Machine learning accelerated geometry optimization in molecular simulation

Yilin Yang, Omar A. Jimenez-Negron, and John R. Kitchin
J. Chem. Phys. 154, 234704 (2021)

Paper Presentation

Shravan Godse

MRL, IIT Bombay

12th December, 2022

Contents

- Motivation

- ML Approach

- Model Details

- Fingerprints: Gaussian Symmetry Functions

- Results

- Conclusions

Motivation

- Structure relaxation/Geometry optimization is an important step in all computational material science calculations because the ground state structure properties are crucial in determining other dynamical properties
- A stable structure implies that the net force on all atoms is 0 and the total energy of the crystal is minimum

Scan Coordinate -Dihedral angle between H4-N2 (degrees)
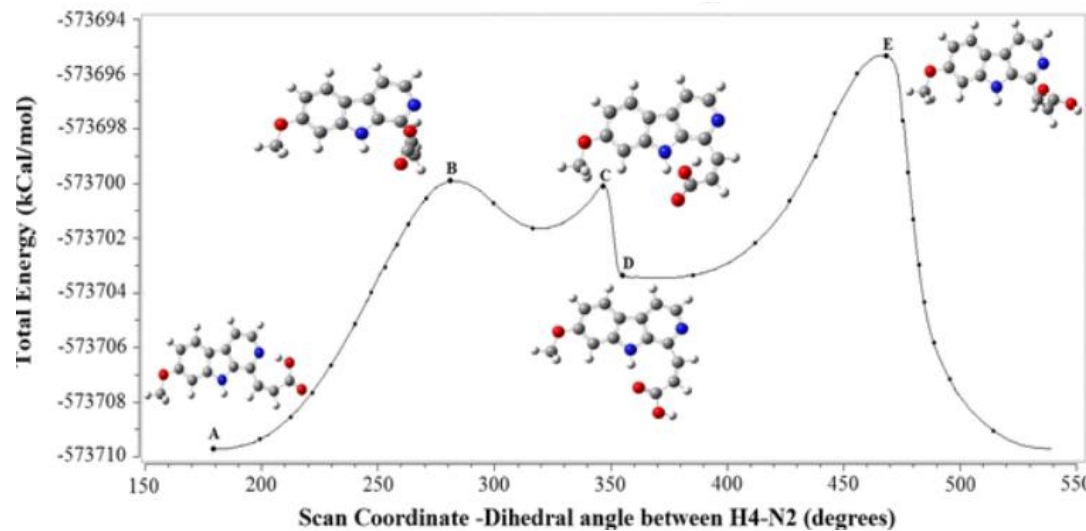
By varying structure parameters (interatomic distances, bond angles, but at the most basic level, atomic coordinates) one can see that the energy of the crystal changes
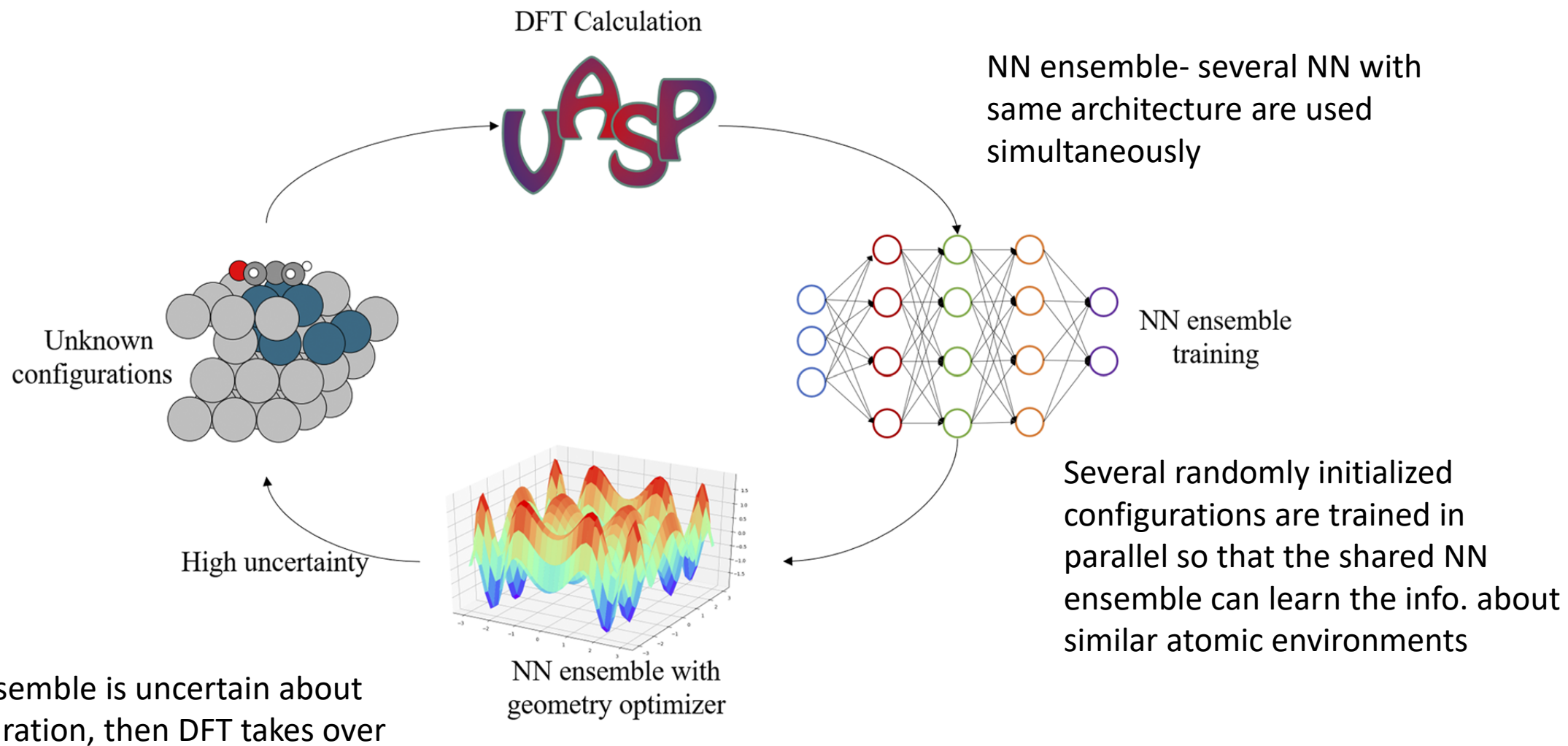
Motivation (continued)

