Minimization with noisy surfaces: geometry optimization in QMC

Lucas K. Wagner, Jeffrey C. Grossman

Computational Nanosciences Group
Berkeley Nanoscience and Nanoengineering Institute
University of California, Berkeley

Geometry optimization in QMC

- Long standing challenge in QMC
- Typically use DFT or experiment, but
  - Transition metals
  - Weak binding/van der Waals
  - Excited states
- Much work on forces[1] and correlated sampling[2]

Our goal:

- Find precise minima in the presence of noise

---

[1] Assaraf & Caffarel, Chiesa et al., Rappe group, Needs group

[2] Filippi & Umrigar, Pierleoni & Ceperley

Geometry optimization in QMC

- Long standing challenge in QMC
- Typically use DFT or experiment, but
  - Transition metals
  - Weak binding/van der Waals
  - Excited states
- Much work on forces[1] and correlated sampling[2]

Our goal:

- Find precise minima in the presence of noise

[1] Assaraf & Caffarel, Chiesa et al., Rappe group, Needs group
[2] Filippi & Umrigar, Pierleoni & Ceperley

Geometry optimization in QMC

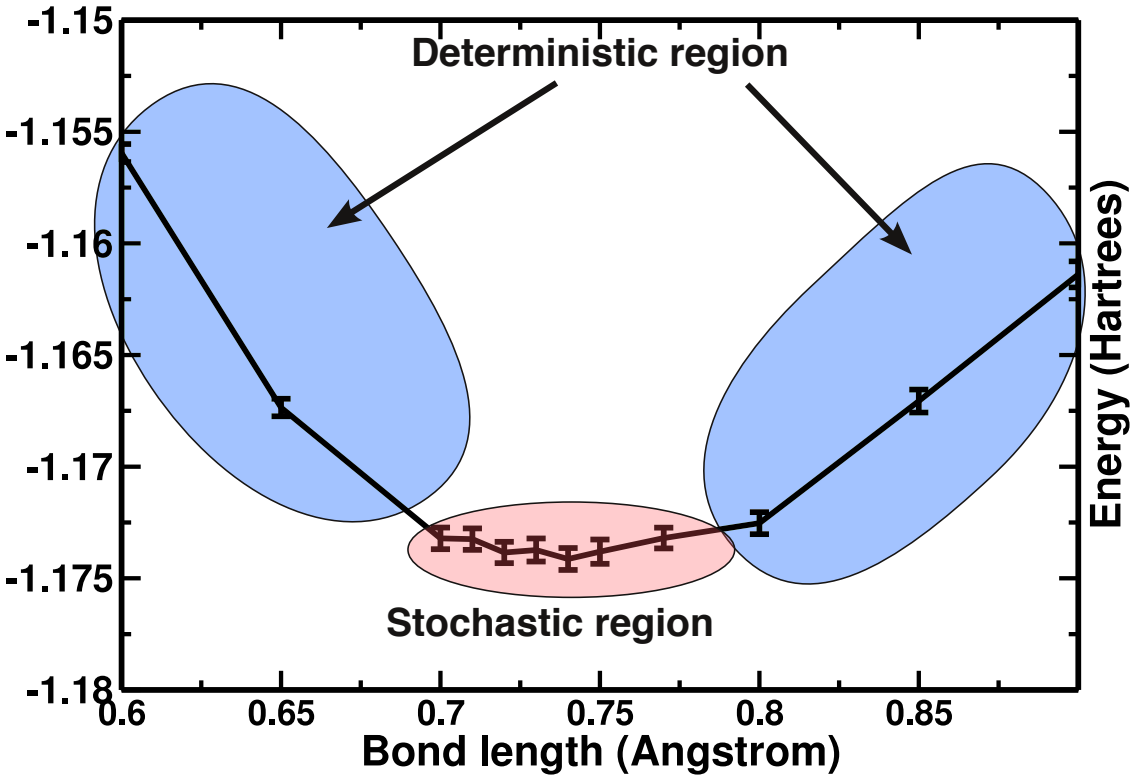
- Long standing challenge in QMC
- Typically use DFT or experiment, but
  - Transition metals
  - Weak binding/van der Waals
  - Excited states
- Much work on forces[1] and correlated sampling[2]

Our goal:

- Find precise minima in the presence of noise

[1] Assaraf & Caffarel, Chiesa et al., Rappe group, Needs group
[2] Filippi & Umrigar, Pierleoni & Ceperley

