



**HAL**  
open science

## Quantum Monte Carlo

Marius Lewerenz

► **To cite this version:**

| Marius Lewerenz. Quantum Monte Carlo. 2013. hal-00832980

**HAL Id: hal-00832980**

**<https://hal.science/hal-00832980>**

Submitted on 11 Jun 2013

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Quantum Monte Carlo

Helsinki 2011

Marius Lewerenz

MSME/CT, UMR 8208 CNRS, Université Paris Est (Marne-la-Vallée)

[Marius.Lewerenz@univ-paris-est.fr](mailto:Marius.Lewerenz@univ-paris-est.fr)

<http://msme.univ-mlv.fr/staff/ct/marius-lewerenz/>

# Table of contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
1.1	What is our problem? . . . . .	2
1.2	What is a Monte Carlo method? . . . . .	5
1.3	What are Monte Carlo methods good for? . . . . .	5
1.4	The Difficulty of Monte Carlo Methods . . . . .	8
1.5	A Classification of Monte Carlo Methods . . . . .	9
<b>2</b>	<b>Review of Probability and Statistics</b>	<b>10</b>
2.1	Probabilities and Random Variables . . . . .	10
2.2	Joint and Marginal Probabilities . . . . .	11
2.3	Random Variables and Expectation Values . . . . .	12

2.4	Moments of a Distribution . . . . .	14
2.5	Variance of a Random Function . . . . .	15
2.6	The Covariance . . . . .	16
2.7	Properties of the Covariance . . . . .	17
2.8	Correlation and Autocorrelation . . . . .	18
2.9	Continuous Distributions . . . . .	19
2.10	Moments of Continuous Distributions . . . . .	20
2.11	Sums of Random Variables . . . . .	21
2.12	Variance of the Sum of Random Variables . . . . .	22
<b>3</b>	<b>Sources of Randomness</b>	<b>23</b>
3.1	Random Walks . . . . .	24
3.2	The Stochastic Matrix . . . . .	25

3.3	Properties of the Stochastic Matrix . . . . .	26
3.4	Detailed Balance . . . . .	28
3.5	Decomposition of the Transition Process . . . . .	29
3.6	Accepting Proposed Transitions . . . . .	30
3.6.1	How to Accept a Proposed Change of State? . . . . .	31
3.7	Random Walks in Continuous Spaces . . . . .	31
3.8	Coordinates of a Random Walk . . . . .	32
3.9	The Transition Function $t$ . . . . .	33
3.10	Summary of Important Random Walk Features . . . . .	34
<b>4</b>	<b>Monte Carlo Integration</b>	<b>35</b>
4.1	The Principle of Monte Carlo Integration . . . . .	35
4.2	Efficiency of Monte Carlo Integration . . . . .	36

5	Variational Quantum Monte Carlo	37
5.1	The energy expectation value . . . . .	37
5.2	The Local Energy . . . . .	39
5.3	Evaluating Expectation Values . . . . .	40
5.4	A Program Skeleton . . . . .	41
5.5	General Criteria for Trial Wave Functions . . . . .	42
5.6	Optimizing Trial Wave Functions . . . . .	43
5.7	Moving Downhill in Parameter Space . . . . .	44
5.8	Variance reduction: Reweighting of Random Walks . . . . .	45
5.9	Iterative Fixed Sampling . . . . .	46
6	Diffusion Quantum Monte Carlo (DMC)	47
6.1	The basic idea . . . . .	47

6.2	The Stationary Solution . . . . .	49
6.3	The effect of imaginary time . . . . .	50
6.4	Importance Sampling . . . . .	51
6.5	Formal Time Evolution . . . . .	52
6.6	Probabilistic Interpretation and Short Time Approximation . . . . .	53
6.7	The Practical Solution . . . . .	54
6.8	The Individual Actions . . . . .	55
6.9	Making a Move: Metropolis algorithm with asymmetric transition probabilities	56
6.10	Why impose detailed balance? . . . . .	57
6.11	Energy Estimators I . . . . .	58
6.12	Trial Functions and Local Energy . . . . .	59
6.13	How to find a good $\Psi_T$ ? . . . . .	60

6.14 Selfconsistent DMC . . . . .	61
6.15 Ensemble Size Error . . . . .	62
6.16 Evolution of Random Walker Weights . . . . .	65
6.17 Handling Random Walker Weights . . . . .	66
6.17.1 Pure DMC . . . . .	66
6.17.2 Population control I . . . . .	67
6.17.3 Population control II: Continuous weights with branching/termination . . . . .	68
6.17.4 Combining small walkers . . . . .	69
6.17.5 Stochastic Ensemble Control . . . . .	70
6.18 Expectation values . . . . .	71
6.18.1 The meaning of the weights . . . . .	71
6.18.2 How to find a measure of $\Phi_0^2$ . . . . .	72

6.18.3 The noise problem . . . . .	72
6.18.4 The cheap and dubious way: Squaring the Weights . . . . .	73
6.18.5 Getting it right: Descendant Weighting/Future Walking I . . . . .	74
6.18.6 Descendant Weighting/Future Walking II . . . . .	75
6.18.7 Perturbational estimates for scalar expectation values . . . . .	77
6.18.8 More good news . . . . .	78
6.19 Correlated Sampling . . . . .	79
6.20 Existing techniques which we have not addressed . . . . .	80
6.21 Scaling behavior of DMC . . . . .	81
6.22 Why is quantum Monte Carlo not more popular? . . . . .	83

**List of Figures**

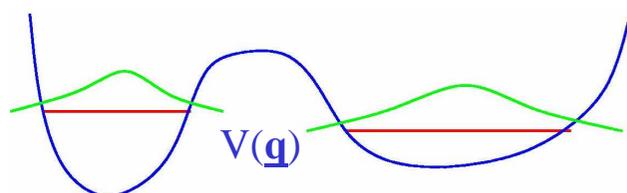
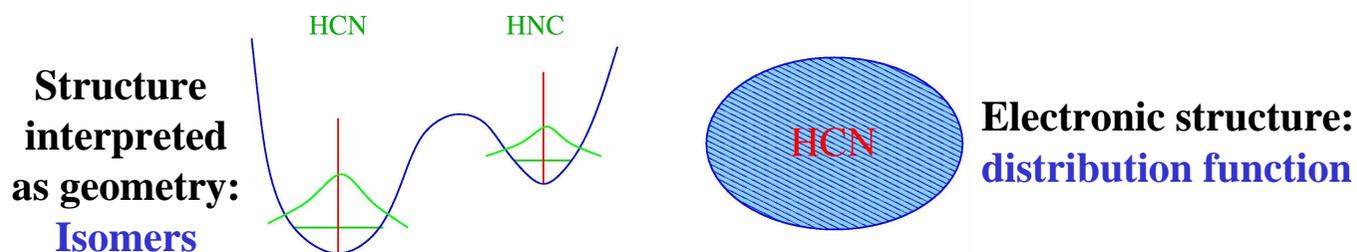
**List of Tables**



# 1 Introduction

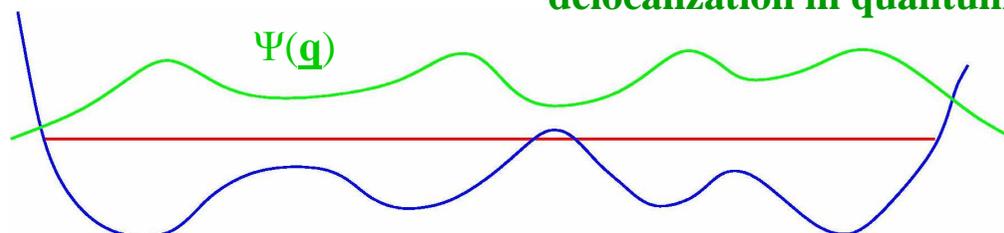
## 1.1 What is our problem?

### Molecular structure and vibrational motion

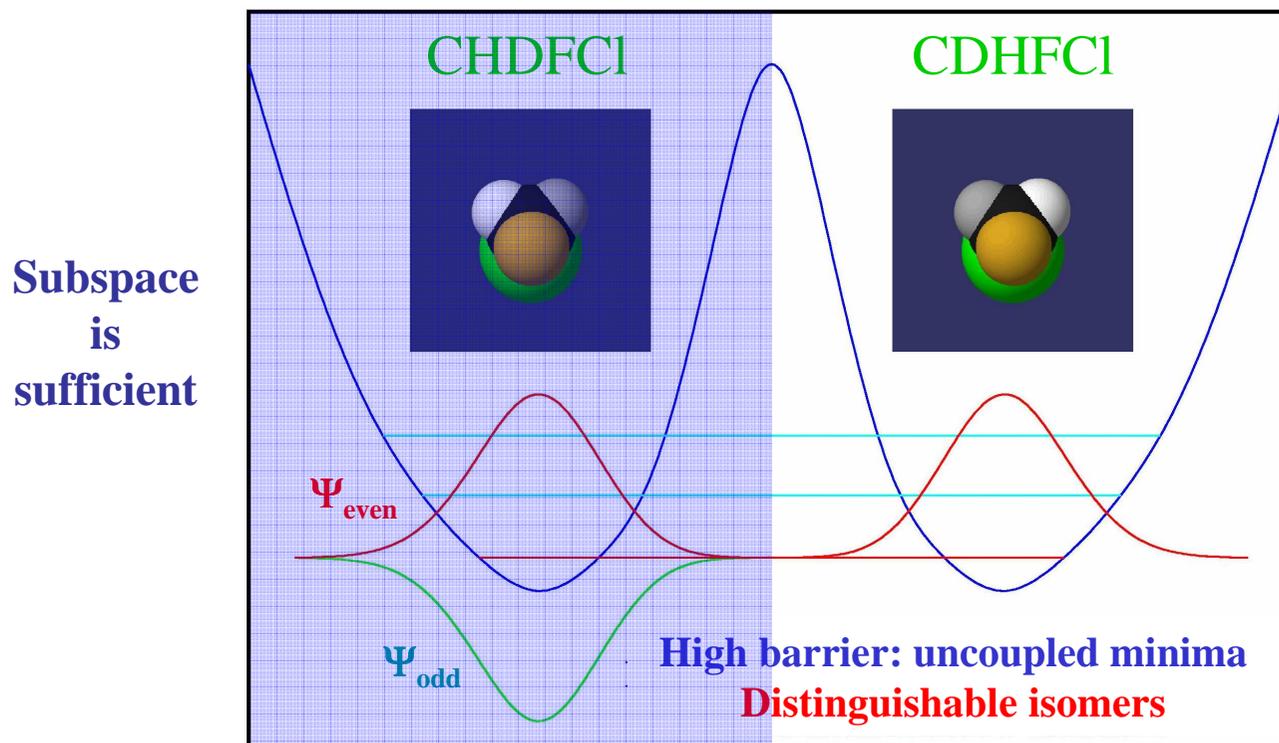


**What is the size of typical vibrational amplitudes?**

**Zero point energy effects on relative stability, delocalization in quantum liquids (LHe, LH<sub>2</sub>)**

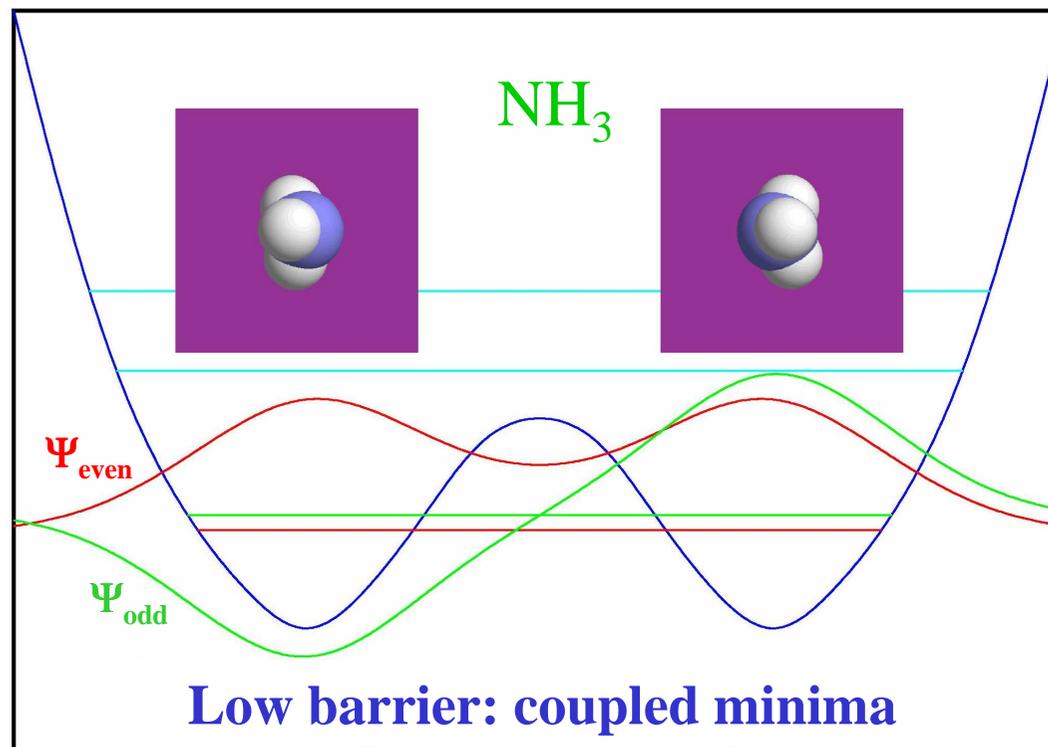


## Equivalent structures with high barriers



Vanishing coupling and practically degenerate states:  
Localized wave functions  $\Psi_r$  and  $\Psi_l \propto (\Psi_{\text{even}} \pm \Psi_{\text{odd}})$

## Equivalent structures with low barriers



Coupling  $\Rightarrow$  splitting between “even” and “odd” states:  
Global wave function  $\Rightarrow$  special techniques required

## 1.2 What is a Monte Carlo method?

Any method which uses random numbers

Our focus will be on Monte Carlo methods of physical and chemical relevance.

### Technical formulation:

- Represent the solution of a mathematical/physical/chemical problem as a parameter of a hypothetical distribution.
- Construct a set of samples from this distribution.
- Use these samples to compute statistical estimators for the desired parameter.

