

Variational and Projector Quantum Monte Carlo Methods in Chemistry and Physics

Cyrus Umrigar

Physics Department, Cornell University, Ithaca.

Email: CyrusUmrigar@cornell.edu

Current Topics in Theoretical Chemistry, Trujillo, Peru, 24-30 Sep 2016

Outline

1. Comparison of Electronic Structure Methods
2. Intro to Monte Carlo Methods
3. The essence of variational and projector quantum MC methods
4. Variational MC (VMC)
5. Projector MC (PMC) methods
 - 1 Full configuration interaction QMC (FCIQMC)
 - 2 Diffusion Monte Carlo (DMC)
6. Sign Problem in various Projector Monte Carlo methods (PMC)
7. Semistochastic Quantum Monte Carlo (SQMC) (extension of FCIQMC)
8. Pros and cons of various Projector Monte Carlo methods in orbital space and real space

Electronic Structure Methods

Quantum Chemistry Methods ~ 200 developers and 2000 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 ~ 400 developers and 10000 users.

Exact in principle, but in practice one uses approximate exchange-correlation functionals and they are not systematically improvable. Scales as $\mathcal{O}(N^3)$.

Quantum Monte Carlo ~ 20 developers and 100 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 well-optimized trial wavefunctions are used. Some observables easier to compute than others. Requires ingenuity to have statistical errors small.

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.

In QMC methods, frequently there is a tradeoff between statistical error and systematic error and one needs to find the best compromise.

In particular, Fermions are antisocial and have a **Fermion Sign Problem**.

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

Central Limit Theorem

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

Let $X_1, X_2, X_3, \dots, X_N$ 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.**

Monte Carlo versus Deterministic Integration methods

Deterministic Integration Methods:

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

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

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

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

Monte Carlo:

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

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

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

Monte Carlo integration

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

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

Quantum Monte Carlo Methods

What is Variational Quantum Monte Carlo?

Monte Carlo evaluation of the many-dimensional integrals needed for computing quantum mechanical expectation values for a trial wavefunction Ψ_T .

Ψ_T may contain many variational parameters which need to be optimized.

The optimized Ψ_T will also be used in Projector Quantum Monte Carlo.

What is Projector 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 or Davidson. 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

Exact wavefunction $|\Psi_0\rangle = \sum_i e_i |\phi_i\rangle$, where, $e_i = \langle \phi_i | \Psi_0 \rangle$

Trial wavefunction $|\Psi_T\rangle = \sum_i t_i |\phi_i\rangle$, where, $t_i = \langle \phi_i | \Psi_T \rangle$

(If basis incomplete then “exact” means “exact in that basis”.)

Variational MC

$$\begin{aligned} E_V &= \frac{\langle \Psi_T | \hat{H} | \Psi_T \rangle}{\langle \Psi_T | \Psi_T \rangle} = \frac{\sum_{ij}^{N_{\text{st}}} \langle \Psi_T | \phi_i \rangle \langle \phi_i | \hat{H} | \phi_j \rangle \langle \phi_j | \Psi_T \rangle}{\sum_i^{N_{\text{st}}} \langle \Psi_T | \phi_k \rangle \langle \phi_k | \Psi_T \rangle} \\ &= \frac{\sum_{ij}^{N_{\text{st}}} t_i H_{ij} t_j}{\sum_k^{N_{\text{st}}} t_k^2} = \sum_i^{N_{\text{st}}} \frac{t_i^2}{\sum_k^{N_{\text{st}}} t_k^2} \frac{\sum_j^{N_{\text{st}}} H_{ij} t_j}{t_i} \\ &= \sum_i^{N_{\text{st}}} \frac{t_i^2}{\sum_k^{N_{\text{st}}} t_k^2} E_L(i) = \frac{\left[\sum_i^{N_{\text{MC}}} E_L(i) \right]_{\Psi_T^2}}{N_{\text{MC}}} \end{aligned}$$

Projector MC

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

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

$$\begin{aligned} E_0 &= \frac{\langle \Psi_0 | \hat{H} | \Psi_T \rangle}{\langle \Psi_0 | \Psi_T \rangle} = \frac{\sum_{ij}^{N_{\text{st}}} \langle \Psi_0 | \phi_i \rangle \langle \phi_i | \hat{H} | \phi_j \rangle \langle \phi_j | \Psi_T \rangle}{\sum_k^{N_{\text{st}}} \langle \Psi_0 | \phi_k \rangle \langle \phi_k | \Psi_T \rangle} \\ &= \frac{\sum_{ij}^{N_{\text{st}}} e_i H_{ij} t_j}{\sum_k^{N_{\text{st}}} e_k t_k} = \sum_i^{N_{\text{st}}} \frac{e_i t_i}{\sum_k^{N_{\text{st}}} e_k t_k} \frac{\sum_j^{N_{\text{st}}} H_{ij} t_j}{t_i} \\ &= \sum_i^{N_{\text{st}}} \frac{e_i t_i}{\sum_k^{N_{\text{st}}} e_k t_k} E_L(i) = \frac{\left[\sum_i^{N_{\text{MC}}} E_L(i) \right]_{\Psi_T \Psi_0}}{N_{\text{MC}}} \end{aligned}$$

Variational Monte Carlo in Real Space

$$\begin{aligned} E_V &= \frac{\int d\mathbf{R} \psi_T(\mathbf{R}) \mathcal{H} \psi_T(\mathbf{R})}{\int d\mathbf{R} \psi_T^2(\mathbf{R})} \\ &= \int d\mathbf{R} \frac{\psi_T^2(\mathbf{R})}{\int d\mathbf{R} \psi_T^2(\mathbf{R})} \frac{\mathcal{H} \psi_T(\mathbf{R})}{\psi_T(\mathbf{R})} \\ &= \frac{1}{N_{\text{MC}}} \sum_i^{N_{\text{MC}}} \frac{\mathcal{H} \Psi_T(\mathbf{R}_i)}{\Psi_T(\mathbf{R}_i)} = \frac{1}{N_{\text{MC}}} \sum_i^{N_{\text{MC}}} E_L(\mathbf{R}_i) \end{aligned}$$

Energy is obtained as an arithmetic sum of the *local energies* $E_L(\mathbf{R}_i)$ evaluated for configurations sampled from $\psi_T^2(\mathbf{R})$ using a generalization of the Metropolis method. If ψ_T is an eigenfunction, the $E_L(\mathbf{R}_i)$ do not fluctuate. Accuracy of VMC depends crucially on the quality of $\psi_T(\mathbf{R})$.

Diffusion MC does better by projecting onto ground state.

Three ingredients for accurate Variational Monte Carlo

1. A method for sampling the square of an arbitrary wave function with small auto-correlation time. [Metropolis-Hastings](#).
2. A functional form for the trial wavefunction 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, Rosenbluth², Teller², 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** probability density.

Samples are sequentially correlated, but, freedom of choice of the proposal probabilities allows one to sample efficiently with very small autocorrelation time.

