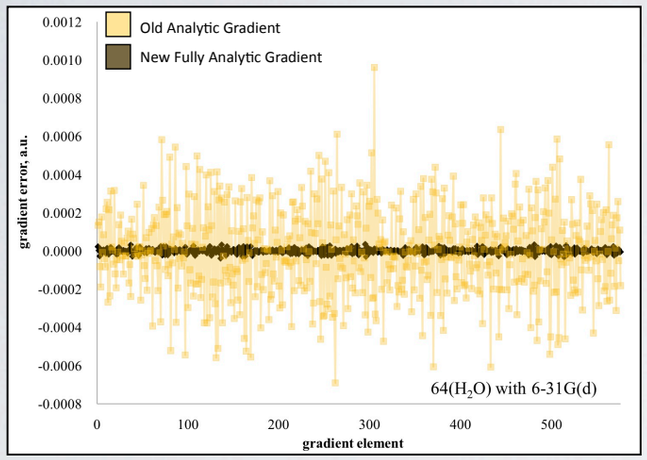
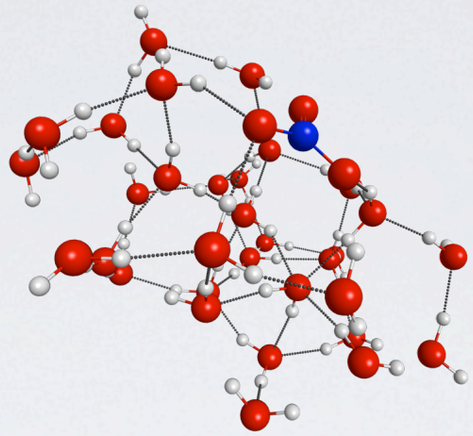
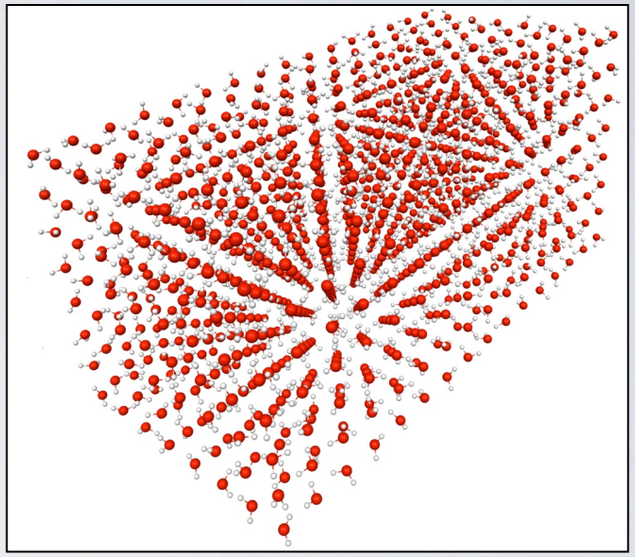
INCITE and ESP Driven Advances in GAMESS

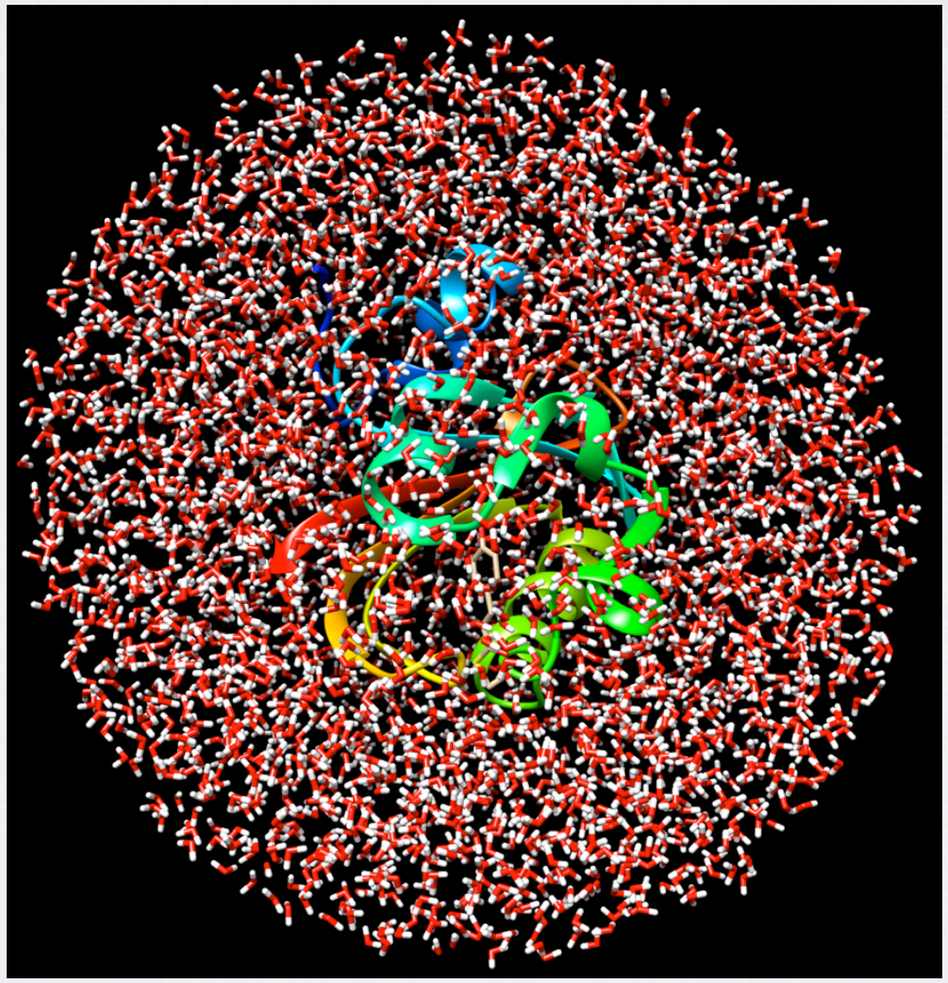
2013 Early Science Program Invesitgators Meeting

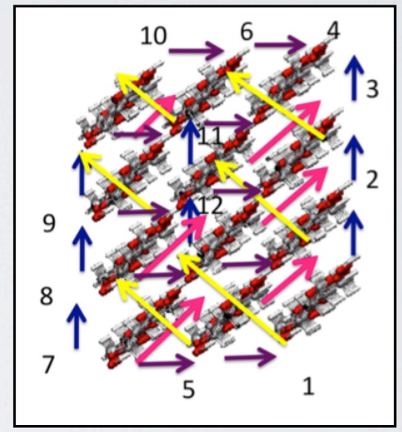
May 15-16th 2013

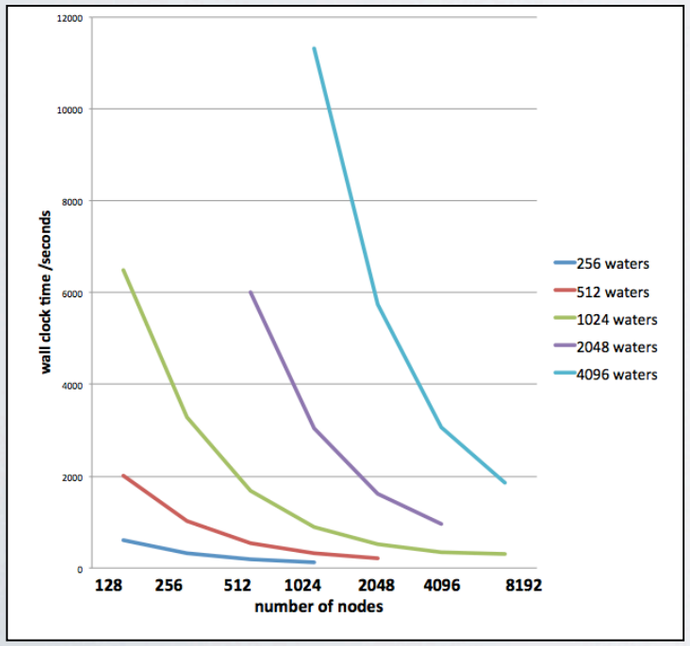


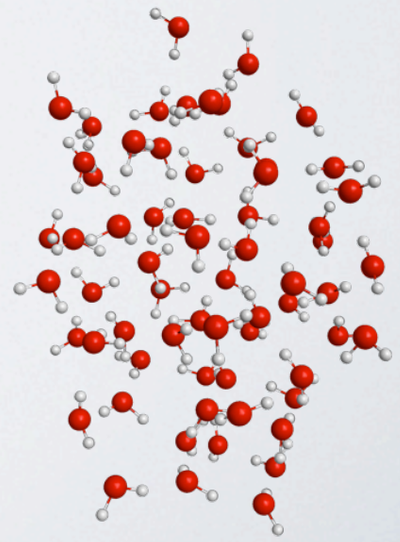












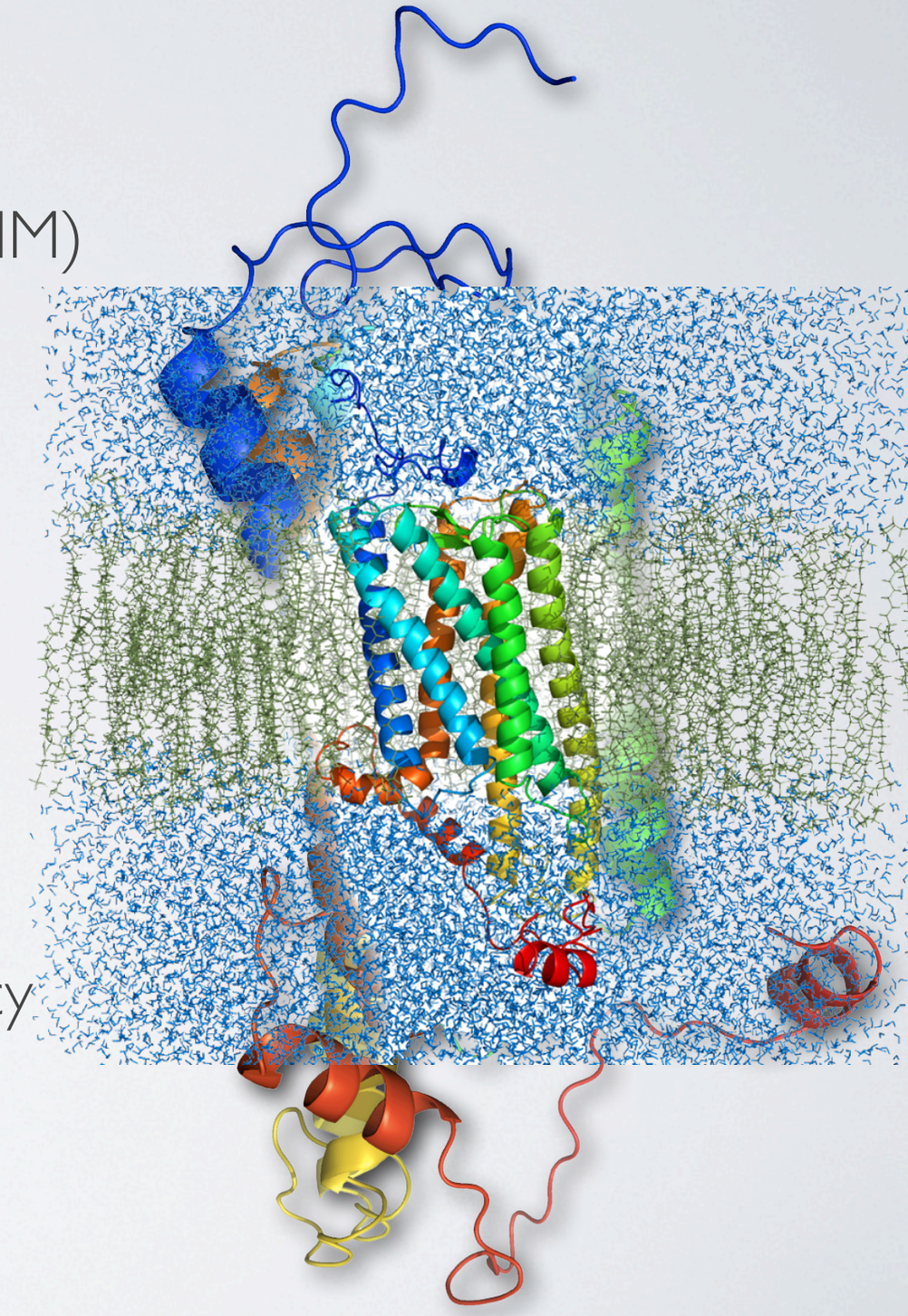
Fragmentation Methods

- Biomolecules contain hundreds or thousands of atoms, making accurate quantum calculations either very difficult or impossible

- Quantum Mechanics/Molecular Mechanics (QM/MM) methods have become popular in recent years, however,

  - As system size grows the QM region can get unwieldy

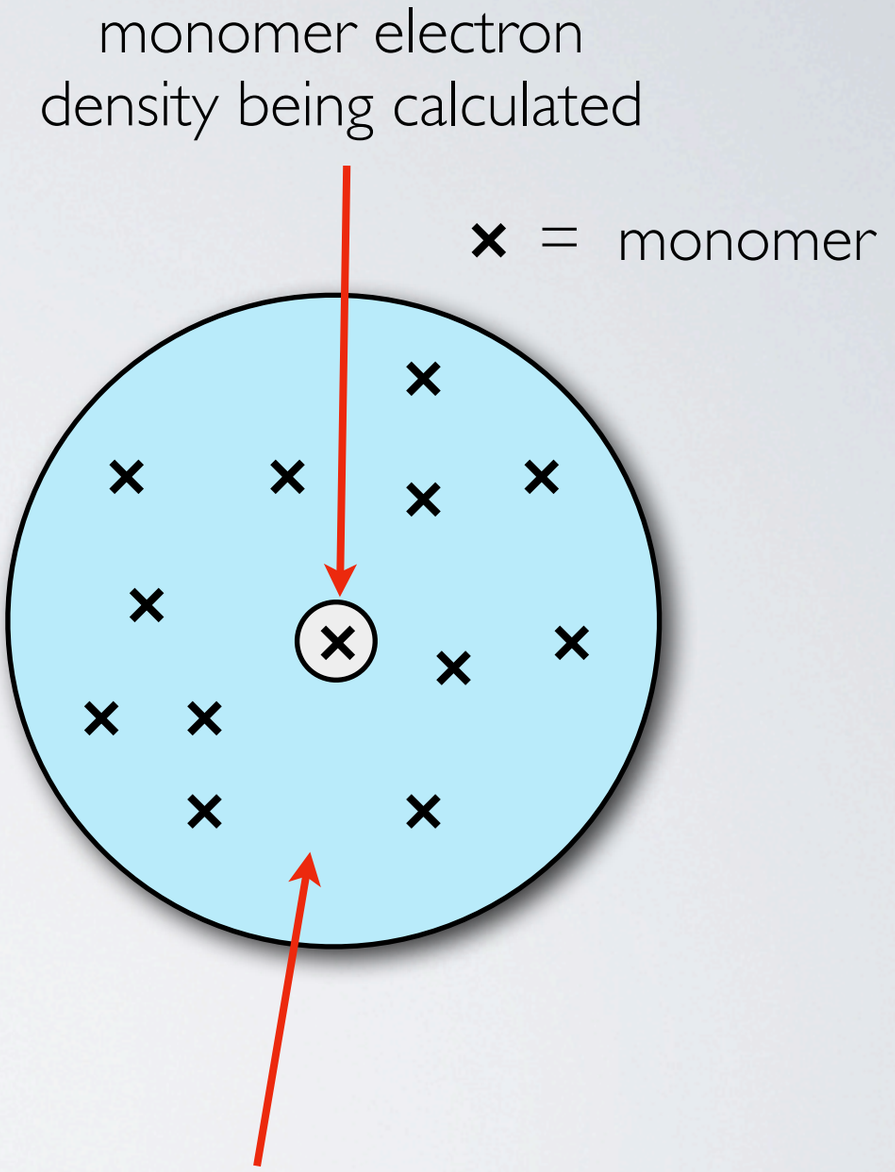
  - The energy contribution from the environment becomes too large to obtain reasonable accuracy from molecular mechanics



