



*The Abdus Salam  
International Centre for Theoretical Physics*



**1929-14**

**Advanced School on Quantum Monte Carlo Methods in Physics and  
Chemistry**

*21 January - 1 February, 2008*

**Optimization of many-body wave functions**

C.J. Umrigar  
*Cornell University*

# Optimization of many-body wave functions

Cyrus J. Umrigar

with contributions by

Julien Toulouse,  
Claudia Filippi, Peter Nightingale, Sandro Sorella, Richard Hennig

QMC winter school, ICTP, Trieste, 21 Jan - 2 Feb 2008

# Optimization of many-body wave functions

1. Importance of optimizing  $\Psi$
2. Variance minimization vs. Variational Energy minimization
3. Energy optimization methods that work well:
  - 1 Newton method
  - 2 Linear method
  - 3 Perturbative method
4. Connection between methods, mixed minimization in the linear method
5. Stabilization of methods
6. Illustrative examples:
  - 1 How quickly do the various methods converge for various parameters?
  - 2 How large are variational and fixed-node errors
  - 3 How do  $E, \sigma, T_{\text{corr}}$  change going from variance minimization to energy minimization
7. Measures of goodness of wave functions (moments of  $\hat{H}$ ):
  - a) Energy, b) fluctuations of energy, c) overlap with exact  $\Psi$

# Accuracy of Diffusion Monte Carlo

## The problem:

Fixed-node error can be **LARGE** for these systems. e.g., the fixed-node error for  $C_2$  total energy is 1.3 eV and for well-depth is 0.7 eV.

## Possible solutions:

1. Solve the Fermion sign problem. **Hard!**
2. Develop:
  - 1 Better forms of trial wave functions.
    - 2.1.1 Casula and Sorella on geminals
    - 2.1.2 Schmidt, and, Bajdich and Mitas on Pfaffians
    - 2.1.3 Rios and Needs on backflow for nonfluid systems
  - 2 Powerful optimization methods to systematically improve the nodes of the trial wavefunctions.

## Almost all other errors reduced too

1. Reduce fixed-node error (nodes move during optimization). Fixed node errors can be **LARGE**.
2. Reduce other systematic errors in the energy – pseudopotential locality error, time-step error, population-control error.
3. Reduce systematic error of observables that do not commute with the Hamiltonian (mixed estimators,  $\langle \Psi_0 | \hat{A} | \Psi_T \rangle$  not exact even for nodeless  $\Psi$ ).
4. Reduce statistical error.

# Functional form of Trial Wave Function

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

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

$D^\uparrow$  and  $D^\downarrow$  are determinants of single-particle orbitals  $\phi$  for up ( $\uparrow$ ) and down ( $\downarrow$ ) spin electrons respectively.

The single-particle orbitals  $\phi$  are given by:

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

- **Jastrow:**  $\mathcal{J}(r_i, r_j, r_{ij}) = \prod_{\alpha i} \exp(A_{\alpha i}) \prod_{ij} \exp(B_{ij}) \prod_{\alpha ij} \exp(C_{\alpha ij})$

$A_{\alpha i} \Rightarrow$  electron-ion correlation

$B_{ij} \Rightarrow$  electron-electron correlation

$C_{\alpha ij} \Rightarrow$  electron-electron-ion correlation

$\sim N_{\text{atomtype}}$  of  $\mathcal{J}$  parms.

$\sim N_{\text{atomtype}}$  of  $\zeta_{k\alpha}$  parms.

$\sim N_{\text{atom}}^2$  of  $c_{k\alpha}$  parms.

$\sim e^{N_{\text{atom}}}$  of  $d_n$  parms.

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

**Power of QMC:**

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

# Progress in optimization of Many-Body Wavefunctions

Naive energy optim. → Variance optim. → Efficient energy optim.

- 1988 naive energy optimization, few ( $\sim 3$ ) parameters
- 1988 — 2001 variance optimization,  $\sim 100$  parameters
- 2001 — efficient energy optimization,  $> 1000$  parameters

# Optimization of Many-Body Wavefunctions

A major advantage of quantum Monte Carlo methods is that there is no restriction on the form of  $\Psi_T(\mathbf{R})$ . Hence any insight one may have, as regards the nature of the many-body correlations, can be built into  $\Psi_T(\mathbf{R})$  and tested. To exploit this freedom it is necessary to have a method for optimizing arbitrary wavefunctions.

**First thought:** Minimize the energy on MC sample.

$$\bar{E} = \sum_{i=1}^{N_{\text{conf}}} \frac{\mathcal{H}\Psi_T(\mathbf{R}_i; \{p\})}{\Psi_T(\mathbf{R}_i; \{p\})} w_i, \quad w_i = \left| \frac{\Psi_T(\mathbf{R}_i)}{\Psi_T^0(\mathbf{R}_i)} \right|^2 \bigg/ \sum_{i=1}^{N_{\text{conf}}} \left| \frac{\Psi_T(\mathbf{R}_i)}{\Psi_T^0(\mathbf{R}_i)} \right|^2$$

**Second thought:** Minimize the variance of the local energy.

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

**Third thought:** Minimize the energy using MC but not on MC sample.

**Fourth thought:** Minimize the linear combination with most of the weight on energy.



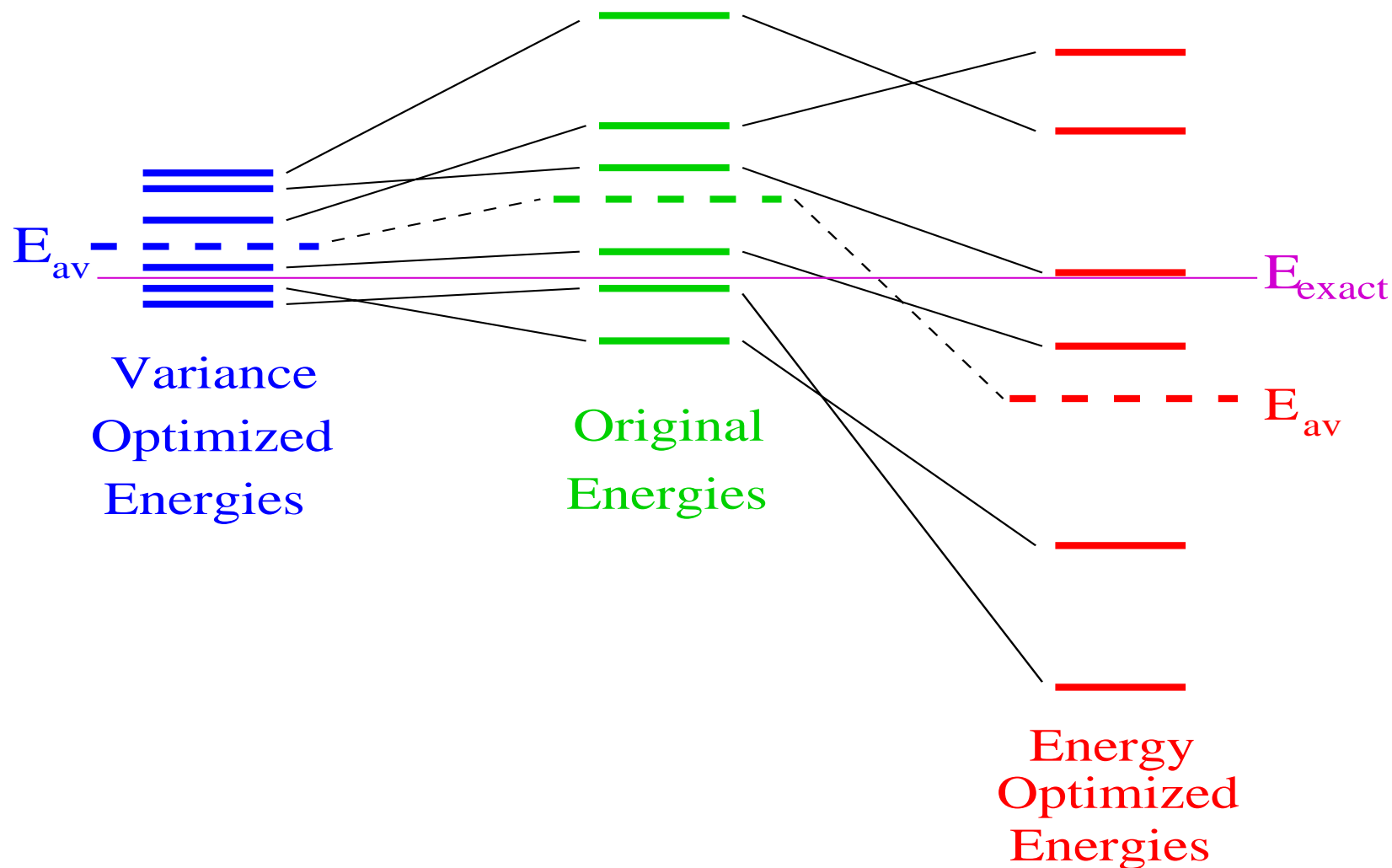
$$\sigma^2 = \sum_{i=1}^{N_{\text{conf}}} \left( \frac{\mathcal{H}\Psi_{\text{T}}(\mathbf{R}_i)}{\Psi_{\text{T}}(\mathbf{R}_i)} - \bar{E} \right)^2$$

**Variance**

vs.

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

**Energy**



## Advantages of Energy (or Mixed) Optim. vs. Variance Optim.

1. Want lowest energy; fluctuations are of secondary importance. Energy and variance do not always go hand-in-hand enough.
2. Some parameters couple more strongly to energy than variance.
3. Some parameters when variance optimized, make the system too extended
4. Energy-optimized  $\Psi$  can be better for other expectation values (Rothstein).
5. Hellman-Feynman theorem can be used for forces (when combined with variance reduction methods).

# Early energy optimization methods that worked well

## Linear method for linear parameters:

M. P. Nightingale and Melik-Alaverdian, Phys. Rev. Lett., **87**, 043401 (2001).

## Effective fluctuation potential method:

- ▶ S. Fahy, in *Quantum Monte Carlo Methods in Physics and Chemistry*, edited by M. P. Nightingale and C. J. Umrigar, [NATO ASI Ser. C. **525** 101, 1999];
- ▶ Filippi and S. Fahy, J. Chem. Phys., **112**, 3523 (2000);
- ▶ F. Schautz and S. Fahy, J. Chem. Phys., **116**, 3533 (2002);
- ▶ D. Prendergast, D. Bevan and S. Fahy, Phys. Rev. B, **66**, 155104 (2002);
- ▶ Friedemann Schautz and Claudia Filippi, J. Chem. Phys., **120**, 10931 (2004).

## Stochastic reconfiguration method:

- ▶ Sandro Sorella, Phys. Rev. B, **64**, 024512 (2001);
- ▶ Casula and Sorella, J. Chem. Phys., **119**, 6500 (2003).
- ▶ Sorella, PRB 71, 241103 (2005).

# Optimization methods to be discussed today

1. Newton method CJU, Filippi, PRL 94, 150201 (2005):  
Add terms to the Hessian that contribute nothing in the limit of an infinite MC sample, but cancel much of the fluctuations for a finite MC sample.  
Gain in efficiency: 3 orders of magnitude for NO<sub>2</sub>, more for C<sub>10</sub>H<sub>12</sub> compared to Newton of Lin-Zhang-Rappe.
2. Linear method (generalized eigenvalue problem):
  - 1 Linear parameters: Nightingale, et al., PRL, **87**, 043401 (2001)  
Use asymmetric  $H$  to have zero variance property in the limit that the basis functions span an invariant subspace.
  - 2 Nonlinear parameters: Toulouse, CJU, JCP (2007)  
CJU, Toulouse, Filippi, Sorella, Hennig, PRL **98**, 110201 (2007).  
Choose freedom of normalization  $\Psi(\mathbf{p}, \mathbf{R}) = N(\mathbf{p}) \Phi(\mathbf{p}, \mathbf{R})$  to make a near optimal change in the parameters.
3. Perturbation theory in an arbitrary nonorthog. basis:  
Toulouse, CJU, J. Chem. Phys., **126**, 084102 (2007).  
(Small modification of Scemama-Filippi (2006) perturbative EFP, modification of the Fahy-Filippi-Prendergast-Schautz EFP method.)

# Newton energy minimization method

Find parameter changes by solving linear equations:

$$\mathbf{h}\Delta\mathbf{p} = -\mathbf{g},$$

$\mathbf{h}$  is the Hessian and  $\mathbf{g}$  is the gradient of the energy.

Two modifications of the straightforward Newton method:

1. Add terms to the Hessian that have zero expectation value on an infinite MC sample but that cancel much of the fluctuations on a finite sample.
2. Replace certain terms in the Hessian by other terms that are approximately proportional to them but that fluctuate less.

These 2 changes improve the efficiency of the simple Newton method of Lin, Zhang and Rappe by about 3 orders of magnitude.

## Newton energy minimization method

$$\bar{E} = \frac{\int d^3\mathbf{R} \psi H \psi}{\int d^3\mathbf{R} |\psi|^2} = \langle E_L \rangle; \quad E_L(\mathbf{R}) = \frac{H\psi(\mathbf{R})}{\psi(\mathbf{R})}$$

$$\bar{E}_i = \left\langle \frac{\psi_i}{\psi} E_L + \frac{H\psi_i}{\psi} - 2\bar{E} \frac{\psi_i}{\psi} \right\rangle = 2 \left\langle \frac{\psi_i}{\psi} (E_L - \bar{E}) \right\rangle \quad (\text{by Hermiticity}).$$

$$\bar{E}_{ij} = 2 \left[ \left\langle \left( \frac{\psi_{ij}}{\psi} + \frac{\psi_i \psi_j}{|\psi|^2} \right) (E_L - \bar{E}) \right\rangle - \left\langle \frac{\psi_i}{\psi} \right\rangle \bar{E}_j - \left\langle \frac{\psi_j}{\psi} \right\rangle \bar{E}_i + \left\langle \frac{\psi_i}{\psi} E_{L,j} \right\rangle \right].$$

Identical to the Hessian in Lin, Zhang, Rappe.

Two changes:

- 1) Symmetrize – eigensystem is real
- 2) Noting that  $\langle E_{L,j} \rangle = 0$ , rewrite in the form of a covariance

$$(\langle ab \rangle - \langle a \rangle \langle b \rangle)$$

The fluctuations of  $\langle ab \rangle - \langle a \rangle \langle b \rangle$  are in most cases smaller than those of  $\langle ab \rangle$ , (e.g. if  $a$  and  $b$  are weakly correlated), and, they are much smaller if  $\sqrt{\langle a^2 \rangle - \langle a \rangle^2} \ll |\langle a \rangle|$  and  $a$  is not strongly correlated with  $1/b$ .

