Parameterization of a reactive force field using a Monte Carlo algorithm

Eldhose Iype

Asst. Prof.
Department of Chemical Engineering,
BITS Pilani Dubai Campus

e.iype@dubai.bits-pilani.ac.in

Who am I?

Asst. Prof. (Currently)

- Department of Chemical Engineering,
- BITS Pilani Dubai Campus

Post-doc (2014-2016)

- Simulation of self-replenishing hydrophilic polyurethane coatings.
- Eindhoven University of Technology

PhD (2014)

- Computational modelling of hydration/dehydration behaviour of energy storage materials.
- Eindhoven University of Technology

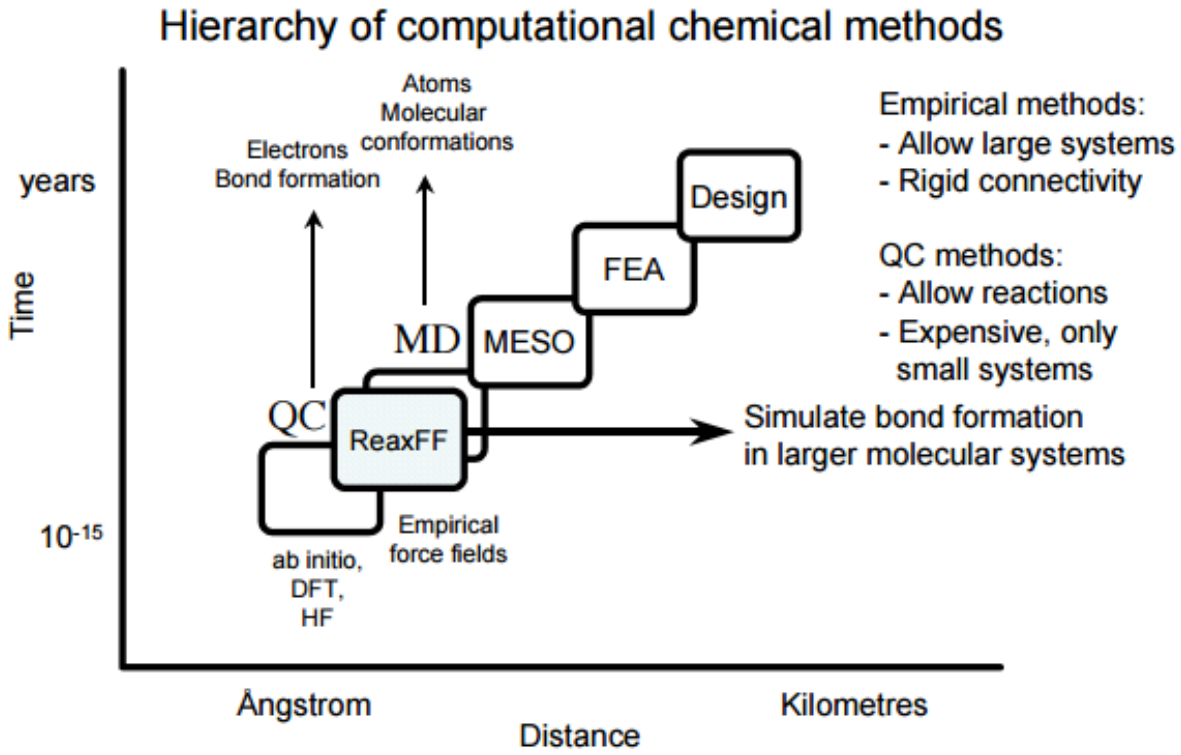
Masters in Chemical Engineering (2008)

- Indian Institute of Technology, Madras

Bachelors in Chemical Engineering (2006)

- Kerala University

Molecular dynamics



Force Field

$$Energy, E = E(r_1, r_2, \cdots r_N)$$

$$Force, F = -\nabla E(r_1, r_2, \cdots r_N)$$

Reax force field

$$E_{system} = E_{vdWaals} + E_{Coulomb} + E_{bond} + E_{val} + E_{tors} + E_{over} + E_{under} + E_{H-bond} + E_{lp} + E_{conj} + E_{pen} + E_{coa} + E_{C2} + E_{triple}$$

Characteristics of ReaxFF

- Dynamic charges are calculated using EEM
- van Der Waal's interaction is calculated using a Morse-type potential
- Energy surface is made continuous
- Connected interactions include
  - Bonded interaction (two body)
  - Valence angle interaction (three body)
  - Torsion interaction (four body)
  - Hydrogen bond interaction

