A Primer on Data-Driven Many-Body Models

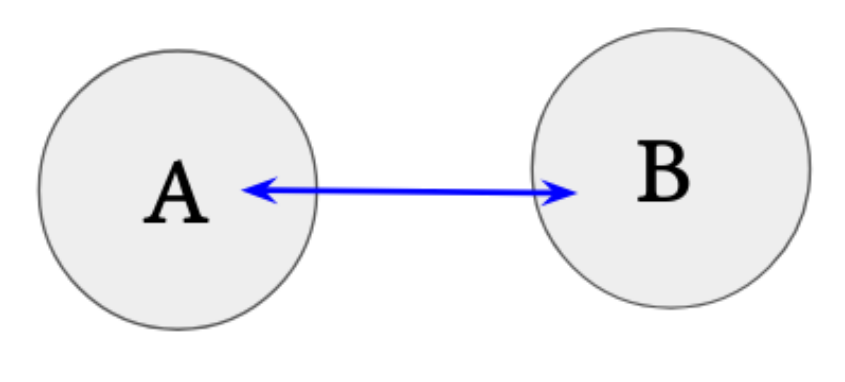
Paesani Research Group

July 14, 2025

Department of Chemistry and Biochemistry,
UC San Diego

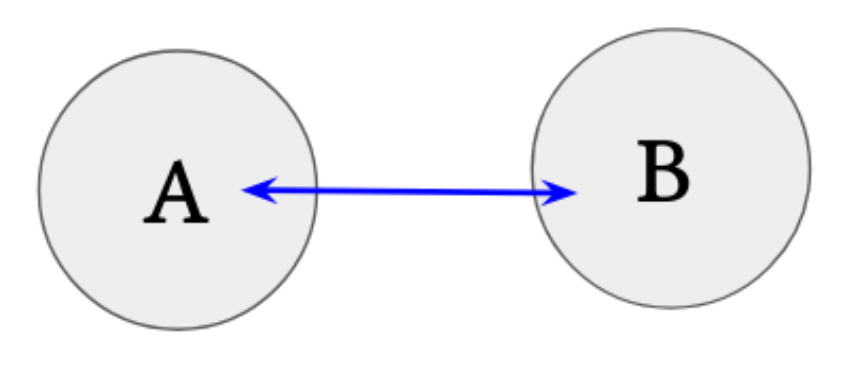
Introduction: A and B

Consider a two-body system comprised by **A** and **B**. What is their interaction energy?



Introduction: A and B

Consider a two-body system comprised by **A** and **B**. What is their interaction energy?



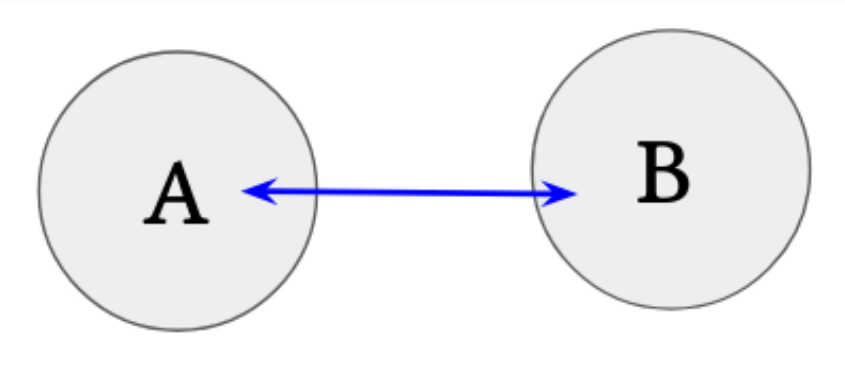
The energy of interaction between **A** and **B** is given by:

$$\Delta E = E_{\mathrm{AB}} - E_{\mathrm{A}} - E_{\mathrm{B}} \tag{1}$$

Introduction: A and B

What is stored in ΔE?

The theory of molecular interactions aims to DECOMPOSE and UNDERSTAND the nature of the interaction.



