MSU.AI

ЛЕКЦИЯ ОТ ПРИГЛАШЕННОГО ЭКСПЕРТА

Deep Learning

for Quantum Chemistry

Кузьма Храбров

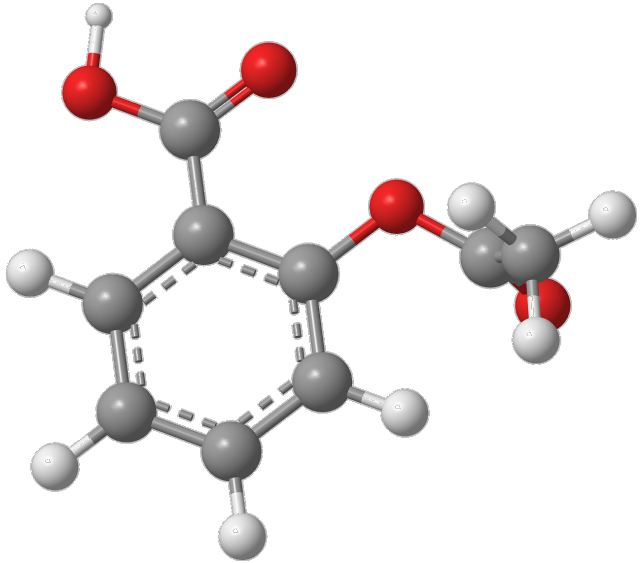
DL in Life Sciences, AIRI

Фонд интеллект

Molecular conformation

Molecular conformation
is a spatial arrangement
of atoms in a molecule

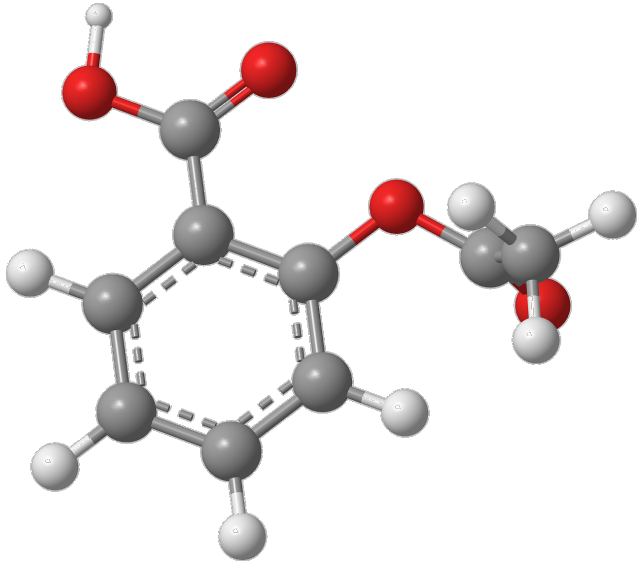
Key characteristic of
a conformation is a potential
energy, which exact form
depends on a method



Molecular conformation. Bolzman Law

The log probability of a conformation is proportional to the negative energy.

$$p_{conf} \propto e^{\frac{-E_{conf}}{kT}}$$



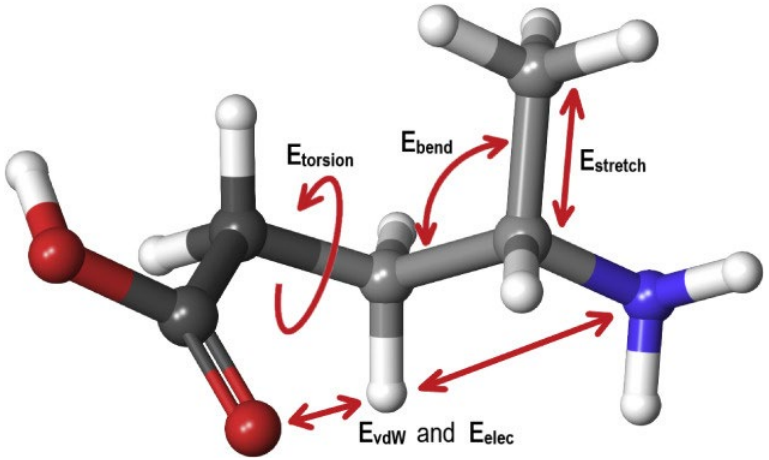
Conformational energy

Handcrafted force fields (eg.: MMFF)

$$E_{MMFF} = \sum EB_{ij} + \sum EA_{ijk} + \sum EBA_{ijk}$$

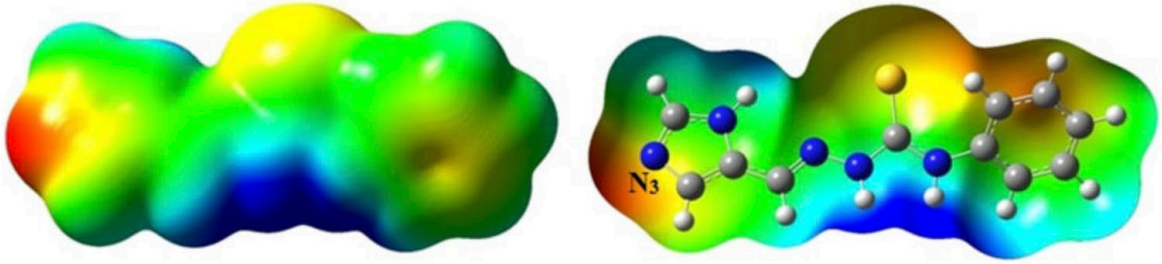
$$+ \sum EOOP_{ijkl} + \sum ET_{ijkl} + \sum EvdW_{ij} + \sum EQ_{ij}$$