Reax force field

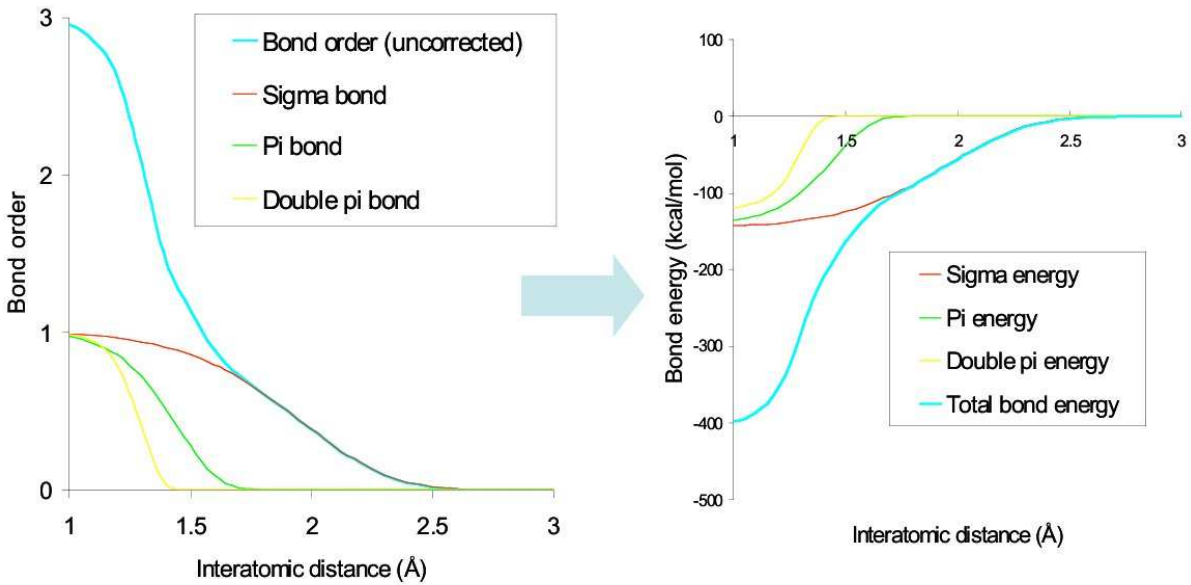
$$E_{system} = E_{vdWaals} + E_{Coulomb} + E_{bond} + E_{val} + E_{tors} + E_{over} + E_{under} + E_{H-bond} + E_{lp} + E_{conj} + E_{pen} + E_{coa} + E_{C2} + E_{triple}$$

Characteristics of ReaxFF

- Dynamic charges are calculated using EEM
- van Der Waal's interaction is calculated using a Morse-type potential
- Energy surface is made continuous
- Connected interactions include
  - Bonded interaction (two body)
  - Valence angle interaction (three body)
  - Torsion interaction (four body)
  - Hydrogen bond interaction

What makes ReaxFF reactive?

- ReaxFF calculates bond order between every pair of atoms
- Bond order is a function of distance of separation
- Every connected interaction is made a function of this bond order
- Thus all the bonds become dynamic



Parameterization of ReaxFF force field

- Quantum Chemical (DFT) data is used to parameterize the force field
- A training data set is prepared which contains the following informations
  - Atomic charges (Mulliken)
  - Equilibrium bond lengths
  - Equilibrium bond angles
  - Torsion angles
  - Energies of the DFT optimized geometries
  - Heat of formation
- Error in the force field is then calculated

$$Err(p_1, p_2, \cdots p_n) = \sum_{i=1}^{n} \left[\frac{x_{i,QM} - x_{i,ReaxFF}}{\sigma_i} \right]^2$$

Parameterization of ReaxFF force field

- Quantum Chemical (DFT) data is used to parameterize the force field
- A training data set is prepared which contains the following informations
  - Atomic charges (Mulliken)
  - Equilibrium bond lengths
  - Equilibrium bond angles
  - Torsion angles
  - Energies of the DFT optimized geometries
  - Heat of formation
- Error in the force field is then calculated

$$Err(p_1, p_2, \cdots p_n) = \sum_{i=1}^{n} \left[\frac{x_{i,QM} - x_{i,ReaxFF}}{\sigma_i} \right]^2$$

Parameterization of ReaxFF force field

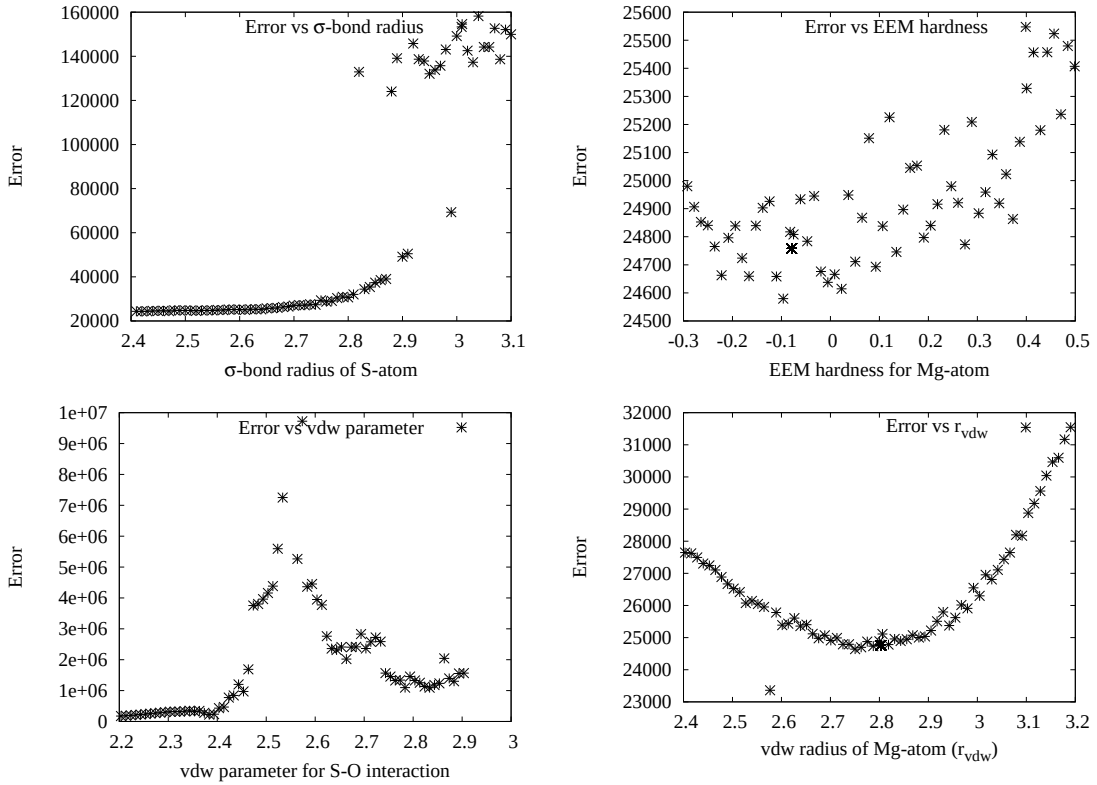
- Quantum Chemical (DFT) data is used to parameterize the force field
- A training data set is prepared which contains the following informations
  - Atomic charges (Mulliken)
  - Equilibrium bond lengths
  - Equilibrium bond angles
  - Torsion angles
  - Energies of the DFT optimized geometries
  - Heat of formation
- Error in the force field is then calculated

