DMSE DEPARTMENT OF MATERIALS SCIENCE & ENGINEERING

Massachusetts Institute of Technology

Inverse design of materials with representation learning and physics-based priors

Coarse-graining atomistic simulations with deep autoencoders.

ML4MS Workshop, Aalto University, Finland
May 9 2019

Rafael Gómez-Bombarelli
rafagb@mit.edu
http://gomezbombarelli.mit.edu/

Virtual discovery

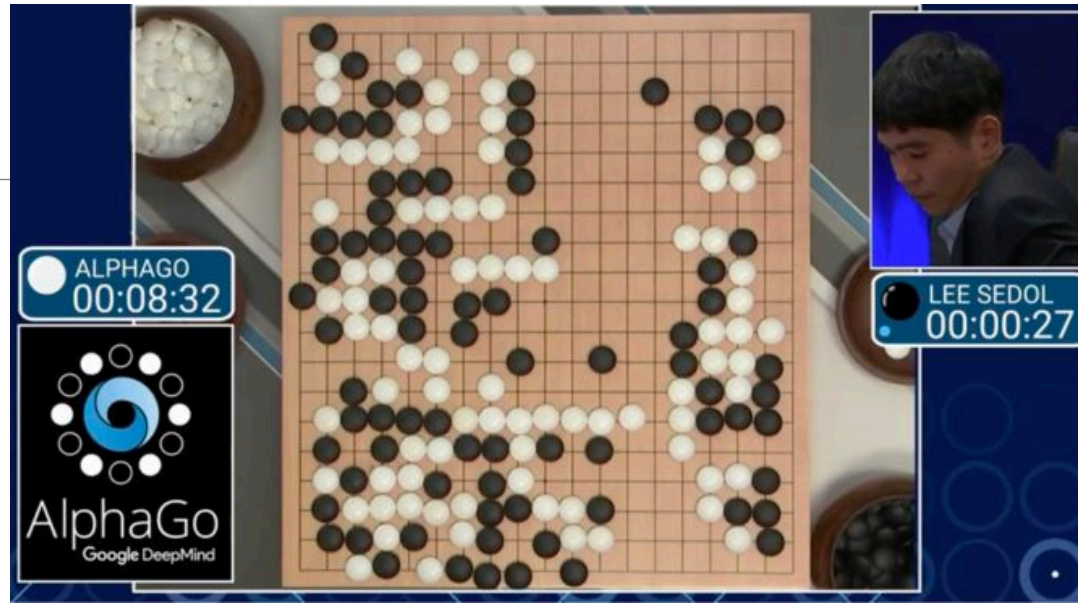
Software is already matching (or beating) humans in performance and speed

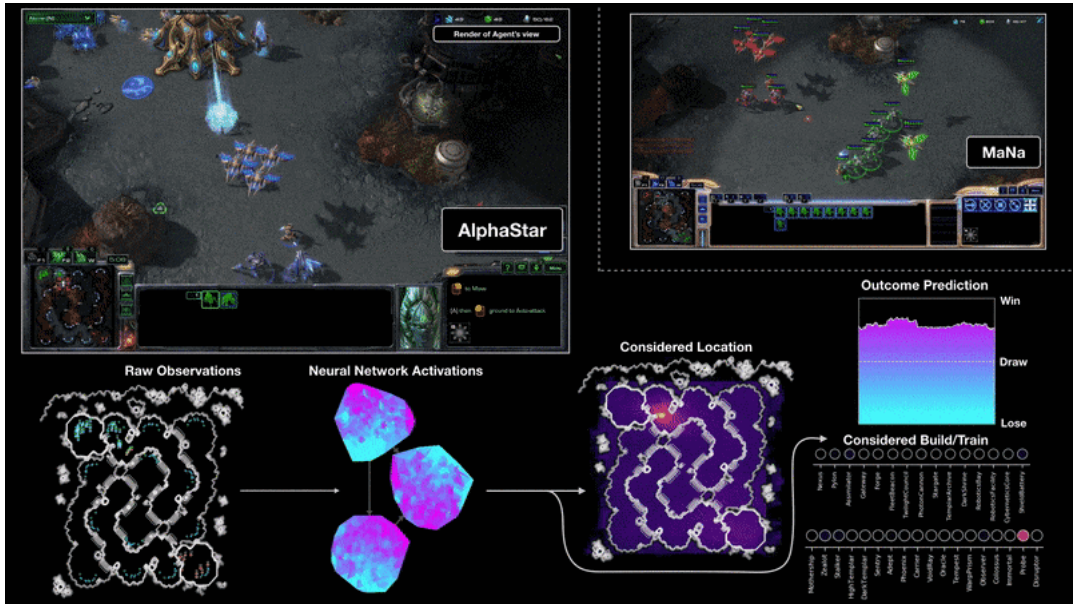
Driverless cars, AlphaGo, virtual assistants, speech & image recognition and generation, …

Increasing computing power

Faster algorithms

More data





Virtual discovery

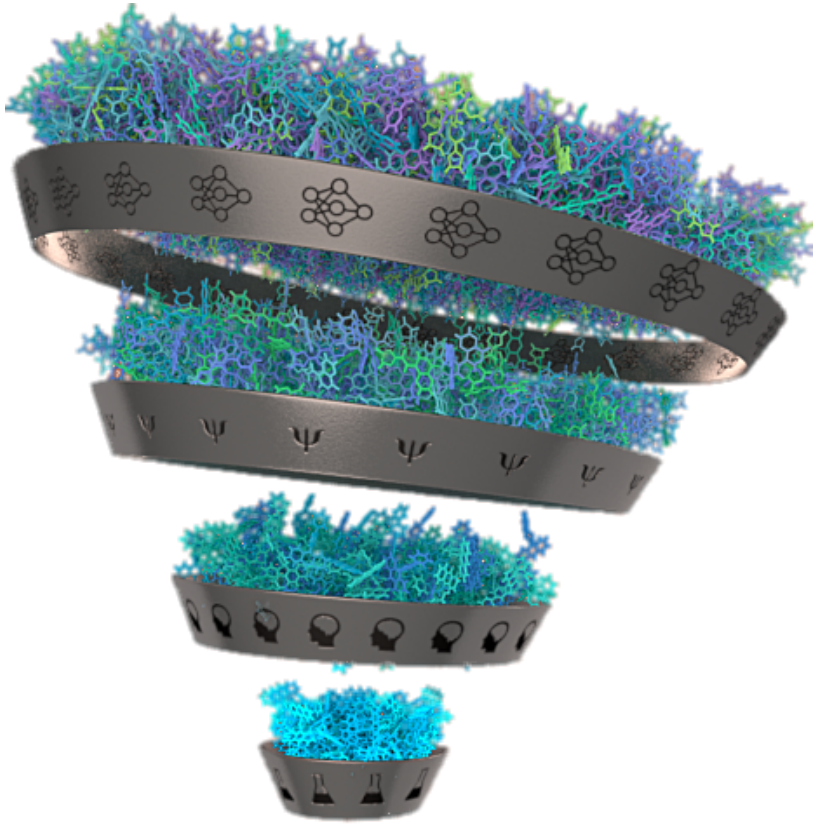
Software is already matching (or beating) humans in performance and speed

Driverless cars, AlphaGo, virtual assistants, speech & image recognition and generation, …

Increasing computing power

Faster algorithms

More data



What works in materials design?

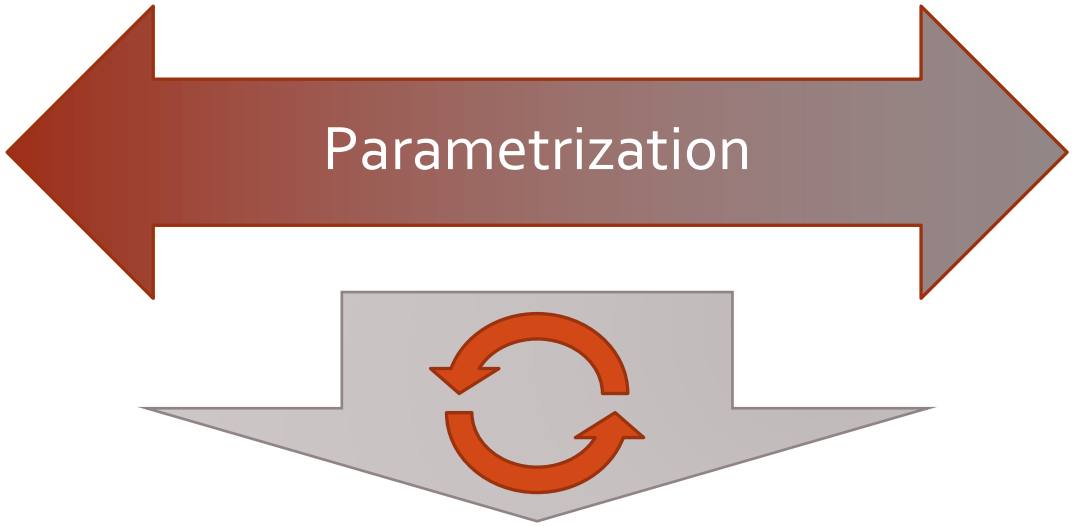
Computational spectrum - virtuous cycle

There is essentially a continuum of higher parametrization and statistical learning connecting first principles (theory-based simulations) to black-box statistical learning over experiments.

First Principles

Out-of-the-box
No/Little fitting
Extrapolates
High-throughput
Sometimes cheap

- but -
As good as model
Sometimes costly



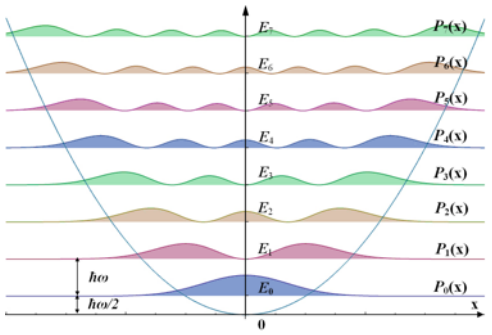
Machine learning

Fast
Uncanny performance
Leverage large data

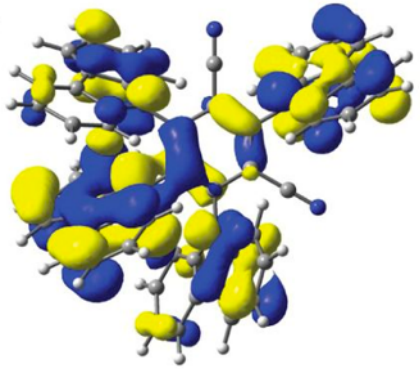
- but -
As good as training data
Expensive / proprietary data

Predictive Simulation

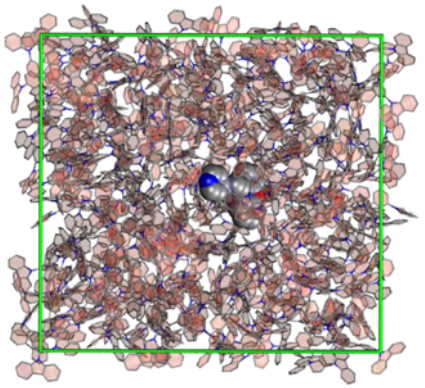
Parametrization from experiments



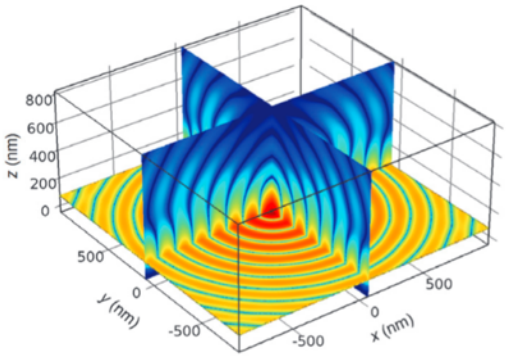
Toy models



Quantum / electronic structure



Molecular dynamics



Continuum models

Experimental observation
Design parameter

Machine Learning for Materials

Training data

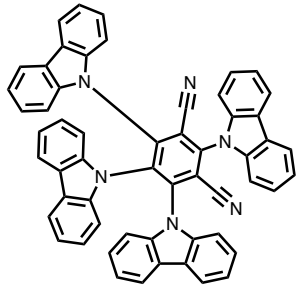
Code

$C_{56}H_{32}N_6$

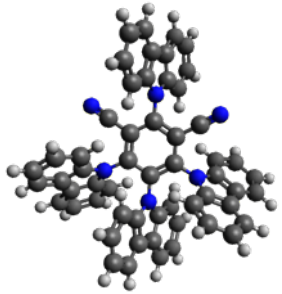
Composition

N#CC1=C(C(N2C3=C(C4=C2C=CC=C4)C=CC=C3)=C(C(C#N)=C1N5C6=C(C7=C5C=CC=C7)C=CC=C6)N8C9=C(C%10=C8C=CC=C%10)C=CC=C9)N%11C%12=C(C%13=C%11C=CC=C%13)C=CC=C%12