## 1.3 What are Monte Carlo methods good for?

Monte Carlo methods work for **any stochastic problem** and for a large class of **deterministic problems**, quadrature being a classical example.

- Classical statistical mechanics: ensemble concept (Gibbs)

**Known distributions**

- Quantum (statistical) mechanics: probabilistic interpretation of wave functions (Born), density matrices  
Usually unknown distributions



## 1.4 The Difficulty of Monte Carlo Methods

- We need to find a proper mapping of our problem onto a stochastic model.
- We need to think in terms of statistics and fluctuations when we analyse Monte Carlo data.
- Each Monte Carlo calculation for the same problem will give a somewhat different answer.
- The notion of convergence needs to be redefined.
- Lack of familiarity can lead to misinterpretations and optimistic ideas about accuracy (correlations).
- Fluctuations can mask subtle systematic errors (random number quality, programming errors etc.).
- Monte Carlo methods look simple but they are not.

Use random numbers to solve your problem but do not produce random results

## 1.5 A Classification of Monte Carlo Methods

	Classical	Quantum	
$T = 0$	Locating the minimum of a multidimensional surface, (Simulated annealing etc.)	Single occupied quantum state of known energy (Minimisation in discrete space)	Single quantum state(s) with unknown properties: variational Monte Carlo ( <b>VMC</b> ), <b>GFMC</b> , diffusion Monte Carlo ( <b>DMC</b> )
$T > 0$	Integration over continuous states Classical Monte Carlo ( <b>CMC</b> ) in various ensembles	Summation over discrete states (lattice model Hamiltonians, Ising etc.), technically <b>similar to CMC</b>	Direct averaging over many quantum states: path integral Monte Carlo ( <b>PIMC</b> )
		known energy levels $E_i$	unknown $E_i$

## 2 Review of Probability and Statistics

### 2.1 Probabilities and Random Variables

We consider a reservoir of possible outcomes  $\{E\}$  for a random event.

$$\{E\} = \{E_1, E_2, E_3 \dots E_n\}$$

We associate a **probability**  $p_k$  with each  $E_k$ :

$$P(E_k) = p_k \quad 1 \geq p_k \geq 0$$

**Properties of  $p_k$ :**

1. The following relations hold for any pair of  $E_i, E_j$ .

$$P(E_i \wedge E_j) \leq p_i + p_j \quad ; \quad P(E_i \vee E_j) \leq p_i + p_j$$

2. If  $E_i$  and  $E_j$  are **mutually exclusive**

$$(E_i \Rightarrow \neg E_j, E_j \Rightarrow \neg E_i):$$

$$P(E_i \wedge E_j) = 0 \quad ; \quad P(E_i \vee E_j) = p_i + p_j$$

3. For a **class of mutually exclusive events**, which contains all possible events we have:

$$P(\text{some } E) = 1 = \sum_i p_i$$

## 2.2 Joint and Marginal Probabilities

Suppose that  $E$  and  $F$  satisfy the conditions defined above

$$P(E_i) = p_{1i} \quad P(F_j) = p_{2j}$$

and we are interested in the probability of the **combined event**  $(E_i, F_j)$ . This probability is called the **joint probability**

$$P(E_i, F_j) = p_{ij}$$

The events  $E_i$  and  $F_j$  are called **independent** if

$$p_{ij} = p_{1i}p_{2j}$$

If the events  $E_i$  and  $F_j$  are not independent, there is a useful decomposition of the **joint probability** as follows:

$$p_{ij} = \left( \sum_k p_{ik} \right) \left[ \frac{p_{ij}}{\sum_k p_{ik}} \right]$$

$$p_{ij} = p(i) \left[ \frac{p_{ij}}{\sum_k p_{ik}} \right]$$

The quantity  $p(i)$  is called the **marginal probability** for the event  $E_i$ , the probability of observing  $E_i$  whatever  $F$  occurs. Clearly  $p(i) = p_{1i}$  and  $\sum_i p(i) = \sum_i \sum_k p_{ik} = 1$ .

The expression  $p_{ij}/\sum_k p_{ik} = p(j|i)$  defines the **conditional probability** of observing  $F_j$ , provided  $E_i$  has occurred. Since the probability for some  $F_j$  should be 1 we have

$$\sum_j p(j|i) = \sum_j \frac{p_{ij}}{\sum_k p_{ik}} = \frac{\sum_j p_{ij}}{\sum_k p_{ik}} = 1$$

## 2.3 Random Variables and Expectation Values

The random events  $E, F$  can be **anything of numerical or non numerical character** (e.g. a noise amplitude or a logical decision). If we can associate a numerical value  $x_i$  with each random event  $E_i$ , we call  $x$  a **random variable**.

We define the **expectation value**  $E\{x\}$  of a random variable  $x$  as

$$E\{x\} = \langle x \rangle = \sum_i p_i x_i$$

Assume that  $g$  is a function of  $x$ ,  $g(x_i) = g_i$ . Then also  $g_i$  will be a random variable and we define

$$E\{g(x)\} = \langle g(x) \rangle = \sum_i p_i g(x_i)$$

Suppose that  $g(x_i) = g(x) = \text{const}$ :

$$E\{g(x)\} = \sum_i p_i g(x_i) = g(x_i) \sum_i p_i = g(x)$$

The expectation value of a constant is a constant.

**Linearity of the expectation value** of two random functions  $g_1(x)$  and  $g_2(x)$ :

$$\begin{aligned} E\{\lambda_1 g_1(x) + \lambda_2 g_2(x)\} &= \langle \lambda_1 g_1(x) + \lambda_2 g_2(x) \rangle \\ &= \sum_i p_i (\lambda_1 g_1(x_i) + \lambda_2 g_2(x_i)) \\ &= \lambda_1 \sum_i p_i g_1(x_i) + \lambda_2 \sum_i p_i g_2(x_i) \\ &= \lambda_1 \langle g_1(x) \rangle + \lambda_2 \langle g_2(x) \rangle \end{aligned}$$

## 2.4 Moments of a Distribution

We define the  $n$ th moment of a distribution as

$$\mu_n = \langle x^n \rangle = \sum_i p_i x_i^n$$

These powers of  $x$  are nothing but special cases of the random functions  $g(x)$ .

Principal moments:

$$\begin{aligned}\mu_1 &= \sum_i p_i x_i && \text{mean of the distribution} \\ \mu_2 &= \sum_i p_i x_i^2\end{aligned}$$

Central moments:

$$\langle m_n(x) \rangle = \langle (x - \mu_1)^n \rangle = \sum_i p_i (x_i - \langle x \rangle)^n$$

The special case of  $n = 2$  is called the **variance**:

$$\text{Var}\{x\} = \langle m_2(x) \rangle = \langle x^2 \rangle - \langle x \rangle^2$$

$\text{Var}\{x\}$  and  $\mu_1$  are sufficient to uniquely specify many important distributions, in particular the important **Gaussian distribution**.

## 2.5 Variance of a Random Function

In a similar way we can define the **variance of a random function  $g(x)$** :

$$\begin{aligned}
 \text{Var}\{g(x)\} &= \langle (g(x) - \langle g(x) \rangle)^2 \rangle \\
 &= \sum_i p_i g^2(x) - 2\langle g(x) \rangle \sum_i p_i g(x_i) + \langle g(x) \rangle^2 \sum_i p_i \\
 &= \langle g^2(x) \rangle - \langle g(x) \rangle^2
 \end{aligned}$$

Variance of a Linear Combination of Random Functions:

$$\begin{aligned}
 \text{Var}\{\lambda_1 g_1(x) + \lambda_2 g_2(x)\} &= \langle (\lambda_1 g_1(x) + \lambda_2 g_2(x) - \langle \lambda_1 g_1(x) + \lambda_2 g_2(x) \rangle)^2 \rangle \\
 &= \langle (\lambda_1 g_1(x) + \lambda_2 g_2(x) - \lambda_1 \langle g_1(x) \rangle - \lambda_2 \langle g_2(x) \rangle)^2 \rangle \\
 &= \langle (\lambda_1 [g_1(x) - \langle g_1(x) \rangle] + \lambda_2 [g_2(x) - \langle g_2(x) \rangle])^2 \rangle \\
 &= \langle \lambda_1^2 [g_1(x) - \langle g_1(x) \rangle]^2 + \lambda_2^2 [g_2(x) - \langle g_2(x) \rangle]^2 \\
 &\quad + 2\lambda_1 \lambda_2 [g_1(x) - \langle g_1(x) \rangle][g_2(x) - \langle g_2(x) \rangle] \rangle \\
 &= \lambda_1^2 \langle [g_1(x) - \langle g_1(x) \rangle]^2 \rangle + \lambda_2^2 \langle [g_2(x) - \langle g_2(x) \rangle]^2 \rangle \\
 &\quad + 2\lambda_1 \lambda_2 \langle g_1(x) g_2(x) - g_1(x) \langle g_2(x) \rangle - \langle g_1(x) \rangle g_2(x) + \langle g_1(x) \rangle \langle g_2(x) \rangle \rangle
 \end{aligned}$$

(Exploitation of the linearity of expectation values)

$$\text{Var}\{\lambda_1 g_1(x) + \lambda_2 g_2(x)\} = \lambda_1^2 \text{Var}\{g_1(x)\} + \lambda_2^2 \text{Var}\{g_2(x)\} + 2\lambda_1 \lambda_2 (\langle g_1(x) g_2(x) \rangle - \langle g_1(x) \rangle \langle g_2(x) \rangle)$$

## 2.6 The Covariance

The mixed last term in the preceding equation defines the **covariance** of  $g_1(x)$  and  $g_2(x)$ .

$$Cov\{g_1(x), g_2(x)\} = \langle g_1(x)g_2(x) \rangle - \langle g_1(x) \rangle \langle g_2(x) \rangle$$

This term **measures the mutual dependence** between the two random functions  $g_1$  and  $g_2$  and can be positive or negative.

Special cases:

$$Cov\{x, y\} = \langle xy \rangle - \langle x \rangle \langle y \rangle$$

$$Cov\{x, x\} = \langle xx \rangle - \langle x \rangle \langle x \rangle = Var\{x\}$$

The variance of a linear combination of random functions or variables can be larger or smaller than the sum of the individual variances, depending on the sign of the covariance.

$$Var\{g_1 + g_2\} = Var\{g_1\} + Var\{g_2\} + 2 Cov\{g_1, g_2\}$$

The possibility of **negative covariance**

$$Var\{g_1 + g_2\} < Var\{g_1\} + Var\{g_2\}$$

is crucial for **correlated sampling techniques**.

## 2.7 Properties of the Covariance

$$\begin{aligned} \text{Cov}\{x, y\} &= \langle xy \rangle - \langle x \rangle \langle y \rangle \\ \langle xy \rangle &= \sum_{ij} p_{ij} x_i y_j \end{aligned}$$

If the random variables  $x$  and  $y$  are **independent**, the  $p_{ij}$  can be decomposed according to  $p_{ij} = p_{1i} p_{2j}$

$$\begin{aligned} \langle xy \rangle &= \sum_{ij} p_{1i} x_i p_{2j} y_j = \left( \sum_i p_{1i} x_i \right) \left( \sum_j p_{2j} y_j \right) = \langle x \rangle \langle y \rangle \\ &\Rightarrow \text{Cov}\{x, y\} = 0 \end{aligned}$$

Independence of two random variables  $x, y$  is a sufficient but not a necessary condition for  $\text{Cov}\{x, y\}$  to be zero!

**Example:**  $x = \{-1, 0, 1\}$ ,  $y = x^2$ , all  $p_i$  equal  
 $\langle x \rangle = 0$ ,  $\langle xy \rangle = \langle x^3 \rangle = 0$   
 $\Rightarrow \text{Cov}\{x, y\} = 0$

The **covariance is zero in spite of a functional relationship** between  $x$  and  $y$  (full correlation).

## 2.8 Correlation and Autocorrelation

The **correlation coefficient**  $r(x, y)$  is the normalized version of the covariance:

$$r(x, y) = \frac{\text{Cov}\{x, y\}}{\sqrt{\text{Var}\{x\}\text{Var}\{y\}}}$$
$$-1 \leq r(x, y) \leq 1$$

If one considers the values of  $y$  as copies of  $x$  with a constant offset  $\delta$  (in time or some pseudotime establishing an order)

$$y_j = x_i = x_{j-\delta}$$

one can compute a correlation coefficient for each offset  $\delta$ .

$$r(x, y; \delta) = A(x; \delta)$$

This function  $A(x; \delta)$  is called the **autocorrelation function** and varies between -1 and +1.

The computation of the autocorrelation function is an important tool to measure the statistical independence of events within a sequence of random events.

All random walk methods require careful autocorrelation analysis

## 2.9 Continuous Distributions

In the preceding section we have assumed discrete random events, but generally random variables can also be continuous.

For a one-dimensional case we have

$$-\infty \leq x \leq \infty$$

We can define a **cumulative distribution function**  $F(x)$  as

$$F(x) = P(\text{randomly selected } y < x)$$

Assume  $x_2 > x_1$ . Then the events  $x_2 > y \geq x_1$  and  $x_1 > y$  are mutually exclusive and we conclude:

$$\begin{aligned} P(x_2 > y \geq x_1) + P(x_1 > y) &= P(x_2 > y) \\ P(x_2 > y) &\geq P(x_1 > y) \end{aligned}$$

It follows that  $F(x)$  is **monotonically increasing**.

$$F(-\infty) = 0 \quad , \quad F(\infty) = 1$$

The function  $F(x)$  is not necessarily smooth. In differentiable regions one can define the **probability density function**  $\rho(x)$ :

$$\rho(x) = \frac{dF(x)}{dx} \geq 0$$

## 2.10 Moments of Continuous Distributions

- Summations are replaced by integrations.
- probabilities  $p_i$  are replaced by  $dF(x)$ .

$$E\{x\} = \langle x \rangle = \int_{-\infty}^{\infty} x dF(x) \quad \left( = \int_{-\infty}^{\infty} x \rho(x) dx \right)$$

$$\int_{-\infty}^{\infty} dF(x) = \left( \int_{-\infty}^{\infty} \rho(x) dx \right) = F(\infty) = 1$$

$$E\{g(x)\} = \langle g(x) \rangle = \int_{-\infty}^{\infty} g(x) dF(x)$$

The variance is now given as

$$Var\{x\} = E\{x^2\} - E\{x\}^2 = \int_{-\infty}^{\infty} x^2 dF(x) - \left[ \int_{-\infty}^{\infty} x dF(x) \right]^2$$

The variance is not a well defined quantity for all  $\rho(x)$ .

