Neural Network Potentials

ANI-1: an extensible neural network potential with DFT accuracy at force field computational cost

Smith, Isayev and Roitberg

Sabri Eyuboglu
February 6, 2018

What are Neural Network Potentials?

Neural Network Potentials are statistical learning models that approximate the potential energy of molecular systems

Why are they significant?

Molecular Dynamics Simulations

Molecular Dynamics Simulations

OBJECTIVE

Simulate the movements of atoms in a molecular system

APPROACH

Use **potential energy** to determine movement of the atoms in the system

```
for each time-step:
    Derive forces acting on each atom using potential energy
    Update position and velocity
```

Citation: CS 279 Slides, Ron Dror 2017

Potential Energy Function

A function mapping a molecular system's **geometry** to its **potential energy**

$$\vec{M} \rightarrow \boxed{\text{Potential Energy Function}} \rightarrow E$$

where

$\vec{M}$	**Molecular Representation** A vector describing the molecular system's geometry. Elements usually consist of atomic numbers and associated 3D coordinates.
E	**Potential Energy** The scalar potential energy of the molecular system.

Potential Energy Function

EXAMPLE

1-Dimensional Molecular Representation for a **Diatomic** Molecule

MOLECULAR REPRESENTATION

Bond Distance, q

$$\overrightarrow{G} = \{q\}$$

Potential Energy Function

EXAMPLE

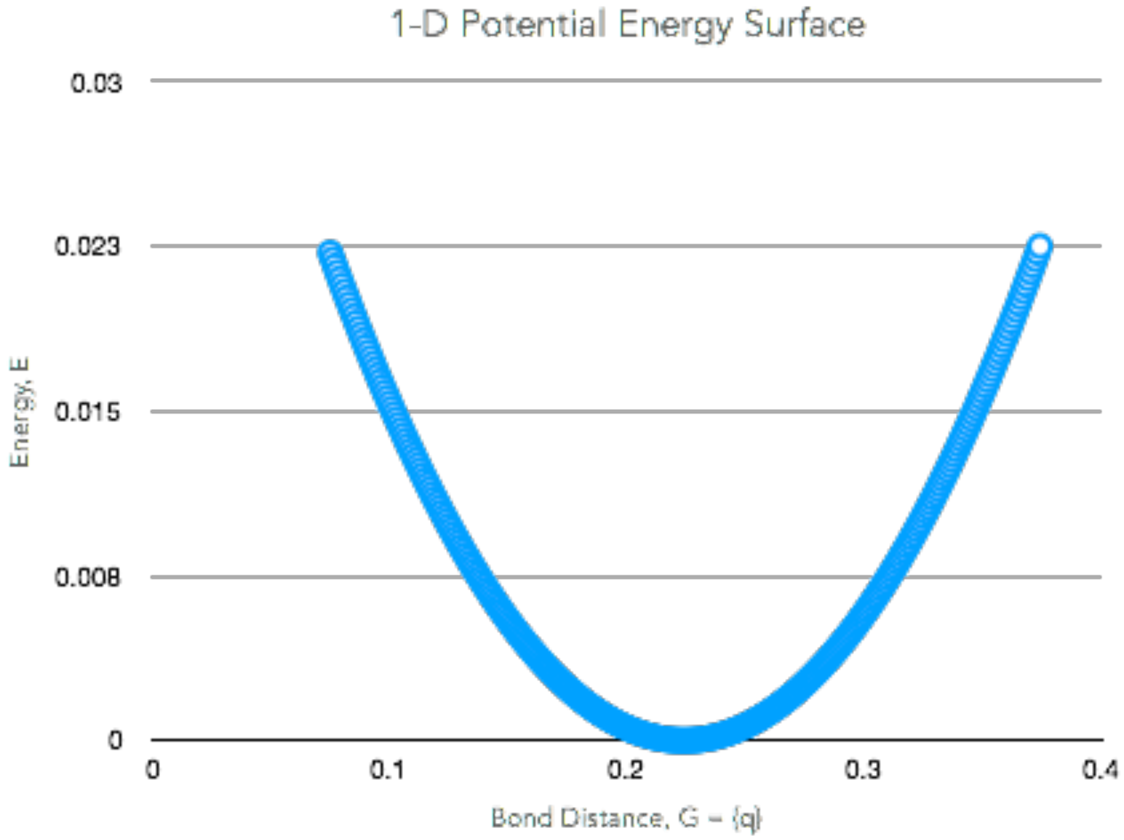
1-Dimensional Molecular Representation for a **Diatomic** Molecule

MOLECULAR REPRESENTATION

Bond Distance, q

$$\overrightarrow{G} = \{q\}$$

POTENTIAL ENERGY FUNCTION



THE PROBLEM

Potential Energy Function Approximation

Real molecular systems require elaborate molecular representations

Real potential energy functions are very **difficult** and **costly** to compute

MD Simulations require

Fast and **Reliable**

Potential Energy Function Approximations

Potential Energy Function Approximations

Method that computes the potential energy from a molecular representation

$$\vec{M} \longrightarrow \boxed{\text{Potential Energy Function Approx.}} \longrightarrow E$$

where

$\vec{M}$ **Molecular Representation**
A vector describing the molecular system's geometry.
Elements can include atom positions, bond lengths and/or angles.

E **Potential Energy**
The scalar potential energy of the molecular system.

METHODS OF
Potential Energy Function Approx.

Density Functional Theory (**DFT**) *ab initio* Methods

Proceed from first principles

ACCURATE
SLOW
TRANSFERABLE

Semi-Empirical Methods

Use empirically determined parameters to speed up DFT computation

LESS ACCURATE
FASTER
TRANSFERABLE

Empirical Methods

Classical Force Fields and Interatomic Potentials

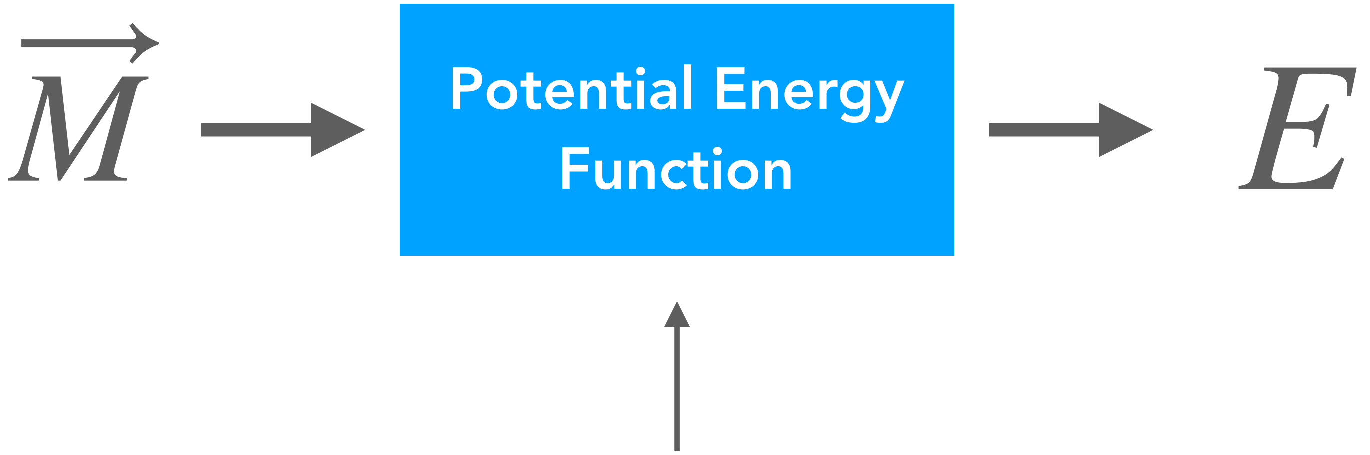
OFTEN INACCURATE
FAST
POOR TRANSFERABILITY

Statistical Learning with Neural Networks

? FAST and ACCURATE and TRANSFERABLE ?

Neural Networks for Regression

Statistical learning models that can learn a **very** diverse class of real-valued functions



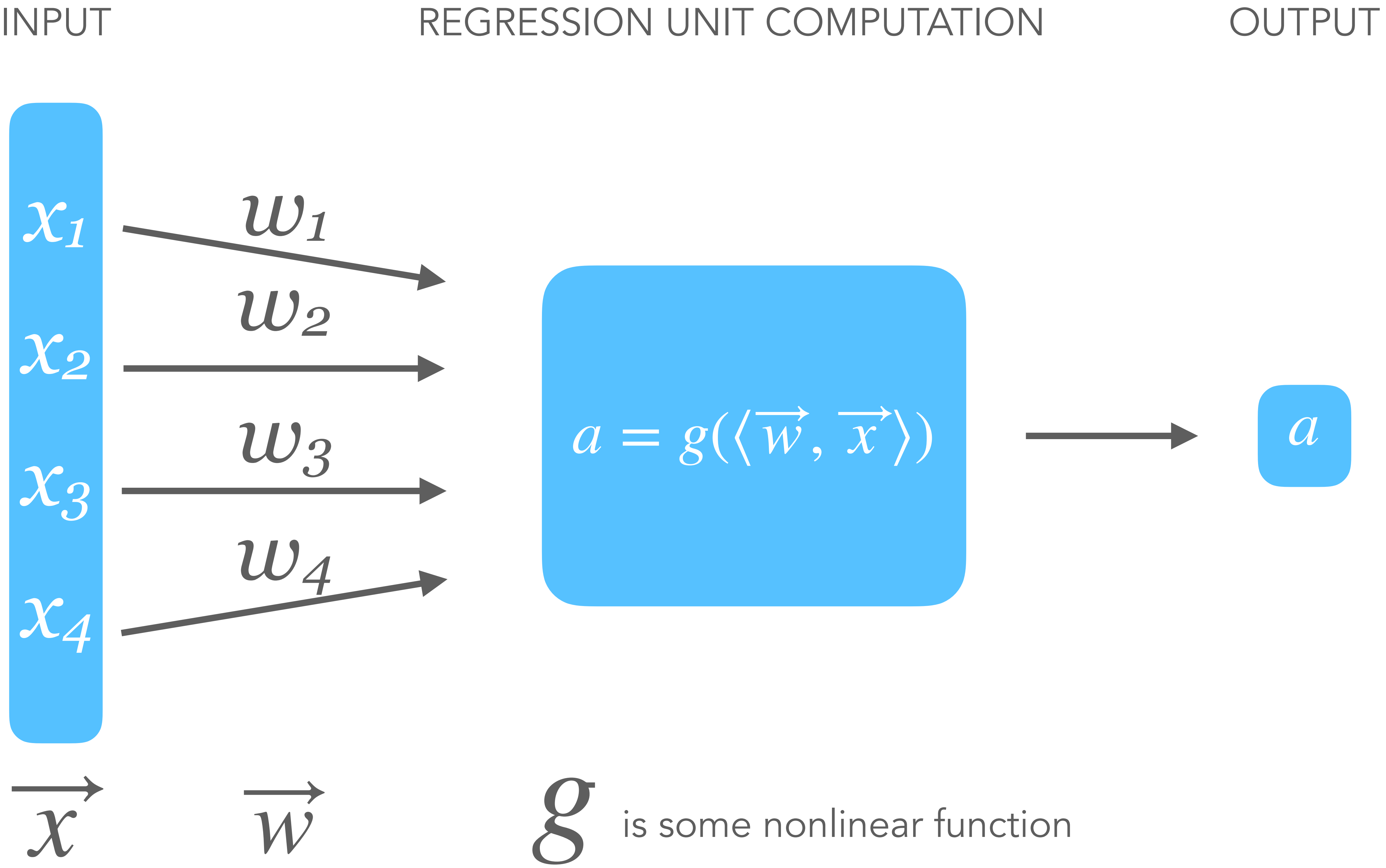
$\vec{M}$

Potential Energy Function

E

Could it be learned from **labeled** molecular data?

