

First Principles and Machine Learning Investigations into the Atomic and Mechanical Properties of Cement Hydrates

J. Gabriel Balarezo, Henry Pinto

CompNano Research Group,
School of Physical Sciences and Nanotechnology,
Yachay Tech University, Ecuador

juan.balarezo@yachaytech.edu.ec

NSBC 2025

May 10, 2025



**SCHOOL OF
PHYSICAL SCIENCES
AND NANOTECHNOLOGY**



Concrete research

- Synthetic material produced in larger volumes than any other material on Earth.
- Annual consumption of ~35 billion tonnes. (Li, Z, *et al.*, 2022)
- Provides the backbone for modern infrastructure. (Monteiro, *et al.*, 2017)
(Biernacki, *et al.*, 2017)

Challenges



Sustainability

- High CO₂ (~8%) emissions from cement production.



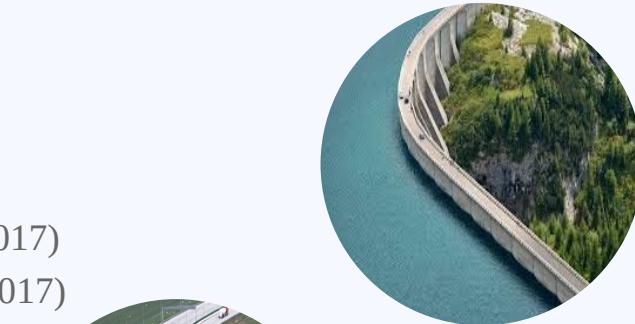
Tailoring Formulations for Specific Applications

- Deep understanding of cement hydration kinetics, admixture interactions and micro-structure control.



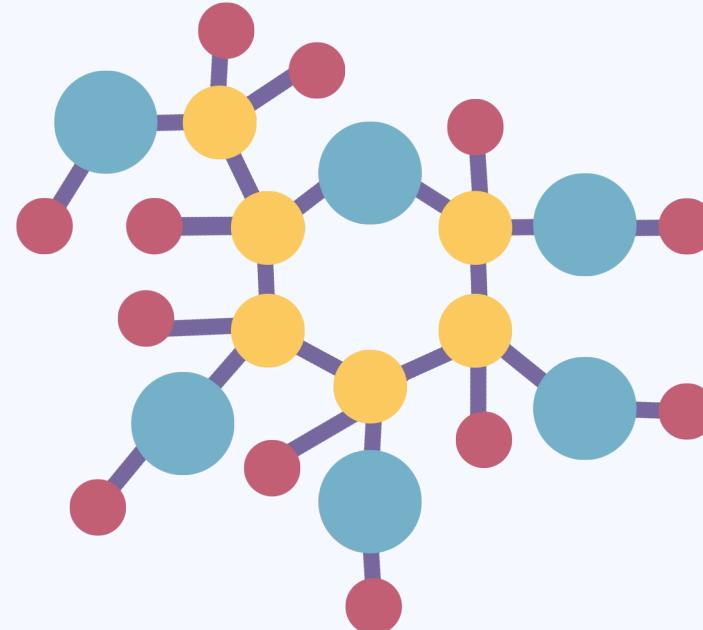
Computational modelling limitations

- Increasing complexity of cementitious materials.



Presentation outline

- Introduction
- Aim and scope
- Methodology
- Results
- Conclusions
- References



Introduction

Cement Hydrates (C-S-H)

- Cement hydrates (C-S-H) are the main product of cement hydration.
- Disordered, gel-like structure with short range order resembling layered tobermorite.
- Responsible for the mechanical strength, chemical and transport properties, and durability of cement, and consequently, of concrete itself.

(Papatzani, *et al.*, 2015)
 (Qomi, *et al.*, 2020)

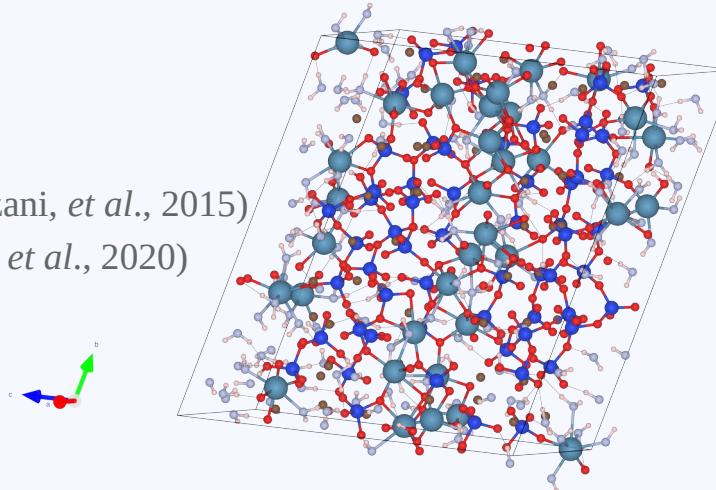
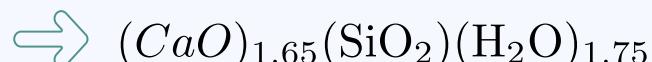


Figure 1. Molecular model of C-S-H proposed by [3]. Lavender and white spheres are oxygen and hydrogen from water molecules respectively; light blue and brown spheres are inter and intra-layer calcium ions, respectively; electric blue and red spheres are silicon and oxygen atoms in silica tetrahedra.

➤ Molecular model (Pellenq, *et al.*, 2009)

Considers only the chemical specificity of the system as the overriding constraint.

