Stony Brook University

Combining DFT and Machine Learning

Towards faster and more accurate ab-initio calculations

Sebastian Dick, Department of Physics and Astronomy, Stony Brook University
Fernandez-Serra Group
Jr. Researcher Award, 07/25/2019

Stony Brook University

Recap, what has changed?

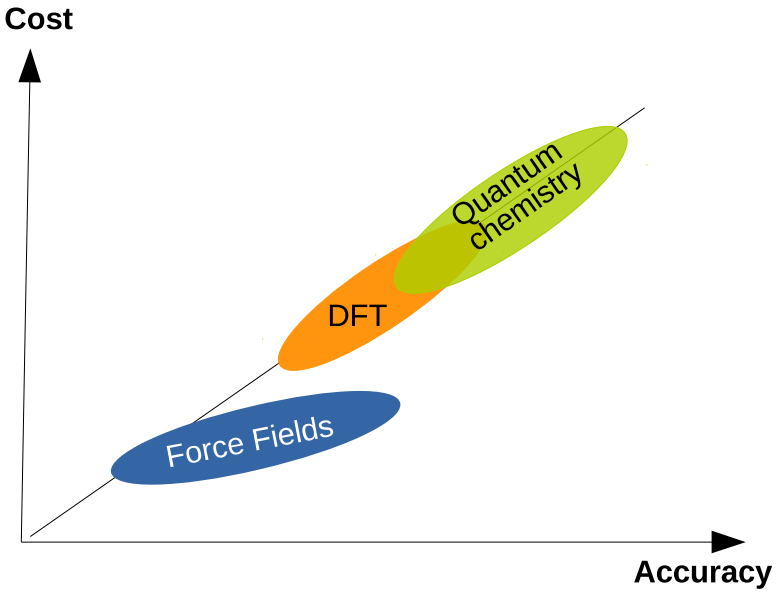
Simulations in Molecular Sciences

Stony Brook University

Atomic coordinates

- **Force Fields**
- **Density Functional Theory (DFT)**
- **Quantum Chemistry**

Energies, **Forces**, Stress, Electron density, Spectra, ...



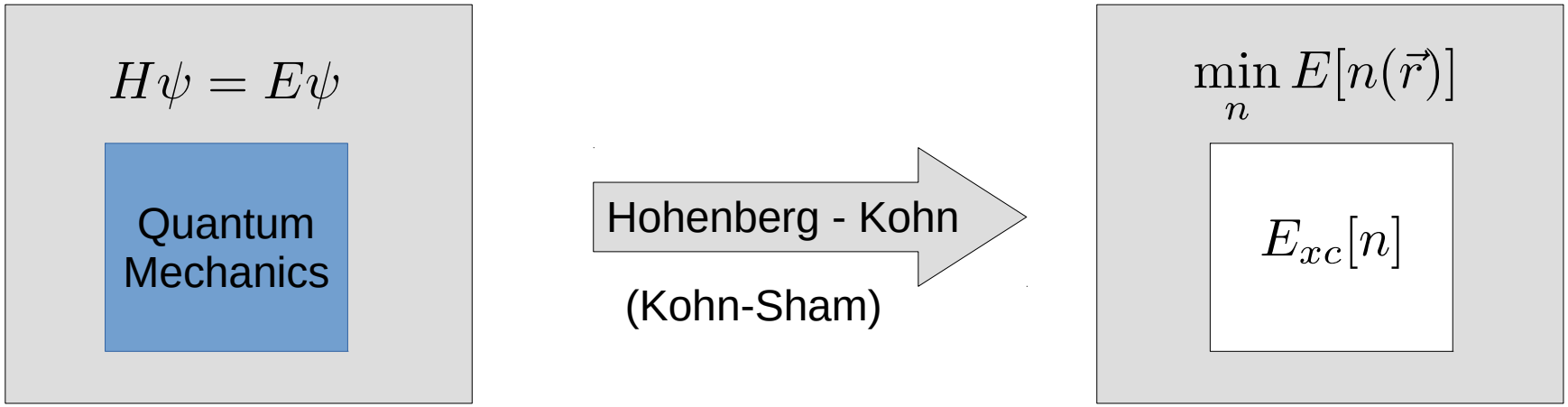
We use DFT because:

- Can scale to large systems sizes (100s to 1000s of atoms) + Periodic boundary conditions

  → Condensed systems

- Non-empirical, hence unbiased
- Fully reactive

How does DFT work ?



$$E[n] = T[n] + E_{xc}[n] + \int d\vec{r}\, V_{ext}(\vec{r}) n(\vec{r}) + E_{Hartree}[n] + E_{II}$$

Instead of solving Schrodinger equation, solve auxiliary system of non-interacting electrons (represented through electron density n).