Fragmentation methods offer a unique solution to accurate calculations on large molecules

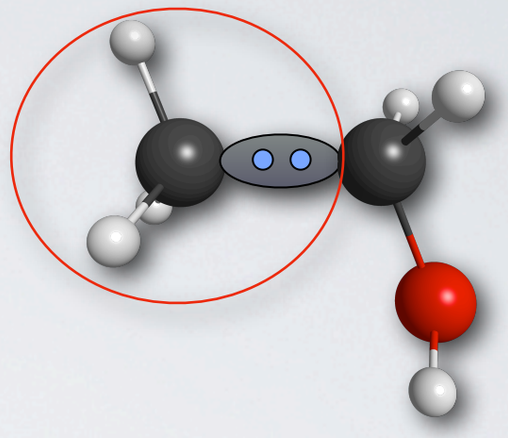
The Fragment Molecular Orbital Method

- Exchange is not long-range in most molecules
- Long-range interactions can be treated using just the Coulomb operator, thereby ignoring exchange
- Perform the molecular calculations individually in the rigorous Coulomb field of the full system
- Improved by explicit many-body corrections for fragment pairs and triples (dimers & trimers)



The Fragment Molecular Orbital Method

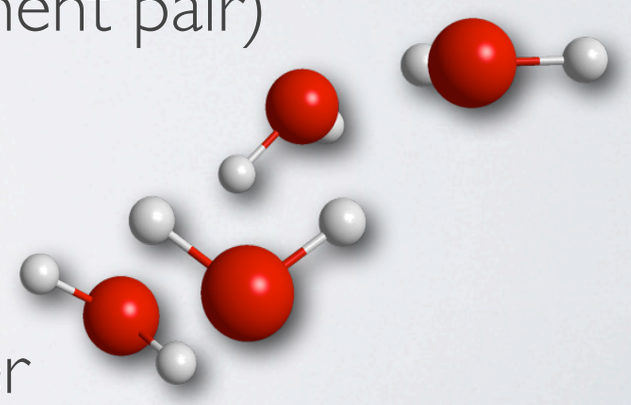
- Bonds are fractioned electrostatically



- Electrons are assigned heterolytically

- FMO fragmentation should be conducted based upon chemical knowledge (not a formal "mathematical exercise")

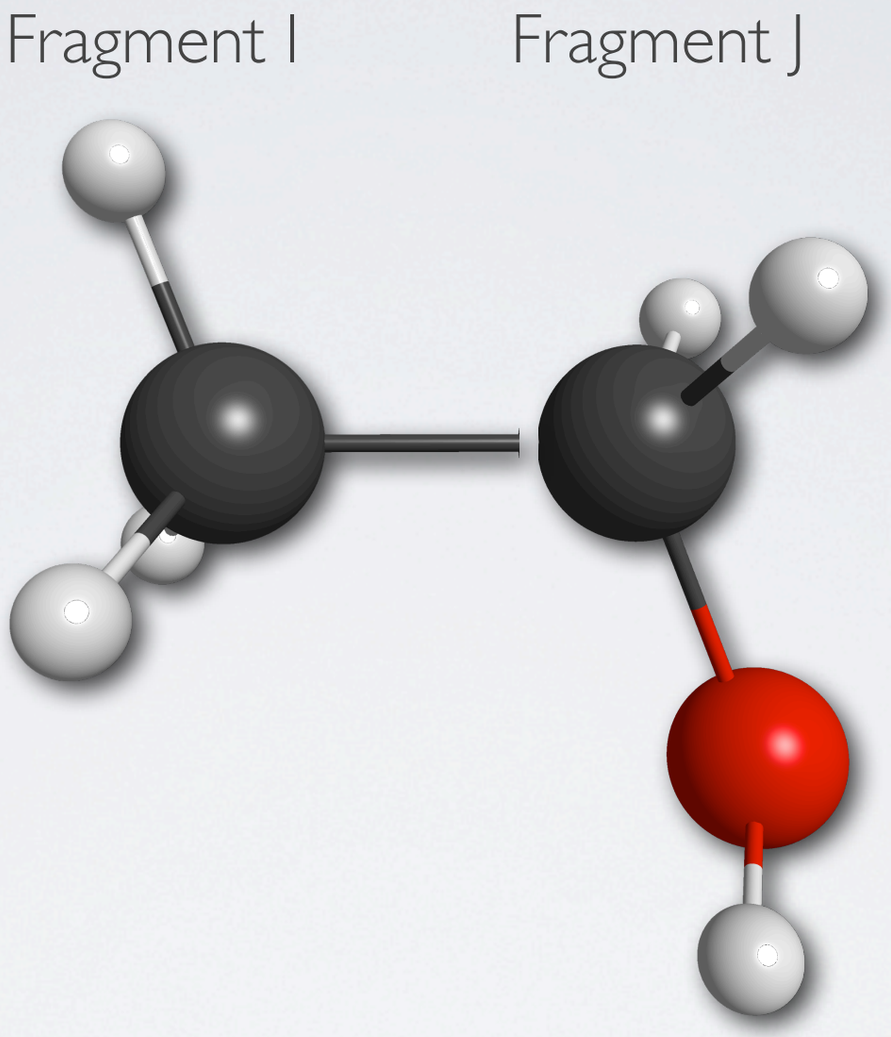
- Hydrogen bonding is accounted for by explicit dimer (fragment pair) calculations



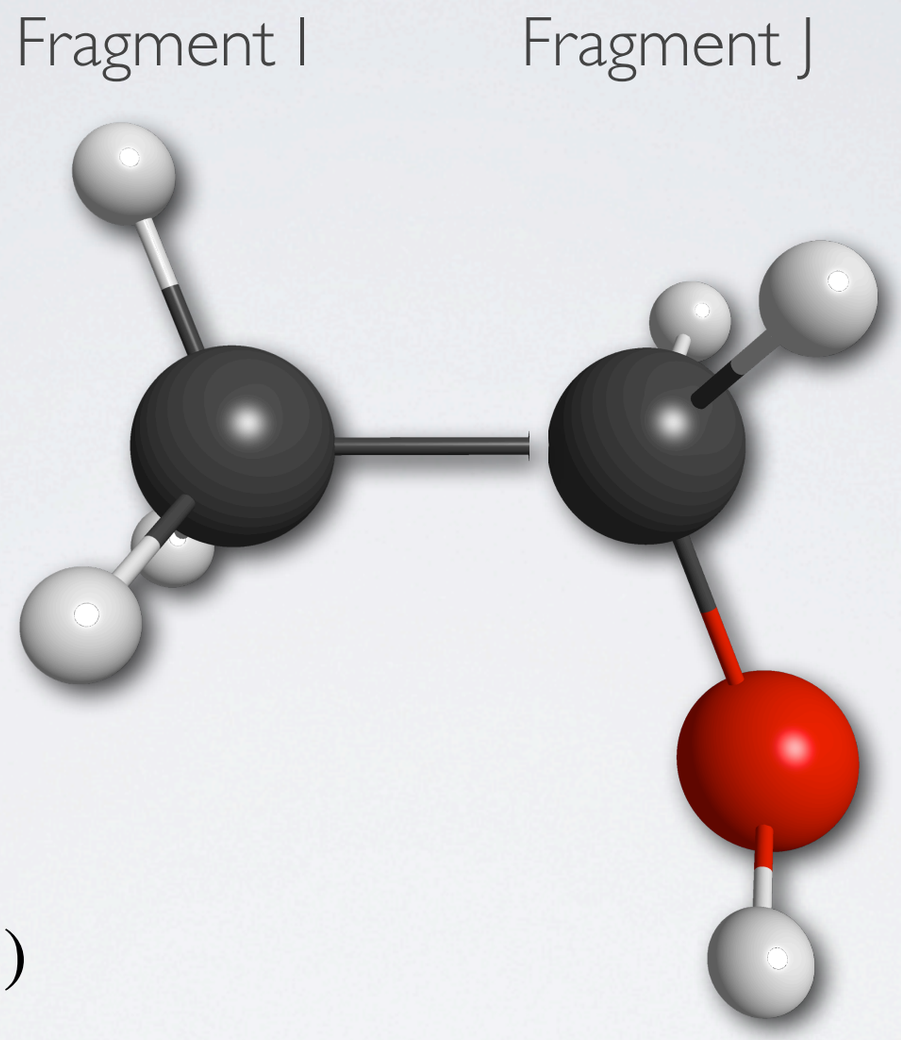
- Dimer & trimer (fragment triple) calculations allow for other quantum effects to be taken into account

The Fragment Molecular Orbital Method

$$H\Psi(r) = E\Psi(r)$$



The Fragment Molecular Orbital Method



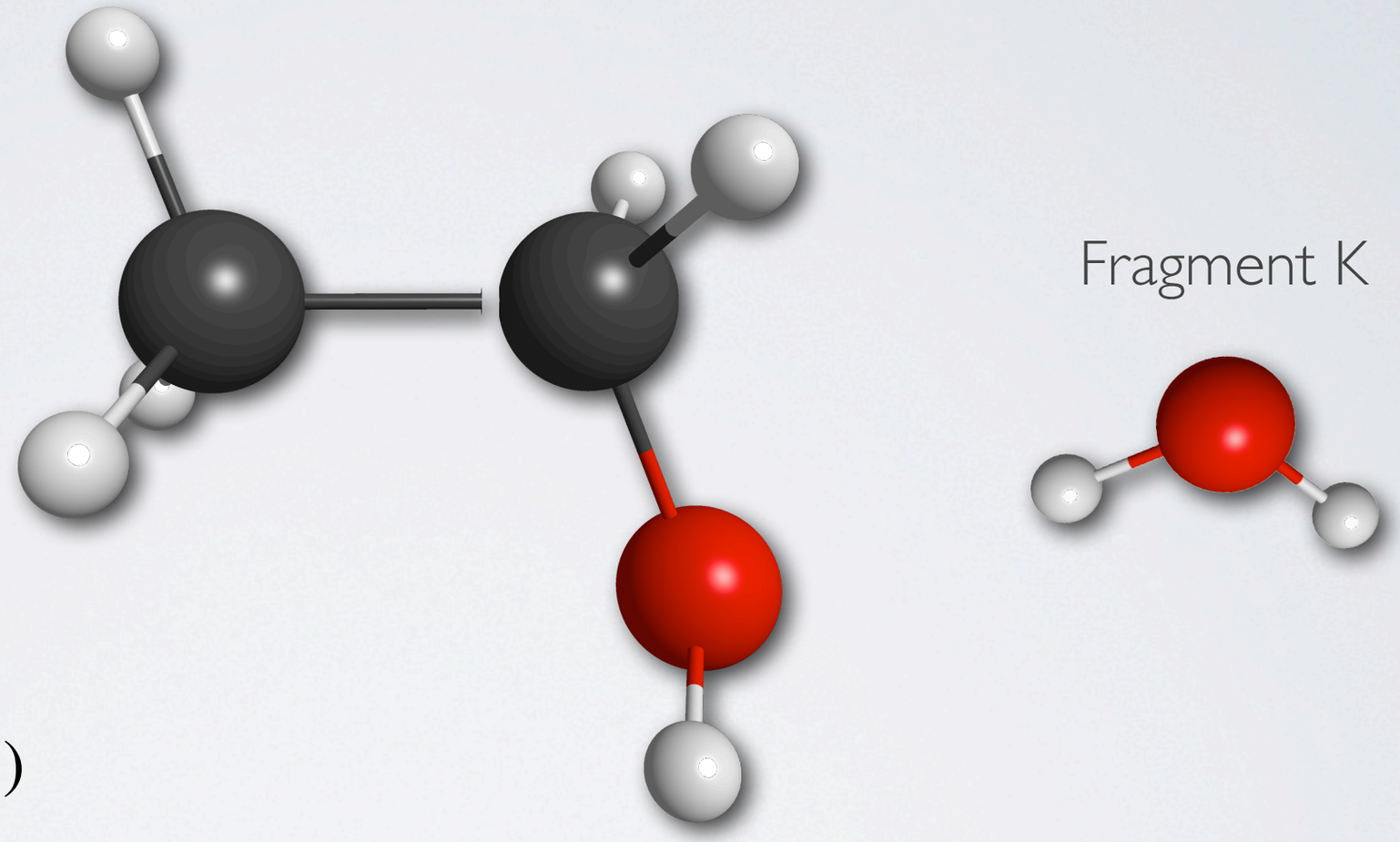
$$E = \sum_{I} E_I + \sum_{I>J} (E_{IJ} - E_I - E_J)$$

