Theory & Practice of Diffusion Quantum Monte Carlo

Neil D. Drummond

Department of Physics, Lancaster University

QMC and the CASINO program, TTI, Vallico Sotto, Italy

Tuesday 5th August, 2014

Introduction

- DMC is the most accurate first-principles total-energy method for systems with more than a few tens of electrons.

- VMC is generally used only as a preliminary to DMC studies.

- In this lecture I will (i) explain the theory behind the DMC method and (ii) give some practical advice on how to run successful DMC calculations.

- You do **not** need to understand all the details in order to **use** DMC.

- You **do** need to be aware of issues such as time-step bias in order to carry out meaningful DMC work.

Imaginary-Time Schrödinger Equation (I)

- *Imaginary-time Schrödinger equation* (ITSE):

$$\left[\hat{H} - E_T\right]\Phi = -\frac{1}{2}\nabla^2\Phi + U\Phi - E_T\Phi = -\frac{\partial\Phi}{\partial t},$$

  where $\Phi(\mathbf{R}, t)$ is a function of configuration $\mathbf{R}$ and imaginary time t, $U(\mathbf{R})$ is potential energy and E_T is a *reference energy*.

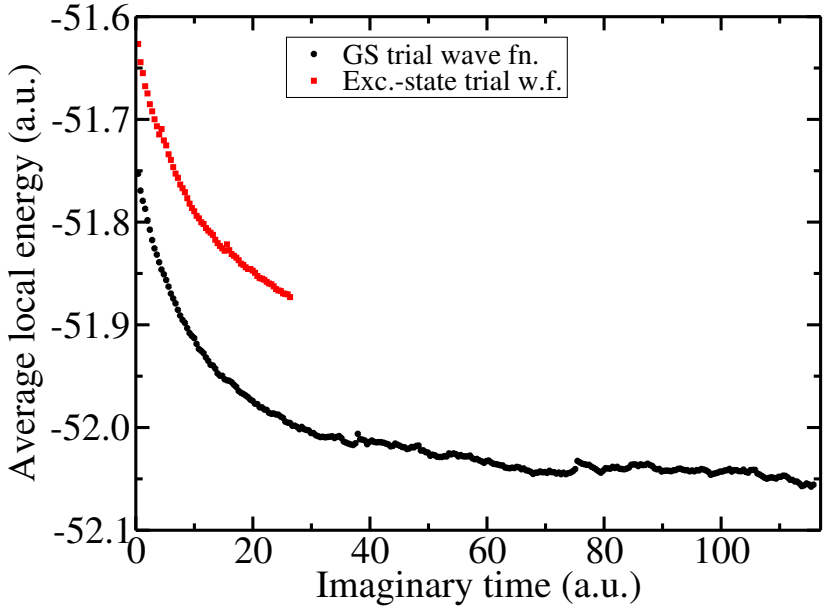
- Time-dependence of ITSE is separable. Can write

$$\Phi = \sum_{n=0}^{\infty} c_n \phi_n \exp[-(E_n - E_T)t],$$

  where E_n and $\phi_n(\mathbf{R})$ are the nth eigenvalue and eigenfunction of Hamiltonian $\hat{H}$.

Imaginary-Time Schrödinger Equation (II)

- Excited states die away exponentially compared with the ground state.
- If $E_T = E_0$ and initial conditions have $c_0 \neq 0$ then, in the limit $t \to \infty$, Φ is proportional to ϕ_0.
- The ground-state component of Φ is "projected out".
- This is true for any reasonable boundary conditions on Φ: see later.



Decay of the average local energy of a set of configurations whose dynamics is governed by the imaginary-time Schrödinger equation.

Importance-Sampling Transformation

- Suppose we have a trial wave function $\Psi(\mathbf{R})$. Let $f = \Phi\Psi$. Then

$$-\frac{1}{2}\nabla^2 f + \nabla \cdot [\mathbf{V} f] + [E_L - E_T] f = -\frac{\partial f}{\partial t},$$

  where

  - $f(\mathbf{R}, t) = \Phi\Psi$ is the *importance-sampled* or *mixed wave function*,
  - $\mathbf{V}(\mathbf{R}) = \Psi^{-1}\nabla\Psi$ is the *drift velocity*,
  - $E_L(\mathbf{R}) = \Psi^{-1}\hat{H}\Psi$ is the *local energy*.

- Proof: substitute $\Phi = \Psi^{-1} f$ into ITSE.

- Consequences of importance sampling:

  1. The term in the ITSE involving potential $U(\mathbf{R})$ is replaced by a term involving the local energy $E_L(\mathbf{R})$, which is relatively uniform. *Makes branching DMC algorithm stable.*
  2. Configurations are distributed according to $f = \Phi\Psi$ rather than Φ. More useful.
  3. *Fixed-node approximation* is introduced (see later).

Importance-Sampled Imaginary Time Schrödinger Equation

- Without the last term on the LHS, the ISITSE is a **Fokker–Planck** equation:

$$-\frac{1}{2}\nabla^2 f + \nabla \cdot [\mathbf{V} f] = -\frac{\partial f}{\partial t}.$$

- This equation describes the time-evolution of the density of a set of "particles" undergoing random diffusion in a $3N$-dimensional fluid of velocity field $\mathbf{V}$.

- Without the first two terms on the LHS, the ISITSE is a **rate** equation:

$$[E_L - E_T]\, f = -\frac{\partial f}{\partial t}.$$

- Gives an exponential increase or decrease in the density of "particles" at each point in the $3N$-dimensional configuration space.

Dirac Notation for Position and Momentum

- Position basis: $|\mathbf{R}\rangle$. NB, $\langle\mathbf{R}'|\mathbf{R}\rangle = \delta(\mathbf{R}-\mathbf{R}')$.
- State vector: $|f\rangle$.
- Spatial wave function: $f(\mathbf{R}) = \langle\mathbf{R}|f\rangle$.
- Position operator is defined such that $\hat{\mathbf{R}}|\mathbf{R}\rangle = \mathbf{R}|\mathbf{R}\rangle$.
- Momentum basis: $|\mathbf{P}\rangle = (2\pi)^{-3N/2}\int \exp(i\mathbf{R}\cdot\mathbf{P})|\mathbf{R}\rangle\, d\mathbf{R}$. NB, $\langle\mathbf{P}'|\mathbf{P}\rangle = \delta(\mathbf{P}-\mathbf{P}')$.
- Momentum operator is defined such that $\hat{\mathbf{P}}|\mathbf{P}\rangle = \mathbf{P}|\mathbf{P}\rangle$.
- $\langle\mathbf{R}|\mathbf{P}\rangle = \exp(i\mathbf{P}\cdot\mathbf{R})/(2\pi)^{3N/2}$. Hence $\nabla_{\mathbf{R}}\langle\mathbf{R}|\mathbf{P}\rangle = i\mathbf{P}\langle\mathbf{R}|\mathbf{P}\rangle = i\langle\mathbf{R}|\hat{\mathbf{P}}|\mathbf{P}\rangle$.
- Completeness: $\int |\mathbf{R}\rangle\langle\mathbf{R}|\, d\mathbf{R} = \int |\mathbf{P}\rangle\langle\mathbf{P}|\, d\mathbf{P} = 1$.

Fokker–Planck Operator

- Let $\mathbf{V}(\hat{\mathbf{R}})|\mathbf{R}\rangle = \mathbf{V}(\mathbf{R})|\mathbf{R}\rangle$ be the drift-velocity operator.
- The first two terms in the ISITSE can be written as

$$\begin{aligned}
-\frac{1}{2}\nabla^2 f(\mathbf{R}) + \nabla\cdot[\mathbf{V}(\mathbf{R})f(\mathbf{R})] &= -\frac{1}{2}\nabla^2\int\langle\mathbf{R}|\mathbf{P}\rangle\langle\mathbf{P}|f\rangle\,d\mathbf{P} \\
&\quad + \nabla\cdot\int\langle\mathbf{R}|\mathbf{P}\rangle\langle\mathbf{P}|\mathbf{V}(\hat{\mathbf{R}})|f\rangle\,d\mathbf{P} \\
&= \frac{1}{2}\int\langle\mathbf{R}|\hat{\mathbf{P}}^2|\mathbf{P}\rangle\langle\mathbf{P}|f\rangle\,d\mathbf{P} \\
&\quad + i\int\langle\mathbf{R}|\hat{\mathbf{P}}|\mathbf{P}\rangle\cdot\langle\mathbf{P}|\mathbf{V}(\hat{\mathbf{R}})|f\rangle\,d\mathbf{P} \\
&= \frac{1}{2}\langle\mathbf{R}|\hat{\mathbf{P}}^2|f\rangle + i\langle\mathbf{R}|[\hat{\mathbf{P}}\cdot\mathbf{V}(\hat{\mathbf{R}})]|f\rangle \\
&= \langle\mathbf{R}|\hat{F}|f\rangle,
\end{aligned}$$

where $\hat{F} = (1/2)\hat{\mathbf{P}}^2 + i\hat{\mathbf{P}}\cdot\mathbf{V}(\hat{\mathbf{R}})$ is the Fokker–Planck operator.

