Some Estimates in Variational Quantum Monte Carlo

J.R. Trail and R. Maezono

School of Information Science, JAIST,

Nomi, Ishikawa 923-1292, Japan

July 2009

- Variational Monte Carlo

- 'Standard' VMC is $P = \psi^2$, and failure of CLT is an artifact of this

- Conditions and expressions for Normally distributed estimates in general sampling

- **Efficient** sampling Monte Carlo implemented - not sampling from $P = \psi^2$

- Estimates for energies, excitation energies, ionisation energies, transition moments

- Results for first row atoms and some molecules

VMC and General Sampling

Sample with $P = \psi^2/w$ and construct an estimate:

$$\overline{\mu} = \frac{\sum w_i E_L(\mathbf{R}_i)}{\sum w_i}$$

- If both variances exist, Fieller's theorem tells us this is a sample from a Normal distribution with:

$$\mu = \frac{\int \psi^2 E_L d\mathbf{R}}{\int \psi^2 d\mathbf{R}} \quad , \quad \sigma^2 = \frac{1}{r} \frac{\int \psi^2/w d\mathbf{R} \int w \psi^2 (E_L - \mu)^2 d\mathbf{R}}{\left[\int \psi^2 d\mathbf{R}\right]^2}$$

- We can estimate the variance:

$$\overline{\sigma}^2 = \frac{r}{r-1} \frac{\sum w_i^2 \left(E_L(\mathbf{R}_i) - \overline{\mu}\right)^2}{\left(\sum w_i\right)^2}$$

- $\overline{\sigma}^2 \neq (\text{sample variance})/r$

- These equations do not follow from the usual (univariate) Central Limit Theorem

- *Zero Variance Principle* is still valid - for exact $\psi \Rightarrow \overline{\sigma} = 0$

$\rightarrow$ The error is **controlled** if the bivariate CLT is **valid** and $\langle w \rangle \neq 0$

Trail JR, Phys. Rev. E. **77**, 016703,016704 (2008)

When is it Normal ?

For standard sampling $P = \psi^2$ ($w = 1$)

● Normal for standard sampling and total energy ($P \propto 1/x^4$)

● Not Normal for standard sampling used with correlated sampling, forces, and many other estimates

● Different choices of P (equivalently w) are possible

● Changes computational cost: flops for evaluating P

● Changes distribution of random errors

● Failure of CLT from singularities in averaged quantities on the nodal surface

For some P (equivalently w) the variance is infinite, and CLT is invalid

$\rightarrow$ Goal is to improve efficiency and reinstate the CLT where it is invalid for standard sampling

Why is Normality so important ?

We want Normally distributed estimates, from a general form of the Central Limit Theorem

If the distribution the estimates are drawn from is *not* normal then it is a Stable Law :
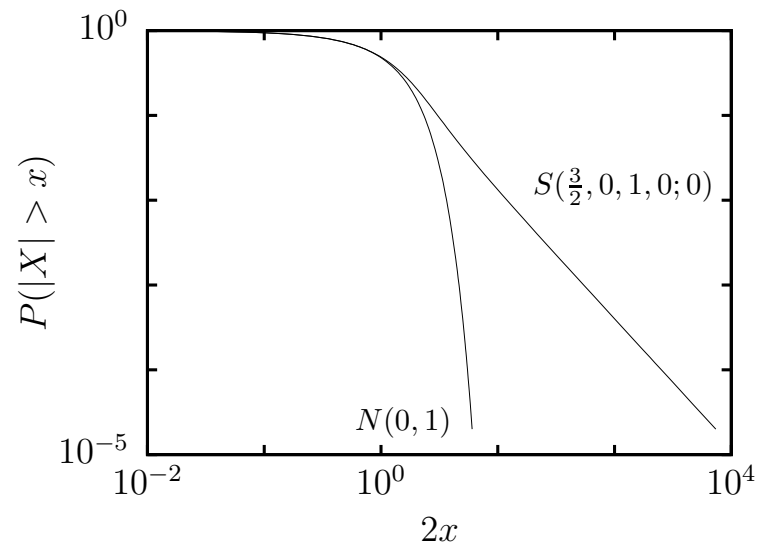
- Normal distribution and an example Stable law

- Probability that a sample fall outside of central interval size $2x$

- Width parameter is representative of error for Normal, not for Stable

- Width parameter is estimateable for Normal - sample standard error

- Width parameter is *not* estimateable for Stable - sample standard error is unrelated

Efficient sampling in VMC

- Draw position vectors from $P = |D_1|^2 + |D_2|^2$ [a], no Jastrow, Backflow

$\rightarrow$ Zero on coalescence planes only, non-zero on rest of nodal surface

- Perform Metropolis accept/reject with P, and use $w = \psi^2/P$

- Using r samples, the Bivariate CLT, and Fiellers theorem provides the Normal estimate

$$\overline{E}_{tot} = \frac{\sum w_i E_L(\mathbf{R}_i)}{\sum w_i} \quad , \quad \overline{\sigma}^2 = \frac{r}{r-1} \frac{\sum w_i^2 \left(E_L(\mathbf{R}_i) - \overline{E}_{tot} \right)^2}{\left(\sum w_i \right)^2}$$