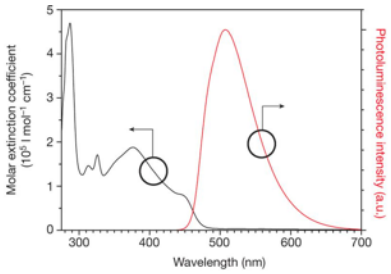
Text



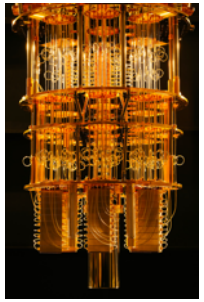
Graph



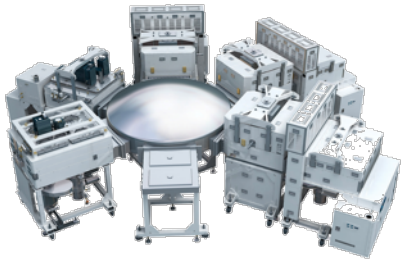
3D



Physical properties



Simulation

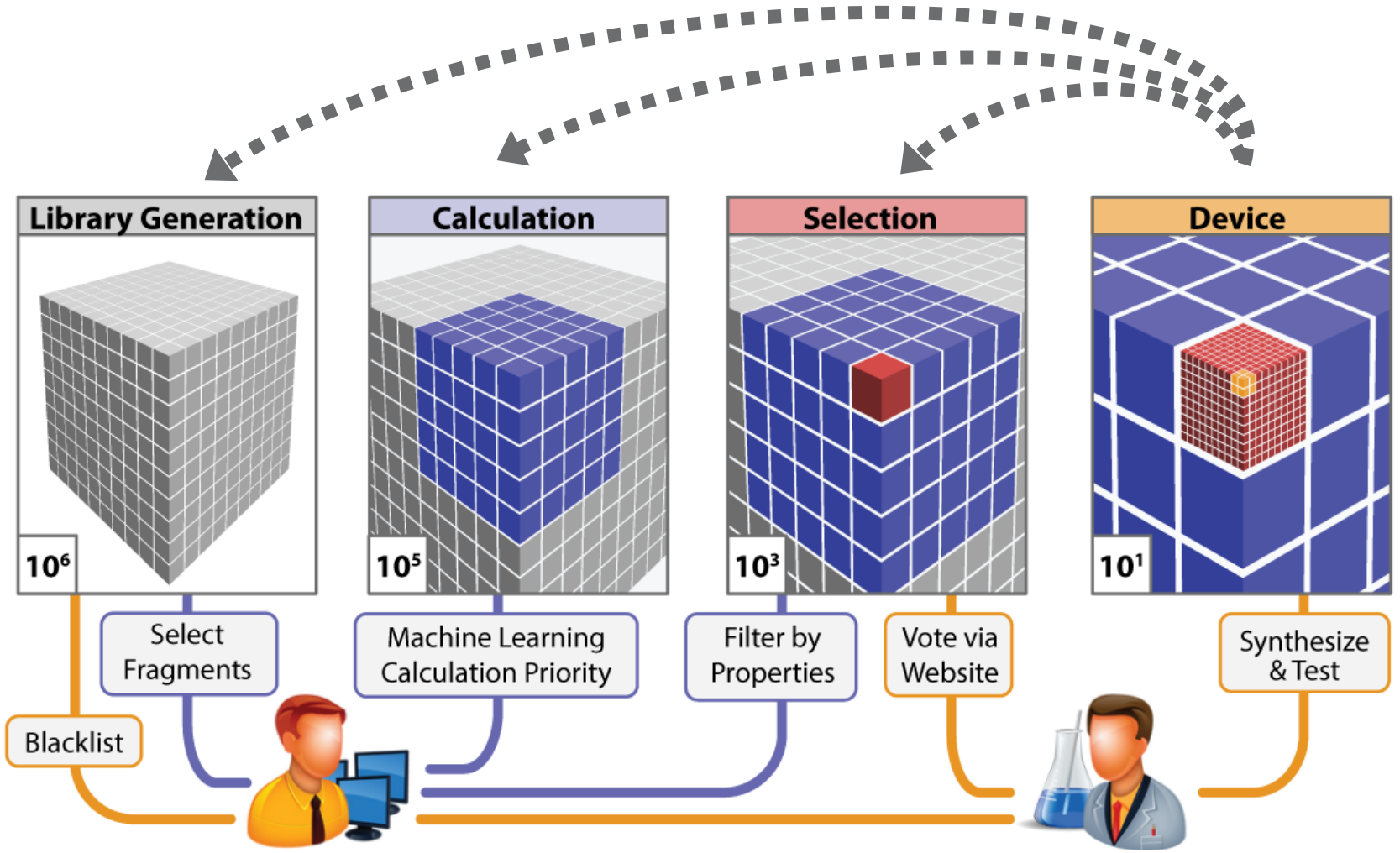


Recipe / processing

Models

Predictions

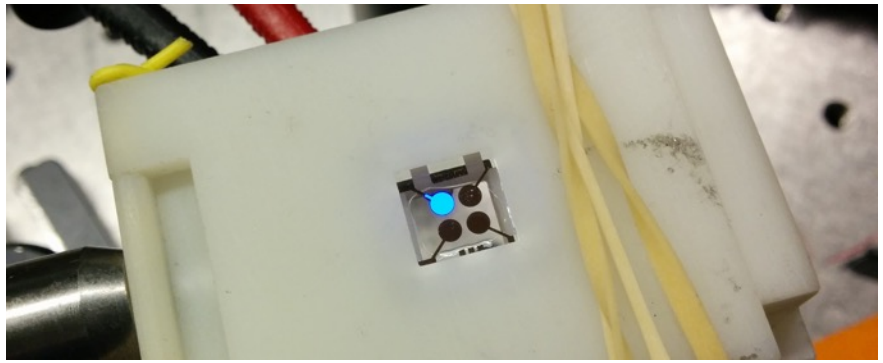
High-throughput virtual screening

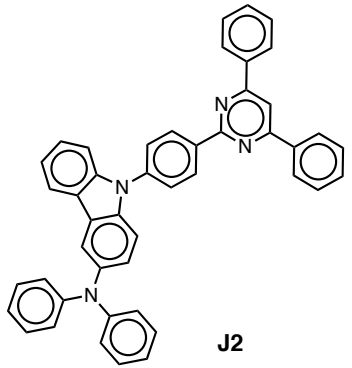


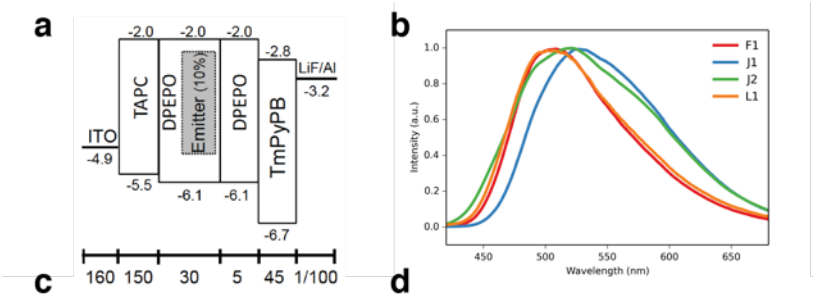
Successful applications

Organic Light Emitting Diodes

- High end displays, potentially lighting.
- Lightweight, flexible, transparent, high contrast, low power

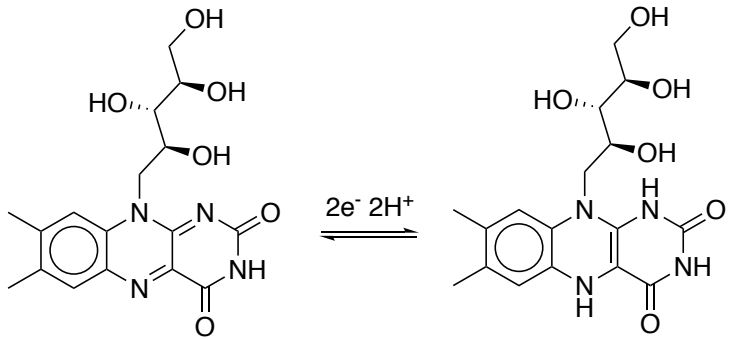






Organic Flow battery electrolytes

- High-scale energy storage
- Emerging technology, promising low-cost



RGB et al. *Nature Materials* 2016

Lin, K. *et al. Nature Energy 1*, 16102 (2015)

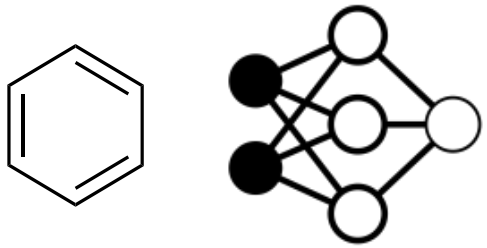
ML for bypassing calculations

Simulations

Are deterministic, reproducible, low-noise, and typically faster and cheaper than experiments.

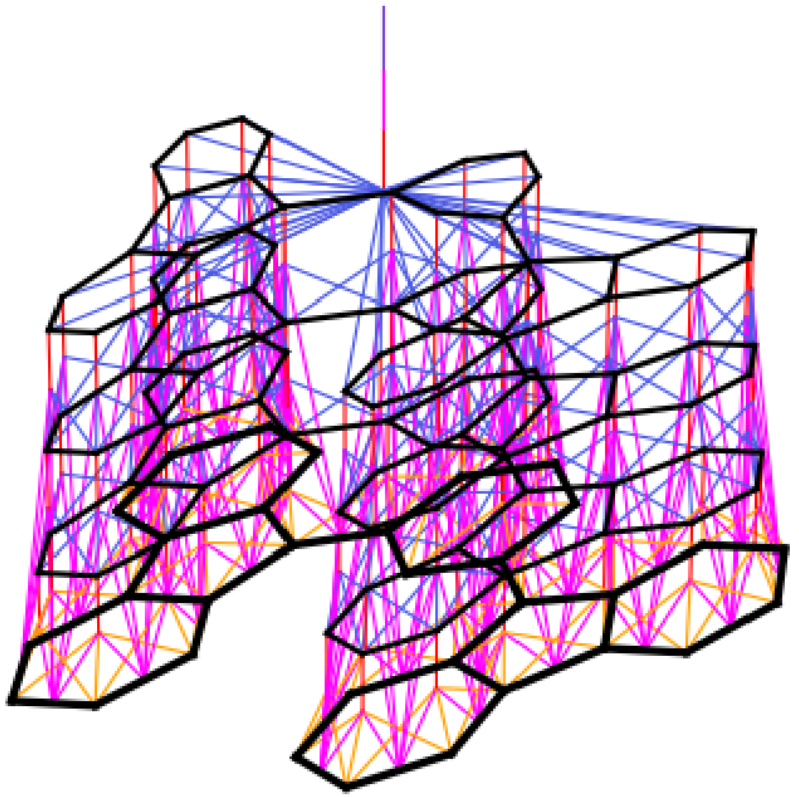
Large amounts of labeled data available.

Representations that capture physics.



Properties from expensive simulation

Message passing neural networks



Duvenaud *et al.* arXiv:1509.09292, https://github.com/HIPS/neural-fingerprint

Choosing a representation

Representation

It is common to embed the physics using *features* extracted from first principles (electronegativity, atomic mass, electronic structure, atomic radius, xyz coordinates …)

Allows higher data efficiency, better transferability.

But they are not degrees of freedom. If they are not reversible, then the search space is not represented.

End-to-end learning from the most basic representations, where we can "move"

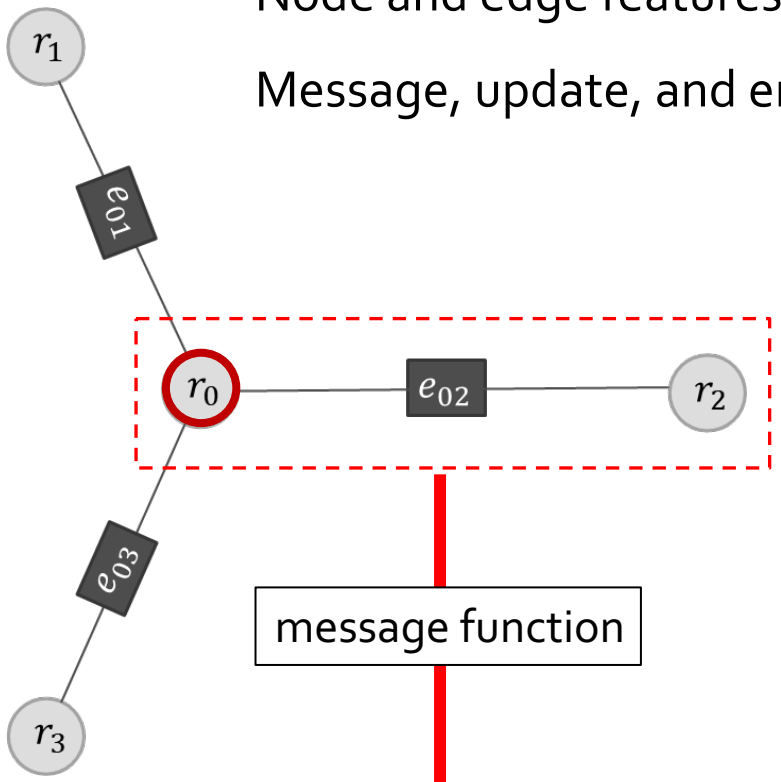
Duvenaud *et al.* arXiv:1509.09292, https://github.com/HIPS/neural-fingerprint

Message-Passing Neural networks

Atoms represented as nodes and bonds as distance-labeled edges

Node and edge features updated iteratively based on learned neighborhood mappings

Message, update, and embedding functions are neural networks



Add messages from all neighbors

$$m_0 = m_{01} + m_{02} + m_{03}$$

update function

r_0'

Update fingerprint

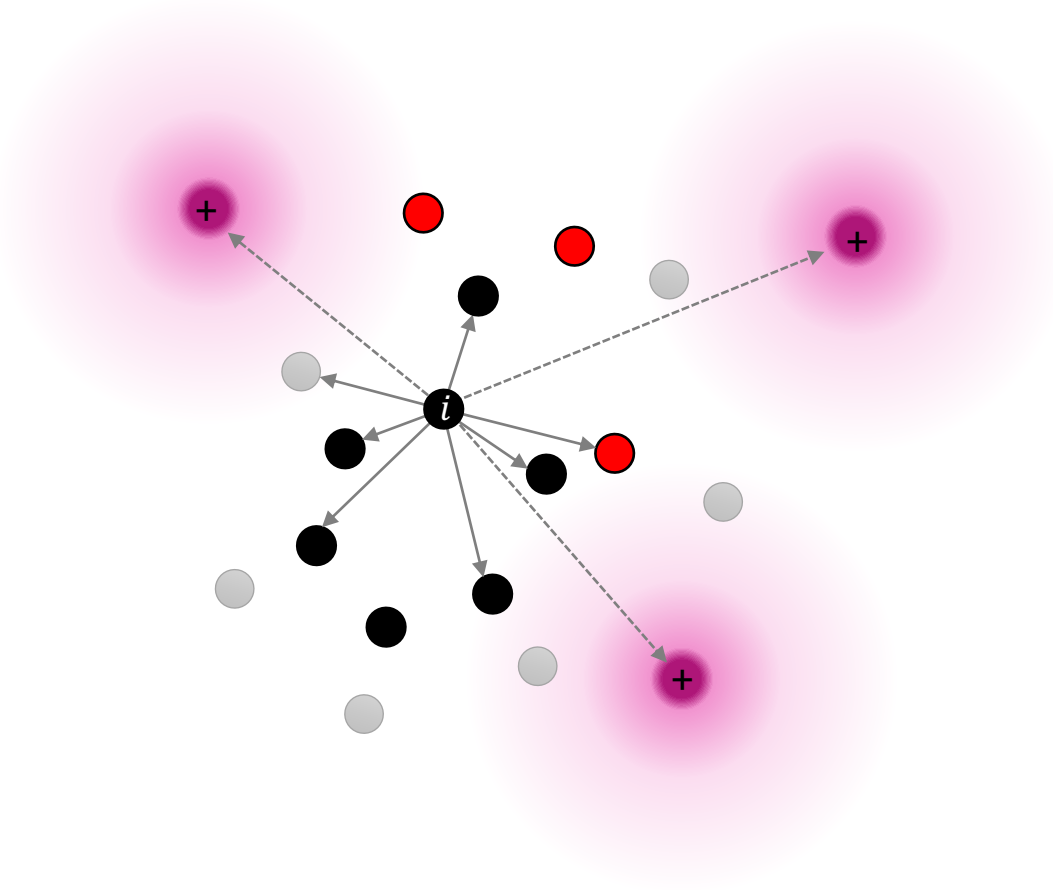
$$\{r_0', r_1', r_2', r_3'\}$$

embedding function

f'

Message-Passing Neural networks

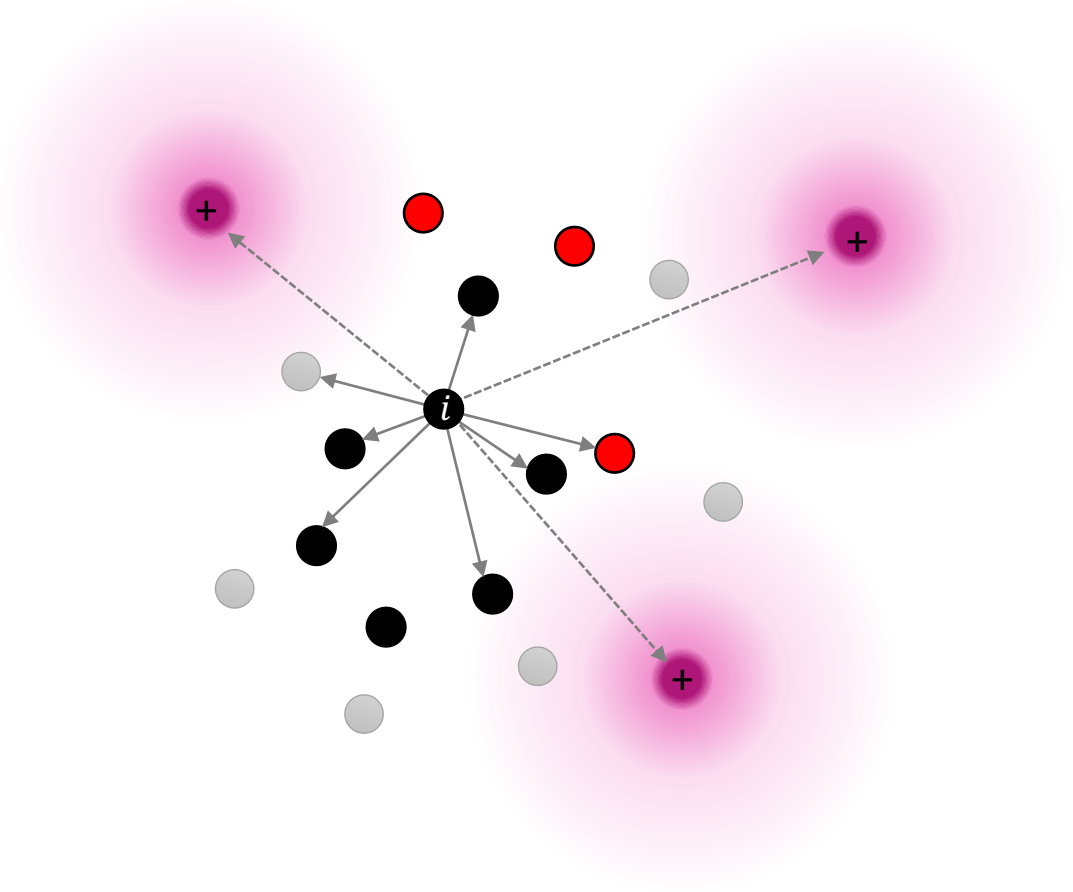
- Typically achieve state of the art performance over atomic QSPR regression problems in chemistry
  - Toxicity, solubility, optical properties, band gap, shear moduli, etc
  - Energy under 1 kcal/mol and forces under 0.1 kcal/mol/A.
- Needing 1,000-10,000 to be really effective
- Customized for a given property
  - Transferability might not be great.



Message-Passing Neural networks

All sorts of bells and whistles.

