

Solving the Schrödinger equation  
by random walks:  
Introduction to quantum Monte Carlo methods

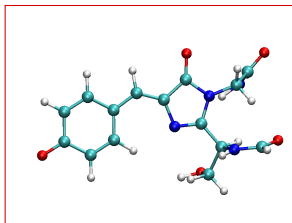
Claudia Filippi

MESA+ Institute for Nanotechnology, Universiteit Twente, The Netherlands

Winterschool on Theoretical Chemistry and Spectroscopy,  
December 12-16, 2011, Han-Sur-Lesse, Belgium

## A quick reminder: What is electronic structure theory?

A quantum mechanical and first-principle approach



→ Collection of ions + electrons



Only input:  $Z_\alpha, N_\alpha$

Work in the Born-Oppenheimer approximation

Solve the Schrödinger equation for the electrons in the ionic field

$$\mathcal{H} = -\frac{1}{2} \sum_i \nabla_i^2 + \sum_i v_{\text{ext}}(\mathbf{r}_i) + \frac{1}{2} \sum_{i \neq j} \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|}$$

## Solving the many-electron Schrödinger equation

$$\mathcal{H} = -\frac{1}{2} \sum_i \nabla_i^2 + \sum_i v_{\text{ext}}(\mathbf{r}_i) + \frac{1}{2} \sum_{i \neq j} \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|}$$

## What do we want to compute?

Fermionic ground state and low-lying excited states

Evaluate expectation values  $\frac{\langle \Psi_n | \mathcal{O} | \Psi_n \rangle}{\langle \Psi_n | \Psi_n \rangle}$

## Where is the difficulty?

Electron-electron interaction → **Non-separable**

## Why quantum simulations?

Why not just doing classical simulations?

Inter-atomic forces are determined by the electrons

Interacting quantum system → Effective inter-atomic potentials

Much simpler and economical but ...

- Empirical potentials → inadequate, non-transferable
- Important phenomena (bond breaking/forming, excitations ...)  
are intrinsically **non** classical

## Is there an optimal theoretical approach?

- Density functional theory methods  
Large systems but approximate exchange/correlation
- Quantum chemistry post-Hartree-Fock methods  
Accurate on relatively small systems  
→ Jungle of approaches: CI, MCSCF, CC, CASPT2 ...

- **Quantum Monte Carlo techniques**

Stochastic solution of the Schrödinger equation

Accurate calculations for medium-large systems

→ Molecules of typically 10-50 1st/2nd-row atoms

→ Relatively little experience with transition metals

→ Solids (Si, C ... Fe, MnO, FeO)

## Is there an optimal theoretical approach?

- Density functional theory methods  
Large systems but approximate exchange/correlation
- Quantum chemistry post-Hartree-Fock methods  
Accurate on relatively small systems  
→ Jungle of approaches: CI, MCSCF, CC, CASPT2 ...

- Quantum Monte Carlo techniques

Stochastic solution of the Schrödinger equation

Accurate calculations for medium-large systems

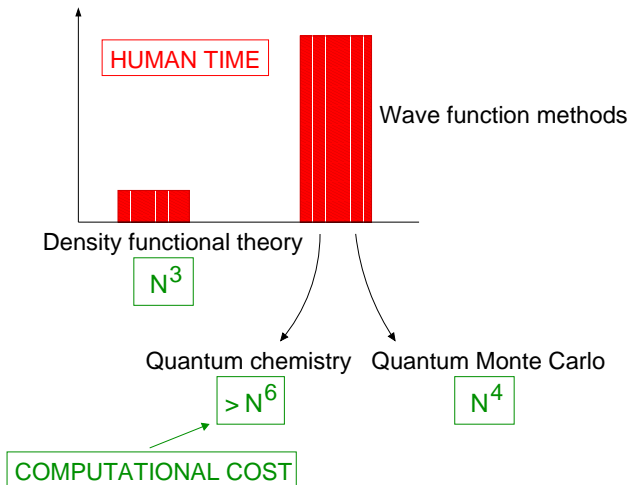
→ Molecules of typically 10-50 1st/2nd-row atoms

→ Relatively little experience with transition metals

→ Solids (Si, C ... Fe, MnO, FeO)

e.g. Geophysics: Bulk Fe with 96 atoms/cell (Alfé 2009)

If you can, use density functional theory!



## Density functional theory: Cheap and painless!

- Hohenberg-Kohn theorem (1964)

Ground-state density  $\rho(\mathbf{r})$  fully characterizes the system

Ground-state energy  $E_0 = E[\rho_0]$

- Kohn-Sham theorem (1965)

Interacting  $\rightarrow$  non-interacting system with same  $\rho(\mathbf{r}) = \sum_{\text{occ}} |\psi_i(\mathbf{r})|^2$

$$\left[ -\frac{1}{2}\nabla^2 + v_{\text{ext}}(\mathbf{r}) + \int \frac{\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} + v_{\text{xc}}(\mathbf{r}) \right] \psi_i(\mathbf{r}) = \epsilon_i \psi_i(\mathbf{r})$$

$$E_0 = -\frac{1}{2} \sum_i \int \psi_i \nabla^2 \psi_i + \int v_{\text{ext}}(\mathbf{r}) \rho(\mathbf{r}) + \frac{1}{2} \iint \frac{\rho(\mathbf{r}) \rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} + E_{\text{xc}}[\rho]$$



## Fast evolution of DFT

State of the art is in constant progression

- 1985 Car-Parrinello molecular dynamics
- 1985 GW calculations → quasiparticle spectra
- 1987 Linear-response for phonons, dielectric tensor
- 1992 Berry-phase approach to polarization
- 1996 Time-dependent DFT for excited states
- 1996 Combine DFT for electrons and quantum ions
- 2000 Using  $O(N)$  → 15000 atoms
- ... Transport, DFT for superconductors etc.

Software engineering: Important part of success

1998 Nobel prize in chemistry to Kohn and Pople

J. Pople → GAUSSIAN70

⋮

GAUSSIAN09

It had a big impact in quantum chemistry community

→ At the origin of the popularity of DFT among chemists

Many commercial, free, open-source codes are now available

Are we theoreticians out of job?

Successful applications of DFT + efficient, user-friendly codes

Can anybody do it?

Better posed questions

Is it always a success story?

Do we have a black-box method close to perfection?

In principle  $\longrightarrow$  DFT is correct

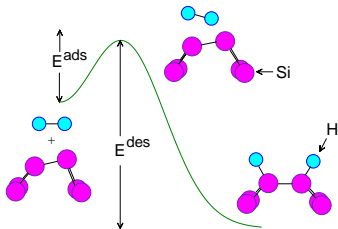
BUT  $E_{xc}[\rho]$  unknown functional of the density

In practice  $\longrightarrow E_{xc}[\rho]$  must be approximated

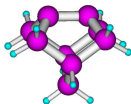
... and sometimes things go wrong

... density functional theory does not always work

A “classical” example: Adsorption/desorption of  $H_2$  on Si(001)



For a small model cluster

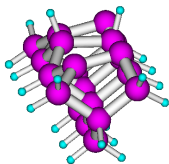


	$E_a^{\text{ads}}$	$E_a^{\text{des}}$	$E_{\text{rxn}}$	
DFT	0.69	2.86	2.17	
QMC	1.01(6)	3.65(6)	2.64(6)	eV

DFT error persists for larger models!

## Favorable scaling of QMC with system size

QMC possible for realistic clusters with 2, 3, 4 ... surface dimers



Accurate QMC calculations doable from small to large scales

Error of DFT is large  $\rightarrow$  0.8 eV on desorption barrier !

Healy, Filippi *et al.* PRL (2001); Filippi *et al.* PRL (2002)

## What about DFT and excited states?

### Density functional methods for excited states

- Restricted open-shell Kohn-Sham method (DFT-ROKS)

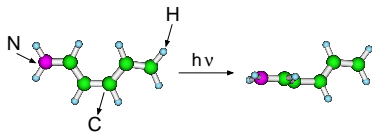
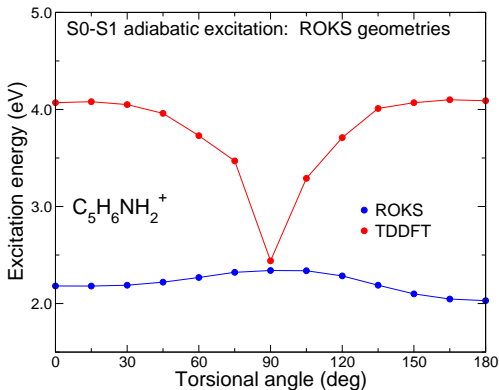
Based on  $\Delta$ SCF approach, efficiently combined with CPMD

- Time-dependent density functional theory (TDDFT)

Formally exact but approximations have limitations

- Lack of two- and higher-electron excitations
- Underestimation of charge-transfer excitations ...

## Problematic example: Minimal model of retinal



Neither approach correctly describes excited-state isomerization

Schautz, Buda, Filippi, JCP (2004)

Several problems remain open

Reaction barriers, weakly bound, strongly correlated systems ...  
Excitations with charge-transfer/multi-configurational character ...

A practical approach or the darkest side of DFT

... Let us get it to work !

And it all began with hybrid schemes such as



$$\frac{1}{3} \text{ Hartree-Fock} + \frac{2}{3} \text{ DFT}$$

which does work better!



# Available functionals in Gaussian09

COMBINATION FORMS		STAND ALONE FUNCTIONALS		
EXCHANGE	CORRELATION	EXCHANGE	PURE	HYBRID
S	VWN	ONLY	VSXC	B3LYP
XA	VWN5	HFS	HCTH	B3P86
B	LYP	XAlpha	HCTH93	B3PW91
PW91	PL	HFB	HCTH147	B1B95
mPW	P86		HCTH407	mPW1PW91
G96	PW91		tHCTH	mPW1LYP
PBE	B95		M06L	mPW1PBE
O	PBE		B97D	mPW3PBE
TPSS	TPSS			B98
BRx	KCIS			B971
PKZB	BRC			B972
wPBEh	PKZB			PBE1PBE
PBEh	VP86			B1LYP
	V5LYP			O3LYP
				BHandH
LONG RANGE				BHandHLYP
CORRECTION				BMK
LC-				M06
				M06HF
				M062X
				tHCTHhyb
				HSEh1PBE
				HSE2PBE
				HSEhPBE
				PBEh1PBE
				wB97XD
				wB97
				wB97X
				TPSSH
				X3LYP
				LC-wPBE
				CAM-B3LYP

... rather distressing

When DFT has problems  $\rightarrow$  Wave function based methods

Wave function  $\Psi(\mathbf{x}_1, \dots, \mathbf{x}_N)$  where  $\mathbf{x} = (\mathbf{r}, \sigma)$  and  $\sigma = \pm 1$

Optimal wave functions and the variational theorem

$\Psi(\mathbf{X}, a)$  with  $\mathbf{X}$  the space-spin variables and  $a$  the parameters

$$E_V(a) = \frac{\langle \Psi(a) | \mathcal{H} | \Psi(a) \rangle}{\langle \Psi(a) | \Psi(a) \rangle} \geq E_0$$

$$E_V(a) = E_0 \Leftrightarrow \Psi(\mathbf{X}, a) = \Psi_0(\mathbf{X})$$

## The variational method and the linear basis approach

Wave function as a linear combination of basis functions  $f_n(\mathbf{X})$

$$\Psi(\mathbf{X}, a) = \sum_n a_n f_n(\mathbf{X}) \Rightarrow E_V(a) = \frac{\sum_{n,m} a_n^* a_m H_{mn}}{\sum_{n,m} a_n^* a_m S_{mn}}$$

where  $H_{nm} = \langle f_n | \mathcal{H} | f_m \rangle$  and  $S_{nm} = \langle f_n | f_m \rangle$

$$\frac{dE}{da} = 2 [Ha - E_V Sa] = 0 \Rightarrow Ha_\lambda = E_\lambda Sa_\lambda$$

Linear basis approach  $\rightarrow$  Generalized eigenvalue problem

$$\Psi(\mathbf{X}, a) = \sum_n a_n f_n(\mathbf{X}) \Rightarrow Ha_\lambda = E_\lambda Sa_\lambda$$

Important properties

▷ For a basis of size  $M$ ,  $\exists$   $M$  eigenvalues and eigenfunctions

▷ McDonald's theorem

$$E_1 \leq E_2 \leq \dots \leq E_M \quad \text{with} \quad E_n^{\text{exact}} \leq E_n$$

## Merits and problems of the variational method

Find approximate solution  $\Psi$  to Schrödinger equation

### Merits

- ▷ Upper bound is guaranteed
- ▷ Linear basis  $\rightarrow$  Generalized eigenvalue problem
- ▷ Linear basis  $\rightarrow$  McDonald's theorem for excited states

### Problems

- ▷ How do we compute the matrix elements  $H_{nm}$  and  $S_{nm}$ ?
- ▷ How do we assess convergence?
- ▷ What goes in, comes out

How do we compute the matrix elements  $S_{nm}$  and  $H_{nm}$ ?

Integrals  $H_{nm}$  and  $S_{nm}$  too slow to perform unless one-particle basis  
→ Problem which can be solved by Monte Carlo integration

Many-body wave functions in traditional quantum chemistry

Interacting  $\Psi(\mathbf{x}_1, \dots, \mathbf{x}_N) \leftrightarrow$  Non-interacting basis  $\psi(\mathbf{x})$

$\Psi$  expanded in determinants of single-particle orbitals  $\psi(\mathbf{x})$

Single-particle orbitals expanded in Gaussian basis

⇒ All integrals can be computed analytically

## Many-body wave functions in traditional quantum chemistry (1)

Starting point  $\rightarrow$  Non-interacting Hartree-Fock wave function

$$D_{\text{HF}}(\mathbf{x}_1, \dots, \mathbf{x}_N) = \begin{vmatrix} \psi_1(\mathbf{x}_1) & \dots & \psi_1(\mathbf{x}_N) \\ \vdots & & \vdots \\ \psi_N(\mathbf{x}_1) & \dots & \psi_N(\mathbf{x}_N) \end{vmatrix}$$

Optimal spin-orbitals  $\psi_i(\mathbf{x}) = \phi_i(\mathbf{r})\chi_{s_i}(\sigma)$  satisfy HF equations

$$\left[ -\frac{1}{2}\nabla^2 + v_{\text{ext}}(\mathbf{r}) + \sum_{j=1}^N \int d\mathbf{r}' \frac{|\phi_j(\mathbf{r}')|^2}{|\mathbf{r} - \mathbf{r}'|} \right] \phi_i(\mathbf{r}) + [\hat{v}_{\text{HF}}\phi_i](\mathbf{r}) = \epsilon_i\phi_i(\mathbf{r})$$

$\Rightarrow$  **occupied** orbitals ( $\psi_1 \dots \psi_N$ ) + **virtual** orbitals ( $\psi_{N+1} \dots$ )

Many-body wave functions in traditional quantum chemistry (2)

A jungle of acronyms: CI, CASSCF, MRCI, CASPT2 ...