$$Err(p_1, p_2, \cdots p_n) = \sum_{i=1}^{n} \left[\frac{x_{i,QM} - x_{i,ReaxFF}}{\sigma_i} \right]^2$$

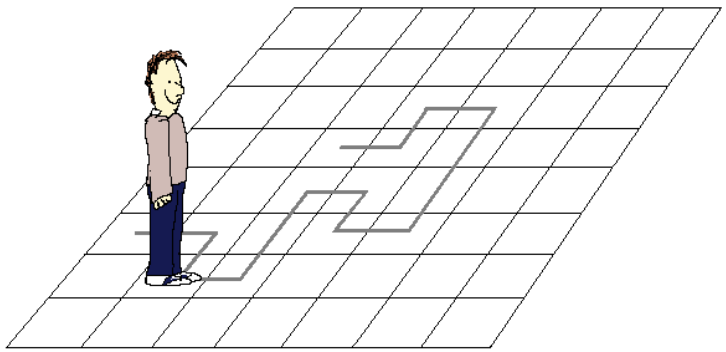
Challenges with parametrization of ReaxFF

- Multiple parameters need to be optimized simultaneously
- Should not be computationally expensive
- Should not get stuck in any local minima
- We do not know a good starting point

Shapes of typical ReaxFF Error surface



Metropolis Monte Carlo algorithm

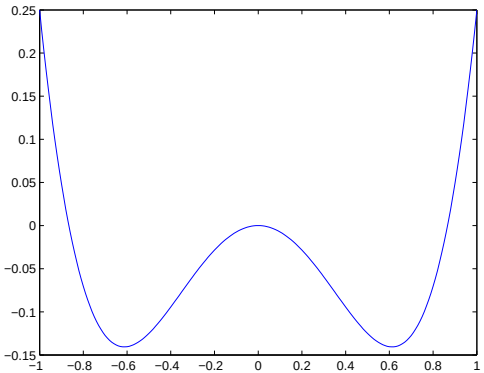


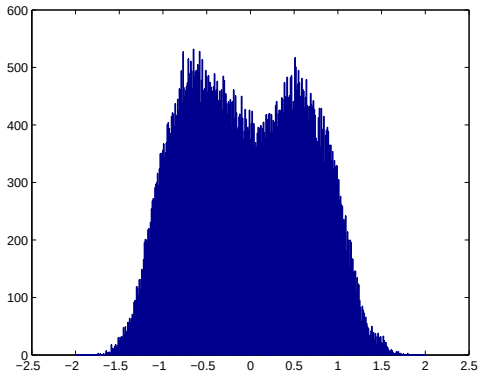
- Start from any random point in space and calculate the initial energy, E_{old}
- Propose a new move in any direction $(random(-1.0, 1.0) * d_{max})$, here d_{max} is the step size.
- Calculate the new energy of the particle, E_{new}
- Calculate the difference in the Energies, $\Delta E = E_{new} - E_{old}$
- Accept the move with a probability given by, $P = min\,[1, exp\,(-\beta \Delta E)]$, where $\beta = \frac{1}{k_B T}$
- Repeat the algorithm

Monte Carlo sampling in a double well potential

$$U(x) = x^4 - 0.75x^2 + 0.01x$$

The distribution has a global maximum near the left well.





MCFF-Optmizer

- Calculate the error of the starting force field, Err_{old}
- Make a new proposition (move) for the parameters
- Calculate the error of the new force field, Err_{new}
- Calculate the difference in the Error, i.e.

