Atomistic simulation on the nucleation of C-S-H

eman ta zabal zazu
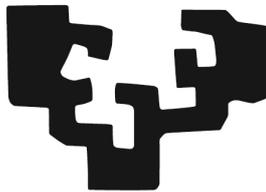
Universidad del País Vasco   Euskal Herriko Unibertsitatea

Xabier Méndez Aretxabaleta

Advisors:
Hegoi Manzano Moro
Iñigo Etxebarria Altzaga

Leioa, 2023

Verbum quod lex.

ABSTRACT

In this work, we have used computational methods to study the early stages of C-S-H nucleation at the atomic scale. Cement is the most widely used construction material in the world, and the C-S-H gel is the most important phase in cement, influencing the hydration rate and the development of the cement microstructure. However, many questions about C-S-H nucleation at the atomic scale remain unanswered. To address this, we have explored two main hypotheses for the formation of C-S-H: the nucleation of C-S-H from a portlandite monolayer, and the formation of C-S-H via pre-nucleation clusters (PNCs) from the main species in solution.

In the first hypothesis, we have studied the stability and phase transition from portlandite of a new theoretical polymorph of $Ca(OH)_2$ with the crystal structure of $Sr(OH)_2$, known as Phase V. We found that Phase V is metastable and that the phase transition between portlandite and Phase V has an energy barrier of 10.13 kcal/mol. However, when we studied the transformation of a portlandite monolayer into a C-S-H layer, we found that the internal energy of the transformation is favorable, but the free energy is unfavorable. This suggests that other mechanisms may also be involved in the formation of C-S-H and that the transformation of a portlandite monolayer into a C-S-H layer may not be the dominant pathway for C-S-H nucleation.

In the second hypothesis, we have used evolutionary algorithms to search for the most stable clusters for different C-S-H stoichiometries. We found that the most stable clusters for small sizes contain a silicate tetrahedron bonded to a 6- or 7-folded Ca atom. For medium size clusters, it was found that silicate dimers are the most stable arrangement and therefore, they would form at the early stages of cement hydration. For larger sizes, we found that the most stable clusters contain silicate trimers but it is already known that trimers are not formed during cement hydration, probably because the aggregation of smaller clusters with dimers does not allow the formation of trimers. We also found that the dehydration of the clusters stabilizes the dimers and that trimers are less likely to form in less hydrated clusters. We also identified a very dehydrated cluster that looks similar to a tobermorite layer, suggesting that such clusters may form at the early stages of cement hydration.

Overall, our research has provided new insights into the mechanisms behind the nucleation of C-S-H and has highlighted the potential im-

portance of PNCs in this process. These findings could have important implications for the design and optimization of cement-based materials and could guide future experimental and computational studies on C-S-H nucleation. In addition, our work has demonstrated the utility of computational methods for studying the atomic-scale processes involved in cement hydration and has shown the potential of evolutionary algorithms for exploring the structure of C-S-H clusters. However, further research is needed to fully understand the role of PNCs in the formation of C-S-H, and to determine the relative importance of the different mechanisms that may be involved in C-S-H nucleation.

RESUMEN

En este trabajo, hemos utilizado métodos computacionales para estudiar las primeras etapas de la nucleación del C-S-H a escala atómica. El cemento es el material de construcción más utilizado en el mundo, y el gel C-S-H es la fase más importante del cemento, ya que influye en la velocidad de hidratación y en el desarrollo de la microestructura del cemento. Sin embargo, muchas cuestiones sobre la nucleación del C-S-H a escala atómica siguen sin respuesta. Para abordar esta cuestión, hemos explorado dos hipótesis principales para la formación del C-S-H: la nucleación del C-S-H a partir de una monocapa de portlandita, y la formación del C-S-H mediante clusters de prenucleación (PNCs) a partir de las principales especies en disolución.

En la primera hipótesis, hemos estudiado la estabilidad y la transición de fase desde la portlandita de un nuevo polimorfo teórico de $Ca(OH)_2$ con la estructura cristalina de $Sr(OH)_2$, conocido como Fase V. Encontramos que la Fase V es metaestable, y que la transición de fase entre la portlandita y la Fase V tiene una barrera energética de 10,13 kcal/mol. Sin embargo, cuando estudiamos la transformación de una monocapa de portlandita en una capa de C-S-H, descubrimos que la energía interna de la transformación es favorable, pero la energía libre es desfavorable. Esto sugiere que otros mecanismos pueden estar también implicados en la formación del C-S-H, y que la transformación de una monocapa de portlandita en una capa de C-S-H puede no ser la vía dominante para la nucleación de C- S-H.

En la segunda hipótesis, hemos utilizado algoritmos evolutivos para buscar los cúmulos más estables para distintas estequiometrías del C-S-H. Encontramos que los clusters más estables para tamaños pequeños contienen un tetraedro de silicato unido a un átomo de Ca de 6 o 7 dímeros. Para los grupos de tamaño medio, se descubrió que los dímeros de silicato son la disposición más estable y, por lo tanto, se formarían en las primeras etapas de la hidratación del cemento. Para tamaños mayores, encontramos que los clusters más estables contienen trímeros de silicato, pero ya se sabe que los trímeros no se forman durante la hidratación del cemento, probablemente porque la agregación de clusters más pequeños con dímeros no permita la formación de trímeros. También descubrimos que la deshidratación de los conglomerados estabiliza los dímeros y que es menos probable que se formen trímeros en los conglomerados menos hidratados. Por último, identificamos un cúmulo muy deshidratado de aspecto similar al de una capa de tobermorita, lo que

sugiere que tales cúmulos pueden formarse en las primeras etapas de la hidratación del cemento.

En general, nuestra investigación ha proporcionado nuevos conocimientos sobre los mecanismos que subyacen a la nucleación de C-S-H, y ha puesto de relieve la importancia potencial de las PNC en este proceso. Estos hallazgos podrían tener importantes implicaciones para el diseño y optimización de materiales basados en cemento, y podrían guiar futuros estudios experimentales y computacionales sobre la nucleación de C-S-H. Además, nuestro trabajo ha demostrado la utilidad de los métodos computacionales para estudiar los procesos a escala atómica implicados en la hidratación del cemento, y ha puesto de manifiesto el potencial de los algoritmos evolutivos para explorar la estructura de las agrupaciones C-S-H. Sin embargo, es necesario seguir investigando para comprender plenamente el papel de los PNCs en la formación del C-S-H, y para determinar la importancia relativa de los diferentes mecanismos que pueden estar implicados en la nucleación del C-S-H.

LABURPENA

Lan honetan, metodo konputazionalak erabili ditugu C-S-H-aren nukleazioaren lehen etapak eskala atomikoan aztertzeko. Zementua da munduan gehien erabiltzen den eraikuntza-materiala, eta C-S-H gela da zementuaren faserik garrantzitsuena. Fase hau hidratazio-abiaduran eta zementuaren mikroegituraren garapenean eragiten du. Hala ere, eskala atomikoan C-S-H-aren nukleazioari buruzko galdera askok erantzunik gabe jarraitzen dute. Gai horri heltzeko, bi hipotesi nagusi aztertu ditugu C-S-H-aren eraketarako: batetik C-S-H-aren nukleazioa portlandita-geruza bakarretik abiatuta, eta bestetik C-S-H-aren sortzea nukleazio aurreko klusterren bidez (PNCen bidez) disoluzioan dauden espezie nagusietatik abiatuta.

Lehenengo hipotesian, Sr(OH)$_2$ren egitura kristalino duen Ca(OH)$_2$ren polimorfo teoriko berri baten egonkortasuna aztertu dugu. Fase berri hau Phase V deitu dugu eta portlanditatik fase-trantsizioa aztertu dugu ere. Phase V fasea metaegonkorra dela ikusten dugu, eta portlanditaren eta Phase V fasearen arteko fase-trantsizioak 10,13 kcal/moleko energia-hesia duela. Hala ere, portlandita-geruza bakar bat C-S-H motako geruza bihurtzea aztertu dugunean, eraldaketaren barne-energia aldekoa dela ohartu gara, baina energia librea kontrakoa dela ere. Horrek iradokitzen du beste mekanismo batzuk ere inplikatuta egon daitezkeela C-S-H-aren eraketan, eta portlandita-geruza bakarra C-S-H motakogeruza bihurtzea agian ez dela bide nagusia C-S-Hren nukleaziorako bide nagusia.

Bigarren hipotesian, algoritmo ebolutiboak erabili ditugu C-S-H-aren estekiometria desberdinetarako kluster egonkorrenak bilatzeko. Tamaina txikietarako kluster egonkorrenek silikatozko tetraedro bat dute, 6 edo 7 koordinazioko Ca atomo bati lotua. Tamaina ertaineko taldeentzat, silikatozko dimeroak antolamendurik egonkorrena direla ikusi da, eta, beraz, zementuaren hidratazioaren lehen etapetan sortuko direal antzeman daiteke. Tamaina handiagoetarako, kluster egonkorrenek silikatozko trimeroak dituzte, baina jakina da trimeroak ez direla zementuaren hidratazioan sortzen, seguru asko, dimeroak dituzten kluster txikiagoak gehitzeak ez duelako trimeroak sortzea ahalbidetzen. Era berean, klusterretan deshidratazioak dimeroak egonkortzen dituela eta hain hidratatuta ez dauden klusterretan trimeroak sortzeko probabilitate gutxiago dagoela deskubritu dugu. Azkenik, tobermoritazko geruza baten antzeko itxura duen kluster deshidratatu bat identifikatu dugu, eta horrek iradokitzen du kluster horiek zementuaren hidratazioaren lehen etapetan sor daitezkeela.

Oro har, gure ikerketak ezagutza berriak eman ditu C-S-H-aren nukleazioan dauden mekanismoei buruz, eta prozesu horretan PNCek izan dezaketen garrantzia nabarmendu du. Aurkikuntza horiek inplikazio garrantzitsuak izan ditzakete zementuan oinarritutako materialen diseinuan eta optimizazioan, eta C-S-H-aren nukleazioari buruzko etorkizuneko azterketa esperimentalak eta konputazionalak gida ditzakete. Gainera, gure lanak erakutsi du metodo konputazionalek zementuaren hidratazioan inplikatutako eskala atomikoko prozesuak aztertzeko duten erabilgarritasuna, eta agerian utzi du algoritmo ebolutiboek C-S-H klusterren egitura aztertzeko duten ahalmena. Hala ere, ikerketa asko egiteko dago, PNCek C-S-H-aren eraketan duten zeregina erabat ulertzeko, eta C-S-Hren nukleazioan inplikatuta egon daitezkeen mekanismoen garrantzi erlatiboa zehazteko.

Note to the reader

Music has been one of my greatest companions in writing this thesis. There are times when it has helped me to concentrate, times when it has made me reflect, and times when it has helped me to overcome difficulties. That is why, at the beginning of each chapter, I have left a suggestion for listening. I understand that for some people reading with music can be distracting, but I hope that for the rest of you it will help you in the laborious task of reading a doctoral thesis. May music always be with us.

Contents

Chapter 1

Introduction

The main objective of this thesis has been to delve into the atomic mechanisms that occur in cement hydration, specifically in the early stages of hydration, where nucleation is really important. For this purpose, different atomistic simulation methods have been used, both *ab initio* and classical methods. These methods allowed us to calculate energy differences, energy barriers, free energy differences, and other thermodynamical properties that let us evaluate different proposals for the crystallization of the main phase of hydrated cement. In other words, the studies presented in this thesis are entirely computational but the obtained results are, as far as possible, compared with experimental results to validate the modeling results.

Cement hydration is a very complex process at the atomic level which contains different sub-processes which are not very well understood. Therefore, let us first introduce the material itself and its most important hydrated phase, the so-called Calcium Silicate Hydrate (C-S-H).

1 | INTRODUCTION TO CEMENTITIOUS MATERIALS

Cement is defined as a substance that acts as a binder and sticks materials together. Because of that property, cementitious materials had been used in construction since the emergence of the first civilizations. The Egyptians already used mortar (a mixture of sand and cementitious material) to join stone blocks and erect their prodigious constructions [1]. Part of one of the pyramids of Giza (2600 B.C.) was built with concrete (hydraulic cement with fine and coarse aggregates), and the mural of Thebes (1950 B.C.) preserves scenes of men making concrete and applying it to a building site. However, the oldest concrete construction is the floor of a hut in Lepensky Vir (nowadays Serbia), dated 5600 BC [2, 3]. Later, Greek and Roman builders discovered that certain materials from volcanic deposits, mixed with limestone, sand, and water, produced a mortar of great strength, capable of resisting the action of fresh and salt water [4]. This is how hydraulic cements were improved and widespread. The Parthenon of Athens (built in 425 BC) and the Roman coliseum (built in 82 AD) are examples of the great durability of these mortars.

After the fall of the western Roman Empire, most of the knowledge about these construction materials was lost. During the dark ages of the middle ages, the quality of mortars became very poor until the year 1200, when it is used again for the construction of large works such as the Salisbury Cathedral in England, whose foundations are made of concrete [5]. It was not until 1754 that cement took a great leap in quality due to the investigations of John Smeaton, who. with the mandate to build a new lighthouse in Eddystone (England), made several experiments and compared the quality of cement produced with lime from different locations [6]. Despite this progress, the mortars used at this time were still made with lime and pozzolana, very similar to those used by the Romans. In 1796 it was discovered that cement could be produced by calcining nodules of argillaceous limestone. Due to this discovery, a lot of patents of different cement recipes were made at the beginning of the XIX. century. One of these was from Joseph Aspdin who in 1824 patented his recipe for cement with a predetermined combination of limestone and clay as well as the manufacturing process [7]. Aspdin called this cement "Portland Cement" because at that time, Portland stone was very famous for its hardness and durability. Although the mineralogy of As-
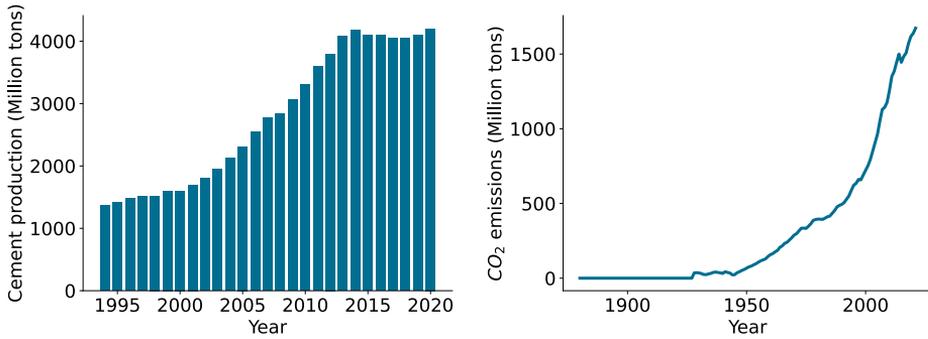
Figure 1.1: The figure on the left shows the production of cement for each year in millions of tons. The figure on the right shows the CO_2 emissions produced by the production of cement for each year in millions of tons.

pdin's Portland cement is very different from the Portland cement we know today, his patent was an essential step in developing it.

Cement and its derivatives are the most used manufactured products in the world with an annual production of 4200 million metric tons (in 2020) [8]. Moreover, cement production has increased substantially in the last 20 years (see Fig. 1.1) and is likely to continue to rise due to the economic development of massive countries such as India, Brazil, and countries in Africa. This massive production has been used to build all kinds of infrastructure that have led the world to an economic development never seen before. However, such a large production generates a considerable environmental impact, accounting for 5 to 10% of all CO_2 emissions. In the last few years, the CO_2 emissions related to the production of cement have gone over 1500 million metric tons [9]. Due to the annual increase in production, substantial advances in sustainability are needed to achieve the USA International National Agency (IEA) proposed emission reductions for the cement industry by 2050 [10]. For this reason, in recent years a tremendous financial effort has been made by governments and private cement companies to encourage cement research.

1.1 | CEMENT PRODUCTION

There are several types of hydraulic cement such as Portland, pozzolan-lime, calcium sulphoaluminate and geopolymeric cement [11]. Among those, Portland cement and its derivatives are the most common ones.

Recolection of raw materials   **Grinding**

- Limestone
- Clays

Raw blend

Gypsum added   **Kiln at 1400°C**

Cement clinker
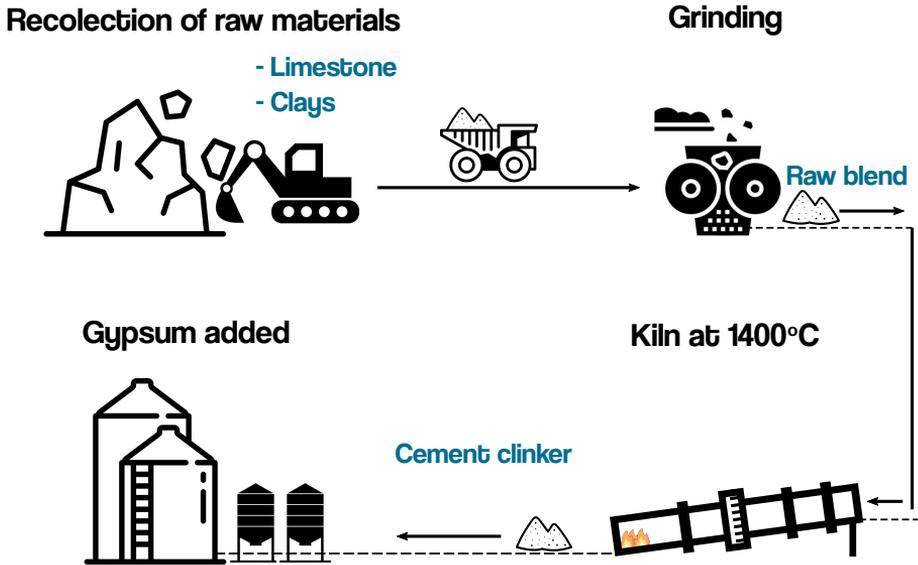
Figure 1.2: Schematic representation of the different processes in cement production.

The production of *Ordinary Portland Cement* (OPC) begins with the rec-
ollection of the raw materials (Limestone and Clays). Then, those materi-
als are mixed with appropriate proportions and ground and stirred until
a homogeneous and fine powder is obtained, usually called a "raw blend".
This raw blend is introduced into a Kiln, which is a cylindrical oven of
about 4 meters in diameter and 150 meters in length that rotates and is
slightly inclined. The mixture enters at one end and gradually moves
until it reaches the other end. Inside the kiln, the mixture is heated up
to 1400°C. This high temperature melts the raw materials and produces
several chemical reactions which create the cement clinker. Finally, the
clinker nodules are milled to get the desired particle size, and additives
such as gypsum are added [12]. A scheme of cement production can be
seen in Fig. 1.2. Depending on the exact composition of the raw blend
and the additives employed, different types of OPC are marketed.

1.2 | CEMENT CLINKER

As mentioned before, a mix of limestone ($CaCO_3$) and clays are intro-
duced in the Klin where it is heated up to 1400°C. As the temperature of
the mixture rises, different chemical reactions occur. From 70 to 110°C
free water previously absorbed in the minerals is evaporated. From 400
to 600° clays decomposed in the basic oxide components such as SiO_2

and Al_2O_3. At higher temperatures, from 600 to 1100°C limestone reacts with the previously released silica dioxide to form belite (Ca_2SiO_4) while the excess limestone decomposes into calcium oxide (CaO) and carbon dioxide (CO_2). This reaction together with the fuel burned for heating up the kiln is the main responsible for the carbon dioxide emissions in the production of cement. Finally, from 1100 to 1400°C some of the Belite is melted and reacts with the calcium oxide to form alite (Ca_3SiO_5). Therefore, the main components of the cement clinker are alite and belite , accounting for 60% and 10%, respectively [13]. The amount of alite and Belite in the composition depends on the type of cement, but alite is predominant in most cases. Despite these majority phases, other mineral phases such as tricalcium aluminate ($Ca_3Al_2O_6$) and aluminoferrite ($Ca_2(Al_{1-x}Fe_x)_2O_5$) are formed in the process.

2 | CEMENT HYDRATION

The term hydration in cement chemistry refers to the chemical reactions, physical changes, and thermodynamic processes that take place when the clinker is put into contact with water. The hydration of cement is a fairly simple process from a technological perspective. It consists on mixing water with cement to obtain a cement paste that will harden over time. However, at the molecular level, it is a really complex process due to the large number of reagents in the clinker. The hydration process results in the formation of an amorphous paste that sets and hardens. The process is complex due to the great number of variables involved. The cement powder contains a wide variety of components, clinker phases and additives, each of them with impurities and random particle sizes and shapes, distributed heterogeneously among the cement, and with different reaction rates. Other external factors, such as the water to cement ratio (w/c, in % wt) and curing temperature, play also a key role in the final properties.

This peculiar process has been investigated from the early 1900s [14] and continues to be the subject of much research [15, 16]. To reduce the level of complexity of cement hydration, the hydration of each of the clinker phases is often studied separately, both experimentally and computationally. Alite and belite similarly react with water. Both dissolve with water and form portlandite ($Ca(OH)_2$) on the one hand and a nearly amorphous calcium silicate hydrate of variable stoichiometry (the
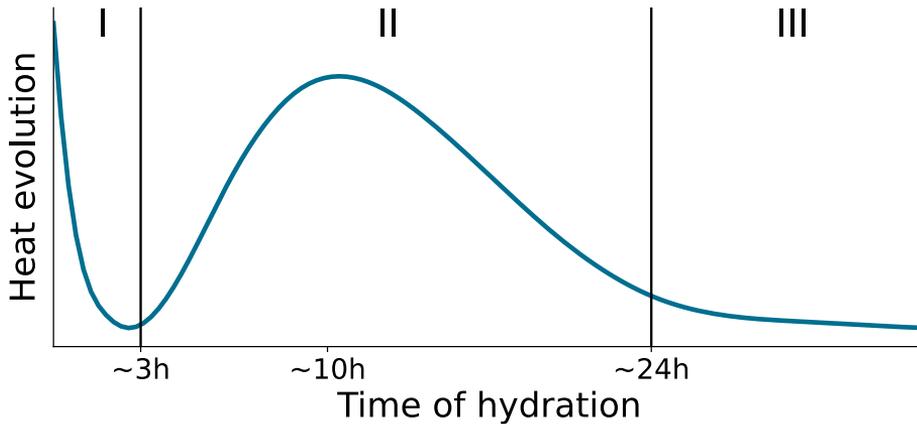
Figure 1.3: Illustration of typical heat evolution in alite hydration. Three different periods can be identified from this heat evolution.

C-S-H) on the other. Despite the similarity in the hydration process, the hydration rates are very different. At 28 days of hydration, 70% of the Al-lite has been consumed. In contrast, only 30% of belite has reacted with water after that time period due to its lower reaction rate. Furthermore, after one year of hydration, the alite has almost completely reacted while 10% of the Belite is still unreacted.

The other two main components of cement clinker, tricalcium alumi-nate ($Ca_3Al_2O_6$) and aluminoferrite ($Ca_2(Al_{1-x}Fe_x)O_5$) react very fast with water in a process called flash setting. This is not desired since it re-duces the workability of the paste and hinders the homogeneous mixing of cement with water. The addition of gypsum solves this problem since it retards the reaction process of these phases. The reaction products of these processes are ettringite and monosulphoaluminates with partial substitution of iron due to the aluminoferrite phase.

2.1 | THE HYDRATION PERIODS

There have been many different techniques used to follow the hydration process, like conductimetry, thermogravimetry, light and electron mi-croscopies, NMR, X-Ray diffraction and many more [16]. Due to the heat evolution curve measured for alite hydration, most of the authors dis-tinguish three different main hydration periods (see Fig. 1.3), although sometimes it has been divided into up to seven periods.
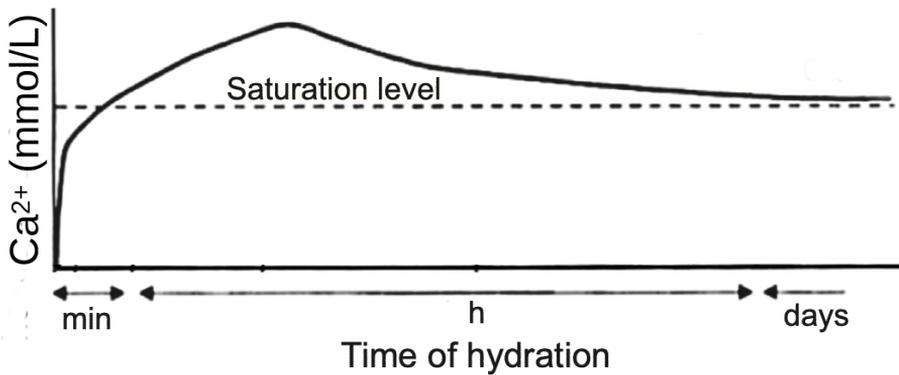
Figure 1.4: Ca^{2+} concentration evolution in solution. Figure adapted from Gartner et al. [17]

2.1.1 | PERIOD I (INDUCTION PERIOD)

The first period goes from the start of the hydration up to the end of the induction period. At the beginning of hydration, the reaction is highly exothermic for the first few minutes and the heat evolution starts in a high pick. This is due to the fast dissolution of the clinker grains releasing Si^{+4}, Ca^{2+} and OH^- ions to solution (see Fig. 1.4). The high pick is followed by a slow deceleration in the heat evolution that lasts until a few hours. At a macroscopic level, the cement mixture remains plastic and workable during this period.

At the atomistic level, it is no fully understood the reason of the deceleration. It was believed that a metastable C-S-H formed in alite surface thus creating an impermeable layer [18]. This layer prevents the alite from hydrating until this layer became more permeable or breaks due to osmotic pressure. However, this protective uniform and tightly bounded layer has never been found even with the most modern methods. Moreover, this approach fails to explain the annealing experiments on alite, which exhibit much longer induction periods without a change in the driving force [19]. That is why this theory is losing the support of the scientific community in the last decades.

Therefore, another theory based on the dissolution controlled by undersaturation is the more plausible one to explain the induction period. Several articles have shown experimentally that the dissolution rate of alite decreases non-linearly as the undersaturation of the system decreases, that is to say, as the system moves towards equilibrium [20, 19,

21]. These experiments have all been done in very diluted conditions where the precipitation of hydrates is prevented. Therefore, they show that the decrease in the heat evolution of the induction period is not related to the formation of a layered hydrated but to the undersaturation of the system.

2.1.2 | PERIOD II (ACCELERATION AND DECELERATION PERIOD)

This period consists of two subperiods, the acceleration period followed by the deceleration period. In this period, the main peak of the heat release corresponds to the massive precipitation of C-S-H and portlandite. This is coupled with an increased dissolution of alite, which is mainly responsible for the heat release measured. There are several theories for the mechanism of heat evolution during this period.

The first one is the diffusion layer hypothesis [22]. Here, the authors state that C-S-H precipitates on the surface of the cement grains, forming a thickening layer that ultimately becomes thick enough to act as a diffusion barrier. However, there is some evidence that does not support this hypothesis. There is no activation energy change at the peak time, thus not making a transition of mechanism plausible [23]. Moreover, it was found that the diffusion coefficients of the reactive species of C-S-H had to be varied over one order of magnitude to fit the particle size distribution (PSD) experiment from Costoya [24, 25].

Another theory proposes that the progressive impingement of C-S-H from neighboring grains leads to the transition from acceleration to deceleration. As the hydrates impinge against each other the surface in contact with water decreases, thus decreasing the reaction rate [25]. However, it has been shown that the main hydration peak barely evolves from water to cement ratio 0.4 to 0.8 [26]. The water to cement ratio is directly related to the interparticle distance and therefore if the impingement was the main mechanism, it would affect the hydration peak.

The confined-growth hypothesis was also presented, where the growth of C-S-H was restricted to one micron from the grain surface in agreement with SEM observations [27]. Nevertheless, the imposition of the distance limit for hydrate formation is completely arbitrary. Nicoleau and Nonat [28] proposed that the dissolution controlled by undersaturation presented for the induction period (period I) continues to determine

the rate of reaction throughout the main heat evolution peak. They claim that the acceleration period is due to the opening of etch pits which increases the alite surface area. At some point, the surface area decreases, for example by etch pits coalescence, thus leading to the deceleration period.

Finally, the theory by Bazzoni [29] states that clusters of C-S-H (needles) progressively nucleate on the cement grains and grow rapidly to a certain length. Then, when the surface of the grain is mostly covered there is a decrease in C-S-H growth, thus creating the deceleration period. The needle model quantitatively shows that the peak coincides with the time at which most of the needles have nucleated and are in their fast growth regime [30, 31].

2.1.3 | PERIOD III (DIFFUSION PERIOD)

This period is defined as the period over the first day of hydration. The slow decrease in the heat evolution from here is mainly explained by the space-filling hypothesis [32]. The are two ways in which the lack of space can affect the hydration of C-S-H, taking into account that it can only precipitate in water-filled regions.

The first way is that the volume of water available for precipitation is depleted. Recently it has been shown, that the gap between the hydrated product and the reacting grain can have an important impact on the water content in the grain surface since the water flows from the external space (the space outside of the C-S-H shells) to the gap [30]. The other way in which the lack of space limits the hydration is by the limiting critical pore size. There is evidence of the importance of the lack of water-filled pores greater than a limiting critical pore size in crystal growth. This means that below the critical pore size, the crystal needs too much energy to continue growing due to the increase of its curvature, which requires a higher concentration of ions in a solution in equilibrium with it.

In this period of continuous deceleration, the microstructure of the C-S-H gel is fully developed, and the gel becomes denser. Moreover, other crystalline phases, specially portlandite, continue growing. It is considered that the hydration is virtually finished after one year, though the process could continue over years.
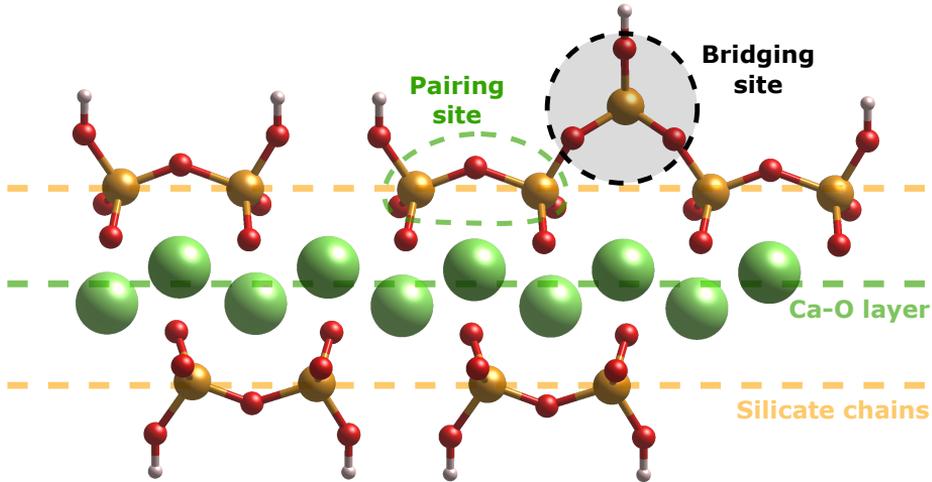
Figure 1.5: Graphical representation of a C-S-H layer model where a Ca-O layer is sandwiched by silicate chains of different lengths. Ca, Si, O and H atoms are represented by green, orange, red and pink atoms

3 | C-S-H STRUCTURE

As presented in the previous sections, the main hydration product of cement is Calcium Silicate Hydrate. In the cement chemistry notation, C stands for CaO, S for SiO_2 and H for H_2O. That is why the Calcium Silicate Hydrate with a lack of a fixed stoichiometry is noted as C-S-H. Following this nomenclature portlandite, alite and belite are called CH, C_3S and C_2S respectively. C-S-H constitutes up to the 70% of the volume of the final solid phase, thus being the main responsible for the properties of the material.

The structure of C-S-H is not fully understood due to its poor crystallinity. This has been observed by X-ray diffraction (XDR) where C-S-H shows very few and diffuse picks, which implies that C-S-H does not have any 3D long-range order [33]. It also means that the exact atomic structure of C-S-H cannot be resolved by conventional XDR methods. The resolution of C-S-H structure is even more complicated as the chemical composition varies not only between different samples but also within the same sample. Therefore, simple parameters like the Ca/Si ratio have been used to characterize the composition of C-S-H. Local energy dispersive X-ray analysis (EDS) has proven that in a 100 nm sample of pure C_3S paste the Ca/Si ratio can vary from 1.2 to 2.1 with a mean Ca/Si ratio of 1.78 [34].

Although C-S-H not having long-range order, different experiments
such as NMR, IR-spectroscopy, X-ray diffraction and SANS [35, 36, 37, 38,
39, 40] have revealed that locally its structure has similarities with crys-
talline calcium silicate hydrates like wollastonite, jennite, hilebrandite,
and specially tobermorite [41]. The tobermorite supergroup is a fam-
ily of minerals with a layered structure. This family contains different
members of the tobermorite group together with similar minerals like
plombierite, clinotobermorite and riversideite. All these minerals con-
tain a layer of seven folded Ca-O sandwiched between silicate chains.
The distinction between the different types of tobermorite minerals is
due to their basal spacing between layers: 1.4 nm, 1.1 nm and 0.9 nm.

Various methods like trimethylsilylation, chromatography or NMR
were used to study the type of silicate species present in C-S-H [42, 43].
These studies revealed that only silica chains with a fixed length of $3n-1$
units were present in C-S-H, where $n = 1, 2, 3, ...$ Each unit is an or-
thosilicic acid group, $Si(OH)_4$, which polymerizes via condensation re-
actions to form the chains. These chains present the so-called drier-
ketten structure where silicate dimers are linked together by a bridging
silicate [34, 44, 45]. In tobermorite, these chains run parallel to the b
crystallographic axis and as denoted by the drierketten structure, the
arrangement repeats every 3 silicates as seen in Fig. 1.6. Two consecu-
tive silicate units are bonded to the central Ca-O layer by sharing two of
their four oxygen atoms. These silicate units are denoted as pairing sites.
The remaining silicate unit, denoted as the bridging site, is not bonded to
the central Ca-O layer and shares two oxygen atoms with adjacent sili-
cate units, thus creating a chain. The structural module, named "complex
module", comprises the central Ca-O layer with a zig-zag arrangement
of the seven-fold coordinated Ca atoms sandwiched between the silicate
chains [46]. It is important to emphasize that although these silicate
chains are infinite in the tobermorite crystal, in C-S-H some bridging
site silicates are missing, thus creating finite length silicates of length
$3n - 1$, as shown in Fig. 1.5. Moreover, it has been shown experimen-
tally that in cured C-S-H 60% of the silicate species are dimers and the
other 40% are longer chains with a mean chain length of about 3.3[42].

One of the problems with tobermorite serving as a C-S-H model is the
disagreement in the Ca/Si ratio. While C-S-H has a mean Ca/Si ratio of
1.78 tobermorite has a Ca/Si ratio of 0.83. To overcome this problem,
Taylor and Howison proposed in 1956 a C-S-H model based on the to-
bermorite crystal but with missing bridging site silicates and replacing
them with Ca ions. This is supported by NMR and infrared spectroscopy
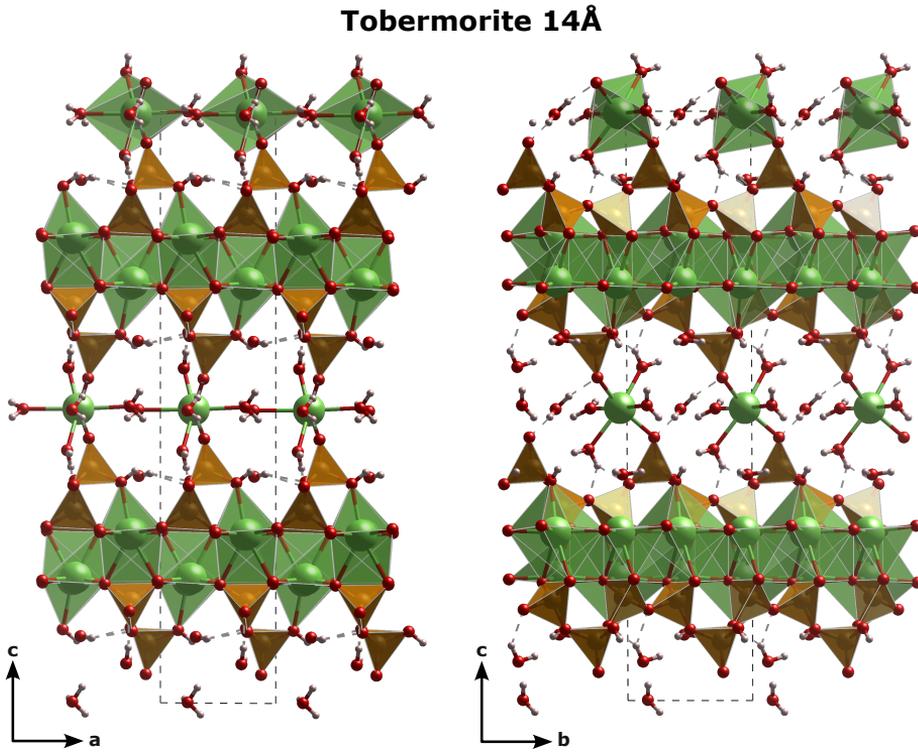
Tobermorite 14Å

Figure 1.6: Experimental atomic structure of Tobermorite

measurements, which indicate the existence of Ca-OH groups in C-S-H [47, 35, 48]. It was found that in a C-S-H with Ca/Si ratio of 1.78, about 23% of the charge of the Ca^{2+} ions is compensated by OH groups. Removing the bridging silicate and adding Ca ions provokes an increase in the Ca/Si ratio from 0.83 up to 1.5 [49]. However, the obtained Ca/Si ratio with this model was still far from the mean C-S-H Ca/Si ratio of 1.78. Therefore, this model was extended by adding a hydrogen atom to the end of the finite silicate chains where the bridging site is missing instead of adding a Ca ion [39]. This way allows us to adjust the Ca/Si ratio to the desired one by replacing as many bridging silicates as needed. This explanation of C-S-H is the basis of all the models presented lately [50, 51].

Recently, the pair distribution function from high energy X-ray experiments on C-S-H suggested that a portlandite signal must be added to the clinotobermorite signal to obtain a compatible result [52, 40]. From these experiments, it is concluded that portlandite mono- and bi-layers coexist between tobermorite-like layers in C-S-H. This implies that the
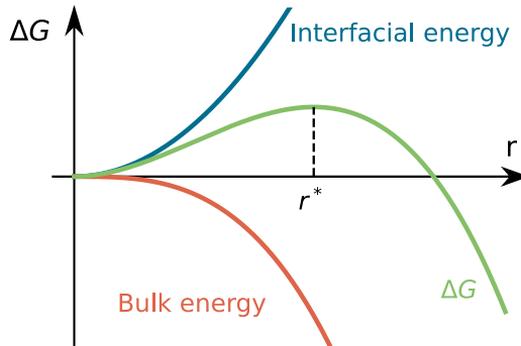
Figure 1.7: Schematic representation of the sum of the bulk free energy and the interfacial free energy. The critical nucleus size r^* is the particle size where the free energy starts to decrease and thus, the cluster can grow spontaneously.

accepted tobermorite-based models may be reconsidered, bringing back the possibility of a mixed tobermorite/portlandite C-S-H. Be that as it may, although there is agreement that C-S-H has tobermorite-like motifs, the exact structure of C-S-H is still far from being unequivocally determined.

4 | NUCLEATION OF THE C-S-H GEL

Nucleation is the first step in the formation of a new phase of matter. This is the process whereby nuclei (seeds) act as templates for later crystal growth. The nucleation processes are classified into two main categories: Homogeneous and heterogeneous nucleation. In homogeneous nucleation, the nucleation probability is homogeneous throughout the system whereas heterogeneous nucleation forms at structural inhomogeneities (container surfaces, impurities, grain boundaries, dislocations). Heterogeneous nucleation, which induces nuclei at the interface of vessel walls, dust particles, and impurities can occur at a lower supersaturation compared to homogeneous nucleation, where large supersaturations are needed to initiate the nucleation [53, 54].

4.1 | CLASSICAL NUCLEATION THEORY (CNT)

Traditionally Classical Nucleation Theory (CNT) has been used to explain nucleation. CNT was developed by Volmer and Weber, Becker and

Döring, and Frenkel [55, 56, 57]. It is based on the condensation of vapor
to a liquid. However, this process can be extended to other liquid-solid
equilibrium systems, such as crystallization from melts and solutions.
The thermodynamic description of the nucleation is based on the free
energy (ΔG) which is defined as the sum of bulk free energy and the
interfacial free energy between the nucleus and the solution. Therefore,
the free energy change of a system during homogeneous nucleation from
an ideal gas of a spherical nucleus of radius r is defined as:

$$\Delta G = 4\pi r^2 \gamma_p + \frac{4}{3}\pi r^3 \Delta G_V \qquad (1.1)$$

where γ_p is the specific surface energy per unit and ΔG_V is the bulk
free energy per volume. This factor is given by

$$\Delta G_V = \frac{-k_B T \ln(S)}{\nu} \qquad (1.2)$$

where k_B is the boltzman constant, S is the vapor supersaturation de-
gree of the solution and ν is the molar volume. Therefore, the first term
in Eq. 1.1 represents the interfacial energy that dominates for small par-
ticles or clusters and the second term represents the bulk energy of the
cluster that dominates for bigger particles. Consequently, there is a crit-
ical nucleus size $r^* = 2\gamma/\Delta G_V$ where the trend of the total free energy
changes and starts to decrease. Particles with a size below the critical
nucleus size will dissolve whereas there would be a tendency to grow
in particles with a bigger size than the nucleus critical size. The nucle-
ation rates predicted by CNT are in fair agreement with water vapor
system [58]. Moreover, the critical supersaturation can be predicted for
the nucleation of ethanol, methanol, water and ethanol-water systems
at low water concentrations [59]. However, the predictions of the nucle-
ation rates for these systems are erroneous by many orders of magnitude.
That is why for most of the systems (including C-S-H) CNT is not valid
and non-classical nucleation has to be used.