Expansion in linear combination of determinants

$$\Psi(\mathbf{x}_1, \dots, \mathbf{x}_N) \longrightarrow D_{\text{HF}} = \begin{vmatrix} \psi_1(\mathbf{x}_1) & \dots & \psi_1(\mathbf{x}_N) \\ \vdots & & \vdots \\ \psi_N(\mathbf{x}_1) & \dots & \psi_N(\mathbf{x}_N) \end{vmatrix}$$

$c_0 D_{\text{HF}} + c_1 D_1 + c_2 D_2 + \dots$  millions of determinants

$$\begin{vmatrix} \psi_1(\mathbf{x}_1) & \dots & \psi_1(\mathbf{x}_N) \\ \vdots & & \vdots \\ \psi_{N+1}(\mathbf{x}_1) & \dots & \psi_{N+1}(\mathbf{x}_N) \end{vmatrix}$$

by constructing single, double, ... up to N-body excitations



## Many-body wave functions in traditional quantum chemistry (3)

Pros and cons of CI expansion in Slater determinants

$$\Psi_{\text{CI}} = c_0 D_{\text{HF}} + \sum_{ab} c_{a \rightarrow b} D^{a \rightarrow b} + \sum_{abcd} c_{ab \rightarrow cd} D^{ab \rightarrow cd} + \dots$$

Optimal CI coefficients by solving generalized eigenvalue equation

$$\Psi_{\text{CI}} = \sum_{i=1}^K c_i D_i \Rightarrow \sum_{j=1}^K \langle D_i | \mathcal{H} | D_j \rangle c_j^{(k)} = E_{\text{CI}}^{(k)} \sum_{j=1}^K \langle D_i | D_j \rangle c_j^{(k)}$$

Orbitals on a Gaussian basis  $\rightarrow$  Integrals computed analytically

... but slowly converging expansion

Can we use a more compact  $\Psi$ ?

We want to construct an accurate and more compact  $\Psi$

Explicit dependence on the inter-electronic distances  $r_{ij}$

How do we compute expectation values if no single-electron basis?

## A different way of writing the expectation values

Consider the expectation value of the Hamiltonian on  $\Psi$

$$\begin{aligned} E_V &= \frac{\langle \Psi | \mathcal{H} | \Psi \rangle}{\langle \Psi | \Psi \rangle} = \frac{\int d\mathbf{R} \Psi^*(\mathbf{R}) \mathcal{H} \Psi(\mathbf{R})}{\int d\mathbf{R} \Psi^*(\mathbf{R}) \Psi(\mathbf{R})} \geq E_0 \\ &= \int d\mathbf{R} \frac{\mathcal{H} \Psi(\mathbf{R})}{\Psi(\mathbf{R})} \frac{|\Psi(\mathbf{R})|^2}{\int d\mathbf{R} |\Psi(\mathbf{R})|^2} \\ &= \int d\mathbf{R} E_L(\mathbf{R}) \rho(\mathbf{R}) = \langle E_L(\mathbf{R}) \rangle_\rho \end{aligned}$$

$\rho$  is a distribution function and  $E_L(\mathbf{R}) = \frac{\mathcal{H} \Psi(\mathbf{R})}{\Psi(\mathbf{R})}$  the local energy

## Variational Monte Carlo: a random walk of the electrons

Use Monte Carlo integration to compute expectation values

- ▷ Sample  $\mathbf{R}$  from  $\rho(\mathbf{R})$  using Metropolis algorithm
- ▷ Average local energy  $E_L(\mathbf{R}) = \frac{\mathcal{H}\Psi(\mathbf{R})}{\Psi(\mathbf{R})}$  to obtain  $E_V$  as

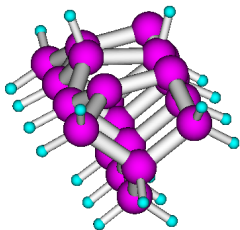
$$E_V = \langle E_L(\mathbf{R}) \rangle_\rho \approx \frac{1}{M} \sum_{i=1}^M E_L(\mathbf{R}_i)$$



Random walk in  $3N$  dimensions,  $\mathbf{R} = (\mathbf{r}_1, \dots, \mathbf{r}_N)$

Just a **trick** to evaluate integrals in many dimensions

Is it really “just” a trick?



$\text{Si}_{21}\text{H}_{22}$

Number of electrons  $4 \times 21 + 22 = 106$

Number of dimensions  $3 \times 106 =$  318

Integral on a grid with 10 points/dimension  $\rightarrow 10^{318}$  points!

MC is a powerful trick  $\Rightarrow$  Freedom in form of the wave function  $\Psi$

## Monte Carlo integration

We want to compute an integral

$$E_V = \int d\mathbf{R} E_L(\mathbf{R}) \rho(\mathbf{R})$$

We sample  $\rho(\mathbf{R}) \rightarrow$   $E_V = \langle E_L(\mathbf{R}) \rangle_\rho \approx \frac{1}{M} \sum_{i=1}^M E_L(\mathbf{R}_i)$

- Does the trick always work?
- How efficient is it?

## The Central Limit Theorem

Consider a probability density  $\rho(x)$  and function  $f(x)$  with a finite mean and variance

$$\boxed{\mu} = \int dx f(x)\rho(x) \quad \boxed{\sigma^2} = \int dx (f(x)^2 - \mu)\rho(x)$$

Sample  $M$  independent random variables  $x_1, \dots, x_M$  from  $\rho(x)$  and define

$$F_M = \frac{1}{M} \sum_{i=1}^M f(x_i)$$

As  $M$  increases,  $F_M$  is normally distributed as  $\frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/2\sigma^2}$

with a mean  $\boxed{\mu}$  and variance  $\boxed{\sigma_M^2 = \sigma^2/M}$

→ Irrespective of the original probability density function

## Monte Carlo versus deterministic integration

Integration error  $\epsilon$  using  $M$  integration/Monte Carlo points

### – Deterministic integration methods

1-dim Simpson rule:  $\epsilon \propto \frac{1}{M^4}$

$d$ -dim Simpson rule:  $\epsilon \propto \frac{1}{M^{4/d}}$

For a given error,  $M$  grows exponentially with  $d$  as  $M \propto (1/\epsilon)^{d/4}$

### – Monte Carlo methods

$\epsilon \propto \frac{1}{\sqrt{M}}$  independent on dimension !

It follows from Central Limit Theorem

→ width of Gaussian decreases as  $\frac{1}{\sqrt{M}}$  for finite variance



## Scaling with number of electrons

Roughly, Monte Carlo integration advantageous if  $d > 8$

... for many-body wave functions  $d = 3N_{\text{elec}}$  !

– Simpson rule integration ( $M_{\text{int}}$  integration points)

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

– Monte Carlo integration ( $M_{\text{MC}}$  Monte Carlo samples)

$$\epsilon = c \sqrt{\frac{N_{\text{elec}}}{M_{\text{MC}}}} \Rightarrow M_{\text{MC}} = \left(\frac{c}{\epsilon}\right)^2 N_{\text{elec}} \quad \text{Linear}$$

## Reminder of variational Monte Carlo

Expectation value of the Hamiltonian on  $\Psi$

$$E_V = \frac{\langle \Psi | \mathcal{H} | \Psi \rangle}{\langle \Psi | \Psi \rangle} = \int d\mathbf{R} \frac{\mathcal{H}\Psi(\mathbf{R})}{\Psi(\mathbf{R})} \frac{|\Psi(\mathbf{R})|^2}{\int d\mathbf{R} |\Psi(\mathbf{R})|^2} = \int d\mathbf{R} E_L(\mathbf{R}) \rho(\mathbf{R})$$

$$E_V = \int d\mathbf{R} E_L(\mathbf{R}) \rho(\mathbf{R})$$

$$\sigma^2 = \int d\mathbf{R} (E_L(\mathbf{R}) - E_V)^2 \rho(\mathbf{R})$$

Estimate  $E_V$  and  $\sigma$  from  $M$  independent samples as

$$\bar{E}_V = \frac{1}{M} \sum_{i=1}^M E_L(\mathbf{R}_i)$$

$$\bar{\sigma}^2 = \frac{1}{M} \sum_{i=1}^M (E_L(\mathbf{R}_i) - \bar{E}_V)^2$$

Are there any conditions on many-body  $\Psi$  to be used in VMC?

Within VMC, we can use any “computable” wave function if

- ▷ Continuous, normalizable, proper symmetry
- ▷ Finite variance

$$\sigma^2 = \frac{\langle \Psi | (\mathcal{H} - E_V)^2 | \Psi \rangle}{\langle \Psi | \Psi \rangle} = \langle (E_L(\mathbf{R}) - E_V)^2 \rangle_\rho$$

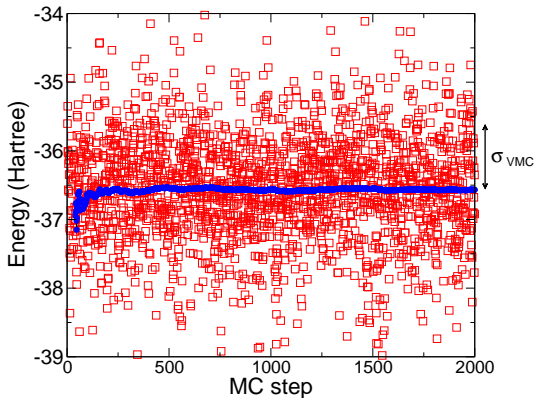
since the Monte Carlo error goes as

$$\text{err}(E_V) \sim \frac{\sigma}{\sqrt{M}}$$

Zero variance principle: if  $\Psi \rightarrow \Psi_0$ ,  $E_L(\mathbf{R})$  does not fluctuate

## Typical VMC run

Example: Local energy and average energy of acetone ( $C_3H_6O$ )



$$E_{\text{VMC}} = \langle E_L(\mathbf{R}) \rangle_{\rho} = -36.542 \pm 0.001 \text{ Hartree (40} \times \text{20000 steps)}$$

$$\sigma_{\text{VMC}} = \langle (E_L(\mathbf{R}) - E_{\text{VMC}})^2 \rangle_{\rho} = 0.90 \text{ Hartree}$$

## Variational Monte Carlo and the generalized Metropolis algorithm

How do we sample distribution function  $\rho(\mathbf{R}) = \frac{|\Psi(\mathbf{R})|^2}{\int d\mathbf{R} |\Psi(\mathbf{R})|^2}$  ?

**Aim** → Obtain a set of  $\{\mathbf{R}_1, \mathbf{R}_2, \dots, \mathbf{R}_M\}$  distributed as  $\rho(\mathbf{R})$

Generate a **Markov chain**

- ▷ Start from arbitrary initial state  $\mathbf{R}_i$
- ▷ Use stochastic transition matrix  $P(\mathbf{R}_f | \mathbf{R}_i)$

$$P(\mathbf{R}_f | \mathbf{R}_i) \geq 0 \quad \sum_{\mathbf{R}_f} P(\mathbf{R}_f | \mathbf{R}_i) = 1.$$

as probability of making transition  $\mathbf{R}_i \rightarrow \mathbf{R}_f$

- ▷ Evolve the system by repeated application of  $P$

## Stationarity condition

To sample  $\rho$ , use  $P$  which satisfies **stationarity condition**:

$$\sum_i P(\mathbf{R}_f | \mathbf{R}_i) \rho(\mathbf{R}_i) = \rho(\mathbf{R}_f) \quad \forall \mathbf{R}_f$$

▷ Stationarity condition

⇒ If we start with  $\rho$ , we continue to sample  $\rho$

▷ Stationarity condition + stochastic property of  $P$  + ergodicity

⇒ Any initial distribution will evolve to  $\rho$

More stringent condition

In practice, we impose detailed balance condition

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

Stationarity condition can be obtained by summing over  $\mathbf{R}_i$

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

Detailed balance is a sufficient but not necessary condition

How do we construct the transition matrix  $P$  in practice?

Write transition matrix  $P$  as proposal  $T \times$  acceptance  $A$

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

$P$  and  $T$  are stochastic matrices but  $A$  is not

Detailed balance condition becomes

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

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



## Choice of acceptance matrix $A$

(1)

Detailed balance condition is

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

For a given choice of  $T$ , infinite choices of  $A$  satisfy this equation

Any function  $A(\mathbf{R}_f|\mathbf{R}_i) = F\left(\frac{T(\mathbf{R}_i|\mathbf{R}_f) \rho(\mathbf{R}_f)}{T(\mathbf{R}_f|\mathbf{R}_i) \rho(\mathbf{R}_i)}\right)$  with

$$\frac{F(x)}{F(1/x)} = x \quad \text{and} \quad 0 \leq F(x) \leq 1$$

will do the job!

Choice of acceptance matrix  $A$

(2)

Original choice by Metropolis *et al.* maximizes the acceptance

$$A(\mathbf{R}_f|\mathbf{R}_i) = \min \left\{ 1, \frac{T(\mathbf{R}_i|\mathbf{R}_f) \rho(\mathbf{R}_f)}{T(\mathbf{R}_f|\mathbf{R}_i) \rho(\mathbf{R}_i)} \right\}$$

Note:  $\rho(\mathbf{R})$  does not have to be normalized

Original Metropolis method

$$\text{Symmetric } T(\mathbf{R}_f|\mathbf{R}_i) = 1/\Delta^{3N} \Rightarrow A(\mathbf{R}_f|\mathbf{R}_i) = \min \left\{ 1, \frac{\rho(\mathbf{R}_f)}{\rho(\mathbf{R}_i)} \right\}$$

Is this the best possible choice for  $T$ ?

