Overview of variational and projector Monte Carlo methods

Cyrus Umrigar

Physics Department, Cornell University, Ithaca.

Email: CyrusUmrigar@cornell.edu

GDR Correlation, Marseille, 8-10 July, 2015

Outline

1. Intro to Monte Carlo Methods
2. The essense of variational and projector MC methods
3. Varational MC (VMC)
    1. Metropolis-Hastings algorithm
    2. Form of trial wavefunctions
    3. Optimization of trial wavefunctions
4. Projector MC (PMC) methods
    1. Full configuration interaction QMC (FCIQMC) (Alavi group)
    2. Diffusion Monte Carlo (DMC)
5. Sign Problem in various Projector Monte Carlo methods (PMC)
6. Semistochastic Quantum Monte Carlo (SQMC) (extension of FCIQMC)
7. Pros and cons of various Projector Monte Carlo methods
8. Applications

Electronic Structure Methods

Quantum Chemistry Methods $\sim 10^2 - 10^3$ developers and users.

Hartree-Fock (HF), Many-Body Perturbation Theory (MBPT), Configuration Interaction (CI), Coupled Cluster (CC)

Systematically improveable (require double limit of ∞ single-particle basis (F12 helps!) and all excitation levels) but computational cost is prohibitive, e.g., Full Configuration Interaction (FCI) is rarely done since cost is $\mathcal{O}(e^N)$, CCSD(T) is popular and scales as $\mathcal{O}(N^7)$.

Density Functional Theory $\sim 10^2 - 10^3$ developers and users.

Exact in principle, but in practice one uses approximate exchange-correlation functionals and they are not systematically improvable.

Quantum Monte Carlo $\sim 10 - 10^2$ developers and users.

Can be used both at zero and at finite temperatures, and, for finite and for periodic systems. Some QMC methods work directly with infinite basis. Low-order polynomial cost if fixed-node or similar approximation is used. FN approximation is often accurate if good well-optimized trial wavefunctions are used. Requires ingenuity to keep statistical errors small. Some observables easier to compute than others.

Monte Carlo Methods

Monte Carlo methods: A class of computational algorithms that rely on repeated random sampling to compute results.
A few broad areas of applications are:

1. physics
2. chemistry
3. engineering
4. social sciences
5. finance and risk analysis

When are MC methods likely to be the methods of choice?

1. When the state space is discrete and very large, say $> 10^{10}$.
2. When the state space is continuous and high dimensional, say > 8.

Obvious drawback of MC methods: There is a statistical error.
Frequently there is a tradeoff between statistical error and systematic error and one needs to find the best compromise.

Physics/Chemistry applications of Quantum Monte Carlo

Some systems to which they have been applied are:

- strongly correlated systems (Hubbard, Anderson, t-J, ... models)
- quantum spin systems (Ising, Heisenberg, xy, ... models),
- liquid and solid helium, liquid-solid interface, droplets
- energy and response of homogeneous electron gas in 2-D and 3-D
- nuclear structure
- lattice gauge theory
- atomic clusters
- electronic structure calculations of atoms, molecules, solids, quantum dots, quantum wires

- both to zero temperature (pure states) and finite temperature problems, but in this lecture we will discuss only zero temperature methods

Early Recorded History of Monte Carlo

1777 Comte de Buffon: If a needle of length L is thrown at random onto a plane ruled with straight lines a distance $d (d > L)$ apart, then the probability P of the needle intersecting one of those lines is $P = \frac{2L}{\pi d}$.
Laplace: This could be used to compute π (inefficiently).

1930s First significant scientific application of MC: Enrico Fermi used it for neutron transport in fissile material.
Segre: "Fermi took great delight in astonishing his Roman colleagues with his "too-good-to-believe" predictions of experimental results."

1940s Monte Carlo named by Nicholas Metropolis and Stanislaw Ulam

1953 Algorithm for sampling any probability density
Metropolis, Rosenbluth, Rosenbluth, Teller and Teller
(generalized by Hastings in 1970)

1962,1974 First PMC calculations, Kalos, and, Kalos, Levesque, Verlet.

1965 First VMC calculations (of liquid He), Bill McMillan.

Central Limit Theorem

de Moivre (1733), Laplace (1812), Lyapunov (1901), Pólya (1920)

Let $X_1, X_2, X_3, \cdots, X_N$ be a sequence of N independent random variables sampled from a probability density function with a finite expectation value, μ, and variance σ^2. The central limit theorem states that as the sample size N increases, the probability density of the sample average of these random variables approaches the normal distribution, $\frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/(2\sigma^2/N)}$, with a mean μ, and variance σ^2/N, irrespective of the original probability density function.

Law of Large Numbers

Cardano, Bernouli, Borel, Cantelli, Kolmogorov, Khinchin

Even if the variance is infinite, if the expected value is finite, the sample means will converge to the expected value but usual error estimates go down slower than $1/\sqrt{N}$ and do not imply usual confidence intervals. Beware of skewed densities that have ∞ variance!

Monte Carlo versus Deterministic Integration methods

Deterministic Integration Methods:

Integration Error, ϵ, using $N_{\rm int}$ integration points:

1-dim Simpson rule: $\epsilon \propto N_{\rm int}^{-4}$, (provided derivatives up to 4^{th} exist)

d-dim Simpson rule: $\epsilon \propto N_{\rm int}^{-4/d}$, (provided derivatives up to 4^{th} exist)

So, for a given error, N and so the computer time increases exponentially with d, since $N \propto (\frac{1}{\epsilon})^{d/4}$.

Monte Carlo:

$\epsilon \sim \sigma(T_{\rm corr}/N_{\rm int})^{1/2}$, independent of dimension!, according to the central limit theorem since width of gaussian decreases as $(T_{\rm corr}/N_{\rm int})^{1/2}$ provided that the variance of the integrand is finite. ($T_{\rm corr}$ is the autocorrelation time.)

Roughly, Monte Carlo becomes advantageous for $d > 8$.

For a many-body wavefunction $d = 3N$ and can be a few thousand!

Scaling with number of electrons

Simpson's rule integration

$$\epsilon = \frac{c}{N_{\rm int}^{4/d}} = \frac{c}{N_{\rm int}^{4/3N_{\rm elec}}}$$

$$N_{\rm int} = \left(\frac{c}{\epsilon}\right)^{\frac{3N_{\rm elec}}{4}} \qquad \text{exponential in } N_{\rm elec}$$

Monte Carlo integration

$$\epsilon = \sigma\sqrt{\frac{N_{\rm elec}}{N_{\rm MC}}}$$

$$N_{\rm MC} = \left(\frac{\sigma}{\epsilon}\right)^2 N_{\rm elec} \qquad \text{linear in } N_{\rm elec}$$

(For both methods, computational cost is higher than this since the cost of evaluating the wavefunction increases with $N_{\rm elec}$, e.g., as $N_{\rm elec}^3$, (better if one uses "linear scaling"; worse if one increases $N_{\rm det}$ with $N_{\rm elec}$.))

Monte Carlo Integration

$$I = \int_V f(x)dx = V\,\overline{f} \pm V\sqrt{\frac{\overline{f^2} - \overline{f}^2}{N-1}}$$

where $\overline{f} = \frac{1}{N}\sum_i^N f(x_i), \quad \overline{f^2} = \frac{1}{N}\sum_i^N f^2(x_i)$

and the points x_i are sampled uniformly in V.

Importance sampling

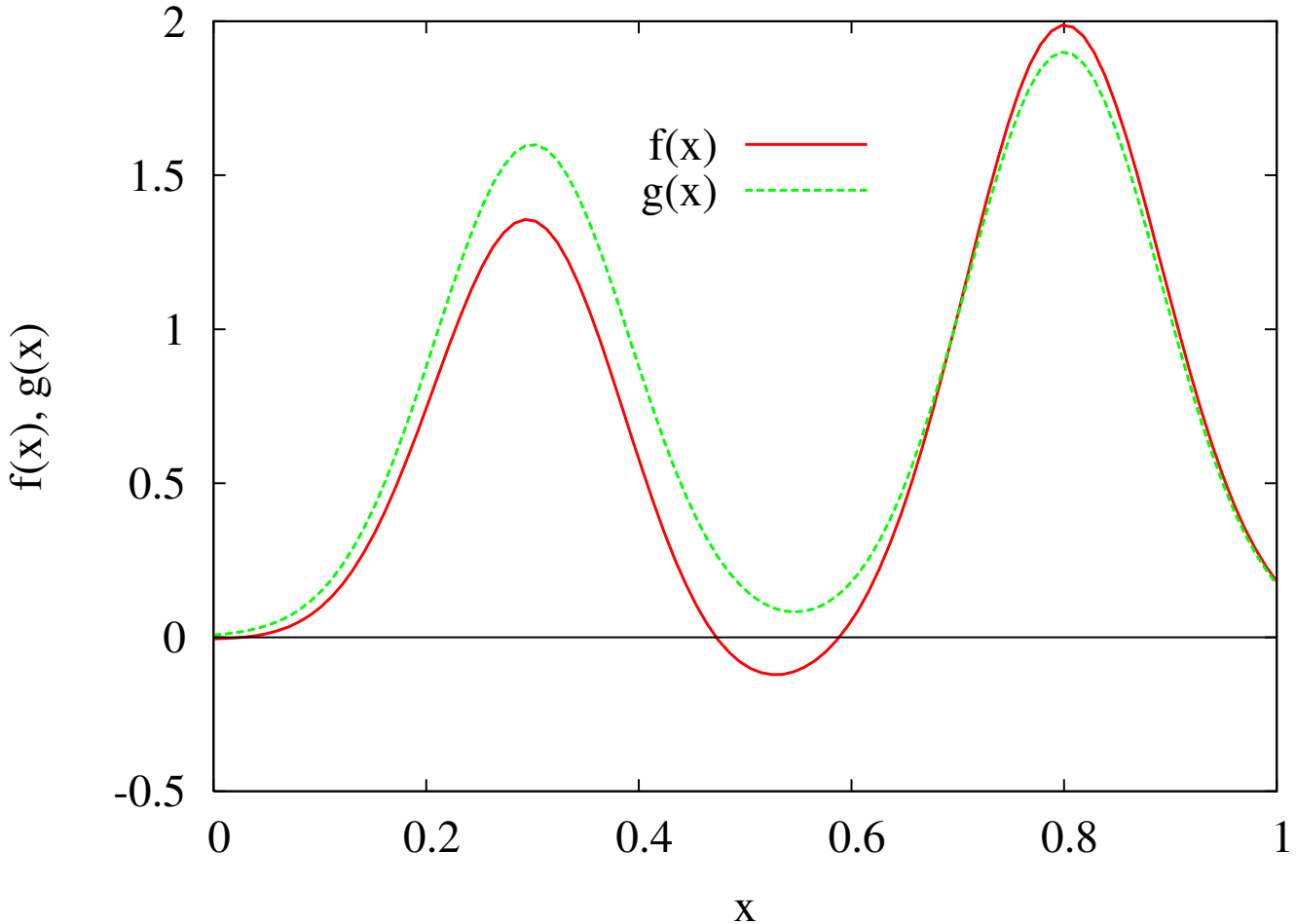
$$I = \int_V g(x)\frac{f(x)}{g(x)}dx = V\overline{\left(\frac{f}{g}\right)} \pm V\sqrt{\frac{\overline{\left(\frac{f}{g}\right)^2} - \overline{\left(\frac{f}{g}\right)}^2}{N-1}}$$

where the probability density function $g(x) \geq 0$ and $\int_V g(x)dx = 1$.
If $g(x) = 1/V$ in V then we recover original fluctuations but if $g(x)$ mimics $f(x)$ then the fluctuations are much reduced. Optimal g is $|f|$. Need: a) $g(x) \geq 0$, b) know integral of $g(x)$, and, c) be able to sample it.

Importance sampling can turn an $\infty-$variance estimator into a finite variance one!

Illustration of Importance Sampling

$f(x)$ is the function to be integrated. $g(x)$ is a function that is "similar" to $f(x)$ and has the required properties: a) $g(x) \geq 0$, b) we know integral of $g(x)$, and, c) we know how to sample it. $\int f(x)dx$ can be evaluated efficiently by sampling $g(x)$ and averaging $f(x)/g(x)$.



Typical probability densities in QMC are highly peaked, so importance sampling is essential.

Quantum Monte Carlo Methods

Cyrus J. Umrigar

What is Quantum Monte Carlo?

Stochastic implementation of the power method for projecting out the dominant eigenvector of a matrix or integral kernel.

"Dominant state" means state with largest absolute eigenvalue.

If we repeatedly multiply an arbitrary vector, not orthogonal to the dominant state, by the matrix, we will eventually project out the dominant state.

QMC methods are used only when the number of states is so large ($> 10^{10}$) that it is not practical to store even a single vector in memory. Otherwise use exact diagonalization method, e.g., Lanczos. So, at each MC generation, only a sample of the states is stored.

QMC methods are used not only in a large discrete space but also in a continuously infinite space. Hence "matrix or integral kernel" above. In the interest of brevity I will use either discrete or continuous language (sums and matrices or integrals and integral kernels), but much of what is said will apply to both situations.

Definitions

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

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

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

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

(If basis incomplete then "exact" means "exact in that basis".)

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.

$\Psi_{\rm G}$ 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 $\Psi_{\rm T}$. Reasons to have $\Psi_{\rm G} \neq \Psi_{\rm T}$ are: a) rapid evaluation of "local energy", b) have finite-variance estimators. To simplify expressions, we use $\Psi_{\rm G} = \Psi_{\rm T}$ or $\Psi_{\rm G} = 1$ in what follows.

Variational MC

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

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

$$= \sum_{i}^{N_{\rm st}} \frac{t_i^2}{\sum_{k}^{N_{\rm st}} t_k^2} E_{\rm L}(i) = \frac{\left[\sum_{i}^{N_{\rm MC}} E_{\rm L}(i)\right]_{\Psi_{\rm T}^2}}{N_{\rm MC}} \xrightarrow{\Psi_G \neq \Psi_T} \frac{\left[\sum_{i}^{N_{\rm MC}} \left(\frac{t_i}{g_i}\right)^2 E_{\rm L}(i)\right]_{\Psi_{\rm G}^2}}{\left[\sum_{k}^{N_{\rm MC}} \left(\frac{t_k}{g_k}\right)^2\right]_{\Psi_{\rm G}^2}}$$

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

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 $\Psi_{\rm T} \to \Psi_0$.

For fixed Ψ_T, $\Psi_G = \Psi_T$ does not minimize statistical fluctuations!

In fact need $\Psi_G \neq \Psi_T$ at times to get finite variance. $\Psi_G = \Psi_T$ does give unbiased estimator.

Projector MC

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

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

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

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

$$= \sum_{i}^{N_{\rm st}} \frac{e_i t_i}{\sum_{k}^{N_{\rm st}} e_k t_k} \, E_{\rm L}(i) = \frac{\left[\sum_{i}^{N_{\rm MC}} E_{\rm L}(i)\right]_{\Psi_{\rm T}\Psi_0}}{N_{\rm MC}} \xrightarrow{\Psi_G \neq \Psi_T} \frac{\left[\sum_{i}^{N_{\rm MC}} \left(\frac{t_i}{g_i}\right) E_{\rm L}(i)\right]_{\Psi_{\rm G}\Psi_0}}{\left[\sum_{k}^{N_{\rm MC}} \left(\frac{t_k}{g_k}\right)\right]_{\Psi_{\rm G}\Psi_0}}$$

Sample $e_i g_i / \sum_k^{N_{\rm st}} e_k g_k$ using projector.

For exact PMC, value indep. of $\Psi_{\rm T}$, $\Psi_{\rm G}$, statistical error depends on $\Psi_{\rm T}$, $\Psi_{\rm G}$.
(For FN-PMC, value depends on $\Psi_{\rm G}$, statistical error on $\Psi_{\rm T}, \Psi_{\rm G}$.)
(For FN-DMC, value depends on nodes of $\Psi_{\rm G}$, statistical error on $\Psi_{\rm T}, \Psi_{\rm G}$.)
Statistical error vanishes as $\Psi_{\rm T} \to \Psi_0$.
For fixed Ψ_T, $\Psi_G = \Psi_T$ does not minimize statistical fluctuations!

Variational and Projector MC

$$E_V = \frac{\left[\sum_i^{N_{\rm MC}} \left(\frac{t_i}{g_i}\right)^2 E_{\rm L}(i)\right]_{\Psi_{\rm G}^2}}{\left[\sum_k^{N_{\rm MC}} \left(\frac{t_k}{g_k}\right)^2\right]_{\Psi_{\rm G}^2}} \quad \text{(Value depends on } \Psi_{\rm T}\text{, error } \Psi_{\rm T}, \Psi_{\rm G}\text{)}$$

$$E_0 = \frac{\left[\sum_i^{N_{\rm MC}} \left(\frac{t_i}{g_i}\right) E_{\rm L}(i)\right]_{\Psi_{\rm G}\Psi_0}}{\left[\sum_k^{N_{\rm MC}} \left(\frac{t_k}{g_k}\right)\right]_{\Psi_{\rm G}\Psi_0}} \quad \text{(Value exact}^\dagger\text{. Error depends on } \Psi_{\rm T}, \Psi_{\rm G}\text{.)}$$

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

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

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 (if space discrete) or semi-diagonal (if space continuous).

† In practice, usually necessary to make approximation (e.g. FN) and value depends on $\Psi_{\rm G}$.

Projector MC

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

Projector is any function of the Hamiltonian that maps the ground state eigenvalue of $\hat{H}$ to 1, and the highest eigenvalue of $\hat{H}$ to an absolute value that is < 1 (preferably close to 0).

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

Linear projector: $\hat{P} = \hat{\mathbf{1}} + \tau(E_T\hat{\mathbf{1}} - \hat{H})$

If spectrum is bounded and $\tau \leq \dfrac{1}{E_{\rm max} - E_{\rm min}}$.

Taxonomy of 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)	$e^{\tau(E_T\hat{\mathbf{1}}-\hat{H})}$ (samp. τ)	$\mathbf{r}$	1^{st}
LRDMC (Sorella, Casula)	$e^{\tau(E_T\hat{\mathbf{1}}-\hat{H})}$ (samp. τ)	$\mathbf{r}_i$	1^{st}
FCIQMC/SQMC	$\hat{\mathbf{1}}+\tau(E_T\hat{\mathbf{1}}-\hat{H})$	ϕ_i^{orthog}	2^{nd}
phaseless AFQMC (Zhang, Krakauer)	$e^{\tau(E_T\hat{\mathbf{1}}-\hat{H})}$	$\phi_i^{\mathrm{nonorthog}}$	2^{nd}

$1+\tau(E_T\hat{\mathbf{1}}-\hat{H})$ and $\frac{1}{\hat{\mathbf{1}}-\tau(E_T\hat{\mathbf{1}}-\hat{H})}$ can be used only if the spectrum of $\hat{H}$ is bounded.

Variational Monte Carlo

Cyrus J. Umrigar

Three ingredients for accurate Variational Monte Carlo

1. A method for sampling an arbitrary wave function Metropolis-Hastings.
2. A functional form for the wave function that is capable of describing the correct physics/chemistry.
3. An efficient method for optimizing the parameters in the wave functions.

Metropolis-Hastings Monte Carlo

Metropolis, Rosenbluth[2], Teller[2], JCP, **21** 1087 (1953)

W.K. Hastings, Biometrika, **57** (1970)

Metropolis method originally used to sample the Boltzmann distribution. This is still one of its more common uses.

General method for sampling **any known** discrete or continuous density. (Other quantum Monte Carlo methods, e.g., diffusion MC, enable one to sample densities that are not explicitly known but are the eigenstates of known matrices or integral kernels.)

Metropolis-Hastings has serial correlations. Hence, direct sampling methods preferable, but rarely possible for complicated densities in many dimensions.

Metropolis-Hastings Monte Carlo (cont)

Construction of M

Impose *detailed balance* condition

$$M(\mathbf{R}_f|\mathbf{R}_i)\ \rho(\mathbf{R}_i) = M(\mathbf{R}_i|\mathbf{R}_f)\ \rho(\mathbf{R}_f)$$

Detailed balance is not necessary but provides way to construct M.
Write elements of M as product of elements of a proposal matrix T and an acceptance Matrix A,

$$M(\mathbf{R}_f|\mathbf{R}_i) = A(\mathbf{R}_f|\mathbf{R}_i)\ T(\mathbf{R}_f|\mathbf{R}_i)$$

$M(\mathbf{R}_f|\mathbf{R}_i)$ and $T(\mathbf{R}_f|\mathbf{R}_i)$ are stochastic matrices, but $A(\mathbf{R}_f|\mathbf{R}_i)$ is not.
Detailed balance is now:

$$A(\mathbf{R}_f|\mathbf{R}_i)\ T(\mathbf{R}_f|\mathbf{R}_i)\ \rho(\mathbf{R}_i) = A(\mathbf{R}_i|\mathbf{R}_f)\ T(\mathbf{R}_i|\mathbf{R}_f)\ \rho(\mathbf{R}_f)$$

$$\text{or} \qquad \frac{A(\mathbf{R}_f|\mathbf{R}_i)}{A(\mathbf{R}_i|\mathbf{R}_f)} = \frac{T(\mathbf{R}_i|\mathbf{R}_f)\ \rho(\mathbf{R}_f)}{T(\mathbf{R}_f|\mathbf{R}_i)\ \rho(\mathbf{R}_i)}\ .$$

Metropolis-Hastings Monte Carlo (cont)

Choice of Acceptance Matrix A

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

Infinity of choices for A. Any function

$$F\left(\frac{T(\mathbf{R}_\mathrm{i}|\mathbf{R}_\mathrm{f})\ \rho(\mathbf{R}_\mathrm{f})}{T(\mathbf{R}_\mathrm{f}|\mathbf{R}_\mathrm{i})\ \rho(\mathbf{R}_\mathrm{i})}\right) = A(\mathbf{R}_\mathrm{f}|\mathbf{R}_\mathrm{i})$$

for which $F(x)/F(1/x) = x$ and $0 \leq F(x) \leq 1$ will do.
Choice of Metropolis *et al.* $F(x) = \min\{1, x\}$, maximizes the acceptance:

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

Other less good choices for $A(\mathbf{R}_\mathrm{f}|\mathbf{R}_\mathrm{i})$ have been made, *e.g.* $F(x) = \frac{x}{1+x}$

$$A(\mathbf{R}_\mathrm{f}|\mathbf{R}_\mathrm{i}) = \frac{T(\mathbf{R}_\mathrm{i}|\mathbf{R}_\mathrm{f})\ \rho(\mathbf{R}_\mathrm{f})}{T(\mathbf{R}_\mathrm{i}|\mathbf{R}_\mathrm{f})\ \rho(\mathbf{R}_\mathrm{f}) + T(\mathbf{R}_\mathrm{f}|\mathbf{R}_\mathrm{i})\ \rho(\mathbf{R}_\mathrm{i})}.$$

Metropolis: $T(\mathbf{R}_\mathrm{i}|\mathbf{R}_\mathrm{f}) = T(\mathbf{R}_\mathrm{f}|\mathbf{R}_\mathrm{i})$, Hastings: $T(\mathbf{R}_\mathrm{i}|\mathbf{R}_\mathrm{f}) \neq T(\mathbf{R}_\mathrm{f}|\mathbf{R}_\mathrm{i})$

Metropolis-Hastings Monte Carlo (cont)

Choice of Proposal Matrix T

So, the optimal choice for the acceptance matrix $A(\mathbf{R}_\mathrm{f}|\mathbf{R}_\mathrm{i})$ is simple and known.

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

However, there is considerable scope for using one's ingenuity to come up with good proposal matrices, $T(\mathbf{R}_\mathrm{f}|\mathbf{R}_\mathrm{i})$, that allow one to make large moves with large acceptances, in order to make the autocorrelation time small.

In fact for electronic structure calculations it is possible to come up with $T(\mathbf{R}_\mathrm{f}|\mathbf{R}_\mathrm{i})$ such that the autocorrelation time T_corr is close to 1.

Functional form of Trial Wave Function

One of the great advantages of QMC is that one has a great deal of freedom in the functional form of the trial wavefunction. This is in contrast to other methods where one is largely restricted to using linear combinations of determinants, and, furthermore the orbitals in these determinants have to be expanded in basis functions, such as gaussians or planewaves, that are amenable to analytic integration.

In QMC one has can utilize one's intuition about the physics or chemistry of the problem to come up with good functional forms for the wavefunction. These functional forms may have several parameters, whose values are not know a priori, so powerful methods for optimizing these parameters have been developed.

Some innovative functional forms that have been used are:

1. Antisymmetrized geminal power times Jastrow Sorella, Casula
2. Pfaffian times Jastrow Schmidt, Mitas and coworkers
3. Inhomogeneous backflow times Jastrow Needs and coworkers

Most common form – multideterminant expansion times Jastrow.

Functional form of Trial Wave Function

$$\Psi_T = \left(\sum_n d_n \mathrm{D}_n^{\uparrow} \mathrm{D}_n^{\downarrow} \right) \times \mathcal{J}(r_i, r_j, r_{ij})$$

- Determinants: $\sum_n d_n \mathrm{D}_n^{\uparrow} \mathrm{D}_n^{\downarrow}$

$\mathrm{D}^{\uparrow}$ and $\mathrm{D}^{\downarrow}$ are determinants of single-particle orbitals ϕ for up ($\uparrow$) and down ($\downarrow$) spin electrons respectively.

The single-particle orbitals ϕ are given by:

$$\phi(\mathbf{r}_i) = \sum_{\alpha k} c_{k_\alpha} \; N_{k_\alpha} \, r_{i\alpha}^{n_{k_\alpha}-1} \, e^{-\zeta_{k_\alpha} r_{i\alpha}} \, \mathrm{Y}_{l_{k_\alpha} m_{k_\alpha}}(\hat{\mathrm{r}}_{i\alpha})$$

- Jastrow: $\mathcal{J}(r_i, r_j, r_{ij}) = \prod_{\alpha i} \exp(A_{\alpha i}) \prod_{ij} \exp(B_{ij}) \prod_{\alpha ij} \exp(C_{\alpha ij})$

  $A_{\alpha i} \Rightarrow$ electron-ion correlation

  $B_{ij} \Rightarrow$ electron-electron correlation

  $C_{\alpha ij} \Rightarrow$ electron-electron-ion correlation

d_n, c_{k_α}, ζ_{k_α} and parms in $\mathcal{J}$ are optimized.