Regression Unit



INPUT

REGRESSION UNIT COMPUTATION

OUTPUT

x_1 x_2 x_3 x_4

w_1 w_2 w_3 w_4

$$a = g(\langle \vec{w}, \vec{x} \rangle)$$

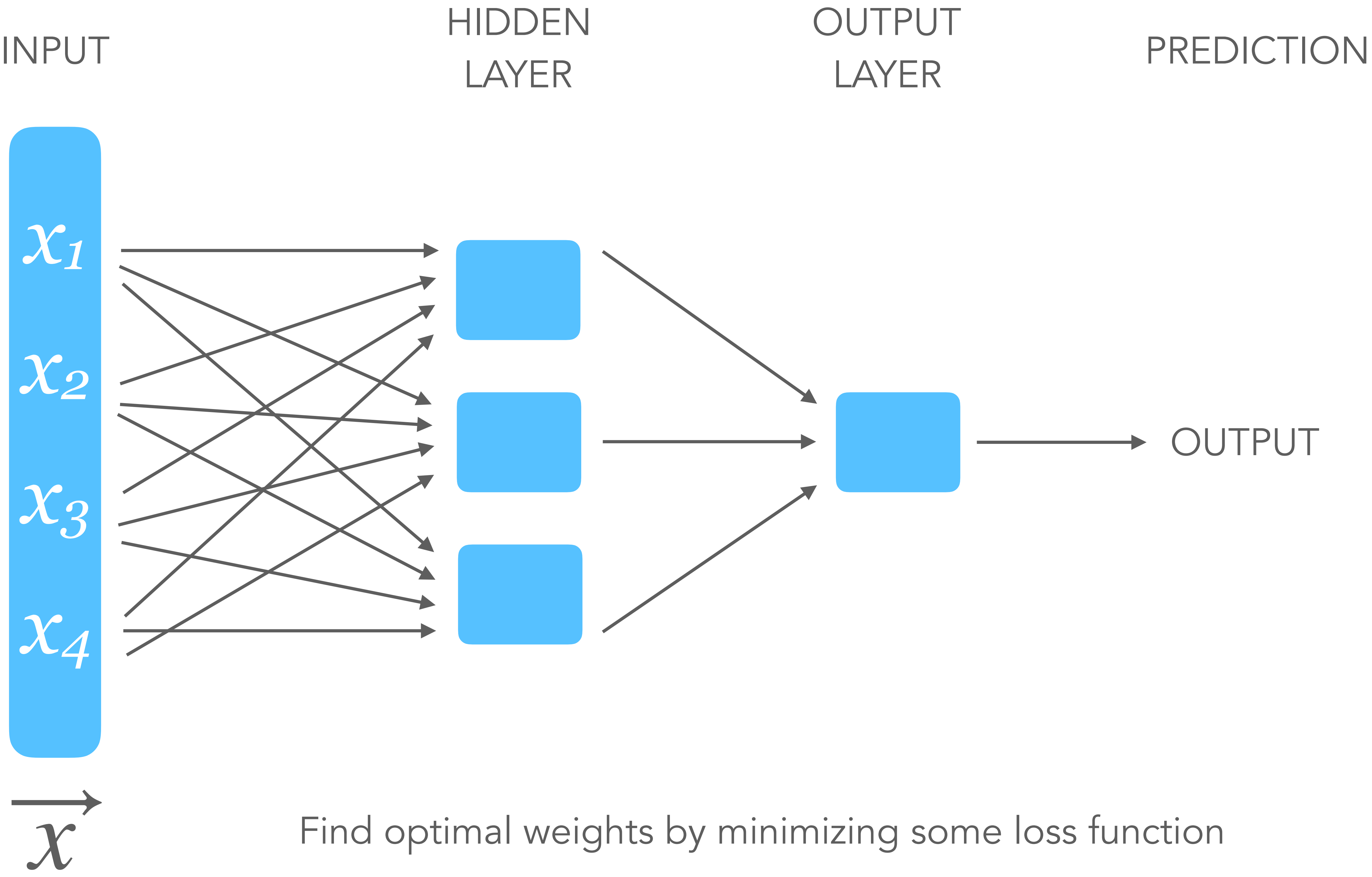
a

$\vec{x}$

$\vec{w}$

g is some nonlinear function

Neural Networks for Regression



INPUT

HIDDEN LAYER

OUTPUT LAYER

PREDICTION

x_1

x_2

x_3

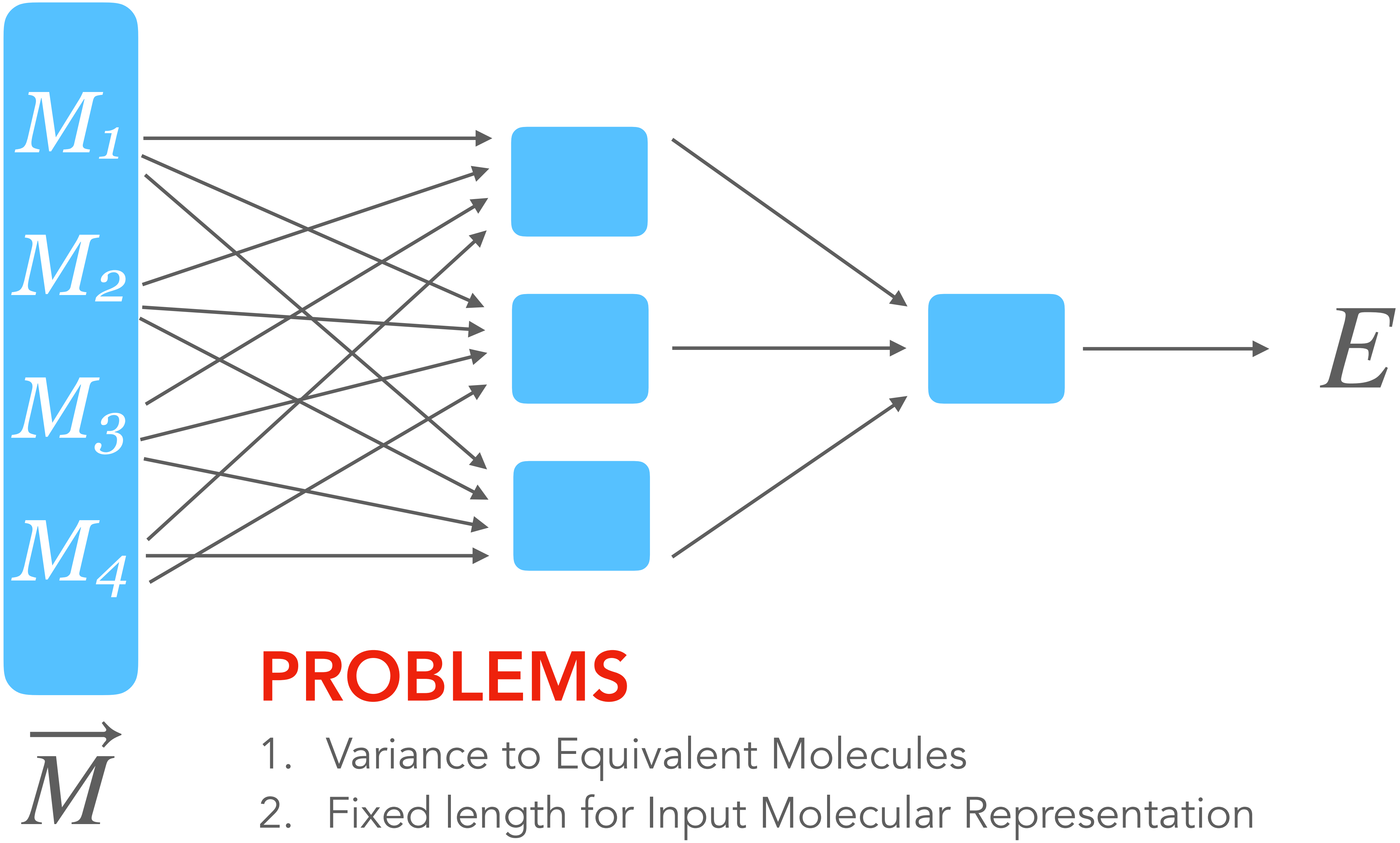
x_4

OUTPUT

$\vec{x}$

Find optimal weights by minimizing some loss function

Naive Neural Network Potential



MOLECULAR REPRESENTATION

HIDDEN LAYER

OUTPUT LAYER

POTENTIAL ENERGY

PROBLEMS

1. Variance to Equivalent Molecules
2. Fixed length for Input Molecular Representation

ANI-1 Neural Network Potential

IDEA: Atomic Decomposition

1. **Decompose** the molecular representation by atom
2. **Decompose** the energy function by atom

ANI-1 Neural Network Potential

Atomic Environment Vectors (AEV)

Decompose molecular representation of the systems total geometry to a sequence of molecular representations capturing the local geometry around an atom

Input: Coordinates of each atom in the system

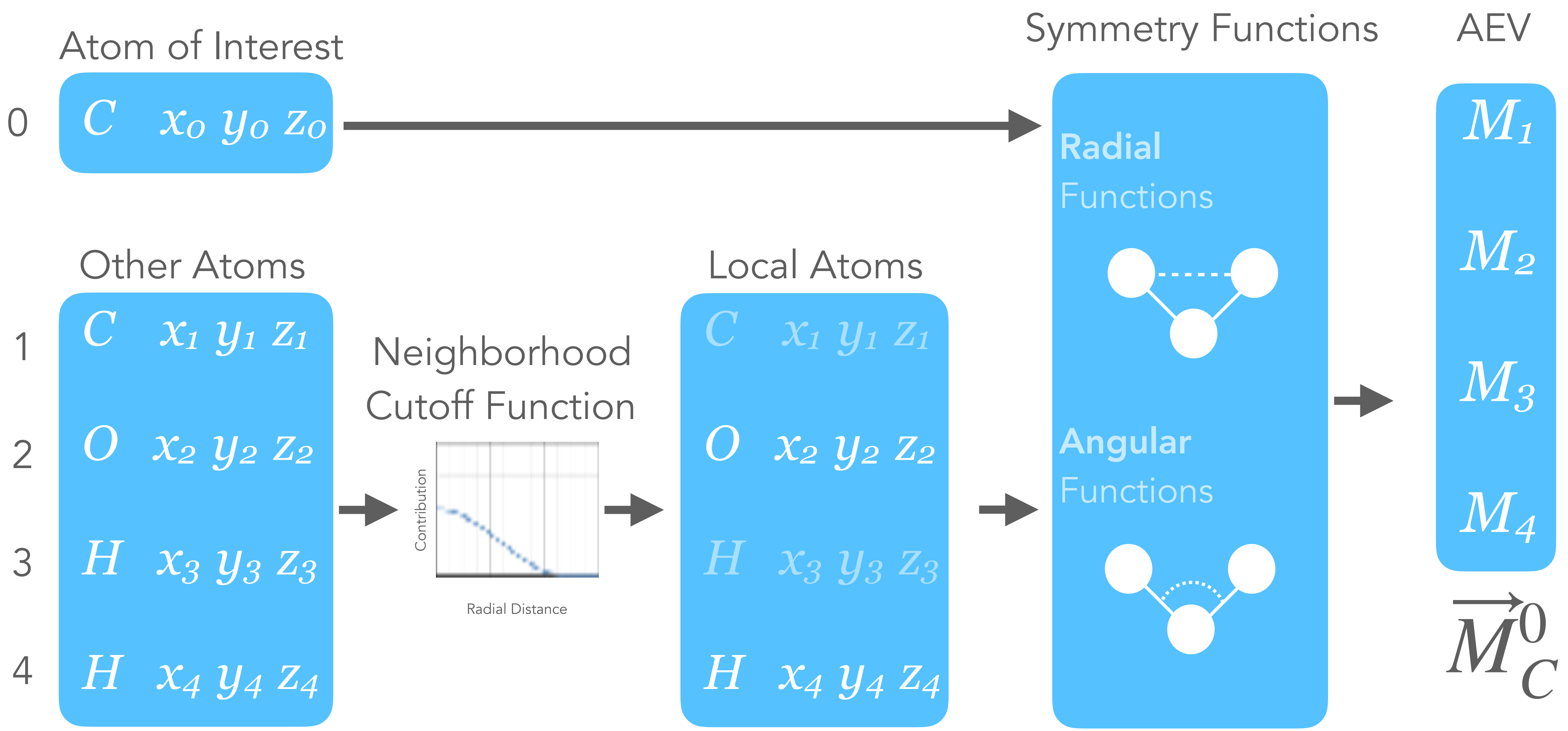
For each: Atom in the system

Build: One AEV factoring in coordinates and atomic number of nearby atoms

ANI-1 Neural Network Potential

Atomic Environment Vectors (AEV)

Computation of AEV



Atom of Interest

0 $C \quad x_0\ y_0\ z_0$

Symmetry Functions

Radial
Functions

Angular
Functions

AEV

M_1

M_2

M_3

M_4

$\vec{M}^{0}_{C}$

Other Atoms

1 $C \quad x_1\ y_1\ z_1$

2 $O \quad x_2\ y_2\ z_2$

3 $H \quad x_3\ y_3\ z_3$

4 $H \quad x_4\ y_4\ z_4$

Neighborhood
Cutoff Function

Local Atoms

$C \quad x_1\ y_1\ z_1$

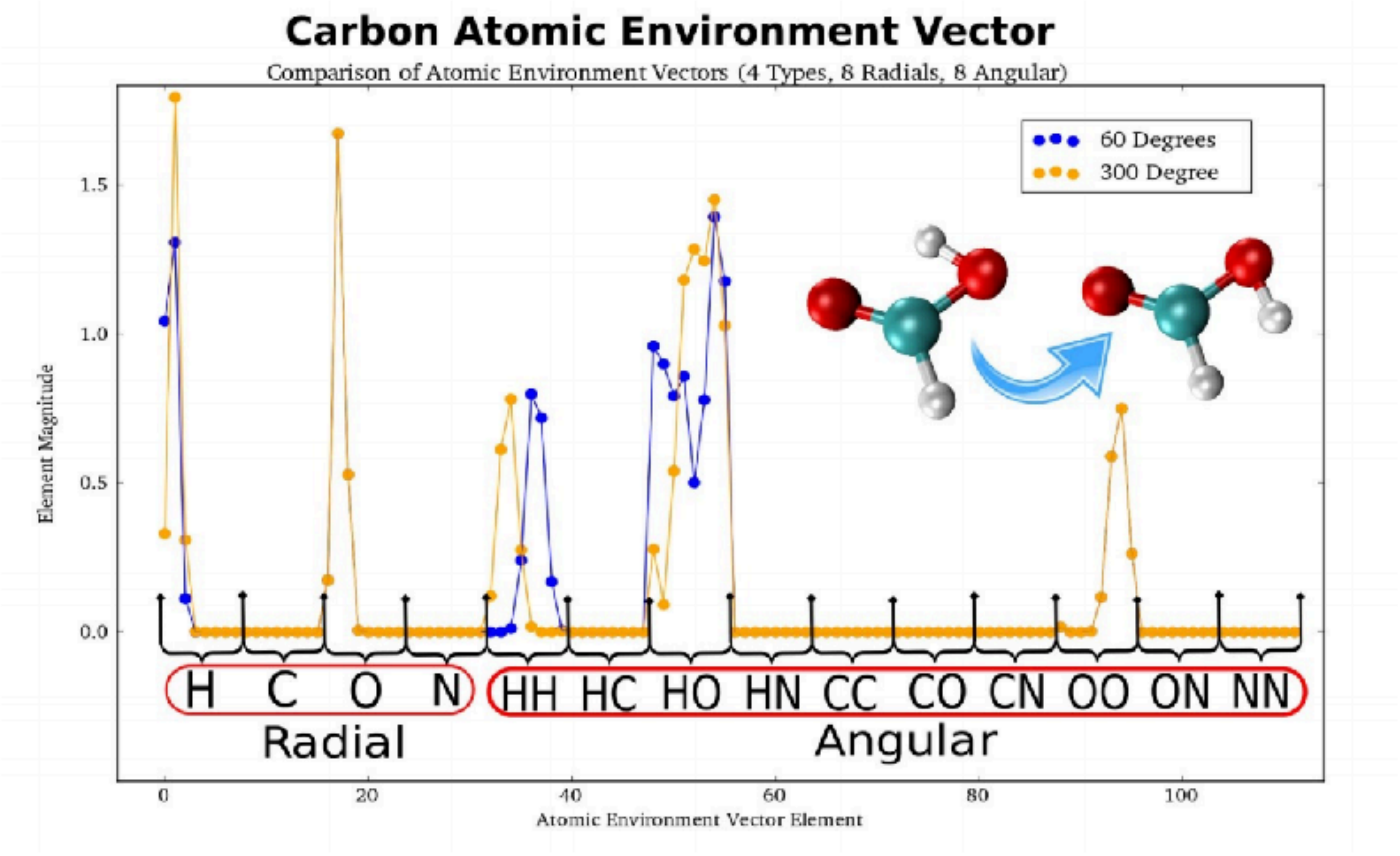
$O \quad x_2\ y_2\ z_2$

$H \quad x_3\ y_3\ z_3$

$H \quad x_4\ y_4\ z_4$

ANI-1 Neural Network Potential

Atomic Environment Vectors (AEV)



Source: J. S. Smith, O. Isayev, and A. E. Roitberg, “ANI-1: an extensible neural network potential with DFT accuracy at force field computational cost.”

