



Pacific  
Northwest  
NATIONAL LABORATORY

# Improving Spectroscopic Analysis using Machine Learning from Atomistic Simulations

April 17, 2019

Eric Bylaska (PNNL)

Raymond Atta-Fynn (UTA)

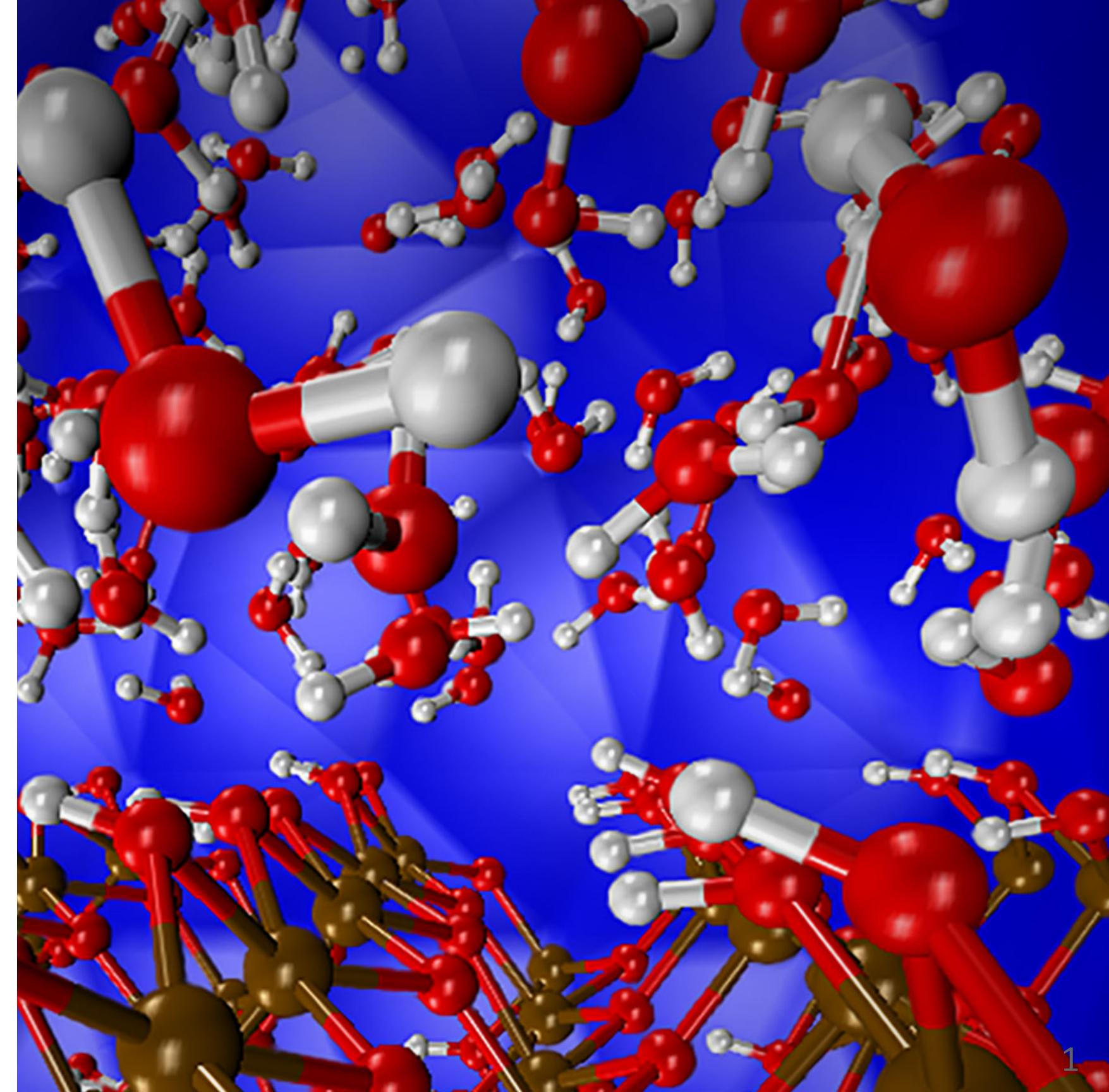
---

"A man dreams of a miracle and wakes up with loaves of bread"

Erich Maria Remarque

U.S. DEPARTMENT OF  
**ENERGY** **BATTELLE**

PNNL is operated by Battelle for the U.S. Department of Energy

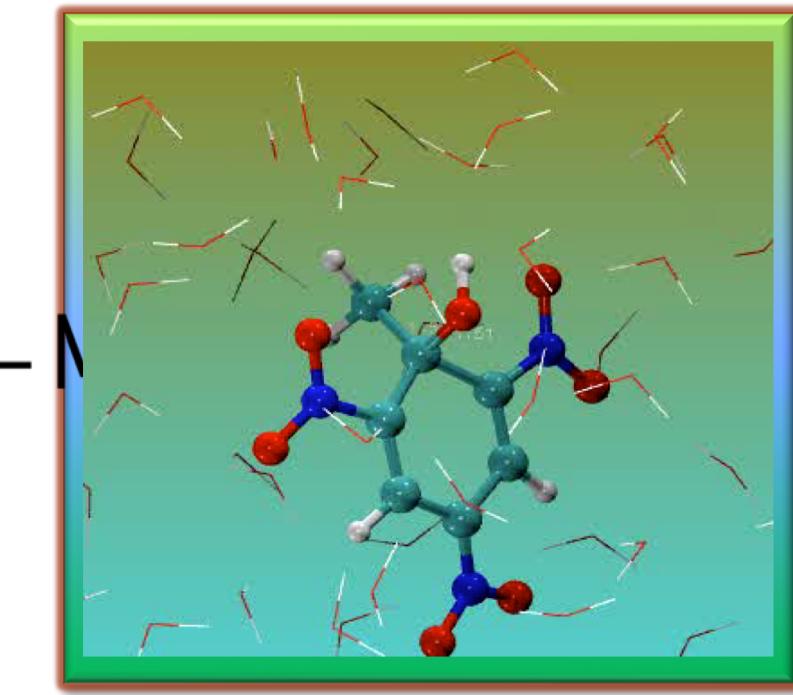
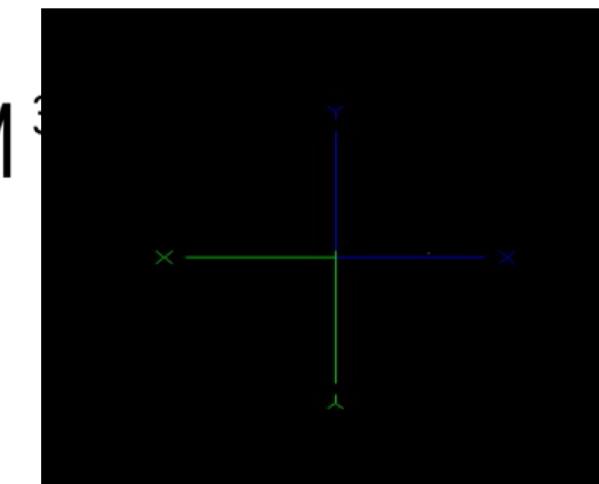
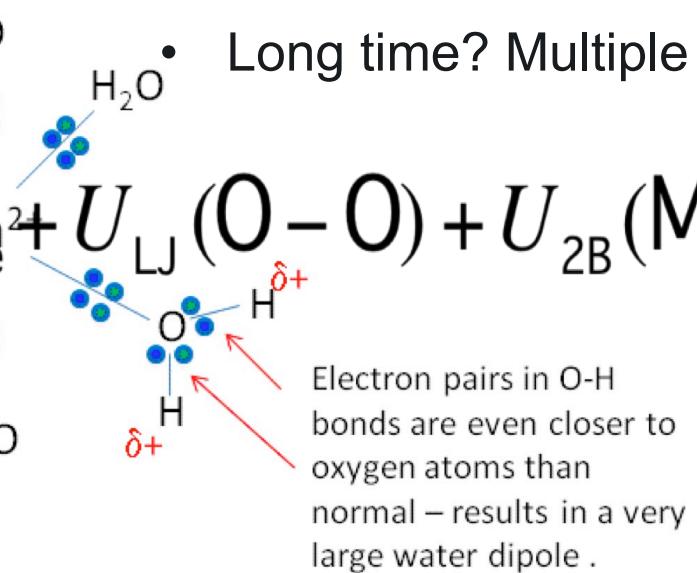
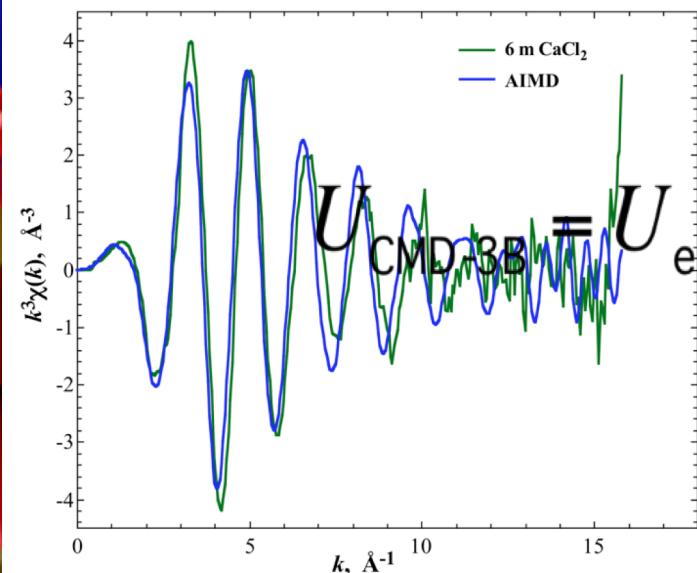
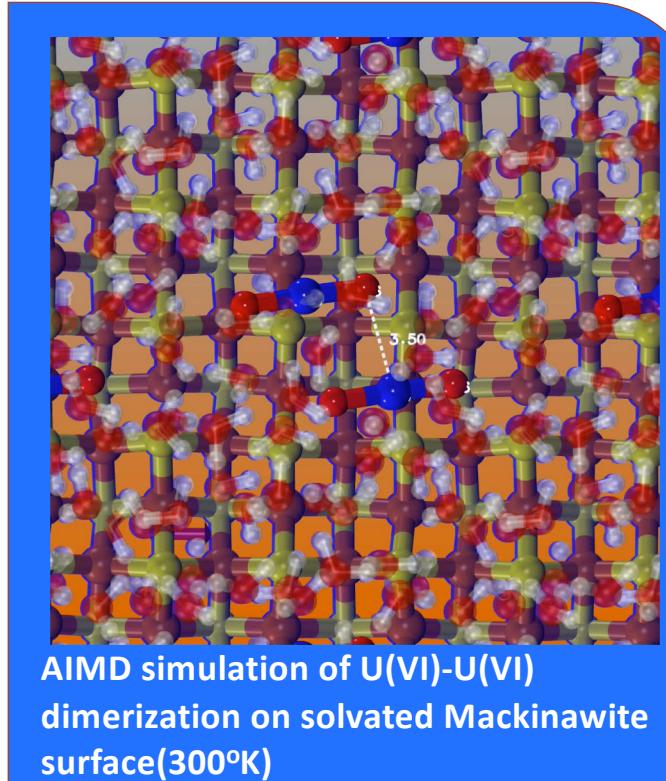


# Outline

- Molecular Modeling in Geochemistry
- Ab initio Molecular Dynamics - Predictive Model for Molecular Geochemistry
- Solving  $dx/dt = F(x)$  faster
- Challenges and Opportunities in Fitting Molecular Dynamics Potentials with ML

# Challenges for Molecular Modeling Of Geochemistry/Actinide Chemistry

- Most (all) geochemistry problems begin at the nano-scale
  - Large numbers of atoms (1000's) and long simulation times are needed to simulate dilute solutions
- Molecular level experiments need interpretation
- Classical two and three body potentials are often used for these type of simulations
- Unfortunately, these classical potentials are often not very good at polarization, e.g. predicting the hydration shells of many aqueous metal (surface) species, or chemical bond breaking/making
- Ab Initio Molecular Dynamics (AIMD) avoids the use of such potentials but is only practical for ~~100's~~ 1000's of atoms.
- Still need Free energy pathways, PMFs
- Long time? Multiple Instances?

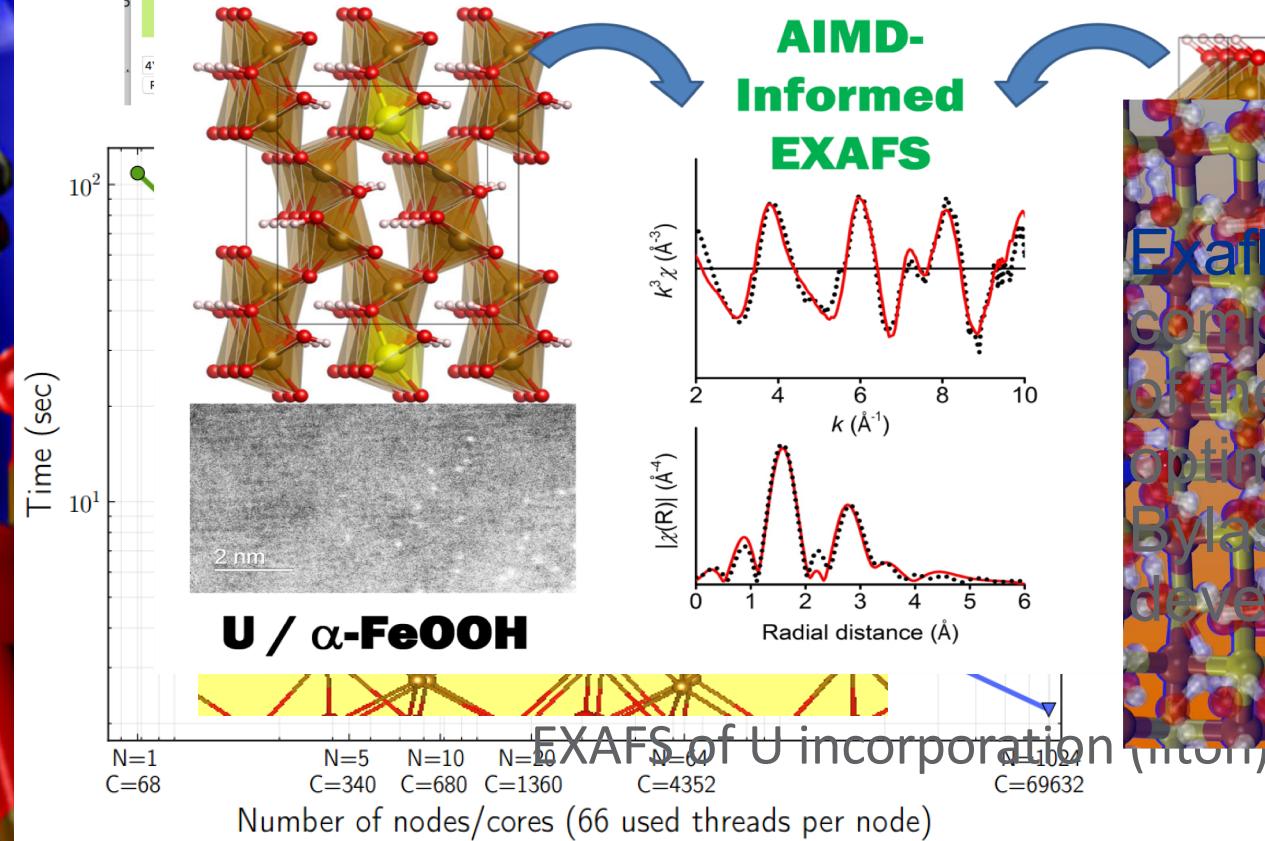




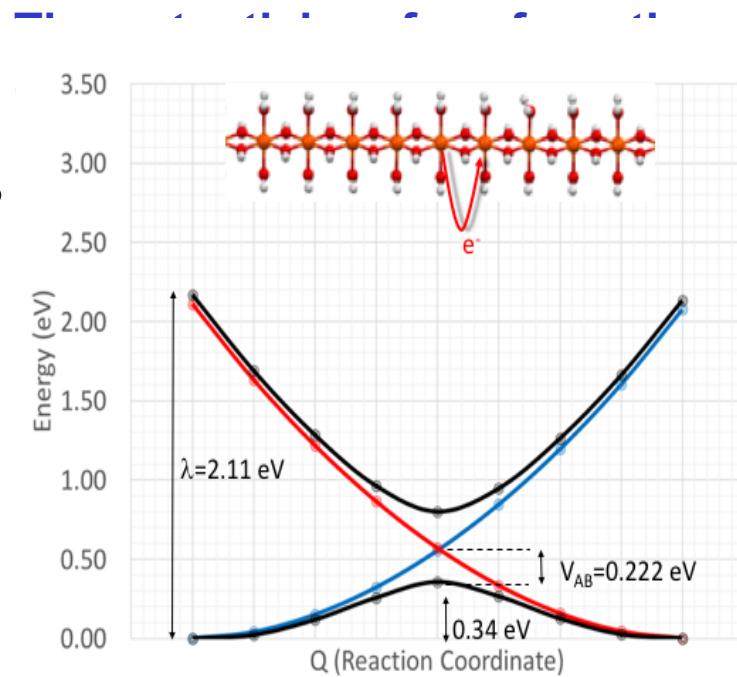
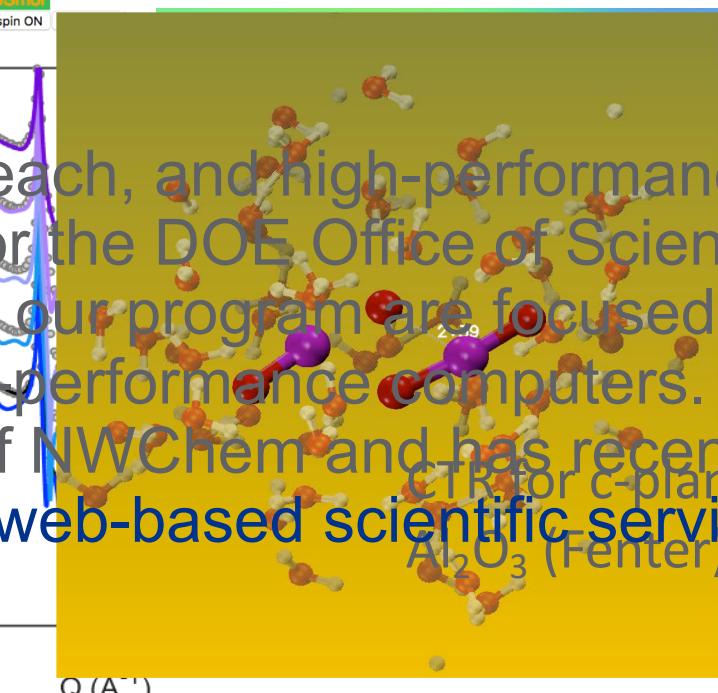
Pacific  
Northwest  
NATIONAL LABORATORY

# Development of Advanced Molecular Models for Geochemistry

Accurate modeling of redox processes, in particular electron transfer (ET) reactions at interfaces and polarization motion in solids, is playing an increasingly important role in modeling and interpretation of spectroscopies. However, these simulations need to be constantly updated to more reliably capture the complex behavior of the properties of the strongly correlated materials containing localized d electrons, coupled with long range processes such as hydration, disorder of the surface/solution interface, the interaction of the solution phase

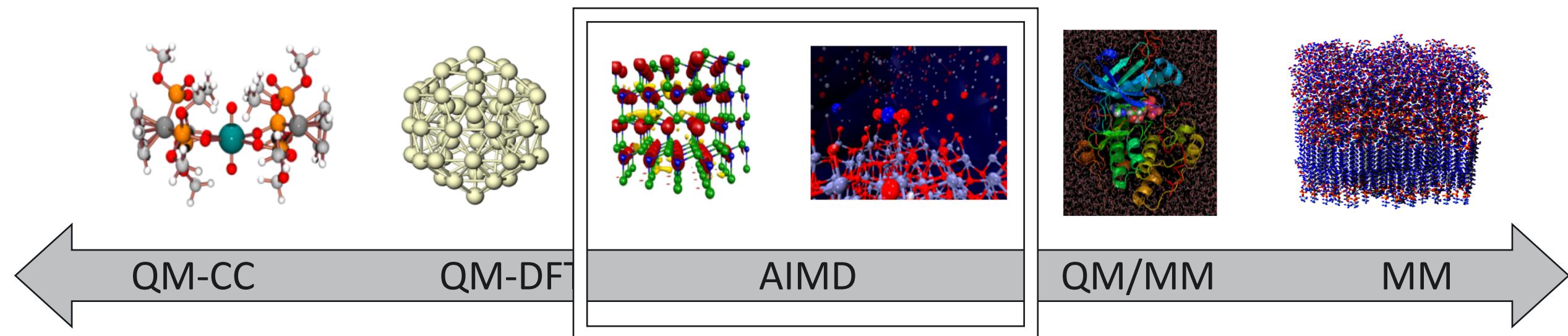


Exaflop computing is within reach, and high-performance computing is a priority area for the DOE Office of Science. All of the simulation advances of our program are focused on the optimal use of emerging high performance computers. Bylaska is a primary author of NWChem and has recently developed the EMSl Arrows web-based scientific service.