The Fragment Molecular Orbital Method

Fragment I

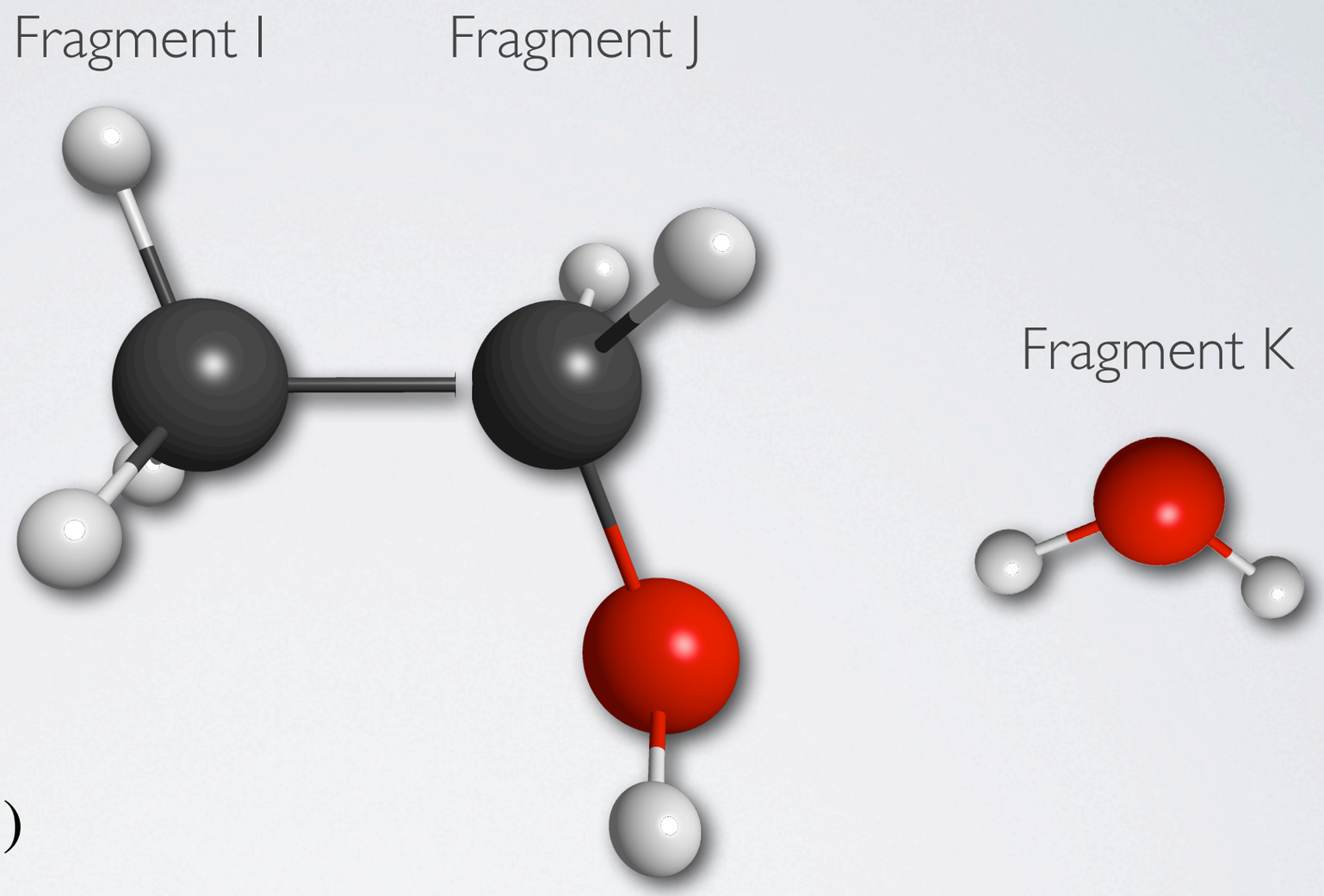
Fragment J

Fragment K



$$E = \sum_{I} E_I + \sum_{I>J} (E_{IJ} - E_I - E_J)$$

The Fragment Molecular Orbital Method



$$E = \sum_{I} E_I + \sum_{I>J} (E_{IJ} - E_I - E_J)$$

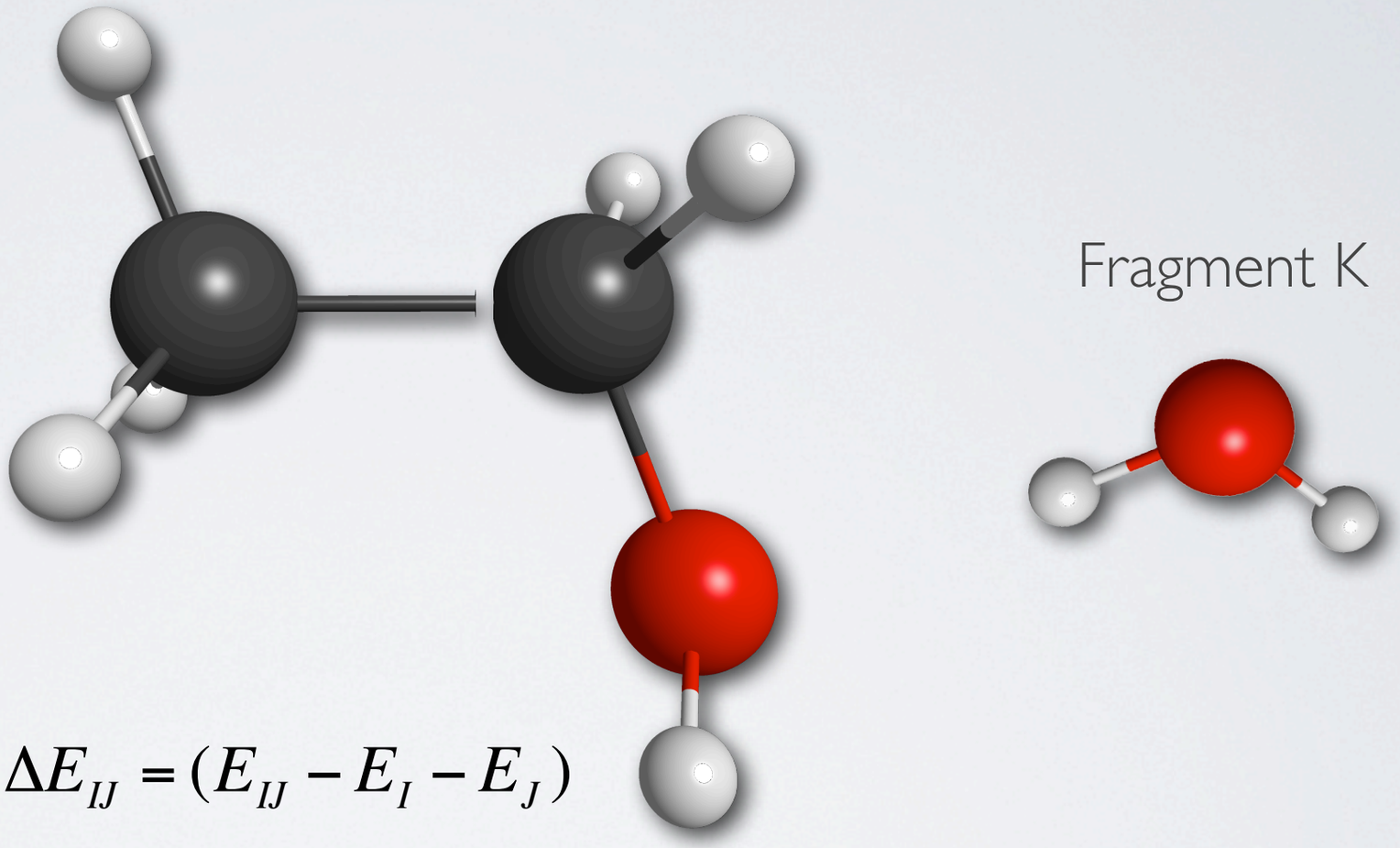
$$+ \sum_{I>J>K} \{(E_{IJK} - E_I - E_J - E_K) - (E_{IJ} - E_I - E_J) - (E_{IK} - E_I - E_K) - (E_{JK} - E_J - E_K)\}$$

The Fragment Molecular Orbital Method

Fragment I

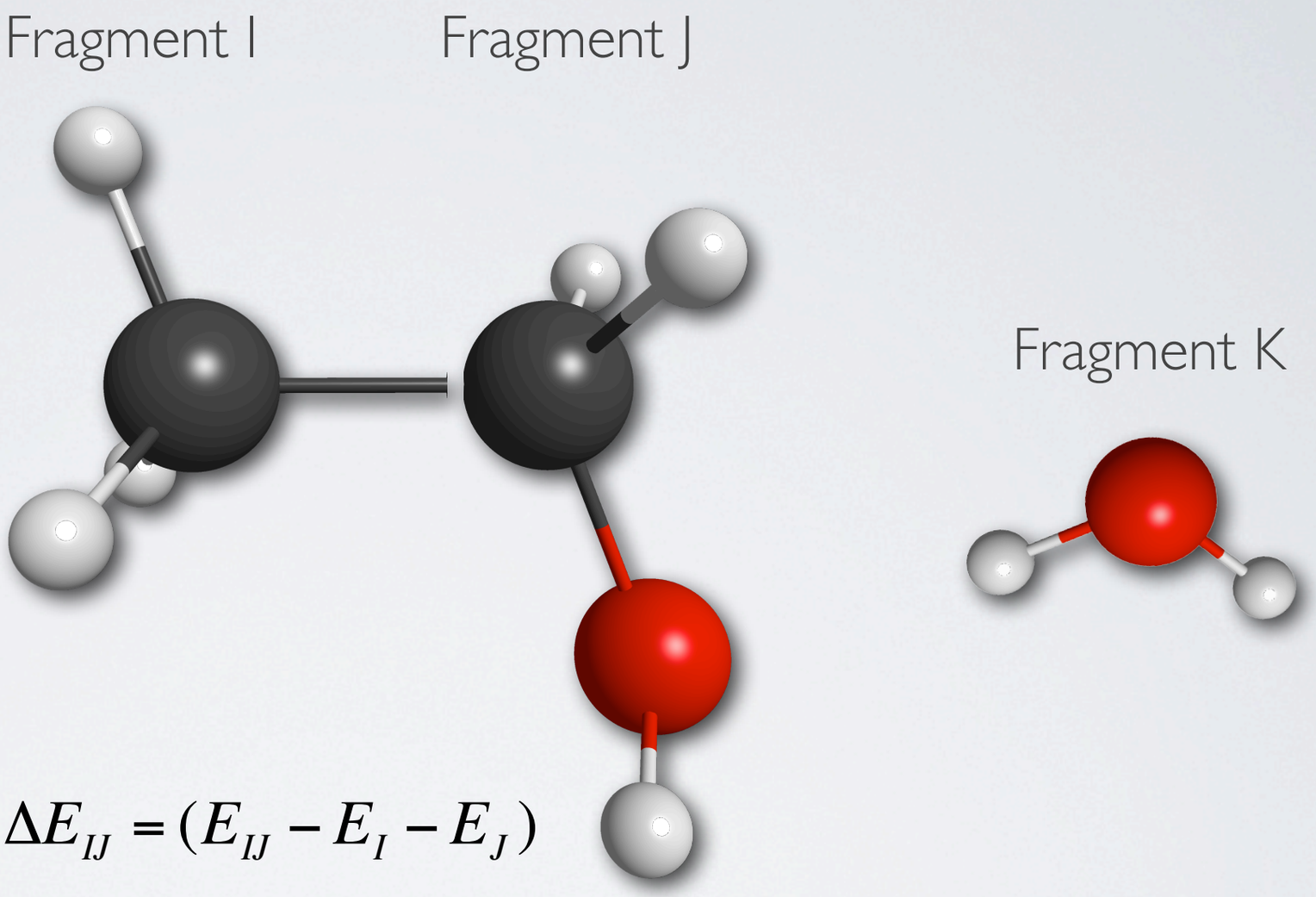
Fragment J

Fragment K



$$E^{FMO2} = \sum_{I} E_I + \sum_{I>J} \Delta E_{IJ} \qquad \Delta E_{IJ} = (E_{IJ} - E_I - E_J)$$

The Fragment Molecular Orbital Method



$$E^{FMO2} = \sum_{I} E_I + \sum_{I>J} \Delta E_{IJ} \qquad \Delta E_{IJ} = (E_{IJ} - E_I - E_J)$$

$$E^{FMO3} = E^{FMO2} + \sum_{I>J>K} \{(E_{IJK} - E_I - E_J - E_K) - (E_{IJ} - E_I - E_J) - (E_{IK} - E_I - E_K) - (E_{JK} - E_J - E_K)\}$$

The Fragment Molecular Orbital Method

1. Divide molecule into fragments and assign electrons to these fragments

2. Calculate initial electron density distribution of the fragments in the Coulomb "bath" of the full system

3. Construct the individual fragment Fock operators using the densities calculated in 2 and solve for the fragment energies

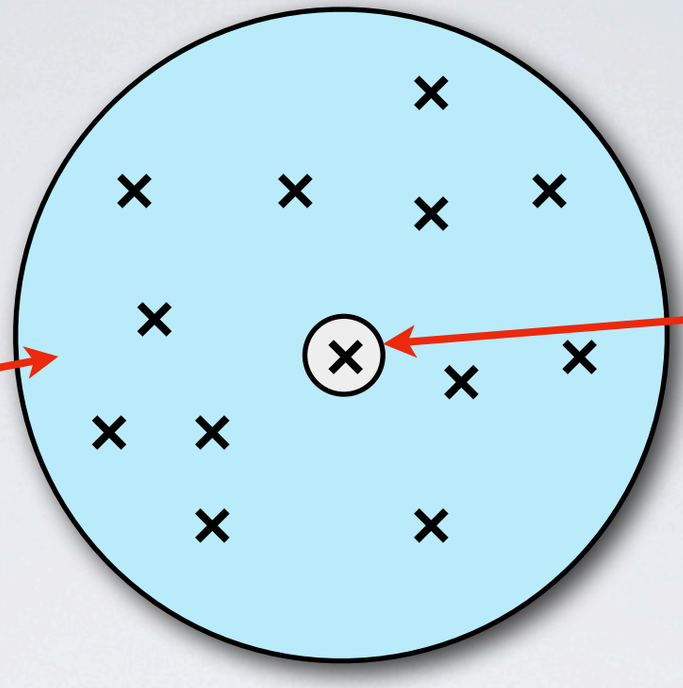
4. Determine if the density has converged for all the fragments. If not, go back to step 3

5. Construct Hamiltonians for each dimer (trimer) calculation using the converged monomer densities from steps 3-4

6. Calculate total energy and electron density

Steps 1-3:

Coulomb "bath" of the full system being taken into account



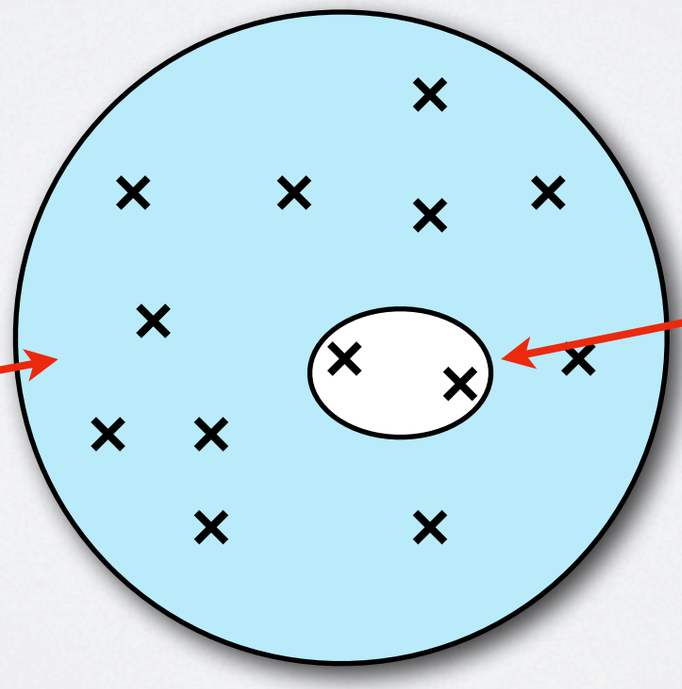
monomer electron density being calculated

× = monomer

Step 4: Steps 1 through 3 are looped until the density of the full system converges to some predetermined threshold