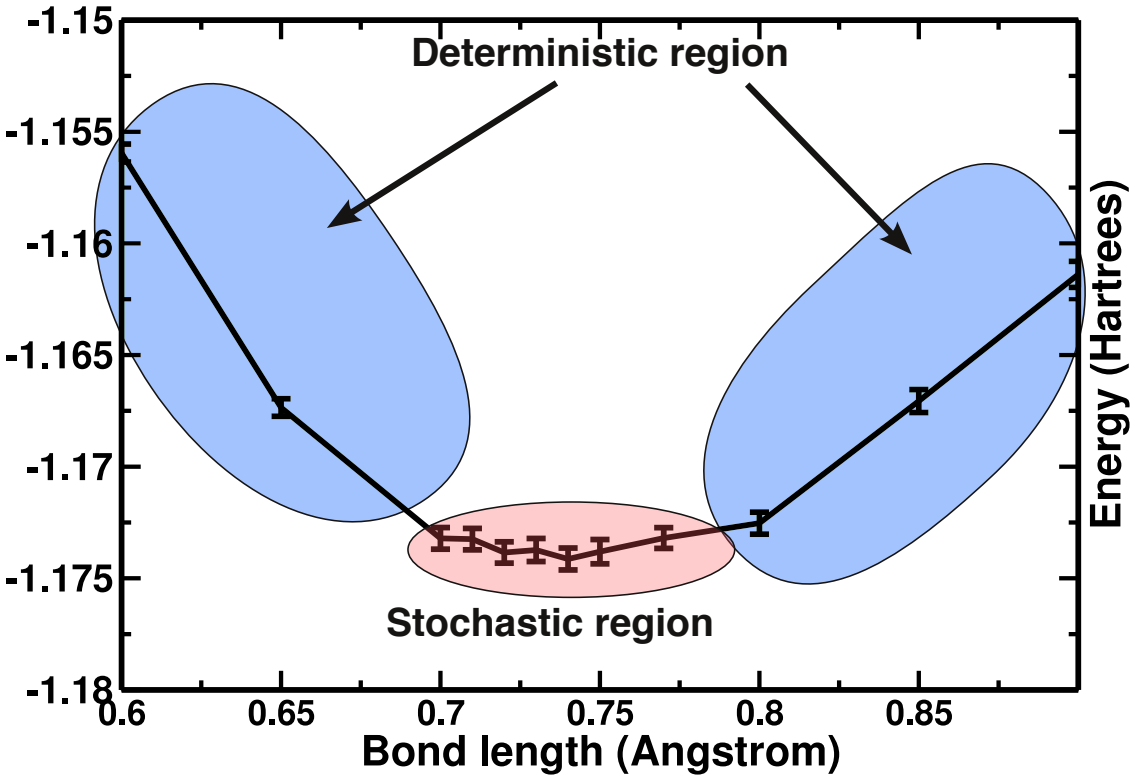
Deterministic minimization vs stochastic



- In the deterministic region, can use normal search methods
- In the stochastic region, noise dominates, must use different algorithms

Most work concentrates on reducing the size of the stochastic region. Notable exception: stochastic gradient approximation (Monro in 50's, Sorella et al. in 90's)

Deterministic minimization vs stochastic



- In the deterministic region, can use normal search methods
- In the stochastic region, noise dominates, must use different algorithms

Most work concentrates on reducing the size of the stochastic region. Notable exception: stochastic gradient approximation (Monro in 50's, Sorella et al. in 90's)

Computational costs to reduce size of the stochastic region

- $E = \alpha \delta x^2$
- Total energy: $\epsilon_x \propto \sqrt{\epsilon_E} \propto T^{-1/4}$
- Forces: $\epsilon_x \propto \epsilon_F \propto T^{-1/2}$
- Fitting not quite so simple, but: $\epsilon_x \propto T^{-1/2}$
- Prefactors are important: if $\alpha\epsilon$ costs α^2

Really want to use forces or fitting; otherwise the stochastic region will be huge for no good reason.

Computational costs to reduce size of the stochastic region

- $E = \alpha \delta x^2$
- Total energy: $\epsilon_x \propto \sqrt{\epsilon_E} \propto T^{-1/4}$
- Forces: $\epsilon_x \propto \epsilon_F \propto T^{-1/2}$
- Fitting not quite so simple, but: $\epsilon_x \propto T^{-1/2}$
- Prefactors are important: if $\alpha\epsilon$ costs α^2

Really want to use forces or fitting; otherwise the stochastic region will be huge for no good reason.

Why operate in the stochastic region?

- We often don't know *a priori* how big the stochastic region will be.
- Can do multiple deterministic runs, but that's somewhat heuristic and unreasonably expensive
- Cost to reduce errors is $\propto T^{-1/2}$

Why am I not using forces?

- Original answer: Forces in QMC are hard!
- More nuanced answer:
  I'll talk in a general way about adding forces here and there.

Why operate in the stochastic region?

- We often don't know *a priori* how big the stochastic region will be.
- Can do multiple deterministic runs, but that's somewhat heuristic and unreasonably expensive
- Cost to reduce errors is $\propto T^{-1/2}$

Why am I not using forces?

- Original answer: Forces in QMC are hard!
- More nuanced answer:
  I'll talk in a general way about adding forces here and there.