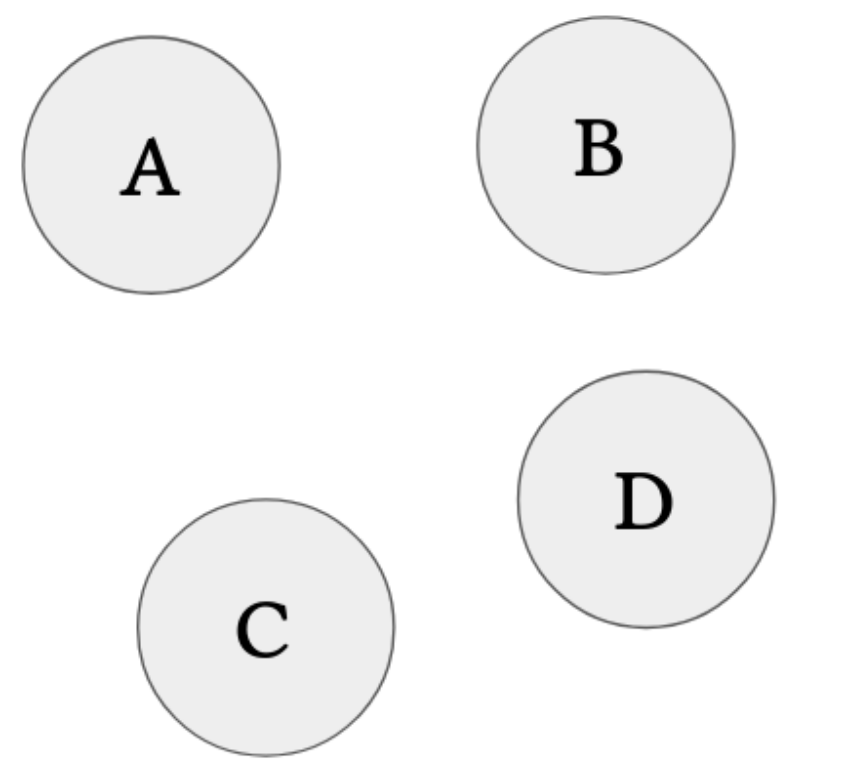
The energy of interaction between **A** and **B** is given by:

$$\Delta E = E_{\mathrm{AB}} - E_{\mathrm{A}} - E_{\mathrm{B}} \tag{2}$$

- How are the properties of **A** affected by the *presence* of **B**?
- Can understanding *single molecule* properties help us predict or rationalize *collective* behavior?

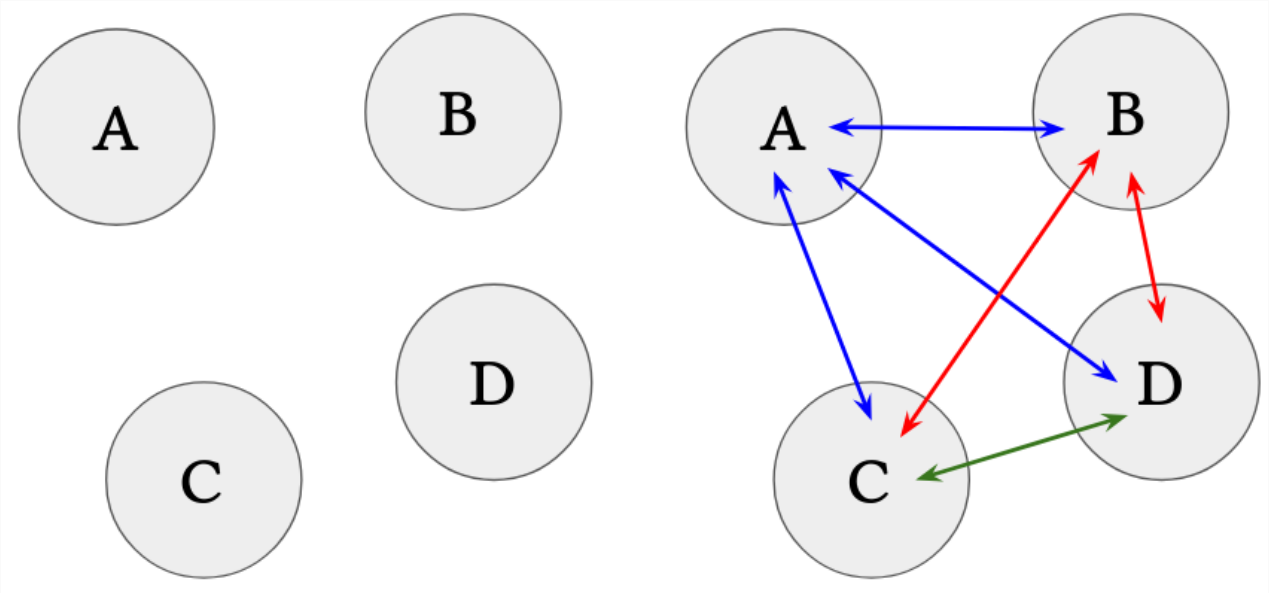
Introduction: A and B

What is the (total) energy of system **ABCD**?