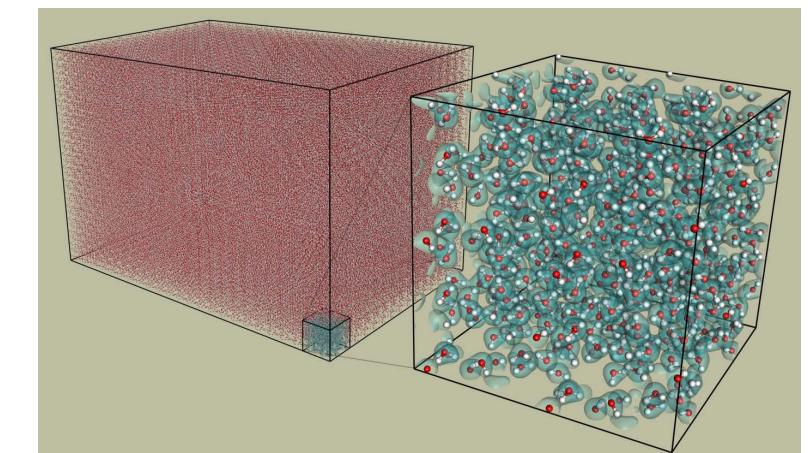


# Ab initio molecular dynamics (AIMD)

Development emphasis is on providing parameter-free predictions applicable to a wide range of temperatures, pressures, and compositions



- 100-1000 atoms,  
uses plane wave basis
- >>10K atoms likely within 2 years
- Many FFTs and  
DGEMM operations
- “Meaty”: Lots of FLOPs,  
but also bandwidth sensitive



# Basic Features of Ab Initio Molecular Dynamics

DFT Equations → MCSCF

$$H\psi_i = \varepsilon_i\psi_i$$

$$H\psi_i(\mathbf{r}) = \left( -\frac{1}{2}\nabla^2 + V_l(\mathbf{r}) + \hat{V}_{NL} + V_H[\rho](\mathbf{r}) \right) \psi_i(\mathbf{r}) - \alpha \sum_j K_{ij}(\mathbf{r}) \psi_j(\mathbf{r}) \\ + (1-\alpha) V_x[\rho](\mathbf{r}) + V_c[\rho](\mathbf{r})$$



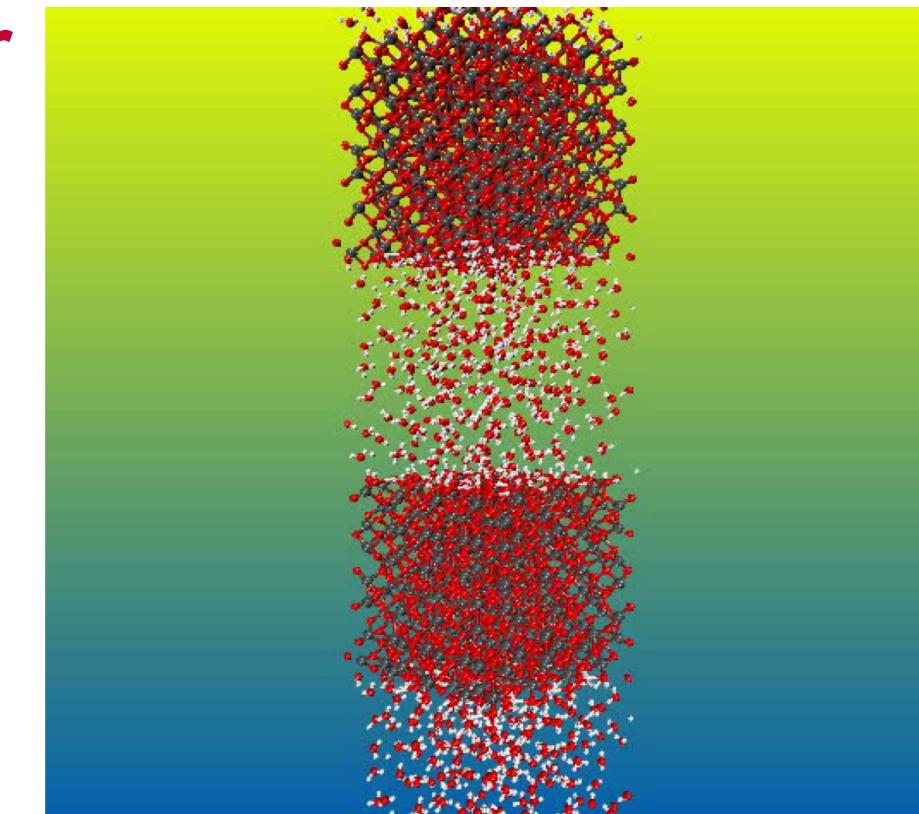
**CP dynamics: Ion and wavefunction motion coupled. Ground state energy  $\mu=0$**

$$\mu\ddot{\psi}_i = H\psi_i - \sum_{j=1}^{N_e} \lambda_{ij} \psi_j$$

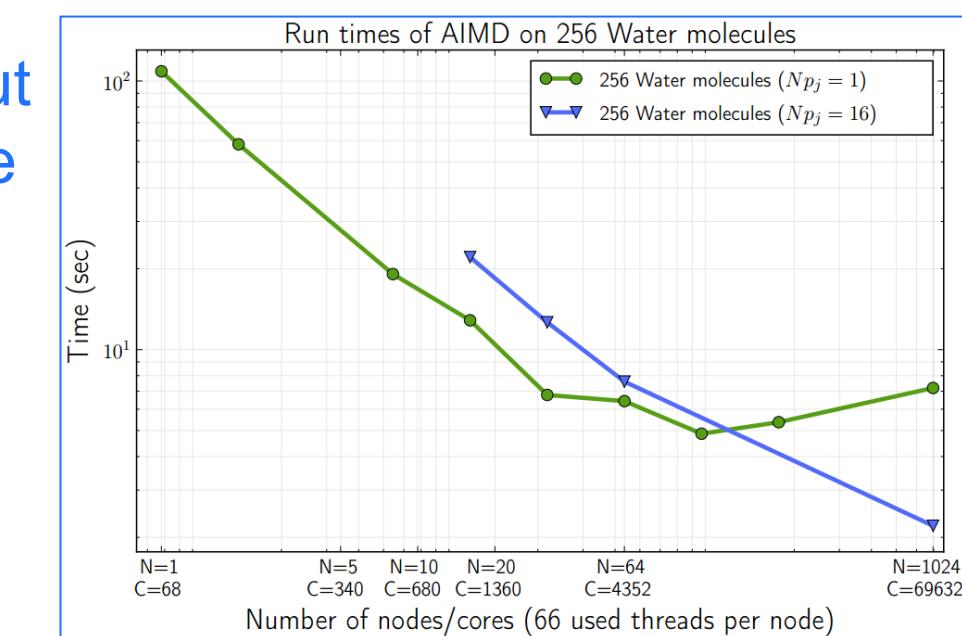
$$M_I \ddot{\mathbf{R}}_I = \mathbf{F}_I \quad \mathbf{F}_I = \sum_{i=1}^{N_e} \left\langle \psi_i \left| \frac{\partial H}{\partial \mathbf{R}_I} \right| \psi_i \right\rangle$$

Doable but  
Expensive

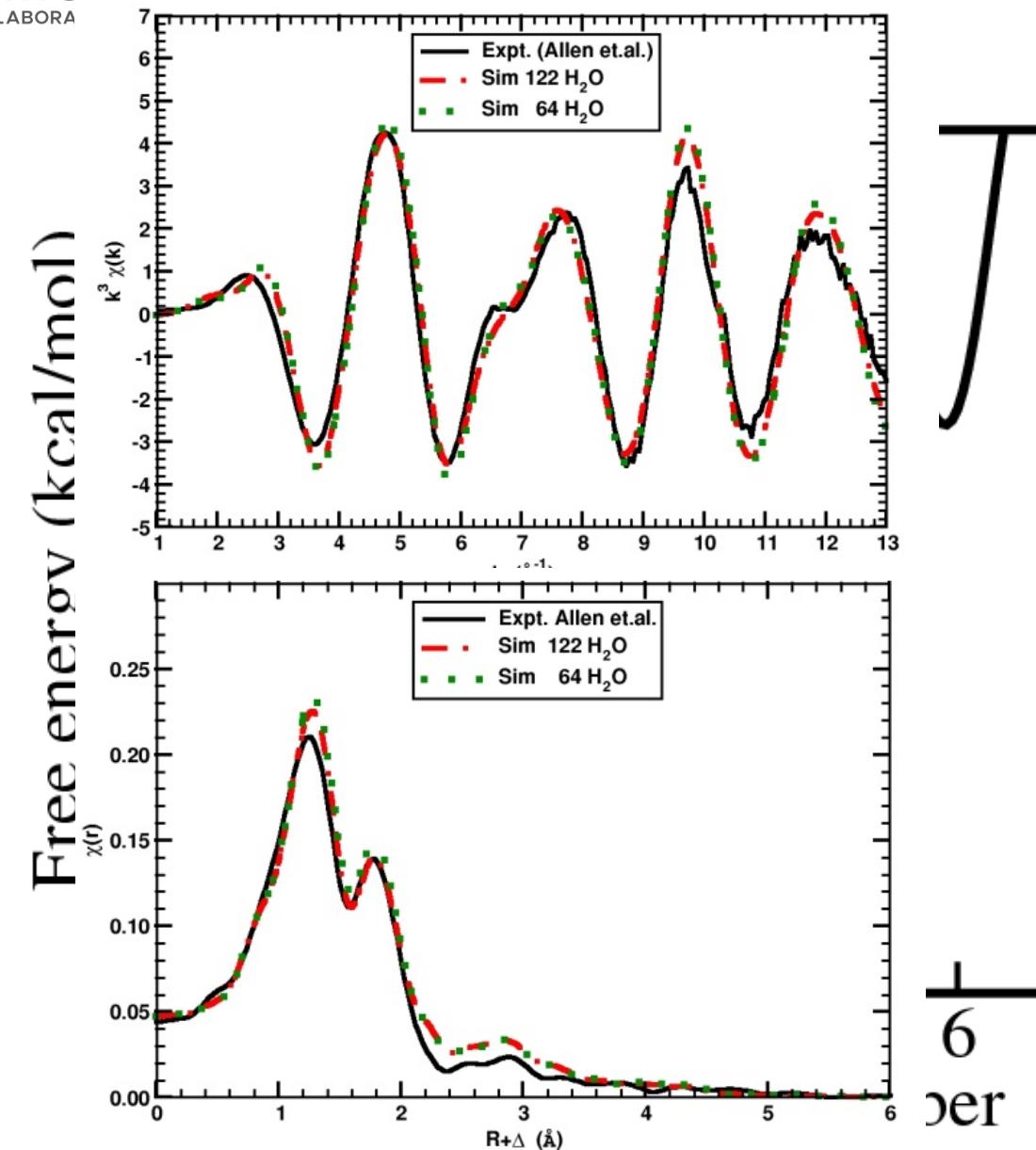
Want to do this in ~1second per step



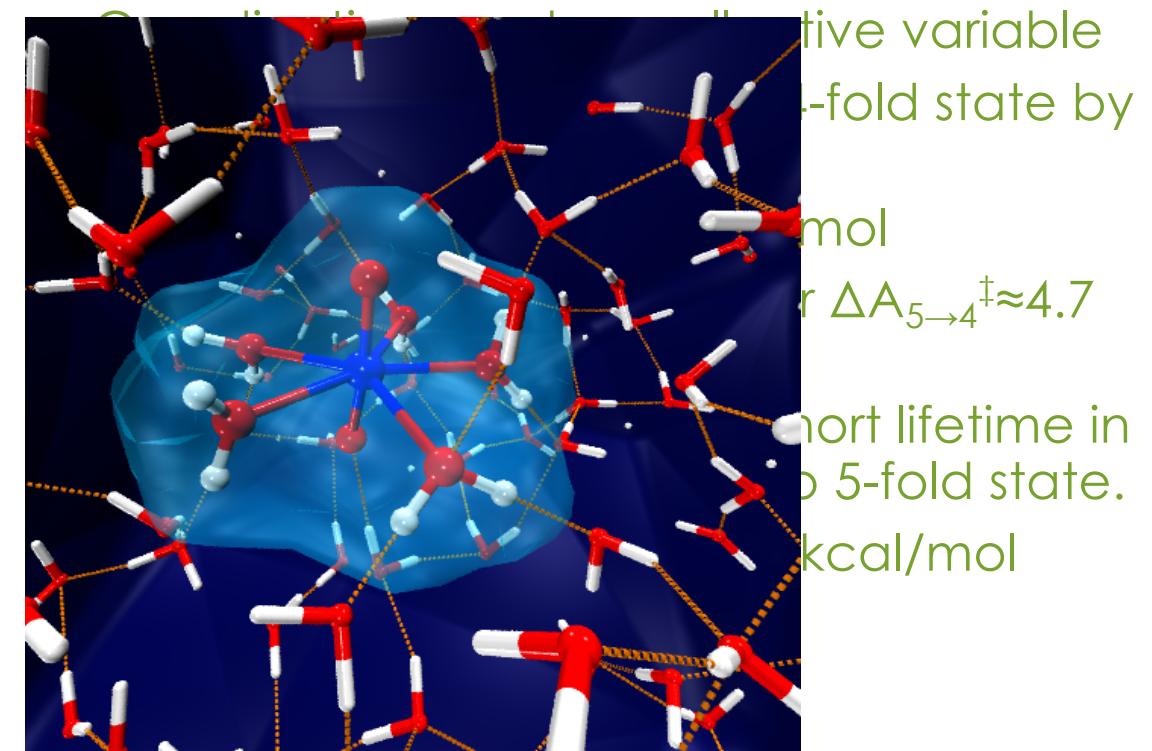
Plane-wave basis sets, pseudopotentials are used to solve PDEs

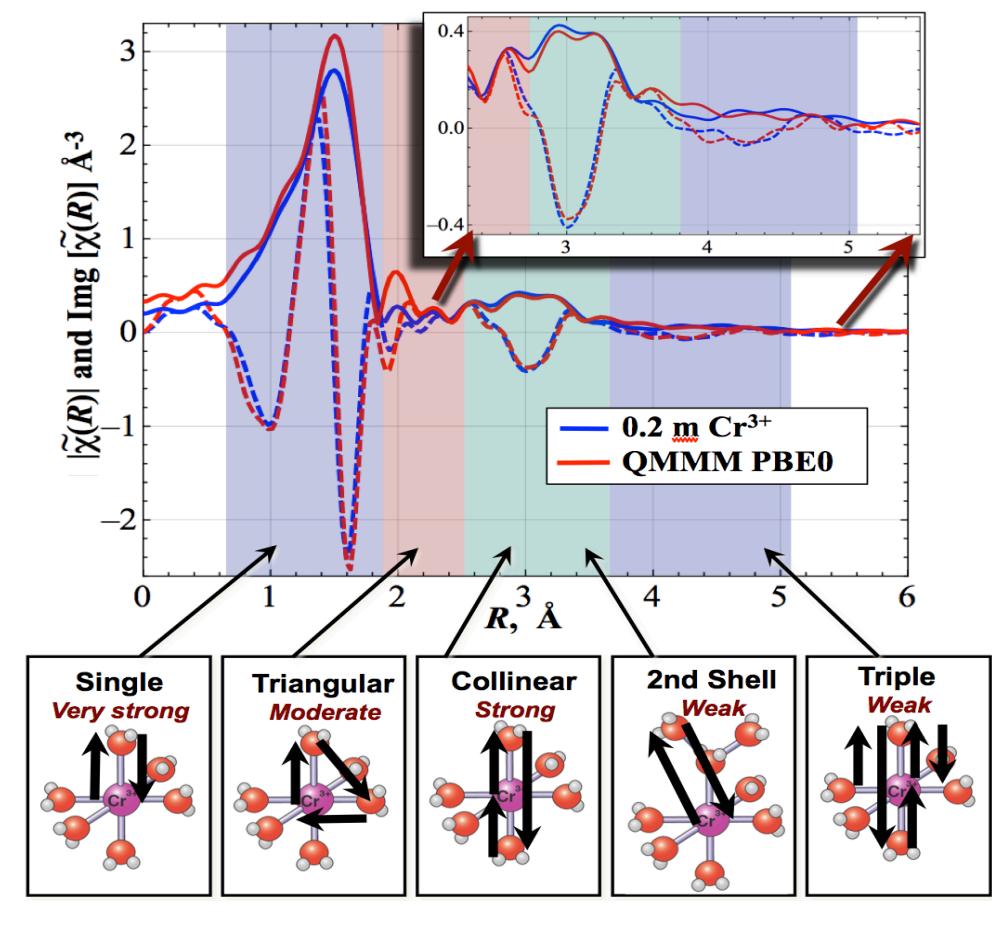
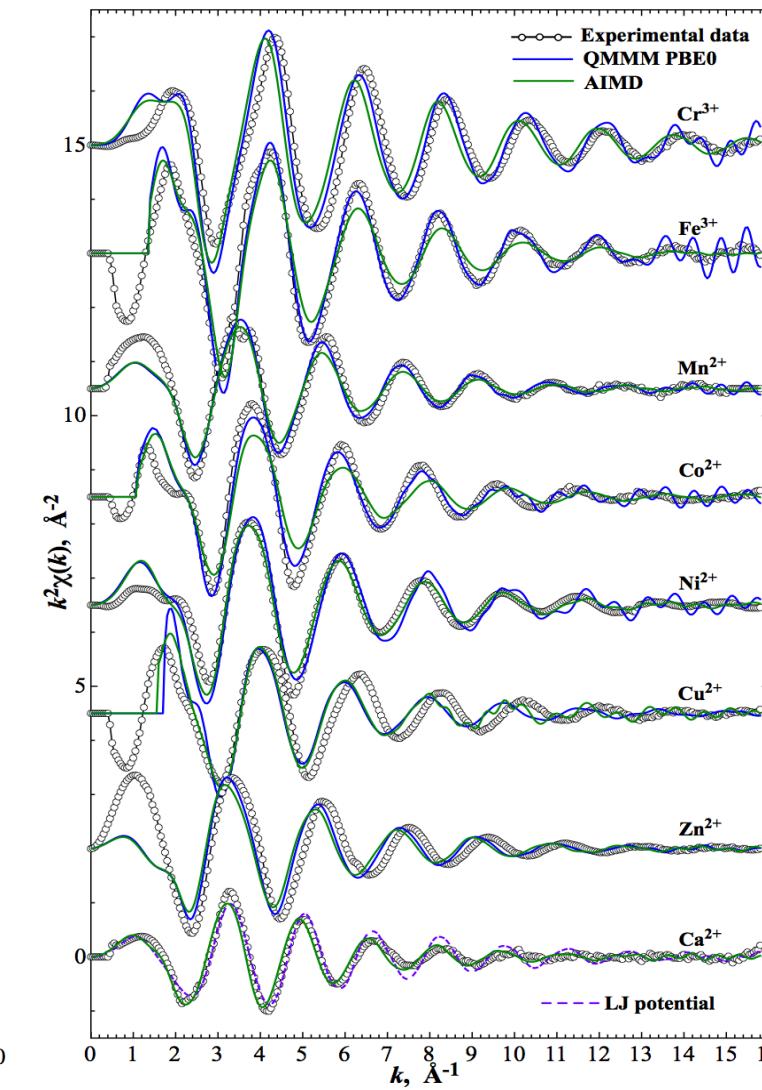
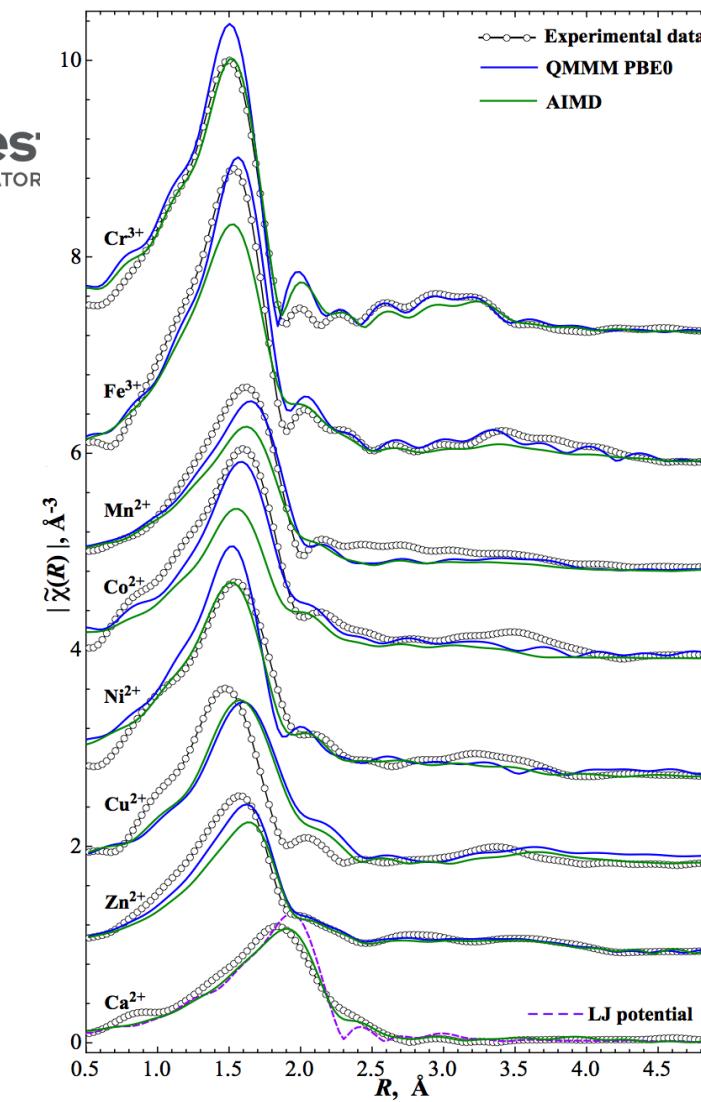