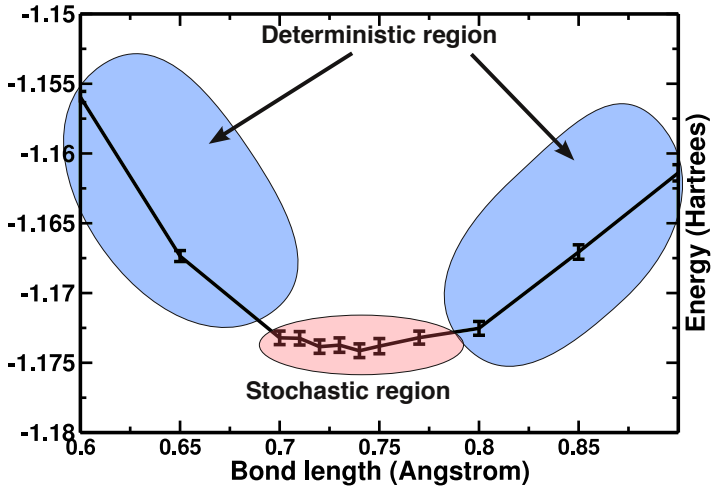
Why operate in the stochastic region?

- We often don't know *a priori* how big the stochastic region will be.
- Can do multiple deterministic runs, but that's somewhat heuristic and unreasonably expensive
- Cost to reduce errors is $\propto T^{-1/2}$

Why am I not using forces?

- Original answer: Forces in QMC are hard!
- More nuanced answer:
  I'll talk in a general way about adding forces here and there.

Advantage of fitting



- Fitting gives uncertainty of .003 Å
- Straight energy differences(estimated): 5000x more expensive
- Forces[3](estimated): 20x more expensive

Drastic decrease in the size of the stochastic region!

$\Longrightarrow$ big gain in efficiency!

But difficult to apply to many dimensions.

---

[3] Assaraf & Caffarel, JCP **113** 4028

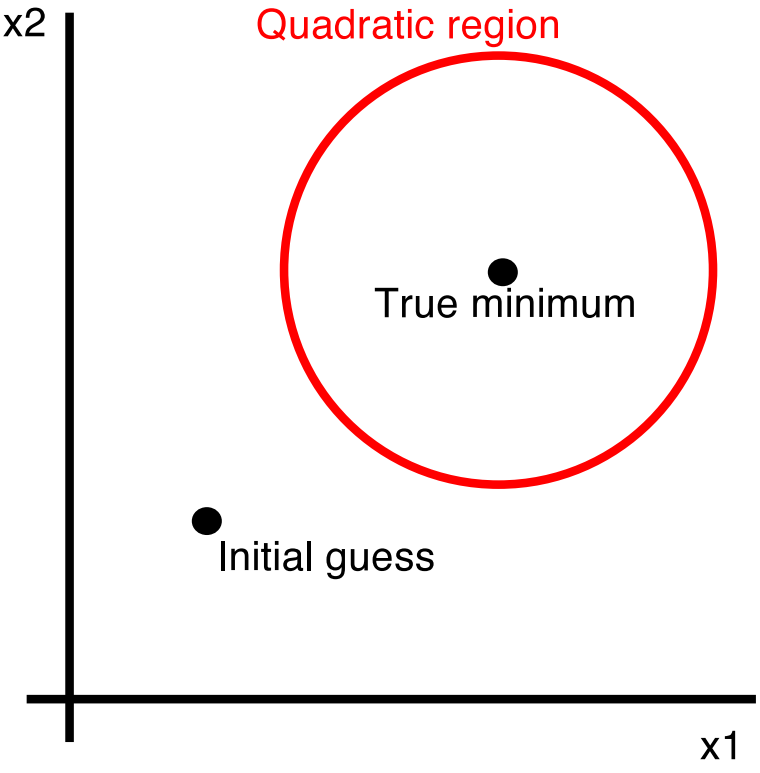
A stochastic process of line minimizations

1. Guess the minimum $\mathbf{x}^{(0)}$.
   Define $\delta\mathbf{x}^{(0)} = \mathbf{x}^{(0)} - m$.
2. Near the minimum,
   $E(\mathbf{x}) = E_0 + \delta\mathbf{x}^T M \delta\mathbf{x}$.
3. minimize along direction i:

$$\delta x_i^{(1)} = \chi_i^{(1)} + \sum_{j\neq i} \frac{M_{ij}}{M_{ii}} \delta\mathbf{x}_j^{(0)}$$