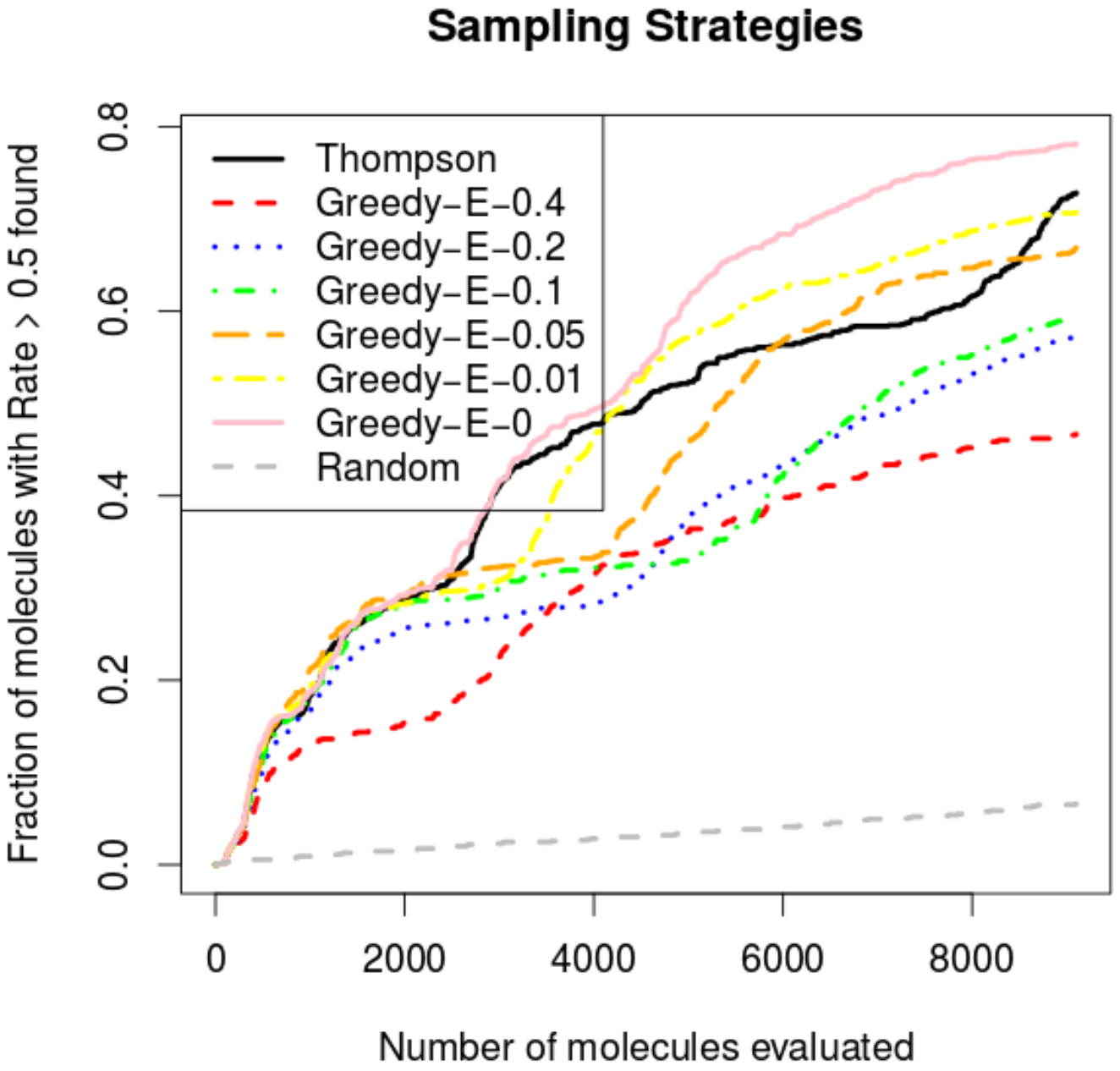
- Edge features with Gaussian basis expression of distance
  Schutt et al J. Chem. Phys. 148, 241722 (2018)
- Edge updates
- k-GNN (combine with k-neighbor rather than first order)
  Morris et al, arXiv:1810.02244
- Use mean/addition pooling function for very additive properties (energies) and concatenation for non-additive properties
- Version of periodic crystals
  T Xie et al, Physical review letters 120 (14), 145301



Explore/exploit

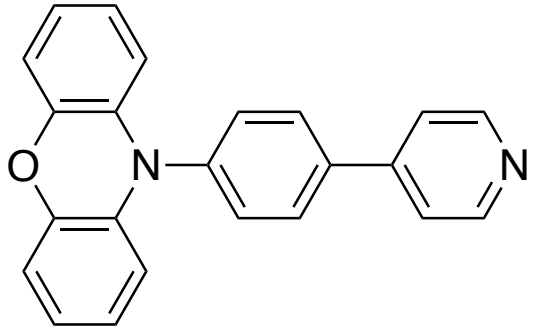
Balancing exploration and exploitation.

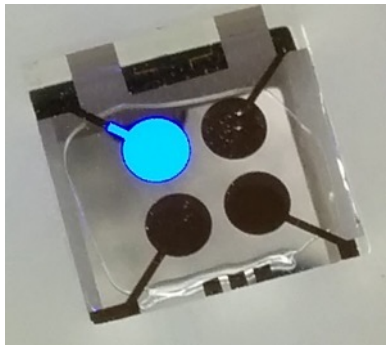
- Greedy and epsilon greedy do a good job
- Thompson sampling is pretty much equivalent



Inverse design

Progress in predicting performance given candidate





Can we generate candidate based on design targets?

?

Deep generative models

Unsupervised learning: Learning from data that has not been labeled, classified or categorized. Find a common denominator in the data.

In a generative model, we then use that commonality to generate novel realistic synthetic samples.

Karras, Aila, Laine, Lehtinen 2018

Nguyen, Dosovitskiy, Yosinski, Brox, Clune 2016

Variational Autoencoder

Artificial neural network used for learning efficient codings.

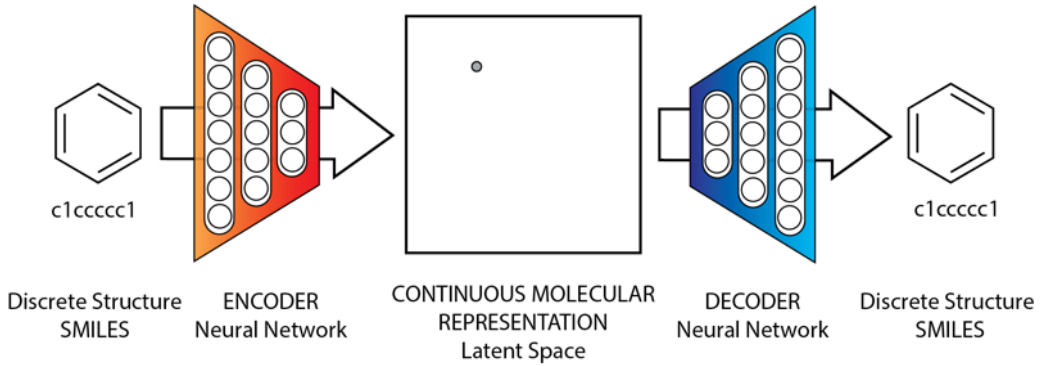
Dimensionality reduction or to learn generative models.

The latent space is continuous, derivable and low dimension.

Try to learn identity function through information bottleneck.

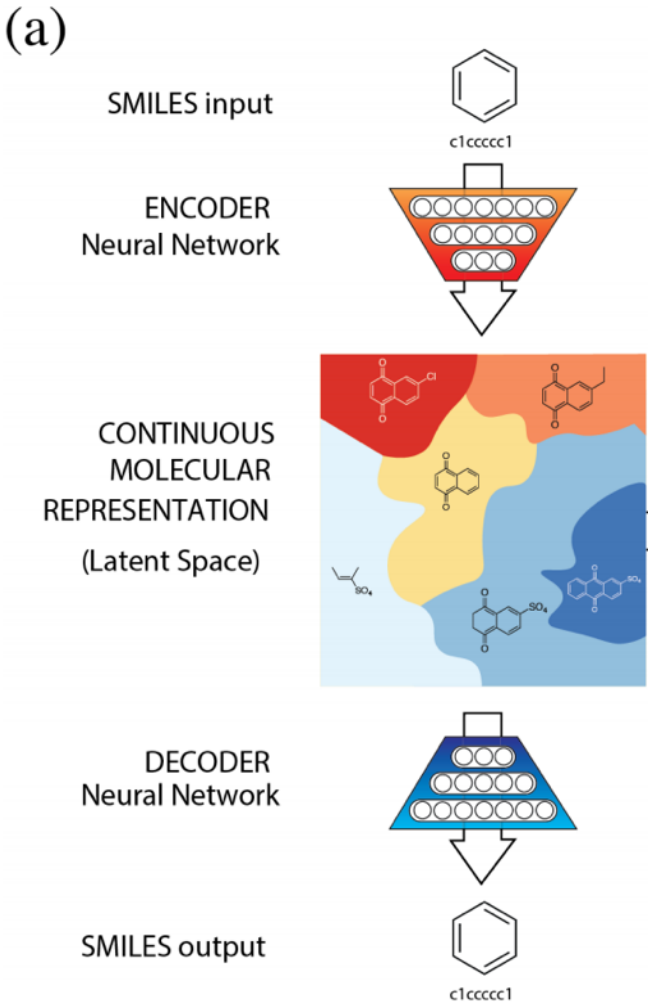
The latent representation of an input is not a single value, but a distribution:

- Acts as a regularizer, enforcing a more homogeneous latent representation
- Need a prior about the distribution of *z* (Gaussian, $\mu = 1$, $\sigma = 1$)

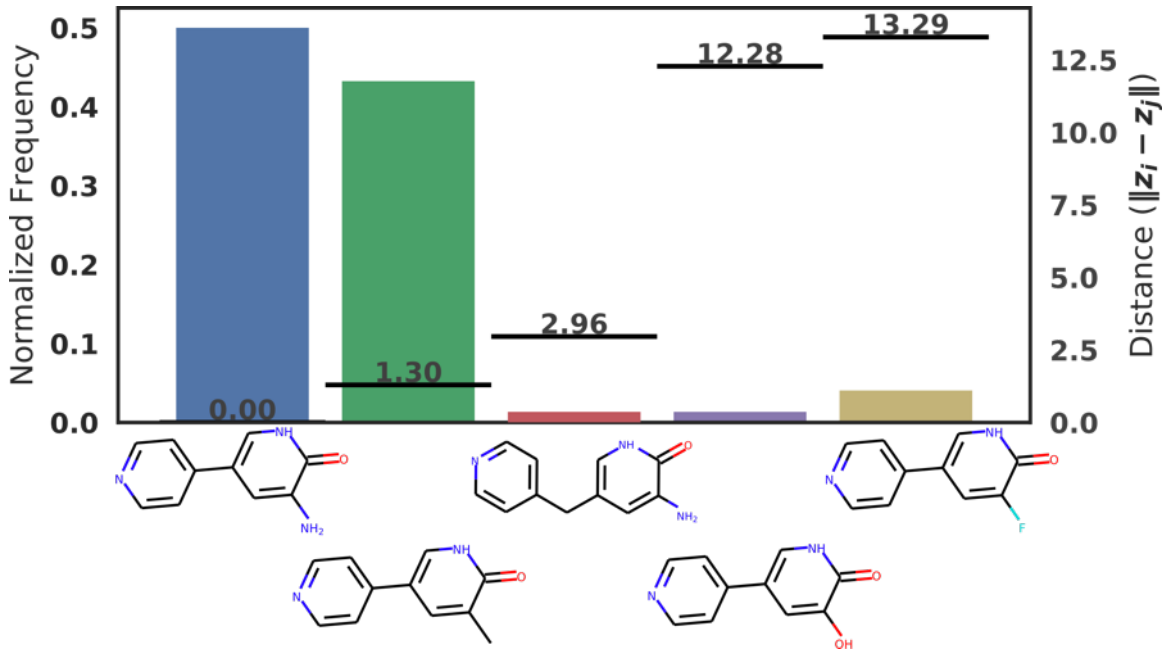


Diederik P. Kingma and Max Welling. 2015.

Chemical Variational Autoencoder

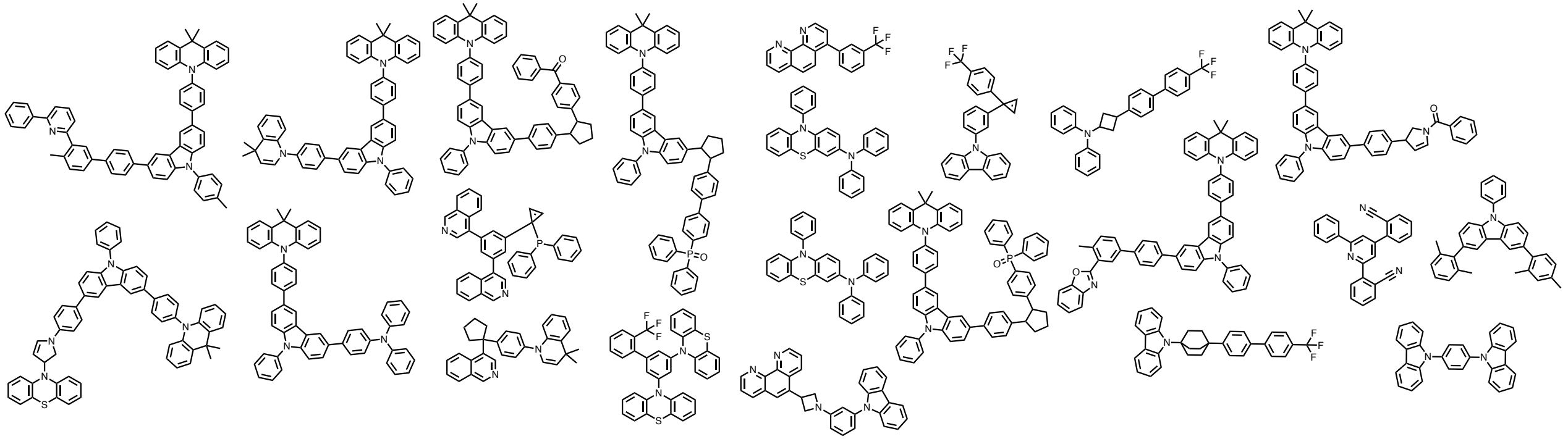


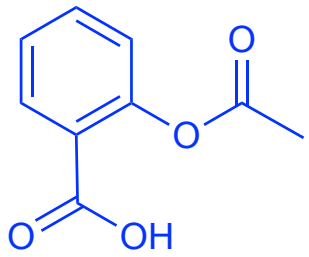
Trained on 250,000 drug-like molecules from ZINC database