# Predictive Model for $\text{UO}_2^{2+}(\text{aq})$



- Good EXAFS agreement but recent HEXS experiments suggest a 4-fold state is energetically nearby
- Results from Metadynamics

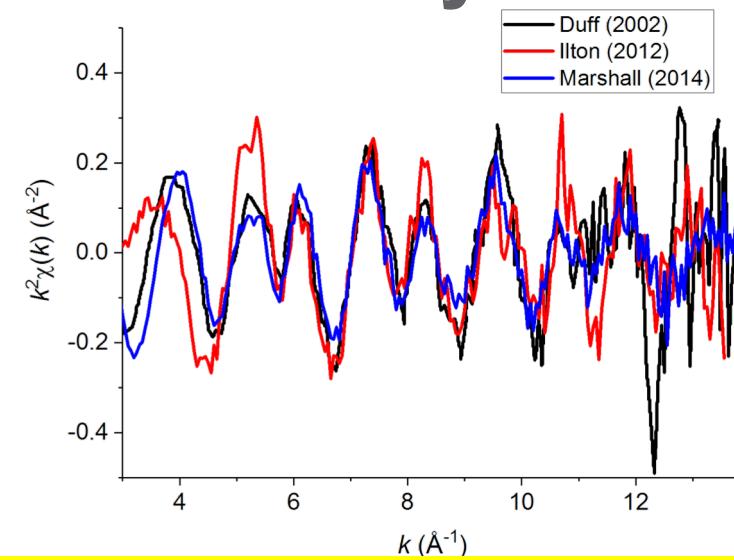




The generally excellent agreement of the 1<sup>st</sup> principle MD-XAFS simulation with the data. The scans are calculated by a parameter free method which can be implemented more efficiently than the use of empirical interactions suggesting that this **method can be used to interpret more XAFS spectra in more complex environments.**

# Octahedral uranyl-like U(VI) incorporated in hematite is accommodated by protonated trans-corner Fe vacancies

## Shell by shell



**UL<sub>3</sub> EXAFS are nearly identical, but three different interpretations:**

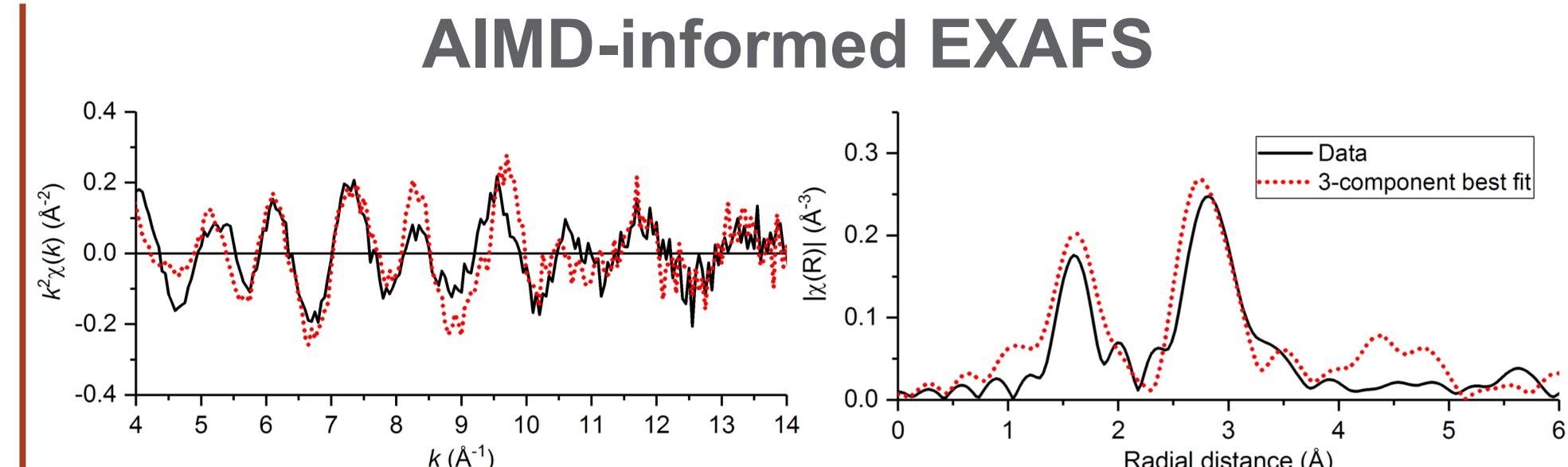
Duff *et al.* (2002)

Ilton *et al.* (2012)

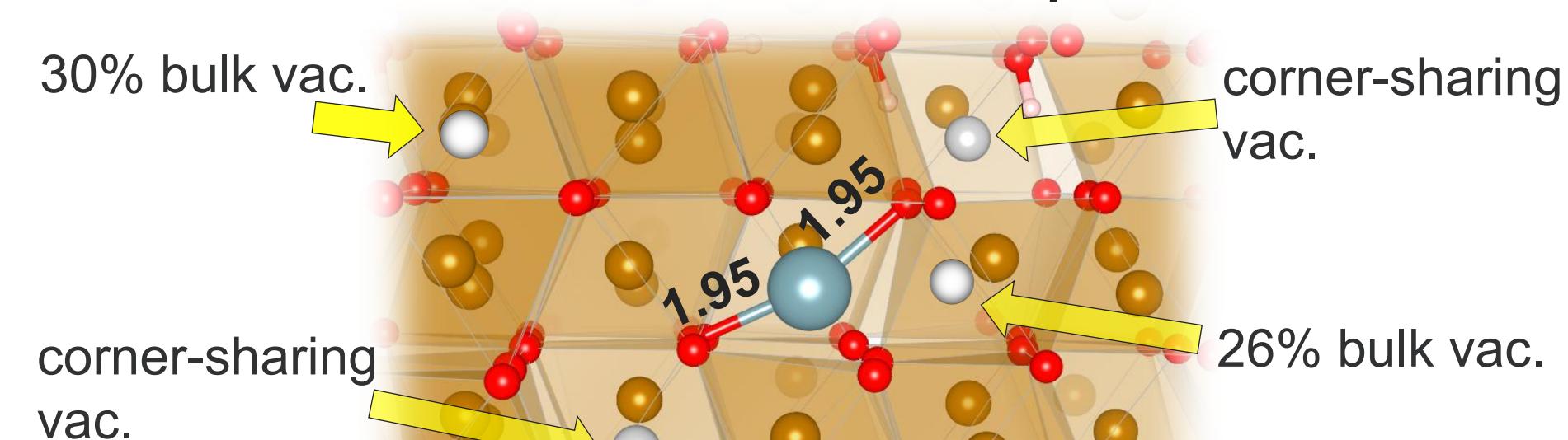
Marshall *et al.* (2014)

All used very high Debye-Waller factors to model the first shell

## AIMD-informed EXAFS



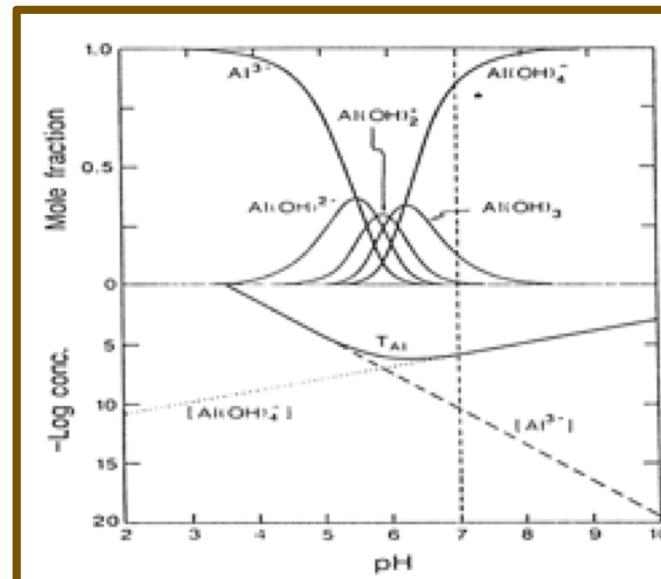
44% adsorbed and 56% incorporated





Pacific  
Northwest  
NATIONAL LABORATORY

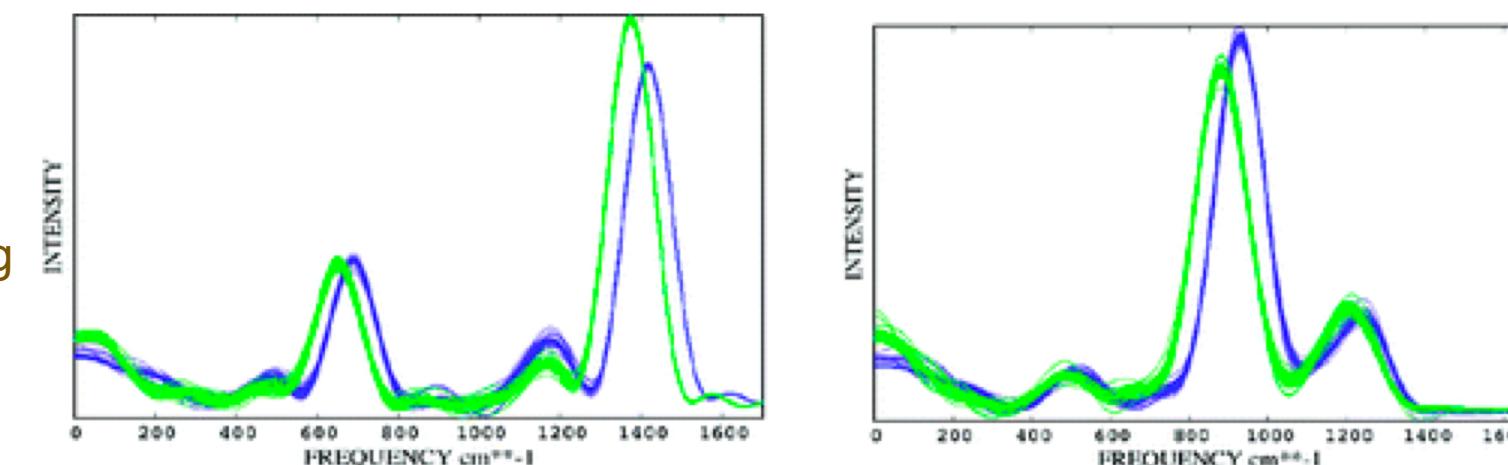
# Future Transformative Applications in Geochemistry



Old Way – The diagram for the distribution of aluminum species was determined entirely from fitting thermodynamic data using an **assumed speciation scheme**.

Synergistic AIMD free energy simulations with XAS spectra can be used to determine the solute structure of environmentally important species in solutions as a function of TPX. Used to develop chemically and thermodynamically highly accurate solution models with exceptional extrapolation properties in TPX.

- Strategies to search configuration space must be developed.



AIMD simulations are already finding relevant solvent structures in a first principles approach to calculating isotope fractionation based on harmonic frequencies.

- If the simulations were faster, it would be possible to directly calculate the fractionation factors using quantum dynamics.

# Other Spectroscopies: Algorithms for IR and Raman spectra interpretation

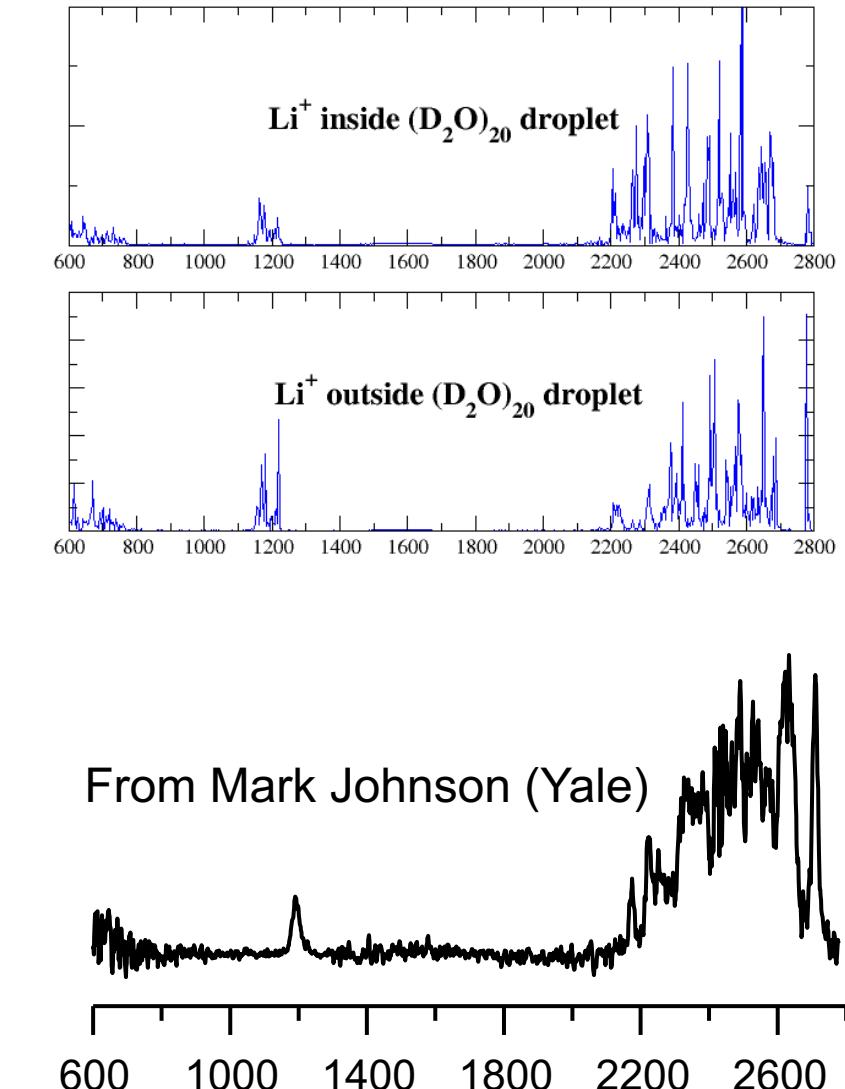
- AIMD simulations have the potential to significantly improve the molecular interpretation of IR and Raman spectroscopies *along recent variants of them where they are combined with instrumentation such as AFM.*
- In principle, AIMD analysis methods can easily be extended to IR and Raman (i.e., AIMD-IR and AIMD-Raman).

$$Absorption(\omega) = \frac{1}{6cVn(\omega)\varepsilon_0 k_B T} \int_{-\infty}^{\infty} e^{i\omega t} \langle \dot{P}(0) | \dot{P}(t) \rangle dt$$

$$P_\mu = \frac{2|e|}{|b_\mu|} \text{Im} \log \det Q^\mu \quad \text{where } Q_{i,j}^\mu = \langle \psi_i | e^{ib_\mu \cdot r} | \psi_j \rangle$$

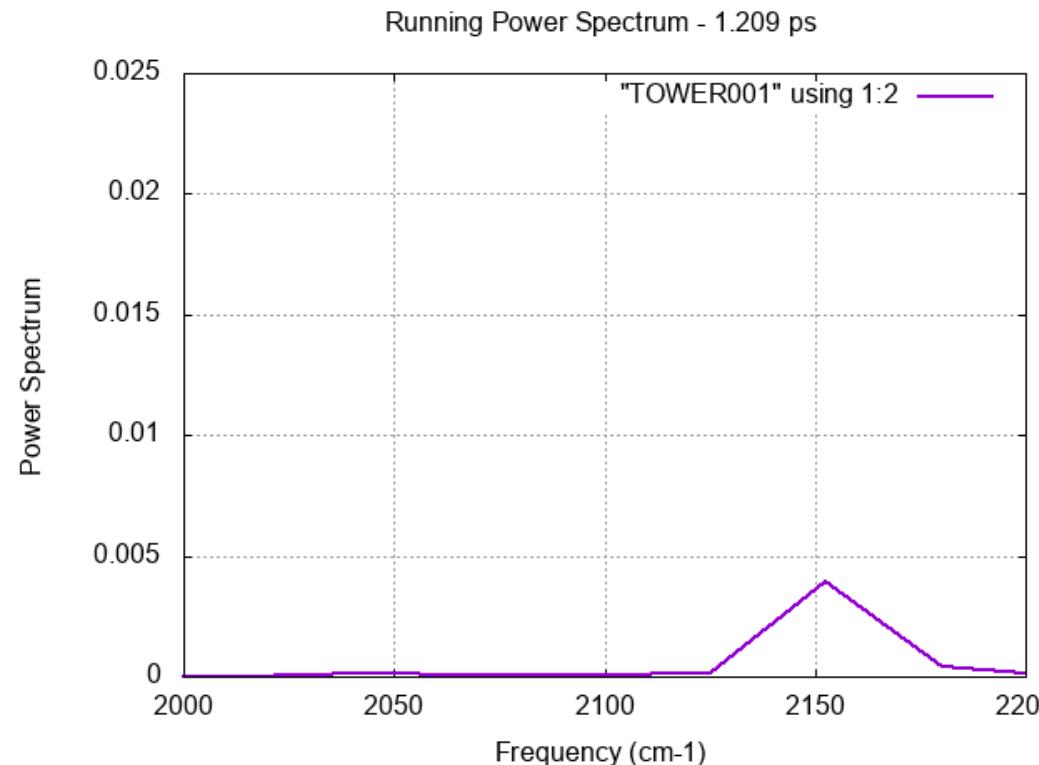
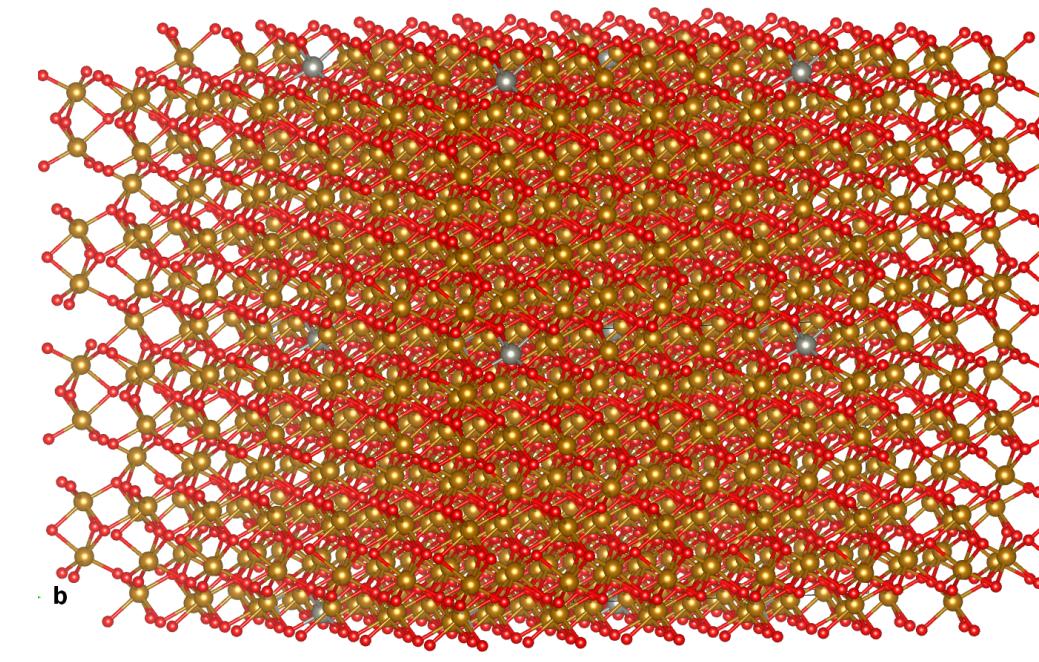
$$\alpha_{\mu\nu} = \frac{2|e|}{|b_\mu|} \text{Im} \left[ \sum_{i,j} \left( \langle \psi_i^{(v)} | e^{ib_\mu \cdot r} | \psi_j \rangle + \langle \psi_i | e^{-b_\mu \cdot r} | \psi_j^{(v)} \rangle \right) [Q^\nu]_{j,i}^{-1} \right]$$

- Improving standard vibrational analysis
  - Larger and more complex systems
  - Better sampling
  - Longer time scales for single molecule