- Calcium/Silicon ratio (C/S) = 1.7
- C-S-H particle density of 2.6 g/cm^2



Aim and scope

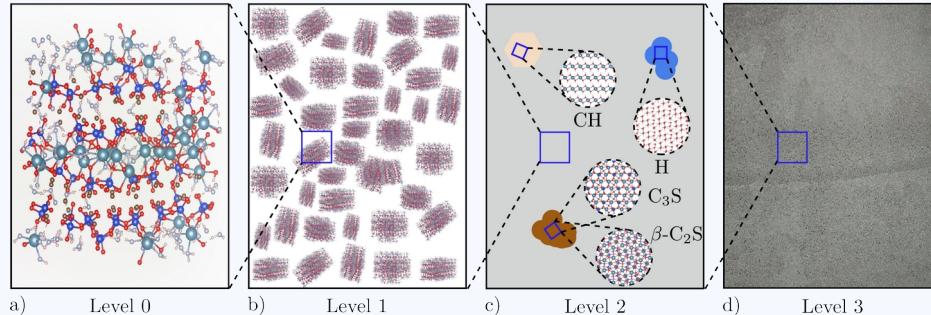
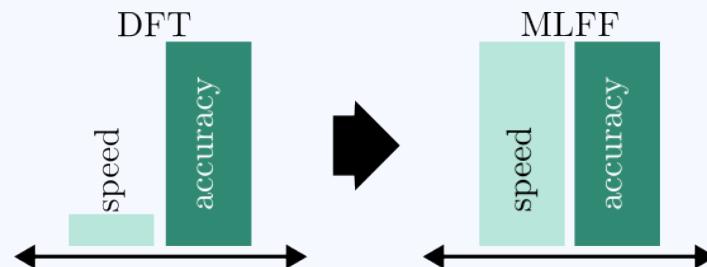


Figure 2. A four-level model to represent the upscaling C-S-H properties from the nanoscale to the engineering scale. (a) Snapshot of the nanostructure of C-S-H. (b) The mesostructure of C-S-H created by agglomeration of randomly oriented C-S-H nanoparticles. (c) The microtexture of hardened paste composed by hydration products. (d) The macrotexture of cement paste at the engineering level.

Naive approach: DFT solely
 Better approach: DFT + ML

(Biernacki, *et al.*, 2017)



Principal Objective

Develop and validate an MLFF for accurate prediction of CSH's equation of state and mechanical properties, achieving DFT-level accuracy with significantly reduced computational cost.

Secondary Objectives

Characterize the electronic and structural properties of CSH through initial DFT relaxation and DOS/PDOS analysis to provide a reliable foundation for MLFF development.

Train an MLFF on-the-fly during a MD simulation at 400 K and evaluate its predictive accuracy through error analysis of energies, forces, and stresses, followed by refitting.

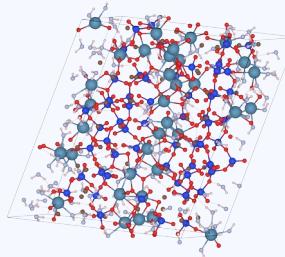
Determine low-energy CSH configurations and mechanical properties using MLFF-based simulated annealing and EOS calculations, preparing for DFT validation.

Methodology

Setup

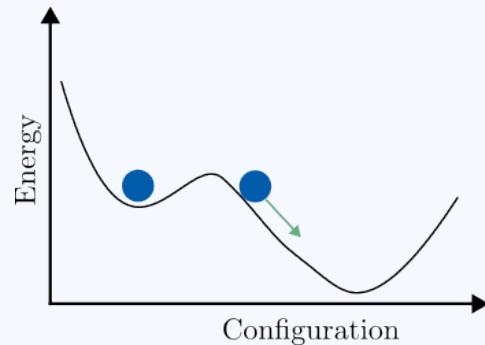
DFTB⁺

VASP



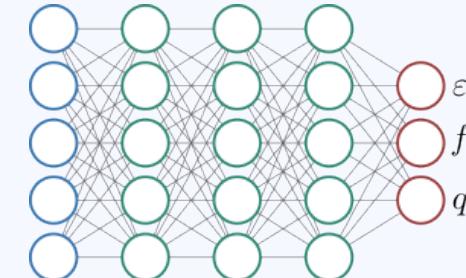
- DFTB+ with GFN1-xTB method for k-points convergence test.
- VASP with PBESol + DFTD3(BJ) for k-point convergence and Cut-off energy convergence test.

Structure optimisation



- K-points convergence test
- Cut-off energy convergence test
- DOS/PDOS calculations

MLFF training and prediction



- Molecular Dynamics and on-the-fly MLFF training.
- MLFF refinement (error analysis and refitting)
- EOS and Simulated Annealing

Structure optimisation

Structure optimisation

The files needed for these simulations are INCAR, POSCAR, KPOINTS, and POTCAR

K-points convergence (DFTB+)