An important example is the **Cauchy-Lorentz-distribution**

$$\rho(x) = \frac{1}{\pi} \frac{a}{x^2 + a^2}$$

for which  $E\{x\} = 0$  and  $E\{x^2\} = \infty$ .

## 2.11 Sums of Random Variables

- Suppose we have random variables  $x_1, x_2, \dots, x_n$  which are distributed according to some probability density function  $\rho(x)$ . The variable  $x_i$  may represent a multidimensional point.
- We evaluate functions  $g_i(x_i)$  for each  $x_i$  where the functions  $g_i$  may or may not be identical. The  $g_i(x_i)$  are random variables.
- We define a **weighted sum**  $G$  over these functions and its expectation value  $E(G)$ :

$$G = \sum_i^n \lambda_i g_i(x_i) \quad \lambda_i \in \mathbf{R}$$

$$E\{G\} = \langle G \rangle = \sum_i^n \lambda_i \langle g_i(x_i) \rangle$$

- A special choice is  $\lambda_i = 1/n$  and all  $g_i$  identical

$$E\{G\} = E \left\{ \frac{1}{n} \sum_i^n g(x_i) \right\} = \frac{1}{n} \sum_i^n E\{g\} = E\{g\}$$

The expectation value for the sum  $G$  is identical with the expectation value  $E\{g\}$  for the function  $g$ .

$G$  can serve as an estimator for  $E\{g(x)\}$

## 2.12 Variance of the Sum of Random Variables

- Assume for simplicity that all  $x_i$  are independent. The covariance is zero for all combinations and the variance of  $G$  can be expressed as the sum of the variances of its terms:

$$\text{Var}\{G\} = \sum_i^n \lambda_i^2 \text{Var}\{g_i(x)\}$$

- Again assume  $\lambda_i = 1/n$ ,  $g_i(x) = g(x)$

$$\text{Var}\{G\} = \sum_i^n \frac{1}{n^2} \underbrace{\text{Var}\{g(x)\}}_{\text{some number}} = \frac{1}{n} \text{Var}\{g(x)\}$$

- The variance of the estimator  $G$  decreases  $\propto 1/n$
- The proportionality factor is  $\text{Var}\{g(x)\}$

The mean of random variables approaches the expectation value of the mean of the distribution.

Statistical convergence:

The deviation  $\delta$  of the estimator from the true value will exceed a specified limit  $\Delta$  with a probability which diminishes as  $n \rightarrow \infty$ .

### 3 Sources of Randomness

Monte Carlo methods require the **creation of random events according to specified probability densities**. There are three classes of sources:

- Natural sources of (true ?) randomness.  
Historically interesting but inefficient, not reproducible, and of hardly quantifiable quality.
- Deterministic **algorithms** producing a sequence of numbers (**pseudorandom numbers**) with properties which are indistinguishable from a true random sequence as measured by a battery of statistical tests.
- Random walks constructed from primitive random events for all complicated multidimensional distributions.

We will look in detail only at **random walk methods** and assume the availability of a good uniform **random number generator** or rather **pseudorandom number generator**.

## 3.1 Random Walks

- General method to generate samples from specified probability density functions  $\rho(x)$ , in particular if the space  $x$  has a high dimensionality.
- For direct generators  $P(x_i)$  is (at least apparently) independent of the previous event  $x_j$ .
- Random walks are a sequence of events  $x_1, x_2, x_3, \dots$ , constructed such that  $P(x_{new}) = f(x_{new}, x_{last})$ .
- The function  $f(x_{new}, x_{last})$  describes a strategy to propagate the walk and is in fact a **conditional probability**.
- The process has a **memory and serial correlation**.

Random walks are a special example for a Markov process

General conditions for random walks which are supposed to generate samples with distribution  $\rho(x)$ :

1. Every point  $x$  must be accessible from any other point in a finite number of steps.
2. It must be possible to revisit the same point  $x$  any number of times.
3. The walk must **not periodically** pass through the same points  $x$  again.

These conditions are equivalent to requiring **ergodicity** of the random walk.

## 3.2 The Stochastic Matrix

Consider for a moment a system with discrete 'states' (position, orientations, quantum numbers etc.)  $x_1, x_2, \dots, x_n$ .

$p_{ij} = p(j|i)$  is the probability to observe  $x_j$  provided that we had  $x_i$  just before.  $p_{ij}$  is the **transition probability** for a Markov process.

All  $p_{ij}$  can be arranged in matrix form:

$$P = \begin{pmatrix} p_{11} & p_{12} & p_{13} & \dots & p_{1n} \\ p_{21} & p_{22} & & & \vdots \\ \vdots & & & & \vdots \\ p_{n1} & \dots & \dots & \dots & p_{nn} \end{pmatrix}$$

The matrix  $P$  is a **stochastic matrix**

- All  $p_{ij} \geq 0$  because they represent probabilities.
- $\sum_j p_{ij} = 1$  for all  $i$  because each transition from  $i$  must lead to one of the available 'states'.

### 3.3 Properties of the Stochastic Matrix

Consider a row vector

$$\underline{\rho}^{(0)} = \{\rho_1^{(0)}, \rho_2^{(0)}, \dots, \rho_n^{(0)}\}$$

which describes an initial state in which  $\rho_i^{(0)}$  is the probability of initially finding the system in state  $i$ .

Each step in the Markov chain can be formulated as a multiplication of this row vector with the stochastic matrix  $P$ :

$$\begin{aligned} \underline{\rho}^{(1)} &= \underline{\rho}^{(0)} P \\ \underline{\rho}^{(2)} &= \underline{\rho}^{(1)} P \\ \dots & \dots \\ \underline{\rho}^{(k)} &= \underline{\rho}^{(0)} P^k \end{aligned}$$

The asymptotic distribution  $\underline{\rho}$  for  $k \rightarrow \infty$  is

$$\underline{\rho} = \lim_{k \rightarrow \infty} \underline{\rho}^{(0)} P^k$$

Repeated multiplication of  $\rho^{(0)}$  with  $P$  converges to a stationary situation if

$$\underline{\rho} = \underline{\rho} \underline{P}$$

$\underline{\rho}$  is an eigenvector of  $\underline{P}$  with eigenvalue 1.

**Note:** The asymptotic distribution  $\underline{\rho}$  is independent of the initial 'state'  $\underline{\rho}^{(0)}$  and depends exclusively on the matrix  $\underline{P}$ . All initial vectors converge to the same asymptotic distribution.

A formal proof requires to show that  $\underline{P}$  has a largest eigenvalue which is exactly 1 and that all other eigenvalues fall between -1 and 1.

## 3.4 Detailed Balance

The remaining question is **how to construct a matrix  $\underline{P}$**  which has an eigenvector corresponding to the desired probability density  $\rho$ .

The eigenvector equation can be written out explicitly as

$$\sum_i^n \rho_i p_{ij} = \rho_j$$

**Ansatz: Detailed balance**

We require

$$\rho_i p_{ij} = \rho_j p_{ji}$$

Therefore we have

$$\sum_i \rho_i p_{ij} = \sum_i \rho_j p_{ji} = \rho_j \sum_i p_{ji} = \rho_j$$

This is exactly the condition required for an eigenvector of  $\underline{P}$  with eigenvalue 1.

**Note:** Detailed balance guarantees  $\underline{\rho P} = \underline{\rho}$  and is therefore a **sufficient condition** to construct a matrix  $\underline{P}$  with the desired asymptotic distribution but it is not necessarily the only possible way!

## 3.5 Decomposition of the Transition Process

We can arbitrarily decompose each  $p_{ij}$  into a factor describing the **probability of proposing** a particular transition  $t_{ij}$  and a factor  $a_{ij}$  describing the **probability of accepting** this choice.

$$p_{ij} = t_{ij}a_{ij}$$

This decomposition is valid if the two processes are **independent**.

Substitution into the detailed balance relationship yields

$$\frac{\rho_j}{\rho_i} = \frac{p_{ij}}{p_{ji}} = \frac{t_{ij}a_{ij}}{t_{ji}a_{ji}}$$

Since we assume that  $\rho$  is a known probability density to be generated by the walk, and since we can pick a transition strategy specifying  $t_{ij}$  according to our taste, it is useful to convert this expression into a form which defines the required acceptance probability:

$$\frac{a_{ij}}{a_{ji}} = \frac{\rho_j t_{ji}}{\rho_i t_{ij}}$$

Note that the construction of the random walk requires only that we are able to compute the **ratio of probability densities**. We can work with densities  $\rho$  which are not normalized!

## 3.6 Accepting Proposed Transitions

There are several choices for the acceptance probabilities  $a_{ij}, a_{ji}$  which satisfy this relation, the most common one being due to [Metropolis et al. \(1953\)](#):

$$a_{ij} = \min \left[ 1, \frac{\rho_j t_{ji}}{\rho_i t_{ij}} \right]$$

Proof by verification of the two possibilities:

$$\begin{aligned} \text{(a)} \quad \rho_j t_{ji} \geq \rho_i t_{ij} &\Rightarrow a_{ij} = 1 \\ &a_{ji} = \frac{\rho_i t_{ij}}{\rho_j t_{ji}} \\ &\frac{a_{ij}}{a_{ji}} = \frac{\rho_j t_{ji}}{\rho_i t_{ij}} \quad q.e.d. \end{aligned}$$

$$\begin{aligned} \text{(b)} \quad \rho_j t_{ji} < \rho_i t_{ij} &\Rightarrow a_{ij} = \frac{\rho_j t_{ji}}{\rho_i t_{ij}} \\ &a_{ji} = 1 \\ &\frac{a_{ij}}{a_{ji}} = \frac{\rho_j t_{ji}}{\rho_i t_{ij}} \quad q.e.d. \end{aligned}$$

### 3.6.1 How to Accept a Proposed Change of State?

- A proposed change of state  $i \rightarrow j$  should be accepted with probability  $0 \leq a_{ij} \leq 1$ .
- Random numbers  $u$  with a uniform probability density  $\rho(u) = 1$  on the interval  $[0, 1]$  have the property
$$P(u \leq x) = F(x) = x$$
- A computed value of  $a_{ij}$  will be greater or equal than a uniform random number  $u$  with probability  $a_{ij}$ .
- **Generate a uniform random number  $u$  and accept the proposed change of state if  $a_{ij} \geq u$ .** The 'new' state  $j$  is the old state  $i$  in case of rejection and has to be counted as a member of the chain of steps!

## 3.7 Random Walks in Continuous Spaces

The fundamental ideas of the construction of random walks via the **stochastic matrix** and a **detailed balance ansatz** can be generalized to an infinite number of states.

$$\frac{a(x \rightarrow x')}{a(x' \rightarrow x)} = \frac{\rho(x')t(x' \rightarrow x)}{\rho(x)t(x \rightarrow x')}$$

$$a(x \rightarrow x') = \min \left[ 1, \frac{\rho(x')t(x' \rightarrow x)}{\rho(x)t(x \rightarrow x')} \right]$$

## 3.8 Coordinates of a Random Walk

The coordinates or 'states' of random walks are defined very broadly and can be

- All discrete (e.g. spins on a lattice)
- All continuous (e.g. atomic coordinates in simulations of liquids)
- Any mixture of the two (e.g. particles with spin and continuous space coordinates)

Specifically also the particle number in physical simulations can be a (discrete) coordinate (Grand canonical Monte Carlo).

### 3.9 The Transition Function $t$

A common simplification consists in the special case

$$t(x' \rightarrow x) = t(x \rightarrow x')$$

specifically the choice  $t(x \rightarrow x') = t(|x - x'|)$ .

The **simplified formula** for the Metropolis acceptance probability for **symmetric transitions** is

$$a(x \rightarrow x') = \min \left[ 1, \frac{\rho(x')}{\rho(x)} \right]$$

Examples:

- Proposing a new position  $x'$  with uniform probability from a volume surrounding  $x$  (e.g. a hypercube of predefined size in the majority of simple random walk methods).
- Picking a new position  $x'$  according to a multidimensional Gaussian centered on  $x$ :

$$t(x \rightarrow x') \propto \exp \left( -\frac{|x - x'|^2}{2\sigma^2} \right)$$

Explicit inclusion of  $t(x \rightarrow x')$  in the formulation allows **guided random walks**:

Force bias Monte Carlo, improved variational Monte Carlo diffusion Monte Carlo with importance sampling

## 3.10 Summary of Important Random Walk Features

### Good features:

- Random sampling from distributions in spaces of high dimension.
- No need to be able to normalize the probability density function (which would involve a multidimensional integration).
- Very general coordinate definition and very broad applicability.

### Troublesome Features:

- The desired distribution is reached only asymptotically.  
When is a random walk in its asymptotic regime?
- Serial correlation between sampling positions.  
Requires careful autocorrelation analysis.

## 4 Monte Carlo Integration

### 4.1 The Principle of Monte Carlo Integration

We have seen that the expectation value of the sum  $G$  of random variables  $g(x_i)$  with  $x_i$  drawn according to the probability density  $\rho(x)$  is identical with the expectation value of  $g(x)$  over the underlying distribution:

$$E\{G\} = E\{g(x)\} \quad , \quad G = \frac{1}{n} \sum_i^n g(x_i) \quad , \quad x_i \propto \rho(x)$$

Now recall the original definition of the expectation value as an integral over the distribution:

$$E\{g(x)\} = \int \rho(x)g(x)dx = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_i^n g(x_i)$$

#### Observations

- The statistical error for the integral is independent of the dimensionality of the integral.
- The statistical error diminishes proportional to  $1/\sqrt{n}$ .
- The proportionality constant controlling the absolute size of the error bar depends on the variance of  $g$ .
- The integrand need not even be smooth.

