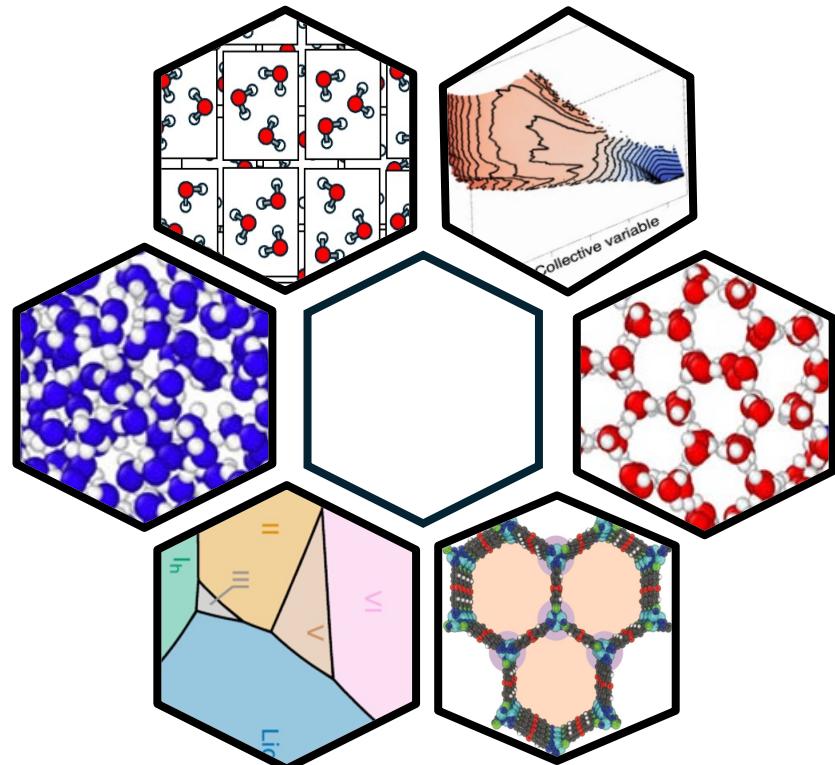


Data-Driven Many-Body Potentials



Suman Saha
Graduate Student @ UC San Diego

28th June 2024

Recap from Molecular Dynamics

$$\frac{dU(q)}{dq} = -\frac{dp}{dt}$$

$$E^{\text{FF}} = \sum_{\text{bonds}} k_b(r - r_0)^2 + \sum_{\text{angles}} k_\theta(\theta - \theta_0)^2 + \sum_{\text{dihedrals}} \frac{V_n}{2} [1 + \cos(n\phi - \gamma)] \\ + \sum_{i < j} \left[4\epsilon_{ij} \left(\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right) + \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \right]$$

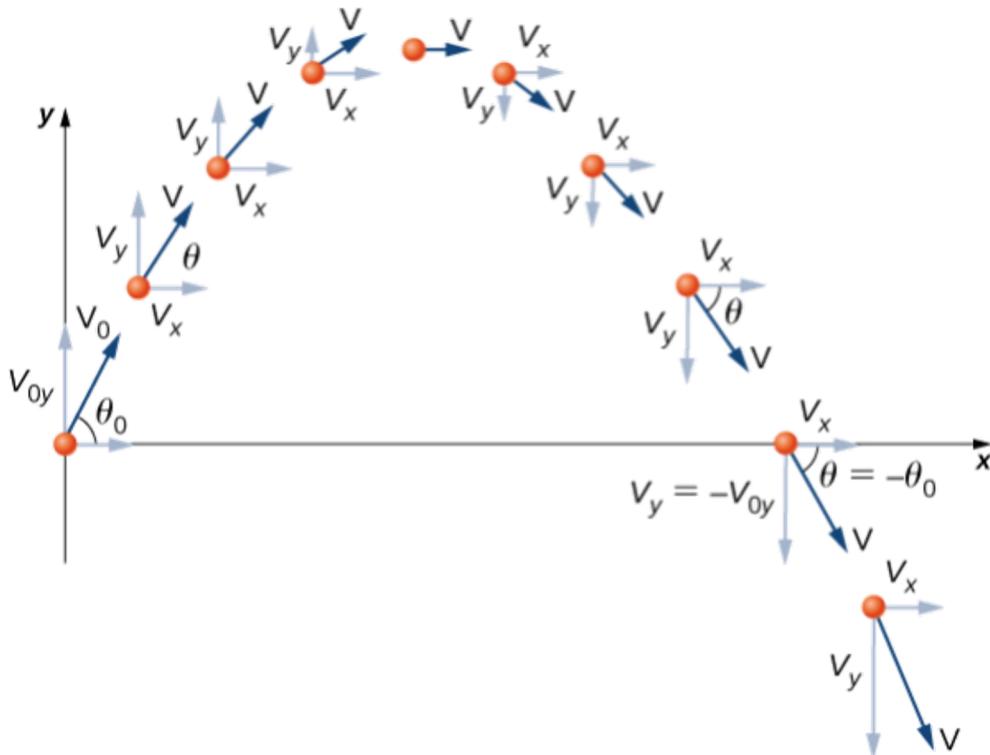
Ensemble Average:

$$\langle f \rangle_e = \frac{1}{N} \sum_i^N f(i)$$

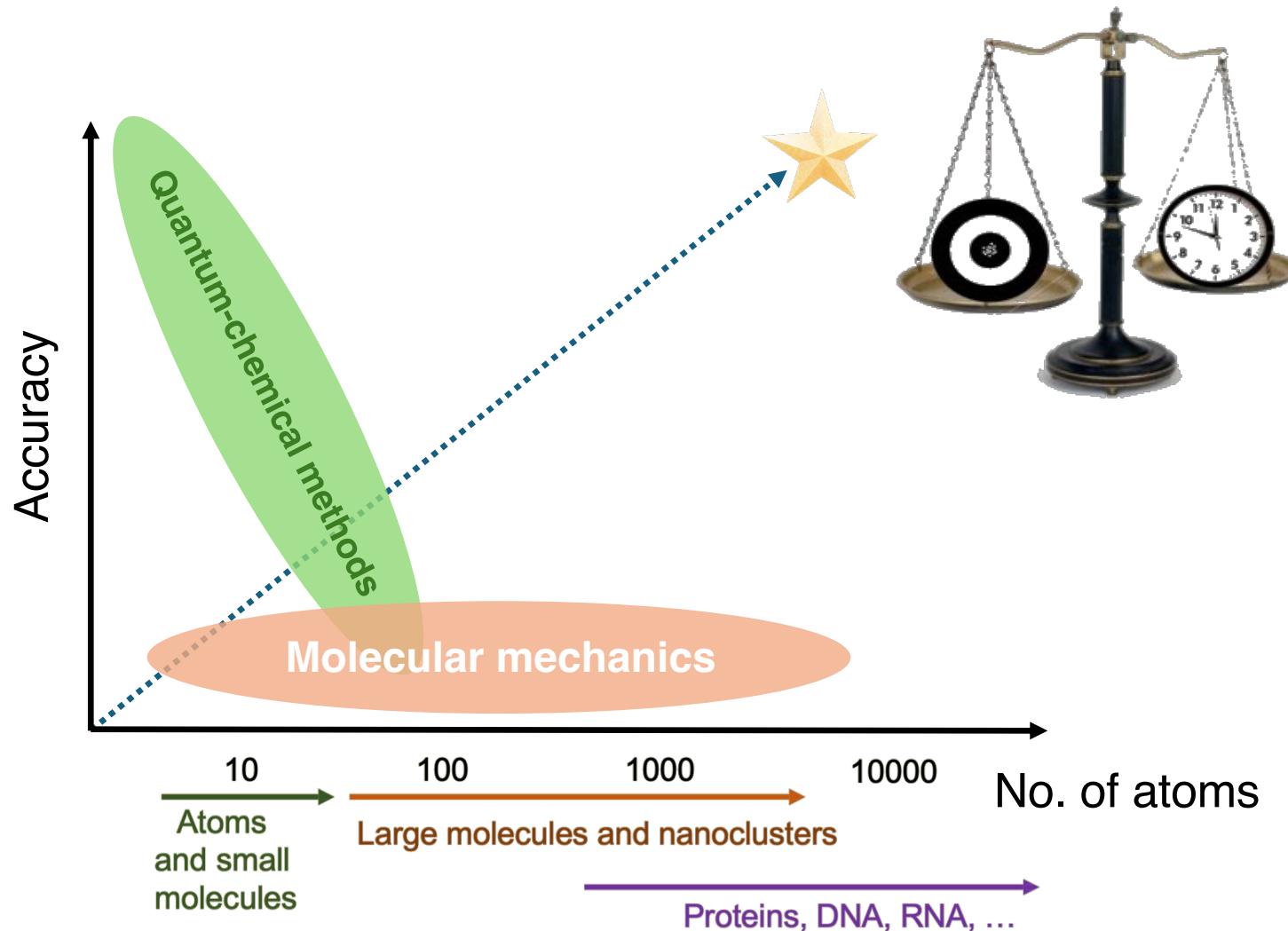
Ergodic Hypothesis

==== Time Average:

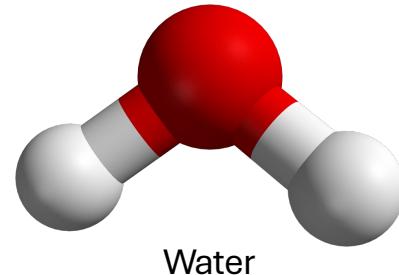
$$\langle f \rangle_t = \frac{1}{T} \sum_i^T f(i)$$