4.2 | NON-CLASSICAL NUCLEATION THEORY

In recent years nucleation has been experimentally studied at the micro-
scopic level and even visualized with high spatial and temporal resolu-
tion. This has resulted in nucleation being revealed as a more diverse and
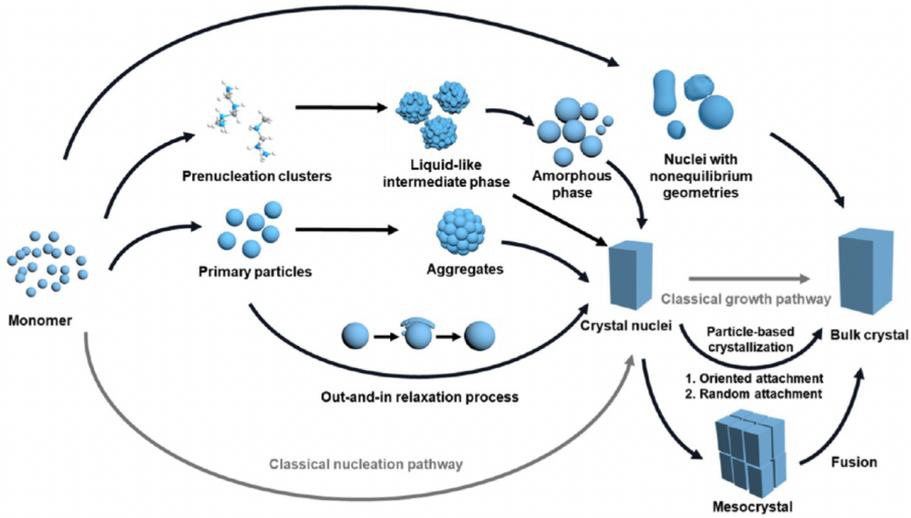
Figure 1.8: Schematic representation of the different possible pathways for non-classical nucleation.

complex process than that represented by CNT [60]. In fact, the main non-classical nucleation pathways are the formation of pre-nucleation clusters PNC, the aggregation of primary particles and the formation of nuclei with nonequilibrium geometries (shown in Fig. 1.8).

The PNC pathway was first discovered in calcium carbonates, where nucleation is not triggered by the energy difference between the bulk and surface energy (as in CNT), but by an entropy increase due to the hydration water loss of hydrated anions. These PNC have the following characteristics [61]:

- They are composed of the constituent atoms, molecules or ions of a forming solid, yet can also contain additional chemical species.
- They are small and thermodynamically stable, thus having no phase boundary between the cluster and the solution.
- PNCs are molecular precursors to the nucleating phase and hence participate in the phase separation.
- They are very dynamic and change their configuration on a scale of picoseconds.
- The structure of the PNCs can encode structural motifs related to the corresponding crystalline polymorphs.

In the mentioned case of calcium carbonate, the PNCs were evidenced employing a combination of potentiometric titrations
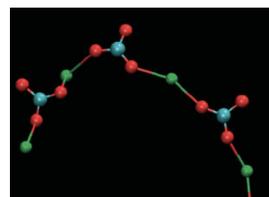
and analytical ultracentrigufation [63, 64].
It was demonstrated that calcium and car-
bonate ions were bonded at a ratio of 1:1
prior to nucleation and that those formed
clusters were larger than simple ion pairs.
The occurrence of larger associates was
corroborated using cryogenic transmission
electron microscopy (cryo-TEM) [65]. Us-
ing the experimentally determined ion as-
sociation it was discovered that those clusters were indeed thermody-
namically stable [63]. All this evidence leads to the definition of PNC
which then would form a liquid-like amorphous calcium carbonate [66].
Later, the use of atomistic simulation allowed the study of these PNCs by
simulating a range of concentrations and pH values [62]. It was revealed
that the PNCs tend to form ionic polymers, composed of alternating cal-
cium and carbonate ions, with a dynamic topology of chains, branches
and rings. By examining the free energy of these PNCs it was observed
that the radius of gyration could be changed by almost a factor of 2 at an
almost zero energy cost. These PNCs were named dynamically ordered
liquid-like oxyanion polymers (DOLLOP) and it is proposed that these
DOLLOPs may appear also in other systems [67].

Another type of non-classical nucleation is the assembly of primary
particles. In CNT the particles with a size below the critical nucleus size
tend to dissolve. However, if the dissolution rate is slower than the colli-
sion rate between those particles, is possible that more than two subcrit-
ical nuclei syncretized into a stable postcritical nucleus. For example, it
has been shown using cryogenic transmission electron microscopy the
existence of primary particles consisting of Fe^{2+} and Fe_3^+ before mag-
netite nucleation. Looking at the free energies of these primary particles
they show that the nucleation from primary particles favors a direct crys-
tallization rather than an amorphous phase [68]. This type of nucleation
pathway has also been observed in the initial crystal growth stages of
gold nanoparticles by in situ small-angle X-ray scattering (SAXS). These
gold primary particles presented a size of 0.8 nm in an aqueous solution.
With the rapid decrease in the concentration of nanoparticles the par-
ticle eventually increased from 0.8 to 1.7 nm, which indicates that the
nucleation of gold is performed by the aggregation pathway [69].

Although CNT assumes closely packed 3D ordered arrangements for
molecules within the nuclei to minimize surface energy, recent research
revealed that deviation from the CNT allows for the nuclei to adopt the

lattice of the emerging crystal with equilibrium shape [70]. This type of nucleation create nuclei with nonequilibrium geometries. This has been seen in the crystallization of protein apoferritin using atomic force microscopy (AFM) [71]. They showed that instead of forming closely pack particles, molecules are arranged along the <110> direction, resulting in arrays with 4-8 molecules. This kind of nucleation pathway has also been found during Au nucleation with aberration-corrected TEMs.

Although in this section the main non-classical nucleation pathways have been explained (shown in Fig. 1.8), there are some other nucleation pathways that have been recently observed [60] such as cluster-cloud mediated nucleation, multistep nucleation in solid-solid transitions or nonclassical homogeneous nucleation from vapor.

4.3 | C-S-H nucleation

The early stages of cement hydration which contains the nucleation of C-S-H have not been studied in depth until a few years ago. In 1999 Garrault-Gauffinet et al. used conductivity results to study the homogeneous nucleation within the CNT framework [72]. They concluded that the interfacial energy of C-S-H is very low giving a very low nucleus critical size. However, it is known that nucleation seeding with calcium silicate hydrate accelerates the hydration period, which means that heterogeneous nucleation is the main nucleation mechanism in C-S-H [73, 74].

The more recent work by Krautwurst et al. [75] used time-resolved potentiometry and turbidimetry combined with dynamic light scattering, small-angle X-ray scattering, and cryo-TEM to study the nucleation of C-S-H. By the characteristic evolution of the turbidity, they claim that the nucleation of C-S-H can be divided in three steps. The first turbidity change in stage I corresponds to the formation of amorphous calcium silicate spheroids with diameters of around 50±10 nm. The chemical composition of the spheroids is that of C-S-H but with a lower calcium concentration. The solution reaches the solubility limit of C-S-H during stage II, where the chemical potential of the spheroids and C-S-H are similar. In stage III, the transformation of the spheroids in C-S-H occurs. These C-S-H crystallites would then aggregate to form C-S-H platelets. In summary, they proposed a two-step nucleation method in which first a dense spheroid forms and from that spheroid C-S-H starts to crystallize. Anyway, these studies were done in synthetic C-S-H and

therefore, the nucleation in real cement hydration could be different. The nucleation of C-S-H was also studied in presence of polycarboxylates and the formation of the spheroids is reported to be similar, yet the crystallization is delayed, probably due to slower diffusion of water from the spheroid [76]. These amorphous spheroids could be framed within the PNC nucleation pathway explained in section 4.2.

Numerical models have also been used to study C-S-H nucleation. Models such as boundary nucleation and growth (BNG) or microstructure-based kinetics (MBK) have been tested to predict cement hydration rate [77, 78, 79, 80]. In addition, a model with colloidal nanoparticles was used to study both primary and secondary nucleations on C-S-H [81, 82]. Primary nucleation represents the nucleation in the absence of parent crystal structures, while secondary nucleation is the birth of new crystals in the presence of parent crystals of the same substance.

Going back to the PNC pathway, a few studies based on atomistic simulation have suggested clusters which then would later participate in nucleation. Precursor nanoclusters have been proposed based on the structure of tobermorite and jennite, and their aggregation has been studied using *ab initio* methods [83]. They suggested that these PNCs could aggregate differently depending on the Ca^{2+} content in the solution. Density functional theory (DFT) was also used to find the lowest energy complexes of the main dissolved species in C-(N)-A-S-H systems [84] (where N stands for nitrogen oxide and A for aluminum oxide) and then their pair-wise interactions Gibbs free energies were computed. They showed that the Ca-Ca interactions were the most favorable followed by the strong ion-paring reaction between Ca and the silicates. Moreover, they claim that the calcium species are largely driving the early-stage formation mechanism of C-S-H and that the Ca-Ca association interaction instigates the formation of the silicate chains. Similar to this work, Li. et al explored the energy difference in terms of Ca coordination in $[Ca(H_2O)_n(SiO_2(OH)_2)]$ and $[Ca(H_2O)_n(SiO(OH)_3)]^+$ complexes by ab initio metadynamics [85]. They show that the most favorable Ca state is that with one Ca-O-Si bond and five Ca-O-water bonds, giving a total of six-fold coordination. However, higher coordination numbers have not been explored.

Although both the real and computational experiments described above make it possible to create some nanoscale models, they offer very little insight into the atomic mechanisms of hydration. The understanding of such atomic mechanisms is crucial because how species in solution in-

teract and rearrange to form the first clusters will set the path for the posterior nucleation and growth.

5 | MOTIVATION AND OBJECTIVES

The nucleation of C-S-H is one of the most important processes in cement hydration and would determine the growth and formation mechanisms. Therefore, understanding how C-S-H nucleates is crucial to improve the characteristics of futuristic cements. As explained in the previous section, in the last two decades several studied have been done around this subject. The experiments carried out provide evidence about whether the nucleation is homogeneous or heterogeneous, about the effects of nucleation seeds or about properties of the environment in which the nucleation process takes place [72, 73, 74]. Moreover, recent studies have stated that C-S-H nucleation is a non-classical nucleation process [75]. All these investigations helped us to understand mainly macroscopic properties of nucleation which have served to make some conjectures about the atomic process of nucleation. However, these studies do not offer any evidence about the atomic processes that occur in C-S-H nucleation.

Therefore, the main motivation of this thesis is to try to understand these processes at the atomic level and for that, there is no better tool than atomistic simulation. As we have seen in section 4.2, there are several nucleation pathways within the non-classical nucleation. That is why in this thesis two hypothetical C-S-H nucleation pathways have been studied.

The first one is based on the similarity of the Ca-O layer in C-S-H and the Sr-O layer in bulk $Sr(OH)_2$. This similarity may seem irrelevant but since Ca and Sr are isoelectronic portlandite may transform into a structure with the atomic arrangement of bulk $Sr(OH)_2$. If this transformation is possible it may be reasonable to think that a portlandite layer may be transformed into a C-S-H layer by the interaction of the silicate in solution. This hypothesis would be further explained later.

The other studied nucleation pathway is the pre-nucleation cluster (PNC) pathway. This type of pathway has been explored for other materials, mostly $CaCO_3$. This pathway is based on the formation of clusters that then would aggregate to form bigger particles of C-S-H. Therefore,

the main goal of this nucleation pathway is to find the structure of the PNCs that could be formed in the early stages of cement hydration. To see how the level of hydration or the size of the cluster affects their structures, different clusters with different stoichiometries have been studied using evolutionary algorithms.

6 | Thesis outline

After a brief introduction to the C-S-H structure and its nucleation, in chapter 2 all the computational methods used in this thesis are explained in detail. From very common methods like Density Functional Theory (DFT) to more specific methods used in this thesis like Targeted Molecular Dynamics (TMD) or sketch-map.

Chapter 3 deals with the transformation of portlandite to Phase V of $Ca(OH)_2$. Phase V has been discovered in this thesis and its stability is analyzed before proposing a transformation mechanism. The transformation mechanism has been studied using Cell Variable Nudge Elastic Band (CV-NEB). The results of this chapter serve as a foundation for the research proposed in Chapter 4.

Following the results of chapter 3, in chapter 4 the same transformation is studied in a portlandite layer. In this chapter, we propose a mechanism for the formation of a C-S-H layer starting from a portlandite layer which undergoes a topochemical transformation by means of silicate dimers. This has been studied using enhanced sampling techniques like targeted molecular dynamics (TMD) and umbrella sampling (US).

Finally, in chapter 5 the possible pre-nucleation clusters (PNC) of C-S-H are studied using evolutionary algorithms. The best clusters for different stoichiometries are studied and a formation mechanism for a tobermorite-like cluster is proposed. Additionally, the aggregation of some of these clusters is studied using molecular dynamics and the results are compared with experimental results.

CHAPTER 2

METHODS

The Tunnel
The Other Side
Tord Gustavsen trio

1 | ATOMISTIC MODELING

The modeling of materials at the atomistic scale is all about defining the interactions between the different atoms in the system by a mathematical model. Once those interactions are defined it is easy to calculate the atomic forces on the system or to calculate its energy. Knowing the forces and total energy of a system will allow us to minimize those quantities or to predict how would the system behave in real life. It goes without saying that to perfectly model systems with more than a few atoms is rather impossible with the computational power we have nowadays. Therefore, some approximations are needed to model larger systems in a faster way. Essentially, all existing methods for modeling materials differ in the degree of approximations they use. That is to say, the more approximations the method uses the larger the system it can model using the same computational time. Moreover, the methods with more approximations are capable of modeling a system for longer time scales. As shown in 2.1 the existing methods can be divided into three categories: the quantum methods, the classical methods and the mesoscale methods. The systems modeled in this thesis range from 5Åto 10nm and the simulated time never exceeded a couple of ns. Therefore, in this thesis, DFT has been used as the main quantum modeling method and the
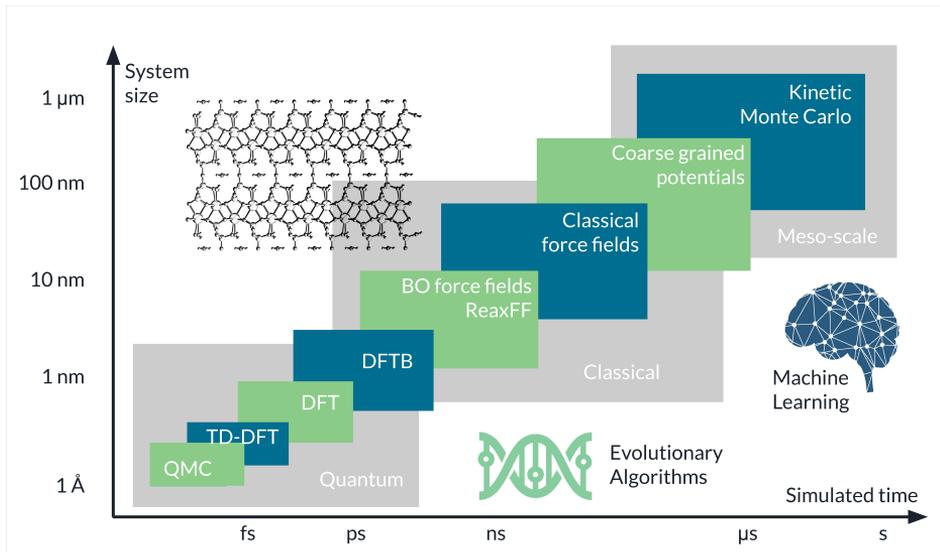
Figure 2.1: Scheme of time and size scales that can be modelled with each simulation technique.

ReaxFF force-field as the main classic method.

2 | DFT

Quantum methods, usually called *ab-initio* methods, are based on directly solving the time-independent Schrödinger equation to get the ground state energy of a system:

$$\hat{H}\ket{\Psi} = E\ket{\Psi} \tag{2.1}$$

where $\hat{H}$ is the hamiltonian operator, E is the energy and $\ket{\Psi}$ is a vector that represents the wave function in the Hilbert space. The most interesting quantity that can be computed by solving this equation is the ground-state energy E_0, which can be calculated by finding the wave function that minimizes the energy in eq. 2.1 as

$$E_0 = \min_\Psi \bra{\Psi}\hat{H}\ket{\Psi} \tag{2.2}$$

where $\bra{\Psi}\hat{H}\ket{\Psi}$ represents the inned product in the Hilbert space,

$$\langle \Psi | \hat{H} | \Psi \rangle = \int \Psi^*(\mathbf{r}) \hat{H}(\mathbf{r}) \Psi(\mathbf{r}) d\mathbf{r}. \qquad (2.3)$$

Any system is composed of M positively charged ions and N negatively charged electrons that interact via the Coulomb interaction. Therefore, the hamiltonian can be separated in different terms as

$$\hat{H} = \sum_i^N \frac{\hat{\mathbf{p}}_i^2}{2} + \sum_I^M \frac{\hat{\mathbf{P}}_I^2}{2M_I} + \sum_i^N \sum_I^M \frac{Z_I}{|\hat{\mathbf{r}}_i - \hat{\mathbf{R}}_I|} + \frac{1}{2} \sum_{i \neq j} \frac{1}{|\hat{\mathbf{r}}_i - \hat{\mathbf{r}}_j|} + \frac{1}{2} \sum_{I \neq J} \frac{Z_I Z_J}{|\hat{\mathbf{R}}_I - \hat{\mathbf{R}}_J|}$$
$$(2.4)$$

where electrons are denoted by lowercase subscripts (i and j), ions are denoted by uppercase subscript (I and J), Z_I and M_I represent the charge and mass of the I-th ion, $\hat{\mathbf{r}}_i$ and $\hat{\mathbf{p}}_i$ express the position and momentum operators of the i-th electron and $\hat{\mathbf{R}}_I$ and $\hat{\mathbf{P}}_I$ denote the position and momentum operators of the I-th ion respectively.

The first two terms in the hamiltonian represent the kinetic operators for the electrons ($\hat{T}_{\text{elec}}$) and the ions ($\hat{T}_{\text{ion}}$) respectively. The remaining terms express the operators for the ion-electron interaction ($\hat{V}_{\text{ion}-\text{elec}}$), electron-electron interaction ($\hat{V}_{\text{elec}-\text{elec}}$) and the ion-ion interaction ($\hat{V}_{\text{ion}-\text{ion}}$). Hence, the previos equation can be condensed as

$$\hat{H} = \hat{T}_{\text{elec}} + \hat{T}_{\text{ion}} + \hat{V}_{\text{ion}-\text{elec}} + \hat{V}_{\text{elec}-\text{elec}} + \hat{V}_{\text{ion}-\text{ion}}. \qquad (2.5)$$

As can be noticed, the many-body problem is very well defined by the fundamentals of quantum physics. However, the analytical solution for most systems is not accessible, those being limited to the two-body hydrogen-like systems. Moreover, the exact numerical solution for the time-dependent Schrodinger equation is computationally very expensive and is mostly limited to very few electrons. Hence, some approximations are needed to solve the Schrodinger equation for molecules and crystals.

One of the mainly used approximations is the Born-Oppenheimen approximation where the wave function in the former Schrödinger equation can be separated into two terms, the electronic term and the ionic term which can be solved separately. This can be justified since the typical electronic velocities are 10^3 times larger than the typical ionic velocities. Hence, ions can be computed fixed with an effective infinite mass ($M \rightarrow \infty$) when computing the electronic states. This approach simplifies dramatically the many-body problem since $\hat{T}_{\text{ion}}$ can be neglected

when calculating the electronic state. Therefore, the electronic Hammiltonian $\hat{H}_{el}$ depending parametrically on the ionic configuration $\hat{\mathbf{R}}$ can be written as:

$$\hat{H}_{el} = \hat{T}_{\text{elec}} + \hat{V}_{\text{ion}-\text{elec}}(\hat{\mathbf{R}}) + \hat{V}_{\text{elec}-\text{elec}}. \tag{2.6}$$

It may be helpful later to define the ion-electron interaction as a sum of a potential v_{ext} acting on each of the electrons in the system.

$$\hat{V}_{\text{ion}-\text{elec}}(\hat{\mathbf{R}}) = \sum_i v_{ext}(\mathbf{r}_i) = \sum_i \left(-\sum_I \frac{Z_I}{|\mathbf{r}_i - \mathbf{R}_I|} \right) \tag{2.7}$$

Once the electronic ground state is computed the ionic hamiltonian can be defined in terms of the ground-state electronic energy $E_{el} * 0(\hat{\mathbf{R}})$ as

$$\hat{H}_{ion} = \hat{T}_{\text{ion}} + \hat{V}_{\text{ion}-\text{ion}} + E_{el}^0(\hat{\mathbf{R}}) \tag{2.8}$$

where $\hat{V}_{\text{ion}-\text{ion}}$ represents the classical pairwise Coulomb interaction between two static ions. The ionic hamiltonian allows us to compute the total ground energy of the system in terms of the ionic configuration which can be used to define a potential energy surface (PES). This is extremely useful to compute the local stability, geometry and transition states of the system.

2.1 | FUNCTIONALS

Before proceeding with the formal formulation of the theory, it is convenient to introduce explicitly what a functional is. As we will see later, the wave function of the system $\Psi(\mathbf{r_1}, ..., \mathbf{r_N})$ can be completely determined by the electron density, or in other words, the wave function is a functional of the electron density. That is, Ψ is a function of the $3N$ spatial coordinates, but a functional of the density $n(\mathbf{r})$, which is represented as follows: $\Psi = \Psi[n](\mathbf{r_1}, ..., \mathbf{r_N})$.

In the same way, a function $f(x)$ gives us a rule that takes us from one number (x) to another $(y = f(x))$, a functional $F[n]$ takes as input a function $n(x)$ and provides the rule to go from this function (or set of functions) to a number. Another tool that will be needed later is the functional derivative. The objective is to determine the change of

the functional $F[n]$ under a variation of the function $n(r)$, which is its argument. Considering a functional of the form [86]

$$F[n] = \int f\left(x, n(x), \frac{dn(x)}{dx}, \frac{d^2n(x)}{dx^2}, ..., \frac{d^mn(x)}{dx^m}\right) dx \qquad (2.9)$$

its functional derivative can be defined as

$$\frac{\delta F[n]}{\delta n(x)} = \frac{\partial f}{\partial n} - \frac{d}{dx}\frac{\partial f}{\partial n'} + \cdots + (-1)^m \frac{d^m}{dx^m}\frac{\partial f}{\partial n^{(m)}}. \qquad (2.10)$$

2.2 | ELECTRONIC DENSITY

Since the final intention is the use of the electron density as the main variable, it is necessary to define it in terms of the wave function, which is obtained after solving the electronic Schrödinger equation (Eq. 2.6). The square modulus of the wave function $|\Psi(\mathbf{r_1}, ..., \mathbf{r_N})|^2$ represents the probability density of finding the electrons in the configuration $(\mathbf{r_1}, ..., \mathbf{r_N})$. Therefore, the electron density can be defined as the number of electrons per unit volume that can be found around the spatial position $\mathbf{r} = \{\mathbf{r_1}, ..., \mathbf{r_N}\}$. Following this definition, the electron density can be written as the mean value of the operator $\hat{n} = \sum_i \delta(\mathbf{r} - \mathbf{r_i})$:

$$n(\mathbf{r}) = \langle\psi|\,\hat{n}\,|\psi\rangle = N \int d^3\mathbf{r_2} \int d^3\mathbf{r_3} \cdots \int d^3\mathbf{r_N}\, |\psi(\mathbf{r}, \mathbf{r_2}, \mathbf{r_3}, \dots, \mathbf{r_N})|^2. \qquad (2.11)$$

This equation implies that the electron density is a functional of the wave function. Since the wave function is normalized to 1, the condition $\int d^3\mathbf{r}n(\mathbf{r}) = N$ is satisfied. It should be noted that by taking the electron charge as the unit, $n(\mathbf{r})$ describes both the electron distribution and the charge distribution.

Using the definition of the electronic density (2.11), the ion-electron interaction $V_{ion-elec} = \langle\psi|\,\hat{V}_{ion-elec}\,|\psi\rangle$ can be redifined as $V_{ext}[n]$ and written as a functional of the electron density:

$$V_{ion-elec}[n] = V_{ext}[n] = \sum_{i=1}^{N} \int d^3\mathbf{r_1}... \int d^3\mathbf{r_N} |\psi(\mathbf{r}, \mathbf{r}_2, \mathbf{r}_3, \ldots, \mathbf{r}_N)|^2 v_{ext}(\mathbf{r}_i)$$

$$= \int d^3\mathbf{r} n(\mathbf{r}) v_{ext}(\mathbf{r})$$

<div align="right">(2.12)</div>

There is not an analytic solution for the kinetic energy ($\hat{T}_{elec}$ of the electrons and the interaction between them ($\hat{V}_{elec-elec}$). However, various methods to solve this problem will be discussed later.

2.3 | THE HOHENBERG-KOHN THEOREMS

The next approximation came in 1964 by Hohenberg and Kohn and defines two fundamental mathematical theorems for the development of Density Functional Theory (DFT)[87]. Shortly, following the work of Thomas and Fermi in 1927, Holenberg and Kohn stated that the many-body problem could be formulated in terms of the electron density $n(\mathbf{r})$. That is to say, they demonstrated that all the properties of the electron system are completely and uniquely determined by the ground-state electron density $n_0(\mathbf{r})$. This approach made the problem easier to solve since it decreases dramatically the dimensionality of the system. Within this new framework, they defined the energy of the electron system in its lowest energy as a functional of the electron density:

$$E_{el} \equiv\,= E_{el}[n] = F[n] + \int v_{ext}(\mathbf{r}) n(\mathbf{r}) d\mathbf{r}, \qquad (2.13)$$

where $F[n]$ is a universal functional of an electron gas and thus, independent of any external potential acting on the electrons. They also showed that the ground-state energy of the electronic system is the global minimum value of the energy functional in Eq. 2.13. Therefore, the electron density that minimizes the energy functional is that of the exact ground state. Although the system was simplified by reformulation of the electronic many-body problem in terms of the electron density no practical methodology was given for calculating the ground-state electron density, since the exact dependence of $F[n]$ respect to $n(\mathbf{r})$ is unknown.

2.4 | KOHN-SHAM EQUATIONS

To overcome this problem, Walter Kohn and Lu Jeu Sham gave a new approach to the problem. This consists of mapping the system of many interacting electrons moving in the external potential onto an auxiliary system made of as many non-interacting electrons moving in an effective potential, i.e. the electronic screened external potential, on the condition that both systems have the same ground-state electron density. Therefore, the energy of this auxiliary system can be expressed by the Kohn-Sham energy:

$$E_{KS}[n] = T_s[n] + E_H[n] + E_{xc}[n] + V_{ext}[n], \qquad (2.14)$$

where T_s is the single-particle kinetic energy, E_H is the Hartree energy and E_{xc} is the exchange-correlation energy. Within the Kohn-Sham representation of non-interacting electrons, the electron kinetic energy can be written as the sum of the kinetic energies of single electrons

$$T_s = -\frac{1}{2} \sum_{i=1}^{N} \langle \Phi_i | \nabla_i^2 | \Phi_i \rangle = -\frac{1}{2} \sum_{i=1}^{N} \int d^3\mathbf{r} \Phi_i^*(\mathbf{r}) \nabla^2 \Phi_i(\mathbf{r}), \qquad (2.15)$$

where $|\Phi_i\rangle$ is the Kohn-Sham single particle state of i-th electron. The many-body electron-electron interaction is replaced by the Hartree and exchange-correlation energy. The Hartree energy represents the electrostatic interaction between the electronic density with itself. Therefore, instead of summing the coulombic interaction between all the electron pairs, we can define the Hartree energy as

$$E_H = \frac{1}{2} \iint d^3\mathbf{r}_1 d^3\mathbf{r}_2 \frac{n(\mathbf{r}_1)n(\mathbf{r}_2)}{|\mathbf{r}_1 - \mathbf{r}_2|}. \qquad (2.16)$$

The exchange-correlation energy accounts for all the quantum many-body effects missing in the Hartree energy. Namely, the exchange energy [88] accounts for the anti-symmetric property of the many-body electronic wave function with respect to particle exchange, since electrons are defined as indistinguishable fermions [89]. The correlation energy [90] accounts for the interaction effects beyond the independent

single-electron picture within the Hartree-Fock approximation. Unfortunately, the exact expression of the exchange-correlation energy functional is unknown, and therefore, appropriate approximations must be made. The success of DFT relies on the ability to account for these energy terms, even if only approximately. One of the most important approximations for the exchange-correlation energy $E_{xc}[n]$ is the local density approximation (LDA) in which the density is treated like a constant locally [91, 92]. Although the electron density in many systems is far from being locally constant, LDA has proven to be very useful. This is mainly since with this approximation the correlation energy is overestimated, but the exchange energy is underestimated, and thus both errors are compensated [93]. Due to the success of the LDA, the generalized-gradient approximation (GGA) was proposed in which the local functional was replaced by a semi-local functional depending also on the neighboring points [94].

Following the Hohenberg-Kohn theorems, the electron density minimizing the electronic energy functional in Eq. 2.13 is the ground-state one. Thereby, we minimize this functional using Lagrange multipliers under the restriction that the Kohn-Sham states are orthonormal, i.e. $\langle\psi_i|\psi_j\rangle = \delta_{ij}$:

$$\frac{\delta}{\delta\psi_i^*(\mathbf{r})}\left[F[n] + \int V_{ext}(\mathbf{r})n(\mathbf{r})d\mathbf{r} - \sum_j \varepsilon_j\left(\int |\psi_j(\mathbf{r}')|^2 d\mathbf{r}' - 1\right)\right] = 0. \tag{2.17}$$

Deriving this equation and taking into account the definition of $F[n]$ in Eq. 2.14 together with Eqs. 2.15 and 2.16 the Kohn-Sham single-electron Schrödinger equation is obtained:

$$\hat{H}_{KS}\psi_i(\mathbf{r}) = \left(\frac{\hat{\nabla}^2}{2} + V_{KS}\right)\psi_i(\mathbf{r}) = \varepsilon_i\psi_i(\mathbf{r}), \tag{2.18}$$

where V_{KS} represents the Kohn-Sham potential whose form is:

$$V_{KS} = V_{ext}(\mathbf{r}) + \frac{\delta E_H[n]}{\delta n(\mathbf{r})} + \frac{E_{xc}[n]}{\delta n(\mathbf{r})} = V_{ext}(\mathbf{r}) + V_H(\mathbf{r}) + V_{xc}(\mathbf{r})$$
$$= -\sum_I \frac{Z_I}{|\mathbf{r} - \mathbf{R}_I|} + \int d\mathbf{r}' \frac{n(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} + V_{xc}(\mathbf{r}). \tag{2.19}$$

Note that the Kohn-Sham potential is the one-body effective counterpart of the many-body electronic potential. It operates independently and equivalently on each non-interacting electron of the Kohn-Sham system at any spatial position $\mathbf{r}$. The solutions of the Kohn-Sham hamiltonian $\hat{H}_{KS}$ are the Kohn-Sham eigenstates $\psi_i(\mathbf{r})$ with energy ε_i for each i-th electron. Finally, the electron density distribution is given in terms of the Kohn-Sham wave functions as follows:

$$n(\mathbf{r}) = \sum_i |\psi_i(\mathbf{r})|^2 \qquad (2.20)$$

In conclusion, the DFT formalism allows us to reformulate the difficult many-body problem in Eq. 2.1 into a set of N_{el} solvable one-body schrödinger equations (Eq. 2.18). Moreover, if the expressión for the exchange-correlation energy $E_{xc}[n]$ was known, the exact ground-state electron density and energy of the interacting electronic system could be obtained. This is why DFT is considered an "exact" quantum many-body method even if in practice some approximations are needed.

2.5 | DFT IN PRACTICE

2.5.1 | SELF-CONSISTENT CYCLE

Once an approximate form of the exchange-correlation functional is chosen a set of $N_e l$ one-body Schrödinger equations (Eq. 2.18), the so-called Kohn-Sham equations, have to be solved. To do so, a self-consistent method is needed since the solutions of Eq. 2.18, the Kohn-Sham wave functions, are at the same time the ones used to calculate the electron density on which the external potential depends. An iterative procedure of 5 steps is needed to solve these equations self-consistently:

1. An arbitrary electron density is chosen to initialize the method.

2. The Kohn-Sham potential is computed using the electron density (Eq. 2.19).

3. The set of N_{el} Kohn-Sham Schrödinger equations is solved by getting the Kohn-Sham eigenstates (Eq. 2.18).

4. A new electron density is computed using the obtained Kohn-Sham eigenstates.

5. The convergence of the obtained results is checked. There are various ways of determining the convergence of the calculations: the difference between the electron density corresponding to two consecutive iterations must be sufficiently small, and the variation of the total energy must be small…If the calculation is not converged the new electron density is used to calculate again the Khon-Sham potential in step 2 and continue with the procedure. Conversely, if the calculation is converged the obtained electron density is considered the ground-state electron density and thus, the electronic energy computed with that electron density will be the ground-state electron energy.

2.5.2 | Base functions

Another tool to facilitate the calculation of the Kohn-Sham schödinger equations is to expand the Kohn-Sham orbitals $\psi_i(\mathbf{r})$ in a function basis $\{f_\alpha(\mathbf{r})\}$ and solve the equations for the coefficients of that expansion:

$$\psi_i(\mathbf{r}) = \sum_\alpha c_\alpha^i f_\alpha(\mathbf{r}). \tag{2.21}$$

Of course, there are many types of bases to choose from, each with its advantages and disadvantages. However, we can divide them into two main types of bases. On the one hand, we have localized atomic orbitals such as the LCAO in which the wave function of the electrons is expanded as a function of the atomic orbitals (of the isolated atom). The problem with this method is that the atomic orbitals must be correctly characterized, which requires further auxiliary calculations. As these orbitals are centered on the position of the atom itself, they are very useful in systems such as isolated molecules or systems where there are isolated atoms. In any case, they are also widely used in the calculation of bulk systems. On the other hand, plane wave bases are also often used for the expansion of Kohn-Sham orbitals. Plane waves are exact solutions to the Schrödinger equation at a constant potential. The expansion in these bases is conceptually very simple since it is only a matter of performing the Fourier series of the wave function. This basis is especially useful when the system to be treated is periodic.

2.5.3 | PSEUDOPOTENTIALS

In the previous sections, the total electron density has been presented as a variable for the description of the system. However, in practice, it is easier to treat the valence electrons and the rest (core electrons) separately. The reason for treating them differently is simple: only the valence electrons interact in chemical bonds, and the rest of the electrons remain located close to the nucleus as if they were frozen. Therefore, the ensemble formed by the nucleus and the non-valence electrons can be characterized by an effective nucleus that, to a large extent, maintains an atomic configuration and whose orbitals hardly change under external agents. However, the interaction between valence electrons and nuclei is no longer solely Coulombian: on the one hand, the charge of the nucleus is shielded by the electrons that remain bound to it; and on the other hand, the interaction between the electrons forming the pseudo nucleus must now also be included here. This new interaction between the valence electrons and the effective nuclei is characterized by an external pseudopotential: $V_{ext}(\mathbf{r}) \rightarrow V_{ext}^{PP}(\mathbf{r})$. Thus, the Kohn-Sham equations are solved only taking into account the contribution of the valence electrons. Therefore, the Kohn-Sham energy functional is also a functional of the valence electron density: $E_{KS}[n_v] = T_s[n_v] + E_H[n_v] + E_{xc}[n_v] + V_{ext}^{PP}[n_v]$.

The advantage of this approximation is obvious since it reduces the number of electrons to be studied explicitly, which simplifies all the calculations to be performed. In addition, by considering the set of nuclei and electrons, the effect of the Coulombian potential produced by the nuclei and the singularity that occurs when the radial distance tends to 0 is smoothed, which is also an improvement for the numerical calculations. However, it also has its disadvantages. In the same way as when incorporating the exchange-correlation functional, we have now introduced a new potential that is unknown. Thus, it requires a new effort to characterize these pseudopotentials in a way that adequately represents the ensemble of the bounded nucleus and electrons.

2.6 | PERIODIC STRUCTURES

The calculation of the ground-state electronic structure can be simplified by taking into account the properties of the system. In this case, the periodicity of the structure can be implemented into the method so that

it is more efficient. In an ideal crystalline structure, the atoms are located in a unit cell that repeats periodically. The unit cell is spatially delimited by the set of primitive real lattice vectors $\{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3\}$. The real or direct Bravais lattice is generated by all the possible translations of the unit cell given by the real primitive lattice vectors [95]. Due to the crystal periodicity, the Kohn-Sahm potential acting on the electrons must be periodic too:

$$V_{KS}(\mathbf{r}) = V_{KS}(\mathbf{r} + \mathbf{T}) \tag{2.22}$$

where $\mathbf{T}$ is the lattice vector defined as $\mathbf{T} = n_1\mathbf{a}_1 + n_2\mathbf{a}_2 + n_3\mathbf{a}_3$ for $n_i \in \mathbb{Z}$ and $\mathbf{r}$ is the vector of the real space defined within the unit cell. Bloch's theorem states that in a system of these characteristics the solution of the Schrödinger equation is the superposition of functions of the form of a plane wave times a crystal periodic function [96]:

$$\psi_{\mathbf{k}} = e^{i\mathbf{k}\cdot\mathbf{r}} u_{\mathbf{k}}(\mathbf{r})$$

where $u_{\mathbf{k}}(\mathbf{r})$ has the same periodicity as the system $u_{\mathbf{k}}(\mathbf{r}) = u_{\mathbf{k}}(\mathbf{r}+\mathbf{T})$. The wave vector $\mathbf{k}$ is commonly known as the crystal momentum. It is sometimes is very useful to work in terms of the crystal momentum $\mathbf{k}$. The space in which the $\mathbf{k}$ wave vectors are defined is called the reciprocal space. In the same manner, the lattice vectors and the unit cell are defined in real space, the same procedure can be done in reciprocal space. In this case, the reciprocal lattice vectors $\{\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3\}$ are defined so that the relation $\mathbf{a}_i\mathbf{a}_j = 2\pi\delta_{ij}$, where δ_{ij} is the Kronecker delta, is satisfied. Similarly, a primitive unit cell known as the first Brillouin zone (1BZ) is defined as the most compact possible unit that repeats periodically to form the reciprocal space. By imposing Born-von Karman boundary conditions on the Kohn-Sham wave functions, it is demonstrated that the momentum k is real and restricted to the following allowed values:

$$\mathbf{k} = \sum_i^3 \frac{m_i}{N_i}\mathbf{b_i} \tag{2.23}$$

where $m_i \in \mathbb{N}$ and $|m_i| < N_i$ being N_i the number of unit cells along the crystal direction i.

In practice, Bloch's theorem allows us to replace some integrals that should be done in the whole real space by integrals done in the 1BZ.

Therefore, those infinite volume integrals can be simplified in into finite volume integrals of the type:

$$g = \frac{V_{cell}}{(2\pi)^3} \int_{1BZ} d^3\mathbf{k}\, g(\mathbf{k}) \tag{2.24}$$

Since these integrals are computed with numerical methods, the integral is performed by only taking into account a finite number of k points. A correct choice of the number of points, as well as the position of the points themselves, is essential for the optimization of resources when making calculations.

Another tool very used in systems with a periodic structure is the expansion of the functions $u_\mathbf{k}(\mathbf{r})$ into plane waves as:

$$u_\mathbf{k}(\mathbf{r}) = \sum_\mathbf{G} C_\mathbf{G}\, e^{i\mathbf{G}\cdot\mathbf{r}} \tag{2.25}$$

where the wave vectors are of the form $\mathbf{G} = m_1\mathbf{b}_1 + m_2\mathbf{b}_2 + m_3\mathbf{b}_3$. Hence, the Block's wave function can be rewritten taking into account this expansion:

$$\psi_\mathbf{k} = \sum_G C_{\mathbf{k}+\mathbf{G}}\, e^{i(\mathbf{G}+\mathbf{k})\cdot\mathbf{r}}. \tag{2.26}$$

As has been mentioned before in sec. 2.5.2, the Konh-Sham orbitals can be expanded in wave planes of the form of Eq. 2.26. The sum in this equation goes over all the possible $\mathbf{G}$ vectors. That is to say, it is an infinite sum. However, in practice the series is cut off at certain $\mathbf{G}_{cut}$ and the vectors with higher modulus are not taken into account. Thus, a cutoff energy is defined, which is the energy corresponding to the plane wave characterized by the vector $\mathbf{G}_{cut}$:

$$E_{cut} = \frac{\hbar^2}{2m_e}\mathbf{G}_{cut}^2 \tag{2.27}$$

The expansion into plane waves is not limited only to the Kohn-Sham orbitals but to other functions that have to be computed in reciprocal space. Therefore, as well as the choice of the k points, the choice of a proper E_{cut} is very important for the accuracy and efficiency of the method.

2.7 | USE OF DFT IN THIS THESIS

DFT has been used to optimize the atomic structure of bulk crystals, layers and clusters. For example in chapter 3 DFT was used to minimize the energy of all the portlandite polymorphs and calculate the forces for the VC-NEB method that will be explained later in Sec. 7. In chapter 4, DFT was used to calculate the forces in the VC-NEB calculation for the silicate reaction on the portlandite surface. Finally, DFT was used in chapter 5 to minimize the structure of every cluster found by the evolutionary algorithms. In total, more than 10.000 DFT calculations have been done.

3 | REAXFF

While traditional force fields use fixed connectivity for the chemical bonds, ReaxFF is a reactive force field that enables bond formation and rupture. To do so, ReaxFF employs the bond order concept to model chemical interactions. This is very helpful when sampling transformations in which chemical reactions occur. This force field requires both quantum and experimental data to parametrize the functions so that the reactivity is accurate.