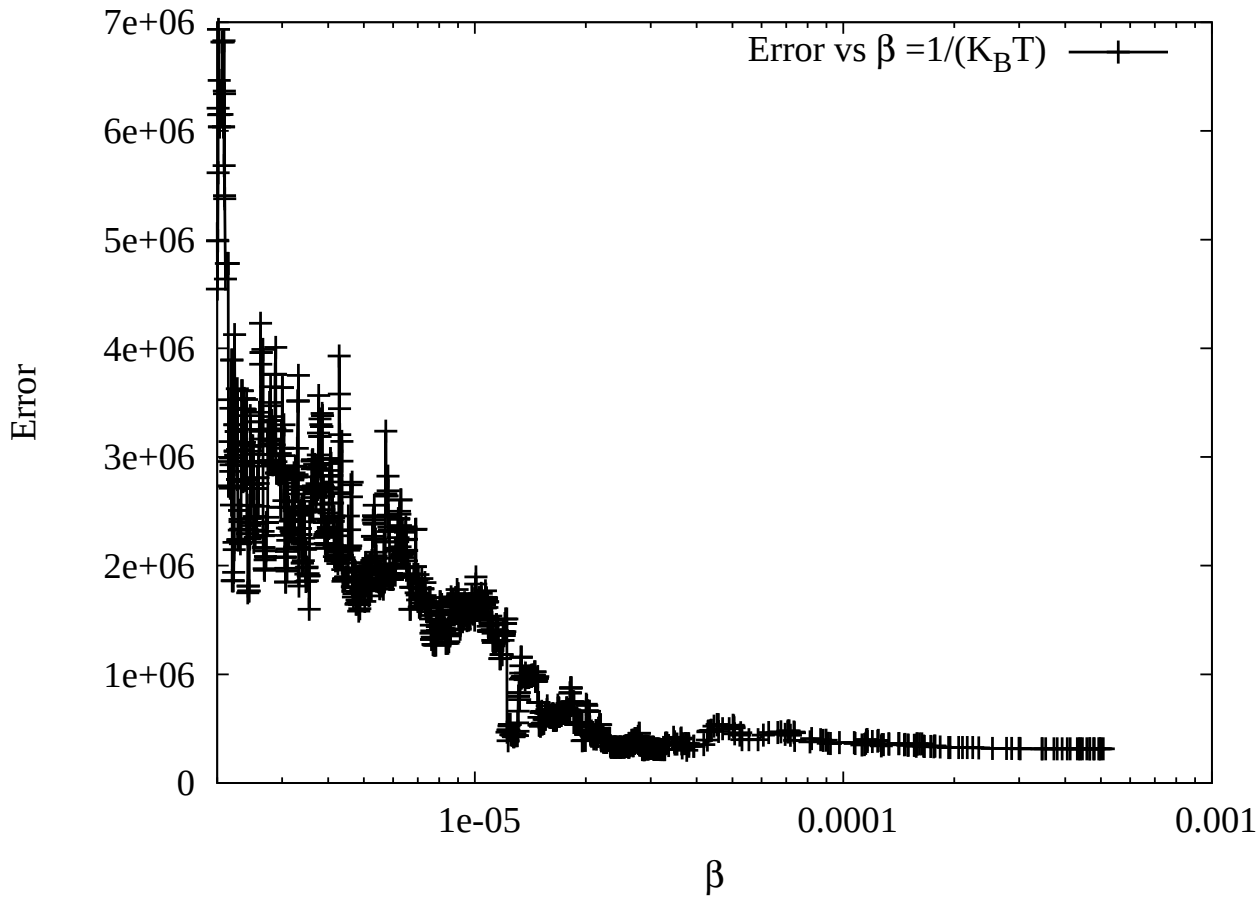
$$\Delta Err = Err_{new} - Err_{old}$$

- Accept the move with a probability given by,

$$P = min\left[1, exp\left(-\beta \Delta Error\right)\right], where, \beta = \frac{1}{k_B T}$$

