



First Principles and Machine Learning Investigations into the atomic and Mechanical Properties of Cement Hydrates (CSH)

Balarezo J. Gabriel^{*1}, Pinto P. Henry¹

¹CompNano Research Group, School of Physical Sciences and Nanotechnology, Yachay Tech University, Urcuqui-Ecuador

*juan.balarezo@yachaytech.edu.ec

Abstract

Concrete is the second-most-used substance in the world after water, with more than 35 billion tons produced yearly[1]. Yet, understanding the atomic and mechanical properties of the main component of concrete, calcium-silicate-hydrate (C-S-H) cement hydrates—the complex binder phase of concrete—still poses a challenge[2][3].

In this work, we aim to investigate cement hydrates' atomic and mechanical properties leveraging density-functional theory (DFT) and machine learning (ML) tools. We use DFT to study the electronic structure, bonding, and mechanical responses of C-S-H at the atomic level [1][4]. Afterwards, we use ab initio molecular dynamics (AIMD) with ML to create a force field on the fly for C-S-H. This will allow us to accurately simulate and capture the complex atomic interactions of cement hydrates while reducing the computation time. By integrating DFT, AIMD, and ML, we seek to provide deeper insights into the fundamental properties of C-S-H and develop a predictive model that could inform the design of new cementitious materials.

Methodology

The calculations were performed using both DFTB+ and the Vienna *ab initio* simulation package (VASP). The GFN1-xTB method was used in DFTB+ for an initial relaxation since it is a fast, yet accurate method. Afterwards, more precise calculations were carried out using VASP, utilising the PBEsol+rVV10 functional and the projector augmented-wave (PAW) method. A kinetic cutoff energy of 800 eV was adopted to expand the wave functions in a basis set of plane waves in combination with a k-point grid of 0.06/Å to converge the total energy to < 1 meV/f.u. Additionally, all the structures were fully relaxed to forces < 0.01 eV/Å.



Cement Hydrates (C-S-H) are the main product of cement hydration. They have a disordered, gel-like structure with short range order resembling layered tobermorite.

