhere. The integral $\int_{\text{circle}}$ would be the most efficient, but you can't sample points from the unit circle since you don’t know what $\pi$ is. In MC, $\int_{-1}^{1} \int_{-1}^{1} dx dy$ is evaluated by sampling random $x$ and $y$ values uniformly from [-1,1], that is, $x, y \in U[-1, 1]$. 
Importance sampling

The goal of importance sampling is to make the integrand as constant as possible. Let’s assume you need to compute

$$ l = \int_{-\infty}^{\infty} dx f(x) , \quad (9) $$

but you don’t know how to sample points $x_i$ from $f(x)$. If there is some $w(x) > 0$, from which you can sample points, then you can write

$$ \int_{-\infty}^{\infty} dx f(x) = \int_{-\infty}^{\infty} dx w(x) \frac{f(x)}{w(x)} \approx \frac{1}{N} \sum_{i=1}^{N} \frac{f(x_i)}{w(x_i)} , \quad (10) $$

with $x_i$ sampled from $w(x)$ and $w(x)$ chosen so that $\frac{f(x)}{w(x)}$ is as constant as possible. We can use a probability distribution function (PDF) $w(x)$,

$$ \int_{-\infty}^{\infty} dx w(x) = 1 \quad \text{normalization} . \quad (11) $$
Using inverse functions

The question is then how to sample points from a PDF $w(x)$.

A) Discrete distributions
We want to sample integers $k$ with $1 \leq k \leq N$ with probability $P_k$.

1. Construct the table of cumulative probabilities, $F_k = F_{k-1} + P_k$.
2. Sample $r \in U[0,1]$, find $k$ so that $F_{k-1} < r < F_k$. Repeat step 2.

B) Continuous distributions (mapping or inverse function method)

1. Calculate the cumulative distribution function (CDF) of $w(x)$,

   $$ F(x) = \int_{-\infty}^{x} dx' \ w(x') \quad SLOW; \ 1D \quad (13) $$

2. Sample $r \in U[0,1]$, solve $x$ from $F(x) = r$; the solutions $x$ sample $w(x)$. 
Rejection method

Simple case: distribution is limited from above, \( w(x) < M, \ x \in [a, b] \).

1. Draw \( r_1 \in U[a, b] \) and \( r_2 \in U[0, M] \) this picks a random point inside the box around \( w(x) \), see figure

2. If \( w(r_1) \geq r_2 \), accept \( x = r_1 \) as the desired random number, else discard both \( r_1 \) and \( r_2 \). Go to 1.

\[
\begin{align*}
    w(x) &= \int_0^M dr_2 \theta(w(x) - r_2) \\
    &= \int_a^b dr_1 \int_0^M dr_2 \delta(x - r_1)\theta(w(r_1) - r_2) . \quad (14)
\end{align*}
\]
Rejection method

Interpreting the identity as an algorithm:

\[ \theta(w(x) - r_2) = \begin{cases} 
0, & \text{if } r_2 > w(x) \\
1, & \text{if } r_2 < w(x) 
\end{cases}, \tag{15} \]

with some fixed \( x \), so integrating \( r_2 \) from 0 to \( M \) gives \( w(x) \). Next we use Dirac delta to move \( x \) out of the argument of \( w \), replacing it with \( r_1 \). The pieces can be now interpreted as an algorithm,

\[
w(x) = \int_a^b dr_1 \int_0^M dr_2 \delta(x - r_1) \theta(w(r_1) - r_2). \tag{16}\]

The rejection algorithm computes this Monte Carlo integral.
Special case: standard normal distribution

A crude algorithm

Draw $N$ random numbers $r_i \in U[0, 1]$ and calculate the sum

$$x = \sqrt{\frac{12}{N}} \sum_{i=1}^{N} (r_i - \frac{1}{2})$$

(17)

As $N$ increases this approaches the standard normal distribution, a consequence of the Central Limit Theorem (see MC-lecture.pdf for proof):

If you draw $N$ random numbers from any distribution that itself has a statistical mean $\mu$ and variance $\sigma$, their average is normally distributed with average $\mu$ and variance $\sigma/\sqrt{N}$, as $N \to \infty$.

Now we draw $N$ random numbers $x_i = r_i - 0.5 \in U[-\frac{1}{2}, \frac{1}{2}]$, a box-shaped distribution $f(x)$ with $\mu = \langle x \rangle = 0$ and $\sigma^2 = \langle (x - \langle x \rangle)^2 \rangle = \langle x^2 \rangle = \int_{-\infty}^{\infty} dx x^2 f(x) = \int_{-\frac{1}{2}}^{\frac{1}{2}} dx x^2 = \frac{1}{12}$, so $\sigma = \frac{1}{\sqrt{12}}$, and hence $\frac{1}{N} \sum_{i=1}^{N} (r_i - \frac{1}{2})$ approaches normal distribution with variance $\frac{1}{\sqrt{12N}}$. Multiply with $\sqrt{12N}$ and you get standard normal distribution.

This gives a rather good $\text{Normal}(x)$ already with small $N$, as $N = 6$ or $N = 12$. The large-$x$ tails are wrong, because you can’t get far with a finite number of steps.
Special case: standard normal distribution

Box-Müller algorithm I

1. Draw $r_1$ and $r_2 \in U[0, 1]$
2. Evaluate $\theta = 2\pi r_1$ and $R = \sqrt{-2 \ln r_2}$
3. $(x_1, x_2) = (R \cos(\theta), R \sin(\theta))$
4. Go to step 1.

Proof:

\[
\int_0^1 dr_1 \int_0^1 dr_2 \delta(x_1 - \sqrt{-2 \ln r_1 \cos(2\pi r_2)}) \delta(x_2 - \sqrt{-2 \ln r_1 \sin(2\pi r_2)})
\]
\[
= \int_0^1 dr_1 \int_0^1 dr_2 \delta(r_1 - e^{-(x_1^2 + x_2^2)/2}) \delta(r_2 - \frac{1}{2\pi} \arctan(\frac{x_2}{x_1})) \left| \frac{\partial(r_1, r_2)}{\partial(x_1, x_2)} \right|
\]
\[
= \frac{1}{2\pi} e^{-(x_1^2 + x_2^2)/2} = \frac{1}{\sqrt{2\pi}} e^{-x_1^2/2} \cdot \frac{1}{\sqrt{2\pi}} e^{-x_2^2/2} = \text{Normal}(x_1) \cdot \text{Normal}(x_2).
\]

The algorithm computes the lhs integrals using MC.

---

Special case: standard normal distribution

Box-Müller algorithm II
Uses the rejection method in order to avoid trig. functions.

1. Draw $r_1$ and $r_2 \in U[0, 1]$
2. Make $R_1 = 2r_1 - 1$ and $R_2 = 2r_2 - 1$ (they sample $\in U[-1, 1]$)
3. If $R = R_1^2 + R_2^2 > 1$ reject the pair and go to step 1
4. Result:
   \[
   (x_1, x_2) = \sqrt{-\frac{2 \ln R}{R}} \ (R_1, R_2) \quad \text{(19)}
   \]
5. Go to step 1
Fernández–Rivero algorithm

Fernández–Rivero algorithm\(^2\) is a small Maxwell-Boltzmann gas simulation with fictitious energy exchanges (step 5). After thermalization "particle" velocities \(v\) have normal distribution (step 8).

1. Initialise \(v_i = 1, i = 1, \ldots, N\) (\(N\) is about 1024) and set \(j = 0\).
2. Draw \(K\) (about 16) values \(\theta_k \in U[0, 2\pi]\).
   Compute \(c_k = \cos(\theta_k)\) and \(s_k = \sin(\theta_k)\).
3. Set \(j = (j + 1) \mod N\), draw \(i \in IU[0, N - 1]\) so that \(i \neq j\).
4. Draw \(k \in IU[0, K - 1]\).
5. Set \(tmp = c_k v_i + s_k v_j\) and \(v_j = -s_k v_i + c_k v_j\).
6. Set \(v_i = tmp\).
7. If still in thermalization go to step 3.
8. Result: \(v_i\) and \(v_j\).
9. Go to 3.

Here \(IU[0, N]\) means random integer between 0 and \(N\).

Markov Chain Monte Carlo

**Markov chain**

Markov Chain is a chain of states, visited with a rule to pick the next state $x_{i+1}$ from the current $x_i$. If we go on forever, states are visited with the weight $w(x)$ (normalized PDF is $\pi(x)$).

A Markov chain forms a **random walk**.

Here $\pi(x)$ is the **asymptotic, stationary, or limiting distribution**.

**Example 1:** Start the chain from $x_1 = 1$. The rule of the next value is $x_{i+1} = (x_i + 1) \mod 5$ (modulo). The chain is 1, 2, 3, 4, 5, 1, 2, 3, 4, 5, 1, .... As a result, the weight of each value is $w(x) = 1$, and the PDF is $\pi(x) = 1/5$.

**Example 2:** Start from $x_1 = 0$. The rule is $x_{i+1} = x_i + \sigma \eta$, where $\eta$ is a random number drawn from the standard normal distribution. The weight of each value is $w(x) = e^{-x^2/(2\sigma^2)}$, and the PDF $\pi(x) = 1/(\sigma \sqrt{2\pi}) e^{-x^2/(2\sigma^2)}$. 
The MC estimate with $M$ samples of an integral is computed as a sum,

$$
\int_{-\infty}^{\infty} dx \frac{1}{\sqrt{2\pi}} e^{-x^2/2} f(x) \approx \frac{1}{M} \sum_{i=1}^{M} f(x_i), \ x_i \text{ sampled from } \frac{1}{\sqrt{2\pi}} e^{-x^2/2}.
$$

We know how to draw *independent* samples from a standard normal distribution (Box-Müller, Fernández-Rivero *etc*.). On the previous slide we stated that the Markov chain with the rule $x_{i+1} = x_i + \sigma \eta$ has

$$
\pi(x) = \frac{1}{\sigma \sqrt{2\pi}} e^{-x^2/(2\sigma^2)}.
$$
Markov chain vs. direct sampling

Think about computing

\[ \int_{-\infty}^{\infty} w(x) f(x) \approx \frac{1}{M} \sum_{i=1}^{M} f(x_i). \]  

In principle, you can draw the \( M \) points \( x_i \) from distribution \( \pi(x) \) using

a) Direct sampling of \( \pi(x) \): Always do this if you can.

b) Markov chain \( \{x_1, \ldots, x_M\} \) that has the limiting distribution \( \pi(x) \).

Here a) is more effective, because the points \( x_i \) are uncorrelated. However, we have algorithms only for sampling a few simple distributions, and many-body wave functions are not among those. Markov chains are inferior in that the points are correlated, but superior in that you can sample an almost arbitrary distribution: All you need is the rule to generate the next point: finding it is our next topic.
**Detailed balance**

Master equation: The weight (probability if normalized) of a state $S'$ varies as a function of the simulation time (treated as continuous):

\[
\frac{d\pi(S')}{dt} = \sum_S P(S \rightarrow S')\pi(S) - \sum_S P(S' \rightarrow S)\pi(S'),
\]

(22)

Here $P(S \rightarrow S')$ is **the rule** to change from $S$ to $S'$, and we need one that gives $\frac{d\pi(S')}{dt} = 0$, because then we get time-independent $\pi(S)$ as the limiting, equilibrium distribution. The **detailed balance condition**\(^3\) is a solution:

\[
P(S \rightarrow S')\pi(S) = P(S' \rightarrow S')\pi(S'),
\]

(23)

where $\pi(S)$ is the weight of state $S$.

The detailed balance condition is a sufficient, but not necessary, condition that the system reaches the correct equilibrium distribution.

---

\(^3\) Also known as **microscopic reversibility**.
Make the change from state $S$ to $S'$ in two stages:

1. Suggest the change $S \to S'$ with a transition rule (matrix) $T(S \to S')$.
2. Accept the change $S \to S'$ with probability $A(S \to S')$.

so the rule can be written as

$$P(S \to S') = \underbrace{T(S \to S')}_{"Try"} \underbrace{A(S \to S')}_{"Accept"}. \quad (24)$$

This way it’s easier to take into account the fact that you may have a nonsymmetric transition rule, $T(S \to S') \neq T(S' \to S)$. Requiring that the detailed balance condition is satisfied still leaves an infinite number of possible rules $P(S \to S')$. **Which one is the most effective?** Meaning, which rule gives $\pi(S)$ fastest? Let’s assume for a while that $T(S \to S') = T(S' \to S)$ and look at the Metropolis answer in the quest of the most effective rule.
Metropolis algorithm

Original problem: sample states, so that they obey the Boltzmann PDF,

\[ \pi(S) = \frac{1}{Z} e^{-E(S)/T}, \quad k_B := 1. \quad (25) \]

Accept moves from state \(S\) to \(S'\) with probability

\[ A(S \rightarrow S') = \min \left[ 1, e^{-\Delta E/T} \right] = \min \left[ 1, \frac{\pi(S')}{\pi(S)} \right], \quad (26) \]

where \(\Delta E = E(S') - E(S)\). This Metropolis algorithm has two merits:

- The partition function \(Z\) is not involved.
  
  At equilibrium the master equation gives \(\sum_S P(S \rightarrow S')\pi(S) - \sum_{S'} P(S' \rightarrow S)\pi(S') = 0\), so any constant factors in \(\pi(S)\) cancels.

- It reaches the equilibrium distribution \(\pi(S)\) fast.
Metropolis-Hastings algorithm

If you try moving right more than left, you should accept left moves more easily. The detailed balance condition with nonsymmetric proposals \( T(S \rightarrow S') \neq T(S' \rightarrow S) \):

\[
\frac{P(S \rightarrow S')}{P(S' \rightarrow S)} = \frac{T(S \rightarrow S')A(S \rightarrow S')}{T(S' \rightarrow S)A(S' \rightarrow S)} = \frac{\pi(S')}{\pi(S)} \quad (27)
\]

\[
\Leftrightarrow \frac{A(S \rightarrow S')}{A(S' \rightarrow S)} = \frac{T(S' \rightarrow S)\pi(S')}{T(S \rightarrow S')\pi(S)} . \quad (28)
\]

The Metropolis-Hastings algorithm chooses

\[
A(S \rightarrow S') = \min \left[ 1, \frac{T(S' \rightarrow S)}{T(S \rightarrow S')} e^{-\Delta E/T} \right] = \min \left[ 1, \frac{T(S' \rightarrow S)\pi(S')}{T(S \rightarrow S')\pi(S)} \right] \quad (29)
\]
Variational Monte Carlo (VMC)

Energy of the trial state $|\varphi_T\rangle$ is

$$E_T = \frac{\langle \varphi_T | H | \varphi_T \rangle}{\langle \varphi_T | \varphi_T \rangle} = \frac{\int dR \varphi^*_T(R) H \varphi_T(R)}{\int dR \varphi_T(R)^2}$$

$$= \frac{\int dR |\varphi_T(R)|^2 \left[ \varphi_T(R)^{-1} H \varphi_T(R) \right]}{\int dR |\varphi_T(R)|^2}.$$  \hspace{1cm} (30)

The quantity

$$w(R) = \frac{|\varphi_T(R)|^2}{\int dR |\varphi_T(R)|^2}$$

is a good PDF $\Rightarrow$ can be sampled in MC.

Use Metropolis (Markov Chain) to draw samples from $|\varphi_T(R)|^2$.

Why: only ratios of weights matter $\Rightarrow \int dR |\varphi_T(R)|^2$ is irrelevant!
Metropolis VMC algorithm

*N* particles, wf $\varphi_T(R)$, $R = (r_1, \ldots, r_N)$

1. Pick initial configuration $R_1$ and set step size $step$.
3. Draw random displacement vector $d = (d_x, d_y, d_z)$, $d_x$, $d_y$ and $d_z \in U[-\frac{1}{2}, \frac{1}{2}]$
4. Suggest new position $r'_i = r_i + step \, d$
5. Compute $ratio = |\varphi_T(R')|^2 / |\varphi_T(R)|^2$
6. Metropolis question: Accept the move with probability $min[1, ratio]$. If $ratio > 1$ accept move. Else, draw $t \in U[0, 1]$ and if $ratio > t$ accept move. If move was accepted set $r_i = r'_i$, else keep old $r_i$.
7. Pick next particle, set $i = i + 1$. If $i < N$ go to 3.
8. If still in thermalization, go to 3. Adjust $step$ to keep acceptance $\in [40, 60]$
9. Accumulate averages, print out progress.
Energy optimization

Let’s look at two flavours, energy optimization and variance optimization. **Energy optimization relies on the Rayleight-Ritz variational principle:**

*Among states with the correct symmetry (Bose or Fermi), the ground state has the lowest energy.*

This ensures, that optimization of $\varphi_T(R; \alpha)$ wrt. energy,

$$\frac{\partial E(\alpha)}{\partial \alpha} = 0 \quad \text{for optimal } \alpha . \quad (33)$$

Caveats:

- The average $E(\alpha)$ cannot always be determined exactly – in MC it has stochastic error
  $$\Rightarrow \text{the minimum of the curve } \langle E(\alpha) \rangle \text{ has stochastic error}$$
  $$\Rightarrow \text{optimal } \alpha \text{ has stochastic error}$$

- The many-body wave function is usually not a simple elementary function
  $$\Rightarrow \text{optimal } \alpha \text{ gives optimal } \varphi_T(R; \alpha) \text{ only among a class of functions.}$$
Energy optimization

More serious caveats:

- The many-body wave function cannot be exactly represented by a finite number of samples \( \{ \mathbf{R} \} \).

- \( E(\alpha) \) is actually evaluated using a finite sample \( \{ \mathbf{R} \} = \{ \mathbf{R}_1, \ldots \mathbf{R}_N \} \), so \( E(\alpha) \approx \langle E(\alpha) \rangle_{\{ \mathbf{R} \}} \).
  
  - There is no natural lower limit to the energy \( \langle E(\alpha) \rangle_{\{ \mathbf{R} \}} \).
    \Rightarrow \) energy optimization may be unstable, stability may require a very big sample \( \{ \mathbf{R} \} \). This usually manifests itself as \( \langle E(\alpha) \rangle_{\{ \mathbf{R} \}} \) getting ridiculously low, because the small sample \( \{ \mathbf{R} \} \) happened to find a low-energy pithole.

This **stability problem** is one of the reasons why variance optimization may be more suitable than energy optimization.
Variance optimization

Variance optimization relies on the zero variance principle:

The variance in the variational energy vanishes for a trial wave function, which coincides with any eigenstate of the Hamiltonian.

Proof: Suppose \( \varphi_T(R; \alpha) = \phi_i(R) \), the \( i \)-th eigenstate of \( \hat{H} \). The variance of energy is

\[
\sigma^2 = \langle E(\alpha)^2 \rangle - \langle E(\alpha) \rangle^2 = 0,
\]

because

\[
\langle E(\alpha)^2 \rangle = \frac{\int dR |\varphi_T(R; \alpha)|^2 E_L(R; \alpha)^2}{\int dR |\varphi_T(R; \alpha)|^2} = \frac{\int dR |\phi_i(R)|^2 [\phi_i^{-1}(R) \hat{H} \phi_i(R)][\phi_i^{-1}(R) \hat{H} \phi_i(R)]}{\int dR |\phi_i(R)|^2}
\]

\[
= \frac{\int dR |\phi_i(R)|^2 E_i^2}{\int dR |\phi_i(R)|^2} = E_i^2
\]

(34)

and

\[
\langle E(\alpha) \rangle^2 = \left[ \frac{\int dR |\varphi_T(R; \alpha)|^2 E_L(R; \alpha)}{\int dR |\varphi_T(R; \alpha)|^2} \right]^2 = \left[ \frac{\int dR |\phi_i(R)|^2 E_i}{\int dR |\phi_i(R)|^2} \right]^2 = E_i^2.
\]

(35)
Energy vs. variance optimization

Comments on variance optimization:

- Variance – unlike energy – has a natural lower limit, $\sigma^2 \geq 0$.
- Variance-optimized energy $\geq$ energy-optimized energy
Energy vs. variance optimization

Energy optimization and variance optimization are not equivalent. Higher energy result can have lower variance and vice versa:

He atom:
Steps 0-25000: optimized energy, $E = -2.878$, $\sigma^2 = 0.11$.
Steps 25001- : optimized variance, $E = -2.869$, $\sigma^2 = 0.08$. 
Minimize an objective function, one possibility is

$$
\sigma_{\text{opt}}^2(\alpha) = \frac{\sum_{i=1}^{N_{\text{opt}}} w(R_i; \alpha) [E_L(R_i) - E_{\text{guess}}]^2}{\sum_{i=1}^{N_{\text{opt}}} w(R_i; \alpha)},
$$

where the weight function is

$$
w(R_i; \alpha) = \left| \frac{\varphi_T(R_i; \alpha)}{\varphi_T(R_i; \alpha_0)} \right|^2,
$$

and $R_i$ are sampled using $\varphi_T(R_i; \alpha_0)$ with a fixed $\alpha_0$, known as correlated sampling, to be explained on the next slide.
Correlated sampling

Direct minimization may be too inaccurate due to statistical fluctuations of data.

The differences of the correlated averages can be much more accurate than the averages themselves.

Energy with parameters $\alpha$:

$$
\langle E(\alpha) \rangle := \frac{\int d\mathbf{R} |\varphi_T(\mathbf{R}; \alpha)|^2 E_L(\mathbf{R}; \alpha)}{\int d\mathbf{R} |\varphi_T(\mathbf{R}; \alpha)|^2}
= \frac{\int d\mathbf{R} |\varphi_T(\mathbf{R}; \alpha_0)|^2 \left[ \frac{|\varphi_T(\mathbf{R}; \alpha)|^2}{|\varphi_T(\mathbf{R}; \alpha_0)|^2} E_L(\mathbf{R}; \alpha) \right]}{\int d\mathbf{R} |\varphi_T(\mathbf{R}; \alpha_0)|^2 \frac{|\varphi_T(\mathbf{R}; \alpha)|^2}{|\varphi_T(\mathbf{R}; \alpha_0)|^2}}.
$$

- Sample points $\mathbf{R}$ using weight $|\varphi_T(\mathbf{R}; \alpha_0)|^2$; parameters are $\alpha_0$
- Evaluate (a) $\frac{|\varphi_T(\mathbf{R}; \alpha)|^2}{|\varphi_T(\mathbf{R}; \alpha_0)|^2} E_L(\mathbf{R}; \alpha)$ (b) $\frac{|\varphi_T(\mathbf{R}; \alpha)|^2}{|\varphi_T(\mathbf{R}; \alpha_0)|^2}$
Correlated sampling

Observations about correlated sampling:

- Variance of individual values $\langle E(\alpha) \rangle$ is larger than sampling from $|\varphi_T(R; \alpha)|^2$ directly
  $\Rightarrow$ Correlated sampling is no good for absolute energy values

- Variance of differences $\langle E(\alpha_1) \rangle - \langle E(\alpha_2) \rangle$ can be smaller than what it would be if $\langle E(\alpha_1) \rangle$ was sampled from $|\varphi_T(R; \alpha_1)|^2$ and $\langle E(\alpha_2) \rangle$ was sampled from $|\varphi_T(R; \alpha_2)|^2$:
  Fluctuations ”cancel in the same direction”, the differences
  \[ \langle E(\alpha_1) \rangle - \langle E(\alpha_2) \rangle, \quad (39) \]
  are more accurate.
  $\Rightarrow$ Correlated sampling is good for hunting energy differences.