$\sim N_{\mathrm{atomtype}}$ of $\mathcal{J}$ parms.

$\sim N_{\mathrm{atomtype}}$ of ζ_{k_α} parms.

$\sim N_{\mathrm{atom}}^2$ of c_{k_α} parms.

$\sim e^{N_{\mathrm{atom}}}$ of d_n parms.

Power of QMC:

$\mathcal{J}$ parms. do work of d_n parms.

Optimization of Wavefunctions

Cyrus J. Umrigar

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 for 1-det wavefn is 1.6 eV for total energy and 0.8 eV for well-depth. However, optimized multidet. wavefn has FN error that is better than chemical accuracy (1 kcal/mole $=$ 0.043 eV/molecule).
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, $\langle \Psi_0 | \hat{A} | \Psi_T \rangle$ not exact even for nodeless ψ_0, $\psi_{\rm T}$) if one does not use forward/side walking.

Choices to be made when optimizing trial wavefunctions

1. What precisely do we want to optimize – the objective function or measure of goodness?
2. What method do we use to do the optimization? If more than one method is applied to the same objective function, they will of course give the same wavefunction, but the efficiency with which we arrive at the solution may be much different.

Measures of goodness of variational wave functions

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

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

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

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

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.

Progress in optimization of Many-Body Wavefunctions

Naive energy optim. $\rightarrow$ Variance optim. $\rightarrow$ Efficient energy optim.

– 1988 naive energy optimization, few (~ 3) parameters

1988 – 2001 variance optimization, ~ 100 parameters

could be used for more, but, variance does not couple strongly to some parameters

R. Coldwell, IJQC (1977)

CJU, Wilson, Wilkins, Phys. Rev. Lett. (1988)

2001 – efficient energy optimization, ~ 1000's of parameters as many as 500,000

M. P. Nightingale and Alaverdian, Phys. Rev. Lett. (2001)

CJU, C. Filippi, Phys. Rev. Lett. (2005)

J. Toulouse, CJU, J. Chem. Phys. (2007)

CJU, J. Toulouse, C. Filippi, S. Sorella, Phys. Rev. Lett. (2007)

S. Sorella, M. Casula, D. Rocca, J. Chem. Phys. (2007)

J. Toulouse, CJU, J. Chem. Phys. (2008)

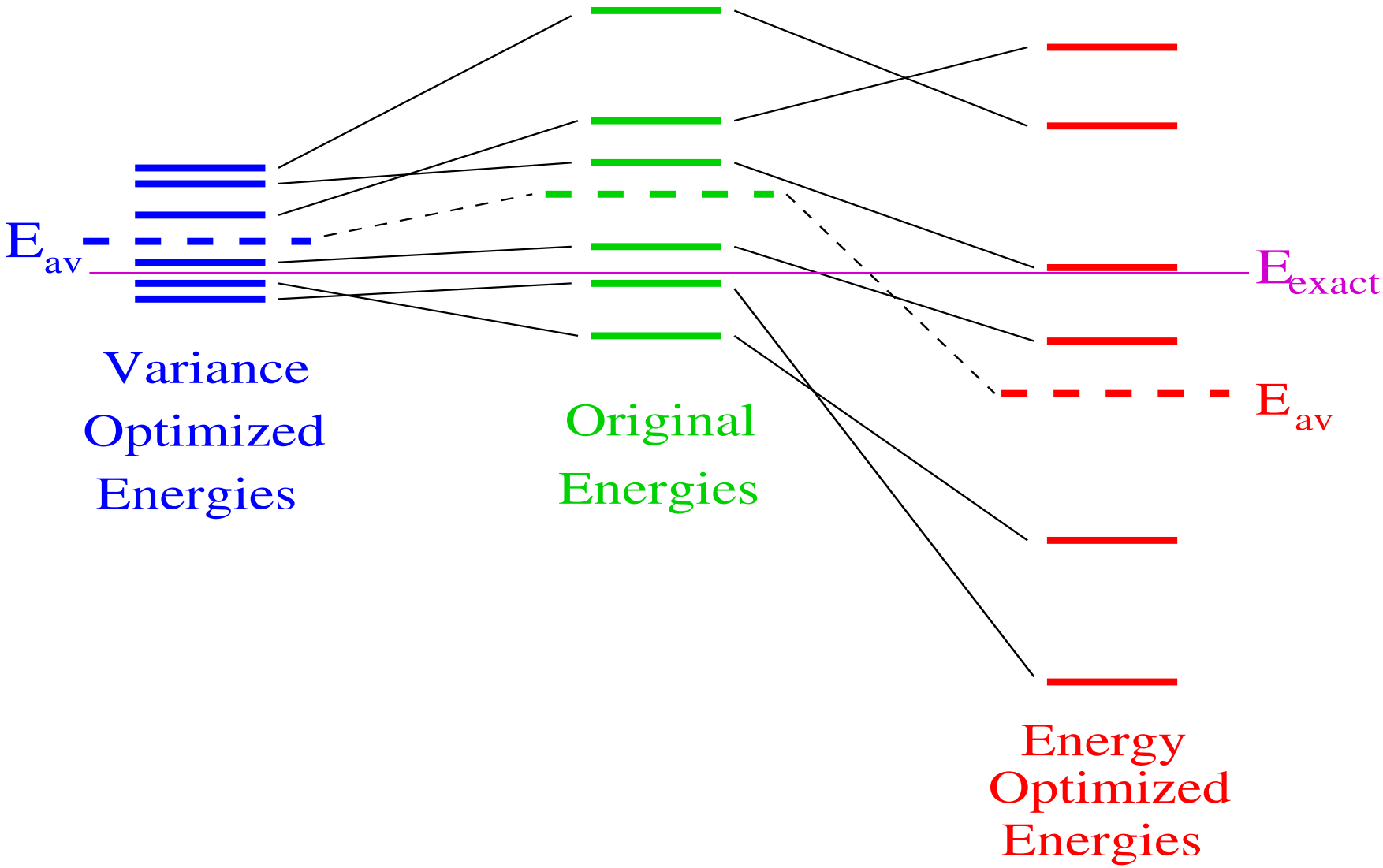
J. Toulouse, CJU, J. Chem. Phys. (2008)

E. Neuscamma, CJU, G. Chan, J. Chem. Phys. (2012)

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.

Advantages of Energy (or Mixed) Optim. vs. Variance Optim.

1. Want lowest energy; fluctuations are of secondary importance. Energy and variance do not always go hand-in-hand enough.
2. Some parameters couple more strongly to energy than variance.
3. Some variance-optimized parameters make wave function too extended.

Optimization Methods

The optimization methods we use are based on standard methods:

1. Levenberg-Marquardt method
2. Newton method
3. Linear method (though with significant extension to nonlinear parameters)
4. Perturbation theory

However, all of them need additional ingredients to work with stochastic methods, and these ingredients improve the efficiency of the method by several orders of magnitude!

Ingredients of efficient optimization methods

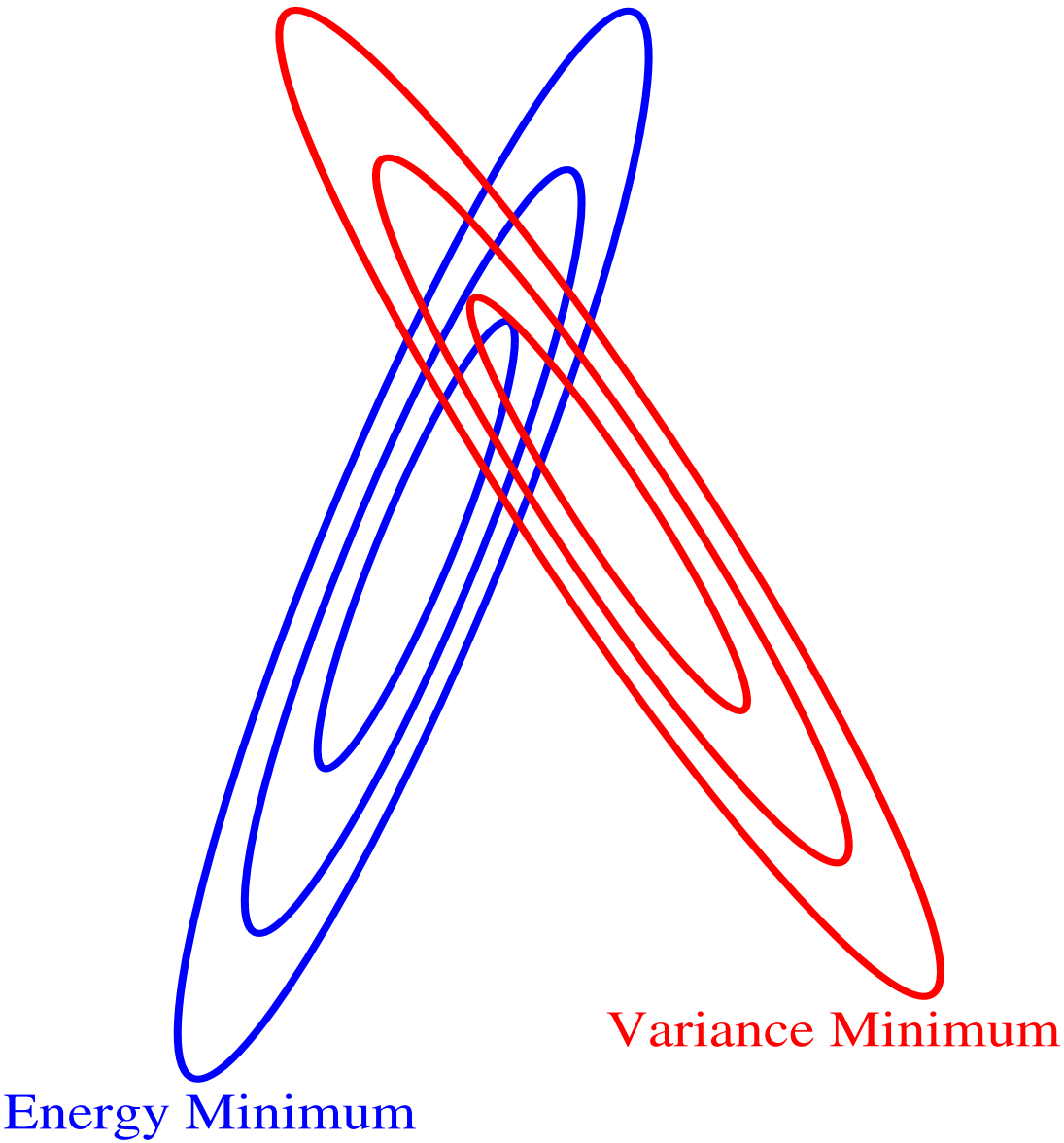
1. Newton: Add terms to the Hessian that have zero expectation value for an infinite sample, but, greatly reduce the noise for a finite sample.
   CJU and C. Filippi, PRL 2005

2. Linear: 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. Linear: Extension of the linear method to nonlinear parameters by using semiorthogonalized parameter derivatives. CJU, J. Toulouse, C. Filippi and S. Sorella, PRL 2007; J. Toulouse and CJU JCP 2007, 2008

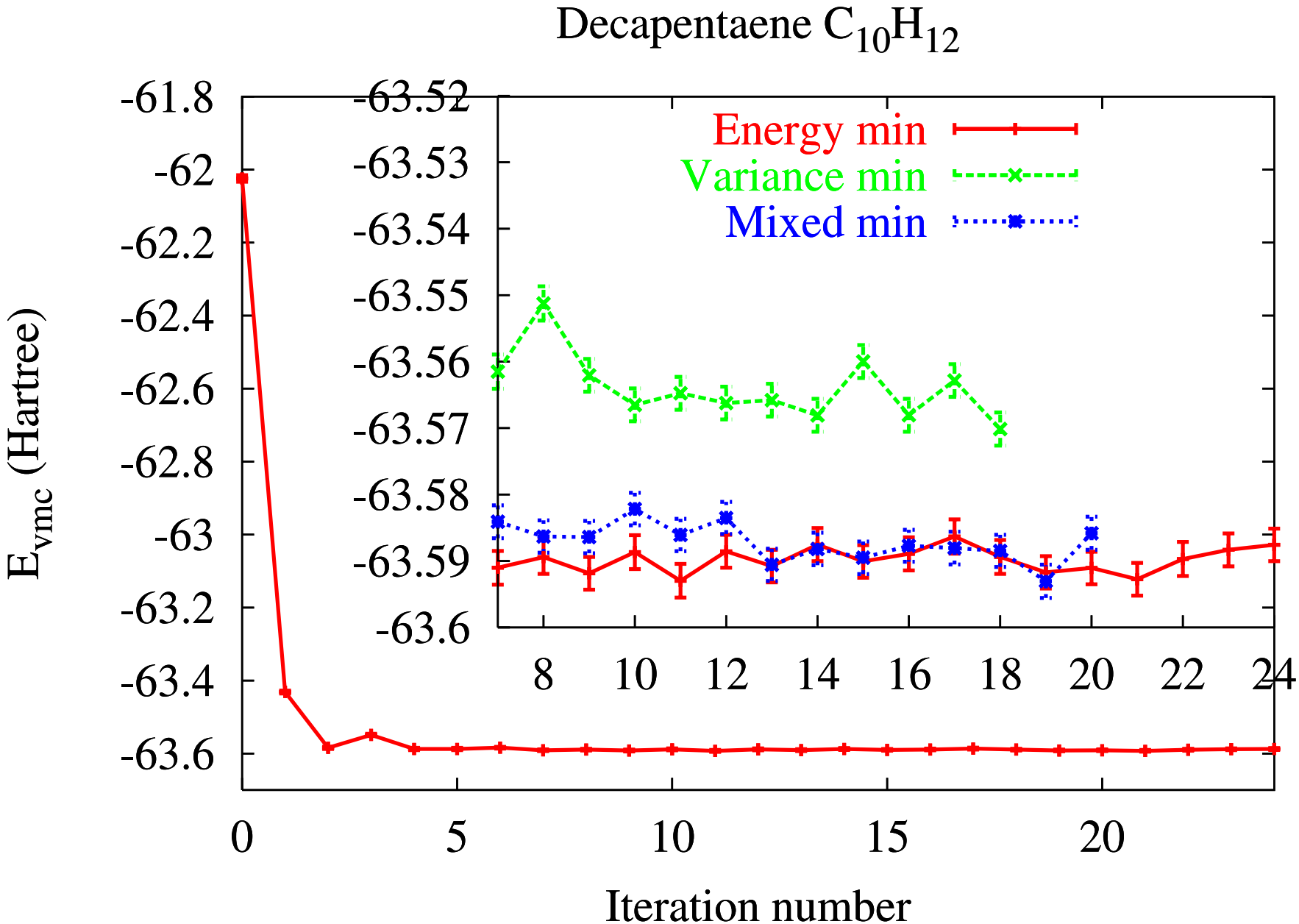
4. Newton and Linear: Automatic procedure for choosing size of moves and recovering from bad moves.

Optimization of linear combination of energy and variance



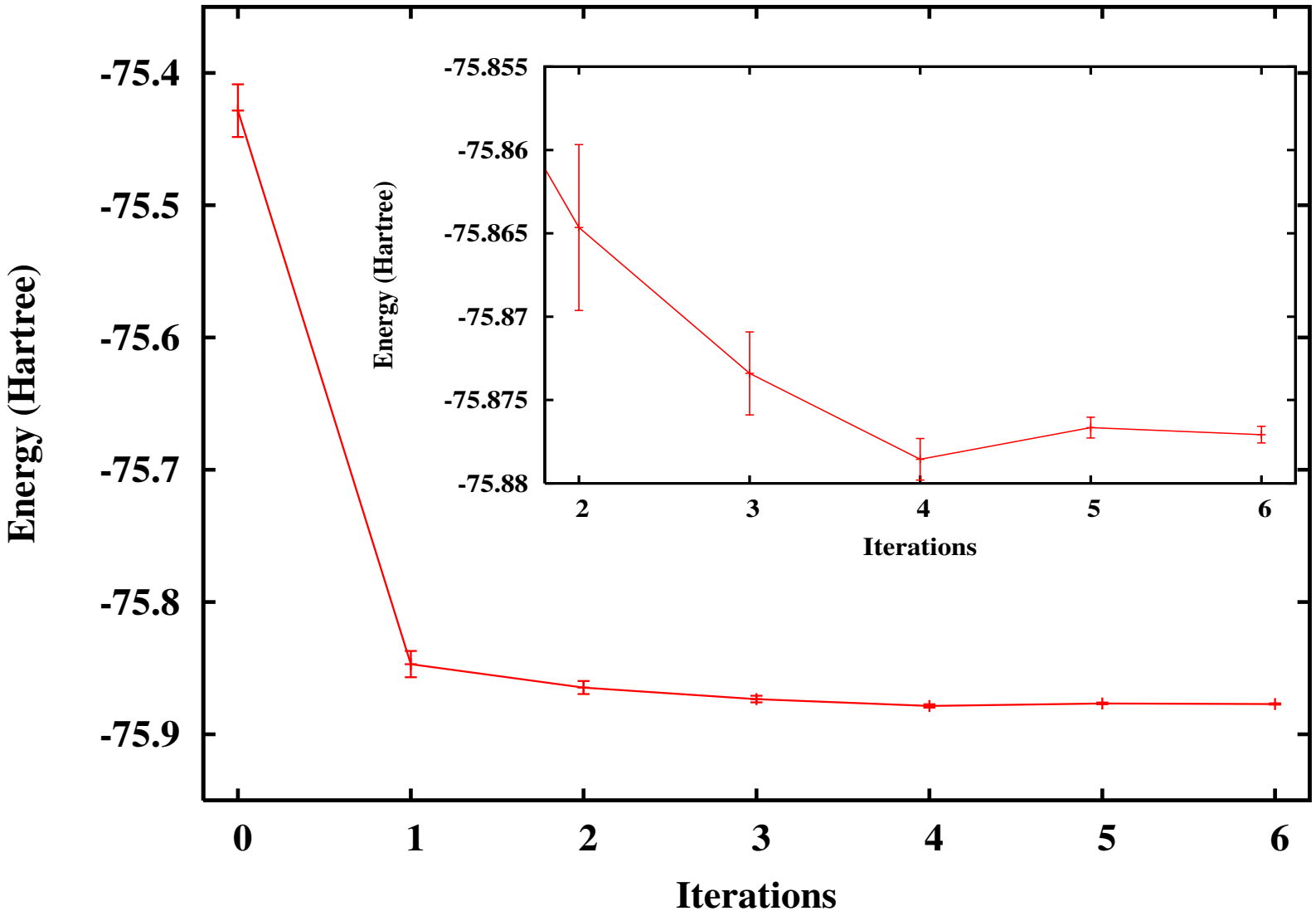
1. Can reduce the variance, without sacrificing appreciably the energy, by minimizing a linear combination, particularly since the ratio of hard to soft directions is 11 orders of magnitude.
2. Easy to do – obvious for Newton. Not obvious, but easy to do for linear method as shown above.
3. Measure of efficiency of the wave function is $\sigma^2 T_{\rm corr}$.

Convergence of energy of decapentaene $C_{10}H_{12}$



Simultaneous optimization of Jastrow, CSFs and orbitals of all-electron C_2 with linear method

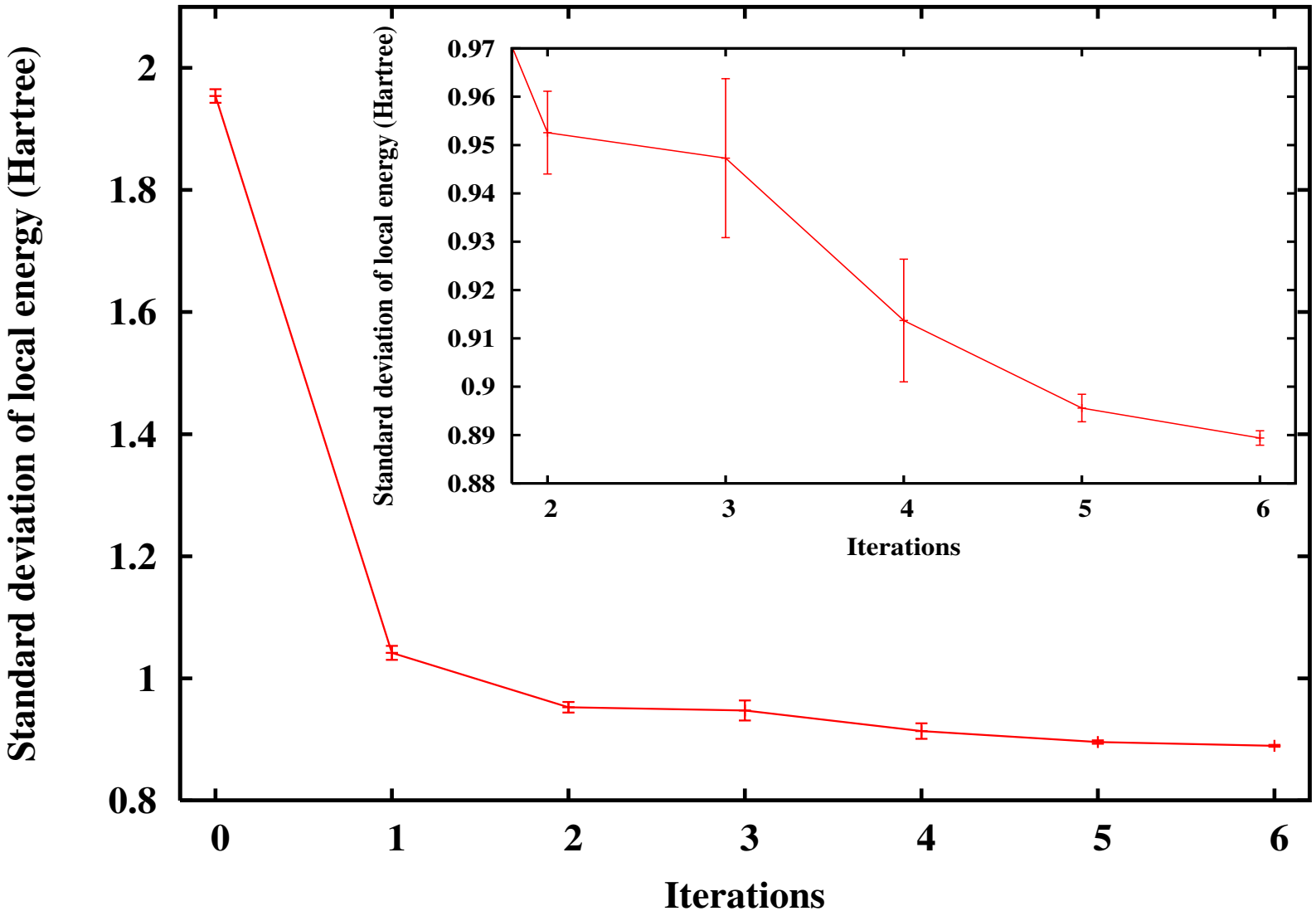
J. Toulouse, CJU, J. Chem. Phys. (2007)



Converges in 4 iterations

Simultaneous optimization of Jastrow, CSFs and orbitals of all-electron C_2 with linear method

J. Toulouse, CJU, J. Chem. Phys. (2007)

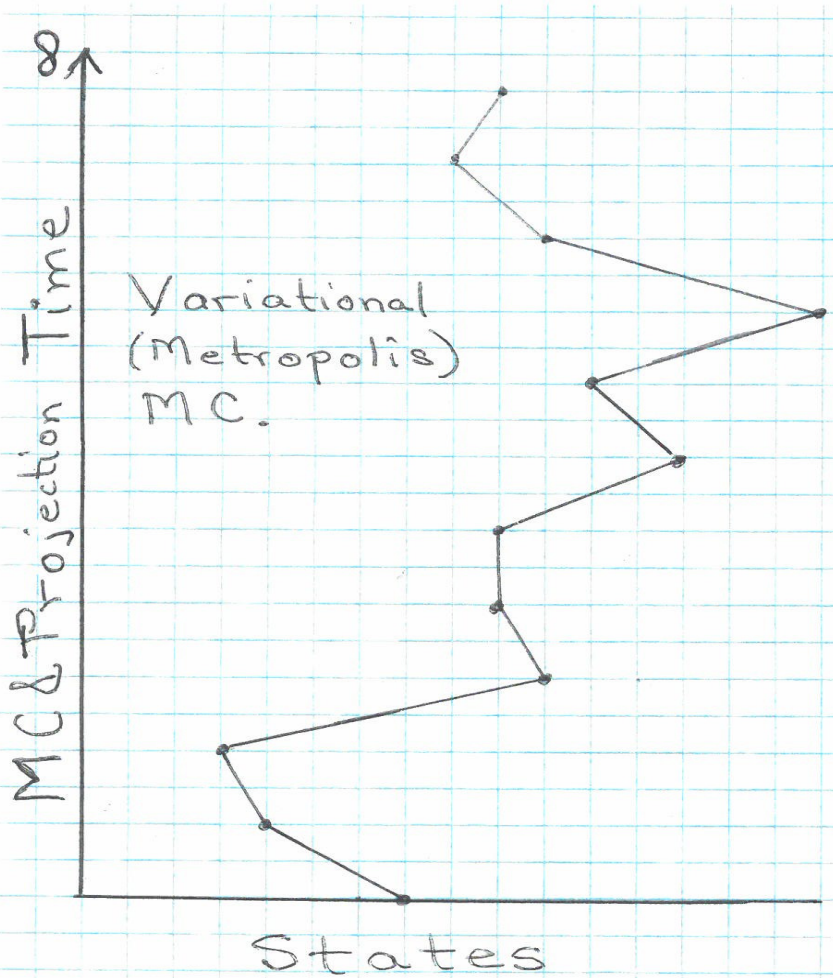


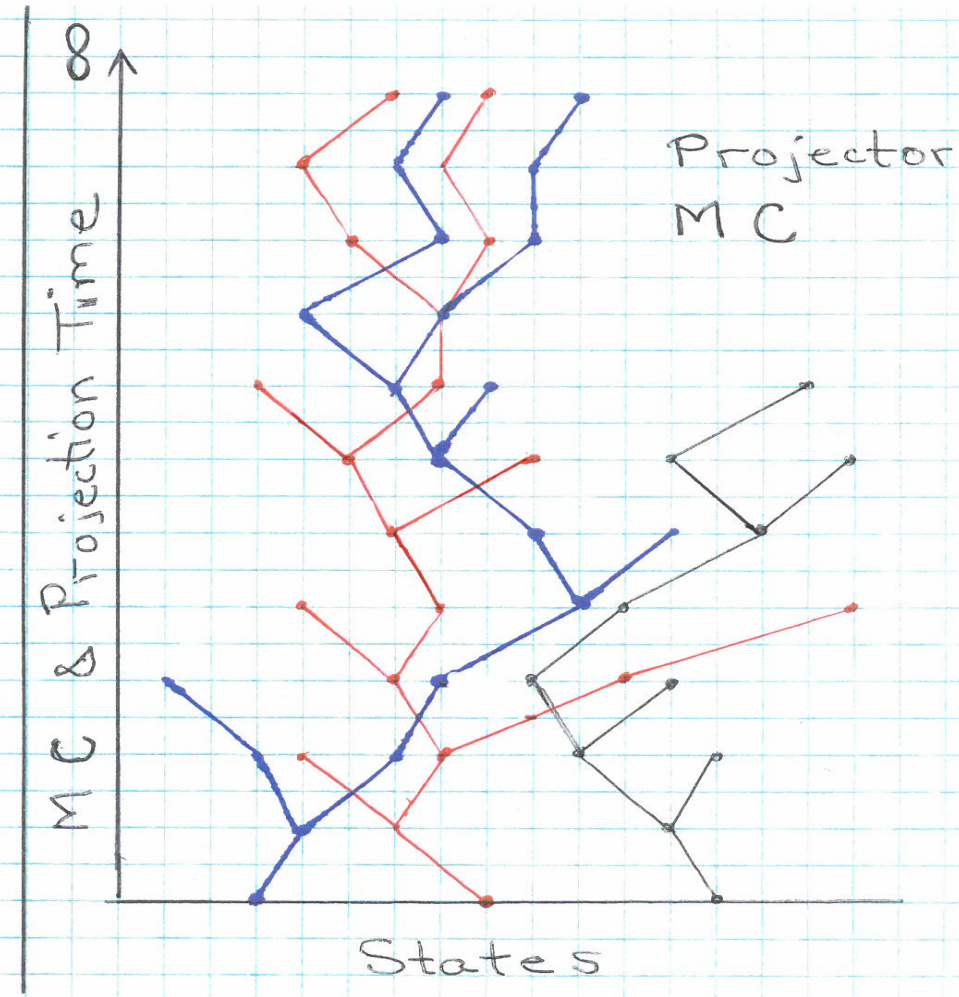
Converges in 5-6 iterations

Projector Monte Carlo Methods

Cyrus J. Umrigar

Schematic of VMC and PMC





Linear Projector in a Discrete Space

$\hat{P} = \hat{\mathbf{1}} + \tau(E_T \hat{\mathbf{1}} - \hat{H})$, space is: 2^{nd}-quant. space of ϕ_i^{orthog}, i.e., determinants

e.g. Full Configuration Interaction Quantum Monte Carlo (FCIQMC)
Booth, Thom, Alavi, JCP (2009), Cleland, Booth, Alavi, JCP (2010)

States are represented as bit-packed orbital occupation numbers.