- (wE_L, w) has no singularities and is bounded $\Rightarrow$ all moments exist $\Rightarrow$ distribution is Normal
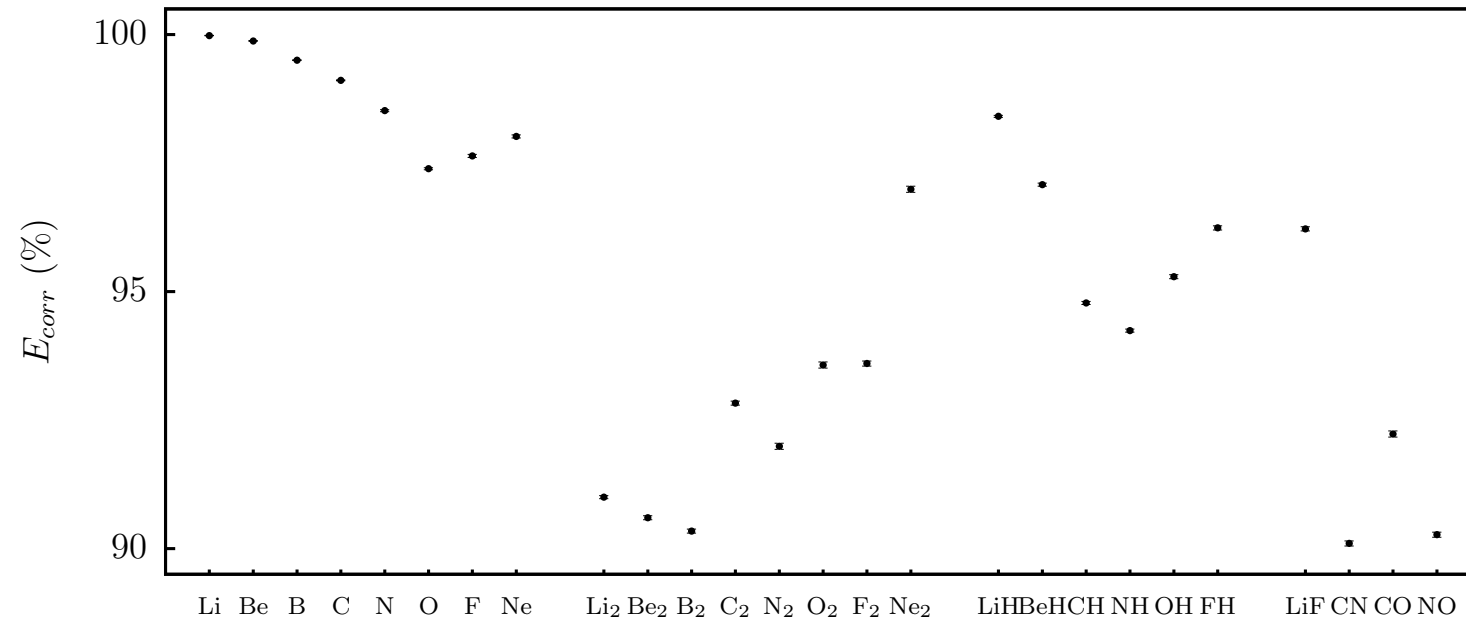
---

[a]Trail JR and Maezono R, JCP (2010)

Efficient sampling in VMC

Optimisation and estimation of total energy:

- All-electron

- First row atoms + some diatomic molecules

- Numerical orbitals from ATSP2K and 2DHF (MCSCF and HF)

- Jastrow, Backflow, and $5 - 86$ CSFs

- 48 h desktop time/system

- r for Final estimate:monitor estimate:optimisation $r = 9000 : 150 : 1$

Efficient sampling in VMC

- Improved results for less computational effort

- For a given error $5 - 35\times$ faster than standard sampling

- Surfaces for energy optimisation are Normal

- Do we need anisotropic Jastrow/Backflow ?

Distribution of more general estimates

What is the distribution for more complex estimates ?

Example: Energy differences

$$\Delta E_{tot} = \frac{\int \psi_1^2 E_1 d\mathbf{R}}{\int \psi_1^2 d\mathbf{R}} - \frac{\int \psi_2^2 E_2 d\mathbf{R}}{\int \psi_2^2 d\mathbf{R}}$$

• Sample with $P = |D_1|^2 + |D_2|^2$ (no nodal surface, only coalescence planes)

$$\begin{aligned} \mathsf{Est}\left[\Delta E_{tot}\right] &= \frac{\sum w_1 E_1}{\sum w_1} - \frac{\sum w_2 E_2}{\sum w_2} \\ &= \frac{\mathsf{S}_2}{\mathsf{S}_1} - \frac{\mathsf{S}_4}{\mathsf{S}_3} \end{aligned}$$

• Elements in sums S_n are correlated only when in same 'time slots'

$\Rightarrow$ Multivariate CLT is true

$\Rightarrow$ Each S_n is Normal

$\Rightarrow$ Correlation between S_n and S_m is linear

Algebra for sums of random variables

What is the distribution of

$$\Delta \mathsf{E}_{tot} = \frac{\mathsf{S}_2 \mathsf{S}_3 - \mathsf{S}_1 \mathsf{S}_4}{\mathsf{S}_1 \mathsf{S}_3} \quad , \quad \mathsf{S}_n = \sum_i X_n(i)$$

with parameters expressed in terms of the estimateable

$$
\begin{aligned}
\mathbb{E}\left[\mathsf{X}_n(i)\right] &= \mu_n \\
Var\left[\mathsf{X}_n(i)\right] &= C_{nn} \\
Cov\left[\mathsf{X}_m(i), \mathsf{X}_n(j)\right] &= C_{nm}\delta_{i,j}
\end{aligned}
$$

?

