From Atoms to Cities: A Bottom-Up Analysis of Infrastructure Materials and Systems

by

Mohammad Javad Abdolhosseini Qomi

B.Sc., University of Tehran (2006)
M.Sc., Sharif University of Technology (2008)

Submitted to the Department of Civil and Environmental Engineering in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in the field of Civil and Environmental Engineering

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2015

Signature redacted

Author .....
Department of Civil and Environmental Engineering
May 7, 2015

Signature redacted

Certified by...
Franz-Josef Ulm
Professor of Civil and Environmental Engineering
Thesis Supervisor

Signature redacted

Certified by..
Roland Pellenq
Senior Research Scientist of Civil and Environmental Engineering
Thesis Supervisor

Signature redacted

Accepted by .....
Heidi Nepf
Donald and Martha Harleman Professor of Civil and Environmental Engineering, Chair, Graduate Program Committee

From Atoms to Cities: A Bottom-Up Analysis of Infrastructure Materials and Systems

by

Mohammad Javad Abdolhosseini Qomi

Submitted to the Department of Civil and Environmental Engineering
on May 7, 2015, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in the field of Civil and Environmental Engineering

Abstract

Civil infrastructure is and continues to be the backbone of our society to meet our needs in housing, transportation, water and electricity supply, and so on. However, its functions are recently revisited in response to rising concerns about its certain sustainability aspects. These aspects include and are not limited to excessive greenhouse gas emissions, unreasonably high energy footprint, relatively short service life, low durability and poor resilience. This presents us with an exclusive opportunity to take these detrimental aspects seriously and turn them into exciting venues for research in the realm of civil and environmental engineering. These opportunities are disseminated across the entire infrastructure landscape, spanning several length scales starting from the molecular structure of construction materials to the entire global transportation network.

From Atoms to Cities is intended to provide a multiscale bottom-up framework to seamlessly connect the heat transport through the molecular structure of construction materials to thermal energy losses at the city level. Separated by twelve orders of magnitude in length scales, from nanometers to kilometers, this provides a chance to link ideas in mechanics and physics of materials to analysis of complex systems. Two major impediments hinder any progress in pursuit of such a hierarchical multiscale model: the absence of a realistic molecular structure of construction materials such as Calcium-Silicate-Hydrates (C-S-H), the glue of concrete, and the multiplicity of factors affecting heat losses at large scales. The first is an indispensable requirement in statistical mechanics as it shapes the energy landscape. The second makes it rather impossible to quantitatively assess the impact of sustainability initiatives at the city scale.

By combining the tool of statistical physics with combinatorial screening technique, we first construct a database of *realistic* molecular structures of C-S-H with varying calcium-to-silicon ratios and compare them against an extensive array of nano-textural and nano-mechanical experiments. A comprehensive analysis of this database reveals a deeper level of connection between cement science and glass physics. This includes the existence of anomalies in mechanical properties similar to that observed

in rigidity transition windows in binary glasses and the presence of extra atomic vibrational modes at low THz regime known as Boson peak. These models are further utilized to calculate the heat capacity and transport properties using Green-Kubo formalism in equilibrium molecular dynamics. While considering other phases in cement paste, the use of mean-field homogenization technique enables us to upscale thermal properties from the nanoscale to the engineering macroscale. The macro-level thermal properties are subsequently compared with those measured experimentally throughout the cement's hydration process. Afterwards, we show that the building envelope's heat transport property is among the set a few influential parameters that affect heat losses at the city scale. This subset of key parameters makes it feasible to construct a high fidelity mechanistic-based reduced order model of heat losses at the building level. Together with energy consumption data of more than 6,200 buildings in Cambridge, MA, this model paves the way to find the shortest path to reduce heat losses in city's building block through retrofit.

Thesis Supervisor: Franz-Josef Ulm
Title: Professor of Civil and Environmental Engineering

Thesis Supervisor: Roland Pellenq
Title: Senior Research Scientist of Civil and Environmental Engineering

Acknowledgments

I would like to take this opportunity to thank my supervisors, Professor Franz-Josef Ulm and Professor Roland Pellenq for their invaluable advice and unwavering support throughout my PhD studies.

Joining Prof. Ulm's group is the best thing happened in my academic career. Prof. Ulm is one of the leading contemporary figures in applied physics and solid mechanics. For me, working with him was a life changing experience. Honestly, I wish I could go back in time and work harder and learn more from him. I am truly humbled by his depth of understanding physics of engineering problems and his degree of enthusiasm about *both* research and education. By setting an example, he refueled the passion for civil engineering in me and I hope to convey this fire to my future students.

I was fortunate enough to begin my PhD career around the same time Prof. Pellenq started his career at MIT. As a physicist, Prof. Pellenq employs statistical mechanics to tackle problems related to multiscale porous materials with a wide range of applications in construction materials and energy harvesting and extraction. I am truly indebted to him as he shaped my understanding of physics during hundreds of *one-to-one* meetings. He is extremely passionate, yet patient enough to discern subtleties in a physical phenomenon or similarities between seemingly distant fields. To me, he is the source of inspiration.

It is with my greatest pleasure to acknowledge listening, learning, talking and working with Professor Sidney Yip. Prof. Yip is a world-renowned figure in computational physics of materials across time and length scales. He is exceptionally knowledgeable, yet very humble to attentively listen to every word you say. To me, he is a role model.

During the course of my PhD studies, I had the unique opportunity to work with Professor Marta Gonzalez. As a brilliant physicist, she is not only passionate about data science, but also open to new lines of research. In spite of MIT's tenure pressure, she is absolutely calm, yet extremely resolved to achieve her research goals.

I am grateful to have such an understanding and supportive mentor.

In the same line, I would like to express my gratitude to Professors Markus Buehler and Martin Bazant, the members of my PhD committee. Without their advice, this work would not have this current structure.

My research and intellectual growth has benefited from stimulating discussions with many individuals including Profs. Hamlin Jennings, Henri Van Damme, Krystyn Van Vliet, Alain Barronet, Amir Reza Khoei, MohammadTaghi Kazemi, Mathieu Bauchy, Drs. Konrad Krakowiak, Benoit Coasne, Jeff Thomas, Hegoi Manzano, Amin Aghaei and other PhD candidates Jake Sobstyl and Jameson Toole. In particular, I acknowledge many lively discussion with Dr. Davoud Ebrahimi that helped me better understand the fundamentals of my research.

For their friendship and collaborative spirit, I am thankful to my peers in Concrete Sustainability Hub (CSHub) at MIT. This work would not have been possible without financial supports provided by CSHub through the Portland Cement Association (PCA) and the NRMCA Research and Education Foundation. I am also thankful to Massachusetts Institute of Technology for bringing all these amazing people and resources together.

Finding a unique interest in nanomechanics was not an act of serendipity in my case. For this, I am indebted to my elder brother, Amir Mahdi Abdolhosseini Qomi, an electrical and computer engineer. Since he knew about my interest in solid mechanics, he gave me a copy of *Nano Mechanics and Materials* by Prof. W.K. Liu and coauthors in fall 2006. Amir Mahdi explained what he thought to be the state-of-the-art in his vision: *linking the mechanics of atoms to the mechanics of continuous matter*. To thirsty ears of a fresh out of college student, this sounded fascinating.

My PhD studies would not have been possible without the lifetime support and encouragement of my family: my father (Mohammad Abdolhosseini Qomi), mother (Maryam Talebian) and sister (Layla Abdolhosseini Rad). Although, I have not had the chance to visit them since the begining of my PhD, they have always been in my heart and mind.

And last but not least, I am grateful to my fiancè , Dr. Mina Baghgar, who is

an experimental condense matter physicist. Despite her special interest in quantized energy levels, she granted me with her continuous love and support during the most hectic days of my PhD studies. I appreciate her taking the time and effort to answer my elementary questions about concepts of modern physics. I am really lucky to have her in my life.

Mohammad Javad Abdolhosseini Qomi

Cambridge, MA 2014

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Societal and Industrial Challenges

Civil infrastructure is and continues to be the backbone of our society to meet our needs for housing, transportation, water and electricity supply, and so on. However, in response to rising public concerns about sustainable and resilient infrastructure development, its functions are recently revisited to encompass roles beyond providing basic services. In fact, the modern infrastructure is assumed to actively cope with excessive greenhouse gas emissions, unreasonably high energy footprint, relatively short service life, low durability and poor resilience. These expectations are placing the United States' infrastructure at the focal point of our Civil and Environmental Engineering (CEE) community. This presents us with an exclusive opportunity to take these threats seriously and turn them into exciting venues for creative research and possibilities to educate the next generation of civil and environmental engineers. These opportunities are disseminated across the entire infrastructure landscape, spanning several length scales starting from the molecular structure of construction materials to the entire global transportation network. The challenge is then to properly recognize opportunities, discern their root causes and develop practical solutions.

As mentioned above, one of the critical issues with civil infrastructure is its contribution to the global warming potential. Global warming is a detrimental threat to our national and global security. Its dire consequences include, and are not limited to,

drastic weather and climate patterns, exacerbating health issues, endangering wildlife, and rising sea levels. The synergies between global warming and climate change entail longer periods of drought in some regions and an increase in the number, duration and intensity of tropical storms in other places, while increasing the potential risk for more frequent wildfires. Heat waves claim thousands of lives across the globe and contribute to smog pollution that intensifies pollen allergies and asthma. Global warming does not only jeopardizes human lives via extreme weather events but also places several species at the verge of extinction. Finally, the melting of glaciers and associated see-level rise accompanies loss of wetlands and places our coastal landmarks at risk of annihilation. These ominous consequences call for different stakeholders ranging from policy makers to scientists and engineers to recognize this worldwide phenomenon as a global threat and take this historic opportunity to curb its unrestrained expansion.

The first step to control global warming is to identify its contributing factors. Most climate scientists agree the main cause of global warming phenomenon is anthropogenic emissions that block the thermal radiations bouncing from earth's surface toward the outer space. The greenhouse gas emissions in 2014 exceed 5,000 million metric tones of carbon dioxide equivalents. Carbon dioxide, methane and nitrous oxide respectively constitute 85%, 8% and 5% of these emissions [Ref: Fig 1-1 from World Resources Institute]. In United States, all sectors supply greenhouse gas discharge to the atmosphere with transportation, electricity and heat, industrial applications and processes, and agriculture and waste accounting respectively for 27.2%, 32.4%, 31.6%, and 8.8% of national emissions [Ref: Fig 1-1 from World Resources Institute]. This holds civil infrastructure, the combination of transportation network, residential and commercial buildings, and productions of construction materials, cement and steel, accountable for more than 60% of national CO_2 emissions.

More specifically, the cement industry faces strict challenges ahead of itself to limit its carbon footprint. The two major sources of emissions in this industry are associated with energy-intensive manufacturing process and calcination of limestone ($CaCO_3$) in a cement kiln. Combined together, manufacturing every ton of cement involves discharging one cubic meter of CO_2 to the atmosphere. This does not per

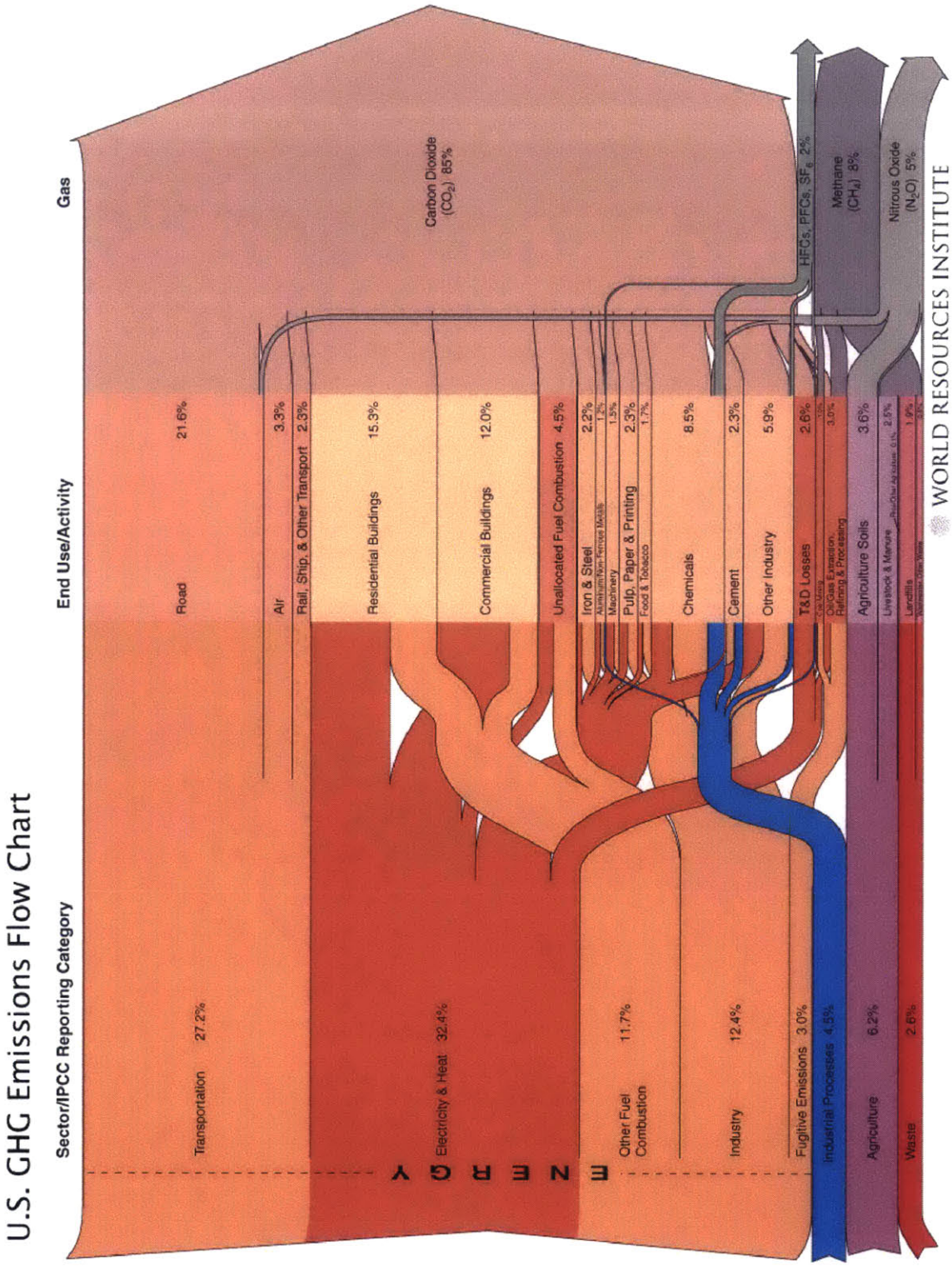

Figure 1-1: United States greenhouse gas (GHG) emissions flowchart across different sectors and end use activity in 2010, courtesy of World Resources Institute.

se make cement a good or a bad material. It is astonishing to know that concrete, the structural backbone of our infrastructure, is the most-used man-made material on Earth. Every human being on earth consumes on average one cubic meter of concrete every year. In fact, it is this unrestrained bulk consumption of concrete that makes it responsible for 4-5% of global and 2.3% of our national CO_2 emissions[358]. Therefore, it is imperative to search for novel approaches to increase bulk properties of cementitious materials in general. To this end, I dedicate part of this thesis to develop a general framework to study the impact of modulating chemistry of cementitious materials on their macroscopic bulk properties.

Another source of infrastructure-related emissions are closely tied to our ever-increasing demand for electric and heating energy consumption in residential and commercial sectors. According to a recent survey, 44% of energy consumption in buildings is used for space heating and cooling, which accounts for 20% of the national CO_2 emissions[2]. Currently, the White House seeks approaches to lower energy consumption of buildings by 40% and reduce the associated greenhouse gas emissions by 150 million metric tons till 2020[3]. At this policy level, it is not always straightforward to discern the fundamental cause of an infrastructure dilemma. As it turns out, several factors affect energy use in buildings, such as human behavior, thermal characteristics of buildings' envelope, HVAC systems, and many others. Therefore, it is instructive to have a quantitative understanding of influential parameters and their impact on network level properties prior to embarking on a specific solution approach. For instance, is it cost-effective to improve the thermal resistance of insulation materials among hundreds of other factors for retrofit purposes? If so, how much decrease in thermal conductance would have a meaningful impact on energy consumption at the city scale? I allocate the other part of this thesis to this class of societal problems with an intention to construct a quantitative model to leverage science-informed decision making in promoting sustainable development.

To conclude, the main industrial and societal challenge here is to mitigate greenhouse gas emissions associated with civil infrastructure. These emissions can be in the form of embedded carbon footprint of construction materials or use phase emissions

associated with energy consumptions.

1.2 Formulating the Science Problems

Now that the industrial and societal problems are identified, the challenge is then to translate them to a set of physically sound science questions, so that answering them could ultimately pave the way to mitigate the discharge of greenhouse gases to the atmosphere. Here, we elaborate on three classes of science problems that have direct implications on sustainability of civil infrastructure. These science problems are usually at the interface of different fields ranging from physics and chemistry, to materials, civil, mechanical, and systems engineering.

The first class of problems tries to enhance the sustainability of cemetitious materials by reducing their carbon footprint and enhancing their bulk properties. Cement's carbon footprint is directly correlated to limestone consumption in the manufacturing process. During limestone's calcination reaction in cement kiln, $CaCO_3$ dissociated into carbon dioxide and calcium oxide. This calcium oxide is the source of calcium in cement clinker phases, mainly different polymorphs of di- and tri-calcium-silicates. Subsequently, this calcium will be involved during cement's hydration process that precipitates calcium-silicate-hydrate (C-S-H). C-S-H is the glue of cement paste responsible for its stiffness, strength, and aging properties. Now that the connection between the limestone consumption and the origins of mechanical properties in cement paste is brought forward, a research question can be formulated as follows: *what does decreasing the limestone usage, and subsequent calcium content, affect mechanical properties of cement paste?* In other words, does modulating the calcium content in the cement paste aggravate or enhance the stiffness, strength, and fracture resistant properties of cement paste? In fact, if it is shown that mechanical properties of the paste do not vary significantly with calcium content, then a factor of β decrease in calcium consumption will result in a β times reduction in emissions. On the other hand, while the emissions scale with volume, the strength of concrete columns, arches and domes, scale with cross sectional area. Therefore, the specific strength of

purely compressive members scales with inverse of length, L^{-1}. This suggests that if the strength of cement paste scales with δ, the carbon emissions would eventually scale with δ^{-1} to maintain the same strength level by resizing the cross section. This reduction in carbon emissions is in addition to the reduction of limestone consumption with the β factor. It is the ultimate goal of my scientific pursuit to identify the correlation between these two factors, β and δ, through a coupled chemo-mechanics analysis.

Second, cement and concrete are multifaceted materials, meaning that their mechanical properties are not the only concerns in the design and construction of civil infrastructure. In fact, durability issues in cementitious materials continue to endanger the serviceability of concrete structures. This triggers the next set of science questions, aiming to explore *the impact of modulating the chemistry of C-S-H on its equilibrium and non-equilibrium physical properties including water and ionic mobility and transport properties.* In other words, are water molecules and cations bound to the surface of the atomic structure of C-S-H or are they free to diffuse within the porous structure of cement paste? More specifically, what is the role of calcium content of C-S-H on the mobility of atomic and molecular species? Another source of concern is cement paste's thermal properties, heat capacity and heat conductivity, which appear to be important during both construction and use phases. The thermally-driven early-age cracking during cement's exothermic hydration process imperils the long-term serviceability of concrete structure. Thermal properties of cement paste affect energy consumption of buildings through heat conduction and thermal mass of the building envelope, the tendency of the building to retain a constant temperature despite outdoor temperature variations. This motivates another set of science problems that try to examine *the impact of chemical modification of C-S-H on its thermo-physical properties.* In other words, how does calcium content affect the heat capacity and thermal conductivity of cement paste? Also, how does this thermal properties evolve during the hydration process? Therefore, another goal of this thesis is to provide a consistent image of the interplay between the chemistry and physics of cement paste.

The third and the last set of problems that we try to address in this thesis pertains to better understanding and hence proposing novel approaches to control excessive energy loss of civil infrastructures at the city scale. Here, the thesis focus is mainly shifted to energy losses associated with thermal conditioning of buildings including both heating and cooling processes. The fundamental challenge is to find *the optimum way that reduces energy losses and the associated carbon footprint of cities through retrofitting its building blocks.* This problem is at the interface of engineering mechanics and systems science as it aims at manipulating the systems level properties through modulating energy losses in elementary building blocks of cities, *i.e.* residential and commercial units. At the building level, the thermal energy losses are affected by several factors, turning it to a complex system that can be regarded as a function with many inputs parameters. *The science challenge would then transform to reducing the complexity of the elementary units to a simplified high-fidelity model that is only affected by a subset of most influential factors, rather than the entire input space.* Answering this challenge is extremely informative and provides the solution to several intermediate questions regarding the importance and impact of different retrofit strategies at material, building and ultimately city levels. For instance, is improving thermal properties of construction materials a necessity for reducing the thermal efficiency of cities? If so, what are the minimum thermal characteristics of a new construction material to place it as a "game-changing" material on the market? Also, how does human behavior affect energy losses in buildings and how to disentangle the thermal efficiency of a unit from residents' behavior? The answer to these questions not only highlight the role of different factors but also facilitate identifying the fastest and most energy efficient way to retrofit a city's building block.

1.3 Thesis Approach and Outline

The above science questions can be rationalized in a multiscale bottom-up framework that involves several length scales starting from the molecular structure of materials at the nanoscale, all the way to the policy and decision forming level at the city

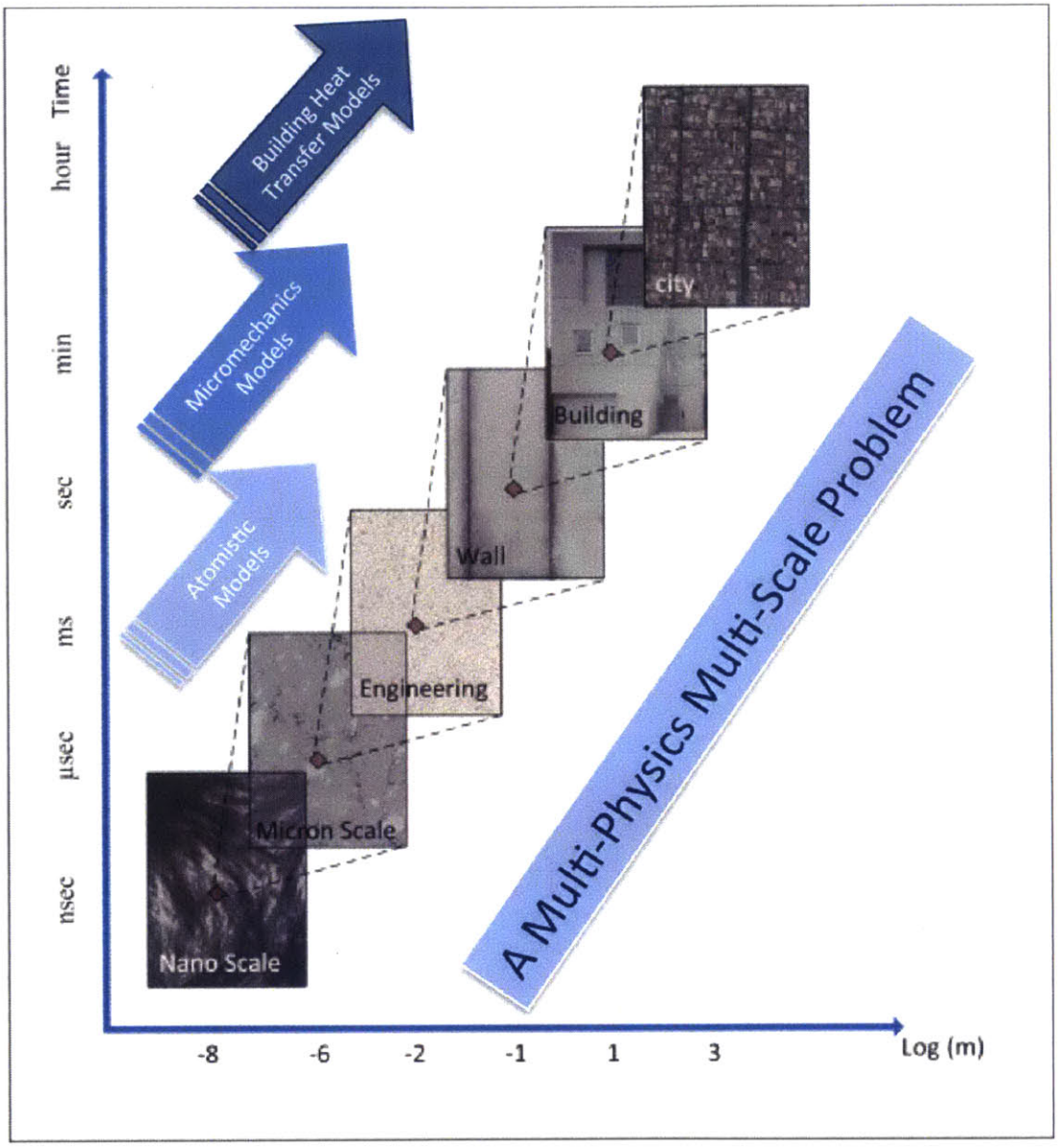

Figure 1-2: Multiscale bottom-up framework for hierarchical infrastructural information flow starting from thermal proeperties of molecular constituents of cement paste all the way to heat loss mechanisms at the city scale.

scale. As discussed extensively in mean field theory, the scale separability condition is absolutely essential in upscaling properties as it ensures the characteristic length l at a particular scale is much smaller than that of the subsequent scale L ($l \ll L$). Given that the scale separability condition is met, a hierarchical multiscale framework provides a versatile tool for information passage in between adjacent scales (see Fig. 1-2). In other words, the ultimate goal of analysis at a particular scale is to provide the necessary information required for better understanding of the physical phenomenon at the larger scale. The hierarchical multiscale modeling approach has a lot in common with renormalization group theory. In fact, both approaches assert that some information becomes increasingly irrelevant as the magnification power of the notional microscope decreases. This means that the information content decreases as the characteristic length scale becomes larger. For instance, the molecular structure of cement paste at the nanoscale is irrelevant in engineering design process. However, the macroscopic strength properties, that are somehow the consequence of molecular arrangements and interatomic interactions, play a substantial role in the design of concrete structure against failure. Here, the aim is to fill this gap by providing a rational framework to sieve the abundant information at the nanoscale and obtain observables of interest at the macroscale.

We use diverse sets of tools ranging from statistical physics and continuum mechanics to complex systems analysis to study each scale in depth. At the nanoscale, we employ tools of computational physics including molecular dynamics, energy minimization, and phonon analysis to unravel the correlation between the atomic structure of cementitious materials and their mechano-physical properties. Subsequently, we employ conventional tools of microporomechanics to upscale physical properties from the nanoscale to the macroscale and compare the upscaled results with experimental measurements. This summarizes my toolbox for studying materials across length scales. Moving toward the system level properties, we employ ideas in the design of experiment, sensitivity analysis, and surrogate modeling and optimization, to connect the energy losses at the city scale to the thermal properties of construction materials at the material's macroscale. The thesis is divided into four chapters. The first two

chapters focus on nanoscale properties of C-S-H. The subsequent chapter connects the nanoscale properties to that of macroscale measurements. The last chapter is dedicated to the study of heat losses in buildings and cities. In the following, we provide a concise overview of every chapter.

Chapter 2 investigates the effect of calcium content or the calcium-to-silicon (Ca/Si) ratio on the nanostructure and nanomechanics of C-S-H. This is achieved by generating a database of 150 C-S-H models with Ca/Si ratio varying between 1.1 to 2.1. There are several steps involved in generating this database. The details of the model construction are extensively discussed in Appendix A. The interested readers are encouraged to study this appendix before proceeding to this chapter. Subsequently, C-S-H models are used to explore the effect of Ca/Si ratio on the nanotexture of C-S-H and comparing it with available experimental data. This is followed by the analysis of mechanical properties of C-S-H, including both elastic and strength properties and their comparison with experimental coupled nano-chemo-mechanical analysis via wavelength dispersive spectroscopy and nanoindentation. The details of the experimental data are provided in Appendix B. Later on, the anomaly in indentation modulus-to-hardness ratio is thoroughly investigated in a dual-defect framework, interpreting the results in terms of both Ca/Si ratio and medium-range order in Ca-O and Si-O bonding in C-S-H. Afterward, this combinatorial database is juxtaposed with Richardson's standard T/CH models and the points of agreement and disparity are brought forward. Finally, discussions on the implications of findings close this chapter.

Chapter 3 probes the effect of Ca/Si ratio on the structure and dynamics of nanoconfined water in the interlayer spacing of C-S-H. The molecular models employed in this chapter are the same as those initially constructed in chapter 2. First, the structural properties of nanoconfined water including the number of hydrogen bonds, dipole moment and Voronoi density of water molecules are studied in all 150 C-S-H models. Subsequently, the anisotropy and inhomogeneity of confined water dynamics are explored by means of mean square displacement analysis. This inhomogeneity in dynamical properties is further characterized through Van Hove space-

time correlation function. This helps us unravel the size and residence times of water molecules in the dynamical cage. Later on, it is shown that the dynamical properties of ultraconfined water in C-S-H are composition-dependent and the self-diffusivity of mobile water molecules are estimated and compared with experimental observations. The conclusive remarks close this chapter.

Chapter 4 deals with thermal properties of cementitious materials across length scales. This chapter starts with a detailed description of the multiscale and multiphase porous structure of cement paste across length scales. The different calcium silicate phases in cement paste, including both clinker and hydration phases, are described afterward. Since the CSH-FF potential used in this thesis was designed to reproduce the structure and elastic properties of C-S-H models, the transferability of this potentials to other crystalline phases is performed afterward and comparisons are made with previous atomistic simulations and experiments. The vibrational density of states of all phases are calculated and analyzed in detail, subsequently. The thermodynamic properties such as heat capacity at constant volume and pressure, coefficient of thermal expansion, and compressibility are calculated next from the density of states and incremental molecular dynamics simulation. The Green-Kubo formalism is employed thereafter to calculate the heat transport properties from equilibrium molecular dynamics simulation. Mean-field homogenization techniques are used later on to upscale heat capacity and conductivity values at the nanoscale to that at the macroscale. These macroscopic quantities are monitored against the hydration degree of cement paste and compared with experimental measurements. This chapter is summarized with some conclusive comments.

Chapter 5 discusses the heat loss mechanisms from buildings and their aggregated form at the city scale. This chapter begins with an overview of the several parameters that affect building heating and cooling energy losses. The interested readers that are not familiar with building energy simulations are refered to Appendix F for an extensive discussion on building energy loss calculations. A global sensitivity analysis is performed afterward to identify the most crucial factors that affect building energy losses. To this end, a Monte Carlo design is employed to propagate uncertainty

from the space of parameters to space of energy losses. This input-output relation is analyzed via the Spearman Partial Rank Correlation Coefficient to identify the aforementioned crucial factors. The interested readers are referred to Appendix D and Appenix C for fundamental definitions and mathematical concepts in design of experiment and sensitivity analysis, respectively. Subsequently, the set of influential parameters are employed to construct a reduced order model of building energy losses via both machine learning and dimensional analysis techniques. Extra discussions of surrogate analysis and design are provided in Appendix E. This study is further enriched by employing real energy consumption data for more than 6,200 buildings in Cambridge, MA. The detailed investigation of the data reveals the intrinsic characteristics of human behavior and building thermal efficiency. Together with the reduced energy loss model, a power-law distribution of potential building energy savings is found that paves the way to find the shortest path to retrofit a city's building block. The conclusions discuss the implication of our findings in this chapter.

Chapter 6 presents the conculsions of this thesis by both summarizing the main findings of each chapter and explaining how they contribute to a new level of understanding about infrastructure materials and systems. Finally, comments on future research and developement directions close the thesis.

Chapter 2

Nanostructure and Nanomechanics of C-S-H

The more than 20 billion tons of concrete produced annually contributes 5-10% to the worldwide anthropogenic carbon dioxide production [358]. One strategy to reduce this environmental footprint is to enhance concrete's specific stiffness or strength [12, 344] by optimizing the molecular-level properties of the calcium-silicate-hydrates [260, 17, 98, 211] (C-S-H), the binding phase of concrete, validated against an array of nano-texture and nanomechanical experiments. In this scheme, a combinatorial database of atomic configurations of C-S-H is generated by simulation, with each configuration having a well-defined set of defect attributes as well as a set of corresponding mechanical properties such as elastic modulus and hardness. Optimization then consists of screening the database for the desired properties against the defect attributes, essentially in the spirit of structure-property correlation. C-S-H paste is comprised of small nano-particles of 5 nm average diameter [17], products of reactions between anhydrous calcium silicates and water that form a gel-like network of variable stoichiometry [285]. Recently, Pellenq and coworkers reported a model molecular structure of C-S-H, attained through computational simulations that were consistent with an experimentally measured average composition, density, scattering and spectroscopic signatures [260]. It is worth mentioning that while Dolado et al. [98] adopted a glass quenching approach to produce an amorphous structure of C-S-

H, we chose to start from the crystalline molecular structure of Tobermorite. In fact, at low C/S ratios, the method of Dolado et al. yields a disordered glassy structure for C-S-H, while our approach leads to dominant layered (hence more crystalline) signatures as seen in experiments. In the present chapter, this model is used to create the database of atomic configurations and corresponding defect attributes and mechanical properties, for a wide range of C-S-H chemical compositions. Before using this database to screen for optimum mechanical behavior, we compare the simulation results to available experiments, consisting of: drying measurements, Small Angle Neutron Scattering (SANS), Inelastic Neutron Scattering (INS), solid-state Nuclear Magnetic Resonance (NMR), and our own wavelength dispersive spectroscopy (WDS), nanoindentation and Transmission Electron Microscopy (TEM) experiments. These comparisons provide model validation and provide insights into the system-level properties of the ensemble. We are especially interested in the effects on the mechanical behavior of two types of defects. The first is the calcium to silica ratio, denoted as C/S ratio, which is well-known in cement chemistry. It can be defined as a measure of points defects (vacancies) in the silicate network [328]. The second type of defects pertains to variations that exist in the medium-range environment of the Si-O and Ca-O networks. It is most simply defined in terms of the first sharp diffraction peak in the Si-O and Ca-O partial structure factor, which is familiar in the study of silica glasses [108].

2.1 Generating the Database of C-S-H Models

The database for the combinatorial screening of mechanical behavior is obtained by creating an ensemble of atomic structures of C-S-H with each member characterized by a known value of the calcium to silica ratio, C/S, over the range of 1.1 to 2.1; (see Appendix A for comprehensive discussions). This is achieved by randomly cutting silica chains (removing a number of charge neutral SiO_2 groups from 11Å Tobermorite) to increase the C/S ratio; allowing an account for reactivity through empirical reactive potentials in the course of this procedure. In applying combinatorial optimization

to better understand the relation between defect attributes and mechanical behavior of C-S-H hydrates, we search the generated database for atomic configurations that have optimum mechanical properties, and meanwhile we determine the defect attributes specifying these configurations. By mechanical properties, we mean the ratio of indentation elastic modulus (M) to indentation hardness (H). This M/H ratio is distinct from indentation toughness that refers typically to fracture toughness and depends also on indentation crack length [20]; here, high M/H describes a material of low elastic strain limit. By defect attributes, we consider C/S as an overall measure of chemical composition and two correlation lengths characterizing medium-range environments of Si-O and Ca-O networks. It is not known a priori whether atomic configurations having optimum M/H actually exist, and if so, which defect attributes are relevant. Thus, we begin our screening by first considering only the defect attribute C/S since the generated database allowed for variations in atomic configurations at fixed values of C/S (see details in Appendix A for details). While others [209, 211, 259, 305] have calculated elastic properties of distinct mineral phases such as Tobermorite polymorphs and Jennite via using atomistic simulation methods, our approach provides a comprehensive screening of mechanical properties for the C-S-H phase as a function of its chemical composition. Regarding experimental aspects, while correlations between chemical composition and mechanical properties of synthetic C-S-H gels [37, 266, 206, 63, 16] (fully cured or calcium leached cement pastes [81]) have been previously reported, our approach provides a venue to directly assess these mechanical properties of the elementary C-S-H nano-particle as a function of C/S ratio.

To be specific about the different defect attributes considered, we show a typical atomic configuration of Tobermorite 11Å [139] in Fig. 2-1.a. The unit cell contains the silica chains Fig. 2-1.b and Ca atoms in two distinct environments, intralayer Ca Fig. 2-1.c and interlayer Ca in the interlayer spacing (between adjacent calcium-silicate sheets as highlighted in Fig. 2-1.d). Hereafter, the intralayer and interlayer calcium atoms are referred to as Ca and Cw, respectively. We introduce two correlation lengths λ and λ' (see Fig. 2-1.a). Here, λ is the medium-range correlation length

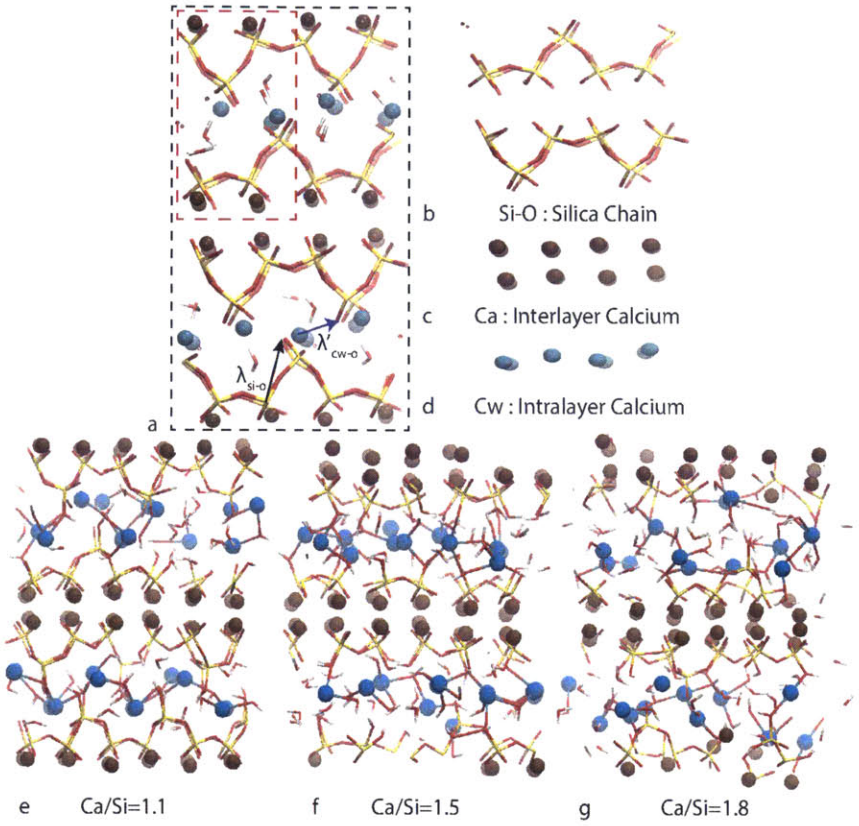

Figure 2-1: Effect of C/S ratio on the molecular structure of C-S-H at the nanoscale. a) The unit cell of Tobermorite 11Å is enclosed by a black dashed line. The brown and cyan spheres represent intra and interlayer calcium ions, respectively. Red and yellow sticks depict Si-O bonds in a silicate tetrahedra. White and red sticks display the hydroxyl groups and water molecules. By repetition of unit cells in all lattice vectors, a (2*2*2) super cell of the molecular structure of Tobermorite 11Å is constructed for representation and is outlined by dashed red line. The medium-range correlation lengths λ and λ' which pertain to Si-O and Cw-O network are represented by dashed black and blue arrows, respectively. The solid skeleton of Tobermorite is consisted of 3 parts b) Silica chains, c) Calcium interlayer and d) Calcium Intralayer. e-g) Molecular model of C-S-H at C/S=1.1, 1.5 and 1.8. e) At C/S=1.1, a 2D lamellar structure is presented with minor defects in silica chains, reminiscent to that of 11Å Tobermorite. The interlayer regions contain counter charge-balancing calcium ions, hydroxyl groups and water molecules. f) At C/S=1.5, several bridging tetrahedra are removed from the silicate chains. The interlayer calcium ions are still organized in well-defined sites. g) The C/S ratio is further increased to 1.8 by removal of more silica tetrahedra. This indicates that, from low to high C/S ratio, the C-S-H's molecular structure changes from 2D crystalline layered to a more amorphous structure.

as measured on Si-O-Si-O motifs and defined for silica glasses [108] [107]. Similarly, λ' is the medium-range correlation length measured on Cw-O-Cw-O motifs. Both λ and λ' are present at around 4.3Å and 4.6Å in experimental total radial distribution function [224]. We will show that each of these two correlation lengths plays a different role in its influence on the mechanical properties, since they pertain to two distinct medium-range order environments in the C-S-H atomic structure.

Our database for the combinatorial screening of mechanical behavior is obtained by creating an ensemble of atomic structures of C-S-H with each member characterized by a known value of the calcium to silica ratio, C/S, lying in the range [1.1, 2.1]. The procedure used to generate this ensemble consists of seven steps of atomistic-scale simulations, which we summarize here (See Appendix A for details). First, a (3x2x2) supercell of the atomic structure of Tobermorite 11Å after Hamid [139] is constructed. With the silicate chains having no defects, there are no hydroxyl groups in the system at this stage. Also, the chains are infinitely long. Then, all water molecules are removed from the interlayer spacing in the super-cell. In step three, 150 samples are prepared by randomly cutting the silica chains (removing a number of charge neutral SiO2 groups). Each removal causes the C/S ratio to increase. By cutting the chain at different locations, several samples with the same C/S ratio are thus generated. In this step, the interlayer calcium atoms are first allowed to relax by energy minimization using CSH-FF potential [306], followed by all the other atoms and the supercell dimensions. In step four, water molecules are introduced back using a Grand Canonical Monte Carlo method using the same potential, simulating equilibrium with bulk water at room temperature. At step five, the inserted water molecules are allowed to react with the interlayer calcium and the silica groups by running semi-classical molecular dynamics simulations using the ReaxFF potential [212]. To accelerate the reaction, the system temperature is raised to 500 K, which is well below the melting point of C-S-H. At this stage, the results show that some of the interlayer water molecules dissociate to hydroxyl groups and protons. Minor condensation/dissociation of silica chains are also observed in some samples. Of the 150 samples generated, all were amenable to equilibration within

ReaxFF over the finite trajectory duration. In step six, a comprehensive topological analysis is performed to identify the local environment of species, in the process of distinguishing between different types of oxygen, hydrogen and calcium atoms. This facilitates the transition from the use of ReaxFF to the non-reactive potential, CSH-FF, which is well suited to the study the mechanical properties of C-S-H phases [306]. In the final step, a 3ns-long simulated annealing is performed on each sample to bring the temperature from 500 K to 300 K at ambient pressure, using the CSH-FF potential. The annealing procedure consists of 1 ns simulation at 500 K followed by a 1 ns ramp to decrease the temperature to 300 K and additional 1 ns at 300 K for relaxation. All the chemical and structural characterizations along with the mechanical properties (indentation modulus and hardness) are calculated from the atomic configurations prepared by this procedure. Fig. 2-1.e-g shows the sensitivity of atomic configurations to variations in C/S ratio in the range of 1.1 to 1.8. Keeping in mind our procedure of removing silica groups in order to increase C/S ratio, we see a progression from a well-ordered 2D lamellar structure at C/S = 1.1 to a more disordered structure at C/S=1.8. In particular, the structural surroundings of the interlayer calcium atoms (Cw) as the C/S ratio increases are of prime importance, as this illustrates the subtle effects of introducing vacancies in the silica chains.

2.2 Effect of C/S on the Nanotexture of C-S-H

Predicted structural properties of C-S-H with C/S ratio are presented in Fig. 2-2, and compared directly against experiments. The number of initially absorbed water molecules is predicted to scale linearly with the C/S ratio Fig. 2-2.a, a behavior that is found in both SANS [17] and drying [78] experiments. There is also consistency with the notion that a removed SiO_2 unit occupies the volume of approximately two H_2O molecules. As a result of using a reactive force field to model the interactions with absorbed water, a part of the initially grand canonically adsorbed water remains structural molecular water with a composition-dependent dynamical anomaly and glassy nature [274, 368] Fig. 2-2.a, while a considerable amount of water molecules

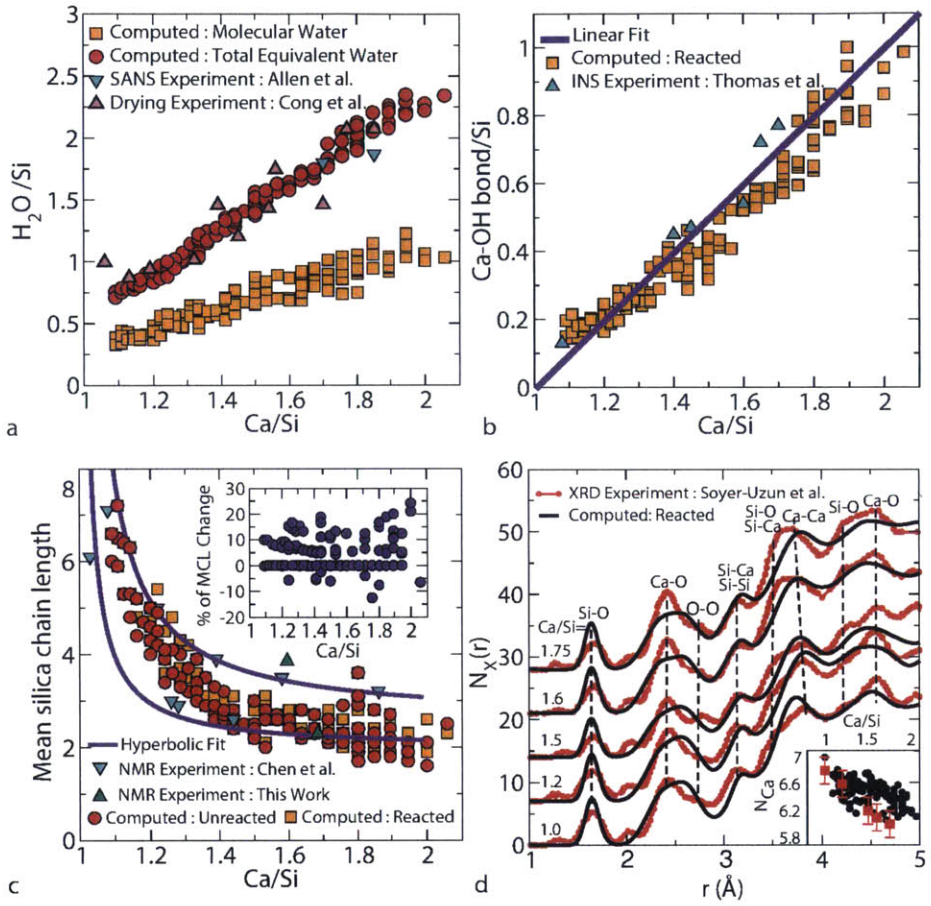

Figure 2-2: Effect of C/S on the nano-texture of C-S-H. a) The state of water in C-S-H interlayers. The total equivalent water contains both the hydroxyl groups and molecular water in the interlayer spacing. The water content is comparable with total equivalent water measured in SANS performed by Allen et al.[17] and a set of controlled drying experiments done by Cong and Kirkpatrick [78]. b) Number of Ca-OH bonds measured via topological analysis and its comparison with INS experiments measured by Thomas et al. [332]. c) The effect of C/S ratio on the mean silicate chain length compared to NMR experiments carried out by Chen et al. [65]. The inset presents the variation of MCL before and after reactive modeling. About 20% of molecular models exhibit extra silica condensation and 5% show silica chains dissociation. d) The total pair correlation function and the comparison with X-ray diffraction experiments of [318]. The inset provides the comparison between coordination number of the calcium ion as calculated from atomistic simulation and measured from X-ray diffraction [318].

dissociate. The hydroxyl groups predominantly react with interlayer calcium ions to form Ca-OH bonds. The protons bond to non-bridging oxygen of defective silicate chains. To a lesser extent, Ca-OH bonds are also produced by condensation of silica chains, which release an oxygen atom that combines with H^+ to form a hydroxyl group. As a consequence of both mechanisms, we find that the number of Ca-OH bonds per Si atom increases linearly with the C/S ratio, Fig. 2-2.b, including zero Ca-OH bond for C/S=1, corresponding to 11Å Tobermorite. These predictions are validated by INS data [332], and provide evidence that the present combinatorial simulation approach is able to describe the local water and Ca environments in C-S-H. The level of agreement achieved between simulations and experiments is largely due to the simulated annealing step incorporated in our database generation procedure (see Appendix A for further details). This means that the stoichiometry of the solid C-S-H phase can be essentially predicted from atomistic simulations, in the form of:

$$Ca_xSiO_{2.75}(OH^{Ca})_{0.85(x-1)+a}(OH^{Si})_{1.15(x-1)+0.5+b}\cdot[0.8(x-1)+0.3+c]H_2O \qquad (2.1)$$

where x is the C/S ratio and a, b, c represent the variability in the nanostructure of C-S-H at a given C/S ratio ($a = b = c = 0$ corresponds to an average pattern for a given C/S ratio and $|a|, |a|, |c| < 0.2$ to only account for polymorphism). The variability in the structure of C-S-H is due to the change in the structure of calcium-silicate backbone at a given C/S ratio. Note that (OH^{Ca}) and (OH^{Si}) show hydroxyl groups in calcium-hydroxide and silanol groups, respectively. In subsection 2.5, we fully compare our approach to that of the T/CH model proposed Richardson [284, 287, 285, 283]. We show that our combinatorial approach to C-S-H not only provide a quantitative agreement with textural data (water content, Ca-OH amount, mean silica chain length and X-ray diffraction pattern) but also has the ability to predict the mechanical properties (elasticity and strength). We term these configurations that share a given C/S composition but differ in atomic level structural details as polymorphs, and later discuss the implications of C-S-H polymorphism. The mean chain length (MCL) of silicates in C-S-H, representative of the degree of polymer-

ization of silicate monomers, is found to decrease non-linearly with C/S Fig. 2-2.c; there is also a considerable range in the chain length for a set of polymorphs (e.g. see C/S=1/8). While some silicate groups, especially monomers, are found through reactive modeling to condense to form longer chains (see Fig. 2-2.c inset and MCL behavior in the inset of Fig. 2-2.c), the amount of monomers at a given C/S before and after condensation reaction varies by less than 15% (see Appendix A for discussions). The existence of a range of chain lengths, including monomers, indicated by our simulations is consistent with ^{29}Si NMR experiments [65] and is the basis of C-S-H polymorphs. Fig. 2-2.d compares simulation and experimental synchrotron X-ray data for the total pair correlation function [318]. Both are qualitatively in good agreement. Then simulation data refines the position of all physical correlation peaks in $N_x(r)$ function for different C/S ratios. However, we note that the Ca-O correlation peak is constantly broader in simulation. But, this does not affect the calculation of the calcium coordination number that is in quantitative agreement with experiment as shown in the inset of Fig. 2-2.d. We note that the existence of secondary small features in the experimental data that correspond to no identifiable correlation distances can be the results of truncation error in the Fourier transform of the original scattering data. An extensive discussion on the calculation procedure of $N_x(r)$ is provided in the Supplementary Methods.

2.3 Effect of C/S on the Mechanical Properties of C-S-H

It is the central goal of this chapter to probe the relation between mechanical properties (stiffness and hardness) and the various defect attributes via a combinatorial database established by atomistic simulations. We are interested in the average elastic properties of the C-S-H solid particles expressed in terms of the indentation modulus (M) which is determined from the full compliance tensor of each C-S-H model, and evaluated by $M = 4G(3K + G)/(3K + 4G)$, where G and K are Voigt-Reuss-Hill

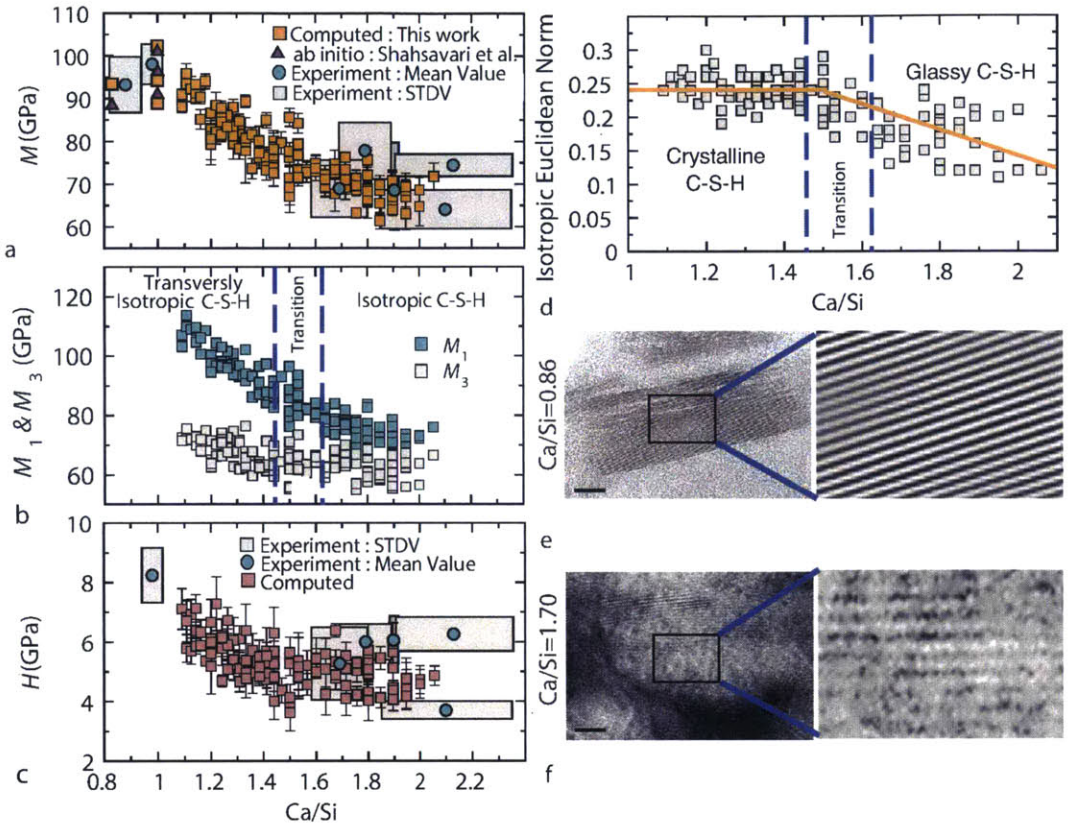

Figure 2-3: Effect of C/S ratio on the mechanical properties of C-S-H at nanoscale. a) C-S-H's solid phase indentation modulus. The computational data (this work, orange squares) are compared with coupled nano-indentation and WDS experiments (cyan circles) together with their confidence bounds (gray squares) and previous ab initio calculations on 11 and 14Å Tobermorite [139]. b) Indentation modulus parallel (M_1) and perpendicular (M_3) to the layers. c) C-S-H's solid phase hardness. The computed data (brown squares) are compared with experimental values following the same convention as in (a). d) Computed isotropic Euclidean norm as an indication of the level of anisotropy in C-S-H. Orange lines are a guide for the eyes. e) TEM image of C-S-H at C/S=0.86 [132]. f) TEM image of C-S-H at C/S=1.7 produced from hydration of C_3S. The error bars in atomistic simulations are calculated via sampling the phase space at independent time frames. The experimental error bars are measured via the statistical clustering technique coupling nanoindentation and WDS data.

shear and bulk moduli, respectively (see Appendix A for mathematical relations). In our atomistic simulations, hardness (H) was determined by using the Mohr-Coulomb failure criterion, applicable to cohesive-frictional materials such as C-S-H [126]. This is achieved through a coupled biaxial (shear-compressive) deformation scheme in several independent deformation paths similar to those utilized for studying validity of the Cauchy-Born rule [13, 172, 171]. The hardness is calculated from the friction angle (ϕ) and cohesion (C) (see Appendix A for methodolgy). Both indentation modulus and hardness are averaged at several statistically independent time frames along the MD trajectory for each polymorph. To provide experimental validation of the predicted changes in C-S-H stiffness and hardness with chemical composition, we synthesized common industrial cement pastes with various amounts of silica additives and curing conditions (see Appendix B). We implemented a statistical chemo-mechanical clustering method of both composition and mechanical properties at the sub-μm scale (see Appendix B for details). The composition for each cement paste was determined by clustering X-ray wavelength dispersive spectroscopy (WDS) data comprising 400 nanoscale volumes (voxels) per sample. Similarly, the indentation modulus and hardness were quantified, at similar length scales, via clustering analysis of nanoindentation results across the cement paste sample surface [126], sampling 400 voxels per sample, while ensuring the compatibility of chemical and mechanical clustering results. The composite properties thus obtained were corrected for the effect of mesoscale porosity to arrive at the properties of the solid C-S-H nanoparticle [81] for direct comparison with the simulation data. Sensitivity analysis of WDS excitation voltage, interaction volumes and nanoindentation extrapolation algorithms were also conducted.

Several key features can be noted in the experimental validation of model predictions shown in Fig. 2-3. Both simulations and experiments indicate a significant decrease of the average stiffness response with increasing C/S ratio (Figs. 2-3.a and 2-3.b). It is not surprising that as C/S increases, the calcium-silicate layers becomes more defective, and as a consequence, mechanical stiffness and anisotropy decrease. A similar trend is found for the hardness, which is related to the mean pressure sustained

beneath the indenter prior to permanent deformation and is proportional to the yield strength in a material that exhibits plastic deformation (Fig. 2-3.c). Compared to typical cement hydrates prepared via usual cement dissolution [284] at a median C/S ratio of 1.7, C-S-H prepared at C/S=1 to 1.1 exhibits on average 31% and 48% superior stiffness and hardness, respectively. Furthermore, the decrease of M and H with C/S follows a bilinear trend intersecting at C/S=1.5. This suggests that C-S-H at low and high C/S ratios respond differently to defect incorporation in the silica chains. To elucidate this particular behavior, we make use of the calculated full compliance tensor of each numerical sample, and determine the orthotropic in-plane and in-layer elasticity constants, M_1 and M_3 (Supplementary Methods) that would be measured in a Hertzian elastic contact loading along orthogonal directions of the calcium-silicate layers [93]. Given the random orientation of C-S-H particles in real cement paste, the experimental assessment of these M_1 and M_3 constants is still out of reach of current indentation technology. The simulations identify a pronounced anisotropic behavior of C-S-H at the nanoscale (Fig. 2-3.b) in terms of a significant difference between the in-plane stiffness, M_1, and layer to layer-stiffness, M_3. While both M_1 and M_3 follow the trend of the indentation stiffness M (Fig. 2-3.a), we can see that the anisotropy, expressed by the ratio M_1/M_3 also decreases with the C/S ratio. That is, a highly anisotropic C-S-H at low C/S ratios (C/S<1.5) becomes gradually isotropic as long silica chains are shortened upon increasing C/S ratio. To further appreciate the impact of texture on properties, it is instructive to employ the isotropic Euclidean norm of the C-S-H stiffness tensor, defined as $d_E(C_{iso}, C_t) = \|C_iso - C_t\|_E$, where C_t and C_{iso} are the full and isotropic parts of the stiffness tensor, respectively [101]. Applied to the C-S-H models, we find that this norm is almost constant for C/S<1.5 (Fig. 2-3.d), which correlates well with the observation that C-S-H maintains a tobermoritic crystalline layered texture for low C/S ratios (see Figs. 2-1.e-g). This predominantly crystalline structure is in agreement with experimental observations by TEM30, [132] and X-ray diffraction [134]. Indeed, C-S-H at C/S=0.86 shows an apparently crystalline lamellar structure (Fig. 2-3.e). In turn, for larger C/S ratios, the Euclidean norm decreases (Fig. 2-3.d). For such compositions, C-S-H retains some long-range

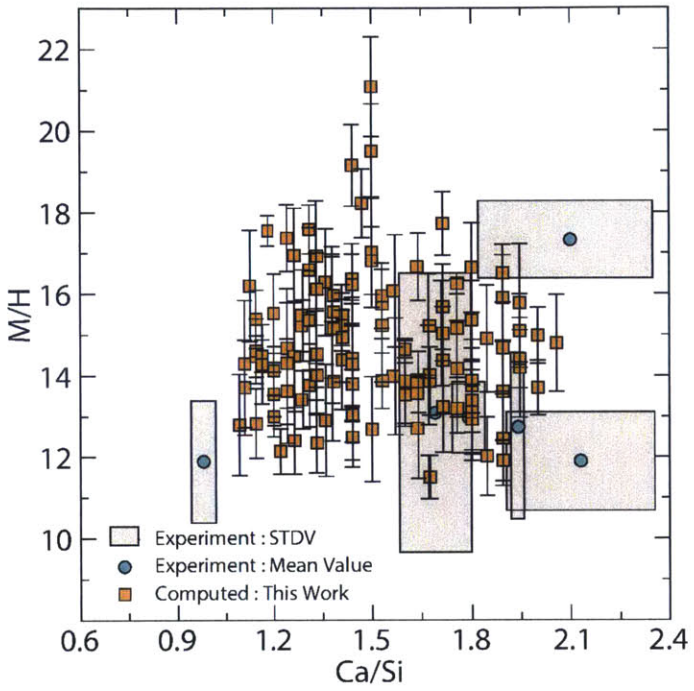

Figure 2-4: M/H as a function of Ca/Si ratio. The simulation results are compared against experimental data. The experimental M/H is calculated from measured M and H by assuming that they are normal independent quantities. The experimental error bars are measured via the statistical clustering technique coupling nanoindentation and WDS data.

layered texture [33, 32, ?] as TEM images show (Fig. 2-3.f) but the increasing amount of defects in the silica chains induces a lower short-range order that leads to textures that are mechanically isotropic (Fig. 2-3.b, $M_1 \approx M_3$).

2.4 Understanding Maximum M/H via Dual-Defect Framework

The goal of our combinatorial optimization is to find atomic configurations that give maximum toughness and correlate those molecular structures with corresponding defect attributes. To this end, we take M/H to be a measure of this property on the grounds that while we want higher stiffness (through M), we also desire higher ductility (through 1/H). The correlation of M/H with C/S ratio is shown in Fig. 2-4. Notably, some C-S-H configurations, with C/S ratio close to 1.5, exhibit maximum nanoindentation toughness. Next, we consider the thermodynamics stability of these polymorphs to consider their relative prevalence in experimentally accessible C-S-H.

Numerous compelling experimental results from this and other experimental studies [17, 65] confirm our conjecture that C-S-H at the nanoscale can exist in different molecular structures for the same oxide composition, i.e., the same C/S ratio. Specifically, results from NMR experiments [65] suggest that C-S-H at a given composition can have different MCL, and results from neutron scattering experiments[17] imply that C-S-H of a given composition can exhibit varying water content. Fig. 2-3 shows that the calculated indentation modulus and hardness exhibit a range of possible values at a fixed oxide composition for a given C/S ratio. Experimentally measured M and H (e.g., C/S 2.1 in Fig. 2-3.a, c) confirm this potential variation in stiffness and hardness for fixed oxide chemistry. Yet, the existence of polymorphism calls for thermodynamics arguments to assess the co-existence of C-S-Hs of different molecular structure at equilibrium for a given composition. This is achieved in our simulation through the calculation of the free energy of the C-S-H models considerably below their melting point via lattice harmonic approximation theory from the phonon density of state. Interestingly, the free energy content of C-S-H polymorphs is almost constant at a given C/S ratio (Supplementary Methods). This means that all polymorphs of a given C/S ratio are equi-probable at equilibrium, and thus thermodynamically competitive. This means that a targeted mechanical property at constant C/S ratio cannot be achieved through equilibrium conditions, but relies on

the kinetics of silica polymerization and associated disorder assimilation.

To consider whether the short-range structural characteristics could explain why M/H is high at a specific C/S ratio, we performed a comprehensive search for correlations among structural characteristics (bond lengths, bond angles, and coordination numbers) and mechanical properties. No such short-range correlations were identified. It was these findings that motivated us to consider medium-range correlation lengths as possible defect attributes as captured by the first sharp diffraction peak (FSDP) in covalent glasses [108, 107], which captures spatial correlations in the medium-range order (Supplementary Methods). While its origin in amorphous silicate solids is not yet fully understood, it is commonly accepted that the FSDP relates to the periodicity of the boundaries between small adjacent structural cages of SiO4 tetrahedra [223]. The associated correlation distance in real space is given by λ =7.7/QFSDP, where QFSDP is the position of the FSDP in the partial Si-O structure factor, and the constant 7.7 is the location of the first maximum of the spherical Bessel function J_0 [227]. Overall, this distance characterizes how well SiO_4 tetrahedra are packed as it corresponds to the distance between Si atoms and their second coordination shell of oxygen atoms [224, 318]. Recently, it has been pointed out that as a coarse-grained defect attribute of network glasses, λ plays a significant role in enabling structural characteristics to be correlated with transport and rigidity properties [227]. Moreover, the lack of covalent bonding and the screening effect of structural water molecules make the chemical environment surrounding interlayer calcium atoms susceptible to deformation. This leads to the localization of the deformation at the interlayer region upon application of uniform strain field [260, 9]. This further motivated us to define, in a similar fashion, the medium-range correlation length λ' in the interlayer calcium environment pertaining to the Cw-O network (see Fig. 2-3.a).

Given the fact that these structural defect attributes, (C/S, λ, λ') may not be independent, it is reasonable to ask whether a broader search by screening two or more defect attributes together could be useful. In fact, we may benefit by extending the screening to correlations of M/H against C/S, λ and λ'. For this purpose, contour plots of M/H values on surfaces of two defect attributes were considered. Fig. 2-5.a-b

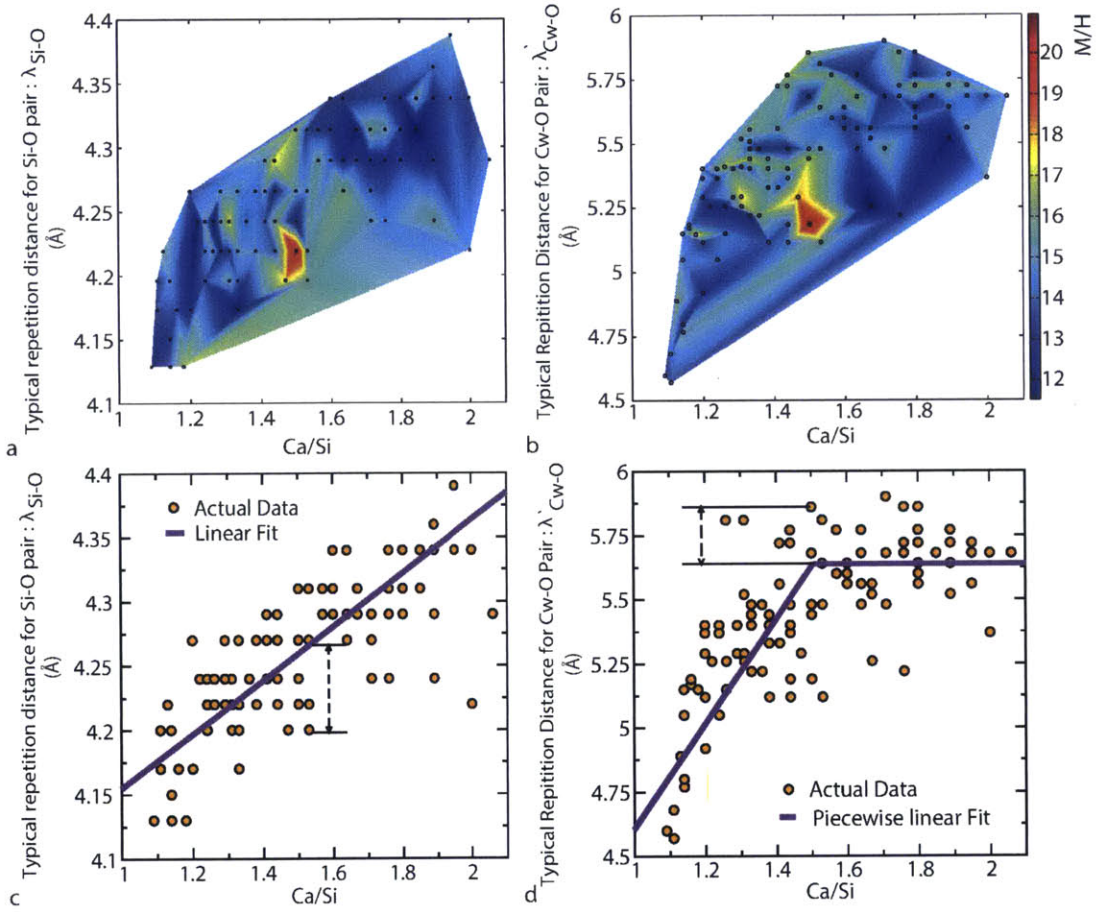

Figure 2-5: Exploring the toughness anomaly in the dual-defect framework. a) Ca/Si-λ(Si-O) and b) Ca/Si-λ'(Cw-O). The both contour plots show a region in which the toughness is maximized. The correlation among defect attributes c) Ca/S-λ(Si-O) and d) λ'(Cw-O) which mutually explain the existence of anomalous toughness configurations by packing of silica chains (responsible for higher stiffness) and openness of Cw-O network (responsible for lower hardness).

shows the contour plots (color coded) of M/H on the surface of [C/S, λ] (Fig. 2-5.a), and [C/S, λ'] (Fig. 2-5.b). In both cases, local regions of peak values are clearly identified. According to Fig. 2-5.a-b, two optimum sets of defect attributes exist, one involving (C/S, λ) and the other (C/S, λ'), which signifies that optimization of M/H can be achieved through the synergistic sensitivity to two sets of dual-defect attributes.

The notion of distributed (synergistic) sensitivity can be examined by correlating the defect attributes, specifically the relations between λ and C/S, and λ' and C/S, which are shown in Fig. 5.c-d. Within the considerable scatter in the data, it appears that λ increases essentially linearly with increasing C/S (Fig 2-5.c), while λ' increases then plateaus (Fig. 2-5.d). The increase in λ with C/S ratio corresponds to an increase in the magnitude of tetrahedron-tetrahedron relaxation in the silica chains - which is spatially distinct from the amount of nanometric SiO_2 vacancy defects that correlates with C/S. Although the range of variation of λ is rather small from 4.1 to 4.4 Å , our results are nevertheless in agreement with recent diffraction studies34 which show that λ corresponding to Si-O peak in the total diffraction pattern (obtained after Fourier transforming the total X-ray scattering signal) does shift to larger distance upon increasing C/S ratio (see Fig. 6 in ref. 12). This indicates that, for a fixed C/S, the low-lying λ points correspond to atomic configurations where the local environment of the silica chains are more compact, and they therefore contribute to higher stiffness (M). Indeed, configurations of greatest M in Fig. 2-3.a at a fixed C/S=1.5 where those that also exhibit lowest λ. Furthermore, in Fig. 2-5.d, we see that λ' reaches its maximum value signaled by a plateau at C/S > 1.5. This corresponds to configurations with relatively open Cw-O medium-range environment, which indicates there is more space for deformation. This explains the hardness plateau in Fig. 2-5.c, for relatively stable and low H for C/S > 1.5. The appearance of M/H an extremum in Fig. 2-5.a-b results from a coupled optimization of λ and λ' at C/S=1.5. Therefore, the golden C-S-H is a C-S-H at C/S=1.5 that simultaneously has a minimum in λ and maximum in λ'. However, it should be kept in mind that these two are correlated. Another way to infer the significance of the

peak region of M/H is to interpret M/H as proportional to the inverse of an effective yield strain. Minimizing this yield strain means greater capacity for plastic energy dissipation and less storage of elastic energy in the atomic configurations, which in turn makes the assembly less brittle.

