Machine Learning of an FeS_{aq} Force Field to Explore Iron:Sulfide Ratio Impact on Nucleation

Janou A. Koskamp, Vincent F. D. Peters, Akshayini Muthuperiyanayagam, Devis Di Tommaso and Mariette Wolthers

Introduction

Mackinawite (FeS) serves as both a scaling agent in geothermal piping and a catalyst in industrial applications, with potential for environmental remediation. Investigating FeS formation under varying $\{Fe^{2+}\}:\{S^{2-}\}$ ratios is crucial for understanding its behavior in natural and engineered systems. However, the influence of this ratio on FeS nucleation remains unknown. Our study aims to characterize the smallest FeS clusters using molecular dynamics simulations (MD) with metadynamics, requiring the development of a suitable force field through first-principles calculations and machine learning. We focus on training and creating an accurate machine-learned force field (MLFF). A mature MLFF enables metadynamics MD to construct the free energy surface (FES) of the association of a charged triple ion cluster (CTIC), revealing the impact of $\{Fe^{2+}\}:\{S^{2-}\}$ ratios on nucleation (Fig. 1).

Method

The MLFF was developed using VASP software. Initially, a water box (Fig. 2) served as input. VASP facilitated the creation of a model linking descriptors (angular and radial distribution functions (ADF/RDF)) to outputs (energies, forces, and stress tensors) (Fig. 3). During MLFF training, the error and learning frequency decreased and the period between learning steps increased (Fig. 4). Upon reaching maturity, the MLFF required refitting to its fast mode. Subsequently, the MLFF was validated through performance evaluation in an MD run (Fig. 5). Successful completion of these steps allowed for the addition of Fe^{2+} and S^{2-} (Fig. 6), initiating the expansion of the MLFF through re-entering the training cycle.

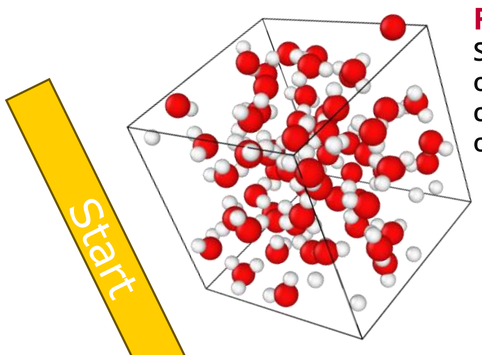

Figure 2
Schematic representation of the initial simulation box containing 63 H_2O to start our on-the-fly training

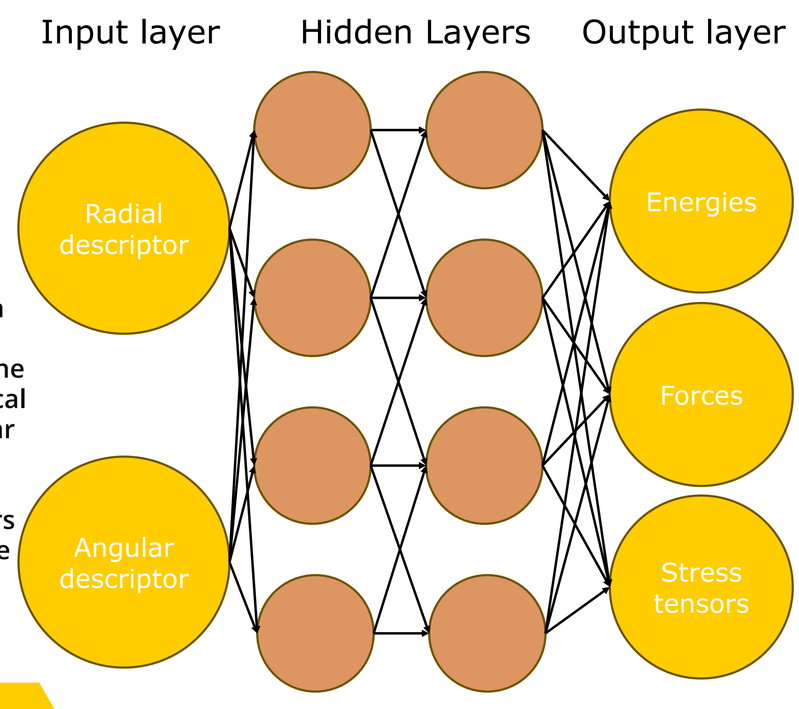

Figure 3
An example of an artificial neural network where the inputs are the local radial and angular distribution; the energies, forces and stress tensors are the output we would like to predict

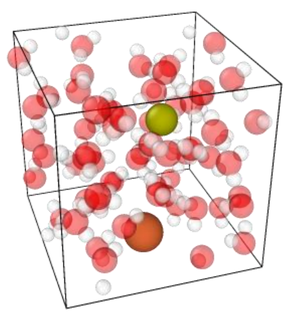

Figure 6
Schematic representation of the simulation box with added species Fe^{2+} and S^{2-} to expand our MLFF and continue training