Dreaming OLEDs

Trained on HTVS library and patented OLED
No bias, just generation

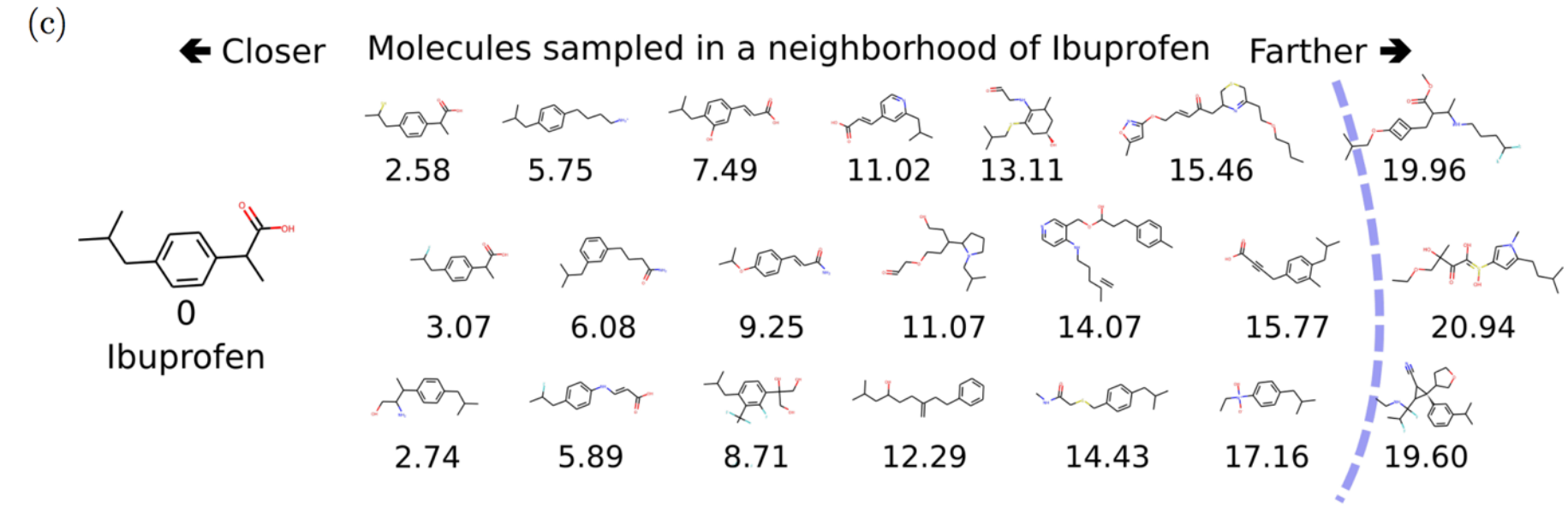




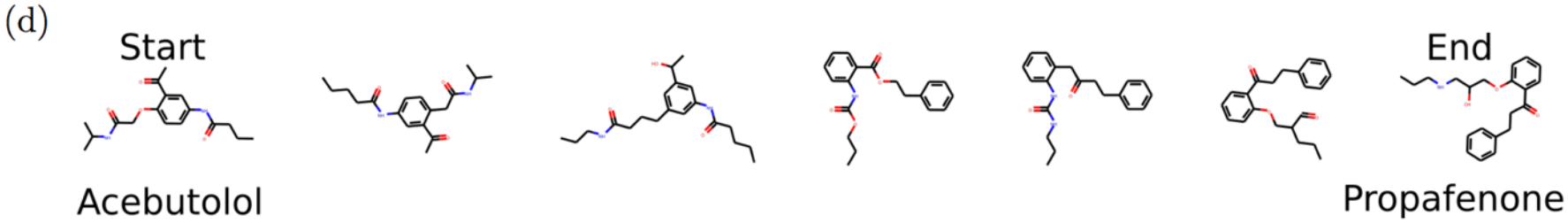
Aspirin

Non-linear avigation

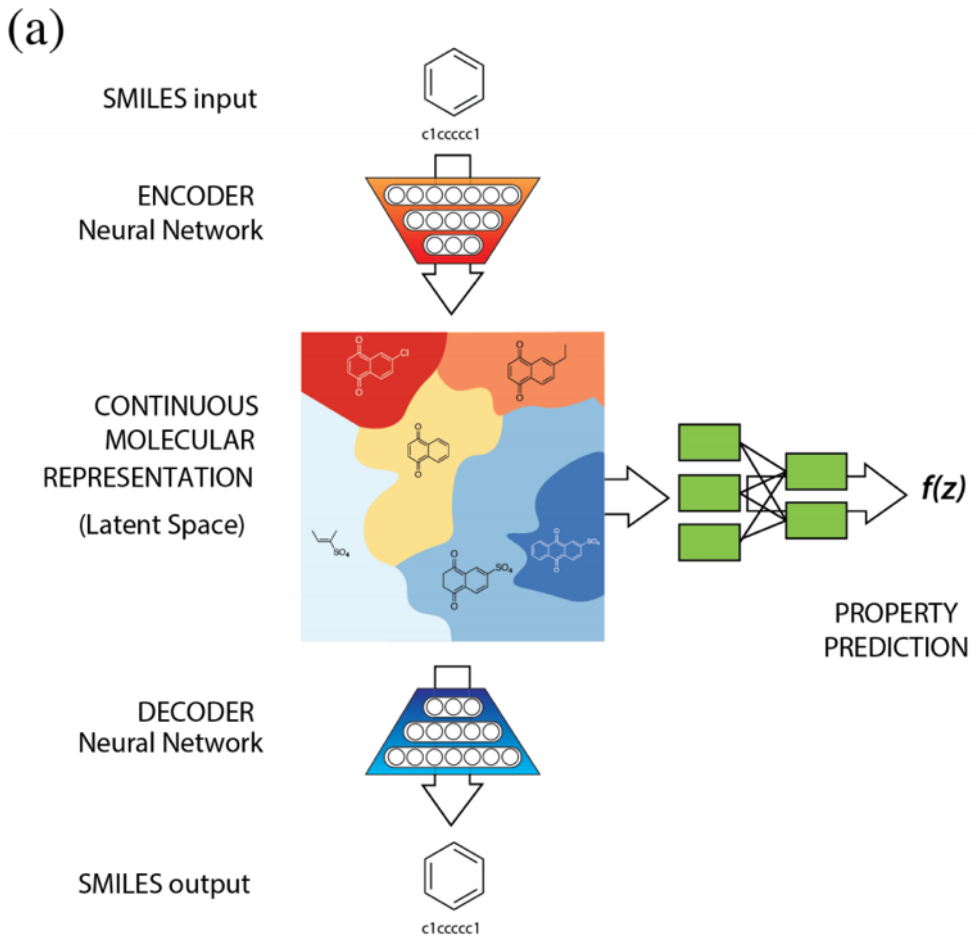
The space is very much not linear. Most molecules are in an annulus far from the mean.



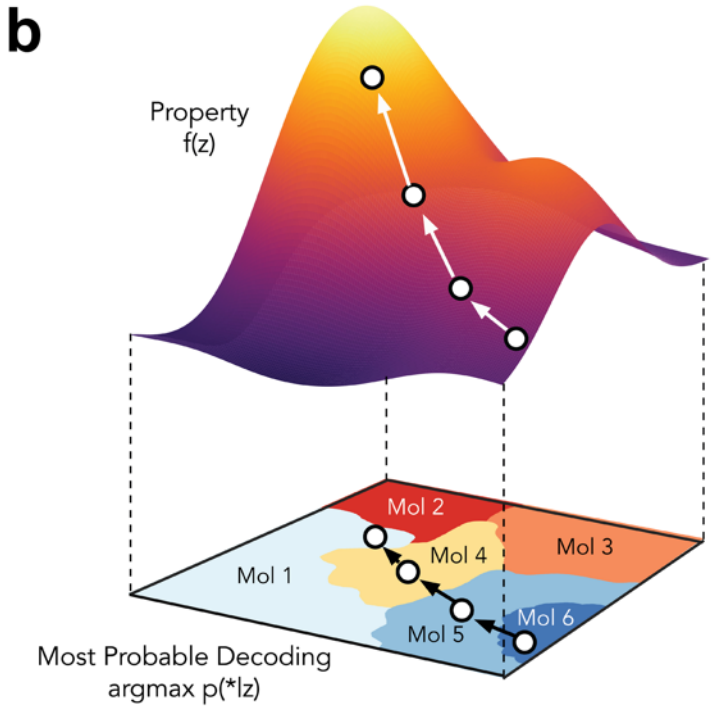
SLERP (Spherical interpolation) allows taking much more sensible steps



Semisupervised Molecular VAE

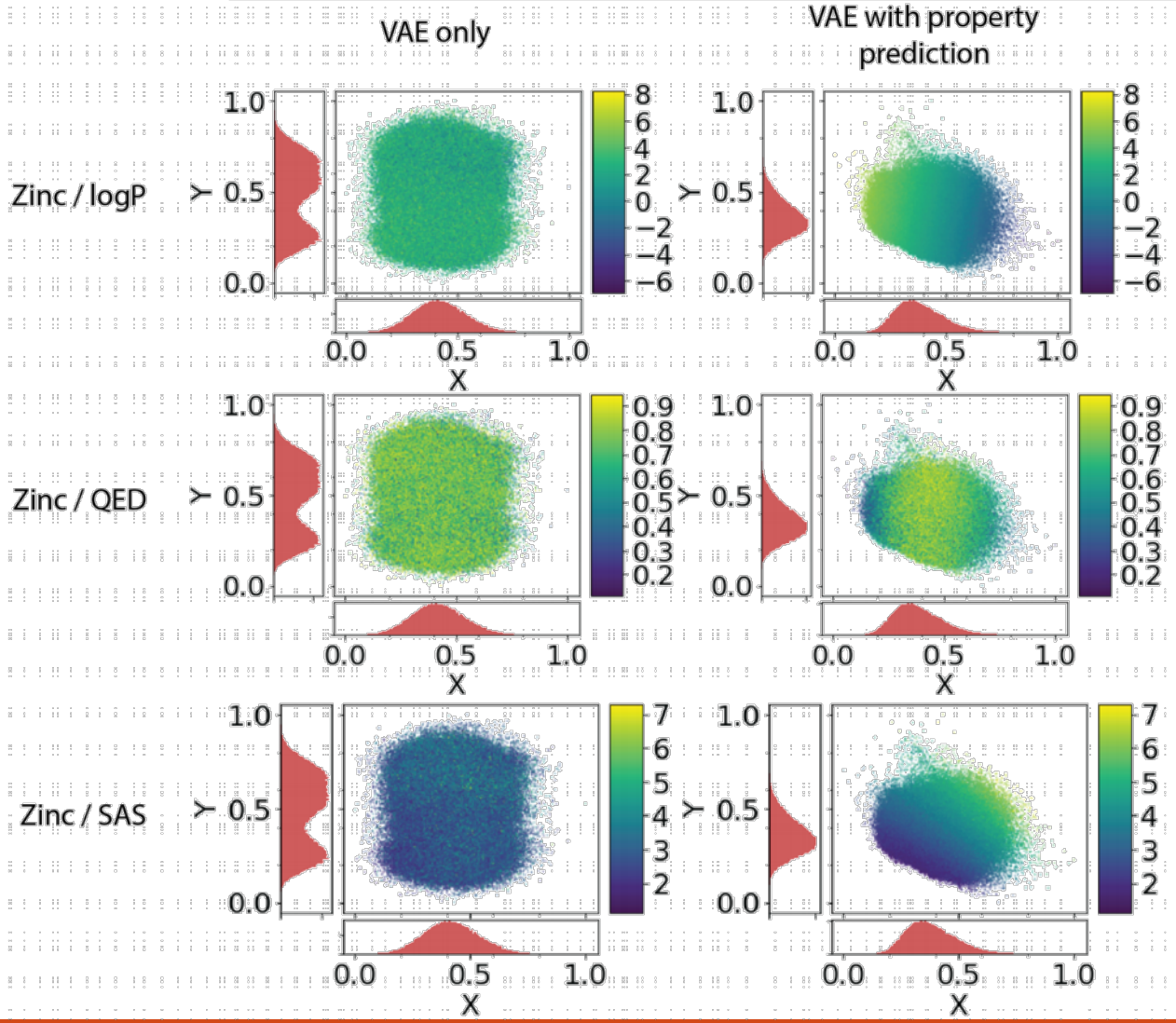


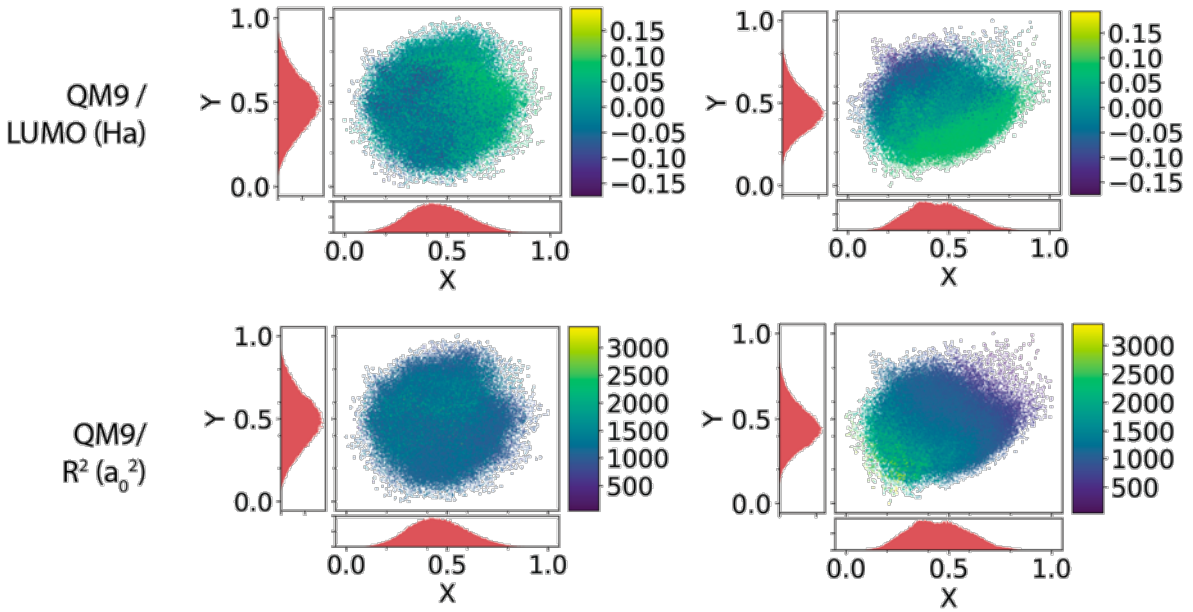
The latent representation now encodes mapping to one or more properties.



Gomez-Bombarelli et al. ACS Central Science **2018**, *4* (2), 268–276

Semisupervised performance

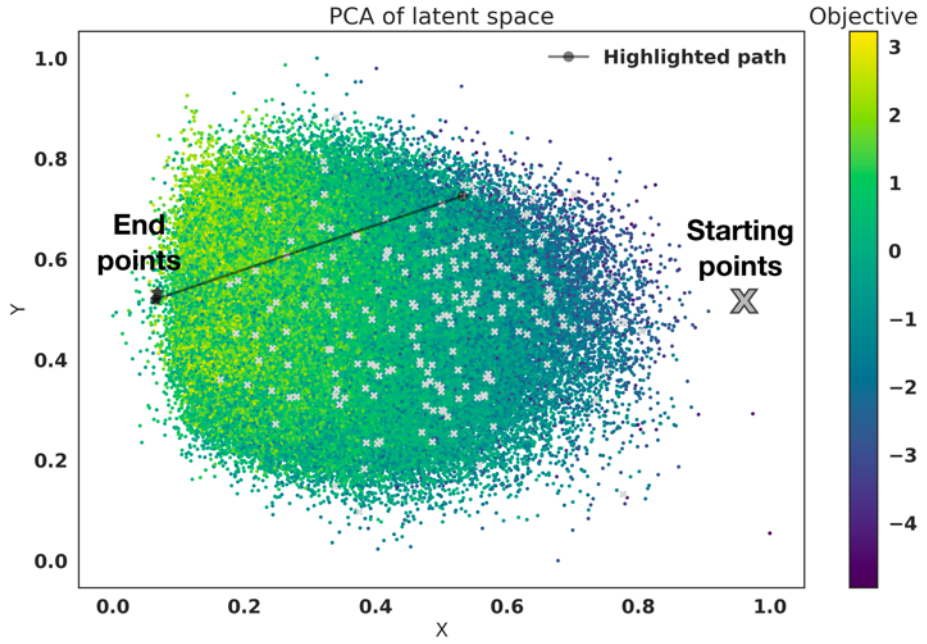


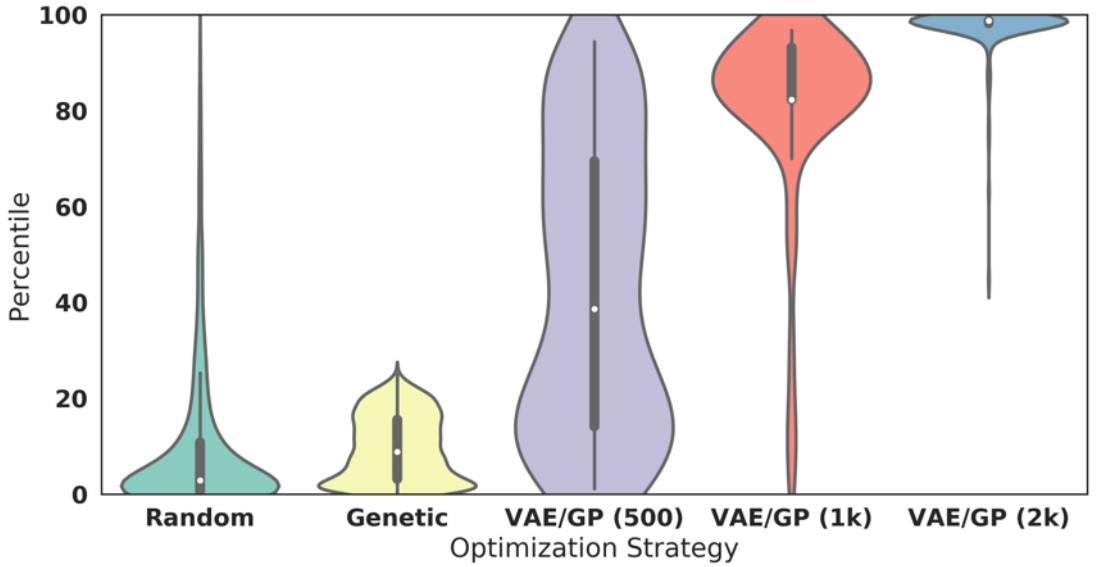


1D Convolutions over SMILES strings afford similar performance to neural convolutions and graph-based representations.

