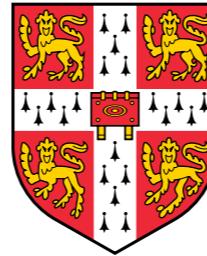


Lecture 4: fitting interatomic potentials I

Gábor Csányi



University of Cambridge
Engineering Laboratory

Interatomic potential

- Interaction of atoms when electrons are in their ground state

difficult electronic problem

$$H = \langle \Psi | \mathcal{H} | \Psi \rangle = \sum_i \frac{p_i^2}{2m_i} + \underbrace{\sum_{ij} \frac{Z_i Z_j}{|q_i - q_j|}}_{\text{"Interatomic potential"} V(q_1, q_2, \dots)} + V_{\text{el}}(q_1, q_2, \dots)$$

“Interatomic potential”
 $V(q_1, q_2, \dots)$

- This is used in *ab initio* molecular dynamics: expensive
 - Electronic problem solved explicitly
 - Potential is “global”, every atom interacts with every other

The locality assumption

- Assume that potential is separable:

$$E_{\text{tot}} = \sum_{\text{atoms } i} \varepsilon(\mathbf{r}_1 - \mathbf{r}_i, \mathbf{r}_2 - \mathbf{r}_i, \dots)$$

Finite range
atomic energy
function

$$E_{\text{tot}} = \sum_{\text{atoms } i} \varepsilon(\mathbf{r}_1 - \mathbf{r}_i, \mathbf{r}_2 - \mathbf{r}_i, \dots) + \frac{1}{2} \sum_{i,j} \hat{L}_i \hat{L}_j \frac{1}{r_{ij}} + \sum_{i,j} \frac{\sigma_{ij}}{|r_{ij}|^6}$$

Coulomb

pair dispersion

- Uncontrolled approximation
- Validity can be studied numerically
- Beginnings of rigorous treatment (C Ortner et al 2015)

Measuring Locality

- At the atom-atom (2-body) level, the decay of the Hessian is informative

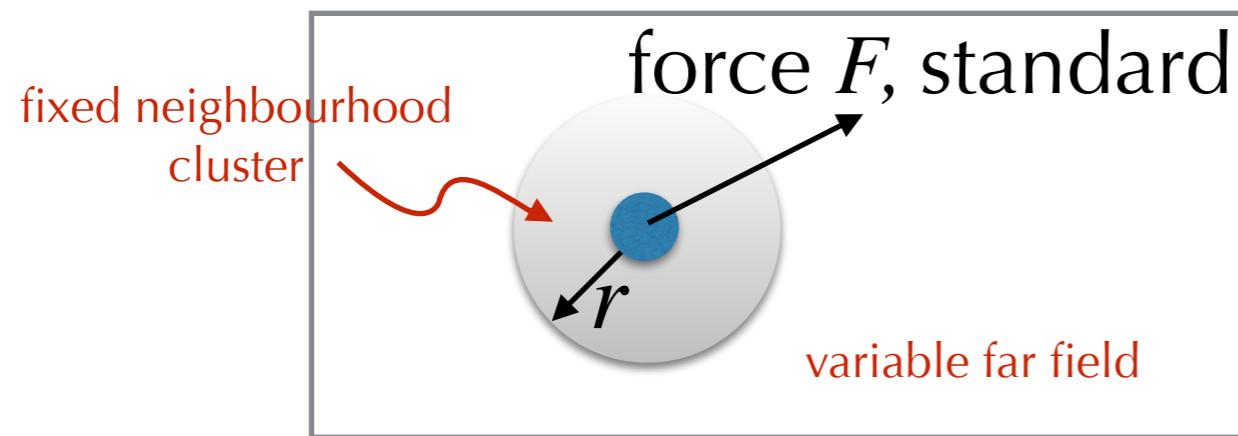
$$\frac{\partial^2 E}{\partial r_i \partial r_j} \rightarrow 0 \quad \text{as } |r_i - r_j| \rightarrow \infty$$

- Not the whole story: 3-body and >3-body effects, but generally smaller and shorter range
- Asymptotically, in simplified QM models:

$$\frac{\partial^2 E}{\partial r_i \partial r_j} \sim e^{-c|r_i - r_j|}$$

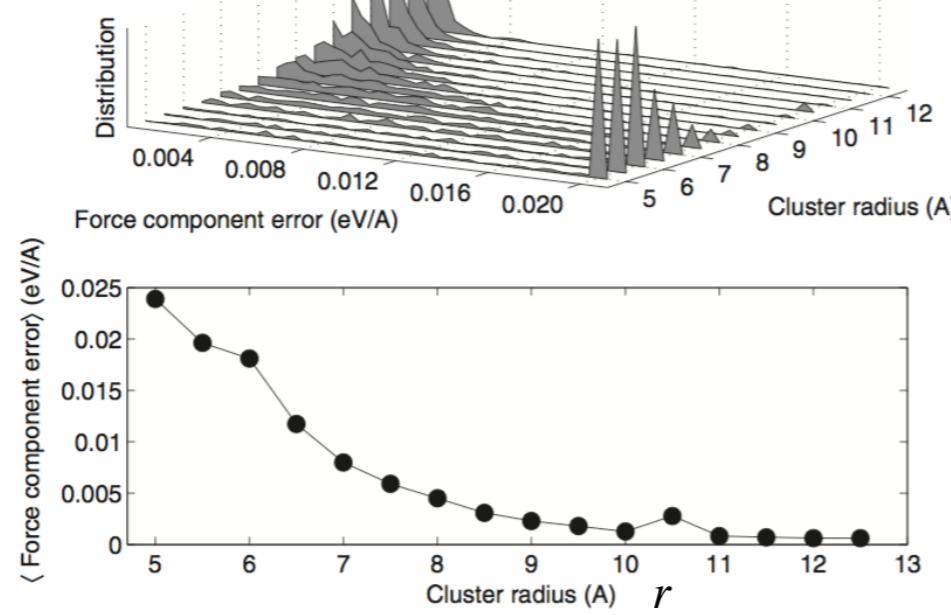
- Insulators: c scales with the band gap
Metals: c^2 scales with temperature