Introduction: A and B

What is the (total) energy of system **ABCD**?

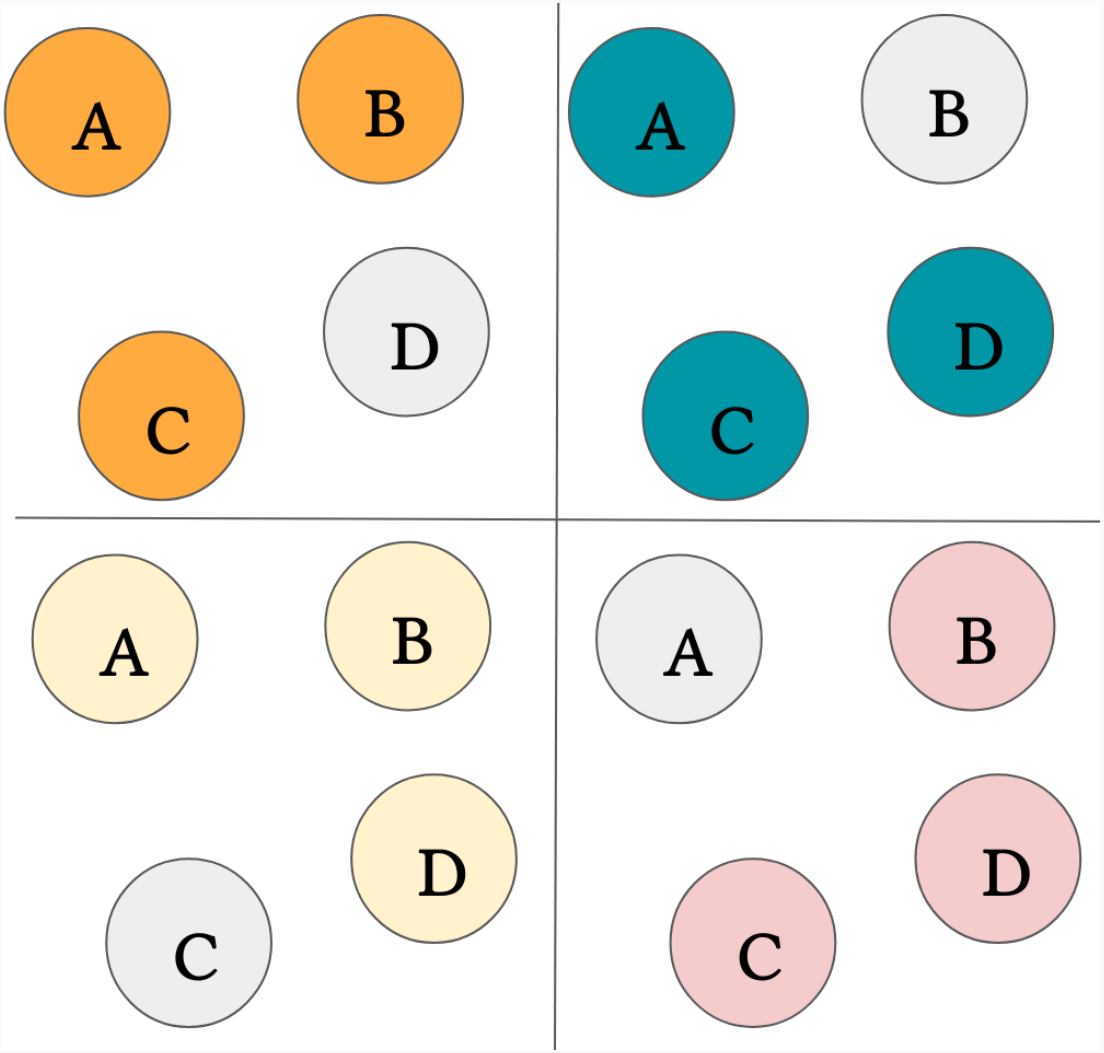


$$E_{\mathrm{ABCD}} \neq E_{\mathrm{A}} + E_{\mathrm{B}} + E_{\mathrm{C}} + E_{\mathrm{D}} + \Delta E^{\mathrm{AB}} + \Delta E^{\mathrm{AC}} + \Delta E^{\mathrm{AD}} + \Delta E^{\mathrm{BC}} + \Delta E^{\mathrm{BD}} + \Delta E^{\mathrm{CD}} \quad (3)$$

This is only a *first approximation*: $E_{\mathrm{ABCD}} \approx \sum_i \varepsilon_{1\mathrm{B}} + \sum_{i>j} \varepsilon_{2\mathrm{B}}$

Introduction: A and B

To better approximate E_{ABCD}, we can *correct* for the next n-body term, and so on.