Although Hilbert space can be huge, since $\hat{H}$ and therefore $\hat{P}$ is sparse in the chosen basis, it is possible to sample from all connected states.

1. Starting from state i, sample state $j \neq i$ with probability T_{ji}.
   ($T_{ji} \neq 0$, if $P_{ji} \neq 0$)
2. Reweight state j by P_{ji}/T_{ji}
3. Reweight state i by P_{ii}
4. Branch states with appropriate probabilities to have unit weight walkers.

If this were the entire algorithm, there would be a fatal sign problem.
Discuss this later.

Diffusion Monte Carlo – Short-time Green's function

$\hat{P}(\tau) = \exp(\tau(E_T\hat{\mathbf{1}} - \hat{H}))$, $\quad |\phi_i\rangle = |\mathbf{R}\rangle$ **, walkers are 1^{st}-quantized**

$$-\frac{1}{2}\nabla^2\psi(\mathbf{R},t) \; + \; (\mathcal{V}(\mathbf{R}) - E_{\rm T})\psi(\mathbf{R},t) \; = \; -\frac{\partial\psi(\mathbf{R},t)}{\partial t}$$

Combining the diffusion Eq. and the rate Eq. Green's functions:

$$\langle \mathbf{R}^{'}|\hat{P}(\tau)|\mathbf{R}\rangle \equiv G(\mathbf{R}^{'},\mathbf{R},\tau) \; \approx \; \frac{1}{(2\pi\tau)^{3N/2}} e^{\left[-\frac{(\mathbf{R}^{'}-\mathbf{R})^2}{2\tau} + \left\{E_{\rm T} - \frac{(\mathcal{V}(\mathbf{R}^{'})+\mathcal{V}(\mathbf{R}))}{2}\right\}\tau\right]}$$

The wavefunction,$\psi(\mathbf{R}^{'},t+\tau)$, evolves according to the integral equation,

$$\psi(\mathbf{R}^{'},t+\tau) \; = \; \int d\mathbf{R}\, G(\mathbf{R}^{'},\mathbf{R},\tau)\psi(\mathbf{R},t).$$

Columns of $G(\mathbf{R}^{'},\mathbf{R},\tau)$ not normalized to 1, so weights and/or branching needed.
Potential energy $\mathcal{V} \to \pm\infty$, so fluctuations in weights and/or population are huge!

Expectation values

There is an additional problem that the contribution that various MC points make to expectation values is proportional to $\Psi_{\rm T}(\mathbf{R})$:

$$
\begin{aligned}
E &= \frac{\int d\mathbf{R}\, \Psi_0(\mathbf{R}) H(\mathbf{R}) \Psi_{\rm T}(\mathbf{R})}{\int d\mathbf{R}\, \Psi_0(\mathbf{R}) \Psi_{\rm T}(\mathbf{R})} \\
&\approx \frac{\sum_i^{N_{\rm MC}} H(\mathbf{R}) \Psi_{\rm T}(\mathbf{R})}{\sum_i^{N_{\rm MC}} \Psi_{\rm T}(\mathbf{R})}
\end{aligned}
$$

This is inefficient for Bosonic systems, and is impossible for Fermionic systems since one gets 0/0.

The problems on previous viewgraph and this one are solved (at the price of biased expectation values) by using importance sampling and fixed-node boundary conditions with the approximate wavefunctions $\Psi_{\rm T}(\mathbf{R})$. In the limit that $\Psi_{\rm T} \to \Psi_0$ the weights of the walkers do not fluctuate at all and every MC point contributes equally to the expectation values.

In order to have finite variance, it is necessary that $\Psi_{\rm T}$ never be nonzero where $\Psi_{\rm G}$ is zero. In fact the usual practice in DMC is $\Psi_{\rm G} = \Psi_{\rm T}$ and so in this section we will not distinguish between them.

Diffusion Monte Carlo – Importance Sampled Fixed-Node Green's Function

Importance sampling: Multiply imaginary-time the Schrödinger equation

$$-\frac{1}{2}\nabla^2\Psi(\mathbf{R},t) \;+\; (\mathcal{V}(\mathbf{R})-E_{\rm T})\Psi(\mathbf{R},t) \;=\; -\frac{\partial\Psi(\mathbf{R},t)}{\partial t}$$

by $\Psi_{\rm T}(\mathbf{R})$ and rearranging terms we obtain

$$-\frac{\nabla^2}{2}(\Psi\Psi_{\rm T}) \;+\; \nabla\cdot\left(\frac{\nabla\Psi_{\rm T}}{\Psi_{\rm T}}\Psi\Psi_{\rm T}\right) \;+\; \left(\underbrace{\frac{-\nabla^2\Psi_{\rm T}}{2\Psi_{\rm T}}+\mathcal{V}}_{E_{\rm L}(\mathbf{R})}-E_{\rm T}\right)(\Psi\Psi_{\rm T}) \;=\; -\frac{\partial(\Psi\Psi_{\rm T})}{\partial t}$$

defining $f(\mathbf{R},t)=\Psi(\mathbf{R},t)\Psi_{\rm T}(\mathbf{R})$, this is

$$\underbrace{-\frac{1}{2}\nabla^2 f}_{\text{diffusion}} \;+\; \underbrace{\nabla\cdot\left(\frac{\nabla\Psi_{\rm T}}{\Psi_{\rm T}}f\right)}_{\text{drift}} \;+\; \underbrace{(E_{\rm L}(\mathbf{R})-E_{\rm T})\,f}_{\text{growth/decay}} \;=\; -\frac{\partial f}{\partial t}$$

Since we know the exact Green function for any one term on LHS, an approximation is:

$$\tilde{G}(\mathbf{R}',\mathbf{R},\tau)\approx\frac{1}{(2\pi\tau)^{3N/2}}e^{\left[-\frac{(\mathbf{R}'-\mathbf{R}-\mathbf{V}\tau)^2}{2\tau}+\left\{E_{\rm T}-\frac{(E_{\rm L}(\mathbf{R}')+E_{\rm L}(\mathbf{R}))}{2}\right\}\tau\right]}$$

Diffusion Monte Carlo with Importance Sampling

$$\tilde{G}(\mathbf{R}',\mathbf{R},\tau) \approx \frac{1}{(2\pi\tau)^{3N/2}} e^{\left[-\frac{(\mathbf{R}'-\mathbf{R}-\mathbf{V}\tau)^2}{2\tau}+\left\{E_{\rm T}-\frac{(E_{\rm L}(\mathbf{R}')+E_{\rm L}(\mathbf{R}))}{2}\right\}\tau\right]}$$

The importance-sampled Green function has $E_{\rm L}(\mathbf{R})$ in the reweighting factor, which behaves MUCH better than the potential, $V(\mathbf{R})$. $V(\mathbf{R})$ diverges to $\pm\infty$ at particle coincidences whereas $E_{\rm L}(\mathbf{R})$ goes to a constant, E_0, as $\Psi_{\rm T} \rightarrow \Psi_0$. In addition it has a drift term that keeps the particles in the important regions, rather than relying on the reweighting to achieve that.

Even this does not always work. Why?

Diffusion Monte Carlo with Importance Sampling

$$\tilde{G}(\mathbf{R}',\mathbf{R},\tau) \approx \frac{1}{(2\pi\tau)^{3N/2}} e^{\left[-\frac{(\mathbf{R}'-\mathbf{R}-\mathbf{V}\tau)^2}{2\tau}+\left\{E_{\rm T}-\frac{(E_{\rm L}(\mathbf{R}')+E_{\rm L}(\mathbf{R}))}{2}\right\}\tau\right]}$$

The importance-sampled Green function has $E_{\rm L}(\mathbf{R})$ in the reweighting factor, which behaves MUCH better than the potential, $V(\mathbf{R})$. $V(\mathbf{R})$ diverges to $\pm\infty$ at particle coincidences whereas $E_{\rm L}(\mathbf{R})$ goes to a constant, E_0, as $\Psi_{\rm T} \rightarrow \Psi_0$. In addition it has a drift term that keeps the particles in the important regions, rather than relying on the reweighting to achieve that.

Even this does not always work. Why?
The above importance sampled Green function leads to an "infinite variance" estimate for systems other than Bosonic ground states!!

Singularities of Green's function

CJU, Nightingale, Runge, JCP 1993

Region	Local energy $E_{\rm L}$	Velocity $\mathbf{V}$
Nodes	$E_{\rm L} \sim \pm\frac{1}{R_\perp}$ for $\Psi_{\rm T}$ $E_{\rm L} = E_0$ for Ψ_0	$V \sim \frac{1}{R_\perp}$ for both $\Psi_{\rm T}$ and Ψ_0
e-n and e-e coincidences	$E_{\rm L} \sim \frac{1}{x}$ if cusps not imposed $E_{\rm L}$ finite if cusps are imposed $E_{\rm L} = E_0$ for Ψ_0	V has a discontinuity for both $\Psi_{\rm T}$ and Ψ_0

All the above infinities and discontinuities cause problems, e.g.,

$$\int_0^a dx\, E_{\rm L} = \int_0^a dx \left(\frac{1}{x}\right) = \pm\infty$$

$$\int_0^a dx\, E_{\rm L}^2 = \int_0^a dx \left(\frac{1}{x}\right)^2 = \infty$$

Modify Green's function, by approximately integrating $E_{\rm L}$ and $\mathbf{V}$ over path, taking account of the singularities, at no additional computational cost.

Nonanalyticity of velocity near a node

CJU, Nightingale, Runge, JCP 1993

Linear approximation to $\Psi_{\rm T}$ (knowing $\mathbf{V} = \nabla\Psi_{\rm T}/\Psi_{\rm T}$):

$$\begin{aligned} \Psi_{\rm T}(\mathbf{R}') &= \Psi_{\rm T}(\mathbf{R}) + \nabla\Psi_{\rm T}(\mathbf{R})\cdot(\mathbf{R}'-\mathbf{R}) \\ &\propto 1 + \mathbf{V}\cdot(\mathbf{R}'-\mathbf{R}) \end{aligned}$$

The average velocity over the time-step τ is:

$$\bar{\mathbf{V}} = \frac{-1+\sqrt{1+2V^2\tau}}{V^2\tau}\mathbf{V} \rightarrow \begin{cases} \mathbf{V} & \text{if } V^2\tau \ll 1 \\ \sqrt{\frac{2}{\tau}}\hat{\mathbf{V}} & \text{if } V^2\tau \gg 1 \end{cases}$$

Infinite local energy near node

Make similar improvement to the growth/decay term of the Green's function by averaging of the local energy over time-step τ.

Discontinuity of velocity at particle coincidences

The e-N coincidence is more important than e-e coincidences because the wavefunction is larger in magnitude there.
Sample from linear combination of drifted Gaussians and exponential centered on nearest nucleus.

Infinite local energy near particle coincidences

Kato, Pure Appl. Math (1957), Pack and Byers-Brown, JCP, (1966), 2^{nd} order, Tew, JCP (2008)

Impose e-N and e-e cusp conditions on the wavefunction, so that divergence in potential energy is exactly canceled by divergence in kinetic energy.

$$\Psi = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} r^l \, f_{lm}(r) \, Y_l^m(\theta, \phi)$$

$$f_{lm}(r) \approx f_{lm}^0 \left[1 + \frac{q_i q_j \mu_{ij} r}{l+1} + O(r^2) \right]$$

with f_{lm}^0 being the first term in the expansion of $f_{lm}(r)$.

Familiar example: e-N cusp for s-state of Hydrogenic atom is $-Z$. e-e cusps are $1/2$ and $1/4$ for $\uparrow\downarrow$ and $\uparrow\uparrow$ respectively. (This is why we chose two of the parameters in the wavefunction in the lab to be -2 and $1/2$.)

Combining with Metropolis to reduce time-step error

Reynolds, Ceperley, Alder, Lester, JCP 1982

$$\underbrace{-\frac{1}{2}\nabla^2 f}_{\text{diffusion}} + \underbrace{\nabla\cdot\left(\frac{\nabla\psi_{\rm T}}{\psi_{\rm T}}f\right)}_{\text{drift}} + \underbrace{\left(E_{\rm L}(\mathbf{R})-E_{\rm T}\right)f}_{\text{growth/decay}} = -\frac{\partial f}{\partial t}$$

If we omit the growth/decay term then $|\Psi_{\rm T}|^2$ is the solution.
But we can sample $|\Psi_{\rm T}|^2$ exactly using Metropolis-Hastings! So, view $G(\mathbf{R}^{'},\mathbf{R},t)$ as being the proposal matrix $T(\mathbf{R}',\mathbf{R})$ and introduce accept-reject step after drift and diffusion steps.

Since some of the moves are rejected, use an effective $\tau_{\rm eff} < \tau$ in the reweighting term. , CJU, Nightingale, Runge, JCP (1993)

Zero-Variance zero-bias estimators for diagonal operators that do not commute with H

Assaraf and Caffarel, JCP (2003)

$$\begin{aligned} \delta E &= \mathcal{O}(|\delta\Psi|^2), \qquad \sigma^2(E_L) = \mathcal{O}(|\delta\Psi|^2) \\ \text{but} & \\ \delta O &= \mathcal{O}(|\delta\Psi|), \qquad \sigma^2(O_L) = \mathcal{O}(1) \end{aligned}$$

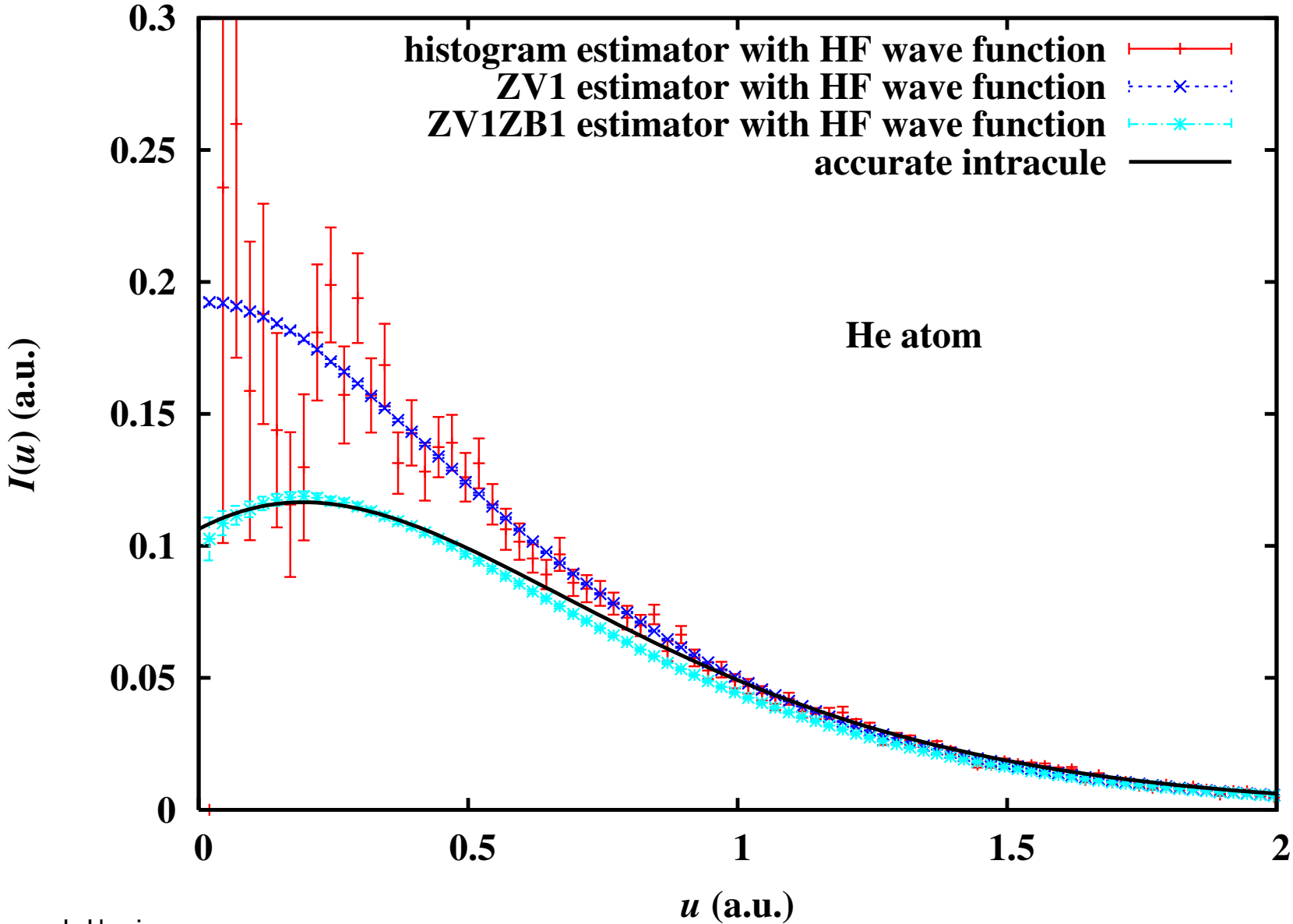
Define a λ-dependent Hamiltonian $\hat{H^\lambda} = \hat{H} + \lambda\hat{O}$ with an associated λ-dependent eigenfunction $\Psi_0^\lambda = \Psi_0 + \lambda\Psi_0' + \cdots$ and an approx. $\Psi^\lambda = \Psi + \lambda\Psi'$, then instead of $\frac{\langle\Psi|\hat{O}|\Psi\rangle}{\langle\Psi|\Psi\rangle}$ compute

$$\frac{\mathrm{d}}{\mathrm{d}\lambda}\left[\frac{\langle\Psi^\lambda|\hat{H}^\lambda|\Psi^\lambda\rangle}{\langle\Psi^\lambda|\Psi^\lambda\rangle}\right] = \frac{\langle\Psi|\hat{O}|\Psi\rangle}{\langle\Psi|\Psi\rangle} + \frac{\langle\Psi'|(\hat{H}-E)|\Psi\rangle}{\langle\Psi|\Psi\rangle} + \frac{\langle\Psi|(\hat{H}-E)|\Psi'\rangle}{\langle\Psi|\Psi\rangle}$$

$$\begin{aligned} \delta O &= \mathcal{O}(|\delta\Psi^2|) + \mathcal{O}(|\delta\Psi\delta\Psi'|) \\ \sigma^2(O_L) &= \mathcal{O}(|\delta\Psi^2|) + \mathcal{O}(|\delta\Psi\delta\Psi'|) + \mathcal{O}(|\delta\Psi'^2|) \end{aligned}$$

Application of ZVZB ideas to compute system averaged pair density (using intentionally bad Ψ)

Toulouse, Assaraf, CJU, JCP (2007)



Expectation values of operators

We wish to compute the pure (as opposed to mixed) expectation value

$$\langle A \rangle_{\rm pure} = \frac{\langle \Psi_0 | \hat{A} | \Psi_0 \rangle}{\langle \Psi_0 | \Psi_0 \rangle}$$

Consider various cases in order of increasing difficulty:

M.P. Nightingale, in Quantum Monte Carlo Methods in Physics and Chemistry, edited by M.P. Nightingale and CJU

1. $\hat{A}$ commutes with with $\hat{G}$ or equivalently $\hat{H}$ and is near-diagonal in chosen basis. (mixed expectation value)
2. $\hat{A}$ is diagonal in chosen basis. (forward/future walking) Liu, Kalos, and Chester, PRA (1974)
3. $\hat{A}$ is not diagonal in chosen basis, but, $A_{ij} \neq 0$ only when $G_{ij} \neq 0$. (forward/future walking)
4. $\hat{A}$ is not diagonal in chosen basis. (side walking) Barnett, Reynolds, Lester, JCP (1992)

Expectation values of operators

Factor the elements of the importance-sampled projector, $\tilde{G}(\mathbf{R}^{'},\mathbf{R})$, as products of elements of a stochastic matrix/kernel (elements are nonnegative and elements of column sum to 1), $\tilde{T}(\mathbf{R}^{'},\mathbf{R})$, and a reweight factor, $w(\mathbf{R}^{'},\mathbf{R})$.

$$\tilde{G}(\mathbf{R}^{'},\mathbf{R}) \;=\; \tilde{T}(\mathbf{R}^{'},\mathbf{R})w(\mathbf{R}^{'},\mathbf{R})$$

In the case of DMC

$$\tilde{T}(\mathbf{R}^{'},\mathbf{R}) \;=\; G_{\rm dif}(\mathbf{R}^{'},\mathbf{R}^{''})\; G_{\rm drift}(\mathbf{R}^{''},\mathbf{R}) \;=\; \frac{1}{(2\pi\tau)^{3N/2}} e^{-\frac{(\mathbf{R}^{'}-\mathbf{R}-\mathbf{V}\tau)^2}{2\tau}}$$

$$w(\mathbf{R}^{'},\mathbf{R}) \;=\; e^{\left\{ E_{\rm T}-\frac{(E_{\rm L}(\mathbf{R}^{'})+E_{\rm L}(\mathbf{R}))}{2} \right\}\tau}$$

For discrete state space and sparse H, define

$$\tilde{T}(\mathbf{R}^{'},\mathbf{R}) \;=\; \frac{\tilde{G}(\mathbf{R}^{'},\mathbf{R})}{\sum_{\mathbf{R}^{''}} \tilde{G}(\mathbf{R}^{''},\mathbf{R})}$$

$$w(\mathbf{R}^{'},\mathbf{R}) \;=\; w(\mathbf{R}) \;=\; \sum_{\mathbf{R}^{''}} \tilde{G}(\mathbf{R}^{''},\mathbf{R})$$

1) $\hat{A}$ commutes with with $\hat{H}$ and is near-diagonal in chosen basis

By *near diagonal* we mean that either:

1. In discrete space $\hat{A}$ is sufficiently sparse that when walker is at state i, $A_{\mathrm{L},i} = \sum_j g_j A j i / g_i$ can be computed sufficiently quickly, or
2. In continuous space $\hat{A}$ has only local and local-derivative terms, e.g., $\frac{-1}{2}\sum_i \nabla_i^2 + V(\mathbf{R})$.