This makes correlated sampling suitable for optimization of the trial wave function.
Correlated sampling
Energy optimization with correlated samples

The energy-optimized value is 1.6875, which can be found also analytically – see next page.
Optimization on paper

With a 1S product wf the optimization can be done by hand. e.g. He,

$$\langle \varphi_T(R)|-\frac{1}{2}\nabla_i^2|\varphi_T(R)\rangle = \frac{1}{2}\alpha^2$$  \hspace{1cm} (40)

$$\langle \varphi_T(R)|-\frac{2}{r_i}|\varphi_T(R)\rangle = -2\alpha$$  \hspace{1cm} (41)

$$\langle \varphi_T(R)|-\frac{1}{r_{ij}}|\varphi_T(R)\rangle = \frac{5}{8}\alpha$$,  \hspace{1cm} (42)

so (this is not the local energy, it’s the energy of the trial state)

$$E(\alpha) = \alpha^2 - 4\alpha + \frac{5}{8}\alpha$$.  \hspace{1cm} (43)

This gives optimized value \( \alpha = 2 - \frac{5}{16} = \frac{27}{16}. \) The first elements give (red indicates violation of Pauli exclusion principle)\(^4\)

<table>
<thead>
<tr>
<th>elem</th>
<th>( \alpha )</th>
<th>( E(\alpha) ) (au)</th>
<th>Exact (au)</th>
</tr>
</thead>
<tbody>
<tr>
<td>H</td>
<td>1</td>
<td>-1/2=-0.5</td>
<td>-0.5</td>
</tr>
<tr>
<td>He</td>
<td>27/16=1.6875</td>
<td>-729/256=-2.8476</td>
<td>-2.90372</td>
</tr>
<tr>
<td>Li</td>
<td>19/8=2.375</td>
<td>-1083/128=-8.4609</td>
<td>-7.47806</td>
</tr>
<tr>
<td>Be</td>
<td>49/16=3.0625</td>
<td>-2401/128=-18.75781</td>
<td>-14.66736</td>
</tr>
</tbody>
</table>

\(^4\)Sympy code energies.py; pretend electrons are bosons and nucleus has infinite mass.
Virial theorem and full optimization

Virial theorem for Coulomb systems (last term is added in the Born-Oppenheimer approximation, nuclear coordinates \( r_I \), electronic energy \( E_e \)).

\[
2\langle \hat{T} \rangle + \langle \hat{V} \rangle + \sum_I r_I \cdot \frac{\partial E_e}{\partial r_I} = 0 .
\] (44)

A good check, but can’t really tell if one trial wave function is better than the other.

**Why not do full optimization?**

Find the *normalized wave function* of lowest energy:

\[
E[\Psi, \Psi^*, \lambda] = \int dR \Psi^*(R) \hat{H} \Psi(R) - \lambda \left( \int dR |\Psi(R)|^2 \right)
\]

vary \( \Psi^* \): \( \delta E = \int dR \left[ \hat{H} \Psi(R) - \lambda \Psi(R) \right] \delta \Psi(R)^* . \) (45)

Take arbitrary \( \delta \Psi(R)^* \), and we are back to Schrödinger equation

\[
\hat{H} \Psi(R) = \lambda \Psi(R) .
\] (46)
Cusp conditions

In the local energy, infinities of the potential energy should be cancelled by those of the kinetic energy. These happen if

- two electrons are too close
- electron get too close to nucleus

Concentrate on two electrons, keeping the rest fixed: relative \( r = r_1 - r_2 \) and center-of-mass \( R_{CM} = \frac{1}{2}(r_1 + r_2) \). Since (see lecture notes for details)

\[
\nabla^2_1 + \nabla^2_2 = 2\nabla^2_r + \frac{1}{2}\nabla^2_{R_{CM}},
\]

(47)

the Hamiltonian is

\[
\hat{\mathcal{H}} = -\left(\frac{1}{4}\nabla^2_{R_{CM}} + \nabla^2_r\right) - \frac{1}{2} \sum_{i=3}^{N} \nabla^2_i + V(r, R_{cm}, r_3, \ldots, r_N).
\]

(48)

---

Cusp conditions

Try a trial wf of Jastrow type,

$$\psi(R) = \psi(r, R_{CM}, r_3, ..., r_N) = e^{-u(r)} f(r) Q(R_{CM}, r_3, ..., r_N),$$

where

1. \(f(0) \neq 0\), if the spins at \(r_1\) and at \(r_2\) are anti-parallel
2. \(f(0) = 0\), and \(f(-r) = -f(r)\), if the spins at \(r_1\) and at \(r_2\) are parallel

The potential part diverges as \(1/r\), so we want to keep this \(h(r)\) finite:

$$h(r) = \frac{1}{e^{-u(r)} f(r)} \left(-\nabla^2_r + \frac{1}{r}\right) e^{-u(r)} f(r).$$

(50)

Expand (leave out the arguments \(r\)),

$$h = \frac{2u'}{r} + u'' - u'^2 + \frac{1}{r} + 2u'\frac{1}{f} \cdot \nabla_r f - \frac{1}{f} \nabla^2_r f,$$

(51)

where \(u' := du/dr\). Look at the two cases separately.
Cusp conditions

Case 1: anti-parallel spins, $f(0) \neq 0$

Only 1. and 4. terms are singular,

$$h = \frac{2u'}{r} + u'' - u'^2 + \frac{1}{r} + 2u' \frac{1}{f} \frac{r}{r} \cdot \nabla_r f - \frac{1}{f} \nabla^2_r f ,$$ (52)

so we require that

$$\frac{2u'}{r} + 1 \text{ finite ,}$$ (53)

and get the cusp condition

$$\left. \frac{du}{dr} \right|_{r=0} = -\frac{1}{2} .$$ (54)

Example: He atom with e-e term (Hartree atomic units)

$$\varphi_T = e^{-u(r)} = e^{\frac{r}{2(1+\beta r)}} , \quad r := |r_1 - r_2| ; \quad 2 \text{ from the cusp.}$$
Cusp conditions

Case 2: parallel spins, \( f(0) = 0 \), and \( f(-r) = -f(r) \)

Taylor expansion of \( f(r) \) around origin is

\[
f(r) = a \cdot r + \mathcal{O}(r^3),
\]

which has only odd terms because \( f(-r) = -f(r) \). Close to origin we get

\[
h \approx \frac{2u'}{r} + u'' - u'^2 + \frac{1}{r} + 2u' \frac{1}{a \cdot r} \cdot a = \frac{4u'}{r} + \frac{1}{r},
\]

so the cusp condition is

\[
\left. \frac{du}{dr} \right|_{r=0} = -\frac{1}{4}.
\]
Current state-of-the-art optimization strategy

Currently the most popular optimization is by Umrigar et al. \(^6\). The wave function is linearized wrt. parameters \(p\) around \(p_0\),

\[
\Psi_{\text{lin}}(p, R) = \underbrace{\Psi_0(R)}_{\Psi(p_0, R)} + \sum_{i=1}^{N_{\text{opt}}} \Delta p_i \underbrace{\Psi_i(R)}_{\frac{\partial \Psi(p, R)}{\partial p_i} |_{p=p_0}} .
\]  

(58)

The parameter updates \(\Delta p_i\) are the lowest eigensolutions of

\[
H \Delta p = ES \Delta p ,
\]

(59)

where the electronic Hamiltonian matrix \(H\) and the overlap matrix \(S\) are approximated by sampled averages

\[
H_{ij} = \left\langle \frac{\Psi_i}{\Psi_0} \hat{H} \frac{\Psi_j}{\Psi_0} \right\rangle_{\psi_0^2} , \quad S_{ij} = \left\langle \frac{\Psi_i}{\Psi_0} \frac{\Psi_j}{\Psi_0} \right\rangle_{\psi_0^2} .
\]

(60)

Current state-of-the-art optimization strategy

After some fine tuning\(^7\) the algorithm has turned out to be very robust. Some results by Umrigar et al.:

![Convergence of VMC total energy of the C\(_2\) molecule](image1)

**FIG. 1** (color online). Convergence of the VMC total energy of the C\(_2\) molecule when simultaneously optimizing 24 Jastrow, 73 CSF and 174 orbital parameters for a truncated CAS(8,14) wave function. The number of MC configurations range from 10,000 for the first iteration to 400,000 for the last iterations.

![Energy graphs for C\(_2\) and Si\(_2\)](image2)

**FIG. 2** (color online). VMC and DMC total energies of the C\(_2\) and Si\(_2\) molecules, versus the number of determinants in truncated Jastrow-Slater CAS(8,14) wave functions, for different levels of optimization. Both VMC and DMC energies decrease monotonically if the CSF coefficients are reoptimized in VMC but not if only the Jastrow factor is optimized. Statistical errors are smaller than the plotted symbols.

\(^7\) Most notably Hermiticity of \(H\) is relaxed.
Diffusion Monte Carlo (DMC)

Schrödinger equation in imaginary time \( \Rightarrow \) diffusion equation

\[
- \frac{\partial |\Psi(\tau)\rangle}{\partial \tau} = \hat{H} |\Psi(\tau)\rangle .
\]  

(61)

formal solution is (if \( \hat{H} \) does not depend on \( \tau \))

\[
|\Psi(\tau)\rangle = e^{-\tau \hat{H}} |\Psi(0)\rangle \quad \text{projection} .
\]  

(62)

The eigenstates \( |\phi_i\rangle \) of \( \hat{H} \) are solutions of (can’t solve, but they do exits)

\[
\hat{H} |\phi_i\rangle = E_i |\phi_i\rangle , \quad E_0 < E_1 < E_2 < \ldots .
\]  

(63)

Expand the initial \( N \)-body state in the basis of eigenstates of \( \hat{H} \),

\[
|\Psi(0)\rangle = \sum_i c_i |\phi_i\rangle ,
\]  

(64)

and operate on this with \( e^{-\tau \hat{H}} \),

\[
|\Psi(\tau)\rangle = \sum_i c_i e^{-\tau E_i} |\phi_i\rangle .
\]  

(65)
Diffusion Monte Carlo (DMC)

Stabilise the evolution by subtracting a trial or reference energy $E_T \approx E_0$:

$$ |\Psi(\tau)\rangle = e^{-\left(\hat{H}-E_T\right)\tau} |\Psi(0)\rangle = \sum_i c_i e^{-\left(E_i-E_T\right)\tau} |\phi_i\rangle. \quad (66) $$

States with $E_i > E_T$ will decay. The projection operator is Green’s function

$$ \hat{G}(\tau) = e^{-\left(\hat{H}-E_T\right)\tau}, \quad (67) $$

satisfies the equation

$$ -\frac{\partial \hat{G}(\tau)}{\partial \tau} = (\hat{H} - E_T)\hat{G}. \quad (68) $$

Represented by a coordinate-space matrix with elements

$$ \langle R' | \hat{G}(\tau) | R \rangle = \langle R' | e^{-\left(\hat{H}-E_T\right)\tau} | R \rangle := G(R', R; \tau). \quad (69) $$
Short time $G(\tau)$: Operator splitting

Usually $\hat{H} = \hat{T} + \hat{V}$, and $[\hat{T}, \hat{V}] \neq 0 \Rightarrow$ we don't know $e^{-(\hat{H}-E_T)\tau}$ exactly. But we can approximate,

$$G(\tau) = e^{-(\hat{H}-E_T)\tau} = e^{-(\hat{T}+\hat{V}-E_T)\tau} = e^{-\hat{T}\tau}e^{-(\hat{V}-E_T)\tau} + \mathcal{O}(\tau^2) . \quad (70)$$

The potential part is easy,

$$G_{\hat{V}}(\mathbf{R}', \mathbf{R}; \tau) = e^{-(V(\mathbf{R})-E_T)\tau} \delta(\mathbf{R}' - \mathbf{R}) . \quad (71)$$

The kinetic part is a bit trickier (to be derived later),

$$G_{\hat{T}}(\mathbf{R}', \mathbf{R}; \tau) = (4\pi\lambda\tau)^{-3N/2} e^{-\frac{(\mathbf{R}-\mathbf{R}')^2}{4\lambda\tau}} . \quad (72)$$

Put together, we get the ”primitive approximation”,

$$G(\mathbf{R}', \mathbf{R}; \tau) \approx (4\pi\lambda\tau)^{-3N/2} e^{-\frac{(\mathbf{R}-\mathbf{R}')^2}{4\lambda\tau}} e^{-(V(\mathbf{R})-E_T)\tau} . \quad (73)$$
**Diffusion $G_{\hat{T}}(\tau)$**

Let’s derive the coordinate-space representation of the diffusion Green’s function. We use the eigenstate expansion ($\mathbf{K}$ is a $3N$-dimensional momentum space vector),

$$
\hat{T} \phi_{\mathbf{K}}(\mathbf{R}) = (-\lambda \nabla_{\mathbf{R}}^2) \phi_{\mathbf{K}}(\mathbf{R}) = \lambda K^2 \phi_{\mathbf{K}}(\mathbf{R})
$$

(74)

$$
\phi_{\mathbf{K}}(\mathbf{R}) = \langle \mathbf{R} | \mathbf{K} \rangle = \frac{1}{\sqrt{L^{3N}}} e^{-i\mathbf{K} \cdot \mathbf{R}},
$$

(75)

and the states $|\mathbf{K}\rangle$ are a complete basis. The coordinate-space representation of $G_{\hat{T}}(\tau)$ is

$$
G_{\hat{T}}(\mathbf{R}', \mathbf{R}; \tau) := \langle \mathbf{R}' | e^{-\tau \hat{T}} | \mathbf{R} \rangle = \sum_{\mathbf{K}} \langle \mathbf{R}' | e^{-\tau \hat{T}} | \mathbf{K} \rangle \langle \mathbf{K} | \mathbf{R} \rangle
$$

(76)

$$
= \sum_{\mathbf{K}} \langle \mathbf{R}' | e^{-\tau \lambda K^2} | \mathbf{K} \rangle \phi_{\mathbf{K}}^*(\mathbf{R}) = \sum_{\mathbf{K}} \phi_{\mathbf{K}}(\mathbf{R}') \phi_{\mathbf{K}}^*(\mathbf{R}) e^{-\tau \lambda K^2}
$$

(77)

$$
= \frac{1}{L^{3N}} \sum_{\mathbf{K}} e^{-i\mathbf{K} \cdot (\mathbf{R}' - \mathbf{R})} e^{-\tau \lambda K^2}.
$$

(78)
Diffusion $G_{\hat{T}}(\tau)$

The momentum values are $2\pi/L$ apart, and for large $L$ the sum can be converted to an integral. \(^8\)

$$\sum_{\mathbf{K}} \approx \frac{L^{3N}}{(2\pi)^{3N}} \int d\mathbf{K}$$  \hspace{1cm} (79)

$$\frac{1}{L^{3N}} \sum_{\mathbf{K}} e^{-i\mathbf{K}\cdot(\mathbf{R}^\prime - \mathbf{R})} e^{-\tau\lambda K^2} \approx \frac{1}{(2\pi)^{3N}} \int d\mathbf{K} e^{-i\mathbf{K}\cdot(\mathbf{R}^\prime - \mathbf{R})} e^{-\tau\lambda K^2},$$  \hspace{1cm} (80)

where the integral is a Fourier transform of a Gaussian, leading to the nice result

$$G_{\hat{T}}(\mathbf{R}', \mathbf{R}; \tau) = (4\pi\lambda\tau)^{-3N/2} e^{-\frac{(\mathbf{R} - \mathbf{R}')^2}{4\lambda\tau}}.$$  \hspace{1cm} (81)

This is usually an excellent approximation, so I left out the "≈".

---

\(^8\) If not, evaluate the sum.
Simulation of $G_{\hat{T}}(\tau)$: Diffusion

We want to compute evolution $|\Psi(\tau)\rangle = G_{\hat{T}}(\tau)|\Psi(0)\rangle$, which is

$$\psi(R', \tau) = \int dRG_{\hat{T}}(R', R; \tau)\psi(R, 0) \approx \frac{1}{M} \sum_{i=1}^{M} \psi(R_i, 0), \tag{82}$$

where $R_i$ is sampled from $G_{\hat{T}}(R', R; \tau)$ (fixed $R'$ and $\tau$).

$G_{\hat{T}}(R', R; \tau)$ is a product of $3N$ Gaussians (in 3D), so all of them can be sampled independently $\Rightarrow 3N$ independent diffusions – but what is $\sigma$?

Compare the exponents:

$$2\sigma^2 = 4\lambda\tau \Rightarrow \sigma = \sqrt{2\lambda\tau}. \tag{83}$$

Result: we can sample points $R_i$ by a random walk with the rule

$$R_{i+1} = R_i + \sqrt{2\lambda\tau} \eta, \tag{84}$$

where $\eta = (\eta_x, \eta_y, \eta_z)$, $\eta_{x,y,z}$ sample standard normal distribution.
MC representation of the many-body wave function

In VMC we had an analytical wave function, but in DMC we don’t have an analytical form for $\Psi(R, \tau)$. Instead, DMC represents $\Psi(R, \tau)$ in terms of a finite number of samples $\{R\}$. These configurations are made of "walkers" (they walk a Markov chain).

Consequences:

1. Need quite a large sample $\{R\}$ to represent $\Psi(R, \tau)$
   - Use of the order of 100–10000 (depending on the problem) configurations $R$, each is one possible configuration in $\Psi(R, \tau)$
   - Slower than VMC by about the number of configurations
   - Many walkers $\Rightarrow$ consumes memory
   - Each configuration is independent $\Rightarrow$ very good parallelization! Only now and then collect data and do branching.

2. Too few configurations means $\Psi(R, \tau)$ is poorly represented.
   $\Rightarrow$ The algorithm must keep the number of configurations large – but not too large, or you’re out of memory.
First DMC algorithm

Pick a large number of configurations \( \{ \mathbf{R} \} \) to begin with.

1. **Diffusion** due to \( G_{\hat{\mathcal{V}}} (\tau) \): Move from \( \mathbf{R} \rightarrow \mathbf{R}' \) diffusively, in other words move each walker according to

   \[ r'_i = r_i + \eta \sqrt{2 \lambda \tau} , \]

   where \( \eta \) is a d-dimensional Gaussian random variable with zero mean and unit variance.

2. **Branching** due to \( G_{\hat{\mathcal{V}}} (\tau) \): The wave function \( \Psi(\mathbf{R}, 0) \) is multiplied by the factor \( e^{-\left( V(\mathbf{R}) - E_T \right) \tau} \). So we calculate \( V(\mathbf{R}) \) for each configuration and see whether this factor is more or less than one, and create/destroy configurations accordingly. Walkers are not moving.

Alas, this is no good if the potential is singular: diffusion takes walkers sooner or later to a near-singular position.
First DMC algorithm

The factor \( e^{-(V(R)-E_T)\tau} \) makes the algorithm problematic:

▶ Repulsive potential, \( V(R) \gg 0 \) if particles come too close. The walker is killed in branching. Easily lots of killed walkers \( \Rightarrow \) only a few may survive.

▶ Attractive potential, \( V(R) \ll 0 \), if particles come too close. The walker branches to very many walkers. \( \Rightarrow \) bad statistics and large fluctuations in the number of walkers.

One way to see where problems are is to realize, that the sole energy measure is the potential energy, so the algorithm would have to produce coordinates \( \{R\} \), where

\[
\langle V(R) \rangle_R = E_0 .
\]  

(85)

Conclusion: Improve the algorithm so that walkers stay away from regions of large \( |V(R)| \). This can be done with importance sampling.
Importance sampling in DMC

Solve the imaginary-time evolution of the product

\[ f(R; \tau) := \psi(R; \tau)\varphi_T(R), \quad (86) \]

knowing that the evolution of \( \psi(R; \tau) \) is given by the imaginary-time Schrödinger equation,

\[ -\frac{\partial \psi(R; \tau)}{\partial \tau} = (\hat{\mathcal{H}} - E_T)\psi(R; \tau). \quad (87) \]

Insert \( \psi(R; \tau) = f(R; \tau)/\varphi_T(R) \) to get

\[ -\frac{\partial f(R; \tau)}{\partial \tau} = -\lambda \nabla^2_R f(R; \tau) + \lambda \nabla_R \cdot [F(R)f(R; \tau)] + [E_L(R) - E_T] f(R; \tau). \quad (88) \]

Define a drift force (or drift velocity, quantum force)

\[ F(R) := \nabla_R \ln |\varphi_T(R)|^2 = 2 \frac{\nabla_R \varphi_T(R)}{\varphi_T(R)}, \quad (89) \]

and the familiar local energy \( E_L(R) := \varphi_T(R)^{-1}\hat{\mathcal{H}}\varphi_T(R) \).
Green’s function for importance sampling DMC

The imaginary-time evolution is

\[
- \frac{\partial f(R, \tau)}{\partial \tau} = -\lambda \nabla_R^2 f(R, \tau) + \lambda \nabla_R \cdot [F(R)f(R, \tau)] + [E_L(R) - E_T] f(R, \tau) \\
:= \hat{H}_F f(R, \tau)
\]  

(90)

which defines the operator $\hat{H}_F$. The formal solution is

\[
f(R, \tau) = e^{-\tau \hat{H}_F} f(R, 0)
\]  

(91)

and the Green’s function is

\[
G(R, R', \tau) = \langle R' | e^{-\tau \hat{H}_F} | R \rangle.
\]  

(92)

Next, we can write the evolution as an integral equation,

\[
f(R, \tau) = \int dR' G(R, R'; \tau) f(R', 0).
\]  

(93)
Green’s function for importance sampling DMC

Without importance sampling we did two things:

1. Decide how accurate Green’s function we want. We settled for the most crude Trotter expansion,

\[ e^{-\tau \hat{H}} \approx e^{-\tau \hat{T}} e^{-\tau \hat{V}} + O(\tau^2) . \]  \hfill (94)

2. Solve \( G_{\hat{T}}(\mathbf{R}', \mathbf{R}; \tau) = \langle \mathbf{R}' | e^{-\tau \hat{T}} | \mathbf{R} \rangle \) and \( G_{\hat{V}}(\mathbf{R}', \mathbf{R}; \tau) = \langle \mathbf{R}' | e^{-\tau \hat{V}} | \mathbf{R} \rangle \)

Importance sampling adds one complication, namely the drift term

\[ \hat{H}_F f(\mathbf{R}, \tau) = -\lambda \nabla^2_{\mathbf{R}} f(\mathbf{R}, \tau) + \lambda \nabla_{\mathbf{R}} \cdot [F(\mathbf{R}) f(\mathbf{R}, \tau)] + [E_L(\mathbf{R}) - E_T] f(\mathbf{R}, \tau) \]

\[ := \hat{T}_F f(\mathbf{R}, \tau) + [E_L(\mathbf{R}) - E_T] f(\mathbf{R}, \tau) . \]  \hfill (95)

Evolution with only diffusion and drift is known as the **Fokker-Planck equation**. Solving the Fokker-Planck evolution is a crucial part of all DMC algorithms.
Fokker-Planck operator

The drift-diffusion a.k.a Fokker-Planck evolution is

\[-\frac{\partial f(R, \tau)}{\partial \tau} = -\lambda \nabla_R^2 f(R, \tau) + \lambda \nabla_R \cdot [F(R)f(R, \tau)] := \hat{T}_F f(R, \tau)\] (96)

Define the drift operator,

\[F(\hat{R})\langle R \rangle := F(R)\langle R \rangle.\] (97)

to get

\[\nabla_R \cdot [F(R)f(R, \tau)] = \nabla_R \cdot \left[ \langle R|F(\hat{R})|f(\tau)\rangle \right] = \nabla_R \cdot F(\hat{R})f(R, \tau).\] (98)

so the Fokker-Planck operator is

\[\hat{T}_F = -\lambda \nabla_R^2 + \lambda \nabla_R \cdot F(\hat{R}).\] (99)
Fokker-Planck operator

We need to evaluate

\[ G_{TF}(R', R; \tau) := \langle R' | e^{-\tau \hat{T}_F} | R \rangle. \quad (100) \]

For diffusion we just solved the eigenfunctions of \( \hat{T} \) and wrote \( \langle R' | e^{-\tau \hat{T}} | R \rangle \) using an eigenfunction expansion, so why not repeat the procedure:

\[ \hat{T}_F Q_Q(R) = q_Q Q_Q(R) \quad (not \ valid \ !) \quad (101) \]
\[ G_{TF}(R', R, \tau) = \sum_Q e^{-\tau q_Q} Q_Q^*(R') Q_Q(R) \quad (not \ valid \ !). \quad (102) \]

So what’s the problem? An eigenfunction expansion is valid only for hermitian operators:

*The Fokker-Planck operator \( \hat{H}_F \) is non-hermitian* \(^9\),

Another fine mess importance sampling got us into!