ANI-1 Neural Network Potential

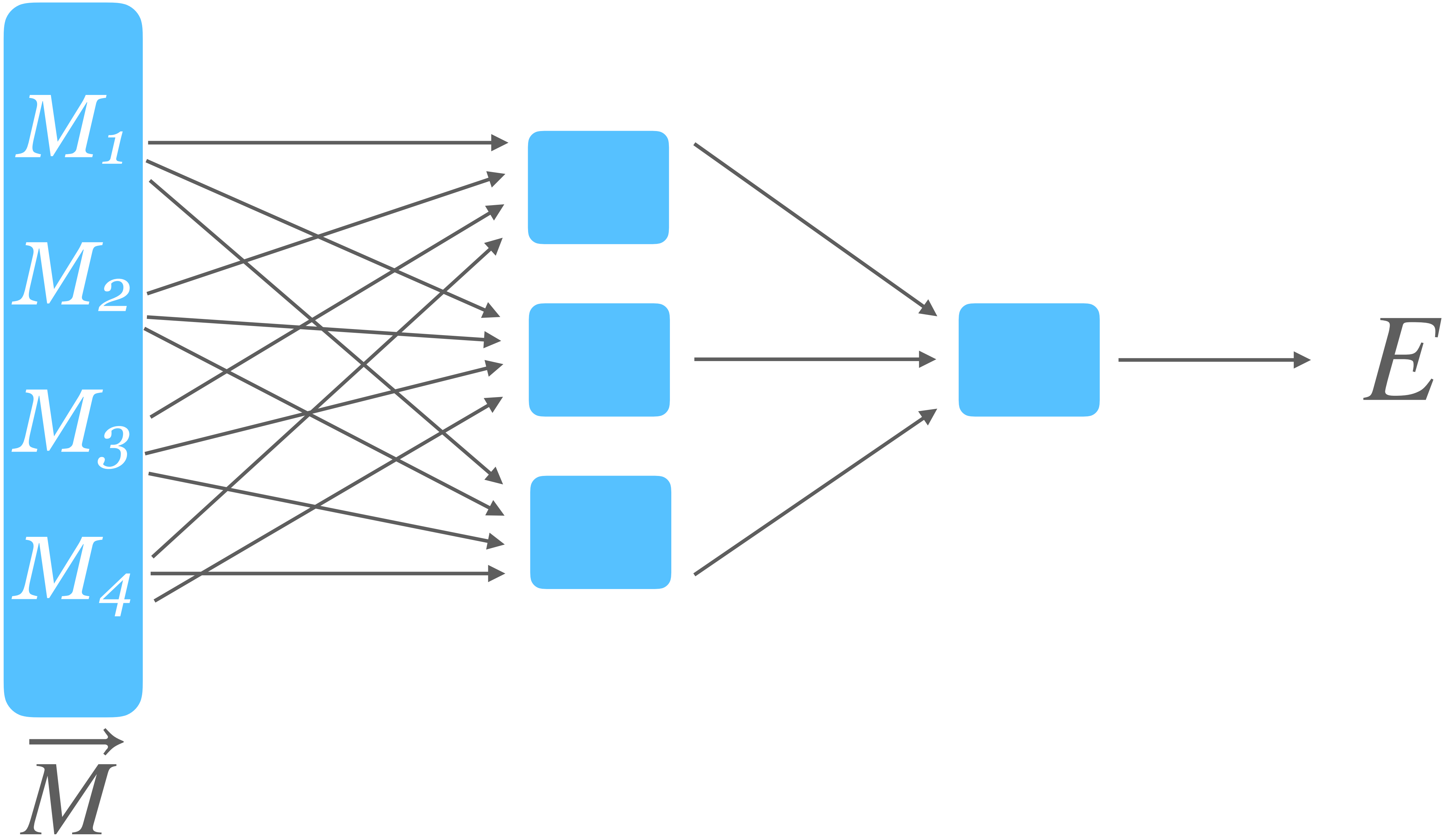
Decomposed Energy Function

Model total energy E as a sum of each atom's contribution E_i

$$E = \sum_{i=1}^{n} E_i$$

where n is the number of atoms in the molecular system

Naive Architecture



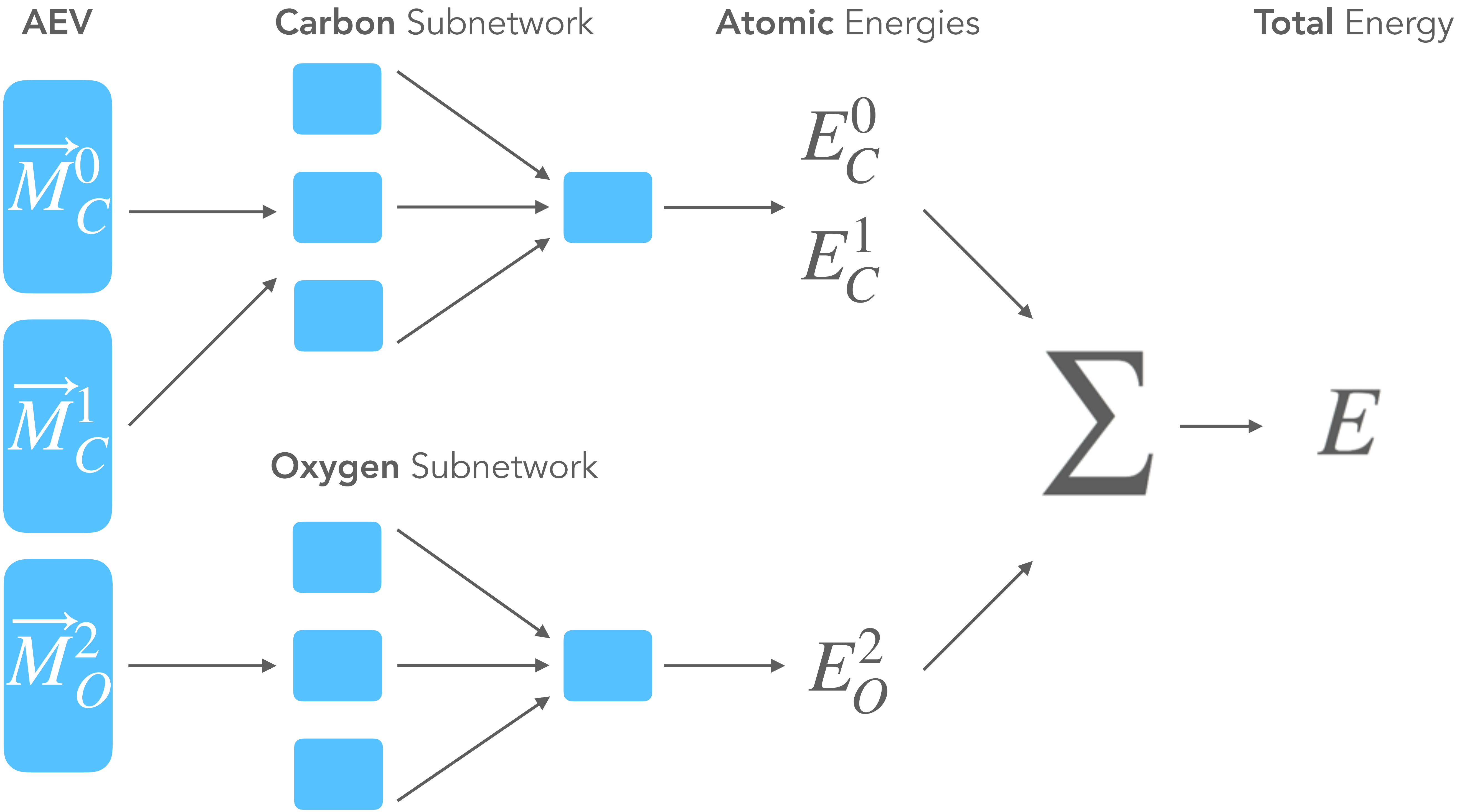
MOLECULAR REPRESENTATION

HIDDEN LAYER

OUTPUT LAYER

POTENTIAL ENERGY

ANI-1 Architecture



AEV

Carbon Subnetwork

Atomic Energies

Total Energy

$\vec{M}_C^0$

$\vec{M}_C^1$

$\vec{M}_O^2$

Oxygen Subnetwork

E_C^0

E_C^1

E_O^2

$\Sigma \rightarrow E$

ANI-1 Neural Network Potential

Training the Network

DATA

Use **GDB-8** database of all possible molecules containing up to **8** atoms of **H**, **C**, **N**, and **O** } ~58k Molecules

Generate likely conformations of each molecule by perturbing the molecule along its **normal modes** } ~17.2mil Conformations

Compute energy of conformation using **DFT** and label the example } ~17.2mil Labeled Examples

Atom Coordinates	DFT Energy
$(x_1\ y_1\ z_1) \ldots (x_n\ y_n\ z_n)$	E
$(x_1\ y_1\ z_1) \ldots (x_n\ y_n\ z_n)$	E
$(x_1\ y_1\ z_1) \ldots (x_n\ y_n\ z_n)$	E
$(x_1\ y_1\ z_1) \ldots (x_n\ y_n\ z_n)$	E
$(x_1\ y_1\ z_1) \ldots (x_n\ y_n\ z_n)$	E
$(x_1\ y_1\ z_1) \ldots (x_n\ y_n\ z_n)$	E
$(x_1\ y_1\ z_1) \ldots (x_n\ y_n\ z_n)$	E
$(x_1\ y_1\ z_1) \ldots (x_n\ y_n\ z_n)$	E

ANI-1 Neural Network Potential

Training the Network

COST FUNCTION

$$C(\overrightarrow{E}^{ANI}) = \exp\left(\sum_j \overrightarrow{E}_j^{ANI} - \overrightarrow{E}_j^{DFT}\right)$$

Find minimize via gradient descent with backpropagation

ANI-1 Neural Network Potential

Testing the Network

Test Set

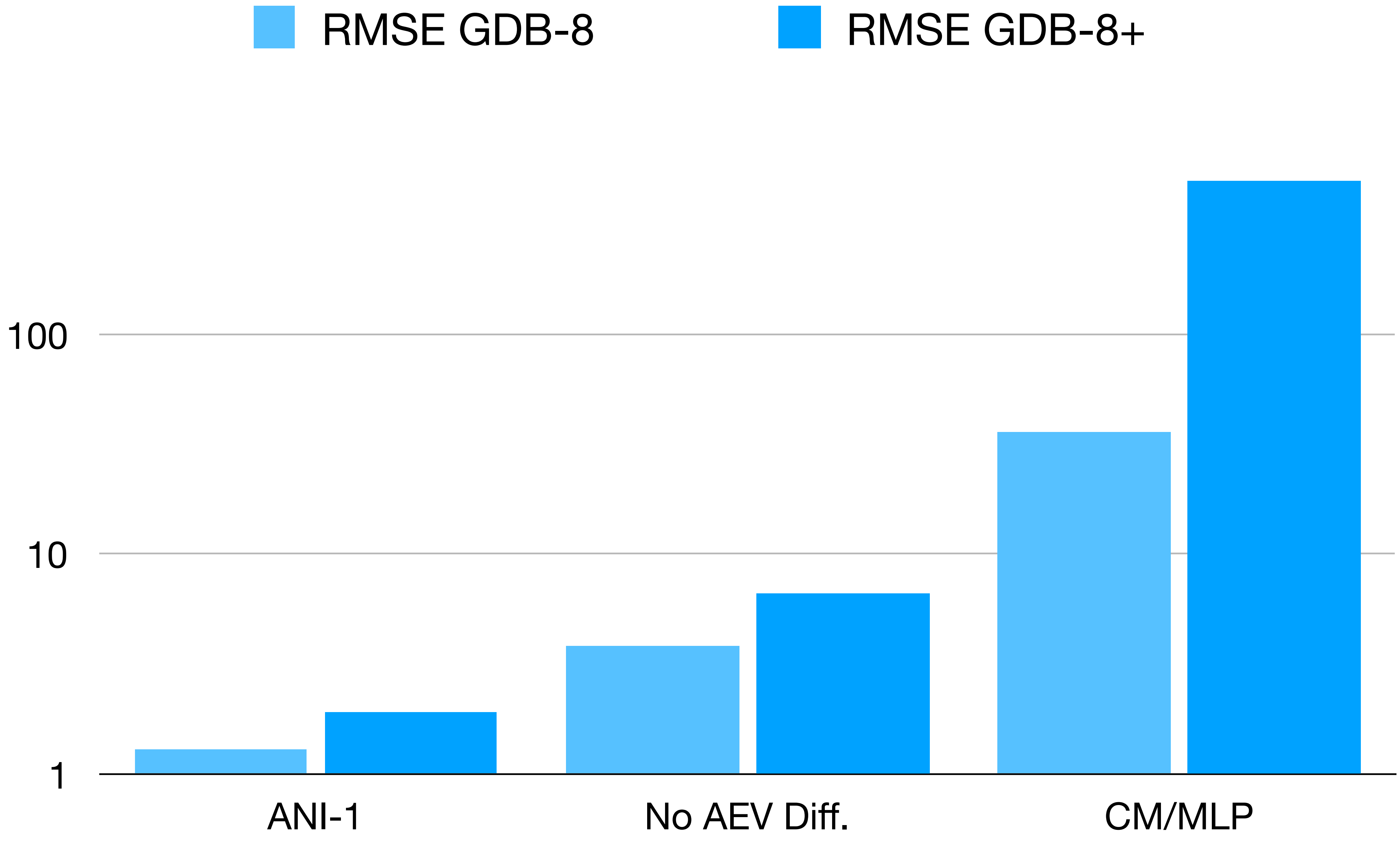
Molecules containing **more** than 8 atoms

