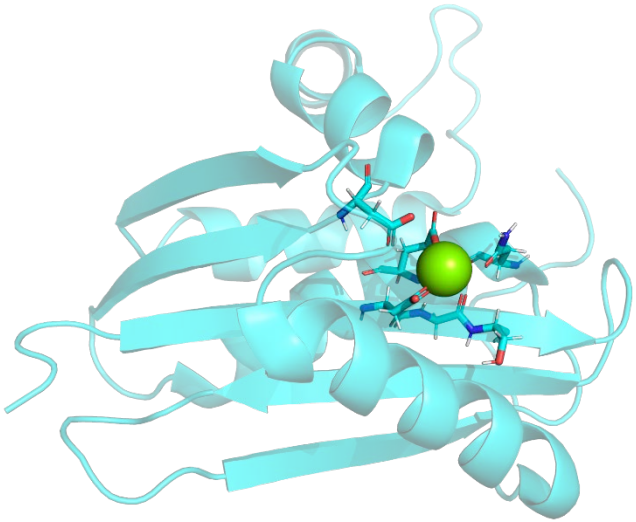
Machine learning force fields: Current status and future directions

Francis

Oct 26, 2018

Deficiencies of current potential energy models

$$E_{\text{total}} = E_{\text{bond}} + E_{\text{angle}} + E_{\text{torsion}} + \cdots$$
$$+E_{\text{vdW}} + E_{\text{elst}}$$



Less accurate for

- Ionic bond
- Covalent bond
- Reactions

Why Neural Networks?

Overfitting
Small error for training set and large error for test set

Less data

More data

More parameters

Fewer parameters

Underfitting
Large errors for both training set and test set

Spline function

$n_{\text{param}} \approx n_{\text{data}}$

Example:

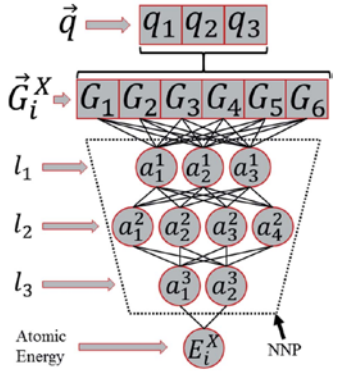
- Grid-based dihedral energy correction (CHARMM CMAP)
- Iterative Boltzmann inversion

Two many parameters for high dimensional functions

Neural networks

$n_{\text{param}}: 10^4 \sim 10^6$

$n_{\text{data}} > 10^6$



Classical force field

$n_{\text{param}} \approx 10^3$

$n_{\text{data}} \approx 10^4$

$$U_{\text{torsion}} = \sum_n K_{n\phi}[1 + \cos(n\phi \pm \delta)]$$

$$U_{\text{vdw}}(ij) = \varepsilon_{ij}\left(\frac{1.07}{\rho_{ij} + 0.07}\right)^7\left(\frac{1.12}{\rho_{ij}^7 + 0.12} - 2\right)$$

Mackerell et al. J Comput Chem 25: 1400 –1415, 2004
Reith et al, J Comput Chem. 2003 Oct;24(13):1624-36.
Smith et al. Chem. Sci., 2017, 8, 3192–3203
Ponder et al. J. Phys. Chem. B 2010, 114, 2549–2564

Neural network energy model

Molecular descriptor

Permutation invariant
Symmetry-preserving
Unique

Neural network

Architecture
Numbers of layers and nodes
Functional form

Reference data

Types of data
Sampling

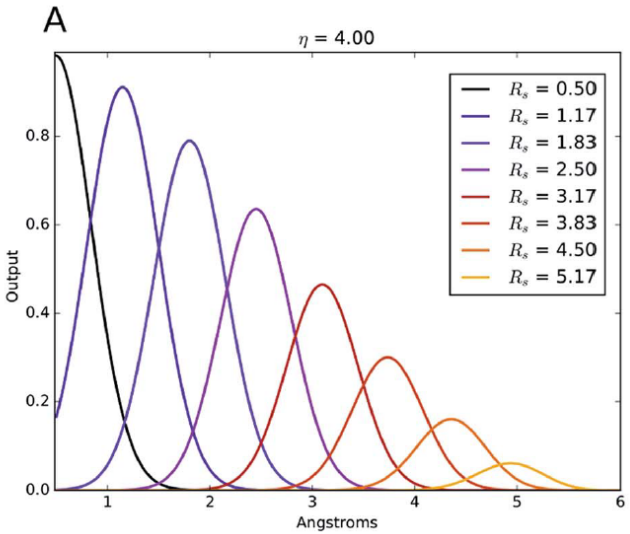
Behler-Parrinello symmetry function

- Atom environment $[G_1^X, G_2^X, \ldots]$

Radial

$$G_m^{\mathrm{R}} = \sum_{j \neq i}^{\text{all atoms}} \mathrm{e}^{-\eta\left(R_{ij}-R_{\mathrm{s}}\right)^2} f_{\mathrm{C}}\left(R_{ij}\right)$$