Choice of proposal matrix  $T$

(1)

Sequential correlation  $\Rightarrow M_{\text{eff}} < M$  independent observations

$$M_{\text{eff}} = \frac{M}{T_{\text{CORR}}} \text{ with } T_{\text{CORR}} \text{ autocorrelation time of desired observable}$$

Aim is to achieve fast evolution of the system and reduce  $T_{\text{CORR}}$

Use freedom in choice of  $T$  to have high acceptance

$$\frac{T(\mathbf{R}_i|\mathbf{R}_f) \rho(\mathbf{R}_f)}{T(\mathbf{R}_f|\mathbf{R}_i) \rho(\mathbf{R}_i)} \approx 1 \Rightarrow A(\mathbf{R}_f|\mathbf{R}_i) \approx 1$$

and small  $T_{\text{CORR}}$  of desired observable

## Choice of proposal matrix $T$

(2)

If  $\Delta$  is the linear dimension of domain around  $\mathbf{R}_i$

$$\frac{A(\mathbf{R}_f|\mathbf{R}_i)}{A(\mathbf{R}_i|\mathbf{R}_f)} = \frac{T(\mathbf{R}_i|\mathbf{R}_f)}{T(\mathbf{R}_f|\mathbf{R}_i)} \frac{\rho(\mathbf{R}_f)}{\rho(\mathbf{R}_i)} \approx 1 - \mathcal{O}(\Delta^m)$$

- ▷  $T$  symmetric as in original Metropolis algorithm gives  $m = 1$
- ▷ A choice motivated by diffusion Monte Carlo with  $m = 2$  is

$$T(\mathbf{R}_f|\mathbf{R}_i) = \mathcal{N} \exp \left[ -\frac{(\mathbf{R}_f - \mathbf{R}_i - \mathbf{V}(\mathbf{R}_i)\tau)^2}{2\tau} \right] \quad \text{with} \quad \mathbf{V}(\mathbf{R}_i) = \frac{\nabla \Psi(\mathbf{R}_i)}{\Psi(\mathbf{R}_i)}$$

- ▷ Other (better) choices of  $T$  are possible

## Acceptance and $T_{\text{corr}}$ for the total energy $E_V$

Example: All-electron Be atom with simple wave function

### Simple Metropolis

$\Delta$	$T_{\text{corr}}$	$\bar{A}$
1.00	41	0.17
0.75	21	0.28
0.50	17	0.46
0.20	45	0.75

### Drift-diffusion transition

$\tau$	$T_{\text{corr}}$	$\bar{A}$
0.100	13	0.42
0.050	7	0.66
0.020	8	0.87
0.010	14	0.94

## Generalized Metropolis algorithm

1. Choose distribution  $\rho(\mathbf{R})$  and proposal matrix  $T(\mathbf{R}_f|\mathbf{R}_i)$
2. Initialize the configuration  $\mathbf{R}_i$
3. Advance the configuration from  $\mathbf{R}_i$  to  $\mathbf{R}'$ 
  - a) Sample  $\mathbf{R}'$  from  $T(\mathbf{R}'|\mathbf{R}_i)$ .
  - b) Calculate the ratio  $\rho = \frac{T(\mathbf{R}_i|\mathbf{R}') \rho(\mathbf{R}')}{T(\mathbf{R}'|\mathbf{R}_i) \rho(\mathbf{R}_i)}$
  - c) Accept or reject with probability  $\rho$ 

Pick a uniformly distributed random number  $\chi \in [0, 1]$

if  $\chi < \rho$ , move accepted  $\rightarrow$  set  $\mathbf{R}_f = \mathbf{R}'$

if  $\chi > \rho$ , move rejected  $\rightarrow$  set  $\mathbf{R}_f = \mathbf{R}$
4. Throw away first  $\kappa$  configurations of equilibration time
5. Collect the averages and block them to obtain the error bars

We compute the expectation value of the Hamiltonian  $\mathcal{H}$  as

$$\begin{aligned} E_V &= \frac{\langle \Psi | \mathcal{H} | \Psi \rangle}{\langle \Psi | \Psi \rangle} \\ &= \int d\mathbf{R} \frac{\mathcal{H}\Psi(\mathbf{R})}{\Psi(\mathbf{R})} \frac{|\Psi(\mathbf{R})|^2}{\int d\mathbf{R} |\Psi(\mathbf{R})|^2} \\ &= \int d\mathbf{R} E_L(\mathbf{R}) \rho(\mathbf{R}) \\ &= \langle E_L(\mathbf{R}) \rangle_\rho \approx \frac{1}{M} \sum_{i=1}^M E_L(\mathbf{R}_i) \end{aligned}$$

Note: a) Metropolis method:  $\rho$  does not have to be normalized

→ For complex  $\Psi$  we do not know the normalization!

b) If  $\Psi \rightarrow$  eigenfunction,  $E_L(\mathbf{R})$  does not fluctuate

## Expectation values in variational Monte Carlo

(2)

The energy is computed by averaging the local energy

$$E_V = \frac{\langle \Psi | \mathcal{H} | \Psi \rangle}{\langle \Psi | \Psi \rangle} = \langle E_L(\mathbf{R}) \rangle_\rho$$

The variance of the local energy is given by

$$\sigma^2 = \frac{\langle \Psi | (\mathcal{H} - E_V)^2 | \Psi \rangle}{\langle \Psi | \Psi \rangle} = \langle (E_L(\mathbf{R}) - E_V)^2 \rangle_\rho$$

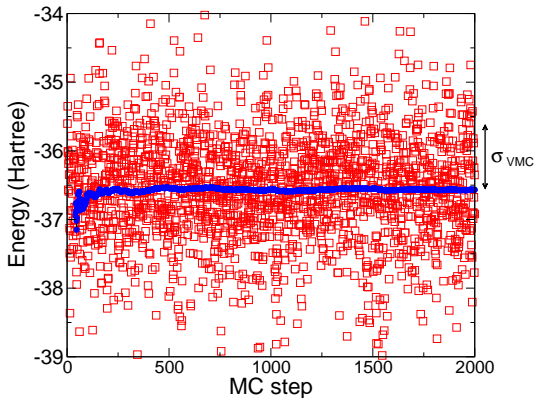
The statistical Monte Carlo error goes as  $\text{err}(E_V) \sim \frac{\sigma}{\sqrt{M}}$

Note: For other operators, substitute  $\mathcal{H}$  with  $\mathcal{X}$



## Typical VMC run

Example: Local energy and average energy of acetone ( $C_3H_6O$ )



$$E_{VMC} = \langle E_L(\mathbf{R}) \rangle_{\rho} = -36.542 \pm 0.001 \text{ Hartree (40} \times \text{20000 steps)}$$

$$\sigma_{VMC} = \langle (E_L(\mathbf{R}) - E_{VMC})^2 \rangle_{\rho} = 0.90 \text{ Hartree}$$

Variational Monte Carlo  $\rightarrow$  Freedom in choice of  $\Psi$

Monte Carlo integration allows the use of complex and accurate  $\Psi$

$\Rightarrow$  More compact representation of  $\Psi$  than in quantum chemistry

$\Rightarrow$  Beyond  $c_0 D_{\text{HF}} + c_1 D_1 + c_2 D_2 + \dots$  millions of determinants

## Jastrow-Slater wave function

$$\Psi(\mathbf{r}_1, \dots, \mathbf{r}_N) = \mathcal{J}(\mathbf{r}_1, \dots, \mathbf{r}_N) \sum_k d_k D_k^\uparrow(\mathbf{r}_1, \dots, \mathbf{r}_{N_\uparrow}) D_k^\downarrow(\mathbf{r}_{N_\uparrow+1}, \dots, \mathbf{r}_N)$$

$\mathcal{J}$   $\longrightarrow$  Jastrow correlation factor

- Positive function of inter-particle distances
- Explicit dependence on electron-electron distances
- Takes care of divergences in potential

$\sum_k c_k D_k$   $\longrightarrow$  Determinants of single-particle orbitals

- **Few** and not millions of determinants as in quantum chemistry
- Determines the nodal surface

What is strange with the Jastrow-Slater wave function?

$$\Psi(\mathbf{r}_1, \dots, \mathbf{r}_N) = \mathcal{J}(\mathbf{r}_1, \dots, \mathbf{r}_N) \sum_k d_k D_k^\uparrow(\mathbf{r}_1, \dots, \mathbf{r}_{N_\uparrow}) D_k^\downarrow(\mathbf{r}_{N_\uparrow+1}, \dots, \mathbf{r}_N)$$

▷ Why is  $\Psi$  not depending on the spin variables  $\sigma$ ?

$$\Psi(\mathbf{x}_1, \dots, \mathbf{x}_N) = \Psi(\mathbf{r}_1, \sigma_1, \dots, \mathbf{r}_N, \sigma_N) \quad \text{with} \quad \sigma_i = \pm 1$$

▷ Why is  $\Psi$  not totally antisymmetric?

Why can we factorize  $D_k^\uparrow D_k^\downarrow$ ?

Consider  $N$  electrons with  $N = N_\uparrow + N_\downarrow$  and  $S_z = (N_\uparrow - N_\downarrow)/2$

$$\Psi(\mathbf{x}_1, \dots, \mathbf{x}_N) = \Psi(\mathbf{r}_1, \sigma_1, \dots, \mathbf{r}_N, \sigma_N) \quad \text{with} \quad \sigma_i = \pm 1$$

Define a spin function  $\zeta_1$

$$\zeta_1(\sigma_1, \dots, \sigma_N) = \chi_\uparrow(\sigma_1) \dots \chi_\uparrow(\sigma_{N_\uparrow}) \chi_\downarrow(\sigma_{N_\uparrow+1}) \dots \chi_\downarrow(\sigma_N)$$

Generate  $K = N!/N_\uparrow!N_\downarrow!$  functions  $\zeta_i$  by permuting indices in  $\zeta_1$

The functions  $\zeta_i$  form a complete, orthonormal set in spin space

$$\sum_{\sigma_1 \dots \sigma_N} \zeta_i(\sigma_1, \dots, \sigma_N) \zeta_j(\sigma_1, \dots, \sigma_N) = \delta_{ij}$$

## Wave function with space and spin variables

Expand the wave function  $\Psi$  in terms of its spin components

$$\Psi(\mathbf{x}_1, \dots, \mathbf{x}_N) = \sum_{i=1}^K F_i(\mathbf{r}_1, \dots, \mathbf{r}_N) \zeta_i(\sigma_1, \dots, \sigma_N)$$

$\Psi$  is totally antisymmetric  $\Rightarrow$

- ▷  $F_i = -F_j$  for interchange of like-spin
- ▷  $F_i = \pm$  permutation of  $F_1$

$$\Psi(\mathbf{x}_1, \dots, \mathbf{x}_N) = \mathcal{A} \{ F_1(\mathbf{r}_1, \dots, \mathbf{r}_N) \zeta_1(\sigma_1, \dots, \sigma_N) \}$$

## Can we get rid of spin variables? Spin-assigned wave functions

Note that if  $\mathcal{O}$  is a spin-independent operator

$$\langle \Psi | \mathcal{O} | \Psi \rangle = \langle F_1 | \mathcal{O} | F_1 \rangle$$

since the functions  $\zeta_j$  form an orthonormal set

More convenient to use  $F_1$  instead of full wave function  $\Psi$

To obtain  $F_1$ , assign the spin-variables of particles:

Particle	1	2	...	$N_{\uparrow}$	$N_{\uparrow+1}$	...	$N$
$\sigma$	1	1	...	1	-1	...	-1

$$F_1(\mathbf{r}_1, \dots, \mathbf{r}_N) = \Psi(\mathbf{r}_1, 1, \dots, \mathbf{r}_{N_{\uparrow}}, 1, \mathbf{r}_{N_{\uparrow+1}}, -1, \dots, \mathbf{r}_N, -1)$$

# Spin assignment: a simple wave function for the Be atom

(1)

Be atom,  $1s^2 2s^2 \Rightarrow N_{\uparrow} = N_{\downarrow} = 2, S_z = 0$

Determinant of spin-orbitals  $\phi_{1s} \chi_{\uparrow}, \phi_{2s} \chi_{\uparrow}, \phi_{1s} \chi_{\downarrow}, \phi_{2s} \chi_{\downarrow}$

$$D = \frac{1}{\sqrt{4!}} \begin{vmatrix} \phi_{1s}(\mathbf{r}_1)\chi_{\uparrow}(\sigma_1) & \dots & \phi_{1s}(\mathbf{r}_4)\chi_{\uparrow}(\sigma_4) \\ \phi_{2s}(\mathbf{r}_1)\chi_{\uparrow}(\sigma_1) & \dots & \phi_{2s}(\mathbf{r}_4)\chi_{\uparrow}(\sigma_4) \\ \phi_{1s}(\mathbf{r}_1)\chi_{\downarrow}(\sigma_1) & \dots & \phi_{1s}(\mathbf{r}_4)\chi_{\downarrow}(\sigma_4) \\ \phi_{2s}(\mathbf{r}_1)\chi_{\downarrow}(\sigma_1) & \dots & \phi_{2s}(\mathbf{r}_4)\chi_{\downarrow}(\sigma_4) \end{vmatrix}$$

Spin-assigned  $F_1(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3, \mathbf{r}_4) = D(\mathbf{r}_1, +1, \mathbf{r}_2, +1, \mathbf{r}_3, -1, \mathbf{r}_4, -1)$

$$F_1 = \frac{1}{\sqrt{4!}} \begin{vmatrix} \phi_{1s}(\mathbf{r}_1) & \phi_{1s}(\mathbf{r}_2) & 0 & 0 \\ \phi_{2s}(\mathbf{r}_1) & \phi_{2s}(\mathbf{r}_2) & 0 & 0 \\ 0 & 0 & \phi_{1s}(\mathbf{r}_3) & \phi_{1s}(\mathbf{r}_4) \\ 0 & 0 & \phi_{2s}(\mathbf{r}_3) & \phi_{2s}(\mathbf{r}_4) \end{vmatrix}$$



## Spin assignment: a simple wave function for the Be atom

(2)

Be atom,  $1s^2 2s^2 \Rightarrow N_{\uparrow} = N_{\downarrow} = 2, S_z = 0$

$$F_1 = \frac{1}{\sqrt{4!}} \begin{vmatrix} \phi_{1s}(\mathbf{r}_1) & \phi_{1s}(\mathbf{r}_2) & 0 & 0 \\ \phi_{2s}(\mathbf{r}_1) & \phi_{2s}(\mathbf{r}_2) & 0 & 0 \\ 0 & 0 & \phi_{1s}(\mathbf{r}_3) & \phi_{1s}(\mathbf{r}_4) \\ 0 & 0 & \phi_{2s}(\mathbf{r}_3) & \phi_{2s}(\mathbf{r}_4) \end{vmatrix}$$
$$= \frac{1}{\sqrt{4!}} \begin{vmatrix} \phi_{1s}(\mathbf{r}_1) & \phi_{1s}(\mathbf{r}_2) \\ \phi_{2s}(\mathbf{r}_1) & \phi_{2s}(\mathbf{r}_2) \end{vmatrix} \times \begin{vmatrix} \phi_{1s}(\mathbf{r}_3) & \phi_{1s}(\mathbf{r}_4) \\ \phi_{2s}(\mathbf{r}_3) & \phi_{2s}(\mathbf{r}_4) \end{vmatrix}$$

$$D(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4) \rightarrow D^{\uparrow}(\mathbf{r}_1, \mathbf{r}_2) \times D^{\downarrow}(\mathbf{r}_3, \mathbf{r}_4)$$

## Jastrow-Slater spin-assigned wave function

To obtain spin-assigned Jastrow-Slater wave functions, impose

Particle	1	2	...	$N_{\uparrow}$	$N_{\uparrow+1}$	...	$N$
$\sigma$	1	1	...	1	-1	...	-1

$$\Psi(\mathbf{r}_1, \dots, \mathbf{r}_N) = F_1(\mathbf{r}_1, \dots, \mathbf{r}_N)$$

$$= \mathcal{J}(\mathbf{r}_1, \dots, \mathbf{r}_N) \sum_k d_k D_k^{\uparrow}(\mathbf{r}_1, \dots, \mathbf{r}_{N_{\uparrow}}) D_k^{\downarrow}(\mathbf{r}_{N_{\uparrow}+1}, \dots, \mathbf{r}_N)$$

## Jastrow factor and divergences in the potential

At interparticle coalescence points, the potential diverges as

$$-\frac{Z}{r_{i\alpha}} \quad \text{for the electron-nucleus potential}$$

$$\frac{1}{r_{ij}} \quad \text{for the electron-electron potential}$$

Local energy  $\frac{\mathcal{H}\Psi}{\Psi} = -\frac{1}{2} \sum_i \frac{\nabla_i^2 \Psi}{\Psi} + \mathcal{V}$  must be **finite**

⇒ Kinetic energy must have opposite divergence to the potential  $\mathcal{V}$

## Divergence in potential and behavior of the local energy

Consider two particles of masses  $m_i, m_j$  and charges  $q_i, q_j$

Assume  $r_{ij} \rightarrow 0$  while all other particles are well separated

Keep only diverging terms in  $\frac{\mathcal{H}\Psi}{\Psi}$  and go to relative coordinates close to  $\mathbf{r} = \mathbf{r}_{ij} = 0$

$$\begin{aligned} -\frac{1}{2\mu_{ij}} \frac{\nabla^2 \Psi}{\Psi} + \mathcal{V}(r) &\sim -\frac{1}{2\mu_{ij}} \frac{\Psi''}{\Psi} - \frac{1}{\mu_{ij}} \frac{1}{r} \frac{\Psi'}{\Psi} + \mathcal{V}(r) \\ &\sim \boxed{-\frac{1}{\mu_{ij}} \frac{1}{r} \frac{\Psi'}{\Psi} + \mathcal{V}(r)} \end{aligned}$$

where  $\mu_{ij} = m_i m_j / (m_i + m_j)$

## Divergence in potential and cusp conditions

Diverging terms in the local energy

$$-\frac{1}{\mu_{ij}} \frac{1}{r} \frac{\Psi'}{\Psi} + \mathcal{V}(r) = -\frac{1}{\mu_{ij}} \frac{1}{r} \frac{\Psi'}{\Psi} + \frac{q_i q_j}{r} = \text{finite}$$

