

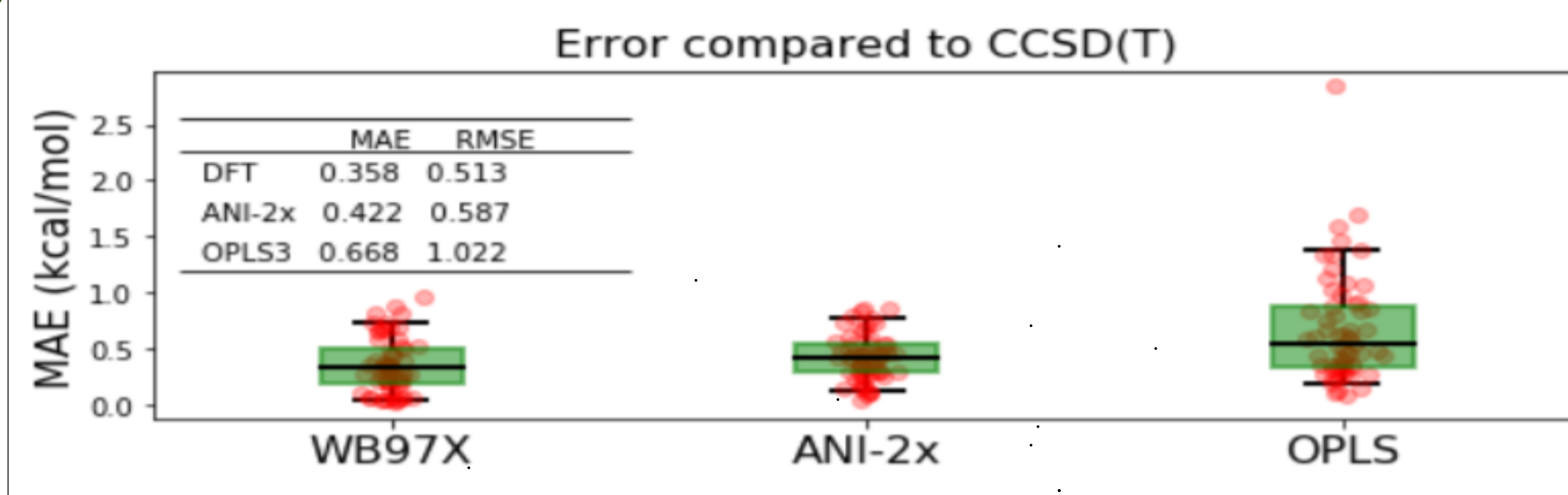
AIM by ANI: The Development of an Atoms-in-Molecules Partition through Machine Learning



Kate K. Davis, Adrian E. Roitberg, and Ramón A. Miranda-Quintana
Department of Chemistry, University of Florida, Gainesville, FL



Introduction

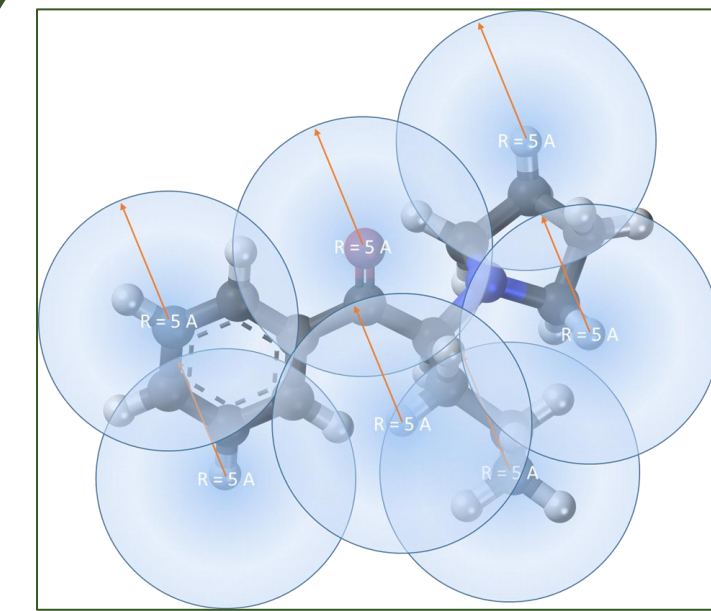


The above figure shows error metrics for energy values in comparison with CCSD(T), a high accuracy *ab initio* quantum calculation method, for ANI-2x, the latest published ANI model, its reference level of theory, DFT's wb97x/6-31g*, and OPLS, a force field method.