Procedure: Calculate $\min_n E[n(\vec{r})]$ self-consistently → obtain energy, forces and other prop. for minimizing density

Problem: $E_{xc}[n]$ not exactly known, needs to be approximated

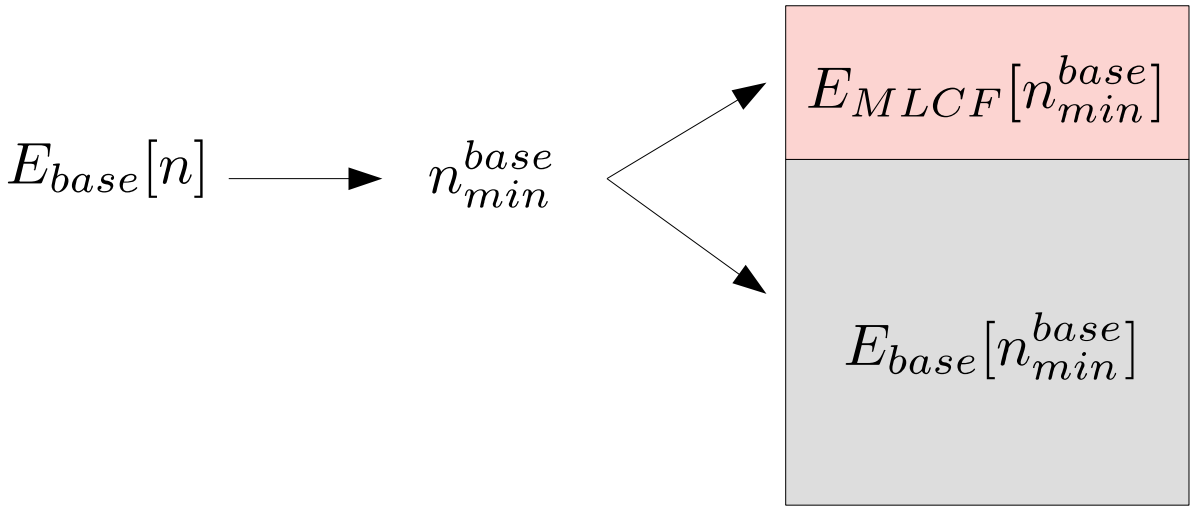
Our (original) method

Procedure: Calculate $\min_n E[n(\vec{r})]$ self-consistently → obtain energy, forces and other prop. for minimizing density

Problem: $E_{xc}[n]$ not exactly known, needs to be approximated

Idea: Do the above with a reasonably cheap baseline approximation $E_{base}[n]$,

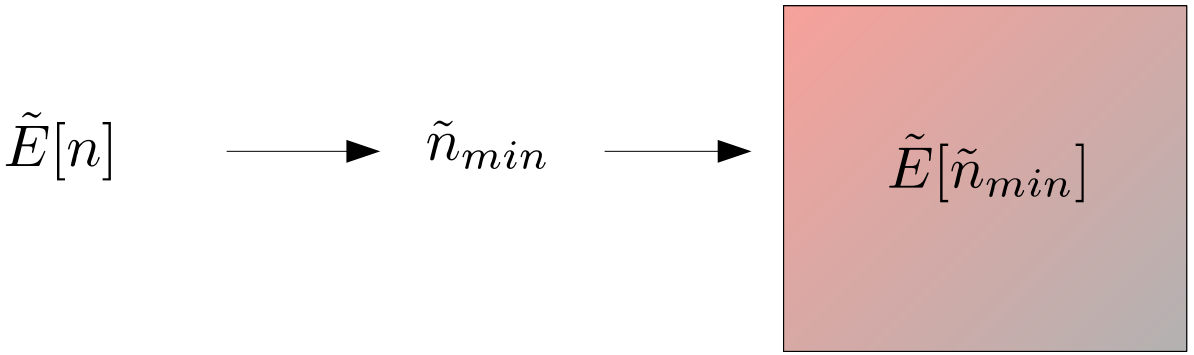
correct the energy with a machine-learned functional of the density $E_{MLCF}[n]$



Modifications

Instead of adding a correcting layer, integrate the machine-learned functional into the minimization process.

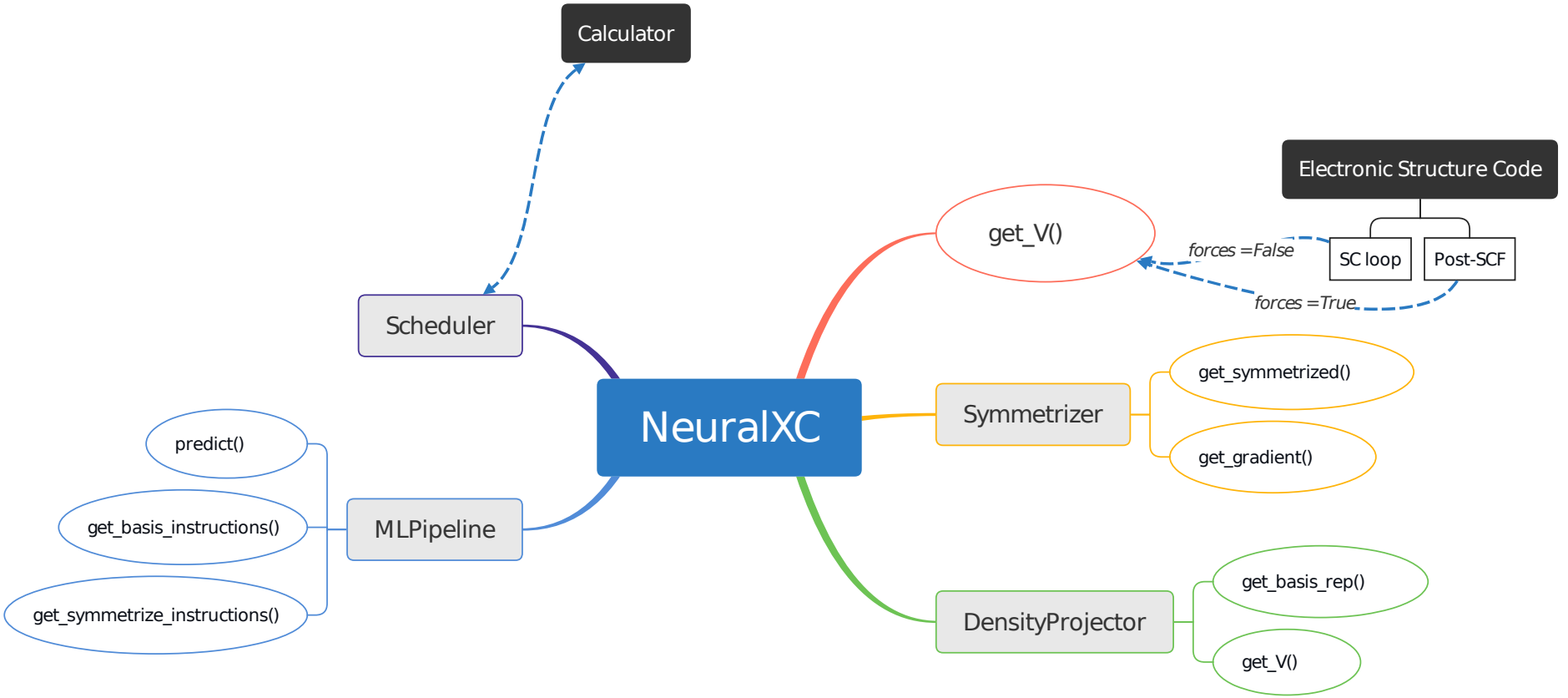
$$\tilde{E}[n] = E_{base}[n] + E_{MLCF}[n]$$

