Argonne NATIONAL LABORATORY

2015 Margaret Butler Celebration

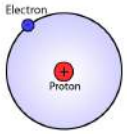
Toward a Better Atomistic Description of Intermolecular Interactions in Materials

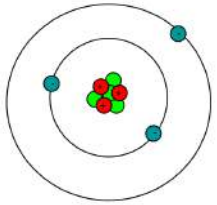
Ying Li
Margaret Butler Postdoctoral Fellow
Leadership Computing Facility
Argonne National Laboratory
yingli@anl.gov

U.S. DEPARTMENT OF ENERGY

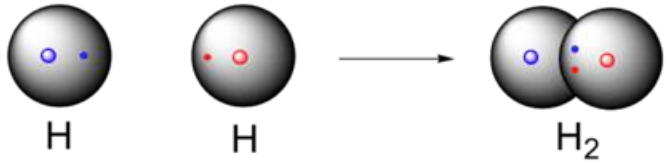
What's all the matter? Atoms and Molecules

- **Atom(s)**: the smallest unit of matter that has the chemical properties.

Hydrogen 


Lithium

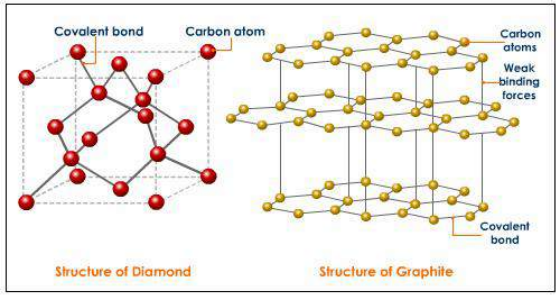
- **Molecule(s)**: chemical compounds held together by two or more atoms.



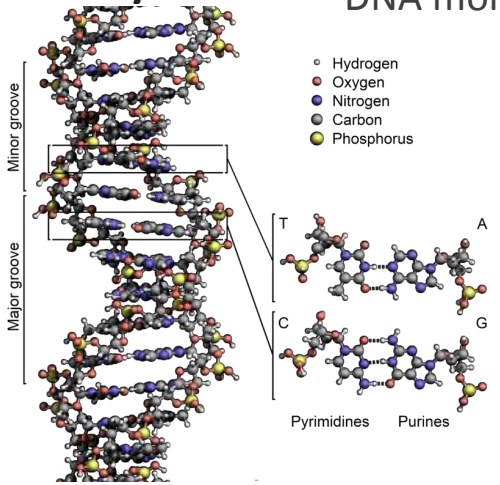
- How atoms make up the world?
  - **(Same or different) atoms arrange in different ways**