Numerical evidence for locality

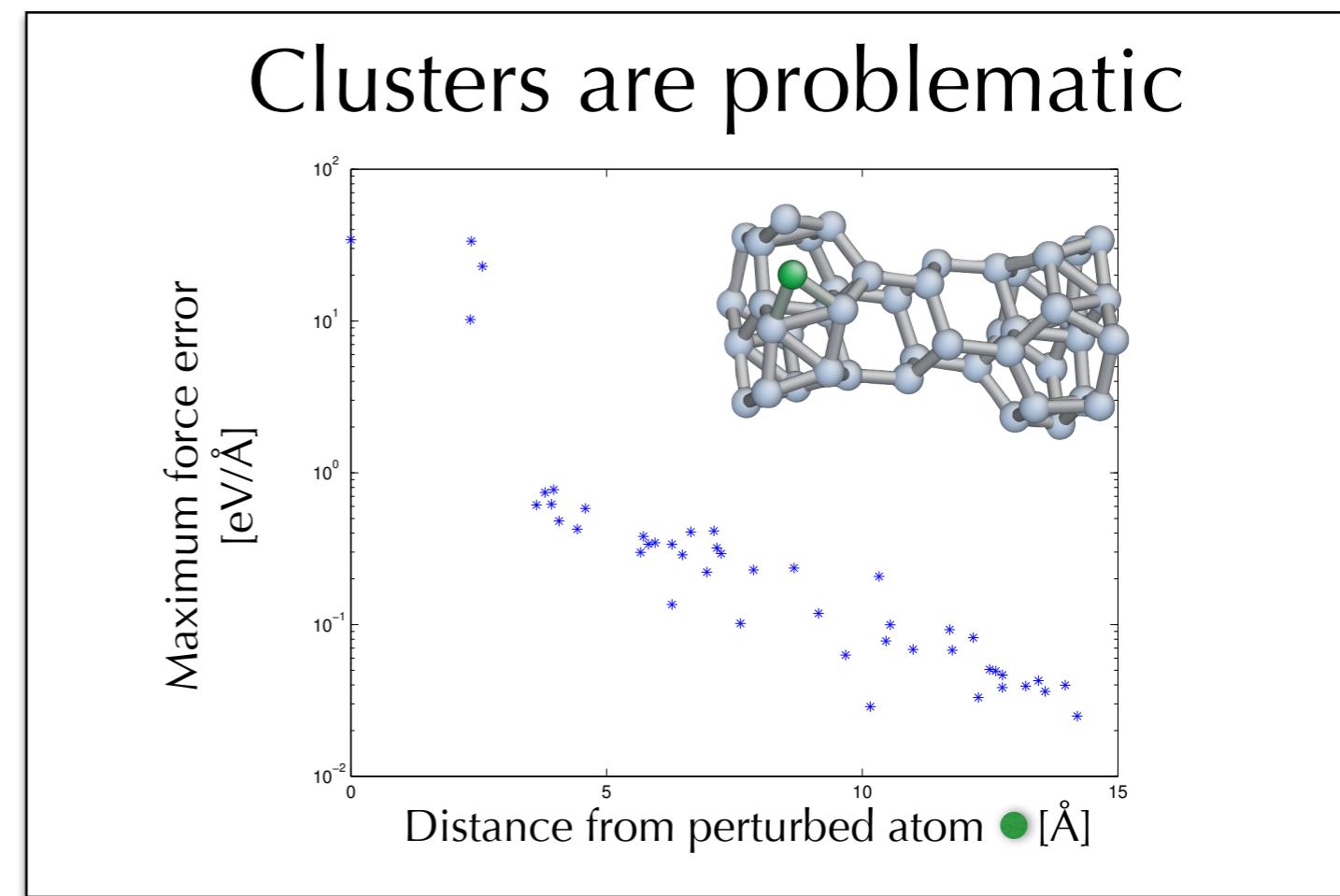


force F , standard deviation \approx error in local potential

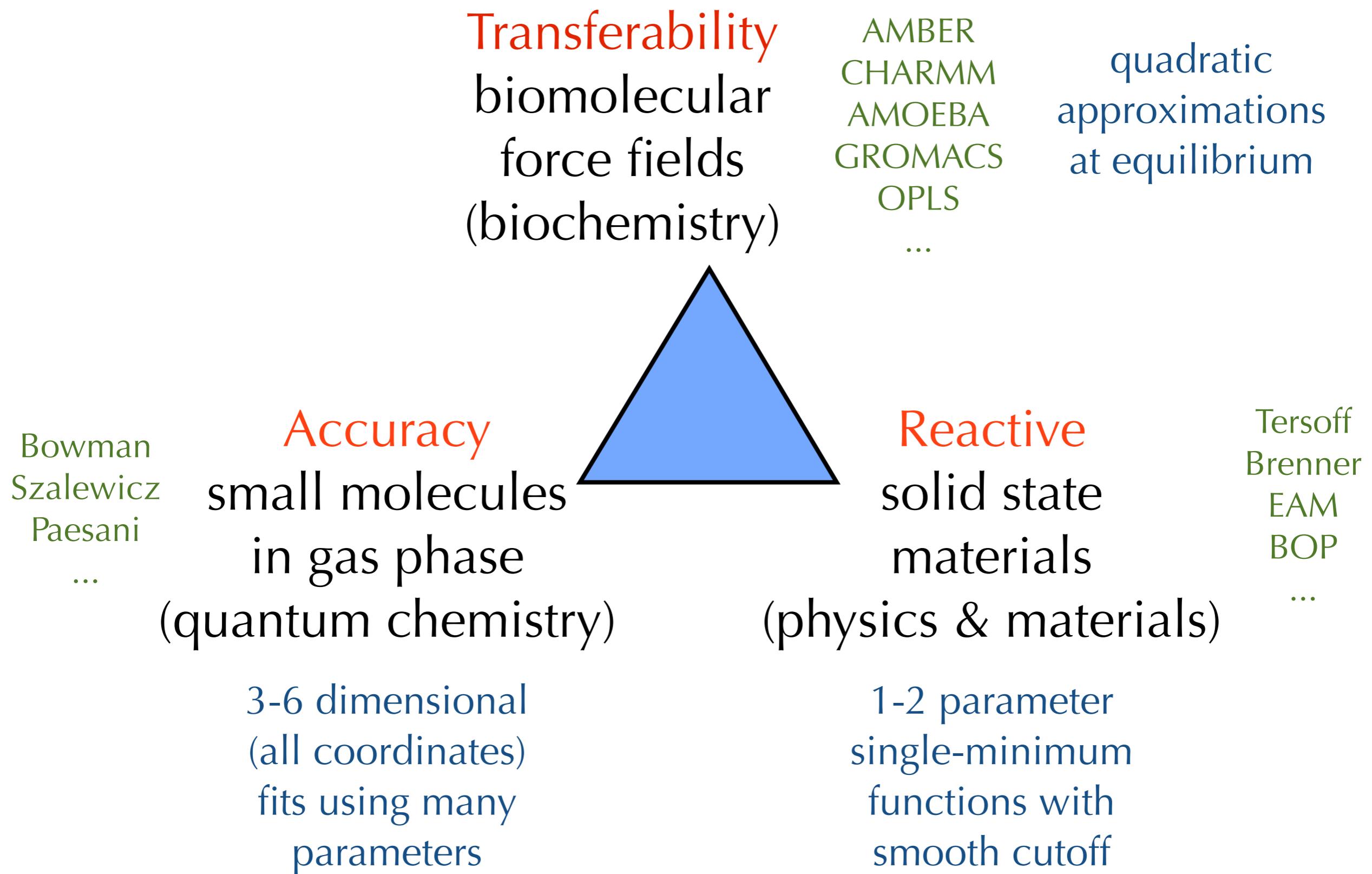
Force errors around
Si self interstitial