Green's Function of ISITSE

- ISITSE in integral form:

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

  where the Green's function $G(\mathbf{R} \leftarrow \mathbf{R}', t)$ is the solution of the ISITSE satisfying the initial condition $G(\mathbf{R} \leftarrow \mathbf{R}', 0) = \delta(\mathbf{R} - \mathbf{R}')$.

- $[\hat{F} + E_L(\hat{\mathbf{R}}) - E_T] \exp[-t(\hat{F} + E_L(\hat{\mathbf{R}}) - E_T)] = -(\partial/\partial t) \exp[-t(\hat{F} + E_L(\hat{\mathbf{R}}) - E_T)]$.

- So, the Green's function of the ISITSE is

$$\begin{aligned} G(\mathbf{R} \leftarrow \mathbf{R}', \tau) &= \langle \mathbf{R} | e^{-\tau(\hat{F} + E_L(\hat{\mathbf{R}}) - E_T)} | \mathbf{R}' \rangle \\ &\simeq \langle \mathbf{R} | e^{-\tau(E_L(\hat{\mathbf{R}}) - E_T)/2} e^{-\tau \hat{F}} e^{-\tau(E_L(\hat{\mathbf{R}}) - E_T)/2} | \mathbf{R}' \rangle, \\ &= e^{-\tau(E_L(\mathbf{R}) - E_T)/2} \langle \mathbf{R} | e^{-\tau \hat{F}} | \mathbf{R}' \rangle e^{-\tau(E_L(\mathbf{R}') - E_T)/2}. \end{aligned}$$

- The error in the approximation is $\mathcal{O}(\tau^3)$ (proof: Taylor-expand the exponentials).

Normal Ordering

- Normal ordering: in any term of an operator, bring all $\hat{\mathbf{P}}$ operators to the left of all $\hat{\mathbf{R}}$ operators (as if they commuted).

- Notation: $:\hat{A}:$ is the normal-ordered version of $\hat{A}$.

- Note that $:\hat{F}:=\hat{F}$, so

$$:\exp(-\tau\hat{F}):=1-\tau\hat{F}+\mathcal{O}(\tau^2)=\exp(-\tau\hat{F})+\mathcal{O}(\tau^2).$$

Drift–Diffusion Green's Function

$$
\begin{aligned}
\langle \mathbf{R}| : \exp(-\tau \hat{F}) : |\mathbf{R}'\rangle &= \int \langle \mathbf{R}|\mathbf{P}\rangle\langle \mathbf{P}| : \exp[-\tau(\hat{\mathbf{P}}^2/2 + i\hat{\mathbf{P}}\cdot\mathbf{V}(\hat{\mathbf{R}}))] : |\mathbf{R}'\rangle \, d\mathbf{P} \\
&= \int \langle \mathbf{R}|\mathbf{P}\rangle \exp[-\tau(\mathbf{P}^2/2 + i\mathbf{P}\cdot\mathbf{V}(\mathbf{R}'))]\langle \mathbf{P}|\mathbf{R}'\rangle \, d\mathbf{P} \\
&= \int \frac{\exp[i\mathbf{P}\cdot(\mathbf{R}-\mathbf{R}')]}{(2\pi)^{3N}} \exp[-\tau(\mathbf{P}^2/2 + i\mathbf{P}\cdot\mathbf{V}(\mathbf{R}'))] \, d\mathbf{P} \\
&= \frac{1}{(2\pi\tau)^{3N/2}} \exp\left(-\frac{|\mathbf{R}-\mathbf{R}'-\tau\mathbf{V}(\mathbf{R}')|^2}{2\tau}\right).
\end{aligned}
$$

- The last line is the **Langevin** or **drift–diffusion** Green's function, describing diffusion in a fluid of constant velocity field $\mathbf{V}(\mathbf{R}')$.

- *Physically, the approximation of using the normal-ordered Green's function is equivalent to assuming the drift velocity is constant between* $\mathbf{R}$ *and* $\mathbf{R}'$.

- NB, $\langle \mathbf{R}| \exp(-\tau \hat{F}) |\mathbf{R}'\rangle = \langle \mathbf{R}| : \exp(-\tau \hat{F}) : |\mathbf{R}'\rangle + \mathcal{O}(\tau^2)$.

DMC Green's Function

$$
\begin{aligned}
G(\mathbf{R} \leftarrow \mathbf{R}', \tau) &= \langle \mathbf{R} | e^{-\tau \hat{F}} | \mathbf{R}' \rangle \exp\left(-\frac{\tau}{2} [E_L(\mathbf{R}) + E_L(\mathbf{R}') - 2E_T] \right) \\
&= \langle \mathbf{R} | : e^{-\tau \hat{F}} : | \mathbf{R}' \rangle \exp\left(-\frac{\tau}{2} [E_L(\mathbf{R}) + E_L(\mathbf{R}') - 2E_T] \right) + \mathcal{O}(\tau^2) \\
&\approx G_{\mathrm{D}}(\mathbf{R} \leftarrow \mathbf{R}', \tau)\, G_{\mathrm{B}}(\mathbf{R} \leftarrow \mathbf{R}', \tau) \equiv G_{\mathrm{DMC}}(\mathbf{R} \leftarrow \mathbf{R}', \tau).
\end{aligned}
$$

- *Drift–diffusion Green's function* G_D describes the evolution of the density of randomly diffusing "particles" in a $3N$-dimensional fluid of constant velocity field $\mathbf{V}(\mathbf{R}')$. Equivalent to diffusion in velocity field $\mathbf{V}(\mathbf{R})$ if the time step τ is sufficiently small.

- *Branching factor* G_B is the solution of the ISITSE without the first two terms on the LHS; G_B represents exponential growth/decay in density of "particles" at each point in config. space.

- The DMC Green's function therefore describes the evolution of the density of a set of "particles" drifting, diffusing and breeding or dying in a $3N$-dimensional space.

Propagation Over a Macroscopic Length of Imaginary Time

Green's function for a macroscopic length of imaginary time $M\tau$:

$$
\begin{aligned}
G(\mathbf{R} \leftarrow \mathbf{R}', M\tau) &= \langle \mathbf{R} | e^{-M\tau(\hat{F}+\hat{E}_L-E_T)} | \mathbf{R}' \rangle \\
&= \int \cdots \int \langle \mathbf{R} | e^{-\tau(\hat{F}+\hat{E}_L-E_T)} | \mathbf{R}'' \rangle \\
&\qquad \cdots \langle \mathbf{R}''' | e^{-\tau(\hat{F}+\hat{E}_L-E_T)} | \mathbf{R}' \rangle \, d\mathbf{R}'' \ldots d\mathbf{R}''' + \mathcal{O}(M\tau^3) \\
&= \int \cdots \int G(\mathbf{R} \leftarrow \mathbf{R}'', \tau) \\
&\qquad \cdots G(\mathbf{R}''' \leftarrow \mathbf{R}', \tau) \, d\mathbf{R}'' \ldots d\mathbf{R}''' + \mathcal{O}(M\tau^3) \\
&= \int \cdots \int G_{\mathrm{DMC}}(\mathbf{R} \leftarrow \mathbf{R}'', \tau) \\
&\qquad \cdots G_{\mathrm{DMC}}(\mathbf{R}''' \leftarrow \mathbf{R}', \tau) \, d\mathbf{R}'' \ldots d\mathbf{R}''' + \mathcal{O}(M\tau^2).
\end{aligned}
$$

The approximation to the Green's function over a finite interval can be made arbitrarily accurate by dividing the interval into sufficiently small slices of imaginary time.