As in non-reactive force fields, the total potential energy in ReaxFF can be expressed as a sum of different contributions:

$$U_{system} = U_{bond} + U_{over} + U_{under} + U_{lp} + U_{val} + U_{tor} + U_{vdWaals} + U_{Coulomb}$$
$$(2.28)$$

where U_{bond} is the bond energy, U_{over} is the over-coordination penalty, U_{under} the under-coordination penalty, U_{lp} the lone-pair energy, U_{val} and U_{tor} the valence angle and torsion energies, $U_{vdWaals}$ the van der Waals energy and $U_{coulomb}$ the coulombic energy. The ReaxFF force field, instead of fixing a certain coordination to an ion, imposes energetic penalties for low and high coordination numbers, which makes the reactions possible. The over-coordination and undercoordination energies, depending on the local environment which is defined by the bond order.

The bond order for a pair of atoms is defined as the sum of the σ bond order, π bond order and the double π order:

$$BO'_{ij} = BO'^{\sigma}_{ij} + BO'^{\pi}_{ij} + BO'^{\pi\pi}_{ij} \qquad (2.29)$$

where each of the bond order is almost 1 under a certain distance and negligible above a certain distance. For instace, for carbon the σ bond order is unity below 1.5 Åand negligible above 2.5 Å, the π bond is unity below 1.2 Åand negligible above 1.75 Åand finally, the double π bond is unity below 1.0 Åand negligible above 1.4 Å.

Once the bond order is calculated, the bond energy E_{bond} can be calculated by multiplying each of the bond order with the energy of a single bond, D^{σ}_e for a σ bond, D^{π}_e for a pi bond and $D^{\pi\pi}_e$ for a double pi bond.

The degree of deviation of the sum of the uncorrected bond orders around an atomic center Δ'_i is calculated by:

$$\Delta'_i = \sum_{j}^{nbond} BO'_{ij} - Val_i \qquad (2.30)$$

where Val_i is the valence of atom i. This deviation together with the over-coordination and under-coordination parameters give the over-coordination and under-coordination energies.

3.1 | Use of ReaxFF in this thesis

As with DFT, ReaxFF was used to minimize a lot of structures when DFT was too slow for our time requirements. Moreover, a lot of calculations like the VC-NEB in chapter 3 were performed first with ReaxFF to see rapid results and the with DFT to obtain more accurate results. In chapter 4, all the molecular dynamic simulations were performed using ReaxFF and in Chapter 5 all the clusters were optimized first with ReaxFF before doing it with DFT. Therefore, ReaxFF has been a very important tool for longer simulations.

4 | Molecular Dynamics

Both with *ab initio* methods like DFT and with the use of Force Fields atomic forces and energies of different configurations can be computed. However, the computation of some equilibrium properties requires the

sampling of other configurations that the system may adopt. In the past, Monte Carlo methods were the only methods that allowed this to be done but in the 1950s, with the development of the early computers, methods to solve the time evolution of the equations of motion for a many-body system were used. For example, in 1964 the Indian physicist Aneesur Rahman published a simulation of liquid Argon using a Lenard-Jones potential [97].

All the configurations a system may adopt are defined by the position $\mathbf{q} = (q_1, ..., q_N)$ and the momentum $\mathbf{p} = (p_1, ..., p_N)$ of all the particles where N is the number of particles. All the different configurations are represented in the phase space $(\mathbf{q}, \mathbf{p})$ and all the configurations together form the *ensemble*. Hence, any system property A can be computed as an ensemble average

$$\langle A \rangle = \iint d\mathbf{q} d\mathbf{p} A(\mathbf{q}, \mathbf{p}) P(\mathbf{q}, \mathbf{p}) \tag{2.31}$$

where $A(\mathbf{q}, \mathbf{p})$ represents the value of that property at that configuration and $P(\mathbf{q}, \mathbf{p})$ is the probability density of the system to be in that configuration. The probability density is given by

$$P(\mathbf{q}, \mathbf{p}) = \frac{1}{Q} e^{-H(\mathbf{q}, \mathbf{p})/k_B T} \tag{2.32}$$

where $H(\mathbf{q}, \mathbf{p})$ is the Hamiltonian, k_B is the Boltzmans constant, T is the temperature and Q is the partition function

$$Q = \int \int d\mathbf{q} d\mathbf{p} e^{-H(\mathbf{q}, \mathbf{p})/k_B T}. \tag{2.33}$$

The calculation of the partition function is very important since several thermodynamical quantities can be derived from it.

This approach requires the totality of the phase space to be sampled to compute the value of A at each configuration and the partition function Q. However, the probability of the system being in most of the configurations is near zero and therefore, those configurations do not contribute to the ensemble average. This is why Monte Carlo methods are based on sampling the most probable configurations to get the ensemble average. On the other hand, molecular dynamics uses an approach more similar to a real experiment in which a series of measurements are performed

during a time interval and then the average of these measurements is determined. This is done by initializing the system in a configuration near the equilibrium and by solving Newton's equations of motion to get a realistic trajectory of the system. In the interest of computing the ensemble average, the "ergodic hypothesis" must be assumed. That is to say, it must be assumed that for a sufficiently long simulation the trajectory will sample the most probable configurations and therefore the ensemble average can be computed by computing the time average

$$\langle A \rangle = \bar{A} = \lim_{\tau \to \infty} \frac{1}{\tau} \int_{\tau=0}^{\infty} A(\mathbf{q}(t), \mathbf{p}(t)) dt. \tag{2.34}$$

All the observable quantities that wanted to be measured with molecular dynamics have to be defined first as a function of the positions and momenta of the particles in the system. For example, the temperature can be defined using the equipartition of energy between all the degrees of freedom. Hence, the total kinetic energy can be expressed as

$$\sum_{i=1}^{N} \frac{\mathbf{p}_i^2}{2m_i} = \frac{1}{2} N k_B T \tag{2.35}$$

where m_i is the mass of the i-th particle. Thus, the instantaneous temperature can be defined as

$$T = \frac{1}{N k_B} \sum_{i=1}^{N} \frac{\mathbf{p}_i^2}{m_i}. \tag{2.36}$$

Since the total kinetic energy fluctuates during the simulation, so does the temperature. Thus, a time average over many fluctuations should be made to get an accurate measurement of temperature.

4.1 | OVERVIEW OF THE ALGORITHM

As stated before, molecular dynamics is all about integrating Newton's equations of motion to get a realistic trajectory of a system. This is done iteratively in a couple of steps:

1. Initialize the system (positions and velocities).

2. Compute forces (DFT or Force Fields).

3. Integrate Newton's equations of motion to get new positions for the next time interval. Go to step 2 again until a maximum number of desired time steps are reached.

4. Compute the average of measured quantities.

The initialization of the system is usually done by giving as an input a structure near the equilibrium. If the starting configuration is very far from equilibrium it can take a substantial computational time to equilibrate the system. The starting velocities are usually chosen to get the desired temperature in the system following Eq. 2.36. To do so, a Gaussian distribution of velocities is taken to produce the desired temperature.

The next two steps are the main core of the molecular dynamics algorithms. The forces can be computed with any force field that describes the particle interaction properly or by an *ab initio* method like DFT. However, *ab initio* methods are much slower than force-field, and therefore, they are only capable of simulating some picoseconds for a small system containing 20-30 atoms. There are several methods for integrating Newton's equations of motion but all of them are based on the Taylor expansion of the coordinate of a particle around time t

$$\mathbf{r}(t + \Delta t) = \mathbf{r}(t) + v(t)\Delta t + \frac{\mathbf{f}(t)}{2m}\Delta t^2 + \frac{\dddot{\mathbf{r}}}{3!}\Delta t^3 + \ldots \qquad (2.37)$$

where r is the position of a particle, Δt is the time step (which has to be small), f is the force experienced by the particle and $\dddot{r}$ is the third time derivative of the positions. Using this expansion various methods were developed but one of the most used ones (and also the one used in this thesis) is called Velocity-Verlet. This method only takes into account the Taylor series to the second order and uses an intermediate time step to compute the positions and velocities of the following time step. This is usually done in three steps as

$$\mathbf{v}(t + \Delta t/2) = \mathbf{v}(t) + \frac{\mathbf{f}(t)}{2m}\Delta t$$
$$\mathbf{r}(t + \Delta t) = \mathbf{r}(t) + \mathbf{v}(t + \Delta t/2)\Delta t \qquad (2.38)$$
$$\mathbf{v}(t + \Delta t) = \mathbf{v}(t + \Delta t/2) + \frac{\mathbf{f}(t + \Delta t)}{2m}\Delta t.$$

This method, allows us to have a deterministic trajectory of a system in the microcanonical ensemble (NVE) where the number of atoms, the volume and the total energy remains constant. However, as done in a real experiment, sometimes it is desired to perform the simulation at constant temperature or constant pressure or both. To do so, a thermostat and a barostat have to be added to the algorithm.

4.2 | PERIODIC BOUNDARY CONDITIONS

When dealing with bulk systems, it is essential to choose boundary conditions that mimic the presence of an infinite bulk surrounding our model of a finite N particle system. This is usually done by the use of periodic boundary conditions where the volume containing the N particles is considered the primitive cell of an infinite periodic lattice of identical cells. That is to say, a particle in the primitive cell not only interacts with the other particles in the primitive cell but also with all the other particles in all other cells (including identical images of the same particle). Thus, assuming that all interactions are pairwise additive, then the total potential energy of the N particles in a periodic cell can be computed as,

$$U_{tot} = \sum_{i,j,\mathbf{n}} u(|\mathbf{r}_{ij} + \mathbf{n}L|) \tag{2.39}$$

where u is the pairwise interaction, $\mathbf{r}_{ij}$ the vector between two particles in the primitive cell, L is the length of the periodic box (assumed cubic for simplicity) and $\mathbf{n}$ is an arbitrary vector of three integers, while the prime over the sum indicates that that the term $i = j$ is excluded when $\mathbf{n} = (0,0,0)$. It is important to note that for a periodic system this sum is infinite and therefore practically unfeasible. However, most of the pairwise interactions are short-range interactions and therefore it is legitimate to truncate all interactions beyond a certain cutoff.

The truncation of the interactions must be done since abrupt truncations will lead to discontinuities in the potential energy surface. Therefore, there are several methods of implementing this truncation. The simplest truncation would be to ignore all interactions beyond a cutoff radius r_c. As mentioned before, this will create discontinuities in the potential energy surface and thus, is not particularly suitable for an MD simulation. Another way of truncating the interaction is to modify slightly the interaction potential adding a smooth tail that turns the

value of the potential interaction almost to zero at r_c. Similarly, the potential can be truncated and shifted u^{tr-sh} so that the potential vanishes at r_c:

$$u^{ts-sh}(r) = \begin{cases} u(r) - u(r_c), & r <= r_c \\ 0, & r > r_c. \end{cases} \tag{2.40}$$

Due to the continuity of these two modified potentials, they are very suitable for MD. Finally, another way of truncating the interaction is the use of the minimum image convention. In this case, instead of having a spherical cutoff, the interaction with the nearest image of all the particles in the simulation box is calculated. As a consequence, the potential is not a constant on the surface of a cube around a given particle. Hence, for the same reasons mentioned in the previous paragraph, the simple minimum image convention should never be used in Molecular Dynamics simulations.

4.3 | THERMOSTAT AND BAROSTAT

Molecular dynamics calculations at constant temperature (NVT ensemble) are done by including a thermostat in the algorithm. Since thermostats simulate the system coupled to an imaginary thermal reservoir, the energy can fluctuate but the temperature and the number of atoms will remain constant. Many different methods have been proposed for this purpose such as velocity rescaling[98], the Andersen thermostat[99], the Berendsen thermostat[100] or the Nose-Hoover[101, 102] thermostat. The latter is the one used in this thesis.

The Nose-Hoover thermostat introduces an additional degree of freedom s together with its conjugate momentum p_s to represent the imaginary thermal bath coupled to the system. Hence, the system is now represented by the Nose Hamiltonian H_n,

$$H_n = \sum_i^N \frac{\mathbf{p}_i^2}{2sm_i} + \frac{1}{2}\sum_{i\neq j} \mathbf{U}(r_i - r_j) + \frac{\mathbf{p}_s^2}{2Q} + gk_BT\ln(s). \tag{2.41}$$

The first two terms in the Nose Hamiltonian H_n describe the usual Hamiltonian for a classical many-body system but the momentum p_i is

replaced by p_i/s since the interaction between the system and s is represented by the scaling of the velocities. This can be interpreted as an exchange of heat between the system and the thermal reservoir. The third term in the Nose Hamiltonian H_n is the kinetic energy of the coordinate s where Q determines the time scale of the temperature fluctuation. Finally, the last term in the Nose Hamiltonian is associated with the potential energy of the thermal reservoir where g is the number of degrees of freedom in the system. This choice of potential ensures that the canonical ensemble (NVT) averages are recovered[101].

Taking into account Nose Hamiltonian, the equations of motion read as

$$
\begin{aligned}
m_i \frac{d^2 \mathbf{r}_i}{dt^2} &= \mathbf{f}_i - \zeta m_i \mathbf{v}_i \\
Q \frac{d\zeta(t)}{dt} &= \frac{1}{2} \sum_i^N m_i \mathbf{v}_i^2 - \frac{3N+1}{2} k_B T
\end{aligned}
\tag{2.42}
$$

where the friction constant is defined as $\zeta = p_s/Q$. These equations of motions can be easily implemented into the velocity Verlet algorithm:

$$
\begin{aligned}
\mathbf{v}(t + \Delta t/2) &= \mathbf{v}(t) + \frac{\Delta t}{2}\left(\frac{\mathbf{f}(t)}{m} - \zeta(t)\mathbf{v}(t)\right) \\
\mathbf{r}(t + \Delta t) &= \mathbf{r}(t) + \mathbf{v}(t)\Delta t + \left(\frac{\mathbf{f}(t)}{m} - \zeta(t)\mathbf{v}(t)\right)\frac{\Delta t^2}{2} \\
\zeta(t + \Delta t/2) &= \zeta(t) + \frac{\Delta t}{2Q}\left[\frac{1}{2}\sum_i^N m_i \mathbf{v}_i(t)^2 - \frac{3N+1}{2} k_B T\right] \\
\zeta(t + \Delta t) &= \zeta(t + \Delta t/2) + \frac{\Delta t}{2Q}\left[\frac{1}{2}\sum_i^N m_i \mathbf{v}_i(t + \Delta t)^2 - \frac{3N+1}{2} k_B T\right] \\
\mathbf{v}(t + \Delta t) &= \left[\mathbf{v}(t + \Delta t/2) + \frac{\mathbf{f}(t + \Delta t)}{2m}\Delta t\right]\left[1 + \frac{\Delta t}{2}\zeta(t + \Delta t)\right]^1.
\end{aligned}
\tag{2.43}
$$

In a similar way to the thermostat, we can use the extended Lagrangian method to calculate the equations of motion in the NPT ensemble. For this, the system pressure P_{ext} must be held constant $\dot{P}_{ext} = 0$. The barostat is introduced by the variables ε, p_ε and W [103, 104]. ε is defined as the natural logarithm of the system volume

$$\varepsilon = \ln \left(\frac{V}{V_0} \right) \tag{2.44}$$

where V_0 is the volume at $t = 0$. p_ε is the conjugate momentum of the epsilon variable and W is the mass associated with the barostat. This associated mass W represents the level of coupling with the barostat.

Therefore, the NPT Hamiltonian H_{NPT} describing the system at constant temperature T and pressure P_{ext} is similar to the NVT Hamiltonian H_{NVT} written in Eq. 2.41 but with two added terms related to the barostat:

$$H_{NPT} = H_{NVT} + P_{ext}V + \frac{p_\varepsilon}{2W}. \tag{2.45}$$

The equations of motion can be easily computed from the hamiltonian. The reader is referred to the original paper by Hoover where the equations of motion are computed [104].

5 | ENHANCED SAMPLING METHODS

Most complex systems have different minima in their free energy surface. This is because these systems contain different phases or conformations with low energy, usually with high energy barriers between them. Transition state theory can be used to compute reaction rates from energy barriers [105]. However, these reaction rates are usually much larger than the periods that can be simulated with MD. While the integration time step of MD is usually limited to femtoseconds or shorter (dictated by the fastest degrees of freedom, such as bond vibrations), the observation of the slow domain motion requires simulations extended to the micro/millisecond time scale. Therefore, conformational changes or phase transitions in a complex system are considered rare events in MD. This is why enhanced sampling techniques are needed to accelerate the dynamics of such systems in simulation. This enhanced sampling technique will allow us to sample the phase space much better and thus, calculate the partition function in Eq. 2.33.

There are many enhanced sampling methods but most of them can be categorized into two main groups: collective variable-based and collective variable-free methods. The collective variable-based method uses

a predefined reaction coordinate to somehow modify the potential energy surface by adding biased potentials along that reaction coordinate so that the sampling around transition regions is increased. Such methods include hyperdynamics [106], conformational space annealing [107], metadynamics [108], umbrella sampling (US) [109] and targeted molecular dynamics (TMD) among others. The drawback of these methods is that for some complex systems, it is really hard to predefine a collective variable to sample the system along its different minima in the free energy landscape. For this reason, collective variable methods do not use any predefined reaction coordinate. In this case, the canonical probability distribution is altered to a distribution that will induce a broader sampling of the potential energy. This category of methods includes temperature-accelerated dynamics [110], parallel tempering [111] and replica exchange molecular dynamics [112] to name a few.

These enhanced sampling methods are very useful to compute the free energy barrier between two local minima. For example, in chapter 4 the following enhanced sampling methods were used to compute the internal and free energies of the transformation from a portlandite monolayer to a C-S-H layer.

From statistical mechanics, it is known that the free energy can be computed as the natural logarithm of the probability of a point in phase space. Therefore, defining a collective variable ρ as a reaction coordinate, the free energy can be defined in terms of this reaction coordinate as:

$$G = \frac{1}{\beta} \ln P(\rho) \tag{2.46}$$

Hence, the difference in free energy can be computed by calculating the probability of the system of being at different points along the reaction coordinate. It is important to note that in this way, only the difference of free energy with respect to the reaction coordinate is calculated and it could be that if the reaction coordinate is not well chosen, this free energy is very different from the real free energy. That is why usually the free energy difference along a reaction coordinate is called the potential of mean force (PMF). For simplicity, in this work, we will just call it free energy difference.

The main problem when calculating the probability along the reaction coordinate is that the high-energy regions are sampled rarely due to the limited running time of the simulations. For rare events, those with an

energy barrier significantly larger than $k_B T$, direct sampling is infeasible. To obtain a profile of $G(\rho)$, however, also those high energy regions, those rare events, are required. This is where enhanced sampling techniques are required allowing us to sample low and high-energy zones equally well.

In this work, both TMD and US have been used to sample the transition regions between two different minima. These methods are very useful to calculate the free energy barrier in a phase transition.

5.1 | TARGETED MOLECULAR DYNAMICS

Targeted molecular dynamics (TMD) is also a method to sample the phase transitions. In this case, the system is pulled from tje starting configuration $\mathbf{x}^A$ to the target configuration $\mathbf{x}_B$ [113, 114]. The distance from each configuration $\mathbf{x}$ to the target configuration is defined as the root mean square distance between the atoms in each configuration and the atoms in the target configuration:

$$\rho = \sqrt{\sum_i (\vec{x}_i - \vec{x}_i^B)^2}, \tag{2.47}$$

where the sum is computed over the atoms involved in TMD. Although the root means square distance would be a free variable in unbiased MD, in TMD the system is forced to be in a certain $\rho_c(t)$ at each timestep and $\rho_c(t)$ linearly evolves towards the final distance with a given velocity v as $\rho_c(t) = \rho(0) + vt$. This drags the system monotonically to the target configuration $\mathbf{x}_B$. This is accomplished by introducing the constraint

$$\Phi(\rho(t)) = \rho(t) - \rho_c(t) = 0 \tag{2.48}$$

by means of a Lagrange multiplier λ, given rise to an additional constraining force

$$f_c(\rho) = \lambda \frac{d\Phi(\rho(t))}{d\rho} = \lambda. \tag{2.49}$$

Here λ has to be chosen so that the system satisfies Eq. 2.48. The work W performed by the forces that are produced by the constraint Φ can be

defined as [115, 116]

$$W(\rho) = \int_{\rho_0}^{\rho} d\rho' f_c(\rho') = \int_{\rho_0}^{\rho} d\rho' \lambda. \tag{2.50}$$

Various ways of computing the free energy difference have been presented for systems in which external forces are encountered. In the quasistatic limit of very slow pulling thermodynamic integration can be performed, which writes the free energy difference using the potential of mean force

$$\Delta F(\rho) = \int_{\rho_0}^{\rho} d\rho' \frac{dF}{d\rho'} = \int_{\rho_0}^{\rho} d\rho' \left\langle f_c(\rho') \right\rangle_{eq}, \tag{2.51}$$

where $\left\langle f_c(\rho') \right\rangle_{eq}$ represents a canonical average of the constrained force f_c at $\rho = \rho'$. Despite being a straightforward way of calculating the free energy difference this method is computationally very expensive since numerous and very long simulations are required to converge to equilibrium.

However, nonequilibrium trajectories can be used to compute the free energy difference using Jarzynski's equality [117]

$$e^{-\beta \Delta F(\rho)} = \left\langle e^{-\beta \Delta W(\rho)} \right\rangle_{neq} \tag{2.52}$$

where $\beta = (k_B T)^{-1}$ and $\langle \cdot \rangle_{neq}$ represents the ensemble average of different trajectories all starting from an equilibrium distribution at $\rho = \rho_0$. This method provides a way to speed up the free energy calculation since non-equilibrium trajectories can be employed.

Because of the difficulty in estimating the exponential average, there are several ways to tackle the numerical computation of Jarzynski's equality. Among those methods, the cumulant expansion used in the present work allows a faster convergence. This method is based on the expansion of the natural logarithm of an exponential average in terms of cumulants $\ln \langle e^x \rangle = \langle x \rangle + 1/2(\langle x^2 \rangle - \langle x \rangle^2) + ...$ Taking the expansion up to the second order and using it in Equation 2.52 the free energy difference can be computed as

$$\Delta F(\rho) = \left\langle W(\rho) \right\rangle_{neq} - \frac{\beta}{2} \left\langle (W(\rho) - \left\langle W(\rho) \right\rangle_{neq})^2 \right\rangle_{neq}. \tag{2.53}$$

This approximation is only exact if the work $W(\rho)$ has a gaussian distribution.

Jarzynki's equality directly reproduces the free energy profile along the pulling coordinate ρ. However, sometimes this coordinate is not a reaction coordinate and in order to find the transition-free energy some collective variables have to be used. Collective variables are widely used in unbiased MD but it is not straightforward to use them in biased non-equilibrium techniques like TMD. Post et al. [116] showed that for a collective variable $\mathbf{r}(\mathbf{x})$ the free energy can be calculated as

$$\beta \Delta F(\mathbf{r}) = -\ln \left[\frac{\int d\rho \left\langle \delta(\mathbf{r} - \mathbf{r}(\mathbf{x}|\rho)) e^{-\beta W(\rho)} \right\rangle_{\text{neq}}}{\int d\rho' \left\langle e^{-\beta W(\rho')} \right\rangle_{\text{neq}}} \right] \tag{2.54}$$

where $\mathbf{r}(\mathbf{x}|\rho)$ represents the collective variable $\mathbf{r}(\mathbf{x})$ restricted to a given value of ρ.

Regarding the application of equation 2.54 to the set of TMD simulations, the integral can be represented by the sum over all the frames in each trajectory which combined with the mean value over all different trajectories the sum can be converted to a sum over all frames in the set:

$$\begin{aligned} \beta \Delta F(\mathbf{r}) &= -\ln \left[\frac{\left\langle e^{-\beta W(\rho)} \delta_n(\mathbf{r}) \right\rangle_{\text{set}}}{\left\langle e^{-\beta W(\rho)} \right\rangle_{\text{set}}} \right] \\ &= -\ln \left[\frac{\sum_n^{N_f} e^{-\beta W(\rho)} \delta_n(\mathbf{r})}{\sum_n^{N_f} e^{-\beta W(\rho)}} \right], \end{aligned} \tag{2.55}$$

where $\langle \cdot \rangle_{\text{set}}$ represents the mean value over the whole set, N_f is the number of frames in the set and $\delta_{k,\rho}(\mathbf{r})$ is a counting function in some bin size $\Delta \mathbf{r}$

$$\delta_{k,\rho}(\mathbf{r}) = \begin{cases} 1 & \text{if} \quad \mathbf{r} - \frac{\Delta \mathbf{r}}{2} \leq \mathbf{r}_k(\rho) < \mathbf{r} + \frac{\Delta \mathbf{r}}{2} \\ 0 & \text{else} \end{cases} \quad . \tag{2.56}$$

Since we are only interested in the difference of the free energy between different structures, all the summing or subtracting constants can be removed from the equation since they will only change the origin. Therefore, taking into account the properties of the logarithm we can remove the denominator inside the logarithm since it does not depend on $\mathbf{r}$.

Apart from that, computing the $\delta_n(\mathbf{r})$ function in the term $\left\langle e^{-\beta W(\rho)}\delta_n(\mathbf{r})\right\rangle_{set}$ will transform the mean value over the whole set to the mean value over the frames in the bin $\delta(\mathbf{r})$:

$$\left\langle e^{-\beta W(\rho)}\delta_n(\mathbf{r})\right\rangle_{set} = \frac{N_{\delta(\mathbf{r})}}{N_f}\left\langle e^{-\beta W(\rho)}\right\rangle_{\delta(\mathbf{r})} \tag{2.57}$$

where $\langle\cdot\rangle_{\delta(\mathbf{r})}$ represents the mean value over the frames in the bin $\delta(\mathbf{r})$, and $N_{\delta(\mathbf{r})}$ the number of frames in that bin. Thus, applying equation 2.57 and the cumulant expansion the free energy difference can be computed for any collective variable with the following formula:

$$\Delta F = \langle W(\rho)\rangle_{\delta(\mathbf{r})} - \frac{\beta}{2}\langle (W(\rho) - \langle W(\rho)\rangle_{\delta(\mathbf{r})})^2\rangle_{\delta(\mathbf{r})} - \frac{\ln N_{\delta(\mathbf{r})}}{\beta} \tag{2.58}$$

5.2 | UMBRELLA SAMPLING

This is very useful when a transition is a rare event and therefore it would need an enormous simulation time to sample this event. In umbrella sampling, to sample the transition state between state $\mathbf{x}^A = (\vec{x}_1^A, ..., \vec{x}_N^A)$ and state $\mathbf{x}_B = (\vec{x}_1^B, ..., \vec{x}_N^B)$, the system is restrained by a bias potential w_i to a certain window i along the reaction coordinate ρ:

$$E^b(\mathbf{x}) = E^u(\mathbf{x}) + w_i(\rho) \tag{2.59}$$

where the superscript 'b' and 'u' denote bias and unbiased quantities respectively and the bias additional energy w_i only depends on the reaction coordinate ρ. In this work, a harmonic bias potential of strength K has been chosen to sample the configurational space of each window with a reference ρ_i^{ref}:

$$w_i = \frac{K}{2}(\rho - \rho_i^{ref}) \tag{2.60}$$

The reference values for each window were taken from a TMD simulation in which the reaction was divided into 40 equidistant windows. The strength of the bias potential has been chosen so that the overlap between the probability of adjacent windows is around 30%.

In order to calculate the free energy of each window $G_i(\rho)$ the unbiased probability distribution $P_i^u(\rho)$ is needed since

$$G(\rho) = \frac{-1}{\beta} \ln P_i^u(\rho). \tag{2.61}$$

From statistical mechanics, it is known that the probability distribution is computed by

$$P_i^u(\rho) = \frac{\int e^{-\beta E(x)} \delta(\rho'(x) - \rho) d^N x}{\int e^{-\beta E(x)}} \tag{2.62}$$

but since umbrella sampling is a biased molecular dynamics method the obtained probability distribution from the simulation is biased. Therefore, the biased distribution can be written as

$$P_i^b(\rho) = \frac{\int e^{-\beta[E(x)+w_i(\rho'(x))]} \delta(\rho'(x) - \rho) d^N x}{\int e^{-\beta[E(x)+w_i(\rho'(x))]}}. \tag{2.63}$$

Combining equations 2.62 and 2.63 a relation between the biased probability distribution and the unbiased probability distribution can be made:

$$P_i^u(\rho) = P_i^b(\rho) e^{\beta w_i(\rho)} \left\langle e^{-\beta w_i(\rho)} \right\rangle. \tag{2.64}$$

Inserting equation 2.64 into 2.61 we can get the Gibbs free energy for each window as

$$G_i(\rho) = \frac{-1}{\beta} \ln P_i^b(\rho) - w_i(\rho) + f_i \tag{2.65}$$

where f_i is the free energy coming from the adding of the biasing potential w_i:

$$f_i = \frac{-1}{\beta} \ln \left\langle e^{-\beta w_i(\rho)} \right\rangle. \tag{2.66}$$

Since f_i does not depend on the reaction coordinate ρ it only describes the origin of the free energy of each window. This is non-important for the calculation of the free energy in each window but it is crucial for the calculation of the free energy $G(\rho)$.

There are a variety of methods to compute f_i but in this work, we used the so-called Weighted Histogram Analysis Method (WHAM) [118]. This method uses a weighted average of the distribution of the individual windows to compute the global distribution:

$$P^u(\rho) = \sum_i p_i(\rho) P_i^u(\rho) \tag{2.67}$$

where $p_i(\rho)$ are the weights that have to be normalized as

$$\sum_i^N p_i(\rho) = 1. \tag{2.68}$$

The weights are calculated by minimazing the statistical error of the total probability distribution, that is to say,

$$\sigma^2[P^u(\rho)] = \sum_i^N p_i^2(\rho) \sigma^2[P_i^u(\rho)]. \tag{2.69}$$

To minimize this statistical error with the constraint in eq. 2.68, the standard method of Lagrange multipliers has to be used. Therefore, the following equation has to be solved:

$$\frac{\partial}{\partial p_i} \left[\sum_i^N p_i^2(\rho) \sigma^2[P_i^u(\rho)] - \mu \left(\sum_i^N p_i(\rho) - 1 \right) \right] = 0 \tag{2.70}$$

where μ is the Lagrange multiplier. Solving eq. 2.68 and 2.70 we get that the weights only depend on the statistical error of the unbiased probability of each window:

$$p_i(\rho) = \frac{(\sigma^2[P_i^u(\rho)])^{-1}}{\sum_i^N (\sigma^2[P_i^u(\rho)])^{-1}}. \tag{2.71}$$

The statistical error of the biased probability distribution can be expressed as

$$\sigma^2[P_i^u(\rho)] = \frac{g_i(\rho)}{n_i \Delta\rho} \overline{P_i^b(\rho)} \tag{2.72}$$

where $\overline{P_i^b(\rho)}$ is the computed histogram for an infinite length simulation, n_i is the number of frames used to compute $P_i^u(\rho)$ and $\Delta\rho$ is the width of the bins used to calculate the histogram. Furthermore,

$$g_i = \left(1 + \frac{\tau_i}{\delta t}\right) \tag{2.73}$$

where τ_i is a correlation time and δt is the time step between two adjacent frames. The biased probability distribution $\overline{P_i^b(\rho)}$ can be represented in terms of the unbiased probability distribution using eq. 2.64 and 2.66. In this case, since the probability densities would be the ones computed for an infinite time P_i^u can be replaced by P^u,

$$\overline{P_i^b(\rho)} = e^{-\beta[w_i(\rho)-f_i]}P^u(\rho) \tag{2.74}$$

6 | EVOLUTIONARY ALGORITHMS

One of the objectives of material science is to find the atomic structure of a material so that its properties can be predicted. Although this is always done experimentally predictions of the atomic structures can be done computationally exploring the potential energy surface of a material and finding the most stable structures [119]. The main problem of exploring the whole potential energy surface is the huge amount of different possibilities that even a few atoms can offer to create a crystal structure. Doing a basic calculation we could compute the number of combinations as

$$C = \binom{V/\delta^3}{N} \prod_i \binom{N}{n_i} \tag{2.75}$$

where V is the volume of the unit cell, N is the number of atoms, δ is the relevant discretization parameter and n_i is the number of atoms of ith type in the unit cell. Taking rounded numbers like a unit cell volume of $V = 10A^3$, a discretization parameter of $\delta = 0.1A$ and $N = 20$ atoms of the same type the number of combinations rises to the order of 10^{61}. This massive number of different combinations makes the search for the absolute minimum of the potential energy landscape rather difficult. One thing that can be done to simplify the problem is to combine

the search of the absolute minimum with local minimum optimization. This reduces the dimensionality of the system since only the different local minima have to be searched to find the absolute minimum. However, the problem still has a huge amount of different combinations for most materials making a random search quite useless.

There are several methods to find an absolute minimum in complex functions like the potential energy landscape but the most used ones are called Evolutionary Algorithms (EA). Like most of the solutions human beings find to any big problem, this one is also based on nature, specifically in Darwin's theory of evolution. Evolutionary algorithms (sometimes also called genetic algorithms) have three main steps. The first one is to create a group of random solutions to the problem. This group of solutions is called a generation. Then, these solutions are compared between them like in nature and only the solutions that fit best the problem survive. This way, the bad solutions are discarded. Finally, the solutions that have survived are used to create a new generation by small mutations or combinations between them. Once a new generation is created the whole process can be repeated by the survival of the fittest and creating new generations over and over again. Every generation, a solution that fits better the problem should appear until the best solution is found. It is now easy to see the analogy with Darwin's evolution theory in which only the species that fit better in their environment will survive optimizing little by little the animals to their conditions.

In this work, evolutionary algorithms have been used to find the best structure of a material with certain stoichiometry both in bulk phase and in a nanocluster. For example, in chapter 5 EA were used to find the best clusters for different C-S-H stoichiometries. For that purpose, the USPEX code was used [120]. The code works as shown in Fig. 2.2. First, given a certain stoichiometry, the first generation of structures is created semi-randomly and locally optimized. Then, after discarding the worst structures, the remaining structures of that generation are used to create a new generation by mutating a parent structure or by the inheritance of various parent structures. Moreover, a small group of the very best structures is directly incorporated into the next generation.

6.1 | INITIALIZATION

The first generation of structures could be generated by fully randomizing the positions of the atoms. However, it has been shown that fully
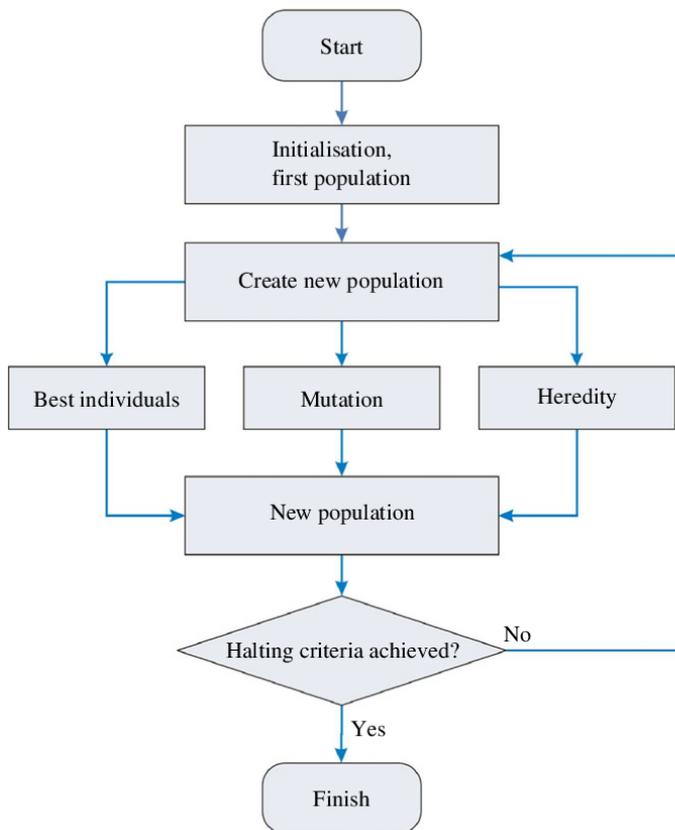
Figure 2.2: Flowchart of the evolutionary algorithm implemented in USPEX. Image from ref. [121]

random structures lead to a generation with structures that have similar (high) energies and a low degree of order, especially for large systems. Therefore, two techniques are applied to get more realistic structures. The first one is to divide the unit cell into smaller subcells [122]. Then, some atoms are placed randomly in the subcell and then replicated in the other subcells. This method provides more ordered structures, which are usually much better from the energetic and chemical points of view. This method really accelerates the process of finding the best structure when dealing with big structures. The other technique used to create a better initial structure is the use of space groups [121]. To do so, one of 230 space groups compatible with the number of atoms is selected and the structure is created following the symmetry operations. A small number of semi-randomly created structures can be added to each generation to improve the diversity of the population.

6.2 | VARIATION OPERATORS

Variation operators play a key role in evolutionary algorithms since they are responsible for creating a better and more diverse population from the last generation. Variation operators are usually only applied to the best-fitted structures of the last generation (frequently around 60% of the best structures from the last generations are used). There are two main variation operators: heredity and mutation.

Two parent structures are needed for the heredity operator. In this operator, spatially coherent blocks are taken from the parents' structure and they are merged into a new single-child structure. To choose the blocks in the parent structures, the order parameter of the environments are calculated. The order parameter defines the degree of order of a certain environment around an atom. Having this information allows the algorithm to choose the blocks depending on the degree of order. For most of the systems, a greater degree of order is preferred for lower energy structures but in some cases, it can be the other way around.

Within the mutation operators, different types of mutations can be found such as softmutation, lattice mutation or permutation. These operators somehow mutate a parent structure to create a new structure for the new generation. Softmutaions are based on slightly moving the atoms along the softest eigenmodes. This is done because low-frequency eigenmodes correspond to low curvature directions in the potential energy surface and probably also low energy barriers [123]. Cell mutation

refers to the transformation of the unit cell so that other parts of the potential energy surface are explored. If evolutionary algorithms are used to find a cluster the cluster will be expanded or contracted in some direction. Another very used mutation operator is the permutation of atoms. In this operator, two atoms of different types are exchanged generating a completely new structure. This can be done a variable number of times for every structure.

7 | VARIABLE CELL NUDGE ELASTIC BAND

Variable cell nudge elastic band (VC-NEB) is a method to study structural phase transitions in solids [124]. To do so, this method creates a set m of images $\{\mathbf{Y}_1, \mathbf{Y}_2, ..., \mathbf{Y}_m\}$ connecting the initial $\mathbf{Y}_1 = \mathbf{Y}_A$ and final $\mathbf{Y}_m = \mathbf{Y}_B$ structures, where A represents de initial structure and B represents de final structure. Each image is described in configuration space by a vector of 9+3M components. That is to say, 3 coordinates for each atom a 9 coordinates to represent the unit cell. This vector can be written as $\mathbf{Y} = (\varepsilon_{1i}, \varepsilon_{2i}, \varepsilon_{3i}, \vec{y}_1, ..., \vec{y}_M)(i = 1, 2, 3)$ where $\boldsymbol{\varepsilon}$ is the stress tensor, $\vec{y}_i$ is the position vector of each atom and M the number of atoms. Every image $\mathbf{Y}_i$ is connected to the previous $\mathbf{Y}_{i-1}$ and following $\mathbf{Y}_{i+1}$ images by a spring.

Under an applied pressure P, the enthalpy $H = E - P\Omega$ of every structure is determined by the (9+3M)-dimensional potential energy surface. Here Ω is defined as $\Omega = det[(1 - \bar{\bar{\varepsilon}})\mathbf{h_0}]$ where $\mathbf{h_0}$ are the initial cell vectors used as a reference configuration and $\bar{\bar{\varepsilon}}$ contains the 9 coordinates of the stress tensor. The expanded force vector $\mathbf{F}$ with 9+3N components can be defined then by the derivative of the enthalpy H as :

$$F = \frac{\partial H}{\partial \mathbf{Y}}\bigg|_P \tag{2.76}$$

This force vector is composed of the forces acting on the lattice $f_{\bar{\bar{\varepsilon}}}$ and the forces acting on the atoms f_v. The tangent vector τ along the path in the NEB method is represented as the unit vector to neighboring images. Therefore, we can decompose the forces in the tangent vector and the vector connecting the images. The VC-NEB force $\mathbf{F}^{\text{VC-NEB}}$ is composed of the tangent force of the potential force acting on the lattice and atoms,

defined as $f_{\bar{\bar{\varepsilon}}}^{\nabla\perp}$ and $f_v^{\nabla\perp}$ respectively, and the spring force linking together the lattice vectors and atoms of consecutive images, $f_{\bar{\bar{\varepsilon}}}^{s\|}$ and $f_v^{s\|}$ respectively. Therefor, the VC-NEB force $\mathbf{F}^{\text{VC-NEB}}$ can be defined as:

$$f_{\bar{\bar{\varepsilon}}}^{\text{VC-NEB}} = f_{\bar{\bar{\varepsilon}}}^{\nabla\perp} + f_{\bar{\bar{\varepsilon}}}^{s\|} \tag{2.77}$$

$$f_v^{\text{VC-NEB}} = f_v^{\nabla\perp} + f_v^{s\|} \tag{2.78}$$

$$\mathbf{F}^{\text{VC-NEB}} = \left(f_{\bar{\bar{\varepsilon}}}^{\text{VC-NEB}}, g f_1^{\text{VC-NEB}}, g f_2^{\text{VC-NEB}}, ..., g f_M^{\text{VC-NEB}} \right) \tag{2.79}$$

where the metric tensor $g = \mathbf{h}^T \mathbf{h}$ is introduced to keep the symmetry during structure relaxation.

Therefore, once an initial path is set up all the images are relaxed minimizing the $\mathbf{F}^{\text{VC-NEB}}$ force which contains the spring forces and the transverse components of the potential forces which were computed with density functional theory (DFT). This method was used both in chapter 3 and 4 to study the transfomation between portlandite and Phase V. The images were relaxed until the forces were below 0.01 eV/Å.