Clusters are problematic



Types of force fields



What makes a good potential

Functional form

Ingredients

- Representation of atomic neighbourhood
- Interpolation of functions
- Database of configurations

Desirable properties

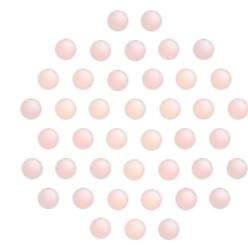
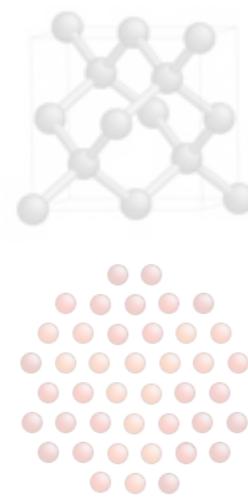
smoothness,
faithfulness,
continuity

flexible but smooth
functional form, few
sensible parameters

predictive power
non-domain specific

Strong interactions

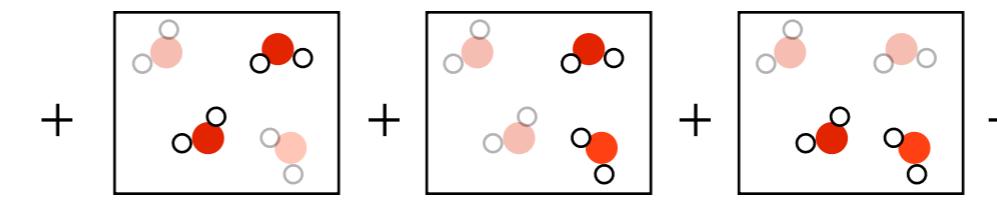
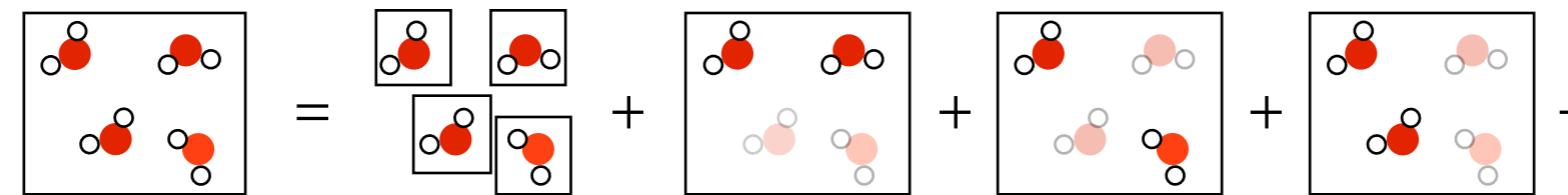
Weak interactions



$$E = \sum_i \varepsilon(q_1^{(i)}, q_2, \dots, q_M^{(i)})$$

+ analytic long range terms

Many-body
(cluster) expansion



+ higher multi-body terms... + analytic long range terms

Need a good kernel between atomic neighbourhoods!

Many body expansion

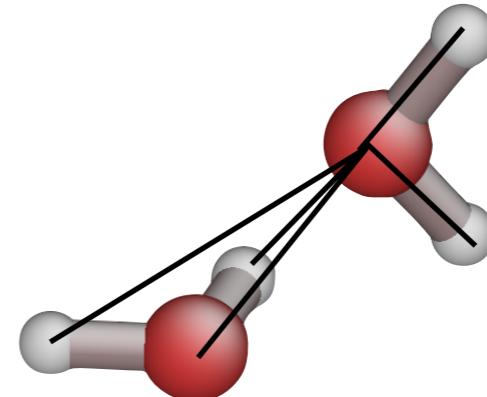
$$\begin{aligned} E^{\text{total}} &= \sum_i E^1 + \sum_{ij} E^2 + \dots \\ &= \sum_i E(i) + \sum_{ij} E(ij) - E(i) - E(j) + \dots \end{aligned}$$

Describe beyond-n-body terms with a baseline model

$$\begin{aligned} E^{\text{total}} &= \sum_i E^1(i) + \sum_{ij} E^2(ij) + \left(E_{\text{base}}^N - \sum_i E_{\text{base}}^1(i) - \sum_{ij} E_{\text{base}}^2(ij) \right) \\ &= \sum_i E^1(i) - E_{\text{base}}^1(i) + \sum_{ij} E^2(ij) - E_{\text{base}}^2(ij) + E_{\text{base}}^N \end{aligned}$$

Baseline $\left\{ \begin{array}{l} 0 \\ \text{Polarisable force field} \\ \text{Density Functional Theory} \end{array} \right.$

GAP model of the H₂O dimer



- Standard representation:
6 atoms \Rightarrow 15 interatomic distances : $x = \{r_{ij}\}$
- Symmetrise Gaussian kernel function:

$$\tilde{K}(x, x') = \sum_{p \in S} K(p(x), x')$$

S : symmetry group of molecules

$|S|=1$ monomer
 $|S|=8$ dimer
 $|S|=48$ trimer

- Use analytic forces if available:

$$K'(x, x') = \frac{\partial}{\partial x} K(x, x')$$

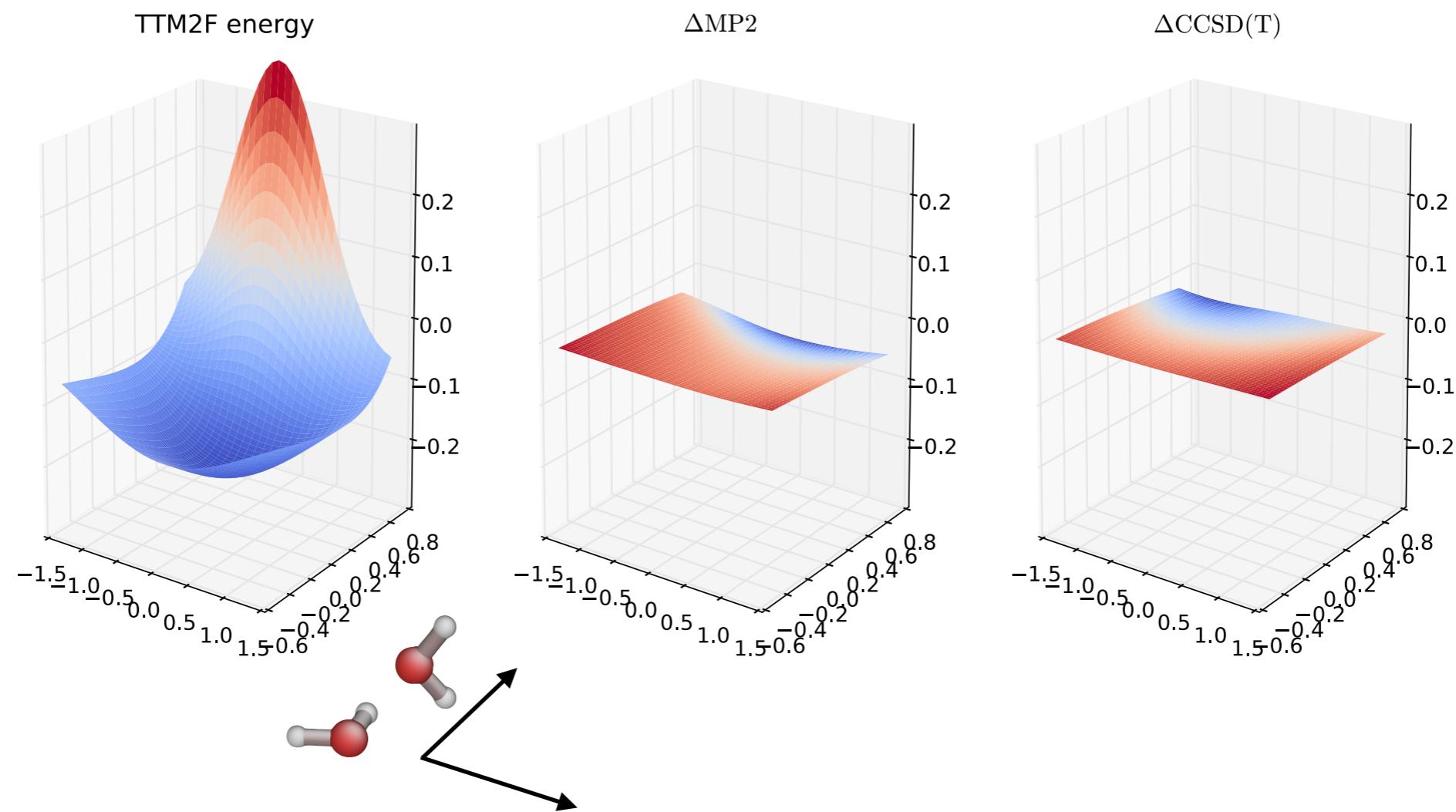
Hierarchy of fits

Target “Gold standard” potential energy surface

Coupled Cluster with Single, Double, (perturbative) Triple excitations: CCSD(T)
Basis set: augmented triple-zeta with f12 correlations : AVTZ-f12

Too expensive to compute directly, use convergence tricks: (still \sim hour/dimer)

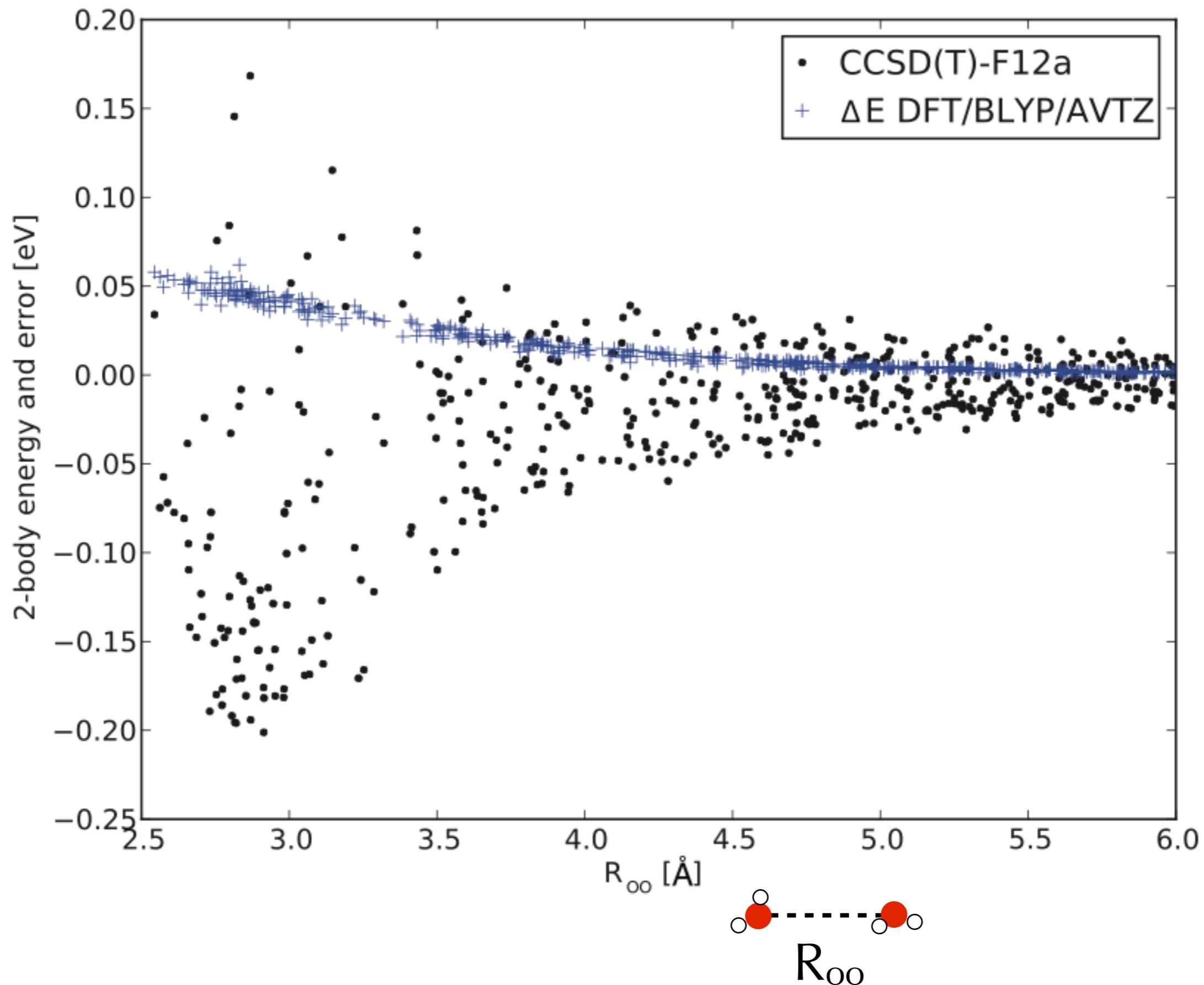
Hartree-Fock@AV5Z + Δ MP2@AVQZ + Δ MP2@AVTZ-f12 + Δ CCSD(T)@AVTZ