$$f_{\mathrm{C}}\left(R_{ij}\right) = \begin{cases} 0.5 \times \cos\left(\dfrac{\pi R_{ij}}{R_{\mathrm{C}}}\right) + 0.5 & \text{for } R_{ij} \leq R_{\mathrm{C}} \\ 0.0 & \text{for } R_{ij} > R_{\mathrm{C}} \end{cases}$$



$$2.7\text{Å} \rightarrow \begin{bmatrix} 0 \\ 0 \\ 0.0177 \\ 0.3111 \\ 0.1509 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

Features corresponding to different distances are independent

Behler et al. PRL 98, 146401 (2007)
Smith et al. Chem. Sci., 2017, 8, 3192–3203

Behler-Parrinello symmetry function

- Atom environment $[G_1^X, G_2^X, \ldots]$

Radial

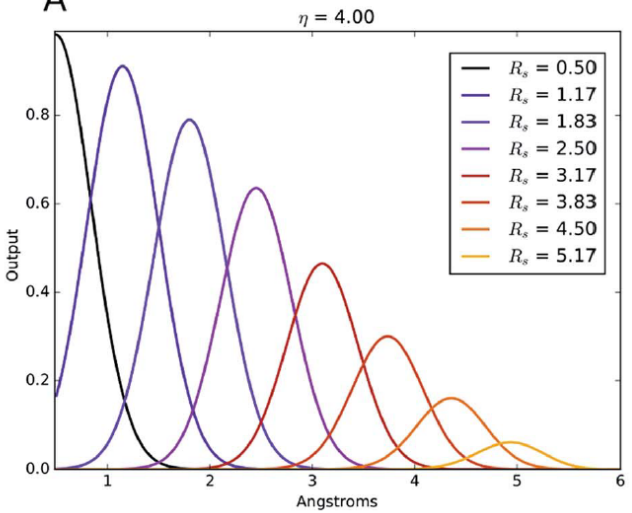
$$G_m^{\mathrm{R}} = \sum_{j \neq i}^{\text{all atoms}} \mathrm{e}^{-\eta\left(R_{ij}-R_{\mathrm{s}}\right)^2} f_{\mathrm{C}}\left(R_{ij}\right)$$

$$f_{\mathrm{C}}\left(R_{ij}\right) = \begin{cases} 0.5 \times \cos\left(\dfrac{\pi R_{ij}}{R_{\mathrm{C}}}\right) + 0.5 & \text{for } R_{ij} \leq R_{\mathrm{C}} \\ 0.0 & \text{for } R_{ij} > R_{\mathrm{C}} \end{cases}$$

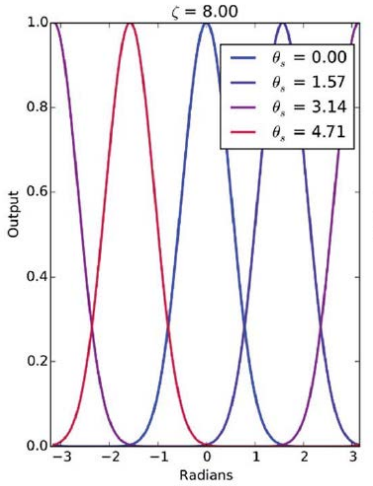
Angular

$$G_m^{\mathrm{A_{mod}}} = 2^{1-\zeta} \sum_{j,k \neq i}^{\text{all atoms}} \left(1 + \cos\left(\theta_{ijk} - \theta_{\mathrm{s}}\right)\right)^{\zeta} \times \exp\left[-\eta\left(\frac{R_{ij} + R_{ik}}{2} - R_{\mathrm{s}}\right)^2\right] f_{\mathrm{C}}\left(R_{ij}\right) f_{\mathrm{C}}\left(R_{ik}\right)$$

A



B



Bispectrum

Atomic density on the center atom

$$\rho_i(\mathbf{r}) = \delta(\mathbf{r}) + \sum_j \delta(\mathbf{r} - \mathbf{r}_{ij}) f_{\text{cut}}(|\mathbf{r}_{ij}|),$$

Projection of atomic density onto spherical coordinate

$$(\phi, \theta, \theta_0) = [\tan^{-1}(y/x), \cos^{-1}(z/|\mathbf{r}|), |\mathbf{r}|/r_0],$$