- Relaxation calculations typically proceed via the computations of energies and forces at each step and employing optimization algorithms such as gradient descent or BFGS to move atoms towards a more stable configuration

- Problem: Computing Forces/Energies at each step is time consuming! Furthermore, time required for DFT calculations is $O(N^3)$ where N is the no. of atoms, so really tedious for supercells or surfaces ☹
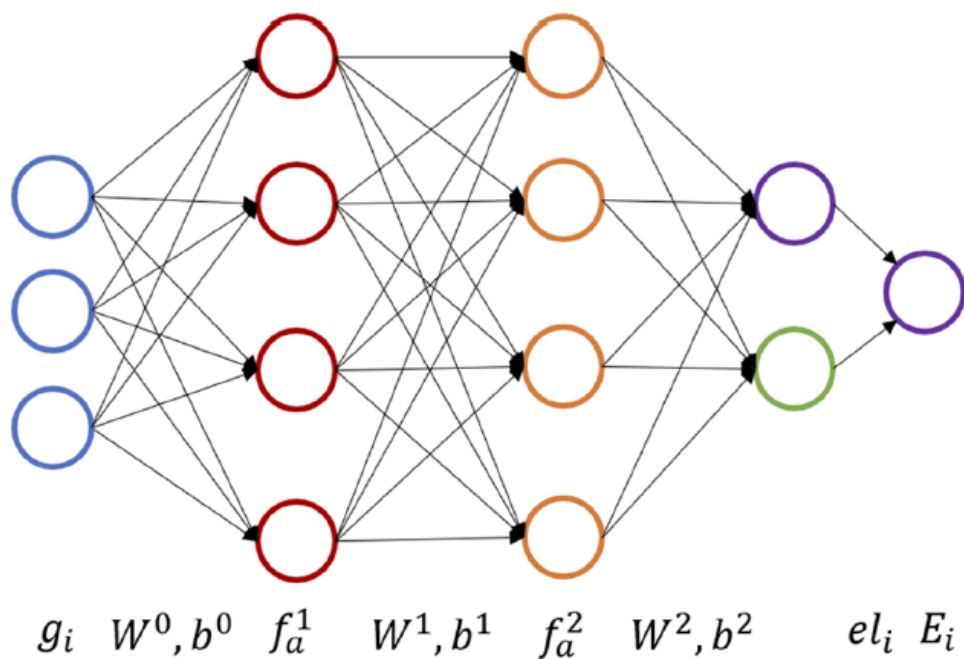
- Try Machine Learning!

How can ML help?