Functional form of Trial Wave Function

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

- **Determinants:** $\sum_n d_n D_n^\uparrow D_n^\downarrow$

D^\uparrow and 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:

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

- **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 $\sim N_{\text{atomtype}}$ of \mathcal{J} parms.
 - $B_{ij} \Rightarrow$ electron-electron correlation $\sim N_{\text{atomtype}}$ of $\zeta_{k\alpha}$ parms.
 - $C_{\alpha ij} \Rightarrow$ electron-electron-ion correlation $\sim N_{\text{atom}}^2$ of $c_{k\alpha}$ parms.
 - $\sim e^{N_{\text{atom}}}$ of d_n parms.

d_n , $c_{k\alpha}$, $\zeta_{k\alpha}$ and parms in \mathcal{J} are optimized.

Power of QMC:

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

Optimization of Wavefunctions

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

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

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

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

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

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. → Variance optim. → 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)

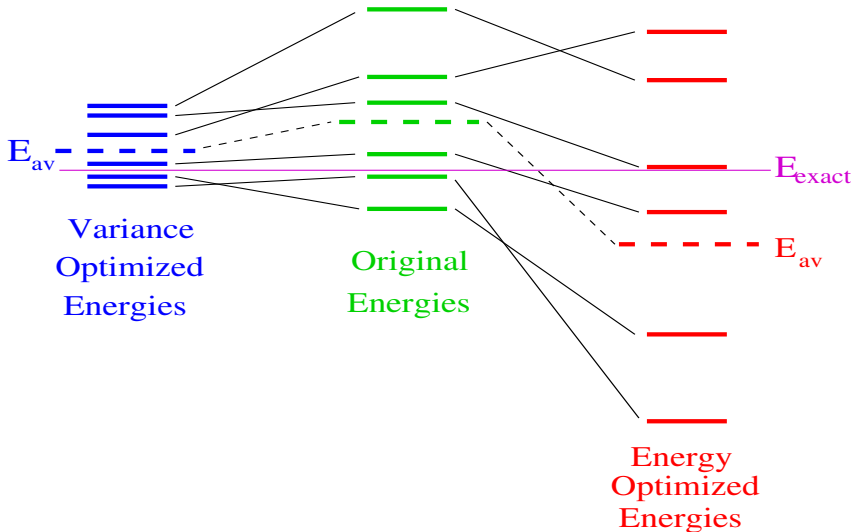
Variance

vs.

Energy

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

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



Projector Monte Carlo Methods

Projector MC

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

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

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

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

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:

- form of the projector, and,
- space in which the walk is done (single-particle basis and quantization).
(1st-quantized \equiv unsymmetrized basis, 2nd-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)	$e^{\tau(E_T \hat{\mathbf{1}} - \hat{H})}$	$\phi_i^{\text{nonorthog}}$	2 nd

$\hat{\mathbf{1}} + \tau(E_T \hat{\mathbf{1}} - \hat{H})$ 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.

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

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

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

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

$$\langle \mathbf{R}' | \hat{P}(\tau) | \mathbf{R} \rangle \equiv G(\mathbf{R}', \mathbf{R}, \tau) \approx \frac{1}{(2\pi\tau)^{3N/2}} e^{\left[-\frac{(\mathbf{R}' - \mathbf{R})^2}{2\tau} + \left\{ E_T - \frac{(\mathcal{V}(\mathbf{R}') + \mathcal{V}(\mathbf{R}))}{2} \right\} \tau \right]}$$

Columns of $G(\mathbf{R}', \mathbf{R}, \tau)$ not normalized to 1, so weights and/or branching needed.
Potential energy $\mathcal{V} \rightarrow \pm\infty$, 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 $\Psi_T(\mathbf{R})$:

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

Inefficient for Bosonic systems because sampled $\Psi_T(\mathbf{R})$ varies by many orders of magnitude, and 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 $\Psi_T(\mathbf{R})$. In the limit that $\Psi_T \rightarrow \Psi_0$ 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 Ψ_T never be nonzero where Ψ_G is zero. In fact the usual practice in DMC is $\Psi_G = \Psi_T$ 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

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

by $\Psi_T(\mathbf{R})$ and rearranging terms we obtain

$$-\frac{\nabla^2}{2}(\Psi\Psi_T) + \nabla \cdot \left(\frac{\nabla\Psi_T}{\Psi_T} \Psi\Psi_T \right) + \underbrace{\left(\frac{-\nabla^2\Psi_T}{2\Psi_T} + \mathcal{V} - E_T \right)}_{E_L(\mathbf{R})}(\Psi\Psi_T) = -\frac{\partial(\Psi\Psi_T)}{\partial t}$$

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

$$\underbrace{-\frac{1}{2}\nabla^2 f}_{\text{diffusion}} + \underbrace{\nabla \cdot \left(\frac{\nabla\Psi_T}{\Psi_T} f \right)}_{\text{drift}} + \underbrace{(E_L(\mathbf{R}) - E_T) f}_{\text{growth/decay}} = -\frac{\partial f}{\partial t}$$

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

$$\tilde{G}(\mathbf{R}', \mathbf{R}, \tau) \approx \frac{1}{(2\pi\tau)^{3N/2}} e^{\left[-\frac{(\mathbf{R}' - \mathbf{R} - \mathbf{V}\tau)^2}{2\tau} + \left\{ E_T - \frac{(E_L(\mathbf{R}') + E_L(\mathbf{R}))}{2} \right\} \tau \right]}$$

Diffusion Monte Carlo with Importance Sampling

$$\tilde{G}(\mathbf{R}', \mathbf{R}, \tau) \approx \frac{1}{(2\pi\tau)^{3N/2}} e^{\left[-\frac{(\mathbf{R}' - \mathbf{R} - \mathbf{V}\tau)^2}{2\tau} + \left\{ E_T - \frac{(E_L(\mathbf{R}') + E_L(\mathbf{R}))}{2} \right\} \tau \right]}$$

The importance-sampled Green function has $E_L(\mathbf{R})$ 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 $E_L(\mathbf{R})$ goes to a constant, E_0 , as $\Psi_T \rightarrow \Psi_0$. 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

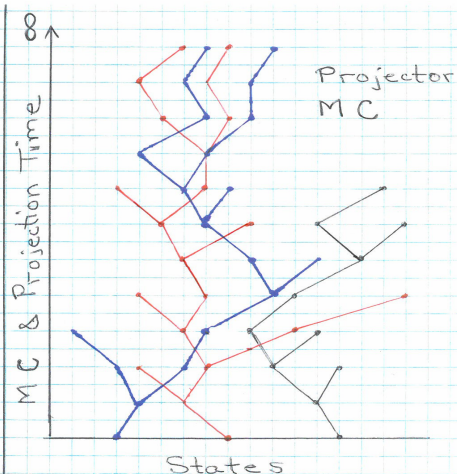
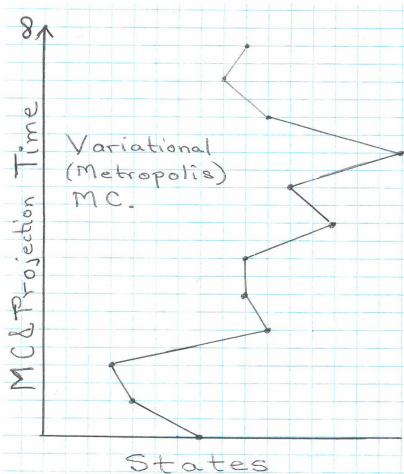
$$\tilde{G}(\mathbf{R}', \mathbf{R}, \tau) \approx \frac{1}{(2\pi\tau)^{3N/2}} e^{\left[-\frac{(\mathbf{R}' - \mathbf{R} - \mathbf{V}\tau)^2}{2\tau} + \left\{ E_T - \frac{(E_L(\mathbf{R}') + E_L(\mathbf{R}))}{2} \right\} \tau \right]}$$

The importance-sampled Green function has $E_L(\mathbf{R})$ 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 $E_L(\mathbf{R})$ goes to a constant, E_0 , as $\Psi_T \rightarrow \Psi_0$. 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!!

