Semistochastic Quantum Monte Carlo – A Hybrid of Exact Diagonalization and QMC Methods

Cyrus Umrigar

Physics Department, Cornell University, Ithaca.

Email: CyrusUmrigar@cornell.edu

June 12, 2013, ES13, College of William and Mary, Williamsburg

Outline

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

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

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

The problem

We wish to find the lowest energy eigenstate(s) of a (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 complete 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 will be used to calculate variational and mixed estimators of operators $\hat{A}$, i.e., $\langle\Psi_{\rm T}|\hat{A}|\Psi_{\rm T}\rangle$, $\langle\Psi_{\rm T}|\hat{A}|\Psi_0\rangle$

Ψ_G will be used to alter the probability density sampled, i.e., $\Psi_{\rm G}^2$ in VMC, and $\Psi_{\rm G}\Psi_0$ in PMC. It affects the statistical error of the mixed and the growth estimators.

$\Psi_{\rm G}$ 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 $\Psi_{\rm T}$. To simplify expressions, we use $\Psi_{\rm G} = \Psi_{\rm T}$ in what follows.

Variational MC

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

$$= \frac{\sum_{ij}^{N_{\rm st}} t_i H_{ij} t_j}{\sum_{k}^{N_{\rm st}} t_k^2} = \sum_{i}^{N_{\rm st}} \frac{t_i^2}{\sum_{k}^{N_{\rm st}} t_k^2} \, \frac{\sum_{j}^{N_{\rm st}} H_{ij} t_j}{t_i}$$

$$= \sum_{i}^{N_{\rm st}} \frac{t_i^2}{\sum_{k}^{N_{\rm st}} t_k^2} \, E_{\rm L}(i) \; = \; \frac{\sum_{i}^{N_{\rm MC}} E_{\rm L}(i)}{N_{\rm MC}}$$

Sample probability density function $\frac{t_i^2}{\sum_{k}^{N_{\rm st}} t_k^2}$ using Metropolis-Hastings.

Value and statistical error depend on $\Psi_{\rm T}$.

Energy bias and statistical error vanish as $\Psi_{\rm T} \to \Psi_0$.

Projector MC

Pure and Mixed estimators for energy are equal:

$$E_0 = \frac{\langle \Psi_0 | \hat{H} | \Psi_0 \rangle}{\langle \Psi_0 | \Psi_0 \rangle} = \frac{\langle \Psi_0 | \hat{H} | \Psi_{\rm T} \rangle}{\langle \Psi_0 | \Psi_{\rm T} \rangle}$$

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

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

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

$$= \sum_{i}^{N_{\rm st}} \frac{e_i t_i}{\sum_{k}^{N_{\rm st}} e_k t_k} \, E_{\rm L}(i) \;=\; \frac{\sum_{i}^{N_{\rm MC}} E_{\rm L}(i)}{N_{\rm MC}}$$

Value exact for Bosons. Statistical error depends on $\Psi_{\rm T}$.

In both VMC and PMC we average the *configuration value of* $\hat{H}$ or *local energy,* $E_{\rm L}(i)$, but from points sampled from different distributions.

Variational and Projector MC

$$E_V = \sum_i^{N_{\rm st}} \frac{t_i^2}{\sum_k^{N_{\rm st}} t_k^2} E_{\rm L}(i) = \sum_i^{N_{\rm MC}} E_{\rm L}(i) \quad \text{(Value and error depend on } \Psi_{\rm T}\text{)}$$

$$E_0 = \sum_i^{N_{\rm st}} \frac{e_i t_i}{\sum_k^{N_{\rm st}} e_k t_k} E_{\rm L}(i) = \sum_i^{N_{\rm MC}} E_{\rm L}(i) \quad \text{(Value exact for Bosons. Error depends on } \Psi_{\rm T}\text{.}$$

$$E_{\rm L}(i) = \frac{\sum_j^{N_{\rm st}} H_{ij} t_j}{t_i}$$

This is practical for systems that are large enough to be interesting if

1. $t_i = \langle \phi_i | \Psi_{\rm T} \rangle$ can be evaluated in polynomial time, say N^3
2. the sum in $E_{\rm L}(i)$ can be done quickly, i.e., $\hat{H}$ is sparse (discrete) or semi-diagonal (continuous).

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).

Method	Projector	SP Basis	Quantiz
Diffusion Monte Carlo	$e^{\tau(E_T\hat{\mathbf{1}}-\hat{H})}$	$\vec{r}$	1^{st}
GFMC (Kalos, Ceperley, Schmidt)	$\frac{1}{\hat{\mathbf{1}}-\tau(E_T\hat{\mathbf{1}}-\hat{H})}$	$\vec{r}$	1^{st}
LRDMC (Sorella, Casula)	$\hat{\mathbf{1}}+\tau(E_T\hat{\mathbf{1}}-\hat{H})$	$\vec{r}_i$	1^{st}
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 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.

