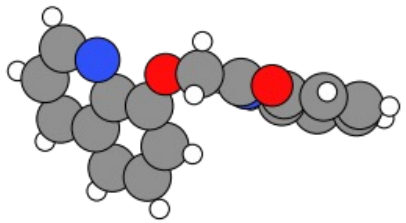
# Gradual Optimization Learning for Conformational Energy Minimization

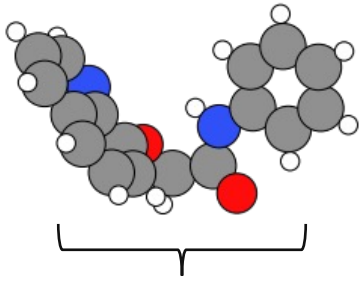
Artem Tsypin et al.

# Low-energy conformations

- Finding low-energy conformations is crucial in computer-aided drug design
- Important molecular properties that define a molecule's medicinal potential are estimated in low-energy conformations
- The transition of a molecule from any given conformation to a low-energy conformation is known as relaxation or geometry optimization



geometry optimization



Low-energy conformation

physical simulator,
neural network, ...

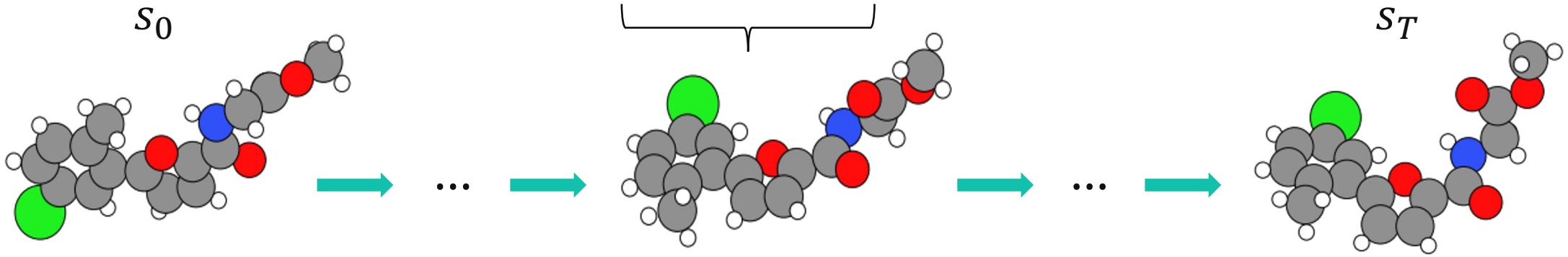
Energy of HOMO
Energy of LUMO
Gap
Dipole moment
Internal energy
Heat capacity
...

# Traditional methods for geometry optimization

- We denote the $t$-th conformation in the optimization trajectory as $s_t = \{X_t, z\}$, where $X_t$ is the matrix of atoms' coordinates, and $z$ is the vector of atomic numbers
- Traditional methods iteratively optimize the geometry using interatomic forces $F_{s_t}$ as anti-gradients
- To obtain reasonably accurate interatomic forces, we employ DFT-based physical simulators[1]
- Can scale as $O(N^4)$, where $N$ is the number of electrons in the system

Computationally expensive!

$$s_{t+1} = s_t + \alpha Opt(F_{s_t}^{DFT})$$




[1] Smith, Daniel GA, et al. "PSI4 1.4: Open-source software for high-throughput quantum chemistry." The Journal of chemical physics 152.18 (2020).



AIRI

# Neural network potentials for geometry optimization

- A Neural Network Potential (**NNP**) is trained to predict energy based on conformation:

$$E_{s_t}^{NNP} = f(s_t; \theta)$$

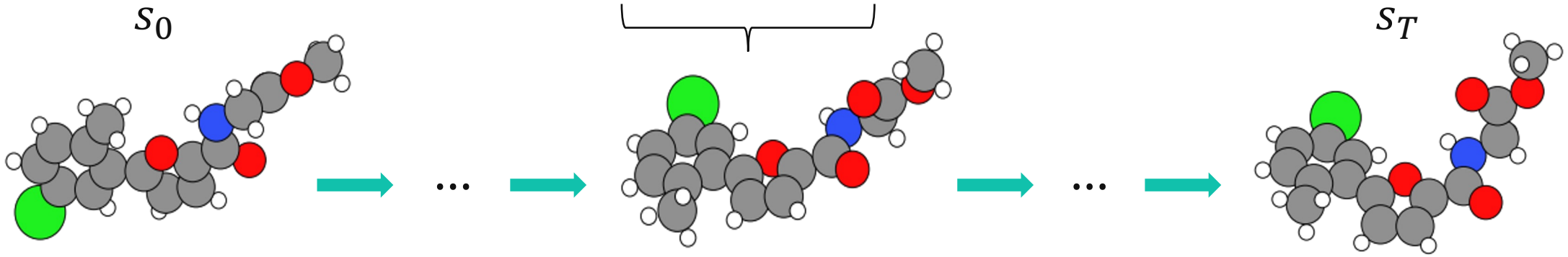
- To predict the forces we take the gradient of the energy w.r.t. atoms' coordinates[1, 2] :

$$F_{s_t}^{NNP} = -\frac{\partial f(s_t;\, \theta)}{\partial\, X_t}$$

- Iterative optimization with NNP is ~2000 times faster in terms of wall-time
- Requires a lot of data!

Requires a lot of training data!