Methods for Comparison

ab initio	*Semi-Empirical*	*NN Potential*
DFT	***DFTB***	*CM Representation*
	PM1	*No AEV Type Diff*
	AM1	

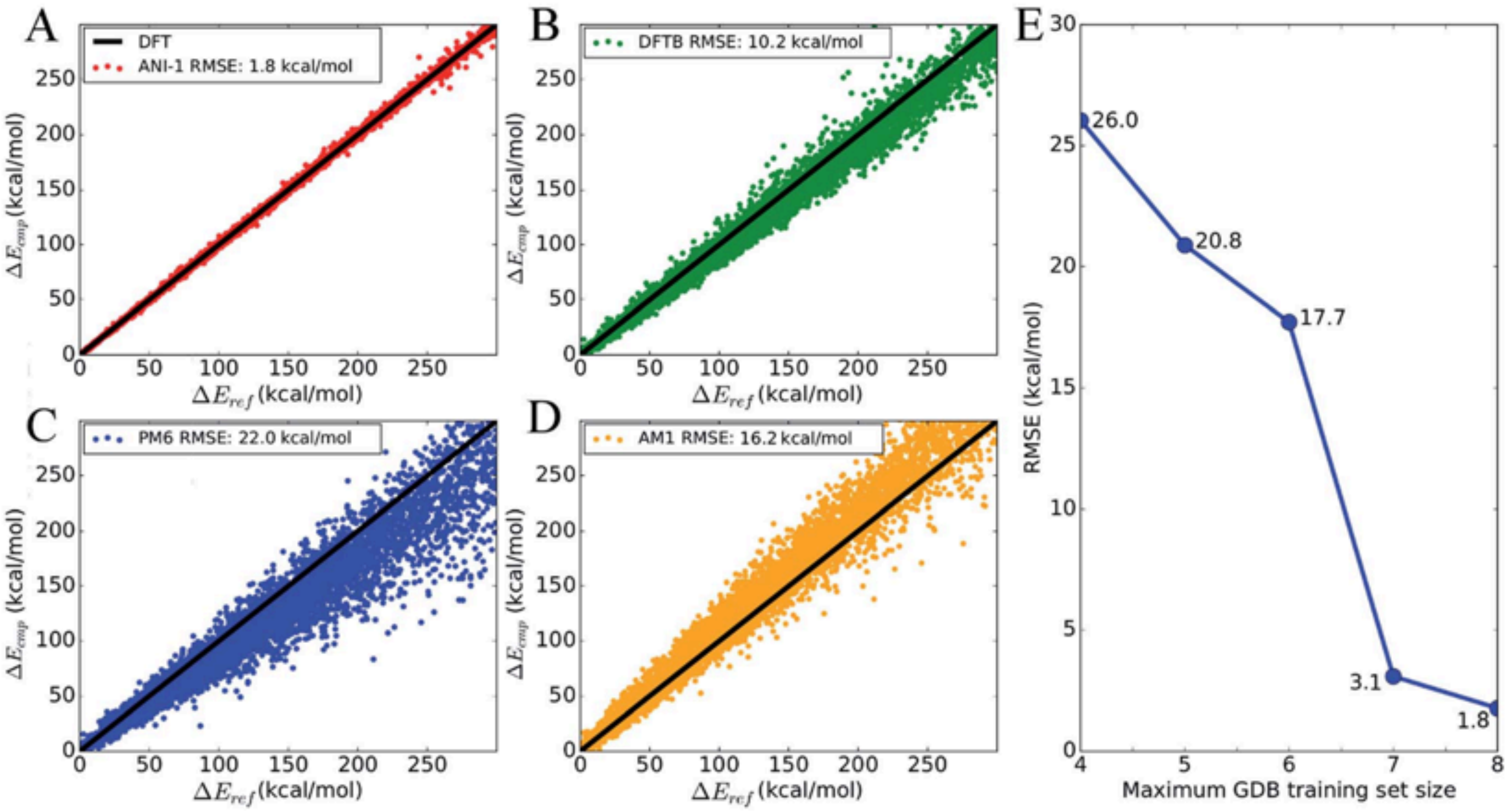
ANI-1 Neural Network Potential

Testing the Network



ANI-1 Neural Network Potential

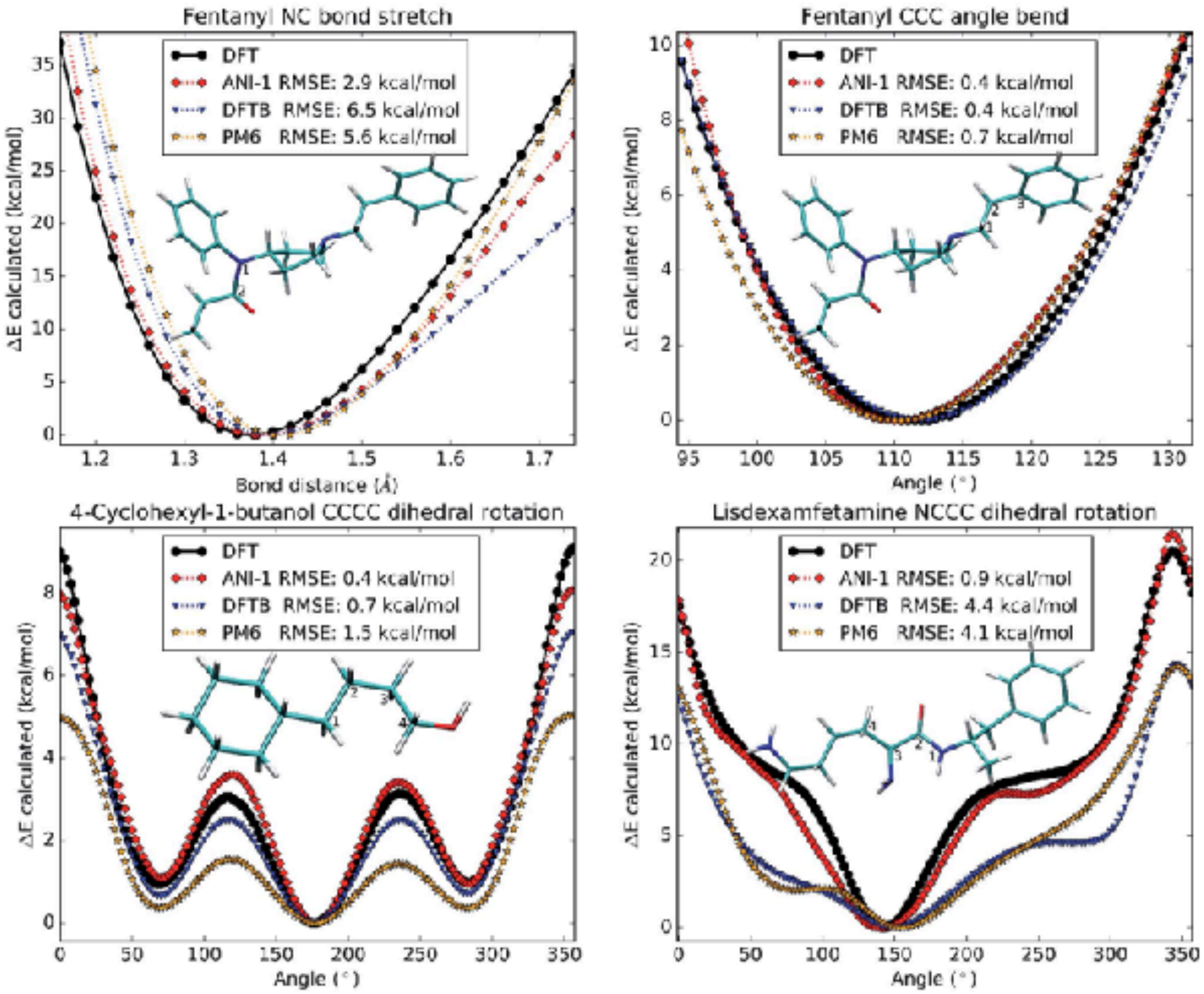
Testing the Network



Source: J. S. Smith, O. Isayev, and A. E. Roitberg, "ANI-1: an extensible neural network potential with DFT accuracy at force field computational cost."

ANI-1 Neural Network Potential

Testing the Network



Source: J. S. Smith, O. Isayev, and A. E. Roitberg, "ANI-1: an extensible neural network potential with DFT accuracy at force field computational cost."

ANI-1 Neural Network Potential

Strengths

Innovative **Architecture**

Highly **Transferable** (works on larger molecules)

Outperforms Baseline Neural Network Potentials

Models DFT Very Accurately

SPEED

ANI-1 Neural Network Potential

Limitations

Lacks Theoretical Justification of Atomic Decomposition?

Mimicking DFT, but…
DFT isn't ground truth

Little to no interpretability of learned function

Only works for C, H, N, O – Scale to more atoms?

Prediction Errors of Molecular Machine Learning Models Lower than Hybrid DFT Error

Lydia Hamburg
2/6/2018

~~Prediction Errors of Molecular Machine Learning Models Lower than Hybrid DFT Error~~

Faster method may be more accurate than traditional method

Lydia Hamburg
1/29/2018

Calculations of chemical properties are useful in chemistry and biology

Knowledge of electronic and thermodynamic properties enables:

- Prediction of chemical reactivity
- Identification of peaks in spectroscopy data
- Design of dyes and fluorophores
- Materials design
- Drug screening

Nearly all quantum chemistry calculations are approximations

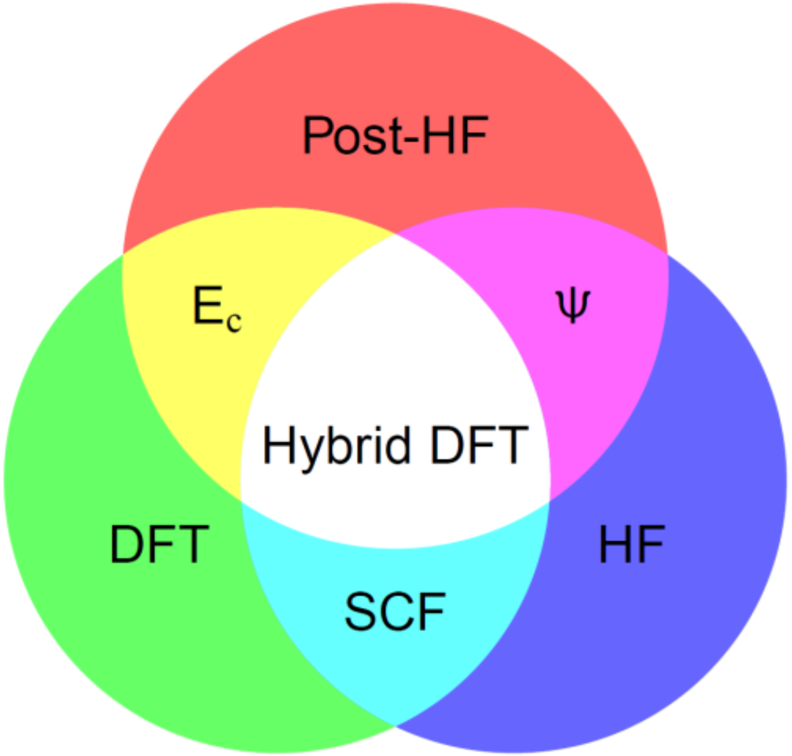
- Schrodinger’s wave equation can’t be solved analytically for more than two particles
- Density Functional Theory (DFT) approximates solution to Schrodinger’s Equation by simplifying the system
- The paper uses data from a hybrid DFT approach called B3LYP (Becke, 3-parameter, Lee-Yang-Parr)

Hybrid DFT (B3LYP) is fast but has flaws

- DFT makes assumptions that intentionally deviate from known quantum theory
- DFT calculations rely on functions that are fit to a limited set of experimental data
- Unable to predict when DFT will fail spectacularly

ML may be able to provide quick quantum chemistry estimates at a higher level of theory

- Density Functional Theory - O(~N^3)
- B3LYP, Hybrid DFT - O(~N^3)
- Hartree-Fock Theory - O(N^2)
- Coupled-cluster theory - O(N^6)
- Configuration interaction - O(>N^6)



“We investigated *all* combinations of regressors and representations…”

