

SAPIENZA
UNIVERSITÀ DI ROMA

Mix, Match, and Assemble:
an inverse design approach to tailor materials

Facoltà di Scienze Matematiche Fisiche e Naturali
Dottorato di Ricerca in Fisica (XXXVIII cycle)

Camilla Beneduce
ID number 1756367

Advisors
Prof. John Russo
Prof. Francesco Sciortino

Academic Year 2025/2026

Mix, Match, and Assemble: an inverse design approach to tailor materials
PhD thesis. Sapienza University of Rome

© 2025 Camilla Beneduce. All rights reserved

This thesis has been typeset by LaTeX and the Sapthesis class.

Author's email: camilla.beneduce@uniroma1.it

Abstract

Inverse self-assembly seeks to design building units, through their shape and interactions, that spontaneously organise into a target structure. This is challenging due to kinetics pathways and the emergence of alternative, often unforeseen, structures. A promising route employs multiple building units with simple shapes and interaction potentials, achieving selectivity through specific, directional interactions. However, mixtures introduce a combinatorial challenge in exploring the space of candidate designs and add thermodynamic complexity: phase behaviour and self-assembly become more intricate, the latter often occurring via non-classical nucleation pathways. The SAT-assembly approach addresses the first challenge by mapping a patchy particle design problem into Boolean equations. Here, we focus on the second challenge: using theoretical analysis and numerical simulations, we study the thermodynamics of mixtures, identify optimal self-assembly conditions, and encode them directly into the design. Crystallisation often proceeds via a two-step nucleation process. While liquid-vapour metastability is known to enhance nucleation through density fluctuations, its coupling with composition fluctuations in mixtures remains poorly understood. We find that higher nucleation rates and larger assemblies emerge when the species concentration in the liquid phase matches that of the target structure. However, during phase separation, mixtures do not preserve the initial component ratio, making it difficult to identify starting conditions for optimal self-assembly. Our strategy is to design azeotropes and incorporate them into the nucleation pathway, since at the azeotrope vapour and liquid phases coexist at the same starting composition, mimicking one-component systems. We derive design rules linking bond connectivity and energy to locate the azeotrope at a desired concentration, integrating these as supplementary clauses into SAT-assembly. In this way, we extend inverse self-assembly into an inverse-thermodynamic framework, capable of engineering not only structural but also thermodynamic properties. Beyond design identification, SAT-assembly enumerates all possible ways to populate a lattice with particles of a given design, thereby listing all polymorphs, which interestingly share the same free energy. We use them to build a falsifiability test for classical nucleation theory (CNT), challenging the assumption that macroscopic properties alone predict nucleation outcomes. These polymorphs can also have distinct photonic properties, motivating us to explore how to design chiral crystals using the SAT-assembly framework. Ultimately, we outline a protocol from the design to the experimental realisation with DNA nanotechnology, towards creating colloidal photonic crystals. DNA is a perfect construction material due to its predictable and selective base-pairing, and the SAT-assembly naturally integrates the concept of complementarity to control interactions.

Acknowledgments

My greatest thanks go to my supervisors, Prof. John Russo and Prof. Francesco Sciortino, who guided me with patience and passion throughout my PhD journey. They have been fundamental in helping me discover the research world, always supporting my first steps in journal writing and conference participation.

I would also like to express my particular gratitude to my referees, Prof. Achille Giacometti and Dr. Fausto Martelli, who provided me with valuable and much appreciated suggestions to improve my thesis.

Finally, I thank Prof. Petr Šulc and his research group for enriching my experience through stimulating collaborations.

Contents

Introduction **1**

1 Self-assembly **7**
1.1 Direct and inverse self-assembly 7
1.2 Soft Matter: the world of tunability 9

2 SAT-assembly **11**
2.1 Patchy Particles 11
2.2 A satisfiability approach to a design problem 14
2.2.1 Variables and clauses 16
2.2.2 Sat-solvers 21
2.3 Cubic diamond crystal 22

3 Patchy Particle's mixtures: theory, simulations and experiments **27**
3.1 DNA nanotechnology 27
3.1.1 From patchy particles to DNA-origami 28
3.1.2 An algorithm to generate DNA strands from the interaction matrix 31
3.2 Wertheim's perturbation theory 36
3.3 Monte Carlo simulations 39
3.3.1 Metropolis algorithm and roto-translational moves 40
3.3.2 AVB moves 41
3.3.3 Umbrella sampling 42

4 Reconstructing phase diagrams **45**
4.1 Isochoric thermodynamics 45
4.2 Gibbs ensemble simulations 47
4.3 Phase behaviour of mixtures and azeotropy 49
4.4 Phase diagram of the N2c8 SAT-designed binary mixture 51

5 Studying nucleation of multi-component systems 59
5.1 Order parameters . . . 59
5.2 Density fluctuations and two step nucleation . . . 62
5.3 Composition fluctuations and azeotropy . . . 66

6 A falsifiability test for Classical Nucleation Theory 73
6.1 CNT . . . 73
6.2 The model system . . . 74
6.2.1 Three isotypic polymorphs . . . 74
6.2.2 Liquid-solid interfacial tension with BSUS . . . 77
6.3 The test . . . 78
6.3.1 A new order parameter . . . 78
6.3.2 Nucleation properties of the isotypes . . . 80
6.4 CNT failure . . . 83

7 Inverse thermodynamics 89
7.1 Designing interactions for targeted phase behaviour . . . 89
7.2 Inverse design of azeotropic points . . . 90
7.3 Azeotropy from bond connectivity rules . . . 92
7.3.1 Bond exclusivity condition . . . 94
7.3.2 Bond multiplicity condition . . . 95
7.3.3 Fully-connected bond condition . . . 97
7.4 Azeotropy from bond energy control . . . 98
7.5 Examples . . . 100
7.5.1 Bond exclusivity . . . 100
7.5.2 Bond multiplicity in a binary mixture . . . 101
7.5.3 Bond multiplicity in a ternary mixture . . . 103
7.5.4 Fully connected . . . 108
7.5.5 Bonding energy control . . . 116

8 Controlling optical activity of multi-component assemblies 121
8.1 The importance of isotypes . . . 121
8.2 A chiral isotype of the cubic diamond crystal . . . 122
8.3 A heterogeneous nucleation approach . . . 124
8.4 An homogeneous nucleation approach . . . 127

9 Conclusions 131

A Boolean clauses in the SAT-assembly framework 137

B Overview of Wertheim perturbation theory **143**

C Stability analysis of mixtures **149**

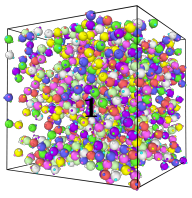
D Overview of Classical Nucleation Theory **151**

E The Kirkwood-Buff theory and the evaluation of chemical potentials in simulations **155**

Bibliography **167**

Introduction

From informatics to medicine, an increasing number of fields depend on the fabrication of highly customised materials. The desired functionalities often manifest at the macroscopic scale, but their emergence is governed by precise structuring at the nano- or meso-scale. Advanced and sophisticated approaches are therefore essential to engineering these properties. Among those, a particularly fascinating one is the inverse self-assembly that combines the goals of nanostructure design with the principles of self-organisation. When the interactions between particles exceed the strength of thermal fluctuations, the system typically becomes unstable. To minimise the free energy, particles spontaneously form bonds, condensing into either liquid or crystalline phases. This process by which particles organise themselves into equilibrium structures is known as self-assembly. But what if we can control the assembly of a structure? This is precisely the aim of inverse self-assembly: determine how to design building blocks that spontaneously aggregate into a desired structure. Unlike conventional chemical synthesis, inverse self-assembly deals not just with atoms and molecules, but with building blocks at the nano- or meso-scale. These units can be engineered in terms of their chemical, physical, and geometric properties to drive specific behaviours to self-assembled structures. Because the interactions involved are not fixed by quantum-mechanics, they can be intentionally designed. These are weak, non-covalent interactions that allow the structures to form without external guidance, relying on the spontaneous formation and breakage of bonds. This results in thermodynamically stable assemblies. The reversible, rather than irreversible, nature of the interactions underlies the reconfigurability of the assembled structures, something rarely observed in materials created through traditional fabrication methods. Moreover, because the assembly process is highly sensitive to external conditions, this technique offers a remarkable degree of tunability. By adjusting environmental parameters, it is possible to guide the formation of different structures and to dynamically influence the properties of the assembled material. The structures can reversibly switch between distinct configurations of functional states, enabling responsive or adaptive behaviour. This expands the design space even further, opening new pathways for materials that can be reconfigured or



controlled through external stimuli.

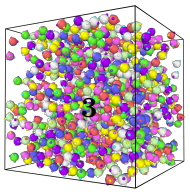
Inverse self-assembly approaches have traditionally focused on tuning two key aspects of the building blocks to achieve selective assembly: their shape and their pair interaction potential. This strategy has led to increasingly complex designs of single-component systems since achieving selectivity requires not only promoting the spontaneous assembly of the desired structure, but also suppressing the formation of competing structures. In contrast, we propose a shift in focus: instead of designing a single particle type with complex shape and interactions, we **mix** simpler but distinct building blocks. Specifically, we employ patchy particles as elementary units. These particles have a fixed spherical shape, while the required anisotropy of interactions is introduced through patches, attractive sites located on their surface. The number and the spatial arrangement of these patches are chosen to **match** the bond geometry of the target structure. Different types of building blocks differ in the type of patches they have, with each patch type encoding specific binding ability. These patch types and interaction rules are what drive the mixture to **assemble** the target structure, and must be designed accordingly. To this end we adopt an approach named SAT-assembly, an **inverse design approach** that determines how different patchy particles must be defined and how they should interact (*i.e.* what are the matchable patches) in order for the target structure to emerge as a free energy minimum. This algorithm iteratively refines the design, providing a systematic route to inverse **tailor** the building units for self-assembly into complex and programmable **materials**.

Crucially, successful and efficient inverse self-assembly requires controlling not only the target structure, but also the intermediate steps leading to its formation. This aspect, often overlooked, is the central focus of this thesis. Our goal is to identify optimal self-assembly pathways and to extend the SAT-assembly framework to include this level of control. To this end, it is fundamentally important to understand how nucleation occurs in mixtures. In particular, we focus on how density fluctuations couple with composition fluctuations, a phenomenon absent in single-component systems. The phase behaviour of mixtures is characterised by a phase separation process where the species concentration of the resulting coexisting phases typically deviates from the one of the starting homogeneous system. In our case, the coexisting vapour and liquid phases reach equilibrium at different concentrations of the constituent species. Without explicitly computing the full phase diagram, it is generally not possible to predict the final equilibrium concentrations from a given initial condition. There is, however, a notable exception: azeotropic mixtures. At the azeotropic point, the mixture separates into phases that retain the initial species ratio, effectively behaving as a single-component system. We explore the possibility

of inverse designing this specific thermodynamic property by tailoring the interactions between particles. Our aim is to identify general rules that inter-particle bonds need to satisfy, both in terms of allowed bonding configurations and their energies, in order to generate azeotropic mixtures. Furthermore, we investigate how this inverse thermodynamic approach to tailor phase behaviour can be integrated into the SAT-assembly framework. This capability is crucial in the context of self-assembly. Molecular dynamics simulations indicate that nucleation typically occurs within the denser phase following phase separation. Hence, for efficient and high-yield self-assembly, it is critical that this dense phase has the same stoichiometric ratio as the target structure. By designing the azeotrope at the same stoichiometric ratio of the crystal, we can optimise the self-assembly process.