8 | COMPARING STRUCTURES

One of the reasons for using the methods explained above like US, TMD or evolutionary algorithms is to explore the potential energy surface to find local minimums, global minimums, minimum energy paths, etc... Linking those properties with structural properties like distances, angles, coordination numbers, type of atoms or many others is vital to understand the behavior of those systems. Some basic structural properties like the distance between two types of atoms or the angle between 3 types of atoms can be easily computed but most of the time the structural properties linking the system to the energetical properties are much more complex. This complexity together with the amount of structural data that we can create with the mentioned methods requires the automation of structural comparison to analyze, classify and represent such a large amount of data.

To do these things, first, the similarity between two structures has to be defined mathematically. The simplest metric for that could be the Eu-

clidean distance between the cartesian coordinates of the atoms, usually
called root mean square displacement (RMSD). Moreover, this distance
could easily be made invariant to rotations and translations, two indis-
pensable features for a global metric. However, the RMSD cannot be
easily extended to systems in which the atoms cannot be unequivocally
mapped between different structures. Therefore, in addition to transla-
tion and rotation invariance, the chosen metric must also be invariant to
permutations. The RMSD metric can be made invariant by considering
all possible rotations of the structure and by finding the smallest RMSD
by a minimization process. However, this increments substantially the
computational cost [125].

In the last decade, a variety of structural descriptors, also called fin-
gerprints, have been proposed using different approaches. Some de-
scriptors have been developed based on graph-theoretic procedures (e.g.
SPRINTS[126]). Others have used analogies with electronic structure
methods using descriptors based on the Hamiltonian matrix, Hessian
matrix, overlap matrix of Gaussian Type Orbitals (GTOs) or even Kohn-
Sham eigenvalue fingerprints [125]. Other methods such as Coulomb
matrices [127], bags of bonds [128], symmetry functions [129], and scat-
tering transformation applied on a linear superposition of atomic densi-
ties [130] have been used to generate descriptors to be used in machine
learning, to predict properties of materials and molecules thus avoiding
computationally expensive *ab initio* calculations.

Although the aforementioned methods are very useful, the most promis-
ing methods, which are also invariant to translations, rotations and per-
mutations, are those that describe the atomic environment of each atom.
These environments are then combined in various ways to create a struc-
tural descriptor that can be compared to that of another structure. One
of the most common ways to compare these structures is to look for the
atomic environment that is most similar in both structures. However,
other ways such as averaging all environments or a mix between av-
eraging and comparing the best environment are also used. Various de-
scriptors based on the atomic environment have been proposed and used
before [131], however, in this thesis the smooth overlap of atomic posi-
tions (SOAP) [132, 133] has been mostly used since it has been proven to
be very robust when comparing materials and molecules. That is why
SOAP has been used to compare the structure of the MD trajectories
in chapter 4 and to compute the similarities between all the clusters in
chapter 5. For this reason, we will now see how the SOAP fingerprints,
its similarity kernel $k(\chi, \chi')$ and the kernel distance $d(\chi, \chi')$ are con-

structed.

8.1 | SMOOTH OVERLAP OF ATOMIC POSITIONS (SOAP)

As its name suggests, SOAP is based on calculating the overlap between different environments in a way that is invariant to rotations, translations and permutations. In this way, the similarity kernel $k(\chi, \chi')$ between two environments centered on an atom can be calculated. The similarity kernel between two environments can then be extended to calculate the similarity between two structures. But let us first start by defining the environment χ around an atom.

8.1.1 | KERNEL DISTANCE BETWEEN DIFFERENT ENVIRONMENTS

The environment centered on an atom can be calculated using the atomic neighbor density. Although the atomic density is usually computed by a sum of Dirac-delta functions, this would lead to a discontinuous similarity kernel. Therefore, the atomic neighbour density $\rho_\chi(\boldsymbol{r})$ of the environment χ is constructed here using gaussian functions of a width σ as

$$\rho_\chi(\boldsymbol{r}) = \sum_{i \in \chi} exp\left(\frac{\boldsymbol{r} - \boldsymbol{r}_i}{2\sigma}\right) \tag{2.80}$$

where $\boldsymbol{r}_i$ represents the position of every atom in the environment within a certain cut-off. For simplicity, the atomic neighbor density can be expanded in terms of spherical harmonic functions:

$$\rho_\chi(\boldsymbol{r}) = \sum_{i \in \chi} exp\left(\frac{\boldsymbol{r} - \boldsymbol{r}_i}{2\sigma}\right) = \sum_{i \in \chi} \sum_{lm} c_{lm}^i(r) Y_l^m(\hat{\boldsymbol{r}}). \tag{2.81}$$

With this simple expansion, we can now describe the atomic neighbor density with the desired accuracy depending on the number of spherical harmonic functions, which is very important when it comes to saving

computing time. However, the computation of the atomic neighbor density is still very dependent on the number of atoms inside the cut-off radius of the environment. Therefore, in the same manner, done with the spherical harmonic functions, we can also expand the atomic neighbor density into a set of radial functions $g_n(r)$ that form an orthonormal basis, i.e., $\int dr\, g_n(r) g_{n'}(r) = \delta_{nn'}$:

$$\rho_\chi(\boldsymbol{r}) = \sum_{i \in \chi} exp\left(\frac{\boldsymbol{r} - \boldsymbol{r}_i}{2\sigma}\right) = \sum_n \sum_{lm} c_{nlm}(r) g_n(r) Y_l^m(\hat{\boldsymbol{r}}). \quad (2.82)$$

Thus, the environment χ can be described by the set of the expansion coefficients c_{nlm}. The overlap between two different environments $S(\rho, \rho')$ can be computed as the inner product of their atomic neighbor densities,

$$S(\rho, \rho') = \int d\boldsymbol{r}\, \rho(\boldsymbol{r}) \rho'(\boldsymbol{r}). \quad (2.83)$$

It has to be noted that from now on $\rho_\chi = \rho$ and $\rho_{\chi'} = \rho'$. This overlap certainly satisfies the invariance under translation or permutation but it does not comply with the invariance under rotations since it depends on the reference axes. This can be overcome by integrating one of the environments under every possible rotation $\hat{R}$ getting the desired similarity kernel

$$k(\rho, \rho') = \int d\hat{R}\, |S(\rho, \hat{R}\rho')|^2 \quad (2.84)$$

By inserting the expansion of the atomic neighbor density in Eq. 2.82 into Eq. 2.84, the similarity kernel can be written in terms of the expansion coefficients c_{nml},

$$k(\rho, \rho') = \sum_{nn'lmm'} c_{nlm} \left(c'_{nlm'}\right)^* \left(c_{n'lm}\right)^* c'_{n'lm'} \quad (2.85)$$

where the power spectrum p of the environment ρ can be defined as:

$$p_{nn'l} = \sum_m c_{nlm} \left(c_{n'lm}\right)^*. \quad (2.86)$$

Finally, since the kernel distance has to be a metric defined between [0,1), the kernel similarity has to be slightly tweaked so that it fulfills that requirement. Thus,

$$d(\rho, \rho') = \sqrt{2 - 2k(\rho, \rho')} \qquad (2.87)$$

8.1.2 | FROM ENVIRONMENTS TO COMPLETE STRUCTURES

Now that the distance between two different environments is defined, we can extend that to the distance between two complete structures. For simplicity, let us first focus on structures with a single specie and the same number of atoms. It is important to notice that a structure with N atoms will have N different environments, each centered on one of the atoms. Therefore, given two different structures A and B with N atoms each a covariant matrix can be computed with every possible paring between environments:

$$C_{ij}(A, B) = k(\chi_i^A, \chi_j^B). \qquad (2.88)$$

The covariant matrix C_{ij} contains all the information on the pair-wise distance between all the environments of the two structures. From the covariant matrix, there are several ways of computing a distance between two complete structures. The first option is to compute the distance between two structures with the average SOAP fingerprints of the environments in each structure. Then, a kernel is computed with this averaged SOAP fingerprint:

$$\bar{K}(A, B) = \left[\frac{1}{N} \sum_i \mathbf{p}(\chi_i^A)\right] \cdot \left[\frac{1}{N} \sum_j \mathbf{p}(\chi_j^B)\right] = \frac{1}{N^2} \sum_{ij} C_{ij}(A, B) \qquad (2.89)$$

where $\mathbf{p}$ contains the elements of the power spectrum $P_{nn'l}$ collected into a unit length vector.

Another way of comparing two structures is to use the best-match structural kernel. In this case, all the environments of structure A are compared with all the environments of structure B so that the best-matching environment is found. The best-match kernel is then constructed with these two environments:

$$\hat{K}(A, B) = \frac{1}{N} \max_{\pi} \sum_{i} C_{i\pi_i}(A, B) \tag{2.90}$$

Although some other methods to compare structures have been proposed like the regularized entropy match kernel or the permutation structural kernel, these methods are combinations or extensions of the methods presented here. In this thesis, the best-match structural kernel has been mostly used.

Dealing with different species is straightforward with the SOAP fingerprints. This is done by building different atomic densities for each specie. That is to say, different environments are defined for each specie. Similarly as done in Eq. 2.85 the kernel distance can now be written as

$$k(\chi, \chi^{\prime}) = \sum_{\alpha\alpha^{\prime}\beta\beta^{\prime}} \mathbf{p}_{\alpha\beta}(\chi)\mathbf{p}_{\alpha^{\prime}\beta^{\prime}}(\chi^{\prime})\kappa_{\alpha\alpha^{\prime}}\kappa_{\beta\beta^{\prime}} \tag{2.91}$$

where $\kappa_{\alpha\beta}$ is defined as the alchemical similarity kernel between species α and β which is one for pairs that should be considered interchangeable, and tend to zero for pairs that one wants to consider as completely unrelated.

8.2 | DIMENSIONALITY REDUCTION ALGORITHMS

One of the problems when analyzing the similarity between a big set of structures is the high dimensionality of the kernel in Eq. 2.90 which contains the pair-wise similarity kernel between all the structures. To visualize the similarity between structures and identify clusters of similar structures which share common structural properties, a dimensionality reduction algorithm is needed. There are several methods of doing a dimensionality reduction but the kernel component principal analysis (KPCA) and Sketchmap have been mostly used in this work.

8.2.1 | KERNEL PRINCIPAL COMPONENT ANALYSIS

KPCA is a dimensionality reduction algorithm based on principal component analysis (PCA). PCA is a linear dimensionality reduction method in which the covariant matrix of a dataset is diagonalized so that the eigenvectors represent the main components of the dataset. However,

in most cases, the dataset is non-linear (as in structural similarity data) and the PCA will not reproduce a correct dimensionality reduction. In these cases, a non-linear dimensionality reduction method like KPCA can be used.

The idea of KPCA relies on the intuition that many datasets, which are not linearly separable in their space, can be made linearly separable by projecting them into a higher dimensional space. The added dimensions are just simple arithmetic operations performed on the original data dimensions. Projecting the dataset into a higher dimensional feature space, it becomes linearly separable and therefore, PCA can be applied to this new dataset. Performing this linear dimensionality reduction in that space will be equivalent to a non-linear dimensionality reduction in the original space.

Practically speaking, the kernel matrix is calculated as a result of calculating the kernel distance between all the pair structures. Then, that matrix is diagonalized and the eigenvalues and eigenvectors are calculated. Since the eigenvectors with higher eigenvalues are the most important components the eigenvectors are sorted from high eigenvalues to low eigenvalues. Choosing the first eigenvectors to represent our data will reduce the dimensionality of our system and represent them in terms of the most important eigenvectors. Let us not forget that the meaning of these eigenvectors is a priori unknown and that more complex analyses have to be performed to know their physical meaning.

8.2.2 | SKETCHMAP

A sketch-map is another dimensionality reduction strategy specifically designed to represent the high-dimensionality data produced in atomistic simulation. This tool is based on reproducing the distance between a set of high-dimensionality data points in a space of lower dimension. As reproducing all distances correctly is a very complicated task, sketch-map takes care of reproducing very accurately the most interesting distances. To do this, it transforms distances in both the high and low dimensional spaces so that structures separated by a distance greater than Rc are considered very far away and structures that are at a distance smaller than Rc are considered very close.

The dimensionality reduction is done by minimizing a stress function χ which calculates the error between the distances in the high and low

dimensional spaces:

$$\chi^2 = \left(\sum_{i \neq j} w_i w_j\right)^{-1} \sum_{i \neq j} w_i w_j \left[F(R_{ij}) - f(r_{ij})\right] \tag{2.92}$$

where R_{ij} and r_{ij} are the distances between structure i and j in high and low dimensionality space respectively, w_i and w_j are the weights of these structures in high and low dimensional spaces and F and f are sigmoid functions of the form:

$$S_{a,b,\sigma}(r) = 1 - (1 + (2^{a/b} - 1)(r/\sigma)^a)^{-b/a} \tag{2.93}$$

where $S_{a,b,\sigma}(\sigma) = 1/2$ and the exponents a and b represent the rate at which the function approaches 0 and 1 respectively.

CHAPTER 3

TRANSFORMATION IN BULK PORTLANDITE

Music for a While
Music for a While
Henry Purcell - Christina Pluhar

A s explained in the Introduction, one of the hypotheses of this thesis is to evaluate if small portlandite monolayers can act as precursors of C-S-H nucleation. Thus, proposing a hydration mechanism that can be complementary to other mechanisms. In our hypothesis, the portlandite monolayers will be transformed by the interaction with the silicate dimers, creating a C-S-H layer. The interaction between the silicate dimers present in cement hydration and the portlandite layer will be studied in chapter 4. For now, we will focus on the similarity between the Ca-O arrangement in tobermorite (the C-S-H crystal structure with space group $P\bar{3}m1$) and the Ca-O arrangement in bulk $Sr(OH)_2$ with space group $Pnma$, as shown in Fig. 3.1. Therefore, in this chapter we will study the possible phase transition between portlandite and $Ca(OH)_2$ with symmetry $Pnma$.

Portlandite, the main phase of bulk $Ca(OH)_2$, is formed together with C-S-H as a result of Alite and Belite hydration. It represents as much as 20-25%[134] of the cement paste volume. Therefore, it is of great interest to the cement and concrete industry since it plays a big role in the mechanical properties of the final product of hydration. In addition to the importance it has for cement, portlandite powder is also used for the flue desulfurization process, providing a low-cost SO_2 control option for

coal-fired power plants [135, 136]. Moreover, it is used for the absorption of atmospheric carbon dioxide which is crucial for controlling CO_2 emissions [137, 138, 139, 140]. Calcium hydroxide can also be used as a thermochemical energy storage by means of reversible gas-solid reactions [141] and it is a very important material to buffer the pH against steel corrosion [142].

Unlike portlandite, whose space group is $P\bar{3}m1$, the structure of bulk $Sr(OH)_2$ is $Pnma$. This structure contains a zig-zag Sr-O structure very similar to that of Ca-O in tobermorite as shown in Fig. 3.1. Since Ca and Sr are isoelectronic, it would be reasonable to think that there could be a $Ca(OH)_2$ phase with the same structure as $Sr(OH)_2$, i.e., hypothetical $Pnma$ $Ca(OH)_2$ structure. Therefore, in this chapter, the stability of the $Ca(OH)_2$ structure with $Pnma$ is analyzed and compared to the other $Ca(OH)_2$ phases found before. Then, the possible structural transformation from portlandite to the $Pnma$ $Ca(OH)_2$ phase is studied since it is of great interest to then study the monolayer transformation that would form the C-S-H layer.

1 | $Ca(OH)_2$ polymorphs

The atomic structure of portlandite has been well established for many years [143]. It belongs to the $P\bar{3}m1$ space group and contains stacked layers of CaO_6 edge-sharing octahedra, that is to say, it comprises six-fold coordinated CaO_6[144]. Each of the O atoms at the edges of the octahedron is bound to an H atom which interacts with three hydroxyl groups from the adjacent layer (see Fig. 3.1).

Besides portlandite [145], the stable phase at room conditions, various high-pressure polymorphs of $Ca(OH)_2$ have been reported [146, 147, 148, 149]. As early as 1963, the pressure decomposition of portlandite was studied at different temperatures determining the phase boundary at 2GPa [144]. Due to the grain size used in the experiments of references [145, 150, 151], these authors claimed that portlandite suffered an amorphization at high pres-
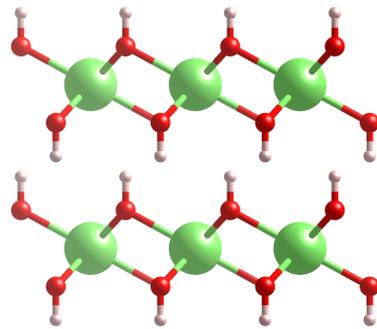
Figure 3.2: Atomic structure of portlandite.
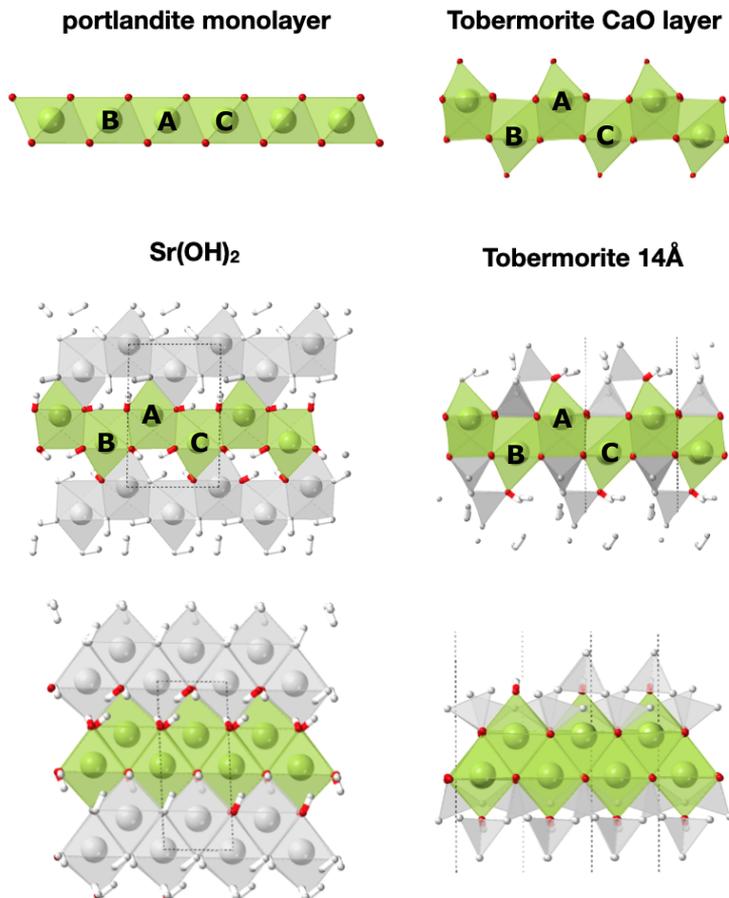
Figure 3.1: In the upper images the difference between the portlandite layer and the tobermorite layer is shown. The bottom images show the similarities between the bulk $Sr(OH)_2$ structure and tobermorite.

sures. However, a high-pressure polymorph with $P2_1/c$ space group was found in 1996 using in situ X-ray diffraction in a multi-anvil high-pressure device (DIA) [152]. This new polymorph was named as Phase II in the literature and although initially it was reported with a similar crystal structure to that of Baddeleyite (ZrO_2), it has been proven later that this was not accurate. 12 years later phase II' was discovered at high pressure and temperature with $I121$ space group [153, 154]. This phase was found as an intermediate step in the phase transition between portlandite (Phase I) and Phase II. Finally, in recent years two more phases were found theoretically using DFT: Phase III and Phase IV [149]. Phase III is a monoclinic $P21/c$ phase which is energetically superior to phase II above 23 GPa at 0 K, where the phase transition pressure becomes higher with increasing temperature. After 78 GPa, it transforms into

Phase	Pressure (GPa)	Lattice Parameters (A)	
Phase I (portlandite) [143] $P\bar{3}m1$ (Exp)	0	a=b=3.5853 c=4.9110	$\alpha=\beta=90°$ $\gamma=120.0°$
Phase II (Baddeleyite) [147] $P2_1c$ (Exp)	0	a=5.3979 b=6.0931 c=5.9852	$\alpha=\gamma=90.0°$ $\beta=103.581$
Phase II' [154] $I121$ (Exp)	11	a=5.793 b=6.733 c=8.845	$\alpha=\gamma=90.0°$ $\beta=104.46$
Phase III [149] $P2_1c$ (Z=4) (Theory)	40	a=4.6333 b=4.9778 c=6.2391	$\alpha=\gamma=90.0°$ $\beta=69.8$
Phase IV [149] $Pnma$ (Theory)	80	a=4.5288 b=3.7000 c=6.7873	$\alpha=\beta=\gamma=90.0$
Phase V $Pnam$ (Theory)	0	a=9.4282 b=5.7342 c=3.5854	$\alpha=\beta=\gamma=90.0$

Table 3.1: Space groups and lattice parameters of the different polymorphs of Ca(OH)$_2$ at different pressures.

phase IV with $Pnma$ space group. Although the structure we are interested in, that of Sr(OH)$_2$, also shares the $Pnma$ space group, Phase IV has a different structure. Therefore, we have named Phase V the Ca(OH)$_2$ phase with $Pnma$ space group with a structure similar to that of Sr(OH)$_2$. In table 3.1 all the phases together with the corresponding space groups and lattice parameters are presented.

2 | PHASE V: STRUCTURE AND (META)STABILITY

among the phases reported in the literature (Phases I-IV) none of them has a similar layered structure to the Ca-O layer found in tobermorite. Nonetheless, such a layered structure can be found in Phase V. Therefore, this phase becomes very interesting to later study the transformation between a portlandite monolayer and a C-S-H layer.

To study the stability of the new Phase V, its energy together with that of portlandite and other polymorphs was calculated using DFT for different pressures. As expected, portlandite is the most stable structure at
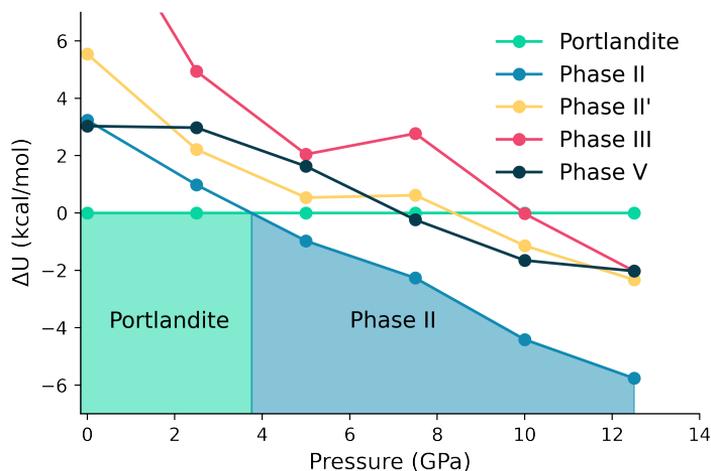
Figure 3.3: Energy difference of the different found polymorphs of $Ca(OH)_2$ compared to portlandite. Phase IV was not included since its energy is too big in this pressure range compared to that of portlandite.

low pressures (see Fig. 3.3). Phase V is a metastable polymorph, with an energy difference with respect to portlandite of 2.78 kcal/mol at 0 GPa. Although Phase V is not the lowest energy polymorph at any pressure, it is always among the lowest energy structures (at least in the 0-12 GPa range). The low energy difference between portlandite and Phase V at 0 Gpa is small compared with the other reported polymorphs at normal pressure [155, 156]. However, the Phase III polymorph presents a similar energy difference to that of Phase V.

Simulation details:
The DFT minimzations were done using VASP [157, 158] with a revised Perdew-Burke-Ernzerhof for solids (PBE-sol) functional [92] of the generalized grandient approximation (GGA) [94]. The energy cut-off for the plane waves was set to 400 eV and the electronic SC-loop was stopped when the energy difference was less than 10^{-4} eV. A Γ-centered k-points mesh was used for the reciprocal lattice with a spacing between the points of 0.025 Å^{-1}. The ionic minimization was performed using a quasi-Newton method (RMM-DIIS) [159] until all the forces were less than 0.01 eV/Å.

The relaxed atomic structure of Phase V can be found in table 3.2.
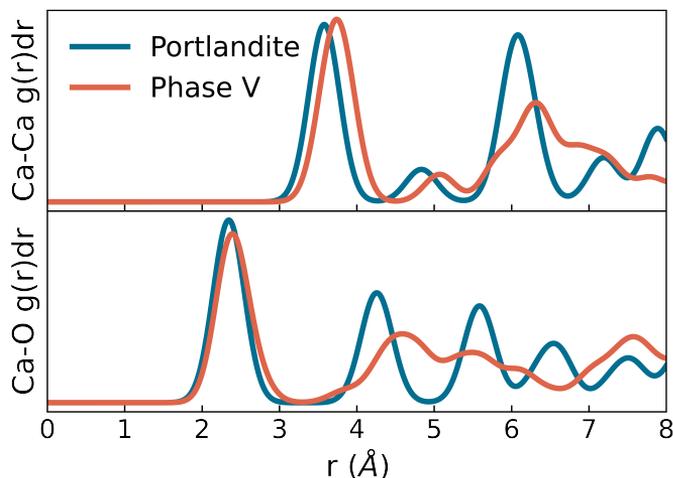
Figure 3.4: Radial distribution function of Ca-Ca distances and Ca-O distances in portlandite and Phase V.

Atom	x	y	z
Ca	0.1606	0.2500	0.9033
O1	0.3973	0.2500	0.6546
O2	0.3716	0.2500	0.1350
H1	0.4769	0.2500	0.1458
H2	0.4088	0.2500	0.8264

Table 3.2: Atomic structure of Phase V in internal coordinates calculated with DFT. The crystal structure has $Pnma$ symmetry with cell parameters a=9.2536 Å, b=3.5751 Å and c=5.7116 Å.

The structure of Phase V contains a seven-folded Ca-O zig-zag layer. Each of the Ca atoms shares a hydroxyl with a Ca atom of the above or below the layer, thus joining the layers together. The radial distribution function has been computed for portlandite and Phase V and compared in Fig. 3.4. It is important to notice that both the Ca-Ca distance and the Ca-O distance are bigger in Phase V. However, the cell volume of Phase V (47.24 Å/f.u) is smaller than that of portlandite (54.41 Å/f.u). This means that phase V is more densely packed despite its larger interatomic distances.

3 | STRUCTURAL TRANSFORMATION

T he structural transformation between portlandite and Phase V
was studied using the variable cell nudge elastic band (VC-NEB)
method. In this method, intermediate images (structures) along
the reaction path are created by interpolation of the atomic coordinates.
To do so, each atom at the beginning of the transformation has to be
mapped with an atom of the final structure. This can be quite delicate
since the mapping of these atoms could lead to different transformation
paths. Two different methods have been used to choose this initial path.
First, since the interest of this study is to investigate if this transforma-
tion could then happen in a monolayer, the atoms at the initial structure
have been mapped with the closest atoms of each type at the end of the
structure maintaining the orientation of the layers (as shown in Fig. 3.5).
Secondly, to search other transformation paths different polymorphs of
$Ca(OH)_2$ have been found in sec 3.2 using evolutionary algorithms and
used as an intermediate step in the transformation.

3.1 | INTRALAYER TRANSFORMATION

As stated just before, the first method to create an initial path is to map
the atoms in portlandite to the closest atoms in Phase V maintaining the
direction of the layer as shown in the upper figures in Fig. 3.1. The cre-
ated images are then optimized simultaneously using VC-NEB to find
the optimal transformation path. More information about how the min-
imization process can be found in section 7 of chapter 2.

> **Simulation details**
> The VC-NEB method [124, 160] was used within the USPEX code [119]. To calcu-
> late the forces for each image the VASP [157, 158] code was used with the GGA-
> PBE-sol functional[92, 94]. The electronic SC-cycle was performed until the energy
> difference was less than 10^{-3} with a Γ-centered k-point mesh with a spacing of
> 0.025Å^{-1} . 25 images were used for the VC-NEB calculation and the spring constant
> was set between 2-6 $eV/\text{Å}^2$ depending on the structural distance between consec-
> utive images. Finally, once the final path is obtained all the final image energies
> were computed using more refined DFT parameters as used in section 2.

In Fig. 3.5(a) the transformation enthalpy is shown as a function of
the image number. The computed transformation path shows a direct

transformation with no local minimum in between. During the transformation, the γ angle between three adjacent Ca atoms in the portlandite layer folds gradually to a 120 zig-zag layer [Fig. 3.5(g)]. This folding together with the Ca-O coordination is what characterizes the difference between portlandite and Phase V. This folding is accompanied by a slight shift between layers and the readjustment of the H atoms to form the structure of Phase V. Several structural parameters were examined to understand the origin of the energy barrier. The evolution of the mean distance between the bonded Ca-O atoms is the parameter that better correlates with the enthalpy [Fig. 3.5(d)]. As the Ca-O distance increases, the enthalpy increases up to the transition state. Then, the Ca-O distance decreases to the final value, accompanied by an increase in the Ca-Ca mean distance. Regarding the volume, it can be stated that until the transition state the volume remains almost constant and then starts to continuously decrease until almost the end where it suffers a small readjustment (final increase in volume). The energy barrier for this transformation is $10.13\,\mathrm{kcal/mol}$. This enthalpy barrier is very similar to other energy barriers calculated for different bulk phase transitions as in bulk thorium monocarbide (ThC) or bulk GaN [161, 124]. The low energy barrier and the fact that this transformation can occur in a portlandite monolayer are promising factors for the study of the proposed mechanism for C-S-H formation.

3.2 | LOOKING FOR OTHER TRANSFORMATION PATHS

The VC-NEB method finds the path of least energy closest to the initially proposed path. This has two main problems: on the one hand, the path depends too much on the initial guess. On the other hand, more complicated paths farther away from the initial guess are not explored and therefore, we cannot know if the path found is the lowest energy path globally. This happens in any kind of minimization where it is impossible to know if the minimum is global unless the whole configurational space is explored. The only thing that can be done is to explore other paths that traverse other parts of the potential energy surface. To do so, first, we try to find metastable polymorphs of $Ca(OH)_2$ that could potentially become intermediate steps of the transformation, using evolutionary algorithms. Evolutionary algorithms allow us to explore the potential energy surface finding different local minima and, in the end,
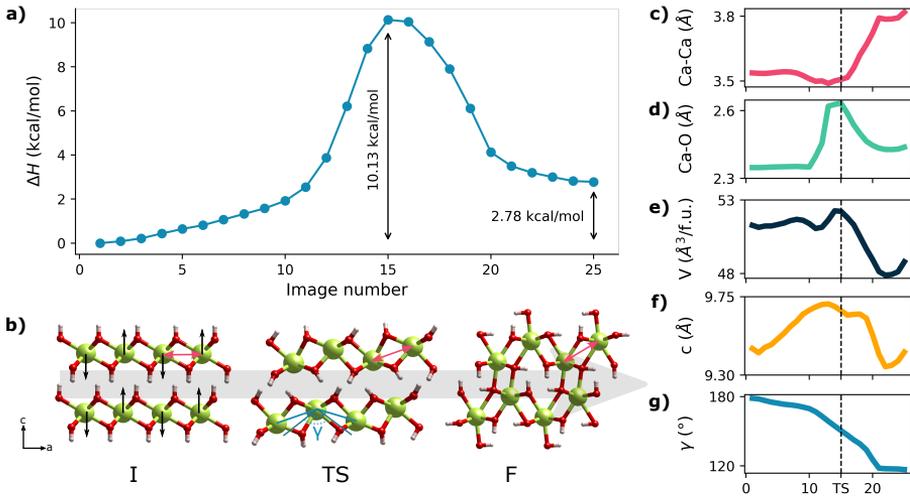
Figure 3.5: Intralayer transformation path from portlandite to Phase V. (a) Enthalpy difference along with the structural transformation for each image used in the VC-NEB method. (b) Graphical representation of the different stages along with the structural transformation. The calcium, oxygen and hydrogen atoms are illustrated with green, red and pink spheres respectively. [(c)-(g)] Respectively, the evolution of the Ca-Ca intralayer distance, Ca-O mean distance, volume, c cell vector and the folding angle γ between three adjacent Ca atoms in the intralayer transformation path.

finding the absolute minimum. In this case, we are only interested in the different local minima since we already know that portlandite is the absolute minimum. To avoid being stuck in one minimum metadynamics can be used together with evolutionary algorithms, that method is called evolutionary metadynamics.

Simulation details:
Evolutionary algorithms were used using the USPEX code [120, 119]. The stoichiometry used for the calculation was $Ca_4O_8H_8$ so that bigger unit cells are also included. Each generation contained 30 individuals and each individual was minimized using DFT in the VASP [157, 158] code with the same parameters used in sec. 2. The best 70% of the individuals of each generation were used to create the next generation. The structures of the next generation were obtained by the following methods: %60 heredity operator, %10 permutation operator, 10% of lattice mutation operator and 20% of random structure generation. After the first 25 generations, the evolutionary metadynamics calculation was triggered by introducing an energy penalty to the already visited structures. This is done by adding a gaussian to the potential energy surface around the already visited structures so that not only the visited structures are penalized but also similar structures. The Gaussian width was set to 0.1 Åand the maximum gaussian height to 100 $Å^3$·kbar. In total, 100 generations were produced.

3.2.1 | RESULTS

After the evolutionary algorithms simulations, the best 40 polymorphs were selected. Then, the structural similarities between them were calculated using the SOAP descriptor [132]. The SOAP similarity gives 0 for completely equal structures and approaches 1 for completely different structures. Since the purpose of looking for these polymorphs is to see if any of them could be an intermediate step in the transformation, it is important to see how different these polymorphs are from the path we have previously found. For that reason, Sketch-map was used to represent the similarities between the different polymorphs and the found transition path in a 2-dimensional plot. In figure 3.6 each circle represents a polymorph of $Ca(OH)_2$ and the blue line is the intralayer path found previously. The distance between the points in the plot represents the structural similarity between the structures. The closer they are the more similar they are, and the farther apart they are the more different they are.

To find new paths, different low-energy polymorphs were selected as intermediate steps. In this case, since the mapping of the atoms was not as easy as in the intralayer path we have found before, this task was done by using the *pmpath* utility from USPEX. This utility uses a geometric method for predicting solid-solid phase transition mechanisms. After mapping the atoms, the VC-NEB method was used to minimize the different transformation paths. Although most of them presented a considerably higher energy barrier, there was one path that had a lower energy barrier than the previous intralayer transformation. This path is represented in Fig. 3.6 as a purple dashed line. Although this transformation path is very interesting since it has the lower energy barrier, it is an interlayer transformation path, thus not being possible to occur in a $Ca(OH)_2$ monolayer.

3.3 | INTERLAYER TRANSFORMATION

The interlayer transformation path has the lower enthalpy barrier of all the paths found from portlandite to Phase V. This path contains an intermediate step (IS) and two transition states. One from portlandite to the IS (TS1) and the other one from the IS to Phase V (TS2). The IS is less stable than both portlandite and Phase V with an enthalpy difference of 6.31 kcal/mol with respect to portlandite. The enthalpy barrier
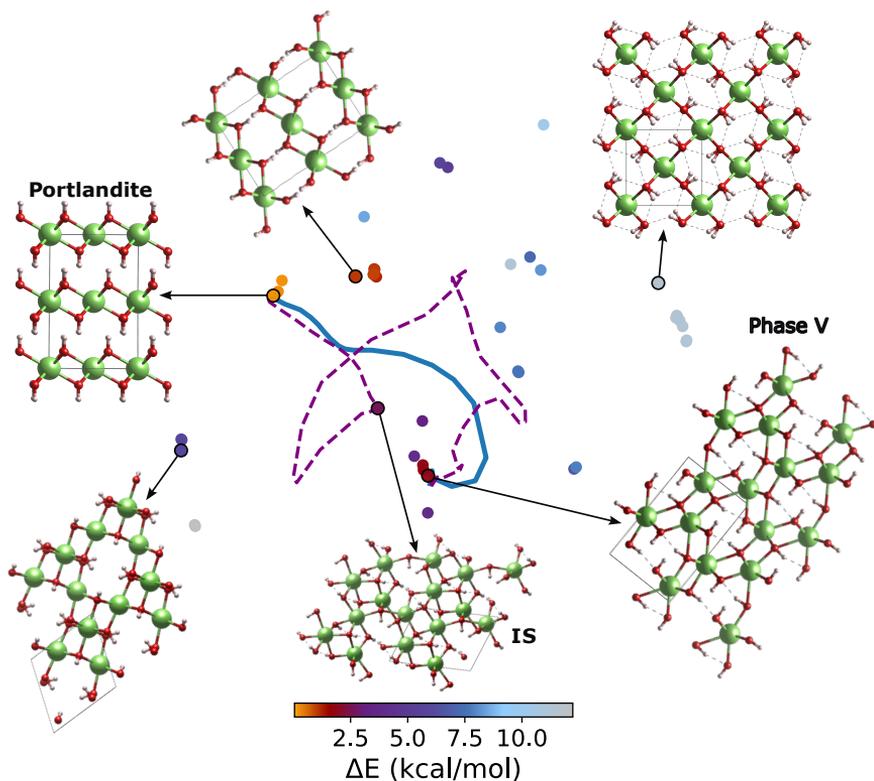
Figure 3.6: Sketch-map representation of different polymorphs and transformation paths of Ca(OH)$_2$. The dots represent the different polymorphs and the distance between the points represents the structural distance. The black line represents the original transition path and the red line is the path found by taking a polymorph as an intermediate step.

from portlandite to the IS and from the IS to Phase V is 8.17 kcal/mol and 9.51 kcal/mol respectively. Therefore, the enthalpy barrier for the whole transformation is 9.51 kcal/mol. This information is shown in Fig. 3.7(a).

To find any correlation between the structural changes in the transformation and the enthalpy change, different structural properties have been plotted in Fig. 3.7(c-i). It can be observed that the layers of portlandite are joined together during the transformation by looking at the Ca_1-Ca_3 and Ca_2-Ca_4 distances. It is important to mention that in portlandite Ca_1 is in a different layer than Ca_3 and Ca_4, as shown in Fig. 3.7(b). Although both Ca_1-Ca_3 and Ca_2-Ca_4 distances decrease during the transformation, thus causing the sheets to move closer together, the approaching of the layers does not occur at the same time in all parts of the sheet. The distance Ca_1-Ca_3 decreases substantially at the TS_1 while the Ca_1-Ca_3 distance decreases gradually between the TS_1 and
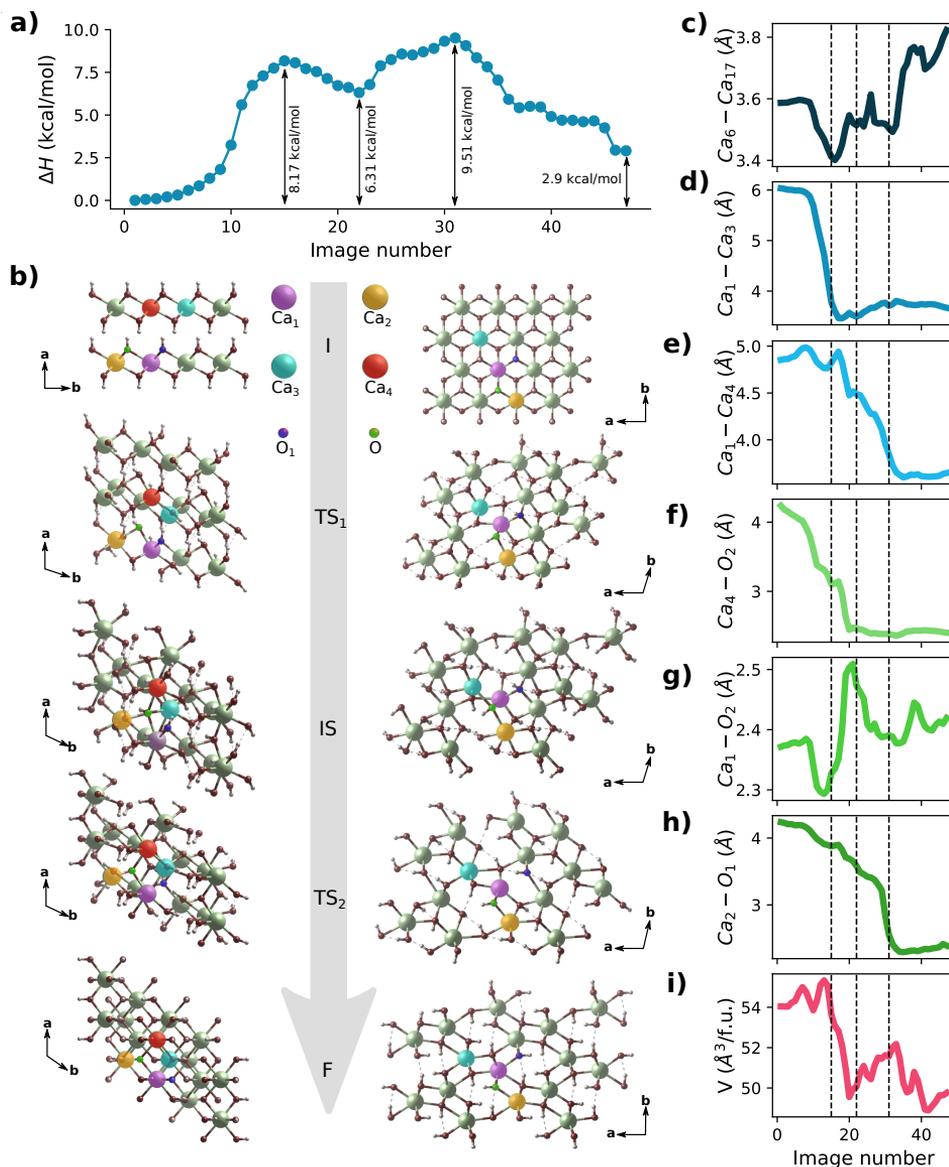
Figure 3.7: Interlayer transformation path from portlandite to Phase V. (a) Enthalpy difference along with the structural transformation for each image used in the VC-NEB method. (b) Graphical representation of the different stages along with the structural transformation. The calcium, oxygen and hydrogen atoms are illustrated with green, red and pink spheres respectively. [(c)-(i)] Respectively, the evolution of different structural properties during the transformation. The atoms are named according to the shiny colors present in (b).

