

# UC Santa Barbara

## UC Santa Barbara Electronic Theses and Dissertations

### Title

Digital Design of Crystals: Predicting Driving Forces For Crystallization Using Atomistic Simulations

### Permalink

<https://escholarship.org/uc/item/9pg5462v>

### Author

Khanna, Vikram

### Publication Date

2021

Peer reviewed|Thesis/dissertation

University of California  
Santa Barbara

# Digital Design of Crystals: Predicting Driving Forces For Crystallization Using Atomistic Simulations

A dissertation submitted in partial satisfaction  
of the requirements for the degree

Doctor of Philosophy  
in  
Chemical Engineering

by

Vikram Khanna

Committee in charge:

Professor Michael F. Doherty, Co-chair  
Professor Baron Peters, Co-chair  
Professor M. Scott Shell  
Professor Ram Seshadri

March 2022

The Dissertation of Vikram Khanna is approved.

---

Professor M. Scott Shell

---

Professor Ram Seshadri

---

Professor Baron Peters, Committee Co-chair

---

Professor Michael F. Doherty, Committee Co-chair

December 2021

Digital Design of Crystals: Predicting Driving Forces For Crystallization Using  
Atomistic Simulations

Copyright © 2022

by

Vikram Khanna

To my late grandmother, Mrs. Shakuntala Sharma  
whose penchant for knowledge & learning continues to inspire me...

## Acknowledgements

*Gratitude* is the dominant feeling I am experiencing as my PhD journey is coming to an end. It does take a village! — at least it did for me — to even embark on this journey, let alone successfully completing it.

To my advisors: Prof. Doherty and Prof. Baron, it is my privilege and honor to be co-advised by you both and be your student. Both of you invested your head, heart, and hands to nurture me, and I am forever grateful for that! Not only have you helped me become a better problem-solver—but also a better human being! Prof. Doherty, your philosophy of optimizing positive outcomes for all stakeholders involved with you rather than just yourself is praiseworthy — and something I will emulate! I would like to thank you for the numerous opportunities you bestowed upon me. The experiences I gained as a result, have been truly one of a kind—and will help me throughout my career! Prof. Baron, your courage to challenge the status quo for the love and benefit of science is inspiring! The enthusiasm you bring to every discussion of ours is infectious! I have thoroughly enjoyed every second working with you both! You have become my role models and it is truly my good fortune that I got to spend 5 fulfilling years with such wonderful humans! Thank you from the core of my heart!

To the rest of my committee — Prof. Shell and Prof. Ram — I am super thankful for your time and nuanced insights. They have helped me navigate my project super smoothly. Each meeting with you all was as stimulating as it was energizing. You are mentors in the truest sense of the word. To my Professors at UCSB, I am thankful to you all for sharing your experience and distilling your knowledge in an extremely palatable form.

To the Professors of my undergraduate university — Institute of Chemical Technology, Mumbai — I am super grateful for the strong foundation you provided me to build

upon — especially Professors Ashwin Patwardhan, Aniruddha Pandit, Pushpito Ghosh, Vishwanath Dalvi, S.S. Bhagwat and J.B. Joshi. You made me fall in love with the subjects you taught and inspired me to pursue a PhD.

Many thanks to Narotam Sekhsaria Foundation and Sinkhai Foundation for the graduate studies scholarships — it is an honor! Beyond grateful to Eli Lilly and Company for the financial support that enabled the work described in this dissertation. To my collaborators at Lilly — Dr. Christopher Burcham & Dr. Susan Reutzal-Edens — thank you for being the fountain of knowledge that you both are and for always asking super intriguing questions! You made me look forward to the 8 a.m. meetings! To my collaborators at University College London — Prof. Sally Price, Prof. Matteo Salvalaglio, Dr. Nicholas Francia, Dr. Louise Price — I couldn't have asked for a better team to work with! Your expertise, collaborative spirit, and constructive criticism has brought out the best outcomes for our project! Also, thank you for hosting me during my visit to London. It was such a pleasure meeting and working with you all in person! To my collaborators, Prof. Daan Frenkel and Prof. Jamshed Anwar, it is one of my highlights to have collaborated and co-authored a publication with you!

To the past and present members of the Doherty, Peters, and Shell labs, you have been a sheer pleasure to work with! I would like to thank Dr. Mark Joswiak, Dr. Carl Tilbury, Dr. Christian Leitold, Dr. Jacob Monroe, and Dr. Kartik Kamat for helping me build the required knowledge base and skillset for my research! Jacob, it was my good fortune to have an excellent scientist such as you, as a co-author on my first publication! Carl, I cannot thank you enough for all the guidance and help you have provided me during your time at UCSB and even after it — helping me land the next job I truly desired!

To my parents, Jaya and Naval, I won the lottery being born as your child! You showered me and Ritam with the best values, resources, and above all unwavering-support

and love. You sacrificed your needs for our wants. I cannot thank you enough and hope to make you proud by doing good and being good. To my brother, Ritam, you are my pillar of support. Without you I couldn't have reached so far. To my rock, Nishita, I am short of words for your love, encouragement, and patience. To my other beloved family — Prasun & Sanjay (Nishita's parents), Kaushalya & Rajendra (Nishita's grandparents), and Viresh (Nishita's brother) — I am lucky to have you by my side — cheering and rooting for me! To my extended family, especially my *masis* (mother's sisters), and Kamala *Maushi*, thank you for the encouragement and best wishes, throughout!

To my friends here at UCSB, thank you so much for making this journey an enjoyable one! It was a pleasure to have such amazing company and I am grateful to have made bonds for life! To NJ, my group of friends from undergrad, you guys mean the world to me. Thank you for sharing all the highs and lows of my journey! To Kartik, Aparna, Anushree, Ashish, Zulkif, Sanj, and Sudarshan — thanks for being just one call away! To Dr. Chandrakanth, Vimisha and Divya, thank you for the wonderful philosophical conversations, they always help me to put things in perspective!

My penultimate thanks — to the staff at ICT and UCSB. The logistics of my academic life have been super smooth because of the fantastic work you all put in!

And finally, thanks to the most important person of my life, my late grandmother (my *nani*—the Hindi word for maternal grandmother), Mrs. Shakuntala Sharma, whose love for knowledge and perpetual desire to learn, continues to inspire me. She learnt English at an old age, and never got tired reaching out to the dictionary whenever she came across a new word! *Nani*, having lost you during my PhD was heart wrenching, but you continue to live with me through your teachings, and I remember you each time I learn something new! Therefore, I will keep seeking new knowledge!

# Curriculum Vitæ

## Vikram Khanna

### Education

- 2021 Ph.D. in Chemical Engineering (Expected), University of California, Santa Barbara.
- 2021 Graduate Program in Management Practice Certificate, University of California, Santa Barbara.
- 2016 B.Chem.Engg, Institute of Chemical Technology, Mumbai.

### Publications

Vikram Khanna, Michael F. Doherty and Baron Peters, “Predicting Driving Forces For Crystallization Using the Absolute Chemical Potential Route,” (manuscript in preparation).

Christopher L. Burcham, Michael F. Doherty, Nicholas Francia, Vikram Khanna, Baron Peters, Sally L. Price, Susan M. Reutzel-Edens, Matteo Salvalaglio, “Digital Design of Crystals: From Molecules to Crystals,” (manuscript in preparation).

Yongsheng Zhao, Vikram Khanna, Robert Gee and Michael F. Doherty, “Morphology Prediction of Crystals using COSMO-based Free Energy Methods,” (manuscript in preparation).

Vikram Khanna, Dennis Robinson Brown and Michael F. Doherty, “Revisiting Population Balances: The Impact of Solid Volume Fraction,” (manuscript in preparation).

Vikram Khanna, Michael F. Doherty, and Baron Peters, “The Ever-kinked Crystal Framework for Studying In Silico Crystal Growth,” (manuscript in preparation).

Jeffery Frumkin, Vikram Khanna, and Michael F. Doherty, “Innovation in Chemical Reactor Engineering Practice and Science,” *Computers & Chemical Engineering*, (accepted).

Vikram Khanna, Jamshed Anwar, Daan Frenkel, Michael F. Doherty and Baron Peters, “Free energies of crystals computed using Einstein crystal with fixed center of mass and differing spring constants,” *Journal of Chemical Physics*, **2021**, 154(16), 164509:1-8. DOI: <https://doi.org/10.1063/5.0044833>.

Vikram Khanna, Michael F. Doherty and Baron Peters, “Absolute chemical potentials for complex molecules in fluid phases: A centroid reference for predicting phase equilibria,” *Journal of Chemical Physics*, **2020**, 153(21), 214504:1-9.

DOI: <https://doi.org/10.1063/5.0025844>.

Vikram Khanna, Jacob I. Monroe, Michael F. Doherty and Baron Peters, "Performing solvation free energy calculations in LAMMPS using the decoupling approach," *Journal of Computer Aided Molecular Design*, **2020**, 34, 641-646.

DOI: <https://doi.org/10.1007/s10822-020-00303-3>.

### **Conference Presentations**

Vikram Khanna, Michael F. Doherty and Baron Peters, "Solubility Prediction of Organic Molecules using Atomistic Simulations," American Institute of Chemical Engineers, Annual Virtual Meeting, San Francisco, CA, USA, Fall 2020.

Vikram Khanna, Michael F. Doherty and Baron Peters, "Towards Digital Design of Crystals: Predicting absolute chemical potentials of solid, solution and gas phases," American Institute of Chemical Engineers, Annual Meeting, Orlando, FL, USA, Fall 2019.

Vikram Khanna and Michael F. Doherty, "Direct Air Capture of CO<sub>2</sub> (DAC) is NOT as Crazy as it Sounds," Carbon Management Workshop at Stanford University, CA, USA, Fall 2019.

Vikram Khanna, Michael F. Doherty and Baron Peters, "Towards Digital Design of Crystals: Predicting solubility via computing absolute chemical potentials of solid, solution and gas phases," The British Association for Crystal Growth, Annual Meeting, London, UK, Summer 2019.

### **Awards and Honors**

Awarded one of top 3 talks (audience rated), Graduate Simulation Seminar Series **2019**

Narotam Sekhsaria Scholarship for Graduate Studies **2016**

Best Student Award **2015**

Sir Ratan Tata Scholarship **2014, 2015, 2016**

## Abstract

Digital Design of Crystals: Predicting Driving Forces For Crystallization Using  
Atomistic Simulations

by

Vikram Khanna

One of the most awe-inspiring class of materials are crystals. These highly ordered groups of atoms/molecules propel our lives through myriad products we humans rely upon. Right from the salt and sugar we consume to elevate the taste of our foods, the silicon chips that form the *brains* of every computing device we use, to the life-saving drugs that have prevented millions if not billions of deaths, all belong to the humble class of materials — the crystal. Therefore, the engineering of this material is crucial to — improve the manufacturing of products that touch our lives daily — ultimately improving the quality of every human life!

The two physical attributes of a crystalline material that have a major impact on its processability and performance are: its shape and size. For e.g., in a crystalline catalyst one wants to engineer the shape of the crystal to maximize the area of its reactive surfaces. In pharmaceutical applications, the crystal size distribution may determine the rate of plasma uptake of a drug when the process is dissolution rate limited. Therefore, engineering the shape and size of the crystals is of immense consequence.

In silico tools that can predict the shape and size of a crystal based on inputs such as crystal structure, temperature, supersaturation, etc. are vital to efficiently navigate the process design space. Such tools help achieve the efficiency gains by being a guiding light to experimentalists, thus enabling cheaper and more effective screening. This dissertation lays out the digital design framework to make these predictions starting from

a molecule. It focuses on the development of a computational toolkit to predict driving forces for crystallization—a key prediction to enable size predictions—of complex molecules harnessing atomistic simulations.

# Contents

<b>Curriculum Vitae</b>	<b>viii</b>
<b>Abstract</b>	<b>x</b>
<b>1 Digital Design of Crystals</b>	<b>1</b>
1.1 Polymorph Prediction . . . . .	7
1.2 Morphology Design . . . . .	11
1.3 Solubility . . . . .	13
1.4 Nucleation and Growth Rates . . . . .	13
1.5 Attainable Particle Size Design . . . . .	16
1.6 Polymorph Selection . . . . .	18
<b>Bibliography</b>	<b>21</b>
<b>2 Crystal Growth and the Solubility Prediction Framework</b>	<b>25</b>
2.1 The crystal growth framework . . . . .	25
2.2 The solubility prediction framework . . . . .	35
<b>Bibliography</b>	<b>41</b>
<b>3 Chemical Potential in the Solid Phase</b>	<b>45</b>
3.1 Introduction . . . . .	45
3.2 The Solid Free Energy Theory . . . . .	48
3.3 Simulation Details . . . . .	53
3.4 Results and Discussion . . . . .	55
3.5 Extension to Molecular Crystals . . . . .	61
3.6 Conclusions . . . . .	62
<b>Bibliography</b>	<b>63</b>
<b>4 Chemical Potential in the Gas Phase</b>	<b>66</b>
4.1 Introduction . . . . .	66

4.2	Computing the Free Energy of the Reference Centroid System . . . . .	68
4.3	Testing the Reference Centroid System . . . . .	70
4.4	Transforming the Centroid into the Molecule . . . . .	72
<b>Bibliography</b>		<b>74</b>
<b>5</b>	<b>Chemical Potential in the Solution Phase</b>	<b>75</b>
5.1	Introduction . . . . .	75
5.2	Thermodynamic Cycle . . . . .	76
5.3	Setting up the Decoupling Route . . . . .	79
5.4	Testing the Computational Strategy . . . . .	82
5.5	Results and Discussion . . . . .	84
5.6	Solvent-Modified Bond Energies—A new application for solvation free energy calculations . . . . .	87
5.7	Conclusions . . . . .	89
<b>Bibliography</b>		<b>90</b>
<b>6</b>	<b>Predicting Solid-Fluid Equilibrium</b>	<b>93</b>
6.1	Introduction . . . . .	93
6.2	Predicting Solid-Vapor Equilibrium . . . . .	94
6.3	Simulation Details . . . . .	96
6.4	Results and Discussion . . . . .	101
6.5	Predicting Solid-Solution Equilibrium . . . . .	110
6.6	Conclusion . . . . .	117
<b>Bibliography</b>		<b>119</b>
<b>7</b>	<b>Revisiting Population Balances: The Impact of Solid Volume Fraction</b>	<b>123</b>
7.1	Introduction . . . . .	123
7.2	The Continuous MSM-PR System . . . . .	124
7.3	Case study: L-glutamic acid ( $\alpha$ and $\beta$ polymorphs) . . . . .	129
7.4	Conclusion . . . . .	138
<b>Bibliography</b>		<b>140</b>
<b>8</b>	<b>Ever-kinked Crystal System</b>	<b>141</b>
8.1	Introduction . . . . .	141
8.2	The Periodic Kink Problem . . . . .	142
8.3	The Ever-kinked Solution: 2-D case . . . . .	144
8.4	The Ever-kinked Solution: 3-D case . . . . .	147
8.5	Conclusions . . . . .	150
<b>Bibliography</b>		<b>152</b>

<b>9</b>	<b>Conclusions and Future Work</b>	<b>153</b>
9.1	Conclusions . . . . .	153
9.2	Future Work . . . . .	155
	<b>Bibliography</b>	<b>157</b>
<b>A</b>	<b>Deriving the finite size correction</b>	<b>158</b>
A.1	Breakdown of the Helmholtz free energy of a crystal . . . . .	158
A.2	Computing $Q_{EC}^{CM}/Q_{EC}$ : . . . . .	160
A.3	Computing $Q_C/Q_C^{CM}$ : . . . . .	162
A.4	Free energy of the crystal: . . . . .	164
<b>B</b>	<b>Ensuring zero net external force due to springs</b>	<b>165</b>
<b>C</b>	<b>T.I. plots</b>	<b>168</b>
C.1	T.I. integrand plot & Data for LiI crystal . . . . .	168
C.2	T.I. integrand plot & Data for NaCl crystal . . . . .	174
<b>D</b>	<b>Error Calculation</b>	<b>180</b>
D.1	Autocorrelation time . . . . .	181
<b>E</b>	<b>Force Field Predicted Crystal Structures</b>	<b>184</b>
<b>F</b>	<b>Sublimation Enthalpy Derivation</b>	<b>197</b>
<b>G</b>	<b>Activity and Activity Coefficients</b>	<b>201</b>
G.1	Activity . . . . .	201
G.2	Activity coefficients . . . . .	202
<b>H</b>	<b>Thermodynamic Checks</b>	<b>206</b>
H.1	$\mu^C(T, P^{sat})$ : . . . . .	206
H.2	Testing the centroid to bond transformations . . . . .	207
H.3	Sampling the COOH dihedral in the gas phase . . . . .	208
H.4	Long range vdW correction for crystalline systems . . . . .	210
	<b>Bibliography</b>	<b>214</b>
<b>I</b>	<b>Computing the free energy of the centroid</b>	<b>215</b>
<b>J</b>	<b>Potential Mean Force - COOH dihedral</b>	<b>219</b>
<b>K</b>	<b>Population Balance Deep Dive</b>	<b>223</b>
K.1	Regularity Condition . . . . .	223

<b>L</b>	<b>Solving the Ever-kinked Crystal System Equations</b>	<b>227</b>
L.1	Solving the 2D Ever-kinked orthorhombic lattice equations . . . . .	227
L.2	Solving the 3D Ever-kinked orthorhombic lattice equations . . . . .	239

# Chapter 1

## Digital Design of Crystals

Reproduced in part with permission from:

Christopher L. Burcham, Michael F. Doherty, Nicholas Francia, Vikram Khanna, Baron Peters, Sally L. Price, Susan M. Reutzel-Edens, Matteo Salvalaglio, “Digital Design of Crystals: From Molecules to Crystals,” *in preparation*.

Jeffery Frumkin, Vikram Khanna, and Michael F. Doherty, “Innovation in Chemical Reactor Engineering Practice and Science,” *Computers & Chemical Engineering*, *accepted*.

*If you can't model it, you don't understand it. If you don't understand it, you can't improve it.*

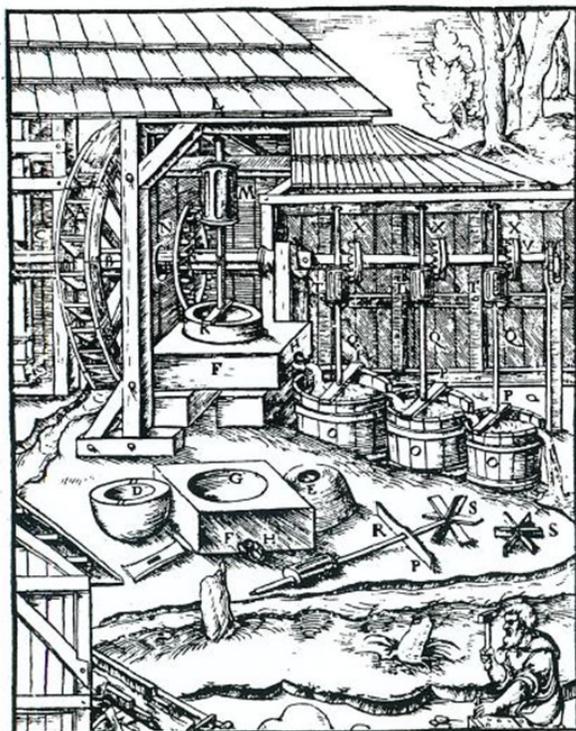
—Prof. Michael Doherty (at the start of every senior design class)

### **A brief history of chemical process design**

From the earliest times humans have discovered ways to use earth's chemical resources to improve the quality of their lives. Pigments were used to create cave paintings, later

leading to brightly colored paints and dyes used by ancient Egyptians to decorate their palaces, sarcophagi, clothing and other objects. Oils, soaps, cleansers, and perfumes were invented for beautification, and embalming. Later, the Chinese discovered saltpeter (potassium nitrate) from which they invented gunpowder. Chemicals were developed for extracting ores and tanning leather. And of course there was always demand for medicines and poisons, which were constantly being developed and improved throughout human history. Alchemists, and their modern incarnation, have been in demand for a very long time.

Along with improved chemical formulations came improved process technology, as seen in a lithograph of ore processing from the 1550's shown in Figure 1(a). Notice the three stirred tub reactors (CSTR's) connected in series. By 1556, our chemical engineering ancestors had already discovered that in some applications it was better to use three smaller stirred tanks in series instead of a single stirred vessel with three times the volume. Although it is unclear from the contemporary description of this chemical process whether reaction is taking place in these tubs (which would thus be a practical embodiment of a liquid-phase plug flow reactor), it is undeniable that by 1556 the concept of staging had already been discovered. While the process looks primitive, it is in fact quite ingenious.



MULTIPURPOSE ORE MILL, powered by a single water wheel at the upper left was illustrated in Georgius Agricola's *De re metallica*, published in 1556. Gold ore was processed in the following steps. First the ore was crushed by a cast-iron stamp (a), just visible to the left of the water wheel. Next the crushed ore was ground to a powder in a pair of millstones to the right of the wheel. Two spare dome-shaped upper millstones (d, e) are on the ground on each side of a

(a)



(b)

Figure 1.1: Chemical processes then and now. (a) Ore mill illustrated in *De re Metallica* by Georgius Agricola (Latinized pen name of author Georg Bauer), published in 1556. (b) View of the Shell Pearl gas-to-liquids plant complex in Qatar, built at a total cost of approximately US\$20 billion between 2010–2015. Figure adapted and reprinted from Frumkin et al.<sup>[1]</sup>

In recent decades, design methods have been created for: heat exchanger networks using pinch technology, membrane cascades, azeotropic & reactive distillation systems, pressure swing adsorption, batch & continuous crystallization systems, and so much more. Simulation, optimization (particularly mixed integer nonlinear programming and global optimization) and control (especially model predictive control) of complex chemical devices and systems have improved enormously. All these advances coupled with the extraordinary improvements in digital computers are what led to the transition from

processes like the one shown in Figure 1(a) to the one shown in Figure 1(b).

In recent years, digital design techniques are being adopted as they offer the following advantages:

- Accelerated innovation: Through a structured exploration of the design space, we reduce experimentation time, thereby allowing for a faster time-to-market new products and processes.
- Better risk management: A digital design workflow serves as a digital twin of the real process. Therefore, the digital design method allows to screen the process parameter space and flag possible high risk outcomes.
- Reduced costs: Through screening of the process parameter space and accessing parameters which are expensive and/or difficult to achieve experimentally, computer simulations reduce the number of physical (expensive and/or time-consuming) experiments to be performed. Thus, ultimately reducing the cost of development.

A new breed of computational tool added to our toolkit—in recent decades—is molecular dynamics. This tool relies on mathematical description of atomic forces and Newton’s third law to study the motion of atoms. The atomic information is then leveraged to compute macroscopic properties via statistical mechanical theories. It is one of the most fundamental approaches to model the world we live in as quoted by the famous Physicist Richard Feynman, *Everything that living things do can be understood in terms of the jigglng and wiggling of atoms*. I would go one step ahead and modify this to: *Everything that living (and non-living) things do can be understood in terms of the jigglng and wiggling of atoms*. The following section lays out our digital design approach to engineer crystals harnessing state-of-the-art digital tools and techniques.

## The Digital Design Workflow for Engineering Crystals

Among the various unit operations developed by us humans, crystallization plays a significant role in various industries (see Figure 1.2) such as: i) food, ii) semiconductor, iii) chemical, and iv) pharmaceutical to name a few. In the pharmaceutical industry, crystallization has arguably a greater importance as: i) more than 90% of the small molecule drugs are delivered in crystalline form; ii) it is not only the product purity but also the shape and size of the crystals that are critical quality determining parameters. In addition, the cost of developing a novel drug candidate is reported to be approximately US\$800 million, increasing at an annual rate of 7.6%.<sup>[2]</sup> Therefore, digitizing the workflow of crystallization has a significant potential to reduce costs and risks.

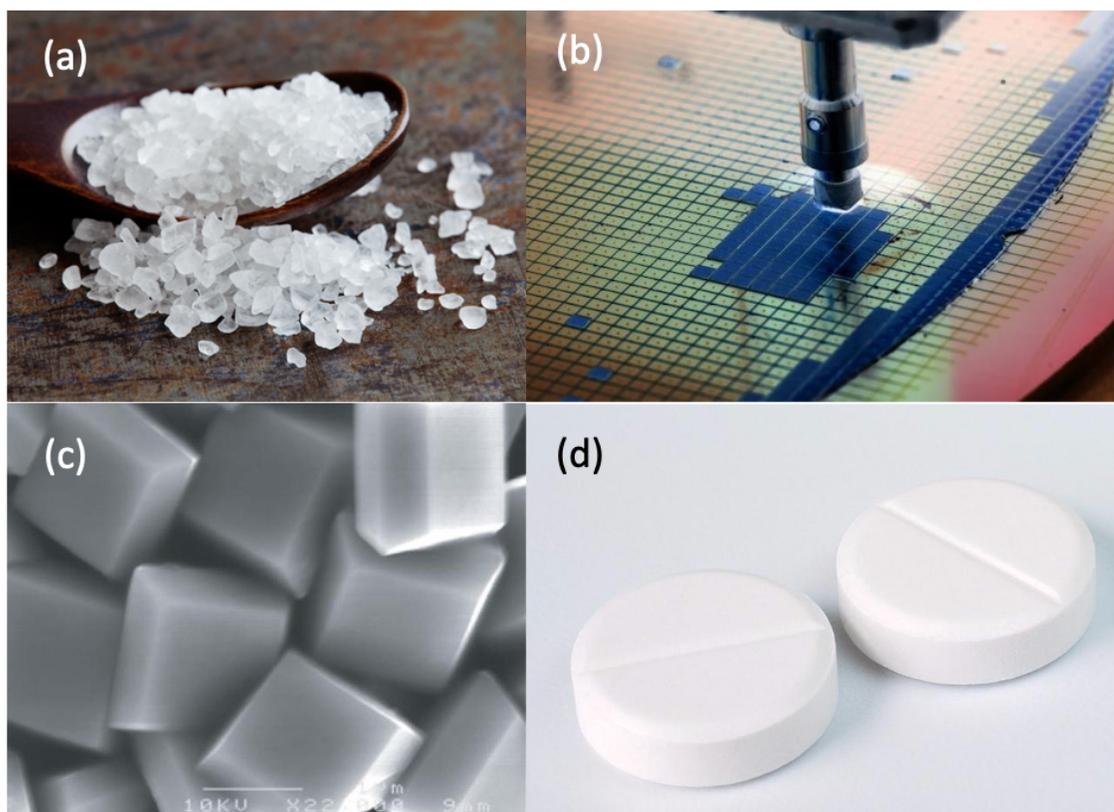


Figure 1.2: Examples of crystalline materials used in various industries: (a) salt crystals (food industry); (b) silicon wafer (semiconductor industry); (c) TiO<sub>2</sub> catalyst crystals (chemical industry); (d) paracetamol tablets (pharmaceutical industry)

Crystal habit (shape) prediction plays an important role in the pharmaceutical industry where we are interested in manufacturing crystals with certain shapes that lead to high bio-availability.<sup>[3-5]</sup> Certain shapes such as needles (high aspect ratio) are undesirable as they impact further processing (see Figure 1.3). Crystal habit affects dissolution rates and stability too.<sup>[6]</sup> The crystal size distribution affects dissolution rates, hence impacts *bioavailability*. It also affects filtration and drying rates<sup>[7]</sup> as well as product formulation parameters such as flow, compactibility, and content uniformity.<sup>[8]</sup>

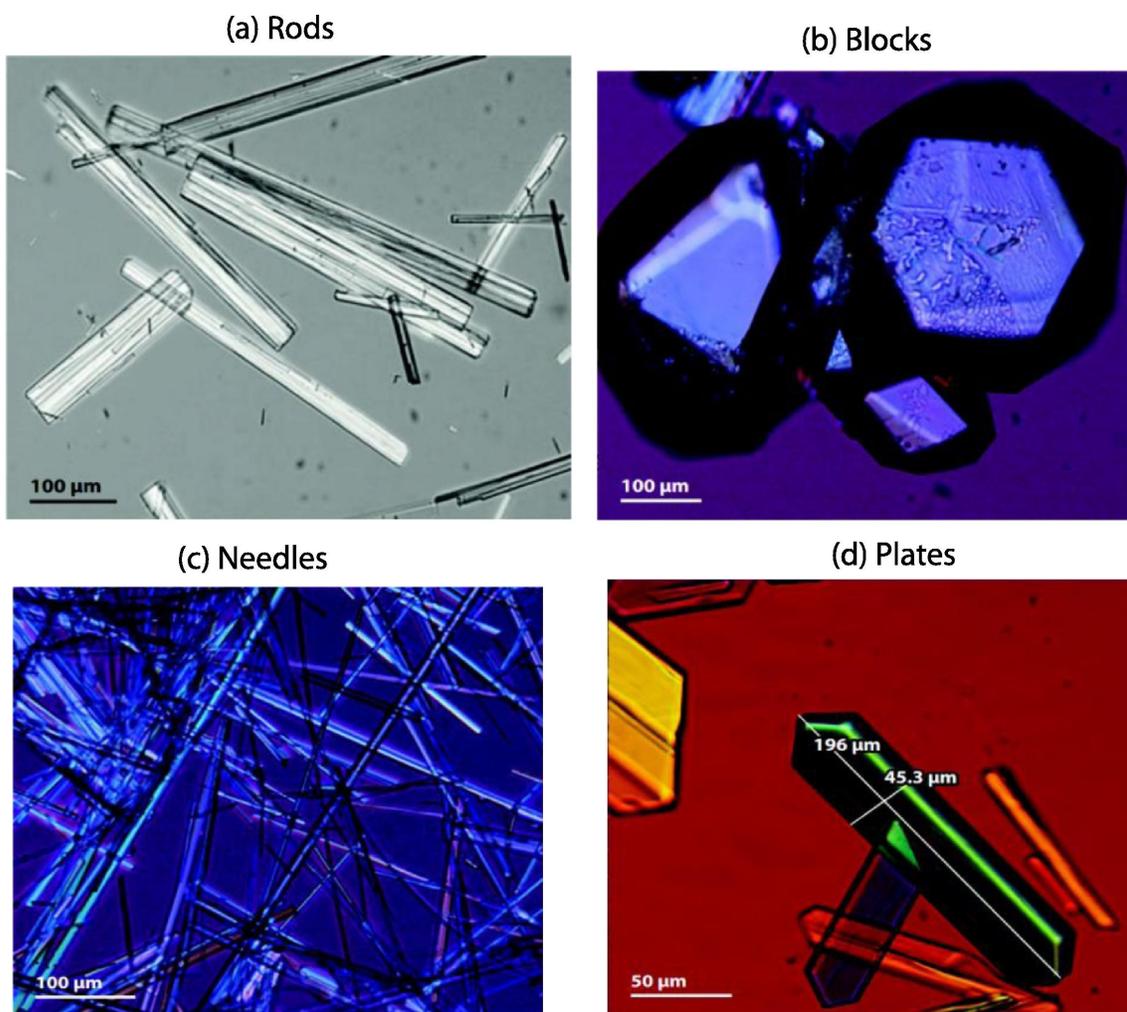


Figure 1.3: Examples of crystal growth shapes under optical microscopy: (a): rods, (b): blocks, (c): needles, (d): plates. Figure adapted and reproduced from Li et al.<sup>[9]</sup>

The most important input parameters to designing a crystal include the molecule being crystallized, its crystal structure, temperature, supersaturation, and growth medium. Therefore, in our digital design approach we begin with the molecular information of the product being crystallized as shown in Figure 1.4.

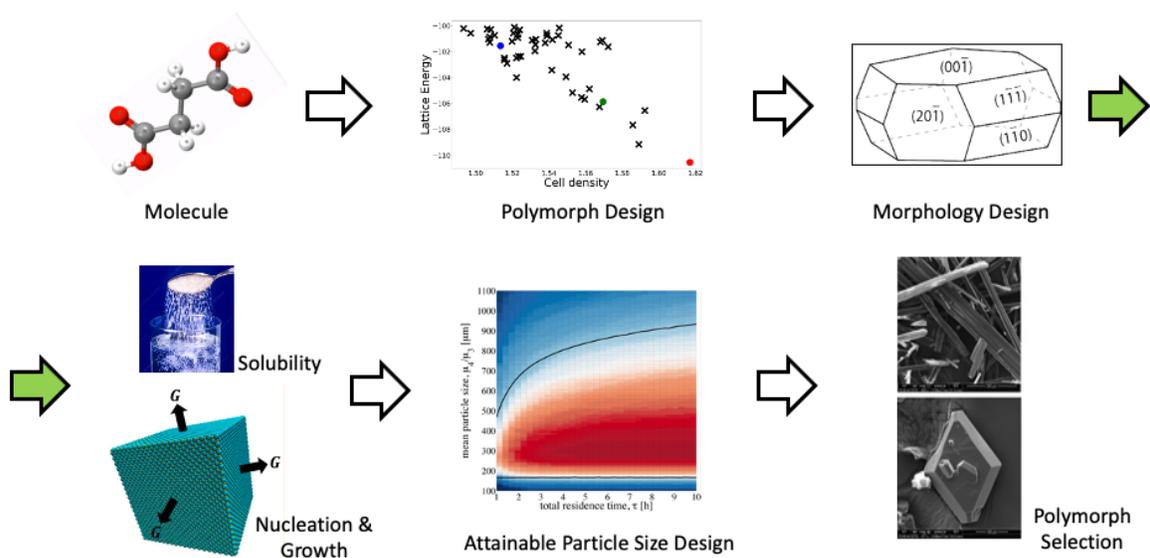


Figure 1.4: The digital design framework for crystal engineering

The next stages of the design are i) Polymorph prediction, ii) Morphology design, iii) Solubility and driving force predictions, iv) Nucleation and Growth rate predictions, v) Particle size distribution and its attainable region, and finally vi) Polymorph selection. Let us delve into each of these in detail.

## 1.1 Polymorph Prediction

One of the most fundamental question since the discovery of polymorphism in 1832 is perhaps—given the molecular structure of a chemical compound, can one predict its crys-

tal structure?<sup>[10–12]</sup> The pharmaceutical industry’s interest in crystal structure prediction (CSP) as a digital design tool is from desire to streamline the search for the possible crystal forms (i.e. the polymorphs of the neat API<sup>[13]</sup> and at least it’s hydrates<sup>[14]</sup>) in order to select the crystal form for manufacture. The main risk to avoid is the late appearance of a more stable form than the one under development, as this may lead to the “disappearance”<sup>[15]</sup> or sudden lack of control of the manufacturing process. However, a CSP study can be a very useful complement to the experimental screening and characterisation of solid forms,<sup>[16]</sup> potentially providing confidence that the most stable form is known, and designing an appropriate experimental search for the metastable forms that will affect the process development. The ideal Crystal Structure Prediction (CSP) computational code would predict not all the polymorphs that can be formed, but only those that can be experimentally realised — and give a recipe for obtaining the first sample of each polymorph. This is indeed the ultimate aim,<sup>[17]</sup> but the series of blind tests of crystal structure prediction, which have been organised by the Cambridge Crystallographic Data Centre<sup>[18]</sup> show that this is still an aspiration in this rapidly developing area. Currently, the type of CSP that is being increasingly applied in industry<sup>[19]</sup> is CSP\_0, a search for the structures that are the most stable minima in the lattice energy. This is the energy required to separate a (hypothetical) static infinite perfect crystal into infinitely separated molecules in their lowest energy conformation, approximating the relative stability at 0 K. Since the relative stability of polymorphs often changes with temperature and pressure, we need to develop the calculation of a crystal energy landscape at ambient conditions. Other thermodynamic factors, such as the balance of bulk and surface energies, reflecting particle size and environment (water activity/solvent) should also be taken into account<sup>[20]</sup> as this can affect the relative stability of polymorphs, and can lead to the observation of new polymorphs in confined crystallisation experiments.<sup>[21]</sup> A major limitation of CSP\_0 is that it usually generates more crystal structures within the likely

energy range of polymorphism than are found experimentally.<sup>[22, 23]</sup> Many of these structures may be artefacts of the neglect of the temperature-dependent molecular motion within the crystals. Other CSP\_0 structures could be kinetically forbidden, because of the relative kinetics of the nucleation and growth of different polymorphs relative to the ability to transform to the most stable form. The current understanding of the kinetic competition involved in apparent polymorph stability is not yet sufficiently mature to be encapsulated into a CSP workflow. Francia et al.,<sup>[24]</sup> have devised a systematic coarsening approach of the CSP\_0 generated crystal energy landscape harnessing MD simulation methods and clustering using custom fingerprints. Figure 1.5 outlines their workflow and Figure 1.6 shows their results for succinic acid in which they successfully reduce the CSP\_0 structures by over 70% at 300 K.

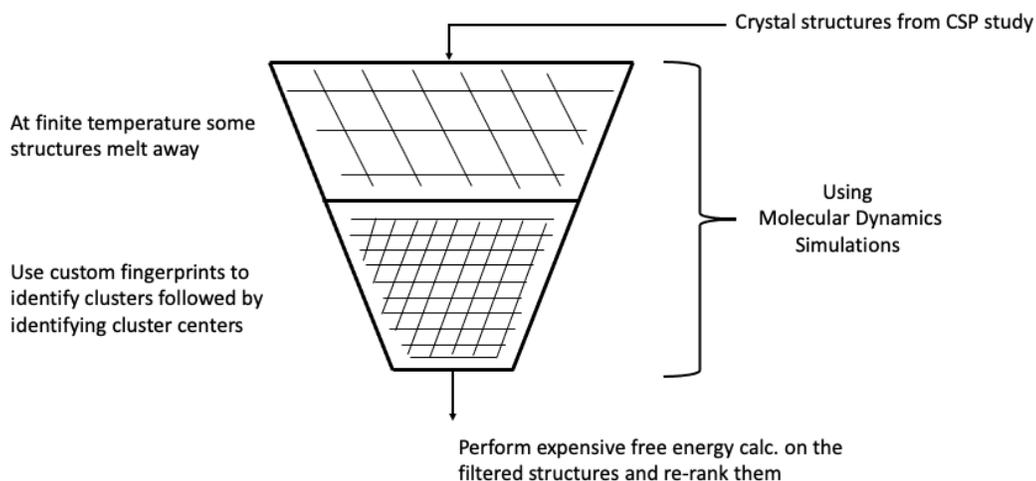


Figure 1.5: The polymorph filtering framework at finite temperatures.

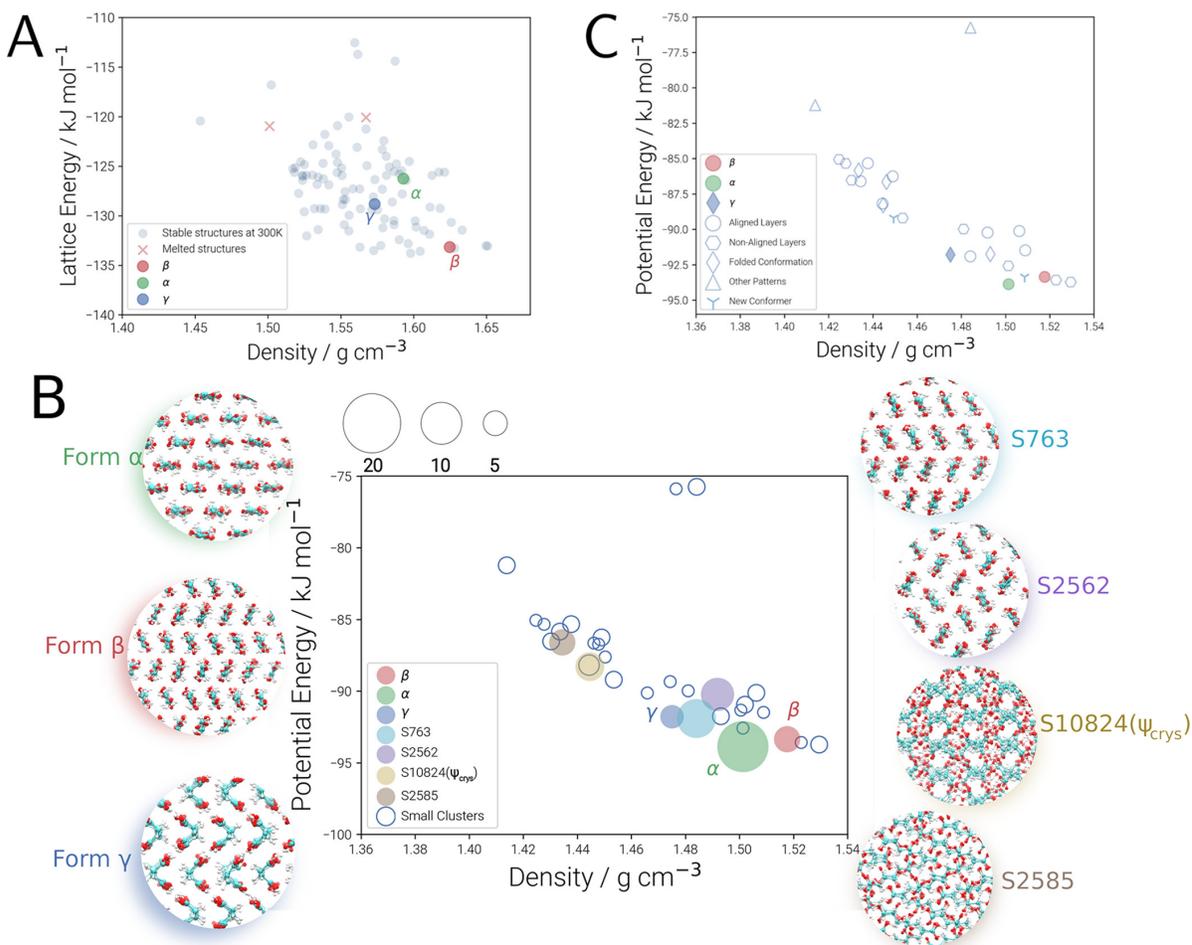


Figure 1.6: Analysis of the finite-temperature structures of succinic acid. (A) Lattice energy landscape of the CSP\_0-generated structures of succinic acid optimized with the GAFF force field at 0 K showing those structures that melt at 300 K as red crosses. (B) Finite-temperature crystal energy landscape of the cluster centers. The size of each point refers to the number of structures that converted to the same geometry. Known structures and most populated clusters are shown on either side of the plot. The different clusters are labeled based on their cluster center. (C) Final finite-temperature crystal energy landscape classified by the motifs observed and including the new structures found during WTmetaD simulations according to the color bar on the lower left side. Figure adapted and reproduced from Francia et al.<sup>[24]</sup>

## 1.2 Morphology Design

Once we have the polymorph of interest, we are interested in predicting the shape and size of the crystal. The size of the crystal requires the use of absolute growth rates of each crystal face, however, to predict the shape, one only needs the relative growth rates.<sup>[25–27]</sup> Our group at University of California Santa Barbara (UCSB) and the Koo group at Sogang University have been developing and upgrading mechanistic frameworks to model complex crystals. Towards predicting shapes, our group has developed AD-DICT—Advanced Design and Development of Industrial Crystallization Technology, a computer software that automates the mechanistic framework and predicts crystal shapes based on the input crystallography, temperature, supersaturation and solvent medium by computing relative growth rates of crystal faces. Figure 1.7 shows the results of shape predictions for a few complex molecules. Chapter 2 will dive into further details of the digital approach to make these predictions.

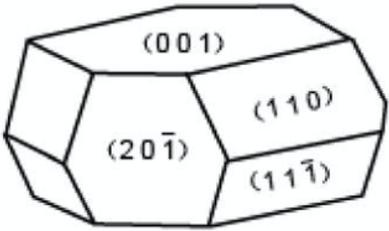
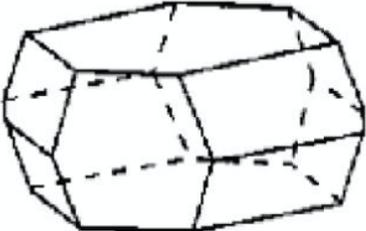
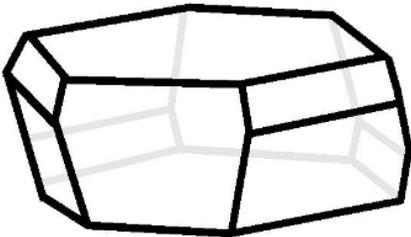
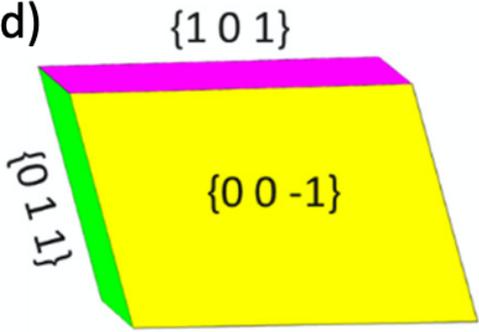
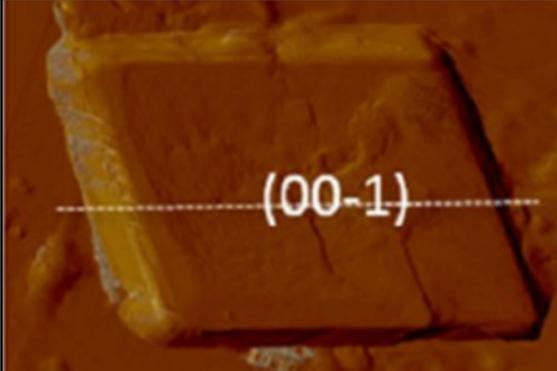
Predicted (ADDICT)	Experimental
<p>(a)</p> 	
<p>(b)</p> 	
<p>(c)</p> 	
<p>(d)</p> 	

Figure 1.7: ADDICT shape predictions compared against experimental shapes: a) Naphthalene grown in ethanol,<sup>[9]</sup> b) Anthracene grown from vapor,<sup>[9]</sup> c) Lovastatin grown in methanol,<sup>[9]</sup> d) Olanzapine dihydrate grown in water.<sup>[28]</sup>

## 1.3 Solubility

The solubility of a solute is its concentration in a solution at which it is in thermal, mechanical and *chemical* equilibrium with the solid phase. Solubility plays a crucial role in the pharmaceutical industry as the solubility of the Active Pharmaceutical Ingredient (API) directly impacts its bioavailability. Additionally, the solubility of a compound is a key prediction for self-consistent in-silico workflow for computing nucleation and growth rates.<sup>[29–31]</sup> This is because, it is the solubility of a compound that determines the driving force of crystallization for a given concentration of solute in solution. Therefore, for in-silico nucleation and growth rate predictions, we need in-silico tools to predict the driving forces for these processes! This dissertation focuses on these tools, and lays out a computational framework to predict driving forces for crystallization using atomistic simulations and advanced statistical thermodynamic frameworks.

## 1.4 Nucleation and Growth Rates

Predicting the nucleation and growth rates of crystals is perhaps one of the most difficult challenges in the digital design workflow. For both these phenomena, there is still need for a fundamental understanding of the underlying processes, in particular, nucleation. In fact, the freezing of water, i.e., the nucleation of ice, continues to yield surprises.<sup>[32–36]</sup> Nucleation plays a decisive role in determining the crystal structure and size distribution for solution crystallization.<sup>[37]</sup> Therefore, its understanding is crucial—to have control over the entire crystallization process. Crystal nucleation can occur in a clear supersaturated solution—primary nucleation—or due to the presence of parent crystals—secondary nucleation. Further, primary nucleation can occur in the bulk volume of a particle free solution —homogeneous primary nucleation — or at inter-

faces (e.g. solution-dust interface, solution-air interface, solution-wall interface, etc.) — heterogeneous nucleation.

Currently, there exists two major theories that describe primary homogeneous nucleation: i) Classical Nucleation Theory (CNT),<sup>[38–40]</sup> and ii) Two-step Nucleation Theory.<sup>[41, 42]</sup> The CNT is a single step theory in which nucleation takes place by the sequential addition of single atoms/molecules to clusters. Whereas, in the two-step nucleation theory, a crystalline nucleus appears inside a dense liquid metastable phase of solute molecules which has intermediate stability. While the understanding of homogeneous primary nucleation is still ongoing, primary heterogeneous nucleation and secondary nucleation (the ubiquitous and dominant nucleation pathways in practical applications) still continue to be an unsolved challenge.

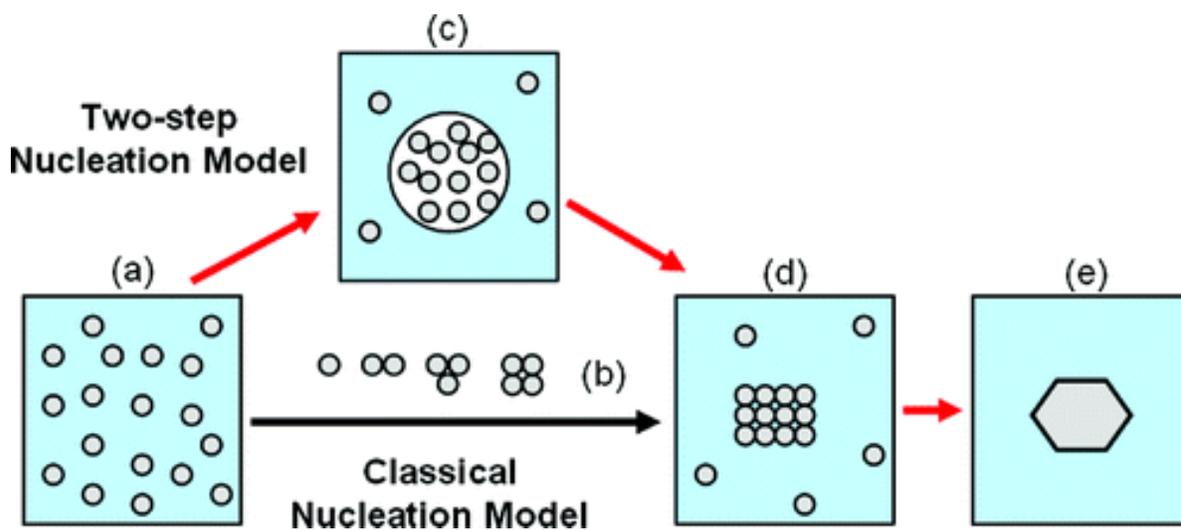


Figure 1.8: Alternative nucleation pathways leading from solution to solid crystal: (a) supersaturated solution; (b) ordered subcritical cluster of solute molecules, proposed by classical nucleation theory; (c) liquid-like cluster of solute molecules, dense precursor proposed by two-step nucleation theory; (d) ordered crystalline nuclei; (e) solid crystal. Figure adapted and reproduced from Erdemir et al.<sup>[37]</sup>

The story on the growth side is a bit more optimistic. It is now well understood that desolvation of the kink site (a feature on the crystal surface where growth units dock, see

Figure 1.9) and the crystal growth unit is the rate-limiting step.<sup>[43, 44]</sup> A systematic approach to understanding the important reaction co-ordinates along which crystal growth can be understood has been laid out by Joswiak et al.<sup>[44]</sup> Also, a key computation of kink attachment rates that feed into a growth model has been demonstrated by Joswiak et al.<sup>[31]</sup> for sodium chloride. Using these attachment rates and the spiral growth model (discussed in Chapter 2) Joswiak et al. show that computing *in silico* absolute growth rates of crystals is now possible. Their work has laid the groundwork needed to analyse more complex crystals.

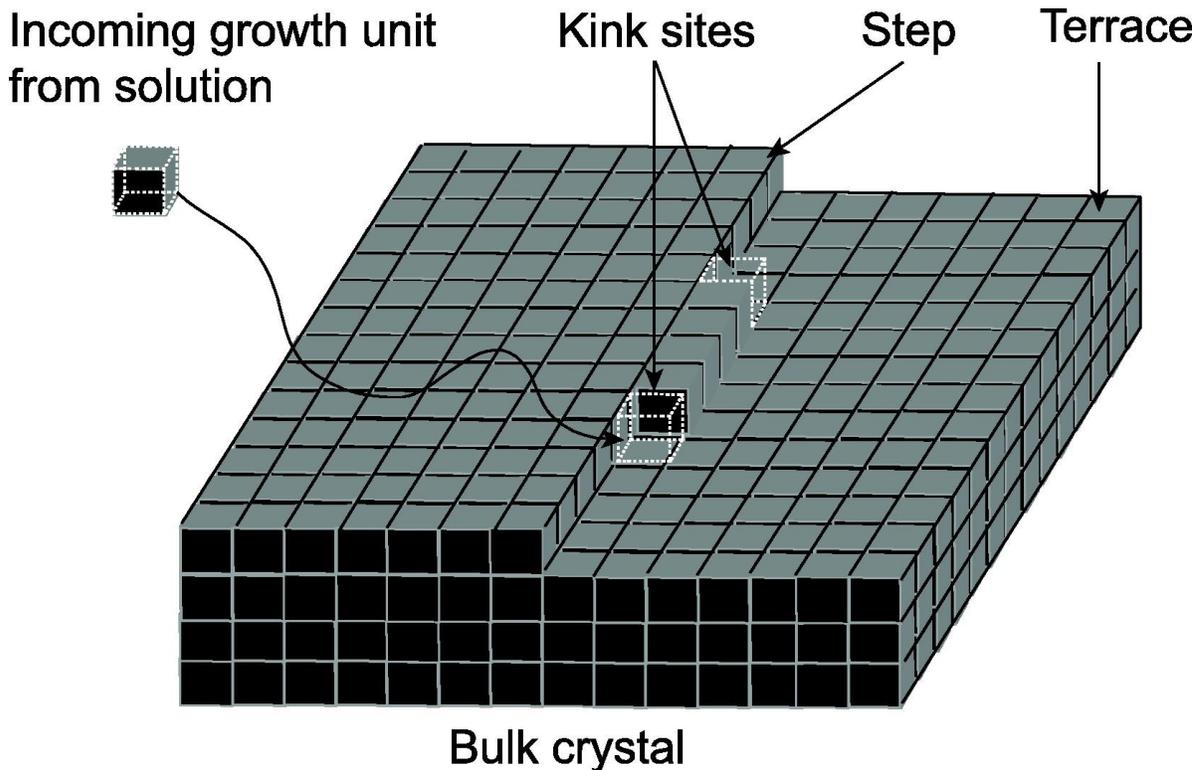


Figure 1.9: Kink sites on a crystal surface. Figure adapted and reproduced from Li et al.<sup>[9]</sup>

## 1.5 Attainable Particle Size Design

Once we have the solubility limits, the nucleation and growth rates, these parameters can then be fed into a population balance model to compute the crystal size distribution. Knowing the attainable region of crystal sizes, it is possible to generate feasible process alternatives that allow specific crystal sizes to be obtained in a given process configuration. Inspired by the attainable region approach used in the design of chemical reactor networks and separation systems, Vetter et al.<sup>[45]</sup> extend the computational methodology to crystallization systems. It is useful to determine whether a desired mean particle size can be achieved in a specific crystallizer type. Figure 1.10 shows the attainable particle size regions for three compounds, a) Paracetamol, b) L-asparagine monohydrate, and c) Aspirin, using different crystallizer configurations.

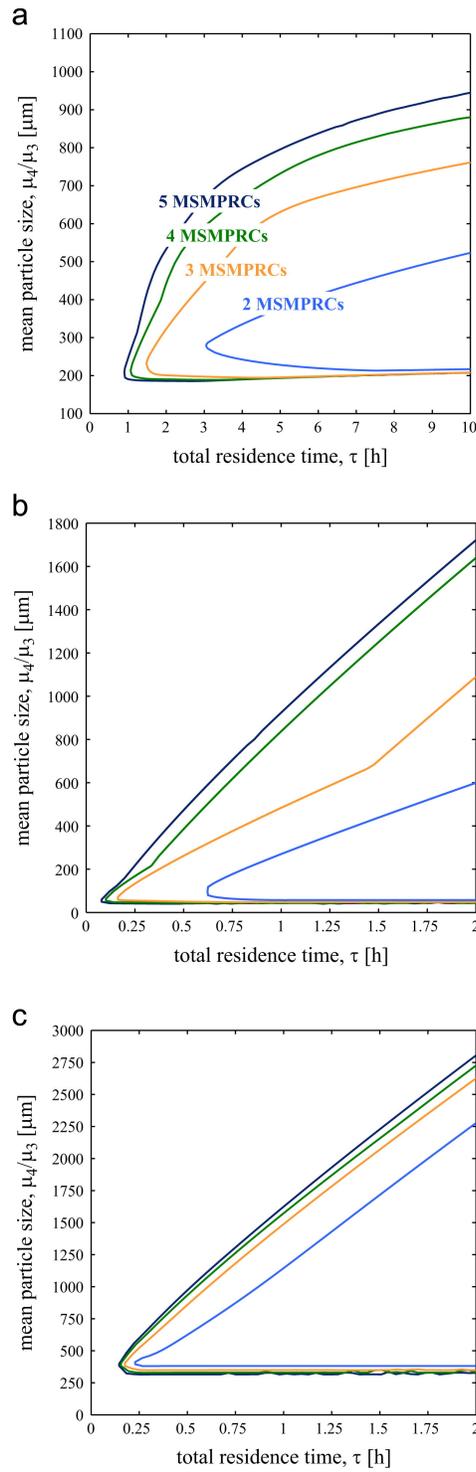


Figure 1.10: Attainable Particle Size Regions for a) Paracetamol, b) L-asparagine, c) Aspirin using cascades of mixed suspension mixed product removal (MSMPRC) crystallizers. Figure adapted and reproduced from Vetter et al.<sup>[45]</sup>

## 1.6 Polymorph Selection

As mentioned earlier, crystalline polymorphism — or the ability of a compound to exist in multiple solid-state forms — has significant impact on a crystal’s physical properties, performance and safety (for ingestible crystals, e.g. Active Pharmaceutical Ingredients, i.e., APIs). Hence, its control is a key objective in crystallization processes. In batch systems, one straightforward method to select and control a polymorph’s crystallization is by seeding the polymorph of interest and preventing primary nucleation in the crystallizer.<sup>[46, 47]</sup> However, polymorph control in continuous crystallization was elusive until 2015 when Tsai and co-workers<sup>[48]</sup> produced the metastable polymorphs of L-glutamic acid and p-aminobenzoic acid in a continuous crystallizer at stable steady-state operation. Learning about this amazing discovery, Farmer and co-workers<sup>[49]</sup> leveraged linear stability analysis to explain why this phenomenon occurs and gave simple design rules to engineer a crystallizer to produce a desired polymorph.<sup>[49]</sup> Figure 1.11 shows the predictive accuracy of their model.

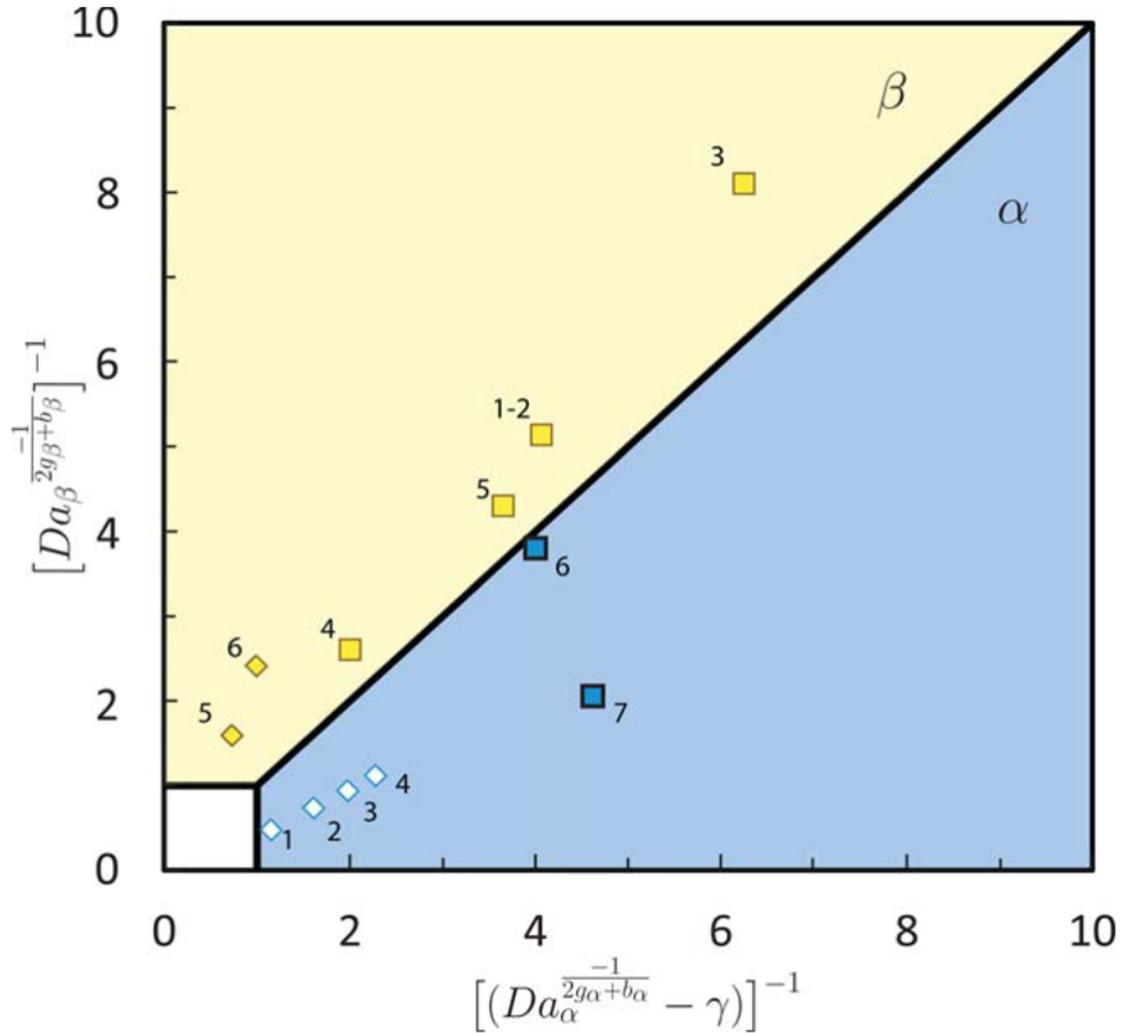


Figure 1.11: There are three stability regions. In the upper left region,  $\beta$  is the dominant polymorph. In the lower right region,  $\alpha$  is dominant, and in the bottom left corner ( $\lim \tau \rightarrow 0$ ) the trivial steady-state is stable. Yellow data points correspond to the observation of the  $\beta$  polymorph experimentally and blue data points correspond to the observation of  $\alpha$  experimentally. Open markers denote thermodynamic metastability of the solid observed at the temperature corresponding to that experiment, while filled markers denote thermodynamically stable solid forms. Figure adapted and reproduced from Farmer et al.<sup>[49]</sup>

The input parameters for their model are: i) nucleation rates, ii) growth rates, and iii) relative solubilities of the potential polymorphs, iv) residence time, v) inlet concentration, and vi) temperature. These parameters are used to compute modified Damkohler

numbers based on which we can predict the polymorph that will be crystallized at steady-state as shown in Figure 1.11. Using this tool, we can dial in the parameters under our control (like residence time, inlet supersaturation, temperature, etc.) to produce our desired polymorph as shown in Figure 1.12.

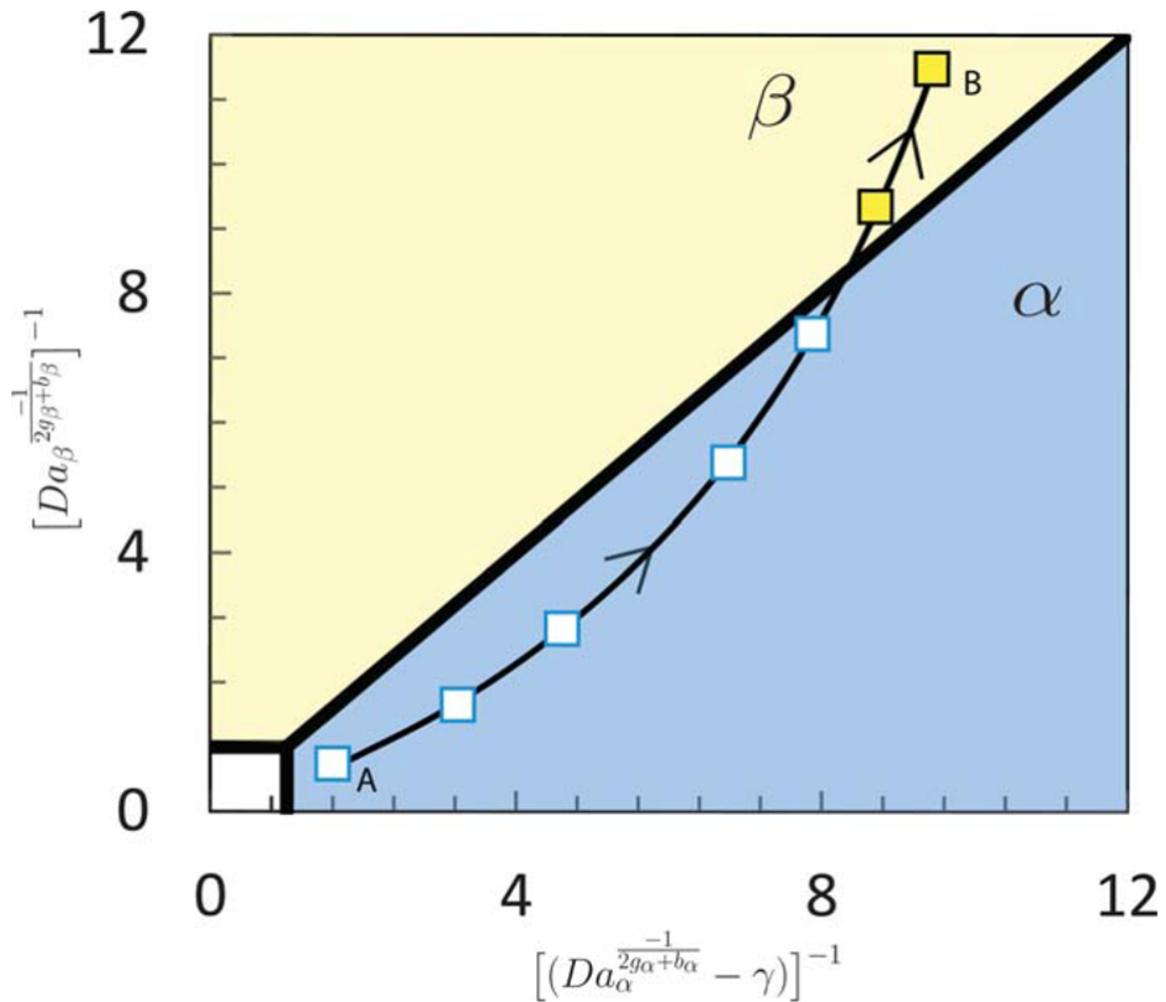


Figure 1.12: The filled square labeled A represents  $\tau = 60$  min, and the filled square label B represents  $\tau = 6000$  min;  $C_o$  (inlet concentration) =  $40 \text{ kg/m}^3$ , and  $T = 25^\circ\text{C}$ . The increasing- $\tau$  solution branch moves toward the bifurcation line and eventually crosses as  $\tau$  increases. Figure adapted and reproduced from Farmer et al.<sup>[49]</sup>

# Bibliography

- [1] J. Frumkin, V. Khanna, and M. F. Doherty, Innovation in chemical reactor engineering practice and science, *Computers & Chemical Engineering*, (*accepted*).
- [2] M. Greiner, E. Elts, J. Schneider, K. Reuter, and H. Briesen, Dissolution study of active pharmaceutical ingredients using molecular dynamics simulations with classical force fields, *Journal of Crystal Growth* **405** (2014) 122–130.
- [3] N. Blagden, M. de Matas, P. T. Gavan, and P. York, Crystal engineering of active pharmaceutical ingredients to improve solubility and dissolution rates, *Advanced drug delivery reviews* **59** (2007), no. 7 617–630.
- [4] S. R. Modi, A. K. Dantuluri, V. Puri, Y. B. Pawar, P. Nandekar, A. T. Sangamwar, S. R. Perumalla, C. C. Sun, and A. K. Bansal, Impact of crystal habit on biopharmaceutical performance of celecoxib, *Crystal Growth & Design* **13** (2013), no. 7 2824–2832.
- [5] R. B. Hammond, K. Pencheva, K. J. Roberts, and T. Auffret, Quantifying solubility enhancement due to particle size reduction and crystal habit modification: case study of acetyl salicylic acid, *Journal of pharmaceutical sciences* **96** (2007), no. 8 1967–1973.
- [6] N. Variankaval, A. S. Cote, and M. F. Doherty, From form to function: Crystallization of active pharmaceutical ingredients, *AIChE journal* **54** (2008), no. 7 1682–1688.
- [7] R. Beck, A. Häkkinen, D. Malthe-Sørensen, and J.-P. Andreassen, The effect of crystallization conditions, crystal morphology and size on pressure filtration of l-glutamic acid and an aromatic amine, *Separation and Purification technology* **66** (2009), no. 3 549–558.
- [8] H. Garekani, F. Sadeghi, A. Badiie, S. Mostafa, A. R. Rajabi-Siahboomi, and A. Rajabi-Siahboomi, Crystal habit modifications of ibuprofen and their physicomechanical characteristics, *Drug development and industrial pharmacy* **27** (2001), no. 8 803–809.
- [9] J. Li, C. J. Tilbury, S. H. Kim, and M. F. Doherty, A design aid for crystal growth engineering, *Progress in Materials Science* **82** (2016) 1–38.
- [10] J. Maddox, Crystals from first principles, *Nature* **335** (1988), no. 6187 201–201.
- [11] A. Gavezzotti, Are crystal structures predictable?, *Accounts of chemical research*

## BIBLIOGRAPHY

---

- 27** (1994), no. 10 309–314.
- [12] J. D. Dunitz, Are crystal structures predictable?, *Chemical Communications* (2003), no. 5 545–548.
- [13] J. Bernstein, *Polymorphism in Molecular Crystals 2e*, vol. 30. International Union of Crystal, 2020.
- [14] S. M. Reutzel-Edens, D. E. Braun, and A. W. Newman, *Hygroscopicity and hydrates in pharmaceutical solids*, vol. 2. Wiley-VCH, 2018.
- [15] D.-K. Bučar, R. W. Lancaster, and J. Bernstein, Disappearing polymorphs revisited, *Angewandte Chemie International Edition* **54** (2015), no. 24 6972–6993.
- [16] S. L. Price and S. M. Reutzel-Edens, The potential of computed crystal energy landscapes to aid solid-form development, *Drug Discovery Today* **21** (2016), no. 6 912–923.
- [17] S. L. Price, Is zeroth order crystal structure prediction (csp\_0) coming to maturity? what should we aim for in an ideal crystal structure prediction code?, *Faraday discussions* **211** (2018) 9–30.
- [18] A. M. Reilly, R. I. Cooper, C. S. Adjiman, S. Bhattacharya, A. D. Boese, J. G. Brandenburg, P. J. Bygrave, R. Bylsma, J. E. Campbell, R. Car, *et. al.*, Report on the sixth blind test of organic crystal structure prediction methods, *Acta Crystallographica Section B: Structural Science, Crystal Engineering and Materials* **72** (2016), no. 4 439–459.
- [19] J. Nyman and S. M. Reutzel-Edens, Crystal structure prediction is changing from basic science to applied technology, *Faraday discussions* **211** (2018) 459–476.
- [20] A. M. Belenguer, G. I. Lampronti, A. J. Cruz-Cabeza, C. A. Hunter, and J. K. Sanders, Solvation and surface effects on polymorph stabilities at the nanoscale, *Chemical science* **7** (2016), no. 11 6617–6627.
- [21] M. D. Ward, Perils of polymorphism: size matters, *Israel Journal of Chemistry* **57** (2017), no. 1-2 82–92.
- [22] D. E. Braun, J. A. McMahon, R. M. Bhardwaj, J. Nyman, M. A. Neumann, J. van de Streek, and S. M. Reutzel-Edens, Inconvenient truths about solid form landscapes revealed in the polymorphs and hydrates of gandotinib, *Crystal Growth & Design* **19** (2019), no. 5 2947–2962.
- [23] S. L. Price, Why don't we find more polymorphs?, *Acta Crystallographica Section B: Structural Science, Crystal Engineering and Materials* **69** (2013), no. 4 313–328.
- [24] N. F. Francia, L. S. Price, J. Nyman, S. L. Price, and M. Salvalaglio, Systematic finite-temperature reduction of crystal energy landscapes, *Crystal Growth & Design* **20** (2020), no. 10 6847–6862.
- [25] F. Frank, The influence of dislocations on crystal growth, *Discussions of the Faraday Society* **5** (1949) 48–54.