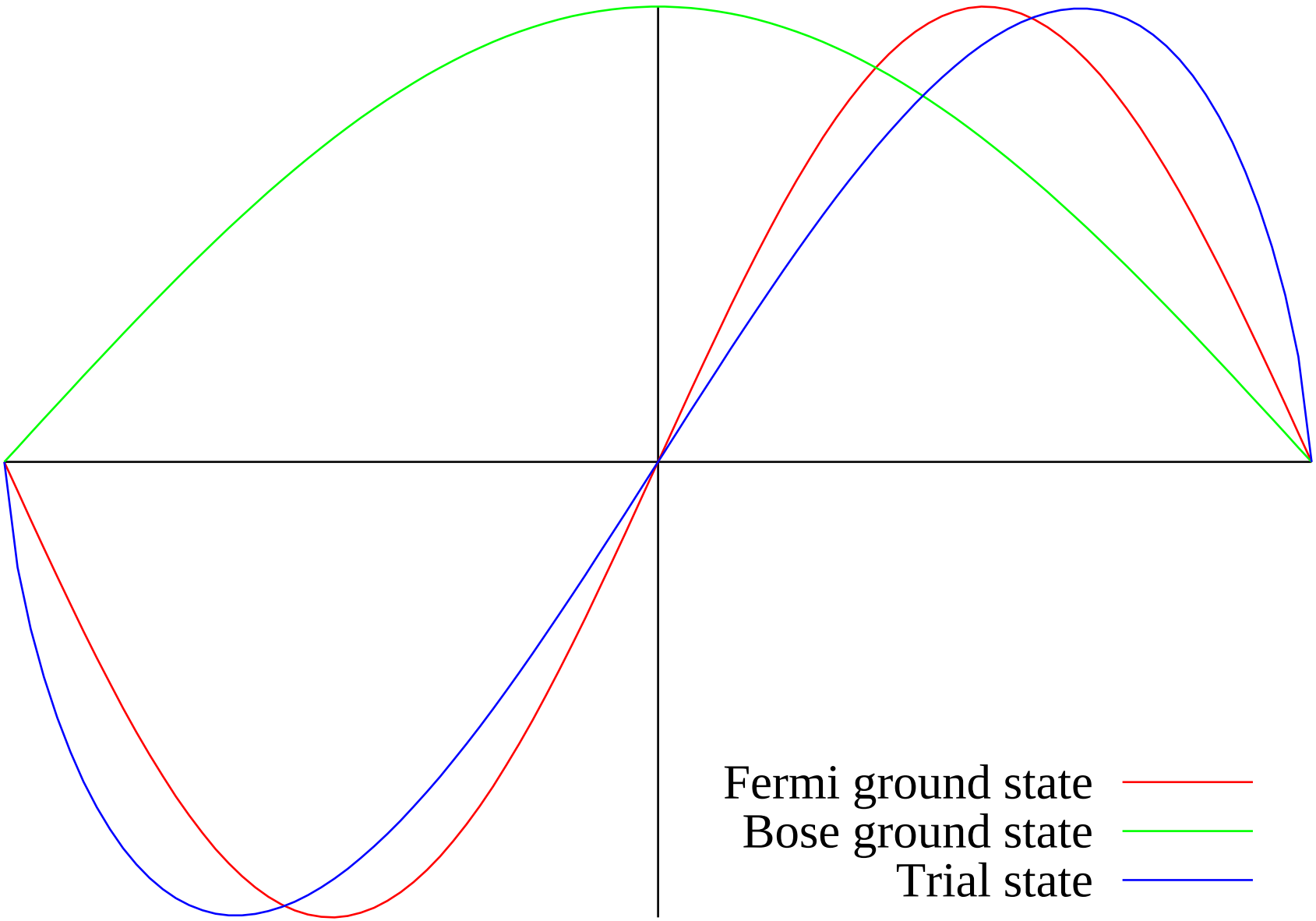
Problem: However, since the Bosonic energy is always lower than the Fermionic energy, the projected state is the Bosonic ground state.

Fixed-node approximation

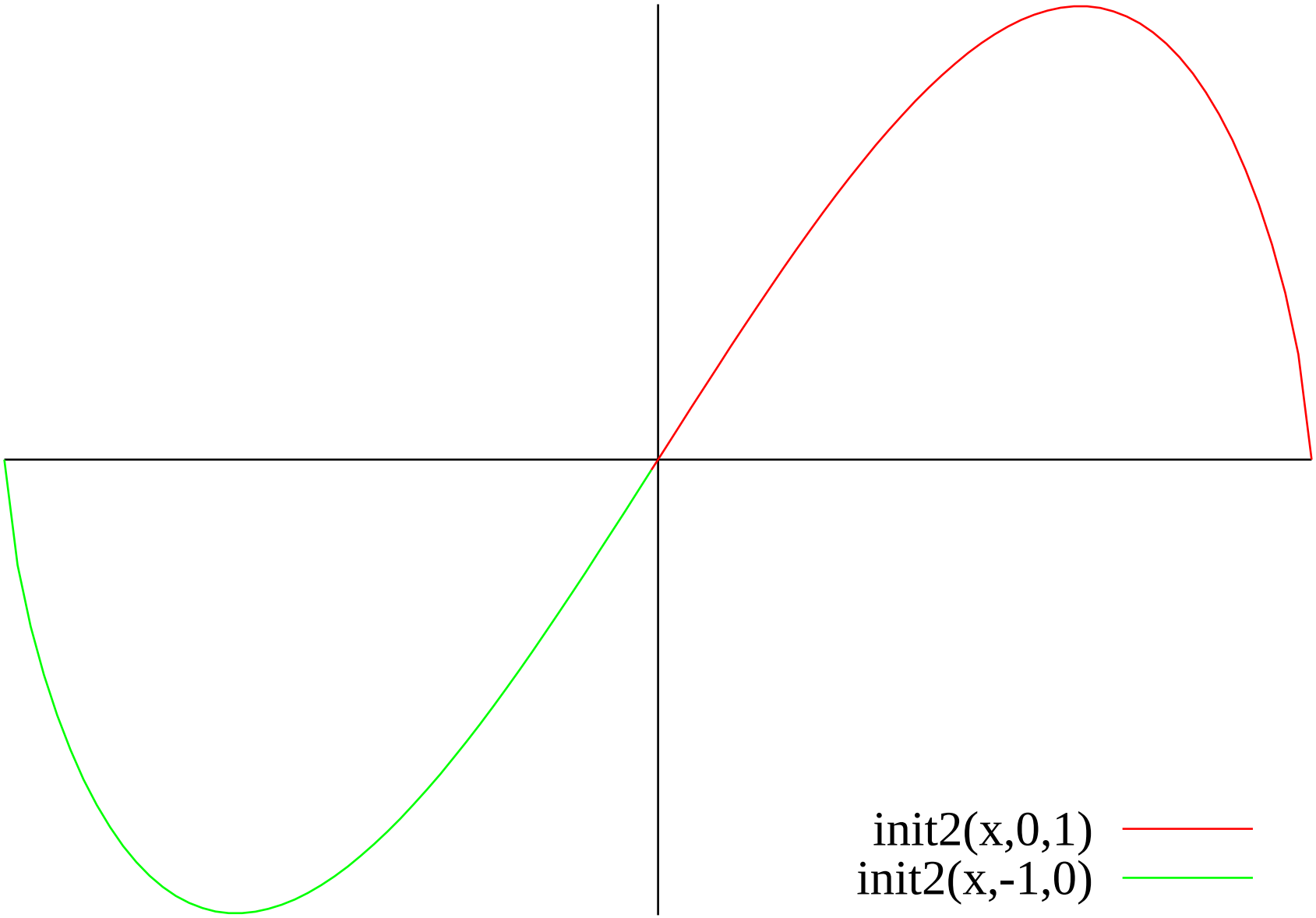
All except a few calculations (release-node, Ceperley) are done using FN approximation. Instead of doing a free projection, impose the boundary condition that the projected state has the same nodes as the trial state $\Psi_{\mathrm{T}}(\mathbf{R})$.