- No new ideas, but useful large scale benchmark
- Central source for results that might instead have been produced in multiple small slightly-different papers

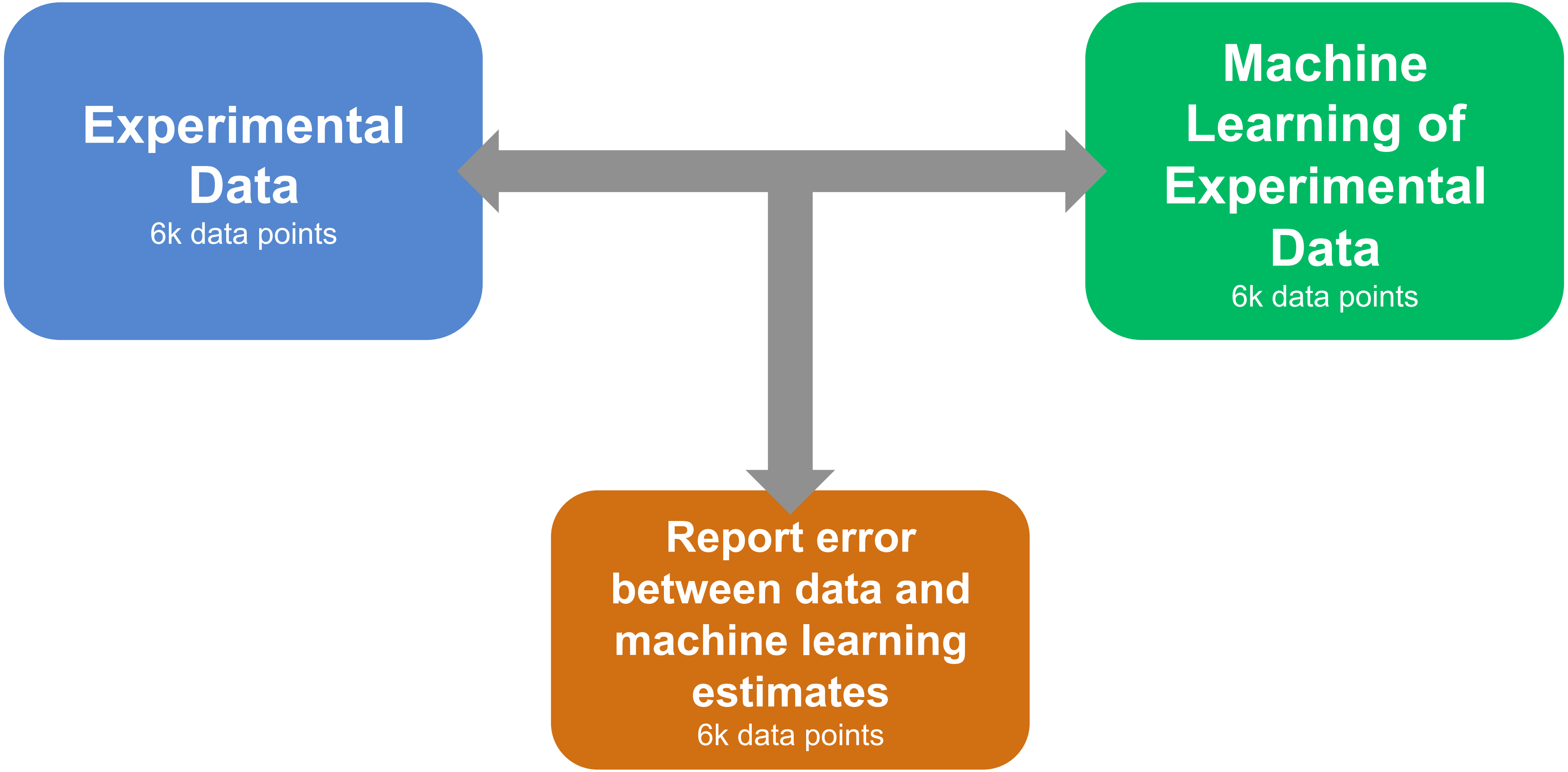
Molecular representations of dataset

- Coulomb matrix (CM)
- Bag of bonds (BOB)
- Molecular graphs (MG)
- Histograms of distances (HD)
- Histograms of dihedrals (HDAD)
- Bonds, angles, machine learning (BAML)
- Extended connectivity fingerprints (ECFP4)
- Molecular atomic radial angular distribution function (MARAD)

Machine learning regressors

- Bayesian ridge regression (BR)
- Elastic net (EN)
- Neural network (NN)
- Graph convolutions (GC)
- Gated graphs (GG)
- Random forest (RF)
- Kernel ridge regression (KRR)

Study was designed to allow the largest possible training set

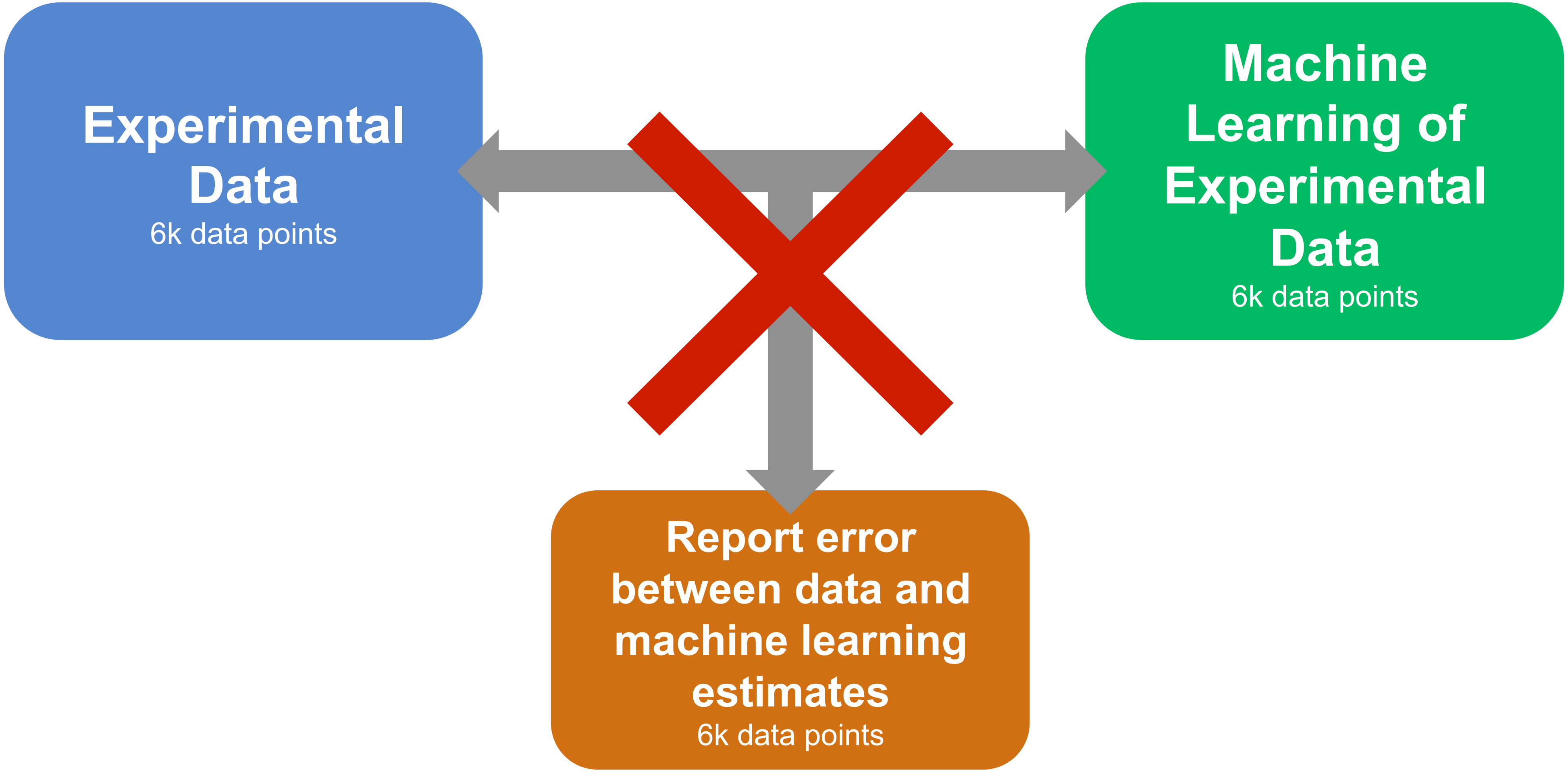


Experimental Data
6k data points

Machine Learning of Experimental Data
6k data points

Report error between data and machine learning estimates
6k data points

Study was designed to allow the largest possible training set



Experimental Data
6k data points

Machine Learning of Experimental Data
6k data points

Report error between data and machine learning estimates
6k data points

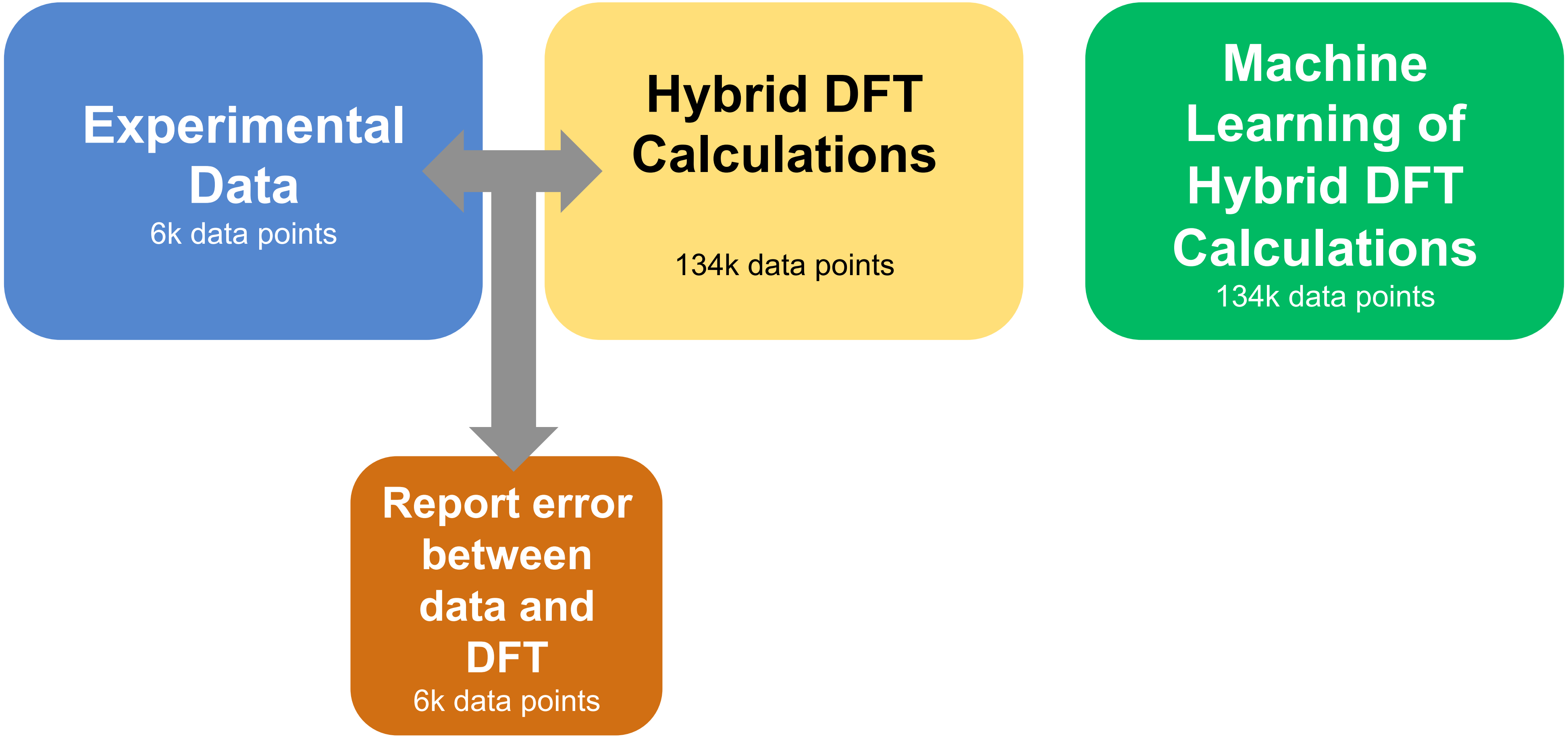
Study was designed to allow the largest possible training set