$\Rightarrow \Psi$  must satisfy Kato's cusp conditions:

$$\left. \frac{\partial \hat{\Psi}}{\partial r_{ij}} \right|_{r_{ij}=0} = \mu_{ij} q_i q_j \Psi(r_{ij} = 0)$$

where  $\hat{\Psi}$  is a spherical average

Note: We assumed  $\Psi(r_{ij} = 0) \neq 0$

## Cusp conditions: example

The condition for the local energy to be finite at  $r = 0$  is

$$\frac{\psi'}{\psi} = \mu_{ij} q_i q_j$$

- Electron-nucleus:  $\mu = 1, q_i = 1, q_j = -Z \Rightarrow$

$$\left. \frac{\psi'}{\psi} \right|_{r=0} = -Z$$

- Electron-electron:  $\mu = \frac{1}{2}, q_i = 1, q_j = 1 \Rightarrow$

$$\left. \frac{\psi'}{\psi} \right|_{r=0} = 1/2$$

## Cusp conditions and QMC wave functions

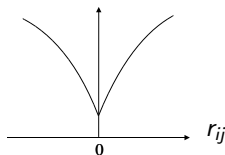
- ▶ Electron-electron cusps imposed through the Jastrow factor

Example: Simple Jastrow factor

$$\mathcal{J}(r_{ij}) = \prod_{i < j} \exp \left\{ b_0 \frac{r_{ij}}{1 + b r_{ij}} \right\}$$

with  $b_0^{\uparrow\downarrow} = \frac{1}{2}$  or  $b_0^{\uparrow\uparrow} = b_0^{\downarrow\downarrow} = \frac{1}{4}$

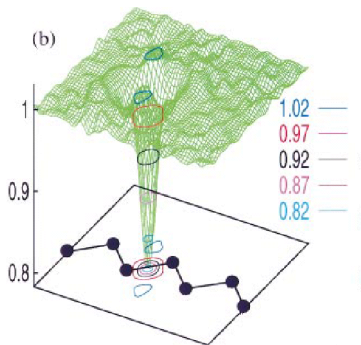
Imposes cusp conditions  
+  
keeps electrons apart



- ▶ Electron-nucleus cusps imposed through the determinantal part

## The effect of the Jastrow factor

Pair correlation function for  $\uparrow\downarrow$  electrons in the (110) plane of Si  
 $g_{\uparrow\downarrow}(\mathbf{r}, \mathbf{r}')$  with one electron is at the bond center



Hood *et al.* Phys. Rev. Lett. **78**, 3350 (1997)



## Simple wave function for the Be atom

Be atom,  $1s^2 2s^2 \Rightarrow N_{\uparrow} = N_{\downarrow} = 2, S_z = 0$

Spin-assigned  $\Psi(\mathbf{r}_1, +1, \mathbf{r}_2, +1, \mathbf{r}_3, -1, \mathbf{r}_4, -1) = \mathcal{J} D$

### ▷ Factorized determinant

$$D = D^{\uparrow} \times D^{\downarrow} = \begin{vmatrix} \phi_{1s}(\mathbf{r}_1) & \phi_{1s}(\mathbf{r}_2) \\ \phi_{2s}(\mathbf{r}_1) & \phi_{2s}(\mathbf{r}_2) \end{vmatrix} \times \begin{vmatrix} \phi_{1s}(\mathbf{r}_3) & \phi_{1s}(\mathbf{r}_4) \\ \phi_{2s}(\mathbf{r}_3) & \phi_{2s}(\mathbf{r}_4) \end{vmatrix}$$

### ▷ Simple Jastrow factor

$$\mathcal{J} = \prod_{ij=13,14,23,24} \exp \left\{ \frac{1}{2} \frac{r_{ij}}{1 + b r_{ij}} \right\} \times \prod_{ij=12,34} \exp \left\{ \frac{1}{4} \frac{r_{ij}}{1 + b r_{ij}} \right\}$$

More general Jastrow form with e-n, e-e and e-e-n terms

$$\prod_{\alpha,i} \exp \{A(r_{i\alpha})\} \prod_{i<j} \exp \{B(r_{ij})\} \prod_{\alpha,i<j} \exp \{C(r_{i\alpha}, r_{j\alpha}, r_{ij})\}$$

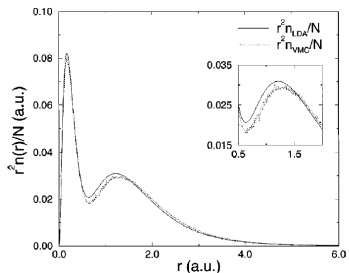
- ▷ Polynomials of scaled variables, e.g.  $\bar{r} = r/(1 + ar)$
- ▷  $\mathcal{J} > 0$  and becomes constant for large  $r_i$ ,  $r_j$  and  $r_{ij}$
- ▷ Electron-electron terms  $B$ 
  - Imposes the cusp conditions and keeps electrons apart
  - More general than simple  $\mathcal{J}(r_{ij})$  gives small improvements
- ▷ Electron-nucleus terms  $A$

Should be included if determinantal part (DFT or HF) is not reoptimized: e-e terms alter the single-particle density

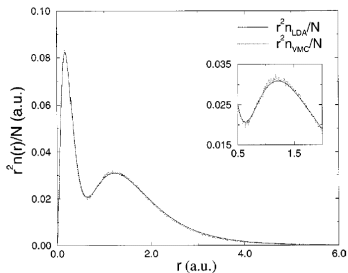
## Role of the electron-nucleus terms

Example: Density of all-electron Carbon atom

DFT determinant + e-e  $\mathcal{J}$



+ e-n  $\mathcal{J}$



Foulkes *et al.* Rev. Mod. Phys. **73**, 33 (2001)

## Some comments on Jastrow factor

(2)

### ▷ Electron-electron-nucleus terms $C$

If the order of the polynomial in the e-e-n terms is infinite,  $\Psi$  can exactly describe a two-electron atom or ion in an  $S$  state

For these systems, a 5<sup>th</sup>-order polynomial recovers more than 99.99% of the correlation energy,  $E_{\text{corr}} = E_{\text{exact}} - E_{\text{HF}}$

### ▷ Is this Jastrow factor adequate for multi-electron systems?

The e-e-n terms are the most important: due to the exclusion principle, it is rare for 3 or more electrons to be close, since at least 2 electrons must necessarily have the same spin

## Jastrow factor with e-e, e-e-n and e-e-e-n terms

	$\mathcal{J}$	$E_{\text{VMC}}$	$E_{\text{VMC}}^{\text{corr}}$ (%)	$\sigma_{\text{VMC}}$	
<b>Li</b>	$E_{\text{HF}}$		-7.43273	0	
		e-e	-7.47427(4)	91.6	0.240
		+ e-e-n	-7.47788(1)	99.6	0.037
		+ e-e-e-n	-7.47797(1)	99.8	0.028
	$E_{\text{exact}}$		-7.47806	100	0
<b>Ne</b>	$E_{\text{HF}}$		-128.5471	0	
		e-e	-128.713(2)	42.5	1.90
		+ e-e-n	-128.9008(1)	90.6	0.90
		+ e-e-e-n	-128.9029(3)	91.1	0.88
	$E_{\text{exact}}$		-128.9376	100	0

Huang, Umrigar, Nightingale, J. Chem. Phys. **107**, 3007 (1997)

## Dynamic and static correlation

$\Psi = \text{Jastrow} \times \text{Determinants} \rightarrow$  Two types of correlation

▷ Dynamic correlation

Described by Jastrow factor

Due to inter-electron repulsion

Always present

▷ Static correlation

Described by a linear combination of determinants

Due to near-degeneracy of occupied and unoccupied orbitals

Not always present

## Static correlation

(1)

Example: Be atom and  $2s$ - $2p$  near-degeneracy

HF ground state configuration  $1s^2 2s^2$

Additional important configuration  $1s^2 2p^2$

Ground state has  $^1S$  symmetry  $\Rightarrow$  4 determinants

$$D = (1s^\uparrow, 2s^\uparrow, 1s^\downarrow, 2s^\downarrow) + c \left[ (1s^\uparrow, 2p_x^\uparrow, 1s^\downarrow, 2p_x^\downarrow) \right. \\ \left. + (1s^\uparrow, 2p_y^\uparrow, 1s^\downarrow, 2p_y^\downarrow) \right. \\ \left. + (1s^\uparrow, 2p_z^\uparrow, 1s^\downarrow, 2p_z^\downarrow) \right]$$

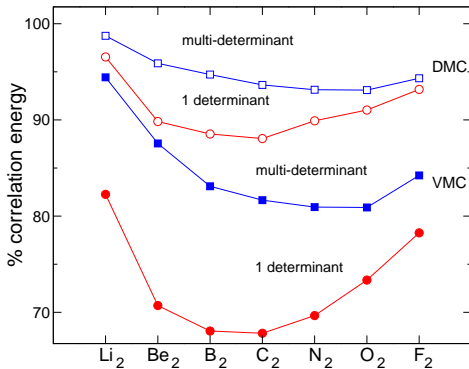
$$1s^2 2s^2 \quad \times \mathcal{J}(r_{ij}) \quad \rightarrow E_{\text{VMC}}^{\text{corr}} = 61\% \\ 1s^2 2s^2 \oplus 1s^2 2p^2 \quad \times \mathcal{J}(r_{ij}) \quad \rightarrow E_{\text{VMC}}^{\text{corr}} = 93\%$$

Example:  $E_{\text{VMC}}^{\text{corr}}$  and  $E_{\text{DMC}}^{\text{corr}}$  for 1<sup>st</sup>-row dimers

MO orbitals with atomic  $s$ - $p$  Slater basis (all-electron)

Active MO orbitals are  $2\sigma_g, 2\sigma_u, 3\sigma_g, 3\sigma_u, 1\pi_u, 1\pi_g$

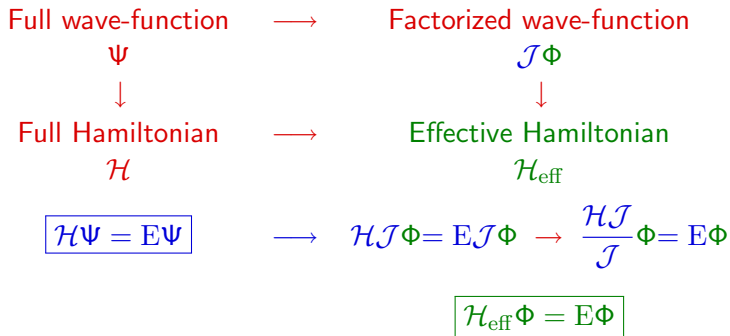
5<sup>th</sup>-order polynomial  $\mathcal{J}$  (e-n, e-e, e-e-n)



Filippi and Umrigar, J. Chem. Phys. **105**, 213 (1996)



Why should  $\Psi_{\text{QMC}} = \mathcal{J}D$  work?



$\mathcal{H}_{\text{eff}}$  weaker Hamiltonian than  $\mathcal{H}$

⇒  $\Phi \approx$  non-interacting wave function  $D$

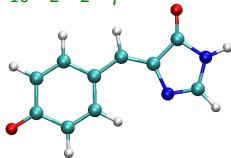
⇒ Quantum Monte Carlo wave function  $\Psi = \mathcal{J}D$

## Construction of the wave function

How do we obtain the parameters in the wave function?

$$\Psi(\mathbf{r}_1, \dots, \mathbf{r}_N) = \mathcal{J} \sum_k d_k D_k^\uparrow D_k^\downarrow$$

$\text{C}_{10}\text{N}_2\text{O}_2\text{H}_7^-$



70 electrons and 21 atoms

VTZ *s-p* basis + 1 polarization

3 *s* + 3 *p* + 1 *d* functions for C, N, O

2 *s* + 1 *p* for H

- ▷ Parameters in the Jastrow factor  $\mathcal{J}$  ( $\approx 100$ )
- ▷ CI coefficients  $d_k$  ( $< 10 - 100$ )
- ▷ Linear coefficients in expansion of the orbitals (5540 !)

## Optimization of trial wave function

How do we find the **best** parameters in  $\Psi = \mathcal{J}\Phi$  ?

**First thought** Let us minimize the energy!

$$E_V = \frac{\int d\mathbf{R} \Psi^*(\mathbf{R}) \mathcal{H} \Psi(\mathbf{R})}{\int d\mathbf{R} \Psi^*(\mathbf{R}) \Psi(\mathbf{R})} = \int d\mathbf{R} \frac{\mathcal{H} \Psi(\mathbf{R})}{\Psi(\mathbf{R})} \frac{|\Psi(\mathbf{R})|^2}{\int d\mathbf{R} |\Psi(\mathbf{R})|^2} = \left\langle \frac{\mathcal{H} \Psi(\mathbf{R})}{\Psi(\mathbf{R})} \right\rangle_{\Psi^2}$$

Straightforward minimization on finite MC sample will **not** work!

## Why problems with straightforward energy minimization ?

Let us write the energy on a finite MC sample

Sample  $N_{\text{conf}}$  configurations from  $|\Psi(\mathbf{R}, \{\alpha_0\})|^2$  with Metropolis

Energy of  $\Psi(\mathbf{R}, \{\alpha\})$  on this set of MC configurations

$$E[\alpha] = \frac{1}{N_{\text{conf}}} \sum_{i=1}^{N_{\text{conf}}} \frac{\mathcal{H}\Psi(\mathbf{R}_i, \{\alpha\})}{\Psi(\mathbf{R}_i, \{\alpha\})} w_i$$

where

$$w_i = \left| \frac{\Psi(\mathbf{R}_i, \{\alpha\})}{\Psi(\mathbf{R}_i, \{\alpha_0\})} \right|^2 \bigg/ \sum_{i=1}^{N_{\text{conf}}} \left| \frac{\Psi(\mathbf{R}, \{\alpha\})}{\Psi(\mathbf{R}, \{\alpha_0\})} \right|^2$$

$E[\alpha]$  on a finite MC sample is not bounded from below

$\Rightarrow$  Straightforward minimization of  $E[\alpha]$  does not work

## Is variance minimization an alternative?

Minimize the variance of the local energy

$$\sigma^2 = \frac{\langle \Psi | (\mathcal{H} - E_V)^2 | \Psi \rangle}{\langle \Psi | \Psi \rangle} = \langle (E_L(\mathbf{R}) - E_V)^2 \rangle_{\Psi^2}$$

Would this work?