Carbon atoms



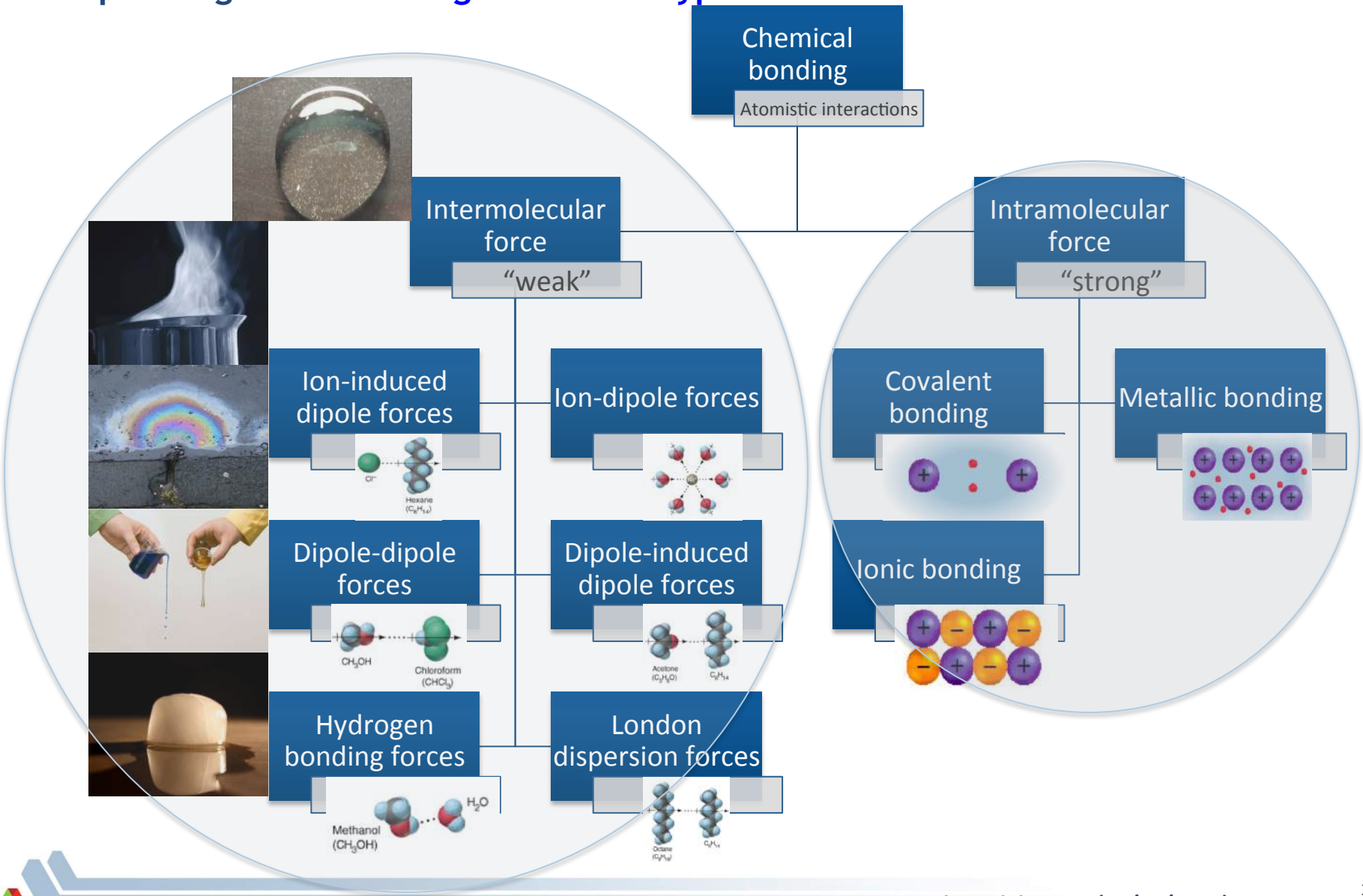


DNA molecule



Types of interactions holding atoms together

Depending on the **arrangement** and **types** of atoms



What is force field?

Used for describing the interactions between atoms/molecules

Region around a body in which it behaviors (experience of force) due to the presence of another body.

In **molecular mechanics simulation**: sets of parameter and functional forms used in simulations to describe the calculate the **potential energy** of the atoms/molecules

Biophysics

- **AMBER** proteins and DNA
- **CHARMM** small molecules and macromolecules
- **OPLS** The parameters were optimized to fit experimental properties of liquids
- **AMEOBA** Biomolecular systems contain polarizable point dioples

Materials

- **EAM** Alloys and metals
- **REBO** e.g., such as semiconductors
- **COMB** e.g., TiO2, Cu, ZnO, etc
- **REAXFF** Various kinds of materials, highly transferable and good parallelism

Water

- **TIP3P, TIP4P, TIP5P and SPC**

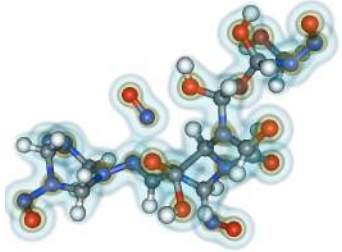
Coarse-grained

- **MARTINI** Group several atoms into a CG interaction site, for lipids, proteins, carbonhydrates, etc.

How to create a force field that is the best of our research project? (3 steps)

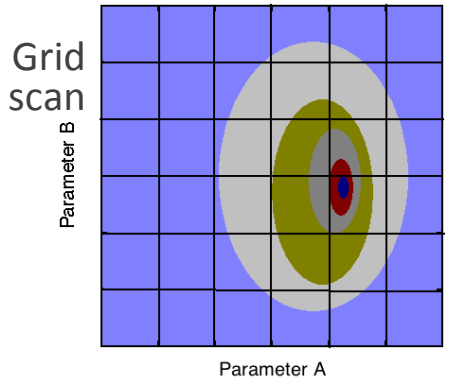
- **Step 1**: Choose the force field functional forms/pick up molecular descriptors
  - E.g., $E_{system} = E\{bond, angle, diahedral\} + E\{vdWaals + Coulomb\}$

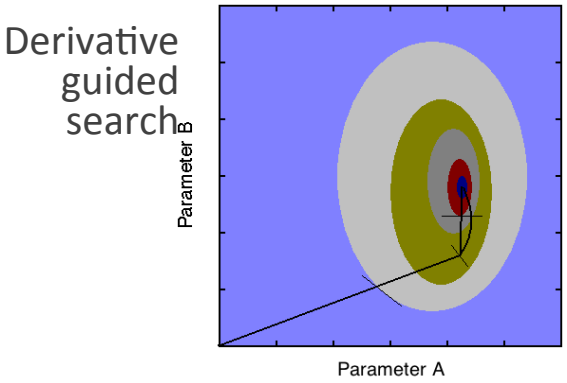
    $E_{system} = E_{\text{intermolecules}} + E_{\text{intramolecules}}$
- **Step 2**: Choose reference data sets from theoretical calculations or experimental measurements.