Since $\hat{A}$ commutes with with $\hat{H}$ the mixed estimator equals the pure estimator

$$\langle A \rangle_{\mathrm{mix}} = \frac{\langle \Psi_0 | \hat{A} | \Psi_{\mathrm{T}} \rangle}{\langle \Psi_0 | \Psi_{\mathrm{T}} \rangle} = \frac{\langle \Psi_0 | \hat{A} | \Psi_0 \rangle}{\langle \Psi_0 | \Psi_0 \rangle} = \langle A \rangle_{\mathrm{pure}}$$

1) $\hat{A}$ commutes with with $\hat{H}$ and is near-diagonal in chosen basis

$$
\begin{aligned}
\langle A \rangle &= \frac{\langle \Psi_{\rm T}|\hat{A}|\Psi_0\rangle}{\langle \Psi_{\rm T}|\Psi_0\rangle} = \frac{\langle \Psi_{\rm T}|\hat{A}G^p(\tau)|\Psi_{\rm T}\rangle}{\langle \Psi_{\rm T}|G^p(\tau)|\Psi_{\rm T}\rangle} \\
&= \frac{\sum_{\mathbf{R}_p\cdots\mathbf{R}_0} A\Psi_{\rm T}(\mathbf{R}_p)\left(\prod_{i=0}^{p-1} G(\mathbf{R}_{i+1},\mathbf{R}_i)\right)\Psi_{\rm T}(\mathbf{R}_0)}{\sum_{\mathbf{R}_p\cdots\mathbf{R}_0} \Psi_{\rm T}(\mathbf{R}_p)\left(\prod_{i=0}^{p-1} G(\mathbf{R}_{i+1},\mathbf{R}_i)\right)\Psi_{\rm T}(\mathbf{R}_0)} \\
&= \frac{\sum_{\mathbf{R}_p\cdots\mathbf{R}_0} \frac{A\Psi_{\rm T}(\mathbf{R}_p)}{\Psi_{\rm T}(\mathbf{R}_p)}\left(\prod_{i=0}^{p-1} \tilde{G}(\mathbf{R}_{i+1},\mathbf{R}_i)\right)(\Psi_{\rm T}(\mathbf{R}_0))^2}{\sum_{\mathbf{R}_p\cdots\mathbf{R}_0} \left(\prod_{i=0}^{p-1} \tilde{G}(\mathbf{R}_{i+1},\mathbf{R}_i)\right)(\Psi_{\rm T}(\mathbf{R}_0))^2} \\
&= \frac{\sum_{t=T_{\rm eq}+1}^{T_{\rm eq}+T} A_L(\mathbf{R}_t)W_t}{\sum_{t=T_{\rm eq}+1}^{T_{\rm eq}+T} W_t} \quad \text{since MC pts. from } \left(\prod_{i=0}^{p-1} \tilde{T}(\mathbf{R}_{i+1},\mathbf{R}_i)\right)(\Psi_{\rm T}(\mathbf{R}_0))^2
\end{aligned}
$$

$W_t = \prod_{i=0}^{p-1} w(\mathbf{R}_{t-i},\mathbf{R}_{t-i-1})$ or better $W_t = \prod_{i=0}^{T_{\rm eq}+t-1} w(\mathbf{R}_{T_{\rm eq}+t-i},\mathbf{R}_{T_{\rm eq}+t-i-1})$.

Branching (described later) is used to prevent inefficiency due wide disparity in weight products.

2) Expectation values of diagonal operators that do not commute with $\hat{H}$

DMC straightforwardly gives us

$$\langle A\rangle_{\rm mix} = \frac{\langle \Psi_0|\hat{A}|\Psi_{\rm T}\rangle}{\langle \Psi_0|\Psi_{\rm T}\rangle} = \frac{\int d\mathbf{R}\ \langle \Psi_0|\mathbf{R}\rangle\ \langle \mathbf{R}|\hat{A}|\mathbf{R}\rangle\ \langle \mathbf{R}|\Psi_{\rm T}\rangle}{\int d\mathbf{R}\ \langle \Psi_0|\mathbf{R}\rangle\ \langle \mathbf{R}|\Psi_{\rm T}\rangle} = \frac{\int d\mathbf{R}\ \Psi_0(\mathbf{R})\ A(R)\ \Psi_{\rm T}(\mathbf{R})}{\int d\mathbf{R}\ \Psi_0(\mathbf{R})\ \Psi_{\rm T}(\mathbf{R})}$$

but we want

$$\langle A\rangle_{\rm pure} = \frac{\langle \Psi_0|\hat{A}|\Psi_0\rangle}{\langle \Psi_0|\Psi_0\rangle} = \frac{\int d\mathbf{R}\ \langle \Psi_0|\mathbf{R}\rangle\ \langle \mathbf{R}|\hat{A}|\mathbf{R}\rangle\ \langle \mathbf{R}|\Psi_0\rangle}{\int d\mathbf{R}\ \langle \Psi_0|\mathbf{R}\rangle\ \langle \mathbf{R}|\Psi_0\rangle} = \frac{\int d\mathbf{R}\ \Psi_0(\mathbf{R})\ A(R)\ \Psi_0(\mathbf{R})}{\int d\mathbf{R}\ \Psi_0(\mathbf{R})\ \Psi_0(\mathbf{R})}$$

Two possibilities: Extrapolated estimator and forward walking

1) Extrapolated estimator

$$\begin{aligned}
\langle A\rangle_{\rm DMC} &= \langle A\rangle_{\rm pure} + \mathcal{O}(||\Psi_{\rm T}-\Psi_0||) \\
\langle A\rangle_{\rm VMC} &= \langle A\rangle_{\rm pure} + \mathcal{O}(||\Psi_{\rm T}-\Psi_0||) \\
2\langle A\rangle_{\rm DMC} - \langle A\rangle_{\rm VMC} &= \langle A\rangle_{\rm pure} + \mathcal{O}(||\Psi_{\rm T}-\Psi_0||)^2
\end{aligned}$$

2) Expectation values of diagonal operators that do not commute with $\hat{H}$

Forward or Future Walking

$$\begin{aligned}
\langle A \rangle &= \frac{\langle \Psi_{\rm T} | G^p(\tau) \hat{A} G^{p'}(\tau) | \Psi_{\rm T} \rangle}{\langle \Psi_{\rm T} | G^{p+p'}(\tau) | \Psi_{\rm T} \rangle} \\
&= \frac{\sum_{\mathbf{R}_{p+p'} \cdots \mathbf{R}_0} A(\mathbf{R}_{p'}) \left(\prod_{i=0}^{p+p'-1} \tilde{G}(\mathbf{R}_{i+1}, \mathbf{R}_i) \right) (\Psi_{\rm T}(\mathbf{R}_0))^2}{\sum_{\mathbf{R}_{p+p'} \cdots \mathbf{R}_0} \left(\prod_{i=0}^{p+p'-1} \tilde{G}(\mathbf{R}_{i+1}, \mathbf{R}_i) \right) (\Psi_{\rm T}(\mathbf{R}_0))^2} \\
&= \frac{\sum_{t=T_{\rm eq}+1}^{T_{\rm eq}+T} A(\mathbf{R}_t) W_{t+p}}{\sum_{t=T_{\rm eq}+1}^{T_{\rm eq}+T} W_{t+p}}
\end{aligned}$$

$W_{t+p} = \prod_{i=0}^{p+p'-1} w(\mathbf{R}_{t+p-i}, \mathbf{R}_{t+p-i-1})$ (product over p' past and p future) or better $W_{t+p} = \prod_{i=0}^{T_{\rm eq}+t+p-1} w(\mathbf{R}_{T_{\rm eq}+t+p-i}, \mathbf{R}_{T_{\rm eq}+t+p-i-1})$, (product over entire past and p future generations).

The contribution to the expectation value is: the local operator at time t, multiplied by the weight at a future time $t+p$. Need to store $A(\mathbf{R}_t)$ for p generations.

Usual tradeoff: If p is small, there is some residual bias since $\Psi_{\rm T}$ has not been fully projected onto Ψ_0, whereas, if p is large the fluctuations of the descendent weights increases the statistical noise. (Since we use branching, weight factors from past are not a problem.) For very large p all walkers will be descended from the same ancestor.

(Mitochondrial Eve! – All humans alive today had same maternal ancestor 10^5 yrs ago.)

3) $\hat{A}$ is not diagonal in chosen basis, but, $A_{ij} \neq 0$ only when $G_{ij} \neq 0$

Forward or Future Walking

$$
\begin{aligned}
\langle A \rangle &= \frac{\langle \Psi_T | G^{p-1}(\tau) \hat{A} G^{p'}(\tau) | \Psi_T \rangle}{\langle \Psi_T | G^{p+p'}(\tau) | \Psi_T \rangle} \\
&= \frac{\sum_{\mathbf{R}_{p+p'} \cdots \mathbf{R}_0} \left(\prod_{i=p'+1}^{p+p'-1} \tilde{G}(\mathbf{R}_{i+1}, \mathbf{R}_i) \right) \tilde{A}(\mathbf{R}_{p'+1}, \mathbf{R}_{p'}) \left(\prod_{i=0}^{p'-1} \tilde{G}(\mathbf{R}_{i+1}, \mathbf{R}_i) \right) (\Psi_T(\mathbf{R}_0))^2}{\sum_{\mathbf{R}_{p+p'} \cdots \mathbf{R}_0} \left(\prod_{i=0}^{p+p'-1} \tilde{G}(\mathbf{R}_{i+1}, \mathbf{R}_i) \right) (\Psi_T(\mathbf{R}_0))^2} \\
&= \frac{\sum_{t=T_{\rm eq}+1}^{T_{\rm eq}+T} W_{t+p-1,t+1} \; a(\mathbf{R}_{t+1}, \mathbf{R}_t) \; W_{t,t-p'}}{\sum_{t=T_{\rm eq}+1}^{T_{\rm eq}+T} W_{t+p}}
\end{aligned}
$$

$$
a(\mathbf{R}_{t+1}, \mathbf{R}_t) = \frac{\tilde{A}(\mathbf{R}_{t+1}, \mathbf{R}_t)}{\tilde{T}(\mathbf{R}_{t+1}, \mathbf{R}_t)} = \frac{A(\mathbf{R}_{t+1}, \mathbf{R}_t)}{T(\mathbf{R}_{t+1}, \mathbf{R}_t)}
$$

$$
W_{t_2,t_1} = \prod_{i=t_1}^{t_2-1} w(\mathbf{R}_{i+1}, \mathbf{R}_i)
$$

Again, the product of p' past weights can be replaced by products of weights over entire past.

$\hat{A}$ 4) is not diagonal in chosen basis, and, $\exists$ some $A_{ij} \neq 0$ where $G_{ij} = 0$

Side Walking

Now it becomes necessary to have side walks that start from the backbone walk.

Just as we did for the importance-sampled projector, we factor $\tilde{A}$ into a Markov matrix and a reweighting factor.

The first transition of the side walk is made using this Markov matrix and and the rest of the side-walk using the usual Markov matrix.

The ends of the side-walks contribute to the expectation values.

This method is even more computationally expensive than forward walking, because one has to do an entire side walk long enough to project onto the ground state to get a single contribution to the expectation value.

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 basis of the $3N$ coordinates of the N electrons.

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

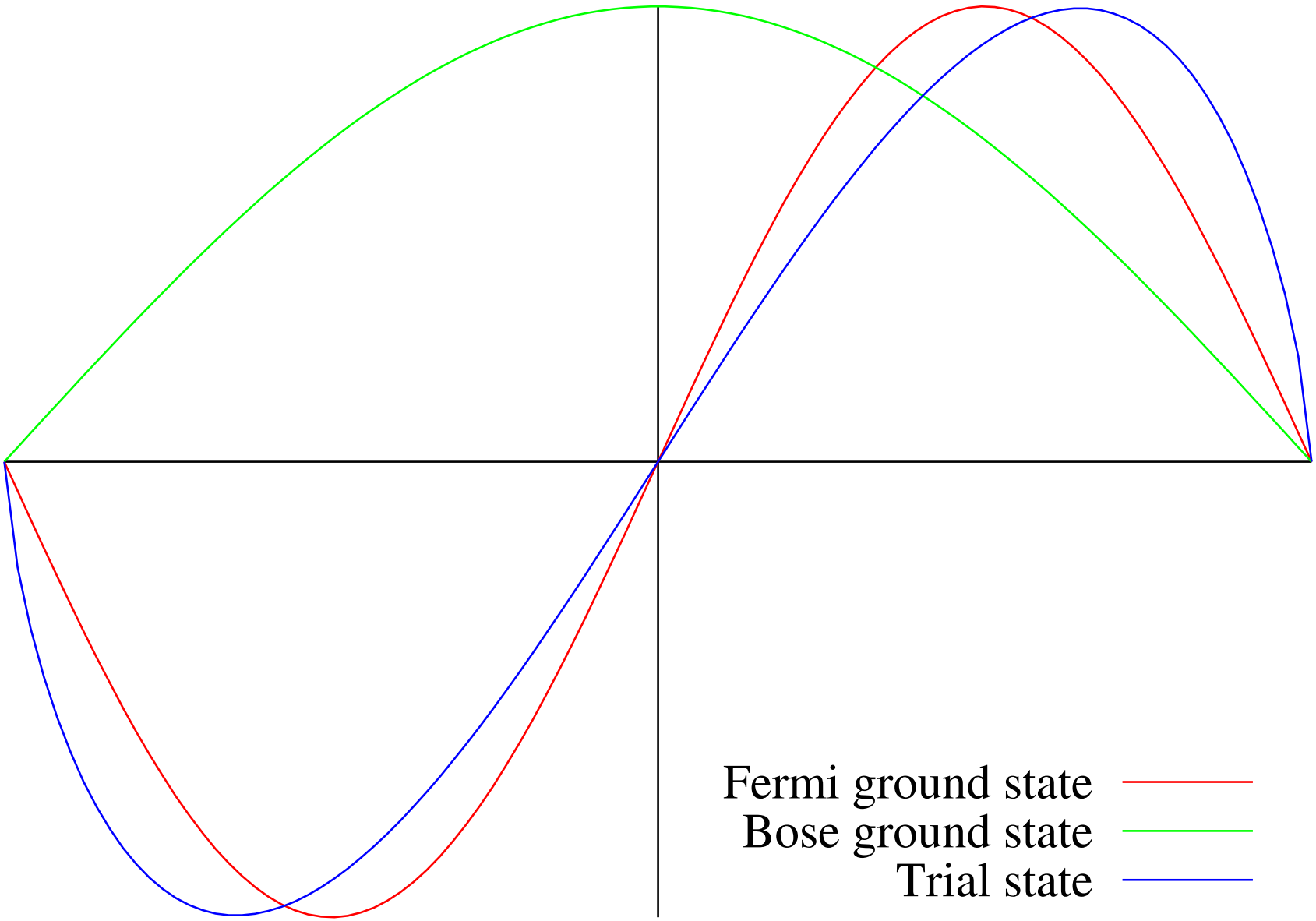
Problem: However, since the Bosonic energy is always lower than the Fermionic energy, the projected state is the Bosonic ground state.

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 $\Psi_{\rm T}(\mathbf{R})$.

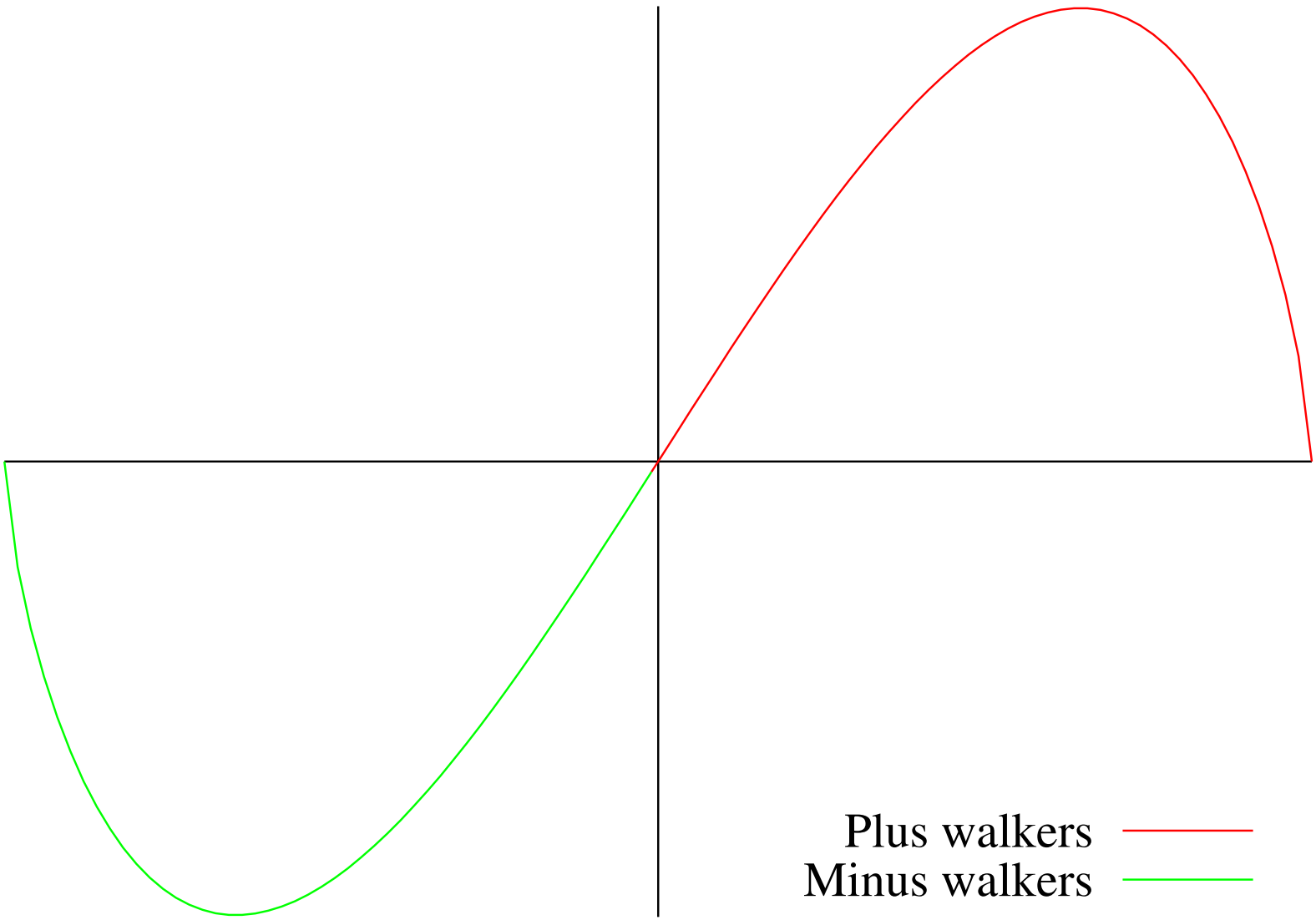
This gives an upper bound to the energy and becomes exact in the limit that $\Psi_{\rm T}$ has the same nodes as Ψ_0.

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

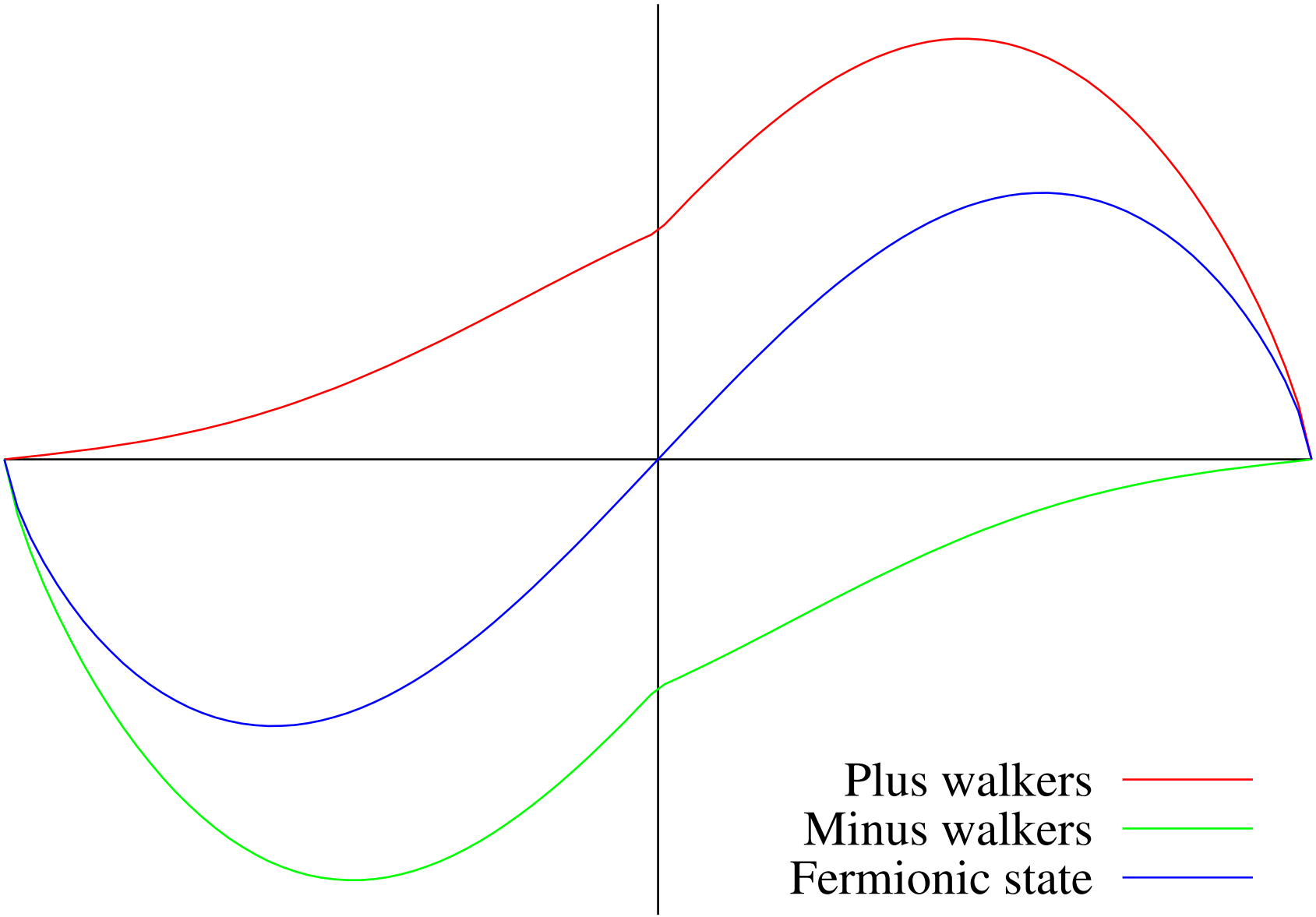


Sign Problem in 1^{st} Quantization and R space

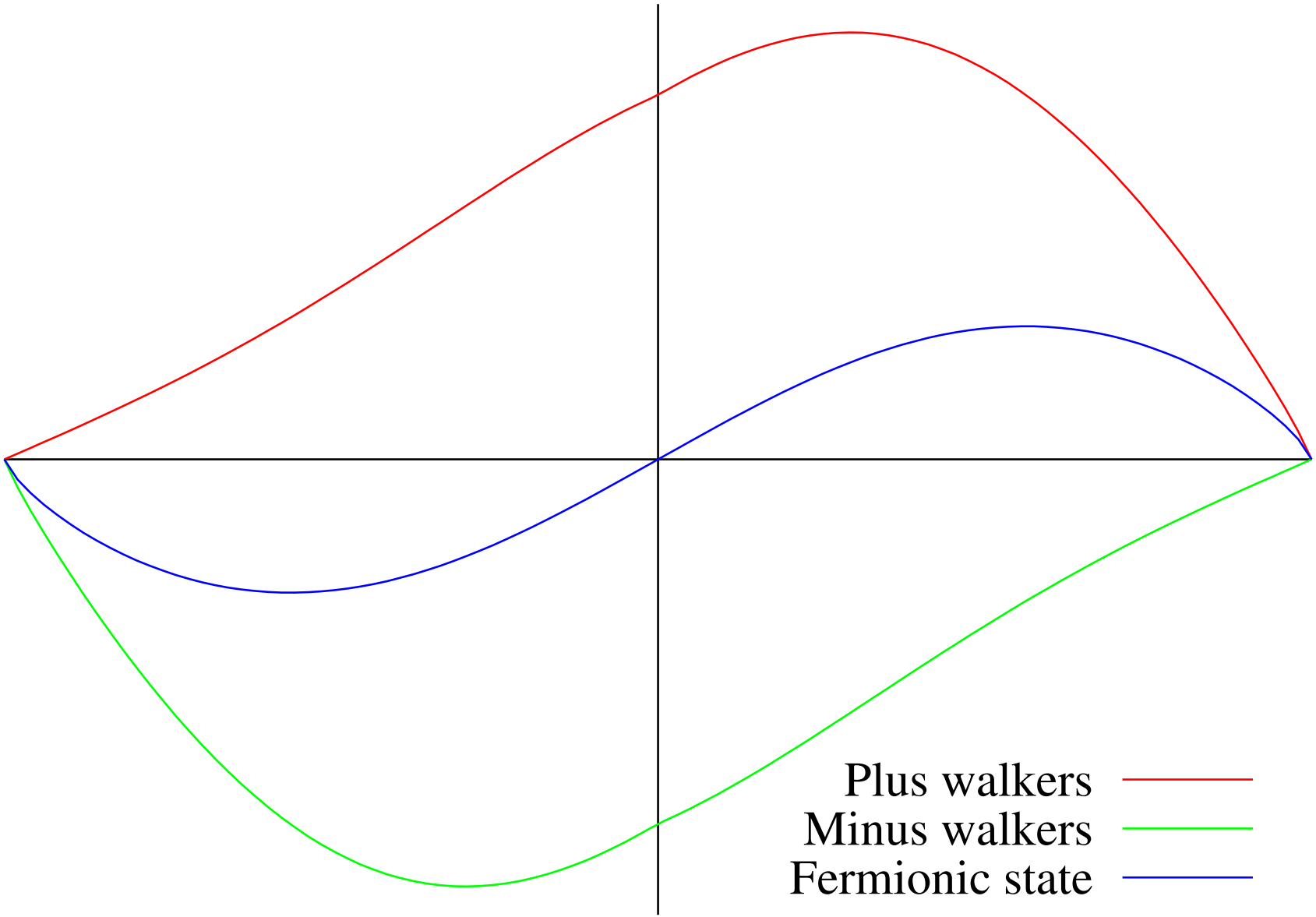
Start with equal + and - walkers, so no Bosonic component.