$$E_{total} = E_{stretch} + E_{bend} + E_{torsion} + E_{vdW} + E_{elec}$$



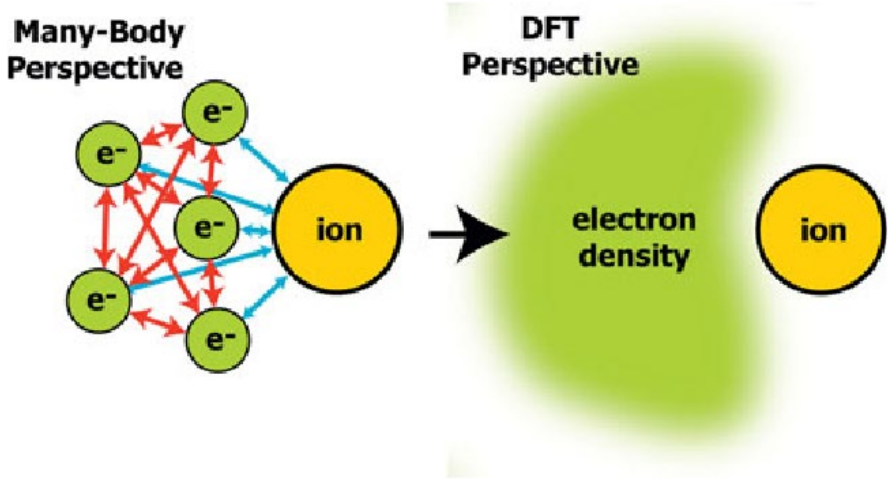
Conformational energy
Quantum force fields

$$\hat{H}\Psi = E\Psi$$



Density Functional Theory (DFT)

The highly nontrivial 3N-variable equation is simplified to 3-variable one via Hohenberg-Kohn theorems.



Density Functional Theory (DFT)

- First Hohenberg-Kohn theorem: The ground state properties of a many-electron system depend only on the electronic density n(x,y,z)
- Second Hohenberg-Kohn theorem: The correct ground state density for a system is the one that minimizes the total energy through the functional E[n(x,y,z)]

$$n(\vec{r}) = 2\sum_i \psi_i^*(\vec{r})\,\psi_i(\vec{r})$$

$$\psi_i = \sum_j C_{ij}\phi_j$$

$$\mathbf{HC} = \boldsymbol{\epsilon}\mathbf{SC}$$

$$H_{ij} = \int \phi_i^*(\mathbf{r})\hat{H}_{\mathrm{el}}\phi_j(\mathbf{r})d\mathbf{r}$$

$$S_{ij} = \int \phi_i^*(\mathbf{r})\phi_j(\mathbf{r})d\mathbf{r}$$

https://www.nature.com/articles/s41467-019-12875-2

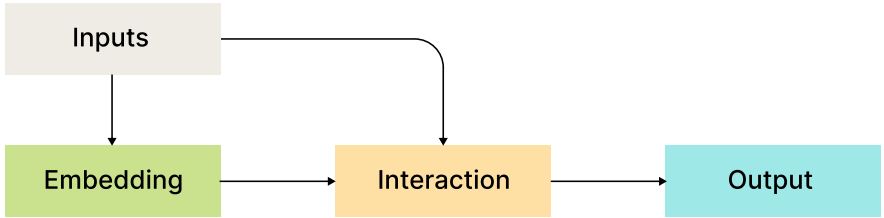
Neural force fields

There are many different neural network models to predict a conformational energy, atomic forces, or a DFT Hamiltonian for a given molecular conformation.

Basic concept

In general, models include three main units:

1. Embedding block
2. Interaction
3. Output



Datasets

Starting with the SchNet vast majority of proposed models are tested in a very simple settings.

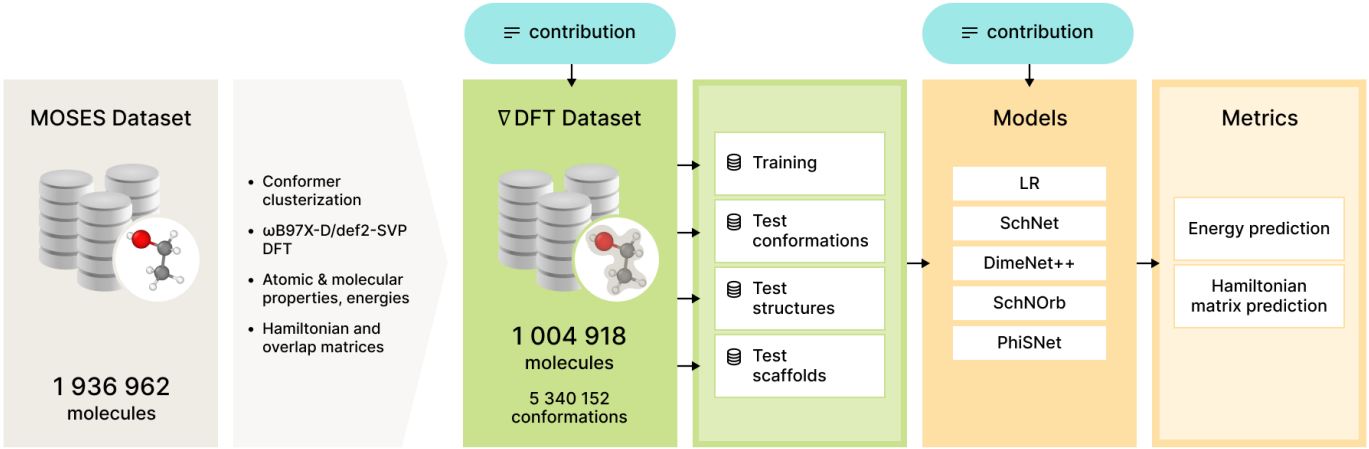
QM9
consists of ≈130k organic molecules with up to 9 heavy atoms of the types {C, O, N, F}