Algebra for sums of random variables : Sums

Adding sums of the random variables, trivial to show that:

● Distribution is Normal

$$
\begin{aligned}
\mathbb{E}\left[\mathsf{S}_1 + \mathsf{S}_2\right] &= r(\mu_1 + \mu_2) \\
Var\left[\mathsf{S}_1 + \mathsf{S}_2\right] &= r(C_{11} + 2.C_{12} + C_{22}) \\
Cov\left[\mathsf{S}_1 + \mathsf{S}_2, \mathsf{S}_3 + \mathsf{S}_4\right] &= r(C_{13} + C_{14} + C_{23} + C_{24})
\end{aligned}
$$

$\rightarrow$ All the $\mathsf{S}_1 + \mathsf{S}_2$ are Normal

$\rightarrow$ Correlation between any $\mathsf{S}_j + \mathsf{S}_k$ and $\mathsf{S}_l + \mathsf{S}_m$ is linear

Algebra for sums of random variables : Quotients

Quotients of sums of the random variables, Fiellers theorem provides:

- Distribution is Normal

$$
\mathbb{E}\left[\frac{S_2}{S_1}\right] = \frac{\mu_2}{\mu_1}
$$

$$
Var\left[\frac{S_2}{S_1}\right] = \frac{1}{r}\frac{1}{\mu_1^2}\left[C_{22} - 2\frac{\mu_2}{\mu_1}C_{12} + \left(\frac{\mu_2}{\mu_1}\right)^2 C_{11}\right]
$$

- No expression for correlation of different quotients

So,

$$
\frac{S_1 + S_2}{S_3 + S_4}
$$

is Normal and parameters are estimateable

Algebra for sums of random variables : Products

Products of sums of the random variables:

- What is the bivariate distribution of $(S_1 S_2, S_3 S_4)$?

- Derive co-moments and compare with bivariate Normal....define $\Delta X_1(i) = X_1(i) - \mu_1$

$$
\begin{aligned}
\Delta(S_1 S_2) &= S_1.S_2 - \mathbb{E}[S_1.S_2] \\
&= \sum_{ij} \Delta X_1(i) \Delta X_2(j) + r \sum_i [\mu_1 \Delta X_2(i) + \mu_2 \Delta X_1(i)] - rC_{12}
\end{aligned}
$$

- Co-moments are defined by $\mu_{m,n} = \mathbb{E}[\Delta(S_1 S_2)^m.\Delta(S_3 S_4)^n]$

$$
\mu_{m,n} = \mathbb{E}\left[\left(\sum_{ij} \Delta X_1(i)\Delta X_2(j) + r\sum_i [\mu_1 \Delta X_2(i) + \mu_2 \Delta X_1(i)] - rC_{12}\right)^m \right.
$$

$$
\left. \times \left(\sum_{ij} \Delta X_3(i)\Delta X_4(j) + r\sum_i (\mu_3 \Delta X_4(i) + \mu_4 \Delta X_3(i)] - rC_{34}\right)^n\right]
$$

- Multiply out and count equivalent terms

- Pick out the dominant r terms.

Algebra for sums of random variables : Products

Powers of r in each term after multiplication:

- 2nd order part: Powers p_1, q_1 from equivalence, none from prefactor

- 1st order part: Powers p_2, q_2 from equivalence, p_2, q_2 from prefactor

- 0th order part: Powers p_3, q_3 from equivalence, p_3, q_3 from prefactor

- Multiplication gives $p_1 + p_2 + p_3 = m$, $q_1 + q_2 + q_3 = n$

- Count powers of r from prefactors $[r]$, and from sums (r)

$\Rightarrow$ Terms have powers of r:

$$[1]^{p_1} (r)^{p_1} [r]^{p_2} (r)^{p_2} [r]^{p_3} (1)^{p_3} . [1]^{q_1} (r)^{q_1} [r]^{q_2} (r)^{q_2} [r]^{q_3} (1)^{q_3} \mathbb{E} [\ldots]$$

But Expectations are zero if any index is unique:

$$\mathbb{E} \left[\Delta \mathsf{X}_1(i) . \Delta \mathsf{X}_2(j) . \Delta \mathsf{X}_3(k) \cdots \right] = \mathbb{E} \left[\Delta \mathsf{X}_1(i) \right] . \mathbb{E} \left[\Delta \mathsf{X}_2(j) . \Delta \mathsf{X}_3(k) \cdots \right]$$
$$= 0$$

- Overcounting due to including zero expectations values

Algebra for sums of random variables : Products

Count only terms with no unique indices:

$$r^{(m+n)+(p_2+q_2)/2}\mathbb{E}\left[\ldots\right] \quad \text{for } (p_2 + q_2) \text{ even}$$

$$r^{(m+n)+(p_2+q_2+1)/2}\mathbb{E}\left[\ldots\right] \quad \text{for } (p_2 + q_2) \text{ odd}$$

$\Rightarrow$ In large r limit $p_2 + q_2 = m + n$ terms dominate ($p_1 = p_3 = q_1 = q_3 = 0$)

$\Rightarrow$ In large r limit the co-moments are same as those for *sums*:

$$\mathbb{E}\left[\Delta(\mathsf{S}_1\mathsf{S}_2)^m.\Delta(\mathsf{S}_3\mathsf{S}_4)^n\right] = r^{m+n}\mathbb{E}\left[\left(\sum_i \left[\mu_1\Delta\mathsf{X}_2(i) + \mu_2\Delta\mathsf{X}_1(i)\right]\right)^m \times \left(\sum_i \left[\mu_3\Delta\mathsf{X}_4(i) + \mu_4\Delta\mathsf{X}_3(i)\right]\right)^n\right]$$

