Semistochastic Quantum Monte Carlo – A Hybrid of Exact Diagonalization and QMC Methods and Optimization of FN-PMC energies and FN-PMC forces

Cyrus Umrigar

Physics Department, Cornell University, Ithaca.

Email: CyrusUmrigar@cornell.edu

Quantum Monte Carlo in the Apuan Alps IX, Towler Institute, Vallico Sotto, 26 July - 2 Aug, 2014

Outline

SQMC:

1. Intro to Variational and Projector Monte Carlo (PMC) methods (zero temperature)
2. Sign Problem in various Projector Monte Carlo (PMC)
3. Semistochastic Quantum Monte Carlo
   with Frank Petruzielo, Hitesh Changlani, Adam Holmes and Peter Nightingale, PRL (2012)

SQMC work motivated by:
a) FCIQMC: Alavi and group (Booth, Thom, Cleland, Spencer, Shepherd, ...)
b) PMC: Ohtsuka and Nagase

Valuable discussions with Bryan Clark, George Booth, Shiwei Zhang, Garnet Chan, Ali Alavi.

Derivatives of FN-PMC Energies:

4. Optimization of many-body wavefunctions
5. Efficient derivatives of FN-PMC energies

The problem

We wish to find the lowest energy eigenstate(s) of a sparse matrix, e.g., the Hamiltonian matrix.

If the number of basis states is sufficiently small that one can store a vector (say $< 10^{10}$), then one can use a deterministic iterative method, such as the power method or the Lanczos method.

Quantum Monte Carlo: If the space is larger than this, even infinite, one can use a stochastic implementation of the power method. At any instant in time only a random sample of the vector is stored in computer memory, and the solution is given by the time-average.

Definitions

Given a basis: $\{|\phi_i\rangle\}$, either discrete or continuous

$$\text{Exact} \quad |\Psi_0\rangle = \sum_i e_i|\phi_i\rangle, \quad \text{where,} \quad e_i = \langle\phi_i|\Psi_0\rangle$$

$$\text{Trial} \quad |\Psi_T\rangle = \sum_i t_i|\phi_i\rangle, \quad \text{where,} \quad t_i = \langle\phi_i|\Psi_T\rangle$$

$$\text{Guiding} \quad |\Psi_G\rangle = \sum_i g_i|\phi_i\rangle, \quad \text{where,} \quad g_i = \langle\phi_i|\Psi_G\rangle$$

Ψ_T used to calculate variational and mixed estimators of operators $\hat{A}$, i.e., $\langle\Psi_{\rm T}|\hat{A}|\Psi_{\rm T}\rangle/\langle\Psi_{\rm T}|\Psi_{\rm T}\rangle$, $\langle\Psi_{\rm T}|\hat{A}|\Psi_0\rangle/\langle\Psi_{\rm T}|\Psi_0\rangle$

Ψ_G used to alter the probability density sampled, i.e., $\Psi_{\rm G}^2$ in VMC, $\Psi_{\rm G}\Psi_0$ in PMC. Affects only the statistical error of VMC and PMC methods.

$\Psi_{\rm G}$ must be such that $g_i \neq 0$ if $e_i \neq 0$. If $\Psi_{\rm T}$ also satisfies this condition then $\Psi_{\rm G}$ can be chosen to be $\Psi_{\rm T}$. However, to speed up computation of mixed energy components, $\Psi_{\rm T}$ often has $t_i = 0$ on most states.

Variational MC

$$E_V = \frac{\langle \Psi_{\rm T}|\hat{H}|\Psi_{\rm T}\rangle}{\langle \Psi_{\rm T}|\Psi_{\rm T}\rangle} = \frac{\sum_{ij}^{N_{\rm st}} \langle \Psi_{\rm T}|\phi_i\rangle \ \langle \phi_i|\hat{H}|\phi_j\rangle \ \langle \phi_j|\Psi_{\rm T}\rangle}{\sum_{i}^{N_{\rm st}} \langle \Psi_{\rm T}|\phi_k\rangle \ \langle \phi_k|\Psi_{\rm T}\rangle}$$

$$= \frac{\sum_{ij}^{N_{\rm st}} t_i H_{ij} t_j}{\sum_{k}^{N_{\rm st}} t_k^2} = \sum_{i}^{N_{\rm st}} \frac{t_i^2}{\sum_{k}^{N_{\rm st}} t_k^2} \underbrace{\frac{\sum_{j}^{N_{\rm st}} H_{ij} t_j}{t_i}}_{E_L(i)}$$

$$= \sum_{i}^{N_{\rm st}} \frac{t_i^2}{\sum_{k}^{N_{\rm st}} t_k^2} E_{\rm L}(i) = \frac{\sum_{i}^{N_{\rm MC}} E_{\rm L}(i)}{N_{\rm MC}} \rightarrow_{\Psi_G \neq \Psi_T} \frac{\sum_{i}^{N_{\rm MC}} \left(\frac{t_i}{g_i}\right)^2 E_{\rm L}(i)}{\sum_{i}^{N_{\rm MC}} \left(\frac{t_i}{g_i}\right)^2}$$

Sample probability density function $\frac{g_i^2}{\sum_k^{N_{\rm st}} g_k^2}$ using Metropolis-Hastings.

Value depends only on $\Psi_{\rm T}$. Statistical error depend on $\Psi_{\rm T}$ and $\Psi_{\rm G}$.
Energy bias and statistical error vanish as $\Psi_{\rm T} \rightarrow \Psi_0$.
For fixed Ψ_T, $\Psi_G = \Psi_T$ minimizes statistical fluctuations of denominator but not of E_V! Making Ψ_G worse may reduce statistical fluctuations!

Projector MC

Pure and Mixed estimators for energy are equal: $E_0 = \dfrac{\langle\Psi_0|\hat{H}|\Psi_0\rangle}{\langle\Psi_0|\Psi_0\rangle} = \dfrac{\langle\Psi_0|\hat{H}|\Psi_T\rangle}{\langle\Psi_0|\Psi_T\rangle}$

Projector: $|\Psi_0\rangle = \hat{P}(\infty)\,|\Psi_T\rangle = \lim_{n\to\infty} \hat{P}^n(\tau)\,|\Psi_T\rangle$

$$E_0 = \frac{\langle\Psi_0|\hat{H}|\Psi_T\rangle}{\langle\Psi_0|\Psi_T\rangle} = \frac{\sum_{ij}^{N_{st}} \langle\Psi_0|\phi_i\rangle\ \langle\phi_i|\hat{H}|\phi_j\rangle\ \langle\phi_j|\Psi_T\rangle}{\sum_{k}^{N_{st}} \langle\Psi_0|\phi_k\rangle\ \langle\phi_k|\Psi_T\rangle}$$

$$= \frac{\sum_{ij}^{N_{st}} e_i H_{ij} t_j}{\sum_{k}^{N_{st}} e_k t_k} = \sum_{i}^{N_{st}} \frac{e_i t_i}{\sum_{k}^{N_{st}} e_k t_k}\ \frac{\sum_{j}^{N_{st}} H_{ij} t_j}{t_i}$$

$$= \sum_{i}^{N_{st}} \frac{e_i t_i}{\sum_{k}^{N_{st}} e_k t_k}\ E_L(i) \ = \ \frac{\sum_{i}^{N_{MC}} E_L(i)}{N_{MC}} \ \to_{\Psi_G \neq \Psi_T} \ \frac{\sum_{i}^{N_{MC}} \left(\frac{t_i}{g_i}\right) E_L(i)}{\sum_{i}^{N_{MC}} \left(\frac{t_i}{g_i}\right)}$$

For exact PMC, value indep. of Ψ_T, Ψ_G, statistical error depends on Ψ_T, Ψ_G.
For FN-PMC, value and statistical error depend on Ψ_G,Ψ_T. (Continuum $\Psi_G = \Psi_T$)
Energy bias and statistical error of E_{mix} vanish as $\Psi_T \to \Psi_0$ and of E_{gr} as $\Psi_G \to \Psi_0$.
For fixed Ψ_T, $\Psi_G = \Psi_T$ minimizes statistical fluctuations of denom. but not of E_V!

Projector Monte Carlo Methods

The amplitudes of Ψ_0 in the chosen basis are obtained by using a "Projector", $\hat{P}$, that is a function of the Hamiltonian, $\hat{H}$, and has Ψ_0 as its dominant state.

Various Projector Monte Carlo Methods differ in:
a) form of the projector, and,
b) space in which the walk is done (single-particle basis and quantization).
(1^{st}-quantized $\equiv$ unsymmetrized basis, 2^{nd}-quantized $\equiv$ antisymmetrized basis.)

Method	Projector	SP Basis	Quantiz
Diffusion Monte Carlo	$e^{\tau(E_T\hat{\mathbf{1}}-\hat{H})}$	$\mathbf{r}$	1^{st}
GFMC (Kalos, Ceperley, Schmidt)	$\frac{1}{\hat{\mathbf{1}}-\tau(E_T\hat{\mathbf{1}}-\hat{H})}$	$\mathbf{r}$	1^{st}
LRDMC (Sorella, Casula)	$\hat{\mathbf{1}}+\tau(E_T\hat{\mathbf{1}}-\hat{H})$	$\mathbf{r}_i$	1^{st}
PMC/FCIQMC/SQMC	$1+\tau(E_T\hat{\mathbf{1}}-\hat{H})$	ϕ_i^{orthog}	2^{nd}
phaseless AFQMC (Zhang, Krakauer)	$e^{\tau(E_T\hat{\mathbf{1}}-\hat{H})}$	$\phi_i^{\mathrm{nonorthog}}$	2^{nd}

Sign Problem

The nature of the sign problem is different in the various methods, depending on the space in which the walk is done.

Sign Problem in DMC

$\hat{P}(\tau) = e^{\tau(E_T\hat{\mathbf{1}}-\hat{H})}$

Walk is done in the space of the $3N$ coordinates of the N electrons.

$\langle \mathbf{R}|\hat{P}(\tau)|\mathbf{R}'\rangle \approx \frac{e^{\frac{-(\mathbf{R}-\mathbf{R}')^2}{2\tau}+\left(E_T-\frac{\mathcal{V}(\mathbf{R})+\mathcal{V}(\mathbf{R}')}{2}\right)\tau}}{(2\pi\tau)^{3N/2}}$ is nonnegative.

