13th U.S. National Congress on Computational Mechanics, San Diego, July 26-30, 2015

Smart use of Density Functional Theory calculations to drive Newtonian dynamics

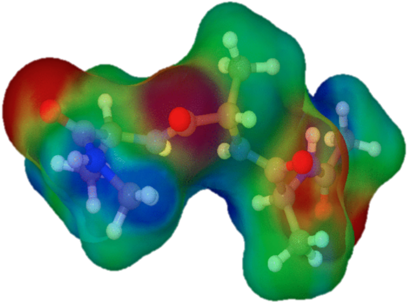
SAND2015-8905C

Concurrent Multi-length Scale Modeling: From Finite Elements to Atoms and Electrons

Monday. 4:30pm

Reese Jones and Mickey Shaughnessy

Sandia National Laboratories

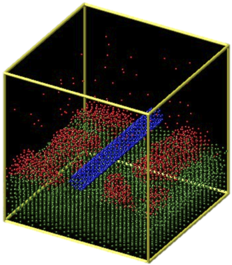


Sandia National Laboratories is a multiprogram laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

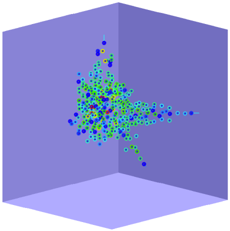
Outline

Motivation: *compositionally* and *structurally complex* materials and processes: e.g. deposition and radiation damage, require the dynamic simulation of *large* atomic systems with *multiple* interspecies interactions.

atomic deposition



radiation damage



Outline:

- Basic idea
- Algorithm
  - Distance metric
  - Metric search
  - Force interpolation
- Results
- Conclusion

please ask questions

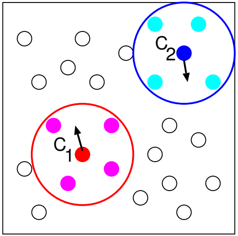
unfortunately the last pretty picture for today

Basic idea

Premise: we can decompose the system into sub-configurations (*clusters*) & if (isolated) sub-configurations are sufficiently large the forces on the central atoms are sufficiently accurate, *i.e.* **locality**

Procedure:

1. as dynamics ensues new (*query*) configurations are generated and compared to a cluster-force database.
2. if sufficient stored clusters are close the query cluster the stored forces are interpolated at the query cluster
3. otherwise, *ab initio* forces are calculated for the new cluster and stored



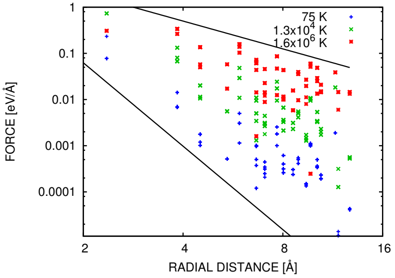
As opposed to the *globally*-tuned *empirical* potential, we explicitly propogate the system in time using *locally interpolated* DFT forces.

We endow the database with a distance measure that facilitates: (a) fast metric-based **searches** and (b) robust **interpolation** of database information at queries.

Q1. Locality

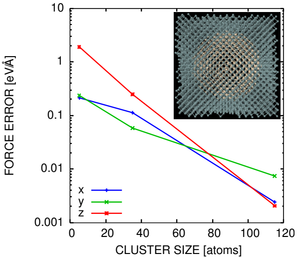
Local dependence of Hellman-Feynman force on configuration is the basic premise of all empirical potentials. **Note: all data shown is from *ab initio* molecular dyanmics**

Spatial decay of force due to a perturbation of Si



Upper trend line is $\frac{1}{r^2}$, the lower $\frac{1}{r^6}$. Sensitive to disorder & displacement

Convergence of the central atom force with cluster size.



for a periodic 1000 atom C system & spherical subconfiguration.