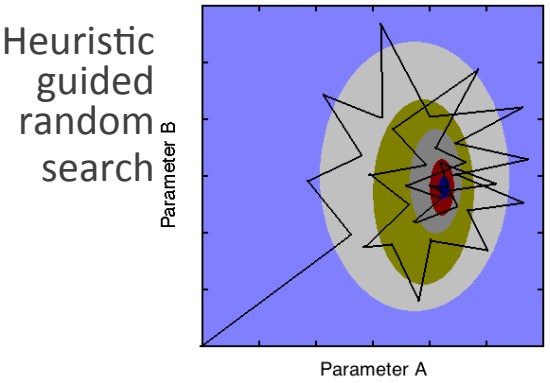


or

- **Step 3**: Optimize (minimize) the objective function, which measures the disagreement between the reference data and the corresponding simulation result.



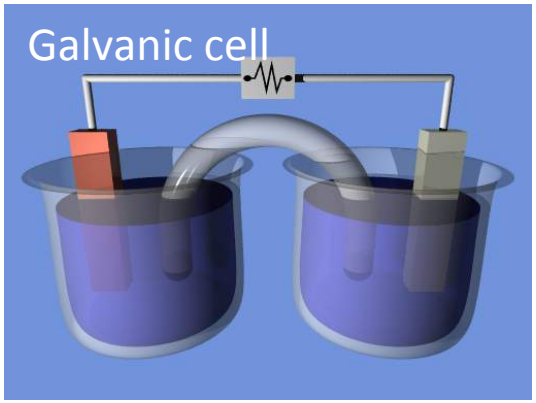




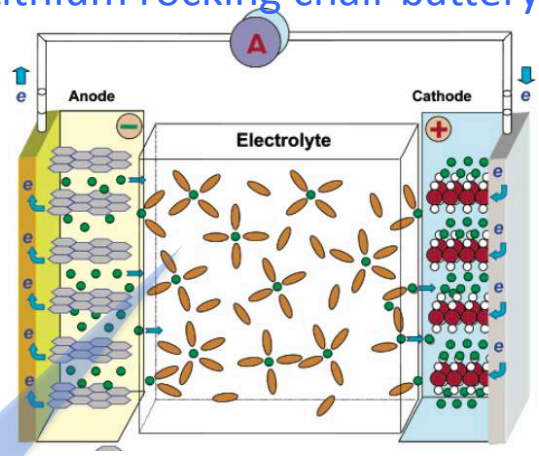
Research motivation

Large-scale molecular dynamics simulations of materials problem

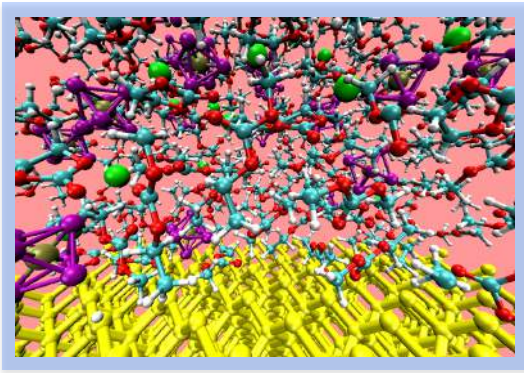
- Electrolytes of battery allow ions to move between electrodes and chemical terminals → current to flow out of the battery to perform work.



