Aslam Kunhi Mohamed, Stephen C. Parker, Paul Bowen, Sandra Galmarini
Laboratoire de Technologie des Poudres, Ecole Polytechnique Fédérale de Lausanne, Switzerland
Computational Solid State Chemistry Group, Department of Chemistry, University of Bath, UK
Building Energy Materials and Components, Empa, Switzerland

Abstract

Understanding the atomistic structure of Calcium silicate hydrate (C-S-H), the main product of cement hydration, is of paramount importance to better formulate sustainable cement. The existing atomistic models are not in total agreement with experimental results and fail to explain the nanosized nature of C-S-H. Here, we present a new approach for describing the complexity of these structures at the molecular level, enabling a detailed comparison of C-S-H models. The new methodology encodes a full, large scale atomistic C-S-H structure by a simple, readable string of characters, similar to the way the base sequence in DNA encodes a vast range of different proteins. We then use the methodology to assess 14 Å tobermorite-defect structures and their stabilities using DFT and classical molecular dynamics. Finally we highlight how the model may be extended to develop reliable atomistic C-S-H models for a range of Ca/Si ratios and conditions.

Keywords: Crystal Structure, Cementitious Systems, Molecular Dynamics, Density Functional Theory, C-S-H

1. Introduction

Hydrated cement is the binder or glue that holds together the sand and aggregates that all together make up concrete - the most used material in the world. Although per kg cement is inherently a low carbon material [1, 2], the vast quantities used mean that cement is responsible for about 5-8% of global CO₂ emissions. The optimisation of the mechanical properties and durability of concrete can therefore aid in the reduction of CO₂ emissions. This would be greatly aided by a better fundamental knowledge of the material at the atomistic structure level.

A better understanding of cementitious systems is complicated by the fact that the exact atomistic and nanoscale structure of the main hydration product (C-S-H) is poorly known. One reason for this is that both X-ray diffraction...
Figure 1: Ca-Si layer of 14 Å tobermorite with characteristic linear dreierketten structure of silicate chains ((a) and (b)) as well as their stacking in 14 Å tobermorite ((c) and (d)). Si: blue, Ca: turquoise, O: red, OH: black. Also shown is the structural abbreviation introduced in section 3.1.1.

(XRD) and selected area diffraction show only two weak signals, indicating repeat distances of about $\sim 1.82 - 1.83$ Å and $\sim 2.7 - 3.2$ Å. This implies that C-S-H does not have any 3D long range order. It also means that the exact atomistic structure of C-S-H cannot be resolved by conventional XRD techniques.

A further complication is the varying chemical composition of C-S-H. The composition varies significantly not only between different samples but also within one and the same sample. In neat tricalcium silicate ($C_3S$) pastes local energy dispersive X-ray analysis (EDS) has shown that on a scale of 100 nm $\frac{Ca}{Si}$ varies between 1.2 and 2.1, with an average calcium to silica ratio $\langle \frac{Ca}{Si} \rangle$ of 1.75.

The type of silicate species present in C-S-H was first studied by Mohan et al. with trimethylsilylation followed by chromatography. This technique revealed that only silica chains with a length of $3n - 1$ (i.e. dimers, pentamers, octamers and some longer chains) are present in C-S-H. This indicates a so called dreierketten structure with silicate dimers that can be linked by bridging silicate tetrahedra (see figure). Subsequent NMR studies revealed, that no connectivity larger than 2 exists between the different silicate chains. The ratio between the two in C-S-H being 2:1. This confirms the dreierketten structure where only chains of length $3n - 1$ can exist.

Additionally, transmission electron microscopy (TEM) images of carefully dried samples show different morpholo-
gies of C-S-H, often with a pronounced texture [4]. The smallest characteristic dimension of C-S-H, corresponding to both the size of the solid regions [4, 14, 15] and of the crystallites [16, 17] is in the order of 3-6 nm. The largest characteristic dimension is less clear. For the crystalline domains in OPC the largest characteristic dimension is typically in the order of 6-15 nm [16, 17], however some order can persist over much longer length scales of >100 nm [4, 18], which could be either due to a partially continuous structure [19] or to a colloidal material [20] with long length scale ordering similar to that observed in e.g. liquid crystals. TEM images with an even higher resolution have revealed layering of the material with an approximate interlayer distance of ~ 12-14 Å [21].

Different water containing, layered calcium-silicate crystal structures containing linear silicate-"dreierketten" have been reported in the literature. The ones most commonly compared to C-S-H are the different tobermorites (e.g. 14 Å tobermorite: \( \frac{Ca}{Si} = 0.83 \), interlayer distance 14 Å, molar water content: 42 %) and jennite (\( \frac{Ca}{Si} = 1.5 \), molar water content: 42 %). Both 14 Å tobermorite and jennite consist of calcium-silicate layers separated by an interlayer space containing both water and calcium. The main layer calcium ions in 14 Å tobermorite form a planar sheet, the calcium layer in the jennite structure is corrugated.

However the calcium to silicon ratio \( Ca/Si \) of all these minerals is significantly smaller than that observed for C-S-H. An indication as to what structural trends lead to an increase of the \( Ca/Si \) ratio in C-S-H can be gleaned from experimental observations. In C-S-H produced from tricalcium silicate (\( C_3S \)) pastes cured for 12 months, about 60 % of the silicate species are present as dimers and 40 % as longer chains [6] with a mean chain length (MCL) of 3.3. Similar results are found for ordinary Portland cement paste (OPC)[4]. The chain length continues to increase over the next couple of decades but even for ~ 20 year old pastes there are still a lot of dimers (~ 40 %[6]) and the MCL is still less than 5 [4]. This means that, compared to the model structures with infinite silicate chains, a significant part of the bridging silicate tetrahedra are missing in C-S-H, which will increase the \( Ca/Si \).

Additionally both \(^1H – ^29Si\) NMR measurements [8] and infrared spectroscopy [22] indicate the existence of Ca-OH groups in C-S-H. Thomas et al. used inelastic neutron scattering to quantify the amount of Ca-OH bonds in C-S-H [23]. They found that at a calcium to silicon ratio of 1.7 about 23 % of the charge of the \( Ca^{2+} \) ions in C-S-H are compensated by hydroxyl groups. This value is in between that of tobermorite (no \( Ca – OH \) groups) and that of jennite (33 %). Thus these results indicate either a mix between jennite and tobermorite or additional \( Ca – OH \) groups incorporated in a tobermorite like structure would be needed to give the appropriate \( Ca/Si \) ratio. Additionally the number of Si-OH
groups decreases with increasing $Ca/Si$. In contrast to a non-defective 14 Å tobermorite, the number of silanol groups in a C-S-H with a $Ca/Si$ ratio of 1.7 is below the detection limit of different experimental techniques [22][11].

While these results would indicate a closer resemblance of C-S-H with a typical $Ca/Si$ ratio of 1.7 to jennite than to tobermorite, structural measurements such as XRD and Raman spectroscopy [24][16][25] as well as theoretical work [26] indicate a closer resemblance of the C-S-H structure to tobermorite rather than to jennite.

In summary, from these results we infer that C-S-H is composed of defective, nanocrystalline tobermorite with missing bridging silicate tetrahedra, leading to a decreased silicate chain length, and deprotonated silanol groups, the charge of which is compensated by additional calcium ions in the water-interlayer and with additional Ca-OH groups. Furthermore, experimental measurements revealed that the Ca-O and Si-O nearest neighbour distances in C-S-H with $Ca/Si$ up to 1.75 are similar to the ones in tobermorite while the Ca-O coordination number decreases almost linearly with increasing $Ca/Si$ (6.0±0.2 at $Ca/Si=1.75$, as compared to 7.1±0.2 for $Ca/Si=0.6$ and 7.0±0.2 at $Ca/Si=0.8$) [27][28]. The approximate stoichiometry is estimated to be $Ca_{1.7}SiO_{3.7} \cdot 1.8H_2O$ with a density of about 2.6 $g cm^{-3}$, excluding the water at the surface [29][15]. XRD measurements indicate that the interlayer distance may be smaller than 14 Å in C-S-H with $Ca/Si=1.7$ (around 11 or 12 Å [16][11][27]).

An atomistic model should take into account all the experimental evidence presented above. Additionally it should give some indications as to why C-S-H displays the rather unique properties it does: How can we explain the consistently nanoparticulate nature of C-S-H? And why does the stoichiometry vary extensively within one and the same C-S-H sample? In recent years, several atomistic models for C-S-H have been proposed [30][21][31][32][33]. Despite capturing some of the salient features of C-S-H, none of the models are entirely satisfactory, when confronted with the experimental evidence available and they leave several questions regarding the properties and behavior of C-S-H unresolved.

Recently, Pellenq et al. [21][31] and Kovačević et al. [32][34] used the Hamid [35] and the Merlino [36] 11 Å tobermorite respectively as the base structure to model C-S-H using classical molecular dynamics simulations. In order to arrive at the required Ca/Si ratio, Pellenq et al. randomly removed $SiO_2$ groups and started with dry tobermorite structure with subsequent water adsorption using Grand Canonical Monte Carlo. This resulted in a very high percentage of monomers and the presence of water within the Ca-Si layers, both inconsistent with the experimental results available. Other limitations of their model were the absence of hydroxyl groups as well as bond lengths and
coordination numbers that are inconsistent with natural calcium-silicate mineral structures [33]. The authors have tried to address some of these shortcomings in a recent work using the ReaxFF potential to split water molecules, making it possible to have hydroxyl groups and thus forming Ca-OH [31]. However the hydrogen of the split water molecules mainly reacted with non-bonding oxygen of the silicate chains, forming protonated silanol groups which according to experimental results is found in very small or negligible amounts at high Ca/Si ratios [11, 22]. Also, in order to eliminate the unexpectedly high concentration of silicate monomers as observed by Pellenq et al. they allowed for (random) polymerisation of the silicates in their revised model structures. However this leads to a silicate chain structure that is in violation of the dreierketten pattern that has been shown to exist in C-S-H, with only chain lengths of 3n − 1 being observed [6, 7, 8, 9, 10, 11].