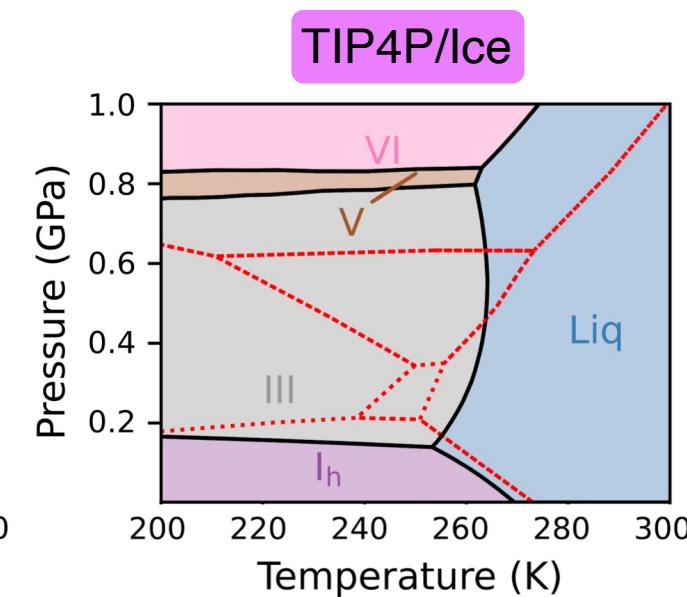
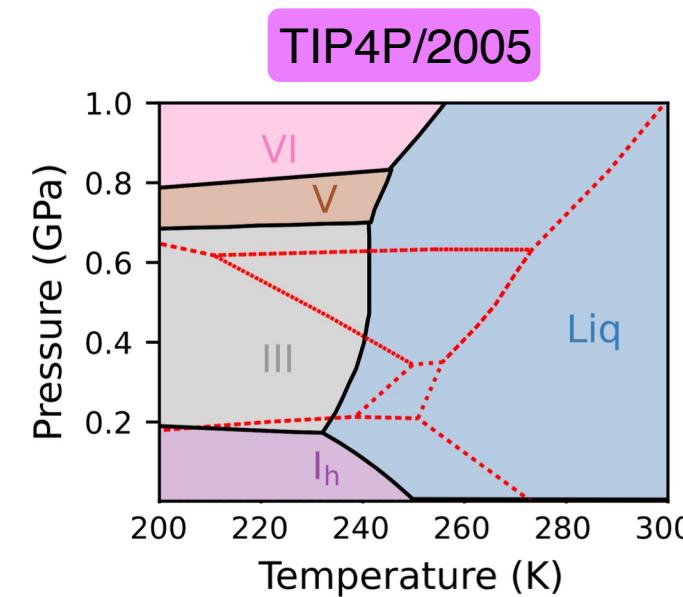
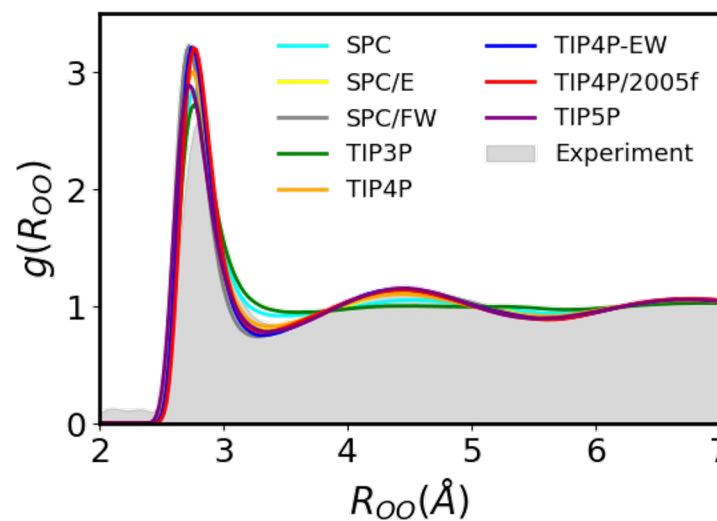
Balancing Accuracy and System Size



Evaluating the Efficacy of Current Force Fields



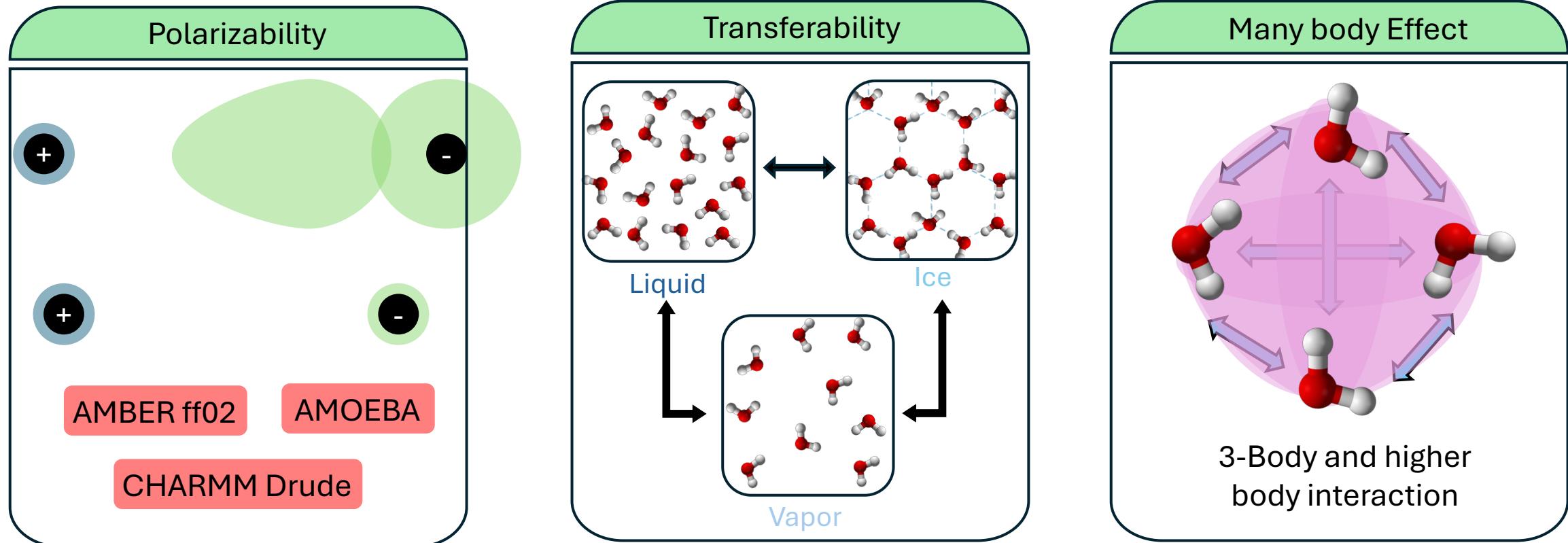
SPC SPC/E SPC/Fw TIP3P TIP3P-Ew TIP5P TIP5P-E
TIP4P TIP4P-Ew TIP4P/Ice TIP4P/2005 TIP4P/2005f q-TIP4P/F



Abascal, Jose LF, and Carlos Vega, The Journal of chemical physics 123.23 (2005).

Bore, Sigbjørn L., et al, The Journal of Chemical Physics 157.5 (2022).

Limitations of Conventional Force Fields



Wang, Zhi-Xiang, et al., *Journal of computational chemistry* 27.6 (2006)

Shi, Yue, et al., *Journal of chemical theory and computation* 9.9 (2013).

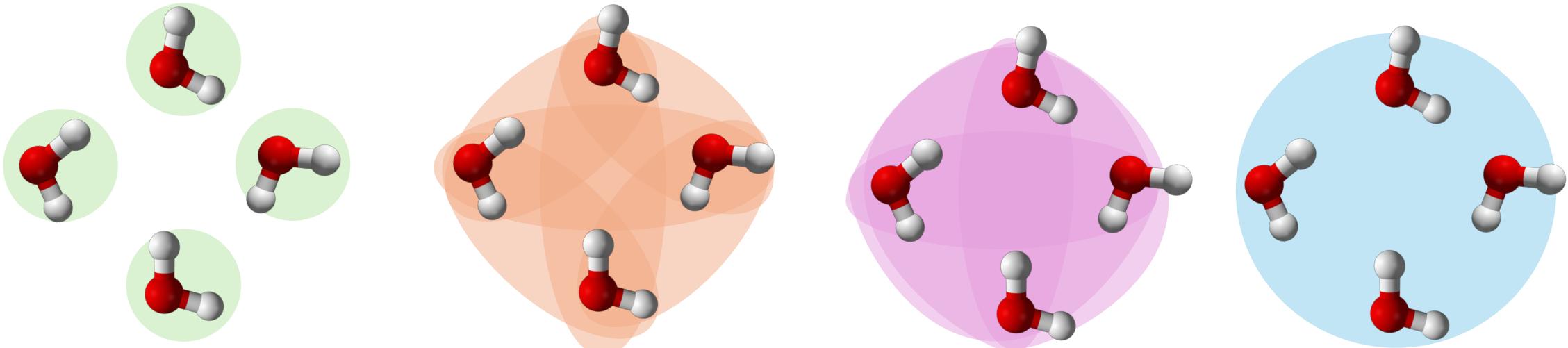
Vanommeslaeghe, Kenno, and A. D. MacKerell Jr., *(BBA)-General Subjects* 1850.5 (2015).

Solution: A different formalism of Energy

$$E^{\text{FF}} = \sum_{\text{bonds}} k_b(r - r_0)^2 + \sum_{\text{angles}} k_\theta(\theta - \theta_0)^2 + \sum_{\text{dihedrals}} \frac{V_n}{2} [1 + \cos(n\phi - \gamma)] + \sum_{i < j} \left[4\epsilon_{ij} \left(\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right) + \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \right]$$

Many Body Formalism

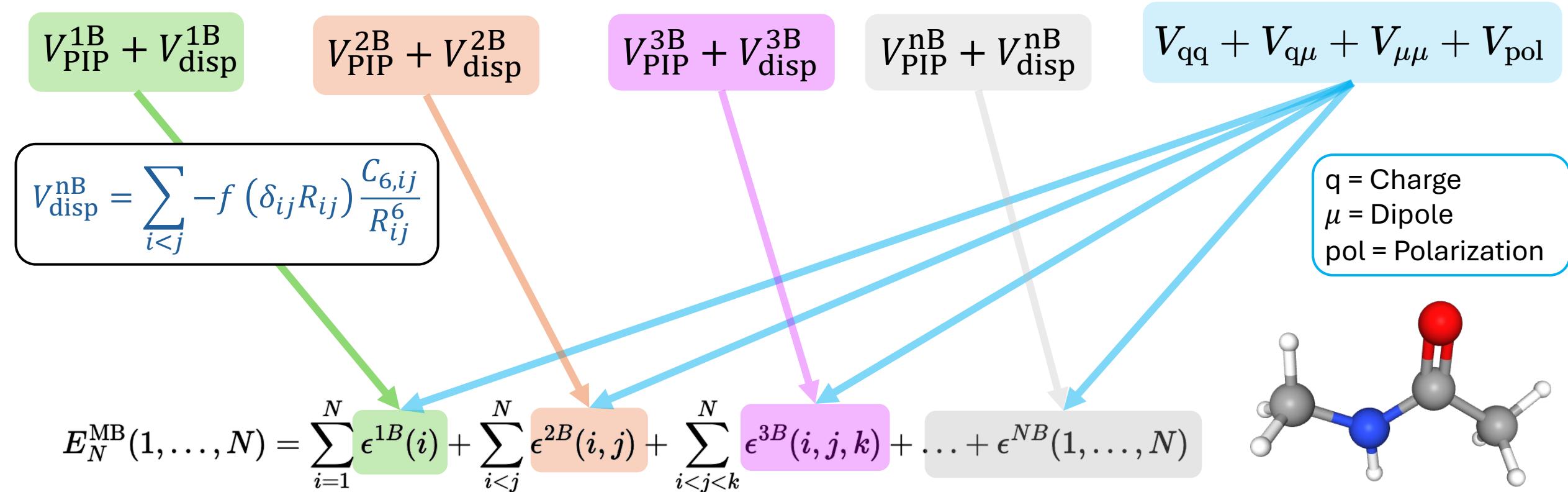
$$E_N^{\text{MB}}(1, \dots, N) = \sum_{i=1}^N \epsilon^{1B}(i) + \sum_{i < j}^N \epsilon^{2B}(i, j) + \sum_{i < j < k}^N \epsilon^{3B}(i, j, k) + \dots + \epsilon^{NB}(1, \dots, N)$$