## BIBLIOGRAPHY

---

- [26] F. Frank, On the kinematic theory of crystal growth and dissolution processes in growth and perfection of crystals (eds. rh doremus, bw roberts & d. turnbull) 411–417, 1958.
- [27] A. Chernov, The kinetics of the growth forms of crystals, *Sov. Phys. Cryst* **7** (1963) 728–730.
- [28] Y. Sun, S. M. Reutzel-Edens, R. M. Bhardwaj, and M. F. Doherty, Crystal morphology modeling of solvates and hydrates of organic molecular crystals: Olanzapine solvate and dihydrate, *Crystal Growth & Design* **21** (2021), no. 9 4871–4877.
- [29] N. E. Zimmermann, B. Vorselaars, D. Quigley, and B. Peters, Nucleation of NaCl from aqueous solution: Critical sizes, ion-attachment kinetics, and rates, *Journal of the American Chemical Society* **137** (2015), no. 41 13352–13361.
- [30] N. E. Zimmermann, B. Vorselaars, J. R. Espinosa, D. Quigley, W. R. Smith, E. Sanz, C. Vega, and B. Peters, NaCl nucleation from brine in seeded simulations: Sources of uncertainty in rate estimates, *The Journal of Chemical Physics* **148** (2018), no. 22 222838.
- [31] M. N. Joswiak, B. Peters, and M. F. Doherty, In silico crystal growth rate prediction for nacl from aqueous solution, *Crystal Growth & Design* **18** (2018), no. 10 6302–6306.
- [32] R. Radhakrishnan and B. L. Trout, Nucleation of hexagonal ice (i h) in liquid water, *Journal of the American Chemical Society* **125** (2003), no. 25 7743–7747.
- [33] B. Slater and D. Quigley, Zeroing in on ice, *Nature materials* **13** (2014), no. 7 670–671.
- [34] E. Sanz, C. Vega, J. Espinosa, R. Caballero-Bernal, J. Abascal, and C. Valeriani, Homogeneous ice nucleation at moderate supercooling from molecular simulation, *Journal of the American Chemical Society* **135** (2013), no. 40 15008–15017.
- [35] G. Bullock and V. Molinero, Low-density liquid water is the mother of ice: on the relation between mesostructure, thermodynamics and ice crystallization in solutions, *Faraday discussions* **167** (2013) 371–388.
- [36] L. Ickes, A. Welti, C. Hoose, and U. Lohmann, Classical nucleation theory of homogeneous freezing of water: thermodynamic and kinetic parameters, *Physical Chemistry Chemical Physics* **17** (2015), no. 8 5514–5537.
- [37] D. Erdemir, A. Y. Lee, and A. S. Myerson, Nucleation of crystals from solution: classical and two-step models, *Accounts of chemical research* **42** (2009), no. 5 621–629.
- [38] M. Volmer and A. Weber, Keimbildung in übersättigten gebilden, *Zeitschrift für Physikalische Chemie* **119U** (1926), no. 1 277–301.
- [39] L. Farkas, Keimbildungsgeschwindigkeit in übersättigten dampfen, *Zeitschrift für Physikalische Chemie* **125U** (1927), no. 1 236–242.

## BIBLIOGRAPHY

---

- [40] F. F. Abraham, *Homogeneous nucleation theory*, vol. 263. Elsevier, 1974.
- [41] P. R. ten Wolde and D. Frenkel, Enhancement of protein crystal nucleation by critical density fluctuations, *Science* **277** (1997), no. 5334 1975–1978.
- [42] V. Talanquer and D. W. Oxtoby, Crystal nucleation in the presence of a metastable critical point, *The Journal of chemical physics* **109** (1998), no. 1 223–227.
- [43] P. G. Vekilov, What determines the rate of growth of crystals from solution?, *Crystal Growth and Design* **7** (2007), no. 12 2796–2810.
- [44] M. N. Joswiak, M. F. Doherty, and B. Peters, Ion dissolution mechanism and kinetics at kink sites on nacl surfaces, *Proceedings of the National Academy of Sciences* **115** (2018), no. 4 656–661.
- [45] T. Vetter, C. L. Burcham, and M. F. Doherty, Regions of attainable particle sizes in continuous and batch crystallization processes, *Chemical Engineering Science* **106** (2014) 167–180.
- [46] N. C. Kee, R. B. Tan, and R. D. Braatz, Selective crystallization of the metastable  $\alpha$ -form of l-glutamic acid using concentration feedback control, *Crystal Growth and Design* **9** (2009), no. 7 3044–3051.
- [47] N. C. Kee, P. D. Arendt, R. B. Tan, and R. D. Braatz, Selective crystallization of the metastable anhydrate form in the enantiotropic pseudo-dimorph system of l-phenylalanine using concentration feedback control, *Crystal Growth and Design* **9** (2009), no. 7 3052–3061.
- [48] T.-T. C. Lai, S. Ferguson, L. Palmer, B. L. Trout, and A. S. Myerson, Continuous crystallization and polymorph dynamics in the l-glutamic acid system, *Organic Process Research & Development* **18** (2014), no. 11 1382–1390.
- [49] T. C. Farmer, C. L. Carpenter, and M. F. Doherty, Polymorph selection by continuous crystallization, *AIChE Journal* **62** (2016), no. 9 3505–3514.

## Chapter 2

# Crystal Growth and the Solubility Prediction Framework

In this chapter we discuss the fundamental processes that lead to crystal growth and the underlying parameters that determine the growth rate of a crystal—solubility being a major one. Next, we discuss the solubility prediction framework that will enable us to make precise *in silico* driving force predictions for crystal growth.

### 2.1 The crystal growth framework

The most basic event leading to the growth of a crystal is the incorporation of building blocks (growth units) which could be atoms, ions, molecules, dimers, etc. from the growth medium (vapor, solution, gel, etc.) into the crystal.<sup>[1]</sup> The growth unit preferentially attaches into the crystal at sites called kinks as shown in Figure 2.1. The series of events (for solution growth of organic crystals) which lead to a growth unit's incorporation into a kink site are as follows:<sup>[2-4]</sup>

1. Solute molecules are transported from the bulk solution to the crystal face via

- convection and diffusion;
2. Adsorption of the solute molecule on the crystal face followed by surface diffusion to the kink site;
  3. Desolvation of kink sites and solute molecules occurs;
  4. The solute molecules are incorporated into the kink sites; and
  5. The latent heat of crystallization is released and transported to the crystal and solution.

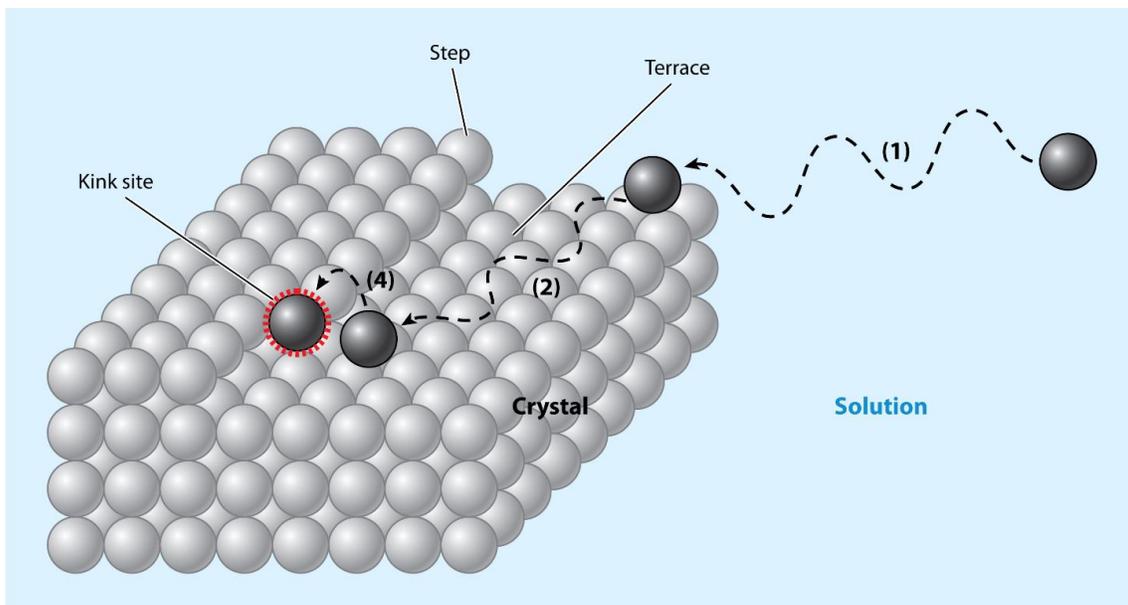


Figure 2.1: Steps during crystal growth: 1) Transport of solute to crystal face, 2) Surface diffusion, 3) Desolvation (not shown for clarity purpose), 4) Surface integration

The third and fourth steps, which comprise of desolvation of the kink site and molecule along with the attachment of the solute at the kink are the rate limiting steps at low to moderate supersaturations.<sup>[1]</sup> At these conditions, the crystal undergoes a layered growth by the following two mechanisms:

1. Spiral growth: spirals emerge from screw dislocations
2. formation of two-dimensional (2D) nuclei and their growth.

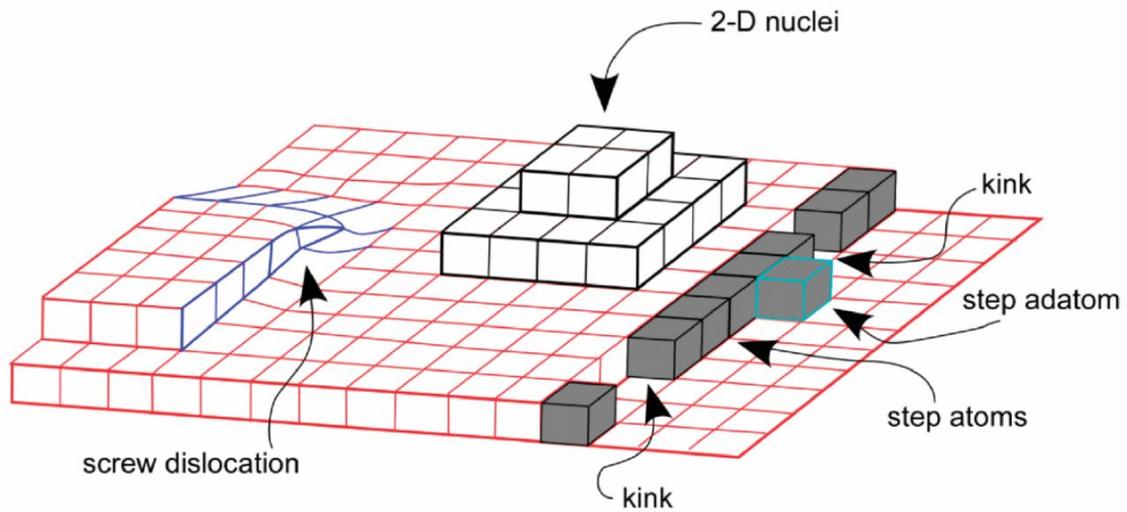


Figure 2.2: Schematic of screw dislocations and 2D nuclei

Mechanistic models such as the spiral growth model require the computation of free energy barriers for kink attachment and detachment as shown in Figure 2.3. When fed with this information, the mechanistic models give us the growth rate of a crystal.

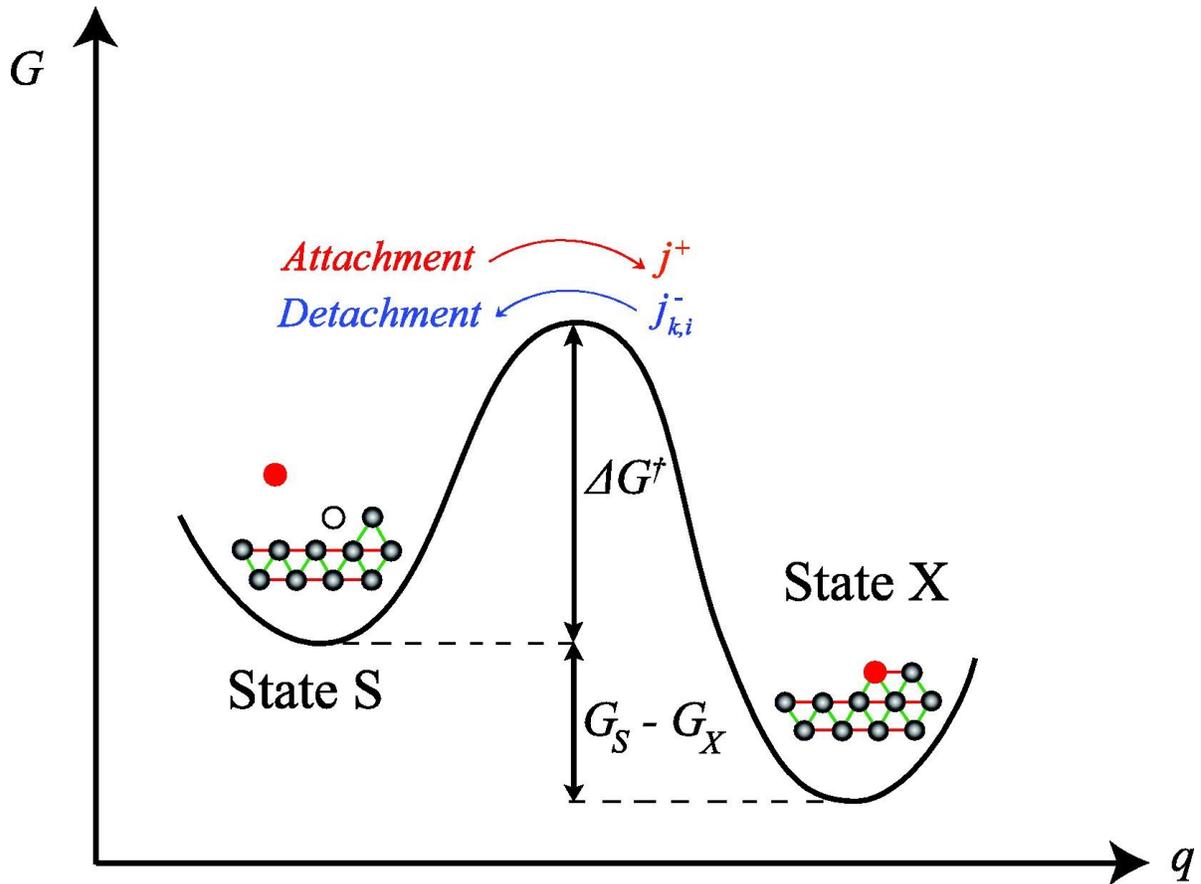


Figure 2.3: Schematic free energy diagram for the attachment and detachment of solute growth units from kink sites. States S and X represent growth units in solution and incorporated to the crystal, respectively;  $j^+$  and  $j_{k,i}^-$  are the corresponding attachment and detachment rates. A simplified reaction coordinate  $q$  could relate to the distance of the growth unit from the kink site. The filled red circles represent the growth unit which is attaching to/detaching from the kink site and the open black circle represents the unoccupied kink site. Image adapted and reprinted from Li et al.<sup>[5]</sup>

### Shape and Size Prediction

The equilibrium shape of a crystal is one which minimizes its Gibbs Free energy for a fixed volume or mass of crystal.<sup>[6]</sup> For a given crystal, the thermodynamically stable

shape is given by the famous Wulff construction,<sup>[7, 8]</sup>

$$\frac{\gamma_1}{H_1} = \frac{\gamma_2}{H_2} = \dots = \frac{\gamma_i}{H_i}, \quad (2.1)$$

where  $\gamma_i$  is the surface free energy of face  $i$ , and  $H_i$  is its perpendicular distance from the center of the crystal. However, it turns out it is often the kinetics and not the thermodynamics that determines the shape of the crystal (see Figure 2.4).

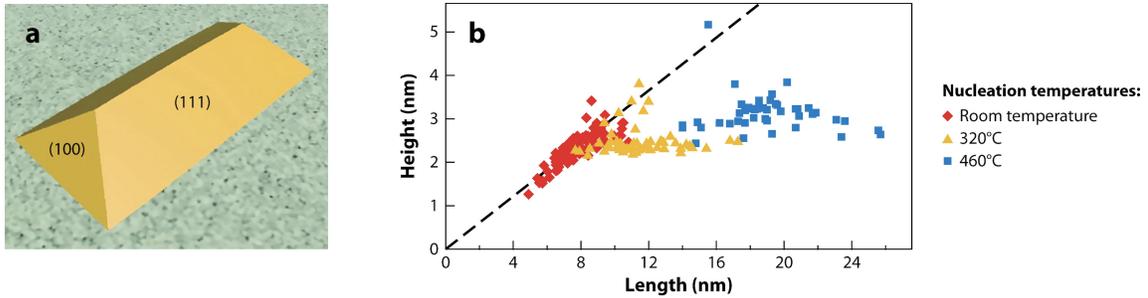


Figure 2.4: (a) 3D model of Pd nanocrystals (golden) grown on a SrTiO<sub>3</sub> (001) substrate (gray) in an ultrahigh-vacuum environment. (b) Evolution of the height and length of the Pd nanocrystals. The dashed line indicates the equilibrium shape, whereas the different markers for the data points indicate different nucleation temperatures.<sup>[4, 9]</sup>

The shape predicted by the kinetics is modeled using the Frank-Chernov<sup>[10, 11]</sup> condition,

$$\frac{G_1}{H_1} = \frac{G_2}{H_2} = \dots = \frac{G_i}{H_i}, \quad (2.2)$$

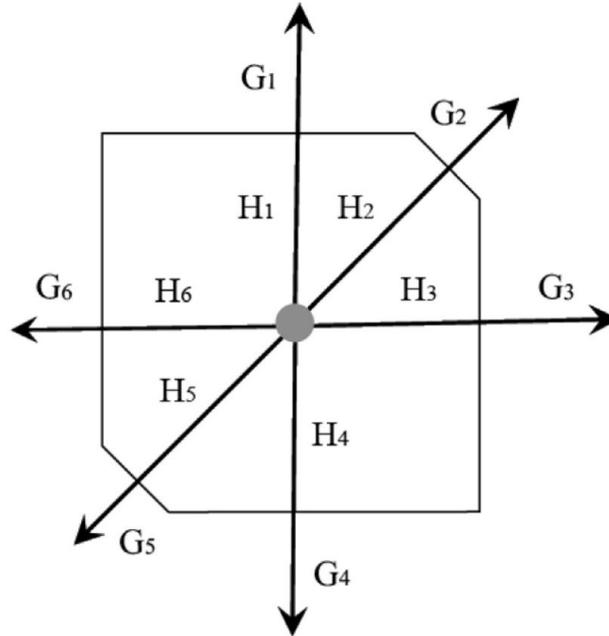


Figure 2.5: Wulff construction using the Frank-Chernov condition

where,  $G_i$  is the absolute growth rate of face  $i$ , and  $H_i$  is its perpendicular distance from the center of the crystal. As we can see from the Frank-Chernov equation, only relative growth rates of the crystal faces and their relative perpendicular distances are required to predict the shape of the crystal.

The growth rate of the crystal face is a function of the kink rate, kink density and the height of the kink.<sup>[12]</sup> The kink rate further is a function of the attachment free energy barrier and detachment work (see Figure 2.3). It so turns out that while calculating relative rates it is a good approximation to assume the attachment rates are the same for each face. Hence they fall off from the equation, and we can get good shape predictions.<sup>[5]</sup>

However, to compute absolute growth rates, there is no alternative, but to compute the attachment rates of growth units at a kink site. We at the Doherty and Peters labs are developing methods to efficiently compute these rates and use them to refine the

shape predictions and more importantly to also get the size of the crystal by computing absolute growth rates and feeding them into a population balance model such as,

$$\frac{\partial n}{\partial t} = -G_1 \left( \frac{\partial n}{\partial H_1} + R_2 \frac{\partial n}{\partial H_2} + \dots + R_n \frac{\partial n}{\partial H_k} \right) + \frac{n_{in}}{\tau} - \frac{n}{\tau} \quad (2.3)$$

where,  $G_1$  is the absolute growth rate of a reference face,  $R_i$  is the relative growth rate of face  $i$  to the reference face,  $n$  is the number density distribution of the crystals, and  $\tau$  is the crystallizer residence time. Figure 2.8 summarizes the workflow to compute the growth rate of crystal faces and ultimately the shape and size of the crystal.

### Spiral Growth Mechanism<sup>[2, 13]</sup>

At low supersaturations, crystals primarily grow via a spiral growth mechanism. A screw dislocation as shown in Figure 2.2 is the origin of the spiral. It exposes an edge on which the growth units dock. This docking of growth units leads to the advancement of the step. After the step has advanced to a critical length ( $l_c$ ), it exposes another edge which grows in the same manner as described above. This process occurs recursively forming a spiral and leads to a layered growth of the crystal face as shown in Figures 2.6 and 2.7.

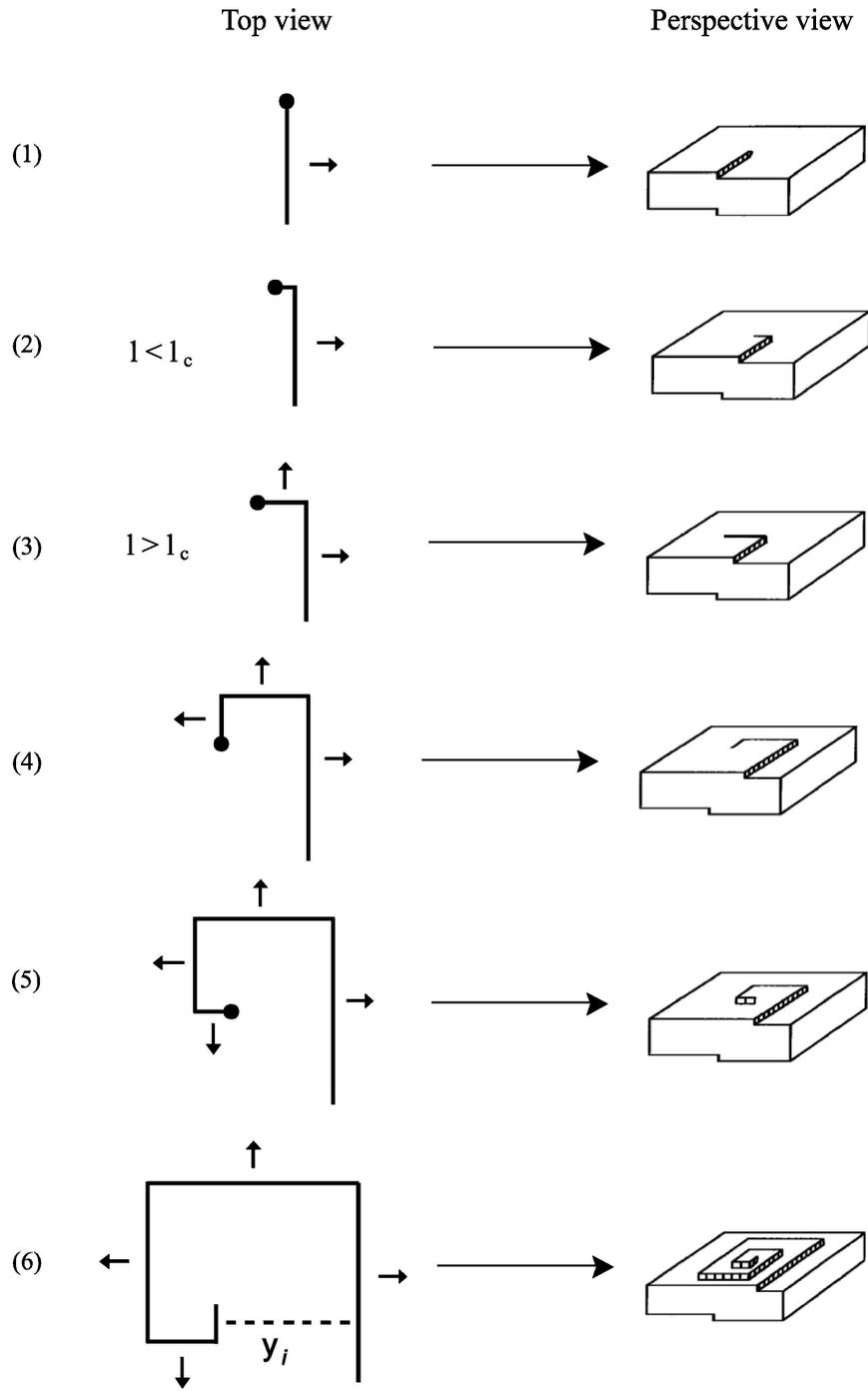


Figure 2.6: The mechanism of spiral growth around a screw dislocation<sup>[5]</sup>

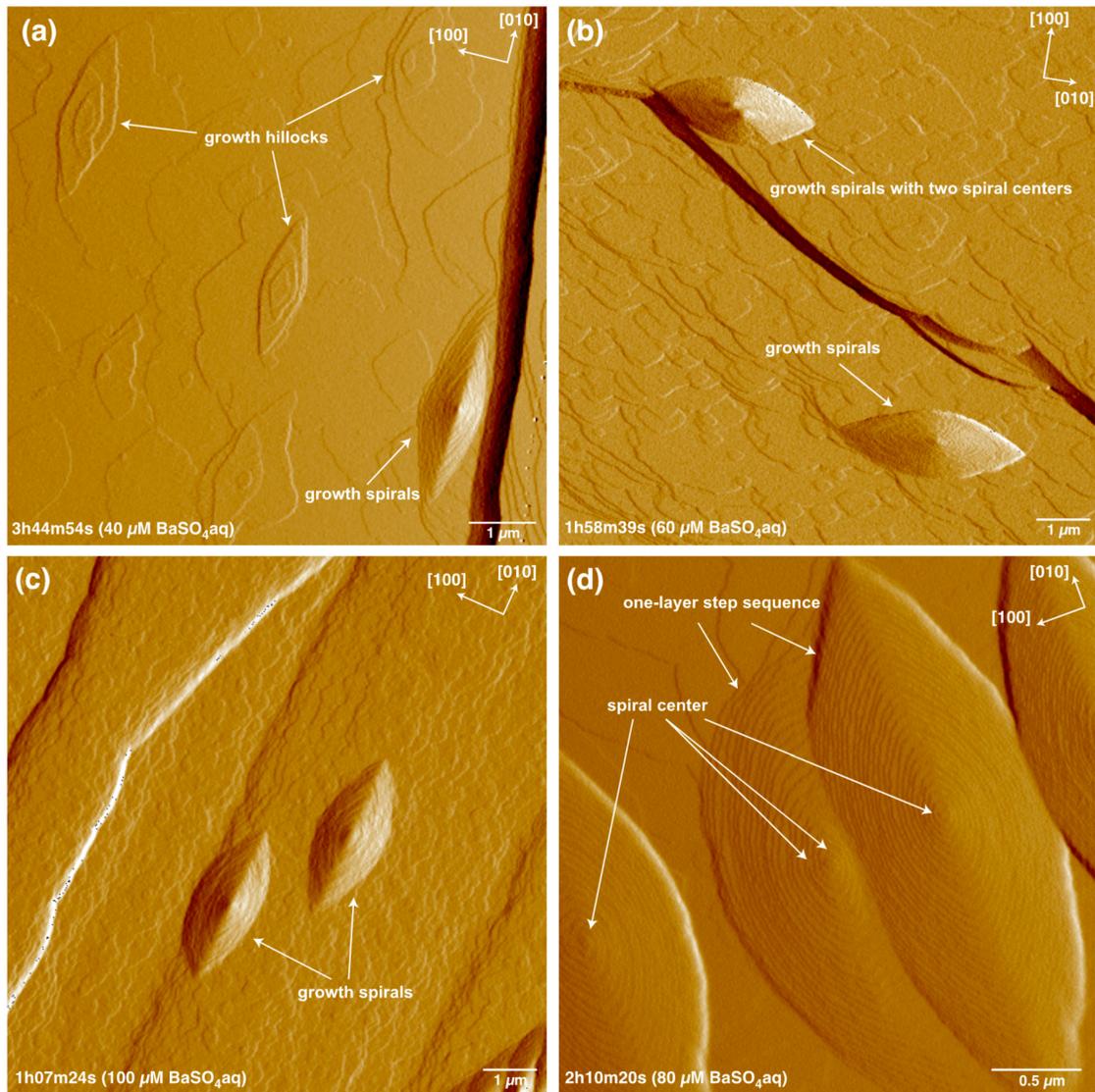


Figure 2.7: (a–c) CMAFM deflection images of a (001) surface in 40 μM, 60 μM, and 100 μM BaSO<sub>4</sub> solutions at 30 °C, respectively, showing the growth spirals that formed from screw dislocations. The growth hillocks that likely formed from edge dislocations are also shown in (a). (d) CMAFM deflection image of a (001) surface in an 80 μM BaSO<sub>4</sub> solution at 30°C showing growth spirals with a one-layer step sequence and regular step spacing. Figure adapted and reproduced from Kuwahara et al.<sup>[14]</sup>

The growth rate ( $G$ ) of a crystal face is given by

$$G = \frac{h}{\tau_s}, \quad (2.4)$$

where  $h$  is the height of the edge/kink and  $\tau_s$  is the spiral rotation time, i.e., the time taken to complete one rotation. The spiral rotation time is a function of the step velocity which in turn is a function of the kink's physical properties and net attachment rate. It follows that<sup>[15, 16]</sup>:

$$G = [k^+ x_{sat}] [(S - 1) \ln(S)] \left[ \frac{h}{\tau_s^*} \right] \quad (2.5)$$

where  $k^+$  is the attachment rate constant of the solute at the kink site (the computation of which requires the calculation of the attachment/detachment barriers),  $x_{sat}$  is the mole fraction of the solute at its solubility limit,  $S$  is the supersaturation,  $h$  is the height of the kink and  $\tau_s^*$  is a function of solvent and temperature as given in equation (54) in Ref. [15].

Thus, the growth rate of a face of a crystal is a function of:

1. solubility & supersaturation; (driving forces)
2. the attachment rate constant;
3. kink density & kink geometry (site availability)

In order to compute growth rates, we need to compute the driving forces for crystallization—and hence solubility of the crystal—predicted via the force field along with the rate constants and kink densities to have a self-consistent growth rate prediction. Therefore, next we look at the solubility prediction framework to compute the driving forces for crystallization processes.

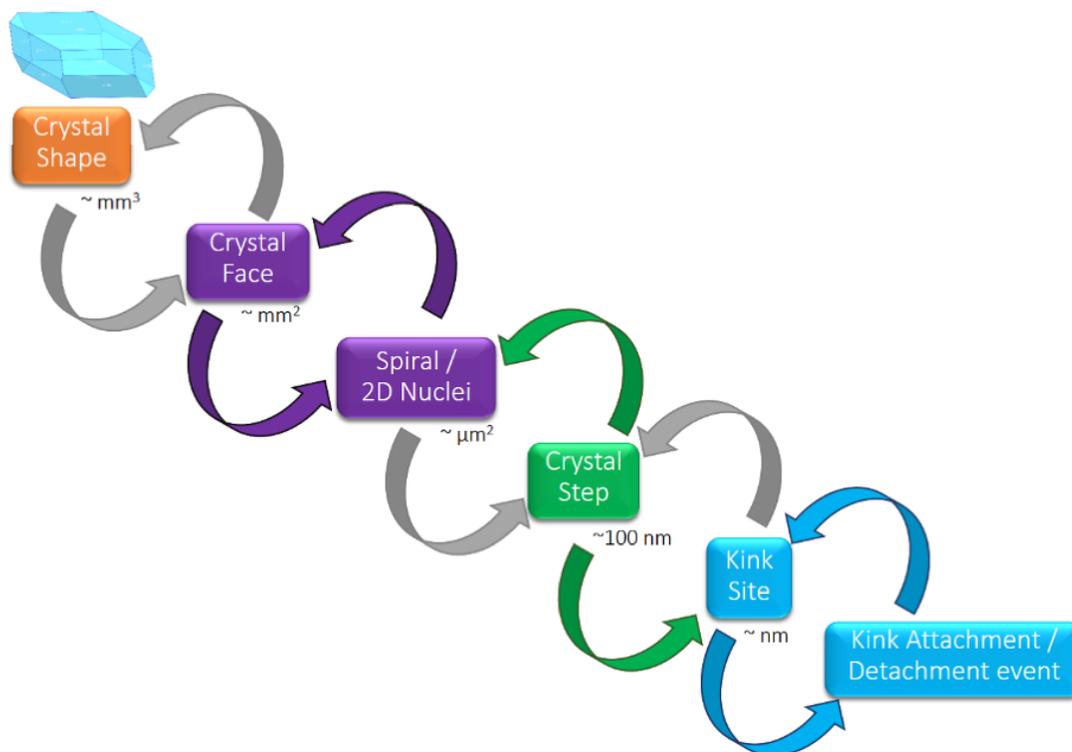


Figure 2.8: The mechanistic framework to compute the shape of a crystal spanning length scales from  $\mathcal{O}(nm)$  to  $\mathcal{O}(mm)$ . Image courtesy: Mark Joswiak and Carl Tilbury

## 2.2 The solubility prediction framework

Computational methods to predict phase equilibria and free energy differences between phases are critical for studies of nucleation and growth.<sup>[17–19]</sup> Some of the first methods predicted phase equilibria for single-component systems and gas–liquid phase equilibria, e.g., histogram reweighting,<sup>[20]</sup> Gibbs ensemble simulations,<sup>[21]</sup> and Gibbs–Duhem integration methods.<sup>[22]</sup> Crystalline solids pose special challenges that require new methods. In particular, all degrees of freedom in a crystalline solid are bound, while the molecules in a fluid phase are all free to translate and rotate. These differences have inspired the development of very different computational methods for free energy calcu-

lations and phase equilibria between fluids and solids. Examples include direct coexistence simulations,<sup>[23, 24]</sup> thermodynamic integration from a harmonic model at 0 K,<sup>[25]</sup> and harmonically mapped average techniques<sup>[26, 27]</sup> and transformations from real crystal to Einstein crystal (EC)/Einstein molecule crystal to atoms or molecules in a fluid phase (the Frenkel–Ladd approach).<sup>[28]</sup>

Direct coexistence simulations have two important limitations. First, direct coexistence simulations are prone to errors arising from long time scale processes such as attachment and detachment at kinks,<sup>[29, 30]</sup> 1D nucleation of kinks,<sup>[31–33]</sup> and 2D island or pit nucleation.<sup>[34–36]</sup> All three activated processes have been implicated as sources of error in direct coexistence simulations.<sup>[23, 37, 38]</sup> Simulations with special “everkinked” crystal orientations can eliminate the row/kink and island/pit nucleation steps,<sup>[38, 39]</sup> but the attachment and detachment rates still set a fundamental limit on the efficiency of direct coexistence results. Second, direct coexistence simulations can only estimate equilibrium concentrations, e.g., the equilibrium solubility limit, the melting temperature, or the partial pressure for evaporation. They cannot estimate the chemical potential differences at non-equilibrium supersaturated conditions that drive crystallization.

Frenkel–Ladd<sup>[28]</sup> and related methods<sup>[40, 41]</sup> can provide chemical potentials and chemical potential differences even under nonequilibrium supersaturated conditions. The free energy difference between the gas and solid phases coupled with the solvation free energy<sup>[42–45]</sup> provides solubilities and driving forces for solute precipitate nucleation and growth. These approaches must either start from two reference systems (e.g., one for the solid and one for the vapor), or else compute the free energy required to transform the solid into a fluid. The solid-to-fluid transformation step is often affected in stages.<sup>[40, 41]</sup> For example, in the first step, one can relax restraints on positions of atoms in the molecular Einstein crystal to obtain a “freely rotatable Einstein crystal”<sup>[41]</sup> of non-interacting molecules with restrained locations. Then, the solid-to-fluid transition is completed by

allowing the molecules to translate throughout the lower density of the fluid. The free energy contribution for the second step is an analytic calculation, but relaxing the rotational and conformational degrees of freedom in the first step requires computation. The difficulty in this first step will increase for large and flexible molecules.

In this dissertation, we present an alternate route to solid–fluid chemical potential differences as shown in Figure 2.9. Instead of performing a solid-to-fluid transformation,<sup>[40, 41]</sup> we compute absolute free energies of the solute in the two phases. For the solid phase, we start with an Einstein crystal and use thermodynamic integration (T.I.) to compute the absolute free energy of the crystal (including the intra-molecular free energy). For the gas phase, we introduce a reference system called the “centroid.” We compute the free energy of the centroid and transform it to the gas phase molecule using thermodynamic integration. We demonstrate this approach by computing a solid–vapor equilibrium prediction. Towards this, we review procedures for computing the solid phase chemical potential and then present procedures using the centroid and T.I. to compute the gas phase chemical potential.

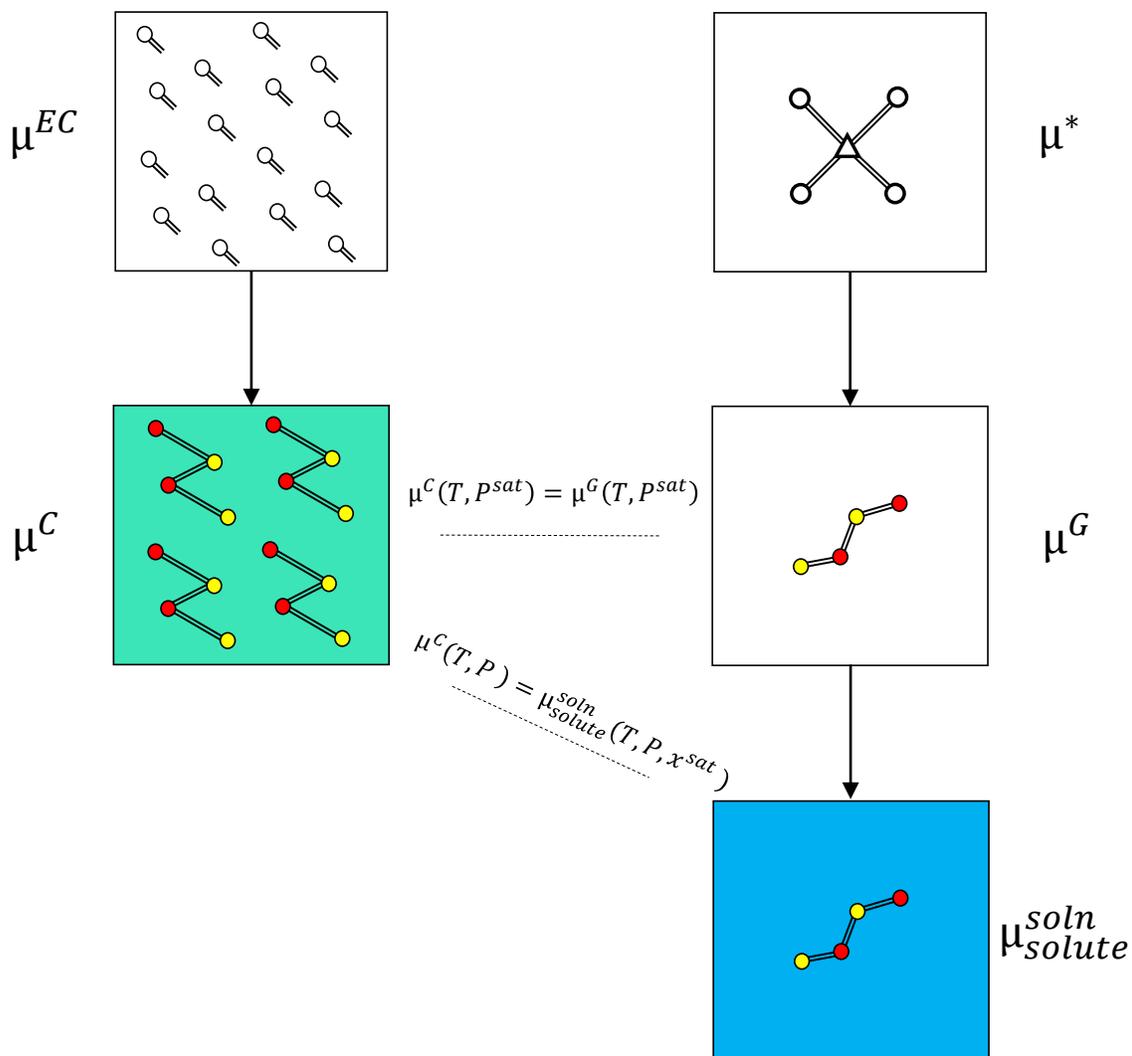


Figure 2.9: The solubility computation framework

To illustrate these methods for floppy gas phase molecules without the need for harmonic approximations, we use the new approach to compute the solid–vapor equilibrium for naphthalene (benchmark the framework with a rigid molecule) and succinic acid in the temperature range of 300 K–350 K. We set the gas phase chemical potential to the solid

chemical potential and solve to obtain the sublimation vapor pressure,  $P^{sat}$ . We then perform additional solvation free energy calculations to compute the chemical potential of the solutes in the solution phase—and ultimately compute the solubility by setting the solution phase chemical potential of the solute equal to the solid. The remaining dissertation is outlined as follows:

- Chapter 3 completes the Frenkel-Ladd theory for computing solid chemical potentials starting from any set of spring constants, this results in ca. 10% increase in phase equilibrium predictions in existing literature. It also extends the Frenkel-Ladd method to compute absolute solid free energies of solids, i.e., including the intra-molecular free energies. This is essential for a decoupled route starting from absolute references.
- Chapter 4 introduces the centroid—a new reference system for computing gas phase chemical potentials pivotal in decoupling the phase equilibrium calculations. It maps out the thermodynamic framework to compute the centroid’s free energy and the free energy difference of the molecule from the centroid. The chapter also discusses the thermodynamic test results to benchmark this new approach by making use of a simple diatomic molecule system.
- Chapter 5 lays out the computational strategy to perform solvation free energy calculations using the decoupling route—an efficient method that saves computational and — more so — implementation time. The chapter also discusses a new application of solvation free energy calculations for computing solvent modified bond energies from first principles—the input parameters for crystal shape prediction that so far relied on experimentally fitted parameters.
- Chapter 6 leverages the tools developed to compute absolute chemical potentials

of the solute in the solid, gas, and solution phases to predict solid-fluid equilibria and driving forces for crystallization from vapor and solution.

- Chapter 7 revisits the population balance and mass balance models for a mixed suspension mixed product removal (MSMPR) crystallizer to study the impact of the solid volume fraction on the accuracy of the current models being used in literature. It provides the accurate population and mass balance models and lays out the approximations made to arrive at the widely used approximate models. It also quantifies the difference between the approximate and the accurate models.
- Chapter 8 lays out the framework to create ever-kinked crystal systems that avoid finite-size effects while studying crystal growth for a general triclinic crystallography.

# Bibliography

- [1] P. G. Vekilov, What determines the rate of growth of crystals from solution?, *Crystal Growth and Design* **7** (2007), no. 12 2796–2810.
- [2] W.-K. Burton, N. Cabrera, and F. Frank, The growth of crystals and the equilibrium structure of their surfaces, *Phil. Trans. R. Soc. Lond. A* **243** (1951), no. 866 299–358.
- [3] K. Chen and P. G. Vekilov, Evidence for the surface-diffusion mechanism of solution crystallization from molecular-level observations with ferritin, *Physical Review E* **66** (2002), no. 2 021606.
- [4] P. Dandekar, Z. B. Kuvadia, and M. F. Doherty, Engineering crystal morphology, *Annual Review of Materials Research* **43** (2013) 359–386.
- [5] J. Li, C. J. Tilbury, S. H. Kim, and M. F. Doherty, A design aid for crystal growth engineering, *Progress in Materials Science* **82** (2016) 1–38.
- [6] J. W. Gibbs, *Scientific Papers: Thermodynamics*, vol. 1. Dover Publications, 1961.
- [7] G. Wulff, Xxv. zur frage der geschwindigkeit des wachstums und der auflösung der krystallflächen, *Zeitschrift für Kristallographie-Crystalline Materials* **34** (1901), no. 1-6 449–530.
- [8] C. Herring, Some theorems on the free energies of crystal surfaces, *Physical review* **82** (1951), no. 1 87.
- [9] F. Silly, A. C. Powell, M. G. Martin, and M. R. Castell, Growth shapes of supported pd nanocrystals on srto 3 (001), *Physical Review B* **72** (2005), no. 16 165403.
- [10] F. Frank, On the kinematic theory of crystal growth and dissolution processes in growth and perfection of crystals (eds. rh doremus, bw roberts & d. turnbull) 411–417, 1958.
- [11] A. Chernov, The kinetics of the growth forms of crystals, *Sov. Phys. Cryst* **7** (1963) 728–730.
- [12] M. A. Lovette, A. R. Browning, D. W. Griffin, J. P. Sizemore, R. C. Snyder, and M. F. Doherty, Crystal shape engineering, *Industrial & engineering chemistry research* **47** (2008), no. 24 9812–9833.
- [13] F. Frank, The influence of dislocations on crystal growth, *Discussions of the*

## BIBLIOGRAPHY

---

- Faraday Society* **5** (1949) 48–54.
- [14] Y. Kuwahara, W. Liu, M. Makio, and K. Otsuka, In situ afm study of crystal growth on a barite (001) surface in baso4 solutions at 30° c, *Minerals* **6** (2016), no. 4 117.
- [15] C. J. Tilbury and M. F. Doherty, Modeling layered crystal growth at increasing supersaturation by connecting growth regimes, *AIChE Journal* **63** (2017), no. 4 1338–1352.
- [16] J. Li, C. J. Tilbury, M. N. Joswiak, B. Peters, and M. F. Doherty, Rate expressions for kink attachment and detachment during crystal growth, *Crystal Growth & Design* **16** (2016), no. 6 3313–3322.
- [17] N. E. Zimmermann, B. Vorselaars, D. Quigley, and B. Peters, Nucleation of NaCl from aqueous solution: Critical sizes, ion-attachment kinetics, and rates, *Journal of the American Chemical Society* **137** (2015), no. 41 13352–13361.
- [18] N. E. Zimmermann, B. Vorselaars, J. R. Espinosa, D. Quigley, W. R. Smith, E. Sanz, C. Vega, and B. Peters, NaCl nucleation from brine in seeded simulations: Sources of uncertainty in rate estimates, *The Journal of Chemical Physics* **148** (2018), no. 22 222838.
- [19] M. N. Joswiak, B. Peters, and M. F. Doherty, In silico crystal growth rate prediction for NaCl from aqueous solution, *Crystal Growth & Design* **18** (2018), no. 10 6302–6306.
- [20] A. M. Ferrenberg and R. H. Swendsen, New monte carlo technique for studying phase transitions, *Physical Review Letters* **61** (1988), no. 23 2635.
- [21] A. Z. Panagiotopoulos, Direct determination of phase coexistence properties of fluids by monte carlo simulation in a new ensemble, *Molecular Physics* **61** (1987), no. 4 813–826.
- [22] D. A. Kofke, Gibbs-duhem integration: a new method for direct evaluation of phase coexistence by molecular simulation, *Molecular Physics* **78** (1993), no. 6 1331–1336.
- [23] A. Benavides, J. Aragonés, and C. Vega, Consensus on the solubility of NaCl in water from computer simulations using the chemical potential route, *The Journal of Chemical Physics* **144** (2016), no. 12 124504.
- [24] J. Espinosa, J. Young, H. Jiang, D. Gupta, C. Vega, E. Sanz, P. Debenedetti, and A. Panagiotopoulos, On the calculation of solubilities via direct coexistence simulations: Investigation of NaCl aqueous solutions and lennard-jones binary mixtures, *The Journal of Chemical Physics* **145** (2016), no. 15 154111.
- [25] J. Kolafa, Free energy of classical molecular crystals by thermodynamic integration from a harmonic reference, *Journal of Chemical Theory and Computation* **15** (2018), no. 1 68–77.
- [26] T. B. Tan, A. J. Schultz, and D. A. Kofke, Suitability of umbrella-and overlap-sampling methods for calculation of solid-phase free energies by molecular

## BIBLIOGRAPHY

---

- simulation, *The Journal of Chemical Physics* **132** (2010), no. 21 214103.
- [27] A. Purohit, A. J. Schultz, and D. A. Kofke, Implementation of harmonically mapped averaging in lammmps, and effect of potential truncation on anharmonic properties, *The Journal of Chemical Physics* **152** (2020), no. 1 014107.
- [28] D. Frenkel and A. J. Ladd, New monte carlo method to compute the free energy of arbitrary solids. application to the fcc and hcp phases of hard spheres, *The Journal of Chemical Physics* **81** (1984), no. 7 3188–3193.
- [29] J. De Yoreo, L. Zepeda-Ruiz, R. Friddle, S. Qiu, L. Wasylenki, A. Chernov, G. Gilmer, and P. Dove, Rethinking classical crystal growth models through molecular scale insights: consequences of kink-limited kinetics, *Crystal Growth & Design* **9** (2009), no. 12 5135–5144.
- [30] J. Li, C. J. Tilbury, M. N. Joswiak, B. Peters, and M. F. Doherty, Rate expressions for kink attachment and detachment during crystal growth, *Crystal Growth & Design* **16** (2016), no. 6 3313–3322.
- [31] A. G. Stack, Molecular dynamics simulations of solvation and kink site formation at the {001} barite- water interface, *The Journal of Physical Chemistry C* **113** (2009), no. 6 2104–2110.
- [32] M. N. Joswiak, M. F. Doherty, and B. Peters, Critical length of a one-dimensional nucleus, *The Journal of Chemical Physics* **145** (2016), no. 21 211916.
- [33] M. De La Pierre, P. Raiteri, A. G. Stack, and J. D. Gale, Uncovering the atomistic mechanism for calcite step growth, *Angewandte Chemie International Edition* **56** (2017), no. 29 8464–8467.
- [34] J. D. Weeks and G. H. Gilmer, Dynamics of crystal growth, *Adv. Chem. Phys* **40** (1979), no. 489 157–227.
- [35] A. Chernov, Present-day understanding of crystal growth from aqueous solutions, *Progress in Crystal Growth and Characterization of Materials* **26** (1993) 121–151.
- [36] A. E. Nielsen, *Kinetics of precipitation*, vol. 18. Pergamon, 1964.
- [37] M. Salvalaglio, C. Perego, F. Giberti, M. Mazzotti, and M. Parrinello, Molecular-dynamics simulations of urea nucleation from aqueous solution, *Proceedings of the National Academy of Sciences* **112** (2015), no. 1 E6–E14.
- [38] J. Kolafa, Solubility of NaCl in water and its melting point by molecular dynamics in the slab geometry and a new bk3-compatible force field, *The Journal of Chemical Physics* **145** (2016), no. 20 204509.
- [39] M. N. Joswiak, M. F. Doherty, and B. Peters, Ion dissolution mechanism and kinetics at kink sites on NaCl surfaces, *Proceedings of the National Academy of Sciences* **115** (2018), no. 4 656–661.
- [40] L. Li, T. Totton, and D. Frenkel, Computational methodology for solubility prediction: Application to the sparingly soluble solutes, *The Journal of Chemical*

## BIBLIOGRAPHY

---

- Physics* **146** (2017), no. 21 214110.
- [41] M. A. Bellucci, G. Gobbo, T. K. Wijethunga, G. Ciccotti, and B. L. Trout, Solubility of paracetamol in ethanol by molecular dynamics using the extended Einstein crystal method and experiments, *The Journal of Chemical Physics* **150** (2019), no. 9 094107.
- [42] M. Mezei, P. Mehrotra, and D. Beveridge, Monte carlo determination of the free energy and internal energy of hydration for the ala dipeptide at 25. degree. c, *Journal of the American Chemical Society* **107** (1985), no. 8 2239–2245.
- [43] M. Cossi, V. Barone, R. Cammi, and J. Tomasi, Ab initio study of solvated molecules: a new implementation of the polarizable continuum model, *Chemical Physics Letters* **255** (1996), no. 4-6 327–335.
- [44] D. Shivakumar, J. Williams, Y. Wu, W. Damm, J. Shelley, and W. Sherman, Prediction of absolute solvation free energies using molecular dynamics free energy perturbation and the opls force field, *Journal of Chemical Theory and Computation* **6** (2010), no. 5 1509–1519.
- [45] D. L. Mobley and J. P. Guthrie, Freesolv: a database of experimental and calculated hydration free energies, with input files, *Journal of computer-aided molecular design* **28** (2014), no. 7 711–720.

# Chapter 3

## Chemical Potential in the Solid Phase

Reproduced in part with permission from:

Vikram Khanna, Jamshed Anwar, Daan Frenkel, Michael F. Doherty and Baron Peters, “Free energies of crystals computed using Einstein crystal with fixed center of mass and differing spring constants,” *Journal of Chemical Physics*, **2021**, 154(16), 164509:1-8. DOI: <https://doi.org/10.1063/5.0044833>. Copyright 2021 AIP Publishing.

Vikram Khanna, Michael F. Doherty and Baron Peters, “Absolute chemical potentials for complex molecules in fluid phases: A centroid reference for predicting phase equilibria,” *Journal of Chemical Physics*, **2020**, 153(21), 214504:1-9. DOI: <https://doi.org/10.1063/5.0025844>. Copyright 2020 AIP Publishing.

### 3.1 Introduction

Solid phase free energy computations<sup>[1-6]</sup> are widely used to predict fluid-solid equilibria, <sup>[7-11]</sup> solid-solid equilibria <sup>[12]</sup> and relative stability of polymorphs.<sup>[13, 14]</sup> The Frenkel-

Ladd method<sup>[3]</sup> computes the free energy difference between the solid under consideration and an Einstein crystal, a reference system whose free energy is analytically known. The method uses thermodynamic integration to transform between the Einstein crystal and real solid. To suppress a weak divergence of the integrand used in computing the free energy, it is recommended to fix the center of mass (COM) of the system.<sup>[3, 8, 15, 16]</sup> To compute the absolute free energy of the unconstrained crystal we need to correct for the effect of imposing this constraint in the calculations. The numerical value of the free energy correction per atom (or molecule) tends to zero as the system size tends to infinity, hence it is a finite-size correction. Polson et al.<sup>[15]</sup> derived an  $\mathcal{O}(\ln N/N)$  term in transformation from constrained to unconstrained crystals as a finite-size correction. Their calculation is based on an Einstein crystal with fixed center of mass and *equal* spring constants. Note that an alternate method, the Einstein molecule method,<sup>[4, 6, 17, 18]</sup> does not require a center of mass constraint or the associated corrections. However, this chapter is aimed at completing the theory for the Einstein crystal method and extending it to molecular solids accounting for intramolecular free energy—making the free energy an absolute free energy estimate in the truest sense.

For crystals comprised of atoms with different atomic masses, different spring constants (in the Einstein crystal approach<sup>[3]</sup>) can help to maintain the center of mass<sup>[8]</sup> constraint (if the molecular dynamics package does not have built-in features to achieve the same) by choosing mass-scaled spring constants that equate the angular frequencies. See supplementary material for details. Alternatively, different spring constants can help to optimize the numerical integration to compute the free energy difference by choosing spring constants that reproduce the mean-squared displacement (MSD) of atoms in the real crystal<sup>[16]</sup> (referred to as MSD-based springs in this article).

In this chapter, we extend the finite-size corrections of Polson et al.<sup>[15]</sup> to the case of *different* spring constants in an Einstein crystal. To test the results we compare the

solid free energies of LiI and NaCl crystals from three calculations: i) mass-scaled springs that exert a null force on the system and help constrain the COM without the need of built-in functions to constrain the COM, ii) MSD-based springs that reproduce the mean-squared displacement (MSD) of atoms in the real crystal that help optimize the numerical integration, and iii) calculations with a set of equal springs. A fourth option, not explored here, is to artificially set all masses and spring constants to equal values, respectively, and then analytically recover the free energies for the real masses, e.g. using equations in Polson et al. <sup>[15]</sup> For the MSD-based and equal spring sets we make use of the built-in center of mass constraint features in the molecular dynamics code LAMMPS<sup>[19]</sup>. These calculations confirm that the two procedures (employing different and equal springs) are equivalent. They also demonstrate the accuracy gained by the use of spring constants that reproduce the MSD of atoms in the real crystal along with the relative magnitudes of the finite-size corrections and other contributions to the absolute free energy of the crystal.

Note that in addition to the finite size correction due to the COM constraint in the calculations, the free energy of solids presents an intrinsic system size dependence as shown by Vega et al.<sup>[4]</sup> These finite size corrections need to be computed by repeating the free energy calculations for several system sizes and extrapolating to infinite size. This intrinsic finite size effect is also computed for the LiI and NaCl crystal systems to emphasize the difference between the two different types of finite size effects entering the solid free energy calculations., i.e., the one stemming due to the COM constraint in the Einstein crystal method and the other, the intrinsic finite size effect related to the cutoff in the phonon spectrum introduced by the finite lattice size.<sup>[20–24]</sup>

## 3.2 The Solid Free Energy Theory

The free energy ( $F$ ) of a real crystal (C) is computed using the following path:

$$F_{EC} \rightarrow F_{EC}^{CM} \rightarrow F_C^{CM} \rightarrow F_C$$

where CM indicates a center of mass constraint and EC refers to an Einstein crystal.

Therefore,

$$F_C = F_{EC} + (F_{EC}^{CM} - F_{EC}) + (F_C^{CM} - F_{EC}^{CM}) + (F_C - F_C^{CM}) \quad (3.1)$$

where the absolute free energy of the Einstein crystal can be obtained from its analytically computable partition function:

$$\beta F_{EC} = -\ln(Q_{EC}) \quad (3.2)$$

and the free energy differences are

$$\beta(F_{EC}^{CM} - F_{EC}) = -\ln\left(\frac{Q_{EC}^{CM}}{Q_{EC}}\right) \quad (3.3a)$$

$$\beta(F_C^{CM} - F_{EC}^{CM}) = \beta \int_{\lambda=0}^{\lambda=1} \left\langle \frac{dU}{d\lambda} \right\rangle_{\lambda}^{CM} d\lambda \quad (3.3b)$$

$$\beta(F_C - F_C^{CM}) = -\ln\left(\frac{Q_C}{Q_C^{CM}}\right) \quad (3.3c)$$

In each of these formulas,  $Q$  is a partition function,  $\beta = (k_B T)^{-1}$ ,  $\lambda$  is a coupling parameter,  $U$  is a  $\lambda$ -dependent potential energy function that interpolates between that of the EC and C systems.

### 3.2.1 Spring Constants and Finite Size Corrections for Einstein Crystals

When all atoms have a common spring constant ( $k$ ), the Helmholtz free energy of a crystal with  $N_{mol}$  molecules composed of a total of  $N$  atoms, is given by Polson et al.<sup>[15]</sup> as:

$$\begin{aligned} \beta F_C = & \sum_{i=1}^N \ln \left( \frac{\beta k \Lambda_i^2}{2\pi} \right)^{3/2} \\ & + \beta \int_{\lambda=0}^{\lambda=1} \left\langle \frac{dU}{d\lambda} \right\rangle_{\lambda}^{CM} d\lambda \\ & - \ln \left[ \left( \frac{\beta k}{2\pi \sum_{i=1}^N \mu_i^2} \right)^{3/2} \left( \frac{V}{N_{mol}} \right) \right] \end{aligned} \quad (3.4)$$

where  $\Lambda_i = (\beta h^2 / (2\pi m_i))^{1/2}$ ,  $h$  is Planck's constant,  $m_i$  is the mass of atom  $i$ ,  $V$  is the volume of the system, and  $\mu_i = m_i / \sum_{i=1}^N m_i$ .

We have combined terms in the result as given in Polson et al. so that all arguments of logarithms are dimensionless.

Note that the corresponding equations in ref. [8], for atoms with different spring constants cannot be combined to give a dimensionless argument to the logarithm. The error introduced in that study, however, was marginal as the spring constants were similar. Using our revised result shown in equation (3.10), the free energy in Ref. [8] is revised to  $-97.44 \pm 0.02 N_{mol} k_B T$  from  $-97.75 \pm 0.02 N_{mol} k_B T$ . This revision results in a shift of the predicted melting point from 1064 K to about 1140 K, some 65 K above the experimental melting of NaCl (1074K). An independent study carried out subsequent to Ref. [8] employing density of states calculations for the same NaCl model gave a melting point of 1050 K.<sup>[25]</sup> In this work we provide a revised version of equation (3.4) for systems

where atoms have different spring constants.

Since the part of the finite-size correction involving spring constants emerges from equation (A.3a), (see Appendix A), we derive the configurational partition function ratio,  $Z_{EC}^{CM}/Z_{EC}$  here for an EC with different spring constants ( $k_i$ ,  $i = 1, 2, \dots, N$ ).

For 1-Dimension (x-direction), the configurational partition function of an Einstein crystal with a COM constraint is:

$$Z_{EC,x}^{CM} = \int dx^N \prod_{i=1}^N \exp \left[ -\frac{\beta k_i}{2} x_i^2 \right] \delta \left( \sum_{i=1}^N \mu_i x_i \right) \quad (3.5)$$

Without loss of generality, we are letting all particles be attached by a spring to a point at the origin. This makes  $x_i$  the displacement of atom  $i$  in the x-direction from the tether point (origin here).

We make the following variable transformation:

$$\xi_i = k_i^{1/2} x_i \quad (3.6)$$

with Jacobian determinant

$$J = \left| \prod_{i=1}^N k_i^{-1/2} \right| \quad (3.7)$$

Therefore,

$$\begin{aligned}
Z_{EC,x}^{CM} &= \int d\xi^N J \prod_{i=1}^N \exp \left[ -\frac{\beta}{2} \xi_i^2 \right] \delta \left( \sum_{i=1}^N \frac{\mu_i}{k_i^{1/2}} \xi_i \right) \\
&= J \left( \frac{\beta}{2\pi \sum_{i=1}^N \frac{\mu_i^2}{k_i}} \right)^{1/2} \prod_{i=1}^N \left( \frac{2\pi}{\beta} \right)^{1/2} \\
&= \left( \frac{\beta}{2\pi \sum_{i=1}^N \frac{\mu_i^2}{k_i}} \right)^{1/2} \prod_{i=1}^N \left( \frac{2\pi}{\beta k_i} \right)^{1/2} \\
&= \left( \frac{\beta}{2\pi \sum_{i=1}^N \frac{\mu_i^2}{k_i}} \right)^{1/2} Z_{EC,x}
\end{aligned} \tag{3.8}$$

where  $Z_{EC,x}$  is the integral in equation (3.5), but without the center of mass constraint. See supplementary material for additional details. The extension to three dimensions just results in three factors of  $(\beta/2\pi \sum_i [\mu_i^2/k_i])^{1/2}$ .

$$\frac{Z_{EC}^{CM}}{Z_{EC}} = \left( \frac{\beta}{2\pi \sum_{i=1}^N \frac{\mu_i^2}{k_i}} \right)^{3/2} \tag{3.9}$$

Using equation (3.9) in place of equation (A.10a) in the derivation shown in Appendix A, we get

$$\begin{aligned}
\beta F_C &= \sum_{i=1}^N \ln \left( \frac{\beta k_i \Lambda_i^2}{2\pi} \right)^{3/2} \\
&+ \beta \int_{\lambda=0}^{\lambda=1} \left\langle \frac{dU}{d\lambda} \right\rangle_{\lambda}^{CM} d\lambda \\
&- \ln \left[ \left( \frac{\beta}{2\pi \sum_{i=1}^N \frac{\mu_i^2}{k_i}} \right)^{3/2} \left( \frac{V}{N_{mol}} \right) \right]
\end{aligned} \tag{3.10}$$

for an Einstein crystal system with different spring constants.

Equation (3.10) gives the free energy of a real crystal computed using an EC system constituting different spring constants. Note that equation (3.10) is properly dimensionless. Moreover, equation (3.10) collapses to equation (3.4) when  $k_i = k$ , i.e., when all spring constants are equal.

### 3.2.2 Constraining the center of mass

In molecular dynamics (MD) simulations with standard force fields, the center of mass can be fixed by beginning with zero total momentum and not adding external forces. For the Einstein crystal, the springs do exert external forces so additional measures are needed.

A simple way to fix the center of mass in a system with springs begins by making the spring constants proportional to the atom masses, so that all atoms have the same frequency. When all oscillators have frequency  $\omega$ , the center of mass evolves as

$$\mathbf{x}_{CM}(t) = \omega^{-1} \mathbf{v}_{CM}^o \sin(\omega t) + \mathbf{x}_{CM}^o \cos(\omega t) \quad (3.11)$$

where,  $\mathbf{v}_{CM}^o$  is the initial center of mass velocity, and  $\mathbf{x}_{CM}^o$  is the initial displacement of the center of mass from the tether points' (lattice positions') center of mass. See Appendix B for details.

Clearly, we can maintain  $\mathbf{x}_{CM} = 0$  by choosing spring constants proportional to the atom masses, so that all atoms have the same frequency and setting initial conditions such that  $\mathbf{v}_{CM}^o = \mathbf{x}_{CM}^o = 0$ .

We note that certain MD packages such as LAMMPS do not need to have a net zero external force on the system to constrain the center of mass. This is achieved by thermostating all degrees of freedom except the center of mass and shifting all atom coordinates after every timestep (equal to the drift in center of mass) to recenter the

system to the initial value of the center of mass. This recentering does not alter the dynamics of the system or change the relative coordinates of any pair of atoms.

### 3.2.3 Thermodynamic Integration

To compute the free energy difference between the crystal of interest and the Einstein crystal, i.e.,  $F_C^{CM} - F_{EC}^{CM}$ , we use thermodynamic integration (T.I.)<sup>[26]</sup> with a linear homotopy<sup>[27]</sup>

$$U(T, V_C; \lambda) = (1 - \lambda)U_{EC}(T, V_C) + \lambda U_C(T, V_C) \quad (3.12)$$

where  $U(T, V_C; \lambda)$  is the potential energy function and  $\lambda$  is the coupling parameter. Also,  $U_{EC}$  and  $U_C$  are the potential energy functions of the Einstein crystal and the crystal of interest (described by the chosen force field), respectively. Therefore

$$\begin{aligned} F_C^{CM} - F_{EC}^{CM} &= \int_{\lambda=0}^{\lambda=1} \left\langle \frac{dU}{d\lambda} \right\rangle_{\lambda}^{CM} d\lambda \\ &= \int_{\lambda=0}^{\lambda=1} \langle U_C - U_{EC} \rangle_{\lambda}^{CM} d\lambda \end{aligned} \quad (3.13)$$

where,  $\langle dU/d\lambda \rangle_{\lambda}^{CM}$  is evaluated by computing an average of  $(U_C - U_{EC})$  over configurations in the  $\lambda$  state canonical ensemble. Note, for  $0 < \lambda < 1$ , equation (3.11) is still applicable. See supplementary material for details.

## 3.3 Simulation Details

We model LiI and NaCl using the Joung Cheatham force field (the version optimized for SPC/E).<sup>[28]</sup> All MD simulations were carried out using LAMMPS.<sup>[19]</sup>

The NVT simulations for computing the free energies were setup using the interionic distances reported by Joung et al.,<sup>[28]</sup> of 3.05 Å, and 2.89 Å for LiI and NaCl, respectively.