Introductions: More than A and B

Contribution	Additive?	Sign	Comment
Long-range ($U \sim R^{-n}$)			
Electrostatic	Yes	±	Strong orientation dependence
Induction	No	−	
Dispersion	approx.	−	Always present
Resonance	No	±	Degenerate states only
Magnetic	Yes	±	Very small
Short-range ($U \sim \mathrm{e}^{-aR}$)			
Exchange-repulsion	approx.	+	Dominates at very short range
Exchange-induction	approx.	−	
Exchange-dispersion	approx.	−	
Charge transfer	No	−	Donor-acceptor interaction

Figure 1: Contributions to the energy of interaction between molecules. Stone, Anthony. *The theory of intermolecular forces.* oUP oxford, 2013.

Many-Body Expansion of the Energy

For any N-body system composed of N atoms or molecules, the total energy of this system, E_N, can be written in terms of the many-body expansion (MBE):

$$
\begin{aligned}
E_N(\mathbf{r}_1, .., \mathbf{r}_N) = & \sum_{i=1}^{N} \varepsilon_{1\mathrm{B}}(\mathbf{r}_i) \\
& + \sum_{i=1}^{N} \sum_{j>i}^{N} \varepsilon_{2\mathrm{B}}(\mathbf{r}_i, \mathbf{r}_j) \\
& + \sum_{i=1}^{N} \sum_{j>i}^{N} \sum_{k>j>i}^{N} \varepsilon_{3\mathrm{B}}(\mathbf{r}_i, \mathbf{r}_j, \mathbf{r}_k) \\
& + ... + \varepsilon_{\mathrm{NB}}(\mathbf{r}_1, .., \mathbf{r}_{\mathrm{N}})
\end{aligned}
\tag{5}
$$

where ε_{1B} represents the energy of an isolated monomer, ε_{2B} a dimer... etc.

Many-Body Expansion of the Energy

For any N-body system composed of N atoms or molecules, the total energy of this system, E_N, can be written in terms of the many-body expansion (MBE):

$$
\begin{aligned}
E_N(\mathbf{r}_1, .., \mathbf{r}_N) = & \sum_{i=1}^{N} \varepsilon_{1\mathrm{B}}(\mathbf{r}_i) \\
& + \sum_{i=1}^{N} \sum_{j>i}^{N} \varepsilon_{2\mathrm{B}}(\mathbf{r}_i, \mathbf{r}_j) \\
& + \sum_{i=1}^{N} \sum_{j>i}^{N} \sum_{k>j>i}^{N} \varepsilon_{3\mathrm{B}}(\mathbf{r}_i, \mathbf{r}_j, \mathbf{r}_k) \\
& + \ldots + \varepsilon_{\mathrm{NB}}(\mathbf{r}_1, .., \mathbf{r}_\mathrm{N})
\end{aligned}
\tag{6}
$$

The MBE is a *finite series* that, when carried to N-th order, yields the exact energy.

Many-Body Expansion of the Energy

The MBE is a *finite series* that, when carried to N-th order, yields the exact energy (fine print!)

- Offers a systematic path to numerically exact models, provided that each n-body term is treated at this level.
- Computationally practical when the MBE converges rapidly, i.e. at small n.

Many-Body Expansion (MBE)

The n-body energies, $\varepsilon_{\mathrm{nB}}$, are defined recursively for $1 \leq n \leq N$ by the expression

$$\begin{aligned}\varepsilon_{\mathrm{nB}} =& E_{\mathrm{n}}(1,...,n) - \sum_{i=1}^{N} \varepsilon_{1\mathrm{B}}(\mathbf{r}_i) - \sum_{i=1}^{N}\sum_{i<j}^{N} \varepsilon_{2\mathrm{B}}(\mathbf{r}_i,\mathbf{r}_j) \\ & - \sum_{i=1}^{N}\sum_{i<j}^{N}\sum_{i<j<k} \varepsilon_{3\mathrm{B}}(\mathbf{r}_i,\mathbf{r}_j,\mathbf{r}_k) - \cdots \\ \cdots & - \sum_{i<j<k<\ldots}^{N} \varepsilon_{(\mathrm{n}-1)\mathrm{B}}(\mathbf{r}_i,\mathbf{r}_j,\mathbf{r}_k,\ldots).\end{aligned} \tag{7}$$

How many *single point calculations do you need to run?*

Many-Body Expansion (MBE)

How many *single point calculations do you need to run?*

Each n-body term subtracts lower-order contributions to avoid double-counting.