Experimental Data
6k data points

Hybrid DFT Calculations
134k data points

Machine Learning of Hybrid DFT Calculations
134k data points

Study was designed to allow the largest possible training set



Experimental Data
6k data points

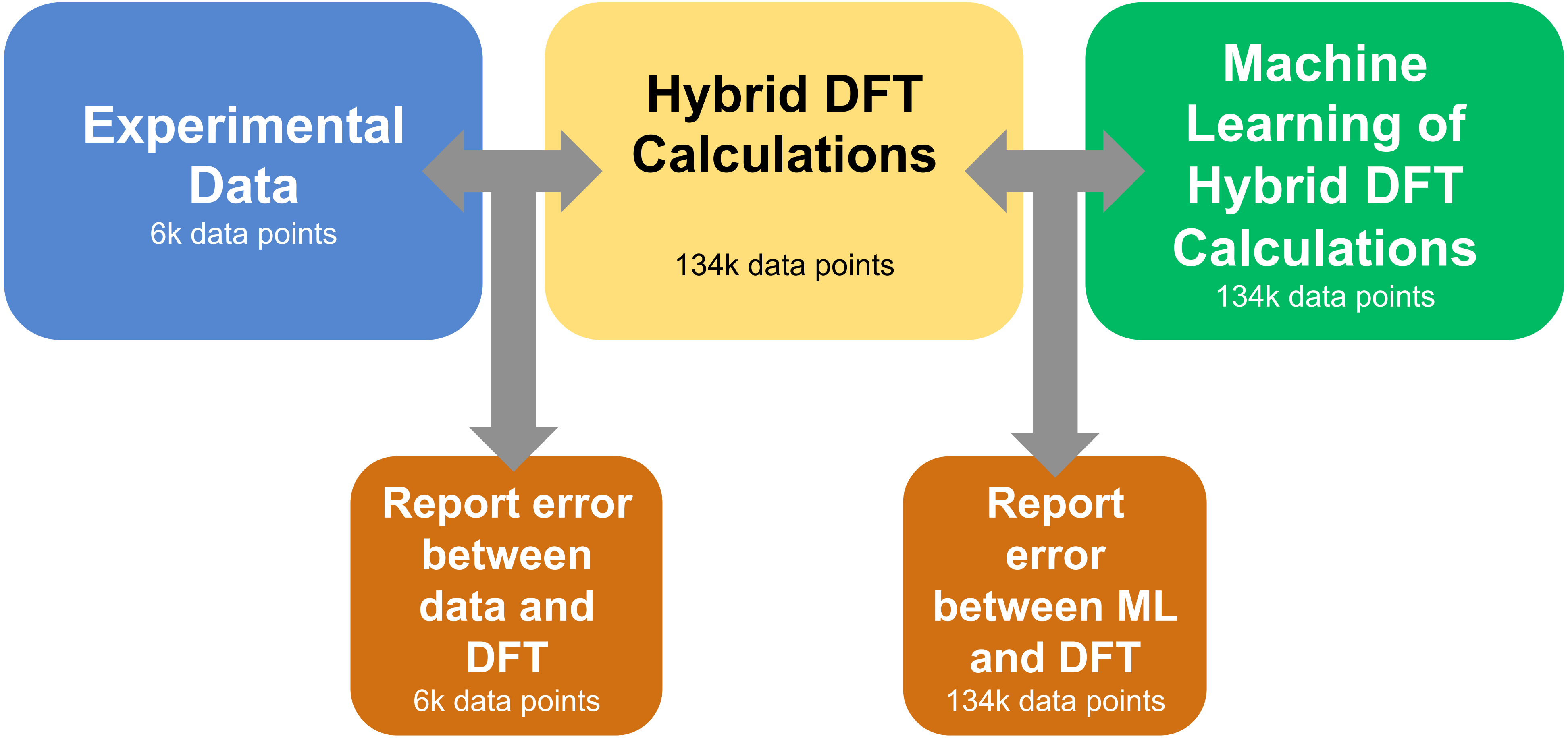
Hybrid DFT Calculations

134k data points

Machine Learning of Hybrid DFT Calculations
134k data points

Report error between data and DFT
6k data points

Study was designed to allow the largest possible training set



Experimental Data
6k data points

Hybrid DFT Calculations

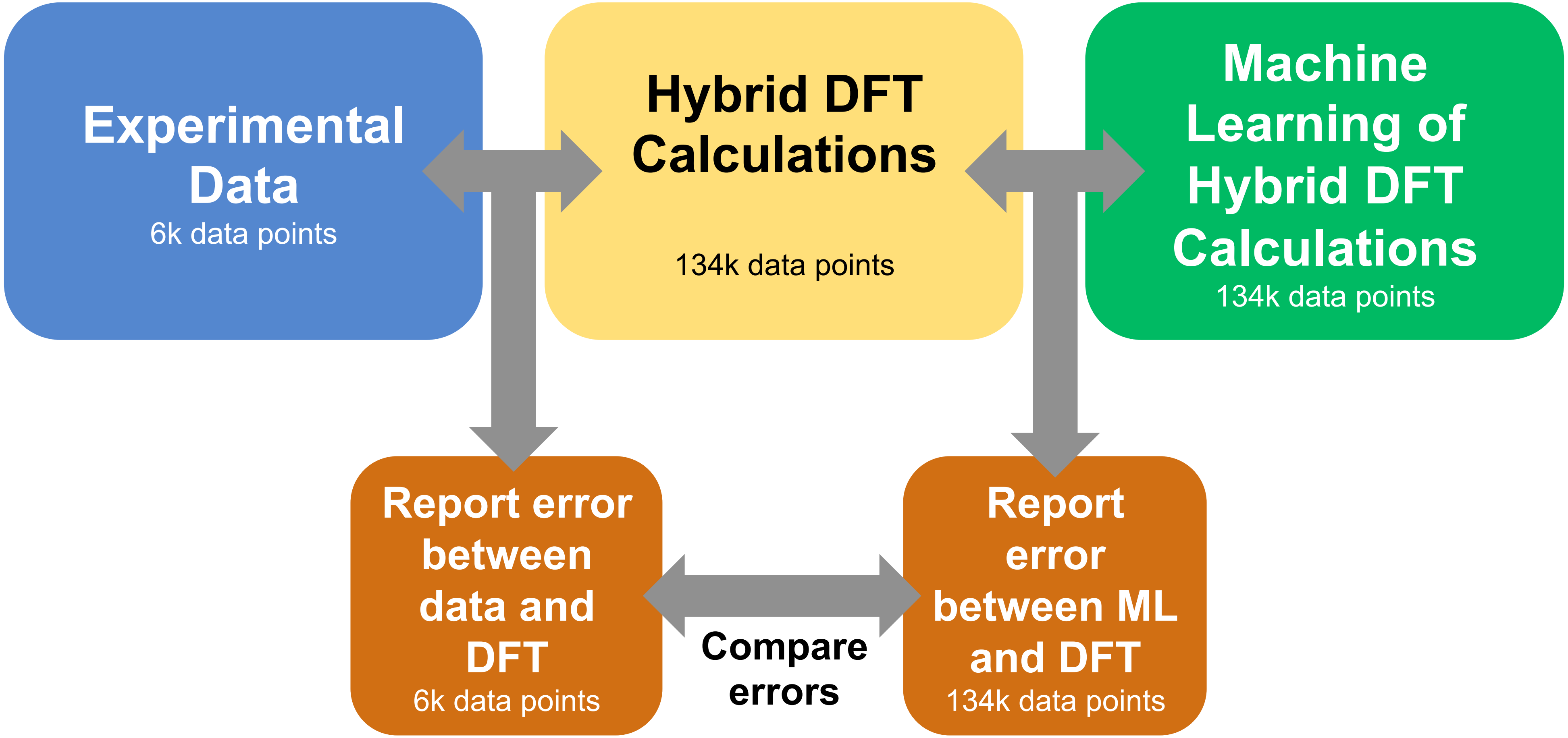
134k data points

Machine Learning of Hybrid DFT Calculations
134k data points

Report error between data and DFT
6k data points

Report error between ML and DFT
134k data points

Study was designed to allow the largest possible training set



Experimental Data
6k data points

Hybrid DFT Calculations

134k data points

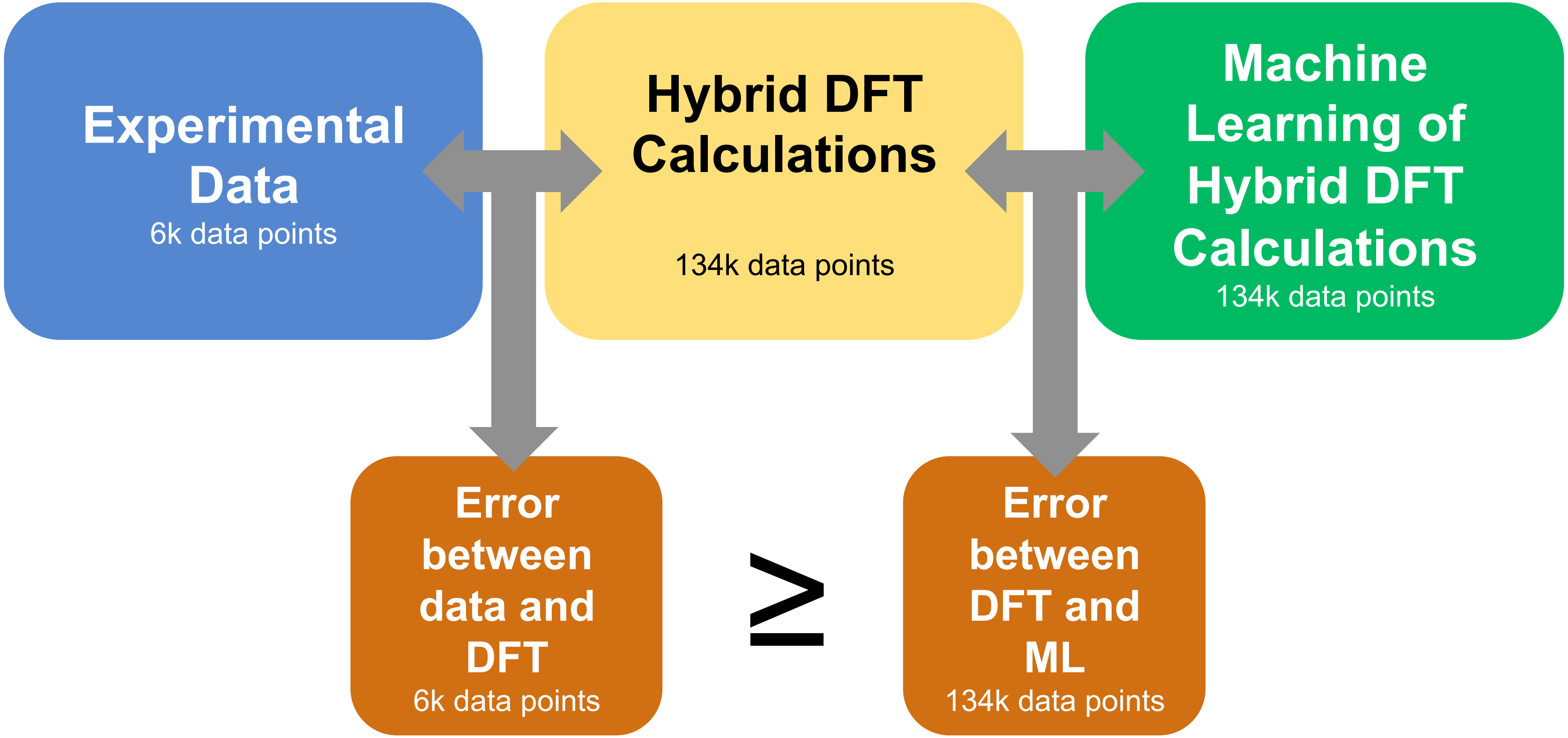
Machine Learning of Hybrid DFT Calculations
134k data points

Report error between data and DFT
6k data points

Compare errors

Report error between ML and DFT
134k data points

Comparison of errors



Experimental Data
6k data points

Hybrid DFT Calculations

134k data points

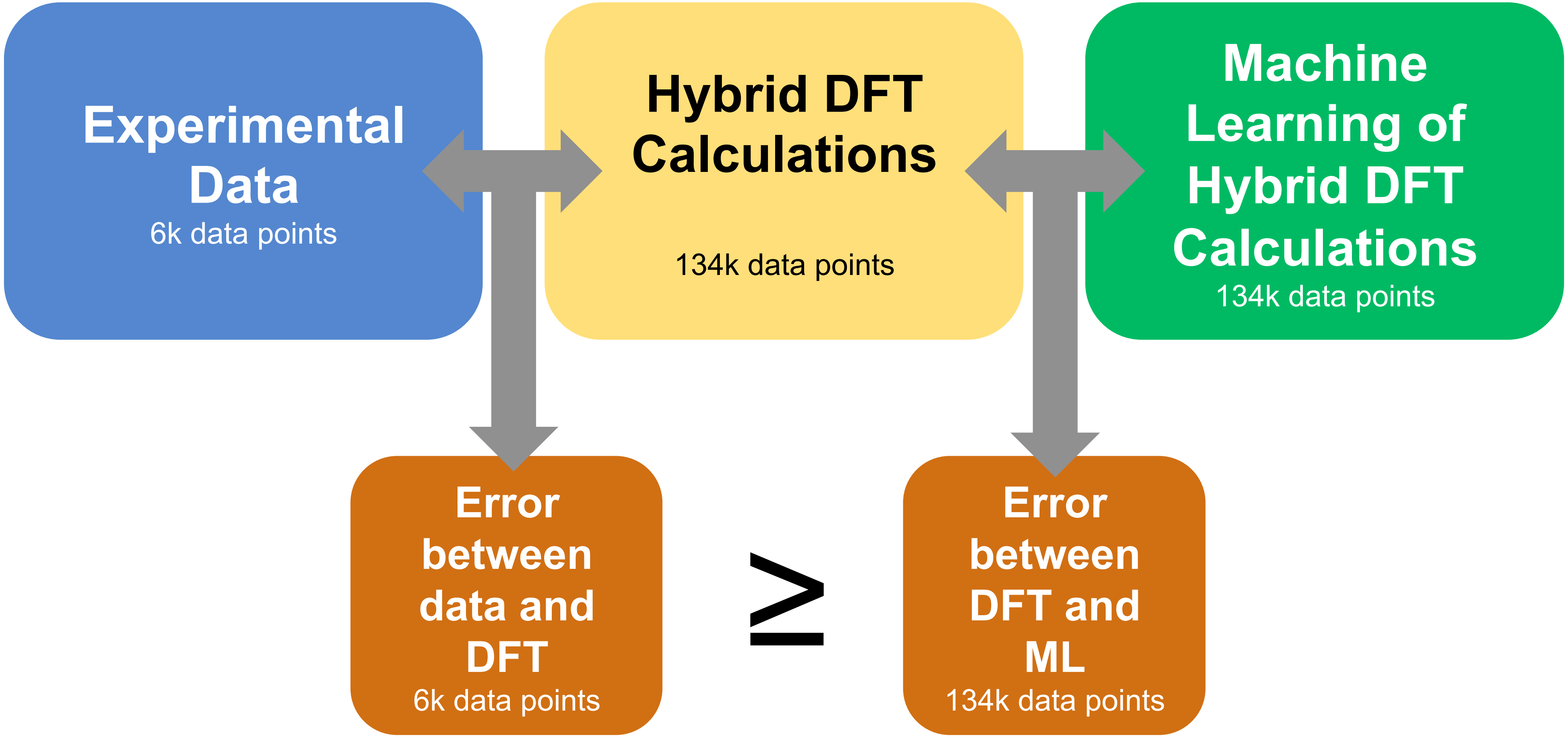
Machine Learning of Hybrid DFT Calculations
134k data points