Kovačević et al. [32, 34] explored three different ways to arrive at a Ca/Si ratio of 1.65 from the Merlino 11 Å tobermorite starting structure [36]. Namely (1) removal of only the bridging silicate tetrahedra and addition of Ca(OH)$_2$, (2) removal of bridging silicate as well as dimers and (3) random removal of all the silicate species as done by Pellenq et al. [31]. This enabled them to compare the energies, calculated via energy minimization [32] and molecular dynamics [34], of many different C-S-H structures with respect to the way the Ca/Si ratio was increased. The authors concluded that the structures generated using the first method, removing only the bridging tetrahedra and adding Ca(OH)$_2$ in the interlayer space are the most stable. The authors conclude that the way of constructing the initial structure has a significant effect on the energy and final properties of the C-S-H model structures. This means that an in-depth discussion about the different ways to construct C-S-H models is needed in order to be able to construct better models leading to a better understanding of C-S-H.

Finally Richardson [33] derived a fully crystalline model for C-S-H based solely on measured XRD structures of different calcium-silicate structures. The model consists of different perfectly ordered layers with different Ca/Si ratios. General guidelines for the type of defects present as well as geometrical reasoning was used to construct these. While the guidelines developed are very valuable for future work, a perfectly ordered bulk structure seems to be unlikely and makes it difficult to understand the nanoparticulate nature of C-S-H.

The aim of the current paper is to thus develop an atomistic structure for C-S-H that can describe the nanostructuring, facilitate the comparison of models and provide some new insights into C-S-H using atomistic modelling. Hence in section 3.1 a building block description of C-S-H is developed, which is able to describe both different base structures...
and different Ca/Si increasing defects. This method of describing C-S-H facilitates the development and construction of new model structures and the comparison of existing models. It allows the fast calculation of a range of characteristics (charge distribution, mean chain length, percentage of dimers, ratio of protonation of silanol groups, OH/Ca ratio). The method will be used to revisit different models reported in the literature. In the second section (3.3.2) the building block description together with both classical molecular dynamics (CMD) and density functional theory (DFT) is used to discuss 14Å tobermorite as another possible basis for C-S-H models by looking at different possible Ca/Si increasing defects, extending the range of the defects considered in the work reported in the literature to date.

It is our view that this novel building block description will significantly help in identifying a more precise atomistic structure of C-S-H and understanding its characteristics, allowing researchers to go one step further from proposing single models which are difficult to compare between each other. However before discussing the model in depth, we briefly describe the simulation methods.

2. Methods

2.1. Density Functional Theory Calculations

DFT calculations were performed using the VASP code [37, 38], in which a plane wave basis set describes the valence electronic states and the projector augmented wave (PAW) approach [39, 40] models the core-valence interaction. The exchange and correlation is treated with the generalized gradient approximation (GGA) via the Perdew-Burke-Ernzerhof (PBE) functional [41]. The inability of standard DFT to account for van der Waals interactions is well-known [42, 43]. In this work, we applied the DFT-D3 method of Grimme [44] which applies dispersion corrections in a pairwise manner. The Monkhorst-Pack k-points mesh 4 x 4 x 2 and a plane wave cut-off of 500 eV were used to ensure convergence. All DFT calculations used a primitive unit cell with half the volume of the original unit cell reported by Bonaccorsi et al. [45]. The cell parameters of 14 Å tobermorite, relaxed with DFT, can be found in table [1] and the simulated XRD patterns in the supplementary materials. Both are compared to the experimentally determined structure [45]. The XRD patterns were generated by the software Mercury [46] using a wavelength of 1.54 Å (Cu-Kα) varying from 5 to 50 degrees at an interval of 0.02 °. The DFT calculations with dispersion correction describe the 14 Å tobermorite structure with good accuracy, the maximum deviation on lattice cell parameters compared to experimental values being 2 % and 1 ° on the angles (see table [1] and figure S5 in the supplementary materials).
2.2. Classical Atomistic Simulations

Molecular dynamics calculations using classical force fields to describe the atom interactions used DL POLY 2.20 [47] and Metadise for energy minimisation [48]. The formal charge force field used for the description of the ions in solution and for the solid is based on an adaptation of Freeman et al. [49] for the use of the force field originally developed by Lewis and Catlow [50] and further developments by various authors [51, 52, 53, 54, 55, 56] for the use with the TIP family of force fields. For water, the TIP4P/2005 rigid model [57] is adopted. The full force field used here has been previously tested for a wide range of calcium-silicate-hydrate structures and an empirical estimation of the error on structural properties and energies has been developed [56]. The error estimations are used here to give an estimation on the accuracy of the results. The full list of parameters used can be found in [56]. Ewald summation [58] was used to estimate the long range electrostatic forces, with a cutoff radius between the long range and the short range domain of 8.5 Å. Energy minimisation calculations have been done under constant pressure (CONP). The time step for the molecular dynamics simulations was 0.7 fs.

A simulation cell of 4x3x1 full 14 Å unit cells (4x3x2 primitive unit cells as considered with DFT) has been used for molecular dynamics calculations. Due to the difference in size of the simulation cells, comparison between DFT and MD is not always possible. The system has been relaxed for > 210 ps before data was collected over 700 ps. After introduction of isolated defects, the system has been first relaxed at zero temperature, followed by equilibration for 35 ps and data collection over a simulation period of 105 ps. The anisotropic NPT ensemble with flexible angles was used. To ensure that the equilibration and simulation period was enough, the simulation was extended to 1 ns for two defects where large relaxations were observed (second structure with Ca/Si ratio of 1.20 and structure with Ca/Si ratio 1.25 in table 2). No further structural relaxation was observed for either defect.

For additional validation, XRD patterns for 14 Å tobermorite were compared for the DFT relaxed structure and after molecular dynamics simulation using the force field potentials (see supplementary material section III for more details). The general structure is well captured and the interatomic distances remain within an acceptable mean deviation of ± 0.13 Å (or 6.4%) which is comparable with the estimated force field error (5% on interatomic distances [56]). Similarly the higher Ca/Si structures considered with DFT (see section 3.3.1) were subsequently energy minimised with the classical force field the resulting simulated XRD spectra compared to the one of the DFT structure, the spectra showed good consistency as shown in the supplementary materials.
Table 1: Unit cell parameters of 14 Å tobermorite calculated with different techniques (molecular dynamics (MD) and density functional theory (DFT)) compared to the experimental values. Errors on MD from [56].

<table>
<thead>
<tr>
<th></th>
<th>a</th>
<th>b</th>
<th>c</th>
<th>α</th>
<th>β</th>
<th>γ</th>
</tr>
</thead>
<tbody>
<tr>
<td>MD [56]</td>
<td>6.85 ± 0.34</td>
<td>7.26 ± 0.36</td>
<td>27.95 ± 1.40</td>
<td>90 ± 10</td>
<td>93 ± 10</td>
<td>123 ± 10</td>
</tr>
<tr>
<td>DFT&lt;sub&gt;vdW&lt;/sub&gt;</td>
<td>6.63</td>
<td>7.37</td>
<td>28.52</td>
<td>89</td>
<td>91</td>
<td>122</td>
</tr>
<tr>
<td>Exp. [45]</td>
<td>6.74</td>
<td>7.43</td>
<td>27.99</td>
<td>90</td>
<td>90</td>
<td>123.25</td>
</tr>
</tbody>
</table>

3. Results and Discussion

The results and discussion part is divided into three sections. In the first section a building block description facilitating the comparison and construction of models is developed. In the second section the formalism is used to examine atomistic structures previously published in the literature. In the final part the building block description is then used to look at different possible 14 Å tobermorite defects. The different types of defects likely to be present are discussed based on structural and energetic considerations. From the insights gained on different defects, a first step is made towards possible full scale models by presenting model structures. The aim here is not to present a final full scale model but rather to present a method on how future models can be constructed, tested and compared to previous models (e.g. those by Kovačević et al. [32] and Richardson [33]).