- Computational chemistry:
 - Methods that approximate the Schrodinger Equation
 - Costly and limiting depending on accuracy demanded and system size
- Using Machine Learning, the ANAKIN-ME method, or ANI, has bridged the gap between computational cost and accuracy.

- ANI
 - transferable, deep learning neural network potential (NNP)
 - predicts energies and forces for organic molecules
- A NNP is a statistical learning model that approximates the potential surface of a system when only given information on the structure and composition.

Atoms in Molecules



The purpose of this work is to continue the extension of ANI by fitting a NNP to a new function and in turn develop a new form of an AIM scheme, while still keeping the original chemical accuracy and efficiency of ANI.

$$E_{QM} = \sum_A E_A \rightarrow E_{QM} = \sum_A E_A + \frac{1}{2} \sum_a q_a \sum_B \frac{q_B}{R_{AB}}$$

- Interested in the trends observed when looking at the atomic contributions to the energy and their corresponding charge values.
- Assume no relationship about the energies and charges
- Expected that we are observing pure interaction energies, so that each atomic energy corresponds to the charge states of the atom in a molecule

AIM populations are widely used in conceptual DFT to compute reactivity descriptors

$$E(q) = E(q_o = 0) + \chi q + \frac{1}{2} \eta q^2$$

Modified Energy Shifter

Atom	SAE (ANI-1x)	GSAE (Gaussian wb97x/6-31g*)
H	-0.600952980000	-0.499321200000
C	-38.08316124000	-37.83383340000
N	-54.70775770000	-54.57328250000
O	-75.19446356000	-75.04245190000

SAE → pure interaction energies

DFT ground state energies, or GSAEs can be used instead!

The table above compares ANI-1x SAE values to GSAE values calculated at the wb97x/6-31g* level of theory. This validates that we must use ground state energies to achieve the prediction of pure interaction energies.

	Average Training RMSE	Std Deviation of TRMSE	Average Validation RMSE	Std Deviation of VRMSE	Average # of Epochs
ANI-2x	1.6329	0.0096	1.8331	0.0372	972.750
ANI-1x	1.9910	0.0272	3.1020	0.0516	1538.125
RXN Modified-1x	1.9914	0.0256	2.3609	0.0199	985.625
D2-GSAE Network	1.9123	0.0139	2.3418	0.0122	1559.875
MV-GSAE Network	1.8546	0.0202	2.2652	0.0262	1644.875

Upon the comparison of different ANI model training statistics, it was found that modifying the Energy Shifter did not effect accuracy but did effect training time.

Conservation Of Charge Methods

ANI only learns neutral molecules, yet it doesn't explicitly know charge neutrality. So, we must implement it!

- Electronegativity: an atom's tendency to attract electrons to itself
- Chemical Hardness: See mathematical definition in c-DFT section; viewed as half of the band gap in a chemical system.

In each method below we are considering the excess charge value, or the sum of the predicted charges:

$$\epsilon = \text{excess charge} = \sum_i q_i$$

- Method 1: Compute charge correction based on electronegativity.
 - Excess charge will be removed from the more electronegative atom.
 - The correction is atom type dependent.