MD17
contains trajectories from molecular dynamics simulations of 8 molecules

ISO17
contains short MD trajectories for 129 isomers of C7O2H10

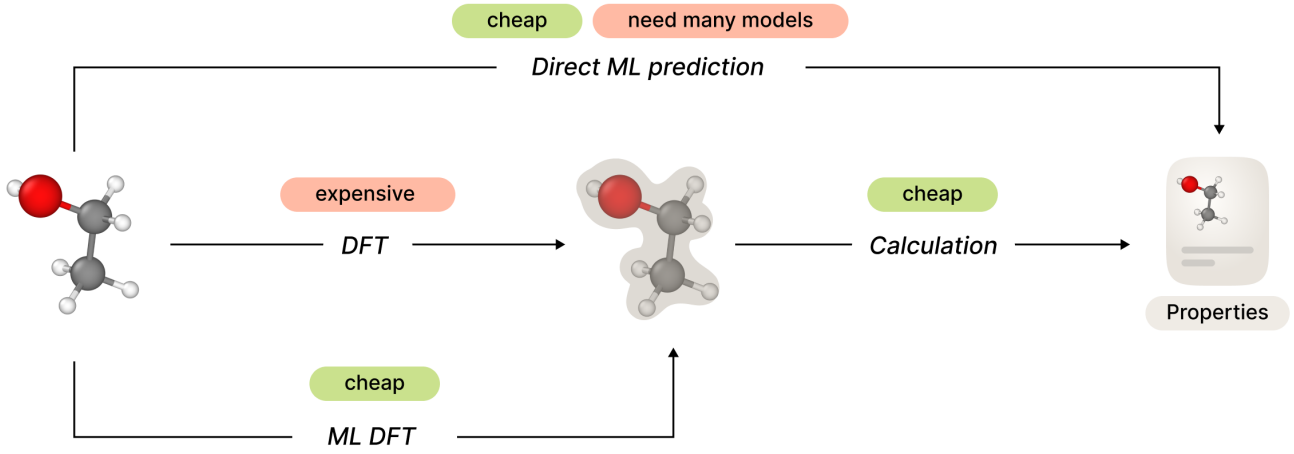
nablaDFT dataset and benchmark

To avoid an overfitting in toy settings we published our own dataset and benchmarked several recent models.

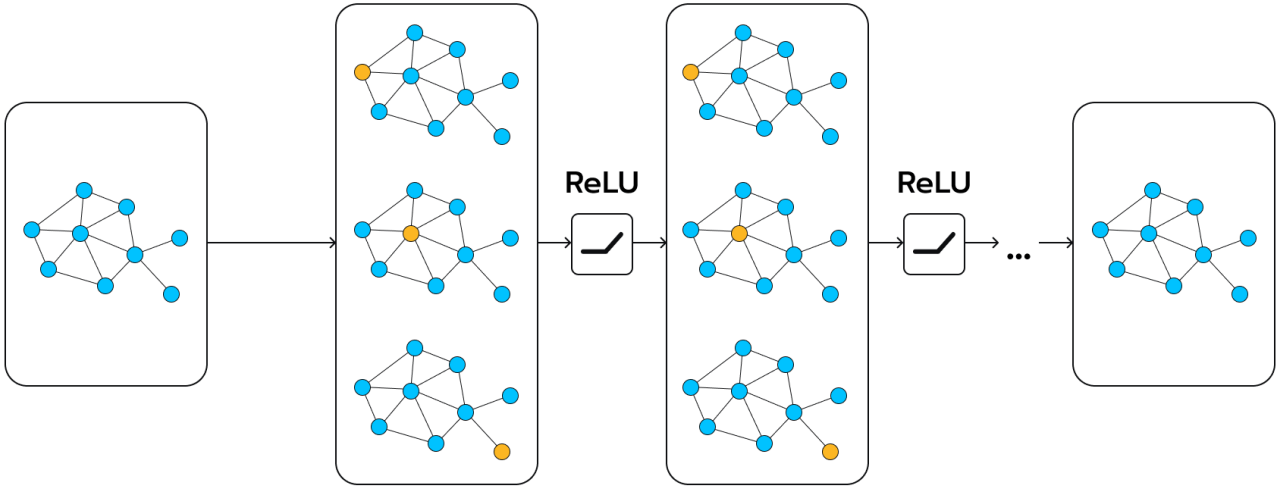


github.com/AIRI-Institute/nablaDFT

Machine Learning Framework



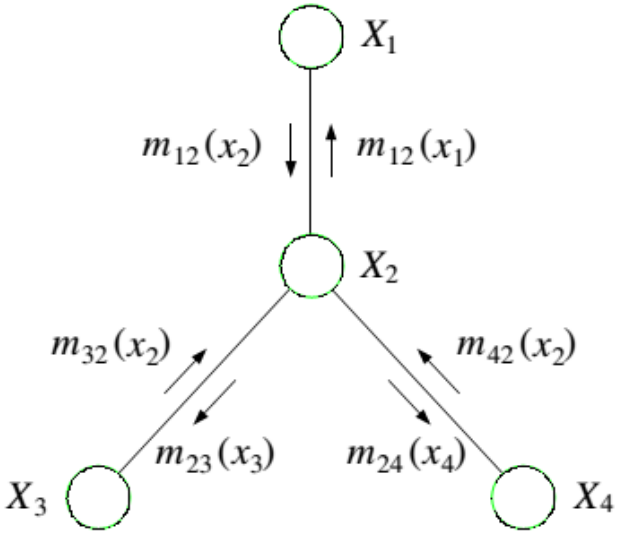
Graph Convolution Nets



Message passing

Message passing is an algorithm for inference on graphical models.

At every step of the message passing, every node has a state, which updates during the process. Each node send a message to all its neighbors computed as function of the node and neighbor states.



$$X_v^{(k)} = \sigma\left(W_0 X_v^{(k-1)} + W_1 \sum_{u \epsilon N(v)} X_u^{(k-1)} + b\right)$$