## Newton energy minimization method (cont.)

$$\begin{aligned}
 \bar{E}_{ij} &= 2 \left[ \left\langle \left( \frac{\psi_{ij}}{\psi} + \frac{\psi_i \psi_j}{|\psi|^2} \right) (E_L - \bar{E}) \right\rangle - \left\langle \frac{\psi_i}{\psi} \right\rangle \bar{E}_j - \left\langle \frac{\psi_j}{\psi} \right\rangle \bar{E}_i \right] \\
 &\quad + \left\langle \frac{\psi_i}{\psi} E_{L,j} \right\rangle - \left\langle \frac{\psi_i}{\psi} \right\rangle \langle E_{L,j} \rangle + \left\langle \frac{\psi_j}{\psi} E_{L,i} \right\rangle - \left\langle \frac{\psi_j}{\psi} \right\rangle \langle E_{L,i} \rangle \\
 &= 2 \left[ \left\langle \left( \frac{\psi_{ij}}{\psi} - \frac{\psi_i \psi_j}{|\psi|^2} \right) (E_L - \bar{E}) \right\rangle \right. \quad (0 \text{ for lin. } p_i \text{ in expon}) \\
 &\quad \left. + 2 \left\langle \left( \frac{\psi_i}{\psi} - \left\langle \frac{\psi_i}{\psi} \right\rangle \right) \left( \frac{\psi_j}{\psi} - \left\langle \frac{\psi_j}{\psi} \right\rangle \right) (E_L - \bar{E}) \right\rangle \right] \\
 &\quad + \left\langle \frac{\psi_i}{\psi} E_{L,j} \right\rangle - \left\langle \frac{\psi_i}{\psi} \right\rangle \langle E_{L,j} \rangle + \left\langle \frac{\psi_j}{\psi} E_{L,i} \right\rangle - \left\langle \frac{\psi_j}{\psi} \right\rangle \langle E_{L,i} \rangle .
 \end{aligned}$$

1) Additional terms =0 for infinite sample but cancel most of the fluctuations for a finite sample.

2) **Red** terms fluctuate less than **blue** terms away from minimum. Ratio of **red** terms to **blue** terms depends on  $\psi$  but is roughly independent of  $i$  and  $j$ . Why?? Exploit that!

## Optimization of determinantal parameters

Different issues arise in optimizing Jastrow parameters and determinantal parameters:

Jastrow: eigenvalues of Hessian have a range of 11 orders of magnitude!

Determinantal parameters: divergences in elements of Hessian and Hamiltonian matrices.

$$\bar{E}_i = 2 \left\langle \frac{\psi_i}{\psi} (E_L - \bar{E}) \right\rangle \quad (\text{by Hermiticity}).$$

$$\begin{aligned} \bar{E}_{ij} = 2 \left[ \left\langle \left( \frac{\psi_{ij}}{\psi} + \frac{\psi_i \psi_j}{|\psi|^2} \right) (E_L - \bar{E}) \right\rangle - \left\langle \frac{\psi_i}{\psi} \right\rangle \bar{E}_j - \left\langle \frac{\psi_j}{\psi} \right\rangle \bar{E}_i \right] \\ + \left\langle \frac{\psi_i}{\psi} E_{L,j} \right\rangle - \left\langle \frac{\psi_i}{\psi} \right\rangle \langle E_{L,j} \rangle + \left\langle \frac{\psi_j}{\psi} E_{L,i} \right\rangle - \left\langle \frac{\psi_j}{\psi} \right\rangle \langle E_{L,i} \rangle \end{aligned}$$

Leading divergences cancel! They cancel in the linear method too.

Remaining divergences do not seem to be problematic. Do 2<sup>nd</sup> order ones cancel?



# Linear method for linear parameters

M. P. Nightingale and Melik-Alaverdian, PRL, **87**, 043401 (2001).

Given  $N = N_{\text{parm}} + 1$  basis functions,  $\Psi_i(\mathbf{R}_\sigma)$ , that are complete, or, less stringently that form an invariant subspace, one can find a linear combination that satisfies the Schrödinger equation exactly for any  $\mathbf{R}$ :

$$\hat{H}(\mathbf{R}) \sum_{j=0}^{N_{\text{parm}}} p_j \Psi_j(\mathbf{R}) = E \sum_{j=0}^{N_{\text{parm}}} p_j \Psi_j(\mathbf{R})$$

Multiplying by  $\Psi_i(\mathbf{R})$  and integ. over  $\mathbf{R}$  gives the gen. eigenvalue eq.,

$$\mathbf{H}\mathbf{p} = E\mathbf{S}\mathbf{p}, \quad (p_0 = 1)$$

where  $H_{ij} = \langle \Psi_i | \hat{H} | \Psi_j \rangle$  and  $S_{ij} = \langle \Psi_i | \Psi_j \rangle$ .

Instead, multiplying by  $\Psi_i(\mathbf{R})/\Psi_0^2(\mathbf{R})$  and averaging over MC points sampled from  $\Psi_0^2$  again gives Eq. 1 but with  $H_{ij} = \left\langle \frac{\Psi_i}{\Psi_0^2} \hat{H} \Psi_j \right\rangle$  and

$S_{ij} = \left\langle \frac{\Psi_i \Psi_j}{\Psi_0^2} \right\rangle$ . It is this **nonsymmetric  $\mathbf{H}$**  that has the **strong zero-variance property** that the parameters are independent of the MC sample.

## Linear method for linear parameters (cont)

Note that minimizing the energy evaluated on a finite sample

$$\begin{aligned} E &= \min_{\mathbf{p}} \frac{\mathbf{p}^T \mathbf{H} \mathbf{p}}{\mathbf{p}^T \mathbf{S} \mathbf{p}} \\ &= \min_{\mathbf{p}} \frac{\sum_{ij} p_i H_{ij} p_j}{\sum_{ij} p_i S_{ij} p_j} \end{aligned}$$

gives a generalized eigenvalue equation with a **symmetric  $\mathbf{H}$**  that lacks the **strong zero-variance property**.

$$\left( \sum_{ij} p_i S_{ij} p_j \right) \left( \sum_j H_{kj} p_j + \sum_i p_i H_{ik} \right) - \left( \sum_{ij} p_i H_{ij} p_j \right) \left( 2 \sum_j S_{kj} p_j \right) = 0$$

$$\frac{(\mathbf{H}^T + \mathbf{H})}{2} \mathbf{p} = E \mathbf{S} \mathbf{p}$$

# Linear method for nonlinear parameters

CJU, Toulouse, Filippi, Sorella, Hennig PRL (2007);  
Toulouse, CJU, JCP (2007).

Make linear order Taylor expansion of wavefunction:

$$\Psi_{\text{lin}} = \Psi_0 + \sum_{i=1}^{N_{\text{parm}}} \Delta p_i \Psi_i$$

$\Psi_0 \equiv \Psi(\mathbf{p}_0, \mathbf{R})$  = current wave function

$\Psi_{\text{lin}}$  = next iteration linearized wave function

$\Psi_i$  = derivative of  $\Psi$  at  $\mathbf{p}_0$ , wrt  $i^{\text{th}}$  parameter.

No unique way to obtain new parameters.

The simplest procedure: is  $p_i^{\text{new}} = p_i + \Delta p_i$ .

More complicated procedure: fit wave function form to the optimal linear combination.

Simpler, yet efficient approach, freedom of norm to make linear approximation better

$$\begin{aligned}\bar{\Psi}(\mathbf{p}, \mathbf{R}) &= N(\mathbf{p}) \Psi(\mathbf{p}, \mathbf{R}), \quad N(\mathbf{p}_0) = 1 \\ \bar{\Psi}_i &= \Psi_i + N_i \Psi_0\end{aligned}$$

## General semiorthogonalization

$$\begin{aligned}\bar{\Psi}(\mathbf{p}, \mathbf{R}) &= N(\mathbf{p}) \Psi(\mathbf{p}, \mathbf{R}), \quad N(\mathbf{p}_0) = 1 \\ \bar{\Psi}_i &= \Psi_i + N_i \Psi_0\end{aligned}$$

Choose  $N_i$  such that the derivatives are orthogonal to a linear combination of  $\Psi_0$  and  $\Psi_{\text{lin}}$ .

$$\left\langle \xi \frac{\Psi_0}{|\Psi_0|} + (1 - \xi) \frac{\Psi_{\text{lin}}}{|\Psi_{\text{lin}}|} \middle| \Psi_i + N_i \Psi_0 \right\rangle = 0$$

Solving for  $N_i$  we get  $[s = 1(-1)$  if  $\langle \Psi_0 | \Psi_{\text{lin}} \rangle = 1 + \sum_j S_{0j} \Delta p_j > 0 (< 0)]$ ,

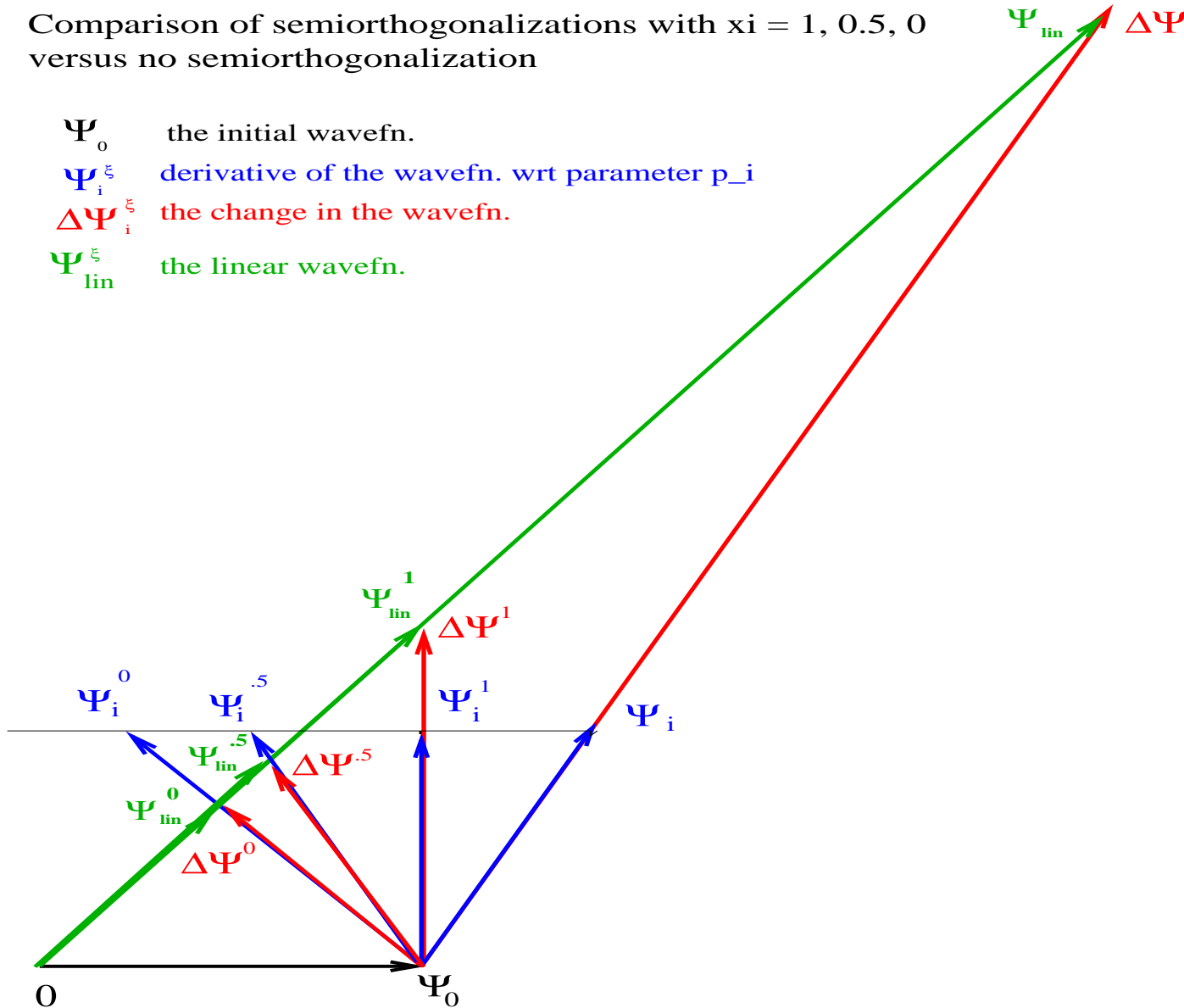
$$N_i = - \frac{\xi D S_{0i} + s(1 - \xi)(S_{0i} + \sum_j S_{ij} \Delta p_j)}{\xi D + s(1 - \xi)(1 + \sum_j S_{0j} \Delta p_j)}$$

$$D = \frac{|\Psi_{\text{lin}}|}{|\Psi_0|} = \left( 1 + 2 \sum_j S_{0j} \Delta p_j + \sum_{i,j} S_{ij} \Delta p_i \Delta p_j \right)^{1/2}$$

# Semiorthogonalization in the linear method

Comparison of semiorthogonalizations with  $\xi = 1, 0.5, 0$  versus no semiorthogonalization

- $\Psi_0$  the initial wavefn.
- $\Psi_i^\xi$  derivative of the wavefn. wrt parameter  $p_i$
- $\Delta\Psi_i^\xi$  the change in the wavefn.
- $\Psi_{lin}^\xi$  the linear wavefn.



## Semiorthogonalization in the linear method

$\Psi_0$  is the initial wave function,  $\Psi_i^\zeta$  is the derivative of the wave function wrt parameter  $p_i$  for  $\zeta$ . If superscript  $\zeta$  is omitted that denotes that no semiorthogonalization is done. Then

$$\Psi_{\text{lin}} = \Psi_0 + \sum_{i=1}^{N_{\text{parm}}} \Delta \Psi_i^\zeta = \Psi_0 + \sum_{i=1}^{N_{\text{parm}}} \Delta p_i^\zeta \Psi_i^\zeta, \quad \Delta p_i^\zeta = \frac{\Delta \Psi_i^\zeta}{\Psi_i^\zeta}$$

Note that  $||\Delta \Psi^\zeta||$  is smallest for  $\zeta = 1$  and that  $||\Psi_{\text{lin}}^{0.5}|| = ||\Psi_0||$ .