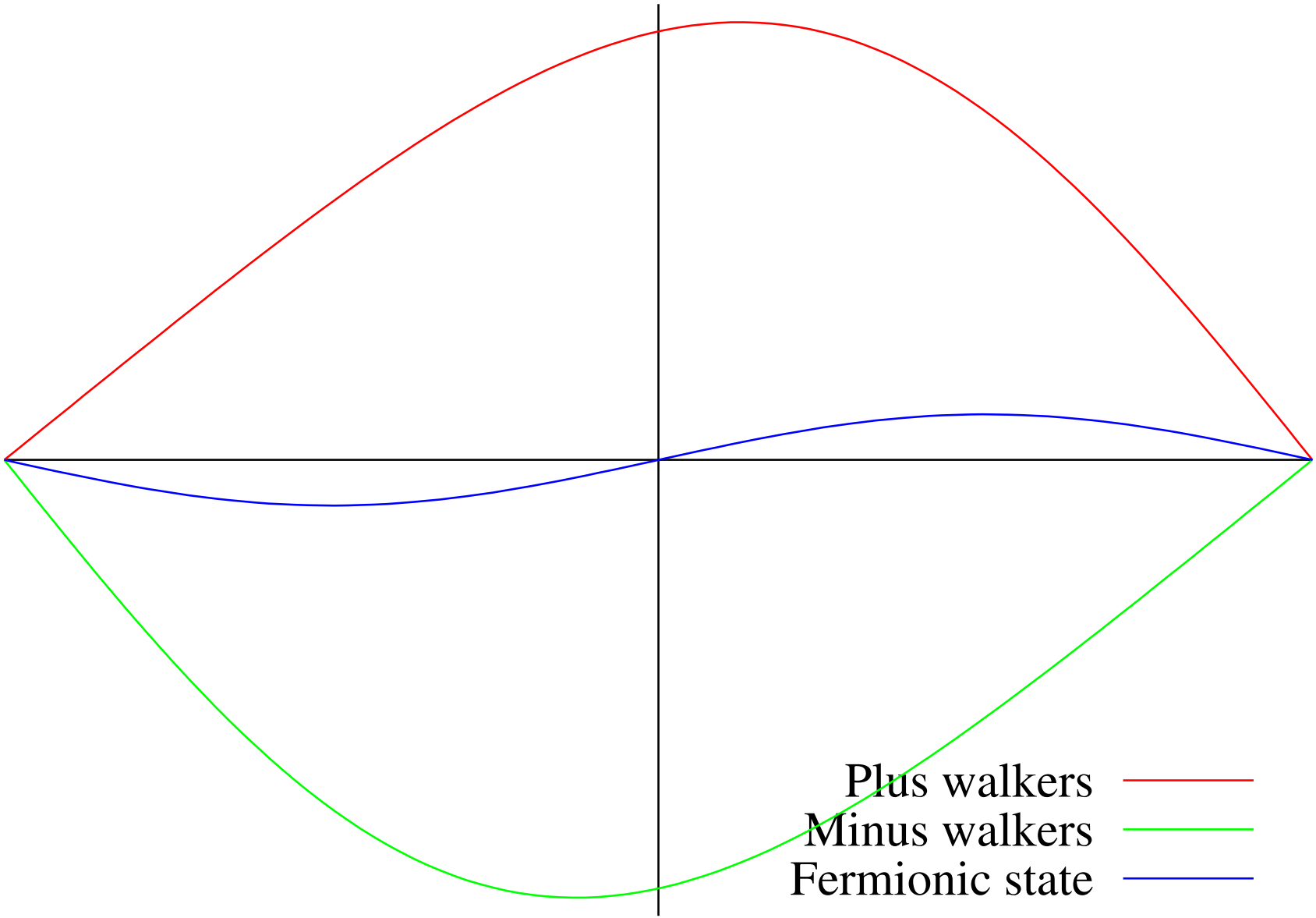
Sign Problem in 1^{st} Quantization and R space



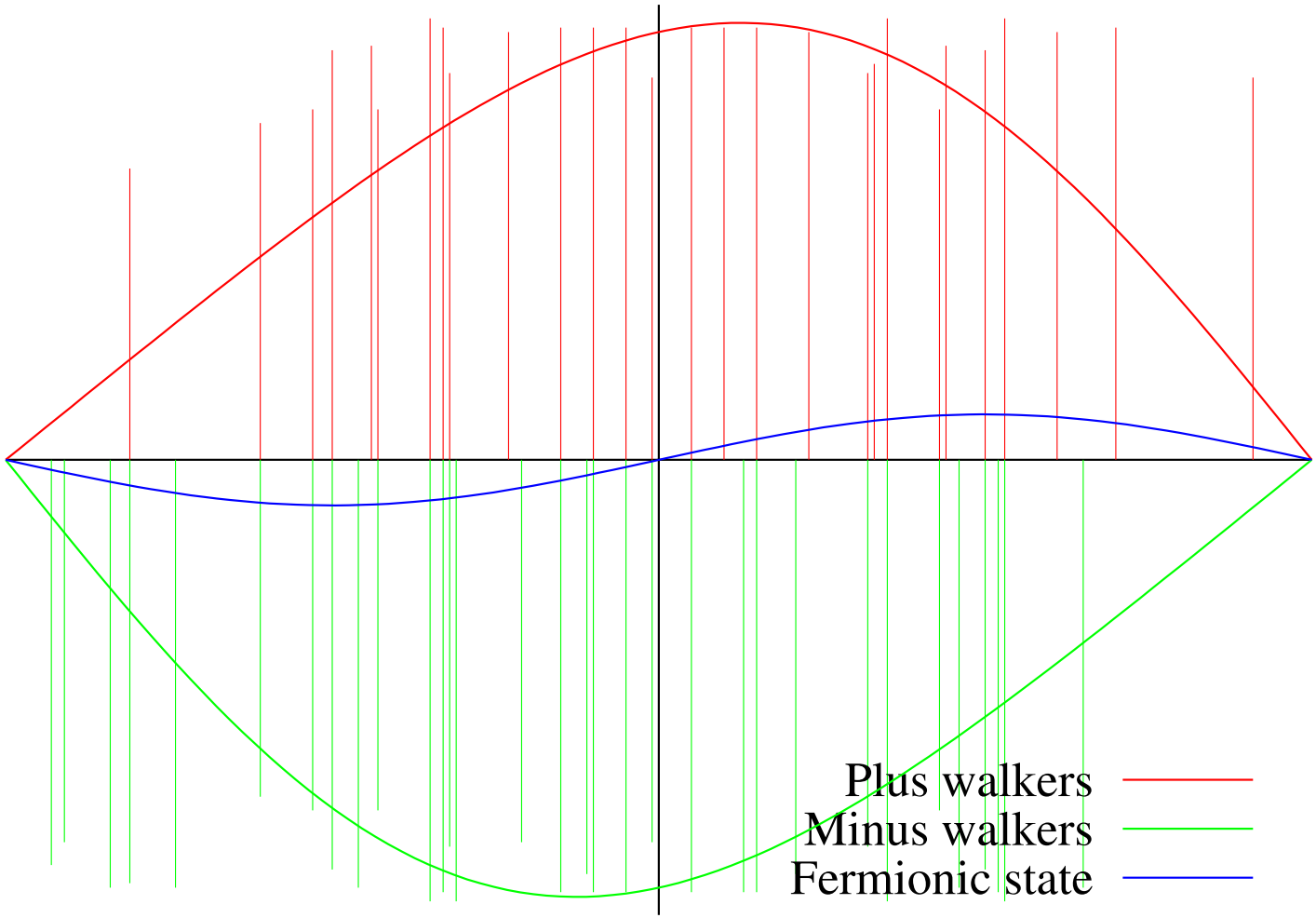
Sign Problem in 1^{st} Quantization and R space



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 component remains!

Sign Problem in 2^{nd} quantization

Walk is done in the basis 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 2^{nd}-quantized basis does not have a sign problem if:
The columns of the projector have the same sign structure aside from an overall sign, e.g.

$$P\Psi = \begin{bmatrix} + & - & + & + \\ - & + & - & - \\ + & - & + & + \\ + & - & + & + \end{bmatrix} \begin{bmatrix} + \\ - \\ + \\ + \end{bmatrix} = \begin{bmatrix} + \\ - \\ + \\ + \end{bmatrix}$$

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, Ohtsuka, Nagase, CPL 2008)

$\hat{P} = \hat{\mathbf{1}} + \tau(E_T\hat{\mathbf{1}} - \hat{H})$, space is: 2^{nd}-quantized ϕ_i^{orthog}, i.e., determinants

It is practical to have a population that is sufficiently large that cancellations in this discrete space can result in a finite signal to noise ratio for small systems in small basis sets. 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.

Semistochastic Quantum Monte Carlo (SQMC)

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

SQMC is hybrid of Exact Diagonalization and QMC

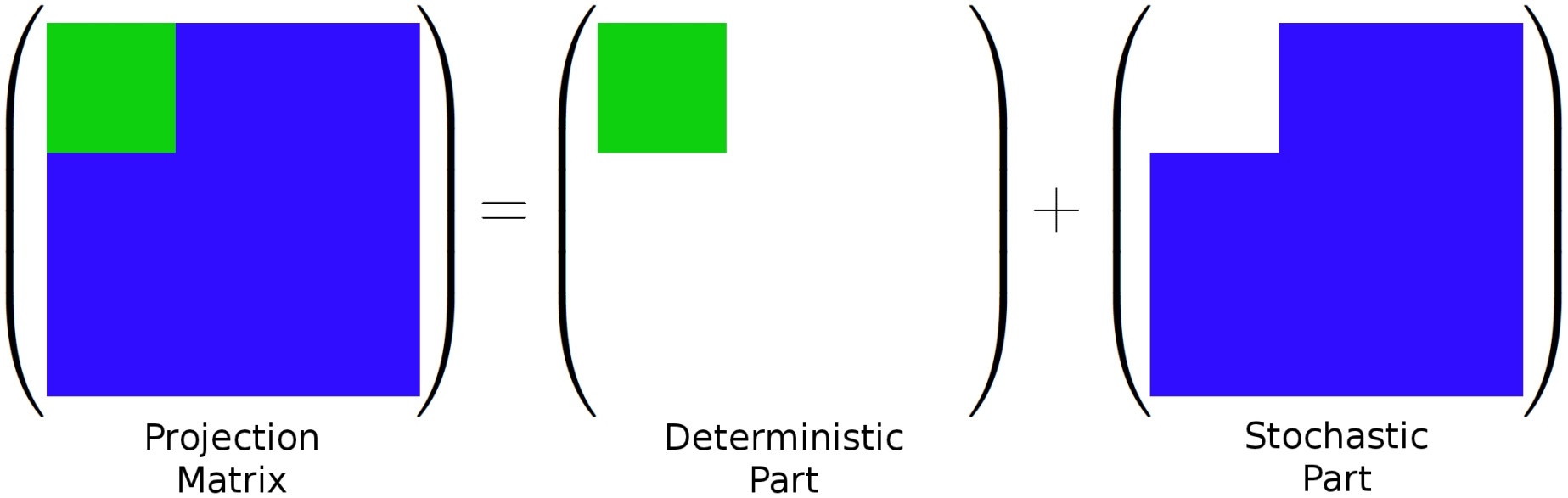
Exact diagonalization has no statistical error or sign problem but is limited to a small number of states ($\sim 10^{10}$ on a single core).

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

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 $P^{\mathcal{S}}$

Since $\quad P_{ij} = \delta_{ij} + \tau(E_T \delta_{ij} - H_{ij})$

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

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

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 states connected to the current wavefunction and keep the ones that make large contributions in 2^{nd}-order perturbation theory
3. Diagonalize in this space
4. iterate

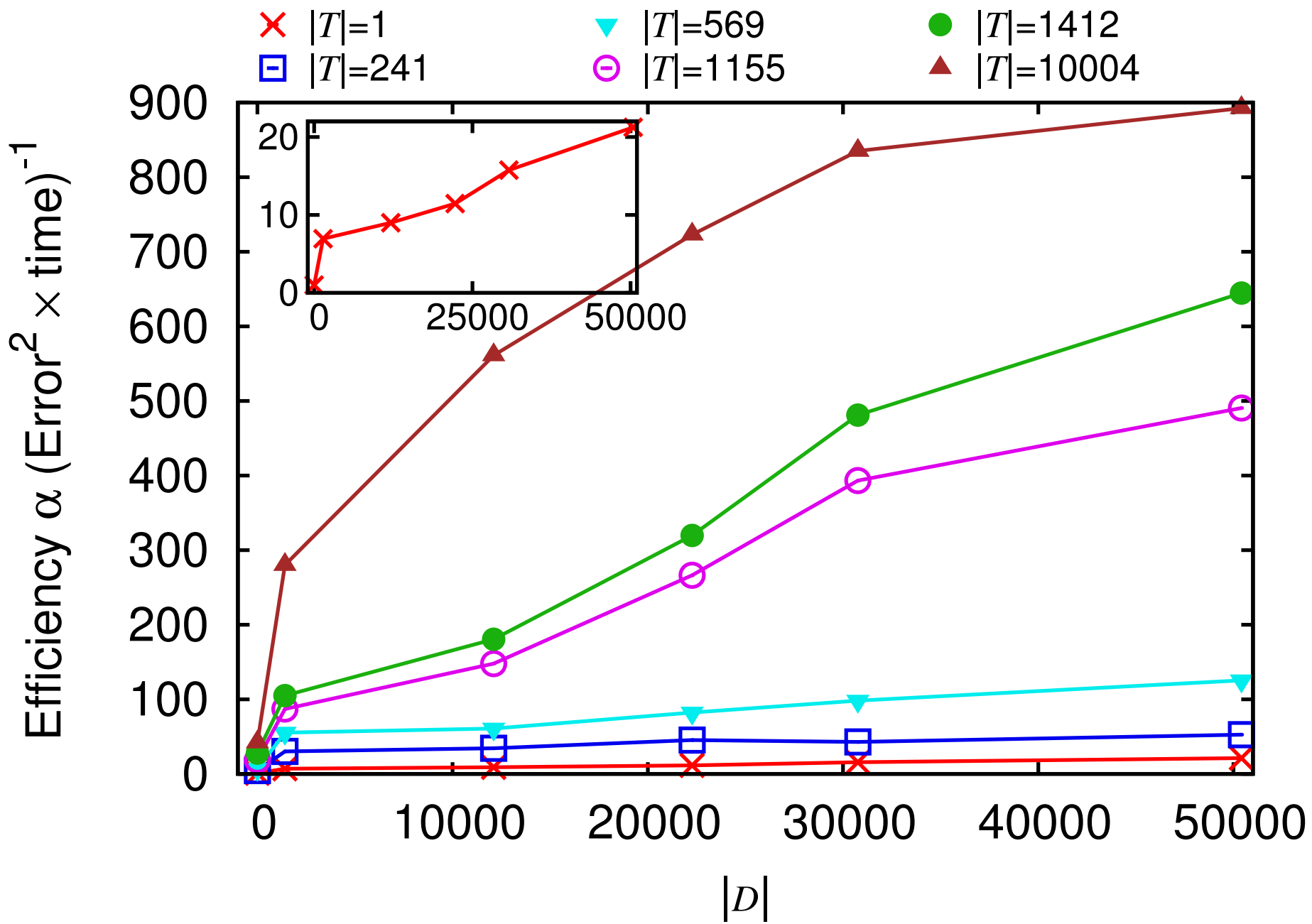
For some systems iterating 2-4 times can give large gain compared to iterating once, i.e., higher order excitations help a lot.

SQMC

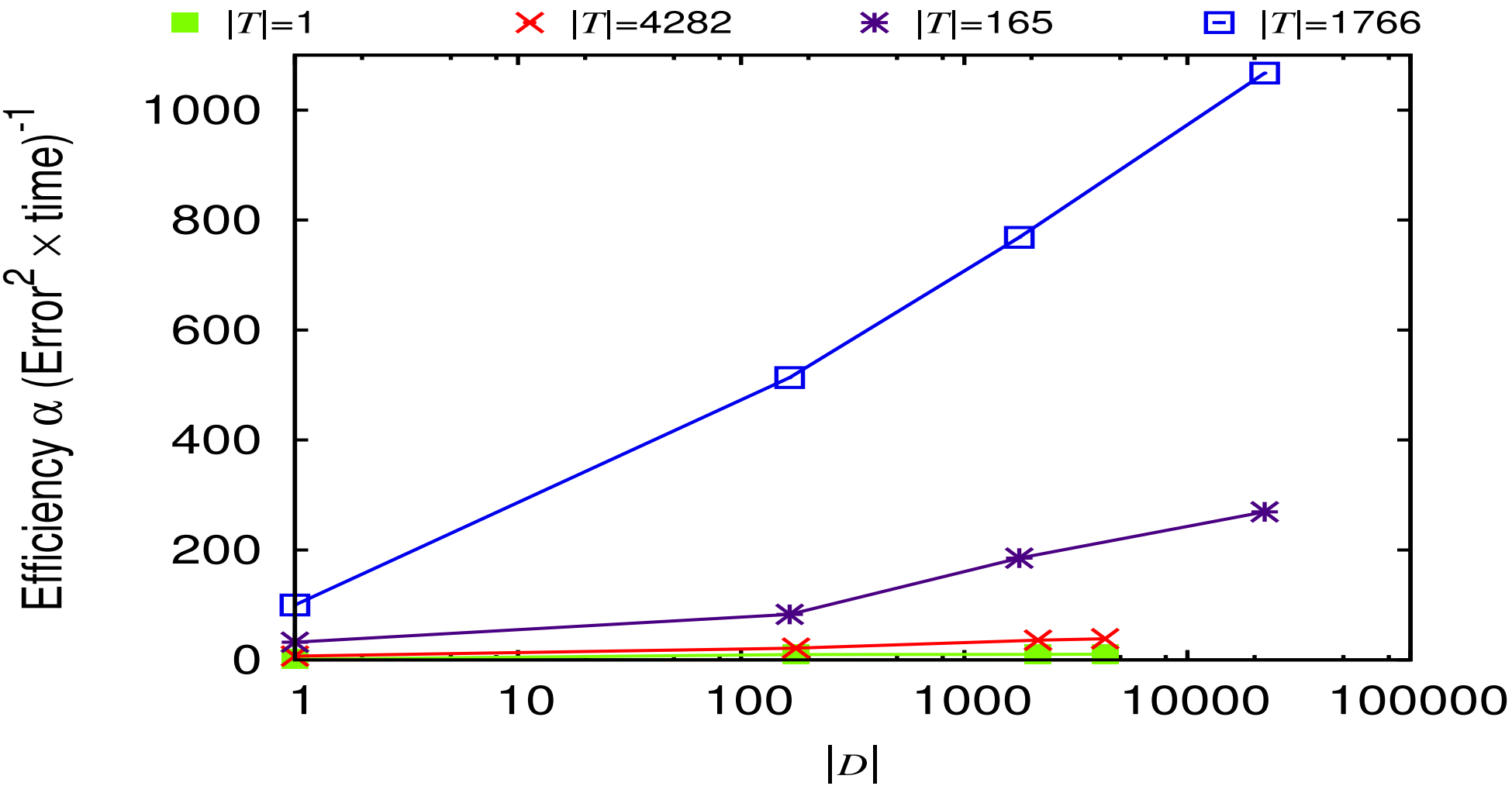
Main differences between SQMC and FCIQMC:

1. Deterministic projection in part of space
2. Multideterminantal $\Psi_{\rm T}$, particularly important for strongly correlated states
3. Real (rather than integer) weights, $|\psi(t)\rangle = \sum_{i=1}^{N} w_i(t)|\phi_i\rangle$

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



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



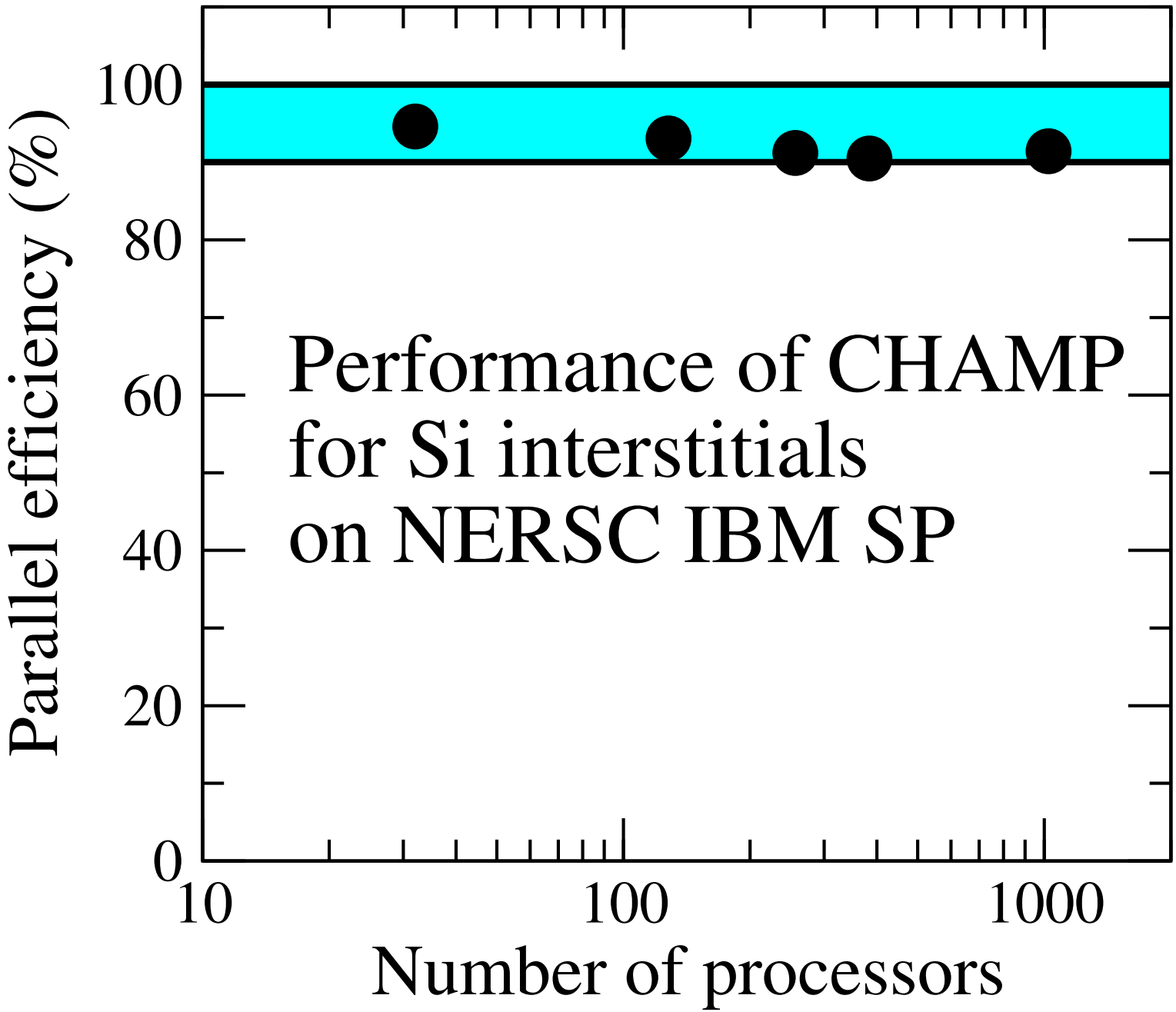
Wavefns. with 165 or 1766 dets. containing some 4^{th}-order excit. are much more efficient than wavefn. with 4282 dets. containing only upto 2^{nd}-order excit.

Comparison of DMC with FCIQMC/SQMC

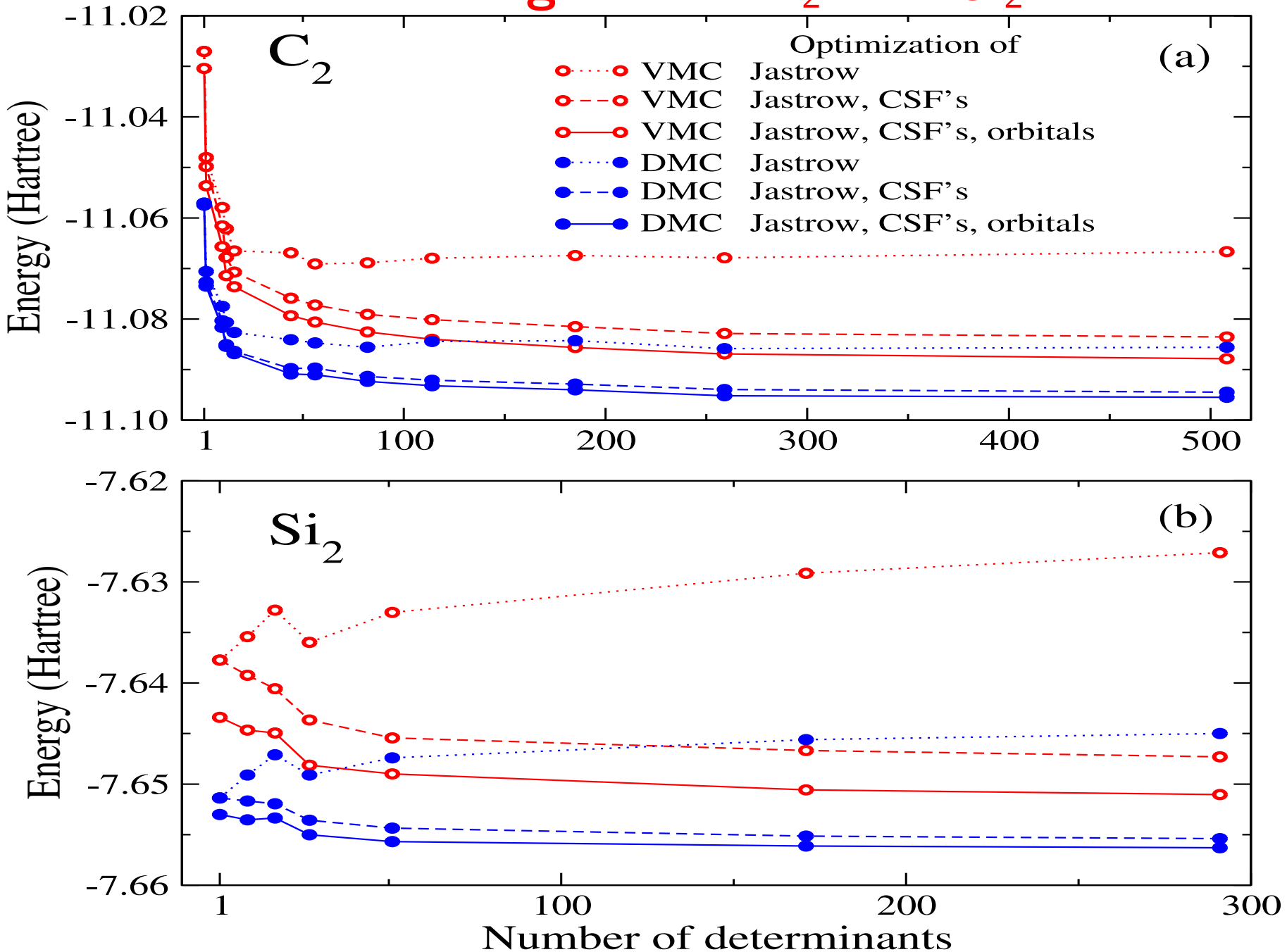
DMC (walk in electron coordinate space)	FCIQMC/SQMC (walk in determinant spac
Severe Fermion sign problem due to growth of Bosonic component relative to Fermionic.	Less severe Fermion sign problem due to opposite sign walkers being spawned on the same determinant
Fixed-node approximation needed for stable 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
Energy is variational	Energy not variational but DM variant is
Need to use pseudopotentials for large Z.	Can easily do frozen-core