Graph convolutional networks

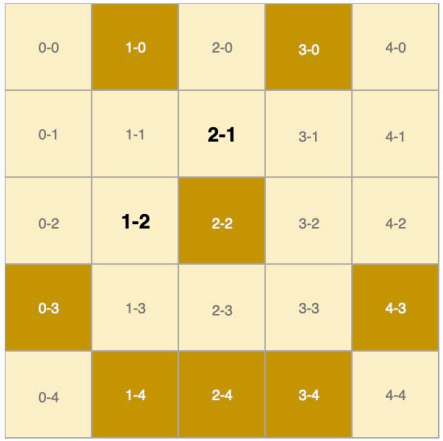
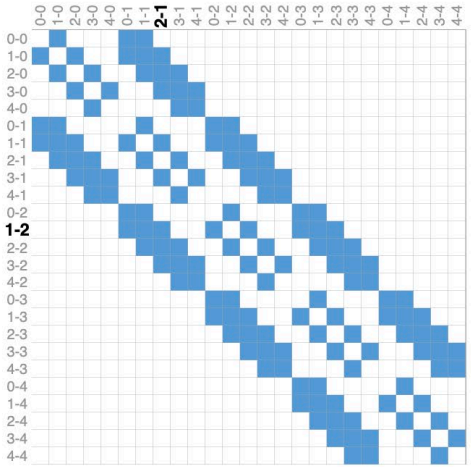
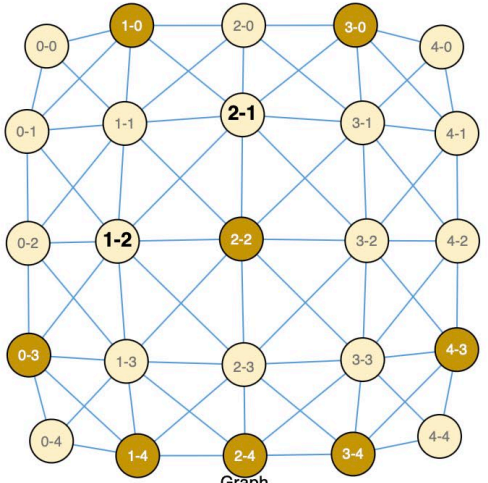


Image Pixels



Adjacency Matrix

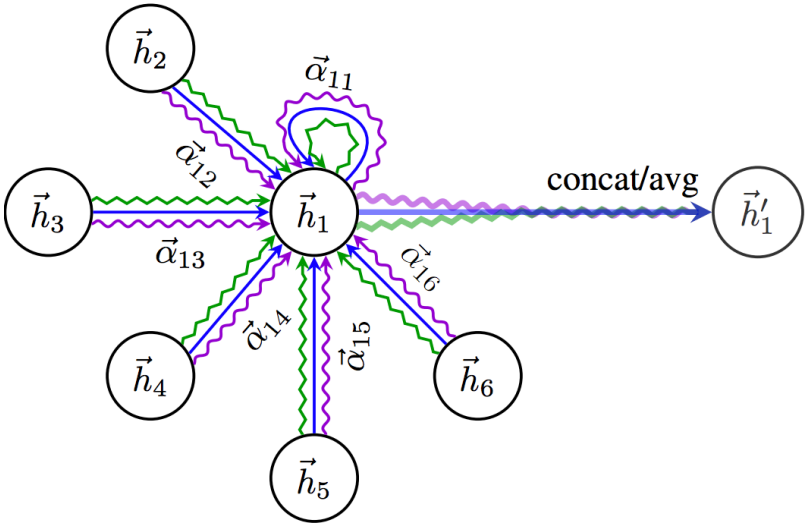


Graph

Message passing

Message passing is an algorithm for inference on graphical models.

At every step of the message passing, every node has a state, which updates during the process. Each node send a message to all its neighbors computed as function of the node and neighbor states.



SchNet

SchNet is trained to predict a conformational energy together with atomic forces given desired force field.

Key features

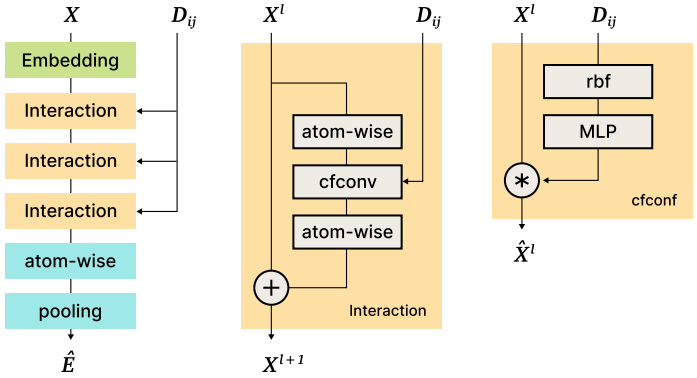
→ E(3) invariant

→ Intermolecular distances embedded with RBFs

→ Atomic forces are computed in the form of an energy gradient

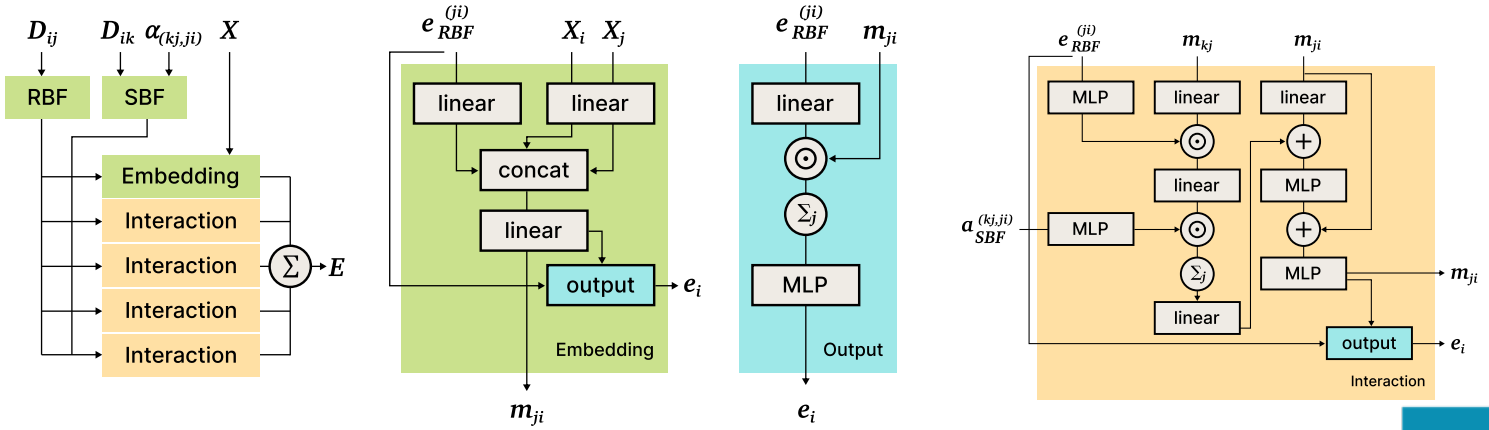
$$\mathcal{L}\left[\left(\hat{E}\right),(E,\boldsymbol{F})\right]=\left\|E-\hat{E}\right\|^2+\frac{\rho}{n}\sum_{i=1}^{n}\left\|\boldsymbol{F}_i-\left(-\frac{\partial\hat{E}}{\partial r_i}\right)\right\|^2$$

$$\widehat{\boldsymbol{F}}_{\boldsymbol{i}}(Z_1,\ldots,Z_n,r_1,\ldots,r_n)=-\frac{\partial\hat{E}}{\partial r_i}(Z_1,\ldots,Z_n,r_1,\ldots,r_n)$$