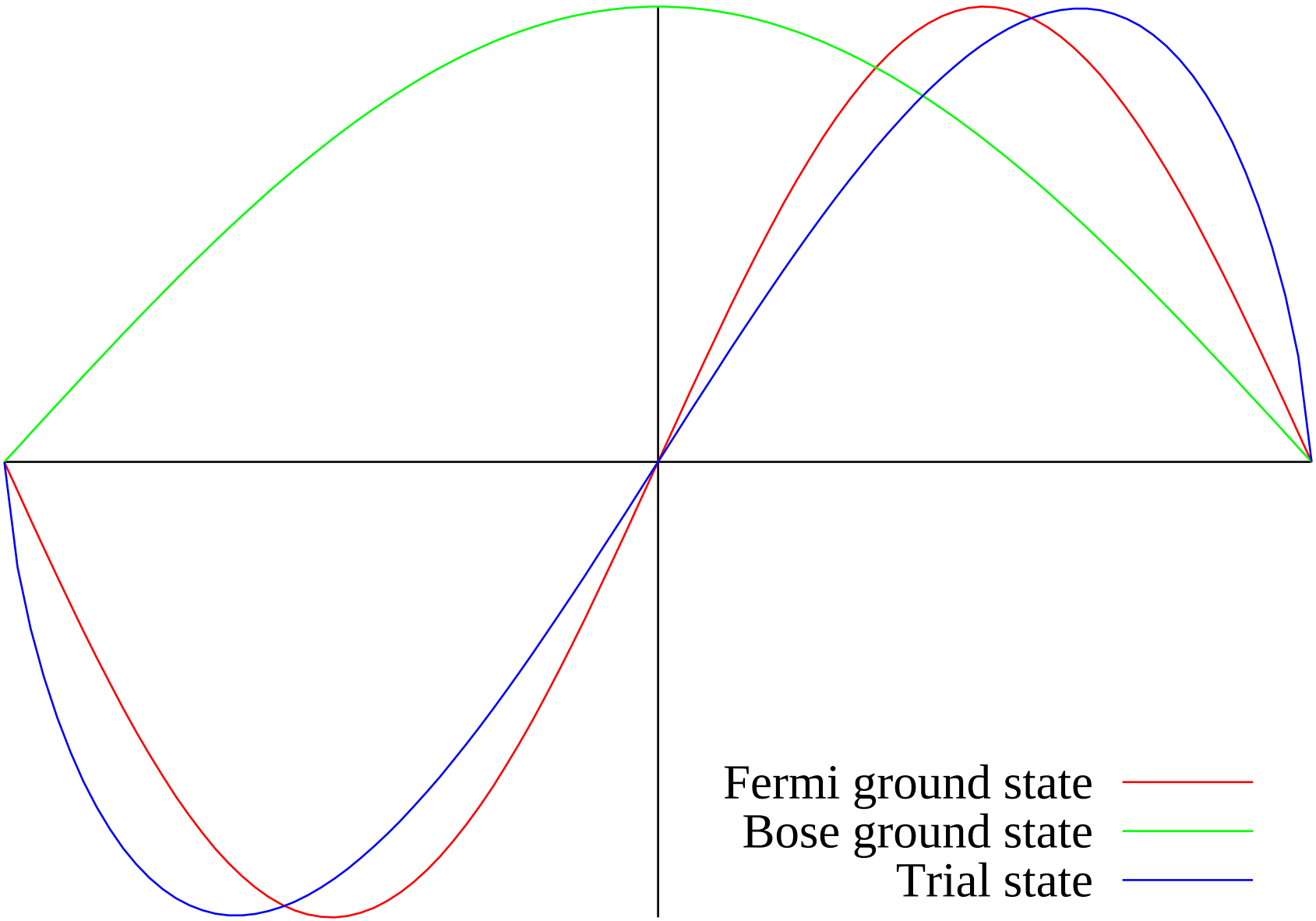
Problem: However, since the Bosonic energy is always lower than the Fermionic energy, the projected state is the Bosonic ground state.

Fixed-node approximation

All except a few calculations (release-node, Ceperley) are done using FN approximation. Instead of doing a free projection, impose the boundary condition that the projected state has the same nodes as the trial state $\Psi_{\mathrm{T}}(\mathbf{R})$.

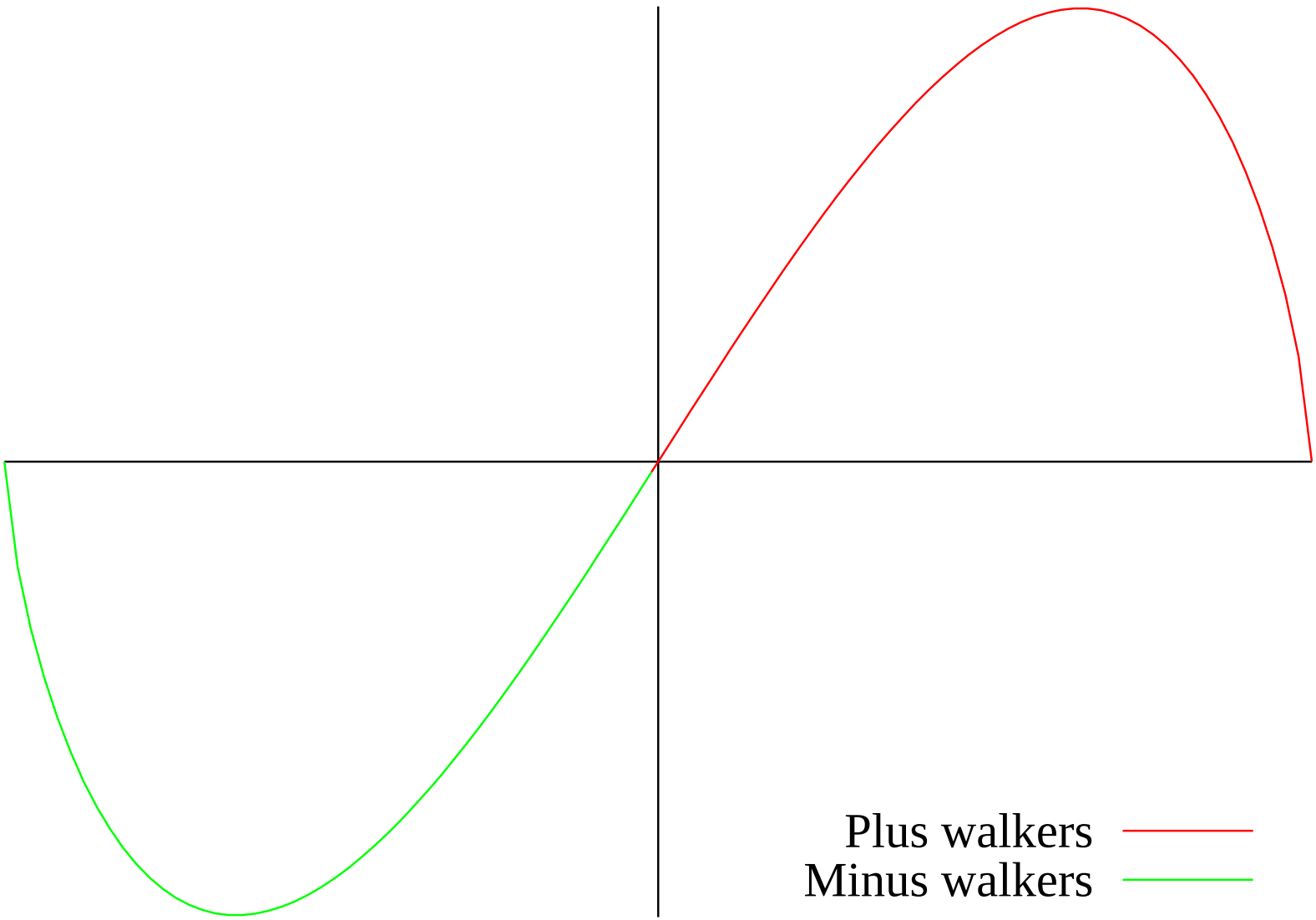
This gives an upper bound to the energy and becomes exact in the limit that Ψ_{T} has the same nodes as Ψ_0.