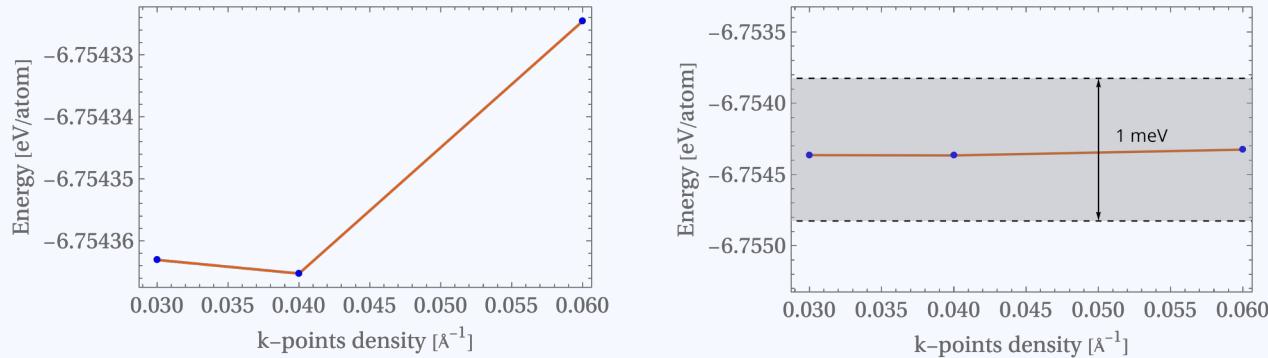


Figure 2. k-point convergence test computed using DFTB+ with the GFN1-xTB method for $0.03 \leq \Delta k (\text{\AA}^{-1}) \leq 0.060$. Due to the large, disordered nature of the C-S-H supercell, the total energy shows negligible variation beyond the Γ -point, justifying its use in subsequent calculations.

» Highlights

- Total energy convergence achieved at a $\Delta k = 0.060 \text{ \AA}^{-1}$ $1 \times 1 \times 1$ k-point grid using GFN1-xTB ($< 1 \text{ meV}$ variations)
- Results used to guide higher-level VASP calculations

Cut-off energy convergence (VASP)

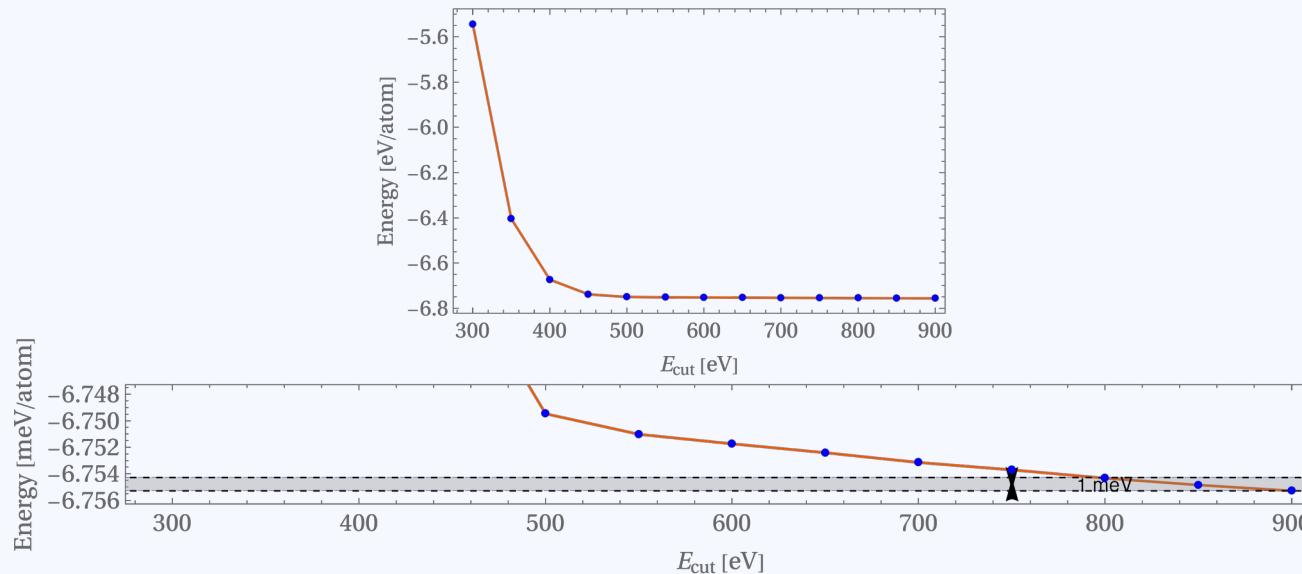


Figure 2. Cut-off energy convergence test computed using VASP with the PBESol+DFTD3(BJ) functional for $300 \leq E_{cut}$ (eV) ≤ 900 . Convergence within 1 meV/atom is reached at 800 eV.

➤ Highlights

- Total energy convergence achieved at $E_{cut} = 800$ eV using PBESol+DFTD3(BJ) (<1meV window)
- High cutoff needed due to oxygen content and structural disorder in C–S–H

Partial Density of States and Total Density of States (PDOS/DOS)

- Total and projected DOS calculated to examine the electronic structure of C–S–H
- VASP calculations performed using the PBE functional with rVV10 dispersion correction
- Projected states onto atomic species (e.g., Ca, Si, O, H) to determine their contributions near the Fermi level

