

metatensor

Training arbitrary models

What we are going to do today?

1. Using *metatrain* to train a *GAP* and *PET* model on ethanol
2. Run MD with LAMMPS
3. Quantify the uncertainty of the trained models

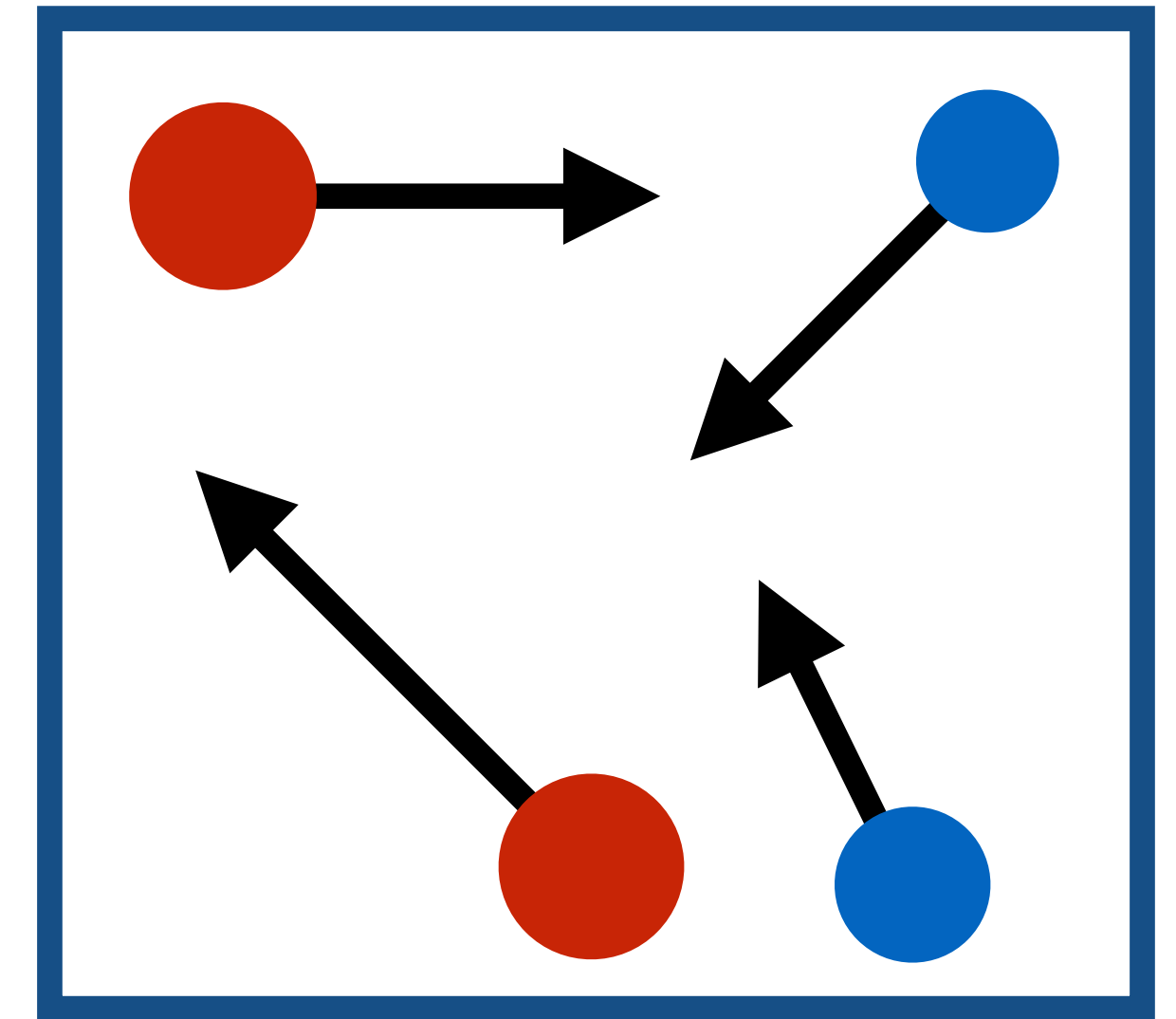
Ethanol dataset

- 1000 ethanol structures taken from the rmd17 dataset
- rmd17: revised molecular dynamics 2017 introduced in Chimela et. al. Sci. Adv. 2017
- Energies and forces for molecular dynamics trajectories of eight organic molecules. Level of theory DFT: PBE+vdW-TS.
- Structure sampled from ab initio molecular dynamics (AIMD)

In short: DO NOT train a model on more than 1000 samples from this dataset. Data already published with 50K samples on the original MD17 dataset should be considered meaningless due to this fact and due to the noise in the original data.

Reminder: Molecular Dynamics

```
call init  
t = 0  
do while t < tmax  
  call force()  
  call integrate()  
  t=t+delta  
enddo
```



from ML

$$\mathbf{F}_i = -\frac{\partial}{\partial \mathbf{r}_i} U(\mathbf{r}_1, \dots, \mathbf{r}_i, \mathbf{r}_N)$$

integrator

$$\mathbf{r}_i(t + \Delta t) \approx 2\mathbf{r}_i(t) - \mathbf{r}_i(t - \Delta t) + \frac{1}{m} \mathbf{F}_i(t) \Delta t^2$$

Reminder: Atomistic ML Architectures

Model	Representation / Fingerprint	Training	Body Order	Elements Supported	Equivariance	Scalability	Summary
Gaussian Approximation Potential (GAP)	Fixed mathematical descriptors (e.g., SOAP)	Linear algebra (sparse regression)	Fixed	Few (< 6 recommended)	✓ Equivariant	⚠ Limited (small datasets)	Good for small datasets, precise with limited data
Point Edge Transformer (PET)	Learned representations (on-the-fly)	Stochastic gradient descent	Unlimited	Many (∞)	✗ Non-equivariant	✓ Scales to large datasets	Good for large datasets, flexible learning

Let's get started

[https://github.com/metatensor/
Workshop-spring-2025](https://github.com/metatensor/Workshop-spring-2025)



github.com metatensor / Workshop-spring-2025

Code Issues Pull requests 2 Actions Projects Wiki

main Workshop-spring-2025 / training-custom-models /

PicoCentauri proofread gap example a185d2d · 10 minutes ago

Name	Last commit message	Last commit date
..		
data	Store raw xyz instead	yesterday
part-1-gap	proofread gap example	10 minutes ago
part-2-pet	Fix path in input file	3 hours ago
README.md	First subsection of PET tutorial	15 hours ago

README.md

Set up

Download the data

Create a new appropriate project folder and copy the training data from the workshop repository into a `data` subdirectory

```
mkdir mts_workshop && cd mts_workshop
```