---

\(^9\) Non-self-adjoint, to be more precise.
Fokker-Planck evolution – again

Let’s make a new start and look again at the Fokker-Planck equation,

\[- \frac{\partial f(R, \tau)}{\partial \tau} = -\lambda \nabla^2_R f(R, \tau) + \lambda \nabla_R \cdot [F(R) f(R, \tau)] \,
\]

Instead of working with the evolution of the probability density \( f(R, \tau) \), we can write an equivalent stochastic PDE for time \textit{evolution of the observable}, \( R(\tau) \). The equation for \( R(\tau) \) is a \textbf{Langevin equation}.

The basic problem is obvious: diffusion and drift don’t commute. We can try splitting them to two separate evolution processes, the way the did with \( \hat{T} \) and \( \hat{V} \). Be careful, you’d wan’t the same accuracy in \( \tau \) as you had in the original operator splitting. A higher-than-necessary accuracy may be too costly and a lower accuracy taints the algorithm. If you are happy with \( O(\tau^2) \) accuracy, then by all means use

\[
G_F(\tau) = e^{-\tau(-\lambda \nabla^2_R + \lambda \nabla_R \cdot F(\hat{R}))} \approx e^{-\tau\hat{T}} e^{-\tau \lambda \nabla_R \cdot F(\hat{R})} + O(\tau^2) ,
\]

where the first part is diffusion, the second is drift. But how to get \( O(\tau^3) \) accuracy?
Drift evolution

Let’s first concentrate on solving the drift evolution,\(^ \text{10} \)

\[
- \frac{\partial f(R, \tau)}{\partial \tau} = \lambda \nabla_R \cdot [F(R)f(R, \tau)] \quad \text{(only drift)}.
\]

(105)

If you interpret

\[f(R, \tau)\] is density \hspace{1cm} (106)

\[
\lambda F(R)f(R, \tau)\] is current \hspace{1cm} (107)

then the drift equation is a \textit{continuity equation}, usually written for density \(\rho\) and current \(j\) as \(\partial \rho / \partial t = \nabla \cdot j\). The drift Green’s function describes propagation of particles along a trajectory,

\[
\langle R' | e^{-\tau \lambda \nabla_R \cdot F(R)} | R \rangle = \delta(R' - R(\tau)),
\]

(108)

so

\textit{what are the trajectories} \(R(\tau)\) of particles under drift \(F(R)\)?

\(^{10}\)I apologize of putting so much time on higher order operator splitting. I’m interested in the DMC algorithm itself, but if you aren’t like me, just skip a few pages.
Drift evolution

The trajectories are solutions of\(^{11}\)

\[
\frac{dR(\tau)}{d\tau} = \lambda F(R(\tau)) .
\] (109)

The solution accurate to linear order in \(\tau\) is

\[
R(\tau) = R(0) + F(R(0))\lambda \tau + O(\tau^2) .
\] (110)

To linear order in \(\tau\) we can assume that the drift doesn’t change. Write this in our DMC notation: starting from \(R\) we want to reach \(R'\), so \(R(0) = R\) and \(R(\tau) = R'\), so the approximate drift solutions is

\[
R' = R + F(R)\lambda \tau .
\] (111)

The Langevin equation of Fokker-Planck evolution

Now we know what drift does (correct to $O(\tau^2)$),

$$\langle R' | e^{-\tau \lambda \nabla_R \cdot F(\hat{R})} | R \rangle = \delta(R' - R - \lambda \tau F(R)) ,$$  \hspace{1cm} (112)

A random walk with both diffusion and drift is (again, up to $O(\tau^2)$)

$$R' = R + \sqrt{2\lambda \tau} \eta + \lambda \tau F(R) .$$  \hspace{1cm} (113)

To summarize, we found an approximate solution to the stochastic Langevin equation, which is equivalent to the Fokker-Planck equation for the probability density $f(R, \tau)$. 
Math behind the first order DMC algorithm

To summarize, the importance sampling DMC is solving

\[ f(R, \tau) := \Psi(R, \tau) \varphi_T(R) \to \phi_0(R) \varphi_T(R) \text{ as } \tau \to \infty, \quad (114) \]

and \( f(R) \) is evolved according to

\[ - \frac{\partial f(R, \tau)}{\partial \tau} = -\lambda \nabla^2_R f(R, \tau) + \lambda \nabla_R \cdot [F(R)f(R, \tau)] \\
+ (E_L(R) - E_T)f(R, \tau) := (\mathcal{H}_F - E_T)f(R, \tau). \quad (115) \]

The solution is

\[ f(R, \tau) = \int dR' \langle R | e^{-\tau(\mathcal{H}_F - E_T)} | R' \rangle f(R', 0), \quad (116) \]

but the Green’s function \( \langle R | e^{-\tau(\mathcal{H}_F - E_T)} | R' \rangle \) is known only approximately for small \( \tau \).
Math behind the first order DMC algorithm

The first order DMC algorithm is based on two approximations:

\[
\langle R' | e^{-\tau(\hat{H}_F - E_T)} | R \rangle := \langle R' | e^{-\tau(\hat{H}_F + E_L(R) - E_T)} | R \rangle \\
\approx \langle R' | e^{-\tau\hat{H}_F} | R \rangle \langle R | e^{-\tau(E_L(R) - E_T)} | R \rangle \\
= \langle R' | e^{-\tau(\lambda \nabla^2 R + \lambda \nabla R \cdot F(R))} | R \rangle \langle R | e^{-\tau(E_L(R) - E_T)} | R \rangle,
\]

and diffusion and drift is approximately

\[
R' = R + \sqrt{2\lambda \tau \eta} + \lambda \tau F(R). \tag{117}
\]
A working first order DMC algorithm

1. Initialise many ($\sim 100 - 500$) configurations, where walkers are distributed according to $|\varphi_T(R)|^2$. We don’t want them all destroyed in the branching process, so we need to generate as diverse a set as possible. You can pick from a VMC run that samples $|\varphi_T(R)|^2$ positions $R$ now and then to a zeroth generation of configurations. The total number of configurations in a generation is its population.

2. In each configuration of the present generation, move the $i^{th}$ walker with

$$r_{i}^{\text{new}} = r_i + \lambda \tau F(R) + \eta \sqrt{2\lambda \tau}, \quad (118)$$

where $\eta$ is a 3-dimensional Gaussian random vector.

3. Calculate a weight for the move

$$\text{ratio} = \frac{|\varphi_T(R')|^2 G(R', R; \tau)}{|\varphi_T(R)|^2 G(R, R', \tau)}, \quad (119)$$

and accept the move $R \rightarrow R'$ with probability $\min[1, \text{ratio}]$ (”Metropolis question”)

A working first order DMC algorithm

4. Take each configuration in turn and make $N_{\text{copies}}$ copies of it,

$$N_{\text{copies}} = \text{int} \left[ e^{-\tau(E_L(R) - E_T)} + \xi \right] \quad \xi \in U[0, 1] , \quad (120)$$

5. Estimate the ground-state energy $E_0$ by taking the average of the local energies of current generation.

6. Keep the population healthy by adjusting the trial energy $E_T$. One way is to use a feedback mechanism,

$$E_T = E_0 + \kappa \ln \frac{N_{\text{Ctarget}}}{N_C} , \quad (121)$$

where $E_0$ is the best guess for the ground-state energy we have so far, $N_{\text{Ctarget}}$ is the population we'd like to have, and $N_C$ is the one we have now. The positive parameter $\kappa$ should have the smallest value that does the job.

7. go back to step 2
A working first order DMC algorithm

Comments:

- The accept/reject test is a Metropolis-Hastings test.
  
  - The Hamiltonian with drift is *non-hermitian*; the accept/reject step ensures that the detailed balance condition is satisfied \(\Rightarrow\) makes sure the points \(\mathbf{R}\) are really sampled from the distribution \(f(\mathbf{R}; \tau) = \Psi(\mathbf{R}; \tau)\varphi_T(\mathbf{R})\).

  - The acceptance *must be* > 99.9 \% – if not, you time step is too large. At your own discretion you may even omit the test and assume that the acceptance is very high anyhow, but then you are on your own.

- Controlling the number of configurations (sets of walkers) should be done with kid gloves. The more you force it, the more fluctuations increase elsewhere. No matter how you control the population you will be upsetting randomness, so be gentle.

- Use as large a population as you can afford. Something like 5000 configurations is common. Debugging can use a smaller set.
Importance sampling from a similarity transformation

Physics remains unchanged in the similarity transformation\(^{12}\)

\[
e^{U(R)} \hat{H} e^{-U(R)} = e^{U(R)} \hat{T} e^{-U(R)} + V(R) = \hat{T} + \hat{D} + E_L(R) ,
\] (122)

where the local energy is \(E_L(R) = e^{-U(R)}(\hat{H} e^{U(R)})\) and the drift operator is defined \(\hat{D}\bullet := \nabla \cdot (\nabla U\bullet) = (\nabla^2 U + \nabla U \cdot \nabla)\bullet\). If \(e^{U(R)}\) is an approximate *bosonic* ground state, you’ll get the familiar importance sampling DMC algorithm with diffusion-drift-branching. Even though \(e^{U(R)}\) is bosonic, this applies to fermions as well.

A related approach - never applied to DMC - is to write

\[
g(R)(\hat{H} - E_n)\phi_n(R) = 0 \text{ scaled Schrödinger equation} ,
\] (123)

with a strictly positive \(g(R)\) that has the property \(g(R)V(R) = \text{finite } \forall R\). This way Nakatsuji et al. stabilized an iterative solution method.\(^{13}\) In DMC, this could be written as a variable step \(\tau g(R)\).

---

\(^{12}\)S. Chin, PRE 101, 043304 (2020).

\(^{13}\)See the recent article by H. Nakatsuji and H. Nakashima, J. Chem. Phys. 157, 094109 (2022) and references therein.
Higher order DMC

The order of DMC comes from how we deal with non-commuting parts in the Hamiltonian $\hat{H}$. With importance sampling, we must write $e^{-\tau \hat{H}_F}$ in terms of simpler, albeit approximate, matrix elements:

1. **First operator splitting**: $\hat{H}_F = \hat{T}_F + \hat{E}_L$, $[\hat{T}_F, \hat{E}_L] \neq 0$
   Separate the Fokker-Planck (drift-diffusion) operator $\hat{T}_F$ and the local energy operator $\hat{E}_L$.

2. **Second operator splitting**: $\hat{T}_F = \hat{T} + \lambda \nabla R \cdot F(R)$,
   $[\hat{T}, \lambda \nabla R \cdot F(R)] \neq 0$
   The Fokker-Planck equation describing probability density evolution in diffusion under drift is cast to a Langevin equation, an equation of motion of a single walker under a velocity field.

To get second or higher order DMC we must improve both splittings in unison, and solve the Langevin equation without losing accuracy.
What Hamiltonian did we actually use?

Let’s ask a question: What "fake Hamiltonian" $\hat{H}'$ has the exact splitting

$$e^{-\tau \hat{H}'} = e^{-\tau \hat{T}} e^{-\tau \hat{V}} ?$$ (124)

The Campbell-Baker-Hausdorff (C-B-H) formula $^{15}$ states that

$$e^{\hat{A}} e^{\hat{B}} = e^{\hat{A}+\hat{B}+\frac{1}{2}[\hat{A},\hat{B}]+\frac{1}{12}[\hat{A}-\hat{B},[\hat{A},\hat{B}]]+...},$$ (125)

so (drop terms beyond $\tau^3$)

$$e^{\tau \hat{T}} e^{\tau \hat{V}} = e^{\tau (\hat{H}+\frac{1}{2} \tau [\hat{T},\hat{V}]+\frac{1}{12} \tau^2 [\hat{T}-\hat{V},[\hat{T},\hat{V}]]}) := e^{\tau \hat{H}'} ,$$ (126)

and read the operator (use $[\hat{T},\hat{V}]= [\hat{H} - \hat{V},\hat{V}]=[\hat{H},\hat{V}]$),

$$\hat{H}' = \hat{H} + \frac{1}{2} \tau [\hat{H},\hat{V}] - \frac{1}{12} \tau^2 [\hat{H} - 2\hat{V},[\hat{H},\hat{V}]] .$$ (127)

---


$^{15}$ An old theorem (1890,1902), has to do with Lie algebra of a Lie group.
Energy accuracy without importance sampling

How good is our first order DMC algorithm? After DMC without importance sampling we get the energy

\[
\langle E \rangle := \langle \phi_0 | \hat{H}' | \phi_0 \rangle
\]

\[
= E_0 + \frac{1}{2} \tau \langle \phi_0 | [\hat{H}, \hat{V}] | \phi_0 \rangle - \frac{1}{12} \tau^2 \langle \phi_0 | [\hat{H} - 2\hat{V}, [\hat{H}, \hat{V}]] | \phi_0 \rangle
\]

\[
= E_0 - \frac{1}{12} \tau^2 \langle \phi_0 | [\hat{H} - 2\hat{V}, [\hat{H}, \hat{V}]] | \phi_0 \rangle ,
\]

where the linear term is zero, because \( |\phi_0\rangle \) is an eigenstate of \( \hat{H} \),

\[
\langle \phi_0 | [\hat{H}, \hat{V}] | \phi_0 \rangle = \langle \phi_0 | \hat{H} \hat{V} - \hat{V} \hat{H} | \phi_0 \rangle = 0 .
\]

*Without importance sampling the approximation \( e^{-\tau \hat{H}} \approx e^{-\tau \hat{H}} e^{-\tau \hat{V}} \) gives \( E_0 \) accurate to \( O(\tau^2) \).*
Energy accuracy with importance sampling

Importance sampling makes DMC sample the product $\phi_0(R)\varphi_T(R)$, and we evaluate the mixed estimator

$$\langle E \rangle := \langle \varphi_T | \hat{H}' | \phi_0 \rangle$$

$$= E_0 + \frac{1}{2} \tau \langle \varphi_T | [\hat{H}, \hat{V}] | \phi_0 \rangle - \frac{1}{12} \tau^2 \langle \varphi_T | [\hat{H} - 2\hat{V}, [\hat{H}, \hat{V}]] | \phi_0 \rangle,$$

which converges only linearly, because $\langle \varphi_T | [\hat{H}, \hat{V}] | \phi_0 \rangle \neq 0$.

With importance sampling the approximation $e^{-\tau \hat{H}} \approx e^{-\tau \hat{T}} e^{-\tau \hat{V}}$ gives $E_0$ accurate only to $O(\tau)$.
Energy accuracy with importance sampling

We saw this linear convergence in the demo program.

![Graph showing energy convergence with importance sampling](image-url)
Propagators with second order accuracy

It’s simple to construct a 2nd order splitting, one popular choice is

\[ e^{-\tau(\hat{A}+\hat{B})} \approx e^{-\frac{1}{2} \tau \hat{A}} e^{-\tau \hat{B}} e^{-\frac{1}{2} \tau \hat{A}} \]  

(133)

Next we must also solve the drift trajectories from Eq. (109)

\[ \frac{dR(\tau)}{d\tau} = \lambda F(R(\tau)) \]  

(134)

to second order. The drift changes as particles move, so Eq. (110) could be improved to (2nd order Runge-Kutta)

\[ R(\tau) = R(0) + \lambda \tau F(R(0)) + \frac{1}{2} \lambda \tau F(R(0)) + \mathcal{O}(\tau^3) \]  

(135)
Propagators with second order accuracy

Figure: Second order algorithm results for He atom, no accept/reject step.

There are several alternatives (see Chin) that differ in how many drift evaluations (expensive) are needed and what is the range of quadratic behaviour. I won’t go into details, but push forward to 4th order algorithms.
Propagators with fourth order accuracy

Sheng-Suzuki theorem: there is no 4th order splitting

\[ e^{-\tau(\hat{A}+\hat{B})} = \prod_{i=1}^{N} e^{-\tau a_i \hat{A}} e^{-\tau b_i \hat{B}} \] (136)

without some \( a_i < 0 \) and \( b_i < 0 \). \(^{16}\)

This we cannot have in DMC, because a negative \( t_i \) means \( \langle R' | e^{-\tau t_i \hat{T}} | R \rangle \) is inverse diffusion (backward in time) and it’s impossible to simulate. Something has to be added, namely some commutator(s) of \( \hat{T} \) and \( \hat{V} \). A convenient one to keep is related to the classical force on particle \( i \), \(^{17}\)

\[ [\hat{V}, [\hat{T}, \hat{V}]] = 2\lambda |\nabla_R V(R)|^2 = 2\lambda \sum_{i=1}^{N} |f_i(R)|^2 , \quad f_i = -\nabla_i V. \] (137)

---


\(^{17}\) The operator \([\hat{V}, [\hat{T}, \hat{V}]]\) acts on some \( \phi(R) \).
The commutator is evaluated like this, \( \hat{T} = -\lambda \nabla^2 \), \( V := V(R) \),

\[
[\hat{V}, [\hat{T}, \hat{V}]] \phi(R) = [-V^2 \hat{T} - \hat{T} V^2 + 2 V \hat{T} V] \phi
\]

\[
= \lambda [V^2 \nabla^2 \phi + \nabla^2 (V^2 \phi) - 2 V \nabla^2 (V \phi)]
\]

\[
= \lambda \left[ V^2 \nabla^2 \phi + \nabla \cdot [2 V (\nabla V) \phi + V^2 \nabla \phi] 
- 2 V \nabla \cdot [(\nabla V) \phi + V \nabla \phi] \right]
\]

\[
= \lambda \left[ V^2 \nabla^2 \phi + 2 (\nabla V)^2 \phi + 2 (V \nabla^2 V) \phi + 2 V (\nabla V) \cdot \nabla \phi 
+ 2 V \nabla V \cdot \nabla \phi + V^2 \nabla^2 \phi 
- 2 V \nabla^2 V \phi - 2 V \nabla V \cdot \nabla \phi - 2 V \nabla V \cdot \nabla \phi - 2 V^2 \nabla^2 \phi \right]
\]

coloured terms cancel, only 2nd term survives

\[
= 2 \lambda (\nabla V)^2 \phi .
\]
Propagators with fourth order accuracy

Caleidoscope of 4th order methods (incomplete list):

- **Trotter** \(^{18}\) (divide \(\tau\) to infinitely many pieces)

\[
e^{-\tau \hat{H}} = \lim_{M \to \infty} \left[ e^{-\frac{\tau}{M} \hat{T}} e^{-\frac{\tau}{M} \hat{V}} \right]^M,
\]

- **Ruth-Forest** (Ruth (1983, private channels), Forest and Ruth (1990), re-discovered several times around 1990 by Campostrini and Rossi and by Candy and Rozmus) \(^{19}\)

- **Takahashi-Imada** (1984)\(^{20}\) and Li-Broughton (1987)\(^{21}\)

- **Chin** (1997)\(^{22}\)


\(^{19}\)E. Forest and R. D. Ruth, Physica D. 43, 105 (1990)


Propagators with fourth order accuracy

Forest-Ruth

\[ \hat{G}_{\text{FR}}(\tau) = e^{-\nu_3 \tau \hat{V}} e^{-t_3 \hat{T}} e^{-\nu_2 \tau \hat{V}} e^{-t_2 \tau \hat{T}} e^{-\nu_1 \tau \hat{V}} e^{-t_1 \tau \hat{T}} e^{-\nu_0 \tau \hat{V}}, \quad (142) \]

with \( s = 2^{1/3} = 1.25992, \)

\[ \nu_0 = \nu_3 = \frac{1}{2} \frac{1}{2 - s} \quad (> 0), \quad \nu_1 = \nu_2 = -\frac{1}{2} \frac{s - 1}{2 - s} \quad (> 0) \quad (143) \]

\[ t_1 = t_3 = \frac{1}{2 - s} \quad (> 0), \quad t_2 = -\frac{s}{2 - s} \quad (< 0!) \quad (144) \]

From right to left, the potential will be evaluated at times 0, \( t_1 \tau, (t_1 + t_2) \tau \) (\(< 0!\)), and \( \tau \). Suzuki was right, there are negative coefficients. These make the application quite tricky. The good thing about FR is that it only needs 6 FFT’s to propagate time \( \tau \),
Propagators with fourth order accuracy

Takahashi-Imada  "not fully fourth order"

\[ \hat{G}_{TI}(\tau) = e^{-\tau \hat{T}} e^{-\tau \hat{V}} - \frac{1}{24} \tau^3 [\hat{V}, [\hat{T}, \hat{V}]] e^{-\tau \hat{T}} \] (145)

Chin  Several possible forms, one is

\[ \hat{G}_{Chin}(\tau) = e^{-v_0 \tau \hat{V}} e^{-t_1 \tau \hat{T}} e^{-v_1 \tau \hat{V}} e^{-t_2 \tau \hat{T}} e^{-v_1 \tau \hat{V}} e^{-t_1 \tau \hat{T}} e^{-v_0 \tau \hat{V}} \] (146)

\[ \hat{W} := \hat{V} + \frac{u_0}{v_1} \tau^2 [\hat{V}, [\hat{T}, \hat{V}]] , \] (147)

\[ v_0 = \frac{6t_1(t_1 - 1) + 1}{12(t_1 - 1)t_1} , \quad t_2 = 1 - 2t_1 \] (148)

\[ v_1 = \frac{1}{2} - v_0 \quad u_0 = \frac{1}{48} \left( \frac{1}{6t_1(1 - t_1)^2} - 1 \right) , \] (149)

and \( t_1 \) is tunable, \( 0 \leq t_1 \leq 1 - 1/\sqrt{3} \). Even 5th order errors may cancel with \( t_1 = 0.35023 \). \textit{All coefficients are positive.}
What can a good propagator do?

We have a good approximation for $\hat{G}(\tau) = e^{-\tau \hat{H}}$, such as

$$\hat{G}_{\text{Chin}}(\tau) = \hat{G}(\tau) + \mathcal{O}(\tau^5).$$  \hfill (150)

Nothing prevents us from overextending the validity, as long as we know that we are improving. Take a trial wave function $\varphi_T(R)$ and propagate it,

$$\langle R | \varphi_T(\tau_0) \rangle = \varphi_T(R; \tau_0) = \int dR' \langle R | G_{\text{Chin}}(\tau_0) | R' \rangle \varphi_T(R').$$  \hfill (151)

We can use VMC to see how far in $\tau_0$ we can go until the energy begins to increase: find $\tau_0$ were

$$E(\tau_0) = \frac{\langle \varphi_T(\tau_0) | \hat{H} | \varphi_T(\tau_0) \rangle}{\langle \varphi_T(\tau_0) | \varphi_T(\tau_0) \rangle}$$  \hfill (152)

is at minimum (approximately).\footnote{In VMC the normalization is irrelevant.}
What can a good propagator do?