Sign Problem in 1^{st} Quantization and R space

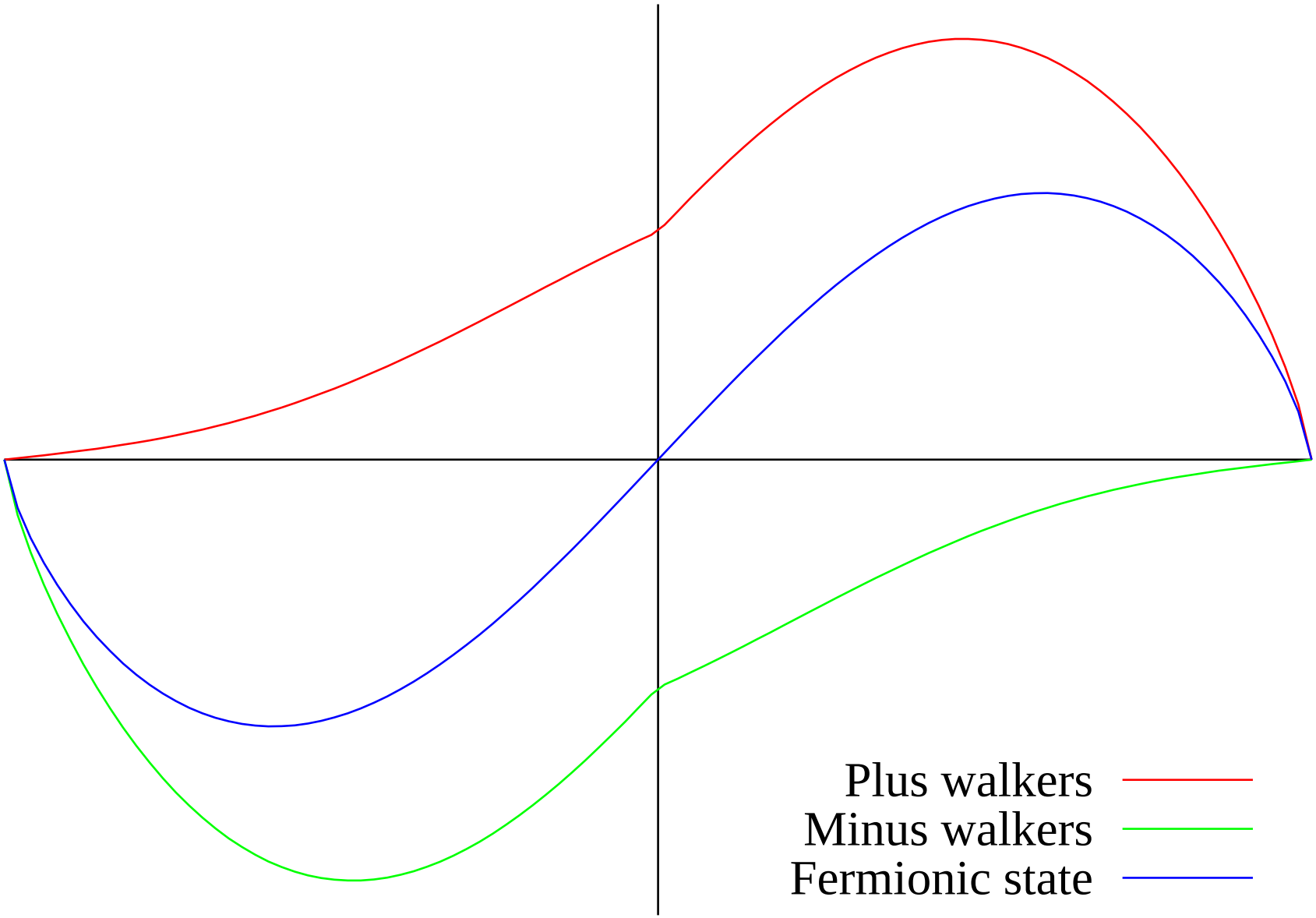


Sign Problem in 1^{st} Quantization and R space

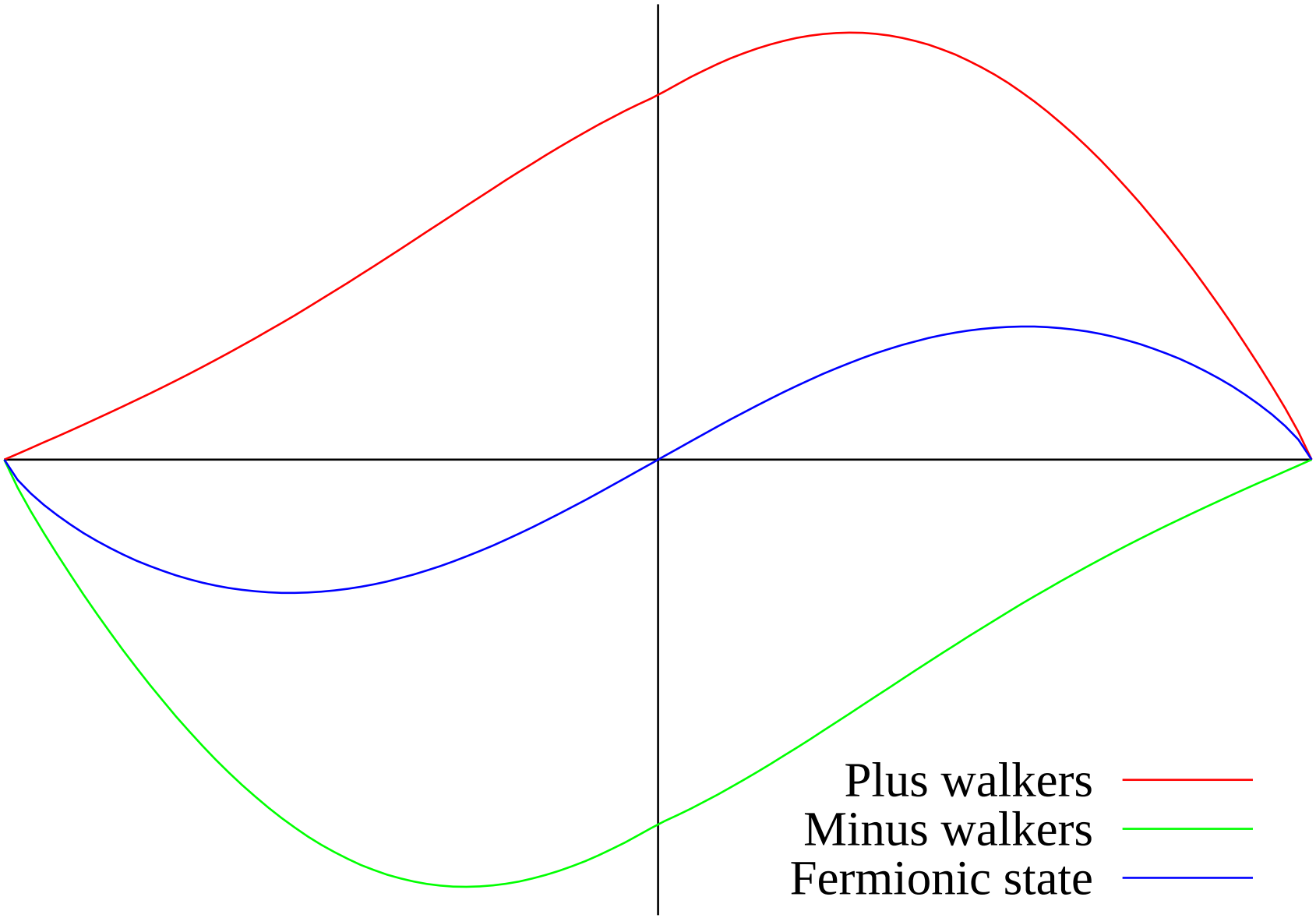
Start with equal + and - walkers, so no Bosonic component.



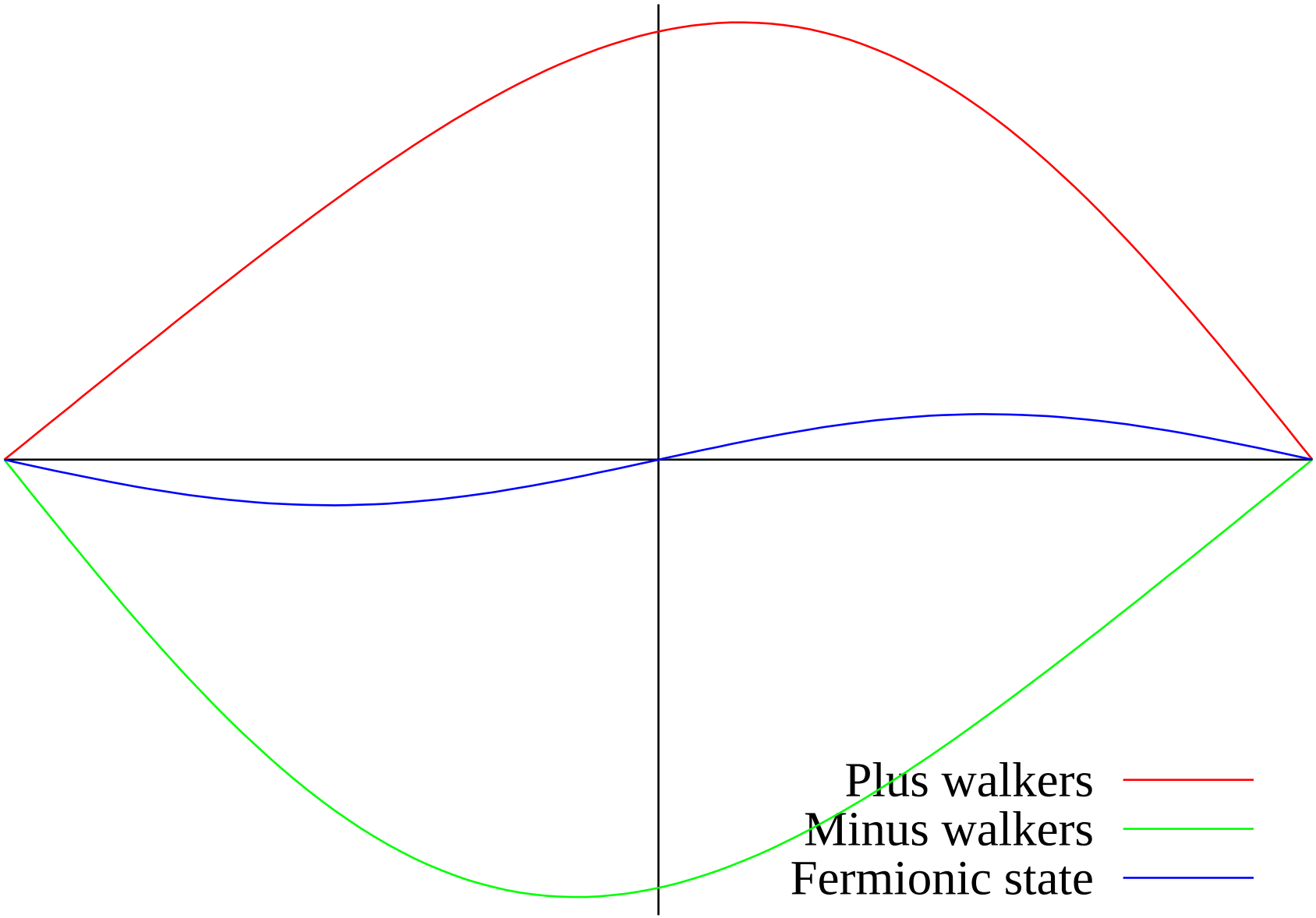
Sign Problem in 1^{st} Quantization and R space