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

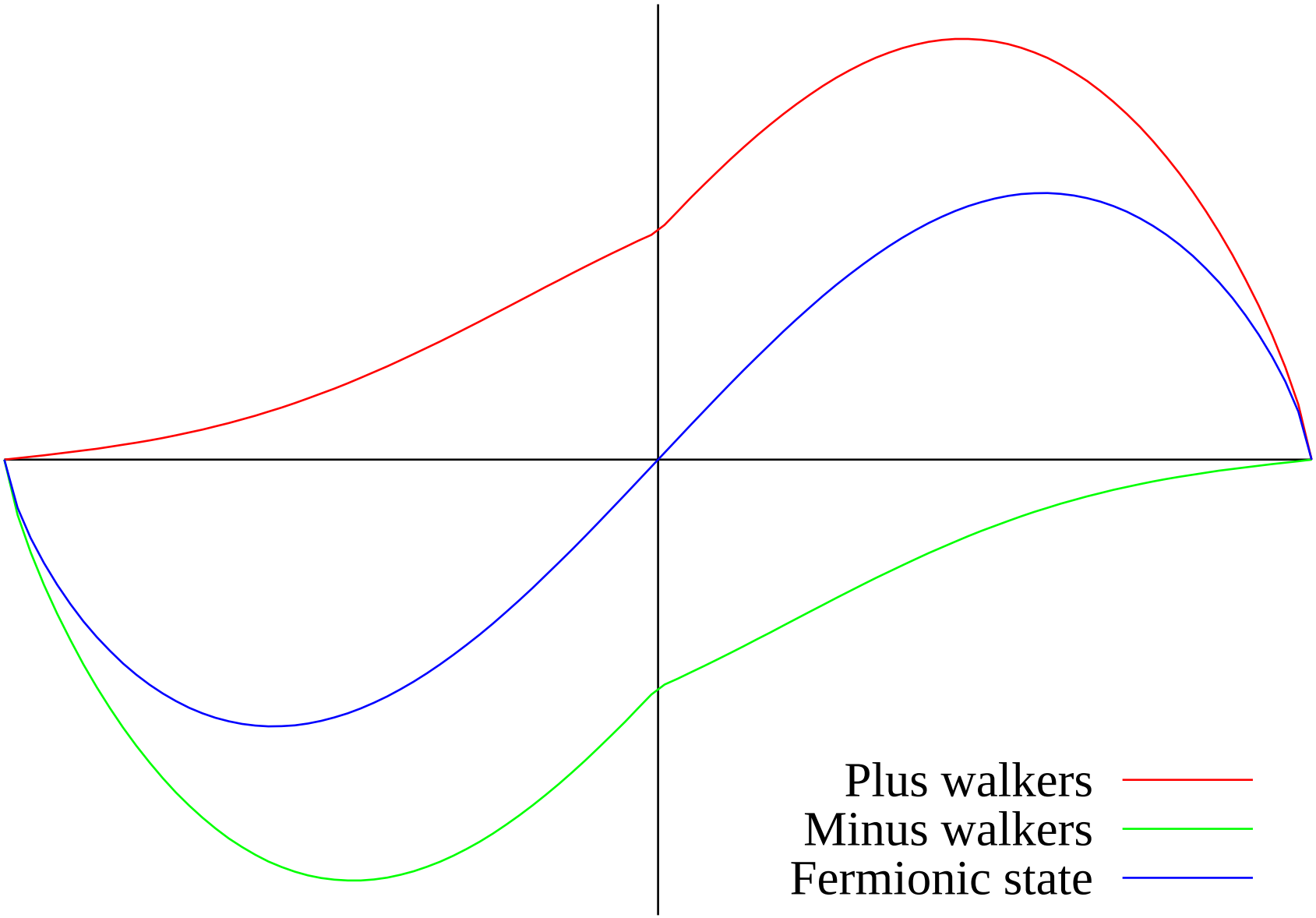
Sign Problem in 1^{st} Quantization and R space