A super cell measuring  $36.6 \text{ \AA} \times 36.6 \text{ \AA} \times 36.6 \text{ \AA}$  comprising 864 ion pairs was used for LiI, and a super cell measuring  $34.68 \text{ \AA} \times 34.68 \text{ \AA} \times 34.68 \text{ \AA}$  comprising 864 ion pairs was used for NaCl. This is equivalent to  $6 \times 6 \times 6$  unit cells for both crystal systems. To compute the intrinsic finite-size effects, systems comprising  $7 \times 7 \times 7$ ,  $8 \times 8 \times 8$ , and  $9 \times 9 \times 9$  unit cells were used for both crystals.

The Lorentz-Berthelot mixing rules were used for computing the interatomic pair coefficients. We used a time step of 1 fs. Nonbonded interactions were cutoff at 1.6 nm, and 1.1 nm for LiI and NaCl, respectively, with long range electrostatics handled by LAMMPS' Particle-Particle-Particle-Mesh (PPPM) summation<sup>[29, 30]</sup> and a switching function applied for Lennard-Jones interactions between 1.4-1.6 nm, and 0.9-1.1 nm for LiI and NaCl, respectively.

A 25 point Gauss Legendre quadrature method<sup>[31]</sup> was used to evaluate the integral in equation 3.13. The NVT simulations were run for a total of 8 ns, of which the initial 2 ns were used for equilibration and then discarded. Data every 1 ps was used to compute the thermodynamic averages for LiI and NaCl systems (the autocorrelation times for  $\langle dU/d\lambda \rangle_\lambda$  were approximately in the range of 50 - 400 fs, therefore a 1 ps sampling frequency provides independent samples for each  $\lambda$  state. See supplementary material for autocorrelation time calculations). The spring constants for the reference Einstein crystals are chosen to reproduce the mean square displacement of lithium and sodium ions at 300K. For the different springs case, respective spring constants are scaled to compute the counter ion's spring constant such that  $k_i/m_i$  is constant. Table 3.1 reports the numerical values used in this study.

Table 3.1: Masses (a.m.u) & spring constants ( $k_B T / \text{\AA}^2$ )

	<i>LiI</i>		<i>NaCl</i>	
	$\text{Li}^+$	$\text{I}^-$	$\text{Na}^+$	$\text{Cl}^-$
mass	6.941	126.904	22.990	35.450
$k_i$ using mass-scaled springs	35.866	655.749	77.543	119.581
$k_i$ using MSD-based springs	35.866	67.121	77.543	85.167
$k_i$ using equal springs	35.866	35.866	77.543	77.543

For the MSD-based and equal springs cases, to constrain the center of mass in LAMMPS, we use the *temp/com* command in conjunction with the langevin thermostat to thermostat all degrees of freedom except the center of mass. We use the *recenter* command to correct for the drift in the center of mass after every time step.

### 3.4 Results and Discussion

For each crystal system (i.e., LiI and NaCl) we compute free energies using mass-scaled, MSD-based and equal spring constants. Table 3.2 shows the dimensionless Helmholtz free energies per ion pair. All components of equation 3.10 are reported to show the relative magnitudes of corrections. See supplementary material for plots of the T.I. integrands. The free energies should match for the two procedures, i.e., with different and equivalent springs. As seen in Table 3.2 the free energies computed using different spring constants are in excellent agreement with the free energies computed using equivalent spring constants, thereby successfully testing the theory. Also, the use of MSD-based spring constants optimizes the calculations as can be seen in the near 24% uncertainty reduction for LiI when compared to the equal spring case. This is because the MSD of

the Li and I atoms differ significantly as can be seen by the MSD-based spring constants. In the NaCl calculations, the MSD of the two atoms are similar, leading to marginal accuracy gains on using MSD-based springs. We also compute the free energy of NaCl crystal at 298 K to compare with Aragonés et al.<sup>[6]</sup> Our free energy estimate at 298 K is  $-311.0340 \pm 0.0002 N_{\text{mol}}k_{\text{B}}T$  (using  $\Lambda_{\text{Na}} = \Lambda_{\text{Cl}} = 1\text{\AA}$ ), which is in excellent agreement with their result of  $-311.10 \pm 0.1 N_{\text{mol}}k_{\text{B}}T$ .

As seen in Figures 3.1 and 3.2, the solid free energies of LiI and NaCl demonstrate a significant system size dependence, i.e. the free energy size dependence is greater than the calculation uncertainty. Tables 3.3 and 3.4 show the accounting of the COM constraint finite size correction for each system size for LiI and NaCl, respectively.

We use a linear model to capture the intrinsic finite size effect and predict free energy in the thermodynamic limit:<sup>[4]</sup>

$$\beta f(N_{\text{mol}}) = \beta f(N_{\text{mol}} \rightarrow \infty) + \frac{d_1}{N_{\text{mol}}} \quad (3.14)$$

The value of the slopes ( $d_1$ ) for LiI and NaCl are  $-14.7 \pm 0.2$ , and  $-13.1 \pm 0.2$ , respectively.

Table 3.2: Absolute dimensionless Helmholtz free energy ( $\beta f = \beta F/N_{mol}$ ) calculation of LiI and NaCl (per ion pair) at 300 K. See supplementary material for details of the column headings shown in this table.

Crystal system	$\beta f_{EC}$	$\beta(f_{EC}^{CM} - f_{EC})^*$	$\beta \Delta f_{EC \rightarrow C}^{CM}$	$\beta(f_C - f_C^{CM})^*$	$\beta f_C$
LiI mass-scaled springs	-0.53908	-0.01990	-307.42991 $\pm$ 0.00028	-0.00467	-307.99356 $\pm$ 0.00028
LiI MSD-based springs	-3.95800	-0.01603	-304.01438 $\pm$ 0.00016	-0.00467	-307.99308 $\pm$ 0.00016
LiI equiv. springs	-4.89807	-0.01494	-303.07519 $\pm$ 0.00021	-0.00467	-307.99286 $\pm$ 0.00021
NaCl mass-scaled springs	-1.81876	-0.01772	-317.07633 $\pm$ 0.00015	-0.00449	-318.91719 $\pm$ 0.00015
NaCl MSD-based springs	-2.32772	-0.01734	-316.56798 $\pm$ 0.00015	-0.00449	-318.91753 $\pm$ 0.00015
NaCl equiv. springs	-2.46850	-0.01723	-316.42744 $\pm$ 0.00016	-0.00449	-318.91755 $\pm$ 0.00016

Table 3.3: Absolute dimensionless Helmholtz free energy ( $\beta f = \beta F/N_{mol}$ ) calculation of LiI (per ion pair) at 300 K for varying system size using MSD-based springs. See supplementary material for details of the column headings shown in this table.

$N_{mol}$	$\beta f_{EC}$	$\beta(f_{EC}^{CM} - f_{EC})^*$	$\beta\Delta f_{EC \rightarrow C}^{CM}$	$\beta(f_C - f_C^{CM})^*$	$\beta f_C$
864	-3.95800	-0.01603	-304.01438 $\pm$ 0.00016	-0.00467	-307.99308 $\pm$ 0.00016
1372	-3.95800	-0.01060	-304.01569 $\pm$ 0.00013	-0.00294	-307.98723 $\pm$ 0.00013
2048	-3.95800	-0.00739	-304.01515 $\pm$ 0.00010	-0.00197	-307.98251 $\pm$ 0.00010
2916	-3.95800	-0.00537	-304.01679 $\pm$ 0.00009	-0.00138	-307.98154 $\pm$ 0.00009
$N_{mol} \rightarrow \infty$	-3.95800	0.00000	-304.01730 $\pm$ 0.00021	0.00000	-307.97607 $\pm$ 0.00014

Table 3.4: Absolute dimensionless Helmholtz free energy ( $\beta f = \beta F/N_{mol}$ ) calculation of NaCl (per ion pair) at 300 K for varying system size computed using MSD-based springs. See supplementary material for details of the column headings shown in this table.

$N_{mol}$	$\beta f_{EC}$	$\beta(f_{EC}^{CM} - f_{EC})^*$	$\beta\Delta f_{EC \rightarrow C}^{CM}$	$\beta(f_C - f_C^{CM})^*$	$\beta f_C$
864	-2.32772	-0.01734	-316.56798 $\pm$ 0.00015	-0.00449	-318.91753 $\pm$ 0.00015
1372	-2.32772	-0.01143	-316.56964 $\pm$ 0.00012	-0.00283	-318.91162 $\pm$ 0.00012
2048	-2.32772	-0.00795	-316.57104 $\pm$ 0.00010	-0.00189	-318.90860 $\pm$ 0.00010
2916	-2.32772	-0.00576	-316.57206 $\pm$ 0.00008	-0.00133	-318.90687 $\pm$ 0.00008
$N_{mol} \rightarrow \infty$	-2.32772	0.00000	-316.57449 $\pm$ 0.00013	0.00000	-318.90223 $\pm$ 0.00013

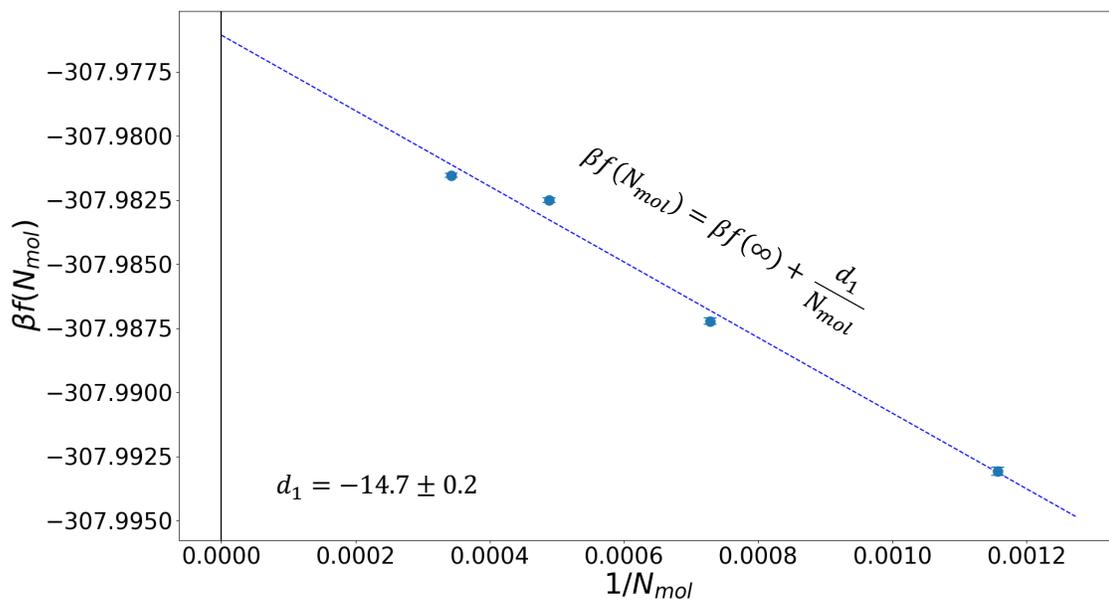


Figure 3.1: Intrinsic system-size dependence of LiI crystal's free energy

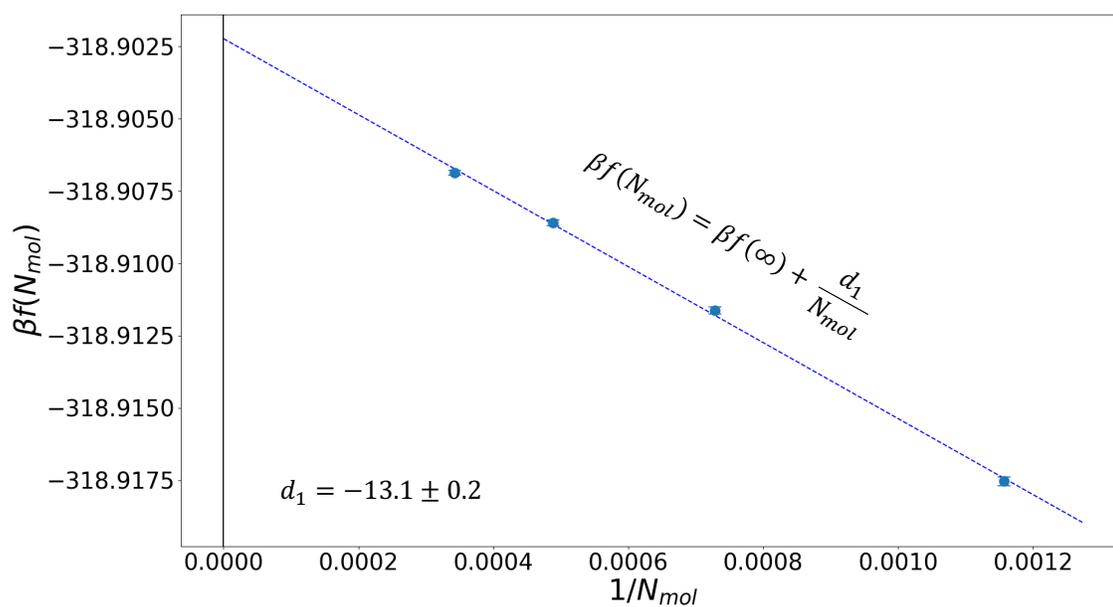


Figure 3.2: Intrinsic system-size dependence of NaCl crystal's free energy

### 3.5 Extension to Molecular Crystals

We apply the Einstein crystal approach for polyatomic molecules differently from previous works<sup>[10, 11, 32]</sup> to obtain an absolute free energy prediction. Li et al.<sup>[10]</sup> and Bellucci et al.<sup>[11]</sup> tether non-collinear atoms in a molecule to their lattice sites and keep the intra-molecular bonded and non-bonded interactions turned on. We tether each atom in the crystal to its lattice site with all inter as well as intra-molecular bonded and non-bonded interactions turned off. This gives a set of simple harmonic oscillators as shown in Figure 3.3. Thus the free energy difference between the EC and the real crystal system includes the intra-molecular and inter-molecular free energies.

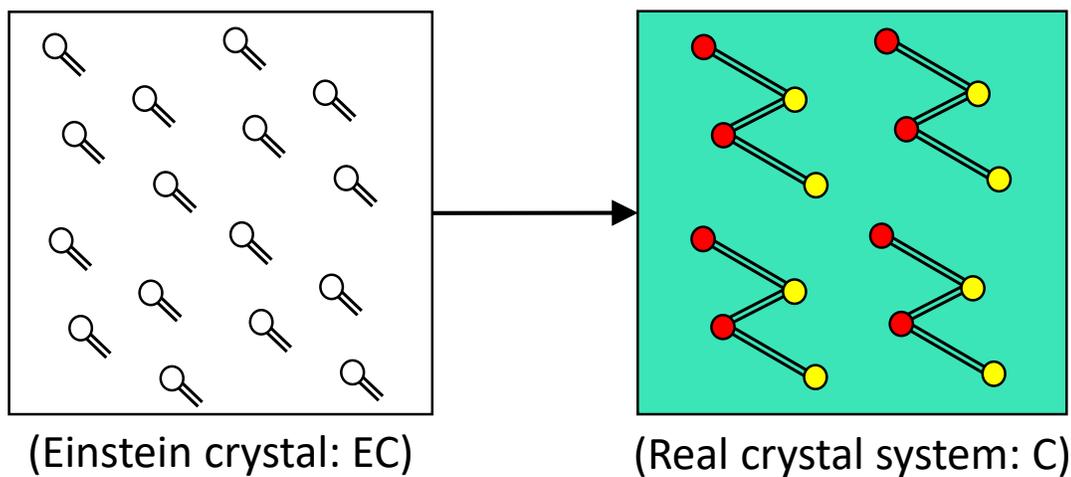


Figure 3.3: Transformation from an all atom Einstein crystal to a real crystal with complete intra-molecular interactions and nonbonded intra and inter-molecular interactions (green). The double lines indicate springs. Between two atoms, they represent a bond, when attached to a single atom they represent tethering to a lattice site.

To compute  $\Delta a^{EC \rightarrow C}(T, v^C)$  we use thermodynamic integration (T.I)<sup>[26]</sup> with a linear homotopy.<sup>[27]</sup>

## 3.6 Conclusions

In this chapter we extend the work of Polson et al.<sup>[15]</sup> to compute free energies of crystals using an Einstein crystal with fixed center of mass and *different* spring constants. For calculations that make use of different spring constants, our result provides a revised equation to be used in place of equation (3) in ref. [8]. We also provide necessary conditions to help constrain the center of mass in a molecular dynamics simulations (using MD packages with no built-in capabilities to do the same) employing springs to tether atoms. The free energies of LiI and NaCl crystals computed using *different* and equal spring constants are in excellent agreement, thus successfully testing the theory. Our results also demonstrate a reduction in error bars when MSD-based springs are used to optimize the calculations. The significant system size dependence of the solid free energy after applying the COM-constraint corrections helps demonstrate the relative magnitudes of the intrinsic and COM-constraint finite-size corrections relative to other contributions to the absolute free energy of the crystals.

# Bibliography

- [1] W. G. Hoover and F. H. Ree, Use of computer experiments to locate the melting transition and calculate the entropy in the solid phase, *J. Chem. Phys.* **47** (1967), no. 12 4873–4878.
- [2] W. G. Hoover and F. H. Ree, Melting transition and communal entropy for hard spheres, *J. Chem. Phys.* **49** (1968), no. 8 3609–3617.
- [3] D. Frenkel and A. J. Ladd, New monte carlo method to compute the free energy of arbitrary solids. application to the fcc and hcp phases of hard spheres, *J. Chem. Phys.* **81** (1984), no. 7 3188–3193.
- [4] C. Vega and E. G. Noya, Revisiting the frenkel-ladd method to compute the free energy of solids: The Einstein molecule approach, *J. Chem. Phys.* **127** (2007), no. 15 154113.
- [5] E. G. Noya, M. Conde, and C. Vega, Computing the free energy of molecular solids by the Einstein molecule approach: Ices xiii and xiv, hard-dumbbells and a patchy model of proteins, *J. Chem. Phys.* **129** (2008), no. 10 104704.
- [6] J. Aragonés, C. Valeriani, and C. Vega, Note: Free energy calculations for atomic solids through the Einstein crystal/molecule methodology using gromacs and lammmps, *J. Chem. Phys.* **137** (2012), no. 14 146101.
- [7] J. M. Polson and D. Frenkel, Calculation of solid-fluid phase equilibria for systems of chain molecules, *J. Chem. Phys.* **109** (1998), no. 1 318–328.
- [8] J. Anwar, D. Frenkel, and M. G. Noro, Calculation of the melting point of NaCl by molecular simulation, *J. Chem. Phys.* **118** (2003), no. 2 728–735.
- [9] A. Benavides, J. Aragonés, and C. Vega, Consensus on the solubility of NaCl in water from computer simulations using the chemical potential route, *J. Chem. Phys.* **144** (2016), no. 12 124504.
- [10] L. Li, T. Totton, and D. Frenkel, Computational methodology for solubility prediction: Application to the sparingly soluble solutes, *J. Chem. Phys.* **146** (2017), no. 21 214110.
- [11] M. A. Bellucci, G. Gobbo, T. K. Wijethunga, G. Ciccotti, and B. L. Trout, Solubility of paracetamol in ethanol by molecular dynamics using the extended Einstein crystal method and experiments, *J. Chem. Phys.* **150** (2019), no. 9 094107.

## BIBLIOGRAPHY

---

- [12] J. Anwar, C. Leitold, and B. Peters, Solid–solid phase equilibria in the NaCl–KCl system, *J. Chem. Phys.* **152** (2020), no. 14 144109.
- [13] E. C. Dybeck, N. P. Schieber, and M. R. Shirts, Effects of a more accurate polarizable hamiltonian on polymorph free energies computed efficiently by reweighting point-charge potentials, *J. Chem. Theory Comput.* **12** (2016), no. 8 3491–3505.
- [14] M. Yang, E. Dybeck, G. Sun, C. Peng, B. Samas, V. M. Burger, Q. Zeng, Y. Jin, M. A. Bellucci, Y. Liu, *et. al.*, Prediction of the relative free energies of drug polymorphs above zero kelvin, *Cryst. Growth Des.* **20** (2020), no. 8 5211–5224.
- [15] J. M. Polson, E. Trizac, S. Pronk, and D. Frenkel, Finite-size corrections to the free energies of crystalline solids, *J. Chem. Phys.* **112** (2000), no. 12 5339–5342.
- [16] D. Frenkel and B. Smit, *Understanding molecular simulation: from algorithms to applications*, vol. 1. Elsevier, 2001.
- [17] J. Aragoes, E. G. Noya, C. Valeriani, and C. Vega, Free energy calculations for molecular solids using gromacs, *J. Chem. Phys.* **139** (2013), no. 3 034104.
- [18] R. K. Reddy A and S. N. Punnathanam, Calculation of excess free energy of molecular solids comprised of flexible molecules using einstein molecule method, *Molecular Simulation* **44** (2018), no. 10 781–788.
- [19] S. Plimpton, Fast parallel algorithms for short-range molecular dynamics, *J. Comput. Phys.* **117** (1995), no. 1 1–19.
- [20] W. G. Hoover, Entropy for small classical crystals, *The Journal of Chemical Physics* **49** (1968), no. 4 1981–1982.
- [21] W. G. Hoover, M. Ross, K. W. Johnson, D. Henderson, J. A. Barker, and B. C. Brown, Soft-sphere equation of state, *J. Chem. Phys.* **52** (1970), no. 10 4931–4941.
- [22] W. G. Hoover, S. G. Gray, and K. W. Johnson, Thermodynamic properties of the fluid and solid phases for inverse power potentials, *J. Chem. Phys.* **55** (1971), no. 3 1128–1136.
- [23] W. G. Hoover, A. C. Hindmarsh, and B. L. Holian, Number Dependence of Small-Crystal Thermodynamic Properties. I, *J. Chem. Phys.* **57** (1972), no. 5 1980–1985.
- [24] R. Freitas, M. Asta, and M. De Koning, Nonequilibrium free-energy calculation of solids using LAMMPS, *Computational Materials Science* **112** (2016) 333–341.
- [25] E. A. Mastny and J. J. de Pablo, Direct calculation of solid-liquid equilibria from density-of-states monte carlo simulations, *J. Chem. Phys.* **122** (2005), no. 12 124109.
- [26] J. G. Kirkwood, Statistical mechanics of fluid mixtures, *J. Chem. Phys.* **3** (1935), no. 5 300–313.
- [27] M. Mezei and D. Beveridge, Free energy simulations, *Ann. New York Acad. Sci.*

## BIBLIOGRAPHY

---

- 482** (1986), no. 1 1–23.
- [28] I. S. Joung and T. E. Cheatham III, Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations, *J. Phys. Chem. B* **112** (2008), no. 30 9020–9041.
- [29] P. P. Ewald *Ann. Phys.* **369** (1921), no. 3 253–287.
- [30] R. W. Hockney and J. W. Eastwood, *Computer simulation using particles*. Adam Hilger, NY, 1989.
- [31] M. Abramowitz and I. A. Stegun, *Handbook of mathematical functions: with formulas, graphs, and mathematical tables*, vol. 55. Courier Corporation, 1965.
- [32] D. R. M. Guilherme and D. L. Mobley, Challenges in the use of atomistic simulations to predict solubilities of drug-like molecules, *F1000Research* **7** (2018).

# Chapter 4

## Chemical Potential in the Gas Phase

Reproduced in part with permission from:

Vikram Khanna, Michael F. Doherty and Baron Peters, “Absolute chemical potentials for complex molecules in fluid phases: A centroid reference for predicting phase equilibria,” *Journal of Chemical Physics*, **2020**, 153(21), 214504:1-9.

DOI: <https://doi.org/10.1063/5.0025844>. Copyright 2020 AIP Publishing.

### 4.1 Introduction

Absolute free energy calculations require a reference system whose free energy is known analytically or could be computed via simulation(s). There are many analytically tractable reference systems for computing absolute free energies of molecules in the gas phase: ideal gas atoms, rigid molecules, harmonic approximations, etc.<sup>[1-4]</sup> Generally, as the reference system becomes simpler, the transformation from the reference model to the detailed molecular system becomes more difficult. For example, the ideal gas atom reference system is trivial, but it has fewer degrees of freedom than a gas of molecules. One could add one gas phase atom per atom in the molecular system, but the atoms and

the molecules will still have different numbers of vibrational, rotational, and translational degrees of freedom. These differences would lead to poor phase space overlap between the two systems,<sup>[5]</sup> and therefore to a difficult free energy calculation. The harmonic approximation works very well for small molecules with stiff vibrations, but not for large molecules with floppy structures and multiple conformational states.<sup>[6]</sup>

We propose a simple and convenient reference that permits accurate numerical absolute partition functions. In the reference entity, all atoms which make up the molecule are attached to their collective center of mass via springs as shown in Figure 4.1. We call this system the “centroid” (represented by the symbol \*) after work on quantum nuclear partition functions.<sup>[7–10]</sup> Tethering all atoms to the center of mass makes them a “single molecular entity” with exactly the same number of vibrations, rotations and translational degrees of freedom as the real molecule.

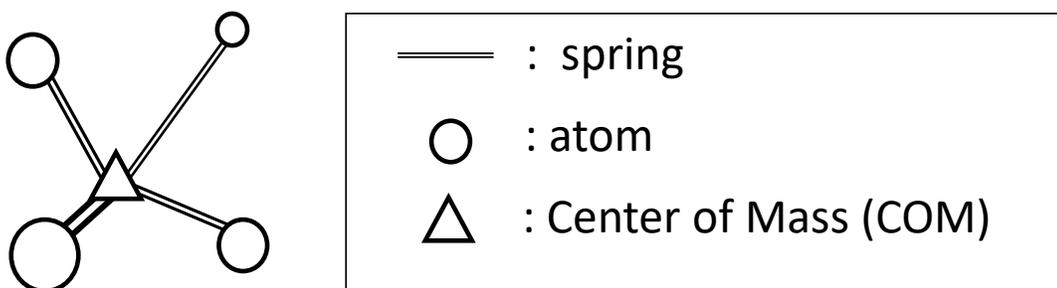


Figure 4.1: The Centroid

To compute the absolute free energy of an isolated molecule, we first compute the free energy of a centroid made up of the same atoms which constitute the molecule.

## 4.2 Computing the Free Energy of the Reference Centroid System

The partition function of an  $N$  atom centroid system is

$$\begin{aligned}
 Q_N &= \frac{1}{\prod_{i=1}^N \Lambda_i^3} \int \prod_{i=1}^N d\mathbf{r}_i e^{-\frac{\beta}{2} \sum_{i=1}^N k_i \|\mathbf{r}_i - \mathbf{r}_{COM}^{(N)}\|^2} \\
 &= \frac{1}{\prod_{i=1}^N \Lambda_i^3} \int \prod_{i=1}^N d\mathbf{r}_i e^{-\frac{\beta}{2} U_N(\mathbf{r}_i; \mathbf{r}_{COM}^{(N)})} \\
 &= \frac{Z_N}{\prod_{i=1}^N \Lambda_i^3}
 \end{aligned} \tag{4.1}$$

where,  $k_i$  is the spring constant between the atom  $i$  and the center of mass,  $\mathbf{r}_i$  and  $\mathbf{r}_{COM}^{(N)}$ , are the coordinates of atom  $i$  and the center of mass of the  $N$  atom system, respectively and  $\Lambda_i$  is the thermal de Broglie wavelength of atom  $i$ .

To compute  $Q_N$  and  $Z_N$ , we start with the analytically tractable centroid partition function for one (or two) atoms. We then recursively build larger centroids and their partition functions by computing the ratios  $Q_n/Q_{n-1}$ .

$$Q_N = Q_1 \prod_{n=2}^N \frac{Q_n}{Q_{n-1}} \tag{4.2}$$

Our task is to compute the ratio  $Q_n/Q_{n-1}$ . The ratios are easily computed,

$$\frac{Q_n}{Q_{n-1}} = \frac{1}{\Lambda_n^3} \frac{Z_n}{Z_{n-1}} = \frac{1}{\Lambda_n^3} \frac{Z_n}{Z_n^o} \frac{Z_n^o}{Z_{n-1}} \tag{4.3}$$

where:

$$\begin{aligned} Z_n^o &= \int \prod_{i=1}^n d\mathbf{r}_i e^{-\frac{\beta}{2} \sum_{i=1}^n k_i \|\mathbf{r}_i - \mathbf{r}_{COM}^{(n-1)}\|^2} \\ &= \int \prod_{i=1}^n d\mathbf{r}_i e^{-\frac{\beta}{2} U_n^o(\mathbf{r}_i; \mathbf{r}_{COM}^{(n-1)})} \end{aligned} \quad (4.4)$$

$Z_n^o$  is the configurational partition function of a  $n$  atom system (green+yellow) where all atoms are tethered to the COM of the  $n - 1$  atom system (yellow).

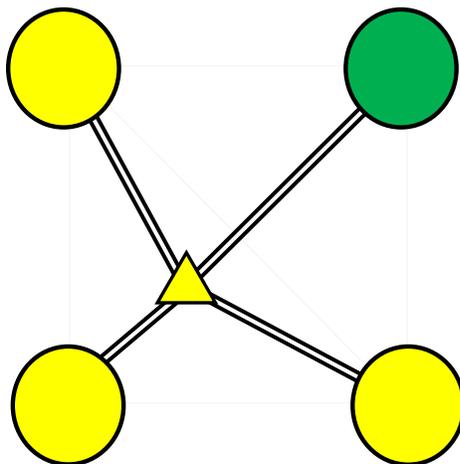


Figure 4.2: The  $n$  atom intermediate system used in computing  $Q_n/Q_{n-1}$  whose partition function is given by  $Z_n^o$ ; here  $n = 4$

We compute the ratio  $Z_n/Z_n^o$  using thermodynamic perturbation i.e., (see S.I. for details)

$$\frac{Z_n}{Z_n^o} = \langle e^{-\beta \Delta U_n} \rangle_{U_n^o} \quad (4.5)$$

where,  $\Delta U_n = U_n - U_n^o$  and,

$$\frac{Z_n^o}{Z_{n-1}^o} = \left( \frac{2\pi}{\beta k_n} \right)^{\frac{3}{2}} \quad (4.6)$$

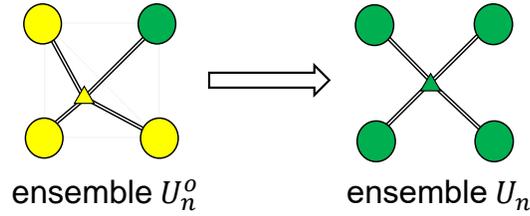


Figure 4.3: The thermodynamic perturbation from  $Z_n^o$  to  $Z_n$ ; here  $n = 4$

Therefore, using equations 4.5 and 4.6 in equation 4.3, we get:

$$\frac{Q_n}{Q_{n-1}} = \frac{1}{\Lambda_n^3} \frac{Z_n}{Z_{n-1}} = \frac{1}{\Lambda_n^3} \langle e^{-\beta \Delta U_n} \rangle_{U_n^o} \left( \frac{2\pi}{\beta k_n} \right)^{\frac{3}{2}} \quad (4.7)$$

Adding particles recursively gives the partition function of the centroid, and hence the free energy

$$\begin{aligned} \beta a^* &= -\ln(Q_N) \\ &= -\ln \left[ \left( \frac{v^G}{\Lambda_1^3} \right) \prod_{i=2}^N \frac{1}{\Lambda_i^3} \langle e^{-\beta \Delta U_i} \rangle_{U_i^o} \left( \frac{2\pi}{\beta k_i} \right)^{3/2} \right] \\ &= -\ln \left( \frac{v^G}{\Lambda_1^3} \right) + \beta a^{*,intra}(T) \end{aligned} \quad (4.8)$$

### 4.3 Testing the Reference Centroid System

For a simple test of the centroid system we compute the free energy of a diatomic molecule using the centroid (see Figure 4.4) and compare it to the analytical solution:

$$\beta a^{molecule} = -\ln \left( \frac{V}{2\Lambda(T)^6} \left[ 2 \left( \frac{\pi}{\beta k_{bond}} \right)^{3/2} + 4r_o^2 \left( \frac{\pi^3}{\beta k_{bond}} \right)^{1/2} \right] \right) \quad (4.9)$$

where, the potential energy of the bond =  $k_{bond}(r - r_o)^2$ ,  $r_o$  is the equilibrium bond distance and  $k_{bond}$  is the force constant. The diatomic molecule is: i) enclosed in a 27,000

$\text{\AA}^3$  volume, ii) assumed to have distinguishable atoms each having a thermal de Broglie wavelength ( $\Lambda$ ) of  $0.292 \text{ \AA}$  at 300 K, and iii) constructed of a spring having a force constant ( $k_{bond}$ ) and bond length ( $r_o$ ) equal to  $786 k_B T / \text{\AA}^2$  and  $1.4 \text{ \AA}$ , respectively.

Further, we compute the free energy of the molecule starting with two centroids having different sets of springs,  $1000 k_B T / \text{\AA}^2$  and  $2000 k_B T / \text{\AA}^2$ . This is because for the centroid to qualify as a reference system, the free energy of the gas phase molecule computed using the centroid as a reference needs to be independent of the strengths of the springs used in the centroid. This is analogous to the free energy calculation of a solid via the Einstein crystal approach. The spring constants used in the Einstein crystal do not affect the final free energy of the solid. The free energies computed via the two centroids were  $-18.03 \pm 0.02 k_B T$  and  $-18.02 \pm 0.03 k_B T$  which are in excellent agreement with the analytical solution of  $-18.037 k_B T$  (see Table 4.1).

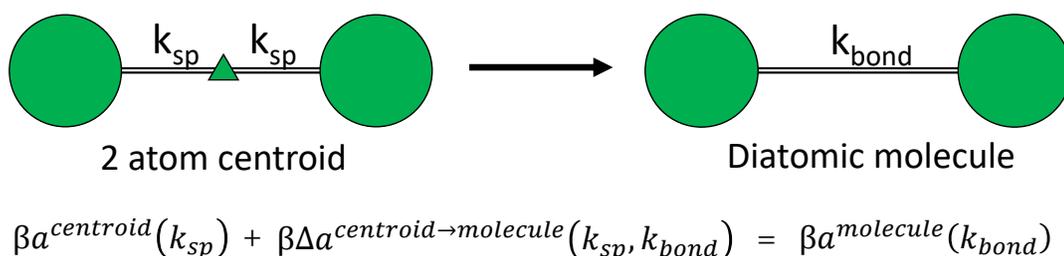


Figure 4.4: Transforming the 2-atom centroid to a diatomic molecule

Table 4.1: Absolute Helmholtz free energy of the diatomic molecule

Centroid	$k_{sp}(kT/\text{\AA}^2)$	$\beta a^*$	$\beta \Delta a^{* \rightarrow molecule}$	$\beta a^{molecule}$	$\beta a_{analytical}^{molecule}$
1	1000	$-11.02 \pm 0.01$	$-7.01 \pm 0.02$	$-18.03 \pm 0.02$	-18.037
2	2000	$-9.99 \pm 0.01$	$-8.03 \pm 0.03$	$-18.02 \pm 0.03$	

## 4.4 Transforming the Centroid into the Molecule

Next, we need to compute the free energy to transform the centroid into an actual molecule. The transformation is done in stages, numbered with integers  $i$  from 1 to 5 as shown in Figure 4.5. The first stage of transformation, 1, involves going from the centroid system to the bonded atoms. The next two stages, 2 and 3, involve turning on the angles and dihedrals, respectively. The last two stages, 4 and 5, comprise turning on the LJ interactions and coulombic interactions, i.e the intra-molecular pairwise interactions, respectively. The transformation free energies are computed using equations

$$\Delta a^i(T) = \int_{\lambda=0}^{\lambda=1} \left\langle \frac{du_i}{d\lambda} \right\rangle_{\lambda} d\lambda \quad (4.10a)$$

where:

$$u_1(T; \lambda) = (1 - \lambda)u^{(1^-)}(T) + \lambda u^{(1^+)}(T) \quad (4.10b)$$

and,

$$u_{i \geq 2}(T; \lambda) = u^{(i^-)}(T) + \lambda u^{(i^+)}(T) \quad (4.10c)$$

where  $\lambda$  is the coupling parameter,  $u^{(i^+)}$  and  $u^{(i^-)}$  are potential energies of states forward and backward comprising transformation  $i$ , respectively. Note, in equation 4.10b we are switching between potential energy terms whereas, in equation 4.10c we are adding a potential energy term during each transformation  $i$ .

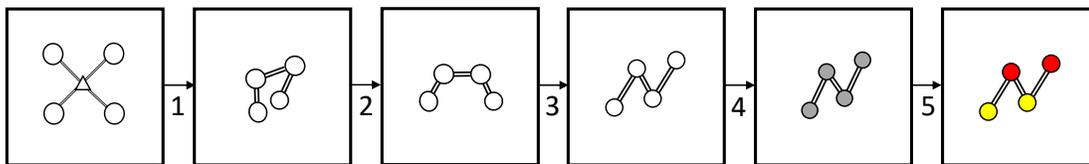


Figure 4.5: Transformation of a centroid to a molecule

Thus, the free energy of the molecule is equal to the sum of the free energy of the reference centroid system and the free energy change to go from the centroid system to the real molecule, i.e.,

$$\mu^G(T, P) = a^*(T, v^G) + \Delta a^{*\rightarrow G}(T) + Pv^G \quad (4.11)$$

# Bibliography

- [1] T. L. Hill, *An introduction to statistical thermodynamics*. Courier Corporation, 1986.
- [2] D. Frenkel and B. Smit, *Understanding molecular simulation: from algorithms to applications*, vol. 1. Elsevier, 2001.
- [3] M. S. Shell, *Thermodynamics and statistical mechanics: an integrated approach*. Cambridge University Press, 2015.
- [4] B. Peters, *Reaction rate theory and rare events*. Elsevier, 2017.
- [5] T. B. Tan, A. J. Schultz, and D. A. Kofke, Suitability of umbrella-and overlap-sampling methods for calculation of solid-phase free energies by molecular simulation, *The Journal of Chemical Physics* **132** (2010), no. 21 214103.
- [6] B. Njagic and M. S. Gordon, Exploring the effect of anharmonicity of molecular vibrations on thermodynamic properties, *The Journal of Chemical Physics* **125** (2006), no. 22 224102.
- [7] J. Cao and G. A. Voth, A new perspective on quantum time correlation functions, *The Journal of Chemical Physics* **99** (1993), no. 12 10070–10073.
- [8] J. Cao and G. A. Voth, The formulation of quantum statistical mechanics based on the feynman path centroid density. ii. dynamical properties, *The Journal of Chemical Physics* **100** (1994), no. 7 5106–5117.
- [9] J. Cao and G. A. Voth, The formulation of quantum statistical mechanics based on the feynman path centroid density. iv. algorithms for centroid molecular dynamics, *The Journal of Chemical Physics* **101** (1994), no. 7 6168–6183.
- [10] R. P. Feynman and H. Kleinert, Effective classical partition functions, *Physical Review A* **34** (1986), no. 6 5080.

# Chapter 5

## Chemical Potential in the Solution Phase

Reproduced in part with permission from:

Vikram Khanna, Jacob I. Monroe, Michael F. Doherty and Baron Peters, “Performing solvation free energy calculations in LAMMPS using the decoupling approach,” *Journal of Computer Aided Molecular Design*, **2020**, 34, 641-646.

DOI: <https://doi.org/10.1007/s10822-020-00303-3>. Copyright 2020 Springer Publishing.

Yongsheng Zhao, ..., Vikram Khanna and Michael F. Doherty, “*First principles based solvent-modified bond energies for crystal engineering* (place holder title)”, *in preparation*.

### 5.1 Introduction

Solvation free energy is the free energy change associated with transferring a solute molecule from an ideal gas phase to a solution phase at a given temperature, pressure, and solute concentration in solution. Solvation free energy computations play an important role in computing solubilities,<sup>[1-8]</sup> partition coefficients,<sup>[9-11]</sup> activity coefficients,<sup>[12]</sup> and

Henry's law constants.<sup>[13]</sup> Established free energy perturbation (FEP)<sup>[14]</sup> and thermodynamic integration (TI)<sup>[15]</sup> methods for computing the solvation free energy fully account for solute-solvent interactions in all solute and solvent arrangements via different alchemical pathways. One approach, called decoupling, modulates only the interactions between the solute and its surrounding, retaining internal interactions. Several molecular dynamics (MD) codes (GROMACS,<sup>[16]</sup> AMBER,<sup>[17]</sup> etc.) include tools to implement the decoupling approach but a key step of charge scaling without altering the intramolecular coulombic interactions cannot be implemented in LAMMPS.<sup>[18]</sup> This chapter demonstrates a procedure to implement the decoupling approach in LAMMPS without altering the source code (an example of altering the source code can be found in the work of Paluch et al.<sup>[19]</sup> where they modify the code of the MD simulation package MDynamix<sup>[20, 21]</sup>). Beyond LAMMPS, our procedure can be implemented with any MD code with equivalent functions as described in this chapter.

## 5.2 Thermodynamic Cycle

In order to perform a free energy calculation we need to define the end states. For solvation free energy calculations one end state (as shown in Figure 5.1) is the isolated molecule (A) in the gas phase along with the solution phase (B) in the background, i.e., state (1). The other end state is the solvated molecule in the solution (B) phase, i.e., state (6). In state (1), the atoms of the isolated molecule interact only with each other. Since the molecule being solvated is isolated from the surrounding, we refer to this state as the **decoupled state**, because the atoms are not interacting with (decoupled from) the surrounding environment. In state (6), the atoms of the molecule interact with all the atoms present in the system, including the solvent and other dissolved solutes. We refer to this state as the **coupled state**. The thermodynamic cycle shown in Figure 5.1

depicts the alternate paths between the end states for solvating a molecule.<sup>[22]</sup> We can construct numerous alternate paths for computing solvation free energies, the two common implemented paths are: i) The **annihilation route**,  $(1) \rightarrow (2) \rightarrow (3) \rightarrow (4) \rightarrow (5) \rightarrow (6)$  and ii) The **decoupling route**,  $(1) \rightarrow (6)$  with an intermediate state  $(6')$  as shown in Figure 5.2. The annihilation route involves annihilation of the solute's pairwise (inter + intra) interactions in the gas and solution phases. The thermodynamic path  $(1) \rightarrow (2) \rightarrow (3)$  is the gas phase annihilation path, where the coulombic interactions are turned off from  $(1) \rightarrow (2)$  followed by turning off of the Lennard-Jones (LJ) interactions from  $(2) \rightarrow (3)$ , in the isolated gas phase. The thermodynamic path  $(6) \rightarrow (5) \rightarrow (4)$  is the solution phase annihilation path, where the coulombic interactions are turned off from  $(6) \rightarrow (5)$  followed by turning off of the LJ interactions from  $(5) \rightarrow (4)$ , in the solution phase. The decoupling route involves decoupling the solute from its surrounding environment by turning off the solute-solution pairwise interactions.

In order to compute the free energy difference between states (1) and (6), we need to sum the free energy differences between successive states along the chosen route. The annihilation route can be implemented in LAMMPS but so far the decoupling route cannot as scaling the charges on the solute to decouple it from the solution does not preserve its intramolecular interactions. However, we have developed a methodology using which the decoupling route can be implemented in LAMMPS. In order to compute the free energy change along path  $(6) \rightarrow (1)$  we need to compute the free energy change of turning off the solute-solution (A-B) interactions leaving the intramolecular solute interactions turned on.

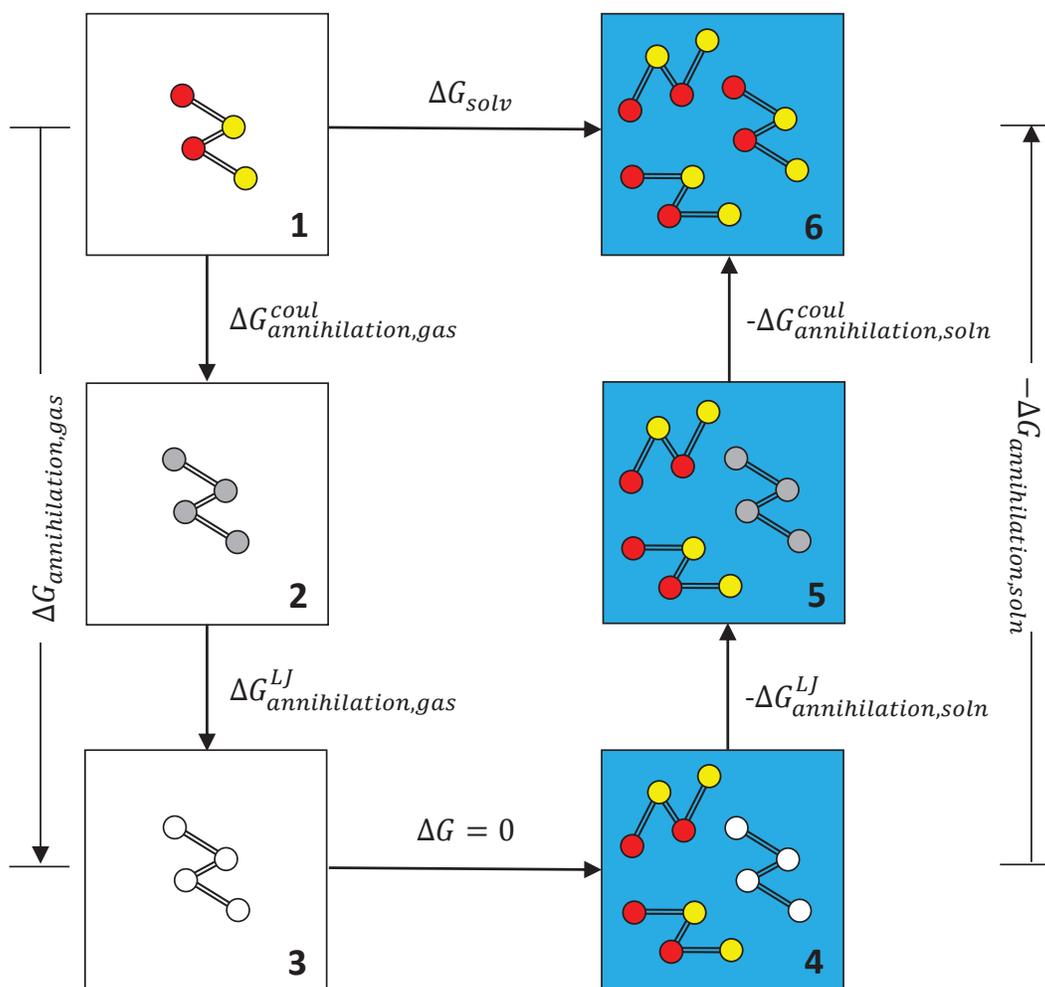


Figure 5.1: The Thermodynamic Solvation Energy Cycle,<sup>[22]</sup> white box: isolated gas phase; blue box: water; colored atoms: LJ=ON, coul=ON, grey atoms: LJ=ON, COUL=OFF, and white atoms: LJ=OFF, coul=OFF. The interactions of the solute under discussion are inter+intra for solution phase and intra only for the isolated gas phase. Note: in state (1), (2) and (3) the solution phase (into which the isolated molecule is being solvated) is present in the background (as explicitly shown in Figure 5.2) to complete the reactant side of the physical reaction of solvation. For clarity, we do not show the solution phase on the left side of the figure

### 5.3 Setting up the Decoupling Route

The free energy change along the path (1)→(6′)→(6) is determined by computing the free energy change of turning on the LJ interactions between the solute (A) and the solution (B), followed by the free energy change of turning on the coulombic interactions between them. We use thermodynamic integration (TI)<sup>[15]</sup> to compute these free energy changes, where

$$\Delta G_{(i)\rightarrow(j)} = \int_{\lambda=0}^{\lambda=1} \left\langle \frac{dU}{d\lambda} \right\rangle_{\lambda} d\lambda \quad (5.1)$$

and  $U(\lambda)$  is a Hamiltonian that continuously changes from that of system ( $i$ ) to system ( $j$ ) as  $\lambda$  goes from 0 to 1. The decoupling route is shown in Figure 5.2. Here we are performing a hydration free energy calculation as the molecule is being solvated in pure water. Throughout states (1), (6′) and (6) the solute’s intramolecular LJ and coulombic interactions are on. State (1) comprises the isolated gas phase molecule and the water box. In state (6′) only the solute-water (intermolecular) LJ interactions are on, and in state (6) the solute-water LJ and coulombic interactions are on. We use Gauss-Legendre quadrature to select intermediate  $\lambda$ 's from (1) to (6′) and from (6′) to (6). The free energy changes computed along these paths,  $\Delta G_{solv}^{LJ}$  and  $\Delta G_{solv}^{coul}$ , respectively, are then added to give the solvation free energy ( $\Delta G_{solv}$ ).

To compute  $\Delta G_{solv}^{LJ}$ , the LJ interactions are turned on via the soft-core potential.<sup>[23]</sup> A coupling parameter ( $\lambda_{LJ}$ ) is varied between 0 and 1 to march from state (1)→(6′), where  $\lambda_{LJ} = 0$  means there are no LJ interactions between A and B, and  $\lambda_{LJ} = 1$  means the LJ interactions are completely turned on. Likewise, for the electrostatics, a coupling parameter ( $\lambda_{coul}$ ) is varied between 0 and 1 to compute  $\Delta G_{solv}^{coul}$ , where

$$\Delta G_{solv}^i = \int_{\lambda_i=0}^{\lambda_i=1} \left\langle \frac{dU}{d\lambda} \right\rangle_{\lambda_i} d\lambda_i \quad (5.2)$$

where,  $i$  represents LJ or coul.

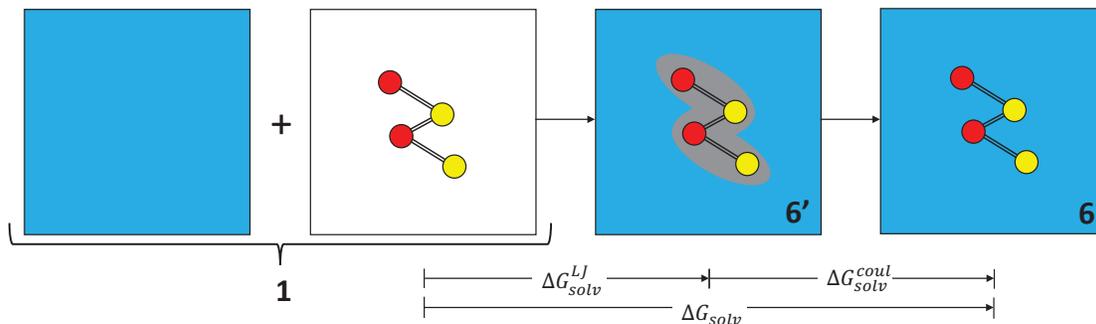


Figure 5.2: The decoupling thermodynamic path. Grey shaded area:  $\text{LJ}_{\text{A-B}}=\text{ON}$ ,  $\text{COUL}_{\text{A-B}}=\text{OFF}$ ; colored atoms: intramolecular LJ and coulombic interactions= $\text{ON}$

### 5.3.1 Staging the Lennard Jones Interactions

Turning on the LJ interactions between A and B, i.e., coupling them, is straightforward in LAMMPS. The soft-core potential used in LAMMPS is given by:

$$U_{LJ}(r_{ij}; \lambda_{ij}) = \lambda_{ij}^2 4\epsilon \left\{ \frac{1}{\left[ \frac{(1-\lambda_{ij})^2}{2} + \left( \frac{r_{ij}}{\sigma_{ij}} \right)^6 \right]^2} - \frac{1}{\frac{(1-\lambda_{ij})^2}{2} + \left( \frac{r_{ij}}{\sigma_{ij}} \right)^6} \right\} \quad (5.3)$$

In LAMMPS we can assign a  $\lambda_{LJ}$  value to each pairwise interaction between the atom types. Therefore, we set

$$\begin{aligned} \lambda_{ij} &= \lambda_{LJ} \quad \forall \quad i \in A, j \in B \\ &= 1 \quad \forall \quad i = j \in A \end{aligned} \quad (5.4)$$

This enables us to scale the A-B LJ interactions without altering the A-A LJ interactions. However, one important point to note while staging the LJ interactions is that we also need to keep the solute's intramolecular coulombic interactions turned on. We achieve

this in our framework by setting the solute’s charge scaling parameter (described in the following subsection),  $\lambda_q$ , equal to  $10^{-9}$ . This is because in the overlay potential (also described in the following subsection), described by equation 5.7, we cannot insert integer zero as the value for  $\lambda_q$ . Hence, we set it to  $10^{-9}$  (which is as an approximate to zero).

### 5.3.2 Staging the Coulombic Interactions

For turning on the coulombic interactions between A and B, we need to scale the charges on A in order to correctly account for the long range electrostatic interactions via the particle mesh Ewald sum. However, in doing so we scale the intramolecular coulombic interactions of A as well. This is not acceptable for the decoupling approach as we want the intramolecular interactions of A to be “unaltered”. Thus, what we want is

$$U_{coul}^{intra}(r_{ij}) = \sum_{i<j} k \frac{q_i q_j}{r_{ij}}; \quad \forall i, j \in A \quad (5.5)$$

where  $k = 1/(4\pi\epsilon)$ ,  $\epsilon$  being the permittivity of free space. However, since we scale  $q_i \rightarrow \lambda_q q_i \forall i \in A$ , the intramolecular (A-A) electrostatic interactions become,

$$\begin{aligned} U_{coul}^{intra,scaled}(r_{ij}; \lambda_q) &= \sum_{i<j} k \frac{(\lambda_q q_i) (\lambda_q q_j)}{r_{ij}} \\ &= \sum_{i<j} \lambda_q^2 k \frac{q_i q_j}{r_{ij}} \end{aligned} \quad (5.6)$$

The scaling of charges on the solute thus reduces the intramolecular interaction from that of an ideal gas. To recover the correct intramolecular Coulomb energy, we overlay,

$$\begin{aligned}
 U_{coul}^{overlay}(r_{ij}; \lambda_q) &= U_{coul}^{intra}(r_{ij}) - U_{coul}^{intra,scaled}(r_{ij}; \lambda_q) \\
 &= \sum_{i<j} (1 - \lambda_q^2) \left\{ k \frac{q_i q_j}{r_{ij}} \right\} \\
 &= \left( \frac{1 - \lambda_q^2}{\lambda_q^2} \right) U_{coul}^{intra,scaled}(r_{ij}; \lambda_q)
 \end{aligned} \tag{5.7}$$

over the  $U_{coul}^{intra,scaled}(r_{ij}, \lambda_q)$  interactions, to recover  $U_{coul}^{intra}(r_{ij})$ . This overlay of potential is implemented in LAMMPS by using the *hybrid overlay* command and overlaying a *coul/cut/soft* potential. This potential has the following functional form (using the scaled charges)

$$U_{coul/cut/soft}(r_{ij}; \lambda_{coul}) = \sum_{i<j} \lambda_{coul} k \frac{(\lambda_q q_i)(\lambda_q q_j)}{r_{ij}} \tag{5.8}$$

where  $\lambda_{coul}$  is the user input parameter which we input as  $(1 - \lambda_q)^2 / \lambda_q^2$  (see Equation 5.7). This exact overlay correction is the main contribution in this work.

## 5.4 Testing the Computational Strategy

In order to test the procedure laid out for computing  $\Delta G_{solv}^{coul}$ , we compute it for two compounds: i) ethanol and ii) biphenyl using LAMMPS and test the results against the results obtained from GROMACS which has a built-in function for performing solvation free energies using the decoupling approach. We compare  $\langle dU/d\lambda \rangle_\lambda$  at each  $\lambda$  state outputted by LAMMPS and GROMACS, as well as the total coulombic free energy change to test our approach.

### 5.4.1 Simulation Details: LAMMPS

We perform hydration free energy calculations using the procedure laid out in the above sections. Ethanol and biphenyl described by the General Amber Force Field (GAFF, version 1.7)<sup>[24]</sup> and AM1-BCC charges<sup>[25, 26]</sup> were solvated in 1288 TIP3P water molecules.<sup>[27]</sup> The Lorentz-Berthelot mixing rules<sup>[28–30]</sup> were used for computing the interatomic pair coefficients. The initial configurations were taken from the FreeSolv database.<sup>[31]</sup> We used a time step of 1.0 fs. Bonds with hydrogen were constrained using SHAKE.<sup>[32, 33]</sup> Nonbonded interactions were cutoff at 1 nm, with long-range electrostatics handled by LAMMPS’ Particle-Particle-Particle-Mesh (PPPM) summation.<sup>[34, 35]</sup> Long-range dispersion corrections to energy and pressure were applied. The NPT ensemble was simulated using Nose-Hoover thermostat and barostat<sup>[36]</sup> at  $T = 298.15$  K and  $P = 1.01325$  bar. All systems were equilibrated for 3 ns followed by 5 ns production runs. A 15 point Gauss-Legendre quadrature method<sup>[37]</sup> was used to evaluate the integral in equation 5.2 for computing the coulombic solvation free energy. For computing the solvation free energy due to the LJ interactions we used a 10, 15, and 15 point Gauss-Legendre quadrature for  $0 < \lambda_{LJ} < 0.2$ ,  $0.2 < \lambda_{LJ} < 0.4$ , and  $0.4 < \lambda_{LJ} < 1.0$ , respectively, amounting to a total of 40 stages. Data every 1 ps was used to compute the thermodynamic averages.

### 5.4.2 Simulation Details: GROMACS

For the GROMACS simulations, we follow procedures as identical as possible to those described in the FreeSolv database,<sup>[31]</sup> only modifying run input files taken from the FreeSolv repository<sup>[38]</sup> minimally in order to ensure compatibility with the GPU-enabled GROMACS version 2016.1.<sup>[39]</sup> Briefly, a neighbor list was updated every 10 steps with a cutoff for its construction of 1.2 nm. Nonbonded interactions were cutoff at 1.0 nm, with long-range electrostatics handled by particle mesh Ewald techniques<sup>[40]</sup> and a switching

function applied for Lennard-Jones interactions between 0.9 and 1.0 nm. Long-range dispersion corrections to energy and pressure were applied, with all bonds involving hydrogens constrained via the LINCS algorithm<sup>[41]</sup> and water kept rigid via SETTLE.<sup>[42]</sup> At each of the  $\lambda$  states described in other sections, we performed 2500 steepest descent steps to minimize fully solvated structure and topology files downloaded from the FreeSolv database.<sup>[31]</sup> After minimization, a Langevin integrator<sup>[43]</sup> was used throughout to propagate dynamics with a 1 fs time step. We first equilibrated temperature at constant density for 25 ps, then simulated in the NPT ensemble for 25 ps to equilibrate density with both a tightly coupled Berendsen thermostat and isotropic barostat,<sup>[44]</sup> and finally performed a 3.0 ns run in the NPT ensemble with a Langevin thermostat (stochastic integrator)<sup>[43]</sup> and isotropic Parinello-Rahman barostat,<sup>[45, 46]</sup> which were also used for production. Production runs for 5.0 ns were performed at a constant temperature of 298.15 K and pressure of 1.01325 bar, with derivatives with respect to  $\lambda$  and potential energies at all states outputted every 1 ps to be averaged for thermodynamic integration (TI).

## 5.5 Results and Discussion

We test our approach to compute the solvation free energies for turning on the coulombic interactions between the solute and solution,  $\Delta G_{solv}^{coul}$ , in LAMMPS, by comparing our results against GROMACS' results (which uses *intermol* functionality to correctly account for the intramolecular interactions). The results are shown in Table 5.1 which are in good agreement. The error bars on the values represent a standard 68% confidence interval (C.I.). This means approximately 1 in every 3 simulations will give results outside the specified interval. Further, Figure 5.3 compares the integrand for turning on the Coulomb interactions,  $\langle dU/d\lambda_q \rangle_{\lambda_q}$  as a function of  $\lambda_q$ . As can be seen, the integrands

are in good agreement too. Thus, we have demonstrated and tested our procedure for performing solvation free energy calculations in LAMMPS using the decoupling approach via counterbalancing the scaling of charges on the solute by overlaying a compensating potential to keep the intramolecular interactions unchanged to that of an ideal gas.

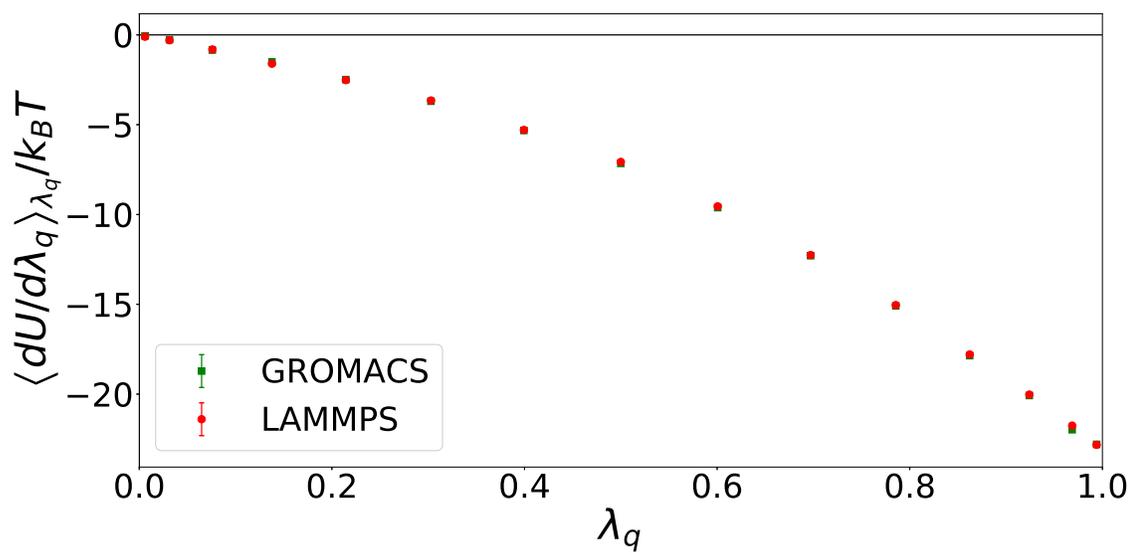
Table 5.1: RESULTS FOR  $\Delta G_{solv}^{coul}/k_B T$ 

Compound	LAMMPS	GROMACS	DIFFERENCE
ethanol	$-8.73 \pm 0.02$	$-8.76 \pm 0.02$	$0.03 \pm 0.03$
biphenyl	$-8.30 \pm 0.02$	$-8.29 \pm 0.02$	$-0.01 \pm 0.03$

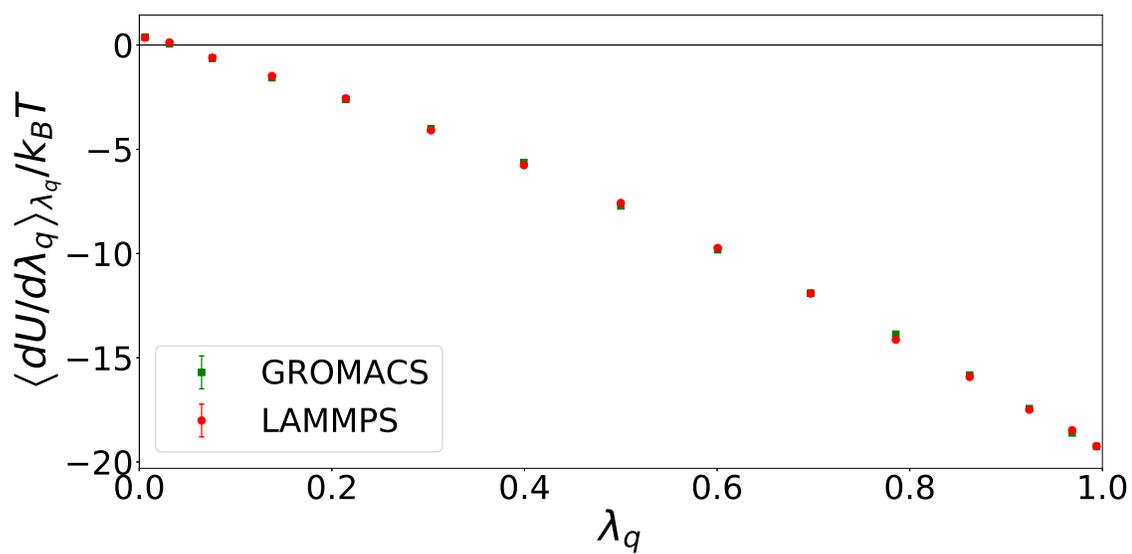
We report the total (dimensionless) solvation free energies ( $\Delta G_{solv}/k_B T$ ) in Table 5.2 for completeness. Our values are consistent across the two tests, and agree with the values computed by Mobley et al.<sup>[47]</sup> considering a 95% C.I. and that Mobley et al. used 15 and 5 stages, whereas we used 40 and 15 stages for turning on the LJ and coulombic interactions, respectively. Note: We use GAFF version 1.7 instead of 1.8 in order to make a comparison to the values reported in [31].

Table 5.2: RESULTS FOR  $\Delta G_{solv}/k_B T$ 

Compound	LAMMPS	GROMACS	FreeSolv <sup>[31]</sup>
ethanol	$-5.76 \pm 0.03$	$-5.79 \pm 0.03$	$-5.72 \pm 0.03$
biphenyl	$-5.22 \pm 0.04$	$-5.27 \pm 0.04$	$-5.30 \pm 0.05$



(a)



(b)

Figure 5.3:  $\langle dU/d\lambda_q \rangle_{\lambda_q} / k_B T$  as a function of  $\lambda_q$  for a) ethanol, b) biphenyl

## 5.6 Solvent-Modified Bond Energies—A new application for solvation free energy calculations

Solvent selection for solution grown crystals is perhaps one of the most critical decisions while designing a crystallization process as it impacts the solubility and morphology of the crystal. The morphology of crystal is determined by the interaction energetics—both i) within the growing crystal and ii) between the crystal and the growth medium. The shape of a crystal grown from solution significantly differs from the one grown from vapor as the presence of a solvent alters the relevant surface interactions due to the solvation of crystal surfaces. Therefore, the predictions of these crystal-solvent interactions, i.e., modification of the crystal surface interactions due to the presence of a solvent, is crucial to accurate morphology predictions.<sup>[48]</sup>

The current interfacial energy models employed in ADDICT make use of empirical data to predict solvent modified bond energies. This approach while fast isn't first principle based and relies on the need for performing experiments for every new molecule synthesized. Also, it is not poised to be extended to solvent mixtures.

We have discovered an application of solvation free energy calculations in computing solvent-modified bond energies which overcomes the drawbacks of the current empirical interfacial energy models.

So what is a solvent-modified bond energy? As shown in Figure 5.4 bond energy is the amount of energy required to break a bond and take the constituent units (atoms/molecules) to infinity—in vacuum. A solvent modified bond energy is the amount of energy required to break a bond and take the constituent units to infinity—in the solvent medium. These solvent-modified bond energies will replace the empirical solvent models currently being employed in the mechanistic modeling of crystal growth.

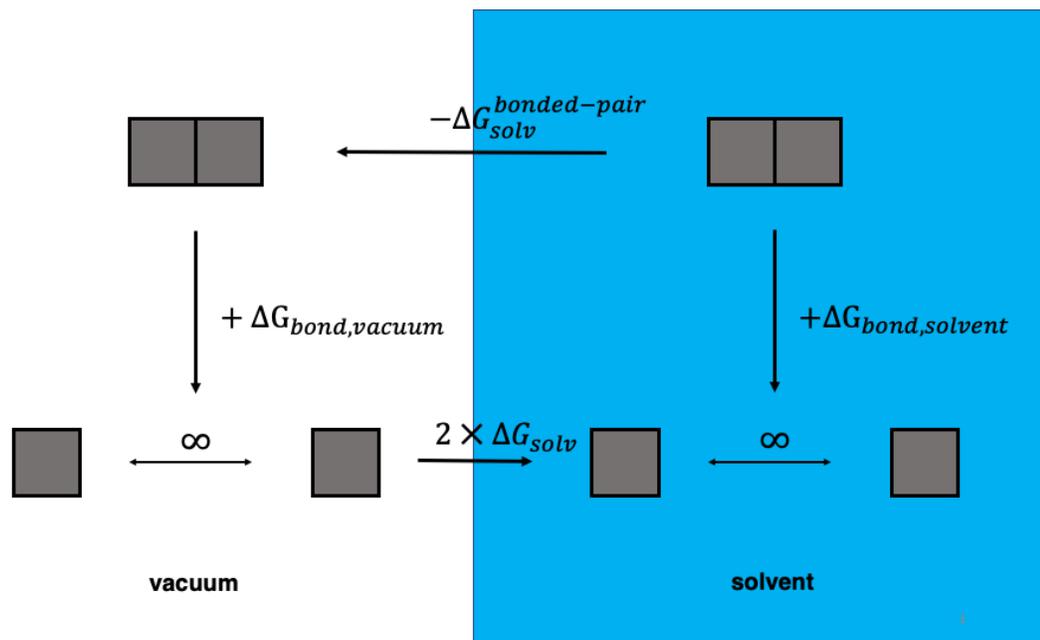


Figure 5.4: The thermodynamic framework to compute solvent-modified bond energies.