Expand in 4D spherical harmonics

$$c^j_{m'm} = \langle U^j_{m'm} | \rho \rangle.$$

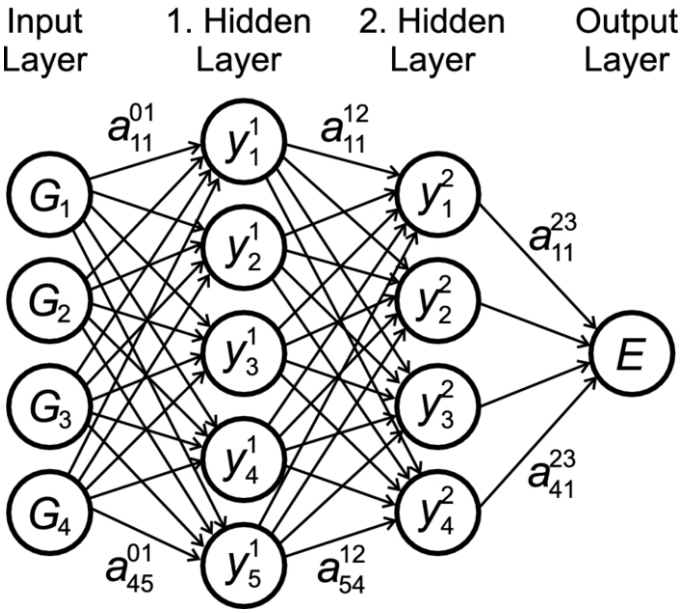
Coulomb matrix

$$M_{IJ} = \begin{cases} 0.5Z_I^{2.4} & \text{for } I = J, \\ \frac{Z_I Z_J}{|\mathbf{R}_I - \mathbf{R}_J|} & \text{for } I \neq J. \end{cases}$$

Atomic energy
Coulomb repulsion

The eigenvalues of Coulomb matrix preserves symmetry

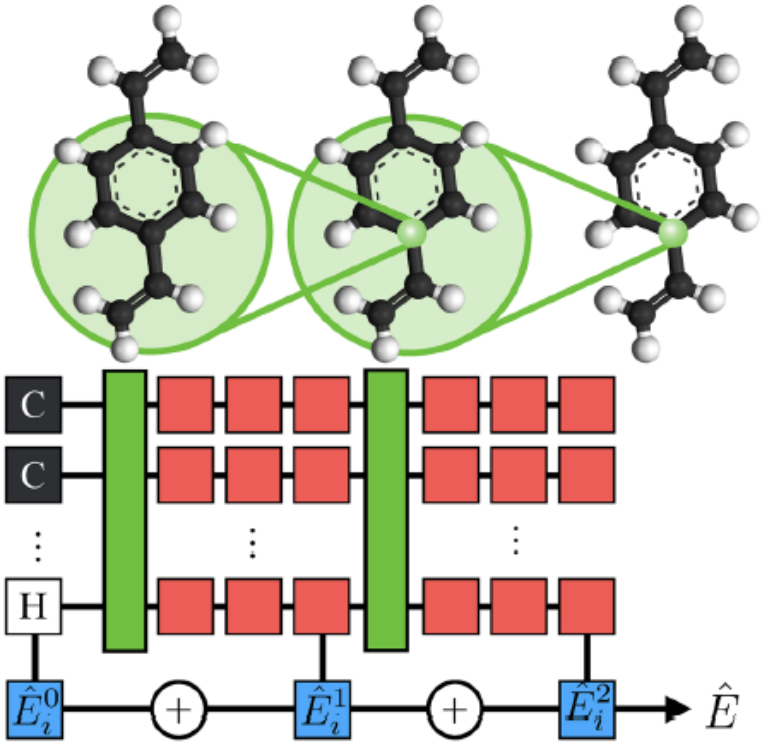
Fully connected network



3-6 hidden layers

32-128 nodes per layer

Hierarchically interacting particle



$$\hat{E}_i^n = \sum_{a=1}^{N_{\text{feature}}} w_a^n z_{i,a}^{\ell_n} + b^n,$$

In each layer, information from neighbor atoms is collected

Kernel ridge regression/GPR

$$\bar{y}(x) = \sum_i w_i K(x, x_i)$$

Linear regression using kernel function as a basis
Kernel function is a similarity measure between input x and training structure x_i