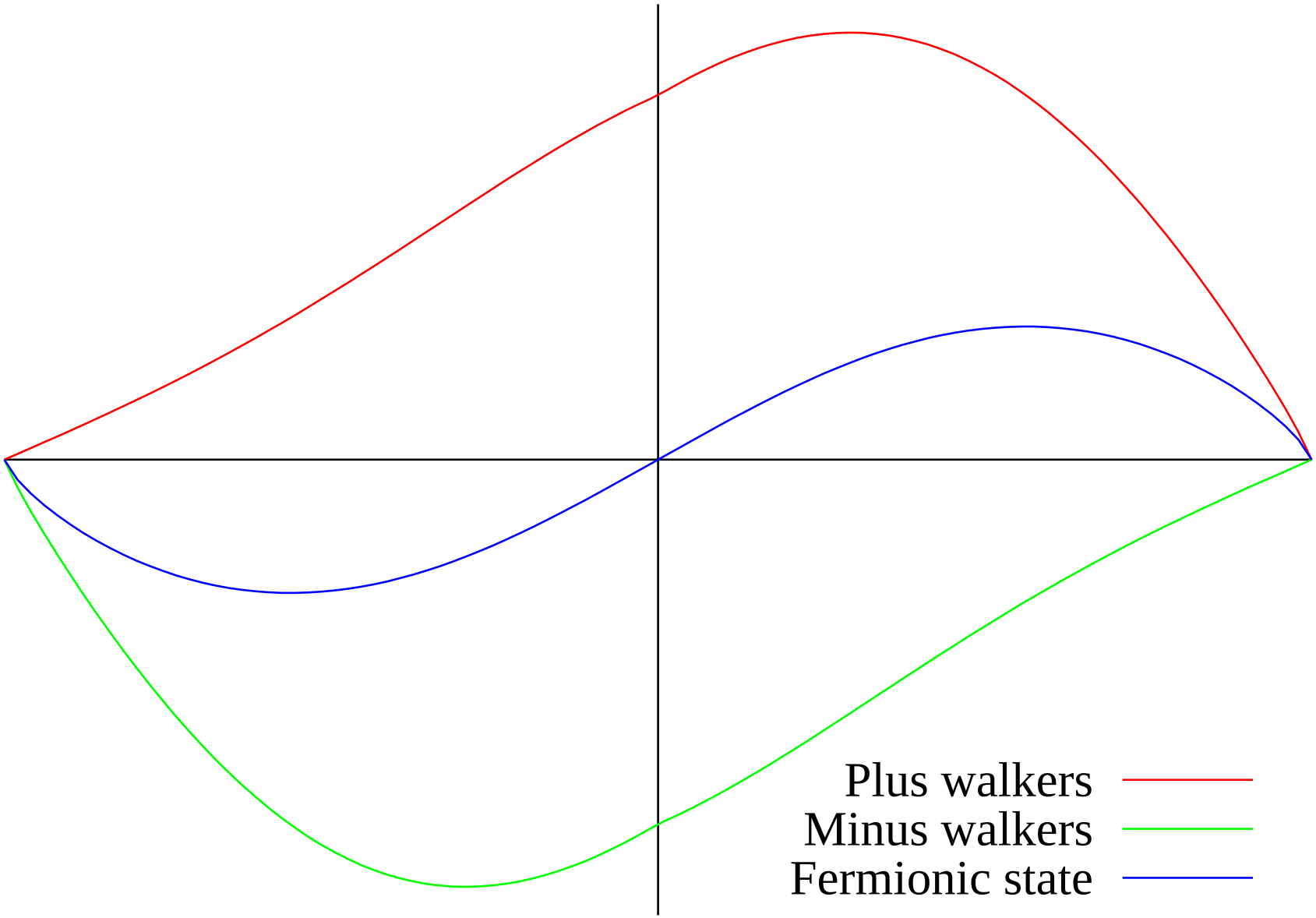
Sign Problem in 1^{st} Quantization and R space



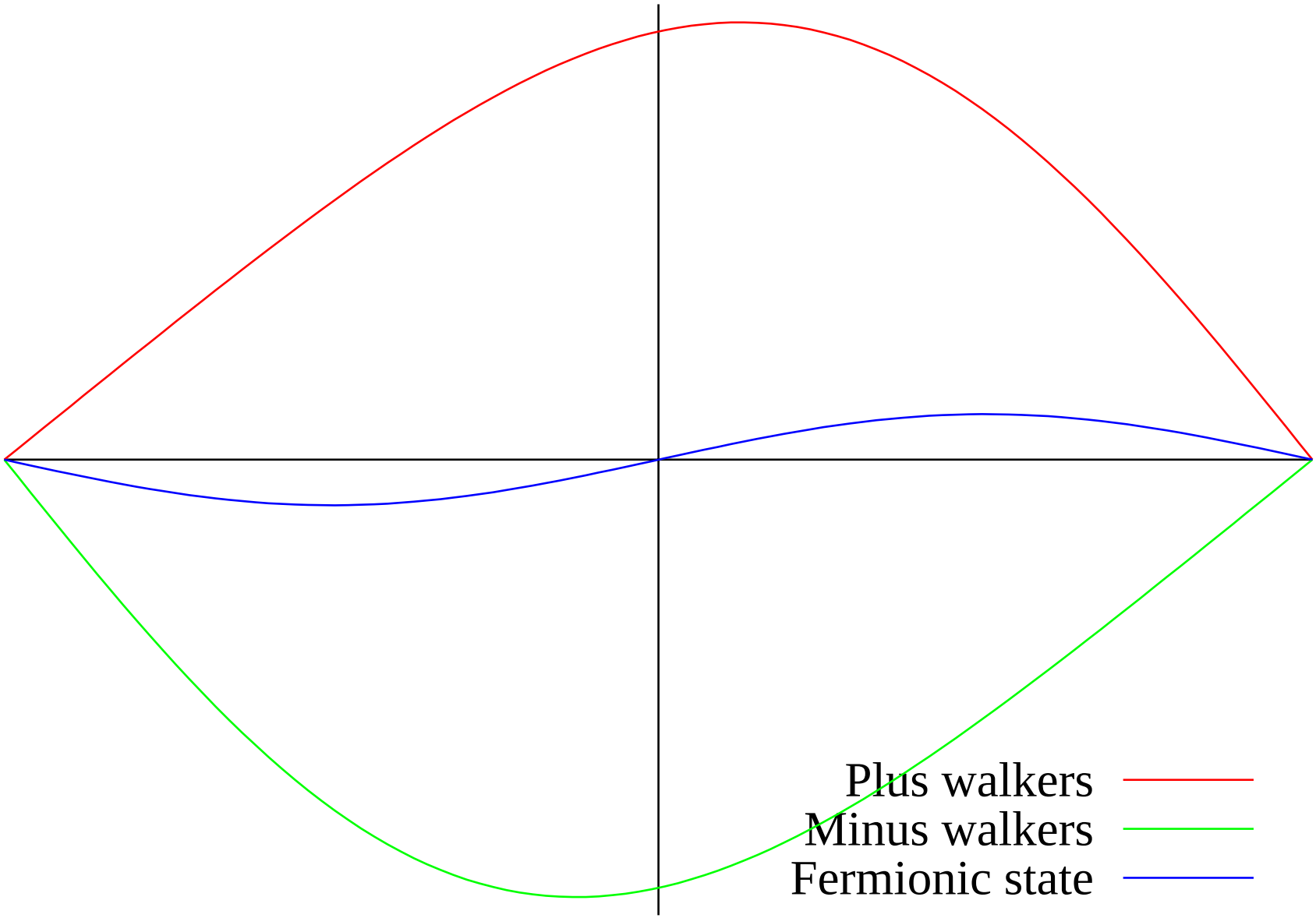
Sign Problem in 1^{st} Quantization and R space