Time-Step Error in Mixed Wave Function

- Use the DMC Green's function to propagate f to large imaginary time [where $f(\mathbf{R},t)=\phi_0(\mathbf{R})\Psi(\mathbf{R})$] using a finite, small time step τ:

$$f(\mathbf{R},t)=\int G(\mathbf{R}\leftarrow\mathbf{R}',t)f(\mathbf{R}',0)\,d\mathbf{R}',$$

  where the propagation is carried out in numerous short steps using G_{DMC}.

- The use of a finite time step continually introduces errors, even as the evolution in time projects out the ground-state component.

- Let $f=\phi_0\Psi+\Delta$.
  - Error in G_{DMC} per time step is $\mathcal{O}(\tau^2)$, so error Δ in f is introduced at rate $\mathcal{O}(\tau)$.
  - Error is removed at a rate that is roughly $-\Delta/T_{\mathrm{corr}}$, where T_{corr} is the correlation period in imaginary time (see statistics talk).
  - In steady state, the rates balance. Hence $\Delta\sim T_{\mathrm{corr}}\tau$.

- **The error in the mixed distribution is $\mathcal{O}(\tau)$.**

Detailed Balance Condition

- Green's function for the ITSE: $\langle \mathbf{R}| \exp[-\tau(\hat{H} - E_T)]|\mathbf{R}'\rangle$.

- Multiply by a constant $[\Psi^{-1}(\mathbf{R}')]$ times $\Psi(\mathbf{R})$ to obtain the Green's function for the ISITSE:

$$G(\mathbf{R} \leftarrow \mathbf{R}', \tau) = \Psi(\mathbf{R})\langle \mathbf{R}| \exp[-\tau(\hat{H} - E_T)]|\mathbf{R}'\rangle \Psi^{-1}(\mathbf{R}').$$

- $\exp[-\tau(\hat{H} - E_T)]$ is Hermitian and the Green's functions are real, so

$$\Psi^2(\mathbf{R}')G(\mathbf{R} \leftarrow \mathbf{R}', \tau) = \Psi^2(\mathbf{R})G(\mathbf{R}' \leftarrow \mathbf{R}, \tau).$$

- Approximation that $\mathbf{V}(\mathbf{R})$ is constant between $\mathbf{R}'$ and $\mathbf{R}$ violates this *detailed-balance* condition for finite time steps.

- Can re-impose this important condition using an accept/reject step (see later).

Propagation of Configuration Population (I)

- At any given moment in a DMC simulation, f is represented by a population of "configurations" or "walkers" in configuration space:

$$f(\mathbf{R},t) = \left\langle \sum_{\alpha} w_{\alpha} \delta(\mathbf{R} - \mathbf{R}_{\alpha}) \right\rangle ,$$

  where $\mathbf{R}_{\alpha}$ is the position of configuration α and w_{α} is its weight and the angled brackets are an ensemble average.

- Ensemble averaging commutes with differentiation, etc.

- Substitute the above expression for f into the integral form of the ISITSE to find the distribution of configurations one time step τ (**dtdmc**) later:

$$f(\mathbf{R},t+\tau) = \left\langle \sum_{\alpha} w_{\alpha} G_B(\mathbf{R} \leftarrow \mathbf{R}_{\alpha},\tau) G_D(\mathbf{R} \leftarrow \mathbf{R}_{\alpha},\tau) \right\rangle .$$

Propagation of Configuration Population (II)

- It is clear that the Green's functions can be treated as **transition-probability densities**; ensemble average then has the correct behaviour.

- To simulate this, configurations drift by $\tau \mathbf{V}(\mathbf{R}_\alpha)$ and diffuse (are displaced by a random vector, Gaussian-distributed with variance τ). Branching factor is then absorbed into a new weight for each configuration.

- Calculation has two phases:
  - Make a number of moves before energy data are accumulated, to allow excited-state components of Φ to die away: *equilibration phase*.
  - Then continue to propagate configurations, but gather energy data: *statistics-accumulation phase*.

Drift, Diffusion and the Accept/Reject Step (I)

- Drift–diffusion: each electron i in each configuration α is moved from $\mathbf{r}'_i(\alpha)$ to $\mathbf{r}_i(\alpha)$ according to

$$\mathbf{r}_i = \mathbf{r}'_i + \boldsymbol{\chi} + \tau \mathbf{v}_i(\mathbf{r}_1, \ldots, \mathbf{r}_{i-1}, \mathbf{r}'_i, \ldots, \mathbf{r}'_N),$$

  where $\boldsymbol{\chi}$ is a three-dimensional vector of Gaussian-distributed numbers with variance τ and zero mean and $\mathbf{v}_i(\mathbf{R})$ denotes those components of the total drift velocity $\mathbf{V}(\mathbf{R})$ due to electron i.

- Hence each electron i is moved from $\mathbf{r}'_i$ to $\mathbf{r}_i$ with transition-probability density

$$\begin{aligned} & t_i(\mathbf{r}_1, \ldots, \mathbf{r}_{i-1}, \mathbf{r}_i \leftarrow \mathbf{r}'_i, \mathbf{r}'_{i+1}, \ldots, \mathbf{r}'_N) \\ & = \frac{1}{(2\pi\tau)^{3/2}} \exp\left(\frac{[\mathbf{r}_i - \mathbf{r}'_i - \tau \mathbf{v}_i(\mathbf{r}_1, \ldots, \mathbf{r}_{i-1}, \mathbf{r}'_i, \ldots, \mathbf{r}'_N)]^2}{2\tau} \right). \end{aligned}$$

- Transition-probability density for move from $\mathbf{R}' = (\mathbf{r}'_1, \ldots, \mathbf{r}'_N)$ to $\mathbf{R} = (\mathbf{r}_1, \ldots, \mathbf{r}_N)$ is probability that each electron i moves from $\mathbf{r}'_i$ to $\mathbf{r}_i$.

Drift, Diffusion and the Accept/Reject Step (II)

- Transition-probability density for the configuration move:

$$T(\mathbf{R} \leftarrow \mathbf{R}') = \prod_{i=1}^{N} t_i(\mathbf{r}_1, \ldots, \mathbf{r}_{i-1}, \mathbf{r}_i \leftarrow \mathbf{r}'_i, \mathbf{r}'_{i+1}, \ldots, \mathbf{r}'_N).$$

- In the limit $\tau \to 0$, the drift velocity $\mathbf{V}$ is constant over the configuration move. Hence

$$T(\mathbf{R} \leftarrow \mathbf{R}') = G_D(\mathbf{R} \leftarrow \mathbf{R}', \tau),$$

  so the drift–diffusion process is indeed described by the drift–diffusion Green's function.

- For finite time steps, the approximation that the drift velocity is constant violates the detailed-balance condition.

- Enforce detailed balance using a Metropolis-style accept/reject step.

Drift, Diffusion and the Accept/Reject Step (III)

- *Electron-by-electron algorithm*: a move of electron i is accepted with probability

$$\min\left\{1, \frac{t_i(\mathbf{r}_1,\ldots,\mathbf{r}_{i-1},\mathbf{r}'_i \leftarrow \mathbf{r}_i,\mathbf{r}'_{i+1},\ldots,\mathbf{r}'_N)\Psi^2(\mathbf{r}_1,\ldots,\mathbf{r}_i,\mathbf{r}'_{i+1},\ldots,\mathbf{r}'_N)}{t_i(\mathbf{r}_1,\ldots,\mathbf{r}_{i-1},\mathbf{r}_i \leftarrow \mathbf{r}'_i,\mathbf{r}'_{i+1},\ldots,\mathbf{r}'_N)\Psi^2(\mathbf{r}_1,\ldots,\mathbf{r}_{i-1},\mathbf{r}'_i,\ldots,\mathbf{r}'_N)}\right\}$$

- *Configuration-by-configuration algorithm*: a move of all N electrons is accepted with probability

$$\min\left\{1, \frac{T(\mathbf{R}' \leftarrow \mathbf{R})\Psi^2(\mathbf{R})}{T(\mathbf{R} \leftarrow \mathbf{R}')\Psi^2(\mathbf{R}')}\right\}$$

- In either case, the RMS distance diffused in configuration space over one time step is $\sqrt{3N\tau p}$, where p is the acceptance probability.

- For a given time step, the electron-by-electron algorithm is more efficient, because the acceptance probability is higher.