TS$_3$. Another way to see how the different layers in portlandite are joined together is to observe the Ca$_4$−O$_2$ distance which corresponds to a Ca atom and an O atom in a different layer at the beginning of the transformation. This distance, shown in 3.7(f), is gradually decreased from the beginning of the transformation to the IS and then remains constant until the end. In summary, the transformation can be split into three parts. In the first part of the transformation, a section of the portlandite layer will join another portlandite layer until getting to the IS. The other part of the layer will join from the IS until the TS$_2$. Finally, the energy will decrease due to the rearrangement of the atoms to form the tobermorite layer in the perpendicular direction of the original portlandite layer.

4 | TRANSFORMATION IN M-S-H

Given that calcium and magnesium are isoelectronic, portlandite and brucite [Mg(OH)$_2$] share the same crystal structure. Hence, the possible stability of Phase V for MgOH$_2$ is also checked. The structure is broken during the energy minimization process, which indicates that the phase is not even metastable. The impossibility of Mg(OH)$_2$ to adopt the crystalline arrangement of Phase V polymorph could be the reason why the M-S-H gels have a completely different structure from the tobermorite-like C-S-H gel [162].

5 | CONCLUSIONS

The structure of Sr(OH)$_2$ contains locally a Sr-O arrangement very similar to the Ca-O arrangement in tobermorite. Therefore, since Ca and Sr are isoelectronic, the stability of the Sr(OH)$_2$ structure replacing the Sr atoms with Ca atoms was studied. This new phase, called Phase V, was found to be stable with an energy difference of 2.78 kcal/mol to respect to portlandite.

The transformation path from portlandite to Phase V was studied using VC-NEB. Two low-energy transformation paths were found: an intralayer one and an interlayer one. The intralayer transformation presents a direct transformation with an enthalpy barrier of 10.13 kcal/mol. The interlayer transformation was found by using a Ca(OH)$_2$ polymorphs as

an intermediate state for the transformation. Therefore, the interlayer transformation presents an IS with two TS, TS_1 and TS_2. The enthalpy barrier for this transformation is 9.51 kcal/mol, thus lower than the intralayer transformation.

Although the interlayer transformation is lower than the intralayer one, the intralayer transformation is more interesting since the portlandite layer and the tobermorite layer present the same direction. This means that a similar transformation could occur in a portlandite monolayer, which is very promising to study the formation of C-S-H from a portlandite monolayer.

CHAPTER 4

C-S-H MONOLAYER FORMATION FROM A PORTLANDITE MONOLAYER

Re:Stacks
Americana
Gregoire Manet - Romain Collin - Bill Frisell

As mentioned in the introduction, the nucleation of the C-S-H gel is crucial in the early stages of cement hydration. Using X-ray scattering together with cryo-TEM Krautwurst et al. [75] have claimed that the nucleation of C-S-H is governed by non-classical nucleation containing at least two main steps. The first step would involve the formation of amorphous spheroids from dense liquid regions, while the second step implies the crystallization of C-S-H from the spheroids. While these experimental studies give very relevant microscopic information about the C-S-H nucleation, the atomic-scale mechanism is still unknown.

In the previous chapter, it has been proven that a phase transition between portlandite and Phase V is feasible. This is very interesting since the Phase V layers are very similar to the Ca-O layers present in C-S-H. Therefore, on this basis, here we propose and study a hypothesis of C-S-H crystallization where portlandite monolayers suffer a topochemical transformation to a tobermorite layer.

In this chapter first, the hypothesis is presented. Then, the proposed formation process is studied in three steps. In the first one, the interaction of a single silicate dimer with a portlandite monolayer is studied using a nudge elastic band (NEB). This would provide insight into
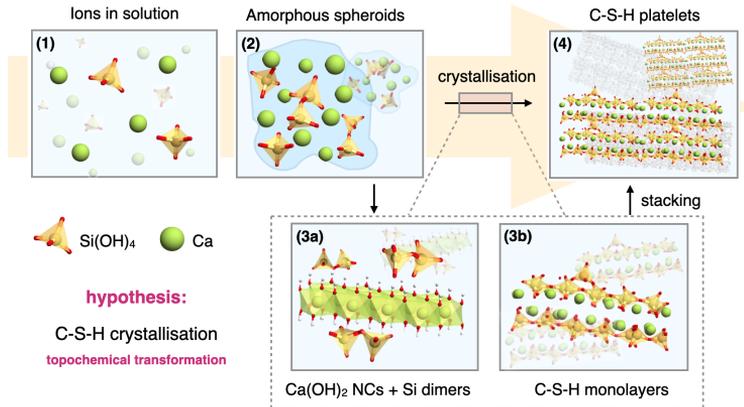
Figure 4.1: Proposed mechanism for the early stage formation of C-S-H. Within this model, the ions in solution would create highly dense amorphous spheroids which then would form portlandite monolayers surrounded by silicate dimers. Finally, these monolayers would suffer a topochemical phase transformation into a C-S-H monolayer which then would create the final C-S-H by stacking them.

the condensation reaction necessary for the transformation and would quantify the energy barrier for these reactions. Then, enhanced sampling techniques are used to sample the whole transformation path from a portlandite monolayer to a C-S-H monolayer. On the one hand, targeted molecular dynamics (TMD) is used to look for different transformation paths. On the other hand, umbrella sampling (US) is used with a previously found path to quantify the free energy barrier for this transformation.

1 | THE HYPOTHESIS

As suggested by Krautwurst et al. [75] the first step in the C-S-H nucleation is the formation of a dense region. It is reasonable to think that this dense region would contain the main species in dissolution, complexes and pre-nucleation clusters formed by these species [Fig. 4.1(2)]. The portlandite monolayers are very stable [144] and therefore, it is very likely that portlandite monolayers are formed during the early stages of C-S-H. Moreover, it is also known that silicate dimers are formed at the beginning of cement hydration.

Therefore, in our hypothesis, those dense regions presented by Krautwurst *et al.* could contain portlandite monolayers and silicate dimers in sub-

stantial quantities. From the previous chapter, we know that portlandite could transform into Phase V by an intralayer transformation. This transformation could also occur in a portlandite monolayer. Therefore, the portlandite monolayer could suffer a similar transformation due to the interaction with the surrounded silicate dimers.

The silicate dimers will approach the portlandite monolayer so that a condensation reaction would occur, releasing water molecules and joining together silicate dimers with the portlandite monolayer. These condensation reactions would provoke the transformation in the portlandite monolayer going from a flat Ca-Ca arrangement to the known zig-zag arrangement found in tobermorite [Fig 4.1(3a)]. That is, the dimers would attach to the portlandite layer, thus creating a tobermorite layer [Fig. 4.1(3b)]. Finally, as shown in Fig. 4.1(4) these tobermorite layers would form the C-S-H gel by stacking them.

2 | Silicate reaction at portlandite surface

The energy barrier of the condensation reaction was studied using nudged elastic band (NEB), taking 47 images between the initial (I) and final (F) stages of the transformation [see Fig. 4.2(b)]. The initial stage consists on a portlandite monolayer and a silicate dimer placed at 5 Å from the surface. In the final stage, the silicate dimer is bonded to the portlandite layer with a Si-O-Ca bond, and a water molecule has been released in the process. The conformations at the initial and final stages were minimized using DFT, and intermediate images were created by interpolation of the atomic coordinates. Only one constraint was imposed on the basis of our experience: As the covalent Si-O bond is generally stronger than the ionic Ca-O bond, the released water was formed by an oxygen atom belonging to the portlandite layer. This assumption was confirmed a posteriori in the MD simulations from Section 3.1, where we verified that in $99\,\%$ of the reactions, water molecules were formed by oxygen atoms from portlandite.

The minimized transition path presents different stages. In Fig. 4.2(a) we show the energy profile, together with bond distances of interest, to characterize the reaction in Fig. 4.2(c-f). The calculated different bond distances are shown in Fig. **??**. First, there is a reorientation of
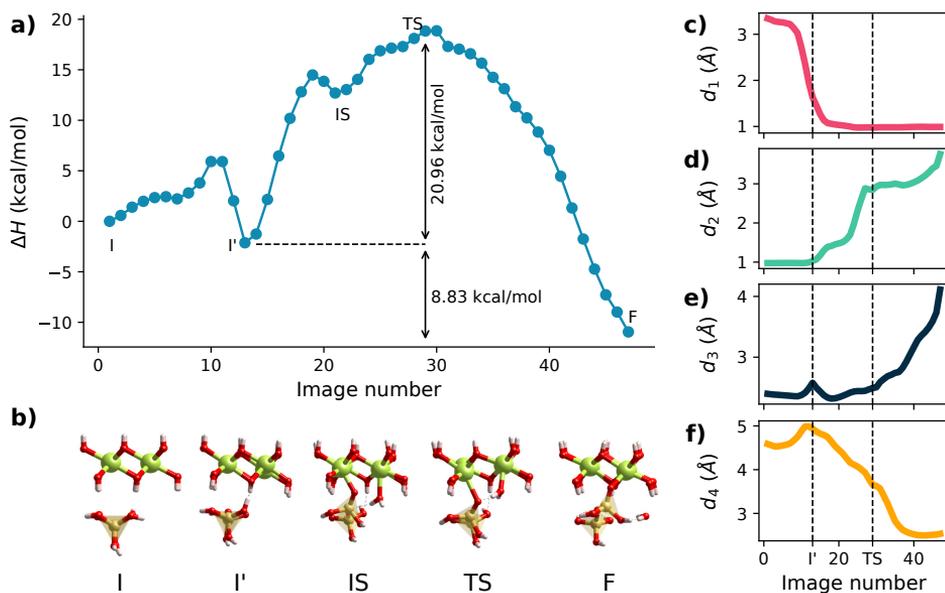
Figure 4.2: Reaction path of the condensation of water in a Portlandite monolayer. (a) Enthalpy difference along the reaction path for each image used in the NEB method. (b) Graphical representation of the different stages along the condensation reaction. The calcium, silicon, oxygen and hydrogen atoms are illustrated with green, orange, red and pink spheres respectively. [(c)-(f)] d_1 represents the distance between the hydrogen atom in the silanol group and the oxygen atoms in the layer, d_2 the distance between the oxygen and hydrogen atoms of the silanol group, d_3 the distance between the oxygen in the layer and the calcium atom and d_4 the distance between the oxygen atom in the silanol group and the calcium atom. All these distances are plotted along the reaction path.

the silanol and hydroxyl groups involved in the condensation reaction, forming a hydrogen bond in which the silanol is the donor and portlandite the acceptor. This can be noticed in a sharp decrease of d_1 in Fig. 4.2(c). The reorientation enthalpy barrier is 5.92 kcal/mol, and the resulting conformation I′ is more favorable than our initial guess I ($\Delta H = -2.13$ kcal/mol). From that stable conformation I′, there is a proton transfer reaction from the silanol to the hydroxyl group to form a water molecule in the intermediate step (IS) [see Fig. 4.2(a)]. This can be observed in the increase of the distance between the oxygen and hydrogen atom of the silanol group d_2 in Fig. 4.2(d), which breaks the O-H bond. The associated activation enthalpy of this concrete process (I′ to IS) is 16.61 kcal/mol. In the transition state (TS) in Fig. 4.2(a) both the silicate and the water molecule are bonded to adjacent Ca atoms from portlandite. Finally, the water molecule is released and the silicate oxygen is bonded to the layer, completing the transition. This is characterized

by the increase in the distance between the acceptor oxygen and the Ca atoms d_3 in Fig. 4.2(e) and the distance between the donor oxygen and the Ca atom d_4 in Fig. 4.2(f). The energy barrier from stage I′ to the TS is 20.96 kcal/mol and the enthalpy gain from stage I′ to the final stage is 8.83 kcal/mol. It is important to notice that the condensation reaction does not alter the portlandite conformation. That is to say, the layer remains almost flat. Further reactions are needed to trigger the transformation, as we will see in the following subsection. The energy barrier is twice as high as that of the bulk portlandite transformation, and hence, the condensation reaction is likely the limiting process.

3 | SAMPLING THE CONFIGURATIONAL SPACE BY TARGETED MOLECULAR DYNAMICS

Now that the mechanism of the condensation reaction between a silica dimer and a portlandite layer is understood, the whole process from a portlandite layer to a tobermorite layer can be sampled using molecular dynamics (MD). To get a tobermorite layer from a portlandite layer all the O-H groups in the layer should be released as water molecules due to the condensation reactions with the silica dimers. Therefore, this transformation is based on the concatenation of condensation reactions. Each of the condensation reactions can be considered a rare event and therefore, the whole process is computationally very hard to replicate in the simulation time accessible for a system of this size. Using the Eyring-Polnayi equation we can estimate the condensation reaction rate k:

$$k = \kappa \frac{k_b T}{h} e^{\frac{\Delta S^{\ddagger}}{R}} e^{-\frac{\Delta H^{\ddagger}}{RT}} \tag{4.1}$$

where κ is the transmission coefficient, k_b is the Boltzmann constant, h is the Planck's constant, $\Delta S^{\ddagger}$ is the activation entropy, $\Delta H^{\ddagger}$ the activation entropy and R the gas constant. Assuming that the entropy remains constant ($\Delta S^{\ddagger}$=0) and the transmission coefficient is $\kappa = 1$, and using the previous result of the activation enthalpy ($\Delta H^{\ddagger} = 20.96$ kcal/mol) at $T = 300$ we can estimate that the reaction rate is $k = 0.003 \ s^{-1}$. That is to say, each reaction will occur every 5+ minutes. Nevertheless,

accessible simulation times are about a few nanoseconds. That is why a sampling technique such as TMD is necessary to speed up the condensation reaction rate and make it comparable to the accessible simulation time.
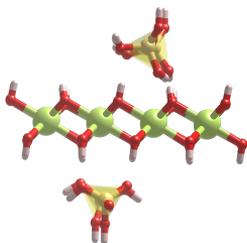
Figure 4.4: Initial configuration of the system for the TMD calculations.

As explained before in chapter 3, TMD uses a time-dependent geometrical constraint to pull the system from the initial configuration towards the target structure [113, 114]. In this case, the initial configuration would be a portlandite layer surrounded by silica dimers and the target configuration would be the tobermorite layer. This method allows sampling the conformational space around a collective variable (a generic reaction coordinate) during the process [116, 163]. To perform these simulations a portlandite monolayer (with periodic boundary conditions) was placed in a simulation box, with silicate dimmers at 3.3 Å above and below it, as shown in Fig. 4.4. This system does not contain the bridging silicate present in tobermorite but it has been proven that at the early stages of cement hydration dimers are the most common silica species and longer chains are very rare [164, 165]. Moreover, the bridging silicate would not participate in the condensation reactions needed for the transformation so they are not needed to study the transformation.

The target position of every atom could be included in the pulling coordinate for TMD. Nevertheless, the target positions of some atoms depend on the transformation path. For example, it is hard to know where the released water molecules would be at the end of the transformation and including them in the pulling coordinate would be too restrictive. Moreover, even if the target positions of Ca and Si atoms in the tobermorite layer are known, we have decided to only include the target positions of the silicon atoms, thus giving full freedom to the rest of the atoms. That is to say, the transformation from $Ca(OH)_2$ to CaO is not predefined nor are the condensation reactions explained in section 2. Therefore, the pulling coordinate

$$\rho = \sqrt{\sum_i (\vec{x}_i - \vec{x}_i^B)^2} \qquad (4.2)$$

was defined as the root mean square distance of the silicon atoms to the targeted position, which was set as the position of the dimers in a tober-
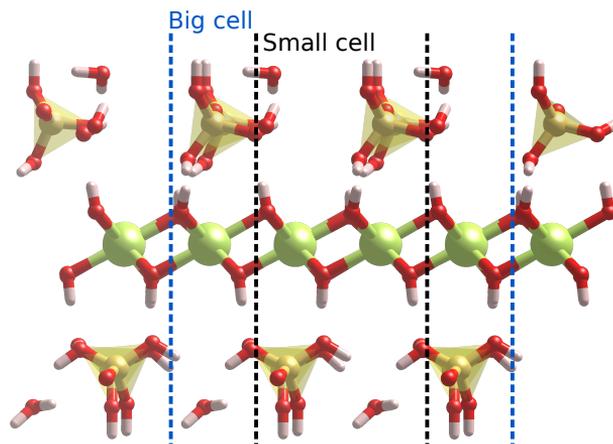
Figure 4.5: Graphical description of the two types of unit cells used for the simulations. The small cell contains a portlandite layer with 4 formula units and 2 dimers. The small cell contains a portlandite layer with 16 formula units and 8 dimers.

morite monolayer. The silicon atoms were dragged over and restrained to the collective coordinate ρ. In this way, the average distance of the silicon atoms is controlled but at the same time, they are able to move freely. In this manner, the system is pulled from $\rho = 3.3$ to $\rho = 0.0$

Due to the difficulty of sampling this process two systems with different sizes have been used. First, a set of small cell calculations containing a portlandite layer of 4 formula units surrounded by 2 silicate dimers have been performed to study the explore the configurational space of the transformation in a reasonable simulation time. Then, to have a better sampling another set of simulations have been done using a 4 times bigger cell with a portlandite layer of 16 formula units surrounded by 8 silicate dimers.

3.1 | Small cell calculations

To study the underlying mechanism of the portlandite to tobermorite layer transformation, the first set of TMD simulations were done with the small system containing a portlandite layer of 4 formula units surrounded by 2 silicate dimers.

3.1.1 | SIMULATION DETAILS

100 trajectories were performed to have enough statistics of the transformation paths and the ReaxFF force field was used for the interaction between atoms. The reaxFF force-field has been used before for this kind of systems [166, 167] since it uses a length-bond order scheme to describe atomic interactions allowing chemical bonds to break and form. This property of the force field is essential to reproduce the condensation reactions needed for the transformation under study.

As the simulations were done in the NVT ensemble there is a cell mismatch between tobermorite and portlandite. The cells differ in their γ angle which is 120° for portlandite and 123° for tobermorite. Therefore, the portlandite unit cell has been modified from $\gamma = 120°$ to $\gamma = 123°$ which barely changes the crystal structure. Another difference is that portlandite unit cell is slightly bigger than that of tobermorite. In the unit cell with 4 formula units of $Ca(OH)_2$ the cell parameters are $a = 7.35$ Å and $b = 7.44$ Å for portlandite and $a = 6.61$ Å and 7.31 Å. To overcome this difference, these cell parameters were modified linearly during the simulations. This change in the unit cell parameters would have an impact on the free energy of the system since an external force is being included but will not change the underlying transformation mechanism. . However, at the moment we are mostly interested in the internal energy.

Another critical point of the TMD calculation is the pulling speed along the coordinate ρ. In principle, any velocity is valid for sampling the configuration space. Faster speeds require more simulations to represent the configuration space correctly. However, in the system under consideration, the condensation reactions limit the pulling speed because if this speed is too fast there will not be time for the reactions to occur. That is why this system is very complicated to sample. On the one hand, we need to accelerate the process with a sufficiently high pulling speed for computational reasons, but on the other hand, it cannot
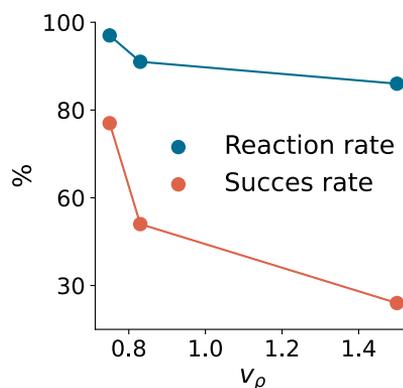
Figure 4.6: Reaction rate and success rate with different pulling velocities.

be too high because otherwise, the reactions will not occur.

To determine a correct pulling velocity first we tried doing 1 ns simulations with a pulling velocity of $v_\rho = 1.5$ Å/ns. This simulation had an 86% of reaction rate but only a few simulations ended up with a full transformation. in other words, the transformation success rate was 36 %. Therefore, we tried another set of 2 ns simulations with a pulling velocity of $v_\rho = 0.75$ Å/ns. These calculations presented a reaction rate of 97% which means that almost all the condensation reactions occurred. The transformation success rate was 77% which was more than enough to have good statistics about the transformation. Slower pulling velocities could have been tried but the simulation time was a limiting factor (see Fig. 4.6). Therefore, after having to change a bit the initial distance between the portlandite layer and the dimers for technical reasons, simulations were performed taking the system from ρ=3.3 Å to ρ=0.1 Å at a pulling speed of $v_\rho = 0.83$ Å/ns. This pulling speed gave a 91% rate of condensation reactions in an affordable simulation time.

> **Simulation parameters:** 100 trajectories were performed with the LAMMPS [168] code. The temperature was set to 500K to accelerate the condensation reaction rate and the simulations were performed in the NVT ensemble using the Nosé-Hoover thermostat [169].The timestep was set to 0.1 fs. The interactions between atoms were modeled with the ReaxFF force-field [170, 171, 172] with the Ca/Si/O/H set of parameters from ref. [172, 173].

3.1.2 | ACCURACY OF REAXFF

To check the accuracy of ReaxFF with respect to ab initio methods such as DFT, we have performed a benchmark by minimizing 6 different structures along the portlandite to C-S-H transformation with ρ =3.3, 2.0, 1.5, 1.0, 0.5, 0.0. The energy differences were computed using the starting point of $\rho = 3.3$ as a reference.

The difference in energy with respect to the portlandite monolayer is almost always underestimated by ReaxFF, with a maximum error of 3.7 kcal/mol. However, the energy difference in the tobermorite monolayer ($\rho = 0.0$) is overestimated by ReaxFF.
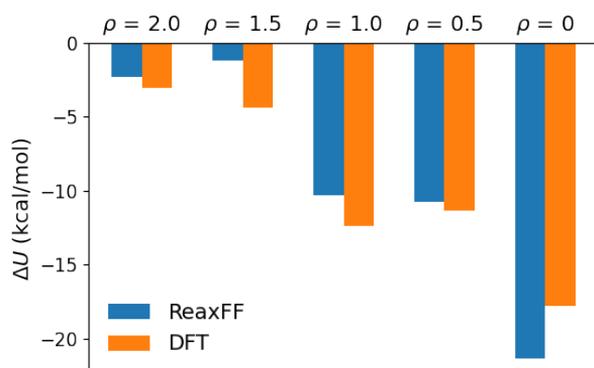
Figure 4.7: Energy differences of different stages during the portlandite to C-S-H transformation respect to the initial state $\rho = 0.0$

3.1.3 | RESULTS

The internal energy of every snapshot in all the 100 trajectories was collected and analyzed. It is shown in Fig. 4.8(a) as a function of the collective variable ρ. The color of each point indicates the percentage of condensation reactions that took place with respect to a fully reacted system. That is to say, in this system, 8 condensation reactions need to take place to have a fully reacted system. From $\rho = 3.3$ Å to $\rho = 1.65$ Å no reactions were observed as the silicate dimmers approach the Ca(OH)$_2$ monolayer. This is because the silicates are not close enough to the portlandite layer for condensation reactions to take place. From here, condensation reactions start to take place following the mechanism described in Sec. 2. The distance ρ at which the first condensation reactions take place is relatively homogeneous across simulations ($\rho \approx 1.65$). Subsequent reactions take place at a different rate, and two subsets of trajectories can be identified: those with a completed transformation and those with an uncompleted transformation. The complete transformation is defined as a transformation where the 8 condensation reactions (100%) needed for the transformation have occurred. The completion of the transition depends strongly on the percentage of chemical reactions the system has gone through. When some hydroxyl groups in the portlandite layer remain unreacted, the structure evolves to a final defective tobermorite-like monolayer with interstitial hydrogen atoms on it. These defects have important energetic consequences. In figure 4.8(a) the solid lines represent the average internal energy of all the completed
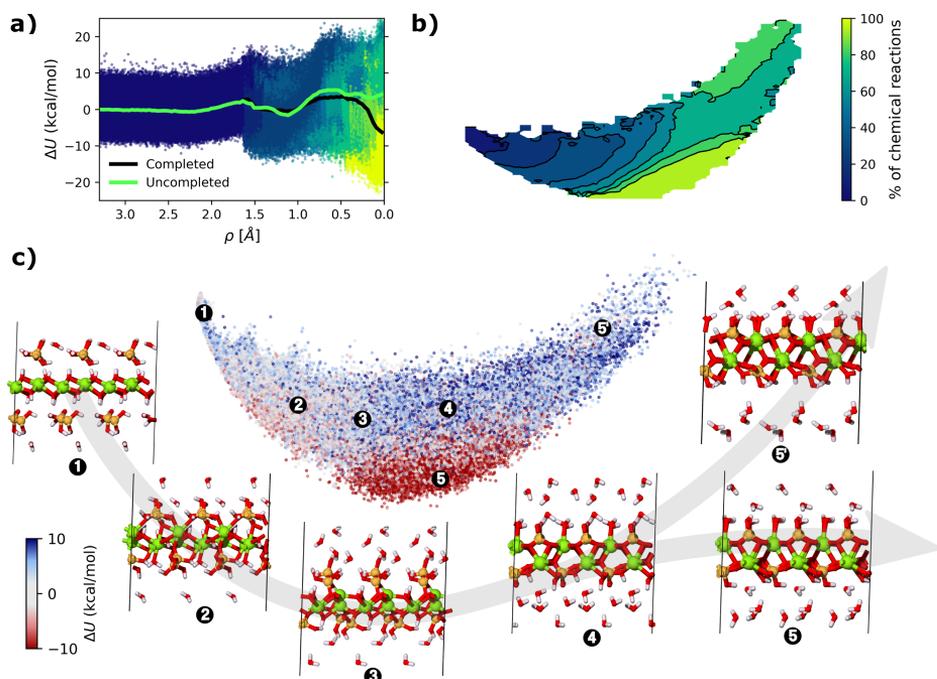
Figure 4.8: Topochemically induced transformation from a portlandite monolayer to
a C-S-H layer computed with TMD. (a) Internal energy difference of the 100 trajectories
along the pulling coordinate. The dots represent different snapshots of the trajectories
and the solid lines represent the mean value of the internal energy for the completed
(black) and uncompleted (green) trajectories. The color scheme represents the per-
centage of condensation reactions needed for the complete process. (b) The average
percentage of the condensation reactions along with the first two components of the
KPCA. The distance in the plot represents the structural distance calculated with SOAP.
The portlandite monolayer is located in the areas where 0 % of the reactions have oc-
curred and the C-S-H monolayer is located in the areas where 100 % of the reactions
have occurred. (c) Internal energy difference of 100.000 snapshots of the trajectories
along with the two first components of the KPCA. Each dot represents a snapshot and
the distance between dots represents the structural distance calculated with SOAP. The
color scheme represents the internal energy difference of each snapshot.

and uncompleted trajectories as a function of ρ. The black color was used for fully reacted systems, and green for uncompleted systems. Up to $\rho = 1.65$ Å, before the chemical reactions start, both sets are equivalent. Then, for $\rho < 1.65$ Å, when the reactions start, the mean internal energies start to deviate. However, it is not until $\rho < 0.25$ Å that there is a significant difference between both sets. The completed trajectories experience a final energy decrease of -6.24 kcal/mol, making them more stable. The uncompleted trajectories end up with a 4.19 kcal/mol increase in energy. It is important to notice that the drop in energy in the completed trajectories only happens when the percentage of condensation reactions is close to 100 %. Hence, the system only stabilizes when the process is totally completed.

To elucidate whether the two sets of trajectories are due to divergent paths or the existence of two final (meta)stable structures, we did a structural distance analysis. We used the smooth overlap of atomic positions descriptor (SOAP) [133, 174] to compute structural distances and the kernel principal component analysis (KPCA) to reduce the dimensionality of the computed distance matrix [175], plotting it in 2D. In the resulting plots [Fig. 4.8(b) and Fig. 4.8(c)] the distance between points represents their structural similarity. In addition, the color scale in Fig. 4.8(b) indicates the percentage of occurred condensation reactions, while the color scheme in Fig 4.8(c) represents the internal energy difference with respect to the starting configuration. The structural distance plotted in Fig. 4.8(b) clearly shows that there is mostly one single path until 60% of reactions occur. However, at the end of the simulations, two limiting regions can be found where a 100% of the chemical reactions have occurred. The upper right area corresponds to uncompleted trajectories, while the bottom area in the plot corresponds to completed trajectories where all the reactions have occurred. Since we have already stated that the transition evolves along with the percentage of chemical reactions, it is clear now that the completed and uncompleted trajectories end up in structurally slightly different regions.

This is illustrated in Fig. 4.8(c), where the internal energy difference and snapshots of some obtained structures during the transformation are shown. The snapshots represent the evolution towards the complete (number 5) and incomplete (number 5') transformation. It can be seen in snapshots 1-3 that for the first 30% of the reactions, the layer is not bent. Until then, both the complete and incomplete trajectories follow the same trend. It is not until 50-60% of the reactions have occurred (snapshot 4) that the layer starts to take the zig-zagged form of tober-
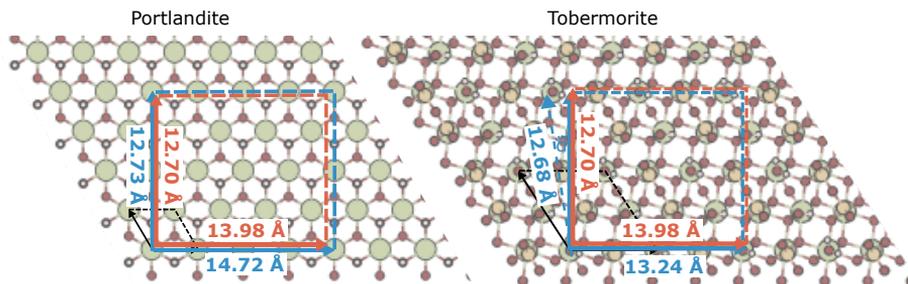
Figure 4.9: Redefinition and transformation of the unit cells of a portlandite and to-
bermorite layer to get a common unit cell. Black arrows represent the original unit cell,
blue arrows the orthogonal unit cell and orange arrows the unified unit cell for both
structures.

morite. Finally, both the complete and incomplete trajectories (snapshots
5 and $5'$ respectively) reach the zig-zag structure of tobermorite. How-
ever, Some hydroxyl groups attached to the layer are still found in the fi-
nal state of the uncompleted trajectories, which is unfavorable. In terms
of energy, it can be observed again that the final state of the completed
trajectory has much lower energy than the final state of the uncompleted
trajectory.

3.2 | BIG CELL CALCULATION

In order to have more realistic results, more trajectories were computed
with a bigger cell containing a portlandite layer of 16 Ca atoms and 8
silicate dimers. The use of bigger unit cells reduces the unwanted corre-
lation between the same atoms of different cells.

3.2.1 | SIMULATION DETAILS

These calculations were done with similar simulation details to that of
sec. 3.1.1. However, since this system is more realistic, it will be valuable
to calculate the free energy difference of the layer transformation using
the Jarzynsky equality [176, 116] along the pulling coordinate ρ. For this
purpose, the approach used in section 2 of changing the unit cell during
the simulation cannot be used in this case. This approach would produce
a contribution to the free energy thus distorting the final result.

Therefore, the mentioned mismatch between the unit cells was over-

	Original Supercell	Transformed Supercell	Difference
Portlandite			
a	14.72 AA	13.98 AA	-5.03%
b	12.73 AA	12.70 AA	+0.24%
α	90	90	0 %
Tobermorite			
a	13.24 AA	13.98 AA	+5.59%
b	12.68 AA	12.70 AA	+0.16%
α	96	90	-6.25%

come by transforming both unit cells into orthogonal unit cells and choosing an intermediate dimension for the calculations. As shown in Fig. 4.9, the portlandite layer can be redefined orthogonal. However, the tobermorite unit cell cannot be redefined completely orthonormal. To overcome this, the unit cell was transformed by the small-angle $\alpha = 6°$. This transformation hardly changes the structure of the tobermorite layer. Then an intermediate cell between the two orthogonal cells was chosen for the TMD calculation. In any manner, the chosen unit cell has a maximum error of 5%.

Due to the bigger size system and limited accessible computational time, the pulling speed has to be increased. In this case, 50 trajectories were done in two steps. First, the system was sampled from $\rho = 2.0$ Å to $\rho = 4.0$ Å at a higher pulling rate of $v_\rho = 2$ m/s. In this part of the trajectories, such a high speed can be used because there is no chemical reaction. Then, the system was sampled with a pulling rate of $v_\rho = 1$ Å/ns from $\rho = 2.0$ Å to $\rho = 0.1$ Å.

3.2.2 | RESULTS

Due to the big size of the system, 36 condensation reactions must occur for the transformation to be completed. The obtained trajectories were unsatisfying due to the low condensation rate. That is, due to the big size of the system and due to the speed limit in the pulling coordinate, the system was not able to react completely in the accessible simulation time. Therefore, from the 50 trajectories computed, only two reached the complete transformation from portlandite to tobermorite. Nevertheless, the free energy was computed using Jarzynsky's identity even knowing that the result will not be entirely correct.

In Fig. 4.10 the potential of mean force along the pulling coordinate ρ is shown with a free energy increase of 40 kcal/mol. As it has been
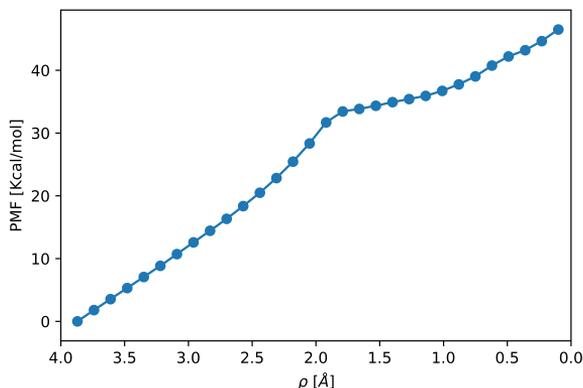
Figure 4.10: Potential of mean force (free energy) computed using Jarzynsky's identity of 50 TMD trajectories.

mentioned before, this result is not accurate since most of the trajectories have not completed the transition. However, it is reasonable to believe that a set of completed trajectories will have a much lower free energy increase or even a decrease since the final arrangement of the complete trajectories are more stable from the internal energy perspective. In any case, this result marks a ceiling of free energy difference and it seems clear that this transformation would have a positive energy difference, i.e., the transformation would not be spontaneous in thermodynamic terms.

Although large cell TMD calculations have not been very useful for free energy calculation, two trajectories present a complete transformation, which can be used as an initial trajectory for other methods of calculating free energy such as umbrella sampling (US). For that reason, in the following section, one of these trajectories has been used and sampled in a more accurate way to obtain a more realistic free energy difference.

4 | CALCULATING THE FREE ENERGY OF THE TRANSFORMATION WITH UMBRELLA SAMPLING

A different way to calculate free energy is to use umbrella sampling (US). US, unlike TMD, uses an initial path to sample the configurations around that path to calculate free energy. In this method, the trajectory is divided into windows along the reaction coordinate. In other words, some snapshots are taken from the initial trajectory and then independent simulations are done for each window starting from each of the snapshots. To maintain the system close to the starting snapshot, the simulations are done using a biased potential restraining the trajectories in a window around the initial snapshot [177, 178]. The free energy is then reconstructed using the weighted histogram analysis method (WHAM) [118].

4.1 | SIMULATION DETAILS

An initial trajectory was taken from the TMD big cell calculations. To perform the calculations more realistically, the simulations have been done on the NPT ensemble, thus allowing the simulation cell to move and avoiding the problem of the unit cell mismatch between portlandite and tobermorite. Moreover, since the initial trajectory already contained the condensation reactions needed for the transformation, the temperature can be now lowered to 300K. To get more freedom to the system, the reaction coordinate ρ has been altered so that it only takes into account the Z coordinate of the atoms. That is to say, ρ now represents the root mean square distance of the height difference to the final tobermorite layer.

In this calculation, 40 equidistant points along the reaction coordinate were taken from the original transformation path calculated with TMD. To have enough data in each window to correctly sample the whole transformation path different spring constants were used starting from 1 kcal mol^{-1}/Å and going up to 1000 kcal mol^{-1}/Å. The form of the gaussian probabilities in each window together with the minimum overlap of 20% between windows was the key to deciding the value of the

spring constant. A spring constant of 50 kcal mol^{-1}/Å was found to be
the most appropriate since there was an overlap of about 30% between
the probability distribution of two trajectories centered in consecutive
points in the reaction coordinate and the probability in each window is
almost perfectly gaussian. Moreover, to fill the gaps where the overlap
between the probability distributions of two consecutive trajectories was
less than 30% 12 more trajectories were computed centered in strategi-
cally chosen points of the reaction coordinate. Later is shown in Fig.
4.13 that reducing the number of windows (which drastically reduces
the overlap between the probability distribution of the consecutive win-
dows) does not alter significantly the reconstructed free energy. To set
the simulation time for each window, different simulation times ranging
from 10 ps to 500 ps were tried out, and the reconstructed free energies
were compared. As seen in Fig. 4.12 the energy converged completely
for 500ps. With the complete set of 52 trajectories, the reaction path was
completely sampled, as shown by the probability distribution in Fig. 4.11.
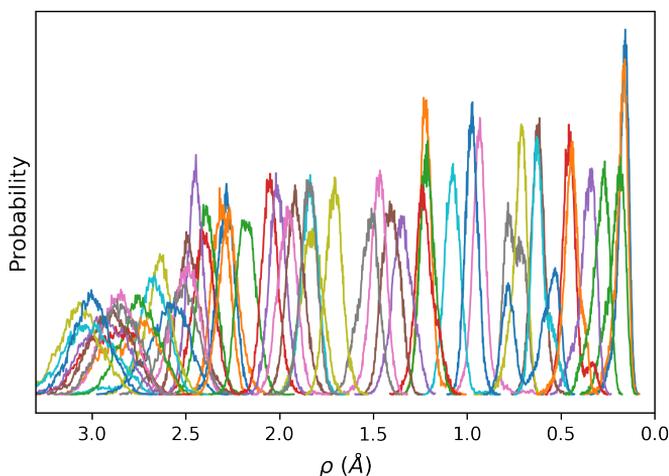
Figure 4.11: Probability distribution in each window of the US calculation.

To reconstruct the free energy profile along the pulling coordinate ρ
the weighted histogram analysis method (WHAM) was used where it
uses a weighted average of the distribution of the individual windows to
compute the global distribution [118].
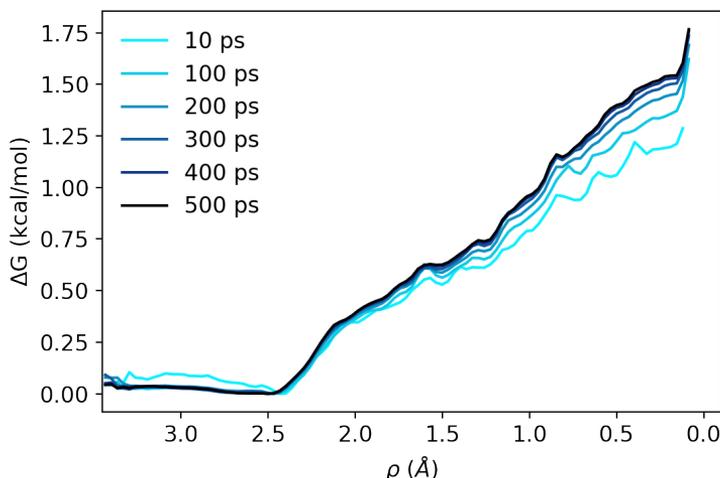
Figure 4.12: Convergence of the free energy computed with US varying the length of each simulation from 10 ps to 500 ps.

4.2 | CONVERGENCE OF THE FREE ENERGY

The calculation of the free energy was done using the Weighted Histogram Analysis Method (WHAM). The trajectories must be sufficiently long to have an accurate result of the free energy so that the free energy is totally converged. In Fig. 4.12 the computation of the free energy change along the reaction coordinate is shown using trajectories of different lengths. The free energy is sufficiently converged using trajectories of 500 ps with a maximum error with the 400 ps trajectories of 0.018 kcal/mol. The convergence of the number of windows was also checked by reducing the number of windows to half. In Fig. 4.13 the difference in free energy is shown with all the windows and with half of the windows. Apart from some fluctuations due to the lower overlap between windows, both free energies look very similar.

4.3 | RESULTS

The results of the umbrella sampling calculations were shown in Figure 4.15. The internal energy difference between the portlandite layer surrounded by silicate dimers ($\rho = 3.3$ Å) and the tobermorite layer ($\rho = 0.0$ Å) is 15 kcal/mol (see Fig. 4.15. This value is larger than that
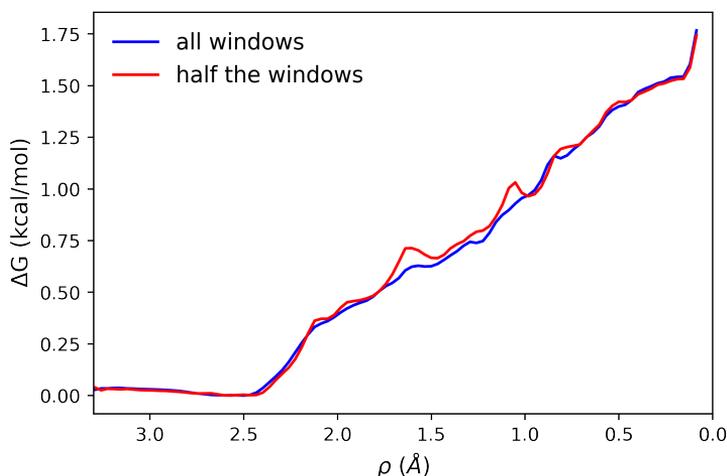
Figure 4.13: Free energy computed with US with all the windows and removing half of the windows. It can be seen that decreasing the number of windows intensifies the fluctuations but does not change its general behavior.

from TMD due to the additional volumetric relaxation. As a matter of fact, the area of the layer decreases by 12.6 % as it reacts. However, this decrease only starts when at least 20% of the chemical reactions have ocurred [see Fig. 4.14(c)]. Then it follows this trend almost constantly until the end of the transition. The compression is not homogeneous since it is mostly found in the direction perpendicular to the silicate dimers, which exhibits an 11.1 % decrease [see Fig. 4.14(a)]. In the other direction, parallel to the silicate dimers, the cell is barely compressed probably due to the difficulty to compress the dimers. The difference in this direction is only around 2%.