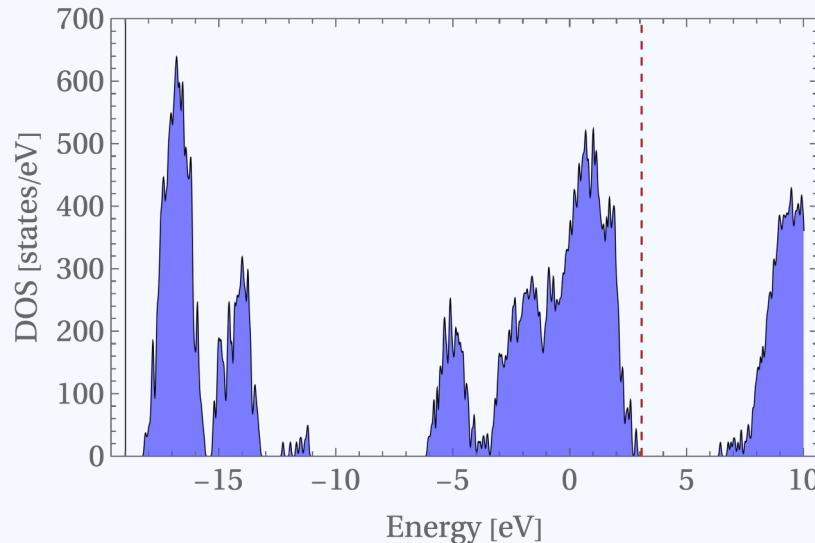
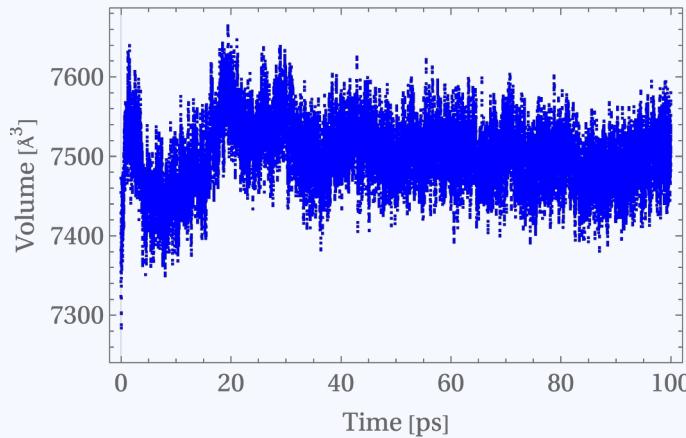
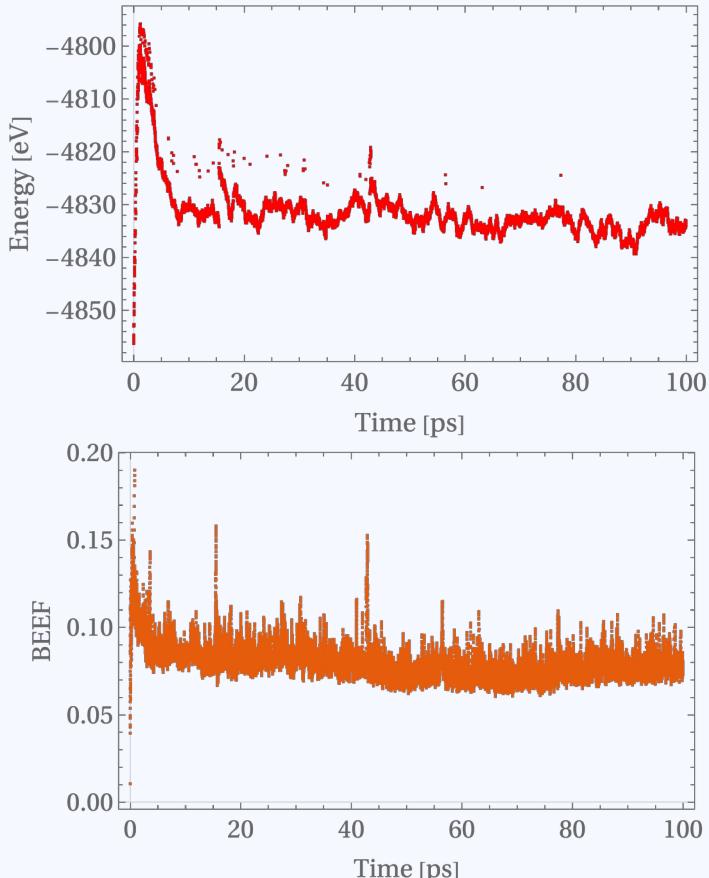


Figure 2. DOS of C–S–H computed after relaxation with the PBESol + rVV10 functional.

MLFF training and prediction

MLFF training and prediction

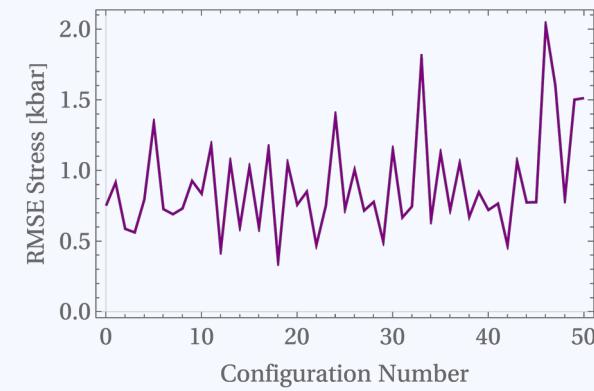
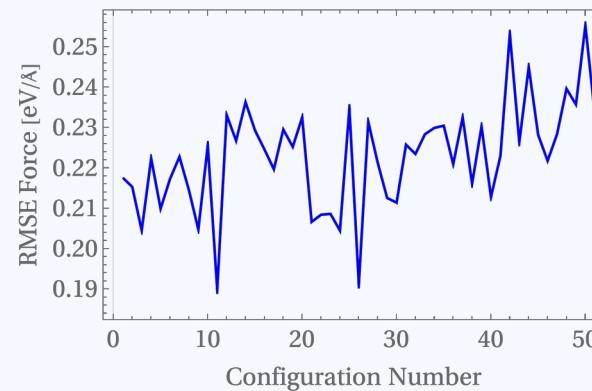
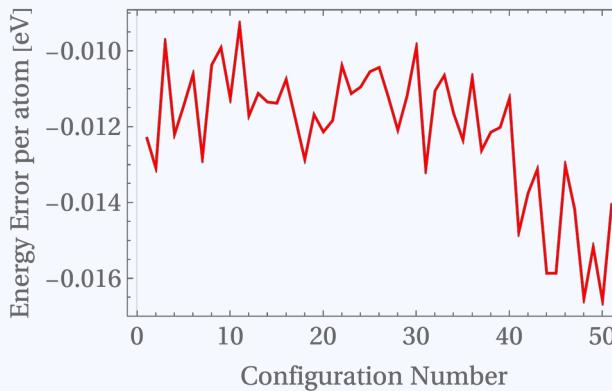
MD at 400 K and on-the-fly MLFF training