3.1. Model for Atomistic Structure Description of tobermorite based C-S-H

What is clear from the different atomistic simulation efforts of C-S-H is that it is unlikely that we can define the exact atomistic structure in the same way a crystalline structure is defined. Rather there will be a large number of possible local arrangements which will give rise to the overall C-S-H properties observed experimentally. This means that entropic effects are likely to be important. However, the first step to be able to include configurational entropy is an efficient way to correctly describe the full phase space of possible arrangements. Within the current paper we will limit ourselves to the description of tobermorite-like structures, although the model may be easily adapted to other structures.
3.1.1. Building Block description of 14 Å tobermorite

Our aim is to be able to describe a larger scale C-S-H structure, comprising of over a thousand atoms, in a simple, readable way. In crystals this is done by defining a unit cell. This unit cell, containing only a few atoms, if repeated in space, is able to represent a larger crystal. For C-S-H, assuming a certain underlying crystal structure, we can define a unit cell as for crystals. However the atoms in each unit cell will vary, meaning that we additionally need a method to specify the atomic arrangement in each consecutive unit cell. For this we will introduce a short-hand notation for the different groups of atoms. Each atom group or chemical unit has a defined stoichiometry and charge represented by a unique symbol containing information about the relative arrangement of the atoms. The symbols for the atom groups are then used to develop short strings which represent the different "defective" unit cells of tobermorite. The atomic structure within the unit cell will be given both by the symbols used and their position within the string. Finally, a sequence of such strings, each representing a unit cell, will be indicative of stacking units with different compositions and will thus describe a larger scale C-S-H structure. In other words these short strings form the building blocks with which to build different C-S-H models. We will first develop the method for 14 Å tobermorite based models, before discussing the extension to other possible base crystals.

We want to represent each unit cell with as small a repeat unit as possible, similar to that chosen for crystals. However, it is important that the limits of the unit cell are well defined and that chemical units are grouped together. Consequently, we choose a primitive crystallographic unit cell with a reduced volume and the following dimensions:

\[ a_{\text{red}}: 6.735(2) \ \text{Å}, \ b_{\text{red}}: 7.425(2) \ \text{Å}, \ c_{\text{red}}: 15.325(7) \ \text{Å}, \ \alpha: 111.34(7) ^\circ, \ \beta: 87.36(8) ^\circ, \ \gamma: 123.25(1) ^\circ \] (see figures 1 and 2). The position of the cell within the 14 Å tobermorite structure is chosen such that; the \[ a_{\text{red}}xb_{\text{red}} \] surface of the unit cell lies in the middle of the the calcium-silicate layers, the \[ a_{\text{red}}xc_{\text{red}} \] is such that the silicate chains are cut in the middle of the silicate dimers and the \[ b_{\text{red}}xc_{\text{red}} \] is between two adjacent linear silicate chains. Stacking of these primitive unit cells or building blocks will lead to the 14 Å tobermorite structure as described by [45] (see e.g. figure 1 of the supplementary materials). An atomistic configuration file for such a unit cell can be found in the supplementary materials.

Now the chemical units and the symbols representing them have to be defined (figure 2). Let us first consider the calcium-silicate layers of the tobermorite structure as shown in figure [1]. These layers contain two Ca-Si sheets of equal composition but different orientation, loosely related by a glide plane symmetry. The sheets themselves are
composed of linear silicate chains and a range of calcium ions. The silicate chains consist of dimers which are linked by a bridging silicate. Each silicate tetrahedra of the silicate dimer is associated with a calcium ion of the sheet. Let us define such an atom group, composed of a dimer silicate with its associated calcium and which has a stoichiometry of $CaSiO_{3.5}$ as one chemical unit. We will choose two different symbols for this unit depending on the relative position of the calcium with respect to the silica chain (above or below) and consequently on whether the unit belongs to the upper sheet of the calcium-silicate layer: $<$ will represent a dimer silicate and the associated calcium above that silicate and $>$ the equivalent unit with the calcium below the silicate (figure 2). We next represent the bridging silicate with a stoichiometry of $SiO_2$ and zero charge by an $S$, independent of which sheet it belongs to. Additionally we have to take into account the possible protonation of the silanol groups. Let us note a protonated silanol group by an apostrophe (‘), stoichiometrically representing an $H$ and a charge of $+1$.

To be able to describe defective structures as well, a few additional units have to be defined. Firstly, in some tobermorite structures the linear silicate chains are cross-linked via the bridging silica tetrahedra. A cross-linked silica tetrahedra is indicated by a vertical line to indicate the shared oxygen: $|S$. Such a cross-linked bridging silica has a stoichiometry of $SiO_{1.5}$ and a charge of $+1$. Finally it is possible to replace a dimer silicate $<$ or $>$ by a calcium hydroxide unit with stoichiometry $Ca(OH)_2$ and zero charge. For this replacing unit we will again distinguish the orientation and represent it by a $/ \text{if the calcium is above the hydroxyls and } \backslash \text{if the calcium is below. If only one of two paired } <$ is replaced, the second will no longer be part of a dimer but will be a monomer (or end-member of a chain). Such a monomer, with a stoichiometry of $CaSiO_4$ and a charge of $-2$, will be represented by $( \text{if the calcium is above and by } ) \text{if the calcium is below the silicate}$.

In addition, we have to define three interlayer species: calcium ions (stoichiometry $Ca$, charge $+2$) are represented by $C$ and hydroxyl ions (stoichiometry $OH$, charge $-1$) by $°$. Finally water molecules are represented by a number, indicating the total number of water molecules in the interlayer. The full overview of the symbols can be found in table 2.

Now that we have defined the symbols for each chemical group, we need to define the string of symbols defining the arrangement of these groups within the unit cell. To do so, we will specify the groups within the unit cell layer by layer first in $b_{\text{red}}$ direction (left to right in figure 2) then in $-c_{\text{red}}$ direction (top to bottom in figure 2). Thus, we will start with the calcium-silicate sheet above the interlayer. For a non-defective 14 Å tobermorite we have a dimeric
Figure 2: Definition and notation of different atom groups used for the description of different structures derived from 14 Å tobermorite (left) as well as a primitive unit cell $<$S$'$<C$>$S$'$ of 14 Å tobermorite with the following dimensions: $a_{red}$: 6.735(2) Å, $b_{red}$: 7.425(2) Å, $c_{red}$: 15.325(7) Å, $\alpha$: 111.34(7) °, $\beta$: 87.36(8) °, $\gamma$: 123.25(1) °. Si: blue. Ca: turquoise, O: red, H$_2$O: white, OH: black.

calcium silicate, a bridging silicate and a second dimeric calcium silicate in this sheet, where the calcium is above the silicate. Thus, we can represent it by $<$S$. There are two silanol groups on the bridging silicate $S$. The one to the right is protonated in 14 Å tobermorite, therefore, we will add a ‘ to the right of the $S$: $<$S$'$. This part of the string fully describes the Ca-Si sheet above the interlayer. Now we add the interlayer. Here, we have one calcium ion ($C$) and 7 water molecules ($\gamma$). Thus, the string for the top Ca-Si sheet and the interlayer becomes: $<$S$'<$C$>$T. Finally, the Ca-Si sheet below the interlayer is equivalent to the top sheet, except for the orientation of the dimeric species: $>$S$'$. This means that for a full unit cell of pristine 14 Å tobermorite our building block string is: $<$S$'<$C$>$T$'$. We can now consider different possible defects within 14 Å tobermorite. A simple defect would be to deprotonate two silanol groups and compensate the charge by an additional calcium ion in the interlayer: $<$S$'<$C$>$T$'$. Similarly, the inverse is also possible, leading to a structure with no calcium in the interlayer: $<$S$'<$C$>$T$'$. Additionally, we can cross-link the silicate chains via the silanol group to the left of the bridging $S$, again leading to Ca-Si layers without a charge and thus, no calcium in the interlayer: $<$S$'<$C$>$T$'$. However, the latter is only possible if we change the geometry of the unit cell, such that the cross-linking is geometrically possible (see section 3.1.4).

Another reported defect [4] is a missing bridging silicate: $<$S$'<$C$>$T$'$. If a silicate unit is missing,
Figure 3: Different possible defective building blocks of 14 Å tobermorite (dimensions: $a_{\text{red}}$: 6.735(2) Å, $b_{\text{red}}$: 7.425(2) Å, $c_{\text{red}}$: 15.325(7) Å, $\alpha$: 111.34(7) °, $\beta$: 87.36(8) °, $\gamma$: 123.25(1) °): from left to right: deprotonation, charge compensated by calcium: $<S<CC>S>$, depolymerisation of the silicate and additional Ca(OH)$_2$: $<C^o<CC>S>$ and exchange of proton between silanol groups: $<S<C>S'>$). Si: blue. Ca: turquoise, O: red, H$_2$O: white, OH: black.

The silanol groups are transferred to the neighbouring silicate groups (one per neighbour). This means that the same sequence of protonation as for the non-defective chains are possible: $<'<q>'S'>$ → $<'<q>S'>$ → $<CC_q>S>$.

A missing S can also be replaced by a calcium (C). Such a calcium or any calcium in the interlayer may be partially charge compensated by a hydroxyl group. For example, a bridging silicate in a fully deprotonated silicate chain being replaced by a calcium and two hydroxyls can be denoted as: $<S<CC>S>$ → $<C^o<CC>S>$ (figure 3). Also, additional calcium hydroxide units can be added to the interlayer, maybe at the same time reducing the number of water molecules: $<S'<C>S'> → <S'<C^o>C>S'>$.

Finally, where the S have already been removed, there is also the possibility of defect clusters involving the silicate dimers of two consecutive building blocks in $b_{\text{red}}$ direction. For example, one possibility would be the substitution of a silicate dimer (<<) by two Ca(OH)$_2$ units (/ /), maybe increasing the number of water molecules at the same time to fill up the interlayer: $<'<C>S'> < '<C>S' > → < '/C_11>S'> / <C_{10}>S'>$. In theory it would also be possible to remove a single silicate of the dimer, and thus creating an end-chain silicate: $<C>S' < '<C>S' > → < '/C_{10}>S'> ( ' '<C>S' >. The extra ' is necessary to preserve charge neutrality as can be seen from figure 2. A number of possible defects are schematically illustrated in figure 4.