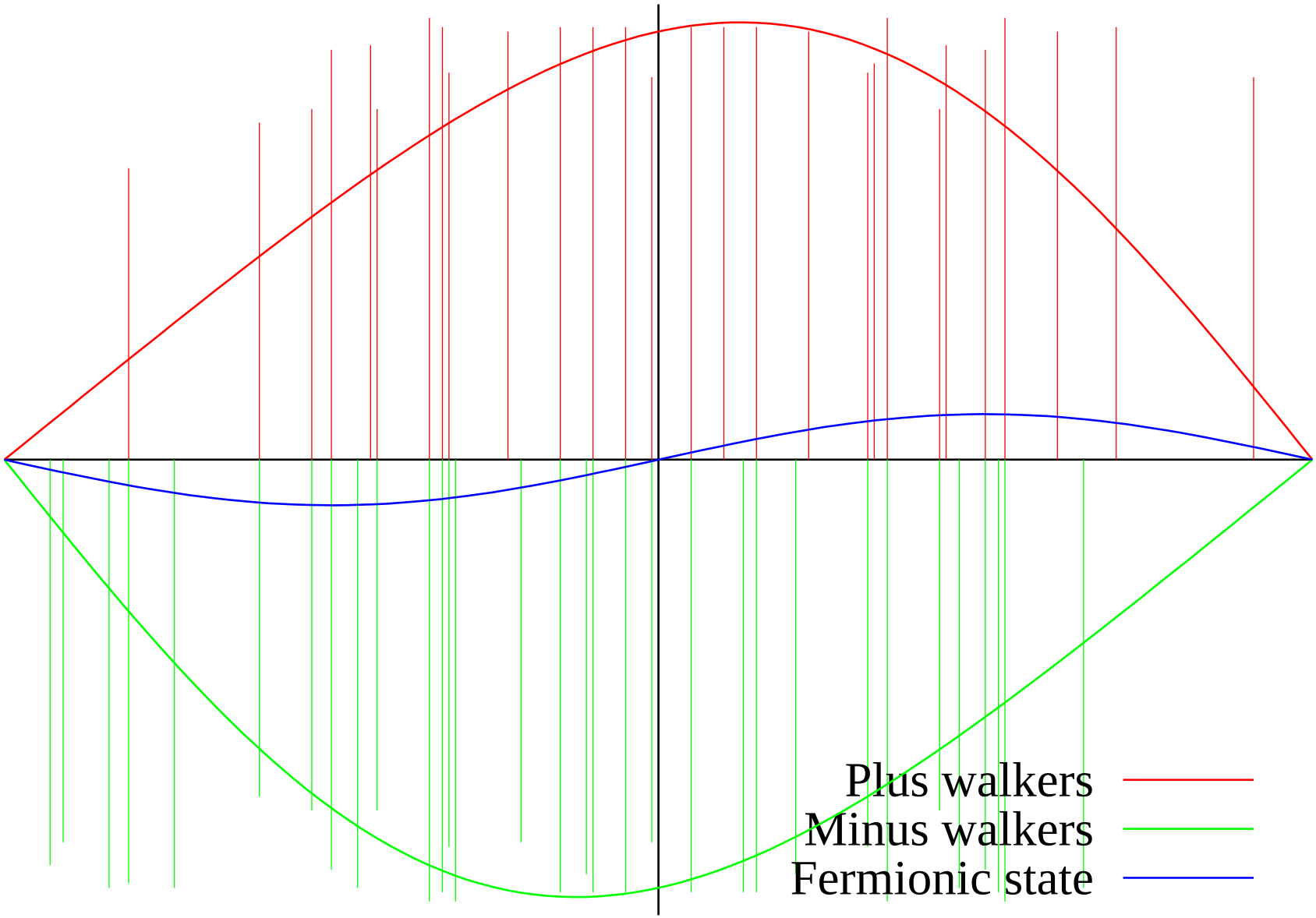
Sign Problem in 1^{st} Quantization and R space



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



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



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 real-space Coulomb systems, real- and momentum-space Hubbard model, but, is different for orbital-space Coulomb systems.

Sign Problem in FCIQMC/SQMC

What these papers did not say, but are important advantages of 2^{nd} quantization even when instability gap is the same:

1. 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.
2. Since the Hilbert space is $N!$ times smaller in 2^{nd} quantization, cancellation are much more effective.

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. Exact if $\Psi_{\rm T}$ nodes exact.	Walker cancellation plus initiator approximation needed for stable algorithm. 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

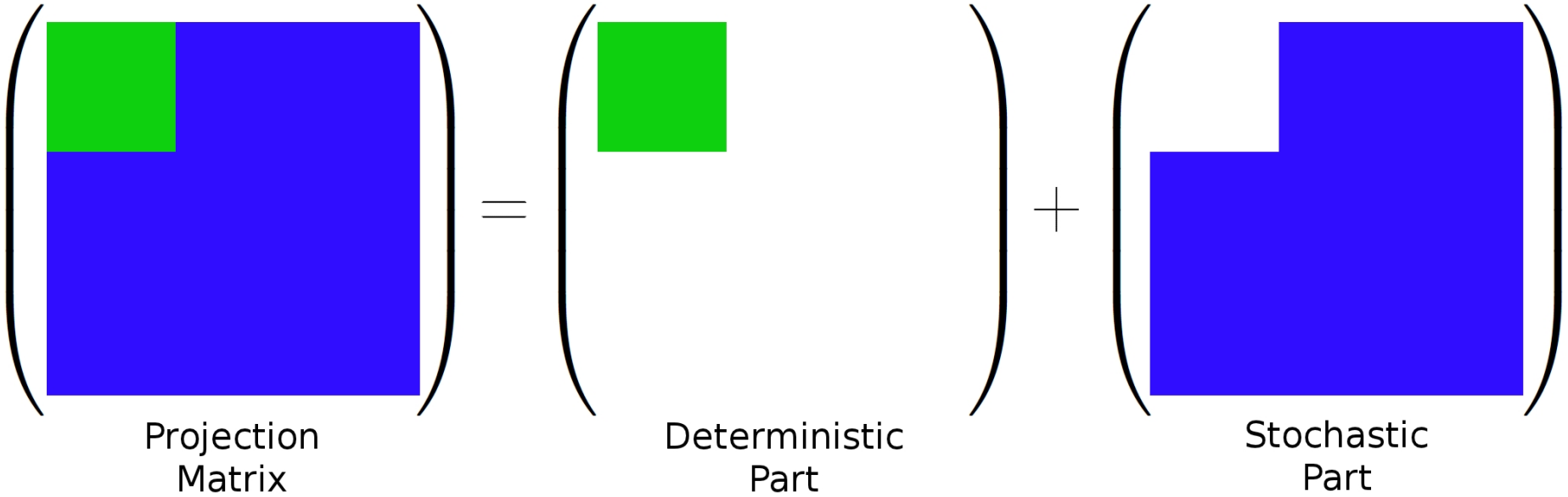
SQMC is hybrid of Exact Diagonalization and QMC

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.