Schematic of VMC and PMC



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

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

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

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

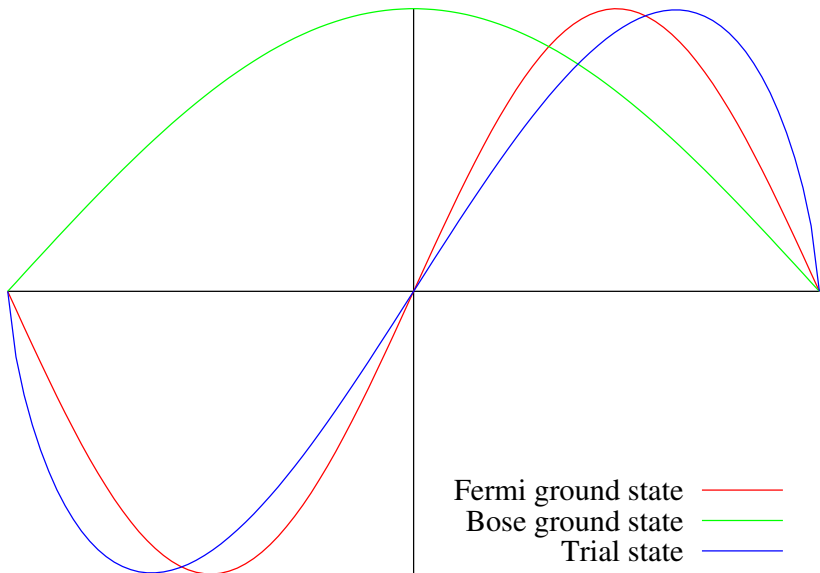
Problem: However, since the Bosonic energy is always lower than the Fermionic energy, the projected state is the Bosonic ground state.

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 $\Psi_T(\mathbf{R})$.

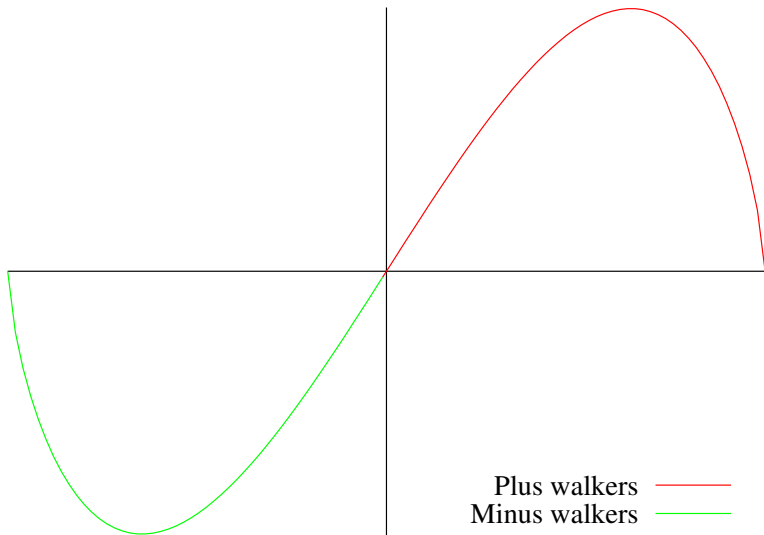
This gives an upper bound to the energy and becomes exact in the limit that Ψ_T has the same nodes as Ψ_0 .

Sign Problem in 1st Quantization and R space

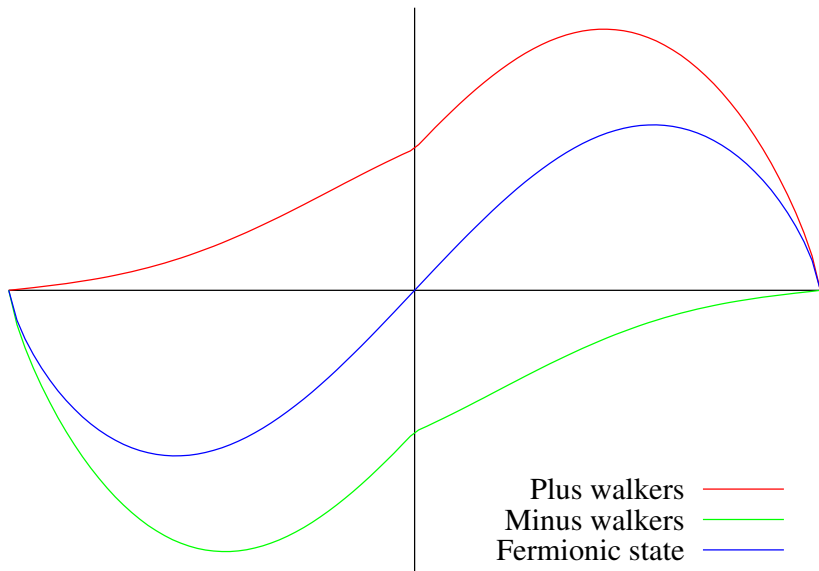


Sign Problem in 1st Quantization and R space

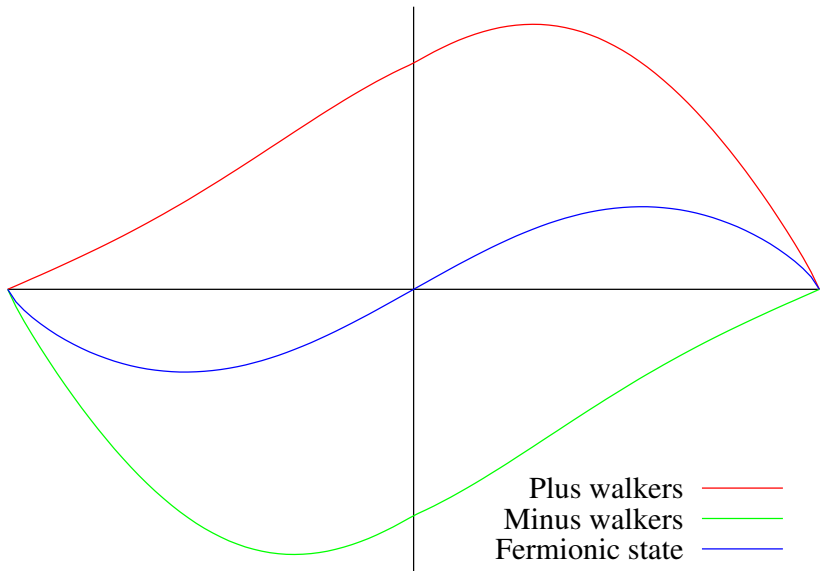
Start with equal + and - walkers, so no Bosonic component.