$$K(A, B) = \sum_{i \in A, j \in B} P_{ij}\, k(x_i, x_j)$$

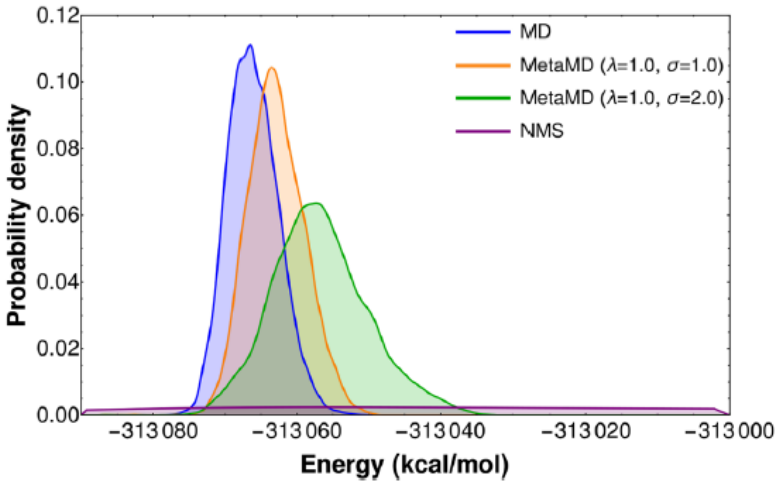
Kernel function between two structures is a sum of atomic contributions

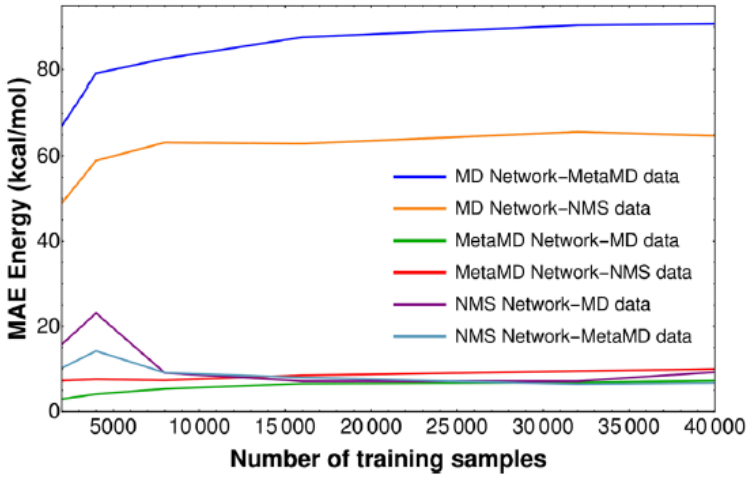
Weight factor can be computed directly

Bartók et al., Sci. Adv. 2017;3: e1701816
Chmiela et al., Sci. Adv. 2017;3: e1603015