## 4.2 Efficiency of Monte Carlo Integration

We consider the following situation

- $g(x)$  is a  $d$ -dimensional function.
- We have fixed resources which allow the evaluation of  $g(x)$  at a fixed number of points  $n$ .
- We dispose of a grid based product integration rule with leading error  $h^q$  in the grid step  $h$ .
- In a balanced treatment of all dimensions the grid step size will then be  $h \propto n^{-1/d}$ .

Efficiency analysis:

- The **product rule quadrature** will have overall accuracy  $\varepsilon_{Grid} \propto n^{-q/d}$
- The **Monte Carlo quadrature** will have overall accuracy  $\varepsilon_{MC} \propto n^{-1/2}$
- Monte Carlo will be more efficient if  $d > 2q$ .

Remarks:

- We have neglected the **prefactors** of the  $n$ -dependence.
- Accurate product rules require (locally) smooth integrands.
- Monte Carlo methods are very **robust** and allow 'spiky' integrands.
- For moderate  $d$  so called **Quasi Monte Carlo Methods** can be even more efficient.

## 5 Variational Quantum Monte Carlo

### 5.1 The energy expectation value

VMC is an approximative approach for quantum many body problems using a multidimensional and usually not factorizable **trial wave function**  $\Psi_T(x; p)$  with free parameters  $p$  and exploiting the variational principle combined with Monte Carlo integration techniques.

$$E(p) = \frac{\int \Psi_T^*(x; p) \hat{H} \Psi_T(x; p) dx}{\int |\Psi_T(x; p)|^2 dx} \geq E_{exact}$$

Recasting of the integrand as "**function · weight**":

$$E(p) = \frac{\int |\Psi_T(x; p)|^2 \Psi_T^{-1}(x; p) \hat{H} \Psi_T(x; p) dx}{\int |\Psi_T(x; p)|^2 dx}$$

$$E(p) = \int \rho(x; p) \Psi_T^{-1}(x; p) \hat{H} \Psi_T(x; p) dx$$

The density  $\rho(x; p)$  is a normalized probability density

$$\rho(x; p) = \frac{|\Psi_T(x; p)|^2}{\int |\Psi_T(x; p)|^2 dx}$$

The second factor under the integral is the **local energy**, an energy-like quantity which varies as a function of  $x$ .

$$E_{local} = \Psi_T^{-1}(x; p) \hat{H} \Psi_T(x; p)$$

This integral can be evaluated by Monte Carlo sampling:

$$E(p) \approx \sum_i E_{local}(x_i; p) \quad , \quad x_i \propto \rho(x; p)$$

## 5.2 The Local Energy

$$E_{local} = \Psi_T^{-1}(x; p) \hat{H} \Psi_T(x; p)$$

The local energy is the key quantity in variational and diffusion Monte Carlo.

### Important Properties:

- $E_{local}$  is computable for any reasonable many body wave function (with or without explicit particle correlation) because it only involves derivatives.
- $E_{local}$  approaches a constant (the exact eigenvalue) as  $\Psi_T$  approaches **any of the eigenfunctions**  $\Phi$  of the system (**zero variance property**).
- $E_{local}$  can have singularities for any  $\Psi_T$  with nodes (ground and excited states of fermionic systems, excited states of bosonic systems).  
This can cause enormous fluctuations in the Monte Carlo sampling and should be avoided. Trial wave functions should respect all known **boundary conditions** near nodes and in the various asymptotes.
- Singularities are absent if  $\Psi_T$  happens to be an eigenfunction.
- $E_{local} = V(x)$  if  $\Psi_T = const$  (limiting case in simple diffusion Monte Carlo).

## 5.3 Evaluating Expectation Values

- The calculation of the energy (and other properties) for a given trial wave function  $\Psi_T(p)$  has a complexity comparable to a classical Monte Carlo simulation,
- Convert any classical canonical MC code into a VMC code by the substitution

$$\exp[-V(x)/k_B T] \rightarrow |\Psi_T(x; p)|^2$$

(Ergodicity problems are usually actually smaller because of a 'softer' distribution)

- Any tricks known in classical Monte Carlo can be immediately exploited (truncation of finite range interactions, fast methods for long range potential summations, smart move generation).

## 5.4 A Program Skeleton

1. Pick a starting point  $x$  in  $3N$ -dimensional space and compute  $|\Psi_T(x; p)|^2$ .
  2. Set all counters to zero.
  3. Propose a new point  $x'$  by randomly modifying one or more components of  $x$  within a hyperbox or a high-dimensional gaussian (symmetric transition).
  4. Compute  $z = |\Psi_T(x'; p)|^2 / |\Psi_T(x; p)|^2$   
(probably taking advantage of unchanged parts of  $\Psi_T$ ).
  5. If  $z \geq u$  with  $u$  random in  $[0, 1]$ , replace  $x$  by  $x'$ .
  6. Otherwise retain  $x$  (possibly counting it thus more than once).
  7. Repeat steps 3 to 6 long enough to allow correlations to die out.
  8. Evaluate  $E_{local}$  and any other property at the present point  $x$  and increment the appropriate counters for mean values and variances.  
Return to step 3
- Accumulate autocorrelation information on all important properties during the entire simulation.  
Autocorrelation lengths are not necessarily identical for all properties!
  - For efficiency reasons (vectorisation or parallelisation) a swarm of parallel walks is an option, if each walk by itself is sufficiently long to ensure that the asymptotic regime is sampled.

## 5.5 General Criteria for Trial Wave Functions

- $\hat{H}\Psi_T$  must exist anywhere in space;  $\Psi_T$  and  $\nabla\Psi_T$  must be smooth in all regions where the potential is not infinite.
- The following three integrals must exist

$$\int |\Psi_T|^2 dV \quad , \quad \int \Psi_T^* \hat{H} \Psi_T dV$$
$$\int |\hat{H}\Psi_T|^2 dV$$

The first two integrals are ordinary **requirements of quantum mechanics**, the last integral is a **supplementary requirement**, which assures a finite variance and the computability of error estimates.

- $\Psi_T$  must possess all required **symmetries** (particle interchange, space inversion).
- $\Psi_T$  should have the physically correct **asymptotic behaviour** for large particle separations.
- $\Psi_T$  should properly describe the approach to zero distance between any two particles (**cusp conditions**).
- For bulk systems appropriate periodic boundary conditions at the limits of the simulation cell have to be satisfied.

## 5.6 Optimizing Trial Wave Functions

The hard part of VMC is optimizing  $\Psi_t(x; p)$  with respect to its parameters  $p$ .

### General Strategy

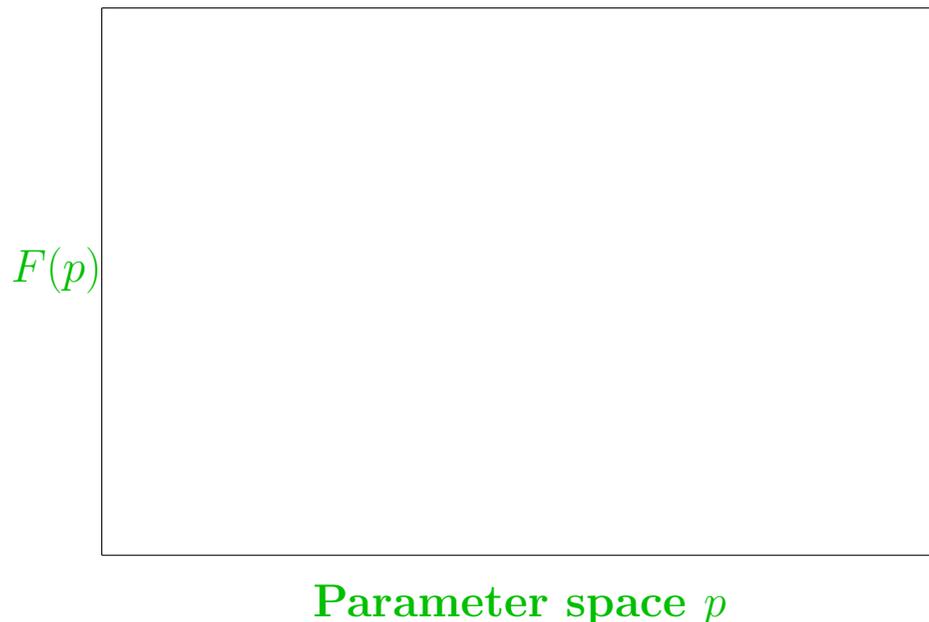
- The variational principle tells us that **energy minimization** as in other quantum chemical techniques would be a good idea.
- The Monte Carlo formulation allows an even better way: **Variance Minimization**.
  - Knowing the target (namely zero variance) we have some sort of measure, how far away we are from an eigenfunction.
  - Variance minimization works in principle also for excited states; energy minimization has contamination problems (mixed wave functions) and is not valid any more unless very difficult orthogonalizations are done.
- Energy minimization is good for coarse optimization, variance minimization can discriminate between energetically equally good functions.
- Variance minimization emphasizes uniform quality and tends to give better expectation values.
- In practice often minimization of a mixed quantity

$$F(p) = \alpha E(p) + \beta \text{Var}\{E(p)\}$$

where  $\beta$  is increased as the optimization progresses.

## 5.7 Moving Downhill in Parameter Space

- As opposed to linear variational problems we have no direct algorithm for locating the minimum.
- Multidimensional nonlinear optimization is known as a **hard problem**. Don't expect too much!
- Random search methods (simulated annealing, threshold accepting etc.) are too expensive and generally not too good in continuous spaces.
- Gradient methods (essentially local) are most common but less global.
- Our biggest problem: **Noise**  
How to compute a gradient from noisy data by finite differencing?



## 5.8 Variance reduction: Reweighting of Random Walks

- In the ordinary case of independent MC integrations at each  $p$ , the error bars of the total energies have to be smaller than the energy difference.
- We should try to directly sample the **change of the energy** as we change  $p$  by a small amount.
- In practice: We evaluate the expectation values  $E(p)$  for a whole set of  $p$ -space points by using sampling positions  $x$  generated by a 'master walk' at some  $p_m$  (usually the best current point).
- Accumulating all estimators over the same set of points guarantees **strong correlation** and reduces relative fluctuations.

$$\begin{aligned}
 E(p) &= \frac{\int \Psi_T(x, p) \hat{H} \Psi_T(x, p) dx}{\int \Psi_T(x, p) \Psi_T(x, p) dx} \\
 &= \frac{\int \Psi_T^2(x, p_m) \frac{\Psi_T^2(x, p)}{\Psi_T^2(x, p_m)} \Psi_T^{-1}(x, p) \hat{H} \Psi_T(x, p) dx}{\int \Psi_T^2(x, p_m) \frac{\Psi_T^2(x, p)}{\Psi_T^2(x, p_m)} dx} \\
 &= \frac{\int \Psi_T^2(x, p_m) W(x; p, p_m) \Psi_T^{-1}(x, p) \hat{H} \Psi_T(x, p) dx}{\int \Psi_T^2(x, p_m) W(x; p, p_m) dx} \\
 &\approx \frac{\sum_i W(x_i; p, p_m) \Psi_T^{-1}(x_i, p) \hat{H} \Psi_T(x_i, p)}{\sum_i W(x_i; p, p_m)}
 \end{aligned}$$

## 5.9 Iterative Fixed Sampling

- In ordinary reweighting a new 'master walk' is generated once a step in  $p$ -space has been taken.
- Actual averages often need only a few thousand uncorrelated samples, originating from very much longer random walks (large delay between samples to suppress serial correlation).
- Reduction of computational cost is possible by storing well decorrelated configurations and some important quantities (e.g.  $\Psi_T(p_m)$  of the generating walk).
- The data volume even for hundreds of particles is quite acceptable (below 3 KByte per configuration and 100 particles).
- Reuse this set of **fixed samples** over many optimization cycles.
- Generate a new set of configurations by a long random walk only when the statistical fluctuations have become too large.

## 6 Diffusion Quantum Monte Carlo (DMC)

### 6.1 The basic idea

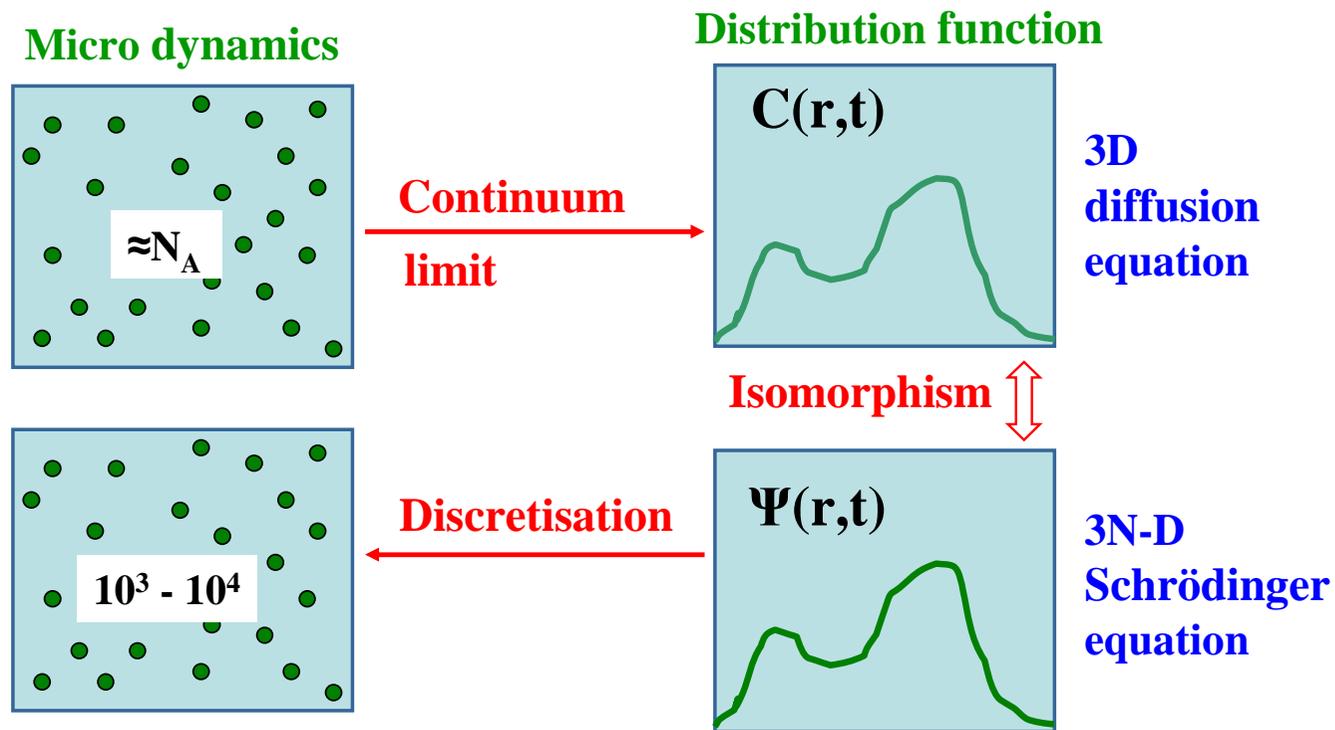
- **Isomorphism** between the time dependent **Schrödinger equation** and a multi dimensional **diffusion equation**. Recognized by Fermi and Ulam in 1940's.
- **Exact solution** limited only by statistical errors.

$$i\hbar \frac{\partial \Psi(\vec{r}, t)}{\partial t} = \left\{ -\frac{\hbar^2}{2} \sum_{j=1}^n \frac{1}{m_j} \nabla_j^2 + \{V(\vec{r}) - E_{ref}\} \right\} \Psi(\vec{r}, t)$$

$$\frac{\partial C(\vec{r}, t)}{\partial t} = \left\{ \sum_{j=1}^n D_j \nabla_j^2 - k(\vec{r}) \right\} C(\vec{r}, t) .$$

- Potential energy function acts as **position dependent reaction term**
- Inverse masses correspond to diffusion coefficients in a  $3N$ -dimensional space.  
**Anisotropic diffusion** with groups of 3 equivalent space dimensions.
- **Reference energy**  $E_{ref}$ : arbitrary shift of the energy scale.