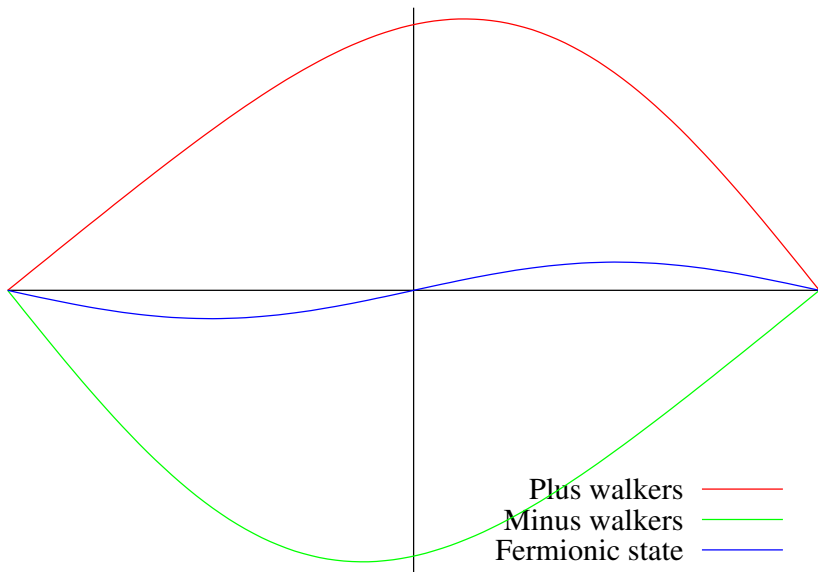
Sign Problem in 1st Quantization and R space



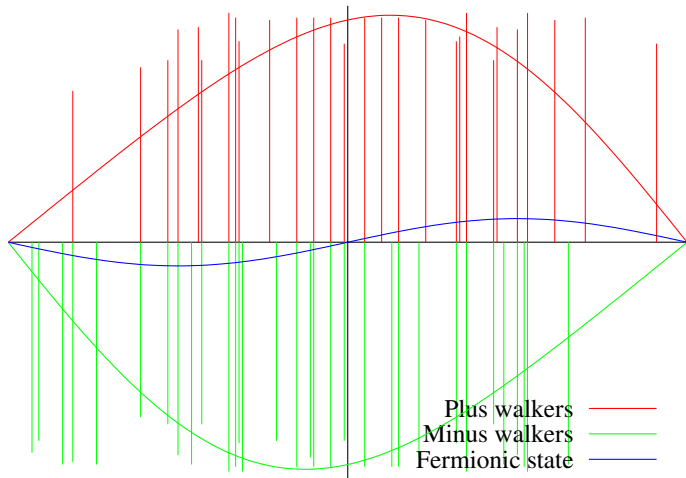
Sign Problem in 1st Quantization and R space



Sign Problem in 1st Quantization and R space



Sign Problem in 1st 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 2nd 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: 2nd-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

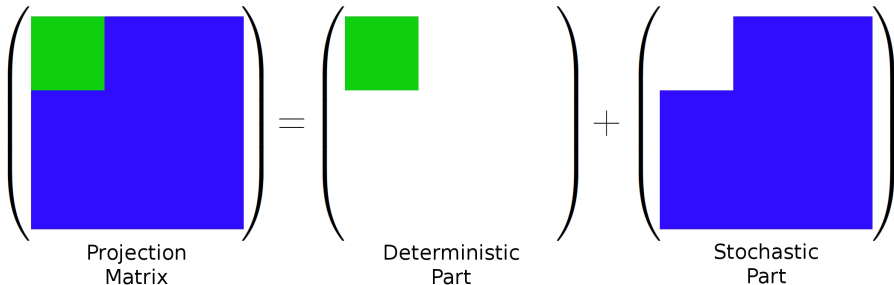
Exact diagonalization has no statistical error or sign problem but is limited to a small number of states ($\sim 10^{10}$ on a single core).

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.

$$P = P^D + P^S$$

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

$$P^S = P - P^D$$

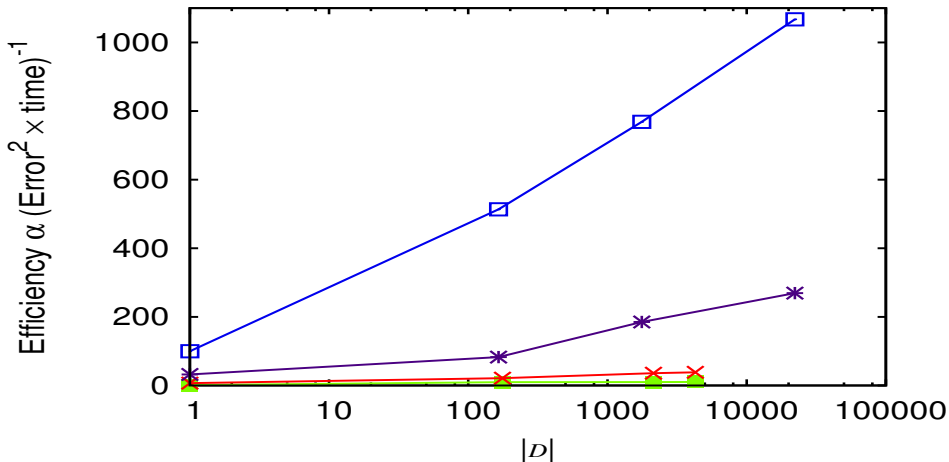
SQMC

Main differences between SQMC and FCIQMC:

1. Deterministic projection in part of space
2. Multideterminantal Ψ_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 gain for C_2 ($3 - \zeta$ basis) from semistochastic projection and Ψ_T

■ $|T|=1$ × $|T|=4282$ * $|T|=165$ □ $|T|=1766$



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

Comparison of DMC with FCIQMC/SQMC

DMC (walk in electron coordinate space)

Severe Fermion sign problem due to growth of Bosonic component relative to Fermionic.

Fixed-node approximation needed for stable algorithm.

Exact if Ψ_T nodes exact.

Infinite basis.

Computational cost is low-order polynomial in N for FN-DMC

Energy is variational

Need to use pseudopotentials for large Z .

FCIQMC/SQMC (walk in determinant space)

Less severe Fermion sign problem due to opposite sign walkers being spawned on the same determinant

Walker cancellation, large population, initiator approximation needed for stable algorithm.

Exact in ∞ -population limit.

Finite basis. (Same basis set dependence as in other quantum chemistry methods.)