- ML can help us leverage the already available data on structure-force/energy by learning the underlying physics and making faster force/energy predictions

- However, ML has its own issues, the most notable one being the generalization issue – due to small sizes of training datasets, sometimes, the models predict erroneous values on previously unseen data

- To overcome this problem, techniques such as Gaussian Process Regression or Neural Network ensemble are used which can estimate the uncertainty associated with model prediction

- If the uncertainty is above a certain threshold, there ML predictions are not trustworthy and it needs to be trained on new data which call for DFT calculations

Approach in the paper – ONLINE ACTIVE LEARNING

DFT Calculation

NN ensemble- several NN with same architecture are used simultaneously

Unknown configurations

NN ensemble training

High uncertainty

Several randomly initialized configurations are trained in parallel so that the shared NN ensemble can learn the info. about similar atomic environments

NN ensemble with geometry optimizer

If the NN ensemble is uncertain about each configuration, then DFT takes over

Model Details

- 10 Neural Networks in the ensemble

- Gaussian Symmetry Functions used as descriptors (explained in next slide)

- Each NN had 2 hidden layers with 50 nodes and tanh activation $\left(\frac{e^z - e^{-z}}{e^z + e^{-z}}\right)$

$$E_i = \left[\mathbf{W}^{(2)} f_a^{(2)} \left(\mathbf{W}^{(1)} f_a^{(1)} \left(\mathbf{W}^{(0)} \mathbf{g}_i + \mathbf{b}^{(0)} \right) + \mathbf{b}^{(1)} \right) + \mathbf{b}^{(2)} \right]_{el_i}, \quad (1)$$

$$E_{tot} = \sum_i^N E_i, \quad (2)$$

$$\mathbf{f}_i = -\frac{\partial E_{tot}}{\partial \mathbf{r}_i}. \quad (3)$$

Loss: $\quad L = \frac{1}{N}\sum_i^N (E_i - \hat{E}_i)^2 + \lambda \frac{1}{\sum_i M_i}\sum_i^N \sum_j^{M_i} (F_{ij} - \hat{F}_{ij})^2, \quad (4)$

$g_i \quad W^0, b^0 \quad f_a^1 \quad W^1, b^1 \quad f_a^2 \quad W^2, b^2 \quad el_i \quad E_i$

Gaussian Symmetry Functions

- We cannot use bare atomic coordinates as material descriptors (input to ML) as they are not invariant under rotation, translation or exchange

Symmetry Functions – A way to transform atomic coordinates to create material fingerprints

$$f_c(R_{ij}) = \begin{cases} 0.5 \cdot \left[\cos\left(\frac{\pi R_{ij}}{R_c}\right) + 1 \right] & \text{for } R_{ij} \leq R_c \\ 0 & \text{for } R_{ij} > R_c. \end{cases}$$

$$G_i^1 = \sum_j f_c(R_{ij}),$$

$$G_i^2 = \sum_j e^{-\eta(R_{ij} - R_s)^2} \cdot f_c(R_{ij}),$$

and

$$G_i^3 = \sum_j \cos(\kappa R_{ij}) \cdot f_c(R_{ij}).$$

$$G_i^4 = 2^{1-\zeta} \sum_{j,k \neq i}^{\text{all}} (1 + \lambda \cos\theta_{ijk})^\zeta \cdot e^{-\eta(R_{ij}^2 + R_{ik}^2 + R_{jk}^2)}$$
$$\cdot f_c(R_{ij}) \cdot f_c(R_{ik}) \cdot f_c(R_{jk}),$$

and

$$G_i^5 = 2^{1-\zeta} \sum_{j,k \neq i}^{\text{all}} (1 + \lambda \cos\theta_{ijk})^\zeta \cdot e^{-\eta(R_{ij}^2 + R_{ik}^2)}$$
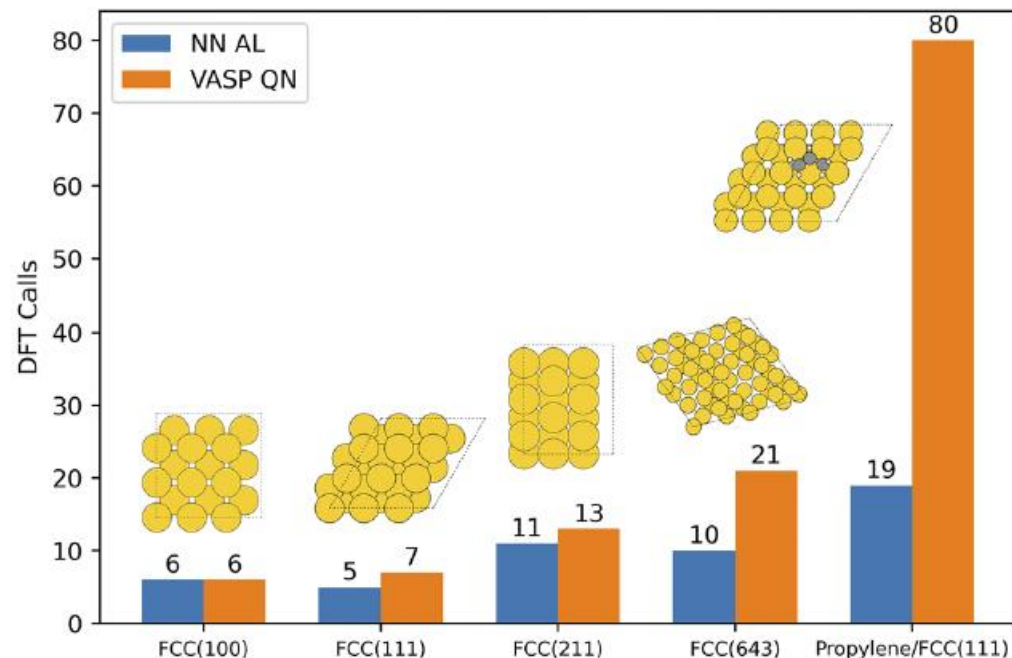$$\cdot f_c(R_{ij}) \cdot f_c(R_{ik}).$$