$$s_{t+1} = s_t + \alpha Opt(F_{s_t}^{NNP})$$

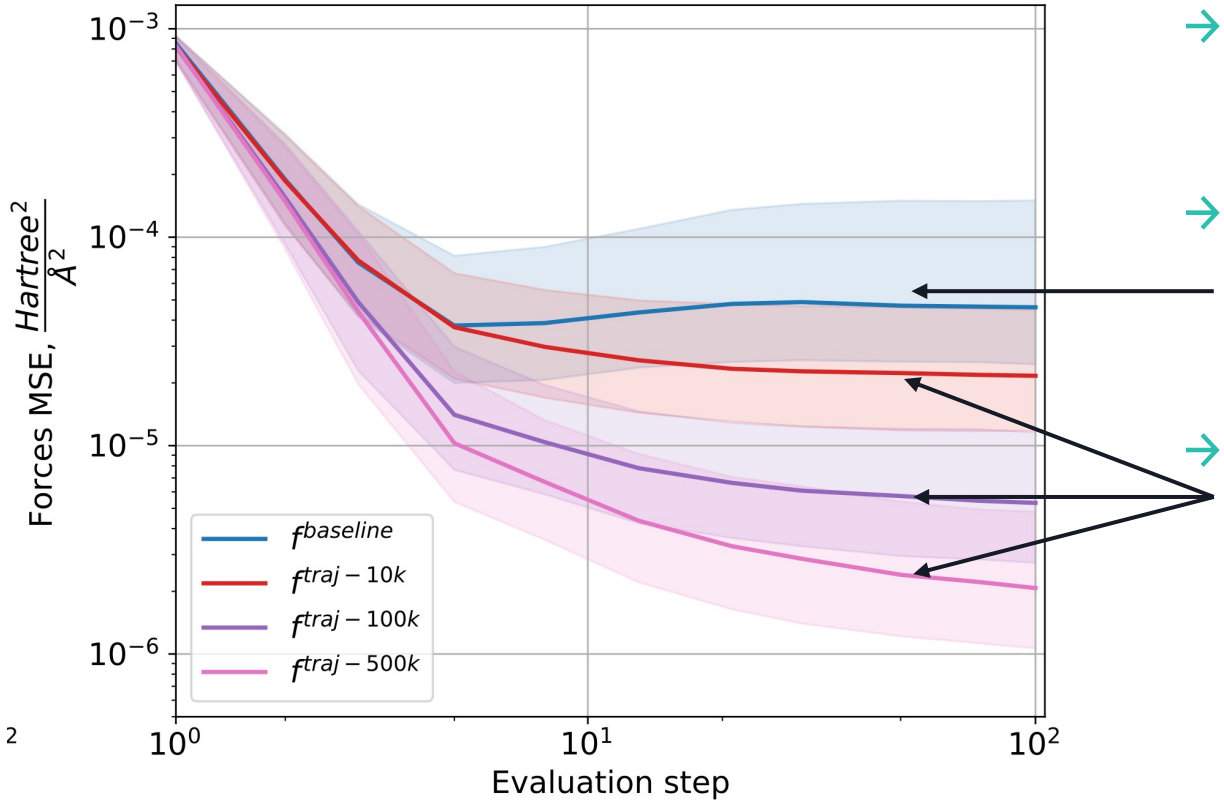



[1] Schütt, Kristof T., et al. "Schnet–a deep learning architecture for molecules and materials." The Journal of Chemical Physics 148.24 (2018).
[2] Schütt, Kristof, Oliver Unke, and Michael Gastegger. "Equivariant message passing for the prediction of tensorial properties and molecular spectra." International Conference on Machine Learning. PMLR, 2021.



AIRI

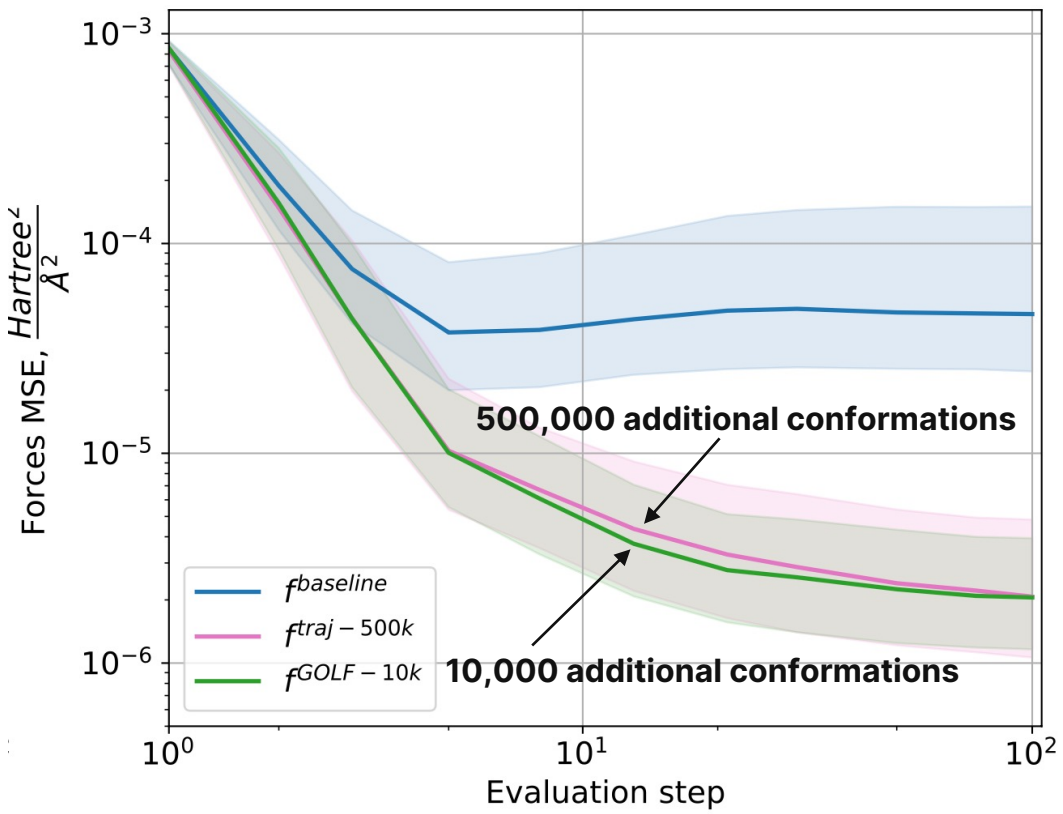
# Training NNPs for the task of geometry optimization




- NNPs trained on publicly available datasets[1, 2] suffer from a distribution shift when used in the optimization task
- To show this, we run iterative optimization with NNP and evaluate forces predicted at each step with DFT-based oracle. The prediction error increases throughout the relaxation
- To alleviate the distribution shift, we extend the training dataset with ground-truth optimization trajectories obtained with DFT-based oracle. The prediction error gradually decreases with the amount of additional training data


[1] Khrabrov, Kuzma, et al. "nabladft: Large-scale conformational energy and hamiltonian prediction benchmark and dataset." Physical Chemistry Chemical Physics 24.42 (2022): 25853-25863.
[2] Eastman, Peter, et al. "Spice, a dataset of drug-like molecules and peptides for training machine learning potentials." Scientific Data 10.1 (2023): 11.