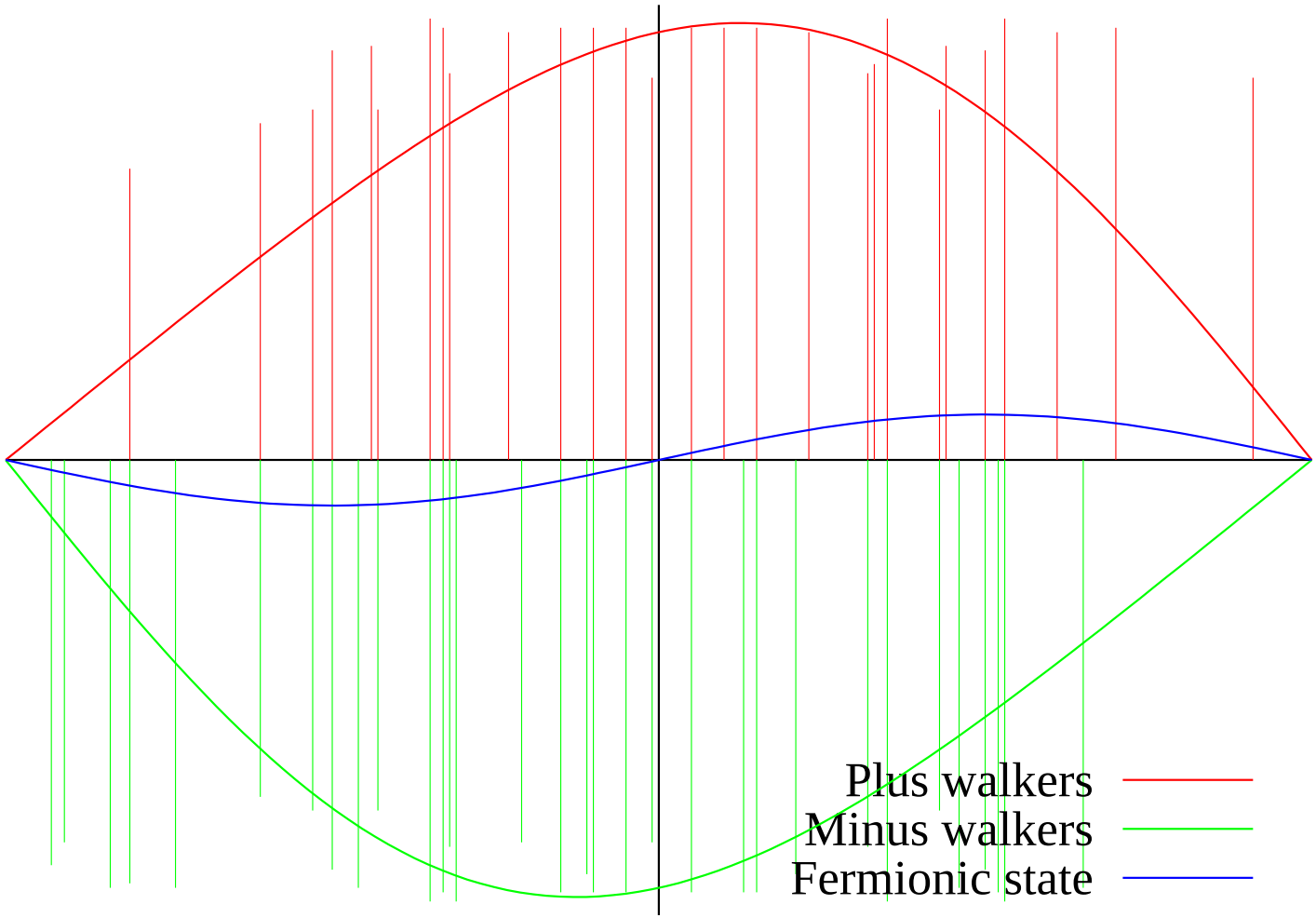
Sign Problem in 1^{st} Quantization and R space



Sign Problem in 1^{st} Quantization and R space



Sign Problem in 1^{st} Quantization and R space



Problem: In large space walkers rarely meet and cancel!

Worse Problem: Eventually + or - walkers dominate, there are no more cancellations and only one Bosonic compoent remains!

Sign Problem in 2^{nd} quantization

Walk is done in the space of determinants.

Since Bosonic and other symmetry states are eliminated, there is some hope of having a stable signal to noise, but there is still a sign problem.

Problem: Paths leading from state i to state j can contribute with opposite sign. Further, Ψ and $-\Psi$ are equally good.

The projector in the chosen basis does not have a sign problem if:
The columns of the projector have the same sign structure aside from an overall sign.
or equivalently:
It is possible to find a set of sign changes of the basis functions such that all elements of the projector are nonnegative.

The sign problem is an issue only because of the stochastic nature of the algorithm. Walkers of different signs can be spawned onto a given state in different MC generations.

Sign Problem in orbital space and 2^{nd} Quantization

FCIQMC (Booth, Thom, Alavi, JCP (2009)

When walk is done is space of determinants of HF orbitals, it is practical to have a population that is sufficiently large that cancellations can result in a finite signal to noise ratio. Once a critical population size is reached the probability of sign flips of the population rapidly become very small.

Initiator approximation (Cleland, Booth, Alavi, JCP (2010)

The required population size can be greatly reduced by allowing only determinants occupied by more than a certain number of walkers to spawn progeny on unoccupied determinants.

Becomes exact in the limit of infinite population size.

In subsequent papers they published FCIQMC calculations on various molecules, the homogeneous electron gas, and, real solids. Largest system has as many as 10^{108} states. (Note, however, that what matters is not the number of states, but, the number of states that have significant occupation.)

Sign Problem in FCIQMC/SQMC

Spencer, Blunt, Foulkes, J. Chem. Phys. (2012)
Kolodrubetz, Spencer, Clark, Foulkes, J. Chem. Phys. (2013)

1. The instability gap is given by the difference in the dominant eigenvalues of the projector, and, those of the projector with all off-diagonal elements replaced by their absolute values.

2. More than 1 Hartree product in a given initial determinant may connect via P (or H) to a given Hartree product in a final determinant. The instability gap is smaller in 2^{nd} quantization than in 1^{st} quantization if there are internal cancellations within these contributions, otherwise it is the same as in 1^{st} quantization.
   For example, it is the same in lattice real-space Coulomb systems, real- and momentum-space Hubbard models, but, is different for orbital-space Coulomb systems.

Sign Problem in FCIQMC/SQMC

These papers did not point out that even when the instability gap is the same, there are two important advantages of 2^{nd} quantization:

1. Since the Hilbert space is $N!$ times smaller in 2^{nd} quantization, cancellation are much more effective.
2. In first quantization, one of the two Bosonic populations will dominate and the signal to noise will go to zero even in the limit of an infinite population, unless additional steps are taken to prevent that.

Using a large population and cancellations, it is possible to get a finite signal to noise ratio in 2^{nd} quantization but not in 1^{st} quantization (unless some further constraints are imposed).

Original attempts at using cancellation to control sign problem (in continuum problems): Mal Kalos and coworkers (David Arnow (1982), Shiwei Zhang, Francesco Pederiva, ...)

Comparison of DMC with FCIQMC/SQMC

DMC (walk in electron coordinate space)	FCIQMC/SQMC (walk in determinant spac
Severe Fermion sign problem due to growth of Bosonic component relative to Fermionic.	Less severe Fermion sign problem due to opposite sign walkers being spawned on the same determinant
Fixed-node approximation needed for stable algorithm.	Walker cancellation, large population, initiator approximation needed for stable algorithm.
Exact if $\Psi_{\rm T}$ nodes exact.	Exact in ∞-population limit.
Infinite basis.	Finite basis. (Same basis set dependence as in other quantum chemistry methods.
Computational cost is low-order polynomial in N	Computational cost is exponential in N but with much smaller exponent than full CI
Need to use pseudopotentials for large Z.	Can easily do frozen-core

Semistochastic Quantum Monte Carlo (SQMC)

Frank Petruzielo, Adam Holmes, Hitesh Changlani, Peter Nightingale, CJU, PRL 2012

SQMC is hybrid of Exact Diagonalization and QMC

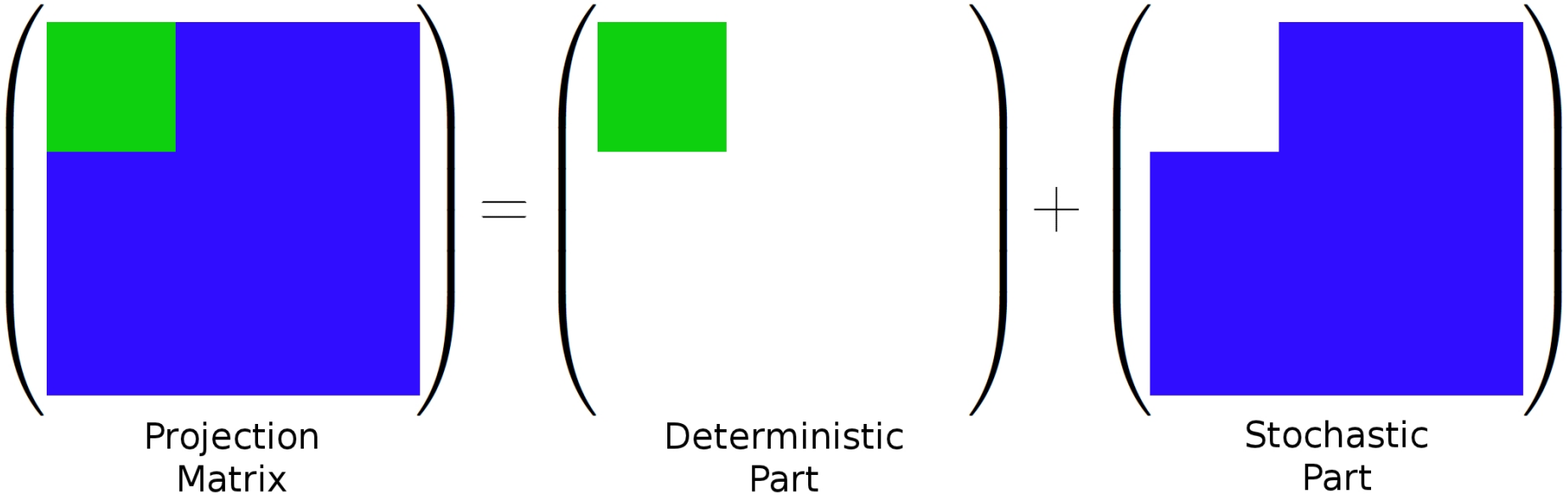
Exact diagonalization has no statistical error or sign problem but is limited to a small number of states ($\sim 10^{10}$ on a single core).

QMC has statistical errors and a sign problem but can employ a much larger number of states.

SQMC combines to some extent the advantages of the above by doing a deterministic projection in a small set of important states and stochastic projection in the rest of the space. It has a much smaller statistical error than stochastic projection and can employ a large number of states.

More generally Semistochastic Projection is an efficient way to find the dominant eigenvalue and corresponding expectation values of any large sparse matrix that has much of its spectral weight on a manageable number of states.

Semistochastic Projection



The part of the projection with both indices in the deterministic part is done deterministically. The part of the projection with either index in the stochastic part is done stochastically.

$$P = P^{\mathcal{D}} + P^{\mathcal{S}}$$

$$P^{\mathcal{D}}_{ij} = \begin{cases} P_{ij}, & i,j \in \mathcal{D} \\ 0, & \text{otherwise} \end{cases}$$

$$P^{\mathcal{S}} = P - P^{\mathcal{D}}$$

Diagonal elements in $P^{\mathcal{S}}$

The contribution to the walker weight on $|\phi_j\rangle$, with $j \in \mathcal{S}$, is

$$P_{jj} w_j(t) \quad = \quad [1 + \tau(E_T - H_{jj})]\, w_j(t)$$

Off-diagonal elements in $P^{\mathcal{S}}$

Weight w_i is divided amongst $n_i = \max(\lfloor w_i \rceil, 1)$ walkers of wt. w_i/n_i.
For each walker on $|\phi_i\rangle$, a move to $|\phi_j\rangle \neq |\phi_i\rangle$ is proposed with probability $T_{ji} > 0$, $(\sum_j T_{ji} = 1)$, where T is the proposal matrix.