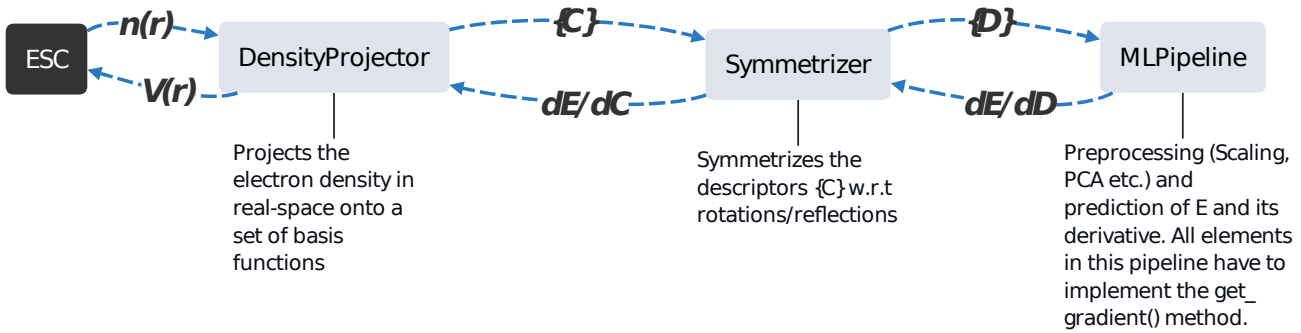


- \+ Can correct density as well
- \+ Energy-conserving forces can be obtained directly
- − Non-trivial implementation

Implementation

Stony Brook University





github.com/semodi/neuralxc

Stony Brook University

Results

Results - Water

Stony Brook University

Dataset:

Training: 640 Monomers, 1600 Dimers, 1200 Trimers

Testing: 160 Monomers, 400 Dimers, 300 Trimers, 50 Tetramers, 50 Pentamers, …

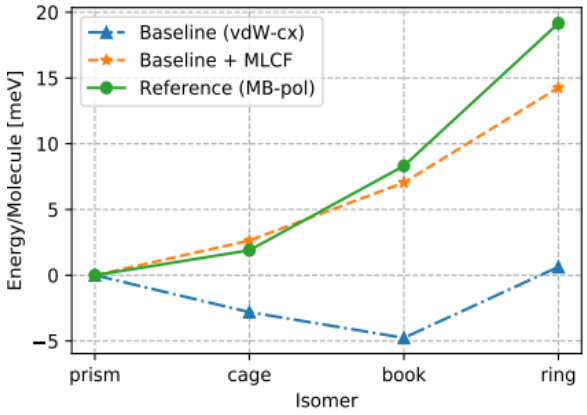
Baseline: vdW-cx (GGA)

Reference: MB-pol

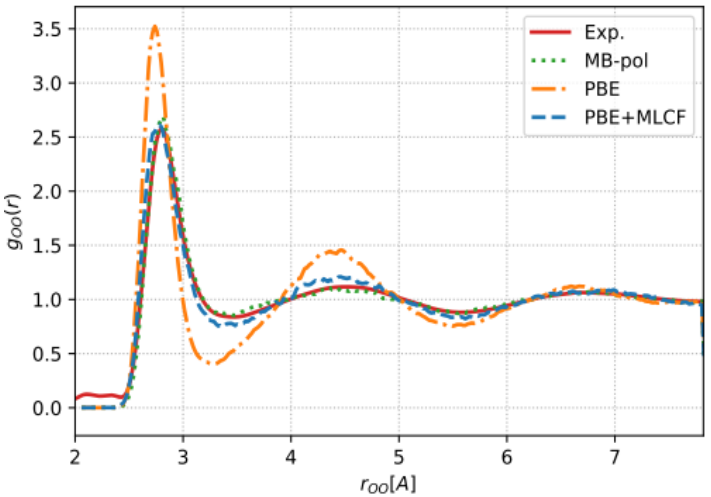
Test set accuracies (all energies in meV/Molecule)