The propagation can now be viewed as a multi-dimensional integral,

$$
\varphi_T(R; \tau_0) = \int dR' dR_1 dR_2 e^{-\nu_0 \tau_0 V(R)} \langle R | e^{-t_1 \tau_0 \hat{T}} | R_2 \rangle e^{-\nu_1 \tau_0 W(R_2)} \langle R_2 | e^{-t_2 \tau_0 \hat{T}} | R_1 \rangle e^{-\nu_1 \tau_0 W(R_1)} \langle R_1 | e^{-t_1 \tau_0 \hat{T}} | R' \rangle e^{-\nu_0 \tau_0 V(R')} \varphi_T(R').
$$

Notice that this applied one $G_{\text{Chin}}$ for a single, long time step $\tau_0$. 24

---

**Figure 1.** Variational energies obtained with a single time step $\tau$ and the proposed wave function. Open circles, squares, and diamonds correspond to $\Psi_m(R)$ of McMillan type, semiclassical model, and 1, respectively. The corresponding filled symbols stand for the asymptotic values using the three models and more than one time step.

---

24 Rota et al. Physical Review E 81, 016707 (2010); The integrated coordinates can also be thought of as "shadow coordinates", as in the variational shadow wave function by L. Reatto, F. Pederiva et al.
What can a good propagator do?

Large $\tau$ means fast projection and less correlated points in DMC:

Liquid He$^4$: Timesteps 0.01 (dots), 0.05 (small points) and 0.2 (large points). Getting to the ground state takes 10 times more updates with a 2nd order propagator than with a 4th order one.

---

Fermions: Spin in expectation values

Does spin $\sigma = \{\uparrow, \downarrow\}$ make computation of expectation values harder? Not really. Take a spin-independent operator $\hat{A} = \hat{A}(\mathbf{R})$. The expectation value in the many-body state with spins $\sigma = (\sigma_1, \ldots, \sigma_N)$

$$\Psi(\mathbf{X}) = \Psi(r_1\sigma_1, r_2\sigma_2, \ldots, r_N\sigma_N)$$ (154)

looks a bit daunting,

$$\langle A \rangle = \frac{\sum_\sigma \int d\mathbf{R} \Psi^*(\mathbf{X}) \hat{A}(\mathbf{R}) \Psi(\mathbf{X})}{\sum_\sigma \int d\mathbf{R} \Psi^*(\mathbf{X}) \Psi(\mathbf{X})}.$$ (155)

The sum $\sum_\sigma$ is constrained, we usually know the spin counts $N_\uparrow$ and $N_\downarrow$, with $N_\uparrow + N_\downarrow = N$. The trick is to re-group particles according to spin,

$$\Psi(\mathbf{X}) = \Psi(r_1 \uparrow, r_2 \uparrow, \ldots, r_{N_\uparrow} \uparrow, r_{N_\uparrow+1} \downarrow, \ldots r_{N_\uparrow+N_\downarrow} \sigma_{N_\uparrow+N_\downarrow})$$, (156)

and the spin summation $\sum_\sigma$ is done!
Fermions: Spin in expectation values

Let's take a look at one term in the spin-summation, assuming there are 5 particles, $N^\uparrow = 3, N^\downarrow = 2$. That term happens to have the wave function

$$\Psi(r_1^\uparrow, r_2^\downarrow, r_3^\uparrow, r_4^\uparrow, r_5^\uparrow).$$

(157)

Shifting all $^\uparrow$ particles to the left can be done using particle exchanges. Each exchange changes the sign of the wave function, so we get$^{26}$

$$\Psi(r_1^\uparrow, r_2^\downarrow, r_3^\uparrow, r_4^\uparrow, r_5^\uparrow) = -\Psi(r_1^\uparrow, r_3^\uparrow, r_4^\uparrow, r_5^\uparrow, r_2^\downarrow).$$

(158)

We do the same in both $\Psi^*$ and $\Psi$, so $\Psi^*\Psi$ has no sign change. Furthermore, coordinates are integrated over, so we may always re-label particle coordinates as 1,2,3,4,5 from left to right. To summarize, we can always use the 5-particle state

$$\Psi(R) := \Psi(r_1^\uparrow, r_2^\uparrow, r_3^\uparrow, r_4^\downarrow, r_5^\downarrow)$$

(159)

and evaluate

$$\langle A \rangle = \frac{\int dR\Psi^*(R)\hat{A}(R)\Psi(R)}{\int dR\Psi^*(R)\Psi(R)}.$$ 

(160)

$^{26}$Need 3 exchanges, hence minus sign.
Fermions: Spin in expectation values

Conclusion:

\[
\langle A \rangle = \frac{\int dR \psi^*(R) \hat{A}(R) \psi(R)}{\int dR \psi^*(R) \psi(R)}
\]

\[
\psi(R) = \psi(r_1 \uparrow, r_2 \uparrow, \ldots, r_{N\uparrow} \uparrow, r_{N\uparrow+1} \downarrow, \ldots, r_{N\uparrow+N\downarrow} \sigma_{N\uparrow+N\downarrow} \downarrow) .
\] (161)

There’s no need to keep track of spins, just to remember that particles 1, ..., \( N\uparrow \) have spins up, the rest have spins down.
Fixed-node DMC

Fixed-node DMC (FN-DMC): use antisymmetric trial wave function,

\[ \varphi_T(...r_i,...,r_j...) = -\varphi_T(...r_j,...,r_i...) \]  \hspace{1cm} (162)

A common way to achieve this is to use

\[ \varphi_T(R) = e^{U(R)} \sum_k d_k D_k[\phi](R) \]  \hspace{1cm} (163)

- Jastrow factor $e^U$
  - particle-particle correlations
  - takes care of particle-particle potentials, especially singularities
- Slater determinants $D_k$
  - antisymmetrized single-particle orbitals \{\phi\}
  - takes care of Pauli principle
  - Functional of orbitals, evaluated at $R$: $D_k[\phi](R)$
Boson ground state

The tiling theorem on the next slide relies on the properties of bosons ground state. If a real and symmetric and normalized function $\Psi(R)$ minimizes the expression

$$\int dR \left[ \left( \frac{\hbar^2}{2m} \nabla_R \Psi(R) \right)^2 + V(R)\Psi(R)^2 \right],$$

(164)

then it’s a solution to the Schrödinger equation for $\hat{H} = -\frac{\hbar^2}{2m} \nabla^2_R + V(R)$. But this applies also to $|\Psi(R)|$, so $|\Psi(R)|$ must have a continuous first derivative wherever $V(R)$ is finite, therefore $\Psi(R)$ cannot change sign. If two different non-negative functions $\Psi_1$ and $\Psi_2$ satisfy the conditions, then also $\Psi_1 - \Psi_2$ is a solution to the Schrödinger equation (it’s a linear equation), so $\Psi_1 - \Psi_2$ cannot change sign, but then both $\Psi_1$ and $\Psi_2$ cannot be normalized, contradicting the assumptions. Therefore, the boson ground state $\Psi(R)$ is a unique, non-negative function. 

---

27 See “why not do full optimization” a few slides ago.

Tiling theorem

In FN-DMC one picks a nodal pocket and inserts fermions. The trial wave function creates a drift that prevent crossing of nodal surfaces. Exact nodes would make FN-DMC an exact ground state method. But why can we ignore the other nodal pockets?

Tiling theorem:  

*Pick any reference point. All points that can be reached from this point following a continuous path without crossing a nodal surface is a nodal pocket. All nodal pockets are equivalent and related to each other by particle permutations.*

The reason is the same as for bosons: *if you add a node, the energy increases.* Find a fermion ground state $\phi(R)$. Pick any nodal pocket and all pockets that are copies of it with particles relabeled, and color them blue. Now assume there’s a nodal pocket in $\phi(R)$ that’s not blue, find it’s permutation copies and color them red. Somewhere there’s a blue and red pocket touching, with a nodal surface in between. If you eliminate the nodal surface and attach the pockets, you can again antisymmetrize the wave function and get a new, normalizable state $\phi'(R)$. But the energy was lowered by removing the nodal surface, so the fermion state $\phi'(R)$ has a lower energy than $\phi(R)$, and consequently $\phi(R)$ couldn’t have been the fermion ground state, a contradiction.

---

Jastrow factor

Often the potential doesn’t depend on spin, so we choose a convenient way to cancel singular potentials. In electron systems, we have singular e-n and e-e potentials, so we need at least two correlation functions $u_{en}$ and $u_{ee}$. With indices $i,j$ for electrons, and $\alpha$ for nuclei,

$$e^{U(R)} = e^{\sum_{i,\alpha} u_{en}(r_{i\alpha}) + \sum_{i<j} u_{ee}(r_{ij})} \quad \text{Jastrow factor} . \quad (165)$$

This can be supplemented with an e-e-n factor,

$$e^{\sum_{i<j,\alpha} u_{een}(r_{i\alpha},r_{j\alpha},r_{ij})} , \quad (166)$$

which should be enough, because e-e-e-n is rare due to the Pauli principle.

DMC evolution optimizes just these correlations to achieve lowest energy
Slater determinants

For spin-1/2 fermions, one often uses two Slater determinants, 30

\[
\varphi_T(R) = e^{U(R)} D^\uparrow(r_1...r_{N\uparrow}) D^\downarrow(r_{N\uparrow+1}...r_N). \tag{167}
\]

Here the Slater determinant for \( N \) fermions with spins up is

\[
D^\uparrow(r_1...r_{N\uparrow}) = \frac{1}{\sqrt{N_{\uparrow}!}} \begin{vmatrix}
\phi_1(r_1) & \phi_2(r_1) & \ldots & \phi_{N\uparrow}(r_1) \\
\phi_1(r_2) & \phi_2(r_2) & \ldots & \phi_{N\uparrow}(r_2) \\
\vdots & \vdots & \ddots & \vdots \\
\phi_1(r_{N\uparrow}) & \phi_2(r_{N\uparrow}) & \ldots & \phi_{N}(r_{N\uparrow}) 
\end{vmatrix}. \tag{168}
\]

The single-particle orbitals \( \{\phi\} \) can be taken from

- Hartree-Fock (HF) ← more popular
- Density functional theory (DFT)

HF has no correlations. This may be a good thing, because DMC has the liberty to optimize correlations in the Jastrow factor. DFT tries to add correlations to orbitals, which may prevent DMC to do it’s best in optimizing correlations. 30

30While not antisymmetric under exchange of spin-up and spin-down fermions, the expectation values of spin-independent operators are valid. See W. M. C. Foulkes, L. Mitas, R. J. Needs and G. Rajagopal, Rev. Mod. Phys. 73 , 33 (2001).
Slater-type orbitals (STO’s)

The first lesson is that hydrogen-type orbitals are not very useful. The complete set contains discrete states plus the continuum, hence $\infty$ number of states. Too many of them are needed to represent core electrons. STO’s have the benefit of having only angular node, $R_{nl}$ is nodeless:

$$\phi_{nlm}(r) = R_{nl}(r)Y_{lm}(\theta, \phi)$$  \hspace{1cm} (169)$$

$$R_{nl}(r) = \frac{(2\zeta)^{3/2}}{\Gamma(2n+1)}(2\zeta r)^{n-1}e^{-\zeta r},$$  \hspace{1cm} (170)$$

with a nuclear-charge-related constant $\zeta$. The notation $nl = 1s, 2s, 2p, 3d, ...$ is re-used. For more flexibility, the exponent can be chosen to depend on $nl, \zeta_{nl}$.

Basis sets often carry the name “double zeta” or “triple zeta”: zeta is the STO exponent. STO’s are computationally expensive, but can be approximated with faster GTO’s (next slide).
Gaussian-type orbitals (GTO’s)

The spherical harmonic GTO’s are

\[ \phi_{nlm}(r) = R_{nl}(r) Y_{lm}(\theta, \phi) \]  \hspace{1cm} (171)

\[ R_{nl}(r) = \frac{2(2\alpha)^{3/4}}{\pi^{1/3}} \sqrt{\frac{2^{2n-l-2}}{(4n - 2l - 3)!!}} \left(\sqrt{2\alpha}r\right)^{2n-l-2} e^{-\alpha r^2}. \] \hspace{1cm} (172)

Here \( n!! \) is a special double factorial,

\[ n!! = \begin{cases} 
1 & n = 0 \\
(n-2)(n-4)\ldots2 & n > 0, \text{ even} \\
(n-2)(n-4)\ldots1 & n > 0, \text{ odd} \\
\frac{1}{(n+2)(n+4)\ldots1} & n < 0 
\end{cases} \] \hspace{1cm} (173)

Typically, a lot more GTO’s than STO’s are needed, but GTO integrals are very fast to evaluate. So, GTO’s have computational, but they have zero cusp (bad) and they die out faster in \( r \) than STO’s (too fast, bad).
Cartesian forms

In cartesian coordinates Slater-type orbitals are

$$\phi_{ijk}^{STO}(x, y, z) = x^i y^j z^k e^{-\zeta r},$$  \hspace{1cm} (174)

and one often uses them as sums of Gaussian primitives,

$$\chi_n(x, y, z) = N_{ijk}^n x^i y^j z^k e^{-\alpha_n r^2},$$  \hspace{1cm} (175)

where the total angular momentum is \( l = i + j + k \), resulting in one s-function \((l = 0)\), three p-functions \((l = 1)\), six d-functions \((l = 2)\) etc.. Simple ones are STO-nG, n Gaussians fitted to make an STO. Improvements are double-zeta (the \( \zeta \)'s in the exponent), two basis functions per atomic orbital, triple-zeta etc.

Example 3: The ground state of Carbon is \( 1s^2 2s^2 2p^2 \), and the orbitals \( 1s, 2s, 2p_x, 2p_y, 2p_z \) need minimal basis of 5 basis functions, or 10 double-zeta basis functions.
Slater-Condon rules

Often the antisymmetric part of the wave function is written as

\[ \Psi = \sum_i x_i D_i, \quad (176) \]

a sum of Slater determinants \( D_i \). Notation: \(|D_{pq}^{mn}\rangle\) means from \(|D\rangle\), change orbital \( m \) to \( p \) and \( n \) to \( q \). Use collective coordinates \( \mathbf{x}_i := (\mathbf{r}_i, \sigma_i) \) and antisymmetrization \( \mathcal{A} \).

\[ |D\rangle := \mathcal{A}[\phi_1(x_1)\ldots\phi_m(x_m)\phi_n(x_n)\ldots\phi_N(x_N)] \quad (177) \]
\[ |D_{m}^{p}\rangle := \mathcal{A}[\phi_1(x_1)\ldots\phi_p(x_m)\phi_n(x_n)\ldots\phi_N(x_N)] \quad (178) \]
\[ |D_{mn}^{pq}\rangle := \mathcal{A}[\phi_1(x_1)\ldots\phi_p(x_m)\phi_q(x_n)\ldots\phi_N(x_N)]. \quad (179) \]

The Slater-Condon rules apply for matrix elements of a one-body operators \( \hat{O}_1 \) and a two-body operators \( \hat{O}_2 \) between Slater determinants.
Slater-Condon rules

The diagonal elements are

$$\langle D | \hat{O}_1 | D \rangle = \sum_{i \in D} \langle \phi_i | \hat{O}_1 | \phi_i \rangle$$

(180)

$$\langle D | \hat{O}_2 | D \rangle = \frac{1}{2} \sum_{(i,j) \in D} \left[ \langle \phi_i \phi_j | \hat{O}_2 | \phi_i \phi_j \rangle - \langle \phi_i \phi_j | \hat{O}_2 | \phi_j \phi_i \rangle \right] .$$

(181)

This is already enough for Hartree-Fock. The electronic Hamiltonian is

$$\hat{H}_{el} = \sum_{i \in \text{electrons}} \hat{h}_i + \sum_{(i<j) \in \text{electrons}} \frac{1}{r_{ij}} \left| 1\text{-body operator} \right|$$

$$+ \sum_{l \in \text{ions}} Z_l \left| 2\text{-body operator} \right| .$$

(182)

(183)
Slater-Condon rules

The Hartree-Fock energy is (use real orbitals)

\[ E_{el} = \langle D | \hat{H}_{el} | D \rangle = \sum_{i \in \text{electrons}} \langle \phi_i | \hat{h}_i | \phi_i \rangle \]  
\[ + \sum_{(i < j) \in \text{electrons}} \left[ \langle \phi_i \phi_j | \frac{1}{r_{ij}} | \phi_i \phi_j \rangle - \langle \phi_i \phi_j | \frac{1}{r_{ij}} | \phi_j \phi_i \rangle \right] \]  
\[ = \sum_{i \in \text{electrons}} \int d\mathbf{x}_1 \phi_i(\mathbf{x}_1)^2 h_i(r_1) \]  
\[ + \sum_{(i < j) \in \text{electrons}} \int d\mathbf{x}_1 d\mathbf{x}_2 \phi_i(\mathbf{x}_1)^2 \frac{1}{r_{12}} \phi_j(\mathbf{x}_2)^2 \]  
\[ - \sum_{(i < j) \in \text{electrons}} \int d\mathbf{x}_1 d\mathbf{x}_2 \phi_i(\mathbf{x}_1) \phi_j(\mathbf{x}_1) \frac{1}{r_{12}} \phi_j(\mathbf{x}_2) \phi_i(\mathbf{x}_2) . \]

Minimize \( E_{el} \) wrt. \( \phi \)'s and you get the Hartree-Fock equations.
Slater-Condon rules

In a multi-determinant wave function we need a few more rules. The elements between one-off orbitals are

\[
\langle D|\hat{O}_1|D_i\rangle = \langle \phi_i|\hat{O}_1|\phi_j\rangle
\]  
(189)

\[
\langle D|\hat{O}_2|D_i\rangle = \sum_{k \in D} \left[ \langle \phi_i\phi_k|\hat{O}_2|\phi_j\phi_k\rangle - \langle \phi_i\phi_k|\hat{O}_2|\phi_k\phi_j\rangle \right],
\]  
(190)

and the elements between two-off orbitals are even simpler,

\[
\langle D|\hat{O}_1|D_{ik}\rangle = 0
\]  
(191)

\[
\langle D|\hat{O}_2|D_{ik}\rangle = \langle \phi_i\phi_k|\hat{O}_2|\phi_j\phi_l\rangle - \langle \phi_i\phi_k|\hat{O}_2|\phi_l\phi_j\rangle.
\]  
(192)
State-of-the-art multi-determinants

Sergio D. Pineda Flores and Eric Neuscamman J. Phys. Chem. A 2019, 123, 1487-1497 summarize the present situation:

2.2. Determinant Expansion. The MSJ wave function can be written as

$$\Psi(\vec{r}) = \Phi(\vec{r}) \psi_j(\vec{r})$$

$$\Phi(\vec{r}) = \sum_{i=0}^{N} c_i D_i(\vec{r})$$

in which $\psi_j$ is the Jastrow factor (to be discussed in the next section) and $\Phi$ is the linear combination of Slater determinants $D_i$. Thanks to Clark's introduction of the table method$^{28,39}$ and its recent enhancement by Filippi and co-workers,$^{29,30}$ it is now possible to evaluate all the MSJ-related quantities that VMC needs at a per-sample cost that scales as the number of particles cubed plus (not times) the product of the number of determinants and the cube of the maximum excitation rank. This is true even for the derivatives of the wave function $\Psi$ and local energy $E_L$ required for wave function optimization, which has recently permitted VMC optimizations with tens of thousands of determinants.$^{30}$

Links to cited articles:


Hartree-Fock + DMC

HF equations are (simplified notation)

\[ \left[ -\frac{1}{2} \nabla^2 + V(r) + V_H(r) \right] \phi_i(r) - \int dr' \frac{\gamma(r, r')}{|r - r'|} \phi_i(r) = \epsilon_i \phi_i(r) , \quad (193) \]

where

\[ \gamma(r, r') := \sum_j \phi_j^*(r) \phi_j(r') \quad (194) \]

\[ V_H(r) = \int dr' \frac{\sum_j \phi_j^*(r') \phi_j(r')}{|r - r'|} \quad \text{Hartree potential} \quad (195) \]

Solution: Pick your favorite (constraint: money) from a long list of software.\(^{31}\) Mike Towler \(^{32}\) pinpoints the problem in feeding single-particle orbitals to a QMC code:

"Often people find that their HF/DFT code of choice is not supported, so they give up before they start."
Hartree-Fock + DMC

HF energy + correlation energy = exact energy (definition)

This much of the correlation energy is typically covered:

Hartree-Fock 0 %
optimized Slater-Jastrow \( \sim 85 \% \)
FN-DMC \( \sim 95 \% \)
Exact 100 %

If HF is OK \( \Rightarrow \) a single-determinant Slater-Jastrow is OK

<table>
<thead>
<tr>
<th>atom</th>
<th>ground state</th>
<th>HF</th>
<th>Exp.</th>
<th>correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td>He</td>
<td>1s^2</td>
<td>-2.861679993</td>
<td>-2.90338683</td>
<td>-0.041706</td>
</tr>
<tr>
<td>Li</td>
<td>1s^22s^1</td>
<td>-7.432726924</td>
<td>-7.478060323</td>
<td>-0.045333</td>
</tr>
<tr>
<td>Be</td>
<td>1s^22s^2</td>
<td>-14.57302313</td>
<td>-14.6673564</td>
<td>-0.094333</td>
</tr>
<tr>
<td>B</td>
<td>1s^22s^22p^1</td>
<td>-24.52906069</td>
<td>-24.541246</td>
<td>-0.012185</td>
</tr>
</tbody>
</table>
Easy start with basis sets

The Basis Set Exchange (www.basissetexchange.org or link to github):\textsuperscript{33}

```
# Basis Set Exchange
# Version v0.9
# https://www.basissetexchange.org
#
# Basis set: STO-3G
# Description: STO-3G Minimal Basis (3 functions/AO)
# Role: orbital
# Version: 1 (Data from Gaussian09)
#
$basis

*li STO-3G
*

3 s
0.1611957475D+02 0.1543289673D+00
0.2936206066D+01 0.5355281423D+00
0.7946504870D+00 0.4446345422D+00

3 s
0.6362897460D+00 -0.9996722019D-01
0.1478600533D+00 0.3995128261D+00
0.4808867840D-01 0.7001154689D+00

3 p
0.6362897469D+00 0.1559162750D+00
0.1478600533D+00 0.06076837186D+00
0.4808867840D-01 0.3919573931D+00

*
$end
```

\textsuperscript{33}Turbomole format.
Easy start with basis sets

The STO-3G basis set gives *three primitive Gaussians* (hence 3G) fit to a Slater-type orbital (hence STO); the s-functions are (always atomic units)

\[
\begin{align*}
\chi_1(r) &= 0.15432897 \times e^{-16.1195750 r^2} \\
\chi_2(r) &= 0.53532814 \times e^{-2.9362007 r^2} \\
\chi_3(r) &= 0.44463454 \times e^{-0.7946505 r^2}.
\end{align*}
\]  

(196)

From the primitives we get *contracted Gaussians*,

\[
G_n(r) = \sum_{\alpha=1}^{N_n} \chi_\alpha(r),
\]

(197)

and from contracted Gaussians the orbitals (need normalization),

\[
\phi_i(r) = \sum_{n=1}^{N_{basis}} c_{in} G_n(r).
\]

(198)

Obviously STO-3G lacks flexibility.
DMC and chemical accuracy

"Chemical accuracy" 1 kcal/mol

1 Hartree = 627.509 kcal/mol
1 kcal/mol = 0.00159360 Hartree = 43.36 meV

Where are we now: 34

Path integral Monte Carlo (PIMC)

Let’s turn our attention to the statistical physics of equilibrium quantum systems. In canonical ensemble\(^{35}\) everything can be computed from the canonical partition function

\[
Z(\beta) = \sum_i e^{-\beta E_i} = \sum_i \langle i | e^{-\beta \hat{H}} | i \rangle
\]