Algebra for sums of random variables : Products

- In large r limit $(\mathsf{S}_1\mathsf{S}_2, \mathsf{S}_3\mathsf{S}_4)$ is bivariate Normal with

$$
\begin{aligned}
\mathbb{E}[\mathsf{S}_1\mathsf{S}_2] &= r^2\mu_1.\mu_2 \\
Var[\mathsf{S}_1\mathsf{S}_2] &= r^3(\mu_1^2 C_{22} + 2\mu_1\mu_2 C_{12} + \mu_2^2 C_{11}) \\
Cov[\mathsf{S}_1\mathsf{S}_2, \mathsf{S}_3\mathsf{S}_4] &= r^3(\mu_1\mu_3 C_{24} + \mu_1\mu_4 C_{23} + \mu_2\mu_3 C_{14} + \mu_2\mu_4 C_{13})
\end{aligned}
$$

$\rightarrow$ We have rules for obtaining the distribution of combinations of sums of random variables (that are Normal)

Energy differences

$$\mathsf{Est}\left[\Delta E_{tot}\right] = \frac{\sum w_1 E_1}{\sum w_1} - \frac{\sum w_2 E_2}{\sum w_2}$$

$$= \frac{\mathsf{S_2 S_3 - S_1 S_4}}{\mathsf{S_1 S_2}}$$

$$\overline{\Delta E}_{tot} = \frac{\overline{\mu}_2}{\overline{\mu}_1} - \frac{\overline{\mu}_4}{\overline{\mu}_3}$$

$$\frac{r-1}{r}\overline{\sigma}^2 = \frac{\sum w_1^2(E_1 - \overline{\mu}_1)^2}{\left[\sum w_1\right]^2} - 2\frac{\sum w_1 w_2(E_1 - \overline{\mu}_1)(E_2 - \overline{\mu}_2)}{\left[\sum w_1\right]\left[\sum w_2\right]} + \frac{\sum w_2^2(E_2 - \overline{\mu}_2)^2}{\left[\sum w_2\right]^2}$$

Sufficient conditions for this estimate to be Normal are

● All moments exist

● All means in the denominator are non-zero

● At least one mean in each product on numerator is non-zero
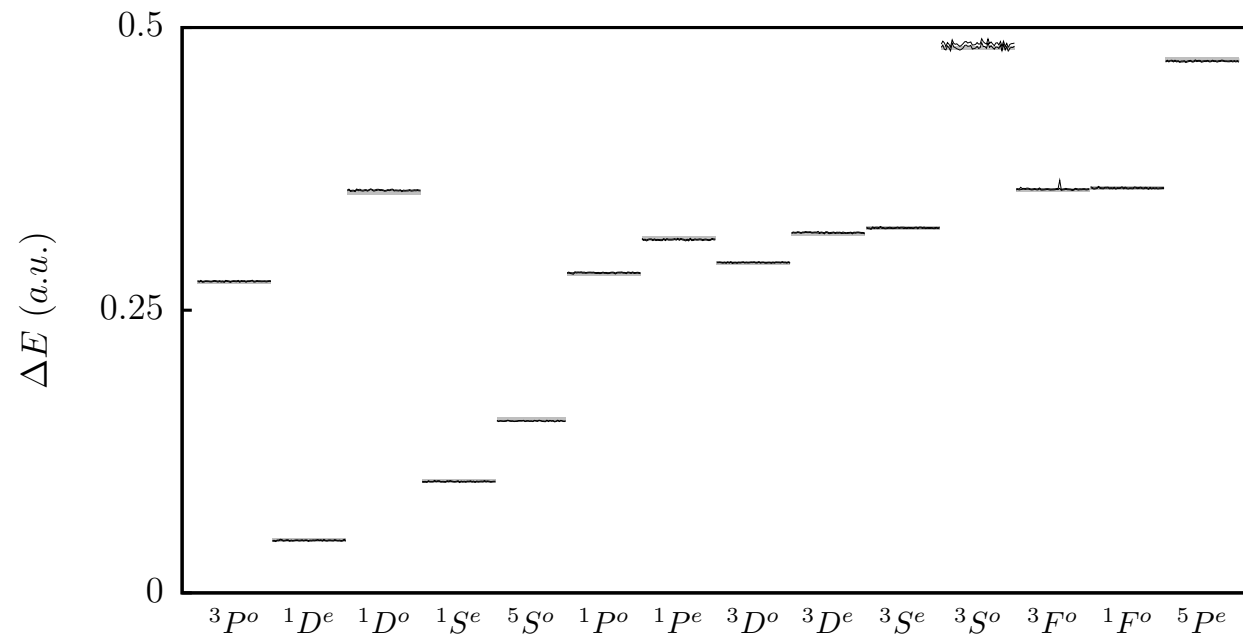
Carbon atom excitation energies

- Calculation as for GS (Multideterminant and numerical orbitals from MCSCF ~ 300 parameters)

- $E_{tot}\left[^{2S+1}L\right] - E_{tot}\left[^3P\right]$

- Lowest two eigenstates for each Term

- Inversion symmetry conserved by Jastrow and Backflow

- Term *approximately* conserved by introduction of Jastrow and Backflow

- Energy minimisation approximately valid for lowest two energies for each Term

- Sample with $P = D_1\left[^{2S+1}L\right]^2 + D_2\left[^{2S+1}L\right]^2 + D_1\left[^3P\right]^2 + D_2\left[^3P\right]^2$