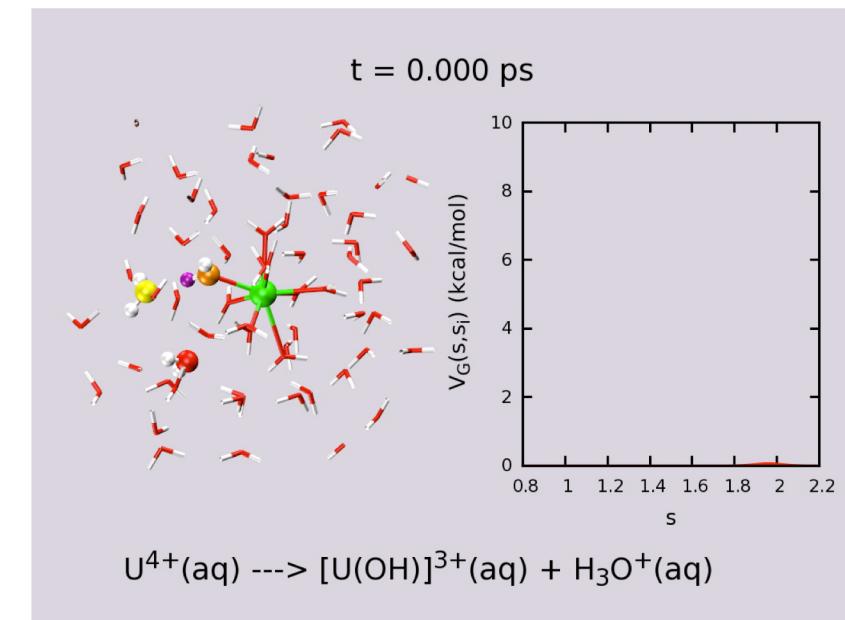
# AIMD simulations will need long trajectories to perform autocorrelation functions

- Ab initio molecular dynamics (AIMD) has transformed how spectroscopic measurements from advanced light sources are analyzed, such as Raman, EXAFS, CTR, XANES, etc.
- Advanced HPC algorithm development has made the first-principles analysis of advanced light sources possible for the first time
- However, the computation cost of AIMD is prohibitive for many projects and other possible predictive analysis, e.g. isotope fractionation using quantum dynamics

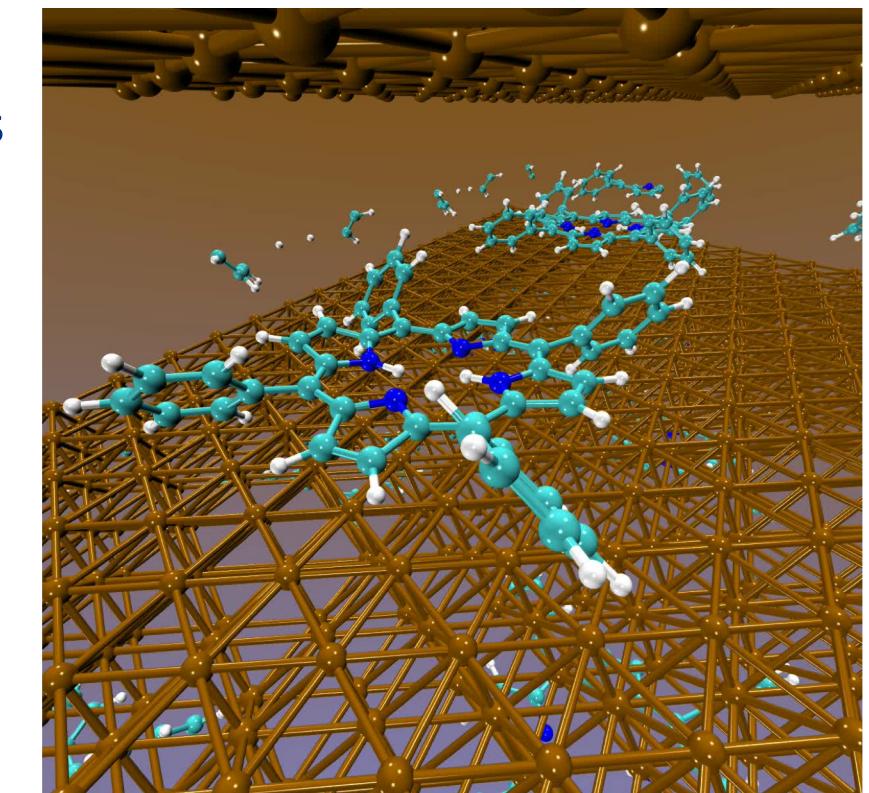


# Strong Scaling is Key

- 20 picoseconds of simulation time ≈ 200,000 steps
  - 1 sec/step = 2-3 days simulation time
  - 10 sec/step = 23 days simulation time
  - 13 sec/step = 70 days simulation time
- Mesoscale phenomena at longer time scales
  - Assume 1 sec/step
  - 100 psec = 10-15 days simulation time
  - 1 nsec = 100 - 150 days simulation time
- Strong scaling required to reduce time per time step as much as possible
  - At least below 1sec/step

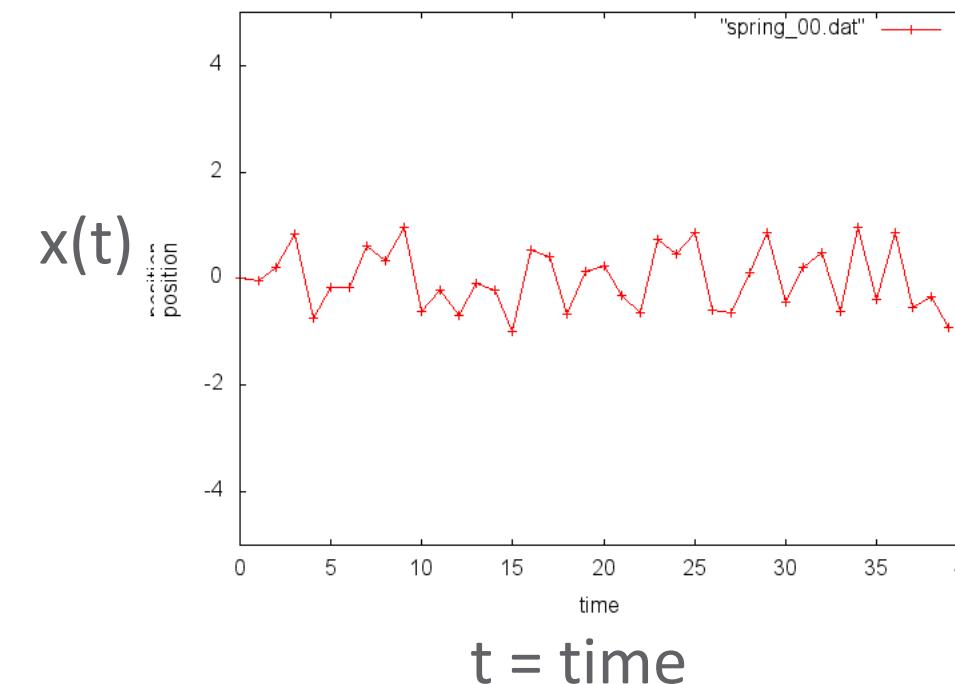


Surface spectroscopies



# Possible Solution: Parallel in Time

- Increasing the time step ( $dt$ ) in time integration quickly becomes unstable
- One approach to bridging these temporal scales is the development of algorithms which parallelize over time, i.e. parallel in time algorithms
- The central philosophy of parallel in time integration is to start with a guess for the trajectory over some fixed time interval and then attempt to relax it until it approximates the “true” trajectory.



Trajectory for a simple spring  
( $K=1, x_0=1, v_0=0$ )

Increasing time step



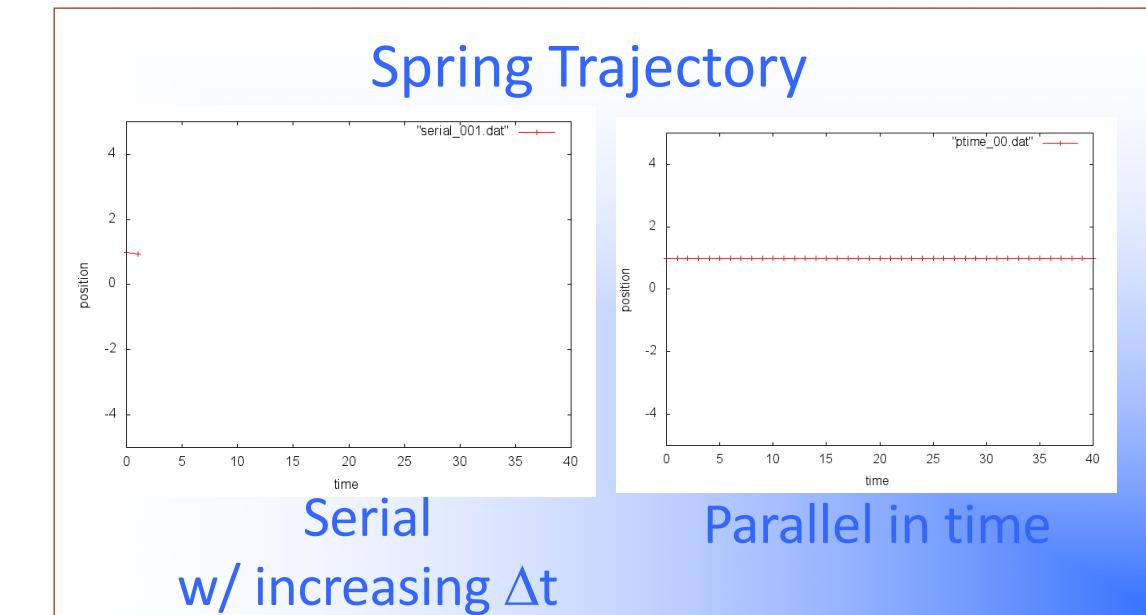
Can this be  
parallelized????



# Parallel In Time Algorithms Without Using Approximate Models: Fixed Point Parallel in Time Algorithms

These algorithms transform standard forward substitution time integration solvers, i.e.  $x_{i+1} \leftarrow f(x_i)$ , into fixed-point root problems

$$F(X) = 0 \text{ or } \begin{pmatrix} x_1 - f(x_0) \\ x_2 - f(x_1) \\ \vdots \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \end{pmatrix}$$



Can be solved using a variety of optimization techniques, including preconditioned fixed-point, quasi-Newton, and preconditioned quasi-Newton optimization methods (i.e. solve  $F(G^{-1}(X))=0$ ). These algorithms can be parallelized since the evaluation of the trial root function  $F(X)$  can be done in parallel. See Bylaska et al. Extending Molecular Simulation Time Scales: Parallel in Time Integrations for High-Level Quantum Chemistry and Complex Force Representations. *J. Chem. Phys.* **2013**, *139*, 074114. DOI: [10.1063/1.4818328](https://doi.org/10.1063/1.4818328).

These algorithms are particularly useful for diffusion based mesoscale models

- Note phase field  $\Delta t \sim \Delta X^4$



# Parallel in Time: Fixed Point Iteration

The serial solution to time integration,

with initial condition

is  $X_{\text{trajectory}} = \begin{bmatrix} f(x_0) \\ f(f(x_0)) \\ f(f(f(x_0))) \\ f(f(f(f(x_0)))) \end{bmatrix}$

$$x_{i+1} = f(x_{i-1}) \quad x_0 = x_0$$

Using column vector to store  
each step in the time iteration  
from i=1,4

This equation can also be solved by a fixed-point iteration (or more advanced root finding algorithms) over the whole path or trajectory

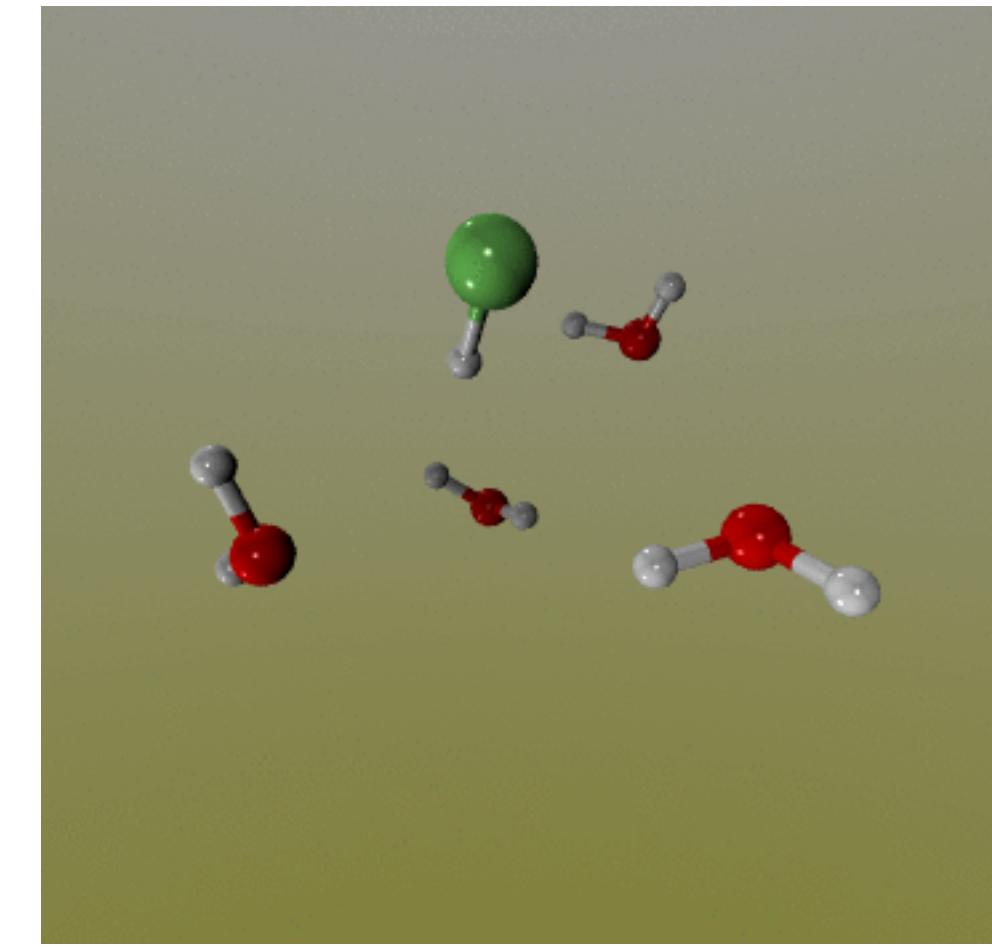
$$X^{(k+1)} = X^{(k)} - F(X^{(k)}) \quad \text{or}$$

Parallelized by distributing work  
over rows

$$\begin{bmatrix} x_1^{(k+1)} \\ x_2^{(k+1)} \\ x_3^{(k+1)} \\ x_4^{(k+1)} \end{bmatrix} = \begin{bmatrix} x_1^{(k)} \\ x_2^{(k)} \\ x_3^{(k)} \\ x_4^{(k)} \end{bmatrix} - \begin{bmatrix} x_1^{(k)} - f(x_0) \\ x_2^{(k)} - f(x_1^{(k)}) \\ x_3^{(k)} - f(x_2^{(k)}) \\ x_4^{(k)} - f(x_3^{(k)}) \end{bmatrix}$$

# Real Example: HCl+4H<sub>2</sub>O MP2 AIMD Simulations

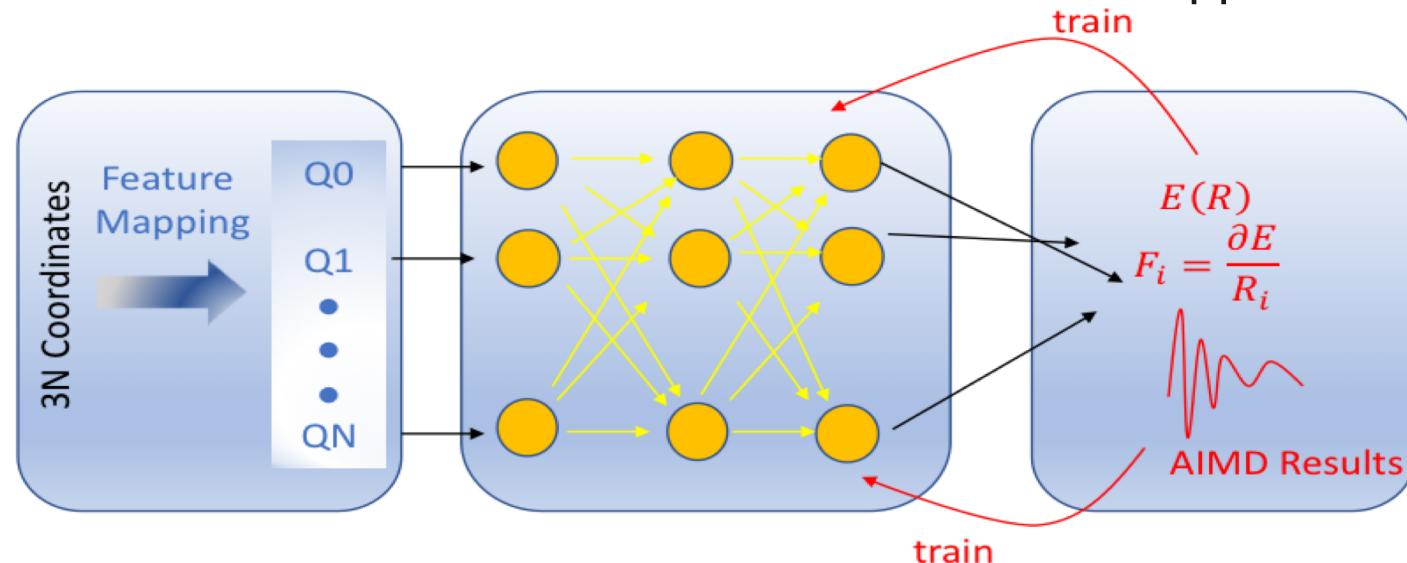
Algorithm	Method	Preconditioner	M	$K_{average}^a$	Speedup	(M/ $K_{average}$ )
QN	MP2/6-31G*		10	2.6	3.7	(3.8)
QN	MP2/6-31G*		20	3.4	5.6	(5.9)
QN	MP2/6-31G*		25	4.0	5.8	(6.3)
QN	MP2/6-31G*		50	6.0	7.7	(8.3)
QN	MP2/6-31G*		100	11.0	8.0	(9.1)
QN	MP2/6-31G*		150	17.0	8.2	(8.8)
PQN	MP2/6-31G*	HF/3-21G	10	2.0	3.7	(5.0)
PQN	MP2/6-31G*	HF/3-21G	20	3.0	4.1	(6.7)
PQN	MP2/6-31G*	HF/3-21G	25	3.0	4.6	(8.3)
PQN	MP2/6-31G*	HF/3-21G	50	4.0	4.8	(12.5)
PQN	MP2/6-31G*	HF/3-21G	100	5.0	4.8	(25.0)
PQN	MP2/6-31G*	HF/3-21G	150	5.0	5.3	(30.0)
QN	MP2/6-311+G*		10	2.5	3.3	(4.0)
QN	MP2/6-311+G*		20	3.5	4.9	(5.7)
QN	MP2/6-311+G*		25	4.0	5.9	(6.3)
QN	MP2/6-311+G*		50	6.0	7.7	(8.3)
QN	MP2/6-311+G*		100	11.0	8.2	(9.1)
QN	MP2/6-311+G*		150	17.0	7.9	(8.8)
PQN	MP2/6-311+G*	HF/3-21G	10	2.0	3.9	(5.0)
PQN	MP2/6-311+G*	HF/3-21G	20	3.0	5.3	(6.7)
PQN	MP2/6-311+G*	HF/3-21G	25	3.0	6.2	(8.3)
PQN	MP2/6-311+G*	HF/3-21G	50	3.5	10.2	(14.3)
PQN	MP2/6-311+G*	HF/3-21G	100	5.0	14.1	(25.0)
PQN	MP2/6-311+G*	HF/3-21G	150	5.0	14.3	(30.0)



Ideal Speedups of 30 seen, and FAS methods show further promise. However, preconditioners really help!

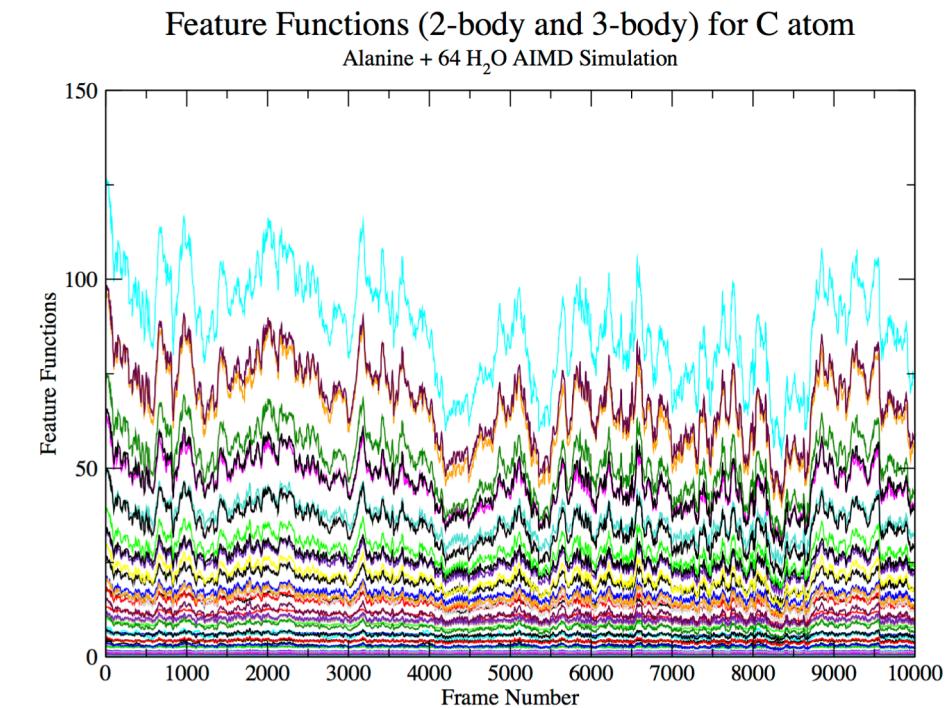
# New machine learning strategies for improving spectroscopic analysis using atomistic simulations

- This development proposes to speed up the AIMD/spectroscopy analysis by generating and using machine-learned atomistic potentials on the fly to speed up the sampling used in spectroscopic analysis, while maintaining the accuracy of the full AIMD analysis. In addition, using machine learning to regress AIMD into effective molecular dynamics potentials has the potential to enable quantum dynamics approaches for isotope fractionation that are able to overcome the limits of inherent structure approaches.