2.5 Comparison of Combinatorial Models with Richardson's T/CH Model

At this stage, it is imparative for us to compare our combinatorial model with the so-called Tobermorite-'solid solution' calcium-hydroxide (T/CH) model. This model is initially proposed by Richardson and Groove [287, 288] and is intended to provide a "crystal-chemical" picture of C-S-H at different C/S starting from 14Å Tobermorite. Following the notation of Richardson [286], the chemical composition of T/CH model can be written as:

$$Ca_xH_{(6n-2x)}Si_{(3n-1)}O_{(9n-2)}.zCa(OH)_2.mH_2O \tag{2.2}$$

where x is the number of Ca^{2+} ions that to charge balance the silica chains, n is the integer number representing the Drierketten structure of the silicate chain (1 for dimer, 2 for pentamer and so on), z and m indicate the number of calcium-hydroxide and intact water molecules in the C-S-H interlayer spacing. In our combinatorial atomistic simulation approach, we start from 11Å Tobermorite and construct our samples carefully following an eight-staged process including proper consideration of chemical reactions via the ReaxFF potential [212]. As we showed in Fig. 2-2 in the main text, our simulations can reproduce drying, elastic and inelastic neutron scattering, ^{29}Si NMR experiments. Given the stoichimetery of the T/CH (eq. 2.2), the scene is properly set to compare the T/CH and the combinatorial models (eq. 2.1). Here, we discuss seven different aspects of these models and emphasize the points of agreement and disparity between them in the forthcoming paragraphs.

1) In the T/CH model, n is an integer number, which gives a mean chain length of

$(3n-1)$. Therefore, T/CH model fails to describe the variability of chain length at the nano-scale for a given C/S ratio. In fact, in a realistic cement paste, a combination of monomers, dimers, pentamers and so on exist, which gives rise to a non-integer mean chain length (MCL). Unlike T/CH, our molecular modeling, along with NMR experiments of Chen et al. [65], assert that MCL can have a value between an upper and lower bound (see Fig. 2-2.c). The variability in MCL is mathematically expressed as:

$$\frac{2y-1.75}{y-1} < \bar{MCL}(y) < \frac{2.5y-2}{y-1} \tag{2.3}$$

where $\bar{MCL}(y)$ denotes the average mean chain length at C/S= y. According to eq. (3), $\bar{MCL}(1)$ is infinite indicating Tobermoritic silica chains. At C/S> 1, $\bar{MCL}$ is between the two limits, which depend on the structural configuration of the calcium silicate backbone. We recognize that the exact prediction of MCL is controlled by the kinetics of the cement hydration process (and not through equilibrium conditions as shown by our detailed free energy calculations), which depends on curing conditions, stoichiometry of reactants, water to cement ratio along with other factors.

2) At 1 <C/S< 1.5, T/CH model proposes that by gradually removing bridging SiO_2 groups, the C/S ratio can be increased from 1 to 1.5. At C/S> 1.5, T/CH model assumes that the molecular structure of C-S-H is combination of Tobermorite devoid of a bridging silica group sandwiched with calcium-hydroxide sheets in the form of nano-scale Portlandite layers (CH). At this point, T/CH model fails to answer concerns regarding the mixing of CH with defective Tobermorite. Where are these layers located? How are they bonded to the calcium-silicate layers? What is the size distribution of these Portlandite nano-layers? What is the molecular structure of C-S-H at C/S=1.75? Recognizing these problems, we adopt a different approach, which resolves this issue. At C/S> 1.5, we start removing the pairing sites in the defected Tobermorite structure such that the number of monomers is kept minimal to agree with NMR experiments. Moreover, we run reactive molecular dynamics simulations to prompt reactions in the silicate regions to polymerize existing monomers if needed (see the inset of Fig 2-2.c). Having done this, we show that the MCL of combinatorial

structure are within the bounds observed in ^{29}Si NMR experiment by Chen et al. [65]. Therefore, our combinatorial model does not need the additional CH layers to describe the high C/S.

3) T/CH model (eq. 2.2) does not provide any relation between the C/S and m (the structural water content). In fact, T/CH model fails to provide quantitative or even qualitative description for the correlation between the defect content in C-S-H and the water adsorbed in the interlayer spacing and voids created by removing silica groups. By contrast, our combinatorial model provides a clear image and identifies a linear correlation between the water content and C/S ratio. This is achieved by performing Grand Canonical Monte Carlo simulation to adsorb water in C-S-H upon creation of defects in the structure of Tobermorite. As discussed in Fig. 2-2.a, the total equivalent water (comprising both structural water and hydroxyl groups) matches perfectly with drying [78] and elastic neutron scattering [17] experiments. We emphasize that both neutron scattering and drying experiments explore the total hydrogen content rather than the hydrogen in structural water molecules per se. This indicates that our combinatorial approach and the proposed chemical composition fully display the interplay between the chemistry of calcium-silicate backbone and the water content, which is missing in T/CH model.

4) The T/CH model (eq. 2.2) does not provide any relation between the C/S ratio and z (the number of calcium-hydroxide bonds). In fact, Thomas et al. [332] showed that there is a linear correlation between Ca-OH content and C/S. Our reactive molecular dynamics simulation without any tuning parameters indicated the very same linear behavior (see Fig. 2-2. b). This means that the environment of Ca is fully captured in our model without the need to introduce artificial CH in the composition. Given that our composition captures both total equivalent H_2O and Ca-OH content, this entails that Si-OH (silanol) content should also be properly captured in our model simply because of conservation of mass. However, there are no experimental data so far for correlation of C/S and silanol content.

5) Unlike the T/CH model which is controlled by 4 parameters (x, n, z, m), the combinatorial model is majorly controlled by 1 parameter (C/S= y). In fact, by

assuming a=b=c=0, one reaches good agreement (in average) with experimental observations. We emphasize that the effect of (a,b,c) is marginal with respect to y, as these parameters are describing local textural variability at a given C/S due to polymorphism.

6) The T/CH model provides a "crystal-chemical" interpretation of the C-S-H structure. As extensively discussed by Richardson [286], the defective Tobermorite remains fairly crystalline and also the additional CH is assumed to be crystalline at high C/S ratios. Therefore, the molecular structure proposed by Richardson and his co-workers has a crystalline nature. This is against the observations by Pellenq and his co-workers [260, 8, 33], that C-S-H at high C/S presents pair-correlation or structure factor responses, which resemble that of defective amorphous and 3D glassy materials.

7) Finally, the very fundamental difference between the T/CH model and the combinatorial models is the predictive power of these two models. While the T/CH gives a description of molecular structure of C-S-H, it is basically incapable of providing qualitative or quantitative insights regarding the physical properties of C-S-H. Unlike T/CH, as shown in Figs. 2-3,2-4 and 2-5, combinatorial models are able to predict physical properties of C-S-H at the nanoscale such as elasticity, strength and fracture toughness indentation elastic modulus and hardness. This is a clear advantage for the combinatorial model, placing it as a foundation for computational material design.

The above seven points fully differentiate the combinatorial approach with respect to the T/CH model proposed by Richardson and co-workers [287, 288, 286, 283, 284].

2.6 Discussions

To our knowledge, the simultaneous screening of mechanical properties against two defect attributes has not been previously considered in the cement science and chemistry literature. By comparing Fig. 2-5.a-b and Fig. 2-5.c-d, we believe this new approach has broad implications regarding property optimization. Simply stated, in Fig. 2-4 one seeks an extremum in a single-variable function $M/H = f(C/S)$,

whereas in Fig. 5 one seeks a function of two variables, $M/H = f(C/S, \lambda)$. Any effects of correlation between C/S and λ would be missed in Fig. 4. Moreover, the results of Fig. 2-4 and Fig. 2-5 suggest that M/H is a property distributed in two defect attributes rather than being fully characterized by the single C/S chemical attribute. This is perhaps to be expected, given the complexity of the molecular structure C-S-H. It is therefore rather gratifying that other structural attributes, the packing of the silica tetrahedra and the medium-range environment of the interlayer Cw-O network, can also affect the mechanical response.

This work has benefitted from studies of network forming glasses [227, 351, 34] which have given further insights into the concept of atom-averaged covalent coordination number, $\langle r \rangle$, as a structural attribute capable of identifying optimum network connectivity[262, 263, 264]. From measurements of microhardness, indentation fracture toughness on binary Ge-Se glasses peak values were found when the properties were screened against $\langle r \rangle$[351], a situation analogous to our findings in Fig. 2-5. Recently it was pointed out that the defect attribute associated with the first sharp diffraction peak (see Introduction) plays a fundamental role in revealing extremum or anomalous behavior in rigidity deformation and transport properties of network glasses [227]. Here again we see a connection with the present findings: A relation can be established between system-level properties and defect attributes in both network glasses and C-S-H, which is potentially universal in a significant overall implication.

In this work, we introduce an approach of combinatorial screening of indentation stiffness and hardness for realistic models of cement hydrates against a set of structural defect attributes. We find peak values in the measured M/H in two defect-attribute sets, (C/S, λ) and (C/S, λ'). Based on considerations of various cross correlations among the defect attributes, we conclude M/H is a distributed property in that the relevant attributes of the underlying atomic configurations are coupled. We interpret the nature of the correlation to lie in the connectivity of medium-range environments, such as packing of silica network and openness of Cw-O regions of the cement hydrates. Moreover, we believe the defect-attribute coupling to be a manifestation of inorganic molecular networks that may also describe other multi-component systems

with structural complexities across the nano- and mesoscale. For future studies, we believe it would be appropriate to focus on the characterization of pores and confined water in cement hydrates. To achieve this, a new class of mesoscale models [215, 102] have to be developed and validated, and appropriate databases generated. These will involve additional defect attributes such as particle packing density, pore size distribution and connectivity, and the effects of water in various spatial confinements [242, 241]. The combinatorial approach introduced here is anticipated to aid those efforts, and to contribute to optimized concrete design that may reduce material consumption and associated CO_2 production. Indeed, while the CO_2 emission of cement clinker production scales with the volume of the concrete structural elements (beams, columns,...), the structural strength scales with their cross-sectional area. By adopting a limit state design approach, it is thus expected that an increase of the material strength by a factor of $\hat{\imath}\acute{t}$ allows reducing the environmental footprint to δ^{-1} for pure compressive members such as columns and shells, $\delta^{-2/3}$ for beams, and $\delta^{-1/2}$ for plates [344]. As contemporary concretes are characterized, on average, by high C/S ratios, it is expected that the found 60% increase in C-S-H strength achieved by reducing C/S from 1.7 to 1.1 (for instance with silica flour additions and proper curing conditions) can entail a 37% reduction of material volume and associated CO_2 emissions for compressive members assuming that microtexture does not alter significantly with C/S ratio; not counting further reductions achieved by diluting cement's calcium clinker [210] with silica or other cement substitutes. In conclusion, the sustainable development of green concrete to meet our societies needs for infrastructure, homes and shelter, becomes possible.

Chapter 3

Structure and Dynamics of Nanoconfined Water in C-S-H

The main focus of this chapter is on the effect of chemical modification of C-S-H on the structure and dynamic properties of water. With growing evidences on the nanogranular texture of C-S-H [17, 82, 68, 215], the C-S-H particles appear to be excellent analogs of nanocapsules with porous surfaces that enables inter-particle water and cationic exchange. However, the difference between the interlayer and inter-particle pore size distribution results in a broad distribution of water molecules' residence times. This leads to partial decomposition and averaged representation of dynamical quantities in Neutron Scattering (NS) [118] and Proton Field-Cycling Relaxometry (PFRC) [180] experiments. This highlights the relevance of atomistic simulation techniques that provide insight into experimentally inaccessible measurements. In fact, these experimentally aggregated descriptions are the origin of classifying water into free, constrained and chemically bound water using Quasi-Elastic Neutron Scattering method [333]. This has led to classifying the dynamical properties of confined water as "glass-like" behavior reminiscent of characteristics of supercooled liquids and glassy phases [368, 125]. Nevertheless, the effect of confinement stoichiometry by modification of surface chemistry and morphology on properties of confined water remains to be investigated. In fact, we aim at answering long-standing questions about the mobility of water in ultraconfined hydrophilic interlaminar spacing and its

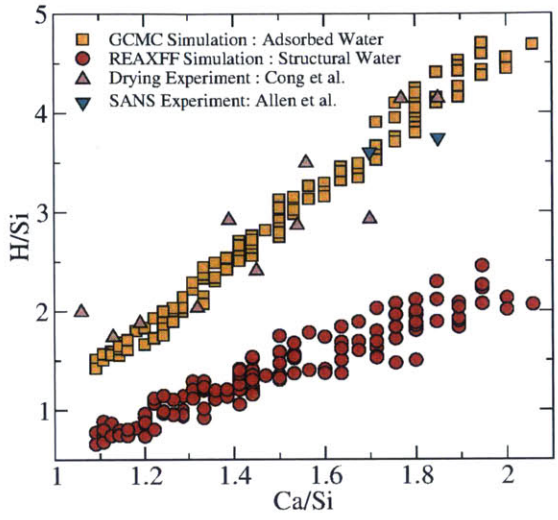

Figure 3-1: The state of hydrogen in C-S-H nanotexture as a function of stoichiometry. The total hydrogen content calculated via a combination of grand canonical Monte Carlo and molecular dynamics simulations based on REAXFF potential [210] compared to drying experiment of Cong et al. [78] and Small Angle Neutron Scattering (SANS) experiment of Allen et al. [17].

comparison with inter-particle and bulk water. The organization of the chapter is as follows. Section 3.1 covers the details of molecular dynamics simulation and molecular structure of C-S-H at different stoichiometry. Section 3.2 discusses the effect of stoichiometry on the structure of water such as bond length, dipolar moment, density and hydrogen bond network. The consecutive three sections (3.3, 3.4and 3.5) delve into anisotropy and inhomogeneity of water dynamics in a confined medium. Section 3.6 presents the effect of confinement chemistry on the self-diffusivity and identifying a new anomalous behavior of water under extreme confinement. Section 3.7 draws final conclusions and summarizes our findings.

3.1 Setting Up the Computational Scene

To describe the disordered molecular structures of C-S-H, we rely on the set of 150 molecular structures for C-S-H constructed by varying Ca/Si ratio ranging from 1.1 to 2.1 in Chapter 2. In the CSH models, the silicate chains are only consisted of

pairing (bridging oxygen atoms) and bridging sites (having two bridging and two dangling oxygen atoms) [9]. These negatively charged calcium-silicate sheets are separated from one another by interlayer spacing which incorporates "interlayer" water molecules and charge-balancing calcium cations (4.8 $\overline{e}/nm^2$). The adsorption of water molecules in the structurally defected Tobermorite models was performed via the Grand Canonical Monte-Carlo (GCMC) method ensuring equilibrium with bulk water at constant volume and room temperature and is denoted by "adsorbed water" in Fig.3-1. The simulated number of hydrogen atoms in water molecules and hydroxyl groups per silicon atom (H/Si) is in full agreement with drying [78] and Small Angle Neutron Scattering (SANS) [17] experiments.

As shown in Fig.3-1, these models provide molecular structures of C-S-H that are consistent with the experimental stoichiometry of $(CaO)_x(SiO_2)(H_2O)_{2(x-1)+0.5+\delta}$ where δ is associated with the polymorphic structure of C-S-H providing a variation of water content based on the combinatorial nature of silicate arrangements at the nano-scale. Some of the abovementioned interlayer water molecules are chemically unstable and dissociates into hydroxyl groups and protons upon first principle or reactive force field modeling. In this work, we use REAXFF potential [210] to react the interlayer water with the defective calcium-silicate sheets. The rest of the water molecules remain unreacted and are denoted in Fig 3-1 as "structural water". The reactive modeling further refines the stoichiometry by adding an extra dimension: $(CaO)_x(SiO_2)(H_2O)^{Unreacted}_{1.2(x-1)+0.3+\beta}(H_2O)^{Hydroxylating}_{0.8(x-1)+0.2+\gamma}$ where extra parameters β and γ introduce the effect of polymorphism to the stoichiometry. It should be emphasized that "Unreacted" and "Hydroxylating" superscripts indicate unreacted and dissociated water molecules during reactive simulations, respectively. It is noteworthy that both drying and SANS experiments measure the total hydrogen content and cannot distinguish between hydrogen atoms in molecular water or hydroxyl groups. The collective topological observation of these 150 numerical samples indicates that while C-S-H at low Ca/Si ratio, below 1.5, has a crystalline molecular structure, at high Ca/Si ratios above 1.5, it has a local glassy structure while still retaining some long-range layered texture [8].

In this work, equilibrium MD simulations are performed on the reacted C-S-H samples with focus on understanding the structure and mobility of unreacted interlayer water molecules. Several studies have suggested the importance of incorporating flexibility in modeling confined water in hydrophobic and hydrophilic confinements and ionic solutions [73, 350, 311]. To resolve the issues related to incorporating rigid water models such as SPC/E, a flexible but unpolarizable version of SPC model was employed which reproduces the structure and properties of water at ambient temperature [330, 302]. A core-only potential, CSH-FF [306] which is a clayff-like potential, is used which reproduces the structure and mechanics of C-S-H at various stoichiometry and is shown to be in good agreement with DFT calculations on crystalline C-S-H such Tobermorite polymorphs [306]. To exclude the expenses incurred by calculating coulombic interactions in real and reciprocal spaces, such interactions are calculated via the wolf method [357]. The further details on the preparation of C-S-H structures and the force fields are presented in the Supplementary Information.

In this work, all MD trajectories for 150 C-S-H models were produced using LAMMPS [267]. The equations of motion are integrated via Velocity-Verlet algorithm. The time steps are set as small as 1 fs to reproduce the dynamics of O-H bonds. All 150 samples were relaxed in isobaric-isothermal ensemble (NPT) with the target temperature of 300 K and pressure of 0 atm. To reproduce the dynamics of water at different time-scales, all the simulations were taken to microcanonical ensemble (NVE) to start two separate sets of simulations for the production phase. The first set of simulations, aiming at describing the dynamics at short time scales, were 100 ps long with configurations saved every 100 fs. The second set of simulations was 10 ns long with outputs recorded on intervals of 10 ps. It is worth mentioning that Churakov predicted that 10ns long simulations are necessary to capture the diffusion of water in the interlayer of Tobermorite minerals [72]. To compare the effect of confinement on water molecules with that of the bulk, a box of water with 1200 water molecules was simulated and referred here as to either "bulk" or "SPC" water. To calculate the structural and dynamical properties of interlayer water, several post-processing scripts were written to analyze the trajectories.

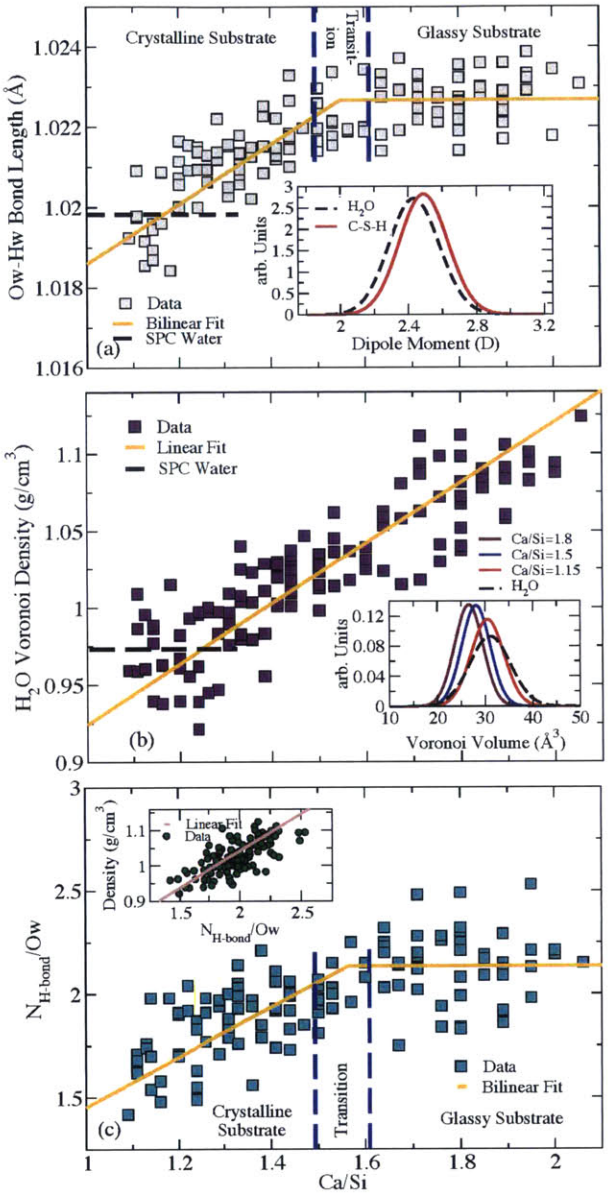

Figure 3-2: The effect of substrate stoichiometry on the characteristics of confined water. (a) Ow-Hw bond length as a function of Ca/Si ratio. The inset shows the distribution of dipole moments in C-S-H (Ca/Si=1.75) against bulk water. The stretch of Ow-Hw results in the increase in dipole moment of water in confined medium. (b) The effect of Ca/Si ratio on the Voronoi density of interlayer water compared to that of bulk SPC water. The inset displays the distribution of Voronoi volume of water molecules for C-S-Hs of varying composition against that of SPC bulk water. (c) The effect of Ca/Si ratio on the number of hydrogen bonds per water molecules. The inset presents the correlation between the number of hydrogen bonds and the Voronoi density of confined water.

3.2 Structure of Confined Water

By closely looking at water density profiles on the vicinity of calcium-silicate sheets, Youssef et al. [368] found excess water concentration at the surfaces. Also, it was realized that hydrogen atoms of water tend toward the surface. These are clear signs of hydrophilicity of calcium-silicate sheets. The hydrophilicity of C-S-H's solid backbone directly affects the structure of water.

Figure 3-2 presents the effect of the stoichiometry of C-S-H on the structure of water in the interlayer spacing. The structure of water is directly affected by the availability of adsorption sites on the calcium-silicate surfaces. Figure 3-2.a presents the effect of stoichiometry on the equilibrium O-H bond length. The O-H bond length in C-S-H microporous medium is larger than that of bulk water. This means that the interaction between C-S-H hydrophilic surfaces and protons in water molecules stretches the bond length. The O-H bond length in SPC water increases linearly from 1.018 Å to 1.022 Å in the crystalline substrate domain and reaches a plateau in the glassy domain. The physical reason behind the plateau response in high Ca/Si ratio is that the extra water is less affected by the surface through a shielding mechanism. In fact, adsorbed, adjacent to the surface, water molecules screen the electrostatic interactions between the additional water molecules and non-bridging oxygen atoms [46]. While the O-H bond length is directly influenced by the hydrophilicity of C-S-H substrate, the H-O-H angular distribution does not show a significant statistical difference. This directly affects the dipole moment of water molecules in the interlayer spacing. The inset in Figure 3-2.a provides the distribution of the dipole moment in a Ca/Si=1.75 C-S-H sample compared to that of bulk water. Statistically, the dipole moment of confined water, 2.56±0.3 D, follows normal distribution with first moment higher than that of bulk water, 2.50±0.3 D. In contrast to hydrophobic zeolite micropores [272, 308], the upshift of dipole moment is a signature of hydrophilicity of C-S-H surfaces. The hydrophilicity of nanopores, not only affects the internal structure of water molecules but also the way that water molecules are packed in the interlayer spacing. This packing affects the accessible volume and density of water in

ultra-confined environment.

The packing of water molecules in the interlayer spacing affects the density of C-S-H particles [334]. This necessitates a fresh look at the definition of density of water in confined geometries. This is achieved by employing topological geometry approaches in the context of nano-scale confinements. The Voronoi tessellation algorithm is one of the most used methods. For instance, it was applied to the structural analysis of liquid glass formers [319], liquid-gas interfaces [185] and ice nucleation [218]. In a periodic system, Voronoi or accessible volume is a portion of the space around a particle in which every point is closer to that particle than any other particle. The Voronoi density can be defined as the mass of particles divided by their Voronoi volume. In the case of bulk water, the conventional and Voronoi densities are identical. However, the ultraconfined medium in the interlayer spacing affects the Voronoi volume of water molecule and hence their Voronoi density. Figure 3-2.b provides the relation between the stoichiometry of C-S-H and Voronoi density of water in the interlayer spacing. The water density increases linearly with the Ca/Si ratio. The adsorption of roughly 1.2 structural water molecules by removing a SiO_2 group asserts that the density of interlayer water is more controlled by the availability of adsorption sites. At low Ca/Si ratios, the density of water is close to that of SPC bulk water. However at high Ca/Si ratios, the density of water is around 1.12 g/cm^3 significantly larger than that of bulk water. This is presented in more details in the inset of the figure 3-2.b which exhibits the Voronoi volume distribution for bulk water against three C-S-Hs at different chemical compositions. All distributions, either bulk or confined, follow normal distribution. At low Ca/Si ratios, Ca/Si=1.15, the first moment of volume distribution for confined and bulk water are identical. However, the second moment is smaller in the case of confined water meaning that the motion of water molecules is very limited in C-S-H interlayer. At high Ca/Si ratios, Ca/Si=1.5 and 1.8, both the first and second moment of volume distribution are smaller than that of bulk. The ultra-packing of water molecules in C-S-H is similar to that of hydration shell of proteins [137, 183].

Generally, higher packing and proper orientation of water molecules increases

the probability of forming hydrogen bonds network. There are numerous ways to identify hydrogen bonds based on either energetic [184, 23] or geometrical criteria [140]. Following the method of Luzar and Chandler [201, 202], the hydrogen bonds are solely distinguished by geometrical rules satisfying d_{O-O}<3.5 Å and θ_{H-O-O}<30° where d_{O-O} is the distance between the oxygen atoms in the donor and acceptor and θ_{H-O-O} is the angle between O-H ray in the donor and O-O ray connecting oxygen in the donor and the acceptor. Figure 3-2.c displays the effect of C-S-h's substrate stoichiometry on the number of H-bonds per water molecule. Similar to glycerol/water mixtures [87] and water at the surface of silicates [21], the number of H-bonds is significantly less than that of bulk water, roughly 3.6 H-bond per water molecules. The number of H-bonds increases linearly in the crystalline domain from 1.5 to 2.15 H-bonds per water molecule and maintains a constant value in the glassy domain. Using a potential-of-mean-force approach (PMF), Youssef et al. [368] calculated the value of 2.3 H-bonds per water molecule for Ca/Si equal 1.7 which is in close agreement with these results. A close investigation of O_{water}-O_{water} partial pair distribution function, $g(r_{O_{water}-O_{water}})$, reveals that the distance between a water molecule and its third and fourth water neighbors diminishes by increasing the Ca/Si ratio from 2.35 to 2.05 Å and 2.8 to 2.4 Å, respectively. This makes the number of H-bonds and density of interlayer water to be linearly correlated, see the inset of Figure 3-2.c. This correlation explains the link between different structural properties of water in confined hydrophilic media. Now that we have demonstrated the effect of chemistry on the structure of water, it is of high interest to investigate the mobility of water in the C-S-H interlayer space.

3.3 Inhomogeneity and Anisotropy in Water Dynamics

Due to strong hydrophilicity of calcium-silicate layers, the width and roughness of interlayer voids, the dynamics of water in the interlayer spacing of C-S-H is expected to

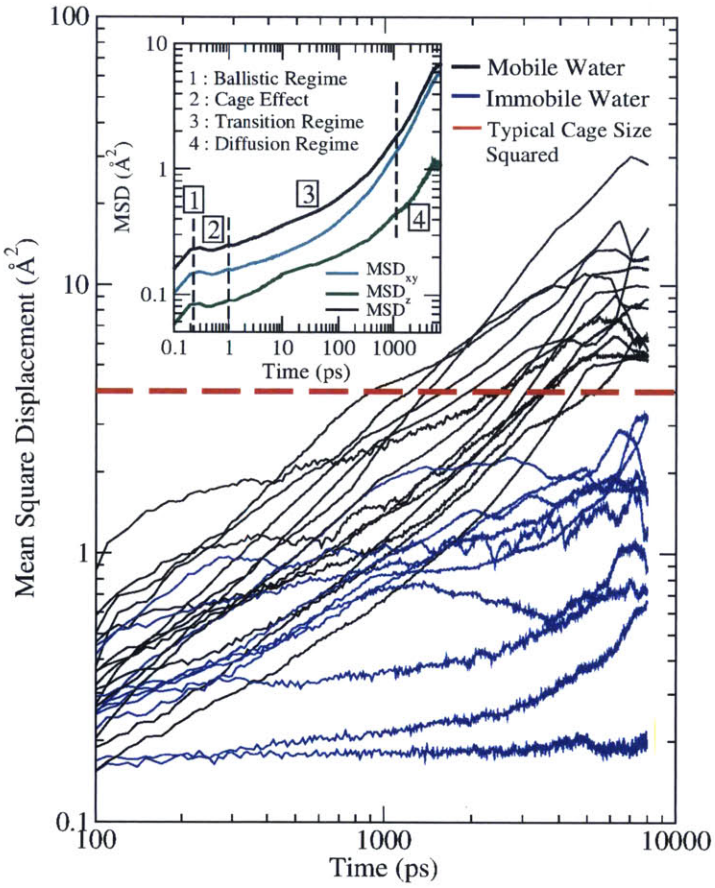

Figure 3-3: Mean square displacement (MSD) as function of time for each water molecule in a sample with Ca/Si=1.4. The inset presents the total average MSD and average MSD parallel and perpendicular to the calcium-silicate layers. The dashed red line at 4Å^2 is used to define the cage size, d_{cage} (see text).

be heterogeneous. This dynamical inhomogeneity can be characterized via both Mean Square Displacement (MSD) and the self part of Van Hove space-time correlation function. By utilizing the principle of invariance under time translation in equilibrium, the MSD of an atom is defined as:

$$\mathrm{MSD}_i = \langle \Delta r^2(t) \rangle = \langle |r_i(t+\tau) - r_i(\tau)|^2 \rangle \tag{3.1}$$

where r_i denotes the coordinate of oxygen atoms in water molecules and bracket represent proper time averaging at multiple time origins, τ. To illustrate the dynamical inhomogeneity of water, the MSDs of individual water molecules across all 150 samples were studied extensively. However, for the sake of brevity, Figure 3-3 presents individual MSDs of water molecules within the range of 100 to 10000ps in a sample with Ca/Si ratio of 1.4. The set of those individual MSDs clearly proves the validity of the aforementioned hypothesis on dynamical inhomogeneity. Hence, water molecules can divided into two categories: mobile and immobile. Immobile water molecules have limited motion around their adsorption sites, which is characteristic of a liberational dynamics in an ultraconfining cage around interlayer calcium or at the vicinity of defective silica chains. Mobile waters exhibit less bounded motion in the interlayer spacing, which resembles to that of diffusive dynamics in supercooled phases. The distinction between mobile and immobile water molecules is based on two arguments. First of all, the root mean square displacement (RMSD) of immobile and mobile waters are significantly different. While the RMSD of immobile water is smaller than 2 Å, the RMSD of mobile water is larger than 2 Å. The 2 Å length scale is referred to as the typical cage radius and will be denoted by d_{cage}. Secondly, the slope of MSD in mobile and immobile water is clearly different. While the slope of MSD versus time of mobile water is significantly larger than zero, the slope of MSD of immobile water is close to 0 (at time scales lower than 1 ns). This inhomogeneity in mobility is also observed in clay-zeolite composites [265].

The inset of Figure 3-3 displays the average MSD in the range of 0.1ps to 10ns. It demonstrates the dynamically slow nature of confined water in C-S-H which can be

partitioned in four stages; analogous to that of dense fluids and supercooled phases [178]. The four stages are ballistic, cage, transition and diffusion regimes. At very short times, t≤200fs, the water molecules move just ballistically, $\langle r^2(t)\rangle \propto t^2$ [290]. In the second stage within intermediate time scales, MSD follows a plateau up to 1ps. This is mainly because water molecules rattle around in the cage formed by neighboring particles colliding with water molecules and calcium layer species. Subsequently, parts of water molecules succeed to escape the cage, which can be associated with an intermediate stage. The motion of water molecules in the transition regime can be described as diffusion with low exponent ($\langle r^2(t)\rangle \propto t^\alpha$ where $\alpha < 1$). The transition with exponent smaller than one is also observed in supercooled liquids and attributed to inhomogeneous non-Gaussian collective hopping of particles [59]. The diffusion exponent gradually increases in the third stage till it reaches that of a diffusion regime, $\langle r^2(t)\rangle \propto t$.

Due to the small width to length ratio of interlayer channels ($\approx \frac{1}{10}$), the diffusion inside C-S-H particles has characteristics of quasi two-dimensional diffusion. This anisotropy in the diffusion can be quantified via the components of average MSD [21]:

$$MSD_{xy} = \frac{1}{N}\sum_{i=1}^{N}\langle \Delta x^2(t) + \Delta y^2(t)\rangle \tag{3.2}$$

$$MSD_z = \frac{1}{N}\sum_{i=1}^{N}\langle \Delta z^2(t)\rangle \tag{3.3}$$

where N is the number of water molecules, MSD_{xy} and MSD_z denote the parallel and perpendicular parts of the MSD. As it is shown in the inset of figure 3-3, the MSD_{xy} is almost one orders of magnitude larger than $\mathrm{MSD}_\perp$. This emphasizes that the diffusion is strongly controlled by the interlayer space width presenting nearly two-dimensional characteristics [46].

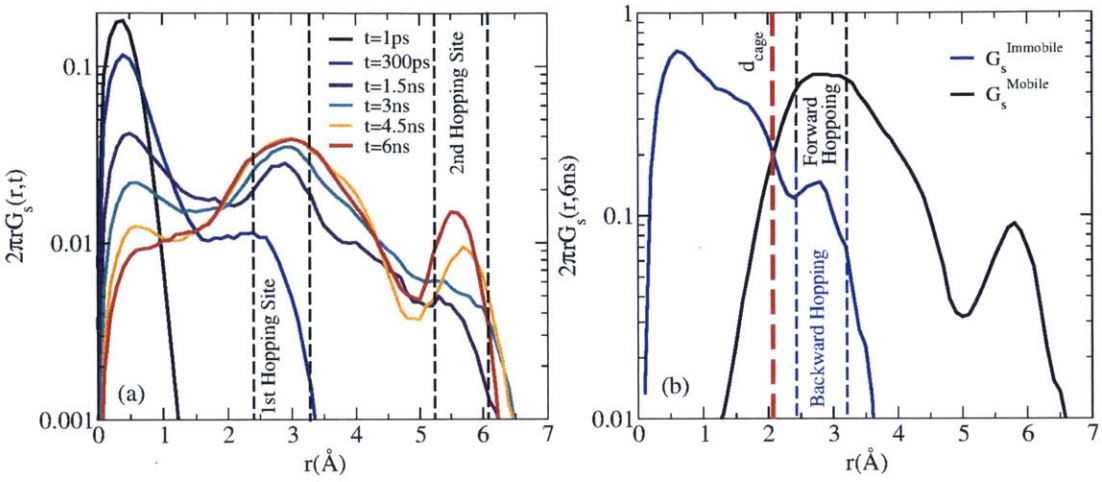

Figure 3-4: Description of the collective diffusion of water molecules using Van Hove space-time correlation function. (a) Van Hove correlation function of water molecules in a C-S-H sample with Ca/Si=1.4 at different time steps. (b) The Van Hove correlation function at 6 ns plotted for mobile and immobile water molecules.

3.4 Inhomogeneity Characterization via the Van Hove Correlation Function

To further characterize the inhomogeneous dynamics of water, the self part of Van Hove space-time correlation function is employed [64]. This correlation is defined as

$$G_s(r,t) = \frac{1}{N}\sum_{i=1}^{N}\langle\delta(r- \mid r_i(t+\tau) - r_i(\tau) \mid)\rangle \tag{3.4}$$

where δ denotes the Kronecker delta. This correlation function allows us to determine the probability $2\pi r G_s(r,t)$ that a water molecule originally at $(t = 0)$ and $(r = 0)$ has moved by distance r after elapsing time t [30]. Following the discussion on the anisotropy of diffusion in the C-S-H, the planar $2\pi r$ factor is adopted instead of $4\pi r^2$, as the motion of mobile water molecules is characterized by quasi two-dimensional diffusion. Presenting only the results for a sample with Ca/Si=1.4 for the sake of brevity, Figure 3-4.a provides the Van Hove correlation function at different time scales. As displayed by the probability distribution, at short time-scales, $(t \approx 0.1ps)$, the motion of water molecules is restricted to ballistic with characteristic length smaller than the

cage size. At a time-scale one order of magnitude longer, ($t \approx 1ps$), water molecules are still in the cage. In longer time-scales (($t \approx 300ps$)), some water molecules have traveled as far as 3Å. In the range of 300 ps to 1.5 ns, the mobile water molecules accumulate at the first hopping site characterized by the length of a water molecule, 3 Å. As the probability of presence in the cage decreases, a peak appears at the position of the first hopping site indicating the accumulation of mobile water molecules. The average jump length in C-S-H ($\approx$ 3 Å) is more than three times larger than that of bulk water at room temperature ($\approx$ 0.9 Å) [329] . Interestingly, the average jump length in supercooled water at 253 K is roughly 2.4 Å [66] which is close to that computed for C-S-H. This is another signature that water in the interlayer space behaves like supercooled liquids. The hopping mechanism to the farther sites occurs at longer time-scales.

The first water molecules arrive at the second hopping site after 1.5ns. At longer time-scales up to 6ns, while the probability of the presence of water molecules in the cage decreases diagnosed by vanishing of the first peak, second and third peaks grow signaling dominant diffusive mode. This probabilistic picture is reminiscent of picosecond local structural fluctuations within dynamical basin and slow inter-basin jumps in supercooled liquids [277]. This brings about the notion of "lower effective temperature" suggested by both neutron diffraction experiment [193] and molecular dynamics simulations of water in hydrophilic surfaces [125]. Specifically, Gallo et al. [125] observed that water molecules adjacent to hydrophilic surfaces behave as they are below mode coupling crossover temperature in pores as large as 4nm even though modeled at ambient conditions. However, this is a homogenous description of water diffusion, which only explains the diffusive motion of mobile water molecules.

The heterogeneity in the dynamics of water molecules in the interlayer spacing of C-S-H is very similar to heterogeneous dynamics of ionic [153] and supercooled liquids [103] in which particles diffuse with significantly different residence times at different regions only a few angstroms apart. To distinguish between the water molecules with different residence times, the self part of Van Hove function is decomposed to that of

mobile, $G_s^M(r,t)$, and immobile, $G_s^I(r,t)$, contributions:

$$G_s^M(r,t) = \frac{1}{N_M}\sum_{i=1}^{N_M}\langle\delta(r- \mid r_i(t+\tau) - r_i(\tau) \mid)\rangle \tag{3.5}$$

$$G_s^I(r,t) = \frac{1}{N_I}\sum_{i=1}^{N_I}\langle\delta(r- \mid r_i(t+\tau) - r_i(\tau) \mid)\rangle \tag{3.6}$$

where N_M and N_I are the number of mobile and immobile water molecules, respectively. The distinction between mobile and immobile species is made based on their individual MSDs. If the average displacement is less than the cage size, $\mathrm{RMSD}_i < \mathrm{d}_{cage}$, then the water molecule is assumed to be immobile. Otherwise, it is classified as mobile species. Figure 3-4.b presents the mobile-immobile decomposition of the Van Hove function. For time-scales as large as 6 ns, the immobile water liberating within the cage might overpass the cage size and briefly spend time in the first hopping zone. This is suggested by the presence of a minor peak in $\mathrm{G}_s^{Immobile}$ in the location of first hopping site in Figure 3-4.b. It should be emphasized that since the MSD of immobile water molecules are less than 3 Å, any transgression beyond cage size is temporal. This is a clear sign of backward hopping mechanism in which water oscillates between the cage and the first hopping site. On the other hand, some of mobile waters might briefly spend sometime at the boundary of the cage. However, mobile water molecules elapse most of their time in the first and second hopping sites. Again, some of the water molecules might oscillate back and forth characterized by a forward hopping mechanism from first to second hoping sites and backward vice versa. This forward and backward hopping is also experimentally observed in the diffusion of hydrogen on TiO_2 surfaces in which forward hopping is favored depending on separation distance [195]. For the case of water in the interlayer of C-S-H, the enhancement of interstitial molecules in the first coordination shell constraints the large displacements. Therefore, diffusion of water molecules necessitates highly correlated displacement of many water molecules referred to as cooperative hopping mechanism [370]. This cooperative hopping mechanism increases the probability of backward hopping as collective diffusion in confined environment attributes to higher

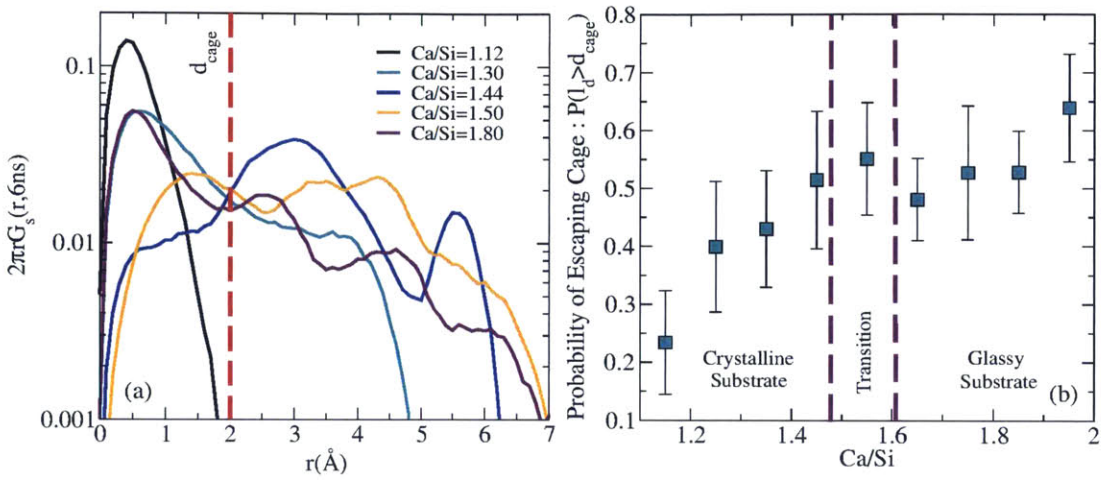

Figure 3-5: The effect of composition on the mobility of water molecules in the C-S-H interlaminar spacing. (a) Van Hove correlation function plotted for C-S-Hs of varying stoichiometry. (b) The probability of water molecules to escape the cage as a function of Ca/Si ratio.

diffusion energy barriers. This collective mobility depends strongly on the chemistry of adsorption surface, which is analyzed in more details in the next section.

3.5 Composition-Dependent Water Mobility

The heterogeneity in water dynamics should substantially depend on the hydrophilicity of nanoconfinements [46, 21]. This can be systematically analyzed by studying mobility across samples of varying stoichiometry. Figure 3-5 provides the Van Hove function at 6 ns for five C-S-H samples with Ca/Si ratios varying from 1.1 to 1.8. At very low Ca/Si ratios distinguished by clear crystalline substrate, Ca/Si=1.12, all of the water molecules are confined within the cage. In this case, the whole Van Hove function is limited within the cage radius, d_{cage}. Figure 3-6 has visualized the motion of water molecules by superposing the positions of water molecules within the structure at different time steps starting at 1 fs up to 6 ns, covering six orders of magnitude in the time domain. As depicted in the first row of Figure 3-6, at low Ca/Si ratios, the motion of water molecules is restricted to liberation adjacent to their adsorption sites. At this stoichiometry, the motion of the water molecules is

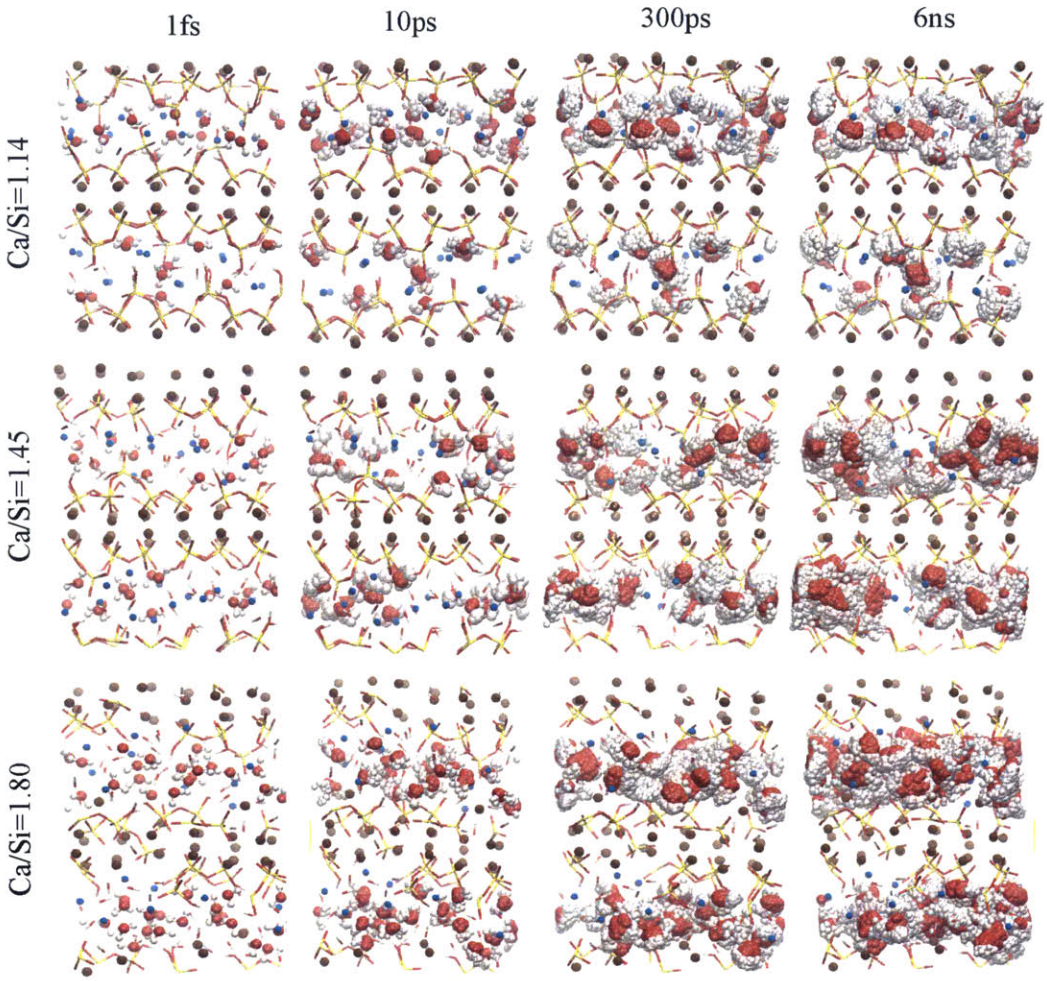

Figure 3-6: Visualization of the space that water molecules explore within the C-S-H interlayer for three different samples with Ca/Si ratios equal to 1.14 (crystalline substrate), 1.45 (transition substrate) and 1.80 (glassy substrate). The water molecules are superposed at 10ps, 300ps and 6ns at the intervals of every 1ps. The brown and cyan spheres present intralayer and interlayer calcium ions, respectively. Red and yellow sticks depict silicate tetrahedra. White-red sticks represent Hydroxyl groups and white and red sticks with white and red spheres display water molecules.

described as "pocket-like" behavior emphasizing their localized motion. This localized behavior is related to abundance of bridging SiO_2 groups protruding into the interlayer spacing which effectively obstructs the diffusive motion of water molecules in the channel. At a slightly higher Ca/Si ratio, Ca/Si=1.30, more birding groups are removed from the silicate chains. This provides the opportunity for some water molecules to escape their cage and enter their first hopping site. At the transition regime, most of the bridging sites are removed from the silica chains, which fully open the diffusion channel. As described before, a significant portion of water molecules diffuse to the first and second hopping sites. As highlighted in the second row of Figure 3-6, the water molecules oscillate within the cage up to 10 ps and start to diffuse at longer residence times in the order of 100 ps. The motion of water molecules in transition regime is individually traceable which resembles to a "patch-like" diffusive motion. As presented in the third row of Figure 3-6, the diffusive channels are fully developed making it hard to distinguish between individual trajectories which indicates a "channel-like" diffusion. To rigorously quantify the impact of substrate stoichiometry on the mobility of water molecules, the probability of a water molecule escaping the cage after 6 ns, $P(d>d_{cage})$, is defined as:

$$P(d > d_{cage}) = \frac{1}{A_n} \int_{d_{cage}}^{\infty} 2\pi r G_s(r, t = 6ns) dr \tag{3.7}$$

where d is displacement and A_n denotes the normalizing factor defined as $\int_0^{\infty} 2\pi r G_s(r, t = 6ns) dr$. This probability of escaping cage increases linearly with Ca/Si ratio within the crystalline C-S-H regime. The probability of escaping the cage reaches to a plateau in transition and glassy regimes. This plateau is mainly due to the absence of bridging sites in the interlayer. However, at large Ca/Si ratio, the mobility increases owing to the fact that extra water molecules provide screening against the columbic interaction between the calcium-silicate substrate and mobile water molecules. Owing to the aforementioned screening effect which was pointed out as the origin of the plateau in the dipole moment of water molecules as well (see Figure 3-2.a), the dipole moment and mobility are strongly correlated as they both follow a bilinear trend.

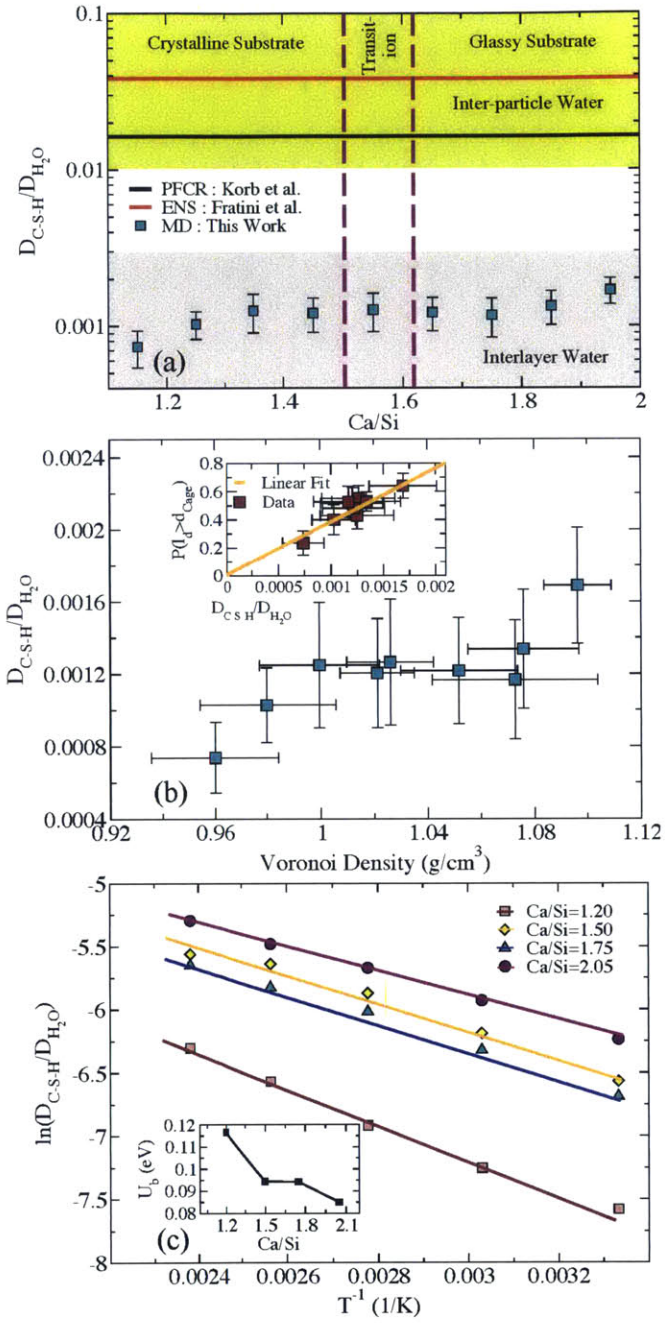

Figure 3-7: Mobility of water in C-S-H's interlayer spacing. (a) The effect of C-S-H composition on the self-diffusivity of interlayer water compared against Elastic Neutron Scattering (ENS) and Proton Field Cycling Relaxometry technique (PFCR). (b) The anomalous correlation between density and self-diffusivity of confined water in C-S-H and the insets provides the relation between the self-diffusivity and probability of escaping the cage. (c) The scaling of self-diffusivity with inverse temperature for four C-S-H models with different Ca/Si ratio. The inset presents the diffusion energy barrier as a function of Ca/Si ratio.

3.6 Self-Diffusion of Ultraconfined Water in C-S-H interlayer

As discussed above, the diffusion of water molecules inside the C-S-H particles is both inhomogeneous and anisotropic. Therefore due to quasi two-dimensional nature of collective diffusive motion of mobile water molecules, the self-diffusivity of water is defined only for planar component of the mobile water species:

$$D_{\parallel} = \frac{1}{4} \lim_{t \to \infty} \frac{msd_{\parallel}(t)}{t} \tag{3.8}$$

Description of water diffusion by $D_{\parallel}$ entails an approximation. However, the ratio of $D_{\perp}/D_{\parallel}$ is quite negligible for mobile water molecules because the width to length ratio of channels is negligible due application of periodic boundary conditions. The effect of substrate composition on the self-diffusivity of water molecules is presented in Figure 3-7. The self-diffusivity of water in the interlayer, D_{C-S-H}, is normalized by that of bulk SPC water, D_{H_2O}, 4.17×10^{-9} m^2s^{-2}. Analogous to the probability of escaping cage, D_{C-S-H} increases linearly in the crystalline domain and maintains a constant value at the transition and glassy regimes. Due to high H_2O/Si ratio at very high Ca/Si ratios, the self-diffusivity increases again because of screening effect. D_{C-S-H} in high Ca/Si, 0.0018× D_{H_2O}, ratio is almost three times higher than that of low Ca/Si, 0.0006× D_{H_2O}. This highlights the effect of substrate stoichiometry on the diffusion of water molecules in the ultraconfined hydrophilic environments. On average, the self-diffusivity of water in the interlayer spacing inside C-S-H particles is a thousand times slower than that of bulk water. This emphasizes that part of the mobile interlayer water diffuses with significantly slow dynamics but distinguishable from immobile water and hydroxyl groups. Bordallo et al. [48] briefly mentioned existence of a dynamics slower than that of glassy water and attributed it to proton exchange in the interlayer. Perhaps, based on the present description of water mobility in C-S-H, this dynamics might be related to diffusion of mobile water in the interlayer. The diffusion of mobile interlayer water is different from inter-particle water molecules

and those on the surface of larger pores.

The values of self-diffusivities presented in this study are at least one order of magnitude less than that measured by Proton Field-Cycling Relaxometry (PFCR) approach by Korb et al. [181] and Incoherent Elastic Neutron Scattering measurements of Fratini et al. [118]. Recent ENS experiments suggest that slow self-diffusion in the paste is associated with inter-particle water [194]. PFCR experiments are interpreted to capture the dynamical behavior of water adsorbed on the pore surfaces [180]. In particular, Kalinichev et al. [168] reported MD simulation of adsorb water on [100] surfaces of anomalous Tobermorite with self-diffusions comparable to that of Korb et al. [180]. Since the inter-particle pore-size distribution is shown to be significantly larger than interlayer spacing [219], the significant difference in characteristic length of confinement explains the difference between the self-diffusion inside and in between C-S-H grains. Recently, Coasne et al. have critically reviewed the state of the art of water in mesoporous systems [76] which sheds light on the dynamics in mesopores in between C-S-H particles. The contrast between the intra- and inter-particle water is also observed in hydrated bundles of imogolite nanotubes [85]. Having explained the difference between intra- and inter-globular diffusion, it is of scientific interest to explore the correlation between the structural and dynamical properties of confined water.

Studying properties of bulk water has led to identification of several unusual behaviors known as anomalous features of water [200]. Prielmeier [270] discovered an anomaly in which self-diffusivity of bulk water increases with increasing pressure with a maximum at 200MPa. Later Scala et al. [299] showed that for supercooled water at 220K, the self-diffusivity of bulk water increases with increasing density with a maximum at 1.15 g/cm^3. This anomalous behavior was also shown to exist for SiO_2 glassy structure as well [110]. Figure 3-7.b presents the correlation between the density and self-diffusivity of water in the interlayer of C-S-H. Having many features in common with supercooled liquids, the self-diffusivity of water in the interlayer of C-S-H increases with the density. However, these anomalies have different sources in supercooled liquids and ultraconfined water in between hydrophilic surfaces. In

supercooled water, the collapse of hydrogen bond network under external pressure is believed to increase the diffusivity of water [92]. However, the physical reason behind the peculiar behavior in C-S-H is related to the increase in probability of escaping cage. As described in the inset of Figure 3-7.b, the correlation between the probability of escaping cage, $P(d > d_{cage})$, and relative self-diffusion, D_{C-S-H}/D_{H_2O} is fairly linear passing through the origin ($P(d>d_{cage}) = 400 \times D_{C-S-H}/D_{H_2O}$). This emphasizes that as the probability of mobility approaches to zero, the self-diffusivity tends to zeros as well. On the contrary, as the effect of surface stoichiometry becomes negligible through screening effect, the mobility increases leading to an increase in self-diffusivity. In fact, both of these mechanisms directly affect the diffusion energy barrier. This energy barrier is equal to the kinetic energy required for a water molecule to break its surrounding cage and diffuse in the ultra-confined medium. This diffusion energy barrier can be estimated via Arrhenius relation

$$D_{C-S-H} = D_{H2O} \times \exp\left(\frac{-U_b}{K_B T}\right) \tag{3.9}$$

where U_b, K_B and T are diffusion energy barrier, Boltzman constant and temperature, respectively. In molecular dynamics, U_b can be estimated by gradually increasing the temperature of the system and monitoring the logarithm of self-diffusivity as function of the inverse temperature. To measure U_b in C-S-H models of varying stoichiometry, five simulations were performed in the range of 300 K to 420 K for four C-S-H models with Ca/Si ratios equal to 1.2, 1.5, 1.75 and 2.05 (Fig3-7.c). It is highlighted that $ln\,(D_{C-S-H})$ scales linearly with the inverse temperature which means that Arrhenius relation is still valid in describing diffusion in nano-confined environment. The slope of these lines are proportional to U_b which are plotted in the inset of Fig3-7.c. The energy barrier is in the order of 0.12 eV at low Ca/Si ratios (Ca/Si=1.2) and decreases to 0.08 at high Ca/Si ratio (Ca/Si=2.05). This further clarifies the abovementioned anomalous behavior in self-diffusivity of nano-confined water in which despite the increase in density, the diffusion energy barrier decreases. The reduction in the energy barrier fully explains the increase in self-diffusivity as it describes the increase in the

probability of escaping the cage with increasing Ca/Si ratio.

3.7 Conclusion

This chapter was dedicated to study the properties of confined water in the interlayer spacing of a class of defective calcium-silicates. The properties of water were carefully studied using atomistic simulation method in a set of 150 samples with the stoichiometry of $(CaO)_x(SiO_2)(H_2O)^{structural}_{1.2(x-1)+0.3+\beta}(H_2O)^{Hydroxylating}_{0.8(x-1)+0.2+\gamma}$. The C-S-H substrate directly affects the structural properties of ultraconfined water such as bond length, dipolar moment, density and average number of hydrogen bonds. It is found that the dipole moment, Voronoi density and the number of hydrogen bonds per water molecules are higher at high Ca/Si ratio. These enhancements are directly attributed to the increase in the hydrophilicity of the substrate with increasing Ca/Si ratio. However, the screening effect between water molecules and calcium-silicate backbone was found to alleviate these enhancements. The dynamics of water is also strongly affected by both the confinement size and hydrophilicity of lamellar sheets and interlayer cations. The water dynamics in the interlayer spacing was characterized as quasi two-dimensional in which the diffusion in the direction of layers is the dominant mode. The mean square displacement shows a four stage dynamics reminiscent of those in supercooled liquids and glassy phases. In addition, both MSD and Van Hove space-time correlation function indicated significant heterogeneity in the water dynamics. The mean hopping distance was found to be around 2.8Å, roughly a size of a water molecule, comparable to that observed in supercooled water. The decomposition of van Hove function into mobile and immobile contributions shed light into the dynamics of immobile waters and revealed the possibility of backward hopping from the first hopping site to the dynamical cage.

More significantly, it was shown that the mobility of water is strongly composition-dependent and increases at high Ca/Si ratio. The increase in the probability of escaping the dynamical cage with increasing Ca/Si ratio physically explained the composition-dependent mobility. Furthermore, this was related to the absence of

dangling silica groups into the diffusion channel. This brought forward the notion of "pocket-like", "patch-like" and "channel-like" diffusion at low, medium and high Ca/Si ratios, respectively. The self-diffusivity of mobile interlayer water is found to be one thousandth of that of bulk water. Both the strong interaction with the C-S-H hydrophilic surfaces and the sub-nanometric diffusion channel dimensions effectively reduced the self-diffusivity of the structural water. The self-diffusivity of water is three times higher in high Ca/Si ratio with regards to low Ca/Si ratio. Furthermore, it was found that the self-diffusivity of water in the interlayer is roughly ten times smaller than those observed in elastic neutron scattering and proton field-cycling relaxometry. This is mainly because these experiment probe the dynamics of inter-particle water. In fact, the size of inter-particle pores is one order of magnitude larger than interlayer spacing reducing the electrostatic interaction with substrate through screening effect. Finally, the correlation between the structural and dynamical properties was brought to attention through correlation between density and self-diffusivity. Similar to the anomalous behavior of bulk water, the self-diffusivity of water increases with increasing the density. Unlike the case of bulk water in which breakage of hydrogen bond network increases the mobility, the confined water is shown to have higher self-diffusivity because of increase in the probability to escape the dynamical cage due opening diffusion channels. Overall, this chapter paves the way to further understand the nuances in the properties of confined water in a broad range of C-S-Hs at nanoscale aiming to shed light on the design of more durable materials.

Chapter 4

Thermal Properties of Cement Paste: From Nano to Macro Level

Being well-known for its appreciable load-bearing capacity, concrete is a material of choice in design and construction of structural units. Despite their importance, the thermal properties of cement paste and concrete are always overshadowed by their mechanical properties, namely stiffness and strength. The thermophysical properties of concrete are important during both construction and use phases. In the course of construction, early-age cracking can occur which would compromise the longer term servicability of the structures. Among various sources of the stress that can initiate such cracking, thermal stress is one of the most influential factors. These stresses are the direct consequence of cement's exothermic hydration process. Basic dimensional arguments assert that the thermal stress is inversely proportional to the thermal diffusivity of hydrating paste. During the use phase, the thermophysical characteristics of concrete (and all the other construction materials) affect the heating and cooling energy consumption of the building via either heat conduction through the envelope (transient, quasi-steady state and steady state) or thermal mass of buildings, i.e. the tendency of a building to maintain a constant temperature despite outdoor temperature oscillations [6]. Notwithstanding the importance of thermal properties of concrete, such studies are truly scarce and are limited to the macroscopic measurement of thermal properties of concrete [90, 173, 94, 95] and

cement paste [39, 367, 230, 214, 363, 5]. In fact, the interplay between chemistry and molecular properties of cement paste's constituents and their relation to the macroscopic properties remain obscure. This requires an in-depth understanding of atomic and molecular processes contributing to thermal properties at the nanoscale along with proper consideration of morphology and multi-phase nature of cement paste as a multiscale porous material.

The main focus of this chapter is to understand the thermo-physical properties of cement paste starting from nanoscale and upscaling them all the way to the engineering macroscale. In fact, we attempt to answer long-standing questions about the physical origins of the macroscopic thermal behavior of a complex material such as cement paste. This chapter is organized into seven sections. Section 4.1 describes the complex multiscale multiphase structure of cement paste. Section 4.2 covers the details of molecular structures of calcium silicate phases in cement paste including cement clinkers and hydration products. Section 4.3 includes the very details of atomistic simulations and discusses the transferability of the inter-atomic potential utilized in this study in comparison with available experimental data and other force fields. Section 4.4 describes vibrational density of state of different phases in cement paste and discusses the existence of boson peak in C-S-H. Section 4.5 presents specific heat capacities of all phases at constant volume and pressure. Section 4.6 explores the nanoscale non-equilibrium thermal properties such as thermal conductivity and the mean free path of phonons for different phases using Green-Kubo relations. Subsequently, section 4.7 is dedicated to the upscaling of the nanoscale thermal properties to the macroscale (Level 3) and their comparison with macroscopic observations. Section 4.8 draws final conclusions and summarizes the findings of this chapter.

4.1 Multiscale Multiphase Porous Structure of Cement Paste

Cement paste is a complex material which is the result of the hydration of cement clinkers. The ordinary Portland cement (OPC) clinker is comprised of different phases such as alite (C_3S), belite (C_2S), calcium aluminate (C_3A), tetracalcium aluminoferrite (C_4AF), gypsum and minor other phases, where in cement chemistry notation [328] C, S, A and F stand for CaO, SiO_2, Al_2O_3 and Fe_2O_3, respectively. C_3S is the major component of OPC, which accounts for 70% of its weight and C_2S is the second most important phase in OPC. Here, we focus on the calcium silicate phases since they are more abundant in OPC. The hydration reactions of C_2S and C_3S can be summarized as:

$$C_2S + xH \rightarrow C_{1.7}SH^s_{1.9} + 0.3CH + (x - 2.2)H^{i,c} \tag{4.1}$$

$$C_3S + xH \rightarrow C_{1.7}SH^s_{1.9} + 1.3CH + (x - 3.2)H^{i,c} \tag{4.2}$$

where x is the number of water moles used in the hydration and $C_{1.7}SH^s_{1.9}$, CH and H are calcium silicate hydrate (CSH), portlandite and water, respectively. The s, i and c superscript are utilized to emphasizes the pore size distribution in the cement paste accommodating (s)tructural, (i)nter-particle and (c)apillary water molecules, respectively. The pore size distribution in the cement paste ranges from the nanoscale to the macroscale. The presence of multiple porosity scales makes a multiscale bottom-up approach the method of choice to study this material.

The hierarchical multiscale models have shown to be effective in portraying a consistent picture of the physical properties of materials with multiple heterogeneity scales ranging from nano to macroscale [142, 273]. These models are based on the notion of the separability of scales, meaning that the heterogeneity at each scale should be significantly smaller than the size of the representative elementary volume (REV) at the larger scale. This approach can be used to study the properties of complex

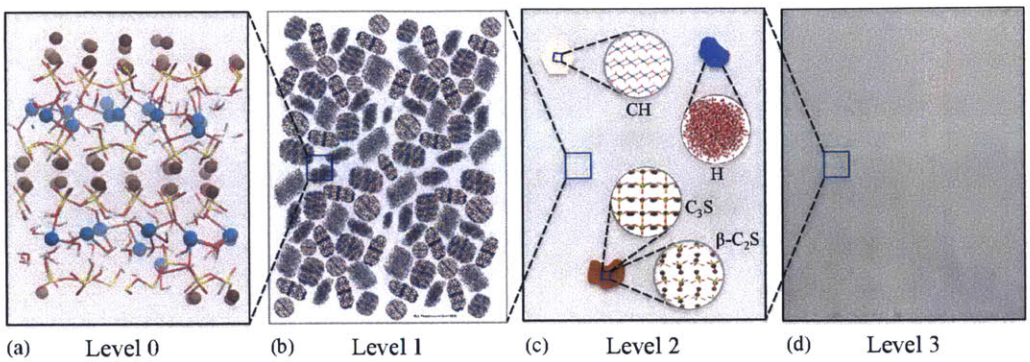

Figure 4-1: A 4-level homogenization thought-model utilized for upscaling the thermal properties of concrete from nano to the engineering scale. a) snapshot of a nano-texture of CSH at Ca/Si=1.5. The brown and cyan spheres represent intra- and inter-layer calcium ions. The Si-O bonds are shown by yellow-red sticks. The water molecules and hydroxyl groups are depicted by red-white sticks. The XY plane is parallel and Z axis is perpendicular to the calcium silicate sheets. b) The meso-texture of CSH constructed by agglomeration of randomly oriented CSH nanoparticles with size poly-dispersity. c) The micro-texture of hardened paste including anhydrated clinker phases (β-C_2S and C_3S), hydration products (CSH and CH) and pore space (saturated or dry). d) The macro-texture of cement paste at the engineering scale.

materials such as cement paste. Here, we employ a 4-Level hierarchical model to investigate the thermal properties of cement paste, see Fig. 4-1. At Level 0 or the nanoscale (Fig. 4-1.a), we calculate the molecular properties of crystalline phases (C_2S, C_3S and CH) and CSH. Understanding molecular properties of CSH is rather important because this phase is the binding component of the cement paste and is responsible for its strength and durability. Unlike crystalline phases, the molecular structure of CSH is rather complex[260, 310] ranging from crystalline to glassy depending on its stoichimetry [7]. The amorphous structure of the solid backbone of CSH (at high Ca/Si ratios) along with the presence of mobile water molecules [274] inside its nanostructure further prohibits the application of Cauchy-Born rule [9, 172, 13, 171]. Therefore, atomistic simulation is critical to reveal the underlying atomic processes governing the thermo-physical behavior of cement paste at the nanoscale.

Both neutron scattering [17, 67] and statistical nanoindentation experiments [82]

provide growing evidences that CSH has a nanogranular meso-texture (see Fig. 4-1.b). In this light, the CSH gel at Level 1 is considered to be formed of poly-disperse nano-scale CSH globules [215]. The space between the CSH particles, known as inter-particle spacing can be empty, filled with the previously mentioned inter-particle water or partially saturated depending on the relative humidity level. The physical properties of CSH gel can be explained via statistical micro-poromechanics theory [99]. The tools of micro-thermo-poromechanics yield the equivalent thermal properties of the CSH matrix at Level 2 (microscale). In fact at the microscale, the CSH gel is regarded as a continuous paste that holds unhydrated cement clinkers and the other hydration products together (see Fig. 4-1.c). As mentioned before, the capillary pores prevail at this scale which can have different saturation degrees depending on the humidity level. Therefore, cement paste at the microscale can be viewed as a multifarious composite material. Similar to Level 1, micro-thermo-poromechanics theory and simple mixture laws can be employed to upscale the thermal properties to Level 3 (macroscale). At the macroscale, we regard the cement paste as an isotropic homogeneous material.