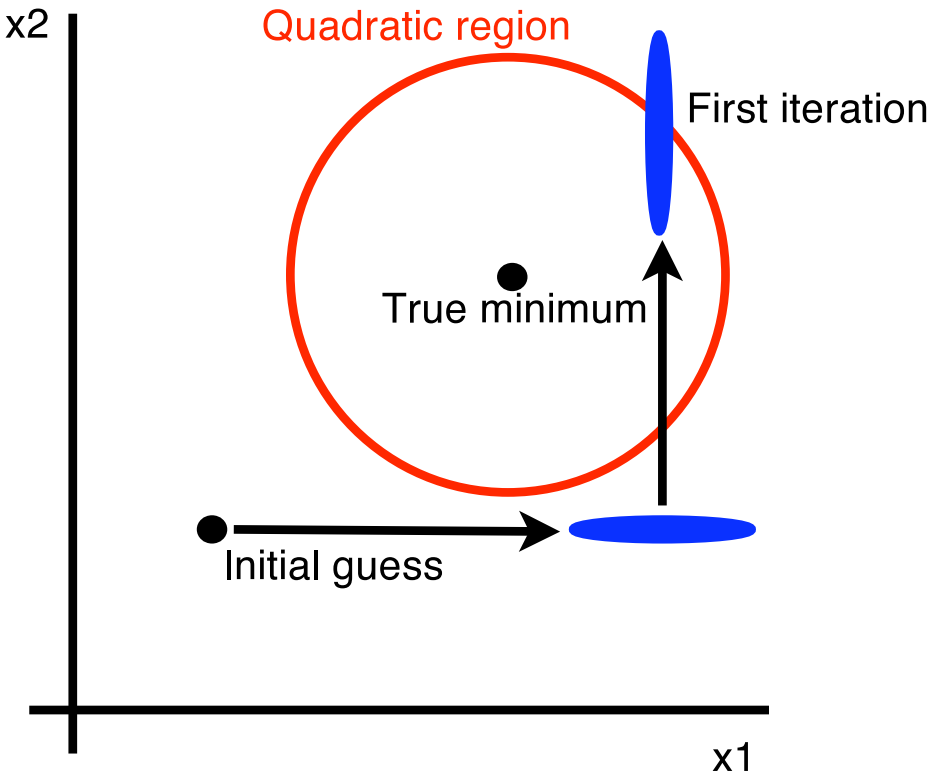
   Total = Noise + Systematic
4. n'th iteration,

$$\delta x_i^{(n)} = \chi_i^{(n)} - \sum_{j\neq i} \frac{M_{ij}}{M_{ii}} \chi_j^{(n-1)} + \sum_{j\neq i} \sum_{k\neq j} \frac{M_{ij}M_{jk}}{M_{ii}M_{jj}} \chi_k^{(n-2)} + \ldots$$



A stochastic process of line minimizations

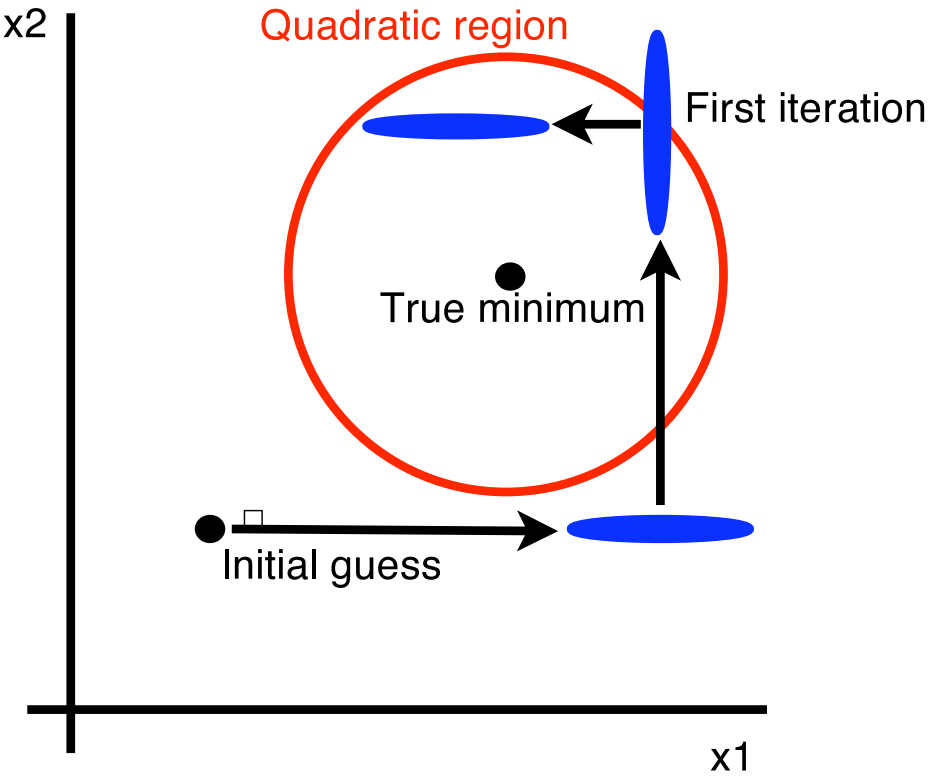
1. Guess the minimum $\mathbf{x}^{(0)}$.
   Define $\delta\mathbf{x}^{(0)} = \mathbf{x}^{(0)} - m$.
2. Near the minimum,
   $E(\mathbf{x}) = E_0 + \delta\mathbf{x}^T M \delta\mathbf{x}$.
3. minimize along direction i:
   $\delta x_i^{(1)} = \chi_i^{(1)} + \sum_{j\neq i} \frac{M_{ij}}{M_{ii}} \delta\mathbf{x}_j^{(0)}$
   Total Noise Systematic
4. n'th iteration,



$$\delta x_i^{(n)} = \chi_i^{(n)} - \sum_{j\neq i} \frac{M_{ij}}{M_{ii}} \chi_j^{(n-1)} + \sum_{j\neq i} \sum_{k\neq j} \frac{M_{ij}M_{jk}}{M_{ii}M_{jj}} \chi_k^{(n-2)} + \ldots$$