Also note that when there is just one parameter (can be generalized to more than one):

1. In the limit that  $\Psi_{\text{lin}} \parallel \Psi_i$ ,  $\Delta p_i = \pm\infty$
2. In the limit that  $\Psi_{\text{lin}} \perp \Psi_0$ ,  $\Delta p_i^1 = \pm\infty$  because  $\Delta \Psi^1 = \infty$ , and,  $\Delta p_i^0 = 0$  because  $\Psi_i^0 = \infty$
3.  $\Delta p_i^{0.5}$  is always finite

Note that  $\Delta p_i^\zeta$  decreases as  $\zeta$  decreases from 1 to 0. In Fig. 1,  $\Delta p_i$  is  $> 1$  for  $\zeta = 1$ , and,  $< 1$  for  $\zeta = 0.5, 0$ .

Also note that in Fig. 1 if we rotate  $\Psi_{\text{lin}}$  such that  $\frac{\nabla \Psi \cdot \Psi_0}{||\nabla \Psi|| ||\Psi_0||} > \frac{\Psi_{\text{lin}} \cdot \Psi_0}{||\Psi_{\text{lin}}|| ||\Psi_0||}$  then  $\Delta p_i$  has the opposite sign as  $\Delta p_i^\zeta$ !

# Perturbation Theory in a nonorthogonal basis

Toulouse, CJU, JCP (2007)

Given a Hamiltonian  $\hat{H}$  and an arbitrary nonorthogonal basis,  $\{|\Psi_i\rangle\}$ , use perturbation theory to get approximate eigenstates of  $\hat{H}$ .

Define dual basis:  $\langle\bar{\Psi}_i|\Psi_j\rangle = \delta_{ij}$  and zeroth order Hamiltonian,  $\hat{H}^{(0)}$ :

$$\langle\bar{\Psi}_i| = \sum_{j=0}^{N_{\text{opt}}} (\mathbf{S}^{-1})_{ij} \langle\Psi_j|, \quad \hat{H}^{(0)} = \sum_{i=0}^{N_{\text{opt}}} E_i |\Psi_i\rangle \langle\bar{\Psi}_i|$$

First order perturbation correction is

$$|\Psi^{(1)}\rangle = - \sum_{i=1}^{N_{\text{opt}}} |\Psi_i\rangle \sum_{j=1}^{N_{\text{opt}}} (\mathbf{S}^{-1})_{ij} \frac{\langle\Psi_j|\hat{H}|\Psi_0\rangle}{E_i - E_0}$$

Want  $\hat{H}$  and  $\hat{H}^{(0)}$  close – choose  $E_i$  so  $\hat{H}$  and  $\hat{H}^{(0)}$  have same diagonals

$$E_i = \frac{\langle\Psi_i|\hat{H}|\Psi_i\rangle}{\langle\Psi_i|\Psi_i\rangle}$$

If the  $E_i$  are evaluated without the Jastrow factor then this is the same as the perturbative eff. fluct. pot. (EFP) method of [Scemama and Filippi 06](#).

# Stabilization

If far from the minimum, or,  $N_{\text{MC}}$ , is small, then the Hessian,  $\bar{E}_{ij}$ , need not be positive definite (whereas variance-minimization Levenberg-Marquardt  $\bar{E}_{ij}$  is positive definite).

Even for positive definite  $\bar{E}_{ij}$ , the new parameter values may make the wave function worse if quadratic approximation is not good.

Determine eigenvalues,  $\epsilon$ , of  $\bar{E}_{ij}$  and add to it  $(\max(0, -\epsilon_{\min}) + a_{\text{diag}})\mathcal{I}$ . This shifts the eigenvalues by the added constant. As  $a_{\text{diag}}$  is increased, the proposed parameter changes become smaller and rotate from the Newtonian direction to the steepest descent direction, but in practice  $a_{\text{diag}}$  is tiny.

The linear method and the perturbative method can be approximately recast into the Newton method. Consequently we can use the same idea for the linear and perturbative methods too.



# Stabilization with Correlated Sampling

Each method has a parameter  $a_{\text{diag}}$  that automatically adjusts to make the method totally stable:

1. Do a MC run to compute the gradient and the Hessian (or overlap and Hamiltonian).
2. Using the above gradient and Hessian (or overlap and Hamiltonian), use 3 different values of  $a_{\text{diag}}$  to predict 3 different sets of updated parameters.
3. Do a short correlated sampling run for the 3 different wave functions to compute the energy differences for the 3 wave functions more accurately than the energies themselves.
4. Fit a parabola through the 3 energies to find the optimal  $a_{\text{diag}}$ .
5. Use this optimal  $a_{\text{diag}}$  to predict a new wave function, using the gradient and Hessian computed in step 1.
6. Loop back

# Comparison of Newton, linear and perturbative methods

Programming effort and cost per iteration:

1. Newton method requires  $\psi$ ,  $\psi_i$ ,  $\psi_{ij}$ ,  $\hat{H}\psi$ ,  $\hat{H}\psi_i$ .
2. Linear method requires  $\psi$ ,  $\psi_i$ ,  $\hat{H}\psi$ ,  $\hat{H}\psi_i$ .
3. Perturbative method requires  $\psi$ ,  $\psi_i$ ,  $\hat{H}\psi$ ,  $\hat{H}\psi_i$ .  
Perturbative method with approximate denominators requires  $\psi$ ,  $\psi_i$ .

Correct to which order:

1. Newton has correct 2nd order terms in  $E$ .
2. Linear does not have correct 2nd order terms in  $E$  but has all order terms coming from linear terms in  $\Psi$ .

Convergence with number of iterations:

1. Newton and linear methods converge in 1-5 iterations for **all** parameters (CSF, orbital and Jastrow).
2. Perturbative method converges in 1-5 iterations for CSF and orbital parameters but is very slow for Jastrow because eigenvalues of Hessian for Jastrow span 9-12 orders of magnitude.

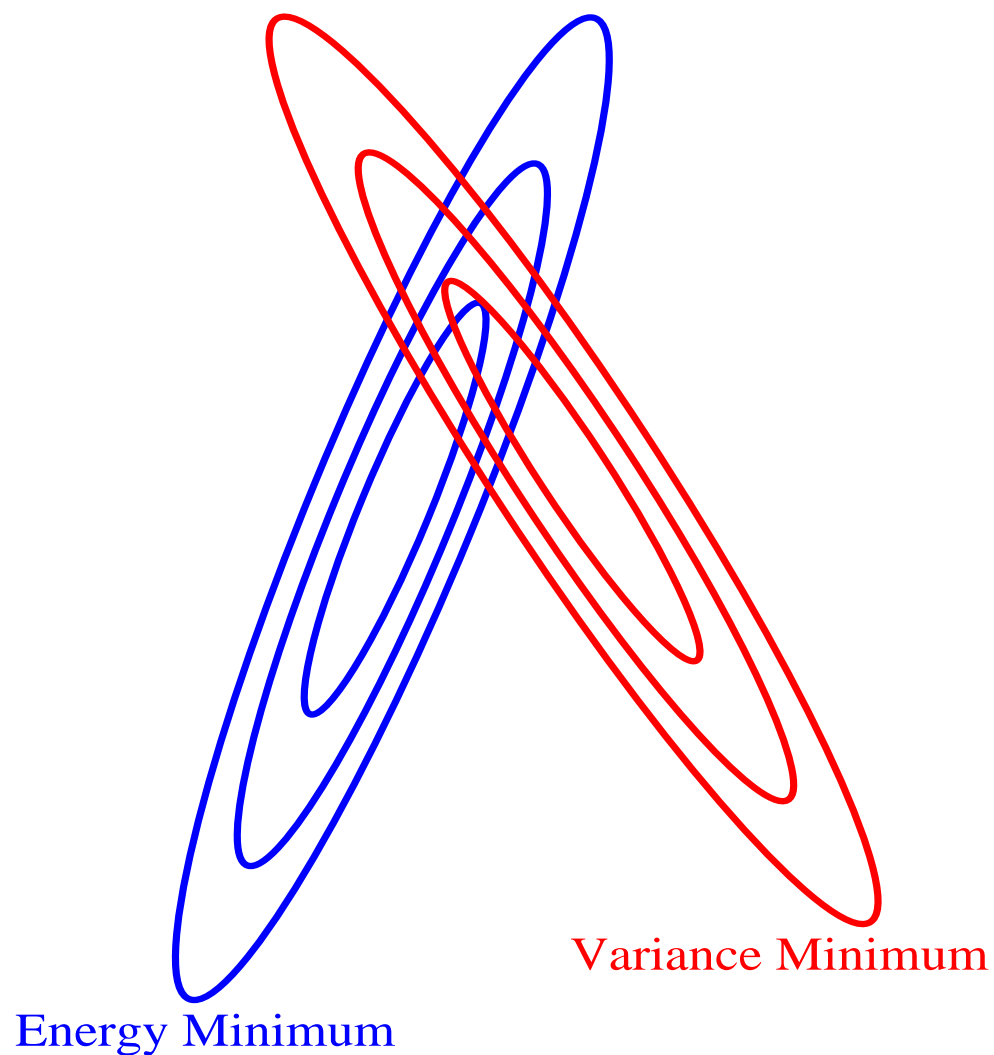
## Things to note

Eigenvalues of  $\bar{E}_{ij}$  for Jastrow parameters span 11 orders of magnitude. So steepest descent would be horribly slow to converge!

Linear and Newton methods can be used for all parameters, even basis-set exponents.

**Take Home Message:** Any method that attempts to minimize the energy, by minimizing the energy evaluated on a set of MC points, will require a very large sample and be highly inefficient.  
Each of the 3 methods presented above avoids doing this.

# Optimization of linear combination of energy and variance



1. Can reduce the variance, without sacrificing appreciably the energy, by minimizing a linear combination, particularly since the ratio of hard to soft directions is 11 orders of magnitude.
2. Easy to do – gradient and Hessian of linear combination are linear combinations of the gradient and Hessian.
3. Measure of efficiency of the wave function is  $\sigma^2 T_{\text{corr}}$ .

## Optim. of linear combin. of $E$ and $\sigma^2$ (Toulouse, CJU)

$$\begin{aligned}
 E &= \min_{\Delta \mathbf{p}} \frac{\begin{pmatrix} 1 & \Delta \mathbf{p}^T \end{pmatrix} \begin{pmatrix} E_0 & H_{0j} \\ H_{i0} & H_{ij} \end{pmatrix} \begin{pmatrix} 1 \\ \Delta \mathbf{p} \end{pmatrix}}{\begin{pmatrix} 1 & \Delta \mathbf{p}^T \end{pmatrix} \begin{pmatrix} 1 & S_{0j} \\ S_{i0} & S_{ij} \end{pmatrix} \begin{pmatrix} 1 \\ \Delta \mathbf{p} \end{pmatrix}}, \\
 &= E_0 + \sum_j (H_{0j} + H_{j0} - 2E_0 S_{0j}) \Delta p_j \\
 &\quad + \sum_{ij} \{H_{ij} - E_0 S_{ij} - 2(H_{0i} + H_{i0}) S_{0j}\} \Delta p_i \Delta p_j
 \end{aligned}$$

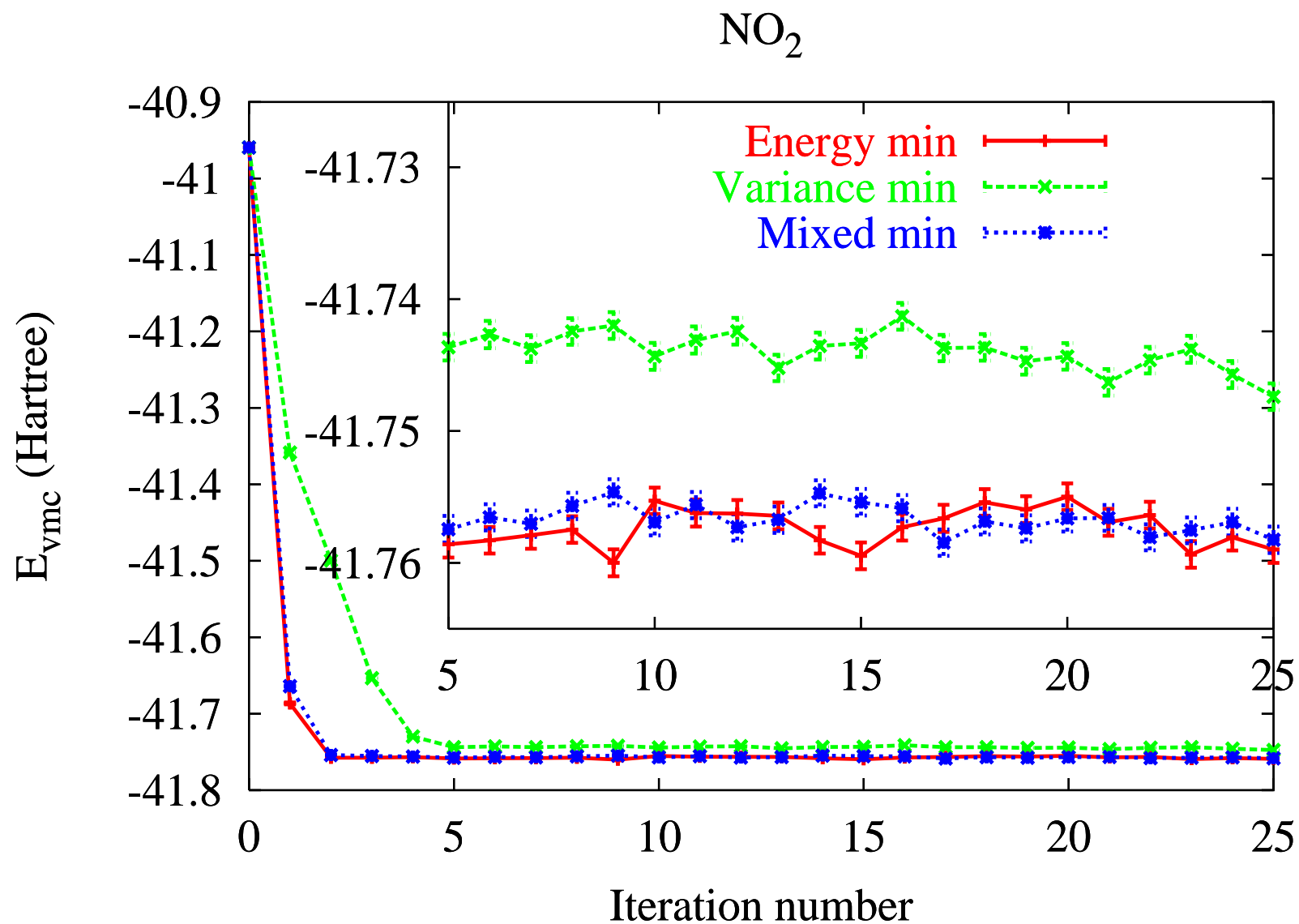
If this is to equal the 2<sup>nd</sup> order expansion in the Newton method then

$$\begin{aligned}
 H_{0j} + H_{j0} - 2E_0 S_{0j} &= -g_j \\
 H_{ij} - E_0 S_{ij} - 2(H_{0i} + H_{i0}) S_{0j} &= \frac{h_{ij}}{2}
 \end{aligned}$$