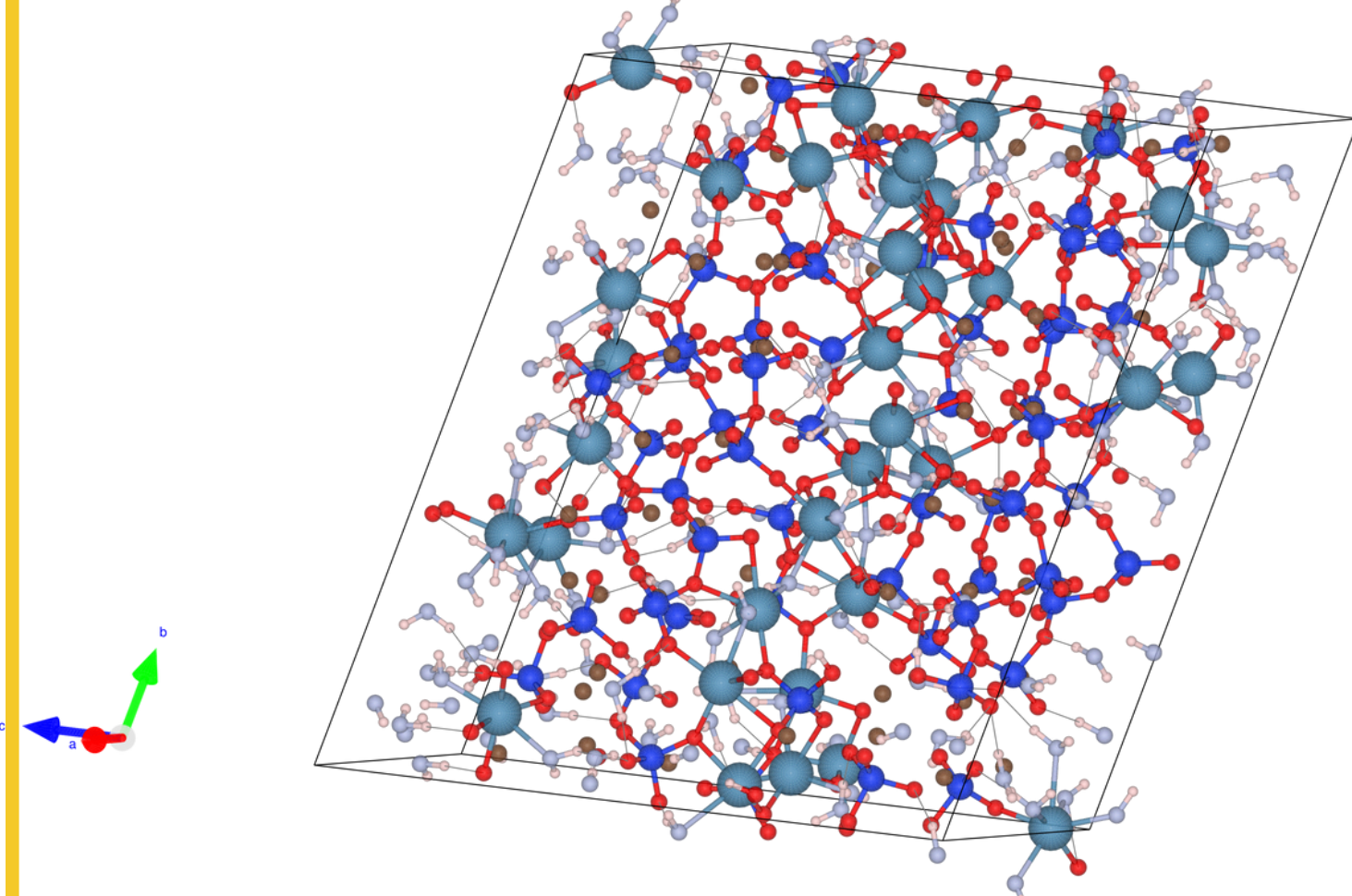


Figure 1. Molecular model of C-S-H proposed by [4]. Lavender and white spheres are oxygen and hydrogen from water molecules respectively; light blue and brown spheres are inter and intra-layer calcium ions, respectively; electric blue and red spheres are silicon and oxygen atoms in silica tetrahedra.

The ML force field (MLFF) was trained on-the-fly using AIMD for 50,000 steps, corresponding to 100 ps. Fifty structures were selected from the dataset for post-processing and model refitting.

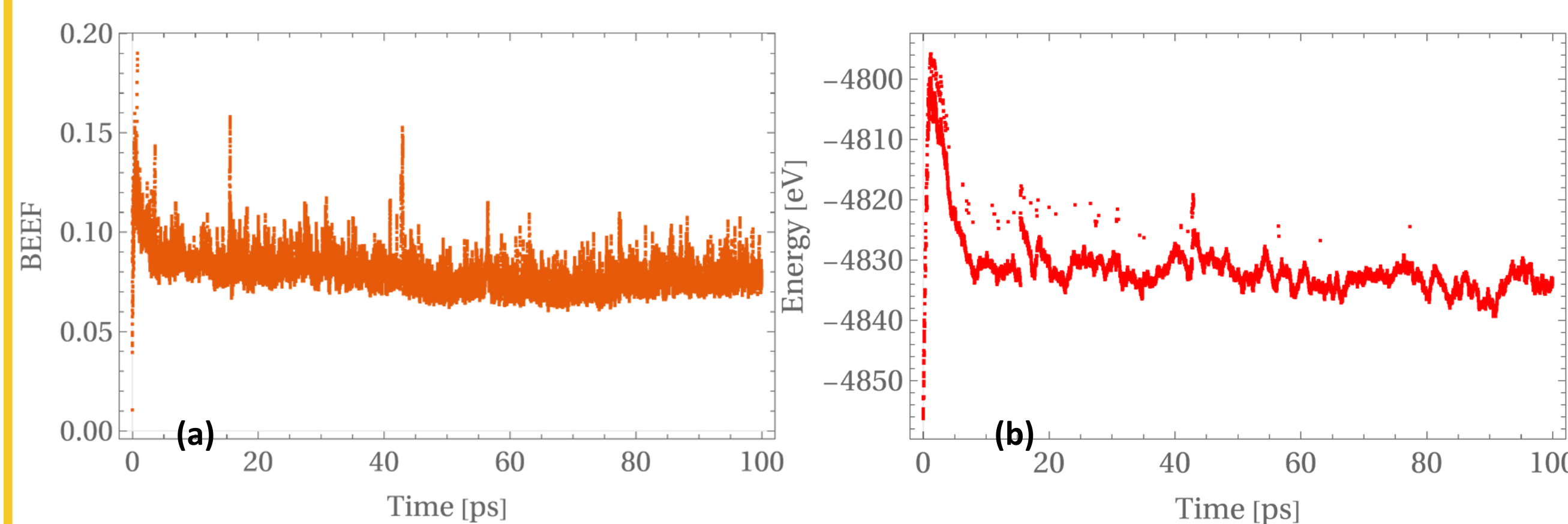


Figure 2. On-the-fly MLFF training. (a) presents the Bayesian error of the MLFF, which serves as an uncertainty metric, guiding the adaptive selection of configurations for *ab initio* evaluation. (b) shows the total energy of the system throughout the AIMD run, showing a fairly stable behaviour during the training phase.