No. Molecules	RMSE	MAE	max. Error
1	2.29 (53.19)	1.25 (43.13)	14.71 (151.19)
2	4.33 (40.17)	2.91 (31.28)	31.03 (136.91)
3	2.89 (28.25)	2.16 (22.29)	12.27 (75.20)
4	2.79 (9.69)	2.15 (7.93)	7.70 (24.95)
5	2.64 (11.24)	2.19 (8.97)	6.23 (36.69)
8	3.43 (9.26)	2.72 (7.34)	8.05 (22.67)
16	2.75 (6.28)	2.15 (5.03)	6.19 (17.15)

Hexamers



Oxygen-Oxygen pair correlation function



Exp: L. B. Skinner et al. J. Chem. Phys. 138, 074506 (2013)

SD, Fernandez-Serra (submitted), arXiv:1812.06572

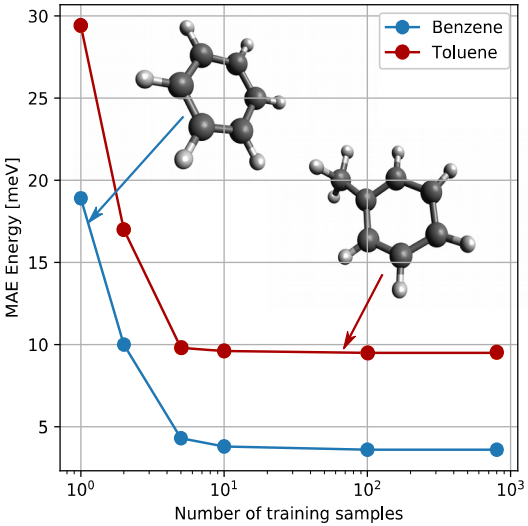
Results - Benzene

Stony Brook University

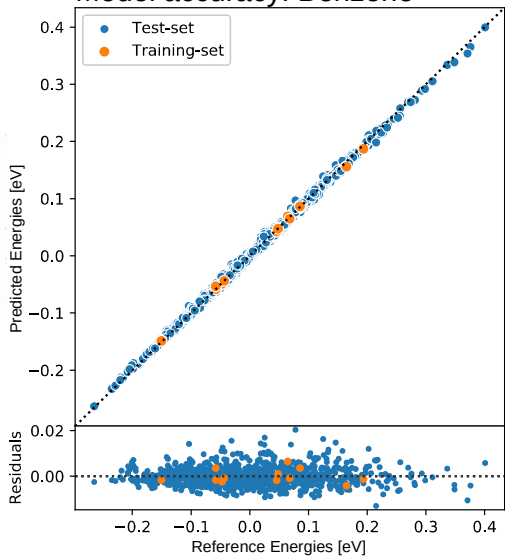
Dataset obtained from [1]:
Training: 1000 snapshots from MD
Testing: 500 snapshots
Baseline: PBE
Reference: CCSD(T) (cc-pVDZ)

- *Data-efficiency:* model reaches full accuracy at only 8-10 training points.
- *Transferability:* model trained on benzene still remains (somewhat) valid for toluene
- *Accurate forces:* self-consistent, conservative forces with MAE of less than 1kcal/mol make molecular dynamics at CCSD(T) level possible

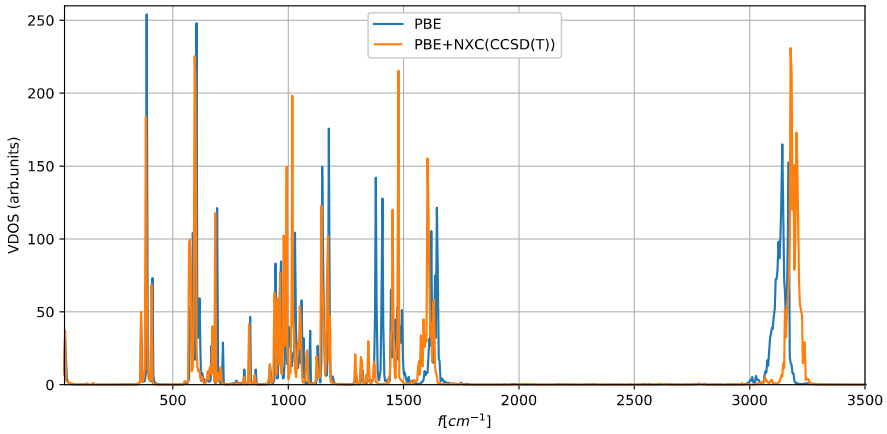
Test set MAE of a model trained on Benzene



Model accuracy: Benzene



VDOS: Benzene, 300K



	PBE(DZP)	PBE(DZP) + MLCF	CCSD(T)(cc-pVDZ)	CCSD(T)(cc-pVTZ)	Experiment
r_{CC}	1.4015	1.4096	1.4096	1.3917	1.397
r_{CH}	1.0997	1.0968	1.0968	1.0779	1.084

SD, Fernandez-Serra (in preparation)

[1] Chmiela, S., Sauceda, H. E., Müller, K.-R., Tkatchenko, A., Nature Communications, 9(1), 2018, 3887