DimeNet++

DimeNet uses angular information during the message passing in addition to pairwise distances. (https://arxiv.org/abs/2011.14115)



$$\boldsymbol{m}_{ji}^{(l+1)} = f_{update}\left(\boldsymbol{m}_{ji}^{(l)}, \sum_{k \in N_j \setminus \{i\}} f_{int}\left(\boldsymbol{m}_{kj}^{(l)}, \boldsymbol{e}_{RBF}^{(ji)}, \boldsymbol{a}_{SBF}^{(kj,ji)}\right)\right)$$

SchNOrb

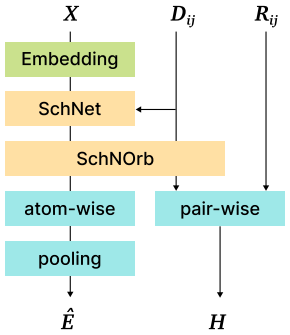
In addition to conformational energy prediction, SchNOrb model is trained to predict DFT Hamiltonian matrix within a given basis set.

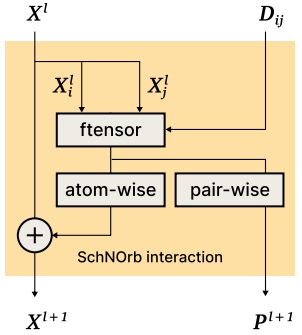
Key features

→ E(3) invariant for energy only

→ Combined loss for energy and Hamiltonian

→ Atomic forces are computed in the form of an energy gradient





$$\mathcal{L}\left[\left(\hat{E},\hat{\boldsymbol{H}}\right),\left(E,\boldsymbol{H},\boldsymbol{F}\right)\right]=\left\|E-\hat{E}\right\|^2+\left\|\boldsymbol{H}-\hat{\boldsymbol{H}}\right\|^2+\frac{\rho}{n}\sum_{i=1}^{n}\left\|\boldsymbol{F}_i-\left(-\frac{\partial\hat{E}}{\partial r_i}\right)\right\|^2$$

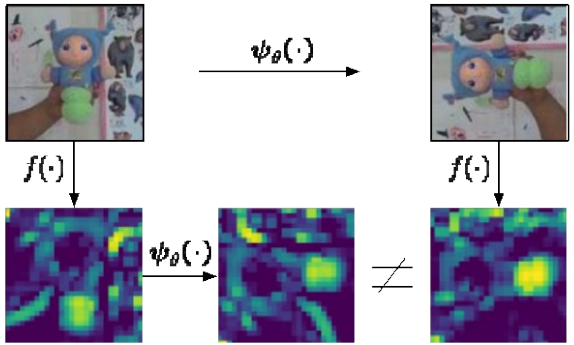
www.nature.com/articles/s41467-019-12875-2

PhiSNet

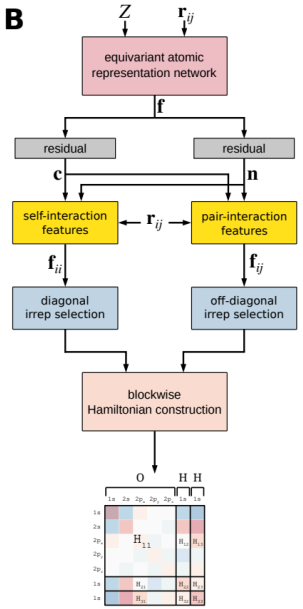
Key features

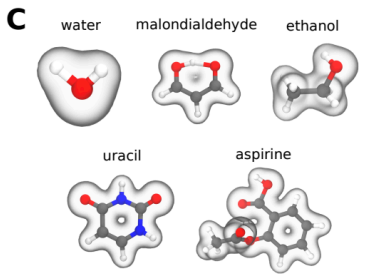
→ E(3)-equivariant features and predictions

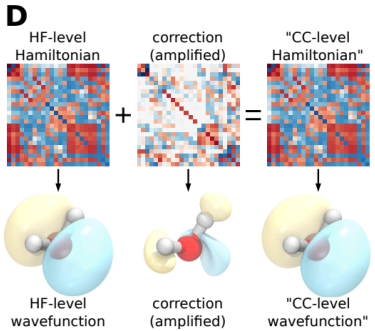
→ Combined loss for energy and Hamiltonian