Results (1)

As a testament that NN models show less variance in the training region and become more and more uncertain when predicting on new data (i.e. predicting on new configurations which is inevitable during relaxation calculations)
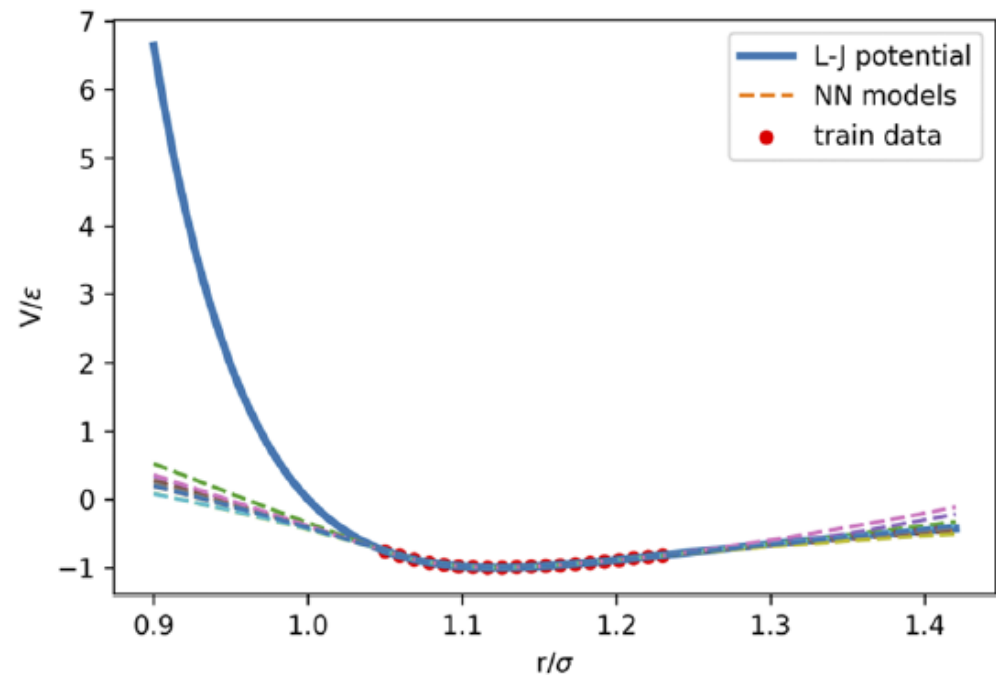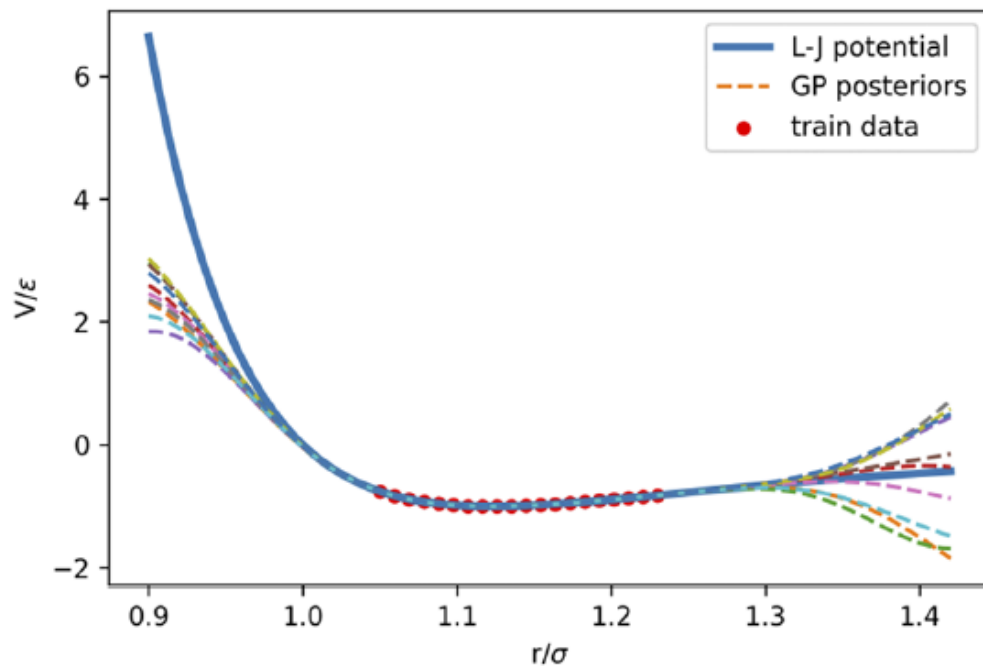
Results (2)