The design rules for azeotropy are identified within the Wertheim first order perturbation theory and validated by computing the vapour-liquid coexistence regions using both numerical integration techniques and Gibbs ensemble simulations. To assess the impact of the azeotropic point on self-assembly, we correlate these phase diagrams with the nucleation behaviour by performing extensive direct nucleation simulations in the canonical ensemble. These simulations take advantage of the use of biased Monte Carlo moves to efficiently capture rare nucleation events. In addition, we employ advanced sampling techniques to compute surface tensions and free energy barriers to investigate the validity of Classical Nucleation Theory (CNT). Our results reveal a significant deviation from CNT predictions: polymorphs that share the same bulk and interfacial free energies exhibit different nucleation rates. This finding challenges the capillarity approximation that underlines CNT. It is the assumption that macroscopic thermodynamic properties alone can predict nucleation rates, and provides insights into nucleation mechanisms in presence of polymorphs. The polymorphs we leverage in our test differ not in crystal structure, but in the arrangement of species on equivalent lattice sites, a consequence of the system's multi-component nature. These subtle arrangement differences lead to variation not only in the nucleation frequencies, but also in the optical properties of the structures. This instance motivates us to explore how constraining specific species to occupy specific lattice sites can be used to impart chirality to the final assembled structure.

While this work is primarily theoretical and computational, it is strongly motivated by the possibility to experimentally realise the designed assemblies. This is why in investigating the thermodynamics and the nucleation process of mixtures, we focus on a case study: a SAT-designed binary mixture able to exclusively self-assemble into a cubic diamond crystal without defect and stacking faults. This crystal lattice is extremely promising for nanotechnology applications as, if realised on colloidal scales, it acts as a photonic crystal capable of manipulating radiation in the visible or



infrared wavelengths. For experimental realisation, the idea is to use DNA-origami whose patchy particles are an ideal coarse-grained model. Our goal is to establish a general framework that support the colloidal realisation of target structures. We believe this work contributes to that goal by extending the SAT-assembly design ability beyond topological requirements alone. Indeed, beyond the SAT-assembly's core ability to automatically identify particle designs that self-assemble exclusively into a target structure while avoiding alternative assemblies, we introduce the capacity to hold additional requirements acting also on thermodynamic and photonic properties, such as the presence of an azeotrope or of chiral structures. This expanded framework offers a powerful tool set for material design, helping to bridge the gap between theoretical models and experimental realisation: SAT-assembly allows us to approach experiments with the optimal and tailored design for a desired structure.

The thesis is structured as follows. Chapter 1 provides a quick overview of self-assembly, specifying the difference between direct and inverse approaches and highlighting the importance of tunable interactions in soft matter. Chapter 2 presents the SAT-assembly framework. We introduce the combinatorial problem that arises when working with mixtures, and show how the design problem can be reformulated, firstly as a colouring task, and then as a Boolean satisfiability problem. To illustrate the method, we focus on the meaningful example of the cubic diamond assembly, introducing the SAT-designed N2c8 binary mixture. Chapter 3 bridges the gap between theory, simulations and experiments. We describe how patchy particles are simulated, and how they can be realised with DNA origami, providing an algorithm to identify possible pairs of single DNA strands satisfying our bond constraints. In Chapter 4, we explore numerical and simulation techniques useful for constructing phase diagrams of mixtures. We apply them to the N2c8 solution, and we underline the thermodynamic differences between multi-component systems with respect to single-component ones. Chapter 5 is devoted to investigating the nucleation behaviour in mixtures. We explore the role of density and composition fluctuations and the relevance of specific thermodynamic conditions, such as the critical and the azeotropic points. In Chapter 6, we present our falsifiability test for CNT, demonstrating the failure of the capillarity approximation on which CNT is based. Chapter 7 is the chapter of inverse thermodynamic, where we define design rules to tailor the phase diagrams, focusing on the appearance of azeotropic points at desired concentrations. Finally, Chapter 8 investigates how assigning designated patchy particle species to specific lattice positions can induce chirality in the assembled crystal structures.

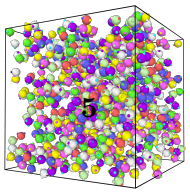
The results presented in this thesis have been published in the following peer-reviewed papers:

1. Beneduce, C., EP Pinto, D., Šulc, P., Sciortino, F., and Russo, J., "Two-step nucleation in a binary mixture of patchy particles", *The Journal of Chemical Physics*, 158(15), 154502, 2023, `https://doi.org/10.1063/5.0140847`, which is primarily discussed in Chapter 5.

2. Beneduce, C., Sciortino, F., Šulc, P., and Russo, J., "Engineering azeotropy to optimize the self-assembly of colloidal mixtures", *ACS nano*, 17(24), 24841-24853, 2023, `https://doi.org/10.1021/acsnano.3c05569`. Chapter 5 and Chapter 7 are the main chapters presenting this work.

3. Beneduce, C., Pinto, D. E., Rovigatti, L., Romano, F., Šulc, P., Sciortino, F., and Russo, J., "Falsifiability test for classical nucleation theory", *Physical Review Letters*, 134(14), 148201, (2025), `https://doi.org/10.1103/PhysRevLett.134.148201`, which is analysed in detail in Chapter 6.

4. Beneduce, C., Mastriani, G., Šulc, P., Sciortino, F., and Russo, J., "Inverse Thermodynamics: Designing Interactions for Targeted Phase Behavior", *The Journal of Physical Chemistry B*, 129(39), 10153–10161, 2025, `https://doi.org/10.1021/acs.jpcb.5c04056`, which is the focus of Chapter 7 .



Chapter 1

Self-assembly

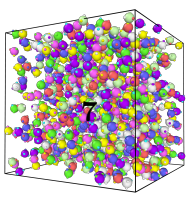
This chapter provides a concise introduction to self-assembly and explores how soft matter systems can be engineered to replicate and control this phenomenon.

1.1 Direct and inverse self-assembly

Self-assembly refers to the spontaneous organisation of single components into defined structures without external guidance. This phenomenon underpins many natural processes such as protein folding, virus capsid formation, liquid bi-layer organisation in cell membranes, shell and coral biomineralisation, as shown in Figure 1.1. The underlying mechanism is increasingly being harnessed in materials science to build complex structures with tailored functionalities.

Association processes occur when particles within a system bind, driven by a balance between the energetic gain from bond formation and the entropic penalty associated with the resulting loss of translational entropy. At high temperatures, the system is in a fluid phase: aggregation does not occur because the thermodynamic behaviour is dominated by entropy, which is maximised when particles remain unbonded. Only when interactions between particles have strength comparable or larger than the thermal energy, the fluid becomes unstable and the particles condense searching for a lower free energy state. The spontaneous formation of inter-particle bonds gives rise to aggregates whose final state can be either that of an ordered lattice or a quasi crystal [4], a connected percolating structure (*e.g.* a liquid), or a collection of finite size clusters, as illustrated in Figure 1.2. This spontaneous search for the lowest free energy state is called *self-assembly* [5, 6]. While the computation of the free energy of a structure is a laborious but solved problem in statistical mechanics, several challenges hamper our understanding of self-assembly and our ability to mimic natural systems. Efforts focused on both direct and inverse self-assembly.

In the *direct* self-assembly problem, one starts from a set of predetermined elementary



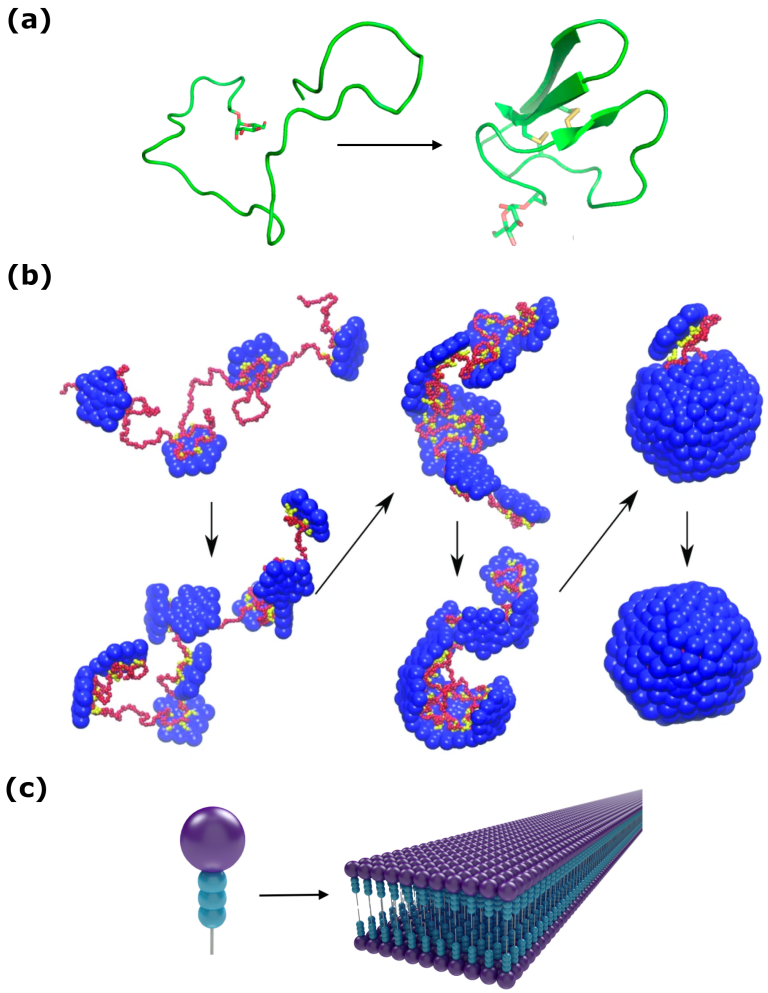


Figure 1.1. Self-assembly in nature. Examples of different spontaneous aggregation processes occurring in biological systems: (a) protein folding; (b) virus capsid assembly (c) amphiphilic phospholipid organisation into a bilayer. Images adapted from [1–3], respectively.

units with known inter-particle interactions and is tasked with selecting structures that correspond to free energy minima. This is done either with intuition (for simple structures), with brute force approaches (direct molecular simulations), or with specialised algorithms [7, 8].