- Total energy reaches a steady state as the training process progresses, showing good model stability and prediction consistency.
- Volume variations reach 2% of the original volume.
- Bayesian error reaches values $\sim 0.07\text{-}0.12 \text{ eV}/\text{\AA}$, indicating moderate uncertainty and showing some peaks corresponding to training phases.

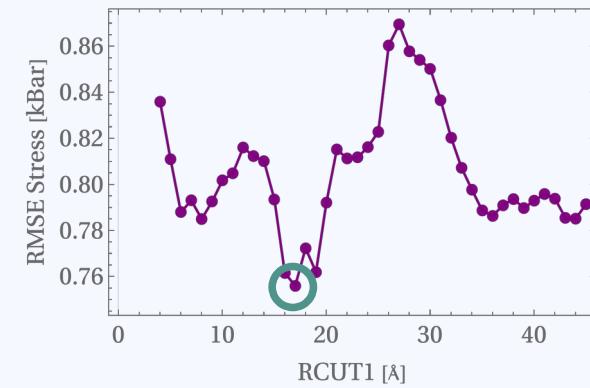
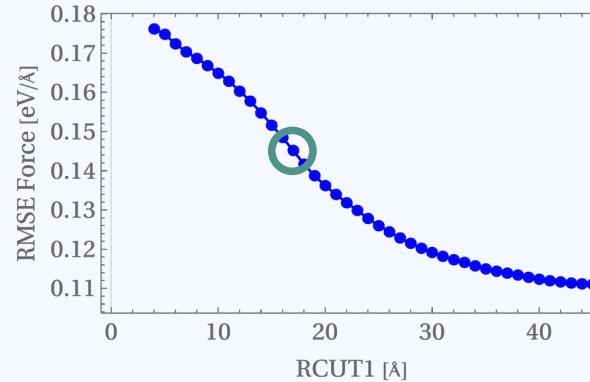
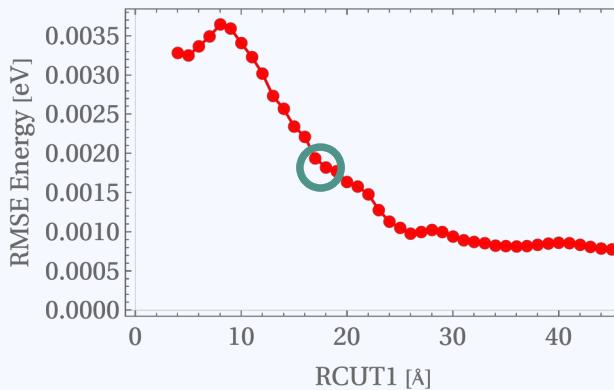
ML refinement (error analysis)

- MD at 400 K this time using the MLFF solely
- We choose 50 structures out of the 50000 structures, uniformly; relax the structures using DFT and MLFF
- Find optimal hyperparameters (RCUT1 & RCUT2) and refit our MLFF model.

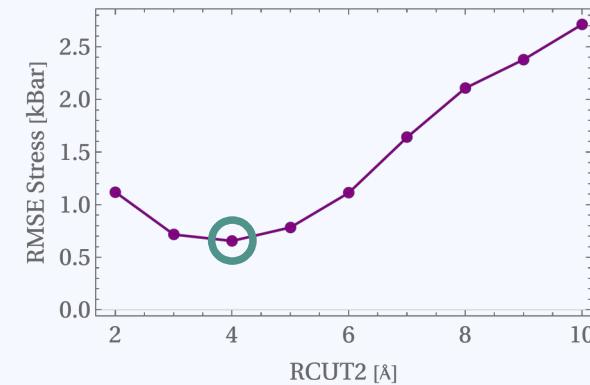
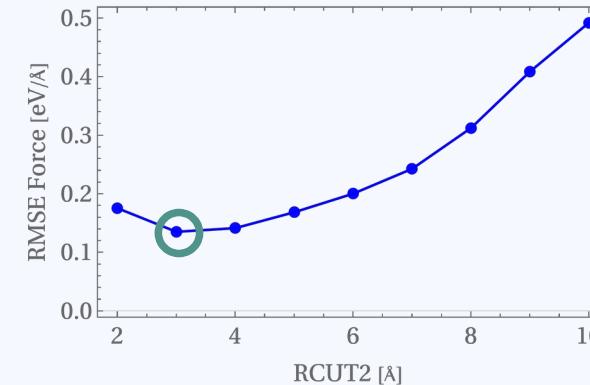
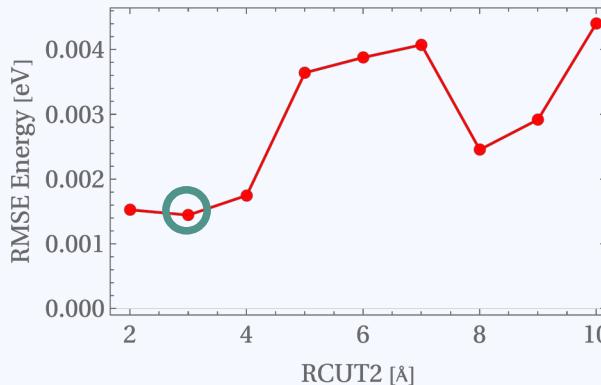