Results

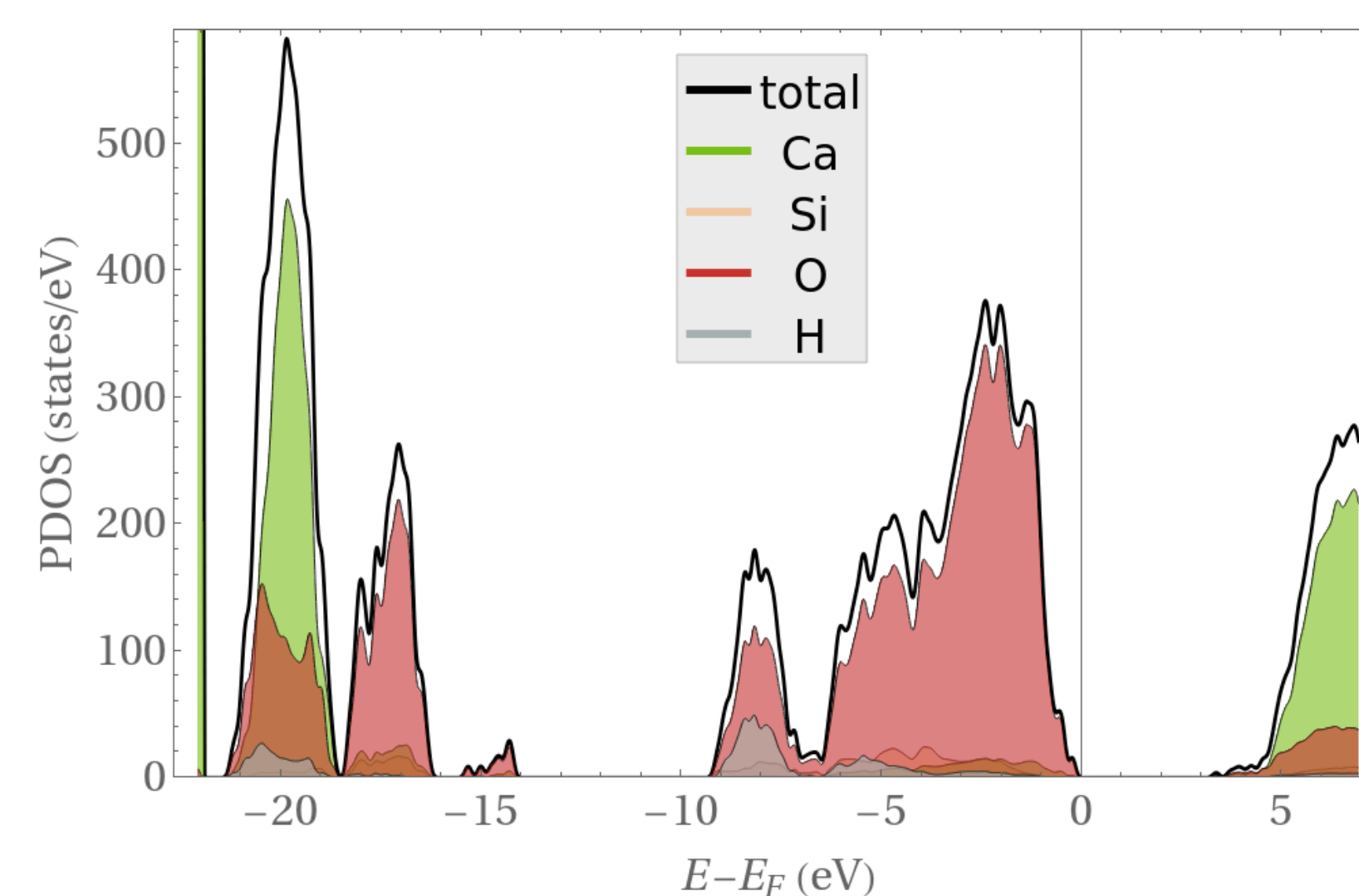


Figure 3. Density of states (DOS) and partial density of states (PDOS) of the C-S-H. Energy axis shifted so $E_f = 0$ eV. A band gap of ~ 3.1 eV is reported. Additionally, a main contribution of O 2p states in the valence band, confirming strong ionic and hydrogen bonding.