χ : electronegativity
 η : chemical hardness

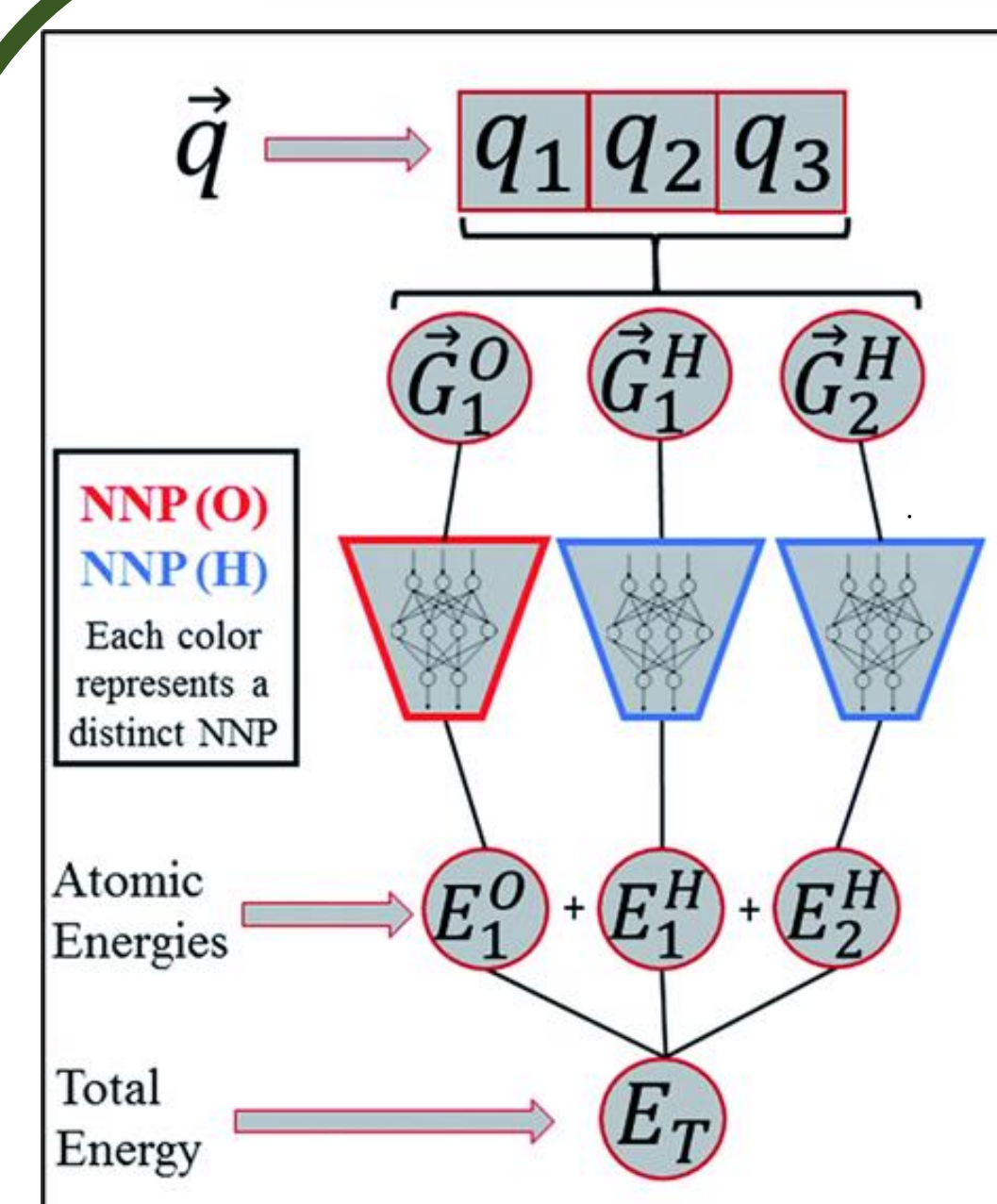
$$\Delta N_{max} = \frac{\chi}{\eta}$$

$$\Delta N_A = \frac{\epsilon}{\sum_B n_B \Delta N_{maxB}} \Delta N_{maxA}$$

Method 2: Compute charge correction in a similar fashion as Method 1, by incorporating the electronegativity. However, different atoms of the same type can be corrected by different magnitude proportional by the amount of charge the network assigns to the atom.

$$\Delta N_A = \frac{\epsilon}{\sum_B [q_B \Delta N_{Bmax}]^2} [q_A \Delta N_{Amax}]^2$$

ANAKIN-ME



The total energy of the system is computed through a simple summation over the atomic contributions. This functional form wouldn't be possible if not due to the nature of the AEVs.

$$E_{total} = \sum_A E_A$$

Coordinates

Coordinates can not act as the input to our network as they are not transferable.

AEV Computer

- Atomic Environment Vectors
 - fixed size vectors that describe a single atoms surroundings within a specific cutoff radius.
- subAEVs
 - radial part: record of atomic species present
 - angular part: record of the atomic pairs present

HD-Atomic NNPs

- HD-atomic NNP model
 - Behler and Parrinello
 - each available atomic number has a distinct network
 - corresponding AEVs act as the input.

Raw Atomic Energies

- Raw Atomic Energies
 - the output of the NNPs
 - small non-physical values
 - require correction through the implementation of the Energy Shifter