(199)

where \(\hat{H} | i \rangle = E_i | i \rangle\) and \(\beta = 1/(k_B T)\), \(k_B\) is the Boltzmann constant, and \(T\) is the temperature. \(Z\) is the trace of the density operator

\[
\rho := e^{-\beta \hat{H}}, \quad \rho(R', R; \beta) := \langle R' | e^{-\beta \hat{H}} | R \rangle,
\]

(200)

where the latter is the density matrix. Trace is basis-independent: In the position basis,

\[
Z(\beta) = \int dR \sum_i \langle i | e^{-\beta \hat{H}} | R \rangle \langle R | i \rangle = \int dR \sum_i \langle R | i \rangle \langle i | e^{-\beta \hat{H}} | R \rangle
\]

\[
= \int dR \langle R | e^{-\beta \hat{H}} | R \rangle = \int dR \rho(R, R; \beta).
\]

(201)

\(^{35}\)Energy and particle number are conserved. Grand canonical ensemble is also possible in PIMC.
Path integral Monte Carlo (PIMC)

In most cases we don’t know \( \langle R | e^{-\beta \hat{H}} | R \rangle \) except for small \( \beta \).\(^{36}\) Split the operator (\( \hat{H} \) obviously commutes with \( \hat{H} \), so this splitting is exact),

\[
Z(\beta) = \int dR \langle R | e^{-\beta \hat{H}} | R \rangle = \int dR \langle R | e^{-\frac{\beta}{2} \hat{H}} e^{-\frac{\beta}{2} \hat{H}} | R \rangle .
\]

(202)

Repeat and insert position states,

\[
Z(\beta) = \int dR dR_1 \ldots dR_{M-1} \langle R | e^{-\frac{\beta}{M} \hat{H}} | R_1 \rangle \langle R_1 | e^{-\frac{\beta}{M} \hat{H}} | R_2 \rangle \ldots \langle R_{M-1} | e^{-\frac{\beta}{M} \hat{H}} | R \rangle .
\]

(203)

This is a multidimensional integral of a product of high-temperature density matrices,

\[
\rho(R', R; \frac{\beta}{M}) = \langle R' | \hat{\rho} \left( \frac{\beta}{M} \right) | R \rangle = \langle R' | e^{-\frac{\beta}{M} \hat{H}} | R \rangle .
\]

(204)

\(^{36}\)Small \( \beta \) (high \( T \)) in PIMC is equivalent to the small timestep \( \tau \) in DMC.
Primitive approximation

The high-temperature density matrix can be approximated using ideas described in the DMC section. The splitting $e^{-\beta \hat{H}} \approx e^{-\beta \hat{T}} e^{-\beta \hat{V}}$ gives the \textit{primitive approximation} for the high-$T$ density matrix,

$$
\langle R' | e^{-\beta \hat{H}} | R \rangle \approx (4\pi \lambda \beta)^{-3N/2} e^{-\frac{(R-R')^2}{4\lambda \beta}} e^{-\beta V(R)}. \tag{205}
$$

In PIMC it’s customary to write the Hamiltonian as

$$
\hat{H} = -\lambda \nabla_R^2 + V(R) \tag{206}
$$

$$
\lambda := \frac{\hbar^2}{2m} = \frac{1}{2} \text{ in a.u.}. \tag{207}
$$

With $M$ time slices it’s nice to define\(^{37}\)

$$
\tau := \frac{\beta}{M} \tag{208}
$$

\(^{37}\)Inequal time slices require minor changes.
Quantum-classical analogy

Define a *link action*

\[
S_m = S(R_{m-1}, R_m; \tau) = -\ln \rho(R_{m-1}, R_m; \tau)
\]

\[
\rho(R_0, R_M; \beta) = \int dR_1 \ldots dR_{M-1} e^{-\sum_{m=1}^{M} S(R_{m-1}, R_m; \tau)}.
\]

Customarily, one defines

\[
S_m := \frac{3N}{2} \ln(4\pi \lambda \tau) + \frac{(R_{m-1} - R_m)^2}{4\lambda \tau} + \frac{K(R_{m-1}, R_m; \tau)}{\beta}
\]

where the inter-action \( U \) depends on the approximation. For example, *the primitive action* is \(^{38}\)

\[
S_m^{(\text{prim.})} = \frac{3N}{2} \ln(4\pi \lambda \tau) + \frac{(R_{m-1} - R_m)^2}{4\lambda \tau} + \frac{\tau V(R_m)}{\beta}
\]
Ring polymers

In polymer analogy the spring term holds the polymer together and the potential term gives the inter-polymer interaction. Example: 5 particles and 4 time slices (not in scale)

Note: The springs in the figure represent the kinetic, spring action 
\[
\frac{(R_{m-1} - R_m)^2}{4\lambda\tau},
\] not the potential!
Ring polymers

Another view of 5 particles in 4 time slices:

Only particles on the same time slice interact:
Blues interact and reds interact, but blues don’t interact with reds.
Energy estimators

Having no trial wave function removes the human bias, but also takes away one obvious energy measure, the trial energy. There are many energy estimators, with the same average, but with different variance:

- The thermodynamic energy estimator

\[
\langle E \rangle_{Th} = -\frac{1}{Z} \frac{\partial Z}{\partial \beta}
\]  

- The direct or Hamiltonian energy estimator

\[
\langle E \rangle_{dir} = \frac{\text{Tr}[\hat{H}e^{-\beta \hat{H}}]}{\text{Tr}[e^{-\beta \hat{H}}]}
\]  

- The virial estimator

\[
\langle E \rangle_{\text{vir}} \quad \text{no simple form, approximation specific}
\]

Here the \( Th, dir, vir \) notation is just to distinguish the three; the averages are the same, only data is collected differently, and they have different variances. You can invent your own, clever estimators.
Energy estimators

Start from the exact action in 3D, Eq. (210),

$$\rho(\mathbf{R}_0, \mathbf{R}_M; \beta) = \int d\mathbf{R}_1 \ldots \mathbf{R}_{M-1} \ e^{-\sum_{m=1}^{M} S(\mathbf{R}_{m-1}, \mathbf{R}_m; \tau)}$$

(216)

$$S(\mathbf{R}_{m-1}, \mathbf{R}_m; \tau) := \frac{3N}{2} \ln(4\pi \lambda \tau) + \frac{(\mathbf{R}_{m-1} - \mathbf{R}_m)^2}{4\lambda \tau} + U_m$$

(217)

$$Z = \int d\mathbf{R}_1 \ldots \mathbf{R}_M \ e^{-\sum_{m=1}^{M} S(\mathbf{R}_{m-1}, \mathbf{R}_m; \tau)} \quad \text{with} \quad \mathbf{R}_0 := \mathbf{R}_M .$$

(218)

At high $T$ the values of $U_m$ average to $\tau \langle V \rangle$. At low $T$ this "inter-action" contains some kinetic terms as well, so it’s really just some unknown term in need of approximation.
Thermal energy estimator

Remember that \( \tau = \beta / M \), \( S_m = S(R_{m-1}, R_m; \tau) \) and use thermodynamics:

\[
\langle E \rangle_{Th} = -\frac{1}{Z} \frac{\partial}{\partial \beta} \int dR_1 \ldots dR_M e^{-\sum_{m=1}^{M} S_m} = \frac{1}{Z} \sum_{m=1}^{M} \int dR_1 \ldots dR_M e^{-\sum_{m=1}^{M} S_m} \frac{\partial S_m}{\partial \beta}
\]

\[
= \frac{1}{Z} \sum_{m=1}^{M} \int dR_{m-1} dR_m \frac{\partial S_m}{\partial \beta} \left( \int dR_1 \ldots dR_{m-2} R_{m+1} R_M e^{-\sum_{m=1}^{M} S_m} \right) \rho(R_{m-1}, R_m; \tau)
\]

\[
= \sum_{m=1}^{M} \int dR_{m-1} dR_m \frac{\partial S_m}{\partial \beta} \frac{1}{Z} \rho(R_{m-1}, R_m; \tau) = \sum_{m=1}^{M} \frac{\partial S_m}{\partial \beta} = \sum_{m=1}^{M} \frac{1}{M} \frac{\partial S_m}{\partial \tau},
\]

which gives

\[
\langle E \rangle_{Th} = \frac{1}{M} \sum_{m=1}^{M} \left\langle \frac{3N}{2\tau} - \frac{(R_{m-1} - R_m)^2}{4\lambda \tau^2} + \frac{\partial U_m}{\partial \tau} \right\rangle.
\]
Thermal energy estimator

The thermal energy estimator has one serious shortcoming, related to the two kinetic terms. The first term,

\[
\frac{1}{M} \sum_{m=1}^{M} \left\langle \frac{3N}{2\tau} \right\rangle = \frac{3MN}{2\beta} = M \times \text{classical kinetic energy ,} \tag{223}
\]

the kinetic energy you get if you start the simulation with \(M\) identical time slices, and only running PIMC for a while the second term

\[
\frac{1}{M} \sum_{m=1}^{M} \left\langle -\frac{(R_{m-1} - R_m)^2}{4\lambda\tau^2} \right\rangle \tag{224}
\]

cancels the excess kinetic energy. The thermal energy estimator is poor even for non-interacting particles. Furthermore, at large \(M\) (small \(\tau\)) the two partly cancelling kinetic terms fluctuate a lot, making the energy expectation value slowly converging. \(^{39}\)

\(^{39}\)Remember there are also natural energy fluctuations proportional to the specific heat.
Virial energy estimator

While the thermal kinetic energy fluctuates a lot, the potential energy doesn’t. According to the quantum virial theorem one can compute the kinetic energy from the potential energy. The energy of a 1D system is

\[
E = \frac{1}{2} x \frac{\partial V(x)}{\partial x} + V(x) .
\]  

(225)

Herman et al. 40 introduced the virial energy estimator in PIMC and lead the way to smaller variance.

---

Virial energy estimator

One way to obtain a better energy estimator is to partially integrate some path variable and get a kinetic terms with smaller fluctuations. The result is\(^{41}\)

\[
\langle E \rangle_{\text{vir}} = \frac{1}{M} \sum_{m=1}^{M} \left\langle \frac{3N}{2L_\tau} - \frac{(R_{L+m} - R_m) \cdot (R_{m+1} - R_m)}{4L_\tau^2 \lambda} \right. \\
- \left. \frac{1}{2} \mathbf{F}_m \Delta_m + \frac{\partial U_m}{\partial \tau} \right\rangle 
\]

\[
\mathbf{F}_m := -\frac{1}{\tau} \nabla (U_{m-1} - U_m) \sim \text{classical force} 
\]

\[
\Delta_m := \frac{1}{2L} \sum_{j=-L+1}^{L-1} (R_m - R_{m+j}) \text{ deviation from average} 
\]

and \(L \in [1, M]\) is a parameter controlling how many time slices are averaged over.

\(^{41}\)Derived in the appendix of Ceperley’s PIMC review (1995).
Virial energy estimator

A good starting point is the identity (partial integration, \(d\) dimensions)

\[
\int d\mathbf{R} \sum_{j=1}^{M} \left( (\mathbf{R}_j - \mathbf{R}_j^*) \cdot \frac{\partial}{\partial \mathbf{R}_j} \right) \rho(\mathbf{R}, \mathbf{R}; \beta) = -\int d\mathbf{R} \sum_{j=1}^{M} \left( dN - \frac{\partial \mathbf{R}_j^*}{\partial \mathbf{R}_j} \right) \rho(\mathbf{R}, \mathbf{R}; \beta),
\]

where \(\mathbf{R}^*\) is a reference configuration\(^{42}\) The choice

\[
\mathbf{R}_j^* = \frac{1}{2M} \sum_{k=0}^{M-1} (\mathbf{R}_{j+k} + \mathbf{R}_{j-k}) \text{ would give } \left( \frac{\partial \mathbf{R}_j^*}{\partial \mathbf{R}_j} = \nabla_{\mathbf{R}_j} \cdot \mathbf{R}_j^* \right)
\]

\[
\sum_{j=1}^{M} \left( dN - \frac{\partial \mathbf{R}_j^*}{\partial \mathbf{R}_j} \right) = (M - 1)dN,
\]

and the rest is computing \(\frac{\partial}{\partial \mathbf{R}_j} \rho(\mathbf{R}, \mathbf{R}; \beta)\).

\(^{42}\)A fixed reference point is crucial in defining positions and continuous paths with periodic boundary conditions.
Virial energy estimator

Special cases:

- $L = 1$, so $\Delta_m = 0$ and the virial estimator reduces to the thermal one,

$$
\langle E \rangle_{\text{vir}} = \frac{1}{M} \sum_{m=1}^{M} \left\langle \frac{3N}{2\tau} - \frac{(R_{m+1} - R_m)^2}{4\tau^2\lambda} + \frac{\partial U_m}{\partial \tau} \right\rangle = \langle E \rangle_{\text{th}}
$$

- $L = M$ (usual choice; notice: $j = 0$ term is zero)

$$
\Delta_m = \frac{1}{2M} \sum_{j=-M+1}^{M-1} (R_m - R_{m+j})
$$

$\langle E \rangle_{\text{vir}} = \frac{1}{M} \sum_{m=1}^{M} \left\langle \frac{3N}{2\beta} - \frac{(R_{m-M} - R_m)}{4M\tau^2\lambda} \cdot (R_{m+1} - R_m) + \frac{1}{2\tau} \nabla(U_{m-1} - U_m)\Delta_m + \frac{\partial U_m}{\partial \tau} \right\rangle
$$
Virial energy estimator in the primitive approximation

The virial kinetic energy in the primitive approximation is

\[
\langle T \rangle = \frac{1}{2M} \sum_{m=1}^{M} \langle \mathbf{R}_m \cdot \nabla V(\mathbf{R}_m) \rangle .
\] (235)

This is already quite good, but writing positions wrt. a reference point reduces the variance further. One choice is to use a centroid, the center of mass of a polymer ring. Equally well one could use just any point on the particles path,

\[
\langle T \rangle = \frac{3N}{2\beta} + \frac{1}{2M} \sum_{m=1}^{M} \langle (\mathbf{R}_m - \mathbf{R}_M) \cdot \nabla V(\mathbf{R}_m) \rangle .
\] (236)

The virial energy estimator for each approximate action has to be elaborated separately, because the inter-action term is different.
Virial energy estimator

The virial theorem can also be derived using scaled coordinates, where the reference points $R_j^*$ are needed to make the scaling well defined. Physically, expanded imaginary time paths correspond to lower temperature $\beta'$, so $s$ is some function of $\beta'/\beta$ with $\beta' \to \beta$ as $s \to 1$. It turns out that $s = \sqrt{\beta'/\beta}$ is a good choice. The thermal estimator is

$$E(\beta) = E_s(\beta')|_{s=1} = -\frac{\partial \ln Z_s(\beta')}{\partial \beta'}|_{s=1}$$

$$= -\frac{1}{Z(\beta)} \frac{\partial}{\partial \beta'} \int dR^s \rho(R^s, R^s; \beta')|_{s=1}$$

$$= -\frac{1}{Z(\beta)} \frac{\partial}{\partial \beta'} \int dR^s(M-1)dN \rho(R^s, R^s; \beta')|_{s=1} .$$

The jacobian, with a fixed reference point reducing the free dimensions by $dN$, has $M$ which cancels with the bad term $dMN/(2\beta)$, and variances reduces considerably. Take case of boson exchange!
Bosons and fermions

We haven’t touched the subject of particle statistics yet. The Hamiltonian (206) doesn’t know about statistics, we have to force it. The density matrix eigenfunction expansion is

\[ \rho(\mathbf{R}', \mathbf{R}; \tau) = \langle \mathbf{R}' | e^{-\beta \hat{H}} | \mathbf{R} \rangle = \sum_i e^{-\beta E_i} \phi_i(\mathbf{R}') \phi_i(\mathbf{R}) , \]  

(241)

so we need is to require that summed states \( \phi_i(\mathbf{R}) \) must be (anti)symmetric. We could use Slater determinants (fermion) or permanents (boson), but the notation is simpler with permutations,

\[ \phi_{i,B/F}(\mathbf{R}) = \frac{1}{N!} \sum_P (\pm 1)^P \phi_i(\hat{P} \mathbf{R}) \begin{cases} + & \text{for bosons (B)} \\ - & \text{for fermions (F)} \end{cases} , \]  

(242)

\( \hat{P} \) permutes particle labels and \( P \) counts pair permutations; \( \sum_P \) denotes the sum over all permutations. For \( N = 3 \):

\[
\begin{align*}
123 : & \quad P = 0 \\
213 : & \quad P = 1 \\
231 : & \quad P = 2 \\
321 : & \quad P = 1 \\
132 : & \quad P = 1 \\
312 : & \quad P = 2 .
\end{align*}
\]
Bosons and fermions

We get the Bose/Fermi density matrix

$$\rho_{B/F}(\mathbf{R}', \mathbf{R}; \tau) = \sum_i e^{-\beta E_i} \phi_i(\mathbf{R}') \left[ \frac{1}{N!} \sum_P (\pm 1)^P \phi_i(\hat{\mathbf{P}} \mathbf{R}) \right]$$

(243)

The symmetrized states give the boson density matrix,

$$\rho_B(\mathbf{R}', \mathbf{R}; \tau) = \frac{1}{N!} \sum_P \rho(\mathbf{R}', \hat{\mathbf{P}} \mathbf{R}; \tau)$$

(244)

and the antisymmetrized states give the fermion density matrix,

$$\rho_F(\mathbf{R}', \mathbf{R}; \tau) = \frac{1}{N!} \sum_P (-1)^P \rho(\mathbf{R}', \hat{\mathbf{P}} \mathbf{R}; \tau)$$

(245)

--

43It doesn’t matter whether the permutation operator operates on \(\mathbf{R}\), on \(\mathbf{R}'\), or on both.
Fermion density matrix

The fermion density matrix as determinant of one-body density matrices \( \rho_1 \):

\[
\rho_F(R', R; \tau) = \begin{vmatrix}
\rho_1(r_1, r'_1; \tau) & \rho_1(r_2, r'_1; \tau) & \ldots & \rho_1(r_N, r'_1; \tau) \\
\rho_1(r_1, r'_2; \tau) & \rho_1(r_2, r'_2; \tau) & \ldots & \rho_1(r_N, r'_2; \tau) \\
\vdots & \vdots & \ddots & \vdots \\
\rho_1(r_1, r'_N; \tau) & \rho_1(r_2, r'_N; \tau) & \ldots & \rho_1(r_N, r'_N; \tau)
\end{vmatrix} .
\] (246)

The equivalent to FN-DMC, the fixed-node trial density matrix needs a reference point \( R^* \). See Militzer: two-particle nodes:

\[
\rho_T(r_1, r_2, r_1^*, r_2^*; \tau) = \begin{vmatrix}
\rho_1(r_1, r_1^*; \tau) & \rho_1(r_2, r_1^*; \tau) \\
\rho_1(r_1, r_2^*; \tau) & \rho_1(r_2, r_2^*; \tau)
\end{vmatrix} ,
\] (247)

which for free particles one has

\[
\rho_T(r_1, r_2, r_1^*, r_2^*; \tau) = (4\pi\lambda\tau)^{6/2} \begin{vmatrix}
e^{-\frac{(r_1-r_1^*)^2}{4\lambda\tau}} & e^{-\frac{(r_2-r_1^*)^2}{4\lambda\tau}} \\
e^{-\frac{(r_2-r_1^*)^2}{4\lambda\tau}} & e^{-\frac{(r_2-r_2^*)^2}{4\lambda\tau}}
\end{vmatrix} ,
\] (248)

which has nodes when \( (r_1 - r_2) \cdot (r_1^* - r_2^*) = 0 \).
Boson superfluidity

Let’s go back on the ring polymer picture. How large are the rings? The spring term is a gaussian \( e^{\frac{(R_{m-1} - R_m)^2}{4\lambda\tau}} \) with variance \( \sigma^2 = 2\lambda\tau \), and \( M \) of them could cover a circle with radius

\[
r \sim \sqrt{2\lambda M\tau} = \sqrt{\frac{\hbar^2\beta}{m}}.
\]  

(249)

Multiply with \( \sqrt{2\pi} \) and you get the deBroglie wavelength,

\[
\lambda_{\text{deBroglie}} = \sqrt{\frac{2\pi\hbar^2\beta}{m}}.
\]  

(250)

Lower \( T \) \( \Rightarrow \) larger \( \beta \) \( \Rightarrow \) larger \( \lambda_{\text{deBroglie}} \): at some low \( T \) two boson polymers may coil together. The inter-polymer spacing in 3D is about \( \rho^{-1/3} \), so quantum statistics is important below the degeneracy temperature

\[
T_D = \frac{\rho^{2/3}\hbar^2}{mk_B} \quad (3D).
\]  

(251)

Remember the periodic boundaries in space.
Winding number and superfluidity

- Experimental superfluidity (1938): Kapitza, Allen and Misener
- Superfluid density is directly related to the response of the free energy to a boundary phase twist.\(^4^4\)
- PIMC on liquid \(^4^He\): Ceperley and Pollock.\(^4^5\)
- Lots of misconceptions and assumptions are about.\(^4^6\)
  The winding number is undefined for Hamiltonians that don’t conserve the particle number.

**Setup of Andronikashvili experiment:** System between two cylinders of radii \(R\) and \(R + d\), and rotate the cylinders **slowly:** a *rotating bucket.*

**PIMC setup:** assume \(d \ll R\) and put the system between two walls that are moving with *small* velocity \(v\) wrt. to the liquid.

**Two-fluid model:** The *normal component*, density \(\rho_N\), responds to the motion of the walls due to *friction*. The rest is *superfluid*, \(\rho_s = \rho - \rho_N\).

---


Winding number and superfluidity

A gas of excitations with energy $\epsilon(p)$ carries momentum (Galilean transf.)

$$\langle \hat{P} \rangle_v = \int dr \ p n(\epsilon - p \cdot v) , \quad (252)$$

where $n(\epsilon)$ is the distribution function for gas at rest. Expands assuming $v$ is small, and you’ll get $\langle \hat{P} \rangle_v = something \times v$, the something is the mass associated with excitations - this is the mass of the normal fluid. 47

Follow Pollock and Ceperley and compute the change in free energy. In the rest frame walls moving with velocity $v$ have the Hamiltonian

$$\hat{H}' = \sum_j \left( -i\hbar \nabla_j - mv \right)^2 + \hat{V} , \quad (253)$$

and the density operator is $\hat{\rho}' = e^{-\beta \hat{H}'}$, the same as $\hat{\rho}_v$ in the rest frame of the system. The momentum of the system can now be computed.