Lithium rocking chair battery



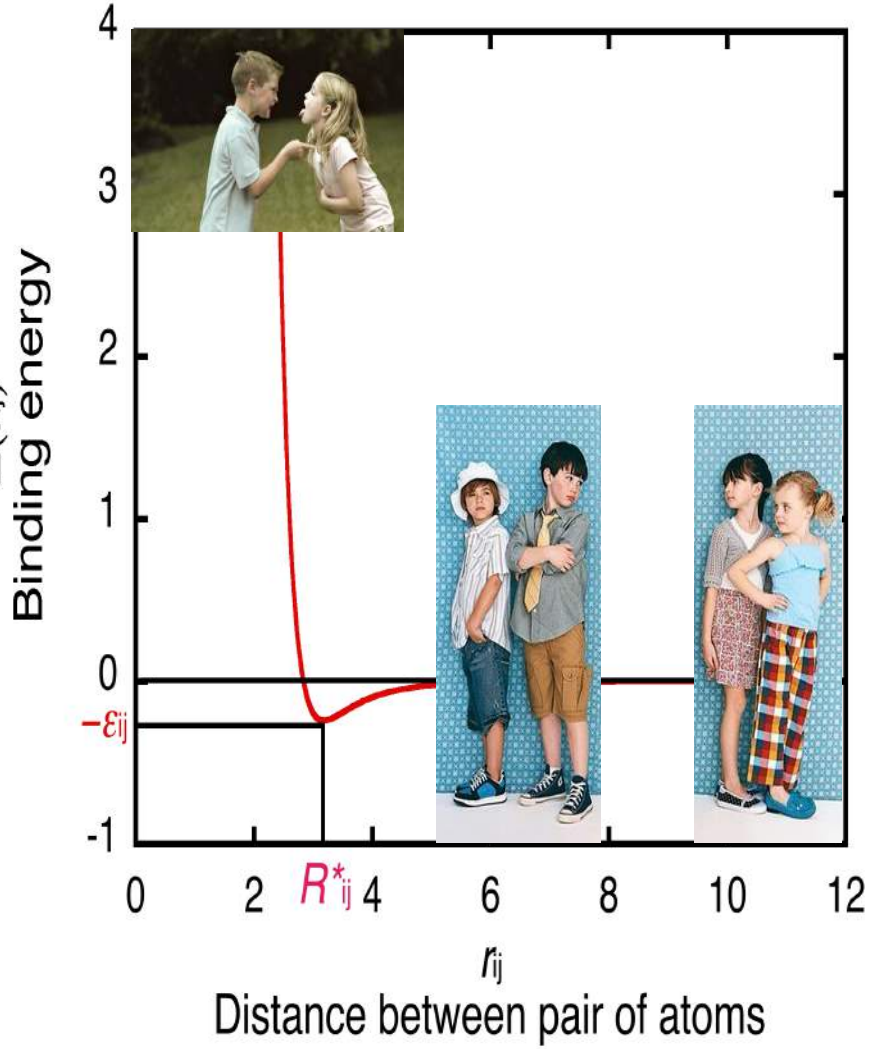
Electrolyte of rechargeable battery often is mixture of organic liquids.



- Initial focus: interatomic potential for electrolyte dynamics
- Leading problem: **non-electrostatic intermolecular interaction** in organic systems

Step 1: Functional (physically meaningful) forms

Binding energy: atomistic representation of non-electrostatic intermolecular interactions in organic systems



- **Form #1:** Classical 12-6 Lennard-Jones potential: (2 tunable parameters)

$$E_{binding} = \sum_{i>j} E(r_{ij}) = \sum_{i>j} \varepsilon_{ij} \left[\left(\frac{R^*_{ij}}{r_{ij}} \right)^{12} - 2 \cdot \left(\frac{R^*_{ij}}{r_{ij}} \right)^{6} \right]$$

$$\varepsilon_{ij} = \sqrt{\varepsilon_{ii}\varepsilon_{jj}} \; ; \quad R^*_{ij} = \frac{1}{2}\left(R^*_{ii} + R^*_{jj} \right)$$

- **Form #2:** Softer Buffered 14-7 potential: (2 tunable parameters)

$$E_{binding} = \sum_{i>j} E(r_{ij}) = \sum_{i>j} \varepsilon_{ij} \left(\frac{1.07}{\rho_{ij} + 0.07} \right)^{7} \left(\frac{1.12}{\rho_{ij}^{7} + 0.12} - 2 \right) ; \quad \rho_{ij} = \frac{r_{ij}}{R^*_{ij}}$$

$$\varepsilon_{ij} = \frac{4\varepsilon_{ii}\varepsilon_{jj}}{\left(\sqrt{\varepsilon_{ii}} + \sqrt{\varepsilon_{jj}} \right)^2} \; ; \quad R^*_{ij} = \frac{\left(R^*_{ii} \right)^3 + \left(R^*_{jj} \right)^3}{\left(R^*_{ii} \right)^2 + \left(R^*_{jj} \right)^2}$$

- **Form #3:** Mie n-m potential: (4 tunable parameters)