Data generation method

- Ab initio Molecular Dynamics
- Metadynamics
- Normal model sampling

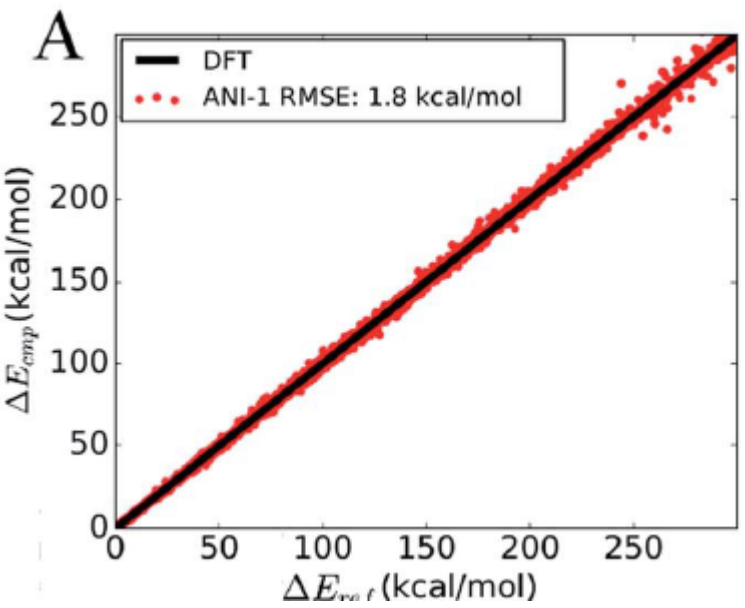


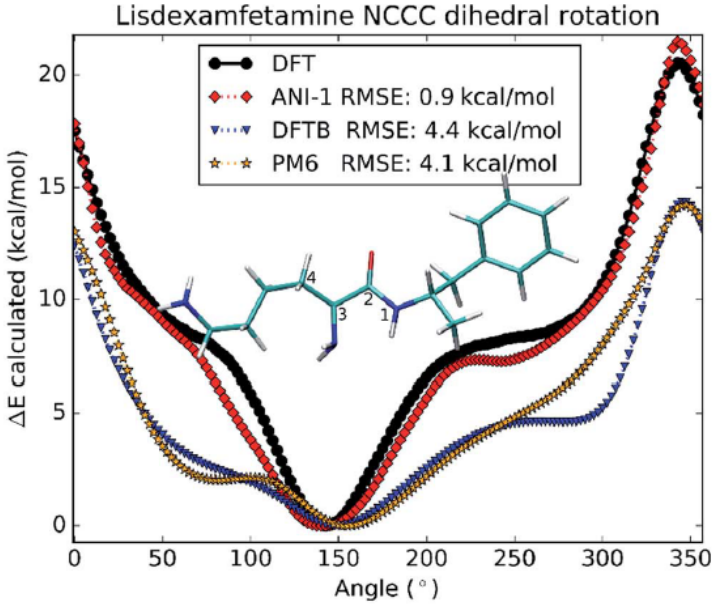


Normal model sampling is a decent method for generating structures
Metadynamics is a more automatic approach

ANI-1

- Modified Behler-Parrinello, 32 radial parameters with cutoff of 4.6 Ang.
- Fully connected network: 768:128:128:64:1
- Ref data: Normal mode sampling
  17.2m conformations of 58k molecules containing up to 8 heavy atoms of CHNO





Test on energies of 10 heavy atoms

TensorMol-0.1

$$E\left(\vec{R}\right) = \sum_i E^{\mathrm{BP}}_{z_i}\left(S_\alpha\left(\vec{R}\right)\right) + \sum_{ij} E^{\mathrm{DSF}}_{ij}\left(q_{ij}\left(S_\alpha\left(\vec{R}\right)\right), R_i, R_j\right) + E^{\mathrm{VDW}}\left(\vec{R}_{ij}\right)$$

$$\begin{cases} E_{\mathrm{DSF}} = E_{\mathrm{DSF(original)}} & R > R_{\mathrm{switch}} \\ q_i q_j \left(a_{\mathrm{elu}} \mathrm{e}^{R - R_{\mathrm{switch}}} + \beta_{\mathrm{elu}}\right) & R < R_{\mathrm{switch}} \end{cases}$$

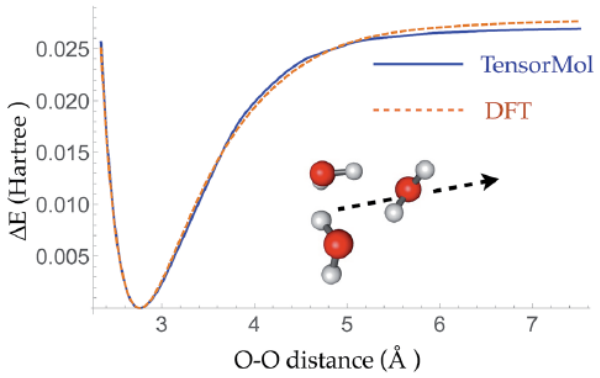
BP Neural net + truncated Coulomb interaction + vdW

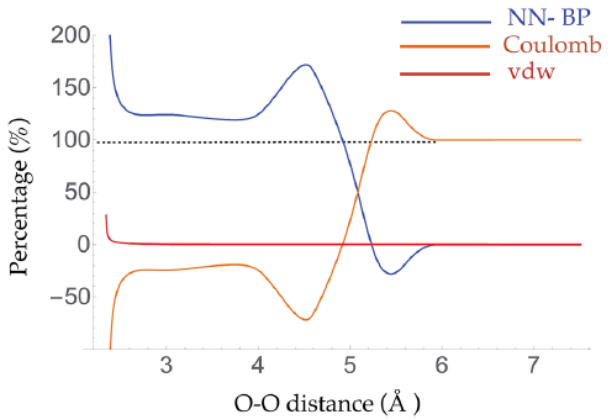
Separate network for charges to reproduce molecular dipoles

$$L_{\mathrm{dipole}} = \sum_A \left(\frac{\mu_A^{\mathrm{DFT}} - \mu_A^{\mathrm{NN}}\left(q_i, q_j, \ldots\right)}{N_{\mathrm{atom}}} \right)^2$$

$$L_{\mathrm{energy}} = \sum_A \left(\frac{E_A^{\mathrm{DFT}} - E_A^{\mathrm{NN}}}{N_{\mathrm{atom}}} \right)^2 + \gamma \sum_A \left(\frac{F_A^{\mathrm{DFT}} - F_A^{\mathrm{NN}}}{N_{\mathrm{atom}}} \right)^2$$