The magnitude of the contribution to the walker weight on $|\phi_j\rangle$ from a single walker on $|\phi_i\rangle$ is

$$\begin{cases} 0, & i,j \in \mathcal{D} \\ \dfrac{P_{ji}}{T_{ji}} \dfrac{w_i(t)}{n_i(t)} = -\tau \dfrac{H_{ji}}{T_{ji}} \dfrac{w_i(t)}{n_i(t)} & \text{otherwise} \end{cases}$$

Elements in $P^{\mathcal{D}}$

The contribution to the weight on a deterministic state, $|\phi_j\rangle$, $(j \in \mathcal{D})$, from all deterministic states is simply

$$w_j(t+1) \;\; = \;\; \sum_{i \in \mathcal{D}} P^{\mathcal{D}}_{ji} \, w_i(t).$$

$P^{\mathcal{D}}$ is stored and applied as a sparse matrix

Construction of deterministic space and $\Psi_{\rm T}$

Construction of deterministic space and $\Psi_{\rm T}$ is done once and for all before start of MC run.

1. Start with a likely state, e.g., Hartree-Fock.
2. Construct all connected states if not too many, or, all connected states that involve excitations to lower lying orbitals
3. Diagonalize using Lanczos
4. Construct connections to the states with the highest absolute weights
5. iterate

For some systems iterating 2 or 3 times can give considerable gain.

Semistochastic Projection

Walkers have a label (bit string of orbital occupation numbers) and signed real weights.

Project Do deterministic and stochastic projection

Sort or Hash Walker labels are sorted.

Merge Walkers on the same determinant are merged

Initiator Use initiator criterion to discard some newly spawned walkers.

Join Because we use real weights, there are many walkers with small weights. Join stochastic space walkers on different determinants using unbiased algorithm.

Update Energy Used stored E_{L} components to update energy estimator. So E_{L} never needs to be computed during body of run.

The only additional steps are the deterministic projection and the "join" step.

SQMC

Some differences between SQMC and FCIQMC or PMC:

1. Deterministic projection in part of space
2. Multideterminantal $\Psi_{\rm T}$, particularly important for strongly correlated states
3. Real (rather than integer) weights, $|\psi(t)\rangle = \sum_{i=1}^{N} w_i(t)|\phi_i\rangle$
4. Graduated initiator, threshold $= i\ d^p$, where d is the number of moves since last visit to deterministic space (Usually choose, $i, p = 1$)

Test Cases

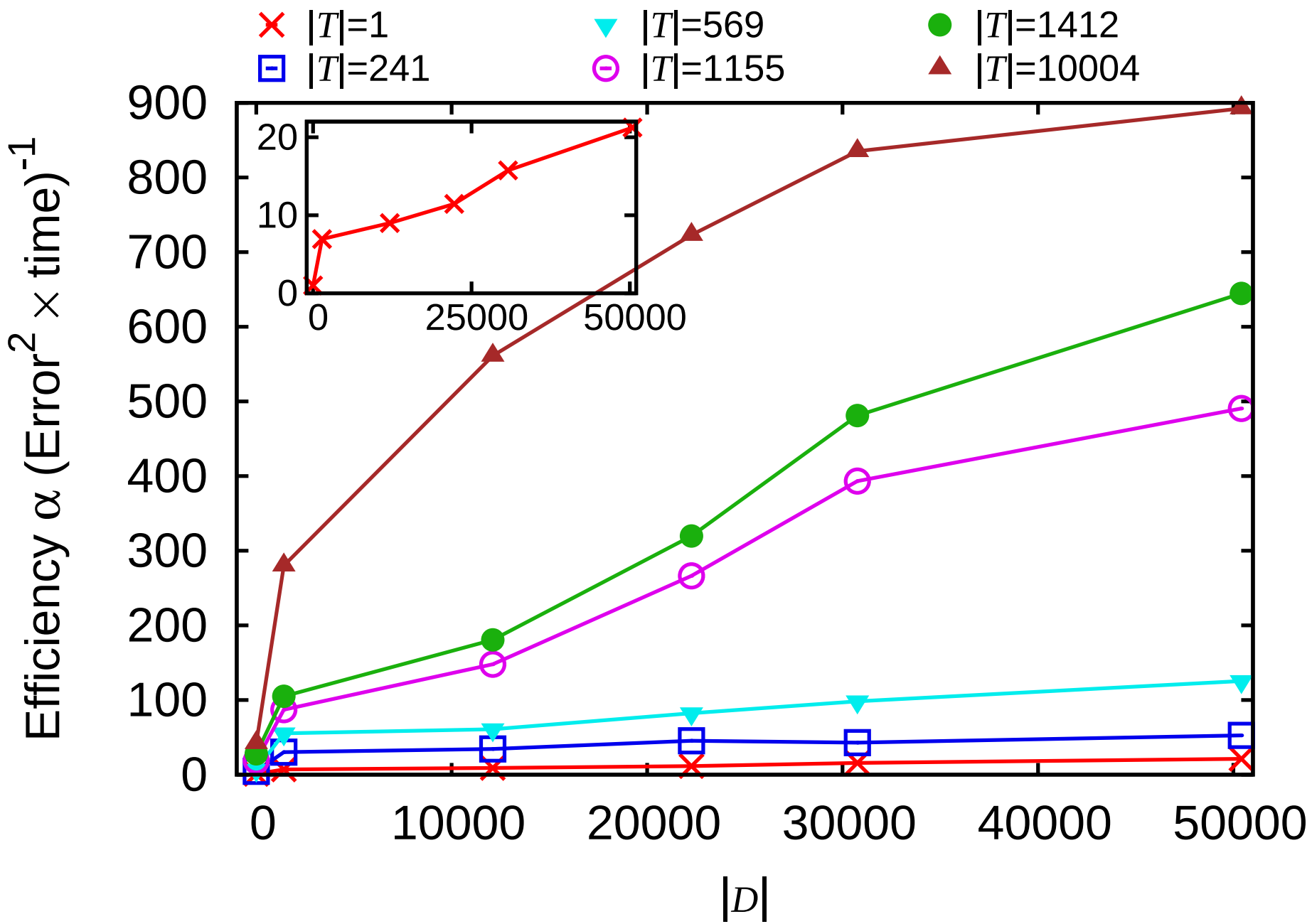
Test the ideas on:

1. 2-D Fermion Hubbard model on 8×8 lattice
2. small molecules

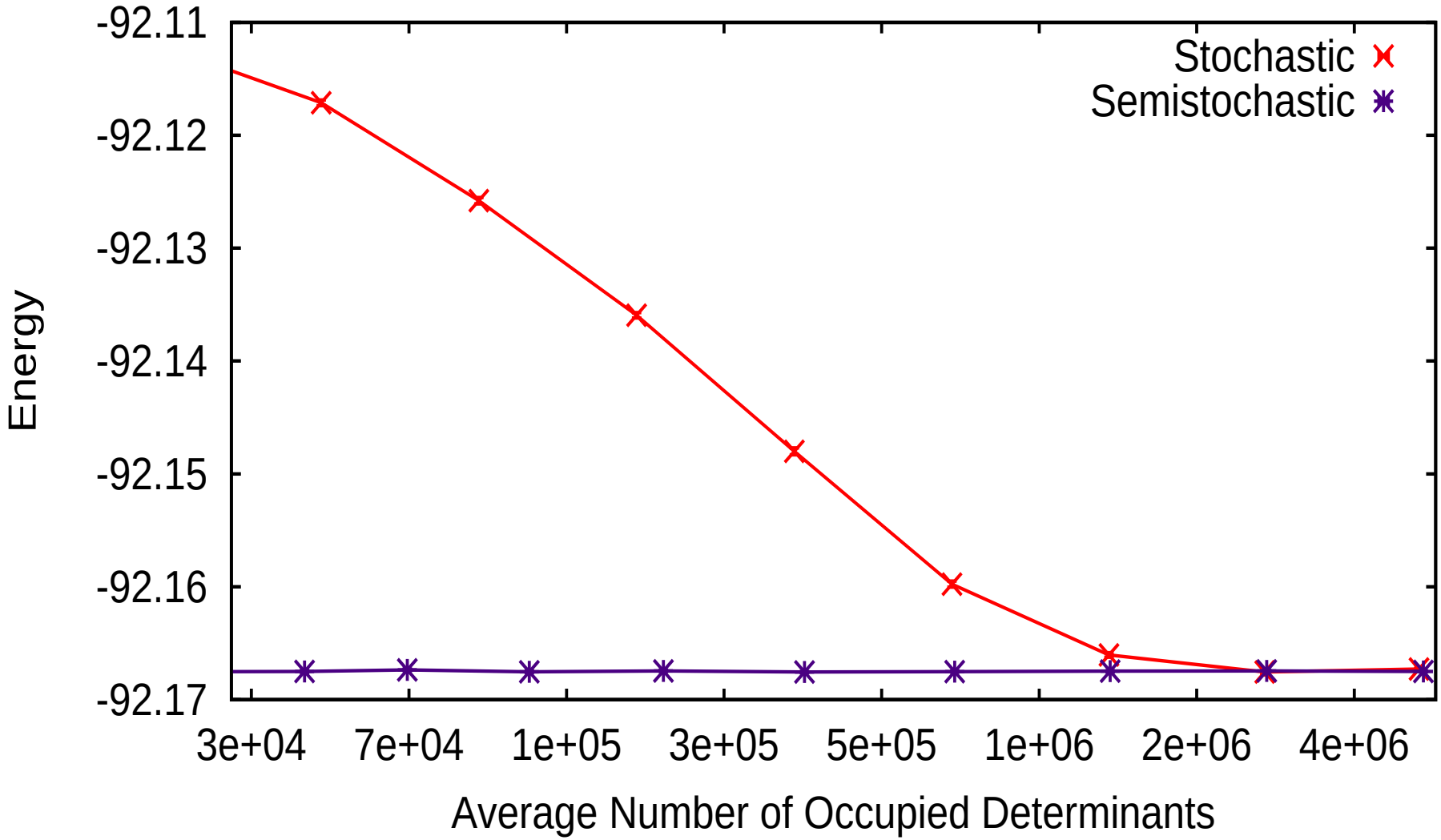
Why Hubbard?