Finally to give the stacking of these building blocks we again first give the sequence in $b_{\text{red}}$ direction:

... $<S<CC>S>$ $<S'<C_11>$ \ $<S'<C_{10}>$ ...

This means that we keep together the building blocks forming the same linear silicate chains. In this way, the ge-
Figure 4: Schematic view of different 14 Å tobermorite structural defects considered leading to changing Ca/Si ratio. The (mostly) unchanging upper part of the unit cell and the water molecules are not shown on the schematics. Si: blue. Ca: turquoise, O: red, H₂O: white, OH: black. Semi transparent oxygen atoms (O): shared oxygen atoms between the silicate dimers forming the Si-O-Si bond as seen in Figure 1.
ometry of one silicate chain can easily be found by looking at only one of the Ca-Si sheets in the building blocks:

\[<S<<S'<<S'... or ...>> S>\]  
... Next, we will indicate stacking in \(a_{\text{red}}\) direction (back to front in figure 2) by consecutive lines of building block sequences. This will keep the units from the same Ca-Si sheets together. Finally, stacking in \(-c_{\text{red}}\) direction is shown by consecutive paragraphs. For example, a structure comprising 3 x 3 x 2 units we have:

\[
<\text{S}<\text{CC}_7>\text{S} \quad <\text{S'}<\text{C}_{11}> \quad <\text{S'}<\text{C}_{10}\>
\]

\[
<\text{S}<\text{CC}_7>\text{S} \quad <\text{S'}<\text{C}_7>\text{S'} \quad <\text{S'}<\text{C}_8>\text{S'}>
\]

\[
<\text{S'}<\text{C}_7>\text{S'} \quad <\text{S'}<\text{C}_7>\text{S'} \quad <\text{S'}<\text{C}_7>\text{S'}>
\]

3.1.2. Disorder and Stacking Faults

The unit cell defined here is closely related to the MDO2 C centered \(a', b', c'\) unit cell defined by Bonaccorsi et al. \[45\], with \(a_{\text{red}} = a', b_{\text{red}} = b'\) and the origin situated at the center of the unit cell. Only the \(c_{\text{red}}\) parameter defined here differs. It is almost equal to the stacking vector \(t_1\) defined by Bonaccorsi et al., but with \(b_{\text{red}}\) added \((c_{\text{red}} = 0.5 \cdot (c' - a') - b')\), to be able to take into account the ions shared between the two adjacent calcium-silicate layers. However, this definition corresponds to one of two possible positions of the calcium ion in the interlayer region. The other one gives rise to a unit cell with \(c_{\text{red},2} = 0.5 \cdot (c' - a') + b'\) (see figure 5). According to Bonaccorsi et al., these two calcium positions in the interlayer are present each at about 50%. Since we are noting the species within the unit cell layer by layer in \(b_{\text{red}}\) direction, in this second unit the Ca-Si sheet below the interlayer will be first, followed by the interlayer and the Ca-Si sheet above the interlayer. This leads to building bricks of the form: \(>S'>C<S'\quad\) (figure 5). In order to preserve the same calcium-silicate layer structure, the two possible unit cells have to be shifted by \(b'/2\).

This could, for example, correspond to alternating layers with different calcium positions (figure 5). It is also possible to shift from one type of unit cell to another within one and the same layer (see figure 2 of the supplementary material). Due to the shift in origin between the two types of unit cells, a transition building block with only a part of the upper chain, e.g. \(<\text{S}<\text{C}_3\) is needed between a \(<\text{S'}<\text{C}_7>\text{S'}\) and a \(>\text{S'}>\text{C}_7\text{S'}\), to complete the upper calcium-silicate layer.
Figure 5: Alternate layers with $a_{\text{red}} = a', b_{\text{red}} = b', c_{\text{red}} = 0.5 \cdot (c' - a') - b'$ and $a_{\text{red},2} = a', b_{\text{red},2} = b', c_{\text{red},2} = 0.5 \cdot (c' - a') + b'$ unit cells respectively.

Similarly, if the two types of unit cells are inverted, a $>S>C_3$ transition unit cell is needed in-between to complete the lower calcium-silicate layer. The origin of these transition unit cells is displaced by $0.75 \cdot b_{\text{red}}$ compared to the full unit cells.

Apart from the disorder coming from the position of the calcium in the interlayer, in their original description of the 14 Å tobermorite structure Bonaccorsi et al. [45] also discuss the extensive presence of stacking faults in the structure. The authors mention that there are two possible ways of stacking which lead to a slightly different symmetry of the calcium silicate layers: in addition to the stacking described above, where layers with the same orientation are stacked without any shift of the origin ($t_1$), there is also a stacking sequence ($t_2$), where the origin is shifted by $b_{\text{red}}/2$.

This leads to calcium silicate layers where bridging silicate tetrahedra on either side of the calcium silicate layer are only displaced by $a_{\text{red}}/2$ but not in the $b_{\text{red}}$ direction (see figure 3 of the supplementary material), in contrast to the $t_1$ stacking where the bridging silicate on either side of the calcium silicate layer are also shifted by $b_{\text{red}}/2$ (see figure 5).

We propose to indicate a $t_2$ stacking sequence by a $T$ at the beginning of the layer. Bonaccorsi et al. report that random sequences of $t_1$ and $t_2$ stacking were observed experimentally with no marked preference or long scale order. However, the refinement of the structure presented by Bonaccorsi et al. could only be done for the $t_1t_1$ stacking sequence.

The defects described above can exist in all the different types of unit cells. All the cells described above, including the transitional units, can be found in the supplementary material. With these units it should be easy to construct a wide
range of different possible larger scale structures with different amount of order and defects.

3.1.3. Surfaces

Due to the nanosized nature of C-S-H, surfaces are likely to be very important. There are two directions in which surfaces can be constructed with the units already defined. The first one is the (100) and the (-100) surface ([1-10] according to the unit cell defined by Bonaccorsi et al. [45]). In this direction the silicate chains run parallel to the surface and consequently do not have to be cut (see figure 4 of the supplementary information). In fact we can use the \(<S’<C_7>S’>\) or the \(>S’>C_7<S’<\) (or any defective version of it) directly for the construction of these surfaces. Depending on the direction \((\vec{b} \times \vec{c} \text{ or } -\vec{b} \times \vec{c})\) of the surface normal, either a (100) or a (-100) surface is created. The fact that the two surfaces differ is easiest to see when considering the orientation of the hydrogen in the \(S’\) group (figure 4 of the SI). In the case of the (100) surface, the hydrogens at the surface point in the direction of the surface normal whereas for the (-100) they point towards the bulk of tobermorite.

Another surface which is easy to construct is the surface perpendicular to the calcium-silicate layers i.e. the (001) surface (same miller index for Bonaccorsi et al. notation [45]). This surface is in all probability cut at the interlayer, consequently it can be constructed using the \(<S<C_3\) or the \(>S>C_3\) units, depending on the orientation of the surface (see SI figure 4). For a (010) surface, the (012) according to Bonaccorsi et al. [45], the silicate chains will be cut. To remain consistent with the silicate speciation reported in literature, this could be done using \(<’>’\) units to complete the dimers on either side of the interlayer.

For surfaces other that the ones described above, the definition of additional surface units would be necessary.

3.1.4. Extension to other calcium-silicate structures

In the section above we have focused on 14 Å tobermorite. However the notation can easily be extended to other layered calcium-silicate structures. Since the other tobermorite structures (9 Å and 11 Å normal and anomalous tobermorite as well as clinotobermorite [59, 36], see figure 1 of the supplementary material) contain essentially the same calcium-silicate layers, they can be described with the same notation as introduced above. Only the unit cell needs to be redefined for each structure. As an example, anomalous 11 Å tobermorite is represented by \(<S’<s>|S’>\) and clinotobermorite by \(<S|<C_3>|S>\). The unit cells for 11 Å normal tobermorite, the Hamid 11 Å tobermorite as well as clinotobermorite, which can be used to construct C-S-H structures according to the notation above, can be found in
For jennite there are additional $\text{Ca(OH)}_{1.5}$ groups in the calcium-silicate layer. We propose to designate them by $\{\text{or}\}J$, depending on the orientation of the calcium-silicate chains. Such a group has a charge of $+0.5$. In other words jennite can be represented by $\{<S(<C_8>)S>\}$ and the corresponding unit cell.

3.2. Characterisation of Previous C-S-H Models

Our approach can also be used to re-examine different C-S-H models previously described in the literature. However, comparison requires a reasonable description of the underlying structure. For the first full scale model reported in literature [21], the authors did not take into account the known silicate speciation to obtain their high calcium to silicon ratio. This lead to very high strains in the structure upon relaxation and to the adsorption of water in the intralayer space. The resulting structure has little resemblance to tobermorite which does not allow us to fully characterize their model. Also in their further work [31], no atomistic configurations were published, so we cannot reconstruct the C-S-H "DNA" used based on the published data. Fortunately, Kovačević et al. [32] provides all relevant data and consequently we can use one of their sample structures to further illustrate our building-block description of C-S-H. However, discussion is limited to the original paper as no atomistic configurations are available for the revised structures reported later [34].