An even more interesting and challenging task is the *inverse* self-assembly problem, which goes beyond merely understanding how self-assembly occurs, and seeks to exert precise control over the process. The goal is to design elementary units that reliably self-assemble exclusively into a desired structure [12, 13]. As Richard Feynman famously pondered, "What would happen if we could arrange the atoms one by one

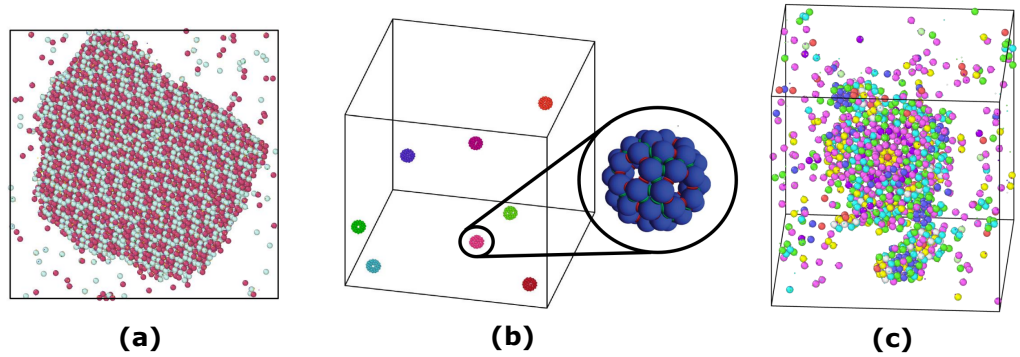
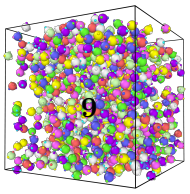


Figure 1.2. Target structures. Examples of different types of assemblies: (a) cubic diamond crystal; (b) snub-dodecahedron cages; (c) quasi crystal. Images adapted from [9–11], respectively.

the way we want them?", achieving a similar level of control over colloidal units promises to unlock new frontiers in material science [14]. This thesis is centred on the problem of inverse self-assembly, which involves addressing two main challenges: first, the identification and design of suitable building blocks which are meso- nano-scale units and not atoms or molecules; second, ensuring that no competing structures can form that would prevent the assembly of the desired one [15]. Indeed, kinetic pathways can drive the assembly towards amorphous aggregates, or structures different from the desired one. The elementary units can be designed by tailoring either their shape or their interaction potential. Both strategies have been explored, and we will briefly describe them in the following section.

1.2 Soft Matter: the world of tunability

A key point one is tasked in the inverse self-assembly problem is the design of the building units. Various strategies have been proposed, and these include geometrical approaches and optimisation algorithms. In the first case, some properties of the building units, such as the shape and the directionality of the bonds, are tuned in order to match the geometric features of the target structure [16–23]. Optimisation algorithms, instead act on the inter-particle potential to establish a pair potential whose free-energy minima is guaranteed to be the desired structure [13, 24–32]. However, the inter-particle interactions resulting from such procedures are often too complex and require a degree of precision that is out of reach in experiments. In contrast, geometrical approaches are experimentally feasible, but demand high degree of geometrical intuition. Moreover, both strategies are strongly system specific, typically rely on *ad-hoc* engineer to make the desired structure the system's free energy minimum and get rapidly complex. In this work, we aim to establish a



general and automated framework capable of systematically addressing the inverse self-assembly problem.

Before introducing our approach, it is important to emphasise that the ability to tune both the geometrical and interaction properties of the building blocks is a unique feature of soft matter systems. Owing to their colloidal dimensions (ranging from nm to μm) these systems are soft, meaning they are easily deformable. Conversely to atomic and molecular systems where interactions are fixed by quantum mechanics and the ability to engineer specific behaviours is severely limited, soft matter systems exist at a mesoscopic scale that grants them a remarkable degree of tunability. They are bigger with respect to atomic and molecular systems to be easily deformable, yet still small to be affected by thermal fluctuations and accurately described by statistical mechanics. This intrinsic tunability allows soft matter systems to exhibit rich phase behaviour, self-organisation, and responsiveness, making them ideal platforms for studying emergent phenomena and for designing materials with programmable structure and function, as we will show in this work.

In summary, our goal is to explore a novel approach to material fabrication that leverages the inverse-self assembly of soft matter systems. Specifically, we aim to establish a general procedure for designing building units at nano- meso-scale that spontaneously and selectively aggregate into a predetermined structure. This framework, referred to as SAT-assembly, is introduced in the next chapter.

Chapter 2

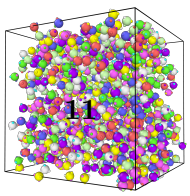
SAT-assembly

In this chapter, we review the SAT-assembly approach to address the inverse self-assembly problem. This framework combines the two routes so far exploited, the geometrical strategies and the optimisation algorithms, and it provides a general and automatic design procedure. The main novelty of this framework is the extension of the number of building blocks. By moving from single-component systems to mixtures, simpler interacting potential can be used, shifting the challenge to introduce specificity through selective bonding between components.

2.1 Patchy Particles

The building units we employ are patchy particles. They are spherical hard core colloids of diameter σ whose surface is decorated by attractive sites named patches. As shown in Figure 2.1, a bond is formed when patchy particles are sufficiently close and properly oriented such that two compatible patches come into proximity. While it is rather uncommon, we note that we can also generalise this definition to non-spherical and poly-disperse systems.
Different species of patchy particles can differ either by the number, the arrangement, and the type of the patches. The bonding rules, *i.e.* which types of patches can establish a bond are encoded in a square matrix, termed the interaction matrix Γ, with dimensions equal to the total number of patches. For example, in a ternary mixture where each species has five patches, Γ is a 15×15 matrix. The first five rows and columns correspond to the patches of the first species, rows and columns from 6 to 10 to those of the second, and the last five to those of the third species. Each element $\Gamma_{(\alpha,i),(\beta,j)}$ is set to 1 if patch α on particle i interacts with patch β on particle j, and to 0 otherwise.

Mathematically, the short-ranged and anisotropic inter-particle interaction between patchy particles, can be modelled by the Kern-Frenkel potential [34,35]. The excluded



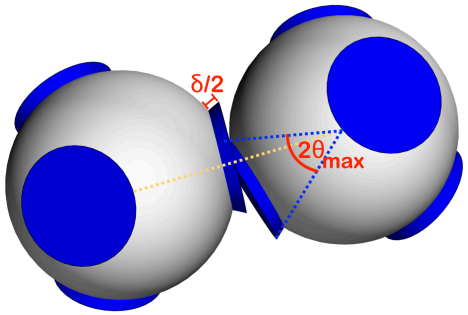


Figure 2.1. Schematisation of patchy particles. Two bonded patchy particles with four equal patches coloured in blue and tetrahedrally arranged. The pair interaction can be modelled by the Kern-Frenkel potential which is based on two key parameters: the depth δ and the wideness θ_{max} of the patches. Image reproduced from [33].

volume interaction is included as an isotropic hard-sphere repulsion embedded within a square well potential V_{SW} of depth ϵ and width δ that provides an attractive range extending from σ to $\sigma + \delta$. To introduce the directional nature of patchy particle interactions, this isotropic potential is modulated by a term F depending on the relative orientation of the particles. Specifically, F is a function of the scalar product between $\mathbf{r}_{i,j}$, the centre-to-centre distance between particles i and j, and $\hat{\mathbf{r}}_{\alpha,i}$ and $\hat{\mathbf{r}}_{\beta,j}$, the unit vectors that define the position of patch α on species i and β on species j, respectively. More precisely, the total interaction potential V is:

$$V(\mathbf{r}_{ij}, \hat{\mathbf{r}}_{\alpha,i}, \hat{\mathbf{r}}_{\beta,j}) = V_{SW}(r_{ij}) F(\mathbf{r}_{ij}, \hat{\mathbf{r}}_{\alpha,i}, \hat{\mathbf{r}}_{\beta,j}) \tag{2.1}$$

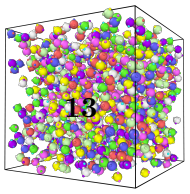
where F is defined as

$$F(\mathbf{r}_{ij}, \hat{\mathbf{r}}_{\alpha,i}, \hat{\mathbf{r}}_{\beta,j}) = \begin{cases} \Gamma_{(\alpha,i)\,(\beta,j)} & \text{if} \quad \hat{\mathbf{r}}_{ij} \cdot \hat{\mathbf{r}}_{\alpha,i} > \cos(\theta_{max}) \\ & \quad\;\; \hat{\mathbf{r}}_{ji} \cdot \hat{\mathbf{r}}_{\beta,j} > \cos(\theta_{max}) \\ 0 & \text{otherwise} \end{cases} \tag{2.2}$$

In summary, at distances shorter than σ, patchy particles i and j interact via isotropic hard sphere repulsion that prevents overlap. In contrast, at intermediate distances, in between σ and $\sigma + \delta$, an attraction between patch α on particle i and patch β on particle j occurs if: (a) the two patches are compatible, that is if $\Gamma_{(\alpha,i)(\beta,j)} = 1$, and (b) they are correctly aligned, *i.e.* the line connecting the centres of the two particles (dotted yellow line in Fig 2.1) intersects the volume of both patches. We notice that $\Gamma_{(\alpha,i)(\alpha,i)} = 1$, *i.e.* a diagonal element of the interaction matrix equal to one, indicates a self-compatible patch. In this case, a bond can form between two identical patches located on two distinct particles of the same species. We notice that the Kern-Frenkel potential is suitable for Monte Carlo simulations,

but must be converted to a differentiable form if Molecular Dynamics simulations are performed. Usually, the hard sphere like repulsion is translated into a Weeks-Chandler-Anderson potential and the overall square well potential into a Lennard-Jones potential truncated with a quadratic smoothing function. At the same time, the dependence on the relative orientation between the particles is given by Gaussian like functions. Importantly, the Kern-Frenkel potential and its continuous versions, are characterised by two independent geometrical parameters δ and θ_{max} that specify the range and the angular width of the patches, respectively (see Figure 2.1). By tuning these values, different thermodynamic and kinetic behaviours are observed. In particular, phase diagram computations varying θ_{max} [36, 37] reveal that both excessively large and overly narrow angular widths can hinder the self-assembly process. When θ_{max} is too large, the nucleation rate is significantly suppressed. This is because the free energy difference between the disordered and ordered phases becomes small, reducing the thermodynamic driving force for nucleation. As a result, even under slow cooling, nucleation may not occur within the accessible observation timescales, and the system may eventually end up in a glass state. Conversely, when θ_{max} is too small, the bonding volume becomes extremely small, leading to a substantial entropic penalty that makes bond formation thermodynamically unfavourable. Moreover, narrow patches can kinetically favour the closure of rings of specific size which can hamper the growth of the desired structure and drive the assembly towards competing aggregates, even when they are metastable with respect to the target one. Narrowness also affects the kinetics of the assembly: narrower patches slow down the process since particle encounters are less likely to result in successful binding. This analysis further highlights the high degree of tunability and flexibility inherent to inter-particle interactions in soft matter systems. In the following, we fix the Kern-Frenkel potential parameters to $\cos(\theta_{max}) = 0.98$ and $\delta = 0.2$, values known to readily access nucleation within simulation times. All energies are expressed in units of the square-well depth (ϵ), and distances in units of the patchy particle diameter (σ).

It is apparent that the Kern-Frenkel potential in Eq. 2.1 is significantly simpler than the *ad-hoc* interaction potentials typically required for controlling self-assembly in single-component systems. However, to reliably aggregate only the desired structure using patchy particles, it becomes necessary to consider multi-component systems. Then, the inverse self-assembly complexity shifts from designing a precise functional form of the interaction potential to determining number, positions, and types of patches of each particle, along with the interaction matrix guaranteeing the formation of the desired assembly. This shift in focus introduces a new combinatorial challenge, which we address by introducing the SAT-assembly framework.