To compute the solvent-modified bond energy, we start by desolvating the bonded-pair ( $-\Delta G_{solv}^{bonded-pair}$ ). This gives us a bonded-pair in vacuum. Now, we break the bond in vacuum ( $\Delta G_{bond,vacuum}$ ), resulting in two atoms/molecules in vacuum. Next, each of these atoms/molecules is solvated yielding two atoms/molecules in the solvent ( $2 \times \Delta G_{solv}$ ). Thus, in this thermodynamic cycle we start from a solvated bonded-pair and end with two solvated molecules—breaking the bond in the solvent medium.

$$\Delta G_{bond,solvent} = -\Delta G_{solv}^{bonded-pair} + \Delta G_{bond,vacuum} + 2\Delta G_{solv} \quad (5.9)$$

Approximating,  $\Delta(PV)_{bond,vacuum} \approx \Delta(PV)_{bond,solvent}$  and  $\Delta(S)_{bond,vacuum} \approx \Delta(S)_{bond,solvent}$ , we get

$$\Delta U_{bond,solvent} = -\Delta G_{solv}^{bonded-pair} + \Delta U_{bond,vacuum} + 2\Delta G_{solv} \quad (5.10)$$

Therefore, we can compute the solvent-modified bond energy using the following three

computations:

- $\Delta G_{solv}^{bonded-pair}$
- $\Delta U_{bond,vacuum}$
- $\Delta G_{solv}$

## 5.7 Conclusions

We have introduced and tested a procedure to carry out solvation free energy calculations in LAMMPS using the decoupling approach. This procedure can be implemented in any MD code which allows the user to define ‘overlay’ potentials. We have only compared solvation free energies between LAMMPS and GROMACS for two molecules to check the implementation because the overlay correction is, in principle, exact. Also, while we have chosen to implement thermodynamic integration here for simplicity, the overlay correction should work equally well with other free energy calculations such as Bennett Acceptance Ratio (BAR)<sup>[49]</sup> and Multistate Bennett Acceptance Ratio (MBAR).<sup>[50]</sup> We further layout a new application of solvation free energy calculations to compute solvent-modified bond energies which will replace the empirical solvent models currently employed in crystal modeling.

# Bibliography

- [1] E. Sanz and C. Vega *The Journal of Chemical Physics* **126** (2007), no. 1 014507.
- [2] J. Aragonés, E. Sanz, and C. Vega *The Journal of Chemical Physics* **136** (2012), no. 24 244508.
- [3] A. Benavides, J. Aragonés, and C. Vega *The Journal of Chemical Physics* **144** (2016), no. 12 124504.
- [4] J. Espinosa, J. Young, H. Jiang, D. Gupta, C. Vega, E. Sanz, P. Debenedetti, and A. Panagiotopoulos *The Journal of Chemical Physics* **145** (2016), no. 15 154111.
- [5] M. J. Schnieders, J. Baltrusaitis, Y. Shi, G. Chattree, L. Zheng, W. Yang, and P. Ren *Journal of Chemical Theory and Computation* **8** (2012), no. 5 1721–1736.
- [6] L. Li, T. Totton, and D. Frenkel *The Journal of Chemical Physics* **146** (2017), no. 21 214110.
- [7] D. R. M. Guilherme and D. L. Mobley *F1000Research* **7** (2018).
- [8] M. A. Bellucci, G. Gobbo, T. K. Wijethunga, G. Ciccotti, and B. L. Trout *The Journal of Chemical Physics* **150** (2019), no. 9 094107.
- [9] N. M. Garrido, I. G. Economou, A. J. Queimada, M. Jorge, and E. A. Macedo *AIChE Journal* **58** (2012), no. 6 1929–1938.
- [10] L. Yang, A. Ahmed, and S. I. Sandler *Journal of Computational Chemistry* **34** (2013), no. 4 284–293.
- [11] D. U. Bapat and V. H. Dalvi *The Journal of Physical Chemistry B* (2019).
- [12] Z. Mester and A. Z. Panagiotopoulos *The Journal of Chemical Physics* **142** (2015), no. 4 044507.
- [13] M. S. Shell, *Thermodynamics and Statistical Mechanics: An Integrated Approach*. Cambridge University Press, 2015.
- [14] R. W. Zwanzig *The Journal of Chemical Physics* **22** (1954), no. 8 1420–1426.
- [15] J. G. Kirkwood *The Journal of Chemical Physics* **3** (1935), no. 5 300–313.
- [16] H. J. Berendsen, D. van der Spoel, and R. van Drunen *Computer Physics Communications* **91** (1995), no. 1-3 43–56.

## BIBLIOGRAPHY

---

- [17] D. A. Pearlman, D. A. Case, J. W. Caldwell, W. S. Ross, T. E. Cheatham III, S. DeBolt, D. Ferguson, G. Seibel, and P. Kollman *Computer Physics Communications* **91** (1995), no. 1-3 1–41.
- [18] S. Plimpton *Journal of Computational Physics* **117** (1995), no. 1 1–19.
- [19] A. S. Paluch, J. K. Shah, and E. J. Maginn *Journal of Chemical Theory and Computation* **7** (2011), no. 5 1394–1403.
- [20] A. P. Lyubartsev and A. Laaksonen *Computer Physics Communications* **128** (2000), no. 3 565–589.
- [21] A. Lyubartsev and A. Laaksonen, Mdynamix: a molecular dynamics program, 2010.
- [22] P. V. Klimovich, M. R. Shirts, and D. L. Mobley *Journal of Computer-Aided Molecular Design* **29** (2015), no. 5 397–411.
- [23] T. C. Beutler, A. E. Mark, R. C. van Schaik, P. R. Gerber, and W. F. Van Gunsteren *Chemical Physics Letters* **222** (1994), no. 6 529–539.
- [24] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, and D. A. Case *Journal of Computational Chemistry* **25** (2004), no. 9 1157–1174.
- [25] A. Jakalian, B. L. Bush, D. B. Jack, and C. I. Bayly *Journal of Computational Chemistry* **21** (2000), no. 2 132–146.
- [26] A. Jakalian, D. B. Jack, and C. I. Bayly, Fast, efficient generation of high-quality atomic charges. am1-bcc model: Ii. parameterization and validation, *Journal of computational chemistry* **23** (2002), no. 16 1623–1641.
- [27] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein *The Journal of Chemical Physics* **79** (1983), no. 2 926–935.
- [28] D. Berthelot *Compt. Rendus* **126** (1898) 1703–1706.
- [29] H. Lorentz *Annalen der physik* **248** (1881), no. 1 127–136.
- [30] M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids*. Oxford University Press, 2017.
- [31] D. L. Mobley and J. P. Guthrie *Journal of Computer-Aided Molecular Design* **28** (2014), no. 7 711–720.
- [32] J.-P. Ryckaert, G. Ciccotti, and H. J. Berendsen *Journal of Computational Physics* **23** (1977), no. 3 327–341.
- [33] H. C. Andersen *Journal of Computational Physics* **52** (1983), no. 1 24–34.
- [34] P. P. Ewald *Annalen der physik* **369** (1921), no. 3 253–287.
- [35] R. W. Hockney and J. W. Eastwood, *Computer simulation using particles*. crc

## BIBLIOGRAPHY

---

- Press, 1988.
- [36] G. J. Martyna, D. J. Tobias, and M. L. Klein *The Journal of Chemical Physics* **101** (1994), no. 5 4177–4189.
- [37] M. Abramowitz and I. A. Stegun, *Handbook of mathematical functions: with formulas, graphs, and mathematical tables*, vol. 55. Courier Corporation, 1965.
- [38] D. L. Mobley, M. Shirts, N. Lim, J. Chodera, K. Beauchamp, and Lee-Ping, Mobleylab/freesolv: Version 0.52, Jan., 2018.
- [39] M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess, and E. Lindahl *SoftwareX* **1** (2015) 19–25.
- [40] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen *The Journal of Chemical Physics* **103** (1995), no. 19 8577–8593.
- [41] B. Hess, H. Bekker, H. J. Berendsen, and J. G. Fraaije *Journal of Computational Chemistry* **18** (1997), no. 12 1463–1472.
- [42] S. Miyamoto and P. A. Kollman *Journal of Computational Chemistry* **13** (1992), no. 8 952–962.
- [43] N. Goga, A. Rzepiela, A. De Vries, S. Marrink, and H. Berendsen *Journal of Chemical Theory and Computation* **8** (2012), no. 10 3637–3649.
- [44] H. J. Berendsen, J. v. Postma, W. F. van Gunsteren, A. DiNola, and J. Haak *The Journal of Chemical Physics* **81** (1984), no. 8 3684–3690.
- [45] M. Parrinello and A. Rahman *Journal of Applied Physics* **52** (1981), no. 12 7182–7190.
- [46] S. Nosé and M. Klein *Molecular Physics* **50** (1983), no. 5 1055–1076.
- [47] G. Duarte Ramos Matos, D. Y. Kyu, H. H. Loeffler, J. D. Chodera, M. R. Shirts, and D. L. Mobley *Journal of Chemical & Engineering Data* **62** (2017), no. 5 1559–1569.
- [48] C. J. Tilbury, D. A. Green, W. J. Marshall, and M. F. Doherty, Predicting the effect of solvent on the crystal habit of small organic molecules, *Crystal Growth & Design* **16** (2016), no. 5 2590–2604.
- [49] C. H. Bennett, Efficient estimation of free energy differences from monte carlo data, *Journal of Computational Physics* **22** (1976), no. 2 245–268.
- [50] M. R. Shirts and J. D. Chodera, Statistically optimal analysis of samples from multiple equilibrium states, *The Journal of chemical physics* **129** (2008), no. 12 124105.

# Chapter 6

## Predicting Solid-Fluid Equilibrium

Reproduced in part with permission from:

Vikram Khanna, Michael F. Doherty and Baron Peters, “Absolute chemical potentials for complex molecules in fluid phases: A centroid reference for predicting phase equilibria,” *Journal of Chemical Physics*, **2020**, 153(21), 214504:1-9.

DOI: <https://doi.org/10.1063/5.0025844>. Copyright 2020 AIP Publishing.

Vikram Khanna, Michael F. Doherty and Baron Peters, “Predicting driving forces for crystallization using the absolute chemical potential route,” In preparation.

### 6.1 Introduction

This chapter leverages the tools we have developed to compute a solute’s chemical potential in the solid, gas, and solution phases to predict solid-fluid equilibria and the driving forces for crystallization. We make predictions for two model compounds, naphthalene (representative of a rigid molecule) and succinic acid (representative of a floppy molecule)—thereby spanning the two terminuses on the rigidity scale. Further, for suc-

cinic acid, we investigate in addition to its known stable polymorph (the  $\beta$  polymorph), the recently discovered  $\gamma$  polymorph, which has sparse experimental data. The  $\gamma$  polymorph predictions demonstrate the future potential of the digital design approach where computational studies can predict thermodynamic properties for difficult to crystallize polymorphs/compounds subject to the accuracy of the employed force fields.

For each model compound, we begin by predicting the solid-vapor equilibrium as this prediction employs only a single force field—the one describing the solute—allowing for a standalone quality (accuracy) assessment of the solute’s force field. Next, we predict the solid-solution equilibrium which involves the use of an additional solvent force field. The accuracy of this prediction relies upon the accuracy of the following three interactions: i) solute-solute (the accuracy of which is already assessed via the solid-vapor equilibrium predictions), ii) solute-solvent, and iii) solvent-solvent.

Therefore, if the accuracy of the solid-vapor predictions is good, the accuracy of the solid-solution predictions can be reduced to the accuracy of the solute-solvent, and solvent-solvent interactions. In the following sections, we make the solid-fluid equilibria predictions and where appropriate make use of the above framework to comment on the accuracy of the predictions.

## 6.2 Predicting Solid-Vapor Equilibrium

At solid-vapor equilibrium we have,

$$\mu^C(T, P^{sat}) = \mu^G(T, P^{sat}). \quad (6.1a)$$

As shown in Figure 6.1, we start with computing the reference system’s free energy in each of the phases followed by computing the free energy change to transform to the

real systems.

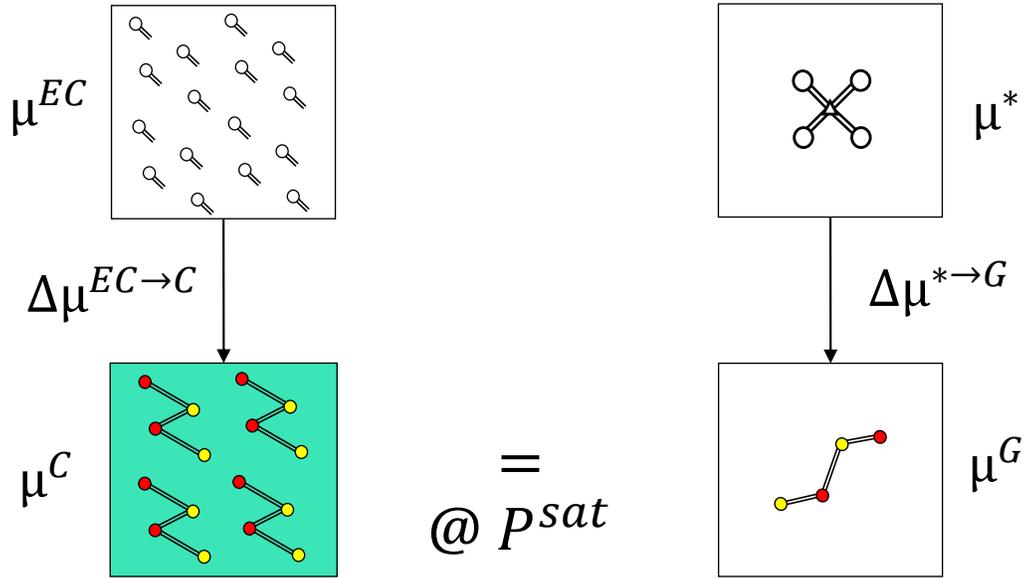


Figure 6.1: Computing the absolute chemical potentials of each phase. The left and right starting points are absolute chemical potentials of the Einstein crystal and Centroid, respectively

For solids with low sublimation vapor pressure ( $\sim$  mPa), we can assume that the ideal gas law is valid, thus, we insert  $k_B T$  in place for  $P^{sat} v^G$  in equation 4.11. Also,  $\mu^C(T, P^{sat}) \approx a^C(T, v^C)$  (see Appendix H.1 for details). This enables us to compute the Helmholtz free energy of the solid equilibrated at 1 atm and use that free energy value in equation 6.1a. Thus, we get

$$\begin{aligned}
 a^C(T, v^C) &= a^*(T, v^G) + \Delta a^{* \rightarrow G}(T) + k_B T + \Delta a^{sym} \\
 &= -k_B T \ln \left( \frac{v^G}{\Lambda_1^3} \right) + a^{*,intra}(T) + \Delta a^{* \rightarrow G}(T) + k_B T + \Delta a^{sym}.
 \end{aligned}
 \tag{6.1b}$$

From independent sets of simulations we can compute  $a^C(T, v^C)$ ,  $\Delta a^{* \rightarrow G}(T)$  and  $a^{*,intra}(T)$ . Inserting them in equation 6.1b and solving for  $v^G$  we get the molar volume of an ideal gas molecule. Finally, we get the vapor pressure by using the ideal gas law,

i.e.,  $P^{sat} = k_B T / v^G$ . Note:  $\Delta a^{sym}$  accounts for the degeneracy in sampling the gas phase for a symmetric molecule.<sup>[1, 2]</sup> For naphthalene and succinic acid,  $\beta \Delta a^{sym} = \ln(4)$ , and  $\beta \Delta a^{sym} = \ln(2)$ , respectively.

In this work, we have not included nuclear quantum effects (NQE). Accordingly, our results are independent of Planck’s constant, but quantized vibrations are potentially important in the free energy. The preferred calculation would include quantized vibrational energy levels but including discrete energy levels would be difficult especially for the centroid calculation. It should be straightforward to include NQE via centroid calculations with path integral molecular dynamics.<sup>[3, 4]</sup> One of the most important applications of free energy calculations is in tuning force fields (FFs). FFs can be tuned via two different strategies:<sup>[5, 6]</sup> i) so that a classical MD simulation directly reproduces experimental properties,<sup>[7–10]</sup> and ii) so that the Born–Oppenheimer potential energy surface closely matches high level ab initio calculations. The purely classical simulations in our study are appropriate for FFs in category (i), but for FFs in category (ii) the methods in this work can be combined with path integral simulations to capture NQE.<sup>[11–13]</sup>

### 6.3 Simulation Details

We model naphthalene using the OPLS-AA force field<sup>[14]</sup> and succinic acid with the General Amber Force Field (version 1.8)<sup>[15]</sup> and AM1-BCC charges<sup>[16, 17]</sup> (see Figure 6.2 for a ball and stick model of these molecules). We compute the vapor pressure of naphthalene at 4 different temperatures—298 K, 308.17 K, 318.17 K, 333.34 K, and that of succinic acid at 6 different temperatures, namely 300K, 305K, 310K, 315K, 330K, and 350K. All molecular dynamics (MD) simulations for the solid and gas phases were performed using the Large-scale Atomic/Molecular Massively Parallel Simulator (LAMMPS) package.<sup>[18]</sup> All MD simulations for the solution phases were performed using OpenMM.<sup>[19]</sup>

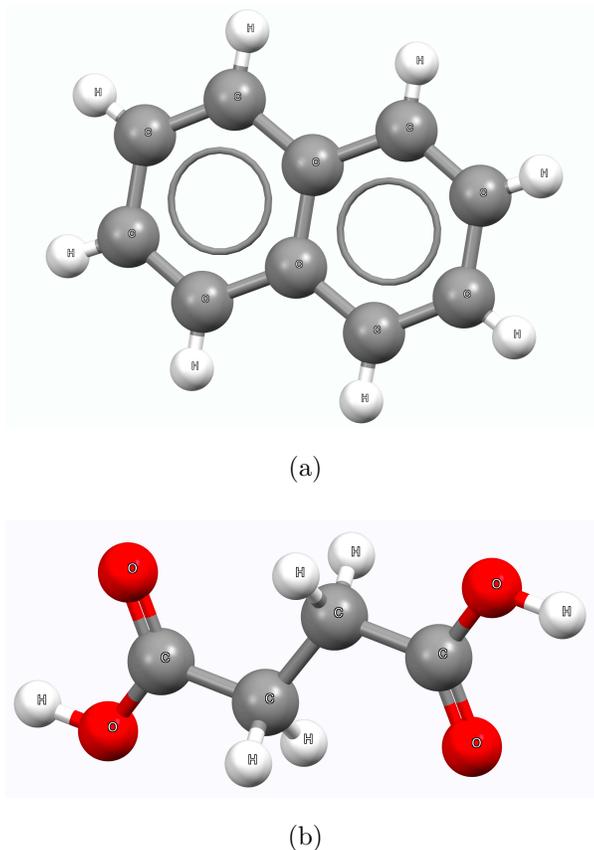


Figure 6.2: (a) Naphthalene molecule, (b) Succinic acid molecule

### 6.3.1 Solid Phase Simulations

For naphthalene, we begin with a naphthalene crystal structure from the Cambridge Crystal Structure Database (NAPHTA52).<sup>[20]</sup> It has a monoclinic space group  $P2_1/a$  with two molecules per unit cell and the following lattice parameters:  $a = 8.2128 \text{ \AA}$ ;  $b = 5.9727 \text{ \AA}$ ;  $c = 8.6745 \text{ \AA}$ ;  $\alpha = 90^\circ$ ;  $\beta = 123.388^\circ$ ;  $\gamma = 90^\circ$ . A  $5 \times 6 \times 4$  supercell is constructed consisting of 240 naphthalene molecules.

For  $\beta$ -succinic acid, we begin with the 300 K crystal structure from the Cambridge Crystal Structure Database (SUCACB02).<sup>[21]</sup> It has a monoclinic space group  $P2_1/c$  with two molecules per unit cell and the following lattice parameters:  $a = 5.519 \text{ \AA}$ ;

$b = 8.862 \text{ \AA}$ ;  $c = 5.101 \text{ \AA}$ ;  $\alpha = 90^\circ$ ;  $\beta = 91.59^\circ$ ;  $\gamma = 90^\circ$ . A  $5 \times 5 \times 5$  supercell is constructed consisting of 250 succinic acid molecules.

For  $\gamma$ -succinic acid, we begin with its recently discovered experimental crystal structure (SUCACB19).<sup>[22]</sup> It has a monoclinic space group  $C2/c$  with four molecules per unit cell and the following lattice parameters:  $a = 5.7015 \text{ \AA}$ ;  $b = 8.4154 \text{ \AA}$ ;  $c = 10.3538 \text{ \AA}$ ;  $\alpha = 90^\circ$ ;  $\beta = 90.374^\circ$ ;  $\gamma = 90^\circ$ . A  $6 \times 4 \times 3$  supercell is constructed consisting of 288 succinic acid molecules.

The simulation box size and atomic positions in the unit cell required for the NVT simulations are determined from an NPT simulation. All systems were equilibrated at their respective temperature and a pressure of 1 atm by simulating an NPT ensemble. A 10 ns equilibration run followed by a 10 ns production run was used to compute the average lattice parameters and atomic positions in the unit cell. The Lorentz-Berthelot mixing rules<sup>[23-25]</sup> were used for computing the interatomic pair coefficients. We used a time step of 0.5-1.0 fs. Nonbonded interactions were cutoff at 1 nm with standard tail corrections applied (see Appendix H.4 for details), with long range electrostatics handled by LAMMPS' Particle-Particle-Particle-Mesh (PPPM) summation.<sup>[26]</sup> The NPT ensemble was simulated using a Langevin thermostat and Nose-Hoover barostat<sup>[27]</sup> with  $P = 1 \text{ atm}$ . A 25 point Gauss Legendre quadrature method<sup>[28]</sup> was used to evaluate the integral in Equation 3.13. The NVT simulations were run for a total of 8 ns, of which the initial 3 ns were used for equilibration. Data every 1 ps was used to compute the thermodynamic averages.

The Einstein crystal method calculations have to be carried out at fixed center of mass for numerical reasons.<sup>[29, 30]</sup> Therefore, to constrain the center of mass we setup the simulation to have zero external force due to springs. To achieve this, we<sup>[31]</sup> i) scale the spring constants with atomic masses so that each oscillator has equal angular frequency,<sup>[32]</sup> ii) initialize the system with zero center of mass velocity, and, iii) initialize

each atom at its lattice position. We first set the spring constant for hydrogen atoms by matching the mean squared displacement and then scale it with atomic masses to get the spring constants for carbon such that  $k_{sp,i}/m_i = constant$ .<sup>[32, 33]</sup> The spring constants for the reference Einstein crystal in simulations at all temperatures are chosen to reproduce the mean square displacement of hydrogen atoms in the real crystal at 298K - 300K. For naphthalene, the spring constants for hydrogen and carbon were set to  $8.09 k_B T/\text{\AA}^2$ , and  $96.44 k_B T/\text{\AA}^2$ , respectively. For succinic acid, the spring constants for hydrogen, carbon, and oxygen were set to  $22 k_B T/\text{\AA}^2$ ,  $262.12 k_B T/\text{\AA}^2$ , and  $349.20 k_B T/\text{\AA}^2$ , respectively.

### 6.3.2 Gas Phase Simulations

For naphthalene, the centroid was constructed by tethering the hydrogens and carbons to the center of mass by springs having force constants  $16.786 k_B T/\text{\AA}^2$  and  $200 k_B T/\text{\AA}^2$ , respectively. For succinic acid, the centroid for succinic acid was constructed by tethering the hydrogens, oxygens, and carbons to the center of mass by springs having force constants  $1.26 k_B T/\text{\AA}^2$ ,  $20.0 k_B T/\text{\AA}^2$ , and  $15.01 k_B T/\text{\AA}^2$ , respectively.

The free energy of the centroid is computed using the procedure laid out in section 4.2. For adding each particle, the ratio  $Z_n/Z_n^o$  is computed using a custom Monte Carlo (MC) code. We run 100 million MC steps with a acceptance ratio between 30%-50%. Data every 1000 MC steps is used to compute thermodynamic averages.

The free energy difference  $\Delta a^{* \rightarrow G}$  was computed in 5 sets of stages (see Tables 6.1 6.2). All simulations were performed in an NVT ensemble with a Langevin thermostat. The simulation box was  $(100\text{\AA})^3$ . No long-range interactions were used in the gas phase simulations. The nonbonded cutoff (LJ and coulombic) was  $20\text{\AA}$ , much larger than the molecule size in order to include all pairwise interactions. A time step of 0.1-0.5 fs was used. For succinic acid, Stages 3,4,5 were simulated using Replica Exchange Molecular

Dynamics to sample the COOH dihedral. Eight replicas at  $T = 300$  K, 305 K, 310 K, 315 K, 330 K, 350 K, 500 K, 600 K were run with swaps attempted every 250-500 fs. Note, LAMMPS log files dump replica indices at the end of the time step after the swap, whereas the thermodynamic output is dumped at the end of the time step before the swap. The remaining simulation details specific to each set of stages are specified in Tables 6.1 and 6.2.

Table 6.1: Simulation details for computing  $\Delta a^i/k_B T$  for naphthalene

	$\beta\Delta a^i$	# of Gauss points	Equilibration period (ns)	Production period (ns)
1	centroid→bonds	95	10	100
2	Angles	30	10	90
3	Dihedrals	30	10	90
4	LJ Soft-core	45	10	90
5	Coulombic	15	10	90

Table 6.2: Simulation details for computing  $\Delta a^i/k_B T$  for succinic acid

	$\beta\Delta a^i$	# of Gauss points	Equilibration period (ns)	Production period (ns)
1	centroid→bonds	75	25	75
2	Angles	30	5	195
3	Dihedrals	30	5	195
4	LJ Soft-core	30	1	399
5	Coulombic	95	5	20

## 6.4 Results and Discussion

Tables 6.3 and 6.4 show the results of the dimensionless solid phase free energies computed at different temperatures for naphthalene and  $\beta$ -succinic acid, respectively. Tables 6.5 and 6.6 show their dimensionless intra-centroid free energies. Note, the computation of these free energies requires simulations described in Figure 4.3 to be performed only at any one temperature. This is because we scale the spring constants of the centroid with temperature. Tables 6.7 and 6.8 show the results of transformation free energies of the centroid to a molecule for naphthalene and succinic acid, respectively. Using  $a^C(T, P^{sim})$ ,  $\Delta a^{* \rightarrow G}(T)$ , and  $a^{*,intra}(T)$  obtained from our simulations we compute the volume ( $v^G$ ) using equation 6.1b. We then use the ideal gas equation of state to compute the vapor pressure. Figures 6.3 and 6.4 show the driving forces for crystallization—from the vapor phase—and the vapor pressures for naphthalene and  $\beta$ -succinic acid, respectively. The vapor pressures at different temperatures for the two systems are reported in Tables 6.9 and 6.10.

## Solid Free Energy Results

Table 6.3: Absolute dimensionless Helmholtz free energies of solid naphthalene

T (K)	$\beta a^{EC}$	$\beta \Delta a^{\delta_1}$	$\beta \Delta a^{EC \rightarrow C}$	$\beta \Delta a^{\delta_2}$	$\beta a^C$
298.00	7.217	-0.066	$53.485 \pm 0.003$	-0.022	$60.615 \pm 0.003$
308.17	6.311	-0.066	$53.013 \pm 0.003$	-0.022	$59.237 \pm 0.003$
318.17	5.449	-0.066	$52.552 \pm 0.003$	-0.022	$57.913 \pm 0.003$
333.34	4.192	-0.066	$51.844 \pm 0.003$	-0.022	$55.947 \pm 0.003$
where: <sup>[31]</sup> $\Delta a^{\delta_1} = a_{COM}^{EC} - a^{EC}$ ; $\Delta a^{\delta_2} = a^C - a_{COM}^C$					

Table 6.4: Absolute dimensionless Helmholtz free energies of solid  $\beta$ -succinic acid

T (K)	$\beta a^{EC}$	$\beta \Delta a^{\delta_1}$	$\beta \Delta a^{EC \rightarrow C}$	$\beta \Delta a^{\delta_2}$	$\beta a^C$
300	26.482	-0.069	$-171.822 \pm 0.005$	-0.020	$-145.429 \pm 0.005$
305	26.135	-0.069	$-169.031 \pm 0.005$	-0.019	$-142.985 \pm 0.005$
310	25.793	-0.069	$-166.325 \pm 0.005$	-0.020	$-140.621 \pm 0.005$
315	25.457	-0.069	$-163.704 \pm 0.004$	-0.020	$-138.336 \pm 0.004$
330	24.480	-0.069	$-156.381 \pm 0.003$	-0.020	$-131.990 \pm 0.003$
350	23.245	-0.069	$-147.639 \pm 0.005$	-0.020	$-124.483 \pm 0.005$
where: <sup>[31]</sup> $\Delta a^{\delta_1} = a_{COM}^{EC} - a^{EC}$ ; $\Delta a^{\delta_2} = a^C - a_{COM}^C$					

**Intra-centroid Free Energies**

Table 6.5: Intra-centroid dimensionless Helmholtz free energies for naphthalene

T (K)	$\beta a^{*,intra}$
298.00	$21.87 \pm 0.01$
308.17	$21.02 \pm 0.01$
318.17	$20.20 \pm 0.01$
333.34	$19.02 \pm 0.01$

Table 6.6: Intra-centroid dimensionless Helmholtz free energies for succinic acid

T (K)	$\beta a^{*,intra}$
300	$55.62 \pm 0.01$
305	$55.30 \pm 0.01$
310	$54.98 \pm 0.01$
315	$54.67 \pm 0.01$
330	$53.76 \pm 0.01$
350	$52.61 \pm 0.01$

Table 6.7: Helmholtz free energy to transform the centroid to naphthalene

T (K)	$\beta\Delta a^{* \rightarrow B}$	$\beta\Delta a^{angles}$	$\beta\Delta a^{dihedrals}$	$\beta\Delta a^{LJ}$	$\beta\Delta a^{coul}$	$\beta\Delta a^{* \rightarrow G}$
298.00	$-50.25 \pm 0.05$	$67.52 \pm 0.03$	$24.62 \pm 0.01$	$16.93 \pm 0.00$	$0.99 \pm 0.00$	$59.82 \pm 0.06$
308.17	$-50.50 \pm 0.05$	$67.16 \pm 0.03$	$24.50 \pm 0.01$	$16.40 \pm 0.00$	$0.96 \pm 0.00$	$58.52 \pm 0.05$
318.17	$-50.77 \pm 0.05$	$66.82 \pm 0.03$	$24.38 \pm 0.01$	$15.90 \pm 0.00$	$0.93 \pm 0.00$	$57.26 \pm 0.05$
333.34	$-51.17 \pm 0.05$	$66.31 \pm 0.03$	$24.22 \pm 0.01$	$15.22 \pm 0.00$	$0.89 \pm 0.00$	$55.46 \pm 0.05$

Table 6.8: Helmholtz free energy to transform the centroid to succinic acid

T (K)	$\beta\Delta a^{* \rightarrow B}$	$\beta\Delta a^{angles}$	$\beta\Delta a^{dihedrals}$	$\beta\Delta a^{L,J}$	$\beta\Delta a^{coul}$	$\beta\Delta a^{* \rightarrow G}$
300.00	3.49 ± 0.03	47.31 ± 0.01	13.041 ± 0.001	2.761 ± 0.001	-143.561 ± 0.003	-76.97 ± 0.03
305.00	3.39 ± 0.02	47.16 ± 0.01	12.931 ± 0.001	2.738 ± 0.001	-141.222 ± 0.003	-75.01 ± 0.02
310.00	3.28 ± 0.02	47.03 ± 0.01	12.825 ± 0.001	2.716 ± 0.001	-138.941 ± 0.003	-73.09 ± 0.02
315.00	3.17 ± 0.02	46.88 ± 0.01	12.722 ± 0.001	2.696 ± 0.001	-136.750 ± 0.003	-71.29 ± 0.02
330.00	2.88 ± 0.02	46.50 ± 0.01	12.427 ± 0.001	2.636 ± 0.001	-130.550 ± 0.003	-66.11 ± 0.02
350.00	2.50 ± 0.02	46.00 ± 0.01	12.055 ± 0.001	2.567 ± 0.001	-123.110 ± 0.002	-59.99 ± 0.02

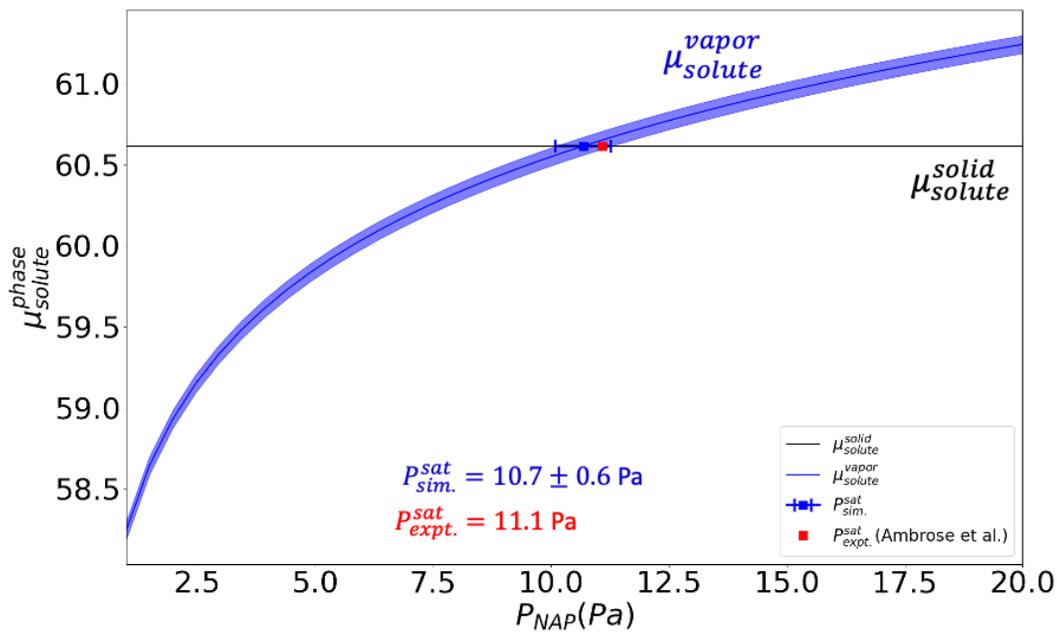


Figure 6.3: The saturation pressure and driving force for crystallization of naphthalene from vapor at 298 K.

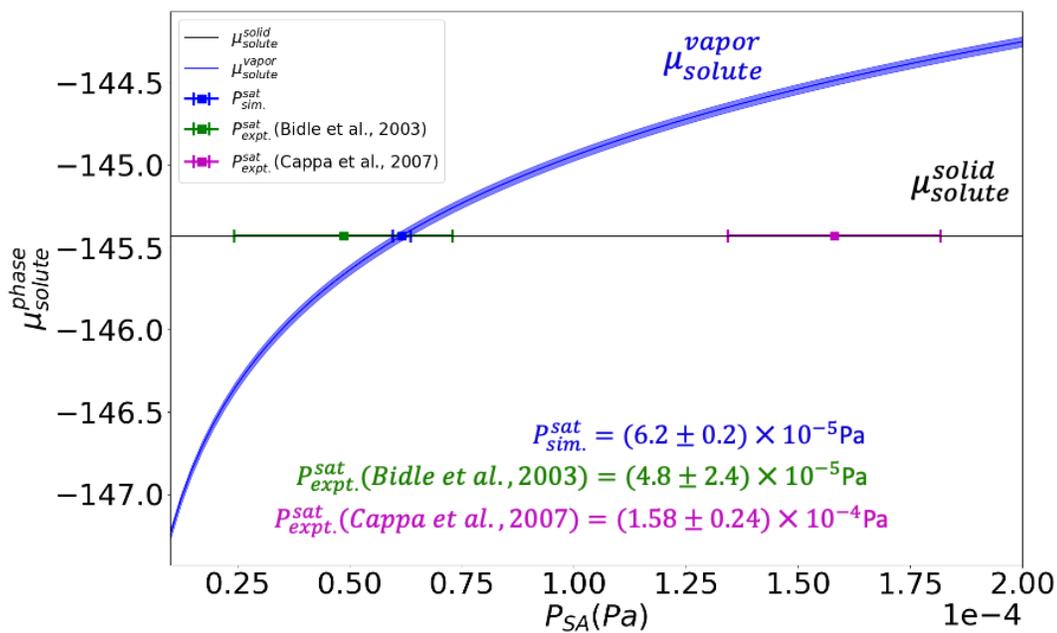


Figure 6.4: The saturation pressure and driving force for crystallization of  $\beta$ -succinic acid from vapor at 300 K.

Table 6.9: Vapor pressure of naphthalene at four different temperatures

T (K)	$P^{sat}$ (mPa)
298.00	$10.7 \pm 0.6$
308.17	$25 \pm 1$
318.17	$58 \pm 3$
333.34	$181 \pm 10$

Table 6.10: Vapor pressure of  $\beta$ -succinic acid at six different temperatures

T (K)	$P^{sat}$ (mPa)
300	$0.062 \pm 0.002$
305	$0.144 \pm 0.003$
310	$0.323 \pm 0.007$
315	$0.74 \pm 0.02$
330	$6.7 \pm 0.2$
350	$97 \pm 2$

Having computed the vapor pressures at various temperature, we plot the natural logarithm of vapor pressure against inverse temperature, i.e., construct a Clausius-Clapeyron plot from the vapor pressure data for each molecule. As seen in Figures 6.5 and 6.6 the simulation points align excellently on a straight line with a regression coefficient of 0.99. The slopes of the Clausius-Clapeyron plots give a sublimation enthalpy ( $\Delta H_{sub}$ ) equal to  $15.8 \pm 0.4$  kcal/mol and  $30.7 \pm 0.1$  kcal/mol for naphthalene and succinic acid, respectively. Direct calculations of sublimation enthalpy using solid and gas phase potential energies (see Appendix F for calculation details) give a sublimation enthalpy of  $16.9 \pm 0.0$  and  $30.7 \pm 0.0$  kcal/mol for naphthalene and succinic acid, respectively, which are in good

agreement with the estimate we get from the Clausius-Clapeyron plots' slopes. This is a self consistency check, as the direct route uses the same force fields as employed in the vapor pressure predictions. The experimental values of sublimation enthalpies ( $\Delta H_{sub}^{expt}$ ) are in agreement with our predictions, too.

We also plot experimental vapor pressure data by Ambrose et al.<sup>[34]</sup> for naphthalene and by Bilde et al.,<sup>[35]</sup> Cappa et al.<sup>[36]</sup> and Saleh et al.<sup>[37, 38]</sup> for  $\beta$ -succinic acid. For  $\beta$ -succinic acid, since the vapor pressures being measured are extremely low ( $\sim$  mPa), their measurement is very difficult. This can be seen from the spread in experimental data between different research groups and the uncertainty in the respective data points. Our vapor pressure predictions lie within the spread of the experiments<sup>[38]</sup> for  $\beta$ -succinic acid and in close agreement with experiments for naphthalene. Note that the vapor pressure predictions exponentially magnify small errors in the free energy calculations (see Equation 6.1b). Approximately, a 0.05% error in absolute free energy estimates translates to a 10% error in vapor pressure predictions. Thus, we can conclude from the solid-vapor equilibrium predictions that OPLS and GAFF accurately capture the solute-solute interactions for naphthalene and  $\beta$ -succinic acid, respectively.

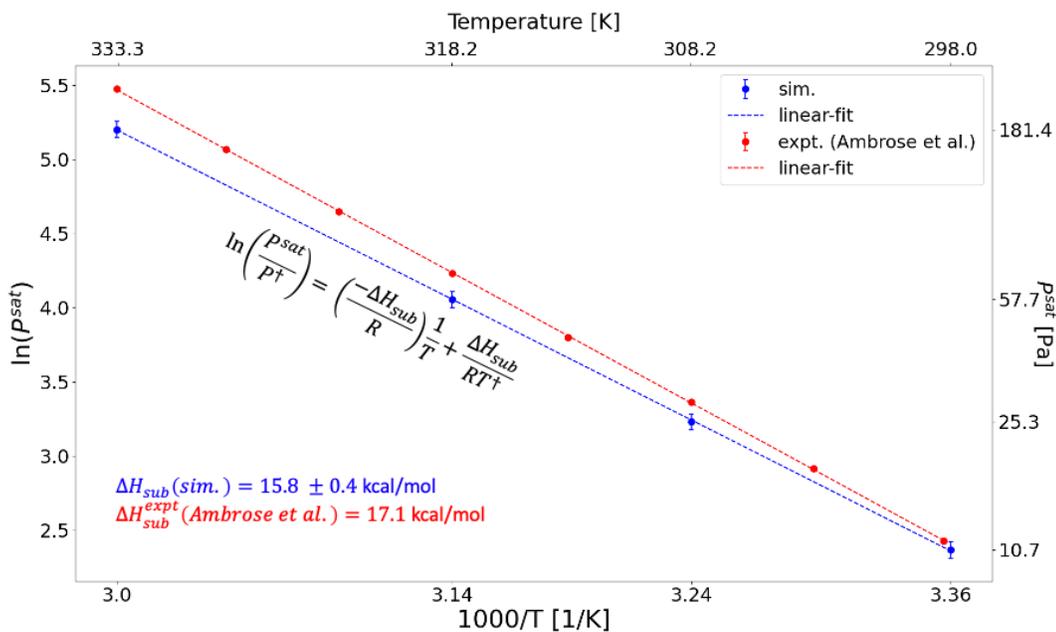


Figure 6.5: The Clausius-Clapeyron plot of naphthalene's sublimation vapor pressure and temperature, where  $P^\dagger = 1$  Pa, and  $T^\dagger$  is the corresponding sublimation temperature ; the secondary axis tick labels are the simulation data points

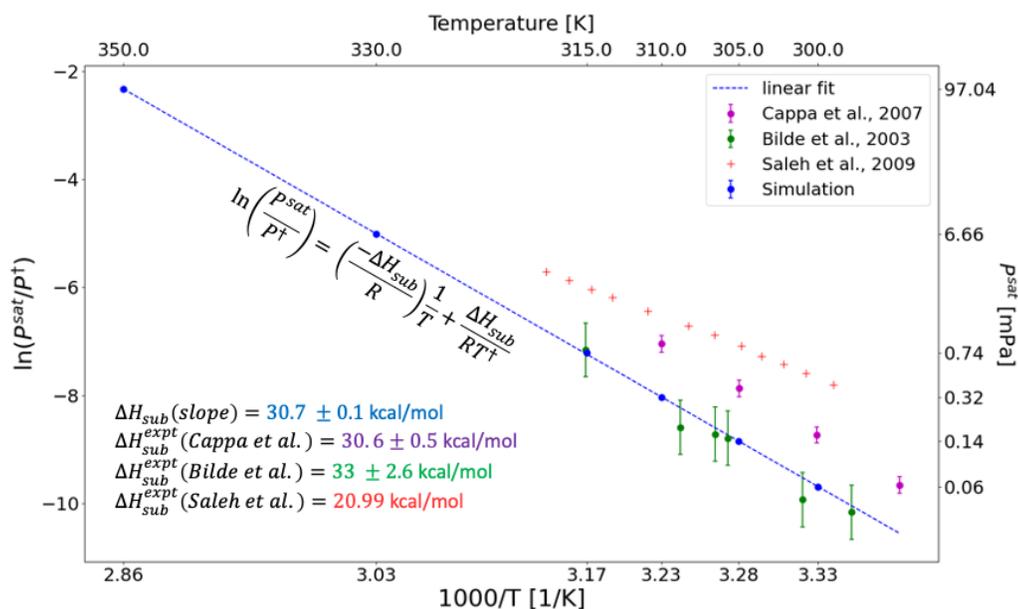


Figure 6.6: The Clausius-Clapeyron plot of  $\beta$ -succinic acid's sublimation vapor pressure and temperature, where  $P^\dagger = 1$  Pa, and  $T^\dagger$  is the corresponding sublimation temperature; the secondary axis tick labels are the simulation data points

## 6.5 Predicting Solid-Solution Equilibrium

At solid-solution equilibrium we have

$$\mu^C(T, P) = \mu_{solute}^{soln}(T, P, x_{sat}) \quad (6.2a)$$

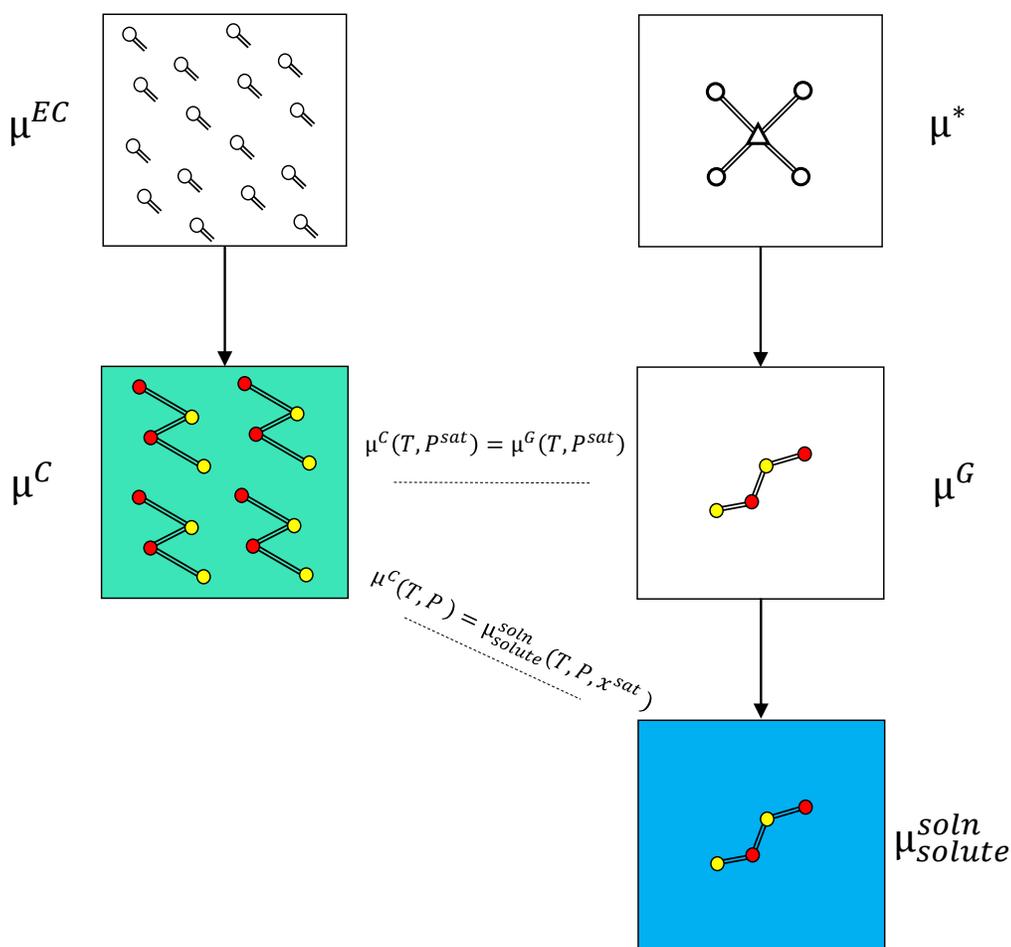


Figure 6.7: Computing the absolute chemical potentials of each phase. The left and right starting points are absolute chemical potentials of the Einstein crystal and Centroid, respectively. For computing the chemical potential in the solution, we need to perform additional solvation free energy calculations.

$$\begin{aligned}
a^C(T, v^C) &= a^*(T, v^{soln}(x_{sat})) + \Delta a^{* \rightarrow G}(T) + \Delta a^{sym} + \Delta G_{solv}(x_{sat}) \\
&= -k_B T \ln \left( \frac{v^{soln}(x_{sat})}{\Lambda_1^3} \right) + a^{*,intra}(T) + \Delta a^{* \rightarrow G}(T) + \Delta a^{sym} + \Delta G_{solv}(x_{sat})
\end{aligned}
\tag{6.2b}$$

As we can see from Equation 6.2b we just need additional solvation free energy computations at different concentrations to obtain the chemical potential of the solute in solution. We harness the solvation free energy computation tools that we developed in chapter 5 to compute the solvation free energy of naphthalene and succinic acid in aqueous solutions.

### 6.5.1 Simulation Details

We perform hydration free energy calculations using the decoupling approach as laid out in chapter 5. Naphthalene as described by the OPLS-AA<sup>[14]</sup> force field was solvated in 864 SPC water molecules.<sup>[39]</sup> Succinic acid as described by the GAFF<sup>[15]</sup> forcefield was solvated in 864 SPCE water molecules.<sup>[40]</sup>

The Lorentz-Berthelot mixing rules<sup>[23-25]</sup> were used for computing the interatomic pair coefficients. The initial configurations were generated using packmol.<sup>[41]</sup> We used a time step of 0.5 fs. Nonbonded interactions were cutoff at 1 nm, with long-range electrostatics handled by OpenMM’s Particle Mesh Ewald (PME) summation<sup>[42]</sup>. Long-range dispersion corrections to energy and pressure were applied.

For naphthalene, the NPT ensemble was simulated using Langevin thermostat and a Monte-Carlo barostat at  $T = 298.0$  K and  $P = 1.01325$  bar. All systems were equilibrated for 4 ns followed by 8 ns production runs. A 15 point Gauss-Legendre quadrature method<sup>[7]</sup> was used to evaluate the integral in equation 5.2 for computing the coulombic

solvation free energy. For computing the solvation free energy due to the LJ interactions we used a 45 point Gauss-Legendre quadrature. Data every 2 ps was used to compute the thermodynamic averages.

For succinic acid, to compute the coulombic solvation free energy a 5 point Gauss-Legendre quadrature was employed. Replica exchange molecular dynamics (REMD) simulations were performed to sample the COOH dihedral (see Appendix J). A total of 22 replicas spanning temperatures 300K - 470K were used. Since, the temperatures employed in the REMD simulations were above the boiling point of water at 1 atm, we approximated the NPT ensemble by simulating a NVT ensemble using the average of the box volume computed at 300K, 1 atm pressure, and  $N_{solute} = 0, 1$ . Each replica was run for 18 ns. The first 8 ns were discarded as equilibration burn, and remaining 10 ns of production run data was sampled every 2 ps to compute thermodynamic averages. For computing the solvation free energy due to the LJ interactions we used a 45 point Gauss-Legendre quadrature. We discarded parts of the trajectory where the COOH dihedral flipped to the trans position by chance as we know from the COOH PMF that it prefers to be in the cis conformation until it is coupled with coulombic interactions with water. All simulations for the LJ part were run for a total of 8 ns, of which the first 4 ns of thermodynamic data was discarded as a part of equilibration burn. Data every 2 ps was sampled to compute thermodynamic averages.

## 6.5.2 Results and Discussion

Tables 6.11 and 6.12 show the results for the dimensionless solvation free energies computed at different temperatures for naphthalene and succinic acid, respectively. Next, using Equation 6.2b, the dimensionless free energy results for the solid, gas and solution phases (Tables 6.3, 6.7, and 6.11, for naphthalene and Tables 6.4, 6.8, and 6.12, for

succinic acid) we generate the solubility plots as shown in Figures 6.8 and 6.9. These plots make use of two approximations for  $x_{NAP} \leq 1e-4$  and  $x_{SA} \leq 0.05$ : i) the volume of mixing is zero, and ii)  $\Delta G_{solv}(x) \approx \text{constant}$ , i.e., the normalized activity coefficients ( $\gamma_{solute}/\gamma_{solute}^\infty$ ) is ca. 1. See Appendix G for details on normalized activity coefficients. The solubility prediction of naphthalene,  $(3.66 \pm 0.2) \times 10^{-6}$  is in good agreement with the predictions of Li et al.,<sup>[1]</sup>  $(4.74 \pm 0.8) \times 10^{-6}$ , and the experimental value of  $4.4 \times 10^{-6}$ . Thus, successfully testing the solubility framework. The predictions for  $\beta$ -succinic acid are off from the experimental values due to  $\Delta\mu_{solution \rightarrow solid} \sim 1 k_B T$ . This difference is less than 1% of the free energies being computed. However, due to exponentiation, the solubility prediction deviate substantially from experimental measurements. Nonetheless, as is discussed in the later results, the solubility trend with temperature is captured well by the employed force fields, i.e. GAFF + SPCE.

Table 6.11: Dimensionless solvation free energies of naphthalene (OPLS) in water (SPC)

$T$ (K)	$\beta\Delta G_{LJ}$	$\beta\Delta G_{coul}$	$\beta\Delta G_{solv}$
298.00	$1.83 \pm 0.05$	$-4.65 \pm 0.01$	$-2.82 \pm 0.05$
308.17	$2.14 \pm 0.05$	$-4.49 \pm 0.01$	$-2.35 \pm 0.05$
318.17	$2.33 \pm 0.05$	$-4.36 \pm 0.01$	$-2.03 \pm 0.05$
333.34	$2.71 \pm 0.04$	$-4.14 \pm 0.01$	$-1.43 \pm 0.05$

Table 6.12: Dimensionless solvation free energies of succinic acid (GAFF) in water (SPCE)

$T$ (K)	$\beta\Delta G_{LJ}$	$\beta\Delta G_{coul}$	$\beta\Delta G_{solv}$
300.00	$1.36 \pm 0.05$	$-25.58 \pm 0.06$	$-24.22 \pm 0.08$
305.00	$1.58 \pm 0.06$	$-25.18 \pm 0.06$	$-23.60 \pm 0.08$
310.00	$1.59 \pm 0.06$	$-24.85 \pm 0.06$	$-23.26 \pm 0.08$
315.00	$1.94 \pm 0.06$	$-24.39 \pm 0.06$	$-22.45 \pm 0.08$
330.00	$2.27 \pm 0.06$	$-23.27 \pm 0.06$	$-21.00 \pm 0.08$

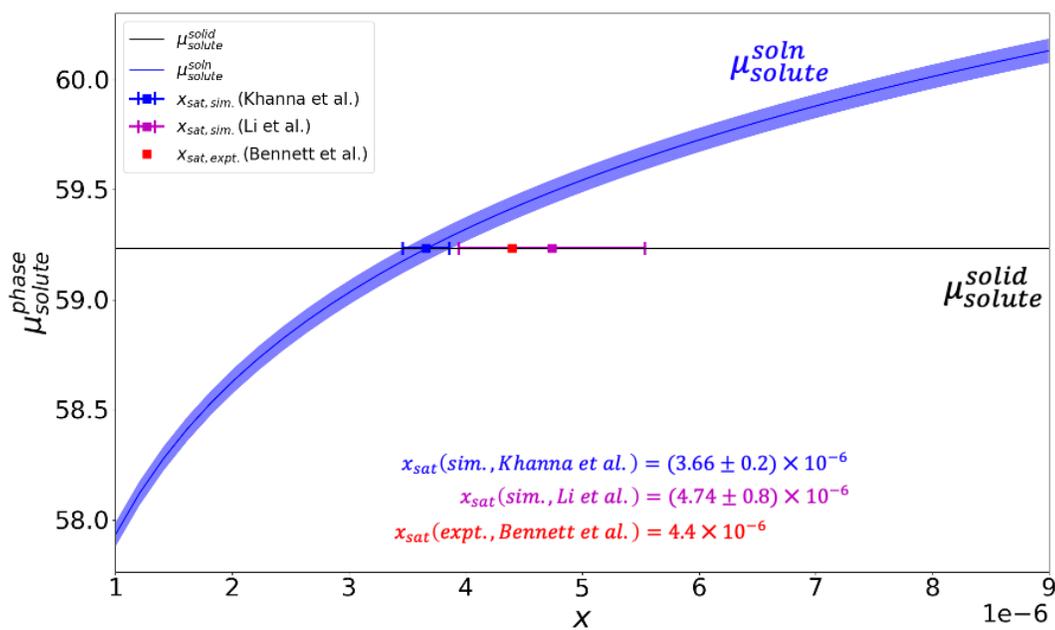


Figure 6.8: The solubility plot for naphthalene in water. This plot shows the driving forces for crystallization as predicted by the employed force fields, i.e., OPLS + SPC

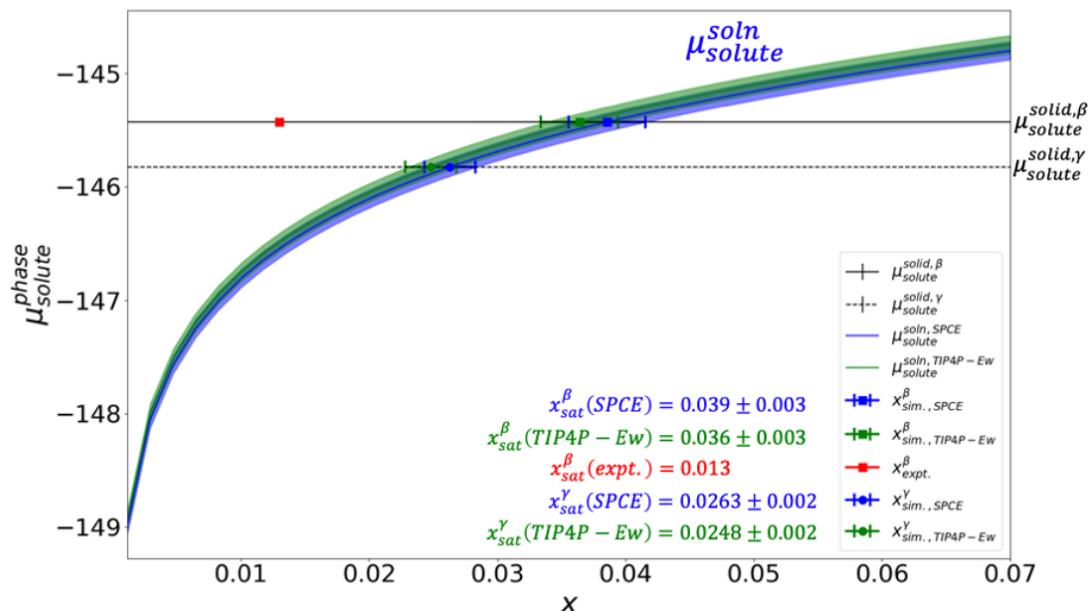


Figure 6.9: The solubility plot for ( $\beta$  and  $\gamma$ )-succinic acid in water. This plot shows the driving forces for crystallization as predicted by the employed force fields, i.e., GAFF + SPCE and GAFF + TIP4P-Ew.

We further compute the solubility of our two model compounds at different temperatures and plot the results on van't Hoff plots ( $\ln(x_{\text{sat}})$  vs.  $1/T$ ) to visualize the solubility trend and compute the enthalpy of dissolution—given by the slope of a van't Hoff plot. The solubility predictions lie on a straight line bolstering the results (see Figures 6.10 and 6.11). The enthalpies of dissolution for naphthalene and  $\beta$ -succinic acid are  $7.8 \pm 0.4$  kcal/mol, and  $8.3 \pm 0.7$  kcal/mol, respectively—which are in good agreement with experimental values of 7.1 kcal/mol and 7.6 kcal/mol, respectively.

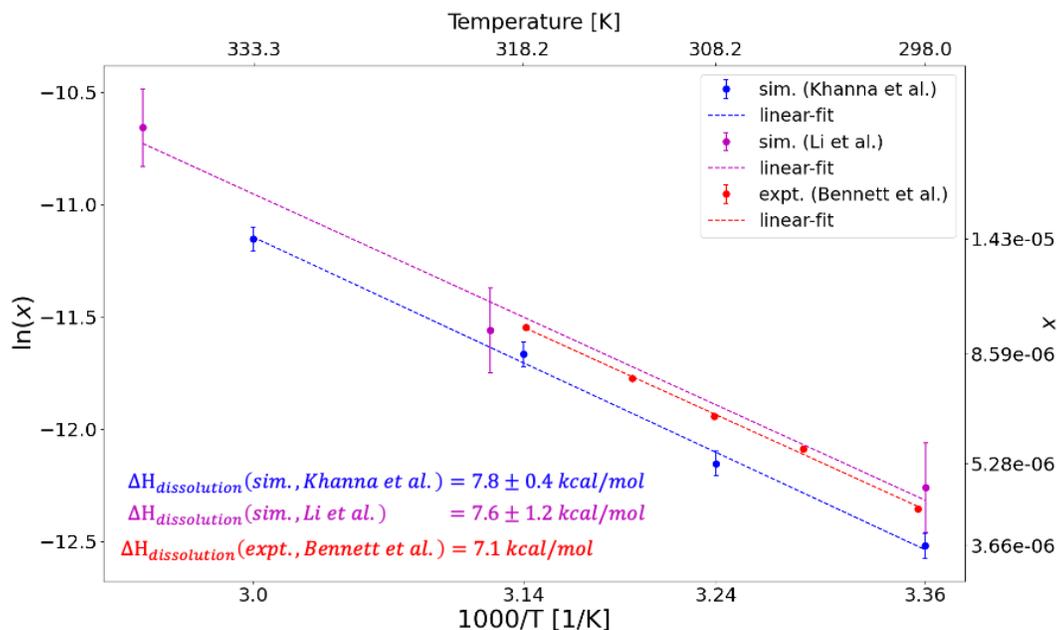


Figure 6.10: The van't Hoff plot for naphthalene (OPLS-AA) in water (SPC).

However, the true power of a prediction can be seen from the results of the  $\gamma$ -polymorph of succinic acid. We predict a vapor pressure of  $0.0142 \pm 0.001$  mPa (using GAFF) and a solubility of  $0.0263 \pm 0.002$  (using GAFF and SPCE) for the  $\gamma$  polymorph—whose vapor pressure and solubility has not yet been measured experimentally! In fact, the  $\gamma$  polymorph has not been synthesized since its first and only appearance as a concomitant with the stable  $\beta$  form during an attempted purification of a peptide by cocrystallization.<sup>[22]</sup> The solid free energy results shown in Figure 6.9 indicate that the  $\gamma$  and  $\beta$  form are close in stability to one another—at 300 K—with the  $\gamma$  form being more stable by  $0.39 k_B T$ /molecule. The only concern with this prediction is—the GAFF force field doesn't accurately predict the crystal structure of the  $\gamma$  form (as shown in Appendix E). Therefore, the true accuracy of these predictions would be unveiled by future experiments which successfully synthesize the  $\gamma$  polymorph and measure its solubility.

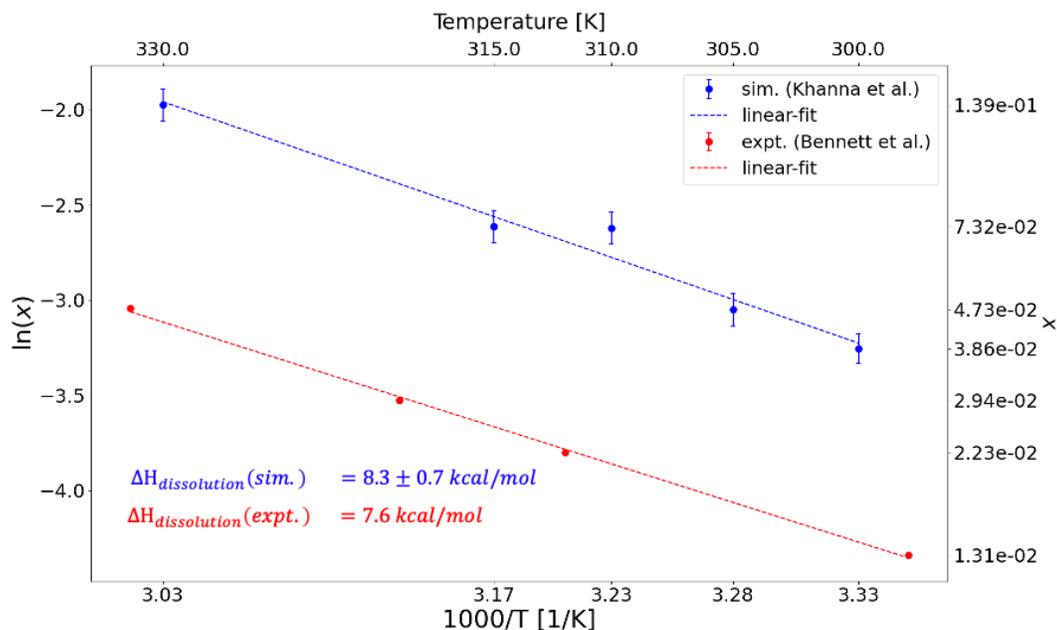


Figure 6.11: The van't Hoff plot for  $\beta$ -succinic acid (GAFF) in water (SPCE).

## 6.6 Conclusion

We have developed a decoupled route to compute fluid-solid equilibria avoiding the fluid-to-solid transformation steps. The decoupled approach requires the computation of *absolute* chemical potential of the solute in the two phases. For the gas phase, this is done by introducing a gas phase “centroid” reference system with the same number of vibrational, rotational and translational degrees of freedom as the real molecule. This enables us to accurately compute *absolute* free energies of ‘floppy’ polyatomic molecules. For the solid phase, the absolute chemical potential is computed by capturing the intramolecular free energy via the Frenkel-Ladd method. For the gas and solid free energy calculations, we use simple thermodynamic perturbation and thermodynamic integration methods. We compute sublimation vapor pressures,  $P^{sat}$ , in temperature ranges 298K-333K, and 300K-350K—for naphthalene and succinic acid—to represent the solid-vapor

equilibrium.

The normal procedure to validate the solid state force field includes computing the lattice energies and comparing it with experiments.<sup>[43, 44]</sup> But this test has no entropic, i.e, free energy contribution. The solid-vapor equilibrium prediction for  $P^{sat}$  does encompass the entropic information. Thus, the calculations used in this work can aid efforts to test and develop force fields. Further, with additional solvation free energy calculations we leverage the absolute chemical potential route to predict the solubilities—of naphthalene and  $\beta$ -succinic acid in aqueous solutions at various temperatures—and thus compute driving forces for crystallization from solution, providing a crucial piece of calculation for a complete and self-consistent in-silico absolute growth rate prediction of organic crystals. Further, this route especially offers advantages when a molecules requires advanced sampling in the gas phase, and driving forces for crystallization of multiple polymorphs is desired. The decoupled route requires a single set of gas phase simulations requiring advanced sampling compared to multiple sets of fluid-to-solid transformation simulations (requiring advanced sampling) for spanning all polymorphs.

# Bibliography

- [1] L. Li, T. Totton, and D. Frenkel, Computational methodology for solubility prediction: Application to the sparingly soluble solutes, *The Journal of Chemical Physics* **146** (2017), no. 21 214110.
- [2] M. A. Bellucci, G. Gobbo, T. K. Wijethunga, G. Ciccotti, and B. L. Trout, Solubility of paracetamol in ethanol by molecular dynamics using the extended Einstein crystal method and experiments, *The Journal of Chemical Physics* **150** (2019), no. 9 094107.
- [3] M. Parrinello and A. Rahman, Study of an f center in molten kcl, *The Journal of chemical physics* **80** (1984), no. 2 860–867.
- [4] S. Habershon, D. E. Manolopoulos, T. E. Markland, and T. F. Miller III, Ring-polymer molecular dynamics: Quantum effects in chemical dynamics from classical trajectories in an extended phase space, *Annual review of physical chemistry* **64** (2013) 387–413.
- [5] L. Pereyaslavets, I. Kurnikov, G. Kamath, O. Butin, A. Illarionov, I. Leontyev, M. Olevanov, M. Levitt, R. D. Kornberg, and B. Fain, On the importance of accounting for nuclear quantum effects in ab initio calibrated force fields in biological simulations, *Proceedings of the National Academy of Sciences* **115** (2018), no. 36 8878–8882.
- [6] T. Fröhlking, M. Bernetti, N. Calonaci, and G. Bussi, Toward empirical force fields that match experimental observables, *The Journal of chemical physics* **152** (2020), no. 23 230902.
- [7] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman, A second generation force field for the simulation of proteins, nucleic acids, and organic molecules, *Journal of the American Chemical Society* **117** (1995), no. 19 5179–5197.
- [8] A. D. MacKerell Jr, J. Wiorkiewicz-Kuczera, and M. Karplus, An all-atom empirical energy function for the simulation of nucleic acids, *Journal of the American Chemical society* **117** (1995), no. 48 11946–11975.
- [9] W. L. Jorgensen and J. Tirado-Rives, The opls [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin, *Journal of the American Chemical Society* **110** (1988), no. 6 1657–1666.

## BIBLIOGRAPHY

---

- [10] C. Oostenbrink, A. Villa, A. E. Mark, and W. F. Van Gunsteren, A biomolecular force field based on the free enthalpy of hydration and solvation: the gromos force-field parameter sets 53a5 and 53a6, *Journal of computational chemistry* **25** (2004), no. 13 1656–1676.
- [11] V. Babin, C. Leforestier, and F. Paesani, Development of a “first principles” water potential with flexible monomers: Dimer potential energy surface, vrt spectrum, and second virial coefficient, *Journal of chemical theory and computation* **9** (2013), no. 12 5395–5403.
- [12] V. Babin, G. R. Medders, and F. Paesani, Development of a “first principles” water potential with flexible monomers. ii: Trimer potential energy surface, third virial coefficient, and small clusters, *Journal of chemical theory and computation* **10** (2014), no. 4 1599–1607.
- [13] G. R. Medders, V. Babin, and F. Paesani, Development of a “first-principles” water potential with flexible monomers. iii. liquid phase properties, *Journal of chemical theory and computation* **10** (2014), no. 8 2906–2910.
- [14] W. L. Jorgensen, D. S. Maxwell, and J. Tirado-Rives, Development and testing of the opls all-atom force field on conformational energetics and properties of organic liquids, *Journal of the American Chemical Society* **118** (1996), no. 45 11225–11236.
- [15] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, and D. A. Case, Development and testing of a general amber force field, *Journal of Computational Chemistry* **25** (2004), no. 9 1157–1174.
- [16] A. Jakalian, B. L. Bush, D. B. Jack, and C. I. Bayly, Fast, efficient generation of high-quality atomic charges. am1-bcc model: I. method, *Journal of Computational Chemistry* **21** (2000), no. 2 132–146.
- [17] A. Jakalian, D. B. Jack, and C. I. Bayly, Fast, efficient generation of high-quality atomic charges. am1-bcc model: Ii. parameterization and validation, *Journal of Computational Chemistry* **23** (2002), no. 16 1623–1641.
- [18] S. Plimpton, Fast parallel algorithms for short-range molecular dynamics, *Journal of Computational Physics* **117** (1995), no. 1 1–19.
- [19] P. Eastman, J. Swails, J. D. Chodera, R. T. McGibbon, Y. Zhao, K. A. Beauchamp, L.-P. Wang, A. C. Simmonett, M. P. Harrigan, C. D. Stern, *et. al.*, Openmm 7: Rapid development of high performance algorithms for molecular dynamics, *PLoS computational biology* **13** (2017), no. 7 e1005659.
- [20] C. P. Brock and J. D. Dunitz, Temperature dependence of thermal motion in crystalline naphthalene, *Acta Crystallographica Section B: Structural Crystallography and Crystal Chemistry* **38** (1982), no. 8 2218–2228.
- [21] J.-L. Leviel, G. Auvert, and J.-M. Savariault, Hydrogen bond studies. a neutron diffraction study of the structures of succinic acid at 300 and 77 k, *Acta Crystallographica Section B: Structural Crystallography and Crystal Chemistry* **37** (1981), no. 12 2185–2189.