4.2 Calcium Silicate Phases in Cement Paste

The molecular structure of solid phases in the hardened cement paste vary from crystalline to glassy depending on the phase and chemical composition (in the case of CSH). The anhydrated clinker phases (alite and belite) and portlandite are crystalline. From a crystallography viewpoint, alite is a chemically modified form of pure tricalcium silicate (Ca_3SiO_5 or C_3S) exhibiting series of reversible phase transitions upon heating and cooling [71, 203, 328]. The atomic structure of Alite polymorphs are similar, differing in the orientation of isolated silica tetrahedra (SiO_4^{4-}), coordination of calcium (Ca^{2+}) and oxygen (O^{2-}) atoms in ionic sites, and hence symmetry groups [247, 43]. In the present study, we choose the M3 polymorph of C_3S, the most abundant polymorph in cement clinker, refined from single crystal by De la Torre et al. [89]. Belite is similar to alite in many aspects. It is a chemically modified form of

dicalcium silicate (Ca_2SiO_5 or C_2S) which also exhibits sequences of reversible phase transitions with temperature [26, 328]. The atomic structure of belite consists of isolated SiO_4^{4-} groups surrounded by Ca^{2+} cations. Belite has several polymorphs such as $\alpha_H^{'}$, $\alpha_L^{'}$ and β which are derived from the α form by decreasing symmetry because of disorder in silica groups and subsequent modification of calcium atoms positions [114, 312, 165]. In the present study, we have employed the β-C_2S polymorph after Midgley [228] as it is the most dominant polymorph in the ordinary OPC systems [154, 106]. The M3 C_3S and β-C_2S phases are rarely found in pure forms and usually contain minor Al^{+3}, Mg^{+2} and Fe^{+3} substitutions that modulate their stability, solubility and chemical reactivity [210]. These substitutions are neglected in the present study as they are attributed to a relatively small mass fraction of the OPC clinker. Beside clinker phases, portlandite ($Ca(OH)_2$ or CH), one of the major hydration products in OPC systems, is also crystalline and is a fairly well-known mineral. It consists of a stack of Ca-O planes with dangling hydroxyl groups pointing to the interlayer spacing. In this work, we adopt a crystalline structure of portlandite after Henderson et al. [148] which has identified the position of protons via the ^1H NMR.

Unlike the above crystalline phases, the nano-texture of CSH strongly depends on its chemical composition measured in terms of calcium-to-silicon ratio (Ca/Si). To realistically describe the atomic structure of CSH, we utilize the set of 150 molecular models in Chapter 2 constructed to cover a wide range of calcium-to-silicon ratios spanning from 1.1 to 2.1 [274]. These models are consistent with the averaged stoichiometry of $Ca_xSiO_{2.75}(OH)^{Ca}_{0.85(x-1)}(OH)^{Si}_{1.15(x-1)+0.5}$[0.8(x-1)+0.3]$H_2O$, where x is the Ca/Si ratio and $(OH)^{Ca}$ and $(OH)^{Si}$ represent hydroxyl groups coordinated to inter-layer calcium atoms and silica groups, respectively. The collective topological observation of these 150 models indicates that, while at low Ca/Si ratios CSH exhibits lamellar crystalline structures with strong transversely isotropic mechanical response [7], the structure of CSH at high Ca/Si ratios (pertinent to the hydration of OPC systems in eqs. 4.1 and 4.2) is fairly amorphous [33] expressing isotropic characteristics with no significant preferential direction in elastic properties.

Table 4.1: Investigation of the CSH-FF transferability to other calcium silicate systems by comparing β-C_2S, C_3S and CH's lattice parameters against experimental measurements and calculated values via the core-shell potential.

	Method	a (Å)	b (Å)	c (Å)	α (°)	β (°)	γ (°)
β-C_2S	CSH-FF	5.44	6.62	9.13	90.5	95.0	90.0
	core-shell [209]	5.51	6.92	9.94	90.0	94.0	90.0
	Exp. [228]	5.50	6.76	9.32	90.0	94.1	90.0
C_3S	CSH-FF	34.70	6.69	18.43	90.0	95.0	90.0
	core-shell [209]	33.45	7.12	18.91	90.0	94.9	90.0
	Exp. [89]	33.08	7.02	18.50	90.0	94.1	90.0
CH	CSH-FF	3.48	3.48	4.82	90.0	90.0	120.0
	core-shell [209]	3.55	3.55	4.94	90.0	90.0	120.0
	Exp. [148]	3.59	3.59	4.91	90.0	90.0	120.0

4.3 Transferability of CSH-FF Potential to Other Phases

In this chapter, atomistic simulation techniques including molecular dynamics (MD), energy minimization and phonon analysis are utilized to study thermal properties of clinker phases and hydration products using LAMMPS [267], Gulp [123, 124] and a series of in-house codes. In MD, the equations of motion were integrated via the Velocity-Verlet algorithm with periodic boundary conditions applied in all directions. After performing extensive convergence studies, the suitable time steps were found to be as small as 0.1 for CSH and portlandite and 1 fs for clinker phases to appropriately capture the behavior of autocorrelation functions at small scales. Initially, all the 150 CSH samples and the crystalline phases were relaxed in isobaric isothermal ensemble (NPT) at room temperature and pressure of 0 atm. To exclude the impact of fictitious forces on the trajectories of atoms affecting velocities and heat fluxes along the MD trajectory, all 150 samples are transferred to microcanonical ensemble (NVE). In order to properly calculate the statistical errors for each model, 10 independent simulations are performed, each 10^6 time steps long. Core-only potential CSH-FF [306], which is a Clayff-like potential [86], is used in the calculations. This potential is shown to

Table 4.2: Comparing elastic properties of β-C_2S, C_3S, CH, $C_{1.0}SH_{0.5}$ and $C_{1.75}SH_{2.0}$ calculated using CSH-FF potential against experimental measurements and calculated values via the core-shell potential.

	Method	k (GPa)	g (GPa)	ν
β-C_2S	CSH-FF	122.3	47.2	0.33
	core-shell [209]	111.0	53.1	0.30
	Exp. [352, 12]	116±8,110±8	54±4,50±4	0.30
C_3S	CSH-FF	104.9	55.0	0.28
	core-shell [209]	103.0	54.5	0.28
	Exp. [352, 12]	112±6,122±4	52±2,56±2	0.3
CH	CSH-FF	28.8	12.4	0.31
	core-shell [209]	31.1	13.4	0.31
	Exp. [236, 356, 36, 12, 81]	29.4-60.7	13.6-28.0	0.315
$C_{1.0}SH_{0.5}$	CSH-FF	68.0	35.0	0.28
	Exp. [7]	60.5±3	38.3±2	-
$C_{1.75}SH_{2.0}$	CSH-FF	55.1	25.3	0.30
	Exp. [7]	52.5±7.5	24.5±3.5	-

reproduce the structure and mechanical properties of CSH at various Ca/Si ratios [7] and has shown to be in good agreement with density functional theory (DFT) calculations on crystalline CSHs such as Tobermorite polymorphs [306]. To avoid the computational expenses incurred by the calculation of coulombic interactions in real and reciprocal spaces, such interactions are calculated via the wolf method [357]. To investigate the transferability of CSH-FF potential to the crystalline phases in cement paste, we juxtapose its predictions including structure and mechanical properties against that of other force fields and experimental measurements. To this end, we compare the crystallographic lattice parameters [89, 228, 148] against atomistic simulation predictions by core-shell [209] and CSH-FF potentials as shown in Table 4.1. The CSH-FF predictions of crystallographic parameters (a, b, c, α, β and γ) after proper energy minimization at constant volume and pressure are very close to those calculated by the core-shell potential and measured experimentally. Therefore, we conclude that CSH-FF can appropriately predict the structure of crystalline phases in the hardened cement paste. There are a few methods to measure the elastic properties of solid phases in the cement paste including nano-indentation [7, 81, 352],

resonance frequency method [352], Brillouin spectroscopy [236], compacting powder samples of different porosity [356, 36] and multiscale micromechanical inverse analysis [12]. These elastic properties include bulk, k, shear, g, and indentation, M, moduli along with Poisson's ratio, ν. In the case of nano-indentation measurements, k and g are estimated from indentation modulus, M, and Poisson's ratio reported in the literature [81] using $k = M \times (1+\nu)(1-\nu)/3 \times (1-2\nu)$ and $g = M \times (1-\nu)/2$. To further explore the transferability of CSH-FF potential, we compare the experimental elastic measurements to those calculated from the atomistic simulations. To avoid misinterpreting the saddle points as local minima in the molecular simulations, we used rational functional optimization (RFO) which ensures the positive-definiteness of the Hessian matrix during the energy minimization process [123, 124]. Subsequently, the full elastic tensor, C_{ijkl}, is calculated from the second derivative of the potential energy $C_{ijkl} = 1/V\,(\partial^2 U_{lat}/\partial\epsilon_{ij}\partial\epsilon_{kl})$ where U_{lat} is the lattice potential energy and ϵ_{ij} is the second order strain tensor. In the atomistic simulations, the equivalent isotropic bulk and shear moduli are calculated within the Voight-Reuss-Hill bounds from the full elastic tensor. The Poisson's ratio is measured from bulk and shear moduli using $\nu = (3k-2g)/2\times(3k+g)$ transformation for isotropic media. These elastic properties are presented in Table 4.2. The CSH-FF potential correctly predicts the experimental elastic properties of the crystalline phases which are in full agreement with core-shell potential. Particularly for portlandite, we note that the simulation measurements via both CSH-FF and core-shell potential are close to that of Brillouin spectroscopy rather than powder compaction method. This is mainly because of the crystalline nature of portlandite which is fully accounted for in both atomistic simulation and Brillouin spectroscopy but rather missing in powder compaction experiments. In addition to crystalline phases, the CSH-FF potential appropriately predict the elastic properties of CSH at different Ca/Si ratios as presented in Table 4.2. Following this extensive transferability study, we ensure that CSH-FF can be properly utilized in the study of bulk properties of the crystalline calcium-silicate phases in the hardened cement paste.

4.4 Vibrational Densities of States of Calcium Silicate Phases

In solids, the vibrational motion of atoms at low temperature around their equilibrium positions can be approximated as harmonic oscillations which increasingly become anharmonic with rising temperature. These collective harmonic vibrations, the so-called phonons, characterize thermodynamic properties of solid materials such as free energy, heat capacity and thermal conductivity. Vibrational density of states, also known as phonon density of states, over frequency, $g(\omega)$, characterizes the number phonons having a frequency in the range of ω and $\omega + d\omega$:

$$g(\omega) = \frac{1}{3N-3} \sum_{s=4}^{3N} \delta(\omega - \omega_s) \tag{4.3}$$

where N is the number of atoms and ω_s is the s^{th} eigen vibrational frequency. The first three frequencies are neglected as they are associated with the translational motion of the solid. Within the harmonic approximation theory (i.e. small atomic displacements from equilibrium positions), $g(\omega)$ can be computed via two different approaches. The first method is eigenvalue decomposition of the dynamical matrix (EDDM) which is based on the notion of the normal mode analysis [229]. Any atomic vibration can be considered as a linear superposition of normal modes, i.e. the eigenvectors of the dynamical matrix $\{e^n, n = 1, \ldots, N\}$:

$$\sum_{q,\beta} D_{p,\alpha;q,\beta}\ e^n_{j\beta} = \omega_n^2 e^n_{p\alpha} \tag{4.4}$$

where p and q denote the atoms, α and β are the Cartesian coordinate indices and $\mathbf{D}$ is the dynamical matrix defined as:

$$D_{pq,\alpha\beta}(k) = \frac{1}{\sqrt{m_p m_q}} \left(\frac{\partial^2 U_{lat}}{\partial\alpha\partial\beta} \right) e^{ikr} \tag{4.5}$$

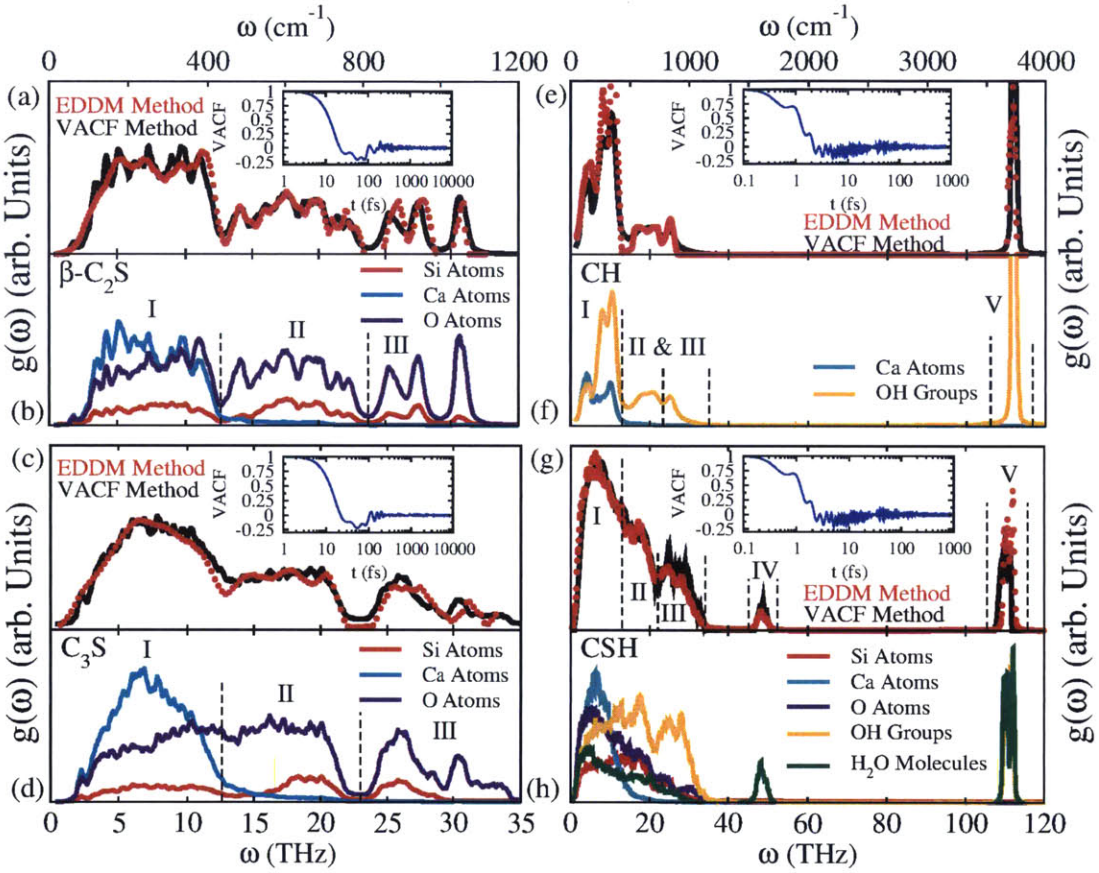

Figure 4-2: Vibrational Density of State (VDOS) for βC_2S, C_3S, CH and a CSH sample with Ca/Si=1.5. VDOS calculated via the eigenvalue decomposition of the dynamical matrix (EDDM) and Fourier transformation of the velocity autocorrelation function (VACF) for a) βC_2S, c) C_3, e) CH and g) CSH. The insets shows the evolution of VACF spanning four orders of magnitude. Decomposition of the VDOS to contributions from different atomic and molecular species for b) βC_2S, d) C_3, f) CH and g) CSH.

where m_p is the mass of p^{th} atom, U_{lat} is the lattice energy and e^{ikr} corresponds to the phase factor. There are a number of possible choices for the set of k-points in the first Brillouin zone, the primitive cell in the reciprocal space. Although CSH supercells are relatively large containing roughly 500 atoms, a fine 15*15*15 grid of k-points with Monkhorst-Pack scheme [235] is utilized to ensure proper sampling of the Brillouin zone.

The other method to compute $g(\omega)$ is via the Fourier transform of the velocity auto-correlation function (VACF) over a long enough molecular dynamics trajectory:

$$g(\omega) = \frac{1}{Nk_BT} \sum_{j=1}^{N} m_j \int_{-\infty}^{\infty} \langle \mathbf{v}_j(t).\mathbf{v}_j(0) \rangle \, e^{i\omega t} dt \tag{4.6}$$

where ω is the frequency, k_B is Boltzman constant, T is temperature, $\mathbf{v}_j(t)$ is its velocity and . denotes dot product. As shown in the inset of Fig. 4-2.a, due to the oscillatory nature of VACF ($\langle \mathbf{v}_j(t).\mathbf{v}_j(0) \rangle$ / $\langle \mathbf{v}_j(0).\mathbf{v}_j(0) \rangle$) femtosecond and even sub-femtosecond resolution is required to fully capture $g(\omega)$ (see the VACF insets in Fig. 4-2).

Fig. 4-2.a, 4-2.c, 4-2.e and 4-2.g present the VDOS of β-C_2S, C_3S, CH and a CSH model with Ca/Si=1.5, respectively. As can be seen in these figures, the two approaches, VACF and EDDM, yield almost identical results for all the molecular models. To better classify the different atomic bonds' contribution to the VDOS in a complex materials such as CSH, we divide the phonon spectrum to five distinct vibrational bands, denoted numerically by I through V in Fig. 4-2.g. We recognize the underlying atomic and molecular vibrations populating the aforementioned bands by decomposing the VACF to its partial contributions, VACF_γ where $\gamma = \{Si, Ca, O, OH, H_2O\}$. These partial contributions to VDOS are presented in Fig. 4-2.b, Fig. 4-2.d, Fig. 4-2.f and Fig. 4-2.h for β-C_2S, C_3S, CH and CSH, respectively. The VDOS partial contributions unanimously assert that the peak in I band is clearly attributed to the vibration of Ca-O bonds. This is in full agreement with the near-infrared spectroscopy experiments of Yu et al. [369] associating $\omega < 400\ cm^{-1}$ to stretching of Ca-O bonds. Unlike Ca-O phonons, vibrations of

SiO_4 units extend over a large band from 0 up to 35 THz. As studied in details for amorphous silica [327], the Si-O frequencies in I band is related to the acoustic-like rigid-body rotation of SiO_4 tetrahedra units in the silica network. In the case of β-C_2S and C_3S, this is associated with the isolated rigid-body rotation of silica monomers. However, in the case of CSH, these vibrations are related to the coupled rotation of SiO_4 units in the silica chains. The simulation results indicate that the II band, 400 $cm^{-1} < \omega <$ 750 cm^{-1}, is dominated by the vibration of silica tetrahedra units. This is in agreement with mid-infrared experiments [369] relating this band to stretching of Si-O bonds, and symmetric and asymmetric bending of Si-O-Si angles. Taraskin and Elliot [327] express that the vibrations in this frequency window are associated with a complex hybridization of acoustic and optical modes manifested in quasi-localized internal stretching and bending of SiO_4 units in silica glasses. The III band in the range of 750 $cm^{-1} < \omega <$ 1200 cm^{-1} is also populated with the silica tetrahedra vibrations. In β-C_2S and C_3S, these vibrations are accumulated into peaks that are associated with highly localized longitudinal and transverse optical modes that correspond to stretching and bending of silica monomers. In particular for CSH, the III band is mainly associated with the vibration of Q_1 and Q_0 sites containing silanol groups (Si-OH) due to the strong presence of THE OH signal in the partial VDOS. The vibrational frequencies in the IV region belong to the in-plane bending of H-O-H angle in water molecules because it is only present in the partial VDOS of water molecules and it is absent in the VDOS of portlandite. The V band represents both the symmetric and asymmetric stretching of O-H bonds in water molecules, stretch of hydoxyl groups coordinated to both Ca and Si atoms in CSH and stretch of hydroxyl groups in portlandite.

One of the most striking features in the VDOS of glasses, glass-forming and supercooled liquids is the presence of a universal excess of states (EoS) compared to that predicted by the Debye model ($g_D \propto \omega^2$) in the low frequency region [304]. This EoS is responsible for the anomalies of the heat capacity and thermal conductivity of glasses at low temperatures [373] and is highlighted by a broad peak in the THz region (<5 THz) when plotting $g(\omega)/\omega^2$ and is referred to as the *Boson Peak* (BP).

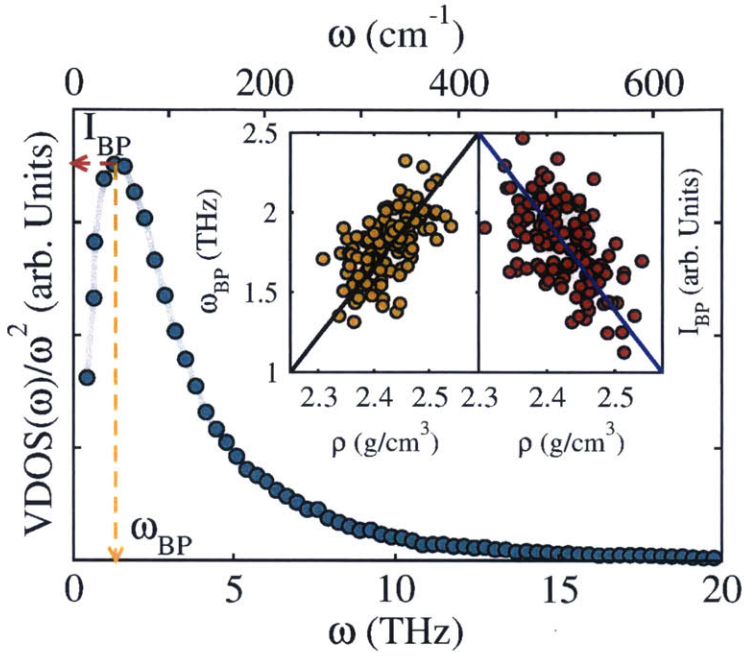

Figure 4-3: Boson peak identification in a CSH sample with Ca/Si=1.5. The inset demonstrates the impact of CSH density on the location (ω_{BP}) and intensity (I_{BP}) of the Boson peak.

The BP can be experimentally identified via Raman spectroscopy and one-phonon scattering cross section in inelastic neutron scattering [316]. Despite decades of work on the origin of the BP, it remains an open question in condensed matter physics and material science [307]. The present theoretical models explain the BP via different mechanisms such as phonon-saddle transition in the energy landscape [135], local vibrational modes of clusters [100], locally favored structures [324], liberation of molecular fragments [54, 136], vibrations in anharmonic potentials [258], and anomaly in transverse phonon propagation related to Ioffe-Regel limit [307]. Exhibiting common features of disordered and amorphous materials [7, 33], the VDOS of CSH presents a BP in the THz region as shown in Fig.4-3. BP is characterized by the peak position, ω_{BP}, and the peak intensity, I_{BP}. As observed in many glass former materials such as polybutadiene, polystyren [316], LiCl [326] and B_2O_3 [109], ω_{BP} is usually shifted to the lower frequencies with increasing temperature. Being an exception to this rule [355], ω_{BP} in silica is left shifted with decreasing density [321, 166]. The

Table 4.3: Equilibrium properties of β-C_2S, C_3S, CH, $C_{1.0}SH_{0.5}$ and $C_{1.75}SH_{2.0}$ including density, compressibility, coefficient of thermal expansion, constant volume and constant pressure specific heat capacities calculated using CSH-FF potential and compared against available experimental measurements.

	c_v (J/g/K)	ρ (g/cm^3)		β (1/Pa)	α (1/K)	c_p (J/g/K))	
	Sim.	Sim.	Exp.	Sim.	Sim.	Sim.	Exp.
β-C_2S	0.69	3.55	3.31 [217],3.36 [228]	8.18E-12	4.99E-5	0.72	0.71 [217],0.75 [338]
C_3S	0.68	3.57	3.13 [217],3.55 [89]	9.53E-12	4.95E-5	0.71	0.72 [217],0.75 [338]
CH	1.14	2.25	2.25 [217],2.08 [148]	3.47E-11	9.91E-5	1.14	1.15 [217]
$C_{1.0}SH_{0.5}$	0.87	2.55	2.48 [335],2.46 [284]	1.47E-11	4.50E-5	0.89	-
$C_{1.75}SH_{2.0}$	0.95	2.40	2.47 [242],2.60 [17]	1.82E-11	4.50E-5	0.97	-

analysis of all CSH models indicates that the composition affects both positions and intensities of BPs. As expressed in terms of density in the inset of Fig. 4-3, CSH's ω_{BP} decreases with increasing density and its I_{BP} behaves conversely. The correlation between CSH's density and its BP properties are similar to pressure-dependence of BP observed in silica [321], $Na_2FeSi_3O_8$ [233], and polymers [149]. To our knowledge, this is the first example of a composition-driven BP shift.

4.5 Nanoscale Heat Capacity Measurements of Calcium Silicate Phases

Heat capacity is a thermal property of materials at equilibrium that links variations of internal energy (U) and temperature.The internal energy of an insulating solid is the sum of lattice, U_{lat}, vibrational, U_{vib}, rotational, U_{rot}, and translational, U_{tra}, energetic contributions. At low temperatures, considerably below Debye temperature [175], the contributions from U_{rot} and U_{tra} are negligible compared to the other two components, hence $U \simeq U_{lat} + U_{vib}$. U_{vib} can be described as the number of phonons occupying a particular state, $n(\omega)$, times their vibrational energy level, $h\omega$. This is written as:

$$U_{vib} = \int_0^\infty h\omega \times \left(n(\omega) + \frac{1}{2}\right) d\omega \tag{4.7}$$

where h is the Plank constant and $h\omega/2$ is zero point energy contribution. Following Bose-Einstein statistics, $f_{BE}(\omega) = \dfrac{1}{e^{-h\omega/k_BT} - 1}$ is the probability distribution of a

boson particles occupying a specific quantum state in thermal equilibrium. Thus, the occupation level is the degeneracy of the vibrational state, described through $g(\omega)$, times the probability of presence of phonons at that specific state, $n(\omega) = g(\omega) \times f_{BE}(\omega)$. This yields the kinetic energy in terms of vibrational density of states in eqs. 4.3 and 4.6. The specific heat capacity at constant volume can be calculated, $c_v = \left(\frac{\partial U}{\partial T}\right)_v = \left(\frac{\partial U_{vib}}{\partial T}\right)_v$, noting that U_{lat} does not depend on T at constant volume. Experimental measurement of c_v is rather difficult as it requires maintaining constant volume throughout the experiment. Therefore, the laboratory measurement of specific heat capacity at constant pressure, c_p, is more popular. From thermodynamics, c_v and c_p are related via:

$$c_p - c_v = T \times \frac{\alpha^2}{\rho\beta} \tag{4.8}$$

where ρ is the density, $\alpha = 1/V\,(\partial V/\partial T)_P$ is the coefficient of thermal expansion and $\beta = -1/V\,(\partial V/\partial P)_T$ is the compressibility which is the reciprocal of bulk modulus. These two coefficients can be calculated via finite difference approach in an isobaric-isothermal (NPT) ensemble. To this end, three 10 ns-long simulations are performed on a number of CSH samples, β-C_2S, C_3S and CH at (300 K, 0 atm), (310 K, 0 atm), and (300 K, 100 atm). The results for β-C_2S, C_3S, CH and two CSH models with stoichiometry of $C_{1.0}SH_{0.5}$ and $C_{1.75}SH_{2.0}$ are summarized in Table 4.3. The compressibility of the crystalline phases are almost identical to those calculated at zero temperature and presented in Table 4.2. The densities of the crystalline phases are compared to experimental values of bulk measurements [217] and crystallographic density [89, 228, 148]. Since the effect of porosity and impurities are neglected in our atomistic simulations, our densities are rather close to crystallographic measurement than those in bulk measurements. CSH's compressibility increases with increasing Ca/Si ratio, 1.5×10^{-11} Pa^{-1} $< \beta_{CSH} <$ 1.8×10^{-11} Pa^{-1}. These results are in full agreement with zero temperature calculations of CSH's bulk modulus [7], k_{CSH}, indicating decrease in κ_{CSH} with increasing Ca/Si ratio, 55 GPa $< k_{CSH} <$ 65 GPa. CSH's coefficient of thermal expansion is calculated to be $4.5(\pm 0.9) \times 10^{-5}$ K^{-1} which

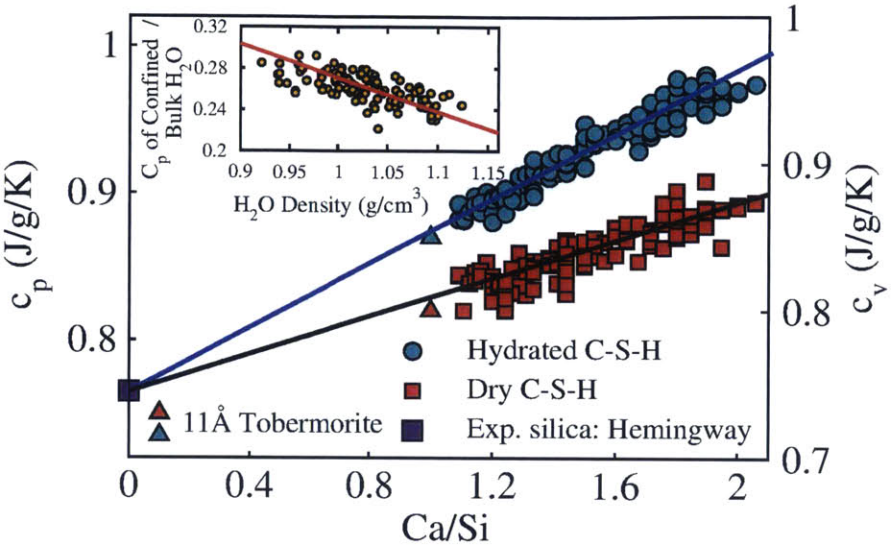

Figure 4-4: The effect of Ca/Si ratio on specific heat capacity values at constant volume and pressure for dry and hydrated CSH models and 11ÅTobermorite. The black and blue lines are fitted to the atomistic simulation data. These lines intersect at a point corresponding to the heat capacity of amorphous silica [147]. The inset indicate the relation between the density of the nanoconfined water and its apparent heat capacity measured as the difference between the heat capacity of dry and hydrated CSH.

is in agreement with 4.2×10^{-5} K^{-1} obtained from micro-thermo-poromechanics back analysis [128]. CSH's density is found to decrease with increasing Ca/Si ratio, 2.55 g/cm^3 $> \rho_{CSH} >$ 2.35 g/cm^3. At low Ca/Si ratio, the density is close to that of experimental density reported for Tobermorite minerals [284, 335]. The density of CSH at high Ca/Si ratio, is slightly lower than those measured by neutron scattering [17] and close to the recent experimetal values reported by Muller et al. [242] after subtracting the mono-layer of water adsorbed on the external surface of CSH nanoparticles. Having computed α, β and ρ, the difference between c_p and c_v is calculated to be in the range of $0.002 - 0.025$ J/g/K for different calcium-silicate systems close to experimental reported values for minerals.

Table 4.3 presents the specific heat capacities at constant volume and pressure for β-C_2S, C_3S and CH. The atomistic simulations prediction of constant pressure heat capacities are in very good agreement with low temperature heat capacity mea-

surements [338] and those calculated from fitted c_p-T relations [217] . This can also be regarded as another validation for the transferability of the CSH-FF potential to other calcium-silicate systems that allows refining not only structural but also vibrational properties. Fig. 4-4 displays the specific heat capacity of both dry (all molecular inter-layer water removed), $c_{v,p}^{dry}$, and hydrated, $c_{v,p}^{hyd}$, CSH as a function of Ca/Si ratio. The specific heat capacities in both dry and hydrated samples increase almost linearly with Ca/Si ratio with a minor scatter due to the polymorphic structure of CSH [7]. Similar to the experimentally observed increase in the heat capacity of rocks [74] and Vycor glass [339], the heat capacity of hydrated CSH is higher than that of dry samples. The constant pressure specific heat capacities can be written in the form of $c_p^{dry} = 0.66x + 0.75$ and $c_p^{hyd} = 1.00x + 0.75$, where x is the Ca/Si ratio. The two lines intersect at 0.75 J/g/K at Ca/Si ratio equal to zero which is the experimental heat capacity of amorphous silica (devoid of calcium) at room temperature [147]. The specific heat capacity of CSH is less than those observed for cement pastes with different water to cement ratios and saturation degrees [39]. The heat capacity of cement paste is a composite response of different phases including unhydrated clinker, portlandite, and water which will be discussed in sec. 4.7.

The difference between dry and hydrated samples indicates the effect of inter-layer water on the heat capacity of CSH. In fact, the apparent heat capacity of nanoconfined water can be considered as the difference between the specific heat capacities of wet and dry CSH, $\Delta c_p = c_p^{hyd} - c_p^{dry} = 0.34x$. According to the average chemical composition (see section 4.2), the inter-layer water content scales with $\Delta M_w = 0.8x$. Since $\Delta c_p / \Delta M_w < 1$, then the heat capacity of the nanoconfined water decreases with increasing Ca/Si ratio. It is more rational to present the variation of the heat capacity of nanoconfined water in terms of its density rather than Ca/Si ratio. To this end, the density of inter-layer water is measured via the Voronoi tessellation algorithm, which finds a portion of space that is closer to a given water molecule than any other atoms in the molecular structure [274, 273]. The inset of Fig. 4-4 shows the relation between the normalized apparent heat capacity and the density of the nanoconfined water. The heat capacity of water in the ultra-confining interlayer spacing of CSH

($d < 1.0$ nm) is noticeably smaller than that of bulk water at room temperature, roughly 4.2 J/g/K [186]. This is in full agreement with Bentz's postulate stating that the heat capacity of chemically and physically bound water with the hydration gel should be significantly lower than that of bulk water [39]. Similar to the dynamical properties [274], the heat capacity of the nanoconfined water in CSH and its trend with density behaves like supercooled water. In fact, although being at the room temperature, Δc_p is in the order of those observed at 150 K [19]. This is mainly due to the strong interactions between the water molecules and the substrate [274].

4.6 Nanoscale Heat Conductivity Measurements of Calcium Silicate Phases

The thermal conductivity, $\mathbf{K}$, of a dielectric material is a second order tensor relating the temperature gradient, ∇T, to the heat flux, $\mathbf{q}$, through Fourier's law, $\mathbf{q} = \mathbf{K}.\nabla T$. There are three methods to calculate $\mathbf{K}$ of an insulating solid via MD simulation. These approaches are the Green-Kubo (GK) which is an equilibrium method, a direct application of Fourier's law within the steady state conduction regime in a non-equilibrium framework (known as direct method) and a transient non-equilibrium method. We utilize the GK approach as it is less size-dependent, does not involve imposing unrealistic temperature gradients in the simulation cell and yields the full thermal conductivity tensor. Heat flux in a multi-particle systems can be simply written as [222]:

$$\mathbf{q} = \frac{1}{V}\frac{d}{dt}\sum_{i=1}^{N}\mathbf{r}_i H_i \tag{4.9}$$

where $\mathbf{r}_i$ and H_i are the position vector and total Hamiltonian (kinetic plus potential energies) of the i^{th} particle in the system and the V is volume respectively. To exclude the numerical errors arising from the calculation of time derivative in finite difference approach, the above expression can be rearranged considering the mathematical form of the inter-atomic potential. Assuming that the angular contributions in the CSH-FF potential are equally distributed among all atoms in a water molecule, H_i can be

expressed in terms of its components via [300]:

$$H_i = \frac{1}{2} m_i \mathbf{v}_i . \mathbf{v}_i + \frac{1}{2} \sum_j u_2(r_{ij}) + \frac{1}{6} \sum_{kj} u_3(\theta_{jik}) \tag{4.10}$$

where u_2 represents pairwise energy terms between the i^{th} particle and its j^{th} neighbor located r_{ij} apart from each other which includes Van der Waals, Coulomb interactions via Wolf summation method, and bond stretching terms. The u_3 energy term considers the bending energy of the central θ_{ijk} angle in water molecules. By substituting eq. 4.10 in eq. 4.9, an alternative expression for heat flux is achieved:

$$\mathbf{q} = \frac{1}{V} \sum_{i=1}^{N} [e_i \mathbf{v}_i - V \mathbf{S}_i . \mathbf{v}_i] \tag{4.11}$$

where e_i is the kinetic energy and $\mathbf{S}_i$ is the atomic Virial stress tensor which can be expressed in terms of two- and three-body potential terms as [192]:

$$\mathbf{S}_i = \frac{1}{2V} \sum_{j=1}^{N_p, N_b} \mathbf{r}_{ij} \otimes \mathbf{F}_{ij} + \frac{1}{6V} \sum_{i=1}^{N_a} (\mathbf{r}_{ij} + \mathbf{r}_{ik}) \otimes \mathbf{F}_{ijk} \tag{4.12}$$

where $\otimes$ is dyadic product in tensor notation, $\mathbf{F}_{ij}$ is the force due to pair potentials, and $\mathbf{F}_{ijk} = - \bigtriangledown_i u_3(\theta_{jik})$ is a three-body force term. We have readily implemented the above heat flux formulation in Gulp [123, 124] source code. In equilibrium, the net heat flux $\mathbf{q}(t)$ fluctuates around zero along the molecular dynamics trajectory. To measure thermal conductivity at the equilibrium, the GK approach links $\mathbf{K}$ to the dissipation rate of heat flux fluctuations at the atomic scale via the fluctuation-dissipation theorem. The GK formalism for an anisotropic material can be written as [222]:

$$\mathbf{K} = \frac{V}{k_B T^2} \int_0^{\infty} \langle \mathbf{q}(t) \otimes \mathbf{q}(0) \rangle \, dt \tag{4.13}$$

where $\bar{\Psi}(t) = \langle \mathbf{q}(t) \otimes \mathbf{q}(0) \rangle \, / \, \langle \mathbf{q}(0) \otimes \mathbf{q}(0) \rangle$ is a second order tensor generally known as heat flux auto-correlation function (HFACF). HFACF elements are unity at the origin and subsequently decays to zero with characteristic time scales proportional to

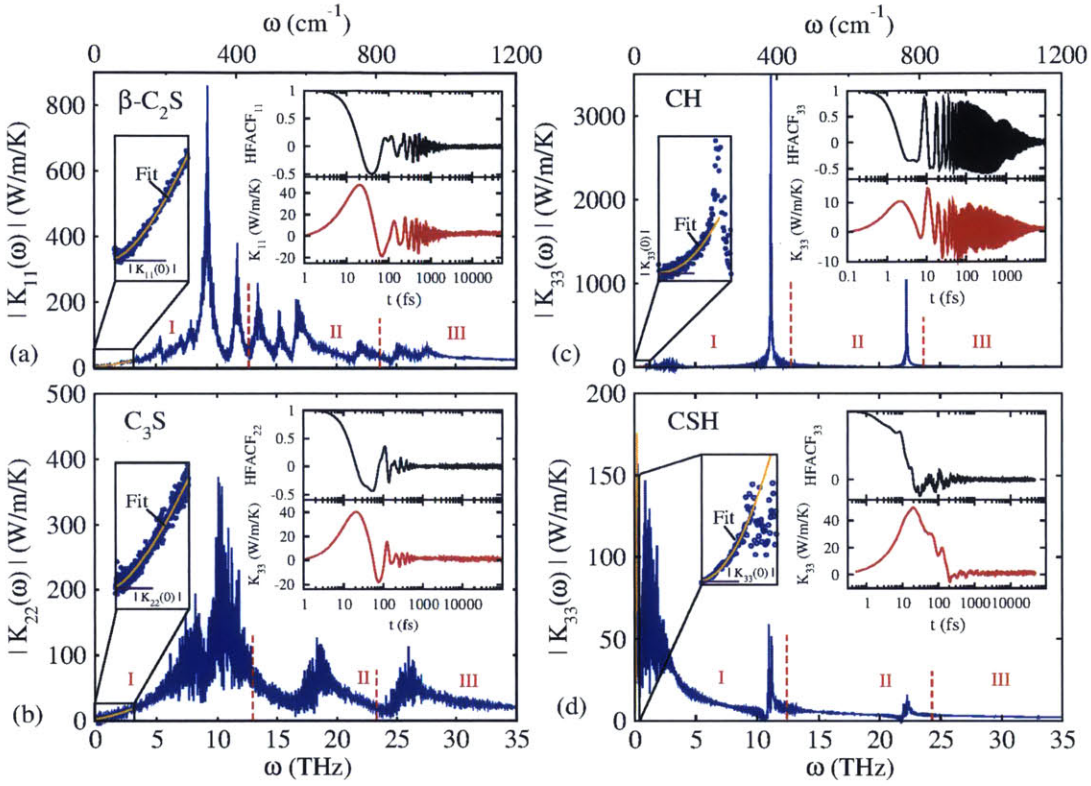

Figure 4-5: The power spectra of different components of the thermal conductivity tensor in frequency domain (Fourier transform of the heat flux autocorrelation function) for β-C_2S, C_3S, CH, and a CSH sample with Ca/Si=1.5. The insets in black and red correspond to the heat flux autocorrelation function and the thermal conductivity in time domain, respectively, exhibiting pronounced oscillations in long time-scale. The zoom-in inset shows the fitting procedure adopted here to estimate the thermal conductivity at the zero frequency which corresponds to Green-Kubo relation.

the thermal conductivity of the material. Since the length of the MD trajectory is finite, the numerical estimation of $\mathbf{K}$ is usually truncated after convergence of HFACF to zero, t_M, and it is calculated via [341]:

$$K_{ij}(t_M) = \frac{\Delta t \times V}{k_B T^2} \sum_{m=1}^{M} \frac{1}{s-m} \sum_{n=1}^{s-m} \frac{q_i(m+n)q_j(n)}{\langle q_i(0)q_j(0)\rangle} \tag{4.14}$$

where Δt is the heat flux outputting time step which is different from the molecular dynamics time step δt, s is the total number of output steps, t_M is given by $M\Delta t$, and $q_j(n)$ is the j^{th} component of the heat flux at n^{th} output step.

A number of complications arise when calculating the thermal conductivity using eq. 4.14. First, due to the finite value of s, imposed by limitations on the disk space and computational expense, thermal conductivity might be erroneously shown to be non-symmetric. However, $\mathbf{K}$ is mathematically shown to be both positive semi-definite and symmetric [269, 88]. This is the artifact of the numerical calculation of HFACF, $\epsilon = |\bar{\Psi}_{ij}(t) - \bar{\Psi}_{ji}(t)| > 0 \mid \text{for } i \neq j$. ϵ can be minimized by increasing s which requires excessive computational expense. Another way to exclude ϵ is to enforce symmetry via $\Psi_{ij} = (\bar{\Psi}_{ij} + \bar{\Psi}_{ji})/2$. In fact, the ergodicity of the equilibrium processes requires $\bar{\Psi}_{ij} = \bar{\Psi}_{ji}$, and, therefore the symmetry approach is more logical in this sense. Second, as shown in the insets of Fig. 4-5.a-d, some oscillations of HFACF might still be present at large time scales ($t > 10^5$ fs). To better understand the nature of these oscillations, we note that GK relation (eq. 4.13) can be expressed in frequency space:

$$\mathbf{K}(\omega) = \frac{V}{k_B T^2} \int_0^\infty \langle \mathbf{q}(t) \otimes \mathbf{q}(0) \rangle \, e^{i\omega t} dt \tag{4.15}$$

where the thermal conductivity can be regarded as the limit of $\mathbf{K}(\omega)$ at the zero frequency. Fig. 4-5.a-d present different elements of $\mathbf{K}(\omega)$ for β-C_2S, C_3S, CH and a CSH model with the Ca/Si=1.5, respectively. It is noted that the high frequencies in the IV and V bands defined in VDOS (Fig. 4-2) are absent in $\mathbf{K}(\omega)$. This simply means that the hydroxyl groups and water molecules do not contribute in the oscillations of heat conduction. This is due to the dangling nature of hydroxyl groups which do not contribute to the propagation of phonons, discussed later in this section. In the case of K_{11} and K_{22} for β-C_2S and C_3S, the contributing frequencies to the oscillation of thermal conductivity extend over I, II and III bands and exhibits distinct peaks. By decomposing eq. 4.9 into $q_u = 1/V \sum_{i=1}^{N} \mathbf{v}_i U_i$ and $q_k = 1/V \sum_{i=1}^{N} \mathbf{r}_i \dot{U}_i$, Landry et al. [187] associated the oscillation in $\mathbf{K}(t)$ to optical zero wave vector phonons. This explanation is consistent with our observation of the absence of such peaks at low frequencies in β-C_2S and C_3S pertaining to the longitudinal and transverse acoustic phonons. The heat conduction across the interlayer spacing of CH and CSH is more

sophisticated. The absence of bonded interaction between the layers, also evidenced in elastic properties ($C_{11} > C_{33}$ and $C_{22} > C_{33}$ following Voight notation), makes the measurement of K_{33} more complicated. As presented in Fig. 4-5.c-d, the optical zero wave vector phonons contributing to the oscillations of $K_{33}(\omega)$ are only present at two sharp peaks in I and II bands. However, unlike β-C_2S and C_3S, significant peaks emerge at low frequencies. This is an evidence of complex interplay between acoustic and optical phonons which is also observed in the study of VDOS of amorphous silica [327]. As shown in the zoomed insets of Fig. 4-5.a-d, $K_{ii}(\omega)$ decays smoothly at small enough frequencies to its limit value. To estimate this limit thermal conductivity (k_{ij}), we fitted a power function of the form $K_{ij}(\omega) = K_{ij} + \xi \times \omega^{\zeta}$ where K_{ij}, ξ and ζ are estimated using the least square approach. In addition, 10 independent simulations with different initial conditions are performed for each molecular model and the mean and standard deviation of thermal conductivity tensor are computed subsequently. Each simulation is 10 ns long and $\mathbf{q}(t)$ is recorded every 0.1 fs for CH and CSHs and every 1.0 fs for β-C_2S and C_3S). The thermal conductivity calculated using this method is found to be close to the average value of thermal conductivity in the time domain. Finally and particularly for CSH, the value of the thermal conductivity is dependent on Δt, the resolution in which $q(t)$ is recorded. In fact, the HFACF plunges very quickly to zero and this suggests the necessity of capturing short time-scale behaviors of $\Psi(t)$. To this end, we measured the thermal conductivity of CSH with different Δt resolutions using the above method. The true thermal conductivity of CSH is then the limit of the $K_{ij}(\Delta t)$ at infinitesimally small Δt values. We find that the thermal conductivity of CSH is linearly dependent on Δt, $K_{ij}(\Delta t) = \zeta * \Delta t + K_{ij}^0$. Therefore, the intercept at the origin is considered to be the actual thermal conductivity value of CSH, $K_{ij} = \lim_{\Delta t \to 0} K_{ij}(\Delta t)$. Following the above procedure, the full thermal conductivity tensor is computed for all models. In the case of β-C_2S and C_3S, we find $\mathbf{K}$ to be diagonal ($K_{ij} = 0\ ,\ i \neq j$). Therefore, the [100],[010] and [001] are the same as the principal thermal conductivity directions. For CH and CSH models, we generally find that $K_{11} > K_{22} > K_{12}$ and $K_{12} \gg K_{13}^0, K_{23}^0$. Therefore, the thermal conductivity of CH and CSH in Cartesian coordinate system,

Table 4.4: Transport properties of β-C_2S, C_3S, CH, $C_{1.0}SH_{0.5}$ and $C_{1.75}SH_{2.0}$ including principal thermal conductivity values, volumetric thermal conductivity, longitudinal, transverse and volumetric acoustic velocities and the mean free path of phonons calculated using CSH-FF potential.

	K_I (W/m/K)	K_{II} (W/m/K)	K_{III} (W/m/K)	K_v (W/m/K)	S_l (km/s)	S_t (km/s)	S_v (km/s)	l_m (Å)
β-C_2S	3.45±0.4	3.45±0.4	3.45±0.4	3.45±0.4	21.4	12.0	13.3	3.2
C_3S	3.35±0.3	3.35±0.3	3.35±0.3	3.35±0.3	22.4	12.6	14.0	2.9
CH	2.00±0.2	1.20±0.2	0.75±0.2	1.32±0.2	14.1	7.0	7.9	1.7
$C_{1.0}SH_{0.5}$	1.25±0.2	0.95±0.2	0.74±0.2	0.98±0.2	19.9	10.2	11.4	1.4
$C_{1.75}SH_{2.0}$	1.25±0.2	0.95±0.2	0.74±0.2	0.98±0.2	19.2	11.67	12.9	1.6

where 1, 2 and 3 directions are respectively [100], [010] and [001] crystallographic directions, can be approximated as:

$$\mathbf{K} \simeq \begin{bmatrix} k_{11} & k_{12} & 0 \\ k_{12} & k_{22} & 0 \\ 0 & 0 & k_{33} \end{bmatrix} \tag{4.16}$$

The form of the thermal conductivity tensor presented in eq. 4.16 is not only symmetric but also positive-definite because of the magnitude of K_{ij} elements. Tensor of thermal conductivity can be expressed in terms of its principal thermal conductivities, the eigenvalues of the **K** tensor. The principal thermal conductivity values for β-C_2S, C_3S, CH and two CSH models are summarized in Table 4.4. The three principal conductivity values for CSH models, $K_{III} < K_{II} < K_I$, are presented in Fig. 4-5.a-c. While the eigenvectors corresponding to K_I and K_{II} are located in the 12 plane (calcium silicate layers), the 3 axis (across the interlayer spacing) is fully aligned with eigenvector associated with K_{III}. Similar to anisotropic nature of mechanical properties in CSH [7], the heat transport properties also show strong anisotropy such that the thermal conductivity in the defective calcium-silicate sheets (12 plane) is statistically larger than that of out of plane direction (3 axis). While the XY plane is populated with fairly heat conductive Si-O bonds, looser Ca-O bonds and water molecules in the interlayer space scatter phonons and diminish the heat transport along the z direction perpendicular to the layers.

Despite the clear presence of anisotropy in heat transport properties, we find no correlation between the chemical composition and thermal conductivity of CSH. To

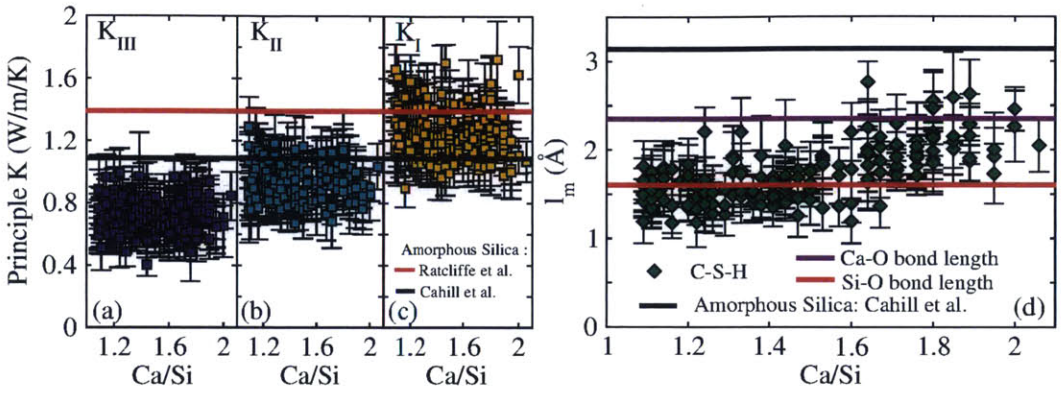

Figure 4-6: The anisotropic nature of thermal conduction in CSH and the effect of stoichiometry on the principal thermal conductivity values a) K_I, b) K_{II} and c) K_{III} compared against the experimental values of thermal conductivity of amorphous silica measured by Cahill et al. [55] and Ratcliffe [280]. d) The mean free path of phonons (l_m) in CSH compared against mean free path of phonons in amorphous silica as measured by Cahill et al. [55, 56] and Si-O bond length.

this end, we investigate the absence of such correlation via a comparative analysis with crystalline phases. The volume thermal conductivity of CSH does not alter with Ca/Si ratio and it is close to that of 11Å Tobermorite, see Table 4.4. The above K_v values for CSH are close to those measured experimentally [280, 340, 112, 55] and calculated numerically [167] and theoretically [18, 57] for amorphous silica. Since the variation of Ca/Si ratio was merely achieved by removing SiO_2 groups in the preparation stage, it can be viewed as a parameter that is inversely proportional to the defect content. In the absence of interface and boundaries in the bulk material, there are two phonon scattering mechanisms: phonon-phonon and phonon-defect scattering. Assuming that different scattering processes are independent, Matthiessen's rule combines the two scattering mechanisms to an effective process using:

$$\frac{1}{l_m} = \frac{1}{\bar{l}_{ph-vacancy}} + \frac{1}{\bar{l}_{ph-ph}} \tag{4.17}$$

where l_m is the mean free path of phonons, and, $\bar{l}_{ph-vacancy}$ is the mean phonon-vacancy scattering length, a distance that a phonon can propagate prior to being scattered off a vacancy defect. $\bar{l}_{ph-ph}$ is the mean phonon-phonon scattering length,

a distance that a phonon travels before being scattered through an Umklapp (U-) or normal (N-) process. The U- and N-process explain the temperature-dependence of thermal conductivity and their l_{ph-ph}^{-1} scales with $T^n \times e^{1/T}$ [268, 14] and T^m [40], respectively, where m and n constants are material specific and are identified by fitting Callaway's model [58] to the experimental data or the results of numerical simulation. The Matthiessen's rule specifies two limiting regimes for phonon transport in a material: vibration-dominated ($\bar{l}_{ph-ph} \ll \bar{l}_{ph-vacancy}$) and defect-dominated ($\bar{l}_{ph-vacancy} \ll \bar{l}_{ph-ph}$) regimes. Therefore, if a material is in the vibration-dominated regime, then the population of defects would not affect the heat transport properties. In such case, the kinetic theory formulation of heat transport can be used to investigate the characteristic length scale of phonons [374]:

$$K_v = \frac{1}{3}\rho C_v \times S_v \times l_m \tag{4.18}$$

where S_v is the speed of sound and is estimated by averaging the longitudinal and transverse sound velocities, S_l and S_t, respectively, using [42]:

$$S_v = \left[\frac{1}{3}\left(\frac{1}{S_l^3} + \frac{2}{S_t^3}\right)\right]^{-1/3} \tag{4.19}$$

where $S_l = \sqrt{\frac{4g+3k}{3\rho}}$ and $S_t = \sqrt{\frac{g}{\rho}}$ are the polycrystalline averages of these acoustic velocities computed from the bulk, k, and shear, g, moduli and the density presented in Tables 4.2 and 4.3. The acoustic velocities and mean free path of phonons for β-C_2S, C_3S, CH and two types of CSH are provided in Table 4.4. The mean free path of in CSHs and CH is almost half of that in β-C_2S and C_3S. The space filling structure of the clinker phases explains their relative high thermal conductivity and the mean free path of phonons. The thermal conductivity values observed for β-C_2S and C_3S are close to the thermal conductivity of crystalline silica [280]. However, the layered structure of CSH and CH scatters phonons across the interlayer spacing and significantly reduces the heat transport properties. Since l_m is in the order of Si-O and Ca-O bond lengths ($d_{Si-O} = 1.62$Å and $d_{Ca-O} = 2.42$Å) across the range of considered Ca/Si ratios in CSH, it can be safely assumed to be in the vibration-

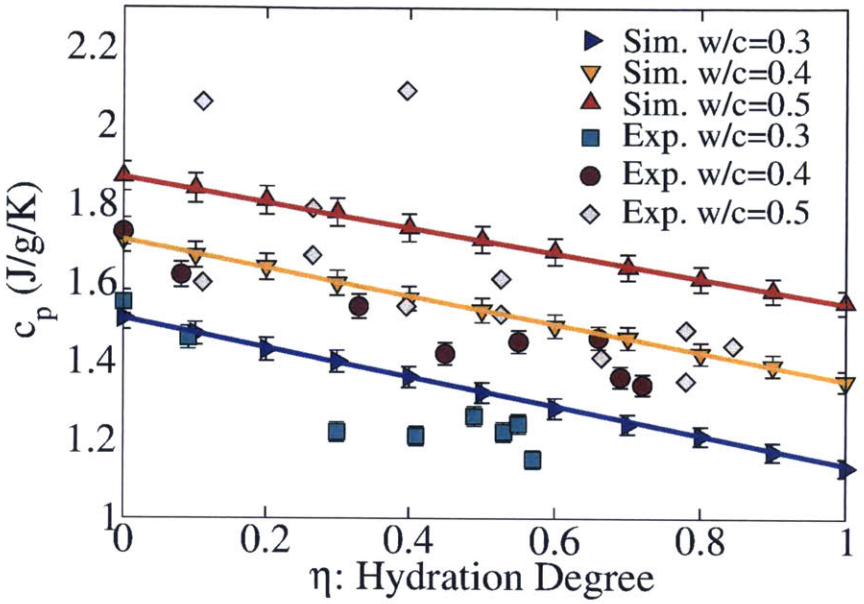

Figure 4-7: The effect of hydration degree on the macroscopic specific heat capacity of hydrating cement paste for three water-to-cement ratios (w/c). The simulation results derived from atomistic simulation and mixture laws are compared with experimental measurements of Bentz [39] for w/c=0.3 and 0.4 and De Schutter et al. [90] for w/c=0.5.

dominated regime (Fig. 4-6.d). Based on experimental data, Cahill et al. [55, 57] concluded that l_m of amorphous silica at room temperature is in the order of two Si-O bonds.

4.7 Upscaling of Thermal Proprties

In sections 4.5 and 4.6, we explored the effect of chemistry on the nanoscale equilibrium and non-equilibrium thermo-physical properties of various calcium-silicate phases present in the cement paste along the dissolution-precipitation reaction path up to the hardened cement paste. In this section, we aim at linking the nanoscale properties of individual constituent phases of cement paste to that observed for their mixture at the macroscale.

4.7.1 Upscaling Heat Capacity from Nano to Macro Scale

The upscaling of specific heat capacity from the atomic scale (Level 0) to the macroscale (Level 3) is rather straightforward. Since energy and volume are extensive thermodynamic quantities, they can be written as the sum of those quantities for individual phases in a composite material. Therefore, the constant pressure specific heat capacity of a composite can be considered as a linear combination of that of individual phases. This would also hold true during the course of chemical reaction involving m reactive agents (R_i) and n reaction products (P_i):

$$\alpha_1 R_1 + \cdots + \alpha_m R_m \rightarrow \beta_1 P_1 + \cdots + \beta_n P_n \tag{4.20}$$

where α_i and β_i are the number of moles of reactants and products in the balanced stoichiometry, respectively. Investigations on the properties of a mixture of reactants and products can be performed in terms of reaction degree η, i.e. the mass of reactants at a given time divided by total mass of reactants and products. Hence, the constant pressure specific heat capacity of a mixture of reactants and products in terms of reaction degree can be written as:

$$c_p^{mix} = (1-\eta) \times \sum_{i=1}^{m} \phi_{R_i} c_p^{R_i} + \eta \times \sum_{i=1}^{m} \phi_{P_i} c_p^{P_i} \tag{4.21}$$

where $c_p^{R_i}$ and $c_p^{P_i}$ are constant pressure specific heat capacities of i^{th} reactant and product, respectively. $\phi_{R_i} = \alpha_i M_{R_i} / \sum_i \alpha_i M_{R_i}$ and $\phi_{P_i} = \beta_i M_{P_i} / \sum_i \beta_i M_{P_i}$ are the mass fractions of the i^{th} reactant and product with molar mass of M_{R_i} and M_{P_i}, respectively. In the light of eqs. 4.20 and 4.21, the constant pressure specific heat capacity of hydrating cement paste can be expressed in terms of hydration degree. Eqs. 4.1-4.2 and and Table 4.3 provide the chemical reactions and individual specific heat capacities required for such calculations.

Fig. 4-7 presents the constant pressure specific heat capacities of three hydrating cement pastes with water-to-cement ratios (w/c) of 0.3, 0.4 and 0.5 and their comparison with macroscopic experimental measurements of De Schutter [90] and Bentz

[39]. According to eq. 4.21, simulation results are necessarily linear in terms of η and are in good agreements with the experimental measurements. This agreement is rather qualitative for the case of w/c=0.5. We note that the materials employed in the study of De Schutter (w/c=0.5) contain significant sources of blast furnace slag and thus the stoichiometries of reactants and hydration products are subjected to be different from eqs. 4.1-4.2. This might specially explain the slight difference between the experimental and simulation values at $\eta > 0.5$. Both simulation and experiments indicate that the heat capacity of the paste decreases with increasing hydration degree. This can be understood by proper consideration of the role of water during the course of hydration. The specific heat capacity of bulk water is considerably high (4.18 J/g/K) which makes the heat capacity of the mixture of water and clinker relatively high. Throughout the hydration process, the water is consumed to precipitate CSH, which has a significantly lower specific heat capacity (see Fig. 4-4). Also, part of this water is trapped within the CSH's molecular structure which exhibits features of supercooled water with a considerably lower specific heat capacity compared to that of room temperature bulk water (see the inset of Fig. 4-4). Following the same line of thought, we would infer that the heat capacity of cement paste would increase with increasing w/c ratio. This is confirmed by both simulation and experiment and can be rationalized in the sense that the portion of water that does not contribute in the hydration is trapped in the inter-particle and capillary pores. The large specific heat capacity of unreacted water in such pores increases the heat capacity of cement paste.

4.7.2 General Micro-thermo-mechanics Formulation

Since heat flux is an intensive thermodynamic quantity, simple composition-based mixture rules would not be suitable for upscaling heat transport properties in composite materials. Fortunately, homogenization techniques present a consistent mechano-physical framework to upscale such properties in multiphase and multiscale materials. These approaches associate the homogenized thermal conductivity, $\mathbf{K}_{hom}(\mathbf{X})$, at Level $l+1$ to that of constituent phases, $\mathbf{K}_l(\mathbf{x})$, at level l, where $\mathbf{X}$ and $\mathbf{x}$ are

position vectors at Level $l+1$ and l, respectively. In micro-thermo-poromechanics, a continuum point at Level $l+1$ is characterized by its temperature, T, temperature gradient, $\mathbf{\Theta} = \bigtriangledown T$, and the associated heat flux, $\mathbf{Q}$. Similarly, a continuum point at Level l is defined by its temperature, τ, temperature gradient, $\boldsymbol{\theta} = \bigtriangledown \tau$, and its relevant heat flux, $\mathbf{q}$. The temperature gradient and heat flux at Level $l+1$ are essentially the average of those observables at Level l. This requires the compatibility of heat flux, $\mathbf{Q} = \langle \mathbf{q} \rangle$, and temperature gradient, $\mathbf{\Theta} = \langle \boldsymbol{\theta} \rangle$, between the two scales. The $\langle . \rangle$ represents the spatial averaging operator defined as $\langle o \rangle = \frac{1}{V} \int_{\Omega} o(\mathbf{x}) dv$, where V is the volume of representative elementary volume (REV), and o is an arbitrary observable at Level l. According to homogenization theory, temperature gradients at levels l and $l+1$ can be assumed to be linearly related via a second order temperature gradient localization tensor or simply localization tensor, $\mathbf{A}(\mathbf{x})$:

$$\boldsymbol{\theta}(\mathbf{x}) = \mathbf{A}(\mathbf{x}) : \mathbf{\Theta}(\mathbf{X}) \tag{4.22}$$

The compatibility condition requires $\langle \mathbf{A} \rangle = \mathbf{I}$. The constitutive relation, here Fourier's law, can be exploited to relate $\mathbf{q}$ and $\mathbf{\Theta}$, $\mathbf{q}(x) = \mathbf{K}_l(\mathbf{x}) : \mathbf{A}(\mathbf{x}) : \mathbf{\Theta}(\mathbf{X})$. By spatially averaging the former identity, the homogenized thermal conductivity is expressed in terms of the localization operator and thermal conductivity of constituent phases at Level l:

$$\mathbf{K}_{hom}(\mathbf{X}) = \langle \mathbf{K}_l(\mathbf{x}) : \mathbf{A}(\mathbf{x}) \rangle \tag{4.23}$$

The mathematical form of the localization tensor is required for the linear upscaling scheme introduced in eq. 4.23. The seminal work of Eshelby [111] revealed that the strain localization tensor is constant for an ellipsoidal inclusion embedded in an infinite matrix. Analogous to the elasticity problem, Hatta and Taya [143] found the localization tensor, $\mathbf{A}_I$, for heat conduction through an isotropic medium with an elliptical anisotropic inclusion:

$$\mathbf{A}_I = \left(\mathbf{I} + \mathbf{S} : \left(\mathbf{K}_M^{-1} \mathbf{R} \mathbf{K}_I \mathbf{R}^T - \mathbf{I} \right) \right)^{-1} \tag{4.24}$$

where $\mathbf{S}$ is the Eshelby-like tensor, $\mathbf{I}$ is the unit tensor, $\mathbf{K}_M$ and $\mathbf{K}_I$ are the thermal conductivity of the matrix and inclusion, respectively. The principal thermal conductivity directions in the inclusion might not necessarily coincide with the ellipsoid's semi-axis. The rotation matrix $\mathbf{R}$ is used to transform the local to global coordinate systems. The analytical form of the second order Eshelby-like tensor $\mathbf{S}$ is found via the Green's function approach:

$$S_{ii} = \frac{a_1 a_2 a_3}{2} \int_0^\infty \frac{1}{(a_i^2 + s)\sqrt{\prod_{k=1}^{3}(a_k^2 + s)}} ds \tag{4.25}$$

where a_1, a_2 and a_3 are semi-axes of ellipsoidal inclusion. It is remarkable to note that $\mathbf{S}$ is strictly diagonal and independent of the position vector $\mathbf{x}$. Therefore, the localization tensor is simply a function of geometrical features at Level l. The above solution for a two-phase matrix-inclusion system can be extended to incorporate n_p phases. Furthermore, to include the effect of morphology, we consider a volume V_0 with thermal conductivity of $\mathbf{K}_0$ surrounding the p^{th} inclusion phase. In this case, the localization tensor can be written as:

$$\mathbf{A}_p = \left(\mathbf{I} + \mathbf{S}_p : \left(\mathbf{K}_0^{-1}\mathbf{R}\mathbf{K}_p\mathbf{R}^T - \mathbf{I}\right)\right)^{-1} \langle \left(\mathbf{I} + \mathbf{S}_m : \left(\mathbf{K}_0^{-1}\mathbf{R}\mathbf{K}_m\mathbf{R}^T - \mathbf{I}\right)\right)^{-1} \rangle^{-1} \tag{4.26}$$

where $\mathbf{K}_r = (\mathbf{K}_0, \mathbf{K}_1, \ldots, \mathbf{K}_{n_p})$ in $(V_0, V_1, \ldots, V_{n_p})$, respectively. The choice of $\mathbf{K}_0$ leads to two distinctly different morphological descriptions of microstructure at Level l. In the case of matrix-inclusion morphology, the natural choice of the reference thermal conductivity is that of the matrix. This approach is known as the Mori-Tanaka (MT) scheme:

$$\mathbf{A}_p^{MT} = \left(\mathbf{I} + \mathbf{S}_p : \left(\mathbf{K}_M^{-1}\mathbf{R}\mathbf{K}_p\mathbf{R}^T - \mathbf{I}\right)\right)^{-1} \langle \left(\mathbf{I} + \mathbf{S}_m : \left(\mathbf{K}_M^{-1}\mathbf{R}\mathbf{K}_m\mathbf{R}^T - \mathbf{I}\right)\right)^{-1} \rangle^{-1} \tag{4.27}$$

In the absence of any dominating matrix-inclusion morphology, all phases may play the role of surrounding homogeneous material. Therefore, an appropriate choice for the reference medium is the homogenized material itself, $\mathbf{K}_{hom} = \mathbf{K}_0$. This approach

is known as the self-consistent (SC) scheme:

$$\mathbf{A}_p^{SC} = \left(\mathbf{I} + \mathbf{S}_p : \left(\mathbf{K}_{hom}^{-1}\mathbf{R}\mathbf{K}_p\mathbf{R}^T - \mathbf{I}\right)\right)^{-1} \langle \left(\mathbf{I} + \mathbf{S}_m : \left(\mathbf{K}_{hom}^{-1}\mathbf{R}\mathbf{K}_m\mathbf{R}^T - \mathbf{I}\right)\right)^{-1} \rangle^{-1} \tag{4.28}$$

where $\mathbf{S}_m$ is the Eshelby tensor of the m^{th} inclusion phase. As demonstrated in eqs. 4.27 and 4.28, the localization tensors and the resultant homogenized thermal conductivity values depend on the orientation of the inclusions within the matrix through $\mathbf{R}$. In fact, an arbitrary observable $\boldsymbol{\chi}$ can be viewed as a random variable that follows a joint probability distribution function (PDF) $P(\boldsymbol{\chi}(\varphi_1, \varphi_2, \varphi_3))$, where φ_1, φ_2 and φ_3 denote the Euler angles [143]. Upon the absence of a preferential orientation for inclusions, the joint PDF becomes uniform with the value of $\frac{1}{8\pi^2}$. The orientation average of a random quantity $\boldsymbol{\chi}$ is denoted by $\langle\langle . \rangle\rangle$ and follows:

$$\langle\langle \boldsymbol{\chi} \rangle\rangle = \frac{1}{8\pi^2} \int_0^{2\pi} \int_0^{\pi} \int_0^{2\pi} \boldsymbol{\chi}(\varphi_1, \varphi_2, \varphi_3) sin\,(\varphi_2)\, d\varphi_1 d\varphi_2 d\varphi_3 = \frac{1}{3} tr\,(\boldsymbol{\chi})\, \mathbf{I} \tag{4.29}$$

where $tr(.)$ denotes the trace operator and the so-called "x_2" convention is utilized in the definition of the rotation matrix (see Mura [243] for details). The orientation averaging can be applied to rotate both the principal thermal conductivity axes and the ellipsoidal inclusion itself. From the homogenized conductivity view point, the system of randomly oriented anisotropic ellipsoidal inclusions embedded in an isotropic matrix, K_v^M, is indistinguishable from the system of spherical inclusions with a volume thermal conductivity K_v. Therefore, after proper orientational averaging, the homogenized Mori-Tanaka thermal conductivity K_v^{MT} would reduce to:

$$K_v^{MT} = \frac{f_M K_v^M + \sum_{s=1}^{n_p} f_s K_v^s A_s^{sph}}{f_v^M + \sum_{s=1}^{n_p} f_s A_s^{sph}} \tag{4.30}$$

where f_s is the volume fraction and $A_s^{sph} = 3K_v^M/(2K_v^M + K_v^s)$ is the spherical localization factor of the s^{th} phase. Similarly, the homogenized self-consistent thermal

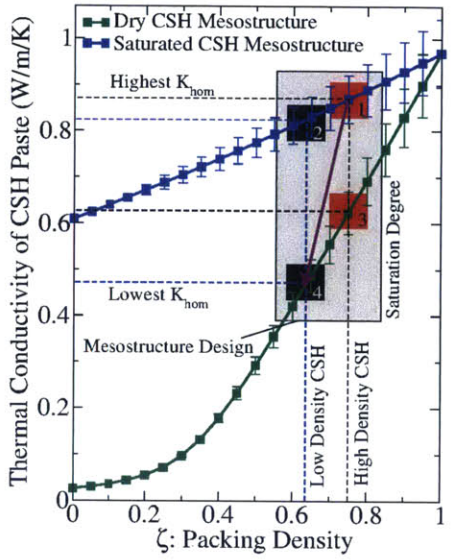

Figure 4-8: The effects of CSH mesostructure (packing density) and saturation degree of mesopores on the thermal conductivity of CSH paste calculated via probabilistic micro-thermo-poromechanics. The error bars are calculated via the Monte Carlo uncertainty propagation method.

conductivity K_v^{SC} would reduce to:

$$K_v^{SC} = \frac{\sum_{s=1}^{n_p} f_s K_v^s B_s^{sph}}{\sum_{s=1}^{n_p} f_s B_s^{sph}} \tag{4.31}$$

where $B_s^{sph} = 3K_v^{SC}/(2K_v^{SC} + K_v^s)$ is the spherical localization factor of the s^{th} phase. It is instructive to note that the mathematical formula for K_v^{SC} has an implicit nature requiring an iterative procedure to obtain the solutions.