2.2 A satisfiability approach to a design problem

The inverse self-assembly goal is to design building units and their interactions to spontaneously self-assemble, without defects, a desired structure. It is a challenging task since the formation of specific structures is not devoid of problems [12, 13]. First of all, even if the target structure is the stable state, kinetic pathways can drive the assembly towards amorphous aggregates, or structures different from the desired one. To address this hurdle, several strategies have been suggested. They focus on single-component systems that are easier to handle from a thermodynamic perspective, and examples include: optimisation of preparation protocols [32, 38, 39], accurate design of inter-particle potentials [13, 24–31], and geometrical approaches where the inter-particle interactions accurately match the geometric features of the target structure [16–23]. However, these strategies are often not experimentally feasible since complex interaction potentials [13] and/or shapes tuning [19] are required to avoid the appearance of alternative structures. Rather than increasing the complexity of the inter-particle interactions, another route is to increase the number of building blocks. Abandoning the one-component approach in favour of multi-component systems shifts the challenge from designing complex single particle potentials to optimising simpler and more geometrical interactions between multiple components [15, 40–45]. We can use less complicated potentials because, expanding the alphabet of building blocks, *i.e.* increasing the number of components, reduces the degree of symmetry in the final structure. This, in turns, decreases the number of competing structures, and simplifies the assembly pathway towards the target design. However, aside from experimental challenges, mixtures introduce a significant drawback compared to single-component systems: a combinatorial problem. As the number of building units grows, so does the existing interaction matrices. Each additional component exponentially increases the space of possible solutions, and with that the computational effort required to find a good design. To tackle this, an approach called SAT-assembly has recently been introduced [41, 46].

SAT-assembly integrates concepts from both geometrical approaches and optimisation algorithms while offering a general and automatic procedure for designing self-assembling systems. Building units are patchy particles and the geometric information is encoded in the number and the placement of the patches selected to match the bond geometry of the target structure. For example, patchy particles with four tetrahedral arranged patches are used to assemble a cubic diamond crystal. The central goal is to design multiple species of patchy particles and their interactions such that only the target structure corresponds to the system's free energy minimum, an objective aligned with traditional optimisation methods. However, rather than engineering complex interaction potentials for a single-component

system, SAT-assembly focuses on designing multi-component systems with simpler pairwise interactions. This is achieved through a colour programming strategy. Each patch type is assigned a specific colour and selectivity towards the target structure is obtained by introducing compatible colours: bonding is allowed only between complementary colours. The challenge then becomes to determine whether a set of particle species, patch types and colour complementarity exist such that the target structure is a fully bonded minimum-free energy state while also ensuring that competing structures cannot form using the same components or any of their subsets. Figure 2.2 presents a simple example to illustrate this colouring approach for the inverse self-assembly task. The purpose is to design patchy particles that can form six-membered rings while preventing the closure of five-membered rings. The simpler solution uses patchy particles with two patches. The key design question becomes: how should the patches be coloured to meet this requirement? An alternating colouring scheme provides a possible solution. Because no particle has two identical patches, odd-membered rings such as pentagons cannot close, while even-membered

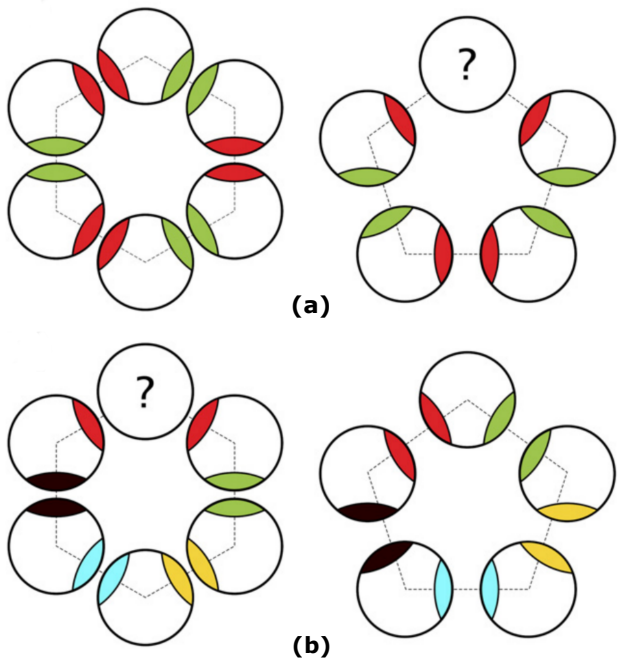
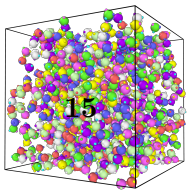


Figure 2.2. Illustration of the colouring problem. The SAT-assembly framework recasts the design problem into a colouring problem that for the simpler situations can be solved manually. (a) In order to close six-membered rings while forbidding the formation of five-membered rings, we can use a single species of bifunctional patchy particles with two different coloured patches. (b) For the inverse task five species and five colours are required. In this case, patches cannot be self-complementary. Hence, in general a given colour does not represent a single patch type but rather a pair of distinct, mutually compatible patch types. As a result, the total number of different patch types is ten. Image reproduced from [41].



rings like hexagons can form. Thus, using just one particle species and two different patch types is sufficient. Interestingly, solving the inverse problem, *i.e.* designing particles that can form five-membered rings while excluding six-membered ones, requires a far more complex design. In this case, five species and five colours are necessary. How the correspondence between colours and patches is defined? In the first case, it can be a one-to-one correspondence where a single colour represent a single patch type. Then the design in (a) consists of two different self-interacting patches (the green and the red). This is not the case in the second example where each of the five colours represents a pair of complementary but distinct patch types. None of these colours can stand for self-complementary patches since each vertex of the pentagon must be occupied by a different species. In the following we use the terms colours and patch types interchangeably, while in figures, to make them quicker and easier to interpret, we often represent two interacting patches using the same colour, even if their types differ.

Although the simple example in Figure 2.2 can be solved manually, this procedure is rarely feasible in more complex scenarios. As the number of different species and patch types increases, structural competition can be more effectively suppressed, but at the cost of expanding the solution space, making manual design impractical. To handle this complexity, the colouring problem is translated into a boolean satisfiability problem (SAT), which can be efficiently solved using SAT solvers. Specifically, SAT-assembly, encodes the bond topology of the target structure into a system of Boolean equations whose solution gives the interaction matrix between different patches. The sophistication of modern SAT solvers [47] allows to effectively tackle the combinatorial problem for complex assemblies, including open crystalline structures, clathrate structures, and finite size aggregates. In the next two sections the SAT-assembly methodology is deeply outlined.

2.2.1 Variables and clauses

In this section we explain how, when the building blocks are patchy colloids, the process of designing the bonding rules to assemble one specific structure (and selectively avoid the formation of any other competing assembly) can be turned into a set of boolean satisfiability (SAT) equations, which can be solved using modern SAT solver algorithms. This strategy, named SAT-assembly, was shown to overcome the combinatorial challenge of manage all the possible solutions for multi-component systems by leading to the successful assembly of several complex target structures, including the coveted colloidal diamond crystal [9, 41, 46]. The SAT-assembly methodology is summarised in Figure 2.3.

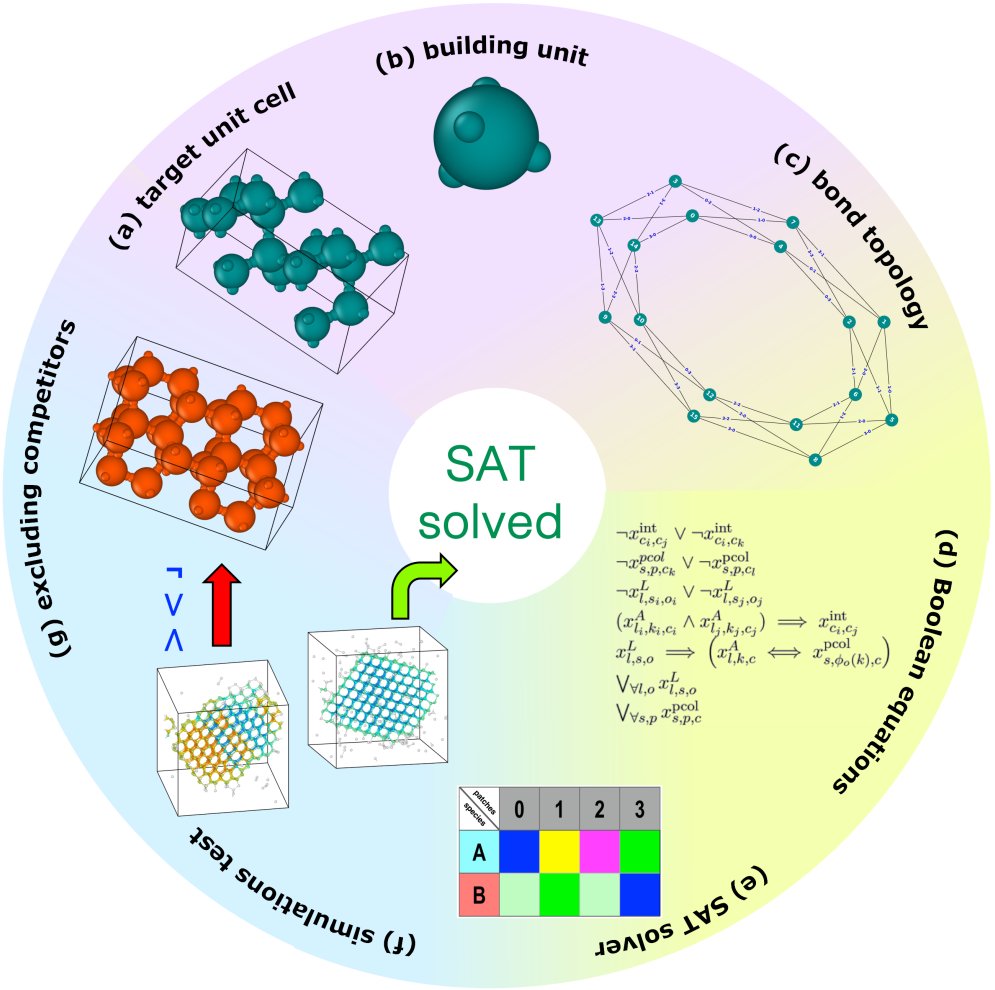
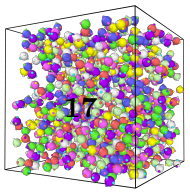


Figure 2.3. Overview of SAT-assembly framework. (a) Definition of the target unit cell, here a 16-particle cubic diamond. (b) Selection of the building units: the valence of the patchy particles and the arrangements of the patches are fixed according to the bond geometry of the target structure. (c) Determination of the bond topology: the neighbours of each particle in the target cell and the specific patches through which bonding occurs are identified. (d) Definition of the boolean variables and clauses: a set of equations is defined for a given number of lattice positions, species, patches, orientations and different patch types. (e) Running of SAT-solvers: if a set of values for the boolean variables making simultaneously true all the clauses exists, the interaction matrix is determined. (f) Design test via molecular dynamics simulations: if trajectories successfully assemble the desired structure the inverse self-assembly problem is solved. If, instead, undesired structures appear, the algorithm goes through points (g), (e) and (f) iteratively. It stops when a good solution is found in (f) or when all possible interaction matrices are not exclusively compatible with the target structure. In this last case, a restart from point (c) by changing the number of patch types or/and species can be attempted. (g) Insertion of additional clauses to avoid competing structures: for each already known, or spotted in (f) undesired topology (here a 32-particle hexagonal diamond), point (e) is executed. If a solution is possible, before running (e) for the target structure, an additional clause preventing the interaction matrix to be the one just found is added.