Implementation of MB-nrg Model

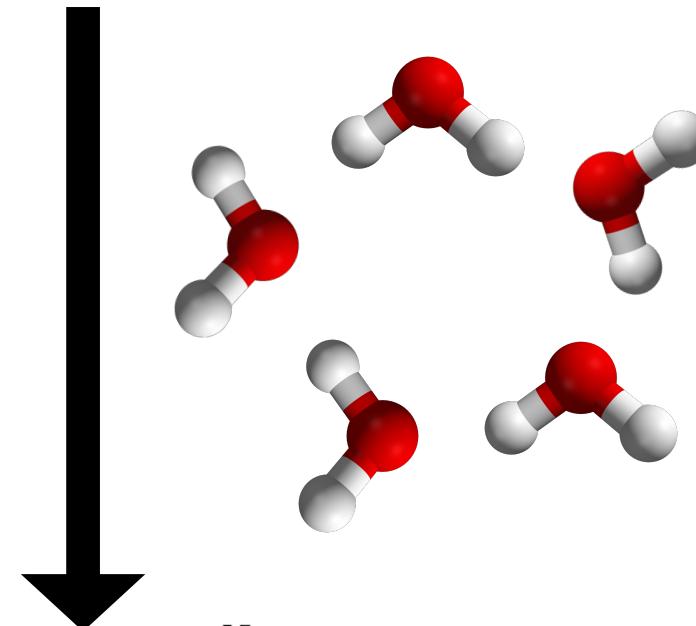
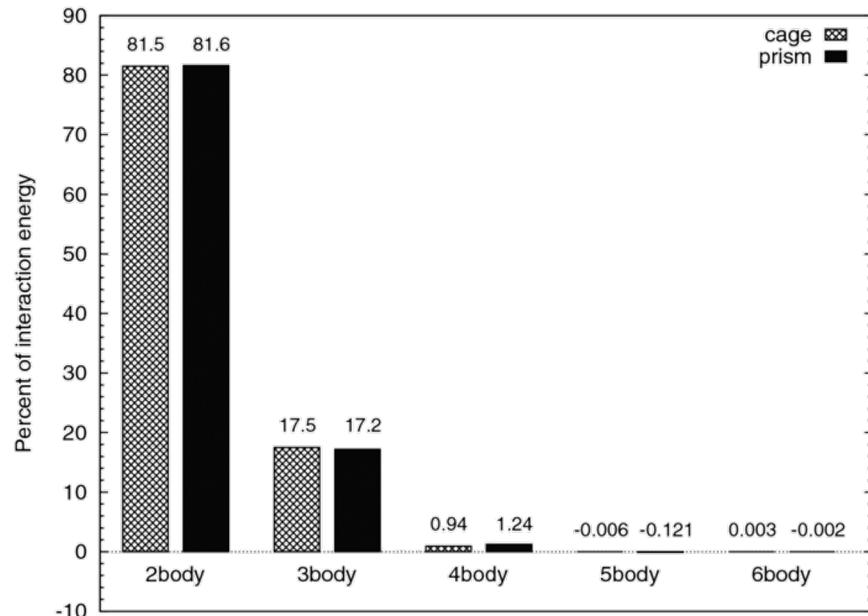
MBX

$$E_N^{MB}(1, \dots, N) = \sum_{i=1}^N V^{1B}(i) + \sum_{i < j} V^{2B}(i, j) + \sum_{i < j < k} V^{3B}(i, j, k) + \dots + \sum_{i < j < k < \dots < n} V^{nB}(i, j, k, \dots, n) + V_{\text{elec}}(1, \dots, N)$$



Energy Approximations in Water Clusters

$$E_N^{MB}(1, \dots, N) = \sum_{i=1}^N V^{1B}(i) + \sum_{i < j} V^{2B}(i, j) + \sum_{i < j < k} V^{3B}(i, j, k) + \dots + \sum_{i < j < k < \dots < n} V^{nB}(i, j, k, \dots, n) + V_{\text{elec}}(1, \dots, N)$$



For water clusters
4B and higher
terms contribute
~1%

$$E_N^{MB}(1, \dots, N) = \sum_{i=1}^N V^{1B}(i) + \sum_{i < j} V^{2B}(i, j) + \sum_{i < j < k} V^{3B}(i, j, k) + V_{\text{elec}}(1, \dots, N)$$

Góra, Urszula, et al, The Journal of chemical physics 135.22 (2011)

Energy Calculations with Polynomials

$$E_N^{MB}(1, \dots, N) = \sum_{i=1}^N V^{1B}(i) + \sum_{i < j} V^{2B}(i, j) + \sum_{i < j < k} V^{3B}(i, j, k) + V_{\text{elec}}(1, \dots, N)$$

$$V_{\text{PIP}}^{\text{nB}} = \sum_{l=1}^L c_l \cdot \eta_l(\xi_1, \xi_2, \dots, \xi_\lambda)$$

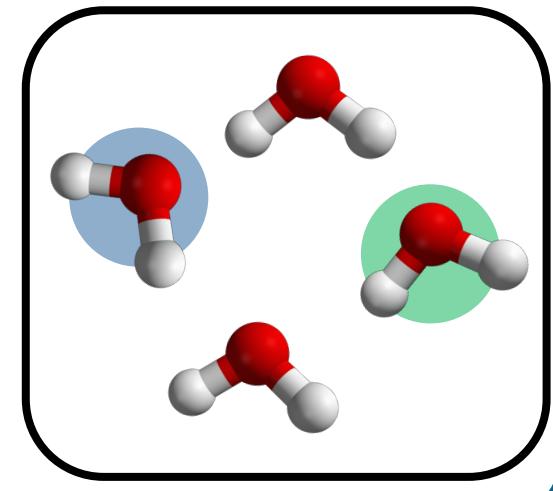
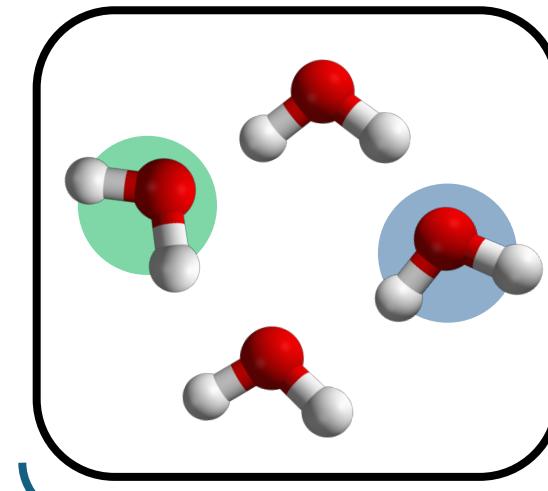
$$\xi^{\text{exp0}}(R_{mn}) = e^{-k_{\tau(mn)}(R_{mn} - d_{0,\tau(mn)})}$$

$$\xi^{\text{exp}}(R_{mn}) = e^{-k_{\tau(mn)} R_{mn}}$$

$$\xi^{\text{coul0}}(R_{mn}) = e^{-k_{\tau(mn)}(R_{mn} - d_{0,\tau(mn)})}/R_{mn}$$

$$\xi^{\text{coul}}(R_{mn}) = e^{-k_{\tau(mn)} R_{mn}}/R_{mn}$$

Permutationally Invariant Polynomials



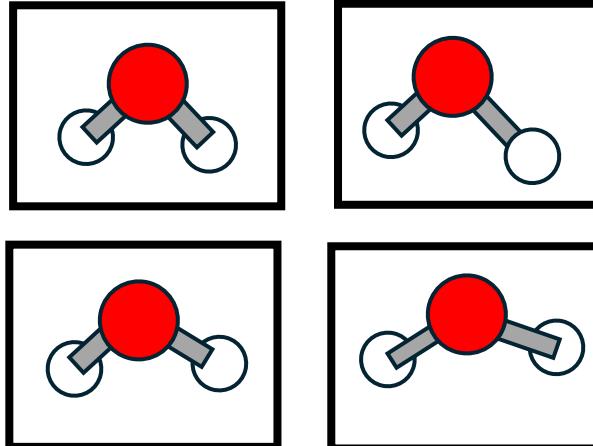
Same Energy

Steps to Construct a MB-nrg Model

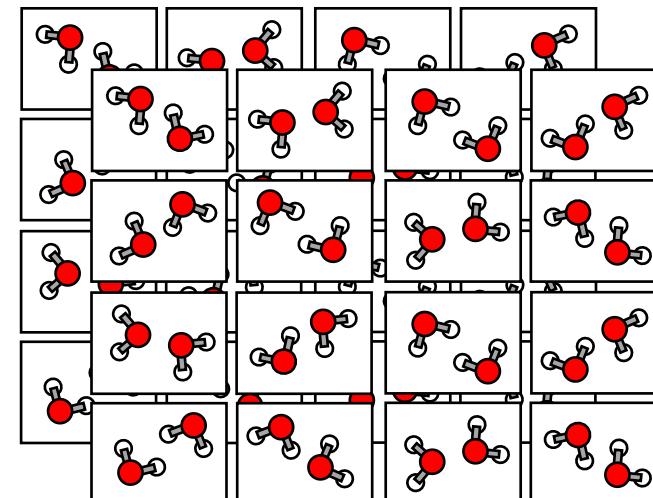
MB-Fit

$$E_N^{MB}(1, \dots, N) = \sum_{i=1}^N V^{1B}(i) + \sum_{i < j} V^{2B}(i, j) + \sum_{i < j < k} V^{3B}(i, j, k) + V_{\text{elec}}(1, \dots, N)$$

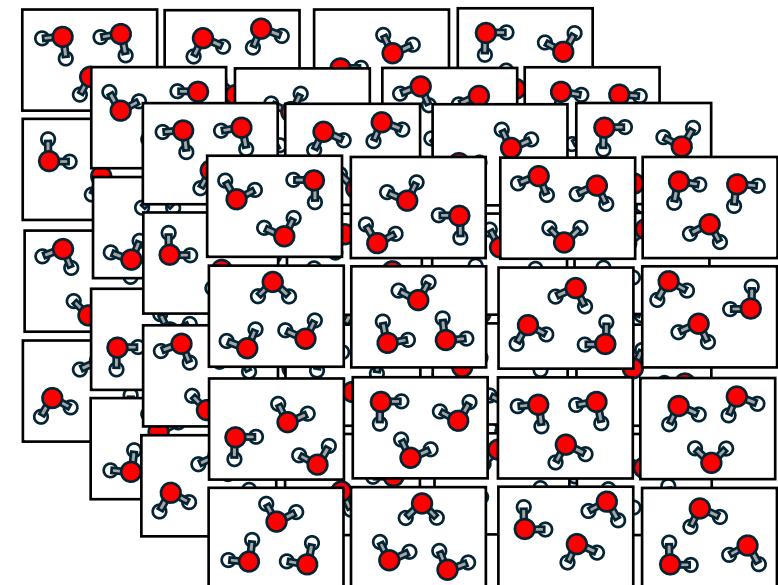
$$V_{\text{PIP}}^{\text{nB}} = \sum_{l=1}^L c_l \cdot \eta_l(\xi_1, \xi_2, \dots, \xi_\lambda)$$