Steps 5-6:

converged Coulomb "bath" of the full system from monomer SCF



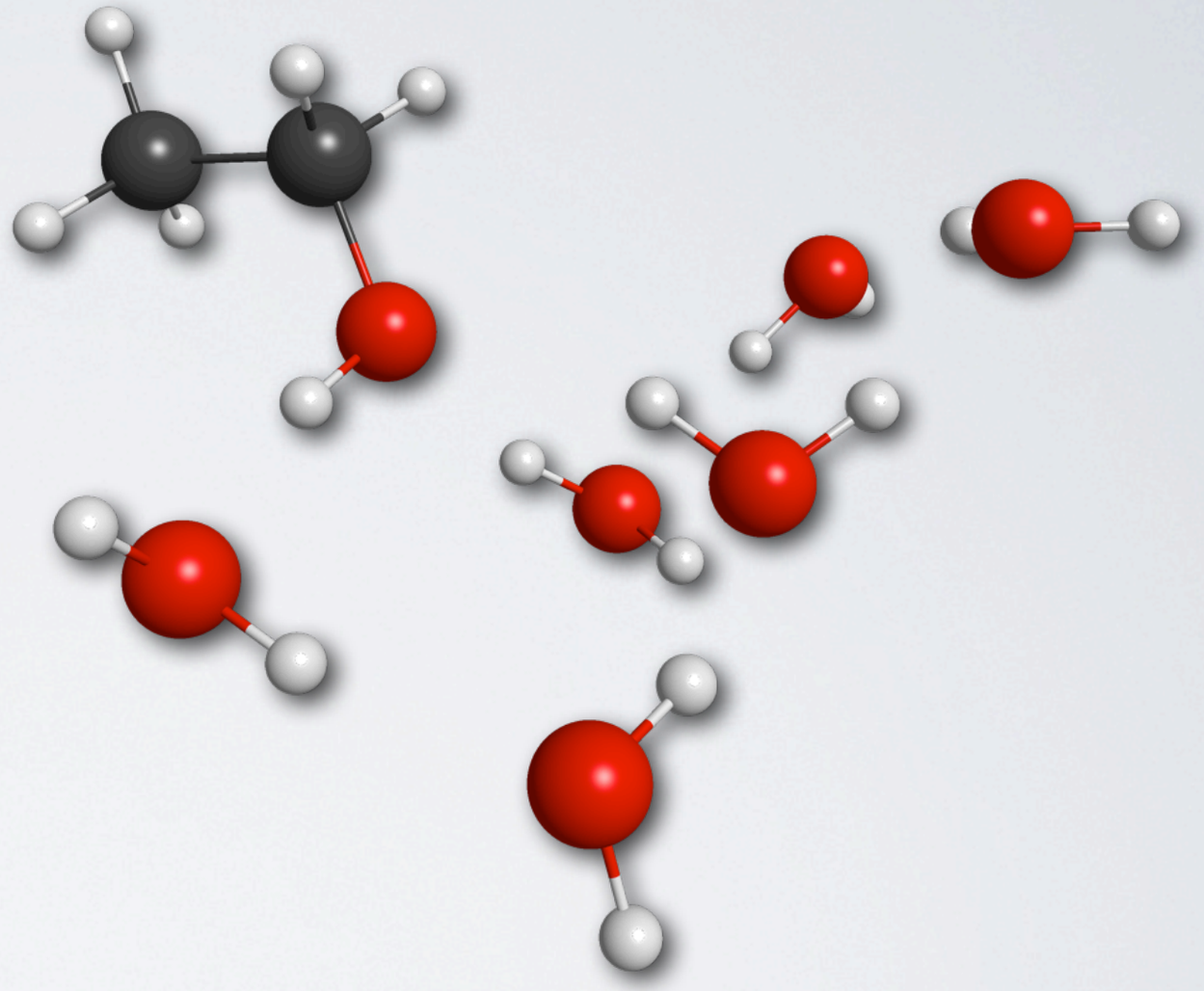
dimer electron density being calculated in the presence of the converged ESP

each dimer calculation is performed once

The Fragment Molecular Orbital Method

The number of dimer calculations increases as "*n* choose 2" where *n* is the number of fragments.

$$\#\ \text{of dimers} = \frac{n!}{2!(n-2)!}$$



7 fragments = 21 dimers

8 fragments = 28 dimers

16 fragments = 120 dimers

32 fragments = 496 dimers

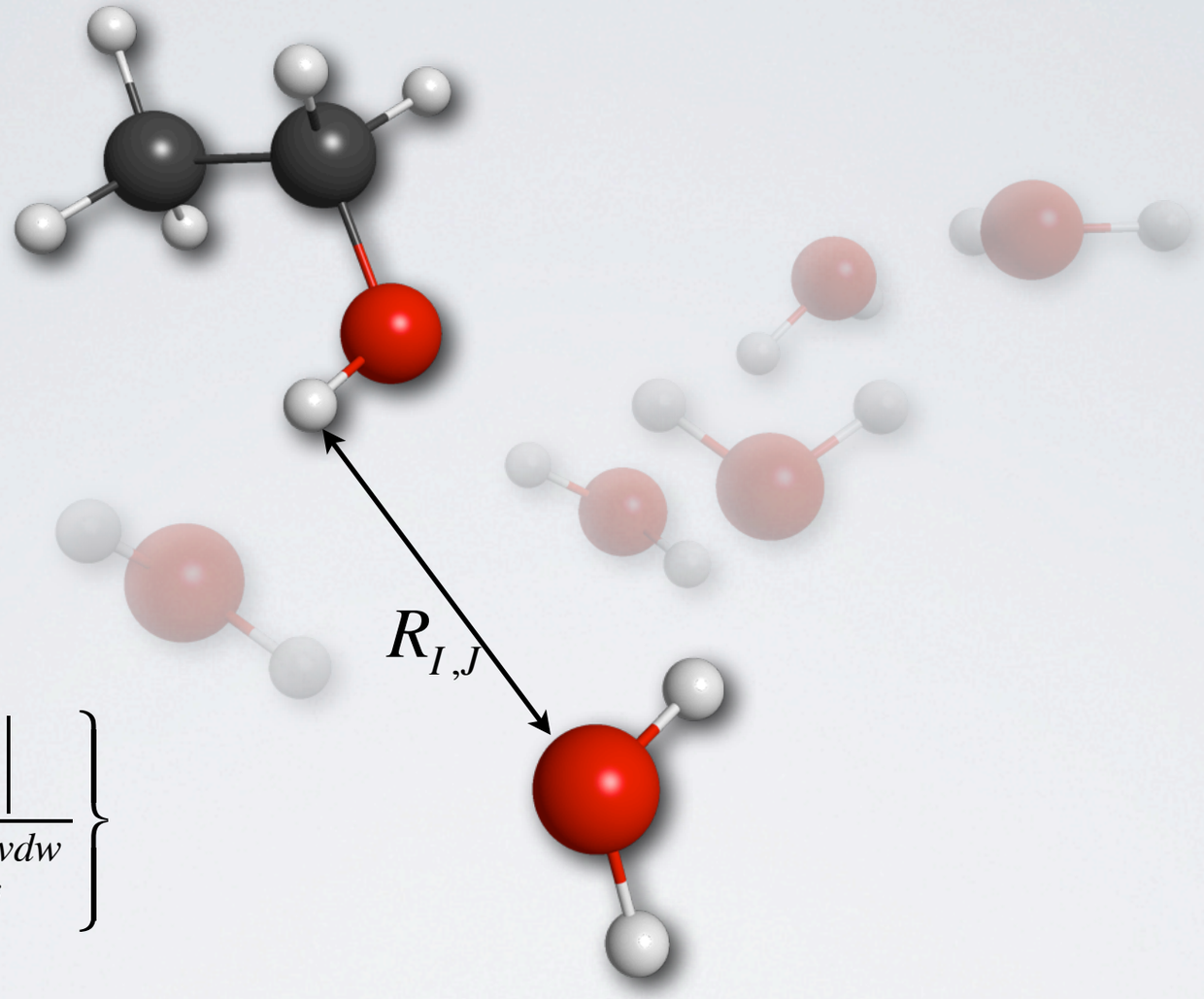
64 fragments = 2016 dimers

128 fragments = 8128 dimers

The total number of dimer calculations increases rapidly!

Two solutions: Approximations and parallelization

The Fragment Molecular Orbital Method: Approximations

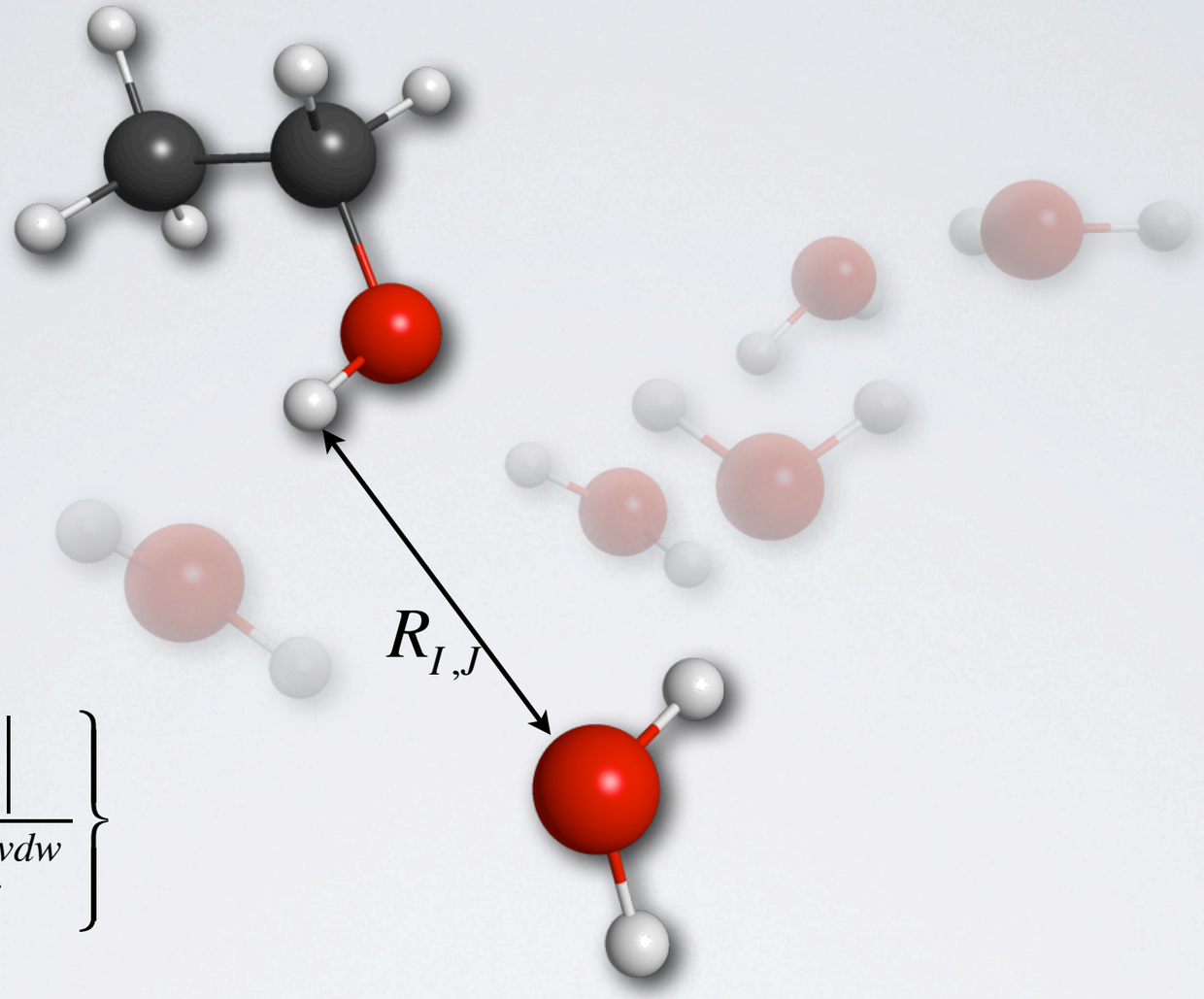


$$R_{I,J} = \min_{i \in I, j \in J} \left\{ \frac{\left| \vec{r}_i - \vec{r}_j \right|}{r_i^{vdw} + r_j^{vdw}} \right\}$$

User defined cut-off value R_{cut}

$$E^{FMO2} = \sum_I E_I + \sum_{I>J} \Delta E_{IJ}$$

The Fragment Molecular Orbital Method: Approximations

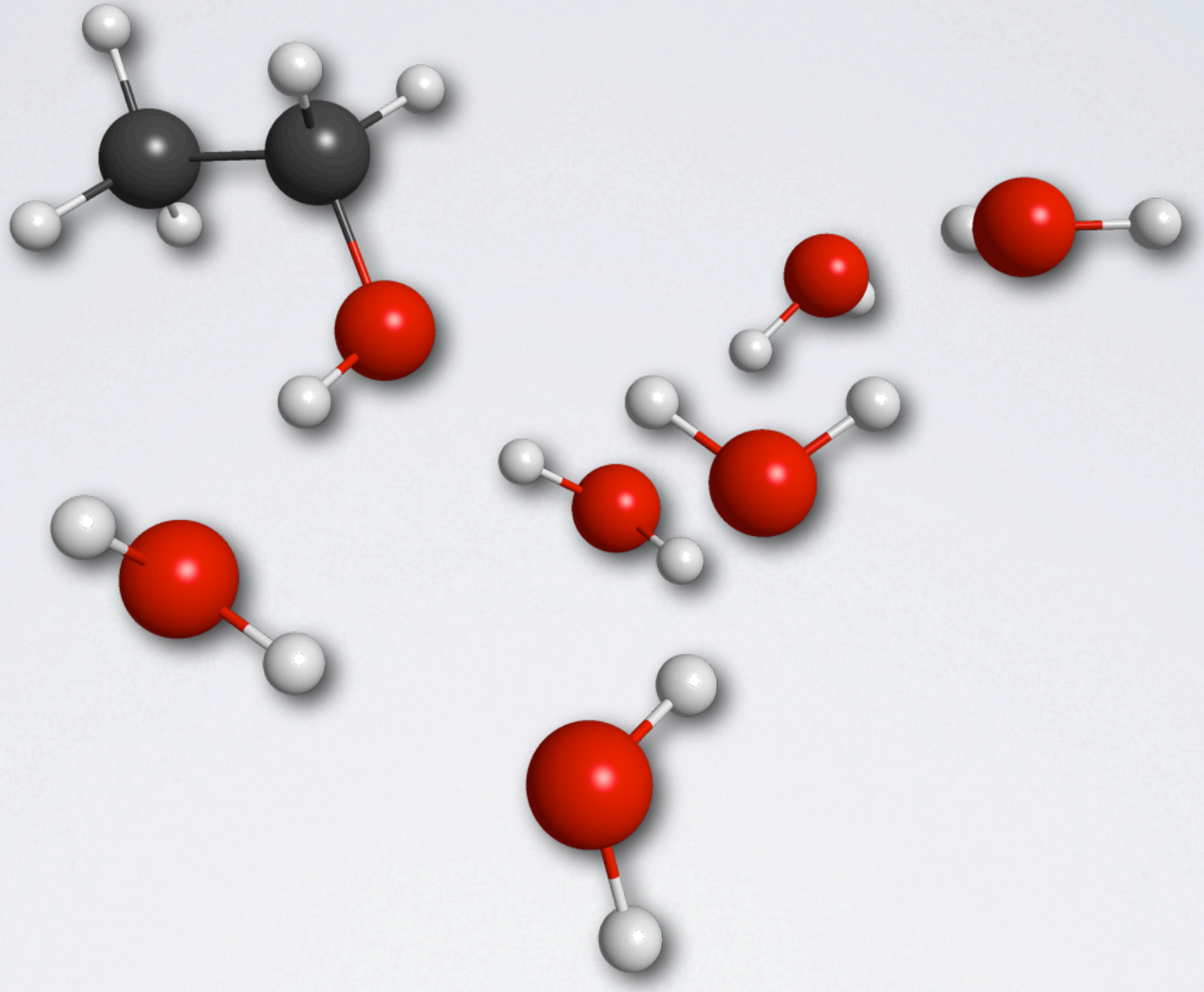


$$R_{I,J} = \min_{i \in I, j \in J} \left\{ \frac{\left| \vec{r}_i - \vec{r}_j \right|}{r_i^{vdw} + r_j^{vdw}} \right\}$$