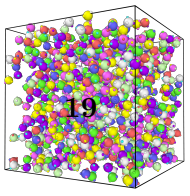
Firstly, the number and the placement of the patches are chosen to match the bond geometry of the target structure. Then, the topology of the target structure unit cell is introduced. It defines *i.e.* where each particle is located, what its neighbours are and through which patches bonds are established. What is left to determine is how many different types of species and patches are needed and how they should interact in such a way that the target structure is the minimum free energy fully-bonded state. This task is converted into a colouring problem: by deciding that different colours encode different binding properties, the goal is to define how each patch of each species is coloured and which colours are complementary allowing the formation of a bond. Since it is impossible to proceed by looking directly for all possible combinations, the colouring problem is transformed into a Boolean satisfiability one (SAT) that can be solved efficiently through SAT solvers [47]. The mapping is done by defining binary variables and binary clauses. The first ones encode the design problem and the second ones impose constraints that variables need to satisfy. A solution to the design problem is found if it exists a set of Boolean variable values ensuring all clauses are simultaneously satisfied. However, the SAT solver guarantees the solution to be compatible with the target structure, but does not prevent particles to arrange themselves into some other competing structure. For this reason, an appropriate design is determined by checking the solution against competitors. An iterative approach is used: each time a solution satisfies also alternative topologies, it is excluded by adding a new clause that prevents the interaction matrix and each patch colouring to be the same as the one just found. If competing structures are already known it is not necessary to test each solution directly, by performing numerical simulations and verifying the nucleation outcomes. Indeed, SAT-assembly can be formulated to check if there is a lattice placement for all or a subset of species that is compatible with an undesired topology, by keeping the interaction matrix and the patch colouring fixed to the one of the solution we want to test. If the SAT solver finds a set of unit cell placement variables, the solution tested must be discarded. Whereas, if the SAT solver gives no solution, we found a design that avoids that particular competing structure.

Given the overall flow of SAT-assembly, we focus on its core aspect, *i.e.* expressing the colouring problem as a set of Boolean equations. The desired topology is given by the position and the orientation of each particle in the lattice. To any lattice position L, V slots representing the possible bonds are assigned. Hence, to each patchy particle we attribute a position l among the L available ones and an orientation o among the R possible ways to orient the particle in such a way that its patches overlap with the slots. Each orientation o defines a mapping ϕ_o from slots to patches, such that $\phi_o(k) = p$, meaning that slot k is occupied by patch p when the particle

is in orientation o. The goal is to determine if it exists a set of patchy particles satisfying the given topology. Is there a design that allows to fill the desired lattice positions such that the neighbouring patches have complementary colours? The different N_c colours encode for different binding properties. Each patchy particle has the same number of patches N_p and the same placement of patches on its surface, but we allow having N_s different species of patchy particles distinguishing them by the colours that their patches have. N_s and N_c are the parameters of the algorithm. Starting from the lowest number of different species and colours we progressively increase them looking for a solution compatible with our target structure while avoiding competing assemblies that are iteratively excluded by means of additional clauses. An alternative approach that avoids reformulating the SAT-assembly problem repeatedly is to vary only the parameters N_s and N_c, rather than systematically excluding previously insufficient designs. While this method can be faster, it likely overlooks the minimal solution, *i.e.* the one that uses the fewest species and colours, is skipped. This is a drawback as minimal designs are generally more practical and cost-effective to implement experimentally. If minimal solutions are also the most efficient from a yield and kinetic point of view, will be the object of further studies. The opposite limit occurs when all patches are different, and we have a fully-addressable design that uses the maximum number of colours (N_s times N_p).
The design is enclosed in the following binary variables that can be either true (equal to 1) or false (equal to 0).

- colour interaction variables $x^{int}_{c_i,c_j}$ defining the way patchy particles interact. There are $\frac{(N_c)(N_c+1)}{2}$ of such variables and if true, colour c_i is complementary to colour c_j;
- patch colouring variables $x^{pcol}_{s,p,c}$ defining the colour assigned to each patch of each species. There are $N_s V N_c$ of such variables and if true, the patch p of particle of species s has colour c;
- unit cell placement variables $x^{L}_{l,s,o}$ defining the occupancy of the desired lattice. There are $N_s L R$ of such variables and if true, the position l is occupied by a patchy particle of species s with orientation o;
- auxiliary variables $x^{A}_{l,k,c}$ helping in formulating the design problem. There are $V L N_c$ of such variables and if true, the particle located at position l is oriented in such a way that the slot k has a patch of colour c.

The correct design rules are established through Boolean clauses that, by coupling the binary variables, constrain their values. The clauses we use are listed below.



- $C^{int}_{c_i,c_j,c_k} = \neg x^{int}_{c_i,c_j} \vee \neg x^{int}_{c_i,c_k}$ ensuring that each colour has a single complementary colour;

- $C^{pcol}_{s,p,c_k,c_l} = \neg x^{pcol}_{s,p,c_k} \vee \neg x^{pcol}_{s,p,c_l}$ imposing that to each patch of each species is assigned only one colour;

- $C^{L}_{l,s_i,o_i,s_j,o_j} = \neg x^{L}_{l,s_i,o_i} \vee \neg x^{L}_{l,s_j,o_j}$ guaranteeing that each lattice position is occupied by a only one particle of specific species and orientation;

- $C^{lint}_{l_i,k_i,l_j,k_j,c_i,c_j} = \neg x^{A}_{l_i,k_i,c_i} \vee \neg x^{A}_{l_j,k_j,c_j} \vee x^{int}_{c_i,c_j}$ assuring neighbour patchy particles in the lattice interact with patches of colour that respect the interaction matrix;

- $C^{LS}_{l,s,o,c,k} = (\neg x^{L}_{l,s,o} \vee \neg x^{A}_{l,k,c} \vee x^{pcol}_{s,\phi_o(k),c}) \wedge (\neg x^{L}_{l,s,o} \vee \neg x^{A}_{l,p,c} \vee x^{pcol}_{s,\phi_o(k),c})$ ensuring that the colour assigned to each patch is compatible with the orientation assigned to the particle;

- $C^{alls.} = \bigvee_{\forall l,o} x^{L}_{l,s,o}$ preventing that only a subset of species is used;

- $C^{allc.} = \bigvee_{\forall s,p} x^{pcol}_{s,p,c}$ avoiding that only a subset of colour is used.

Importantly, in SAT clauses, the variables or their negation (obtained with the NOT ($\neg$) logical operator) must be connected only through OR ($\vee$) operators since AND ($\wedge$) operators are used to link together the different clauses. Indeed, the design problem has a solution if there is an assignment of truth values (true or false) to the variables so that every clause is satisfied (is true). Technically speaking, when this condition is met, the Boolean problem is said to be satisfiable. Finally, we note that the last two clauses $C^{alls.}$ and $C^{allc.}$ must be removed when checking a particular solution against a competing structure to take into account that it can arise from the self-assembly of a subset of patchy particles. Further discussion of Boolean clauses, aimed at supporting non-expert readers, is provided in Appendix A.

The SAT-assembly framework is an effective, general, and versatile approach for determining designs that lead to desired assemblies without defects and stacking faults. Notably, by combining it with the patchy particles model, we outline a design strategy that is not system specific. Moreover, we emphasise that additional clauses can always be incorporated to enforce specific properties, allowing for precise tailoring of the design. This ability for customisation is the central theme of the thesis.

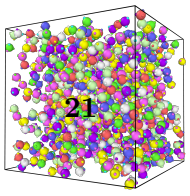
Once the Boolean satisfiability problem is defined, the seeking of determining if it is satisfiable or unsatisfiable is committed to SAT-solvers. The following section is devoted to these powerful solvers.

2.2.2 Sat-solvers

The colouring problem defined above falls into the class of NP complete problems, problems for which no polynomial time algorithm is currently known. NP problems are a class of decision problems where each instance has a "yes" or "no" answer. While a proposed solution can be verified in polynomial time, it is not known whether a solution can be found in polynomial time. In practice, solving these problems typically involves algorithms that systematically explore the space of possible solutions in an intelligent way, *i.e.* by using constraint-solving strategies, rather than performing a brute-force search, in order to find a variable assignment that satisfies all constraints or to determine that none exists. This is precisely the strategy employed in our SAT-assembly framework.

Although challenging, the development of modern SAT-solvers allows us to determine a solution to problems of our size in times ranging from few second to few hours. Then, it is worth nothing that different solvers perform differently, and the best choice depend on the specific design problem. While deepening into the details of each solver is out of the scope of this thesis, where only the MiniSAT [47] solver is used, we believe it is worthwhile to briefly outline the common underlying mechanisms shared by different solvers.

The Boolean formula to determine if a problem is satisfiable is represented in Conjunctive Normal Form (CNF): it is a conjunction (AND $\wedge$ operators) of clauses, where each clause is a disjunction (OR $\vee$ operators) of variables or of their negations (NOT $\neg$ operators). The first variables assigned with a value are the ones entering the so called unit-clauses, clauses where only one variable appear. Assigning each of these variables with the value making true the clauses, can generate new unit-clauses that are readily considered. When there are no unit-clauses left, the solver makes a decision by picking a variable and assign it a value. It continues until a conflict is detected, that is when a clause is false due to the values assigned to the variables that define it. In this case, the solver backtracks to the most recent decision point and try the opposite assignment. Additionally, it learns a new clause that prevents the same conflict to happens again. For each detected conflict a new clause is inserted. This procedure is named CDCL (Conflict Driven Clause Learning). In particular, when the algorithm hits a conflict, it constructs an implication graph that illustrates all the assignments that led to the conflict, starting from the current decision level. Within this graph, the first UIP (unique implication point) is identified. This point corresponds to the variable assignment closest to the conflict that lies on all the paths from the decision assignment to the conflict. Finally, to avoid repeating the same conflict, the solver learns a new clause that is added to the Boolean formula. This clause prevents the variable involved in the first UIP to assumes the same value



under the same conditions, *i.e.* when the other variables leading to the UIP are assigned to the same values. After the new clause is added the solver back jumps to the highest decision level that appears in that clause. The algorithm continues until it finds an assignment that makes the formula true, or it determines none exists. When the solvers have to make a decision, heuristic methods are used to choose both the variable to assign and its value. While they do not ensure to find the optimal solution, they can significantly speed up the process. Firstly assigned variables are the ones appearing in the largest number of clauses since in this way the formula is simplified more quickly. Other prioritised variables are the ones that have been involved in recent conflicts or have been assigned frequently. Indeed, it means that they are active variables that have an important influence on the outcome. Finally, regarding the value assigned, usually a variable is provided with a value according to the one that it has assumed most often in previous conflicts.
These are the most common heuristic strategies that allow to reduce the number of decisions and to improve the chance of find a satisfiable assignment. While using heuristic strategies during the resolution, these algorithms are not stochastic; they follow well-defined logical procedures and given the same input and decision heuristic, the result is the same. It is worth to mention that it exists also stochastic solvers. Unlike deterministic ones, they do not rely on clause learning but rather perform a random search over the assignment space. In essence, they start by randomly assigning values to all variables and while there are unsatisfied clauses, they randomly select one of them and flip the value of a randomly chosen variable within that clause. They are generally faster in proving the satisfiability of a specific variable assignment with respect to deterministic algorithms, however they can get stuck and, above all, cannot prove that a problem is unsatisfiable.

Having defined our methodology that uses patchy particle mixtures, SAT clauses and deterministic SAT solvers, we put it to work on the inverse self-assembly task of the cubic diamond crystal.