Carbon atom excitation energies
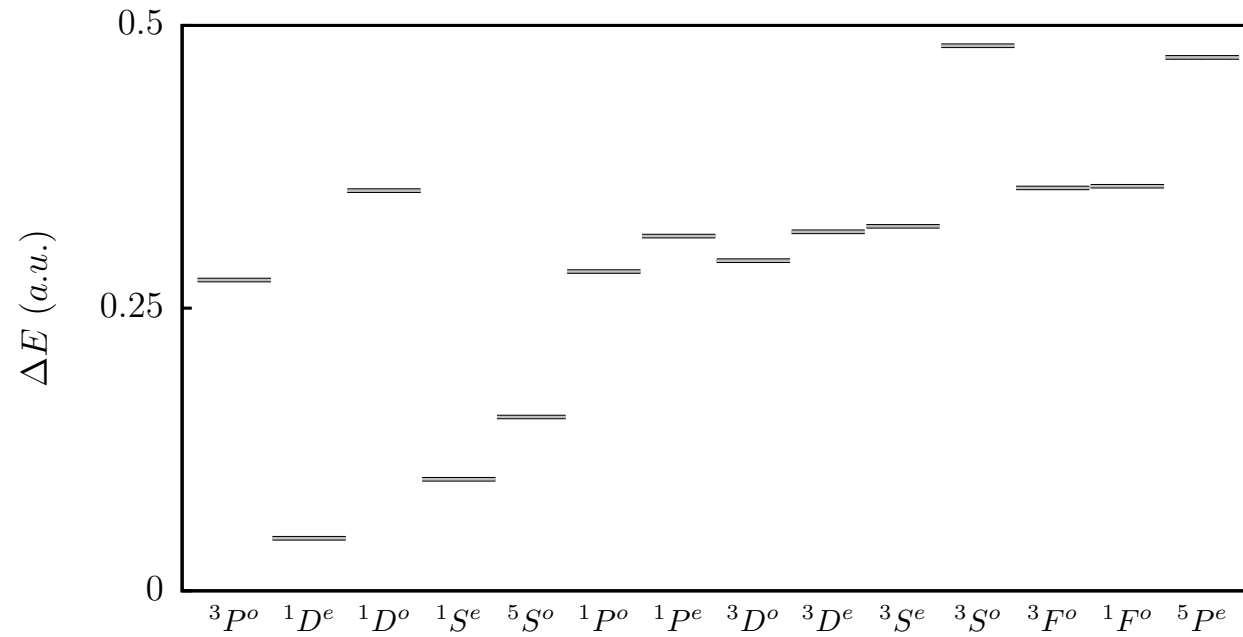
- C $^{2S+1}L \rightarrow$ C ^3P

- Grey: experimental Spectroscopic values $\pm$ 'chemical accuracy'
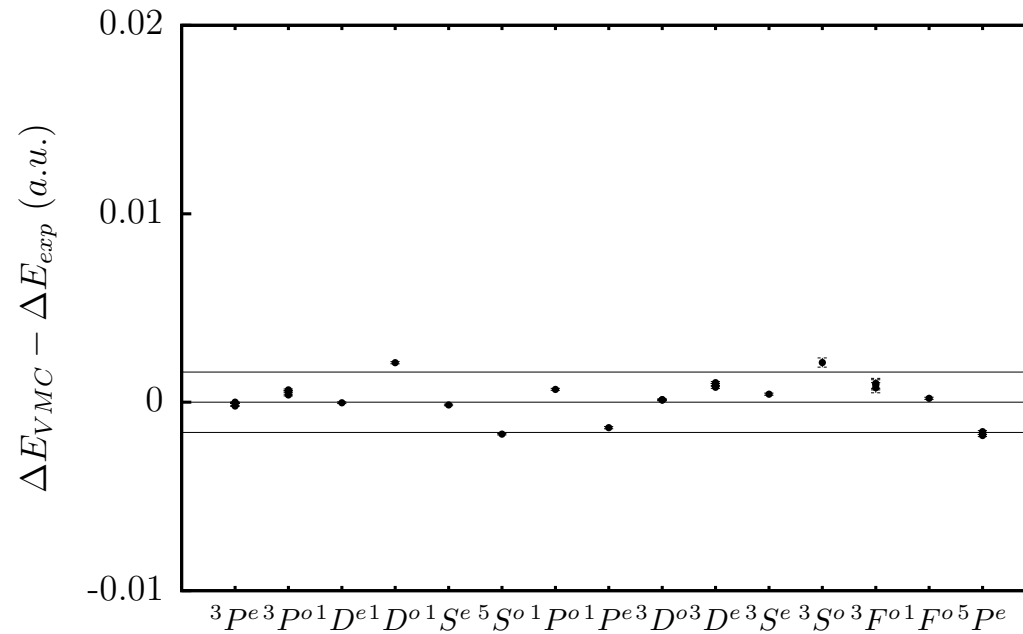
Carbon atom excitation energies

- C $^{2S+1}L \to$ C ^{3}P

- Grey: experimental Spectroscopic values $\pm$ 'chemical accuracy'