- CASINO can perform either e-by-e or c-by-c calculations (choose with the **dmc_method** keyword), but the former are clearly preferable in general.

Branching

- After all the electrons have attempted to move, the branching factor of configuration α is calculated as:

$$M_b(\alpha) = G_B(\mathbf{R}_\alpha \leftarrow \mathbf{R}'_\alpha, \tau) = \exp\left[\left(-\frac{1}{2}\left[E_L(\mathbf{R}_\alpha) + E_L(\mathbf{R}'_\alpha)\right] + E_T\right)\tau\right],$$

  where $\mathbf{R}'_\alpha$ and $\mathbf{R}_\alpha$ are the electron coordinates before and after the complete drift–diffusion step.

- The "weight" of configuration α should be increased by a factor of $M_b(\alpha)$ (on average).

- "Unweighted" DMC: number of copies of configuration α in next time step is

$$M(\alpha) = \mathrm{int}[\eta + M_b(\alpha)],$$

  where η is a random number drawn from a uniform distribution on $[0, 1]$.

- Expected number of daughter configurations after branching: $\langle M(\alpha)\rangle = M_b(\alpha)$.

Adjusting the Reference Energy

- Let $E_{\rm best}$ be the best available estimate of the ground-state energy.
  - During equilibration $E_{\rm best}$ is the average configuration local energy over the last **ebest_av_window** iterations.
  - During statistics accumulation, $E_{\rm best}$ is the average local energy since the start of statistics accumulation.

- The branching factors depend exponentially on the reference energy E_T.

- E_T is adjusted so that the total population does not deviate too much from a *target population* M_0 (**dmc_target_weight**).

- Specifically, the reference energy is adjusted so that the population attempts to return to M_0 after $1/c_T$ a.u. of imaginary time (or one time step, if that is greater).

Population-Control Bias (I)

- The DMC Green's function was derived under the assumption that E_T is constant. Adjusting E_T leads to *population-control* bias.
  - Suppose the local energies are mostly less than E_0. The population will try to increase. But the population-control mechanism counteracts this.
  - Suppose the local energies are mostly greater than E_0. The population will try to decrease. But the population-control mechanism counteracts this.

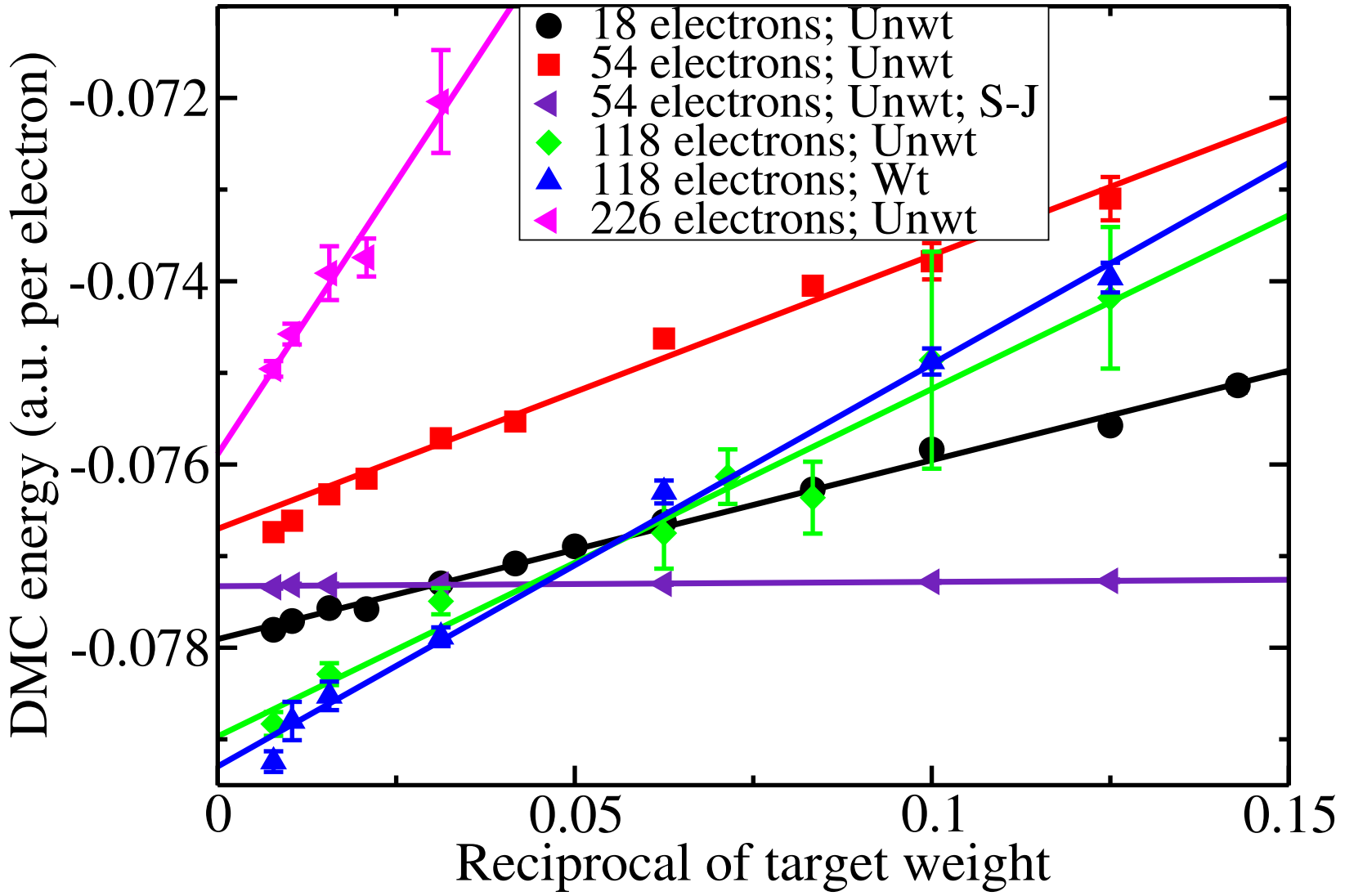
- In either case, the average local energy increases as a result. Population control introduces a positive bias into the DMC energy.

- Since fluctuations in the average local energy and branching factor are proportional to $1/\sqrt{M_0}$, population-control bias goes as $1/M_0$.

- Population-control bias increases very slowly as the system-size increases.

- *Improve the trial wave function to reduce population-control bias.*

Population-Control Bias (II)



Population-control bias in a paramagnetic Fermi fluid with $r_s = 4$ a.u. No Jastrow factor was used, except where indicated by "S-J".

Wave-Function Antisymmetry

- We want to find the **Fermionic** (antisymmetric) ground state.
- Lowest-energy wave function is the **Bosonic** (symmetric) ground state.
- Therefore we have to constrain DMC to preserve antisymmetry.
- Constraint is "automatic" in the importance-sampled DMC algorithm:
  - If Φ and Ψ have different nodal surfaces, there must exist regions where f is negative.
  - Our algorithm is based on interpreting f as a *probability density*.
  - Can never have a negative f in our algorithm.
  - So we cannot describe a change in the nodal surface of Φ.
- *By importance sampling and not permitting weights to become negative, we have introduced the fixed-node approximation.*

Fixed-Node Approximation (I)

- The nodes of Ψ divide configuration space into *nodal pockets*.
  - Within each nodal pocket λ we solve the Schrödinger equation subject to the boundary condition that the wave function Φ_λ is zero outside the pocket.
  - So $\hat{H}\Phi_\lambda = E_0^\lambda \Phi_\lambda + \delta_\lambda$, where δ_λ are δ functions at the pocket boundary arising from the discontinuity of the derivative of Φ_λ and E_0^λ is the *pocket energy*.

- Consider antisymmetric wave function $\tilde{\Phi}_\lambda(\mathbf{R}) = \hat{A}\Phi_\lambda(\mathbf{R}) \equiv \sum_{\hat{P}} (-1)^p \hat{P}\Phi_\lambda(\mathbf{R})$, where $\{\hat{P}\}$ permute like-spin coordinates and $\{p\}$ are the corresponding parities.