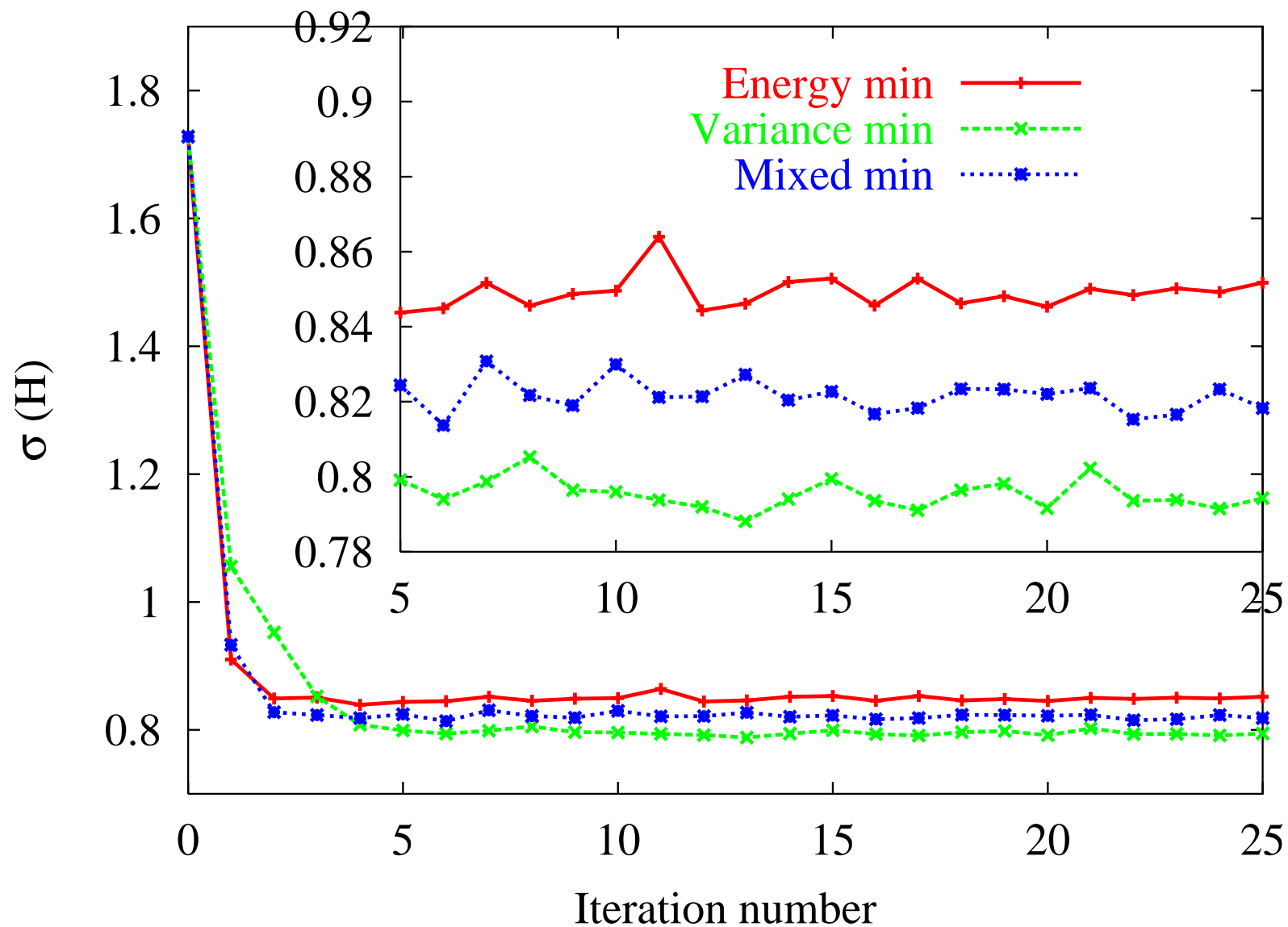
So choose,

$$\begin{aligned}
 H_{0j} + H_{j0} &= 2E_0 S_{0j} - g_j \\
 H_{ij} &= E_0 S_{ij} + 2(H_{0i} + H_{i0}) S_{0j} + \frac{h_{ij}}{2}
 \end{aligned}$$

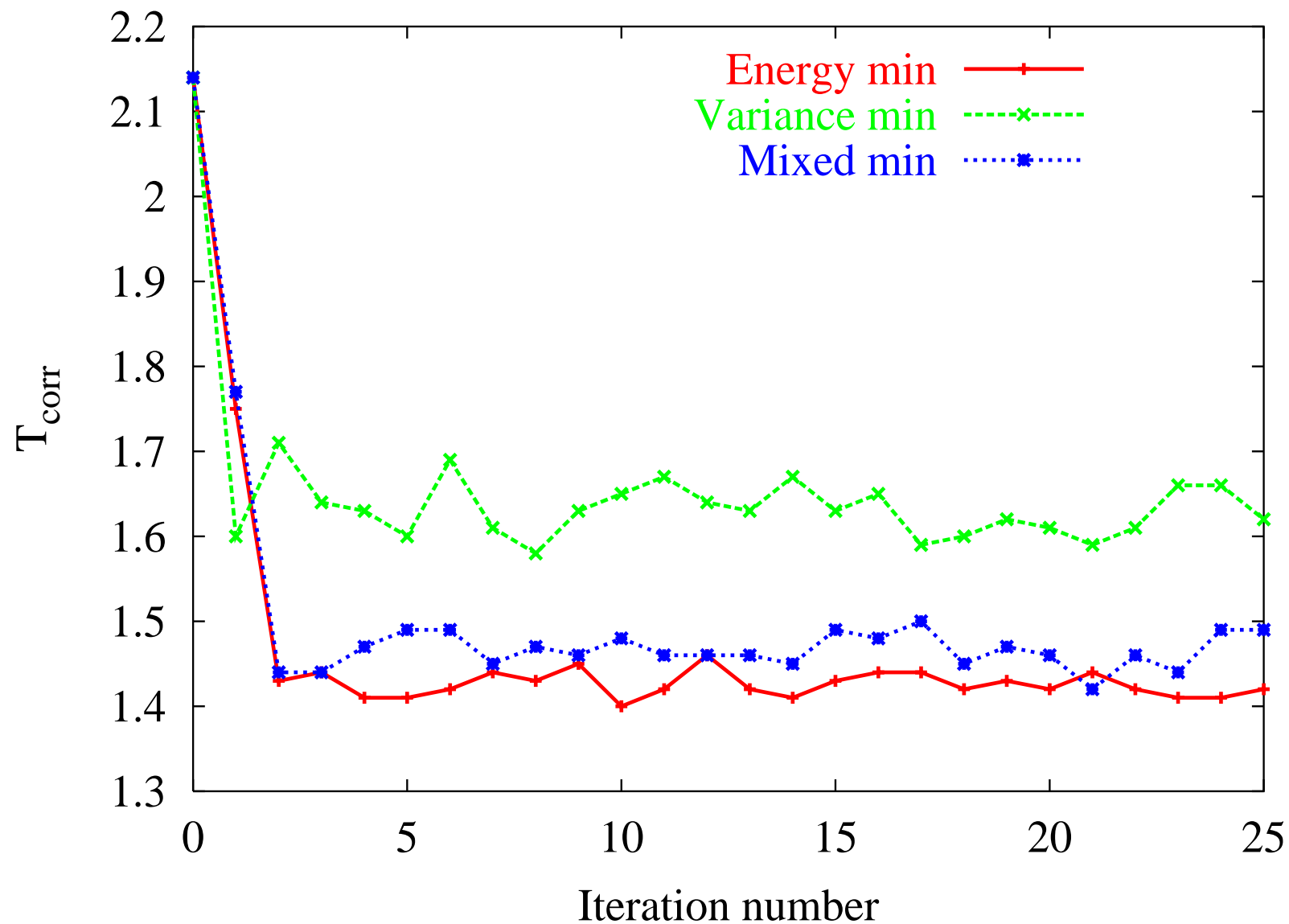
# Convergence of energy of NO<sub>2</sub>