1 body training set



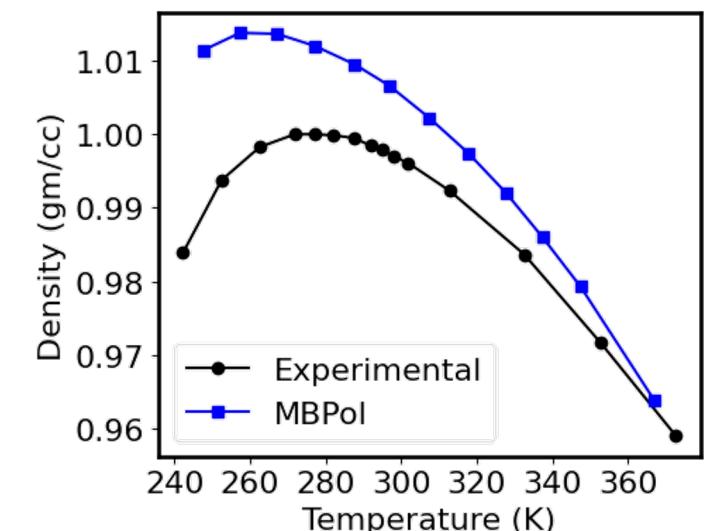
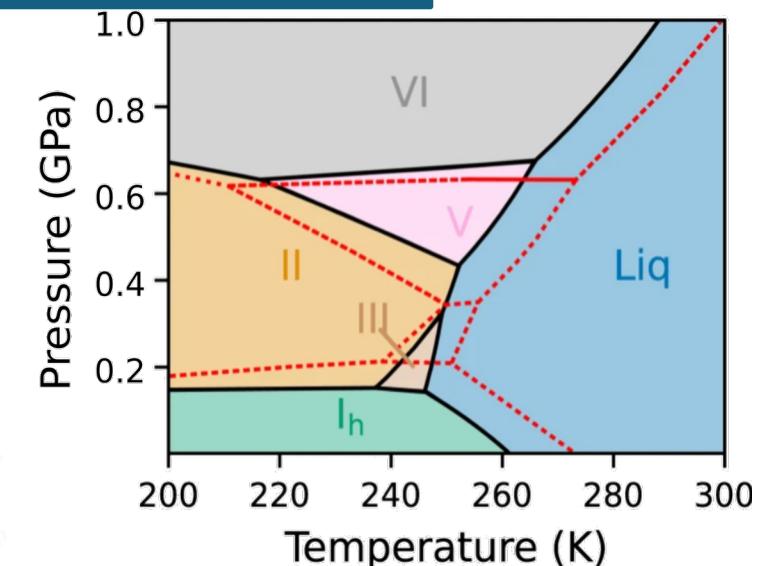
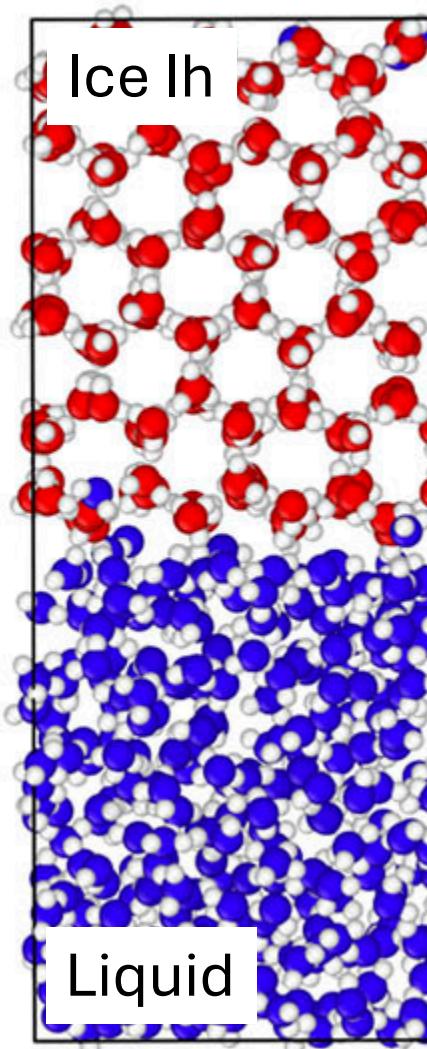
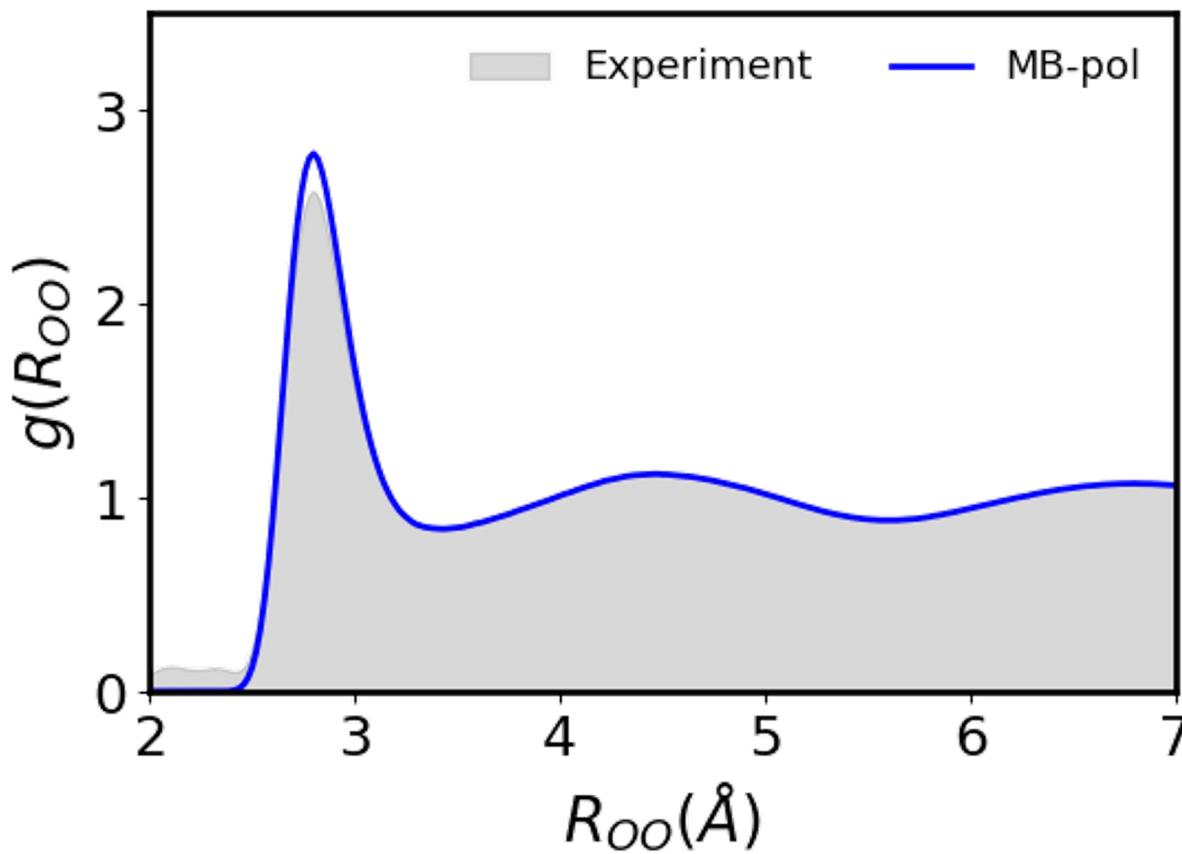
2 body training set



3 body training set

Bull-Vulpe, Ethan F., et al. The Journal of Chemical Physics 155.12 (2021)

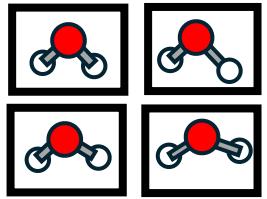
MB-nrg Water Model: MB-pol



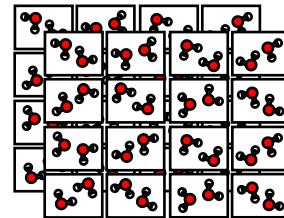
Bore, Sigbjørn Løland, and Francesco Paesani, Nature communications 14.1 (2023): 3349.

Dasgupta, Saswata, et al., Nature communications 12.1 (2021): 6359.

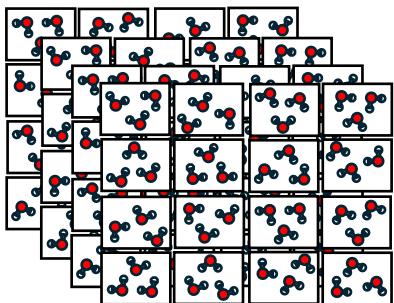
MB-nrg Water Model: MB-pol(2023)