Error between data and DFT
6k data points

≥

Error between DFT and ML
134k data points

“ML models could be more accurate than hybrid DFT if…data were available”



Experimental Data
6k data points

Hybrid DFT Calculations

134k data points

Machine Learning of Hybrid DFT Calculations
134k data points

Error between data and DFT
6k data points

≥

Error between DFT and ML
134k data points

Weaknesses

- Not a new concept (but the thoroughness is very satisfying)
- Generalizability unknown: explored 134K/10^{60} of chemical space
- Molecules types of interest unlikely to be well represented in training set
- Calculations of higher levels of theory might get faster
- The transitive nature of the conclusion slightly weakens confidence in the findings

Strengths

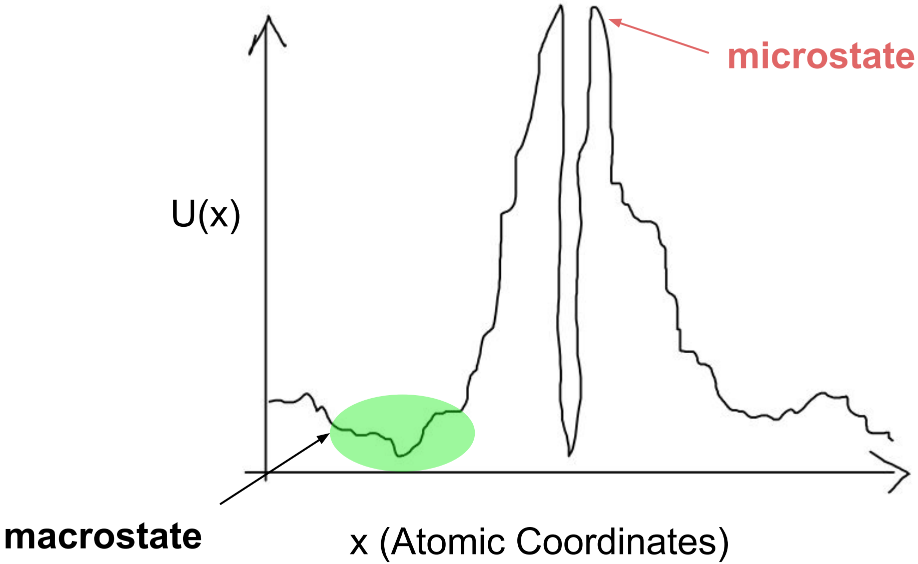
- Competently and thoroughly explored the space
- Made use of a huge percentage of all of the quantum chemistry data known to mankind
- Great example of a cross-disciplinary collaboration – chemists supplied the descriptions and interpretations, Google did the ML
- Straightforward about shortcomings

Simultaneous Optimization of Biomolecular Energy Functions on Features from Small Molecules and Macromolecules

Anvita Gupta
CS371

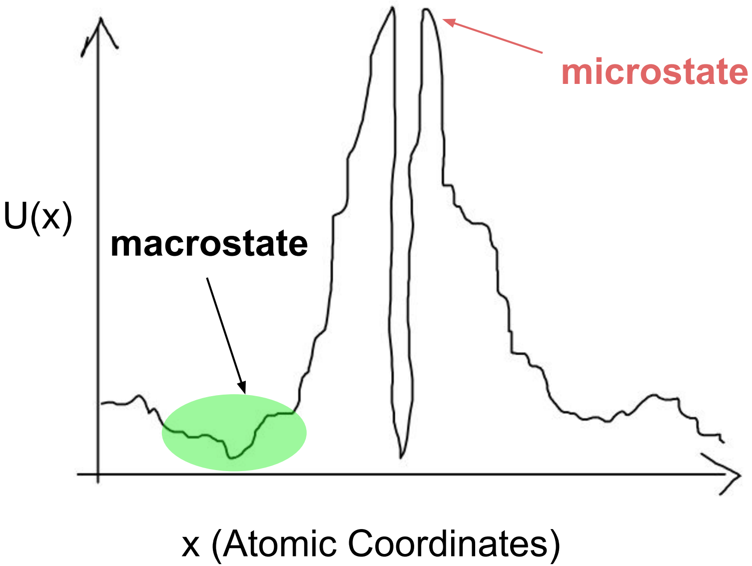
Free Energy vs Potential Energy

U(x)* ~** Potential energy based on exact coordinates of every atom in system ***(x)



Free Energy vs Potential Energy

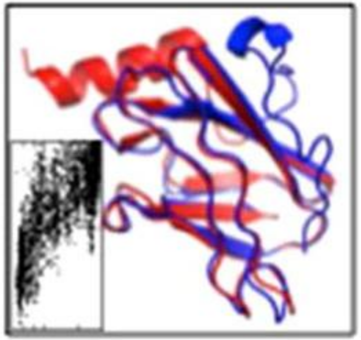
U(x) ~ Potential energy based on exact coordinates of every atom in system *(x)*



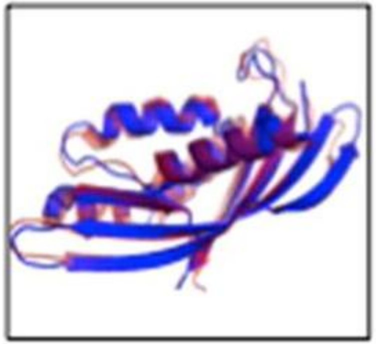
ΔG (Free Energy) gives a **penalty** to macrostates which are statistically unlikely

Goal of Authors: Predict **free energy** of macromolecular complex

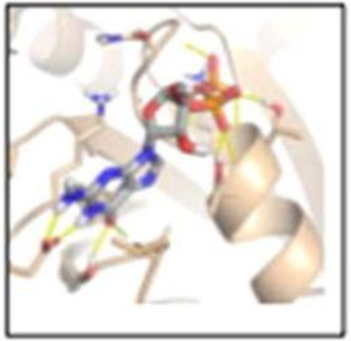
Develop free-energy function based on both **physics modelling** and **statistics** from **empirical data**



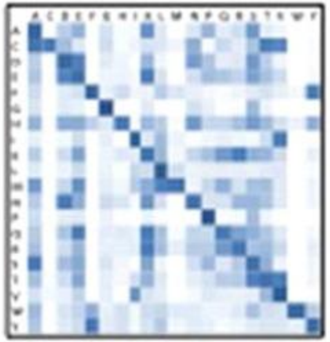
Decoy discrimination



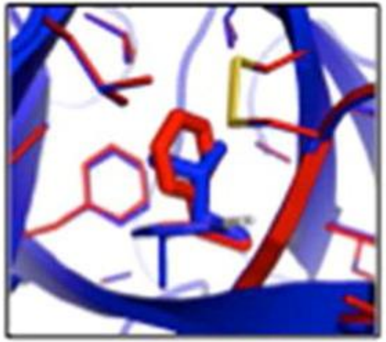
Homology modeling



Molecular docking



Sequence prediction



Mutational ΔΔG

Energy Function Development Pipeline

Modelling

Pick Terms for Energy Function - **Physics** and **Statistics**

Energy ~ w*[electrostatic] + w*[bond lengths] + w*[protein torsion from PDB] + ...

Training

Extract terms from training data

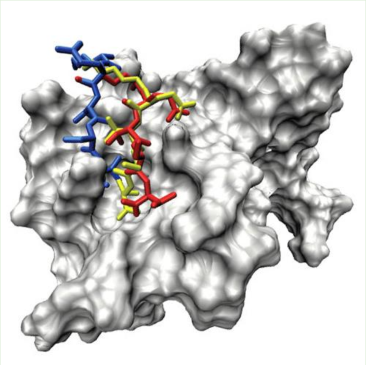
Feature Recovery Benchmarks

Atom Pair Distance = F[Energy(molecule)]

Optimize Weights for Energy Function

Evaluation

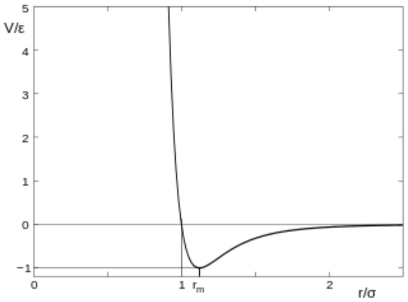
Scientific Benchmarks



Docking Scores, etc.

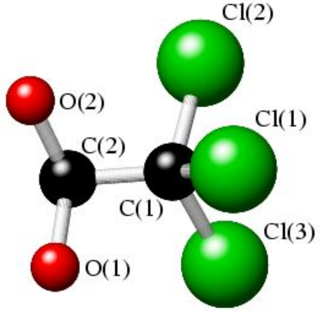
Simplex Optimization

Modelling Energy Function (~100 parameters)



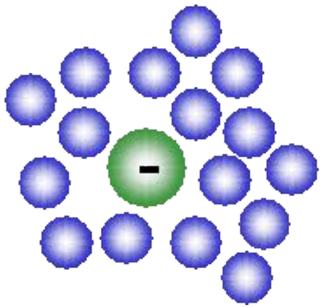
Nonbonding Interactions

- Lennard Jones Potential
    - **Improved!**
- Coulomb's Law (Electrostatic)
- Van Der Waal's
- Hydrogen Bonding



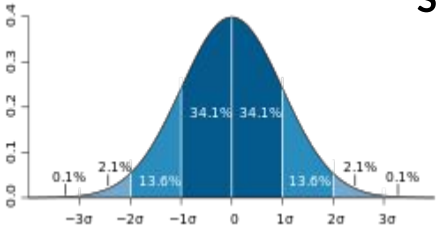
Bonding Terms

- Bond Torsion
    - **Improved!**
- Bond lengths



Solvation Energy

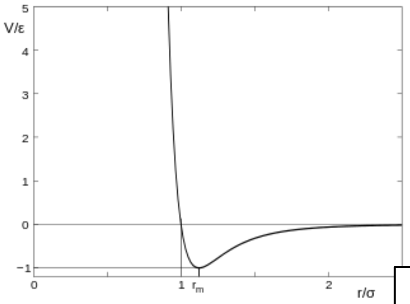
- Anisotropic (Asymmetric) Solvation Model
    - **Improved!**



Statistical Terms

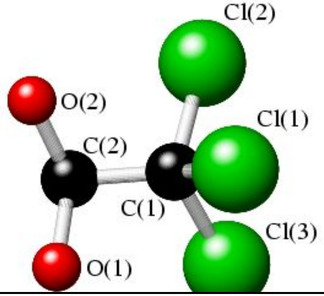
- Small Molecule and Macromolecular Data
- -log(Prob) ~ Energy

Modelling Energy Function (~100 parameters)



Nonbonding Interactions

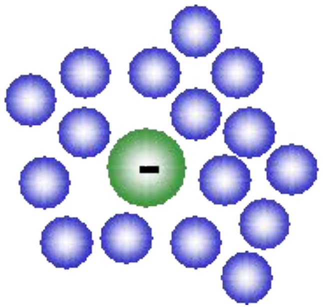
- Lennard Jones Potential
    - **Improved!**
- Coulomb's Law (Electrostatic)
- Van Der Waal's



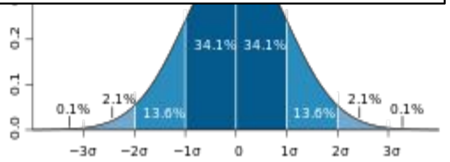
Bonding Terms

- Bond Torsion
    - **Improved!**
- Bond lengths

$$
\begin{aligned}
E_{\text{total}} = {} & E_{\text{LJ\_atr}} + W_{\text{LJ\_rep}} E_{\text{LJ\_rep}} + E_{\text{Coulomb}} + E_{\text{Hbond}} \\
& + E_{\text{solv\_iso}} + E_{\text{solv\_aniso}} + W_{\text{dun}} E_{\text{dun}} + W_{\text{rama}} E_{\text{rama}} \\
& + W_{\text{p\_aa\_pp}} E_{\text{p\_aa\_pp}} + W_{\text{bonded}} E_{\text{bonded}} + E_{\text{ref}} \qquad (5)
\end{aligned}
$$