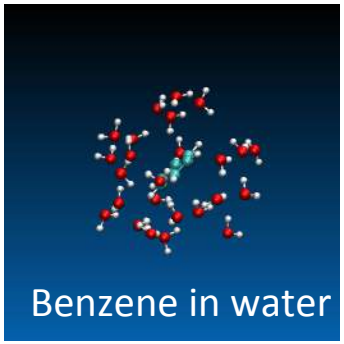
$$E_{binding(Buf\_n-m)} = \sum_{i<j} E(r_{ij}) = \sum_{i<j} k_{ij}\varepsilon_{ij} \left[\left(\frac{1}{\rho_{ij}} \right)^{n_{ij}} - \left(\frac{1}{\rho_{ij}} \right)^{m_{ij}} \right]$$

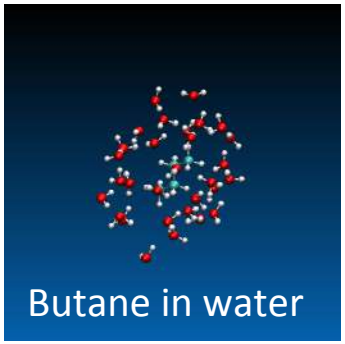
$$\varepsilon_{ij} \; ; \; \rho_{ij} = \frac{r_{ij}}{R^*_{ij}} \; ; \; k_{ij} = \frac{n_{ij}}{n_{ij} - m_{ij}} \left(\frac{n_{ij}}{m_{ij}} \right)^{m_{ij} / (n_{ij} - m_{ij})}$$ with explicit pair parameters

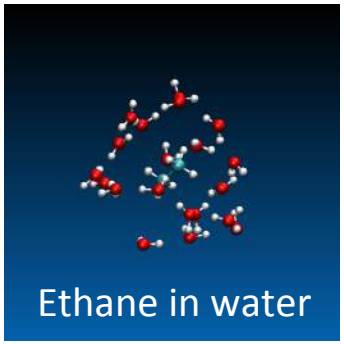
Step 2: Reference data

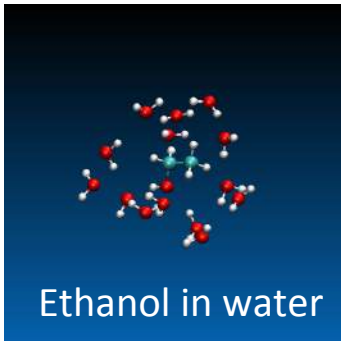
Big data: the more (accurate), the merrier

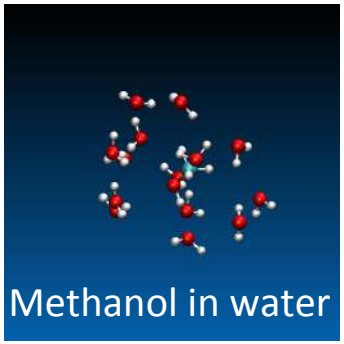
- High level quantum mechanics calculations $E_{binding} = E_{solution} - \left(E_{solute} + E_{solvent}\right)$
- ~1400 configurations and corresponding intermolecular binding energies for various organic molecules
  - Organic molecule in water:


Benzene in water


Butane in water

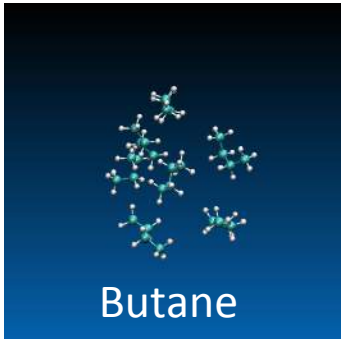

Ethane in water

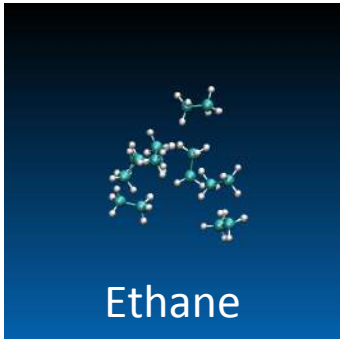

Ethanol in water

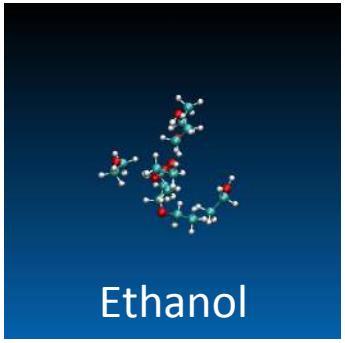

Methanol in water

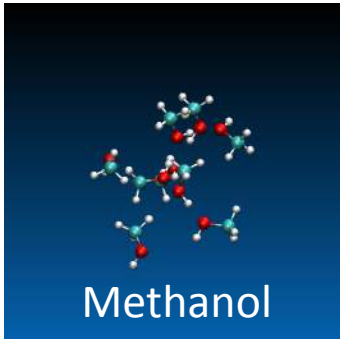
  - Pure organic molecules:


Benzene


Butane


Ethane


Ethanol


Methanol

Step 3: Optimization method: genetic algorithm

GA is a heuristic search that mimics the process of natural selection.