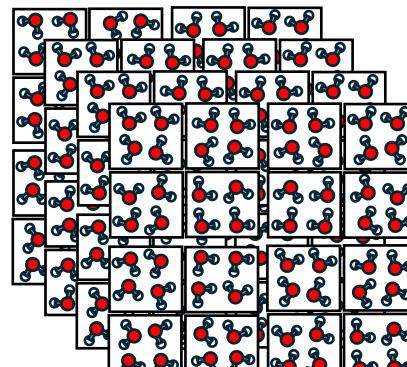
1 body training set



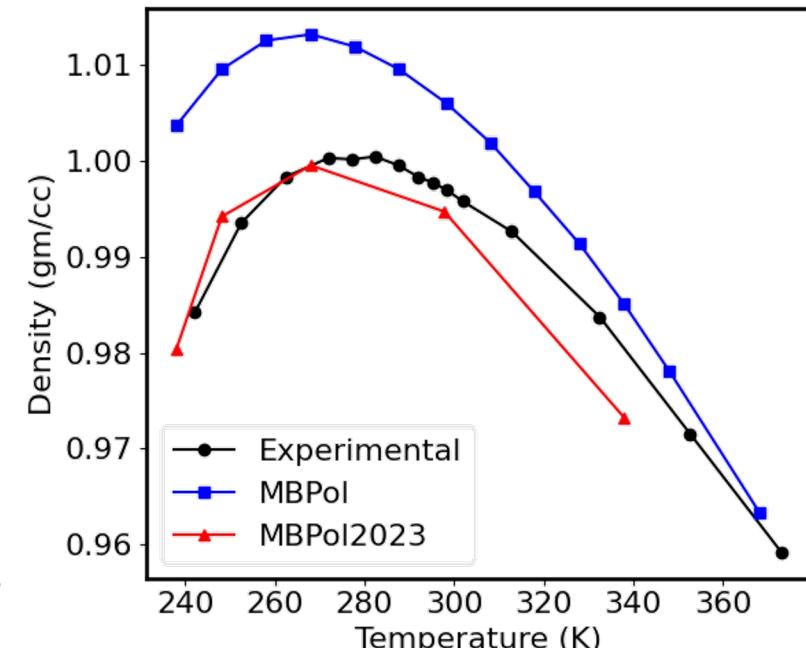
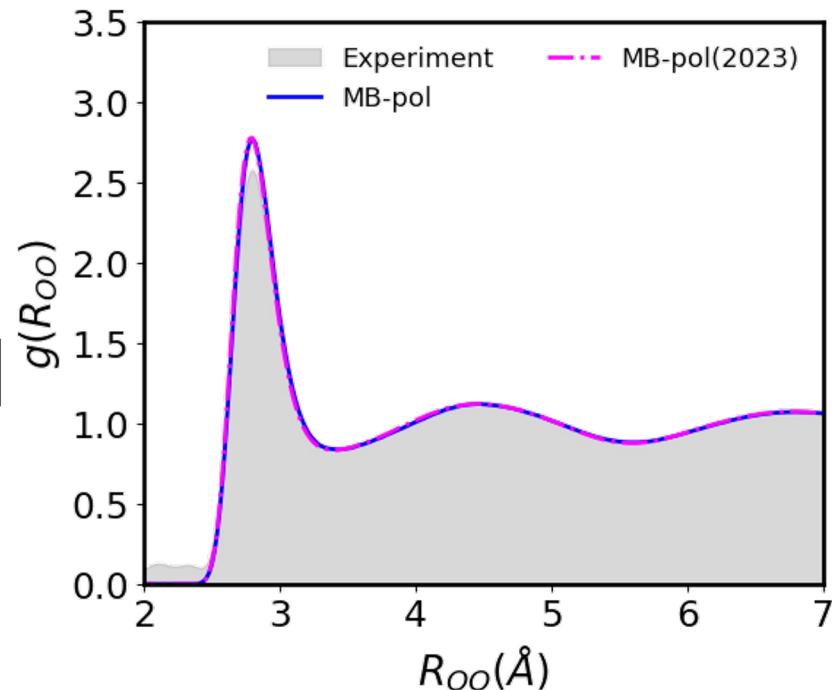
2 body training set



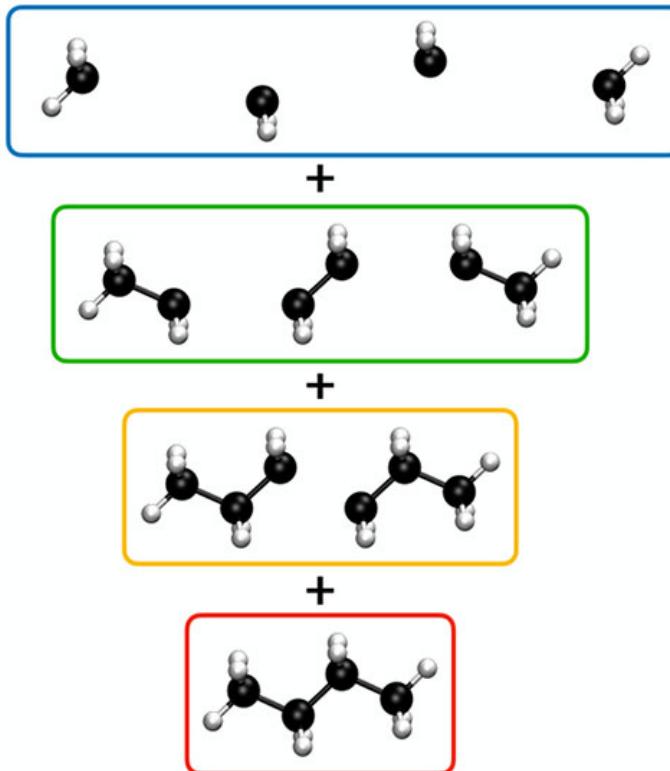
3 body training set



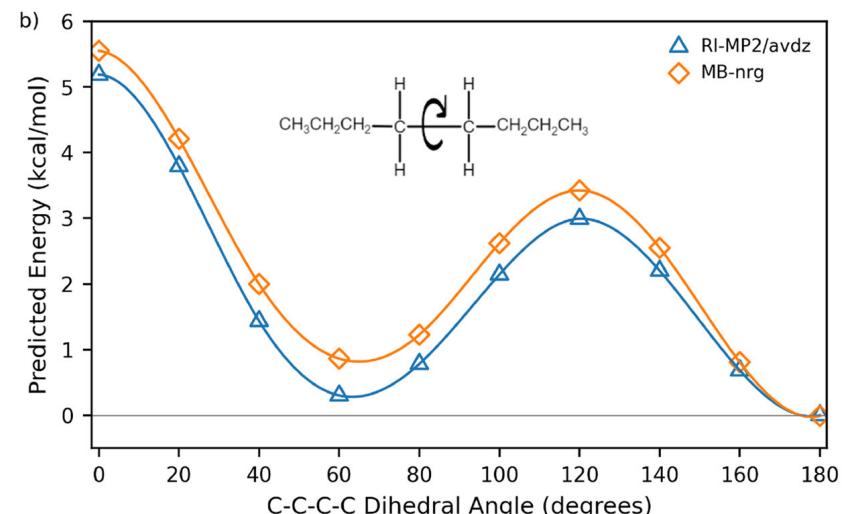
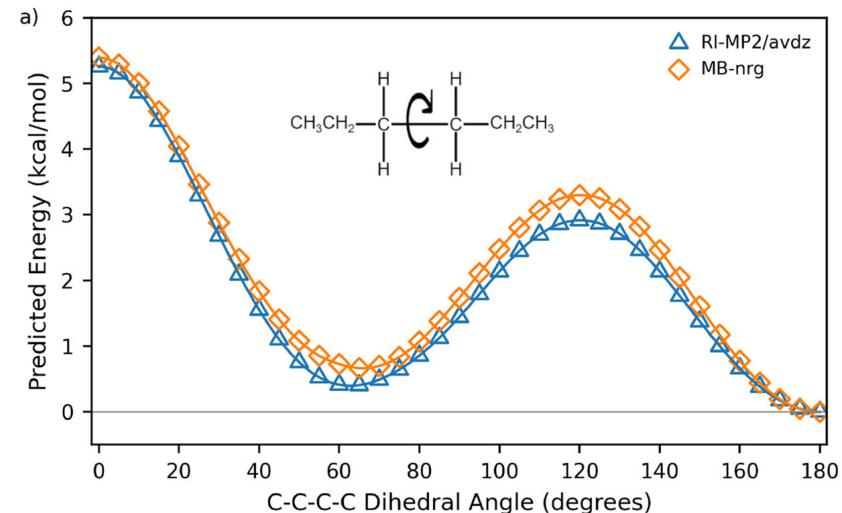
4 body training set



MB-nrg model for Alkanes



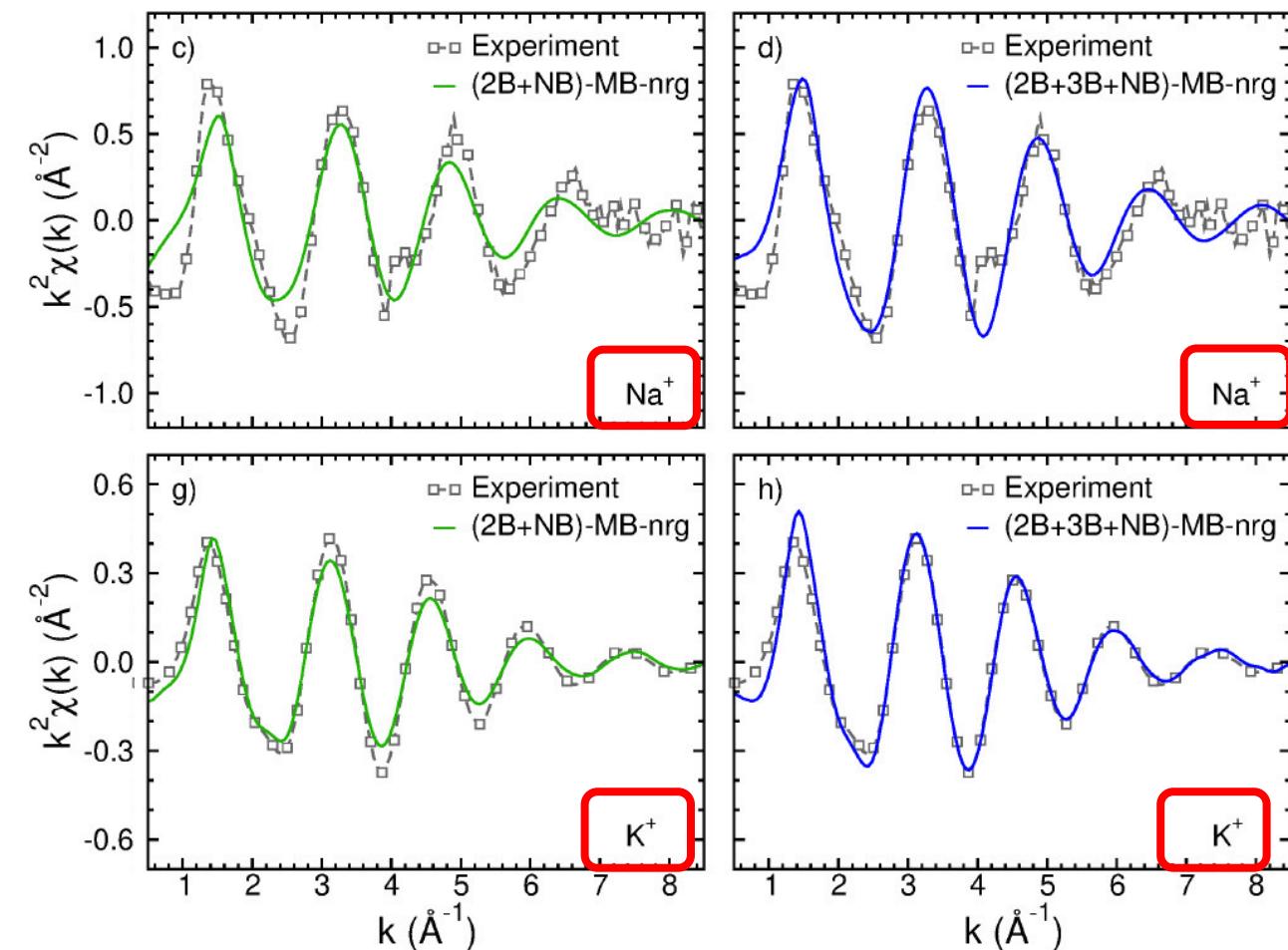
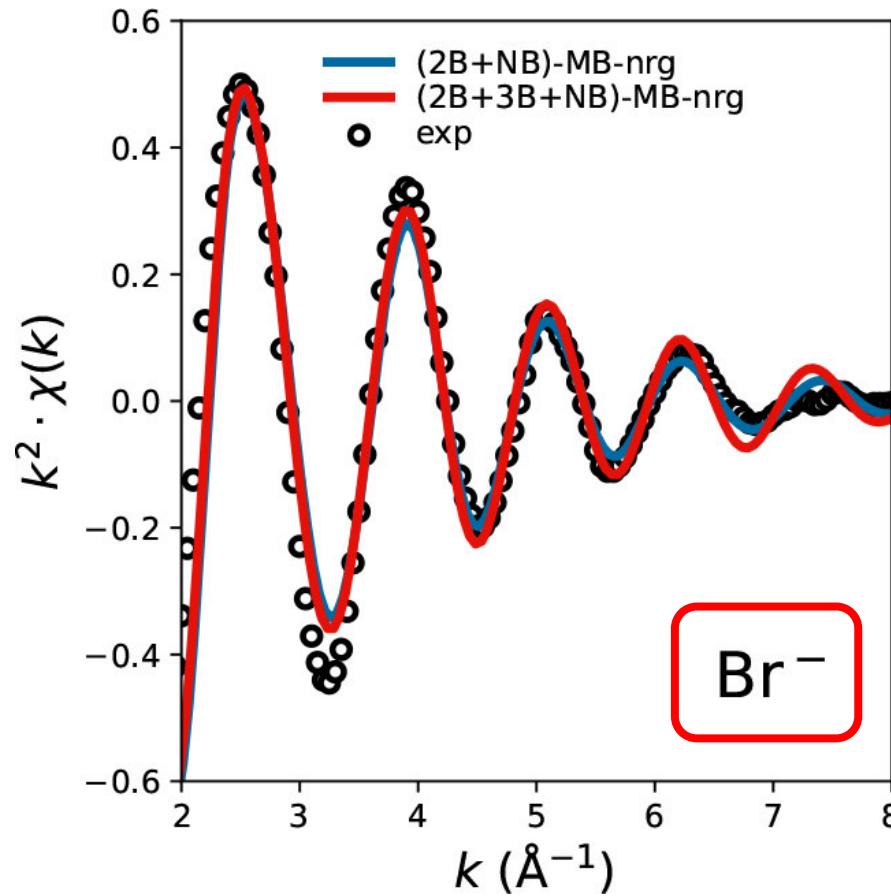
$$\begin{aligned} E = & \sum_i \epsilon^{1\text{B}}(i) \\ & + \sum_{i < j} \epsilon^{2\text{B}}(i, j) \\ & + \sum_{i < j < k} \epsilon^{3\text{B}}(i, j, k) \\ & + \dots \\ & + \epsilon^{\text{NB}}(1, 2, 3, 4, \dots, N) \end{aligned}$$