- 48 energy difference estimates for each excitation

Error in Carbon atom excitation energies
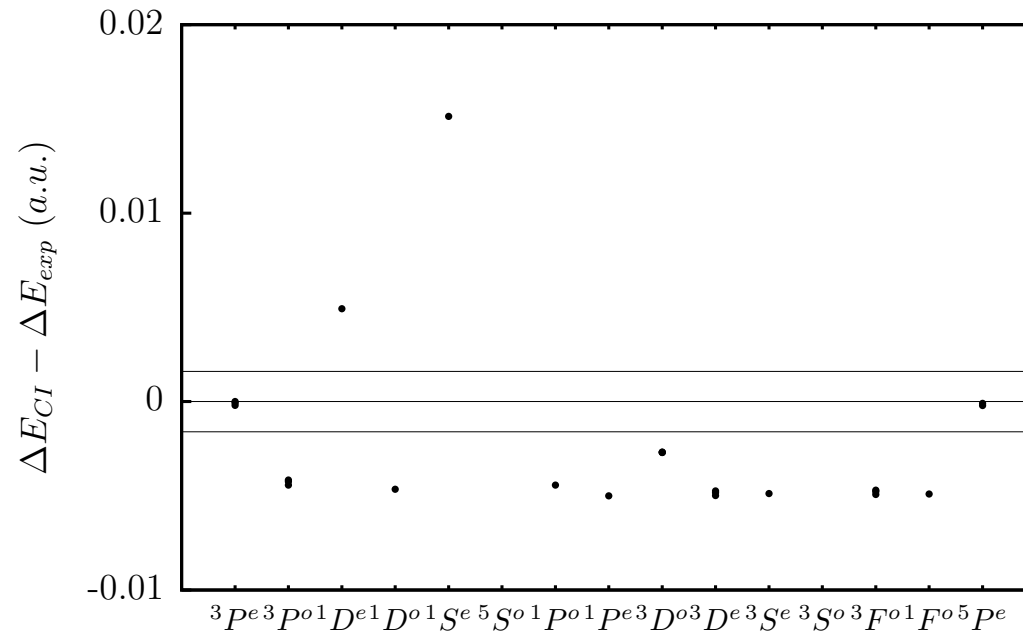
- CI results used for allocating transitions to lines

- CI AS and orbitals chosen empirically to reproduce spectrosopy

Error in Carbon atom excitation energies

$\Delta E_{VMC} - \Delta E_{exp}$ (a.u.) versus excitation states $^3P^e\,^3P^o\,^1D^e\,^1D^o\,^1S^e\,^5S^o\,^1P^o\,^1P^e\,^3D^o\,^3D^e\,^3S^e\,^3S^o\,^3F^o\,^1F^o\,^5P^e$

- Chemical accuracy from VMC + efficient sampling

- *NOT* spectroscopic accuracy

- Estimate of difference *not* difference of estimates

- Correlation reduces error by $10 - 70\%$

Ionization energies

- How do we deal with changes in electron number?

$$
\begin{aligned}
E_{ion} &= E_n - E_{n-1} \\
&= \frac{\int \psi_1^2 E_1 d\mathbf{R}_n}{\int \psi_1^2 d\mathbf{R}_n} - \frac{\int \psi_2^2 E_2 d\mathbf{R}_{n-1}}{\int \psi_2^2 d\mathbf{R}_{n-1}}
\end{aligned}
$$

- Sampled $P_n(\mathbf{R}_n) = D_1(n)^2 + D_2(n)^2$

- For estimating E_{n-1} ignore *one* $3d$ sample vector, so $\mathbf{R}_n \to \mathbf{R}_{n-1}$

- Distribution of $(w_2 E_2, w_2)$ given by integrating P_n analytically

$$
\begin{aligned}
P_{n-1} &= \int D_1(n)^2 + D_2(n)^2 d^3\mathbf{r}_n \\
&= \int \left[\phi_1(\mathbf{r}_n).C_1(1,n) + \phi_2(\mathbf{r}_n).C_1(2,n) + \ldots\right]^2 + \left[\phi_1(\mathbf{r}_n).C_2(1,n) + \phi_2(\mathbf{r}_n).C_2(2,n) + \ldots\right]^2 d^3\mathbf{r}_n \\
&= \left[C_1(1,n)^2 + C_1(2,n)^2 + \ldots\right] + \left[C_2(1,n)^2 + C_2(2,n)^2 + \ldots\right]
\end{aligned}
$$

Ionization energies

- Provides distributions of $(w_1 E_1, w_1, w_2 E_2, w_2)$ in terms of weights

$$
\begin{aligned}
w_1(\mathbf{R}_n) &= \psi_1^2(\mathbf{R}_n)/P_n(\mathbf{R}_n) \\
w_2(\mathbf{R}_{n-1}) &= \psi_2^2(\mathbf{R}_{n-1})/P_{n-1}(\mathbf{R}_{n-1})
\end{aligned}
$$