User defined cut-off value R_{cut}

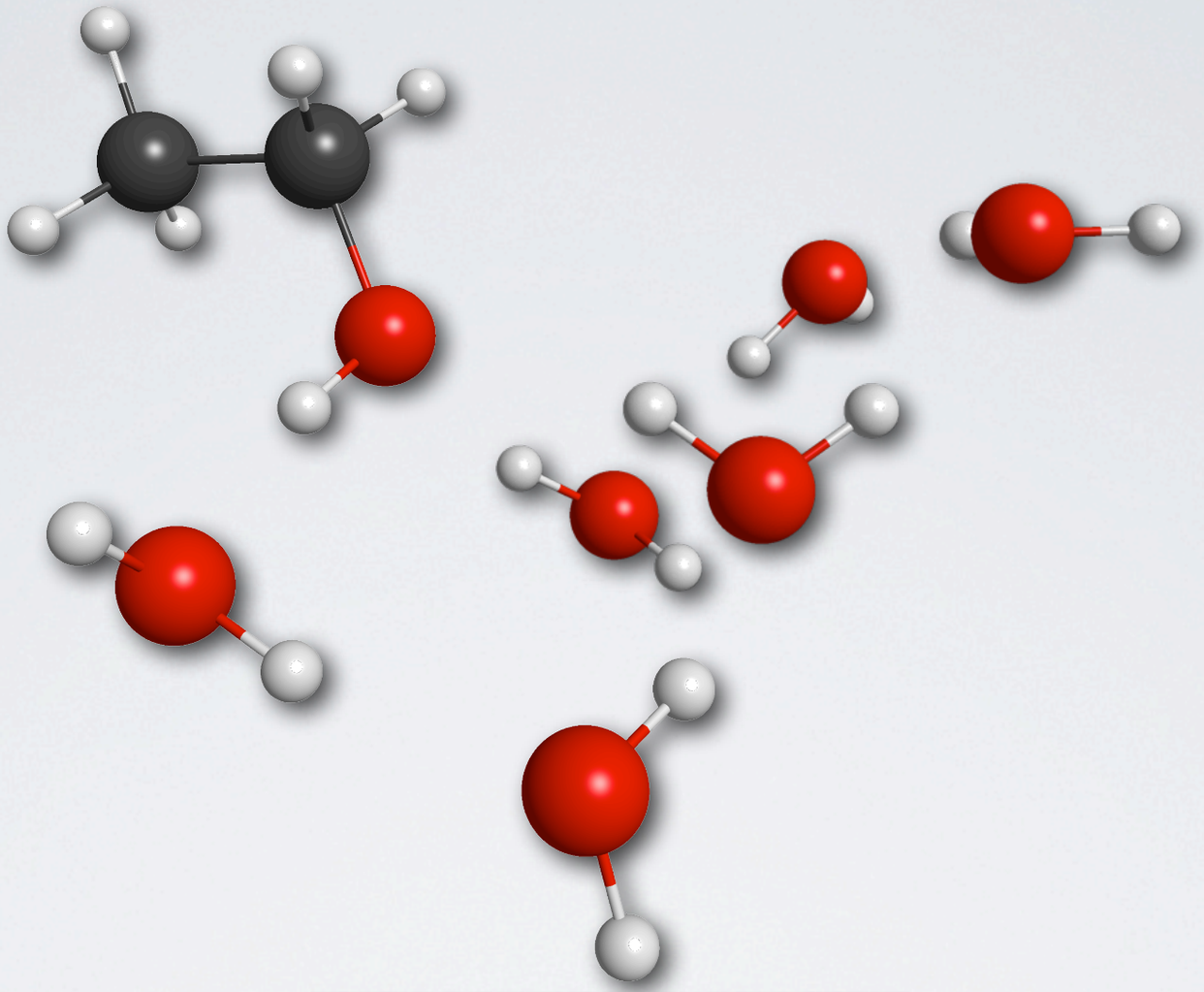
$$E^{FMO2} = \sum_{I} E_I + \sum_{I>J}^{R_{I,J} \le R_{cut}} \Delta E_{IJ} + \sum_{I>J}^{R_{I,J} > R_{cut}} \Delta E_{IJ}^{sep}$$

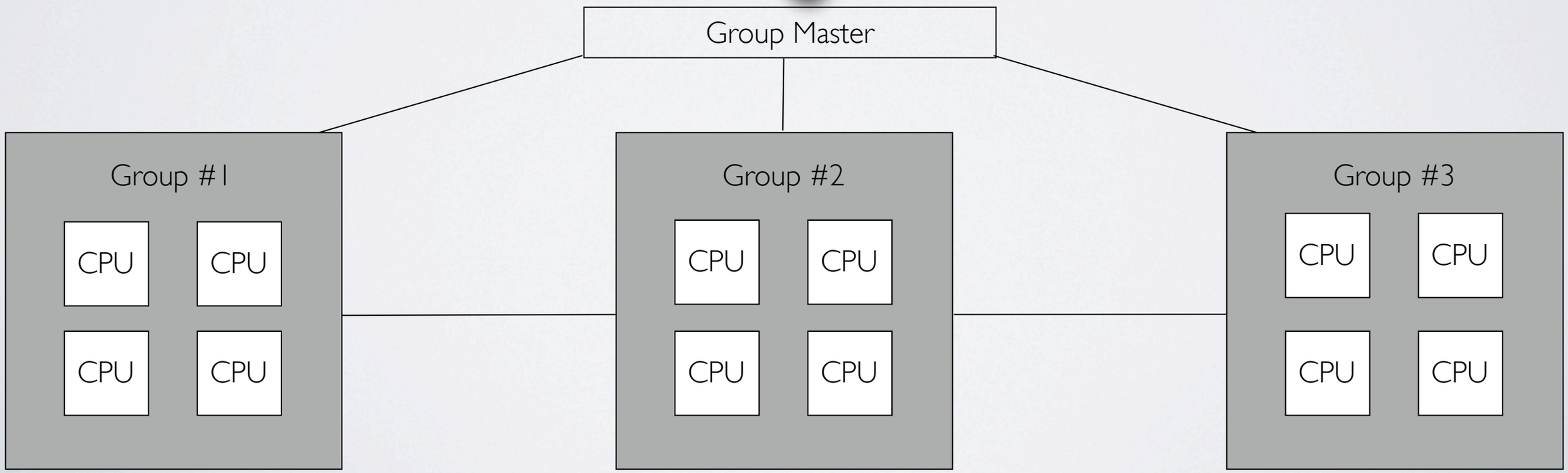
The Fragment Molecular Orbital Method: Parallelization



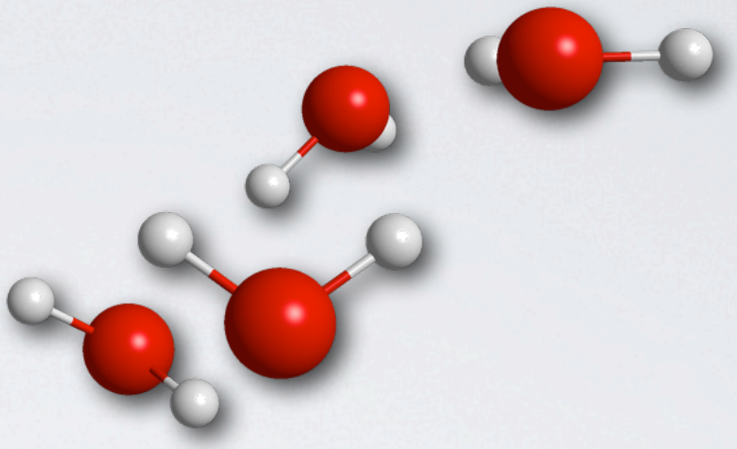
- The Generalized Distributed Data Interface (GDDI)
- GDDI allows for massively parallel calculations on clusters of computers or supercomputers
  - After the molecule is divided into fragments, each fragment is sent to a group which is composed of more than one processor or SMP enclosure
  - Each fragment is then run in parallel in each group
  - This provides two levels of parallelization, greatly speeding up the calculation

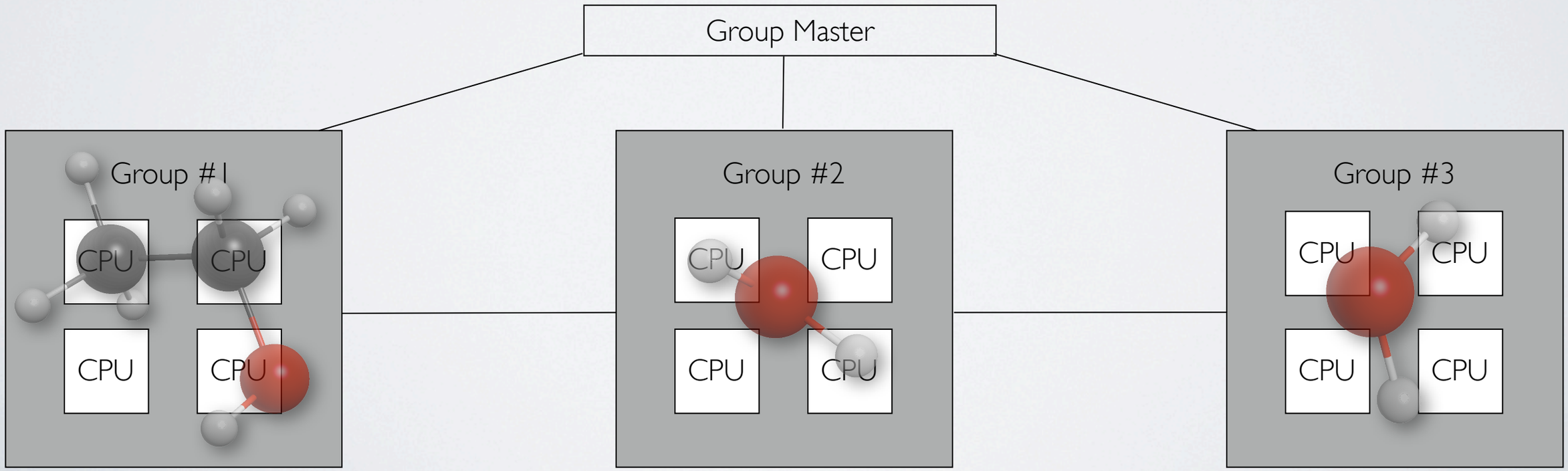
The Fragment Molecular Orbital Method: Parallelization





The Fragment Molecular Orbital Method: Parallelization



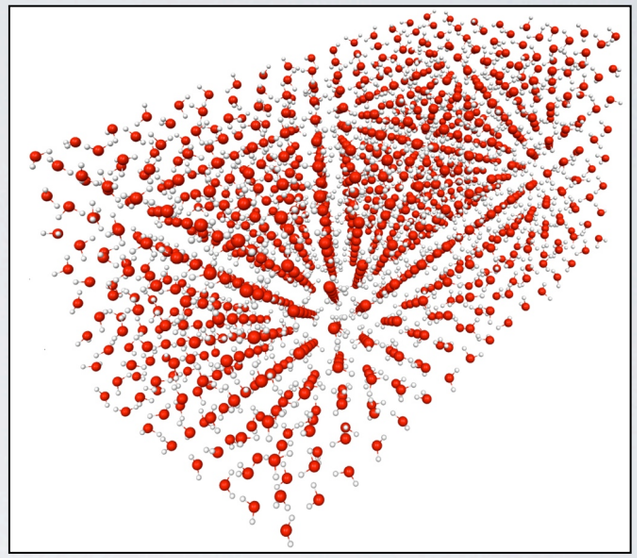


Benchmarks and Algorithmic Alterations

In previous FMO implementations, the master process of each group created a direct-access file in which the densities of all fragments are stored.

The new approach is based on a large array containing all fragment densities created in shared memory distributed among nodes.

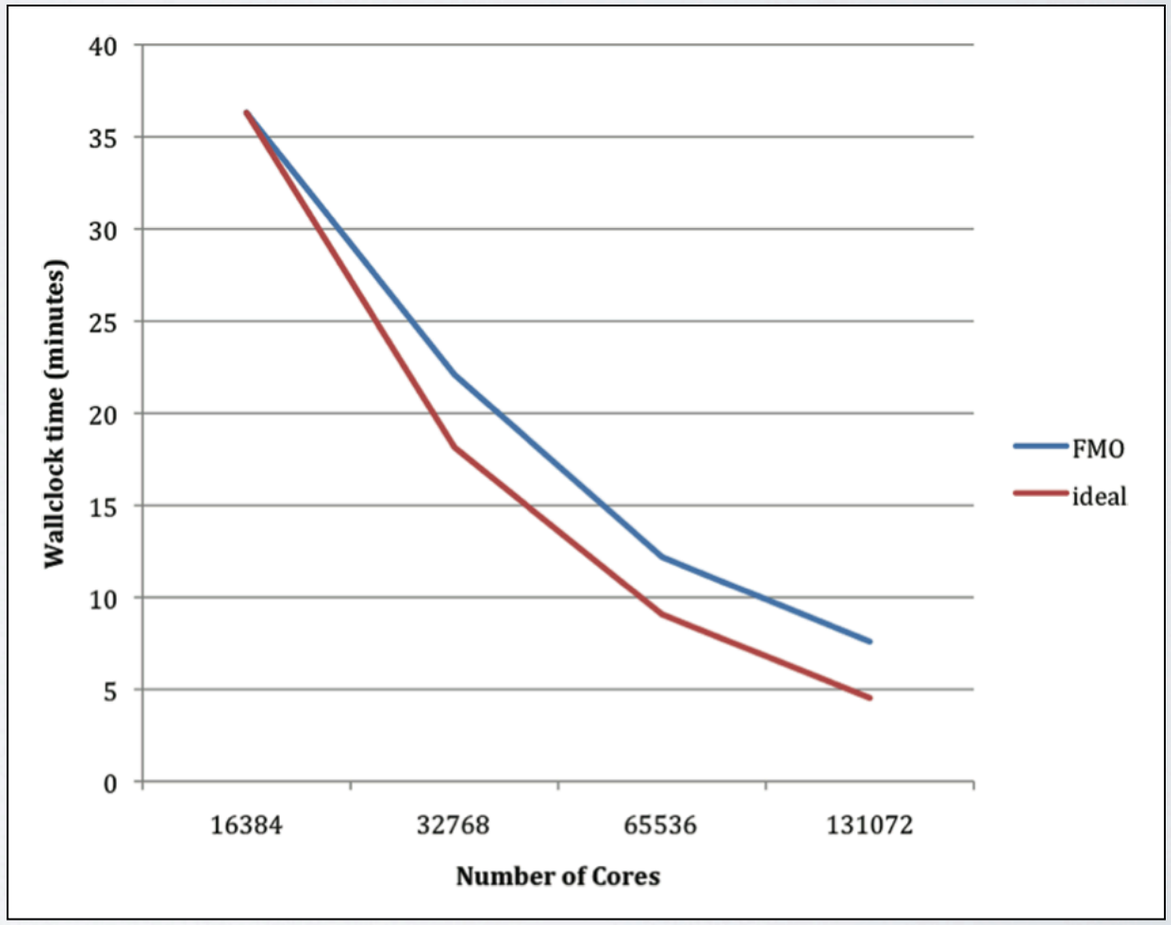
Fragment densities are stored on data servers and sent on demand to compute nodes directly, with these communications sometimes involving intergroup operations.



The energy of a cluster of 1024 water molecules was calculated using FMO2 with MP2 and the 6-31G(d,p) basis set, both with the previous disk-based implementation ("FMOd") and the new implementation ("FMOm").