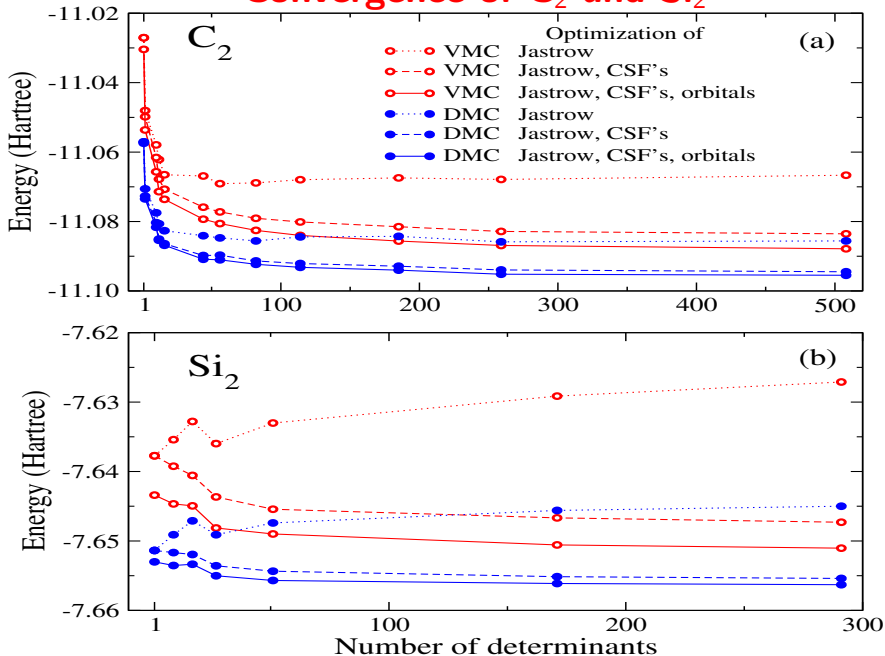
Computational cost is exponential in N but with much smaller exponent than full CI

Energy not variational but DM variant is

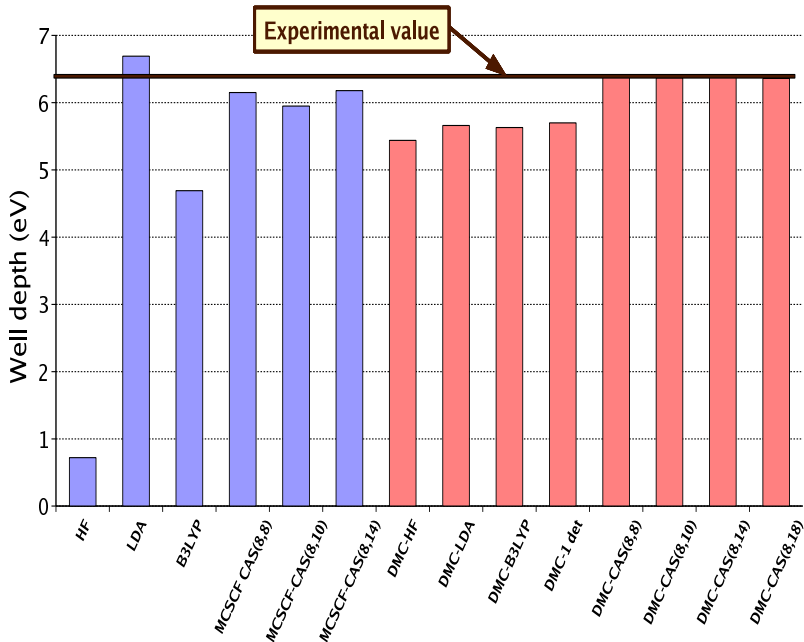
Can easily do frozen-core

Applications

Convergence of C_2 and Si_2

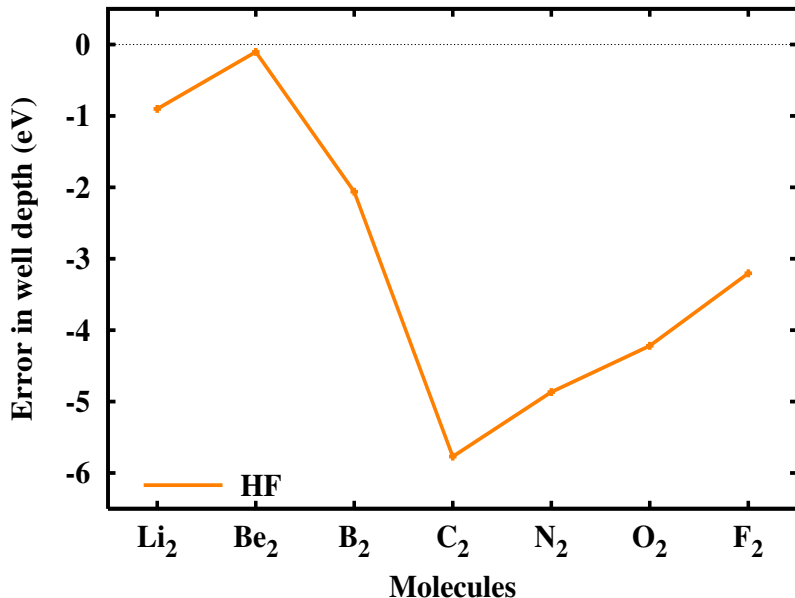


Well-depth of C₂



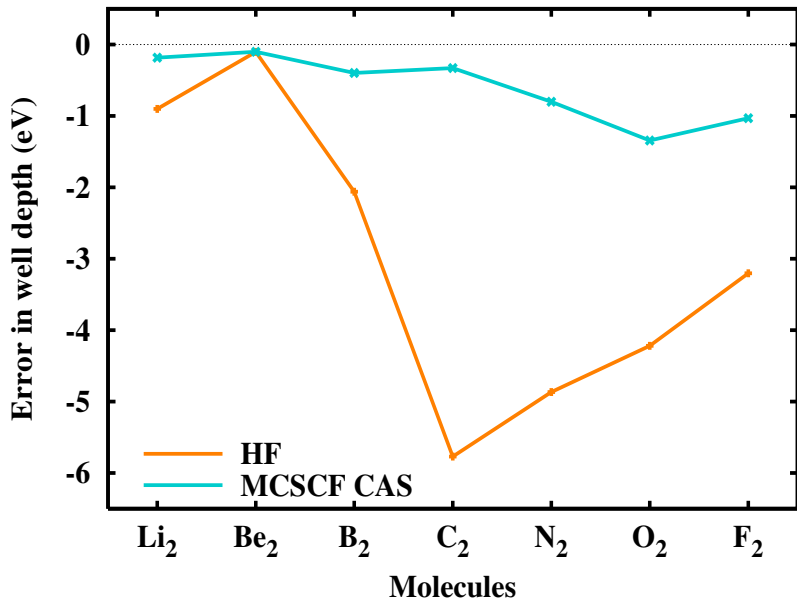
Error in Well-Depth of 1st-Row Diatomic Molecules

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



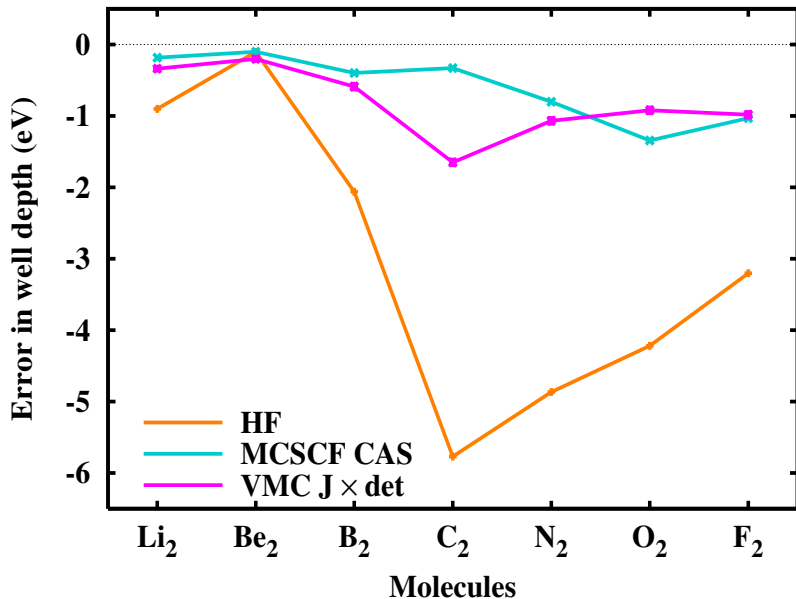
Error in Well-Depth of 1st-Row Diatomic Molecules