- P_n is zero on coalescence planes only

- P_{n-1} is zero on coalescence planes only

- All sums are Normal and linearly correlated $\rightarrow$ Normal estimate

Ionization energies

- First row neutral atoms and ions

- Orbitals from same source as before

- Jastrow/Backflow and computational cost as before

- Optimise neutral atom and ion seperately, with energy minimisation

- Estimate energy difference/error as for excitation energies

Ionization energies
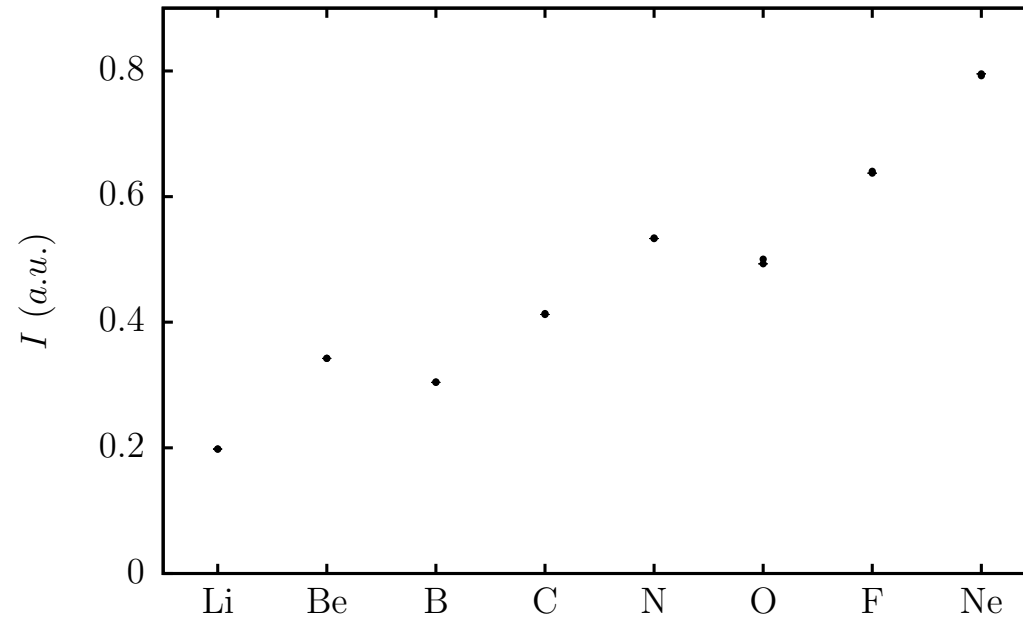
- Experimental ionization energies

- VMC estimated energy difference

Error in Ionization energies
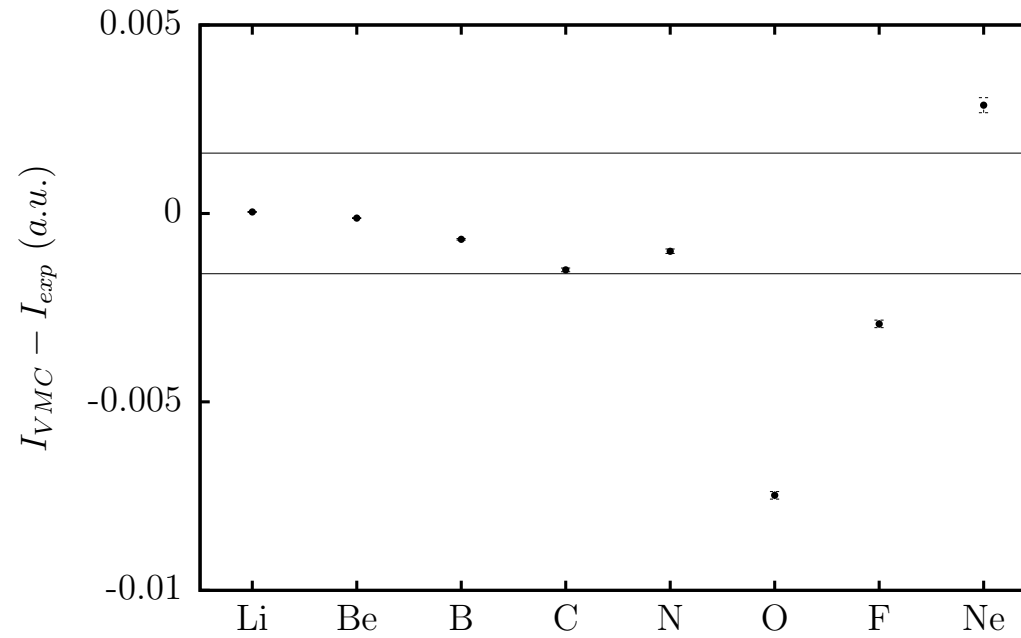
- Close but *not* chemical accuracy

- Maybe we need T/Q excitations ?

- Correlation reduces error by $0 - 40\%$

Transition moments

Spectroscopic line widths characterised by transition dipole moments

$$\Delta_\omega = \frac{3}{2}\Delta E \sum_m |\langle\psi_0|\mathbf{R}|\psi_m\rangle|^2$$

with sum over total angular momentum eigenstates.

Not done yet ... start with estimates for transition dipole moments:

$$t_{12} = |\langle\psi_1|\sum_i \mathbf{r}_i|\psi_2\rangle|^2$$

- Sample using $P = D_1(1)^2 + D_2(1)^2 + D_1(2)^2 + D_2(2)^2$

- Get an error estimate from random variable algebra

- Not zero variance...

Transition moments

Estimate:

$$t_{12} = |\langle\psi_1|\sum \mathbf{r}_i|\psi_2\rangle|^2$$

$$= \frac{\left[\int \psi_1\psi_2(x_1 + \ldots + x_n)d\mathbf{R}\right]^2 + \left[\int \psi_1\psi_2(y_1 + \ldots + y_n)d\mathbf{R}\right]^2 + \left[\int \psi_1\psi_2(z_1 + \ldots + z_n)d\mathbf{R}\right]^2}{\int \psi_1^2 d\mathbf{R} \int \psi_2^2 d\mathbf{R}}$$