Reference data
QM calculations to represent the reality (Structures, energies, …)

$X_{reference\ data}$

$X_{potential\ functional}$

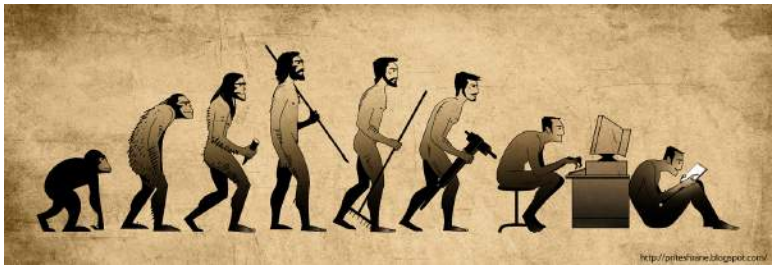
Random distributed parameters sets (populations) of potential functional forms

$F(X_{reference\ data}, X^{(better)}_{potential\ functional})$

Objective functions

Genetic operations
- Reproduction
- Crossover
- Mutation

Strategic selection



$X^{better}_{potential\ functional}$

Better populations

Converged ?

No

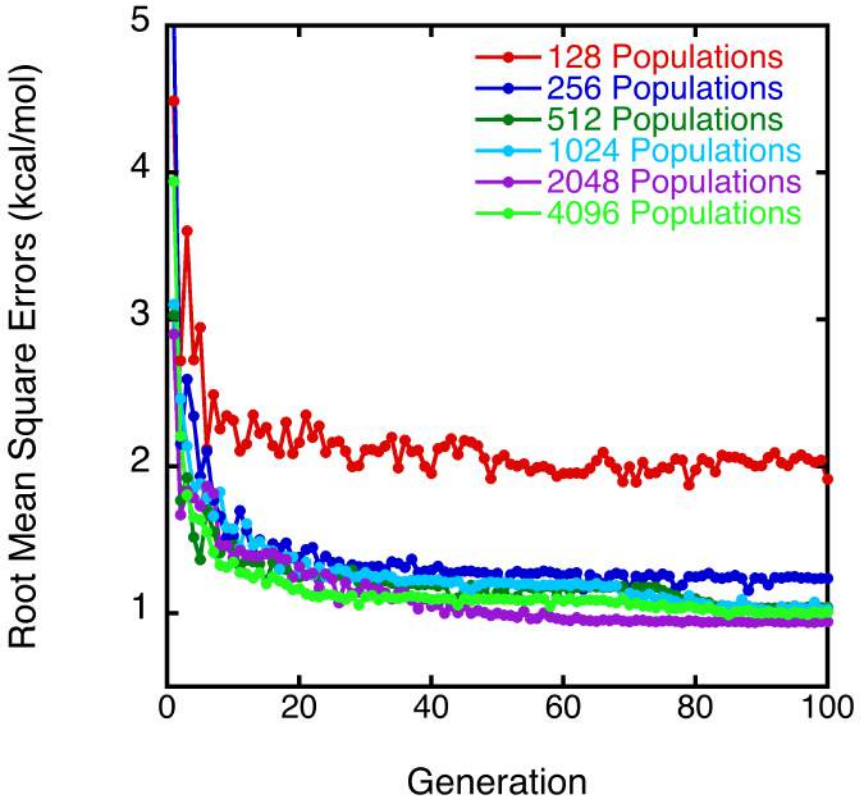
Yes

Optimized solution (cross validation, regularization …)

Parallel genetic algorithm (PGA) efficiency

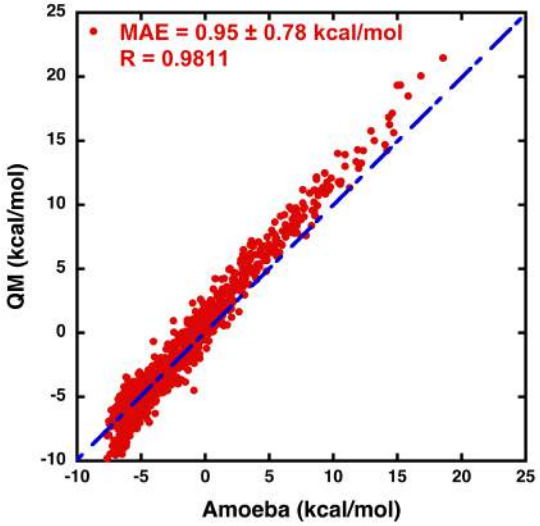
In order to make GA faster, parallel GA is implemented.