Optimization

Leverage gradients to inform optimization
Leverage smooth, continuous QSPR
Mix with optimizer of choice





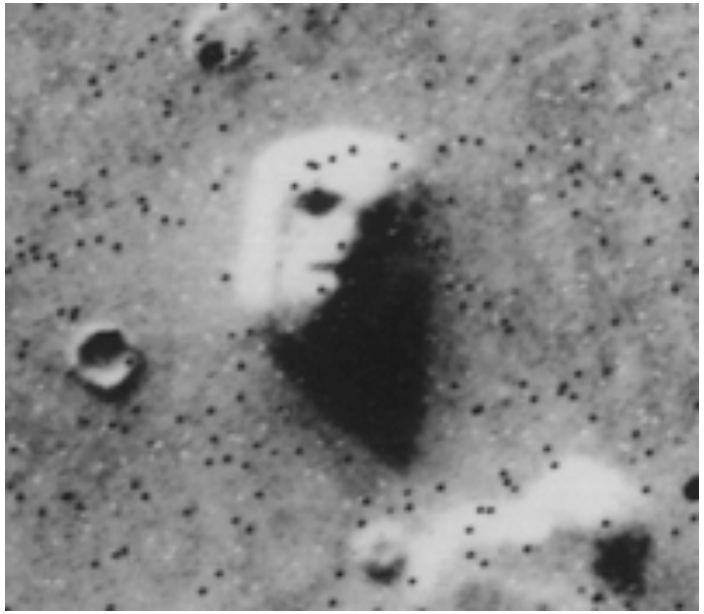
Chemical space is different

Generative models of images rely of human's prowess in pattern-recognition.

This works …

Radford, Metz, Chintala arXiv:1511.06434v2 2016

… a bit like this works



Seeing Jesus in toast: Neural and behavioral correlates of face pareidolia. Jiangang Liu, Jun Li, Lu Feng, Ling Li, Jie Tian, Kang Lee

Extensions, upgrades and alternatives

Character-by-character RNN

Generating Focused Molecule Libraries for Drug Discovery with Recurrent Neural Networks, ACS Cent. Sci., **2018**, *4* (1), pp 120–131

SMILES Enumeration as Data Augmentation for Neural Network Modeling of Molecules, arXiv:1703.07076 [cs.LG]

Molecular de-novo design through deep reinforcement learning, Journal of Cheminformatics **2017** *9*:48

Generative Recurrent Networks for De Novo Drug Design, Molecular informatics **2018** *37*, - Whole issue!

Better decoder

Grammar Variational Autoencoder, arXiv:1703.01925

Bayesian molecular design with a chemical language model, J Comput Aided Mol Des **2017** *31* 379.

Syntax-Directed Variational Autoencoder for Structured Data arXiv:1802.08786

Population-based de novo molecule generation, using grammatical evolution, arXiv:1804.02134

Extensions, upgrades and alternatives

Generative adversarial networks

Optimizing distributions over molecular space.An Objective-Reinforced GenerativeAdversarial Network for Inverse-designChemistry (ORGANIC) ChemRxiv

Objective-Reinforced Generative Adversarial Networks (ORGAN) for Sequence Generation Models, arXiv:1705.10843

Adversarial Threshold Neural Computer for Molecular de Novo Design, Mol. Pharmaceutics

Graph decoding

Junction Tree Variational Autoencoder for Molecular Graph Generation, arXiv:1802.04364

GraphRNN: A Deep Generative Model for Graphs, arXiv:1802.08773

Designing Random Graph Models Using Variational Autoencoders With Applications to Chemical Design, arXiv:1802.05283

Towards Variational Generation of Small Graphs, ICLR 2018

Optimization

Latent Constraints: Learning to Generate Conditionally from Unconditional Generative Models, arXiv:1711.05772

Sequence to Better Sequence: Continuous Revision of Combinatorial Structures PMLR 70:2536-2544, 2017

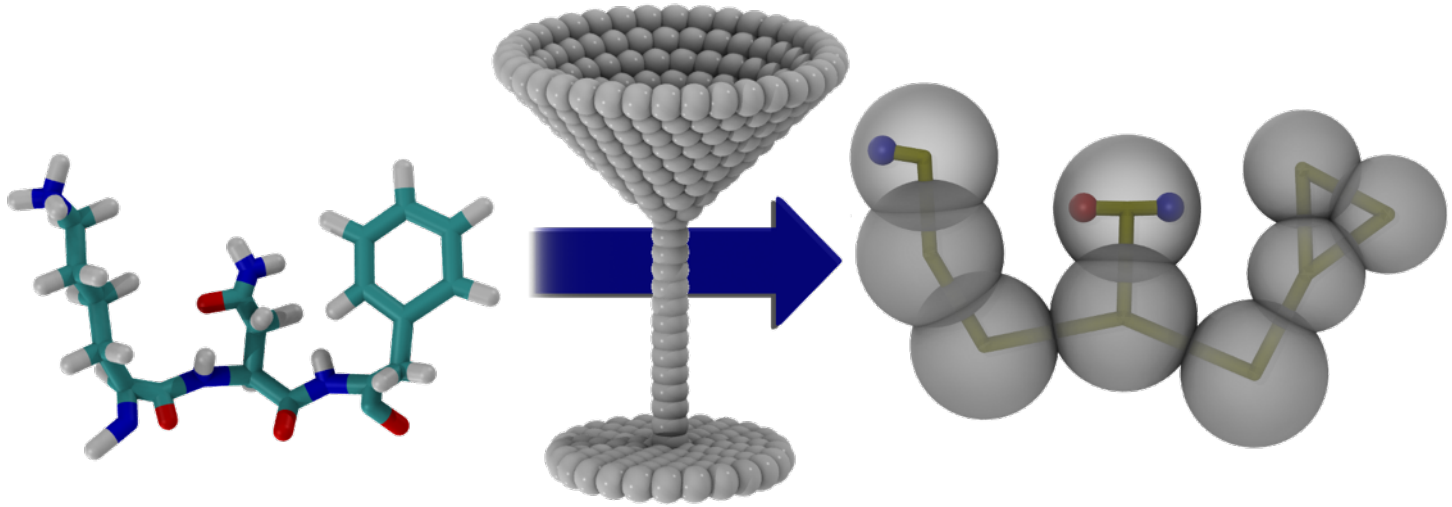
Coarse-graining MD

INVERSE DESIGN IN 3D

Coarse Grained Methods

Coarse Graining MD simulates coarse grained variables that represents **slow collective atomistic motions** derived from full atomistic simulations

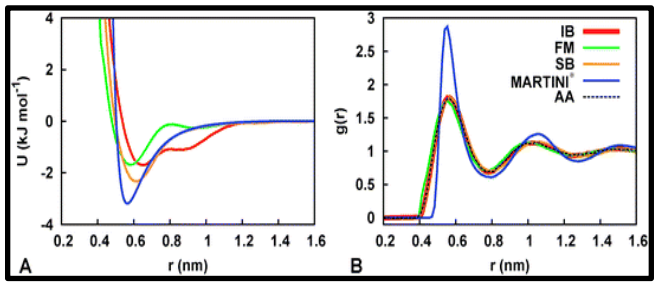
Coarse Graining methods find the "effective" coarse grained potential **given** a predetermined coarse graining mapping



S. J. Marrink et al. J. Phys. Chem. B **111**, 7812 (2007).

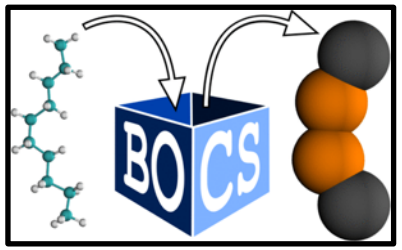
Coarse Grained Methods

Extensively studied how to find the coarse graining potentials that reproduces ***equilibrium*** structural correlation function from atomistic simulations given a pre-determined CG mapping



[1]

Methods to approximate Coarse Grained Force Fields: Relative Entropy, Force Matching, g-YBG (implemented in VOTCA, BOCS, etc.)



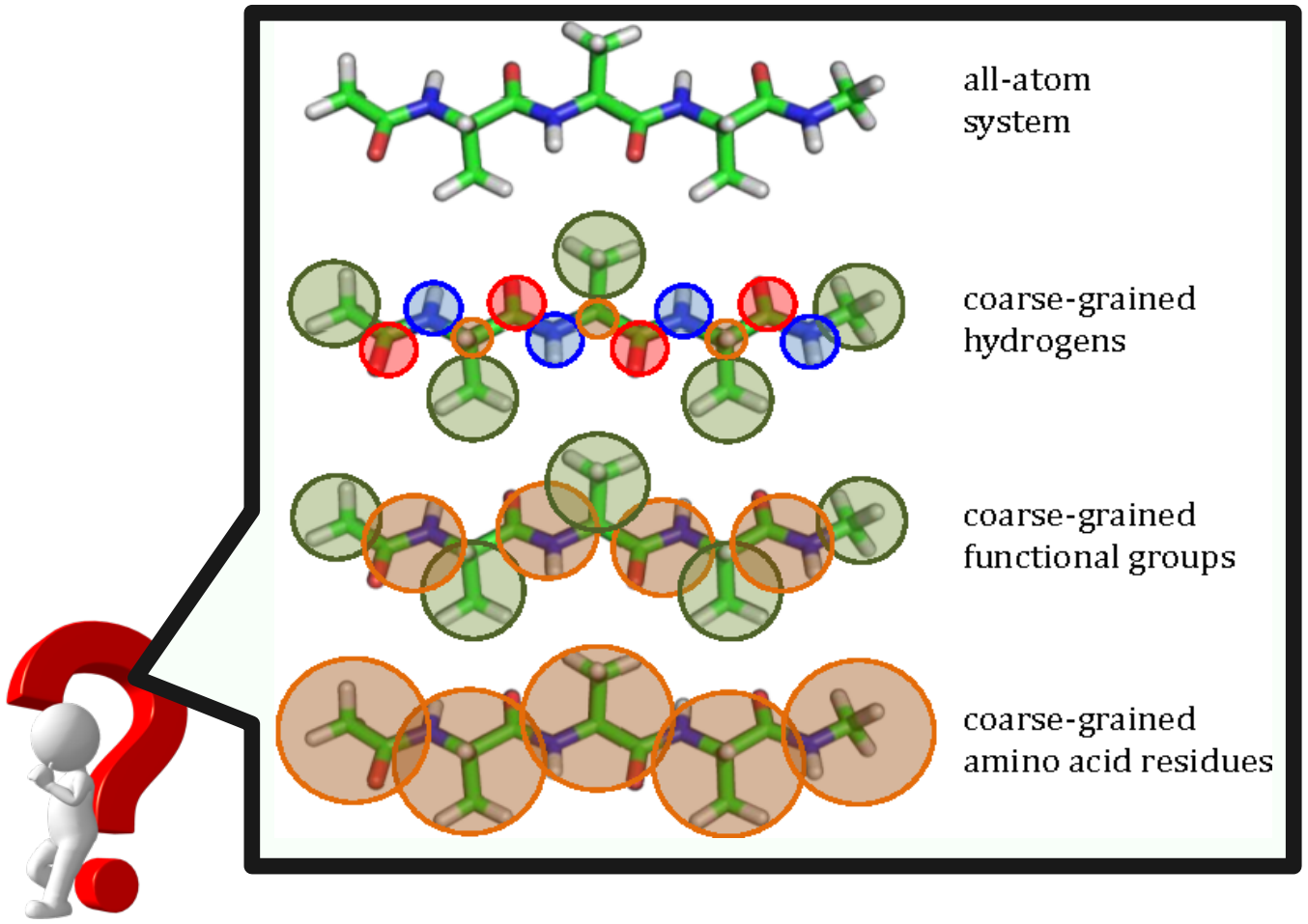
[2]

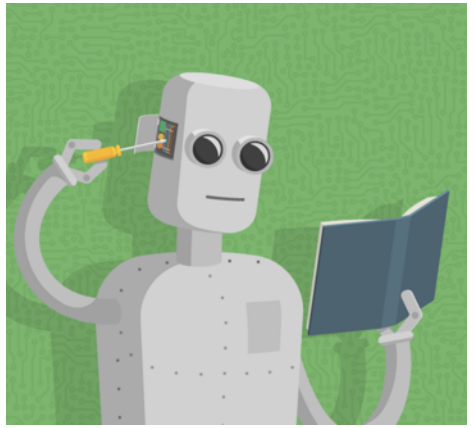
Systematic Coarse Grained force fields for Biomolecules: MARTIN

[3]

[1] A. J. Rzepiela et al. Phys. Chem. Chem. Phys. **13**, 10437 (2011). [2] N. J. H. Dunn et al. J. Phys. Chem. B **122**, 3363 (2018). [3] S. J. Marrink et al. J. Phys. Chem. B **111**, 7812 (2007).

Learning to Coarse-Grain

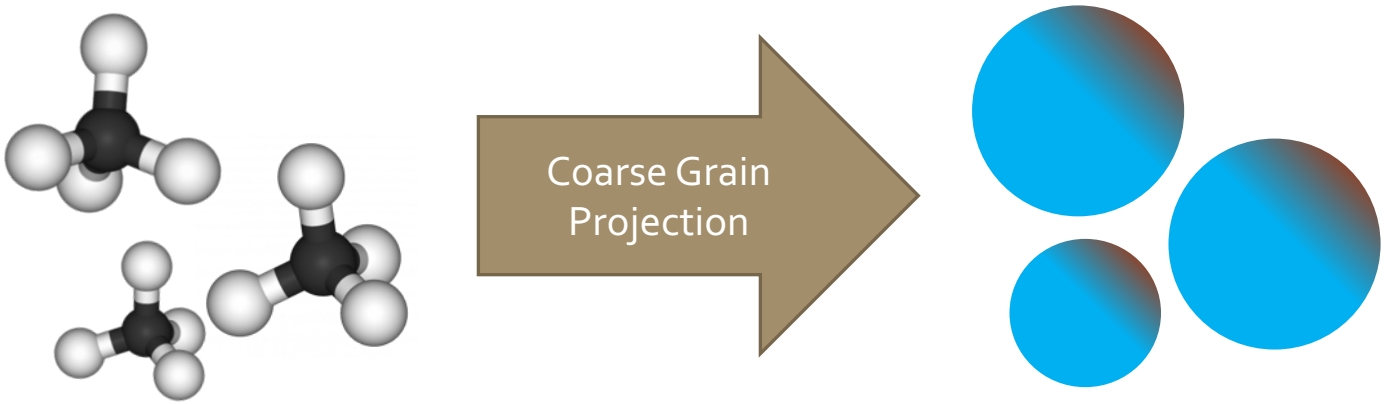




A learning problem

M. S. Shell, in Adv. Chem. Phys. (Wiley-Blackwell, 2016), pp. 395–441.

Coarse-graining framework



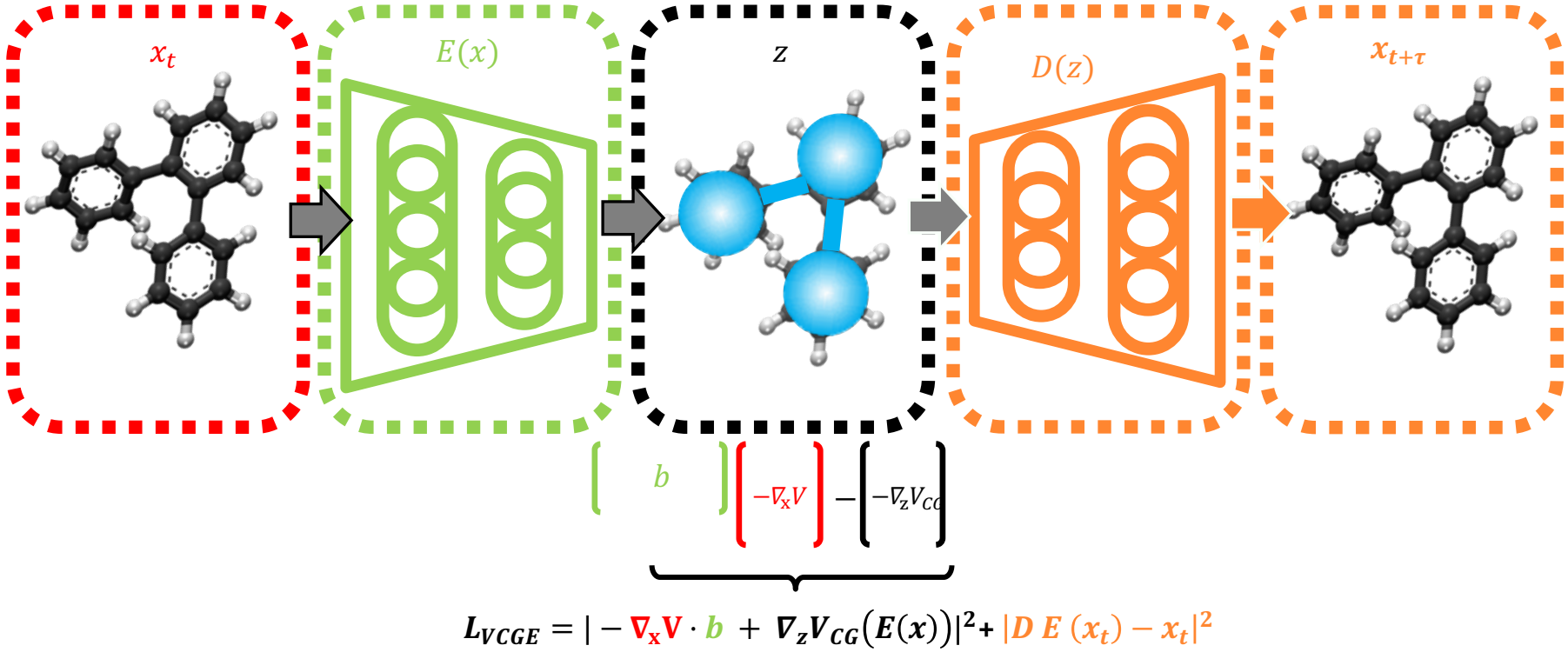
$\{x, V(x)\}$ $\quad z = E(x)$ $\quad \{z, V_{CG}(z)\}$