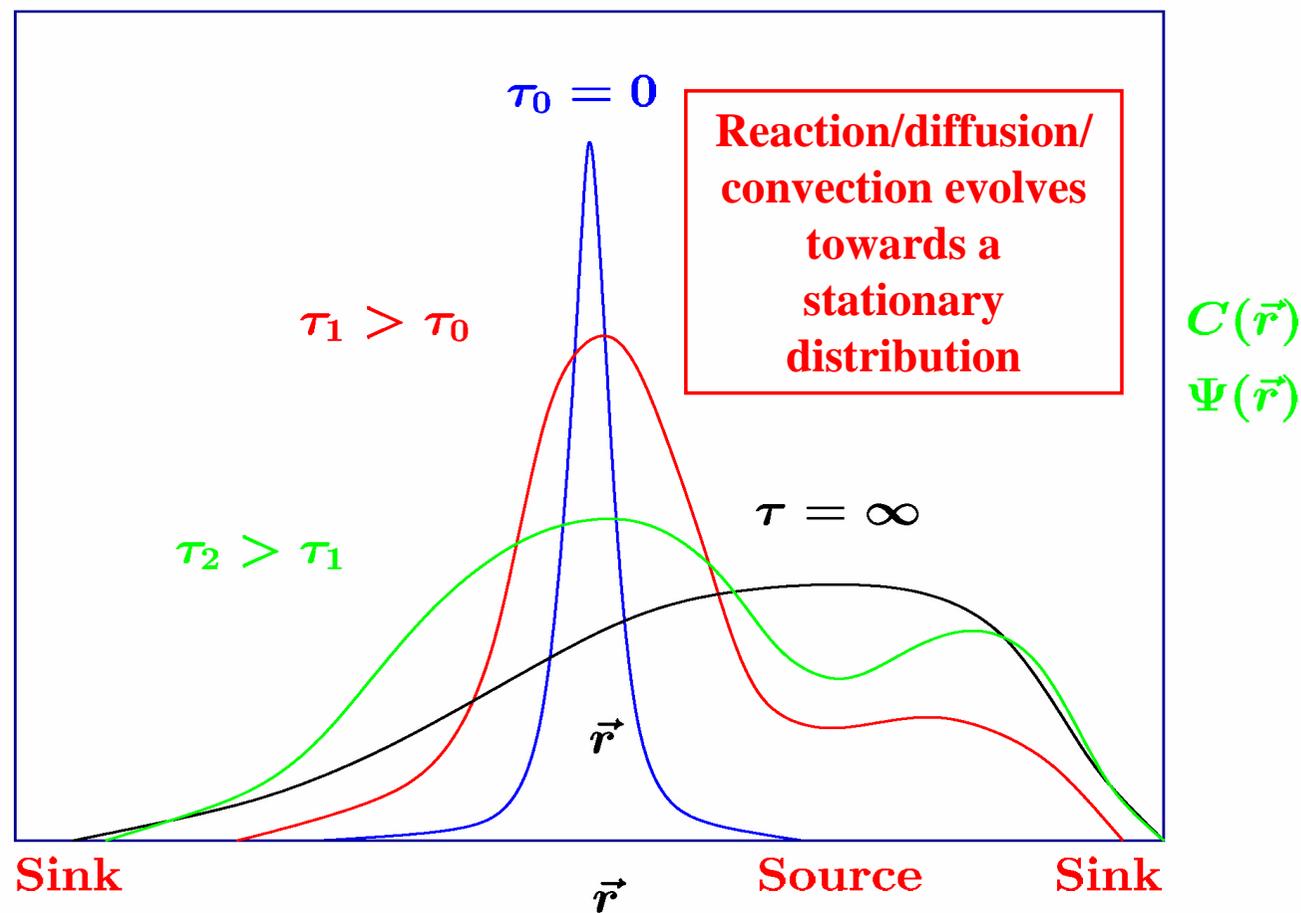
## Key concept of DMC



**Brownian Dynamics  $\Rightarrow$  Ensemble of 'Random Walkers'**

## 6.2 The Stationary Solution

### The diffusion analogy I



## 6.3 The effect of imaginary time

- **Imaginary time**  $\tau = it/\hbar$   
 $\Rightarrow$  mapping of periodic processes on relaxations.
- Propagation of arbitrary initial distribution  $\Psi(\vec{r}, 0)$  leads to **steady state ground state distribution**  $\Phi_0(\vec{r})$ .
- Asymptotic decay with a **rate proportional to**  $E_0 - E_{ref}$ :

$$\Psi(\vec{r}, \tau) = \sum_k A_k \exp\{-(E_k - E_{ref})\tau\} \Phi_k(\vec{r})$$
$$\lim_{\tau \rightarrow \infty} \Psi(\vec{r}, \tau) = A_0 \exp\{-(E_0 - E_{ref})\tau\} \Phi_0(\vec{r})$$

Diffusion Monte Carlo is in principle a **projection method** which projects the exact ground state wave function out of an arbitrary initial function.

The difference with respect to other quantum mechanical projection techniques is the implementation of the projection by **sampling** which has several **advantages and disadvantages!**

Once the ground state has been projected out we can continue the projection and accumulate an arbitrary amount of information.

**Combination of ensemble and time average**

## 6.4 Importance Sampling

Improvement of statistical accuracy and introduction of boundary conditions through **trial wave functions**  $\Psi_T$ :  
**diffusion equation with convection** for  $\Psi\Psi_T$ :

$$\frac{\partial(\Psi\Psi_T)}{\partial\tau} = \sum_{j=1}^n \left\{ \frac{1}{2m_j} \nabla_j^2(\Psi\Psi_T) - \overbrace{\frac{1}{m_j} \nabla_j(\Psi\Psi_T \nabla \ln \Psi_T)}^{\text{Drift}} \right\} - \left\{ \underbrace{\Psi_T^{-1} \hat{T} \Psi_T + V(\vec{r})}_{E_{local}} - E_{ref} \right\} (\Psi\Psi_T)$$

**Advantages:**

- $E_{local}(\vec{r})$  is smoother than  $V(\vec{r})$ .
- Statistical error approaches zero for  $\Psi_T \rightarrow \Psi$ .
- Distribution  $\Psi\Psi_T$  is closer to  $|\Psi|^2$ .  $\Rightarrow$  Simpler evaluation of expectation values.
- Introduction of nodal constraints.

**New Computational Quantities:**

- $E_{local}(\vec{r})$  by analytical differentiation of  $\Psi_T$ .
- "Quantum force" from gradient of  $\Psi_T$ .

## 6.5 Formal Time Evolution

Consider the sequence of trial functions defined by

$$\Psi_{n+1}(\vec{r}) = \exp\{-\Delta\tau(\hat{H} - E_{ref})\}\Psi_n(\vec{r}) = \int d\vec{r}' G(\vec{r}, \vec{r}') \Psi_n(\vec{r}')$$

Expand into eigenfunctions  $\Phi_k$ :

$$\Psi_0(\vec{r}) = \sum_k \langle \Phi_k | \Psi_0 \rangle \Phi_k(\vec{r})$$

$$\Psi_1(\vec{r}) = \sum_k \langle \Phi_k | \Psi_0 \rangle \Phi_k(\vec{r}) \exp[-\Delta\tau(E_k - E_{ref})]$$

$$\Psi_n(\vec{r}) = \sum_k \langle \Phi_k | \Psi_0 \rangle \Phi_k(\vec{r}) \exp[-n\Delta\tau(E_k - E_{ref})]$$

$$\lim_{n \rightarrow \infty} \Psi_n(\vec{r}) = \langle \Phi_0 | \Psi_0 \rangle \Phi_0(\vec{r}) \exp[-n\Delta\tau(E_0 - E_{ref})]$$

In the case of importance sampling:

$$f_n(\vec{r}) = \Psi_T(\vec{r}) \Psi_n(\vec{r})$$

$$f_{n+1}(\vec{r}) = \Psi_T(\vec{r}) \exp\{-\Delta\tau(\hat{H} - E_{ref})\} \Psi_n(\vec{r}) = \int d\vec{r}' \tilde{G}(\vec{r}, \vec{r}') f_n(\vec{r}')$$

$$\lim_{n \rightarrow \infty} f_n(\vec{r}) = \langle \Phi_0 | \Psi_0 \rangle \Phi_0(\vec{r}) \Psi_T(\vec{r}) \exp[-n\Delta\tau(E_0 - E_{ref})]$$

## 6.6 Probabilistic Interpretation and Short Time Approximation

- Propagation requires the evaluation of the integral over the Green's function  $G(\vec{r}, \vec{r}')$ .
- We need a coordinate space representation of  $G$ .
- Positive  $G(\vec{r}, \vec{r}')$  would allow a probabilistic interpretation and **sampling of the integral**.

$$G(\vec{r}, \vec{r}'; \Delta\tau) = \prod_j^N \left[ \left( \frac{m_j}{2\pi\Delta\tau} \right)^{3/2} \exp \left\{ -\frac{m_j}{2\Delta\tau} (\vec{r}_j - \vec{r}'_j)^2 \right\} \right] \times \exp \left\{ -\Delta\tau \left( \frac{V(\vec{r}) + V(\vec{r}')}{2} - E_{ref} \right) \right\} + \mathcal{O}(\Delta\tau^3)$$

In the case of importance sampling:

$$\tilde{G}(\vec{r}, \vec{r}'; \Delta\tau) = \prod_j^N \left[ \left( \frac{m_j}{2\pi\Delta\tau} \right)^{3/2} \exp \left\{ -\frac{m_j}{2\Delta\tau} \left( \vec{r}_j - \vec{r}'_j - \frac{\Delta\tau}{2m_j} \vec{F}_j(\vec{r}) \right)^2 \right\} \right] \times \\ \exp \left\{ -\Delta\tau_{eff} \left( \frac{E_{local}(\vec{r}) + E_{local}(\vec{r}')}{2} - E_{ref} \right) \right\} + \mathcal{O}(\Delta\tau^2)$$

## 6.7 The Practical Solution

- Represent solution by an ensemble of random walkers with weights  $w_i$  ( $\delta$ -functions). Typical ensemble size 1000-10000.
- Cut propagation time into slices  $\Delta\tau$ .
- Solve each part of the diffusion equation (drift, diffusion, growth/decay) separately.  
 $\Rightarrow$  **Short time approximation**, nothing else but a Trotter factorisation.

Drift displaces each  $\delta$ -function  
Diffusion broadens  $\delta$ -function to Gaussian  
Growth/decay scales weight of  $\delta$ -function

Sample the Gaussian to preserve the  $\delta$ -representation by picking new position from a multidimensional Gaussian centered on the position reached after the drift step

- Random walkers move independently  $\Rightarrow$  vectorization or parallelisation.
- Communication only for ensemble control and averaging; minimal overhead.