$$f(D^{\mathcal{X}}(g)\mathbf{x}) = D^{\mathcal{Y}}(g)f(\mathbf{x}).$$

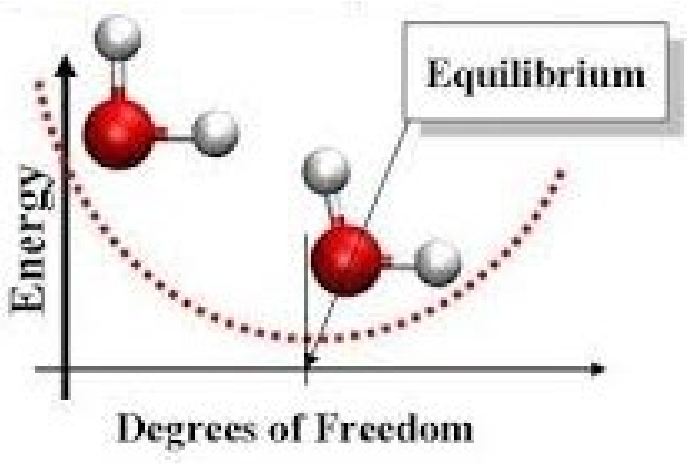






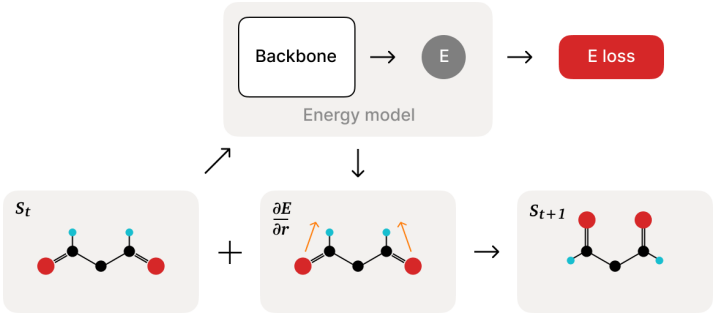
openreview.net/forum?id=auGY2UQfhSu

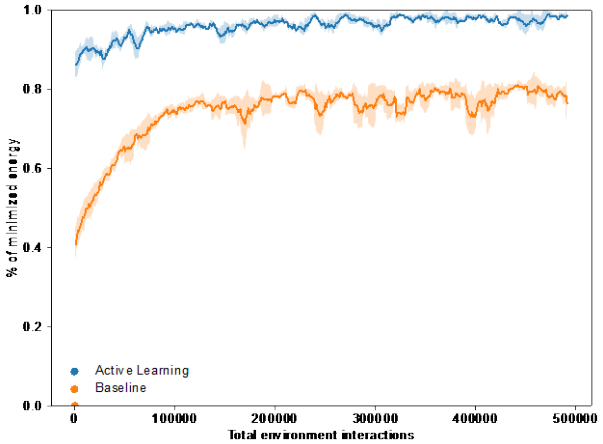
Geometry optimization



Molecular geometry optimization with NNs.

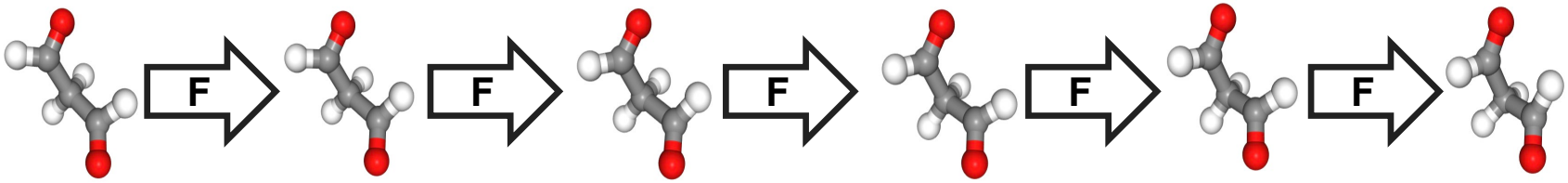
Based on our research of quantum properties of small molecules we develop a model for molecular geometry optimization with the quality on par with traditional physical simulators.



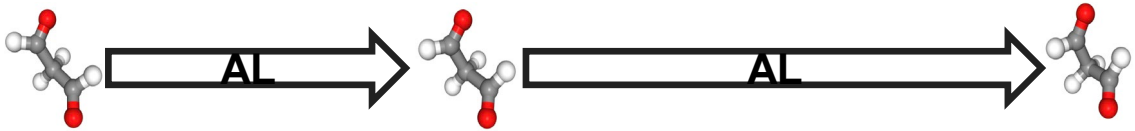


Idea

Classical geometry optimization pipeline is an analog of the gradient descent algorithm



The idea is to replace the the gradient descent with Active learning or Reinforcement learning

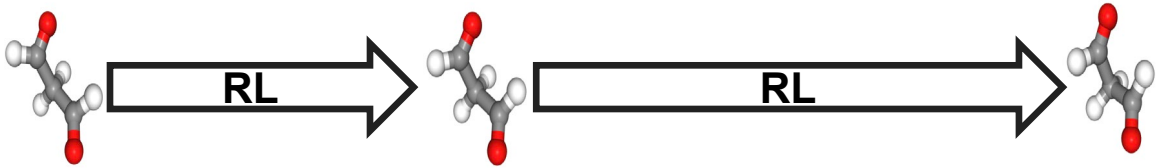


Conformation optimization as RL problem

States: $s \in \mathbb{R}^{3\times n_{atoms}}$ — atoms' 3D-coordinates.

Actions: $a \in \mathbb{R}^{3\times n_{atoms}}$ — coordinates shifts.

Reward: $r = -(E_{s'} - E_s)$ - $\alpha\|F_{s'} - F_s\|$, so the agent's goal is to minimize the energy and forces difference. Energies and forces for the initial and final states are taken from physical simulators or baseline models.
Transition function: $P(s'|s,a) = \delta(s+a)$