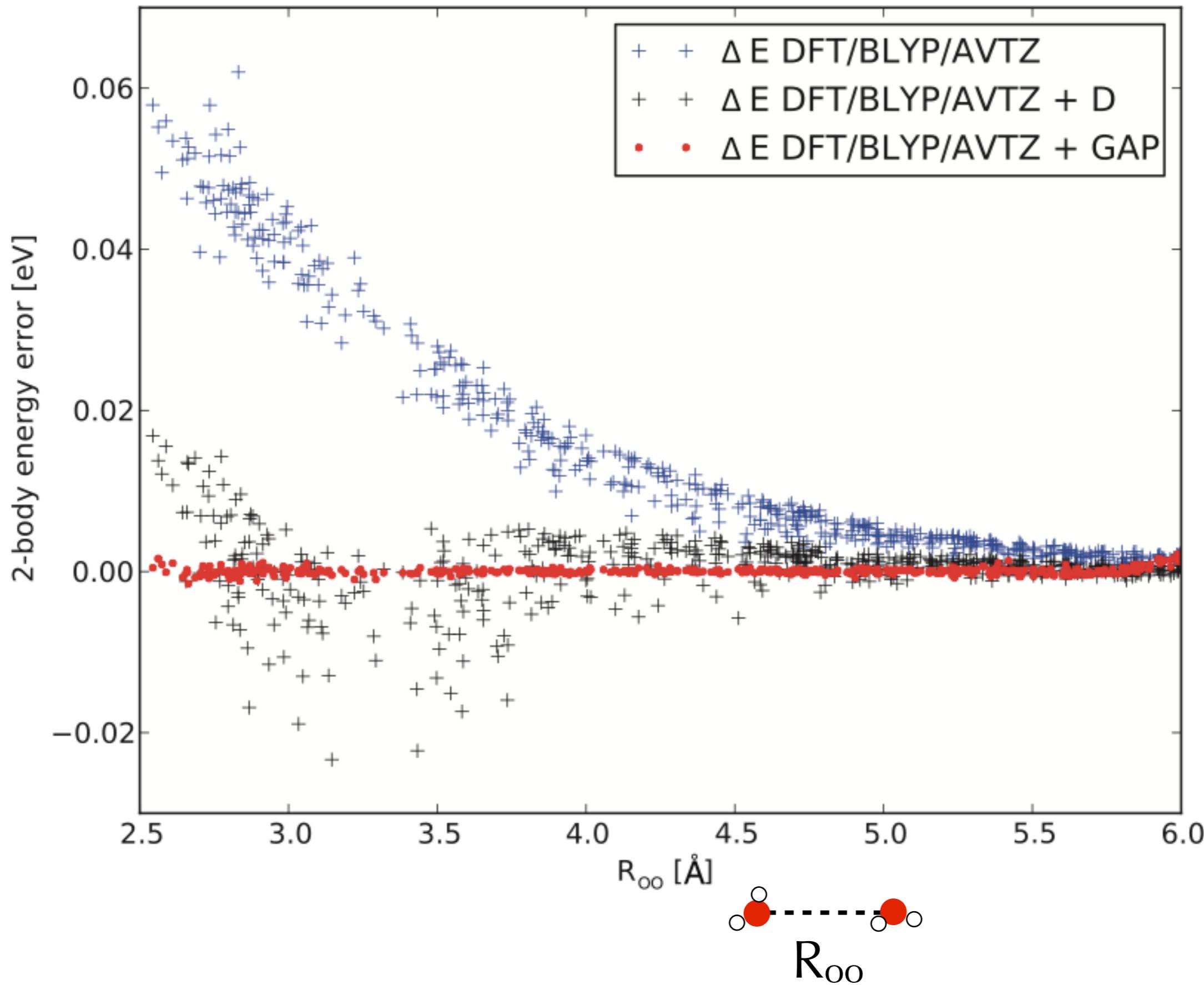
Database protocol

- >2-body terms described by baseline
- All configurations generated by MD
- GAP1: DFT → MP2 AVQZ (energies and forces)
 - 6000 configs $r_{OO} < 4.5 \text{ \AA}$, AMOEBA bulk phase MD 300K
 - 1000 configs $r_{OO} < 6.0 \text{ \AA}$, AMOEBA bulk phase MD 300K
 - 2000 configs confined dimers, DFT MD, 4000K
- GAP2: MP2 → CCSD(T) AVTZ/f12 (only energies)
 - 800 configs $r_{OO} < 4.5 \text{ \AA}$, AMOEBA bulk phase MD 300K
 - 500 configs $r_{OO} < 6.0 \text{ \AA}$, AMOEBA bulk phase MD 300K

H₂O dimer energy



H₂O dimer corrections

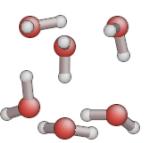
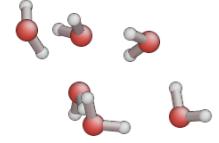
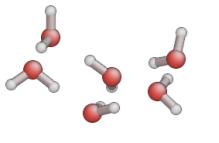
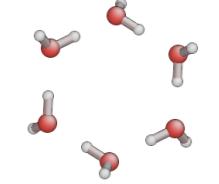


How do we validate such a model?

- Two tasks: *verification* & *validation*
(ISO 9000 Quality Management System)
- Verification: product meets specification
 - Potentials for monomer, dimer, trimer etc. meet accuracy goal, database covers necessary range of configurations
- Validation: product satisfies requirements
 - Many-body-expanded and -corrected model performs acceptably for systems of scientific interest: large water clusters, liquid properties, ice properties
 - Need widely accepted benchmarks on large systems. Not easy!