## BIBLIOGRAPHY

---

- [22] P. Lucaioli, E. Nauha, I. Gimondi, L. S. Price, R. Guo, L. Iuzzolino, I. Singh, M. Salvalaglio, S. L. Price, and N. Blagden, Serendipitous isolation of a disappearing conformational polymorph of succinic acid challenges computational polymorph prediction, *CrystEngComm* **20** (2018), no. 28 3971–3977.
- [23] D. Berthelot *Compt. Rendus* **126** (1898) 1703–1706.
- [24] H. Lorentz *Annalen der physik* **248** (1881), no. 1 127–136.
- [25] M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids*. Oxford University Press, 2017.
- [26] P. P. Ewald, Die berechnung optischer und elektrostatischer gitterpotentiale, *Annalen der physik* **369** (1921), no. 3 253–287.
- [27] G. J. Martyna, D. J. Tobias, and M. L. Klein, Constant pressure molecular dynamics algorithms, *The Journal of Chemical Physics* **101** (1994), no. 5 4177–4189.
- [28] M. Abramowitz and I. A. Stegun, *Handbook of mathematical functions: with formulas, graphs, and mathematical tables*, vol. 55. Courier Corporation, 1965.
- [29] D. Frenkel and A. J. Ladd, New monte carlo method to compute the free energy of arbitrary solids. application to the fcc and hcp phases of hard spheres, *The Journal of Chemical Physics* **81** (1984), no. 7 3188–3193.
- [30] J. M. Polson, E. Trizac, S. Pronk, and D. Frenkel, Finite-size corrections to the free energies of crystalline solids, *The Journal of Chemical Physics* **112** (2000), no. 12 5339–5342.
- [31] V. Khanna, J. Anwar, D. Frenkel, M. F. Doherty, and B. Peters, Free energies of crystals computed using einstein crystal with fixed center of mass and differing spring constants, *The Journal of Chemical Physics* **154** (2021), no. 16 164509.
- [32] J. Anwar, D. Frenkel, and M. G. Noro, Calculation of the melting point of NaCl by molecular simulation, *The Journal of Chemical Physics* **118** (2003), no. 2 728–735.
- [33] D. Frenkel and B. Smit, *From algorithms to applications*, 1996.
- [34] D. Ambrose, I. Lawrenson, and C. Sprake, The vapour pressure of naphthalene, *The Journal of Chemical Thermodynamics* **7** (1975), no. 12 1173–1176.
- [35] M. Bilde, B. Svenningsson, J. Mønster, and T. Rosenørn, Even- odd alternation of evaporation rates and vapor pressures of c3- c9 dicarboxylic acid aerosols, *Environmental Science & Technology* **37** (2003), no. 7 1371–1378.
- [36] C. D. Cappa, E. R. Lovejoy, and A. Ravishankara, Determination of evaporation rates and vapor pressures of very low volatility compounds: a study of the c4- c10 and c12 dicarboxylic acids, *The Journal of Physical Chemistry A* **111** (2007), no. 16 3099–3109.
- [37] R. Saleh, A. Shihadeh, and A. Khlystov, Determination of evaporation coefficients of semi-volatile organic aerosols using an integrated volume—tandem differential

## BIBLIOGRAPHY

---

- mobility analysis (iv-tdma) method, *Journal of Aerosol Science* **40** (2009), no. 12 1019–1029.
- [38] M. Bilde, K. Barsanti, M. Booth, C. D. Cappa, N. M. Donahue, E. U. Emanuelsson, G. McFiggans, U. K. Krieger, C. Marcolli, D. Topping, *et. al.*, Saturation vapor pressures and transition enthalpies of low-volatility organic molecules of atmospheric relevance: from dicarboxylic acids to complex mixtures, *Chemical reviews* **115** (2015), no. 10 4115–4156.
- [39] H. J. Berendsen, J. P. Postma, W. F. van Gunsteren, and J. Hermans, Interaction models for water in relation to protein hydration, in *Intermolecular forces*, pp. 331–342. Springer, 1981.
- [40] H. Berendsen, J. Grigera, and T. Straatsma, The missing term in effective pair potentials, *Journal of Physical Chemistry* **91** (1987), no. 24 6269–6271.
- [41] L. Martínez, R. Andrade, E. G. Birgin, and J. M. Martínez, Packmol: a package for building initial configurations for molecular dynamics simulations, *Journal of computational chemistry* **30** (2009), no. 13 2157–2164.
- [42] T. Darden, D. York, and L. Pedersen, Particle mesh ewald: An nlog (n) method for ewald sums in large systems, *The Journal of chemical physics* **98** (1993), no. 12 10089–10092.
- [43] A. Hagler, E. Huler, and S. Lifson, Energy functions for peptides and proteins. i. derivation of a consistent force field including the hydrogen bond from amide crystals, *Journal of the American Chemical Society* **96** (1974), no. 17 5319–5327.
- [44] S. W. Bunte and H. Sun, Molecular modeling of energetic materials: the parameterization and validation of nitrate esters in the compass force field, *The Journal of Physical Chemistry B* **104** (2000), no. 11 2477–2489.

# Chapter 7

## Revisting Population Balances: The Impact of Solid Volume Fraction

Reproduced in part with permission from:

Vikram Khanna, Dennis Robinson Brown and Michael F. Doherty, “*Revisting Population Balances: The Impact of Solid Volume Fraction*”, *in preparation*.

### 7.1 Introduction

The continuous mixed suspension mixed product removal (MSMPR) crystallizer is ubiquitous in the crystallization field. Modeling the MSMPR requires solving the population balance equation coupled with the mass balance equation, along with rate expressions for crystal growth, nucleation, and other birth and death mechanisms as appropriate. Over the decades, the models used to describe the MSMPR often have unstated approximations, especially the solid volume fraction in the mass balance and the regularity boundary condition. In this chapter, we lay out the full model used to describe the MSMPR and the approximations required to arrive at its widely used form. We quantify

the approximations using examples from the literature. While there is no inherent need to make the approximations for numerical solutions of the model, they are essential for gaining insights via analytical steady-state and dynamic stability analyses.

## 7.2 The Continuous MSMPR System

Consider a continuous MSMPR crystallizer (as shown in Figure 7.1) which is fed a clear supersaturated solution of concentration,  $c_{in}$ , temperature,  $T_{in}$  at a volumetric flow rate,  $q_{in}$ . The crystallizer is maintained at a temperature,  $T$ . For cooling crystallization,  $T < T_{in}$ . Inside the crystallizer, the dissolved solute crystallizes (nucleation + growth), and hence, at the outlet of the crystallizer, we have a mixed suspension with a volumetric flow rate,  $q$ , comprising: i) the mother liquor at concentration,  $c$  and ii) solids making up  $1 - \epsilon$  volume fraction, where  $\epsilon = V_{liq}/V$ . Note, it is the suspension volume,  $V$ , that is assumed constant in a MSMPR (equivalent to assuming zero volume of mixing between solvent and solute).

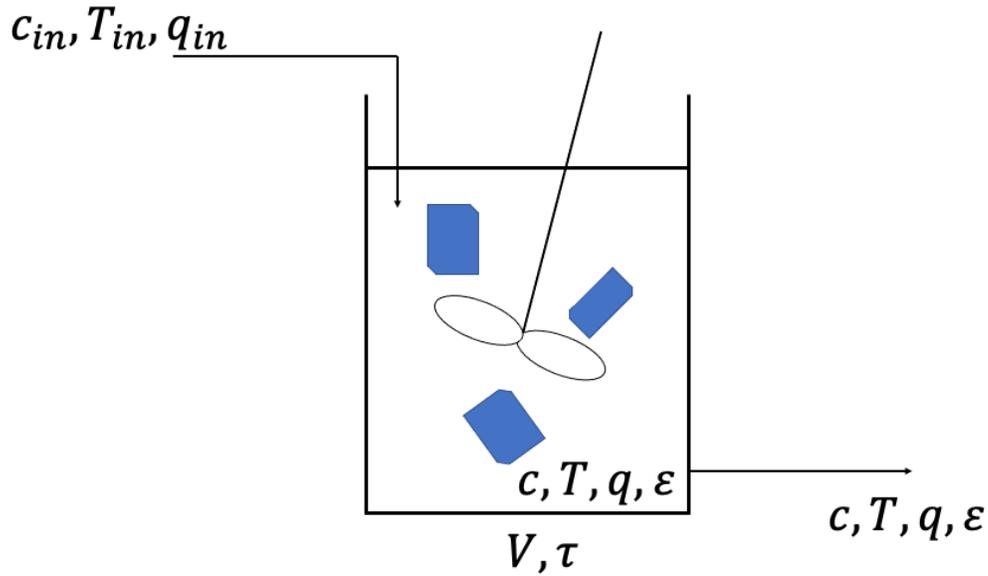


Figure 7.1: A mixed-suspension mixed-product removal crystallizer

## 7.2.1 Population Balance

The population balance for a MSMPR crystallizer is given by:

$$\frac{\partial n}{\partial t} = \frac{n_{in}}{\tau} - \frac{n}{\tau} + Q_B - Q_D - \frac{\partial(G \cdot n)}{\partial r} + \frac{\partial(\delta(r) \cdot G \cdot n)}{\partial r} \quad (7.1a)$$

$$n(r, t = 0) = n_{seed}(r) \quad (7.1b)$$

$$n(r \rightarrow \infty, t) = 0 \quad (7.1c)$$

where:

- $n$  is the population density of the crystals [ $\#/(m \cdot m^3 \text{ of suspension})$ ]

- $n_{in}$  is the population density of the crystals in the inlet stream [ $\#/(m \cdot m^3 \text{ of suspension})$ ]
- $\tau$  is the residence time [s]
- $Q_B$  is the birth rate of crystals inside the crystallizer [ $\#/(m \cdot m^3 \text{ of suspension} \cdot s)$ ]
- $Q_D$  is the death rate of crystals inside the crystallizer [ $\#/(m \cdot m^3 \text{ of suspension} \cdot s)$ ]
- $G$  is the linear growth rate of crystals [m/s]

Equation 7.1a reduces to

$$\frac{\partial n}{\partial t} = -\frac{n}{\tau} - G \frac{\partial n}{\partial r} + G \frac{\partial(\delta(r) \cdot n)}{\partial r} + \epsilon B \delta(r) \quad (7.2)$$

when:

- $n_{in} = 0$ , i.e., no crystals are present in the feed
- $Q_B = \epsilon B$ , where B is the nucleation rate having units  $\#/V_{liq}/s$
- $Q_D = 0$
- $G$  is independent of the crystal size

Equation 7.2 can be reformulated with the source terms removed and a regularity boundary condition inserted, as shown in Appendix K.1 to:

$$\frac{\partial n}{\partial t} = -\frac{n}{\tau} - G \frac{\partial n}{\partial r} \quad (7.3a)$$

$$n(r = 0^+, t) = \epsilon \frac{B}{G} \quad (7.3b)$$

The regularity boundary condition (given by Equation 7.3b) is approximated in the literature to have  $\epsilon = 1$ . However, industrial crystallizers are operated at  $\epsilon$  between 0.85–0.95 to have an economical yield of crystals.

The above population balance needs to be solved for in conjunction with the mass balance to model the MSMPR. Next, we formulate the mass balance across the MSMPR.

## 7.2.2 Mass Balance

The mass balance across the crystallizer is given by:

$$\text{Accumulation} = \text{In} - \text{Out} \quad (7.4a)$$

$$\frac{d(V_{liq}c)}{dt} + \frac{dM}{dt} = q_{liq,in}c_{in} - q_{liq,out}c - (1 - \epsilon)\rho q \quad (7.4b)$$

where:

- $d(V_{liq}c)/dt$  = Accumulation of solute mass in the liquid phase
- $dM/dt$  = Accumulation of solute mass in the solid phase
- $M = \rho V(1 - \epsilon)$  = mass of crystals
- $\rho$  = density of crystals
- $q_{liq,in}c_{in}$  = mass of solute entering in the liquid phase
- $q_{liq,out}c = \epsilon qc$  = mass of solute exiting in the liquid phase
- $(1 - \epsilon)\rho q$  = mass of solute exiting in the solid phase

Assuming  $q_{liq,in} = q$  and defining  $\tau = V/q$ , equation 7.4b is transformed to:

$$\frac{dc}{dt} = \frac{c_{in}}{\epsilon\tau} - \frac{c}{\tau} - \left(\frac{1-\epsilon}{\epsilon}\right) \frac{\rho}{\tau} + \left(\frac{\rho-c}{\epsilon}\right) \frac{d\epsilon}{dt} \quad (7.5)$$

The above equation can be re-written as

$$\frac{dc}{dt} = \frac{1}{\epsilon} \left[ \left(\frac{c_{in}-c}{\tau}\right) - 3k_v G m_2 (\rho - c) \right] \quad (7.6)$$

where,  $k_v$  is the shape volume factor and  $m_2$  is the second moment of the particle size distribution.

Now, it is assuming  $\epsilon \approx 1$  and  $\rho \gg c$  that we get

$$\frac{dc}{dt} = \left(\frac{c_{in}-c}{\tau}\right) - 3k_v G m_2 (\rho) \quad (7.7)$$

which is the widely used mass balance equation when modeling MSMPRs. The widely used model—in essence—discounts the accumulation of solute inside the crystallizer in the solid phase. Therefore, we now have the following two models for a MSMPR:

### The Full Model of a MSMPR

Thus, we have the following model for an MSMPR,

$$\frac{\partial n}{\partial t} = -\frac{n}{\tau} - G \frac{\partial n}{\partial r} \quad (7.8a)$$

$$n(r = 0^+, t) = \epsilon \frac{B}{G} \quad (7.8b)$$

$$n(r, t = 0) = n_{seed}(r) \quad (7.8c)$$

$$\frac{dc}{dt} = \frac{1}{\epsilon} \left[ \left( \frac{c_{in} - c}{\tau} \right) - 3k_v G m_2 (\rho - c) \right] \quad (7.9)$$

### The Approximate Model of a MSMPR

Thus, we have the following model for an MSMPR,

$$\frac{\partial n}{\partial t} = -\frac{n}{\tau} - G \frac{\partial n}{\partial r} \quad (7.10a)$$

$$n(r = 0, t) = \frac{B}{G} \quad (7.10b)$$

$$n(r, t = 0) = n_{seed}(r) \quad (7.10c)$$

$$\frac{dc}{dt} = \frac{c_{in} - c}{\tau} - 3k_v G m_2 \rho \quad (7.11)$$

## 7.3 Case study: L-glutamic acid ( $\alpha$ and $\beta$ polymorphs)

We use the L-glutamic acid crystal system modeled using the approximate model by Li et al.<sup>[1]</sup> to compute the quantitative impact of the approximations. We employ two quantitative metrics to capture the approximations: i) Productivity of the crystallizer, i.e., mass of crystals produced per unit time in the MSMPR, and ii) the volume weighted mean particle size.

Next, we setup the model rate expressions and parameters. The nucleation ( $B$ ) and

growth rates ( $G$ ) employed to model this system are:

$$B_i = k_{b,i} \left( \frac{c}{c_{sat,i}} - 1 \right)^{b_i} \mu_{2,i}, \quad (7.12)$$

and,

$$G_i = k_{g,i} \left( \frac{c}{c_{sat,i}} - 1 \right)^{g_i}, \quad (7.13)$$

respectively, where  $\mu_{2,i}$  is the second moment of the population density of polymorph  $i$ ; where  $i = \alpha$  or  $\beta$  phase. The parameter values used in the model are:<sup>[1]</sup>

Table 7.1: L-glutamic acid parameters

Parameter	Value	
	$i=\alpha$	$i=\beta$
$k_{b,i}$ (#/m <sup>3</sup> /s)	$3.81 \times 10^4$	$7.95 \times 10^2$
$b_i$	2.62	2.81
$k_{g,i}$ (m/s)	$1.30 \times 10^{-8}$	$1.60 \times 10^{-9}$
$g_i$	1.31	1.1
$c_{sat,i}$ (kg/m <sup>3</sup> )	11.36	8.51
$\rho_i$ (kg/m <sup>3</sup> )	1532	1569
$k_{v,i}$	$\pi/6$	0.01
$T$ (°C)	25	

## Results and Discussion

For an inlet supersaturation of 5.84 (with respect to the  $\beta$  polymorph,  $S_{in,\beta}$ ) and a residence time ( $\tau$ ) of 2 hrs, we solve the population balance equations coupled with

the mass balance equations for the two models. For these inlet conditions the modified Damkohler numbers ( $\Phi_\alpha, \Phi_\beta$ )<sup>[2]</sup> for the  $\alpha$  and  $\beta$  polymorphs are 3.653, and 1.324, respectively. Since,  $\Phi_\alpha > \Phi_\beta$ ,  $\alpha$  is the steady-state polymorph as predicted by Figure 7.2 and validated in Figure 7.3.

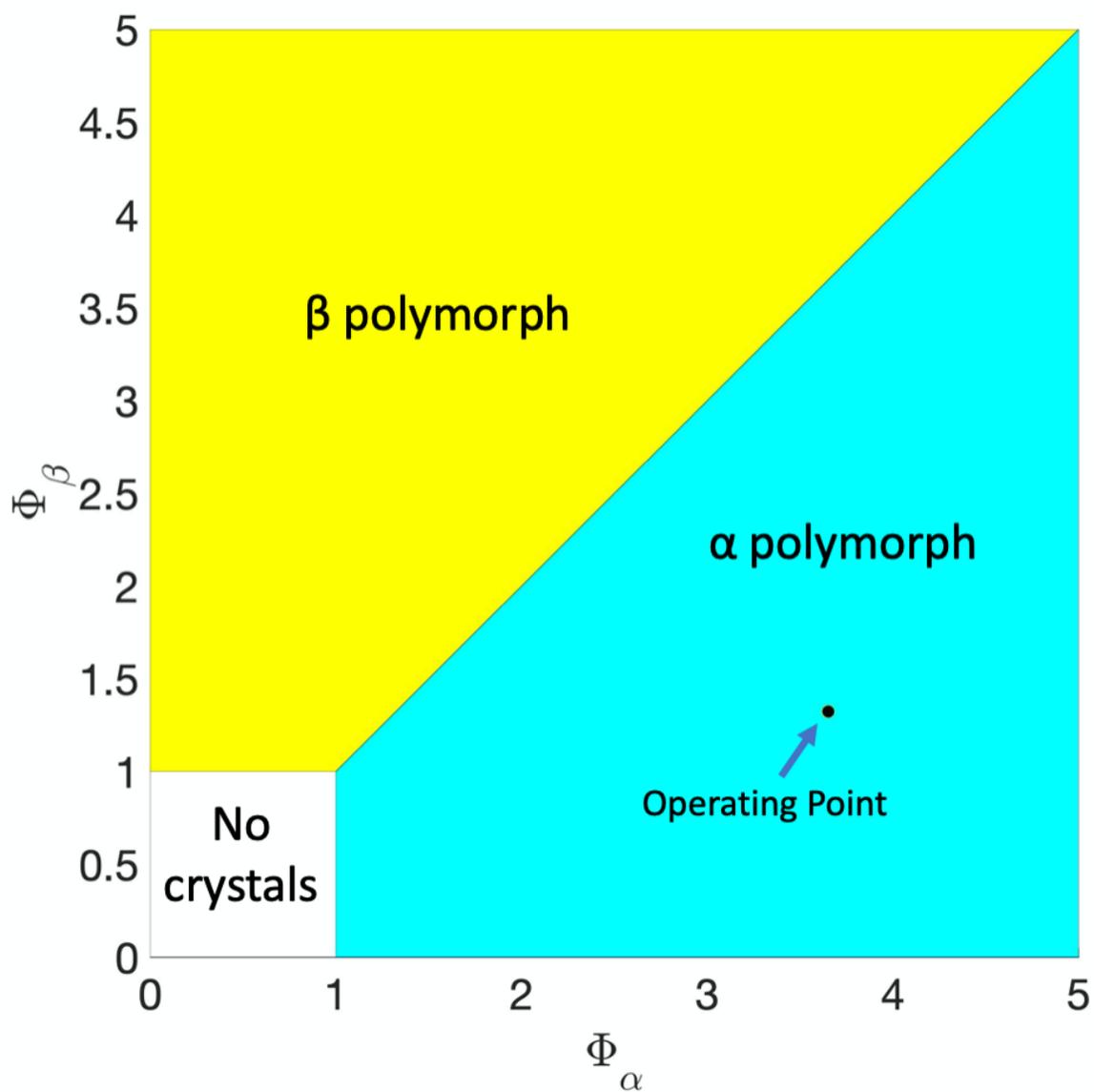


Figure 7.2: Stability diagram for polymorph selection of L-glutamic acid grown from aqueous solution

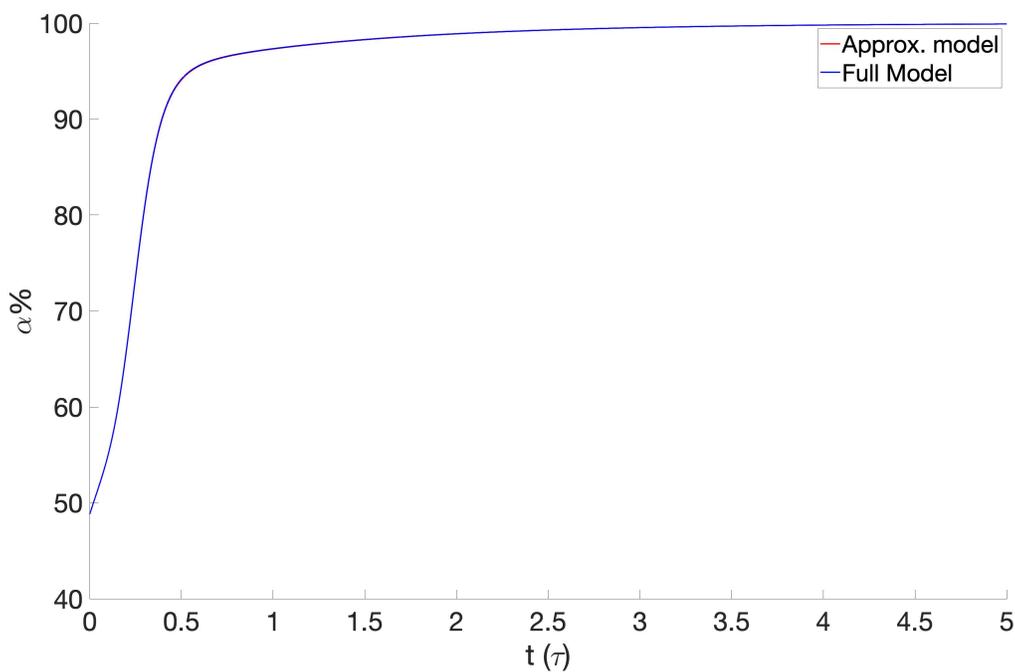
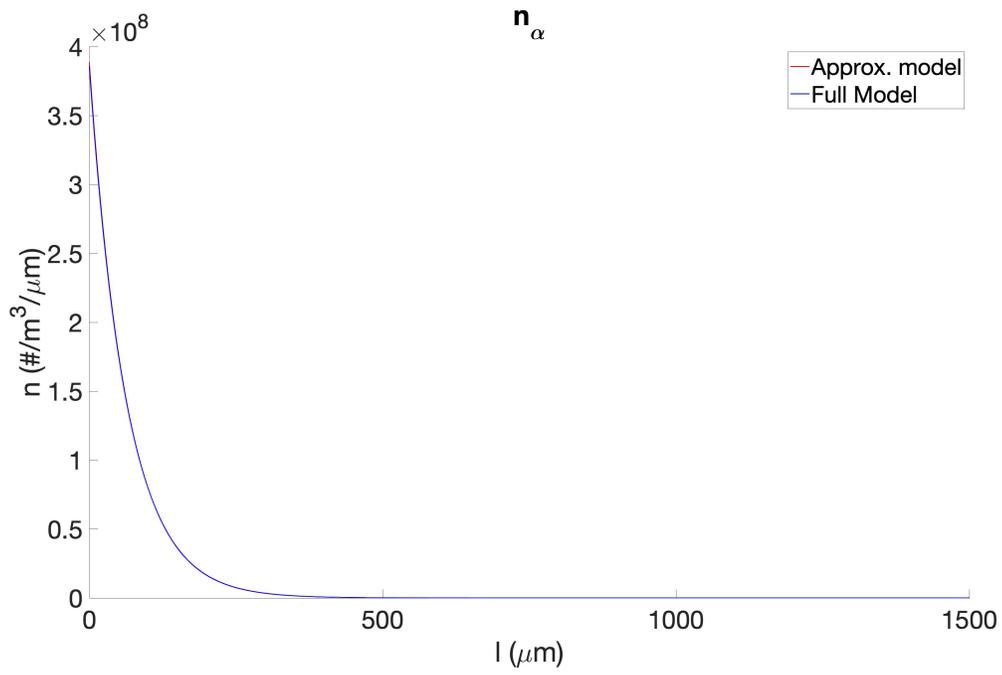
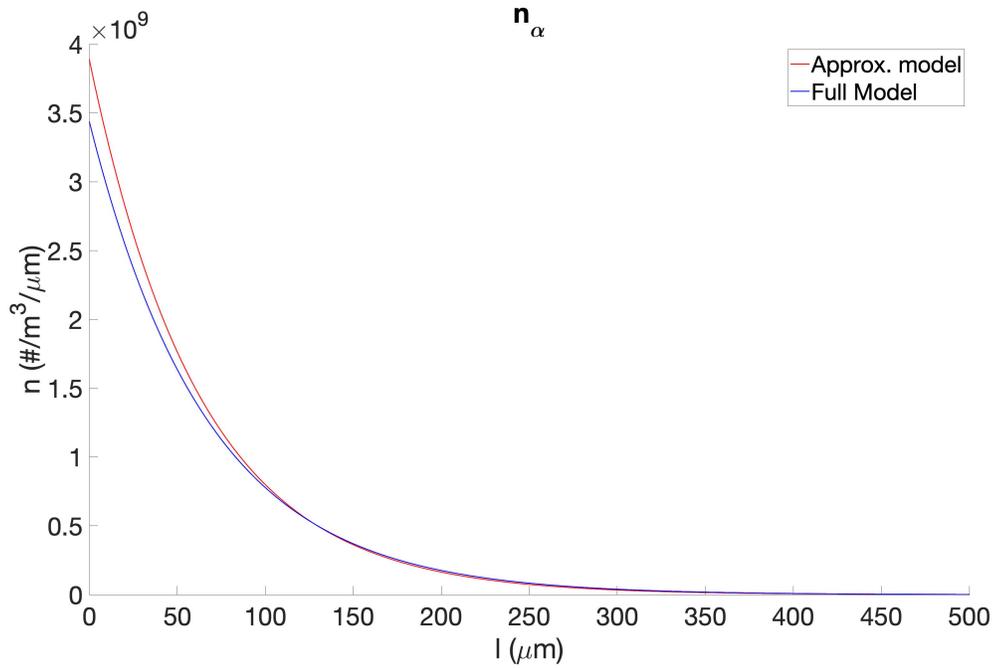


Figure 7.3: Polymorph percentage in MSMPR as a function of time. The crystallizer is seeded with equal percentages of each polymorph.

Figures 7.4a, 7.5a, 7.6a, and 7.7a show the steady-state population density, steady-state crystal size distribution, liquid volume fraction, and solute concentration in the MSMPR computed using the two models, respectively. Table 7.2 gives the difference between the approximate and the full model results for the two quantitative metrics: productivity and volume weighted mean particle size.



(a)



(b)

Figure 7.4: Population density of  $\alpha$  polymorph for case: a) original  $c_{sat}$ , b)  $c_{sat}$  increased by a factor of 10.

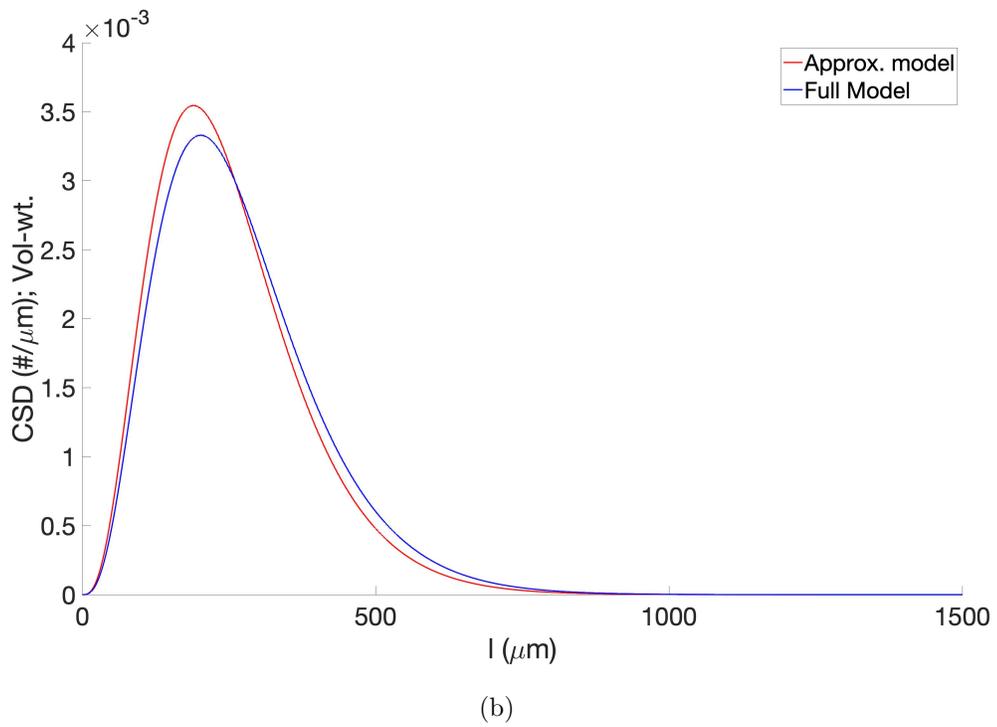
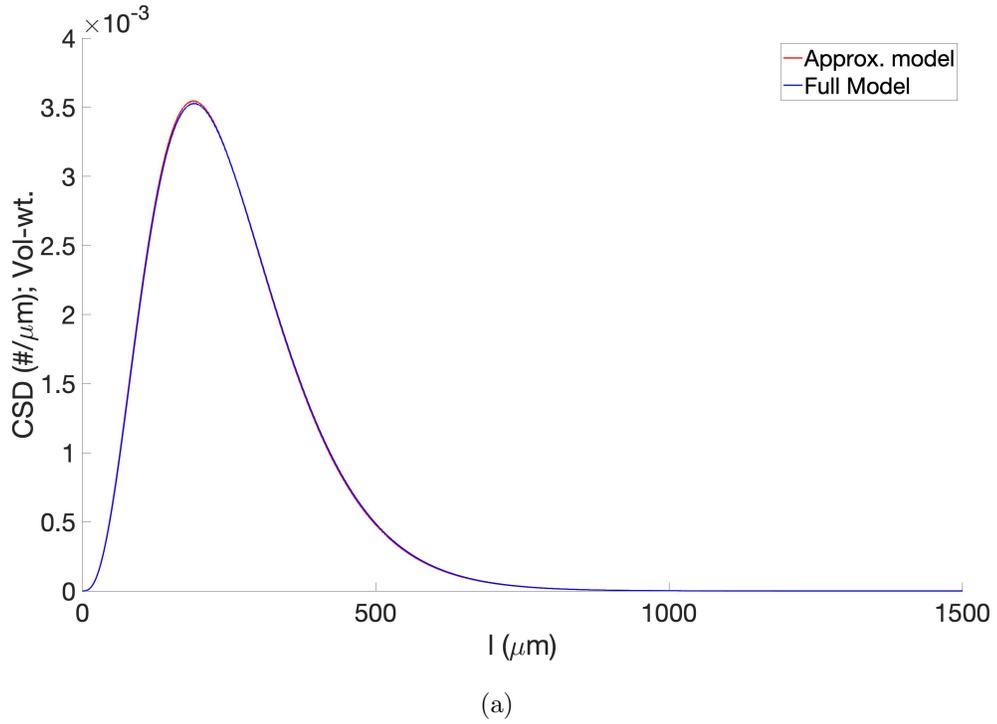


Figure 7.5: Crystal size distribution of  $\alpha$  polymorph for case: a) original  $c_{sat}$ , b)  $c_{sat}$  increased by a factor of 10.

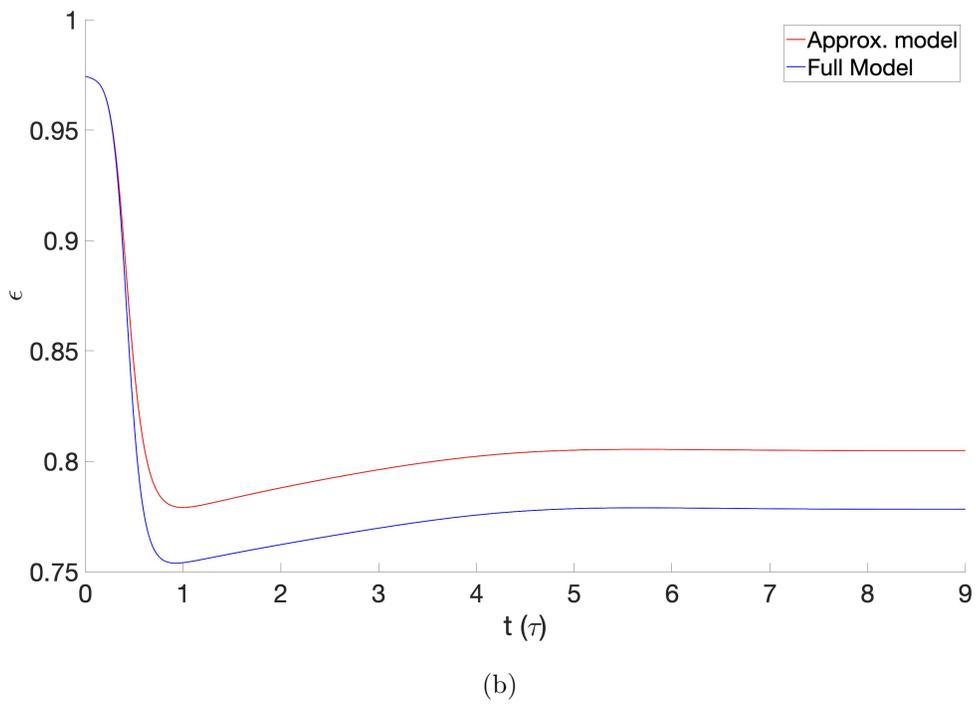
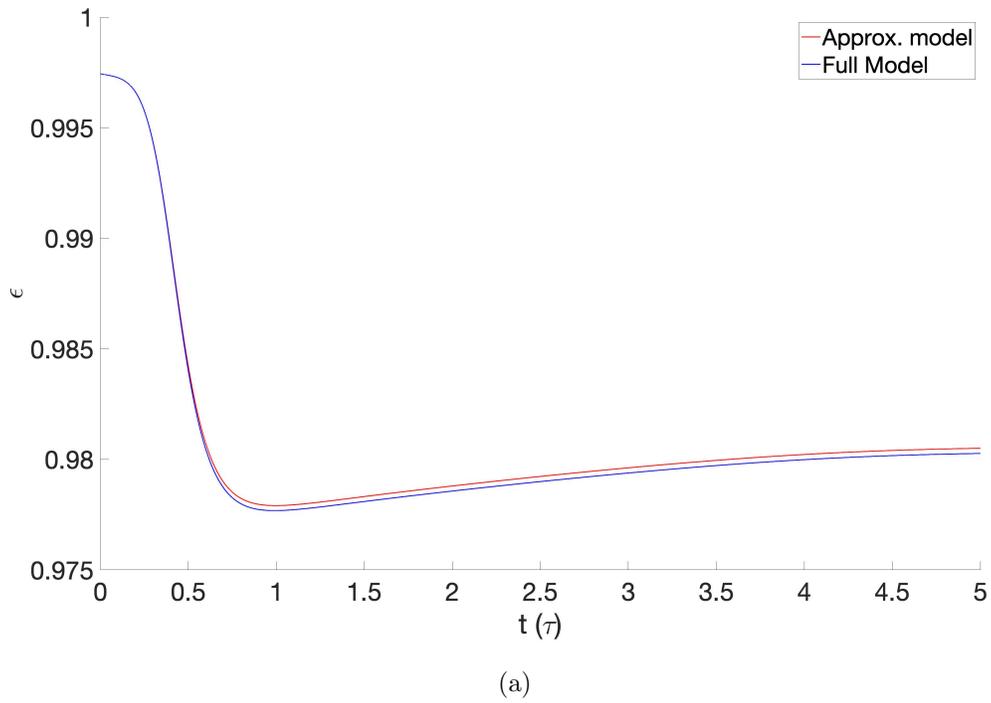


Figure 7.6: Liquid volume fraction in the MSMPR for case: a) original  $c_{sat}$ , b)  $c_{sat}$  increased by a factor of 10.

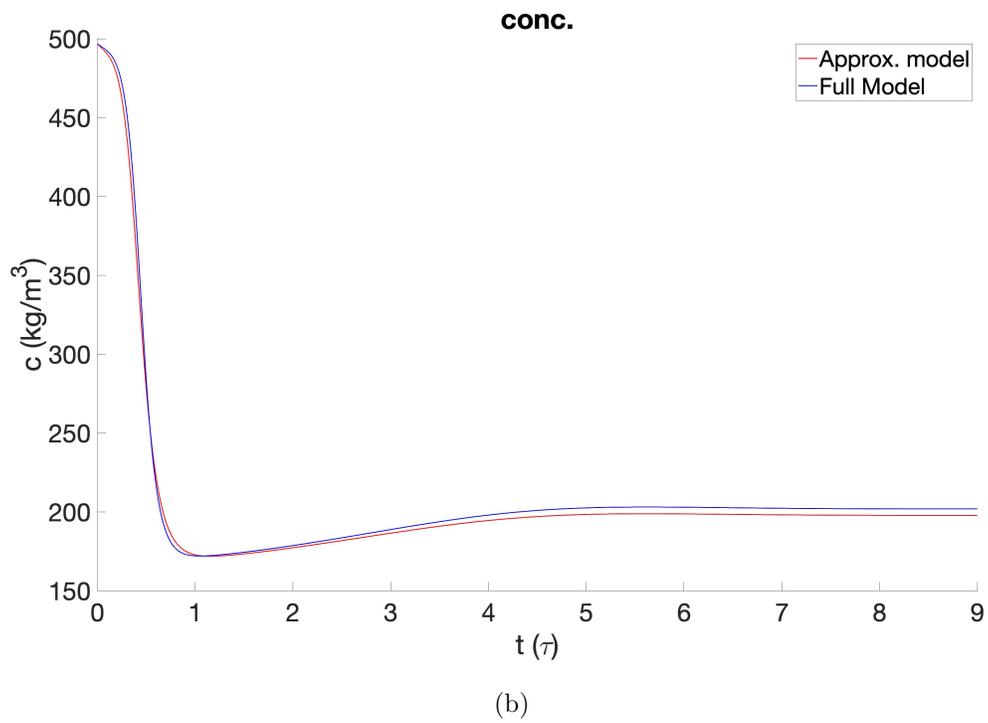
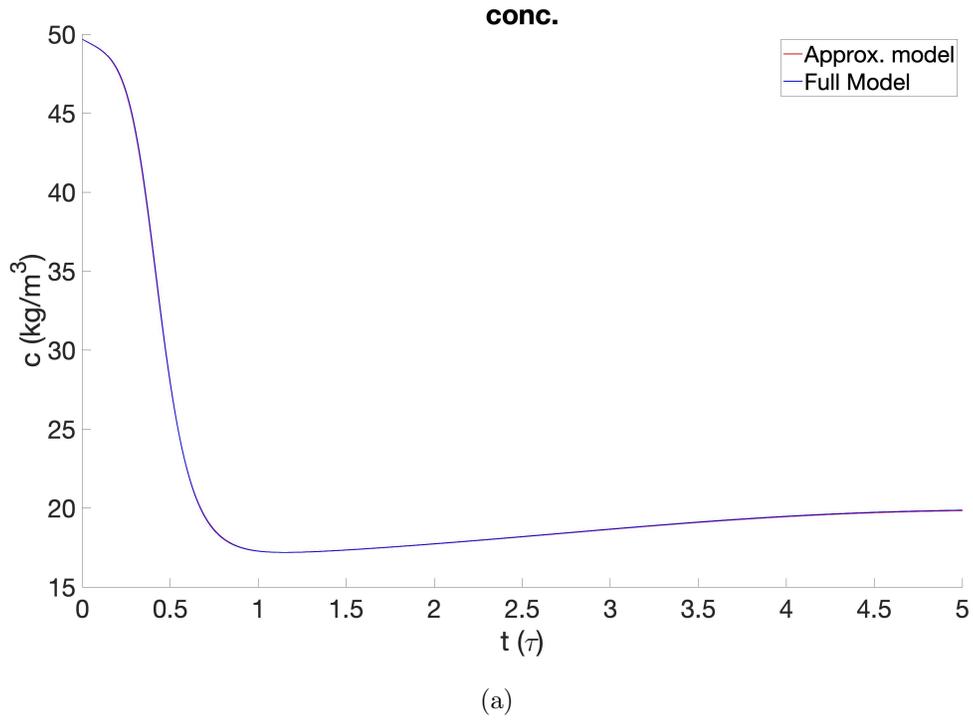


Figure 7.7: Concentration of solute in the MSMPR for case: a) original  $c_{sat}$ , b)  $c_{sat}$  increased by a factor of 10.

Table 7.2: Comparing the Approximate model against the Full model: original  $c_{sat}$ 

Test Metric	Approximate Model	Full Model	Diff(%)
Productivity/q (kg/m <sup>3</sup> of suspension)	29.8975	30.2567	-1.19%
Volume weighted Mean Particle Size ( $\mu\text{m}$ )	252.86	254.12	-0.50%

As we can see, the approximate model does an excellent job for these system parameters. However, we can see there is a small difference between the two models. Next, we test if the ratio  $c_{sat}/\rho$  is higher, does the difference between the approximate and the full model get amplified. For this, we artificially scale the solubility of the  $\alpha$  and  $\beta$  polymorphs by a factor of 10, and re-compute the results. As seen in Figures 7.4b, 7.5b, 7.6b, and 7.7b, the difference between the approximate and full model widens for a higher  $c_{sat}/\rho$  ratio. Table 7.3 quantifies the difference between the two models. As seen, the difference in the mass balance is now significant, ca. 12%, due to the higher  $c_{sat}/\rho$  ratio. Therefore, the ratio  $c_{sat}/\rho$  does impact the accuracy of the approximate model. To test the sensitivity of the approximate model with this ratio we solve the models for varying  $c_{sat}/\rho$  ratios. The results shown in Table 7.4 indicate that for systems where the solubility of the solute is high (ca.  $\geq 3\%$  of the solute's solid density), the deviation between the approximate and the full model maybe significant.

Table 7.3: Comparing the Approximate model against the Full model:  $c_{sat}$  scaled by 10

Test Metric	Approximate Model	Full Model	Diff(%)
Productivity/q (kg/m <sup>3</sup> of suspension)	298.9753	339.6082	-11.96%
Volume weighted Mean Particle Size ( $\mu\text{m}$ )	252.8603	269.2093	-6.07%

Table 7.4: Difference between Approximate and Full Models for varying  $c_{sat}/\rho$ 

$c_{sat}/\rho$	Metric	Approximate Model	Full Model	Diff(%)
0.0074	Productivity/q (kg/m <sup>3</sup> )	29.90	30.26	-1.19%
	Volume weighted Mean Particle Size ( $\mu\text{m}$ )	252.86	254.12	-0.50%
0.0148	Productivity/q (kg/m <sup>3</sup> )	59.80	61.25	-2.38%
	Volume weighted Mean Particle Size ( $\mu\text{m}$ )	252.86	255.45	-1.01%
0.0222	Productivity/q (kg/m <sup>3</sup> )	89.69	93.01	-3.57%
	Volume weighted Mean Particle Size ( $\mu\text{m}$ )	252.86	256.85	-1.55%
0.0296	Productivity/q (kg/m <sup>3</sup> )	119.59	125.57	-4.76%
	Volume weighted Mean Particle Size ( $\mu\text{m}$ )	252.86	258.32	-2.11%
0.037	Productivity/q (kg/m <sup>3</sup> )	149.49	158.95	-5.95%
	Volume weighted Mean Particle Size ( $\mu\text{m}$ )	252.86	259.88	-2.70%
0.074	Productivity/q (kg/m <sup>3</sup> )	298.98	339.61	-11.96%
	Volume weighted Mean Particle Size ( $\mu\text{m}$ )	252.86	269.21	-6.07%

## 7.4 Conclusion

In this chapter we rederive the accurate population balance and mass balance equations for a MSMPR. We emphasize the approximations made to in the accurate model to arrive at the widely used models in literature.<sup>[3]</sup> The approximations are excellent for majority crystal systems, however, for crystals with high solubility the differences between the approximate and accurate model may be significant as demonstrated by the L-glutamic

acid case study. Note that the approximations are not necessary for numerical simulations but they are for analytical steady-state and dynamic stability analysis.

# Bibliography

- [1] Y. Li, S. O'Shea, Q. Yin, and T. Vetter, Polymorph selection by continuous crystallization in the presence of wet milling, *Crystal Growth & Design* **19** (2019), no. 4 2259–2271.
- [2] T. C. Farmer, C. L. Carpenter, and M. F. Doherty, Polymorph selection by continuous crystallization, *AIChE Journal* **62** (2016), no. 9 3505–3514.
- [3] A. Randolph, *Theory of particulate processes: analysis and techniques of continuous crystallization*. Elsevier, 2012.

# Chapter 8

## Ever-kinked Crystal System

Reproduced in part with permission from:

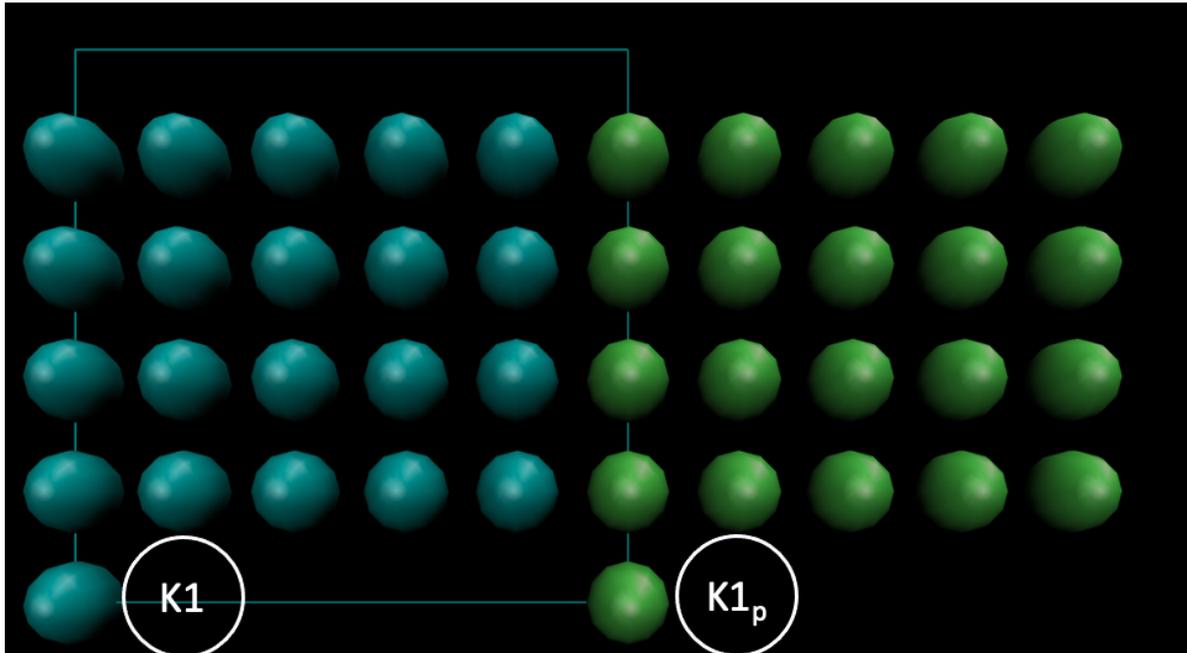
Vikram Khanna, Michael F. Doherty, and Baron Peters, “The Ever-kinked Crystal Framework for In Silico Crystal Growth Studies”, *in preparation*.

### 8.1 Introduction

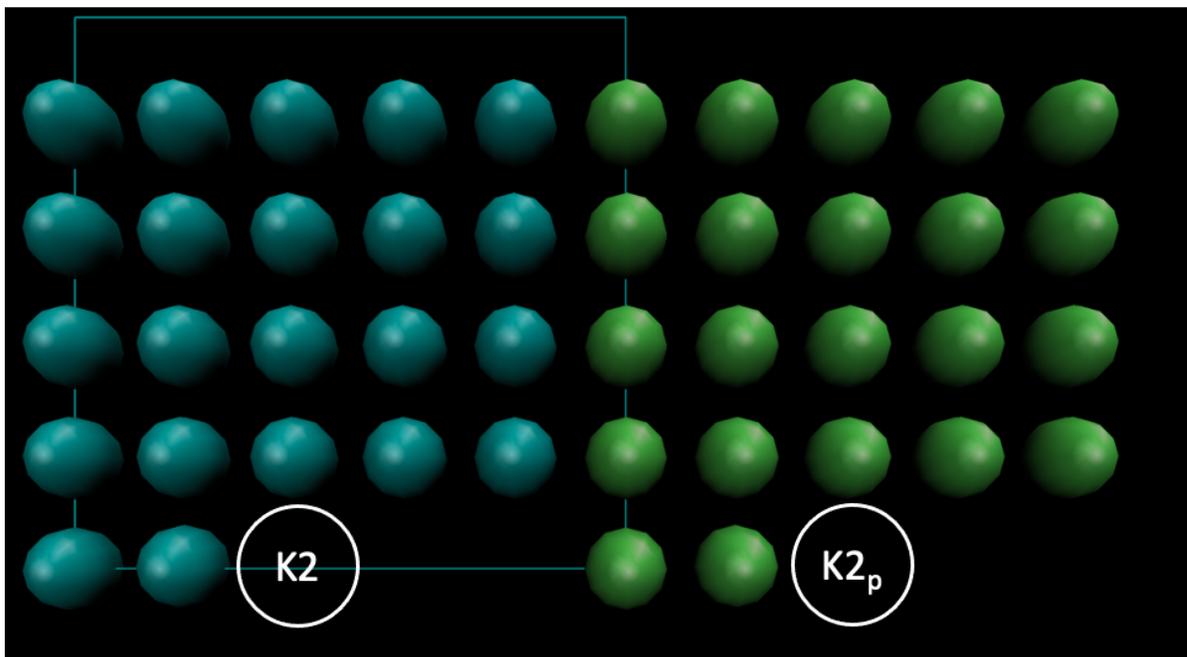
Having developed a computational framework to predict driving forces for crystallization the next piece of the puzzle to predict growth rates is computing the rate constants for attachment and detachment events at kink sites. Joswiak et al.<sup>[1]</sup> have developed a framework for ionic crystals, laying the necessary groundwork to compute rate constants. The first step in their approach is to create an ever-kinked crystal slab to avoid finite-size effects. For this, they developed a custom framework for cubic crystals. However, most of the organic crystals (> 99.5%) belong to non-cubic crystal systems. Thus, we need to develop a framework to develop an ever-kinked crystal slab for a general triclinic crystal system. This section focuses on the development of this framework.

## 8.2 The Periodic Kink Problem

In order to avoid finite size effects while studying attachment and detachment processes using simulations, we need to make sure that post either of the events, the kink which is generated has the same environment. Consider the kink system shown in Figure 8.1. As one fills up the incomplete row, the distance between successive kinks decreases.



(a)

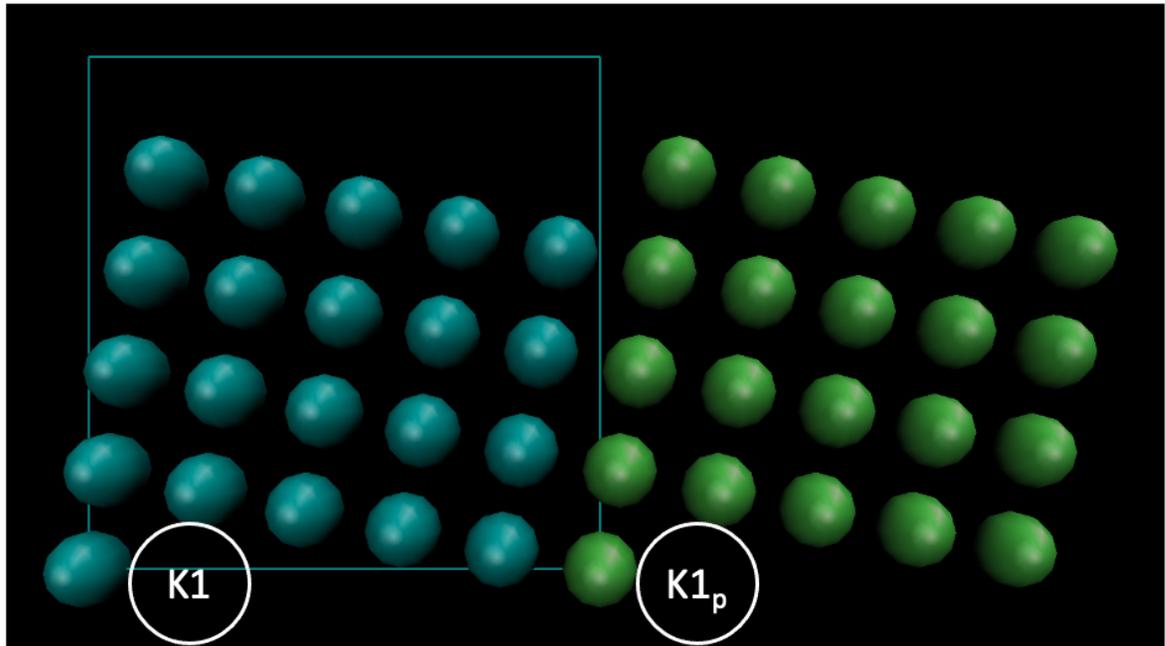


(b)

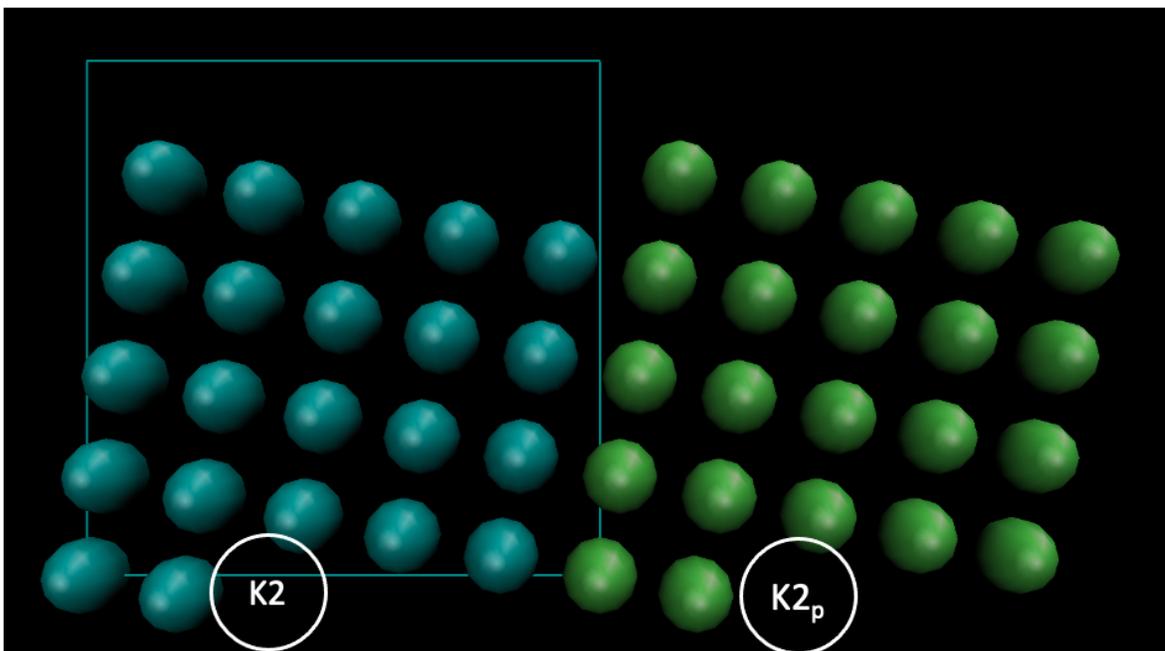
Figure 8.1: Distance between kinks on addition of the growth unit reduced by one growth unit from (a) to (b); Lines indicate the periodic boundary and lime spheres are the periodic images of the cyan spheres.  $K1$  and  $K2$  are kinks, and  $K1_p$  and  $K2_p$  are their periodic images, respectively.

### 8.3 The Ever-kinked Solution: 2-D case

The solution is to rotate the lattice in a particular way (as shown in Figure 8.2) so that we maintain a constant distance between successive kinks post an attachment/detachment event. It can be seen in Figure 8.2 that on adding a growth unit (indicated by the pointer in Figure 8.2b), the distance between the kinks does not change. Hence, we recover the original kink system.



(a)



(b)

Figure 8.2: Distance between kinks on addition of the growth unit remains constant from (a) to (b); Blue lines indicate the periodic boundary and lime spheres are the periodic images of the grey spheres

To achieve this, we can take a crystal system and tilt it about one of the lattice axes by one unit cell. Mathematically this transforms to solving the following set of equations:

$$(m.a)\mathbf{i}_{new} + b\mathbf{j}_{new} = L_{x,new}\mathbf{i} \quad (8.1a)$$

$$-a\mathbf{i}_{new} + (n.b)\mathbf{j}_{new} = L_{y,new}\mathbf{j} \quad (8.1b)$$

$$\mathbf{i}_{new} \cdot \mathbf{i}_{new} = 1 \quad (8.1c)$$

$$\mathbf{j}_{new} \cdot \mathbf{j}_{new} = 1 \quad (8.1d)$$

$$\mathbf{i}_{new} \cdot \mathbf{j}_{new} = \cos(\gamma) \quad (8.1e)$$

where  $\mathbf{i}$  and  $\mathbf{j}$  are original unit lattice vectors of the 2D lattice having values:

$$\mathbf{i} = (1, 0),$$

$$\mathbf{j} = (\cos\gamma, \sin\gamma),$$

$\gamma$  is the angle between the vectors, and  $\mathbf{i}_{new}, \mathbf{j}_{new}$  are new unit lattice vectors for the rotated lattice,

$(m, n)$  = no. of unit cells in  $\mathbf{i}$  and  $\mathbf{j}$  direction,

$(L_{x,new}, L_{y,new})$  = periodic box length in  $(\mathbf{i}, \mathbf{j})$  direction, respectively.

Solving the above set of equations for a 2-D orthogonal lattice (i.e.  $\gamma = 90^\circ$ ) we get (see Appendix L for details):

$$a^2m - b^2n = 0 \quad (8.2)$$

This result places a restriction on the supercell size we can create as  $m$  and  $n$  can only

admit integer values. Equation 8.2 is a diophantine equation as the unknowns  $m$  and  $n$  can only admit integer values. This result indicates that for certain crystal lattice parameters the supercell size can be large enough to be computationally infeasible.

For the case of a cubic lattice, i.e., ( $a = b$ ) equation 8.2 implies  $m = n$ . Therefore, for cubic systems we can always create computationally feasible supercells as long as the supercell is also cubic, i.e.,  $m = n$ .

For a general  $\gamma$  value, we don't get a neat result as equation 8.2.

## 8.4 The Ever-kinked Solution: 3-D case

For the 3-D case we can take a crystal system and tilt it about one of the lattice axes by one unit cell, this would lead to the formation of steps. Another rotation about another axes would lead to the formation of periodic kinks. Mathematically this transforms to solving the following set of equations:

$$(m.a)\mathbf{i}_{new} - b\mathbf{j}_{new} = L_{x,new}\mathbf{i} \quad (8.3a)$$

$$a\mathbf{i}_{new} + (n.b)\mathbf{j}_{new} + c\mathbf{k}_{new} = L_{y,new}\mathbf{j} \quad (8.3b)$$

$$\mathbf{i}_{new} \cdot \mathbf{i}_{new} = 1 \quad (8.3c)$$

$$\mathbf{j}_{new} \cdot \mathbf{j}_{new} = 1 \quad (8.3d)$$

$$\mathbf{k}_{new} \cdot \mathbf{k}_{new} = 1 \quad (8.3e)$$

$$\mathbf{i}_{new} \cdot \mathbf{j}_{new} = \cos(\gamma) \quad (8.3f)$$

$$\mathbf{j}_{new} \cdot \mathbf{k}_{new} = \cos(\alpha) \quad (8.3g)$$

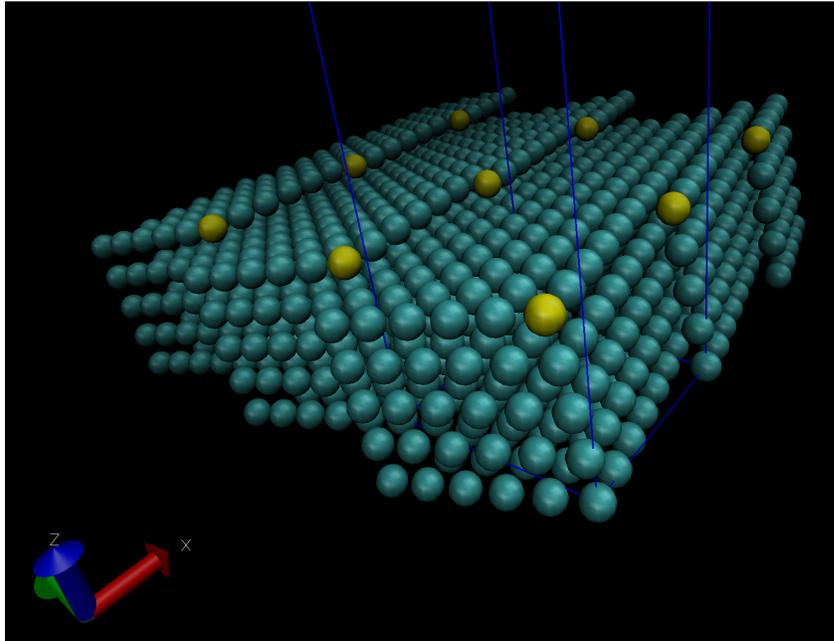
$$\mathbf{k}_{new} \cdot \mathbf{i}_{new} = \cos(\beta) \quad (8.3h)$$

where  $\mathbf{i}$ ,  $\mathbf{j}$  and  $\mathbf{k}$  are original unit lattice vectors of the 3D lattice, and  $\mathbf{i}_{new}$ ,  $\mathbf{j}_{new}$ ,  $\mathbf{k}_{new}$  are new unit lattice vectors for the rotated lattice. Variables  $m$ , and  $n$  are no. of unit cells in  $\mathbf{i}$  and  $\mathbf{j}$  direction, and  $(L_{x,new}, L_{y,new})$  are periodic box lengths in  $(\mathbf{i}, \mathbf{j})$  direction, respectively.

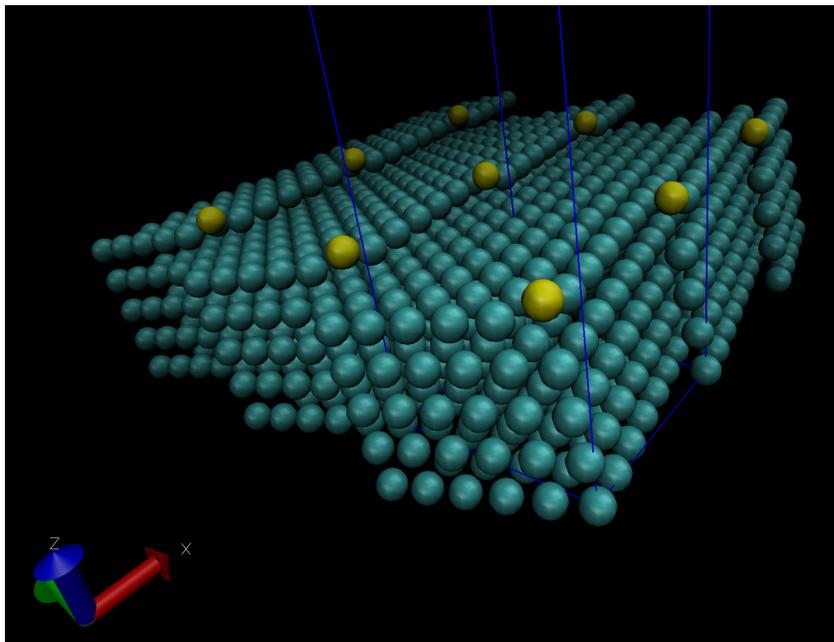
Solving the above set of equations for a 3-D orthogonal lattice (i.e.  $\alpha=\beta=\gamma=90^\circ$ ) we get:

$$a^2m - b^2n = 0 \quad (8.4)$$

The above result is same as the 2D case (see equation 8.2 and see Appendix L for details). Therefore, it is likely for many crystal systems the supercell size would be computationally infeasible. However, by making some approximations in the lattice vectors, it may be possible to create ever-kinked crystal systems that satisfy the above equations. Also, note that creating the ever-kinked crystal system by only tilting and twisting a crystal leads to the formation of holes in the system (see Appendix L). We solve this problem by adding an extra column of unit cells as shown in Appendix L. The final result is shown in Figure 8.3.



(a)



(b)

Figure 8.3: Ever-kinked cubic lattice: (a) Yellow atoms are the ones sitting in a potential kink site, (b): Periodic kinks being displayed via dislodging an atom sitting in a potential kink site.

### 8.4.1 The Approximate Solution

Consider the case of the 3-D orthogonal lattice. Solving the ever-kinked framework leads us to equation 8.4. Since,  $m$  and  $n$  are integers, equation 8.4 implies  $b^2/a^2$  must be a rational number. Now, since  $a$  and  $b$  are lattice constants, this is not guaranteed for every crystal system. Therefore, for some crystal systems, equation 8.4 will not be satisfied. For such cases, we propose an approximate solution, i.e., we approximate either  $a$ ,  $b$ , or both, such that  $b_{new}^2/a_{new}^2$  is a rational number; where

$$b_{new} = b + \delta_b \quad (8.5)$$

and,

$$a_{new} = a + \delta_a \quad (8.6)$$

Therefore, by choosing  $\delta_a$  and  $\delta_b$ , we approximate the lattice constants such that

$$\frac{(b + \delta_b)^2}{(a + \delta_a)^2} = \frac{m}{n} \quad (8.7)$$

This strategy enables us to pick supercell sizes that are computationally feasible and satisfy the ever-kinked framework. See Appendix L for details.

## 8.5 Conclusions

In this chapter we develop a general framework to create ever-kinked crystals for studying crystal growth using atomistic simulations that alleviates the finite-size effects. In summary, for a given crystal system one needs to choose from the following three system-setups to investigate growth rates:

- For computationally feasible supercell sizes create an ever-kinked system using the

ever-kinked framework developed in this chapter.

- For computationally infeasible sizes one needs to either approximate the lattice parameters to enable a feasible supercell size, however this can develop stresses within the crystal system and would not be the recommended route for large approximations.
- The alternative is to work with a periodic kinked system that suffers from finite size effects and then accurately account and correct for the finite size effects.

# Bibliography

- [1] M. N. Joswiak, B. Peters, and M. F. Doherty, In silico crystal growth rate prediction for NaCl from aqueous solution, *Crystal Growth & Design* **18** (2018), no. 10 6302–6306.

# Chapter 9

## Conclusions and Future Work

### 9.1 Conclusions

A digital design approach to engineer crystals is one of the most prudent ways to reduce the risks and costs of pharmaceutical drug development. This thesis outlines a digital design strategy to predict crystal polymorphs, shapes, and sizes starting from a target molecule. It focuses on developing tools to predict driving forces for crystallization—accurately and precisely making use of high fidelity molecular simulations. In summary:

Chapter 1 outlines the major design elements of the digital design strategy: namely, polymorph prediction, shape prediction, nucleation & growth rate predictions, solubility prediction, attainable particle size prediction, and lastly polymorph selection. The design element with one of the most pressing challenges is nucleation, in particular, heterogeneous nucleation and secondary nucleation.

Chapter 2 discusses the crystal growth models to compute crystal shapes and sizes. It presents the roadmap to predict driving forces for crystallization—a key prediction for predicting crystal sizes—via atomistic simulations, using a decoupled approach—avoiding

difficult to compute fluid-to-solid transitions.

Chapter 3 completes the theory for accurately computing solid free energies using the Einstein crystal approach making use of any general set of spring constants. It delineates the computational strategy to setup solid free energy simulations for atomic systems and extends it for molecular systems to account for intramolecular free energies—yielding an absolute free energy prediction.

Chapter 4 introduces a new gas phase reference system for complex molecules—the centroid—pivotal for decoupling the phase equilibrium calculation. It lays out the computational strategy to compute the free energy of the centroid and gas phase molecules, testing it via a diatomic molecule system.

Chapter 5 develops the computational strategy to compute solvation free energies using the decoupling approach for molecular dynamics software such as LAMMPS. It also develops a thermodynamic pathway to compute solvent-modified bond energies—leveraging solvation free energy calculations—which are key computations for morphology predictions of crystals grown from solution.