4.7.3 Upscaling Heat Conductivity from Nano to Macro Scale

As discussed in section **??** and schematically shown in Fig. 4-1.b, the CSH paste at the mesoscale is comprised of randomly oriented polydisperse CSH particles with an average size of roughly 5 nm. The characteristic length of these nanoparticles are much smaller than the size of the microscale REV and much larger than the mean free path of phonons in the order of Si-O and Ca-O bonds. This means that the conditions of scale separability are fully met between the nano, meso and microscales.

Therefore, the micro-thermo-poromechanics formulation introduced above is suitable for homogenization of thermal conductivity of CSH at the mesoscale. Based on the results presented in Fig. 4-6, the principal thermal conductivity values of CSH are normally distributed and uncorrelated. Hence, we can treat eq. 4.31 in a probabilistic fashion by randomly choosing the principal thermal conductivity values of the CSH particles from normal distributions. Here, we utilized Monte Carlo uncertainty propagation scheme to measure both the expectations and standard deviations of the homogenized thermal conductivity values at the microscale. Fig. 4-8 presents the homogenized thermal conductivity of CSH paste at microscale for two limiting cases of fully saturated and dry mesopores as a function of the CSH's packing density, $\zeta = 1 - \varphi$, where φ is the mesoporosity. Since the bulk thermal conductivity values of air and water are lower than that of CSH, the thermal conductivity of CSH paste decreases with increasing porosity. Note that in nanoporous media in contact with the outside air, the composition of confined air will not be that in the outside. In the realm of engineering CSH's heat conduction properties, there are two design parameters: the mesostructure design and the saturation degree of the mesopores. Constantinides and Ulm [82] showed CSH nanoparticles coalesce at 64% and 76% packing fractions attributing them to the low density (LD) and high density (HD) CSHs. Given the two limiting saturation levels and the two packing fractions, the upper and lower bounds of the CSH paste's thermal conductivity can be identified. These bounds are displayed by numbered red and black rectangles in Fig. 4-8. While the saturated HD (rectangle 1) and dry LD (rectangle 4) CSH have the highest and lowest thermal conductivity values, respectively.

The cement paste at microscale can be assumed to be a multiphase composite consisted of CSH paste embracing anhydrated clinker phases, portland and the fluid inside capillary pores (see Fig. 4-1.c). The micro-structure of cement paste at this scale is akin to that of matrix-inclusion morphology. Therefore, the Mori-Tanaka scheme would be the method of choice for upscaling the thermal conductivity of individual phases at the microscale (Level 2) to that of macroscale (Level 3). Over the course of cement hydration (see eqs. 4.1,4.2,4.20), the volume fraction of individual

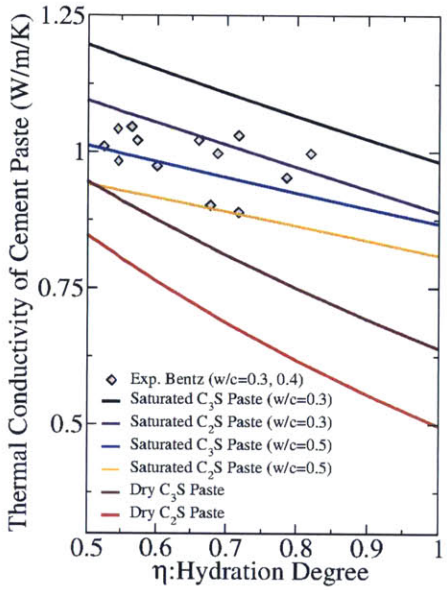

Figure 4-9: The effect of w/c ratio, saturation degree and the type of clinker phase on the thermal conductivity of hydrating cement paste. The simulation results calculated via micro-thermo-poromechanics are compared against experimental measurements of Bentz [39] for hydrating cement pastes.

phases (f_i) (reactant or product) are known from the reaction degree η and their mass density ρ_i (see table. 4.3). Therefore, the homogenized thermal conductivity of cement paste can be implicitly expressed in terms hydration degree and the thermal conductivity values of individual phases (table 4.4) through eq. 4.30. Fig. 4-9 presents the thermal conductivity as a function of hydration degree for the resulting cement pastes produced from the hydration of C_2S and C_3S pure clinkers at different w/c ratios (w/c=0.3 and 0.5) and two limiting cases of fully saturated and dry meso and micropores compared with experimental measurements of Bentz [39] for cement of varying w/c ratio and curing conditions. Generally, we observe that the thermal conductivity of cement paste decreases with increasing hydration degree. This is consistent with experimental observations of Mounanga et al. [239]. This is due to the fact that the thermal conductivity of clinker phases are higher than those of the hydration products (CSH and CH). Also, the thermal conductivity of cement paste decreases with w/c ratio, in agreement with recent experiments of Maruyama and

Igarashi [214]. The reason is that while the volume fraction of CH decreases with increasing w/c ratio and the volume fraction of capillary water increases with it; and since the thermal conductivity of bulk water is less than that of portlandite, then the thermal conductivity of the overall cement paste decreases with increasing w/c ratio. The thermal conductivity of dry cement paste is lower than that of saturated paste because the thermal conductivity of the relatively negligible thermal conductivity of air.

4.8 Conclusions

The thermal properties of cement paste were studied via a multiscale approach, starting from properties of individual constituent phases at the nanoscale and upscaling them all the way to the macroscale via the mixtures rules and micro-thermo-poromechanics theory. At the nanoscale, the phonons density of states of CSH, CH, β-C_2S and C_3S and their components are studied in detail via the diagonalization of the dynamical matrix and the velocity autocorrelation functions. Analogous to glassy materials, CSH was found to exhibit an excess of vibrational states characterized by the Boson peak at the lower frequencies region. The position of the Boson peak is found to be shifted to higher frequency values with increasing CSH density. The constant volume and pressure specific heat capacities are calculated for each individual phase. We found that the specific heat capacity of dry and saturated CSH increases linearly with Ca/Si ratio. The apparent heat capacity of nanoconfined water in the interlayer spacing of CSH is found to be significantly lower than that of bulk water at room temperature due to strong interaction with the calcium-silicate substrate. The full thermal conductivity tensor of each individual phase was measured using the autocorrelation of heat flux and the Green-Kubo relation. The Fourier transformation of the heat flux autocorreltion function revealed long-lasting sharp frequencies associated with optical phonons that do not contribute to the thermal conduction of these complex calcium-silicate systems. In addition, the mean free path of phonons in different phases are estimated to be in the order of Si-O and Ca-O bond lengths.

The small mean free path of phonons in the cement paste constituent phases meets the condition of scale separability required for upscaling of the thermal properties to larger scales. We employed simple mixture laws to demonstrate the variation of the heat capacity of cement paste at the macroscale as function of the hydration degree for different water to cement ratios. With no fitting parameter, the macroscopic heat capacity values calculated via our multiscale approach are in good quantitative agreement with experiments. The homogenized heat conductivity of CSH paste at the microscale was calculated via the self-consistent theory and is shown to be only affected by the packing density and the saturation level of CSH's mesopores. This homogenized thermal conductivity was subsequently exploited to estimate the thermal conductivity of hydrating cement paste at Level 3. Our analysis indicates that the heat conductivity of cement paste decreases with increasing w/c ratio and hydration degree in agreement with experiments. Overall, this work paves the way for thermal engineering of multiscale porous materials by providing a seamless connection between their properties at the nano and macroscales.

Chapter 5

Heat Loss Analysis: From a Single Buildings to the Entire City

The most immediate application of heat capacity and heat conductivity of materials is in heat loss in buildings. Depending on the climate, building heat losses constitute one of the significant portions of building's total energy consumption. The energy consumption patterns of residential, commercial, and transportation sectors are driven by human activities, public perception and decisions and physical constraints. The common urban energy models utilize micro-simulations to predict future energy usage by emulating behavior of urban dwellers (agents) and converting their decisions to respective energy demands [38, 69]. Thus, a building's energy usage becomes a function of the choice of technology (e.g. insulation conditions, or efficiency of the heating system), its utilization (e.g. choice of the internal temperature set point), and additional extrinsic variables such as weather and neighborhood patterns, which affect a building's energy performance [281]. Limiting heat losses from building envelope, i.e. the act of retrofit, is primarily concerned with the choice of technology, namely the insulation to enhance the thermal resistance of its envelope (walls, windows, doors, roof, and floors) and to reduce losses due to heat conduction, Q_{cond}, and infiltration, Q_{inf}. Here, we aim to identify, based on building energy simulation and data analysis, the most influential set of physical parameters that affects heat loss of buildings and the associated shortest path to reduce heat losses at the city level [41, 179].

This chapter is divided into seven sections. Section 5.1 explains why we should care about heat losses in buildings. Section 5.2 discusses the factors that affect heat losses in buildings and their associated uncertainties. Section 5.3 aims at finding the most influential parameters by performing global sensitivity analysis on the space of all parameters that affect heat losses in a single building. Using these influential parameters, section 5.4 constructs two types of metamodels, based on dimensional homogeneity and machine learning techniques, that estimate the heat losses in a building. We use different machine learning techniques in subsection 5.4.1 to emulate the response of a famous building energy loss modeling software, Energyplus. In subsection 5.4.2, we construct a mechanistic model based on conservation of energy and dimensional analysis to predict energy losses at the building level. Subsequently, we use our knowledge on the heat losses at the single building level to better understand energy losses at the city scale. To this end, we first analyze energy consumption of more than 6,200 buildings in Cambridge, MA for three consecutive years during 2007-2009 in section 5.5. Afterwards, we combine the results of data analysis with our surrogate models presented in section 5.4 to find the shortest path to minimize heat losses at the city level. Section 5.7 summarized our findings and concludes the paper.

5.1 Why Does Heat Loss Matter?

In 2012, the aggregate home energy expenditure of the 130 million households in the United States [2] topped $250 billion [3], nearly 37% of which was wasted due to heat losses. This translates to an average of $1,000 per household spent on space heating and cooling. With over 85% of population concentrated in urban areas, governments have embraced important initiatives to reduce this waste and the associated carbon footprints of cities by providing stimulus funds to adopt energy efficiency programs. These programs, however, operate with limited financial resources; and can renovate only a selected number of buildings per year. These limitations call for fast and accurate methods to inform smart, citywide retrofit plans that pinpoint to buildings

with the greatest energy saving potential in order to maximize returns on investment.

One challenge is to develop reliable methodologies and toolsets to estimate energy losses in buildings. There is a vast body of literature on this subject since the 1970's. The readers are referred to Swan, and Ugursal [323], Kavgic et al. [169] and references therein for a comprehensive review of various modeling techniques. Another important challenge is concerned with developing a quantitative framework that supports informed decisions on energy-saving policies relevant to retrofitting at the urban scale. Such framework requires a scalable and sufficiently representative description of urban energy consumption that avoids redundant details while carrying enough physical information at the building level to inform retrofit options. This paper is motivated by the latter challenge. Current approaches to evaluate retrofits use rating or audit tools to help energy-saving investments. Bardhan et al. [27] present an updated review of current practices and methods. The level of complexity and required information of these methods vary significantly. The general results to inform policy are presented in terms of scores, characterizing the relative energy efficiency of a house in the region, or the recommendation of actions and the estimate of their consequent energy savings.

Current tools are often best suited for building-wise assessment. Upscaling the results to the community level relies on considering a "typical" or "average" house as a building block. For instance, in the Home Energy Saver (HES) designed by the Lawrence Berkeley National Laboratory, cities are analyzed at the zip code level within a sizeable resident population from each urban area. These models are used to estimate annual energy consumption of standardized house in cities, to provide upgrade recommendations and to perform cost-benefit analysis of each specific retrofit. Although these tools show promise in helping energy-saving investment by filling the informational gap to a great extent, there are limitations in their accuracy and scalability when used to inform retrofitting energy efficiency at city scale. These limitations deter their application as an effective and robust decision support methodology at urban scale. We herein propose a novel ranking algorithm that allows upscaling energy consumption from relevant information on building level to the size of na-

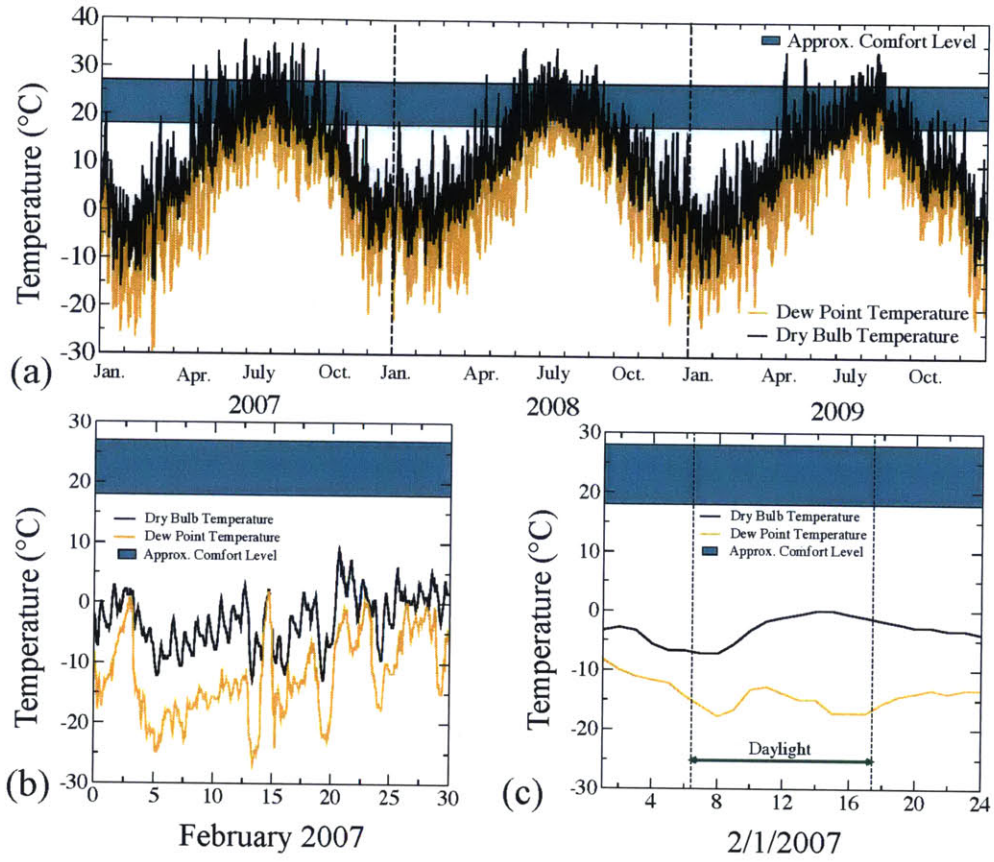

Figure 5-1: Dry and wet bulb temperatures at a) 2007 to 2009 period, b) in February 2007 and c) hourly data for 2/1/2007 measured by the National Oceanic and Atmospheric Administration (NOAA)'s weather station at Boston Logan International Airport. The time period for the weather data corresponds to the same period as energy consumption readings.

tional and global sustainability goals. We perform analysis of variance combining data from energy bills, building footprint and physical simulations to avoid statistical bias introduced by "typical houses". We use this information to construct a simple yet efficient physics-based description of heating energy demand based on a model reduction scheme that encloses the most relevant parameters for the observed consumption. The model provides means to identify buildings with the greatest potential for improvement, and quantifies the aggregated energy savings.

5.2 Factors that Affect Heat Loss at the Building Level

Several internal and external factors affect heat losses in buildings. The most important external factor is climate or in other words the thermal loading. In fact, the monthly energy consumption or heat loss in buildings is highly correlated to the average outdoor temperature. In this work, we only concentrate on the climate of Cambridge, MA, as we have partial access to energy consumption of roughly 6,200 buildings during 2007-2009. Therefore, we construct hourly energy consumption data from the weather records provided by National Oceanic and Atmospheric Administration (NOAA) [?] for Boston Logan airport for the period that corresponds to the available energy consumption readings. Fig. 5-1 provides an example of T_{db} and T_{dp} at yearly, monthly and daily scales. We employ this weather record in all heat loss calculations using Energyplus software [84]. Appendix F provides the theoretical background behind building heat loss modeling in Energyplus software and the readers are encouraged to review that section before proceeding to read this chapter. The other set of external factors that affect heat losses of existing buildings are their sizes, namely their equivalent length, width and height. For a given building, dimensions have rather a deterministic nature. However, at large scales, we can think of building dimensions in a probabilistic fashion. In fact, the size distributions in the city of Cambridge MA, for which we have the monthly energy consumption, is highly correlated (see Fig. 5-2). By analyzing the GIS data (Fig. 5-2.a), It turns out that both building length (L_x) and width (L_y) follow bi-lognormal distributions (Fig. 5-2.b). However as suggested by Fig. 5-2.c, the joint probability distribution function $P(L_x, L_y)$ is highly localized meaning that L_x and L_y are highly correlated. As shown by Fig. 5-2.d, building height (L_z) is also bi-lognormally distributed and it is highly correlated to L_x and L_y. Therefore, the building size distributions are both non-uniform and correlated.

Another set of external parameters that affect heat losses is the urban texture that a building is embedded in. We rationalize this urban texture in terms of inter-building

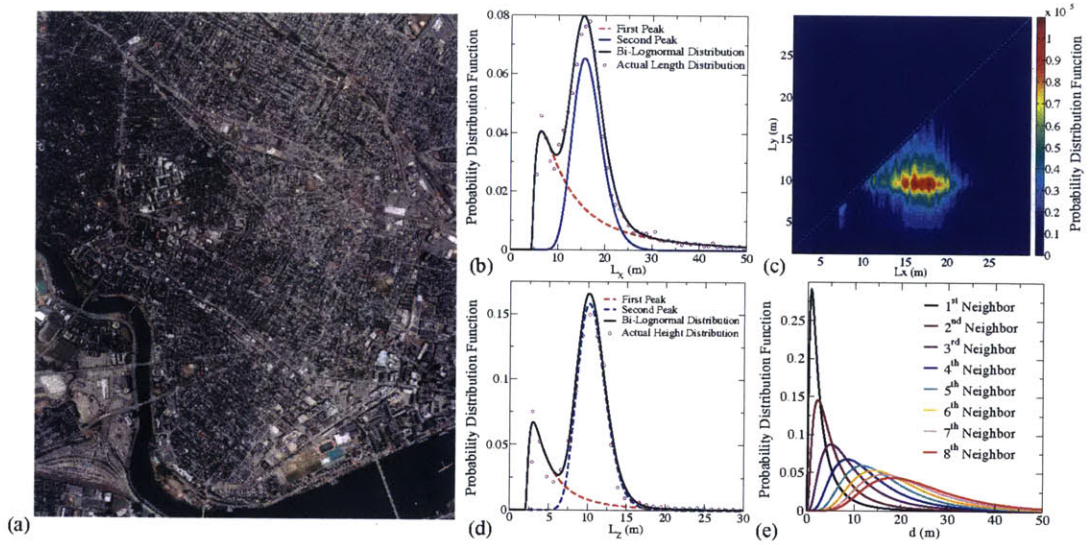

Figure 5-2: a) The Arial map of Cambridge, MA courtesy of Google maps and simplifying the complexity of urban texture via probability distribution functions derived from the spatial analysis of building parcel information in GIS data. b) building length (L_x) distribution c) joint probability distribution function between building length and width (L_y). d) The distribution of building heights. e) the distribution of the closest distance to first and up to eighth neighbor.

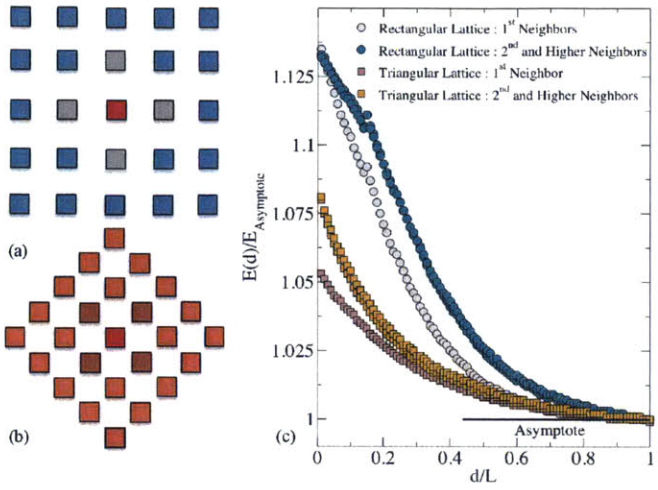

Figure 5-3: Local sensitivity analysis on the effect of neighborhood on the heating energy consumption in Cambridge, MA. The top view of the tow considered patterns a) rectangular lattice b) triangular lattice with d being the closest distance between the two buildings. The central building is shown in red for which the energy consumption is calculated. c) The energy consumption of the building divided by its energy consumption in the absence of neighbors.

distances that affect shadowing between buildings. Lighting and shadowing in the other aspects that affect heat gains in buildings. Therefore, the number of neighbors, their arrangement in the neighborhood and their relative distance influences energy consumption. Then, we would wonder up to how many neighbors should we consider around a building to properly estimate the effect of neighbors on heat gains. In other words, if the closest building is labeled as 1^{st} and second closest as 2^{nd} and so on and so forth up to the n^{th} neighbor, what is the value of n for which heat losses approach their asymptotic values. To address this problem, we utilize a one-at-a-time (OAT) sensitivity analysis to understand the effective system size for modeling heat loss of a single building embedded in an urban environment (readers are referred to Appendix C for a brief introduction to sensitivity analysis). To this end, we consider two building arrangements namely rectangular and triangular lattices as depicted in Fig. 5-3.a and 5-3.b. The inter-building distance d denotes the closest distance in between a building and its first neighbors, which is kept constant over the entire lattice. The buildings are considered to be three stories tall 10m$\times$10m$\times$10m with single-pane window in all

walls covering the 20% of the wall. The thermal resistance of the building envelope is set to $4m^2K/W$ and infiltration rate (air change per hour) is assumed to be $1/h$. In the OAT analysis, all the above design variables are kept constant except d which is systematically increased from $0.2m$ to $10m$. Fig. 5-3.c presents the effect of inter-building distance and the number of neighbors on the energy consumption of the building. First, we observe that energy consumption reduces with the increase in the inter-building distance. This is intuitively correct because in cold climate the unblocked direct and diffuse solar radiations increase the temperature of the outer surface of the wall and solar radiation reached inside the zone through the windows. Secondly, once the buildings get enough distant from one another, the heat losses will not be affected by the neighborhood pattern and reach an asymptotic value of $E_{asymptote}$. Therefore, beyond a certain distance ($d/L > 1$) buildings do not interact. Finally, OAT analysis shows that the number of shell of neighbors considered in the model affects the heat losses of a building. The heat loss with considering only 4 neighbors (first shell of neighbors) is different from considering 8 neighbors (first and second shell of neighbors). However, the heat loss would not change in either of the patterns by adding extra neighbors (third, forth and fifth shell of neighbors).

The above sensitivity analysis indicates that considering the first eight neighbors would be enough for modeling energy consumption in the urban environment. This significantly reduces the computational cost associated with modeling building heat transfer at the city-scale. In fact, similar to the building dimensions, we can consider inter-building distances for the eight closest neighbors in a probabilistic fashion. As shown in Fig. 5-2.e, the closest distance between a building and its i^{th} neighbors d_i is distributed following a lognormal distribution. Since d_i should be less or equal to d_{i+1}, these lognormal distributions are not independent. We can heuristically resolved this dependence by noting that $COV(d_i, d_j)$ is negligible and hence sampling independently from $d_{ij} = d_j - d_i$ instead of d_i and d_j. As it was mentioned before, d_1

follows lognormal distribution:

$$y = P(x|\mu,\sigma) = \frac{1}{x\sigma\sqrt{2\pi}} \times exp\left(\frac{-(ln\,(x)-\mu)^2}{2\sigma^2}\right) \tag{5.1}$$

where μ and σ are 3.8 and 25.0m, respectively for the d_1. The distribution of difference between neighbor distances d_{ij} follows an exponential distribution:

$$y = P(x|\mu) = \frac{1}{\mu} \times exp\left(\frac{-x}{\mu}\right) \tag{5.2}$$

where μ is 2.8, 3.9, 3.4, 2.9, 2.6, 2.4 and 2.3, respectively correspond to d_{12}, d_{23}, d_{34}, d_{45}, d_{56}, d_{78} and d_{78}.

In addition to the climate, building size and neighborhood texture, heat losses are affected by several other parameters ranging from building thermal properties to human behavior. The comprehensive list of all parameters employed in our study is provided in Table. 5.2. This list includes and is not limited to the thermal properties of wall, roof, separator, partition, floor, slab, characteristics of the windows, HVAC, temperature set points, air infiltration rate, and number of residents. For each of the envelope sections, we consider roughness, thickness, thermal conductivity, specific heat capacity, density, thermal adsorptance, solar adsorptance, and light adsorption in the visible range. These parameters can be deterministic or probabilistic depending on our knowledge about them. For instance, if we measure the thermal properties of a wall, they are known to us and hence we can treat them as deterministic variables. Otherwise, they are unknown and can be assumed to have probabilistic nature. In practice, after proper sampling (measuring the thermal properties of enough number of walls), we can identify the probability distributions describing those properties. In the absence of such samples usually due to practical reasons, we can use engineering judgment to identify minimum and maximum bounds for those parameters and assume that such parameters are distributed uniformly in between those bounds. Such assumptions directly influence the outcome of any probabilistic study and, therefore, they should be treated with special attention. In the following section, we use the list

Num	Name	PDF	Min	Max	Num	Name	PDF	Min	Max
1	Wall Roughness	Uniform	1	5	38	Partition Roughness	Uniform	1	5
2	Wall Thickness (m)	Uniform	0.1	0.4	39	Partition Thickness (m)	Uniform	0.1	0.2
3	Wall Conductivity (W/mK)	Uniform	0.05	2	40	Partition Conductivity (W/mK)	Uniform	0.05	2
4	Wall Specific Heat (J/KgK)	Uniform	500	3000	41	Partition Specific Heat (J/KgK)	Uniform	500	3000
5	Wall Desnsity (Kg/m^3)	Uniform	1500	3000	42	Partition Desnsity (Kg/m^3)	Uniform	1500	3000
6	Wall Thermal Absorptance	Uniform	0.55	0.85	43	Partition Thermal Absorptance	Uniform	0.55	0.85
7	Wall Solar Absorptance	Uniform	0.55	0.85	44	Partition Solar Absorptance	Uniform	0.55	0.85
8	Wall Visible Absorptance	Uniform	0.55	0.85	45	Partition Visible Absorptance	Uniform	0.55	0.85
9	roof Roughness	Uniform	1	5	46	floor Roughness	Uniform	1	5
10	roof Thickness (m)	Uniform	0.1	0.4	47	floor Thickness (m)	Uniform	0.3	0.5
11	roof Conductivity (W/mK)	Uniform	0.05	2	48	floor Conductivity (W/mK)	Uniform	0.2	2
12	roof Specific Heat (J/KgK)	Uniform	500	3000	49	floor Specific Heat (J/KgK)	Uniform	500	3000
13	roof Desnsity	Uniform	1500	3000	50	floor Desnsity (Kg/m^3)	Uniform	1500	3000
14	roof Thermal Absorptance	Uniform	0.55	0.95	51	floor Thermal Absorptance	Uniform	0.55	0.85
15	roof Solar Absorptance	Uniform	0.55	0.95	52	floor Solar Absorptance	Uniform	0.55	0.85
16	roof Visible Absorptance	Uniform	0.55	0.95	53	floor Visible Absorptance	Uniform	0.55	0.85
17	Seperator Roughness	Uniform	1	5	54	floo Roughness	Uniform	1	5
18	Seperator Thickness (m)	Uniform	0.1	0.2	55	Shared Slab Thickness (m)	Uniform	0.1	0.2
19	Seperator Conductivity (W/mK)	Uniform	0.05	2	56	Shared Slab Conductivity (W/mK)	Uniform	0.05	2
20	Seperator Specific Heat (J/KgK)	Uniform	500	3000	57	Shared Slab Specific Heat (J/KgK)	Uniform	500	3000
21	Seperator Desnsity (Kg/m^3)	Uniform	1500	3000	58	Shared Slab Desnsity (Kg/m^3)	Uniform	1500	3000
22	Seperator Thermal Absorptance	Uniform	0.55	0.85	59	Shared Slab Thermal Absorptance	Uniform	0.55	0.85
23	Seperator Solar Absorptance	Uniform	0.55	0.85	60	Shared Slab Solar Absorptance	Uniform	0.55	0.85
24	Seperator Visible Absorptance	Uniform	0.55	0.85	61	Shared Slab Visible Absorptance	Uniform	0.55	0.85
25	Window Type	Dummy	1	2	62	Angle w.r.t North (°)	Uniform	0	180
26	Window Percentage	Uniform	0.2	0.4	63	Distance from 1^{st} Neighbor (m)	Lognormal	-	-
27	Internal Mass	Uniform	0.5	2	64	Distance from 2^{nd} Neighbor (m)	Exponential	-	-
28	Number of Residents	Uniform	1	5	65	Distance from 3^{rd} Neighbor (m)	Exponential	-	-
29	Lighting Power (W/m^2)	Uniform	10	40	66	Distance from 4^{th} Neighbor (m)	Exponential	-	-
30	Air Infiltration Rate (ACH)	Uniform	0.3	3.5	67	Distance from 5^{th} Neighbor (m)	Exponential	-	-
31	Set Point Temperature (°C)	Uniform	12	25	68	Distance from 6^{th} Neighbor (m)	Exponential	-	-
32	Air Infiltration per Resident (ACH)	Uniform	0.005	0.015	69	Distance from 7^{th} Neighbor (m)	Exponential	-	-
33	Ground Temperature (°C)	Uniform	15	25	70	Distance from 8^{th} Neighbor (m)	Exponential	-	-
34	Number of Attached Buildings	Dummy	0	2	71	Max Supply Heating Temperature (°C)	Uniform	30	50
35	Building Length (m)	Correlated Data	-	-	72	Min Supply Cooling Temperature (°C)	Uniform	10	20
36	Building Width (m)	Correlated Data	-	-	73	Max Supply Heating Humidity	Uniform	0.012	0.018
37	Building Height (m)	Correlated Data	-	-	74	Min Supply Cooling Humidity	Uniform	0.005	0.015

Table 5.1: The list of the random variables including the type of the distribution and their min and max if they are uniformly distributed.

of variables in Table. 5.2 to construct a probabilistic heat loss model. This probabilistic model will be crucial in determining the number of influential parameters that affect heat losses the most among all these 74 variables.

5.3 Heat Loss Sensitivity Analysis at the Building Level

Now that we have identified parameters that affect heat loss, we can proceed and construct a probabilistic model that relates the uncertainty in parameter domain to uncertainty in the heat loss domain. To this end, we employ the Energyplus Software, a fairly standard deterministic heat loss calculation package. The "*deterministic*" term means that if the software is executed several times with the same input parameters,

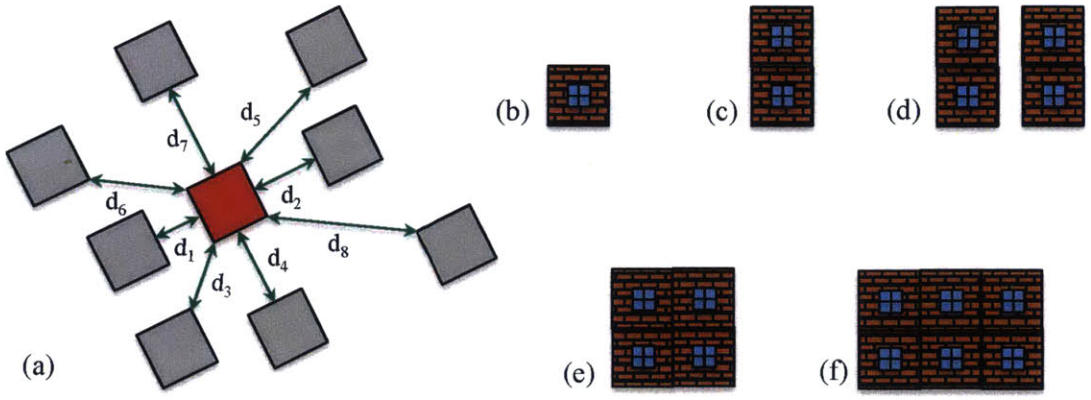

Figure 5-4: The geometrical configuration of the probabilistic model. a) The top view of the models indicating the central building in red and the neighbors in grey. The inter-building distances are shown by d_i with i ranging from 1 (closest) to 8 (furthest). The pre-processor can generate different configurations ranging from b) single-story detached house, c) multiple-story buildings, d) unattached buildings e) attached building in one side c) attached buildings in two sides of the building.

it yields exactly the same results. Therefore, our probabilistic model is constructed by feeding uncertain input parameters into this deterministic software and collecting the output, so they have a probabilistic nature. Here, we have defined the type of the distribution in the column next to the variable in Table 5.1. If the variable is assumed to follow a uniform distribution, then we have included the minimum and maximum of that variable in the two columns in front of it. In addition to characteristics of building properties, the human behavior, HVAC specifications, angle of the building with north direction, and ground temperature are also assumed to be uniformly distributed. As mentioned in Section 5.2, the neighborhood texture has a probabilistic nature and such a probabilistic nature can affect heat loss. The distribution of inter-building distance follows lognormal and exponential distributions as discussed in Section 5.2. We also demonstrated that building heat gains are only affected by its first eight immediate neighbors. Therefore, this suggests that we can build a probabilistic model for a single building embedded in an urban environment with only considering eight neighbors that surround it and such information should be sufficient to describe the urban texture which affects the energy consumption. The size of the buildings are

also identified as discussed in Section 5.2 by sampling from correlated data derived from the analysis of GIS data. Fig. 5-4.a provides the schematic of our probabilistic model for modeling heat losses considering the neighborhood surrounding a building. The central building for which the heat loss is calculated is shown in red. The rest of the buildings which only affect the shadowing in the model are shown in grey. We have considered several combinations in our automatized modeling procedure. The pre-processor designed to generate Energyplus inputs can produce both single- (Fig. 5-4.b) and multiple-story (Fig. 5-4.c) buildings. Considering the neighboring building, the model allows detached (Fig. 5-4.d), only one attached building (Fig. 5-4.e) and two attached buildings (Fig. 5-4.b) in modeling a group of buildings. Therefore, if such a model is sampled enough via Monte Carlo uncertainty propagation method, the superposition of model together would yield a probabilistic picture of heat loss in buildings.

The above probabilistic model can provide an invaluable insight about energy loss at the building scale. This includes the answer to the very fundamental question regarding the degree of importance of these 74 variables. Do they have the same strength in altering the heat loss are some of them more influential than others? This is achieved here by combining the Global Sensitivity Analysis (GSA) technique with the probabilistic simulation approach of the block of nine interacting buildings (see Fig. 5-5.a for schematic explanation and Appendix C for definition and different flavors of GSA). To reduce the dimension of our parameter space, we perform a sampling-based sensitivity analysis using the Monte Carlo method to propagate the uncertainty from all parameters containing all possible building sizes and specifications (Fig. 5-5.a) into energy consumption space. This Monte Carlo sampling provides a probabilistic mapping necessary to infer the contribution of each uncertain variable to the variance of heat loss using analysis of variance (ANOVA). In particular, we employ the Spearman Rank Correlation Coefficient (SRCC) to characterize the sensitivity of energy loss norms with respect to all uncertain variables. Having proposed a neighborhood energy consumption model and the pertinent ranges for uncertain variables (see Table. 5.1), we perform a Monte Carlo simulation to propagate

uncertainty from building and neighborhood property space to energy consumption space. At this stage, we generated 75,000 input samples and calculated the relevant heat loss values using Energyplus.

Having access to the input-output relation for this probabilistic energy loss model, we perform GSA on the outcomes using the SRCC method as presented in Fig. 5-5.b. As presented in Fig. 5-5.b, building size contributes the most to the variance of the energy loss. Therefore, the energy loss by itself cannot be informative about the efficiency of a building as it contains more information about the size of the building ($L_{x,y,z}$) rather than building thermal properties. However, building heat loss per envelope surface area is more informative as it is less affected by the building size and hence contains more information regarding thermal resistance of the envelope ($R_{env} = \frac{d_{env}}{k_{env}}$ where d_{env} and k_{env} are equivalent thickness and thermal conductivity of building envelope), air infiltration rate (I_{env}), internal temperature set point (T_{set}), window type (W_{typ} which identifies the number of glazing used in the window) and whether the building is attached or detached (N_c). Therefore, we note that while the largest contribution to the variance of heat loss (E) is accounted through the variability in building size, the Spearman Rank Correlation Coefficients (SRCC) in heat loss per surface (E/S) is mainly the result of the interplay between individual activity and envelope properties (T_{set}, I_{env}, R_{env}). Therefore, E/S is more informative about the thermal efficiency of the building envelope per se. Second, most of the contribution to the variance of E_{summer}/S is attributable to the consumer's set point temperature, T_{set}. This is likely the result of the fluctuating temperatures of the summers in the northeastern United States. In other words, the average monthly temperature is close to T_{comf}, making the temperature difference between the inside and outside sensitive to individual choices and not the building's thermal efficiency. However, the average monthly temperature in cold seasons drops significantly and steadily below T_{comf}, which, as shown in Fig. 5-5.b, results in winter heat loss (E_{winter}) that can be expressed in terms of building variables (I_{env}, R_{env}) and different choices per building (T_{set}). When inspecting analysis of variance's (ANOVA) results of R_{eff}, we notice that R_{eff} has the same characteristics as E_{winter}/S with the exception of

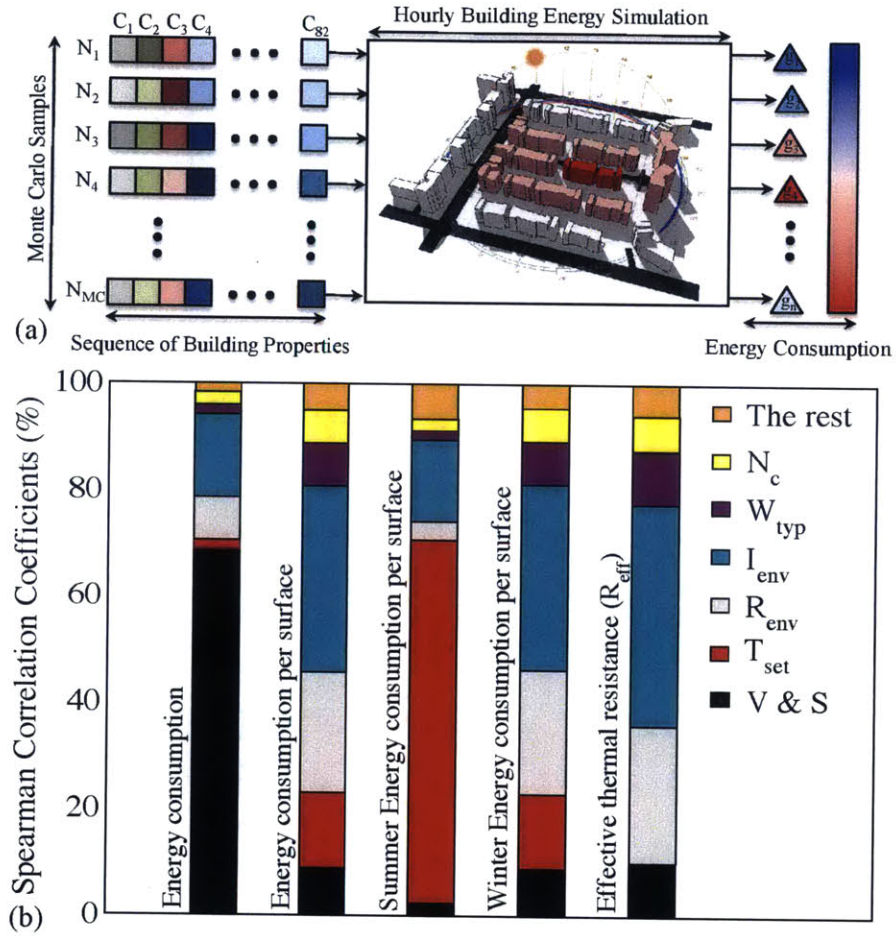

Figure 5-5: Reducing the complexity and identification of the most influential parameters in the building heat loss via the analysis of variance. a) The schematic representation of Monte Carlo uncertainty propagation method. Given the probabilistic nature of building characteristics at the city-scale denoted by 82 squares (C_1 to C_{82}), the energy consumption can be viewed as an uncertain parameter. Here, building heat loss modeling bridges the gap between the building characteristic space and the heat loss space. We randomly sample from the building characteristic space and calculate the relevant heat loss (denoted by triangles). b) Identification of the most influential parameters from the global sensitivity analysis on different heat loss norms. While the gas consumption is strongly affected by building volume and surface (V, S), the gas consumption per surface contains more information about the air infiltration rate (I_{env}), thermal resistance of walls (R_{env}), and Window type (W_{typ}). Unlike winter energy consumption, the summer consumption depends strongly on the internal temperature set point (T_{set}) due to the proximity of the outdoor and indoor temperatures. More importantly, T_{set} does not affect R_{eff} because if T_{set} is relatively constant inside a building then the temperature gradient is only due to the variation of the outdoor temperature.

being completely independent of residents' choices (T_{set}). This reflects the preferences of households to maintain T_{set} regardless of T_{out}. As a consequence, the gradient of energy consumption as a function of T_{out} is independent of T_{set} and captures solely the building efficiency properties. Thus, we confirm from ANOVA and heat loss simulations that R_{eff} is the most convenient norm of choice for capturing the physical response of buildings, reducing the consumption to only six physical variables in buildings.

5.4 Heat Loss Surrogate Modeling at the Building Level

At this stage, we have identified the influential parameters and the relevant component of heat loss to the energy efficiency of a building. However, the relation between influential parameters and the heat loss value still remains unknown. We have comprehensively addressed this issue in the current section using conventional approaches in surrogate analysis also known as metamodeling, emulator design, and response function theory. Subsection 5.4.1 discusses the application of machine learning techniques which establish the input-output relation in a mathematical sense without considering principles of physics. On the other hand, subsection 5.4.2 provides a mechanistic model based on dimensional analysis enriched with a simplified surrogate function which not only establishes the input-output relation but also provides a physical understanding of the underlying heat loss process.

5.4.1 Response Surface Models Based on Machine Learning

The purpose of Response Surface Modeling (RSM) is to construct a relation, either parametric or non-parametric, between design variables (influential parameters or factors) and observations (see Appendix D for fundamental definitions and different flavors of design of experiment). The literature on RSM and Machine Learning (ML) contains several approaches for relating inputs and outputs in a mathematically rig-

orous fashion. In Appendix E, we provide the basic definitions of response surface methodology and three main falvors of it, namely polynomial regression (PRG), Kriging and Multiple Adaptive Regression Splines (MARS), which are employed in our study. Readers are encouraged to review this Appendix before proceeding to read the rest of this subsection. The challenge here is to identify the most reliable and cost effective ML method that is suitable for heat loss modeling. Researchers have compared different machine learning techniques to predict heat loss of buildings. Recently, Tso et al. [342] compared the performance of PRG, decision tree and neural networks. Here, we aim at comparing the performance of the above three methods and discover their advantages and disadvantages.

As thoroughly discussed in section 5.2, we can identify the most influential parameters affecting energy consumption and related heat loss norms. Here, we employ the aforementioned machine learning techniques to relate these influential parameters to heat loss per surface. We have performed a comparative study to analyze the performance of linear polynomial, pure quadratic polynomial, linear Kriging, cubic Kriging, linear MARS and cubic MARS against results of energy consumption modeling provided by Energyplus. To this end, we performed Pseudo-Monte Carlo sampling in the reduced space of the most influential parameters. As discussed in sec. 5.2, this critical parameter space consists of building envelope thermal resistance (R_{env}), the air infiltration rate (I_{env}), envelope temperature set point (T_{set}), building length, width and height (L_x, L_y and L_z), window type (W_{typ}) and umber of buildings sharing a wall (N_c). The rest of parameters are kept at their mean values provided the table 5.1. Subsequently, we have created a database of 80,000 Monte Carlo samples to study the performance and convergence of different learning techniques. We split the database into two groups: learning set and validation set. The learning set was solely used for training purposes and was not used for the validation of meta models. We have used sample sizes of 100, 200, 500, 1000 and 5000 for this study. The validation set was used to evaluate the prediction power of the surrogate models upon the absence of the training set. Fig. E.2 provides the learning curves for the above surrogate models. Fig. E.2.a displays the coefficient of determination

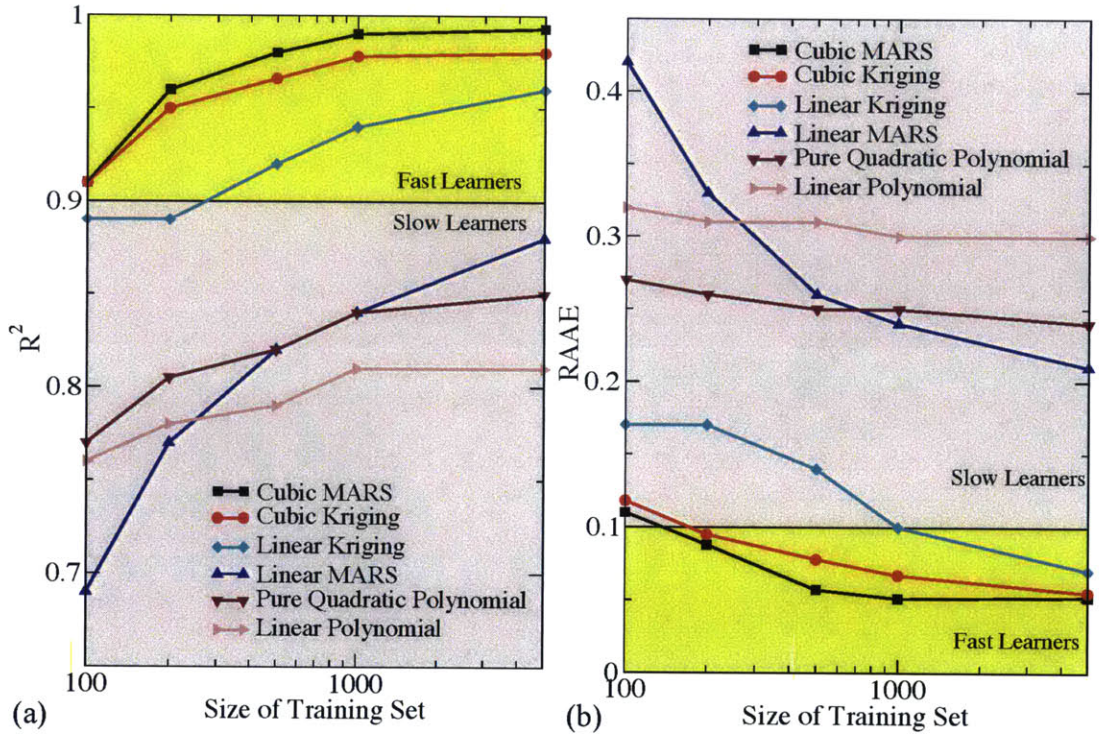

Figure 5-6: Comparative study of different machine learning methods as a function of the size of training set utilized to construct the surrogate model. a) R^2 and b) RAAE values for linear and pure quadratic polynomials, linear and cubic Kriging and MARS techniques.

as a function of the size of the training set. Generally, for all surrogate models, R^2 increases with the size of the training set and reaches a plateau at large values. Also, we find that in each type of the surrogate model, the higher order (quadratic and cubic) versions perform better as they can smoothly describe the curvature of the space. Among all metamodels, linear and cubic Kriging and cubic MARS perform exceptionally even with training sets as small as 100 samples. With training sets as large as 5000 samples, both cubic MARS and Kriging achieve R^2 values in the order of 0.99. Linear Kriging also performs well and reaches an R^2 of 0.96 at 5000 samples. The R^2 of Linear MARS increases significantly with the training set size and reaches 0.88 for the largest size. However, R^2 analysis indicates that the polynomial models cannot precisely describe the variation of energy consumption in the reduced space. This is readily seen in the magnitude of RAAE as function of the size of the training set provided in Fig. E.2.b. In fact, only the three surrogate models in the fast learning region reach the acceptable level of 95% accuracy within the range of simulated training sets. In fact, linear and pure quadratic polynomials do not have sufficient degrees of freedom to explain the variability of heat loss in the design space. To further investigate the performance of the surrogate models, we have plotted the performance of the surrogate models in R_{env}-I_{env} space. Fig. 5-7 compares the color maps of cumulative heat loss per surface (E/S) in R_{env}-I_{env} space. While Figs. 5-7.a,b,c display the color maps of the quadratic polynomial, cubic Kriging and cubic MARS, respectively, Fig. 5-7.d provides results obtained by performing the analysis in Energyplus. In fact, Fig. 5-7.d can be regarded as the true response while the rest of the plots are different approximations to the exact evaluations. As can be seen in Fig. 5-7.a, the pure quadratic polynomial completely fails to approximate the true response due to fictitious curvature imposed by its parametric form. Fig. 5-7.b indicates that cubic Kriging performs nicely. However, from the iso-performance lines, we can clearly notice that it is an extension of the pure quadratic model. This is resonated in the change of curvature at high R-values which is reminiscent of pure quadratic model behavior. In fact, the Kriging formulation is intended to provide a mathematical framework for the description of deviation from simple polynomial

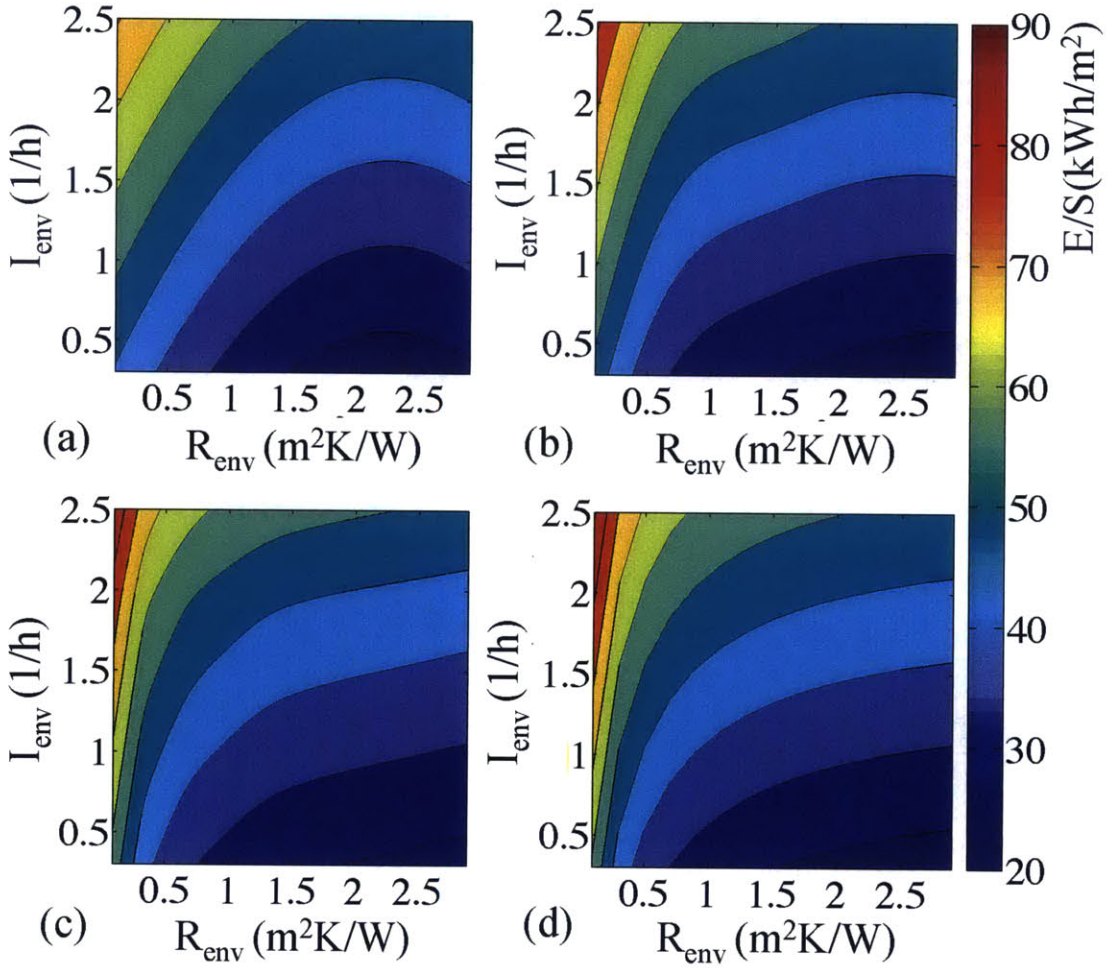

Figure 5-7: Comparing the energy consumption per surface predicted via a) Polynomial regression b) Kriging and c) MARS against the response of d) Energyplus software in R_{env}-I_{env} space.

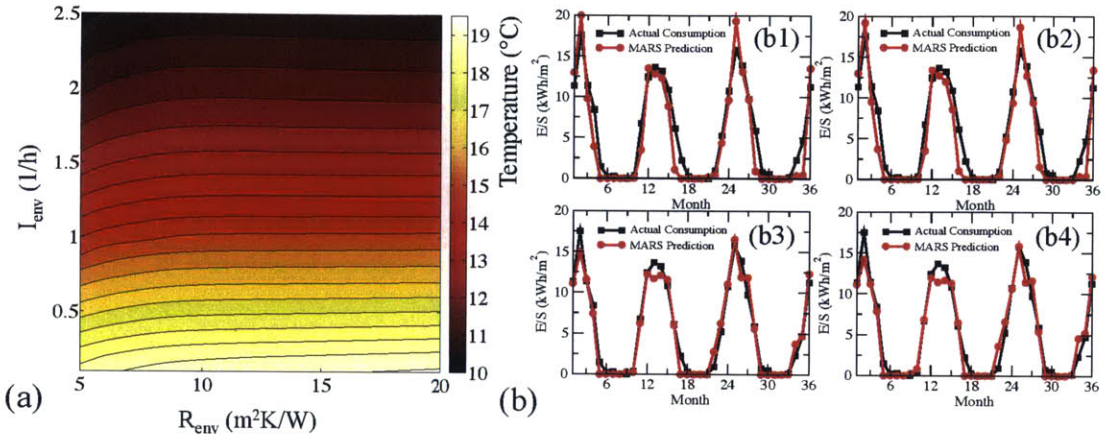

Figure 5-8: Application of building surrogate models in identifying building thermal characteristics and residents' choices from monthly energy consumption data. a) Building iso-performance surface derived from MARS surrogate model. The iso-performance surface indicates all triplets of (R_{env},I_{env}, T_{set}) that flourish the same monthly energy consumption patterns. b) The predictions of MARS surrogate function at the corners of the iso-performance surface compared against actual energy consumption data.

models. The visualization indicates the effectiveness of such an approach. Despite its powerful predictive strength, Kriging suffers from a significant disadvantage. To calculate the optimal Kriging coefficients, we need to find the inverse of a $N_s \times N_s$ (N_s is the size of the training set) non-sparse matrix which is computationally demanding and requires a lot of memory for construction of models with a relatively large number of observations. As shown in Fig. 5-7.c, cubic MARS can predict almost identical responses produced by Energyplus. Cubic MARS is not only more accurate than Kriging but also computationally much less intensive. This makes cubic MARS a great candidate for constructing response surfaces for the purpose of heat loss modeling and optimization. After controlling the quality of the surrogate models through performance metrics, we can utilize these high fidelity models such as MARS or Kriging for fast and accurate estimation of heat losses in different contexts such as system identification and strategic scenario planning. We denote these surrogate

models in a functional form:

$$\frac{E}{S} = \Psi\left(R_{env}, I_{env}, T_{set}, L_x, L_y, L_z, W_{typ}, N_c\right) \tag{5.3}$$

where Ψ is the verified surrogate model. A properly trained MARS surrogate function addresses the direct analysis $((R_{env}, I_{env}, T_{set}, L_x, L_y, L_z, W_{typ}, N_c) \mapsto \frac{E}{S})$. However, in retrofit planning (reducing heat losses), we aim at linking E/S to the actual building specifications $(I_{env}, R_{env}, W_{typ})$ and individual choices (T_{set}). Such an objective involves an inverse analysis analogous to problems in the system identification domain [197]. For a building with a given set of dimensions and attached neighbors, the inverse analysis maps heat losses to the envelope properties $(E/S|_{L_x,L_y,L_z,N_c} \mapsto (T_{set}, I_{env}, R_{env}, W_{typ}))$. The resulting multi-dimensional inverse problem can be solved via the gradient-based contour schemes following [91]. Here, we minimize the L_2-norm of the goodness of the fit, $|\psi - E/S|_2$, to search for the admissible solutions of parameters. Similar to the majority of the multi-dimensional inverse problems, the solution to this heat loss identification problem is non-unique. In fact, for a given window type, there are a range of $(I_{env}, R_{env}, T_{set})$ triplets that correspond to a fixed $\frac{E}{S}$. As shown in Fig. 5-8.a, the set of admissible triplets forms a 3-dimensional plane known as an iso-performance surface [91]. In fact, the sets of parameters located on the iso-performance surface respond identically to the thermal loading at the monthly scale. For instance, corner points in Fig. 5-8.a denote four iso-performance points, with different infiltration and conduction rates. As denoted in Fig.5-8.b, the $\frac{E}{S} - T$ relation predicted by the direct evaluation of the MARS surrogate function at these four points is the same. Therefore, a building can have infinite states $(I_{env}, R_{env}, T_{set}, W_{typ})$ corresponding to an identical $\frac{E}{S}$.

5.4.2 Mechanistic Model Based On Dimensional Analysis

The main argument against machine learning methods for predicting the behavior of physical systems is that it does not provide physical insight and its predictions might even violate the principles of physics. This is actually true as machine learning meth-

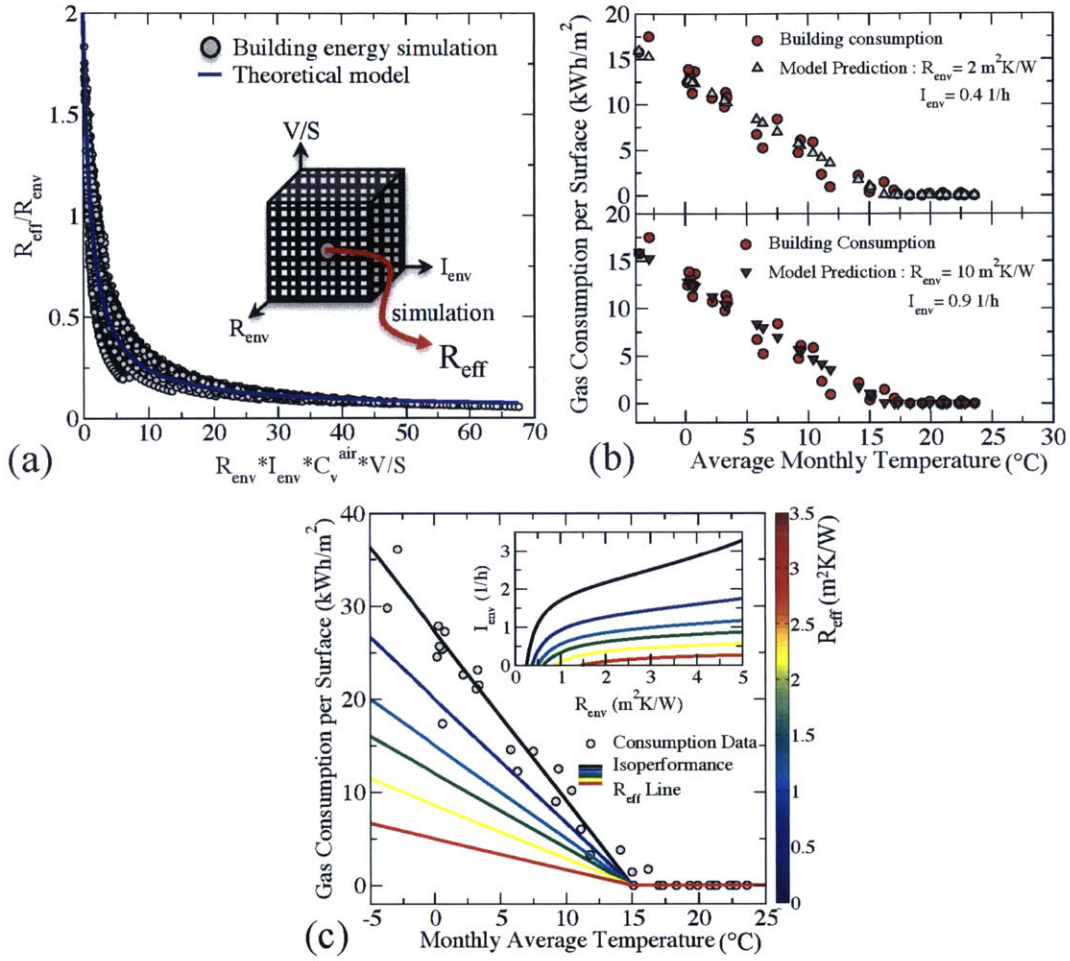

Figure 5-9: Estimation of building energy consumption in dimensionless space via the response surface methodology. a) Construction of the surrogate model in dimensionless form. Considering a dense full-factorial grid in (I_{env}, R_{env}, V/S) space (as denoted in the inset), R_eff can be calculated at each point via building energy simulation using Energyplus. For a building with a given W_{typ} and N_c, the dimensional analysis yields that there only exist two dimensionless quantities relating R_{eff} to the rest of the influential parameters: I_{env}, R_{env}, V/S and C_v^{air} (volume heat capacity of air) which are $\Pi_1 = R_{eff}/R_{env}$ and $\Pi_2 = I_{env} \times R_{env} \times C_v^{air} \times V/S$. A surrogate model of the form $\Pi_1 = 1/(A_1 + A_2 \times \Pi_2)$ as shown by the blue line is fitted to the simulation results. b) Building thermal properties identification from monthly gas consumption data using the surrogate model. The solution of inverse problems to identify the building characteristics such as I_{env} and R_{env} is non-unique. For instance, the surrogate model predicts the same R_{eff} value at the two alternative cases of (I_{env} = 0.4 $1/h$, R_{env} = 2 m^2K/W) and (I_{env} = 0.9 $1/h$, R_{env} = 10 m^2K/W) located on an iso-performance line. c) Estimation of retrofit energy savings at the building level. Given the monthly gas consumption of the building, R_{eff} is estimated by fitting a line denoted in black. The associated R_{eff} attributes to the black iso-performance line in the inset in (R_{env}, I_{env}) space.

ods are only designed based mathematical arguments rather than physics principles. In this subsection, our task is reduced to quantitatively describe R_{eff} as a function of physical properties of each building and to quantify the impact of retrofits at the city-scale. From a dimensional perspective, the effective thermal resistance of an envelope with dominant conduction and infiltration heat transfer mechanisms can be written as $R_{eff} = \psi(R_{env}, I_{env}, V, S, C_v^{air})$. Dimensional analysis reduces this functional form to a simple relation between a reduced number of dimensionless parameters. This is achieved here by rationalizing the problem through a dimensional analysis of the physical quantities involved that possibly affect R_{eff}, namely the effective thermal resistance of the building envelope (R_{env}), the air infiltration rate (I_{env}), the volumetric heat capacity of air (C_v^{air}) emphasizing that the heat exchange is performed through air, and the building's characteristic dimension expressed by the volume-to-surface ratio (V/S). The rank of the exponent matrix, the matrix formed by the exponents of variables' dimensions, is 4. Thus according to the Π-theorem [51], there is only one independent dimensionless variable among the initial six parameters. Therefore, dimensional analysis allows us to further reduce the dimension of the problem to a 1-parameter relation between the dimensionless thermal resistance and the ratio of infiltration (Q_{inf}) to conduction losses (Q_{cond}).The dimensionless relation is:

$$\Pi_1 = \frac{R_{eff}}{R_{env}} = F\left(\Pi_2 = R_{env} \times I_{env} \times C_v^{air}\frac{S}{V} = \frac{Q_{inf}}{Q_{cond}}\right) \tag{5.4}$$

where Π_1 is the ratio of the effective thermal resistance of the system to the conduction resistance of the envelope and Π_2 will be shown to be the ratio of the infiltration heat transfer to the conductive heat transfer. This implies that a simple functional form of $\Pi_1 = F(\Pi_2)$ is sufficient to describe the physical response of the system without the need of extensive simulations. Therefore, dimensional analysis effectively reduces the number of variables but it does not quantify the relation between them. To this end, we utilize the conservation of energy law to propose a functional form of F. To determine this functional relation, a full factorial design in the (R_{env}, I_{env}, V/S) space is performed by means of simulations (Fig. 5-9.a). The conservation of energy

for the control volume (volume inside an envelope), which exchanges heat with the surrounding media through conduction and infiltration, can be written as:

$$Q_{tot} = \frac{A_1 \times S \times \Delta T}{R_{env}} + A_2 \times I_{env} \times C_v^{air} \times V \times \Delta T \tag{5.5}$$

where Q_{tot} is the total heat loss through the envelope and thus is equal to $(S \times \Delta T)/R_{eff}$, which can be rearranged in the following form:

$$\Pi_1 = \frac{R_{eff}}{R_{env}} = \frac{1}{(A_1 + A_2 \times \Pi_2)} \tag{5.6}$$

where A_1 and A_2 are the degrees of the freedom in the model. These parameters need to be calibrated with the results of physical simulations in the urban settlement under consideration. Here, we used Energyplus software to numerically estimate these coefficients. We have performed a full factorial simulation varying R_{env}, I_{env} and V/S. R_{eff} is computed as the derivative of predicted heat losses with respect to average monthly temperature. The results are plotted in the $\Pi_1 - \Pi_2$ space and A_1 and A_2 are derived by fitting eq. 5.6 to the results via the least-square approach. For instance, for the case of detached buildings with double-glazed windows; $A_1 = 0.49$ and $A_2 = 0.30$; see Fig. 5-9.a. The dimensionless form (eq. 5.6) provides insights into a building's thermal properties and energy performance from reading its monthly energy consumption data alone. For instance, as shown in Fig. 5-9.b, higher energy efficiency at the building scale can be equally achieved by increasing the thermal resistance of the envelope or by decreasing the air-infiltration rate. That is, all retrofit solutions are located on an iso-performance line [91] in the (R_{env}, I_{env}) space where any point corresponds to a unique value of R_{eff}. This conclusion was further tested via standard machine learning methods, namely by means of Multiple Adaptive Regression Splines (MARS) [119], well suited for capturing response surfaces of multi-parametric problems [291, 226]. These results demonstrate that given the nature of heat loss response to the parameters space, the response of the system as a surrogate function is always solvable.

5.5 Data Analytics of Heat Loss at the City Level

The first step in analyzing city scale heat loss is to understand the sensitivity of energy loss of the city's building stock to changes in temperature. Utility companies record the total monthly energy consumption, E, per parcel in kilowatt-hour (kWh) across the entire city. Herein, we use a 3-year-long record (2007-2009) collected for billing purposes and anonymized by the data source, capturing the consumption pattern of more than 6,200 individual residential buildings in Cambridge, MA. When matching this record, for each home, with the buildings' footprint using data from geographic information system (GIS), and the average monthly temperature calculated by averaging hourly temperature records from the closest weather station [?], the energy loss is found particularly to exhibit a characteristic piecewise linear form (Fig. 5-10.a), separated by a cut-off temperature, T_0. We found no statistically significant correlation between electricity consumption and the outdoor temperature at the building level, most likely the result of insufficient record length. Whereas the gas-related energy demand increases almost linearly below this outdoor cut-off temperature, it does not vary significantly for higher temperatures; thus defining a temperature-insensitive baseline gas consumption (E_0), which is most likely due to hot water production. We identify this cut-off temperature, as the temperature below which consumers turn on the home's heating system to maintain indoor spaces at a desired comfort temperature, T_{comf}. In other words, the outdoor cut-off temperature is representative of individual choices or behavioral patterns. The probability density function of T_0 for the analyzed building sample (Fig. 5-10.b) sheds some light on this behavioral pattern, in the form of three major peaks at 13, 15 and 17 (with 1°C standard deviation) elucidating the core of the distribution (93% of the overall distribution, while the remaining 7% have medians at 4.7, 8.5 and 23°C). These empirical findings give us means to account in energy simulations at city scale for distributions of the comfort temperature as a function of the external temperature.