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



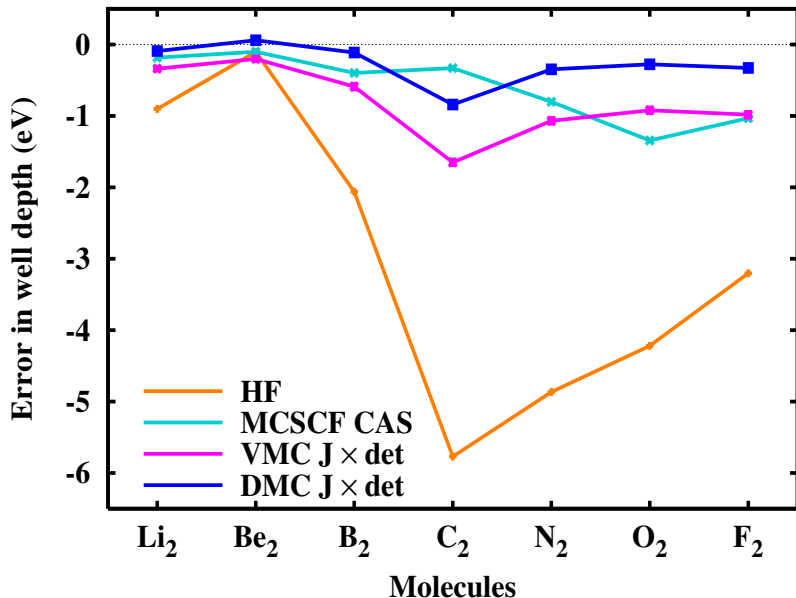
Error in Well-Depth of 1st-Row Diatomic Molecules

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



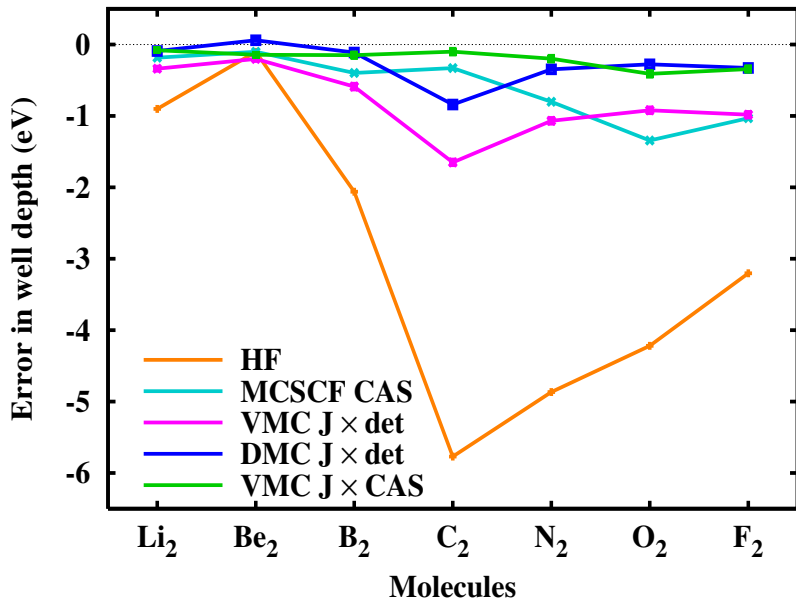
Error in Well-Depth of 1st-Row Diatomic Molecules

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



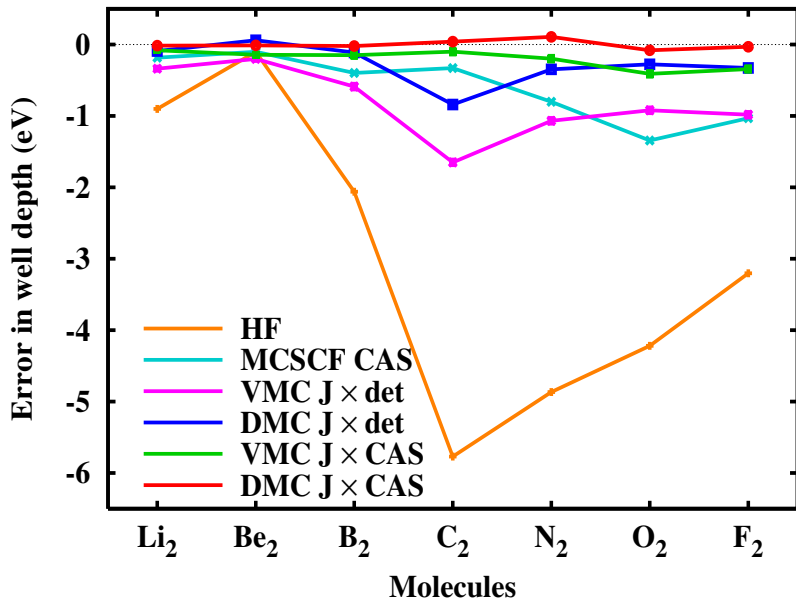
Error in Well-Depth of 1st-Row Diatomic Molecules

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



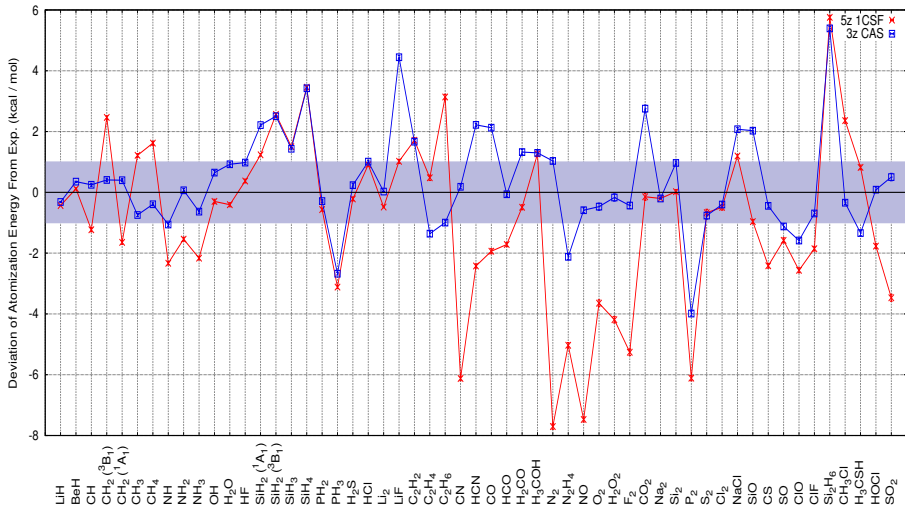
Error in Well-Depth of 1st-Row Diatomic Molecules