- Variational principle:

$$E_0^F \leq \frac{\langle \tilde{\Phi}_\lambda | \hat{H} | \tilde{\Phi}_\lambda \rangle}{\langle \tilde{\Phi}_\lambda | \tilde{\Phi}_\lambda \rangle} = \frac{\langle \Phi_\lambda | \hat{A}\hat{H}\hat{A} | \Phi_\lambda \rangle}{\langle \Phi_\lambda | \hat{A}^2 | \Phi_\lambda \rangle} = \frac{\langle \Phi_\lambda | \hat{A}^2\hat{H} | \Phi_\lambda \rangle}{\langle \Phi_\lambda | \hat{A}^2 | \Phi_\lambda \rangle} = E_0^\lambda,$$

  so each pocket energy is greater than the Fermion ground-state energy E_0^F.
  - We have used the fact that $\hat{A}$ is Hermitian and that it commutes with $\hat{H}$, and that the contribution due to δ_λ vanishes because $\tilde{\Phi}_\lambda = 0$ at nodes.

Fixed-Node Approximation (II)

- Within each nodal pocket λ, the mixed estimator gives energy E_0^λ.
- Configuration populations in high-energy pockets tend to die out, so the DMC energy is $\min\{E_0^\lambda\} \geq E_0^F$.
- **Hence the fixed-node DMC energy exceeds the Fermion ground-state energy, becoming equal in the limit that the nodal surface is exact.**
- *FNA is the only fundamental approximation in DMC.* The error in the DMC energy is second order in the error in the nodal surface.
- Drift velocity diverges at nodal surface, carrying away configurations.
- At finite time steps, configurations can drift/diffuse across surface: source of time-step bias.

Fixed-Node Approximation (III)

- The FNA with an antisymmetric trial wave function gives us a variational principle for the lowest-energy antisymmetric eigenstate.
  - Likewise, the FNA with a trial wave function of any given symmetry gives us a variational principle for the lowest-energy state with that symmetry.

- DMC always gives the energy of an excited state exactly if the nodal surface is exact for that state.
  - Hence we can use fixed-node DMC to calculate excited-state energies by using an appropriate trial wave function.
  - So we can calculate excitation energies (points on the band structure for periodic systems) via differences in total energy.
  - The variational principle does not hold for excited-state energies in general.

Mixed Estimator of the Energy

- Fixed-node ground state satisfies $\hat{H}\phi_0 = E_0\phi_0$. Hence

$$E_0 = \frac{\langle\phi_0|\hat{H}|\Psi\rangle}{\langle\phi_0|\Psi\rangle} = \frac{\int \phi_0\Psi E_L\, d\mathbf{R}}{\int \phi_0\Psi\, d\mathbf{R}} = \langle E_L\rangle_{\phi_0\Psi}.$$

- After equilibration, the DMC configuration population has the mixed distribution $\phi_0\Psi$.

- Average the local energies of the configurations generated in the statistics-accumulation phase to obtain an estimate of E_0.

- At each iteration, the local energies are averaged over the configuration population (weighted by the branching factors) and written to `dmc.hist`, along with the total weight.

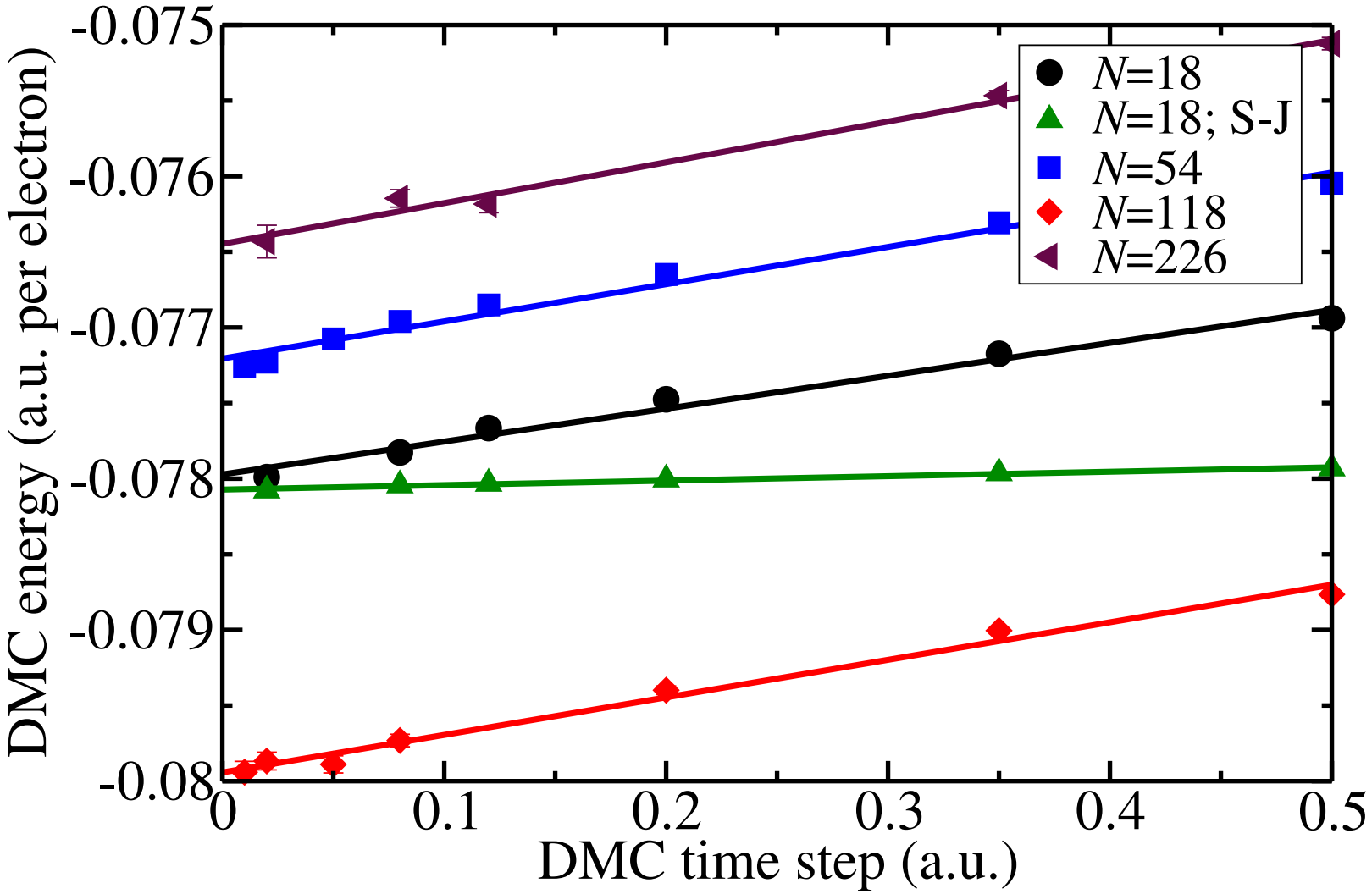
Time-Step Bias (I)

- The $\mathcal{O}(\tau)$ error in the mixed distribution gives an $\mathcal{O}(\tau)$ **time-step bias** in the mixed estimator.
  - Time-step bias vanishes in the limit of zero time step and is linear for sufficiently small time steps.
  - Time-step bias does not get more severe in larger systems.
  - The bias is greatly reduced if the trial wave function is good.

- Must either (i) use a sufficiently small time step that the bias is negligible or (ii) perform simulations at different time steps and extrapolate to zero time step.
  - The EXTRAPOLATE_TAU utility exists to help you do the latter.

- Time-step biases may cancel out of energy differences, but one must not assume this without checking.

Time-Step Bias (II)



Time-step bias in a paramagnetic Fermi fluid with $r_s = 4$ a.u. No Jastrow factor was used, except where indicated by "S-J".

Choosing a DMC Time Step

- **Always** check for time-step bias by performing simulations at different time steps.
- First guess at time step: *one fiftieth of optimised VMC time step.*[1]
- The RMS distance diffused by each electron each time step ($\sqrt{3\tau}$) should be $\leq$ the *smallest* length scale.
- For "typical" time steps in "typical" systems, local energies separated by 1024 iterations can typically be regarded as independent, i.e., the block length in REBLOCK can be chosen to be 1024.
- Typically need tens of thousands of steps for equilibration and statistics accumulation.
- RMS distance diffused by each electron over equilibration period ($\sqrt{3N_{\rm eq}\tau}$, where $N_{\rm eq}$ is number of equilibration iterations) should be $\geq$ *longest* length scale in problem.