Water hexamers - relative energies

(no zero-point correction)

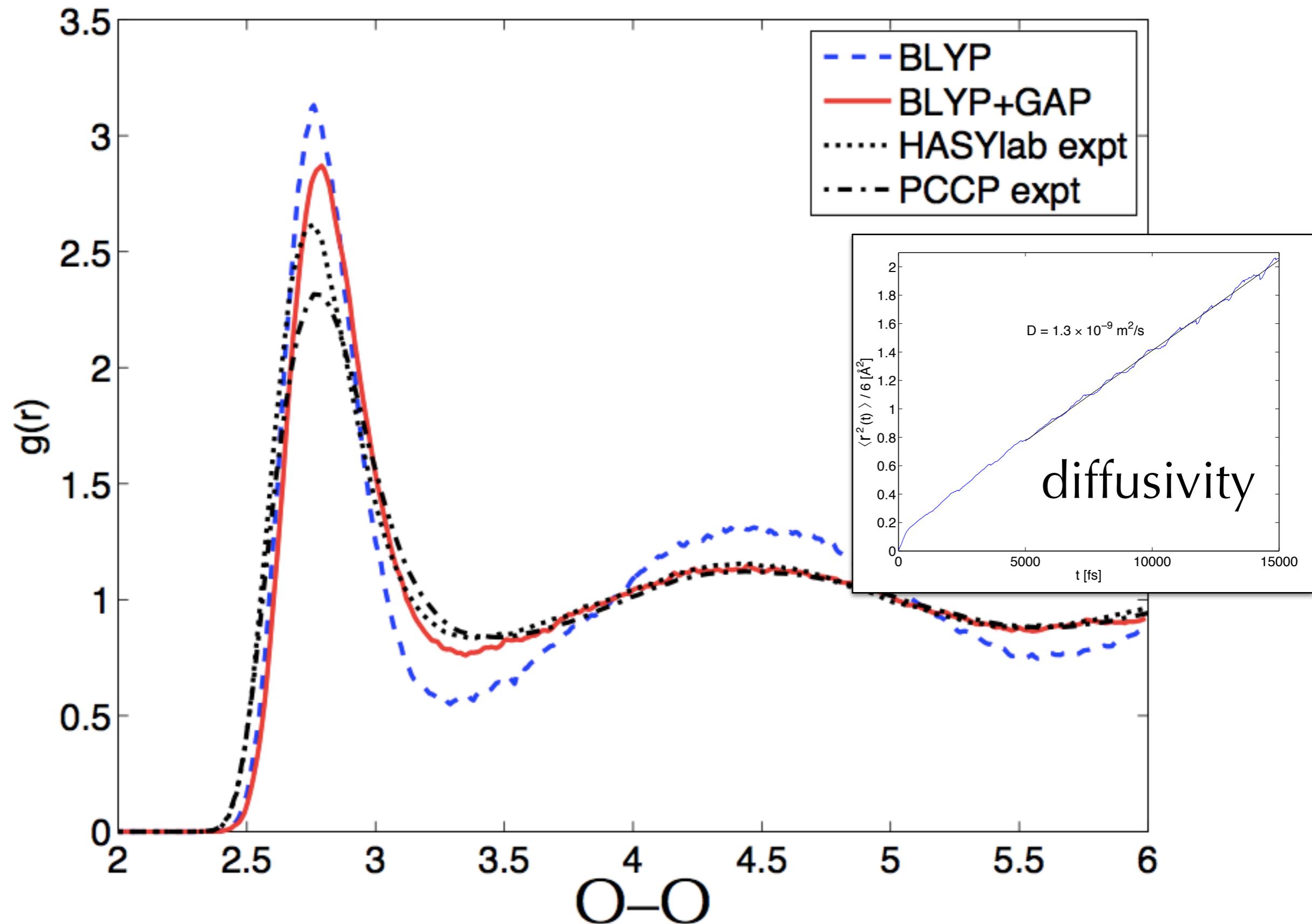
CCSD(T)	DFT/BLYP	DMC (Diffusion Monte Carlo)	WHBB (Symmetrized polynomial fit)	DFT/BLYP +GAP
	0.00	0.00	0.00	0.00
	0.24	-0.59	0.33	0.15
	0.70	-2.18	0.56	0.52
	1.69	-2.44	1.53	1.77

[kcal / mol]

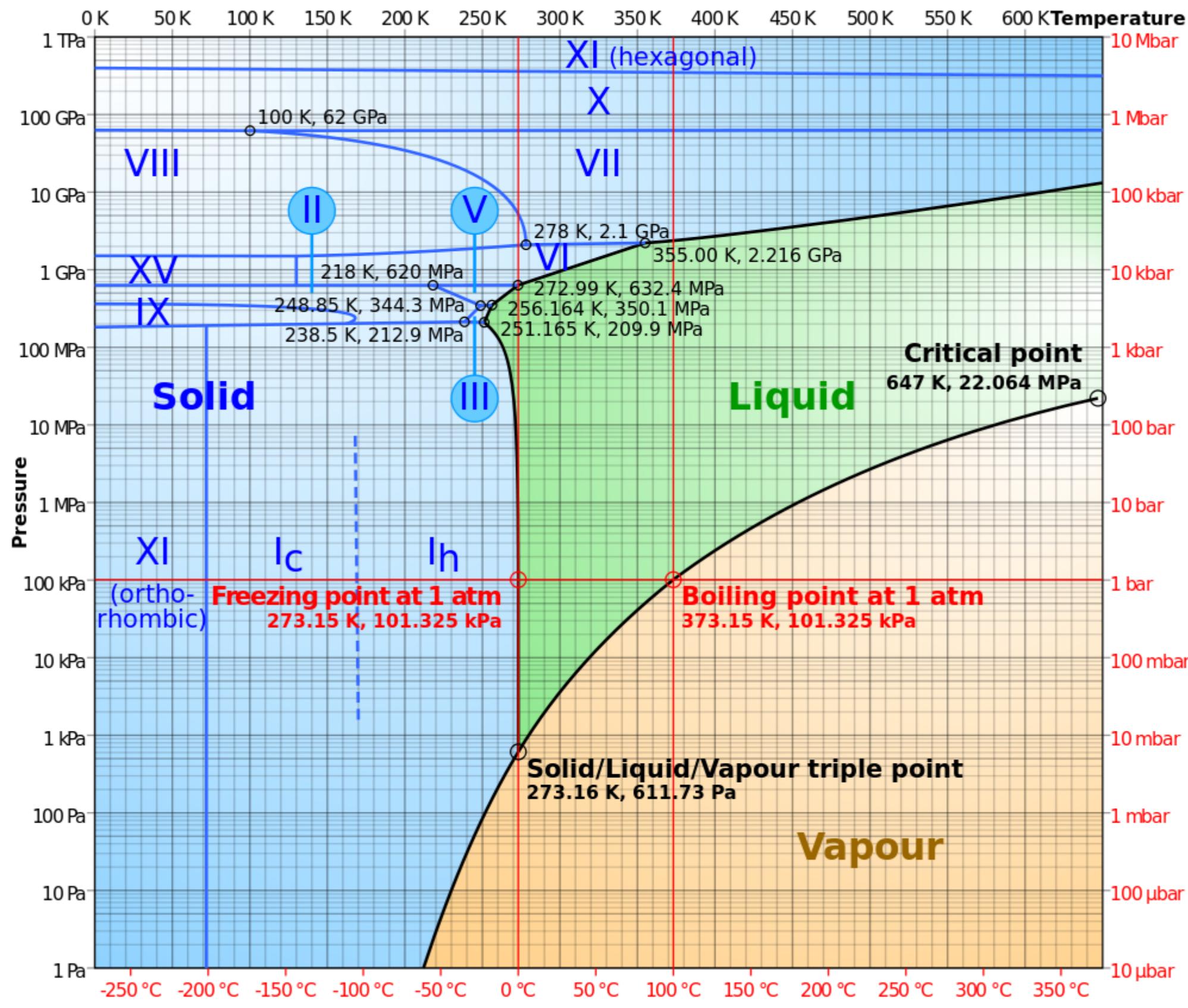
(B3LYP optimised
geometries from
Tschumper)