2.3 Cubic diamond crystal

One of the most interesting and challenging bottom-up realisation of a target structure is that of the cubic diamond crystal [48, 49]. Realising a perfect cubic diamond on colloidal scale would enable the creation of a photonic crystal. Just like semiconductors have an electronic band-gap, photonic crystals have an optical band gap that allow for light manipulation in a controlled way [50–52]. The wavelengths of photons that are unable to propagate through the crystal are of the same order as the cavity sizes of the cubic diamond lattice.

Cubic diamond self-assembly is as fascinating as it is complex. Indeed, its lattice is an open structure which competes with the hexagonal diamond structure preventing the cubic diamond from forming without defects, such as stacking faults [53]. Several studies have been performed to overcome these difficulties [53,54], including solutions obtained within the SAT-assembly framework [41, 46]. In order to convert the cubic diamond inverse self-assembly task into a colouring problem, we follow the procedure outlined in the previous section. Specifically, we select patchy particles of valence four with a tetrahedral arrangement of the patches as the building units, since each particle in the cubic diamond lattice form four bonds oriented at angles of 109.5 degrees. Since cubic and hexagonal diamonds share the same tetrahedral coordination, we already know that the hexagonal diamond is also likely to assemble from these building units. Hence, each design that is tested for the cubic diamond is also tested against the hexagonal one. While the two crystals have the same local bond geometry, they differ in crystal symmetry and in the long range stacking. This constrains the cell size on which the topology must be defined. Picking a 8 or 16 particle cell will result in designs that are not able to discard the hexagonal lattice since both crystals share the same first and second neighbours. We need to go up to third neighbours to observe a difference. Hence, we construct our SAT problem around a cubic diamond 16-particle cell and a hexagonal diamond 32-particle cell.

The minimal SAT-designed solution is found by progressively increasing the number

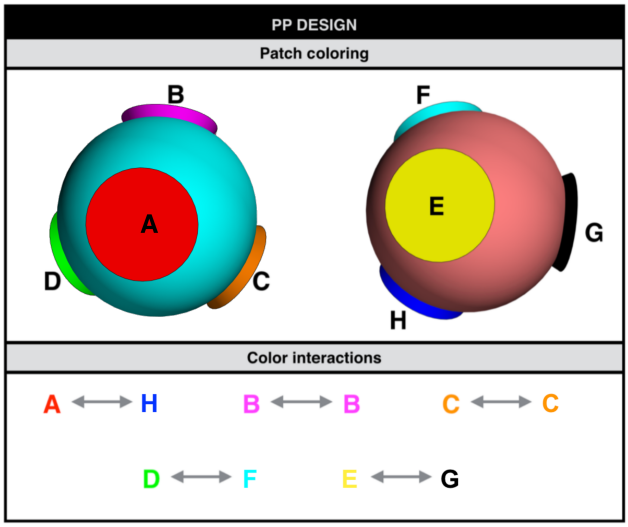
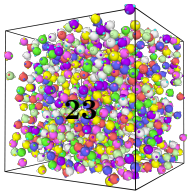


Figure 2.4. 3D representation of the N2c8 solution, SAT-designed to self-assemble a cubic diamond crystal. It is a mixture consisting of two species of patchy particles (N2), the cyan and the red one, that differ in their patch types. There are eight (c8) different patch types, *i.e.* colours, referred to by letters from A to H. Complementary colours identify patches that can bind and are reported in the bottom part of the table. Image adapted from [55].



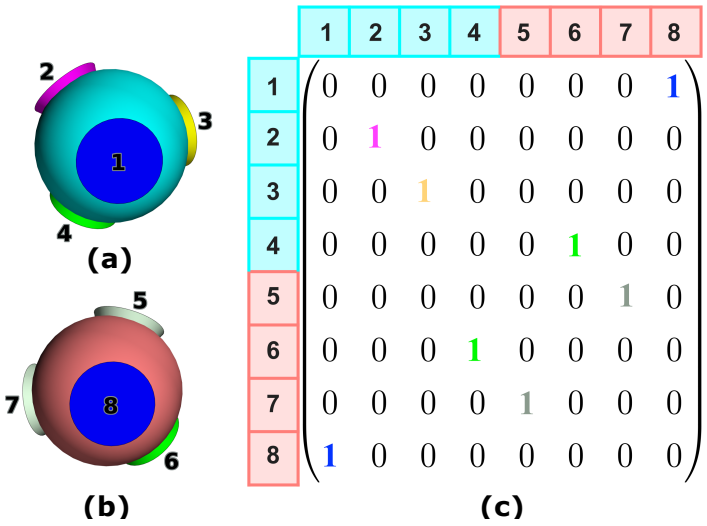


Figure 2.5. Interaction matrix of the N2c8 solution, SAT-designed to self-assembly a cubic diamond crystal . The two species composing the binary mixture are depicted in cyan (a) and red (b). They both have four different patches (numbered from 1 to 8) tetrahedrally arranged that bind according to the interaction matrix (c) where the ones indicate the interacting patches. Matching colours appear for complementary patches (off-diagonal ones), while unique colours represent self-complementary interactions (diagonal ones). Image reproduced from [56].

of colours and species. All single-component systems are compatible with both cubic and hexagonal topologies. However, a binary mixture is sufficient to exclusively select the cubic diamond lattice. The SAT-designed binary mixture identified is referred to as N2c8 solution [9] as it consists of two species (N2) and eight colours (patch types) (c8). Figure 2.4 illustrates this system.

The bonding rules are conveniently captured by the interaction matrix, which mathematically encodes patch compatibility. The interaction matrix for the SAT-designed N2c8 binary mixture is shown in Figure 2.5. It is a square matrix whose dimension is equal to the total number of patches and whose elements m_{ij} are equal to 1 only if patches i and j can bind. Hence, it is a 8×8 square matrix that can be partitioned into four 4×4 blocks. The upper left block is defined by the first four rows and columns, representing the patches of the first species as highlighted in cyan in Figure 2.5. It defines the complementarity between distinct particles of the first species. The same information but for the second species is enclosed in the lower right block made by the last four rows and columns, depicted in red. Finally, the upper-right (first four rows-last four columns) and lower-left (last four rows-first four columns) blocks specify the possible bonds that two different species can establish. In summary, the diagonal blocks express the intraspecies bonds, while the anti-diagonal ones define the interspecies bonds. This is why the matrix is symmetric, diagonal ones signal patches that are self-complementary, and off-diagonal one the simply

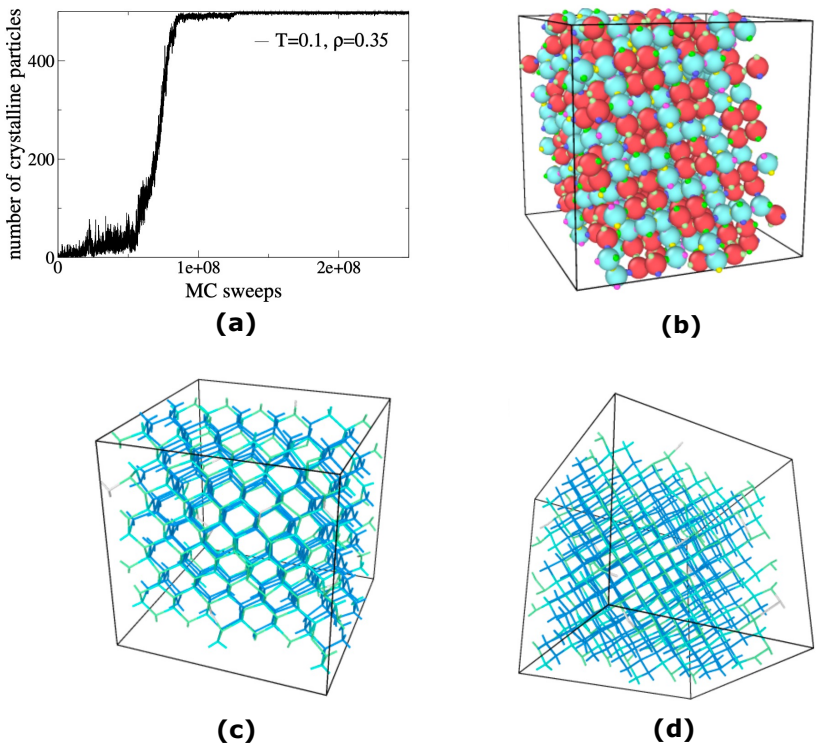
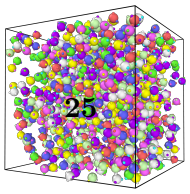


Figure 2.6. Cubic diamond self-assembly from simulations. The N2c8 SAT-designed binary mixture is tested via Monte Carlo simulations in the canonical ensemble. (a) The evolution of the number of crystalline particles indicates a nucleation event. (b) Final configuration snapshot showing the successful self-assembly of the two species of patchy particles into the cubic diamond crystal. Snapshots of the final configuration viewed from a hexagonal plane (c) and a cubic plane (d), where only bonds between particles are shown. Their colour indicate the phase each particle belong to; blue/green corresponds to the cubic diamond phase. All images are taken from a trajectory at temperature $T = 0.1$, density $\rho = 0.35$, and with 500 patchy particles.

complementary ones. Additionally, we note that the number of self-interacting patches is also the trace of the matrix.

By means of molecular simulations, we verified that this mixture crystallises in the diamond-cubic phase, and notably cannot form the hexagonal diamond. A snapshot of the final configuration from a nucleating trajectory is reported in Figure 2.6. A detailed discussion about the nucleation process is presented in the following chapters. Here, we only aim to provide a general overview of the inverse self-assembly task for the cubic diamond and introduce the minimal design identified through the SAT-assembly approach.

In summary, inverse self-assembly sets out to design building units in such a way they spontaneously and exclusively aggregate into a target assembly. However, the route to assembling a specific structure is riddled with obstacles. Even when



the desired structure represents the system's solely thermodynamically stable state, kinetic traps can prevent successful nucleation, leading instead to amorphous aggregate or alternative structures. To overcome these challenges, we propose to rely on the SAT-assembly framework, a methodology that combines patchy particles with Boolean satisfiability equations, to systematically search for successful designs. Among the most sought-after targets is the cubic diamond crystal, valued for its potential application as a photonic crystal. For this crystal, we found the N2c8 design: a binary mixture of patchy particles that not only promotes the formation of the cubic diamond lattice but also suppress the competing hexagonal phase, which can cause stacking faults detrimental to the material's photonic properties. In the following chapters, we delve deeper into this design, using it as a case study to derive valuable, broad insights into the nucleation behaviour and the thermodynamic properties of multi-component systems.

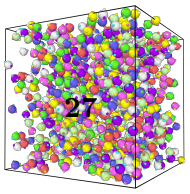
Chapter 3

Patchy Particle's mixtures: theory, simulations and experiments

In this chapter we provide a comprehensive overview of the theoretical methods and computational tools employed to investigate the thermodynamics and the nucleation processes in mixtures of patchy particles. These techniques serve to confirm the effectiveness of the SAT-designs and offer deeper, generalised understanding of multi-component systems, an area less explored compared to single-component systems. Special attention is also devoted to explain why patchy particle level of coarse-graining is meaningful and how colloidal scale building units with such specificity in binding can be experimentally realised.