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 FV-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, (JCP 2009)

States of methylene (CH_2) are:

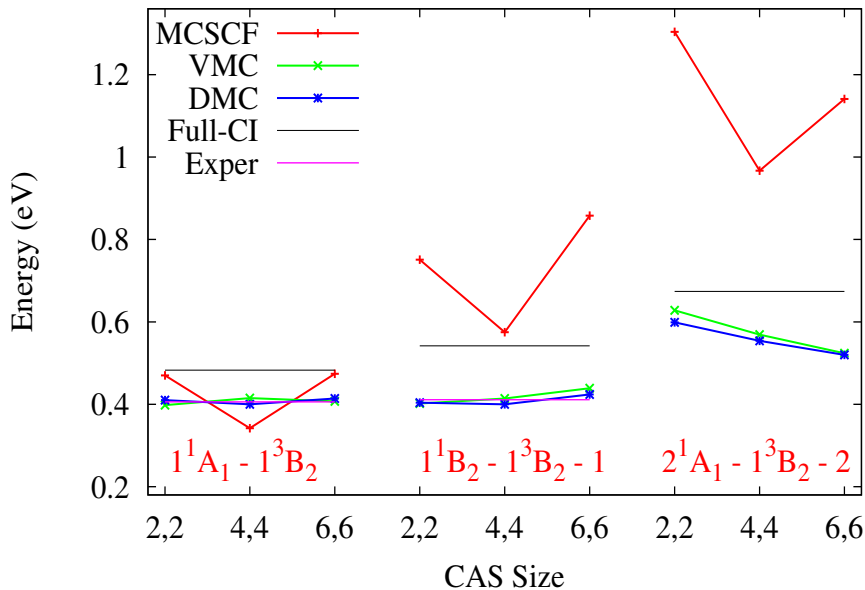
1^3B_2 , ground state, single reference

1^1A_1 , 1st excited state, multi reference

1^1B_2 , 2nd excited state, single reference

2^1A_1 , 3rd excited state, multi reference, true excited state

Convergence of excitation energies of CH₂ with CAS Size

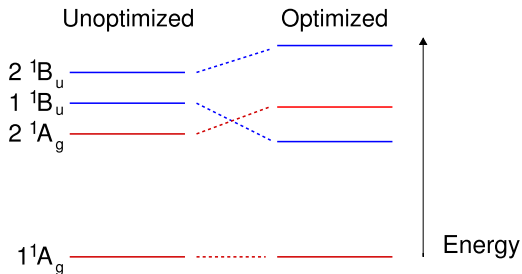


EFP method for ground and excited states

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

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

DMC excitation energies

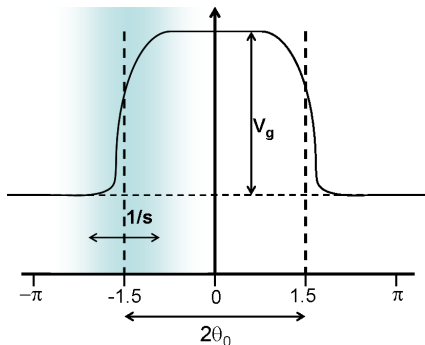
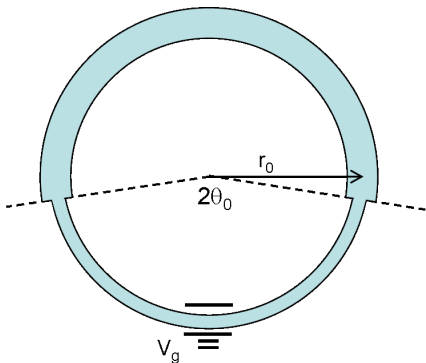


State	Unoptimized	Optimized	expt.	(eV)
1^1B_{1u}	8.45(2)	7.93(2)	>7.7	
2^1A_g	7.96(2)	8.36(2)	8.29	
2^1B_{1u}	9.05(2)	9.37(2)	9.33	

Localization in inhomogeneous quantum wires

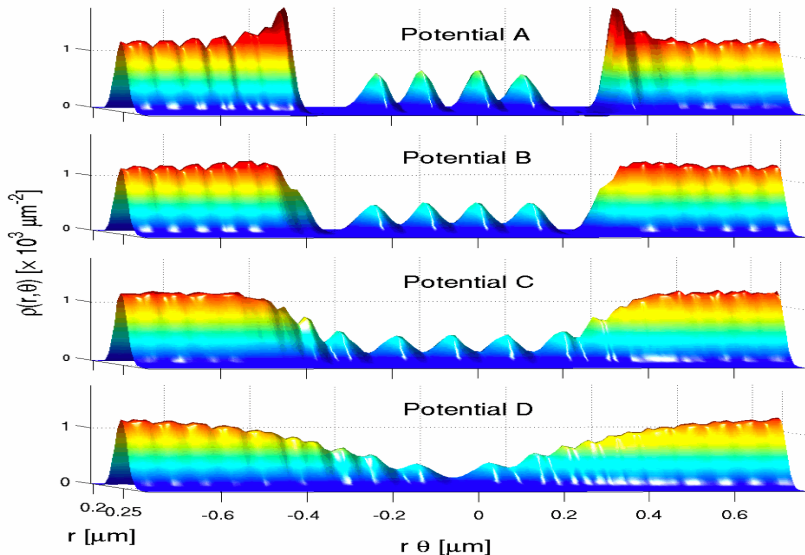
Güçlü, Jiang, CJU, Baranger

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



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.

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

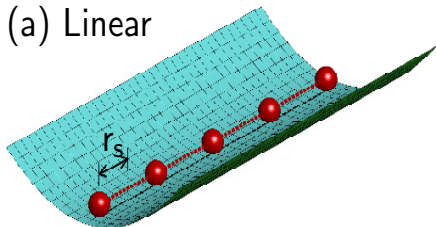
At low densities electrons form linear Wigner crystal. Two length scales:

$r_s = 1/(2n)$, and,

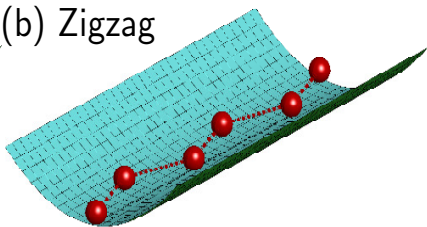
r_0 : confinement and Coulomb energies are equal $(1/2)m\omega^2 r_0^2 = e^2/(\epsilon r_0)$.

As density n is raised, expect a transition to a zigzag phase when $r_s \approx r_0$ before transition to liquid phase.