Bull-Vulpe, Ethan F., et al., Journal of Chemical Theory and Computation 19.14 (2022): 4494-4509.

MB-nrg Model for Ions

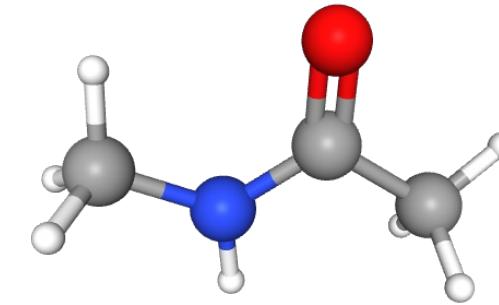
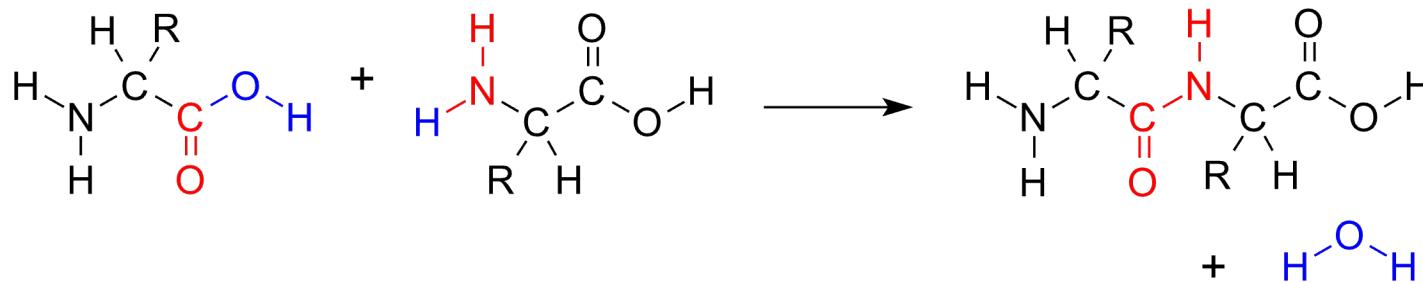
K-edge EXAFS spectrum



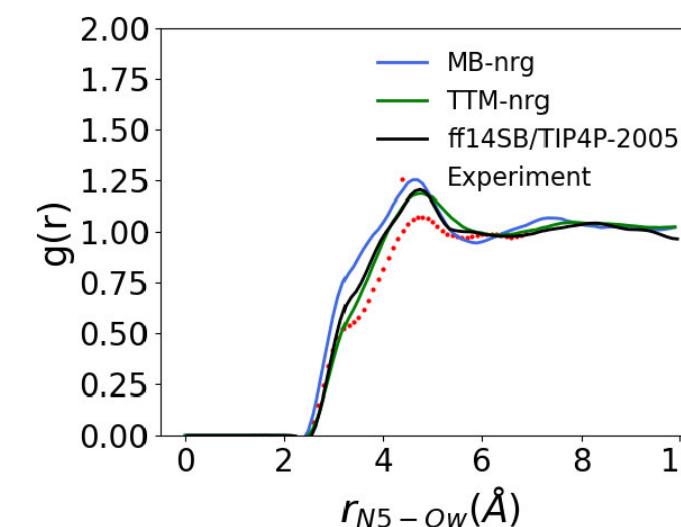
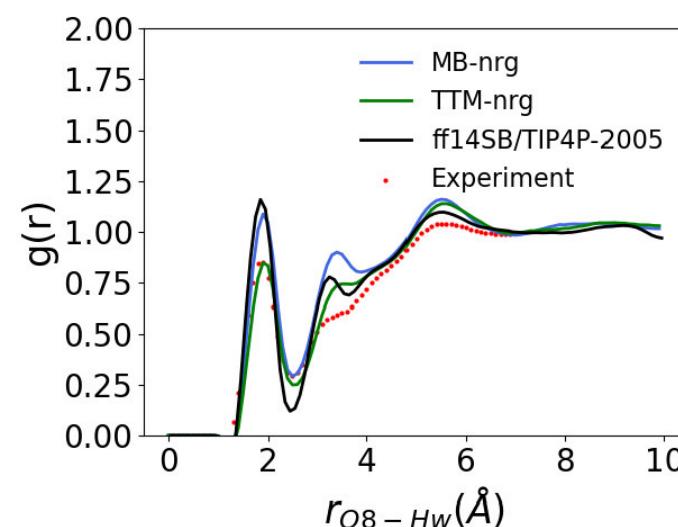
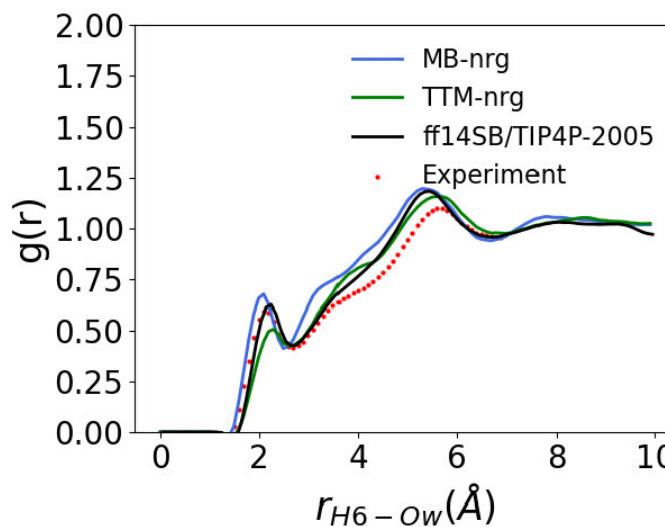
Caruso, Alessandro, et al., The Journal of Physical Chemistry B 126.41 (2022): 8266-8278.

Zhuang, Debbie, et al., The Journal of Physical Chemistry B 126.45 (2022): 9349-9360.

MB-nrg Model for Protein Backbone



N-Methylacetamide (NMA)



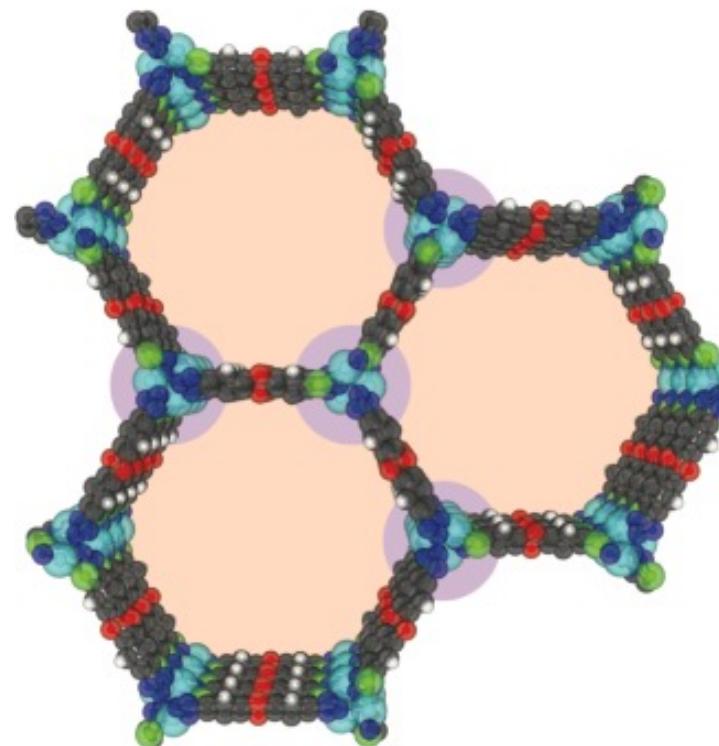
Zhou, Ruihan, Marc Riera, and Francesco Paesani, Journal of Chemical Theory and Computation 19.13 (2023): 4308-4321.