3.1 DNA nanotechnology

Patchy particle models are particularly suited for addressing the inverse self-assembly problem, as they allow precise control over the number (valence), the placement, and the type of patches that can effectively encode the desired topology requirements. The successful experimental realisation of such systems requires ensuring the binding specificity defined by the interaction matrix. Various strategies have been developed to create the short ranged, anisotropic, and selective interactions characteristic of patchy particles. Most approaches rely on particle shape modelling [57], and on chemical surface patterning [58–62]. One of the most promising technique involves the use of DNA nanotechnology. DNA is an ideal construction material: nucleic acids, through predictable and programmable Watson-Crick base pairing, enable the design of highly specific interactions [63–65]. For example, col-



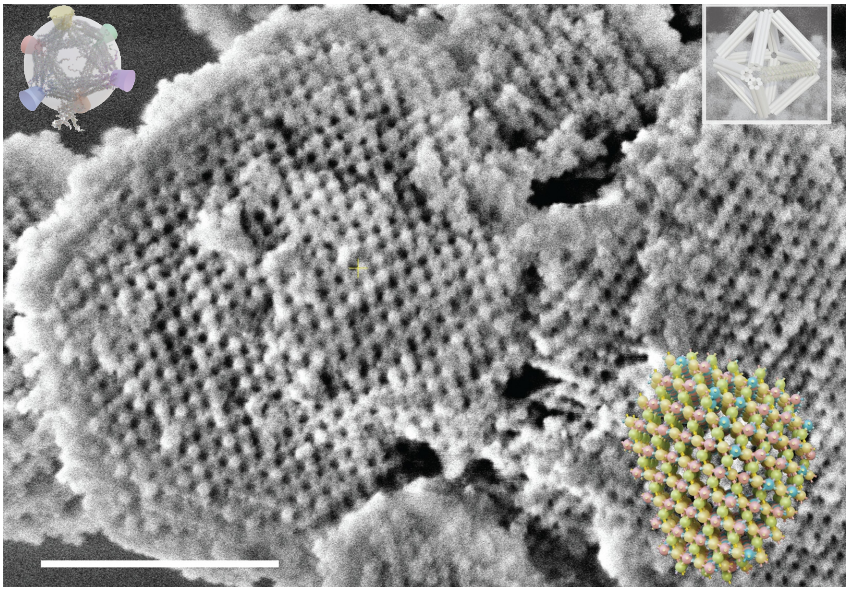


Figure 3.1. Self-assembly of a pyrochlore lattice with DNA origami. SEM image showing the successful assembly of a pyrochlore lattice, schematised in the bottom right using patchy particles. The lattice is formed from octahedral DNA origami, whose shape is highlighted in the upper right inset. The upper left corner shows the position of the single stranded DNA overhangs at the vertices of the origami via coloured cones. The white scale bar represents $1\mu m$. Image adapted from [77].

loidal surfaces can be patterned by attaching single DNA strands to well-defined positions [62, 66–71]. Alternatively, DNA-origami offers a powerful approach for fabricating nanoparticles with exact shape and interaction specificity [48, 72–76]. This methodology has recently been successfully employed to self-assemble a pyrochlore lattice, as shown in Figure 3.1. This achievement demonstrates the feasibility of the method [77], and the following section highlights its key concepts.

3.1.1 From patchy particles to DNA-origami

DNA-origami are three-dimensional structures obtained by a nanoscale folding of a DNA sequence. In Figure 3.2 the main steps towards their synthesis are illustrated. The fabrication of DNA origami involves folding a long single stranded DNA molecule, known as the scaffold strand (ssDNA), into a desired shape using many short DNA strands, called staple strands. These staples provide high precision in directing where the scaffold bends and connects, allowing for the formation of complex three-dimensional structures. The first step is the identification of the scaffold and the staple strands required to obtain the target DNA origami, which consists of tightly packed and interconnected DNA helices. Software tools such as caDNAno [78] facilitate this design process.

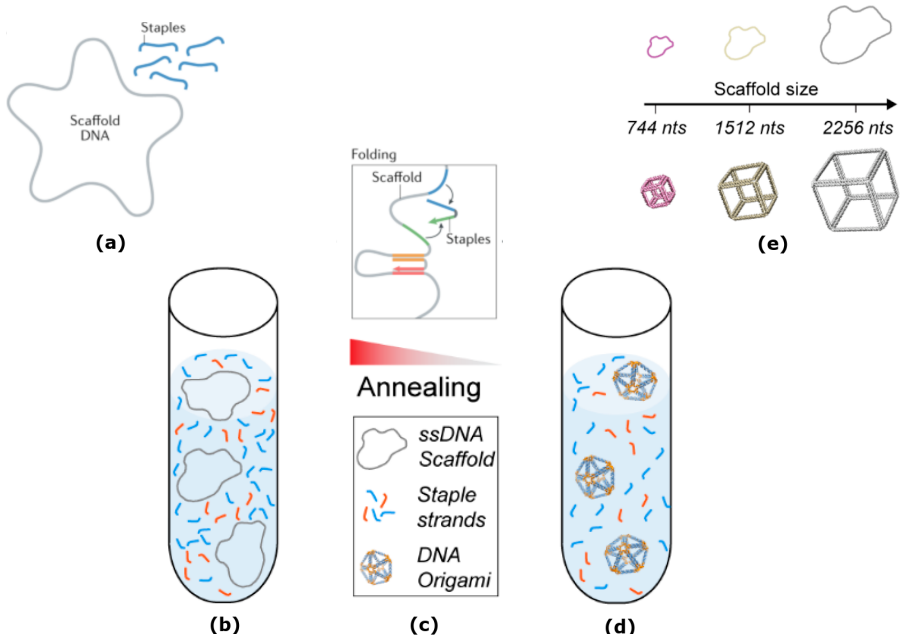
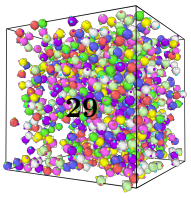


Figure 3.2. Overview of the main steps in DNA origami fabrication. (a) Two types of DNA sequences are required: a long scaffold sequence and numerous short staple sequences. The nucleotide sequences of each staple are designed, based on the target origami, using specialised software. (b) The scaffold and the staple strands are mixed together in a buffer solution. (c) Thermal annealing enables the staples to bind to specific sites on the scaffold leading to the correct folding of the scaffold into the desired shape. (d) The assembled DNA origami are isolated through purification procedures. (e) The size of the DNA origami is determined by the length of the scaffold. Image adapted from [79, 80].

In practice, the user defines the desired structure by placing DNA helices on a grid, specifying the path of the scaffold along this grid so that it spans all parts of the structure, and indicating the crossover points between helices where staples will be placed. The software then automatically generates the nucleotides sequence for each staple strand. These staples are designed to be complementary to specific regions of the scaffold and to have similar melting temperatures to ensure thermodynamic stability, while also minimising sequence repetition to reduce the risk of misfolding. Typically, hundreds of staple strands are required, each composed of tens of nucleotides, while the scaffold is a single long strand made up to thousands of nucleotides. The scaffold sequence is usually known in advance: a common choice is the circular M13mp18 viral DNA, consisting of 7249 nucleotides. Alternatives choices can be made to control the finale size of the origami [79]. Once the DNA strands have been synthesised, they are mixed together in a buffer (a water based solution containing salts and ions that neutralise the negatively charged DNA backbone and promote base pairing between scaffold and staples) and subjected to a



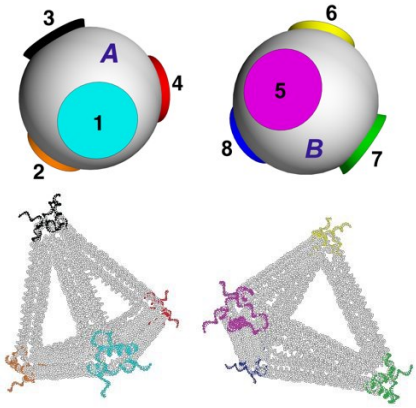


Figure 3.3. Sketch of two patchy particles realised through DNA origami. The four patches tetrahedrally arranged are mapped in single-stranded overhangs at each vertex of a tetrahedron DNA origami made by a nanoscale folding of the DNA scaffold sequence. Different interacting patches correspond to complementary single DNA strands and the self-interacting ones to palindromic DNA strands. Image reproduced from [33].

thermal annealing process to induce folding. Annealing begins with a denaturation step, where the solution is heated to high temperatures (80–90°C) to disrupt any pre-existing folding in the scaffold and break any unintended connection between staple strands. This step is essential to avoid partial folding, secondary structures, or incorrect pairings that can lead to misfolding or low yield. After denaturation, the temperature is gradually reduced to 65°C at a rate of about 1°C per minute. During this phase, staple strands hybridise with their complementary regions on the scaffold. Slow cooling is required to prevent kinetic traps. Small misfolding are corrected during a much slower cooling to room temperature in a range of 10-70 hours that fine-tunes the structure. A final purification step removes excess staples and other contaminants. One common method is precipitation, which isolates the folded DNA origami while leaving unbonded staples in solution. After purification, the assembled origami are typically stored at 4°C until use and remain stable for days to weeks.

In the context of inverse self-assembly DNA-origami can be used as building units. The idea is to map patchy particles into DNA origami, whose shape mirrors the patchy particle bond geometry, and the interacting patches into single complementary DNA strands located at the vertices of each origami. In particular, compatible patches [81] correspond to complementary single DNA strands and self-compatible patches to self-complementary sequences. Figure 3.3 shows a possible realisation of a binary mixture of patchy colloids with eight different patches (colours). The decorated hard-sphere colloidal model (which can be closely experimentally realised [82]) is

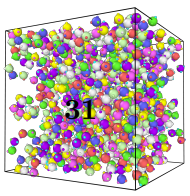
displayed together with a DNA-origami implementation [48]. The vertices of the tetrahedron are functionalised with DNA strands, exploiting DNA addressability to encode patch-patch interactions. Therefore, we propose using DNA-origami as the experimental building units. However, patchy particles simulations remain essential, as current coarse-grained DNA model, such us oxDNA [83–85], are not coarse-grained enough to observe the full self-assembly process within practical simulation times. The detailed nature of the interactions in oxDNA makes it computationally infeasible to simulate the complete self-assembly dynamics. Instead, oxDNA can be effectively used to assess the overall stability of the final assembly and that of individual origami units. This limitation underscores the importance of employing patchy particle models, which offer a more abstracted representation, allowing the study of large-scale assembly behaviour.

In conclusion, DNA origami represents an ideal approach for the realisation of colloidal assemblies, offering accurate control over both geometry and interaction specificity. Beyond their structural versatility, DNA origami also enables the tuning of the photonics properties of the assemblies. This is achieved by designing nanoscale scaffolds capable of hosting different nanoparticles, thereby enabling the fabrication of structures with a periodically modulated refractive index.
In the following section, we present an algorithm we developed to determine DNA sequences that fulfil a specified interaction matrix. This method supports the engineering of a wide range of interaction schemes, including same-colour, distinct-colour and multiple-colour interactions, the latter occurring when colour complementarity is not exclusive.

3.1.2 An algorithm to generate DNA strands from the interaction matrix

We explained that the interaction matrix encodes the binding rules which must be satisfied by the mixture of patchy particles. We also briefly outlined how to use single-stranded DNA sequences to translate such rules into experimental systems, either in DNA wireframe origami [48] or in DNA-functionalised patchy colloids [49]. To bridge the gap between patchy particle models and their DNA based experimental realisation, we present an algorithm that identifies single DNA strands satisfying a desired interaction matrix.