1. Generally accepted as an interesting many-body system that exhibits a variety of phenomena and is extremely hard to solve.
2. Matrix elements can be computed quickly
3. Can go from very weakly correlated to very strongly correlated by turning a single knob, U. Large U model much more challenging than small molecules.
4. Can study effect of changing number of electrons, N, easily.

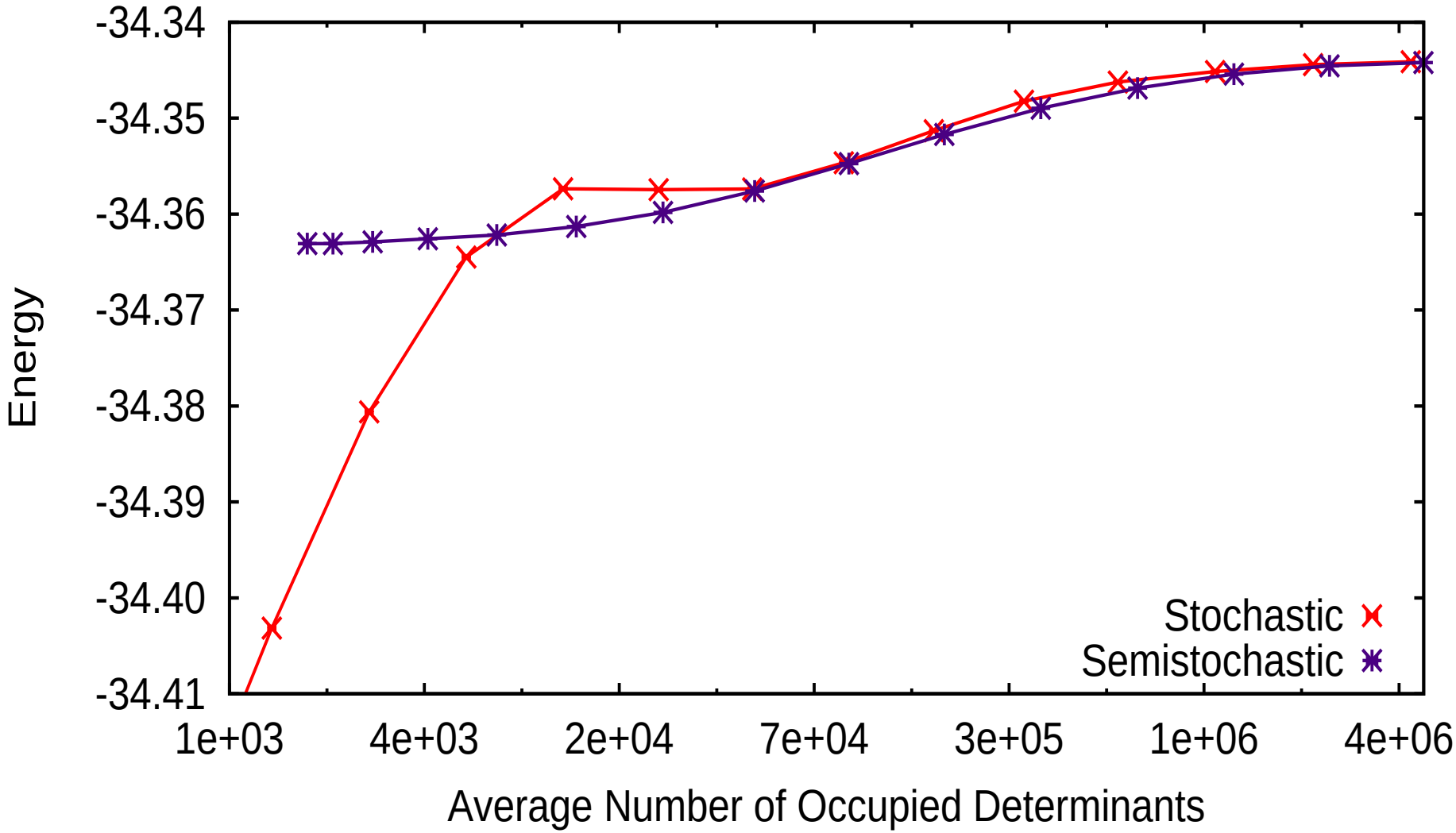
Efficiency Gains in 8×8 Hubbard Model, $N = 10$



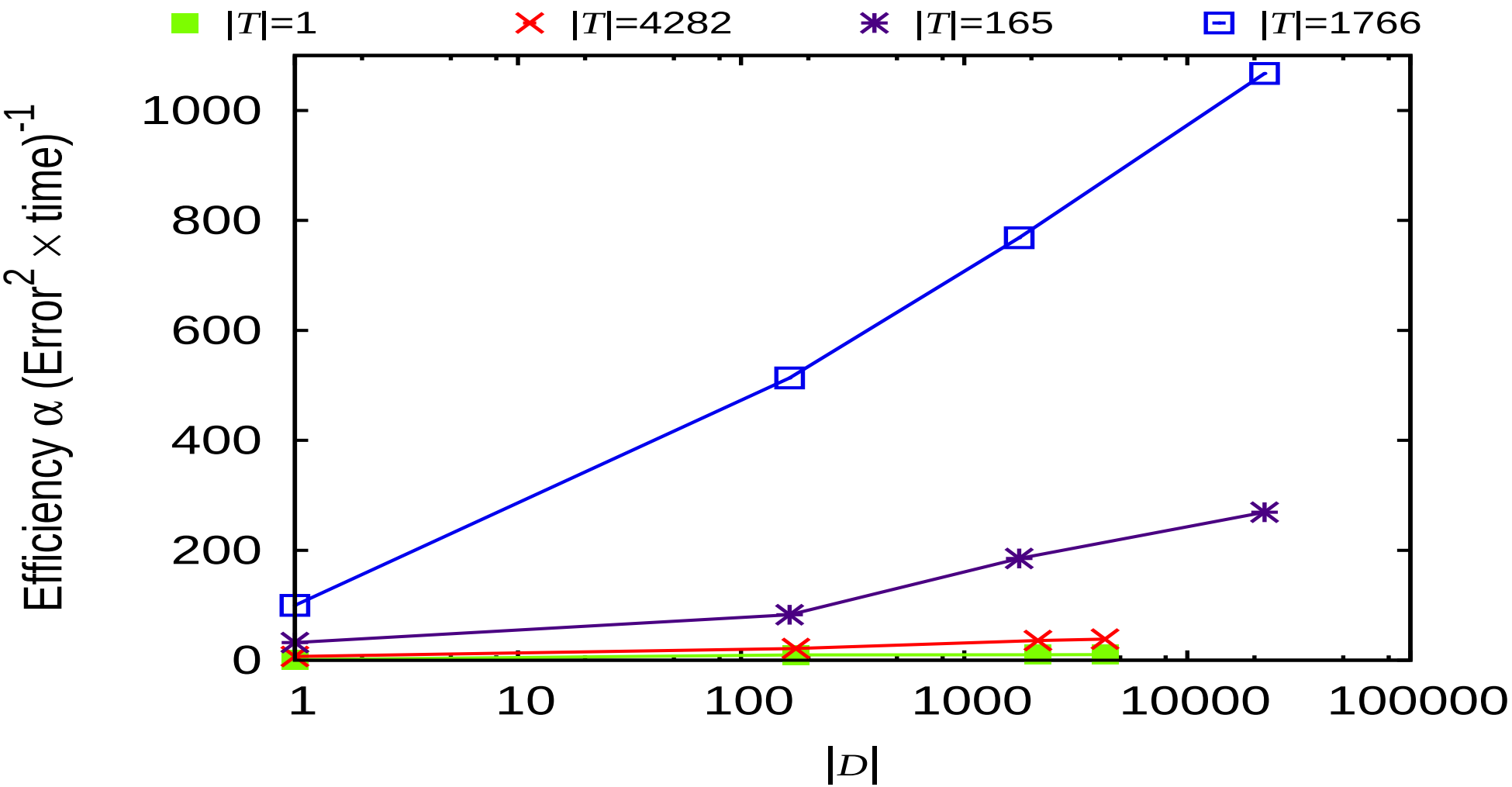
Energy versus average number of occupied determinants, 8×8 Hubbard, $N = 50$, $U = 1$



Energy versus average number of occupied determinants, 8×8 Hubbard, $N = 10$, $U = 4$



Efficiency gain for C_2 ($3-\zeta$ basis) from semistochastic projection and Ψ_T



Wavefns. with 165 or 1766 dets. containing some 4^{th}-order excit. are much more efficient than wavefn. with 4282 dets. containing only upto 2^{nd}-order excit.

Ongoing/Future Work on SQMC

Semistochastic projection plus multideterminantal $\Psi_{\rm T}$ results in about 3 orders of magnitude gain in efficiency.
In addition the initiator bias is often reduced.
Even with these improvements the method is very expensive.
However, there are still many improvements that can be made, including:

1. choice of basis, including using $\Psi_{\rm T}$ as a basis state
2. better deterministic space, trial wavefunctions, $\Psi_{\rm T}$, and, guiding wavefunctions, $\Psi_{\rm G}$, e.g., Matrix Product States (Garnet Chan), Coupled Cluster (Alessandro Roggero and Francesco Pederiva)
3. use exponential projector to increase average time step (Bryan Clark, Alessandro Roggero and Francesco Pederiva)
4. use F12 methods to improve basis convergence (with Takeshi Yanai, Garnet Chan, George Booth, Sandeep Sharma, Miguel Morales)
5. embedding (Garnet Chan, George Booth)
6. excited states: 1) projecting out lower states (Ohtsuka and Nagase,
   2) dividing Hilbert space into a small and a large piece and calculating an effective Hamiltonian in the small space, Ten-no,
   3) using modified projector, $1+\tau(E_T-\hat{H})^2$, to target desired state, Booth and Chan.