47 L. Landau, E. M. Lifshitz, and L. P. Pitaevskii, Statistical Physics part II. If a thing has momentum, it has mass.
Winding number and superfluidity

Using statistical physics,

\[
\langle \hat{A} \rangle = \frac{\text{Tr}[\hat{A}\hat{\rho}]}{\text{Tr}[\hat{\rho}]} = \frac{\sum_i \langle i | \hat{A} \hat{\rho} | i \rangle}{\sum_i \langle i | \hat{\rho} | i \rangle} = \frac{1}{Z} \sum_i \langle i | \hat{A} \hat{\rho} | i \rangle ,
\]

(254)

after some work we find the momentum of the system with moving walls (total density \( \rho \), volume is \( N/\rho \)),

\[
\langle \hat{P} \rangle_v = \frac{\text{Tr}[\hat{P}\hat{\rho}_v]}{\text{Tr}[\hat{\rho}_v]} \approx \text{normal mass } \times \nu = (N N m) \nu = \left( \rho_N \frac{N}{\rho} m \right) \nu .
\]

(255)

Apparently this proportionality can hold only if \( \nu \) is small, but that’s all we need.
Winding number and superfluidity

Compute the momentum and expand it to linear order in $\nu$. The free energy is

$$F_\nu = -\frac{1}{\beta} \ln Z_\nu = -\frac{1}{\beta} \ln \text{Tr}[\hat{\rho}_\nu] ,$$

so you finally get (details in Pollock-Ceperley article),

$$\frac{\Delta F_\nu}{N} = \frac{1}{2} m \nu^2 \frac{\rho S}{\rho} + O(\nu^4) .$$

In PIMC, we don’t need actual moving walls to measure $\Delta F_\nu$. The density matrix with moving walls, $\rho_\nu$, satisfies the Bloch equation,

$$- \frac{\partial \rho_\nu(R', R; \beta)}{\partial \beta} = \hat{H}' \rho_\nu(R', R; \beta) .$$

where we use periodic boundary conditions, $r_i + L = r_i$. 
Winding number and superfluidity

Define a modified density matrix $\tilde{\rho}$ via a phase factor,

$$\phi(R) := \frac{m}{\hbar} \sum_j (r_j - r'_j)$$  \hspace{1cm} (259)

$$\rho_v(R', R; \beta) = e^{i\phi(R)} \tilde{\rho}(R', R; \beta),$$  \hspace{1cm} (260)

and insert it to the Bloch equation. After some manipulation you get

$$- \frac{\partial \tilde{\rho}(R', R; \beta)}{\partial \beta} = \hat{H} \tilde{\rho}(R', R; \beta),$$  \hspace{1cm} (261)

but the boundary conditions come with a phase factor,

$$\tilde{\rho}(R', r_1...r_i + L, ...r_N; \beta) = e^{-imv\cdot L/\hbar} \tilde{\rho}(R', R; \beta).$$  \hspace{1cm} (262)

The \textit{winding number} $W$ counts how many times particle paths wind around periodically. Each winding adds a factor $e^{-imv\cdot L/\hbar}$. 
Winding number and superfluidity

The principle of how the Galilean transformation

\[ \mathbf{P}_j = -i\hbar \nabla_j - mv_j \]  \hspace{1cm} (263)

in the Hamiltonian turns into a boundary condition for the density matrix, phase \( e^{-imv \cdot \mathbf{L}/\hbar} \), is widely applicable. The unitary operator that does the Galilean transformation is the phase,

\[ \hat{U} = e^{-imv \cdot \mathbf{L}/\hbar}, \quad \hat{U}^\dagger \hat{P} \hat{U} = \hat{P} - mv\hat{N}. \]  \hspace{1cm} (264)

Electrons in magnetic field have the momentum operator

\[ \hat{\mathbf{P}}_j = -i\hbar \nabla_j - e\mathbf{A}, \]  \hspace{1cm} (265)

and if they move a full circle around a flux tube with flux \( \Phi \), their wave function picks up the Aharonov-Bohm phase \( e^{-ie\Phi/\hbar} \). 48

\[ ^{48}\text{The funny thing is that even if the circulating electron visits only regions with } \mathbf{A} = 0, \text{ the wave function still picks up the phase.} \]
Winding number and superfluidity

From the winding number we can compute the superfluid fraction, in 3D

\[
\frac{\rho_S}{\rho} = \frac{m}{\hbar^2} \frac{\langle W^2 \rangle L^{-1}}{3 \rho \beta}. \tag{266}
\]

On the other hand, the ideal Bose gas Bose-Einstein condensation (BEC) temperature is

\[
T_c = \frac{3.31 \hbar^2}{g m k_B} \rho^{2/3}, \tag{267}
\]

where \( g \) is the \( k \)-state degeneracy factor. For liquid \( ^4 \text{He} \) density this would give \( T_c \sim 3.14 \text{ K} \), \(^{49}\) a bit above the experimental superfluid transition temperature at 2.17 K. The fact that the temperatures are so close led Fritz London to propose that BEC is related to superfluidity, and that He-He interactions suppress \( T_c \) down to 2.17 K. Let’s see visually what superfluidity requires, next pages contain imaginary time plots of paths.

Boson superfluidity: not superfluid
Boson superfluidity: not superfluid
Boson superfluidity: not superfluid
Boson superfluidity: not superfluid
Boson superfluidity: not superfluid
Boson superfluidity: Start again, not superfluid
Boson superfluidity: *Superfluid!*

![Diagram of boson superfluidity]
Boson superfluidity: *Superfluid!*

- Net flux in imaginary time over *any* boundary in space (such as the green boundary, periodic boundaries aren’t special)
- Four particles in exchange loop, one particle not participating
- Order $N$ particles, not necessarily all particles, is needed for the exchange loop to reach through the box.
Ideal bosons

The simple case of ideal bosons is interesting. We can interpret the particle exchange at $\beta$ as a single particle at the lower temperature $2\beta$:

This is allowed, since there are no interaction links to break.
Ideal bosons

We can now reduce all exchange loops to stretched single particle paths, starting from the single-particle partition function $Z_1$,

\[ Z_1(\beta) = "(1)" = Z_1(\beta) \]  \hspace{1cm} (268)

\[ Z_2(\beta) = "(12) + (21)" = \frac{1}{2} [Z_1(\beta)^2 + Z_1(2\beta)] \]  \hspace{1cm} (269)

\[ Z_3(\beta) = "(123) + (132) + (213) + (312) + (231)" \]

\[ = \frac{1}{3!} [Z_1(\beta)^3 + 3Z_1(\beta)Z_1(2\beta) + 2Z_1(3\beta)] \]  \hspace{1cm} (270)

and so on. Some combinatorics leads to the recursion relation

\[
Z_N(\beta) = \frac{1}{N} \sum_{n=1}^{N} Z_1(n\beta)Z_{N-n}(\beta) .
\]  \hspace{1cm} (271)

For example (init recursion with $Z_1(0) = 1$),

\[ Z_3(\beta) = \frac{1}{3} \left[ Z_1(\beta)Z_2(\beta) + Z_1(2\beta)Z_1(\beta) + Z_1(3\beta) \right] \]

\[ = \frac{1}{3} \left[ Z_1(\beta) \frac{1}{2} \left( Z_1(\beta)^2 + Z_1(2\beta) \right) + Z_1(2\beta)Z_1(\beta) + Z_1(3\beta) \right] \]

\[ = \frac{1}{6} \left[ Z_1(\beta)^3 + 3Z_1(\beta)Z_1(2\beta) + 2Z_1(3\beta) \right] \text{ OK} . \]  \hspace{1cm} (272)
Ideal bosons

The (inefficient) bose.py demonstrates how the specific heat peak forms at $T_c$ as $N$ increases - and how important it is to use the correct $Z_1(\beta)$.  

*Spoiler:* $Z_1(\beta)$ must count the ground state accurately, or you won’t get

$$\lim_{T \to 0} C_V(T) = 0 .$$  \hspace{1cm} (273)

One choice is to *fix the density*, so that more particles occupy larger volume,

$$\rho = \frac{N}{\Omega} = \text{constant} \Rightarrow \Omega = \frac{N}{\rho} \propto N .$$  \hspace{1cm} (274)

---


51 We need this limit or we can’t define the entropy of a given state, only entropy differences of states.
Ideal bosons

Once $Z_1(\beta)$ is set, the recursion to any $N$ is determined. It doesn't depend on any spatial quantities, in particular possible periodic boundaries. PBC allows the same permutation, say "(312)", to cover more than one spatial interpretations. Say particles 1 and 2 won't cross the boundary. Particle 3 may end up at the original position of particle 1 ($r_3(\beta) = r_1(0)$)

1. without crossing the periodic boundary; $r_3(\tau)$ is continuous in the box.

2. crossing the periodic boundary; $r_3(\tau)$ is discontinuous in the box. The path winds (=jumps) from side to side at some $\tau'$,

$$r_3(\tau') = (x_3(\tau'), y_3(\tau'), z_3(\tau')) = (x_3(\tau') - L, y_3(\tau'), z_3(\tau')).$$

The path is continuous in the extended view of periodic boxes. Case 1 gives zero winding, but case 2 does wind. From the exactness of the recursion we must conclude, that both paths are present in the computed $Z_3(\beta)$, if $Z_1(\beta)$, the "seed", has PBC.
Ideal bosons in a box

Let’s look at the calculation of $Z_1(\beta)$. For a particle in a box, without PBC and $V = \infty$ outside the box, one has\textsuperscript{52}

$$Z_1(\beta) = \left[ \sum_{q=1}^{\infty} e^{-\beta q^2(\frac{\pi^2 \lambda}{\Omega^2/3})} \right]^3.$$  \hspace{1cm} (275)

Approximately, the sum can be converted to an integral, and taking the ground state separately one finds the form used in bose.py.

$$Z_1(\beta) \approx 1 + \frac{\Omega}{(4\pi \lambda)^{3/2}} \beta^{-3/2} = 1 + \frac{\rho}{(4\pi \lambda)^{3/2}} \beta^{-3/2},$$  \hspace{1cm} (276)

The reason I’m going through all this is that

As $N$ increases, $C_V(T)$ has the peaked signature of BEC transition at $T_c$, even though there are no periodic boundaries and no winding paths.

\textsuperscript{52}Again, see K. Glaum, H. Kleinert, and A Pelster, arXiv 0707.2715v2 (2007). I’ve changed the notation a bit.
Ideal bosons and superfluidity

Earlier we described superfluidity as winding around periodic boundaries, and from the ideal bosons-in-a-box calculation we concluded, that BEC happens regardless of PBC or winding. So what gives? The dilemma is in relating two phenomena,

- superfluidity as zero viscosity and finite critical velocity.
- Bose-Einstein condensation to the \( k = 0 \) state.

For He\(^4\), neutron scattering experiments have confirmed that the momentum distribution shows BEC (about 10 % condensed) at \( T = 0 \).

**Ideal bosons have zero critical velocity, so there's BEC but no superfluidity.** As Pethick and Smith put it,\(^{53}\)

However, the connection between Bose-Einstein condensation and superfluidity is a subtle one. A Bose-Einstein condensed system does not necessarily exhibit superfluidity, an example being the ideal Bose gas for which the critical velocity vanishes, as demonstrated in Sec. 10.1 below. Also lower-dimensional systems may exhibit superfluid behavior in the absence of a true condensate, as we shall see in Chapter 15.

---

Pair product approximation to the density matrix

We’ve already discussed 4th order propagators, which serve as system-independent work horses and get us down to low $T$. One more method deserves attention, the *pair product approximation*. For a hard sphere potential the exact two-body density matrix would obviously be useful, knowing the merits of the pair-product Jastrow wave function. In general, we have two guidelines for density matrices:

- The density matrix satisfies the Bloch equation (see also Eq. (258))

$$\frac{\partial \rho(R, R'; \beta)}{\partial \beta} = \hat{H} \rho(R, R'; \beta) ,$$  \hspace{1cm} (277)

- The Feynman-Kac formula for density matrices,

$$\rho(R, R'; \beta) = \rho_0(R, R'; \beta) \langle e^{-\int_0^\beta dt V(R(t))} \rangle_{BRW} ,$$  \hspace{1cm} (278)

where $\rho_0$ is the free-particle density matrix, and the average is over all Brownian random walk paths.
Pair product approximation to the density matrix

The Feynman-Kac formula is easy to derive, just take the primitive approximation and use $M - 1$ intermediate points,

$$\rho(R, R'; \beta) \approx \langle R | e^{-\beta \hat{T}} e^{-\beta \hat{V}} ... e^{-\beta \hat{T}} e^{-\beta \hat{V}} | R' \rangle$$  \hspace{1cm} (279)

$$= \int dR_1...R_{M-1} \rho_0(R, R_1; \beta)...\rho_0(R_{M-1}, R'; \beta) e^{-\beta \frac{1}{M} \sum_{i=1}^{M} V(R_i)}$$  \hspace{1cm} (280)

$$= \rho_0(R, R'; \beta) \frac{\int dR_1...R_{M-1} \rho_0(R, R_1; \beta)...\rho_0(R_{M-1}, R'; \beta) e^{-\beta \frac{1}{M} \sum_{i=1}^{M} V(R_i)}}{\rho_0(R, R'; \beta)} \langle e^{-\beta \frac{1}{M} \sum_{i=1}^{M} V(R_i)} \rangle_{BRW_M}$$

with $R_M = R'$. Let $M \rightarrow \infty$,

$$\langle e^{-\beta \frac{1}{M} \sum_{i=1}^{M} V(R_i)} \rangle_{BRW_M} \rightarrow \langle e^{-\int_0^\beta dt V(R(t))} \rangle_{BRW} ,$$  \hspace{1cm} (281)

and you have the Feynman-Kac formula, an exact relation.

\textsuperscript{54}The resulting density matrix is not symmetric, but it will be in the limit $M \rightarrow \infty$.\textsuperscript{175}
Pair product approximation to the density matrix

If the potential in the Feynman-Kac formula is a pair potential, then

$$e^{-\int_0^\beta dt V(R(t))} = \prod_{i<j} e^{-\int_0^\beta dt V(r_{ij}(t))}.$$  \hfill (282)

In Jastrow spirit, a good approximation is to assume that particles are correlated only in pairs, so that you can average each pair independently,

$$\langle \prod_{i<j} e^{-\int_0^\beta dt V(r_{ij}(t))} \rangle_{\text{BRW}} \approx \prod_{i<j} \langle e^{-\int_0^\beta dt V(r_{ij}(t))} \rangle_{\text{BRW}}$$ \hfill (283)

$$\implies U(R, R'; \beta) \approx \prod_{i<j} u(r_{ij}, r_{ij}'; \beta).$$  \hfill (284)

This is the pair product approximation, which ignores three-body correlations and beyond. The point is that $u(r_{ij}, r_{ij}'; \beta)$ can be computed accurately either using matrix squaring or eigenfunction expansion.
Matrix squaring

1. Write the two-body problem in center-of-mass and relative coordinates

2. Use spherical coordinates, angular momentum becomes a parameter $l$,

$$\rho(r, r'; \beta) = \sum_l \rho_l(r, r'; \beta) P_l(\cos(\theta))$$  \hspace{1cm} (285)

3. Solve the density matrix at high $T$ and iterate, for each $l$, down to desired $T$:

$$\rho_l(r, r''; 2\beta) = \int_0^\infty dr'' \rho_l(r, r'', \beta) \rho_l(r'' r'; \beta)$$  \hspace{1cm} (286)

4. Compute the final result by summing over $l$.

The actual numerical implementation needs a few tricks to keep the result accurate, but *the two-body problem is solved exactly*. 
Pair product approximation to the density matrix

Making further approximations,\(^{55}\) the *cumulant approximation* is

\[
\prod_{i<j} \langle e^{-\int_0^\beta dt V(r_{ij}(t))} \rangle_{\text{BRW}} \approx \prod_{i<j} e^{-\int_0^\beta dt \langle V(r_{ij}(t)) \rangle_{\text{BRW}}} \tag{287}
\]

and, assuming classical paths we get the *semi-classical approximation*

\[
\prod_{i<j} \langle e^{-\int_0^\beta dt V(r_{ij}(t))} \rangle_{\text{BRW}} \approx \prod_{i<j} e^{-\int_0^\beta dt V(r_{ij}^{(\text{class})}(t))} , \tag{288}
\]

and for a very small \(\beta\) one recovers the *primitive approximation*

\[
\prod_{i<j} \langle e^{-\int_0^\beta dt V(r_{ij}(t))} \rangle_{\text{BRW}} \approx \prod_{i<j} e^{-\beta \frac{1}{2} \left[ V(r_{ij}(0)) + V(r_{ij}(\beta)) \right]} . \tag{289}
\]

Liquid $^4$He

Pollock and Ceperley (1987):

\[ \frac{\rho_N}{\rho} \]

\[ T \text{ (K)} \]

**FIG. 1.** Computed $\rho_N/\rho$ values along the SVP curve (lower panel). The solid line is the experimental curve (Ref. 13). The solid dots, for 125 atoms, are calculated from Eq. (33). The estimates based on the winding number [Eq. (22)] (open circles) are for 64 atoms. The upper panel shows the $x$-direction winding-number distribution at $T = 2.0$ and $1.18 \text{ K}$. 
Liquid $^4$He

D. Ceperley, Rev. Mod. Phys. 67, 279 (1995):

**FIG. 10.** The extended trace of six $^4$He atoms at a temperature of 0.75 K and with 53 time slices. The dashed square represents the periodic boundary conditions. Three of the atoms are involved in an exchange which winds around the boundary in the $x$ direction.

**FIG. 11.** The specific heat of $^4$He: solid line, experiment at saturated vapor pressure (Wilks, 1967); triangles with error bars, PIMC calculations (Ceperley and Pollock, 1986); open circles, Feynman-Kikuchi model with $20^3$ sites (Elser, 1984). In Elser’s calculation, only the fluctuation term in the specific heat has been included, and the temperature has been scaled to match the experimental transition temperature.
Liquid $^4$He

D. Ceperley, Rev. Mod. Phys. 67, 279 (1995):

FIG. 13. The energy/atom of $^4$He: solid line, experiment at saturated vapor pressure; symbols with error bars, PIMC calculations (Ceperley and Pollock, 1986). The upper panel is the potential energy (solid line and left scale) and kinetic energy (dashed line and right scale).
Liquid $^4$He

The density has a kink at $T_\lambda$, so it's present "already" in the liquid structure:

\[
\rho(T) = \frac{N}{\Omega(T)}
\]

The equilibrium density of liquid He$^4$ as a function of temperature $^a$


For pair interactions, the finite-$T$ potential energy of homogeneous liquid is ($\rho(T) = N/\Omega(T)$ with volume $\Omega(T)$)

\[
\langle V \rangle_{\beta} = \frac{1}{Z} \int dR \langle R | e^{-\beta \hat{H}} | R \rangle = \frac{1}{2} \int dr_1 dr_2 V(r_{12}) \frac{1}{Z} \sum_i e^{-\beta E_i} N(N-1) \int dr_3 ... dr_N | \phi_i(R) |^2
\]

\[
= \frac{1}{2} \int dr_1 dr_2 V(r_{12}) \langle \rho_2(r_1, r_2) \rangle_{\beta} = \frac{1}{2} \Omega(T) \rho(T)^2 4\pi \int dr^2 V(r) g(r; T)
\]

\[
= \Omega(T) 4\pi \rho(T)^2 \int_0^\infty dr r^2 V(r) g(r; T) = 2\pi N \rho(T) \int_0^\infty dr r^2 V(r) g(r; T)
\]

(290)

Comparison with the PIMC data shows that most of the $T$-dependence comes from $\rho(T)$. The integral is roughly a constant, about $-978 \text{KÅ}^{-3}$, and the pair distribution $g(r; T)$ has a very weak temperature dependence.
Path sampling

We want to sample different terms in $Z$ as effectively as possible, including permutations of indistinguishable particles.

- Single bead move: move just one random particle in a random time slice
- Rigid move: move the whole imaginary time world line of a random particle
- Recreate path(s) for multiple time slices, say slices 2, 3, 4, keeping slices 1 and 5 fixed; accept or reject at any stage:
  - Bisection: sample 3 as midpoint 1-3-5, sample 2 as midpoint 1-2-3 and 4 as midpoint 3-4-5.
    - Easy permutation sampling, but insufficient for some topologies
  - Staging: sample 2 from 1-2-5, then 3 from 2-3-5, then 4 from 3-4-5.
- Worm move: open a path and let the ends move and recombine
  - Global updates of paths
  - *The ultimate permutation sampler*
Path sampling: single-bead move

Update: two $G_\hat{T}$’s and potential energy in the time slice.

*Ineffective but cheap*
Path sampling: rigid move

Update: potential energy of all time slices. No $G_{\vec{t}}$ change.

*Effective but expensive*
Path sampling: bisection
Path sampling: bisection
Path sampling: bisection
Path sampling: bisection
Path sampling: bisection
Path sampling: bisection
Path sampling: Adding permutations to bisection
Path sampling: Adding permutations to bisection
Path sampling: Adding permutations to bisection
Path sampling: Adding permutations to bisection
Path sampling: Adding permutations to bisection
Path sampling: Adding permutations to bisection
Path sampling: Adding permutations to bisection
Bisection

Comments on the bisection update

▶ Effective, path changes a lot
  ▶ If the new midpoint is accepted, it may be far from the original point
  ▶ If the new midpoint is rejected (large potential) we don’t waste time on the rest
  ▶ Does the algorithm early-reject a move that could be finally accepted? In principle yes, but without any practical consequences

▶ Move to the temporary mid-point (red) is deterministic

▶ Move from the temporary mid-point to new point (green) can be done using random Gaussian move. The length can be adjusted for optimal acceptance.

▶ Long $G_f$’s of previous stage can be divided away in the next stage. Intermediate steps can be free-spirited, but in the final step you have to use the correct weights.

▶ Adjustable parameters: length of the re-sampled path and amplitudes of random Gaussian moves on each bisection level.

\[56\] Llorenç Brualla i Barberà, Thesis (Barcelona, 2002).
Worm

The worm algorithm\textsuperscript{57} owns to the advances made in the Ising model MC sampling in the 80’s. It solves many problems in PIMC path sampling:

- **SOLVED: Global changes** are more efficient than local ones
- **SOLVED: Long permutation loops** are easily created with high acceptance
  - other path sampling methods *explicitly construct* permutations
  - acceptance of suggested permutations is very low $\Rightarrow$ permutation sampling is very inefficient
  - on a torus-shaped space you practically never get a path winding around $\Rightarrow$ wrong result
- **SOLVED: Evaluation of off-diagonal quantities**: condensate fraction and imaginary-time correlations, without any ambiguity in normalization.

Worm

Start with a configuration that describes one possible term in the partition function $Z$. This term is in the $Z$ sector. The basic worm updates are:

1. **Open**: Open a path, this leaves two dangling ends, *Head* and *Tail*. You generated a term that’s not in $Z$, and entered the $G$ sector.
2. **Advance/recede**: Head moves ahead (pulls back) and the worm gets longer (shorter)
3. **Insert/remove**: Add a completely new piece of path or remove one
   ⇒ *Grand Canonical Ensemble* (variable $N$, fixed chemical potential $\mu$)
4. **Swap**: Head jumps to another path, close the old path
5. **Close**: Close the dangling ends of the worm and return to the $Z$ sector. This must be in balance with the open update – *detailed balance!*

While in the $Z$ sector, measure anything you would from $Z$.
While in the $G$ sector, measure off-diagonal quantities: Head and Tail correspond to the field operators $\psi(r; \tau)$ and $\psi^\dagger(r'; \tau')$. 
Path sampling: worm open

\[ \hat{\psi}(r; \tau_H) \]

\[ \hat{\psi}^\dagger(r; \tau_T) \]
Path sampling: worm recede
Path sampling: worm advance
Path sampling: worm close
Path sampling: worm insert
Path sampling: worm advance
Path sampling: worm advance
Path sampling: worm advance
Path sampling: worm advance
Path sampling: worm close
Path sampling: worm swap
Path sampling: worm swap PERMUTATIONS!
Worm partition function

The $Z$ and $G$ sectors are also known with less colourful names: *diagonal* and *off-diagonal* sector. The worm samples an "extended" partition function,

$$
Z_{\text{worm}} = \sum_{N=0}^{\infty} Z_N \left[ e^{\beta \mu N} + C \sum_{i_H,i_T} \int d r_H d r_T G(r_H, r_T; (i_H - i_T) \tau) \right],
$$

(291)

where the integrand is the *Matsubara Green’s function* (thermal average),

$$
G(r, r'; \tau - \tau') = \langle \hat{T}_\tau \hat{\psi}(r, \tau) \hat{\psi}^\dagger(r'; \tau') \rangle.
$$

(292)

Here $\hat{T}_\tau$ is the imaginary-time ordering operator and $Z_N$ is the canonical partition function. The $Z_{\text{worm}}$ is sum of the grand canonical partition function and sum over all worms.
Worm partition function

- Worm’s natural ensemble is grand canonical, but also a canonical implementation has been done. The grand canonical partition function is

\[
Z_G = \text{Tr} \left[ e^{-\beta (\hat{H} - \mu N)} \right] = \sum_{N=0}^{\infty} Z_N e^{\beta \mu N},
\]

(293)

and we approximate the canonical partition function \( Z_N \) with the multidimensional integral, \((X := \{ R_1, R_2 \ldots R_M \})\)

\[
Z_G \approx \sum_{N=0}^{\infty} e^{\beta \mu N} \int dX A(X; \tau) e^{-U(X)} .
\]

(294)

Here \( A(X; \tau) = \rho_0(R_1, R_2; \tau) \ldots - \rho_0(R_{M-1}, R_M; \tau) \) is the free-particle contribution and \( U(X) \) is the rest, all interaction terms.