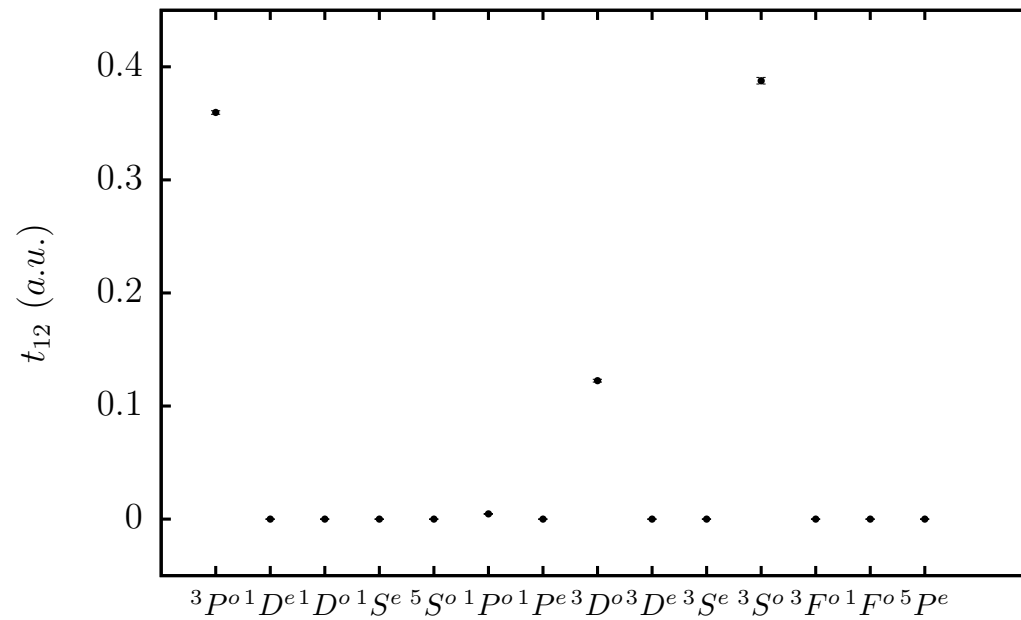- Define weights $w_1 = \psi_1^2/P$, $w_2 = \psi_2^2/P$

- Analyse as before ...

$$\text{Est}\,[t_{12}] = \frac{\mathsf{S}_1^2 + \mathsf{S}_2^2 + \mathsf{S}_3^2}{\mathsf{S}_4.\mathsf{S}_5}$$

- Normally distributed with estimateable mean and variance:

$$\bar{t}_{12} = \frac{\mu_1^2 + \mu_2^2 + \mu_3^2}{\mu_4\mu_5}$$

$$r\mu_4^2\mu_5^2 Var[\bar{t}_{12}] = \left[4\mu_1^2 C_{11} + 4\mu_2^2 C_{22} + 4\mu_3^2 C_{33} + 8\mu_1\mu_2 C_{12} + 8\mu_1\mu_3 C_{13} + 8\mu_2\mu_3 C_{23}\right]$$

$$-2\bar{t}_{12}\left[2\mu_1\mu_5 C_{14} + 2\mu_2\mu_5 C_{24} + 2\mu_3\mu_5 C_{34} + 2\mu_1\mu_4 C_{15} + 2\mu_2\mu_4 C_{25} + 2\mu_3\mu_4 C_{35}\right]$$

$$+\bar{t}_{12}^2\left[\mu_5^2 C_{44} + 2\mu_4\mu_5 C_{45} + \mu_4^2 C_{55}\right]$$
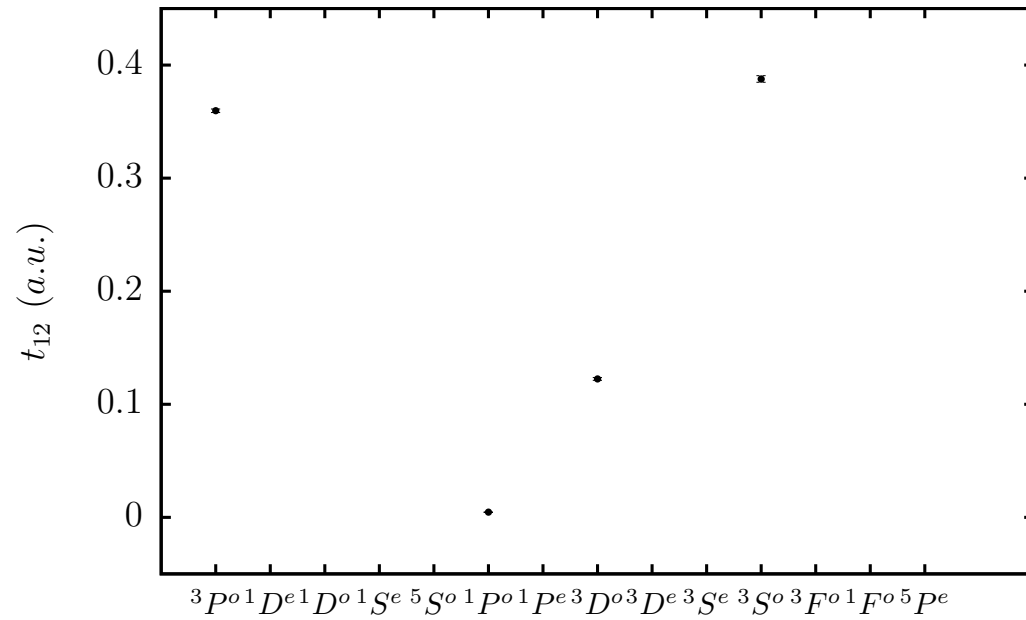
- Replace C's and μ's with unbiased estimates

Transition moments

- Many are non-normal and zero from symmetry considerations...

Transition moments

- ... so we drop them

- Random error $\sim 1\%$

Conclusions

- Normal errors can be reintroduced

- More computationally efficient than standard sampling

- Optimisation is on a Normal surface, unlike standard sampling

- Distribution of random error can be derived for general estimates and sampling

Normal and efficient estimates implemented for:

- Total energies

- Energy differences and Ionization energies

- Transition moments

Next?

- Spectroscopic line widths

- Electron affinities

- Optimise orthogonalised trial wavefunctions for more excited states

- Normal force estimates and geometery optimisation surfaces

- Generalised DMC