[1] By default, the VMC time step is optimised to ensure that $\sim 50\%$ of moves are accepted.

Population-Explosion Catastrophes (I)

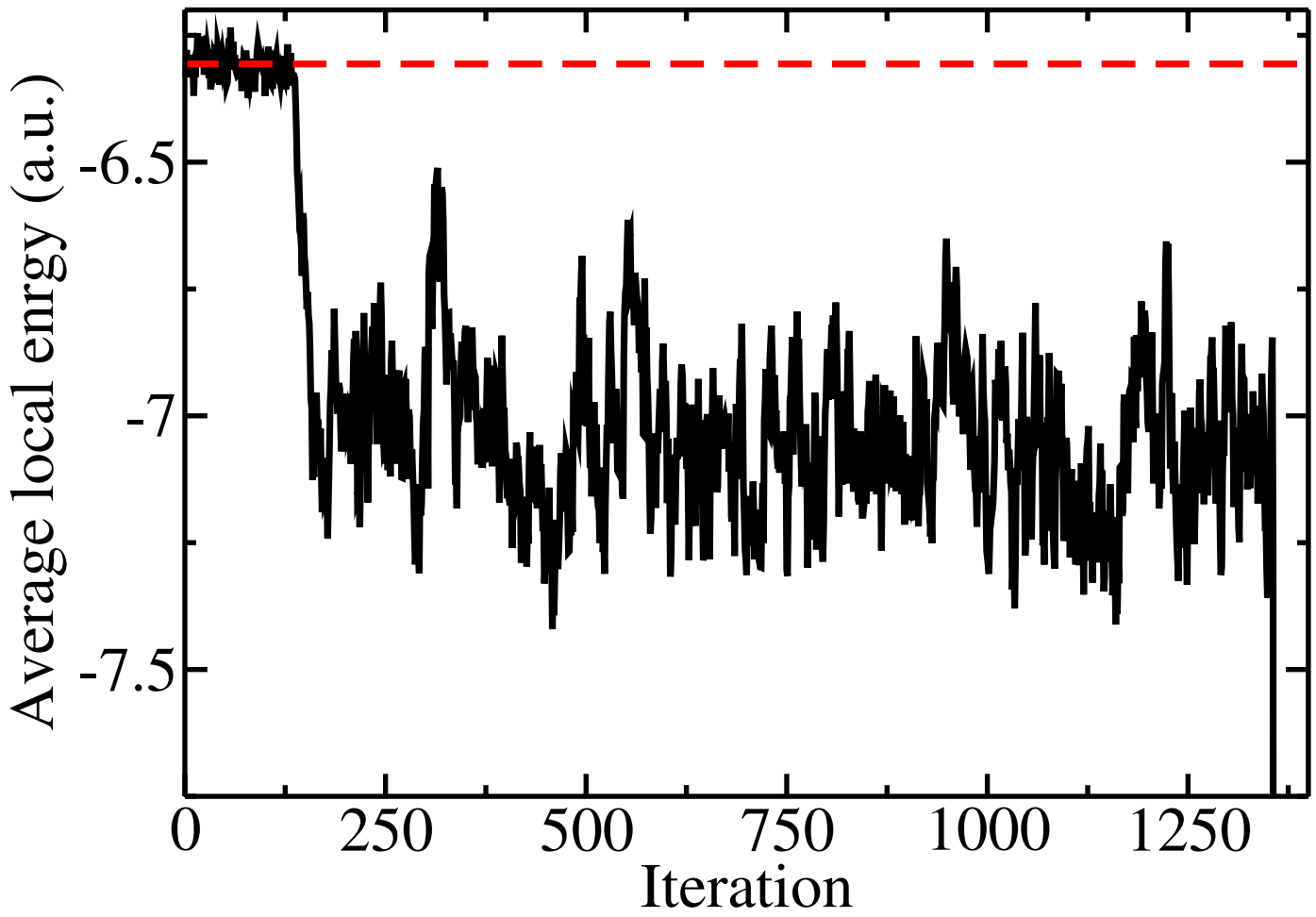
- Configuration-population explosions are liable to occur whenever the local energy shows singular behaviour.

- Large local energies can invalidate the short-time approximation and the branching factor can diverge.

- Usual signature: unphysically low average configuration energy, accompanied by a jump in the population.

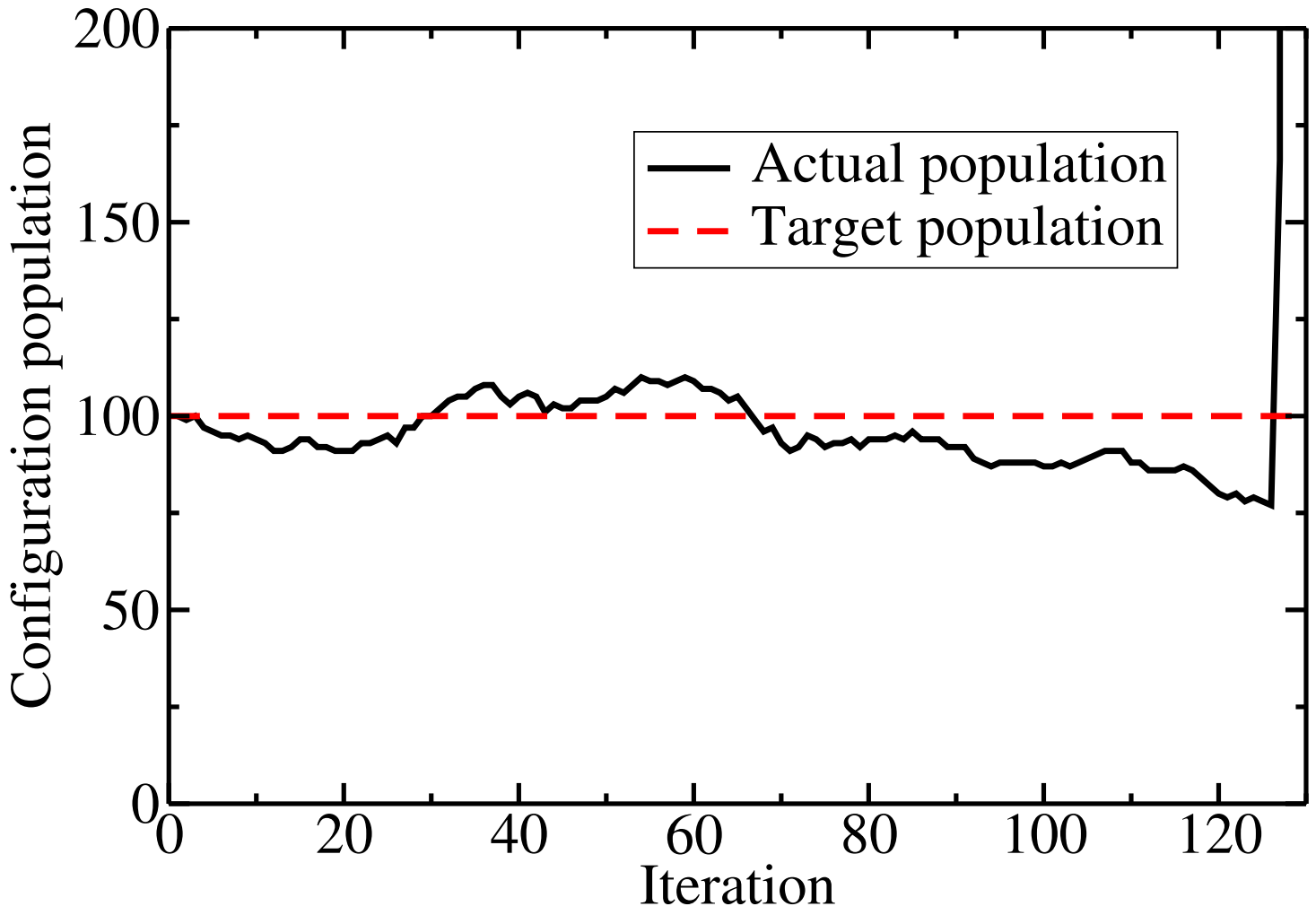
- CASINO halts if the population on a processor exceeds $5 \times$ **dmc_target_weight**$/P$, where P is the number of processors.

Population-Explosion Catastrophes (II)



Average local energy during a DMC simulation of SiH_4. Dashed line shows DMC ground-state energy as found using simulations with a much smaller time step. Electron–nucleus cusp condition not satisfied at hydrogen nuclei.