MB-nrg Model in MOFs

Metal Organic Frameworks

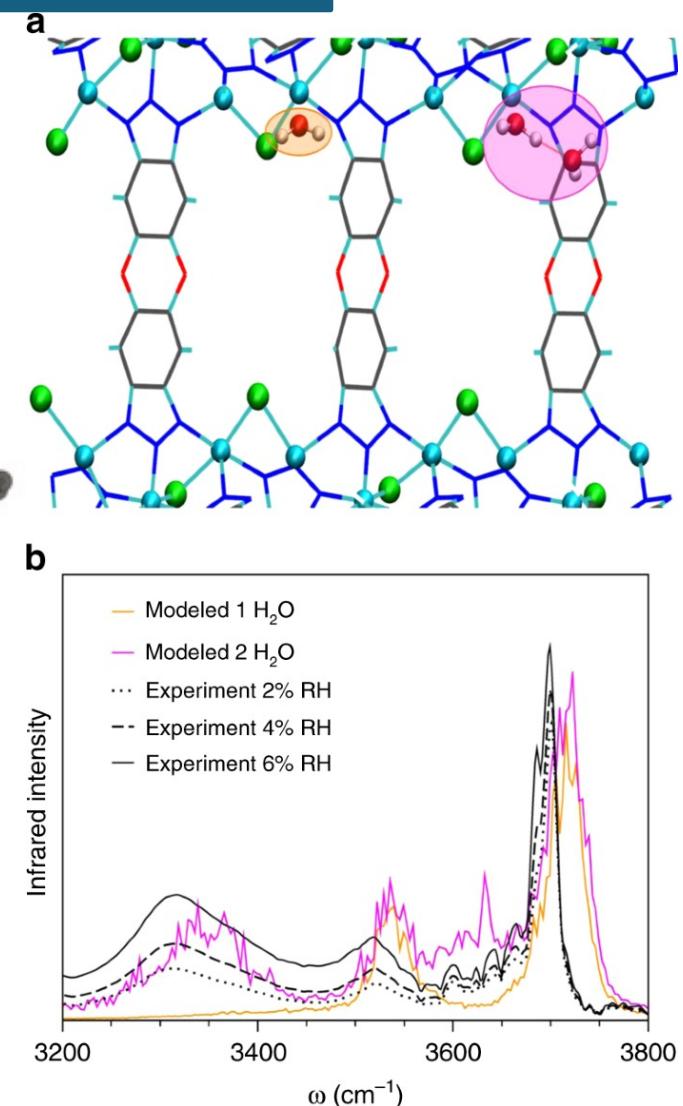
It will try to polarize the anion and water molecules

Polarizable model is needed



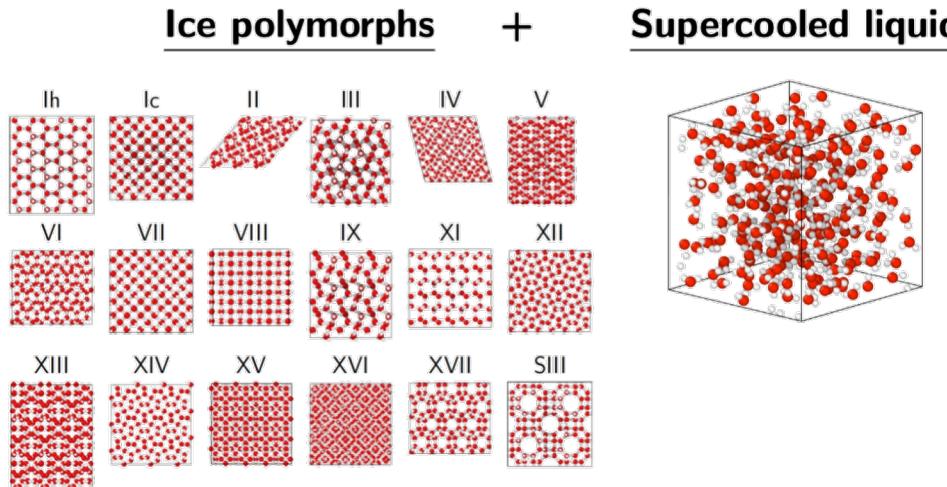
$\text{Co}_2\text{Cl}_2\text{BTDD}$

Rieth, Adam J., et al., Nature communications 10.1 (2019): 4771.

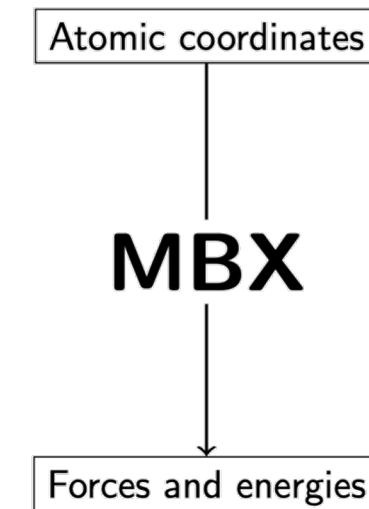


Machine Learned Potentials: DNN@MB-pol

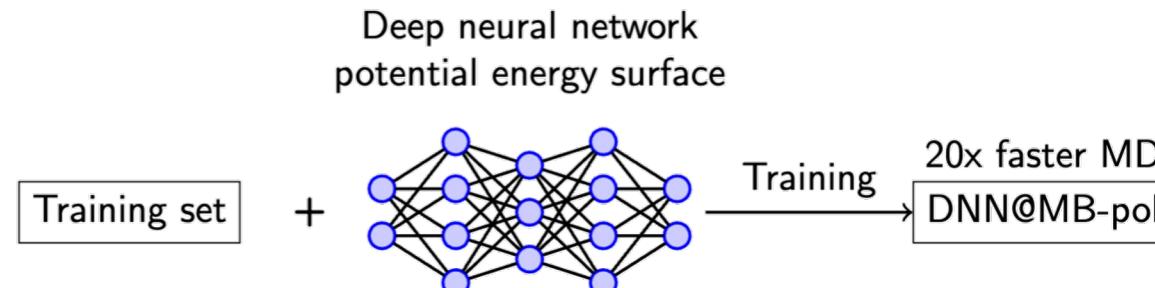
(a) Training set



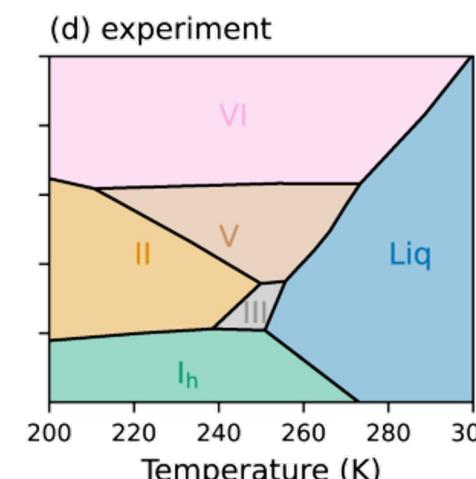
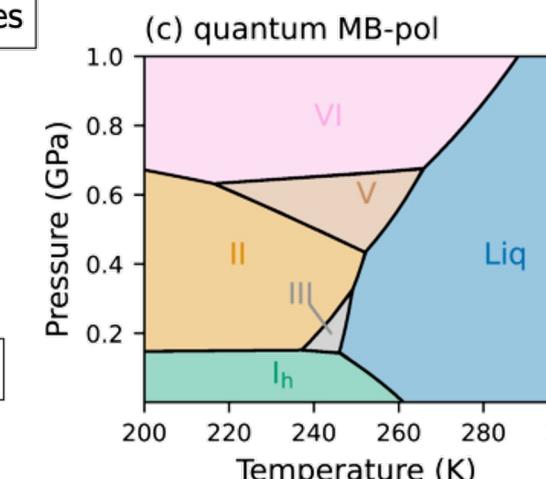
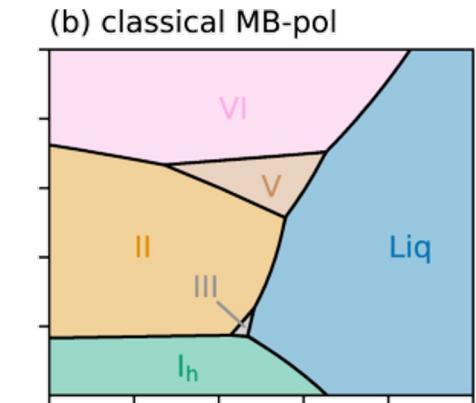
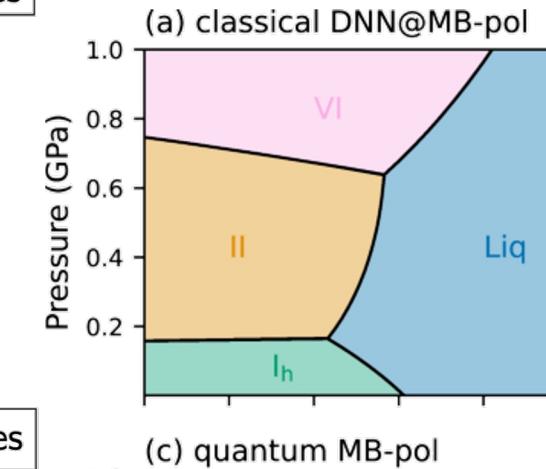
(b) MB-pol reference



(c) DNN@MB-pol model:

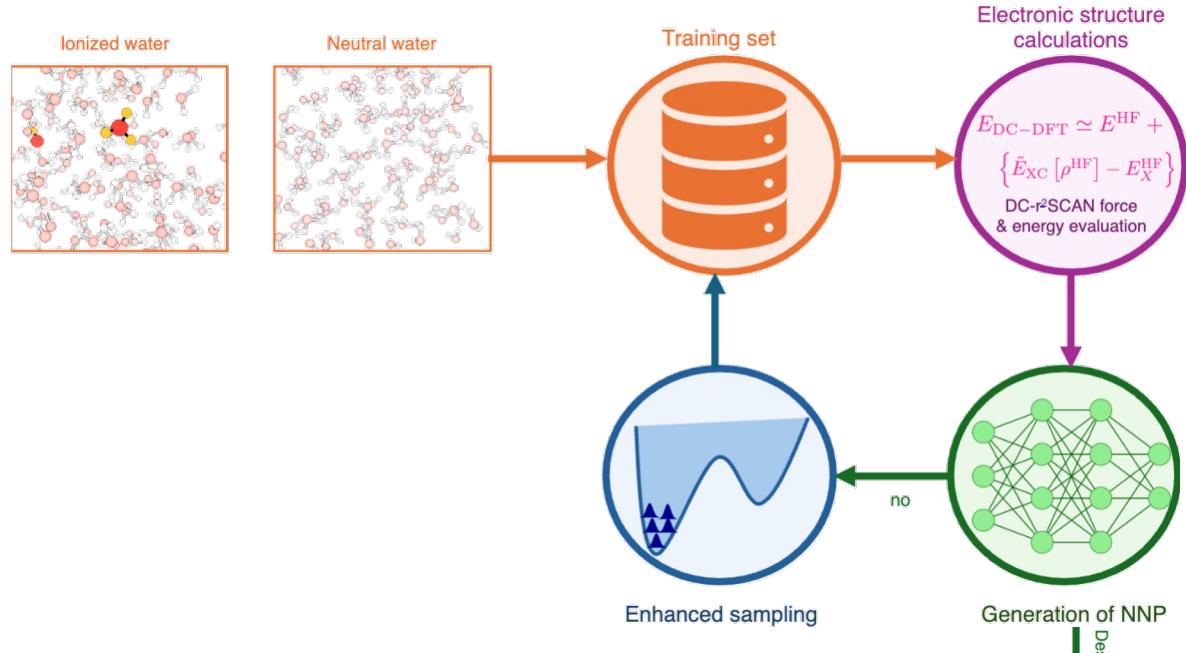


Deep Neural Network

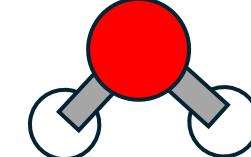
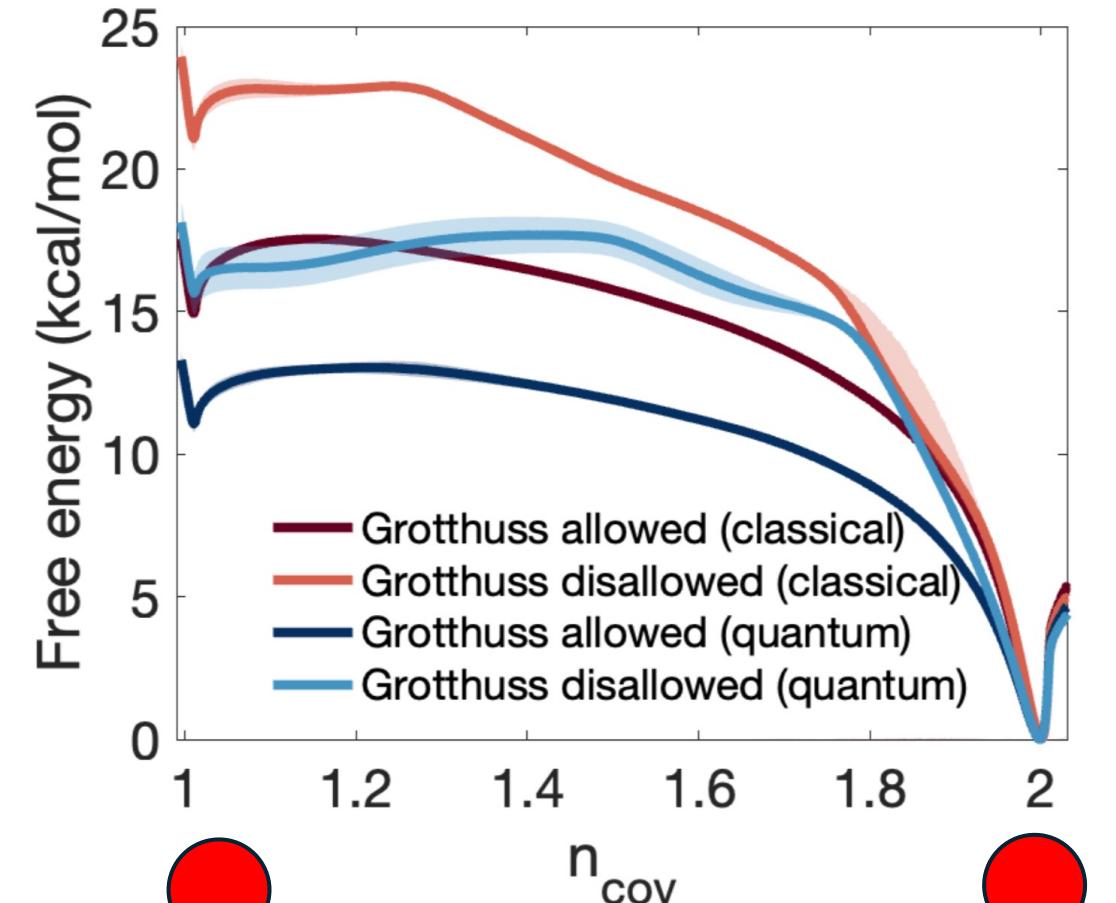
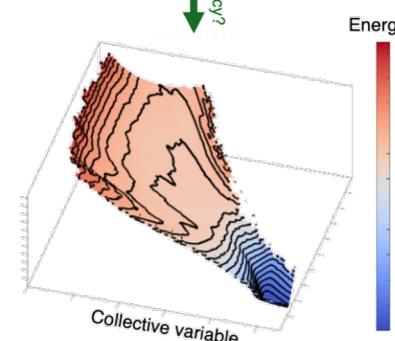


Bore, Sigbjørn Løland, and Francesco Paesani, Nature communications 14.1 (2023): 3349.

Machine Learned Potentials: DNN@DC-r²SCAN



$$pK_w = 13.91 \pm 0.04$$

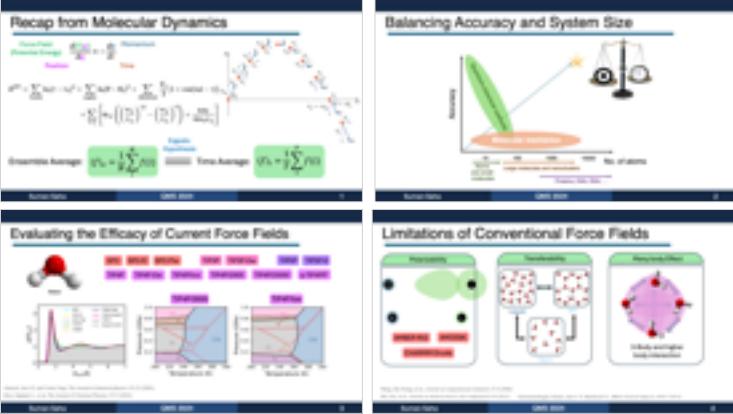


Dasgupta, Saswata, et al. Journal of Chemical Theory and Computation 18.8 (2022): 4745-4761

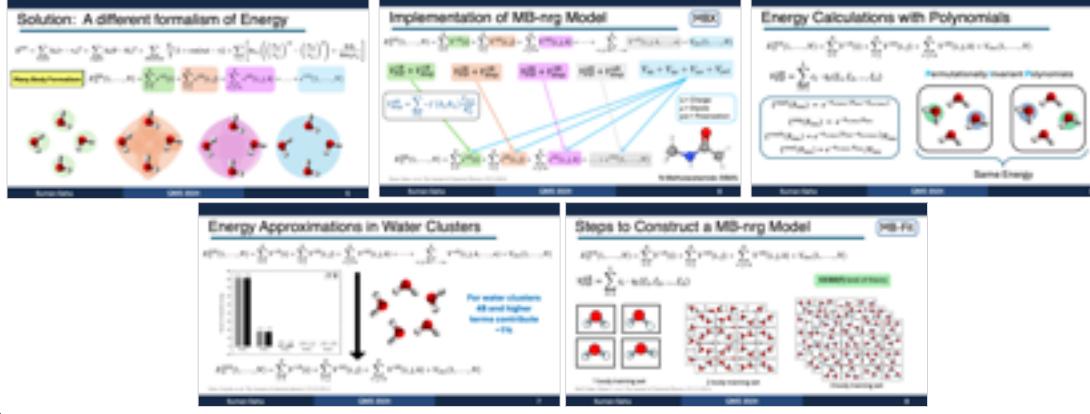
Dasgupta, Saswata, Giuseppe Cassone, and Francesco Paesani. DOI: 10.26434/chemrxiv-2024-zkz7v (2024).

Overview

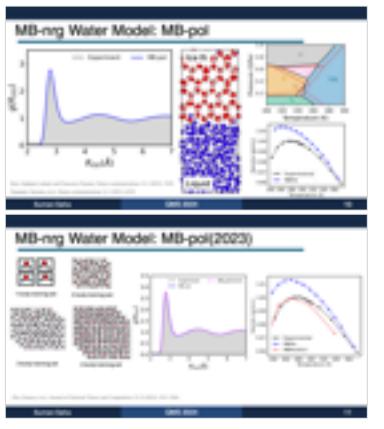
Conventional Force Fields



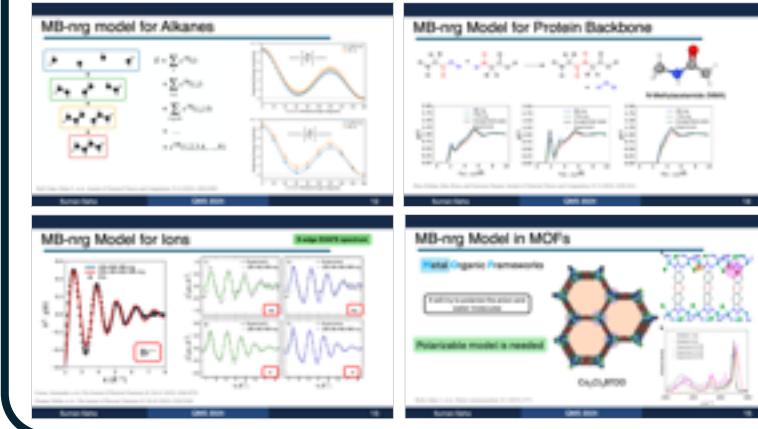
MB-nrg Theory



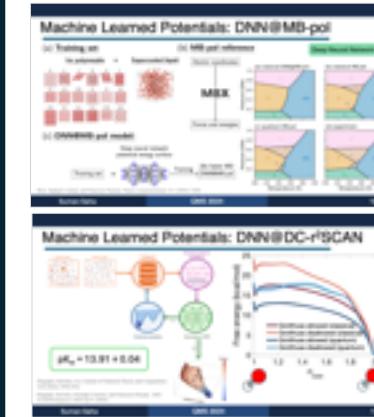
MB-pol Water



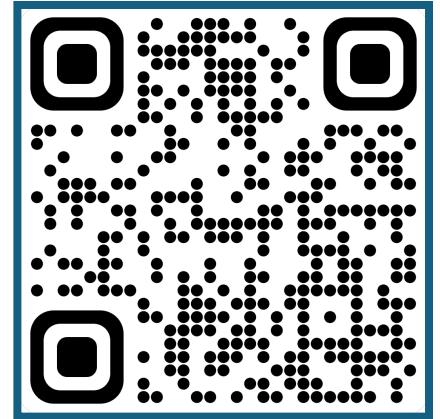
MB-nrg Models



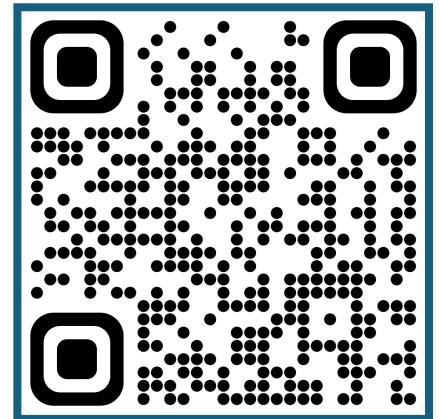
MLPs



Acknowledgements



MB-Fit



MBX

Extra Slides

Dispersion Energy

$$V_{\text{disp}}^{\text{nB}} = \sum_{i < j} -f(\delta_{ij} R_{ij}) \frac{C_{6,ij}}{R_{ij}^6}$$

$$f(\delta_{ij}, R_{ij}) = 1 - \exp(-\delta_{ij} R_{ij}) \sum_{n=0}^6 \frac{(\delta_{ij} R_{ij})^n}{n!}$$



Tang-Toennies damping function

V_elec Term

$$V_{qq} = \sum_i^N \sum_{j>i} q_i \hat{T}_{ij} q_j$$

$$V_{q\mu} = \sum_i^N \sum_{j>i} \left(\mu_i^\alpha \hat{T}_{ij}^\alpha q_j - q_i \hat{T}_{ij}^\alpha \mu_j^\alpha \right)$$

$$V_{\mu\mu} = - \sum_i^N \sum_{j>i} \mu_i^\alpha \hat{T}_{ij}^{\alpha\beta} \mu_j^\beta$$

$$V_{\text{pol}} = \frac{1}{2} \sum_{i=1}^N \mu_i \hat{\alpha}_i^{-1} \mu_i$$