Cluster distance

A *cluster* $\mathcal{C}_A$ is a set $\mathcal{C}_A = \{\Delta\mathbf{x}_{1A}, \Delta\mathbf{x}_{2A}, \Delta\mathbf{x}_{3A}, \ldots, \Delta\mathbf{x}_{N_A A}\}$ of distance vectors $\Delta\mathbf{x}_{\alpha A} \equiv \mathbf{x}_\alpha - \mathbf{x}_A$ relative to a central atom $\mathbf{x}_A$,

A distance $d(A, B)$ between clusters $\mathcal{C}_A$ & $\mathcal{C}_B$, needs (a) basic **metric** properties:

- M1 **coincidence**: $d(A, B) = 0$ iff $A = B$
- M2 **positivity**: $d(A, B) > 0$
- M3 **symmetry**: $d(A, B) = d(B, A)$
- M4 **triangle inequality**: $d(A, C) \leq d(A, B) + d(B, C)$

and (b) physical **invariances** $d(A, B) = d(A, B')$:

- I1 **translation** $\mathcal{C}_B \to \mathcal{C}_{B'} = \mathcal{C}_B + \mathbf{a}$
- I2 **rotation** $\mathcal{C}_B \to \mathcal{C}_{B'} = \mathbf{R}\mathcal{C}_B$
- I3 **permutation** $\mathcal{C}_B \to \mathcal{C}_{B'} = \mathrm{P}\mathcal{C}_B$

Root Mean Square Distance

Assuming $N_A = N_B$, the root mean square deviation (RMSD) comparison metric is

$$d_{\mathrm{RSMD}}(A,B) = \min_{\mathbf{R},\mathrm{P}} \|\mathrm{X}_A - \mathrm{P}\mathrm{X}_B\mathbf{R}^T\|$$
$$= \min_{\mathbf{R},\mathrm{P}} \sqrt{(\mathrm{X}_A - \mathrm{P}\mathrm{X}_B\mathbf{R}^T)\cdot \mathrm{W}\,(\mathrm{X}_A - \mathrm{P}\mathrm{X}_B\mathbf{R}^T)}$$
$$\equiv \min_{\mathbf{R},\mathrm{P}} \sqrt{\sum_{\alpha,\beta=1}^{N_b} \|\Delta\mathbf{x}_{\alpha A} - \mathrm{P}_{\alpha\beta}\mathbf{R}\Delta\mathbf{x}_{\beta B}\|^2\,\mathrm{W}_{\alpha\beta}}$$

where:

- X_A and X_B are matrices of the relative position vectors $\Delta\mathbf{x}_\alpha$,
- $\mathbf{R} \in \mathrm{Orth}^+$ is a rotation of the cluster &
- P is a permutation of the cluster ordering, *i.e.* a binary orthogonal matrix which is simply the rearrangement of the rows or columns of the identity matrix.

Optimal rotation

$$d(A,B) = \min_{\mathbf{R},\mathrm{P}} \sqrt{(\mathrm{X}_A - \mathrm{P}\mathrm{X}_B\mathbf{R}^T)\cdot \mathrm{W}\,(\mathrm{X}_A - \mathrm{P}\mathrm{X}_B\mathbf{R}^T)}$$

$$= \min_{\mathbf{R},\mathbf{P}} \sqrt{\|\mathrm{X}_A\|^2 + \|\mathrm{X}_B\|^2 - 2\mathrm{X}_A \cdot \mathrm{WP}\mathrm{X}_B\mathbf{R}^T}$$

To determine the rotation **R**, [Kabsch,1976] noticed the last term

$$2\mathrm{X}_A \cdot \mathrm{WP}\mathrm{X}_B\mathbf{R}^T = 2\mathrm{X}_A^T\mathrm{WP}\mathrm{X}_B\mathbf{R}^T$$

is the only one dependent on **R**.

A (3x3) singular value decomposition SVD $\left[\mathrm{X}_A^T\mathrm{WP}\mathrm{X}_B\right] = \mathbf{USV}^T$ gives the solution $\mathbf{R} = \mathbf{UV}^T$ and hence

$$d(A,B) = \min_{\mathrm{P}} \sqrt{\|\mathrm{X}_A\|^2 + \|\mathrm{X}_B\|^2 - 2\,\mathrm{tr}\,\mathbf{S}(\mathrm{P})}$$

Distance from Gaussian densities

Finding the optimal *permutation* P is considerably harder, requiring *e.g.* the $\mathcal{O}(n^3)$ Hungarian/branch & bound algorithms.