Optimal hyperparameters (RCUT1 & RCUT2)

RCUT1: Radial cutoff (\AA) for pairwise interactions in MLFF descriptors, setting the maximum distance for radial terms.

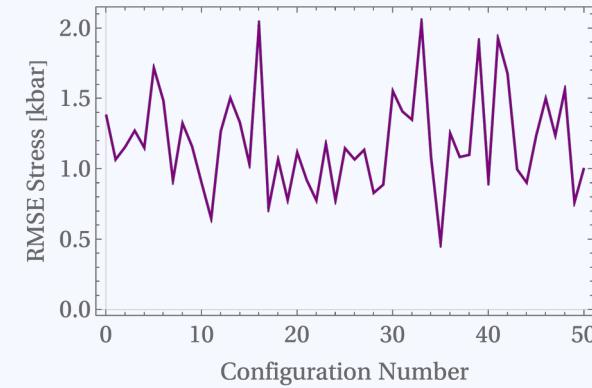
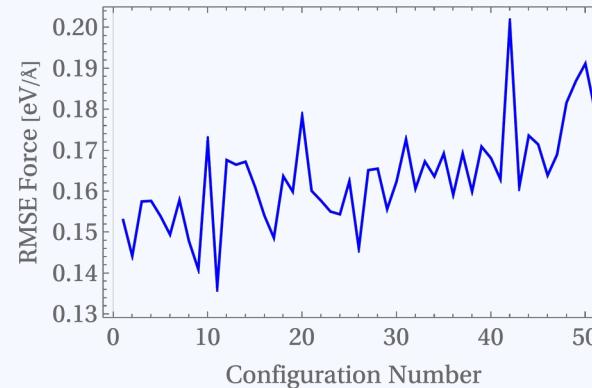
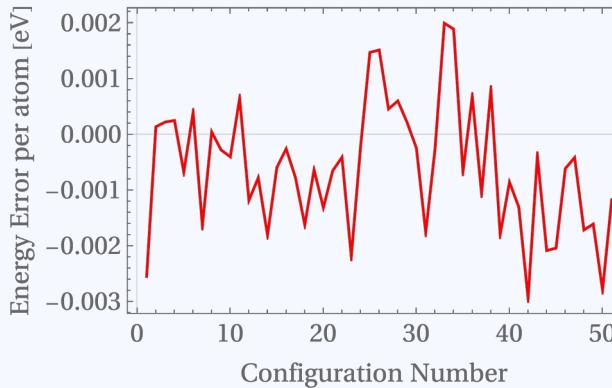


RCUT2: Radial cutoff (\AA) for three-body angular interactions, defining the range for triplets in angular descriptors.



Refitted MLFF model

- Run a refit process setting the optimal parameters (RCUT1=17 & RCUT2=3)
- Recompute energies, forces and stresses.



EOS and Simulated Annealing

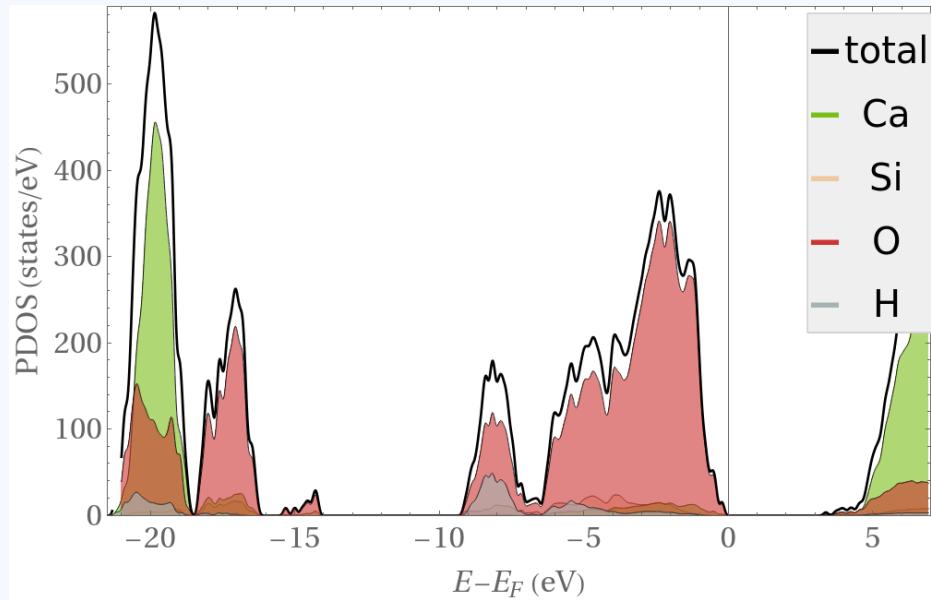
- 1% volume variation leftwards and rightwards, achieving 16 points in total, then find the optimal structure using our MLFF model.
- Perform simulated annealing to cool the C-S-H structure from 400 K to 0 K, and compute EOS as described above.