Energy Shifter

Compute Total Energy

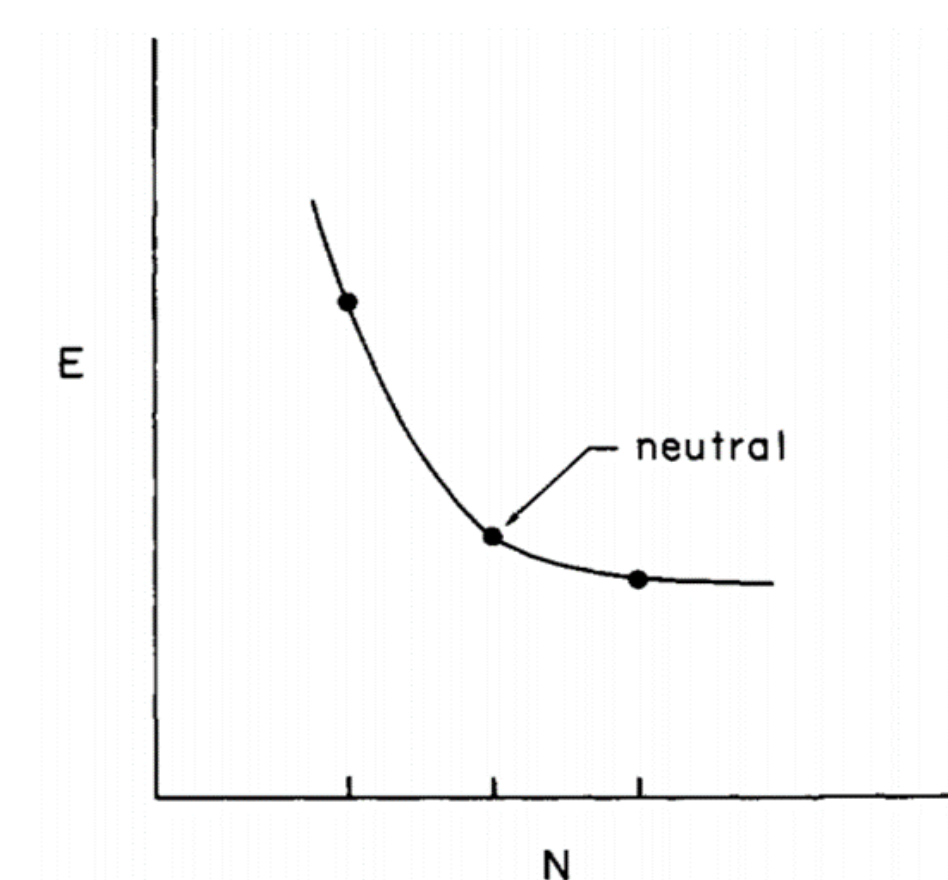
Conceptual DFT

$$E[\Delta N] = E[N_A] + \Delta N \left(\frac{\delta E[N_A]}{\delta N_A} \right) + \frac{(\Delta N)^2}{2} \frac{\delta^2 E[N_A]}{\delta N_A^2} \dots$$

$$N_A = Z_A - q_A$$

$$-\chi = \mu = \frac{\delta E[N_A]}{\delta N_A}$$

$$\eta = \frac{\delta^2 E[N_A]}{\delta N_A^2}$$



Example plot for the Parabolic Model, one of many models in Conceptual DFT. Here the electronic energy of an atom is plotted as a function of N, the total number of electrons.

Consider this: Experimentally you only have integral numbers of electrons. Consider a smooth curve where the electron population of an atom in a molecule would not necessarily be of integral value.

Conceptual Density Functional Theory, or c-DFT, is a theory in which the main idea is that a systems reactivity is defined by its response to perturbations.

- Remember we are making no assumptions about the E vs q relationship in our model!
- Non-parametric fit of our energy and charge data
- Check correspondence with many models in c-DFT

Conclusion

The extension of ANI to the development of an AIM partition with the methods and ideas presented would allow for a novel scheme and the opportunity for flexible non-parametric fitting of our predicted energy and charge values.

References

- Devereux, C. *et al.* Extending the Applicability of the ANI Deep Learning Molecular Potential to Sulfur and Halogens. *Journal of Chemical Theory and Computation* **16**, 4192–4202 (2020).
- Smith, J. S., Isayev, O. & Roitberg, A. E. ANI-1: an extensible neural network potential with DFT accuracy at force field computational cost. *Chemical Science* **8**, 3192–3203 (2017).
- Parr, R. G. & Pearson, R. G. Absolute Hardness: Companion Parameter to Absolute Electronegativity. *Journal of the American Chemical Society* **105**, 7512–7516 (1983).
- Behler, J. & Parrinello, M. Generalized neural-network representation of high-dimensional potential-energy surfaces. *Physical Review Letters* **98**, (2007).

Acknowledgments

Thank you to UF Informatics Institute, Funding provided by: HiperGator, NVIDIA, Adrian Roitberg and the Roitberg Group, Ramón Miranda Quintana, Olexander Isayev and his research group.