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.

$$P = P^{\mathcal{D}} + P^{\mathcal{S}} \tag{1}$$

$$P^{\mathcal{D}}_{ij} = \begin{cases} P_{ij}, & i,j \in \mathcal{D} \\ 0, & \text{otherwise} \end{cases} \tag{2}$$

$$P^{\mathcal{S}} = P - P^{\mathcal{D}} \tag{3}$$

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

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

$$P_{jj} w_j(t) \quad = \quad [1 + \tau(E_T - H_{jj})]\, w_j(t) \tag{4}$$

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

$$\begin{cases} 0, & i,j \in \mathcal{D} \\ \dfrac{P_{ji}}{T_{ji}} \dfrac{w_i(t)}{n_i(t)} = -\tau \dfrac{H_{ji}}{T_{ji}} \dfrac{w_i(t)}{n_i(t)} & \text{otherwise} \end{cases} \tag{5}$$

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

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

$$\sum_{i \in \mathcal{D}} P^{\mathcal{D}}_{ji} w_i(t). \tag{6}$$

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

Semistochastic Projection

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

Project Do deterministic and stochastic projection

Sort Walker labels are sorted.

Merge Walkers on the same determinant are merged

Initiator The initiator criterion is used to discard some walkers.

Join Because we use real weights, there are many walkers with small weights. Join 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

Precompute:

Before MC part of the calculation do following:

1. Choose the deterministic space $\mathcal{D}$ and precompute matrix elements of projector, P, between all pairs of deterministic determinants.
2. Choose the trial wave function, Ψ_{T}, and precompute the local energy components of all determinants connected to those in Ψ_{T}.

Some differences between SQMC and FCIQMC or PMC:

1. Deterministic projection in part of space
2. Real (rather than integer) weights, $|\psi(t)\rangle = \sum_{i=1}^{N} w_i(t)|\phi_i\rangle$
3. Graduated initiator, threshold $= i\, d^p$, (Usually choose, $i, d = 1$)
4. Multideterminantal Ψ_{T}, particularly important for strongly correlated states

Test Cases

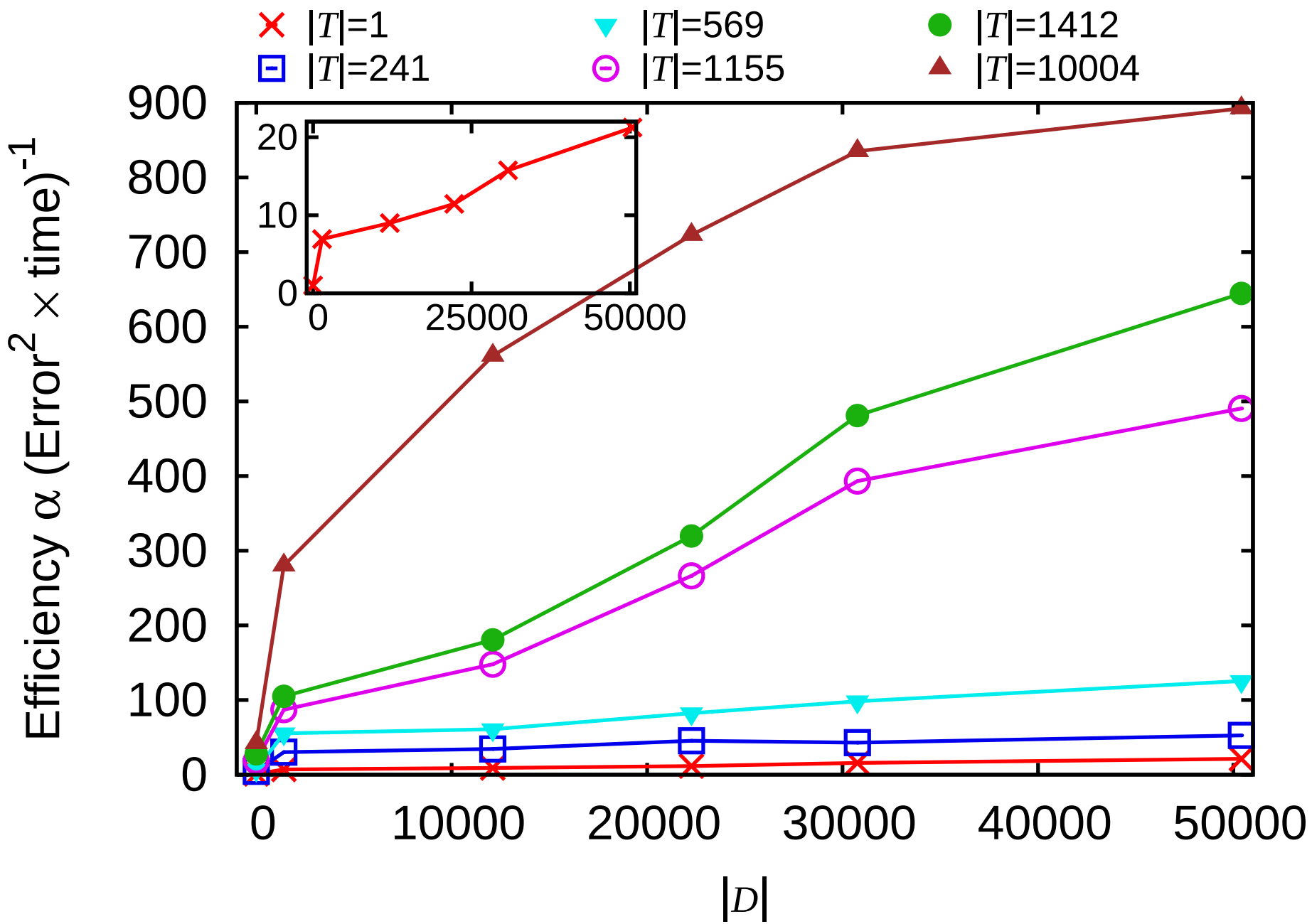
Test the ideas on:

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

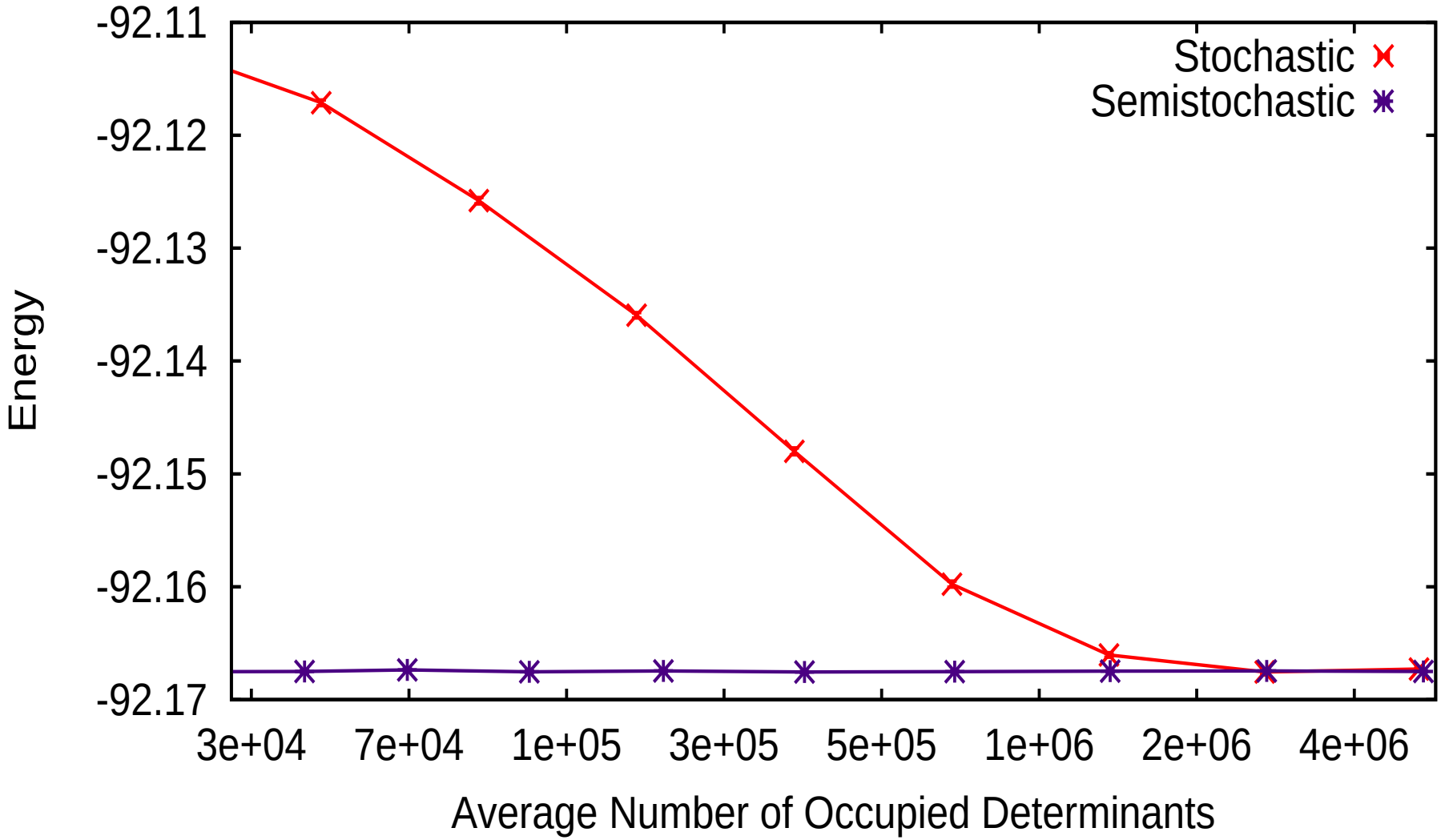
Why Hubbard?

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.

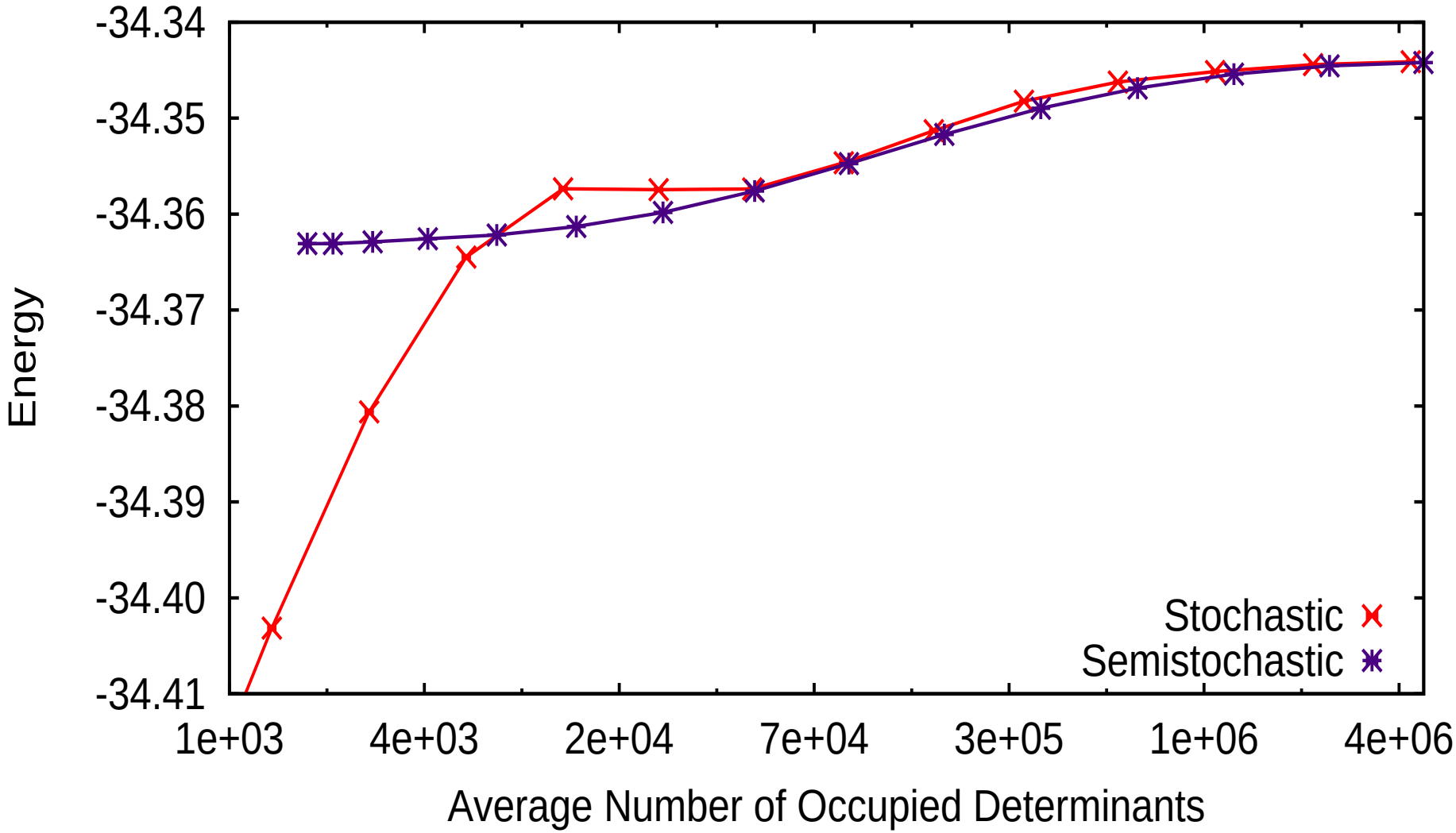
Efficiency Gains in 8×8 Hubbard Model, $N = 10$