Kovačević et al. used the 11 Å normal tobermorite reported by Merlino et al [36] as a model structure, this means their model can be described using the corresponding primitive crystallographic unit cell (reported in the supplementary materials: 11_a_norm_tobermorite.config) as the basis for building blocks. Their Sample 1 structure, corresponding to one of the lowest energy family of structures, has been divided into 4x4x2 building blocks with dimension a: 7.28 Å, b: 7.68 Å, c: 12.42 Å, $\alpha$: 81°, $\beta$: 107°, $\gamma$: 122°. Some of the interlayer water molecules, hydroxyl or calcium ions could be assigned to either of two neighbouring cells. However, this assignment is not so important as it does not significantly change the overall properties of the structure. The final description of their structure can be seen in figure 6.

Based on our proposed building-block description, we can now calculate different properties of their model. The stoichiometry of the sample structure can be calculated using equation 1 (where $n_X$ is the number of the units of type $X$ (see figure 2 for the list of symbols), different symbols for different orientations are taken together, i.e. $n_c$ is the total
number of < and >: the resulting stoichiometry is $Ca_{1.68}SiO_{3.7} \cdot 2.2H_2O$.

$$Ca_{\kappa_1}SiO_{\kappa_2} \cdot \kappa_3H_2O$$ (1)

where,

$$\kappa_1 = \frac{n_\angle + n_C + n_f + n_i}{n_\angle + n_S + n_i}$$ (2)

$$\kappa_2 = \frac{3.5 \cdot n_\angle + 2 \cdot n_S + 4 \cdot n_i - 0.5 \cdot n_\prime + 0.5 \cdot n_\circ + n_f}{n_\angle + n_S + n_i}$$ (3)

$$\kappa_3 = \frac{0.5 \cdot (n_\prime + n_\circ) + n_f + n_{H_2O}}{n_\angle + n_S + n_i}$$ (4)

Another quantity that can be easily calculated is the mean chain length $MCL$. The number of silicate chains corresponds to half the number of silicate chain ends (where a monomer counts double) as every silicate chain has two end units. This leads to equation 5 which for the Kovačević model 1 structure gives a mean chain length of 2.4.

$$MCL = \frac{n_\angle + n_S + n_i}{0.5 \cdot n_\angle - n_S + n_i}$$ (5)
There are no silicate monomers in the Kovačević structure (figure 6); the structure consists entirely of dimers and longer chains. The ratio between dimers and longer chains depends on the spacial arrangement of the different units and hence cannot be calculated via a generally applicable formula. Rather the number of dimers and longer chains in the structure have to be counted explicitly. If we look for instance on the frontmost silicate chain of the upper chain of layer 1 of the sample structure we get the following: $< << << <$S$< < <$. This means that, if we suppose the structure to be periodic and hence repeating, there are two dimers per pentamer in this chain. In this way we can identify the total number of dimers (48) and longer chains (8) in the structure. The percentage of dimers (86 % of all chains) is higher than that measured for OPC systems (60 % [6]) and the mean chain length is shorter than that indicated from experimental data (around 3 for fresh cement [4]). However, in the revised model Kovačević et al. report a slightly lower percentage of dimers of 72 % [34].

We can also calculate the extent of silanol protonation $Q_{prot}$ (equation 6): 18 % of the silanol groups in the structure are protonated. This is very high compared to the very small or negligible amount of protonated silanol groups reported in the literature for the high Ca/Si ratio of 1.68 [11, 22]. Similarly, the percentage of the charges of the calcium ions that is compensated by hydroxyl groups $Q_{OH}^{comp}$ (equation 7) is calculated to be: 21 % of the charge of the calcium ions which is compensated by hydroxyl groups in the interlayer. This is consistent with experimental measurements [23]. Some of these hydroxyl groups are likely to be formed upon the splitting of water where a hydrogen is donated to a silanol group. This indicates that, in order to reach both the correct hydroxyl and protonated silanol groups concentration, more hydroxyl has to be introduced implicitly in the interlayer.

$$Q_{prot} = \frac{n'}{n' + 2 \cdot n_c}$$ (6)

$$Q_{OH}^{comp} = \frac{n_o + 2 \cdot n_f}{2 \cdot (n_c + n_C + n_f + n_i)}$$ (7)

Another interesting property to consider is the distribution of charges within the model. The total charge of each unit cell has been calculated based on the building block description (equation 8, where $Z_X$ is the charge of unit $X$). Although the overall structure is charge neutral, we can observe significant charge fluctuations for this Sample 1 structure reported by Kovačević et al. For example in the third column from the left (figure 6) we calculate that a total of -8 e is building up in the upper and -5 e in the lower layer (see supplementary material). We can consider
this as a plane with a charge of 1.75 e/nm$^2$ which is gradually compensated over a distance of 24 Å on either side (i.e. the distance between subsequent charged layers). While charge fluctuations are likely to occur, their magnitude and length scale are likely to be restricted due to the associated Coulombic energy penalty. As no information on the charge fluctuations in C-S-H is currently available, this warrants further study. It would be interesting to see whether these charge fluctuations remained even after the more stringent relaxation procedure used for the revised model of Kovačević et al. [34].

$$Z_{\text{block}}^{\text{tot}} = \sum_{X \in \{<, C, H, O, \ldots \}} n_x \cdot Z_X$$ (8)

### 3.3. Development and Discussion of 14 Å Tobermorite Based C-S-H-Model

#### 3.3.1. Structural Properties

The development of atomistic models for C-S-H has so far mostly been focused on 11 Å tobermorite structures [21, 31, 32]. As noted above, the reasoning for this was mostly the interlayer distance is closer to the one measured for real cements (11-12 Å), and indeed the model structure reported by Kovačević et al. and discussed above [32] has an interlayer distance of 11.87 which is consistent with values reported in literature, although some scatter and controversy about the influence of sample preparation exists [16, 33, 11, 27]. From a compiled dataset of interlayer distance measurements via XRD, Richardson [33] observed that the interlayer distance of C-S-H decreases from ~13-14 Å at Ca/Si = 0.8 to ~10 Å at Ca/Si = 1.5. Recently, Roosz et al. [60] has shown that severe drying conditions during sample preparations can underestimate the interlayer distance measurement using XRD by about 2.8 Å for a C-S-H of Ca/Si = 1.2 having an interlayer distance of 12.5 Å in the fully hydrated state. It seems difficult to reconcile an underlying 11 Å tobermorite structure with the decrease of the interlayer distance observed for C-S-H with increasing Ca/Si ratio. Also the interlayer distance is likely to relax more readily than the calcium-silicate layers. All in all it seems therefore a reasonable assumption that, at least for lower Ca/Si ratios 14 Å might be an equally valid or better base structure than the 11 Å tobermorites. Therefore we consider possible 14 Å tobermorite based C-S-H models and compare them to the preexisting models using the formalism described above.

Richardson recently published some guidelines for the construction of new models, based on summarized properties of different experimentally determined tobermorite structures [33]. Based on the geometry of possible oxygen environments in the crystals and the change of the interlayer distance he concluded that the best model structure for
C-S-H is clinotobmorite. However no further relaxation of the structure other than a change of the interlayer distance was considered. Also part of the measurements found in literature was excluded based on the hypothesis that the larger interlayer distances found were due to an intermixture with calcium hydroxide like phases leading to a higher $\frac{Ca}{Si}$ ratio than is actually observed for the C-S-H phase. However this hypothesis does not seem to be entirely validated and other hypotheses can be put forward to explain different interlayer distances, such as different underlying tobermorite model structures depending on the exact conditions under which the C-S-H was formed.

Figure 7: Structural properties of different (defective) tobermorite structures derived from 14 Å tobermorite as calculated by DFT: $<S'<C_7>S'>$, $<S<CC_7>S>$, $<S<CC_7>G>$, $<S<C_7>C_9>$, $<S<C_7>C_9>$, $<S<C_7>G>$, $<S<C_9>°C°>$, $<S<CC_7>G>$, $<S<CC_7>G>$, $<S<CC_7>G>$ with Ca/Si ratios 0.83, 1.0, 1.2, 1.6 1.5 and 2.25 respectively. Trends reported by Ian Richardson shown in black [33].

To confirm the hypothesis, that other tobermorite model phases cannot a priori be excluded, DFT minimizations of a periodically repeated single, defective building block of 14 Å tobermorite [45] structures with varying $\frac{Ca}{Si}$ ($<S'<C_7>S'>$, $<S<CC_7>S>$, $<S<CC_7>G>$, $<S<C_7>C_9>$, $<S<C_7>C_9>$, $<S<C_7>G>$, $<S<C_9>°C°>$, $<S<CC_7>G>$, $<S<CC_7>G>$) were relaxed and characterized structurally according to the characteristics defined by Richardson [33]. As one can see from the building block description, the number of oxygen atoms in the structure was kept constant, removal of, for example, a silicon tetrahedron being compensated by the addition of either water or hydroxyl molecules.

For the non-defective 14 Å tobermorite ($<S'<C_7>S'>$), the interlayer distance is only overestimated by about 0.3 Å compared to the experimental structure (see table [1]) and the average Ca-O distance is underestimated by 0.06 Å (see figure [7]), giving us confidence in the DFT calculated structures. The Average Ca-O coordination and distances as well as the polyhedral volume is entirely consistent with the structural integrity checks suggested by Richardson [33], except for the highest $\frac{Ca}{Si}$ ratio studied here ($<C°<C°C_7>C°>$). If we accept the result of the integrity check,
this indicates either that the DFT calculations found an unfavorable local minimum or that this defect structure is not present in this form or only in low concentrations. In addition, up to a $\frac{Ca}{Si}$ of 1.2 the coordination number decreases from 6.6 to 6. This is consistent with the reported decrease of the coordination number reported by Soyer-Uzun et al. However, for $<^{°}C°<C_{9}>'$ ($\frac{Ca}{Si}$: 1.5) the coordination number is again 6.7.