$$E(V) = E_0 + \frac{9V_0B_0}{16} \left\{ \left[\left(\frac{V_0}{V} \right)^{2/3} - 1 \right]^3 B'_0 + \left[\left(\frac{V_0}{V} \right)^{2/3} - 1 \right]^2 \left[6 - 4 \left(\frac{V_0}{V} \right)^{2/3} \right] \right\}$$

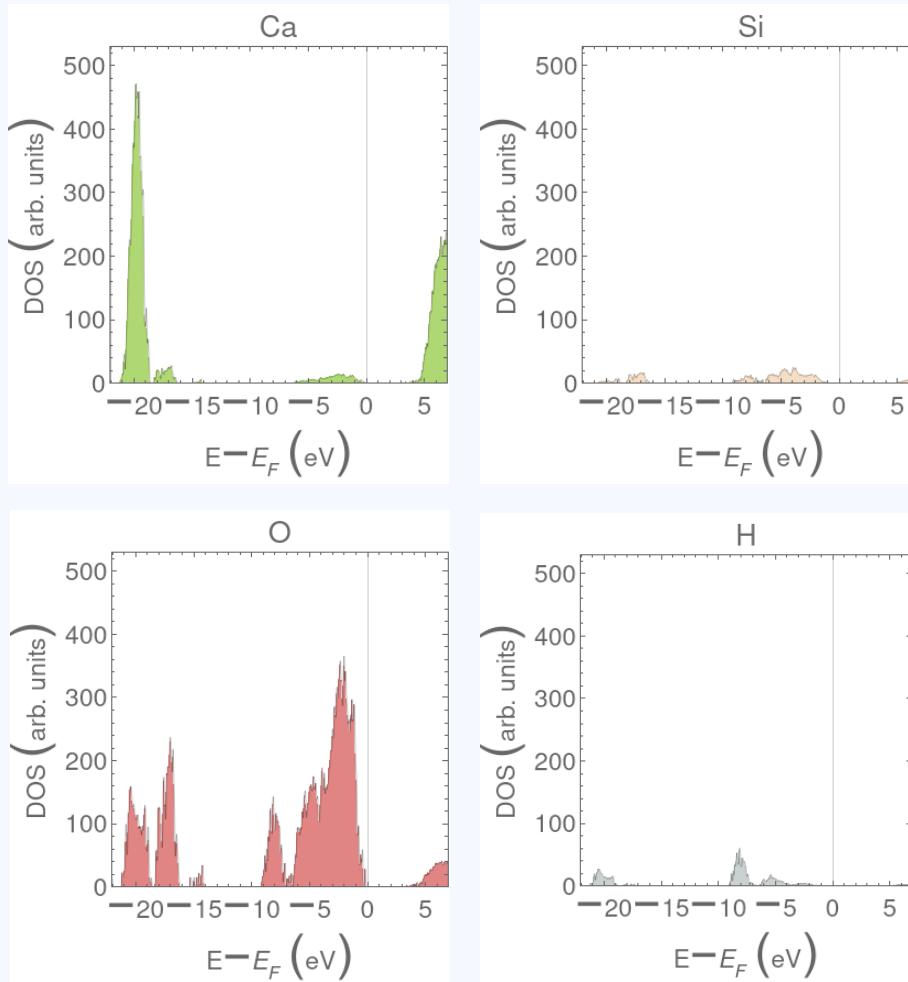
Third-order Birch-Murnaghan equation of state
(Katsura, et al., 2019)

Results

DOS & PDOS

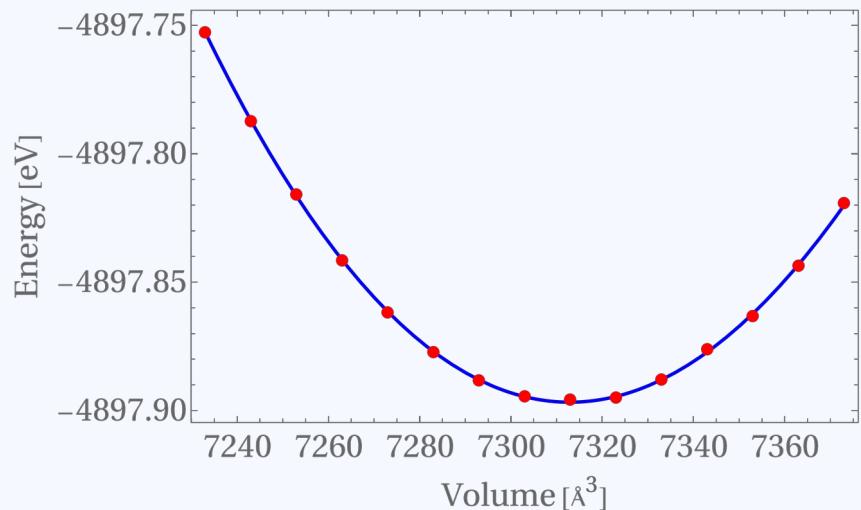


- CSH is an insulator (band gap ~ 3.1 eV), with localized O 2p states in the valence band, confirming strong ionic and hydrogen bonding.
- Si-O (covalent), Ca-O (ionic), and O-H (hydrogen bonding) interactions dominate, validated by Si 3p, Ca 3d, and H 1s contributions.