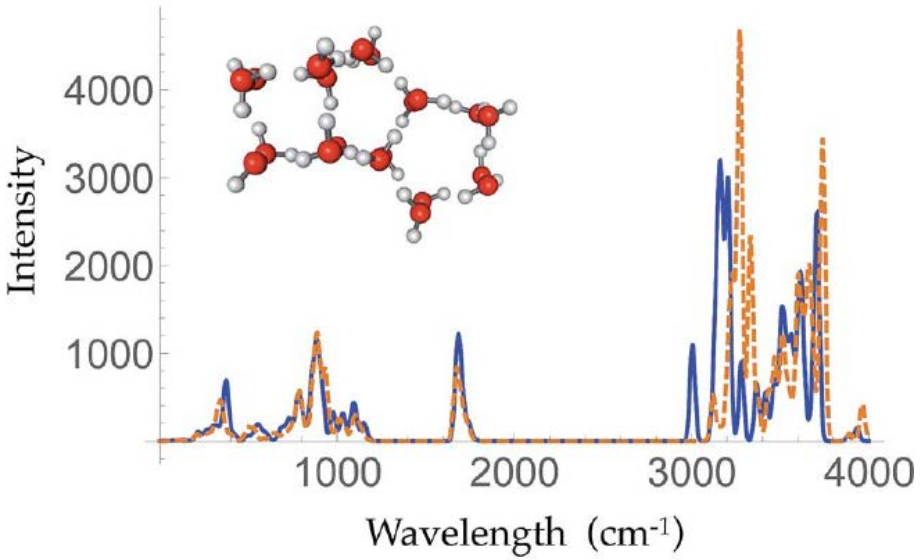
TensorMol-0.1



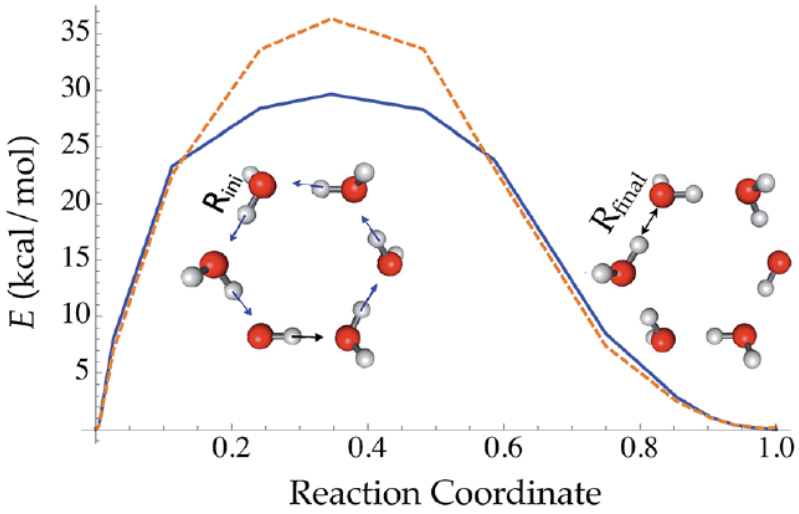


TensorMol-0.1

IR spectrum

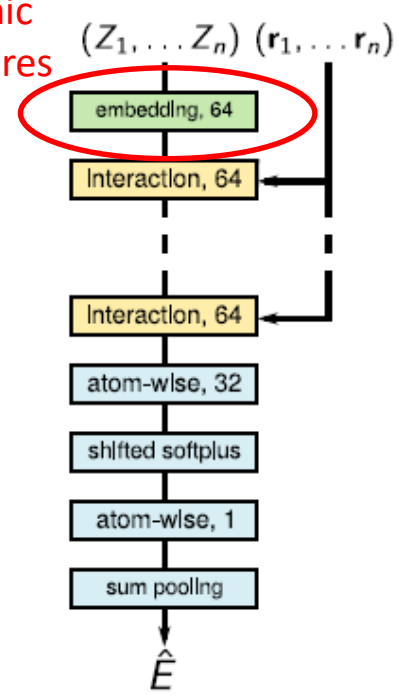


Proton transfer

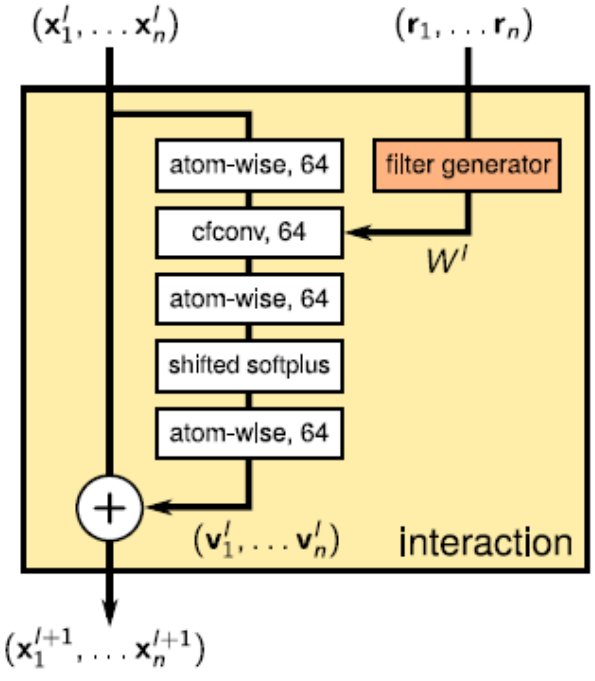


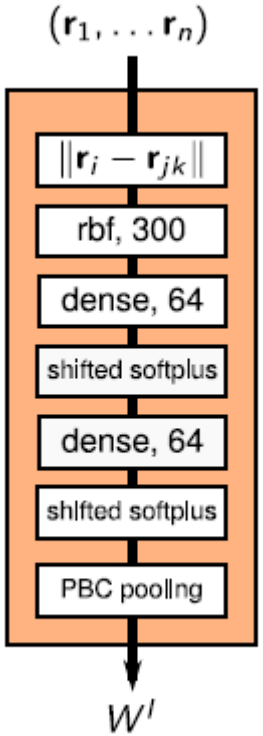
SchNet

Atomic features



6 interaction blocks





cfconv $$\mathbf{x}_i^{l+1} = (X^l * W^l)_i = \sum_{j=0}^{n_{\text{atoms}}} \mathbf{x}_j^l \circ W^l(\mathbf{r}_j - \mathbf{r}_i),$$