Both the deprotonation ($<S'<-><S<->^{°}S><°><°>$) and the removal of a silica tetrahedron ($<S<->^{°}S><°>$) results in a decrease in the interlayer distance. In contrast, the replacement of a silica tetrahedron with a calcium group ($<S<->^{°}S><°>$) does not seem to have an influence on the interlayer distance, while the addition of calcium hydroxyl groups to the interlayer ($^{°}C°$) leads to an increase in the interlayer distance. Taking into account the accuracy of the method, as well as the limited number of structures considered (e.g. the water content has not been varied at all) the results, at least up to a $\frac{Ca}{Si}$ of 1.2 are entirely consistent with the trends reported by Richardson et al. [33] (see figure 7). The interlayer distance, although at the upper range of the measured experimental data reported [33] decreases linearly for the structures $\frac{Ca}{Si}<1.2$ and even the $<^{°}C°<C_{9}>'$ structure with a $\frac{Ca}{Si}$ of 1.5 has a reasonable, although slightly higher than expected interlayer distance of 13.2 Å. Additionally, the interlayer distance might be further decreased by decreasing the number of water molecules in the unit cell.

This is the first time a decrease in the interlayer distance with increasing $\frac{Ca}{Si}$ from 0.83 to 1.2-1.5 as observed experimentally was predicted by atomistic simulations. In addition, the structures conform with the structural guidelines proposed by Richardson [33]. Even though the decrease in interlayer distance is somewhat less than expected, we conclude that 14 Å tobermorite should be further considered as a model structure for C-S-H at the very least for low $\frac{Ca}{Si}$. However to further explore this route, larger scale defects and larger scale structures should be considered. Consequently in the next section, $\frac{Ca}{Si}$ increasing defects in larger scale structures are studied with molecular dynamics and compared to the results in this section.

3.3.2. Structure and Energies of Defective Building Blocks

In the preceding section a single building block ($<S '<C_{7}>S '>$) is used for DFT calculations. This means that the defect concentrations are very high and will be strongly interacting, which will reflect in the DFT results. Consequently, to assess the energy and properties of non-interacting defects an isolated defective building block in a supercell of 4x3x2 14 Å tobermorite building blocks ($<S '<C_{7}>S '>$) is considered, resulting in a simulation cell of 27x21x28 Å. Classical MD was used to take advantage of dynamics and its rapid calculation of interatomic forces to efficiently probe the
energy landscape, and thus, evaluate the relaxation of larger scale structures compared to DFT.

Additionally, to start addressing the large structural disorder seen in tobermorite structures - leading to positional disorder, stacking faults and polymorphism - a second structure was investigated in addition to the \(<S' <C_7>S'>\) 14 Å tobermorite reported by Bonaccorsi et al. \[45\]. For the second 14 Å polymorph \(<S' <C_7>S'>\), the orientation of the protonated silanol group is different on either side of the interlayer, leading to a different orientation of the silicate tetrahedron, similar to that observed for clinotobermorite (see figure 1 of the supplementary materials).

Structural Relaxation within the Building Blocks: A substantial amount of relaxation was observed in the interlayer. Both the positions of the ions and the water molecules in the interlayer as well as the orientation of the silicate tetrahedron changed significantly during relaxation \[61\]. Consequently, choosing the exact starting position of the species in the interlayer does not appear to be critical that, e.g., a \(<'S' <C_7>S'>\) unit will be equivalent to a \(<C_5<S'C_7>S'>\) unit. However, if more complicated structures are studied, this point needs to be verified, and a more detailed search of configurational space is required to identify the relevant local and global minima.

The structure of the silicate layers, on the other hand, does not relax as readily. In particular, due to the low symmetry of the system, the positioning of the hydroxyl groups on the two available sites is not equivalent. This means that a \(<S' <C_7>S'>\) unit is not equivalent geometrically to a \(<S' <C_7>S'>\) unit. In fact, the first corresponds to the polymorph used for some of the calculations as described above. Similarly a \(<'S' <C_7>S'>\) unit is not equivalent to a \(<'C_7>S'>\) unit, even after extensive relaxation. Also the relaxation of the whole structure can be very important. Depending on the energy minimization and relaxation path, the structure can relax to a different minimum with a different hydrogen bonding network and will no longer relax to the more pronounced minimum even after more than 1 ns of molecular dynamics. In table 2 the different energies of two differently relaxed \(<S' <C_7>S'>\) structures can be seen. This underlines the importance of testing a number of different structures and extensive structure relaxation to get a realistic estimation of the structure and energies. In this context, the current results, especially for the classical atomistic calculations, are meant as an initial point. A more in-depth study of these effects is currently under way and results will be presented in a separate article.

Defect Energies: The enthalpy of the different defects are difficult to compare, not least due to their different stoichiometries. Consequently, the raw energies are converted to reaction enthalpies, by calculating the enthalpy of formation from quartz \((SiO_2)\), lime \((CaO)\) and water \((H_2O)\) \[9\]. For the atomistic MD calculations, a correction had
to be used for any $H_2O + O \rightarrow 2 \cdot OH$ occurrences [56].

$$n_1 \cdot CaO + n_2 \cdot SiO_2 + n_3 \cdot H_2O \rightarrow [(CaO)_{n_1}(SiO_2)_{n_2}(H_2O)_{n_3}]_{tot}$$ (9)

Generally, reaction enthalpies are given per unit or mole of product. However, as there is no clear structural unit of C-S-H and its stoichiometry is variable, the notion of a mole or unit of C-S-H is not well defined. Therefore, in the literature energies are generally normalized per silicon [32]. The disadvantage of such a normalization is that this makes the reaction enthalpy undefined for silica free phases such as portlandite, the comparison of which with C-S-H is however often interesting. Therefore, we chose another way to normalize the energy by fixing $n_1 + n_2 + n_3 = 1$.

While it is not possible to directly compare energies of C-S-H structures with different stoichiometries, it makes it easy to estimate relative stabilities using possible reactions between different structures.

Table 2 shows the defect energies for a series of more than 10 different defects with different Ca/Si ratios. All the defects shown are energetically stable (negative enthalpies) compared to pure CaO, SiO2 and water for the configurations studied (i.e. an isolated defect in a block of 24 unit cells for MD calculations). The enthalpies have been calculated for both polymorphs considered here: ($<S'<C_7>S'$) and ($<S'<C_7>S>$). The enthalpies for the same defect are very similar in both structures (see table 2). The differences between the structures are smaller than the expected errors for all defect energies. This again confirms the hypothesis that there will be some structural variety for C-S-H, similar to the polymorphism and order-disorder character observed for tobermorites [36, 45, 59].

Where possible, the defect energies calculated with classical molecular dynamics have been compared to the DFT calculations described in section 3.3 (table 2). As mentioned before, a single building block with half the volume of the original tobermorite unit cell is used to do all the DFT calculations. This means that the defects are strongly interacting and consequently the DFT results are not necessarily expected to be equivalent to the classical molecular dynamics results, where defects are considered to be non-interacting as isolated defects in an non-defective 14 Å tobermorite matrix. However, the results are generally in relatively good agreement. The biggest difference is observed for the $°C_{3.5}'$ $'$ defect. For this defect, large structural relaxations were observed leading to a decrease of the interlayer distance to 13.24 Å (see figure 7). For an isolated defect on the other hand, the interlayer distance is fixed by the surrounding non-defective structure. Consequently, the energy of the isolated defect is expected to be less favorable, which is what is observed (table 2).
Table 2: Table of calculated enthalpies for different defects in 14 Åtobermorite in eV. An eV corresponds to 96.49 kJ/mol. The corresponding table in kJ/mol can be found in the supplementary materials. All the defects shown here are energetically favourable compared to pure CaO, SiO$_2$ and water. Although some of these defects were previously proposed to model C-S-H, this is the first time the energetic stabilities of these defect structures have been calculated. ‡ Different structure obtained from different relaxation process will not relax to the lower energy structure even after > 1 ns of molecular dynamics. †Note: shows large interlayer contraction. * calculated from an equivalent unit where a calcium is shared between two neighbouring building blocks: $<$S<$CC_7$>S'> $<$S'$<$C_7$>S$

