# Optimum sampling in Quantum Monte Carlo

*J.R. Trail and R.J. Needs*

*Theory of Condensed Matter Group, Cavendish Laboratory,*

*University of Cambridge, UK*

*ESDG February 2007*

## Monte Carlo in Quantum Mechanics

- $N$-body QM is about solving integral problems in $3N$ dimensions
- Monte Carlo provides stochastic estimates of integrals

For example, for VMC we usually choose to:

- Take $r$ random samples in $3N$-d space, $\mathsf{R}$, from distribution $\psi^2(\mathbf{R})$
- Gives random variable $E_L(\mathsf{R}) = \psi^{-1}\hat{H}\psi$

Then Monte Carlo with this choice of sampling gives

$$\begin{aligned} \mathsf{Est}_r\left[\langle\psi|\hat{H}|\psi\rangle\right] &= \mathsf{Est}_r\left[\int \psi^2 E_L d\mathbf{R}\right] \\ &= \frac{1}{r}\sum_{i=1}^{r} E_L(\mathsf{R}_i) \end{aligned}$$

- The estimate provided is a *sample* value of a *random* variable

## What is its distribution?

Starts with samples from a random number generator of distribution $\propto \psi^2$

$\longrightarrow r$ samples of R

$\longrightarrow r$ samples of $E_L(\mathsf{R})$

$\longrightarrow 1$ sample of $\mathsf{Est}_r\left[E_{tot}\right] = 1/r \sum_{i=1}^{r} E_L$

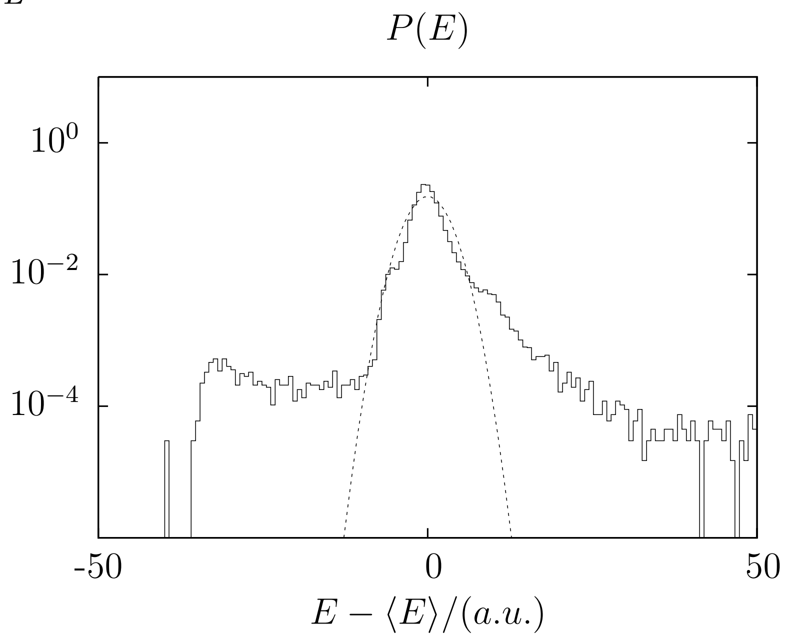
are all random variables, with different distributions

$r$ samples of R provide one sample of the total energy estimate

- Knowledge of its distribution is neccesary to obtain statistical error bars

## Central Limit Theorem

Distribution of $E_L$ is far from Gaussian:




CLT shows[a] that $1/r \sum_{i=1}^{r} E_L$ is a sample from a distribution that:

- is Gaussian (in large $r$ limit)
- has mean $\langle \psi | \hat{H} | \psi \rangle$
- has standard deviation $\sigma = \sqrt{\text{Var}[E_L]/r}$
- $68.26895$ % Confidence interval is $\pm\sigma$ about mean
- $99.99994$ % Confidence interval is $\pm 5\sigma$ about mean

[a] Except when it doesn't

## Is $\psi^2$ the 'best' distribution to sample with?

Previous two slides sampled R with distribution $\psi^2$ - 'standard sampling'

- We can sample with a wide range of functions of R - average $wE_L$ sampled with $\psi^2/w$
- Which function gives the smallest random error for $r$ samples - the optimum function ?
- How much better than standard sampling is it ?
- Does the CLT still apply for optimum sampling ?