Each calculation was run on 1024 nodes (4096 cores) on the BG/P. The wall time required for the FMOd calculation was 335.4 min, whereas the corresponding FMOm wall time was 10.7 min.

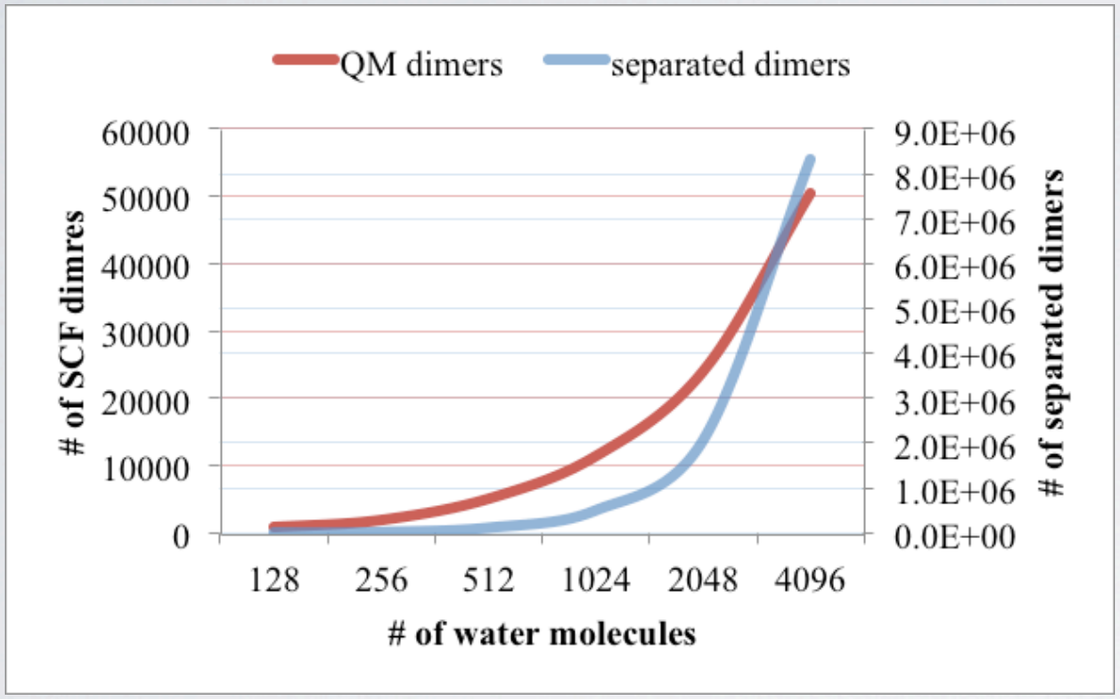
The 31-fold speed-up demonstrates that the DDI-based density storage is paramount to running FMO calculations effectively on large-scale parallel computers.



FMO2-MP2/6-31G(d) forces calculation of 12288 atoms on BG/P

Fletcher, G.D.; Fedorov, D.G.; Pruitt, S.R.; Windus, T.L.; Gordon, M.S. J. Chem. Theor. Comp., 2012, *8*, 75

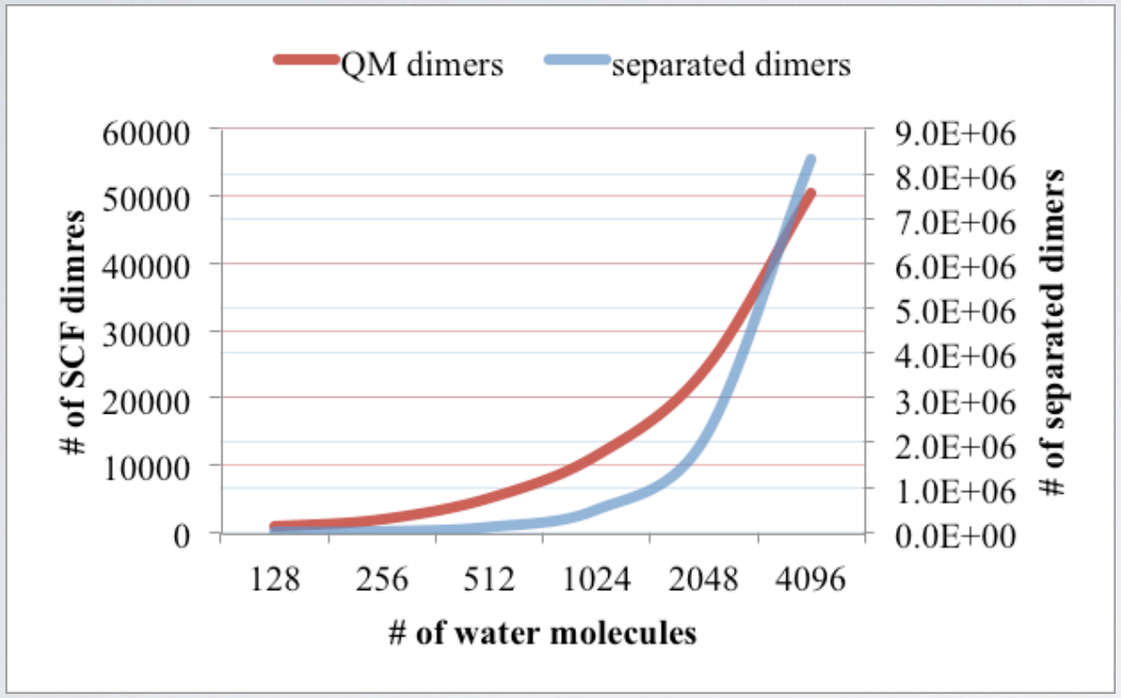
Benchmarks and Algorithmic Alterations



The total number of dimer calculations required increases dramatically with system size. Specifically, the number of separated dimer calculations for the benchmark calculations increases by a factor of ~4 when the system size is increased from 2048 to 4096 water molecules.

While the separated dimer calculations take a fraction of the time compared to QM dimer calculations, the sheer number of these calculations requires more computational effort than the QM calculations after a certain system size is reached.

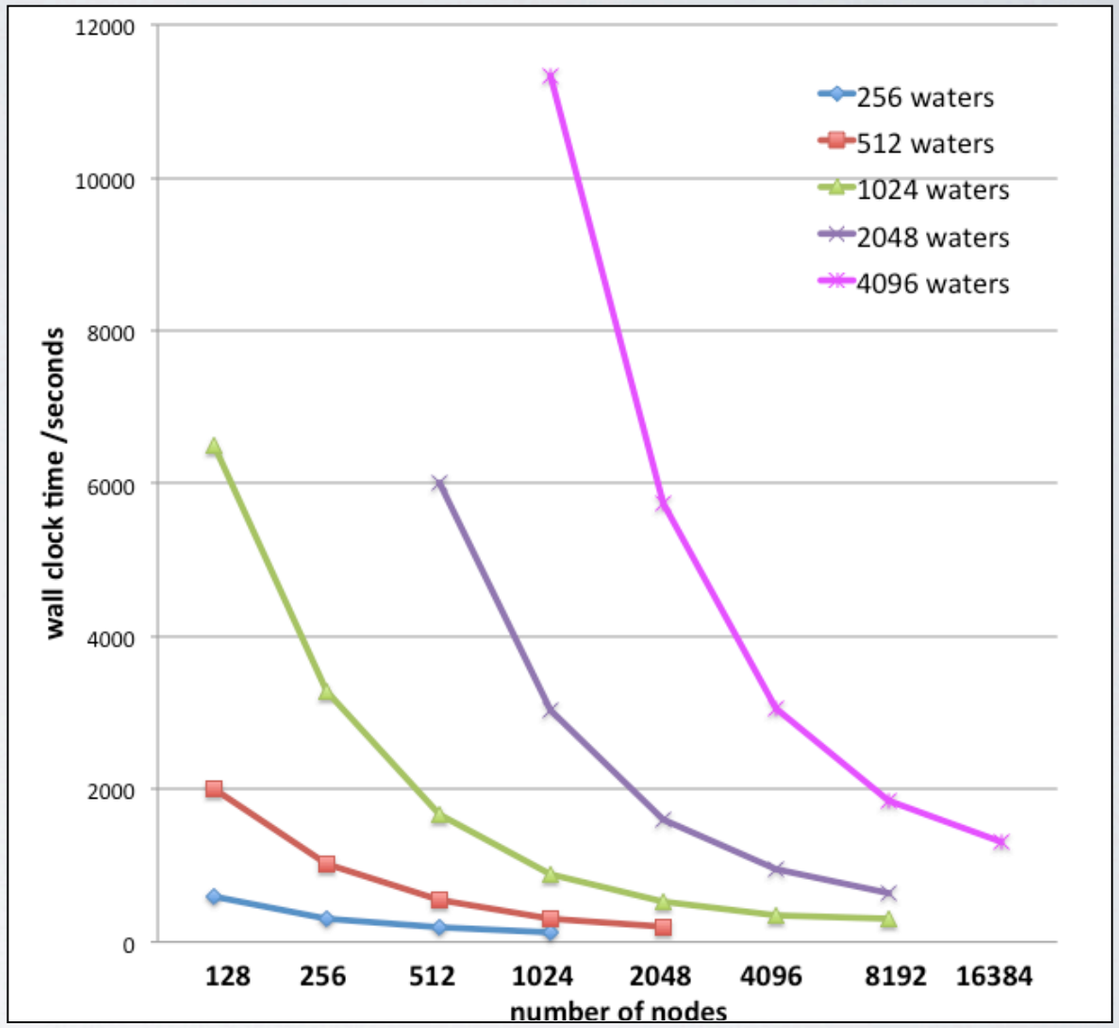
BG/Q Benchmarks



The best timing on 1024 water molecules from BG/P was ~7-8 minutes on 131,072 cores using the 6-31G(d) basis set.

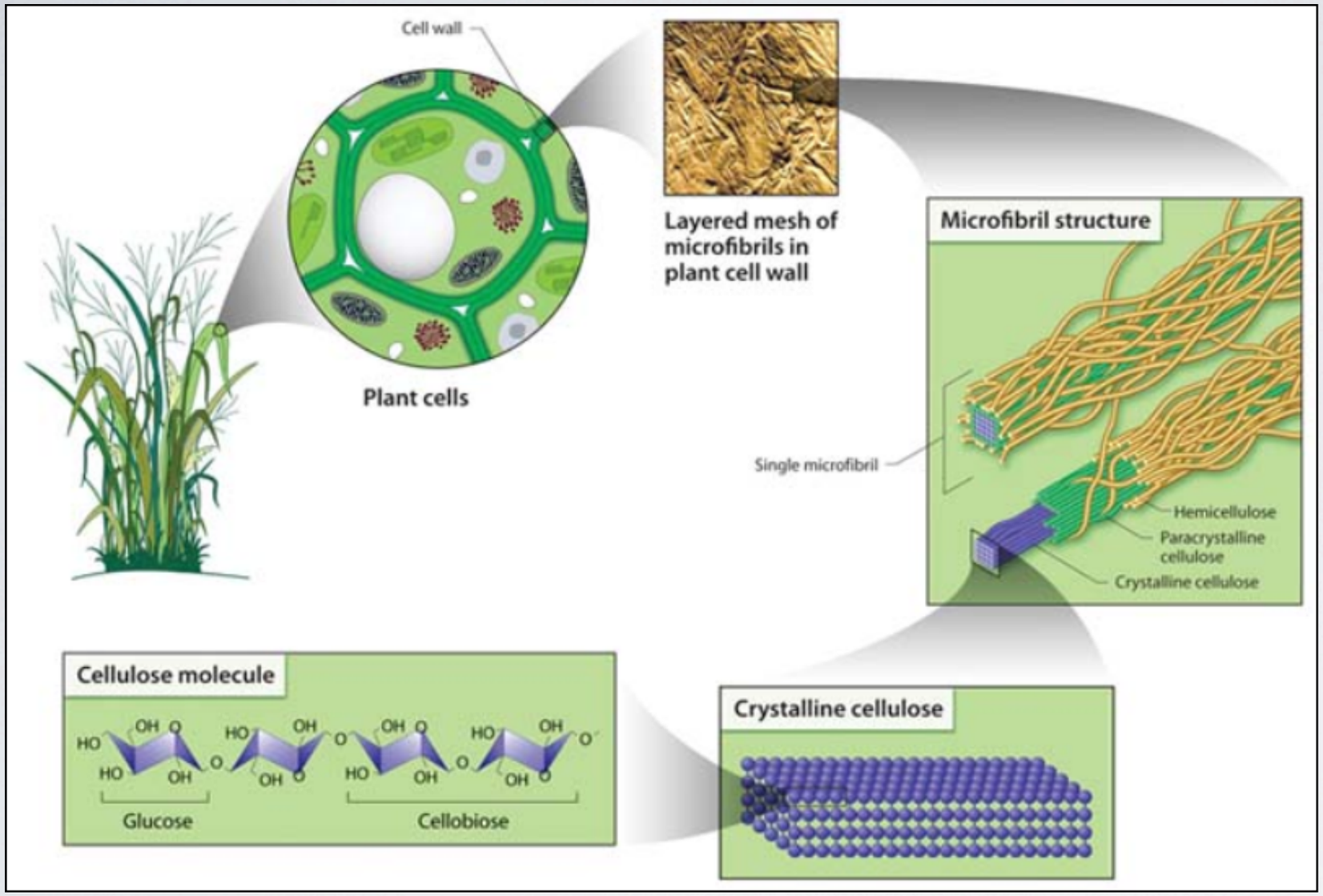
BG/Q provides faster timings on 1024 nodes (16,384 cores) using the **aug-cc-pVDZ** basis set.