## Convergence of energy fluctuations, $\sigma$ , of NO<sub>2</sub>

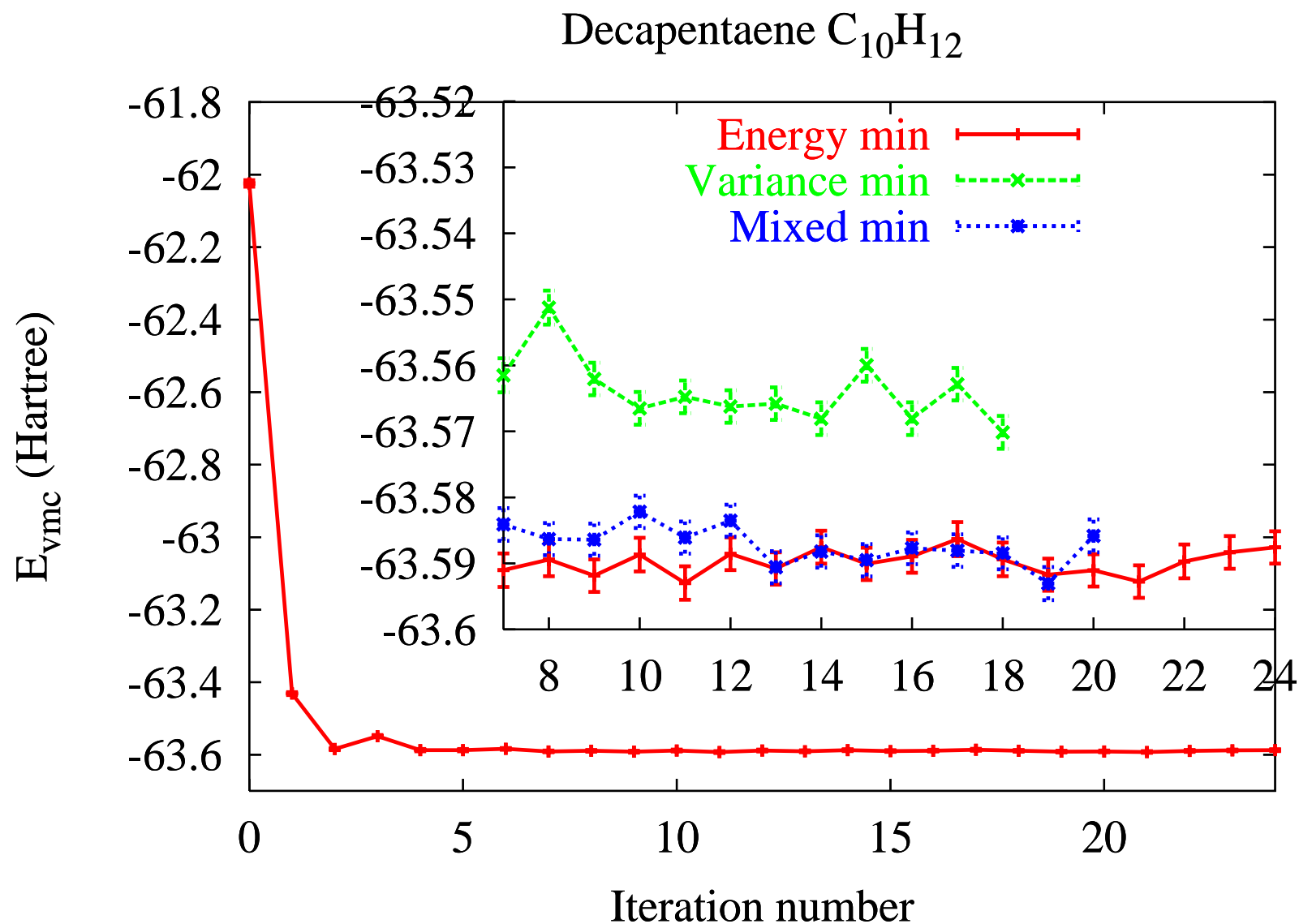


## Convergence of autocorrelation time, $T_{\text{corr}}$ , of $\text{NO}_2$

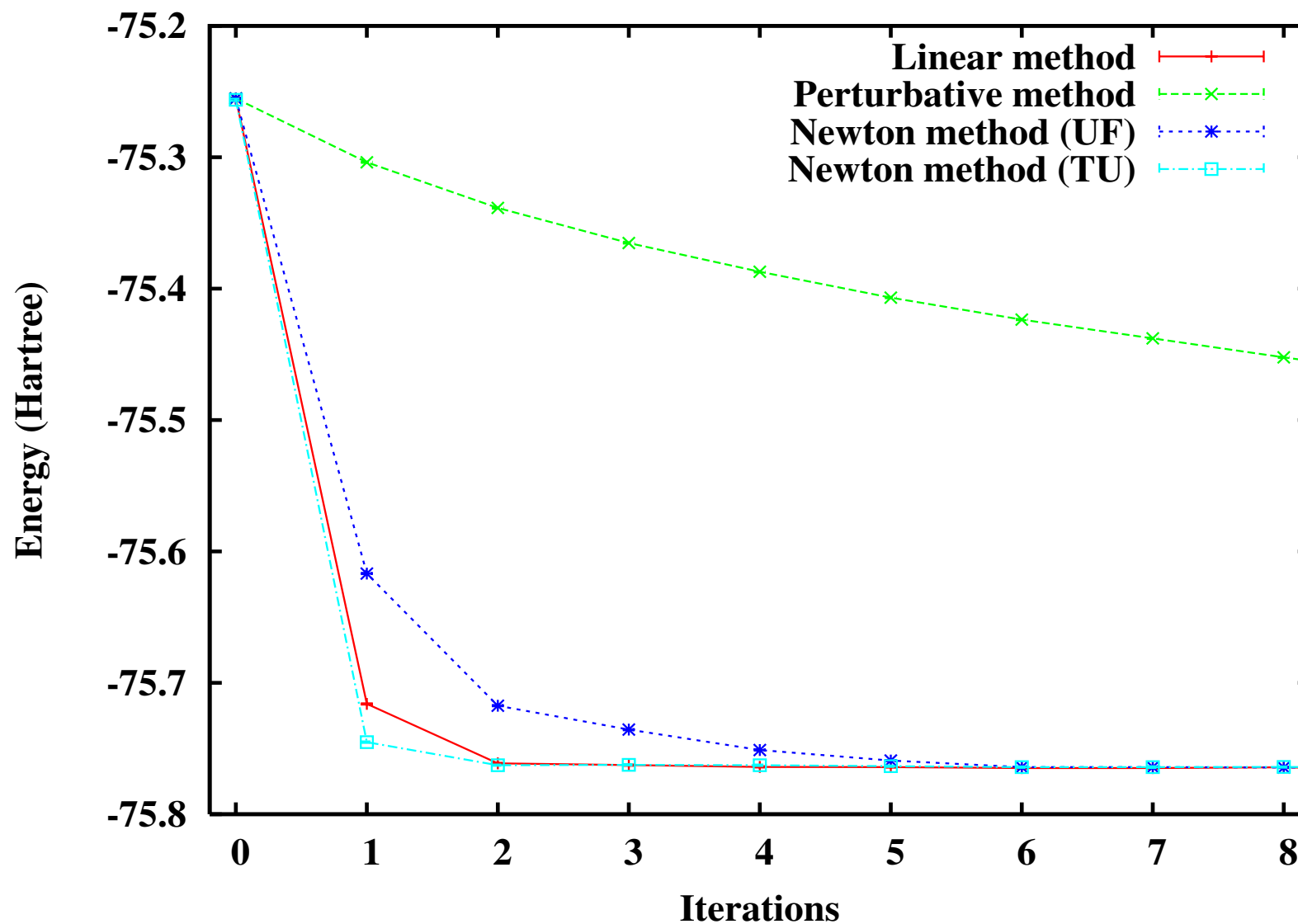




# Convergence of energy of decapentaene $C_{10}H_{12}$

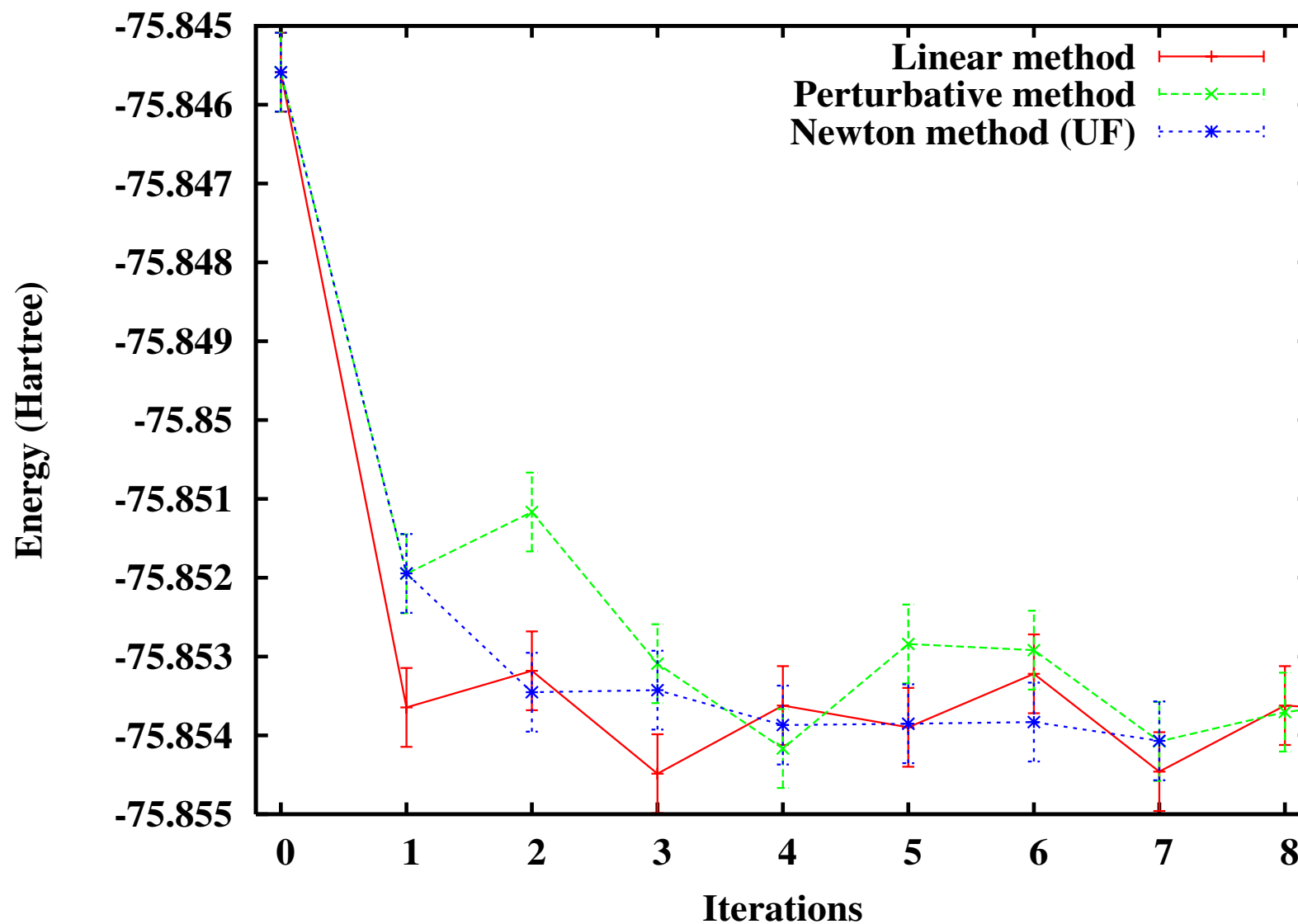


## Optimization of Jastrow of all-electron $C_2$



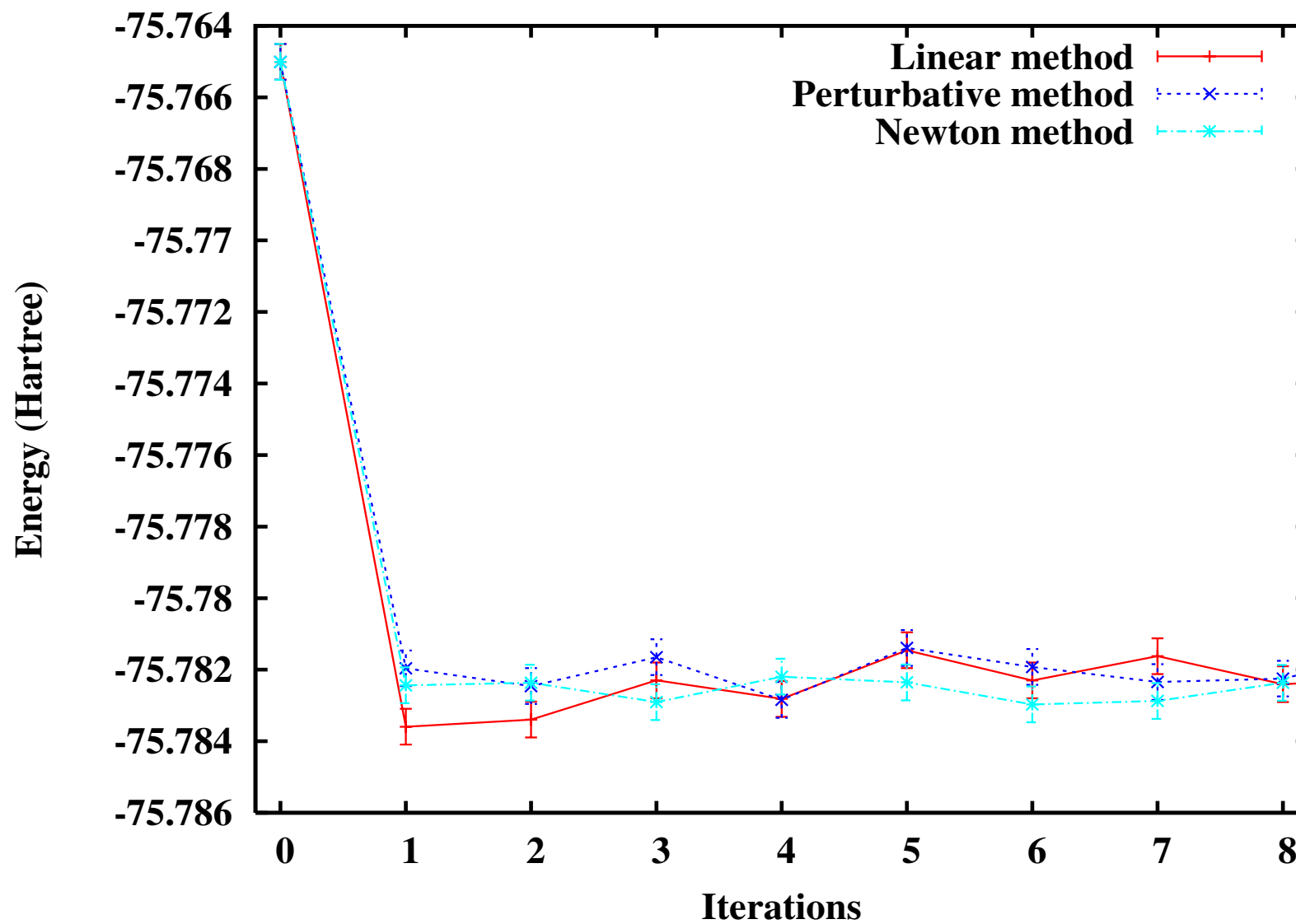
Linear and Newton methods converge in 2-6 iterations, perturbative is slow!

## Optimization of CSF coefs. of all-electron $C_2$



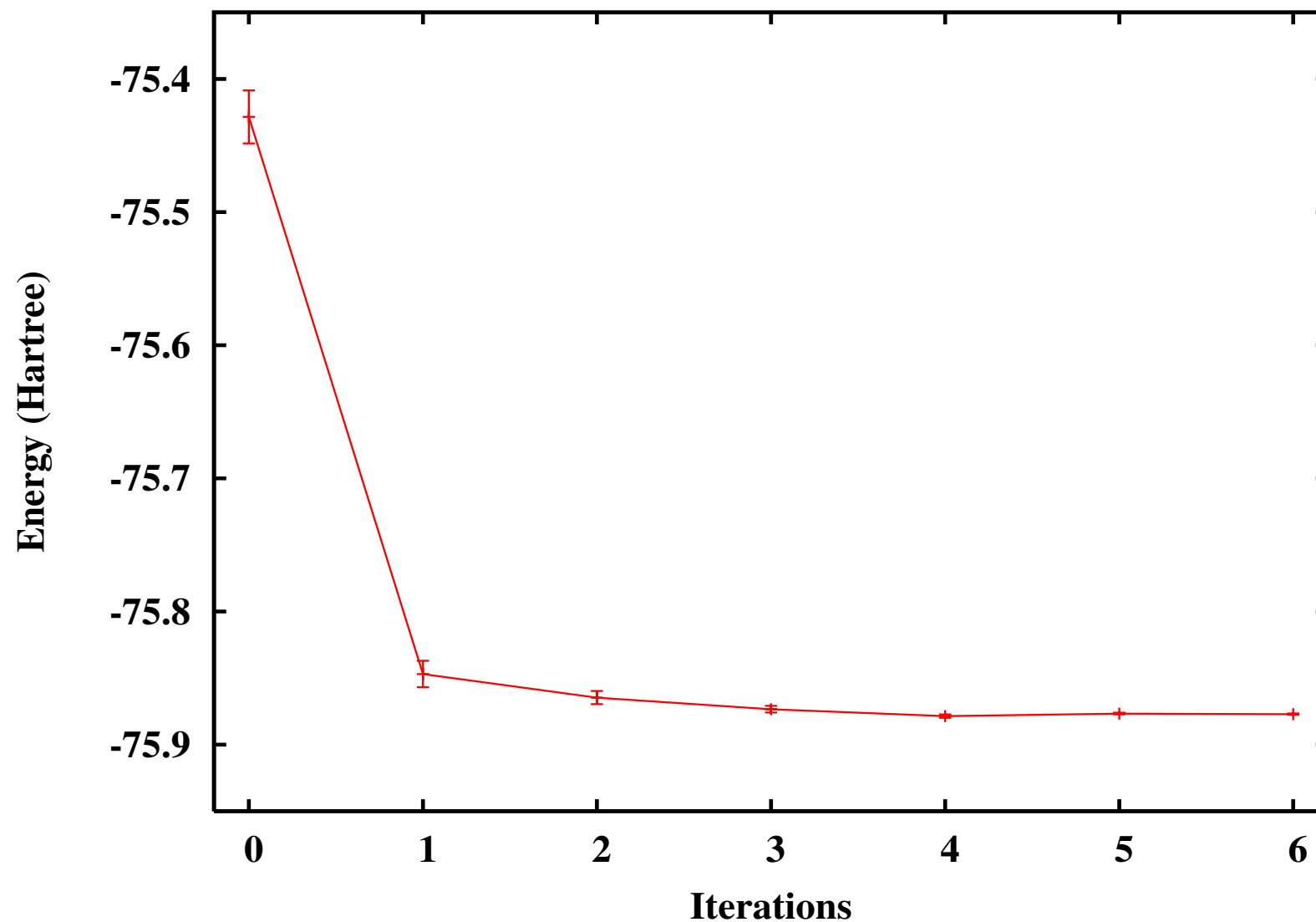
All 3 methods converge in 1-3 iteration!