In addition to the internal energy, the sampling made with US allowed us to compute the main thermodynamical properties of the process, i.e., the enthalpy, the free energy and the entropy. As seen in the previous section, the internal energy drop starts when most (more than 80%) of the chemical reactions have taken place. As expected from an NPT simulation, the enthalpy and the internal energy are mostly the same. However, the reaction free energy increases monotonically for the entire transition as due to the entropic contribution ($T\Delta S$), see Fig 4.15(b). This increase has also been observed in the previous TMD simulations but this time the free energy increase is much lower, making it much more accessible. However, it implies that the transformation is not spontaneous from a thermodynamical point of view. The free energy starts
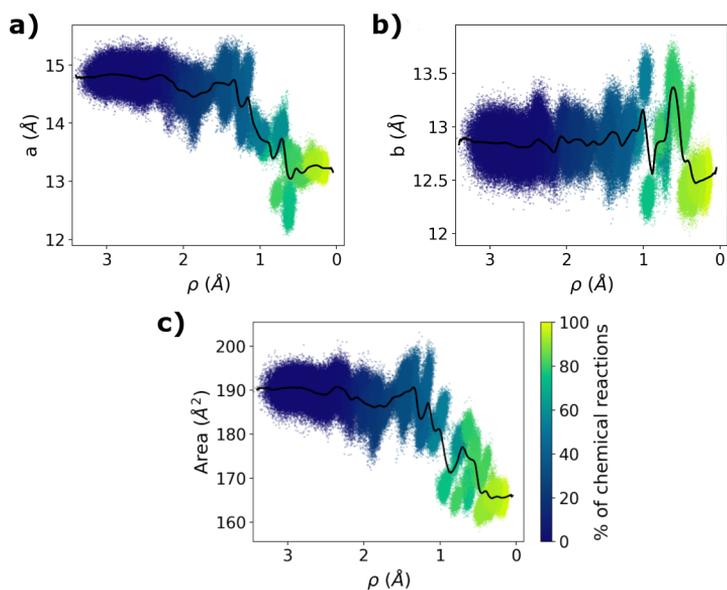
Figure 4.14: Change in the cell parameters a and b in terms of the pulling coordinate ρ. **c)** The area of the simulation box in terms of the pulling coordinate. Each dot represents a snapshot of the simulations and the color of the dot is the percentage of condensation reactions that had occurred until that snapshot. The black line represents the mean value.
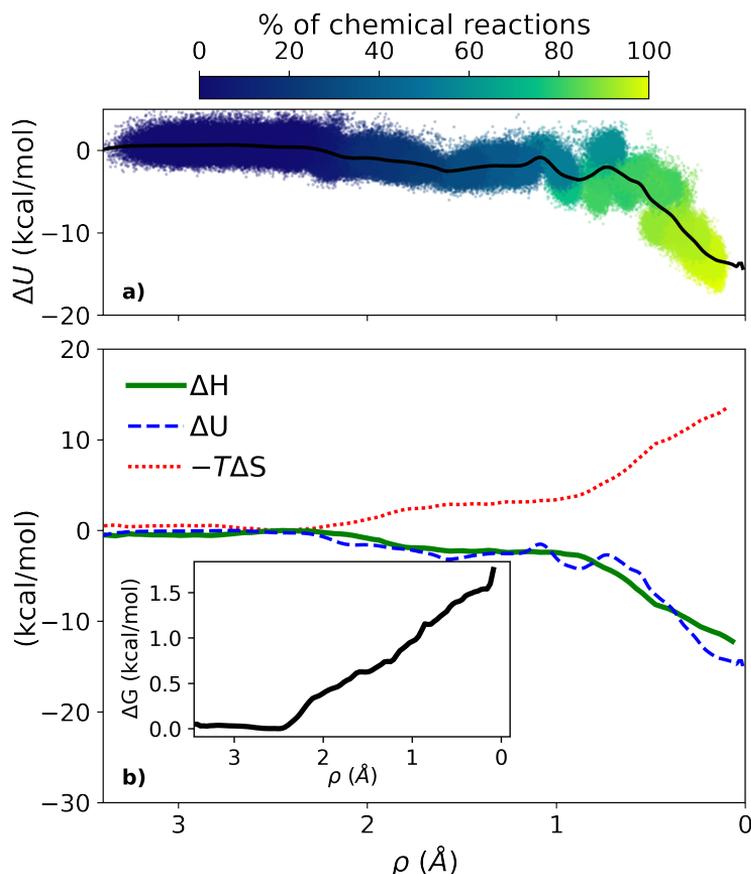
Figure 4.15: (a) Internal energy difference of the umbrella sampling trajectories along the pulling coordinate. The color scheme represents the percentage of condensation reactions needed for the complete process. (b) The thermodynamical properties of the process along the pulling coordinate.

to grow just when the first condensation reactions take place, and no energy barrier was found. The reason behind the increase in ΔG is a decrease of the system entropy, $\Delta G = \Delta H - T\Delta S$. We suggest that the condensation reactions impose a decrease in the translational and vibrational degrees of freedom, decreasing the entropy. This scenario should be usual for any nucleation and growth process, yet in the present case, the gain on enthalpy cannot compensate for such an entropy diminution.

According to our simulations, the free energy of the transformation from a portlandite monolayer to a tobermorite-like monolayer is $\Delta G_r^0 = 1.5$ kcal/mol. In the present case, the internal energy and the enthalpy are both negative and with a very similar magnitude (-14.95 and -12.21 kcal/mol), yet there is a decrease in the term $T\Delta S_r^0$ of 13.5 kcal/mol as-

sociated to the reduction of the translational and vibrational degrees of freedom during the C-S-H monolayer formation.

4.4 | COMPARING THE RESULTS WITH THERMO-DYNAMIC DATA

We can now compare the studied reaction with the expected results from thermodynamic data. The comparison can only be qualitative due to the different thermodynamic conditions. In the proposed transformation path the reactants are solid portlandite and fully protonated silicate dimers. From our calculations, we obtain that the free energy for two mols of $Ca(OH)_2$ is:

$$2Ca(OH)_{2(s)} + Si_2O(OH)_6 \longrightarrow Ca_2Si_2O_5(OH)_2 + 4H^+H_2O$$
$$\Delta G_1^0 = 3 \text{ kcal/mol.}$$
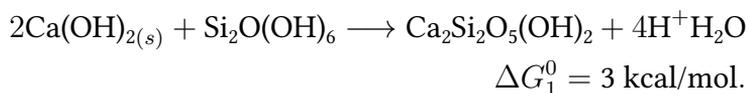
From the literature the standard free energy of reactions for portlandite [179, 180] and the silicate dimmers [181, 182] are:

$$Ca^{2+} + 2H_2O \longrightarrow Ca(OH)_{2(s)} + 2H^+$$
$$\Delta G_2^0 = 31.2 \text{ kcal/mol}$$

and

$$2Si(OH)_4 \longrightarrow Si_2O(OH)_6 + H_2O$$
$$\Delta G_3^0 = -3.5 \text{ kcal/mol.}$$
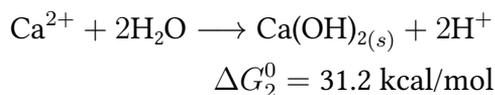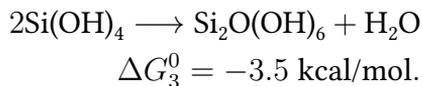
Therefore, constructing a Hess cycle the ΔG_f for the reaction from the monomeric species in solution to form the C-S-H monolayer with the intermediate step of portlandite formation and silica polymerization is:

$$2Ca^{2+} + 2Si(OH)_4 \longrightarrow Ca_2Si_2O_5(OH)_2 + H_2O + 4H^+$$
$$\Delta G_f^0 = 61.9 \text{ kcal/mol.}$$
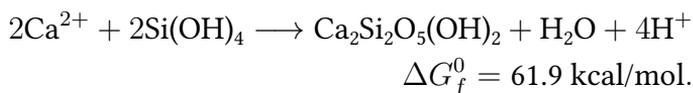
The result $\Delta G_r = 61.9$ kcal/mol is considerably higher, ≈ 20 kcal/mol, than the value calculated by Haas and Nonat for the same reaction to form a C-S-H with the same composition (C-S-H-β, $\Delta G_r^0 = 40.4$ kcal/mol) [183]. The discrepancy can be ascribed to computational aspects or to the model. Methodological aspects like a correct trajectory sampling and the accuracy of the ReaxFF energies have been evaluated, and do not represent a relevant source of error as discussed before in section 3.1.2. However, the thermodynamic conditions of the simulations might account for the difference. The proposed reaction takes place on a portlandite monolayer, far from the standard thermodynamic conditions that assume infinite crystals. In our system, we can expect that the contribution of the surface free energy to the total free energy will be important. Furthermore, the composition of a monolayer will differ from a bulky crystal of C-S-H where layers are stacked by interlaminar Ca atoms not present in our monolayer. Moreover, our simulations are done in the gas phase, i.e., water as a solvent is not included.

This work aimed to test a hypothetical mechanism for the C-S-H formation at the nanoscale. In this regard, we proved that the mechanism is feasible and that portlandite can react with silicate species in solution via condensation reactions to form C-S-H monolayers. The internal energy and enthalpy of the formation reactions indicate that it is a favorable transformation. Nevertheless, the exact thermodynamic conditions must be carefully evaluated to understand if this could be a feasible step in C-S-H formation, at least under certain conditions.

The results suggest that the reaction free energy of the proposed scenario for the formation of C-S-H is principally governed by the formation of portlandite nanoclusters, since the transformation from a portlandite monolayer to a C-S-H monolayer does not contribute substantially to the overall free energy of formation. As a matter of fact, at high pH both the formation of portlandite and C-S-H become negative, $\Delta G_f^0 = -214.4$ kcal/mol [184] and $\Delta G_f^0 = -855.54$ kcal/mol [185, 180] respectively, while the silicate dimerization free energy becomes positive [181], supporting the key role of portlandite formation in the process. Finally, we must also take into account that Ca and silicic acid form complexes in solution [186, 187] which could hinder the formation of nanoportlandite, leading to a different mechanism based on the aggregation of complexes.

5 | Conclusion

As stated in the introduction, the nucleation of the C-S-H gel plays a key role in the early stages of cement hydration, influencing the hydration rate and the cement microstructure development. In this chapter, we have proposed a model for C-S-H crystallization process where a portlandite monolayer would suffer a topochemically induced transformation into a tobermorite layer which then will lead to C-S-H by the stacking of them.

First, the condensation reaction necessary for the transformation has been studied in isolation. For that, the condensation reaction between a silicate dimer and a portlandite layer has been studied using CV-NEB. The reaction activation enthalpy is found to be 20.96 kcal/mol. This activation energy is twice as high as the transformation energy found in Chapter 3 from bulk portlandite to Phase V. Therefore, the condensation reaction may be the limiting process of the portlandite monolayer to tobermorite layer transformation.

With the condensation reaction activation enthalpy the reaction rate has been calculated and it was found that the condensation reaction would occur every 5 minutes. Since these long times are not accessible from a simulation point of view, enhanced sampling techniques have to be used to study the proposed transformation. Therefore, both TMD and umbrella sampling have been used to study the system with different system sizes. It was found that the internal energy remains almost constant until the last condensation reaction occurs and the internal energy decreases by -15 kcal/mol. Although the internal energy of the transformation is favorable, the free energy of the process is unfavorable with a free energy increase of 1.5 kcal/mol. From a qualitative point of view, this free energy value is in a reasonable range that makes the studied process a potential C-S-H formation pathway. The proposed formation pathway does not exclude other possible formation mechanisms that could complement our hypothesis and may exist alongside others.

CHAPTER 5

LOOKING FOR C-S-H PNCS

Slumber
Sun, Cloud
Luke Howard

U ntil here, a specific formation mechanism has been studied in which portlandite monolayers transform into a C-S-H layer through the interaction with silicate dimers. However, other nucleation mechanisms are possible. By comparison with other materials, the PNC pathway explained in section 4.2 can also be present in the nucleation of C-S-H in which PNCs may form from Ca and Si ions during the early stages of cement hydration. The structural properties of those PNCs might be very important to understand the nucleation and growth of C-S-H. They would form a dense liquid-like phase before crystallizing into C-S-H. For instance, an spherical dense phase with a diameter of 50 $\pm$ 10 nm has been identified using Cryo-TEM, STEM and EDX [75] which reinforces the importance of PNC pathway in C-S-H nucleation.

The structure of C-S-H PNCs has been studied before using *ab initio* methods. Manzano et al. identified a basic building block of C-S-H by isolating a small section of tobermorite and jennite layer comprising 4 Ca and Si ions and minimizing its structure using Hartree-Fock [83] (see Fig. 5.1). The aggregation of these building blocks was also studied and found as a promising candidate for the formation of a tobermorite-like layer. Yang et al. used DFT to calculate the formation free energies of different complexes formed at the early
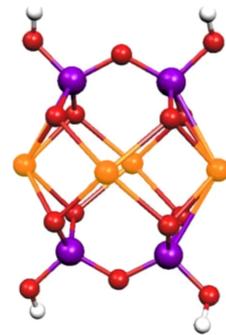
Figure 5.1: Atomic structure of C-S-H nanoparticle found by Manzano et. al. [83].

stages of hydration including the complexes with
one Ca and one Si ions [84]. They found that the
formation of complexes from a silicate and calcium species in dissolution
is favorable in terms of free energy. Finally, similar clusters containing
one Ca and one Si ion were studied by *ab initio* metadynamics by Li et
al. [85]. They looked for the best structures for the complex with Ca co-
ordination from 2 to 6 and found that those with 6-fold Ca coordination
are the most stable.
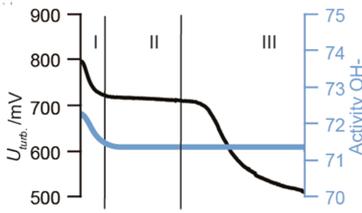
Figure 5.2: Evolution of turbidity and solution concentration from ref. [75].

From turbidity results [75] it is known
that the nucleation of C-S-H is governed
by non-classical nucleation (see Fig. 5.2).
Dense spheroids are found to be formed
as an intermediate step in C-S-H nucle-
ation. PNCs are probably formed within
those spheroids from the calcium and sil-
icate species in dissolution. Therefore,
to elucidate what type of cluster can be
formed it is important to know which
are the major species in dissolution. Using inductively coupled plasma
and optical emission spectrometry (ICP-OES) it was determined that the
major Ca species in cement at its commonly observed pH of 12.5 are
$Ca(OH)^+$ and Ca^{+2} [188]. For the silicates, the major species in disso-
lution were found using deprotonation constant and are $H_3SiO_4^{-1}$ and
$H_2SiO_4^{-2}$ [189]. Therefore, the complexes must be formed from these
mentioned species.

In this chapter, we use evolutionary algorithms to explore the con-
figurational space of some PNCs. Although PNCs of any stoichiometry
may exist in principle, this study is centered on clusters with a Ca/Si
ratio equal to 1. Hence, even though in theory charged clusters may
form, for simplicity we only studied clusters with neutral charge valance.
Therefore, the studied clusters would have to be formed by the aggre-
gation of $Ca(OH)^+$ and $H_3SiO_4^{-1}$ species or by the aggregation of Ca^{+2}
and $H_2SiO_4^{-2}$. Therefore, the selected stoichiometries are of the form
$nCaO + nSiO_2 + xH_2O$. For simplicity, we will refer to these stoichiome-
tries as $C_nS_nH_x$ using cement chemistry notation, where C stands for
CaO, S stands for SiO_2 and H stands for water molecules. In terms of
their size, clusters with 1, 2 and 4 Ca and Si ions have been studied and
different hydration levels have been chosen to study its effect in the clus-
ters. To sum up, the following stoichiometries have been studied: CSH_{10},
CSH_6, CSH_2, $C_2S_2H_{20}$, $C_2S_2H_{12}$, $C_2S_2H_4$, $C_4S_4H_8$, $C_4S_4H_4$ and $C_4S_4H_2$.
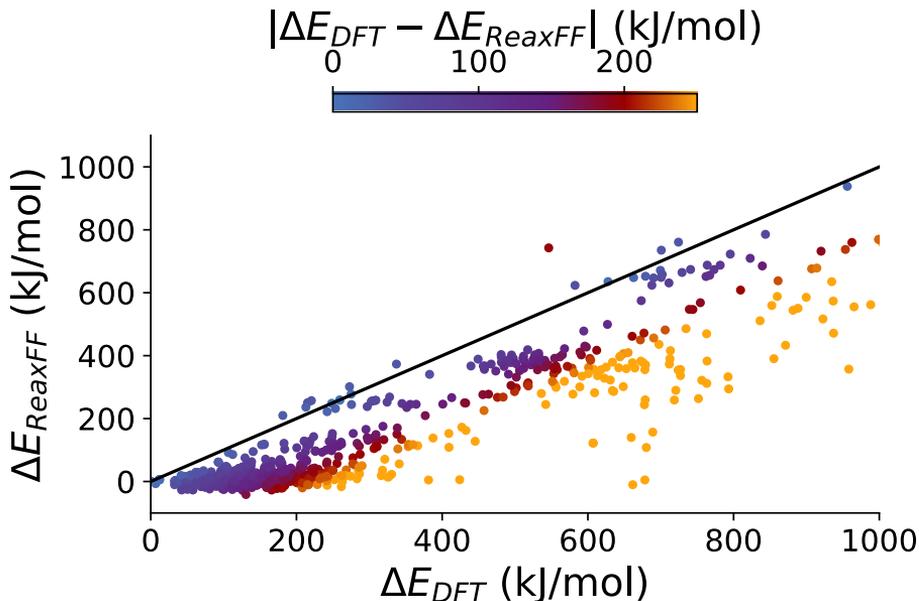
Figure 5.3: Comparision of ReaxFF and DFT energies for different CSH_{10} clusters.

1 | SIMULATION DETAILS

The procedure for finding the best clusters was similar for each stoichiometry. First, USPEX was used to create a large number of semi-random structures of the selected stoichiometry and they were relaxed using the ReaxFF force field. From those relaxed structures, USPEX was used to select a small group of the best structures taking into account that all the selected structures are structurally different. The number of selected structures depends on the cluster size. This selection was used as the first generation for an evolutionary algorithm using US-PEX as explained in sec 6. The minimizations in the evolutionary algorithms were done first using ReaxFF and then DFT for each structure. This is done in order to save computational time since reaxFF handles very quickly the first part of the minimization while DFT accurately finishes the minimization. The difference in energy between ReaxFF and DFT can be seen in Fig. 6.1 which shows that ReaxFF underestimates the energy of most clusters. That is why de use of "ab initio" methods is necessary to obtain accurate results.

An example of the energy evolution of clusters after generations can be seen in Fig. 5.4. This figure represents the energy evolution of the

$C_2S_2H_{12}$ cluster but will be similar for any other cluster. It can be observed that the randomly generated structures usually have very high energy. However, the clusters created by heredity and mutation from the randomly created clusters present much lower energy. Therefore, the randomly created clusters are very important to explore different areas of the configurational space and the heredity and mutation operators are essential to obtain low-energy clusters.

> **Simulation parameters:** 500 structures were created by USPEX for the first generation. This first generation was relaxed only with ReaxFF [170, 171, 172, 173] using the GULP code [190]. The relaxations of the structures were done using the conjugate gradient method. For the next generations, each generation contains between 15 and 25 structures depending on the size of the cluster. 60% of the best structures were used from the last generation to create the new generation. The structures were relaxed using ReaxFF first and then DFT within the SIESTA /citesoler2002siesta code. The DFT calculations were done using the GGA-PBE [94] functional together with the Grimme dispersion correction [191] to include the van der waals forces. The conjugate gradient was used to relax the structure until the maximum force is less than 0.04 ev/
> A to a maximum of 400 steps.

Once the calculation of the evolutionary algorithm is finished, we ended up with a set of thousand of structures and their energies. Although the founding of the energetically best structure is important, it is also interesting to investigate common properties of the whole set to find structural properties that correlate with their stability. To do so, different characterization parameters like the coordination numbers, gyration radius, number of water molecules or internal distances were calculated for each structure. To look for different correlations, the structural similarity between all the structures was calculated using SOAP [133, 174], thus creating a distance matrix between all of them. This information is used then to represent the set and the similarities between its structures in a 2d plot. This is done using sketch-map. In sketch-map [192] each structure is represented by a dot and the distance between dots represents the structural distance between the structures they represent. In this way, by plotting the same sketch-map side by side and using a colormap to represent different structural properties, correlations between these properties (including energy) can be found.
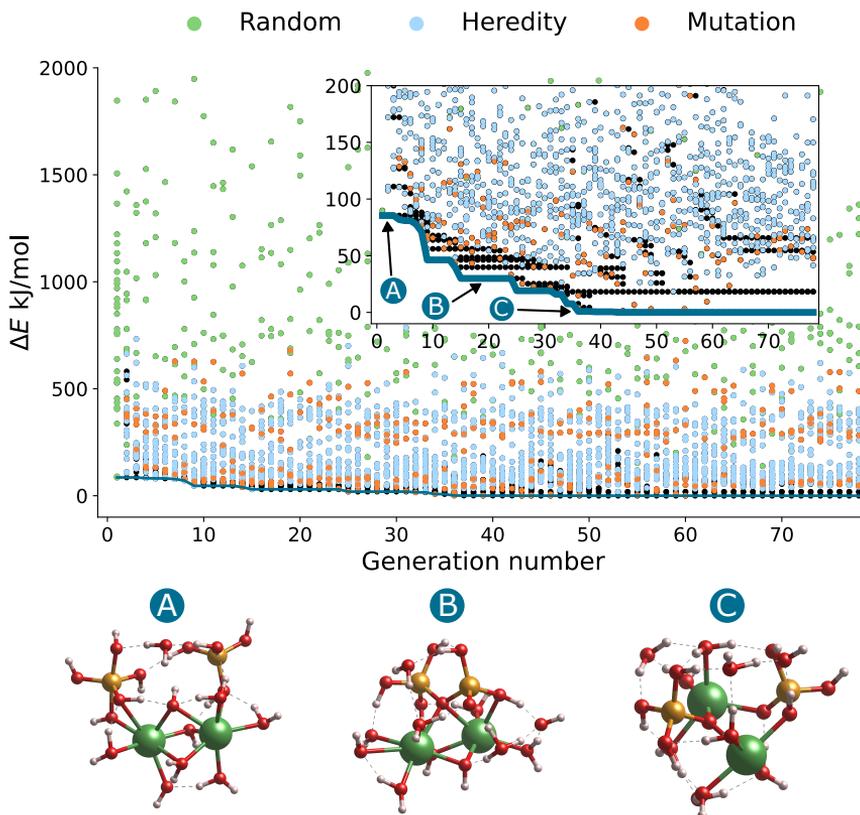
Figure 5.4: Evolution of the energies of each cluster in each generation. The green dots represent clusters created randomly, the blue dots represent clusters created by heredity and the orange dots represent clusters created by permutation of atoms or sofmutation. The black dots are clusters that have been kept the same since they were the best of the previous generation.

2 | CSH$_x$ CLUSTERS

I t is reasonable to believe that the first clusters that would form in the early stages of cement hydration are aggregates from the species in dissolution. The smallest of these clusters would then be clusters with one Ca atom and one Si atom. These complexes are studied in this section with different levels of hydration.
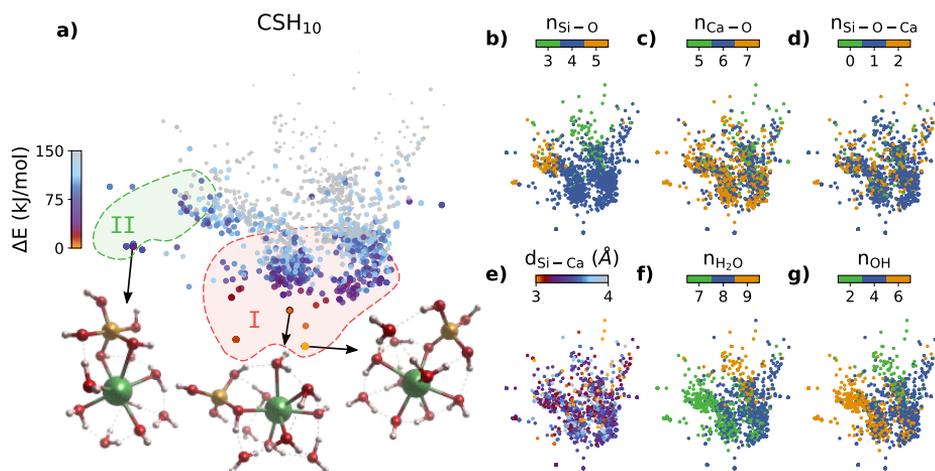
Figure 5.5: Sketch-map representation of the complex structures found by evolutionary algorithms. Each point represents a different complex, and the distances between points represent the structural similarity between the complexes. The closer the points are, the more similar the structures they represent are and vice versa. In each figure, the color represents different properties of the structures. The atomic structure of the best complexes is also shown in **(a)** where Ca atoms are represented in green, Si atoms in orange, O atoms in red and H atoms in white.

2.1 | CSH_{10}

Figure 5.5 shows a sketch-map of the clusters found during the global minimization with energy lower than E_0 + 150 kJ/mol, where E_0 is the energy of the lowest energy cluster. In Fig. 5.5(a) the color scale represents the energy difference while in Figs. 5.5(b-g) the colors represent a mapping of several structural properties. The sketch-map shows 2 different regions that can be distinguished in Fig. 5.5(a). In region I. the low energy complexes all contain a silicate tetrahedron bonded to a 6 or 7-fold coordinated Ca atom [fig. 5.5(c)]. The silicon and calcium atoms are mostly bonded by a single Si-O-Ca bond, and 7 or 8 water molecules are created around the complex, with 4 of 5 filling the Ca hydration shell and 3 in the second hydration shell. In region II the complexes include 5-coordinated silicate groups, with energies at least 1.35 kJ/mol higher than those of region I. In this case, the silicon and calcium atoms are mostly bonded by two Si-O-Ca bonds [Fig. 5.5 (d)]. The appearance of pentacoordinate silicate is not a surprise as it is a common transient structure during condensation or hydrolysis reactions, especially in high pH environments [193].

The complexes presented here are very similar to those studied before
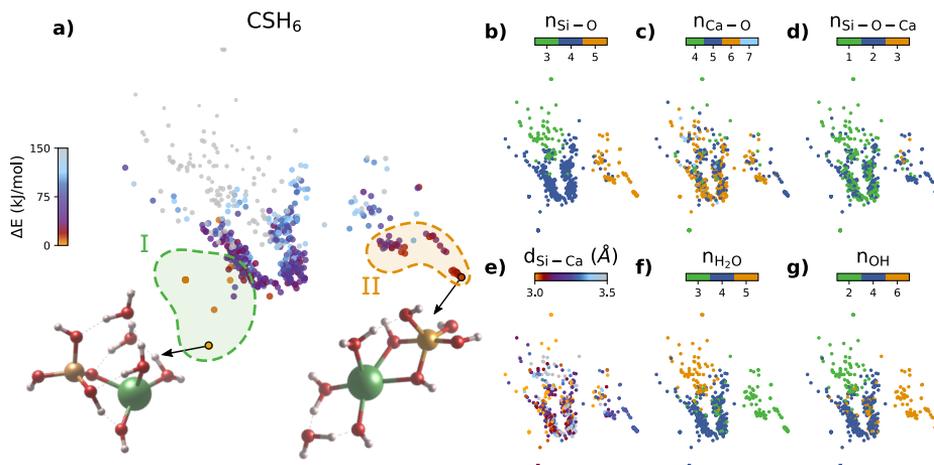
Figure 5.6: Sketch-map representation of the complex structures found by evolutionary algorithms. Each point represents a different complex, and the distances between points represent the structural similarity between the complexes. The closer the points are, the more similar the structures they represent are and vice versa. In each figure, the color represents different properties of the structures. The atomic structure of the best complexes is also shown in **(a)** where Ca atoms are represented in green, Si atoms in orange, O atoms in red and H atoms in white.

by ab-initio metadynamics [85]. All those clusters are also formed by a silicate tetrahedron bonded to the Ca atom by 1 or 2 Ca-O-Si bonds. It is also important to notice that many complexes within accessible energies present 7-coordinated atoms (see figure 1c). That is the coordination of Ca atoms in the C-S-H gel intralayer region [83, 194], and it is common to use such a structural feature to characterize the early formation of C-S-H from Synchrotron X-ray powder diffraction [52, 195]. Our simulations indicate that the complexes may already display such coordination, and therefore they may interfere with the characterization of the solid phase.

To see how dehydration affects the structure of the complexes and their stability, similar calculations have been made with less water for stoichiometries CSH$_6$ and CSH$_2$, as shown in the following sections.

2.2 | CSH$_6$

The structure of the best clusters does not change much. The two same groups can be identified in the sketchmap shown in Fig. 5.6(a). The first one comprises the complexes with less energy and they all have a silicate tetramer while the second group is characterized for having a
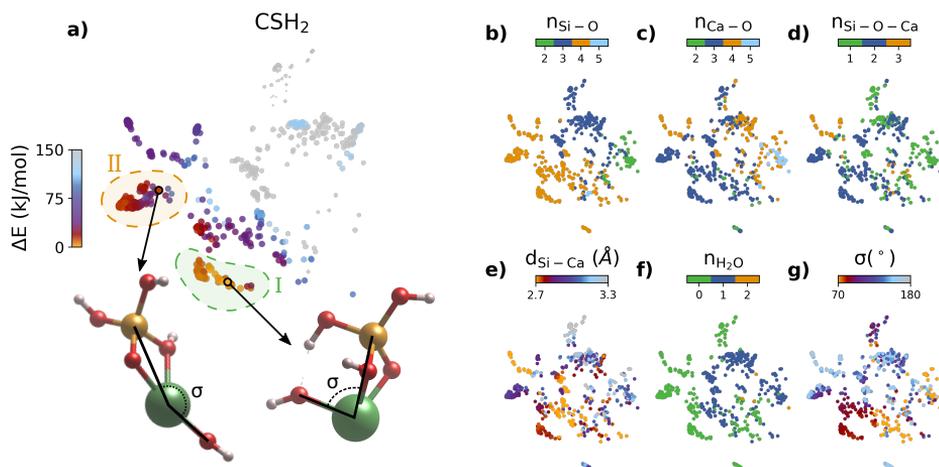
Figure 5.7: Sketch-map representation of the complex structures found by evolutionary algorithms. Each point represents a different complex, and the distances between points represent the structural similarity between the complexes. The closer the points are, the more similar the structures they represent areand vice versa. In each figure, the color represents different properties of the structures. The atomic structure of the best complexes is also shown in **(a)** where Ca atoms are represented in green, Si atoms in orange, O atoms in red and H atoms in white.

silicate pentamer with at least 12.86 kJ/mol more energy. Both groups are very similar to the groups identified in the more hydrated CSH_{10} stoichiometry.

The calcium coordination has been reduced compared to the CSH_{10} clusters, as there is less water to be coordinated with. Group I presents a Ca coordination of 4 or 5 having at least always one water molecule in the second hydration shell. One could expect that the cluster will be more stable with a higher 6-fold coordinated Ca and no water molecules in the second hydration layer, but forming hydrogen bonds seems to be equally favorable from an energetic point of view. The same happens in group II where Ca is coordinated with 5 oxygen atoms and water is also at the second hydration layer. Two of the coordinated oxygens are also coordinated to the Si atoms thus creating two Ca-O-Si bonds like in the more hydrated cluster.

2.3 | CSH_2

In the fully dehydrated stoichiometry, there are no clusters with silicate pentamers due to the small amount of oxygen available. In this stoi-

chiometry, for Si to be coordinated to 5 O atoms, Ca have to remain very poorly coordinated, which is known to be unfavorable [85]. Therefore, all the low energy complexes contain a silicate pentamer bonded to the Ca ion by two Ca-O-Si bonds, leaving the clusters with only one Ca-O-Si bond much more unstable.

In the sketch-map shown in Fig. 5.7 two types of clusters are identified, although the difference is so small that they could be categorized in the same group. The structural difference is based on the Si-Ca-O bond angle σ shown in Fig 5.7(g). The clusters in group I are the most stable ones and present a σ angle of around 70°. The clusters in group II contain a σ angle between 145° and 180° which contributes to a gain of at least 10 kJ/mol.

3 | C$_2$S$_2$H$_x$ CLUSTERS

N ow that the one Ca atom complexes have been studied, it is interesting to investigate how would these complexes aggregate. For that reason, we executed the same procedure but with bigger stoichiometries containing two Ca ions. That is, the following stoichiometries have been studied: C$_2$S$_2$H$_{20}$, C$_2$S$_2$H$_{12}$ and C$_2$S$_2$H$_4$.

3.1 | C$_2$S$_2$H$_{20}$

This stoichiometry contains two Ca atoms, two Si atoms and enough water so that the clusters are fully hydrated. Since these clusters contain more ions, the number of different structures is also increased. Besides the Ca and Si coordination, now the clusters can be also characterized by silicate polymerization, that is, if it contains two monomers or a dimer. Moreover, how the two calcium atoms are bonded to each other is also a parameter that characterizes these clusters.

As can be seen in the sketch-map in Fig 5.8(a), two main groups of low-energy clusters can be identified. Group I contains the lowest energy clusters whose main characteristic is that they contain a silicate dimer. The dimer is usually formed by two silicate tetrahedra although silica pentahedra are also present. One of the oxygens in the tetrahedra is used to create the Si-O-Si bond and another one to create a Si-O-Ca bond leaving each tetrahedron with two hydroxyl groups. The Ca atoms are
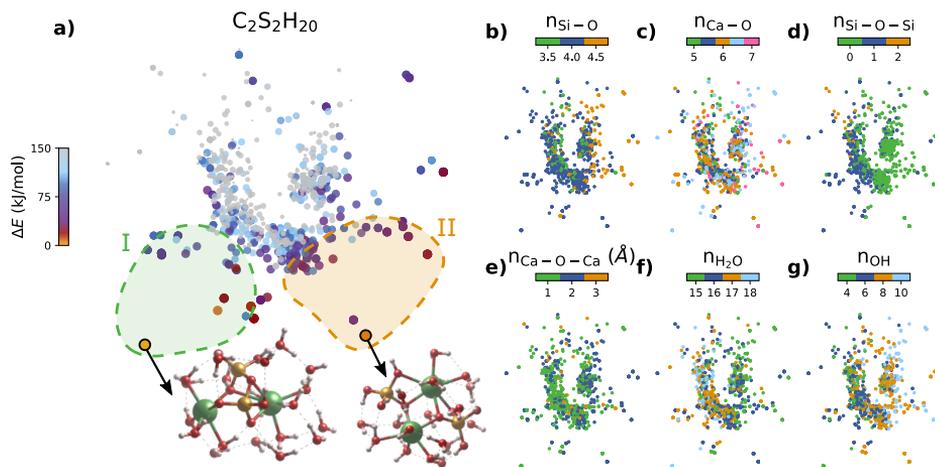
Figure 5.8: Sketch-map representation of the clusters found by evolutionary algorithms. Each point represents a different cluster, and the distances between points represent the structural similarity between them. The closer the points are, the more similar the structures they represent are and vice versa. In each figure, the color represents different properties of the structures. The atomic structure of the best clusters is also shown in **(a)** where Ca atoms are represented in green, Si atoms in orange, O atoms in red and H atoms in white.

mainly bonded by one Ca-O-Ca bond yet some structures contain two of those bonds. It is interesting that in this group the Ca-O-Ca bonds are almost always perpendicular to the Si-O-Si bonds. On the other hand, group II contains clusters with two silicate monomers. These two silicate monomers are usually far apart from each other on one end of the cluster. The two Ca ions are bonded mostly by two Ca-O-Ca bonds and the oxygen atoms participating in those bonds are also bonded to a silicon atom. Therefore, in these clusters, there are always two oxygen atoms that are each bonded to two calcium atoms and one silicon atom. The mean Ca coordination is 6 in group II while in group I is 6.5, thus meaning that there are some Ca atoms with 7-fold coordination.

In terms of energy, the best cluster in group II is 5.41 kJ/mol higher in energy. This difference is very low, meaning that both clusters with a silicate dimer and clusters with two silicate monomers are energetically very stable and accessible in terms of energy. Moreover, the transformation from a cluster with two silicate monomers to a cluster with two silicate dimers is would imply a condensation reaction whose transformation energy would be of 87 kJ/mol.
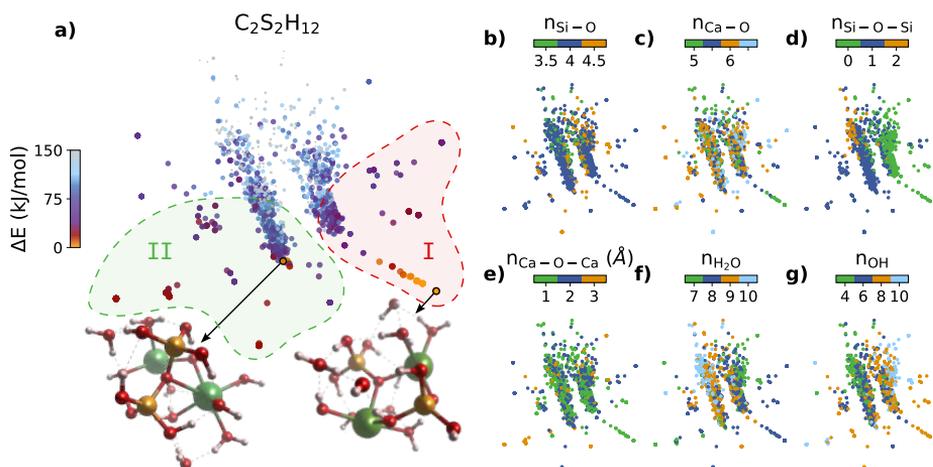
Figure 5.9: Sketch-map representation of the clusters found by evolutionary algorithms. Each point represents a different cluster, and the distances between points represent the structural similarity between them. The closer the points are, the more similar the structures they represent are and vice versa. In each figure, the color represents different properties of the structures. The atomic structure of the best clusters is also shown in **(a)** where Ca atoms are represented in green, Si atoms in orange, O atoms in red and H atoms in white.

3.2 | C$_2$S$_2$H$_{12}$

Two well-differentiated regions can be distinguished for the C$_2$S$_2$H$_{12}$ see [Fig. 5.9(d)]: Region I (where the most stable candidates are found) with independent silicate monomers, and Region II with the two silicates joined by a Si-O-Si bond forming dimmers. The energy difference between the best candidate from group I and the best candidate from group II is just 4.34 kJ/mol, which indicates that although structures with independent monomers are more stable, the energy difference is low and both types of structures are accessible. This is very similar to the results obtained in the more hydrated cluster C$_2$S$_2$H$_{20}$ although in that cluster the structures containing a silicate dimer were more stable.

The best structures of region I contain mostly 5-fold coordinated Ca atoms although some 6-fold coordinated Ca are also present. The two Ca atoms are bonded by a 1 Ca-O-Ca bond and the silicate monomers are mostly tetrahedra. In region II, the Ca-O coordination is usually bigger ranging from a mean value of 5.5 to 6.5. The two calcium atoms are mostly bonded with 1 or 2 Ca-O-Ca bonds and although silicate dimers mostly contain 4-fold coordination, some 5-fold coordinated silica are present as seen in Fig. 5.9(b).
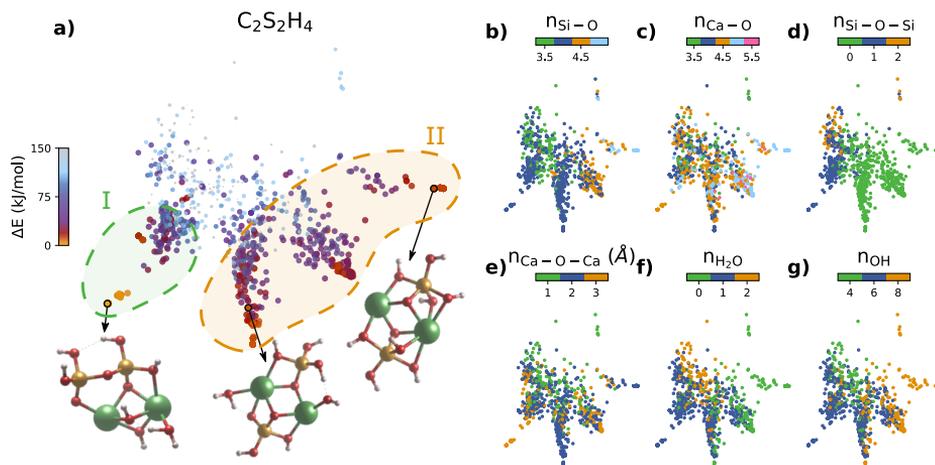
Figure 5.10: Sketch-map representation of the clusters found by evolutionary algorithms. Each point represents a different cluster, and the distances between points represent the structural similarity between them. The closer the points are, the more similar the structures they represent are and vice versa. In each figure, the color represents different properties of the structures. The atomic structure of the best clusters is also shown in **(a)** where Ca atoms are represented in green, Si atoms in orange, O atoms in red and H atoms in white.