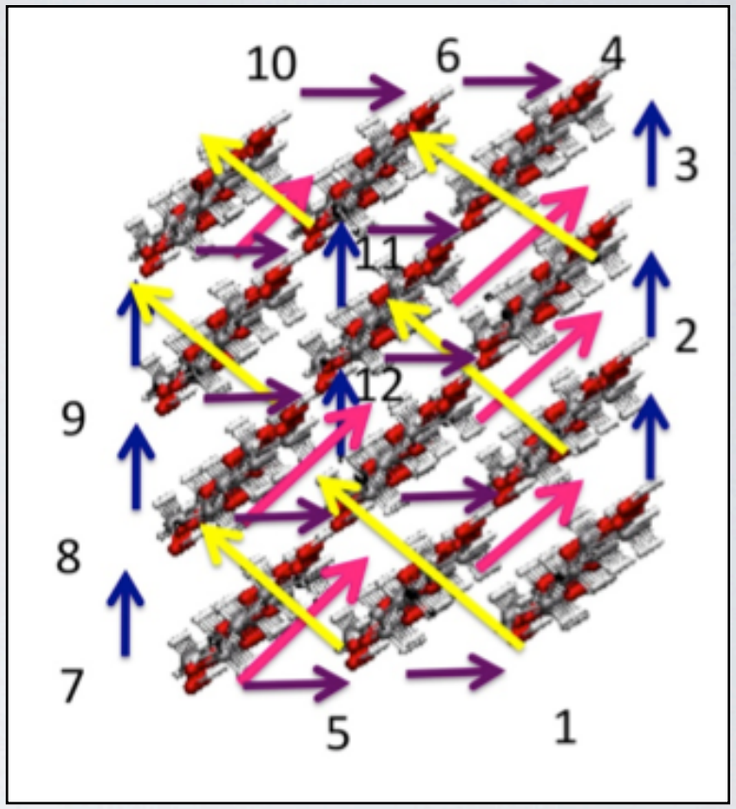
Previous BG/P benchmarks also did not use the fully analytic gradient (increases computational cost by ~20%)

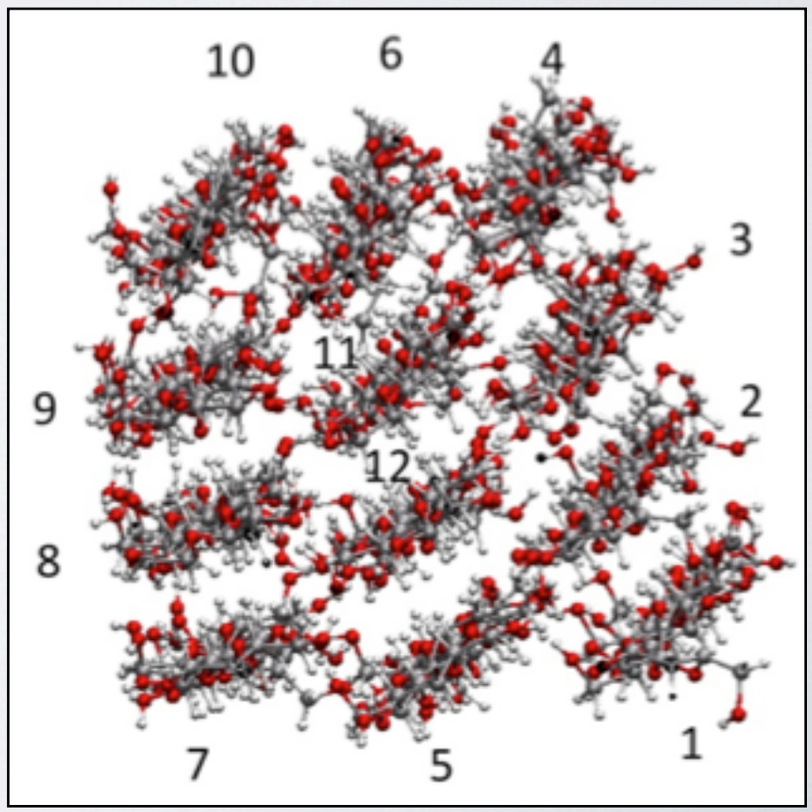


FMO2-MP2/aug-cc-pVDZ gradient calculation on BG/Q

Biomass Recalcitrance: Crystalline Cellulose I-alpha Microfiber Chain





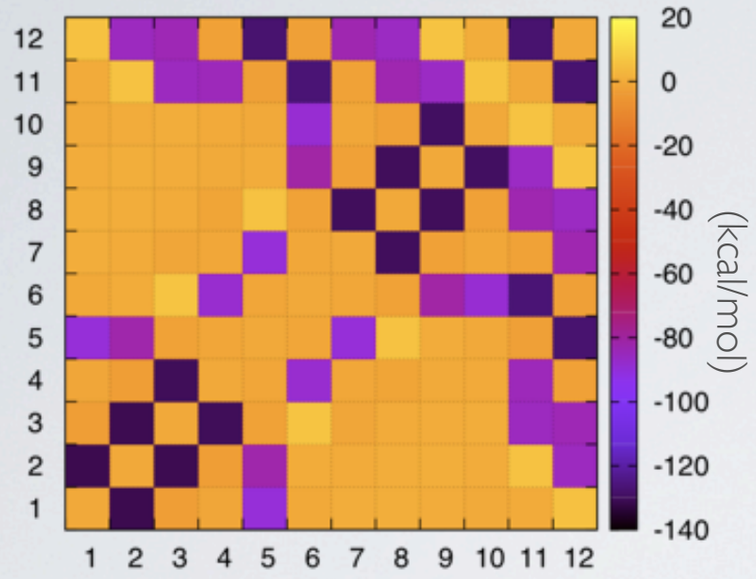


FMO2/MP2 Energy Calculations

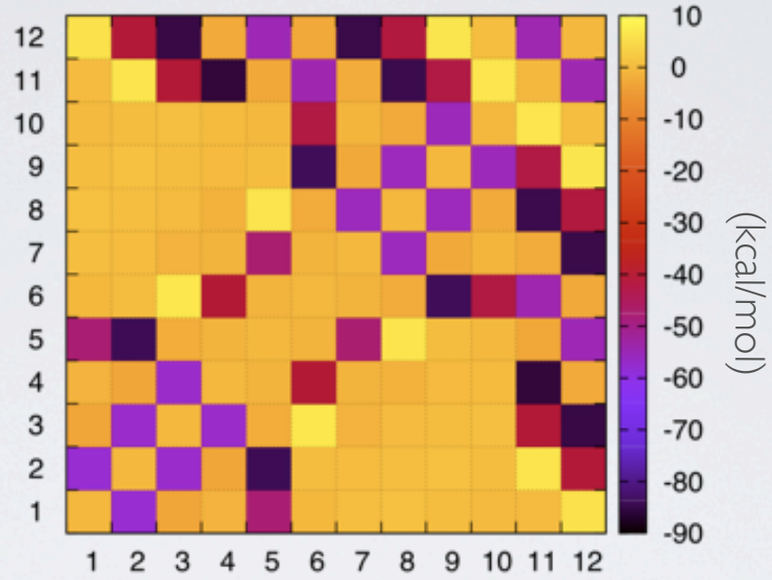
Pair Interaction Energy Decomposition Analysis (PIEDA) peformed to gain insight into what types of molecular interactions are dominant

Crystalline Cellulose I-alpha Microfiber Chain

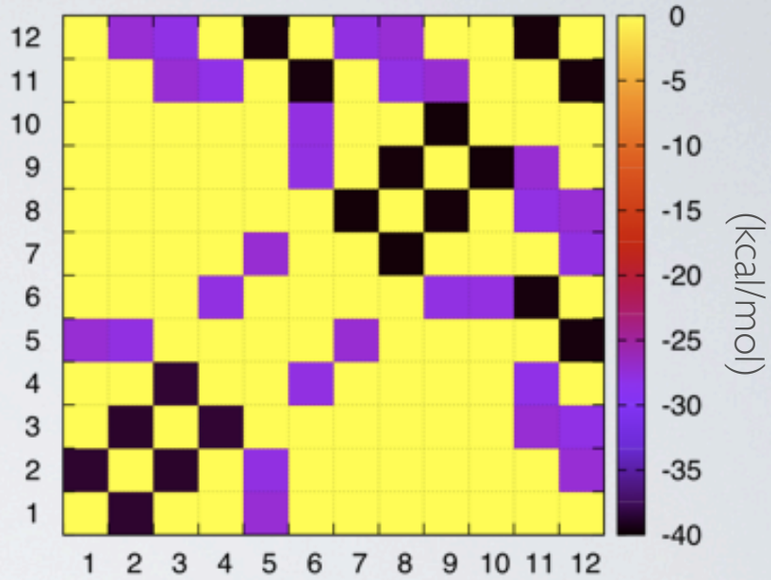
Pair Interaction Energy Decomposition Analysis



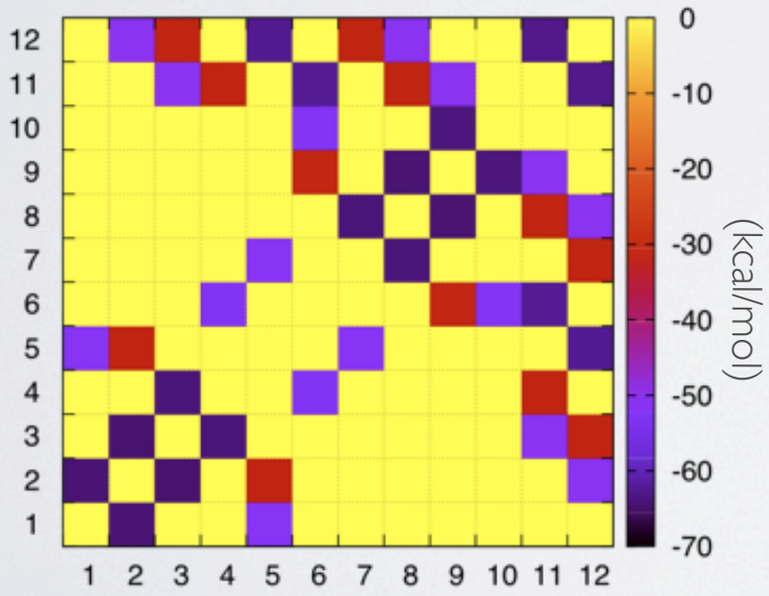
Total Energy



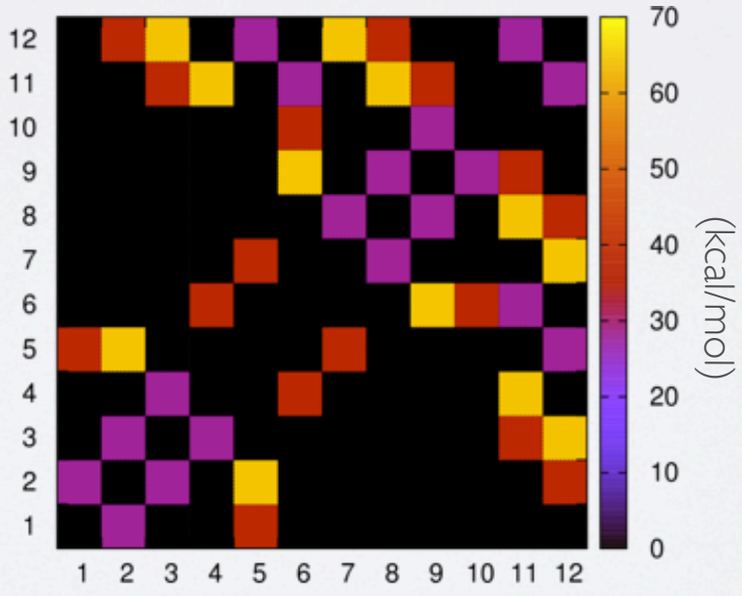
Electrostatic



Charge Transfer



Dispersion



Exchange Repulsion

X & Y axis = Chain Number

Total interaction shows stronger intersheet binding

Intrasheet interactions are dominated by electrostatics

Intersheet interactions are dominated by dispersion and charge transfer

Quantum contributions to the interaction energy play an important role.

FMO Fully Analytic Gradient

The gradient for the FMO method was derived at the same time as the original FMO equations.

Since the FMO method is not fully variational, a fully analytic gradient required the solution of the Coupled Perturbed Hartree-Fock (CPHF) equations.

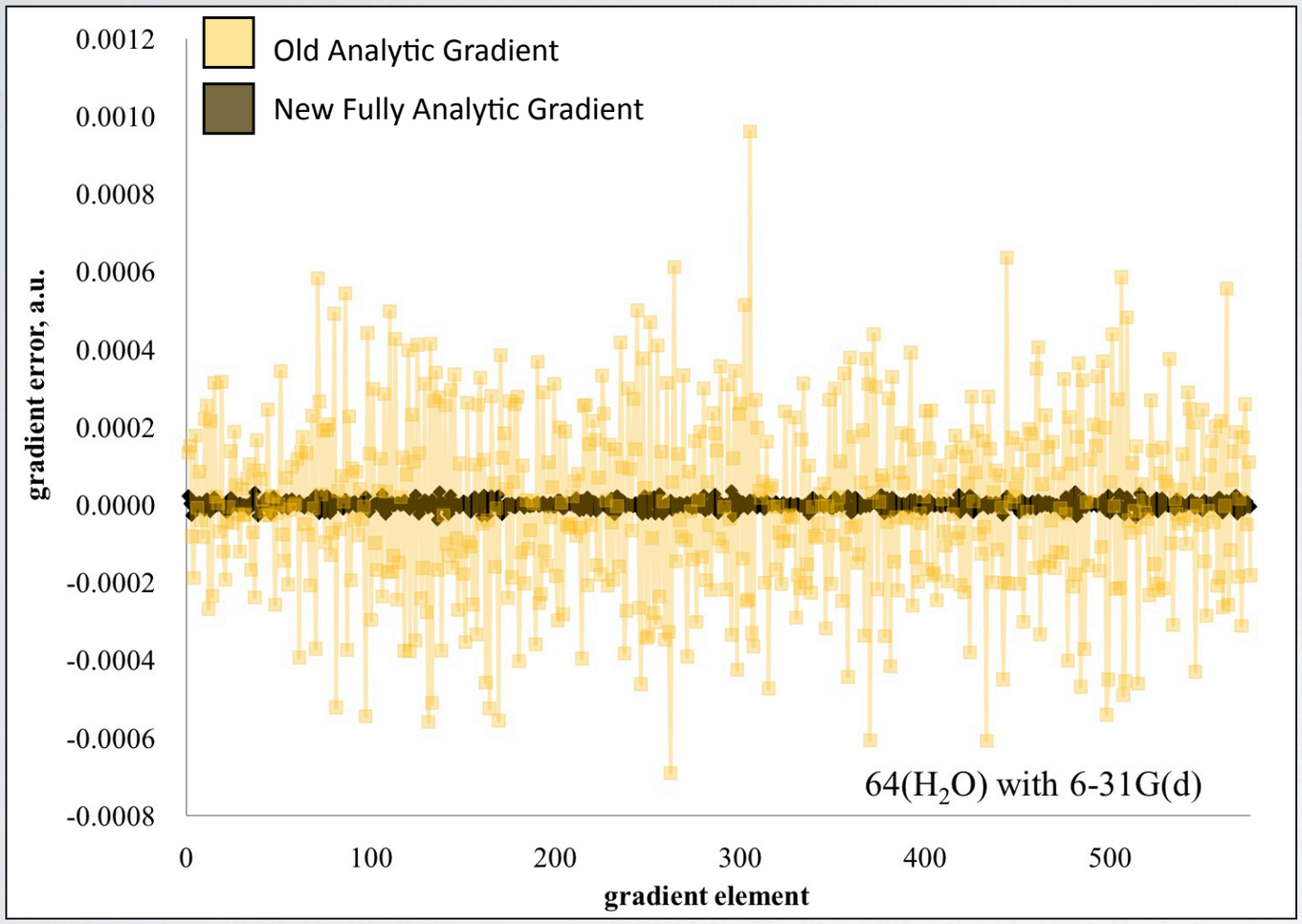
Due to the difficulty in solving the CPHF equations, their contribution to the gradient was ignored (termed "near fully analytic").

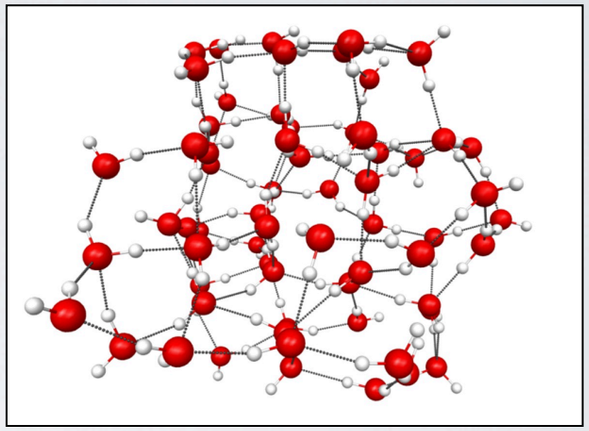
The original gradient was checked against numeric results, and it was determined it was sufficient for geometry optimizations, but the accuracy was not high enough for FMO-MD.