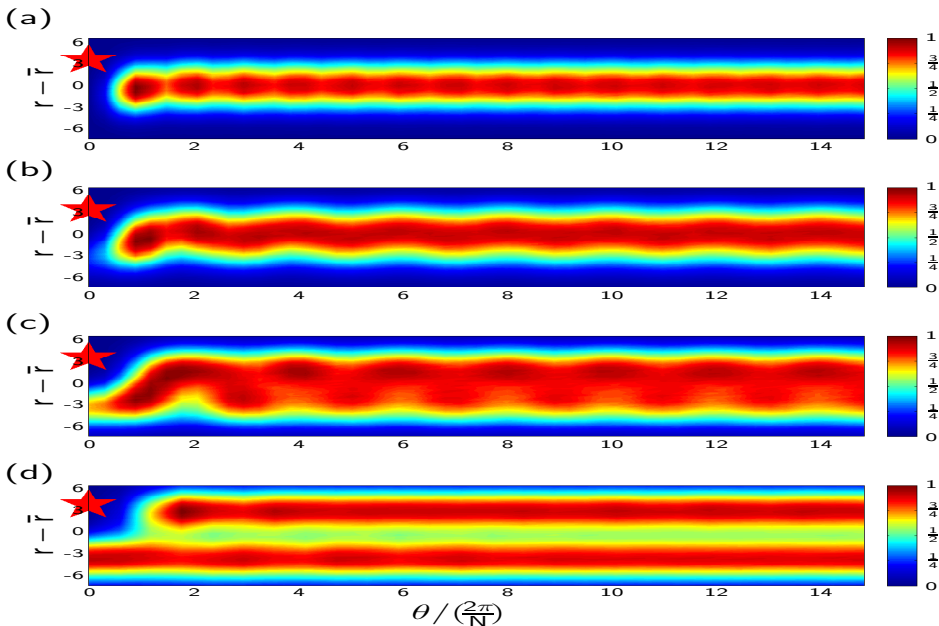
(a) Linear



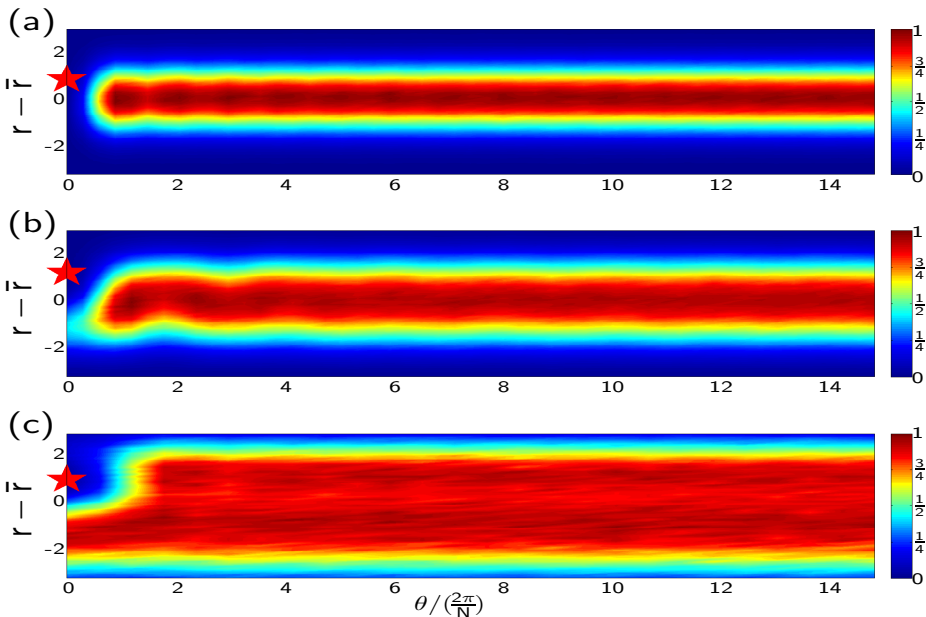
(b) Zigzag



Pair densities at $\omega = 0.1$



Pair densities at $\omega = 0.6$



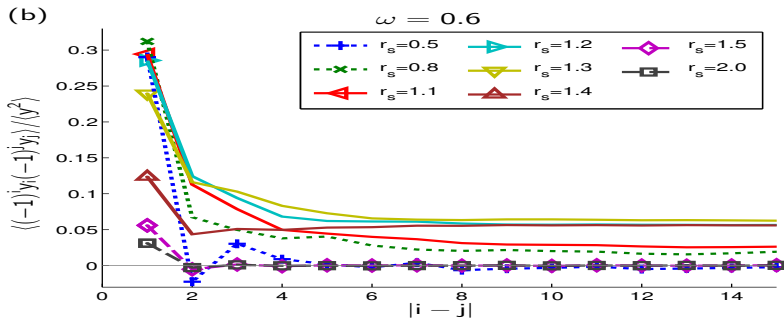
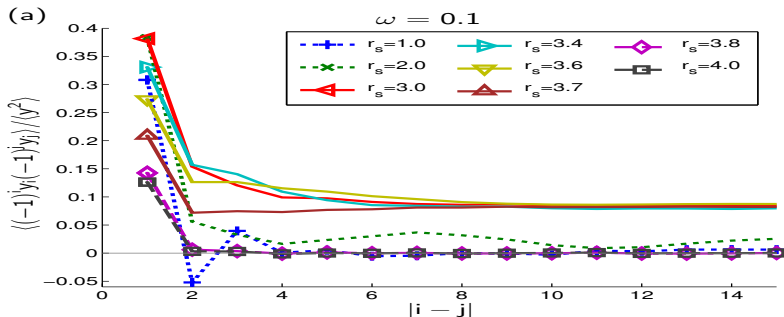
Zigzag Correlation Function

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

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

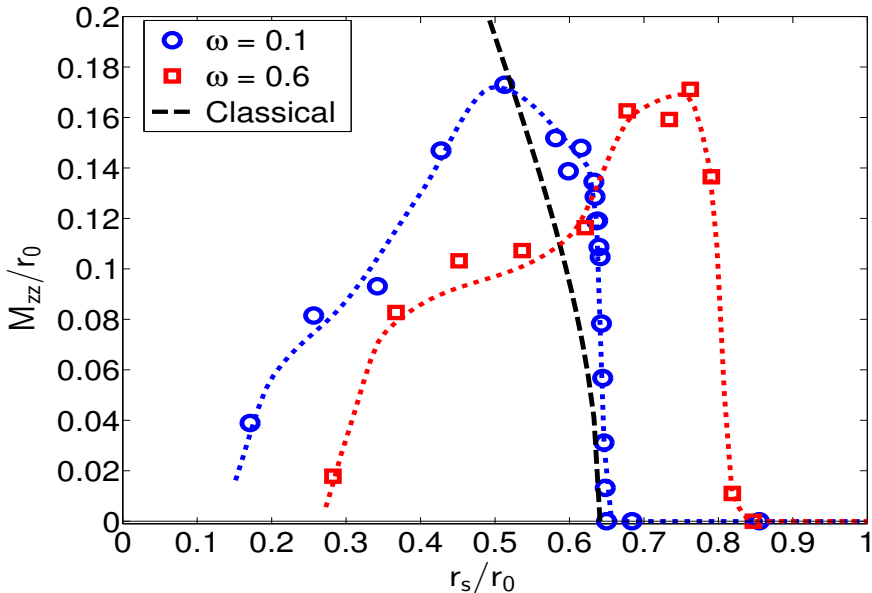
$$C_{ZZ}(|i-j|) = \langle (-1)^i (y_i - \langle y \rangle) (-1)^j (y_j - \langle y \rangle) \rangle$$

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