- Paths (world lines) are created sampling exactly the free-particle density matrices \( \rho_0(r_1, r_2; \tau) \). This can be done using staging.
Staging algorithm

We generate a random walk path between two known endpoints \( \mathbf{r}_1 \) and \( \mathbf{r}_m \), marked just 1 and \( m \). The imaginary-time difference is \( \tau_m - \tau_1 := \tau_{1m} \).

Split a long-leap Gaussian to short-step Gaussians (here in 3D):

\[
\rho_0(\mathbf{r}_1, \mathbf{r}_m; \tau_{1m}) := \langle \mathbf{r}_1 | e^{-\tau_{1m} \hat{T}} | \mathbf{r}_m \rangle = \frac{1}{(4\pi \lambda \tau_{1m})^{3/2}} e^{-\frac{(\mathbf{r}_1 - \mathbf{r}_m)^2}{4\lambda \tau_{1m}}} 
\]

\[= \int d\mathbf{r}_2 \ldots d\mathbf{r}_{m-1} \langle \mathbf{r}_1 | e^{-\tau_{12} \hat{T}} | \mathbf{r}_2 \rangle \langle \mathbf{r}_2 | e^{-\tau_{23} \hat{T}} | \mathbf{r}_3 \rangle \ldots \langle \mathbf{r}_{m-1} | e^{-\tau_{m-1,m} \hat{T}} | \mathbf{r}_m \rangle \]

\[= \int d\mathbf{r}_2 \ldots d\mathbf{r}_{m-1} \frac{1}{(4\pi \lambda \tau_{12} \tau_{23} \ldots \tau_{m-1,m})^{3/2}} e^{-\frac{(\mathbf{r}_1 - \mathbf{r}_2)^2}{4\lambda \tau_{12}}} e^{-\frac{(\mathbf{r}_2 - \mathbf{r}_3)^2}{4\lambda \tau_{23}}} \ldots e^{-\frac{(\mathbf{r}_{m-1} - \mathbf{r}_m)^2}{4\lambda \tau_{m-1,m}}} . \]

Here the unknown, intermedia coordinates \( \mathbf{r}_2, \ldots, \mathbf{r}_{m-1} \) are entangled, but we can re-arrange terms.
Staging algorithm

\[ \rho_0(r_1, r_m; \tau_{1m}) = N \int dr_2 \cdots dr_{m-1} e^{-\frac{(r_1 - r_m)^2}{4\lambda \tau_{1m}}} \]

\[
\left[ e^{\frac{-\left(r_1-r_2 \right)^2}{4\lambda \tau_{12}}} \frac{-\left(r_2-r_m \right)^2}{4\lambda \tau_{2m}} e^{\frac{-\left(r_1-r_m \right)^2}{4\lambda \tau_{1m}}} \right] \left[ e^{\frac{-\left(r_2-r_3 \right)^2}{4\lambda \tau_{23}}} \frac{-\left(r_3-r_m \right)^2}{4\lambda \tau_{3m}} e^{\frac{-\left(r_2-r_m \right)^2}{4\lambda \tau_{2m}}} \right] \left[ e^{\frac{-\left(r_3-r_4 \right)^2}{4\lambda \tau_{34}}} \frac{-\left(r_4-r_m \right)^2}{4\lambda \tau_{4m}} e^{\frac{-\left(r_3-r_m \right)^2}{4\lambda \tau_{3m}}} \right] \left[ e^{\frac{-\left(r_m-3-r_m-2 \right)^2}{4\lambda \tau_{m-3,m-2}}} \frac{-\left(r_m-2-r_m \right)^2}{4\lambda \tau_{m-2,m}} e^{\frac{-\left(r_m-3-r_m \right)^2}{4\lambda \tau_{m-3,m}}} \right] \left[ e^{\frac{-\left(r_m-2-r_m-1 \right)^2}{4\lambda \tau_{m-2,m-1}}} \frac{-\left(r_m-1-r_m \right)^2}{4\lambda \tau_{m-1,m}} e^{\frac{-\left(r_m-2-r_m \right)^2}{4\lambda \tau_{m-2,m}}} \right], \quad (296) \]

and all extra colored terms cancel. The brackets simplify, for example

\[ e^{\frac{-\left(r_1-r_2 \right)^2}{4\lambda \tau_{12}}} \frac{-\left(r_2-r_m \right)^2}{4\lambda \tau_{2m}} e^{\frac{-\left(r_1-r_m \right)^2}{4\lambda \tau_{1m}}} = e^{\frac{-\left(r_2-r^*_2 \right)^2}{4\lambda \tau_{12m}}} \]

\[ \tau_{12m} := \frac{\tau_{12} \tau_{2m}}{\tau_{1m}} \quad \tau^*_2 := \frac{\tau_{2m} r_1 + \tau_{12} r_m}{\tau_{1m}}. \quad (297) \]
Staging algorithm

We now have a formula to generate a random walk between fixed end points $r_1$ and $r_m$ – don’t forget the blue term! –

$$\rho_0(r_1, r_m; \tau_{1m}) = e^{-(r_1-r_m)^2/4\lambda\tau_{1m}} \mathcal{N} \int dr_2...dr_{m-1} e^{-(r_2-r_2^*)^2/4\lambda\tau_{12m}} ... e^{-(r_{m-1}-r_{m-1}^*)^2/4\lambda\tau_{m-2,m-1,m}} ,$$

where

$$\tau_{ijk} := \frac{\tau_{ij} \tau_{jk}}{\tau_{ik}}, \quad \tau_{ij} := \tau_j - \tau_i$$

$$r_j^* := \frac{\tau_{jk} r_i + \tau_{ij} r_k}{\tau_{ik}} .$$

(299) (300)

Usage: compute $r_2^*$ using $r_1$ and $r_m$, and set

$$r_2 = r_2^* + \sqrt{2\lambda\tau_{12m}} \eta .$$

(301)

Compute $r_3^*$, sample $r_3$, etc. This samples the integrand exactly.

Notice that this staging algorithm doesn’t assume that the imaginary times are evenly spaced.
Worm partition function

The acceptance probabilities of worm updates are given by the Metropolis-Hastings formula:

\[ P_{\text{old} \to \text{new}} = \min \left\{ 1, \frac{T(\text{new} \to \text{old})W(\text{new})}{T(\text{old} \to \text{new})W(\text{old})} \right\}. \]  (302)

Free-particle paths between fixed \( r' \) and \( r \) are created with a random walk, which samples the product

\[ \rho_0(r', r_1; \tau)\rho_0(r_1, r_2; \tau)\ldots\rho_0(r_{k-1}, r; \tau). \]  (303)

From the \( Z_{\text{worm}} \) in Eq. (291) we should be able to compute the update probabilities.
Let’s go through the worm updates one by one. For more details, see the article by Boninsegni et al. from 2006.
Open worm

Adjustable parameters are blue. From now on, \textit{above} means up, and \textit{below} means down in imaginary time: imaginary time is periodic, with period $\beta$.

\begin{enumerate}
\item Set Head to a random bead, there are $N_{\text{beads}}$ to choose from.
\item Pick $k \in U[1, K]$ and set Tail $k$ steps above Head; $\tau_{HT} := \tau_H - \tau_T$.
\item Remove beads between Head and Tail.
\item Accept with probability
\end{enumerate}

$$P_{\text{open}} = \min \left\{ 1, \frac{C K N_{\text{beads}} e^{-\Delta U - \mu \tau_{HT}}}{\rho_0(r_H, r_T; \tau_{HT})} \right\} . \quad (305)$$

where

$$\Delta U = U(\text{new}) - U(\text{old}) = U(\text{open}) - U(\text{close}) .$$
Worm open, probability details

Let’s look how this creates the worm term in Eq. (291), repeated here in a simplified form,

\[
Z_{\text{worm}} = \sum_{N=0}^{\infty} \left[ Z_N \exp(\beta \mu N) + C \sum_{i_H, i_T} \int d\boldsymbol{r}_H d\boldsymbol{r}_T (Z_N \exp(\beta \mu N) - \text{path H-T}) \right].
\]

Head was picked from \( N_{\text{beads}} \) beads, which is the current value of \( NM \); and the reverse can be done in 1 way. Tail was picked from \( K \) beads, reverse has 1 possible way. Putting together, we have so far

\[
\frac{T(\text{new} \rightarrow \text{old})}{T(\text{old} \rightarrow \text{new})} = KN_{\text{bead}} = KNM.
\]

(306)

The old weight is \( \exp(\beta \mu N) \), while the new weight has a factor \( C \) and there are \( N - 1 \) particles for the duration \( \tau_{HT} \), \( \exp(\mu N(\beta - \tau_{HT}) + \mu (N-1)\tau_{HT}) \), so

\[
\frac{W(\text{new})}{W(\text{old})} = C \exp(-\mu \tau_{HT}).
\]

(307)
Open worm, details

Now things gets interesting. Deleting the path between Head and Tail also deletes the interactions this particle had between Head and Tail. Those are present in $Z_N$, so the weights have changed by the amount of the potential action; the explicit form depends on the approximation. Using $U$ defined in Eq. (294), and marking $\Delta U := U(\text{new}) - U(\text{old})$, we get

$$
\frac{W(\text{new})}{W(\text{old})} = e^{-\Delta U} .
$$

(308)

The free-particle paths are also different,

$$
\cdots \rho_0(\mathbf{r}_{H-1}, \mathbf{r}_H; \tau) \rho_0(\mathbf{r}_H, \mathbf{r}_{H+1}; \tau) \cdots \rho_0(\mathbf{r}_{T-1}, \mathbf{r}_T; \tau) \rho_0(\mathbf{r}_T, \mathbf{r}_{T+1}; \tau) \cdots .
$$

In closing the path we recreate the path segment with staging, which has the Head-Tail factor that’s not taken care of:

$$
\frac{W(\text{new})}{W(\text{old})} = \frac{1}{\rho_0(\mathbf{r}_H, \mathbf{r}_T; \tau_{HT})} .
$$

(309)

Now we have covered all terms in Eq. (305).
Open worm, details

- If $C = 0$ we never open a worm. Fine, because then there is none.
- If $C$ increases, we accept more worms.
- More particles, more worms – unless $C$ cancels $N$.
- More time slices, more worms – unless $C$ cancels $M$.
- We accept less worms with Head and Tail farther apart in a random walk.
Close worm

Reverse of open, must satisfy the detailed balance condition.

**Close**

1. Reject the update, if Tail is zero or more than $K$ steps above Head.
2. Generate a path from Head to Tail.
3. Accept with probability

$$P_{\text{close}} = \min \left\{ 1, \frac{\rho_0(r_H, r_T; \tau_{HT})}{CK N_{\text{bead}} e^{\Delta U - \mu \tau_{HT}}} \right\} ,$$

(311)

where $N_{\text{bead}}$ is the number of beads after closing, and

$$\Delta U = U(\text{new}) - U(\text{old}) = U(\text{closed}) - U(\text{open}) ,$$

hence it’s $-\Delta U$ in used in the opening acceptance.
Close worm, details

The probability of closing a worm may get very small if the random walk distance of Head and Tail is large, i.e., \( \rho_0(r_H, r_T; \tau_{HT}) \) is very small. Such worms are rarely created, but they are also hard to close. In order to avoid these sticky worms Boninsegni et al. used a hard limit, that keeps \( \rho_0(r_H, r_T; \tau_{HT}) \) (actually \( \ln[\rho_0(r_H, r_T; \tau_{HT})] \)) more reasonable: If

\[
- \frac{(r_H - r_T)^2}{4\lambda \tau_{HT}} > X
\]  

(312)

opening and closing are rejected.\(^\text{58}\) This choice is deterministic, but also symmetric, so \( T(new \rightarrow old) = T(old \rightarrow new) \) and no harm is done to stochasticity.

\(^\text{58}\) Boninsegni et al. used \( X = 4 \) in liquid He calculations.
Insert worm

This seeds a new world line strand of length $k$.

Insert

1. Set Tail to a random $r$ from the simulation volume $\Omega$ and random time slice
2. Pick $k \in U[1, K]$ generate a $k$-step path above Tail; Put Head there.
3. Accept with probability

$$P_{\text{insert}} = \min\left\{ 1, CK M\Omega e^{\Delta U - \mu \tau_{HT}} \right\},$$  \hspace{1cm} (314)
The reverse update of insert.

**Remove**

1. If the worm length is $> K$ reject remove, the worm is too long.
2. Accept with probability

$$P_{\text{remove}} = \min \left\{ 1, \frac{e^{\Delta U - \mu T_{HT}}}{CKM\Omega} \right\}, \quad (316)$$
Advance

This moves Head up in imaginary time.

Advance

1. Pick $k \in U[1, K]$
2. Generate a $k$-step path from old Head up in imaginary time and set the new Head there.
3. Accept advance with probability

$$P_{\text{advance}} = \min \left\{ 1, e^{\Delta U + \mu \tau_H} \right\}.$$  (318)
Recede

This moves Head down in imaginary time.

Recede

1. Pick $k \in U[1, K]$. If the worm has less than $k$ beads reject the update.
2. Move Head $k$ steps down in imaginary time, delete beads on the way.
3. Accept with probability

$$P_{\text{recede}} = \min \left\{ 1, e^{\Delta U - \mu^T H T} \right\}, \quad (320)$$
1. Pick $k \in U[1, K]$. Build a list $\{S_i\}$ of beads $k$ steps above Head on another world line.

2. Build a weight list $\{\rho_0(r_H, r_{S_i}, \tau_{HS_i})\}$. Normalization of the list is $\Sigma_S = \sum_i \rho_0(r_H, r_{S_i}, \tau_{HS_i})$.

3. Pick a bead $S$ with a probability of the normalized weight list.

4. Find the bead $T$ that is $k$ steps down in imaginary time from bead $S$. If Tail is met along the way, reject the update.

5. Build another weight list $\{\rho_0(r_S, r_{T_i}, \tau_{IT_i})\}$. Normalization of the list is $\Sigma_T = \sum_i \rho_0(r_S, r_{T_i}, \tau_{ST_i})$.

6. Generate a free-particle random path from Head to $S$.

7. Accept swap and move Head to $T$ with probability

$$P_{\text{swap}} = \min \left\{ 1, \frac{\Sigma_S}{\Sigma_T} e^{\Delta U} \right\} . \quad (322)$$
Swap details

The first weight list and sampling bead $S$ from it favors swap attempts to beads that are not too far in a random walk from Head. This is $T(\text{old} \rightarrow \text{new})$. Another weight list is needed because the swap is its own reverse, so we must compute also $T(\text{new} \rightarrow \text{old})$. 
Head/tail wiggle
Head/tail wiggle
Head/tail wiggle

Wiggling the head (Head): In free-particle terms the red coordinates change,

\[ \cdots \rho_0(r_{H-k-2}, r_{H-k-1}; \tau) \rho_0(r_{H-k-1}, r_{H-k}; \tau) \cdots \rho_0(r_{H-2}, r_{H-1}; \tau) \rho_0(r_{H-1}, r_{H}; \tau). \]

Here \( r_{H-k-1} \) is fixed, and the new path segment can be sampled exactly, point by point:

\[
\begin{align*}
  r_{H-k-1}' &= r_{H-k-1} + \sqrt{2\lambda \tau} \eta, \quad \text{samples exactly } \rho_0(r_{H-k-1}, r_{H-k}'; \tau) \\
  r_{H-k+1}' &= r_{H-k}' + \sqrt{2\lambda \tau} \eta, \quad \text{samples exactly } \rho_0(r_{H-k}', r_{H-k+1}'; \tau) \\
  \cdots \\
  r_H' &= r_{H-1}' + \sqrt{2\lambda \tau} \eta, \quad \text{samples exactly } \rho_0(r_{H-1}', r_H'; \tau). \quad (323)
\end{align*}
\]

where green coordinates are new. The move is accepted/rejected based on the potential change in the wiggled time slices. It’s a good idea to add a ”hard limit”, and reject all wiggles that take Head too far from Tail, \( \rho_0(r'_H, r_T, \tau_{HT}) < X \), so that the wiggled worm can close.
Another way to sample the new path is to first sample new Head from the fixed point $\mathbf{r}_{l-k}$, which is a long time jump $\tau_H - \tau_k = k\tau$ from Head,

$$r'_H = r_k + \sqrt{2\lambda k\tau} \eta$$, samples exactly $\rho_0(r_k, r'_H; k\tau)$, \hspace{1cm} (324)

and fill the path between the new Head and the fixed point using staging. Be careful to use the minimum image convention properly, so that you don’t sample the new path across the whole system. The wiggle move is its own inverse.
Some PIMC researchers

Just to give you a few pointers to some useful names. In no particular order:

- D. M. Ceperley, E. L. Pollock, B. Militzer, H. Kleinert
- T. Dornheim, V. S. Filinov, M. Bonitz, V. E. Fortov, T. Schoof, S. Groth
- M. Boninsegni, N. Prokof’ev, B. Svistunov, S. A. Chin, W. Janke, T. Sauer
- D. E. Galli, C.H. Mak, S. Zakharov, M. F. Herman, E. J. Bruskin, B. J. Berne
- Finland (TUT): T. T. Rantala and his students (I. Kylänpää, M. Leino, J. Tiihonen etc.)

And, of course, R. P. Feynman.
Fermion sign problem: the Nemesis of QMC.

The expectation value of some quantity $\hat{A}$ is

$$\langle \hat{A} \rangle = \frac{\text{Tr}[\hat{A} \hat{\rho}]}{\text{Tr}[\hat{\rho}]} ,$$

which is evaluated in MC using weights $w_i$,

$$\langle A \rangle_w = \frac{\sum_i w_i A_i}{\sum w_i} ,$$

but the weights $w_i$ may be negative, unsuitable for MC. Separate the sign as $w_i = \text{sgn}(w_i)|w_i|$, 

$$\langle A \rangle_w = \frac{\sum_i w_i A_i}{\sum w_i} = \frac{\sum_i \text{sgn}(w_i)|w_i|A_i}{\sum_i \text{sgn}(w_i)|w_i|} = \frac{\sum_i |w_i|\text{sgn}(w_i)A_i}{\sum_i |w_i|} \cdot$$

The factor

$$p_i := \frac{|w_i|}{\sum_i |w_i|} ,$$

is a good probability, so we should just evaluate two averages. Easy, ha?
Fermion sign problem: the Nemesis of QMC.

Using the probabilities \( p_i \) made of absolute values we get

\[
\langle A \rangle_w = \frac{\sum_i p_i \text{sgn}(w_i)A_i}{\sum_i p_i \text{sgn}(w_i)} = \frac{\langle \text{sign}A \rangle_{|w|}}{\langle \text{sign} \rangle_{|w|}}.
\] (329)

In principle, this can be easily done in MC, but there’s a catch. The denominator is related to two partition functions, one with weights \( w_i \), and another one with weights \( |w_i| \),

\[
\langle \text{sign} \rangle_{|w|} = \sum_i p_i \text{sgn}(w_i) = \frac{\sum_i w_i}{\sum_i |w_i|} = \frac{Z_w}{Z_{|w|}} = \frac{e^{-\beta F_w}}{e^{-\beta F_{|w|}}} = e^{-\beta(F_w-F_{|w|})},
\] (330)

where the free energy is \( F = -\frac{1}{\beta} \ln Z \). As always in MC, the error of the mean diminishes as \( 1/\sqrt{N_{MC}} \) with \( N_{MC} \) samples, but the pre-factor grows,

\[
\frac{\Delta(\langle \text{sign} \rangle_{|w|})}{\langle \text{sign} \rangle_{|w|}} = \frac{1}{\sqrt{N_{MC}}} \sqrt{\langle \text{sign}^2 \rangle_{|w|} - \langle \text{sign} \rangle_{|w|}^2} \sim \frac{e^{\beta(F_w-F_{|w|})}}{\sqrt{N_{MC}}} = \frac{e^{\beta N(f_w-f_{|w|})}}{\sqrt{N_{MC}}}.
\] (331)
Fermion sign problem: the Nemesis of QMC.

The error behaves like

$$\frac{\Delta \langle \text{sign} \rangle_{|w|}}{\langle \text{sign} \rangle_{|w|}} \sim \frac{e^{\beta N(f|w|-f_{|w|})}}{\sqrt{N_{MC}}} ,$$

where $f$’s are free energies per particle – remember $F$ is extensive. The difference $f_{|w|} - f_{|w|}$ is positive, because the boson energy with weights $|w|$ is below the fermion energy with weights $w$. The factor grows exponentially both with decreasing temperature, and with increasing number of fermions $N$. This is the fermion sign problem.

$\Rightarrow$ Can’t sum up signed quantities, so guess signs, that is, nodes

- FN-DMC: use fixed nodes (a guess made at $T = 0$).
- FP-DMC: use fixed phase\(^{59}\)
- PIMC: (nodes at $T = 0$) $\neq$ (nodes at finite-$T$). Instead, the density matrix nodes depend on the imaginary time:

$$\rho(R', R_0; t) = 0 \text{ with fixed (reference point) } R_0 \text{ and } t .$$

Fermion sign problem: strategies

- **Restricted PIMC (RPIMC)**: Force nodes via a trial density matrix and keep paths from crossing to another sign
  - determinant of free-particle density matrices – accurate at high $T$
  - variational trial density matrix (Militzer and Ceperley)
  - Jellium: RPIMC does not work at high density ($r_s < 1$) \(^{60}\)

- **Direct PIMC (DPIMC)** Accurate, but extremely costly. \(^{61}\)

- **Configuration PIMC (CPIMC)** \(^{62}\)
  - no timestep error: continuous time QMC
  - excellent at high density

- **Permutation Blocking PIMC (PB-PIMC)** \(^{63}\)

---


\(^{61}\) See, for example, T. Dornheim, J. Phys. A: Math. Theor. 54, 335001 (2021) [link to open access article].


Fermion sign problem: strategies

Dornheim et al.\textsuperscript{64} computed the structural properties and collective excitations of liquid $^3$He without fixed nodes using the direct PIMC method.\textsuperscript{65} Therefore, our simulations are free of uncontrolled approximations, but computationally extremely costly when the temperature is decreased.

For 38 atoms (or should one dare say, for only 38 atoms) ...which means we need 1000 times as much computer time for the fermionic calculations as we do for the bosonic case without a sign problem. While this is still feasible with $\mathcal{O}(10^5)$ CPU hours, it constitutes the limit of the present investigation.

Experiments on $^3$He are not very popular, a liter costs about $1000 and is on short supply.