Applications

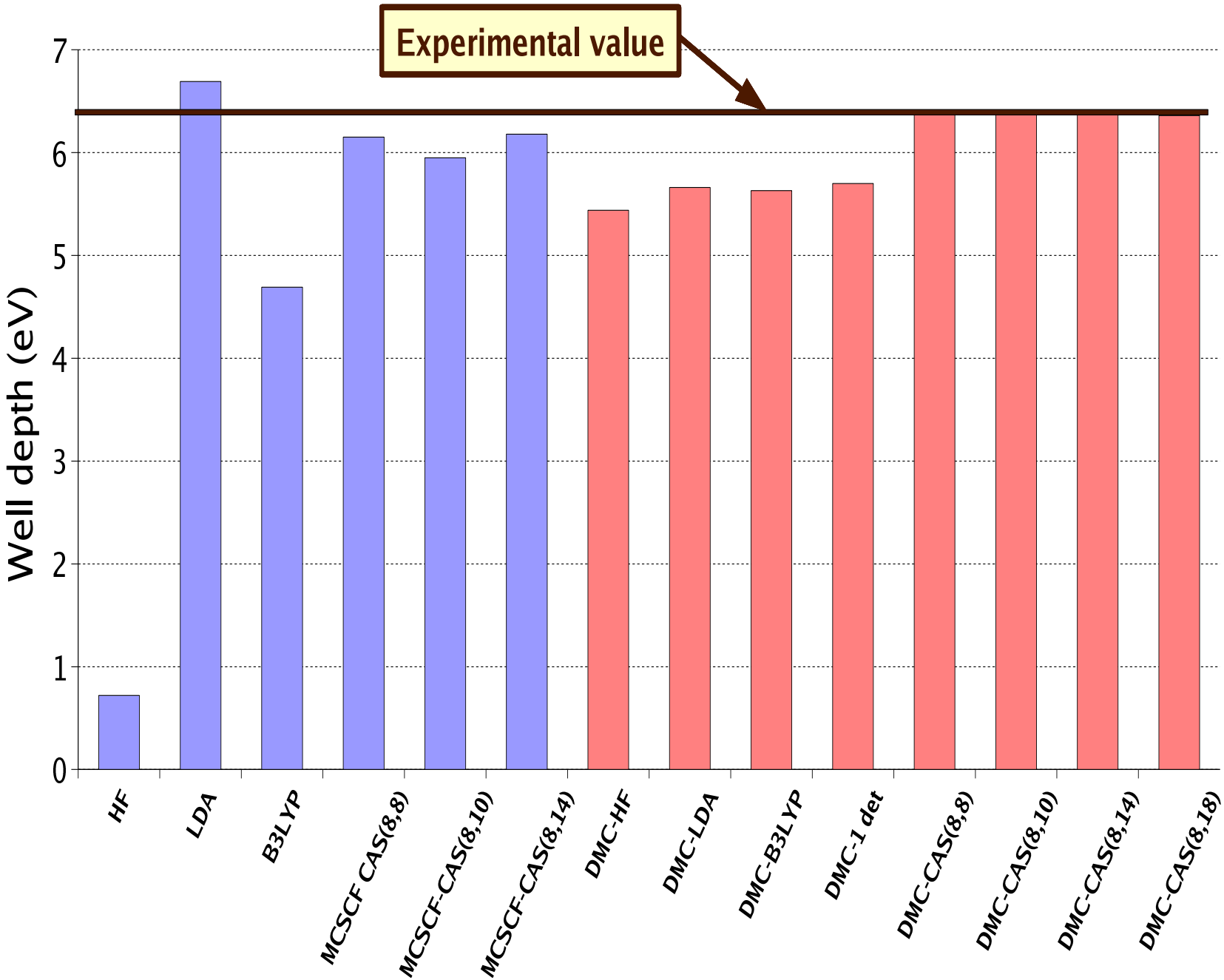
Parallel Efficiency of DMC



Convergence of C_2 and Si_2

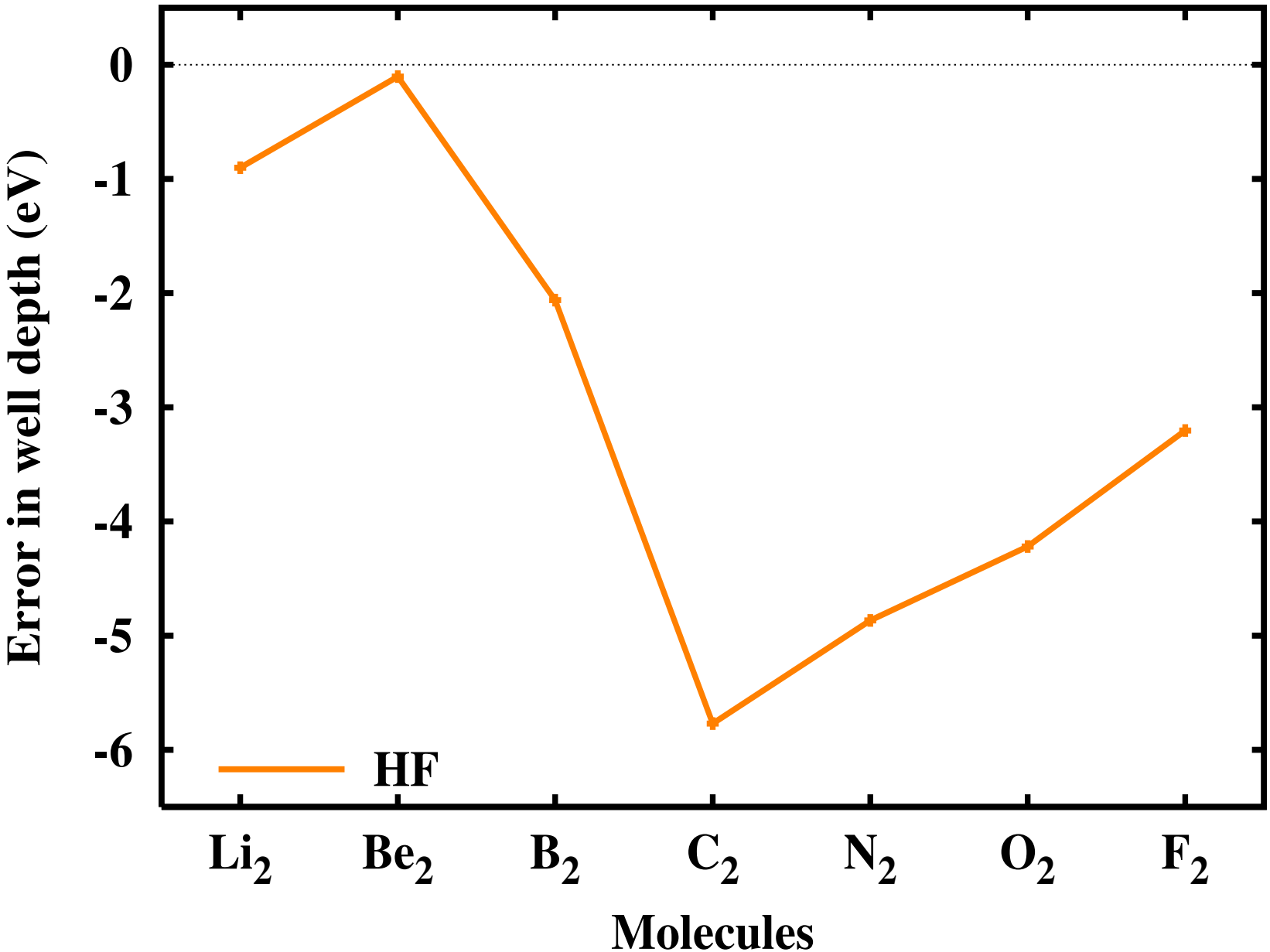


Well-depth of C_2



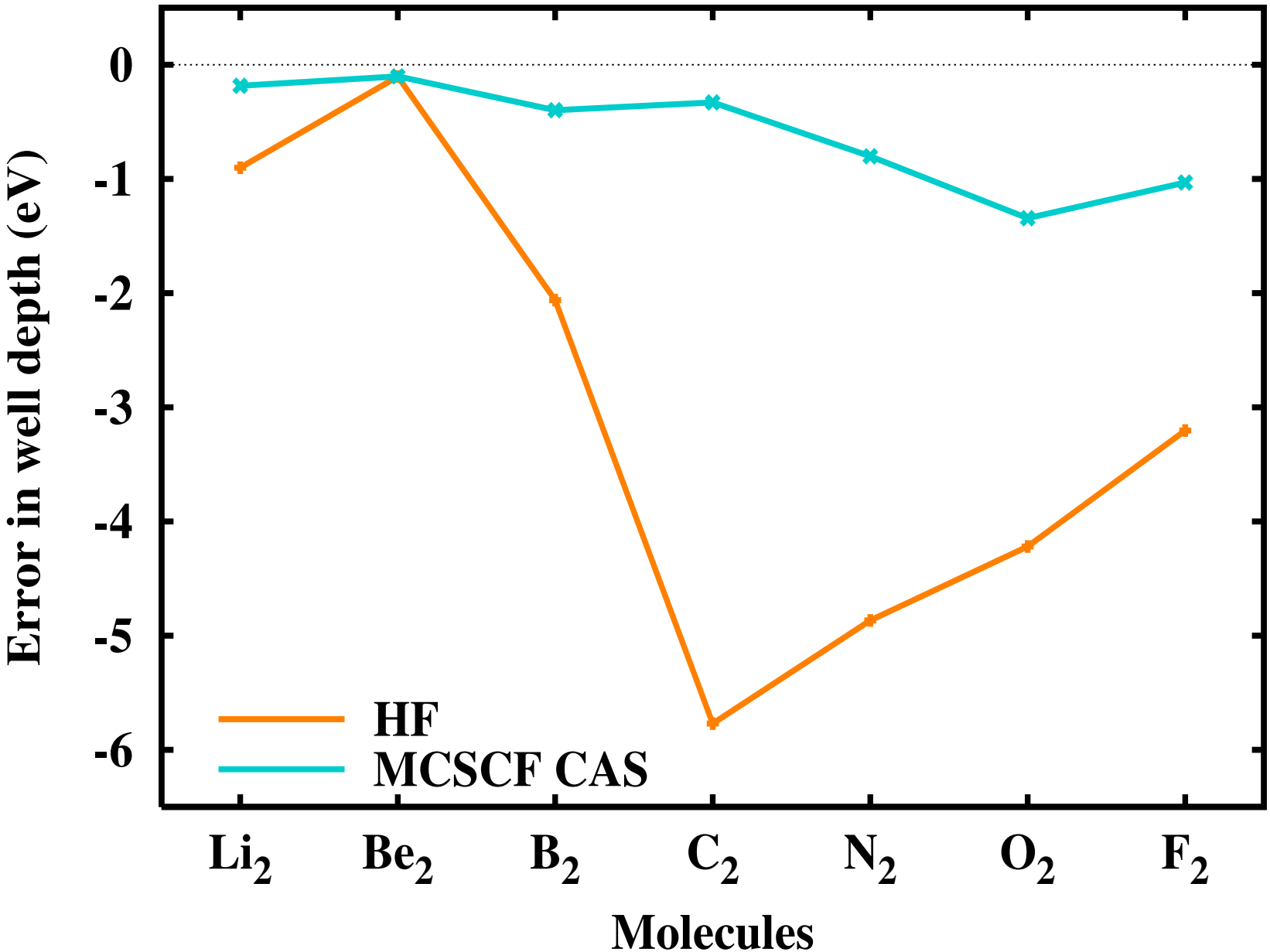
Error in Well-Depth of 1^{st}-Row Diatomic Molecules

Julien Toulouse and CJU, J. Chem. Phys. (2008)



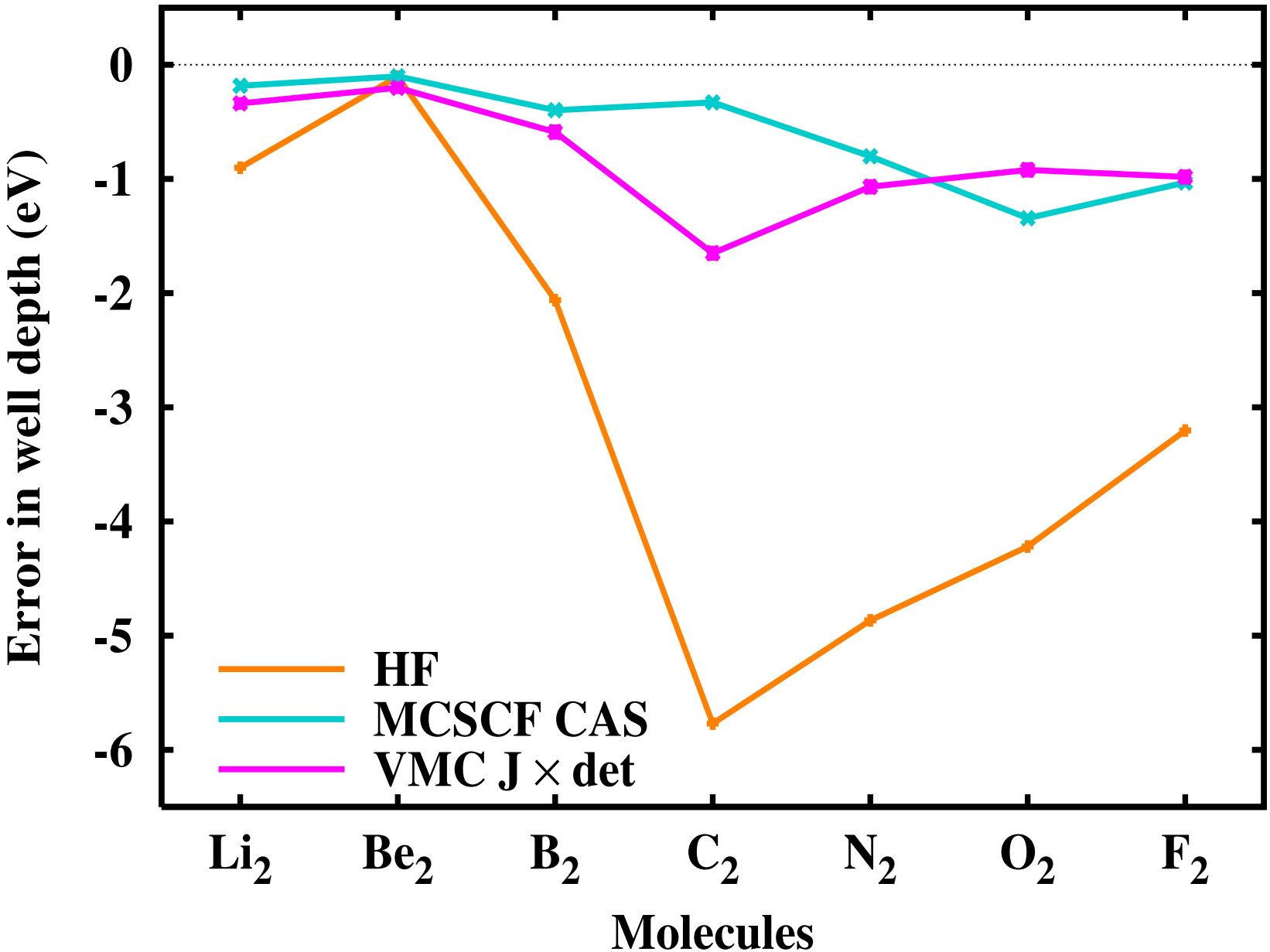
Error in Well-Depth of 1^{st}-Row Diatomic Molecules

Julien Toulouse and CJU, J. Chem. Phys. (2008)



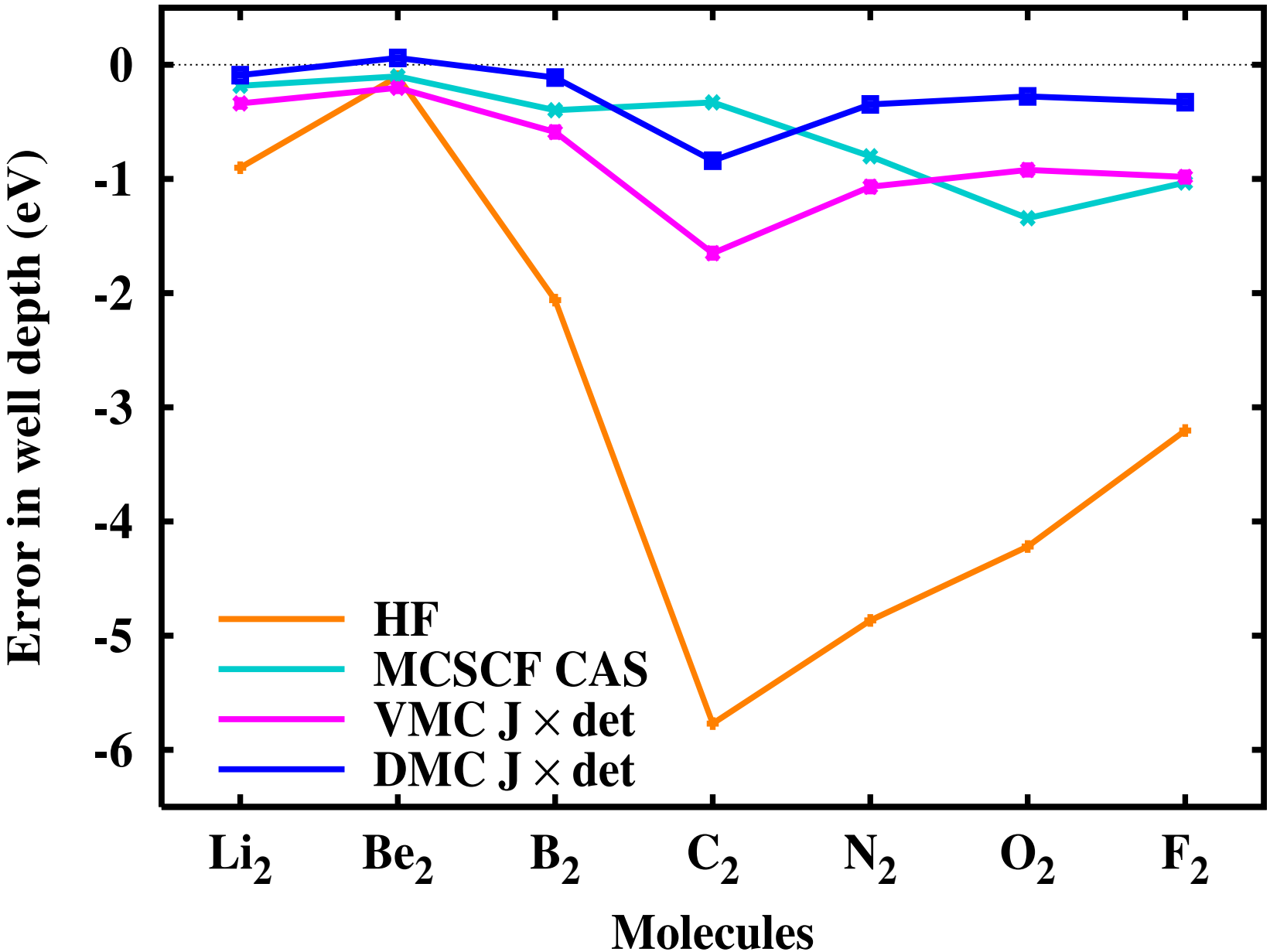
Error in Well-Depth of 1^{st}-Row Diatomic Molecules

Julien Toulouse and CJU, J. Chem. Phys. (2008)



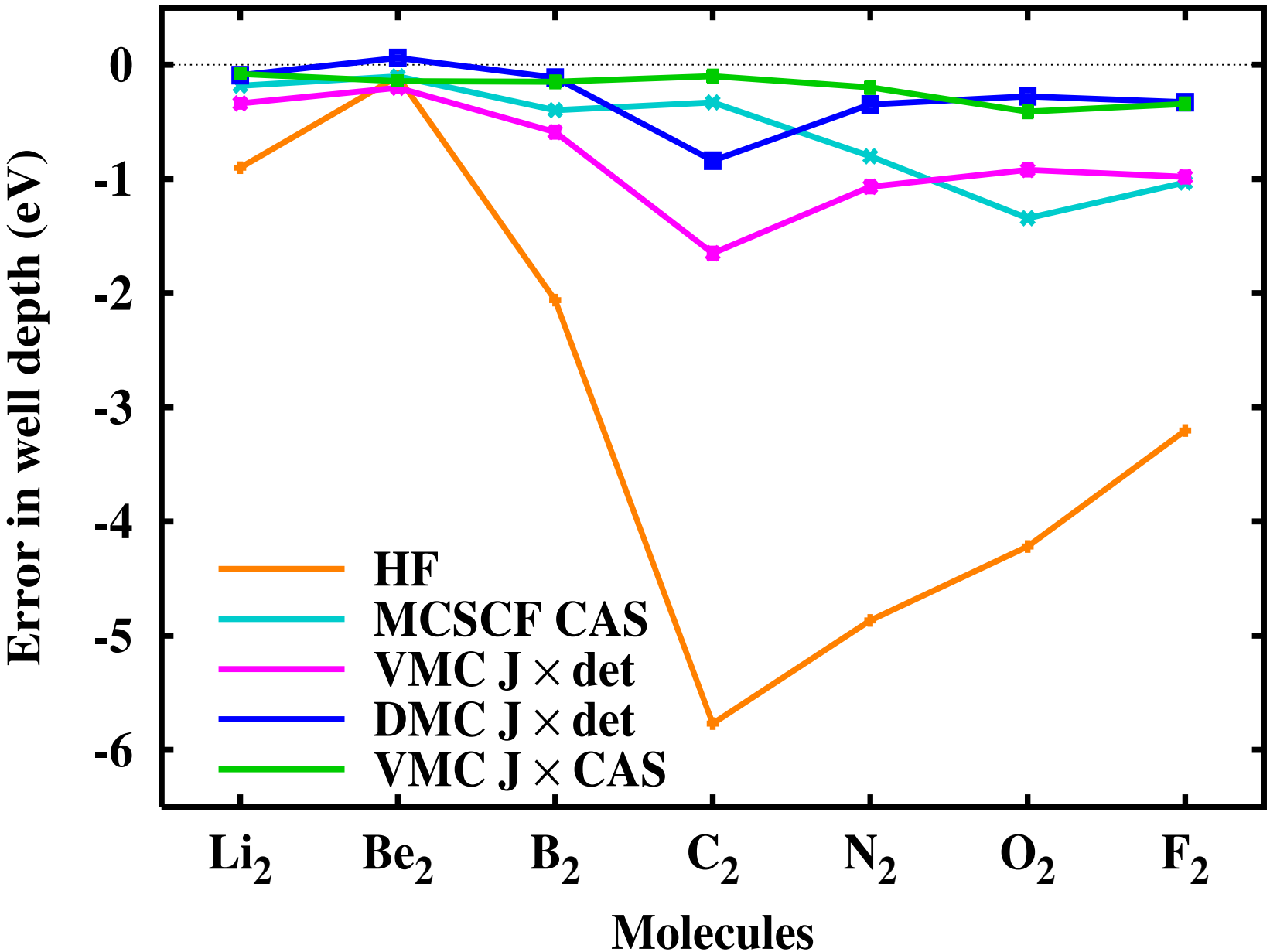
Error in Well-Depth of 1^{st}-Row Diatomic Molecules

Julien Toulouse and CJU, J. Chem. Phys. (2008)



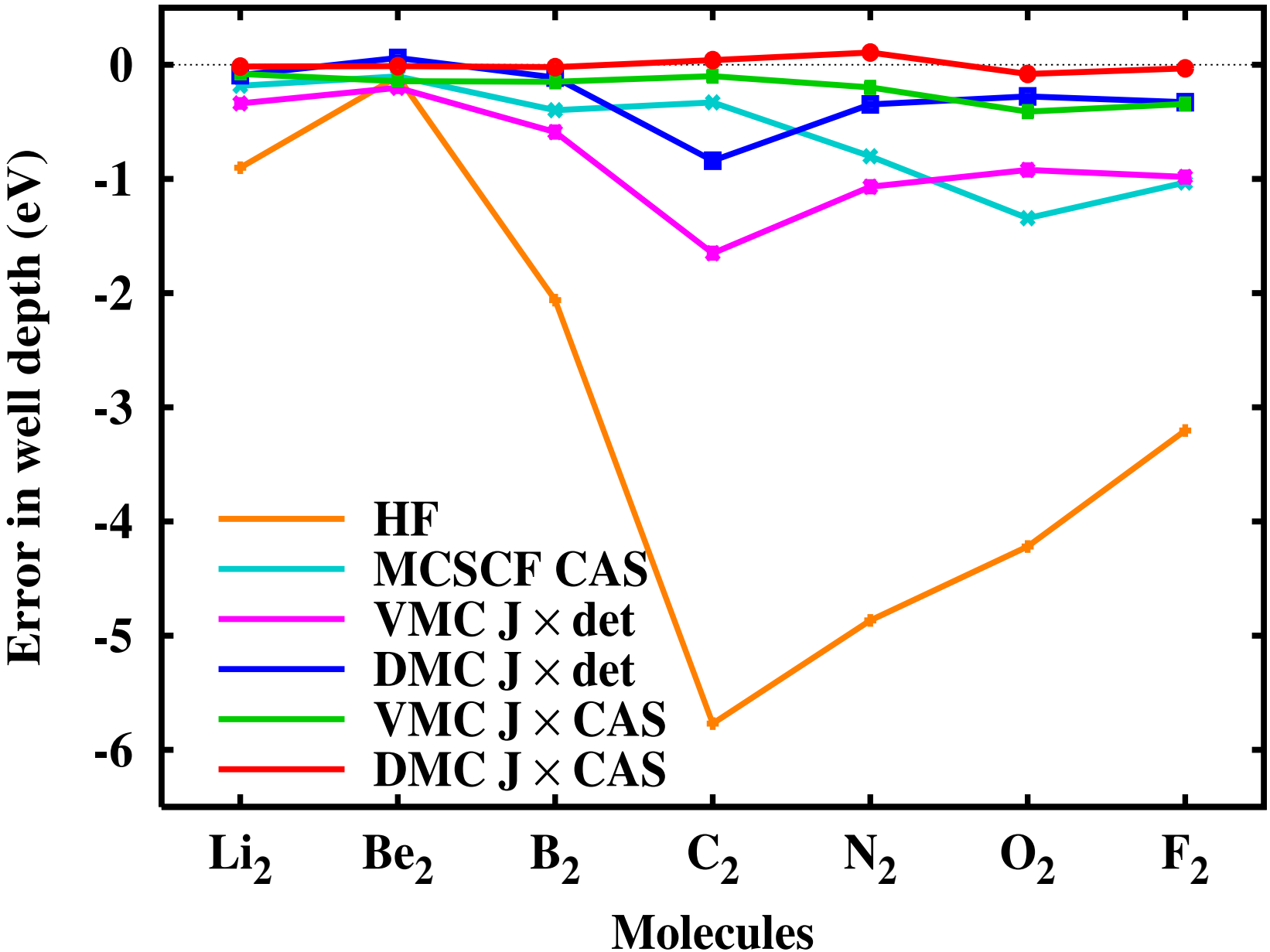
Error in Well-Depth of 1^{st}-Row Diatomic Molecules