## Optimization of orbitals of all-electron C<sub>2</sub>

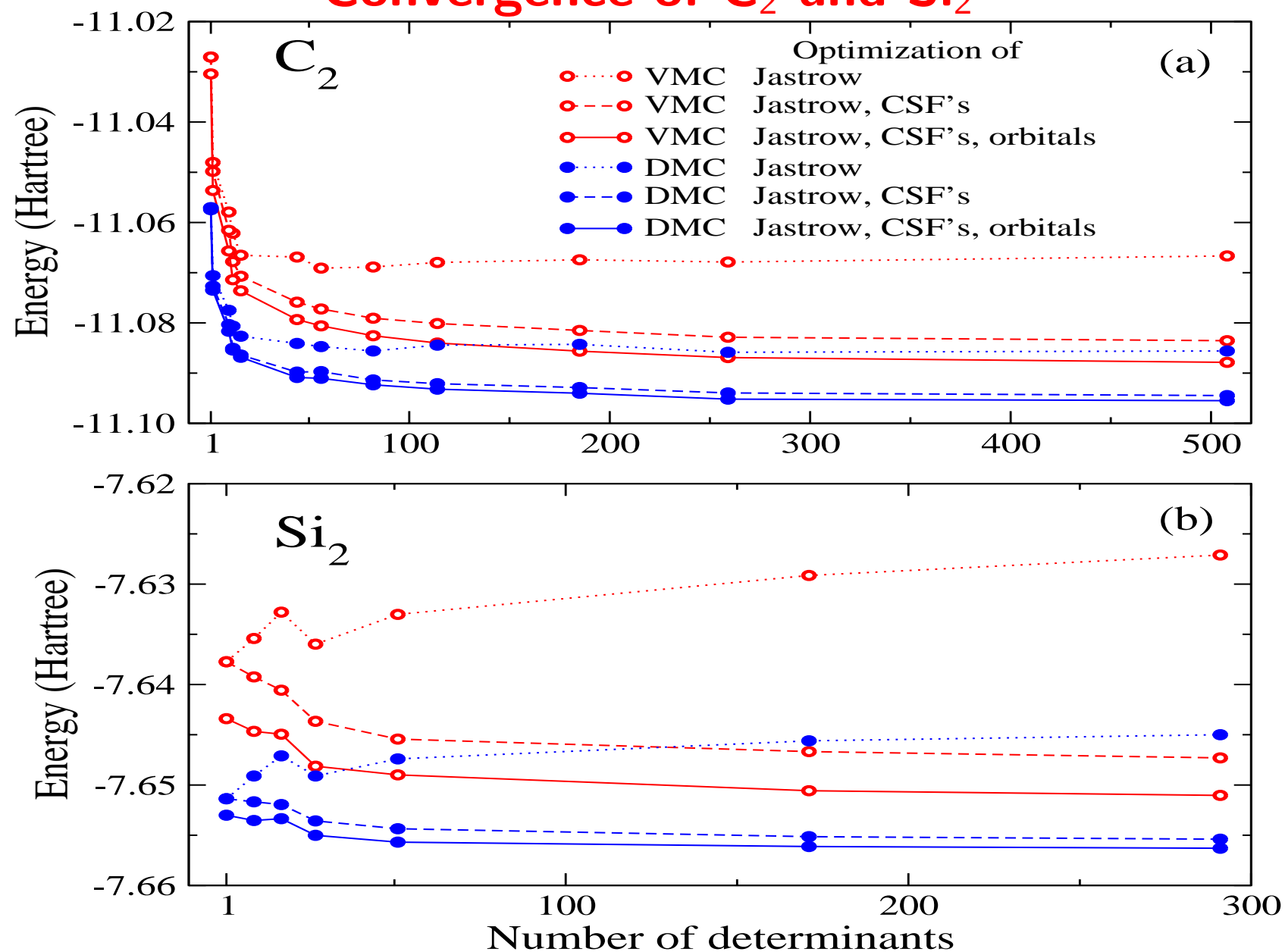


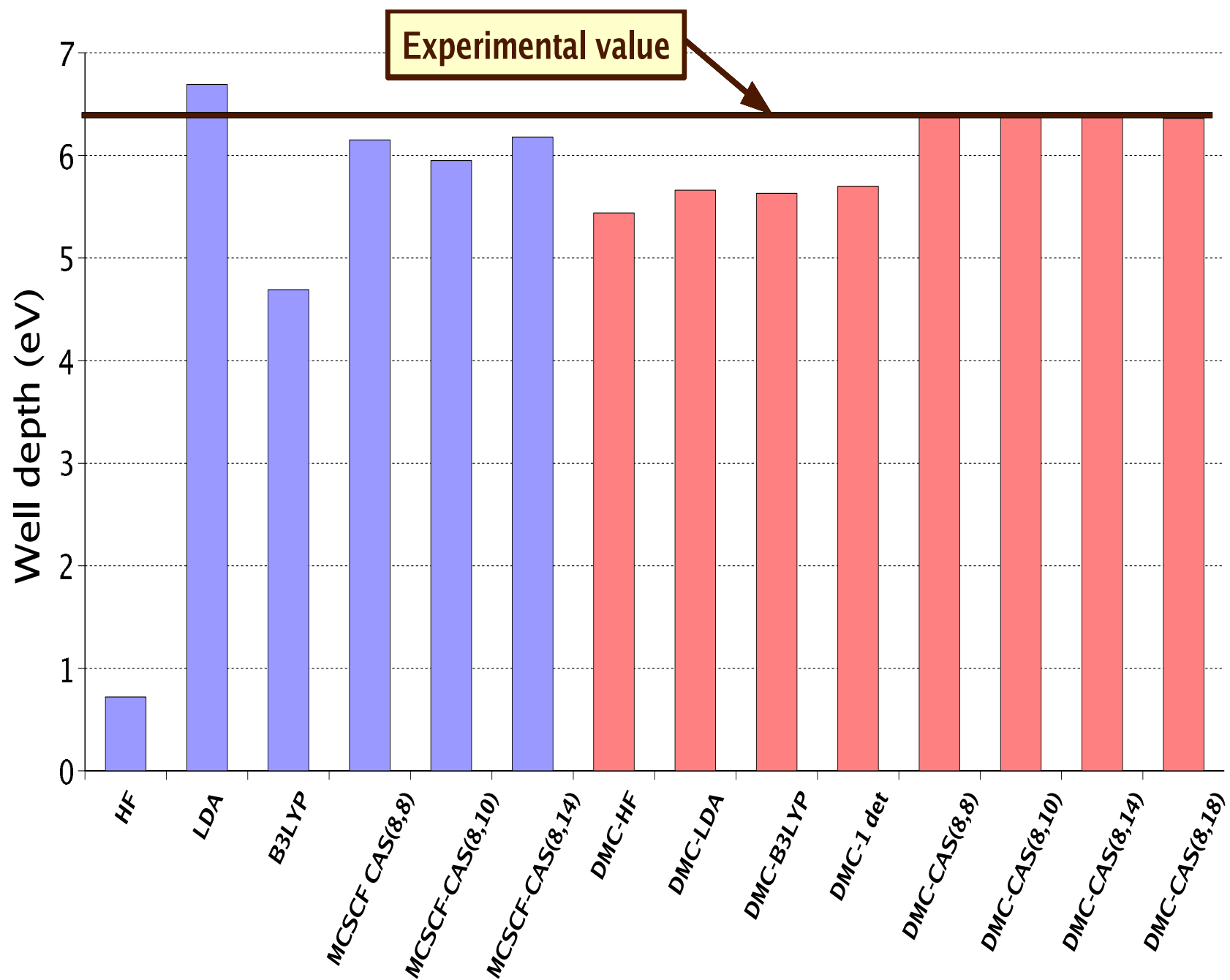
All 3 methods converge in 1 iteration!