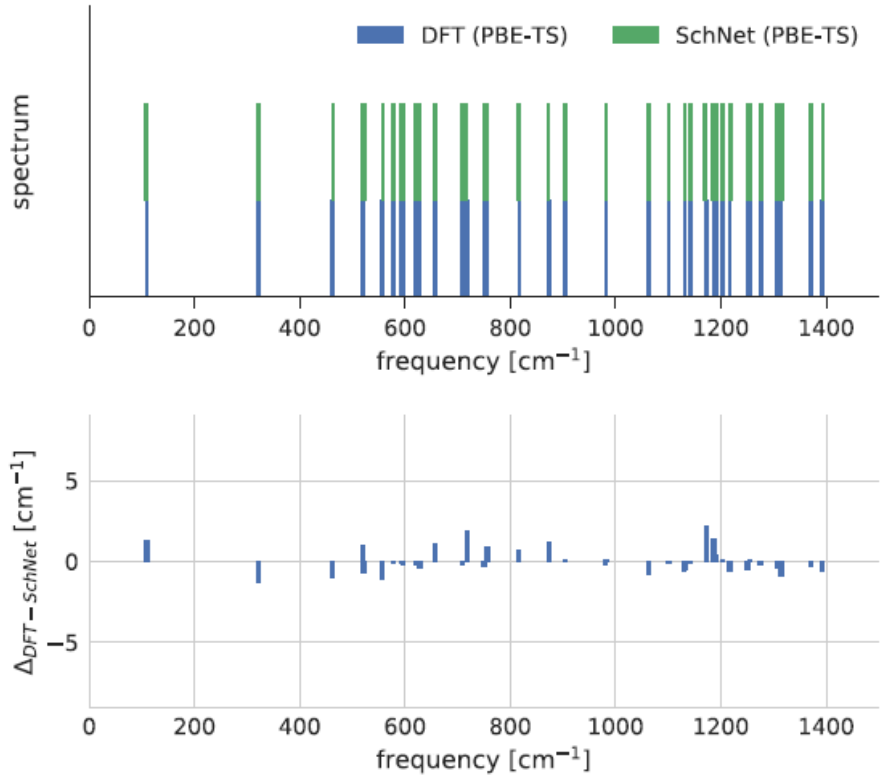
rbf $$K(\mathbf{x}, \mathbf{x}') = \exp(-\gamma \|\mathbf{x} - \mathbf{x}'\|^2)$$

SchNet

TABLE III. Mean absolute errors for total energies (in kcal/mol). GDML,[21] DTNN,[28] and SchNet[30] test errors for N = 1000 and N = 50 000 reference calculations of molecular dynamics simulations of small organic molecules are shown. Best results are given in bold.