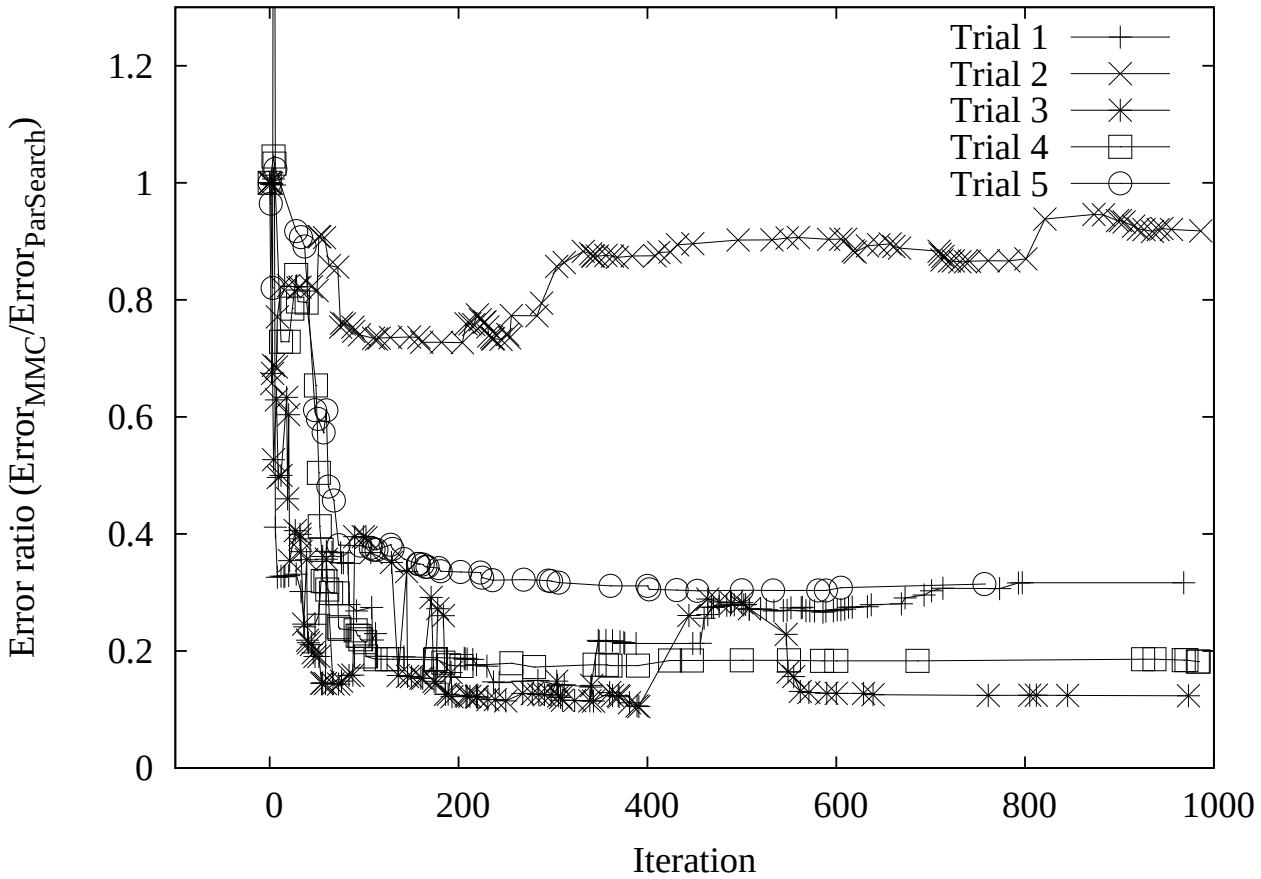
- Repeat the algorithm

Error vs β for a Simulated Annealing run



Comparison between MMC and ParSearch

For five random starting force fields



Comparison of the error values

Initial and final errors of the five simulations. The average $< Error >$ and the coefficient of variation, $\frac{std(sigma)}{<Error>}$, of the final errors are shown.

Trial	$Error_{initial} \times 1E-6$	$Error_{final} \times 1E-6$	
		ParSearch	MMC
1	4.3	0.9	0.31
2	1.1	0.4	0.44
3	8.1	2.5	0.31
4	2.0	1.4	0.26
5	4.3	1.3	0.42
$< Error >$		1.3	0.35
$\frac{std(Error)}{<Error>}$		0.6	0.21

Technical details

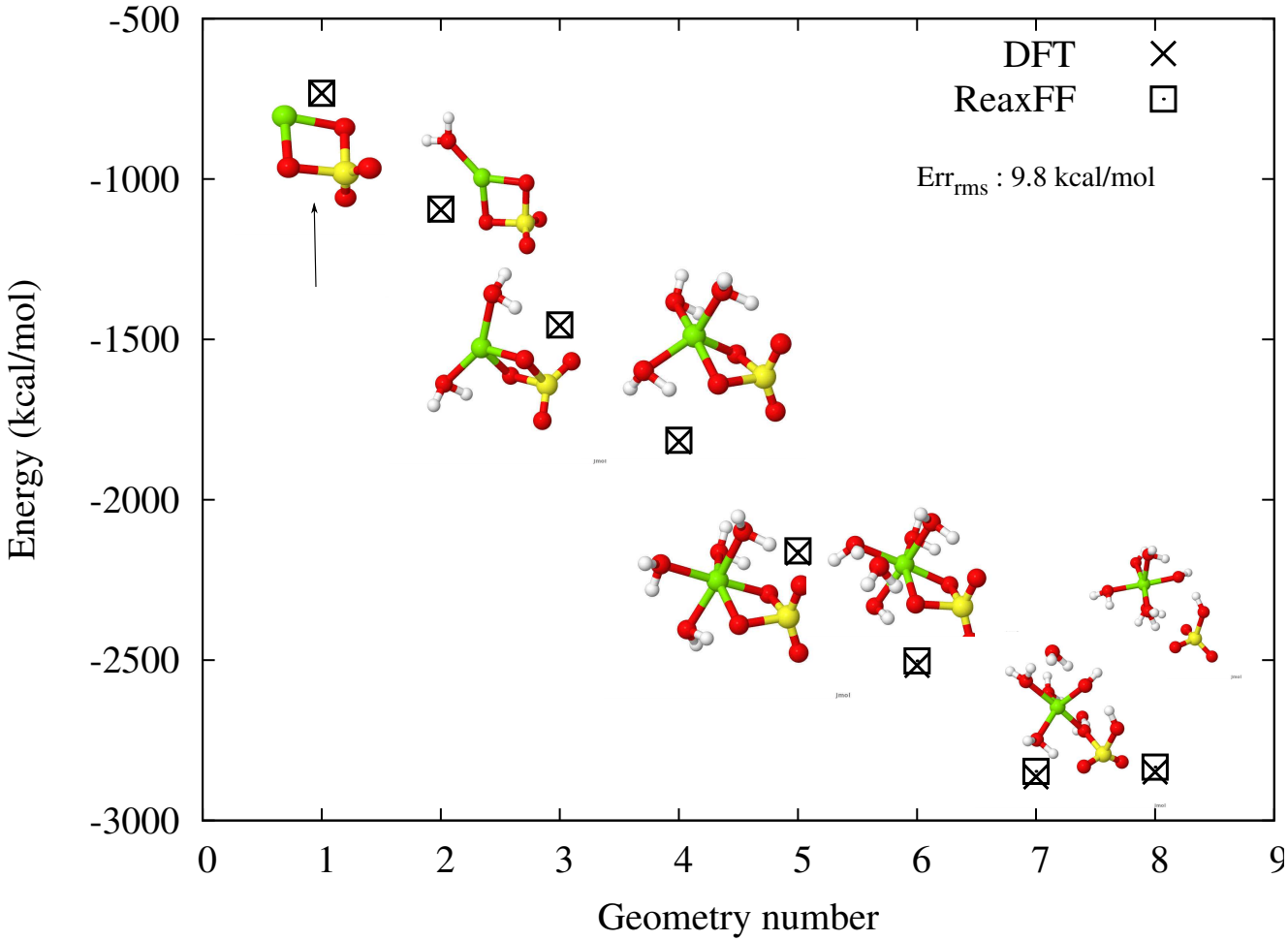
- Maximum step size (d_{max})
  - Maximum change in parameter value at every step
  - It is different for each parameter in a force field
  - Smaller step size results in more acceptance and vice versa, for a given temperature
- Acceptance rate
  - The ratio of accepted moves to the total number of moves
  - Set an acceptance rate of your choice (typically, I use around 20%)
  - Lower, the better (but computationally expensive)
- Cooling rate
  - Slow cooling rate is good
  - At any temperature, the acceptance rate is maintained by scaling up or down d_{max} value
  - So ideally, one would like to have a slow cooling rate and low acceptance rate. But this come at a cost.