## Incomplete solution...

- General sampling, use $P(\mathsf{R}) = \beta\psi^2/w$, with $\beta$ a normalising constant

$$\mathsf{Est}_r\left[E_{tot}\right] \quad = \quad \frac{1}{\beta}\frac{1}{r}\sum_{i=1}^{r} wE_L(\mathsf{R}_i)$$

- CLT gives standard deviation in estimate of $E_{tot}$

$$\sigma^2 = \frac{1}{r\beta^2}\left[\langle (wE_L)^2\rangle - \langle wE_L\rangle^2\right]$$

- use definition of averages and $\beta$ as integrals:

$$\Rightarrow \qquad r\sigma^2 = \int \psi^2/w d\mathbf{R} \int w\psi^2 E_L^2 d\mathbf{R} - \left[\int \psi^2 E_L d\mathbf{R}\right]^2$$

- Find stationary values w.r.t variations in function $w$:

$$\frac{\delta\sigma^2}{\delta w} = 0 \quad \Rightarrow \quad w = \frac{1}{|E_L|}$$

The solution?

## Incomplete solution...

Estimate of total energy:

$$\mathsf{Est}_r\left[E_{tot}\right] = \frac{1}{\beta}\langle \mathsf{sgn}(E_L)\rangle_r$$

Estimate of standard error in total energy:

$$\mathsf{Est}_r\left[\sigma^2\right] = \frac{1}{\beta^2} - \frac{1}{\beta^2}\langle \mathsf{sgn}(E_L)\rangle_r^2$$

- Optimum sampling depends on zero of energy
- $\beta$ is treated as a fixed variable, but is defined by

$$\frac{1}{\beta} = \int \psi^2 |E_L| d\mathbf{R}$$

so we need a MC estimate to evaluate it

- Most of the random error will be in $\beta$

$\Rightarrow$ This is *not optimum sampling*

## A more complete solution...

- Including the normalisation from the start, with unspecified $w$, gives

$$\mathsf{Est}_r\left[E_{tot}\right] = \frac{\langle wE_L\rangle_r}{\langle w\rangle_r}$$

- Same as standard sampling for $w = 1$

- Optimum sampling is the $w$ that minimises the random error in the estimate of this quotient

So, we need something more than the usual CLT, as this estimate of the total energy is a quotient of two sums of random variables

## Are they independent ?

Considering the distribution of the numerator and denominator separately,

$$\mathsf{Y} = \sum_{i=1}^{r} w(\mathsf{R}_i) E_L(\mathsf{R}_i) \qquad \mathsf{X} = \sum_{i=1}^{r} w(\mathsf{R}_i)$$

so

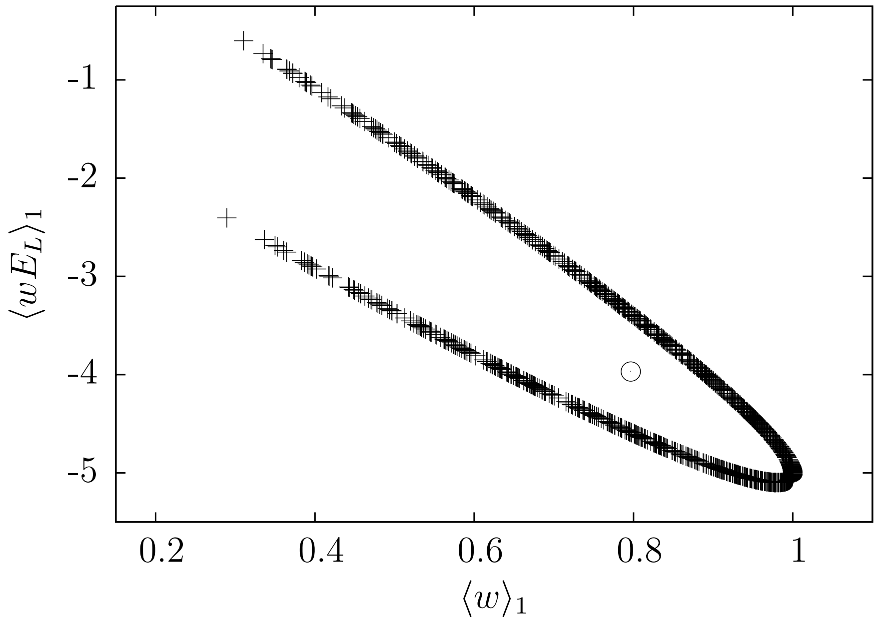
$$\mathsf{Est}_r\left[E_{tot}\right] = \frac{\mathsf{Y}}{\mathsf{X}}$$

Both $Y$ and $X$ depend of $\mathsf{R}$, so do they depend on each other?