GAP error < 0.03 kcal/mol / H₂O

Liquid O-O Radial Distribution Function



Water phase diagram



Ice crystal polymorphs

Relative energies

[meV / H ₂ O]	Expt	DMC±5 (Diffusion Monte Carlo)	DFT/PBE0 +vdW ^{TS}	DFT/BLYP +GAP±3
Ih	0	0	0	0
II	I	-4	6	-5
IX	4		3	-3
VIII	33	30	76	30

error < 0.15 kcal/mol / H₂O

Ice crystal polymorphs

Relative volumes

[Å ³ / H ₂ O]	Expt	DMC (Diffusion Monte Carlo)	DFT/PBE0 +vdW ^{TS}	DFT/BLYP +GAP
Ih	0	0	0	0
II	-7.1	-7.0	-6.2	-6.7
IX	-6.4		-6.0	-6.5
VIII	-12.0	-12.2	10.2	-11.9

error < 5%

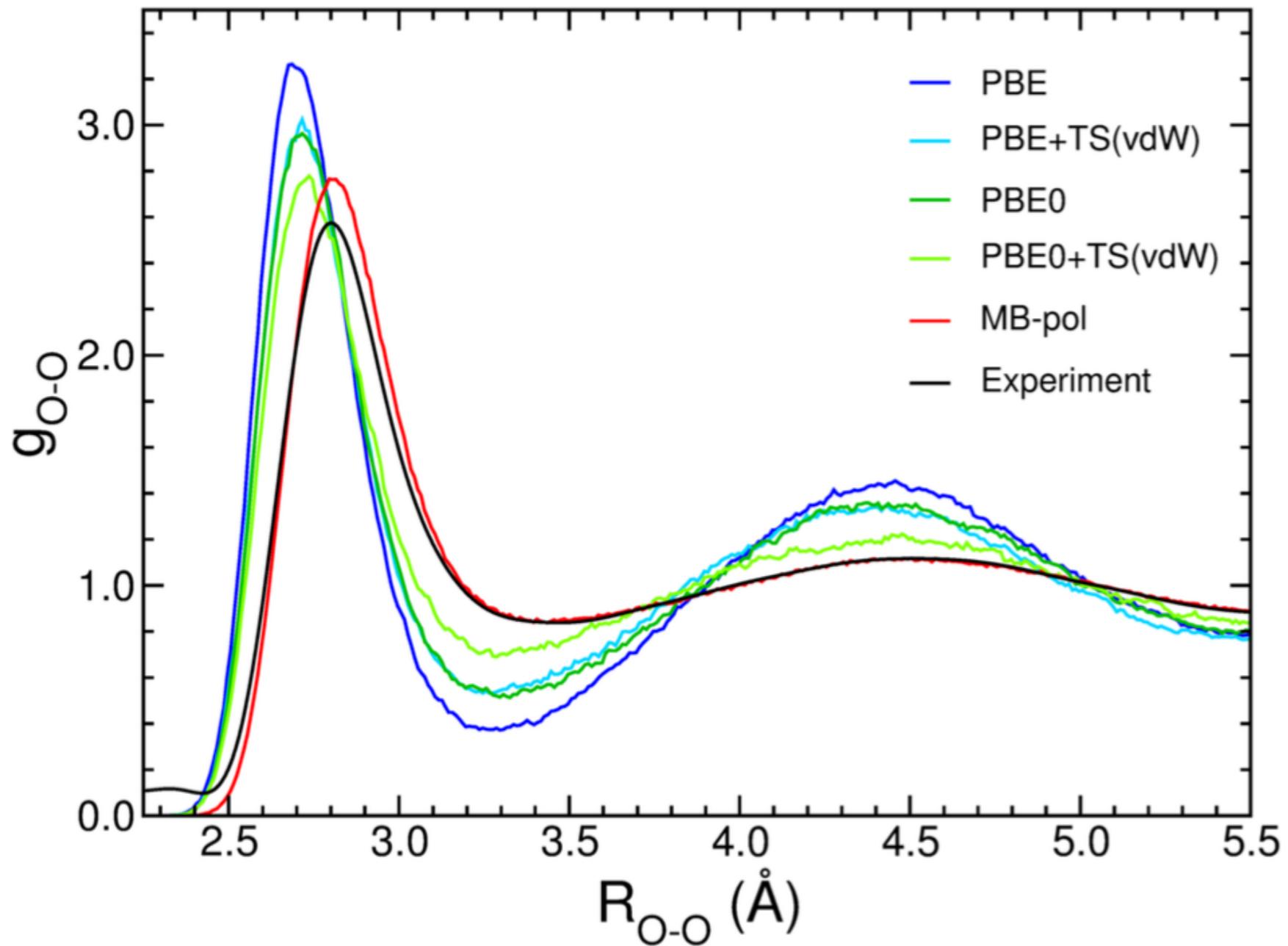
Ice crystal polymorphs

Absolute binding energies

[meV / H ₂ O]	Expt	DMC \pm 5 (Diffusion Monte Carlo)	DFT/PBE0 +vdW ^{TS}	DFT/BLYP +GAP \pm 3
Ih	610	605	672	667
II	609	609	666	672
IX				
VIII	577	575	596	637

BLYP+GAP overbinds *uniformly* by ~ 60 meV
 \Rightarrow 3-body terms!