Results – small molecules

Stony Brook University

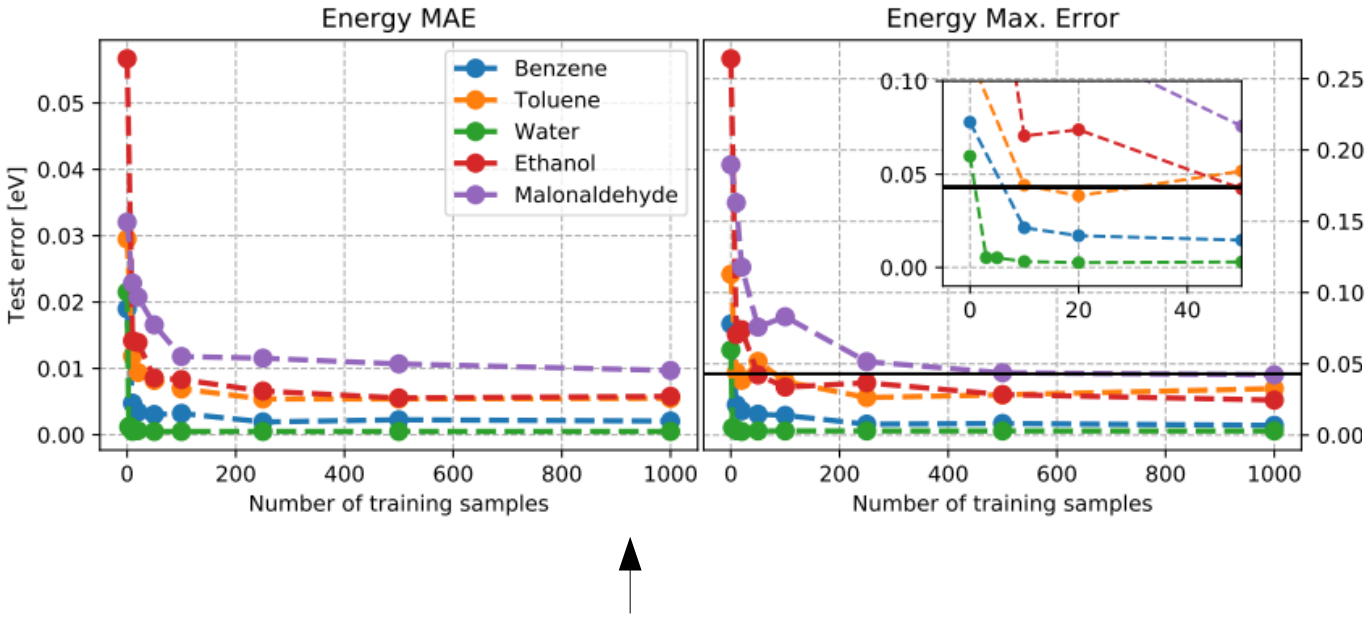
Dataset obtained from [1]:

Training: 1000 snapshots from MD

Testing: 500 snapshots

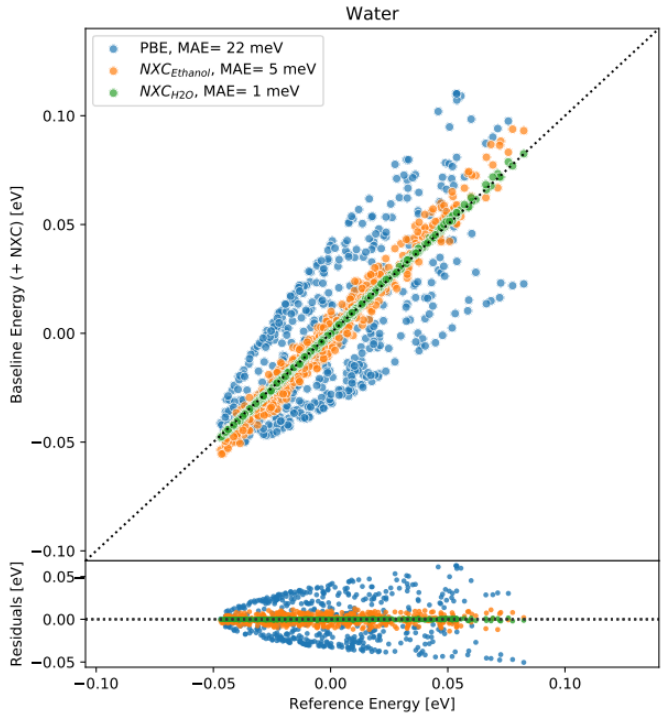
Baseline: PBE

Reference: CCSD(T) (cc-pVDZ)



50 training samples sufficient to reach max. errors below 1kcal/mol (“chemical accuracy”) for all molecules except Malonaldehyde (requires 250 – 500 samples)

Model trained on ethanol works well for water



SD, Fernandez-Serra (in preparation)

[1] Chmiela, S., Sauceda, H. E., Müller, K.-R., Tkatchenko, A., Nature Communications, 9(1), 2018, 3887