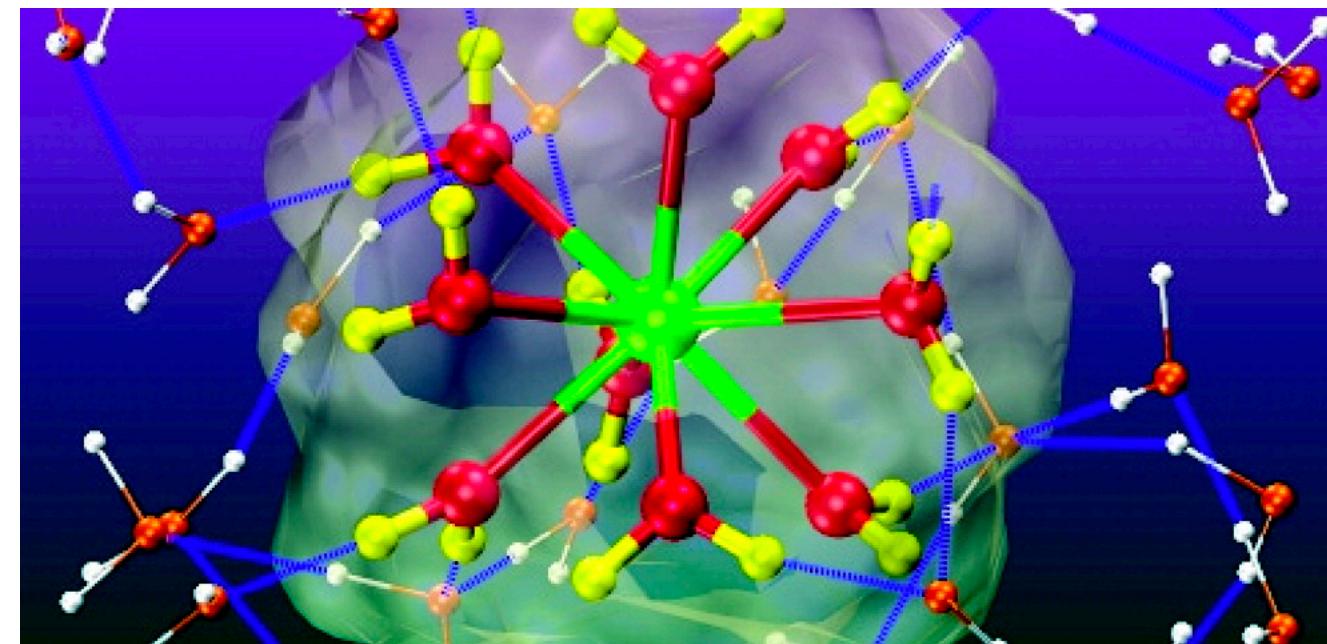


In these approaches, 2-body and 3-body functions are used to define the feature mapping or basis that is input into the feed-forward neural network. These functions are similar to 2- and 3-body molecular dynamics potentials but with varying parameters, e.g.,

$$q_i^{(\eta_\alpha)} = \sum_{\substack{j=1 \\ i \neq j}}^{\text{localatoms}} e^{-\eta_\alpha(R_{ij}-R_s)^2} f_c(R_{ij}) \quad q_i^{(\eta_\alpha)} = 2^{1-\xi} \sum_{\substack{j,k=1 \\ i \neq j, i \neq k, j \neq k}}^{\text{localatoms}} (1 - \lambda \cos \theta_{ijk})^\xi e^{-\eta_\alpha(R_{ij}^2 + R_{ik}^2 + R_{jk}^2)} f_c(R_{ij}) f_c(R_{ik}) f_c(R_{jk}).$$



# Traditional Fitting of MD Potentials

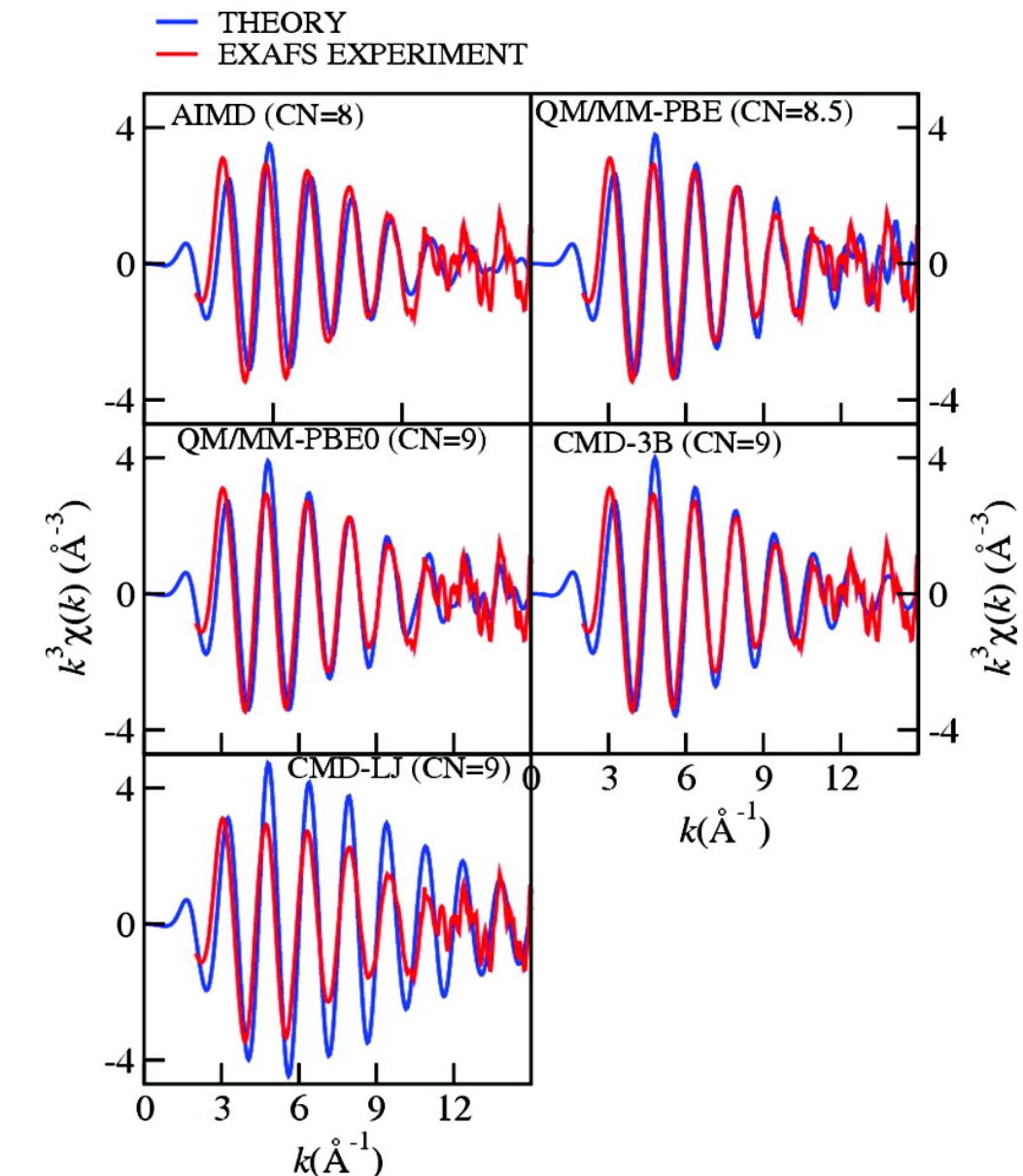


Postulate a potential form

$$U_{2B}(\text{Cm}^{3+} - \text{H}_2\text{O}) = \sum_{\alpha=\text{O}, \text{H}1, \text{H}2} \left[ A_{\text{Cm}\alpha} \exp(-B_{\text{Cm}\alpha} r_{\text{Cm}\alpha}) + \frac{C_{\text{Cm}\alpha}}{r_{\text{Cm}\alpha}^4} + \frac{D_{\text{Cm}\alpha}}{r_{\text{Cm}\alpha}^6} \right]$$

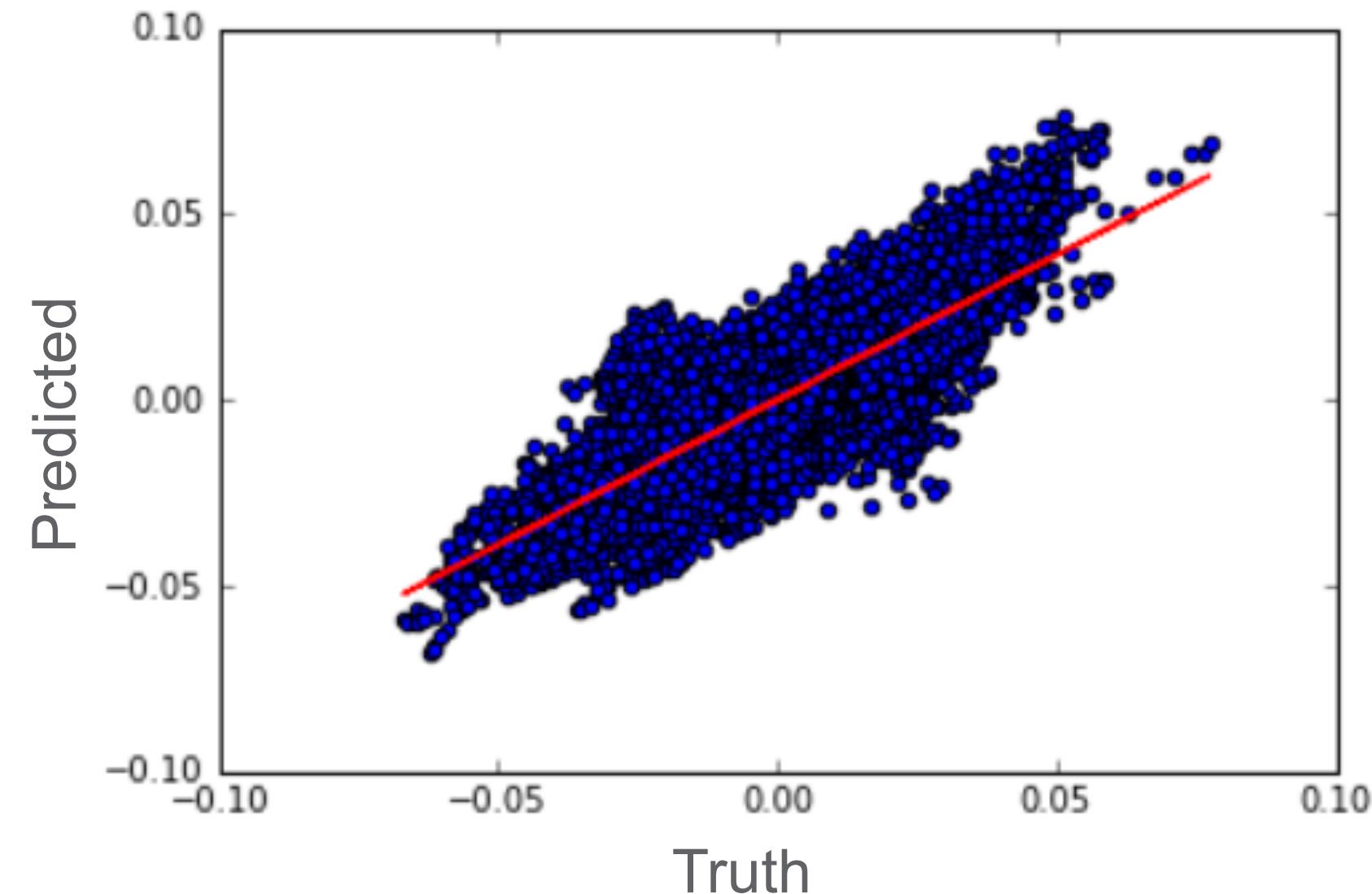
$$U_{3B}(\text{O} - \text{Cm}^{3+} - \text{O}) = \alpha \exp(-\beta r_1 - \beta r_2 - \gamma r_3)$$

And use non-linear regression (mrqmin) of simulation/experimental data to find parameters



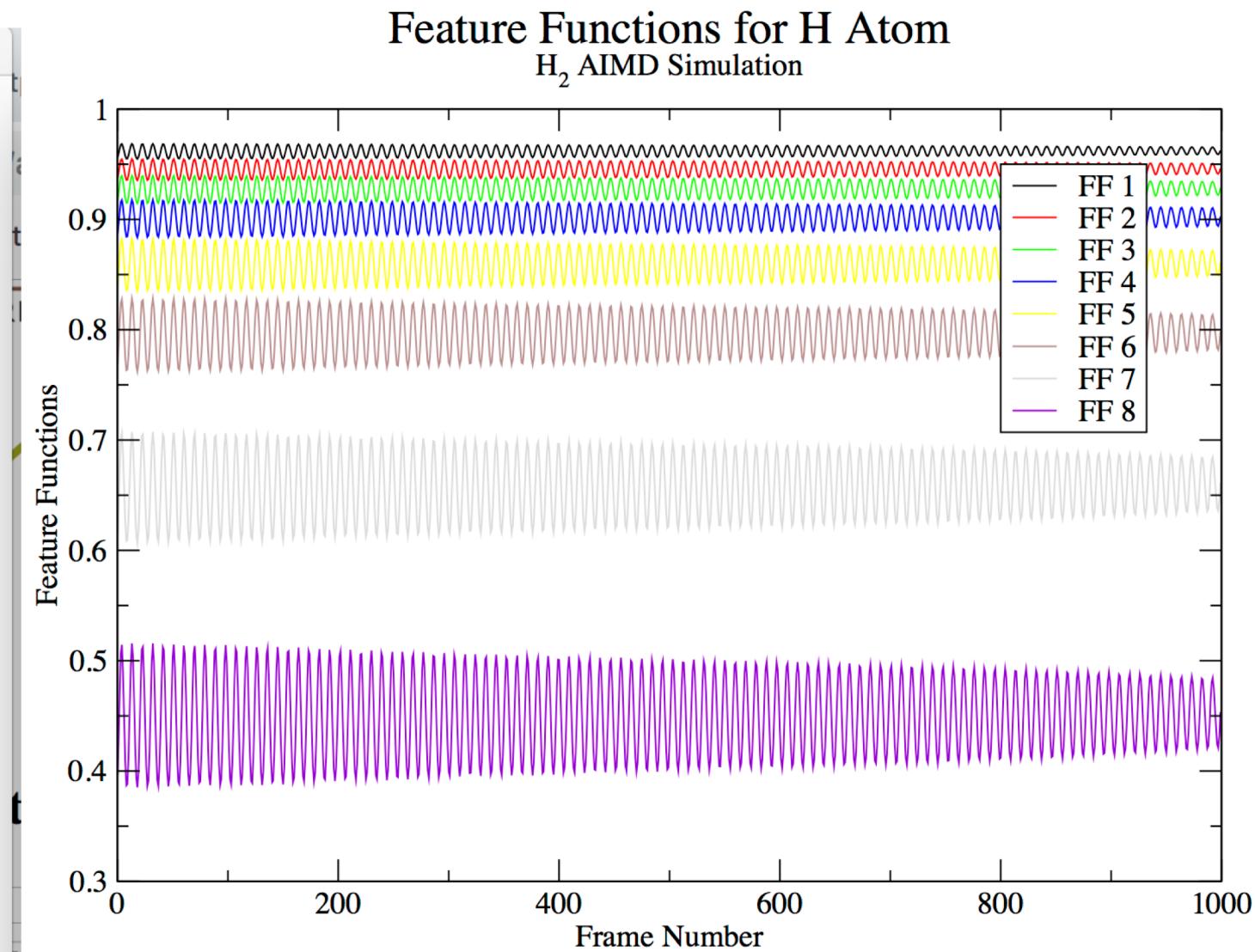
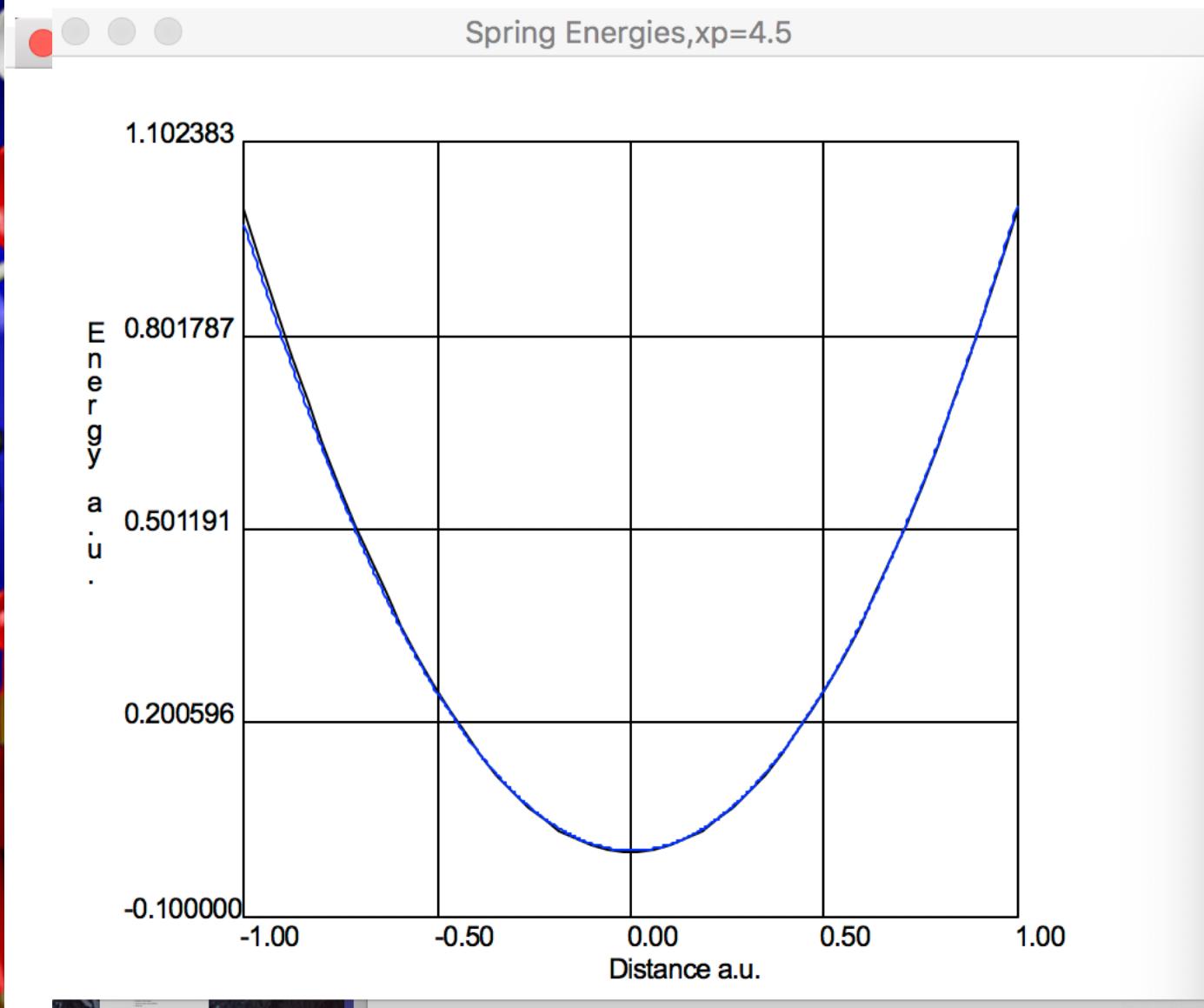
# First Attempt – Fitting forces with TensorFlow

We were able to reduce the mean absolute error w down to 0.006, which represents roughly 5% of the range of the force. However, this error is too large and , additional research into choosing feature functions as well as the use of longer AIMD simulations is required for this method to be truly competitive.