A stochastic process of line minimizations

1. Guess the minimum $\mathbf{x}^{(0)}$.
   Define $\delta\mathbf{x}^{(0)} = \mathbf{x}^{(0)} - m$.
2. Near the minimum,
   $E(\mathbf{x}) = E_0 + \delta\mathbf{x}^T M \delta\mathbf{x}$.
3. minimize along direction i:
   $\delta x_i^{(1)} = \chi_i^{(1)} + \sum_{j\neq i} \frac{M_{ij}}{M_{ii}} \delta\mathbf{x}_j^{(0)}$
   Total Noise Systematic
4. n'th iteration,



$$\delta x_i^{(n)} = \chi_i^{(n)} - \sum_{j\neq i} \frac{M_{ij}}{M_{ii}} \chi_j^{(n-1)} + \sum_{j\neq i} \sum_{k\neq j} \frac{M_{ij}M_{jk}}{M_{ii}M_{jj}} \chi_k^{(n-2)} + \ldots$$

Properties of the sequence

$$\delta x_i^{(n)} = \chi_i^{(n)} - \sum_{j \neq i} \frac{M_{ij}}{M_{ii}} \chi_j^{(n-1)} + \sum_{j \neq i} \sum_{k \neq j} \frac{M_{ij} M_{jk}}{M_{ii} M_{jj}} \chi_k^{(n-2)} + \ldots$$

- Conditions
  - Hessian is diagonally dominant: $\sum_{j \neq i} \left| \frac{M_{ij}}{M_{ii}} \right| < 1, \forall i$ (sufficient, but not necessary)
  - Noise is not biased: $\langle \chi \rangle = 0$

- Reduces to a geometric series
  - Finite autocorrelation time
  - Average of points equals the exact minimum
  - Variance is finite

Properties of the sequence

$$\delta x_i^{(n)} = \chi_i^{(n)} - \sum_{j \neq i} \frac{M_{ij}}{M_{ii}} \chi_j^{(n-1)} + \sum_{j \neq i} \sum_{k \neq j} \frac{M_{ij} M_{jk}}{M_{ii} M_{jj}} \chi_k^{(n-2)} + \ldots$$

- Conditions
  - Hessian is diagonally dominant: $\sum_{j \neq i} \left| \frac{M_{ij}}{M_{ii}} \right| < 1, \forall i$ (sufficient, but not necessary)
  - Noise is not biased: $\langle \chi \rangle = 0$

- Reduces to a geometric series
  - Finite autocorrelation time
  - Average of points equals the exact minimum
  - Variance is finite

Scaling analysis

- Each pass is $\mathcal{O}(N_{DOF})$

$$N_{pass} \propto \frac{\sigma_{\delta x}^2}{\epsilon^2}$$

$$\sigma_{\delta x_i}^2 = \sigma_{\chi_i}^2 + \sum_{j \neq i} \frac{M_{ij}^2}{M_{ii}^2} \sigma_{\chi_j}^2 + \sum_j \sum_k \cdots$$

- Scaling we see is dependent on the matrix elements of the Hessian
- For a diagonalized matrix, scaling is $\mathcal{O}(N_{DOF})$

Choosing the search directions

- Intuition
- Phonons from DFT
- Powell/Brent's method
- A few other unexplored ideas..

Traditional determistic methods methods *will not work* in the stochastic region.

Choosing the search directions

- Intuition
- Phonons from DFT
- Powell/Brent's method
- A few other unexplored ideas..

Traditional determistic methods methods *will not work* in the stochastic region.

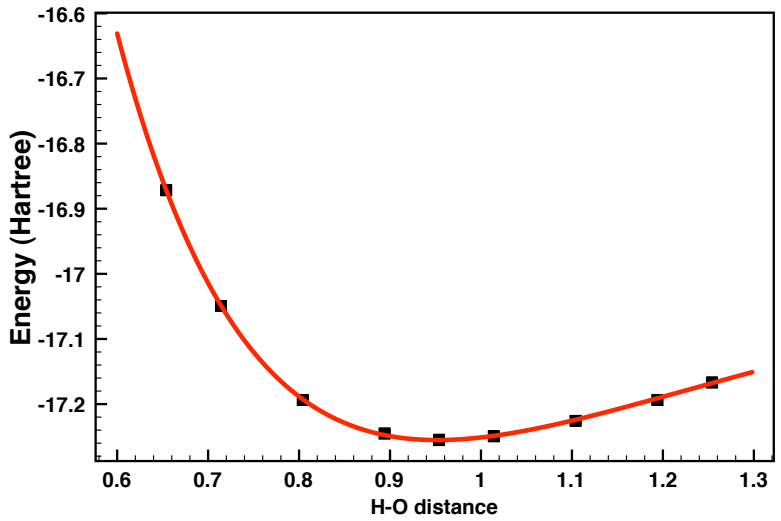
Method summary

1. Use determistic methods to get near the minimum and find search directions
2. Loop through directions
   - Evaluate energy along a line
   - Fit to find minimum
3. Add new point to average, go back to 2.