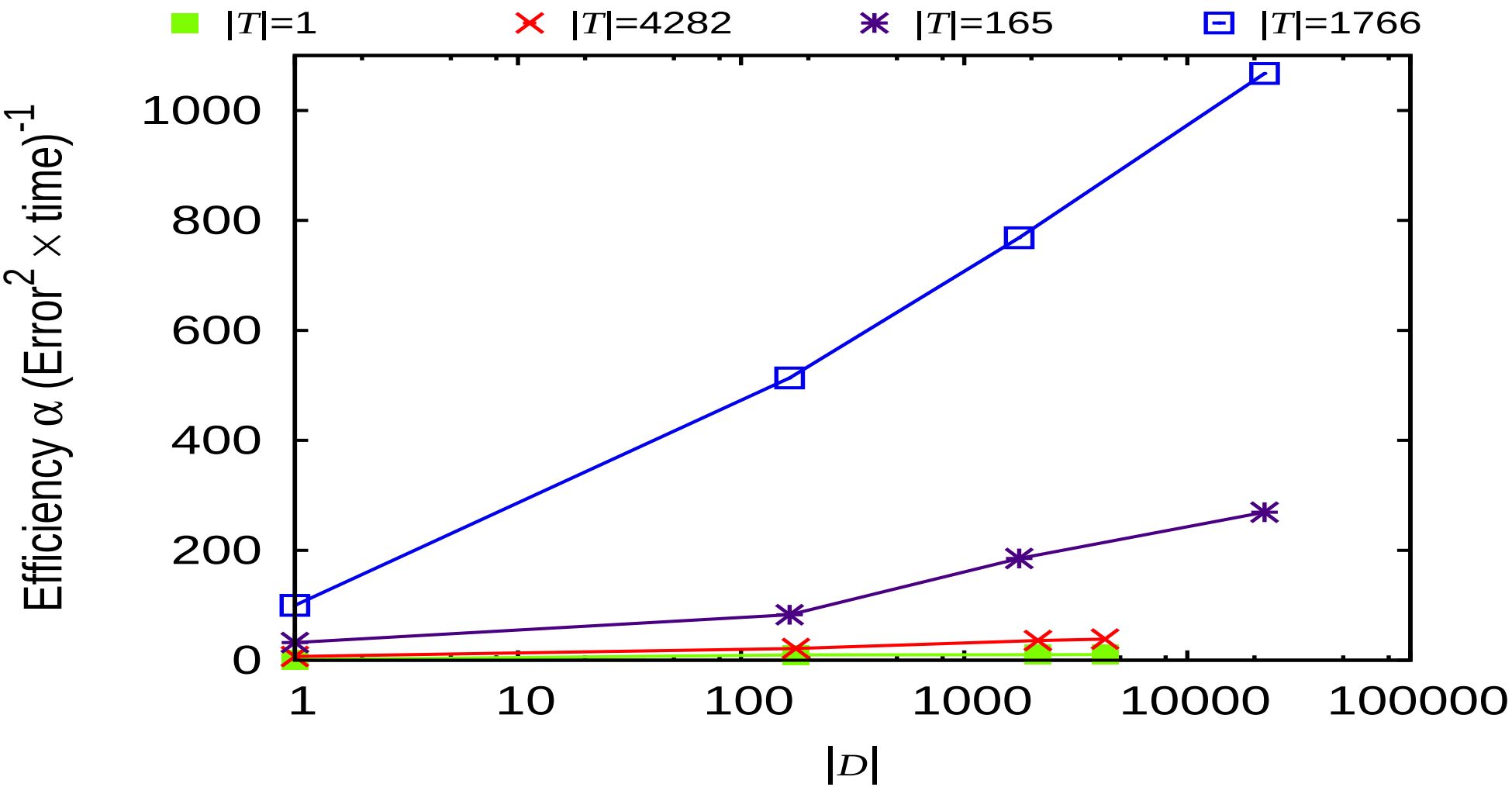
Energy versus average number of occupied determinants, 8×8 Hubbard, $N = 50$, $U = 1$



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 trial wave functions, $\Psi_{\rm T}$ and deterministic space
3. use F12 methods to improve basis convergence (with Takeshi Yanai, Garnet Chan, George Booth)
4. embedding (Garnet Chan, George Booth)
5. excited states (Ohtsuka and Nagase by projecting out lower states, Booth and Chan by modified projector to target desired state)