# Back to the Basics

- $1 \times 20 \times 40 \times 1$  – Adam solver

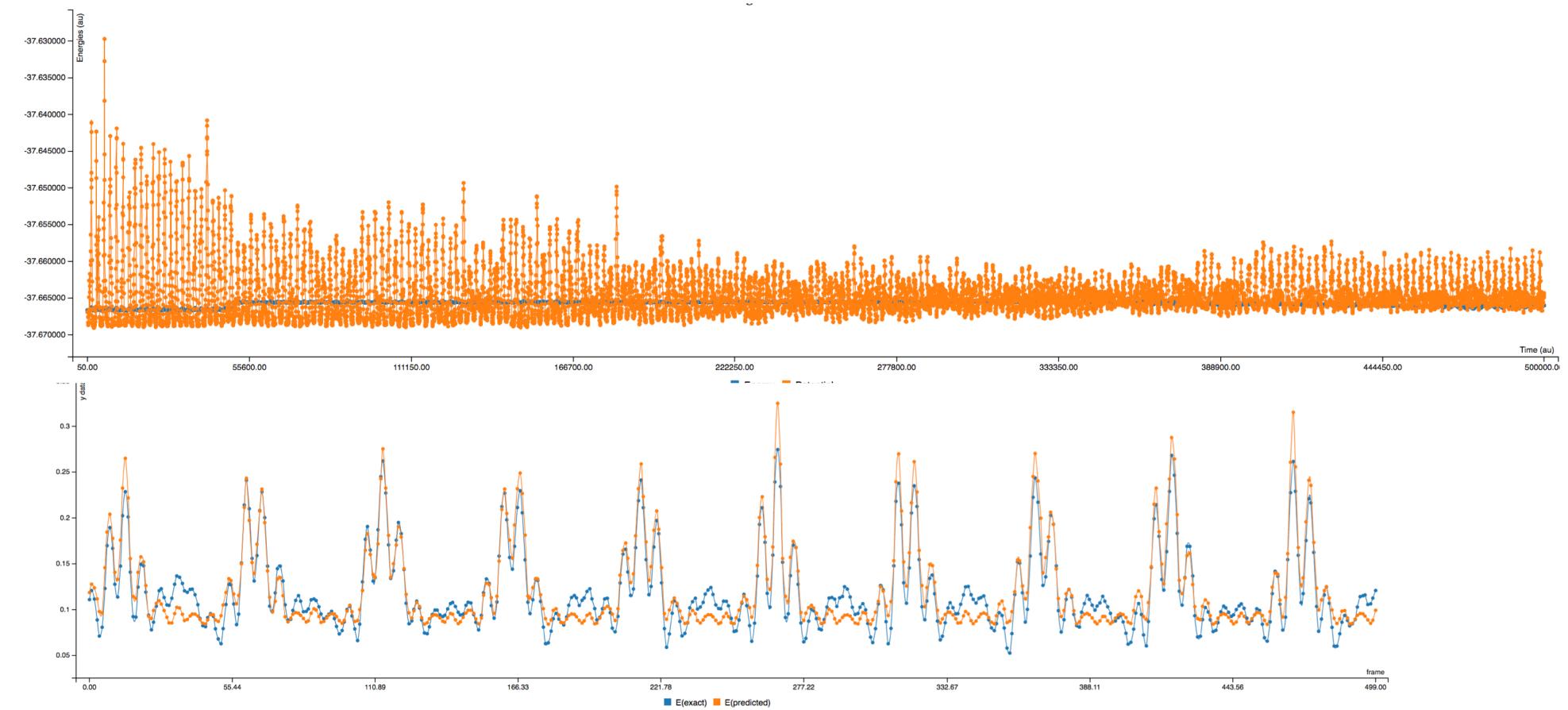
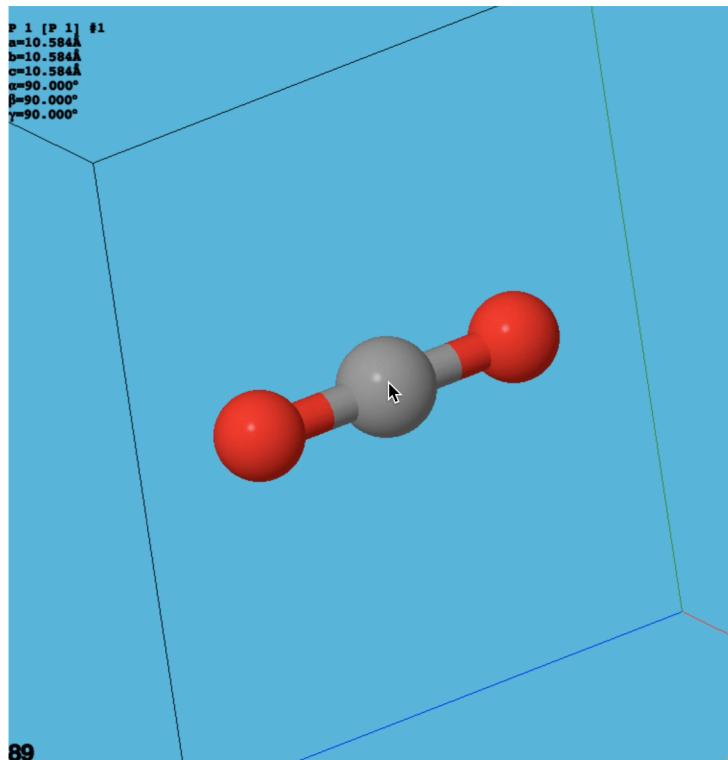




Pacific  
Northwest  
NATIONAL LABORATORY

# ML Potential for CO<sub>2</sub> Molecule

- 3x60x120x1
- Described by a 3d space
- 10,000 points  $\approx 21^3$



While the ML approach looks feasible and automatable there are several challenges going forward

- High-dimensional spaces
- Charges, dipoles, polarization, bond breaking, ...
- More model input will probably be needed for this approach to be predictive
- Start with fitted MD potentials and correct with ML



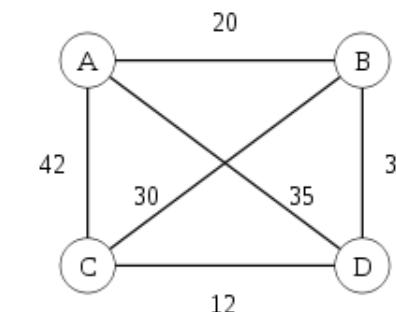
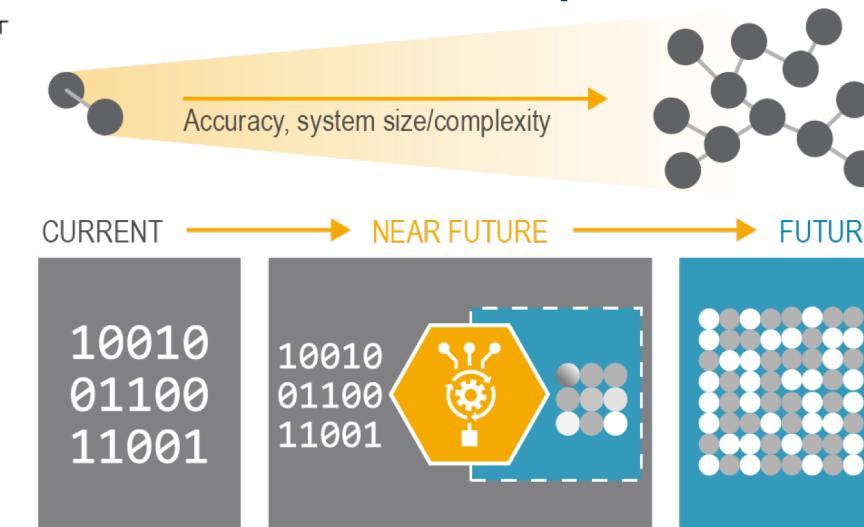
Pacific  
Northwest  
NATIONAL LABORATORY

# Computing Horizons – The Next Wave

- Classical computing is becoming power bound
- All DOE supercomputers are going to have more GPUs than they know what to do with.....
- Planning for many cycles to be used for ML



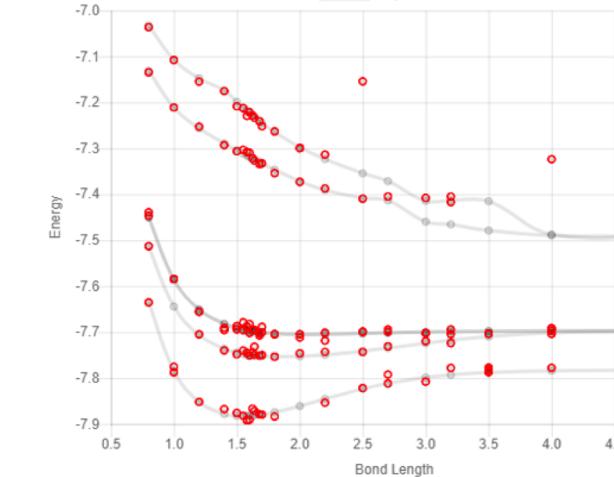
Exact quantum chemistry not computable on classical computers



$$e^{-iHt} = \prod_{j=1}^M e^{-ih_j P_j t} + O(M^2 \max_{j,k} \| [P_j, P_k] \| \max |h_j|^2 t^2).$$

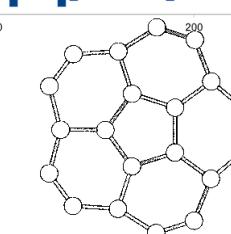
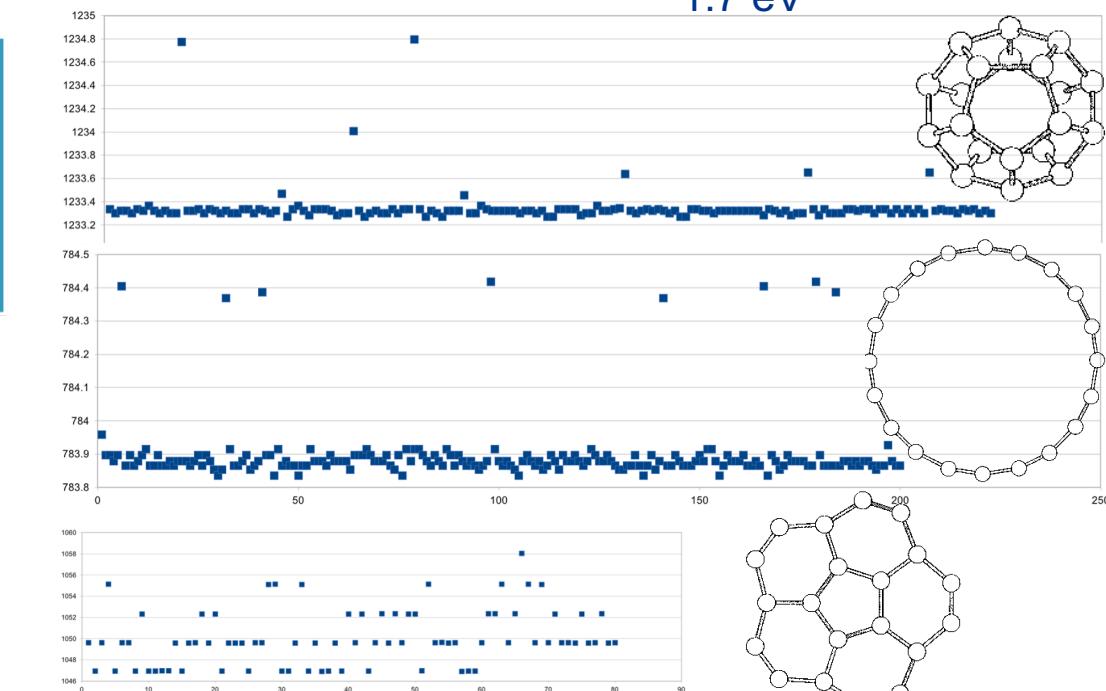
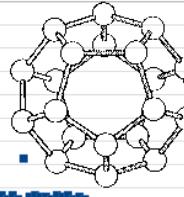
LiH Simulation Demo

Estimated |G| |E1| |E2| |E3| |E4| |E5|



The smaller the molecule in a published paper the greater its importance

- Trotterization predicts the fullerene lower than the ring by 1.75 eV
- Taylor, Bylaska, Kawai, Weare – CCSD(T) early 1990s, Fullerene lower than ring by 1.7 eV



# Conclusions

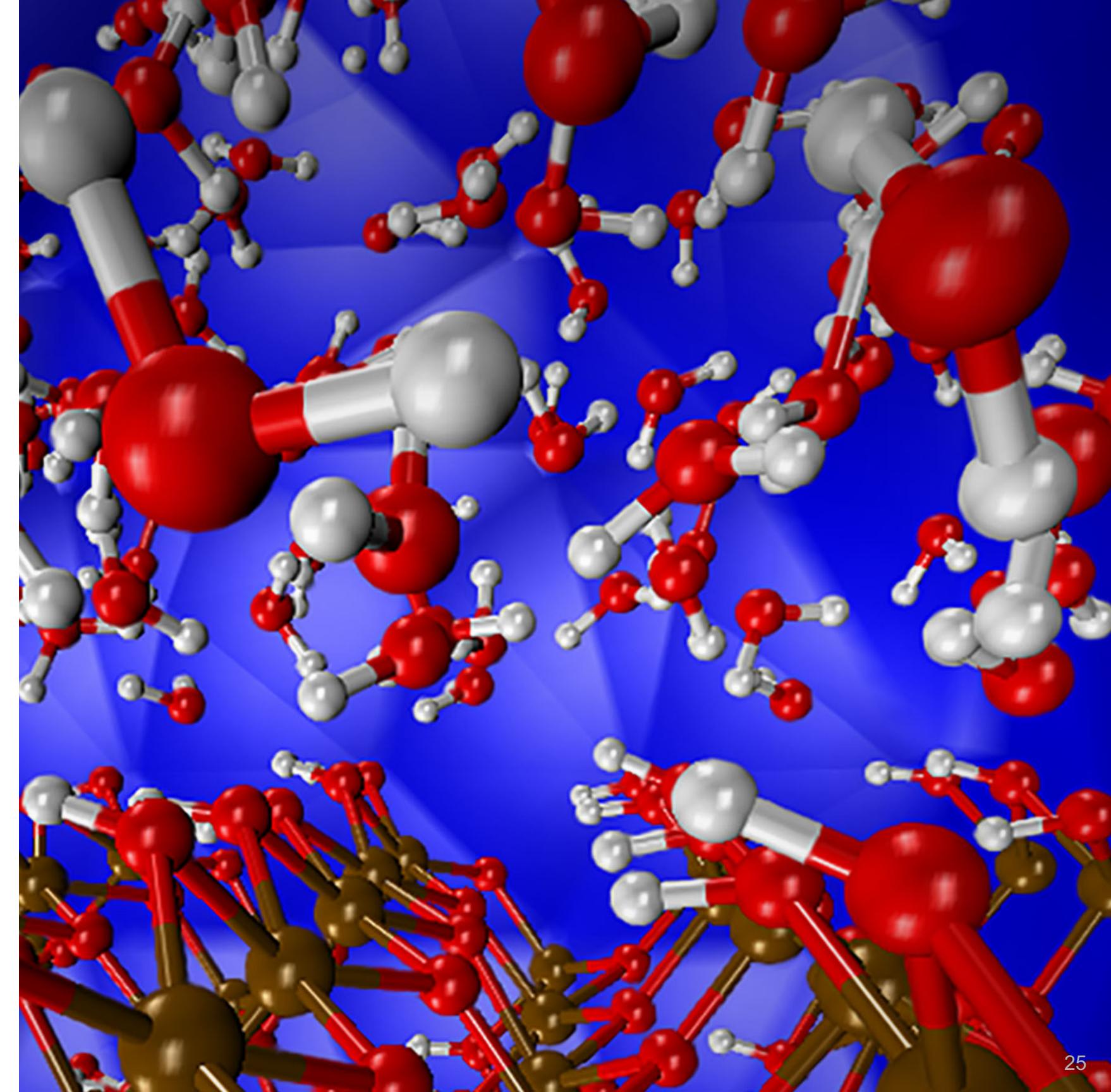
- Computation chemistry methods are becoming truly predictive, rather than just rationalizing existing knowledge. Synergistic use of AIMD and spectroscopies is already changing many spectroscopies.
- New machine-learning methods for developing MD potentials will support longer dynamical simulations and improved phase sampling methods, which will provide new models of chemical mechanisms in complex brines, defected solids and interfaces.
  - Inverse modeling expertise from the chemistry and condensed matter communities needs to be better incorporated.
- All these developments will be available to the wider geochemistry, chemistry, and materials communities via inclusion in the NWChem (NWChemEx) program or the EMSL Arrows scientific service.

**Acknowledgements** This work was primarily supported by the DOE BES Geosciences program. Also thanks to DOE BER EMSL, DOE ECP, DOE BES Chemistry. A portion of this research was performed using the NERSC Computing Facility at LBL. EMSL is a national scientific user facility sponsored by the Department of Energy's Office of Biological and Environmental Research located at Pacific Northwest National Laboratory, DE-AC06-76RLO 1830. We also acknowledge EMSL for supporting the development of NWChem. The Pacific Northwest National Laboratory is operated by Battelle Memorial Institute.



Pacific  
Northwest  
NATIONAL LABORATORY

Thank you





# Is Fixed Point Iteration Stable?

Solving  $X^{(k+1)} = X^{(k)} - F(X^{(k)})$

**Guess the Initial Path**

$$X^{(1)} = \begin{bmatrix} b \\ c \\ d \\ e \end{bmatrix}$$



# Is Fixed Point Iteration Stable?

Solving  $X^{(k+1)} = X^{(k)} - F(X^{(k)})$

**k=1**

$$\begin{bmatrix} f(x_0) \\ f(b) \\ f(c) \\ f(d) \end{bmatrix} = \begin{bmatrix} b \\ c \\ d \\ e \end{bmatrix} - \begin{bmatrix} b - f(x_0) \\ c - f(b) \\ d - f(c) \\ e - f(d) \end{bmatrix}$$

Each step of this global iteration can be parallelized by evaluating each row on a different cpu



# Is Fixed Point Iteration Stable?

Solving  $X^{(k+1)} = X^{(k)} - F(X^{(k)})$

**k=2**

$$\begin{bmatrix} f(x_0) \\ f(f(x_0)) \\ f(f(b)) \\ f(f(c)) \end{bmatrix} = \begin{bmatrix} f(x_0) \\ f(b) \\ f(c) \\ f(d) \end{bmatrix} - \begin{bmatrix} f(x_0) - f(x_0) \\ f(b) - f(f(x_0)) \\ f(c) - f(f(b)) \\ f(d) - f(f(c)) \end{bmatrix}$$

Each step of this global iteration can be parallelized by evaluating each row on a different cpu



# Is Fixed Point Iteration Stable?

Solving  $X^{(k+1)} = X^{(k)} - F(X^{(k)})$

**k=3**

$$\begin{bmatrix} f(x_0) \\ f(f(x_0)) \\ f(f(f(x_0))) \\ f(f(f(b))) \end{bmatrix} = \begin{bmatrix} f(x_0) \\ f(f(x_0)) \\ f(f(b)) \\ f(f(c)) \end{bmatrix} - \begin{bmatrix} f(x_0) - f(x_0) \\ f(f(x_0)) - f(f(x_0)) \\ f(f(b)) - f(f(f(x_0))) \\ f(f(c)) - f(f(f(b))) \end{bmatrix}$$

Each step of this global iteration can be parallelized by evaluating each row on a different cpu



# Is Fixed Point Iteration Fast?

Solving  $X^{(k+1)} = X^{(k)} - F(X^{(k)})$

**k=4**

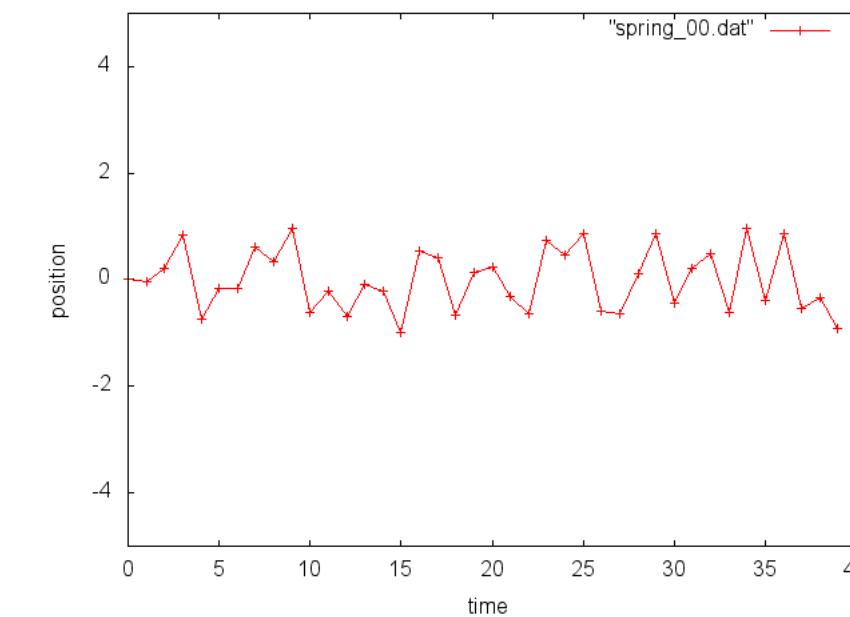
$$\begin{bmatrix} f(x_0) \\ f(f(x_0)) \\ f(f(f(x_0))) \\ f(f(f(f(x_0)))) \end{bmatrix} = \begin{bmatrix} f(x_0) \\ f(f(x_0)) \\ f(f(f(x_0))) \\ f(f(f(f(b)))) \end{bmatrix} - \begin{bmatrix} f(x_0) - f(x_0) \\ f(f(x_0)) - f(f(x_0)) \\ f(f(f(x_0))) - f(f(f(x_0))) \\ f(f(f(f(b)))) - f(f(f(f(x_0)))) \end{bmatrix}$$

Exact solution in 4 steps, but no speedup!