	$N = 1000$			$N = 50\,000$		
	GDML	SchNet		DTNN	SchNet	
Trained on	Forces	Energy	Energy + forces	Energy	Energy	Energy + forces
Benzene	**0.07**	1.19	0.08	**0.04**	0.08	0.07
Toluene	**0.12**	2.95	**0.12**	0.18	0.16	**0.09**
Malondialdehyde	0.16	2.03	**0.13**	0.19	0.13	**0.08**
Salicylic acid	**0.12**	3.27	0.20	0.41	0.25	**0.10**
Aspirin	**0.27**	4.20	0.37	–	0.25	**0.12**
Ethanol	0.15	0.93	**0.08**	–	0.07	**0.05**
Uracil	**0.11**	2.26	0.14	–	0.13	**0.10**
Naphthalene	**0.12**	3.58	0.16	–	0.20	**0.11**

SchNet



Normal mode analysis of C_{20}

HIP-NN

$$N_{\text{feature}} = 80$$

$$z_{i,a}^{\ell+1} \underset{\text{inter.}}{=} f\left(\sum_{j,b} v_{ab}^{\ell}(r_{ij}) z_{j,b}^{\ell} + \sum_{b} W_{ab}^{\ell} z_{i,b}^{\ell} + B_a^{\ell}\right),$$

Pairwise distances

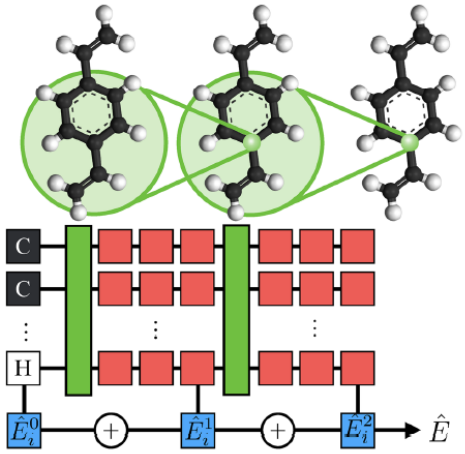


TABLE I. QM9 performance (MAE in kcal/mol) for various models reported in the literature. Best results for each training category are shown in bold.

$N_{\text{train}} + N_{\text{validate}}$	HIP-NN	MTM[22]	SchNet[33]	MPNN[31]	HDAD + KRR[20]	DTNN[32]
110 426	**0.256** ± 0.003	...	0.31	0.42	0.58	...
100 000	**0.261** ± 0.002	...	0.34	...	...	0.84
50 000	**0.354** ± 0.004	0.41	0.49	...	...	0.94

HIP-NN

TABLE II. QM9 performance for HIP-NN models with varying complexity.

N_{feature}	Parameter count	MAE (kcal/mol)	RMSE (kcal/mol)	Errors above 1 kcal/mol (%)
5	1.6k	1.177 ± 0.014	1.851 ± 0.019	42.18 ± 0.58
10	4.9k	0.653 ± 0.007	1.077 ± 0.015	18.95 ± 0.37
20	17k	0.398 ± 0.005	0.706 ± 0.025	6.60 ± 0.16
40	61k	0.274 ± 0.003	0.539 ± 0.014	2.65 ± 0.06
60	134k	0.261 ± 0.004	0.552 ± 0.024	2.37 ± 0.09
80	234k	0.256 ± 0.003	0.527 ± 0.020	2.26 ± 0.07
60 (no hierarchy)	134k	0.278 ± 0.008	0.522 ± 0.013	2.75 ± 0.21
80 (no hierarchy)	234k	0.293 ± 0.008	0.539 ± 0.013	3.10 ± 0.21

HIP reduces the number of outliers

Neural network energy model

Molecular descriptor

Behler-Parrinello
Bispectrum
Coulomb matrix

Neural network

Fully connected network
ResNet
HIP-NN

Reference data

DFT or CC
Energy, force
Normal mode sampling, metadynamics
Active sampling

Neural network energy model

- Molecular descriptors based on symmetry functions (radial and angular bases) can achieve chemical accuracy
- More recent work shows that using pairwise distances combined with atomic feature vectors can further improve the accuracy
- Using physical constraint, e.g. many-body energy expansion, is an interesting approach to address the overfitting problem
- Challenges are mainly on how to reduce number of parameters, improve transferability and scale to more elements

Problems and future directions

- Quadratic increase in number of parameters (and training data size) when adding new elements, since each element has independent parameters
  - Using vectorized features or physical properties to represent atoms
  - Adding physical constraint between parameters

- When augmented with physical long-range forces, the accuracy is limited by the physical force (especially atomic charge)
  - Better ways for representing long-range interactions

- Non-local effect, e.g. for extended aromatic system
  - HIP-NN
  - Convolution NN
  - Better functions and longer cutoff distance