Derivatives of FN-PMC energies

Examples of derivatives of interest are:

1. derivative wrt parameters of the wavefn., needed to optimize the VMC or PMC energy for given geometry
2. derivative wrt nuclear coordinates, needed to optimize the geometry

Almost all errors reduced by optimizing trial wavefunctions

1. Statistical error in $E_{\rm VMC}$ and $E_{\rm DMC}$ (both the rms fluctuations of $E_{\rm L}$ and the autocorrelation time)
2. $E_{\rm VMC}$
3. Fixed-node error in $E_{\rm DMC}$ (nodes move during optimization). Fixed node errors can be LARGE. For C_2, FN error is 1.3 eV for total energy and 0.7 eV for well-depth. However, optimized multidet. wavefn has FN error that is better than chemical accuracy.
4. Time-step error in DMC (from Trotter-Suzuki approximation)
5. Population control error in DMC
6. Pseudopotential locality error in DMC when using nonlocal pseudopotentials
7. Error of observables that do not commute with the Hamiltonian (mixed estimators, $\langle\psi_{\rm FN}|\hat{A}|\psi\rangle$ not exact even for nodeless $\psi_{\rm FN}$, ψ).

Measures of goodness of variational wave functions

$$\min \; E_{\rm VMC} = \frac{\langle\psi|H|\psi\rangle}{\langle\psi|\psi\rangle} = \langle E_{\rm L}\rangle_{\psi^2}$$

$$\min \; \sigma^2_{\rm VMC} = \frac{\langle\psi|(H-E_{\rm T})^2|\psi\rangle}{\langle\psi|\psi\rangle} = \langle E^2_{\rm L}(\mathbf{R}_i)\rangle_{\psi^2} - \langle E_{\rm L}(\mathbf{R}_i)\rangle^2_{\psi^2}$$

$$\max \quad \Omega^2 = \frac{|\langle\psi_{\rm FN}|\psi\rangle|^2}{\langle\psi_{\rm FN}|\psi_{\rm FN}\rangle\langle\psi|\psi\rangle} = \frac{\left\langle \frac{\psi_{\rm FN}}{\psi}\right\rangle^2_{\psi^2}}{\left\langle \left|\frac{\psi_{\rm FN}}{\psi}\right|^2\right\rangle_{\psi^2}}$$

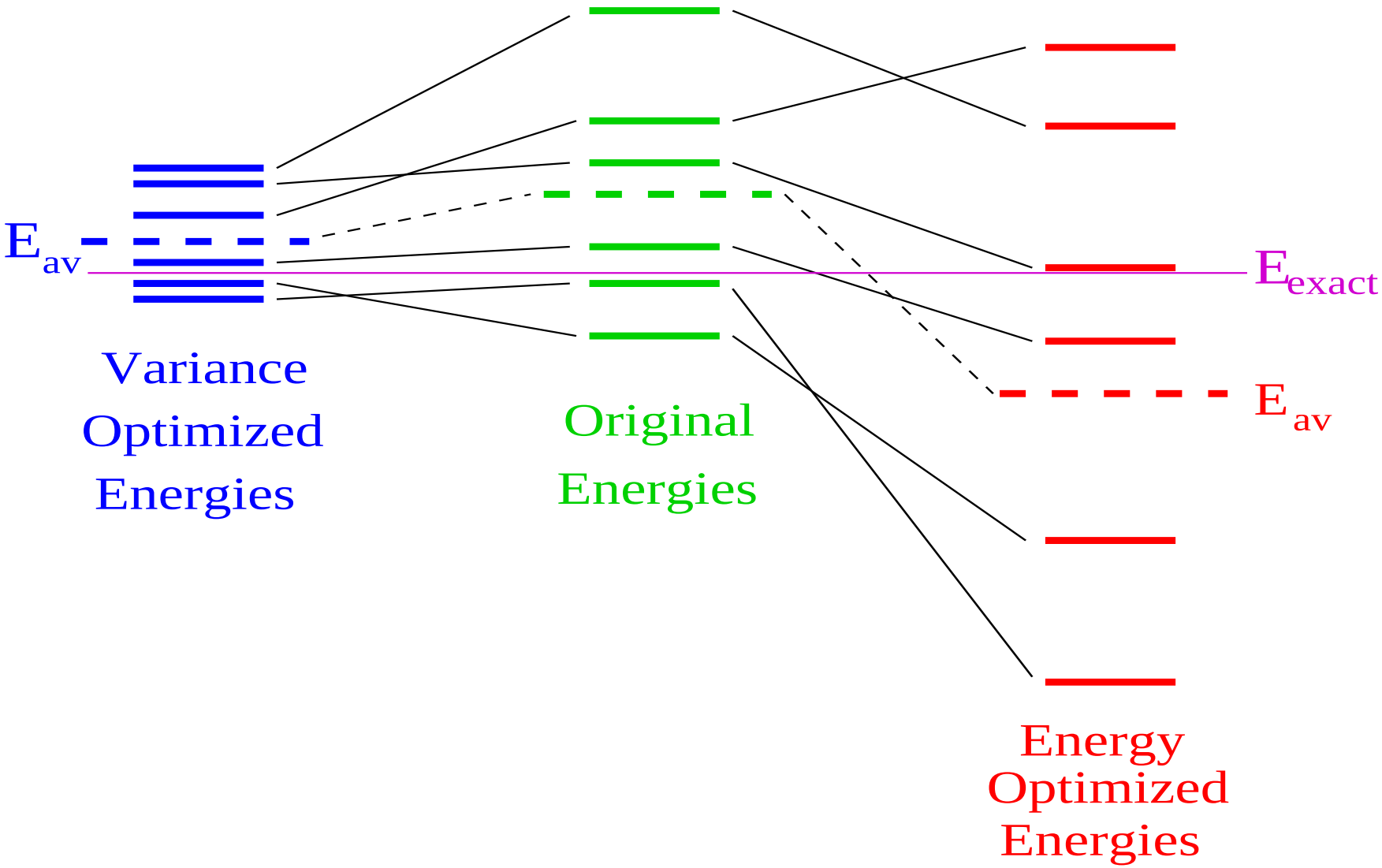
$$\min \; E_{\rm DMC} = \frac{\langle\psi_{\rm FN}|H|\psi\rangle}{\langle\psi_{\rm FN}|\psi\rangle} = \langle E_{\rm L}\rangle_{|\psi\psi_{\rm FN}|}$$

For an infinitely flexible wave function all optimizations will yield the exact wavefunction (except that minimizing σ could yield an excited state) but for the imperfect functional forms used in practice they differ.

Variance vs. Energy

$$\sigma^2 = \sum_{i=1}^{N_{\rm conf}} \left(\frac{\mathcal{H}\Psi_{\rm T}(\mathbf{R}_i)}{\Psi_{\rm T}(\mathbf{R}_i)} - \bar{E} \right)^2$$

$$\bar{E} = \sum_{i=1}^{N_{\rm conf}} \frac{\mathcal{H}\Psi_{\rm T}(\mathbf{R}_i)}{\Psi_{\rm T}(\mathbf{R}_i)}$$



Take-home Message

Energy optimization methods that minimize the energy evaluated on finite sample will yield poor energies on other samples, unless the sample used to do the minimization is very large.

So, efficient energy optimization methods do NOT optimize the energy evaluated on a finite sample, although they do minimize the energy in the limit of an infinite sample.

Ingredients of efficient optimization methods

Efficient energy and variance optimization methods are based on standard optimization methods, the Newton method, and, the linear method (generalized eigenvalue problem), but with significant extensions, mostly to allow them to work in a stochastic approach.

1. Add terms that have zero expectation value for an infinite sample, but, greatly reduce the noise for a finite sample. CJU and C. Filippi, PRL 2005
2. Although the true Hamiltonian is symmetric, for a finite sample a nonsymmetric Hamiltonian satisfies a strong zero-variance principle and gives much smaller fluctuations. If the space is closed under the action of H then there is no noise the parameters, regardless of the sample, provided that it is larger than the number of parameters. M.P. Nightingale and Melik-Alaverdian, PRL 2001
3. Extension of the linear method to nonlinear parameters. CJU, J. Toulouse, C. Filippi and S. Sorella, PRL 2007; J. Toulouse and CJU JCP 2007, 2008
4. Automatic procedure for choosing size of moves and recovering from bad moves.

Derivatives of the Energy

Examples of derivatives of interest are:

1. derivative wrt parameters of the wavefn., needed to optimize the VMC or PMC energy for given geometry
2. derivative wrt nuclear coordinates, needed to optimize the geometry

$$E = \frac{\int d\mathbf{R}\ \rho(\mathbf{R})\ E_{\rm L}(\mathbf{R})}{\int d\mathbf{R}\ \rho(\mathbf{R})} \approx \frac{\sum_{n=N_{\rm eq}+1}^{N_{\rm eq}+N_{\rm MC}} E_{\rm L}(\mathbf{R}_n)}{N_{\rm MC}} \equiv \langle E_{\rm L}(\mathbf{R}_n)\rangle_\rho$$

where $\langle X(\mathbf{R}_n)\rangle_\rho$ denotes $\frac{\sum_{n=N_{\rm eq}+1}^{N_{\rm eq}+N_{\rm MC}} X(\mathbf{R}_n)}{N_{\rm MC}}$, with points $\mathbf{R}_n$ sampled from $\frac{\rho(\mathbf{R})}{\int d\mathbf{R}\ \rho(\mathbf{R})}$.