Applications: calculation details

- Single determinant trial function from GAMESS with gaussian basis set, pseudopotentials[4]
- Two or three-body Jastrow factor
- Diffusion Monte Carlo as implemented in QWalk (www.qwalk.org)
- Bayesian fitting framework[5]
- Modified Morse potential for fitting

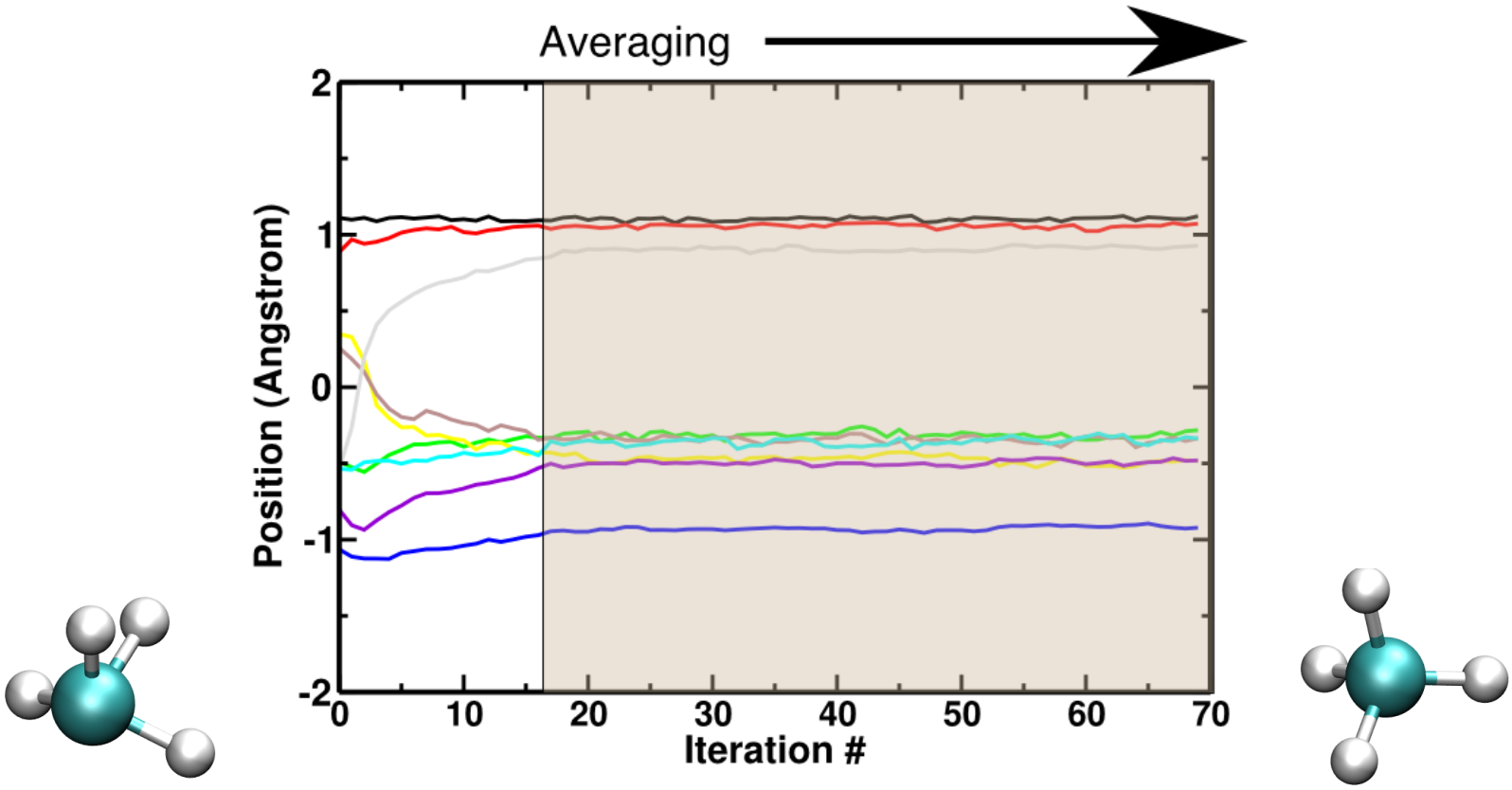
$$E(x) = E_0 + a(1 - \exp(-b(x - x_0)))^2$$