\textsuperscript{64} T. Dornheim, Z. A. Moldabekov, J. Vorberger, and B. Militzer, Nature 12, 1-12 (2022) link to open access article.

\textsuperscript{65} A follow-up to uniform electron gas work by T. Dornheim, J. Vorberger, B. Militzer, and Z. A. Moldabekov, PRE 104, 055206 (2021).
Fermion sign problem; perspectives

Troyer and Wiese \textsuperscript{66}

Quantum Monte Carlo simulations, while being efficient for bosons, suffer from the “negative sign problem” when applied to fermions—causing an exponential increase of the computing time with the number of particles. A polynomial time solution to the sign problem is highly desired since it would provide an unbiased and numerically exact method to simulate correlated quantum systems. Here we show that such a solution is almost certainly unattainable by proving that the sign problem is nondeterministic polynomial (NP) hard, implying that a generic solution of the sign problem would also solve all problems in the complexity class NP in polynomial time.

Fermion sign problem; perspectives

Troyer and Wiese study the NP hard problem of determining whether there exists a state with energy $E \leq E_0$, for a fixed bound $E_0$, of the Ising Hamilton function

$$H = - \sum_{\langle j, k \rangle} J_{jk} \sigma_j \sigma_k .$$

(334)

What makes this NP hard is the evaluation of the sum over all configurations $c$,

$$Z = \sum_c e^{-\beta E(c)} .$$

(335)

The quantum version has $\sigma_j$ and $\sigma_k$ replaced by Pauli spin matrices $\sigma_j^x$ and $\sigma_k^x$. The bosons problem is easy, it’s a ferromagnet with all couplings $J_{jk} \geq 0$. The fermion problem has a "random" sign, which can be mapped on random off-diagonal signs in $J$, so Troyer and Wiese conclude that the sign problem is the origin of the NP hardness. So if NP $\neq$ P (solvable in polynomial time): tough luck!
Fermion sign problem; perspectives

Are we out of luck? If we could solve the eigenstates and eigenvalues of \( \hat{H}, \hat{H}|i\rangle = E_i|i\rangle \), we could calculate finite temperature expectation values from the normalized density operator \( \hat{\rho} = \frac{1}{Z} e^{-\beta \hat{H}} \),

\[
\langle \hat{A} \rangle = \text{Tr}(\hat{\rho} \hat{A}) = \frac{\sum_i e^{-\beta E_i} \langle i|\hat{A}|i\rangle}{\sum_i e^{-\beta E_i}}.
\] (336)

Here the weights \( e^{-\beta E_i} \) are positive, so there is no sign problem. Alas, solving \( \hat{H}|i\rangle = E_i|i\rangle \) is an exponentially hard problem.

A question one should next ask is, just how accurately do we need to know the solution to \( \hat{H}|i\rangle = E_i|i\rangle \) in order to avoid the sign problem? Is it enough to have the approximate solution hidden in an evolution algorithm?
Here we argue that: (1) The sign problem is not intrinsic to solving a fermion problem; (2) it is only a consequence of a poor approximation to the exact propagator; and (3) it can be automatically minimized by using better, higher-order approximate propagators. By following up on the last point, this work shows that by using optimized fourth-order propagators, accurate results can be obtained for up to 20 spin-polarized electrons in a two-dimensional (2D) circular, parabolic quantum dot.

Chin argues that the "sign problem" is due to the large number of antisymmetric free-fermion propagators that are needed to extract the ground state wave function at large imaginary time.

**Goal:** Get to a large $\tau$ with less $e^{-\tau \hat{T}}$'s, by using e.g. Eq. (146):

$$\hat{G}_{\text{Chin}}(\tau) = e^{-v_0 \tau \hat{V}} e^{-t_1 \tau \hat{T}} e^{-v_1 \tau \hat{W}} e^{-t_2 \tau \hat{T}} e^{-v_1 \tau \hat{W}} e^{-t_1 \tau \hat{T}} e^{-v_0 \tau \hat{V}}. \quad (337)$$

---

Fermion sign problem; perspectives

Let’s see how the antisymmetric free-fermion propagator looks like. Define, as usual

\[ G_{\hat{T}}(\mathbf{R}', \mathbf{R}; \tau) = \langle \mathbf{R}' | e^{-\tau \hat{T}} | \mathbf{R} \rangle . \]  

(338)

Antisymmetry comes from the spectral expansion in eigenstates of \( \hat{T} \) with antisymmetric wavefunctions, explicit either as a permutation sum

\[ \phi_K(\mathbf{R}) = \sum_P (-1)^P \phi_{\text{boltzmann}}(\hat{P} \mathbf{R}) \]

or as a determinant. We know the boltzmann result, just collect them to a determinant:

\[ G_{\hat{T}}(\mathbf{R}', \mathbf{R}; \tau) = (4\pi \lambda \tau)^{-3N/2} \det M , \]  

(339)

where the matrix \( M \) has elements

\[ M_{ij} = e^{-\frac{(r_i - r'_j)^2}{4\lambda \tau}} . \]  

(340)

The almost random signs originate from \( \det M \).
Fermion sign problem; perspectives

The idea is to get the result before the sign problem appears

FIG. 4. (Color online) Convergence of the Hamiltonian energy for \( N = 20 \) polarized electrons using the optimized, fourth-order three- and four-bead propagators. Error bars are computed from 200 to 300 block average of \( 5 \times 10^4 \) configurations of all 20 electrons.
Alternatives with no timestep error: SSE

There are ways to do QMC without any timestep error from splitting $e^{-\beta \hat{H}}$.

One is the *Stochastic Series Expansion (SSE)*. Invented by Anders Sandvik and J. Kurkijärvi\textsuperscript{68}, as an extension to Handscomb’s method. Make a series expansion

$$Z = \text{Tr}[e^{-\beta \hat{H}}] = \sum_{\alpha} \sum_{n=0}^{\infty} \frac{(-\beta)^n}{n!} \langle \alpha | \hat{H}^n | \alpha \rangle$$  \hspace{1cm} (341)

See O. F. Syljåsen, A. W. Sandvik, Phys. Rev. E 66, 046701 (2002) for *SSE with directed loop update*. SSE has been mostly used in lattice spin systems, such as Ising and Hubbard models.

SSE with the Hubbard model

The Hubbard Hamiltonian with nearest neighbour jumps $t$, on-site interaction $U$, and chemical potential $\mu$ is

$$\hat{H} = \sum_{\langle ij \rangle} \left[-t(a_i^\dagger a_j + a_j^\dagger a_i)\right] + \sum_i \left[-\mu \hat{n}_i + \frac{U}{2} \hat{n}_i(\hat{n}_i - 1)\right]$$  \hspace{1cm} (342)

*Figure:* Standing waves in crossing laser fields create an optical potential, where bosons arrange in a lattice and obey the Hubbard Hamiltonian.  \(^{70}\)

---

\(^{69}\) Figure: Krissia de Zawadzka, how-to-draw-an-optical-lattice @stackoverflow

\(^{70}\) Figure: Krissia de Zawadzka, how-to-draw-an-optical-lattice @stackoverflow
SSE with the Hubbard model

Hamiltonian as a sum over bond operators (coordination number $Z$),

$$\hat{H} = - \sum_i \hat{K}_i$$

$$= - \sum_{\langle ij \rangle} \left[ t(a_i^\dagger a_j + a_j^\dagger a_i) + \frac{\mu}{Z}(\hat{n}_i + \hat{n}_j) - \frac{U}{2Z}[\hat{n}_i(\hat{n}_i - 1) + \hat{n}_j(\hat{n}_j - 1)] \right],$$

and sample a random product of $\hat{H}$ terms in the partition function,

$$Z = \sum_s \sum_s \sum_{n=0}^{\beta} \frac{\beta^n}{n!} \sum_{i_1,\ldots,i_n} \langle s | \hat{K}_{i_1} \ldots \hat{K}_{i_n} | s \rangle$$
SSE Diagonal update

- The number of operators is controlled by the temperature.
- Only diagonal operators can be inserted/removed locally.
- It’s easier to deal with a constant $M$ operators: use $I$ for filling.

Example: 8 sites

```
12324211 12324211
12324211 12324211
```

- Identity operator
- Diagonal operator

**Site occupations**
SSE Diagonal update

- **Start**: Pick any reasonable state $|s\rangle$ and insert $M$ identity operators $I$:

  
  \[
  \ldots \quad 12314152 \quad \ldots \quad 12314152 \quad \ldots \quad 12314152 \quad \ldots
  \]

- **Diagonal update**: Change identity ops to diagonal ops or vice versa with acceptance probabilities

  \[
  A_{\text{ins}}(I \rightarrow \hat{K}_d) = \min \left\{ \frac{\langle s|\hat{K}_d|s\rangle \beta N_{\text{bonds}}}{M - n}, 1 \right\} \quad (345)
  \]

  \[
  A_{\text{rem}}(\hat{K}_d \rightarrow I) = \min \left\{ \frac{M - n + 1}{\langle s|\hat{K}_d|s\rangle \beta N_{\text{bonds}}}, 1 \right\} \quad (346)
  \]
SSE Diagonal update

Derivation: Use the detailed balance condition for $I \rightarrow \hat{K}_d$ and $\hat{K}_d \rightarrow I$:

$$I : \quad M - n + 1 \leftrightarrow M - n$$

$$\hat{K}_d : \quad n - 1 \leftrightarrow n$$

$$\frac{P(S \rightarrow S')}{{P(S' \rightarrow S)}} = \frac{W(S')}{{W(S)}} = \frac{\langle s' | \beta \hat{K}_d | s' \rangle}{\langle s | I | s \rangle} = \beta \frac{\langle s' | \hat{K}_d | s' \rangle}{\langle s | I | s \rangle} = \frac{T(S \rightarrow S') A_{\text{ins}}(I \rightarrow \hat{K}_d)}{T(S' \rightarrow S) A_{\text{rem}}(\hat{K}_d \rightarrow I)}, \quad (347)$$

where

$$T(S \rightarrow S') = \text{Prob. of picking a } I = \frac{M - n + 1}{N_{\text{bonds}}} \quad (348)$$

$$T(S' \rightarrow S) = \text{Prob. of picking the } \hat{K}_d = 1, \quad (349)$$

so

$$\frac{A_{\text{ins}}(I \rightarrow \hat{K}_d)}{A_{\text{rem}}(\hat{K}_d \rightarrow I)} = \frac{\langle s | \hat{K}_d | s \rangle \beta N_{\text{bonds}}}{M - n + 1}, \quad (350)$$

and use the Metropolis solution. Here operations are exact reverses, insert is $n - 1 \rightarrow n$ and remove is $n \rightarrow n - 1$, but in practice $n$ is the current number of $\hat{K}_d$’s, so insert is, instead, $n \rightarrow n + 1$:

$\rightarrow$ Insert acceptance has $M - n$, not $M - n + 1$!
SSE Directed loop update

- Task: find an effective update, such as boson add/remove loop update \((+1)/(-1)\)
- Update changes the type of an operator, but not their total number
- The *directed loop update*\(^{71}\) is a global, worm update.

---

SSE Directed loop update

- Loop update can add/remove bosons ⇒ Grand Canonical Ensemble
- The end state in $Z$ is sampled automatically!

Example: (2 sites)

2 bosons
(Density = 1)

3 bosons
(Density = 1.5)

Knowing the in-leg, how to pick the out-leg so that the $Z$ is sampled?
SSE Directed loop update

Again, it’s up to detailed balance to choose the out-leg. The weights are matrix elements,

$$W(S) = \langle s_1 | \hat{K} | s_2 \rangle, \quad W(S') = \langle s'_1 | \hat{K}' | s'_2 \rangle$$

and create probability tables, coded $P(update, in\text{−}leg, out\text{−}leg)$, such as

$$P(S \rightarrow S') = P(+1; i, j), \quad P(S \rightarrow S') = P(-1; i, j) \text{ or } P(+1; j, i).$$

One possible normalization is the ”heat bath”,

$$P(S \rightarrow S') = \frac{W(S')}{\sum_S W(S)},$$

which assumes positive weights.
SSE Directed loop update

- Loop update may "bounce", up-going adds a boson, down-going removes one.

- Loop continues until it reaches the starting point (worm closes). This usually happens soon enough.
SSE Forcing positive weights

Add a constant $C$ to the Hamiltonian,

$$\hat{H} = -\sum_{\langle ij \rangle} t(a_i^\dagger a_j + a_j^\dagger a_i) + C + \frac{\mu}{Z}(\hat{n}_i + \hat{n}_j) - \frac{U}{2Z}[\hat{n}_i(\hat{n}_i - 1) + \hat{n}_j(\hat{n}_j - 1)]$$

- $\langle s|\text{off-diagonal}|s' \rangle \geq 0$, OK
- $\langle s|\text{diagonal}|s \rangle \geq 0$, if we choose a maximum allowed occupation $n_{max}$ and a large enough $C$ to match it.
- If $n_{max}$ turns out to be too low, start anew with a larger one.
SSE some observables

- Energy, thermodynamical estimator

\[
\langle E \rangle = -\frac{1}{Z} \frac{\partial Z}{\partial \beta}
\]

\[
= -\frac{1}{\beta} \sum_s \sum_{n=0}^{\infty} \frac{n \beta^n}{n!} \sum_{i_1, \ldots, i_n} \langle s | \hat{K}_{i_1} \cdots \hat{K}_{i_n} | s \rangle = -\frac{\langle n \rangle}{\beta},
\]

where \( \langle n \rangle \) is the average number of operators (vertices).

- The ground state energy is a finite, fixed number, so \( \langle n \rangle \) increases with \( \beta \Rightarrow \) low- \( T \) simulations have more terms and they are slower.

- Energy fluctuates around \( \langle E \rangle \), so there exists a finite number of terms \( M \) that is reached during any finite time simulation.

- Density and compressibility are trivial to measure.

- Superfluid density \( \rho_s \) is again measured via the winding number, the net amount of hopping across a boundary.
SSE Bose results

Figure: The average particle number per site in a 2D $8 \times 8$ lattice, at a very low temperature, as a function of the hopping parameter $t$ and the chemical potential $\mu$. The numbered plateaus are Mott insulating phases with the indicated particle number per site. Error bars are smaller than the line width.$^{73}$


$^{73}$V. Apaja (ca. 2004, unpublished).
SSE Bose results

Figure: Density $\rho$, compressibility $\kappa$ and the superfluid density $\rho_s$ in a 2D $32 \times 32$ lattice, as a function of the chemical potential $\mu$. In Mott insulating phases the density is constant, while compressibility and the superfluid density are zero.  

---

$^{75}$O. Syljuåsen, PRE 67, 046701 (2003).
Alternatives with no timestep error: Continuous time QMC

Continuous time QMC has multiple variants. For example, solve $\hat{U}(\beta)$ from

$$e^{-\beta \hat{H}} = e^{-\beta \hat{T}} \hat{U}(\beta) \Leftrightarrow \hat{U}(\beta) = e^{\beta \hat{T}} e^{-\beta \hat{H}} \quad (355)$$

$$\Rightarrow \frac{\partial \hat{U}(\beta)}{\partial \beta} = \hat{T} e^{\beta \hat{T}} e^{-\beta \hat{H}} - e^{\beta \hat{T}} \hat{H} e^{-\beta \hat{H}}$$

$$= \hat{T} e^{\beta \hat{T}} e^{-\beta \hat{H}} - e^{\beta \hat{T}} (\hat{T} + \hat{V}) e^{-\beta \hat{H}}$$

$$= -e^{\beta \hat{T}} \hat{V} e^{-\beta \hat{H}} = -e^{\beta \hat{T}} \hat{V} e^{-\beta \hat{T}} [e^{\beta \hat{T}} e^{-\beta \hat{H}}]$$

$$= -e^{\beta \hat{T}} \hat{V} e^{-\beta \hat{T}} \hat{U}(\beta) := -\hat{V}(\beta) \hat{U}(\beta), \quad (356)$$

and integrate, 

$$\int_0^\beta \frac{\partial \hat{U}(t)}{\partial t} \, dt = \hat{U}(\beta) - \hat{U}(0) = - \int_0^\beta dt \hat{V}(t) \hat{U}(t) \quad (357)$$
Alternatives with no timestep error: Continuous time QMC

Insert and iterate, $\hat{T}_\tau$ is the imaginary-time ordering operator,

\[
e^{-\beta \hat{H}} = e^{-\beta \hat{T} \hat{U}(\beta)} = e^{-\beta \hat{T}} \left[ 1 - \int_0^\beta d\tau \hat{V}(\tau) \hat{U}(\tau) \right] \tag{358}
\]

\[
e^{-\beta \hat{T}} \left[ 1 - \int_0^\beta d\tau_1 \hat{V}(\tau_1) + \int_0^\beta d\tau_1 \int_0^{\tau_1} d\tau_2 \hat{V}(\tau_1) \hat{V}(\tau_2) \ldots \right]
\]

\[
e^{-\beta \hat{T}} \left[ 1 - \int_0^\beta d\tau_1 \hat{V}(\tau_1) + \frac{1}{2} \int_0^\beta d\tau_1 \int_0^\beta d\tau_2 \hat{T}_\tau [\hat{V}(\tau_1) \hat{V}(\tau_2)] \ldots \right].
\]

This is often written in the form\(^{76}\)

\[
e^{-\beta \hat{H}} = e^{-\beta \hat{T} \hat{T} \hat{T} \hat{T}} \left[ e^{-\int_0^\beta d\tau \hat{V}(\tau)} \right], \quad \hat{V}(\tau) := e^{\tau \hat{T} \hat{V} e^{-\tau \hat{T}}} . \tag{359}
\]

This can be viewed as the interaction picture representation wrt. $\hat{V}$, and the expanded form does perturbation theory wrt. $\hat{V}$.

\(^{76}\)Notice the similarity to the Feynman-Kac formula (278).
Alternatives with no timestep error: Continuous time QMC

Explicitly,

\[ e^{-\beta \hat{H}} = e^{-\beta \hat{T}} - \int_0^\beta d\tau_1 e^{-(\beta - \tau_1) \hat{T}} \hat{V} e^{-\tau_1 \hat{T}} \]

\[ + \frac{1}{2} \int_0^\beta d\tau_1 d\tau_2 e^{-(\beta - \tau_1) \hat{T}} \hat{V} e^{(\tau_1 - \tau_2) \hat{T}} \hat{V} e^{-\tau_2 \hat{T}} \ldots . \] (360)

This gives the canonical partition function, \( \rho_0(R', R; \tau) := \langle R' | e^{-\tau \hat{T}} | R \rangle, \)

\[ Z = \int dR \langle R | e^{-\beta \hat{H}} | R \rangle = \int dR \rho_0(R, R; \beta) \]

\[ - \int_0^\beta d\tau_1 \int dR dR_1 \rho_0(R, R_1; \beta - \tau_1) V(R_1) \rho_0(R_1, R; \tau_1) \]

\[ + \frac{1}{2} \int_0^\beta d\tau_1 d\tau_2 \int dR dR_1 dR_2 \hat{V} \left[ \rho_0(R, R_1; \beta - \tau_1) V(R_1) \rho_0(R_1, R; \tau_1) \rho_0(R_1, R_2; \tau_1 - \tau_2) V(R_2) \rho_0(R_2, R; \tau_2) \right] \ldots . \] (361)
Alternatives with no timestep error: Continuous time QMC

- This $Z$ is made of DVD-kinks, free diffusion - potential - free diffusion. Free diffusion combined with a singular potential is always bad.

- Terms are positive definite only for an attractive potential – again the Hydrogen atom problem works perfectly. MC sampling of $Z$ could be done by avoiding positive potential regions using \textit{domain Green's functions} \cite{Schmidt2005}. Without any tricks the large $|V(R)|$ occasions overwhelm the MC sampling.

Eq. (361) uses $\{R\}$ basis, where $\hat{V}$ is diagonal. You could as well choose an occupation number basis, and divide $\hat{H}$ to diagonal ($\hat{D}$) or off-diagonal ($\hat{Y}$) contributions in this basis. Next, in Eq. (361), replace $\hat{T} \rightarrow \hat{D}$ and $\hat{V} \rightarrow \hat{Y}$. This leads to CPIMC formulation by Schoof \textit{et al.} as in Eqs. (92)—(98) in \textit{Introduction to Configuration Path Integral Monte Carlo}, discussed on the next slide.

\footnote{Schmidt \textit{et al.}, Phys. Rev. E 71, 016707 (2005).}
In occupation number basis $|n_0, n_1, ...⟩ := \{|\{n\}\rangle$ we have

$$\sum_{\{n_i\}} |\{n_i\}\rangle\langle \{n_i\}| = I \quad , \quad Z = \sum_{\{n\}} \langle\{n\}|e^{-\beta\hat{H}}|\{n\}\rangle , \quad (362)$$

and the perturbation series for $\hat{H} = \hat{D} + \hat{Y}$, using latter as perturbation, is

$$Z = \sum_{\{n\}} \langle\{n\}|e^{-\beta\hat{D}}|\{n\}\rangle$$

$$- \int_0^\beta d\tau_1 \sum_{\{n\}} \langle\{n\}|e^{-(\beta-\tau_1)\hat{D}}|\{n\}\rangle \langle\{n\}|\hat{Y}|\{n\}\rangle \langle\{n\}|e^{-\tau_1\hat{D}}|\{n\}\rangle$$

$$+ \int_0^\beta d\tau_1 \int_0^{\tau_1} d\tau_2 \sum_{\{n\}} \sum_{\{n_1\}} \left[ \langle\{n\}|e^{-(\beta-\tau_1)\hat{D}}|\{n\}\rangle \langle\{n\}|\hat{Y}|\{n_1\}\rangle \langle\{n_1\}|e^{-(\tau_1-\tau_2)\hat{D}}|\{n_1\}\rangle \langle\{n_1\}|\hat{Y}|\{n\}\rangle \langle\{n\}|e^{-\tau_2\hat{D}}|\{n\}\rangle \right] \ldots .$$

I left out the time ordering operator and factorials, and used explicit ordering in time integrals.
Configuration Path Integral Monte Carlo

Using $n$ to mark occupation numbers $\{n\}$ and marking matrix elements shortly as

\[
\langle \{n\} | e^{-\beta \hat{D}} | \{n\} \rangle := e^{-\beta D(n)} 
\]  \hspace{1cm} (364)

\[
\langle \{n_1\} | \hat{Y} | \{n_1\} \rangle := Y(n_1, n_2) , 
\]  \hspace{1cm} (365)

one gets (usually $Y(n, n) = 0$)

\[
Z = \sum_n e^{-\beta D(n)} - \int_0^\beta d\tau_1 \sum_n e^{-\beta D(n)} Y(n, n) 
\]

\[
+ \int_0^\beta d\tau_1 \int_0^{\tau_1} d\tau_2 \sum_{n,n_1} 
\left[ e^{-(\beta-\tau_1)D(n)} Y(n, n_1) e^{-(\tau_1-\tau_2)D(n_1)} Y(n_1, n) e^{-\tau_2 D(n)} \right] ... . 
\]  \hspace{1cm} (366)

Fermion statistics is built-in in second quantization, but the sign problem prevails, because weights are not as such positive definite.
Things that were not done and things to do

I didn’t talk about

▶ Path integrals in quantum field theory (Dirac and Majorana fermions etc.), just path integrals in condensed matter physics. If you want to delve into field theoretical aspects, a good place to start is Edward Witten’s *Fermion Path Integrals And Topological Phases*. A youtube video is also available here.

▶ Feynman path integrals (see lecture notes, although they are lagging behind these slides)

▶ Fermion PIMC in detail; I recommend Burkhard Militzer’s excellent presentations on the topic

▶ Ground state path integrals

Things you should do

▶ find a way around the sign problem