Denoting the derivative wrt the i^{th} parameter by subscript i,

$$\begin{aligned} E_i &= \frac{\int d\mathbf{R}\ \rho(\mathbf{R})\ \left(\left(E_{\rm L}(\mathbf{R}) - E\right) \frac{\rho_i(\mathbf{R})}{\rho(\mathbf{R})} \ +\ E_{{\rm L},i}(\mathbf{R}) \right)}{\int d\mathbf{R}\ \rho(\mathbf{R})} \\ &= \frac{\sum_{n=N_{\rm eq}+1}^{N_{\rm eq}+N_{\rm MC}} \left(\left(E_{\rm L}(\mathbf{R}_n) - E\right) \frac{\rho_i(\mathbf{R}_n)}{\rho(\mathbf{R}_n)} + E_{{\rm L},i}(\mathbf{R}_n) \right)}{N_{\rm MC}} \\ &\equiv \left\langle \left(E_{\rm L}(\mathbf{R}_n) - E\right) \frac{\rho_i(\mathbf{R}_n)}{\rho(\mathbf{R}_n)} \right\rangle_\rho + \left\langle E_{{\rm L},i}(\mathbf{R}_n) \right\rangle_\rho \end{aligned}$$

Derivatives of the VMC Energy

$$E_i = \left\langle \left(E_{\rm L}(\mathbf{R}_n) - E \right) \frac{\rho_i(\mathbf{R}_n)}{\rho(\mathbf{R}_n)} \right\rangle_\rho + \langle E_{{\rm L},i}(\mathbf{R}_n) \rangle_\rho$$

VMC: ρ known explicitly: $\rho(\mathbf{R}) = \psi^2(\mathbf{R})$

$$E_i = 2 \left\langle \left(E_{\rm L}(\mathbf{R}_n) - E \right) \frac{\psi_i(\mathbf{R}_n)}{\psi(\mathbf{R}_n)} \right\rangle_{\psi^2} + \langle E_{{\rm L},i}(\mathbf{R}_n) \rangle_{\psi^2}$$

If $\hat{H}$ is indep of parameter i, then $\langle E_{{\rm L},i}(\mathbf{R}_n) \rangle_{\psi^2} = 0$ by hermiticity, so

$$E_i = 2 \left\langle \left(E_{\rm L}(\mathbf{R}_n) - E \right) \frac{\psi_i(\mathbf{R}_n)}{\psi(\mathbf{R}_n)} \right\rangle_{\psi^2}$$

Straightforward derivatives of the PMC Energy

$$E_i = \left\langle \left(E_{\rm L}(\mathbf{R}_n) - E\right) \frac{\rho_i(\mathbf{R}_n)}{\rho(\mathbf{R}_n)} \right\rangle_{\psi\psi_{\rm FN}} + \langle E_{{\rm L},i}(\mathbf{R}_n)\rangle_{\psi\psi_{\rm FN}}$$

PMC: ρ from path integral: $\rho(\mathbf{R}_n) = \int \left(\prod_{k=0}^{n-1} d\mathbf{R}_k \; P(\mathbf{R}_{k+1}, \mathbf{R}_k)\right) \tilde{\rho}(\mathbf{R}_0)$

$$E_i = \left\langle \left(E_{\rm L}(\mathbf{R}_n) - E\right) \left(\sum_{k=0}^{n-1} \frac{P_i(\mathbf{R}_{k+1}, \mathbf{R}_k)}{P(\mathbf{R}_{k+1}, \mathbf{R}_k)} + \frac{\tilde{\rho}_i(\mathbf{R}_0)}{\tilde{\rho}(\mathbf{R}_0)}\right) \right\rangle_{\psi\psi_{\rm FN}} + \left\langle E_{{\rm L},i}(\mathbf{R}_n) \right\rangle_{\psi\psi_{\rm FN}}$$

$$= \left\langle \left(E_{\rm L}(\mathbf{R}_n) - E\right) \sum_{k=0}^{n-1} \frac{P_i(\mathbf{R}_{k+1}, \mathbf{R}_k)}{P(\mathbf{R}_{k+1}, \mathbf{R}_k)} \right\rangle_{\psi\psi_{\rm FN}} + \left\langle E_{{\rm L},i}(\mathbf{R}_n) \right\rangle_{\psi\psi_{\rm FN}} \quad \because \mathrm{covar}\left(E_{\rm L}(\mathbf{R}_n), \frac{\tilde{\rho}_i(\mathbf{R}_0)}{\tilde{\rho}(\mathbf{R}_0)}\right) = 0$$

Log derivatives of both the stochastic part (drift-diffusion for DMC) and the reweighting part of P appear.

Note that although $\left(\sum_{k=0}^{n-1} \frac{P_i(\mathbf{R}_{k+1},\mathbf{R}_k)}{P(\mathbf{R}_{k+1},\mathbf{R}_k)}\right) \to \infty$ for $n \to \infty$,

$\left\langle \left(E_{\rm L}(\mathbf{R}_n) - E\right) \left(\sum_{k=0}^{n-1} \frac{P_i(\mathbf{R}_{k+1},\mathbf{R}_k)}{P(\mathbf{R}_{k+1},\mathbf{R}_k)}\right) \right\rangle_{\psi\psi_{\rm FN}}$ is finite. However it has infinite variance, so replace $\sum_{k=0}^{n-1}$ by $\sum_{k=n-m}^{n-1}$, where m is a few times $T_{\rm corr}$.

Exact but noisy.

Efficient derivatives of the PMC Energy

More efficient method can be derived using:

1. The FN projector can be written as the product of a stochastic matrix, T, and a diagonal reweighting matrix. (In continuum, a diagonal matrix on each side of stochastic matrix.)
2. The equilibrium distribution of the stochastic matrix is known!

Continuous Real Space (DMC): equil. dist.: $\psi^2(\mathbf{R})$

Follows from plugging in $\psi^2(\mathbf{R})$ into importance-sampled Schrödinger Eq. omitting reweighting term.

Discrete Space: equil. dist.: $\psi^2(\mathbf{R})\, P_{\rm L}(\mathbf{R})$

$$T(\mathbf{R}_i, \mathbf{R}_j) = \frac{\tilde{P}(\mathbf{R}_i, \mathbf{R}_j)}{P_{\rm L}(\mathbf{R}_j)}, \quad P_{\rm L}(\mathbf{R}_j) = \sum_i \tilde{P}(\mathbf{R}_i, \mathbf{R}_j) = \frac{\sum_i \psi(\mathbf{R}_i) P(\mathbf{R}_i, \mathbf{R}_j)}{\psi(\mathbf{R}_j)}$$ by def. of stochastic matri

Proof:

$$\sum_j T(\mathbf{R}_i, \mathbf{R}_j)\, \psi^2(\mathbf{R}_j)\, P_{\rm L}(\mathbf{R}_j) = \sum_j \psi(\mathbf{R}_i) \frac{P(\mathbf{R}_i, \mathbf{R}_j)}{\cancel{P_{\rm L}(\mathbf{R}_j)}} \frac{1}{\cancel{\psi(\mathbf{R}_j)}} \psi^{\cancel{2}}(\mathbf{R}_j) \cancel{P_{\rm L}(\mathbf{R}_j)} = \psi(\mathbf{R}_i) \sum_j P(\mathbf{R}_i, \mathbf{R}_j) \psi(\mathbf{R}_j)$$

$$= \psi^2(\mathbf{R}_i) P_{\rm L}(\mathbf{R}_i) \qquad \text{QED}$$

Efficient derivatives of the PMC Energy

Continuous Space:

$$E = \frac{\left\langle \psi^2(\mathbf{R}_n)\, W_n\, E_{\rm L}(\mathbf{R}_n)\right\rangle}{\left\langle\, \psi^2(\mathbf{R}_n)\, W_n\,\right\rangle} = \left\langle E_{\rm L}(\mathbf{R}_n)\right\rangle_{\psi\psi_{\rm FN}} \tag{1}$$

where $W_n = \prod_{k=n-m}^{n-1} w(\mathbf{R}_{k+1},\mathbf{R}_k)$

$$E_i = \left\langle \Big(E_{\rm L}(\mathbf{R}_n) - E\Big)\left(2\frac{\psi_i(\mathbf{R}_n)}{\psi(\mathbf{R}_n)} + \sum_{k=n-m}^{n-1}\frac{w_i(\mathbf{R}_{k+1},\mathbf{R}_k)}{w(\mathbf{R}_{k+1},\mathbf{R}_k)}\right)\right\rangle_{\psi\psi_{\rm FN}} + \left\langle E_{{\rm L},i}(\mathbf{R}_n)\right\rangle_{\psi\psi_{\rm FN}}$$

Discrete Space:

$$E = \frac{\left\langle \psi^2(\mathbf{R}_n)\, P_{\rm L}(\mathbf{R}_n)\, W_n\, E_{\rm L}(\mathbf{R}_n)\right\rangle}{\left\langle\, \psi^2(\mathbf{R}_n)\, P_{\rm L}(\mathbf{R}_n)\, W_n\,\right\rangle} = \left\langle E_{\rm L}(\mathbf{R}_n)\right\rangle_{\psi\psi_{\rm FN}} \tag{2}$$

where $W_n = \prod_{k=n-m}^{n-1} w(\mathbf{R}_{k+1},\mathbf{R}_k) = \prod_{k=n-m}^{n-1} P_{\rm L}(\mathbf{R}_k)$, so, $P_{\rm L}(\mathbf{R}_n)\, W_n = \prod_{k=n-m}^{n} P_{\rm L}(\mathbf{R}_k)$

$$E_i = \left\langle \Big(E_{\rm L}(\mathbf{R}_n) - E\Big)\left(2\frac{\psi_i(\mathbf{R}_n)}{\psi(\mathbf{R}_n)} + \sum_{k=n-m}^{n}\frac{P_{{\rm L},i}(\mathbf{R}_k)}{P_{\rm L}(\mathbf{R}_k)}\right)\right\rangle_{\psi\psi_{\rm FN}} + \left\langle E_{{\rm L},i}(\mathbf{R}_n)\right\rangle_{\psi\psi_{\rm FN}}$$

Questions

1. Are the above expressions for E_i exact or approximate? When Claudia Filippi and I presented the finite-difference version of this for forces in 2000, we said it was approximate, and everyone else subsequently said it was approximate, but I don't see why any more. In Eqs. 2 and 2 the only factors that do not depend on $\mathbf{R}_n$ are the W_n. So one can average over all paths that end at $\mathbf{R}_n$ and there is no issue of product of averages not being the average of products.
2. Of the 3 terms in the expressions for E_i, the 2^{nd} term is the noisiest one since it involves a sum over the path. However, I have 2 interesting observations, based on which it may be possible to make an approximation that eliminates that term.
   1. When the fixed-node projector is independent of the parameter being varied, the 2^{nd} term equals the 3^{rd} term and each equals $-1/2$ the 1^{st} term. (Of course the sum must be zero, since, if projector is independent of the parameter, the derivative of the energy must be zero.
   2. When $\psi = \psi_0$, of course $E_i = 0$ and in fact each of the 3 terms is 0. As $\psi \to \psi_0$, the 2^{nd} and 3^{rd} terms are equal to linear order in the deviation. This is an empirical observation. Can we prove it?

   Based on these 2 observations (first one is provable), it may be a reasonably good approximation to just replace the 2^{nd} term by the 3^{rd} term.

Collaborators

SQMC

Idea born in discussion with:
M. Peter Nightingale, Physics, University of Rhode Island

Most of the implementation done by graduate students:
Frank Petruzielo, Physics, Cornell
Hitesh Changlani, Physics, Cornell
Adam Holmes, Physics, Cornell

Discussions with Bryan Clark, George Booth, Ali Alavi, Garnet Chan, Shiwei Zhang.

Funding from NSF and DOE.

Derivatives of the FN-PMC energy

Graduate student: John Elton
Discussions with Roland Assaraf, Julien Toulouse and Saverio Moroni.