Consider variance on a finite number of MC configurations

$$\sigma^2[\alpha] = \sum_{i=1}^{N_{\text{conf}}} \left( \frac{\mathcal{H}\Psi(\mathbf{R}_i, \{\alpha\})}{\Psi(\mathbf{R}_i, \{\alpha\})} - \bar{E} \right)^2 w_i$$

$\sigma^2$  has a known lower bound  $\sigma^2 = 0$

Robust and stable optimization for very small values of  $N_{\text{conf}}$

Energy minimization to optimize the trial wave function

How do we find the **best** parameters in  $\Psi = \mathcal{J}\Phi$  ?

But it seems simple !?!

Let us compute gradient + Hessian of the energy in VMC

## What is so difficult about wave function optimization?

Statistical error: Both a blessing and a curse!

$\Psi_{\{\alpha_k\}}$  → Energy and its derivatives wrt parameters  $\{\alpha_k\}$

$$E_V = \int d\mathbf{R} \frac{\mathcal{H}\Psi(\mathbf{R})}{\Psi(\mathbf{R})} \frac{|\Psi(\mathbf{R})|^2}{\int d\mathbf{R} |\Psi(\mathbf{R})|^2} = \langle E_L \rangle_{\Psi^2}$$

$$\partial_k E_V = \left\langle \frac{\partial_k \Psi}{\Psi} E_L + \frac{\mathcal{H} \partial_k \Psi}{\Psi} - 2E_V \frac{\partial_k \Psi}{\Psi} \right\rangle_{\Psi^2}$$

$$= 2 \left\langle \frac{\partial_k \Psi}{\Psi} (E_L - E_V) \right\rangle_{\Psi^2}$$

The last expression is obtained using Hermiticity of  $\mathcal{H}$

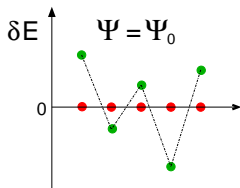
Use gradient/Hessian expressions with smaller fluctuations

Two mathematically equivalent expressions of the energy gradient

$$\partial_k E_V = \left\langle \frac{\partial_k \Psi}{\Psi} E_L + \frac{\mathcal{H} \partial_k \Psi}{\Psi} - 2E_V \frac{\partial_k \Psi}{\Psi} \right\rangle_{\Psi^2} = 2 \left\langle \frac{\partial_k \Psi}{\Psi} (E_L - E_V) \right\rangle_{\Psi^2}$$

Why using the last expression?

Lower fluctuations  $\rightarrow 0$  as  $\Psi \rightarrow \Psi_0$



If you play similar tricks with the Hessian as with the gradient

$\rightarrow$  5 orders of magnitude efficiency gain wrt using original Hessian

C. Umrigar and C. Filippi, PRL **94**, 150201 (2005)



## Energy minimization is possible: Three most successful methods

- Newton method (Umrigar and Filippi, 2005)

$$E(\alpha) \approx E(\alpha^0) + \sum_i \frac{\partial E(\alpha^0)}{\partial \alpha_i} \Delta \alpha_i + \frac{1}{2} \sum_{ij} \frac{\partial^2 E(\alpha^0)}{\partial \alpha_i \partial \alpha_j} \Delta \alpha_i \Delta \alpha_j$$

- Linear method (Umrigar, Toulouse, Filippi, and Sorella, 2007)

$$\Psi(\alpha) \approx \Psi(\alpha^0) + \sum_i \frac{\partial \Psi(\alpha^0)}{\partial \alpha_i} \Delta \alpha_i$$

Solution of  $\mathcal{H} \Delta \alpha = E S \Delta \alpha$  in the basis of  $\{\Psi(\alpha^0), \frac{\partial \Psi(\alpha^0)}{\partial \alpha_i}\}$

- Perturbative approach (Scemama and Filippi, 2006)

## Customary practice for optimizing wave function

### Jastrow-Slater wave function

$$\Psi(\mathbf{r}_1, \dots, \mathbf{r}_N) = \mathcal{J} \sum_k d_k D_k^\uparrow D_k^\downarrow$$

- ▷ Jastrow factor optimized in variance/energy minimization
- ▷ Orbitals and  $d_k$  coefficients in determinantal part are from
  - Hartree-Fock or DFT (LDA, GGA, B3LYP ...)
  - CI or multi-configuration self-consistent-field calculation
  - Optimized in energy minimization (very simple for  $d_k$ )

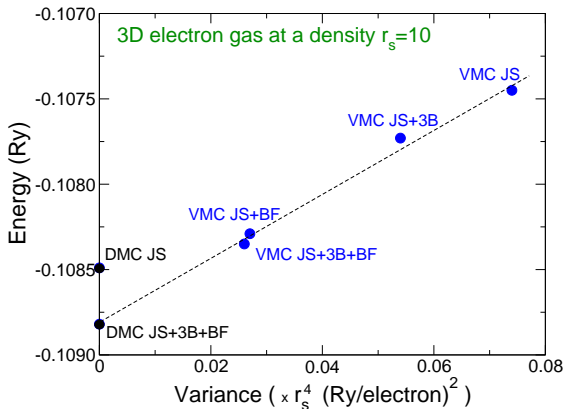
## Beyond VMC?

Removing or reducing wave function bias?

⇒ Projection Monte Carlo methods

## Why going beyond VMC?

Dependence of VMC from wave function  $\Psi$



Kwon, Ceperley, Martin, Phys. Rev. B **58**, 6800 (1998)

## Why going beyond VMC?

- ▷ Dependence on wave function: What goes in, comes out!
- ▷ No automatic way of constructing wave function  $\Psi$   
Choices must be made about functional form (human time)
- ▷ Hard to ensure good error cancelation on energy differences  
e.g. easier to construct good  $\Psi$  for closed than open shells

Can we remove wave function bias?

## Projector Monte Carlo methods

- ▷ Construct an operator which inverts spectrum of  $\mathcal{H}$
- ▷ Use it to stochastically project the ground state of  $\mathcal{H}$

Diffusion Monte Carlo

$$\exp[-\tau(\mathcal{H} - E_T)]$$

Green's function Monte Carlo

$$1/(\mathcal{H} - E_T)$$

Power Monte Carlo

$$E_T - \mathcal{H}$$

## Diffusion Monte Carlo

Consider initial guess  $\Psi^{(0)}$  and repeatedly apply projection operator

$$\Psi^{(n)} = e^{-\tau(\mathcal{H}-E_T)}\Psi^{(n-1)}$$

Expand  $\Psi^{(0)}$  on the eigenstates  $\Psi_i$  with energies  $E_i$  of  $\mathcal{H}$

$$\Psi^{(n)} = e^{-n\tau(\mathcal{H}-E_T)}\Psi^{(0)} = \sum_i \Psi_i \langle \Psi^{(0)} | \Psi_i \rangle e^{-n\tau(E_i-E_T)}$$

and obtain in the limit of  $n \rightarrow \infty$

$$\lim_{n \rightarrow \infty} \Psi^{(n)} = \Psi_0 \langle \Psi^{(0)} | \Psi_0 \rangle e^{-n\tau(E_0-E_T)}$$

If we choose  $E_T \approx E_0$ , we obtain

$$\lim_{n \rightarrow \infty} \Psi^{(n)} = \Psi_0$$

How do we perform the projection?

Rewrite projection equation in integral form

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

where  $G(\mathbf{R}', \mathbf{R}, \tau) = \langle \mathbf{R}' | e^{-\tau(\mathcal{H} - E_T)} | \mathbf{R} \rangle$

▷ Can we sample the wave function?

For the moment, assume we are dealing with **bosons**, so  $\Psi > 0$

▷ Can we interpret  $G(\mathbf{R}', \mathbf{R}, \tau)$  as a transition probability?

If yes, we can perform this integral by Monte Carlo integration



## VMC and DMC as power methods

VMC Distribution function is given  $\rho(\mathbf{R}) = \frac{|\Psi(\mathbf{R})|^2}{\int d\mathbf{R} |\Psi(\mathbf{R})|^2}$

Construct  $P$  which satisfies stationarity condition  $P\rho = \rho$

→  $\rho$  is eigenvector of  $P$  with eigenvalue 1

→  $\rho$  is the dominant eigenvector  $\Rightarrow \lim_{n \rightarrow \infty} P^n \rho_{\text{initial}} = \rho$

DMC Opposite procedure!

The matrix  $P$  is given  $\rightarrow P = \langle \mathbf{R}' | e^{-\tau(\mathcal{H} - E_T)} | \mathbf{R} \rangle$

We want to find the dominant eigenvector  $\rho = \Psi_0$

What can we say about the Green's function?

$$G(\mathbf{R}', \mathbf{R}, \tau) = \langle \mathbf{R}' | e^{-\tau(\mathcal{H} - E_T)} | \mathbf{R} \rangle$$

$G(\mathbf{R}', \mathbf{R}, \tau)$  satisfies the imaginary-time Schrödinger equation

$$(\mathcal{H} - E_T)G(\mathbf{R}, \mathbf{R}_0, t) = -\frac{\partial G(\mathbf{R}, \mathbf{R}_0, t)}{\partial t}$$

with  $G(\mathbf{R}', \mathbf{R}, 0) = \delta(\mathbf{R}' - \mathbf{R})$

Can we interpret  $G(\mathbf{R}', \mathbf{R}, \tau)$  as a transition probability?

(1)

$$\mathcal{H} = \mathcal{T}$$

Imaginary-time Schrödinger equation is a diffusion equation

$$-\frac{1}{2}\nabla^2 G(\mathbf{R}, \mathbf{R}_0, t) = -\frac{\partial G(\mathbf{R}, \mathbf{R}_0, t)}{\partial t}$$

The Green's function is given by a Gaussian

$$G(\mathbf{R}', \mathbf{R}, \tau) = (2\pi\tau)^{-3N/2} \exp\left[-\frac{(\mathbf{R}' - \mathbf{R})^2}{2\tau}\right]$$

Positive and can be sampled

Can we interpret  $G(\mathbf{R}', \mathbf{R}, \tau)$  as a transition probability? (2)

$$\mathcal{H} = \mathcal{V}$$

$$(\mathcal{V}(\mathbf{R}) - E_T)G(\mathbf{R}, \mathbf{R}_0, t) = -\frac{\partial G(\mathbf{R}, \mathbf{R}_0, t)}{\partial t},$$

The Green's function is given by

$$G(\mathbf{R}', \mathbf{R}, \tau) = \exp[-\tau(\mathcal{V}(\mathbf{R}) - E_T)] \delta(\mathbf{R} - \mathbf{R}'),$$

Positive but does not preserve the normalization

It is a factor by which we multiply the distribution  $\Psi(\mathbf{R}, t)$

$\mathcal{H} = \mathcal{T} + \mathcal{V}$  and a combination of diffusion and branching

Trotter's theorem  $\rightarrow$   $e^{(A+B)\tau} = e^{A\tau} e^{B\tau} + \mathcal{O}(\tau^2)$

$$\begin{aligned}\langle \mathbf{R}' | e^{-\mathcal{H}\tau} | \mathbf{R}_0 \rangle &\approx \langle \mathbf{R}' | e^{-\mathcal{T}\tau} e^{-\mathcal{V}\tau} | \mathbf{R}_0 \rangle \\ &= \int d\mathbf{R}'' \langle \mathbf{R}' | e^{-\mathcal{T}\tau} | \mathbf{R}'' \rangle \langle \mathbf{R}'' | e^{-\mathcal{V}\tau} | \mathbf{R}_0 \rangle \\ &= \langle \mathbf{R}' | e^{-\mathcal{T}\tau} | \mathbf{R}_0 \rangle e^{-\mathcal{V}(\mathbf{R}_0)\tau}\end{aligned}$$

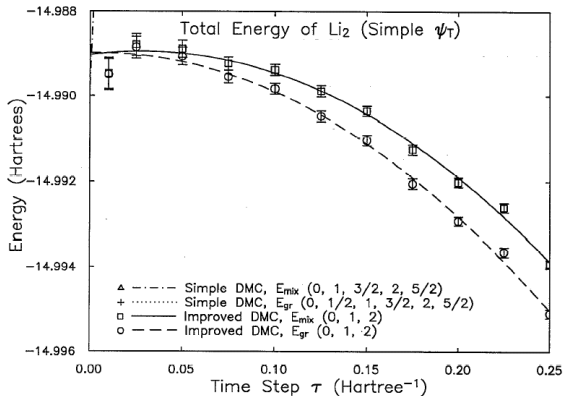
The Green's function in the short-time approximation to  $\mathcal{O}(\tau^2)$  is

$$G(\mathbf{R}', \mathbf{R}, \tau) = (2\pi\tau)^{-3N/2} \exp\left[-\frac{(\mathbf{R}' - \mathbf{R})^2}{2\tau}\right] \exp[-\tau(\mathcal{V}(\mathbf{R}) - E_T)]$$

DMC results must be extrapolated at short time-steps ( $\tau \rightarrow 0$ )

## Time-step extrapolation

Example: Energy of  $\text{Li}_2$  versus time-step  $\tau$



Umrigar, Nightingale, Runge, J. Chem. Phys. **94**, 2865 (1993)

The basic DMC algorithm is rather simple:

1. Sample  $\Psi^{(0)}(\mathbf{R})$  with the Metropolis algorithm

Generate  $M_0$  walkers  $\mathbf{R}_1, \dots, \mathbf{R}_{M_0}$  (zeroth generation)

2. Diffuse each walker as  $\mathbf{R}' = \mathbf{R} + \xi$

where  $\xi$  is sampled from  $g(\xi) = (2\pi\tau)^{-3N/2} \exp(-\xi^2/2\tau)$

3. For each walker, compute the factor

$$p = \exp[-\tau(\mathcal{V}(\mathbf{R}) - E_T)]$$

Branch the walker with  $p$  the probability to survive

Continue →

4. Branch the walker with  $p$  the probability to survive

▷ If  $p < 1$ , the walker survives with probability  $p$

▷ If  $p > 1$ , the walker continues and new walkers with the same coordinates are created with probability  $p - 1$

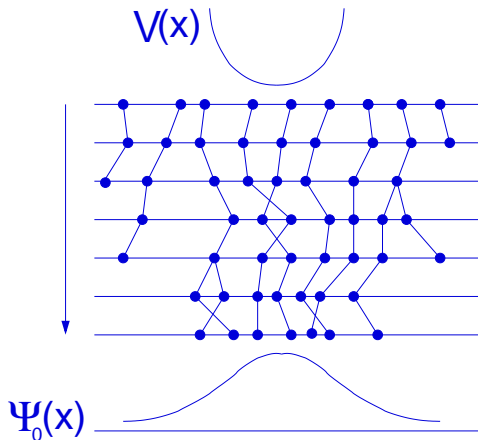
⇒ Number of copies of the current walker equal to  $\text{int}(p + \eta)$   
where  $\eta$  is a random number between  $(0,1)$

5. Adjust  $E_T$  so that population fluctuates around target  $M_0$

→ After many iterations, walkers distributed as  $\Psi_0(\mathbf{R})$



## Diffusion and branching in a harmonic potential



Walkers proliferate/die in regions of lower/higher potential than  $E_T$

## Problems with simple algorithm

The simple algorithm is inefficient and unstable

- ▷ Potential can vary a lot and be unbounded  
e.g. electron-nucleus interaction → Exploding population
- ▷ Branching factor grows with system size

## Importance sampling

Start from integral equation

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

Multiply each side by trial  $\Psi$  and define  $f(\mathbf{R}, t) = \Psi(\mathbf{R})\Psi(\mathbf{R}, t)$

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

where the importance sampled Green's function is

$$\tilde{G}(\mathbf{R}', \mathbf{R}, \tau) = \Psi(\mathbf{R}') \langle \mathbf{R}' | e^{-\tau(\mathcal{H} - E_T)} | \mathbf{R} \rangle / \Psi(\mathbf{R})$$

We obtain  $\lim_{n \rightarrow \infty} f(\mathbf{R}) = \Psi(\mathbf{R})\Psi_0(\mathbf{R})$

## Importance sampled Green's function

The importance sampled  $\tilde{G}(\mathbf{R}, \mathbf{R}_0, \tau)$  satisfies

$$-\frac{1}{2}\nabla^2\tilde{G} + \nabla \cdot [\tilde{G}\mathbf{V}(\mathbf{R})] + [E_L(\mathbf{R}) - E_T]\tilde{G} = -\frac{\partial\tilde{G}}{\partial\tau}$$

with the quantum velocity  $\mathbf{V}(\mathbf{R}) = \frac{\nabla\Psi(\mathbf{R})}{\Psi(\mathbf{R})}$

We now have **drift** in addition to diffusion and branching terms

Trotter's theorem  $\Rightarrow$  Consider them separately for small enough  $\tau$

## The drift-diffusion-branching Green's function

Drift-diffusion-branching short-time Green's function is

$$\begin{aligned} \tilde{G}(\mathbf{R}', \mathbf{R}, \tau) &= (2\pi\tau)^{-3N/2} \exp \left[ -\frac{(\mathbf{R}' - \mathbf{R} - \tau\mathbf{V}(\mathbf{R}))^2}{2\tau} \right] \times \\ &\times \exp \left\{ -\tau \left[ (E_L(\mathbf{R}) + E_L(\mathbf{R}'))/2 - E_T \right] \right\} + \mathcal{O}(\tau^2) \end{aligned}$$

What is new in the drift-diffusion-branching expression?