## Simultaneous optimization of Jastrow, CSFs and orbitals of all-electron $C_2$ with linear method

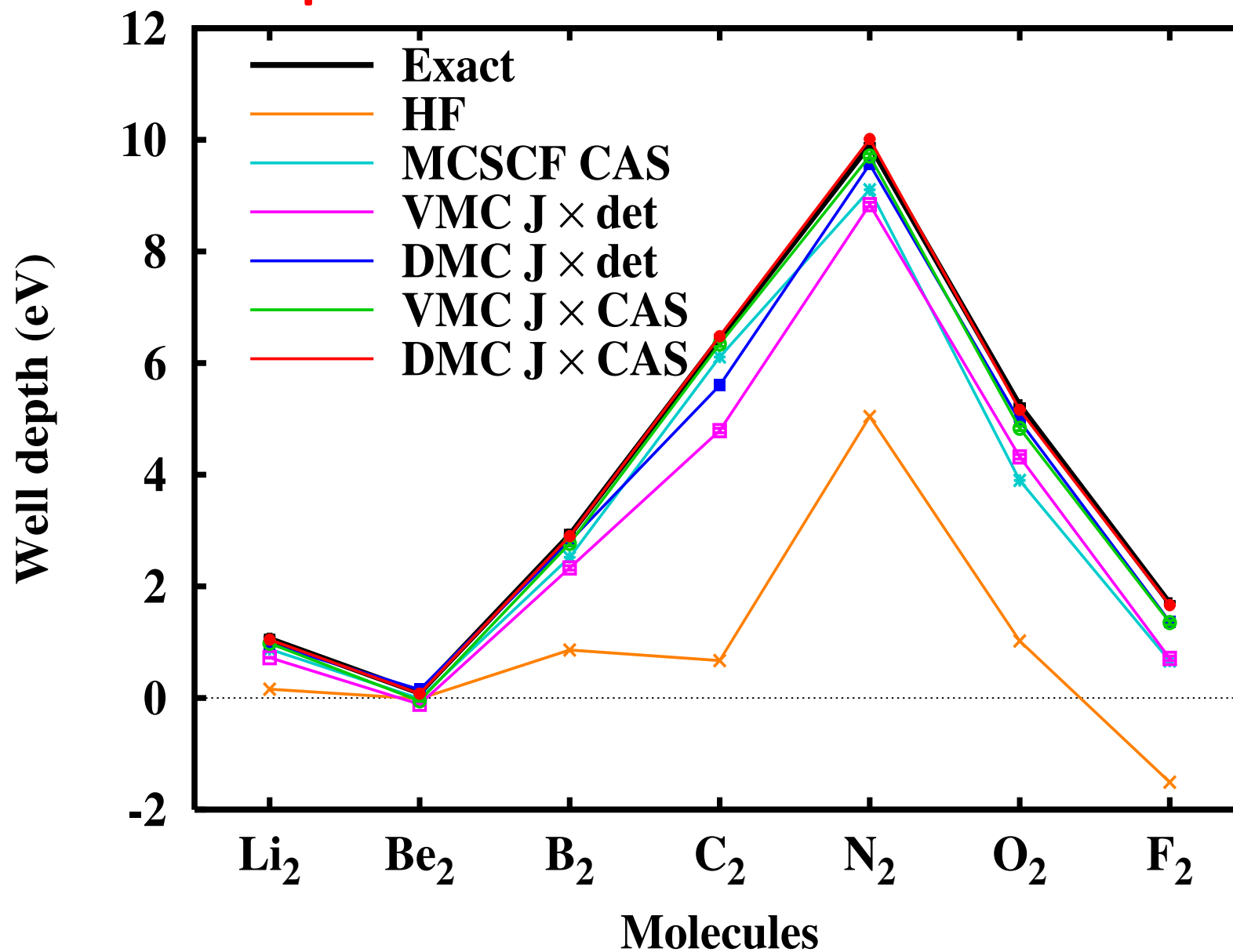


# Convergence of C<sub>2</sub> and Si<sub>2</sub>



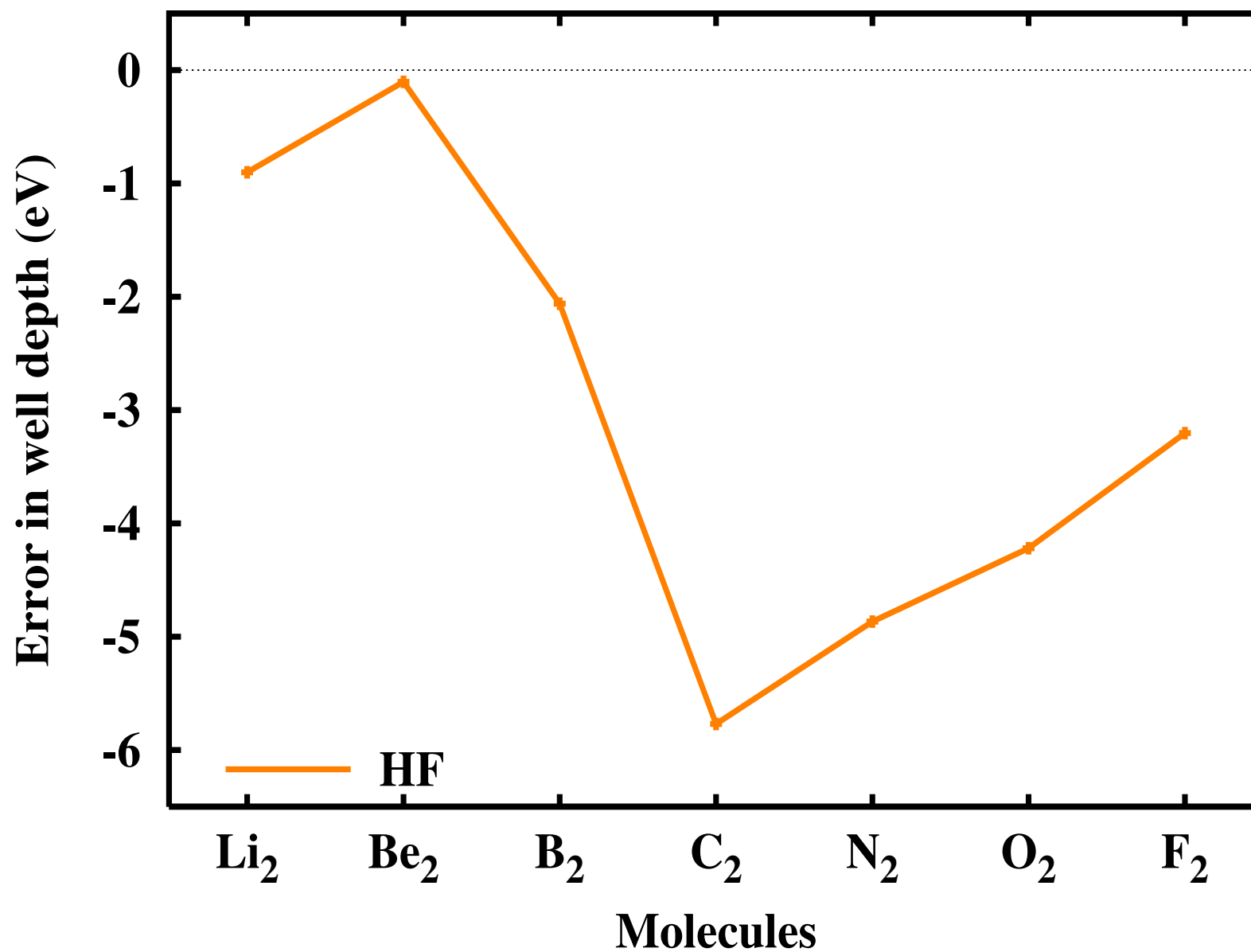


## Well-Depth of 1<sup>st</sup>-Row Diatomic Molecules

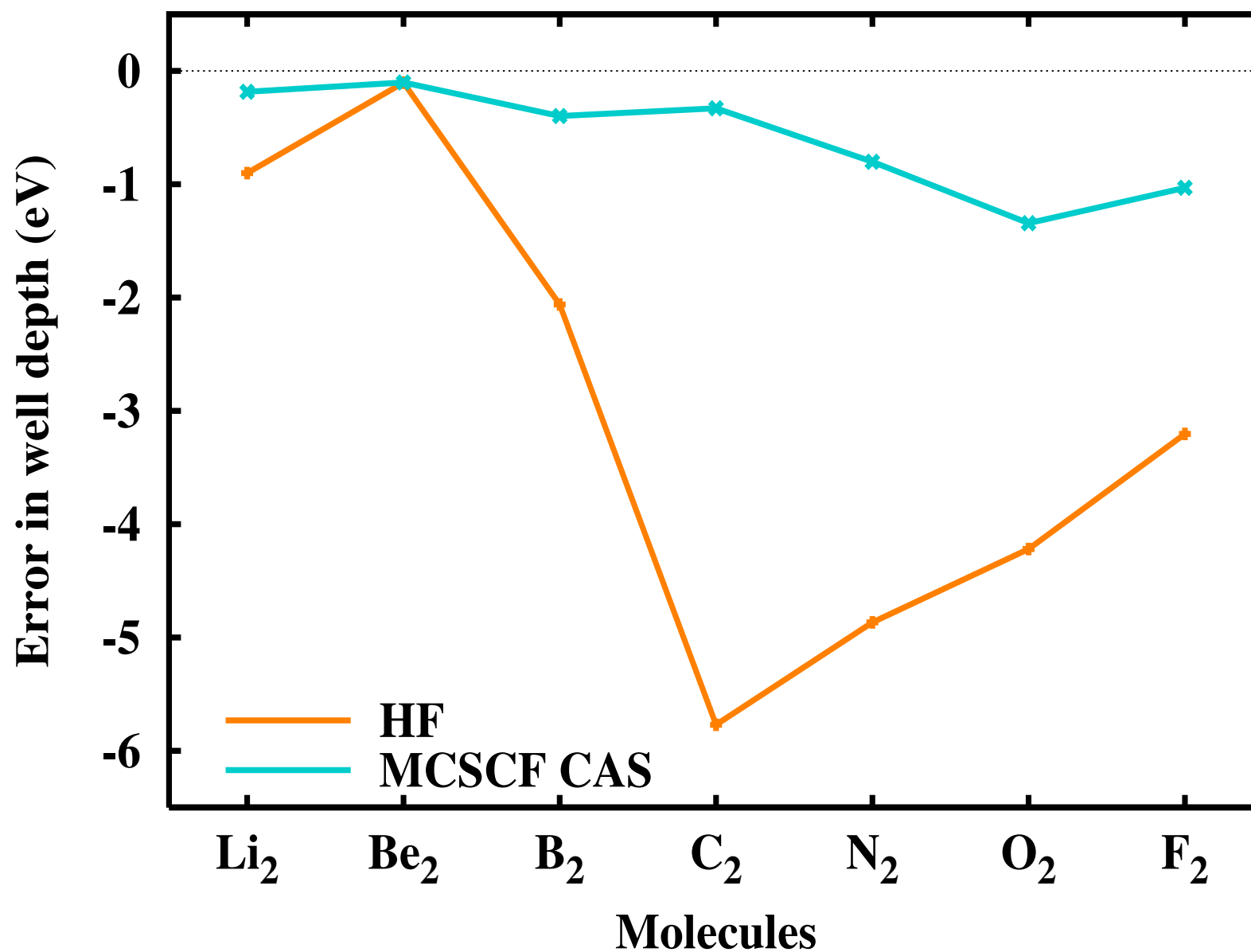




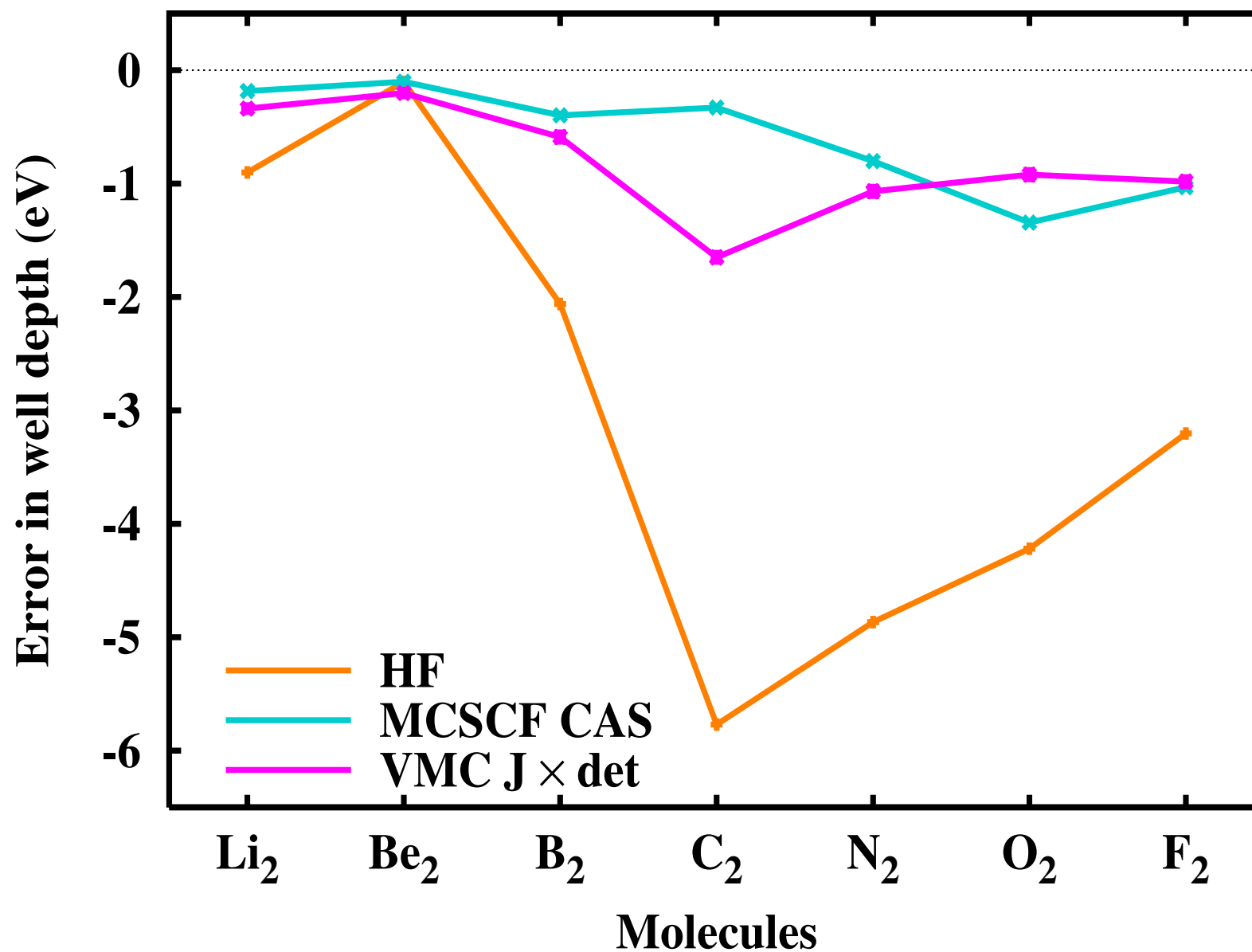
## Error in Well-Depth of 1<sup>st</sup>-Row Diatomic Molecules



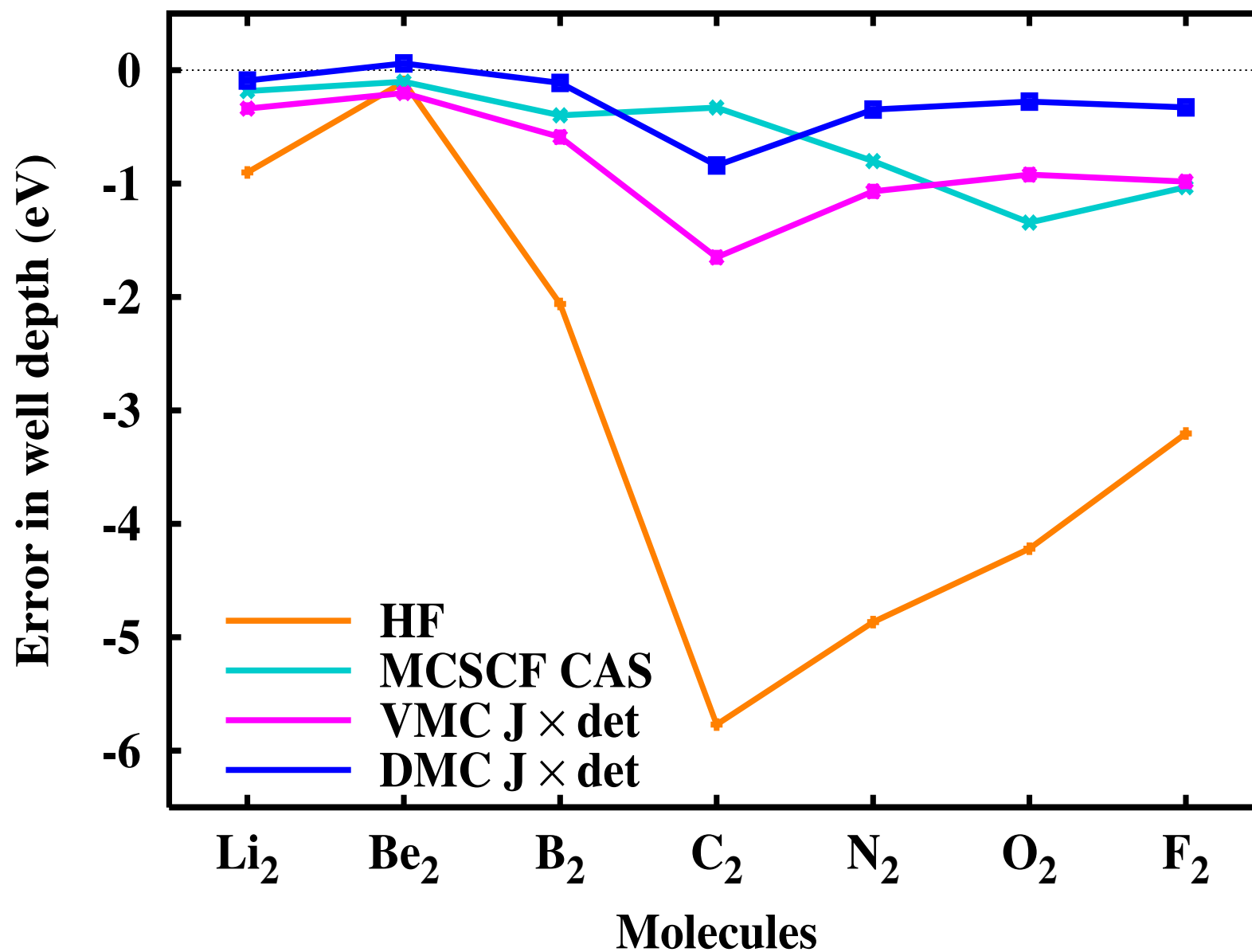
## Error in Well-Depth of 1<sup>st</sup>-Row Diatomic Molecules



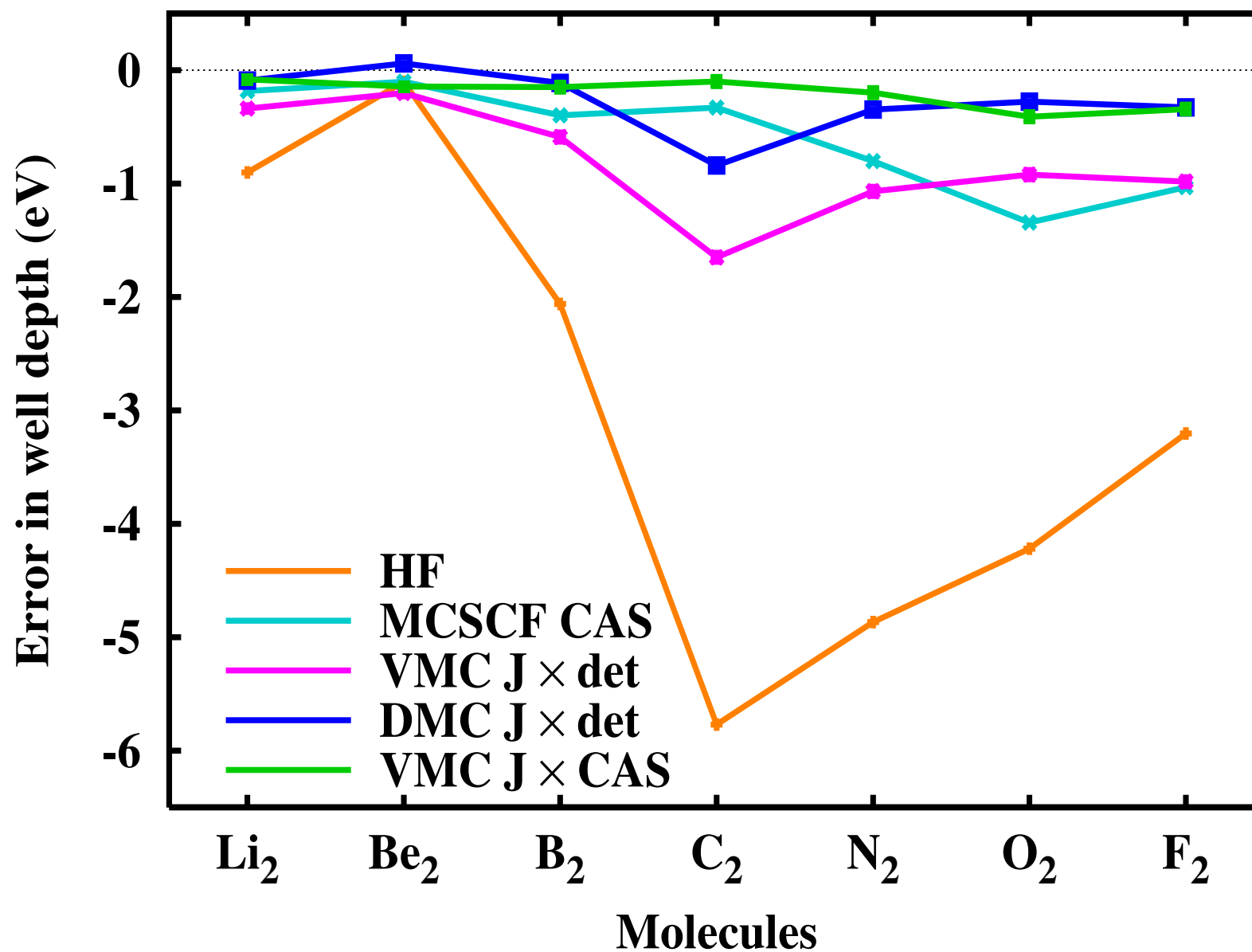
## Error in Well-Depth of 1<sup>st</sup>-Row Diatomic Molecules