- Objective functions converge quicker with larger number of populations
  - Each population runs on one CPU to get the objective function
  - Comparison of objective functions on 128 ~ 4096 CPUs on Mira/Vesta

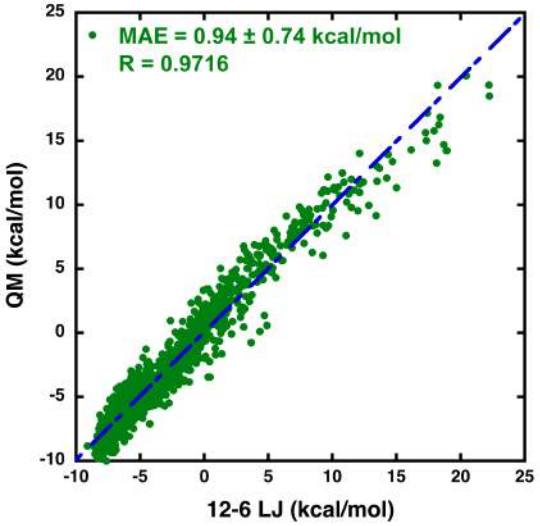


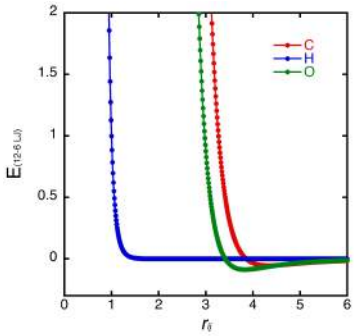
- With more complicated objective functions (such as, multiobjective functions) and multidimensional parameters, the requirement of parallelism will increase.

Better fitted results

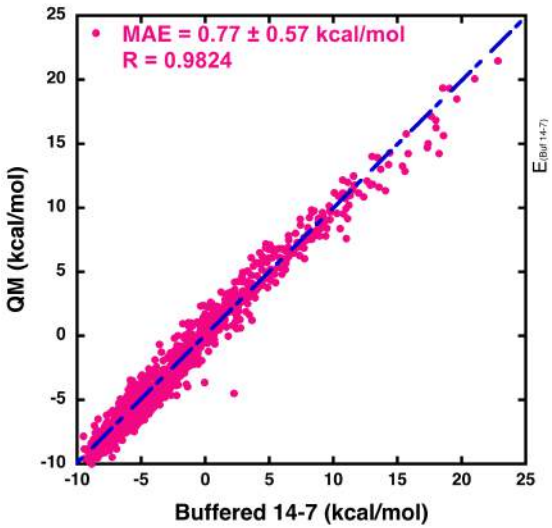


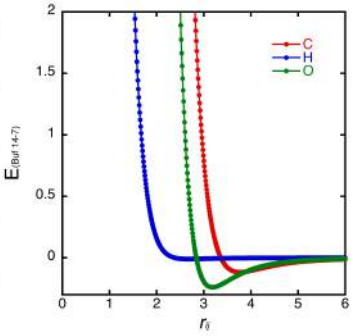
Original Amoeba [1]: 8 explicit atom types (24 parameters)



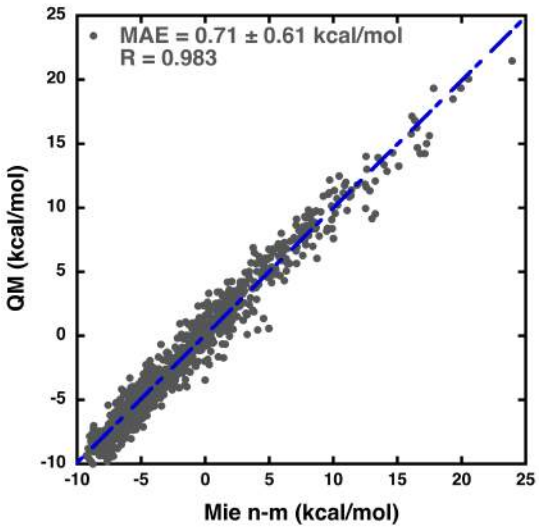


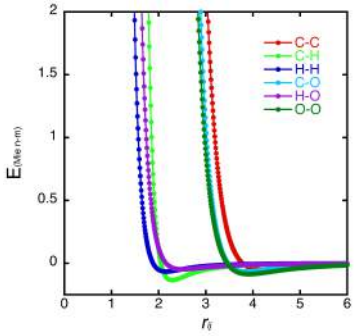
Form #1
(6 parameters)