- ▷  $\mathbf{V}(\mathbf{R})$  pushes walkers where  $\Psi$  is large
- ▷  $E_L(\mathbf{R})$  is better behaved than the potential  $\mathcal{V}(\mathbf{R})$   
Cusp conditions  $\Rightarrow$  No divergences when particles approach  
As  $\Psi \rightarrow \Psi_0$ ,  $E_L \rightarrow E_0$  and branching factor is smaller

## DMC algorithm with importance sampling

1. Sample initial walkers from  $|\Psi(\mathbf{R})|^2$
2. Drift and diffuse the walkers as  $\mathbf{R}' = \mathbf{R} + \tau\mathbf{V}(\mathbf{R}) + \xi$   
where  $\xi$  is sampled from  $g(\xi) = (2\pi\tau)^{-3N/2} \exp(-\xi^2/2\tau)$
3. Branching step as in the simple algorithm but with the factor

$$\rho = \exp\{-\tau[(E_L(\mathbf{R}) + E_L(\mathbf{R}'))/2 - E_T]\}$$

4. Adjust the trial energy to keep the population stable

→ After many iterations, walkers distributed as  $\Psi(\mathbf{R})\Psi_0(\mathbf{R})$

## Evolution equation of the probability distribution

$$\triangleright \quad \Psi(\mathbf{R}', t + \tau) = \int d\mathbf{R} G(\mathbf{R}', \mathbf{R}, \tau) \Psi(\mathbf{R}, t)$$

where  $G(\mathbf{R}', \mathbf{R}, \tau) = \langle \mathbf{R}' | e^{-\tau(\mathcal{H} - E_T)} | \mathbf{R} \rangle$

$$(\mathcal{H} - E_T)G(\mathbf{R}, \mathbf{R}_0, t) = -\frac{\partial G(\mathbf{R}, \mathbf{R}_0, t)}{\partial t}$$

$$\triangleright \quad \Psi(\mathbf{R}, t) = \int d\mathbf{R}_0 G(\mathbf{R}, \mathbf{R}_0, t) \Psi^{(0)}(\mathbf{R}_0)$$

satisfies the imaginary-time Schrödinger equation

$$(\mathcal{H} - E_T)\Psi(\mathbf{R}, t) = -\frac{\partial \Psi(\mathbf{R}, t)}{\partial t}$$

Electrons are fermions!

We assumed that  $\Psi_0 > 0$  and that we are dealing with bosons

Fermions  $\rightarrow \Psi$  is antisymmetric and changes sign!

Fermion Sign Problem

All fermion QMC methods suffer from sign problems

These sign problems look different but have the same “flavour”

Arise when you treat something non-positive as probability density



## The DMC Sign Problem

How can we impose antisymmetry in simple DMC method?

**Idea** Evolve separate positive and negative populations of walkers

### Simple 1D example

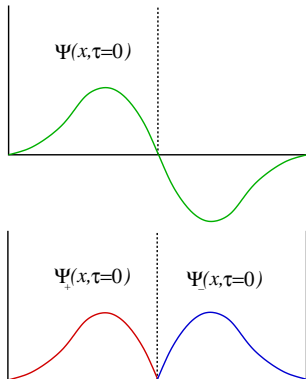
Rewrite  $\Psi(x, \tau = 0)$  as

$$\Psi = \Psi_+ - \Psi_-$$

where

$$\Psi_+ = \frac{1}{2}(|\Psi| + \Psi)$$

$$\Psi_- = \frac{1}{2}(|\Psi| - \Psi)$$



## Particle in a box and the fermionic problem

(1)

The imaginary-time Schrödinger equation

$$\mathcal{H}\Psi = -\frac{\partial\Psi}{\partial t}$$

is linear, so solving it with the initial condition

$$\Psi(x, t = 0) = \Psi_+(x, t = 0) - \Psi_-(x, t = 0)$$

is equivalent to solving

$$\mathcal{H}\Psi_+ = -\frac{\partial\Psi_+}{\partial t}$$

and

$$\mathcal{H}\Psi_- = -\frac{\partial\Psi_-}{\partial t}$$

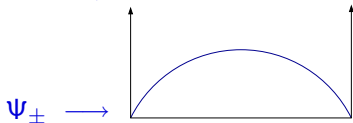
separately and subtracting one solution from the other

- ▷ Expand  $\Psi_{\pm}$  at  $t = 0$  in eigenfunctions

$$\text{At } t = 0, \quad \Psi_{\pm}(t = 0) = c_0^s \Psi_0^s \pm c_0^a \Psi_0^a + \dots$$

$$\text{As } t \rightarrow \infty, \quad \Psi_{\pm}(t) \rightarrow c_0^s e^{-E_0^s t} \Psi_0^s \pm c_0^a e^{-E_0^a t} \Psi_0^a + \dots$$

- ▷ Since  $E_0^s < E_0^a$ , both  $\Psi_+$  and  $\Psi_-$  evolve to  $\Psi_0^s$



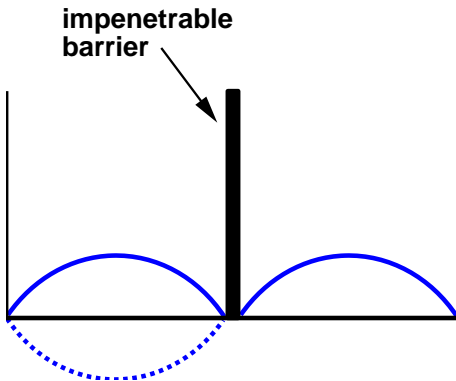
- ▷ Antisymmetric component exponentially harder to extract

$$\frac{|\Psi_+ - \Psi_-|}{|\Psi_+ + \Psi_-|} \propto \frac{e^{-E_0^a t}}{e^{-E_0^s t}} \quad \text{as } t \rightarrow \infty$$

## The Fixed-Node Approximation

**Problem** Small antisymmetric part swamped by random errors

**Solution** Fix the nodes! (If you don't know them, guess them)



## Fixed-node algorithm in simple DMC

(1)

How do we impose that additional boundary condition?

- ▷ Distribute walkers according to any positive initial  $\Psi^{(0)}$
- ▷ Evolve according to imaginary-time Schrödinger equation
- ▷ Annihilate walkers that bump into barrier (and into walls)
  - This step enforces  $\Psi = 0$  boundary conditions

In each nodal pocket, evolution to ground state in that pocket

## Fixed-node algorithm in simple DMC

(2)

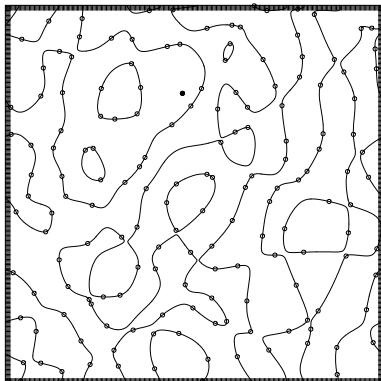
Numerically stable algorithm (no exponentially growing noise)

- Solution is exact if nodes are exact
- The computed energy is variational if nodes approximate
- Best solution consistent with the assumed nodes

For many electrons, what are the nodes? A complex beast

Many-electron wave function  $\Psi(\mathbf{R}) = \Psi(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N)$

Node  $\rightarrow$  surface where  $\Psi = 0$  and across which  $\Psi$  changes sign



A 2D slice through the 321-dimensional nodal surface  
of a gas of 161 spin-up electrons.

## Some known properties of the nodes

Physical space has  $d$  ( $=1,2,3$ ) dimensions

- ▶ Node is  $(dN - 1)$ -dimensional surface in  $dN$  dimensions

One constraint ( $\Psi = 0$ )  $\Rightarrow$   $(dN - 1)$ -dimensional node

- ▶ Equations as  $\mathbf{r}_i = \mathbf{r}_j$  define  $(dN - d)$ -dimensional coincidence surfaces and do not define the node completely if  $d > 1$
- ▶ If  $d = 1$ , coincidence points  $x_i = x_j$  define the ground-state node completely  $\rightarrow$  One-dim problems are easy to simulate

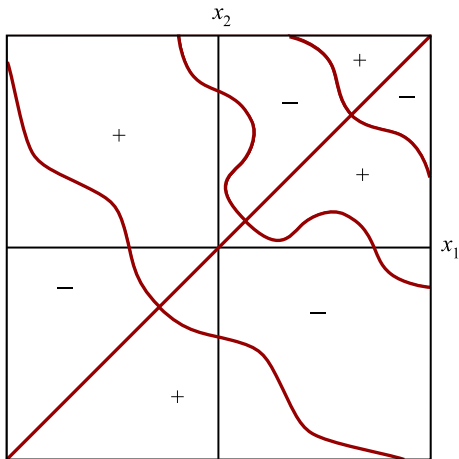


## Nodal pockets can be divided up into classes

Start from  $\mathbf{R}_0$  and continuously reach all points with  $\Psi(\mathbf{R}) \neq 0$

$\Rightarrow$  Nodal pocket accessible from  $\mathbf{R}_0$

Map this subvolume over rest of the space with permutations

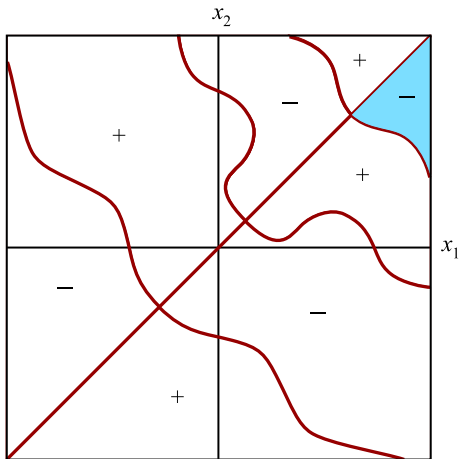


## Nodal pockets can be divided up into classes

Start from  $\mathbf{R}_0$  and continuously reach all points with  $\Psi(\mathbf{R}) \neq 0$

$\Rightarrow$  Nodal pocket accessible from  $\mathbf{R}_0$

Map this subvolume over rest of the space with permutations

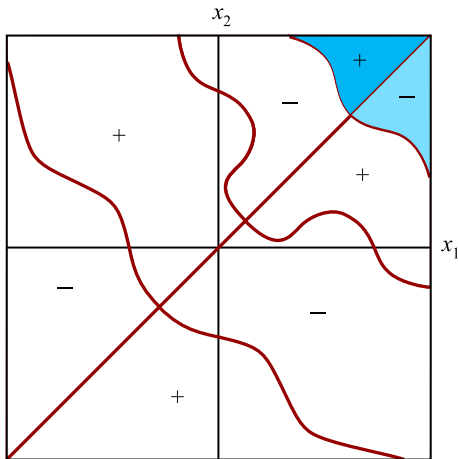


Nodal pockets can be divided up into classes

Start from  $\mathbf{R}_0$  and continuously reach all points with  $\Psi(\mathbf{R}) \neq 0$

$\Rightarrow$  Nodal pocket accessible from  $\mathbf{R}_0$

Map this subvolume over rest of the space with permutations

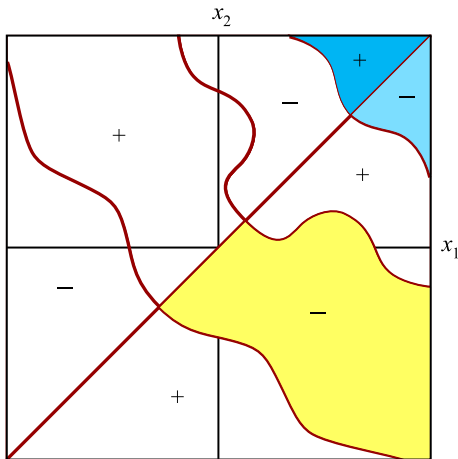


## Nodal pockets can be divided up into classes

Start from  $\mathbf{R}_0$  and continuously reach all points with  $\Psi(\mathbf{R}) \neq 0$

$\Rightarrow$  Nodal pocket accessible from  $\mathbf{R}_0$

Map this subvolume over rest of the space with permutations

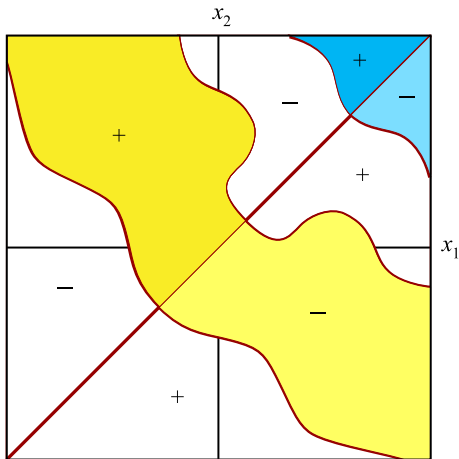


## Nodal pockets can be divided up into classes

Start from  $\mathbf{R}_0$  and continuously reach all points with  $\Psi(\mathbf{R}) \neq 0$

$\Rightarrow$  Nodal pocket accessible from  $\mathbf{R}_0$

Map this subvolume over rest of the space with permutations

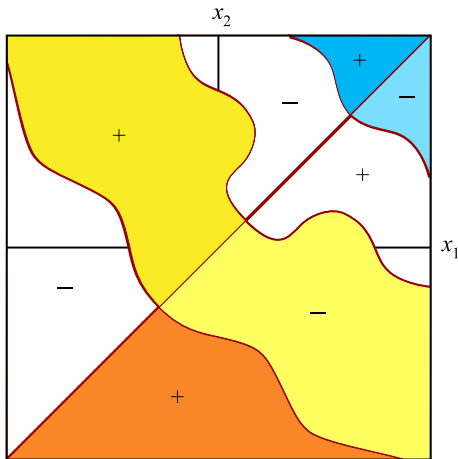


## Nodal pockets can be divided up into classes

Start from  $\mathbf{R}_0$  and continuously reach all points with  $\Psi(\mathbf{R}) \neq 0$