Total number of terms: $\sum_{n=1}^{N} \binom{N}{n} = 2N - 1$

Example 1: Water Hexamer ($N = 6$)

Many-Body Expansion (MBE)

How many *single point calculations do you need to run?*

Each n-body term subtracts lower-order contributions to avoid double-counting.

Total number of terms: $\sum_{n=1}^{N} \binom{N}{n} = 2N - 1$

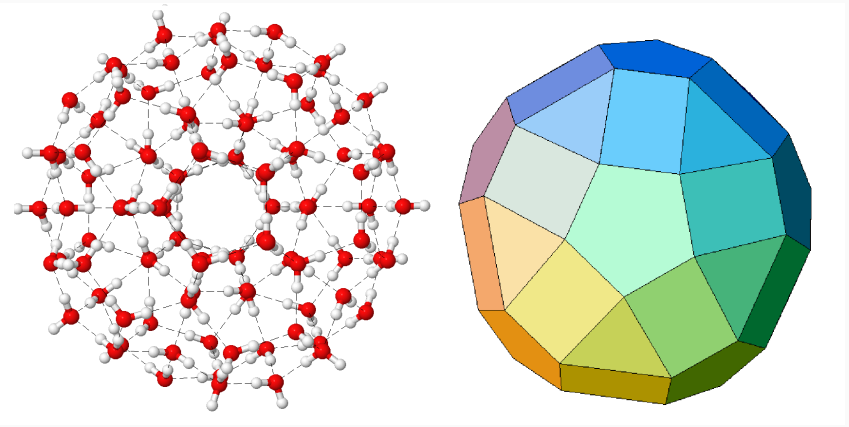
Example 1: Water Hexamer ($N = 6$)

- 6 one-body terms: $\binom{6}{1}$
- 15 two-body terms: $\binom{6}{2}$
- 20 three-body terms: $\binom{6}{3}$
- 15 four-body terms: $\binom{6}{4}$
- 6 five-body terms: $\binom{6}{5}$
- 1 six-body term: $\binom{6}{6}$

Total: 63 total energy calculations.

Many-Body Expansion (MBE)

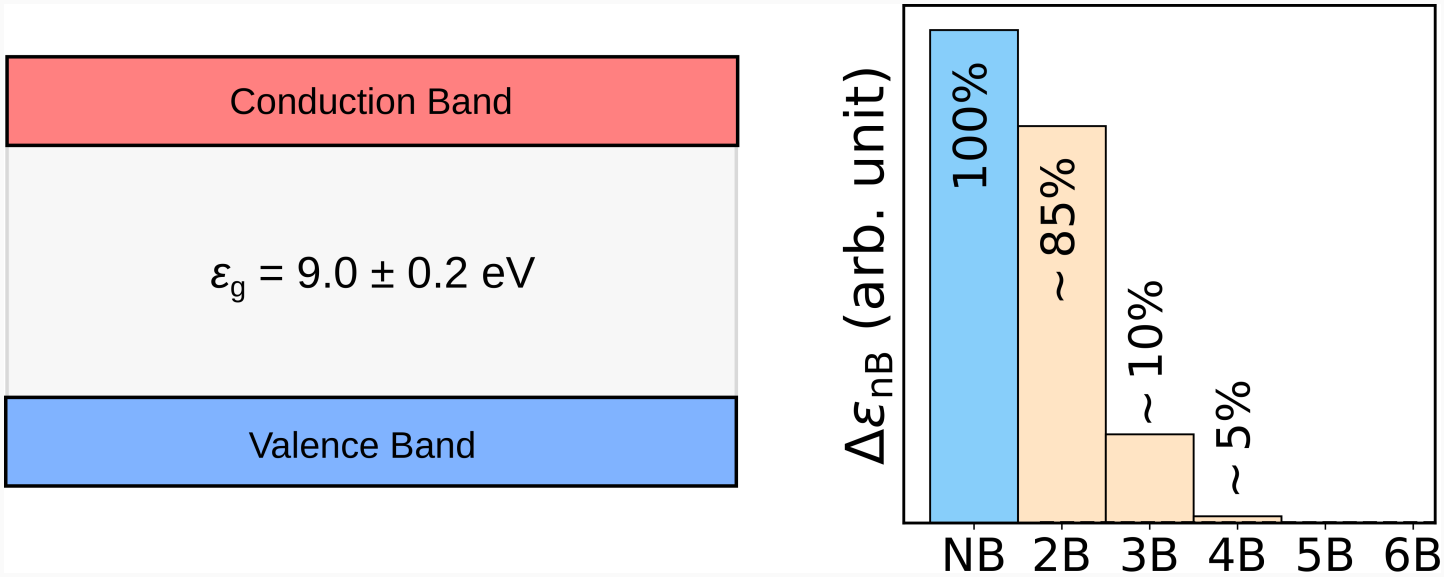
Example 2: $N = 100$



Total: $\sim$ 1.4-million for 4-body approximation

Many-Body Expansion (MBE)