Interestingly, each building has a constant rate in the increase of heat loss below the offset temperature, suggesting a linear form between the heating energy and

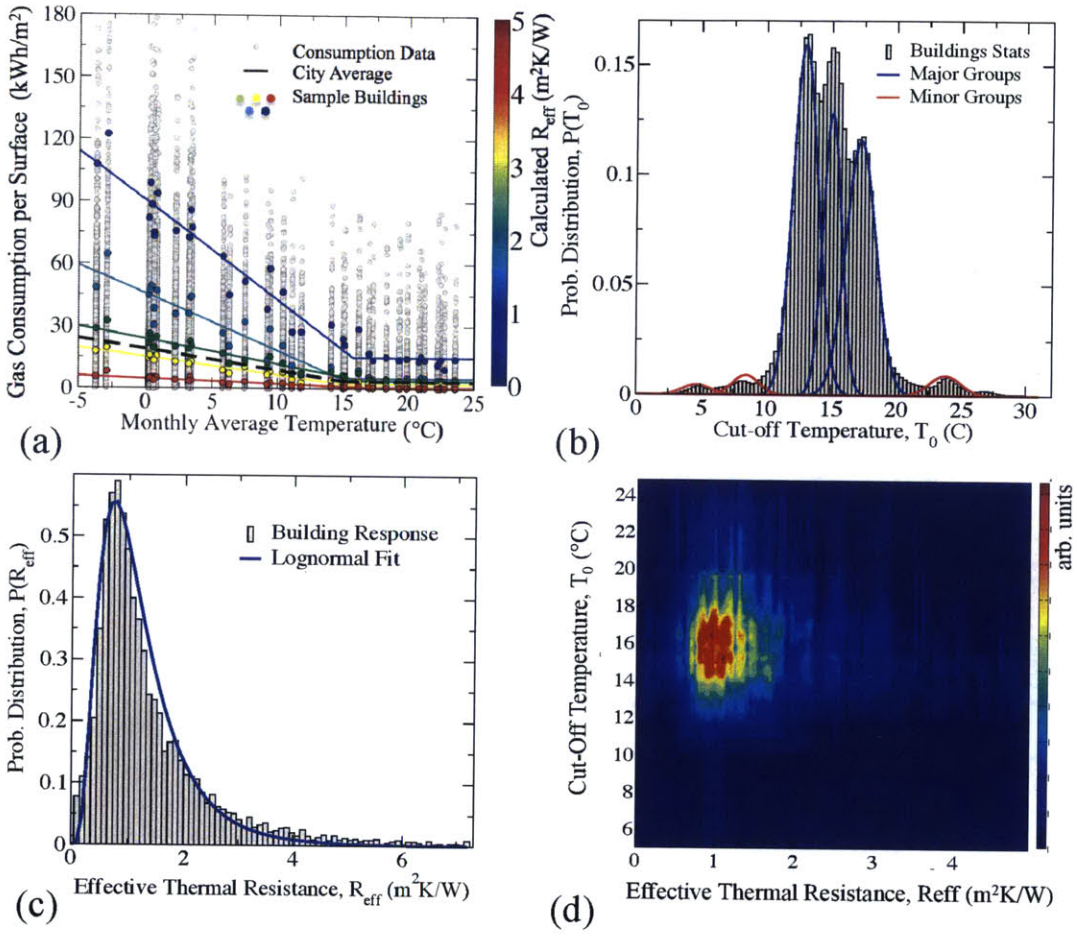

Figure 5-10: Data assimilation and analysis by integration of various data sources including buildings footprints, weather data, and energy consumption bills. a) Gas consumption per surface as a function of monthly average temperature for more than 6204 buildings in a period of 2007-2009 in Cambridge, MA. The average city consumption per surface is shown in dashed black line. Five sample buildings are highlighted based on their effective thermal resistance defined in eq. 5.7. The gas consumption of buildings exhibits a piecewise linear trend with energy consumption increasing below a certain temperature threshold. b) Distribution of the cutoff temperature indicating the variability of consumers' perception of and resistance to cold weather. c) Distribution of effective thermal resistance for all the buildings. The Gaussian mixture analysis identifies the three major populations of the consumers that turn on their heating system at 17, 15, and 13°C on average. The distribution of effective normal resistance follows a lognormal distribution, with an average value of 1 m^2K/W. d) the joint probability distribution function of the T_0 and R_{eff}. The absence of correlation suggests that the energy consumption of buildings and the consumers' behavior are not correlated.

outside temperature in excess of the baseline gas consumption (E_0), in the form:

$$E - E_0 = \frac{1}{R_{eff}} \langle T_0 - T_{out} \rangle \times \Delta T \times S \tag{5.7}$$

where S is the building's envelope surface area and E is the heating energy necessary to maintain an inside temperature of T_{comf} when the outside temperature T_{out} is bellow T_0 during the time interval of exposure, Δt, corresponding to the numbers of hours between two consecutive energy readings (E) by the utility company. Moreover, the linearity between the temperature difference $T_0 - T_{out}$ and the heating energy defines a linear coefficient, R_{eff} (in m^2K/W), that can be viewed as an effective thermal resistance representative of the energy efficiency of a building. Unlike the consumption coefficient in the degree-day approach [294, 207], we have shown that R_{eff} is the most effective heat loss norm and depends only on physical attributes of a building's envelope; namely heat transport, and infiltration properties. For the sample of 6,204 homes in Cambridge, MA, the effective thermal resistance that is obtained by a linear fitting of the energy readings according to eq. 5.7, is found to follow a lognormal distribution (Fig. 5-10.c). The fact that this distribution is uncorrelated with residents' preference for temperature (Fig. 5-10.d), establishes a means to separate behavioral and physical attributes in the choice of retrofitting at city scale.

5.6 Minimizing Heat Losses at the City Level

It is precisely this data analysis combined with a reduced model with a single parameter that allows us to employ the dimensionless model, eq. 5.6, to explore scalable retrofit options of buildings at city scale in function of collected records of energy consumption. We start at the building level, by considering the gas consumption of an arbitrary building and its pertinent linear fit (Fig. 5-9.c). The so obtained effective heat resistance, R_{eff}^{-}, is situated on an iso-performance line (black solid line in the inset of Fig. 5-9.c), which captures the current energy performance of the building

in terms of envelope heat resistance (R_{env}) and infiltration rate (I_{env}) according to eq. 5.6. Any retrofit option, such as increasing insulation (R_{env}), reducing air infiltration rate (I_{env}) or installing multiple-paned windows (W_{typ}), while retaining the preferred behavioral choices (same value of T_0 in Fig. 5-9.c) would entail an increase of $R^-_{eff} \rightarrow R^+_{eff}$, to higher iso-performance levels (smaller slope of energy consumption in Fig. 5-9.c), and thus, in the light of eq. 5.7, to an energy saving after retrofitting of $\Delta E = E^- - E^+ = (1 - \frac{R^-_{eff}}{R^+_{eff}}(E^- - E_0)$, which is independent of the particular choice of retrofit.

This simple form provides a straightforward means to upscale the heat loss reduction potential from the building to the city scale in view of urban policy choice and implementation. That is, the challenge pertaining to city-scale strategic retrofit planning is concerned with finding the shortest retrofit path that achieves the highest energy savings with the least number of retrofitted buildings. In this regard, an important feature emerges from the ranking of the potential energy savings of buildings (Fig. 5-11.a); in that the rank and magnitude of energy saving follows over a large range a power law with an exponent of 0.75; much alike Zipf's law [122, 75]. While the deviation of the tail from Zipf's law attributes to buildings with insignificant energy savings, the tail of the energy saving distribution follows a power law with the exponent of 2.2 (inset of Fig 5-11.a). Given the significance of a Zipf-type data distribution, it appears to us that such a ranking based on energy saving potential will provide the shortest path for city scale energy savings. To test our hypothesis, we compare this ranking with other selection criteria associated with urban policy choices; starting with a random retrofit of buildings at city scale, performed upon requests of building owners, in which case the achieved energy saving scales linearly with the number of retrofits. The results of this analysis displayed in Fig. 5-11.b show that an informed selection based on ranking the energy saving of buildings ($rank\,(\Delta E)$) provides indeed the highest rate of energy saving per retrofit; followed by an informed selection based on ranking of building energy-per-surface consumption: $\left(rank\left(\frac{E^- - E_0}{S}\right)\right)\left(rank\left(\frac{1}{R^-_{eff}}\right)\right)$. By way of example, if Cambridge, MA targeted a 60% overall energy reduction related to heating, it would suffice, with such

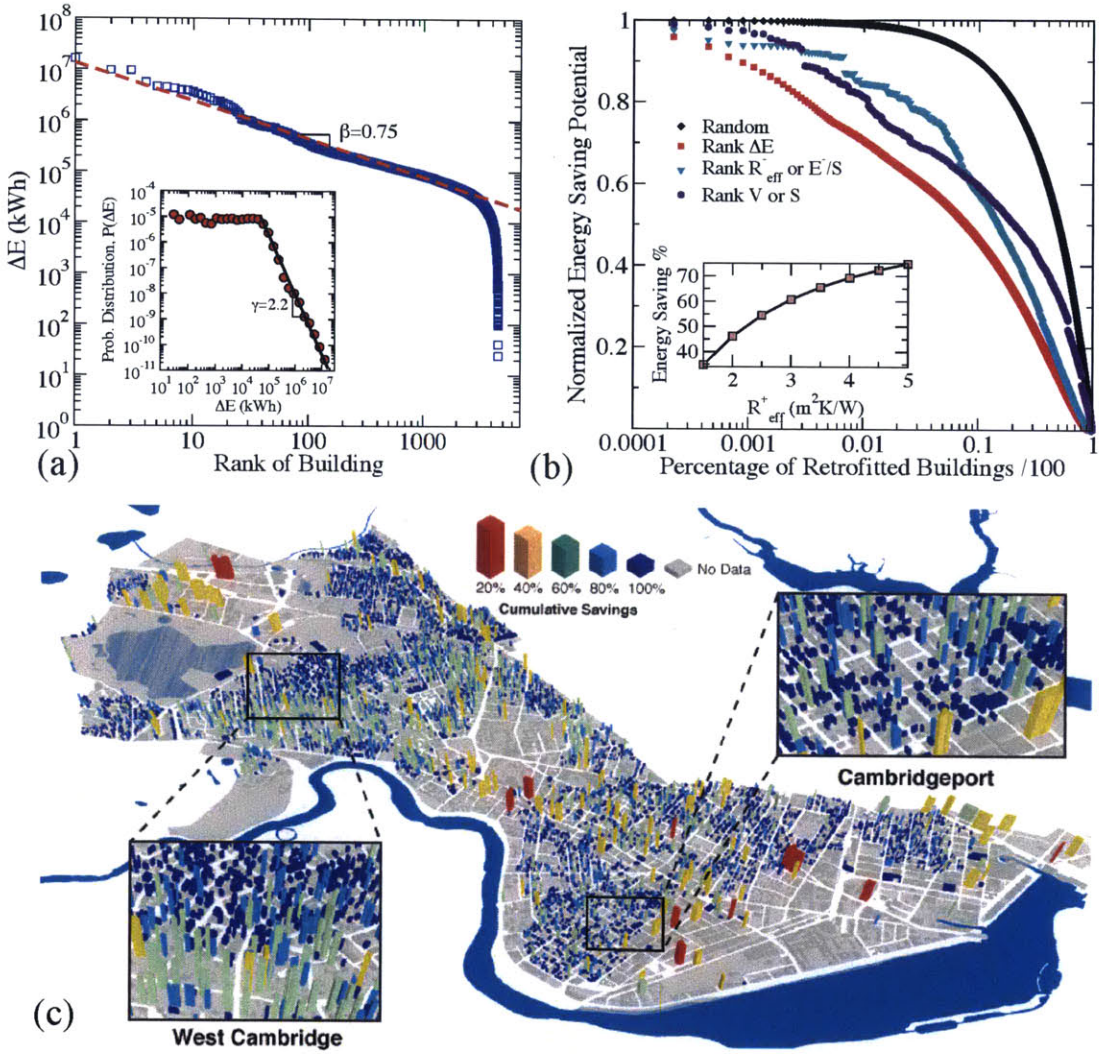

Figure 5-11: Citywide retrofitability analysis by combining GIS, weather, energy consumption data, and surrogate modeling at individual building level. a) Energy saving as a function of retrofit rank of building which follows a Zipf-like relation with the exponent of $\beta = 0.75$. The inset displays the distribution of ΔE, which indicates a linear tail in log-log scale akin to scaling laws. b) Comparative study between different retrofit scenarios based on ranking of ΔE, E^-, R^-_{eff}, V or S and demonstrating their effectiveness against random scenario (blind retrofit) at the city-scale. The calculation of energy saving potential at the city level is based on assumption of $R^+_{eff} = 3.0m^2K/W$ after retrofit. If buildings are retrofitted on a random basis, the gas consumption of the city decreases linearly with the number of retrofits. If the buildings are retrofitted based on their potential energy saving, the city-scale gas consumption decreases non-linearly through the shortest path strategy. The inset c) The GIS map of buildings in Cambridge, MA with colors representing the retrofitability potential at the city-level. The colors are randomized to respect the privacy laws and terms of non-disclosure agreement (NDA) required by the utility company (NSTAR).

an informed selection process, to retrofit 16% of the entire building stock as mapped in Fig. 5-11.c in order to achieve a 40% reduction of current energy use; in contrast to 67% of buildings with a random selection procedure to achieve the same target. That is, the proposed selection scheme based on ranking potential energy savings provides an efficient means to achieve the shortest path for substantial energy savings at city scale.

5.7 Conclusions

To conclude, we propose a method of analysis that combines data on energy, climate, and buildings footprints with surrogate energy modeling. This powerful framework reduces the complexity of the problem to a simple functional form to estimate the thermal response of buildings. Calibrated with utility data, this functional form allows us to easily estimate potential energy savings per building under different retrofit scenarios with minimal computational expense. When applied at the urban scale, we can make informed selections towards the reduction of the energy consumption footprint by identifying the shortest path to the desired goal. This approach has profound environmental implications for the associated greenhouse gas emissions of energy consumption and its global warming potentials. The method is portable to any city in the world, requiring solely data that is readily available for billing and urban planning purposes. More generally, similar model reduction approaches combining large data with statistical analysis and physical simulations to gain predictive understanding of the system's response appear to us most promising for urban energy solutions such as patterns of hourly electricity demand and the adoption of alternative sources for generation of electricity. These data driven methods have the premise to help cities optimize decisions that make them more environmentally and economically sustainable.

Chapter 6

Conclusions and Future Works

6.1 Defects Really Matter.

Actually, there is a common theme that glues our findings in different chapters together. Regardless of the system and its size, whether it is the molecular structure of C-S-H or heat losses at the city scale, defects are of high importance. Let us remind ourselves that the accumulation of defects in the C-S-H is responsible for decreasing mechanical properties of C-S-H's nanostructure. These defects not only reduce the stiffness of molecular network but also diminish its strength properties. These defects are also responsible for creating adsorption sites that increase water content with increasing Ca/Si ratio. Although it is not influenced by atomic-scale defects, the thermal conductivity of C-S-H is strongly affected by meso- and microscale porosity values. The sensitivity analysis at the city level identifies the infiltration as the most important factor for reducing heat losses in buildings. In fact, excessive infiltration rate is rather a construction defect that should be limited during the design phase or retrofit process. Therefore, to reduce energy losses in buildings, one might simultaneously increase the defects to intentionally scatter conduction of heat and decrease construction defects to restrain excessive energy losses through other heat loss mechanisms. Defects at the material level are a bit more complex and entail other consequences with them. While mechanical properties, elastic and strength, are maximized with lowering defects, the thermal properties are minimized with increasing

the defect content. Therefore, a sound material design relies on characterizing the optimum defect content at each scale rather than simply maximizing or minimizing it across all length scales.

6.2 The Notion of C-S-H As a Quasi-glass.

After reading previous chapters, one might conclude that the biggest contribution of this thesis is providing a realistic molecular structure of C-S-H at varying Ca/Si ratio. In fact, this has some truth to it as this step brings together many pieces of experimental data from different research groups in the past two decades. We showed that high-throughput combinatorial molecular modeling can explain drying, elastic and inelastic neutron scattering, nuclear magnetic resonance, extended X-ray Absorption fine structure, transmission electron microscopy and coupled nanoindenation and wavelength dispersive spectroscopy analysis. Without this database of molecular structures, it would have been impossible to study the evolution of mechanical properties, structure and dynamics of ultraconfined water, coefficient of thermal expansion, heat capacity and heat conductivity of C-S-H, the binding phase of cement paste. However, the creation of this database, its verification against available experimental data, and analysis of various physical properties serve as a means for a greater purpose.

The contribution of this work is finding similarities between cement science and other branches of materials science and engineering. This relieves cement from the scientific barriers that have held it back for many decades by placing it in the same footing as glassy materials and supercooled liquids. There are five figures in the thesis that support this claim. Figures 2-4 and 2-5 are the fist set of computational observations that support the presence of intermediate phases similar to that of rigidity transition window in binary glasses. To support this notion, ideas from topological constraint theory were applied to the C-S-H database. This revealed that the intermediate phases pertain to those molecular structures having three constraints per atoms analogous to isostatic molecular structures of glasses at the transition window. In

addition to this, the presence of the Boson peak as presented in Figure 4-3 in the low THz region reconfirms that C-S-H has a lot more in common with glassy structures than crystalline materials. The inhomogeneous diffusive motion of inetrlayer water with anomalous dynamical properties, summarized in Figures 3-3 and 3-7, is akin to that of glassy phases and supercooled liquids.

There are some evidences that do not allow C-S-H to be fully considered as a glassy material. The most notable example of which is Figure 2-3 indicating that even at large Ca/Si ratios, C-S-H still retains some layered order. We emphasize that quantitative indices indicate that the magnitude of this order is very small. For this reason, we would rather adopt a conservative position by classifying C-S-H as a *"quasi-glass"*. We coin this term for the molecular structure of cement paste to highlight a potential connection with glass physics. Indeed, exploiting the ideas of glass science in studying cement paste could be a promising research direction in the future.

6.3 The Road Ahead.

This work would benefit in many ways from follow-up research studies. As for the molecular structure of C-S-H, there are a lot more to be done. The mechanical properties calculated in this work are at zero temperature. Therefore, it would be very insightful to measure these properties at finite temperature. Also, these molecular models provide a robust foundation for elemental incorporation in C-S-H. This elemental substitutions may include aluminum, magnesium, mercury and even radionuclide elements if possible. The mobility and transport of ions in the structure of CSH remains to be further investigated in the future. With possible advancement of computational speed in the near future, full scale first principle calculations of C-S-H models will become computationally more tractable. This flow of quantum information opens a new horizon for studying electronic structure, charge distribution, magnetic properties and polarization phenomenon in C-S-H to name a few. At the city-scale, the energy loss simulations were performed using the most simple HVAC

system. These studies would strongly benefit from coupled building and HVAC modeling to identify both static and dynamic scheduling schemes.

Appendix A

Recipe for Constructing Realistic C-S-H Models

This appendix is intended to explain the details required to reproduce the results of the atomistic simulations provided in Chapter 2. We have described in details different stages in construction of realistic C-S-H models. Furthermore, we have extensively discussed how these numerical samples can be utilized to calculate elastic properties, strength, free energy, structure factor and total X-ray pair correlation function.

A.1 Atomistic Simulation and Model Construction

The computational strategy used here, aiming to vary Ca/Si systematically and explore resulting properties, is a combination of several classical atomistic simulation techniques. These methods are deployed at different stages of the model preparation. The complex hydrate materials have different types of atoms with varying chemical environments, which is discussed in details in following subsections. These simulations, which have classical nature, are performed using ReaxFF [212] and CSH-FF [306] potentials. As schematically presented in Fig. A-1, there are eight explicit stages to construct a proper molecular structure of calcium-silicate-hydrates (C-S-H). To eliminate the finite size effect in computational modeling, a super-cell of the molecular structure of Tobermorite 11Å due to Hamid [139] is prepared in the first

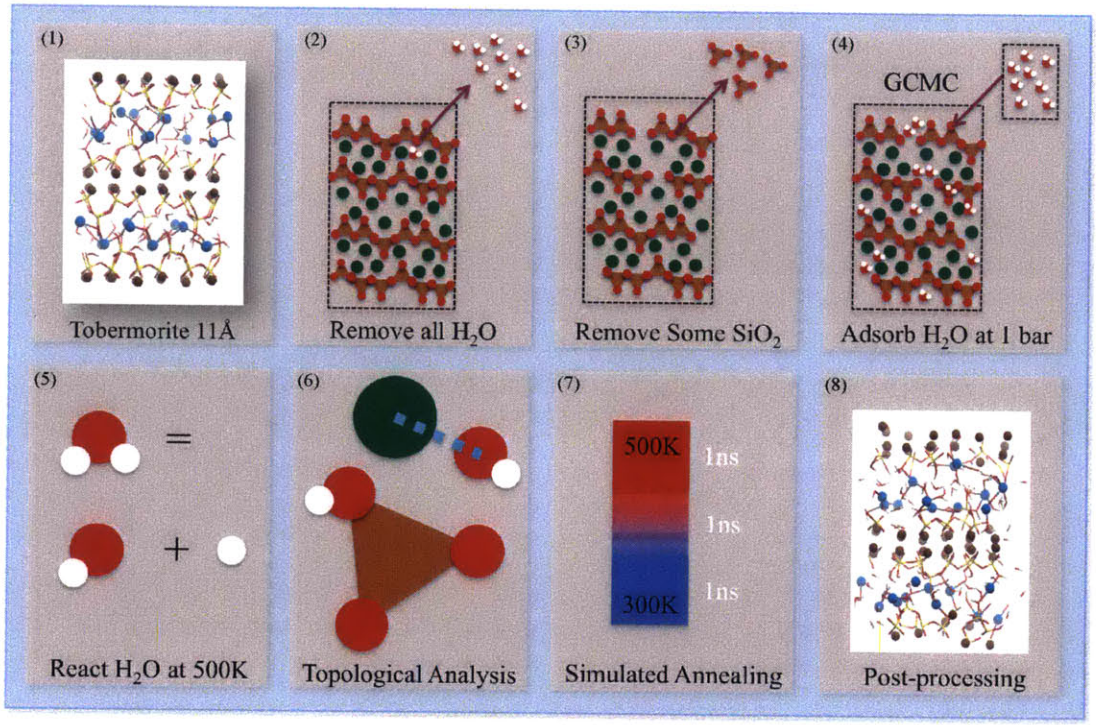

Figure A-1: The eight stages of model construction. These stages are strictly followed 150 times to produce C-S-H samples with varying Ca/Si ratio ranging from 1.1 to 2.1.

step. 11Å Tobermorite has no hydroxyl groups and the length of its silica chains is infinite. In the second step, the entire water molecules are removed from the interlayer spacing. In the third stage, 150 numerical models are constructed by randomly cutting the silica chains (removing charge-neutral SiO_2 groups). This increases the Ca/Si ratio (C/S) of the models. This provides a wide range of C/S to investigate the effect of stoichiometry on physical properties of C-S-H. At this stage, first the interlayer calcium atoms and subsequently all the atomic species and cell dimensions are relaxed using energy minimization at 0 K. In the next stage, water molecules are adsorbed back to the interlayer spacing and nano-voids created by removing SiO_2 groups using the Grand Canonical Monte Carlo simulation. At this point, the interlayer water is reacted at 500 K with the defective calcium-silicate backbone of C-S-H using ReaxFF potential. In this stage part of the interlayer water dissociates in to hydroxyl groups and proton. Minor condensations of silica chains are also observed in some samples. In the sixth stage, a comprehensive topological analysis is performed to identify the local environment of each atom. This facilitates the transfer of model from reactive environments provided via ReaxFF potential to non-reactive environment in CSH-FF force field. In stage seven, a 3ns long simulated annealing is performed on each individual sample to reduce the temperature from 500 K to 300 K at ambient pressure. Finally, the samples are relaxed and proper sampling of energy space is performed to measure elastic constants, hardness measurements, free energy, medium-range order and total X-ray pair correlation function.

Table A.1: partial charges of species in C-S-H-FF potential.

spec	charge
Si	1.722357
Ca	1.435466
Cw	1.705529
Ow	-0.820000
Hw	0.410000
H	0.425000
O	[-1.15,-1.26]
Oh	[-0.93,-0.97]
Ob	[-1.01,-1.13]

A.2 Descriptions of Atomic Species

Although minor impurities might be existing in C-S-H [9, 210], the main binding phase in hardened cement paste is generally considered to be consisted of Si, Ca, O and H atoms. These atoms adopt different characteristics depending on their local chemical environments. To appropriately address such variations, classical potentials utilize different labels for a given element to describe these chemical environments. However, ReaxFF potential shows the ability to calculate the interatomic interactions without the need to designate different labels for various chemical environments. For instance, due to charge compensation, the O-H bonding would be different inside a water molecule and a hydroxyl group. Therefore, for classical simulations using CSH-FF potential, one requires the a priori knowledge about the local atomic environments in the molecular structure of material. More specifically, CSH-FF uses nine labels to describe the four types of elements in C-S-H. Our convention for these nine labels is as follows: Ca (intralayer calcium), Cw (intralayer calcium), Si (silicon), Oh (oxygen in hydroxyl group), H (hydrogen in hydroxyl groups), Ob (bridging oxygen in silica chain), O (non-bridging oxygen in silica chain or calcium layers not in a hydroxyl group), Ow (oxygen in water molecules), Hw (hydrogen in water molecule). The partial charges of all species are provided in Table B.1. The partial charges of all species are constant except oxygen atoms. In fact, the partial charges of oxygen species in calcium-silicate layers are set to vary in order to ensure charge-neutrality within numerical samples. Since ReaxFF calculations are performed using electronegativity equalization method (EEM) [238] and the species of atoms are subjected to change as a result of chemical reactions, the unit cells might not be charge-balanced after reactions. A comprehensive sensitivity analysis on the elastic properties is carefully performed with different charge-balancing schemes. Due to the small charge-imbalance in the samples, charge-neutralization schemes do not strongly affect the elastic properties. In this work, the excess charge is equally distributed among O, Ob and Oh atoms. Interatomic Potentials In this work, we use two interatomic potentials: ReaxFF and CSH-FF. ReaxFF is a reactive force field originally developed on

the same notion as bond order potentials [331, 52, 10] and it is further enriched with polarizable charge model [238]. ReaxFF is a powerful tool for modeling formation and breakage of bonds during chemical reactions and modification of charge distribution during molecular dynamics (MD) simulation. In this force field, the total potential energy is written as a sum of various energetic contributions:

$$E_{tot}^{ReaxFF} = E_{Col} + E_{VdW} + E_{radial} + E_{angular} + E_{lp} + E_{OverC} + E_{UnnderC} \quad \text{(A.1)}$$

where E_{Col}, E_{VdW}, E_{radial}, $E_{angular}$ and E_{lp} are Coulomb, Van der Waals, radial bond stretching, angular bond bending and lone pair contributions to the potential energy, respectively. E_{OverC} and $E_{UnnderC}$ are defined as over- and under-coordination energy contributions to impose energetic penalties for high and low coordination numbers. Each of these terms depends on the local environment of atoms described via a bond order expression. In addition, a screened taper function is applied on the non-bonded Coulomb and Van der Waals interactions. In addition to previously defined Si-OH interactions [116], ReaxFF was further developed to include Ca-OH interactions. This was particularly done to incorporate calcium within ReaxFF framework. This facilitates the description of chemical reactions within C-S-H ultraconfined medium. For more detailed discussion on the ReaxFF potential, interested readers are referred to the work of Russo et al. [292].

The CSH-FF interatomic/intermolecular potential is a core-only model for hydrated calcium-silicates, in which atoms are described as interacting points through Coulombic and short-range potential functions. The total interatomic potential for CSH-FF potential reads:

$$E_{tot}^{CSHFF} = E_{Col} + E_{lj} + E_{radial} + E_{angular} \quad \text{(A.2)}$$

where E_{Col}, E_{lj}, E_{radial}, and $E_{angular}$ are the electrostatic, short-range Van der Waals in the form of Lennard-Jones potential, bond stretching and bond bending contributions to the potential energy, respectively. CSH-FF potential is a customized force field for hydrated calcium-silicates, and it is not transferable to other oxides with dif-

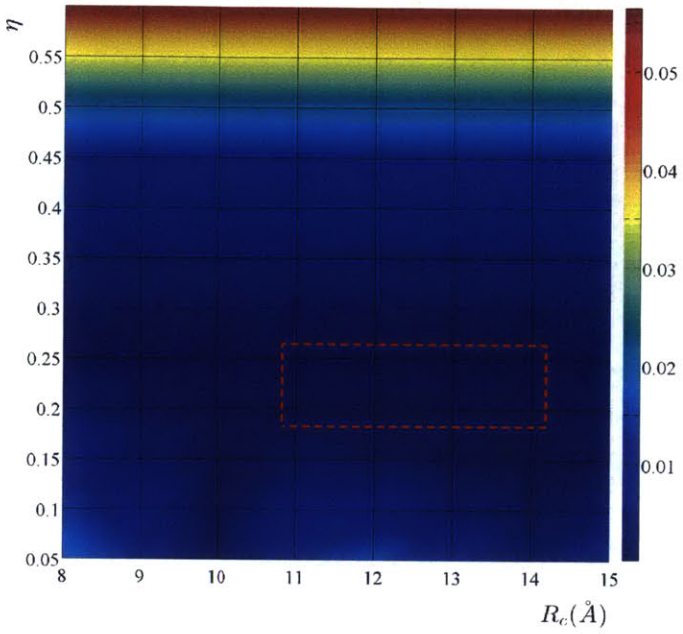

Figure A-2: Distribution of in $R_c - \eta$ space used for identification of optimal parameters of Wolf potential for C-S-H samples. The dashed rectangles located the region in which the normalized error is less than 10^{-4}. R_c and η are set to 12Å and 0.25 in all simulations, respectively.

ferent chemical environments. Therefore, CSH-FF can be seen as an improved version of the generic ClayFF potential [86] (originally developed for structural properties of clay minerals) for hydrated calcium-silicates such as cement hydrates. To describe the molecular structure of water and its interaction with other water molecules, both CSH-FF and Clay-FF potential models are built around the flexible SPC model for water. In particular, CSH-FF was designed to reproduce both structural and elastic data of Tobermorite minerals obtained from ab initio calculations [306]. This was achieved by adjusting a set of 29 potential parameters as described in Ref 15. The original form of CSH-FF was employed in the stage 1 through 4 of the model preparation where the hydroxyl groups were absent in the model.

In order to extend the predictability of the CSH-FF model, the latter was slightly modified to include a couple of new features. Primarily to reduce the computational costs incurred by the calculation of long-range forces, the columbic interactions are

calculated using Wolf method [357]. In Ewald summation scheme, the inverse distance in electrostatic interaction is rewritten as its Laplace transformation and then split into two moderately fast converging summations between real and reciprocal spaces. The distribution of the summation between reciprocal and real space is controlled by γ. The final Ewald expression reads:

$$E_{Col}^{Ewald} = \frac{2}{\pi V}\sum_{G}\frac{exp\left(\frac{-G^2}{4\gamma}\right)}{G^2}\sum_{i,j} q_i q_j exp\left(-iGr_{ij}\right) + \sum_{i,j}\frac{\left(q_i q_j erfc\left(\gamma^{1/2}r_{ij}\right)\right)}{2r_{ij}} \quad \text{(A.3)}$$

where q is charge, r is the interatomic distance, V is the volume of the crystal's unit cell and G is a vector of the reciprocal lattice. The erfc stands for complimentary error function in mathematics. The notion of Wolf summation is based on the idea of charge neutralization of a spherical subset of atoms in condensed ionic systems. Originally, Wolf proposed a two-body real space summation, which significantly reduces the computational cost of Ewald summation in real and reciprocal spaces. The Wolf summation reads:

$$E_{Col}^{Wolf} = \frac{1}{2}\sum_{i,j(r_{ij}<R_c)} q_i q_j \left(\frac{\text{erfc}\left(\gamma^{1/2}r_{ij}\right)}{r_{ij}} - \lim_{(r_{ij}\to R_c)}\left(\frac{\text{erfc}\left(\gamma^{1/2}r_{ij}\right)}{r_{ij}}\right)\right) - \left(\frac{\text{erfc}(\gamma^{1/2}r_{ij})}{2R_c} + \frac{\eta}{\pi^{1/2}}\right)\sum_{i} q_i^2 \quad \text{(A.4)}$$

where N is the number of atoms in the unit cell, q is the charge, r is the interatomic distance, R_c is the cut-off distance for which the neutralization technique is imposed and η is the damping factor. While Ewald summation scales with $O(N^{3/2})$, Wolf summation scales with $O(N)$. This is computationally advantageous, especially for extensive MD studies involving thousands of simulations. However, the success of the Wolf summation is tied to the proper identification of the Wolf's free parameters, R_c and η. To obtain the proper values for C-S-H's molecular structure, the absolute difference between the Ewald and Wolf summations is calculated for different values

of R_c and η. The normalized error is given as:

$$\lambda = \frac{|E_{Col}^{Ewald} - E_{Col}^{Wolf}|}{(E_{Col}^{Ewald}} \tag{A.5}$$

As it is shown in Fig. A-2, λ is quite negligible (in the order of 10-4) in the window of 11Å< R_c <14Å and 0.2< η <0.25. To apply a uniform cut-off radius in the simulation box for both long- and short-range interactions, R_c is set to 12Å. The damping factor, η, is set to 0.25 in all simulations. The second modification in the CSH-FF potential pertains to incorporation of hydroxyl groups. The force field parameters for hydroxyl groups are included in the extended form of the CSH-FF potential.

Table A.2: ϵ and σ for all species in modified C-S-H-FF potential.

$spec_i$	$spec_j$	ϵ (Kcal/mol)	σ (Å)	$spec_i$	$spec_j$	ϵ (Kcal/mol)	σ (Å)
O	Ca	8.694E-04	4.365	Oh	Oh	6.180E-02	3.448
Ca	Ob	7.057E-04	5.452	O	Ob	4.543E-02	3.626
Ca	Oh	8.694E-04	4.365	O	Oh	4.543E-02	3.626
O	Cw	1.460E-03	4.365	Oh	Ob	6.180E-02	3.448
Cw	Ob	1.040E-03	4.446	Ca	Ow	8.763E-04	4.365
Cw	Oh	1.040E-03	4.446	Cw	Ow	6.042E-04	4.472
O	Si	5.604E-04	3.270	O	Ow	5.258E-03	4.241
Si	Ob	5.950E-04	3.261	Ow	Ob	8.717E-01	2.895
Si	Oh	5.950E-04	3.261	Ow	Oh	8.717E-01	2.895
O	O	1.243E+00	2.735	Ow	Ow	1.540E-01	3.163
Ob	Ob	6.180E-02	3.448				

The short-range interactions in CSH-FF and ClayFF force fields are described via Lennard-Jones potential:

$$E_{lj} = 4\epsilon \sum_i \sum_{j>i} \left[\left(\frac{\sigma}{r_{ij}} \right)^{12} - \left(\frac{\sigma}{r_{ij}} \right)^{6} \right] \tag{A.6}$$

where σ and ϵ have the dimensions of distance and energy, respectively. These parameters are provided for short-range interactions between pairs of elements in the model. The potential parameters are provided in Table B.2. Similar to ClayFF potential, the hydrogen species do not have short-range interactions.

The radial and angular excursions in water molecules and hydroxyl groups are

Table A.3: radial stiffness and equilibrim bond length in O-H bonds.

$spec_i$	$spec_j$	k_r (Kcal/mol/ ^{2})	r_0 (Å)
Hw	Ow	554.135	1
H	O	554.135	1

Table A.4: angular stiffness and equilibrim angle in H_2O molecules.

$spec_i$	$spec_j$	$spec_k$	k_θ (Kcal/mol/θ^2)	θ_0
Hw	Ow	Hw	45.770	109.47

described via simple harmonic potentials:

$$E_{radial} = \sum_{i=1}^{N_{OH}} \frac{1}{2} k_r \left(r_{ij} - r_0\right)^2 \tag{A.7}$$

$$E_{angular} = \sum_{i=1}^{N_{H_2O}} \frac{1}{2} k_\theta \left(\theta_{ijk} - \theta_0\right)^2 \tag{A.8}$$

where k_r and k_θ are radial and angular stiffness, respectively. Parameters r_0 and θ_0 denote the equilibrium hydroxyl bond length and Hw-Ow-Hw angle in water molecules, respectively. N_{OH} and N_{H_2O} are the total number of O-H bonds in hydroxyl groups and water molecules and total number of water molecules, respectively. The radial and angular constants are provided in the Table A.3 and A.4.

A.3 Generating different C/S

In this work, the C-S-H models describe cement hydrates as defective calcium-silicates with short- to medium-range order. These models are suggestive of Tobermorite layer-to-layer long-range correlations. In contrast to Tobermorite polymorphs, cement hydrates have a different stoichiometry with significantly larger calcium (and water) content: C/S in Tobermorite is unity or smaller while it is around 1.7 for ordinary Portland cement hydrates9. This large C/S can be achieved by engineering the silica chains lengths (these chains are infinite in the case of Tobermorite minerals). In other words, the C-S-H model of cement hydrates contains many defects and also a larger water content compared to that of Tobermorite polymorphs. Silica chain

defects in the C-S-H models were implemented by randomly removing SiO_2 groups from an anhydrous 11Å Tobermorite crystal so the targeted C/S was obtained. This is precisely the strategy that was deployed in the present work aiming at exploring the effect of varying this C/S on texture and mechanical properties. It is fairly well known that in real cement paste the C/S spans majorly in the range of 1 to 2.1.

In this section, we implemented the above strategy to impose a given chemistry in terms of C/S. As for all the C-S-H models, we started from a periodic simulation box containing $2 \times 3 \times 1$ unit cells of Hamid's structure of 11Å Tobermorite [139]. At this stage, we constructed an anhydrous version of this super-cell as we were only interested in creating the solid backbone of the model itself; the addition of water was carried out at the fourth stage using Grand Canonical Monte Carlo simulation utilized to describe adsorption/desorption processes [245]. The simulation boxes are large enough to accommodate silica dimers, pentamers and even octamers as these types of silica chain are known to be dominantly present in real cement paste. One can map out each initial infinite silica chain onto a line that contains "1" for pairing tetrahedral and "2" for the bridging ones. In Tobermorite, bridging and pairing silica tetrahedrals form the so-called Drierketten pattern that in this simple notation writes "112". Thus within our periodic simulation box, silica structure of Tobermorite writes "112112112". Juxtaposing all the eight chains (2 per side of each layer, two layers), we can assemble the following matrix for the abovementioned 11ÅTobermorite super-cell:

$$\begin{pmatrix} 2 & 1 & 1 & 2 & 1 & 1 & 2 & 1 & 1 \\ 2 & 1 & 1 & 2 & 1 & 1 & 2 & 1 & 1 \\ 1 & 2 & 1 & 1 & 2 & 1 & 1 & 2 & 1 \\ 1 & 2 & 1 & 1 & 2 & 1 & 1 & 2 & 1 \\ 2 & 1 & 1 & 2 & 1 & 1 & 2 & 1 & 1 \\ 2 & 1 & 1 & 2 & 1 & 1 & 2 & 1 & 1 \\ 1 & 2 & 1 & 1 & 2 & 1 & 1 & 2 & 1 \\ 1 & 2 & 1 & 1 & 2 & 1 & 1 & 2 & 1 \end{pmatrix} \tag{A.9}$$

Note that the above super-cell contains 72 calcium and silicon atoms (C/S=1). To

construct a model with a give C/S, we first developed a simple computer program that uses a random number generator to replace some elements of the above matrix by "0" with the constraint of matching a given C/S (within ± 0.1) and producing as much dimers ("0110") pentamers ("0112110") and octamers ("0112112110") as possible. For instance, the "cutting" matrix for C/S=1.30 can be written as:

$$\begin{pmatrix} 0 & 1 & 1 & 0 & 1 & 1 & 0 & 1 & 1 \\ 2 & 1 & 1 & 2 & 1 & 1 & 2 & 1 & 1 \\ 1 & 2 & 1 & 1 & 0 & 1 & 1 & 2 & 1 \\ 1 & 0 & 1 & 1 & 2 & 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 & 1 & 1 & 0 & 1 & 1 \\ 2 & 1 & 1 & 2 & 1 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 & 1 & 1 & 2 & 1 \\ 1 & 2 & 1 & 1 & 0 & 0 & 1 & 0 & 1 \end{pmatrix} \tag{A.10}$$

The second step is to realize the combinatorial aspect of the approach, as there are many ways to "cut" for a given stoichiometry. Here, we considered 10 to 15 different samples for each C/S value. Second, each "cutting matrix" is projected onto our simulation box. Each "0" corresponds to a silicon atom that has to be removed along with two of its neighboring oxygen atoms. Since SiO_2 is a charge-neutral group, the unit-cells electro-neutrality is maintained. We have made sure that silicon in ending sites (Q_1 in solid state NMR nomenclature) had the necessary four oxygen neighbors. The preparation of monomers in C/S>1.5 is only plausible if the pairing sites are removed from the silica chains. In fact, upon removal of all the bridging silica sites from the initial crystalline structure, the C/S ratio is found to be 1.5. Therefore, some monomers, Q_0, inevitably exist in the initial solid skeleton of C-S-H samples. However, the percentage of monomers in the samples is reduced via reactive modeling through which condensation of silicate chains is energetically plausible.

A.4 Grand Canonical Monte-Carlo technique for water adsorption

In this work, we first produced an anhydrous version of C-S-H models by removing all the interlayer water. We subsequently calculated the maximum amount of water that can be accommodated in its pore voids via the Grand Canonical Monte-Carlo (GCMC) simulation technique. GCMC is a well-suited technique to study adsorption/desorption processes in nanoporous materials. GCMC determines the properties of a system at a constant volume V (the pore with the adsorbed phase) in equilibrium with an infinite fictitious reservoir of particles imposing its chemical potential μ and its temperature T 27. For different values of Îij, the absolute adsorption isotherm can be determined as an ensemble average of the adsorbed atom numbers in the system versus the pressure of the gas reservoir P (the latter can be obtained from the chemical potential according to the equation of state for the bulk gas). The adsorption and desorption processes can be respectively simulated by increasing or decreasing the chemical potential of the reservoir; the final configuration of a simulation is the initial state for the next point. Periodic boundary conditions were used in all directions of space as for the energy minimization procedure. An equal number attempt for translation, rotation, creation or destruction of molecules has been chosen. The isotherm has been calculated for 300 K. Acknowledging the very restricted available space in between Tobermorite layers, one should not expect capillary condensation to occur by contrast to larger pore systems such as Vycor [271, 272] or in the larger inter C-S-H grain pore space [47]. In our case, the adsorption/desorption process is expected to be close to that observed for microporous zeolite as far as water adsorption is concerned [272]. We did not calculate the entire water adsorption/desorption isotherm but perform a single GCMC simulation with the water chemical potential fixed to a value that corresponds to the bulk liquid phase with a density of 1 g/cm3 at room temperature (μ=0 eV for flexible SPC water model). CSH-FF potential was used in all GCMC calculations via Gulp [124, 123].

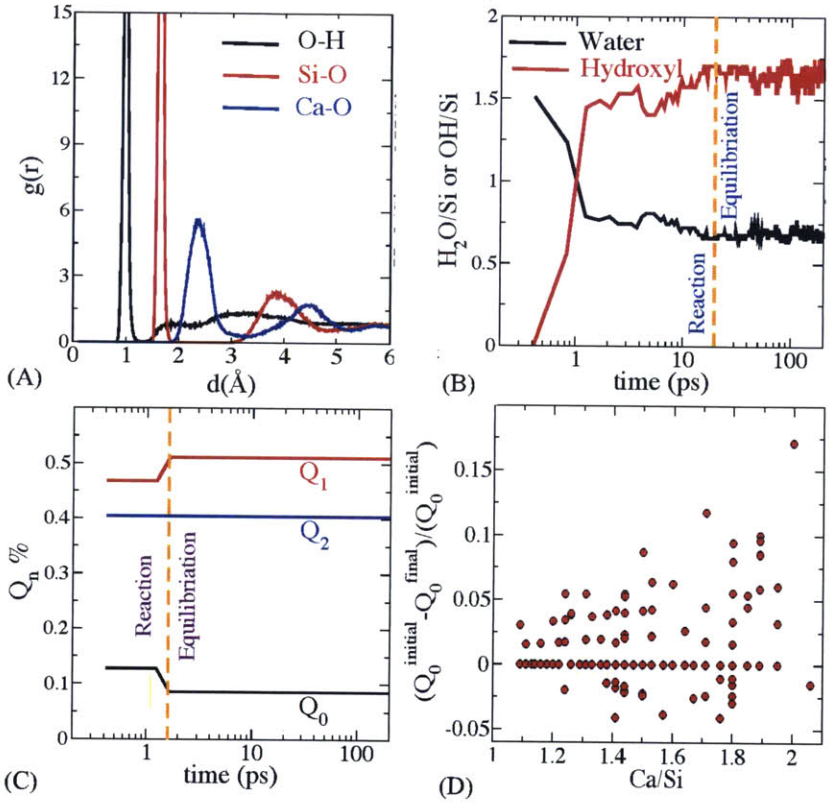

Figure A-3: Reactive force field modeling in canonical ensemble. (A) Radial distribution function for Si-O, Ca-O and O-H bonds. (B) Dissociation of water molecules in C-S-H and production of hydroxyl groups. (C) Condensation of silica groups in the silicate chain. (D) Condensation of silicate chains as a function of Ca/Si ratio.

A.5 Reactive and Non-Reactive MD in the NPT and NVT ensembles

All reactive simulations are performed using LAMMPS [267]. We employ the velocity Verlet algorithm for our time integration schemes. In order to appropriately describe chemical reactions, the time step is set to 0.1 fs. To further trigger chemical reactions, ReaxFF simulations are carried at 500 K to facilitate the reaction of water molecules with defective calcium-silicate layers. According to Arrhenius law, this assists the water molecules to overcome the reaction energy barriers. This reduces the computational expense of reactive modeling by reducing the required simulation time to reach equilibrium. Initially, all samples are relaxed in canonical ensemble (NVT) for at 200 ps using Nose-Hoover thermostat. Subsequently, the samples are further relaxed in isothermal-isobaric ensemble (NPT) at the 500 K and 1atm using Parrinello-Rahman method [257]. The overall performance of a C-S-H sample with Ca/Si = 1.5 is presented in Fig. A-3.

Fig. A-3.A provides the radial distribution function, RDF or $g(r)$, of Si-O, Ca-O and O-H bonds. Since the peaks in $g(r)$ are intact, the slightly elevated temperature has not affected the solid backbone of C-S-H. It should be also noted that while Si-O and O-H RDFs have a very defined first sharp peak, the Ca-O RDF exhibits a wider peak. This means that the radial excursions of Si-O and H-O bonds are very limited. However, Ca-O bond has a much larger radial excursion. This means that Ca-O bond is not as stiff as Si-O and H-O bonds. As it can be seen in Fig. A-3.B, in the very first few picoseconds of the simulation, a large portion of water molecules dissociates to form OH^- and H^+ ionic pairs. While the hydroxyl groups coordinate with Ca^{2+} ions, protons react with non-bridging oxygen atoms in silicate groups. Non-bridging oxygen atoms are only attached to one silicon atom in the silicate chain. No hydrogen atom was found coordinated to the bridging oxygen shared between adjacent silicate groups. In fact, over-coordination energy penalty makes such a reaction nearly energetically implausible. This hydrophobic behavior was originally seen in siloxane oxygen atoms [83] and further confirmed for C-S-

H [209]. The dissociation mechanism affects the large portion of interlayer water molecules in the C-S-H and the dissociation percentage depends on the C/S which is directly proportional to the number of non-bridging oxygen atoms in silicate chains. Reactive modeling also affects the solid skeleton of C-S-H.

Figure A-3.C shows the effect of reactivity on condensation of silicate groups. Unlike the recent work of Manzano et al. [211], it is observed that silicate groups condensate in a number of C-S-H models. In the sample presented in Fig. A-3, a pair of monomers has condensed to form a dimer. This is in agreement with energetic arguments provided by DFT calculations of isolated silicate groups by Manzano et al. [208] showing dimers have high chemical stability index. The observed condensation mechanism is triggered by two silicate tetrahedra pairing through which a dimer is produced and an oxygen atom is released. The free oxygen pairs with a proton and the resultant hydroxyl group coordinate with a Cw^{2+} cation to complete the reaction path. Fig. A-3.D presents the difference in percentage of silica monomers before and after reactive modeling for all the C-S-H models. It is observed that some samples are intact, in some condensation occurs and dissociations of longer instable chains increases monomer content slightly in a few samples. As shown in Fig. A-3.C, the condensation occurs in the very first couple of picoseconds of simulation. This signifies that due to random "cutting" of chains, some monomers have extremely unrealistic configuration. This unrealistic state is relaxed via polymerization of monomers producing energetically more stable dimers. The energy barriers for the condensation of silica chains are large, in the order of a few eVs, and the relevant simulation time required to capture such a rare condensation events exponentially increases according to the well-known transition state theory. The methods and approaches to overcome the barriers in energy landscape and subsequent time marching strategies such as Activation-Relaxation Technique (ART) [28] or Autonomous Basin Climbing (ABC) [188] are beyond the scope of this thesis.

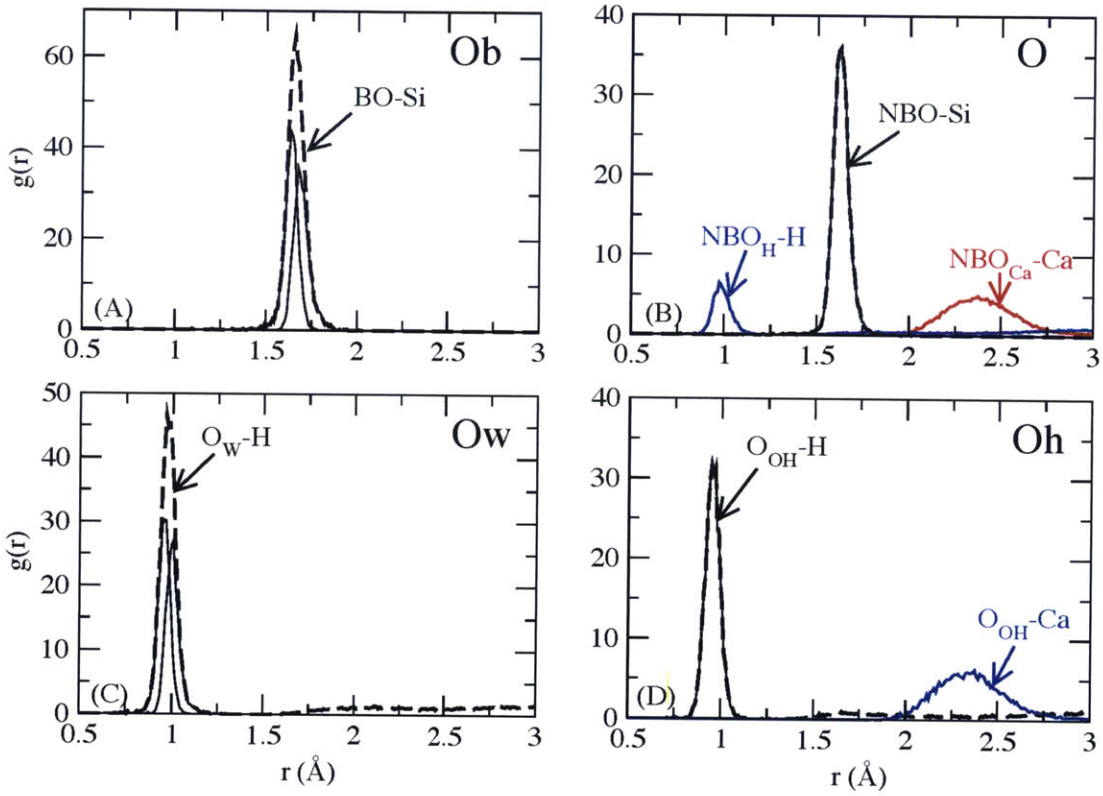

Figure A-4: Determination of the local environment of oxygen. Radial distribution functions of Si, Ca, and H cations around Ob (A), O (B), Ow (C) and Oh (D). Figures (A) and (B) show the contributions of the two nearest Si neighbors and figures (C) and (D) the contributions of the two nearest H neighbors.

A.6 Topological Analysis

ReaxFF simulations are computationally expensive to undertake. In order to overcome the prohibiting computational cost of reactive modeling, required for extensive mechanical testing on each individual sample, all samples are transferred to non-reactive environment after completion of reactions. As the direct consequence of reactive modeling, the local environment of atoms undergoes drastic changes. This necessitates a precise characterization of local environment of each single atom. Such a precise analysis can be performed using the fundamental ideas of topological constraint analysis [31]. In C-S-H's molecular structure, Si, Ca and H cations are always connected to O atoms, the latter showing different local environment (Fig. A-4). In order to differentiate between these species, we performed a careful topological analysis of the coordination of each O atom. Bridging oxygen atoms (Ob) are thus defined as oxygen atoms showing exactly two Si neighbors in their O-Si first coordination shell (FCS), which limit has been fixed as the first minimum of the g_{O-Si} RDF, typically 2.5Å. On the contrary, non-bridging oxygen atoms (O) show only one Si in their O-Si FCS. Those species arise from the depolymerisation of the silicate network by Ca and H cations and can belong to hydroxyl groups O-H (Oh) if they show an H neighbor in their O-H FCS (minimum of the gO-H RDF, typically 1.3Å) or are connected to Ca atoms otherwise. Remaining O atoms that are not part of the silicate network have then been respectively differentiated into oxygen in water molecules (Ow) and hydroxyl group (Oh) when they show two and one hydrogen atom in their O-H FCS, respectively.

A.7 Simulated Annealing

Following the topological analysis, all samples are transferred to the non-reactive environment of CSH-FF potential. Since the conditions at the reactive simulations were performed at elevated temperature, simulated annealing technique (SAT) was used to lower the temperature to that of ambient conditions (Fig. A-5). During SAT, each

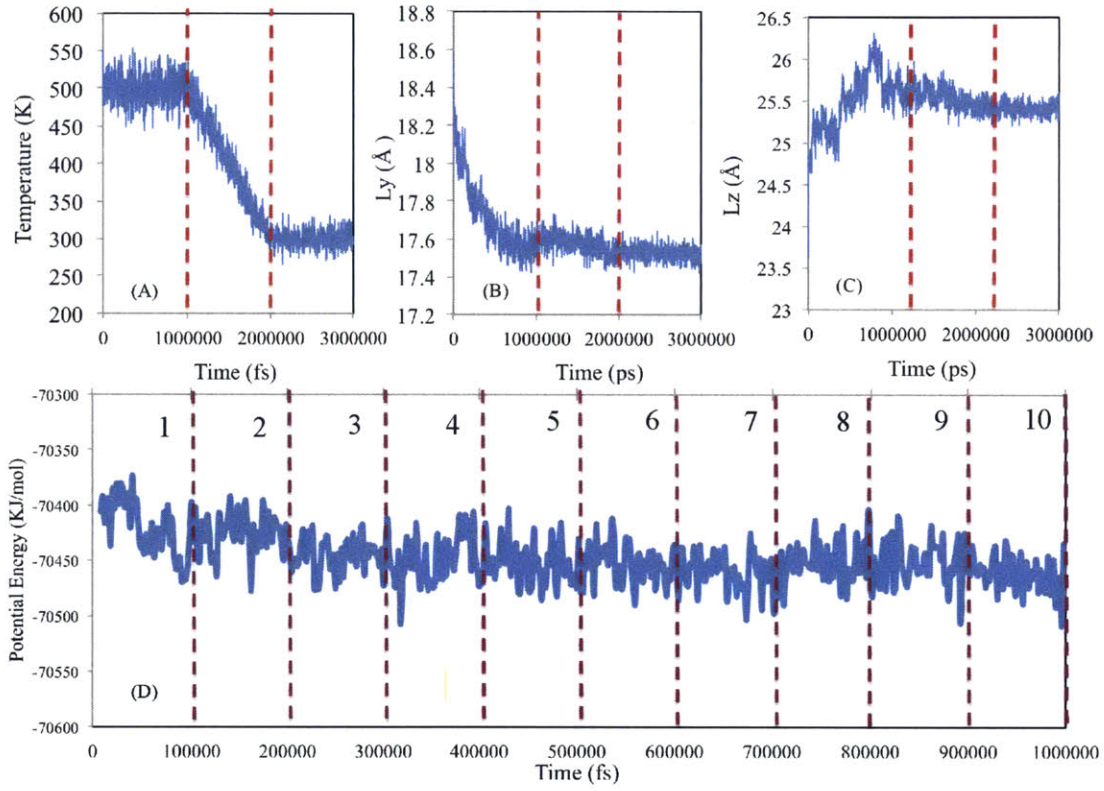

Figure A-5: Simulated annealing and sampling process from the energy landscape. (A) The temperature profile in annealing process. (B) and (C) the simulation box size during the quenching period. (E) Sampling from the energy landscape for estimation of mechanical properties.

sample is relaxed for 1 ns at 500 K using CSH-FF potential in isobaric-isothermal ensemble with time steps of 1 fs. All samples are subsequently quenched over 1 ns to 300K using 100 annealing steps. Finally, each sample is further relaxed in NPT ensemble for another 1 ns at 300K. The temperature profile during the quenching stage is provided in Fig. A-5.A. The simulation box sizes were tracked to ensure proper convergence of SAT (Fig. A-5.B and A-5.C). We note that the most of box length adjustments occur during the model relaxation at 500 K and not during the quenching stage. These slight adjustments are mainly due to the discrepancies between the ReaxFF and CSH-FF potentials. However, as shown in Fig. A-5.A-C, these differences are negligible.

A.8 Calculation of Elastic Properties

After the simulated annealing stage, all models are further relaxed for 1 ns. Next, 10 independent frames each 100 ps apart are consecutively extracted from MD trajectories. This procedure is strictly followed to ensure proper unbiased statistics for each sample. Afterwards, all samples are relaxed via energy minimization at constant volume (CONV) [124, 123] with rational functional optimization (RFO) [124, 123] module for residuals less than 10^{-3}. This ensures that the Hessian matrix is positive definite, i.e. all the eigenvalues of the Hessian matrix are positive. All the relaxed instances of all C-S-H models are further relaxed via energy minimization at constant pressure (CONP) [124, 123] with RFO option. Through this multi-step scheme, it is made sure that all the available degrees of freedom, both the atomic positions and the unit cell, are properly relaxed. Subsequently, the elastic properties are calculated for each of 1500 abovementioned frames. The stiffness tensor, C_{ijkl} in tensor notation and C_{ij} in Voigt notation, is subsequently calculated from the second derivative of the energy density.

$$C_{ijkl} = \frac{1}{V}\left(\frac{\partial^2 E}{\partial \epsilon_{ij} \partial \epsilon_{kl}}\right) \tag{A.11}$$

where E and V are internal energy and volume of the unit cell, respectively. ϵ_{ij} denotes the strain tensor. It should be noted that positive-definiteness of Hessian

matrix entails positive-definiteness of the stiffness tensor. A wide range of elastic constants can be calculated from stiffness and compliance ($S = C^{-1}$) tensors. Two of the commonly used isotropic measures of the stiffness are bulk (K) and shear (G) moduli. The Voight-Reuss-Hill bounds of bulk and shear modulus are given by:

$$K_{Voigt} = \frac{1}{9}\left(C_{11} + C_{22} + C_{33} + 2\left(C_{12} + C_{13} + C_{23}\right)\right) \tag{A.12}$$

$$K_{Reuss} = \left(S_{11} + S_{22} + S_{33} + 2\left(S_{12} + S_{13} + S_{23}\right)\right) \tag{A.13}$$

$$G_{Voigt} = \frac{1}{15}\left(C_{11} + C_{22} + C_{33} + 3\left(C_{44} + C_{55} + C_{66}\right) - C_{12} - C_{13} - C_{23}\right) \tag{A.14}$$

$$G_{Reuss} = \frac{15}{4\left(S_{11} + S_{22} + S_{33} - S_{12} - S_{13} - S_{23}\right) + 3\left(S_{44} + S_{55} + S_{66}\right)} \tag{A.15}$$

The Hill bounds are simply the average of Voigt and Reuss bounds, which are called Voight-Reuss-Hill bounds31. The indentation modulus (M) can be directly calculated from bulk and shear modulus:

$$M = 4G\frac{3K + G}{3K + 4G} \tag{A.16}$$

In fact these bounds are isotropic descriptions of a generally anisotropic medium. A wide range of materials in nature including clay [101] and Tobermorite minerals [9] exhibit anisotropic behavior due to their molecular structure.

There are numerous approaches to quantify anisotropy at nanoscale. The stiffness of materials under the uniaxial tension is called Young Modulus. In anisotropic materials, the Young Modulus in different directions is different in uniaxial deformation path ($Y_i = d\sigma_{ii}/d\epsilon_{ii}$). Another method to assess the level of anisotropy in C-S-H is through the measurement of orthotropic indentation modulus in the plane and perpendicular to the calcium-silicate layers. Following the derivation of Delafargue and Ulm [93]:

$$M_1 = \sqrt{\frac{C_{11}C_{33} - C_{13}^2}{C_{11}}\left(\frac{1}{C_{44}} + \frac{2}{\sqrt{C_{11}C_{33}} + C_{13}}\right)^{-1}} \tag{A.17}$$

$$M_3 = \sqrt{(\sqrt{\frac{C_{11}}{C_{33}}}\frac{C_{11}^2 - C_{22}^2}{C_{11}}M_1} \tag{A.18}$$

One can think of $\frac{|M_1 - M_3|}{M_1}$ as a measure of anisotropy in the structure of materials. There are many other ways to assess the level of anisotropy in materials such as Thompson index based on anisotropy in acoustic velocities of materials [251]. In this work, C^{iso} is mathematically defined by enforcing isotropic conditions on a general stiffness tensor:

$$C_{11}^{iso} = C_{22}^{iso} = C_{33}^{iso} = \frac{C_{11} + C_{22} + C_{33}}{3} \tag{A.19}$$

$$C_{44}^{iso} = C_{55}^{iso} = C_{66}^{iso} = \frac{C_{44} + C_{55} + C_{66}}{3} \tag{A.20}$$

$$C_{12}^{iso} = C_{12}^{iso} = C_{23}^{iso} = C_{11}^{iso} - 2C_{44}^{iso} \tag{A.21}$$

The rest of the entries in the C^{iso} are set to zero. A new index for anisotropy is defined using Euclidean, dE, and Riemannian, dR, Norms of a general stiffness tensor from its isotropic part, C^{iso}, defined as:

$$d_E(C^{iso}, C) = ||C^{iso}, C||_E = \sqrt{tr\left((C^{iso} - C)^T(C^{iso} - C)\right)} \tag{A.22}$$

$$d_R(C^{iso}, C) = ||ln\left(C^{1/2}\left(C^{iso}\right)^(- 1)C^{1/2}\right)||_E = \sqrt{\sum_{i=1}^{n} ln^2\lambda_i} \tag{A.23}$$

where λ denotes set of eigenvalues and $tr()$ stands for the trace of the matrix.

A.9 Hardness Prediction via Biaxial Deformation

In this work, a new method is designed for assessment of hardness at nano-scale. The notion behind this approach is originated from laboratory tri-axial testing of materials. Very similar to tri-axial testing, the atomistic simulation cell can be isotropically deformed in different deformation paths. Due to the application of periodic boundary conditions, the simulation cell is representative of the bulk of the material. There are six independent deformation paths to deform a material, which consists of three nor-

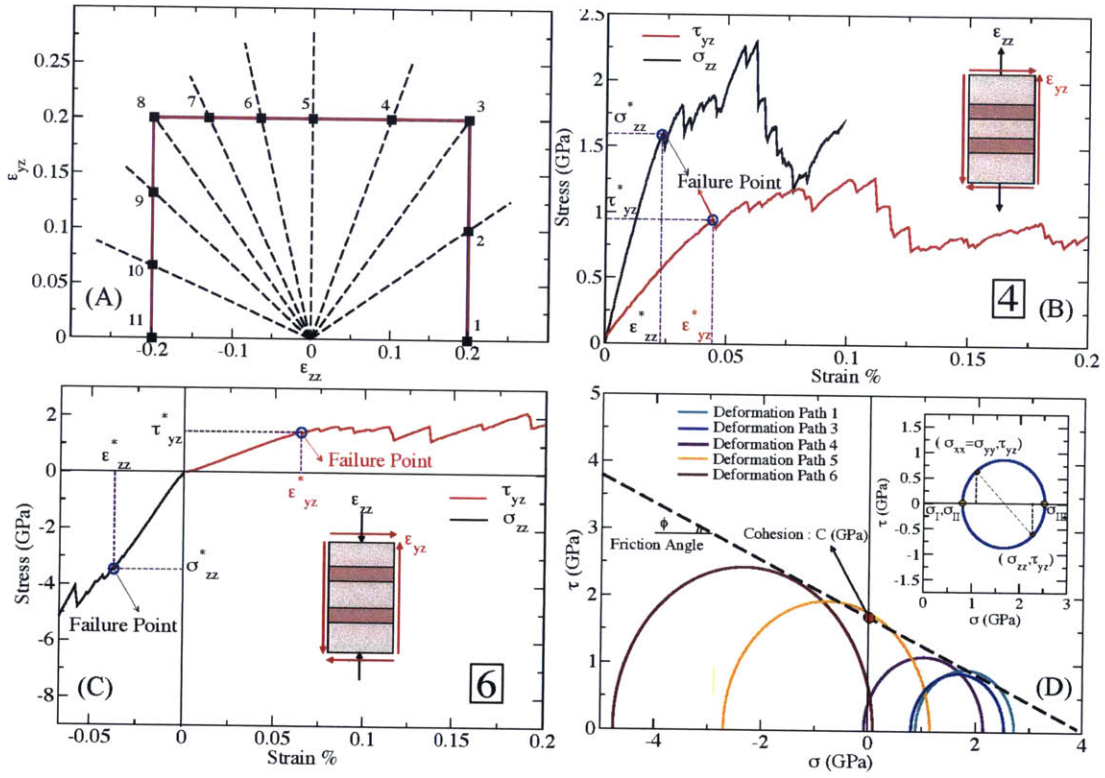

Figure A-6: calculation of hardness in atomistic simulation. (A) 11 independent deformation paths to deform C-S-H sample. Biaxial strain-control deformation analysis of C-S-H in (B) deformation path #4 and (C) deformation path #6. (D) Calculation of cohesion and friction angle from failure analysis using Mohr-Coulomb criterion.

mal, and three shear paths [205]. In continuum mechanics, the deformation gradient tensor, F, describes the state of deformation in the deformed configuration compared to the undeformed body as:

$$dx = F_{ij} dX \tag{A.24}$$

where dX and dx denote the infinitesimal vectors in undeformed and deformed configurations, respectively. The Green-Lagrange strain tensor is a rotation-independent measure of local relative deformation defined as

$$\epsilon = \frac{1}{2} \left(F_{ki} F_{kj} - \delta_{ij} \right) \tag{A.25}$$

where δ_{ij} is the Kronecker delta. Following the assumption of Cauchy-Born, it is assumed that the unit cell of the material at the atomic scale deforms with the macroscopic deformation gradient [13, 172, 171, 273]. In other words, the deformation at the molecular scale can be stated as:

$$a_i = F_{ij} A_j \tag{A.26}$$

where A and a denote the lattice vectors of the original undeformed lattice and the deformed structure, respectively. This entails an approximation due to the finite size of the lattice. In fact, Cauchy-Born rule implies that the macroscopic deformation gradient minimizes the free energy, Ψ, of the lattice at atomic-scale. In solid mechanics, the Cauchy stress is defined as:

$$\sigma_{ij}^{continuum} = \frac{d\Psi}{d\epsilon_{ij}} \tag{A.27}$$

The above formulation defines the stress tensor in a hyperplastic sense for energy-conserving solid undergoing either linearly or non-linearly elastic deformation. As defined in thermodynamics, the free energy, $\Psi = E - TS$, is the sum of molecular interaction and entropic contributions where T and S are temperature and entropy of the system. At considerably low temperatures (below Debye temperature), both

configurational and vibrational entropy can be assumed to be negligible. Therefore, the stress can be measured as the derivative of internal energy with respect to strain. In molecular mechanics, the internal energy can be stated as a function describing interaction between particles. This gives rise to a semi-analytical formulation for stress tensor at the atomic scale known as Virial stress:

$$\sigma_{ij}^{Virial} = \left\langle \frac{1}{\Omega} \sum_{k} \sum_{l} \left(r_{kl}^{i} f_{kl}^{j} \right) \right\rangle \tag{A.28}$$

where Ω is the unit cell volume, r_{kl}^{i} and f_{kl}^{j} is the distance and pair-wise force between the pair of k^{th} and l^{th} atom in the i^{th} direction. The <> sign represents the ensemble average which is extensible to the time average under the ergodic assumption. Considering kinematic equations, eqs. A.24 and A.25, along with Virial stress , eq. A.28, the stress-strain relationship at the scale of unit cell can be described via atomistic simulation. The two-body form of Wolf summation makes it possible to calculate stress for ionic systems. The concise theory presented above builds the foundation for measuring hardness at the atomic-scale.

Due to transversely isotropic nature of C-S-H, especially at low Ca/Si ratios, the direction perpendicular to the calcium silicate layers, z direction, shows lower strength. This is directly attributed to the lack of covalent bonding, Q_3 and Q_4 sites, bridging between the layers. This suggests that the failure analysis should be conducted in ϵ_{zz} and the two other relevant shear strains ϵ_{xz} and ϵ_{yz}. The response of the C-S-H against the two shear paths is almost identical as they both describe shear localization at the interlayer spacing. Therefore, the six dimensional strain spaces reduced effectively to a two dimensional space. This considerably reduces the number of computations associated with calculation of hardness. Subsequently as displayed in Fig. A-6.A, eleven independent deformation paths are considered in $\epsilon_{zz} - \epsilon_{yz}$ space. Each line is a linear combination of ϵ_{zz} and ϵ_{yz}. The normal strain, ϵ_{zz}, covers the range of -0.2 to 0.2, which includes both tensile and compressive modes. In fact, mutual application of shear deformation along with tensile or compressive deformation resonates the confinement effect at the atomic-scale. This mechanism is

at a level below the scale of friction and wear, describing how solid bodies are rubbed against each other.

Meanwhile the processes of 1 ns relaxation during quenching process, all 150 C-S-H models are sampled at six different frames in the phase space. Each frame is at least 100 ps apart from the other frames making sure that the frames are statistically independent making a total of 900 samples. Each sample is subsequently deformed following the above 11 deformation paths in a strain-control fashion. To eliminate the rate effect, all simulations are performed using energy minimization approach. An extensive convergence analysis is performed to identify the size of the strain increments. It is found that strain steps as large as 10-3 are somehow large enforcing the stress path to deviate from the equilibrium deformation path. This leads to improper identification of the failure point. Therefore, the strain increment in all simulations is set to 10^{-4}.

