Data-Driven Models for Predictive Molecular Simulations

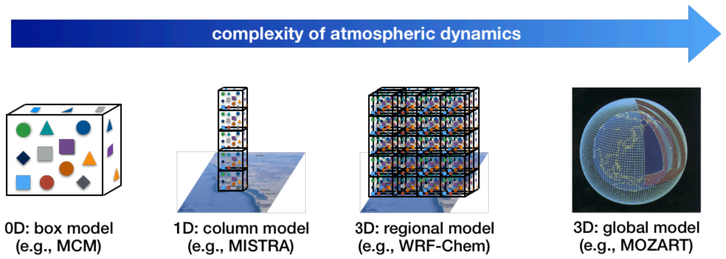
PI: Francesco Paesani; Co-PIs: Andreas Götz and Andrea Zonca

UC San Diego

SDSC SAN DIEGO SUPERCOMPUTER CENTER

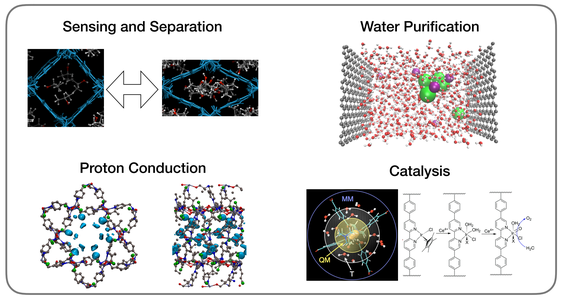
Need for Accurate Molecular Models

From Fundamental Chemistry to Climate Models



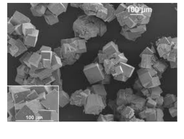
- *J. Phys. Chem. Lett.* 4, 779 (2013)
- *J. Phys. Chem. A* 117, 7131 (2013)
- *J. Phys. Chem. B* 118, 8081 (2014)
- *J. Phys. Chem. A* 119, 1859 (2015)
- *Langmuir* 31, 2147 (2015)
- *J. Phys. Chem. A* 120, 1822 (2016)
- *Phys. Chem. Chem. Phys.* 19, 10481 (2017)
- *J. Am. Chem. Soc.* 140, 4905 (2018)

From Molecular Properties to Materials Discovery



- molecular properties
- machine learning
- *in silico* screening

de novo materials



- *J. Am. Chem. Soc.* 134, 4207 (2012)
- *J. Phys. Chem. C* 117, 19508 (2013)
- *J. Phys. Chem. Lett.* 5, 2897 (2014)
- *J. Am. Chem. Soc.* 138, 6123 (2016)
- *J. Phys. Chem. Lett.*7, 4022 (2017)
- *J. Am. Chem. Soc.* 139, 13973 (2017)
- *Sci. Adv.* 3, 1701508 (2017)
- *J. Am. Chem. Soc.* 140, 1348 (2018)

Many-Body Molecular Dynamics: Chemical Accuracy Across Different Phases

$$E_N = \sum_i^N V^{1B}(i) \quad \text{1-body}$$

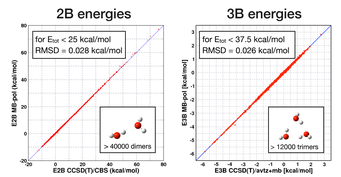
$$+ \sum_{i<j}^N V^{2B}(i,j) \quad \text{2-body}$$

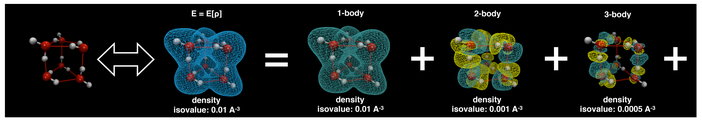
$$+ \sum_{i<j<k}^N V^{3B}(i,j,k) \quad \text{3-body}$$

$$+ \ldots$$

$$+ V^{NB}(1,\ldots,N) \quad \text{N-body}$$

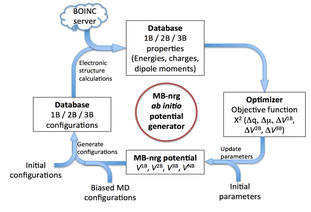
explicit 1B term
+
N-body induction
+
explicit 2B and 3B terms from CCSD(T)/CBS