Stony Brook University

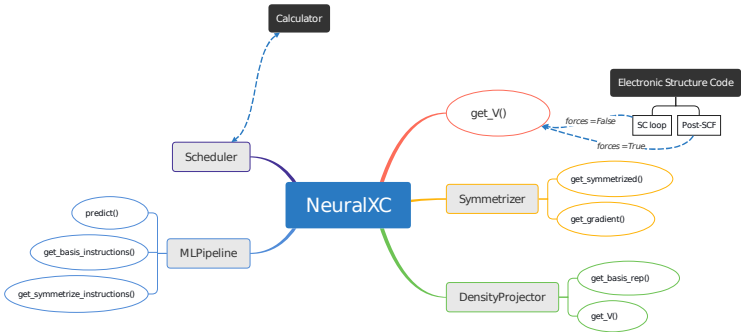
Fellowship/Collaborations

MolSSI Seed Fellowship

Stony Brook University

The 2019 Seed Fellows

My mentor,
Samuel Ellis

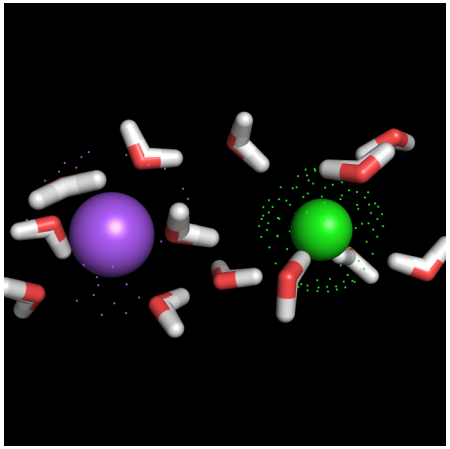


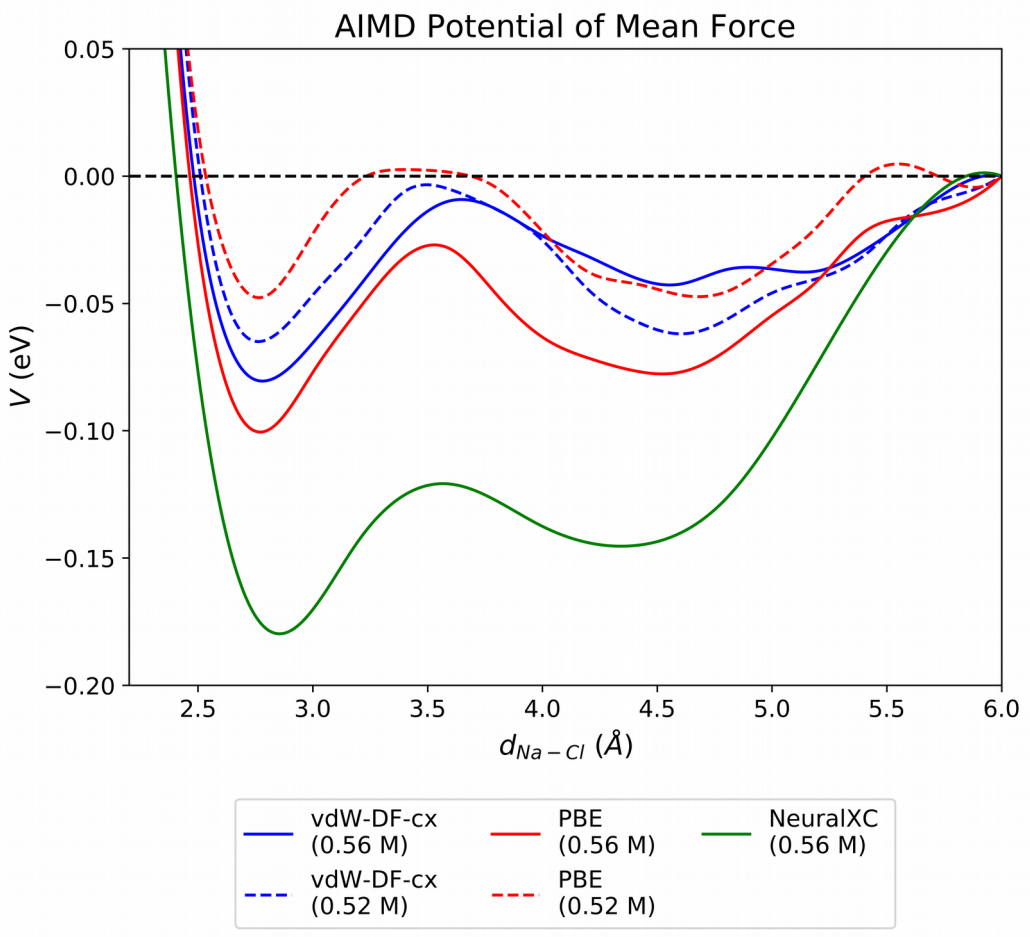
- Financial support and mentoring over a 6 month period (January to July 2019)
- Bootcamp in January introducing fellows to agile and sustainable software development
- Weekly Skype meetings with my mentor Sam Ellis to set short-term goals and report progress

Water + Ion calculations

Stony Brook University

Alec Wills





Collaborations

Stony Brook University

Alberto Torres

Luana Pedroza

unesp

“Investigating the behavior of water on gold surfaces”