For $r = 1$ they do - $1$ random $\mathsf{R}$ provide $2$ random variables

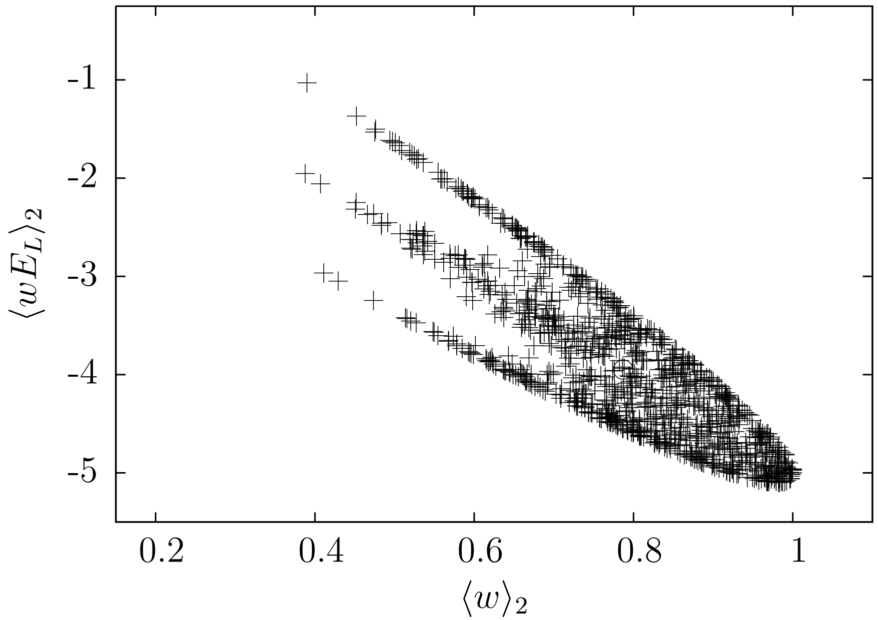
For $r > 1$ they do not - $r$ random $\mathsf{R}$s are reduced to $2$ random variables

## An example: r=1



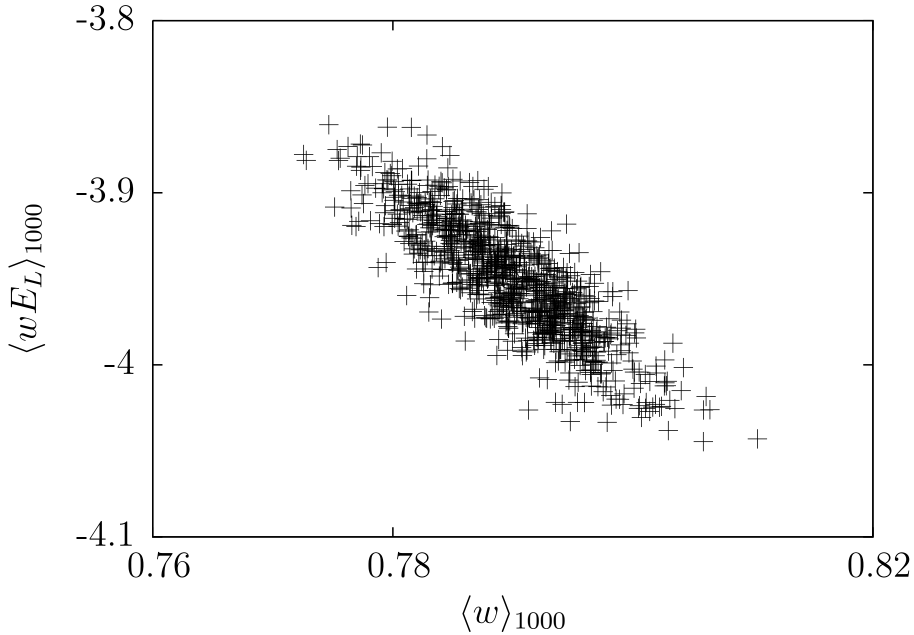

- $1000$ samples of $(Y, X)$ each constructed from $1$ sample of R

## An example: r=2




- $1000$ samples of $(Y, X)$ each constructed from $2$ samples of R

## An example: r=1000




- $1000$ estimates of $(Y, X)$ each constructed from $1000$ samples of R

## Bivariate CLT

- $Y$ and $X$ alone both have a Gaussian distribution
- only partial linear correlation remains between Y and X

$$P(y,x) = \frac{1}{2\pi} \frac{r^{1/2}}{|C|^{1/2}} \exp(-s^2/2)$$

$$s^2 = r \left[ c_{22}(x - \langle X \rangle)^2 - 2c_{12}(x - \langle X \rangle)(y - \langle Y \rangle) + c_{11}(y - \langle Y \rangle)^2 \right] / |C|$$