# Motivation and the goal

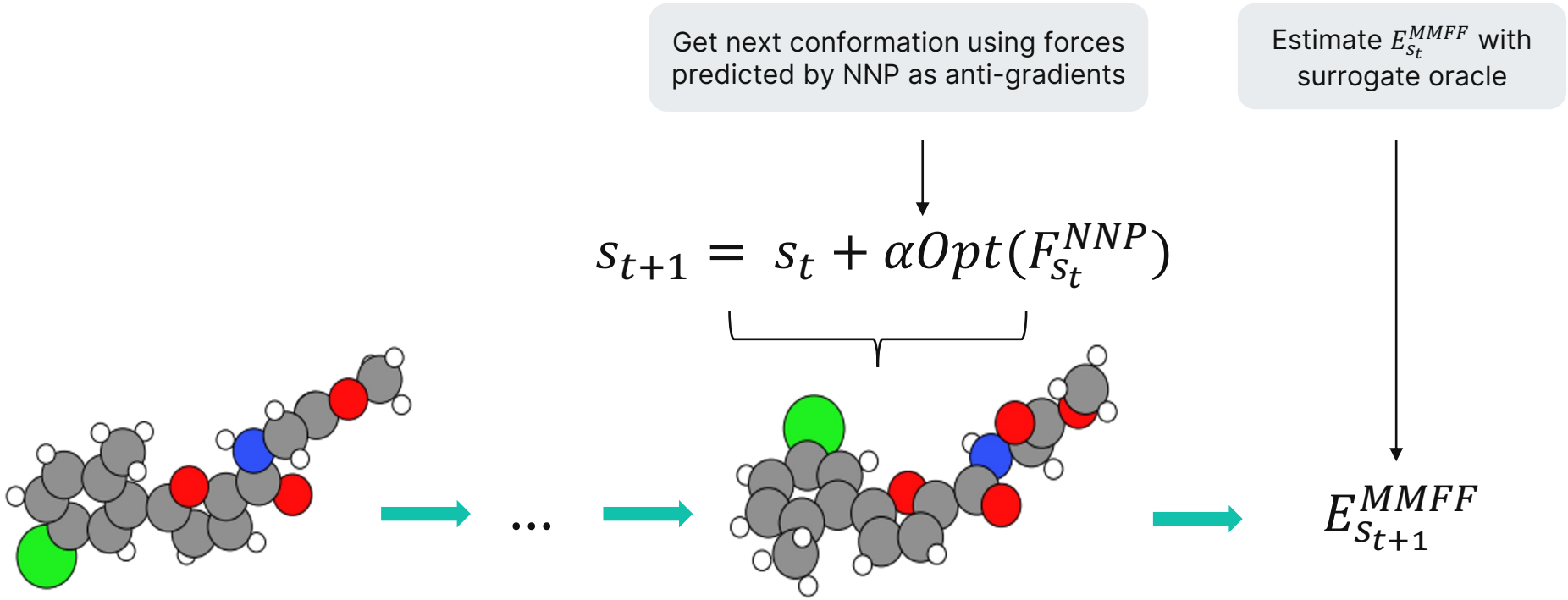



- We show that it requires approximately 500,000 additional conformations to reach optimization quality comparable with the DFT-based oracle
- For the molecules from the nablaDFT dataset and our selected physical simulator, this amounts to about **9 CPU-years of compute**

Our primary goal is to reduce the amount of additional data while maintaining optimization quality comparable to the DFT-based physical simulator.