Challenges

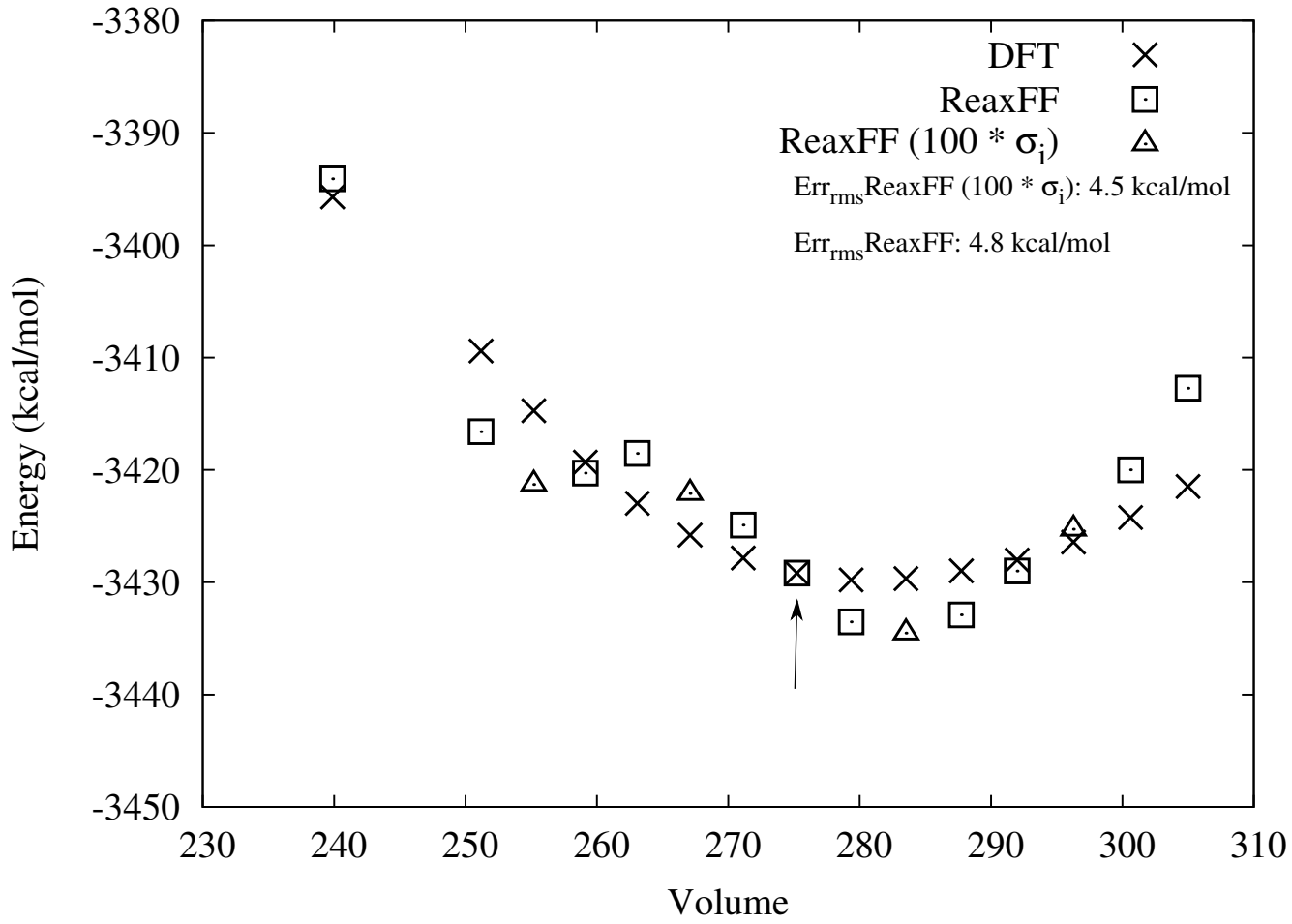
- Bounds for the parameters
  - Taken from the values of each parameter from a set of 11 force fields
  - This would mean, two different elements have same bounds for a given parameter
  - No way to monitor if the optimized value for the parameter is physically meaningful