$\mathbf{x}$: atomistic coordinates
$\mathbf{V}(\mathbf{x})$: All-Atom Potential
$\boldsymbol{z}$: coarse grained coordinates
$\boldsymbol{V_{CG}(z)}$: coarse grained Potential

Instead of
given $\boldsymbol{x}, \boldsymbol{V(x)}, \boldsymbol{E(x)} \rightarrow$ find $\boldsymbol{V_{CG}(z)}$
We propose
given $\boldsymbol{x}, \boldsymbol{V(x)}$ $\rightarrow$ find $\boldsymbol{E(x)}$ and $\boldsymbol{V_{CG}(z = E(x))}$

$V_{CG}(z)$ can have an arbitrary functional form
- Classica l
  $\boldsymbol{V_{CG}(z) = V_{bonded}(z) + V_{non-bonded}(z)}$
- Neural (using MPNN)

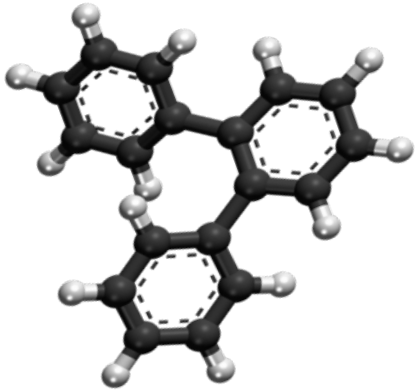
Coarse Graining Auto-Encoding Framework



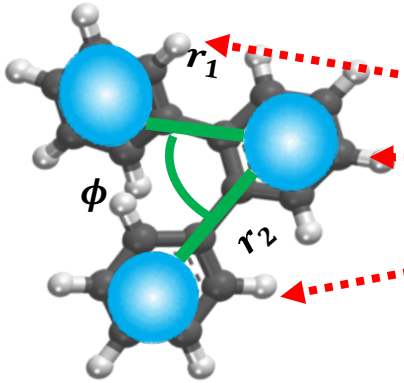
$$L_{VCGE} = | - \nabla_x V \cdot b + \nabla_z V_{CG}(E(x))|^2 + |D\,E\,(x_t) - x_t|^2$$

- **AutoEncoder** automatically coarse-grains atomistic coordinates to CG coordinates in a data-driven way
- **Force matching** also helps to shape the learning of CG and obtain $V_{CG}(z = E(x))$ for CG simulations

Automatic CG for small molecules

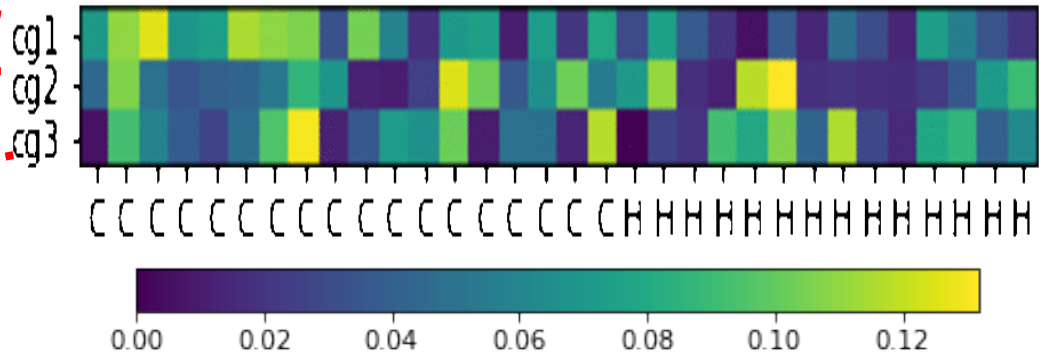


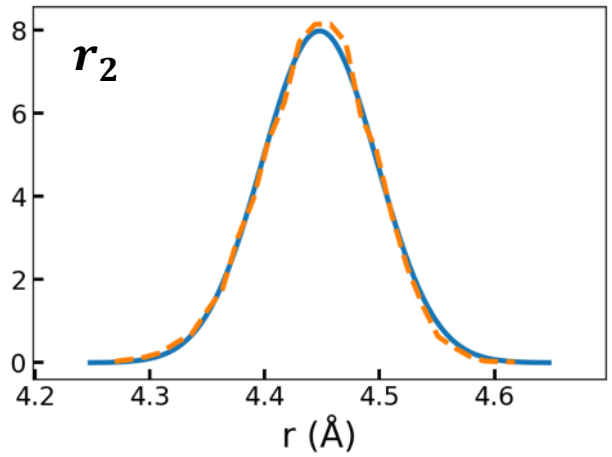
$$V(x) = V_{bond}(\mathrm{x}) + \mathrm{V_{angle}}(\mathrm{x}) + \mathrm{V_{dihe}}(\mathrm{x}) + \mathrm{V_{pair}}(\mathrm{x}) + \mathrm{V_{coulomb}}(\mathrm{x})$$

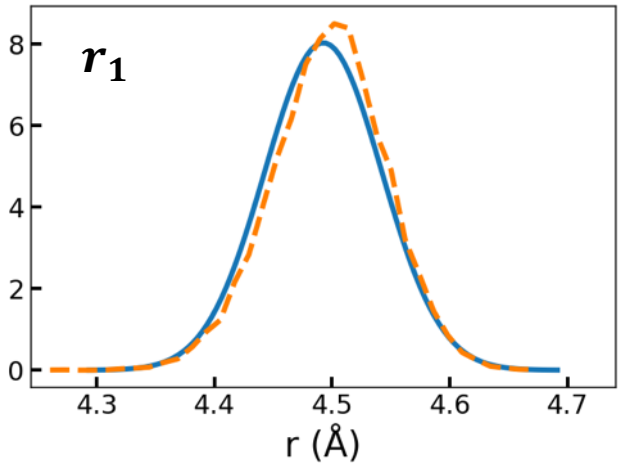


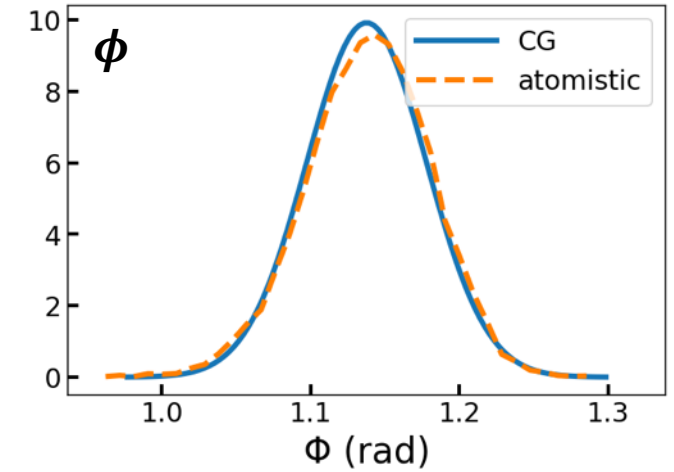
$$V_{CG}(z) = V_{bond}(\mathrm{z}) + \mathrm{V_{angle}}(\mathrm{z})$$

CG Encoder In Training

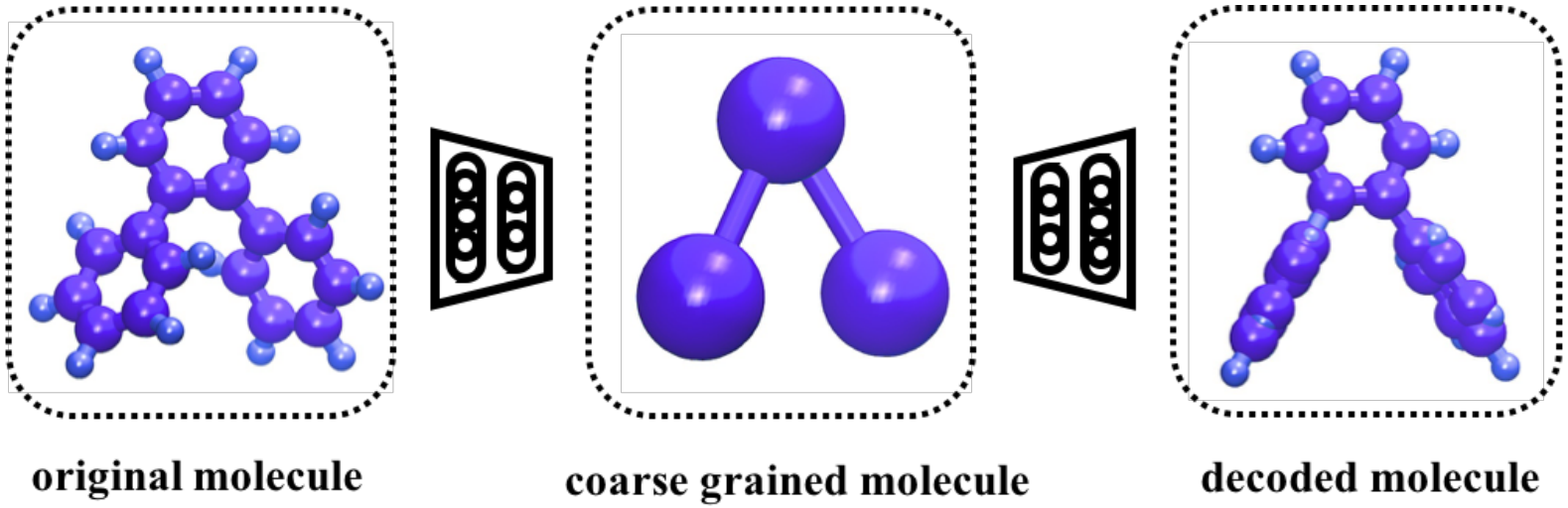








Decoding structures

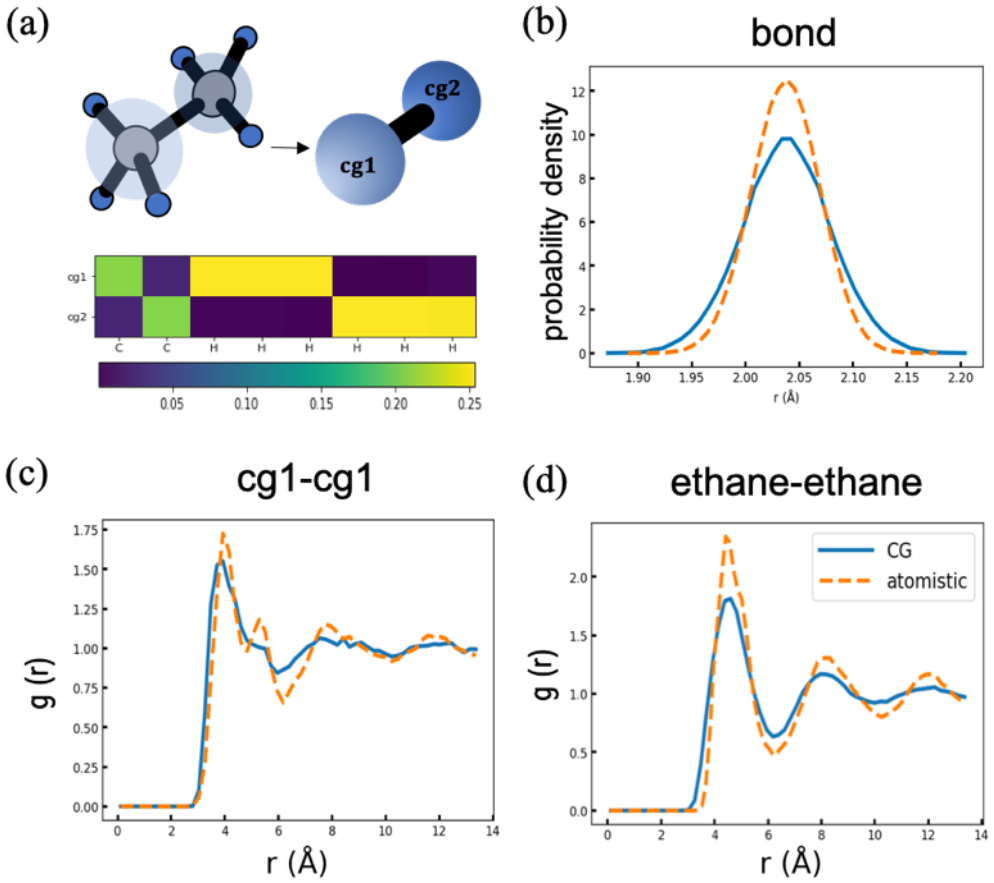


We have sacrificed degrees of freedom, the information is gone and only an average is decoded.

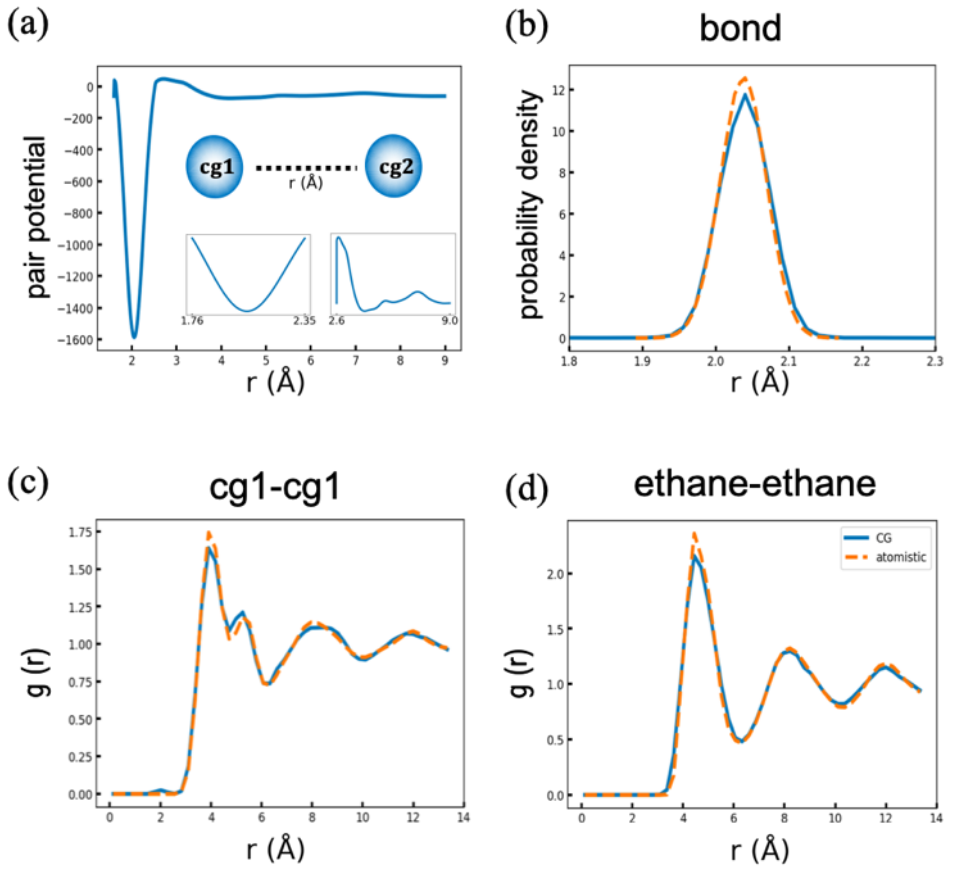
Adding stochastic noise to the decoder is a promising avenue to sample instantaneous configurations.