Form #2
(6 parameters)





Form #3
(24 parameters)

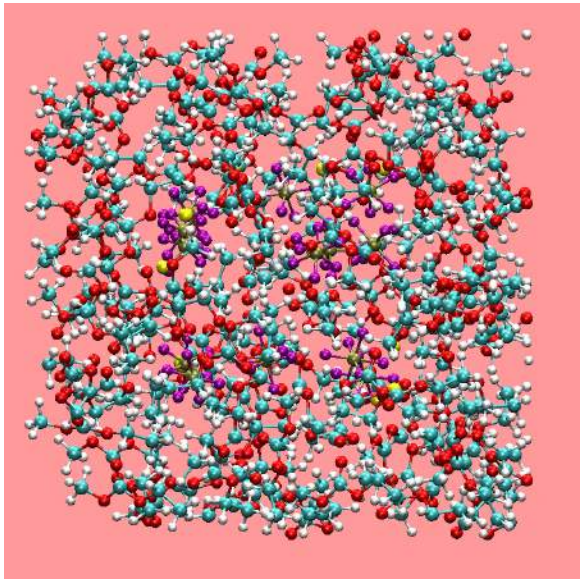
- **Better:** accuracy (mean absolute errors: $0.95^{24} \rightarrow 0.94^{6} \rightarrow 0.77^{6} \rightarrow 0.71^{24}$ kcal/mole) improves, but no complexity of model increases

[1] Yue Shi, Zhen Xia, Jiajing Zhang, Robert Best, Chuanjie Wu, Jay W. Ponder, and Pengyu Ren Polarizable Atomic Multipole-Based AMOEBA Force Field for Proteins. *Journal of Chemical Theory and Computation* 2013 *9* (9), 4046-4063 DOI: 10.1021/ct4003702

Challenges

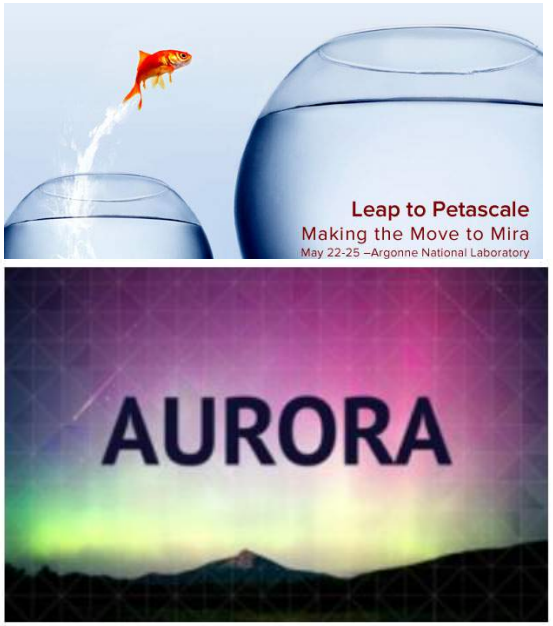
using leadership computing facility to solve materials problems

- Scientific challenges:
  - More (accurate) reference data
  - More complicated functional forms
  - Large (10^9~10^{12} atoms) and long time (ms~s) real scale simulations



Molecular Dynamics simulation with random pair force field: electrolyte salt ($LiPF_6$) in organic liquids (ethylene carbonate ($(CH_2O)_2CO$) + methylene carbonate ($OC(OCH_3)_2$))

- Computational challenges:
  - Better algorithmic solvers
  - New computing architectures
  - Software platforms



Aurora is coming. Step one: Theta

Acknowledgement

- Margaret Butler Postdoctoral Fellowship
- Women In Science and Technology program, Argonne
- Director's Discretionary Program, Argonne Leadership Computing Facility
- Dr. Maria Chan, Dr. Subramanian Sankaranarayanan, Center for Nanoscale Materials, Argonne
- Prof. Benoit Roux, Department of Biochemistry and Molecular Biology, University of Chicago
- Dr. Peter Zapol, Materials Science Division, Argonne
- Dr. John Low, Computing, Environment and Life Sciences, Argonne
- Dr. Nichols Romero, Dr. Wei Jiang, Argonne Leadership Computing Facility, Argonne
- Fusion/Blues Laboratory Computing Resource Center, Argonne

Thank You