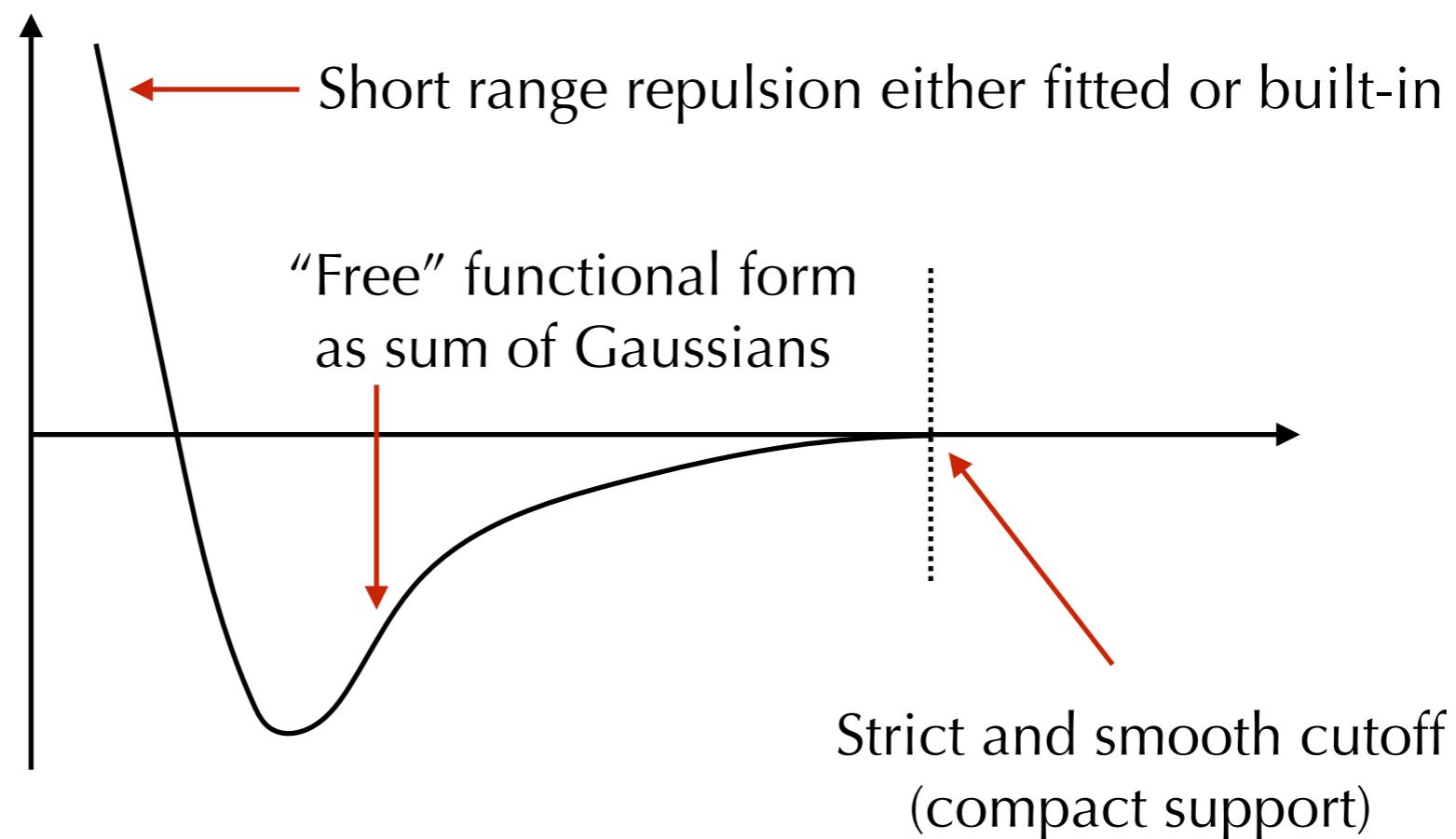
Current best water model (but not for ice!)



Francesco Paesani (2014), 1+2+3 body corrections to TTM4F baseline
(symmetrized polynomials as basis, ~50,000 data points)

Can we do the same for crystalline materials?

- Many-body expansion could still make sense
 - Evidence: 2-body and 2+3-body potentials have some usefulness
- What representation to use?
- Example: fitting a 2-body (= 2-atom) potential



How to fit to the sum of terms

Usual kernel fit, now for atomic energy function

$$\varepsilon(x) = \sum_i \alpha_i k(x, x_i)$$

Total energy for 2 atoms:

$$E = \varepsilon(x_1) + \varepsilon(x_2)$$

$$= \sum_i \alpha_i k(x_1, x_i) + \sum_i \alpha_i k(x_2, x_i)$$

$$= \sum_i \alpha_i [k(x_1, x_i) + k(x_2, x_i)]$$

$$\tilde{k}(X, x_i)$$

$$X = \{x_1, x_2\}$$

N configurations in database

$$E^n = \sum_i \alpha_i \tilde{k}(X^n, x_i)$$

How to fit a function of atom-atom distance?

Database configurations contain many “dimers”

$$E = \sum_{j < j'} \varepsilon_2(r_{jj'}) \quad \varepsilon_2(d) = \sum_i \alpha_i k(d, d_i) \quad d = |r|$$

$$E = \sum_i \alpha_i \sum_{j < j'} k(|r_{jj'}|, d_i)$$

Distance
between
atom pairs

$$E^n = \sum_i \alpha_i \tilde{k}(R^n, d_i) \quad R = \{|r|\} \quad \begin{matrix} \text{set of all} \\ \text{distances} \\ \text{between} \\ \text{atom pairs} \end{matrix}$$

Test: fit to a known simple potential

$$E = \sum_{p \in \text{ pairs}} \varepsilon_p^{(2)} + \sum_{t \in \text{ triplets}} \varepsilon_t^{(3)}$$

Target: Stillinger-Weber (SW) model of silicon with 2-body bond and 3-body angle terms

Database: Several 64-atom configurations from MD at finite T