The stress-strain response is plotted accordingly for each individual deformation test. Fig. A-6.A-B presents the stress-strain diagram in deformation path 4 and 6, respectively. Fig. A-6.A-B are instances of C-S-H under biaxial tensile-shear and compressive-shear mixed-mode failure analysis, respectively. Both tensile and shear modes exhibit linear elastic response before the failure point. As indicated by circles locating the failure point, upon the irreversible deformation at the atomic scale, the stress relaxes in both deformation modes simultaneously. This is an indicative of plastic deformation, which can be verified by unloading the sample before and after the yielding point and identifying the extent of residual strain by unloading the sample to its original state. As illustrated in Fig. A-6.D, the failure analysis identifies the strain, $(\epsilon^*_{zz}, \epsilon^*_{yz})$, and stress $(\sigma^*_{zz}, \sigma^*_{yz})$ at which the material fails upon loading in a particular deformation path. Identification of failure is certainly subjected to user-bias if performed manually. To eliminate such an effect from the simulation results, a simple conditional statement is implemented such that if the slope of stress-strain curve falls below a certain threshold, that point is taken as the failure limit. Since the deformation is confined in the other axial directions and because of Poisson effect, a planar stress evolves in the material while undergoing constrained biaxial

deformation. The planar stress is direction-independent, $\sigma^*_{xx} = \sigma^*_{yy} = \sigma^*_{pl}$, due to isotropy in calcium-silicate sheets. Therefore, the stress tensor at failure point, σ^*, can be viewed as:

$$\sigma^* = \begin{pmatrix} \sigma^*_{pl} & 0 & 0 \\ 0 & \sigma^*_{pl} & \sigma^*_{yz} \\ 0 & \sigma^*_{yz} & \sigma^*_{zz} \end{pmatrix} \tag{A.29}$$

The failure stress tensor can be stated in terms of principle failure stress components ($\sigma^*_I = \sigma^*_{II} = \sigma^*_{III}$) by simply finding the eigenvalues of failure stress tensor. Therefore, the principle stresses can be written as:

$$\sigma^*_I = \sigma^*_{pl} \tag{A.30}$$

$$\sigma^*_{II} = \frac{1}{2}\left(-\sqrt{\left(\sigma^*_{pl} - \sigma^*_{zz}\right)^2 + 4\left(\sigma^*_{pl}\right)^2} + \sigma^*_{pl} - \sigma^*_{zz}\right) \tag{A.31}$$

$$\sigma^*_{III} = \frac{1}{2}\left(\sqrt{\left(\sigma^*_{pl} - \sigma^*_{zz}\right)^2 + 4\left(\sigma^*_{pl}\right)^2} + \sigma^*_{pl} - \sigma^*_{zz}\right) \tag{A.32}$$

The transformation of normal and shear stresses to the principle stress components is graphically represented in the inset of Fig. A-6.D. Using the Mohr representation, the principle components of the stress can be identified as the intersection of Mohr circles with σ axis. The maximum and minimum of principle stresses for k^{th} deformation path are σ^k_I and σ^k_{II}, respectively. These two principle stresses and enclosing Mohr circles define the limit stress state at which atomic structures of C-S-H undergoes an irreversible deformation. As shown in Fig. A-6.D, a set of Mohr failure circles on independent deformation paths can define the failure mechanism and the plastic limit of the material. Fig. A-6.D denotes that C-S-H at the atomic scale follows a Mohr-Coulomb-like failure criterion. Application of Mohr-Coulomb in mechanics of granular materials is prevalent due to proper description of inter-granular friction. However, the increase in strength of C-S-H at compressive domain is attributed to the confinement effect. There is no dissipative mechanism such as friction or wear at the molecular scale and the increase in the strength is mainly because of elastic confinement effect making the failure harder for materials at high pressures. Particularly,

in the case of C-S-H, the failure envelope is linear in the stress domains related to nano-indentation of cementitious materials. Therefore, C-S-H at the nano-scale can be described as a cohesive-frictional material identified via cohesion (C) constant and friction angle (ϕ). Cohesion is the strength of the C-S-H at zero confinement pressure. Following the work of Ganneau et al. [126], the hardness (H) of a cohesive-frictional material can be stated as:

$$\frac{H}{C} = \frac{\delta(\phi,\theta)}{\tan\phi} = \frac{1}{\tan\phi}\sum_{m=1}^{N}(a_k(\theta)\ \tan\phi)^k \tag{A.33}$$

where θ is the indenter apex angle and a_k coefficients are fitting parameter for given indenter geometry. For the special case of C-S-H materials, the friction angle is on the order of 6-9 degrees in which $\tan(\phi)$ is quite negligible and as a result, the hardness can be approximated with 5.8C. In current work, friction angle, cohesion and hardness were independently measured at 6 independent points for each sample. The reported hardness and the error bars are the mean and standard deviations for each sample.

A.10 Free Energy Calculations

Due to vibrational entropy at finite temperature in solids, the atoms constantly oscillate around their equilibrium position. At low temperatures, these oscillations can be decomposed to a linear superposition of vibrational modes known as 'phonons'. In the case of infinite 3D solids, there are infinite numbers of phonons. The phonons are characterized by their values at a point in the reciprocal space, usually called k-point, within the first Brillouin zone. Hence, there are 3N phonons per k-point. The lowest three phonons are known as acoustic branches, which tend to zero at the center of Brillouin zone, known as Γ-point. The rest of the phonons belong to the optical branches, which have a finite frequency at the Γ-point. Identification of phonons in a 3D periodic solid involves numerical calculation of the eigenvalues of dynamical

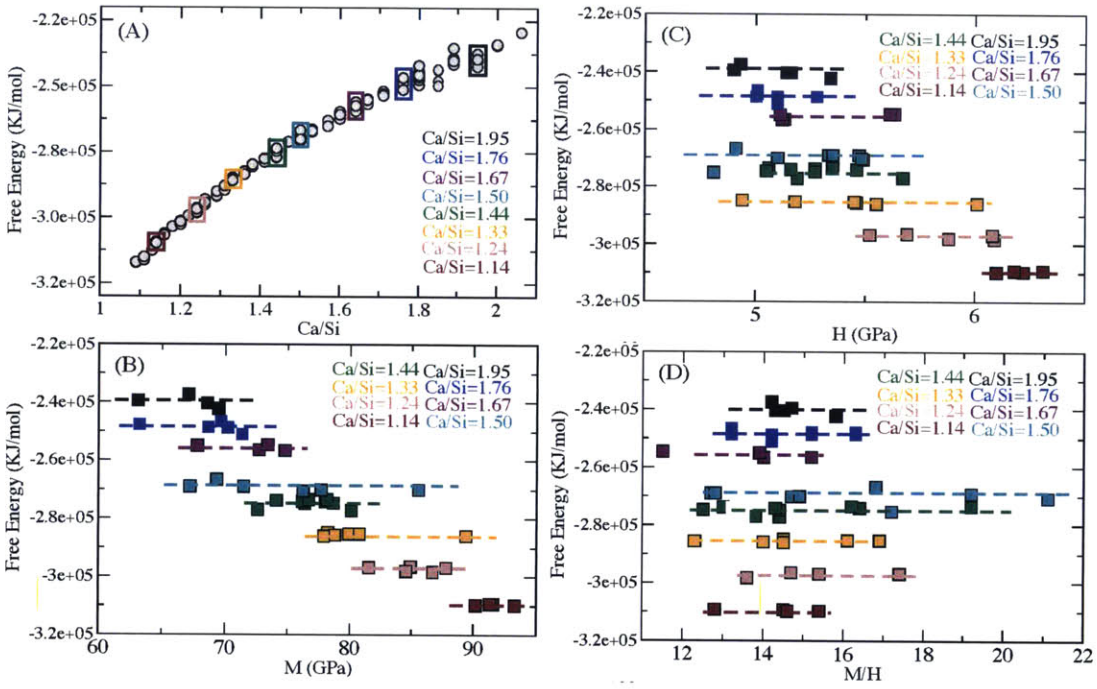

Figure A-7: Polymorphism in C-S-H. (A) Free energy as function of Ca/Si ratio. Free Energy as a function of B) and M C) H and D) M/H.

matrix at each k-point:

$$D_{i\alpha j\beta} = \frac{1}{\sqrt{m_i m_j}} \sum_R \left(\frac{\partial^2 U}{\partial\alpha\partial\beta} exp\left(ik\left(r_{ij} + R\right)\right)\right) \tag{A.34}$$

where m is the mass of atom, i and j indices denote two arbitrary atoms within the cut-off radius, α and β show the components in the real space, k is the reciprocal vector and R represents the sum over lattice vectors within the cut-off radius. Having phonons in hand, statistical mechanics can be employed as a powerful tool to bridge microscopic energy levels to macroscopic thermodynamical observables such as entropy, Helmholtz and Gibbs free energies and many others. The ubiquitous assumption in molecular physics is that all forms of the energy are independent and as a result the total energy can be decomposed to translational, rotational, vibrational and the contributions to the electronic state of the system. In the special case of solids at low temperatures (considerably below the melting point), the rotational and translational contributions can be neglected. In atomistic simulations, the electronic contribution is directly calculated via interatomic force field. Therefore, the challenge simplifies to the measurement of vibrational contributions. This is achieved via calculation of partition function, Z_{vib}, in the framework of lattice harmonic approximation. The vibrational energy level for a harmonic oscillator of the mth mode can be written as

$$U_{vib}^m(n,k) = (n + 1/2)h\omega(m,k) \tag{A.35}$$

where n is the number of phonons occupying that particular energy level. The number of phonons occupying an energy state is given by Bose-Einstein statistics:

$$n = \frac{1}{exp\left(\frac{-h\omega}{k_B T}\right) - 1} \tag{A.36}$$

where T is the temperature and k_B is Bolzman constant. The internal vibrational energy, U_{vib} can be written as the superposition of all the energy levels by considering that the energy at each state is the number of phonons occupying that state times

the energy of each phonon:

$$U_{vib} = \sum_m \sum_k w_k \left(\frac{1}{2} h\omega + \frac{h\omega}{exp\left(\frac{-h\omega}{k_B T}\right) - 1} \right) \tag{A.37}$$

where w_k represents the weight of a particular point in the Brillouin zone such that the sum of the weights equal one. To calculate the internal energy, the summation should be over all oscillation modes, m, and reciprocal lattice vectors, k. It should be noted that the fist term in the internal energy is just due to zero point vibrational energy. The partition function in canonical ensemble can be written as:

$$Z_{vib} = \sum_m \sum_k w_k \frac{exp\left(\frac{-h\omega}{k_B T}\right)}{exp\left(\frac{-h\omega}{k_B T}\right) - 1} \tag{A.38}$$

It should be emphasized that the frequencies are explicit function of k and m, which are omitted from the above equation for brevity. The Helmholtz free energy, Ψ_{vib}, and Gibbs free energy, L_{vib}, are readily calculated using:

$$\Psi_{vib} = \sum_m \sum_k w_k \left(\frac{1}{2} h\omega + ln\left(1 - exp\left(\frac{-h\omega}{k_B T}\right)\right)\right) \tag{A.39}$$

$$L_{vib} = \Psi_{vib} + PV \tag{A.40}$$

where P and V are pressure and volume of the unit cell, respectively. In practice, the above double summation is reduced to integration over the phonon density of states (DOS). Generally, the integration involves a numerical scheme over a discrete set of points. Due to the lack of symmetry in C-S-H's molecular structures, the most general set of points in the first Brillouin zone known as Monkhorst-Pack scheme is used in this work. A set of 1099 evenly spaced mesh of k-points, given by shrinking factors along each axis in reciprocal space, is adopted in all the simulations. Subsequently, the DOS is distributed over 60 equal bins and all the thermodynamic observables are calculated following the above-mentioned approach. The free energy as a function of

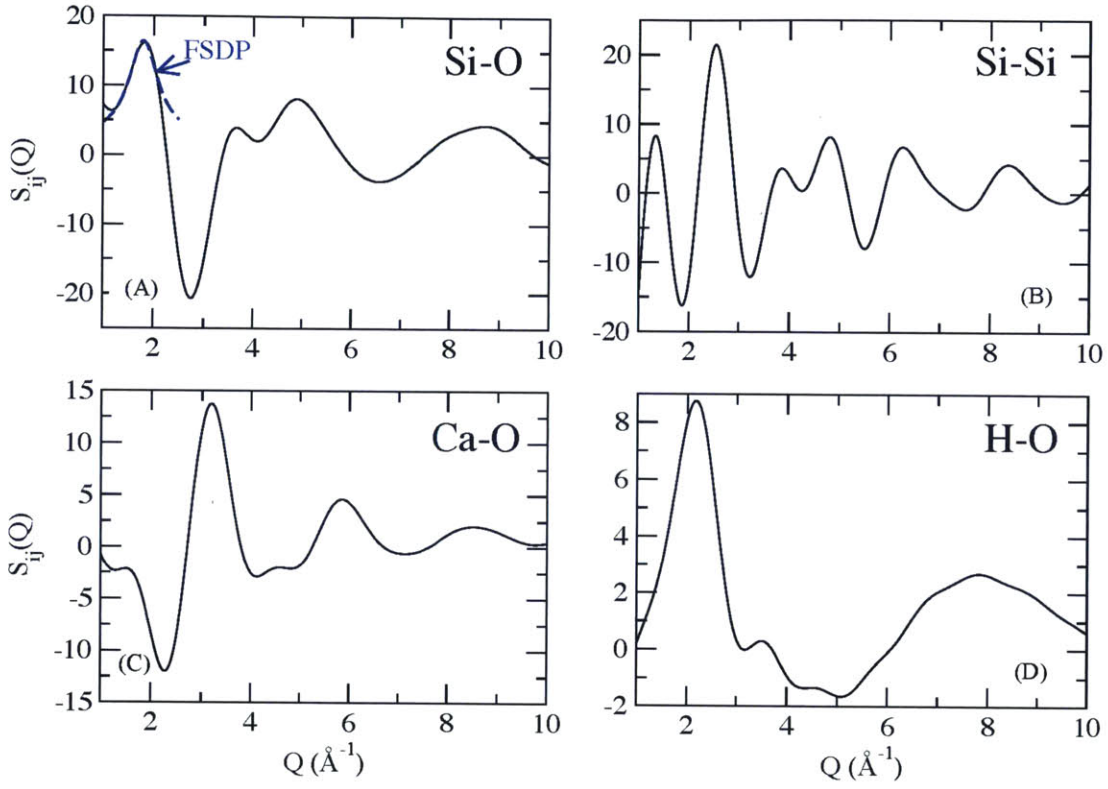

Figure A-8: Partial Structure factors in C-S-H. Computed (A) Si-O, (B) Si-Si, (C) Ca-O and (D) H-O structure factors at Ca/Si=1.7. The blue broken line shows a fit of the first sharp diffraction peak with a Lorentzian function.

C/S is presented in Fig. A-7.A. This shows that the free energy content decreases with C/S. Fig. A-7.B-Ds present the correlation between free energy and mechanical response of the C-S-H (M, H and M/H). They all show that at a given free energy level, the mechanical properties can vary which is a clear sign of polymorphism in C-S-H molecular structures.

A.11 Medium Range Order Analysis

As presented before, radial distribution functions (RDF) give insights into the short-range order of the structure of the material, this length scale being mostly driven by coordination numbers and bond lengths. As C-S-H polymorphs are characterized by the same short range order but show different densities, we analyzed their medium

range order to understand the origin of this density variability. To do so, we largely relied on tools introduced in glass science. The medium range order can be captured by the partial structure factors, which can be computed from the RDFs $g_ij(r)$ [8, 33, 34]

$$S_{ij} = 1 + \rho_0 \int_0^R 4\pi r^2 (g_{ij}(r) - 1) \frac{sin(Qr)}{Qr} F_L dr \tag{A.41}$$

where $Q = 2\pi/r$ is the scattering length, ρ_0 is the average atom number density and R is the maximum value of the integration in real space (here $R = 6$ Å). The $FL(r) = sin\left(\pi\frac{r}{R}\right) / \left(\pi\frac{r}{R}\right)$ term is a Lortch-type window function used to reduce the effect of the finite cutoff of r in the integration [359]. As discussed elsewhere [159], the use of this function reduces the ripples at low Q but can induce a broadening of the structure factor peaks. Fig. A-8 shows Si-O, Si-Si, Ca-O and H-O partial structure factors at Ca/Si=1.7, which are by all means similar to typical structure factors observed in silicate glasses [159].

Among the different features of structure factors, the so-called first sharp diffraction peak (FSDP) has received a lot of attention in glass science. Starting from the principal peak, which position $Q_P = 2\pi/r_P$ is associated to the nearest-neighbor distance r_P, the FSDP is defined as the next peak located at lower $Q < Q_P$. It is thus related to the structural correlations in the medium range order at $r > r_p$. This feature has been observed both in covalent [107, 317] and ionic [354] amorphous system. In ionic systems, this medium-range order has been associated to the forced separation between cations because of their mutual Coulomb repulsion, thus producing a pre-peak in the cation-cation structure factor58. Pre-peaks can also arise from size effects of the atoms of the network [158]. However, the network formation itself can have a major role since the FSDP is also observed in the monoatomic tetravalent systems a-Si and a-Ge [97, 343]. The FSDP origin is now usually explained by using a void-based model [108, 372] in which ordering of interstitial voids occurs in the structure. It is worth noticing that FSDPs are not simply the first of the many peaks of any diffraction pattern but display many anomalous behaviors as a function of temperature, pressure and composition [108, 227].

The FSDPs we obtained from simulations were further studied by fitting them with Lorentzian functions (example of fitted function can be seen on Fig. A-8). This choice is supported by the fact that the experimental results in neutron scattering factor of silica can be better fitted with a Lorentzian function than with a Gaussian one [361]. It should be noted that the fit has been done on the low Q part of the FSDP to avoid the contribution of the following peaks. This allows us to track precisely the FSDP peak intensity, position and full-width at half maximum (FWHM) for each sample.

Coming back to the real space structural correlations, the FSDP peak position QFSDP is usually related to a characteristic repetition distance $d = 2\pi/Q_{FSDP}$ and the FWHM to a correlation length $L = 2\pi/FWHM$, sometimes also called coherence length, due to atomic density fluctuations [375]. The effect of irradiation [361, 322], water content [375, 375] and alkali content 54 on the FSDP have been studied, leading to the idea that a depolymerization of the network (a decrease of the atomic order) is associated to a decrease of the intensity of the FSDP and a decrease in the characteristic distance [29]. A global understanding of the correlation length L is still lacking and seems to be highly system dependent.

A.12 Calculation of X-ray total pair correlation from atomistic simulation

We calculated the total X-ray pair correlation function $N_x(r)$ from atomistic simulation. $N_x(r)$ is related to the total X-ray pair distribution function $G_x(r)$ via:

$$N_x(r) = 4\pi r^2 \rho_0 G_x(r) \tag{A.42}$$

where ρ_0 is the average number density in the medium. $G_x(r)$ is related to the partial pair correlation functions $g_{ij}(r)$ where i and j indicate different species (Ca, Si, O

and H) in the simulation:

$$G_x(r) = \sum_{ij}(c_i c_j \frac{f_i(Q) f_j(Q)}{\langle f(Q)\rangle} g_{ij}(r) \tag{A.43}$$

where c_i is the concentration of i^{th} specie, $f_i(Q)$ is the atomic form factor describing the Q-dependent scattering length for the i^{th} specie and $\langle f(Q)\rangle = [\sum_i c_i f_i(Q)]^2$ [318]. The atomic form factor is calculated using:

$$f(Q) = \sum_{i=1}^{4} a_i exp\left(-b_i \left(\frac{q}{4\pi}\right)^2\right) + c \tag{A.44}$$

where a_i, b_i and c are element-dependent form factor constants. Due to the broadening effect induced by termination of X-ray structure factor at high Q values in experiments, we follow the suggestion of Wright [361, 359, 360] to calculate the broadened molecular dynamics response $G_x^b(r)$ as follow:

$$G_x^b(r) = \int_{-\infty}^{\infty} G_x(x) * H(r-x)dx \tag{A.45}$$

where H is a Gaussian with standard deviation set to 0.055. We use $G_x^b(r)$ to calculate $N_x(r)$ from eq. A.42. Subsequently, we calculated the coordination number of Ca atoms in the simulation using:

$$N_{Ca} = \int_0^{r_{min}} 4\pi r^2 \rho_O g_{CaO}(r)dr \tag{A.46}$$

where ρ_O is the average number density of oxygen atoms in the simulation box.

Appendix B

Coupled Nano-chemo-mechanical Testing

For the purpose of this study, cement based systems with different Ca/Si ratio were synthesized in the laboratory. Cement pastes based on the ordinary Portland cement Class I, with water-to-cement ratio [0.4, 0.5, 0.6] and cured at room temperature were produced as representative of the C-S-H with overall Ca/Si=1.75. C-S-H with lower concentrations of calcium were synthesized from a mixture of Portland cement Class G and crystalline silica, under hydrothermal conditions at the temperatures 130°C, 175°C and 200°C and the pressure of 3000 psi. The silica flour was added to achieve 35% by weight with the cement that resulted in the overall Ca/Si≈1 of the cementitious system.

B.1 Surface Preparation Protocol

The nanoindentation technique used requires a flat polish of the sample surface to minimize the role of surface geometry on measured mechanical properties. This was achieved using a semi-automatic polishing procedure developed and successfully applied to each specimen. In the first pass, cylindrical sub-samples, ≈1.0cm in diameter and 0.5cm in height, were cut from each bulk sample, with the use of the low speed diamond saw IsoMet, Buehler (Lake Bluff, USA). The cylindrical samples where then

mounted separately with cyanoacrylate glue on the standard AFM base (Ted Pella, USA), to be later automatically leveled and ground in a specially designed jig and the post system. Details of this system may be found elsewhere [11, 231]. The sample, inside the jig, was ground on a hard perforated cloth TexMetP (Buehler, USA) with a 9μm] oil-based diamond suspension MetaDi (Buehler, USA). In case of high-temperature cured specimens, the jig was loaded with an additional mass of 150g. Grinding pass was perfomed for approximately 20 minutes. This preparation time does not include the 3 min ultrasonic cleaning with n-Decane, (TCI, Tokyo, Japan) which is performed after every 5 min of grinding. In the second step, the ground and cleaned samples were manually polished on the alumina abrasive disks FibrMet, Buehler (Lake Bluff, USA), with the respective reduced abrasive sizes 9, 3 and 1μm. At each pad of different abrasive size the samples were polished for around 5min. In this step the used pad was exachanged for a fresh one every 2.5min. These changes of the pad, as well as abrasive size, were accompanied by 1min cleaning of the sample in the ultrasonic bath with n-Decane. So obtained final surface was characterized by rare occurrence of the scratches on its surface, visible under 20 and 50$\times$ objective of optical microscope, and high mirror-like reflectivity of the surface.

B.2 Indentation Measurement Protocol

Indentation tests in this study were conducted using the CSM Nanoindentation platform, equipped with the vibration isolation table as well as an environmental chamber to control testing conditions, temperature (18°C) and relative humidity (RH$\approx$20%), respectively. All indentations were conducted using the Berkovich diamond indenter, a three-sided pyramid with the half angle of 65.35 deg. and an area-to-depth function which is the same as that of well-known Vickers probe. Certified fused silica sample, with an elastic modulus of 73.1$\pm$0.7 GPa was used to calibrate the tip area function in the load range 0.5-100 mN. The indentation protocol consisted of three phases: linear loading and unloading at constant rate 8 mN/sec. This was performed to a maximum force of 2mN, with a holding period of 5 sec at the maximum load. Both properties

of interest, indentation modulus, M, and indentation hardness, H, were determined from the measured penetration depth-force curves according to the method proposed by Oliver and Pharr [249, 250], with each indentation curve being corrected for point of contact and frame compliance. All indentation measurements were conducted in the grid fashion, with the minimum number of indentation points set up to 400 and the grid spacing large enough for the successive measurements to be considered as statistically independent based on prior works [80, 121].

B.3 Statistical Analysis of the Indentation Data

Statistical analysis of the indentation data was carried out in the framework of Finite Mixture Models (see Eq. B.1) with the component probability densities of the mixture assumed to be of Gaussian type (Eqs. B.2-B.3) [337, 221].

$$f(\mathbf{y_i}, \mathbf{\Psi}) = \sum_{i=1}^{g} \pi_i f_i(\mathbf{y_i}, \theta_\mathbf{i}); 0 \leq \pi_i \leq 1 (i = 0, ..., g); \sum_{i=1}^{g} \pi_i = 1 \tag{B.1}$$

$$f_i(\mathbf{y_i}, \theta_\mathbf{i}) = \phi(\mathbf{y_i}, \mu_\mathbf{i}, \sigma_\mathbf{i}) \tag{B.2}$$

$$\phi(\mathbf{y_i}, \mu_\mathbf{i}, \sigma_\mathbf{i}) = (2\pi_i)^{-p/2} |\sigma_\mathbf{i}|^{-1/2} exp\left(\frac{-1}{2}(\mathbf{y_i} - \mu_\mathbf{i})^T {\sigma_\mathbf{i}}^{-1} (\mathbf{y_i} - \mu_\mathbf{i})\right) \tag{B.3}$$

Therefore, by design of the indentation experiment, at each location on the indentation grid the pair of properties under study is measured $\mathbf{y_i}$=(H_j,M_j) where j=1,...,N with N the total number of measurement points. A possible outcome of single realization of random variable $\mathbf{Y}$ is $\mathbf{y_j}$ belonging to one of g mechanically active phases within material microstructure e.g. residual clinker grains, calcium-silica-hydrate gel (C-S-H). Therefore, each of such phases has a uniquely defined pair of indentation hardness and modulus represented by the mean vector of properties μ_i=(H_i,M_i), with i=1,...,g. Additionally, in the framework of this statistical model, variances on both phase properties as well as correlations among them are captured in the covariance

matrices σ_i. The last parameter of the model, π_i, represents the surface fraction of a given phase. All unknown parameters $\mathbf{\Psi}$=(π_1,...,π_g,μ_1,σ_1,...,μ_g,σ_g) of the assumed GMM model are estimated with the maximum likelihood (ML) fitting approach via the expectation maximization (EM) algorithm [96]. The estimation of model parameters as well as the clustering of observations based on the Bayes rule of allocation is carried out with the use of non-commercial program EMMIX[?]

Estimation of the parameters of the assumed Gaussian Mixture Model requires, a priori, selection of the number of all g mixture components. In this study, the selection of model order, which corresponds to the number of mechanically active phases is assessed with Bayesian Information Criterion (BIC), Eq. B.4 [221, 96, 301]. Each grid indentation data set is analyzed for the range of components g=1,...,8, and the one which maximizes penalized log likelihood statistics is selected as optimum fit; k_g is the number of model parameters and $L_g(\Psi)$ corresponds to the likelihood of a model with g components.

$$logL_g(\Psi) - \frac{1}{2}k_g log(N) \tag{B.4}$$

A typical example of the applied statistical data analysis in the GMM framework and the inference of the optimum model are presented on the Fig.B-1. Presented indentation results were obtained on the ordinary Portland cement paste with w/c=0.45.

B.4 Electron Microprobe Measurement Protocol

Back-scattered electron (BSE) images, elemental X-ray maps, and chemical compositions of phases were obtained by a JEOL JXA-8200 Superprobe electron probe micro-analyzer (EPMA) at Massachusetts Institute of Technology, Cambridge, USA. The polished samples were analyzed by Wavelength Dispersive Spectrometry (WDS) using an accelerating voltage of 15 keV, a beam current of 10 nA, a 1 μm beam diameter, and 10s counting times per element with background correction at every point. The raw data were corrected for matrix effects with the CITZAF package [22]

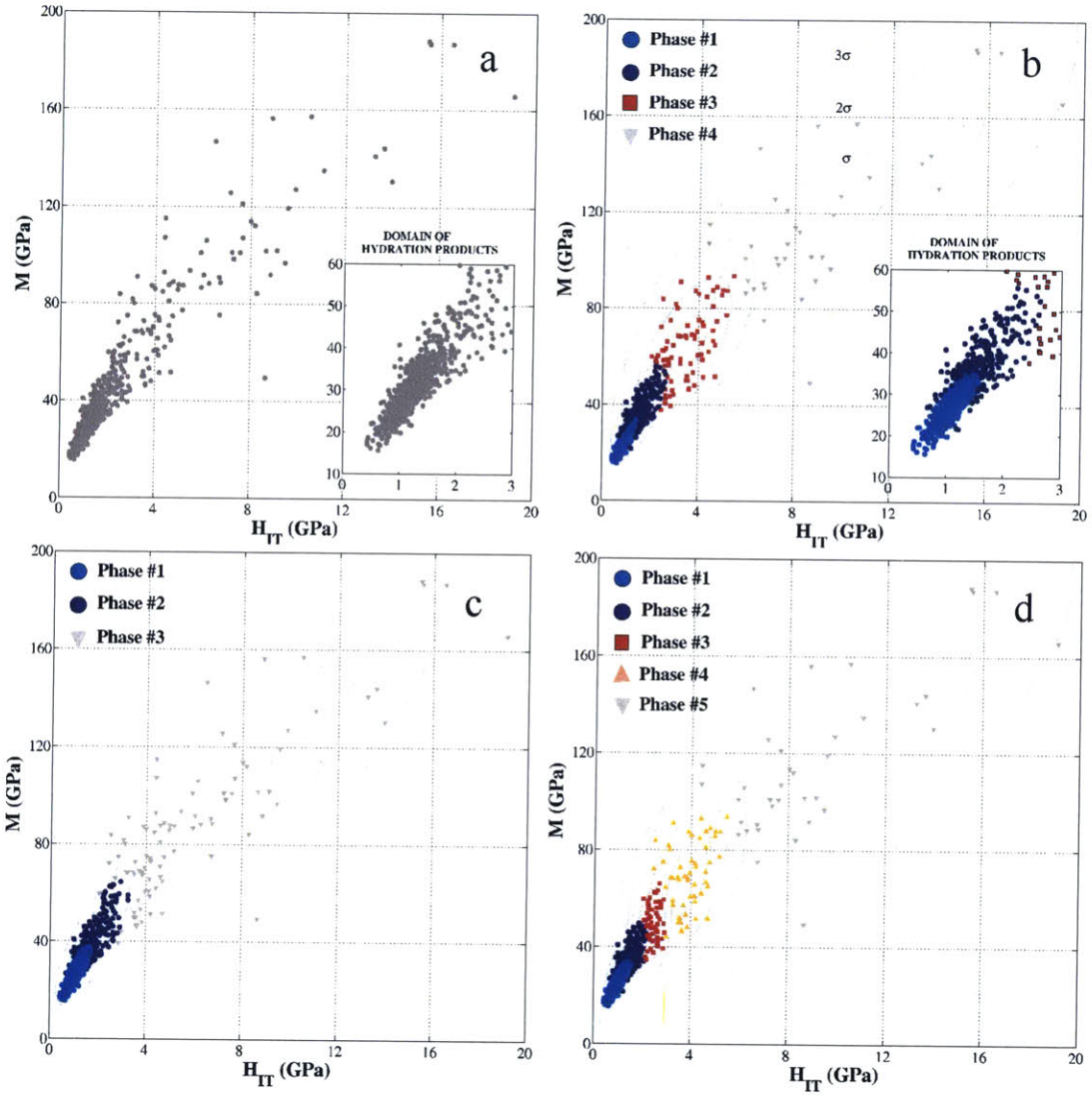

Figure B-1: Maximum Likelihood based deconvolution of the grid indentation data on OPC: a) global plot of the experimental records, with a focus on the expected domain of hydration products, b) clustering of the experimental data into four component Gaussian Mixture Model corresponding to maximized penalized likelihood BIC statistics, c-d) clustering of the experimental data with respect to the model with ± 1 number of optimum mechanically active phases. H-indentation hardness, M-indentation modulus, data courtesy of Dr. Matthieu Vandamme, Universite Paris-Est.

using the $\phi(\rho z)$-method (stopping power, back-scatter corrections and $\phi(0)$ from Love and Scott, mean ionization potential from Berger and Seltzer, Heinrich's tabulation of mass absorption coefficients, and Armstrong's modified equations for absorption

Table B.1: List of measured elements with associated standard materials and diffracting crystals used in the WDS measurements.

Element	Standard	Crystal
Si	Synthetic Diopside-Jadeite	PET
Ca	Synthetic Diopside-Jadeite	TAP
Al	Synthetic Diopside-Jadeite	TAP
Mg	Synthetic Diopside-Jadeite	TAP
Na	Synthetic Diopside-Jadeite	TAP
Fe	Hematite	LiF
O	Hematite	LDE1
K	Synthetic Orthoclase	PET
Ti	Rutile	LiF
S	NiS	PET

and fluorescence corrections [146, 1]. The set of measured elements, their associated standard materials, and the diffraction crystals are presented in Table B.1.

The obtained experimental data were statistically analyzed in the framework of Gaussian Mixture Models (eqs. B.1-B.3), as in the case of the nanoindentation results. However, contrary to the indentation data analysis results, chemical phases were inferred assuming multivariate model with the principal variables being four elements dominant in concentration:Ca;Si;Al;Fe. Such selection of variables provided sufficient "statistical contrast" for the chemical phases to be accurately inferred and clustered on a purely probabilistic basis. An example of electron microprobe data analysis in the assumed framework, obtained on the system cement Class G and crystalline silica in the amount 35% by weight of cement, is presented in Fig. B-2.

B.5 Assumed Micromechanics Model of Cement-Based Matrix, and Fitting Procedure

The intrinsic properties of the calcium-silicate-hydrate skeleton are determined from an indentation experiment using a micromechanics model based on the assumption of a granular solid (see Fig. B-3) [346]. The relationship between the phase hardness, H, measured in an indentation test on a porous solid at the scale of a REV, and the

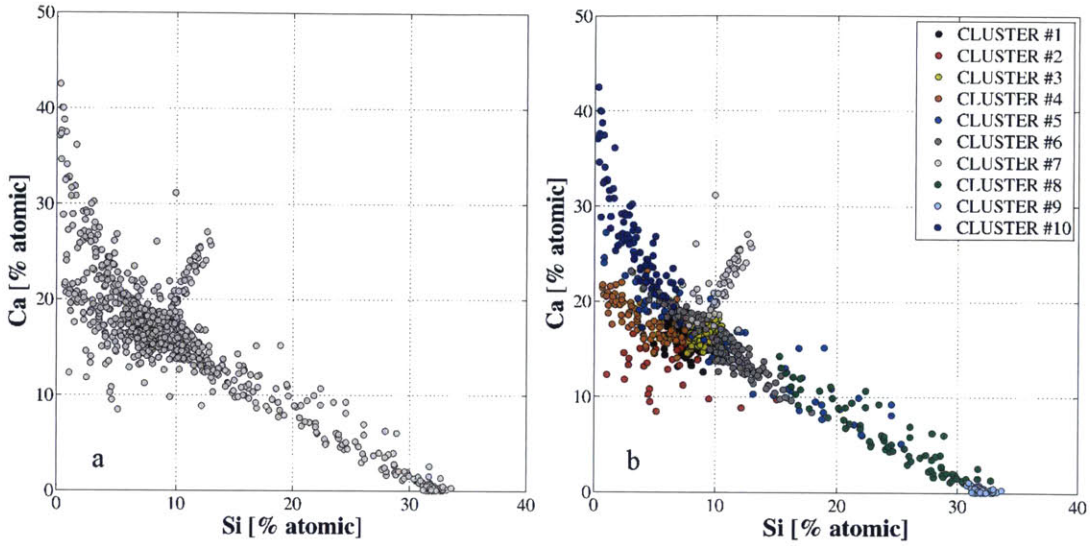

Figure B-2: Statistical analysis of the data obtained in the grid electron microprobe measurement on the system cement Class G with silica addition: a) a global plot of the experimental record in the Si-Ca domain, b) Maximum - Likelihood estimation and clustering of the experimental data into Gaussian Mixture Model corresponding to maximized penalized likelihood BIC statistics. Data courtesy of X-CEM project, research collaboration between Schlumberger and Massachusetts Institute of Technology.

intrinsic properties of the isotropic solid, cohesion c_s and friction coefficient (friction angle) α, is given by:

$$\frac{H}{h_s(c_s,\alpha)} = \Pi_h(\alpha,\eta,\eta_0) \tag{B.5}$$

The dependence of H on the local packing density η, percolation threshold η_0, and friction coefficient α, is captured in the dimensionless function Π_h. In the limiting case, $h_s = h_s(c_s,\alpha) = lim_{\eta\to 1)}H$ is the asymptotic hardness of a cohesive-frictional solid phase that obeys the Drucker-Prager criterion [346, 60]. The solid hardness relates to the solid's cohesion c_s, and friction coefficient, α, by the model function:

$$h_s(c_s,\alpha) = c_s \times a\left(1 + b\alpha + (c\alpha)^3 + (d\alpha)^1 0\right) \tag{B.6}$$

In eq. B.6 a, b, c, and d are the fitted parameters: $a = 4.7644, b = 2.5934, c = 2.1860, d = 1.6777$ respectively. The dimensionless function Φ_h depends on the local

solid packing density η and the friction coefficient α of the solid, for more details see ref. [346, 127].

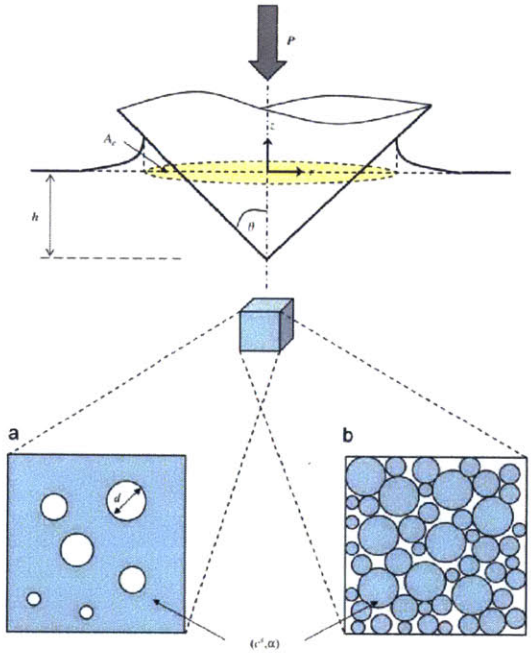

Figure B-3: Conical indentation in a cohesive-frictional porous material half-space: P is the indentation force, Ac is the projected contact area (projected on the initial sample surface). The porous composite (REV) is composed of a solid phase (cohesion cs and friction angle α) and pore space. Two solid-pore morphologies are presented: (a) matrix-pore inclusion morphology ($\eta 0$=0), (b) polycrystal morphology ($\eta 0$=1/2) assumed in the present modeling, after [346].

The indentation modulus M, extracted from the indentation test, relates to the solid modulus m_s, packing density η of the solid within REV, percolation threshold η_0, and Poisson's ratio ν_s. This relation is written in the dimensionless form as:

$$\frac{M}{m_s} = \Pi_M(\nu_s, \eta, \eta_0) \tag{B.7}$$

$$\Pi_M(\nu_s = 0.2, \eta, \eta_0 = 0.5) = 2\eta - 1 \tag{B.8}$$

For a perfectly disordered porous material system (see Fig B-4.b), the solid percolation threshold, representing the solid volume fraction required to provide a continuous force path through the system, approaches η_0=0.5. Assuming Poisson's ratio ν_s=0.2,

this relationship reduces to linear scaling of the indentation modulus with the packing density eq. B.8.

B.6 Fitting Procedure

The micromechanics model chosen links the measured indentation hardness H and modulus M to the set of three intrinsic properties m_s,c_s,α_s and packing density distribution η of the solid phase. Therefore, provided the existence of the unique solid C-S-H phase, the scaling relationships may be used to probe the microstructure sensed in the grid indentation experiment.

The solid properties and the packing density were determined by minimizing the quadratic error (eq. B.9) between the experimentally measured values $X = (M, H)$ and the model predictions using the scaling relationships. Because of different orders of magnitude between measured hardness and indentation modulus, the error minimization is carried out in the normalized space with the normalization factors $max(H, M)$. Such normalization prevents bias in the fitted model toward one of the variables, since errors on both have been assigned equal importance.

$$min \sum_{i=1}^{N} \sum_{X=(M,H)} \left(\frac{X_i - \kappa_X^S \Pi_X(\chi_X^s, \eta_i)}{max(X)} \right)^2 \tag{B.9}$$

The orthogonal least square fitting algorithm has been coded and executed in the Matlab environment using the trust-region-reflective algorithm for solving the nonlinear minimization [79]. In order to minimize the risk of obtaining the solution corresponding to local minimum, the minimization problem have been initiated from n=50 random initial vectors of the form m_s^0,c_s^0,α_s^0 drawn from the parameter space m_s^0=[0;140], c_s^0=[0;2], α_s^0=[0;1/$\sqrt{3}$[, and the solution with minimum error was selected. The termination tolerance on the objective function value as well as model parameters was set to 10^{-7}. An illustrative example of the results of error minimization procedure is presented on Fig. B-5.

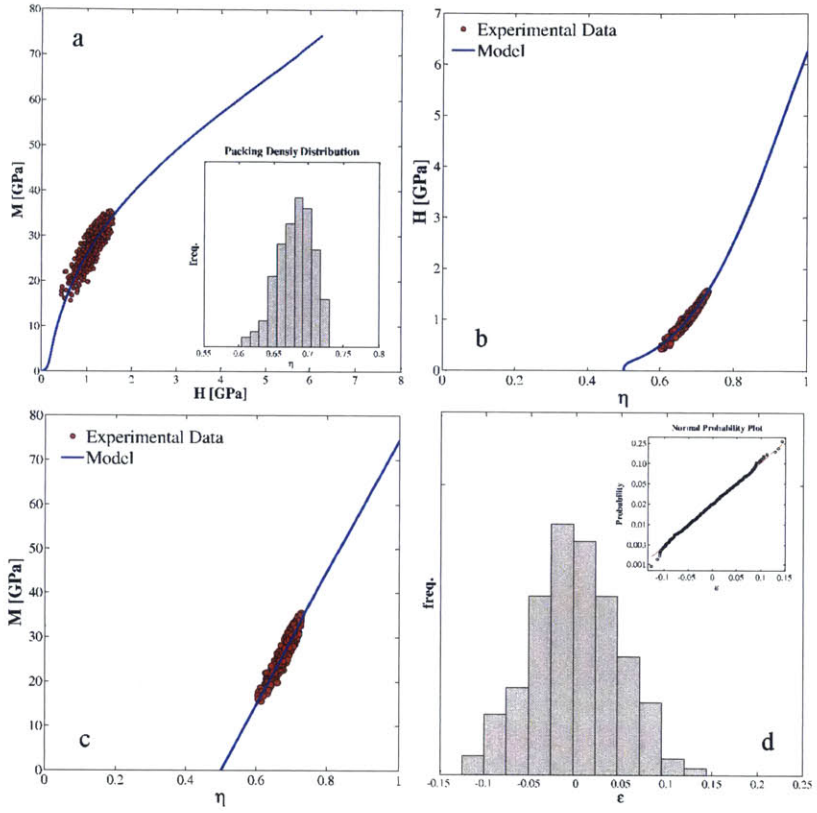

Figure B-4: Micromechanics based inference of the skeleton solid properties m_s,c_s,α_s of the C-S-H dominated matrix for Class G neat cement paste with w/c=0.45 cured at room temperature under 3000psi pressure; (a) scatter diagram of indentation results with overlaid optimum micromechanics fit m_s=74.5,c_s=0.57,α_s=0.34, and corresponding packing density distribution with mean $\overline{\eta}$=0.68 and standard deviation s_η=0.03, the estimated solid hardness of C-S-H particles is h_s (c_s,α_s)=6.3 GPa, (b) scaling of the indentation hardness and indentation modulus (c) with packing density, (d) distribution of model residuals ϵ_i

B.7 In-direct coupling of phase chemistry and mechanics: general step-by-step procedure and the cross-validation of the results

The linking of the mechanical properties of the solid C-S-H skeleton with its chemistry was done via in-direct coupling procedure, in which the results of EMPA and grid indentation are cross-validated by comparison of the estimated phase volume

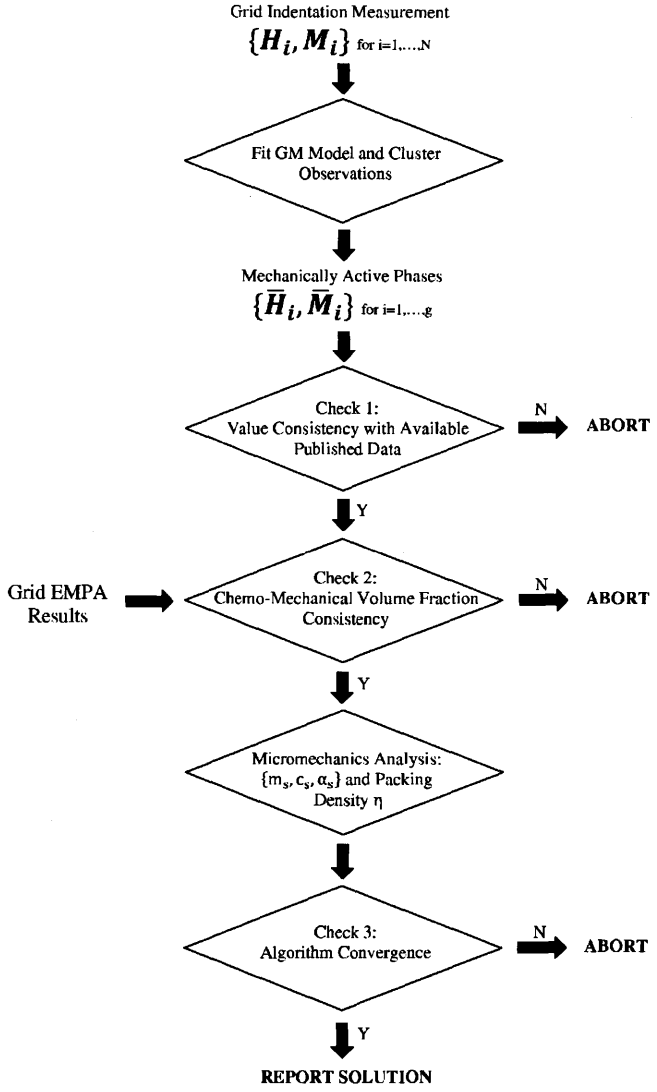

Figure B-5: Schematic of the step-by-step process of data analysis and indirect chemo-mechanic coupling.

fractions. These two experimental techniques have different objectives, chemistry assessment (e.g. Ca/Si ratio) on one side and mechanical properties (H, M) on the other. However, because of A- reliance on probing the area large enough to be considered representative of the bulk B- material volume sensed by both techniques are comparable ($\approx$1.5 μm) C- shared technique for statistical modeling and analysis of the experimental data The estimated volume fractions of the material phases obtained in

both approaches should be in general agreement. Such cross-validation is one of the main check-points in the general step-by-step rigorous procedure for the correlation of the indentation results with phase chemistry assessed by EMPA, as outlined on the flow chart on Fig. B-5. To ensure confidence, the threshold of 10% relative discrepancy in the estimated volume fractions of the hydration products by both techniques was enforced. A pair of measurements fulfilling this criterion was considered for the micromechanics analysis. Two other validations were implemented (see Fig. B-5): i) comparison against available literature data collected by the authors in the internal database system and b) verification of the convergence of error minimization algorithm in the estimation of solid properties. The final results reported in this work fulfill all three criteria.

Table B.2: Experimental data for nanoindentation and EMPA/WDS analysis. Data indicated as mean ± standard deviation, where standard deviation is calculated by bootstrap method. * - This sample type is typically rich in unreacted clinker, leading to unresolved mechanical properties of C-S-H with our cluster approach. Those samples have pronounced composite effect of CSH-clinker phases and therefore they are in the main text. As it can be seen in possible ranges of H and M values, they are close to those measured for clinker phases C2S and C3S [234, 352]. ** - values obtained from direct measurement of chemistry and indentation properties at the grid point location on C-S-H dominated phase

*0.78±0.28	115.8±24.7	13.8±4.9
0.88±0.06	93.3±6.6	12.4±1.8
0.98±0.04	98.1±4.9	8.2±0.7
1.69±0.11	69.0±6.8	5.3±0.9
**1.79±0.10	77.8±6.6	6.0±0.6
1.9±0.15	69.1±13.3	6.0±2.1
2.13±0.23	74.5±2.6	6.3±0.6
*2.1±0.18	100.9±8.3	8.5±1.0
**2.10±0.25	64.1±4.5	3.7±0.3

B.8 Calculation of uncertainty in M, H, and Ca/Si

In Table B.2, uncertainty is indicated in M, H, M/H or Ca/Si. For M, H we report uncertainty as standard errors on each parameter estimated in nonlinear regression analysis. The uncertainty on Ca/Si ratio is reported as sample standard

deviation of the experimental measurements attributed to C-S-H product, or dominated by this phase. Additionally, the statistical bootstrapping [204] was employed for selected samples in order to get more insight into the probability distribution of investigated parameters, and estimates of standard errors. Briefly, measured set of n nano-indentation values, $[, H]$ representing porous composite, was sampled with replacement in order to create m independent bootstrap samples. The size of bootstrap samples was equal to the size of the original sample. For each bootstrap sample the estimate of the monolithic C-S-H properties was obtained following the fitting procedure of micromechanics model described previously. The number of bootstrap samples, m, was typically > 50 in order to assure convergence of the standard error. The uncertainty in the quotient M/H was obtained according to propagation of errors theorem, assuming M and H are statistically independent and normally distributed.

B.9 TEM Imaging Conditions

The cement samples are prepared to obtain 30 microns thick blades (stuck on optical glass blades). A micro-pincher (O'Medenbach) is used to drill 3 mm diameters cement paste discs. Cu slots for TEM are stuck with a fine layer of araldite glue on these discs. The unit 'disc+slot' is then unstuck from the optical glass. After drying, the unit is subjected to ionic abrasion using a PIPS slimming ionic beam (Gatan) until the complete drilling of the cement in the central zone of the slot. The edges of the home are then ready for TEM Observations. Carbon metallization is not necessary to obtain the electronic conduction of the cement paste samples. Observation are performed on a high resolution field emission gun TEM JEOL 2010F (structural resolution of 0.18 nm, probe size of 0.5 nm, KEVEW EDS system, GIF loss of energy system, vacuum conditions 10^{-6} torr).

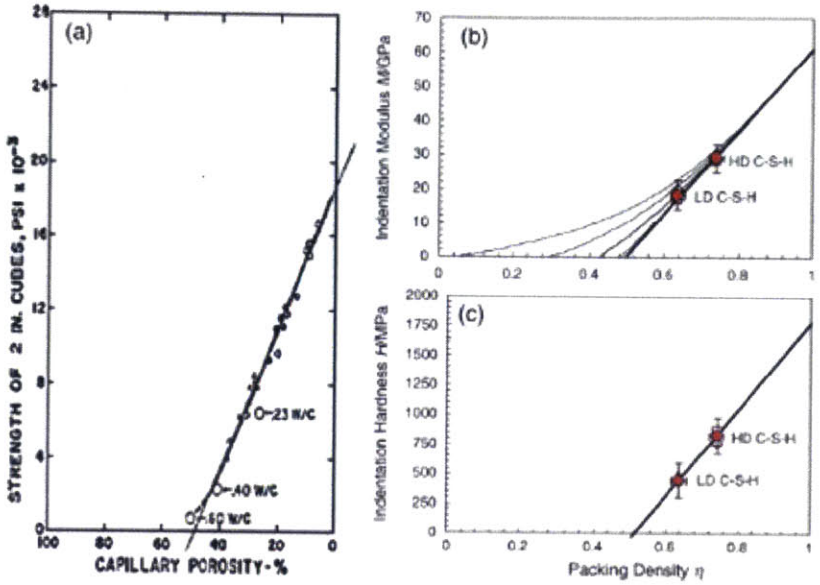

Figure B-6: Scaling of the mechanical properties of cement paste with porosity (one minus packing density) of a) strength b) indentation modulus, and c) hardness. (Figure adapted from Ulm and Jennings [345])

B.10 Scaling of macroscopic strength with C-S-H strength

Since the beginning of its industrialization in the early 20^{th} century, the strength of concrete (i.e. cement paste plus sand and aggregate) has been recognized to be determined by the strength of the binder phase [115] the cement paste. In its turn, groundbreaking work by Powers and collaborators at the Portland Cement Association in the 1950s and 1960s90 established functional relations between the cement paste strength and the porosity, see Fig. B-6); while assuming a constant strength of the elementary solid phase (C-S-H); a result that has been substantiated by means of nanoindentation in the last decade (see. E.g. Ulm and Jennings [345]). The result of this body of knowledge shows that for the same porosity of the cement paste, the engineering strength of concrete is scaled - in first order - by the strength of the elementary C-S-H particle. It is this scaling we employ to forecast the impact of raising the C-S-H strength on the environmental impact of concrete in the conclusion of the main text.

Appendix C

Sensitivity Analysis

The purpose of sensitivity Analysis (SA) is to measure, qualitatively or quantitatively, the variation of the response (observation) of an experiment (laboratory or numerical simulation) with regards to the variation of design parameters. In the realm of deterministic computer simulations, one can determine the underlying connection between the model output and the information fed into it [297, 298]. There are several reason why both experimentalist and modelers should perform sensitivity analysis [296]. The first reason is to clarify whether the proposed model properly represents the system or processes under investigation. The second reason is to identify the significant and insignificant factors (design variables) that contribute the output variability. The third feature of SA is the ability to determine the regions in the design space, which the variation of the response is either maximal or minimal. Finally, SA can identify which design variables interact with each other for producing the response.

The application of SA in deterministic computer simulations aims at determining how sensitive the model is with respect to the input parameters. When inputs are relatively certain, the partial derivatives of the output with respect to the input factors. Such a sensitivity measure can be easily calculated performing a few calculations around nominal values and estimating the derivatives via finite difference method. This method of SA yields the local impact of the inputs on the output and hence these kinds of methods are classified as local sensitivity analysis (LSA). However, we

usually want to identify which of the uncertain parameters are more important in in determining the uncertainty in the output. This plausible only through the global sensitivity analysis (GSA), which are usually implemented using pseudo-Monte Carlo sampling in the design space.

C.1 Different flavors of sensitivity analysis

SA has several different flavors, each suitable for a specific class of models and different stage of model construction. At early stages and for models with several uncertain input factors, a screening approach such as Morris method can be applied to qualitatively identify the significant parameters from unimportant inputs. Also, a natural step after MC sampling is to examine scatter plots to visually quantify the correlations between the input and output parameters. Such graphical method can also assist with identification of nonlinearity and non-monotonicity in input-output relation. For linear models, linear relationship measures like Pearson product moment correlation coefficient (CC), partial correlation coefficient (PCC) and standardized regression coefficients (SRC) would perform well. For monotonic nonlinear models, measures like rank transformation like Spearman rank correlation coefficient (RCC), partial rank regression coefficient (PRCC) and standardized rank regression coefficients (SRRC) would be suitable. For nonlinear non-monotonic models, methods founded on the decomposition of the variance are the proper choice. The examples of such methods include Sobol's method, Jansen's alternative, Fourier amplitude sensitivity test (FAST) and the extended Fourier amplitude sensitivity test (EFAST).

This section is divided in three parts. In the first part, we provide examples of One-at-a-time sensitivity analysis on building energy consumption through which we intend to build intuition on building heat transfer modeling. Based on this intuition, we propose a general probabilistic model for quantifying the effect of uncertainty in building parameters on the energy consumption. This includes model description,

identification of uncertain variables and the type of uncertainty. In the last part, we perform global sensitivity analysis via sampling based methods to identify the influential parameters on the energy consumption and other related energy consumption norms.

C.1.1 One-at-a-time sensitivity analysis

As mentioned earlier for LSA, all the design variables are kept constant at a nominal value except only one. This yields the relation between that particular design variable and response (observation) noting that the interaction with other design variables are neglected. Since only one variable is systematically varied, this approach is also called as One-At-a-Time (OAT) SA. Not being conclusive at all for systems with large number of design variables, OAT can be used to provide initial insight. Sailor and Munoz [294] performed OAT sensitivity analysis on the effect of climate on the electricity and fuel consumption of cities across the united states.

C.1.2 Sampling-based sensitivity snalysis

Sampling methods are commonly utilized for performing SA. Eisenhower et al. [104] performed sampling based analysis on more than one thousand parameters to identify the influential parameters, construct surrogate models and study their effect on cost of energy consumption. These methods are founded on the notion of generating input-output relations via Monte Carlo sampling. Sampling methods are categorized as the GSA approaches because these methods quantify the effect of x_i design variables, while simultaneously varying $x_j (j \neq i)$. In this approach, all the variables are varied in their entire domain. In case of existence of correlation between design variables, special attention is required as explained in sec. D.2. There exist various sampling-based GSA methods. Scatter plots are commonly used in the early phases of post-processing data to visually and qualitatively identify the correlations between the input and output. In complex high-dimensional systems and as a result of interaction

between design variables, graphical methods might turn out to be not especially informative. In such conditions, other sampling-based methods such as regression analysis, correlation coefficients and rank transformations are shown to be extremely insightful.

A powerful global sensitivity measure can be obtained via multivariate regression to fit the design variables (or their combinations, e.g. higher order terms) to a given empirical or theoretical formula that can yield the observations (response or output) as close as possible with minimal error. The most common regression technique for SA is least square approach (see sec. E.1.1). Since the coefficients of regression in linear fits are dependent on the dimension of the design variables and outcomes, the normalized form of the regression model is used in sensitivity analysis:

$$\frac{\hat{y}_i - \bar{y}_i}{\hat{s}} = \sum_{j=1}^{k} \frac{\beta_j \hat{s}_j}{\hat{s}} \frac{x_{ij} - \bar{x}_j}{\hat{s}_j}; \ \hat{s} = \left[\sum_{i=1} N \frac{(y_i - \hat{y})^2}{N-1} \right]^{1/2} \ \& \ \hat{s}_j = \left[\sum_{i=1} N \frac{(x_i j - \hat{x}_j)^2}{N-1} \right]^{1/2} \tag{C.1}$$

where bar above variables represents average operator and hat indicates the estimate of the linear regression response. In eq. C.1, subscript i permutes on the number of observations while subscript j is varied over the k terms in the linear regression. In GSA, the *standardized coefficients* (SRCs) $\frac{\beta_j \hat{s}_j}{\hat{s}}$ are considered as sensitivity measures. If x_j are independent, SRCs show the significance by moving each design variable away from its expectation value by a given fraction of its standard deviation while keeping all other variables at their expected values. Therefore, despite performing sampling and because of using linear regression, SRC is a local sensitivity measure rather than a global measure.

Correlation Coefficients (CC) usually known as Pearson's product moment correlation coefficients, provides a measure of the strength of the linear relation between two variables. CC is defined as:

$$R(x,y) = \frac{cov(x,y)}{\sigma(x) \times \sigma(y)} \tag{C.2}$$

where $cov(x,y)$ is the covariance between the design variable x and the response y and

σ denotes the standard deviation. Therefore, CC can be regarded as a normalized covariance between the two sets and provides a unitless index between -1 and +1. In fact, CC is equivalent of the coefficient of determination as defined in sec. E.2. Like SRC, CC only measures the linear relation between two variables by neglecting the effect of the other variables.

When more than one design variable is under consideration, like our probabilistic energy consumption model, *partial correlation coefficients* (PCCs) can be used to obtain a measure of linear relationship between the variables when all linear effects are removed. PCC between an a design variable x_i and response y can be obtained from the use of regression models. Denoting the PCC of x_i and Y by removing the effect of $Z = x_j (j = 1, ..., N \ \ \& \ \ j \neq i)$, $R(x_i, y|Z)$, then:

$$R(x_i, y|Z) = \frac{R(x_i, y) - R(x_i, Z) \times R(y, Z)}{\sqrt{(1 - R^2(x_i, Z)) \times (1 - R^2(y, Z))}} \tag{C.3}$$

In fact, PCC designates the strength of the linear correlation between the input and output after a correlation has been considered for the linear effect of the other design variables in the model. As a result, when the input factors are uncorrelated, results from PCC and SRC are identical.

The premise is based on the assumption of linear relationship between the design variables and observations. Therefore, those methods will perform poorly in the presence of nonlinearity in the model. In fact, as shown in Fig. 5-3, the inter-building distance affects the energy consumption of a building in nonlinear but monotonic fashion. Rank transformations are powerful tools to transform nonlinear monotonic relationships to linear correlations. Upon utilization of rank transformations, the samples are replaced with their corresponding ranks. The values are ranked from the smallest to the largest and the ties are equally distributed in between the values. Subsequently, the normal regression and correlation procedures are performed on the ranks of the variables rather than their actual values. *Standardized rank regression coefficients* (SRRC) are SRCs calculated based on ranks, *spearman rank correlation coefficient* (RCC) are corresponding to calculation of CCs on ranks and partial rank

correlation coefficients (PRCC) are PCC based om ranks. Statistics based on rank transformations are more robust as we obtain a great insight upon the existence of long tails in input-output distributions. The rank transformations are not only more linear but also more additives. This reduces the complexity by relatively increasing the weight of the first-order terms and decreasing the strength of higher-order terms and interactions.

Appendix D

Design of Experiment

As described in the previous section, energy consumption is influenced by several parameters ranging from building envelope properties to climate data. Here, we follow a minimalistic strategy. The first step in such an approach involves identifying the most influential parameters. This necessitates varying different parameters and analyzing how much they alter the energy consumption. Researchers have proposed several approaches to vary input parameters. These methods are classified under *design of experiment (DOE)*. In this section, we briefly elaborate on the basic concepts of DOE and describe two conventional methods, namely full factorial and pseudo-Monte Carlo designs that we have utilized in this work.

Before proceeding with the descriptions, we define a few common terms in the design of experiment literature. *Design variable (factor)* is any parameter or quantity that is thought to alter the observations, hence being varied during the (laboratory or numerical) experiment. We represent a design variable as an element of an N_d-dimensional vector, x_i, where $i = 1, 2, \ldots, N_d$. *Design space* is an N_d-dimensional space restricted by the lower and upper bound of the design variables. Typically, the design space bounds are scaled to range from -1 to +1 or 0 to +1 to take advantage of existing designs or avoid singularity in particular design. The domain of the design space is denoted by $[-1, +1]^{N_d}$ or $[0, +1]^{N_d}$ representing an N_d-dimensional hypercube. Therefore, a design variable is considered a real vector in the design space (i.e.

$\mathbf{x} \in \Re^{N_d}$). *Design point (sample)* is a specific instance of the $\mathbf{x}$ within the design space hypercube. The design variable is shown as a vector of length N_d or in an ordered t-tuple, $(x_1, x_2, \ldots, x_{N_d})$. *Response (also observation or outcome)* characterizes a dependent quantity that is measured at a particular design point. A single observation is mathematically represented as $y(\mathbf{x})$. A vector response, describing different dependent attributes at a design point, are shown as $\mathbf{y}(\mathbf{x})$, $\mathbf{Y}(\mathbf{x})$ or simply $\mathbf{Y}$.

DOE is a process of choosing a set of N_s design points in the design space with the intent of maximizing the information inferred from a limited number of observations [276]. The DOE strategies were originally developed for model construction in the laboratory and field experiments usually referred to as the classical DOE. The term "*modern* DOE" is rather forged more recently referring to a group of techniques specifically developed for applications in the computational design of deterministic computer simulations [50]. The very fundamental difference between modern and classical DOE pertains to the assumption that the random error exists in the laboratory experiments, but it does not occur in deterministic computer simulations [131]. Usually, random error is associated with the lab observations to explain non-repeatability in the experiments (e.g. observations in experimental chemistry and material science). This non-repeatability is mathematically stated as:

$$y_m(\mathbf{x}) = y_t(\mathbf{x}) + \epsilon \tag{D.1}$$

where y_m is the actual measured response, y_t is the true response and ϵ is an independent and identically distributed random (IID) error. Due to the random term, the repetition of the experiment at the very same design point x yields different results. This random error is always presents due to sources such as measurement error, inherent fluctuation in the response quantity (e.g. turbulence in experimental fluid dynamics), or other sources. To quantify the extent of the random error ϵ, the classical DOE employs replicate sampling. Replicate sampling is merely repeated measurements taken at identical design points, which permit a lack-of-fit statistical

analysis and measurement of random term.

Usually in the classical experimental design, it is assumed that the experimenter has a general knowledge of the underlying trends of the true response, $y_t(\mathbf{x})$. This *apriori* knowledge is utilized to place a fixed number of samples in the design space so as to minimize the relative influence of random error. This is ensured by spreading the design points as much as possible so that the ratio of $\epsilon/(y_m(\mathbf{x_i}) - y_m(\mathbf{x_j})$ is minimized. In fact, if the $\mathbf{x_i}$ and $\mathbf{x_j}$ sample points are close enough such that $y_m(\mathbf{x_i}) - y_m(\mathbf{x_j})$ becomes relatively small, then the relative error dominates resulting in poor description of the underlying model. Therefore, in classical DOEs, the samples are located at the boundaries with a few points in the middle of the design space.

Unlike classical DOEs, the observations in the deterministic computer simulations are repeatable meaning that the outcomes of independent simulations are identical up to the machine precision. Another feature that distinguishes between the classical and modern DOE is the choice of the probability distribution function (PDF) associated with each design variable [315, 246]. In classical DOE, it is typically assumed that the design variables are equally probable and restricted between an upper and a lower bound. However, in the modern DOE, design variables are allowed to follow uniform and non-uniform PDFs (e.g. Gaussian or Lognormal distributions). In addition, in contrast to classical DOEs, the modern DOE methods are inclined to insert design points in the interior of the design space, which is usually referred to as "space-filling tendency". Despite these differences, a common feature of classical and modern DOE is that the design points are independent meaning that the simulations or experiments can be performed independently from one another. In the case of computer simulations, the use of distributed processing can significantly decrease the computational costs as it scales with the inverse of the number of processors.

To properly understand the goodness of modern or classical DOE, one should understand different statistical errors such as bias and variance component of the

empirical error. In what follows, we assume that the function values (observations or data sets) have noise or random component associated with it. *Bias error* measures the extent to which the metamodel estimation (i.e. $\hat{f}(x)$) deviates from the actual observations (i.e. $f(x)$) calculated over the whole set of the DOE. Considering a quadratic loss function, the bias error is:

$$E_{bias}(x) = \{E_{ADS}[\hat{f}(x)] - f(x)\}^2 \tag{D.2}$$

where E_{ADS} denotes the expected value considering all possible data sets. Variance, on the other hand, quantifies how the surrogate model ($\hat{f}(x)$) is sensitive to a given training data set D, a random selection of the total observation set. The variance is mathematically stated as:

$$E_{var} = E_{ADS}[\hat{f}(x) - E_{ADS}[\hat{f}(x)]]^2 \tag{D.3}$$

There is a mutual trade-off between the bias and variance errors. In fact, a surrogate model that tends to fit a given data set (low bias) entails a higher variance [276]. Smoothing the surrogate function can decrease the variance. However, if surrogate function is smoothed excessively, the bias increases as the surrogate function poorly predicts the actual observations. Basically, both bias and variance can be simultaneously decreased by changing the surrogate model (choosing more appropriate metamodel based on the specific behavior of the observations) to decrease the bias and by increasing the size of training sets used to construct the surrogate model to lower the variance error. Usually, the numbers of the observations are extremely restricted because of the experimental costs or computational expenses associated with performing each observation. In practice, the design of experiment involves finding a balance between the bias and variance to minimize the associated costs. This balance is achieved by minimizing bias while imposing restrictions on the model complexity for example by using Tikhonov regularization. However, in the design of experiments for deterministic computer simulations where the numerical noise is negligible, the main focus is shifted to the minimization of the bias error. In fact, the space-filling

tendency of the modern DOEs is intended to assist in minimization of the bias error.

The bias minimization can be rigorously performed for some simple cases. For example, if the underlying model is thought to be a polynomial of a given order, there exist a theoretical design which minimized the bias design [244, 275]. In general, the bias can be significantly reduced if the design points are scattered uniformly in the design space [49, 293, 325]. Although the full factorial design is relatively uniformly distributed in the design space, the associated computational cost associated for such a dense design brings forward the idea of choosing only an adequate fraction of the factorial design. Such a design is usually referred to as fractional factorial. In general, the uniformity of the design space can be ensured via maximization of the minimum distance [163] or minimization of a correlation measure between the design points [156, 254].

D.1 Different Flavors of design of experiment

The choice of the design of experiment can strongly affect both the accuracy and the cost of the construction of the metamodel. Here, we elaborate on two DOE strategies, namely full factorial design and psudo-Monte Carlo design as they are commonly used in uncertainty propagation and design of experiments in problems involving many variables and mixed dummy-continuous design spaces.

D.1.1 Full and Fractional Factorial Design

Full factorial design (FUFD) is a classical DOE in which design variables are varied together, instead of one at a time. Therefore, FUFD is capable of capturing the interactions between design variables. In FUFD, the upper and lower bounds of all N_d design variables needs to be identified in advance. Subsequently, the admissible range of each design variable is discretized at s levels. If the design variables are only

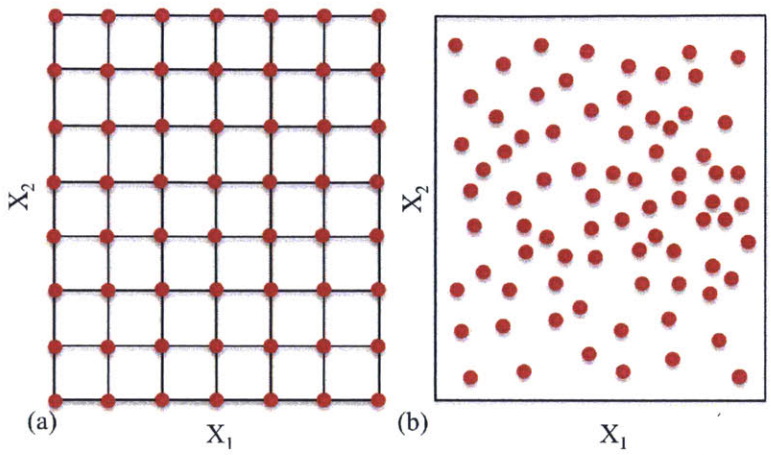

Figure D-1: The schematics of design of experiment in two dimensional space. a) Full factorial design with seven and eight intervals in X_1 and X_2 design variables, respectively. b) Pseudo-Monte Carlo sampling in X_1-X_2 space.

restricted to the upper and lower bounds (2-level FUFD), the FUFD consists of 2^s design points. Similarly, if the midpoints are also included in the DOE as well, the number of design points will be 3^s which is called 3-level FUFD. FUFD designs can be employed to construct higher order polynomials (second and more) which reproduce the interaction between design variables and the response surface curvatures. The main drawback of FUFD is that the number of design points grows exponentially (s^{N_d}) which for complex systems with large number of variables is intractable.

If the set of design variables becomes significantly large, only a fraction of FUFD can be used at the cost of sacrificing the ability to capture the interaction between the variables. This approach is called fractional factorial design (FRFD) which is extensively used in screening stage. For a s^{N_d} FUFD, a $(\frac{1}{s})^p$ fraction can be constructed, resulting in $s^{(N_d-p)}$ designs.

D.1.2 Central Composite Design (CCD)

Another classical DOE is the central composite design (CCD) which can be utilized to construct full quadratic (second-order) surrogate models. CCD is constructed of 2-level FUFD augmented by center and axial points to allow estimation of interaction

parameters in a quadratic model. In CCD, the number of design points is $(2^{N_d} + 2 \times N_d + 1)$, where 2^{N_d} corresponds to 2-level FUFD points, $2 \times N_d$ is the number of composite points in each design directions plus 1 point in the center of the design. CCD is usually viewed as an alternative for the 3-level FUFD (15 points in CCD and 27 point in FUFD). A common drawbacks of CCD and other classical DOEs is that they tend to place design points at or near the boundaries of the design space. This leaves the interior of the design space mostly un-explored.

D.1.3 D-optimal Designs (DOD)

As stated by Myers and Montgomery [244] for construction of modern DOE, the D-optimality criterion enables an efficient building of quadratic surrogate models. In D-optimal designs, the objective is to select N_s design points from a larger set of potential design points. If the polynomial regression is written in the matrix form as $\mathbf{f} = \mathbf{Z}\alpha + \epsilon$ such that $\mathbf{Z}$ is the matrix of the values of design variables at the potential design points (see section E.1.1 for polynomial regression), then the D-optimality criterion states that the best set of points in the experimental design maximizes the $det(\mathbf{Z^T Z})$ where the "D" stands for the matrix determinant. From a statistical point of view, a D-optimal design is attributed to a design which minimized the maximum variance in the predicted response. In other words, the choice of a D-optimal design reduces the statistical error in the estimation of the coefficients of the polynomial model. The advantages of the DOD are the ability to use irregular shapes and to add extra design points.