## 6.8 The Individual Actions

- **Propose a new position** vector  $r'$  for each random walker by updating the components related to particle  $j$  according to

$$r'_j = r_j + \frac{\Delta\tau}{m_j} \vec{F}_j(r) + \sqrt{\frac{\Delta\tau}{m_j}} \vec{\zeta}$$

$$\vec{F}_j(r) = \frac{\nabla_j \Psi_T(\vec{r})}{\Psi_T(\vec{r})}$$

$$\vec{\zeta} = \text{Gaussian random vector}$$

- **Check acceptance** of  $r'$  to ensure detailed balance by computation of backward drift  $\vec{F}(r')$  and new probability density  $\Psi_T^2(r')$ .
- **Apply weight update** according to local energy values:

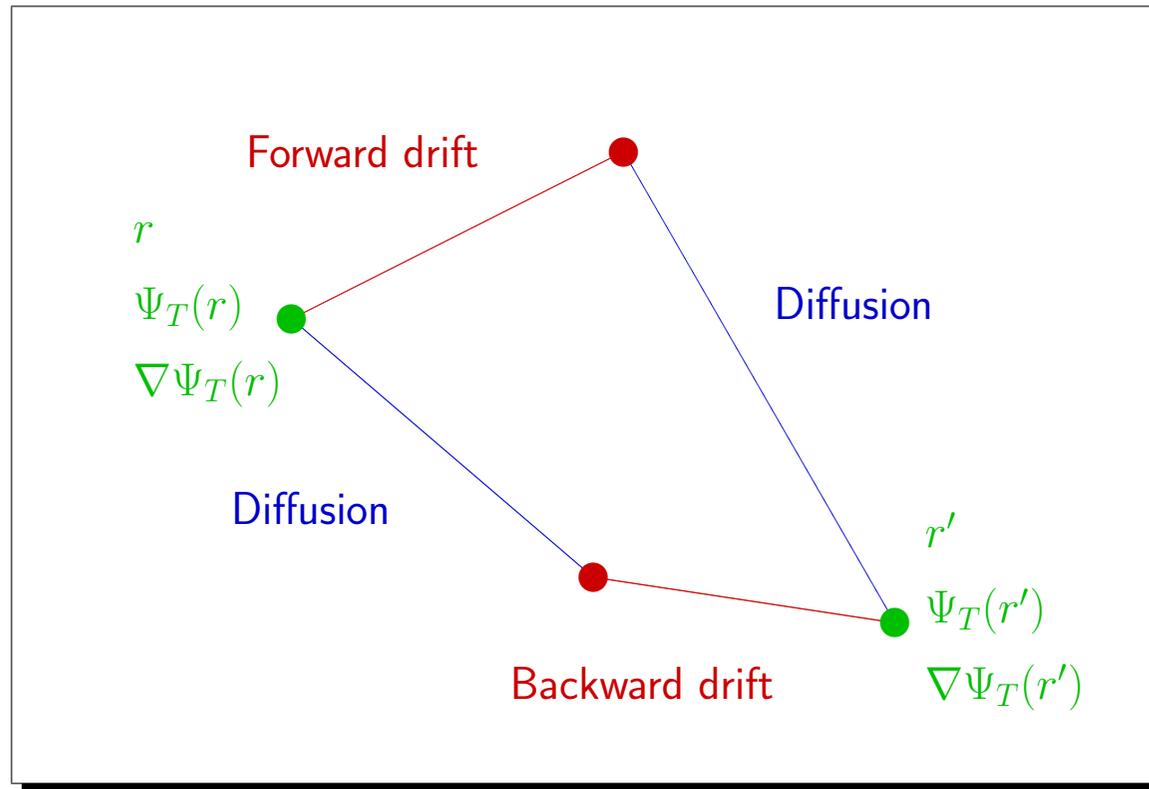
$$w_k^{new} = w_k^{old} \exp \left[ -\Delta\tau_{eff} \left( \frac{E_{local}(\vec{r}) + E_{local}(\vec{r}')}{2} - E_{ref} \right) \right]$$

- Keep track of actual displacements:

$$\langle (r_j - r'_j)^2 \rangle_{prop.} = \frac{\Delta\tau}{m_j}$$

$$\langle (r_j - r'_j)^2 \rangle_{acc.} = \frac{\Delta\tau_{eff}}{m_j}$$

## 6.9 Making a Move: Metropolis algorithm with asymmetric transition probabilities



$$P(r \rightarrow r') = \min \left\{ 1, \frac{|\Psi_T(r)|^2 t(r' \rightarrow r)}{|\Psi_T(r')|^2 t(r \rightarrow r')} \right\}$$

$$t(r \rightarrow r') = \exp \left\{ - \sum_j \frac{m_j}{2\Delta\tau} \left( \vec{r}_j - \vec{r}'_j - \frac{\Delta\tau}{2m_j} \vec{F}_j(r) \right)^2 \right\} ; \quad \vec{F}_j(r) = \nabla_j \ln \Psi_T$$

## 6.10 Why impose detailed balance?

- Distribution of random walkers is supposed to represent  $\Phi_k \Psi_T$  (usually  $k = 0$ ).
- If  $\Psi_T$  is an eigenfunction  $\Phi_k$  (perfect importance sampling)  $E_{local}$  is a constant (the eigenvalue!) and the DMC energy is exactly  $E_k$  for any  $\Delta\tau$  (**vanishing time step error!**).  
All weights  $w_i$  are equal and the **space density** of random walkers alone carries the information on  $\Phi_k^2$ .
- Low order integration of drift leads to a distorted distribution after each time step  $\Delta\tau > 0$  even with perfect importance sampling!

Benefits of detailed balance:

- Imposing detailed balance ensures that the space density of random walkers is proportional to  $\Psi_T^2$  **for any  $\Delta\tau$** .
- Clean interpretation of weights: **Space density** of walkers is always  $\Psi_T^2$  and the weighted ensemble represents  $\Phi_k \Psi_T \Rightarrow$  the **weights are a statistical measure of  $\Phi_k / \Psi_T$** .
- Well defined weights are important for expectation values!
- Significant reduction of **time step error** (empirical observation, no known formal proof).
- Very low rejection rate at typical  $\Delta\tau$ .  $\Rightarrow$  Large enhancement of efficiency at negligible cost.
- Shorter effective time step  $\tau_{eff}$  due to rejected proposals computed from Einstein relationship.

## 6.11 Energy Estimators I

Method 1:  $E_{mean}$ , local energy or mean energy estimator

$$\langle E_{local} \rangle = \frac{\sum_i w_i E_{local}(r_i)}{\sum_i w_i} = \frac{\int \Phi_0 \Psi_T \Psi_T^{-1} \hat{H} \Psi_T d\vec{r}}{\int \Phi_0 \Psi_T d\vec{r}} = \frac{\int \Phi_0 \hat{H} (\sum_k c_k \Phi_k) d\vec{r}}{\int \Phi_0 (\sum_k c_k \Phi_k) d\vec{r}} = E_0$$

Special case  $\Psi_T = const \Rightarrow E_0 = \langle V \rangle$

Method 2:  $E_{grow}$ , growth estimator

$$\Phi_0 \Psi_T(\vec{r}, \tau + \Delta\tau) = \exp\{-(E_0 - E_{ref})\Delta\tau\} \Phi_0 \Psi_T(\vec{r}, \tau)$$

Analyse weight evolution:

$$E_{grow} = E_0 = E_{ref} - \frac{d \ln W(\tau)}{d\tau} ; \quad W = \sum_i w_i$$

## 6.12 Trial Functions and Local Energy

Standard form for atomic systems:

$$\Psi_T(\vec{r}; \vec{p}) = \prod_{i < j \in Rg} \Phi_{ij}(r_{ij}; \vec{p})$$

Permutation symmetry!

$$\Psi_T^{-1} \hat{T} \Psi_T = - \sum_i \frac{\hbar^2}{2m_i} \left\{ \sum_{q=x,y,z} \left[ \sum_{j \neq i} \frac{1}{r_{ij}} \frac{\partial \ln \Phi_{ij}}{\partial r_{ij}} (q_i - q_j) \right]^2 + \sum_{j \neq i} \left[ \frac{2}{r_{ij}} \frac{\partial \ln \Phi_{ij}}{\partial r_{ij}} + \frac{1}{\Phi_{ij}} \frac{\partial^2 \Phi_{ij}}{\partial r_{ij}^2} - \left( \frac{\partial \ln \Phi_{ij}}{\partial r_{ij}} \right)^2 \right] \right\}$$

Computational complexity  $\propto N^2$

Extension for rare gas-molecule clusters with anisotropy:

$$\Psi_T(\vec{r}) = \prod_{i \in Rg} \chi_i(r_i, \cos \theta_i) \cdot \prod_{i < j \in Rg} \Phi_{ij}(r_{ij})$$

## 6.13 How to find a good $\Psi_T$ ?

### 1. Get an idea about the functional form:

- Physical intuition or luck
- Better: Use whatever is known about **asymptotic behavior and cusp conditions**
- Let DMC generate it for you: **Selfconsistent DMC**

### 2. Optimize free parameters in $\Psi_T$

- Variational Monte Carlo

$$E_{var} = \frac{\int E_{local} \Psi_T^2 dr}{\int \Psi_T^2 dr} \approx \langle E_{local} \rangle_{walk} \geq E_0$$

- Better: Exploit zero variance property of  $E_{local}$  as  $\Psi_T$  approaches an eigenfunction:

$$E_{local}(r) = \Psi_T^{-1}(r) \hat{H} \Psi_T(r)$$

**Minimize**  $\langle E_{local}^2 \rangle - \langle E_{local} \rangle^2$

Valid for all eigenstates:

Optimisation toward excited states possible

Least squares type algorithm with iterative fixed samples

## 6.14 Selfconsistent DMC

- Let an unbiased DMC run explore space and settle into important wells  
⇒ Extension to minimization problems
- Collect histograms of the distribution  $\propto \Psi_0$
- Inspect histograms and develop a suitable analytical form for their representation
- Feed the result as  $\Psi_T$  into an improved DMC run

Very efficient for anisotropic cases  
Avoids bias in DMC calculation caused by bad choices of  $\Psi_T$

## 6.15 Ensemble Size Error

In the asymptotic regime we ideally expect to have a random walker ensemble representing  $\Phi_0\Psi_T$ , contributions from higher  $\Phi_k$  having died out.

$$\lim_{\tau \rightarrow \infty} \Psi(\vec{r}, \tau)\Psi_T \propto \Phi_0(\vec{r})\Psi_T \exp\{-(E_0 - E_{ref})\tau\}$$

Problem: **Fluctuations in a finite random walker ensemble**

$$[\Psi(\vec{r}, \tau)\Psi_T]^{finite} \propto (\Phi_0(\vec{r})\Psi_T + \delta(\tau)) \exp\{-(E_0 - E_{ref})\tau\}$$

with a **random perturbation**  $\delta(\tau)$  with zero mean.

The rms amplitude of  $\delta$  scales with  $n_{walk}^{-1/2}$ .  $\delta$  can be expanded in the eigenfunctions  $\Phi_k, k > 0$ .

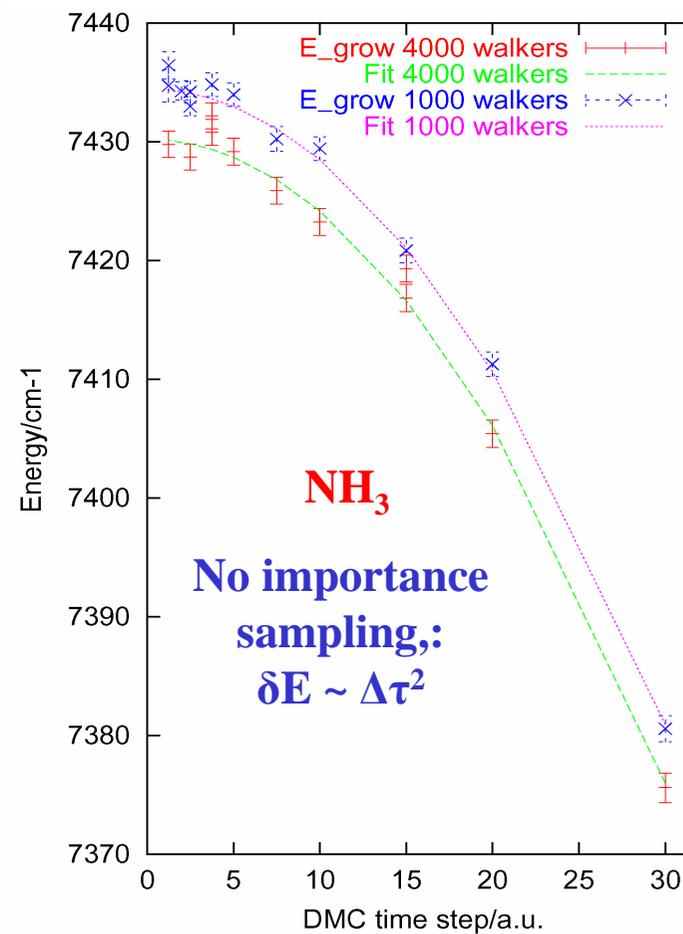
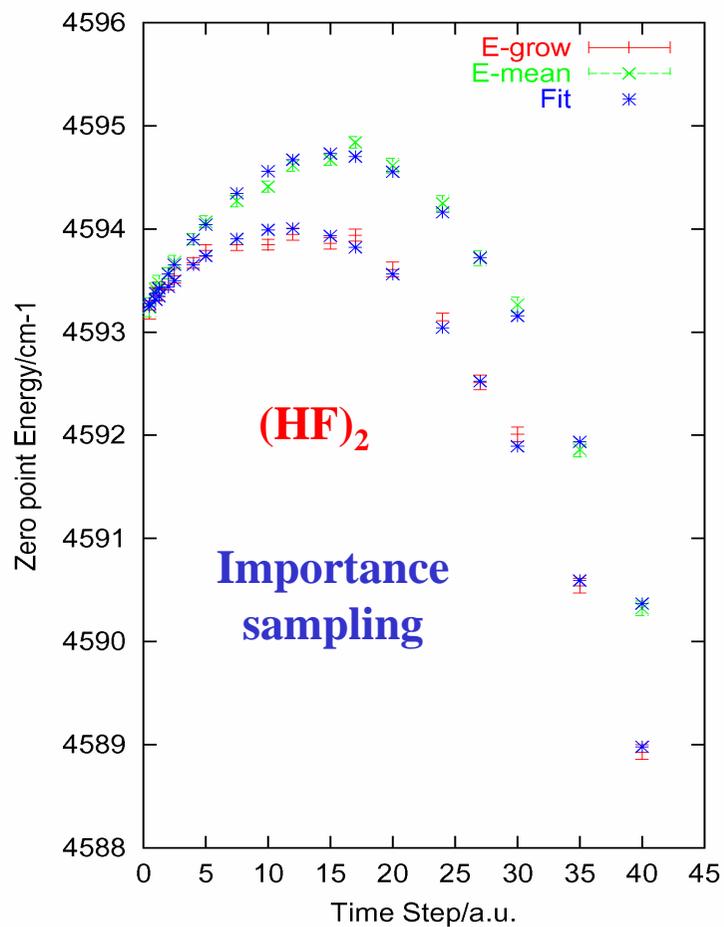
$$\delta = \frac{1}{\sqrt{n_{walk}}} \sum_k^{\infty} c_k \Phi_k$$