Active learning for geometry optimization of a single configuration:

Relaxation of different Au FCC surfaces and Au FCC with propylene on top

- The authors compared the performance of active learning model with VASP's quasi-Newton optimizer for relaxing Au FCC surfaces with increasing complexity
- The active learning model required fewer DFT calculations compared to VASP even when the structure complexity increase

Results (3)

Different possible approaches:
1. Train using a single configuration (as in the previous slide) and use NN ensemble only when uncertainty is low (single method)
2. Train using multiple configurations in parallel, but using a single NN ensemble (multiple method)
3. If you have access to previous relaxation data, you can pretrain the model and then use it (warm-up method)
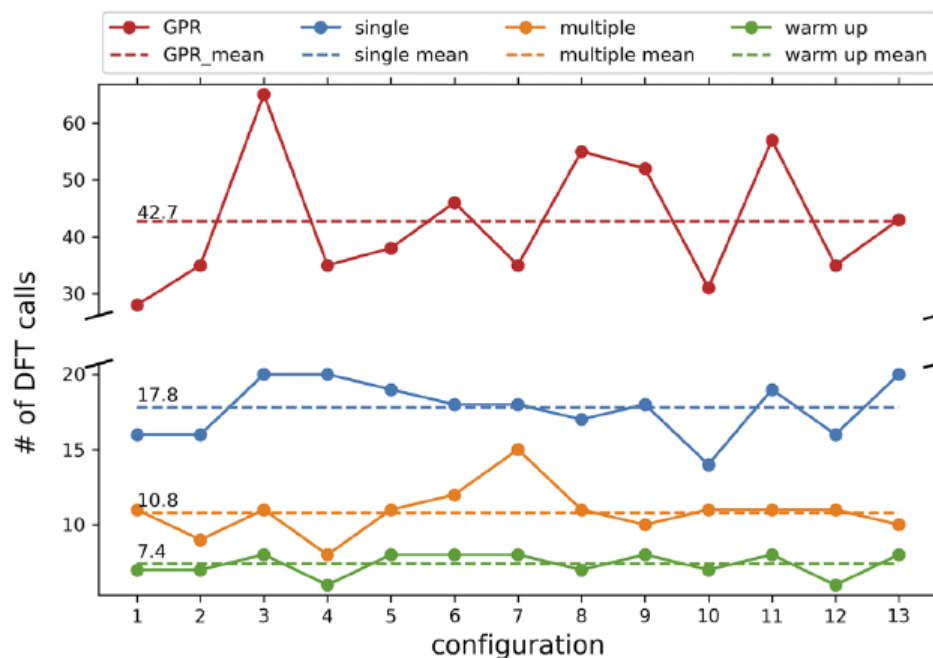

Number of DFT calls for three different active learning settings for the relaxation of acrolein/AgPd(111)

- Relaxation using plain VASP requires 193 DFT calculations
- GPR can perform the same relaxation with just about 43 DFT calculations
- The warm-up method seems to be most effective, but it needs prior relaxation data

Results (4)