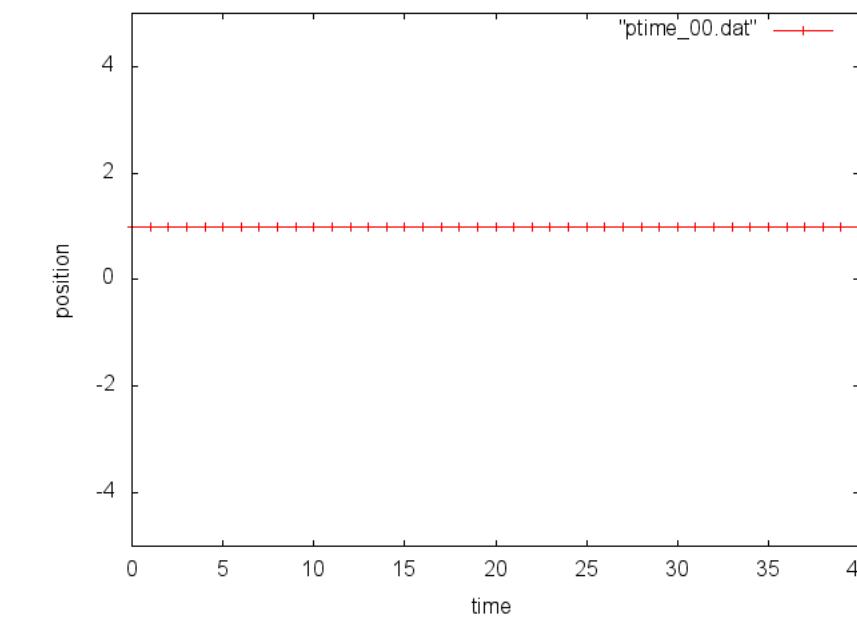
# Improving Fixed Point Parallel in Time Algorithm

- For speedup to occur the root finding iterations needs to converge in less than  $M$  iterations, where  $M$  is the path length!
- Our strategy is to use a variety of optimization techniques, including preconditioned fixed-point, quasi-Newton, and preconditioned quasi-Newton optimization methods to obtain speedup

Fixed Point



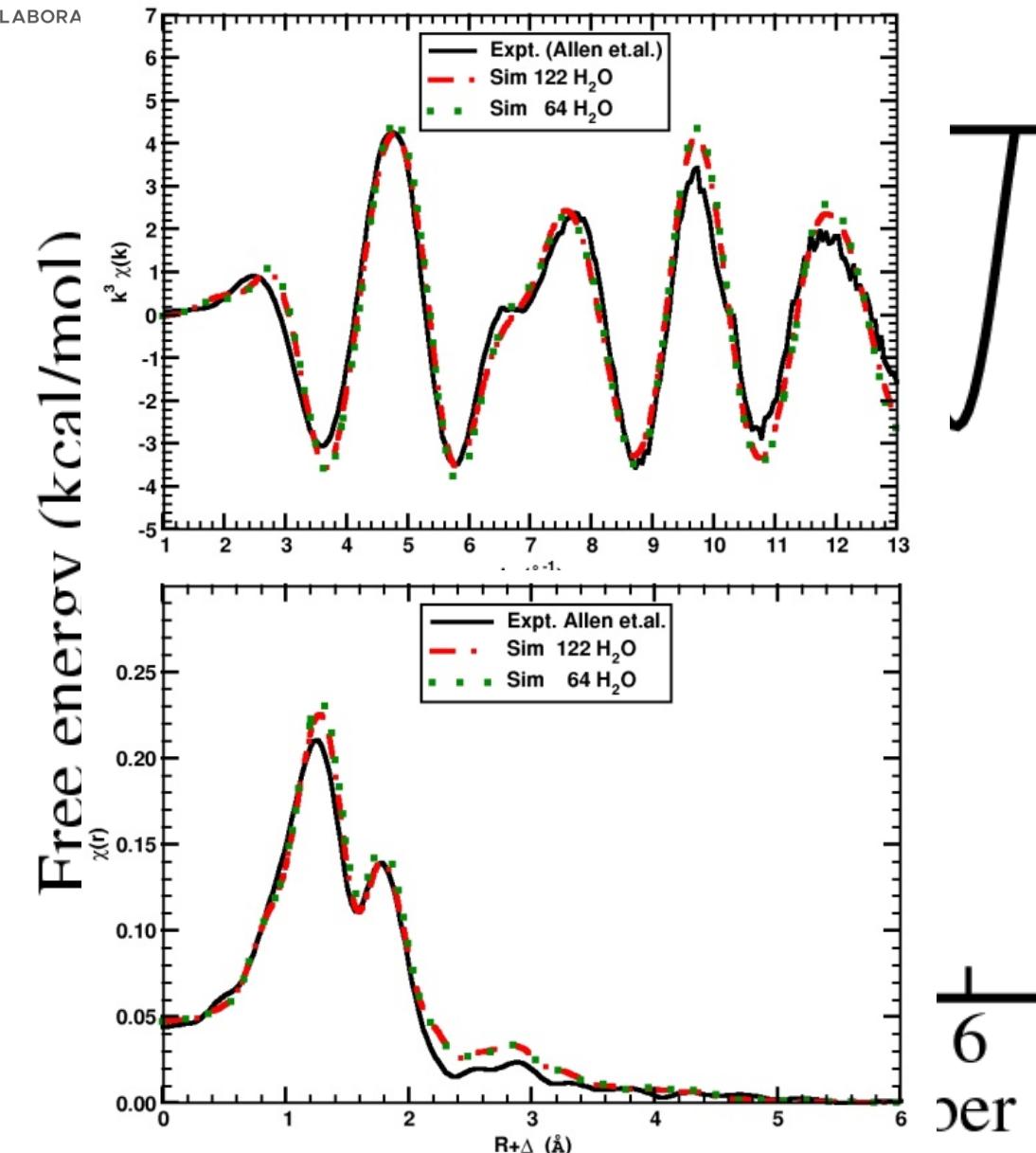
Quasi-Newton



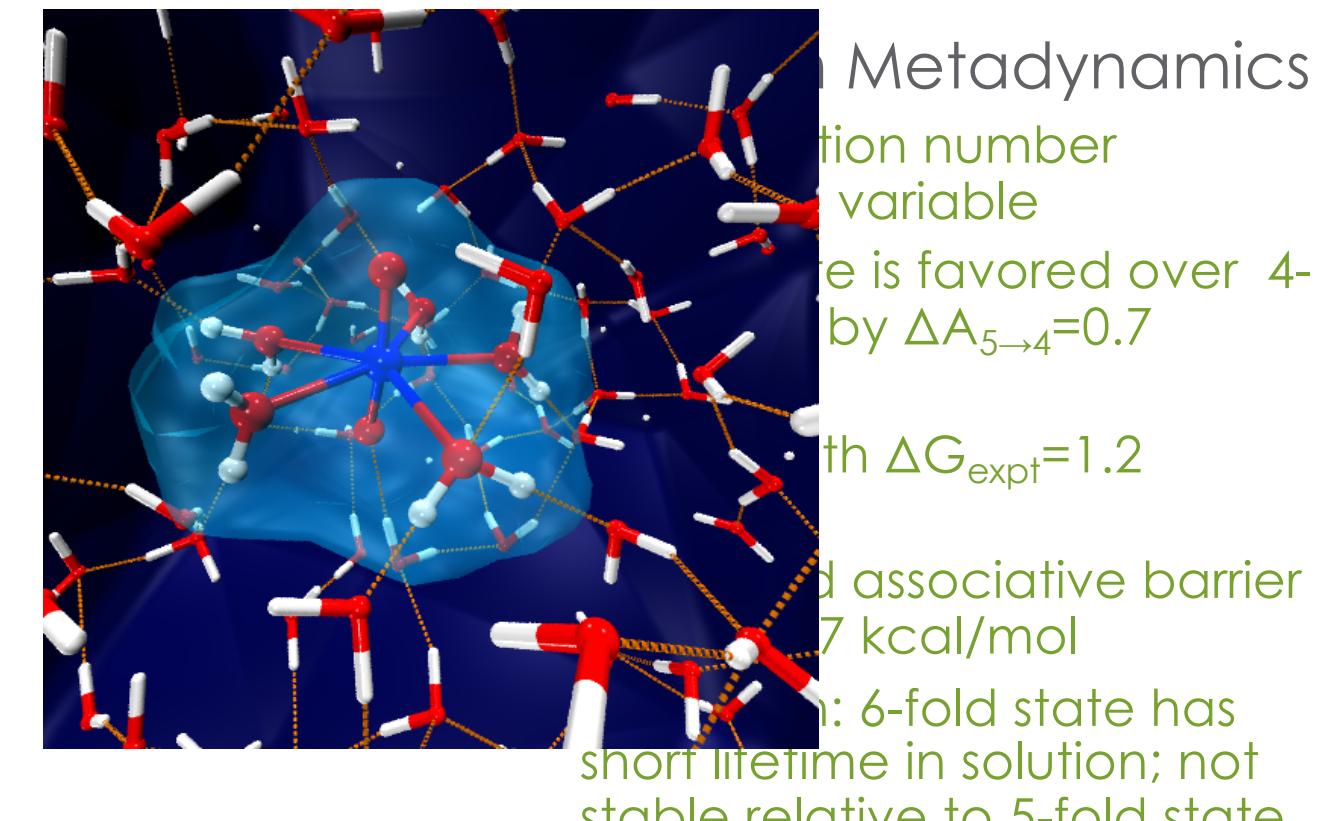
# Exchange Free Energies and Barriers in

~~UO<sub>2</sub><sup>2+</sup>(aq)~~

Pacific  
Northwest  
NATIONAL LABORATORY



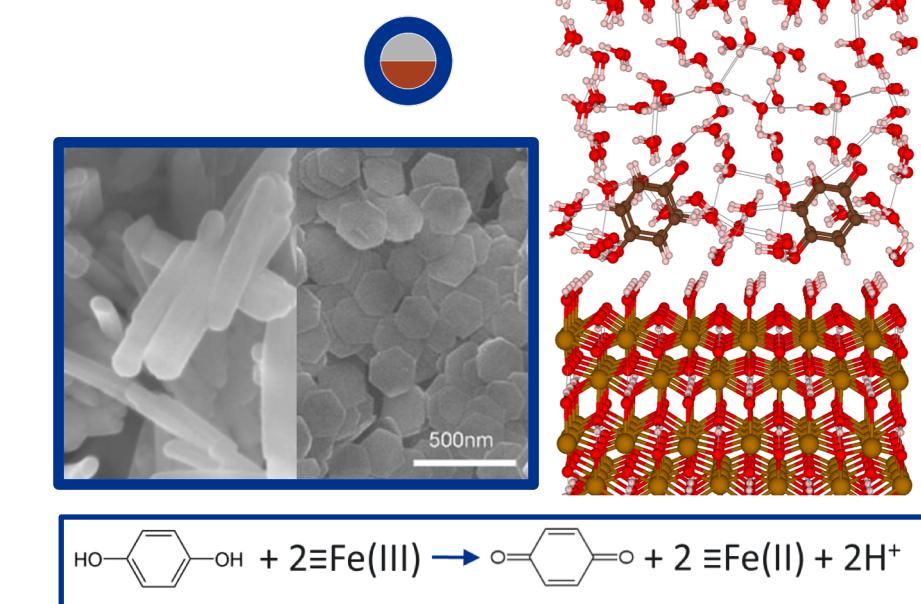
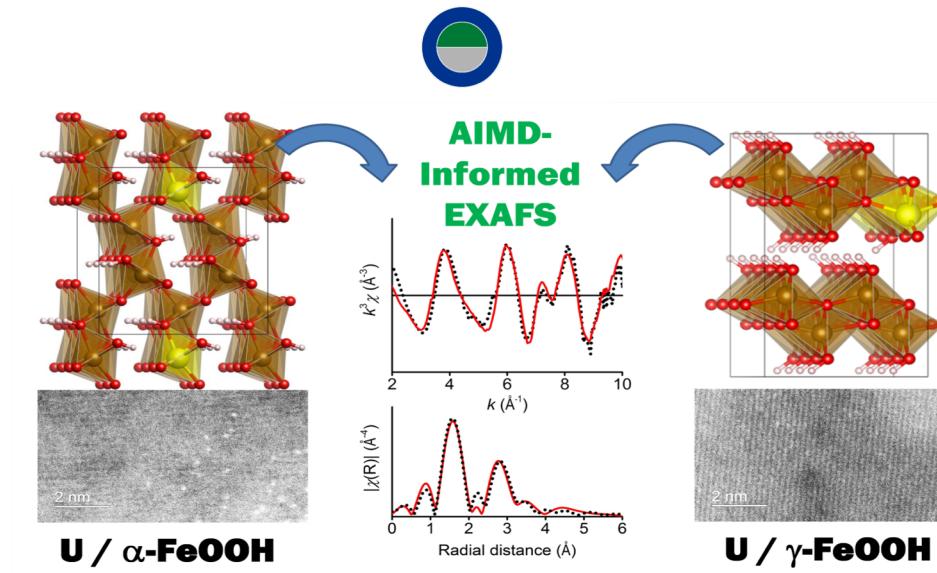
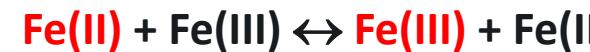
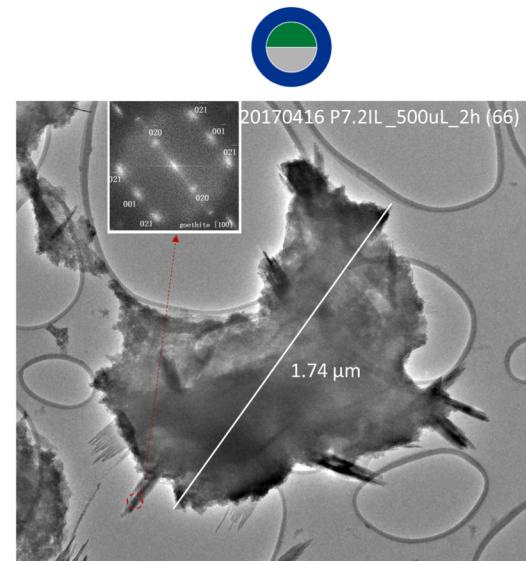
- Good EXAFS agreement but recent HEXS experiments suggest a 4-fold state is energetically nearby



- $\Delta A_{5 \rightarrow 6} = 8$  kcal/mol;  $\Delta A_{5 \rightarrow 6}^\ddagger \approx 9$  kcal/mol

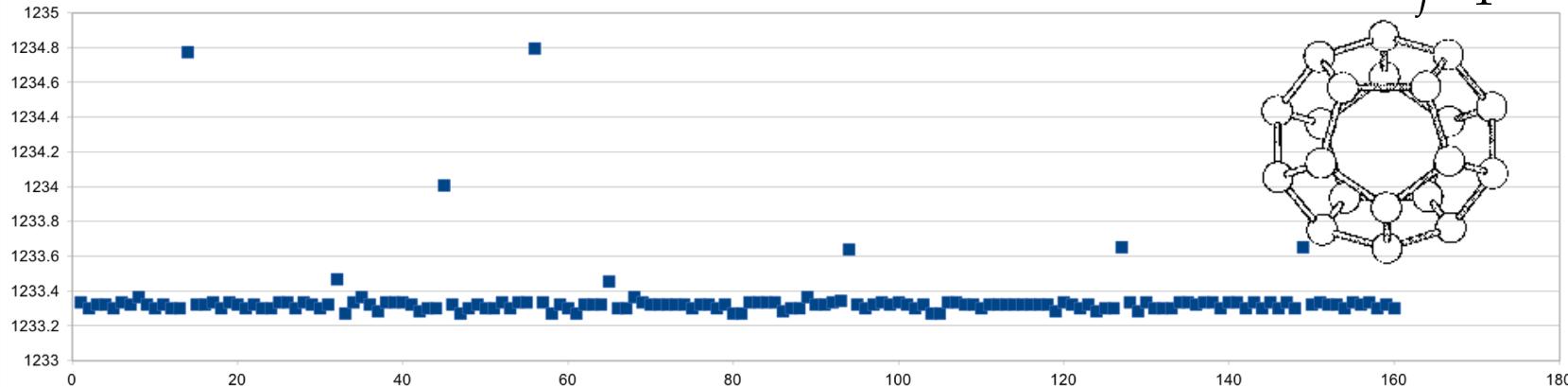
# Background and significance

- Previous work on this project has focused on developing molecular modeling methods for complex solutions
- Moving towards mineral interactions with reservoir fluids
  - Electron transfer modeling in condensed phase system
  - Sorption of species and reactions at complex geochemical interfaces
  - Direct simulation methods being developed to more fully exploit the structural sensitivity accessible in new surface experimental tools. Increasing the level of quantitative detail provided by these spectroscopies (surface/tip-enhanced Raman spectroscopies, XAS, Mössbauer)
  - Examples of specific simulations have been discussed in the subtasks of Rosso/Scherer (ST1), Ilton (ST2), and Loring (ST5).



# Future Quantum Computing Applications – Microsoft QDK

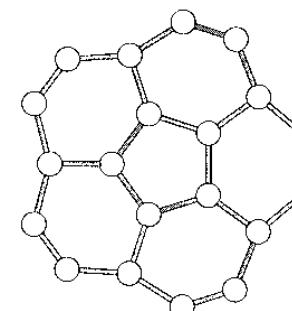
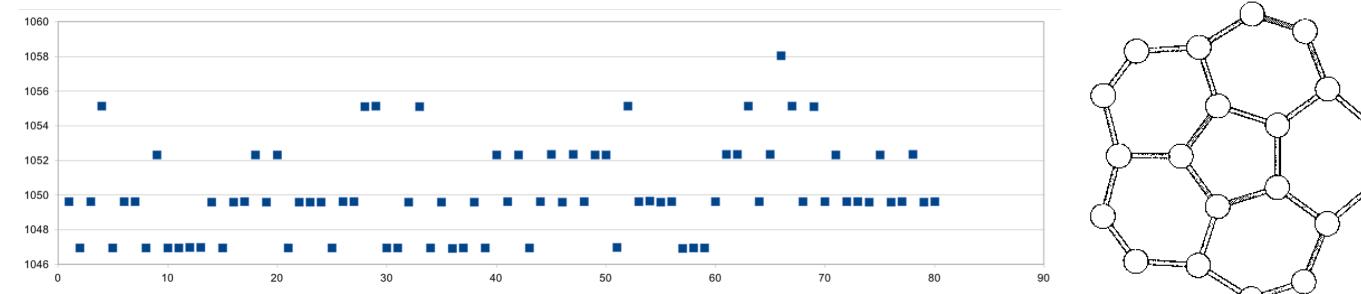
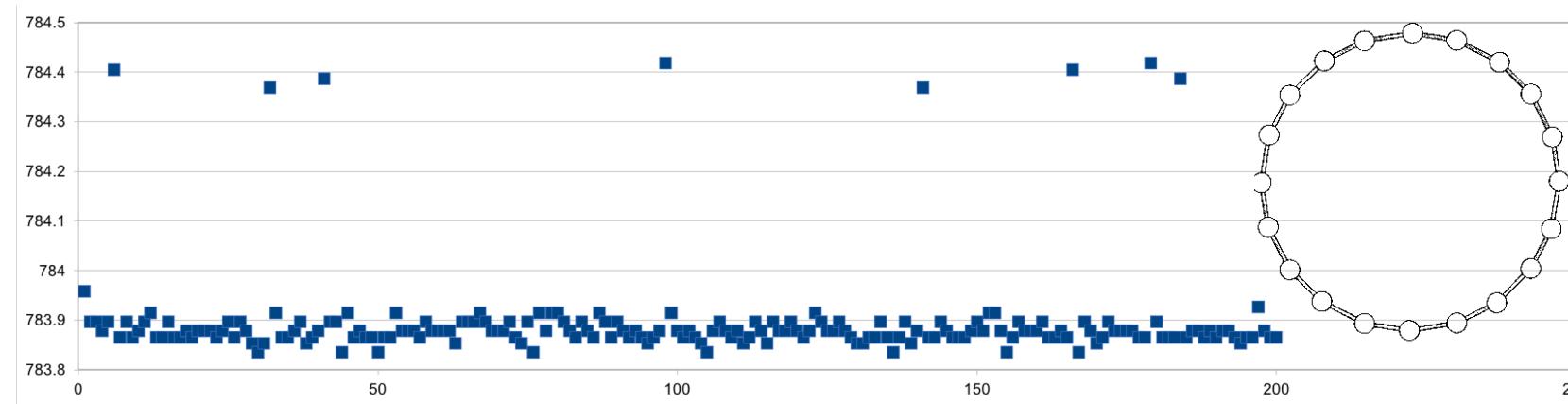
## Trotterization – Cannon's Phase Estimation