- $C$ is the co-variance matrix, with elements $c_{11}, c_{12}, c_{22}$
- This is analogous to the variance in $1d$
- $c_{12}$ measure the partial correlation between Y and X
- They can be estimated an analogous way to the variance for one variable

## Estimate and Error for arbitrary sampling

- Estimated total energy, for sampling with $P \propto \psi^2/w$

$$\mathsf{Est}_r\left[E_{tot}\right] = \frac{\langle wE_L\rangle_r}{\langle w\rangle_r} = \frac{\mathsf{Y}}{\mathsf{X}}$$

- Confidence intervals from bivariate Gaussian via Fieller's theorem:

$$l_{u,l} = \frac{(rYX - c_{12}) \pm \sqrt{(rYX - c_{12})^2 - (rX^2 - c_{11})(rY^2 - c_{22})}}{rX^2 - c_{11}}$$

For $68.3\%$ probability that $l_l < \mathsf{Est}_r[E_{tot}] < l_u$

- This is the equivalent of 'standard error = standard deviation$/\sqrt{r}$' for the distribution of gradients that give the estimate of the total energy

## Optimum Sampling

- What $w$ gives the smallest confidence interval for $r$ samples?

  Using definition of $C$, $Y$, $X$, size of confidence interval is

$$(l_u - l_l)^2 = \frac{q^2}{r} \frac{\int \psi^2 / w d\mathbf{R} \int w\psi^2 (E_L - \mathsf{Est}_r\left[E_{tot}\right])^2 d\mathbf{R}}{\left[\int \psi^2 d\mathbf{R}\right]^2}$$

- Use the $\mathsf{Est}_r\left[E_{tot}\right]$ that we estimate, with mean $E_{tot}$ and variance $(l_u - l_l)^2/4$
- Take the mean of the equation for $(l_u - l_l)^2$

$$\left[\int \psi^2 d\mathbf{R}\right]^2 \frac{r}{q^2}(l_u - l_l)^2 = \int \frac{\psi^2}{w} d\mathbf{R} \int w\psi^2 \left[(E_L - E_{tot})^2 + (l_u - l_l)^2/4\right]$$

- Find stationary values w.r.t variations in function $w$:

$$\frac{\delta (l_u - l_l)^2}{\delta w} = 0 \quad \Rightarrow \quad w = \frac{1}{\left[(E_L - E_{tot})^2 + (l_u - l_l)^2/4\right]^{1/2}}$$

## Optimum Sampling

- Sample $3N$-d space with $P \propto \psi^2/w$, where

$$w = \frac{1}{\left[(E_L - E_0)^2 + \epsilon^2\right]^{1/2}}$$

where $E_0$ is an estimate of the total energy, and $\epsilon$ is its standard error

- Independent of zero of energy
- No unknown normalisation constants

To apply it in practice:

- Estimates of $E_0$ and $\epsilon$ required, but these do not bias results
- or self consistency

## Theoretical limit

- Results provides a theoretical limit to how small the statistical error can be for $r$ samples (in the large $r$ limit)
- For standard sampling, the $68.3\%$ confidence interval is $2\epsilon$ in size, with

$$\epsilon = \frac{1}{\sqrt{r}}\left[\int \psi^2 (E_L - E_{tot})^2 d\mathbf{R}\right]^{1/2}$$

- For optimum sampling, the $68.3\%$ confidence interval is $2\epsilon$ in size, with

$$\epsilon = \frac{1}{\sqrt{r}}\int \psi^2 |E_L - E_{tot}| d\mathbf{R}$$

- For a Gaussian distribution of $E_L$ optimum sampling error is $0.8\times$ standard sampling error
- This ratio will get smaller the ‘fatter’ the distribution of $E_L$

## A final point....

- We could sample with $P = \delta(E_L - E_{tot})$, which gives the exact answer for one sample
- This is not a stochastic estimate, since exact $E_{tot}$ is required exactly

## What does optimum sampling give us ?

- Provides a theoretical limit to random error as a function of $r$
- Provides a measure of how close to optimum any sampling strategy is
- Bivariate analysis is applicable to most QMC methods
- Applicable to estimates of any operator
- If the CLT is invalid for standard sampling, then optimum sampling should reinstate it

What does it not tells us...

- Relative computational cost

Acknowledgements

*Financial support was provided by the Engineering and Physical Sciences Research Council (EPSRC), UK, and computational resources were provided by the Cambridge High Performance Computing Facility.*