MBE[1] converges quickly for systems with large bandgaps and/or localized electron densities, such as H_2O ($\varepsilon_g \sim 9$ eV)[2].



[1] *J. Chem. Phys. 53, 4544 (1970)*

Theory: MB-nrg potential energy functions

MB-nrg: Functional form

The **MB-nrg** theoretical/computational framework generalizes the MB formalism introduced with the MB-pol data-driven many-body potential for water.

MB-nrg approximates the MBE defined in Eq. 6 as:

$$E_N = V^{1\mathrm{B}} + V^{2\mathrm{B}} + V^{3\mathrm{B}} + \ldots + V^{n\mathrm{B}} + V_{\mathrm{elec}} \tag{8}$$

where $n \leq N$ and N is the total number of 1-mers in the system.

MB-nrg: Functional Form

Each $V^{n\mathrm{B}}$ term of an MB-nrg PEF includes an n-body machine-learned term ($V^{n\mathrm{B}}_{\mathrm{ML}}$) for each n-mer.

Each $V^{n\mathrm{B}}_{\mathrm{ML}}$ is expressed as a product of a switching function and a PIP (i.e., $V^{n\mathrm{B}}_{\mathrm{ML}} = s^{n\mathrm{B}} V^{n\mathrm{B}}_{\mathrm{PIP}}$).

The switching function ($s^{n\mathrm{B}}$) ensures contributions from associated $V^{n\mathrm{B}}_{\mathrm{ML}}$ go to zero as any subset of the 1-mers in an n-mer is separated from the rest.

MB-nrg: Functional Form

Following MB-pol: a given n-body PIP takes the following form:

$$V_{\mathrm{PIP}}^{n\mathrm{B}}(\mathrm{M}_1,\mathrm{M}_2,\ldots,\mathrm{M}_n|\nu(\mathrm{M}_1,\mathrm{M}_2,\ldots,\mathrm{M}_n)) = \sum_{l=1}^{L} c_l \cdot \eta_l(\xi_1,\xi_2,\ldots,\xi_\lambda) \tag{9}$$

$\mathrm{M}_1,\mathrm{M}_2,\ldots,\mathrm{M}_n$ are n 1-mers which compose an n-mer of type $\nu(\mathrm{M}_1,\mathrm{M}_2,\ldots,\mathrm{M}_n)$,

L : number of linear coefficients

c_l : the linear coefficients

η_l: symmetrized monomials built from the variables $\xi_{1-\lambda}$

Frame Title

The 1-body term is a machine-learned PIP representing the 1-body energy:

$$V_{\mathrm{ML}}^{1\mathrm{B}}\left(\mathrm{M}_i|\nu(\mathrm{M}_i)\right) = V_{\mathrm{PIP}}^{1\mathrm{B}}\left(\mathrm{M}_i|\nu(\mathrm{M}_i)\right) \tag{10}$$

The 2-body term of an MB-nrg PEF, $V^{2\mathrm{B}}$ in Eq. 8, is explicitly written as follows:

$$V^{2\mathrm{B}} = \sum_{\substack{i=1 \\ j>i}}^{N} V_{\mathrm{ML}}^{2\mathrm{B}}\left(\mathrm{M}_i, \mathrm{M}_j|\nu(\mathrm{M}_i, \mathrm{M}_j)\right) + V_{\mathrm{disp}}^{2\mathrm{B}} \tag{11}$$

$$V_{\mathrm{ML}}^{2\mathrm{B}}\left(\mathrm{M}_i, \mathrm{M}_j|\nu(i,j)\right) = s^{2\mathrm{B}}\left(\mathrm{M}_i, \mathrm{M}_j|\nu(\mathrm{M}_i, \mathrm{M}_j)\right) V_{\mathrm{PIP}}^{2\mathrm{B}}\left(\mathrm{M}_i, \mathrm{M}_j|\nu(\mathrm{M}_i, \mathrm{M}_j)\right) \tag{12}$$