Energies of the hydrates of $MgSO_4.xH_2O$

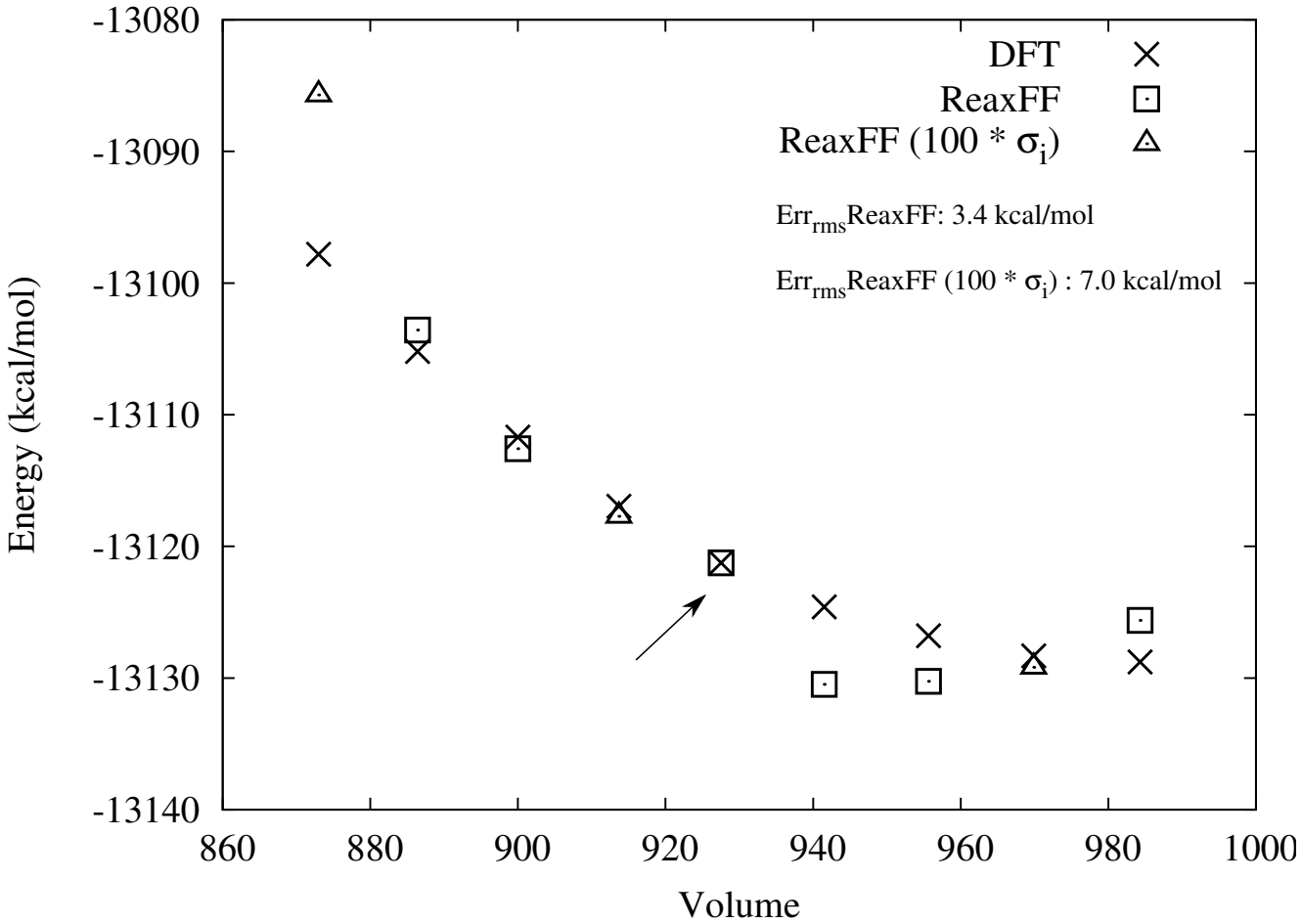
x ranging from 0 to 6



Equation of state for $MgSO_4$

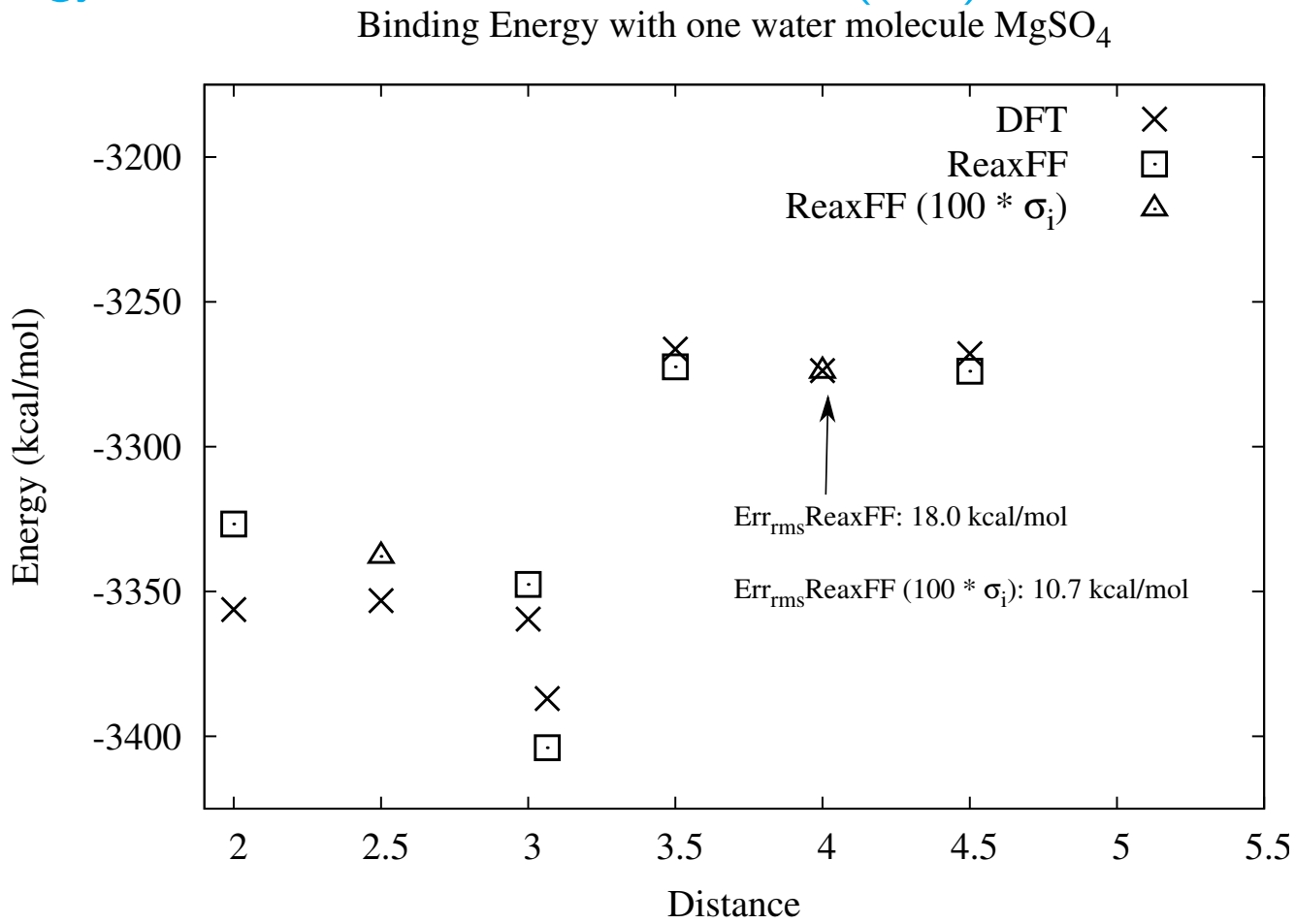


Equation of state for $MgSO_4.7H_2O$

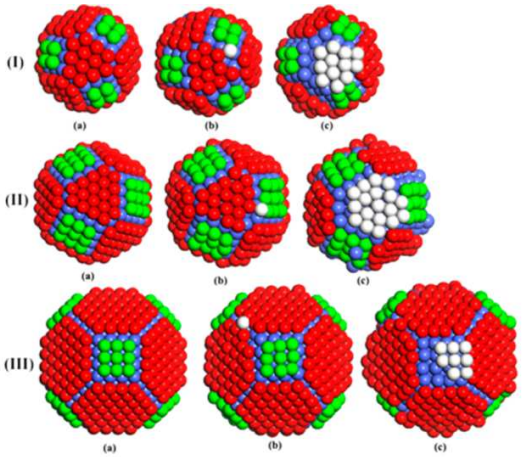


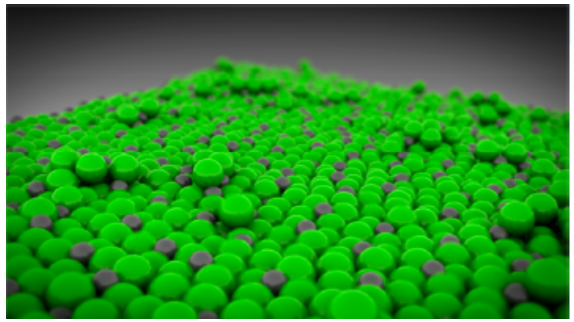
Binding energy

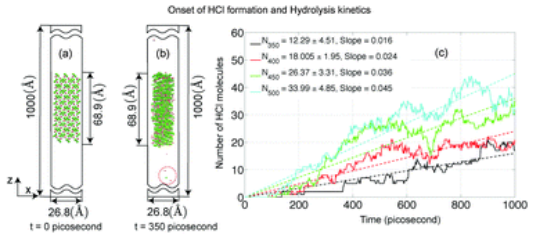
Binding energy of one water molecule on the (100) surface of $MgSO_4$



Examples of works based on MCFF-optimizer







J. Phys. Chem. C. 2014. 118. 6882-6886

ACS Catal. 2015. 5. 596601

PCCP 2016. 18. 15838

Conclusion

- Metropolis MC algorithm is used to parameterize the ReaxFF force field.
- The method shows good improvement over the traditional optimization scheme.
- The stochastic nature of the method allows one to arrive at the global minima in the parameter space.
- The method is efficient and cheaper
- One does not need a good starting point

Acknowledgement

- dr. M. (Markus) Hütter, dr. A.P.J. (Tonek) Jansen, dr. Adri van Duin
- SCM, NCF-SARA, EGS, NSCCS

Conclusion

- Metropolis MC algorithm is used to parameterize the ReaxFF force field.
- The method shows good improvement over the traditional optimization scheme.
- The stochastic nature of the method allows one to arrive at the global minima in the parameter space.
- The method is efficient and cheaper
- One does not need a good starting point

Acknowledgement

- dr. M. (Markus) Hütter, dr. A.P.J. (Tonek) Jansen, dr. Adri van Duin
- SCM, NCF-SARA, EGS, NSCCS

Thank You!!