$\Rightarrow$  Nodal pocket accessible from  $\mathbf{R}_0$

Map this subvolume over rest of the space with permutations

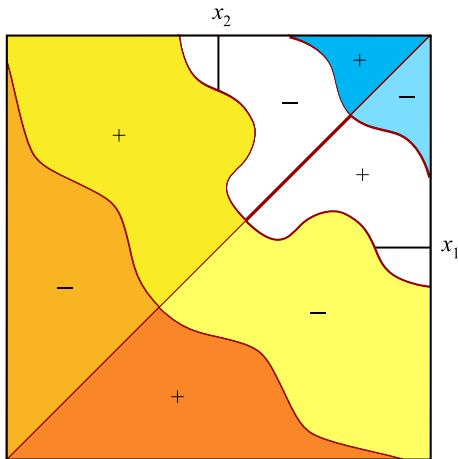


## Nodal pockets can be divided up into classes

Start from  $\mathbf{R}_0$  and continuously reach all points with  $\Psi(\mathbf{R}) \neq 0$

$\Rightarrow$  Nodal pocket accessible from  $\mathbf{R}_0$

Map this subvolume over rest of the space with permutations

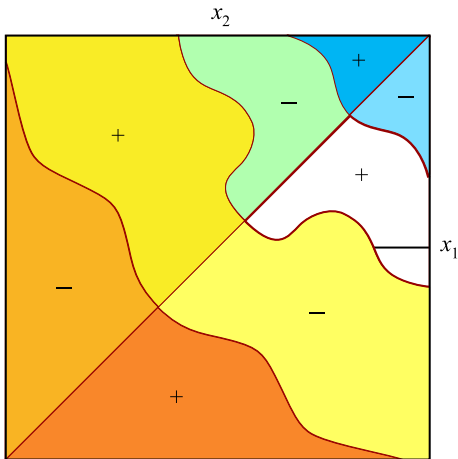


## Nodal pockets can be divided up into classes

Start from  $\mathbf{R}_0$  and continuously reach all points with  $\Psi(\mathbf{R}) \neq 0$

$\Rightarrow$  Nodal pocket accessible from  $\mathbf{R}_0$

Map this subvolume over rest of the space with permutations



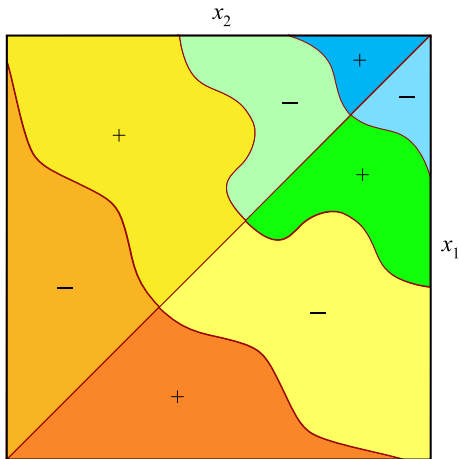


## Nodal pockets can be divided up into classes

Start from  $\mathbf{R}_0$  and continuously reach all points with  $\Psi(\mathbf{R}) \neq 0$

$\Rightarrow$  Nodal pocket accessible from  $\mathbf{R}_0$

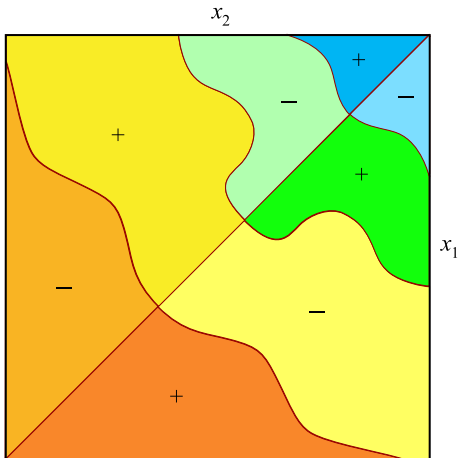
Map this subvolume over rest of the space with permutations



## The Tiling Theorem

Consider Hamiltonian with a local potential

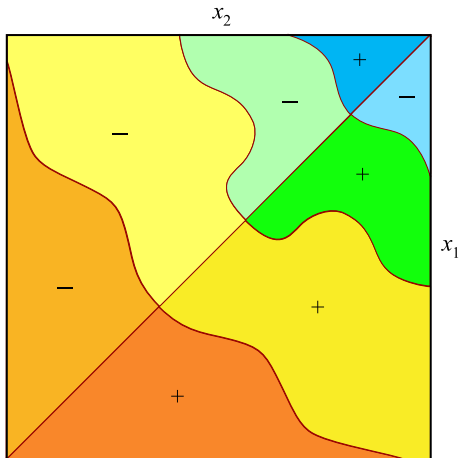
For ground-state wavefunction, all pockets are in the same class



## The Tiling Theorem

Consider Hamiltonian with a local potential

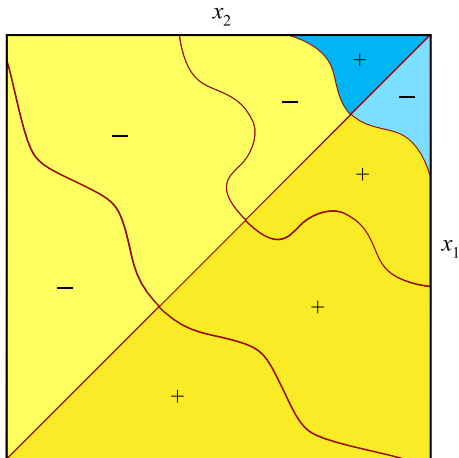
For ground-state wavefunction, all pockets are in the same class



## The Tiling Theorem

Consider Hamiltonian with a local potential

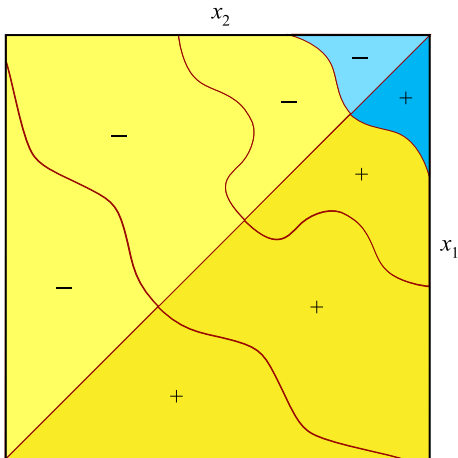
For ground-state wavefunction, all pockets are in the same class



## The Tiling Theorem

Consider Hamiltonian with a local potential

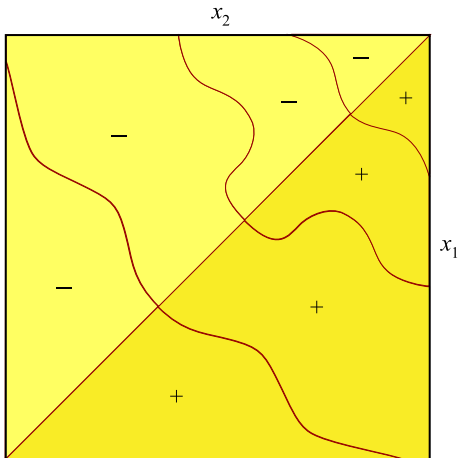
For ground-state wavefunction, all pockets are in the same class



## The Tiling Theorem

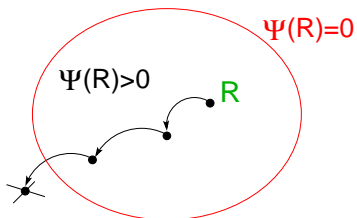
Consider Hamiltonian with a local potential

For ground-state wavefunction, all pockets are in the same class



Use the nodes of trial  $\Psi \rightarrow$  Fixed-node approximation

Use the nodes of the best available trial  $\Psi$  wave function



Find best solution with same nodes as trial wave function  $\Psi$

Fixed-node solution exact if the nodes of trial  $\Psi$  are exact

## Fixed-node solution and importance-sampling DMC

Given trial  $\Psi(\mathbf{R})$ , evolve  $f(\mathbf{R}, t) = \Psi(\mathbf{R})\psi(\mathbf{R}, t)$  as

$$-\frac{1}{2}\nabla^2 f + \nabla \cdot [f \mathbf{V}(\mathbf{R})] + [E_L(\mathbf{R}) - E_T] f = -\frac{\partial f}{\partial \tau}$$

with  $\mathbf{V}(\mathbf{R}) = \frac{\nabla \Psi(\mathbf{R})}{\Psi(\mathbf{R})}$  and  $E_L(\mathbf{R}) = \frac{\mathcal{H}\Psi(\mathbf{R})}{\Psi(\mathbf{R})}$

Fixed-node approximation  $\rightarrow f(\mathbf{R}, t) \geq 0$



## Fixed-node solution and behavior at the nodes

Within the nodes  $\mathcal{H}\Psi_{\text{FN}}(\mathbf{R}) = E_{\text{FN}}\Psi_{\text{FN}}(\mathbf{R})$

If the nodes not exact  $\rightarrow \Psi_{\text{FN}} \neq \Psi_0$

If the nodes not exact  $\rightarrow$  Discontinuity of derivatives at the nodes

$$\mathcal{H}\Psi_{\text{FN}}(\mathbf{R}) = E_{\text{FN}}\Psi_{\text{FN}}(\mathbf{R}) + \delta \quad \text{for } \mathbf{R} \in \delta\Omega$$

Note that the  $\delta$  function does not affect the computed energy

$$\int \Psi_{\text{FN}}\mathcal{H}\Psi_{\text{FN}} = \int \Psi_{\text{FN}}(E_{\text{FN}}\Psi_{\text{FN}} + \delta) = \int \Psi_{\text{FN}}E_{\text{FN}}\Psi_{\text{FN}} = E_{\text{FN}}$$

Fixed-node solution is an upper bound to exact energy

In a nodal pocket  $\Omega$  of the trial wave function  $\Psi$

$$\mathcal{H}\Psi_{\text{FN}}(\mathbf{R}) = E_{\text{FN}}\Psi_{\text{FN}}(\mathbf{R}) \quad \mathbf{R} \in \Omega$$

with  $\Psi_{\text{FN}}(\mathbf{R}) = 0$  for  $\mathbf{R} \notin \Omega \rightarrow$  Extend solution over all space

$$\tilde{\Psi}_{\text{FN}}(\mathbf{R}) = \frac{1}{N!} \sum_{\mathbf{P}} (-1)^{\mathbf{P}} \Psi_{\text{FN}}(\mathbf{P}\mathbf{R})$$

which satisfies

$$\frac{\int d\mathbf{R} \tilde{\Psi}_{\text{FN}}^*(\mathbf{R}) \mathcal{H} \tilde{\Psi}_{\text{FN}}(\mathbf{R})}{\int d\mathbf{R} \tilde{\Psi}_{\text{FN}}^*(\mathbf{R}) \tilde{\Psi}_{\text{FN}}(\mathbf{R})} = E_{\text{FN}} \geq E_0$$

Have we solved all our problems?

Results depend on the nodes of the trial wave function  $\Psi$

How well are we doing with a simple  $\Psi$ ?

One determinant of natural orbitals, 6-311++G(2d,2p) basis

$\epsilon_{\text{MAD}}$  for atomization energy of the 55 molecules of the G1 set

	DMC	CCSD(T)/aug-cc-pVQZ
$\epsilon_{\text{MAD}}$	2.9	2.8 kcal/mol

Grossman, J. Chem. Phys. **117**, 1434 (2002)

We are doing very well without much effort on  $\Psi$  !

## Diffusion Monte Carlo as a black-box approach?

Not always but with some more effort ... QMC can do better!

Example: Problematic in G1 set → Atomization energy of  $P_2$

DMC one-det	107.9(2)	kcal/mol
DMC multi-det	115.9(2)	
Experiment	116.1(5)	

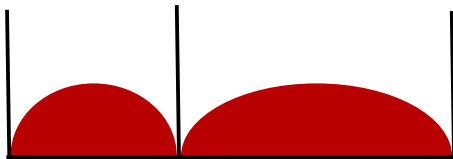
Grossman, J. Chem. Phys. **117**, 1434 (2002)

## Fixed-node DMC and excited states

(1)

No general fixed-node variational principle for excited states

$\tau = 0$ :

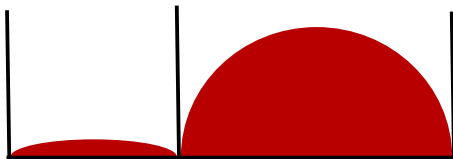


## Fixed-node DMC and excited states

(1)

No general fixed-node variational principle for excited states

$\tau > 0$ :

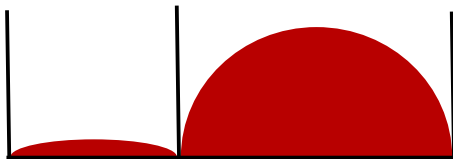


## Fixed-node DMC and excited states

(1)

No general fixed-node variational principle for excited states

$\tau > 0$ :



For  $t \rightarrow \infty$ , only pockets of the lowest energy class are occupied

It can happen that  $E_{\text{FN}} < E_{\text{exact}}$

## Fixed-node diffusion Monte Carlo and excited states

(2)

Is fixed-node diffusion Monte Carlo variational?

For lowest state in each 1-dim irreducible representation

What about “real” excited states?