Chapter 6 is where the above three computational tools—developed to compute chemical potentials of a solute in the solid, gas, and solutions phases, respectively—are leveraged to compute solid-vapor and solid-solution equilibrium and predict the driving forces for crystallization from vapor and solution of two model compounds, namely, naphthalene and succinic acid. It also gives a workflow for bench marking the solute’s (target molecule) force field while making sublimation phase equilibrium predictions via the Clausius-Clapeyron plot.

Chapter 7 revisits the population and mass balance models employed for modeling the ubiquitously used mixed suspension, mixed product removal (MSMPR) crystallizer. It identifies the approximations made to arrive at the widely used models in the literature and quantifies their impact on tangible output parameters of the crystallizer—namely,

productivity and volume weighted mean crystal size.

Chapter 8 lays out the framework to create ever-kinked crystal systems that avoid finite-size effects while studying crystal growth for a general triclinic crystallography.

## 9.2 Future Work

### Solubility predictions

On the solubility prediction front, the following three directions should be explored:

- The absolute chemical potential route framework must be leveraged to compute solubilities using different combinations of force fields for a variety of compound types. This will enable us to understand which combination works best for a given class of materials. For example, from morphology prediction results we believe that the Lifson force field works well for dicarboxylic acids. Therefore, testing it for the solubility prediction of succinic acid would be a good next step. Similarly, a well curated database of solubility predictions would aid the future users to make wise choices while using off the shelf force fields.
- Scaling the solute-solvent interactions (mixing rules) to match solubility predictions with experimental data. However, this approach will add the use of experimental data to the digital design strategy steering it away from a purely in-silico approach.
- Another addition to the solubility prediction workflow would be direct coexistence calculations—to have a self-consistency check. We could leverage the ever-kinked framework discussed in Chapter 8 to potentially speed up direct coexistence simulations.

## Growth rate predictions

As noted in Chapter 8, for a given crystal system one needs to choose from the following three system setups to investigate growth rates:

- For computationally feasible supercell sizes create an ever-kinked system using the ever-kinked framework developed in Chapter 8.
- For computationally infeasible sizes one needs to either approximate the lattice parameters to enable a feasible supercell size, however this can develop stresses within the crystal system and would not be the recommended route for large approximations.
- The alternative is to work with a periodic kinked system that suffers from finite size effects and then accurately account and correct for the finite size effects.

The next step to compute growth rates would be to identify the important solvent and solute collective variables (CVs) for the system under study. For this, one needs to harvest an ensemble of Transition paths via the flexible length<sup>[1]</sup> and aimless shooting<sup>[2]</sup> variety of Transition Path Sampling (TPS). Then prepare an extensive list of trial reaction coordinates and rank them using the inertial version of likelihood maximization.<sup>[3]</sup> Next, perform committor tests<sup>[4]</sup> to see how well peaked the committor distribution for the chosen reaction coordinate comes out to be. For the chosen solute and solvent reaction coordinate system compute free energies using harmonic umbrella sampling method. Followed by creating global free energy surfaces and profiles. Also one needs to calculate transmission coefficients for correcting the transition state theory predicted rates. This framework will enable to calculate attachment and detachment rates, and gain significant insight into the mechanism of the crystal growth process for complex molecules.

# Bibliography

- [1] R. G. Mullen, J.-E. Shea, and B. Peters, Easy transition path sampling methods: Flexible-length aimless shooting and permutation shooting, *Journal of chemical theory and computation* **11** (2015), no. 6 2421–2428.
- [2] B. Peters and B. L. Trout, Obtaining reaction coordinates by likelihood maximization, *The Journal of chemical physics* **125** (2006), no. 5 054108.
- [3] B. Peters, Inertial likelihood maximization for reaction coordinates with high transmission coefficients, *Chemical Physics Letters* **554** (2012) 248–253.
- [4] B. Peters, Reaction coordinates and mechanistic hypothesis tests, *Annual review of physical chemistry* **67** (2016) 669–690.

# Appendix A

## Deriving the finite size correction

### A.1 Breakdown of the Helmholtz free energy of a crystal

The Helmholtz free energy calculation of a crystal is given by

$$f_C = f_{EC} + (f_{EC}^{CM} - f_{EC}) + (f_C^{CM} - f_{EC}^{CM}) + (f_C - f_C^{CM}) \quad (\text{A.1})$$

where the absolute free energy of the Einstein crystal can be obtained from its analytically computable partition function:

$$\beta f_{EC} = \frac{-1}{N_{mol}} \ln(Q_{EC}) = \frac{1}{N_{mol}} \sum_{i=1}^N \ln \left( \frac{\beta k_i \Lambda_i^2}{2\pi} \right)^{3/2} \quad (\text{A.2})$$

and the free energy differences are

$$\beta(f_{EC}^{CM} - f_{EC}) = \frac{-1}{N_{mol}} \ln \left( \frac{Q_{EC}^{CM}}{Q_{EC}} \right) = \frac{-1}{N_{mol}} \ln \left[ \left( \frac{\beta}{2\pi \sum_{i=1}^N \frac{\mu_i^2}{k_i}} \right)^{3/2} \left( \frac{\beta h^2}{2\pi M} \right)^{3/2} \right] \quad (\text{A.3a})$$

$$\beta(f_C^{CM} - f_{EC}^{CM}) = \frac{\beta}{N_{mol}} \int_{\lambda=0}^{\lambda=1} \left\langle \frac{dU}{d\lambda} \right\rangle_{\lambda}^{CM} d\lambda \quad (\text{A.3b})$$

$$\beta(f_C - f_C^{CM}) = \frac{-1}{N_{mol}} \ln \left( \frac{Q_C}{Q_C^{CM}} \right) = \frac{-1}{N_{mol}} \ln \left[ \left( \frac{V}{N_{mol}} \right) \left( \frac{\beta h^2}{2\pi M} \right)^{-3/2} \right] \quad (\text{A.3c})$$

In each of these formulas,  $Q$  is a partition function,  $\beta = (k_B T)^{-1}$ ,  $\lambda$  is a coupling parameter,  $U$  is a  $\lambda$ -dependent potential energy function that interpolates between that of the EC and C systems,  $h$  is Planck's constant,  $V$  is volume of the system,  $\mu_i = m_i / \sum_{i=1}^N m_i$ ,  $M = \sum_{i=1}^N m_i$ , where  $m_i$  is the mass of atom  $i$ ,  $\Lambda_i = (\beta h^2 / (2\pi m_i))^{1/2}$ , and  $k_i$  is the spring constant of the spring attached to atom  $i$ .

Therefore, the Helmholtz free energy of the crystal is

$$\begin{aligned} \beta f_C &= \frac{1}{N_{mol}} \sum_{i=1}^N \ln \left( \frac{\beta k_i \Lambda_i^2}{2\pi} \right)^{3/2} \\ &\quad - \frac{1}{N_{mol}} \ln \left[ \left( \frac{\beta}{2\pi \sum_{i=1}^N \frac{\mu_i^2}{k_i}} \right)^{3/2} \left( \frac{\beta h^2}{2\pi M} \right)^{3/2} \right] \\ &\quad + \frac{\beta}{N_{mol}} \int_{\lambda=0}^{\lambda=1} \left\langle \frac{dU}{d\lambda} \right\rangle_{\lambda}^{CM} d\lambda \\ &\quad - \frac{1}{N_{mol}} \ln \left[ \left( \frac{V}{N_{mol}} \right) \left( \frac{\beta h^2}{2\pi M} \right)^{-3/2} \right] \end{aligned} \quad (\text{A.4})$$

The  $\beta h^2/(2\pi M)$  part of the log terms cancel out, giving

$$\begin{aligned}
\beta f_C &= \frac{1}{N_{mol}} \sum_{i=1}^N \ln \left( \frac{\beta k_i \Lambda_i^2}{2\pi} \right)^{3/2} \\
&\quad - \frac{1}{N_{mol}} \ln \left( \frac{\beta}{2\pi \sum_{i=1}^N \frac{\mu_i^2}{k_i}} \right)^{3/2} \\
&\quad + \frac{\beta}{N_{mol}} \int_{\lambda=0}^{\lambda=1} \left\langle \frac{dU}{d\lambda} \right\rangle_{\lambda}^{CM} d\lambda \\
&\quad - \frac{1}{N_{mol}} \ln \left( \frac{V}{N_{mol}} \right)
\end{aligned} \tag{A.5}$$

Note, combining the log terms in the above equation gives a dimensionless argument to the logarithm.

Thus, the terms in Table 6.4 are:

Table A.1: Breakdown of the Helmholtz free energy calculation shown in Table II

Column heading	Free energy term
$\beta f_{EC}$	$\frac{1}{N_{mol}} \sum_{i=1}^N \ln \left( \frac{\beta k_i \Lambda_i^2}{2\pi} \right)^{3/2}$
$\beta(f_{EC}^{CM} - f_{EC})^*$	$-\frac{1}{N_{mol}} \ln \left( \frac{\beta}{2\pi \sum_{i=1}^N \frac{\mu_i^2}{k_i}} \right)^{3/2}$
$\beta \Delta f_{EC \rightarrow C}^{CM}$	$\frac{\beta}{N_{mol}} \int_{\lambda=0}^{\lambda=1} \left\langle \frac{dU}{d\lambda} \right\rangle_{\lambda}^{CM} d\lambda$
$\beta(f_C - f_C^{CM})^*$	$-\frac{1}{N_{mol}} \ln \left( \frac{V}{N_{mol}} \right)$

## A.2 Computing $Q_{EC}^{CM}/Q_{EC}$ :

For an  $N$  atom Einstein crystal system with uniform spring constants ( $k$ ) and distinguishable particles,

$$Q_{EC} = \frac{1}{h^{3N}} Z_{EC} P_{EC} \tag{A.6}$$

where,

$$Z_{EC} = \int d\mathbf{r}^N \prod_{i=1}^N \exp \left[ -\frac{\beta k}{2} \|\mathbf{r}_i\|^2 \right] = \prod_{i=1}^N \left( \frac{2\pi}{\beta k} \right)^{3/2} \quad (\text{A.7a})$$

$$P_{EC} = \int d\mathbf{p}^N \prod_{i=1}^N \exp \left[ -\frac{\beta}{2m_i} \|\mathbf{p}_i\|^2 \right] = \prod_{i=1}^N \left( \frac{2\pi m_i}{\beta} \right)^{3/2} \quad (\text{A.7b})$$

Similarly,

$$Q_{EC}^{CM} = \frac{1}{h^{3(N-1)}} Z_{EC}^{CM} P_{EC}^{CM} \quad (\text{A.8})$$

Note that in the above equation we have  $h^{3(N-1)}$  because with a hard constraint of a fixed center of mass, the system “lives in” a  $6N - 6$  dimensional phase space. Equation (A.8) differs from a harmonically restrained system in  $6N$  dimensions, in which case we would still have  $h^{3N}$ .

$Z_{EC}^{CM}$  and  $P_{EC}^{CM}$  are

$$\begin{aligned} Z_{EC}^{CM} &= \int d\mathbf{r}^N \prod_{i=1}^N \exp \left[ -\frac{\beta k}{2} r_i^2 \right] \delta \left( \sum_{i=1}^N \mu_i \mathbf{r}_i \right) \\ &= \left( \frac{\beta k}{2\pi \sum_{i=1}^N \mu_i^2} \right)^{3/2} \prod_{i=1}^N \left( \frac{2\pi}{\beta k} \right)^{3/2} \\ &= \left( \frac{\beta k}{2\pi \sum_{i=1}^N \mu_i^2} \right)^{3/2} Z_{EC} \end{aligned} \quad (\text{A.9a})$$

$$\begin{aligned} P_{EC}^{CM} &= \int d\mathbf{p}^N \prod_{i=1}^N \exp \left[ -\frac{\beta}{2m_i} p_i^2 \right] \delta \left( \sum_{i=1}^N \mathbf{p}_i \right) \\ &= \left( \frac{\beta}{2\pi M} \right)^{3/2} \prod_{i=1}^N \left( \frac{2\pi m_i}{\beta} \right)^{3/2} \\ &= \left( \frac{\beta}{2\pi M} \right)^{3/2} P_{EC} \end{aligned} \quad (\text{A.9b})$$

where  $M = \sum_{i=1}^N m_i$ .

Therefore,

$$\frac{Z_{EC}^{CM}}{Z_{EC}} = \left( \frac{\beta k}{2\pi \sum_{i=1}^N \mu_i^2} \right)^{3/2} \quad (\text{A.10a})$$

$$\frac{P_{EC}^{CM}}{P_{EC}} = \left( \frac{\beta}{2\pi M} \right)^{3/2} \quad (\text{A.10b})$$

Now, using (A.6) and (A.8),

$$\begin{aligned} \frac{Q_{EC}^{CM}}{Q_{EC}} &= h^3 \left( \frac{Z_{EC}^{CM}}{Z_{EC}} \right) \left( \frac{P_{EC}^{CM}}{P_{EC}} \right) \\ &= \left( \frac{\beta k}{2\pi \sum_{i=1}^N \mu_i^2} \right)^{3/2} \left( \frac{\beta h^2}{2\pi M} \right)^{3/2} \end{aligned} \quad (\text{A.11})$$

### A.3 Computing $Q_C/Q_C^{CM}$ :

$$Q_C = \frac{1}{h^{3N}} Z_C P_C \quad (\text{A.12})$$

where:

$$Z_C = \int d\mathbf{r}^N \prod_{i=1}^N \exp[-\beta U(\mathbf{r}^N)] \quad (\text{A.13a})$$

$$P_C = \int d\mathbf{p}^N \prod_{i=1}^N \exp\left[-\frac{\beta}{2m_i} \|\mathbf{p}_i\|^2\right] = \prod_{i=1}^N \left( \frac{2\pi m_i}{\beta} \right)^{3/2} \quad (\text{A.13b})$$

now using equation (13) from Polson et al.<sup>[?] ],</sup>

$$\begin{aligned}
\frac{Z_C}{Z_C^{CM}} &= \frac{\int d\mathbf{r}^N \prod_{i=1}^N \exp[-\beta U(\mathbf{r}^N)]}{\int d\mathbf{r}^N \prod_{i=1}^N \exp[-\beta U(\mathbf{r}^N)] \delta\left(\sum_{i=1}^N \mu_i \mathbf{r}_i\right)} \\
&= \frac{1}{\left\langle \delta\left(\sum_{i=1}^N \mu_i \mathbf{r}_i\right) \right\rangle} \\
&= \frac{1}{\mathcal{O}(\mathbf{r}_{CM} = 0)} \\
&= \frac{V}{N_{mol}}
\end{aligned} \tag{A.14}$$

Note<sup>[?] ]</sup> that  $\mathcal{O}(\mathbf{r}_{CM})$  is the probability distribution function of the center of mass. Since, the probability distribution of the center of mass of the lattice is evenly distributed over a volume equal to that of the Wigner–Seitz cell of the lattice positioned at the center of the volume over which we carry out the integration in the partition function. It follows for one molecule per Wigner-Seitz cell:  $\mathcal{O}(\mathbf{r}_{CM} = 0) = N_{mol}/V$ .

Since  $P_C = P_{EC}$  we have from equation (A.10b)

$$\frac{P_C}{P_C^{CM}} = \left(\frac{\beta}{2\pi M}\right)^{-3/2} \tag{A.15}$$

Therefore,

$$\begin{aligned}
\frac{Q_C}{Q_C^{CM}} &= h^{-3} \left(\frac{Z_C}{Z_C^{CM}}\right) \left(\frac{P_C}{P_C^{CM}}\right) \\
&= \left(\frac{V}{N_{mol}}\right) \left(\frac{\beta h^2}{2\pi M}\right)^{-3/2}
\end{aligned} \tag{A.16}$$

## A.4 Free energy of the crystal:

Using equations (A.11) and (A.16), we have

$$\begin{aligned}
 \beta F_C = & \sum_{i=1}^N \ln \left( \frac{\beta k \Lambda_i^2}{2\pi} \right)^{3/2} \\
 & - \ln \left[ \left( \frac{\beta k}{2\pi \sum_{i=1}^N \mu_i^2} \right)^{3/2} \left( \frac{\beta h^2}{2\pi M} \right)^{3/2} \right] \\
 & + \beta \int_{\lambda=0}^{\lambda=1} \left\langle \frac{dU}{d\lambda} \right\rangle_{\lambda} d\lambda \\
 & - \ln \left[ \left( \frac{V}{N_{mol}} \right) \left( \frac{\beta h^2}{2\pi M} \right)^{-3/2} \right]
 \end{aligned} \tag{A.17a}$$

$$\begin{aligned}
 \beta F_C = & \sum_{i=1}^N \ln \left( \frac{\beta k \Lambda_i^2}{2\pi} \right)^{3/2} \\
 & + \beta \int_{\lambda=0}^{\lambda=1} \left\langle \frac{dU}{d\lambda} \right\rangle_{\lambda}^{CM} d\lambda \\
 & - \ln \left[ \left( \frac{\beta k}{2\pi \sum_{i=1}^N \mu_i^2} \right)^{3/2} \left( \frac{V}{N_{mol}} \right) \right]
 \end{aligned} \tag{A.17b}$$

where,  $\Lambda_i = (\beta h^2 / (2\pi m_i))^{1/2}$

Note that the log terms in equation (A.17b) have been combined to give a dimensionless argument to the logarithms.

# Appendix B

## Ensuring zero net external force due to springs

The equation describing the displacement in x-direction of a simple harmonic oscillator (SHO)  $i$  is,

$$\mathbf{x}_i(t) = A_i \sin(\omega_i t + \phi_i) \quad (\text{B.1})$$

where  $\mathbf{x}_i(t)$  is the displacement of SHO in x-direction from the tether position,  $A_i$  is the amplitude,  $\omega_i = \sqrt{k_i/m_i}$  is the angular frequency of SHO, and  $\phi_i$  is the phase.

Taking the time derivative of equation (B.1), we get

$$\mathbf{v}_i(t) = A_i \omega_i \cos(\omega_i t + \phi_i) \quad (\text{B.2})$$

where  $\mathbf{v}_i(t)$  is the velocity at time  $t$ .

Using equation (B.2), and plugging in  $t = 0$  gives a relation between  $A_i$ ,  $\omega_i$ ,  $\phi_i$ , and the initial velocity ( $\mathbf{v}_i^o$ ):

$$A_i = \frac{\mathbf{v}_i^o}{\omega_i \cos(\phi_i)} \quad (\text{B.3})$$

Therefore,

$$\mathbf{x}_i(t) = \frac{\mathbf{v}_i^o}{\omega_i \cos(\phi_i)} \sin(\omega_i t + \phi_i) \quad (\text{B.4})$$

Now, the net force on the system ( $\mathbf{F}_{CM}$ ) is

$$\begin{aligned} \mathbf{F}_{CM} &= - \sum_{i=1}^N k_i \mathbf{x}_i \\ &= - \sum_{i=1}^N k_i \frac{\mathbf{v}_i^o}{\omega_i \cos(\phi_i)} \sin(\omega_i t + \phi_i) \\ &= - \sum_{i=1}^N k_i \frac{\mathbf{v}_i^o}{\omega_i \cos(\phi_i)} [\sin(\omega_i t) \cos(\phi_i) + \cos(\omega_i t) \sin(\phi_i)] \\ &= - \sum_{i=1}^N \omega_i \frac{m_i \mathbf{v}_i^o}{\cos(\phi_i)} [\sin(\omega_i t) \cos(\phi_i) + \cos(\omega_i t) \sin(\phi_i)] \end{aligned} \quad (\text{B.5})$$

Now if  $\omega_i = \omega$ , i.e.,  $k_i/m_i = \omega^2 = \text{constant}$ , we have

$$\begin{aligned} \mathbf{F}_{CM} &= -\omega \sin(\omega t) \sum_{i=1}^N m_i \mathbf{v}_i^o - \omega \cos(\omega t) \sum_{i=1}^N m_i \mathbf{v}_i^o \tan(\phi_i) \\ &= -\omega M \mathbf{v}_{CM}^o \sin(\omega t) - \omega^2 M \mathbf{x}_{CM}^o \cos(\omega t) \end{aligned} \quad (\text{B.6})$$

Where we have used equation (B.4) to write second term in terms of  $\mathbf{x}_{CM}^o$ , the initial displacement of the center of mass from the tether points' center of mass given by

$$\mathbf{x}_{CM}^o = \frac{1}{M} \sum_{i=1}^N m_i \frac{\mathbf{v}_{CM}^o}{\omega} \tan(\phi_i) \quad (\text{B.7})$$

Thus, if we initialize the system such that  $\mathbf{v}_{CM}^o = 0$  and  $\mathbf{x}_{CM}^o = 0$ , then the center of mass remains fixed  $\forall t$ .

Therefore, a *straightforward* way to ensure zero net external force on a system due to springs is to i) make spring constants proportional to atomic masses such that all oscillators have the same angular frequency, ii) initialize the system with zero center

of mass velocity, and iii) initialize the system with all SHOs on their tethered (lattice) positions.

## Equation describing the time evolution of the center of mass

For a system having springs with different spring constants such that  $k_i/m_i$  is constant, the center of mass is given by:

$$\begin{aligned}
\mathbf{x}_{CM}(t) &= \frac{1}{M} \sum_{i=1}^N m_i \mathbf{x}_i \\
&= \frac{1}{M} \sum_{i=1}^N m_i A_i \sin(\omega t + \phi_i) \\
&= \frac{1}{M} \sum_{i=1}^N m_i \frac{\mathbf{v}_i^o}{\omega \cos(\phi_i)} \sin(\omega t + \phi_i) \\
&= \frac{1}{M} \sum_{i=1}^N m_i \frac{\mathbf{v}_i^o}{\omega \cos(\phi_i)} [\sin(\omega t) \cos(\phi_i) + \cos(\omega t) \sin(\phi_i)] \quad (\text{B.8}) \\
&= \omega^{-1} \sin(\omega t) \left( \frac{1}{M} \sum_{i=1}^N m_i \mathbf{v}_i^o \right) \\
&\quad + \cos(\omega t) \left( \frac{1}{M} \sum_{i=1}^N m_i \frac{\mathbf{v}_i^o}{\omega} \tan(\phi_i) \right) \\
&= \omega^{-1} \mathbf{v}_{CM}^o \sin(\omega t) + \mathbf{x}_{CM}^o \cos(\omega t)
\end{aligned}$$

Note that  $M\ddot{\mathbf{x}}_{CM}(t) = \mathbf{F}_{CM}$ . Therefore equation (B.8) can be used to get the same result as obtained in equation (B.6).

Note that during T.I. when  $0 < \lambda < 1$ , we apply the force field (FF) on the atoms in addition to the spring forces. Since,  $\sum_{i=atoms} \mathbf{F}_{i,FF} = 0$  (where  $\mathbf{F}_{i,FF}$  is the force on atom  $i$  due to the force field), no additional force is exerted on the center of mass due to it, therefore, equations (B.6) and (B.8) are still applicable.

# Appendix C

## T.I. plots

### C.1 T.I. integrand plot & Data for LiI crystal

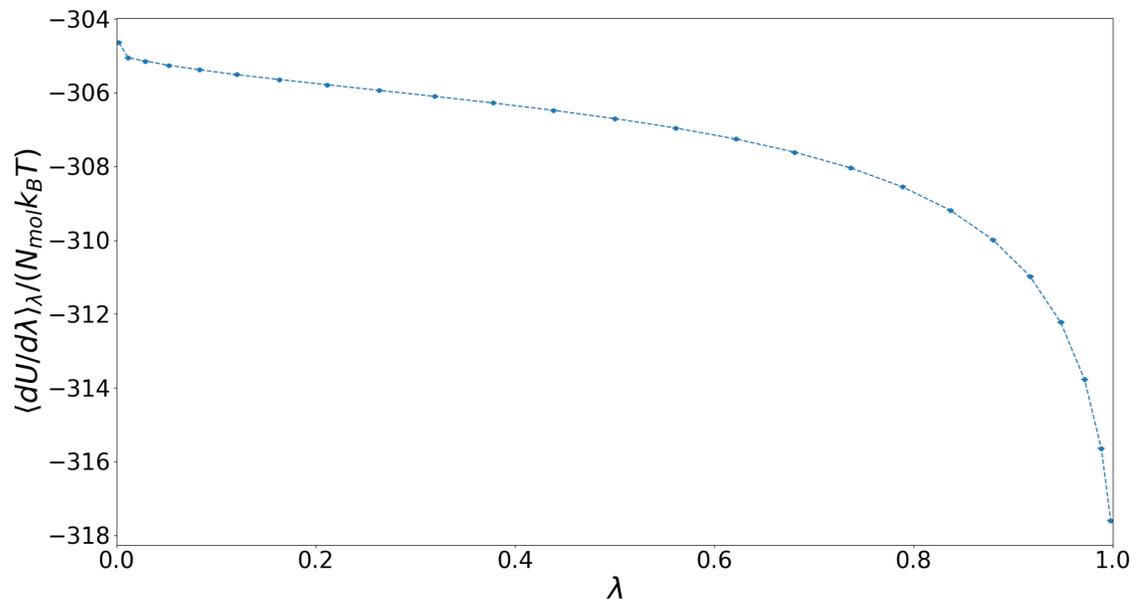


Figure C.1: T.I integrand plot for LiI crystal (w/ mass-scaled springs)

Table C.1: T.I. data for LiI (w/ mass-scaled springs)

$i$	$\lambda_i$	$w_i$	$\beta \langle du/d\lambda \rangle_\lambda$
1	0.00222	0.0057	-304.63627 $\pm$ 0.00095
2	0.01167	0.01318	-305.0528 $\pm$ 0.00093
3	0.02851	0.02047	-305.14515 $\pm$ 0.00088
4	0.0525	0.02745	-305.25958 $\pm$ 0.00087
5	0.08328	0.03402	-305.38006 $\pm$ 0.00081
6	0.12037	0.04007	-305.5126 $\pm$ 0.00077
7	0.16322	0.04551	-305.64401 $\pm$ 0.00076
8	0.21117	0.05027	-305.78636 $\pm$ 0.00074
9	0.2635	0.05426	-305.93919 $\pm$ 0.00074
10	0.31941	0.05743	-306.10326 $\pm$ 0.00076
11	0.37807	0.05973	-306.28122 $\pm$ 0.00078
12	0.43857	0.06112	-306.48054 $\pm$ 0.00082
13	0.5	0.06159	-306.70411 $\pm$ 0.00086
14	0.56143	0.06112	-306.96066 $\pm$ 0.00093
15	0.62193	0.05973	-307.25967 $\pm$ 0.00102
16	0.68059	0.05743	-307.61391 $\pm$ 0.00113
17	0.7365	0.05426	-308.03824 $\pm$ 0.00129
18	0.78883	0.05027	-308.55432 $\pm$ 0.00148
19	0.83678	0.04551	-309.18988 $\pm$ 0.00172
20	0.87963	0.04007	-309.98973 $\pm$ 0.00208
21	0.91672	0.03402	-310.97968 $\pm$ 0.00253
22	0.9475	0.02745	-312.22696 $\pm$ 0.00314
23	0.97149	0.02047	-313.77678 $\pm$ 0.004
24	0.98833	0.01318	-315.64407 $\pm$ 0.00534
25	0.99778	0.0057	-317.6001 $\pm$ 0.00761

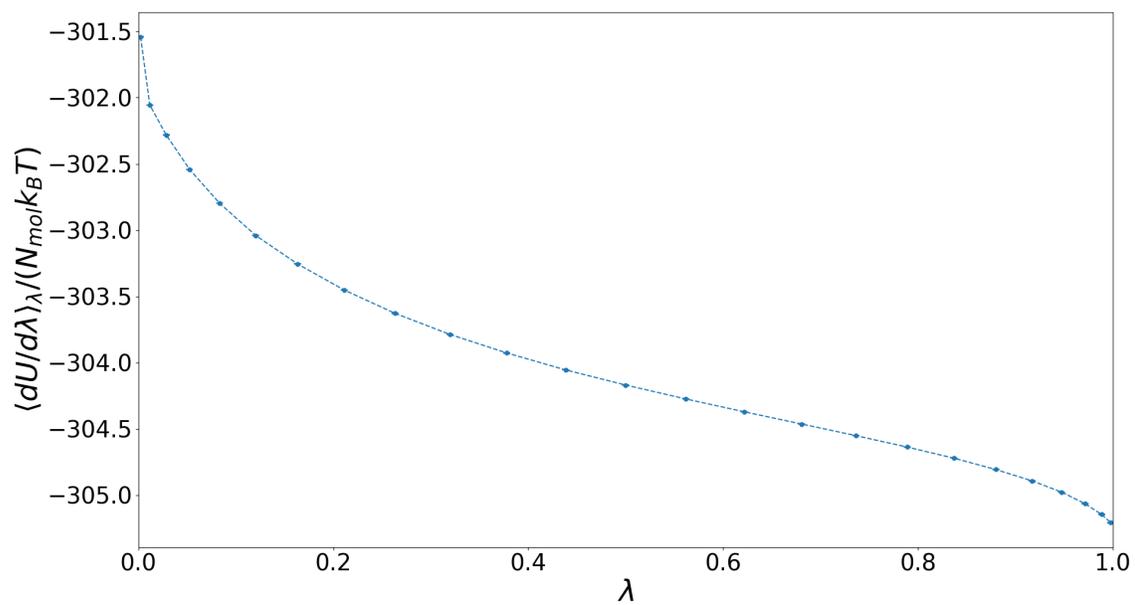


Figure C.2: T.I integrand plot for LiI crystal (w/ MSD-based springs)

Table C.2: T.I. data for LiI (w/ MSD-based springs)

$i$	$\lambda_i$	$w_i$	$\beta \langle du/d\lambda \rangle_\lambda$
1	0.00222	0.0057	-301.54089 $\pm$ 0.00178
2	0.01167	0.01318	-302.0557 $\pm$ 0.00162
3	0.02851	0.02047	-302.28181 $\pm$ 0.00147
4	0.0525	0.02745	-302.54136 $\pm$ 0.0013
5	0.08328	0.03402	-302.79664 $\pm$ 0.00118
6	0.12037	0.04007	-303.04005 $\pm$ 0.00106
7	0.16322	0.04551	-303.25461 $\pm$ 0.00094
8	0.21117	0.05027	-303.4518 $\pm$ 0.00085
9	0.2635	0.05426	-303.62807 $\pm$ 0.00077
10	0.31941	0.05743	-303.78601 $\pm$ 0.00071
11	0.37807	0.05973	-303.92669 $\pm$ 0.00065
12	0.43857	0.06112	-304.05476 $\pm$ 0.00061
13	0.5	0.06159	-304.16861 $\pm$ 0.00059
14	0.56143	0.06112	-304.27293 $\pm$ 0.00057
15	0.62193	0.05973	-304.3709 $\pm$ 0.00055
16	0.68059	0.05743	-304.46273 $\pm$ 0.00054
17	0.7365	0.05426	-304.55084 $\pm$ 0.00055
18	0.78883	0.05027	-304.63638 $\pm$ 0.00057
19	0.83678	0.04551	-304.72041 $\pm$ 0.0006
20	0.87963	0.04007	-304.80548 $\pm$ 0.00063
21	0.91672	0.03402	-304.89086 $\pm$ 0.0007
22	0.9475	0.02745	-304.97683 $\pm$ 0.0008
23	0.97149	0.02047	-305.06433 $\pm$ 0.0009
24	0.98833	0.01318	-305.14504 $\pm$ 0.00102
25	0.99778	0.0057	-305.20683 $\pm$ 0.00117

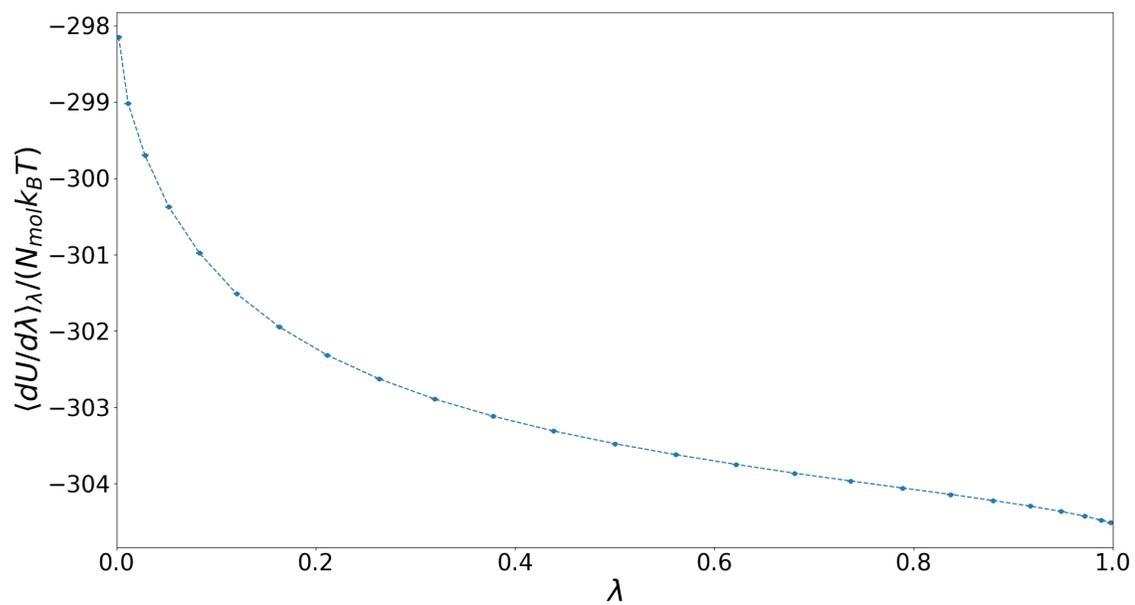


Figure C.3: T.I integrand plot for LiI crystal (w/ equal springs)

Table C.3: T.I. data for LiI (w/ equal springs)

$i$	$\lambda_i$	$w_i$	$\beta \langle du/d\lambda \rangle_\lambda$
1	0.00222	0.0057	-298.14578 $\pm$ 0.00343
2	0.01167	0.01318	-299.01976 $\pm$ 0.00302
3	0.02851	0.02047	-299.69576 $\pm$ 0.00249
4	0.0525	0.02745	-300.37508 $\pm$ 0.0021
5	0.08328	0.03402	-300.97966 $\pm$ 0.00177
6	0.12037	0.04007	-301.50976 $\pm$ 0.0015
7	0.16322	0.04551	-301.94525 $\pm$ 0.00133
8	0.21117	0.05027	-302.31415 $\pm$ 0.00113
9	0.2635	0.05426	-302.62698 $\pm$ 0.001
10	0.31941	0.05743	-302.89214 $\pm$ 0.0009
11	0.37807	0.05973	-303.1163 $\pm$ 0.0008
12	0.43857	0.06112	-303.30992 $\pm$ 0.00073
13	0.5	0.06159	-303.47602 $\pm$ 0.00068
14	0.56143	0.06112	-303.62078 $\pm$ 0.00065
15	0.62193	0.05973	-303.74829 $\pm$ 0.00062
16	0.68059	0.05743	-303.86337 $\pm$ 0.0006
17	0.7365	0.05426	-303.964 $\pm$ 0.00059
18	0.78883	0.05027	-304.05703 $\pm$ 0.00058
19	0.83678	0.04551	-304.14066 $\pm$ 0.00059
20	0.87963	0.04007	-304.22022 $\pm$ 0.00061
21	0.91672	0.03402	-304.29352 $\pm$ 0.00064
22	0.9475	0.02745	-304.36287 $\pm$ 0.00068
23	0.97149	0.02047	-304.42474 $\pm$ 0.00074
24	0.98833	0.01318	-304.47781 $\pm$ 0.00082
25	0.99778	0.0057	-304.51219 $\pm$ 0.00087

## C.2 T.I. integrand plot & Data for NaCl crystal

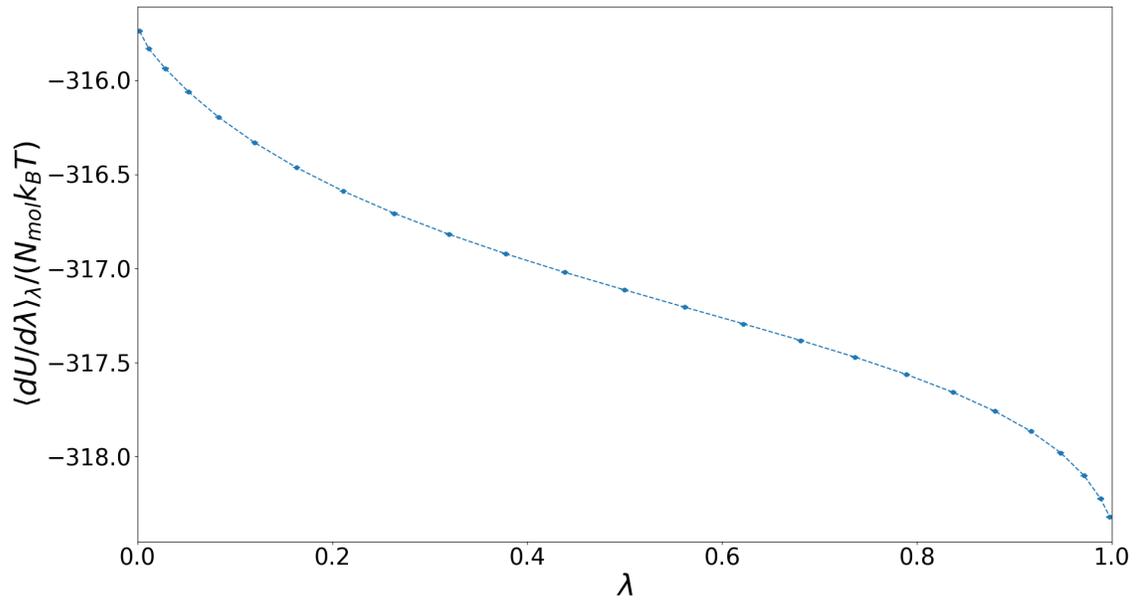


Figure C.4: T.I integrand plot for NaCl crystal (w/ mass-scaled springs)

Table C.4: T.I. data for NaCl (w/ mass-scaled springs)

$i$	$\lambda_i$	$w_i$	$\beta \langle du/d\lambda \rangle_\lambda$
1	0.00222	0.0057	-315.73798 $\pm$ 0.00116
2	0.01167	0.01318	-315.8311 $\pm$ 0.00111
3	0.02851	0.02047	-315.93651 $\pm$ 0.00104
4	0.0525	0.02745	-316.06102 $\pm$ 0.00096
5	0.08328	0.03402	-316.19545 $\pm$ 0.00088
6	0.12037	0.04007	-316.33113 $\pm$ 0.00081
7	0.16322	0.04551	-316.46303 $\pm$ 0.00074
8	0.21117	0.05027	-316.58845 $\pm$ 0.00069
9	0.2635	0.05426	-316.7066 $\pm$ 0.00064
10	0.31941	0.05743	-316.81745 $\pm$ 0.00061
11	0.37807	0.05973	-316.9217 $\pm$ 0.00058
12	0.43857	0.06112	-317.02025 $\pm$ 0.00056
13	0.5	0.06159	-317.11347 $\pm$ 0.00078
14	0.56143	0.06112	-317.2051 $\pm$ 0.00054
15	0.62193	0.05973	-317.29391 $\pm$ 0.00054
16	0.68059	0.05743	-317.3821 $\pm$ 0.00055
17	0.7365	0.05426	-317.47112 $\pm$ 0.00057
18	0.78883	0.05027	-317.56238 $\pm$ 0.0006
19	0.83678	0.04551	-317.65739 $\pm$ 0.00064
20	0.87963	0.04007	-317.75764 $\pm$ 0.00071
21	0.91672	0.03402	-317.86456 $\pm$ 0.00079
22	0.9475	0.02745	-317.97925 $\pm$ 0.00091
23	0.97149	0.02047	-318.10113 $\pm$ 0.00108
24	0.98833	0.01318	-318.22324 $\pm$ 0.0013
25	0.99778	0.0057	-318.32016 $\pm$ 0.00152

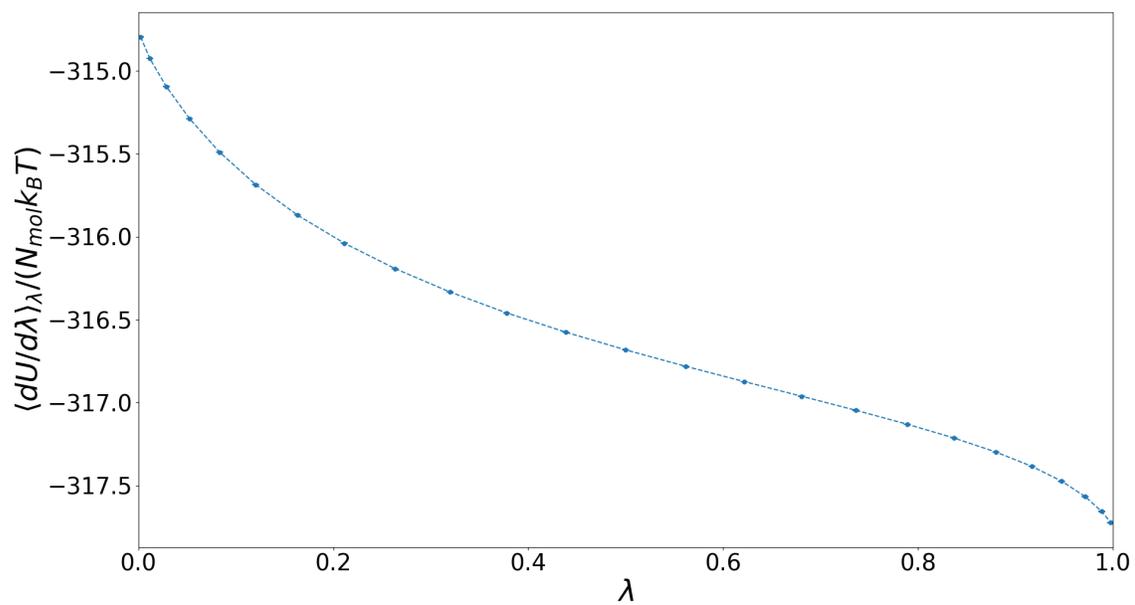


Figure C.5: T.I integrand plot for NaCl crystal (w/ MSD-based springs)

Table C.5: T.I. data for NaCl (w/ MSD-based springs)

$i$	$\lambda_i$	$w_i$	$\beta \langle du/d\lambda \rangle_\lambda$
1	0.00222	0.0057	-314.79526 $\pm$ 0.00154
2	0.01167	0.01318	-314.92536 $\pm$ 0.00145
3	0.02851	0.02047	-315.09434 $\pm$ 0.00133
4	0.0525	0.02745	-315.28877 $\pm$ 0.0012
5	0.08328	0.03402	-315.49023 $\pm$ 0.00108
6	0.12037	0.04007	-315.68599 $\pm$ 0.00097
7	0.16322	0.04551	-315.86963 $\pm$ 0.00088
8	0.21117	0.05027	-316.03855 $\pm$ 0.0008
9	0.2635	0.05426	-316.19249 $\pm$ 0.00073
10	0.31941	0.05743	-316.33221 $\pm$ 0.00068
11	0.37807	0.05973	-316.45929 $\pm$ 0.00064
12	0.43857	0.06112	-316.57518 $\pm$ 0.0006
13	0.5	0.06159	-316.6817 $\pm$ 0.00058
14	0.56143	0.06112	-316.78046 $\pm$ 0.00056
15	0.62193	0.05973	-316.87307 $\pm$ 0.00055
16	0.68059	0.05743	-316.9611 $\pm$ 0.00055
17	0.7365	0.05426	-317.04607 $\pm$ 0.00055
18	0.78883	0.05027	-317.1294 $\pm$ 0.00057
19	0.83678	0.04551	-317.21249 $\pm$ 0.0006
20	0.87963	0.04007	-317.2967 $\pm$ 0.00064
21	0.91672	0.03402	-317.38332 $\pm$ 0.0007
22	0.9475	0.02745	-317.4732 $\pm$ 0.0008
23	0.97149	0.02047	-317.56563 $\pm$ 0.00093
24	0.98833	0.01318	-317.65477 $\pm$ 0.00109
25	0.99778	0.0057	-317.72238 $\pm$ 0.00125

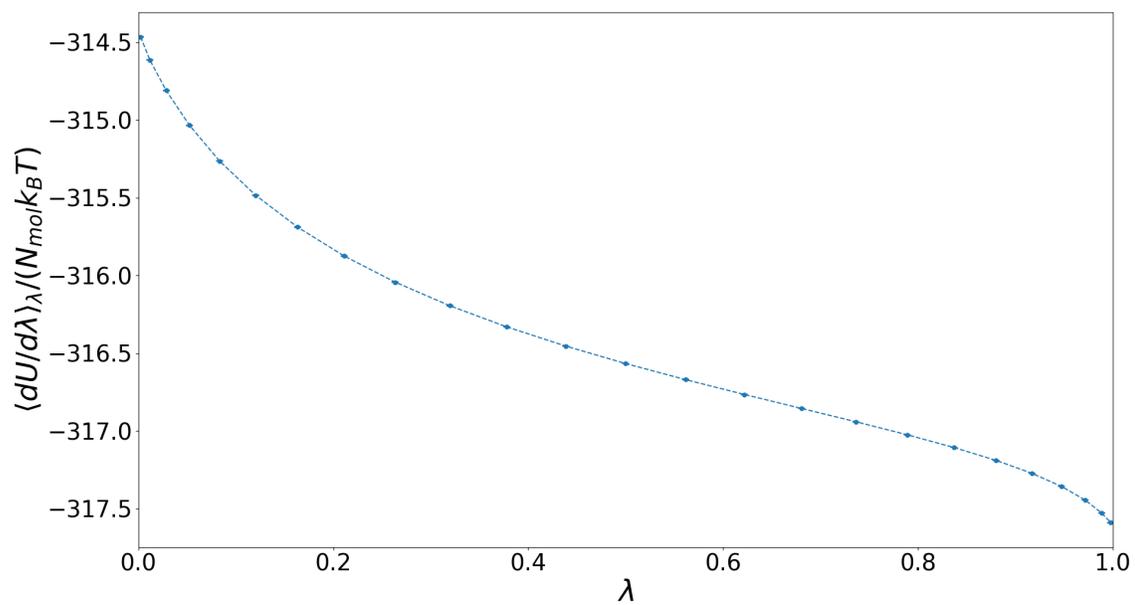


Figure C.6: T.I. integrand plot for NaCl crystal (w/ equal springs)

Table C.6: T.I. data for NaCl (w/ equal springs)

$i$	$\lambda_i$	$w_i$	$\beta \langle du/d\lambda \rangle_\lambda$
1	0.00222	0.0057	-314.46466 ± 0.00168
2	0.01167	0.01318	-314.61311 ± 0.00157
3	0.02851	0.02047	-314.81036 ± 0.00143
4	0.0525	0.02745	-315.03466 ± 0.00128
5	0.08328	0.03402	-315.26395 ± 0.00114
6	0.12037	0.04007	-315.4841 ± 0.00103
7	0.16322	0.04551	-315.68848 ± 0.00093
8	0.21117	0.05027	-315.8747 ± 0.00084
9	0.2635	0.05426	-316.04292 ± 0.00076
10	0.31941	0.05743	-316.19427 ± 0.00071
11	0.37807	0.05973	-316.33077 ± 0.00066
12	0.43857	0.06112	-316.45415 ± 0.00062
13	0.5	0.06159	-316.56652 ± 0.00059
14	0.56143	0.06112	-316.66968 ± 0.00057
15	0.62193	0.05973	-316.76542 ± 0.00056
16	0.68059	0.05743	-316.85542 ± 0.00055
17	0.7365	0.05426	-316.94128 ± 0.00055
18	0.78883	0.05027	-317.02446 ± 0.00057
19	0.83678	0.04551	-317.1064 ± 0.00059
20	0.87963	0.04007	-317.18845 ± 0.00063
21	0.91672	0.03402	-317.27186 ± 0.00069
22	0.9475	0.02745	-317.35747 ± 0.00077
23	0.97149	0.02047	-317.44453 ± 0.0009
24	0.98833	0.01318	-317.52746 ± 0.00105
25	0.99778	0.0057	-317.58957 ± 0.0012

# Appendix D

## Error Calculation

At each  $\lambda_i$  the derivative,  $\langle du/d\lambda_i \rangle_\lambda$  is computed using independent samples. See auto-correlation subsection for details.

$$\begin{aligned} \left\langle \frac{du}{d\lambda} \right\rangle_\lambda &= \overline{du/d\lambda} + \sigma_{\overline{du/d\lambda}} \\ &= \overline{du/d\lambda} + \frac{\sigma_{du/d\lambda}}{\sqrt{N_{data}}} \end{aligned} \tag{D.1}$$

where  $\overline{du/d\lambda}$  is the sample mean,  $\sigma_{\overline{du/d\lambda}}$  is the standard error in the mean,  $\sigma_{du/d\lambda}$  is the sample standard deviation, and  $N_{data}$  is the number of independent samples used in computing the sample mean.

The free energy difference ( $\Delta f$ ) is

$$\Delta f = \sum_{i=1}^{25} w_i \left\langle \frac{du}{d\lambda} \right\rangle_{\lambda_i} \tag{D.2}$$

However, the estimate (mean) of the free energy difference ( $\overline{\Delta f}$ ) is

$$\overline{\Delta f} = \sum_{i=1}^{25} w_i \overline{\frac{du}{d\lambda_i}} \tag{D.3}$$

and the error in the estimate (mean) of the free energy difference is

$$\sigma_{\Delta f} = \sqrt{\sum_{i=1}^{25} w_i^2 \sigma_{\frac{du}{d\lambda_i}}^2} \quad (\text{D.4})$$

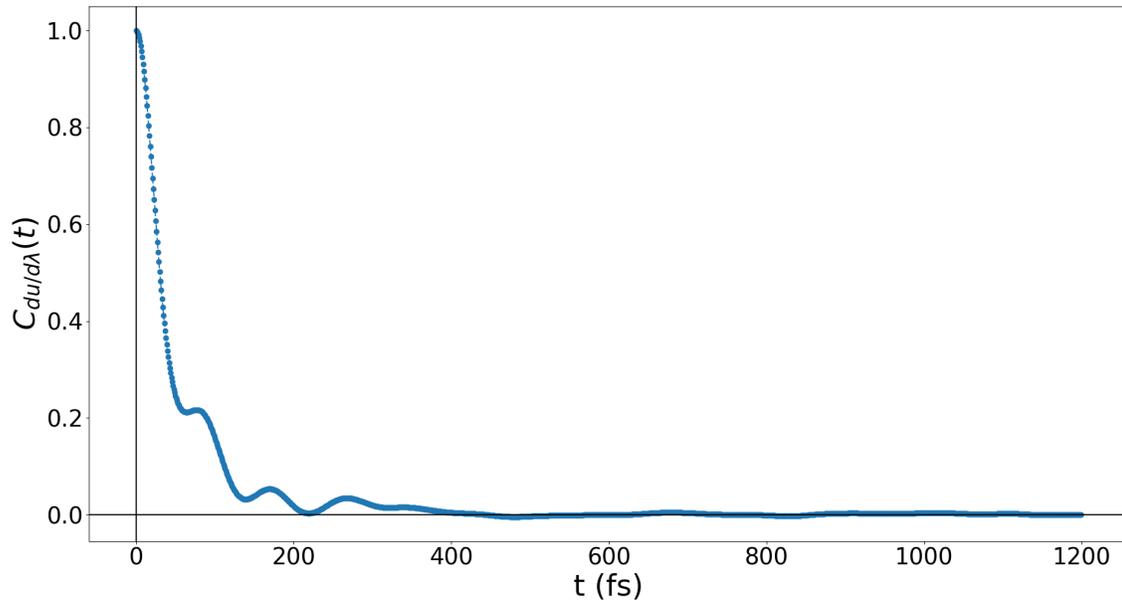
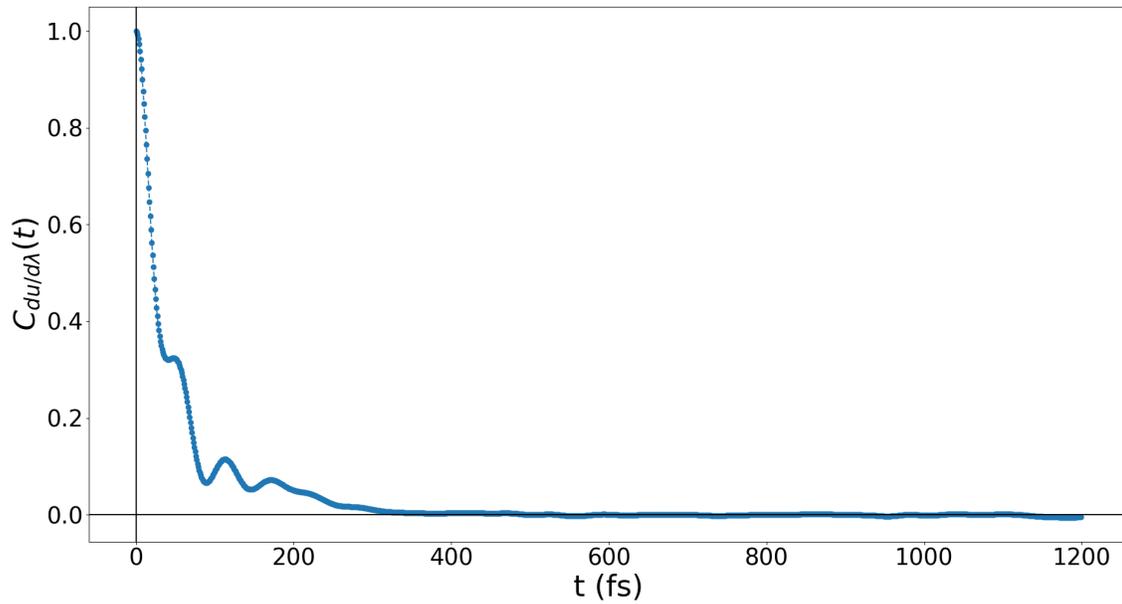
## D.1 Autocorrelation time

We compute the autocorrelation time for  $du/d\lambda$  ( $\tau_{du/d\lambda}$ ) using the time autocorrelation function, where:

$$\tau_{du/d\lambda} = \int_{t=0}^{t=\infty} C_{du/d\lambda} dt \quad (\text{D.5})$$

Table D.1: Autocorrelation times of  $du/d\lambda$  for LiI (w/ MSD springs) at three  $\lambda$  values.

$\lambda$	$\tau_{du/d\lambda}$ (fs)
0.00222	51
0.5	47
0.997778	401

Figure D.1: Autocorrelation function at  $\lambda = 0.00222$ Figure D.2: Autocorrelation function at  $\lambda = 0.5$

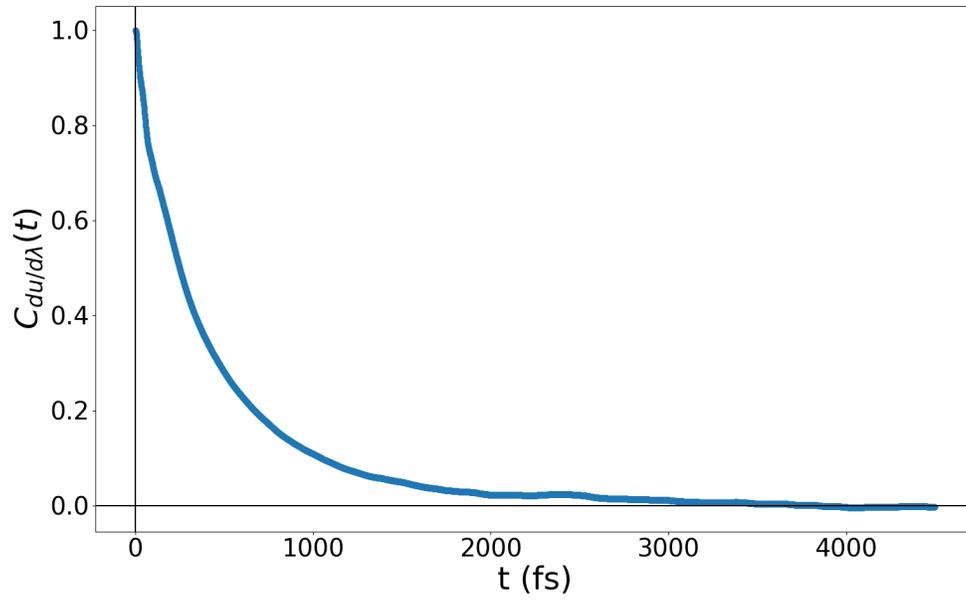


Figure D.3: Autocorrelation function at  $\lambda = 0.99778$

# Appendix E

## Force Field Predicted Crystal Structures

### Naphthalene

Table E.1 shows a comparison of the force field predicted lattice parameters of naphthalene against the experimental ones. Also, Figures E.1, E.2,, and E.3, show the force field predicted unit cell superimposed on the experimental unit cell from the x, y, and z directional views, respectively. As seen from these predictions OPLS-AA does a good job reproducing the naphthalene crystal structure.

Table E.1: OPLS-AA predicted lattice parameters of naphthalene crystal

Lattice parameter	Expt.	OPLS-AA	% $\Delta$
$a$ (Å)	$8.213 \pm 0.002$	$8.278 \pm 0.001$	0.79%
$b$ (Å)	$5.973 \pm 0.001$	$5.991 \pm 0.001$	0.30%
$c$ (Å)	$8.675 \pm 0.002$	$8.565 \pm 0.001$	-1.27%
$\alpha$ (°)	$90.00 \pm 0.00$	$89.998 \pm 0.004$	0.00%
$\beta$ (°)	$123.39 \pm 0.02$	$119.14 \pm 0.01$	-3.44%
$\gamma$ (°)	$90.00 \pm 0.00$	$90.005 \pm 0.003$	0.01%

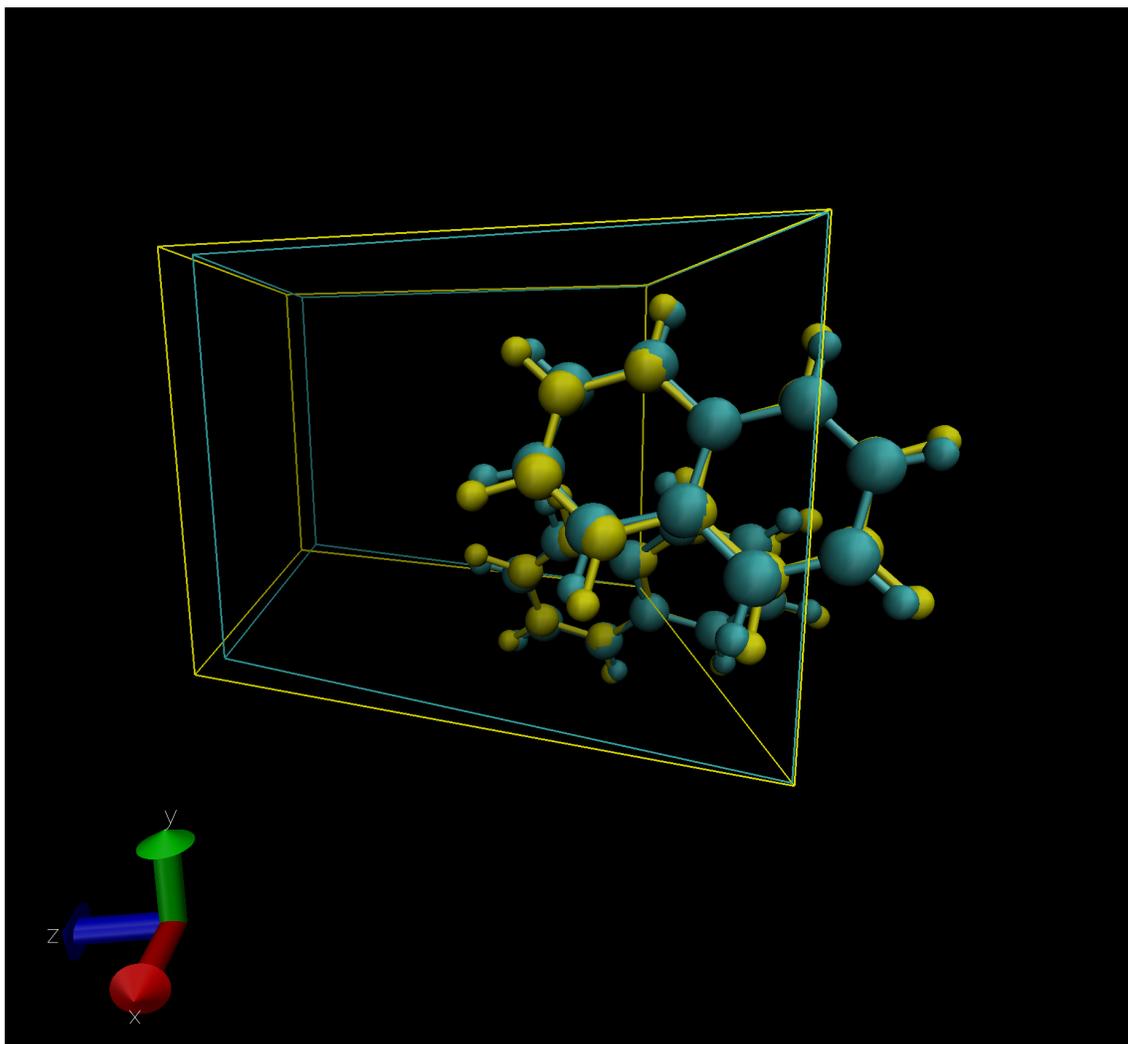


Figure E.1: OPLS-AA predicted crystal structure (yellow) of naphthalene superimposed on the experimental structure (cyan). X-direction view.

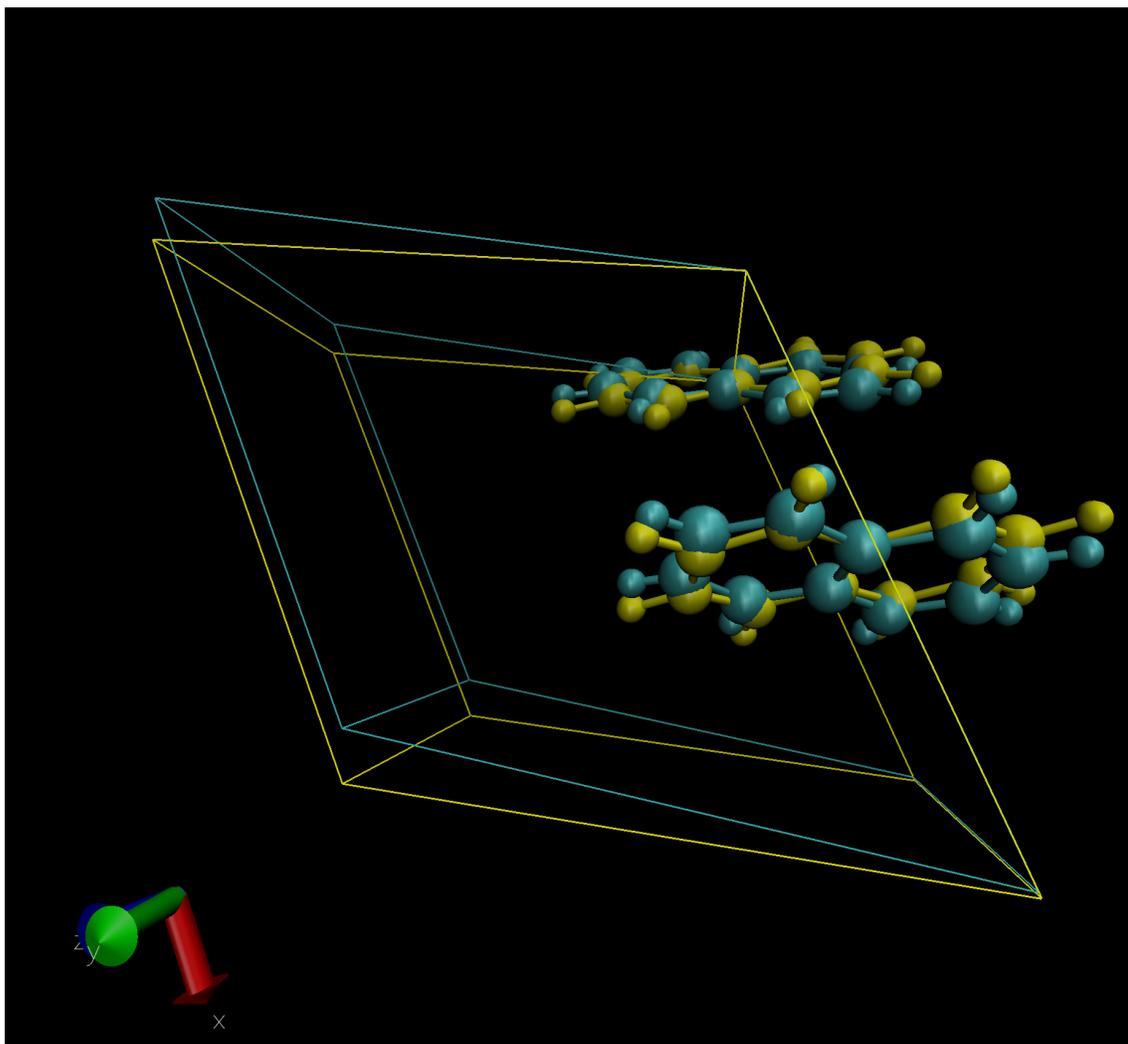


Figure E.2: OPLS-AA predicted crystal structure (yellow) of naphthalene superimposed on the experimental structure (cyan). Y-direction view.

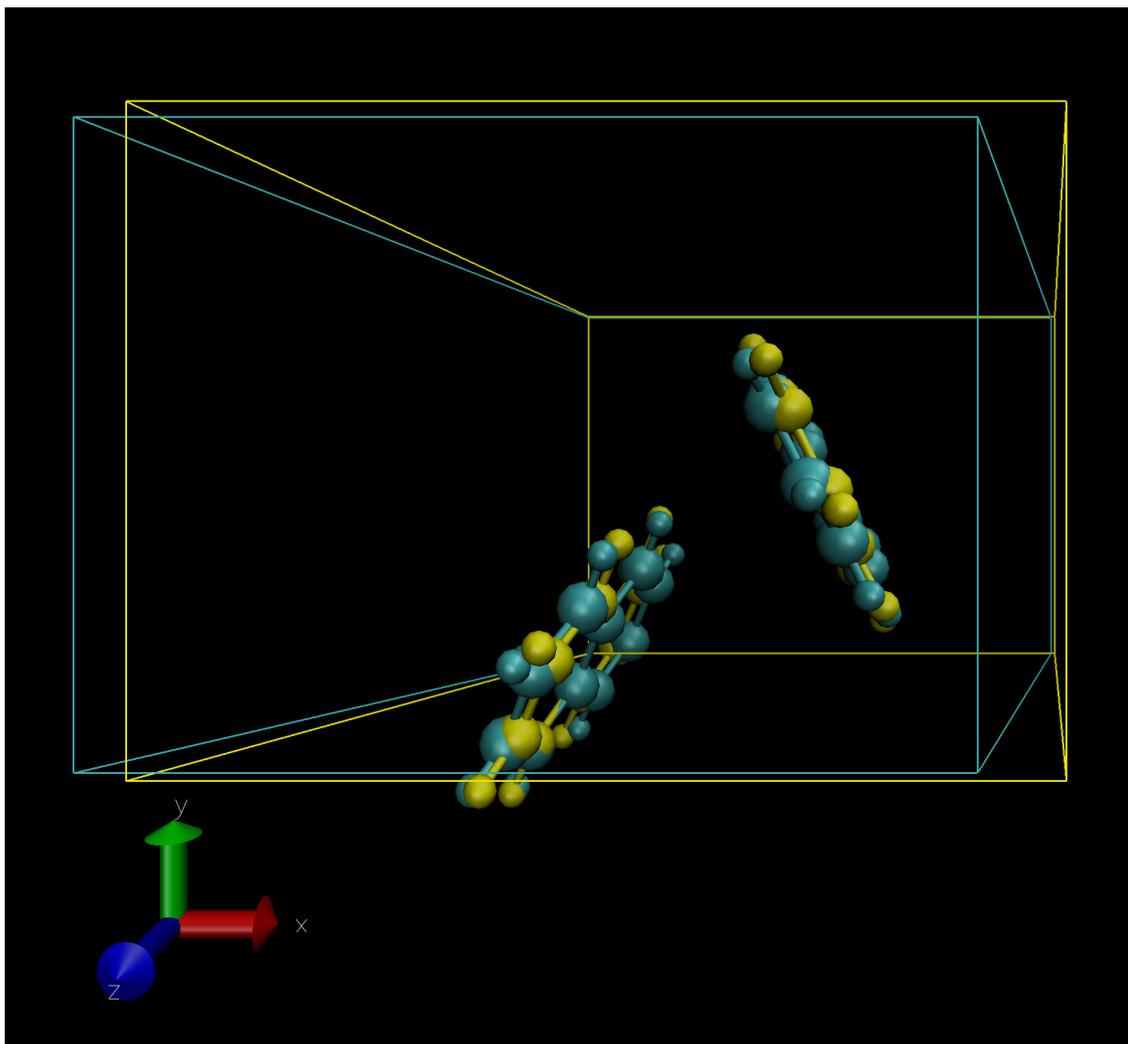


Figure E.3: OPLS-AA predicted crystal structure (yellow) of naphthalene superimposed on the experimental structure (cyan). Z-direction view.

### Succinic acid ( $\beta$ and $\gamma$ polymorphs)

Tables E.2 and E.3 show a comparison of the force field predicted lattice parameters against the experimental ones. Also, Figures E.4, E.5, E.6, Figures E.7, E.8, and E.9, show the force field predicted unit cells superimposed on the experimental unit cells from the x, y, and z directional views, for  $\beta$  and  $\gamma$  polymorphs, respectively. As seen from these predictions GAFF does a good job reproducing the succinic acid crystal structure for the

$\beta$  polymorph, however, it doesn't accurately predict the structure of the  $\gamma$  polymorph.