Linfeng Zhang

Roberto Car

Chemistry in Solution and at Interfaces
Computational Chemical Science Center

PRINCETON UNIVERSITY

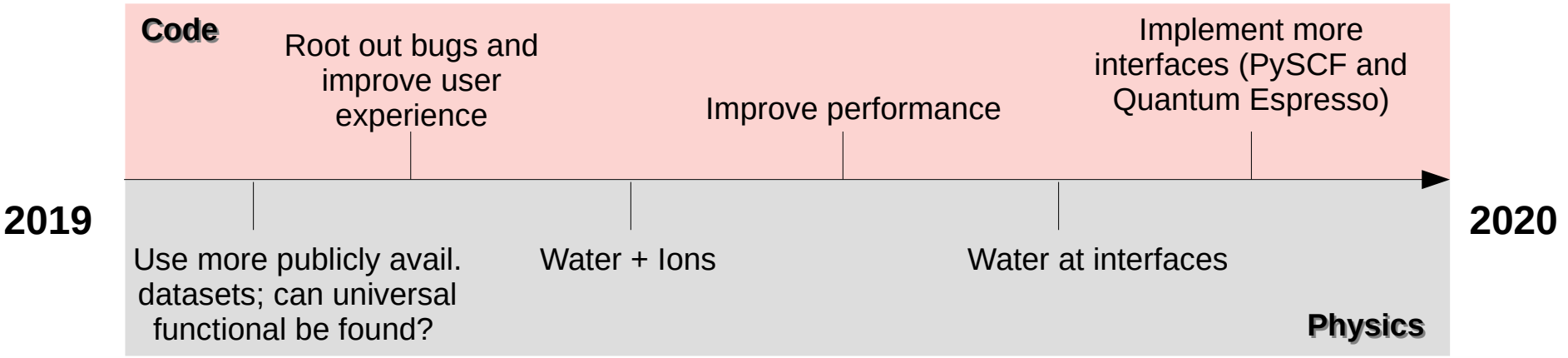
Collaboration in planning

Conclusion and Outlook

Stony Brook University

- We successfully trained a ML-based density functional that lifts PBE to the accuracy of a higher level method
- The method is highly data efficient
- Once trained, it can be used at small additional computational cost
- It respects physical symmetries and conserves energy
- First results on transferability seem to indicate that method learns underlying physics

What remains to be done

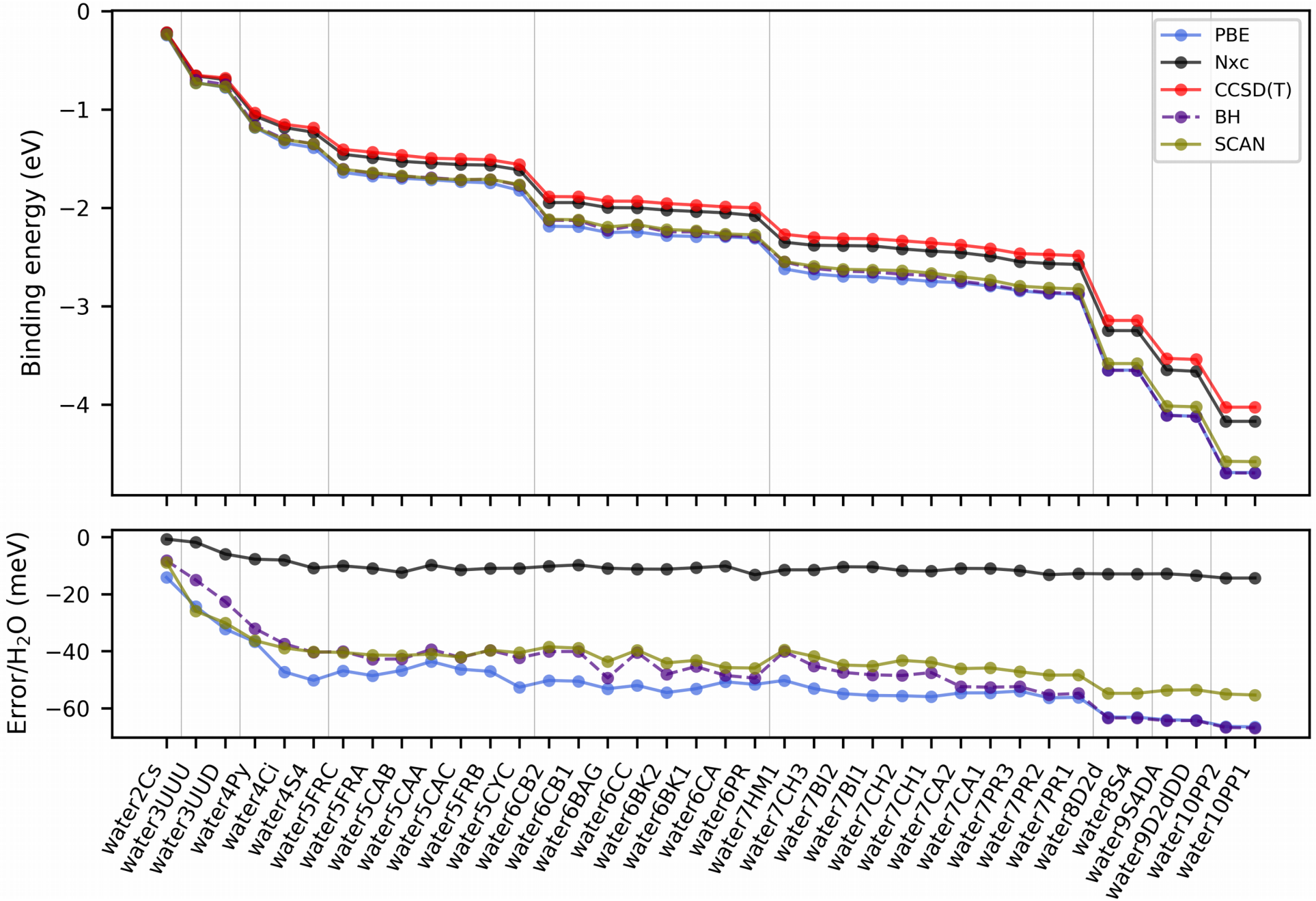


Stony Brook University

Thank you!