Julien Toulouse and CJU, J. Chem. Phys. (2008)



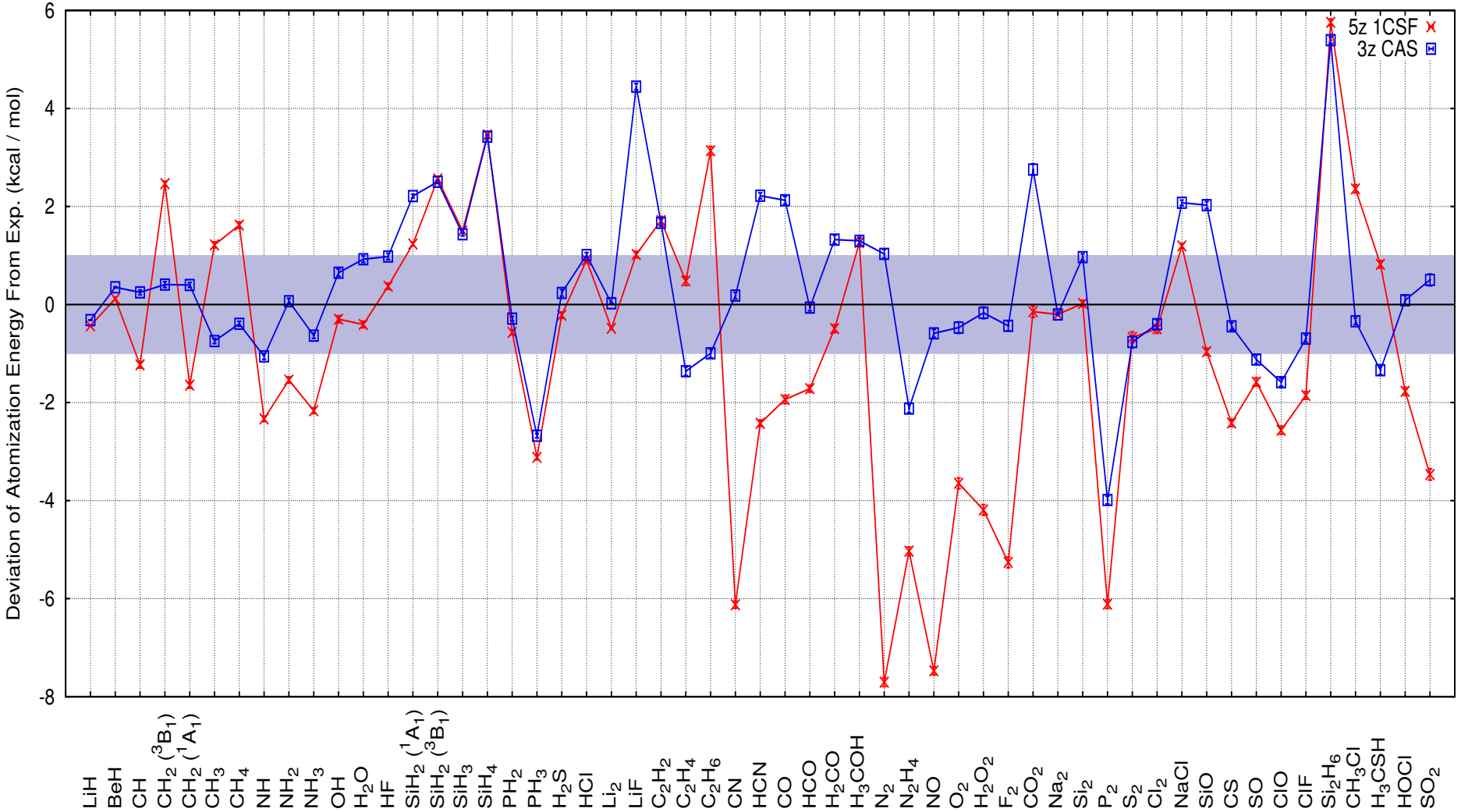
Error in Well-Depth of 1^{st}-Row Diatomic Molecules

Julien Toulouse and CJU, J. Chem. Phys. (2008)



Atomization energies of the G2 set

F.R. Petruzielo, Julien Toulouse and CJU, J. Chem. Phys. (2012)



The mean absolute deviation from experiment for the DMC energies using the CAS trial wave functions is 1.2 kcal/mole.

Excited states

Previous work: Ceperley & Bernu; Nightingale; Filippi

Excited states that are the lowest state of that symmetry are no harder than ground states.

True excited states are harder to compute. Options for computing true excited states:

1. Take determinantal part of wavefunction from another method, e.g, MCSCF, multiply it by Jastrow factor and rely on fixed-node constraint to prevent collapse to ground state. Not very satisfactory.
2. Do state-averaged optimization of ground and excited states, alternating between calculating the linear coefficients of the states and optimizing the nonlinear (orbital, basis exponent and Jastrow) coefficients. Guarantees upper bound. (M. P. Nightingale et al. for bosonic clusters; Claudia Filippi et al. for molecules)
3. Do a state-specific optimization of each state separately. Sometimes has root-flipping problems but sometimes it works.

Excited states of methylene (CH_2)

Zimmerman, Toulouse, Zhang, Musgrave, CJU, (submitted to JCP)

see talk COMP 0344 for more details

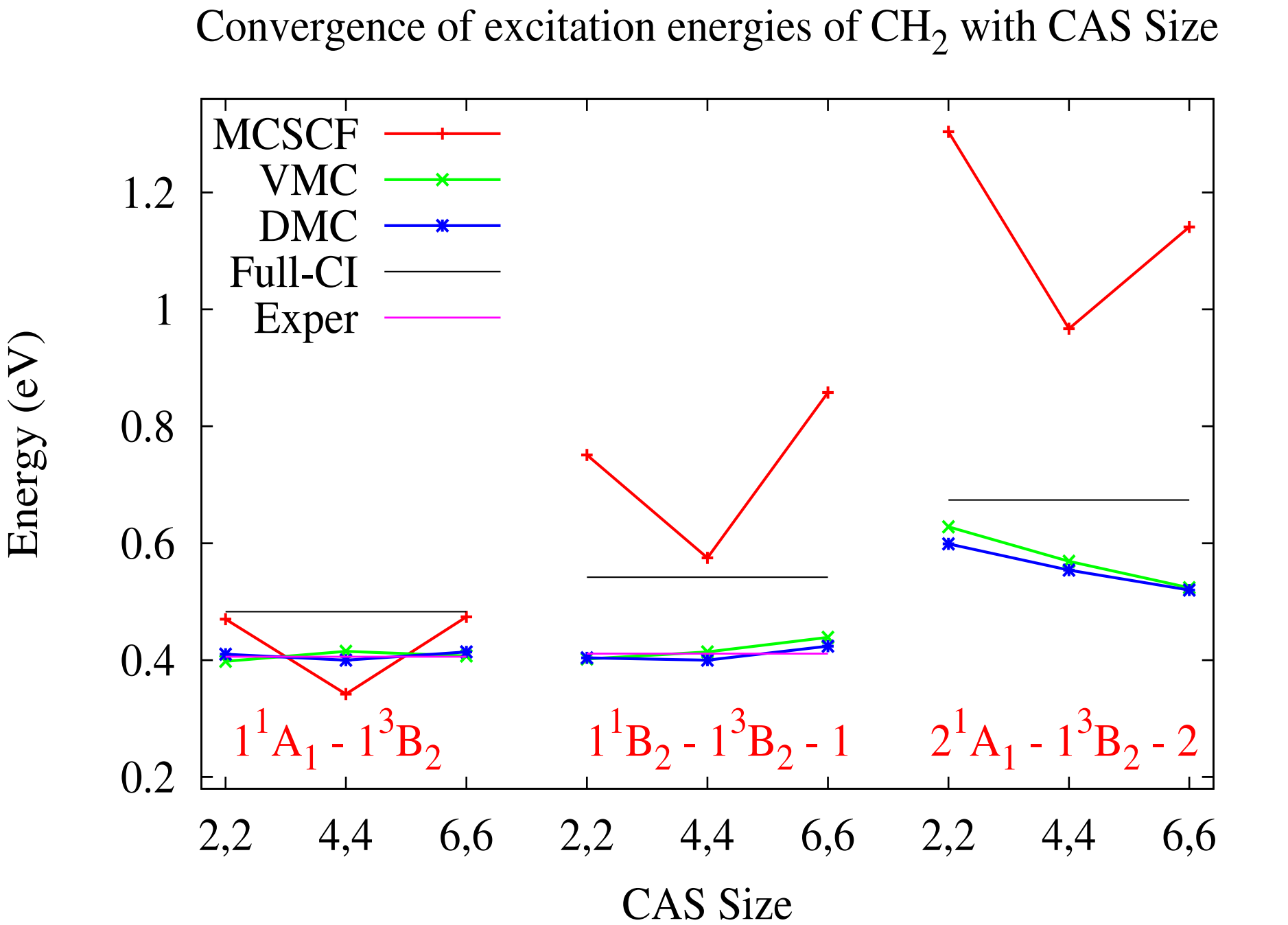
States of methylene (CH_2) are:

$1\ ^3B_2$, ground state, single reference

$1\ ^1A_1$, 1^{st} excited state, multi reference

$1\ ^1B_2$, 2^{nd} excited state, single reference

$2\ ^1A_1$, 3^{rd} excited state, multi reference, true excited state

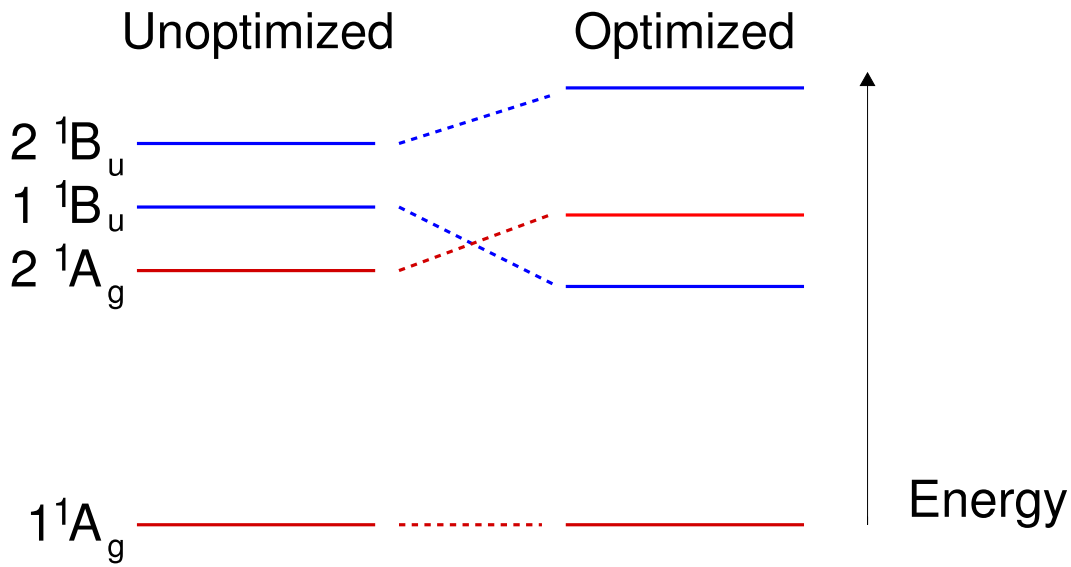

Convergence of excitation energies of CH_2 with CAS Size
MCSCF
VMC
DMC
Full-CI
Exper
Energy (eV)
1.2
1
0.8
0.6
0.4
0.2
$1^1A_1 - 1^3B_2$
$1^1B_2 - 1^3B_2 - 1$
$2^1A_1 - 1^3B_2 - 2$
2,2
4,4
6,6
2,2
4,4
6,6
2,2
4,4
6,6
CAS Size

EFP method for ground and excited states

F. Schautz and C. Filippi, JCP **120**, 10931 (2004)

- Excitations of ethene C_2H_4 $\rightarrow$ Up to $\boxed{858}$ optimized parameters

DMC excitation energies



State	Unoptimized	Optimized	expt.	(eV)
1^1B_{1u}	8.45(2)	7.93(2)	$>$7.7	
2^1A_g	7.96(2)	8.36(2)	8.29	
2^1B_{1u}	9.05(2)	9.37(2)	9.33	

Localization in planar quantum dots

Devrim Güçlü, Amit Ghosal, CJU, Harold Baranger

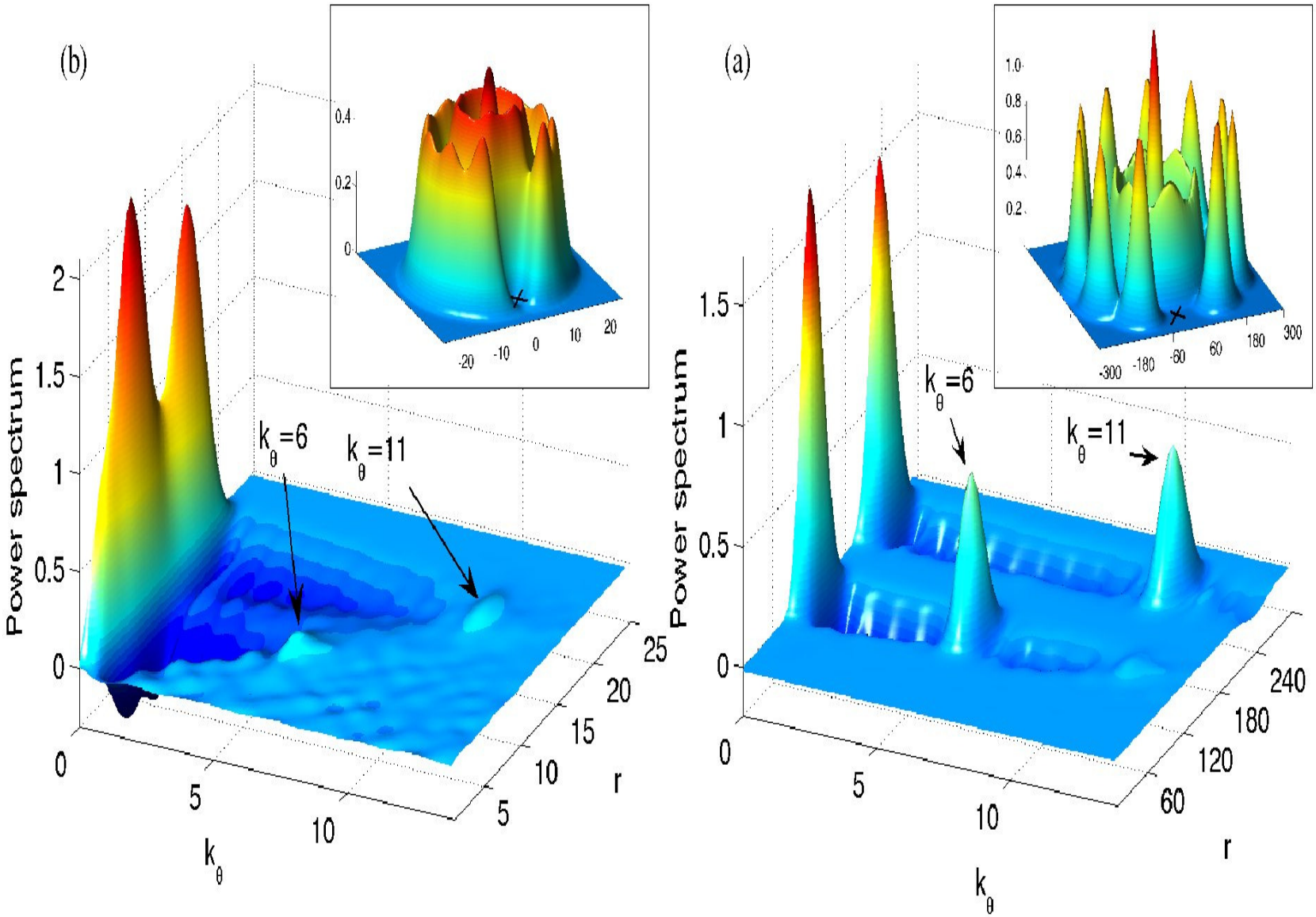
Signatures of localization

1. Pair densities.
2. Addition energy spectrum evolving from noninterating limit to classical limit.

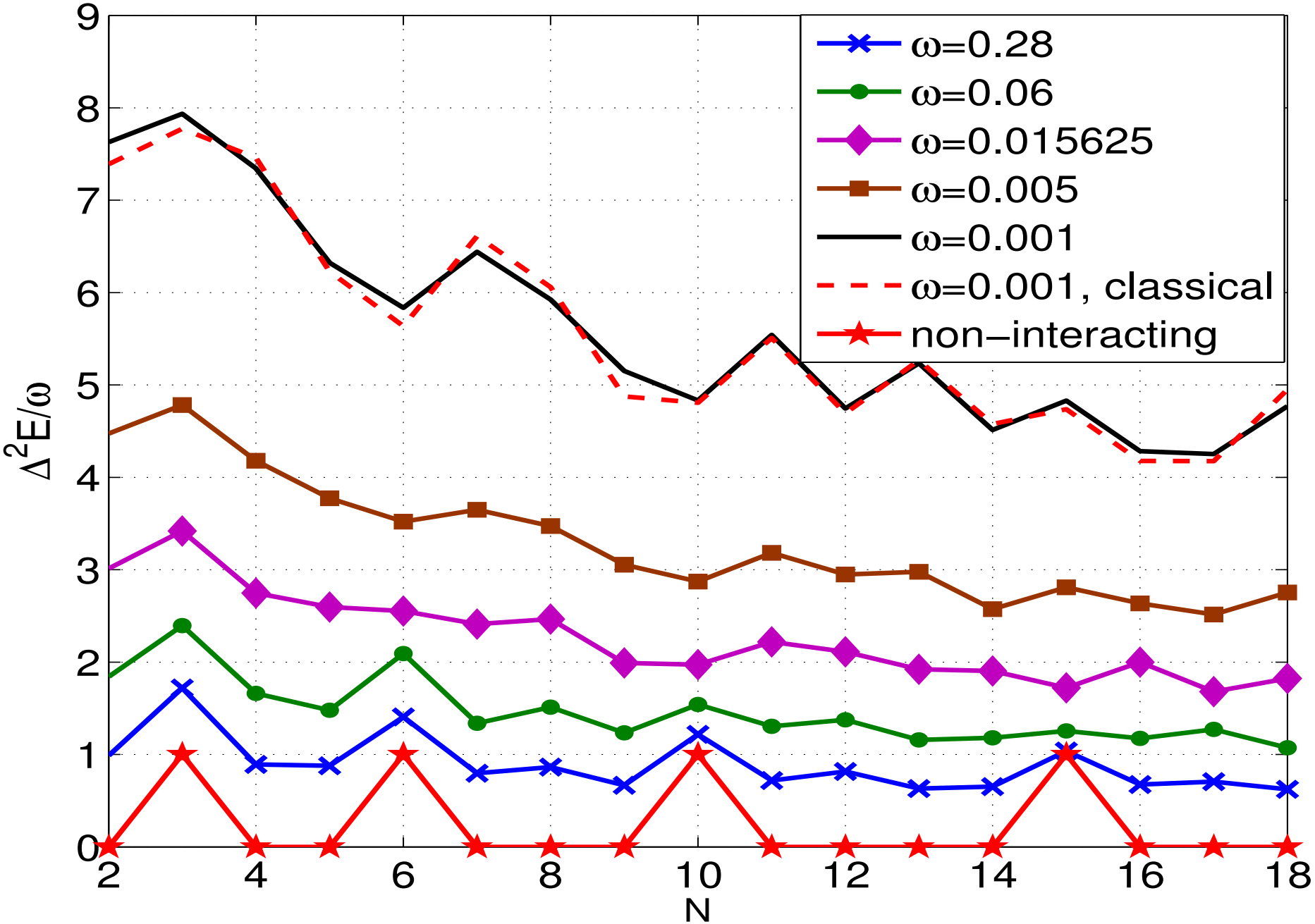
Pair densities and power spectrum of $N = 18$ dot

$r_s = 4.8$

$r_s = 52$



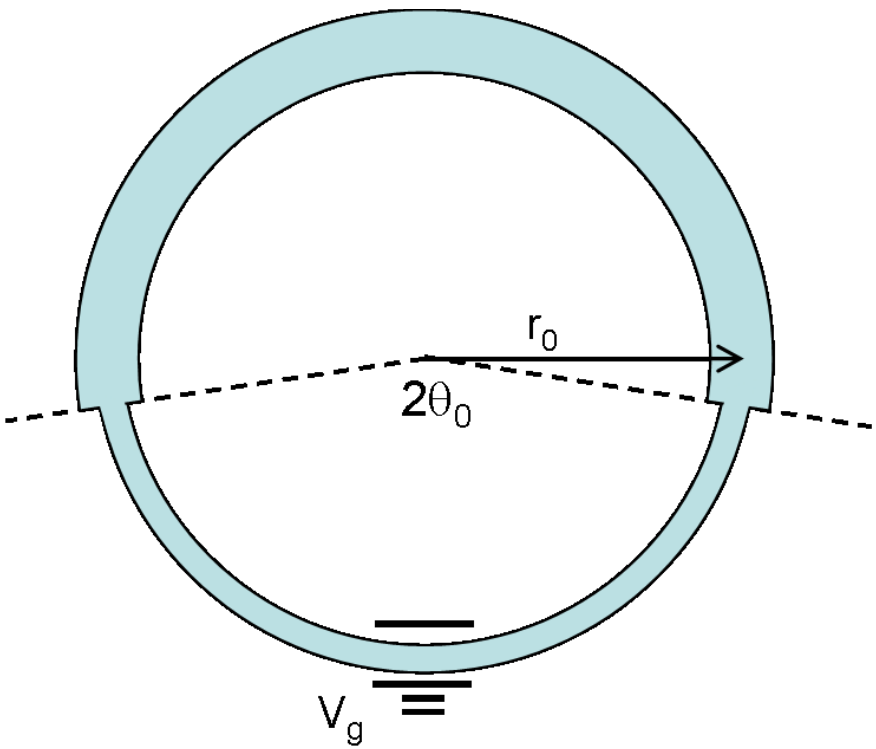
$\Delta^2 E/\omega$ for different ω or r_s (strong in-plane B)