Table E.2: GAFF predicted lattice parameters of succinic acid crystal —  $\beta$  polymorph

Lattice parameter	Expt.	GAFF	% $\Delta$
$a(\text{\AA})$	$5.519 \pm 0.002$	$5.473 \pm 0.001$	-0.83%
$b(\text{\AA})$	$8.862 \pm 0.006$	$9.068 \pm 0.001$	2.32%
$c(\text{\AA})$	$5.101 \pm 0.001$	$5.314 \pm 0.001$	4.18%
$\alpha$ ( $^\circ$ )	$90.00 \pm 0.00$	$89.999 \pm 0.003$	0.00%
$\beta$ ( $^\circ$ )	$91.59 \pm 0.04$	$89.624 \pm 0.004$	-2.15%
$\gamma$ ( $^\circ$ )	$90.00 \pm 0.00$	$89.997 \pm 0.003$	0.00%

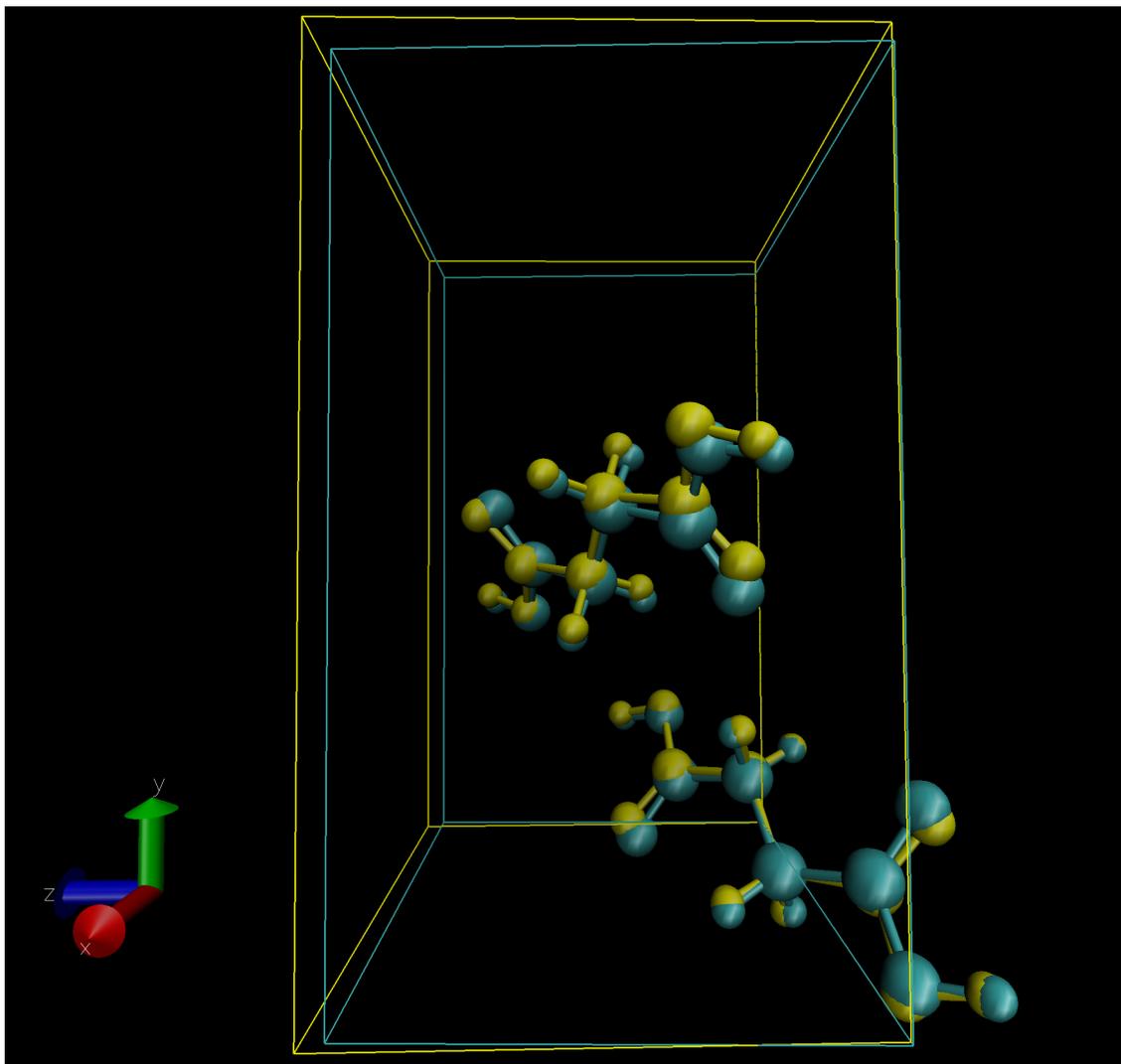


Figure E.4: GAFF predicted crystal structure (yellow) of  $\beta$ -succinic acid superimposed on the experimental structure (cyan). x-direction view.

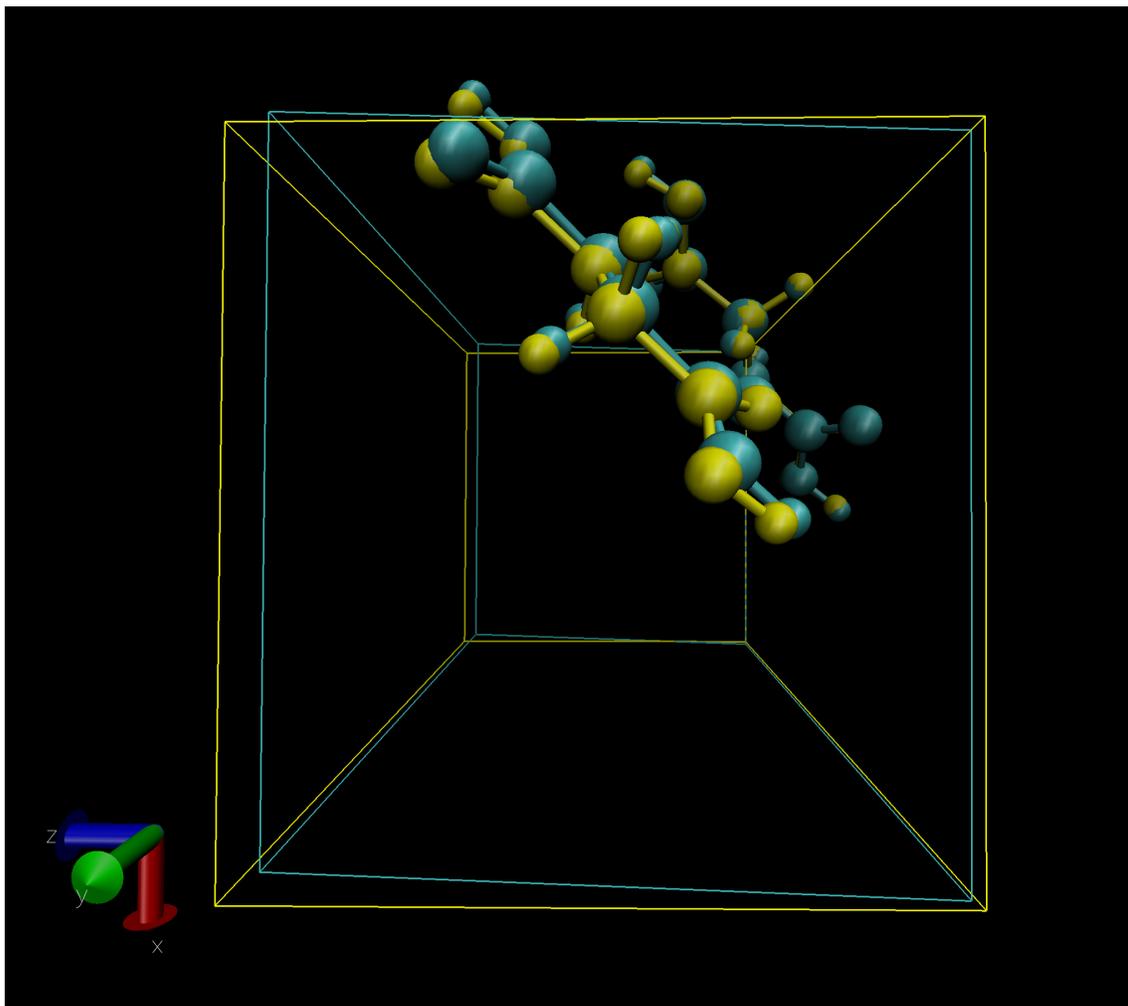


Figure E.5: GAFF predicted crystal structure (yellow) of  $\beta$ -succinic acid superimposed on the experimental structure (cyan). y-direction view.

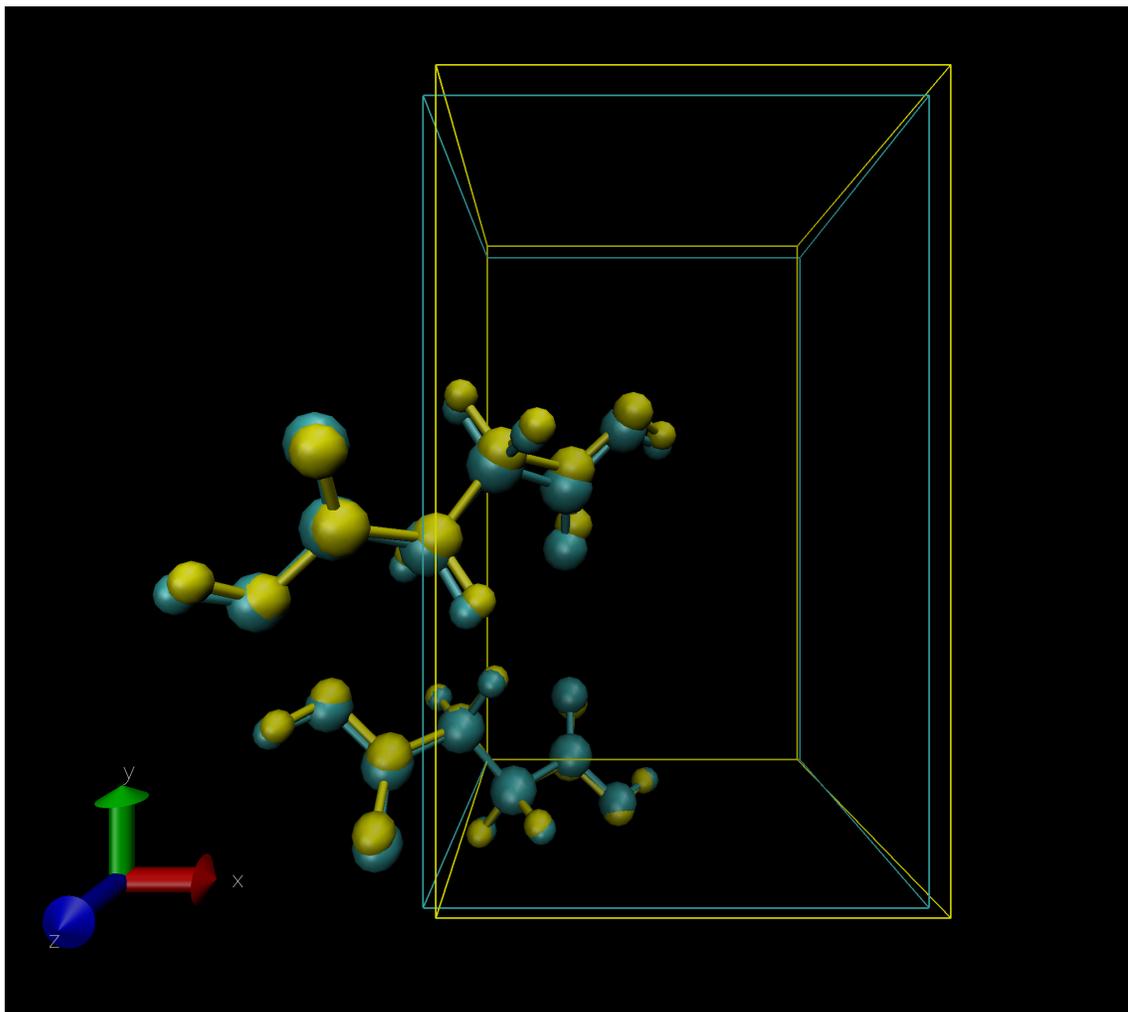


Figure E.6: GAFF predicted crystal structure (yellow) of  $\beta$ -succinic acid superimposed on the experimental structure (cyan). z-direction view.

Table E.3: GAFF predicted lattice parameters of succinic acid crystal —  $\gamma$  polymorph

Lattice parameter	Expt.	GAFF	% $\Delta$
$a(\text{\AA})$	$5.702 \pm 0.001$	$5.696 \pm 0.001$	-0.10%
$b(\text{\AA})$	$8.415 \pm 0.001$	$9.229 \pm 0.001$	9.67%
$c(\text{\AA})$	$10.354 \pm 0.001$	$10.318 \pm 0.001$	-0.35%
$\alpha$ ( $^\circ$ )	$90.00 \pm 0.00$	$89.993 \pm 0.004$	-0.01%
$\beta$ ( $^\circ$ )	$90.37 \pm 0.00$	$99.438 \pm 0.004$	10.03%
$\gamma$ ( $^\circ$ )	$90.00 \pm 0.00$	$90.001 \pm 0.004$	0.00%

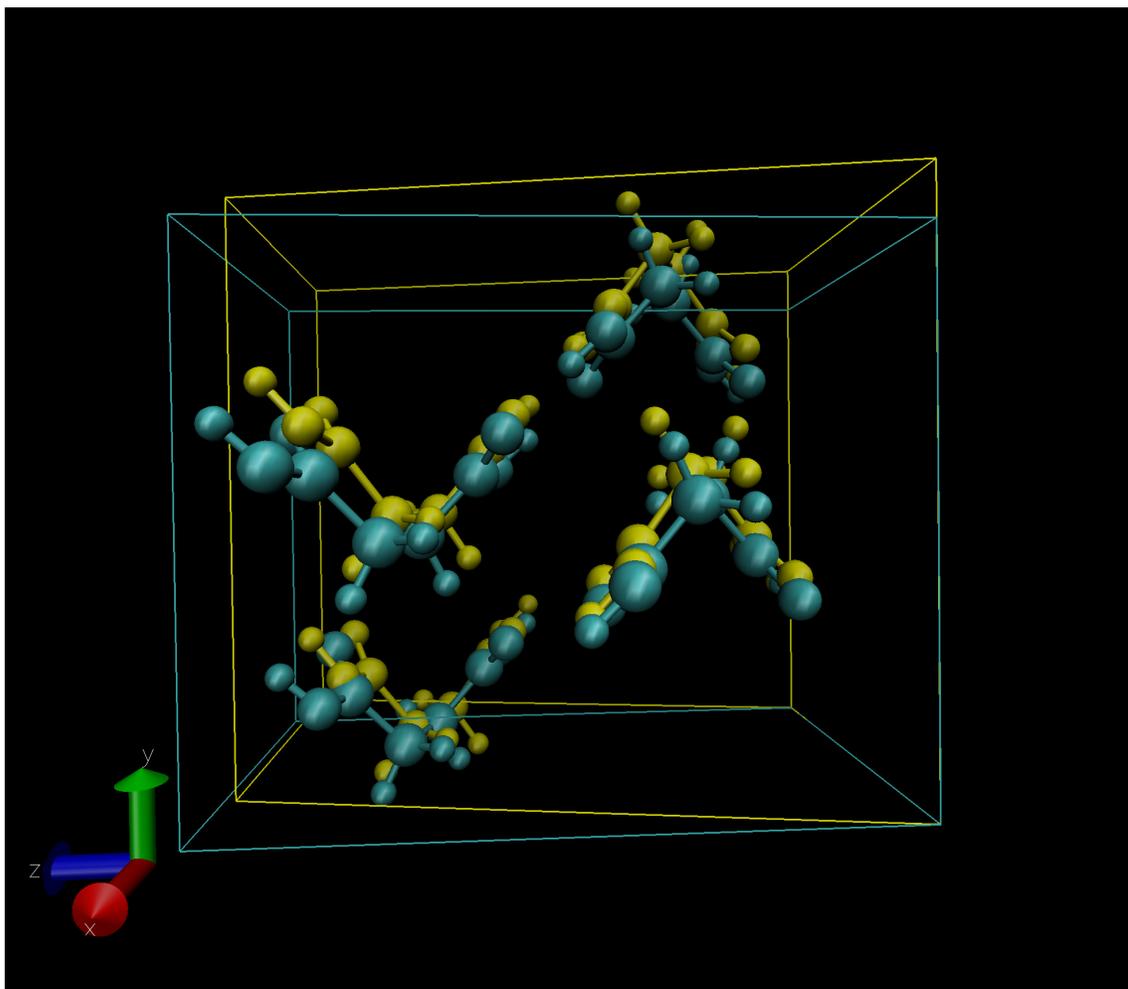


Figure E.7: GAFF predicted crystal structure (yellow) of  $\gamma$ -succinic acid superimposed on the experimental structure (cyan). x-direction view.

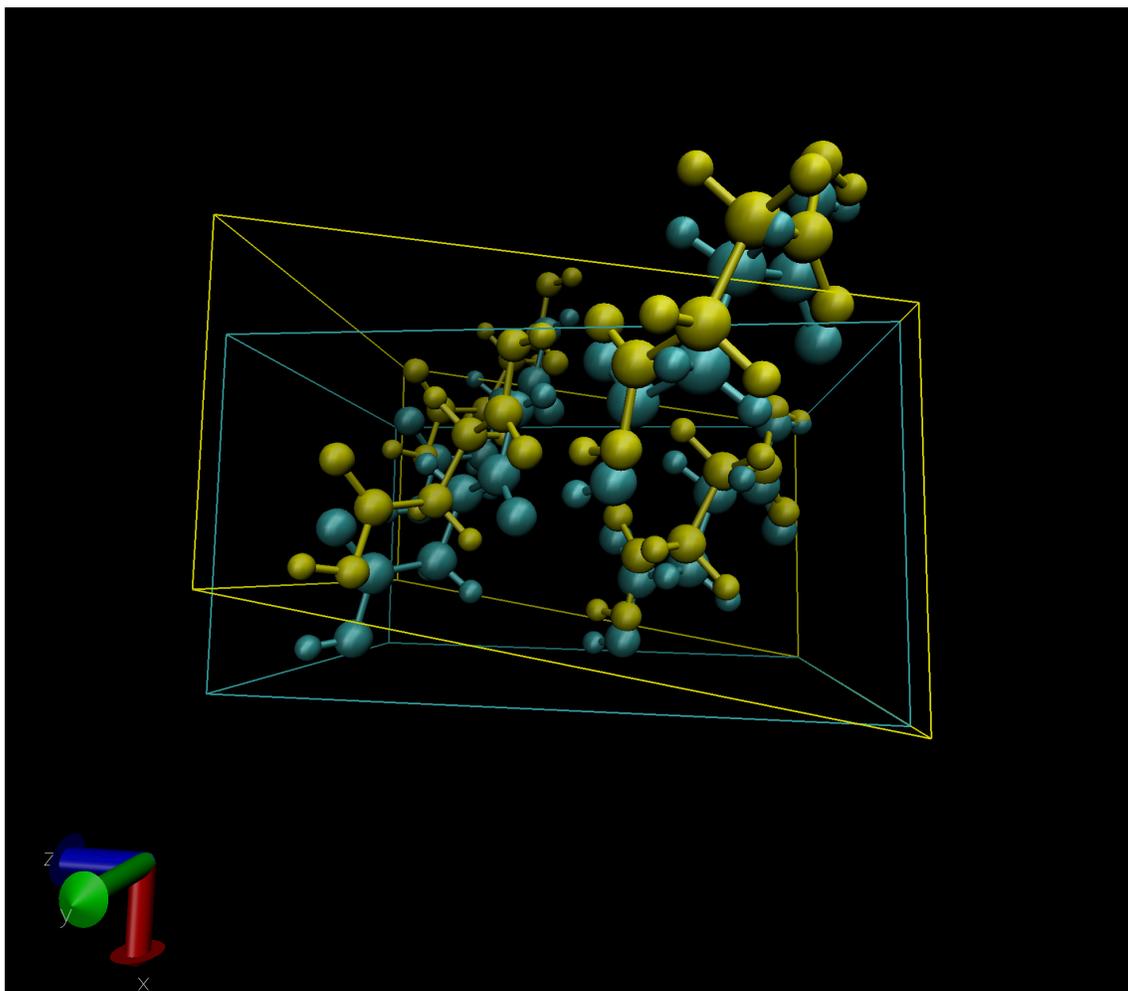


Figure E.8: GAFF predicted crystal structure (yellow) of  $\gamma$ -succinic acid superimposed on the experimental structure (cyan). *y*-direction view.

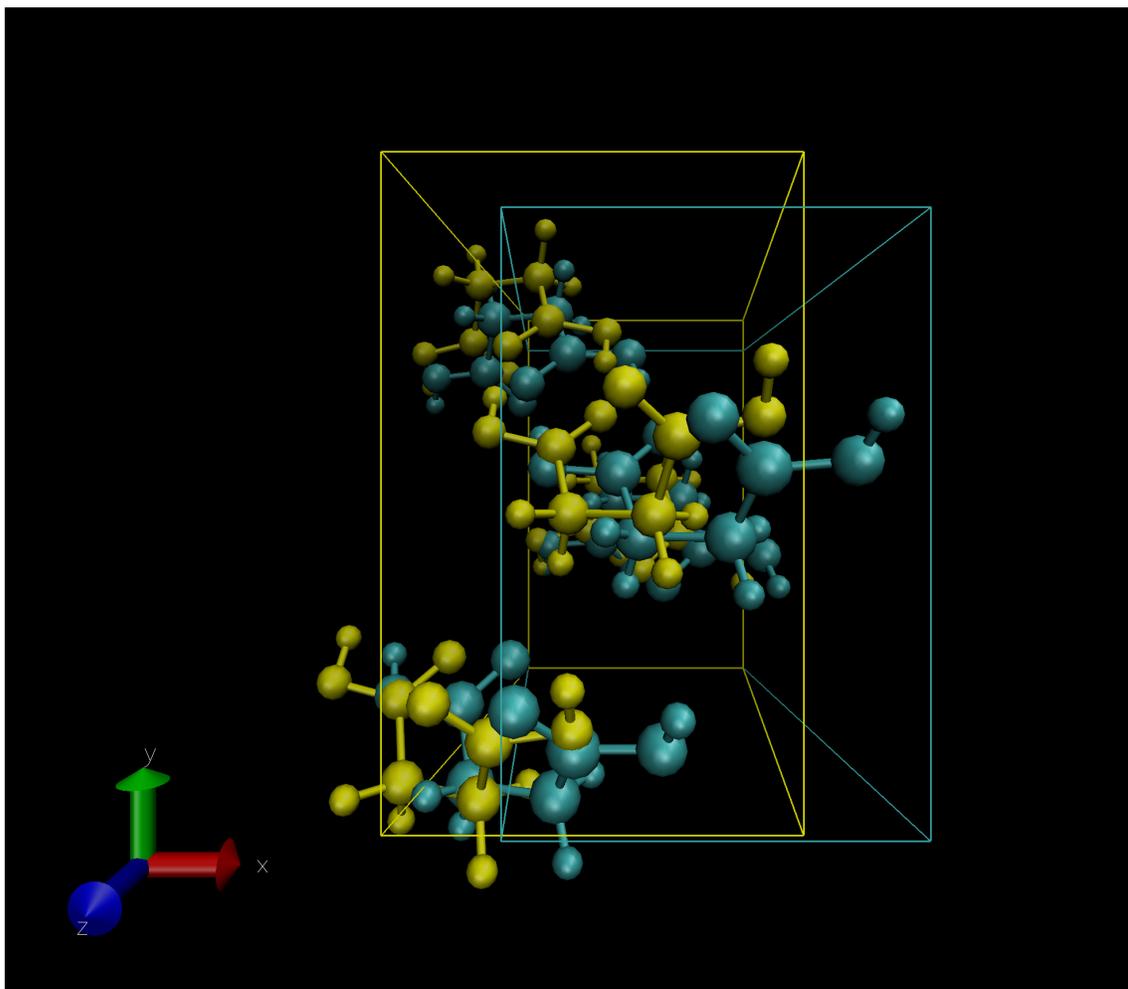


Figure E.9: GAFF predicted crystal structure (yellow) of  $\gamma$ -succinic acid superimposed on the experimental structure (cyan). z-direction view.

# Appendix F

## Sublimation Enthalpy Derivation

The sublimation enthalpy for a solid is given by:

$$\Delta h_{sub}(T_{eqm}, P_{eqm}) = h_{gas}(T_{eqm}, P_{eqm}) - h_{solid}(T_{eqm}, P_{eqm}) \quad (\text{F.1})$$

We simulate the crystal at 1 atm. since the crystallographic data is assumed to be reported at 1 atm. Also, for this direct calculation of sublimation pressure, we haven't calculated the equilibrium solid-vapor curve using the force field and are therefore unaware of the true equilibrium conditions. Hence to get  $h_{solid}$  at eqm we use:

$$h_{solid}(T_{eqm}, P_{eqm}) = h_{solid}(T_{eqm}, 1 \text{ atm}) + \int_{1 \text{ atm}}^{P_{eqm}} \left( \frac{\partial h}{\partial P} \right)_T dP \quad (\text{F.2})$$

Now we need to evaluate the integral: From the fundamental equation of thermodynamics we know that:

$$dh = Tds + vdP \quad (\text{F.3})$$

Taking the partial derivative of eqn (F.3) w.r.t P at constant T we get:

$$\left(\frac{\partial h}{\partial P}\right)_T = T \left(\frac{\partial s}{\partial P}\right)_T + v \quad (\text{F.4})$$

Using a Maxwell relation :

$$\left(\frac{\partial s}{\partial P}\right)_T = - \left(\frac{\partial v}{\partial T}\right)_P \quad (\text{F.5})$$

eqn (F.4) becomes:

$$\left(\frac{\partial h}{\partial P}\right)_T = -T \left(\frac{\partial v}{\partial T}\right)_P + v \quad (\text{F.6})$$

Since:

$$\left(\frac{\partial v}{\partial T}\right)_P = v\alpha_p \quad (\text{F.7})$$

where  $\alpha_p$  is the thermal expansion coefficient, therefore eqn (F.6) becomes:

$$\left(\frac{\partial h}{\partial P}\right)_T = v(1 - T\alpha_p) \quad (\text{F.8})$$

Therefore the integral evaluates to:

$$\int_{1atm}^{P_{eqm}} \left(\frac{\partial h}{\partial P}\right)_T dP = \int_{1atm}^{P_{eqm}} v(1 - T\alpha_p)dP \quad (\text{F.9})$$

Now consider:  $v(1 - T\alpha_p)$

We know that  $\alpha_{p,liq} > \alpha_{p,solid}$

And for water,  $\alpha_p = 0.0002K^{-1}$

hence, at 300 K,  $(1 - T\alpha_{p,solid}) > 0.96$

Therefore,  $(1 - T\alpha_{p,solid}) \approx 1$ , gives

$$\int_{1 \text{ atm}}^{P_{eqm}} \left( \frac{\partial h}{\partial P} \right)_T dP \approx \int_{1 \text{ atm}}^{P_{eqm}} v dP \quad (\text{F.10})$$

Assuming that:  $v$  doesn't vary much with  $P$ ,

$$\int_{1 \text{ atm}}^{P_{eqm}} v dP \approx v(P_{eqm} - 1) \approx v(0 - 1) \approx -8.0 \text{ J/mol} \quad (\text{F.11})$$

However, the sublimation enthalpies we are dealing with are  $\approx 10^5 \text{ J/mol}$ , hence we can safely neglect the integral. Therefore, eqn (F.2) evaluates to:

$$h_{solid}(T_{eqm}, P_{eqm}) = h_{solid}(T_{eqm}, 1 \text{ atm}) \quad (\text{F.12})$$

i.e. the enthalpy of the solid is invariant of Pressure. Also the vapor pressure of a solid ( $P_{eqm}$ ) is so small ( $\approx 1 \text{ mPa}$ ) that ideal gas law is valid.

Therefore,

$$\Delta h_{sub}(T_{eqm}, P_{eqm}) = h_{gas}(T_{eqm}, P_{eqm}) - h_{solid}(T_{eqm}, 1 \text{ atm}) \quad (\text{F.13})$$

and,

$$\Delta h_{sub}(T_{eqm}, P_{eqm}) = u_{gas}(T_{eqm}) + Pv_{gas} - u_{solid}(T_{eqm}, 1 \text{ atm}) - Pv_{solid} \quad (\text{F.14})$$

Since,  $v_{gas} \gg v_{solid}$  and  $Pv_{gas} = RT$ ,  $P(v_{gas} - v_{solid}) \approx RT$

Thus:

$$\Delta h_{sub}(T_{eqm}, P_{eqm}) = u_{gas}(T_{eqm}) - u_{solid}(T_{eqm}, 1 \text{ atm}) + RT \quad (\text{F.15})$$

# Appendix G

## Activity and Activity Coefficients

In this appendix we iron out a few key definitions crucial in the context of chemical potential calculations and applications.

### G.1 Activity

When one refers to activity, there are two *types* of activity as laid out by the International Union of Pure and Applied Chemistry (IUPAC):

1. *Absolute activity* ( $a$ ): The exponential of the ratio of the chemical potential,  $\mu$ , to  $RT$  where  $R$  is the gas constant and  $T$  the thermodynamic temperature, i.e.,

$$a = \exp\left(\frac{\mu}{RT}\right). \quad (\text{G.1})$$

<https://doi.org/10.1351/goldbook.A00019>

2. *Relative activity* ( $a^\circ$ ): The relative activity, is defined by the equation

$$a^\circ = \exp\left(\frac{\mu - \mu^\circ}{RT}\right) \quad (\text{G.2})$$

where  $R$  is the gas constant,  $T$  the thermodynamic temperature,  $\mu$  the chemical potential, and  $\mu^\circ$  the standard chemical potential—the definition of which depends on the choice of standard state.

<https://doi.org/10.1351/goldbook.A00115> Note: Here  $\mu^\circ$  is a standard state, which is essentially a reference state. A user of this concept is “free” to define their own reference state ( $\dagger$ ). For e.g., the above equation can be re-written as

$$a^\dagger = \exp\left(\frac{\mu - \mu^\dagger}{RT}\right) \quad (\text{G.3})$$

where  $\mu^\dagger$  is the reference chemical potential, and  $a^\dagger$  is the relative activity defined relative to the reference state,  $\dagger$ .

## G.2 Activity coefficients

An activity coefficient is a thermodynamic factor used to account for deviations from the reference state, which is usually a state that represents *ideality*, i.e.,

$$RT \ln(\gamma_i^\dagger x_i) = \mu_i - \mu_i^\dagger \quad (\text{G.4})$$

Therefore, using equations G.3 and G.4 we have,

$$\gamma_i^\dagger = \frac{a_i^\dagger}{x_i} \quad (\text{G.5})$$

### G.2.1 Revisiting the Ideal Solution Model

The Gibbs free energy,  $G$ , of a mixture is given by,

$$G = \sum_i G_i^* + \Delta G_{mix}, \quad (\text{G.6})$$

where,  $G_i^*$  is the free energy of  $i$  in the pure state, and for an ideal solution,

$$\Delta G_{mix} = Nk_B T \sum_i x_i \ln x_i. \quad (\text{G.7})$$

Therefore, we have

$$G = \sum_i G_i^* + Nk_B T \sum_i x_i \ln x_i. \quad (\text{G.8})$$

Now, we know that

$$\mu_i = \left( \frac{\partial G}{\partial N_i} \right)_{T,P,N_{j \neq i}} \quad (\text{G.9})$$

Thus, taking the partial derivative of  $G$  wrt  $N_i$ , we get

$$\mu_i^{id} = \mu_i^* + k_B T \ln x_i \quad (\text{G.10})$$

where,  $\mu_i^*$  is the chemical potential of pure species  $i$  in the same state of aggregation as that of the solution, i.e., in a liquid mixture,  $\mu_i^*$  is the chemical potential of pure liquid  $i$  at temperature  $T$  and pressure  $P$ . Thus, the limiting behavior of Equation G.10 is,

- $\lim_{x_i \rightarrow 1} \mu_i = \mu_i^*$ , i.e., the chemical potential of solute  $i$  tends to the pure  $i$ 's chemical potential as  $x_i$  tends to 1.
- $\lim_{x_i \rightarrow 0} \mu_i = -\infty$ , i.e., the chemical potential of solute  $i$  tends to negative infinity as  $x_i$  tends to 0. This is because, the number of configurations of solute  $i$  in the solution tend to infinity as  $x_i$  tends to 0.

## G.2.2 Non-ideal solution

For non-ideal solutions, the chemical potential of species  $i$  can be broken as:

$$\mu_i = \mu_i^{id} + \mu_i^{ex} \quad (\text{G.11})$$

where,  $\mu_i^{ex}$  is the real solution's deviation from the ideal one, and, we have

$$\mu_i = \mu_i^* + k_B T \ln x_i + \mu_i^{ex} \quad (\text{G.12})$$

Now, by defining

$$\mu_i^{ex} = k_B T \ln \gamma_i, \quad (\text{G.13})$$

we get:

$$\mu_i = \mu_i^* + k_B T \ln x_i + k_B T \ln \gamma_i, \quad (\text{G.14})$$

or,

$$\mu_i = \mu_i^* + k_B T \ln \gamma_i x_i. \quad (\text{G.15})$$

where,  $\gamma_i(T, P, \mathbf{x})$  is the activity coefficient of species  $i$ . Thus, the activity coefficient gives us a measure of deviation from an ideal solution. Note, this is because in Equation G.11, we broke down the species' chemical potential into an ideal part and an excess part, i.e., the ideal solution is the reference state/system against which the non-ideality is being measured.

Thus, the limiting behavior of Equation G.15 is,

- $\lim_{x_i \rightarrow 1} \gamma_i = 1$ , since the chemical potential of solute  $i$  ( $\mu_i$ ) tends to the pure  $i$ 's chemical potential ( $\mu_i^*$ ) as  $x_i$  tends to 1,  $\gamma_i$  tends to 1.
- $\lim_{x_i \rightarrow 0} \gamma_i = \gamma_i^\infty$ , as we solvate a species in a solution, the excess chemical potential

of solute  $i$  tends to a constant value as  $x_i$  tends to 0. Imagine solvating a molecule in a lake vs. in an ocean, the *excess* potential—over the ideal—would be the same! This is because, the molecule is only seeing and interacting with solvent molecules (think spherical cutoffs for the interactions) in both the cases, so the free energy change is the same.

Note,  $\mu_i^*$  is the chemical potential of pure species  $i$  in the same state of aggregation as that of the solution. So, if we have a solution with ions,  $\mu_i^*$  is the chemical potential of the pure ion in the liquid phase at temperature  $T$  and pressure  $P$ , which is a hypothetical state—not measurable experimentally! Therefore, for such systems we adopt a different *convention*. We define the activity coefficient at infinite dilution of the solute to be 1. We do this by normalizing the activity coefficient in Equation G.15 by the infinite dilution activity coefficient, i.e.,

$$\gamma_i^\dagger = \frac{\gamma_i}{\gamma_i^\infty} \quad (\text{G.16})$$

which gives us,

$$\mu_i = \mu_i^* + k_B T \ln \left( \gamma_i^\infty \frac{\gamma_i}{\gamma_i^\infty} x_i \right) \quad (\text{G.17a})$$

$$\mu_i = \mu_i^\dagger + k_B T \ln \gamma_i^\dagger x_i \quad (\text{G.17b})$$

where,

$$\mu_i^\dagger = \mu_i^* + k_B T \ln \gamma_i^\infty \quad (\text{G.18})$$

Thus, the limiting behavior of Equation G.17 is,

- $\lim_{x_i \rightarrow 0} \gamma_i^\dagger = 1$ , by construction. Here the non-ideality is being measured from a state of an infinitely diluted solution, i.e., a reference state.
- $\lim_{x_i \rightarrow 1} \gamma_i^\dagger = 1/\gamma_i^\infty$ .

# Appendix H

## Thermodynamic Checks

### H.1 $\mu^C(T, P^{sat})$ :

$$\mu^C(T, P^{sat}) = a^C(T, v^C) + P^{sat}v^C \quad (\text{H.1})$$

since,  $v^C$  doesn't vary much with P we can assume it is unchanged. Also, since  $P^{sat} \sim \text{mPa}$ ,  $P^{sat}v^C \sim 10^{-11}k_B T/\text{molecule}$  and can be ignored. Thus,

$$\mu^C(T, P^{sat}) = a^C(T, v^C) \quad (\text{H.2})$$

## H.2 Testing the centroid to bond transformations

Table H.1: Centroid to bond transformations for naphthalene

$k_{sp,C}(k_B T/\text{\AA}^2)$	# of Gauss points	$\beta a^{*,intra}$	$\beta a^{*\rightarrow bonded}$	$\beta a^{bonded,\dagger}$
2000	45	80.60	-108.53	$-27.93 \pm 0.17$
	70	80.60	-108.97	$-28.36 \pm 0.14$
200	45	21.87	-50.17	$-28.30 \pm 0.09$
	70	21.87	-50.24	$-28.36 \pm 0.07$
	95	21.87	-50.18	$-28.31 \pm 0.06$
20	45	-36.84	8.51	$-28.33 \pm 0.05$

Table H.2: Centroid to bond transformations for succinic acid

$k_{sp,O}(k_B T/\text{\AA}^2)$	# of Gauss points	$\beta a^{*,intra}$	$\beta a^{*\rightarrow bonded}$	$\beta a^{bonded,\dagger}$
2000	45	55.62	-85.65	$-30.03 \pm 0.10$
200	75	10.72	-41.57	$-30.85 \pm 0.05$
200	95	10.72	-41.56	$-30.84 \pm 0.05$
20	75	-34.18	3.47	$-30.72 \pm 0.03$
20	95	-34.18	3.49	$-30.70 \pm 0.03$

## H.3 Sampling the COOH dihedral in the gas phase

### REMD fingerprint

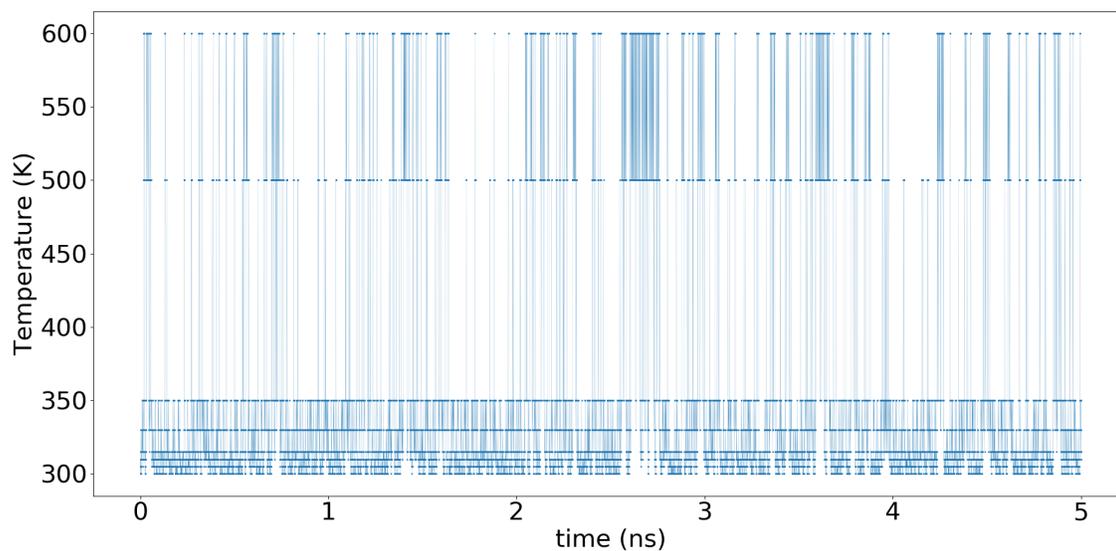


Figure H.1: Temperature walk of first replica ( $\lambda = 0.5$ ), only first 5 ns shown for clarity. The lower temperatures (300 K- 350 K) are chosen where we evaluate the vapor pressures, the higher temperatures (500 K, 600 K) are for enhancing the sampling of the COOH dihedral

### COOH dihedral sampling improvement while switching ON coulombic interactions

The integration time step in both the MD and REMD simulations is 0.5 fs, the dihedral data is shown at an interval of 10 ps for clarity.

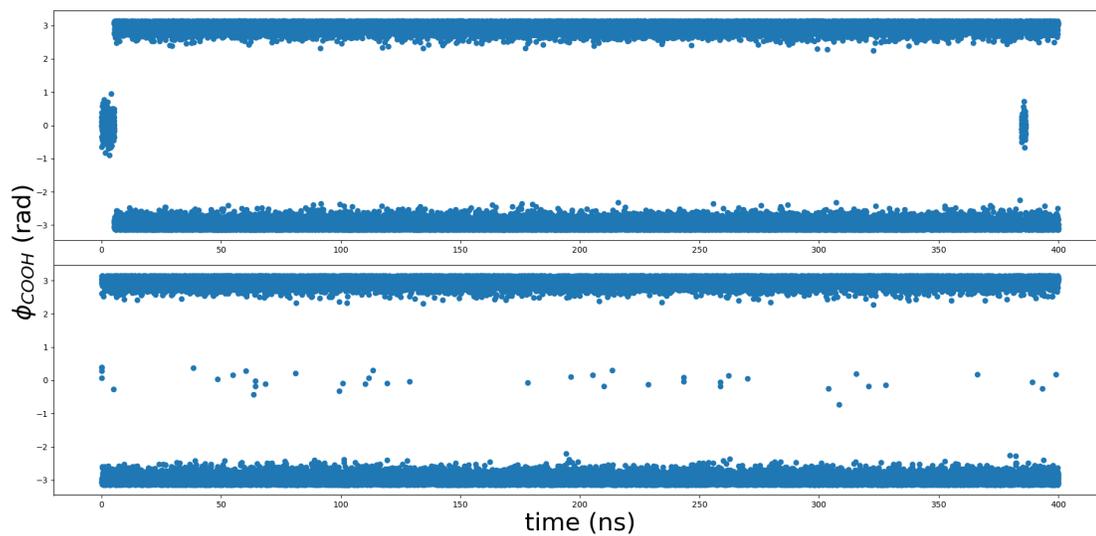


Figure H.2: MD (upper) vs REMD (lower) COOH dihedral timeseries ( $\lambda = 0.047$ )

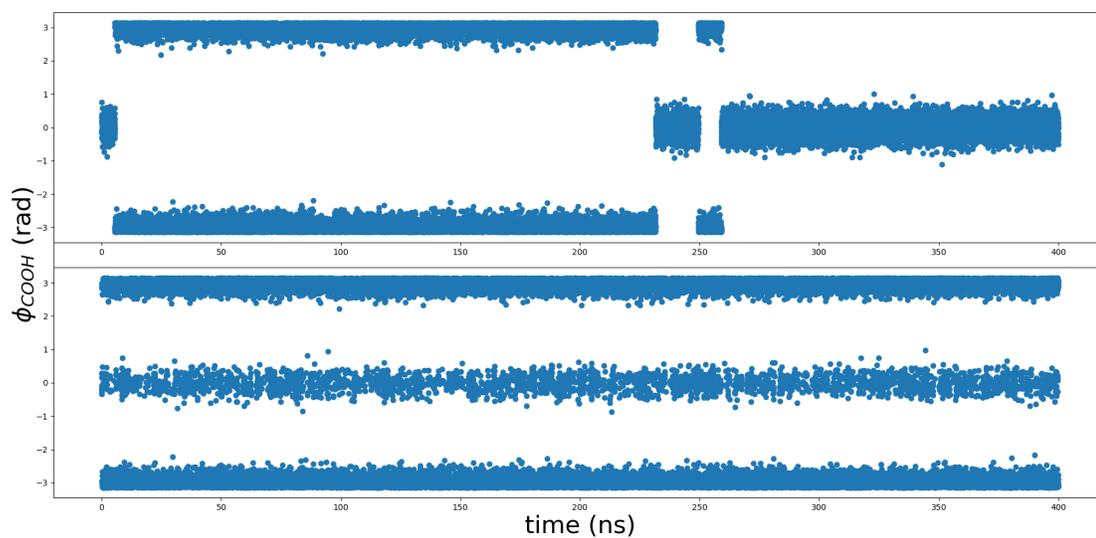


Figure H.3: MD (upper) vs REMD (lower) COOH dihedral timeseries ( $\lambda = 0.5$ )

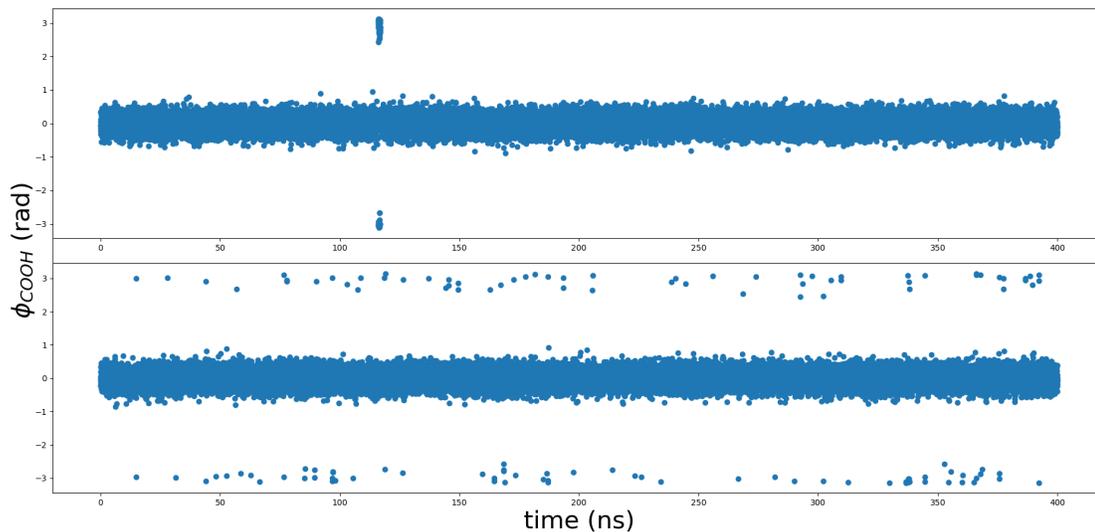


Figure H.4: MD (upper) vs REMD (lower) COOH dihedral timeseries ( $\lambda = 0.953$ )

## H.4 Long range vdW correction for crystalline systems

For long range vdW interactions the standard tail correction for  $n$  different nonbond atom types interacting with pair potential  $E_{ij}(r)$  is given by,

$$E_{tail-corrected} = \frac{1}{2} \sum_{i=1}^n N_i \sum_{j=1}^n \rho_j 4\pi \int_{r_c}^{\infty} g_{ij}(r) E_{ij}(r) r^2 dr, \quad (\text{H.3})$$

where  $g_{ij}(r)$  is the pair radial distribution function and  $r_c$  is the cutoff value.<sup>[1]</sup> For liquids,  $g(r) = 1$  is a good approximation for large values of  $r_c$ . For molecular crystals, though the radial distribution function  $g(r)$  does not converge to unity, the definition of  $g(r)$  implies that at large separation, the mean value of  $g(r)$  must be 1. Since the vdW function and its first derivative are very flat in the region of long separation, it is plausible to assume  $g(r) = 1$  and use equation H.3 to calculate the tail corrections.<sup>[1]</sup> To

test this plausibility, we compute the potential energy per unit molecule—for the initial configuration—at various cutoffs with and without tail corrections for the naphthalene and succinic acid crystals as shown in Figure H.5.

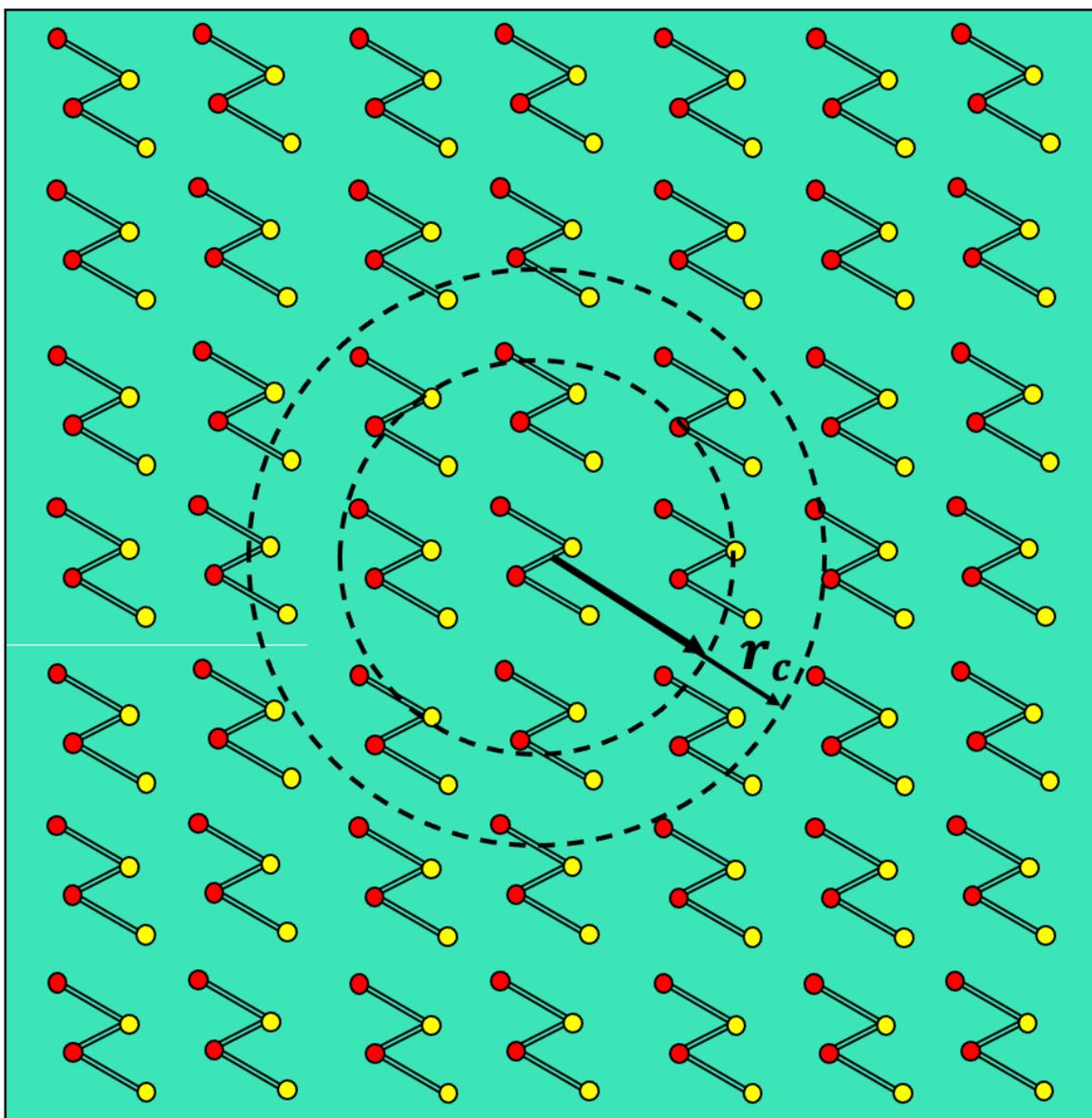


Figure H.5: Test system schematic for testing long range vdW corrections.

We setup a  $10 \times 14 \times 14$  and  $15 \times 10 \times 15$  system for naphthalene and  $\beta$ -succinic

acid, respectively to compute the potential energy per molecule with and without tail corrections for  $r_c = 10\text{-}35 \text{ \AA}$ . As seen from figures H.6 and H.7, the tail corrections computed using equation H.3 are an excellent approximation.

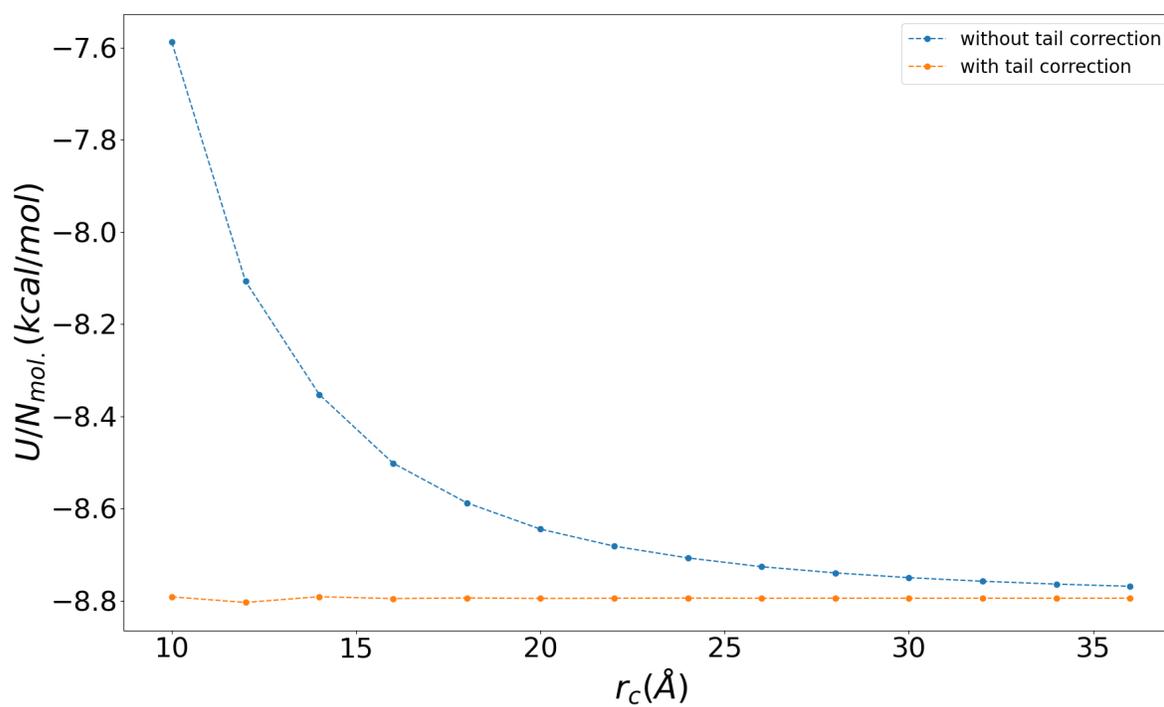


Figure H.6: Testing tail corrections for the naphthalene crystal system.

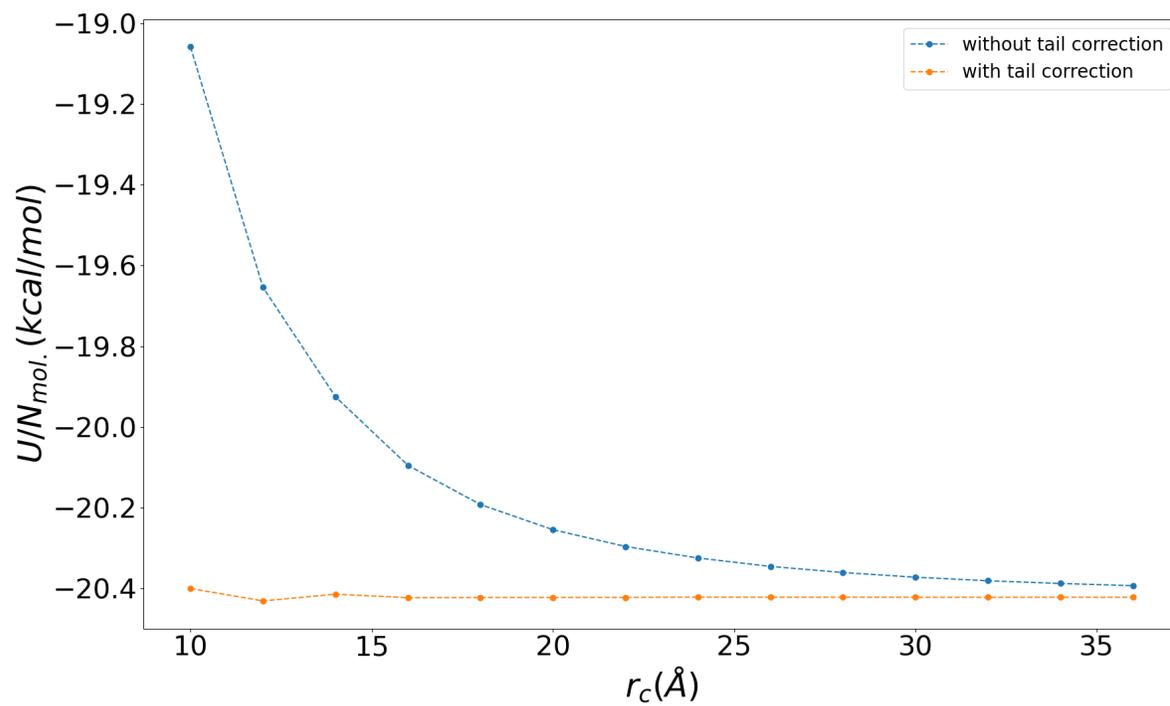


Figure H.7: Testing tail corrections for the  $\beta$ -succinic acid crystal system.

# Bibliography

- [1] H. Sun, *Compass: an ab initio force-field optimized for condensed-phase applications overview with details on alkane and benzene compounds*, *The Journal of Physical Chemistry B* **102** (1998), no. 38 7338–7364.

# Appendix I

## Computing the free energy of the centroid

Computing  $Z_n/Z_n^o$ :

$$\begin{aligned}\frac{Z_n}{Z_n^o} &= \frac{\int \prod_{i=1}^n d\mathbf{r}_i e^{-\frac{\beta}{2} \sum_{i=1}^n k_i \|\mathbf{r}_i - \mathbf{r}_{COM}^{(n)}\|^2}}{\int \prod_{i=1}^n d\mathbf{r}_i e^{-\frac{\beta}{2} \sum_{i=1}^n k_i \|\mathbf{r}_i - \mathbf{r}_{COM}^{(n-1)}\|^2}} \\ &= \frac{\int \prod_{i=1}^n d\mathbf{r}_i e^{-\frac{\beta}{2} U_n(\mathbf{r}; \mathbf{r}_{COM}^{(n)})}}{\int \prod_{i=1}^n d\mathbf{r}_i e^{-\frac{\beta}{2} U_n^o(\mathbf{r}; \mathbf{r}_{COM}^{(n-1)})}} \\ &= \langle e^{-\beta(U_n - U_n^o)} \rangle_{U_n^o} \\ &= \langle e^{-\beta \Delta U_n} \rangle_{U_n^o}\end{aligned}\tag{I.1}$$

## Naphthalene

The results for adding atoms starting with a carbon (C) atom are:

Table I.1: Simulation results for successively adding atoms to build the centroid for naphthalene

Atom being added	n	$\langle e^{-\beta\Delta U_n} \rangle_{U_n^o}$
C	2	$2.8417 \pm 0.0273$
	3	$1.8374 \pm 0.0041$
	4	$1.5378 \pm 0.0022$
	5	$1.3976 \pm 0.0014$
	6	$1.3144 \pm 0.0011$
	7	$1.2608 \pm 0.0008$
	8	$1.2209 \pm 0.0007$
	9	$1.1931 \pm 0.0006$
	10	$1.1710 \pm 0.0005$
H	11	$1.0126 \pm 0.0000$
	12	$1.0125 \pm 0.0000$
	13	$1.0124 \pm 0.0000$
	14	$1.0123 \pm 0.0000$
	15	$1.0123 \pm 0.0000$
	16	$1.0121 \pm 0.0000$
	17	$1.0121 \pm 0.0000$
	18	$1.0120 \pm 0.0000$

**Succinic acid**

The results for adding atoms starting with an oxygen (O) atom are:

Table I.2: Simulation results for successively adding atoms to build the centroid for succinic acid

Atom being added	n	$\langle e^{-\beta\Delta U_n} \rangle_{U_n^o}$
O	2	$2.8403 \pm 0.0097$
	3	$1.8381 \pm 0.0013$
	4	$1.5395 \pm 0.0007$
C	5	$1.2947 \pm 0.0003$
	6	$1.2464 \pm 0.0002$
	7	$1.2112 \pm 0.0002$
	8	$1.1853 \pm 0.0002$
H	9	$1.0135 \pm 0.0000$
	10	$1.0134 \pm 0.0000$
	11	$1.0133 \pm 0.0000$
	12	$1.0132 \pm 0.0000$
	13	$1.0131 \pm 0.0000$
	14	$1.0129 \pm 0.0000$

Computing  $Z_n^o/Z_{n-1}$ :

$$\begin{aligned}
\frac{Z_n^o}{Z_{n-1}} &= \frac{\int \prod_{i=1}^n d\mathbf{r}_i e^{-\frac{\beta}{2} \sum_{i=1}^n k_i \|\mathbf{r}_i - \mathbf{r}_{COM}^{(n-1)}\|^2}}{\int \prod_{i=1}^{n-1} d\mathbf{r}_i e^{-\frac{\beta}{2} \sum_{i=1}^{n-1} k_i \|\mathbf{r}_i - \mathbf{r}_{COM}^{(n-1)}\|^2}} \\
&= \frac{\int d\mathbf{r}_n \left[ \int \prod_{i=1}^{n-1} d\mathbf{r}_i e^{-\frac{\beta}{2} \sum_{i=1}^{n-1} k_i \|\mathbf{r}_i - \mathbf{r}_{COM}^{(n-1)}\|^2} \right] e^{-\frac{\beta}{2} k_n \|\mathbf{r}_n - \mathbf{r}_{COM}^{(n-1)}\|^2}}{\int \prod_{i=1}^{n-1} d\mathbf{r}_i e^{-\frac{\beta}{2} \sum_{i=1}^{n-1} k_i \|\mathbf{r}_i - \mathbf{r}_{COM}^{(n-1)}\|^2}} \quad (\text{I.2}) \\
&= \int_{-\infty}^{+\infty} d\mathbf{r}_n e^{-\frac{\beta}{2} k_n \|\mathbf{r}_n - \mathbf{r}_{COM}^{(n-1)}\|^2} \\
&= \left( \frac{2\pi}{\beta k_n} \right)^{3/2}
\end{aligned}$$

# Appendix J

## Potential Mean Force - COOH dihedral

In this appendix we examine the relative stability of the cis and trans conformations of the COOH dihedral in the solution phase along the coupling path shown in the figure below.

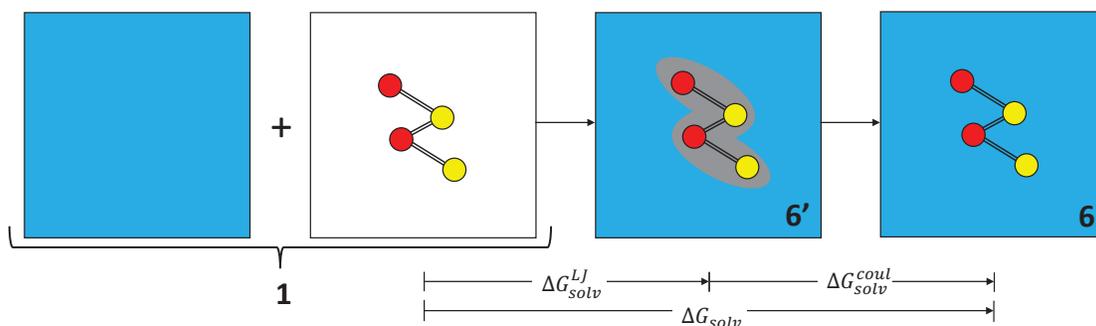


Figure J.1: The decoupling thermodynamic path. Grey shaded area:  $LJ_{A-B}=ON$ ,  $COUL_{A-B}=OFF$ ; colored atoms: intramolecular LJ and coulombic interactions=ON

## Gas Phase

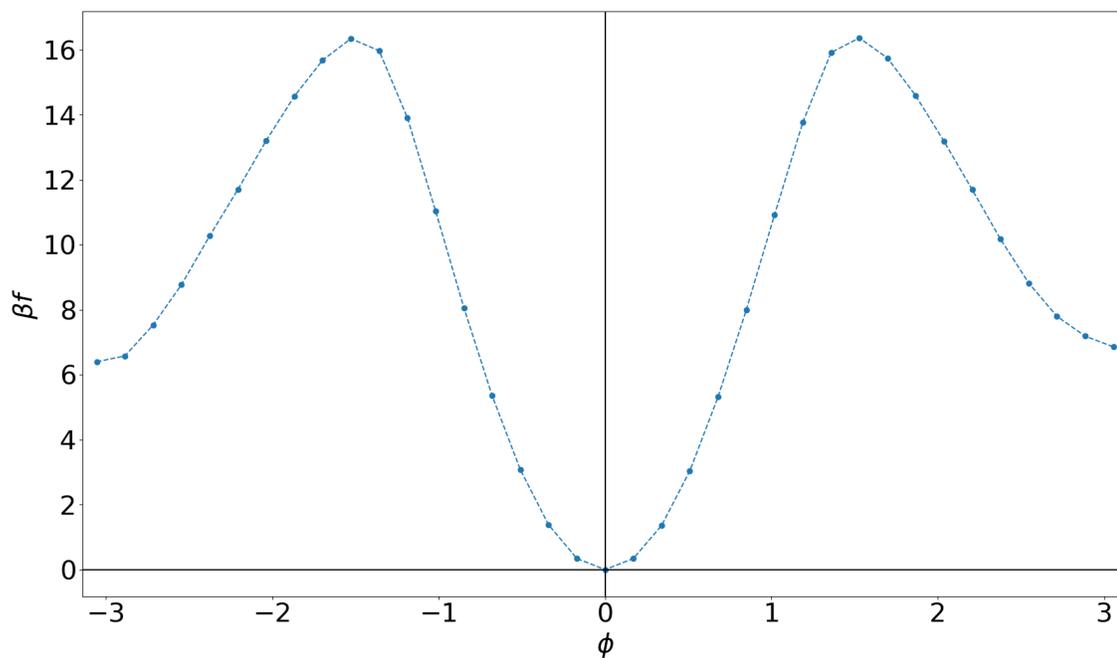


Figure J.2: The potential mean force along the COOH dihedral in the gas phase—the decoupled state—where  $\lambda_{LJ} = 0$  and  $\lambda_q = 0$ .

Figure J.2 shows the potential mean force of the COOH dihedral in the decoupled state, i.e., the gas state. Since the free energy difference between the cis ( $\phi = 0^\circ$ ) and trans ( $\phi = 180^\circ$ ) is  $\sim 6 k_B T$ —it is clear that the dihedral exists in the cis conformation.

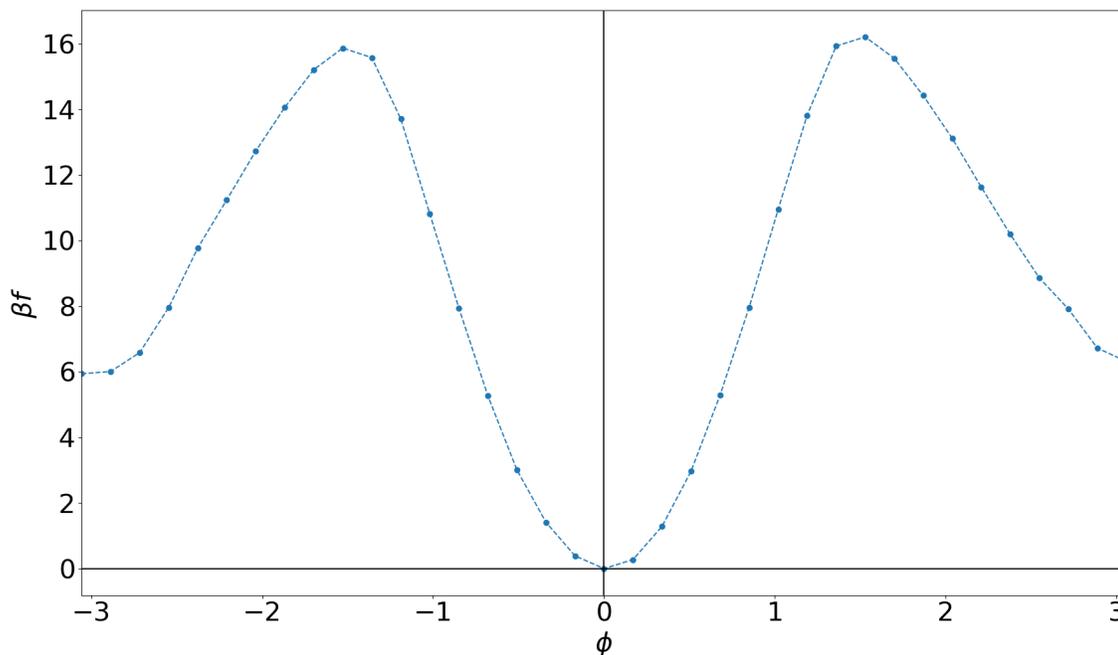
**LJ-coupled State**

Figure J.3: The potential mean force along the COOH dihedral in the LJ-coupled state—where  $\lambda_{LJ} = 1$ ,  $\lambda_q = 0$ .

Figure J.3 shows the potential mean force of the COOH dihedral in the LJ-coupled state, i.e., the solute molecule is only coupled to the solution via the Lennard-Jones interactions. Since the free energy difference between the cis ( $\phi = 0^\circ$ ) and trans ( $\phi = 180^\circ$ ) is once again  $\sim 6 k_B T$ —it is clear that the dihedral exists in the cis conformation.