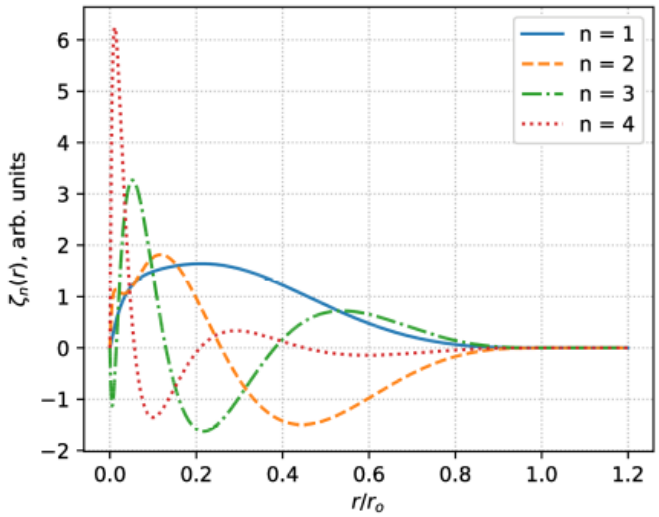
Water clusters

Stony Brook University



- **Input:** Expansion of electron density around each atom into orthonormal basis functions:

$$\psi_{nlm}(\vec{r}) = Y_l^m(\theta,\phi)\zeta_n(r)$$



Electronic descriptors:

$$c_{nlm}^{\alpha,I} = \int_{\vec{r}} \rho(\vec{r} - \vec{r}_{\alpha,I})\psi_{nlm}^{*\alpha}(\vec{r})$$

Atomic species

Atom index

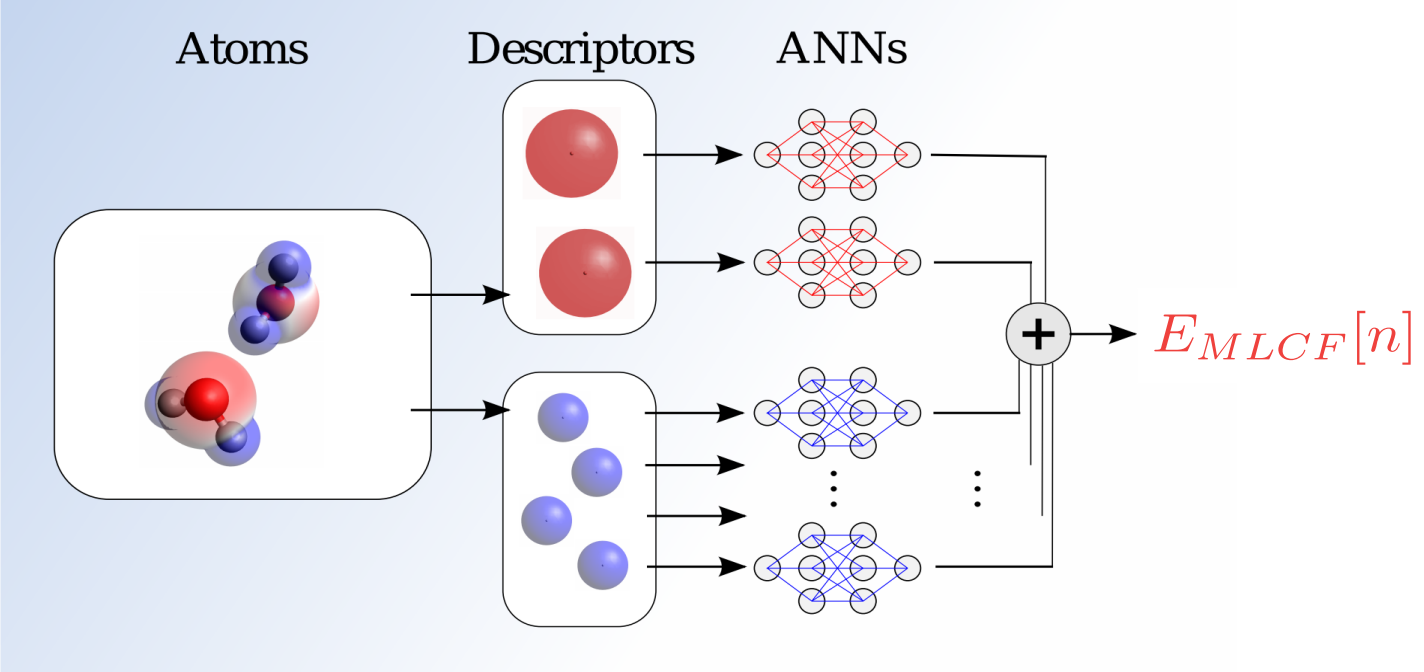
- **Targets:** Difference between reference and baseline energies

$$E_{MLCF}(\mathbf{c}[\rho]) = E^{ref}[\rho] - E^{base}[\rho]$$

- **Potential:**

$$V_{MLCF}[\rho(\vec{r})] = \frac{\delta E_{MLCF}}{\delta\rho(\vec{r})} = \sum_\beta \frac{\partial E_{MLCF}(\mathbf{c}[\rho])}{\partial c_\beta}\psi_\beta^*(\vec{r})$$

$$\tilde{V}_{xc}[\rho(\vec{r})] = V_{xc}^{base}[\rho(\vec{r})] + V_{MLCF}[\rho(\vec{r})]$$


Atoms
Descriptors
ANNs
$E_{MLCF}[n]$