3.3 | C$_2$S$_2$H$_4$

With the less hydrated stoichiometry, two different regions can be identified. As in the more hydrated clusters, the groups are differentiated due to silicate polymerization. Group I contains clusters with silicate dimers. However, this time, in contrast to the more hydrated clusters, the Si-O-Si bond and the Ca-O-Ca bonds are parallel. In this way, although the cluster is less hydrated, greater coordination of calcium ions is achieved, which is energetically favorable. The mean coordination of the calcium ions is indeed between 3 and 4. Among those, 2 or 3 oxygen atoms participate in the Ca-O-Ca bonds bonding the two Ca ions. In the case of 3 Ca-O-Ca bonds, one of those oxygen atoms is also bonded to one of the silicate tetrahedra. The silicate ions are bonded by one oxygen thus creating a Si-O-Si bond. This type of cluster is very similar to the Ca and Si arrangement in tobermorite.

In group II the clusters contain two silicate monomers and the best cluster of this group has 6.29 kJ/mol more energy than the best cluster in group I. The two Ca ions are bonded mostly by 2 Ca-O-Ca bonds although some 3 Ca-O-Ca bonds can be found in this group. The group itself can be subdivided into 3 subgroups depending on the silicate co-

ordination. The subgroup with the more stable cluster contains clusters with two silicate tetrahedra. There is another subgroup with one silicate tetrahedra and one silicate pentahedra and the final subgroup contains two silicate pentahedra. These two last subgroups have around 12 kJ/mol more energy and therefore are accessible from an energetic perspective.

Looking at all the clusters with 2 Ca ions there can be seen that although different hydration levels both the clusters with silicate dimers and clusters with two silicate monomers are stable and accessible from an energetical point of view. It is also interesting to see that as the water content decreases dimers are more stable. This is in agreement with the appearance of silicate dimers in C-S-H, which are created when the structures are dehydrated.

4 | C$_4$S$_4$H$_x$ CLUSTERS

Now let us concentrate on the clusters with 4 Ca and Si atoms. These clusters would correspond to the aggregation of two C$_2$S$_2$H$_n$ clusters. These clusters contain 4 Si ions, thus can create different polymerization of silicate. The different arrangements of the silicate tetrahedra are 4 silicate monomer, one silicate dimer and two silicate monomers, 2 silicate dimers, 1 silicate trimer and 1 silicate monomer, and a closed chain of a silicate tetramer. These different arrangements are represented in Fig. 5.11(i).

4.1 | C$_4$S$_4$H$_8$

In this case, we divided the similarity map again into two regions. Region I comprises low-energy PNCs with heterogeneous characteristics in terms of silicate chain network, the number of water molecules and OH groups, as shown in Fig. 5.11(a). It is interesting to notice that the lowest energy structures are well-differentiated within region I, and share similar characteristics: they are formed by a silicate monomer and a trimer, and their gyration radius is considerably smaller than that of clusters of other regions. Therefore, a more compact structure is energetically favorable. However, it is known that trimers do not appear at any stage of C-S-H formation [196, 44], and free energy simulations of CaCO$_3$ PNCs indicate the opposite trend in the free energy with the gy-
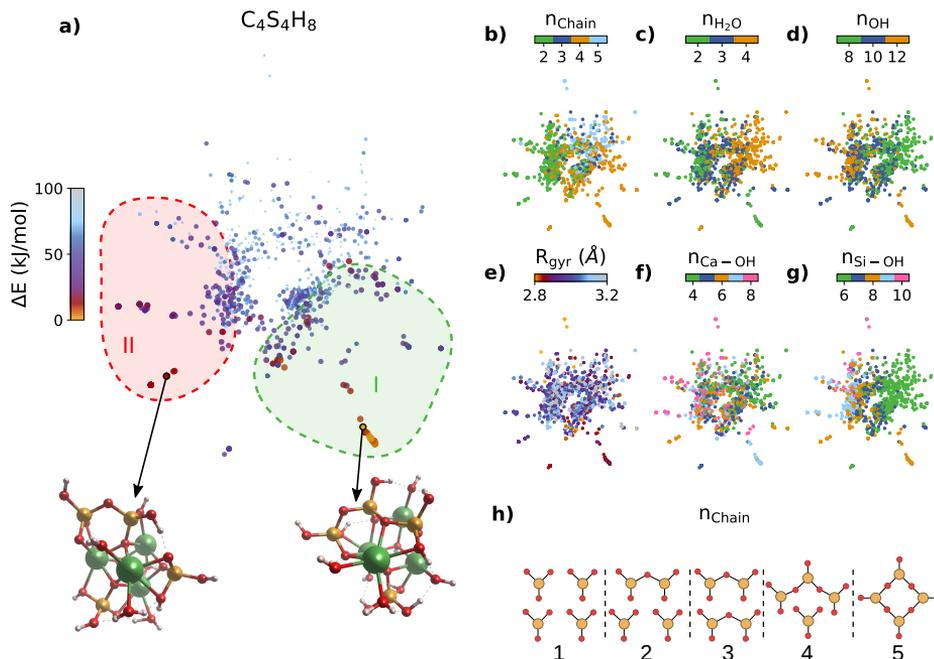
Figure 5.11: Sketch-map representation of the clusters found by evolutionary algorithms. Each point represents a different cluster, and the distances between points represent the structural similarity between them. The closer the points are, the more similar the structures they represent are and vice versa. In each figure the color represents different properties of the structures. The atomic structure of the best clusters is also shown in **(a)** where Ca atoms are represented in green, Si atoms in orange, O atoms in red and H atoms in white.

ration radius [62]. This could be because larger clusters are formed as aggregates of smaller clusters that already contain a dimer and therefore, the previously formed dimers limit the formation of trimers. Region II contains clusters with a silicate dimer and 2 monomers. The best cluster in region II has 11.28 kJ/mol more energy than the best cluster in region I. These clusters agree much more with the experimental results since dimers appear from the early stages of cement hydration. Additionally, as mentioned above, the remaining two monomers in these clusters can easily form another dimer by a condensation reaction. Another interesting thing is that all the clusters present a structure in which Ca ions are sandwiched by the Si tetrahedra like in the tobermorite layer. Therefore, this suggests that the arrangement of Ca ions between Si dimers may appear from the early stages of cement hydration.
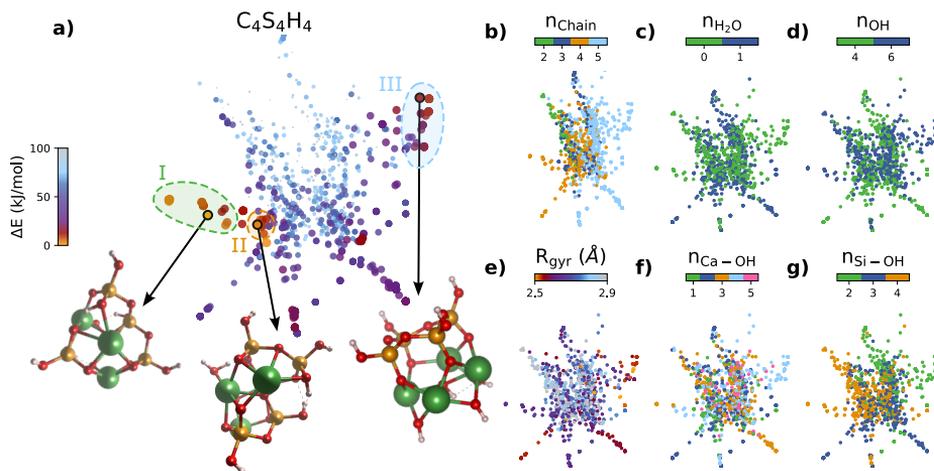
Figure 5.12: Sketch-map representation of the clusters found by evolutionary algorithms. Each point represents a different cluster, and the distances between points represent the structural similarity between them. The closer the points are, the more similar the structures they represent are and vice versa. In each figure, the color represents different properties of the structures. The atomic structure of the best clusters is also shown in **(a)** where Ca atoms are represented in green, Si atoms in orange, O atoms in red and H atoms in white.

Between all the C$_4$S$_4$H$_8$ clusters no low-energy cluster with two dimers in its structure has been found. This may be because that structure is not very stable or because evolutionary algorithms have not been able to explore properly that area of the configuration space.

4.2 | C$_4$S$_4$H$_4$

The C$_4$S$_4$H$_4$ can be subdivided into three regions depending on the arrangement of the silicate tetramers. The low energy clusters are comprised in region I and as in the more hydrated cluster, they all contain a silicate trimer and a monomer. Again, these type of PNCs are not very realistic since they have not been observed in any experiments.

Region II contains a small number of clusters in which two dimers are present. These two dimers sandwich the Ca ions in a similar manner to what happens in tobermorite. The energy difference with the best cluster of region I is 5.28 kJ/mol, which makes these clusters energetically accessible at room temperature and therefore, this is in agreement with the dimers found experimentally at the early stages of cement hydration [196].
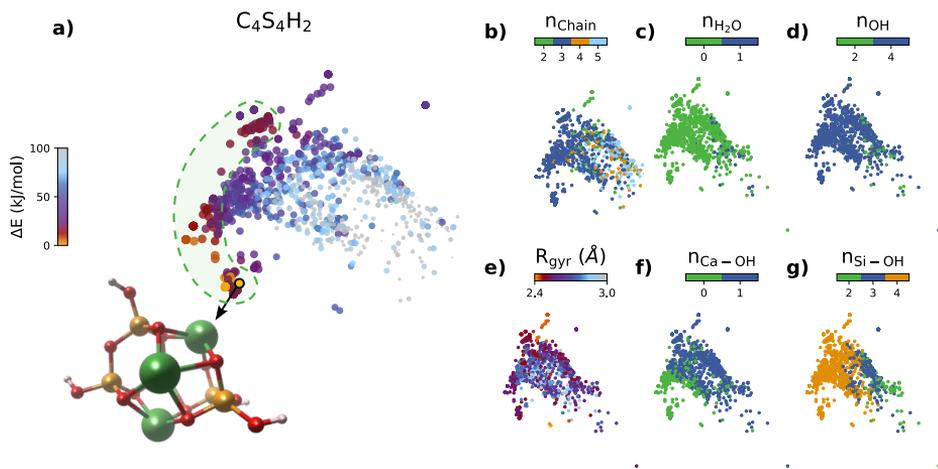
Figure 5.13: Sketch-map representation of the clusters found by evolutionary algorithms. Each point represents a different cluster, and the distances between points represent the structural similarity between them. The closer the points are, the more similar the structures they represent are and vice versa. In each figure, the color represents different properties of the structures. The atomic structure of the best clusters is also shown in **(a)** where Ca atoms are represented in green, Si atoms in orange, O atoms in red and H atoms in white.

Finally, region III comprises clusters with a ring of silicates and have 10.08 kJ/mole more energy than the best cluster in region I. This is the only type of cluster in which the silicate is not sandwiching the Ca ions. This arrangement of silicate has not been observed experimentally and therefore, this result is probably due to the simulation conditions and will not be very realistic.

4.3 | $C_4S_4H_2$

Within the clusters with less water content. we see that all the low-energy clusters belong to the same region. These clusters contain two dimers, each on one side of the cluster, and between them a layer of Ca ions, very similar to tobermorite although the two dimers are perpendicular in the cluster and parallel in a tobermorite layer (see Fig. 5.14). As a matter of fact, this result is very promising since this cluster is very similar to the one found by Manzano et al. [83]. They found a C-S-H nanoparticle by extracting a tobermorite building block and minimizing its structure. The structure of this nanoparticle is very similar to our cluster but with parallel silicate dimers. However, we have found that the cluster found in this work is energetically more stable.
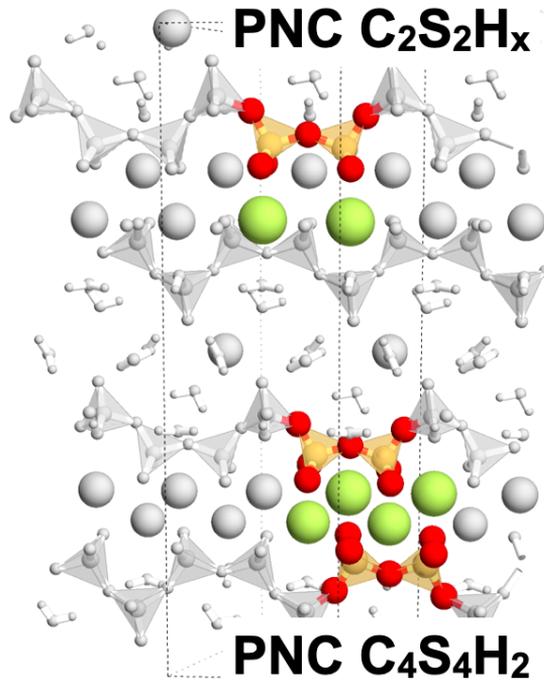
Figure 5.14: Tobermorite structure in which its building blocks are highlighted. These highlighted building blocks have beend found as C-S-H PNCs.

PNCs that have similar structures to the bulk structure have been seen before in other materials such as apoferritin protein [71]. Therefore, the cluster found here is very promising and will probably participate in the nucleation of a C-S-H layer.

5 | FORMATION PATH

A fter determined the best clusters for the different sizes, we discuss a possible formation pathway from the ions in solution up to the C-S-H prenucleation cluster $C_4S_4H_2$. For that, we computed the sequential enthalpy of formation of the complexes/PNCs formed by merging smaller structures. For each size, the lowest energy clusters were further relaxed using DFT using the COSMO implicit solvation model with an epsilon value of $\epsilon = 56$ to simulate the pH=12.5 found in cement hydration [197].

$Ca(OH)(H_2O)^+$ $Ca(H_2O)_6^{2+}$ $SiO_4H_3(H_2O)^-$ $SiO_4H_2(H_2O)_5^{2-}$
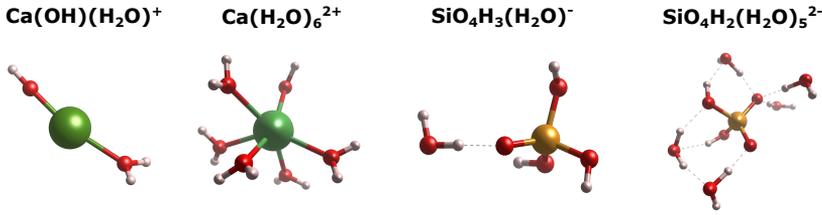
Figure 5.15: Main Ca and Si species in dissolution at high pH.

Product	Reagents	ΔG (KJ/mol)
CSH_{10}	$CaO_2H_3^+ + SiO_5H_5^- + 6H_2O$	-91.1
	$CaO_6H_{12}^{+2} + SiO_9H_{12}^{-2} - 2H_2O$	-156.7
$C_2S_2H_{20}$	$2CaO_2H_3^+ + 2SiO_5H_5^- + 12H_2O$	-127.5
	$2CaO_6H_{12}^{+2} + 2SiO_9H_{12}^{-2} - 4H_2O$	-193.2
	$2CSH_{10}$	-36.5
$C_4S_4H_8$	$4CaO_2H_3^+ + 4SiO_5H_5^- - 8H_2O$	-132.7
	$4CaO_6H_{12}^{+2} + 4SiO_9H_{12}^{-2} - 40H_2O$	-198.4
	$4CSH_{10} -32H_2O$	-41.7
	$2CSH_{10} + C_2S_2H_{12} - 30H_2O$	-25.1
	$C_2S_2H_{12}-I + C_2S_2H_{12}-II - 16H_2O$	-14.7
$C_4S_4H_2$	$4CaO_2H_3^+ + 4SiO_5H_5^- - 14H_2O$	-140.4
	$4CaO_6H_{12}^{+2} + 4SiO_9H_{12}^{-2} - 46H_2O$	-206.0
	$4CSH_{10} -38H_2O$	-49.3
	$2CSH_{10} + C_2S_2H_{12} - 30 H_2O$	-32.8
	$2C_2S_2H_{12} - 22H_2O$	-16.2

Table 5.1: Formation free energies.

The first step is the formation of complexes from the ions in solution. The ions have been selected based on the predominant speciation states of Ca and silicic acid in high-pH environments. According to [188, 189], Ca is present mainly as Ca^{2+} and as $Ca(OH)^+$ in a lesser extent. In the case of Si, the main species in dissolution is $H_2SiO_4^{-2}$ followed by $H_3SiO_4^{-1}$. To simulate better these species in dissolution a certain amount of water molecules have been added surrounding the species. The number of water molecules needed to simulate correctly this species was already calculated by Yang et al. [84]: 6 water molecules for Ca^{2+}, 1 for $Ca(OH)^+$, 5 for $H_2SiO_4^{-2}$ and 1 for $H_3SiO_4^{-1}$.

All the enthalpies of formation for the complex formation are shown in table 5.1, and figure 5.17 shows only the most favorable ones, which correspond to the pathways starting from the ionic species with the highest charges $Ca^{2+}x_6H_2O$ and $H_2SiO_4^{-2}x_5H_2O$ as reactants. This can be
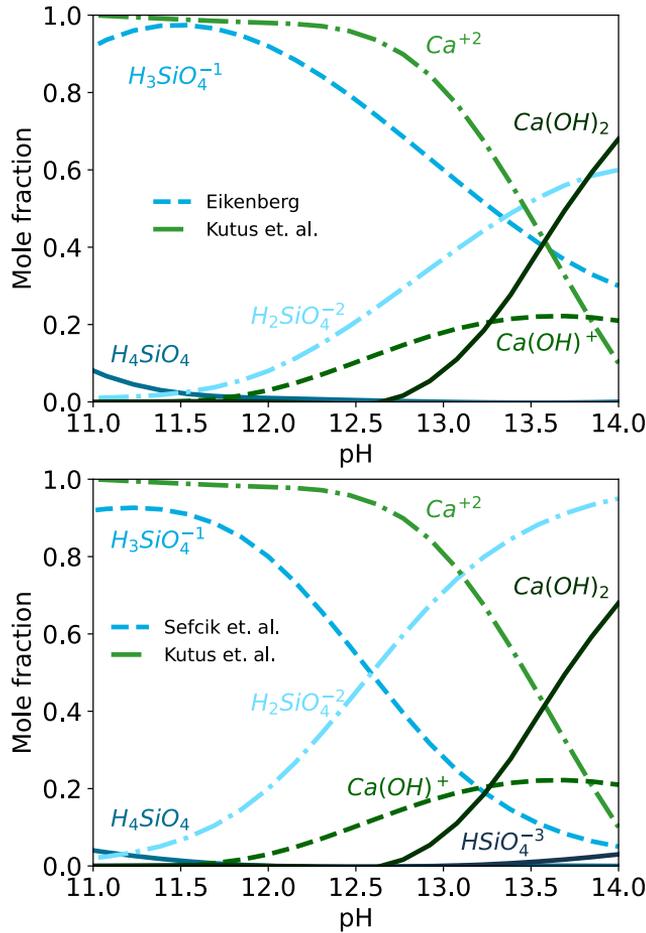
Figure 5.16: Molar fraction of the Ca and Si species in dissolution at different pH values. Data from Refs. [188, 181, 189].

due to a worse stabilization of the charge in the simulations, even with the implicit solvent approximation.

The enthalpy of formation for the CSH_{10} complex is 156.75 kJ/mol. The formation of the complex is very exothermic, indicating a considerable stabilization of the ions. It must be noted that even though the initial species in solution are charged, the complex is neutral, with a $H_3SiO_4^-$ silicate sharing its deprotonated oxygen atom with the Ca atom, which completes its hydration sphere with 5 water molecules and a hydroxyl group. Nevertheless, we can suggest that the formation of the CSH_{10} complex is very favorable and this specie could be dominant in solution soon after the dissolution.

The second step is the formation of the $C_2S_2H_{12}$ PNCs by the associ-
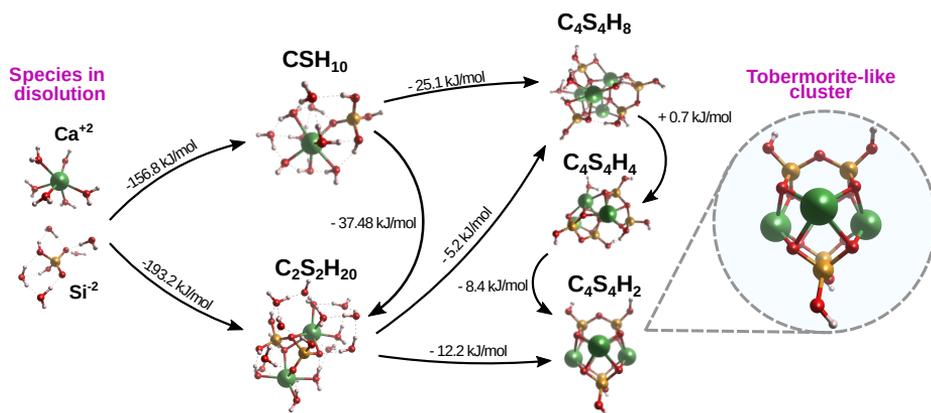
Figure 5.17: Formation mechanism of the pre-nucreation cluster in C-S-H hydration. Free energy results are given in kJ/mol.

ation of complexes. The formation enthalpy from the CSH_{10} complex is -66.11 kJ/mol and therefore merging the complexes into a PNC is favorable. The formation enthalpy of the best $C_2S_2H_{20}$ PNC was computed, which contains a silicate dimer. It is well known that silicate dimers appear from the early stages of C-S-H formation. In particular, it has been reported by CPMAS Si-NMR [196] that dimers form mainly within the first hour of OPC hydration. Experimentally we cannot determine at which formation stage the condensation reactions take place, but based on our computational results we suggest that it can happen at this PNC characteristic size.

The last step in the formation path involves the formation of $C_4S_4H_x$ PNCs. Based on the fact that EA predicts clusters with silicate trimers as the most favorable ones, we suggest that $C_4S_4H_x$ clusters are not present in solution. However, the $C_4S_4H_2$ structures can be seen as C-S-H basic building blocks, as shown in figure 5.17, so we propose that they will be formed by progressive dehydration within the large $C_2S_2H_{20}$ aggregates. Dehydration has been suggested to be a key step in $CaCO_3$ nucleation, as releasing water leads to an increase of the entropy, reducing the free energy of the nuclei [62, 61, 198]. Our DFT simulations of isolated clusters indicate that dehydration is slightly favorable. More interesting is the evolution of the $C_2S_2H_{20}$ aggregates in solution. As can be seen in figure 4f, the mean number of water molecules attached to the aggregates decreases with time. The decrease is due to the aggregation itself since the PNCs lose water molecules from their hydration shells, and also due to rearrangements within clusters, including condensation reactions between the $C_2S_2H_{20}$ and $Ca(OH)_2$. Despite our simulations being too short

to observe full dehydration, we suggest that $C_4S_4H_2$ clusters will be progressively formed within the aggregates leading to crystallization, following the experimental steps found experimentally by Kraitwurst et al. [75].

6 | CONCLUSIONS

E volutionary algorithms have been found to be very useful for exploring the most stable clusters for different C-S-H stoichiometries. Moreover, as the configuration space is explored, correlations can be found between the structural properties of the clusters and their energies. That allows us to group the different types of clusters that can be formed for each stoichiometry.

Exploring the first complexes that would appear in cement hydration, it has been shown that those complexes are formed by a silicate tetrahedron bonded to a 6- or 7-folded Ca atom. In addition, the decrease in water content does not alter the main structure of the cluster, that is, the silicate tetrahedron bonded to the Ca ion. Another group with silicate pentahedrons has also been found but with larger energy than those with a tetrahedron.

The clusters with two Ca and Si ions show that the most stable clusters contain a silicate dimer. However, clusters with two monomers are also very stable. Due to this result, it can be stated that silicate dimers may form in the formation of clusters of this size.

With the exploration of larger PNCs, there are two main conclusions. The first one is that although knowing that trimers do not appear in cement hydration, in this work it is shown that some of the best clusters contain silicate trimers within their structures. This difference may be attributed to the fact that the cluster search calculations have been made in vacuum, to the fixed stoichiometry of the calculations or to the difficulty to find an energy minimum cluster with a high number of atoms. Bigger clusters with more Si atoms may present more dimers or pentamers, which are very common in cement hydration. The second conclusion is that the dehydration of the clusters stabilizes the dimers, and therefore, monomers and trimers are less likely to form in less hydrated clusters.

Finally, a very dehydrated cluster has been found that looks very simi-

lar to a tobermorite layer. This result is very remarkable since a proposal of a tobermorite-like cluster was presented before taking a small section of a tobermorite layer and minimizing it [83]. The fact that we have now found a similar result with evolutionary algorithms reaffirms the possibility that these clusters form at the early stages of cement hydration. Whether these clusters could form larger tobermorite-like clusters is unknown, but undoubtedly, it would be a study of great interest, as it could broaden the knowledge of the C-S-H nucleation mechanisms.

CHAPTER 6

MAIN CONCLUSIONS

Jupiter
The Planets
Gustav Holst

C ement is the most used manufactures material in the entire world, and therefore, due to the impact it has in the gloval economy and in the CO_2 emissions, the study of its structure and properties has great interest to improve its behavior. Once hydrated, the most important phase in cement is the C-S-H gel, which is the major responsible for the mechanical properties of the final product, together with its durability and impermeability. C-S-H has been studied for years. However, there are still a lot of questions to be answered at the atomic scale. One of the most interesting issues is how the species in solution nucleate into the solid C-S-H gel.

The main objective of this thesis was to study the early stages of C-S-H at the atomic scale. To do so, computational methods are a very useful tool as this scale is beyond the scope of the main experimental methods. However, computational methods require some approximations to be able to simulate the required systems in an affordable simulation time. For that reason, DFT has been used to relax the structures and to calculate transition paths with VC-NEB, but ReaxFF force field has been used to perform long MD simulations in order to calculate transition free energies. Several computational tools have also been used for the analysis of the results of this thesis. SOAP was used to calculate the structural similarity between different structures, and KPCA and Sketch-map were used to perform a dimensionality reduction of all the calculated structural similarities so that correlations between structural properties and

energy can be made. This is one of the keys to being able to analyze large databases of structures, such as those generated in this thesis.

As the atomic process of C-S-H nucleation is not yet known, it is open to hypotheses of all kinds. This thesis is based on two hypothesis. The first one being the nucleation of C-S-H from a portlandite mnonolayer (studied in chapters 3 and 4) and the second one being the PNC pathway from the main species in dissolution (studied in chapter 5).

The first hypothesis was based on the similarity of the Ca-O structure in the C-S-H and that of the Sr-O structure in the $Sr(OH)_2$ crystal. Although it seems that there might be little relevance to this similarity, it must be remembered that Sr and Ca are isoelectronic atoms and therefore have similar behavior. That is why the first thing that has been investigated in chapter 3 is the stability of Phase V, a new theoretical polymorph of $Ca(OH)_2$ with the crystal structure of $Sr(OH)_2$. This new polymorph, Phase V, has been found to be metastable with 2.78 kcal/mol higher energy compared to that of portlandite. The phase transition between portlandite and Phase V has been studied using VC-NEB and an energy barrier of 10.13 kcal/mol has been determined. This energy barrier is similar to that of phase transitions in other materials and therefore energetically accessible. This result opens the door to the possibility that the C-S-H formation may originate from portlandite. This process should take place in a single portlandite layer, so it is plausible to think that portlandite monolayers are formed in the early stages of cement hydration, and then modified by the interaction with silicate dimers into a C-S-H layer.

To investigate the possibility that a portlandite layer could be converted into a C-S-H layer by interaction with silicate dimers, the condensation reaction that will trigger this transformation was first studied with NEB in chapter 4. This reaction is absolutely necessary for the transformation to occur since it is what creates the Ca-O-Si bonds that we observe in the C-S-H. The activation energy for this reaction has been determined to be 21.0 kcal/mol. Then, the transition from a portlandite layer to a C-S-H layer was studied. Although the formation of portlandite layers in the early stages of cement hydration has not been investigated, it is plausible to think that they would be formed due to their stability. To study the complete transformation we first studied its mechanism and internal energy using TMD and found that the transformation has no energy barrier and that the energy only decreases at the end of the transformation when all the condensation reactions oc-
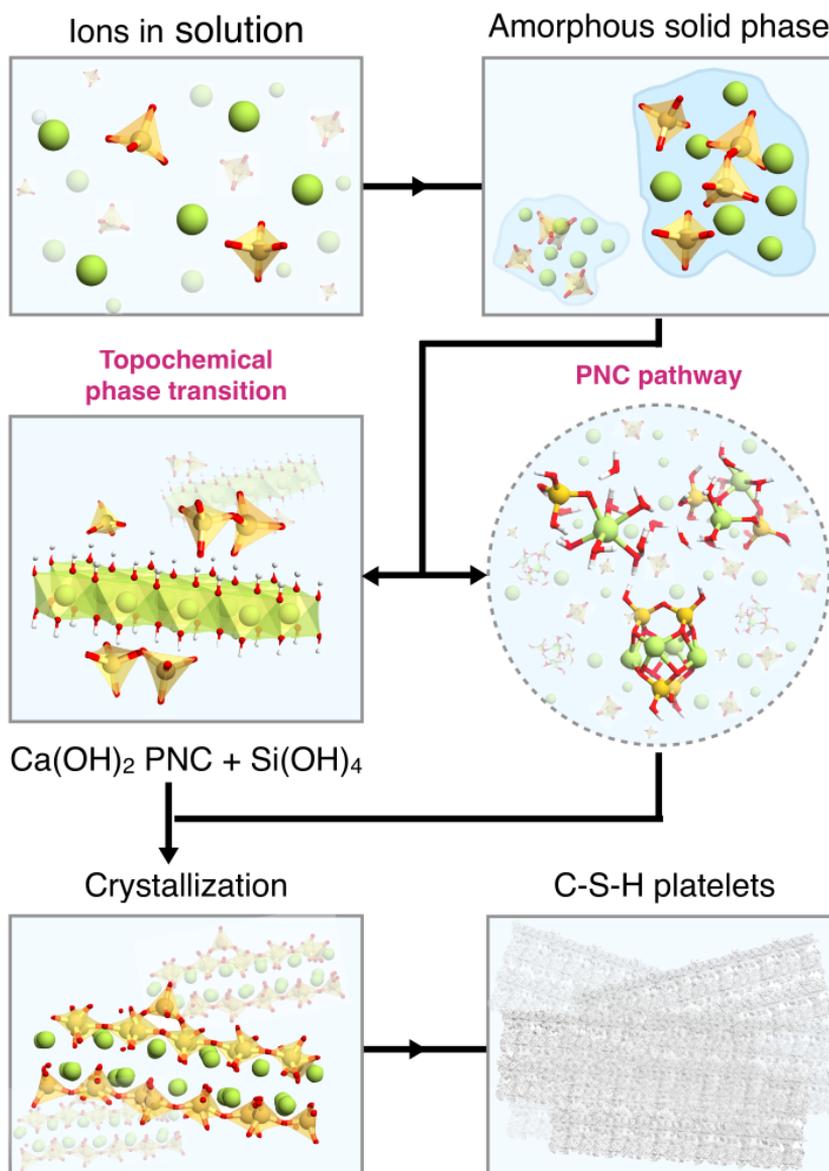
Figure 6.1: Graphical representation of the two formation mechanisms explored in this thesis for C-S-H.

cur. Finally, the free energy of the transition has been measured using US and it has been observed that although the free energy is positive, its increase is only 1.5 kcal/mol. This result indicates that although a priori this hypothesis of the transformation of the portlandite layer into a C-S-H layer is not spontaneous, it cannot be totally discarded.

The hypothesis that portlandite layers could play a key role in the formation of C-S-H has been always floating around. Back in 1952, Bernal et al. proposed that the formation of C-S-H was the result of a modified portlandite layer by condensation of monomeric silicate anions [199, 200]. Shpynova et al. [201] claimed that C-S-H was the result of the condensation of the 4 hydroxyl groups of an orthosilicic acid molecule with portlandite. However, these models were discarded due to the incongruencies with the experimental distribution of silicate groups found by ^{29}Si NMR measurements. In this thesis, this hypothesis was again restated and studied at the atomistic level for the first time. This could lead to future studies of this exact process or similar processes. Moreover, This study opens the door to the reconsideration that the formation of C-S-H from a portlandite layer may be one of the mechanics of C-S-H formation among others.

The second hypothesis is based in the recent advances in the nucleation of materials such as $CaCO_3$. It has been proposed that PNCs may be one of the nucleation kernels. For that reason, it is interesting to know what kind of C-S-H PNCs could be formed in the early stages of cement hydration. For this purpose, we have used genetic algorithms to find the best clusters for several stoichiometries: CSH_x, $C_2S_2H_x$ and $C_4S_4H_x$. The study of the smaller clusters with 1 Ca and Si ions showed that Ca is more stable when it is coordinated to 6-7 oxygen atoms. In addition, silicates prefer to be tetracoordinated although some pentacoordinated silicates have also been found with low energy. As for the polymerization of silicates, the study of clusters with 2 Ca and Si ions showed that both dimers and monomers are very similar in energy stability. Studying the bigger clusters, it is shown that although experimentally we know that trimers cannot appear in C-S-H, some low-energy clusters have been found which contained these trimers. This is probably since the earlier formation of clusters with dimers limits the formation of trimers. The most interesting finding is that as the larger clusters dehydrate, they increasingly resemble the tobermorite structure. Moreover, a very dehydrated cluster has been found that is basically a building block of tobermorite. This makes the result very promising as it has been seen in other materials that PNCs often adopt crystal-like shapes before nucleation.

On the one hand, the study of the PNCs in C-S-H helps to understand and propose models for the formation of C-S-H since these clusters could be the key to nucleation. Bigger clusters with different stoichiometries should be studied in the future to have a more complete understanding of the aggregates that could form in the early stages of cement hydration. On the other hand, this work sets a reliable method to look for PNCs of any kind an opens a whole new opportunity to study the structures of the PNCs of any other material. As future work, it would be interesting to see how these PNCs aggregate using MD to see what kind of structures they create. The experiments we have done so far point to the fact that the PNCs would aggregate into a kind of DOLLOPS chains.

In conclusion, this thesis has been useful to test successfully different computational methods never used until now to study the nucleation of C-S-H. This has helped us to study on the one hand the formation of a C-S-H layer from a portlandite layer and to investigate on the other hand the structures of the best PNCs in C-S-H that could be part of the nucleation. However, this thesis also opens new avenues for research and generates many other questions about C-S-H nucleation.

Anyway, like anything else in life, this thesis also requires an ending. Therefore, just as Mozart left his Requiem unfinished, we will leave (for a different reason) this research for its composition to be continued in the future.

ACKNOWLEDGEMENTS

Baldorba
Hitaz oroit
Benito Lertxundi

It has been long years since I started this thesis. Before I started, I was not convinced to start a thesis on cement. I considered cement a material too messy to be able to do real science. However, Hegoi Manzano convinced me from the beginning by showing me his great passion for this subject. I never thought it would take me so long to finish this thesis but looking back I realize how many things I have done in these 5 and a half years besides the thesis. Undoubtedly, these years will mark my life forever as I feel that I leave as a different person than the one I entered: with more knowledge, more maturity, more passion and with the feeling that I still know nothing (which will probably haunt me throughout my life).

First of all, I would like to thank my directors Hegoi Manzano and Iñigo Etxebarria who have taught me everything I know about research. I met Iñigo while I was taking the optics course in the physics degree at the UPV/EHU. I have to admit that this subject was not one of my favorite ones but I discovered in Iñigo a great passion for physics as well as a fun and relaxed way of doing science. On the other hand, I met Hegoi while I was doing my final degree work because he was my work director. In that period, Hegoi taught me that computational techniques were very useful to do science and thanks to that I decided to do a master's degree purely dedicated to modeling and computation. After finishing my master, Hegoi trusted me completely and offered me without hesitation to do a PhD with him. I have to say that I was not very passionate about the topic he proposed to me, but the confidence he had in me and the good relationship with him did not make me doubt that it would be a good place to do my doctoral thesis. From the beginning, both directors have treated me as an equal and I was surprised that I could talk about science with them giving my opinion without it being discredited simply because I was a young PhD student. With them, I have discovered that research can be done in a healthy environment where pressure, confrontation, competitiveness and hierarchy are replaced by motivation, understanding, group work and respect. Moreover, with them, I have learned to look at research from a broader and interdisciplinary point of

view where all disciplines are interested to understand the world around us. I hope to continue working with you in the future.

I also wanted to thank all the colleagues who have passed through the department and who have done the thesis with me. To Paul for teaching me the basics of crystallography, to Santos for the fun conversations we had in the office, to Iñigo for the long political conversations, to Lafu for singing James Taylor on any Tuesday at 17:00, to Telmo for his musical recommendations, to Haritz for sharing my desire to abandon the thesis in the hardest moments and to Oscar, the Basqueized Castilian. I also wanted to thank all the young people who have come after me and who are the future of this department: Malen, Jon Otegi, Jon Gabirondo, Aitor and Mireia. I would especially like to thank two people who have passed through here. First of all to Peio who has been the person with whom I have shared more moments in this office. For the long conversations we have had about politics and whole afternoons singing loudly in the office. You are a reference for me Peio. The other person I want to thank personally is Jon (aka Balma) for helping me during the thesis and for working side by side on the most complicated scientific problems I encountered during the thesis. It is a pleasure to work with you.

I wanted to thank Gose Naiz Taldea (GNT) for the support I have received over the years and for those interesting conversation meals. I would also like to thank Bariku Forala and all its members for the eloquent and endless research we have done on the Basque people every Friday at 13:00. To Calvo, to Delri, to Eukeni, to Mikel, to Alain, to Irune and despite being a traitor to foral values, I also want to thank Balma for being our infiltrated liberal. I especially want to thank the most important member of Bariku Forala, Mr. Simon, for being the perfect traveling companion throughout this thesis. No decision I have made without consulting him first. We have spent this thesis studying everything but what we had to do for the thesis. For example history, religion, mythology, Latin, music theory and much more. It has been a great learning experience to share with him all these years.

I would also like to thank those who made my stay in Grenoble more pleasant. To Alejandro and Alexander for helping me in my research and to Rogier for being such a good co-worker and for accompanying me in all my meals in Grenoble. I especially want to thank Alicia for her time with me, who from the first moment was willing to help me with all the problems that a Basque can encounter in France. I hope we will meet again some other time.

Finalmente me gustaría agradecer a la gente más cercana. A mi cuadrilla por ofrecerme una vía de escape cuando la he necesitado. A toda la gente con la que comparto mucho tiempo haciendo música juntos. Ya lo sabéis pero sois el motor que me hace seguir siempre adelante. Gracias por vuestro apoyo constante. Pero sobretodo quería agradecer a los que me han apoyado a diario. A Inés, por estar conmigo en los buenos momentos pero sobretodo por soportarme en los malos. Es imposible que hubiera acabado esta tesis sin tu apoyo diario. Eskertu nahiko nuke ere familia osoa eta bereziki nire gurasoak. Beti euki duzue nigan konfiantza osoa eta animatu didazue gustatzen zitzaidana egitera. Horren ondorioz nago orain nagoen lekuan. Benetan eskerrik asko. Azkenik, amama eskertzea bakarrik falta zait. Zure etxea beti izan da ikasteko leku aproposa eta zurekin sofan eduki ditudan elkarrizketak ez ditut inoiz ahaztuko. Danoi bihotz bihotzez, eskerrik asko!

LIST OF PUBLICATIONS

Published in refereed Journals

- Zuniga, F. J., Cruz-Cabeza, A. J., Aretxabaleta, X. M., De La Pinta, N., Breczewski, T., Quesada-Moreno, M. M., ... & Elguero, J. (2018). *Conformational aspects of polymorphs and phases of 2-propyl-1H-benzimidazole.* IUCrJ, 5(6), 706-715.

- Aretxabaleta, X. M., Etxebarria, I., & Manzano, H. (2020). *Electronic and elastic properties of brownmillerite.* Materials Research Express, 7(1), 015516.

- Aretxabaleta, X. M., Gonchenko, M., Harshman, N. L., Jackson, S. G., Olshanii, M., & Astrakharchik, G. E. (2020). *The Dynamics of Digits: Calculating Pi with Galperin's Billiards.* Mathematics, 8(4), 509.

- Aretxabaleta, X. M., López-Zorrilla, J., Labbez, C., Etxebarria, I., & Manzano, H. (2022). *A potential CSH nucleation mechanism: atomistic simulations of the portlandite to CSH transformation.* Cement and Concrete Research, 162, 106965.

Conference proceedings

- Klar, P.B.; Aretxabaleta, X.M.; Madariaga, *G. Exploiting superspace to enable DFT calculations of modulated structures with disordered sites using the example of mullite,* in Proceedings of the 1st International Electronic Conference on Crystals, 21–31 May 2018, MDPI: Basel, Switzerland, doi:10.3390/IECC_2018-05253

- De la Pinta, N., Cruz-Cabeza, A. J., TomaszBreczewski, Aretxabaleta, X. M., Zuniga, J., Mar Quesada-Moreno, M., ... & Elguero, J. (2018, August). *Polymorphism in 2-propyl-1H-benzimidazole.* In Acta Crystallographica A-Foundation and advances (Vol. 74, pp. E339-E340). 2 Abbey Sq, Chester, CH1 2HU, England: Int Union Crystallography.

- Aretxabaleta, X. M., Etxebarria, I., & Manzano, H. (2018).Brownmilleritaren (Ca2AlFeO5) propietate elastikoak, Materialen Zientzia eta Teknologia IV. Kongresuko lanen bilduma. ISBN 978-84-8081-608-3

- Aretxabaleta, X. M., Etxebarria, I., & Manzano, H. (2019). *Investigation of the calcium silicate hydrate potential energy surface.* 15th International Congress on the Chemistry of Cement, 33, Research Institute of Binding Materials Prague, ISBN 978-80-906541-4-3

Submitted

- Aretxabaleta, X.M., López-Zorrilla, J., Etxebarria, I., & Manzano,

H. (2023). *C-S-H nucleation pathway from atomistic simulations.* Currently in formal peer review in Nature Communications.