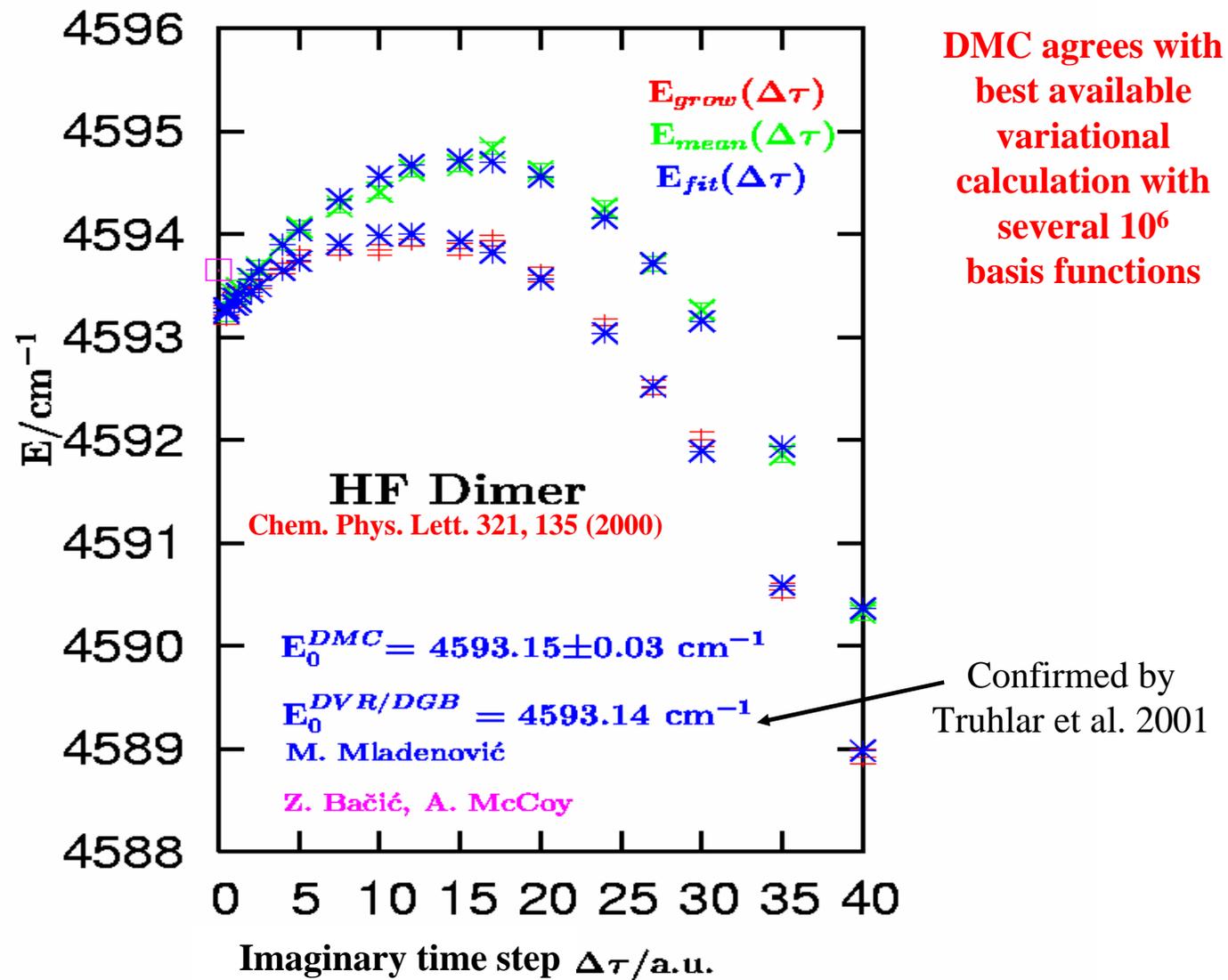
The energy expectation value should then take the form:

$$E(n_{walk}) = E_0 + \frac{1}{n_{walk}} \sum_k^{\infty} c_k^2 E_k$$

- The ensemble size bias is **always positive** and scales as  $1/n_{walk}$ .
- It can be suppressed by any noise reduction technique, e.g. a good trial wave function, but can be large for simple DMC without  $\Psi_T$ .

## Time step and ensemble size errors





## 6.16 Evolution of Random Walker Weights

After  $n$  steps the weight  $w_i$  of walker  $i$  with initial weight  $w_i(\tau = 0)$  (usually 1) is

$$w_i(n\Delta\tau) = w_i(0) \prod_{j=1}^n g_{ij} \quad ; \quad \ln w_i(n\Delta\tau) = \ln w_i(0) + \sum_{j=1}^n \ln g_{ij}$$

with **random growth factor**  $g_{ij}$  due to random position  $\vec{r}_{i;j}$  in step  $j$  (unless  $\Psi_T = \Phi_k$ ,  $E_{local}$  depends on  $\vec{r}_i$ ):

$$g_{ij} = \exp\left\{-\frac{\Delta\tau}{2}[E_{local}(\vec{r}_{i;j}) + E_{local}(\vec{r}_{i;j-1}) - 2E_{ref}]\right\}$$

- Weights  $w_i$  **grow or decay exponentially** and reflect the life history of random walker: Walkers which have explored regions of large  $E_{local}$  have accumulated low  $w_i$  and walkers which accidentally remained in regions of low  $E_{local}$  carry exponentially larger weights.
- The sum of random variables tends towards a Gaussian distribution (central limit theorem):  $g_{ij}$  and  $\ln g_{ij}$  are random variables and the asymptotic distribution of weights has to be log-normal!
- The width of the initial  $\delta$ -distribution of weights becomes exponentially large and the ensemble is dominated by a **few walkers with large  $w_i$**  (asymptotically one walker).  $\Rightarrow$  Unstable simulation with **increasingly noisy ensemble and time averages**.
- Loss of computational efficiency and creeping in of systematic **ensemble size errors**.

How to handle the instability?

## 6.17 Handling Random Walker Weights

### 6.17.1 Pure DMC

- Just **ignore the instability!**
- Requires **very good**  $\Psi_T$  with weight divergence slower than satisfactory statistical convergence of system properties.
- Automatically maintains a **constant ensemble size**.
- Clearly impossible without importance sampling.  
No restoring "quantum force"  $\Rightarrow$  diffusive spatial divergence of random walker ensemble.
- Commonly used in **electronic structure DMC** where sufficiently accurate  $\Psi_T$  **from conventional methods** are available.
- Keep your fingers crossed!

What to do when pure DMC does not work?

"Population control"

## 6.17.2 Population control I

The name of the game: Keep all weights in a reasonable range

This is an **artificial intervention** without motivation in the original diffusion equation and requires great care to **avoid systematic sampling biases**.

Method I (historical, Anderson 1975): Discrete weights

- Assign a standard weight  $w_s$  to each walker.
- Compute evolution of weight  $w_s$  using growth factor  $g_i$  for walker  $i$ :  $w_i = w_s g_i$
- Replace walker  $i$  with  $n_{walk}$  **copies** with identical properties and standard weight  $w_s$  using a uniform random number  $0 < u \leq 1$ .  
 $n_{walk} = \text{int}(w_i + u)$
- All information carried by **number density of random walkers**.
- Leads to **fluctuating ensemble size**. Nonuniform ensemble size errors; hard to deal with.
- Requires careful adjustment of  $E_{ref}$  to avoid collapse or explosion of ensemble size.
- Introduces **discretization noise** into the simulation.

### 6.17.3 Population control II: Continuous weights with branching/termination

- **Branching**: Replacing a walker with weight  $w_i$  by  $k$  walkers with weight  $w_i/k$  leaves both energy estimators unchanged:

$$W(\tau) = \sum_i w_i(\tau) \quad \rightarrow E_{grow}$$

$$E_{mean}(\tau) = \frac{1}{W(\tau)} \sum_i w_i(\tau) E_{local}(r_i, \tau)$$

- Branching whenever  $w_i$  exceeds an arbitrary threshold  $w_{max}$  avoids excessive weights.
- Branching does not introduce artificial noise but systematically increases the number of random walkers.

#### Termination to avoid ensemble explosion

- Walkers with small  $w_i$  contribute very little to the physically interesting property averages.  
 $\Rightarrow$  Eliminate walkers with small  $w_i$ .
- Small walkers tend to be in regions with above average  $E_{local} \Rightarrow$  Systematic elimination of small walkers biases both energy estimators towards lower values.

#### Invent unbiased termination rules:

Find rules which ensure that  $W(\tau)$  and  $E_{mean}(\tau)$  are conserved strictly or at least on average.

## 6.17.4 Combining small walkers

- Pick two small walkers with weights  $w_1$  and  $w_2$ .
- Assign their total weight  $w = w_1 + w_2$  to walker  $i = 1, 2$  with probability

$$p_i = \frac{w_i}{w_1 + w_2}$$

and eliminate the loser.

- $W(\tau)$  and thus  $E_{grow}$  are **strictly conserved**.

- $E_{mean}$  is conserved in an **average sense**:

$$\text{Before: } E_{mean}^{(1,2)} = w_1 E_{local}^{(1)} + w_2 E_{local}^{(2)}$$

$$\text{After: } \langle E_{mean}^{(1,2)} \rangle = w(p_1 E_{local}^{(1)} + p_2 E_{local}^{(2)}) = E_{mean}^{(1,2)}$$

- Minimal introduction of artificial noise.
- Provides true reversibility also for branching but cannot be generalized for correlated sampling.

## 6.17.5 Stochastic Ensemble Control

### One weight per walker:

1. Eliminate walkers with relative weight  $w < w_{min}$  with probability  $p$  (usually  $p = 1 - w$ ).
2. Increase the weight of a walker surviving an elimination attempt by  $w_+ = wp/(1 - p)$ .
3. Split walkers with high weight  $w > w_{max}$  into  $k$  walkers with weights  $w_{new} = w/k$ .

No systematic sources or sinks because  $-pw + (1 - p)w_+ = 0$

### Extension to multiple weights:

$p$  is in principle arbitrary  $\Rightarrow$  generalization

$$p = 1 - \frac{1}{N} \sum_{j=1}^N w_j = 1 - w_{mean}$$

- Explicit evaluation of new weights in all "states" if the walker remains in the ensemble ( $w_{new} = w/w_{mean}$ ).
- All walkers are active in all simultaneous simulations.
- Slowest possible decay of correlation, no systematic errors.

## 6.18 Expectation values

### 6.18.1 The meaning of the weights

- The solution function of the diffusion equation is the distribution  $\Phi_0\Psi_T$ .
- The number density of walkers in space is  $\propto \Psi_T^2$  if we handled the walks and the weights correctly (This is true for only a few implementations of the DMC algorithm!).
- Instantaneous weights  $w_i$  therefore provide a statistical measure for  $\Phi_0(r_i)/\Psi_T(r_i)$ .

Weighted property average yields "mixed" expectation value:

$$\langle \Phi_0 | \hat{A} | \Psi_T \rangle = \left\langle \frac{\sum_i w_i A(\vec{r}_i)}{\sum_i w_i} \right\rangle_\tau$$

Cheap extrapolation to exact expectation value:

$$\langle \Phi_0 | A | \Phi_0 \rangle = 2\langle \Phi_0 | A | \Psi_T \rangle - \langle \Psi_T | A | \Psi_T \rangle + \mathcal{O}(\Delta^2)$$

$$\Delta = \Phi_0 - \Psi_T$$

## 6.18.2 How to find a measure of $\Phi_0^2$

Can we just square the weights?

$$\langle A \rangle = \frac{\sum_i w_i^2 A(\vec{r}_i)}{\sum_i w_i^2}$$

The hard problem: we only have a stochastic representation of  $\Phi_0$  via the ratio  $\frac{\Phi_0}{\Psi_T}$ .

## 6.18.3 The noise problem

The weight  $w_i$  of a walker at position  $\vec{r}_i$  depends on the history of the walker and does not give an exact measure of  $\Phi_0(\vec{r}_i)/\Psi_T(\vec{r}_i)$ . The actual weight  $w_i$  has a **random deviation**  $\delta$  from the true ratio  $\Phi_0(\vec{r}_i)/\Psi_T(\vec{r}_i)$

$$w(\vec{r}_i) = \frac{\Phi_0(\vec{r}_i)}{\Psi_T(\vec{r}_i)} + \delta$$

Only the **imaginary time average** of weights of walkers visiting  $\vec{r}_i$  provides  $\Phi_0(\vec{r}_i)/\Psi_T(\vec{r}_i)$ .

$$\langle w(\vec{r}_i) \rangle_\tau = \left\langle \frac{\Phi_0(\vec{r}_i)}{\Psi_T(\vec{r}_i)} \right\rangle + \langle \delta \rangle_\tau = \frac{\Phi_0(\vec{r}_i)}{\Psi_T(\vec{r}_i)}$$

because of  $\langle \delta \rangle_\tau = 0$  for random noise.

## 6.18.4 The cheap and dubious way: Squaring the Weights

$$w^2(\vec{r}_i) = \frac{\Phi_0^2(\vec{r}_i)}{\Psi_T^2(\vec{r}_i)} + \delta^2 + 2\frac{\Phi_0(\vec{r}_i)}{\Psi_T(\vec{r}_i)}\delta$$

$$\langle w^2(\vec{r}_i) \rangle_\tau = \left\langle \frac{\Phi_0^2(\vec{r}_i)}{\Psi_T^2(\vec{r}_i)} \right\rangle_\tau + \langle \delta^2 \rangle_\tau + 2 \left\langle \frac{\Phi_0(\vec{r}_i)}{\Psi_T(\vec{r}_i)} \delta \right\rangle_\tau$$

$$\langle w^2(\vec{r}_i) \rangle_\tau = \frac{\Phi_0^2(\vec{r}_i)}{\Psi_T^2(\vec{r}_i)} + \langle \delta^2 \rangle_\tau$$

The average of the square is **biased by a positive quantity** whose amplitude depends on the noise.

We need a **2nd uncorrelated estimate** of  $\Phi_0(\vec{r}_i)/\Psi_T(\vec{r}_i)$ :

$$w'(\vec{r}_i) = \frac{\Phi_0(\vec{r}_i)}{\Psi_T(\vec{r}_i)} + \delta'$$

to evaluate the quantity

$$\langle w(\vec{r}_i)w'(\vec{r}_i) \rangle_\tau = \frac{\Phi_0^2(\vec{r}_i)}{\Psi_T^2(\vec{r}_i)} + \langle \delta\delta' \rangle_\tau$$

where  $\langle \delta\delta' \rangle_\tau = 0$  for uncorrelated random deviations  $\delta$  and  $\delta'$ .