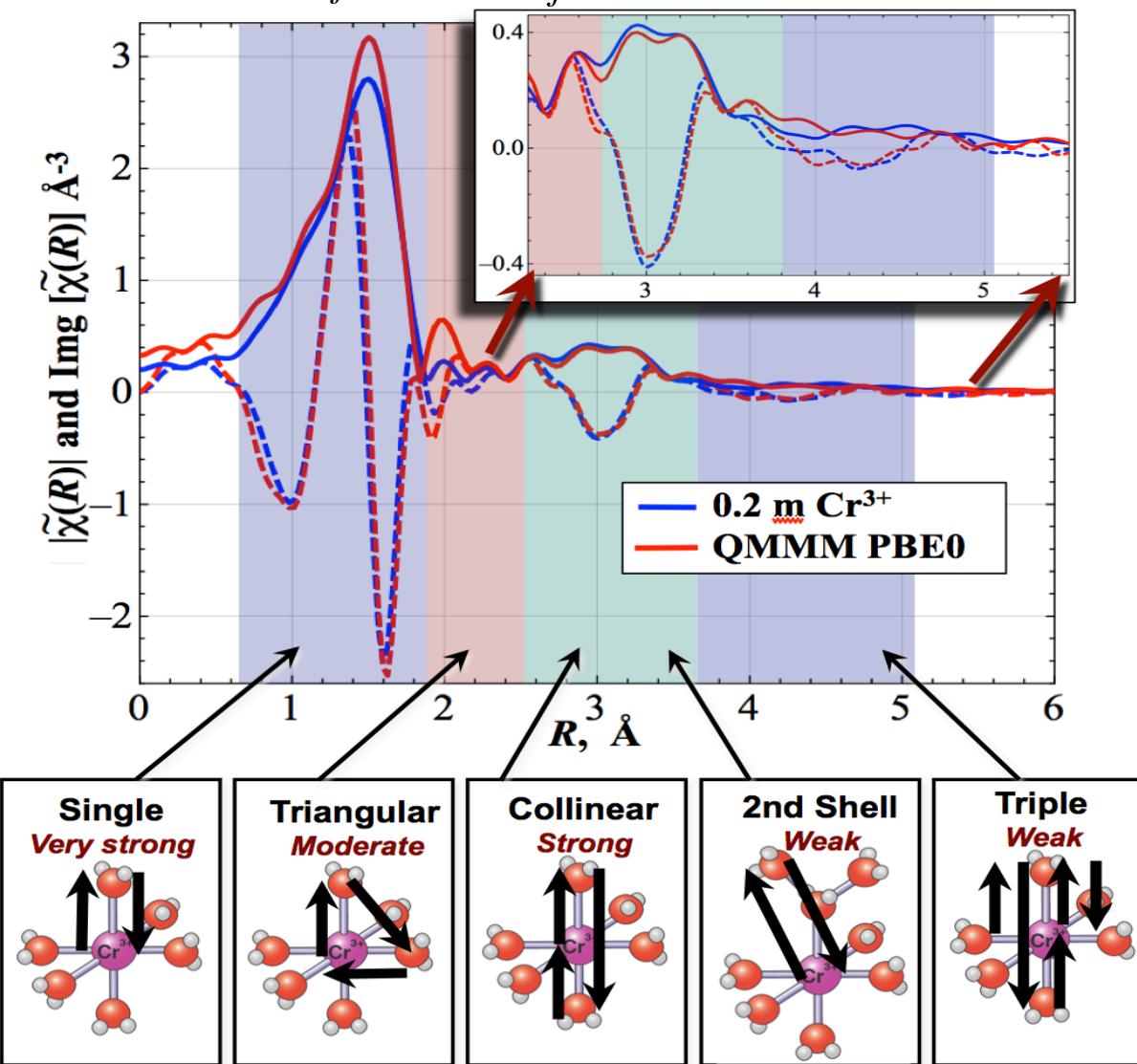
$$e^{-iHt} = \prod_{j=1}^M e^{-ih_j P_j t} + O(M^2 \max_{j,k} \|[P_j, P_k]\| \max |h_j|^2 t^2).$$

The idea behind Trotter–Suzuki formulas is simple: express the Hamiltonian as a sum of easy to simulate Hamiltonians and then approximate the total evolution as a sequence of these simpler evolutions

- Hartree-Fock predicts the ring lower than fullerene by 2.2 eV
- Taylor, Bylaska, Kawai, Weare – CCSD(T) early 1990s
  - Fullerene lower than ring by 1.7 eV
- Trotterization predicts the fullerene lower than the ring by 1.75 eV
  - 8 orbitals (4 occupied, 4 virtual)



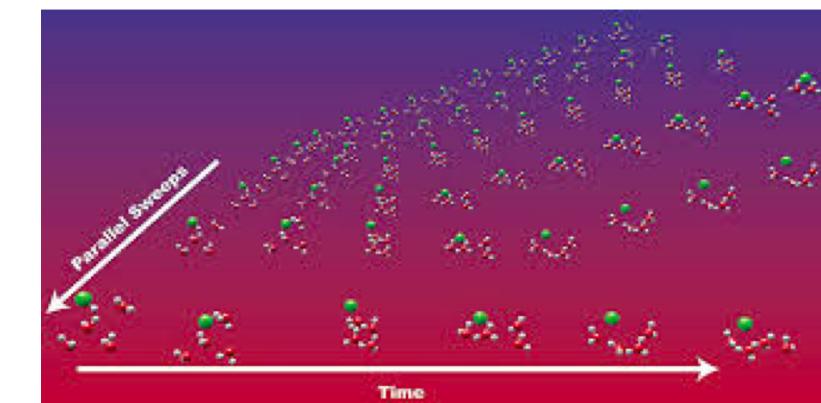
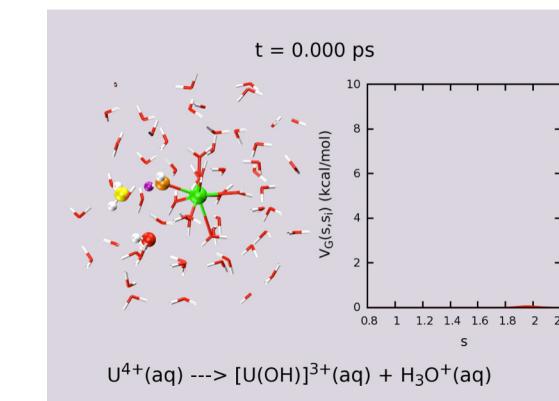
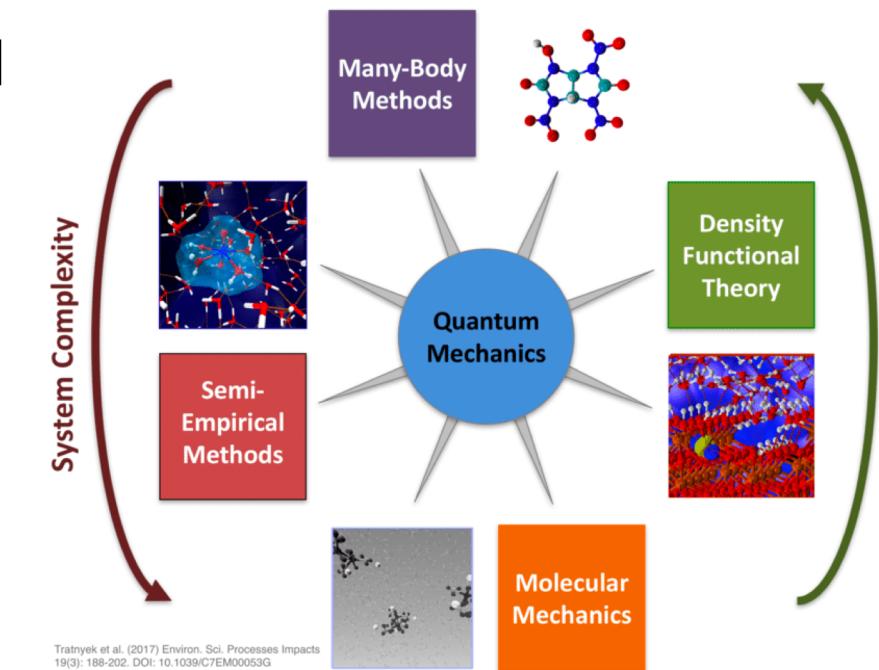
$$\chi(k) = \frac{1}{N} \sum_i \sum_j^{struc.\ paths} S_0^2 \frac{N_j}{kR_{ij}^2} |F(k)| \sin[2R_{ij} - \varphi_j(k)] e^{-2R_{ij}/\lambda}$$



Illustrates the agreement with data obtained using the full 1<sup>st</sup> principle MD-XAFS method. That is direct simulation of spectra using ab-initio MD. All scattering paths are used. Slide illustrates the contributions from the various scattering paths. **The parameter free 1<sup>st</sup> principle simulation will very accurately reproduce both the structural parameters and disorder effects.**

# Background and significance

- The theoretical developments of this program will implement advanced computational methods to ensure that PNNL is a leader in computational geochemistry
- Focus is on first-principles development and methods for improving time-scales
  - Fast electronic structure
  - Rare event
- HPC exascale development (e.g., O(N) DFT), also development started in quantum computing
- Machine learning methods
  - Learning MD potentials
- Exploring Web 4.0-5.0 technologies
  - EMSL Arrows



# ~~AIMD and MD Still Need Faster Time To Solutions: The Irrational Pursuit Of Solving The Schrödinger Equation In less than a Second Per Step~~

Pacific Northwest  
NATIONAL LABORATORY

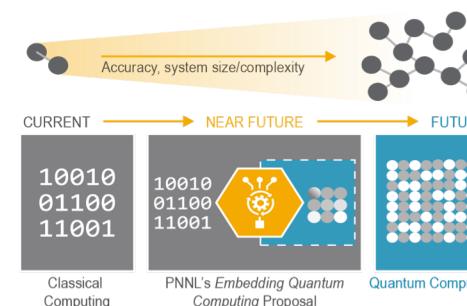
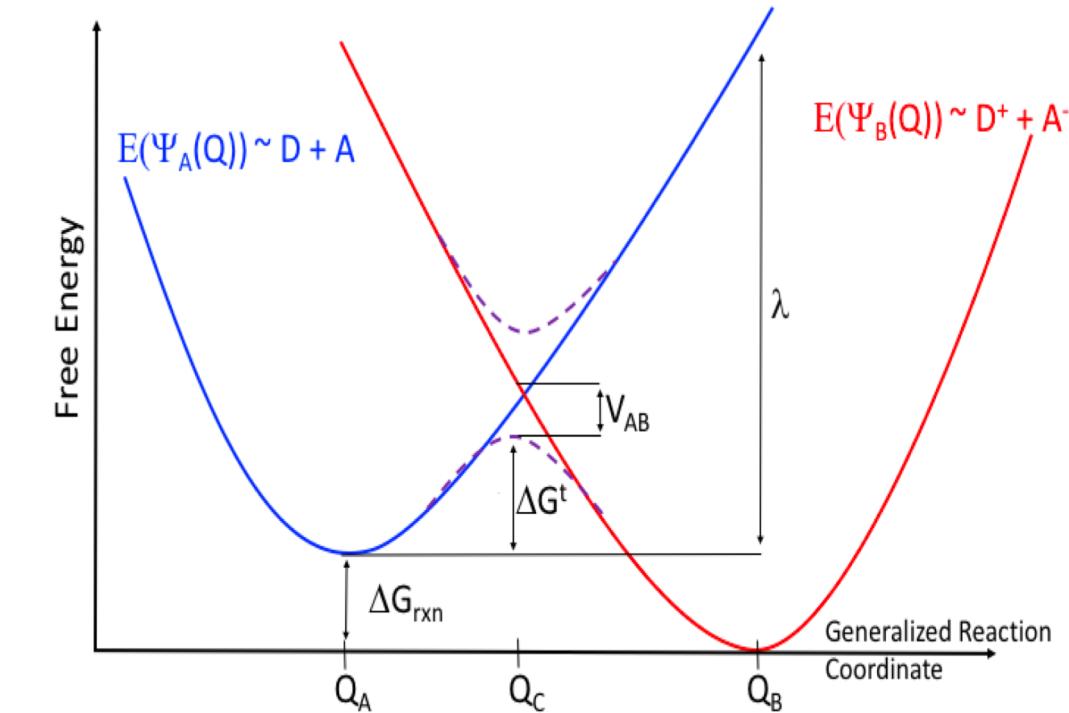
- Current *ab initio* molecular dynamics simulations for 10 to 100 picoseconds can take several months to complete
- The step length in *ab initio* molecular dynamics simulation is on the order of 0.1...0.2 fs/step
  - 20 ps of simulation time → 200,000 steps
    - At 1 second per step → 2-3 days
    - At 10 seconds per step → 23 days
    - At 30 seconds per step → 70 days
  - 1 ns of simulation time → 10,000,000 steps
    - at 1 second per step → 115 days of computing time
    - At 10 seconds per step → 3 years
    - At 30 seconds per step → 9 years
    - At 0.1 seconds per step → 11.5 days
- For classical molecular dynamics, time step  $\sim= 1\text{fs}/\text{step}$ 
  - 1 $\mu\text{s}$  of simulation time → 1,000,000,000 steps
    - 1 millisecond per step → 11.6 days of computing time
    - 1 second per step → 31 years of computing time
    - 10 seconds per step → 310 years of computing time
    - 1  $\mu\text{s}$  per step → 16.6 minutes

HCl+4H<sub>2</sub>O MP2/6-311++G(2d,2p) force evaluation takes 32 seconds!

# Objective #1: Development of beyond DFT electronic structure methods for electron transfer

The periodic ET method we have previously developed can be generalized to implement a periodic PW MCSCF and CASSCF method fairly easily. This method can be formulated as a two-step procedure:

1. Calculate the select CI expansion coefficients,  $C_I$ , by diagonalizing the select CI matrix.
2. Using the  $C_I$  coefficients, calculate the gradient with respect to the one-electron orbitals, and then update with a conjugate gradient or similar method

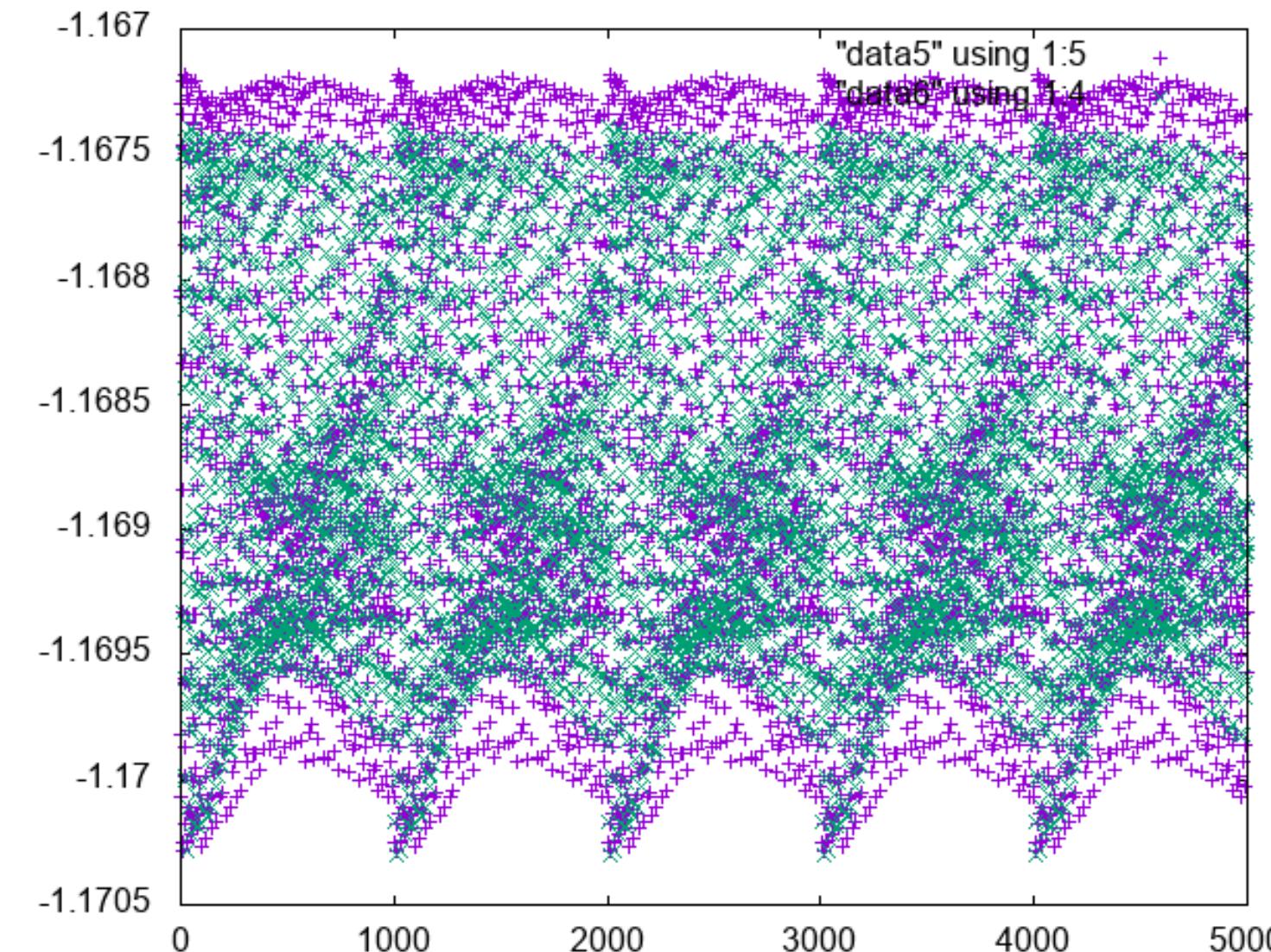


Number of configurations becomes exceedingly large as a function of the number of redox atoms, *quickly approaching a googol of configurations*

This software will support the prediction of the mechanisms of ET and the structure of defects in oxides, and will identify charge transfer mechanisms occurring at the mineral/water interface.

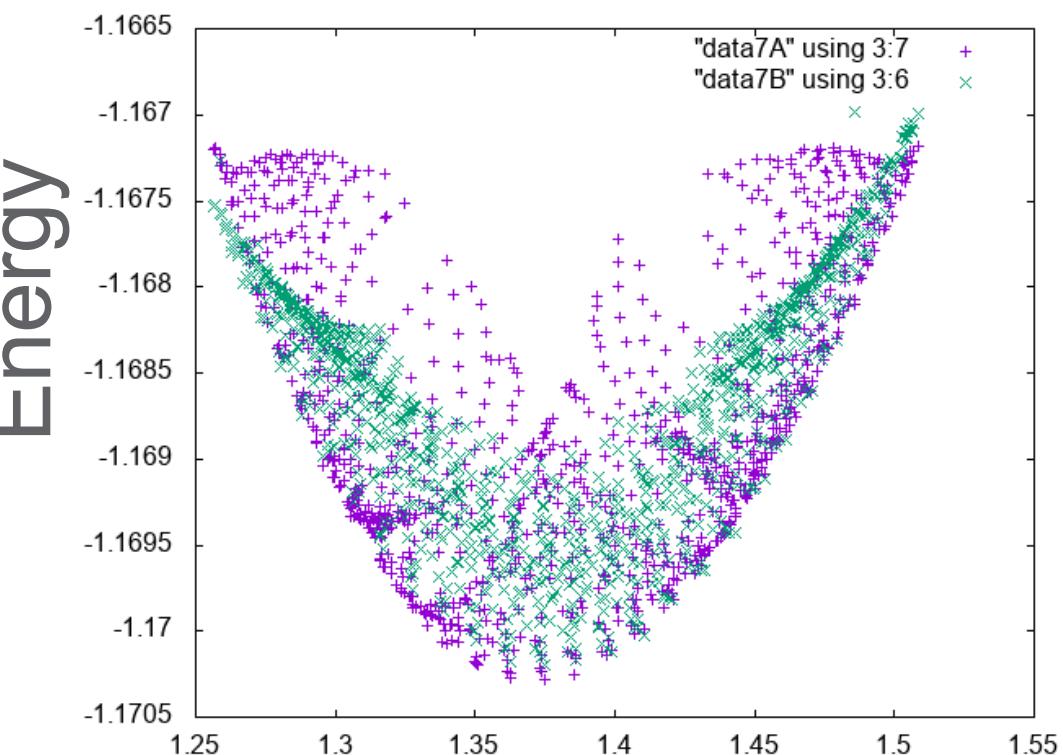
# Example of fitting trajectory

Energy



Trajectory

84 Feature functions per atom  
3 Layers



$d_{12}$

# Background and significance

- Previous work on this project has focused on developing molecular modeling methods for complex solutions
- Moving towards mineral interactions with reservoir fluids
  - Electron transfer modeling in condensed phase system
  - Sorption of species and reactions at complex geochemical interfaces
  - Direct simulation methods being developed to more fully exploit the structural sensitivity accessible in new surface experimental tools. Increasing the level of quantitative detail provided by these spectroscopies (surface/tip-enhanced Raman spectroscopies, XAS, Mössbauer)
  - Examples of specific simulations have been discussed in the subtasks of Rosso/Scherer (ST1), Ilton (ST2), and Loring (ST5).