<table>
<thead>
<tr>
<th>Ca/Si</th>
<th>Formula</th>
<th>$n_{CaO}$</th>
<th>$n_{SiO_2}$</th>
<th>$n_{H_2O}$</th>
<th>$\Delta H_{reac}^{&lt;S'&lt;C_7&gt;S'}$ [eV]</th>
<th>$\Delta H_{reac}^{&lt;S'&lt;C_7&gt;S'}$ [eV]</th>
<th>$\Delta H_{DFT}^{reac}$ [eV]</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.83</td>
<td>$&lt;S'&lt;C_7&gt;S'&gt;$</td>
<td>0.26</td>
<td>0.32</td>
<td>0.42</td>
<td>-0.37±0.06</td>
<td>-0.27</td>
<td></td>
</tr>
<tr>
<td>0.83</td>
<td>$&lt;S'&lt;C_7&gt;S'&gt;$‡</td>
<td>0.26</td>
<td>0.32</td>
<td>0.42</td>
<td>-0.16±0.04</td>
<td>-0.16±0.04</td>
<td></td>
</tr>
<tr>
<td>1.00</td>
<td>$&lt;S&lt;CC_7&gt;S&gt;$</td>
<td>0.32</td>
<td>0.32</td>
<td>0.37</td>
<td>-0.30±0.04</td>
<td>-0.28±0.03</td>
<td>-0.25</td>
</tr>
<tr>
<td>1.00</td>
<td>$&lt;S&lt;CC_7&gt;S&gt;$*</td>
<td>0.32</td>
<td>0.32</td>
<td>0.37</td>
<td>-0.30±0.04</td>
<td>-0.34±0.03</td>
<td></td>
</tr>
<tr>
<td>1.00</td>
<td>$&lt;S'&lt;C_9&gt; '&gt;$</td>
<td>0.25</td>
<td>0.25</td>
<td>0.50</td>
<td>-0.15±0.04</td>
<td>-0.14±0.03</td>
<td></td>
</tr>
<tr>
<td>1.00</td>
<td>$&lt;S'&lt;C_8&gt; '&gt;$</td>
<td>0.25</td>
<td>0.25</td>
<td>0.50</td>
<td>-0.02±0.04</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.20</td>
<td>$&lt;S&lt;CC_9&gt; &gt;$</td>
<td>0.30</td>
<td>0.25</td>
<td>0.45</td>
<td>-0.25±0.03</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.20</td>
<td>$&lt;S&lt;CC_9&gt;C'&gt;$</td>
<td>0.30</td>
<td>0.25</td>
<td>0.45</td>
<td>-0.15±0.05</td>
<td>-0.17±0.04</td>
<td>-0.28†</td>
</tr>
<tr>
<td>1.20</td>
<td>$&lt;S'&lt;C_8&gt;C'&gt;$</td>
<td>0.30</td>
<td>0.25</td>
<td>0.45</td>
<td>-0.21±0.04</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.25</td>
<td>$&lt;S'&lt;CC_{10}&gt; S'&gt;$ $S'&lt;CC_{10.5}&gt; $</td>
<td>0.24</td>
<td>0.20</td>
<td>0.56</td>
<td>-0.26±0.04</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.40</td>
<td>$&lt;S&lt;CC_7&gt;C'&gt;$</td>
<td>0.35</td>
<td>0.25</td>
<td>0.40</td>
<td>-0.36±0.05</td>
<td>-0.31±0.04</td>
<td>-0.29</td>
</tr>
<tr>
<td>1.50</td>
<td>$&lt;C'&gt;C_9&gt; '&gt;$</td>
<td>0.29</td>
<td>0.19</td>
<td>0.43</td>
<td>-0.26±0.04</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.25</td>
<td>$&lt;C&gt;C'C_7&gt;C'&gt;$</td>
<td>0.39</td>
<td>0.17</td>
<td>0.43</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.67</td>
<td>full model see figure 9</td>
<td>0.39</td>
<td>0.23</td>
<td>0.38</td>
<td>-0.29±0.08</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
As mentioned above, the energies themselves can only be directly compared between defects with the same stoichiometry. What is observed when comparing defects with equivalent stoichiometry (i.e., \(<S'\langle C_9\rangle'\) with \(<S'\langle C_8\rangle'\)) as well as the series \(<S\langle C_9\rangle \rangle, <S\langle C_8C_7\rangle' >, <S'\langle C_8\rangle >\) and \(<S'\langle C_7\rangle >\langle C_8\rangle >\) is that defective structures with deprotonated silanol groups are preferred. This is consistent with experimental observations, where a decrease of the concentration of protonated silanol groups is observed with increasing Ca/Si ratio. Also at a Ca/Si ratio of 1.2, an additional calcium ion in the interlayer (\(<S\langle C_8C_7\rangle' >\)) seems to have a very similar energy to a calcium ion substituting a bridging silica, although the two defects are structurally distinct. This second defect has often been omitted from previous modelling of C-S-H yet has been shown to be important by Kumar et al. 

Another property we can look at is the local charge neutrality of the structure. All, except one of the building blocks discussed above, (6.7 x 6.2 x 14.0 Å before relaxation) are charge neutral. Consequently, any larger structure constructed from those units will also be charge neutral. However, in reality, charged building blocks may also exist, compensated by a space charge region. Consequently, we have looked at charged units in one instance: in addition to the charge neutral defect \(<S\langle C_7\rangle \rangle, <S'\langle C_7\rangle >\), we have also looked at an equivalent defect, where the calcium ion is shared between two units: \(<S'\langle C_7\rangle \rangle, <S\langle C_7\rangle >\). These charged building blocks have very similar energies (see figure 2). Thus, the building blocks do not need to be charge neutral even if the total structure is. This means the magnitude and the length scale of realistic charge fluctuations also remain to be studied.

To compare the defect energies between different stoichiometries, reaction enthalpies between the different defects, and where appropriate related phases such as \(SiO_2\), \(CaO\), \(Ca(OH)_2\) and \(H_2O\), should be calculated. As an example, to compare the enthalpy of \(<S\langle C_7\rangle \rangle, <S'\langle C_7\rangle >\) we can look at the reaction 0.79 \(\{[CaO]_{0.316}(SiO_2)_{0.316}(H_2O)_{0.368}\}_{lab} + \)

\(0.21 \cdot H_2O \rightarrow [(CaO)_{0.25}(SiO_2)_{0.25}(H_2O)_{0.50}]_{lab}\). As the reaction enthalpy for pure water is zero, the enthalpy for this reaction is equal to \(\Delta H_{\text{reac}}^{S'\langle C_7\rangle} <S'\langle C_7\rangle > - (-0.79) \cdot \Delta H_{\text{reac}}^{<S\langle C_7\rangle} <S\langle C_7\rangle > = 0.09 \pm 0.07\) eV (8.7 kJ/mol), indicating that, neglecting any entropic effects, the \(<S\langle C_7\rangle \rangle, <S'\langle C_7\rangle >\) defect appears to be slightly more favorable. However, these reaction enthalpies can be more readily estimated and trends seen if the reaction enthalpies are plotted on a ternary diagram (figure 8). For the atomistic MD calculations, only the lowest energy for a specific stoichiometry is reported. From the ternary diagram, points lying on the straight line formed by three points or a point inside a triangle formed by three other points can be readily compared, as reactions between them can be defined. A thorough knowledge of the energy landscape within
Figure 8: Variation of the calculated reaction enthalpy in eV for the building blocks considered from (a) DFT and (b) classical MD calculations compared with different thermodynamic models of C-S-H; (c) the Tobermorite/jennite solid solution model from the CEMDATA thermodynamic database [62] and (d,e) the CSH3T and CSHQ models from [63]. An eV corresponds to 96.49 kJ/mol. Important cement phases C₃S, C₂S and Ca(OH)₂ are also shown on (c, d and e) diagrams. The bottom edge of the ternary diagram shows increasing water content from left to right, the edge on the right hand side shows increasing silica content from bottom to top and the edge on the left hand side shows increasing lime content from top to bottom. For illustration, an example composition with stoichiometry (CaO)₀.₂(SiO₂)₀.₆(H₂O)₀.₂ is shown in (a) (red star) together with a guide as to how to read compositions. The reaction enthalpies calculated for 14 Å tobermorite (26 % CaO, 32 % SiO₂, 42 % H₂O) from DFT, MD and the Tobermorite/Jennite model are -0.27, -0.37 and -0.31 eV respectively. Thus allowing a reliable comparison between the different techniques.

Same colour scale for the reaction enthalpy (see legend) is used for all plots, positive reaction enthalpies are shown in gray.
this ternary diagram would not only help immensely to understand the thermodynamic behavior of C-S-H but would also serve as a quick check whether new theoretical atomistic models are reasonable. However, plotting the previously reported experimentally based thermodynamic models from literature (the tobermorite/jennite solid solution model from the CEMDATA thermodynamic database [62] as well as the CSH3T and CSHQ models from [63]) illustrates that more work is still needed to reach this goal. Due to the uncertainties of the atomistic structure of C-S-H, the resulting energies from the different thermodynamic models will largely depend on the underlying model assumptions. But globally, the results from the experimental models indicate that the enthalpy likely lies between -0.33 to -0.25 eV (-31.8 to -24.1 kJ/mol) with less negative energies being expected for structures containing more water.

For both the DFT and the molecular dynamics results, the calculated energies largely follow these trends. Even for the defect comprising a missing dimer ($<S'<\text{CC}_{10}> \setminus <S'<\text{C}_{10.5} \setminus >$) the energy does not seem to be outside the expected range. At the same time, the defect is unlikely to be stable at large concentrations, as this would significantly change the nature of the calcium-silicate sheets, which is not observed experimentally [24, 16]. This makes this defect difficult to investigate with DFT. However, based on the defect energy calculated here, it seems likely that missing dimers occur to some extent at higher $\frac{Ca}{Si}$ ratios. The biggest difference between the thermodynamically expected and calculated enthalpies are observed for the defective structure involving the removal of a silicate tetrahedron without replacement and a protonated silanol group on a silicate dimer ($<S'<\text{C}_9> \setminus$). Based on the comparison with DFT, large structural relaxation of the interlayer distance is expected when removing a silicate dimer without replacement, making an isolated defect much less energetically favorable than a fully defective structure. Also, as discussed above, protonated silanol groups for higher Ca/Si ratio structures do not seem to be energetically favorable. Even though the removal of a silicate dimer is a relatively large scale defect, the structural relaxation of the position of the calcium within the calcium-silicate layer is small [61].