Bibliography

[1] Dieter Arnold. *Building in Egypt: pharaonic stone masonry*. Oxford University Press on Demand, 1991.

[2] Hilal Tuğba ÖRMECİOĞLU. The historic concrete: Construction and architecture. *ENGINEERING AND TECHNOLOGY MANAGEMENT*, page 210, 2021.

[3] Lepenski Vir a mesolithic site on the iron gates gorge of the danube. https://www.donsmaps.com/lepenski.html. Accessed: 2022-05-23.

[4] Amelia Carolina Sparavigna. Some notes on ancient concrete. *International Journal of Sciences February*, 2014.

[5] WS Richard. History of concrete. *PDF). The Aberdeen Group. Archived from the original (PDF) on*, 28, 2015.

[6] Robert G Blezard. The history of calcareous cements. *Lea's chemistry of cement and concrete*, 4, 1998.

[7] Geoffrey Martin. A hundred, years of portland cement, 1824–1924. *Nature*, 115(2883):150–151, 1925.

[8] United States Geological Service cement statistics and information. https://www.usgs.gov/centers/national-minerals-information-center/cement-statistics-and-information. Accessed: 2022-05-23.

[9] Robbie M Andrew. Global co 2 emissions from cement production, 1928–2017. *Earth System Science Data*, 10(4):2213–2239, 2018.

[10] Daniele Poponi, Tyler Bryant, Keith Burnard, Pierpaolo Cazzola, John Dulac, Araceli Fernandez Pales, Joerg Husar, Peter Janoska, Eric R Masanet, Luis Munuera, et al. *Energy technology perspectives 2016: towards sustainable urban energy systems*. International Energy Agency, 2016.

[11] Peter Hewlett and Martin Liska. *Lea's chemistry of cement and concrete*. Butterworth-Heinemann, 2019.

[12] Harry FW Taylor et al. *Cement chemistry*, volume 2. Thomas Telford London, 1997.

[13] HT Cao, Liana Bucea, A Ray, and S Yozghatlian. The effect of cement composition and ph of environment on sulfate resistance of portland cements and blended cements. *Cement and Concrete Composites*, 19(2):161–171, 1997.

[14] Henri Le Chatelier. *Recherches expérimentales sur la constitution des mortiers hydrauliques*. Dunod, 1904.

[15] Karen Scrivener, Alexandre Ouzia, Patrick Juilland, and Aslam Kunhi Mohamed. Advances in understanding cement hydration mechanisms. *Cement and Concrete Research*, 124:105823, 2019.

[16] Karen L Scrivener and André Nonat. Hydration of cementitious materials, present and future. *Cement and concrete research*, 41(7):651–665, 2011.

[17] Ellis Gartner. Discussion of the paper "dissolution theory applied to the induction period in alite hydration" by p. juilland et al., cem. concr. res. 40 (2010) 831–844. *Cement and Concrete research*, 41(5):560–562, 2011.

[18] Ellis M Gartner and Hamlin M Jennings. Thermodynamics of calcium silicate hydrates and their solutions. *Journal of the American Ceramic Society*, 70(10):743–749, 1987.

[19] Patrick Juilland, Emmanuel Gallucci, Robert Flatt, and Karen Scrivener. Dissolution theory applied to the induction period in alite hydration. *Cement and Concrete Research*, 40(6):831–844, 2010.

[20] Patrick Juilland, Luc Nicoleau, Rolf S Arvidson, and Emmanuel Gallucci. Advances in dissolution understanding and their implications for cement hydration. *RILEM Technical Letters*, 2:90–98, 2017.

[21] Frank Bellmann, Thomas Sowoidnich, Horst-Michael Ludwig, and Denis Damidot. Dissolution rates during the early hydration of tricalcium silicate. *Cement and Concrete Research*, 72:108–116, 2015.

[22] Klaas Van Breugel. Numerical simulation of hydration and microstructural development in hardening cement-based materials (i) theory. *Cement and concrete research*, 25(2):319–331, 1995.

[23] Jeffrey J Thomas. The instantaneous apparent activation energy of cement hydration measured using a novel calorimetry-based method. *Journal of the American Ceramic Society*, 95(10):3291–3296, 2012.

[24] Maria Mercedes Costoya Fernández. Effect of particle size on the hydration kinetics and microstructural development of tricalcium silicate. Technical report, EPFL, 2008.

[25] Shashank Bishnoi and Karen L Scrivener. Studying nucleation and growth kinetics of alite hydration using μic. *Cement and Concrete Research*, 39(10):849–860, 2009.

[26] David M Kirby and Joseph J Biernacki. The effect of water-to-cement ratio on the hydration kinetics of tricalcium silicate cements: Testing the two-step hydration hypothesis. *Cement and Concrete Research*, 42(8):1147–1156, 2012.

[27] Joseph J Biernacki and Tiantian Xie. An advanced single particle model for c3s and alite hydration. *Journal of the American Ceramic Society*, 94(7):2037–2047, 2011.

[28] L Nicoleau and A Nonat. A new view on the kinetics of tricalcium silicate hydration. *Cement and Concrete Research*, 86:1–11, 2016.

[29] Amélie Bazzoni. Study of early hydration mechanisms of cement by means of electron microscopy. Technical report, EPFL, 2014.

[30] Alexandre Reda Constantin William Caleb Ouzia. Modeling the kinetics of the main peak and later age of alite hydration. Technical report, EPFL, 2019.

[31] Alexandre Ouzia and Karen Scrivener. The needle model: a new model for the main hydration peak of alite. *Cement and Concrete Research*, 115:339–360, 2019.

[32] Karen L Scrivener, Patrick Juilland, and Paulo JM Monteiro. Advances in understanding hydration of portland cement. *Cement and Concrete Research*, 78:38–56, 2015.

[33] JA Gard and HFW Taylor. Calcium silicate hydrate (ii)("csh (ii)"). *Cement and Concrete Research*, 6(5):667–677, 1976.

[34] IG Richardson and GW Groves. Microstructure and microanalysis of hardened ordinary portland cement pastes. *Journal of Materials science*, 28(1):265–277, 1993.

[35] Ping Yu, R James Kirkpatrick, Brent Poe, Paul F McMillan, and Xiandong Cong. Structure of calcium silicate hydrate (c-s-h): Near-, mid-, and far-infrared spectroscopy. *Journal of the American Ceramic Society*, 82(3):742–748, 1999.

[36] Andrew J Allen, Jeffrey J Thomas, and Hamlin M Jennings. Composition and density of nanoscale calcium–silicate–hydrate in cement. *Nature materials*, 6(4):311–316, 2007.

[37] Paulo JM Monteiro, Guoqing Geng, Delphine Marchon, Jiaqi Li, Prasanth Alapati, Kimberly E Kurtis, and Mohammad Javad Abdolhosseini Qomi. Advances in characterizing and understanding the microstructure of cementitious materials. *Cement and Concrete Research*, 124:105806, 2019.

[38] LB Skinner, SR Chae, CJ Benmore, HR Wenk, and PJM Monteiro. Nanostructure of calcium silicate hydrates in cements. *Physical review letters*, 104(19):195502, 2010.

[39] HFW Taylor. Nanostructure of c-s-h: Current status. *Advanced cement based materials*, 1(1):38–46, 1993.

[40] Ana Cuesta, Jesus D Zea-Garcia, Diana Londono-Zuluaga, Angeles G De la Torre, Isabel Santacruz, Oriol Vallcorba, Monica Dapiaggi, Susana G Sanfélix, and Miguel AG Aranda. Multiscale understanding of tricalcium silicate hydration reactions. *Scientific reports*, 8(1):1–11, 2018.

[41] Ian G Richardson. The calcium silicate hydrates. *Cement and concrete research*, 38(2):137–158, 2008.

[42] K Mohan and HFW Taylor. A trimethylsilylation study of tricalcium silicate pastes. *Cement and Concrete Research*, 12(1):25–31, 1982.

[43] André Nonat. The structure and stoichiometry of csh. *Cement and concrete research*, 34(9):1521–1528, 2004.

[44] Ian G Richardson. The nature of csh in hardened cements. *cement and concrete research*, 29(8):1131–1147, 1999.

[45] IG Richardson and JG Cabrera. The nature of c-s-h in model slag-cements. *Cement and Concrete composites*, 22(4):259–266, 2000.

[46] Elena Bonaccorsi and Stefano Merlino. Modular microporous minerals: cancrinite-davyne group and csh phases. *Reviews in Mineralogy and Geochemistry*, 57(1):241–290, 2005.

[47] F Brunet, Ph Bertani, Th Charpentier, A Nonat, and J Virlet. Application of 29si homonuclear and 1h- 29si heteronuclear nmr correlation to structural studies of calcium silicate hydrates. *The Journal of Physical Chemistry B*, 108(40):15494–15502, 2004.

[48] Aslam Kunhi Mohamed, Stephen C Parker, Paul Bowen, and Sandra Galmarini. An atomistic building block description of csh-towards a realistic csh model. *Cement and Concrete Research*, 107:221–235, 2018.

[49] HFW Taylor and JW Howison. Relationships between calcium silicates and clay minerals. *Clay Minerals Bulletin*, 3(16):98–111, 1956.

[50] IG Richardson and GW Groves. Models for the composition and structure of calcium silicate hydrate (c-s-h) gel in hardened tricalcium silicate pastes. *cement and concrete research*, 22(6):1001–1010, 1992.

[51] Xiandong Cong and R James Kirkpatrick. 17o mas nmr investigation of the structure of calcium silicate hydrate gel. *Journal of the American Ceramic Society*, 79(6):1585–1592, 1996.

[52] Sylvain Grangeon, Alejandro Fernandez-Martinez, Alain Baronnet, Nicolas Marty, Agnieszka Poulain, Erik Elkaim, Cédric Roosz, Stéphane Gaboreau, Pierre Henocq, and Francis Claret. Quantitative x-ray pair distribution function analysis of nanocrystalline calcium silicate hydrates: a contribution to the understanding of cement chemistry. *Journal of applied crystallography*, 50(1):14–21, 2017.

[53] S Karthika, TK Radhakrishnan, and P Kalaichelvi. A review of classical and nonclassical nucleation theories. *Crystal Growth & Design*, 16(11):6663–6681, 2016.

[54] XY Liu. Heterogeneous nucleation or homogeneous nucleation? *The Journal of Chemical Physics*, 112(22):9949–9955, 2000.

[55] Martin Volmer and A Weber. Keimbildung in übersättigten gebilden. *Zeitschrift für physikalische Chemie*, 119(1):277–301, 1926.

[56] Joseph A Becker. Thermionic electron emission and adsorption part i. thermionic emission. *Reviews of Modern Physics*, 7(2):95, 1935.

[57] Julius Frenkel. A general theory of heterophase fluctuations and pretransition phenomena. *The Journal of Chemical Physics*, 7(7):538–547, 1939.

[58] Y Viisanen, R Strey, and H Reiss. Homogeneous nucleation rates for water. *The Journal of chemical physics*, 99(6):4680–4692, 1993.

[59] P Mirabel and Joseph L Katz. Condensaton of a supersaturated vapor. iv. the homogeneous nucleation of binary mixtures. *The Journal of Chemical Physics*, 67(4):1697–1704, 1977.

[60] Haoyang Fu, Xing Gao, Xin Zhang, and Lan Ling. Recent advances in nonclassical crystallization: Fundamentals, applications, and challenges. *Crystal Growth & Design*, 22(2):1476–1499, 2021.

[61] Denis Gebauer, Matthias Kellermeier, Julian D Gale, Lennart Bergström, and Helmut Cölfen. Pre-nucleation clusters as solute precursors in crystallisation. *Chemical Society Reviews*, 43(7):2348–2371, 2014.

[62] Raffaella Demichelis, Paolo Raiteri, Julian D Gale, David Quigley, and Denis Gebauer. Stable prenucleation mineral clusters are liquid-like ionic polymers. *Nature communications*, 2(1):1–8, 2011.

[63] Denis Gebauer, Antje Volkel, and Helmut Colfen. Stable prenucleation calcium carbonate clusters. *Science*, 322(5909):1819–1822, 2008.

[64] Denis Gebauer and Helmut Cölfen. Prenucleation clusters and non-classical nucleation. *Nano Today*, 6(6):564–584, 2011.

[65] Emilie M Pouget, Paul HH Bomans, Jeroen ACM Goos, Peter M Frederik, Gijsbertus de With, and Nico AJM Sommerdijk. The initial stages of template-controlled caco3 formation revealed by cryo-tem. *Science*, 323(5920):1455–1458, 2009.

[66] Adam F Wallace, Lester O Hedges, Alejandro Fernandez-Martinez, Paolo Raiteri, Julian D Gale, Glenn A Waychunas, Stephen Whitelam, Jillian F Banfield, and James J De Yoreo. Microscopic evidence for liquid-liquid separation in supersaturated caco3 solutions. *Science*, 341(6148):885–889, 2013.

[67] Martina Sammer, Cees Kamp, Astrid H Paulitsch-Fuchs, Adam D Wexler, Cees JN Buisman, and Elmar C Fuchs. Strong gradients in weak magnetic fields induce dollop formation in tap water. *Water*, 8(3):79, 2016.

[68] Jens Baumgartner, Archan Dey, Paul HH Bomans, Cécile Le Coadou, Peter Fratzl, Nico AJM Sommerdijk, and Damien Faivre. Nucleation and growth of magnetite from solution. *Nature materials*, 12(4):310–314, 2013.

[69] Jorg Polte, Robert Erler, Andreas F Thunemann, Sergey Sokolov, T Torsten Ahner, Klaus Rademann, Franziska Emmerling, and Ralph Kraehnert. Nucleation and growth of gold nanoparticles studied via in situ small angle x-ray scattering at millisecond time resolution. *ACS nano*, 4(2):1076–1082, 2010.

[70] Peter G Vekilov. Crystallization tracked atom by atom, 2019.

[71] S-T Yau and Peter G Vekilov. Direct observation of nucleus structure and nucleation pathways in apoferritin crystallization. *Journal of the American Chemical Society*, 123(6):1080–1089, 2001.

[72] S Garrault-Gauffinet and A Nonat. Experimental investigation of calcium silicate hydrate (csh) nucleation. *Journal of crystal growth*, 200(3-4):565–574, 1999.

[73] Rouhollah Alizadeh, Laila Raki, Jon M Makar, James J Beaudoin, and Igor Moudrakovski. Hydration of tricalcium silicate in the presence of synthetic calcium–silicate–hydrate. *Journal of Materials Chemistry*, 19(42):7937–7946, 2009.

[74] Elisabeth John, Thomas Matschei, and Dietmar Stephan. Nucleation seeding with calcium silicate hydrate–a review. *Cement and Concrete Research*, 113:74–85, 2018.

[75] Nina Krautwurst, Luc Nicoleau, Michael Dietzsch, Ingo Lieberwirth, Christophe Labbez, Alejandro Fernandez-Martinez, Alexander ES Van Driessche, Bastian Barton, Sebastian Leukel,

and Wolfgang Tremel. Two-step nucleation process of calcium silicate hydrate, the nanobrick of cement. *Chemistry of Materials*, 30(9):2895–2904, 2018.

[76] Markus Schönlein and Johann Plank. A tem study on the very early crystallization of csh in the presence of polycarboxylate superplasticizers: Transformation from initial csh globules to nanofoils. *Cement and Concrete Research*, 106:33–39, 2018.

[77] J Plank, M Schönlein, and V Kanchanason. Study on the early crystallization of calcium silicate hydrate (csh) in the presence of polycarboxylate superplasticizers. *Journal of Organometallic Chemistry*, 869:227–232, 2018.

[78] Enrico Masoero, Jeffrey J Thomas, and Hamlin M Jennings. A reaction zone hypothesis for the effects of particle size and water-to-cement ratio on the early hydration kinetics of c3s. *Journal of the American Ceramic Society*, 97(3):967–975, 2014.

[79] Enrico Masoero. Mesoscale mechanisms of cement hydration: Bng model and particle simulations, 2018.

[80] Jeffrey J Thomas. A new approach to modeling the nucleation and growth kinetics of tricalcium silicate hydration. *Journal of the American Ceramic Society*, 90(10):3282–3288, 2007.

[81] R González-Teresa, Jorge S Dolado, Andrés Ayuela, and Jean-Christophe Gimel. Nanoscale texture development of csh gel: A computational model for nucleation and growth. *Applied Physics Letters*, 103(23):234105, 2013.

[82] Igor Shvab, Laurent Brochard, Hegoi Manzano, and Enrico Masoero. Precipitation mechanisms of mesoporous nanoparticle aggregates: off-lattice, coarse-grained, kinetic simulations. *Crystal Growth & Design*, 17(3):1316–1327, 2017.

[83] H Manzano, A Ayuela, and JS Dolado. On the formation of cementitious c–s–h nanoparticles. *Journal of computer-aided materials design*, 14(1):45–51, 2007.

[84] Kengran Yang and Claire E White. Modeling of aqueous species interaction energies prior to nucleation in cement-based gel systems. *Cement and Concrete Research*, 139:106266, 2021.

[85] Yunjian Li, Hui Pan, and Zongjin Li. Ab initio metadynamics simulations on the formation of calcium silicate aqua complexes prior to the nuleation of calcium silicate hydrate. *Cement and Concrete Research*, 156:106767, 2022.

[86] Robert G Parr. Density functional theory of atoms and molecules. In *Horizons of quantum chemistry*, pages 5–15. Springer, 1980.

[87] Pierre Hohenberg and Walter Kohn. Inhomogeneous electron gas. *Physical review*, 136(3B):B864, 1964.

[88] Vladimir Fock. Näherungsmethode zur lösung des quantenmechanischen mehrkörperproblems. *Zeitschrift für Physik*, 61(1-2):126–148, 1930.

[89] John C Slater. The theory of complex spectra. *Physical Review*, 34(10):1293, 1929.

[90] Eugene Wigner. On the interaction of electrons in metals. *Physical Review*, 46(11):1002, 1934.

[91] David M Ceperley and Berni J Alder. Ground state of the electron gas by a stochastic method. *Physical review letters*, 45(7):566, 1980.

[92] John P Perdew and Alex Zunger. Self-interaction correction to density-functional approximations for many-electron systems. *Physical Review B*, 23(10):5048, 1981.

[93] Klaus Capelle. A bird's-eye view of density-functional theory. *Brazilian journal of physics*, 36:1318–1343, 2006.

[94] John P Perdew, Kieron Burke, and Matthias Ernzerhof. Generalized gradient approximation made simple. *Physical review letters*, 77(18):3865, 1996.

[95] Neil W Ashcroft and N David Mermin. Solid state physics (saunders college, philadelphia, 1976). *Appendix N*, 166, 2010.

[96] Felix Bloch. Über die quantenmechanik der elektronen in kristallgittern. *Zeitschrift für physik*, 52(7):555–600, 1929.

[97] Aneesur Rahman. Correlations in the motion of atoms in liquid argon. *Physical review*, 136(2A):A405, 1964.

[98] Giovanni Bussi, Davide Donadio, and Michele Parrinello. Canonical sampling through velocity rescaling. *The Journal of chemical physics*, 126(1):014101, 2007.

[99] Hans C Andersen. Molecular dynamics simulations at constant pressure and/or temperature. *The Journal of chemical physics*, 72(4):2384–2393, 1980.

[100] Herman JC Berendsen, JPM van Postma, Wilfred F van Gunsteren, ARHJ DiNola, and Jan R Haak. Molecular dynamics with coupling to an external bath. *The Journal of chemical physics*, 81(8):3684–3690, 1984.

[101] Shūichi Nosé. A molecular dynamics method for simulations in the canonical ensemble. *Molecular physics*, 52(2):255–268, 1984.

[102] William G Hoover. Canonical dynamics: Equilibrium phase-space distributions. *Physical review A*, 31(3):1695, 1985.

[103] Shuichi Nosé and ML Klein. Constant pressure molecular dynamics for molecular systems. *Molecular Physics*, 50(5):1055–1076, 1983.

[104] William G Hoover. Constant-pressure equations of motion. *Physical Review A*, 34(3):2499, 1986.

[105] Svante Arrhenius. Über die reaktionsgeschwindigkeit bei der inversion von rohrzucker durch säuren. *Zeitschrift für physikalische Chemie*, 4(1):226–248, 1889.

[106] Arthur F Voter. Hyperdynamics: Accelerated molecular dynamics of infrequent events. *Physical Review Letters*, 78(20):3908, 1997.

[107] Jooyoung Lee, Harold A Scheraga, and Shelly Rackovsky. New optimization method for conformational energy calculations on polypeptides: conformational space annealing. *Journal of computational chemistry*, 18(9):1222–1232, 1997.

[108] Alessandro Laio and Michele Parrinello. Escaping free-energy minima. *Proceedings of the National Academy of Sciences*, 99(20):12562–12566, 2002.

[109] Glenn M Torrie and John P Valleau. Nonphysical sampling distributions in monte carlo free-energy estimation: Umbrella sampling. *Journal of Computational Physics*, 23(2):187–199, 1977.

[110] Mads R So/rensen and Arthur F Voter. Temperature-accelerated dynamics for simulation of infrequent events. *The Journal of Chemical Physics*, 112(21):9599–9606, 2000.

[111] Robert H Swendsen and Jian-Sheng Wang. Replica monte carlo simulation of spin-glasses. *Physical review letters*, 57(21):2607, 1986.

[112] Yuji Sugita and Yuko Okamoto. Replica-exchange molecular dynamics method for protein folding. *Chemical physics letters*, 314(1-2):141–151, 1999.

[113] J. Schlitter, M. Engels, and P. Krüger. Targeted molecular dynamics: A new approach for searching pathways of conformational transitions. *Journal of Molecular Graphics*, 12(2):84 – 89, 1994.

[114] Jürgen Schlitter, Wolfgang Swegat, and Thomas Mülders. Distance-type reaction coordinates for modelling activated processes. *Molecular modeling annual*, 7(6):171–177, 2001.

[115] T Mülders, P Krüger, W Swegat, and J Schlitter. Free energy as the potential of mean constraint force. *The Journal of chemical physics*, 104(12):4869–4870, 1996.

[116] Matthias Post, Steffen Wolf, and Gerhard Stock. Principal component analysis of nonequilibrium molecular dynamics simulations. *The Journal of Chemical Physics*, 150(20):204110, 2019.

[117] Christopher Jarzynski. Nonequilibrium equality for free energy differences. *Physical Review Letters*, 78(14):2690, 1997.

[118] Marc Souaille and Benoıt Roux. Extension to the weighted histogram analysis method: combining umbrella sampling with free energy calculations. *Computer physics communications*, 135(1):40–57, 2001.

[119] Artem R Oganov, Andriy O Lyakhov, and Mario Valle. How evolutionary crystal structure prediction works and why. *Accounts of chemical research*, 44(3):227–237, 2011.

[120] Artem R Oganov and Colin W Glass. Crystal structure prediction using ab initio evolutionary techniques: Principles and applications. *The Journal of chemical physics*, 124(24):244704, 2006.

[121] Andriy O Lyakhov, Artem R Oganov, Harold T Stokes, and Qiang Zhu. New developments in evolutionary structure prediction algorithm uspex. *Computer Physics Communications*, 184(4):1172–1182, 2013.

[122] Andriy O. Lyakhov, Artem R. Oganov, and Mario Valle. How to predict very large and complex crystal structures. *Computer Physics Communications*, 181(9):1623–1632, 2010.

[123] Michael Sicher, Stephan Mohr, and Stefan Goedecker. Efficient moves for global geometry optimization methods and their application to binary systems. *The Journal of chemical physics*, 134(4):044106, 2011.

[124] Guang-Rui Qian, Xiao Dong, Xiang-Feng Zhou, Yongjun Tian, Artem R. Oganov, and Hui-Tian Wang. Variable cell nudged elastic band method for studying solid–solid structural phase transitions. *Computer Physics Communications*, 184(9):2111 – 2118, 2013.

[125] Ali Sadeghi, S Alireza Ghasemi, Bastian Schaefer, Stephan Mohr, Markus A Lill, and Stefan Goedecker. Metrics for measuring distances in configuration spaces. *The Journal of chemical physics*, 139(18):184118, 2013.

[126] Fabio Pietrucci and Wanda Andreoni. Graph theory meets ab initio molecular dynamics: atomic structures and transformations at the nanoscale. *Physical review letters*, 107(8):085504, 2011.

[127] Matthias Rupp, Alexandre Tkatchenko, Klaus-Robert Müller, and O Anatole Von Lilienfeld. Fast and accurate modeling of molecular atomization energies with machine learning. *Physical review letters*, 108(5):058301, 2012.

[128] Katja Hansen, Franziska Biegler, Raghunathan Ramakrishnan, Wiktor Pronobis, O Anatole Von Lilienfeld, Klaus-Robert Müller, and Alexandre Tkatchenko. Machine learning predictions of molecular properties: Accurate many-body potentials and nonlocality in chemical space. *The journal of physical chemistry letters*, 6(12):2326–2331, 2015.

[129] Jörg Behler. Atom-centered symmetry functions for constructing high-dimensional neural network potentials. *The Journal of chemical physics*, 134(7):074106, 2011.

[130] Matthew Hirn, Stéphane Mallat, and Nicolas Poilvert. Wavelet scattering regression of quantum chemical energies. *Multiscale Modeling & Simulation*, 15(2):827–863, 2017.

[131] Mostafa Yaghoobi and Mojtaba Alaei. Machine learning for compositional disorder: A comparison between different descriptors and machine learning frameworks. *Computational Materials Science*, 207:111284, 2022.

[132] Albert P Bartók, Risi Kondor, and Gábor Csányi. On representing chemical environments. *Physical Review B*, 87(18):184115, 2013.

[133] Sandip De, Albert P Bartók, Gábor Csányi, and Michele Ceriotti. Comparing molecules and solids across structural and alchemical space. *Physical Chemistry Chemical Physics*, 18(20):13754–13769, 2016.

[134] Sergio Speziale, Hans Josef Reichmann, FR Schilling, HR Wenk, and PJM Monteiro. Determination of the elastic constants of portlandite by brillouin spectroscopy. *Cement and Concrete Research*, 38(10):1148–1153, 2008.

[135] A Irabien, F Cortabitarte, J Viguri, and MI Ortiz. Kinetic model for desulfurization at low temperatures using calcium hydroxide. *Chemical engineering science*, 45(12):3427–3433, 1990.

[136] Shin-Min Shih, Ch'un-Sung Ho, Yuen-Sheng Song, and Jyh-Ping Lin. Kinetics of the reaction of ca (oh) 2 with co2 at low temperature. *Industrial & engineering chemistry research*, 38(4):1316–1322, 1999.

[137] Dario T Beruto and Rodolfo Botter. Liquid-like h2o adsorption layers to catalyze the ca (oh) 2/co2 solid–gas reaction and to form a non-protective solid product layer at 20° c. *Journal of the European Ceramic Society*, 20(4):497–503, 2000.

[138] RM Dheilly, J Tudo, Y Sebaibi, and M Queneudec. Influence of storage conditions on the carbonation of powdered ca (oh) 2. *Construction and building materials*, 16(3):155–161, 2002.

[139] AM Kalinkin, EV Kalinkina, OA Zalkind, and TI Makarova. Chemical interaction of calcium oxide and calcium hydroxide with co 2 during mechanical activation. *Inorganic Materials*, 41(10):1073–1079, 2005.

[140] G Montes-Hernandez, Antoine Pommerol, F Renard, P Beck, E Quirico, and O Brissaud. In situ kinetic measurements of gas–solid carbonation of ca (oh) 2 by using an infrared microscope coupled to a reaction cell. *Chemical Engineering Journal*, 161(1-2):250–256, 2010.

[141] Min Xu, Xiulan Huai, and Jun Cai. Agglomeration behavior of calcium hydroxide/calcium oxide as thermochemical heat storage material: a reactive molecular dynamics study. *The Journal of Physical Chemistry C*, 121(5):3025–3033, 2017.

[142] Chris L Page. Initiation of chloride-induced corrosion of steel in concrete: role of the interfacial zone. *Materials and corrosion*, 60(8):586–592, 2009.

[143] DM Henderson and HS Gutowsky. A nuclear magnetic resonance determination of the hydrogen positions in ca (oh) 2. *American Mineralogist: Journal of Earth and Planetary Materials*, 47(11-12):1231–1251, 1962.

[144] Y Aierken, H Sahin, Fadıl Iyikanat, S Horzum, A Suslu, B Chen, Ramazan Tugrul Senger, Sefaattin Tongay, and François M Peeters. Portlandite crystal: Bulk, bilayer, and monolayer structures. *Physical Review B*, 91(24):245413, 2015.

[145] T Nagai, T Ito, T Hattori, and T Yamanaka. Compression mechanism and amorphization of portlandite, ca (oh) 2: structural refinement under pressure. *Physics and Chemistry of Minerals*, 27(7):462–466, 2000.

[146] Romain Dupuis, Jorge S Dolado, Magali Benoit, Jose Surga, and Andrés Ayuela. Quantum nuclear dynamics of protons within layered hydroxides at high pressure. *Scientific reports*, 7(1):1–7, 2017.

[147] Kurt Leinenweber, Dan E Partin, Udo Schuelke, Michael O'Keeffe, and Robert B Von Dreele. The structure of high pressure ca (od) 2ii from powder neutron diffraction: Relationship to the zro2and eui2structures. *Journal of Solid State Chemistry*, 132(2):267–273, 1997.

[148] Riko Iizuka, Takehiko Yagi, Kazuki Komatsu, Hirotada Gotou, Taku Tsuchiya, Keiji Kusaba, and Hiroyuki Kagi. Crystal structure of the high-pressure phase of calcium hydroxide, portlandite:

In situ powder and single-crystal x-ray diffraction study. *American Mineralogist*, 98(8-9):1421–1428, 2013.

[149] Sen Shao, Jingkai Bi, Pengyue Gao, Guangtao Liu, Mi Zhou, Jian Lv, Yu Xie, and Yanchao Wang. Stability of ca (oh) 2 at earth's deep lower mantle conditions. *Physical Review B*, 104(1):014107, 2021.

[150] MB Kruger, Q Williams, and R Jeanloz. Vibrational spectra of mg (oh) 2 and ca (oh) 2 under pressure. *The Journal of Chemical Physics*, 91(10):5910–5915, 1989.

[151] Charles Meade and Raymond Jeanloz. Static compression of ca (oh) 2 at room temperature: observations of amorphization and equation of state measurements to 10.7 gpa. *Geophysical Research Letters*, 17(8):1157–1160, 1990.

[152] M Kunz, Kurt Leinenweber, JB Parise, T-C Wu, WA Bassett, K Brister, DJ Weidner, MT Vaughan, and Y Wang. The baddeleyite-type high pressure phase of ca (oh) 2. *International Journal of High Pressure Research*, 14(4-6):311–319, 1996.

[153] Krystle Catalli, S-H Shim, and VB Prakapenka. A crystalline-to-crystalline phase transition in ca (oh) 2 at 8 gpa and room temperature. *Geophysical Research Letters*, 35(5), 2008.

[154] Riko Iizuka, Kazuki Komatsu, Hiroyuki Kagi, Takaya Nagai, Asami Sano-Furukawa, Takanori Hattori, Hirotada Gotou, and Takehiko Yagi. Phase transitions and hydrogen bonding in deuterated calcium hydroxide: High-pressure and high-temperature neutron diffraction measurements. *Journal of Solid State Chemistry*, 218:95–102, 2014.

[155] Diana Denice, A Arya, Manoj Kumar, and Gopika Vinod. First-principles study of electronic, cohesive and elastic properties of silica polymorphs. *Materials Today Communications*, 31:103607, 2022.

[156] John Trail, Bartomeu Monserrat, Pablo López Ríos, Ryo Maezono, and Richard J Needs. Quantum monte carlo study of the energetics of the rutile, anatase, brookite, and columbite tio 2 polymorphs. *Physical Review B*, 95(12):121108, 2017.

[157] Georg Kresse and Jürgen Hafner. Ab initio molecular dynamics for liquid metals. *Physical review B*, 47(1):558, 1993.

[158] Georg Kresse and Jürgen Furthmüller. Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set. *Computational materials science*, 6(1):15–50, 1996.

[159] Péter Pulay. Convergence acceleration of iterative sequences. the case of scf iteration. *Chemical Physics Letters*, 73(2):393–398, 1980.

[160] Xiang-Feng Zhou, Guang-Rui Qian, Xiao Dong, Lixin Zhang, Yongjun Tian, Hui-Tian Wang, et al. Ab initio study of the formation of transparent carbon under pressure. *Physical Review B*, 82(13):134126, 2010.

[161] RF Zhang, SH Zhang, YQ Guo, ZH Fu, Dominik Legut, Timothy Clark Germann, and S Veprek. First-principles design of strong solids: Approaches and applications. *Physics Reports*, 826:1–49, 2019.

[162] Ellina Bernard, Barbara Lothenbach, Christophe Chlique, Mateusz Wyrzykowski, Alexandre Dauzeres, Isabelle Pochard, and Céline Cau-Dit-Coumes. Characterization of magnesium silicate hydrate (msh). *Cement and concrete research*, 116:309–330, 2019.

[163] Philippe Ferrara, Joannis Apostolakis, and Amedeo Caflisch. Computer simulations of protein folding by targeted molecular dynamics. *Proteins: Structure, Function, and Bioinformatics*, 39(3):252–260, 2000.

[164] Xiandong Cong and R James Kirkpatrick. 29si mas nmr study of the structure of calcium silicate hydrate. *Advanced Cement Based Materials*, 3(3-4):144–156, 1996.

[165] Andrés Ayuela, Jorge S Dolado, I Campillo, YR De Miguel, Edurne Erkizia, Daniel Sánchez-Portal, Angel Rubio, A Porro, and Pedro M Echenique. Silicate chain formation in the nanostructure of cement-based materials. *The Journal of chemical physics*, 127(16):164710, 2007.

[166] Jie Zhu, Dejian Shen, Wei Wu, Baosheng Jin, and Shengxing Wu. Hydration inhibition mechanism of gypsum on tricalcium aluminate from reaxff molecular dynamics simulation and quantum chemical calculation. *Molecular Simulation*, 47(17):1465–1476, 2021.

[167] Alexei I Abrikosov, Goran Kovačević, and Valera Veryazov. Multi-scale study of crystal and electronic structure of al defects in concrete. In *AIP Conference Proceedings*, volume 2040, page 020007. AIP Publishing LLC, 2018.

[168] Steve Plimpton. Fast parallel algorithms for short-range molecular dynamics. *Journal of computational physics*, 117(1):1–19, 1995.

[169] Denis J Evans and Brad Lee Holian. The nose–hoover thermostat. *The Journal of chemical physics*, 83(8):4069–4074, 1985.

[170] Adri CT Van Duin, Siddharth Dasgupta, Francois Lorant, and William A Goddard. Reaxff: a reactive force field for hydrocarbons. *The Journal of Physical Chemistry A*, 105(41):9396–9409, 2001.

[171] Adri CT Van Duin, Alejandro Strachan, Shannon Stewman, Qingsong Zhang, Xin Xu, and William A Goddard. Reaxffsio reactive force field for silicon and silicon oxide systems. *The Journal of Physical Chemistry A*, 107(19):3803–3811, 2003.

[172] Hegoi Manzano, Roland JM Pellenq, Franz-Josef Ulm, Markus J Buehler, and Adri CT van Duin. Hydration of calcium oxide surface predicted by reactive force field molecular dynamics. *Langmuir*, 28(9):4187–4197, 2012.

[173] Hegoi Manzano, Sina Moeini, Francis Marinelli, Adri CT Van Duin, Franz-Josef Ulm, and Roland J-M Pellenq. Confined water dissociation in microporous defective silicates: mechanism, dipole distribution, and impact on substrate properties. *Journal of the American Chemical Society*, 134(4):2208–2215, 2012.

[174] Bernhard Schölkopf, Alexander Smola, and Klaus-Robert Müller. Kernel principal component analysis. In *International conference on artificial neural networks*, pages 583–588. Springer, 1997.

[175] Benjamin A Helfrecht, Rose K Cersonsky, Guillaume Fraux, and Michele Ceriotti. Structure-property maps with kernel principal covariates regression. *Machine Learning: Science and Technology*, 1(4):045021, 2020.

[176] Mehrnoosh Arrar, Fernando Martín Boubeta, Maria Eugenia Szretter, Mariela Sued, Leonardo Boechi, and Daniela Rodriguez. On the accurate estimation of free energies using the jarzynski equality. *Journal of Computational Chemistry*, 40(4):688–696, 2019.

[177] Glenn M Torrie and John P Valleau. Nonphysical sampling distributions in monte carlo free-energy estimation: Umbrella sampling. *Journal of Computational Physics*, 23(2):187–199, 1977.

[178] Johannes Kästner. Umbrella sampling. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 1(6):932–942, 2011.

[179] S Börjesson, A Emrén, and C Ekberg. A thermodynamic model for the calcium silicate hydrate gel, modelled as a non-ideal binary solid solution. *Cement and Concrete Research*, 27(11):1649–1657, 1997.

[180] Josep M Soler. Thermodynamic description of the solubility of csh gels in hydrated portland cement. literature review. 2007.

[181] Jost Eikenberg. On the problem of silica solubility at high ph. Technical report, Paul Scherrer Inst.(PSI), 1990.

[182] Kenneth R Applin. The diffusion of dissolved silica in dilute aqueous solution. *Geochimica et Cosmochimica Acta*, 51(8):2147–2151, 1987.

[183] Jeremy Haas and André Nonat. From c–s–h to c–a–s–h: Experimental study and thermodynamic modelling. *Cement and Concrete Research*, 68:124–138, 2015.

[184] Colin S. Walker, Shunkichi Sutou, Chie Oda, Morihiro Mihara, and Akira Honda. Calcium silicate hydrate (c-s-h) gel solubility data and a discrete solid phase model at 25°c based on two binary non-ideal solid solutions. *Cement and Concrete Research*, 79:1–30, 2016.

[185] Fredrik P Glasser, Eric E Lachowski, and Donald E Macphee. Compositional model for calcium silicate hydrate (c-s-h) gels, their solubilities, and free energies of formation. *Journal of the American Ceramic Society*, 70(7):481–485, 1987.

[186] Miho Tanaka and Kazuya Takahashi. The identification and characterization of silicate complexes in calcium chloride solution using fast atom bombardment mass spectrometry. *Analytica chimica acta*, 411(1-2):109–119, 2000.

[187] Miho Tanaka and Kazuya Takahashi. Characterization of silicate monomer with sodium, calcium and strontium but not with lithium and magnesium ions by fast atom bombardment mass spectrometry. *Journal of mass spectrometry*, 37(6):623–630, 2002.

[188] Bence Kutus, Attila Gácsi, Attila Pallagi, István Pálinkó, Gábor Peintler, and P Sipos. A comprehensive study on the dominant formation of the dissolved ca (oh) 2 (aq) in strongly alkaline solutions saturated by ca (ii). *RSC advances*, 6(51):45231–45240, 2016.

[189] Ján Šefčík and Alon V McCormick. Thermochemistry of aqueous silicate solution precursors to ceramics. *AIChE Journal*, 43(S11):2773–2784, 1997.

[190] Julian D Gale. Gulp: A computer program for the symmetry-adapted simulation of solids. *Journal of the Chemical Society, Faraday Transactions*, 93(4):629–637, 1997.

[191] Stefan Grimme. Semiempirical gga-type density functional constructed with a long-range dispersion correction. *Journal of computational chemistry*, 27(15):1787–1799, 2006.

[192] Michele Ceriotti, Gareth A Tribello, and Michele Parrinello. Simplifying the representation of complex free-energy landscapes using sketch-map. *Proceedings of the National Academy of Sciences*, 108(32):13023–13028, 2011.

[193] JD Kubicki, Y Xiao, and AC Lasaga. Theoretical reaction pathways for the formation of [si (oh) 5] 1- and the deprotonation of orthosilicic acid in basic solution. *Geochimica et cosmochimica acta*, 57(16):3847–3853, 1993.

[194] Xabier M Aretxabaleta, Jon López-Zorrilla, Christophe Labbez, Iñigo Etxebarria, and Hegoi Manzano. A potential csh nucleation mechanism: atomistic simulations of the portlandite to csh transformation. *Cement and Concrete Research*, 162:106965, 2022.

[195] Ana Cuesta, Isabel Santacruz, G Angeles, Monica Dapiaggi, Jesus D Zea-Garcia, and Miguel AG Aranda. Local structure and ca/si ratio in csh gels from hydration of blends of tricalcium silicate and silica fume. *Cement and Concrete Research*, 143:106405, 2021.

[196] Elizaveta Pustovgar, Rahul P Sangodkar, Andrey S Andreev, Marta Palacios, Bradley F Chmelka, Robert J Flatt, and Jean-Baptiste d'Espinose de Lacaillerie. Understanding silicate hydration from quantitative analyses of hydrating tricalcium silicates. *Nature communications*, 7(1):1–9, 2016.

[197] Isabel Galan, Bernhard Müller, Lukas G Briendl, Florian Mittermayr, Torsten Mayr, Martin Dietzel, and Cyrill Grengg. Continuous optical in-situ ph monitoring during early hydration of cementitious materials. *Cement and Concrete Research*, 150:106584, 2021.

[198] Paolo Raiteri and Julian D Gale. Water is the key to nonclassical nucleation of amorphous calcium carbonate. *Journal of the American Chemical Society*, 132(49):17623–17634, 2010.

[199] JD Bernal. The structures of cement hydration compounds. In *Proc. 3rd Int. Symp. Chem. Cem., London: Cement and Concrete Association*, pages 216–236, 1952.

[200] JD Bernal, JW Jeffery, and HFW Taylor. Crystallographic research on the hydration of portland cement. a first report on investigations in progress. *Magazine of Concrete research*, 4(11):49–54, 1952.

[201] LG Shpynova, NABITOVI. ID, and NV Belov. Microstructure of alite cement stone (hydrated tricalcium silicate). *Soviet Physics Crystallography, USSR*, 11(6):747, 1967.