In general, exact excited state for exact nodal structure

For excited states, even bigger role of the trial wave function

→ Enforces fermionic antisymmetry + selects the state

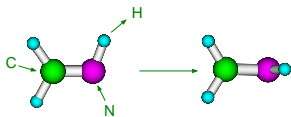
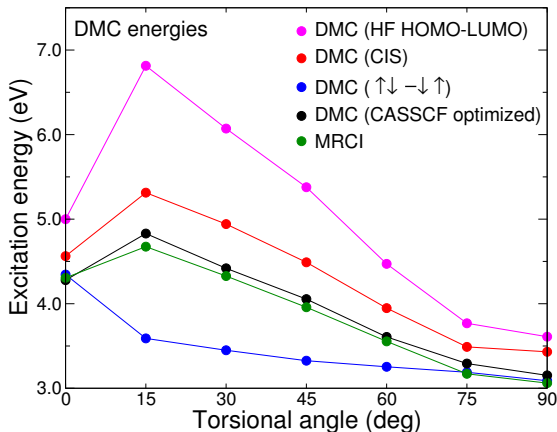
In practice, fixed-node DMC most often works very well



## Excited states and the trial wave function

Dependence of DMC energy from wave function  $\Psi = \mathcal{J} \left[ \sum_i c_i D_i \right]$

Lowest singlet excitation along torsional path of formaldimine



## Excited state optimal wave function

Wave functions for multiple states of the same symmetry

$$\Psi_I(\mathbf{r}_1, \dots, \mathbf{r}_N) = \sum_i c_i^I \mathcal{J}(\mathbf{r}_1, \dots, \mathbf{r}_N) \times D_i(\mathbf{r}_1, \dots, \mathbf{r}_N)$$

Common set of parameters in  $\mathcal{J}$  and  $D_i$  but different coefficients  $c_i^I$

Optimize parameters in  $\mathcal{J}$  and  $D_i$  by state averaging

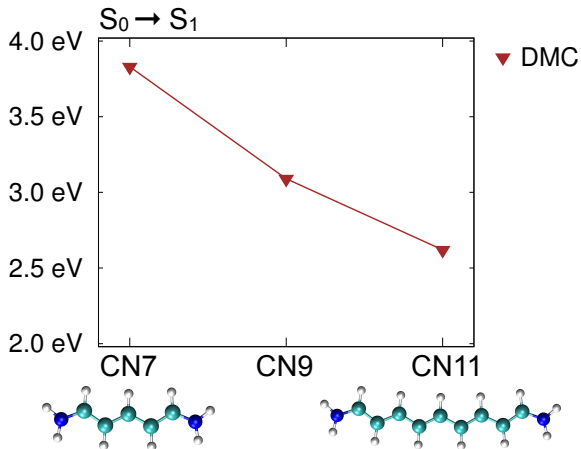
$$E_{\text{SA}} = \sum_I w_I \frac{\langle \Psi_I | \mathcal{H} | \Psi_I \rangle}{\langle \Psi_I | \Psi_I \rangle}$$

and preserve orthogonality through coefficients  $c_i^I$

Filippi, Zaccheddu and Buda, JCTC (2009)

In practice, fixed-node DMC for excited states works well

Example: Cyanines dyes

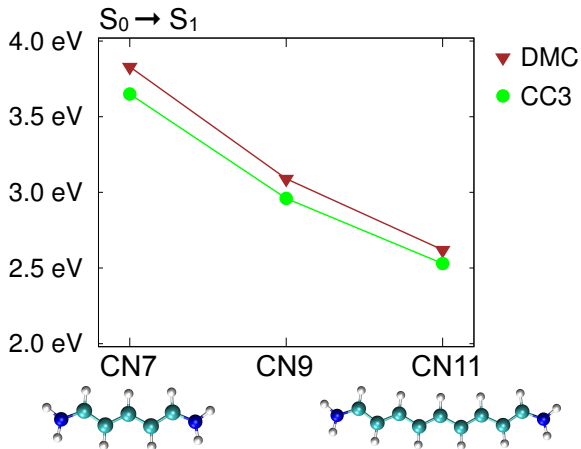


DMC/aug-cc-pVTZ, CC and CASPT2/ANO-L-VTZP calculations

Send, Valsson, Filippi, JCTC (2011).

In practice, fixed-node DMC for excited states works well

Example: Cyanines dyes

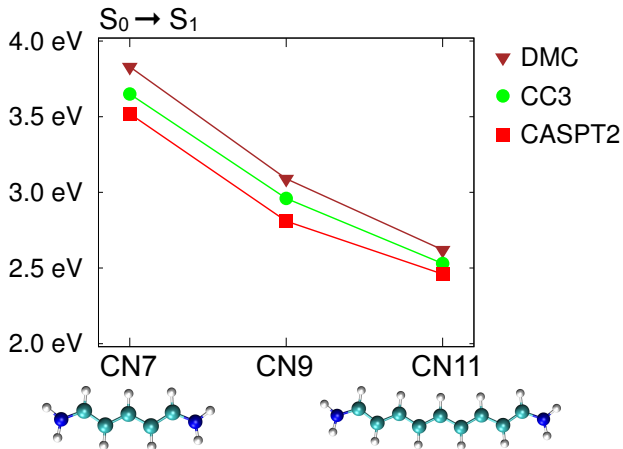


DMC/aug-cc-pVTZ, CC and CASPT2/ANO-L-VTZP calculations

Send, Valsson, Filippi, JCTC (2011).

In practice, fixed-node DMC for excited states works well

## Example: Cyanines dyes

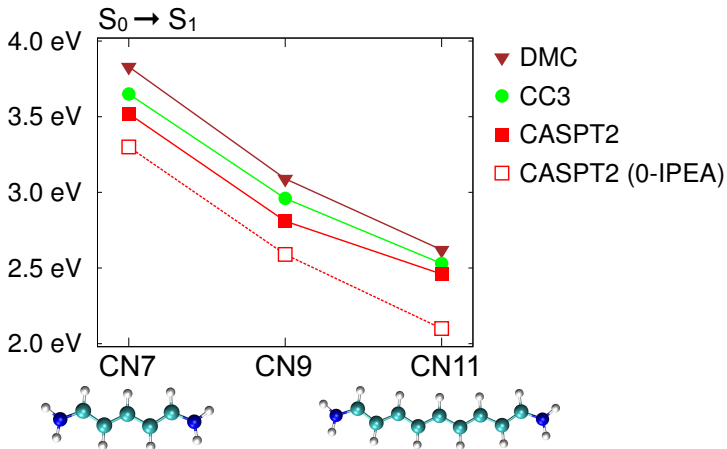


DMC/aug-cc-pVTZ, CC and CASPT2/ANO-L-VTZP calculations

Send, Valsson, Filippi, JCTC (2011).

In practice, fixed-node DMC for excited states works well

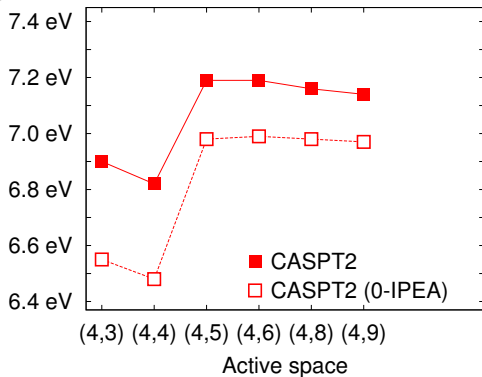
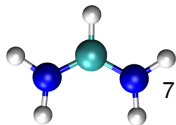
Example: Cyanines dyes



DMC/aug-cc-pVTZ, CC and CASPT2/ANO-L-VTZP calculations

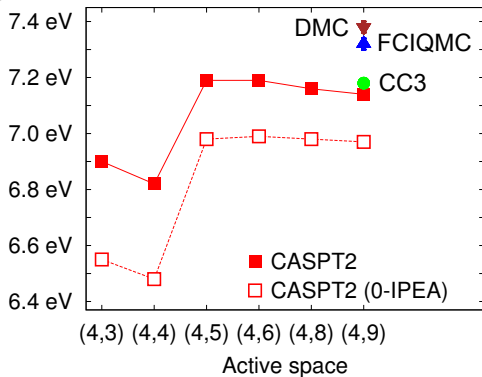
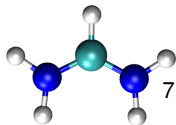
Send, Valsson, Filippi, JCTC (2011).

## Comparison with other theories? A headache



FCIQMC/ANO-L-VDZP calculations

## Comparison with other theories? A headache



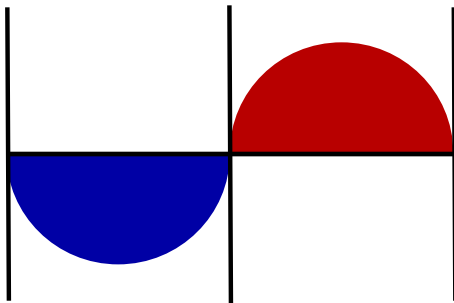
FCIQMC/ANO-L-VDZP calculations



## Alternatives to fixed-node DMC: Releasing the nodes

(1)

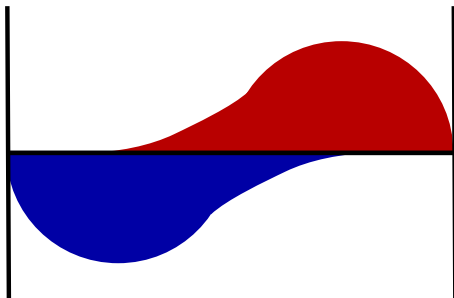
First do a fixed-node DMC simulation



## Alternatives to fixed-node DMC: Releasing the nodes

(1)

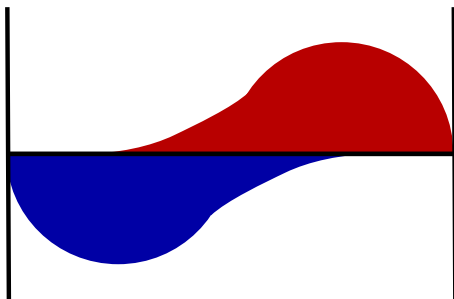
Then release the nodes



## Alternatives to fixed-node DMC: Releasing the nodes

(1)

Then release the nodes



- ▶ Red and blue solutions collapse to boson ground state, but their difference approaches the fermion ground state
- ▶ Back to the sign problem: exponentially growing noise

## Alternatives to fixed-node DMC: Determinantal QMC

(2)

Given single-particle basis, perform projection in determinant space

Different way to deal with fermionic problem

– Determinantal QMC by Zhang and Krakauer

Appears less plagued by fixed phase than DMC by FN

– Full-CI QMC by Alavi

Start from  $\Psi_{\text{CI}} = \sum_i c_i D_i$

$$\mathcal{H}\Psi = -\frac{\partial\Psi}{\partial t} \rightarrow H_{ij}c_j = -\frac{\partial c_i}{\partial t}$$

## DMC in summary

The fixed-node DMC method is

- ▶ Easy to do
- ▶ Stable
- ▶ Accurate enough for many applications in quantum chemistry  
... especially in large systems
- ▶ Not accurate enough for subtle correlation physics

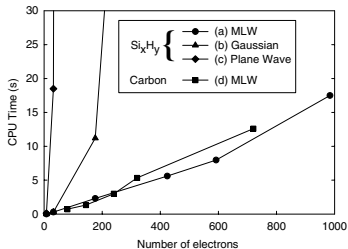
## Beauty of quantum Monte Carlo $\rightarrow$ Highly parallelizable

$\Psi(\mathbf{r}_1, \dots, \mathbf{r}_N) \rightarrow$  Ensemble of walkers diffusing in  $3N$  dimensions

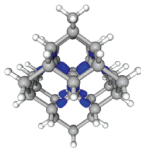
VMC  $\rightarrow$  Independent walkers  $\Rightarrow$  Trivial parallelization

DMC  $\rightarrow$  Nearly independent walkers  $\Rightarrow$  Few communications

Easily take great advantage of parallel supercomputers!



Up to  $\text{Si}_{123}\text{H}_{100}$  and  $\text{C}_{180}$  !



Williamson, Hood, Grossman (2001)

## Human and computational cost of a typical QMC calculation

Task	Human time	Computer time
Choice of basis set, pseudo etc.	10%	5%
DFT/HF/CI runs for $\Psi$ setup	65%	10%
Optimization of $\Psi$	20%	30%
DMC calculation	5%	55%

## Outlook on QMC → Subjects of ongoing research

- ▷ Search for different forms of trial wave function
- ▷ Interatomic forces → Relaxation and dynamics
- ▷ Let us attack transition metals!
- ▷ Alternatives to fixed-node diffusion Monte Carlo



## Other applications of quantum Monte Carlo methods

- ▶ **Electronic structure calculations**
- ▶ Strongly correlated systems (Hubbard, t-J, ...)
- ▶ Quantum spin systems (Ising, Heisenberg, XY, ...)
- ▶ Liquid-solid helium, liquid-solid interface, droplets
- ▶ Atomic clusters
- ▶ Nuclear structure
- ▶ Lattice gauge theory

Both zero (ground state) and finite temperature

## The drift-branching components: Reminder

### Diffusion term

$$-\frac{1}{2}\nabla^2\tilde{G}(\mathbf{R}, \mathbf{R}_0, t) = -\frac{\partial\tilde{G}(\mathbf{R}, \mathbf{R}_0, t)}{\partial t}$$

$$\Rightarrow \tilde{G}(\mathbf{R}', \mathbf{R}, \tau) = (2\pi\tau)^{-3N/2} \exp\left[-\frac{(\mathbf{R}' - \mathbf{R})^2}{2\tau}\right]$$

### Branching term

$$(E_L(\mathbf{R}) - E_T)\tilde{G}(\mathbf{R}, \mathbf{R}_0, t) = -\frac{\partial\tilde{G}(\mathbf{R}, \mathbf{R}_0, t)}{\partial t}$$

$$\Rightarrow \tilde{G}(\mathbf{R}', \mathbf{R}, \tau) = \exp[-\tau(E_L(\mathbf{R}) - E_T)] \delta(\mathbf{R} - \mathbf{R}')$$

## The drift-diffusion-branching Green's function

$$-\frac{1}{2}\nabla^2\tilde{G} + \nabla \cdot [\tilde{G}\mathbf{V}(\mathbf{R})] + [E_L(\mathbf{R}) - E_T]\tilde{G} = -\frac{\partial\tilde{G}}{\partial\tau}$$

### Drift term

Assume  $\mathbf{V}(\mathbf{R}) = \frac{\nabla\Psi(\mathbf{R})}{\Psi(\mathbf{R})}$  constant over the move (true as  $\tau \rightarrow 0$ )

The drift operator becomes  $\mathbf{V} \cdot \nabla + \nabla \cdot \mathbf{V} \approx \mathbf{V} \cdot \nabla$  so that

$$\mathbf{V} \cdot \nabla \tilde{G}(\mathbf{R}, \mathbf{R}_0, t) = -\frac{\partial\tilde{G}(\mathbf{R}, \mathbf{R}_0, t)}{\partial t}$$

with solution  $\tilde{G}(\mathbf{R}, \mathbf{R}_0, t) = \delta(\mathbf{R} - \mathbf{R}_0 - \mathbf{V}t)$

## An important and simple improvement

If  $\Psi = \Psi_0$ ,  $E_L(\mathbf{R}) = E_0 \rightarrow$  No branching term  $\rightarrow$  Sample  $\Psi^2$

Due to time-step approximation, we only sample  $\Psi^2$  as  $\tau \rightarrow 0$  !

**Solution** Introduce accept/reject step like in Metropolis algorithm

$$\tilde{G}(\mathbf{R}', \mathbf{R}, \tau) \approx \underbrace{\mathcal{N} \exp \left[ -\frac{(\mathbf{R}' - \mathbf{R} - \mathbf{V}(\mathbf{R})\tau)^2}{2\tau} \right]}_{T(\mathbf{R}', \mathbf{R}, \tau)} \exp \left[ -(E_L(\mathbf{R}) + E_L(\mathbf{R}')) \frac{\tau}{2} \right]$$

Walker drifts, diffuses and the move is accepted with probability

$$p = \min \left\{ 1, \frac{|\Psi(\mathbf{R}')|^2 T(\mathbf{R}, \mathbf{R}', \tau)}{|\Psi(\mathbf{R})|^2 T(\mathbf{R}', \mathbf{R}, \tau)} \right\}$$

$\rightarrow$  Improved algorithm with smaller time-step error

## Finite and infinite variance

$$F_M = \frac{1}{M} \sum_{i=1}^M f(x_i)$$

Finite variance  $\sigma \Rightarrow$  **The Central Limit Theorem**

Since we have a Gaussian distribution for  $F_M$ , the probability of

$F_M$  being within  $1 \sigma_M$  of the true mean is 68.3%

$F_M$  being within  $2 \sigma_M$  of the true mean is 95.4%

$F_M$  being within  $3 \sigma_M$  of the true mean is 99.7%

Infinite variance  $\sigma \Rightarrow$  **The law of large numbers**

The sample mean converge to the expected value (if finite)

... but statistical error goes down slower than  $1/\sqrt{M}$

Beware of densities with  $\infty$  variance!