[4] Burkatzki et al. JCP **126**, 234105

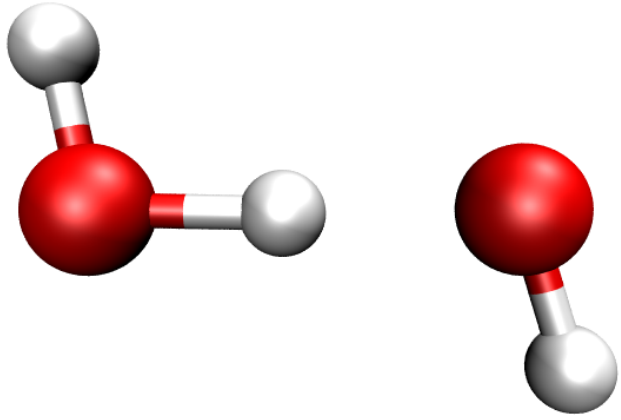
[5] Wagner & Mitas, JCP **126**, 034105

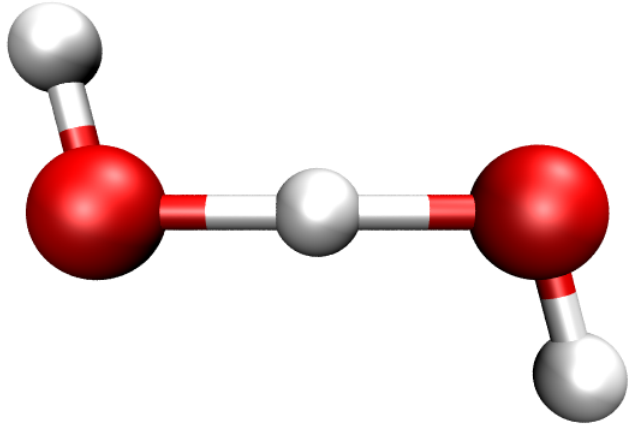
Line minimization example: Methane



- Each line represents a coordinate
- No special direction choices

H_2O-OH^- complex



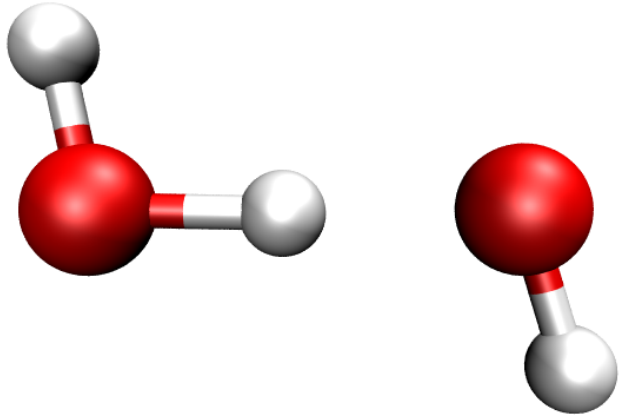


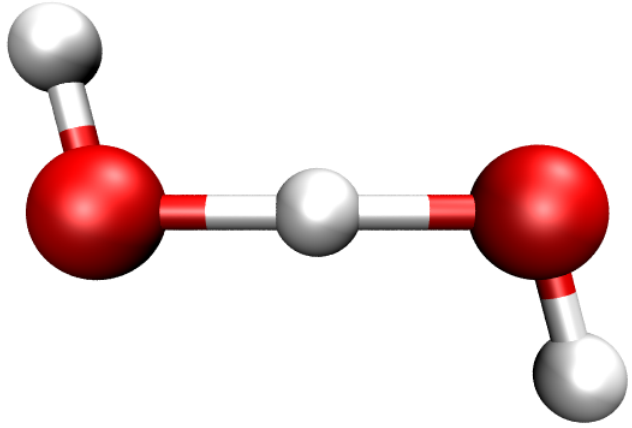
- Traditional picture (structure A)
- Hartree-Fock, MP2

- Centrosymmetric minimum? (structure B)
- LDA, BLYP, PBE

- What is the actual ground state structure?
- And what is the barrier to exchange hydrogen?