MB-nrg: Dispersion

$V_{\mathrm{disp}}^{\mathrm{2B}}$ is the total 2-body dispersion energy calculated as a sum of pairwise additive contributions associated with each pair of atoms located on the two 1-mers in a 2-mer:

$$V_{\mathrm{disp}}^{\mathrm{2B}} = \sum_{\substack{i=1 \\ j>i}}^{N} \left[\sum_{k \in \mathrm{M}_i} \sum_{l \in \mathrm{M}_j} -f(\mathrm{b}_{kl} R_{kl}) \frac{C_{6,kl}}{R_{kl}^6} \right] \tag{13}$$

R_{kl} is the distance between atoms k and l located on 1-mers M_i and M_j.

$\mathrm{C}_{6,kl}$ is the corresponding dispersion coefficient.

$f(\mathrm{b}_{kl} R_{kl})$ is the Tang-Toennies damping function.

MB-nrg: Arbitrary ML n-body terms

All other explicit $V^{n\mathrm{B}}$ in Eq. 8 take the following form:

$$V^{n\mathrm{B}} = \sum_{\substack{i=1 \\ j>i \\ \cdots \\ l>k}}^{N} V^{n\mathrm{B}}_{\mathrm{ML}}\left(\mathrm{M}_i, \mathrm{M}_j, \ldots, \mathrm{M}_l | \nu(\mathrm{M}_i, \mathrm{M}_j, \ldots, \mathrm{M}_l)\right) \qquad (14)$$

where each $V^{n\mathrm{B}}_{\mathrm{ML}}\left(\mathrm{M}_i, \mathrm{M}_j, \ldots, \mathrm{M}_l | \nu(\mathrm{M}_i, \mathrm{M}_j, \ldots, \mathrm{M}_l)\right)$ is built as the product of a switching function $s^{n\mathrm{B}}$ and a PIP, as shown previously.

MB-nrg PEFs are systematically improvable with $V^{n\mathrm{B}}$

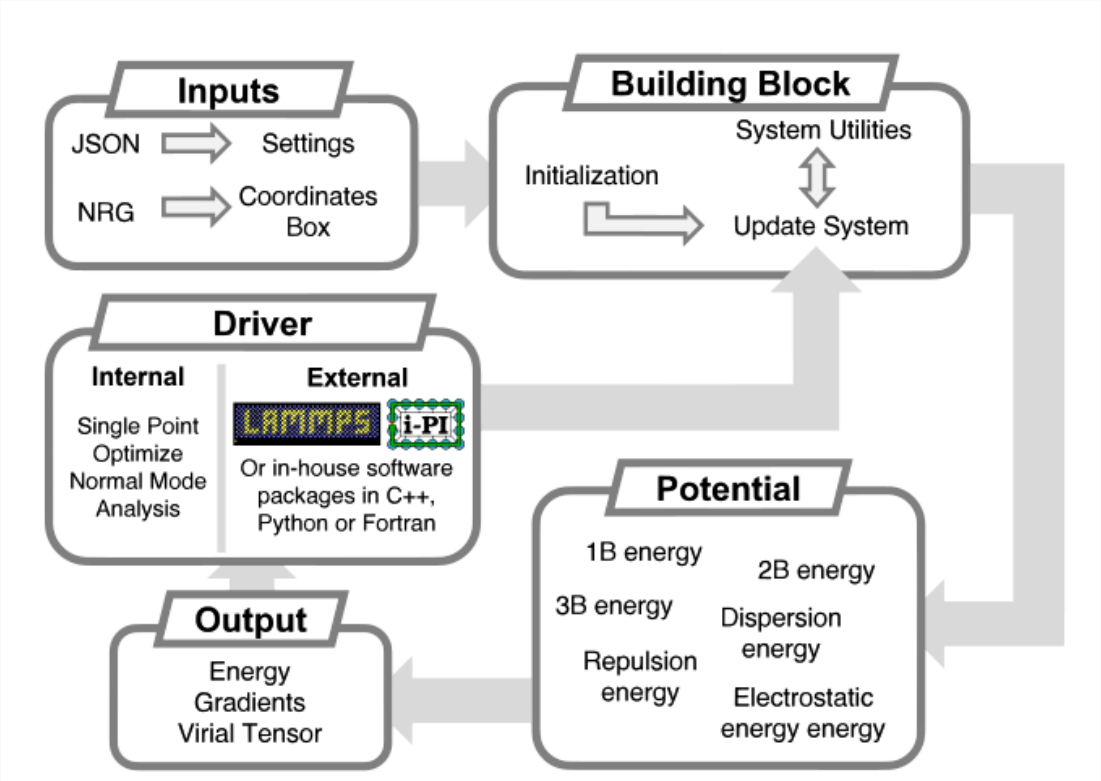
MB-nrg: Electrostatics

In MBX, V_{elec} is represented by four terms describing **charge-charge** interactions (V_{qq}), **charge-dipole** interactions ($V_{\mathrm{q}\mu}$), **dipole-dipole** interactions ($V_{\mu\mu}$), and the **polarization energy** (V_{pol}), respectively:

$$
\begin{aligned}
V_{\mathrm{qq}} &= \sum_{i}^{N} \sum_{j>i} q_i \widehat{T}_{ij} q_j \\
V_{\mathrm{q}\mu} &= \sum_{i}^{N} \sum_{j>i} \left(\mu_i^{\alpha} \widehat{T}_{ij}^{\alpha} q_j - q_i \widehat{T}_{ij}^{\alpha} \mu_j^{\alpha} \right) \\
V_{\mu\mu} &= - \sum_{i}^{N} \sum_{j>i} \mu_i^{\alpha} \widehat{T}_{ij}^{\alpha\beta} \mu_j^{\beta} \\
V_{\mathrm{pol}} &= \frac{1}{2} \sum_{i=1}^{N} \mu_i \widehat{\alpha}_i^{-1} \mu_i
\end{aligned}
\tag{15}
$$

MBX

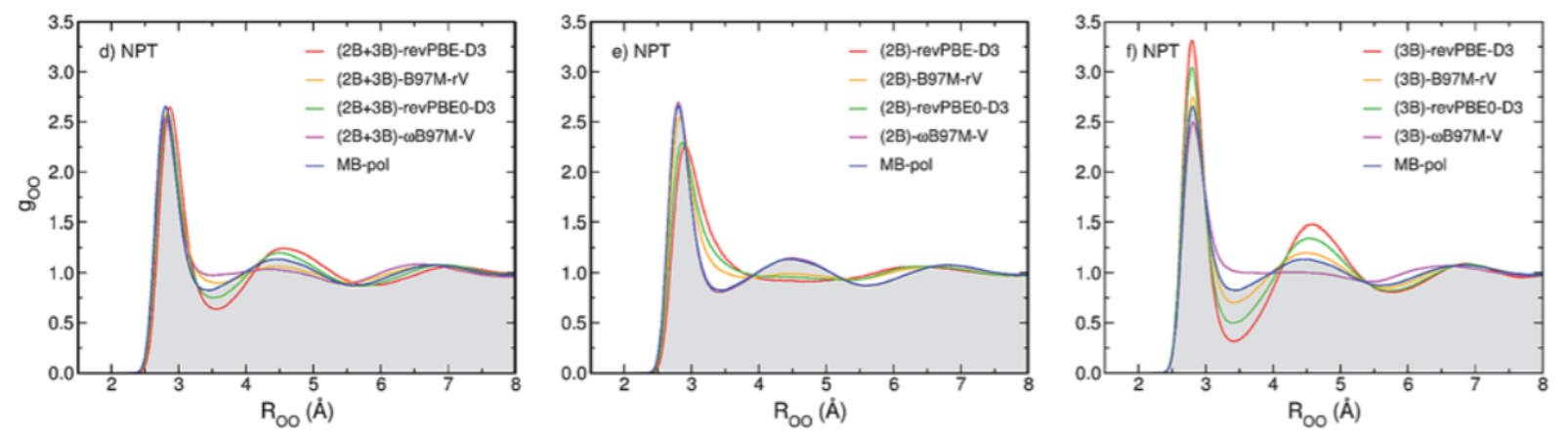
MBX is a C++ library for the calculation of n-body energies (it is an energy *calculator*, not a *force* engine).



It has over 20+ MB-nrg PEFs implemented for several systems...

and it is open source!

We can *track* the contributions of n-body terms to structural and, consequentially, thermodynamic properties.



Enables going *beyond the blackbox* in "first-principles" simulations, term-by-term design and understanding. Chem. Sci., 2019, 10, 8211

Summary

What have we discussed?

- MBE offers a path towards numerically-exact models
- MB-nrg very accurately approximates MBE, by integrating ML terms with physics-based terms
- rigorously represents short-range and long-range interactions
- systematically improvable
- open source

What is today's goal?

- Learn to *use* MB-nrg models in MBX

What is tomorrow's goal?

- Learn to *make/implement* MB-nrg models in MBX

[Cambria-Math.ttf]