- *J. Phys. Chem. Lett.* 3, 3765 (2012)
- *J. Chem. Theory Comput.* 9, 1103 (2013)
- *J. Chem. Theory Comput.* 9, 4844 (2013)
- *J. Chem. Theory Comput.* 9, 5395 (2013)
- *J. Chem. Theory Comput.* 10, 1599 (2014)
- *J. Chem. Theory Comput.* 10, 2906 (2014)
- *J. Chem. Theory Comput.* 11, 1145 (2015)
- *J. Chem. Phys.* 145, 194594 (2016)

Machine Learning and Software Development

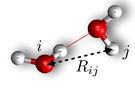


Permutationally Invariant Polynomials

Descriptors

Symmetrized monomials up to 4th degree

$$\xi = e^{-kR_{ij}} \text{ or } \xi = e^{-kR_{ij}}/R_{ij}$$



Output energy

$$V_{short}^{2B/3B} = \sum_n c_n P_n$$

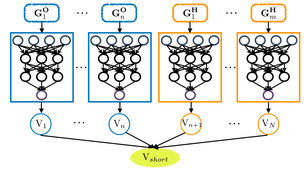
Neural Networks

Descriptors

Symmetry functions

$$G_i^{rad} = \sum_{j \neq i} e^{-\eta(R_{ij}-R_s)^2}$$

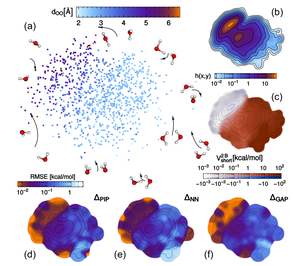
$$G_i^{ang} = 2^{1-\zeta} \sum_{\substack{j \neq i \\ k \neq i,j}} (1+\lambda\cos\theta_{ijk})^{\zeta} e^{-\eta(R_{ij}+R_{ik}+R_{jk})^2}$$

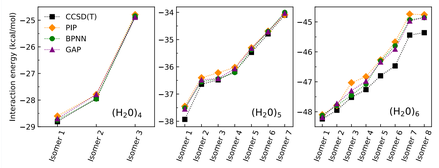


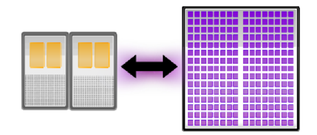
Output energy

$$V_{short}^{2B/3B} = \sum_i^N \tanh\left(\tanh\left(\tanh\left(\mathbf{G}_i\mathbf{W}_{i1}\right)\mathbf{W}_{i2}\right)\mathbf{W}_{i3}\right)\mathbf{W}_{i4}$$

Implementation







Multicore CPU
Fast Serial Processing

Latency-optimized cores

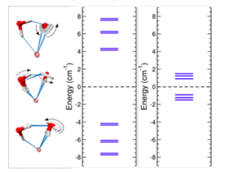
Manycore GPU
Scalable Parallel Processing

Throughput-optimized cores

- *J. Chem. Phys.* 148, 241725 (2018)
- *MB-MD in OpenMM: http://paesanigroup.ucsd.edu/software/mbpol_openmm.html*
- *MB-MD in i-PI: http://paesanigroup.ucsd.edu/software/mbpol_ipi.html*
- *MB-MD in dedicated parallel software: https://github.com/chemphys/clusters_ultimate*

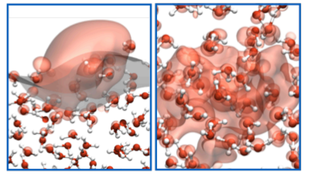
Applications

Ion-Mediated Hydrogen-Bond Tunneling



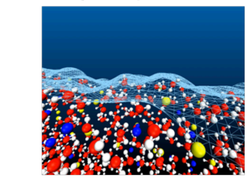
- *J. Chem. Theory Comput.* 12, 2698 (2016)
- *J. Chem. Phys.* 148, 102321 (2018)

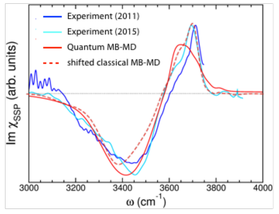
Electrons in Water



- *Nat. Commun.* 9, 247 (2018)

Chemistry at Aqueous Interfaces





- *J. Am. Chem. Soc.* 138, 3912 (2016)
- *J. Phys. Chem. Lett.* 8, 2579 (2017)
- *J. Phys. Chem. B* 122, 4356 (2018)

OAC-1642336

NSF

MolSSI

ACI-1642336

XSEDE
Extreme Science and Engineering Discovery Environment