**Combination with the number density of walkers ( $\propto |\Psi_T|^2$ ) gives the correct  $\Phi_0^2$  distribution.**

## 6.18.5 Getting it right: Descendant Weighting/Future Walking I

- The density of random walkers at any time origin is given by:

$$f(\vec{r}, 0) = \Psi(\vec{r}, 0)\Psi_T(\vec{r})$$

- Assume a single random walker at position  $\vec{R}$  at time 0.

$$f(\vec{r}, 0) = \delta(\vec{r} - \vec{R})$$

- Expand  $\Psi$  into eigenfunctions  $\Phi$ :

$$\delta(\vec{r} - \vec{R}) = \Psi_T(\vec{r}) \sum_k c_k \Phi_k(\vec{r})$$

- Multiply with  $\Phi_i/\Psi_T$  and integrate:

$$\int \delta(\vec{r} - \vec{R}) \frac{\Phi_i(\vec{r})}{\Psi_T(\vec{r})} d\vec{r} = c_i = \frac{\Phi_i(\vec{R})}{\Psi_T(\vec{R})}$$

- $c_0$  is a measure of the ratio  $\Phi_0/\Psi_T$  needed to transform the simulated distribution  $\Phi_0\Psi_T$  into the desired  $\Phi_0^2$ .

$$c_0(\vec{R}) = \frac{\Phi_0(\vec{R})}{\Psi_T(\vec{R})}$$

## 6.18.6 Descendant Weighting/Future Walking II

The asymptotic distribution of the random walks is

$$f(\vec{r}, \tau \rightarrow \infty) = c_0 \Phi_0(\vec{r}) \Psi_T(\vec{r}) \exp[-\tau(E_0 - E_{ref})]$$

The total population is given by the integral:

$$P(\tau) = \int f(\vec{r}, \tau) d\vec{r} = \sum_i w_i(\tau)$$

The contribution  $P_\infty(\vec{R})$  originating from an ancestral walker at position  $\vec{R}$  is:

$$\begin{aligned} P_\infty(\vec{R}) &= \int c_0(\vec{R}) \Phi_0(\vec{r}) \Psi_T(\vec{r}) \exp[-\tau(E_0 - E_{ref})] \\ &= \frac{\Phi_0(\vec{R})}{\Psi_t(\vec{R})} \exp[-\tau(E_0 - E_{ref})] \langle \Phi_0 | \Psi_T \rangle \end{aligned}$$

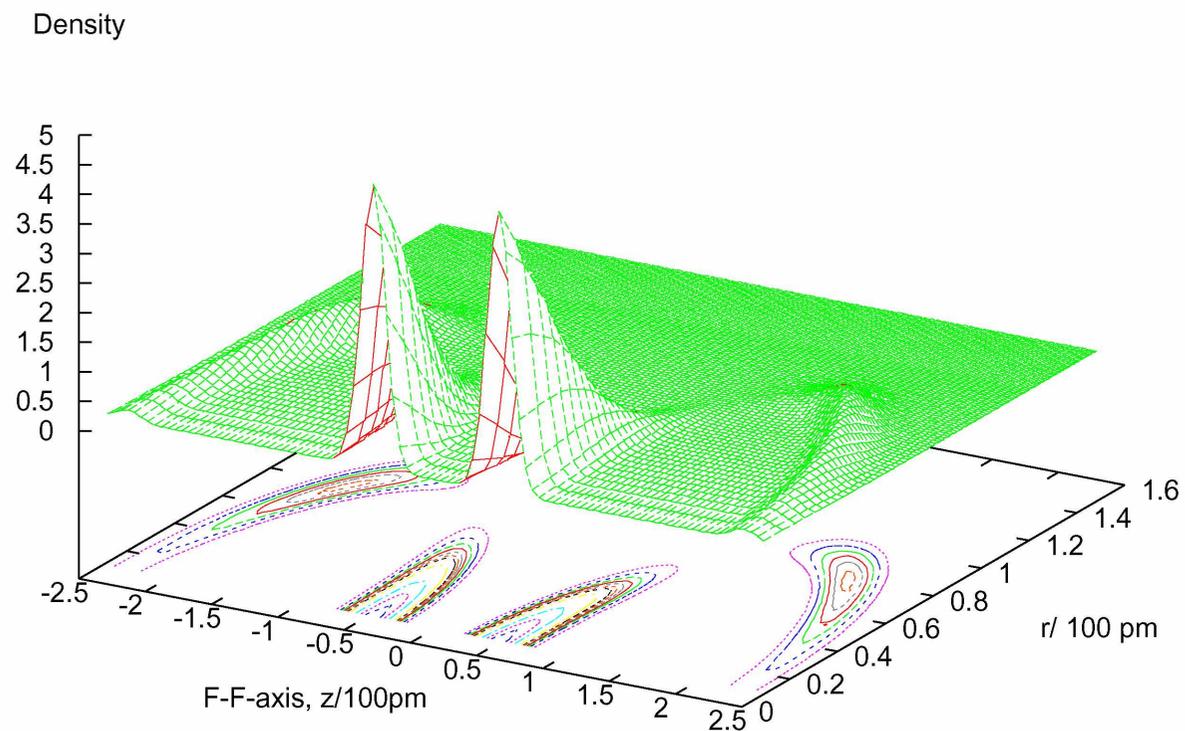
The asymptotic population originating from a given walker is proportional to the desired ratio  $\Phi_0(\vec{R})/\Psi_T(\vec{R})$ .

Since the other factors in  $P_\infty$  are identical for all walkers one can simply correct the weights by multiplication with the total weight of the asymptotic descendants.

This requires storing old coordinates and weights and a tagging algorithm for each branching event.

$$\langle \hat{A} \rangle \approx \frac{\sum_i A(\vec{R}_i, \tau) w_i(\vec{R}_i, \tau) P(\vec{R}_i, \tau)}{\sum_i w_i(\vec{R}_i, \tau) P(\vec{R}_i, \tau)}$$

## Hydrogen density in $(\text{HF})_2$ from DMC



## 6.18.7 Perturbational estimates for scalar expectation values

Add a perturbation  $A$  to the original hamiltonian  $H_0$  with eigenvalue  $E_0$  and eigenfunction  $\Phi_0$  according to

$$H_\lambda = H_0 + \lambda A$$

1st order perturbation theory gives for small  $\lambda A$

$$E_p(\lambda) \approx \langle \Phi_0 | H_0 + \lambda A | \Phi_0 \rangle = E_0 + \lambda \langle \Phi_0 | A | \Phi_0 \rangle$$

Use DMC to compute  $E_0$  and the eigenvalue  $E_\lambda$  of  $H_\lambda$  with eigenfunction  $\Phi_\lambda$ . For small  $\lambda$  we have

$$E_p(\lambda) = E_\lambda$$

$$\langle \Phi_0 | A | \Phi_0 \rangle = \frac{E_\lambda - E_0}{\lambda}$$

**Problem:** Extract a small energy difference from noisy data

**Solution:** Correlated sampling

Run correlated DMC on a **set of "potential surfaces"**  $V_\lambda = V_0 + \lambda A$  to get  $E_\lambda$ .

- Compute entire curves  $E_\lambda$  in a single simulation also for several  $A_i$ .
- Correlation is highest for small  $\lambda$ , just where we want to be.

## 6.18.8 More good news

### Origins of time step error

1. Neglect of commutators in split operator.
2. Low order integration of drift term.

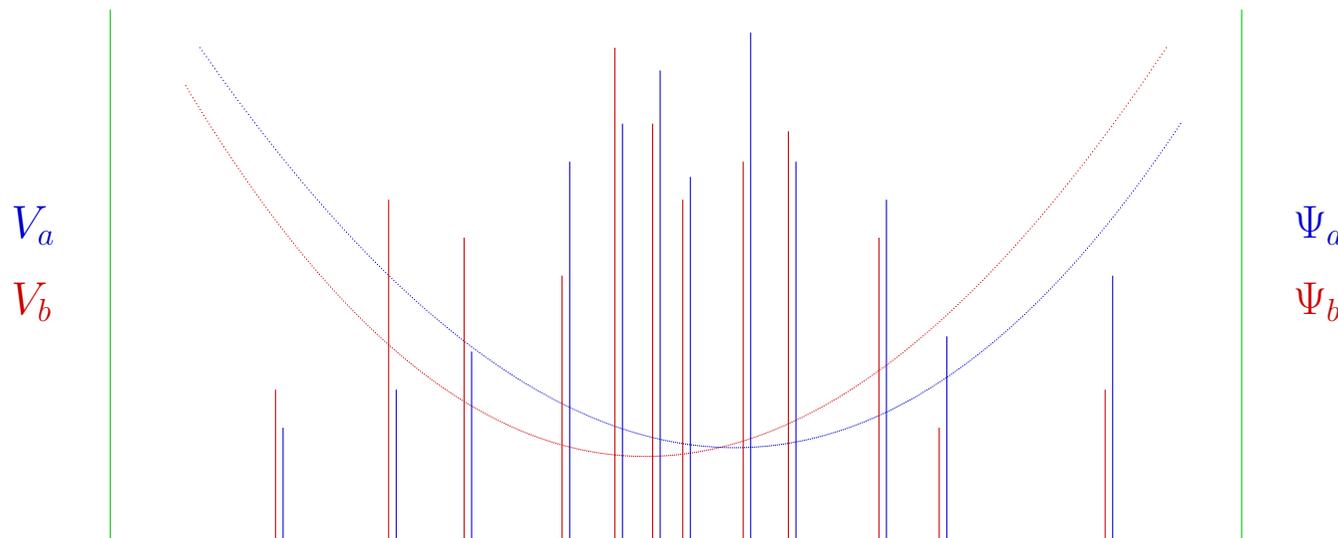
### Consequences for perturbation method:

$$E_\lambda(\Delta\tau) = E_\lambda + \sum_k c_{\lambda k} (\Delta\tau)^k$$

- Replacement of  $V_0$  by  $V_\lambda$  has only a small effect on  $c_2$ , all  $c_{\lambda 2} \approx$  equal.
- Linear time step error is strictly independent of  $\lambda$ . All  $c_{\lambda 1}$  are equal.
- The curves  $E_\lambda(\Delta\tau)$  are essentially parallel.
- **Energy differences  $E_\lambda - E_0$  are robust** wrt  $\Delta\tau$ .
- We can use a large time step and speed up the calculation.

## 6.19 Correlated Sampling

- Kinetic energy and  $\Psi_T$  control the random walkers position.
- Potential  $V$  affects only the weights  $w$ .



- **Simultaneous simulation** on several surfaces through introduction of a **set of weights**.
- Identical sampling positions cause **large covariance**.
- High accuracy for **energy difference sampling**.
- Computational savings (common terms in several potentials etc.).

## 6.20 Existing techniques which we have not addressed

- Approximations for excited states in special situations:
  - Adiabatic decompositions leading to rigid body constraints which can be handled by methods from classical molecular dynamics (SHAKE, quaternion techniques etc.)  
RB-DMC = Rigid body diffusion Monte Carlo
  - Beyond the adiabatic decomposition: allowing high frequency and low frequency modes to influence each other  
SB-DMC = Soft body diffusion Monte Carlo

Useful for the calculation of matrix shifts

- General approaches to excited states.
  - POITSE: Projection operator imaginary time spectral evolution  
very expensive, requires careful choice of the projection operator
  - CFMC: Correlation function Monte Carlo  
Uses diffusion Monte Carlo projections of wave functions to construct sequences of non-orthogonal basis sets which are injected into a diagonalisation method.  
Troubles with precision of matrix elements and resulting ill-conditioned matrices
- Finite temperature techniques.  
PIMC: Path integral Monte Carlo

## 6.21 Scaling behavior of DMC

### General computational issues

- Storage requirements only  $\propto N$ .
- Large cpu-requirements: typically  $> 10^8$  random walker steps.
- Very large volume of step by step data: only storage of histograms and running averages is feasible.

### Cost per step

- Time consuming steps are potential evaluation and trial function evaluation and derivatives
- Potential effort scales as in any other simulation method: e.g.  $N^2$  for pairwise models.
- Trial wave function effort depends on functional form, usually  $N^2$ .
- Long range cutoffs are possible in potential and  $\Psi_T$ . Larger cutoff radius than in classical simulations.
- Usually all particles move in each step. Propagation cost per step  $\propto N$ .
- Multiple 'time'-step techniques are not obvious.

## Statistical convergence

- Absolute error grows with total energy  $\propto N$ .
- Number of steps for similar absolute error  $\propto N^2$ .  
 $\Rightarrow$  Overall cost  $\propto N^3 - N^4$ .

## Parallelization issues

- All cpu-intensive parts are distributable and do not require communication.
- Communication only for average accumulation and ensemble control  $\propto N$ .
- Computation grows much faster than communication.  
Ratio can be adjusted within certain limits.
- Low I/O bandwidth.

## 6.22 Why is quantum Monte Carlo not more popular?

Top reasons why QMC is not generally used in chemistry:

[archive.ncsa.uiuc.edu/Apps/CMP/topten/topten.html](http://archive.ncsa.uiuc.edu/Apps/CMP/topten/topten.html) (D. M. Ceperley)

- We need forces, dummy!
- Try getting O<sub>2</sub> to bind at the variational level.
- How many graduate students lives have been lost optimizing wavefunctions?
- It is hard to get 0.01 eV accuracy by throwing dice.
- Most chemical problems have more than 50 electrons.
- Who thought LDA or HF pseudopotentials would be any good?
- How many spectra have you seen computed by QMC?
- QMC is only exact for energies.
- Multiple determinants. We can't live with them, we can't live without them.
- After all, electrons are fermions.
- Electrons move.
- QMC isn't included in Gaussian 90. Who programs anyway?

My personal additions from nuclear dynamics:

- We need good trial wave functions for arbitrary potentials
- We want many excited states.
- How to handle almost degenerate states?

There is room for a lot of future research!