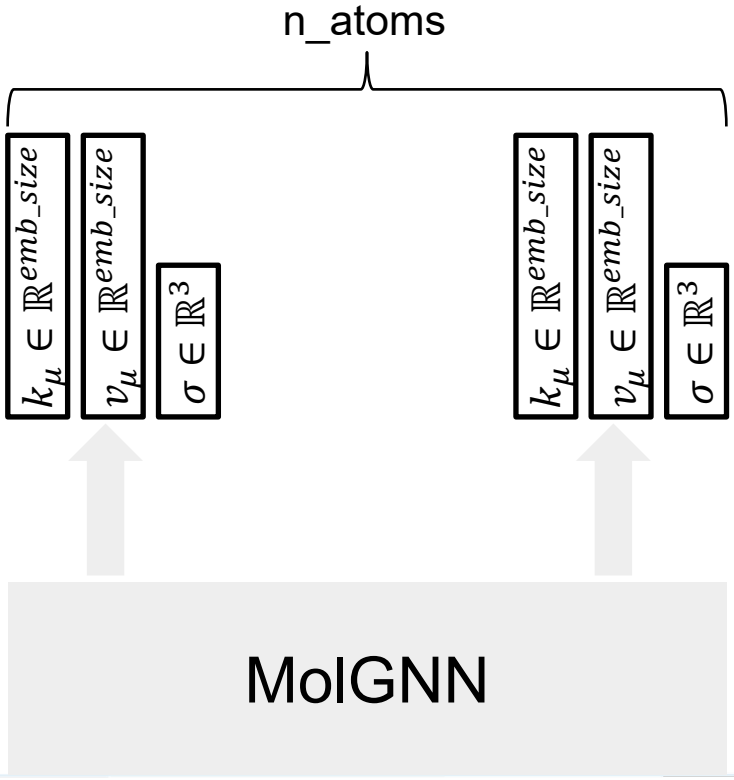
RL-agent architecture

We use one of molecular GNN as a backbone, and compute pairwise shifts in relative coordinates, so the transition becomes SE(3)-equivariant.

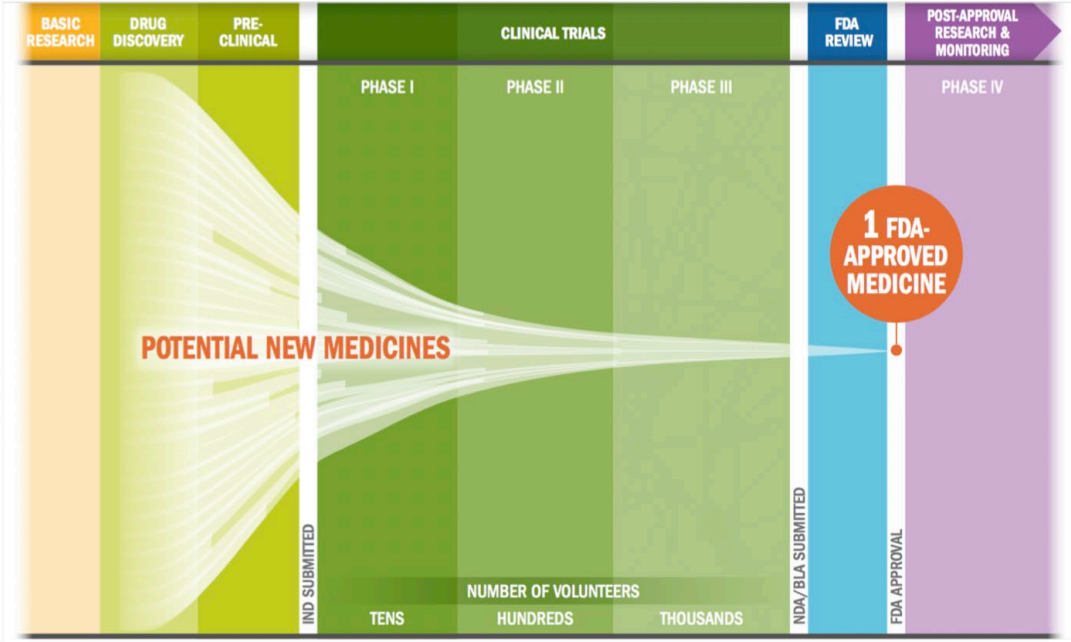
Shifts are computed the following way:

1. At first we calculate the matrix of pairwise directions:
   $\boldsymbol{P} \in \mathbb{R}^{n_{atoms} \times n_{atoms}}$, where $\boldsymbol{P}_{i,j} = \frac{\boldsymbol{r}_i - \boldsymbol{r}_j}{\|\boldsymbol{r}_i - \boldsymbol{r}_j\|}$
2. Afterwards the coefficients are predicted :
   $\boldsymbol{A}_{shift} = \boldsymbol{K}_\mu \boldsymbol{V}_\mu^T$, where $\boldsymbol{K}_\mu$ and $\boldsymbol{V}_\mu \in \mathbb{R}^{n\_atoms \times emb\_size}$
3. Finally the shifts are calculated in the following way:
   $a \sim tanh\left(Normal\left(\boldsymbol{A}_{shift} * \boldsymbol{P}, \sigma\right)\right) * action\_scale$, during training
   $a \sim tanh\left(\boldsymbol{A}_{shift} * \boldsymbol{P}\right) * action\_scale$, during inference.