Additionally, with the original gradient, errors tended to increase for large basis sets.

FMO Fully Analytic Gradient

$(H_2O)_{64}$ RHF/6-31G(d)



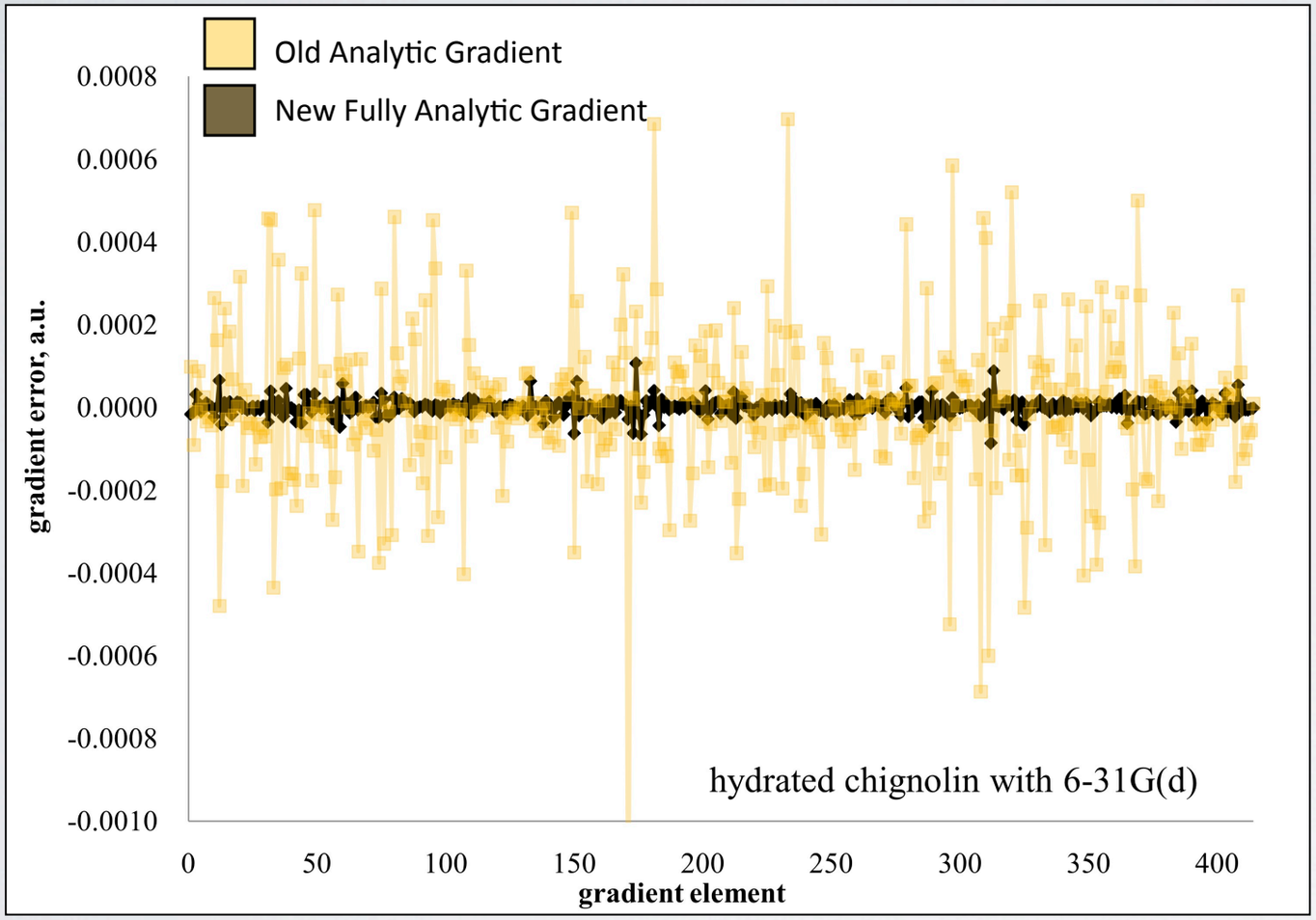


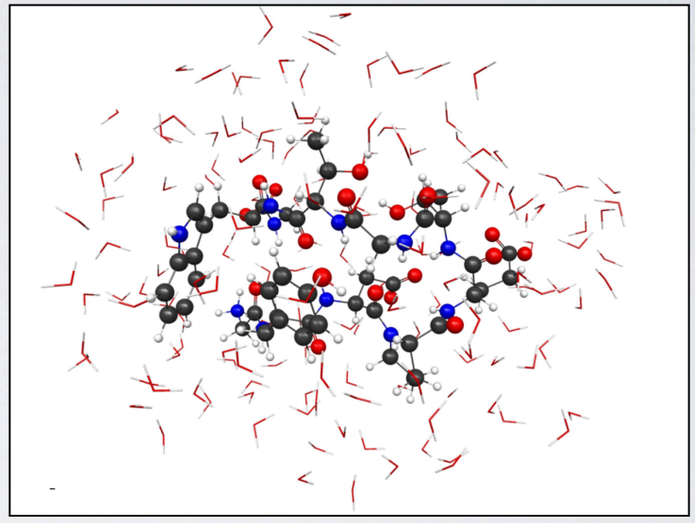
	Old Analytic Gradient	New Fully Analytic Gradient
Max Gradient Error (a.u.)	0.000961	0.000035
RMS Gradient Error (a.u.)	0.000231	0.000011

Nagata, T.; Brorsen, K.; Fedorov, D.G.; Kitaura, K.; Gordon, M.S. J. Chem. Phys., 2011, *134*, 124115

FMO Fully Analytic Gradient

Chignolin/EFP RHF/cc-pVDZ

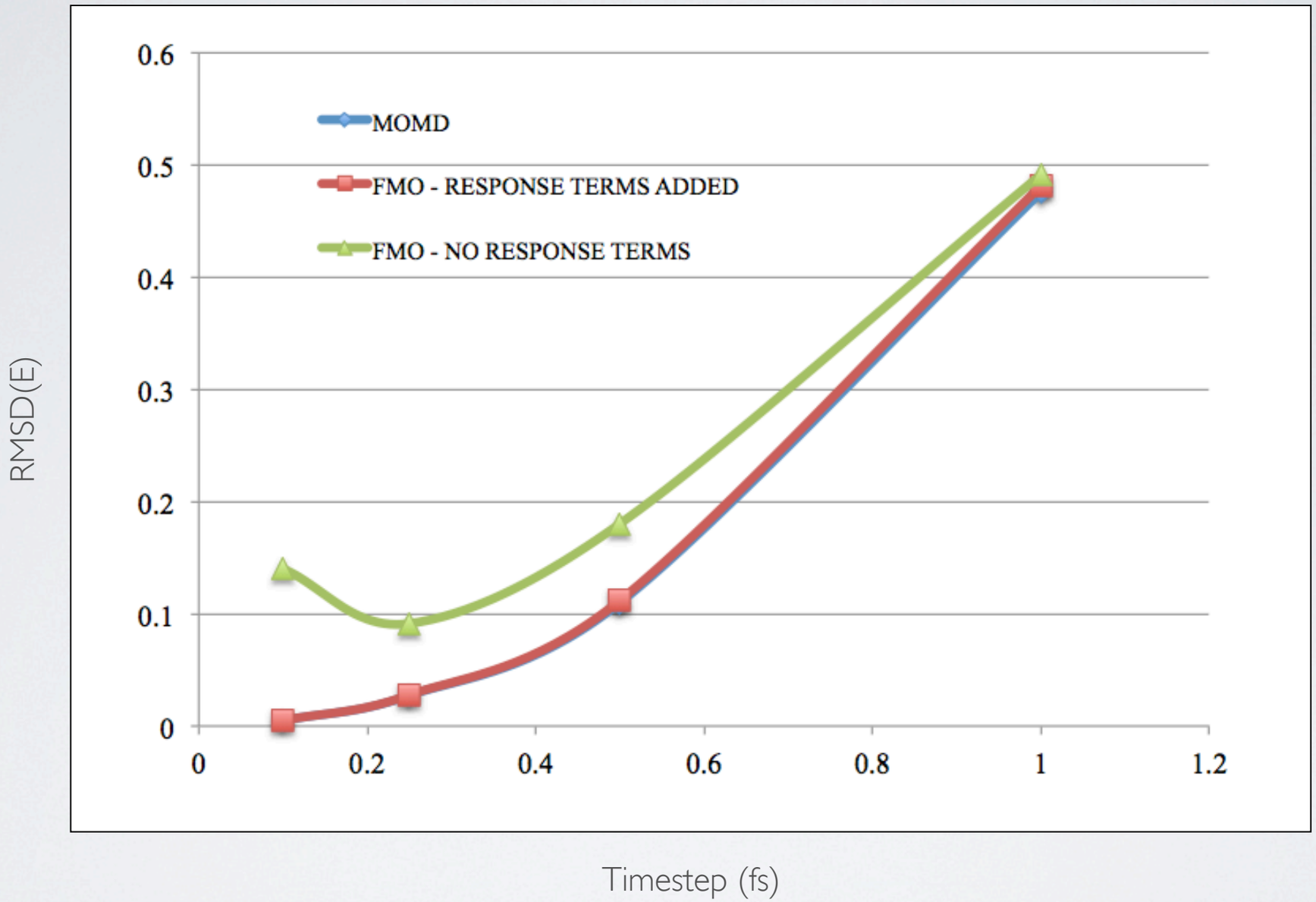




	Old Analytic Gradient	New Fully Analytic Gradient
Max Gradient Error (a.u.)	0.001501	0.000092
RMS Gradient Error (a.u.)	0.000191	0.000017

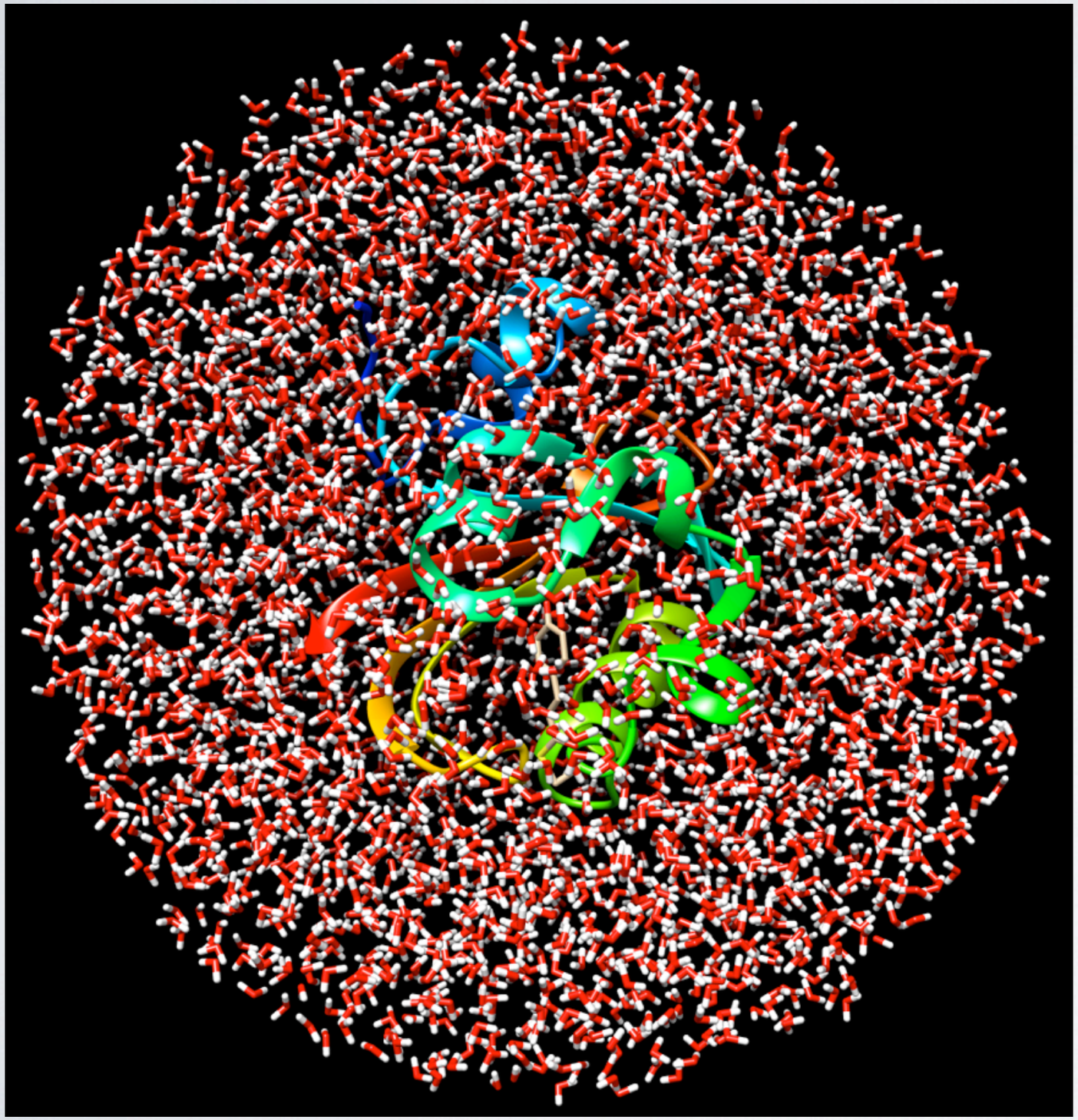
FMO Molecular Dynamics with the Fully Analytic Gradient

$(H_2O)_{32}$ RHF/6-31G(d), 200 FS



Brorsen, K.; Minezawa, N.; Feng, X.; Windus, T.L.; Gordon, M.S. J. Chem. Theor. Comp., 2012, *8*, 5008

GAMESS Scaling on "That Other Supercomputer"



- 124 amino acid residues
- 3460 water molecules
- one residue or water per fragment
- FMO2-MP2/6-311G(d,p) energy calculation
- 131,432 basis functions
- 16,384 nodes (262,144 cores)
- ~10 hours
- ~72% of Blue Waters

Conclusions

Highly scalable algorithms have been implemented and improved thanks to the INCITE and ESP access to Blue Gene

These new and improved algorithms have allowed specific scientific questions that could not be answered to be addressed at a high level of accuracy.

Due to work coming directly from INCITE and ESP, accurate QM molecular dynamics simulations are now possible with GAMESS. (DOE CSGF - Kurt Brorsen)

Future areas of improvement involve reductions in amount of I/O during FMO calculations and faster algorithms for dealing with the increasing number of separated n-mer interactions.

Acknowledgements

Mark Gordon
Graham Fletcher
Yuri Alexeev
Maricris Mayes
Dmitri Fedorov
Mike Schmidt
Theresa Windus
Monica Lamm
Kurt Brorsen
Seung Ha Kim
Federico Zahariev
Ajitha Deverajan

IOWA STATE UNIVERSITY
Department of Chemistry

DOE CSGF
DOE Computational Science Graduate Fellowship

THE Ames Laboratory
Creating Materials & Energy Solutions
U.S. DEPARTMENT OF ENERGY

AIR FORCE OFFICE OF SCIENTIFIC RESEARCH
TODAY'S BREAKTHROUGH SCIENCE FOR TOMORROW'S AIR FORCE
Air Force Office of Scientific Research

Argonne NATIONAL LABORATORY

DEPARTMENT OF ENERGY · UNITED STATES OF AMERICA
Department of Energy