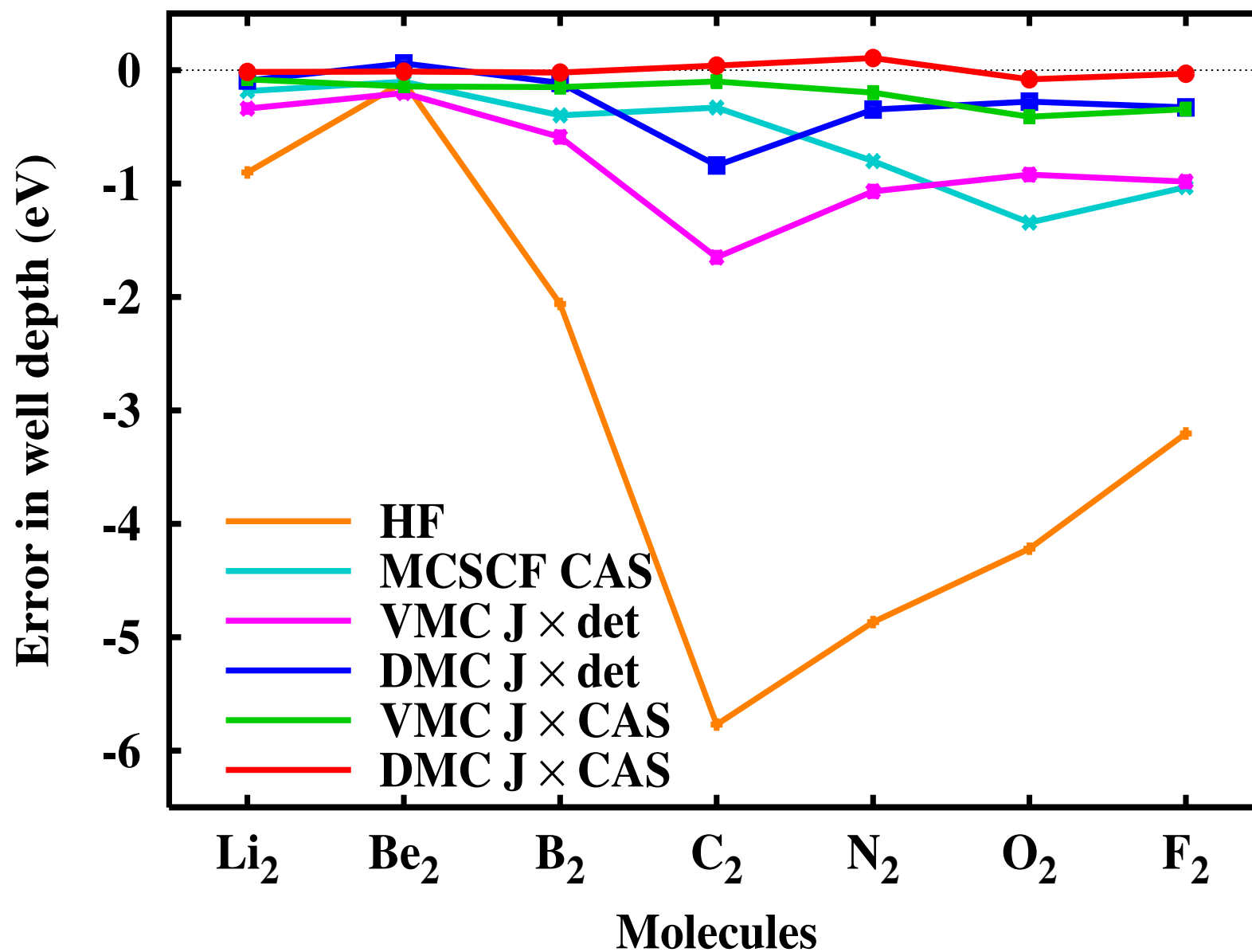
## Error in Well-Depth of 1<sup>st</sup>-Row Diatomic Molecules



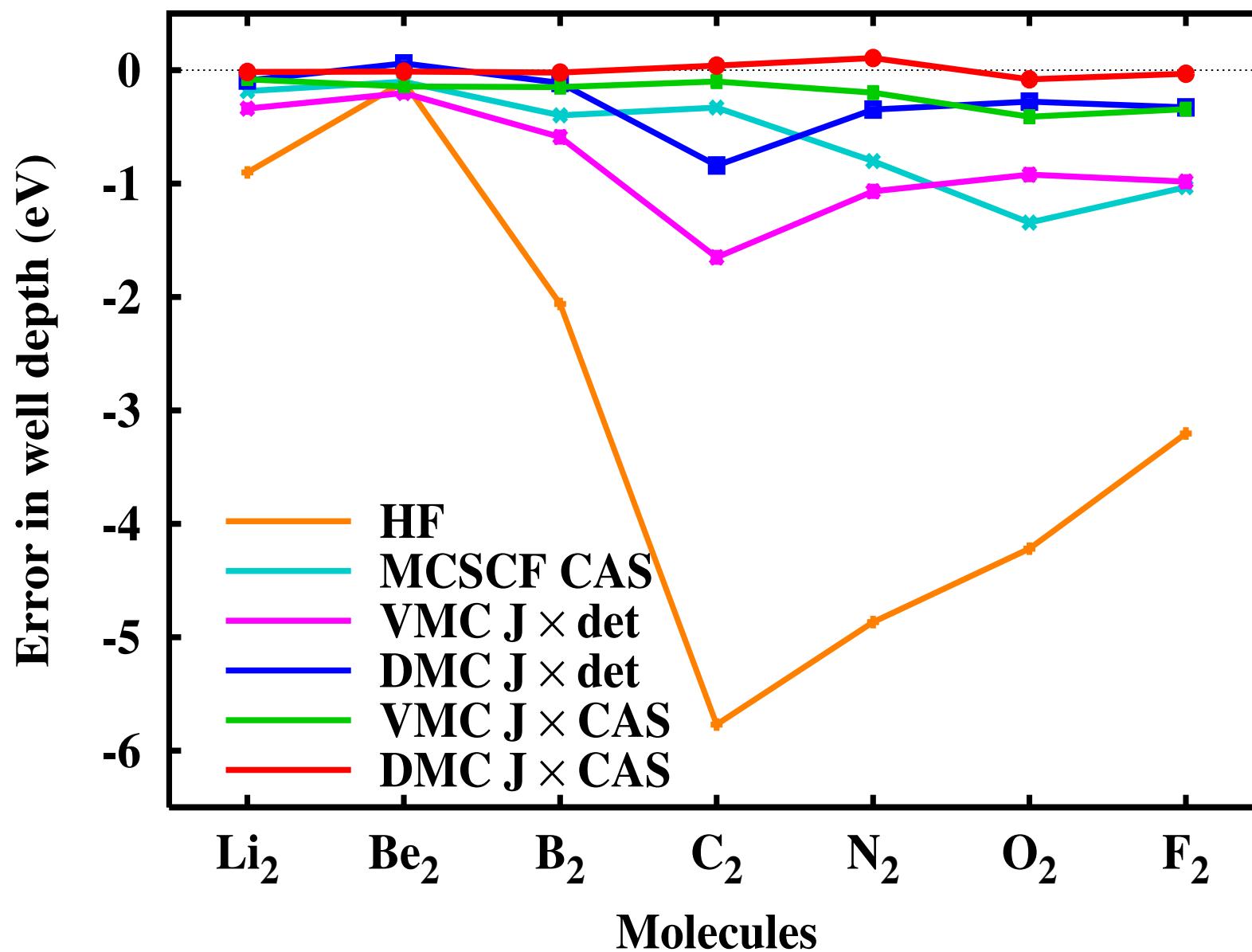
## Error in Well-Depth of 1<sup>st</sup>-Row Diatomic Molecules



## Error in Well-Depth of 1<sup>st</sup>-Row Diatomic Molecules



## Error in Well-Depth of 1<sup>st</sup>-Row Diatomic Molecules



## Measures of goodness of variational wave functions

Expectation value of the energy, variance of the local energy and square of overlap with the ground state ( $1^{st}$ ,  $2^{nd}$  and  $0^{th}$  moments of  $\hat{H}$ ) are:

$$E_T = \frac{\langle \psi_T | H | \psi_T \rangle}{\langle \psi_T | \psi_T \rangle} = \langle E_L \rangle_{|\psi_T|^2} = \frac{1}{N} \sum_i E_L(\mathbf{R}_i)$$

$$\begin{aligned} \sigma^2 &= \frac{\langle \psi_T | (H - E_T)^2 | \psi_T \rangle}{\langle \psi_T | \psi_T \rangle} = \langle E_L^2(\mathbf{R}_i) \rangle_{|\psi_T|^2} - \langle E_L(\mathbf{R}_i) \rangle_{|\psi_T|^2}^2 \\ &= \frac{1}{N} \sum_i E_L^2(\mathbf{R}_i) - E_T^2 \end{aligned}$$

$$\Omega^2 = \frac{|\langle \psi_0 | \psi_T \rangle|^2}{\langle \psi_0 | \psi_0 \rangle \langle \psi_T | \psi_T \rangle} = \frac{\left| \left\langle \frac{\psi_0}{\psi_T} \right\rangle_{|\psi_T|^2} \right|^2}{\left\langle \left| \frac{\psi_0}{\psi_T} \right|^2 \right\rangle_{|\psi_T|^2}}$$

Since  $\psi_0$  is not known, we cannot evaluate  $\Omega$  as written above. In principle we can optimize the wave function by minimizing either  $E_T$  or  $\sigma$  or by maximizing  $\Omega$ . For an infinitely flexible wave function all three will yield the same result, except that minimizing  $\sigma$  could yield an excited state.



# Measures of goodness of variational wave funcs

Define (Mora and Waintal, PRL 2007)

$$\begin{aligned}\Psi(\tau) &= e^{-H\tau} \psi_T = \sum_{i=0}^{\infty} e^{-E_i \tau} c_i \Psi_i \\ Z(\tau) &= \langle \psi_T | e^{-H\tau} | \psi_T \rangle \\ E(\tau) &= \frac{\langle \psi_T | H e^{-H\tau} | \psi_T \rangle}{\langle \psi_T | e^{-H\tau} | \psi_T \rangle} = -\frac{\partial}{\partial \tau} \ln Z(\tau)\end{aligned}$$

Then

$$\lim_{\tau \rightarrow \infty} \Psi(\tau) = \lim_{\tau \rightarrow \infty} e^{-E_0 \tau} c_0 \Psi_0$$

Integrating  $E(\tau)$  we get

$$\ln Z(\tau_2) - \ln Z(\tau_1) = - \int_{\tau_1}^{\tau_2} E(\tau) d\tau$$

$$\Omega^2 = \frac{|\langle \psi_0 | \psi_T \rangle|^2}{\langle \psi_0 | \psi_0 \rangle \langle \psi_T | \psi_T \rangle} = \frac{\lim_{\tau \rightarrow \infty} \left| \frac{e^{E_0 \tau}}{c_0} Z(\tau) \right|^2}{\lim_{\tau \rightarrow \infty} \frac{e^{2E_0 \tau}}{c_0^2} Z(2\tau) Z(0)} = \lim_{\tau \rightarrow \infty} \frac{|Z(\tau)|^2}{Z(2\tau) Z(0)}$$

## Measures of goodness of variational wave functions

$$\begin{aligned}\ln \Omega^2 &= \lim_{\tau \rightarrow \infty} [2 \ln(Z(\tau)) - \ln Z(2\tau) - \ln Z(0)] \\ &= \lim_{\tau \rightarrow \infty} [\ln(Z(\tau)) - \ln(0) - \{\ln Z(2\tau) - \ln Z(\tau)\}] \\ &= - \lim_{\tau \rightarrow \infty} \left[ \int_0^\tau E(\tau) d\tau - \int_\tau^{2\tau} E(2\tau) d\tau \right] = - \int_0^\infty (E(\tau) - E_0) d\tau\end{aligned}$$

So,

$$\Omega = e^{-\kappa}, \quad \text{where,} \quad \kappa = \frac{1}{2} \int_0^\infty (E(\tau) - E_0) d\tau$$

So, overlap can be deduced from area underneath  $(E(\tau) - E_0)$  curve.

# Measures of goodness of variational wave functions

Some things to note:

1.  $E(\tau)$  must be a monotonically decreasing function of  $\tau$ , but,  $\partial E/\partial \tau$  need not be monotonic. If  $\psi_T$  is close to an excited state and has only a very tiny component in the ground state then  $E(\tau)$  will have a tiny negative slope at  $\tau = 0$  and the magnitude of the slope will increase as the ground state component increases before it eventually decreases to zero. If we start off with no component in the ground state then we will converge to the lowest excited state for which  $\psi_T$  has a nonzero component.
2. If two wave functions have the same variational energy but different  $\sigma$ 's, then the one with the **larger**  $\sigma$  will likely have the larger overlap with the ground state.
3. In practice this overlap is useful only for choosing between trial wave functions for bosonic ground states. For other states one gets the overlap of each trial wave function with its own fixed-node wave function, which is not very useful.

# Conclusions

1. Three simple, robust and highly efficient methods for energy optimization.  
Newton and linear methods work for all the parameters.  
Perturbation theory is least expensive method for optimizing a large number of orbital parameters but slow for Jastrow.
2. Methods have been used on large systems – 1000 electrons in solid phases of Silicon.
3. All 3 methods have been used to optimize hundreds of parameters and perturbative method for a few thousand orbital parameters (Sorella, Filippi).
4. Optimization of the determinantal coefficients can be very important for reducing fixed-node error of  $E_{\text{DMC}}$ .
5. Seemingly similar molecules, e.g.  $\text{C}_2$  and  $\text{Si}_2$  have fixed-node errors for single-determinant wave functions that differ by a factor of 10.