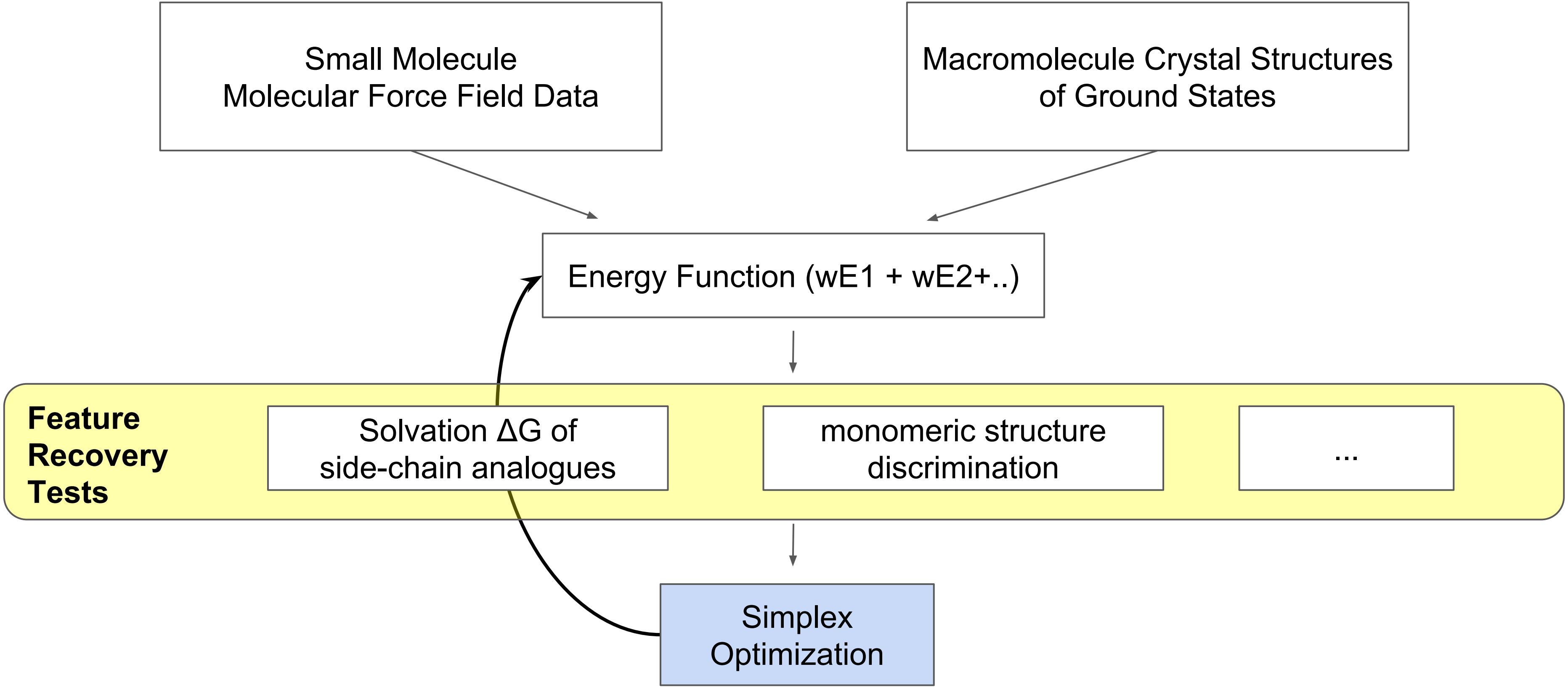
- Anisotropic (Asymmetric) Solvation Model
    - **Improved!**



tatistical Energy Terms

- Small Molecule and Macromolecular Data
- -log(Prob)

Training Energy Function



Small Molecule
Molecular Force Field Data

Macromolecule Crystal Structures
of Ground States

Energy Function (wE1 + wE2+..)

Feature Recovery Tests

Solvation ΔG of
side-chain analogues

monomeric structure
discrimination

...

Simplex
Optimization

Evaluating Energy Function (Results)

- Divided into test set and training set
- **Decoy Detection**
  - First allow structures to move (relax) in current energy function)
- **Improvement of:**
  - 20.8% (36.3% to 57.1%) on training set
  - 14.1% (53.1% to 67.2%) on test set
- **Homology Modeling**
  - Small but consistent improvements when using this energy function with Rosetta

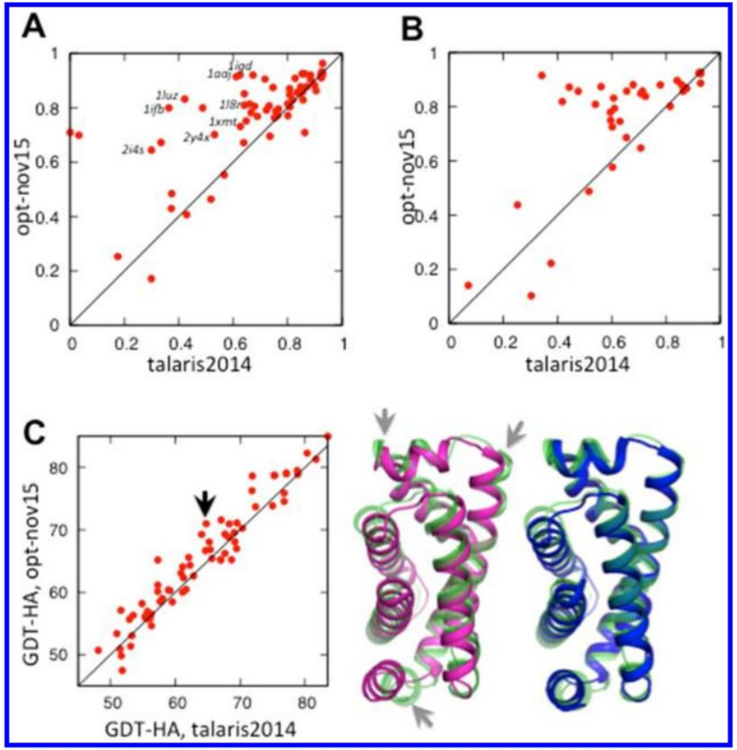
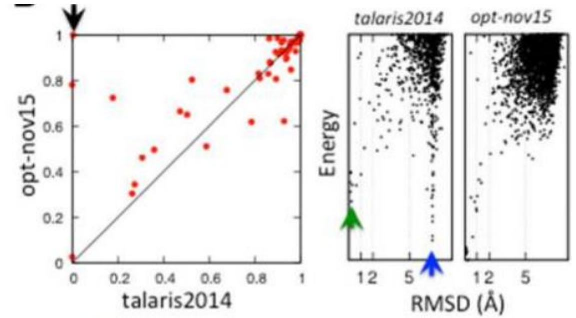
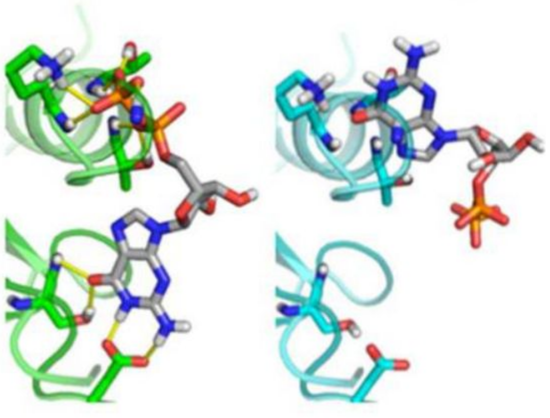


Figure 2. Improvements in monomeric structure prediction from independent tests.

Results: Docking Studies

- Improvements in both **protein-protein** and **protein-ligand** docking
- Protein-ligand not used in optimization!
- Demonstrates success in balancing:
  - Nonbonded interaction terms
  - Solvation energy
- Key successes of new function:
  - Correct protein-protein docked pose with smaller buried surface area but more favorable interactions
  - Correct protein-ligand pose with greater Solvation energy but more interactions

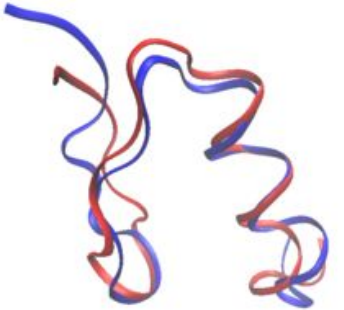




Left: Correct structure found by *optnov15*, Right: non-native structure selected by *talaris*

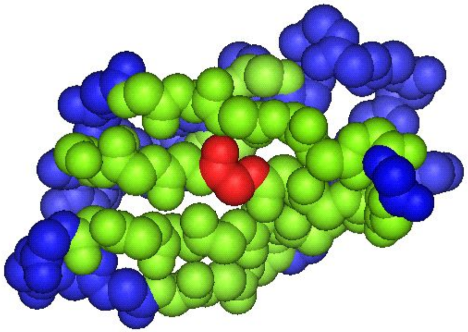
Results: Various Other Tasks

Protein Design



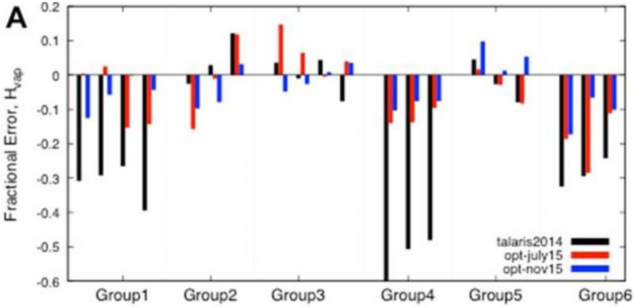
- Small improvements (on the order of 1 percentage point)
- Better balanced preferences for different amino acids

Free Energy Change from Mutations



- R^2 between predicted and experimental ΔΔG improves by 4% to 0.743
- <1% improvement in classification accuracy for stabilizing mutations

Small Molecule Thermo Data



- To be expected.
- Improved estimates of heat of vaporization
- Original function not enough weight on nonbonded interaction strength

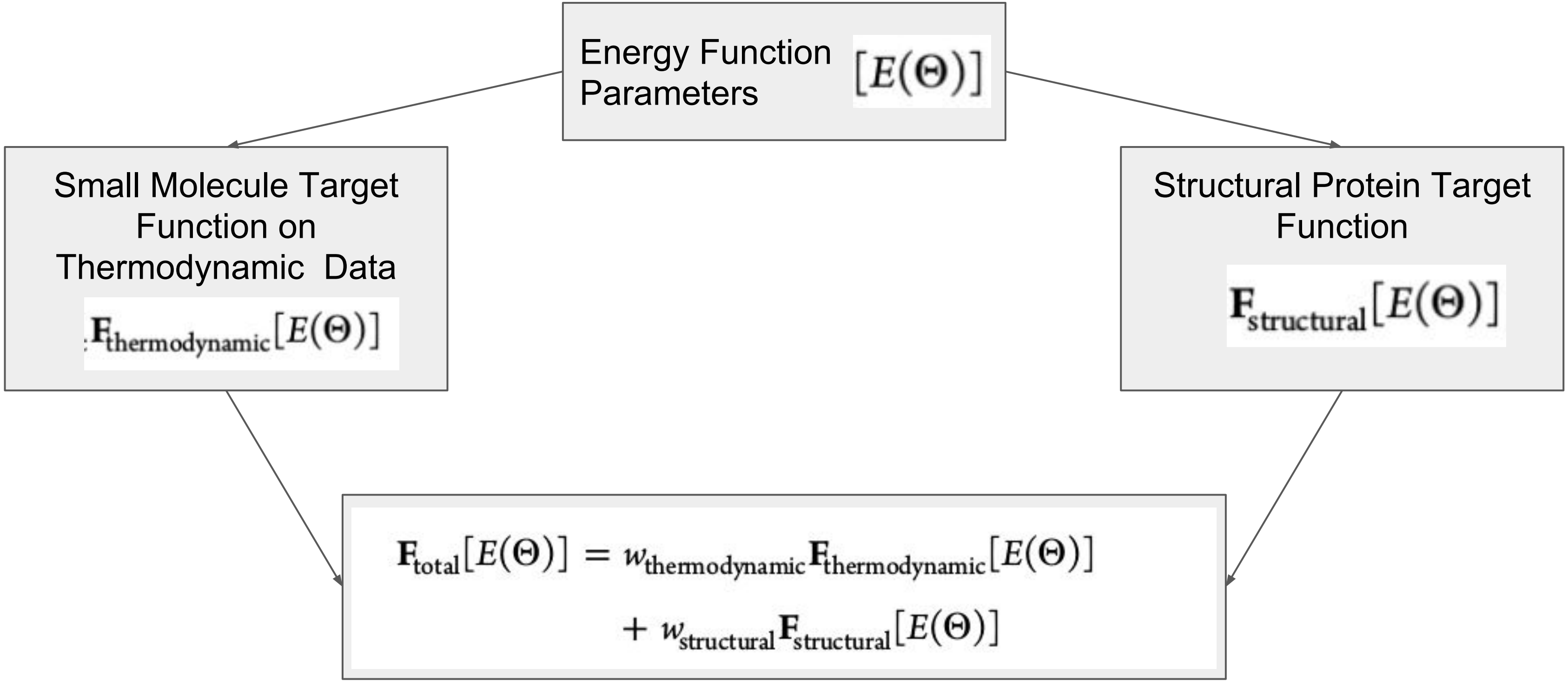
Key Takeaways

1. Integrated both small molecule force field data and macromolecular structural data to improve energy function
2. Thoroughly evaluated energy function
   a. Results from almost every task Energy Function could be used for
   b. Good benchmarking of all computational software
   c. Tested dualOptE (simplex optimization) on existing energy function to make sure it was performing correctly
3. Good interpretation of cases when new energy function improved upon older energy function

Limitations

1. No standard evaluation on several tasks for benchmarking
2. Conflation of improved energy terms and increased training data
3. “Black magic”
    a. Setting weights for target functions
    b. Statistics/Sources of Training Data
4. Order of 100 parameters
    a. Authors call this a “high dimensional subspace” when nowadays this is not true
5. Important to see: more careful analysis of cases when initial energy function (*talaris)* performed better than new energy function (*opt-nov-15)*
6. In general: too much jargon makes paper unclear, generally disorganized.

Methods: Overview of Approach



Energy Function Parameters $[E(\Theta)]$

Small Molecule Target Function on Thermodynamic Data

$\mathbf{F}_{\text{thermodynamic}}[E(\Theta)]$

Structural Protein Target Function

$\mathbf{F}_{\text{structural}}[E(\Theta)]$

$$\mathbf{F}_{\text{total}}[E(\Theta)] = w_{\text{thermodynamic}}\mathbf{F}_{\text{thermodynamic}}[E(\Theta)] + w_{\text{structural}}\mathbf{F}_{\text{structural}}[E(\Theta)]$$