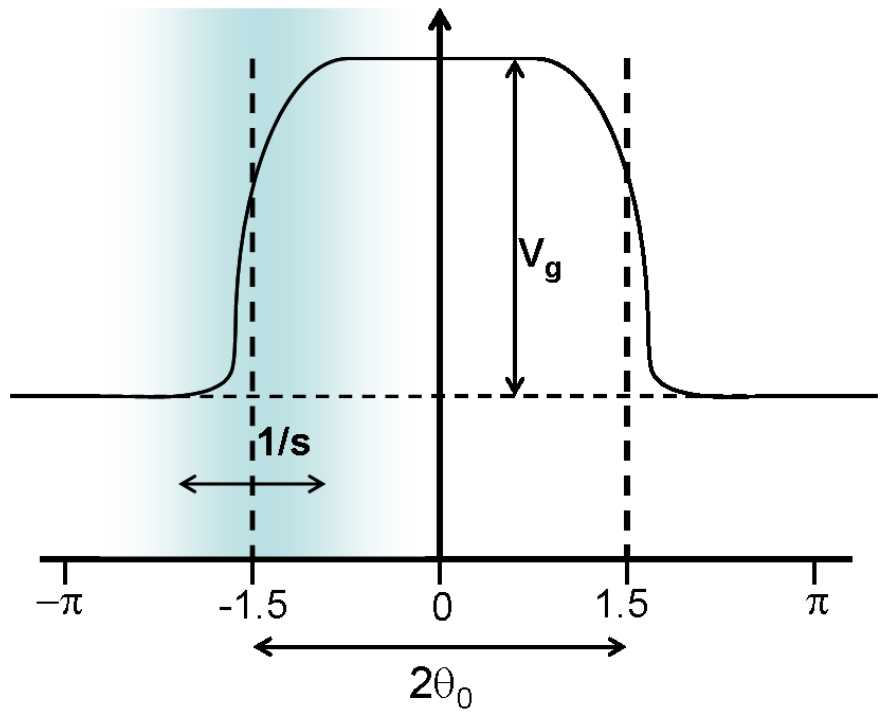


Localization in inhomogeneous quantum wires

Güçlü, Jiang, CJU, Baranger

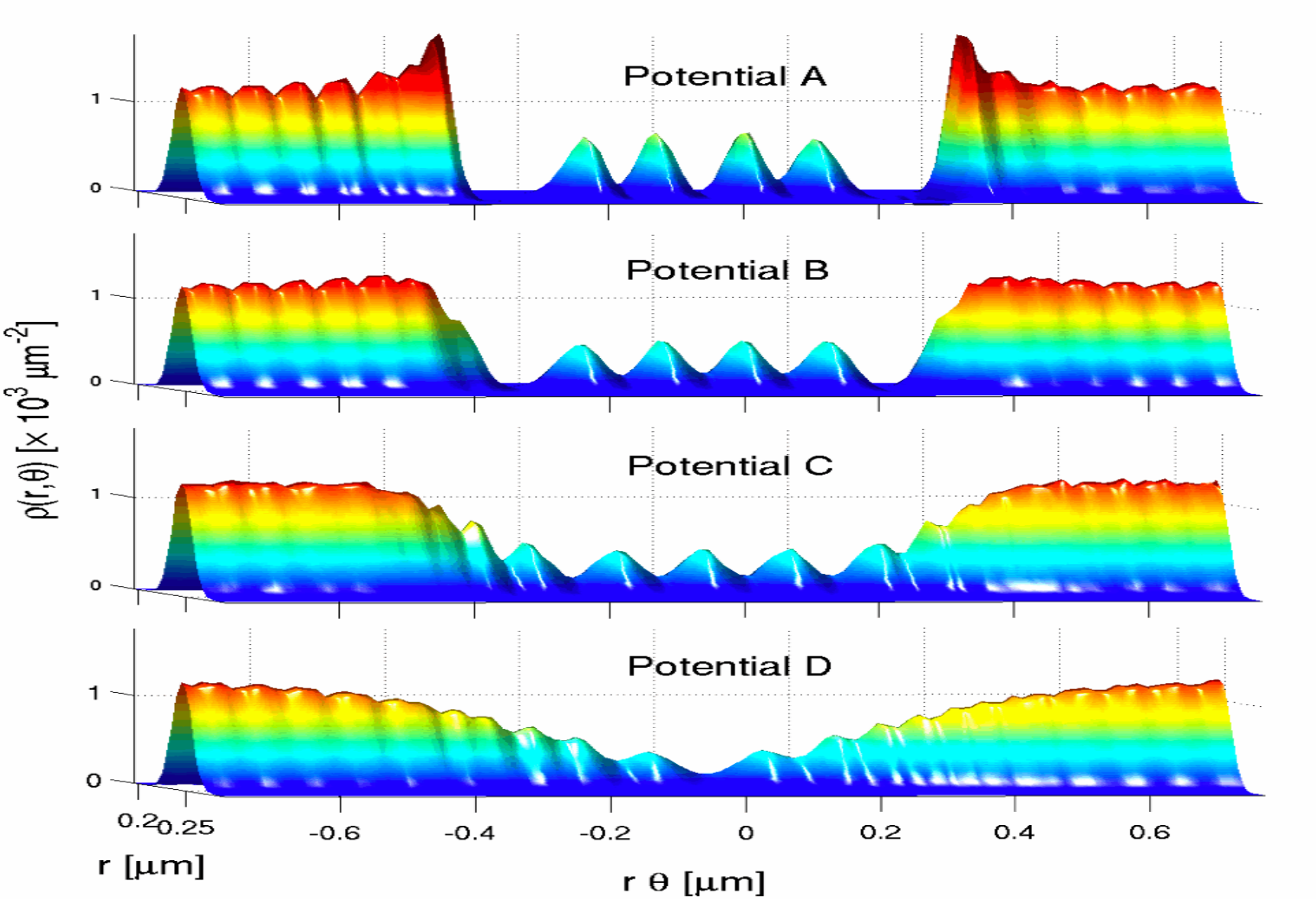
$$H = -\frac{1}{2}\sum_i^N \nabla_i^2 + \frac{1}{2}\sum_i^N \omega^2 (r_i - r_0)^2 + \sum_{i<j}^N \frac{1}{r_{ij}}$$
$$+ V_g \left\{ \tanh\left[s(\theta_i + \theta_0)\right] - \tanh\left[s(\theta_i - \theta_0)\right] \right\}$$





Localization in inhomogeneous quantum wires

Güçlü, Jiang, CJU, Baranger



Quantum Zigzag Phase Transition in Quantum Wires

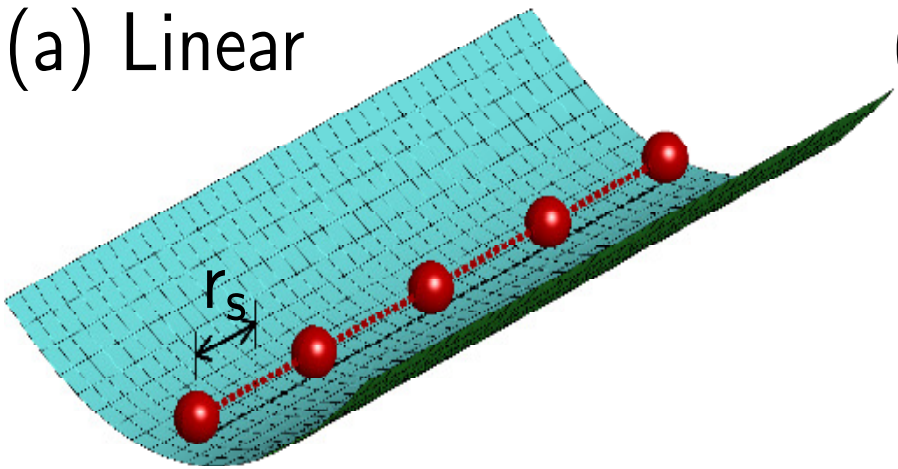
Abhijit Mehta, CJU, Julia Meyer, Harold Baranger

Consider a 2-d wire, along the x direction with a finite width along the y direction.

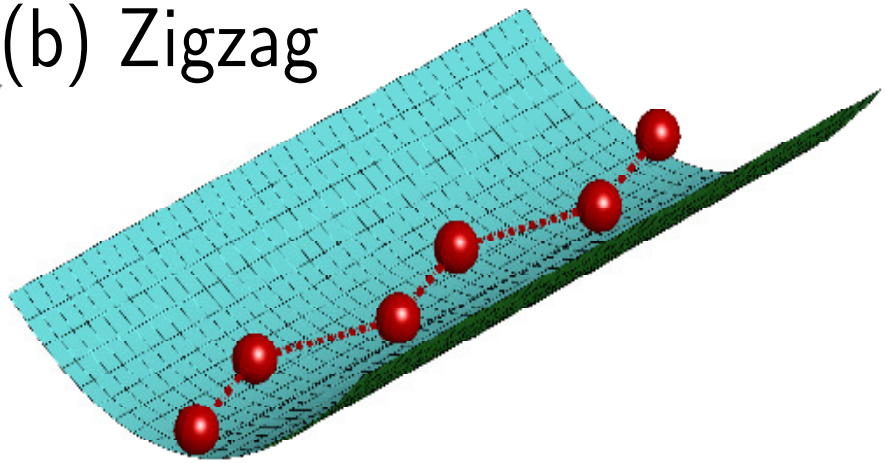
$$H = -\frac{1}{2}\sum_{i=1}^{N}\nabla_i^2 + \frac{1}{2}\sum_{i=1}^{N}\omega y_i^2 + \sum_{i<j\leq N}\frac{e^2}{\epsilon|\mathbf{r}_i - \mathbf{r}_j|} \tag{1}$$

At low densities electrons form linear Wigner crystal. Two length scales:
$r_s = 1/(2n)$, and,
r_0: confinement and Coulomb energies are equal $(1/2)m\omega^2 r_0^2 = e^2/(\epsilon r_0)$.
As density n is raised, expect a transition to a zigzag phase when $r_s \approx r_0$ before transition to liquid phase.

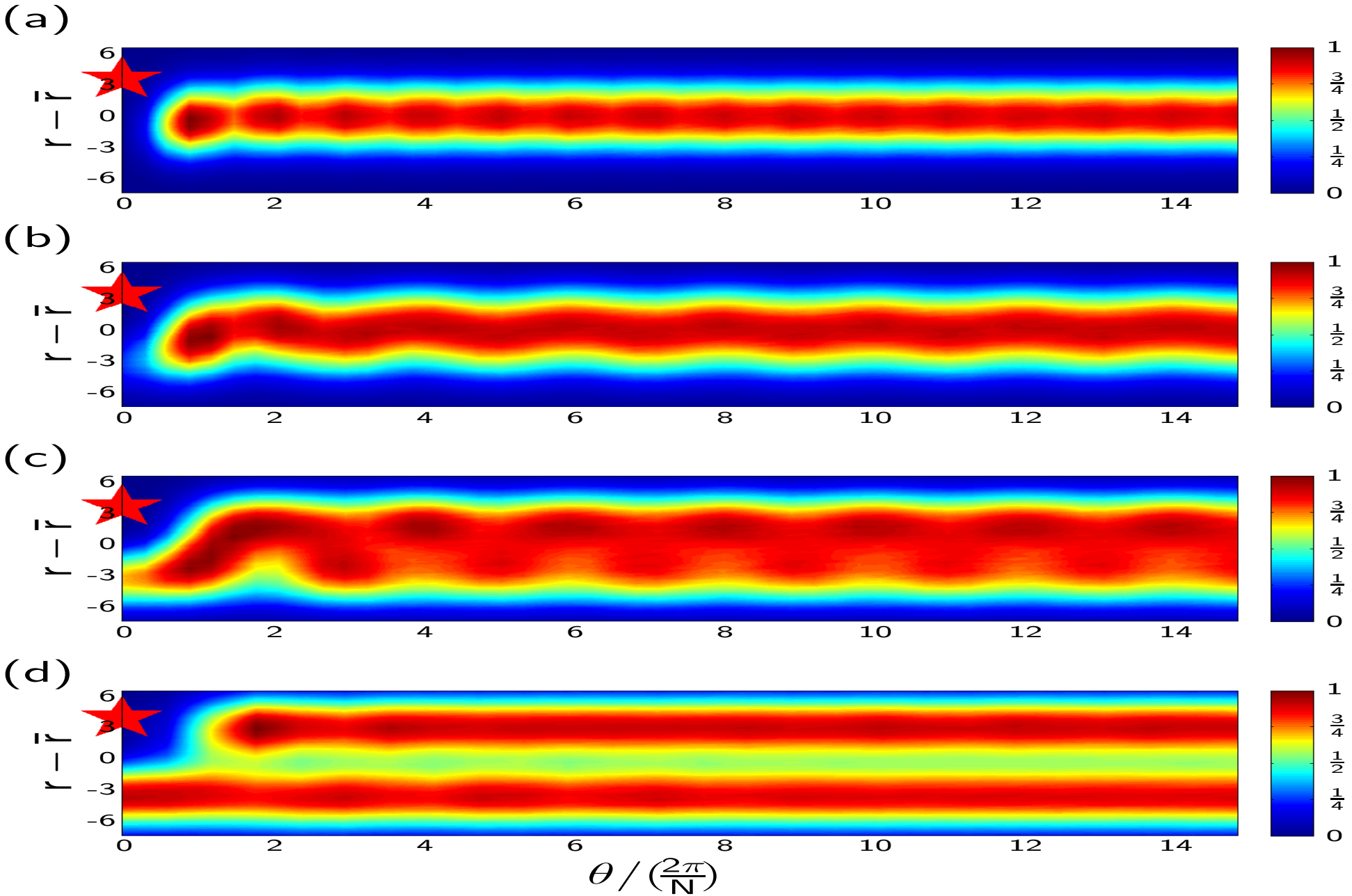
(a) Linear



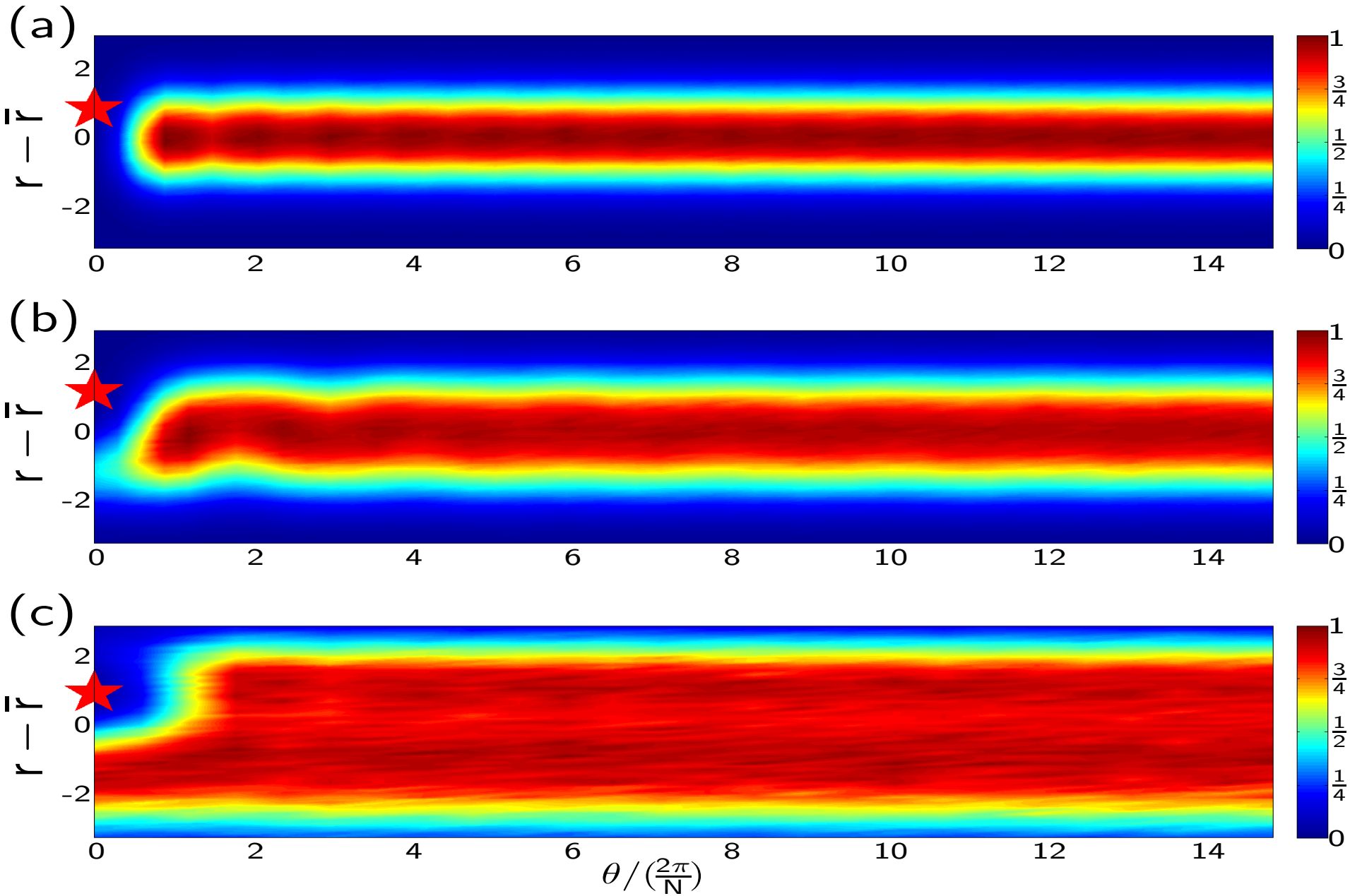
(b) Zigzag



Pair densities at $\omega = 0.1$



Pair densities at $\omega = 0.6$



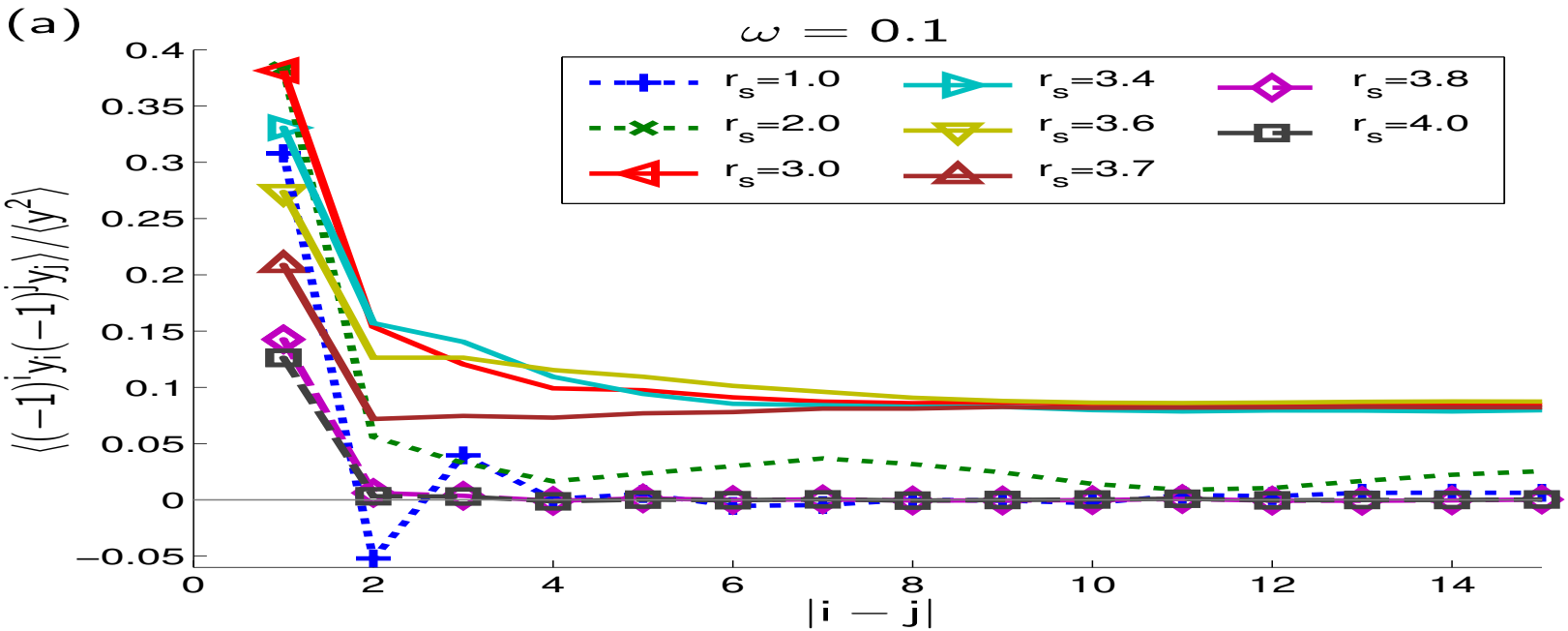
Zigzag Correlation Function

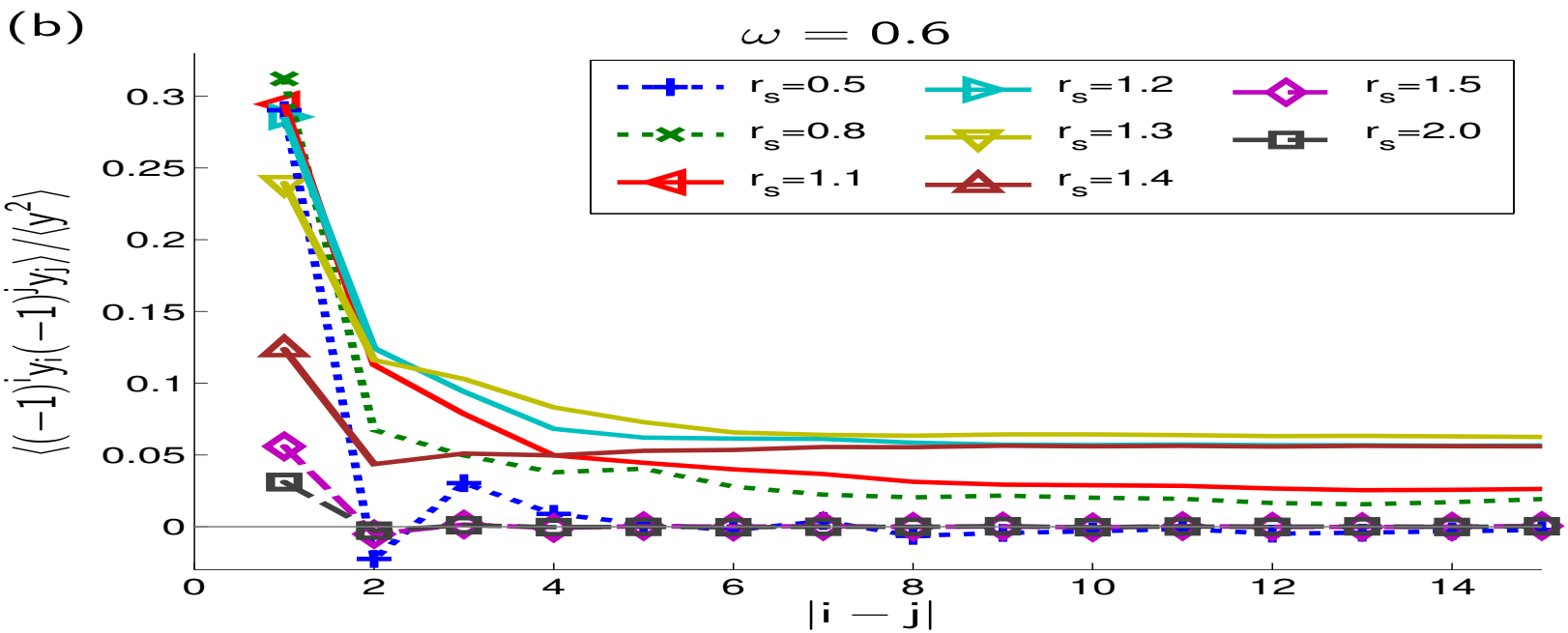
$$C_{ZZ}(|i-j|) \;=\; \left\langle (-1)^i \left(y_i - \langle y \rangle\right) (-1)^j \left(y_j - \langle y \rangle\right) \right\rangle$$

Order electrons along the length of the wire.

Zigzag order is tied to the ordering of the electrons, not their position along the wire.

Zigzag Correlation Function $\omega = 0.1,\ 0.6$





Zigzag Order Parameter

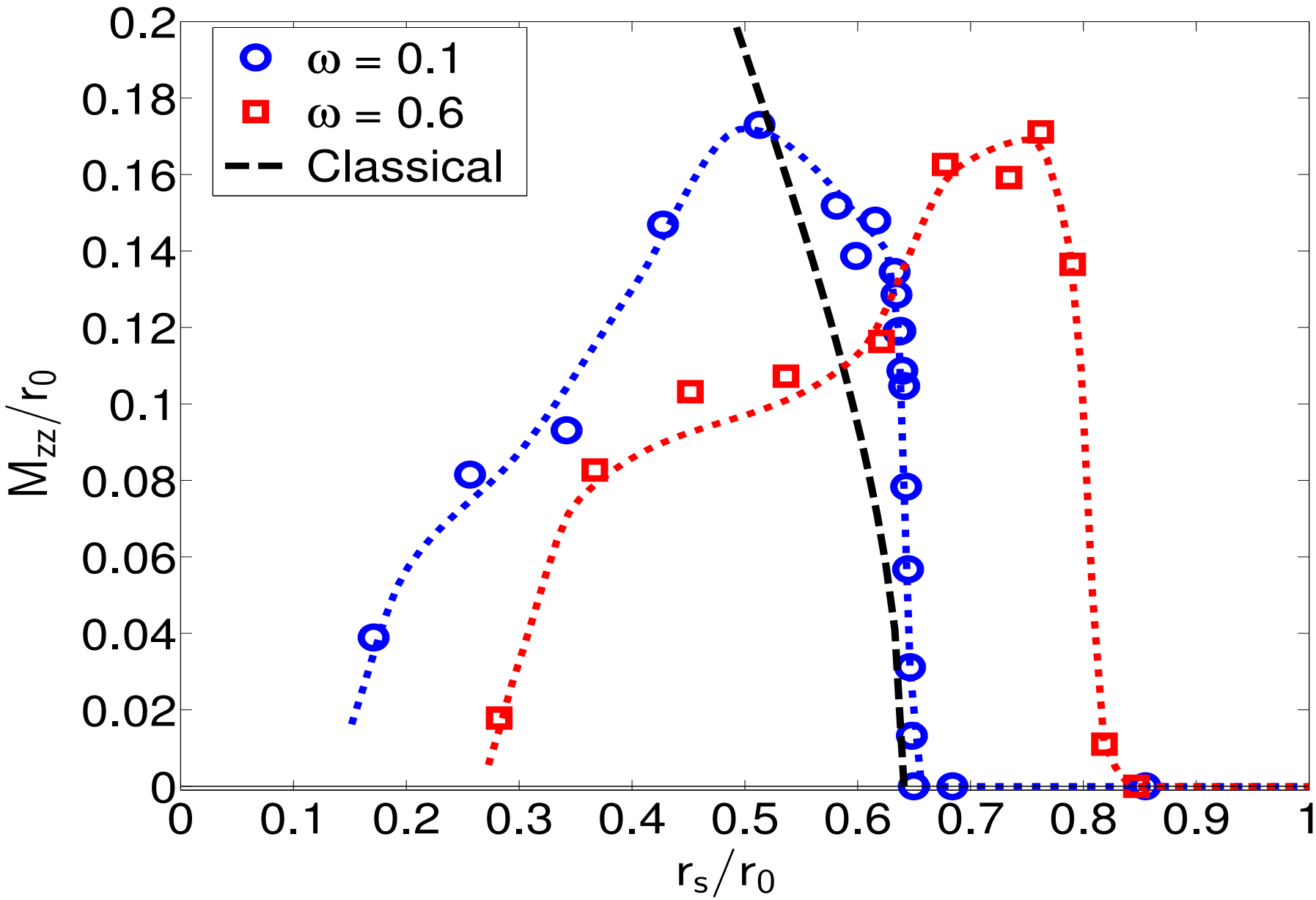
$$C_{ZZ}(|i-j|) \;=\; \left\langle (-1)^i \left(y_i - \langle y \rangle\right) (-1)^j \left(y_j - \langle y \rangle\right) \right\rangle$$

Zigzag order parameter, M_{ZZ}. M_{ZZ}^2, is average of zigzag correlation function, $C_{ZZ}(|i-j|)$, for electrons far from the fixed reference electron.

$$M_{ZZ}^2 \;=\; \langle C_{ZZ}(|i-j|) \rangle_{|i-j|>N/4}$$

Zigzag Order Parameter

$$M_{ZZ}^2 = \langle C_{ZZ}(|i-j|)\rangle_{|i-j|>N/4}$$



Collaborators

Julien Toulouse, Paris
Claudia Filippi, Leiden
Frank Petruzielo, Physics, Cornell
Hitesh Changlani, Physics, Cornell
Adam Holmes, Physics, Cornell
M. Peter Nightingale, Rhode Island
Roland Assaraf, Paris
Sandro Sorella, Trieste
Harold Baranger, Duke
Devrim Güçlü, NRC
Amit Ghosal, Kolkata
Hong Jiang, China
Paul Zimmerman, Stanford
Zhiyong Zhang, Stanford