A related scenario is when we have some data about the target configurations that we want to relax, for eg., if we have the active learning relaxation trajectories for many configurations of acrolein/AgPd and we want to relax the remaining configurations
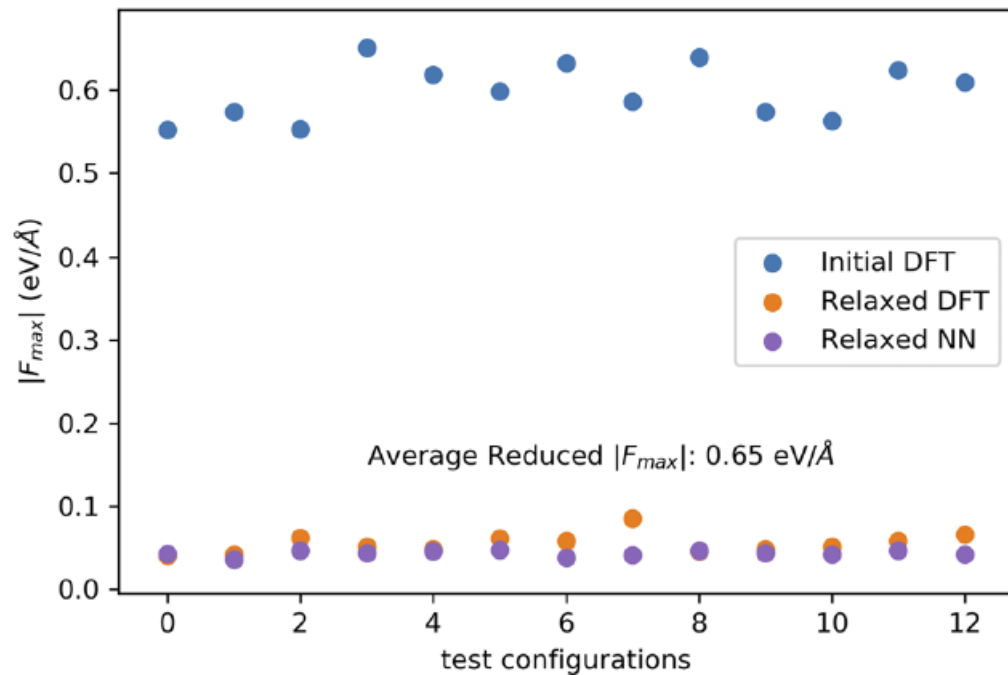
The authors tried offline learning in this case
They had information about relaxation trajectories of 243 configurations of acrolein/AgPd
They trained their NN on these 243 trajectories and tried to relax 13 new configurations
They find that this trained NN could reduce the forces from 0.7 eV/Å to 0.1 eV/ Å
Although this may not be sufficient, offline training can provide us with better initial structures which when relaxed using standard tools like VASP or Espresso will require fewer DFT steps

Results (5)

**Performance of the active learning on more complex systems**
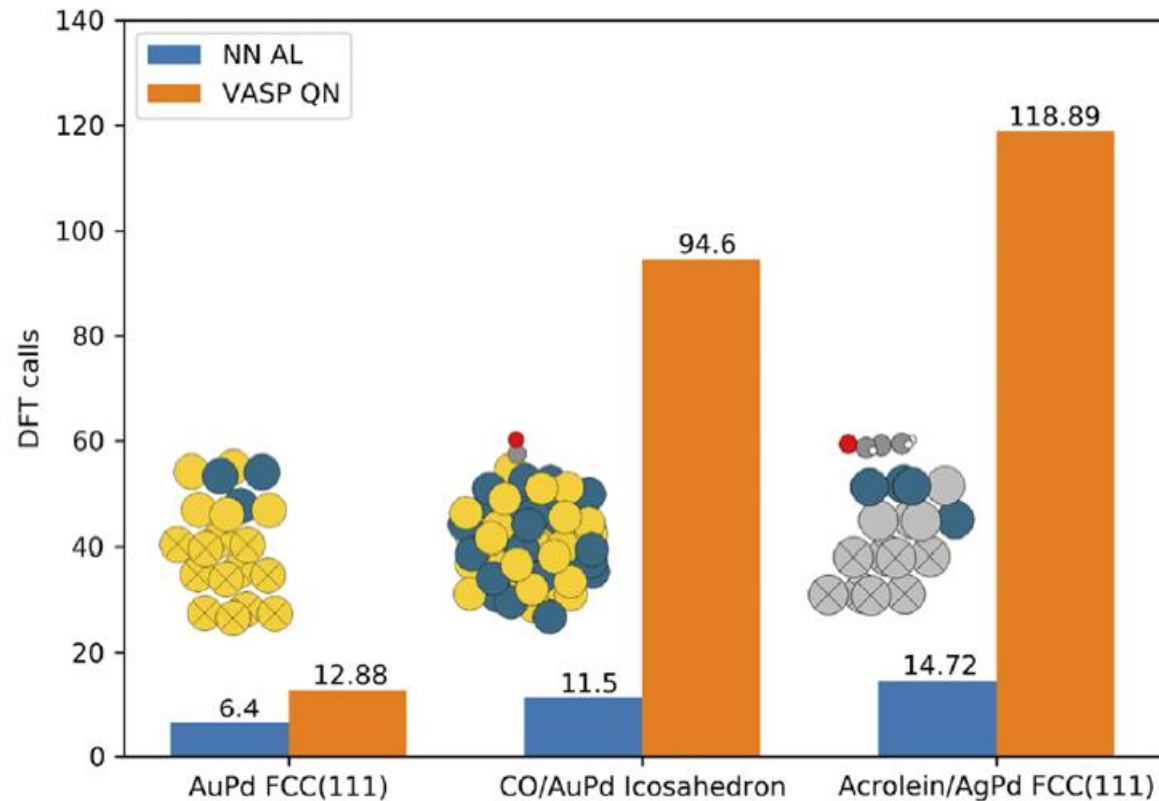
To see advantage of active learning, the authors performed relaxations on surfaces with increasing complexity – (a) AuPd FCC(111) (b) CO on Au/Pd Icosahedron (c) Acrolein/Ag/Pd FCC(111)