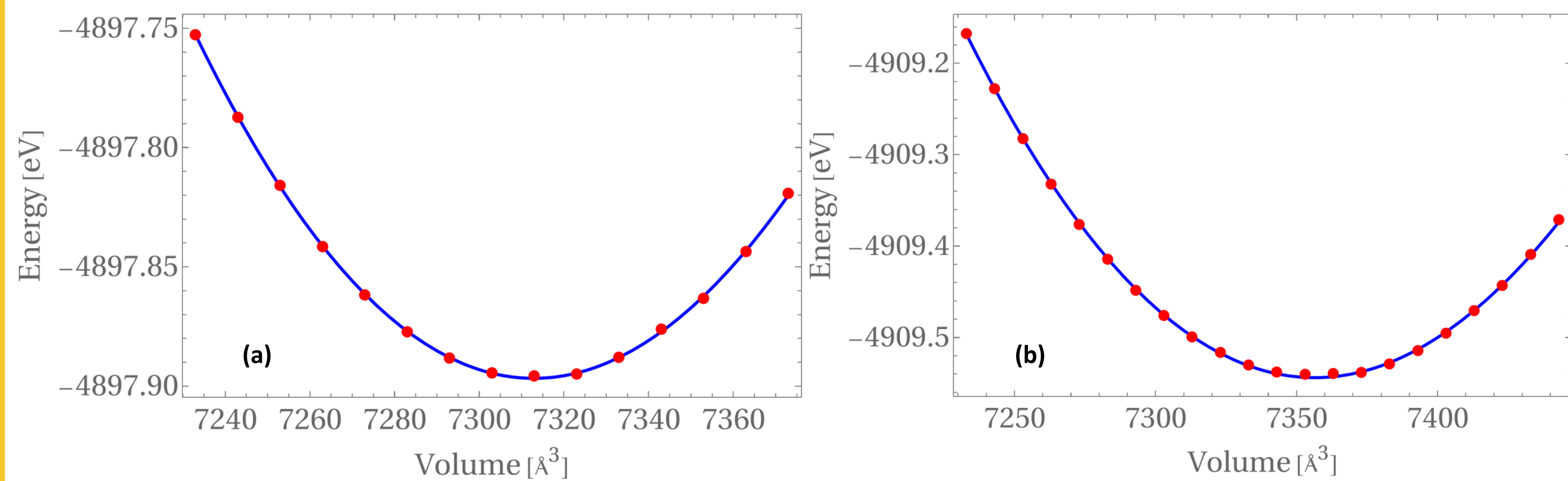


Figure 4: Equation of state (EOS) derived by fitting the volume-energy data to the third-order Birch-Murnaghan EOS. (a) shows the fitted curve using data obtained after full structural relaxation with our MLFF. (b) presents the corresponding fit using data generated following the same protocol as in (a), but subsequent to a simulated annealing process.

In Table 1 we present values for the optimal volume, total energy, bulk modulus and bulk modulus derivative for C-S-H, computed using our MLFF, and simulated annealing + MLFF. Additionally, we contrast our results with experimental values reported in the literature.

Table 1. Results obtained from calculations using a refitted MLFF and a simulated annealing + MLFF approach. The fourth column includes available experimental data reported by [5].

Parameter	Rifitted MLFF	Simulated Annealing	Literature
Volume (V)	7312.85 Å ³	7356.38 Å ³	N/A
Energy (E)	-4897.9 eV	-4909.54 eV	N/A
Bulk Modulus (B_0)	51.0563 GPa	55.3869 GPa	47±3 GPa
Bulk Modulus Derivative (B'_0)	8.52598	7.48163	4

It is evident that the bulk moduls computed with both methods are fairly in agree with values reported in the literature, providing a initial validation for our MLFF models.

Acknowledgments

The authors acknowledge the support of the School of Physical Sciences and Nanotechnology at Yachay Tech University.

References

- [1] Li, Z.; et al. 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.; et al. Cements in the 21st Century: Challenges, Perspectives, and Opportunities. Journal of the American Ceramic Society 2017, 100, 2746–2773.
- [4] Pellenq, R. J.-M.; et al. A Realistic Molecular Model of Cement Hydrates. Proceedings of the National Academy of Sciences 2009, 106, 16102–16107.

Conclusions

- 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.
- 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.
- 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.