D.1.4 Lattice Hypercube Sampling (LHS)

In stratified sampling, it is made sure that all the design variables are explored uniformly at all levels which usually leads relatively small variances [293]. The idea of Latin squares and specially their generalization to Latin hypercube were performed by Kishen [174]. In modern DOE, LHS [220] is a flavor of stratified sampling with the condition that all the portions of any dimension of the design space denoted by

x_k should be sampled. As noted by McKay et al. [220] and Queipo et al. [276] for deterministic computer simulations, an LHS design of size N_s can be constructed by dissecting all dimensions in the design space into N_s strata with each stratum having a probability of $\frac{1}{N_s}$ to be sampled (each partition is sampled only once). An LHS sample can be denoted by $x_k^{(j)}$ where k denoted the dimensionality of the design space ($k = 1, \ldots, K$) and j represents the design point ($k = 1, \ldots, N_s$). In this case, an arbitrary LHS sample is composed of $x_k^{(j)}$ with their components being matched at random. An attractive aspect of LHS is that it allows tailoring the number of design points with regards to the available computational budget. While the cost of a LHS DOE scales with $O(N_d)$, the cost of classical DOEs scale with $O(s^{N_d})$. This makes LHS very appealing in modern DOE. Although LHS provides an improvement over unrestricted stratified sampling [320], it can produce samples with considerable range of performances. For instance, an LHS constructed by placing the design points along the diagonal of the design space, is not only a poor arrangement in terms of uniformity criterion but also is co-linear (high spatial correlation) which leads to ill-conditioned linear system of equations (e.g. least square scheme in polynomial regression, see section E.1.1). This is mainly because of the nature of the randomness in the distribution of LHS design points, as they might not meet the requirements for the uniformity. As discussed above, this shortcoming can be overcome using maximizing the minimum distance or minimizing the correlation schemes.

D.1.5 Orthogonal Arrays (OA)

OA is a generalization of LHS which was originally suggested in the work of Ruo [278, 279]. In the field of combinatorial design, an orthogonal arrays is a table whose entries are chosen from a set of symbols (usually $\{1, 2, \ldots, n\}$). The symbols are arranged in such a way that there exist an integer t, called strength, so that every selection of t columns forms exactly γ t-tuples (known as OA index) of the symbols, constructed by taking symbols from each rows as restricted by the specified t columns. In DOE, the size of the OA table if $N_c \times N_d$ where N_c is the size of the training

sample and N_d is the number of design variables. Here, $N_c = q^t$ relates the number of samples (N_s) to the number of symbols (q) and the strength. An OA can be shown in a vector form as $\text{OA}(N_s, N_d, q, t)$. In fact, $\text{OA}(N_s, N_d, q, 1)$ is nothing but an LHS design. There are two major limitations with OA. The first is the lack of flexibility as imposed to N_s as it cannot have an arbitrary integer vale depending the strength and number of levels. See Hedayat et al. [145] and Owen [254], for further details. Secondly, the OA design projected to the subspace spanned by an effective factors (set of the most influential parameters) can results in replication points for $\gamma > 1$. This is undesirable as it influences the bias and might cause mathematical singularity in constructing surrogate models.

D.1.6 Optimal LHS, OA-based LHS and optimal OA-based LHS

various methods have been proposed to resolve the lack of uniformity in LHS designs with the main idea being to optimize some relevant criteria (minimum distance or correlation) over the sample points [163, 254, 156]. As noted by Ye et al. [366], these methods are insensitive with respect to the optimal design criteria. Tang [325] has proposed an strength t OA-based LHS and has compared it with the standard LHS which the former is shown to be far superior. In optimized OA-based LHS, a measure of the spread of the sample points is optimized and the LHS search is restricted to orthogonal arrays [190]. In adaptive DOEs are proposed to insert additional points in the experimental design which is the outcome of the model appraisal step [164, 256]. Finally, having identified the surrogate model, the DOE can be optimized in such a way to suit that particular choice of the metamodel. Myers et al. [244] have extensively discussed the minimization of the variance in polynomial regression models. Qu et al. [275] have proposed a minimum-bias central composite designs for the polynomial regression models. The computational expense for optimization of the more advanced surrogate models are almost prohibitive and it is rarely performed.

D.1.7 Audze-Eglais' Design (AED)

Audze and Eglais [?] have suggested an optimal LHS with a non-traditional criterion (correlation) for optimization of DOE regardless of the mathematical description of the problem under consideration. The procedure to construct an AED consists of two steps. In the first step and analogous to the LHS design, the N_d design variable are discretized to N_s levels (equal to the sample size). In the second stage, the uniformity in the desig space is ensured as much as possible. This is analogous to an energy minimization scheme in a multi-particle system (particle $\leftrightarrow$ design point) with repulsive forces arising from electrical or gravitational origins. In both cases, if the charges or the mass is assumed to be identical at all the design points and equal to 1, then the total potential energy U_p can be written as:

$$U_p = \sum_{i=1}^{N_s} \sum_{i=i+1}^{N_s} \frac{1}{L_{ij}^2} \tag{D.4}$$

where L_{ij} denotes the euclidean distance between the i^{th} and j^{th} point in the design space. Simlar to degeneracy of the states in the statistical physics, the minimization of eq. D.4 with respect to L_{ij}s is non-unique. Regardless of the non-uniqueness, AED invloves an optimization in a high-dimensional space (large N_d) which computationally scales with $(N_s \times N_d)^3$ using Gauss-Jordan elimination which is the main disadvantage of AED. The other disadvantage of this method is that once the optimal design is achieved no extra points can be inserted to the final set without reading the second step.

D.1.8 Quasi-Monte Carlo Samling (QMC)

Quasi-Monte Carlo sampling (QMC) or low-descrepency sampling (also called quasi-random sequences or sub-random sequences) employs deterministic algorithms to generate design points in an N_s-dimensional space, so that the outcome design is fairly uniform. The term *discrepancy* referes to a quantitative measure of deviation from an ideal uniform distribution. Hence, discrepancy is a desired feature of this class

of modern DOEs. In other words, analogous to Auzde-Eglais' Design which enforces uniformity by minimization of a fictional potential energy (see section D.1.7), the final design in QCM is unstructured but nearly equally spaced in the design space. QMC has a few variants such as Halton [138], Sobol [314] and Hammersly Sequential Sampling [141] which all have the same convergence rate.

It is mathematically shown that quasi-MC has a convergence rate of $O(\frac{1}{N_s} \times (\log_{10} N_s)^{N_d-1})$ where as the convergence rate of psudo-MC method is $O(\frac{1}{\sqrt{N_s}})$ [253]. Niederreitee [246] has discussed the error bounds on multi-dimensional integration Comparing pseudo- and quasi-MC. Based on the conclusion of the Niederreitee's work, for low-dimensional design spaces ($N_d < 5$) quasi-MC sampling leads to lower integration errors than pseudo-MC sampling. for $N_d > 5$, the *average* error of pseudo-MC sampling is lower than the *exact* error of quasi-MC over the most reasonable values of the number of the samples (Note that QMC is deterministic while MC is probabilistic). However, since the error bound of the pseudo-MC is a probabilistic quantity, there is no guarantee that a particular set of pseudo-MC samples obtain this error bound. For this particular reason, some researchers in the numerical integration community prefer using quasi-MC as the error bound is exactly known.

D.1.9 Pseudo-Monte Carlo Sampling (MC)

Psudo-random sampling was first applied to computer simulations by Metropolis and Ulam [225] introducing the idea of psudo-random sampling for solving integral differential equations [113]. The prefix psudo refers to psudo-random number generation algorithms which are intended to emulate a true random behavior. Historically, the term "psudo" has dropped off and this technique is commonly referred to as Monte Carlo (MC) method. Given an interval $[\mathbf{x}_L \mathbf{x}_U]^{N_d}$ MC sampling invloves generation of the design point at random which the samples are ordered N_d-tuples. The idea of MC sampling can be extended to general convex design spaces by enforcing geometrical restrictions. However, the high-dimensional MC sampling in non-convex spaces can become nontrivial. While MC sampling is easy to implement, it might leave a large

portion of the design space unexplored due to random and independent nature of psudo-random number generation. The non-uniformity problem can be resolved by techniques such as stratified Monte-Carlo sampling [129]. Similar to LHS, the design variable domain is discretized in to s equally probable bins. Subsequently, a design point in each bin is randomly selected to ensure uniformity in the design space. The Main disadvantage of the stratified MC is that the computational cost of such design is of order $O(s^{N_d})$ which makes it unappealing for the study of complex systems with large number of design variables.

D.2 Sampling Non-uniform and Correlated Distributions

Depending on the system under investigation, the design variables might not be uniformly distributed in the design space. In such conditions, MC is advantaged over FFD as it can handle non-rectangular distributions. There are few methods for sampling random variables from a general probability distribution function ($P(x_i)$). Here, we use the most common approach known as inverse method. According to inverse approch, β (a random number from P distribution) can be generated using:

$$\beta = C^{-1}(\alpha) \tag{D.5}$$

where α is a uniformly-distributed random number and C^{-1} denotes the inverse of the cumulative distribution function (CDF).Another challenge in the analysis of complex systems is that the design variables might turn out to be correlated. To generate random variables with correlated structure, we consider $\Gamma = (L_1, L_2, L_3)$ as vector of random variables which its elements attribute to the L_x, L_y and L_z. The covariance matrix identifies the correlation between different elements of Γ and is defined as:

$$COV(L_i, L_j) = E\{[L_i - E\{L_i\}] \times [L_j - E\{L_j\}]\}, \quad i, j = 1, 2, 3 \tag{D.6}$$

where $E\{\ \}$ denotes the expectation operator. To transform non-uniform to uniform distributions, we use the CDF (F_i) of Γ elements leading to:

$$\lambda_i = F_i(L_i) \tag{D.7}$$

Subsequently, the fractile correlation matrix for λ reads [248]:

$$R_{mn} = R(\lambda_m, \lambda_n) = \frac{COV(\lambda_m, \lambda_n)}{\sqrt{\sigma(\lambda_m) \times \sigma(\lambda_n)}} = 12E\{\lambda_m \times \lambda_n\} - 3 \tag{D.8}$$

where σ is the variance operator. It is readily deduced from the definition that the fractile correlation is invariant with respect to monotonic transformation. Now, assuming that a translation form for the random variable Γ exists such that Γ can be written as a transformation of a Gaussian variate:

$$L_i = F_i^{-1}(G(\psi_i)) \tag{D.9}$$

where ψ_i is the zero-mean unit-variance random vector and G is the CFD of the standard Gaussian variate. Applying the monotonic transformation of the form given by Eq. D.7 to the both sides of Eq. D.9 reads:

$$F_i(L_i) = G(\psi_i) = \lambda_i \tag{D.10}$$

which implies that the fractile correlations of the translation variate and underlying Gaussian are matching. The moment-product correlation function (also covariance function) of the standard Gaussian variate can be obtained from its fractile correlation using the following simple relationship, originally proposed by Pearson in 1904 for the case of random variables [150]:

$$COV(\psi_m, \psi_n) = 2sin\left(\frac{\pi}{6} \times R(L_m, L_n)\right) \tag{D.11}$$

Note that Eq. D.11 results in a valid covariance function only if it is non-negative definite. Once a valid covariance function is in hand, the simulation of building length

random variable can be done by first generating the samples of underlying zero-mean Gaussian variate with a prescribed covariance matrix (generation of standard multivariate normal distribution with known covariance is performed using MATLAB®), and then applying the transformation given by Eq. D.9.

Appendix E

Response Surface Methodology

Any function that represents the trend of an observation over the range in design space is referred to as response surface. Traditionally, while response surface denotes the use of low-order polynomials in engineering applications, it is usually referred to as the true observation in statistics community where "response surface approximation" is the estimation of the true response. Response Surface is also called by many other names such as metamodel, surrogate model, approximation model, emulator or simply the model in different fields and contexts. Nonetheless, the common goal of response surface modeling is to estimate and predict the underlying trends in observed responses. While there are many forms for the approximation function, a generic model can be written as:

$$\hat{f}(\mathbf{x}) = f(\mathbf{x}, \mathbf{x}_s, \mathbf{y}_m(\mathbf{x}_s)) \tag{E.1}$$

where the sample design points are represented as $\mathbf{x}_s$, the already measured response data is denoted as $\mathbf{f}_m(\mathbf{x}_s)$, f is a pre-specified function and $\hat{f}(\mathbf{x})$ is the estimated response at an arbitrary point $\mathbf{x}$ in the design space. In this light, the surrogate modeling can be viewed as a non-linear inverse problem for which we aim to identify the continuous function (f) from a limited number of observations in the design space. The available observations $(\hat{f}(\mathbf{x}))$ can have either deterministic or probabilistic nature and therefore cannot possess sufficient information to uniquely identify (f). In fact,

the solution to the inverse problem is non-unique and multiple surrogate functions can be found to be consistent with the existing set of observations. Therefore, the surrogate modeling is a coupled problem of mutually constructing a model $\hat{f}$ (model estimation) and evaluating the errors associated with it (model appraisal) [313].

The parametric and non-parametric RSMs are the two alternatives to construct a surrogate model. While the global functional form between the design and response variables are known in the parametric approach, the non-parametric ones utilize different types of rudimentary local models in different regions of the design space to build up the metamodel. Polynomial regression and Kriging [182] are among the most notable variants of the parametric RSM. Radial basis function, multiple adaptive regression splines [119], projection pursuit regression [120] and neural networks [144] are the most common non-parametric RSM. A robust approach to investigate the appropriateness of a particular model ($\hat{f}$) is provided by regularization theory [336, 237] which imposes extra restrictions to the estimated model. Hence, $\hat{f}$ can be selected as the solution to the following Tikhonov regularization problem:

$$\min_{\hat{f}\in H} Z(\hat{f}) = \frac{1}{N_s}\sum_{i=1} N_s L\left(f_i - \hat{f}\left(\mathbf{x}^{(i)}\right)\right) + \lambda \int \|D^m \hat{f}\|_H dx \qquad \text{(E.2)}$$

where H is the set of surrogate models under consideration, $L(x)$ is the loss function (e.g. quadratic, Laplace, Huber, and ϵ-insensitive) used to quantify the empirical error, $\lambda \in \Re^+$ is a regularization parameter, and $D^m \hat{f}$ is the m-derivative of the proposed model. While the first term in the right hand side of eq. E enforces closeness by characterizing a proximity error, the second term ensures the smoothness of the proposed model by penalizing large derivatives (e.g. large curvatures for the case of $m = 2$). In fact, the role of λ is to enforce Occam's razor principle (principle of parsimony, economy or succinctness) by favoring simplicity in the model construction [45].

E.1 Different Flavors of Metamodels

Here, we discuss the model estimation and appraisal components corresponding to different parametric and non-parametric RSMs including polynomial regression, radial basis function, Kriging and multiple adaptive regression splines.

E.1.1 Polynomial Regression (PRG)

One of the basic parametric methods of the surrogate analysis is the polynomial regression (PRG). Recently, Brown [53] used this method for constructing emulators based on the MIT Design advisor online application. Hygh et al. [155] constructed a surrogate model for building energy simulation using Energyplus simulation software. PRG quantitatively associates a function of interest f to a set of N_{poly} basis functions $z^{(j)}$. In a design of experiment with a total N_s observations, PRG relates an observation f_i to a set of basis functions $z_i^{(j)}$ at the i^{th} point. This rises to a set of N_s linear equations:

$$f_i(z) = \sum_{j=1}^{N_{poly}} \alpha_j \times z_i^{(j)} + \epsilon_i \leftrightarrow \mathbf{f} = \mathbf{Z}\alpha + \epsilon \tag{E.3}$$

where ϵ_i is the independent error with expectation value of zero ($E(\epsilon_i) = 0$) and variance of σ^2 ($Var(\epsilon_i) = \sigma^2$). In the matrix formulation in eq. E.3, Z is a $N_s \times N_{poly}$ sets of polynomial evaluated at DOE points, α is a $N_{poly} \times 1$ vector of the coefficients of the regression and $\mathbf{f}$ is a $N_s \times 1$ vector of observations. The most common polynomial forms in metamodeling are linear, pure quadratic (no cross-correlation terms) and full quadratic. For example, a full quadratic polynomial in a two dimensional design of experiment ($N_{design} = 2, N_{poly} = 6$) can be written as:

$$\hat{f}(x) = \alpha_0 + \sum_{i=1}^{N_{design}} \alpha_i x_i + \sum_{i=1}^{N_{design}} \sum_{j \geq i}^{N_{design}} \alpha_{ij} x_i x_j \tag{E.4}$$

The coefficients of regression can be estimated using the least square method ($\hat{\alpha}$) which entail their unbiasedness and minimum variance. Therefore, the vector of

estimated regression parameters can found as:

$$\hat{\alpha} = (\mathbf{Z^T Z})^{-1} \mathbf{Z^T f} \tag{E.5}$$

E.1.2 Kriging Method

Kriging was originally developed in geostatistics (spatial statistics) by Danie Krige [182], a South African mining engineer. Later on, Matheron [216], a french mathematician, further developed the mathematical foundation of Kriging technique. It has recieved a popular acceptance in geostatistics applications ranging from hydrology (e.g. advective transport in underground water), environmental science (e.g. the air quality [35]), natural resources (resource evaluation [133] and ore selection in mining [289]) to health sciences (e.g. spatial mapping of sexually transmitted disease [189]). Kriging was first applied to the field of deterministic computer simulations by Sacks et al. [293]. Afterward, it was applied to problems in structural optimization and reliability [295, 170], metal forming process [160] and design of aerospike nuzzel in aerospace applications [309]. Recently, Kleijnan [176] has reviewed Kriging and its more recent variants. Martin et al. [213] utilized Kriging to perform test studies on different problems. Among few implementations of Kriging method, we use DACE MATLAB® toolbox provided by Lophaven et al. [198].

Kriging is a statistical method designed to estimate the response surface given a set, S, of already observed measurements including n observation points. While $y(s_i)$ is the response at the point s_i, it can be generally assumed that $y(s_i)$ is merely a single realization of a much fundamental random process, $Y(S)$. Therefore, if x represents an untried point in the admissible domain, X, $\hat{y}(x)$ and $\hat{Y}(x)$ are the estimate of the observation and the random process producing that observation at x. The estimate of the observation at untried point can be simply written as a linear superposition of already measured observations:

$$\hat{Y}(x) = \sum_{i=1}^{n} C_i'(x) \times Y(s_i) \tag{E.6}$$

where $C_i'(x)$ is the strength of observation at point S_i at point x. It is obvious that the expectation from $\hat{Y}$ and Y is the same:

$$E[Y(x)] = E[C'(x)Y(s)] \tag{E.7}$$

The Kriging method intends to provide the best estimate for $C_i'(x)$ by minimizing the mean square error (MSE) defined as:

$$\begin{aligned} MSE[\hat{Y}x] = E[\hat{Y}(x) - Y(x)]^2 = E[C'Y(s) - Y(x)]^2 = E[C'Y(s) - Y(x)]^2 \\ = E[C'(x)Y(s)Y'(s)C(x) + Y^2(x) - 2Y(x)C'(x)Y(s)] \end{aligned} \tag{E.8}$$

Statistically speaking, $C_i'(x)$ denoted the degree of correlation between the at point s_i and x. Two correlation matrices are defined subsequently:

$$r(x) = [R(s_i, x)_{n\times 1}] = [R(s_1, x)R(s_2, x)\cdots R(s_n, x)]' \tag{E.9}$$

and

$$R = [R(s_i, s_j)_{n\times n}] = \begin{bmatrix} R(s_1, s_1) & \cdots & R(s_1, s_n) \\ \vdots & \ddots & \vdots \\ R(s_n, s_1) & \cdots & R(s_n, s_n) \end{bmatrix} \tag{E.10}$$

where r and R are the correlation matrices between observed-untried and observed-observed sets, respectively. Also, $R(s_i, s_j)$ is defined as the $COV(s, x)/\sigma^2$ where COV is the covariance matrix in statistics defined as $COV(s, x) = E((s - E[s])(x - E[x]))$, and σ^2 denoted the standard deviation. There are many choices to adapt for the correlation functions but the most prominent which is stationary, (i.e. $R(s_i, s_j) = R(s_i - s_j)$) as well, is the multiplicative Gaussians:

$$R(s_i, x) = \prod_k \exp(-\theta_k(s_{i,k} - x_k)^2) \tag{E.11}$$

where k permutes over all the dimensions of x. In this functional form, $theta_k$ represents the correlation strength in the k^{th} direction. It should be emphasized that

since the Kriging does not use an Euclidean or Riemannian norms, then the issue of dimensional homogeneity will not occur as different dimensions are treated separately. By employing the above mentioned correlation functions, the MSE (eq.E.8) can be simplified to:

$$MSE = \sigma^2(C'_{1\times n}(x)R_{n\times n}C(x)_{n\times n} + 1 - 2C'_{1\times n}(x)r(x)_{n\times 1}) \tag{E.12}$$

Therefore, the $C'(x)$ coefficient can be estimated by minimizing the simplified form of MSE, introduced above. However, a constraint should be satisfied in the meantime; That is the observation at the untried point x should match the interpolation of some test functions. This test function can be linear interpolation between the data-set or other suitable fitted functions:

$$f(x) = [f_1(x)f_2(x)\cdots f_m(x)]' \tag{E.13}$$

and

$$F = \begin{bmatrix} f_1(s_1) & \cdots & f_m(s_1) \\ \vdots & \ddots & \vdots \\ f_1(s_n) & \cdots & f_m(s_n) \end{bmatrix}_{n\times m} \tag{E.14}$$

then the constraint would be in the form of:

$$F'C(x) = f(x) \tag{E.15}$$

Therefore, the problem of finding the best coefficients reduces to a constraint minimization problem, as defined below:

$$\begin{cases} \min_{C(x)} : 1 + C'(x)RC(x) - 2C'(x).r \\ F'C(x) = f(x) \end{cases} \tag{E.16}$$

This constrained minimization can be solved using Lagrange Multiplier method (please note that the multiplier is assumed to be 2λ instead of λ as this produces a nicer final

expressions):

$$\omega(c,\lambda) = (1 + C'(x)RC(x) - 2C'(x).r) + 2\lambda'(F'C(x) - f(x)) \tag{E.17}$$

Then, ω should be minimized with respect to both $C(x)$ and λ:

$$\begin{cases} \dfrac{\partial\omega}{\partial\lambda} = 0 \longrightarrow F'C(x) - f(x) = 0 \\ \dfrac{\partial\omega}{\partial C(x)} = 0 \longrightarrow RC - r + \lambda F = 0 \end{cases} \tag{E.18}$$

The above system of equations can be written in the matrix form as:

$$\begin{bmatrix} O_{m\times m} & F'_{m\times n} \\ F_{n\times m} & R_{m\times m} \end{bmatrix}_{(m+n)\times(m+n)} \begin{bmatrix} \lambda'_{m\times 1} \\ C(x)_{n\times 1} \end{bmatrix}_{(m+n)\times 1} = \begin{bmatrix} f(x)_{m\times 1} \\ r(x)_{n\times 1} \end{bmatrix}_{(m+n)\times 1} \tag{E.19}$$

The solution of the above system yields $C(x)$ and subsequently $y(s)$ can be estimated via eq.E.6.

$$\begin{cases} \hat{Y}(X) = F'(x)\hat{\beta} + r'(x)R^{-1}(Y(s) - F\hat{\beta}) \\ \hat{\beta} = (F'R^{-1}F)^{-1}F'R^{-1}Y(s) \end{cases} \tag{E.20}$$

In fact, the first term in the right-hand side of E.20 is the contribution of constraint functions to the response at x, while the second uncorrelated second term provides a smooth response of the residuals at the untried point, x.

It is emphasized that MSE is a strong function of the correlation parameter θ. Obviously, for any given set of θ's, Kriging yields an estimate of the observation at point x. But, the best Kriging solution is achieved if the optimum set of θ's are utilized. Ideally, this can be reached by minimizing the MSE with respect to θ's. However, there are two other approaches to estimate θ's using Cross-Validation (CV) and Maximum Likelihood Estimator (MLE). Especially in the MLE approach, one maximizes the likelihood of observations at already observed points by varying θ's.

This is mathematically stated as:

$$\theta_{MLE} \subseteq \{arg\max_{\theta \subset \Theta} \hat{l}(\theta|Y(s_1), Y(s_2), \ldots, Y(s_n))\} \tag{E.21}$$

However, there is a strict rule in MLE that the observations of $Y(s)$ should come from a Gaussian distribution. This is not very much the case for deterministic computer simulations. However, as we know the residuals are distributed normally which can be tested via Shapiro-Wilk normality test. Therefore, one can write the above formulation for MLE in more mathematically sound form:

$$\theta_{MLE} \subseteq \{arg\max_{\theta \subset \Theta} \hat{l}(\theta|Y(s) - F(s)\} \tag{E.22}$$

where F is a simple response surface such as polynomial regression. Using the MLE in describing the residuals, the Gaussian process can be assumed to describe the normality of residuals when estimating the response surface via the Kriging method [130]:

$$\begin{cases} L(\gamma|Y) = \dfrac{1}{(2\pi\sigma)^{n/2}\sqrt{det(R)}} \prod_{i=1}^{n} \prod_{j=1}^{n} exp[-\dfrac{1}{2\sigma^2}(Y(s_i) - F(s_i)\beta)R_{ij}^{-1}(Y(s_j) - F(s_j)\beta)] \\ \gamma = (\beta, \sigma, \theta) \end{cases} \tag{E.23}$$

γ is the vector of unknowns in the likelihood function. The log-likelihood is written as:

$$\hat{l}(\gamma|Y) = -\frac{n}{2}\ln(2\pi\sigma^2) - \frac{1}{2}\ln(det(R)) - \frac{1}{2\sigma^2}(Y - F\beta)^T R^{-1}(Y - F\beta) \tag{E.24}$$

Now, if $\hat{l}(\gamma|Y)$ is minimized with respect to σ^2 and β, $\dfrac{\partial\hat{l}(\gamma|Y)}{\partial\sigma^2} = 0$ and $\dfrac{\partial\hat{l}(\gamma|Y)}{\partial\beta} = 0$, the optimal β is found to be identical to $\hat{beta}$ in eq. E.20. and the estimate of the variance is:

$$\hat{\sigma}^2 = \frac{1}{n}(Y - F\hat{\beta})^T R^{-1}(Y - F\hat{\beta}) \tag{E.25}$$

Having $\hat{\beta}$ and $\hat{\sigma}^2$ known and at hand, the problem reduces to numerically maximize

$l(\gamma|\theta)$ which yields the estimate $\hat{\theta}$.

E.1.3 Multiple Adaptive Regression Spline (MARS)

The next variant of RSM is called Multiple Adaptive Regression Spline (MARS). MARS was initially proposed by Freidman [119] to efficiently approximate a response surface in a piece-wise regression scheme. This method has found an unprecedented acceptance in different fields of science and applications in engineering including speech modeling [152], mobile radio-channel prediction [105], intrusion detection in information system security [240], the correlation between retention indices and molecular description of Alkanes [362], pesticide transport in soils [364]. It has also got some attention in the finance for credit scoring [191] and also medical science by data mining for breast cancer patterns [70]. Related to energy domain, Zareipour et al. [371] used MARS to forecast the hourly Ontario energy price.

It should be emphasized that MARS is a non-parametric RSM approach. Global parametric methods such as polynomial regression methods are easy to develop and interpret. However, they have limited flexibility and work well only when the true underlying relation is close to the pre-specified functional form. To overcome the weaknesses of the global parametric methods, non-parametric models are developed locally over specified sub-regions of the data.

In MARS, the data is searched for the optimal number of sub-regions and simple function is optimally fit to the actual response in each sub-domain. Consider an n-dimensional variable X with N realizations of the system connecting X to y, therefore one can write the response surface as

$$y = f(X) + \epsilon = f(X_1 + X_2 + X_3 + ... + X_n) + \epsilon \tag{E.26}$$

where ϵ denotes the normal fitting error with zero mean. Briefly, the RSM can be approximated using MARS

$$\hat{f}(X) = C_0 + \sum_{m=1}^{M} a_m B_m(X) \tag{E.27}$$

where a_m are the coefficients of the fitting, M is the number of sub-domains or the number of basis functions in the approximated model. The main core of MARS modeling approach is the hockey stick spline which maps X to ξ

$$\xi = max\,(X - C, 0) \tag{E.28}$$

where C is referred to as the basis function knot. The mirror of the basis function is actually covering the values below the knot

$$\xi^* = max\,(-(X - C), 0) \tag{E.29}$$

Therefore, the basis functions in $B_m(X)$ can be developed as

$$B_m(X) = \prod_{k=1}^{k_m} \left[max(s_{k,m}(X_{v(k,m)-t_{k,m}}, 0))\right] \tag{E.30}$$

where k_m is the level of interaction between $v(k, m)$, $t_{k,m}$ is the knot location, $s_{k,m}$ takes +1 for the hockey stick basis function and -1 for its mirror image. Therefore, the MARS overall functional form can be written as

$$\hat{f}\,(X) = C_0 + \sum_{m=1}^{M} a_m \{\prod_{k=1}^{k_m} \left[max(s_{k,m}(X_{v(k,m)-t_{k,m}}, 0))\right]\} \tag{E.31}$$

MARS models are developed through a two-stage forward/backward step-wise regression procedure. In the forward stage, the entire domain is split into overlapping sub-domains and parameters are optimized by lack-of-fit criterion. If the maximum number of sub-regions is not specified, an over-fitted model will be produced. In the backward stage, the basis functions, which no longer attribute to the accuracy of the fit, will be removed. Usually, the users and limited to the second order problems set the level of interaction between the variables. A modified version Generalized

cross-validation criterion is used in MARS as the lack-of-fit criterion:

$$MGCV = \frac{1}{N}\sum_{i=1} N \frac{\left[y_i - \hat{f}(X_i)\right]^2}{\left[1 - \overline{C}(M)/N\right]^2} \tag{E.32}$$

where $\left[1 - \overline{C}(M)/N\right]^2$ is a penalty factor avoiding over-fitting. $\overline{C}(M)$ is defined as

$$\overline{C}(M) = C(M) + d \times M \tag{E.33}$$

where $C(M)$ is the number of parameters being used for fitting and d is the complexity penalty usually set between 2 and 3. To perform MARS modeling, we employ a MATLAB® toolbox (AresLab) developed by Jakobsons [161].

E.2 Metamodel Performance Metrics

In surrogate modeling, usually two DOE sets are constructed, one is training data set used for construction of the metamodel and the other one is testing set utilized for investigating the performance of the surrogate model. One can define different parameters to assess the fitness of the surrogate model [162]. These parameters known as performance metrics should have different characteristics to be suitable for studies. The first characteristic is accuracy, which implies the capability of predicting the system response over the proposed design of experiment space. Second, the robustness is also crucial as it entails the capability of achieving good accuracy in different problems. The third attribute is efficiency because of the computational expense required for constructing the surrogate model and predicting untried points. The fourth feature is the transparency which is the capability of providing the information concerning contributions of different variables and interactions among them. The last aspect of these metrics is conceptual simplicity which translates directly to the ease of implementation with having less user inputs and adaptability to new problem sets.

To quantify the performance of a surrogate model, its performance should be analyzed against a training set. There are three metrics, while the first two being

global, the third is rather local. The first metric is coefficient of determination or R square:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} = 1 - \left(\frac{e}{\sigma}\right)^2 \tag{E.34}$$

where y_i, $\hat{y}_i$, $\bar{y}$ are the actual observation, model prediction and the mean of the observations in all untried points, respectively. This metric can also be stated in terms of more familiar statistical parameters such as mean square error (e^2) and variance (σ^2) of the test data set. The second variable is relative average absolute error ($RAAE$) defined as:

$$RAAE = \frac{\sum_{i=1}^{n}|y_i - \hat{y}_i|}{n \times \sigma} \tag{E.35}$$

The third metric is the relative average absolute error ($RMAE$) which is defined as:

$$RMAE = \frac{max(|y_1 - \hat{y}_1|, |y_2 - \hat{y}_2|, \ldots, |y_i - \hat{y}_i|)}{\sigma} \tag{E.36}$$

In practice, it turns out that $RMAE$ and $RAAE$ are highly correlated and therefore, one of them along with R^2 would be enough to investigate the performance of the surrogate model.

Appendix F

Building Energy Consumption Modeling

Researchers have worked on computational tools for estimating building energy consumption for the past few decades. As a result of these numerous efforts, several software packages are developed. Depending on the supporting services, these programs can be proprietary, free or licensed under GNU. Some of these tools might be more specialized and hence more reliable for some applications than other. Endorsed by the US Department of Energy, Energyplus [84] is an energy analysis and thermal load simulation program intended to improve the core computational elements embedded in DOE-2 and BLAST programs. Energyplus calculations are based on combined heat and moisture balance using sub-hourly weather data. Although the source code is not public, it is offered free of charge and the developers assert that it is extensively before release. TRNSYS [177] is a proprietary software developed at University of Wisconsin-Madison and is composed of computational modules for the transient heat transfer simulation at the building level. Using steady-state heat flow calculations, MIT Design Advisor (MEA) [348, 347] is a web-based building energy simulation tool which aims to help reduce building energy demand at early design stage. Benchmark studies suggest that MEA's results are within 5-10% of the Energyplus predictions [349]. In this work, we utilize a UNIX version of the Energyplus. Unix operating system not only helps with easy manipulation of input and output

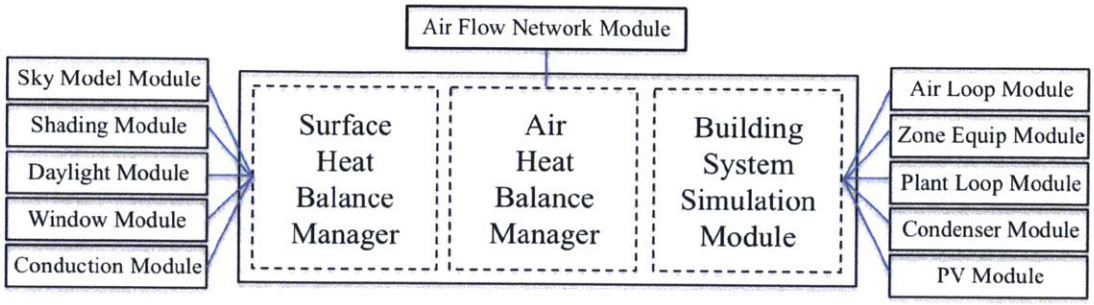

Figure F-1: The schematics of the solution engine of Energyplus which presents a complex system comprised of several modeules. Through interaction between different modules, the solution manager calculates energy consumption considering climate and human behavior and fundamental physical laws of heat transfer.

text files but also enables us to run several Energyplus calculations at a time on High Performance Computing resources. In this section, we elaborate briefly on Energyplus's computational engine. Eventually, This helps us better understand physics of building heat transfer.

F.1 Zone Heat Balance

Energyplus simultaneously solves the envelope, the system energy, and the air balance by considering multiple thermal zones (see Fig. F-1.). Energyplus solves implicit finite difference equations to enforce conservation of mass and energy. While the conservation of mass law controls the humidity level, the heat balance is enforced by the conservation of energy. The energy and humidity balance equations are simultaneously solved at hourly or even sub-hourly scales in all zones, air system and either side of the building envelope. Without undermining the effect of humidity and for the sake of brevity, we particularly focus on the energy term. Assuming a control volume for the total volume of a thermal zone, the heat balance can be written as (see Fig. F-2):

$$\dot{Q}_{air} = \dot{Q}_{pl} + \dot{Q}_{conv} + \dot{Q}_{mix} + \dot{Q}_{inf} + \dot{Q}_{HVAC} \tag{F.1}$$

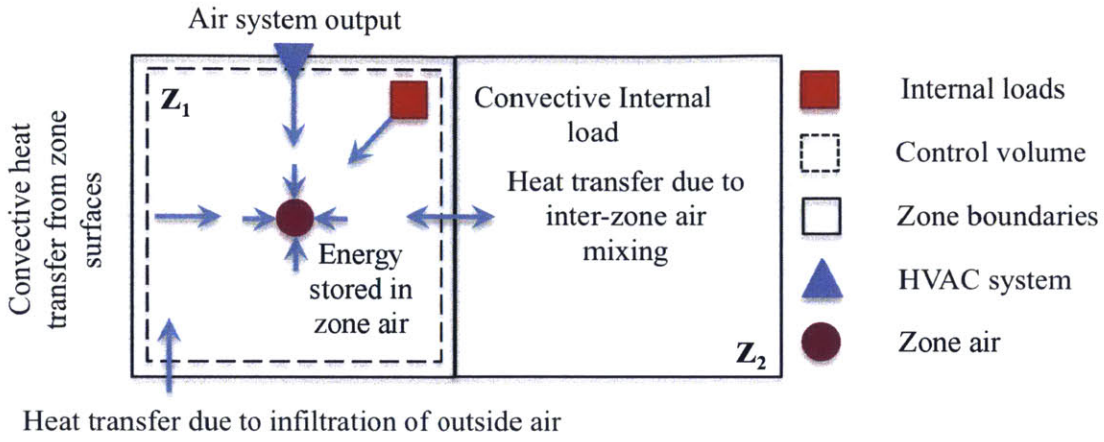

Figure F-2: Heat balance in Z_1 zone of a two zone system. The two zones are separated with a wall. The heat balance considers convective, infiltration, system input air and the energy stored in the air.

where $\dot{Q}_{air}$, $\dot{Q}_{pl}$, $\dot{Q}_{conv}$, $\dot{Q}_{mix}$, $\dot{Q}_{inf}$, and $\dot{Q}_{HVAC}$ are the rates of indoor air thermal energy, internal point load (residents, lighting, etc.), the heat convection from the envelope surface, inter-zone mixing energy, infiltration energy contribution, and energy source from the HVAC system, respectively. The energy rates in eq. F.1 can be separately stated in terms of the thermal oscillations and thermal properties of building surfaces and air:

$$\begin{aligned}
\dot{Q}_{air} &= C_v^{air} \times \frac{dT_z^{(t)}}{dt} \\
\dot{Q}_{conv} &= \sum_{j=1}^{N_{surf}} h_j \times A_j \times \left(T_{iwj}^{(t)} - T_z^{(t)}\right) \\
\dot{Q}_{mix} &= \sum_{i=1}^{N_{zone}} I_i \times V \times C_v^{air} \times \left(T_{zi}^{(t)} - T_z^{(t)}\right) \\
\dot{Q}_{inf} &= I_{env} \times V \times C_v^{air} \times \left(T_{\infty}^{(t)} - T_z^{(t)}\right) \\
\dot{Q}_{HVAC} &= I_{sys} \times C_v^{air} \times \left(T_{sup}^{(t)} - T_z^{(t)}\right)
\end{aligned} \tag{F.2}$$

$\dot{Q}_{air}$ is simply stated as a function volume heat capacity of air C_v^{air} times the rate of indoor temper oscillation T_z. Using a Newtonian convection model, $\dot{Q}_{air}$ is the

effective convection coefficient h_j times the surface A_j multiplied to the temperature difference between that surface T_{iwj} and the indoor temperature for the whole building envelope. $\dot{Q}_{mix}$ is the inter-zone infiltration rate I_i times the zone volume V multiplied by the heat capacity of air and temperature difference between the neighboring zones and the zone under consideration. $\dot{Q}_{inf}$ describes the air exchange rate between indoor and outdoor I_{env} with the outdoor temperature denoted as T_∞. $\dot{Q}_{HVAC}$ provides air at the rate of I_{sys} and the temperature of T_{sup} to balance energy loss or gain in the zone. This represents a simple HVAC system with constant mass flow rate and variable temperature. For more complex HVAC models (e.g. systems with variable air volume), the relevant equation for $\dot{Q}_{HVAC}$ changes accordingly.

The form of conservation of energy presented in eq. F.1 is an ordinary differential equation. While the analytical solution might exist for simple cases (e.g. assuming $\frac{dT_z^{(t)}}{dt} = 0$), finite difference (FD) method can be utilized to numerically solve this ODE. Due to the cyclic nature of the building heat transfer, FD error does not accumulate and moderately cancels for repeating cycles. The solution engine in Energyplus provides the option to choose between Euler and 3^{rd} order backward finite difference methods. Employing the second approach in this work, the rate of indoor temperature oscillations is expanded in the following form:

$$\frac{dT_z^{(t)}}{dt} = \frac{\frac{11}{6}T_z^{(t)} - 3T_z^{(t-\Delta t)} + \frac{3}{2}T_z^{(t-2\Delta t)} - \frac{1}{3}T_z^{(t-3\Delta t)}}{\Delta t} \tag{F.3}$$

where Δt is finite difference time steps. By applying the above finite difference scheme to the equations of energy conservation, the indoor air temperature at time t is derived:

$$T_z = \frac{\dot{Q}_{pl} + \sum_{i=1}^{N_{surf}} h_i A_i T_{wi}^t + \sum_{i=1}^{N_{zone}} I_i V C_v^{air} T_{zi}^t + I_{env} V C_v^{air} T_\infty^t + I_{sys} C_v^{air} T_{sup}^t}{\frac{11}{6} \times \frac{C_v^{air}}{\Delta t} + \sum_{i=1}^{N_{surf}} h_i A_i + \sum_{i=1}^{N_{zone}} I_i V C_v^{air} + I_{env} V C_v^{air} + I_{sys} C_v^{air}} + \frac{\frac{C_v^{air}}{\Delta T}\left(3T_z^{(t-\Delta t)} - \frac{3}{2}T_z^{(t-2\Delta t)} + \frac{1}{3}T_z^{(t-3\Delta t)}\right)}{\frac{11}{6} \times \frac{C_v^{air}}{\Delta t} + \sum_{i=1}^{N_{surf}} h_i A_i + \sum_{i=1}^{N_{zone}} I_i V C_v^{air} + I_{env} V C_v^{air} + I_{sys} C_v^{air}} \tag{F.4}$$

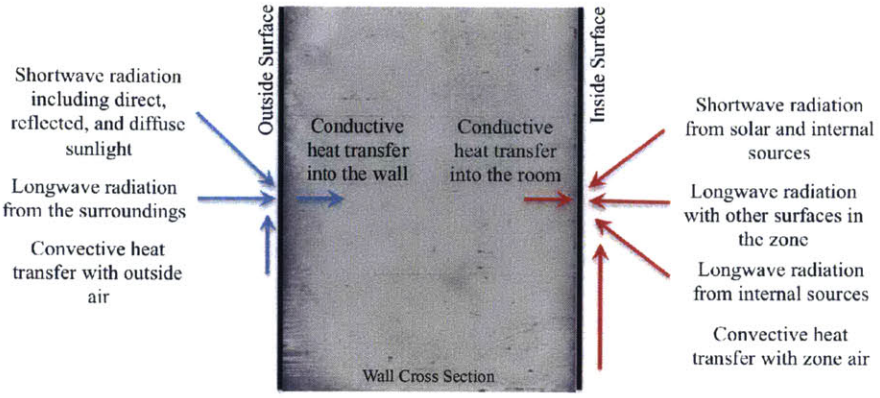

Figure F-3: Heat balance on the outside and inside surface of the envelope. According to the energy conservation energy law, sum of all conductive, convective, and radiation heat fluxes should be zero at any surface.

where the superscripts show the solution time step. In the right hand side of eq. F.4, all the variables are known except T^t_{sup}. In fact, the HVAC system adapts the supplied air temperature to keep the indoor air temperature at the desired level (T_{set}). In Energyplus, the implicit dependence of T_z to T_{sup} is resolved by a predictor-corrector scheme to iteratively calculate both unknown variables. In the first step of the solution, it is assumed that $\frac{dT_z^{(t)}}{dt} = 0$ and T_{sup} is calculated subsequently from eqs. F.1 and F.2. In the next step, $T_z^{(t)}$ is calculated using eq. F.4 and the estimated T_{sup}. This solution is used to re-evaluate T_{sup}. This iterative procedure is continued till the convergence is achieved.

According to eq. F.4, the calculation of T_z at any step of the solution is dependent on knowing the temperature of all inner surfaces of (T_iw) enclosing that zone. The calculation of T_iw entails satisfaction of heat balance at the inner surface of the walls which is related to heat conduction through the envelope and heat balance at the outer surfaces of the building envelope. Next, we describe the heat balance at the either surfaces of the building envelope.

F.2 Heat Balance on the Outer Surface of the Envelope

According to conservation of energy law, the heat flux rate leaving a surface should be equal to the heat flux rate entering that surface. The heat balance on the outer surface of the envelope can be written as (see Fig. F-3):

$$\dot{Q}_{asol} + \dot{Q}_{lwr} + \dot{Q}_{oconv} - \dot{Q}_{ocond} = 0 \tag{F.5}$$

where $\dot{Q}_{asol}$, $\dot{Q}_{lwr}$, $\dot{Q}_{oconv}$, and $\dot{Q}_{ocond}$ are absorbed direct and diffuse shortwave solar radiation, net longwave (thermal radiation flux exchange between surface, ground and sky) radiation, convective heat transfer from the outer surface and conductive heat transfer into the envelope, respectively. $\dot{Q}_{asol}$ is a function of location (longitude and latitude), surface angel and tilt, surface adsorption properties and atmospheric conditions (sky visibility and etc). $\dot{Q}_{lwr}$ depends on the surface adsorption properties and temperatures of wall, ground, and sky. The relevant surface properties of the envelope, emissivity ϵ, absorptivity α, reflectivity δ, and transmissivity γ (with $\alpha + \delta + \gamma = 1$), are complex functions of temperature, angle, and wavelength. However for building loads calculations, some simplifying assumptions are made (see Chapman [62] and Lienhard [196] for details). According to these assumptions, all surfaces are assumed to reflect or emit diffusely and identically (angle-independent and $\epsilon = \alpha$), are gray (wavelength-independent) and opaque ($\gamma = 0$ and hence $\epsilon = 1 - \delta$). In addition, surfaces are assumed to have uniform temperature and the heat flux is uniformly distributed across their surface. Furthermore, it is considered that the medium within the enclosure does not involve in the process. Therefore, $\dot{Q}_{lwr}$ can be written as:

$$\dot{Q}_{lwr} = \dot{Q}_{gnd} + \dot{Q}_{sky} + \dot{Q}_{air} \tag{F.6}$$

where $\dot{Q}_{gnd}$, $\dot{Q}_{sky}$, and $\dot{Q}_{air}$ describe long-wave radiative heat transport between the surfaces and ground, sky and surrounding air, respectively. Therefore, the long-wave

radiation can be stated in terms of Stephan-Boltzmann law:

$$\dot{Q}_{lwr} = \epsilon \times \sigma \left[VF_{gnd} \times \left(T_{gnd}^4 - T_{ow}^4 \right) + VF_{sky} \times \left(T_{sky}^4 - T_{ow}^4 \right) + VF_{air} \times \left(T_{\infty}^4 - T_{ow}^4 \right) \right] \tag{F.7}$$

where σ is Stephan-Boltzmann constant and T_{gnd}, T_{sky}, T_{∞}, and T_{ow} are environmental ground surface temperature, sky effective temperature, outside air temperature, and outside surface temperature, respectively. In Energyplus, T_{gnd} is set to be equal to the outside air temperature T_{∞}. Also, VF_{gnd}, VF_{sky}, and VF_{air} are respectively the view factor of the envelope surface to ground, sky and surrounding air. These view factors are given by:

$$\begin{aligned} VF_{grd} &= \frac{1}{2} \times (1 - \cos(\phi)) \\ VF_{sky} &= \frac{1}{2\sqrt{2}} \times (1 + \cos(\phi))^{\frac{3}{2}} \\ VF_{air} &= \frac{1}{2} \times (1 + \cos(\phi)) \times \left(1 - \sqrt{\frac{1}{2} \times (1 + \cos(\phi))} \right) \end{aligned} \tag{F.8}$$

where ϕ is the tilt angel of the surface.

F.3 Heat Balance on the Inner Surface of the Envelope

Similar to the heat balance on the outer surface, the heat balance on the inner surface can be written as (see Fig. F-3):

$$\dot{Q}_{lwx} + \dot{Q}_{sw} + \dot{Q}_{lws} + \dot{Q}_{iconv} + \dot{Q}_{sol} + \dot{Q}_{icond} = 0 \tag{F.9}$$

where $\dot{Q}_{lwx}$ is net long-wave radiant exchange between zone surfaces, $\dot{Q}_{sw}$ is net short-wave radiation to surface from lights, $\dot{Q}_{lws}$ is long-wave radiation flux from internal sources in the zone, $\dot{Q}_{iconv}$ is the conductive heat flux rate to the inner surface, and

$\dot{Q}_{sol}$ is transmitted solar radiation adsorbed at the surface.

There are two limiting cases for long-wave radiation among zone surfaces. The first limit is to assume that the zone air is completely transparent to the long-wave radiation. The other assumes that the zone air completely absorbs the longwave radiation from all surfaces. Energyplus takes the first assumption in heat balance calculation, which is a reasonable assumption as long as the moisture content in the zone is fairly low. Being founded on the traditional notion of the view factor (VF), Energyplus uses ScriptF method [151] to consider emission, re-emission and absorption modes of radiative heat transfer. ScriptF method contains matrix of exchange coefficients between pairs of surfaces that includes all exchange paths with the assumption that all surfaces are gray (wavelength-independent radiation properties) and diffuse (direction-independent radiation properties). It is written as:

$$\dot{Q}_{lwx}(p,q) = \sigma A_p \times VF(p,q) \times \left(T_{iwp}^4 - T_{iwq}^4\right) \tag{F.10}$$

where T_{iwp} and T_{iwq} are the temperature of the inner surface on the p^{th} and q^{th} surfaces. There are some complications regarding radiative heat transfer modeling in building energy simulation. The first problem is related to the uncertainty in the position of furniture and partitions. The second issue is the computational expense related to the exact calculation of VF. To simplify VF calculations, it is assumed that surface is not seen if it is not within 10° from the other surface and that all surfaces can see the floor, ceiling and roofs. Furthermore, Energyplus approximates the view factor in three steps. In the first step, Energyplus determines $A(i,j)$ which is the total area of surface j seen by surface i. In the next stage, the view factor is approximated as:

$$VF(i,j) \approx \frac{A(i,j)}{A_j} \tag{F.11}$$

It should be noted that no surface can see itself and thus $A(i,i) = 0$. The view factor calculation in Energyplus is so approximate that a third correction step is required to ensure reciprocity (two surfaces should exchange equal amounts of energy) and completeness ($\sum_j VF(i,j) = 1$) in radiative heat transfer.

The long-wave radiation from internal sources ($\dot{Q}_{lws}$) need the exact knowledge of all devices and their relevant radiative/convective model. Such considerations are extremely case-dependent. Energyplus distributes internal shortwave radiation ($\dot{Q}_{sw}$) from light sources in a prescribed manner among all surfaces. Transmitted solar radiation ($\dot{Q}_{sol}$) is distributed over the surfaces in the zone in a preassigned procedure. It is be possible to calculate the actual position of beam solar radiation, but it involves partial surface irradiation [281, 282]. This is however inconsistent with the rest of the zone model that assumes uniform conditions over an entire surface. Next, we elaborate on the convective terms in the energy balance on the inner and outer surfaces (eqs. F.5 and F.9).

F.4 Convective Heat Transfer

The complex turbulent nature of convective heat transfer complicates the measurement of convection coefficient (h) and therefore it has been the subject of many theoretical, numerical and experimental studies in the past several decades [77, 255, 157, 199, 25, 24, 44]. The nature of heat convection in outdoor (h_{out}) and indoor are different and they are treated with separate convection models in Energyplus. When the weather data indicates raining condition, Energyplus assumes that all the outer surfaces of the building are wet and subsequently sets $T_{ow} = T_{wet-bulb}$ and $h_{out} = 1000\frac{W}{m^2K}$. Energyplus provides several models for calculation of h_{out} including simple combined, TARP (Thermal Analysis Research Program) algorithm [353], MoWiTT algorithm [365], DOE-2 model and Adaptive Convection algorithm (ACV). In simple combined model employed in this work, h_{out} is a function of surface roughness and the local wind speed (V_z) as follows:

$$h_{out} = D + E \times V_z + F \times V_z^2 \tag{F.12}$$

where D, E and F are roughness coefficient and their descriptive norms vary from very rough to very smooth surfaces. In Energyplus, the wind speed is automatically

calculated for the building surface centroids (see details in section F.6).

Energyplus provides several categories and options for calculation of indoor convection coefficient (h_{in}). The inside surface classification is based on the flow regime (dynamical properties) and the orientation of the surface (static properties). The dynamical properties controlling the flow regime are Richardson number Ri and convective stability (ΔT sign). Ri is a dimensionless quantity relating Grashof (Gr) number and Reynolds number (Re) through $Ri = Gr/Re^2$. In buoyancy dominant flow regimes $Ri >> 1$ and in forced dominated flows $Ri << 1$. Some of them are dynamics a. The static properties include zone dimensions, HVAC type, hydraulic diameter of horizontal surfaces, HVAC airflow rate, surface classification, and tilt angle. Having implicitly or explicitly identified the static and dynamic convective properties, Energyplus offers seven major indoor convection models. These models are simple buoyancy, in-floor heating or in-ceiling cooling, wall panel heating, convective zone heater, mechanical central diffuser, mechanical zone fan distribution and mixed algorithm. Here, we employ simple buoyancy methods in our work due to the absence of knowledge about the building HVAC system. Simple buoyancy method uses Fohanno-Polidori relations [117] for calculation of internal heat convection coefficient of vertical surfaces (h_{in}^{ver}):

$$h_{in}^{ver} = 1.332 \times \left(\frac{|\Delta T|}{H}\right)^{1/4}; \qquad Ra_H^* \leq 6.3 \times 10^9 \tag{F.13}$$

$$h_{in}^{ver} = 1.235 \exp\left(0.0467H\right) |\Delta T|^{0.316}; \qquad Ra_H^* > 6.3 \times 10^9 \tag{F.14}$$

where $DeltaT$, H and Ra_H^* are the temperature difference between wall surface and ambient, wall height and modified Reighley number based on wall height, respectively. For horizontal surfaces convection, simple buoyancy model utilizes Alamdari-Hammond models [15] for both stable and unstable regimes.

$$h_{in}^{hor} = 0.6 \times \left(\frac{\Delta T}{D_n^2}\right)^{1/5}; \qquad \Delta T > 0 : stable \tag{F.15}$$

$$h_{in}^{hor} = \left[7.53 \left(\frac{|\Delta T|}{D_n^2}\right)^{3/2} + 18.75|\Delta T|^2\right]^{1/6}; \qquad \Delta T < 0 : unstable \tag{F.16}$$

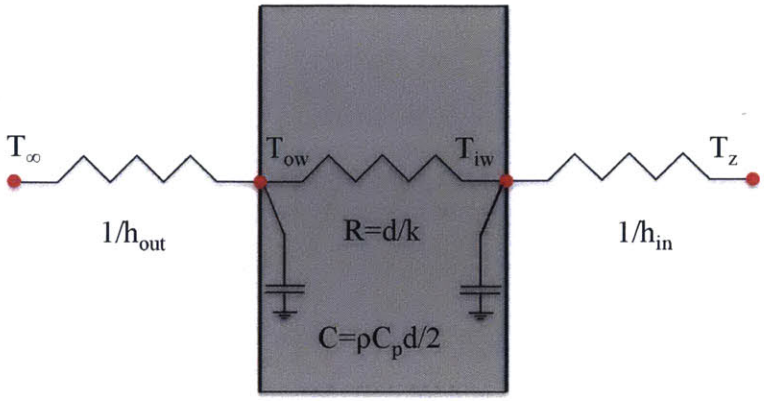

Figure F-4: The schematics of state space method with only considering 2 nodal points at the inner and outer surfaces of the wall. The conductive and convective thermal resistences are denoted as resistors and the thermal energy storge in the wall is represented by capacitors.

where D_n is the hydraulic diameter defined as $D_n = 4A/p$ where A and p denote area and perimeter, respectively. Finally, simple buoyancy model employs ISO 15099:2003 [4] for estimating convection coefficients from glazing using:

$$h_{in}^{gl} = N_u \times \left(\frac{\lambda}{H_w}\right) \tag{F.17}$$

where N_u, λ and H_w are Nusselt number, thermal conductivity of air and window height, respectively. For extra information on convective heat transfer, see Energyplus engineering manual and relevant references therein. In the next section, we discuss the conductive heat transfer through the building envelope.

F.5 Conductive Heat Transfer

The temperature of building's inner and outer surfaces plays a significant role in heat balance of conditioned zones and heat exchange with surroundings. These temperature, namely T_{iw} and T_{ow}, are strongly dependent on the conductive thermal properties of the walls. Such dependence is addressed via numerical modeling of heat conduction through the wall thickness. In building heat transfer modeling, the

conductive heat transport PDE ($\frac{\partial Q}{\partial T} = -\alpha \frac{\partial^2 Q}{\partial x^2}$ with α being the heat diffusivity of the wall) is solved with the boundary conditions satisfying the heat balance equations on the either sides of the wall (Dirichlet boundary conditions). Energyplus uses state space method which is a flavor of conduction transfer function (CTF) approach [61, 303, 252] to calculate the nodal temperatures and heat conduction through the envelope (see Fig. F-4). For instance by considering two nodal points at the inner and outer surfaces of the wall, the heat balance can be written as:

$$C\frac{dT_1}{dt} = h_{out} \times (T_\infty - T_{ow}) + \frac{T_{iw} - T_{ow}}{R} \tag{F.18}$$

$$C\frac{dT_2}{dt} = h_{in} \times (T_z - T_{iw}) + \frac{T_{ow} - T_{iw}}{R} \tag{F.19}$$

where C is the thermal heat capacitance of the wall defined as $C = C_p \rho d/2$ where C_p is the heat capacity of the wall, ρ is the density, d is the wall thickness and 2 in denominator indicates that each node pertains to half of the wall volume. R is the thermal resistance of the wall and is defined as $R = d/K$ where k is the thermal conductivity of the wall. These sets of state space equations can be generalized to matrix formulation:

$$\begin{aligned} \frac{d\mathbf{x}}{dt} &= \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{u} \\ \mathbf{y} &= \mathbf{C}\mathbf{X} + \mathbf{D}\mathbf{u} \end{aligned} \tag{F.20}$$

where $\mathbf{x}$ is the vector of state variables (nodal temperature), $\mathbf{u}$ is the vector of inputs (indoor and outdoor temperature), $\mathbf{y}$ is the vector of outputs (resulting heat fluxes) and $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$ and $\mathbf{D}$ are coefficient matrices. Next, we discuss the climate data as we have encountered it in numerous occasions so far such as outdoor temperature, humidity, and wind velocity and radiation terms.

F.6 Atmospheric conditions: solar radiation, shading and weather data

There are several challenges in measurement of climate conditions and hence preparation of weather data for building energy consumption modeling. For instance, proper measurement of solar lighting fluctuations under partly cloudy conditions is fairly sophisticated. Also, wind direction and speed can be source of error as it might frequently change in between measurement intervals. More importantly, the effect of urban dwellings on atmospheric conditions known as heat island effect [232] is not properly captured by the weather stations far from urban centers. Recognizing all these limitations, Energyplus weather file contains hourly readings of dry bulb temperature, dew point temperature, relative humidity, solar radiation (horizontal infrared, direct normal and diffuse horizontal), luminance, wind speed and direction, sky cover and current weather indicator. Also, the weather file includes information regarding the location (longitude, latitude and elevation), ground temperature and daylight saving. The weather records provided by weather stations are not always complete. For instance, horizontal infrared radiation (I_{ir}) is required by Energyplus for sky radiation modeling. However, it is not a typically observed value reported by the meteorological offices. We have estimated I_{ir} via the following relations:

$$I_{ir} = \sigma \times E_{sky} \times T_{db}^4$$
$$E_{sky} = \left(0.787 + 0.764 \times ln\left(\frac{T_{dp}}{273}\right)\right)\left(1 + 0.224N_c + 0.0035N_c^2 + 0.00028N_c^3\right) \quad \text{(F.21)}$$

where this E_{sky}, T_{db}, T_{dp} and N_c are Sky emissivity, dry bulb temperature, wet bulb temperature and sky cover index. The sky radiance model in Energyplus is based on the empirical model proposed by Perez et al. [261]. Considering the shadowing effect, sky diffuse solar radiation (I_{sky}) can be written as:

$$I_{sky} = R_{hr} \times I_{hr}(I_{dh}, \phi) + R_{dm} \times I_{dm}(I_{dh}, \phi) + R_{cr} \times I_{cr}(I_{dh}, \phi) \quad \text{(F.22)}$$

where I_{dm}, I_{cr} and I_{hr} are isotropic distribution that covers the sky dome, circumsolar brightening concentrated at the center of the sun and a horizontal brightening which is considered as a linear source at the horizon, respectively. These three contributions are functions of sky diffuse solar radiation (I_{dh}) and the surface tilt angle (ϕ). The R_{hr}, R_{dm} and R_{cr} are shadowing coefficients considering how much of the radiation is received by the surface. Since the circumsolar radiation is assumed to be concentrated at the solar disk, R_{cr} is simply the ratio of circumsolar radiance with obstruction to circumsolar radiance without obstruction. For estimating R_{hr}, the horizon line is divided into 24 intervals of equal length. For the case of R_{dm}, the sky dome is divided to 144 sections. R_{hr} and R_{dm} are estimated using:

$$R_{hr} = \frac{\sum_{i=1}^{24} I_{hr}^{i} \times SF_i}{\sum_{i=1}^{24} I_{hr}^{i}} \tag{F.23}$$

$$R_{dom} = \frac{\sum_{i=1}^{24} \sum_{j=1}^{6} I_{dm}^{ij} \times SF_{ij}}{\sum_{j=1}^{6} I_{dm}^{ij}} \tag{F.24}$$

where SF is the sunlight fraction coming from a given interval. We emphasize that the summations are only considered for the intervals that are in front of the surface.

In Energyplus, temperature and wind speed are automatically calculated for the building zones and surface centroids. Given a weather data for a meteorological station at the altitude h_0, the temperature at the altitude z can be estimated using:

$$T_{h=z} = T_{h=h_0} + \lambda \times (H_z - H_b) \tag{F.25}$$

where λ is temperature drop constant and $T_{h=z}$ and $T_{h=b}$ are temperatures at two different altitudes of z and h_0, respectively. H_z is the geopotential altitude as given by $H_z = R_e z/(R_e + z)$ where R_e is the radius of the earth. The relation provided in eq. F.25 can be used to calculated temperature distribution along the height of a high-rise building given the temperature at the meteorological station. Similarly, the wind speed V_z depends on the wind speed at the station through:

$$V_z = V_{met} \times \left(\frac{\delta_{met}}{h_0}\right)^{\varsigma_{met}} \left(\frac{z}{\delta}\right)^{\varsigma} \tag{F.26}$$

where δ, δ_{met}, ς and ς_{met} are wind speed profile boundary layer thickness at the site and meteorological station and wind speed profile exponent at the site and the meteorological station, respectively. These variables are dependent on the roughness characteristics of the surrounding terrain.

Bibliography

[1] *Scanning Electron Microscopy and X-ray Microanalysis.*

[2] American Housing Survey for the United States: 2007. Technical report, U.S. Census Bureau, Current Housing Reports, Series H150/07., U.S. Government Printing Office, Washington, DC, 20401 Printed in (2008), 2008.

[3] Buildings and their Impact on the Environment: A Statistical Summary. Technical report, Green Building Workshop, U.S. Environmental Protection Agency (Revised April 22, 2009), 2009.

[4] ISO 15099:2003. Thermal performance of windows, doors and shading devicesâĂŞdetailed calculations. Technical report, International Organization for Standardization, 2003.

[5] A. Abdelalim, S. Abdallah, K. Easawi, S. Negm, and H. Talaat. Thermal properties of hydrated cement pastes studied by the photoacoustic technique. *Journal of Physics: Conference Series*, 214(1):012136, March 2010.

[6] Mohammad Javad Abdolhosseine Qomi, Arash Noshadravan, Jake Sobstyl, Jameson Toole, J Ferreira, R. J-M. Pellenq, F.-J. Ulm, and Marta C. Gonzalez. Data Analytics for Energy Efficient Retrofits of Cities. *Under Review*, 2014.

[7] M. J. Abdolhosseini Qomi, K. J. Krakowiak, M. Bauchy, K. L. Stewart, R. Shahsavari, D. Jagannathan, D. B. Brommer, A. Baronnet, M. J. Buehler, S. Yip,

F.-J. Ulm, K. J. Van Vliet, and R. J.-M. Pellenq. Combinatorial molecular optimization of cement hydrates. *Nature Communications*, 5:4960 doi: 10.1038/ncomms5960, September 2014.

[8] Mohammad Javad Abdolhosseini Qomi, M. Bauchy, R. J-M. Pellenq, and F-J. Ulm. Applying Tools from Glass Science to Study Calcium-Silicate- Hydrates. In *Mechanics and Physics of Creep, Shrinkage, and Durability of Concrete: A Tribute to Zdenk P. Bazant*, pages 78–85. American Society of Civil Engineers, September 2013.

[9] Mohammad Javad Abdolhosseini Qomi, Franz-Josef Ulm, and Roland J.-M. Pellenq. evidence on the dual nature of aluminum in the calcium-silicate-hydrates based on atomistic simulations. *Journal of the American Ceramic Society*, 95(3):evidence on the dual nature of aluminum in the calcium–silicate–hydrates based on atomistic simulations, 2012.

[10] G. C. Abell. Empirical chemical pseudopotential theory of molecular and metallic bonding. *Physical Review B*, 31(10):6184–6196, May 1985.

[11] Muhannad Abuhaikal. *Nano-ChemoMechanical assessment of Rice Husk Ash cement by wavelength dispersive spectroscopy and nanoindentation.* MSc Thesis, Massachusetts Institute of Technology, 2011.

[12] Paul Acker. Micromechanical analysis of creep and shrinkage mechanisms. In *Creep, Shrinkage, and Durability Mechanics of Concrete and Other Quasi-Brittle Materials*, pages 15–25, London, July 2001. Elsevier.

[13] A. Aghaei, M. J. Abdolhosseini Qomi, M. T. Kazemi, and A. R. Khoei. Stability and size-dependency of CauchyâĂŞBorn hypothesis in three-dimensional applications. *International Journal of Solids and Structures*, 46(9):1925–1936, May 2009.

[14] Bal Krishna Agrawal and G. S. Verma. Lattice Thermal Conductivity of Solid Helium. *Physical Review*, 128(2):603–605, October 1962.

[15] F. Alamdari and G. P. Hammond. Improved data correlations for buoyancy-driven convection in rooms. *Building Services Engineering Research and Technology*, 4(3):106–112, August 1983.

[16] 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. WOS:000270942100020.

[17] Andrew J. Allen, Jeffrey J. Thomas, and Hamlin M. Jennings. composition and density of nanoscale calciumâĂŞsilicateâĂŞhydrate in cement. *Nature Materials*, 6(4):311–316, 2007.

[18] Philip B. Allen and Joseph L. Feldman. Thermal Conductivity of Glasses: Theory and Application to Amorphous Si. *Physical Review Letters*, 62(6):645–648, February 1989.

[19] C. Austen Angell. Insights into Phases of Liquid Water from Study of Its Unusual Glass-Forming Properties. *Science*, 319(5863):582–587, February 2008.

[20] G.r. Anstis, P. Chantikul, B.r. Lawn, and D.b. Marshall. a critical evaluation of indentation techniques for measuring fracture toughness: I, direct crack measurements. *Journal of the American Ceramic Society*, 64(9):533–538, 1981.

[21] Dimitrios Argyris, David R. Cole, and Alberto Striolo. Dynamic Behavior of Interfacial Water at the Silica Surface. *The Journal of Physical Chemistry C*, 113(45):19591–19600, November 2009.

[22] J. T. Armstrong. CITZAF - A package of correction programs for the quantitative electron microbeam X-ray analysis of thick polished materials, thinfilms, and particles. *Microbeam Analysis*, 4(3):177–200, 1995.

[23] B. Auer, R. Kumar, J. R. Schmidt, and J. L. Skinner. Hydrogen bonding and Raman, IR, and 2d-IR spectroscopy of dilute HOD in liquid D2o. *Proceedings of the National Academy of Sciences*, 104(36):14215–14220, September 2007.

[24] H. B. Awbi and A. Hatton. Natural convection from heated room surfaces. *Energy and Buildings*, 30(3):233–244, August 1999.

[25] Hazim B. Awbi. Calculation of convective heat transfer coefficients of room surfaces for natural convection. *Energy and Buildings*, 28(2):219–227, October 1998.

[26] J. Barbier and B. G. Hyde. The structures of the polymorphs of dicalcium silicate, Ca_2SiO_4. *Acta Crystallographica Section B Structural Science*, 41(6):383–390, December 1985.

[27] Ashok Bardhan, Dwight Jaffee, Cynthia Kroll, and Nancy Wallace. Energy efficiency retrofits for U.S. housing: Removing the bottlenecks. *Regional Science and Urban Economics*.

[28] G. T. Barkema and Normand Mousseau. Event-Based Relaxation of Continuous Disordered Systems. *Physical Review Letters*, 77(21):4358–4361, November 1996.

[29] M. Bauchy. Structural, vibrational, and thermal properties of densified silicates: Insights from molecular dynamics. *The Journal of Chemical Physics*, 137(4):044510–044510–12, July 2012.

[30] M. Bauchy and M. Micoulaut. from pockets to channels: Density-controlled diffusion in sodium silicates. *Physical Review B*, 83(18):184118, May 2011.

[31] M. Bauchy and M. Micoulaut. transport anomalies and adaptative pressure-dependent topological constraints in tetrahedral liquids: evidence for a reversibility window analogue. *Physical Review Letters*, 110(9):095501, February 2013.

[32] M. Bauchy, M. J. Abdolhosseini Qomi, C. Bichara, F.-J. Ulm, and R. J.-M. Pellenq. Topological Origin of Fracture Toughening in Complex Solids: the Viewpoint of Rigidity Theory. *arXiv:1410.2916 [cond-mat]*, October 2014. arXiv: 1410.2916.

[33] M. Bauchy, M. J. Abdolhosseini Qomi, F.-J. Ulm, and R. J.-M. Pellenq. Order and disorder in calciumâĂŞsilicateâĂŞhydrate. *The Journal of Chemical Physics*, 140(21):214503, June 2014.

[34] Mathieu Bauchy, Mohammad Javad Abdolhosseini Qomi, Christophe Bichara, Franz-Joseph Ulm, and Roland J.-M. Pellenq. Nanoscale Structure of Cement: Viewpoint of Rigidity Theory. *The Journal of Physical Chemistry C*, 118(23):12485–12493, June 2014.

[35] Hanefi Bayraktar and F. Sezer Turalioglu. A Kriging-based approach for locating a sampling siteâĂŤin the assessment of air quality. *Stochastic Environmental Research and Risk Assessment*, 19(4):301–305, October 2005.

[36] J. J. Beaudoin. Comparison of mechanical properties of compacted calcium hydroxide and portland cement paste systems. *Cement and Concrete Research*, 13(3):319–324, May 1983.

[37] J. J. Beaudoin, P. Gu, and R. E. Myers. the fracture of C-S-H and C-S-H/CH mixtures. *Cement and Concrete Research*, 28(3):341–347, March 1998. WOS:000072784400004.

[38] Moshe E. Ben-Akiva and Steven R. Lerman. *Discrete Choice Analysis: Theory and Application to Travel Demand.* MIT Press, 1985.

[39] D. P. Bentz. Transient plane source measurements of the thermal properties of hydrating cement pastes. *Materials and Structures*, 40(10):1073–1080, December 2007.

[40] R. Berman and J. C. F. Brock. The Effect of Isotopes on Lattice Heat Conduction. I. Lithium Fluoride. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, 289(1416):46–65, December 1965.

[41] Luis Bettencourt and Geoffrey West. A unified theory of urban living. *Nature*, 467(7318):912–913, October 2010.