## Coupled State

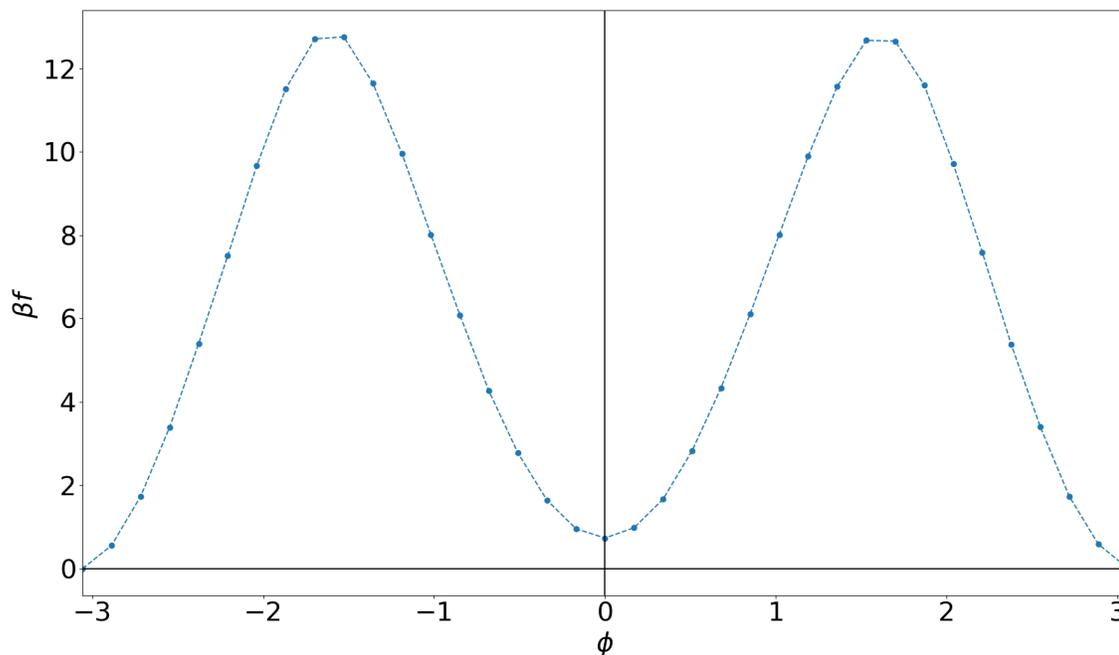


Figure J.4: The potential mean force along the COOH dihedral in the fully-coupled state—where  $\lambda_{LJ} = 1, \lambda_q = 1$ .

Figure J.4 shows the potential mean force of the COOH dihedral in the coupled state, i.e., the solute molecule is fully coupled to the solution via the Lennard-Jones and coulombic interactions. In this case, since the free energy difference between the cis ( $\phi = 0^\circ$ ) and trans ( $\phi = 180^\circ$ )  $\leq 1 k_B T$  — the dihedral exists in the cis and trans conformations, both.

# Appendix K

## Population Balance Deep Dive

### K.1 Regularity Condition

Here in this appendix we derive the regularity boundary condition for the PDE:

$$\frac{\partial n}{\partial t} = \frac{n_{in}}{\tau} - \frac{n}{\tau} + Q_B - Q_D - \frac{\partial(G \cdot n)}{\partial r} + \frac{\partial\delta(r) \cdot G \cdot n}{\partial r} \quad (\text{K.1})$$

with the initial condition and boundary conditions as  $n(r, 0) = n_0(r)$  &  $n(\infty, t) = n(0, t) = 0$ , respectively

where:

1.  $c$  = concentration of the solute in the crystallizer
2.  $n = n(r, t)$  = # of crystals per unit volume of SUSPENSION with radius  $r$  and time  $t$
3.  $Q_B = Q_B(r, c)$  = # of crystals born at radius  $r$  per unit volume of SUSPENSION per unit volume of time, i.e., the birth rate

$Q_B = \epsilon B \delta(r)$ , where:

$B$  = nucleation rate per unit volume of LIQUID

$\epsilon$  = fraction of liquid in the crystallizer, i.e.,  $\epsilon = V_{liq}/V_{sus}$

4.  $Q_D = Q_D(r, c) = \#$  of crystals which die with radius  $r$  per unit volume of SUSPENSION per unit volume of time, i.e., the death rate
5.  $G = G(r, c) =$  the rate at which a finite size particle increases its radius, i.e. the growth rate

Note,  $n(0, t) = 0$ , as nuclei/particles of size 0 do not exist. The smallest size nuclei are the critical size nuclei, i.e.  $r = r_{crit}$ .

Assuming  $n_{in} = 0$  &  $Q_D = 0$ , we integrate PDE K.1 around  $r = 0$  and take the limit as  $r \rightarrow 0$ ,

$$\lim_{r \rightarrow 0} \left[ \int_{0-r}^{0+r} \frac{\partial n}{\partial t} dr = \int_{0-r}^{0+r} -\frac{n}{\tau} dr + \int_{0-r}^{0+r} \epsilon B \delta(r) dr - \int_{0-r}^{0+r} \frac{\partial(G \cdot n)}{\partial r} dr + \int_{0-r}^{0+r} \frac{\partial(\delta(r) \cdot G \cdot n)}{\partial t} dr \right] \quad (\text{K.2})$$

Going left to right term wise we get:

$$\lim_{r \rightarrow 0} \left[ \int_{0-r}^{0+r} \frac{\partial n}{\partial t} dr \right] = \frac{\partial}{\partial t} \lim_{r \rightarrow 0} \left[ \int_{0-r}^{0+r} n dr \right] = 0 \quad (\text{K.3a})$$

$$\lim_{r \rightarrow 0} \left[ \int_{0-r}^{0+r} -\frac{n}{\tau} dr \right] = 0 \quad (\text{K.3b})$$

$$\lim_{r \rightarrow 0} \left[ \int_{0-r}^{0+r} \epsilon B \delta(r) dr \right] = \epsilon B \quad (\text{K.3c})$$

$$\lim_{r \rightarrow 0} \left[ \int_{0-r}^{0+r} \frac{\partial(G \cdot n)}{\partial r} dr \right] = G [n(0^+) - n(0^-)] \quad (\text{K.3d})$$

$$\lim_{r \rightarrow 0} \left[ \int_{0-r}^{0+r} \frac{\partial(\delta(r) \cdot G \cdot n)}{\partial r} dr \right] = G \cdot n(0, t) = 0 \quad (\text{K.3e})$$

From the above set of equations we get the “jump condition”:

$$n(0^+) - n(0^-) = \frac{\epsilon B}{G} \quad (\text{K.4})$$

which translates into the regularity boundary condition as  $n(0^-) = 0$ :

$$n(0^+, t) = \frac{\epsilon B}{G} \quad (\text{K.5})$$

with the corresponding PDE being

$$\frac{\partial n}{\partial t} = -\frac{n}{\tau} - \frac{\partial(G \cdot n)}{\partial r} \quad (\text{K.6})$$

The regularity expression can be arrived from a “continuation” at boundary point of view giving the “incorrect” expression in literature as:

$$G \cdot n(0, t) = B \quad (\text{K.7})$$

‘Dimensionally’ this is incorrect as

$$\frac{\#}{V_{sus} \cdot t} \neq \frac{\#}{V_{liq} \cdot t} \quad (\text{K.8})$$

However, the correct regularity condition (equation K.5) gives the ‘dimensions’

$$\frac{\#}{V_{sus} \cdot t} = \frac{\#}{V_{sus} \cdot t} \quad (\text{K.9})$$

Another footprint for the correct regularity condition can be seen in the zeroth moment equation (integrating equation K.6 wrt  $r$  from 0 to  $\infty$ ),

$$\int_0^\infty \frac{\partial n}{\partial t} dr = \int_0^\infty -\frac{n}{\tau} dr - \int_0^\infty \frac{\partial(G \cdot n)}{\partial r} dr \quad (\text{K.10})$$

which gives

$$m_0 = -\frac{m_0}{\tau} + G \cdot n(0) = -\frac{m_0}{\tau} + \epsilon B \quad (\text{K.11})$$

which is the same zeroth moment equation arrived from the ‘delta function PDE’.

Figure K.1 shows the schematic of the boundary condition. The jump in the boundary condition at 0 is the manifestation of the delta nucleation function. Note,  $n(0^+) = \epsilon B/G$ , is the regularity boundary condition. Since the smallest size nuclei are of size  $r_{crit}$ , therefore  $n(0^+) = n(r_{crit})$ . For numerical simulations, we approximate  $n(0^+) \approx n(0)$ .

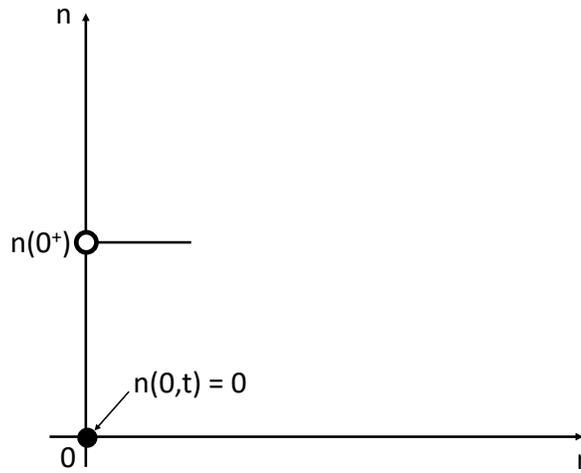


Figure K.1: Boundary Condition Schematic

# Appendix L

## Solving the Ever-kinked Crystal System Equations

### L.1 Solving the 2D Ever-kinked orthorhombic lattice equations

```

In[1]:= (*Initializing crystallography*)
 $\gamma = \pi / 2;$  (*90.0 deg*)
a = a;
b = b;

(*Setting Supercell size*)
m = m;
n = n;

(*Initializing original lattice vectors*)
i = {1, 0};
j = {Cos[ $\gamma$ ], Sin[ $\gamma$ ]};

(*Initializing new lattice vectors*)
iNew = {iNew1, iNew2};
jNew = {jNew1, jNew2};

(*Setting up equations for a Ever-kinked crystal surface*)
eq4 = m a iNew + b jNew == Lx i;
eq5 = n b jNew - a iNew == Ly j;

"Solving m a iNew + b jNew == Lx i && n b jNew - a iNew == Ly j"
ans = Solve[{eq4, eq5}, {iNew1, iNew2, jNew1, jNew2}]
iNew = {ans[[1, 1, 2]], ans[[1, 2, 2]]};
jNew = {ans[[1, 3, 2]], ans[[1, 4, 2]]};

"iNew.iNew:"
iNew.iNew // Simplify
"jNew.jNew:"
jNew.jNew // Simplify

"Solving iNew.jNew==cos[ $\gamma$ ]:"
ans2 = Solve[{iNew.jNew == Cos[ $\gamma$ ]}, {Ly}]
"Ly=f(Lx):"
Ly2 = ans2[[2, 1, 2]]

"New iNew after plugging in Ly = f(Lx)"
iNew = iNew /. Ly  $\rightarrow$  Ly2

"New jNew after plugging in Ly = f(Lx)"
jNew = jNew /. Ly  $\rightarrow$  Ly2

"Solving iNew.iNew==1:"

```

```
ans3 = Solve[{iNew.iNew == 1}, {Lx}] // Simplify
```

```
"Lx: (new periodic box length in i-direction)"
```

```
Lx2 = ans3[[2, 1, 2]]
```

```
"Ly: (new periodic box length in j-direction)"
```

```
Ly2 = Ly2 /. Lx → Lx2
```

```
iNew = iNew /. Lx → Lx2;
```

```
jNew = jNew /. Lx → Lx2 /. Ly → Ly2;
```

```
"New iNew after plugging Ly as a f(Lx):"
```

```
iNew
```

```
"New jNew after plugging Ly as a f(Lx):"
```

```
jNew
```

```
"New iNew.iNew: should be equal to 1"
```

```
iNew.iNew // Simplify
```

```
"New jNew.jNew: should be equal to 1"
```

```
jNew.jNew // Simplify
```

```
"New iNew.jNew: should be equal to 0"
```

```
iNew.jNew // Simplify
```

```
Out[11]= Solving m a iNew + b jNew == Lx i && n b jNew - a iNew == Ly j
```

```
Out[12]=  $\left\{ \left\{ iNew1 \rightarrow \frac{Lx n}{a (1 + m n)}, iNew2 \rightarrow -\frac{Ly}{a (1 + m n)}, jNew1 \rightarrow \frac{Lx}{b (1 + m n)}, jNew2 \rightarrow \frac{Ly m}{b (1 + m n)} \right\} \right\}$ 
```

```
Out[15]= iNew.iNew:
```

```
Out[16]= 
$$\frac{Ly^2 + Lx^2 n^2}{(a + a m n)^2}$$

```

```
Out[17]= jNew.jNew:
```

```
Out[18]= 
$$\frac{Lx^2 + Ly^2 m^2}{(b + b m n)^2}$$

```

```
Out[19]= Solving iNew.jNew==cos[γ]:
```

```
Out[20]=  $\left\{ \left\{ Ly \rightarrow -\frac{Lx \sqrt{n}}{\sqrt{m}} \right\}, \left\{ Ly \rightarrow \frac{Lx \sqrt{n}}{\sqrt{m}} \right\} \right\}$ 
```

```
Out[21]= Ly=f(Lx):
```

```
Out[22]= 
$$\frac{Lx \sqrt{n}}{\sqrt{m}}$$

```

```
Out[23]= New iNew after plugging in Ly = f(Lx)
```

$$\text{Out[24]= } \left\{ \frac{Lx n}{a (1 + m n)}, - \frac{Lx \sqrt{n}}{a \sqrt{m} (1 + m n)} \right\}$$

Out[25]= New jNew after plugging in Ly = f(Lx)

$$\text{Out[26]= } \left\{ \frac{Lx}{b (1 + m n)}, \frac{Lx \sqrt{m} \sqrt{n}}{b (1 + m n)} \right\}$$

Out[27]= Solving iNew.iNew==1:

$$\text{Out[28]= } \left\{ \left\{ Lx \rightarrow - \frac{1}{\sqrt{\frac{n}{a^2 m (1+m n)}}} \right\}, \left\{ Lx \rightarrow \frac{1}{\sqrt{\frac{n}{a^2 m (1+m n)}}} \right\} \right\}$$

Out[29]= Lx: (new periodic box length in i-direction)

$$\text{Out[30]= } \frac{1}{\sqrt{\frac{n}{a^2 m (1+m n)}}}$$

Out[31]= Ly: (new periodic box length in j-direction)

$$\text{Out[32]= } \frac{\sqrt{n}}{\sqrt{m} \sqrt{\frac{n}{a^2 m (1+m n)}}}$$

Out[35]= New iNew after plugging Ly as a f(Lx):

$$\text{Out[36]= } \left\{ \frac{n}{a \sqrt{\frac{n}{a^2 m (1+m n)}} (1 + m n)}, - \frac{\sqrt{n}}{a \sqrt{m} \sqrt{\frac{n}{a^2 m (1+m n)}} (1 + m n)} \right\}$$

Out[37]= New jNew after plugging Ly as a f(Lx):

$$\text{Out[38]= } \left\{ \frac{1}{b \sqrt{\frac{n}{a^2 m (1+m n)}} (1 + m n)}, \frac{\sqrt{m} \sqrt{n}}{b \sqrt{\frac{n}{a^2 m (1+m n)}} (1 + m n)} \right\}$$

Out[39]= New iNew.iNew: should be equal to 1

Out[40]= 1

Out[41]= New jNew.jNew: should be equal to 1

$$\text{Out[42]= } \frac{a^2 m}{b^2 n}$$

Out[43]= New iNew.jNew: should be equal to 0

Out[44]= 0

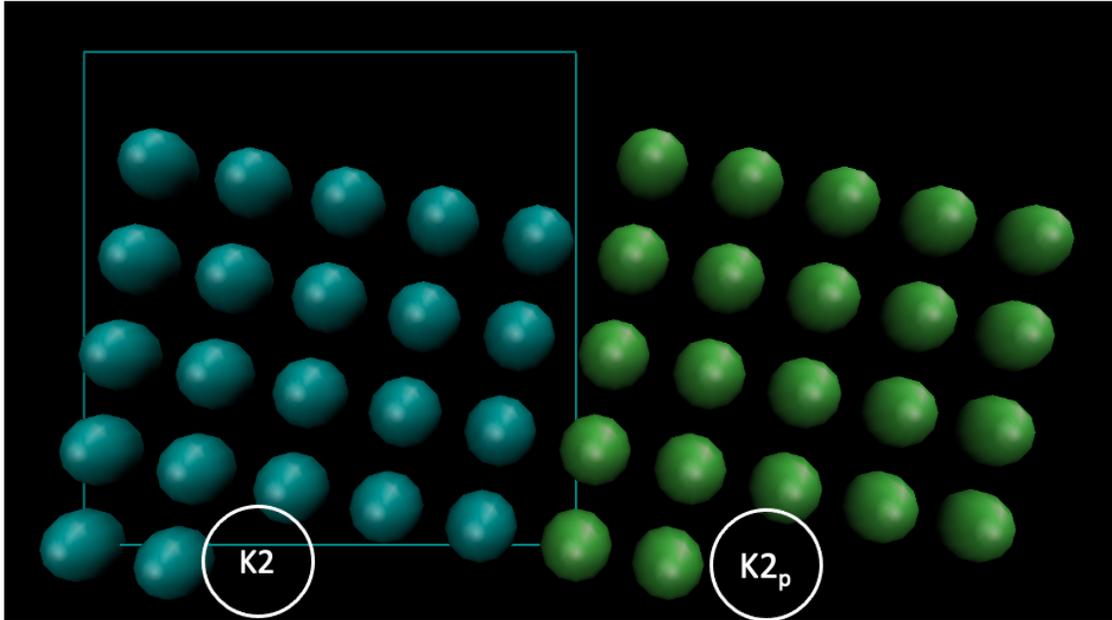
**L.1.1** Numeric Solution for  $a=b=1$  &  $\gamma=90^\circ$ 

Figure L.1: The 2D everkinked lattice for  $a=b=1$ ,  $\gamma=90^\circ$

```

In[1]:= (*Initializing crystallography*)
 $\gamma = \pi / 2;$  (*90.0 deg*)
a = 1;
b = 1;

(*Setting Supercell size*)
m = 5;
n = 5;

(*Initializing original lattice vectors*)
i = {1, 0};
j = {Cos[ $\gamma$ ], Sin[ $\gamma$ ]};

(*Initializing new lattice vectors*)
iNew = {iNew1, iNew2};
jNew = {jNew1, jNew2};

(*Setting up equations for a Ever-kinked crystal surface*)
eq4 = m a iNew + b jNew == Lx i;
eq5 = n b jNew - a iNew == Ly j;

"Solving m a iNew + b jNew == Lx i && n b jNew - a iNew == Ly j"
ans = Solve[{eq4, eq5}, {iNew1, iNew2, jNew1, jNew2}]
iNew = {ans[[1, 1, 2]], ans[[1, 2, 2]]};
jNew = {ans[[1, 3, 2]], ans[[1, 4, 2]]};

"iNew.iNew:"
iNew.iNew // Simplify
"jNew.jNew:"
jNew.jNew // Simplify

"Solving iNew.jNew==cos[ $\gamma$ ]:"
ans2 = Solve[{iNew.jNew == Cos[ $\gamma$ ]}, {Ly}]
"Ly=f(Lx):"
Ly2 = ans2[[2, 1, 2]]

"New iNew after plugging in Ly = f(Lx)"
iNew = iNew /. Ly  $\rightarrow$  Ly2

"New jNew after plugging in Ly = f(Lx)"
jNew = jNew /. Ly  $\rightarrow$  Ly2

"Solving iNew.iNew==1:"

```

```
ans3 = Solve[{iNew.iNew == 1}, {Lx}] // Simplify
```

```
"Lx: (new periodic box length in i-direction)"
```

```
Lx2 = ans3[[2, 1, 2]]
```

```
"Ly: (new periodic box length in j-direction)"
```

```
Ly2 = Ly2 /. Lx → Lx2
```

```
iNew = iNew /. Lx → Lx2;
```

```
jNew = jNew /. Lx → Lx2 /. Ly → Ly2;
```

```
"New iNew after plugging Ly as a f(Lx):"
```

```
iNew
```

```
"New jNew after plugging Ly as a f(Lx):"
```

```
jNew
```

```
"New iNew.iNew: should be equal to 1"
```

```
iNew.iNew // Simplify
```

```
"New jNew.jNew: should be equal to 1"
```

```
jNew.jNew // Simplify
```

```
"New iNew.jNew: should be equal to 0"
```

```
iNew.jNew // Simplify
```

```
Out[12]= Solving m a iNew + b jNew == Lx i && n b jNew - a iNew == Ly j
```

```
Out[13]=  $\left\{ \left\{ iNew1 \rightarrow \frac{5 Lx}{26}, iNew2 \rightarrow -\frac{Ly}{26}, jNew1 \rightarrow \frac{Lx}{26}, jNew2 \rightarrow \frac{5 Ly}{26} \right\} \right\}$ 
```

```
Out[16]= iNew.iNew:
```

```
Out[17]=  $\frac{1}{676} (25 Lx^2 + Ly^2)$ 
```

```
Out[18]= jNew.jNew:
```

```
Out[19]=  $\frac{1}{676} (Lx^2 + 25 Ly^2)$ 
```

```
Out[20]= Solving iNew.jNew==cos[γ]:
```

```
Out[21]=  $\{ \{ Ly \rightarrow -Lx \}, \{ Ly \rightarrow Lx \} \}$ 
```

```
Out[22]= Ly=f(Lx):
```

```
Out[23]= Lx
```

```
Out[24]= New iNew after plugging in Ly = f(Lx)
```

```
Out[25]=  $\left\{ \frac{5 Lx}{26}, -\frac{Lx}{26} \right\}$ 
```

```
Out[26]= New jNew after plugging in Ly = f(Lx)
```

$$\text{Out[27]} = \left\{ \frac{Lx}{26}, \frac{5 Lx}{26} \right\}$$

Out[28]= Solving  $i_{\text{New}}.i_{\text{New}}=1$ :

$$\text{Out[29]} = \left\{ \{Lx \rightarrow -\sqrt{26}\}, \{Lx \rightarrow \sqrt{26}\} \right\}$$

Out[30]=  $Lx$ : (new periodic box length in  $i$ -direction)

$$\text{Out[31]} = \sqrt{26}$$

Out[32]=  $Ly$ : (new periodic box length in  $j$ -direction)

$$\text{Out[33]} = \sqrt{26}$$

Out[36]= New  $i_{\text{New}}$  after plugging  $Ly$  as a  $f(Lx)$ :

$$\text{Out[37]} = \left\{ \frac{5}{\sqrt{26}}, -\frac{1}{\sqrt{26}} \right\}$$

Out[38]= New  $j_{\text{New}}$  after plugging  $Ly$  as a  $f(Lx)$ :

$$\text{Out[39]} = \left\{ \frac{1}{\sqrt{26}}, \frac{5}{\sqrt{26}} \right\}$$

Out[40]= New  $i_{\text{New}}.i_{\text{New}}$ : should be equal to 1

Out[41]= 1

Out[42]= New  $j_{\text{New}}.j_{\text{New}}$ : should be equal to 1

Out[43]= 1

Out[44]= New  $i_{\text{New}}.j_{\text{New}}$ : should be equal to 0

Out[45]= 0

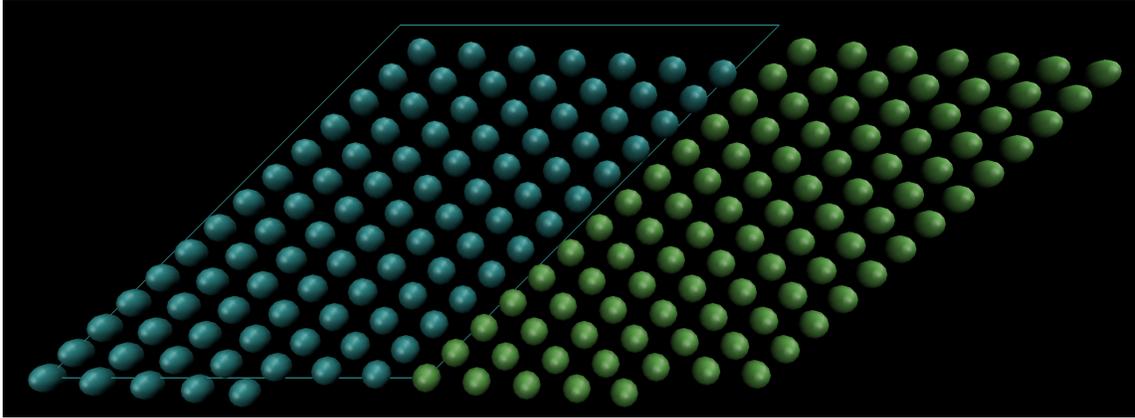
**L.1.2** Numeric Solution for  $a=1.3168$ ,  $b=1$  &  $\gamma=45^\circ$ 

Figure L.2: The 2D everkinked lattice for  $a=1.3168$ ,  $b=1$  &  $\gamma=45^\circ$

```

In[1]:= (*Initializing crystallography*)
 $\gamma = \pi / 4;$  (*45 deg*)
a = 1.3168;
b = 1.0;

(*Setting Supercell size*)
m = 7;
n = 14;

(*Initializing original lattice vectors*)
i = {1, 0};
j = {Cos[ $\gamma$ ], Sin[ $\gamma$ ]};

(*Initializing new lattice vectors*)
iNew = {iNew1, iNew2};
jNew = {jNew1, jNew2};

(*Setting up equations for a Ever-kinked crystal surface*)
eq4 = m a iNew + b jNew == Lx i;
eq5 = n b jNew - a iNew == Ly j;

"Solving m a iNew + b jNew == Lx i && n b jNew - a iNew == Ly j"
ans = Solve[{eq4, eq5}, {iNew1, iNew2, jNew1, jNew2}]
iNew = {ans[[1, 1, 2]], ans[[1, 2, 2]]};
jNew = {ans[[1, 3, 2]], ans[[1, 4, 2]]};

"iNew.iNew"
iNew.iNew // Simplify
"jNew.jNew:"
jNew.jNew // Simplify

"Solving iNew.jNew==cos[ $\gamma$ ]:"
ans2 = Solve[{iNew.jNew == Cos[ $\gamma$ ]}, {Ly}]
"Ly=f(Lx)"
Ly2 = ans2[[1, 1, 2]]
iNew = iNew /. Ly  $\rightarrow$  Ly2

"Solving iNew.iNew==1:"
ans3 = Solve[{iNew.iNew == 1}, {Lx}] // Simplify

"Lx: (new periodic box length in i-direction)"
Lx2 = ans3[[2, 1, 2]]
N[Lx2]

```

"Ly: (new periodic box length in j-direction)"

Ly2 = Ly2 /. Lx → Lx2

N[Ly2]

iNew = iNew /. Lx → Lx2;

jNew = jNew /. Lx → Lx2 /. Ly → Ly2;

"New iNew after plugging Ly as a f(Lx):"

iNew

"New jNew after plugging Ly as a f(Lx):"

jNew

"New iNew.iNew: should be equal to 1"

iNew.iNew // Simplify

N[iNew.iNew]

"New jNew.jNew: should be equal to 1"

jNew.jNew // Simplify

N[jNew.jNew]

"New iNew.jNew: should be equal to 0"

iNew.jNew // Simplify

Out[11]= Solving m a iNew + b jNew == Lx i && n b jNew - a iNew == Ly j

Out[12]=  $\{ \{ iNew1 \rightarrow 0.107392 Lx - 0.00542413 Ly, iNew2 \rightarrow 0. - 0.00542413 Ly, jNew1 \rightarrow 0.010101 Lx + 0.0499974 Ly, jNew2 \rightarrow 0. + 0.0499974 Ly \} \}$

Out[15]= iNew.iNew

Out[16]=  $0. + 0.0115331 Lx^2 - 0.00116502 Lx Ly + 0.0000588423 Ly^2$

Out[17]= jNew.jNew:

Out[18]=  $0. + 0.00010203 Lx^2 + 0.00101005 Lx Ly + 0.00499949 Ly^2$

Out[19]= Solving iNew.jNew==cos[γ]:

 **Solve:** Solve was unable to solve the system with inexact coefficients. The answer was obtained by solving a corresponding exact system and numericizing the result.

Out[20]=  $\left\{ \left\{ Ly \rightarrow 2.89462 \times 10^{-20} \left( 1.69253 \times 10^{20} Lx - 2.64575 \sqrt{-2.22278 \times 10^{41} + 4.43338 \times 10^{39} Lx^2} \right) \right\}, \left\{ Ly \rightarrow 2.89462 \times 10^{-20} \left( 1.69253 \times 10^{20} Lx + 2.64575 \sqrt{-2.22278 \times 10^{41} + 4.43338 \times 10^{39} Lx^2} \right) \right\} \right\}$

Out[21]= Ly=f(Lx)

Out[22]=  $2.89462 \times 10^{-20} \left( 1.69253 \times 10^{20} Lx - 2.64575 \sqrt{-2.22278 \times 10^{41} + 4.43338 \times 10^{39} Lx^2} \right)$

Out[23]=  $\left\{ 0.107392 Lx - 1.57008 \times 10^{-22} \left( 1.69253 \times 10^{20} Lx - 2.64575 \sqrt{-2.22278 \times 10^{41} + 4.43338 \times 10^{39} Lx^2} \right), 0. - 1.57008 \times 10^{-22} \left( 1.69253 \times 10^{20} Lx - 2.64575 \sqrt{-2.22278 \times 10^{41} + 4.43338 \times 10^{39} Lx^2} \right) \right\}$

```
Out[24]= Solving iNew.iNew==1:  
Out[25]= {{Lx → -13.1354}, {Lx → 9.94986}}  
Out[26]= Lx: (new periodic box length in i-direction)  
Out[27]= 9.94986  
Out[28]= 9.94986  
Out[29]= Ly: (new periodic box length in j-direction)  
Out[30]= 13.102  
Out[31]= 13.102  
Out[34]= New iNew after plugging Ly as a f(Lx):  
Out[35]= {0.997472, -0.071067}  
Out[36]= New jNew after plugging Ly as a f(Lx):  
Out[37]= {0.755571, 0.655067}  
Out[38]= New iNew.iNew: should be equal to 1  
Out[39]= 1.  
Out[40]= 1.  
Out[41]= New jNew.jNew: should be equal to 1  
Out[42]= 1.  
Out[43]= 1.  
Out[44]= New iNew.jNew: should be equal to 0  
Out[45]= 0.707107
```

## **L.2 Solving the 3D Ever-kinked orthorhombic lattice equations**

```

In[206]:= (*Initializing crystallography*)
 $\gamma = \pi / 2; (*90.00 \text{ deg}*)$ 
a = a;
b = b;
c = c;

(*Setting Supercell size*)
m = m;
n = n;

(*Initializing original lattice vectors*)
i = {1, 0, 0};
j = {0, 1, 0};

(*Initializing new lattice vectors*)
iNew = {iNew1, iNew2, iNew3};
jNew = {jNew1, jNew2, jNew3};
kNew = Cross[iNew, jNew];

(*Setting up equations for a Ever-kinked crystal surface*)
eq4 = m a iNew - b jNew == Lx i;
eq5 = n b jNew + a iNew + c kNew == Ly j;

"Solving m a iNew- b jNew==Lx i && n b jNew + a iNew+ c kNew==Ly j"
ans = Solve[{eq4, eq5}, {iNew1, iNew2, iNew3, jNew1, jNew2, jNew3}]
iNew = {ans[[1, 1, 2]], ans[[1, 2, 2]], ans[[1, 3, 2]]};
jNew = {ans[[1, 4, 2]], ans[[1, 5, 2]], ans[[1, 6, 2]]};

"iNew.iNew:"
iNew.iNew // Simplify
"jNew.jNew:"
jNew.jNew // Simplify

"Solving iNew.jNew==cos[\gamma]:"
ans2 = Solve[{iNew.jNew == Cos[\gamma]}, {Ly}]
"Ly=f(Lx):"
Ly2 = ans2[[2, 1, 2]]

"iNew with Ly=f(Lx):"
iNew = iNew /. Ly -> Ly2

"jNew with Ly=f(Lx):"

```

```

jNew = jNew /. Ly → Ly2

"iNew.iNew:"
iNew.iNew // Simplify

"jNew.jNew:"
jNew.jNew // Simplify

"Solving jNew.jNew==1:"
ans3 = Solve[{jNew.jNew == 1}, {Lx}] // Simplify

"Lx: (new periodic box length in i-direction)"
Lx2 = ans3[[2, 1, 2]]
"Ly: (new periodic box length in j-direction)"
Ly2 = Ly2 /. Lx → Lx2

iNew = iNew /. Lx → Lx2;
jNew = jNew /. Lx → Lx2 /. Ly → Ly2;

"New iNew after plugging Lx as a f(Ly):"
iNew
"New jNew after plugging Lx as a f(Ly):"
jNew
"New iNew.iNew: should be equal to 1"
iNew.iNew // Simplify
"New jNew.jNew: should be equal to 1"
jNew.jNew // Simplify
"New iNew.jNew: should be equal to 0"
iNew.jNew // Simplify

```

Out[219]= Solving  $m a iNew - b jNew = Lx$  and  $n b jNew + a iNew + c kNew = Ly$  j

$$\text{Out[220]= } \left\{ \left\{ iNew1 \rightarrow \frac{Lx n}{a (1 + m n)}, iNew2 \rightarrow \frac{a b^2 Ly (1 + m n)}{a^2 b^2 + c^2 Lx^2 + 2 a^2 b^2 m n + a^2 b^2 m^2 n^2}, \right. \right.$$

$$iNew3 \rightarrow -\frac{b c Lx Ly}{a^2 b^2 + c^2 Lx^2 + 2 a^2 b^2 m n + a^2 b^2 m^2 n^2},$$

$$jNew1 \rightarrow -\frac{Lx}{b (1 + m n)}, jNew2 \rightarrow \frac{a^2 b Ly m (1 + m n)}{a^2 b^2 + c^2 Lx^2 + 2 a^2 b^2 m n + a^2 b^2 m^2 n^2},$$

$$\left. \left. jNew3 \rightarrow -\frac{a c Lx Ly m}{a^2 b^2 + c^2 Lx^2 + 2 a^2 b^2 m n + a^2 b^2 m^2 n^2} \right\} \right\}$$

Out[223]= iNew.iNew:

$$\text{Out[224]= } \frac{c^2 Lx^4 n^2 + a^2 (b + b m n)^2 (Ly^2 + Lx^2 n^2)}{(a + a m n)^2 (c^2 Lx^2 + a^2 (b + b m n)^2)}$$

Out[225]= **jNew.jNew:**

$$\text{Out[226]= } \frac{c^2 Lx^4 + a^2 (Lx^2 + Ly^2 m^2) (b + b m n)^2}{(b + b m n)^2 (c^2 Lx^2 + a^2 (b + b m n)^2)}$$

Out[227]= **Solving iNew.jNew==cos[\gamma]:**

$$\text{Out[228]= } \left\{ \left\{ Ly \rightarrow - \left( (Lx \sqrt{n}) / \left( \sqrt{a} \sqrt{b} (1 + m n) \sqrt{\left( \frac{a^3 b^3 m}{(a^2 b^2 + c^2 Lx^2 + 2 a^2 b^2 m n + a^2 b^2 m^2 n^2)^2} + \frac{a b c^2 Lx^2 m}{(a^2 b^2 + c^2 Lx^2 + 2 a^2 b^2 m n + a^2 b^2 m^2 n^2)^2} + \frac{2 a^3 b^3 m^2 n}{(a^2 b^2 + c^2 Lx^2 + 2 a^2 b^2 m n + a^2 b^2 m^2 n^2)^2} + \frac{a^3 b^3 m^3 n^2}{(a^2 b^2 + c^2 Lx^2 + 2 a^2 b^2 m n + a^2 b^2 m^2 n^2)^2} \right)^2} \right) \right\}, \right. \\ \left. \left\{ Ly \rightarrow (Lx \sqrt{n}) / \left( \sqrt{a} \sqrt{b} (1 + m n) \sqrt{\left( \frac{a^3 b^3 m}{(a^2 b^2 + c^2 Lx^2 + 2 a^2 b^2 m n + a^2 b^2 m^2 n^2)^2} + \frac{a b c^2 Lx^2 m}{(a^2 b^2 + c^2 Lx^2 + 2 a^2 b^2 m n + a^2 b^2 m^2 n^2)^2} + \frac{2 a^3 b^3 m^2 n}{(a^2 b^2 + c^2 Lx^2 + 2 a^2 b^2 m n + a^2 b^2 m^2 n^2)^2} + \frac{a^3 b^3 m^3 n^2}{(a^2 b^2 + c^2 Lx^2 + 2 a^2 b^2 m n + a^2 b^2 m^2 n^2)^2} \right)^2} \right) \right\} \right\}$$

Out[229]= **Ly=f(Lx):**

$$\text{Out[230]= } (Lx \sqrt{n}) / \left( \sqrt{a} \sqrt{b} (1 + m n) \sqrt{\left( \frac{a^3 b^3 m}{(a^2 b^2 + c^2 Lx^2 + 2 a^2 b^2 m n + a^2 b^2 m^2 n^2)^2} + \frac{a b c^2 Lx^2 m}{(a^2 b^2 + c^2 Lx^2 + 2 a^2 b^2 m n + a^2 b^2 m^2 n^2)^2} + \frac{2 a^3 b^3 m^2 n}{(a^2 b^2 + c^2 Lx^2 + 2 a^2 b^2 m n + a^2 b^2 m^2 n^2)^2} + \frac{a^3 b^3 m^3 n^2}{(a^2 b^2 + c^2 Lx^2 + 2 a^2 b^2 m n + a^2 b^2 m^2 n^2)^2} \right)^2} \right)$$

Out[231]= **iNew with Ly=f(Lx):**

$$\begin{aligned}
 \text{Out[232]} = & \left\{ \frac{Lx\ n}{a\ (1 + m\ n)}, (\sqrt{a}\ b^{3/2}\ Lx\ \sqrt{n}) / \left( (a^2\ b^2 + c^2\ Lx^2 + 2\ a^2\ b^2\ m\ n + a^2\ b^2\ m^2\ n^2) \right. \right. \\
 & \left. \left. \sqrt{\left( \frac{a^3\ b^3\ m}{(a^2\ b^2 + c^2\ Lx^2 + 2\ a^2\ b^2\ m\ n + a^2\ b^2\ m^2\ n^2)^2} + \frac{a\ b\ c^2\ Lx^2\ m}{(a^2\ b^2 + c^2\ Lx^2 + 2\ a^2\ b^2\ m\ n + a^2\ b^2\ m^2\ n^2)^2} + \right. \right. \\
 & \left. \left. \frac{2\ a^3\ b^3\ m^2\ n}{(a^2\ b^2 + c^2\ Lx^2 + 2\ a^2\ b^2\ m\ n + a^2\ b^2\ m^2\ n^2)^2} + \frac{a^3\ b^3\ m^3\ n^2}{(a^2\ b^2 + c^2\ Lx^2 + 2\ a^2\ b^2\ m\ n + a^2\ b^2\ m^2\ n^2)^2} \right) \right\}, \\
 & - \left( (\sqrt{b}\ c\ Lx^2\ \sqrt{n}) / \left( \sqrt{a}\ (1 + m\ n)\ (a^2\ b^2 + c^2\ Lx^2 + 2\ a^2\ b^2\ m\ n + a^2\ b^2\ m^2\ n^2) \right. \right. \\
 & \left. \left. \sqrt{\left( \frac{a^3\ b^3\ m}{(a^2\ b^2 + c^2\ Lx^2 + 2\ a^2\ b^2\ m\ n + a^2\ b^2\ m^2\ n^2)^2} + \right. \right. \right. \\
 & \left. \left. \frac{a\ b\ c^2\ Lx^2\ m}{(a^2\ b^2 + c^2\ Lx^2 + 2\ a^2\ b^2\ m\ n + a^2\ b^2\ m^2\ n^2)^2} + \frac{2\ a^3\ b^3\ m^2\ n}{(a^2\ b^2 + c^2\ Lx^2 + 2\ a^2\ b^2\ m\ n + a^2\ b^2\ m^2\ n^2)^2} + \right. \right. \\
 & \left. \left. \left. \frac{a^3\ b^3\ m^3\ n^2}{(a^2\ b^2 + c^2\ Lx^2 + 2\ a^2\ b^2\ m\ n + a^2\ b^2\ m^2\ n^2)^2} \right) \right) \right\}
 \end{aligned}$$

Out[233]= jNew with Ly=f(Lx) :

$$\begin{aligned}
 \text{Out[234]} = & \left\{ -\frac{Lx}{b\ (1 + m\ n)}, (a^{3/2}\ \sqrt{b}\ Lx\ m\ \sqrt{n}) / \left( (a^2\ b^2 + c^2\ Lx^2 + 2\ a^2\ b^2\ m\ n + a^2\ b^2\ m^2\ n^2) \right. \right. \\
 & \left. \left. \sqrt{\left( \frac{a^3\ b^3\ m}{(a^2\ b^2 + c^2\ Lx^2 + 2\ a^2\ b^2\ m\ n + a^2\ b^2\ m^2\ n^2)^2} + \frac{a\ b\ c^2\ Lx^2\ m}{(a^2\ b^2 + c^2\ Lx^2 + 2\ a^2\ b^2\ m\ n + a^2\ b^2\ m^2\ n^2)^2} + \right. \right. \\
 & \left. \left. \frac{2\ a^3\ b^3\ m^2\ n}{(a^2\ b^2 + c^2\ Lx^2 + 2\ a^2\ b^2\ m\ n + a^2\ b^2\ m^2\ n^2)^2} + \frac{a^3\ b^3\ m^3\ n^2}{(a^2\ b^2 + c^2\ Lx^2 + 2\ a^2\ b^2\ m\ n + a^2\ b^2\ m^2\ n^2)^2} \right) \right\}, \\
 & - \left( (\sqrt{a}\ c\ Lx^2\ m\ \sqrt{n}) / \left( \sqrt{b}\ (1 + m\ n)\ (a^2\ b^2 + c^2\ Lx^2 + 2\ a^2\ b^2\ m\ n + a^2\ b^2\ m^2\ n^2) \right. \right. \\
 & \left. \left. \sqrt{\left( \frac{a^3\ b^3\ m}{(a^2\ b^2 + c^2\ Lx^2 + 2\ a^2\ b^2\ m\ n + a^2\ b^2\ m^2\ n^2)^2} + \right. \right. \right. \\
 & \left. \left. \frac{a\ b\ c^2\ Lx^2\ m}{(a^2\ b^2 + c^2\ Lx^2 + 2\ a^2\ b^2\ m\ n + a^2\ b^2\ m^2\ n^2)^2} + \frac{2\ a^3\ b^3\ m^2\ n}{(a^2\ b^2 + c^2\ Lx^2 + 2\ a^2\ b^2\ m\ n + a^2\ b^2\ m^2\ n^2)^2} + \right. \right. \\
 & \left. \left. \left. \frac{a^3\ b^3\ m^3\ n^2}{(a^2\ b^2 + c^2\ Lx^2 + 2\ a^2\ b^2\ m\ n + a^2\ b^2\ m^2\ n^2)^2} \right) \right) \right\}
 \end{aligned}$$

Out[235]= **iNew.iNew:**

$$\text{Out[236]= } \frac{Lx^2 n}{a^2 m (1 + m n)}$$

Out[237]= **jNew.jNew:**

$$\text{Out[238]= } \frac{Lx^2}{b^2 (1 + m n)}$$

Out[239]= **Solving jNew.jNew==1:**

$$\text{Out[240]= } \left\{ \left\{ Lx \rightarrow -b \sqrt{1 + m n} \right\}, \left\{ Lx \rightarrow b \sqrt{1 + m n} \right\} \right\}$$

Out[241]= **Lx: (new periodic box length in i-direction)**

$$\text{Out[242]= } b \sqrt{1 + m n}$$

Out[243]= **Ly: (new periodic box length in j-direction)**

$$\text{Out[244]= } \left( \sqrt{b} \sqrt{n} \right) / \left( \sqrt{a} \sqrt{1 + m n} \sqrt{\left( \frac{a^3 b^3 m}{(a^2 b^2 + 2 a^2 b^2 m n + a^2 b^2 m^2 n^2 + b^2 c^2 (1 + m n))^2} + \frac{2 a^3 b^3 m^2 n}{(a^2 b^2 + 2 a^2 b^2 m n + a^2 b^2 m^2 n^2 + b^2 c^2 (1 + m n))^2} + \frac{a^3 b^3 m^3 n^2}{(a^2 b^2 + 2 a^2 b^2 m n + a^2 b^2 m^2 n^2 + b^2 c^2 (1 + m n))^2} + \frac{a b^3 c^2 m (1 + m n)}{(a^2 b^2 + 2 a^2 b^2 m n + a^2 b^2 m^2 n^2 + b^2 c^2 (1 + m n))^2} \right)} \right)$$

Out[247]= **New iNew after plugging Lx as a f(Ly):**

$$\begin{aligned}
\text{Out[248]= } & \left\{ \frac{b n}{a \sqrt{1+m n}}, (\sqrt{a} b^{5/2} \sqrt{n} \sqrt{1+m n}) / \left( (a^2 b^2 + 2 a^2 b^2 m n + a^2 b^2 m^2 n^2 + b^2 c^2 (1+m n)) \right) \right. \\
& \sqrt{\left( \frac{a^3 b^3 m}{(a^2 b^2 + 2 a^2 b^2 m n + a^2 b^2 m^2 n^2 + b^2 c^2 (1+m n))^2} + \right.} \\
& \frac{2 a^3 b^3 m^2 n}{(a^2 b^2 + 2 a^2 b^2 m n + a^2 b^2 m^2 n^2 + b^2 c^2 (1+m n))^2} + \\
& \frac{a^3 b^3 m^3 n^2}{(a^2 b^2 + 2 a^2 b^2 m n + a^2 b^2 m^2 n^2 + b^2 c^2 (1+m n))^2} + \\
& \left. \left. \frac{a b^3 c^2 m (1+m n)}{(a^2 b^2 + 2 a^2 b^2 m n + a^2 b^2 m^2 n^2 + b^2 c^2 (1+m n))^2} \right) \right), \\
& - \left( (b^{5/2} c \sqrt{n}) / \left( \sqrt{a} (a^2 b^2 + 2 a^2 b^2 m n + a^2 b^2 m^2 n^2 + b^2 c^2 (1+m n)) \right) \right. \\
& \sqrt{\left( \frac{a^3 b^3 m}{(a^2 b^2 + 2 a^2 b^2 m n + a^2 b^2 m^2 n^2 + b^2 c^2 (1+m n))^2} + \right.} \\
& \frac{2 a^3 b^3 m^2 n}{(a^2 b^2 + 2 a^2 b^2 m n + a^2 b^2 m^2 n^2 + b^2 c^2 (1+m n))^2} + \\
& \frac{a^3 b^3 m^3 n^2}{(a^2 b^2 + 2 a^2 b^2 m n + a^2 b^2 m^2 n^2 + b^2 c^2 (1+m n))^2} + \\
& \left. \left. \frac{a b^3 c^2 m (1+m n)}{(a^2 b^2 + 2 a^2 b^2 m n + a^2 b^2 m^2 n^2 + b^2 c^2 (1+m n))^2} \right) \right) \left. \right\}
\end{aligned}$$

Out[249]= New jNew after plugging Lx as a f(Ly):

$$\begin{aligned}
\text{Out[250]} = & \left\{ -\frac{1}{\sqrt{1+mn}}, (a^{3/2} b^{3/2} m \sqrt{n} \sqrt{1+mn}) / \left( (a^2 b^2 + 2 a^2 b^2 m n + a^2 b^2 m^2 n^2 + b^2 c^2 (1+mn)) \right) \right. \\
& \sqrt{\left( \frac{a^3 b^3 m}{(a^2 b^2 + 2 a^2 b^2 m n + a^2 b^2 m^2 n^2 + b^2 c^2 (1+mn))^2} + \right.} \\
& \frac{2 a^3 b^3 m^2 n}{(a^2 b^2 + 2 a^2 b^2 m n + a^2 b^2 m^2 n^2 + b^2 c^2 (1+mn))^2} + \\
& \frac{a^3 b^3 m^3 n^2}{(a^2 b^2 + 2 a^2 b^2 m n + a^2 b^2 m^2 n^2 + b^2 c^2 (1+mn))^2} + \\
& \left. \left. \frac{a b^3 c^2 m (1+mn)}{(a^2 b^2 + 2 a^2 b^2 m n + a^2 b^2 m^2 n^2 + b^2 c^2 (1+mn))^2} \right) \right\}, \\
& - \left( (\sqrt{a} b^{3/2} c m \sqrt{n}) / \left( (a^2 b^2 + 2 a^2 b^2 m n + a^2 b^2 m^2 n^2 + b^2 c^2 (1+mn)) \right) \right. \\
& \sqrt{\left( \frac{a^3 b^3 m}{(a^2 b^2 + 2 a^2 b^2 m n + a^2 b^2 m^2 n^2 + b^2 c^2 (1+mn))^2} + \right.} \\
& \frac{2 a^3 b^3 m^2 n}{(a^2 b^2 + 2 a^2 b^2 m n + a^2 b^2 m^2 n^2 + b^2 c^2 (1+mn))^2} + \\
& \frac{a^3 b^3 m^3 n^2}{(a^2 b^2 + 2 a^2 b^2 m n + a^2 b^2 m^2 n^2 + b^2 c^2 (1+mn))^2} + \\
& \left. \left. \frac{a b^3 c^2 m (1+mn)}{(a^2 b^2 + 2 a^2 b^2 m n + a^2 b^2 m^2 n^2 + b^2 c^2 (1+mn))^2} \right) \right) \left. \right\}
\end{aligned}$$

Out[251]= New iNew.iNew: should be equal to 1

$$\text{Out[252]} = \frac{b^2 n}{a^2 m}$$

Out[253]= New jNew.jNew: should be equal to 1

Out[254]= 1

Out[255]= New iNew.jNew: should be equal to 0

Out[256]= 0

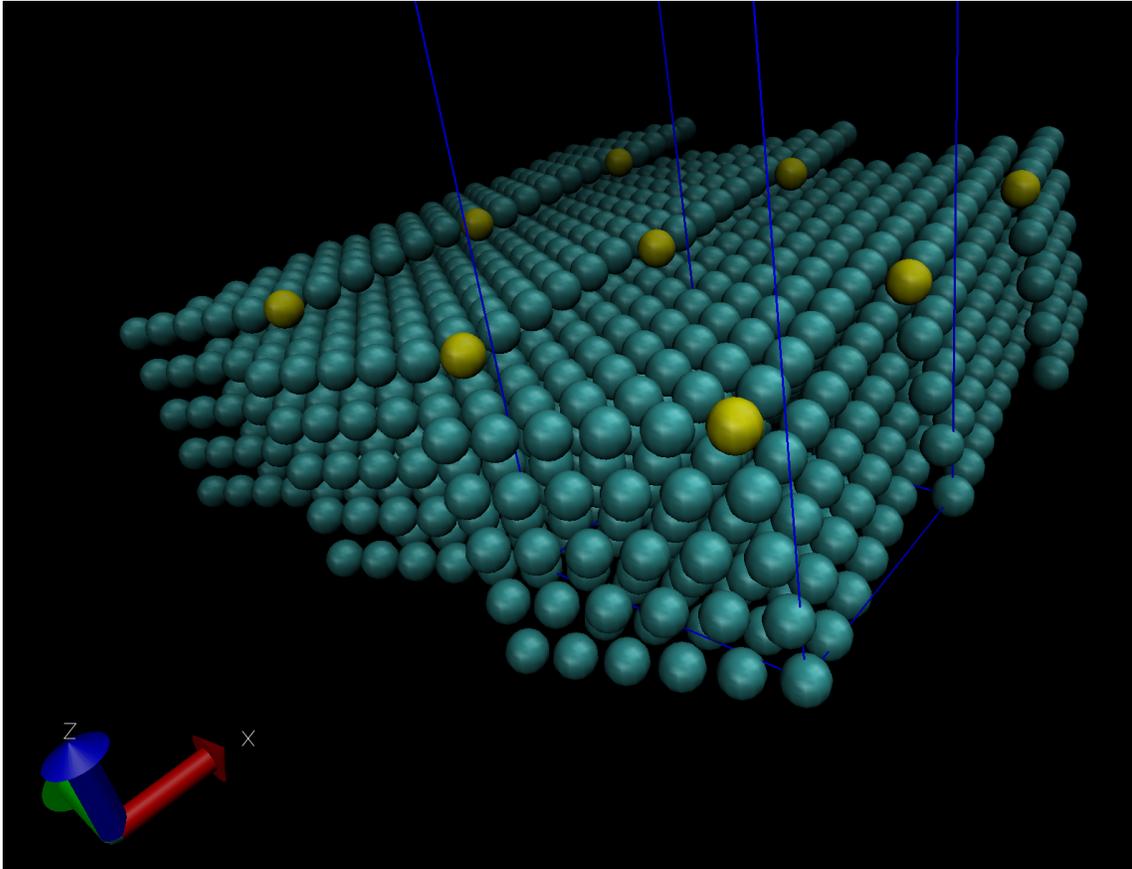
**L.2.1 Numeric Solution for  $a=b=c=1$  &  $\alpha=\beta=\gamma=90^\circ$** 

Figure L.3: The 3D everkinked lattice for  $a=b=c=1$ ,  $\alpha=\beta=\gamma=90^\circ$

```

In[54]:= (*Initializing crystallography*)
 $\gamma = \pi / 2;$  (*90.00 deg*)
a = 1;
b = 1;
c = 1;

(*Setting Supercell size*)
m = 5;
n = 5;

(*Initializing original lattice vectors*)
i = {1, 0, 0};
j = {0, 1, 0};

(*Initializing new lattice vectors*)
iNew = {iNew1, iNew2, iNew3};
jNew = {jNew1, jNew2, jNew3};
kNew = Cross[iNew, jNew];

(*Setting up equations for a Ever-kinked crystal surface*)
eq4 = m a iNew - b jNew == Lx i;
eq5 = n b jNew + a iNew + c kNew == Ly j;

"Solving m a iNew - b jNew == Lx i && n b jNew + a iNew + c kNew == Ly j"
ans = Solve[{eq4, eq5}, {iNew1, iNew2, iNew3, jNew1, jNew2, jNew3}]
iNew = {ans[[1, 1, 2]], ans[[1, 2, 2]], ans[[1, 3, 2]]};
jNew = {ans[[1, 4, 2]], ans[[1, 5, 2]], ans[[1, 6, 2]]};

"iNew.iNew:"
iNew.iNew // Simplify
"jNew.jNew:"
jNew.jNew // Simplify

"Solving iNew.jNew==cos[ $\gamma$ ]:"
ans2 = Solve[{iNew.jNew == Cos[ $\gamma$ ]}, {Ly}]
"Ly=f(Lx):"
Ly2 = ans2[[2, 1, 2]]

"iNew with Ly=f(Lx):"

```

```

iNew = iNew /. Ly → Ly2

"jNew with Ly=f(Lx):"
jNew = jNew /. Ly → Ly2

"iNew.iNew:"
iNew.iNew // Simplify

"jNew.jNew:"
jNew.jNew // Simplify

"Solving jNew.jNew==1:"
ans3 = Solve[{jNew.jNew == 1}, {Lx}] // Simplify

"Lx: (new periodic box length in i-direction)"
Lx2 = ans3[[2, 1, 2]]
N[Lx2]
"Ly: (new periodic box length in j-direction)"
Ly2 = Ly2 /. Lx → Lx2
N[Ly2]

iNew = iNew /. Lx → Lx2;
jNew = jNew /. Lx → Lx2 /. Ly → Ly2;

"New iNew after plugging Lx as a f(Ly):"
iNew
"New jNew after plugging Lx as a f(Ly):"
jNew
"New iNew.iNew: should be equal to 1"
iNew.iNew // Simplify
"New jNew.jNew: should be equal to 1"
jNew.jNew // Simplify
"New iNew.jNew: should be equal to 0"
iNew.jNew // Simplify

```

Out[67]= Solving  $ma iNew - b jNew = Lx i$  &&  $n b jNew + a iNew + c kNew = Ly j$

$$\text{Out[68]= } \left\{ \left\{ iNew1 \rightarrow \frac{5 Lx}{26}, iNew2 \rightarrow \frac{26 Ly}{676 + Lx^2}, iNew3 \rightarrow -\frac{Lx Ly}{676 + Lx^2}, \right. \right. \\
\left. \left. jNew1 \rightarrow -\frac{Lx}{26}, jNew2 \rightarrow \frac{130 Ly}{676 + Lx^2}, jNew3 \rightarrow -\frac{5 Lx Ly}{676 + Lx^2} \right\} \right\}$$

Out[71]= iNew.iNew:

$$\text{Out[72]= } \frac{16\,900\,Lx^2 + 25\,Lx^4 + 676\,Ly^2}{676\,(676 + Lx^2)}$$

Out[73]= **jNew.jNew:**

$$\text{Out[74]= } \frac{676\,Lx^2 + Lx^4 + 16\,900\,Ly^2}{676\,(676 + Lx^2)}$$

Out[75]= **Solving iNew.jNew==cos[\gamma]:**

$$\text{Out[76]= } \left\{ \left\{ Ly \rightarrow -\frac{Lx}{26\sqrt{\frac{676}{(676+Lx^2)^2} + \frac{Lx^2}{(676+Lx^2)^2}}} \right\}, \left\{ Ly \rightarrow \frac{Lx}{26\sqrt{\frac{676}{(676+Lx^2)^2} + \frac{Lx^2}{(676+Lx^2)^2}}} \right\} \right\}$$

Out[77]= **Ly=f(Lx):**

$$\text{Out[78]= } \frac{Lx}{26\sqrt{\frac{676}{(676+Lx^2)^2} + \frac{Lx^2}{(676+Lx^2)^2}}}$$

Out[79]= **iNew with Ly=f(Lx):**

$$\text{Out[80]= } \left\{ \frac{5\,Lx}{26}, \frac{Lx}{(676 + Lx^2)\sqrt{\frac{676}{(676+Lx^2)^2} + \frac{Lx^2}{(676+Lx^2)^2}}}, -\frac{Lx^2}{26\,(676 + Lx^2)\sqrt{\frac{676}{(676+Lx^2)^2} + \frac{Lx^2}{(676+Lx^2)^2}}} \right\}$$

Out[81]= **jNew with Ly=f(Lx):**

$$\text{Out[82]= } \left\{ -\frac{Lx}{26}, \frac{5\,Lx}{(676 + Lx^2)\sqrt{\frac{676}{(676+Lx^2)^2} + \frac{Lx^2}{(676+Lx^2)^2}}}, -\frac{5\,Lx^2}{26\,(676 + Lx^2)\sqrt{\frac{676}{(676+Lx^2)^2} + \frac{Lx^2}{(676+Lx^2)^2}}} \right\}$$

Out[83]= **iNew.iNew:**

$$\text{Out[84]= } \frac{Lx^2}{26}$$

Out[85]= **jNew.jNew:**

$$\text{Out[86]= } \frac{Lx^2}{26}$$

Out[87]= **Solving jNew.jNew==1:**

$$\text{Out[88]= } \left\{ \left\{ Lx \rightarrow -\sqrt{26} \right\}, \left\{ Lx \rightarrow \sqrt{26} \right\} \right\}$$

Out[89]= **Lx: (new periodic box length in i-direction)**

$$\text{Out[90]= } \sqrt{26}$$

Out[91]= **5.09902**

Out[92]= **Ly: (new periodic box length in j-direction)**

$$\text{Out[93]= } 3\sqrt{3}$$

Out[94]= 5.19615

Out[97]= New iNew after plugging Lx as a f(Ly):

$$\text{Out[98]= } \left\{ \frac{5}{\sqrt{26}}, \frac{1}{3\sqrt{3}}, -\frac{1}{3\sqrt{78}} \right\}$$

Out[99]= New jNew after plugging Lx as a f(Ly):

$$\text{Out[100]= } \left\{ -\frac{1}{\sqrt{26}}, \frac{5}{3\sqrt{3}}, -\frac{5}{3\sqrt{78}} \right\}$$

Out[101]= New iNew.iNew: should be equal to 1

Out[102]= 1

Out[103]= New jNew.jNew: should be equal to 1

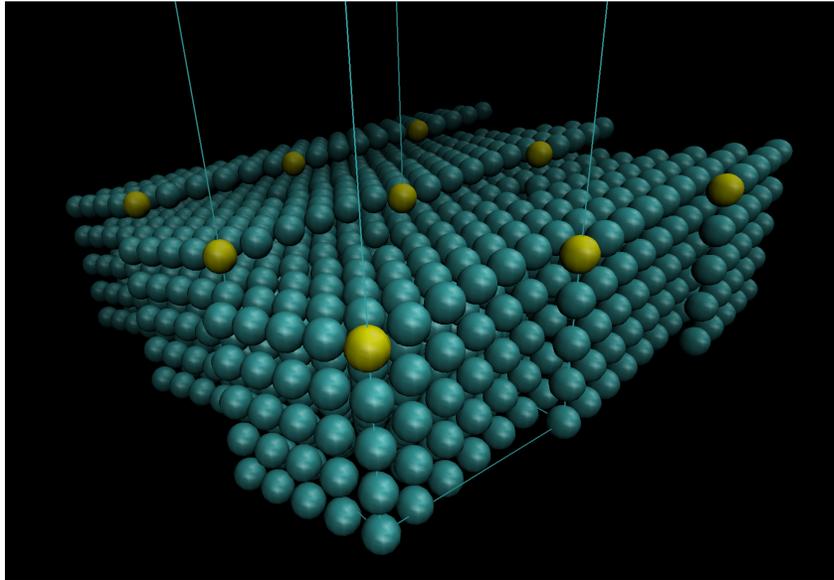
Out[104]= 1

Out[105]= New iNew.jNew: should be equal to 0

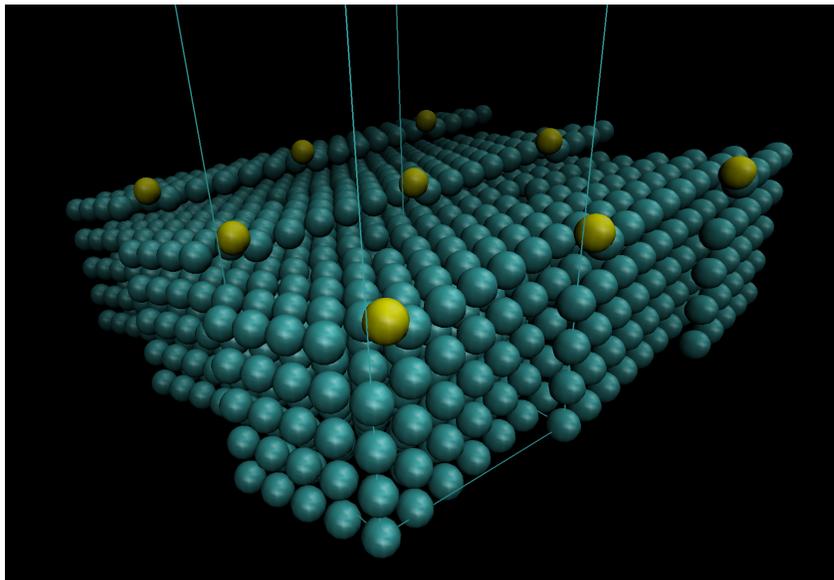
Out[106]= 0

## L.2.2 Solving the 3D Ever-kinked “Hole”

The tilting and twisting of a crystal system leads to the formation of holes in the system as shown in Figure L.4. This is fixed by adding an extra column of unit cells as shown in Figure L.5.



(a)



(b)

Figure L.4: Ever-kinked cubic lattice with holes next to the kinks: (a) Yellow atoms are the ones sitting in a potential kink site, (b): Periodic kinks being displayed via dislodging an atom sitting in a potential kink site.

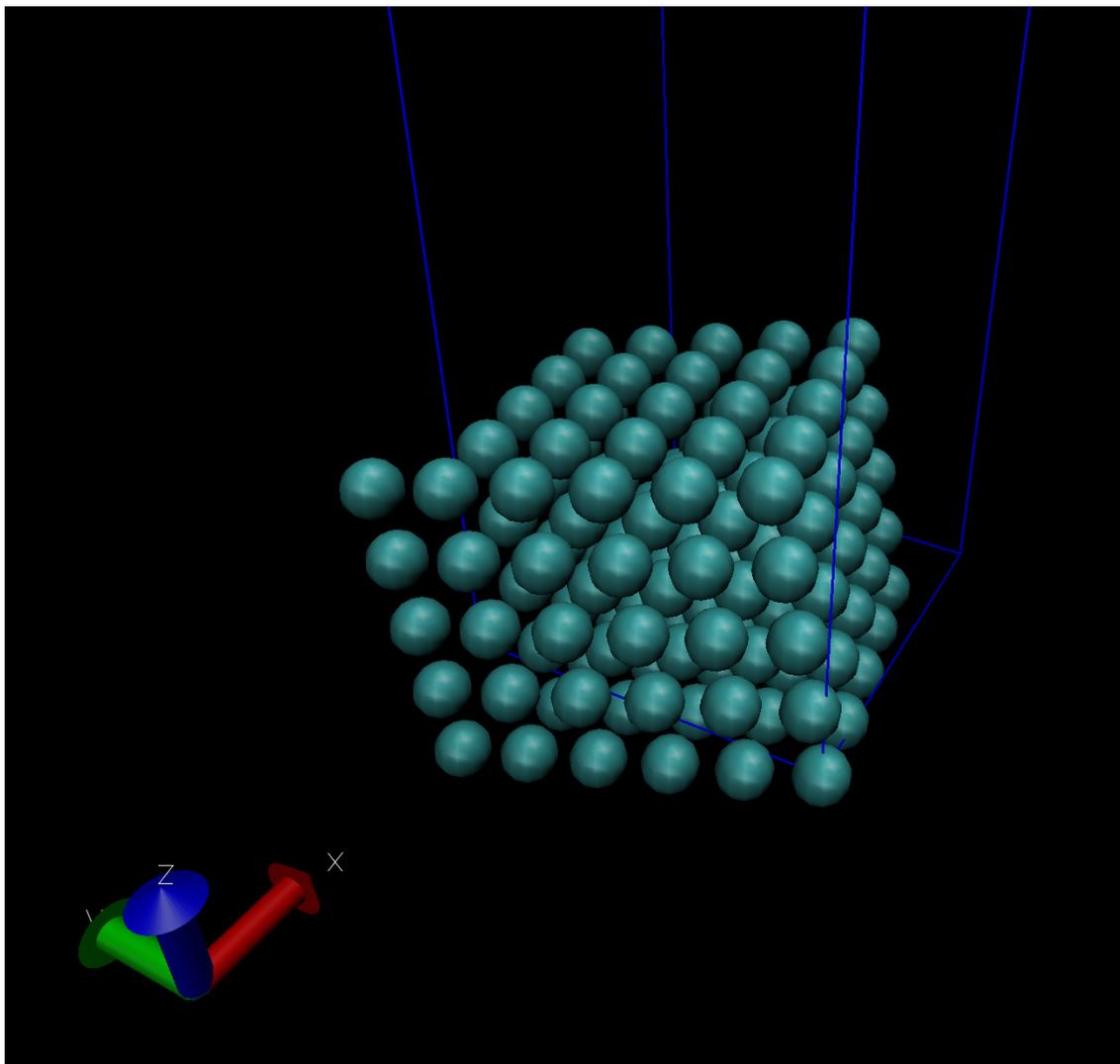


Figure L.5: The 3D ever-kinked lattice for  $a=b=c=1$ ,  $\alpha=\beta=\gamma=90^\circ$ , with an added column of unit cells to fix the “hole”

### L.2.3 A case study to demonstrate the approximate solution strategy for creating ever-kinked crystal systems

Consider the *beta*-succinic acid crystal system. The GAFF predicted lattice constants are shown in Table L.1. To ease the problem we first approximate the  $\beta$  angle to  $90^\circ$ .

Therefore, now we just need to choose  $m$  and  $n$  such that

$$\frac{b^2}{a^2} = \frac{m}{n} \quad (\text{L.1})$$

However, for the GAFF predicted lattice constants, the ratio  $b^2/a^2$  is irrational. Therefore, we need to approximate  $a$ ,  $b$  or both, such that  $b_{approx}^2/a_{approx}^2$  is rational. Table L.2 shows the results for different values of  $\delta_a$  and  $\delta_b$ , that rationalize  $b_{approx}^2/a_{approx}^2$  within  $b^2/a^2 \pm \delta$ . As seen, as we relax the tolerance on the approximation ( $\delta$ ), we achieve computationally feasible supercell sizes.

Table L.1: GAFF predicted lattice parameters of succinic acid crystal —  $\beta$  polymorph

Lattice parameter	Expt.	GAFF	% $\Delta$
$a(\text{\AA})$	$5.519 \pm 0.002$	$5.473 \pm 0.001$	-0.83%
$b(\text{\AA})$	$8.862 \pm 0.006$	$9.068 \pm 0.001$	2.32%
$c(\text{\AA})$	$5.101 \pm 0.001$	$5.314 \pm 0.001$	4.18%
$\alpha$ ( $^\circ$ )	$90.00 \pm 0.00$	$89.999 \pm 0.003$	0.00%
$\beta$ ( $^\circ$ )	$91.59 \pm 0.04$	$89.624 \pm 0.004$	-2.15%
$\gamma$ ( $^\circ$ )	$90.00 \pm 0.00$	$89.997 \pm 0.003$	0.00%

Table L.2: Generating approximate ever-kinked solutions within  $\delta$  neighborhood

$\delta$	0.1	0.01	0.001	0.00001	0.000001
$\frac{b^2}{a^2} _{approx}$	$\frac{8}{3}$	$\frac{11}{4}$	$\frac{129}{47}$	$\frac{140}{51}$	$\frac{571}{208}$
$\delta_b\%$	-1.44	0.09	-0.01	-0.00	0.00
$\delta_a\%$	1.46	-0.09	0.01	0.00	-0.00