# GOLF

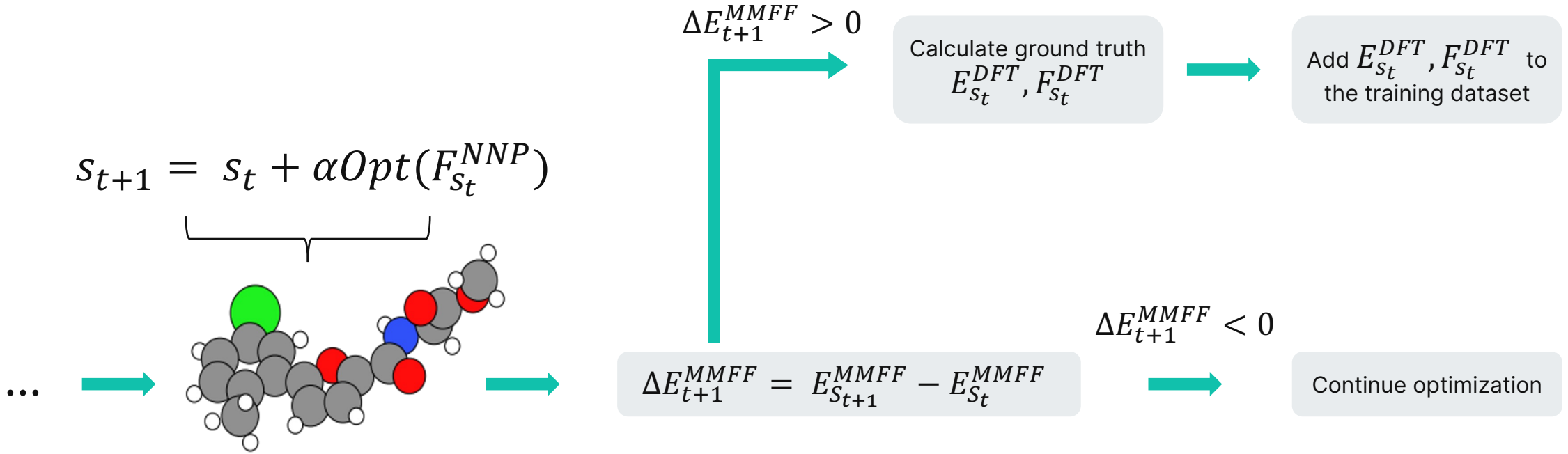
We use the NNP to perform iterative optimization.
After every step, we estimate the energy in $s_{t+1}$ with the **surrogate oracle** (RDKit's[1] MMFF in our work).




[1] RDKit: Open-source cheminformatics. https://www.rdkit.org

# GOLF

Then we estimate whether the energy has decreased. If the energy decreased, we continue the optimization; otherwise, we consider the NNP's prediction of interatomic forces incorrect and add the conformation from the previous step to the training dataset.



# Results

- We test the algorithm on a subset of nablaDFT[1] dataset $\mathcal{D}^{test}$ (~20000 conformations for ~10000 molecules)
- We use the following metrics:

$$E^{residual} = \frac{1}{|\mathcal{D}^{test}|} \sum_{s \in \mathcal{D}^{test}} E_{s_T}^{DFT} - E_{optimal}^{DFT}$$

$$pct_{success} = \frac{1}{|\mathcal{D}^{test}|} \sum_{s \in \mathcal{D}^{test}} I\left[\left(E_{s_T}^{DFT} - E_{optimal}^{DFT}\right) < 1\right]$$

Chemical precison, kcal/mol

- Training NNPs with GOLF reduces the amount of additional conformations required to match the optimization quality of DFT from 500,000 to 10,000

| | $\overline{E^{\text{res}}}_T$(kcal/mol) ↓ | $\overline{\text{pct}}_T$(%) ↑ |
|---|---|---|
| $f^{\text{baseline}}$ | 8.6 | 8.2 |
| $f^{\text{traj-500k}}$ | 0.5 | 73.4 |
| $f^{\text{GOLF-10k}}$ | 0.5 | 77.3 |

AIRI

airi.net

airi_research_institute

AIRI Institute

AIRI Institute

AIRI_inst

artificial-intelligence-research-institute

Telegram

Artificial Intelligence
Research Institute