Instead, the cluster atomic densities can be represented as an order-independent sum of Gaussian smeared point **densities**

$$\rho(\mathbf{x}) = \Delta(\mathbf{0}) + \sum_{\alpha} \Delta(\mathbf{x} - \mathbf{x}_\alpha)$$

where
$$\Delta_\sigma(\mathbf{x}) = \exp\left(-\frac{\|\mathbf{x}-\mathbf{x}_\alpha\|^2}{2\sigma^2}\right)$$

And hence

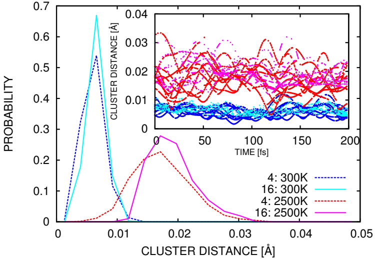
$$d_{\mathrm{OGTO}}(A, B) = \sqrt{\pi}^3 \sigma^3 \sum_{\alpha,\beta} \exp\left(-\frac{r_{\alpha\beta}^2}{4\sigma^2}\right)$$

Finding the optimal rotation $\mathbf{R}$ and Gaussian width σ is possible but the details are omitted here.

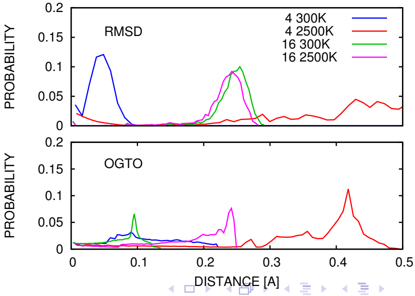
Distribution of cluster distances

The distribution of inter-cluster distances as a function of:

- **metric**: RMSD *vs.* OGTO, the two metrics produce qualitatively different distributions *e.g.* RMSD permutation sensitivity results in an artificial gap
- **number of neighbors**: 4 (1st) *vs.* 16 (1st & 2nd shells), more neighbors spreads the distance distribution & is more discriminatory
- **temperature**: 300 K (crystalline) *vs.* 2500 K (melt), both metrics indicate the expected structural changes with the shift from lattice to amorphous



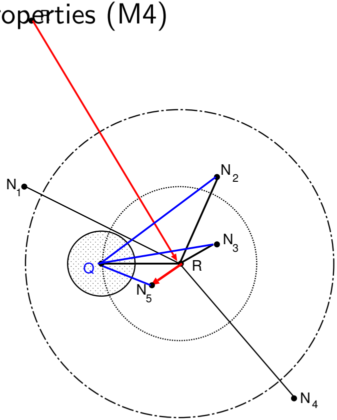
⇑ single trajectory ⇓ all atoms & all times



Metric database search

A key ingredient is searching a large, dense database efficiently [FOGOLARI,2012] - utilize metric properties (M4)

- **Initialize:** *select* a set of points $\{R_i\}$ in the database randomly or from previous time step.
- **Loop:**
  1. *Find* closest in set: $\text{argmin}_R\, d(Q, R_i)$, where Q is the query configuration
  2. *Retrieve* neighboring set $\{N\}$ of points about the point R for which $2d(N_i, R) < d(Q, R)$
  3. *Stop* if the neighborhood around Q is the desired radius or size.



Query point Q, **reference point** R and its neighbors N_i in the database. The subset { N_2, N_3, N_5 } are candidates for configurations in the interpolation ball around Q. After computation of distances of the subset to Q, N_5 will be selected as the best

Interpolation of forces

Given $\{d_{QN} \mid N \in \mathcal{B}_Q\}$, an accurate force $\mathbf{f}_Q$ can be obtained from interpolation via **radial basis functions** $\phi(r)$

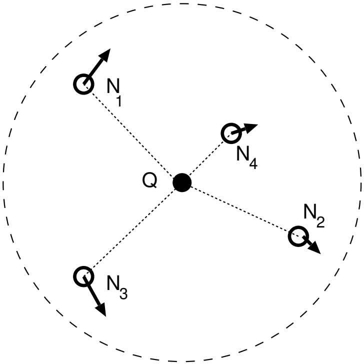
$$\mathbf{f}_Q = \sum_{N \in \mathcal{B}_Q} \phi(d_{QN})\, \mathbf{a}_N$$

where $\mathbf{a}_N$ are given by the consistency conditions

$$\mathbf{R}_{QN}\mathbf{f}_N = \sum_{B \in \mathcal{B}_Q} \phi(d_{NB})\, \mathbf{a}_B$$

for every $N \in \mathcal{B}_Q$

Cluster space



this is the (shaded) trust ball around the query point Q in the previous slide

Questions

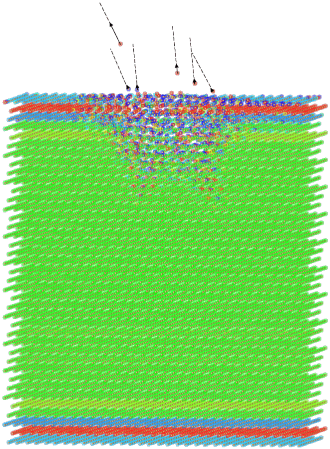
The algorithm is viable if:

Q1. **sensitivity** of the forces to the configuration is **local**

Q2. inter-cluster **distances** are **correlated** with **forces** on the central atoms

Q3. the search is sufficiently **efficient**

Q4. the **error** in interpolating the force is **controllable**



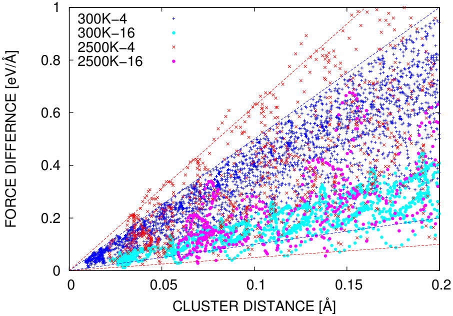
Q2. Correlation of cluster distance and forces

The difference in forces on the central atom for two clusters

$$\|\mathbf{f}_A - \mathbf{f}_B\| \propto d(A, B)$$

is highly & linearly correlated with the inter-cluster distance $d(A, B)$.

The correlation is dependent on temperature and less so on number of neighbors.



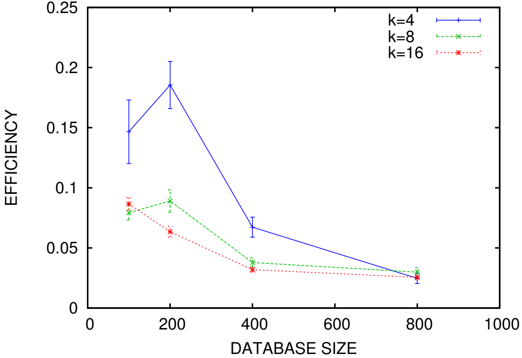
$\Longleftarrow$ **better match**

The sensitivity is ≈ 10 eV/Å^2

Q3. Metric search efficiency

Measuring search efficiency as: $\dfrac{d}{s}$

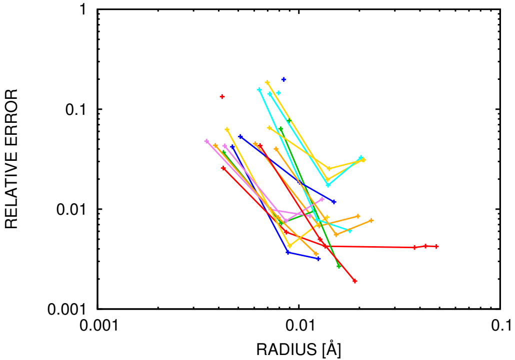
where s=database size & d=number of distance computations



This is a worst-case for dynamics, subsequent steps would use the previous as a starting guess.

Q4. Force interpolation error

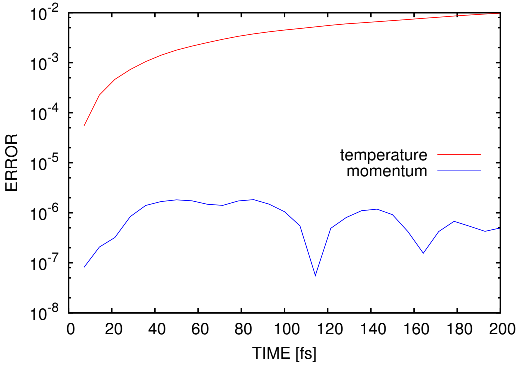
Measuring interpolation of error as $\dfrac{\|\hat{\mathbf{f}} - \mathbf{f}\|}{\|\mathbf{f}\|}$



where $\mathbf{f}$=true force & $\hat{\mathbf{f}}$=interpolated force (not using the true force sample) and R is the radius of the trust ball in cluster space

Conservation properties

Momentum and temperature as a function of time:



we don't have a direct means of measuring potential energy.

We can control temperature with a thermostat.

Current work

- Application to atomic deposition, radiation damage, and structurally interesting materials like metal-organic frameworks is upcoming.
- Since the method reduces to *ab initio* molecular dynamics in the limit to full system clusters and no interpolation, we are trying to make stronger connections to it.
- We are also exploring using machine learning techniques to enhance the search and interpolation methodology.

Reese Jones **rjones@sandia.gov**