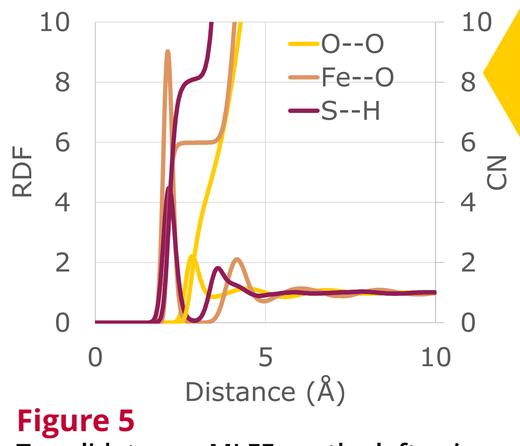

Figure 5
To validate our MLFF: on the left axis the RDF (solid lines) and on the right axis the coordination number (CN) (dashed lines)

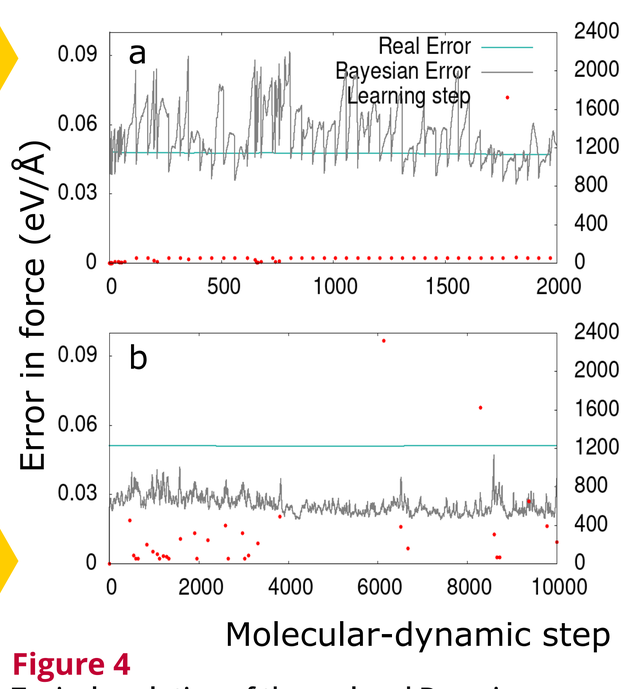

Figure 4
Typical evolution of the real and Bayesian errors (green and gray lines, primary y-axis) for the forces at the start of the training (a) and later in the learning procedure (b)

Figure 1
Illustrative 3D representation of a free energy surface of triple ion cluster with a 2D projection of the surface at the bottom of the graph

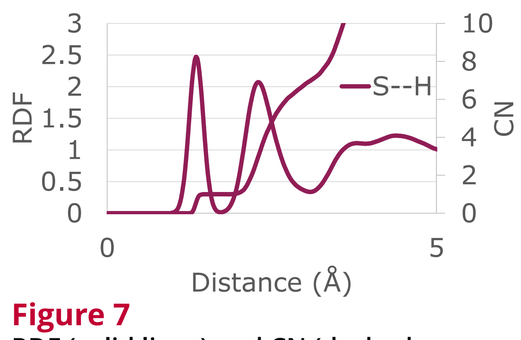

Figure 7
RDF (solid lines) and CN (dashed lines) of S—H in *ab initio* MD during MLFF training

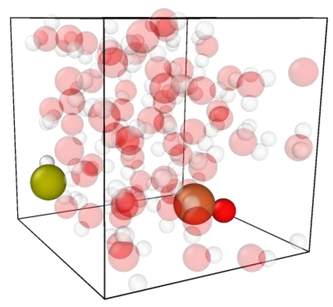

Figure 8
Snapshot of simulation box with species FeO (orange and red, respectively) and SH^- (green and white). Revealing a vacuum around the O of FeO

Results:

- Developed MLFF for FeS_{aq}
- MLFF accurately reproduces RDFs and (low) self-diffusion coefficient of water, consistent with RPBE functional used for training
- MLFF significantly faster than ab initio methods, enabling larger box sizes (e.g., 1105 or 480 water molecules).

Challenges:

- Reproducing SH^- and $Fe^{2+}\cdots OH^-$ species observed experimentally, especially:
  - Instability of certain configurations, as found in *ab initio* force field training (Fig. 7, with HS^-, closer to experiments instead of S^{2-}).
  - However, $FeOH^+$ transforms into FeO and causes physical-chemical unrealistic behaviour of the water molecules, generating the equivalent of a vacuum around FeO (Fig. 8)