Finally, it has to be mentioned that, while the range of expected enthalpies seem to make it possible to exclude some unrealistic energies, for a complete estimation of the relative stability of structures the Gibbs free energy and not the enthalpy should be used. Especially the configurational entropy, favoring more disordered structures, is likely to significantly influence results. The study of these effects is currently ongoing and will be the subject of a follow-up article.
3.3.3. Full Atomistic Model

We can use the building block description to predict and build a larger scale model, inspired by the defect energies calculated above and experimental observation by different authors. Let us consider a C-S-H like slab 4x3 units in $a_{red}$ and $b_{red}$ direction respectively and 2 units in $c_{red}$ direction. In order to take into account the nanosized character of C-S-H, let us assume that the slab is terminated on both sides by a surface in $c_{red}$ direction. We can now describe each consecutive layer in the above described notation (see figure 9).

This structure was generated ad hoc using the building block description such that its Ca/Si ratio, stoichiometry and the chemical features are comparable to that of an industrially relevant C-S-H or high Ca/Si ratio synthetic C-S-H that has recently been produced for the first time [13]. The exact stoichiometry of the structure can be calculated according to equation [1]. For the complete sample structure, we get a stoichiometry of $C_{1.7}SiO_{3.7} \cdot 1.3H_2O$ with a calcium to silicon ratio $r_{Ca/Si}$ of 1.67. This is consistent with measurements by different authors of the calcium to silicon ratio of C-S-H in Portland cement [4, 29, 15]. The 1.3 formula units of water in the average stoichiometry is partly present as both molecular water and $OH^-$ ions. In fact 22% of the charge of the calcium ions is compensated by hydroxyl groups (see equation [7]). The amount of hydroxyl groups is consistent with measurements by Thomas et al. [23], who
reported that around 23% of the charge of the calcium ions is compensated by hydroxyl groups. However, the overall water content is somewhat lower than what has been reported in literature [29, 15]. Following experimental results in the literature for high Ca/Si ratio C-S-H [22, 11], which are confirmed by our enthalpy calculations, defects with protonated silanol groups are not included in the full model. Based on our results, two defects are included which have not been considered in (most) previously reported atomistic models in the literature [21, 31, 32]: the replacement of a bridging silica tetrahedra ($<S<$ or $>S>$) by a calcium ion ($<C<$ or $>C>$) as well as the replacement of a silicate dimer by hydroxyls ($<S’<CC_{10}> \backslash <S’<C_{10.5}> \backslash$). The inclusion of the former is in accordance with recent work by Kumar et al. [13], where they found calcium at the bridging site as a defining feature of C-S-H. All building blocks are charge neutral.

The mean chain length of the model is 3.3 (eq 5). This is consistent with reported values for fresh cement [7, 4, 8, 9, 10, 11]. There are 23 dimers and 12 longer chains in the structure, which corresponds to 63% dimers. Again, this is more or less consistent with results reported for fresh cement in literature [6], where 60% dimers were reported.

Molecular dynamics calculations of the structure lead to an interlayer distance of 13.76 Å, which is larger than generally expected. However, as discussed above, there is still some debate about the interlayer distance of C-S-H [16, 33, 11, 27]. The Ca-O environment with a coordination number of 5.8 and an average Ca-O distance of 2.3 Å (nearest neighbour cut-off of 3.0 Å) is both consistent with the trends reported in figure 7a and with the coordination number of around 6 reported by Soyer-Uzun [27]. The definition of the density is complicated by the corrugated surface. However, we can give the density of the central units, which is 2.35 g/cm$^3$. If on the other hand we assume that the building blocks at the surface only occupy about one third of the volume of a full building block and that all water at the surface is part of the pore water and not of the C-S-H structure [29], we can calculate a density of 2.70 g/cm$^3$ for the whole structure, which should give an upper limit. Given the uncertainties, the results seem reasonable compared to experimental results [29, 15].

Compared to previous models in the literature (see section 3.2), our model includes surfaces, an extended range of different defects (replacement of bridging silicate tetrahedron with calcium and of a silicate dimer with hydroxyl groups), is entirely composed of neutral building blocks and is based on the 14 Å tobermorite reported by Merlino et al. [45]. This leads to a model structure where the concentration of protonated silanol groups as well as the silicate chain speciation and the mean chain length is more consistent with experimental measurements than previous models.
However the water content of the model is lower and the interlayer distance larger than expected, areas where previous models were performing better (see section 3.2). As for the magnitude and length scale of charge fluctuations, these still have to be further looked at.

Finally, we have estimated the reaction enthalpy for this structure and compared it to the DFT and MD results of single defects discussed above (table 2 and figure 8). The reaction enthalpy has been calculated excluding all water at the surface. Although the error on this energy is likely to be quite large (> 0.08 eV or 8 kJ/mol, the estimated force field error), both to the thermodynamic ensemble used for the calculations (NVT) and due to the uncertainties concerning the inclusion or not of the surface water, the reaction enthalpy of -0.29 eV (-28 kJ/mol) is well within the expected values (see table 2). The full atomistic model reported here serves as an example to illustrate the feasibility and ease of the bottom up building block approach introduced in this paper. A more thorough study on different combinations of defective building blocks to better understand the disordered, stochastic nature of C-S-H structures is currently underway and will be the subject of a future publication.

4. Conclusions

The complexity of C-S-H structures has meant that the generation of reliable atom-level structural models has been elusive. One of the major difficulties in C-S-H model construction is the generation of the correct silicate-chain speciation with a low concentration of protonated silanol groups and an appropriate Ca/Si ratio. Additionally, full atomistic models are often developed for a single stoichiometry or Ca/Si ratio and results are difficult to extrapolate to other C-S-H stoichiometries [32, 21]. In the first part of this paper, we described a novel way for generating and describing C-S-H structures using a building block approach. The building block description encodes all important features of a full atomic C-S-H model in a simple, readable string of characters. The most relevant properties of a C-S-H model structure, such as calcium silicate ratio, silica chain length distribution, number of interlayer hydroxyl groups and degree of protonation of the silanol groups can be directly calculated from the building block description. In addition, the building block description can be used for C-S-H models based on different model crystal structures (14 Å tobermorite, 11 Å tobermorite, jennite etc.) and can deal with larger scale defects such as stacking faults and surfaces. This makes it easier to construct and compare structures without having to analyse and manipulate the full atomic structure.
was illustrated for a model structure taken from the literature.

We then investigated different defective 14 Å tobermorite based building blocks. We applied both DFT and classical MD simulations to calculate structural relaxations and energies. DFT calculations on high defect concentrations show structural evolutions consistent with previously reported experiments, at least for \( \frac{Ca}{Si} \leq 1.5 \). Normalising reaction energies and plotting them on a ternary diagram additionally allows the comparison of a range of C-S-H building blocks and structures. Based on these results, we propose that two new variations should be included for the construction of future models, namely the replacement of bridging silicate tetrahedra with calcium and of silicate dimers with hydroxyl groups. Also, we find that deprotonated silanol groups are more stable than protonated ones at high Ca/Si ratio, which is consistent with experimental results [22][11]. Many of the defective building blocks studied here have similar energies, emphasizing the disordered character of C-S-H, where many different local defects are likely to be distributed randomly within the overall structure.

Finally to emphasize the capabilities of the building block description and based on the insights of the calculations on single defective building blocks, an “ad-hoc”-constructed, larger-scale C-S-H example structure is presented and discussed. The model has a reasonable energy and both the concentration of protonated silanol groups and the silicate chain length distribution is more consistent with experimental measurements compared to the previous models. The model includes surfaces and other defects (replacement of bridging silicate tetrahedra with calcium and of a silicate dimer with hydroxyl groups) that have not been considered so far in full scale atomistic models. However, further work is clearly needed to address the discrepancies on water content and the interlayer distance.

In summary, the development of the building block description of C-S-H now makes it easier to propose new models with relative ease and to compare existing models in much more detail than previously possible. In addition, the normalization of the defect energies proposed here and the ternary energy diagrams should make it easier to compare the energies of different structures. This will significantly facilitate the choice, construction and improvement of C-S-H models for a range of Ca/Si ratios in the near future. We further suggest that this will facilitate discussions of what are the characteristics that make C-S-H the material with the very distinctive properties observed experimentally and provide a better fundamental understanding in growth mechanisms, mechanical properties and ageing of cementitious systems.
Acknowledgement

The authors would like to thank Ellis Gartner from Lafarge, Prof. Karen Scrivener from EPFL, Patrick Juilland from SIKA and Uli Aschauer from ETH Zurich for their input and helpful discussions. Additionally the authors would like to thank the industrial-academic research network Nanocem for the funding of the work of Sandra Galmarini. The authors would like to thank Swiss National Foundation for financing the contribution of Aslam Kunhi Mohamed (Grant No. 153044). University of Bath acknowledges the Engineering and Physical Sciences Research Council (EPSRC) for financial support through project EP/K025597/1. Access to HPC computing resources on ARCHER for DFT calculations was provided via the membership of the Materials Chemistry Consortium, which is funded by EPSRC (EP/L000202) and on HPC Belena through the the University of Bath.

References


[38] G. Kresse, J. Furthmüller, Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set, Physical review B 54 (1996) 11169.