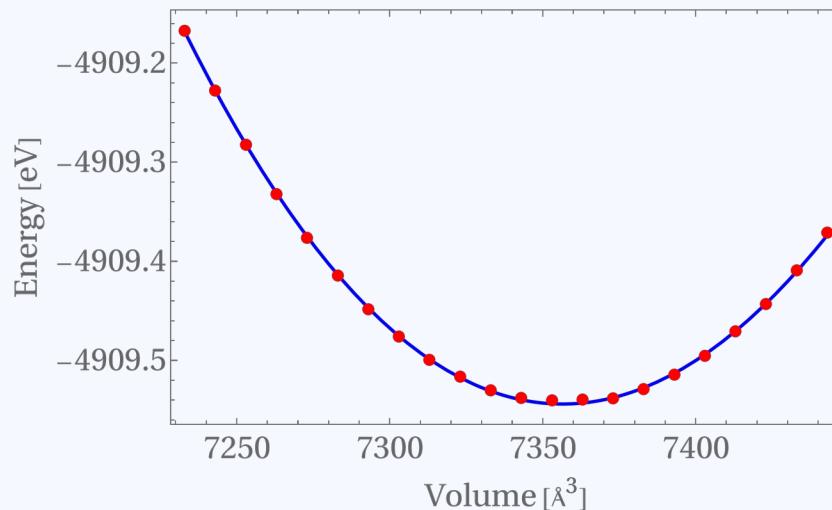


EOS: Bulk modulus, derivative of the bulk modulus & optimal volume

Refitted MLFF



Original MLFF & Simulated annealing



Parameter	Rifitted MLFF	Simulated Annealing	Literature
Volume (V)	7312.85 \AA^3	7356.38 \AA^3	N/A
Energy (E)	-4897.9 eV	-4909.54 eV	N/A
Bulk Modulus (B_0)	51.0563 GPa	55.3869 GPa	$47 \pm 3 \text{ GPa}$
Bulk Modulus Derivative (B'_0)	8.52598	7.48163	4

Conclusions and outlook

- **CSH's Electronic Structure Informs Bonding**

DOS/PDOS reveals a ~3.1 eV band gap and dominant O 2p states, confirming CSH's insulating nature and strong Si-O, Ca-O, and O-H bonding, providing a solid DFT foundation for MLFF training.

- **MLFF Achieves Promising Accuracy**

On-the-fly MLFF training over 50,000 MD steps, followed by error analysis of 50 structures, yields RMSE values of ~1–3 meV/atom (energies), ~0.13–0.20 eV/Å (forces), and ~0.2–2 kbar (stresses), improved further by refitting.

- **MLFF EOS Predicts Mechanical Properties**

Simulated annealing and EOS calculations using MLFF estimate CSH's bulk modulus at ~51–55 GPa and optimal volume at ~7312–7356 Å³, demonstrating MLFF's efficiency for large-scale cement simulations.

- **DFT Validation Underway for Accuracy**

Ongoing DFT relaxations of 12 volumes will validate MLFF predictions, ensuring reliability of CSH's mechanical properties and highlighting MLFF's potential for sustainable cement research.

References

- [1] Li, Z.; Yoon, J.; Zhang, R.; Rajabipour, F.; Srubar III, W. V.; Dabo, I.; Radlińska, A. Machine Learning in Concrete Science: Applications, Challenges, and Best Practices. *npj Computational Materials* **2022**, *8*, 127.
- [2] Monteiro, P. J. M.; Miller, S. A.; Horvath, A. Towards Sustainable Concrete. *Nature Materials* **2017**, *16*, 698–699.
- [3] Biernacki, J. J.; Bullard, J. W.; Sant, G.; Brown, K.; Glasser, F. P.; Jones, S.; Ley, T.; Livingston, R.; Nicoleau, L.; Olek, J.; Sanchez, F.; Shahsavari, R.; Stutzman, P. E.; Sobolev, K.; Prater, T. Cements in the 21st Century: Challenges, Perspectives, and Opportunities. *Journal of the American Ceramic Society* **2017**, *100*, 2746–2773.
- [3] Pellenq, R. J.-M.; Kushima, A.; Shahsavari, R.; Van Vliet, K. J.; Buehler, M. J.; Yip, S.; Ulm, F.-J. A Realistic Molecular Model of Cement Hydrates. *Proceedings of the National Academy of Sciences* **2009**, *106*, 16102–16107.
- [4] Papatzani, S.; Paine, K.; Calabria-Holley, J. A Comprehensive Review of the Models on the Nanostructure of Calcium Silicate Hydrates. *Construction and Building Materials* **2015**, *74*, 219–234.
- [5] Qomi, M. J. A.; Bauchy, M.; Pellenq, R. J.-M. In *Handbook of Materials Modeling: Applications: Current and Emerging Materials*; Andreoni, W., Yip, S., Eds.; Springer International Publishing: Cham, 2020; pp 1761–1792.
- [6] Katsura, T.; Tange, Y. A Simple Derivation of the Birch–Murnaghan Equations of State (EOSs) and Comparison with EOSs Derived from Other Definitions of Finite Strain. *Minerals* **2019**, *9*, 745.
- [7] Oh, J. E.; Clark, S. M.; Wenk, H.-R.; Monteiro, P. J. M. Experimental Determination of Bulk Modulus of 14 Å Tobermorite Using High Pressure Synchrotron X-ray Diffraction. *Cement and Concrete Research* **2012**, *42*, 397–403.