H_2O-OH^- complex



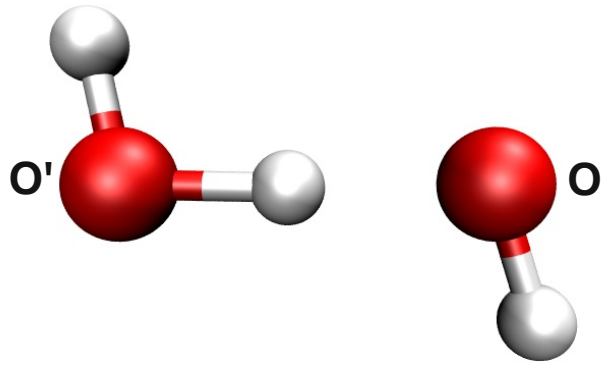


- Traditional picture (structure A)
- Hartree-Fock, MP2

- Centrosymmetric minimum? (structure B)
- LDA, BLYP, PBE

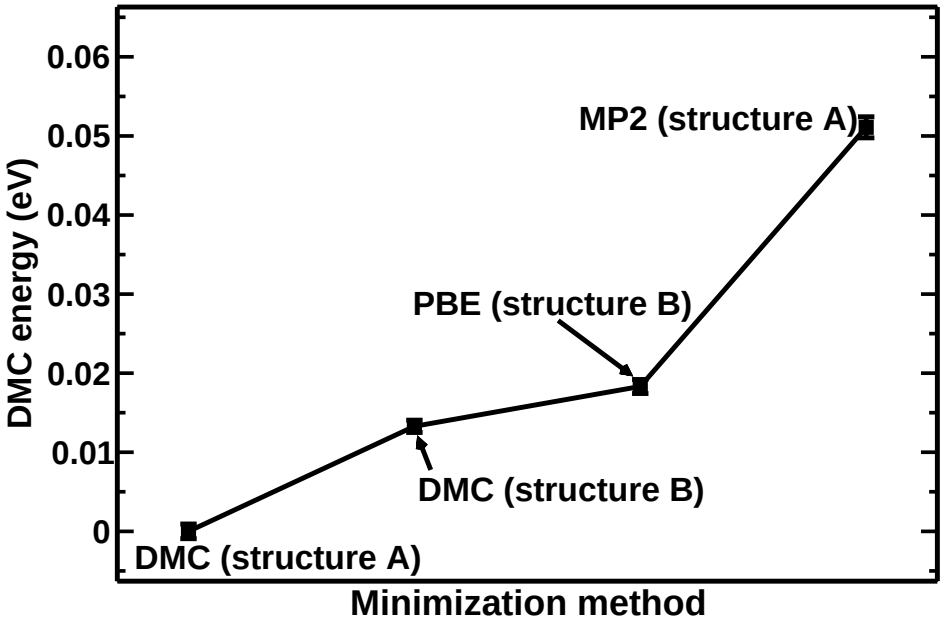
- What is the actual ground state structure?
- And what is the barrier to exchange hydrogen?

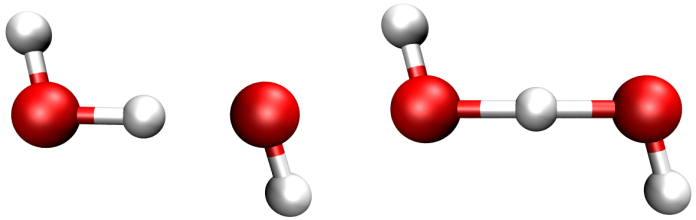
Results



Method	O-O’	O’-H	Structure type
LDA	2.448	1.224	B
PBE	2.470	1.235	B
MP2	2.469	1.123	A
DMC	2.491(2)	1.111(3)	A
DMC*	2.469(3)	1.235(2)	B

H_2O-OH^- complex





Structure A Structure B

- DMC gets different symmetry from GGA/LDA geometries
- MP2 structure is right symmetry, but high energy
- DMC geometry optimization important for finding real minima
- The barrier is quite small: $\sim$0.01 eV!

Conclusion

- New way to obtain precise minimum energy geometries with QMC
- Takes advantage of many low-precision calculations and only energies
- Works for many dimensions (have done nine)
- In some cases, DMC minimum can differ significantly from mean-field theories

Funded by the National Science Foundation

Conclusion

- New way to obtain precise minimum energy geometries with QMC
- Takes advantage of many low-precision calculations and only energies
- Works for many dimensions (have done nine)
- In some cases, DMC minimum can differ significantly from mean-field theories

Funded by the National Science Foundation

Thoughts on adding forces

- An extra $\mathcal{O}(N_{DOF})$ information!
- Can use it to determine search directions and decrease cost of fits
- Could construct a conjugate-gradients type of technique that's stochastically sound
- Still get a huge prefactor improvement for fitting
- Need to be careful about cross-correlations between different components

Thoughts on adding forces

- An extra $\mathcal{O}(N_{DOF})$ information!
- Can use it to determine search directions and decrease cost of fits
- Could construct a conjugate-gradients type of technique that's stochastically sound
- Still get a huge prefactor improvement for fitting
- Need to be careful about cross-correlations between different components

Thoughts on adding forces

- An extra $\mathcal{O}(N_{DOF})$ information!
- Can use it to determine search directions and decrease cost of fits
- Could construct a conjugate-gradients type of technique that's stochastically sound
- Still get a huge prefactor improvement for fitting
- Need to be careful about cross-correlations between different components

Thoughts on adding forces

- An extra $\mathcal{O}(N_{DOF})$ information!
- Can use it to determine search directions and decrease cost of fits
- Could construct a conjugate-gradients type of technique that's stochastically sound
- Still get a huge prefactor improvement for fitting
- Need to be careful about cross-correlations between different components

Thanks, Mike!