[42] Somnath Bhowmick and Vijay B. Shenoy. Effect of strain on the thermal conductivity of solids. *The Journal of Chemical Physics*, 125(16):164513, October 2006.

[43] M. BigarÃĬ, A. Guinier, C. MaziÃĺres, M. Regourd, N. Yannaquis, W. Eysbl, Th. Hahn, and E. Woermann. Polymorphism of Tricalcium Silicate and Its Solid Solutions. *Journal of the American Ceramic Society*, 50(11):609–619, November 1967.

[44] B. Blocken, T. Defraeye, D. Derome, and J. Carmeliet. High-resolution CFD simulations for forced convective heat transfer coefficients at the facade of a low-rise building. *Building and Environment*, 44(12):2396–2412, December 2009.

[45] Anselm Blumer, Andrzej Ehrenfeucht, David Haussler, and Manfred K. Warmuth. Occam's Razor. *Information Processing Letters*, 24(6):377–380, April 1987.

[46] P. A. Bonnaud, B. Coasne, and R. J.-M. Pellenq. Molecular simulation of water confined in nanoporous silica. *Journal of Physics: Condensed Matter*, 22(28):284110, July 2010.

[47] Patrick A. Bonnaud, Qing Ji, and Krystyn J. Van Vliet. Effects of elevated temperature on the structure and properties of calciumâĂŞsilicateâĂŞhydrate gels: the role of confined water. *Soft Matter*, 9(28):6418–6429, June 2013.

[48] Heloisa N. Bordallo, Laurence P. Aldridge, and Arnaud Desmedt. Water Dynamics in Hardened Ordinary Portland Cement Paste or Concrete: From Quasielastic Neutron Scattering. *The Journal of Physical Chemistry B*, 110(36):17966–17976, September 2006.

[49] G. E. P. Box and Norman R. Draper. A Basis for the Selection of a Response Surface Design. *Journal of the American Statistical Association*, 54(287):622–654, 1959.

[50] George. E. P. Box and Norman R. Draper. *Empirical model-building and response surfaces*, volume xiv of *Wiley series in probability and mathematical statistics.* John Wiley & Sons, Oxford, England, 1987.

[51] Louis Brand. The Pi theorem of dimensional analysis. *Archive for Rational Mechanics and Analysis*, 1(1):35–45, January 1957.

[52] Donald W. Brenner. Empirical potential for hydrocarbons for use in simulating the chemical vapor deposition of diamond films. *Physical Review B*, 42(15):9458–9471, November 1990.

[53] Carrie Ann Brown. *Toward zero net energy buildings: optimized for energy use and cost.* PhD thesis, Massachusetts Institute of Technology, Cambridge, 2012.

[54] U. Buchenau, N. NÃijcker, and A. J. Dianoux. Neutron Scattering Study of the Low-Frequency Vibrations in Vitreous Silica. *Physical Review Letters*, 53(24):2316–2319, December 1984.

[55] D. G. Cahill and R. O. Pohl. Lattice Vibrations and Heat Transport in Crystals and Glasses. *Annual Review of Physical Chemistry*, 39:93–121, October 1988.

[56] David G. Cahill, S.-M. Lee, and Torbjorn I. Selinder. Thermal conductivity of k-Al2o3 and a-Al2o3 wear-resistant coatings. *Journal of Applied Physics*, 83(11):5783–5786, June 1998.

[57] David G. Cahill and R. O. Pohl. Heat flow and lattice vibrations in glasses. *Solid State Communications*, 70(10):927–930, June 1989.

[58] Joseph Callaway. Model for Lattice Thermal Conductivity at Low Temperatures. *Physical Review*, 113(4):1046–1051, February 1959.

[59] D. Caprion, J. Matsui, and H. R. Schober. Dynamic Heterogeneity of Relaxations in Glasses and Liquids. *Physical Review Letters*, 85(20):4293–4296, November 2000.

[60] Sophie Cariou, Franz-Josef Ulm, and Luc Dormieux. HardnessâĂŞpacking density scaling relations for cohesive-frictional porous materials. *Journal of the Mechanics and Physics of Solids*, 56(3):924–952, March 2008.

[61] H. T. Ceylan and G. E. Meyers. Long-Time Solutions to Heat-Conduction Transients with Time-Dependent Inputs. *J. Heat Transfer; (United States)*, 102, February 1980.

[62] Alan J Chapman. *Heat transfer*. Macmillan, New York, 1974.

[63] S. Chatterji. concrete durability and Cao/Sio2 mole ratio of csh. *Cement and Concrete Research*, 25(5):929–932, July 1995. WOS:A1995RH49500003.

[64] Pinaki Chaudhuri, Ludovic Berthier, and Walter Kob. Universal Nature of Particle Displacements close to Glass and Jamming Transitions. *Physical Review Letters*, 99(6):060604, August 2007.

[65] Jeffrey J. Chen, Jeffrey J. Thomas, Hal F.W. Taylor, and Hamlin M. Jennings. solubility and structure of calcium silicate hydrate. *Cement and Concrete Research*, 34(9):1499–1519, September 2004.

[66] S.-H. Chen, P. Gallo, and M.-C. Bellissent-Funel. Slow dynamics of interfacial water. *Canadian Journal of Physics*, 73(11-12):703–709, November 1995.

[67] W.-S. Chiang, G. Ferraro, E. Fratini, F. Ridi, Y.-Q. Yeh, U-S. Jeng, S.-H. Chen, and P. Baglioni. Multiscale structure of calcium- and magnesium-silicate-hydrate gels. *Journal of Materials Chemistry A*, 2(32):12991, July 2014.

[68] Wei-Shan Chiang, Emiliano Fratini, Piero Baglioni, Dazhi Liu, and Sow-Hsin Chen. Microstructure Determination of Calcium-Silicate-Hydrate Globules by Small-Angle Neutron Scattering. *The Journal of Physical Chemistry C*, 116(8):5055–5061, March 2012.

[69] Franco Chingcuanco and Eric J. Miller. A microsimulation model of urban energy use: Modelling residential space heating demand in ILUTE. *Computers, Environment and Urban Systems*, 36(2):186–194, March 2012.

[70] Shieu-Ming Chou, Tian-Shyug Lee, Yuehjen E. Shao, and I-Fei Chen. Mining the breast cancer pattern using artificial neural networks and multivariate adaptive regression splines. *Expert Systems with Applications*, 27(1):133–142, July 2004.

[71] S. ChromÃ¡ and I. Maki. Determination of the quantitative phase composition of alite and belite in portland cement clinker by microphotometry. *Cement and Concrete Research*, 12(4):511–516, July 1982.

[72] Sergey V. Churakov. Structural position of H2o molecules and hydrogen bonding in anomalous 11 ÃĚ tobermorite. *American Mineralogist*, 94(1):156–165, January 2009.

[73] Giancarlo Cicero, Jeffrey C. Grossman, Eric Schwegler, Francois Gygi, and Giulia Galli. Water Confined in Nanotubes and between Graphene Sheets: A First Principle Study. *Journal of the American Chemical Society*, 130(6):1871–1878, February 2008.

[74] Christoph Clauser and Ernst Huenges. Thermal Conductivity of Rocks and Minerals. In Thomas J. Ahrens, editor, *Rock Physics & Phase Relations*, volume 1, pages 105–126. American Geophysical Union, New York, NY, first edition, July 1995.

[75] A. Clauset, C. Shalizi, and M. Newman. Power-Law Distributions in Empirical Data. *SIAM Review*, 51(4):661–703, November 2009.

[76] Benoit Coasne, Anne Galarneau, Roland J. M. Pellenq, and Francesco Di Renzo. Adsorption, intrusion and freezing in porous silica: the view from the nanoscale. *Chemical Society Reviews*, 42(9):4141–4171, April 2013.

[77] David A Coley and Stefan Schukat. Low-energy design: combining computer-based optimisation and human judgement. *Building and Environment*, 37(12):1241–1247, December 2002.

[78] Xiandong Cong and R.James Kirkpatrick. 29si MAS NMR study of the structure of calcium silicate hydrate. *Advanced Cement Based Materials*, 3(3âĂŞ4):144–156, April 1996.

[79] Andrew R. Conn, Nicholas I. M. Gould, and Philippe L. Toint. *Trust-Region Methods*. Society for Industrial and Applied Mathematics, January 1987.

[80] G Constantinides. *Invariant mechanical properties of calcium-silicate-hydrates (C-S-H) in cement-based materials: Instrumented nanoindentation and micro-poromechanical modeling*. PhD Thesis, Massachusetts Institute of Technology, 2006.

[81] G. Constantinides and F. J. Ulm. the effect of two types of C-S-H on the elasticity of cement-based materials: Results from nanoindentation and micromechanical modeling. *Cement and Concrete Research*, 34(1):67–80, January 2004. WOS:000224016100010.

[82] Georgios Constantinides and Franz-Josef Ulm. the nanogranular nature of CâĂŞSâĂŞH. *Journal of the Mechanics and Physics of Solids*, 55(1):64–90, January 2007.

[83] FranÃğois-Xavier Coudert, Rodolphe Vuilleumier, and Anne Boutin. Dipole Moment, Hydrogen Bonding and IR Spectrum of Confined Water. *ChemPhysChem*, 7(12):2464–2467, 2006.

[84] Drury B. Crawley, Linda K. Lawrie, Frederick C. Winkelmann, W. F. Buhl, Y. Joe Huang, Curtis O. Pedersen, Richard K. Strand, Richard J. Liesen, Daniel E. Fisher, Michael J. Witte, and Jason Glazer. EnergyPlus: creating a new-generation building energy simulation program. *Energy and Buildings*, 33(4):319–331, April 2001.

[85] BenoÃőt Creton, Daniel Bougeard, Konstantin S. Smirnov, Jean Guilment, and Olivier Poncelet. Molecular dynamics study of hydrated imogolite : 2. Struc-

ture and dynamics of confined water. *Physical Chemistry Chemical Physics*, 10(32):4879, 2008.

[86] Randall T. Cygan, Jian-Jie Liang, and Andrey G. Kalinichev. Molecular Models of Hydroxide, Oxyhydroxide, and Clay Phases and the Development of a General Force Field. *The Journal of Physical Chemistry B*, 108(4):1255–1266, January 2004.

[87] Jennifer L. Dashnau, Nathaniel V. Nucci, Kim A. Sharp, and Jane M. Vanderkooi. Hydrogen Bonding and the Cryoprotective Properties of Glycerol/Water Mixtures. *The Journal of Physical Chemistry B*, 110(27):13670–13677, July 2006.

[88] W. A. Day and Morton E. Gurtin. On the symmetry of the conductivity tensor and other restrictions in the nonlinear theory of heat conduction. *Archive for Rational Mechanics and Analysis*, 33(1):26–32, January 1969.

[89] ÃĄngeles G De La Torre, SebastiÃąn Bruque, Javier Campo, and Miguel A. G Aranda. The superstructure of C3s from synchrotron and neutron powder diffraction and its role in quantitative phase analyses. *Cement and Concrete Research*, 32(9):1347–1356, September 2002.

[90] G. De Schutter and L. Taerwe. Specific heat and thermal diffusivity of hardening concrete. *Magazine of Concrete Research*, 47(172):203–208, January 1995.

[91] Olivier L. de Weck and Marshall B. Jones. Isoperformance: Analysis and design of complex systems with desired outcomes. *Systems Engineering*, 9(1):45–61, March 2006.

[92] Pablo G. Debenedetti and Frank H. Stillinger. Supercooled liquids and the glass transition. *Nature*, 410(6825):259–267, March 2001.

[93] Antoine Delafargue and Franz-Josef Ulm. explicit approximations of the indentation modulus of elastically orthotropic solids for conical indenters. *International journal of solids and structures*, 41(26):7351–7360, 2004.

[94] Ramazan Demirboga and Rustem Gul. The effects of expanded perlite aggregate, silica fume and fly ash on the thermal conductivity of lightweight concrete. *Cement and Concrete Research*, 33(5):723–727, May 2003.

[95] Ramazan DemirboÇğa. Influence of mineral admixtures on thermal conductivity and compressive strength of mortar. *Energy and Buildings*, 35(2):189–192, February 2003.

[96] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *JOURNAL OF THE ROYAL STATISTICAL SOCIETY, SERIES B*, 39(1):1–38, 1977.

[97] Jean Dixmier. Hole generation of prepeaks in diffraction patterns of glasses. *Journal de Physique I*, 2(6):1011–1027, June 1992.

[98] Jorge S. Dolado, Michael Griebel, and Jan Hamaekers. a molecular dynamic study of cementitious calcium silicate hydrate (CâĂŞSâĂŞH) gels. *Journal of the American Ceramic Society*, 90(12):3938–3942, 2007.

[99] Luc Dormieux, Djimedo Kondo, and Franz-Jozef Ulm. *Microporomechanics*. Wiley, Chichester, West Sussex, England ; Hoboken, NJ, 1 edition edition, August 2006.

[100] E. Duval, A. Boukenter, and T. Achibat. Vibrational dynamics and the structure of glasses. *Journal of Physics: Condensed Matter*, 2(51):10227, December 1990.

[101] Davoud Ebrahimi, Roland J.-M. Pellenq, and Andrew J. Whittle. nanoscale elastic properties of montmorillonite upon water adsorption. *Langmuir*, 28(49):16855–16863, December 2012.

[102] Davoud Ebrahimi, Andrew J. Whittle, and Roland J.-M. Pellenq. Mesoscale properties of clay aggregates from potential of mean force representation of interactions between nanoplatelets. *The Journal of Chemical Physics*, 140(15):154309, April 2014.

[103] M. D. Ediger. Spatially Heterogeneous Dynamics in Supercooled Liquids. *Annual Review of Physical Chemistry*, 51(1):99–128, 2000.

[104] Bryan Eisenhower, Zheng ONeill, Satish Narayanan, Vladimir A. Fonoberov, and Igor Mezic. A methodology for meta-model based optimization in building energy models. *Energy and Buildings*, 47:292–301, April 2012.

[105] T. Ekman and G. Kubin. Nonlinear prediction of mobile radio channels: measurements and MARS model designs. In , *1999 IEEE International Conference on Acoustics, Speech, and Signal Processing, 1999. Proceedings*, volume 5, pages 2667–2670 vol.5, 1999.

[106] H. El-Didamony, A. M. Sharara, I. M. Helmy, and S. Abd El-Aleem. Hydration characteristics of Îš-C2s in the presence of some accelerators. *Cement and Concrete Research*, 26(8):1179–1187, August 1996.

[107] S. R. Elliott. origin of the first sharp diffraction peak in the structure factor of covalent glasses. *Physical Review Letters*, 67(6):711–714, August 1991.

[108] S.R. Elliott. extended-range order, interstitial voids and the first sharp diffraction peak of network glasses. *Journal of Non-Crystalline Solids*, 182(1âĂŞ2):40–48, March 1995.

[109] D. Engberg, A. Wischnewski, U. Buchenau, L. BÃűrjesson, A. J. Dianoux, A. P. Sokolov, and L. M. Torell. Origin of the boson peak in a network glass B2o3. *Physical Review B*, 59(6):4053–4057, February 1999.

[110] Jeffrey R. Errington and Pablo G. Debenedetti. Relationship between structural order and the anomalies of liquid water. *Nature*, 409(6818):318–321, January 2001.

[111] J. D. Eshelby. The Determination of the Elastic Field of an Ellipsoidal Inclusion, and Related Problems. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, 241(1226):376–396, August 1957.

[112] A. Eucken. ÃIJber die TemperaturabhÃd'ngigkeit der WÃd'rmeleitfÃd'higkeit fester Nichtmetalle. *Annalen der Physik*, 339(2):185–221, January 1911.

[113] Michael Evans and Tim Swartz. *Approximating integrals via Monte Carlo and deterministic methods*. Oxford University Press, Oxford; New York, 2000.

[114] Feng Xiuji, Min Xinmin, and Tao Congxi. Study on the structure and characteristic of dicalcium silicate with quantum chemistry calculations. *Cement and Concrete Research*, 24(7):1311–1316, 1994.

[115] RenÃl' Feret. *Ã L'tude expÃlrimentale du ciment armÃl'*. Gauthier-Villars, 1906.

[116] Joseph C. Fogarty, Hasan Metin Aktulga, Ananth Y. Grama, Adri C. T. van Duin, and Sagar A. Pandit. A reactive molecular dynamics simulation of the silica-water interface. *The Journal of Chemical Physics*, 132(17):174704–174704–10, May 2010.

[117] StÃl'phane Fohanno and Guillaume Polidori. Modelling of natural convective heat transfer at an internal surface. *Energy and Buildings*, 38(5):548–553, May 2006.

[118] Emiliano Fratini, Antonio Faraone, Francesca Ridi, Sow-Hsin Chen, and Piero Baglioni. Hydration Water Dynamics in Tricalcium Silicate Pastes by Time-Resolved Incoherent Elastic Neutron Scattering. *The Journal of Physical Chemistry C*, 117(14):7358–7364, April 2013.

[119] Jerome H. Friedman. Multivariate Adaptive Regression Splines. *The Annals of Statistics*, 19(1):1–67, March 1991. Mathematical Reviews number (MathSciNet): MR1091842; Zentralblatt MATH identifier: 0765.62064.

[120] Jerome H. Friedman and Werner Stuetzle. Projection Pursuit Regression. *Journal of the American Statistical Association*, 76(376):817–823, December 1981.

[121] K. S. Ravi Chandran G. Constantinides. grid indentation analysis of composite microstructure and mechanics: Principles and validation. *Materials Science and*

Engineering a-Structural Materials Properties Microstructure and Processing, 430:189–202.

[122] Xavier Gabaix. Zipf's Law for Cities: An Explanation. *The Quarterly Journal of Economics*, 114(3):739–767, August 1999.

[123] J. D. Gale. GULP: A computer program for the symmetry-adapted simulation of solids. *Journal of the Chemical Society-Faraday Transactions*, 93(4):629–637, February 1997. WOS:A1997WK27100016.

[124] J. D. Gale and A. L. Rohl. The General Utility Lattice Program (GULP). *Molecular Simulation*, 29(5):291–341, April 2003. WOS:000182882300001.

[125] P. Gallo, M. Rovere, and E. Spohr. Supercooled Confined Water and the Mode Coupling Crossover Temperature. *Physical Review Letters*, 85(20):4317–4320, November 2000.

[126] F.P. Ganneau, G. Constantinides, and F.-J. Ulm. dual-indentation technique for the assessment of strength properties of cohesive-frictional materials. *International Journal of Solids and Structures*, 43(6):1727–1745, March 2006.

[127] B. Gathier. *Multiscale strength homogenization âĂŞ application to shale nanoindentation.* MSc Thesis, Massachusetts Institute of Technology, 2008.

[128] Siavash Ghabezloo. Micromechanics analysis of thermal expansion and thermal pressurization of a hardened cement paste. *Cement and Concrete Research*, 41(5):520–532, May 2011.

[129] Subir Ghosh and Calyampudi Radhakrishna Rao. *Handbook of Statistics 13: Design and analysis of experiments.* Elsevier, Amsterdam, 1996.

[130] Mark N Gibbs. *Bayesian Gaussian processes for regression and classification.* PhD Thesis, University of Cambridge, 1997.

[131] Anthony A Giunta, Steven F Wojtkiewicz, and Michael S Eldred. Overview of modern design of experiments methods for computational simulations. In

Proceedings of the 41st AIAA Aerospace Sciences Meeting and Exhibit, pages AIAA–2003–0649, 2003.

[132] Ahmed Gmira. *etude texturale et thermodynamique dâĂŹhydrates modÃĺles du ciment.* PhD Thesis, UNIVERSITÃĽ DâĂŹORLÃĽANS, 2004.

[133] Pierre Goovaerts. *Geostatistics for natural resources evaluation.* Oxford University Press, New York, 1997.

[134] Sylvain Grangeon, Francis Claret, Catherine Lerouge, Fabienne Warmont, Tsutomu Sato, Sohtaro Anraku, Chiya Numako, Yannick Linard, and Bruno Lanson. on the nature of structural disorder in calcium silicate hydrates with a calcium/silicon ratio similar to tobermorite. *Cement and Concrete Research*, 52:31–37, October 2013.

[135] T. S. Grigera, V. MartÃŋn-Mayor, G. Parisi, and P. Verrocchio. Phonon interpretation of the âĂŸboson peakâĂŹ in supercooled liquids. *Nature*, 422(6929):289–292, March 2003.

[136] Bertrand Guillot and Yves Guissani. Boson Peak and High Frequency Modes in Amorphous Silica. *Physical Review Letters*, 78(12):2401–2404, March 1997.

[137] Bertil Halle and Monika Davidovic. Biomolecular hydration: From water dynamics to hydrodynamics. *Proceedings of the National Academy of Sciences*, 100(21):12135–12140, October 2003.

[138] J. H. Halton. Algorithm 247: Radical-inverse Quasi-random Point Sequence. *Communication of the ACM*, 7(12):701–702, December 1964.

[139] Sa Hamid. the crystal-structure of the 11a natural tobermorite Ca2.25[Si3o7.5(OH)1.5].1h2o. *Zeitschrift Fur Kristallographie*, 154(3-4):189–198, 1981. WOS:A1981KW47700002.

[140] A. D. Hammerich and V. Buch. An alternative near-neighbor definition of hydrogen bonding in water. *The Journal of Chemical Physics*, 128(11):111101, 2008.

[141] J. M. Hammersley. Monte Carlo Methods for Solving Multivariable Problems. *Annals of the New York Academy of Sciences*, 86(3):844–874, 1960.

[142] Su Hao, Wing Kam Liu, Brian Moran, Franck Vernerey, and Gregory B. Olson. Multi-scale constitutive model and computational framework for the design of ultra-high strength, high toughness steels. *Computer Methods in Applied Mechanics and Engineering*, 193(17âĂŞ20):1865–1908, May 2004.

[143] Hiroshi Hatta and Minoru Taya. Effective thermal conductivity of a misoriented short fiber composite. *Journal of Applied Physics*, 58(7):2478–2486, October 1985.

[144] Simon Haykin. *Neural Networks: A Comprehensive Foundation.* Prentice Hall PTR, Upper Saddle River, NJ, USA, 1st edition, 1994.

[145] A Hedayat, N. J. A Sloane, and John Stufken. *Orthogonal arrays: theory and applications.* Springer, New York, 1999.

[146] K. F. J. Heinrich and D. Newbury. *Electron Probe Quantitation.* Springer, June 1991.

[147] Bruce S. Hemingway. Quartz; heat capacities from 340 to 1000 K and revised values for the thermodynamic properties. *American Mineralogist*, 72(3-4):273–279, April 1987.

[148] D Henderson, M and H Gutowsky, S. A nuclear magnetic resonance determination of the hydrogen positions in Ca(OH)2. *American Mineralogist*, 47:1231–1251, July 1962.

[149] L. Hong, B. Begen, A. Kisliuk, C. Alba-Simionesco, V. N. Novikov, and A. P. Sokolov. Pressure and density dependence of the boson peak in polymers. *Physical Review B*, 78(13):134201, October 2008.

[150] Harold Hotelling and Margaret Richards Pabst. Rank Correlation and Tests of Significance Involving No Assumption of Normality. *The Annals of Mathematical Statistics*, 7(1):29–43, March 1936.

[151] Hoyt Clarke Hottel and Adel F. Sarofim. *Radiative transfer.* McGraw-Hill, 1967.

[152] Yi Hu and P.C. Loizou. Evaluation of Objective Quality Measures for Speech Enhancement. *IEEE Transactions on Audio, Speech, and Language Processing*, 16(1):229–238, 2008.

[153] Zhonghan Hu and Claudio J. Margulis. Heterogeneity in a room-temperature ionic liquid: Persistent local environments and the red-edge effect. *Proceedings of the National Academy of Sciences of the United States of America*, 103(4):831–836, January 2006.

[154] K. E. Hudson and G. W. Groves. The structure of alite in Portland cement clinker âĂŤ TEM evidence. *Cement and Concrete Research*, 12(1):61–68, January 1982.

[155] Janelle S. Hygh, Joseph F. DeCarolis, David B. Hill, and S. Ranji Ranjithan. Multivariate regression as an energy assessment tool in early building design. *Building and Environment*, 57:165–175, November 2012.

[156] Ronald L. Iman and W. J. Conover. A distribution-free approach to inducing rank correlation among input variables. *Communications in Statistics - Simulation and Computation*, 11(3):311–334, 1982.

[157] N. Ito. Field Experiment Study on the Convective Heat Transfer Coefficient on Exterior Surface of a Building. *ASHRAE Trans.; (United States)*, 78, January 1972.

[158] H. Iyetomi and P. Vashishta. Atomic-size effects on medium-range order in glasses. *Physical Review B*, 47(6):3063–3069, February 1993.

[159] L. R. Corrales J. Du. Compositional dependence of the first sharp diffraction peaks in alkali silicate glasses: A molecular dynamics study. *Journal of Non-Crystalline Solids*, pages 3255–3269, 2006.

[160] J. Jakumeit, M. Herdy, and M. Nitsche. Parameter optimization of the sheet metal forming process using an iterative parallel Kriging algorithm. *Structural and Multidisciplinary Optimization*, 29(6):498–507, June 2005.

[161] Gints Jekabsons. ARESLab: adaptive regression spline toolbox for Matlab/Octave. Technical report, Institute of Applied Computer Systems, Riga Technical University, Riga, Latvia, 2010.

[162] R. Jin, W. Chen, and T. W. Simpson. Comparative studies of metamodelling techniques under multiple modelling criteria. *Structural and Multidisciplinary Optimization*, 23(1):1–13, December 2001.

[163] M. E. Johnson, L. M. Moore, and D. Ylvisaker. Minimax and maximin distance designs. *Journal of Statistical Planning and Inference*, 26(2):131–148, October 1990.

[164] Donald R. Jones, Matthias Schonlau, and William J. Welch. Efficient Global Optimization of Expensive Black-Box Functions. *Journal of Global Optimization*, 13(4):455–492, December 1998.

[165] K. H. Jost, B. Ziemer, and R. Seydel. Redetermination of the structure of Îš-dicalcium silicate. *Acta Crystallographica Section B Structural Crystallography and Crystal Chemistry*, 33(6):1696–1700, June 1977.

[166] P. Jund and R. Jullien. Densification effects on the Boson peak in vitreous silica: A molecular-dynamics study. *The Journal of Chemical Physics*, 113(7):2768–2771, August 2000.

[167] Philippe Jund and RÃľmi Jullien. Molecular-dynamics calculation of the thermal conductivity of vitreous silica. *Physical Review B*, 59(21):13707–13711, June 1999.

[168] Andrey G. Kalinichev, Jianwei Wang, and R. James Kirkpatrick. Molecular dynamics modeling of the structure, dynamics and energetics of mineralâĂŞwater

interfaces: Application to cement materials. *Cement and Concrete Research*, 37(3):337–347, March 2007.

[169] M. Kavgic, A. Mavrogianni, D. Mumovic, A. Summerfield, Z. Stevanovic, and M. Djurovic-Petrovic. A review of bottom-up building stock models for energy consumption in the residential sector. *Building and Environment*, 45(7):1683–1697, July 2010.

[170] Irfan Kaymaz. Application of kriging method to structural reliability problems. *Structural Safety*, 27(2):133–151, April 2005.

[171] A. R. Khoei, P. Ghahremani, M. J. Abdolhosseini Qomi, and P. Banihashemi. Stability and size-dependency of temperature-related CauchyâĂŞBorn hypothesis. *Computational Materials Science*, 50(5):1731–1743, March 2011.

[172] A. R. Khoei, M. J. Abdolhosseini Qomi, M. T. Kazemi, and A. Aghaei. An investigation on the validity of CauchyâĂŞBorn hypothesis using Sutton-Chen many-body potential. *Computational Materials Science*, 44(3):999–1006, January 2009.

[173] Kook-Han Kim, Sang-Eun Jeon, Jin-Keun Kim, and Sungchul Yang. An experimental study on thermal conductivity of concrete. *Cement and Concrete Research*, 33(3):363–371, March 2003.

[174] K Kishen. On latin and hyper-graeco cubes and hypercubes. *Current Science*, 11:98–99, 1942.

[175] Charles Kittel. *Introduction to Solid State Physics*. Wiley, Hoboken, NJ, 8 edition edition, November 2004.

[176] Jack P. C. Kleijnen. Kriging metamodeling in simulation: A review. *European Journal of Operational Research*, 192(3):707–716, February 2009.

[177] S.A Klein. TRNSYS 17: a transient system simulation program, 2010.

[178] Walter Kob. Computer simulations of supercooled liquids and glasses. *Journal of Physics: Condensed Matter*, 11(10):R85, March 1999.

[179] Jeremy Z. Kolter and Joseph Ferreira Jr. A large-scale study on predicting and contextualizing building energy usage. *Proceedings of the Twenty-Fifth AAAI Conference on Artificial Intelligence*, August 2011. National Science Foundation (U.S.) (NSF Computing Innovation Fellowship).

[180] J.-P. Korb, P.J. McDonald, L. Monteilhet, A.G. Kalinichev, and R.J. Kirkpatrick. Comparison of proton field-cycling relaxometry and molecular dynamics simulations for protonâĂŞwater surface dynamics in cement-based materials. *Cement and Concrete Research*, 37(3):348–350, March 2007.

[181] J.-P. Korb, L. Monteilhet, P.J. McDonald, and J. Mitchell. Microstructure and texture of hydrated cement-based materials: A proton field cycling relaxometry approach. *Cement and Concrete Research*, 37(3):295–302, March 2007.

[182] Dg Krige and G. Matheron. 2-Dimensional Weighted Moving Average Trend Surfaces for Ore Valuation. *Journal of the South African Institute of Mining and Metallurgy*, 67(12):687–&, 1967. WOS:A19679853900005.

[183] Anna Kuffel and Jan Zielkiewicz. Why the Solvation Water around Proteins Is More Dense than Bulk Water. *The Journal of Physical Chemistry B*, 116(40):12113–12124, October 2012.

[184] R. Kumar, J. R. Schmidt, and J. L. Skinner. Hydrogen bonding definitions and dynamics in liquid water. *Journal of Chemical Physics*, 126:4107, May 2007.

[185] I.-Feng W. Kuo and Christopher J. Mundy. An ab Initio Molecular Dynamics Study of the Aqueous Liquid-Vapor Interface. *Science*, 303(5658):658–660, January 2004.

[186] Takahiro Kuroki, Noboru Kagawa, Harumi Endo, Seizou Tsuruno, and Joseph W. Magee. Specific Heat Capacity at Constant Volume for Water,

Methanol, and Their Mixtures at Temperatures from 300 K to 400 K and Pressures to 20 MPaâĂă. *Journal of Chemical & Engineering Data*, 46(5):1101–1106, September 2001.

[187] E. S. Landry, M. I. Hussein, and A. J. H. McGaughey. Complex superlattice unit cell designs for reduced thermal conductivity. *Physical Review B*, 77(18):184302, May 2008.

[188] Timothy T. Lau, Akihiro Kushima, and Sidney Yip. Atomistic Simulation of Creep in a Nanocrystal. *Physical Review Letters*, 104(17):175501, April 2010.

[189] D. C. G. Law, M. L. Serre, G. Christakos, P. A. Leone, and W. C. Miller. Spatial analysis and mapping of sexually transmitted diseases to optimise intervention and prevention strategies. *Sexually Transmitted Infections*, 80(4):294–299, August 2004.

[190] Stephen Leary, Atul Bhaskar, and Andy Keane. Optimal orthogonal-array-based latin hypercubes. *Journal of Applied Statistics*, 30(5):585–598, 2003.

[191] Tian-Shyug Lee and I-Fei Chen. A two-stage hybrid credit scoring model using artificial neural networks and multivariate adaptive regression splines. *Expert Systems with Applications*, 28(4):743–752, May 2005.

[192] Young Hee Lee, R. Biswas, C. M. Soukoulis, C. Z. Wang, C. T. Chan, and K. M. Ho. Molecular-dynamics simulation of thermal conductivity in amorphous silicon. *Physical Review B*, 43(8):6573–6580, March 1991.

[193] Hua Li, Wei-Shan Chiang, Emiliano Fratini, Francesca Ridi, Francesco Bausi, Piero Baglioni, Madhu Tyagi, and Sow-Hsin Chen. Dynamic crossover in hydration water of curing cement paste: the effect of superplasticizer. *Journal of Physics: Condensed Matter*, 24(6):064108, February 2012.

[194] Hua Li, Emiliano Fratini, Wei-Shan Chiang, Piero Baglioni, Eugene Mamontov, and Sow-Hsin Chen. Dynamic behavior of hydration water in calcium-silicate-

hydrate gel: A quasielastic neutron scattering spectroscopy investigation. *Physical Review E*, 86(6):061505, December 2012.

[195] Shao-Chun Li, Zhenrong Zhang, Daniel Sheppard, Bruce D. Kay, J. M. White, Yingge Du, Igor Lyubinetsky, Graeme Henkelman, and Zdenek Dohnalek. Intrinsic Diffusion of Hydrogen on Rutile TiO2(110). *Journal of the American Chemical Society*, 130(28):9080–9088, July 2008.

[196] John H Lienhard. *A heat transfer textbook*. Prentice-Hall, Englewood Cliffs, N.J., 1987.

[197] Lennart Ljung. *System identification: theory for the user*. Prentice-Hall, Upper Saddle River, NJ, 2009.

[198] SÃÿren Nymand Lophaven, Hans Bruun Nielsen, and SÃÿndergaard Jacob. Aspects of the matlab toolbox DACE. Technical report, DTU, Technical University of Denmark, 2002.

[199] D. L. Loveday and A. H. Taki. Convective heat transfer coefficients at a plane surface on a full-scale building facade. *International Journal of Heat and Mass Transfer*, 39(8):1729–1742, May 1996.

[200] Ralf Ludwig. Water: From Clusters to the Bulk. *Angewandte Chemie International Edition*, 40(10):1808–1827, 2001.

[201] Alenka Luzar and David Chandler. Effect of Environment on Hydrogen Bond Dynamics in Liquid Water. *Physical Review Letters*, 76(6):928–931, February 1996.

[202] Alenka Luzar and David Chandler. Hydrogen-bond kinetics in liquid water. *Nature*, 379(6560):55–57, January 1996.

[203] Iwao Maki and Stanislav ChromÃ¡. Microscopic study on the polymorphism of Ca3sio5. *Cement and Concrete Research*, 8(4):407–414, July 1978.

[204] JohnÂăM. Maloney, Eric Lehnhardt, AlexandraÂăF. Long, and KrystynÂăJ. VanÂăVliet. Mechanical Fluidity of Fully Suspended Biological Cells. *Biophysical Journal*, 105(8):1767–1777, October 2013.

[205] L. E. Malvern. INTRODUCTION TO THE MECHANICS OF A CONTINUOUS MEDIUM. 1969.

[206] P. Mandaliev, R. Daehn, J. Tits, B. Wehrli, and E. Wieland. EXAFS study of Nd(III) uptake by amorphous calcium silicate hydrates (C-S-H). *Journal of Colloid and Interface Science*, 342(1):1–7, February 2010. WOS:000273665900001.

[207] Massimiliano Manfren, Niccolo Aste, and Reza Moshksar. Calibration and uncertainty analysis for computer models âĂŞ A meta-model based approach for integrated building energy simulation. *Applied Energy*, 103:627–641, March 2013.

[208] H. Manzano, J. S. Dolado, and A. Ayuela. aluminum Incorporation to Dreierketten Silicate Chains. *The Journal of Physical Chemistry B*, 113(9):2832–2839, March 2009.

[209] H. Manzano, J.S. Dolado, and A. Ayuela. elastic properties of the main species present in Portland cement pastes. *Acta Materialia*, 57(5):1666–1674, March 2009.

[210] Hegoi Manzano, Engin Durgun, Mohammed Javad Abdolhosseine Qomi, Franz-Josef Ulm, Roland. J. M. Pellenq, and Jeffrey. C. Grossman. impact of chemical impurities on the crystalline cement clinker phases determined by atomistic simulations. *Crystal Growth & Design*, 11(7):2964–2972, July 2011.

[211] Hegoi Manzano, Sina Moeini, Francis Marinelli, Adri C. T. 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, February 2012.

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

[213] Jay Martin and Timothy Simpson. Use of Kriging Models to Approximate Deterministic Computer Models. *AIAA Journal*, 43:853–863, April 2005.

[214] Ippei Maruyama and Go Igarashi. Cement Reaction and Resultant Physical Properties of Cement Paste. *Journal of Advanced Concrete Technology*, 12(6):200–213, 2014.

[215] E. Masoero, E. Del Gado, R. J.-M. Pellenq, F.-J. Ulm, and S. Yip. nanostructure and nanomechanics of cement: polydisperse colloidal packing. *Physical Review Letters*, 109(15):155503, October 2012.

[216] Georges Matheron. Principles of geostatistics. *Economic Geology*, 58(8):1246–1266, December 1963.

[217] Thomas Matschei, Barbara Lothenbach, and Fredrik P. Glasser. Thermodynamic properties of Portland cement hydrates in the system CaOâĂŞAl2o3âĂŞSiO2âĂŞCaSO4âĂŞCaCO3âĂŞH2o. *Cement and Concrete Research*, 37(10):1379–1410, October 2007.

[218] Masakazu Matsumoto, Shinji Saito, and Iwao Ohmine. Molecular dynamics simulation of the ice nucleation and growth process leading to water freezing. *Nature*, 416(6879):409–413, March 2002.

[219] P.J. McDonald, V. Rodin, and A. Valori. Characterisation of intra- and inter-CâĂŞSâĂŞH gel pore water in white cement based on an analysis of NMR signal amplitudes as a function of water content. *Cement and Concrete Research*, 40(12):1656–1663, December 2010.

[220] M. D. McKay, R. J. Beckman, and W. J. Conover. Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code. *Technometrics*, 21(2):239–245, 1979.

[221] Geoffrey McLachlan and David Peel. *Finite Mixture Models.* Wiley-Interscience, 1 edition, October 2000.

[222] Donald A McQuarrie. *Statistical mechanics.* Harper & Row, New York, first edition, November 1975.

[223] Q. Mei, C. J. Benmore, S. Sen, R. Sharma, and J. L. Yarger. intermediate range order in vitreous silica from a partial structure factor analysis. *Physical Review B*, 78(14):144204, October 2008.

[224] Cagla Meral, C.J. Benmore, and Paulo J.M. Monteiro. the study of disorder and nanocrystallinity in CâĂŞSâĂŞH, supplementary cementitious materials and geopolymers using pair distribution function analysis. *Cement and Concrete Research*, 41(7):696–710, July 2011.

[225] Nicholas Metropolis and S. Ulam. The Monte Carlo Method. *Journal of the American Statistical Association*, 44(247):335–341, September 1949.

[226] Ryszard S. Michalski, Ryszard StanisÅĆaw Michalski, Jaime Guillermo Carbonell, and Tom Michael Mitchell. *Machine Learning: An Artificial Intelligence Approach.* Morgan Kaufmann, 1986.

[227] Matthieu Micoulaut and Mathieu Bauchy. anomalies of the first sharp diffraction peak in network glasses: Evidence for correlations with dynamic and rigidity properties. *physica status solidi (b)*, 250(5):976–982, 2013.

[228] C. M. Midgley. The crystal structure of Îš dicalcium silicate. *Acta Crystallographica*, 5(3):307–312, May 1952.

[229] B. Mihailova, N. Zotov, M. Marinov, J. Nikolov, and L. Konstantinov. Vibrational spectra of rings in silicate glasses. *Journal of Non-Crystalline Solids*, 168(3):265–274, March 1994.

[230] D. Mikulic, B. Milovanovic, and I. Gabrijel. Analysis of Thermal Properties of Cement Paste During Setting and Hardening. In Oguz Gunes and YÄślmaz

Akkaya, editors, *Nondestructive Testing of Materials and Structures*, number 6 in RILEM Bookseries, pages 465–471. Springer Netherlands, January 2013.

[231] Mahalia Miller, Christopher Bobko, Matthieu Vandamme, and Franz-Josef Ulm. Surface roughness criteria for cement paste nanoindentation. *Cement and Concrete Research*, 38(4):467–476, April 2008.

[232] Parham A. Mirzaei and Fariborz Haghighat. Approaches to study Urban Heat Island âĂŞ Abilities and limitations. *Building and Environment*, 45(10):2192–2201, October 2010.

[233] A. Monaco, A. I. Chumakov, G. Monaco, W. A. Crichton, A. Meyer, L. Comez, D. Fioretto, J. Korecki, and R. RÃijffer. Effect of Densification on the Density of Vibrational States of Glasses. *Physical Review Letters*, 97(13):135501, September 2006.

[234] Paramita Mondal. *Nanomechanical properties of cementitious materials.* PhD Thesis, Northwestern University, Chicago, IL, USA., December 2008.

[235] Hendrik J. Monkhorst and James D. Pack. Special points for Brillouin-zone integrations. *Physical Review B*, 13:5188–5192, June 1976.

[236] Paulo J. M. Monteiro and C. T. Chang. The elastic moduli of calcium hydroxide. *Cement and Concrete Research*, 25(8):1605–1609, December 1995.

[237] Vladimir Alekseevich Morozov. *Methods for solving incorrectly posed problems.* Springer verlag, New York, 1984.

[238] Wilfried J. Mortier, Swapan K. Ghosh, and S. Shankar. Electronegativity-equalization method for the calculation of atomic charges in molecules. *Journal of the American Chemical Society*, 108(15):4315–4320, July 1986.

[239] P. Mounanga, A. Khelidj, and G. Bastian. Experimental study and modelling approaches for the thermal conductivity evolution of hydrating cement paste. *Advances in Cement Research*, 16(3):95–103, January 2004.

[240] Srinivas Mukkamala, Andrew H. Sung, Ajith Abraham, and Vitorino Ramos. Intrusion Detection Systems Using Adaptive Regression Spines. In Isabel Seruca, JosÃĺ Cordeiro, Slimane Hammoudi, and Joaquim Filipe, editors, *Enterprise Information Systems VI*, pages 211–218. Springer Netherlands, January 2006.

[241] A.C.A. Muller, K.L. Scrivener, A.M. Gajewicz, and P.J. McDonald. use of bench-top NMR to measure the density, composition and desorption isotherm of CâĂŞSâĂŞH in cement paste. *Microporous and Mesoporous Materials*, 178:99–103, September 2013.

[242] Arnaud C. A. Muller, Karen L. Scrivener, Agata M. Gajewicz, and Peter J. McDonald. densification of CâĂŞSâĂŞH measured by 1h NMR relaxometry. *The Journal of Physical Chemistry C*, 117(1):403–412, January 2013.

[243] T. Mura. *Micromechanics of Defects in Solids.* Springer, Dordrecht, Netherlands ; Boston : Hingham, MA, USA, 2nd edition edition, November 1987.

[244] Raymond H Myers, Christine M Anderson-Cook, and Douglas C Montgomery. *Response surface methodology: process and product optimization using designed experiments.* Wiley, Hoboken, NJ, 2009.

[245] David Nicholson and Neville George Parsonage. *Computer simulation and the statistical mechanics of adsorption.* Academic Press, 1982.

[246] Harald Niederreiter, editor. *Random number generation and quasi-Monte Carlo methods.* Society for Industrial and Applied Mathematics, Philadelphia, Pa., 1992.

[247] Fumito Nishi, Yoshio Takeuchi, and Iwao Maki. Tricalcium Silicate Ca3o[SIO4]: The monoclinic superstructure. *Zeitschrift fÃijr Kristallographie - Crystalline Materials*, 172(1-4):297–314, 1985.

[248] Arash Noshadravan, Roger Ghanem, Johann Guilleminot, Ikshwaku Atodaria, and Pedro Peralta. VALIDATION OF A PROBABILISTIC MODEL FOR

MESOSCALE ELASTICITY TENSOR OF RANDOM POLYCRYSTALS. *International Journal for Uncertainty Quantification*, 3(1):73–100, 2013.

[249] W.c. Oliver and G.m. Pharr. An improved technique for determining hardness and elastic modulus using load and displacement sensing indentation experiments. *Journal of Materials Research*, 7(06):1564–1583, 1992.

[250] W.c. Oliver and G.m. Pharr. Measurement of hardness and elastic modulus by instrumented indentation: Advances in understanding and refinements to methodology. *Journal of Materials Research*, 19(01):3–20, 2004.

[251] J. Alberto Ortega, Franz-Josef Ulm, and Younane Abousleiman. The nanogranular acoustic signature of shale. *Geophysics*, 74(3):D65–D84, May 2009.

[252] Kunze Ouyang and Fariborz Haghighat. A procedure for calculating thermal response factors of multi-layer wallsâĂŤState space method. *Building and Environment*, 26(2):173–177, 1991.

[253] A.B. Owen. Monte Carlo extension of quasi-Monte Carlo. In *Simulation Conference Proceedings, 1998. Winter*, volume 1, pages 571–577 vol.1, 1998.

[254] Art B. Owen. Controlling Correlations in Latin Hypercube Samples. *Journal of the American Statistical Association*, 89(428):1517–1522, December 1994.

[255] J. A. Palyvos. A survey of wind convection coefficient correlations for building envelope energy systemsâĂŹ modeling. *Applied Thermal Engineering*, 28(8âĂŞ9):801–808, June 2008.

[256] Panos Papalambros, Pierre Goovaerts, and Michael J. Sasena. Exploration of Metamodeling Sampling Criteria for Constrained Global Optimization. *Engineering Optimization*, 34(3):263–278, 2002.

[257] M. Parrinello and A. Rahman. polymorphic transitions in single crystals: a new molecular dynamics method. *Journal of Applied Physics*, 52(12):7182–7190, December 1981.

[258] D. A. Parshin, H. R. Schober, and V. L. Gurevich. Vibrational instability, two-level systems, and the boson peak in glasses. *Physical Review B*, 76(6):064206, August 2007.

[259] R. J.-M. Pellenq, N. Lequeux, and H. van Damme. engineering the bonding scheme in C-S-H: The iono-covalent framework. *Cement and Concrete Research*, 38(2):159–174, February 2008. WOS:000252623800006.

[260] Roland J.-M. Pellenq, Akihiro Kushima, Rouzbeh Shahsavari, Krystyn J. Van Vliet, Markus J. Buehler, Sidney Yip, and Franz-Josef Ulm. a realistic molecular model of cement hydrates. *Proceedings of the National Academy of Sciences*, 106(38):16102–16107, September 2009.

[261] Richard Perez, Pierre Ineichen, Robert Seals, Joseph Michalsky, and Ronald Stewart. Modeling daylight availability and irradiance components from direct and global irradiance. *Solar Energy*, 44(5):271–289, 1990.

[262] J.C. Phillips. topology of covalent non-crystalline solids I: short-range order in chalcogenide alloys. *Journal of Non-Crystalline Solids*, 34(2):153–181, October 1979.

[263] J.C. Phillips. topology of covalent non-crystalline solids II: medium-range order in chalcogenide alloys and AîŮÿSi(Ge). *Journal of Non-Crystalline Solids*, 43(1):37–77, January 1981.

[264] J.C. Phillips and M.F. Thorpe. constraint theory, vector percolation and glass formation. *Solid State Communications*, 53(8):699–702, February 1985.

[265] Michael C. Pitman and Adri C. T. van Duin. Dynamics of Confined Reactive Water in Smectite ClayâĂŞZeolite Composites. *Journal of the American Chemical Society*, 134(6):3042–3053, February 2012.

[266] C. Plassard, E. Lesniewska, I. Pochard, and A. Nonat. investigation of the surface structure and elastic properties of calcium silicate hy-

drates at the nanoscale. *Ultramicroscopy*, 100(3-4):331–338, August 2004. WOS:000222716900026.

[267] S. Plimpton. FAST PARALLEL ALGORITHMS FOR SHORT-RANGE MOLECULAR DYNAMICS. *Journal of computational physics*, 117(1):1–19, 1995.

[268] Robert O. Pohl. Influence of F Centers on the Lattice Thermal Conductivity in LiF. *Physical Review*, 118(6):1499–1508, June 1960.

[269] Joseph M. Powers. On the Necessity of Positive Semi-Definite Conductivity and Onsager Reciprocity in Modeling Heat Conduction in Anisotropic Media. *Journal of Heat Transfer*, 126(5):670–675, November 2004.

[270] F. X. Prielmeier, E. W. Lang, R. J. Speedy, and H.-D. Ludemann. The pressure dependence of self diffusion in supercooled light and heavy water. *Berichte der Bunsengesellschaft fur Physikalische Chemie*, 92(10):1111–1117, 1988.

[271] J. Puibasset and R. J. M. Pellenq. Water adsorption in disordered mesoporous silica (Vycor) at 300 K and 650 K: A Grand Canonical Monte Carlo simulation study of hysteresis. *Journal of Chemical Physics*, 122(9), March 2005. WOS:000227483300059.

[272] JoeÌĹl Puibasset and Roland J.-M. Pellenq. Grand Canonical Monte Carlo Simulation Study of Water Adsorption in Silicalite at 300 K. *The Journal of Physical Chemistry B*, 112(20):6390–6397, May 2008.

[273] M. J. Abdolhosseini Qomi, A. Aghaei, and A. R. Khoei. Multi-scale modeling of surface effect via the boundary CauchyâĂŞBorn method. *International Journal for Numerical Methods in Engineering*, 85(7):827–846, February 2011.

[274] Mohammad Javad Abdolhosseini Qomi, Mathieu Bauchy, Franz-Josef Ulm, and Roland J.-M. Pellenq. anomalous composition-dependent dynamics of nanoconfined water in the interlayer of disordered calcium-silicates. *The Journal of Chemical Physics*, 140(5):054515, February 2014.

[275] X. Qu, G. Venter, and R. T. Haftka. New formulation of minimum-bias central composite experimental design and Gauss quadrature. *Structural and Multidisciplinary Optimization*, 28(4):231–242, October 2004.

[276] Nestor V. Queipo, Raphael T. Haftka, Wei Shyy, Tushar Goel, Rajkumar Vaidyanathan, and P. Kevin Tucker. Surrogate-based analysis and optimization. *Progress in Aerospace Sciences*, 41(1):1–28, January 2005.

[277] Johan Qvist, Helmut Schober, and Bertil Halle. Structural dynamics of supercooled water from quasielastic neutron scattering and molecular simulations. *The Journal of Chemical Physics*, 134(14):144508–144508–20, April 2011.

[278] C. R. Rao. Hypercubes of strength "d" leading to confounded designs in factorial experiments. *Bulletin of the Calcutta Mathematical Society*, 38:67–78, 1946.

[279] C. Radhakrishna Rao. Factorial Experiments Derivable from Combinatorial Arrangements of Arrays. *Supplement to the Journal of the Royal Statistical Society*, 9(1):128, 1947.

[280] E. H. Ratcliffe. Thermal conductivities of fused and crystalline quartz. *British Journal of Applied Physics*, 10(1):22, January 1959.

[281] Christoph F. Reinhart and Sebastian Herkel. The simulation of annual daylight illuminance distributions âĂŤ a state-of-the-art comparison of six RADIANCE-based methods. *Energy and Buildings*, 32(2):167–187, July 2000.

[282] Christoph F. Reinhart, John Mardaljevic, and Zack Rogers. *Dynamic Daylight Performance Metrics for Sustainable Building Design.* 2006.

[283] I. G Richardson. the nature of the hydration products in hardened cement pastes. *Cement and Concrete Composites*, 22(2):97–113, April 2000.

[284] I. G. Richardson. the calcium silicate hydrates. *Cement and Concrete Research*, 38(2):137–158, February 2008.

[285] I. G. Richardson and G. W. Groves. the incorporation of minor and trace elements into calcium silicate hydrate (C-S-H) gel in hardened cement pastes. *Cement and Concrete Research*, 23(1):131–138, January 1993.

[286] I.G. Richardson. tobermorite/jennite- and tobermorite/calcium hydroxide-based models for the structure of C-S-H: applicability to hardened pastes of tricalcium silicate, Îš-dicalcium silicate, Portland cement, and blends of Portland cement with blast-furnace slag, metakaolin, or silica fume. *Cement and Concrete Research*, 34(9):1733–1777, September 2004.

[287] Ig Richardson and Gw Groves. microstructure and microanalysis of hardened cement pastes involving ground granulated blast-furnace slag. *Journal of Materials Science*, 27(22):6204–6212, November 1992. WOS:A1992JZ83700034.

[288] Ig Richardson and Gw Groves. microstructure and microanalysis of hardened ordinary portland-cement pastes. *Journal of Materials Science*, 28(1):265–277, January 1993. WOS:A1993KM82100037.

[289] Andrew Richmond. Financially Efficient Ore Selections Incorporating Grade Uncertainty. *Mathematical Geology*, 35(2):195–215, February 2003.

[290] Alexandra Roder, Walter Kob, and Kurt Binder. Structure and dynamics of amorphous silica surfaces. *Journal of Chemical Physics*, 114:7602–7614, May 2001.

[291] Stuart J Russell, Peter Norvig, and Ernest Davis. *Artificial intelligence: a modern approach*. Prentice Hall, Upper Saddle River, NJ, 2010.

[292] Michael F. Russo Jr. and Adri C.T. van Duin. Atomistic-scale simulations of chemical reactions: Bridging from quantum chemistry to engineering. *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms*, 269(14):1549–1554, July 2011.

[293] Jerome Sacks, William J. Welch, Toby J. Mitchell, and Henry P. Wynn. Design and Analysis of Computer Experiments. *Statistical Science*, 4(4):409–423,

November 1989. Mathematical Reviews number (MathSciNet): MR1041765; Zentralblatt MATH identifier: 0955.62619.

[294] David J. Sailor and J. Ricardo Munoz. Sensitivity of electricity and natural gas consumption to climate in the U.S.A.âĂŤMethodology and results for eight states. *Energy*, 22(10):987–998, October 1997.

[295] S. Sakata, F. Ashida, and M. Zako. Structural optimization using Kriging approximation. *Computer Methods in Applied Mechanics and Engineering*, 192(7âĂŞ8):923–939, February 2003.

[296] A Saltelli. *Global sensitivity analysis: the primer.* John Wiley, Chichester, England; Hoboken, NJ, 2008.

[297] A. Saltelli, K. Chan, and E. M. Scott. *Sensitivity Analysis: Gauging the Worth of Scientific Models.* Wiley, October 2000.

[298] A. Saltelli, Stefano Tarantola, Francesca Campolongo, and Marco Ratto. *Sensitivity Analysis in Practice: A Guide to Assessing Scientific Models.* John Wiley & Sons, July 2004.

[299] Antonio Scala, Francis W. Starr, Emilia La Nave, Francesco Sciortino, and H. Eugene Stanley. Configurational entropy and diffusivity of supercooled water. *Nature*, 406(6792):166–169, July 2000.

[300] Patrick K. Schelling, Simon R. Phillpot, and Pawel Keblinski. Comparison of atomic-level simulation methods for computing thermal conductivity. *Physical Review B*, 65(14):144306, April 2002.

[301] Gideon Schwarz. Estimating the Dimension of a Model. *The Annals of Statistics*, 6(2):461–464, March 1978. Mathematical Reviews number (MathSciNet): MR468014; Zentralblatt MATH identifier: 0379.62005.

[302] Felix Sedlmeier, Dominik Horinek, and Roland R. Netz. Spatial Correlations of Density and Structural Fluctuations in Liquid Water: A Comparative Sim-

ulation Study. *Journal of the American Chemical Society*, 133(5):1391–1398, February 2011.

[303] J E Seem. *modeling of heat transfer in buildings.* PhD Thesis, University of Wisconsin, Madison, WI, 1987.

[304] Francesco Sette, Michael H. Krisch, Claudio Masciovecchio, Giancarlo Ruocco, and Giulio Monaco. Dynamics of Glasses and Glass-Forming Liquids Studied by Inelastic X-ray Scattering. *Science*, 280(5369):1550–1555, June 1998.

[305] Rouzbeh Shahsavari, Markus J. Buehler, Roland J.-M. Pellenq, and Franz-Josef Ulm. first-principles study of elastic constants and interlayer interactions of complex hydrated oxides: case study of tobermorite and jennite. *Journal of the American Ceramic Society*, 92(10):2323–2330, October 2009. WOS:000270188300025.

[306] Rouzbeh Shahsavari, Roland J-M Pellenq, and Franz-Josef Ulm. empirical force fields for complex hydrated calcio-silicate layered materials. *Physical chemistry chemical physics: PCCP*, 13(3):1002–1011, January 2011.

[307] Hiroshi Shintani and Hajime Tanaka. Universal link between the boson peak and transverse phonons in glass. *Nature Materials*, 7(11):870–877, November 2008.

[308] Katsuhiro Shirono and Hirofumi Daiguji. Dipole moments of water molecules confined in NaâĂŞLSX zeolites âĂŞ Molecular dynamics simulations including polarization of water. *Chemical Physics Letters*, 417(1âĂŞ3):251–255, January 2006.

[309] Timothy W. Simpson, John J. Korte, Timothy M. Mauery, and Farrokh Mistree. Kriging Models for Global Approximation in Simulation-Based Multidisciplinary Design Optimization. *AIAA Journal*, 39:2233–2241, December 2001.

[310] L. B. Skinner, S. R. Chae, C. J. Benmore, H. R. Wenk, and P. J. M. Monteiro. nanostructure of calcium silicate hydrates in cements. *Physical Review Letters*, 104(19):195502, May 2010.

[311] K.S Smirnov and D Bougeard. Including the polarization in simulations of hydrated aluminosilicates. Model and application to water in silicalite. *Chemical Physics*, 292(1):53–70, July 2003.

[312] D. K. Smith, A. Majumdar, and F. Ordway. The crystal structure of Î§-dicalcium silicate. *Acta Crystallographica*, 18(4):787–795, April 1965.

[313] Roel Snieder. The role of nonlinearity in inverse problems. *Inverse Problems*, 14(3):387, June 1998.

[314] Ilya M. Sobol. On the distribution of points in a cube and the approximate evaluation of integrals. *USSR Computational Mathematics and Mathematical Physics*, 7(4):86–112, 1967.

[315] Ilya M. Sobol. *A Primer for the Monte Carlo Method.* CRC Press, New York, NY, 1994.

[316] A. P. Sokolov, U. Buchenau, W. Steffen, B. Frick, and A. Wischnewski. Comparison of Raman- and neutron-scattering data for glass-forming systems. *Physical Review B*, 52(14):R9815–R9818, October 1995.

[317] A. P. Sokolov, A. Kisliuk, M. Soltwisch, and D. Quitmann. Medium-range order in glasses: Comparison of Raman and diffraction measurements. *Physical Review Letters*, 69(10):1540–1543, September 1992.

[318] Sezen Soyer-Uzun, Sejung Rosie Chae, Chris J. Benmore, Hans-Rudolf Wenk, and Paulo J. M. Monteiro. compositional evolution of calcium silicate hydrate (CâĂŞSâĂŞH) structures by total X-ray scattering. *Journal of the American Ceramic Society*, 95(2):793–798, 2012.

[319] Francis W. Starr, Srikanth Sastry, Jack F. Douglas, and Sharon C. Glotzer. What Do We Learn from the Local Geometry of Glass-Forming Liquids? *Physical Review Letters*, 89(12):125501, August 2002.

[320] Michael Stein. Large Sample Properties of Simulations Using Latin Hypercube Sampling. *Technometrics*, 29(2):143–151, 1987.

[321] S. Sugai and A. Onodera. Medium-Range Order in Permanently Densified SiO2 and GeO2 Glass. *Physical Review Letters*, 77(20):4210–4213, November 1996.

[322] S. Susman, K. J. Volin, R. C. Liebermann, G. D. Gwanmesia, and Yanbin Wang. Structural changes in irreversibly densified fused silica : implications for the chemical resistance of high level nuclear waste glasses. *Physics and chemistry of glasses*, 31(4):144–150, 1990.

[323] Lukas G. Swan and V. Ismet Ugursal. Modeling of end-use energy consumption in the residential sector: A review of modeling techniques. *Renewable and Sustainable Energy Reviews*, 13(8):1819–1835, October 2009.

[324] Hajime Tanaka. Physical Origin of the Boson Peak Deduced from a Two-Order-Parameter Model of Liquid. *Journal of the Physical Society of Japan*, 70(5):1178–1181, May 2001.

[325] Boxin Tang. Orthogonal Array-Based Latin Hypercubes. *Journal of the American Statistical Association*, 88(424):1392–1397, 1993.

[326] N. J. Tao, G. Li, X. Chen, W. M. Du, and H. Z. Cummins. Low-frequency Raman-scattering study of the liquid-glass transition in aqueous lithium chloride solutions. *Physical Review A*, 44(10):6665–6676, November 1991.

[327] S. N. Taraskin and S. R. Elliott. Nature of vibrational excitations in vitreous silica. *Physical Review B*, 56(14):8605–8622, October 1997.

[328] H. F. W Taylor. *cement chemistry.* T. Telford, London, July 1997.

[329] J. Teixeira, M.-C. Bellissent-Funel, S. H. Chen, and A. J. Dianoux. Experimental determination of the nature of diffusive motions of water molecules at low temperatures. *Physical Review A*, 31(3):1913–1917, March 1985.

[330] Olle Teleman, Bo Jonsson, and Sven Engstrom. A molecular dynamics simulation of a water model with intramolecular degrees of freedom. *Molecular Physics*, 60(1):193–203, 1987.

[331] J. Tersoff. Empirical Interatomic Potential for Carbon, with Applications to Amorphous Carbon. *Physical Review Letters*, 61(25):2879–2882, December 1988.

[332] Jeffrey J. Thomas, Jeffrey J. Chen, Hamlin M. Jennings, and Dan A. Neumann. Ca-OH bonding in the C-S-H gel phase of tricalcium silicate and white portland cement pastes measured by inelastic neutron scattering. *Chemistry of materials*, 15(20):3813–3817, 2003.

[333] Jeffrey J. Thomas, Stephen A. FitzGerald, Dan A. Neumann, and Richard A. Livingston. State of Water in Hydrating Tricalcium Silicate and Portland Cement Pastes as Measured by Quasi-Elastic Neutron Scattering. *Journal of the American Ceramic Society*, 84(8):1811–1816, 2001.

[334] Jeffrey J. Thomas, Hamlin M. Jennings, and Andrew J. Allen. Relationships between Composition and Density of Tobermorite, Jennite, and Nanoscale CaO-SiO2-H2o. *Journal of Physical Chemistry C*, 114(17):7594–7601, May 2010. WOS:000277053600011.

[335] John A. Thomas, Joseph E. Turney, Ryan M. Iutzi, Cristina H. Amon, and Alan J. H. McGaughey. Predicting phonon dispersion relations and lifetimes from the spectral energy density. *Physical Review B*, 81(8):081411, February 2010.

[336] A. N Tikhonov and V. IïÿăAïÿą Arsenin. *Solutions of ill-posed problems.* Winston ; Distributed solely by Halsted Press, Washington; New York, 1977.

[337] D. M. Titterington, Adrian F. M. Smith, and U. E. Makov. *Statistical Analysis of Finite Mixture Distributions.* John Wiley & Sons, 1 edition, January 1986.

[338] S. S. Todd. Low-temperature Heat Capacities and Entropies at 298.16ÂřK. of Crystalline Calcium Orthosilicate, Zinc Orthosilicate and Tricalcium Silicate. *Journal of the American Chemical Society*, 73(7):3277–3278, July 1951.

[339] E. Tombari, G. Salvetti, C. Ferrari, and G. P. Johari. Heat capacity of water in nanopores. *The Journal of Chemical Physics*, 123(21):214706, December 2005.

[340] Y. S. Touloukian. *Thermal Conductivity: Nonmetallic Solids.* Springer, New York, 1 edition edition, January 1971.

[341] Konstantin V. Tretiakov and Sandro Scandolo. Thermal conductivity of solid argon from molecular dynamics simulations. *The Journal of Chemical Physics*, 120(8):3765–3769, February 2004.

[342] Geoffrey K. F. Tso and Kelvin K. W. Yau. Predicting electricity energy consumption: A comparison of regression analysis, decision tree and neural networks. *Energy*, 32(9):1761–1768, September 2007.

[343] A. Uhlherr and S. R. Elliott. extended-range propagated order in amorphous solids. *Journal of Physics: Condensed Matter*, 6(8):L99, February 1994.

[344] Franz-Josef Ulm. concrete innovation potential: from atoms to green infrastructure. *Beton- und Stahlbetonbau*, 107(8):504–509, 2012.

[345] Franz-Josef Ulm and Hamlin M. Jennings. Does CâĂŞSâĂŞH particle shape matter? A discussion of the paper âĂŸModelling elasticity of a hydrating cement pasteâĂŹ, by Julien Sanahuja, Luc Dormieux and Gilles Chanvillard. CCR 37 (2007) 1427âĂŞ1439. *Cement and Concrete Research*, 38(8âĂŞ9):1126–1129, August 2008.

[346] Franz-Josef Ulm, Matthieu Vandamme, Chris Bobko, and Jose Alberto Ortega. Statistical indentation techniques for hydrated nanocomposites: Concrete,

bone, and shale. *Journal of the American Ceramic Society*, 90(9):2677–2692, September 2007. WOS:000249276800001.

[347] Bryan Urban. *The MIT Design Advisor: Simple and Rapid Energy Simulation of Early-Stage Building Designs.* Doctoral dissertation, Massachusetts Institute of Technology, Cambridge, 2007.

[348] Bryan Urban and Leon Glicksman. The MIT Design Advisor-a fast, simple tool for energy efficient building design. In *IBPSA-USA Conference Proceedings*, pages 270–276, 2006.

[349] Bryan Urban and Leon Glicksman. A simplified rapid energy model and interface for nontechnical users. In *10th ORNL Thermal Performance of the Exterior Envelopes of Whole Buildings International Conference*, 2007.

[350] Paul J. van Maaren and David van der Spoel. Molecular Dynamics Simulations of Water with Novel Shell-Model Potentials. *The Journal of Physical Chemistry B*, 105(13):2618–2626, April 2001.

[351] Arun K. Varshneya and Daniel J. Mauro. microhardness, indentation toughness, elasticity, plasticity, and brittleness of GeâĂŞSbâĂŞSe chalcogenide glasses. *Journal of Non-Crystalline Solids*, 353(13âĂŞ15):1291–1297, May 2007.

[352] Karine Velez, Sandrine Maximilien, Denis Damidot, Gilbert Fantozzi, and Fransois Sorrentino. Determination by nanoindentation of elastic modulus and hardness of pure constituents of Portland cement clinker. *Cement and Concrete Research*, 31(4):555–561, April 2001.

[353] George N Walton. Thermal Analysis Research Program, 1983.

[354] M. Wilson and P. A. Madden. âĂŸâĂŸPrepeaksâĂŹâĂŹ and âĂŸâĂŸfirst sharp diffraction peaksâĂŹâĂŹ in computer simulations of strong and fragile ionic liquids. *Physical Review Letters*, 72(19):3033–3036, May 1994.

[355] A. Wischnewski, U. Buchenau, A. J. Dianoux, W. A. Kamitakahara, and J. L. Zarestky. Sound-wave scattering in silica. *Physical Review B*, 57(5):2663–2666, February 1998.

[356] F. H. Wittmann. Estimation of the modulus of elasticity of calcium hydroxide. *Cement and Concrete Research*, 16(6):971–972, November 1986.

[357] D. Wolf, P. Keblinski, S. R. Phillpot, and J. Eggebrecht. Exact method for the simulation of Coulombic systems by spherically truncated, pairwise râĹŠ1 summation. *The Journal of Chemical Physics*, 110(17):8254–8282, May 1999.

[358] Ernst Worrell, Lynn Price, Nathan Martin, Chris Hendriks, and Leticia Ozawa Meida. carbon dioxide emissions from the global cement industry. *Annual Review of Energy and the Environment*, 26(1):303–329, 2001.

[359] Adrian C. Wright. Neutron and X-ray amorphography. *Journal of Non-Crystalline Solids*, 106(1âĂŞ3):1–16, December 1988.

[360] Adrian C Wright. The comparison of molecular dynamics simulations with diffraction experiments. *Journal of Non-Crystalline Solids*, 159(3):264–268, 1993.

[361] Adrian C. Wright, Robert A. Hulme, David I. Grimley, Roger N. Sinclair, Steve W. Martin, David L. Price, and Frank L. Galeener. The structure of some simple amorphous network solids revisited. *Journal of Non-Crystalline Solids*, 129(1âĂŞ3):213–232, March 1991.

[362] Qing-Song Xu, D. L. Massart, Yi-Zeng Liang, and Kai-Tai Fang. Two-step multivariate adaptive regression splines for modeling a quantitative relationship between gas chromatography retention indices and molecular descriptors. *Journal of Chromatography A*, 998(1âĂŞ2):155–167, May 2003.

[363] Yunsheng Xu and D. D. L. Chung. Effect of sand addition on the specific heat and thermal conductivity of cement. *Cement and Concrete Research*, 30(1):59–61, January 2000.

[364] Chun-Chieh Yang, Shiv O. Prasher, RenÃľ Lacroix, and Seung Hyun Kim. A Multivariate Adaptive Regression Splines Model for Simulation of Pesticide Transport in Soils. *Biosystems Engineering*, 86(1):9–15, September 2003.

[365] M Yazdanian and J H Klems. Measurement of the exterior convective film coefficient for windows in low-rise buildings. *ASHRAE Transaction*, 100(1):1–15, 1994.

[366] Kenny Q Ye, William Li, and Agus Sudjianto. Algorithmic construction of optimal symmetric Latin hypercube designs. *Journal of Statistical Planning and Inference*, 90(1):145–159, September 2000.

[367] Seyoon Yoon, Donald E. Macphee, and Mohammed S. Imbabi. Estimation of the thermal properties of hardened cement paste on the basis of guarded heat flow meter measurements. *Thermochimica Acta*, 588:1–10, July 2014.

[368] Mostafa Youssef, Roland J.-M. Pellenq, and Bilge Yildiz. glassy nature of water in an ultraconfining disordered material: the case of calciumâĹŠsilicateâĹŠhydrate. *Journal of the American Chemical Society*, 133(8):2499–2510, March 2011.

[369] 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, March 1999.

[370] Ronen Zangi and Stuart A. Rice. Freezing transition and correlated motion in a quasi-two-dimensional colloid suspension. *Physical Review E*, 68(6):061508, December 2003.

[371] H. Zareipour, K. Bhattacharya, and C.A. Canizares. Forecasting the hourly Ontario energy price by multivariate adaptive regression splines. In *IEEE Power Engineering Society General Meeting, 2006*, pages 7 pp.–, 2006.

[372] Joseph M. Zaug, Alan K. Soper, and Simon M. Clark. Pressure-dependent structures of amorphous red phosphorus and the origin of the first sharp diffraction peaks. *Nature Materials*, 7(11):890–899, November 2008.

[373] R. C. Zeller and R. O. Pohl. Thermal Conductivity and Specific Heat of Noncrystalline Solids. *Physical Review B*, 4(6):2029–2041, September 1971.

[374] J. M. Ziman. *Electrons and Phonons: The Theory of Transport Phenomena in Solids.* Oxford University Press, New York, NY, first edition, April 2001.

[375] N. Zotov, Y. Yanev, M. Epelbaum, and L. Konstantinov. Effect of water on the structure of rhyolite glasses - X-ray diffraction and Raman spectroscopy studies. *Journal of Non-Crystalline Solids*, 142:234–246, 1992.