CG of liquid ethane

Classical potential

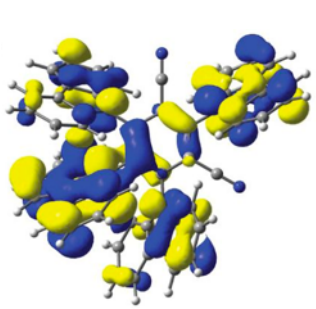


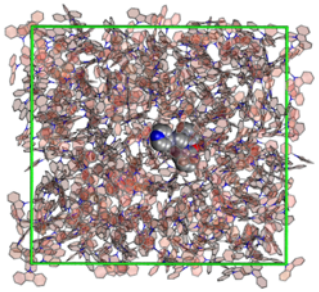
MPNN neural potential

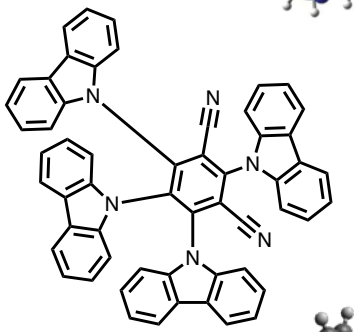


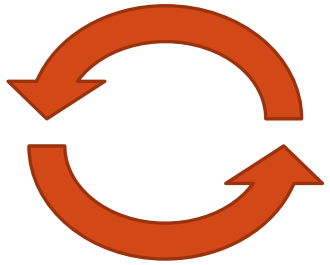
Conclusions and outlook

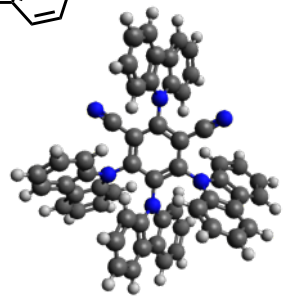
Blurring lines between ML and simulation

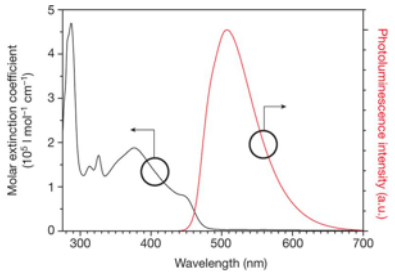










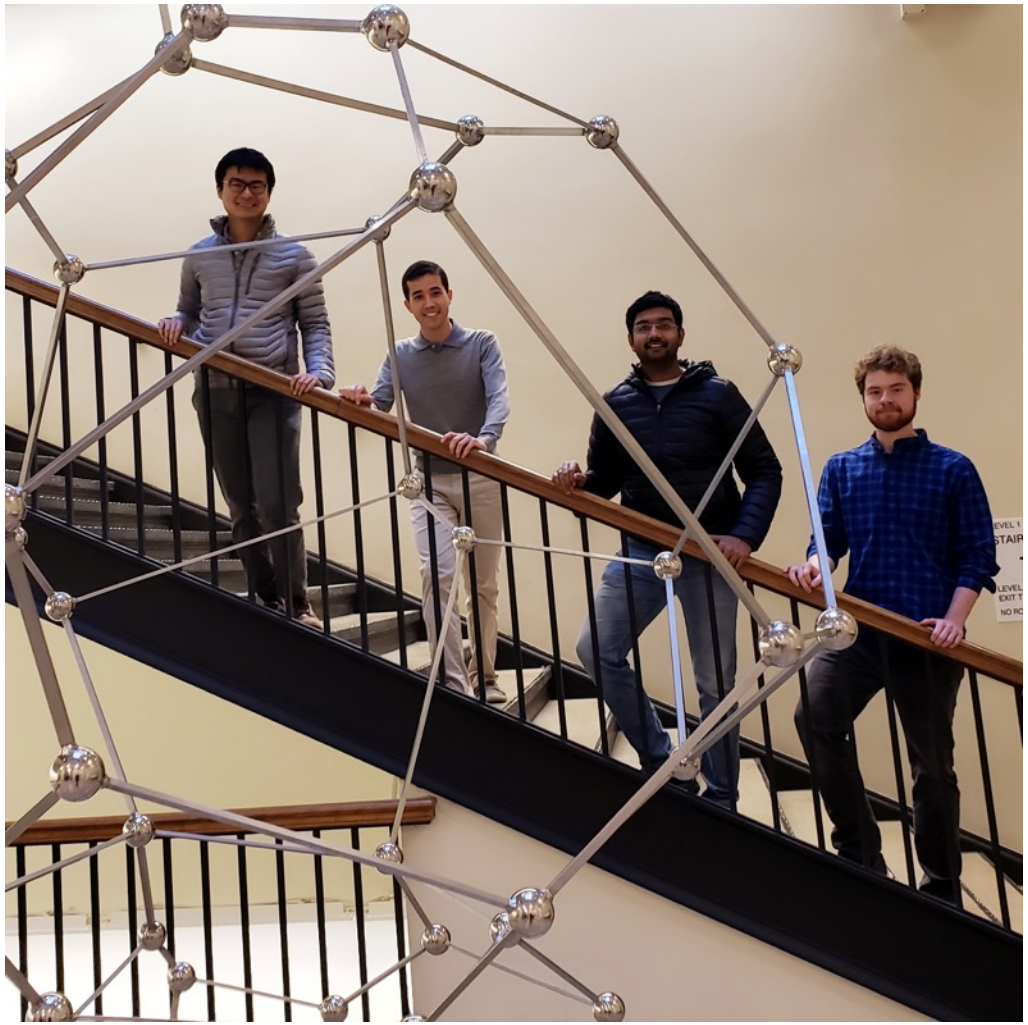


Parametrization of a physics model is a learning problem. Even building a physics model may be a learning problem

Machine learning over matter is an issue of representation: how to input a system in a way that captures known physics and chemistry

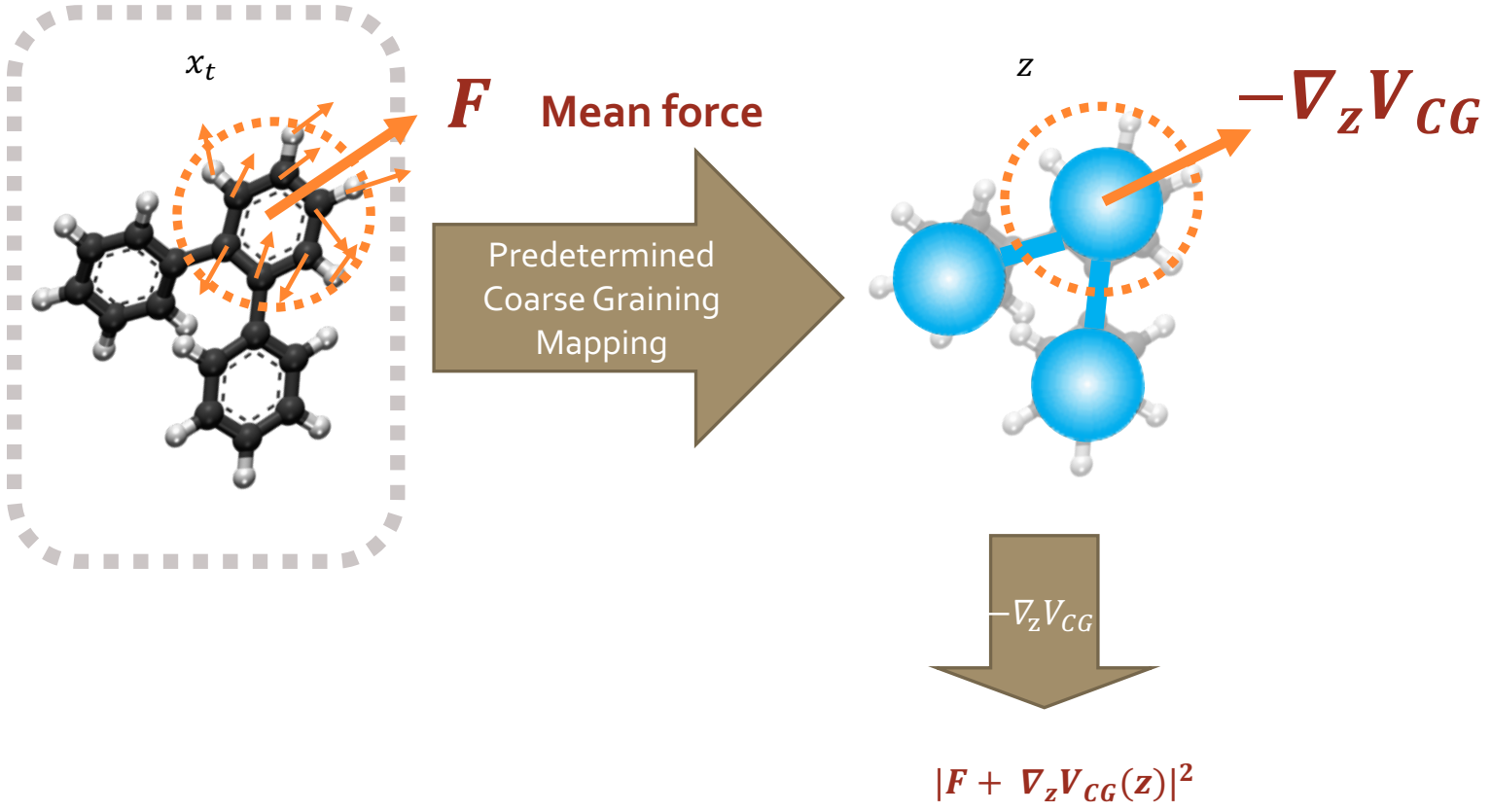
Simulations and machine learning are two sides of the same coin.

Thanks!



Wujie Wang (CG)
Wil Harris
Daniel Schwalbe Koda
Somesh Mohapatra
James Damewood
Shi Jun Ang

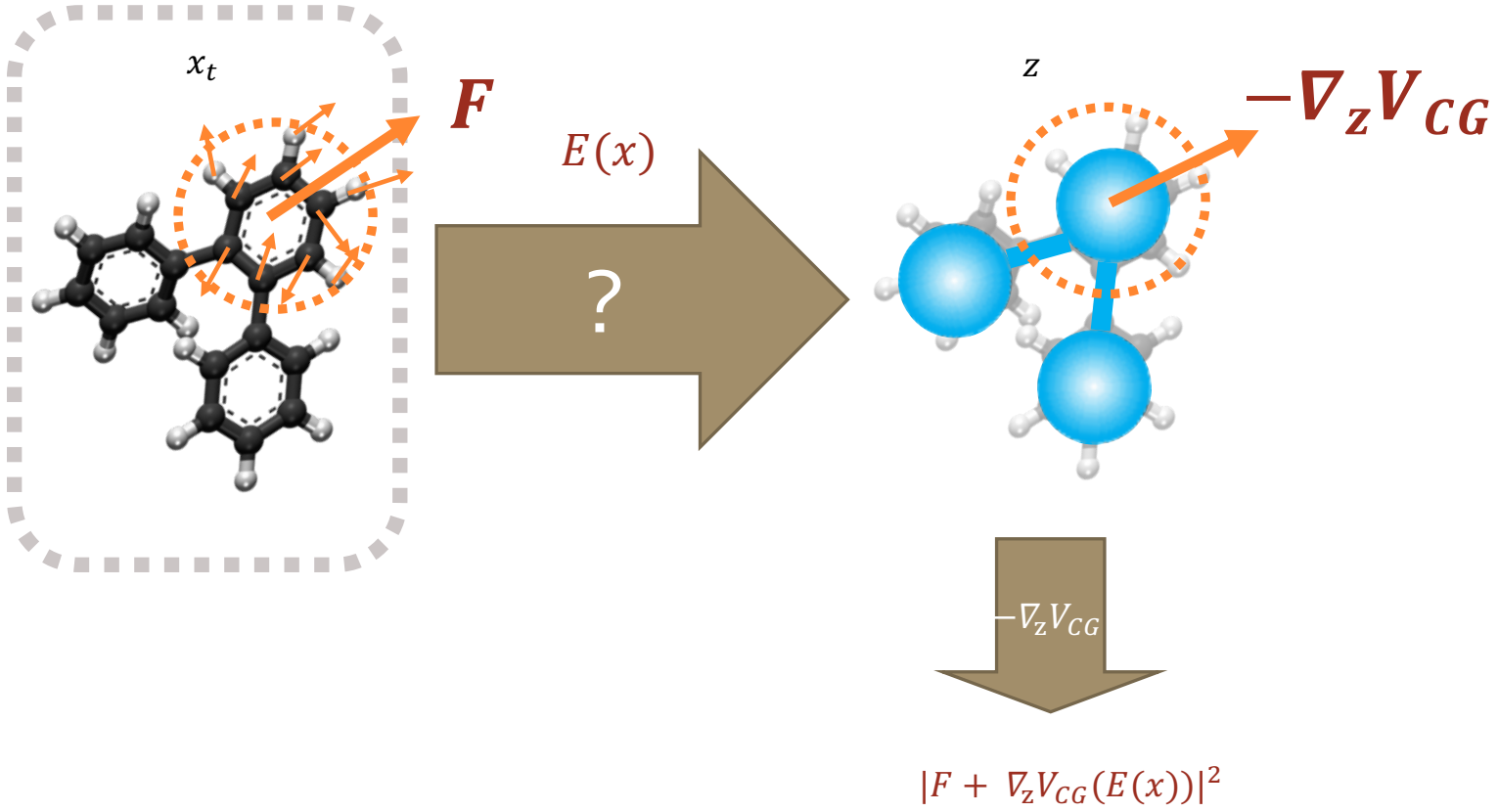
Force matching



- Force matching finds $\nabla V_{CG}(z)$ that best approximates the mean force F.
- We also want to optimize the optimization target L :

$$L = |F + \nabla_z V_{CG}(E(x))|^2$$

Towards learned CG

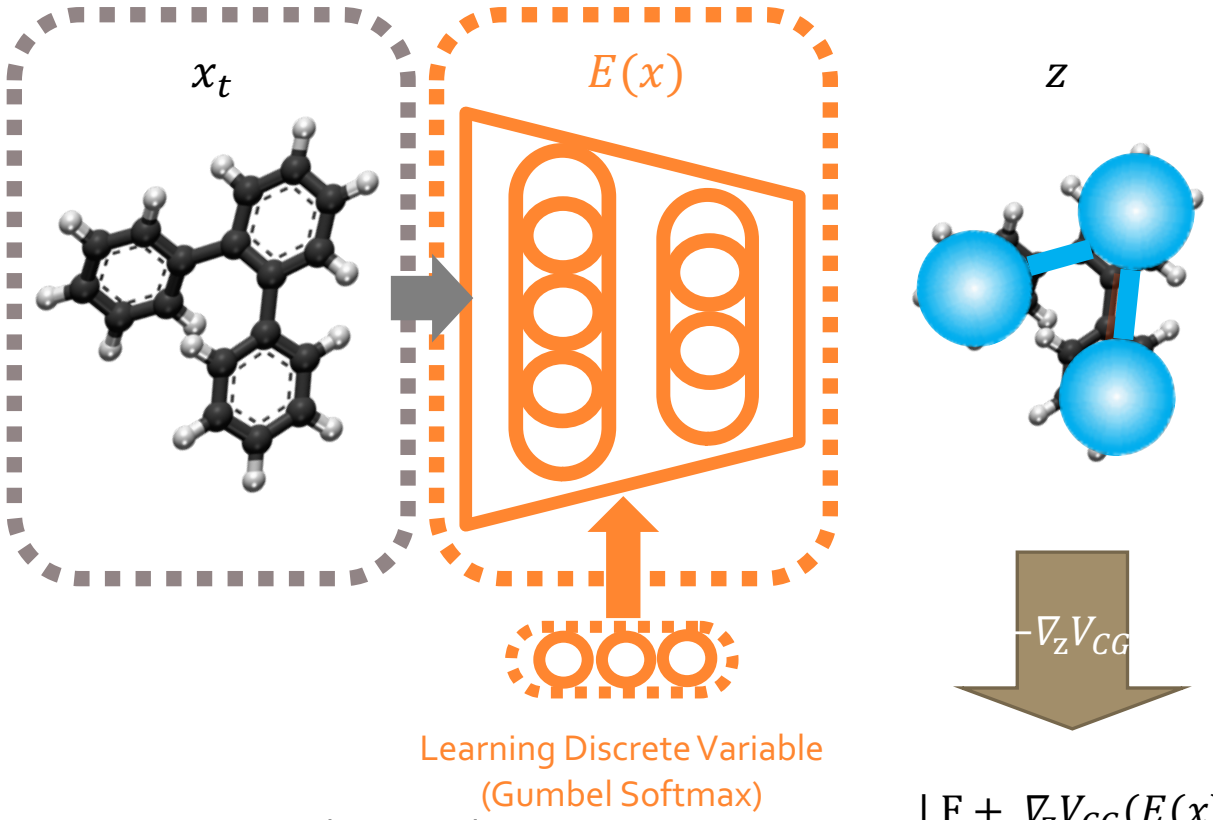


The requirements to design a physically intuitive and rigorous encoding function:

1) The coarse graining variable is a statistical averaging of the positions/motions from contributing atoms
2) Perform a dimension reduction : $z_I = E_I(x), E\colon \mathbb{R}^{3n} \rightarrow \mathbb{R}^{3N}$
3) Need to systematically redefine coarse grained mass to conserve kinetic energy
4) Each atom contributes to at most one coarse grained variable, to ensure consistency in the momentum space

W. G. Noid e al. Chem. Phys. **128**, 243116 (2008).

The encoding function



$|\mathrm{F} + \nabla_z V_{CG}(E(x))|^2$

In order to make a physically intuitive and rigorous encoding function, we propose a neural network like encoding with the following constraints

1) $\mathrm{z_I} = \mathrm{E_I(x)} = \sum_j \mathrm{E_{Ij}\ x_j}$ [1]

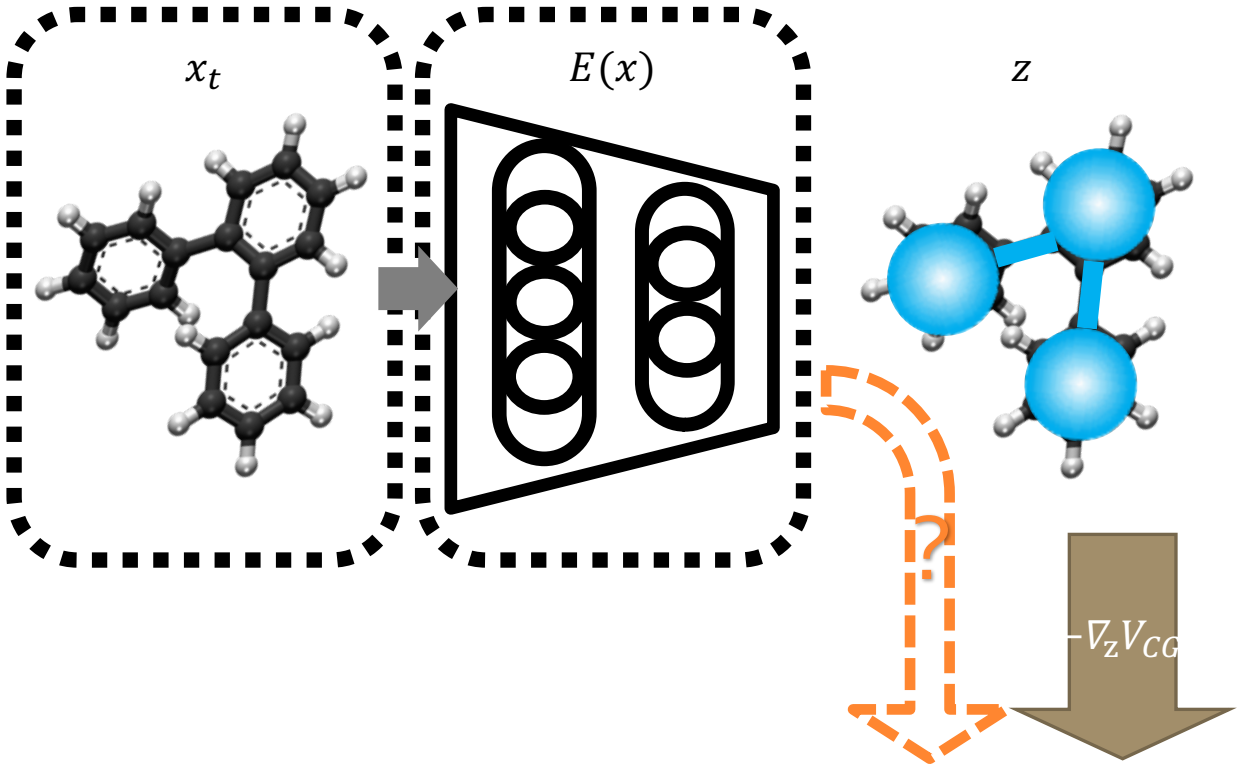
2) $\sum_j \mathrm{E_{jI}} = 1$ and $\mathrm{E_{jI}} \geq 0$

3) $\mathrm{M_I} = (\nabla\ \mathrm{E^{-1}})^{\mathrm{T}} \mathrm{M_j}\ \nabla\ \mathrm{E^{-1}}$

4) We use tricks in training neural network to enforce the learning of **discrete** coarse graining variables to ensure that each atom only contributes to one CG atom

[1] I: 1 … N coarse grained atom index
j: 1 … n atom index

“Coarse Graining” Forces

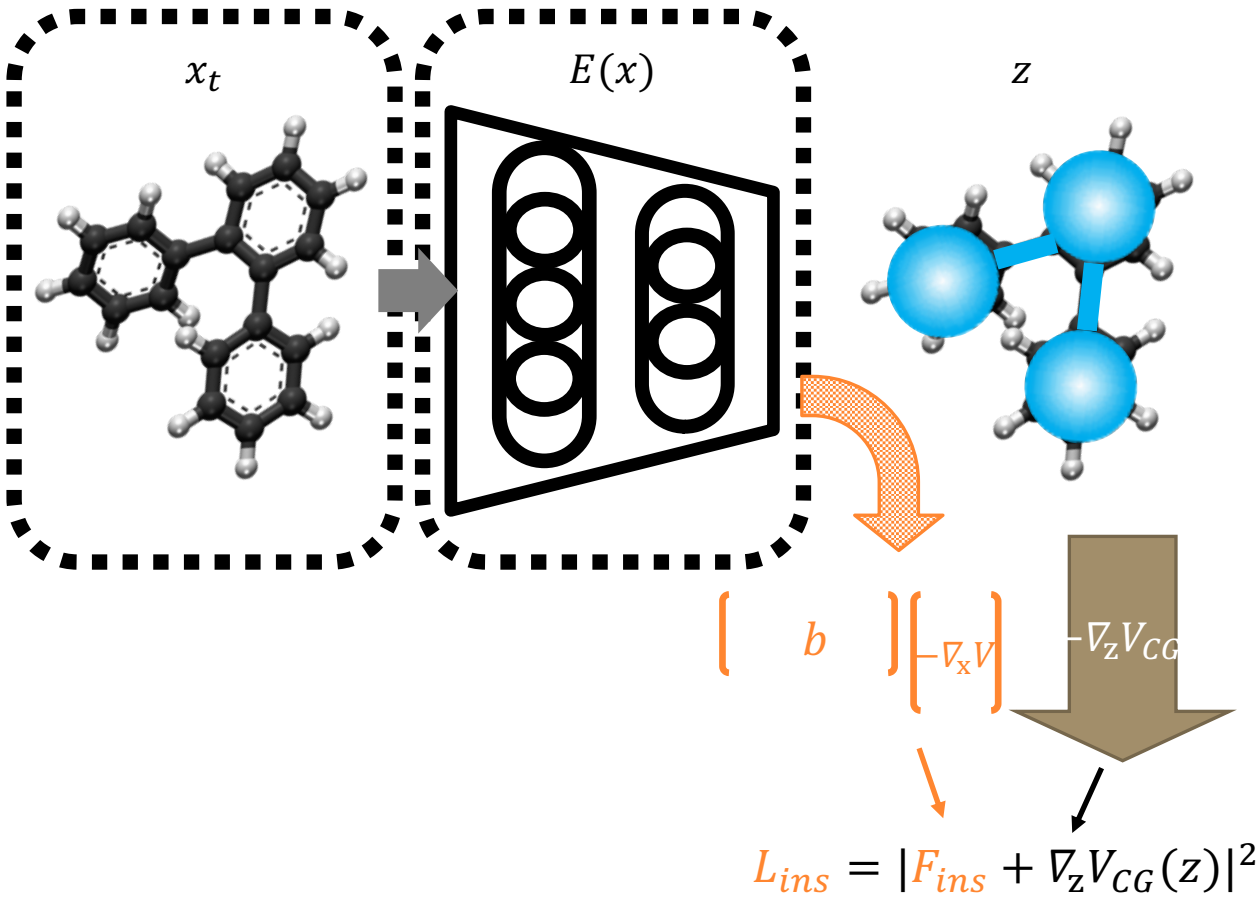


$$|\langle -\mathrm{b}\cdot\nabla V\rangle + \nabla_z\, \mathrm{V_{CG}}(\mathrm{E(x)})|^2$$

- We also need a function that variationally determines F
- $F = \langle -\mathrm{b}\cdot\ \nabla V(x)\rangle_{E(x)=z}$ where b is the **force coarse graining function**
- A consistent choice for b from statistical mechanics: $\mathrm{b} = \frac{\nabla \mathrm{E(x)}^{\mathrm{T}}}{\nabla \mathrm{E(x)}^{\mathrm{T}}\cdot\nabla \mathrm{E(x)}}$ [1]
- Computing the mean force F requires constrained dynamics.
- However, we want a one-shot optimization stack without running extra MD simulations.

[1] E. Darve, *Numerical Methods for Calculating the Potential of Mean Force.*

Stochastic Force Matching



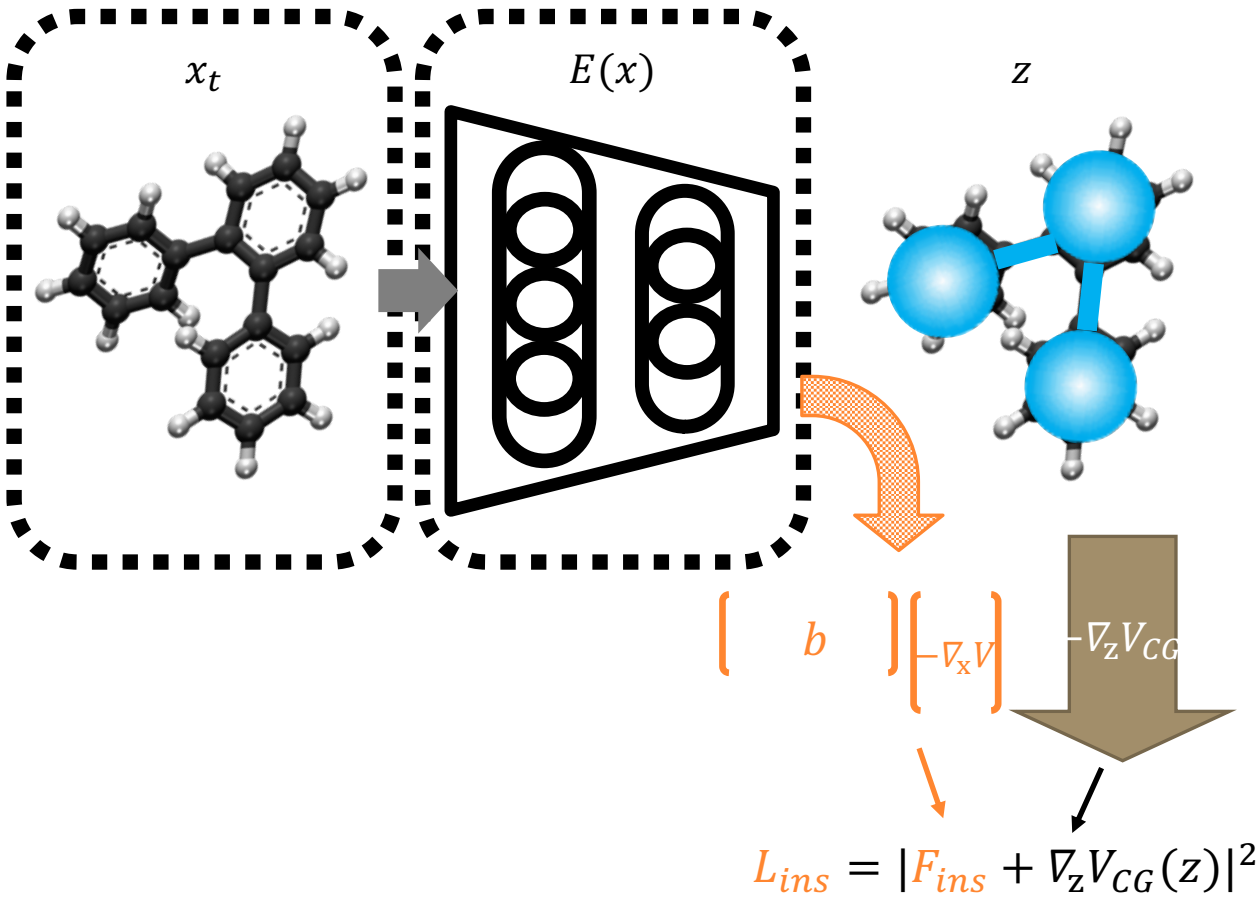
- Instead we compute the instantaneous stochastic "coarse grained" force F_{ins} [1]
- $F_{ins} = -\mathrm{b} \cdot \nabla V(x)$ (instantaneous force)
- $F = \langle F_{ins} \rangle_{E(x)=z}$ (mean force)

$$L = |F + \nabla_z V_{CG}(E(x))|^2$$

$$L_{ins} = |F_{ins} + \nabla_z V_{CG}(E(x))|^2$$

[1] L. Zhang et al. J. Chem. Phys. **149**, 034101 (2018).

Stochastic Force Matching



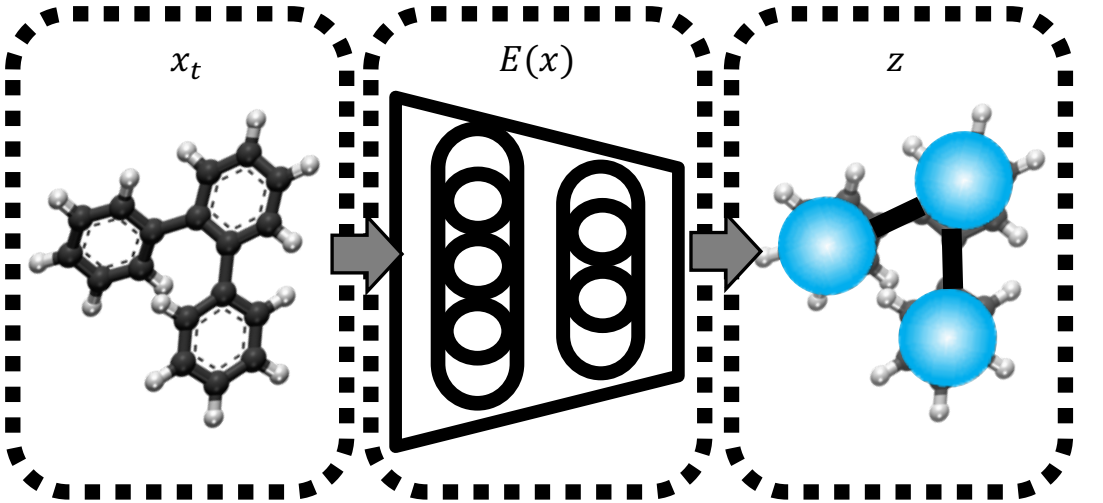
- We train by **propagating forward** the atomistic trajectories in the model to obtain the energy
- The optimization is done by **backpropagation** using auto-differentiation framework

$$L = |F + \nabla_z V_{CG}(E(x))|^2$$

$$L_{ins} = |F_{ins} + \nabla_z V_{CG}(E(x))|^2$$

[1] L. Zhang et al. J. Chem. Phys. **149**, 034101 (2018).

Supervised Coarse Graining Framework



$$\begin{bmatrix} b \end{bmatrix} \begin{bmatrix} -\nabla_x V \end{bmatrix} - \begin{bmatrix} -\nabla_z V_{CG} \end{bmatrix}$$

$$\boldsymbol{L_{ins} = |F_{ins} + \nabla_z V_{CG}(E(x))|^2}$$