Molecular insight into crystal nucleation during cement hydration from ab initio machine-learning simulations

Shulai Guo [1,2], Wanqi Zhou [1], Pablo M. Piaggi [1,3]
[1] CIC nanoGUNE, Tolosa Hiribidea 76, San Sebastian 20018, Spain
[2] Tongji University, Shanghai, 201804, China
[3] Ikerbasque, Basque Foundation for Science, Bilbao 48013, Spain

s.guo@nanogune.eu & pm.piaggi@nanogune.eu

Abstract

Understanding the nucleation of layered silicate minerals from aqueous solution is of central importance in geochemistry, materials science, and engineering [1]. In particular, calcium silicate hydrate (C–S–H), one of the primary phases governing the mechanical properties of cement-based materials, forms through a crystallization process whose molecular mechanism remains poorly understood. While the thermodynamic endpoints, i.e, the dissolved ions in solution and the crystalline solid, are well characterized, the pathways by which hydrated ions assemble into clusters with emerging crystalline order remain under active debate [2].

Molecular simulations have provided valuable insight into this process. However, capturing nucleation requires large system sizes and long simulation times, alongside an accurate description of complex reactive events, pushing beyond the limits of standard simulation approaches. Here, we introduce an advanced framework that integrates reactive molecular dynamics with ab initio machine-learning potentials (MLPs) to address this challenge (Fig. 1).

We show that, although the classical reactive force field ReaxFF [3] is a useful tool for simulating large-scale chemical reactivity, it lacks the accuracy needed to capture the subtle energy differences governing ion association, proton transfer, silicate polymerization, and ultimately crystallization. To overcome these limitations, we follow the approach of Ref. [4] and combine ReaxFF-driven molecular dynamics with enhanced sampling methods to generate a diverse and representative training dataset, with energies and forces computed at the level of the SCAN density functional theory (DFT) approximation. From this dataset, we train a machine-learning potential, termed DP-SCAN, capable of accurately describing both solvated ions and their chemical transformations along the pathway leading to the formation of the solid.

Our results reveal substantial differences between ReaxFF and DP-SCAN in the free energy landscapes describing ion association and reactivity, leading to qualitatively distinct reactive pathways. These findings underscore the importance of quantum-accurate models for studying reactive crystallization in solution. More broadly, the development of such accurate and scalable potential opens the door to directly observing the full mineral nucleation process, from dissolved ions to layered crystalline structures, a regime that has so far remained inaccessible to both experiments and conventional simulations. By bridging electronic-structure accuracy with multi-nanosecond sampling, this work moves toward a definitive molecular-level understanding of how mineral phases emerge from disordered precursors in both geological and engineering environments.

References

[1] J. J. De Yoreo et al., Science, 349 (2015) aaa6760.
[2] X. M. Aretxabaleta et al., Nat Commun, 14 (2023) 7979.
[3] H. Manzano et al., J. Am. Chem. Soc., 134 (2012) 2208.
[4] P. M. Piaggi et al., Proc. Natl. Acad. Sci. U.S.A., 122(41) (2025) e2415663122.

Figures

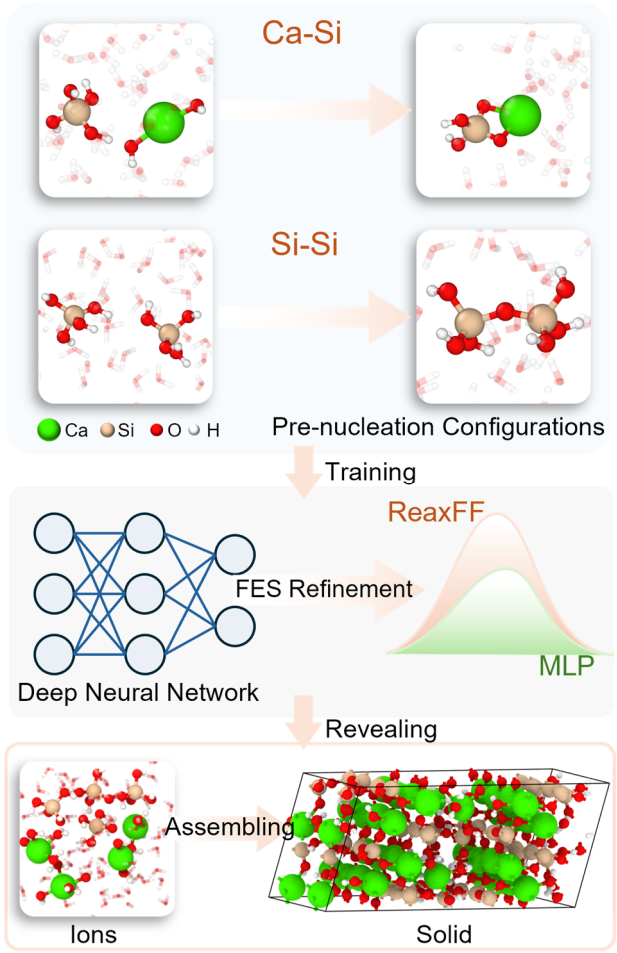

Figure 1. A machine-learning-driven framework for C-S-H nucleation study.