Population-Explosion Catastrophes (III)

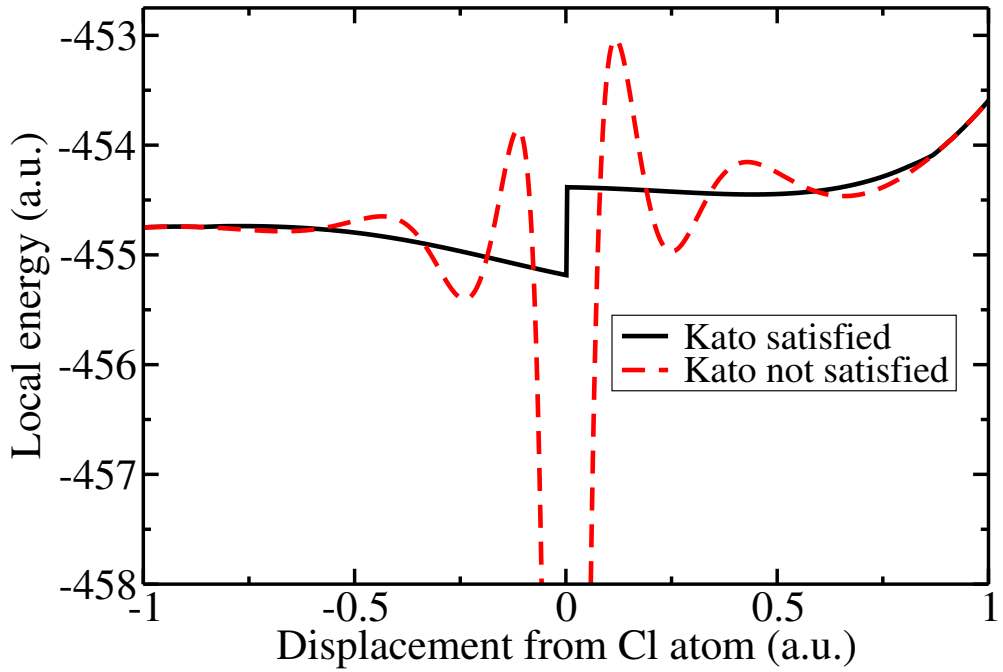


Config population in a simulation of H_2O. Electron–nucleus cusp condition not satisfied.

Nuclear Persistent-Electron Catastrophes (I)

- If the electron–nucleus cusp condition isn't satisfied, the local energy diverges as r^{-1} when an electron approaches a nucleus.

- The divergence is negative, causing a positive divergence of the branching factor.



Local energy as electron moves through a bare nucleus. Note that local energy does not change sign at divergence, unlike divergence at a node.

Nuclear Persistent-Electron Catastrophes (II)

- Probability density of electrons being at a nucleus is finite.
- Accept/reject step tends to prevent electron moves away from the nucleus, so electrons may become trapped.
- Possible for simulation to proceed with a population of configurations containing a "persistent" electron, but there will be a large negative bias.
- If the multiplicity is sufficiently high, an unbounded population explosion occurs.
- Use a trial wave function satisfying electron–nucleus cusp conditions: nuclear persistent-electron catastrophes are eliminated.
- E.g., if all-electron calculations are performed with a Gaussian basis set, **cusp_correction** should be `T` (as it is by default).

Nonlocal Pseudopotentials

- Nonlocal pseudopotentials are awkward in DMC: when we derived the DMC Green's function we tacitly assumed that the potential was local.

- Locality approximation: replace nonlocal pseudopotential $\hat{V}_{\rm NL}$ by local operator $\hat{V}_{\rm LA}(\mathbf{R}) = \Psi^{-1}(\mathbf{R})\hat{V}_{\rm NL}\Psi(\mathbf{R})$.

- Clear that $\langle\hat{V}_{\rm LA}\rangle_{\rm VMC} = \langle\Psi|\hat{V}_{\rm LA}|\Psi\rangle/\langle\Psi|\Psi\rangle = \langle\Psi|\hat{V}_{\rm NL}|\Psi\rangle/\langle\Psi|\Psi\rangle = \langle\hat{V}_{\rm NL}\rangle_{\rm VMC}$.

- Likewise, $\langle\phi_{\rm LA}|\hat{V}_{\rm LA}|\Psi\rangle/\langle\phi_{\rm LA}|\Psi\rangle = \langle\phi_{\rm LA}|\hat{V}_{\rm NL}|\Psi\rangle/\langle\phi_{\rm LA}|\Psi\rangle$, where $\phi_{\rm LA}$ is the ground state in the locality approximation.

- However, $\langle\phi_{\rm LA}|\hat{V}_{\rm NL}|\Psi\rangle/\langle\phi_{\rm LA}|\Psi\rangle \neq \langle\phi_{\rm NL}|\hat{V}_{\rm NL}|\Psi\rangle/\langle\phi_{\rm NL}|\Psi\rangle$, where $\phi_{\rm NL}$ is the actual ground state (for the nonlocal pseudopotential).

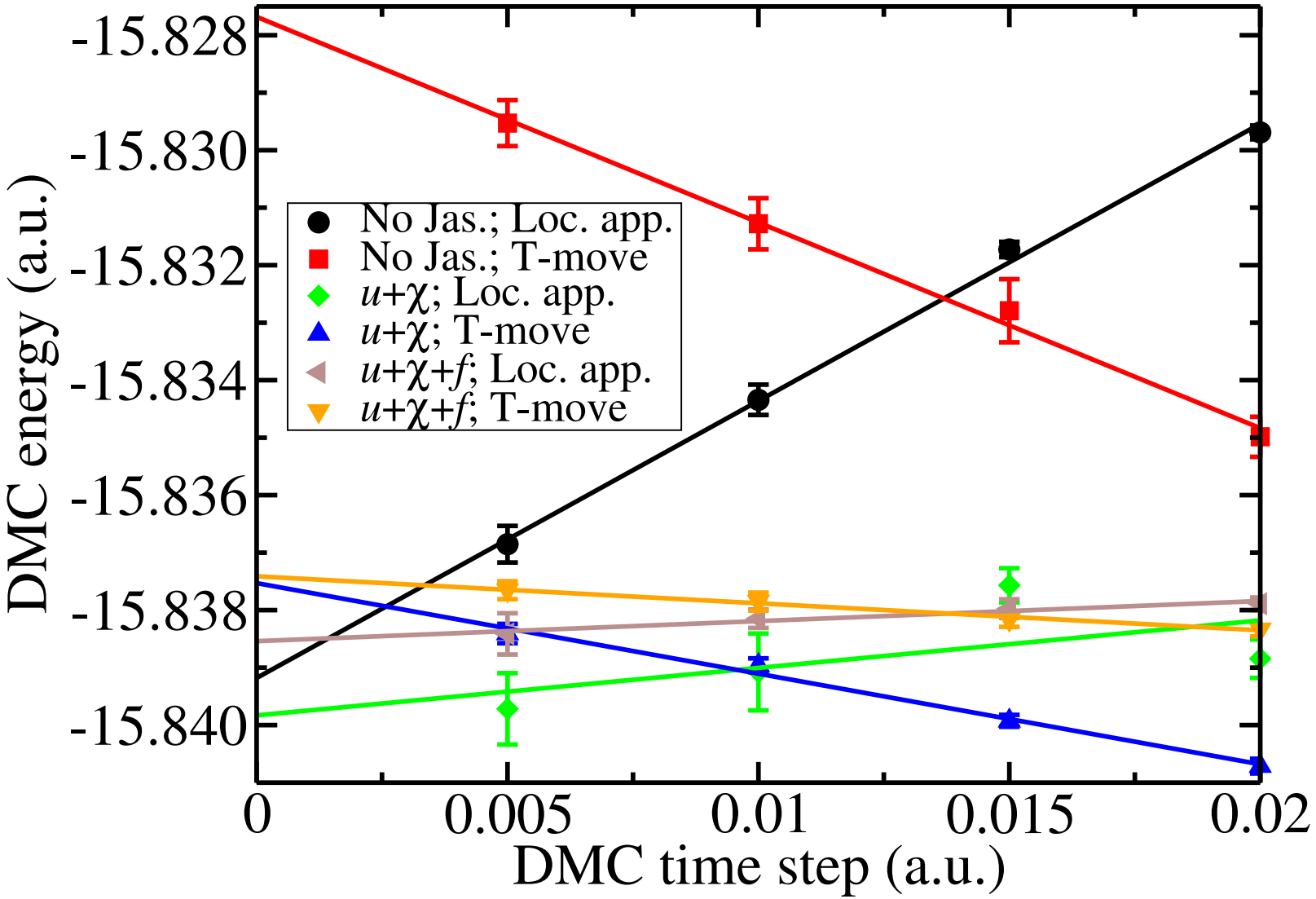
- Can show that error in $\langle\phi_{\rm LA}|\hat{V}_{\rm NL}|\Psi\rangle/\langle\phi_{\rm LA}|\Psi\rangle$ is second order in error in Ψ. However, DMC energies in locality approximation are not guaranteed to exceed GS energy.