RL-agent architecture



Drug discovery



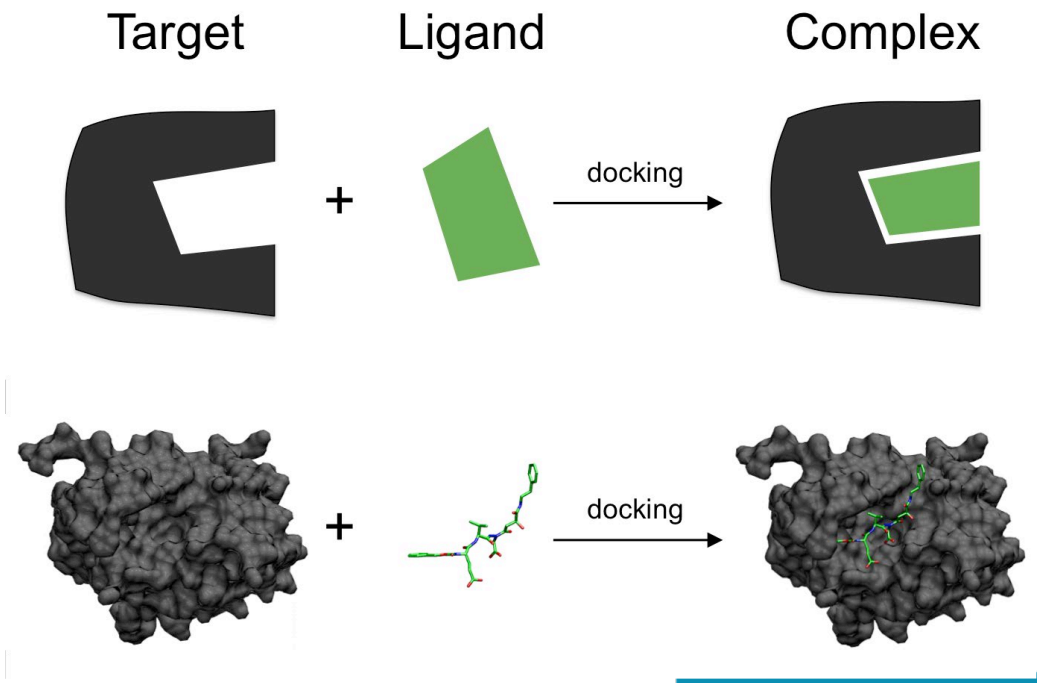
DL in drug discovery

- The traditional process of new drug discovery is notoriously **long** and **expensive**.
- Key goal is to construct new chemical structures possessing **desired properties**.
- Traditional virtual screening approach **limited** to the databases of already known drug candidates and is not capable of designing novel drugs
- Most of existing DL approaches **do not take** into consideration the **target protein**

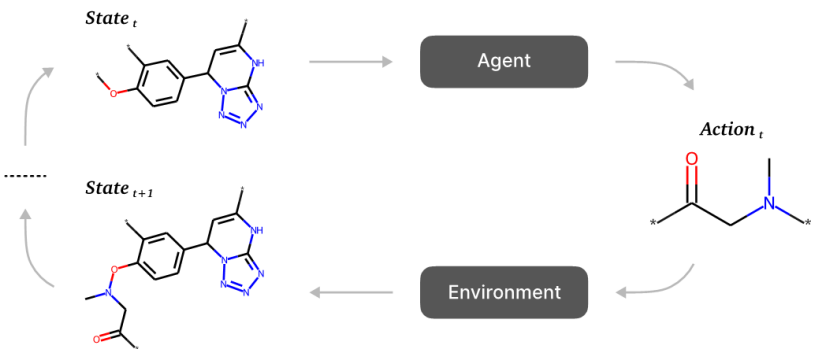
DL problem

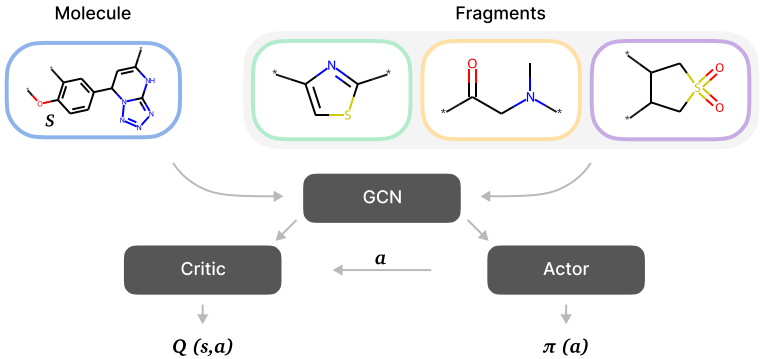
Design a framework for generation of novel molecules with desired objectives.

One possible objective is Docking score, which is an approximate prediction of the binding affinity.



Theory and Algorithms

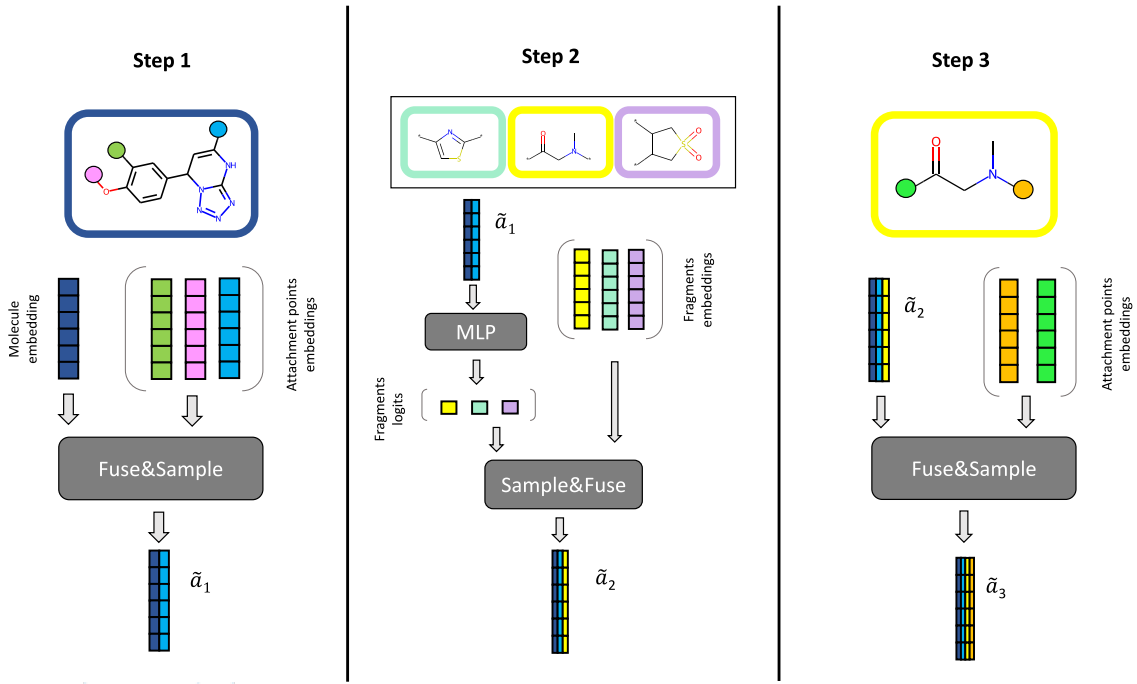




O - objective to be optimized

$$r_{t+1} = O(s_{t+1}) - O(s_t) \text{ or } r_{t+1} = \begin{cases} 0 & \text{if } s_{t+1} \text{ is non-terminal;} \\ O(s_{t+1}) & \text{if } s_{t+1} \text{ is terminal.} \end{cases}$$

Theory and Algorithms



@AIRI_Research_Institute

- airi_research_institute
- AIRI Institute
- AIRI Institute
- AIRI_inst
- artificial-intelligence-research-institute