We can see that as system complexity increases (more types of atoms, less symmetry), the active learning technique offers a significant advantage

Results (6)

Climbing Image NEB calculations using Active Learning
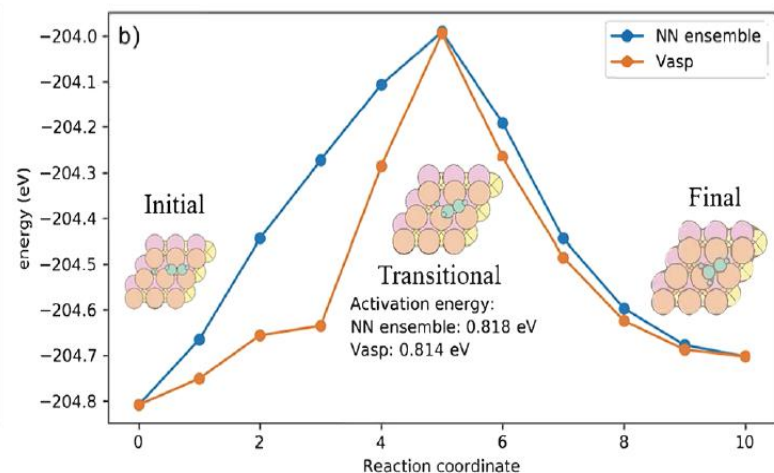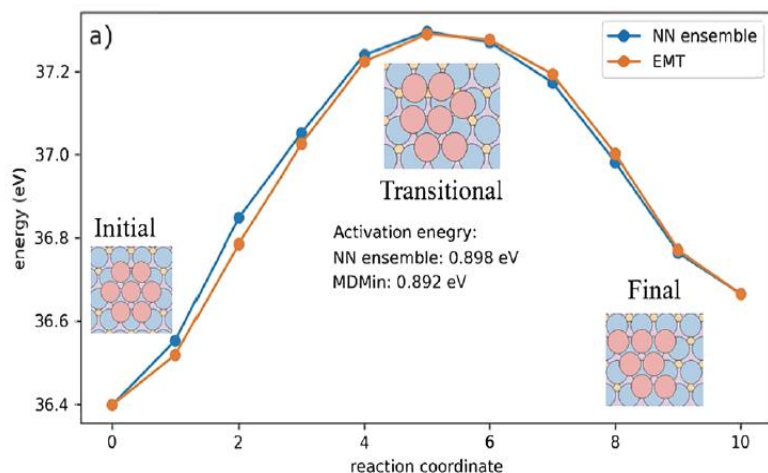
CINEB uses initial and final configurations to determine the path taken to achieve them

For the Pt heptamer rearrangement over Pt FCC(111), the authors use EMT for NEB and for acetylene hydrogenation over Pd FCC(111), they use DFT

Irrespective of DFT or EMT, the active learning method can accurately predict the transition state and thus, the activation barrier

	Pt heptamer rearrangement (EMT)	Acetylene hydrogenation (VASP)
Calculator	596 calls	1109 calls
NN ensemble with AL	9 calls	30 calls

Furthermore, it requires far fewer DFT/EMT calls showing rapid acceleration

Results (7)