- The use of the locality approximation can lead to catastrophic behaviour.

T-Move Scheme (I)

- Split nonlocal pseudopotential into a part with negative matrix elements w.r.t. position basis and a part with positive matrix elements.
  - Simulate a process corresponding to negative elements (so-called *T-moves*).
  - Treat positive matrix elements within locality approximation.
- Restores property that DMC energy is greater than the ground-state energy.
- Tends to move electrons away from nodes on nonlocal integration grid: eliminates instabilities.
- Set **use_tmove** to `T` to use the T-move scheme.

T-Move Scheme (II)



DMC energy against time step for a pseudo-oxygen atom, with and without the T-move scheme. Three different trial wave functions are used.

If You Encounter a Catastrophe. . . (I)

- If you encounter catastrophic behaviour:
  1. Check that your wave function satisfies the Kato cusp conditions by using **qmc_plot** to examine the local energy as an electron is moved through each nucleus[2].
  2. Use the **jastrow_plot** block to examine the Jastrow factor. The u term should increase monotonically to zero.
  3. Consider using the T-move scheme for nonlocal pseudopotentials.
  4. Consider using a smaller time step.

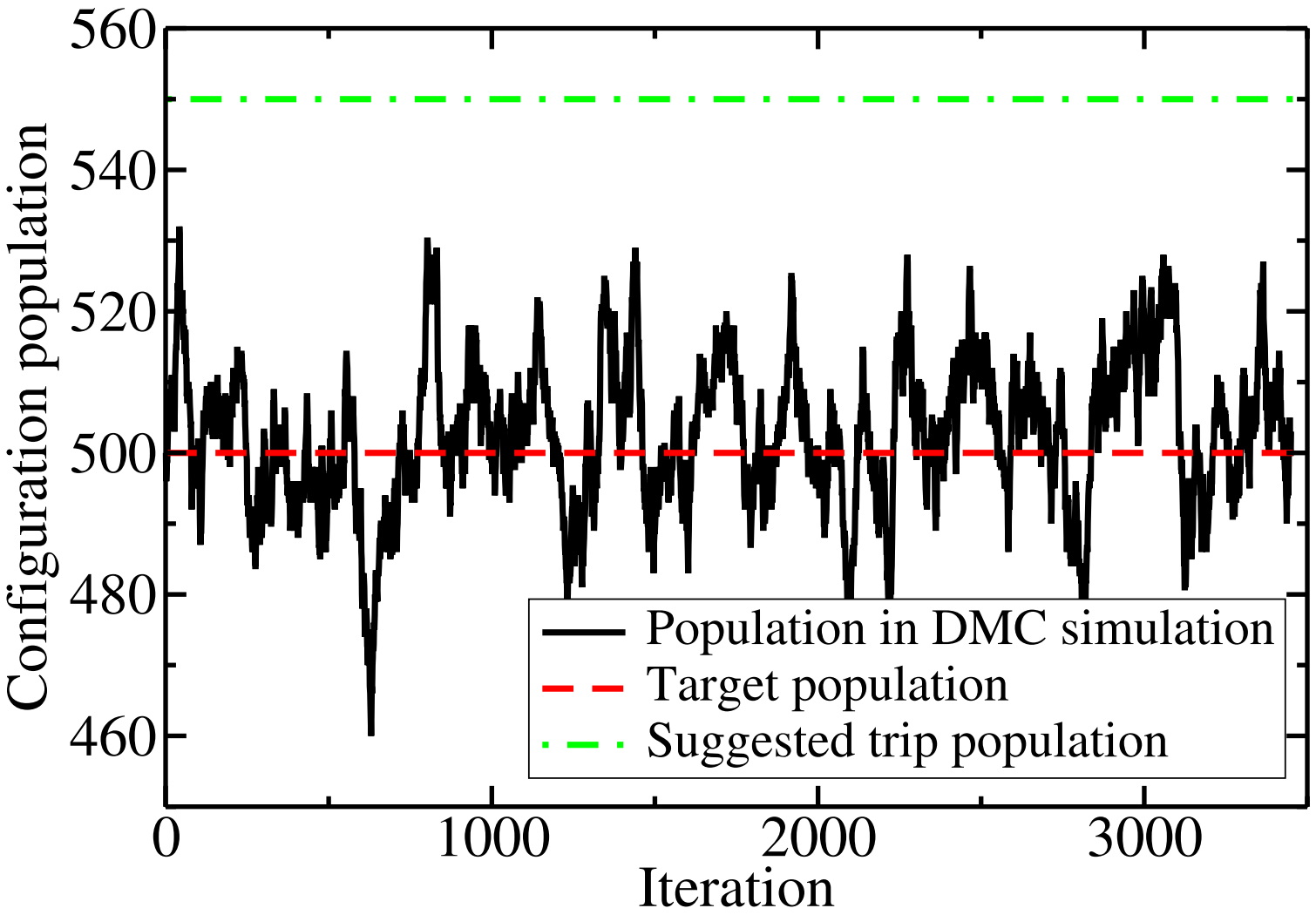
- If you still encounter occasional catastrophic behaviour, you can set an upper limit on the population (**dmc_trip_weight**).

- CASINO will jump back to an earlier point in the simulation and change the random number sequence if the upper limit is exceeded.

- Choose the upper limit to be slightly higher than the maximum population that would be encountered in an ordinary fluctuation.

[2] Type *casinohelp qmc_plot* to find out how to use **qmc_plot**.

If You Encounter a Catastrophe. . . (II)



Suggested trip population in a DMC simulation of SiH_4.

Summary (I)

To carry out successful DMC calculations:

- Test for time-step bias by performing simulations at different time steps (for a representative system).
  - If necessary, remove time-step bias by extrapolation to zero time step.
  - "Typical" DMC time step in a pseudopotential calculation: $\tau \approx 0.01$ a.u.
  - "Typical" DMC time step in an all-electron calculation for first-row atoms: $\tau \approx 0.001$ a.u.
  - Time steps vary widely, especially in HEGs. Make sure the distance diffused in one time step $\sqrt{3\tau}$ is smaller than the shortest relevant length scale.
  - In general, a reasonable first guess at the DMC time step is about **one fiftieth** of the **optimised** VMC time step.
  - Time-step biases in energy differences are often much smaller than the biases in the individual energies: can often exploit this cancellation of bias to use a relatively large time step.

Summary (II)

- Ensure the equilibration period is sufficiently long.
  - Ensure the RMS distance diffused over the equilibration period $\sqrt{3N_{\rm eq}\tau}$ is larger than the longest relevant length scale.
  - Typically use several thousand or tens of thousands of equilibration moves.

- Use a sufficiently large target population.
  - Typically use a target population of more than 1,000 configurations.
  - Test for population-control bias by halving or doubling the population for a representative calculation.

- Use a highly optimised trial wave function.
  - Statistical error bars are reduced; time-step and population-control biases are reduced; locality errors are reduced; and (if backflow or multideterminant wave functions are used) fixed-node errors are reduced when the wave function is well-optimised; the chance of encountering population-control problems is reduced.
  - Make sure the Kato cusp conditions are satisfied.

Summary (III)

- Ensure that your statistical error bars are an order of magnitude smaller than the energy difference you are trying to resolve.
  - If error bars were exact and underlying statistics were Gaussian then one in three points would be out by one error bar, one in twenty would be out by two error bars, etc.
  - For a given system, error bars fall off as the square root of the amount of data you gather. Can easily estimate the computational requirements for achieving a given error bar by performing a short test calculation.