Limiting training data to recent configurations

While NN training takes almost no time compared to DFT, it may create issues if the training size increases. Moreover, the authors found that the correlation between two configurations decrease as the number of steps between them increase. Thus, one can use only the recent configurations to perform locally geometry optimizations
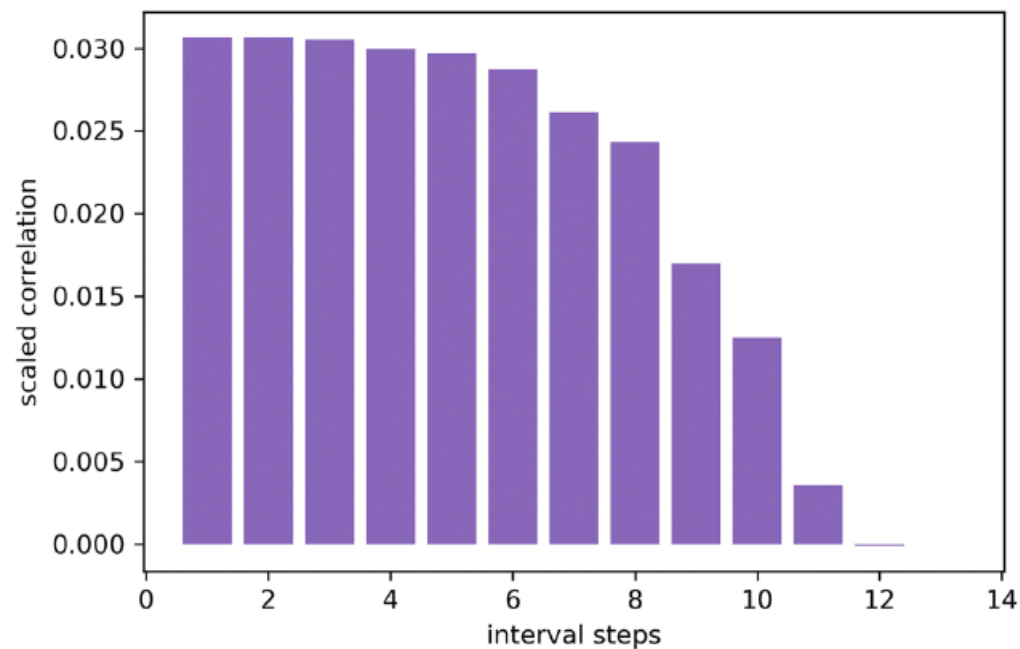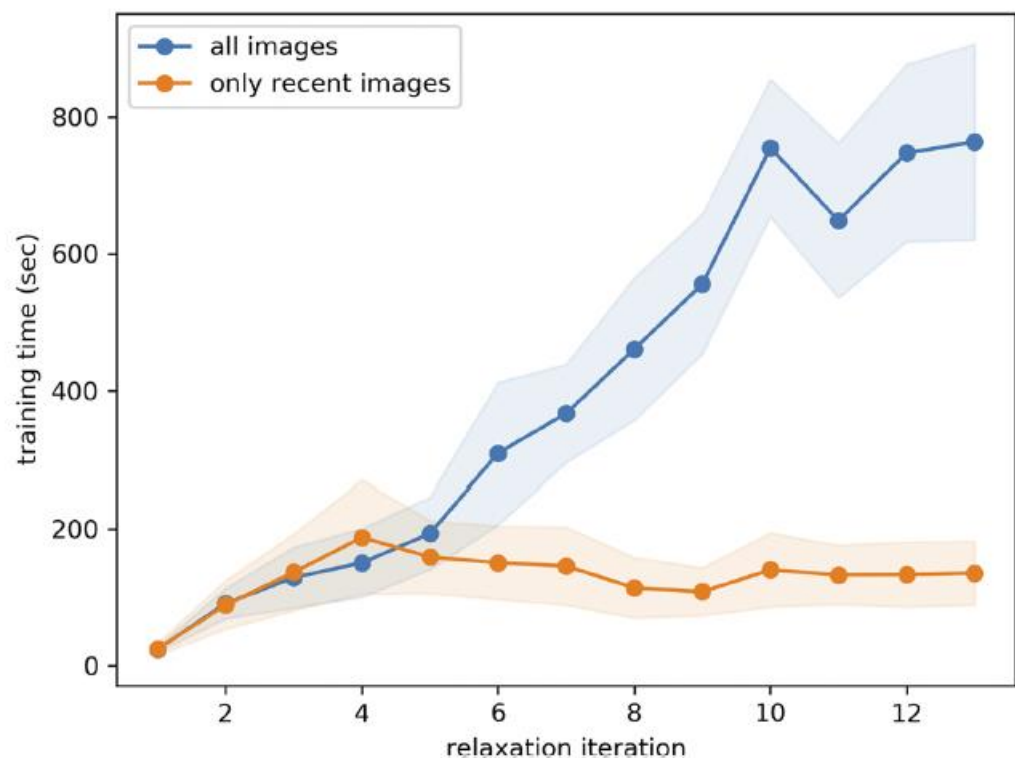
Conclusions

- The authors demonstrated the advantage of NN ensemble based active learning by relaxing surfaces and doing NEB calculations

- The acceleration offered by this technique is more pronounced as one considers systems of increasing complexity

- This technique reduces calls to DFT/EMT by 50-90% depending on the system