We recall that DNA is a double helix composed of two strands, each of which is a sequence of four types of nucleobases: adenine A, guanine G, thymine T and cytosine C. The nucleobases can selectively bind to each other forming hydrogen bonds and the only possible base pairs are $A-T$ and $C-G$. We also recall that the melting temperature of a DNA oligomer is a function of the length of the



DNA complementary sequences. For example, at a temperature at which DNA complementary sequences of length four are bonded, DNA complementary sequences of length two rarely bind.
We illustrate and demonstrate the algorithm on a specific interaction matrix, but the method is general and can be applied to satisfy arbitrary binding rules. Specifically, we focus on a binary mixture of particles with four patches each, interacting with the interaction matrix in Eq. 3.1

$$\begin{pmatrix} 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \end{pmatrix} \tag{3.1}$$

It is a 8×8 square matrix, as the one of the N2c8 solution, introduced in the previous chapter (see Figure 2.5). However, there is a key difference in between the two. In the N2c8 interaction matrix, each row contains a single element equal to one, indicating that each patch has a unique complementary partner. In contrast, matrix 3.1 features two elements equal to one per row, meaning that each patch has two possible complementary partners. This implies that, unlike the previous case, colour complementarity is not exclusive; a patch does not form multiple bonds

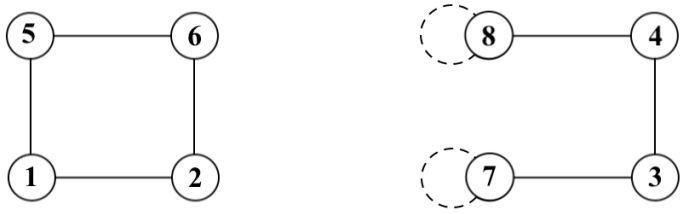


Figure 3.4. Bond topology of the double binding matrix. The patch connections required by the interaction matrix in Eq. 3.1 result into two disconnected bonded "clusters": a ring of four patches, and a chain of four patches.Image reproduced from [33].

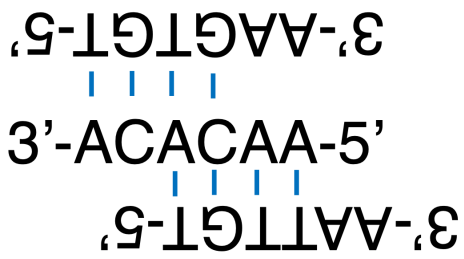


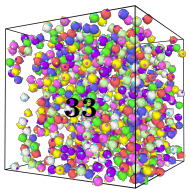
Figure 3.5. Schematic example of a single DNA strand satisfying the double binding condition. Three single DNA strands in which the central one is able to binds through four consecutive nucleotides with two other strands (reversed in their 3'-5' order). Image reproduced from [33].

simultaneously but can choose between two valid binding partners. For example, the colour of the first patch of the first species is complementary both to the colour of the second patch of the first species and to the colour of the first patch of the second species. We are therefore considering a more general scenario where multiple binding options are possible. In particular, the interaction matrix 3.1 encodes a design in which each patch can bind either to another patch of the same species, or to one of a different species. As we will see in the next chapters, this type of multiple binding introduces an interesting thermodynamic behaviour in the system. The interaction matrix 3.1 can also be represented as a list of nodes (the eight patches, labelled from 1 to 8 in Figure 3.3) connected by lines representing the 1 in the interaction matrix, resulting in the connected graphs in Figure 3.4. We name the left graph in Figure 3.3 ring graph while the right one chain graph. In the ring graph each patch binds to two different patches, while in the chain graph the first and the last node bind to one identical patch and to a different patch.

Our goal is to find single DNA strands that represent, as close as possible, the interaction matrix in Eq. 3.1 or equivalently the bond topology in Figure 3.4. Specifically, each strand must be able to bind with up to two different other strands. This double binding condition can be realised by defining as bonded two single DNA sequences having a number of n_b consecutive base pairs (a realistic value could be $n_b = 4$) that is smaller than the total number of nucleotides n_s of the sequences. An example for $n_s = 6$, is shown in Figure 3.5.

To implement interactions among multiple colours, we design each complementarity using only a specific subsequence of nucleotides. To identify eight DNA single strands of length n_s which satisfy the interaction matrix in Eq. 3.1, we propose the following algorithm:

- we generate all the 4^{n_s} different oligomers of length n_s and evaluate, for each



pair of them the maximum number of consecutive bases n_{max} that bind to each other. It is important to remind that DNA has directionality and the two complementary strands that form it have opposite directions: one goes from the five-prime end to the three-prime end $5^{'} \longrightarrow 3^{'}$ and the other one from the three-prime end to the five-prime end $3^{'} \longrightarrow 5^{'}$. We then construct a $4^{n_s} \times 4^{n_s}$ matrix whose elements are the strength of the binding between the two strands. For clarity, we identify here the strength with the n_{max} value. A more elaborate formulation could use the binding Gibbs free energy or the strand pair melting temperature, calculated for example with the SantaLucia nearest-neighbour model [86].

- We set a threshold $n_{\text{threshold}}$ for the largest value of n_{max} which can be safely assumed as non-bonding. In the case of $n_b = 4$, this can be chosen as 2. Indeed, at the melting temperature of sequences of length 4, the binding probability of sequences of length 2 is negligible. Put zero in the $4^{n_s} \times 4^{n_s}$ matrix of n_{max} for all elements $n_{\text{max}} \leq n_{\text{threshold}}$. In this way, the $4^{n_s} \times 4^{n_s}$ matrix has non-zero elements only for pairs of sequences which bind to each other.

- Eliminate from the set of all possible strands (randomly) one of the two strands for which n_{max} is larger than n_b ($n_{\text{max}} = 5$ and 6 in our example). This eliminates, among others, one of each pair of complementary strands and one of the two self-complementary sequences (palindromic in the DNA nomenclature). To eliminate the strand, it is sufficient to fill with zero the row and column associated to that strand. This makes sure that, among the remaining set of strands, the strongest binding is indeed the ones with $n_{max} = n_b$.

- By now, the $4^{n_s} \times 4^{n_s}$ matrix contains, weighted by their strength, all possible bonds between all possible strands, with values of n_{max} from $n_{\text{threshold}} + 1$ to n_b (from 3 to 4 in this specific case). This matrix can also be seen as a network of bonds (links) between the 4^{n_s} strands (nodes). We then run a search on this network to identify the desired bonding clusters. In the specific case (Figure 3.4)we first identify all rings of four nodes connected by links with $n_{\text{max}} = n_b$. We then eliminate the rings which can be short-cut by a intra-ring bond and all rings in which one of the node is able to self-bind.

- For each of the remaining ring, we eliminate from the list of all possible DNA strands, the strands (i.e. the nodes) which are connected to the ring strands. This guarantee that the four strands defining the ring do not interact with any other remaining strand. Among the remaining nodes, we search for all chains of length four (in this specific case) on the network starting and ending with

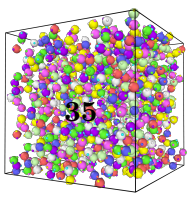
a node which is able to self-bind. As for the ring, we eliminate all chains in which intra-chain bonds are present.

- Iterating this procedure for each ring results in a list of DNA strands, all good candidates to experimentally realise the required interaction matrix. In a more refined treatment, each of these possible solutions can be examined to select the smallest variance in the ΔG of binding between different patches (compensating the different strength of the A-T and G-C pairing), and further restrict sequence selection so that ΔG distance between pairs with n_b and pairs with $n_{\text{threshold}}$ is as large as possible to minimise crosstalk. ΔG of binding between any pair of sequences would be calculated using SantaLucia's model, and can be obtained from with available strand analysis and optimisation tools such as NUPACK [87].

In Eq. 3.2 we report one of the possible sets of 6 bases for 8 sequences that bind according to the requested topology in Figure 3.4.

$$\begin{aligned} ① &= AAGGGG \\ ② &= CCCCCC \\ ③ &= ACACAA \\ ④ &= AAGTGT \\ ⑤ &= CCCTCA \\ ⑥ &= AGGGGA \\ ⑦ &= AATTGT \\ ⑧ &= CACTAG \end{aligned} \tag{3.2}$$

These sequences correspond to the following matrix of maximum number of paired base pairs.



$$\begin{pmatrix} 0 & 4 & 1 & 1 & 4 & 0 & 2 & 2 \\ 4 & 0 & 0 & 1 & 0 & 4 & 1 & 1 \\ 1 & 0 & 0 & 4 & 1 & 1 & 4 & 1 \\ 1 & 1 & 4 & 1 & 2 & 1 & 2 & 4 \\ 4 & 0 & 1 & 2 & 1 & 4 & 2 & 2 \\ 0 & 4 & 1 & 1 & 4 & 0 & 1 & 2 \\ 2 & 1 & 4 & 2 & 2 & 1 & 4 & 2 \\ 2 & 1 & 1 & 4 & 2 & 2 & 2 & 4 \end{pmatrix} \tag{3.3}$$

Considering as bonded only the elements filled with a 4 (i.e. neglecting pairing of one or two nucleotides that are smaller than the threshold we set), this matrix coincides with the matrix in Eq. 3.1. We call attention on the fact that Eq. 3.1 has two identical row/columns. This makes it possible to reduce, if needed, the number of distinct colours from eight to six without altering the connectivity table. Strand ⑤ in Eq. 3.2 can be substituted with strand ②, and strand ① can be substituted with strand ⑥ (or vice versa).

Having established both the computational advantages and the experimental feasibility of patchy particle systems, we now introduce the theoretical and computational frameworks we employ in the study of multi-component systems of patchy particles.

3.2 Wertheim's perturbation theory

Theoretical predictions about the thermodynamic behaviour of mixture of patchy particles can be obtained by the Wertheim first order perturbation theory [88]. While it was originally developed to derive a mean-field theory for associating fluids, it can be easily generalised to patchy particles [89, 90]. A brief overview of the key aspects of Wertheim's theory is provided in Appendix B. In several works [91–97], the theory has been adopted to study, in detail, the static (*e.g.* percolation) and thermodynamic (*e.g.* phase behaviour) properties of patchy particle systems, both in pure components and in mixtures, showing excellent qualitative agreement with numerical simulations.

Wertheim developed a perturbative method that estimates the effect of the attractive patches on the Helmholtz free energy of a reference system of hard spheres. By following the conventions of Refs. [92, 95], the Helmholtz free energy per particle in units of k_BT of an n-component mixture can be expressed as the sum of a reference and a bonding contribution:

$$\beta f = \beta f_{\text{reference}} + \beta f_{\text{bonding}} \tag{3.4}$$

The reference contribution does not account for association process and contains an ideal gas term and the hard sphere excess term:

$$\begin{aligned} &\beta f_{\text{reference}} = \beta f_{\text{ideal}} + \beta f_{\text{HS}} \quad \text{with} \\ &\beta f_{\text{ideal}} = \ln \rho - 1 + \sum_{i=1}^{n} x^{(i)} \ln \left(x^{(i)} V_i \right) \\ &\beta f_{\text{HS}} = \frac{4\phi - 3\phi^2}{(1-\phi)^2} \end{aligned} \tag{3.5}$$

where ρ is the density, $x^{(i)}$ is the molar fraction of species i, V_i is the thermal volume, and ϕ is the packing fraction that is equal to ρV_s, where $V_s = \sigma^3 \pi / 6$ is the volume of a single particle. The bonding contribution contains the sum over the species ($\sum_{i=1}^{n}$) and the sum over the patches of a certain species i ($\sum_{\alpha \in \Gamma(i)}$); the number of patches of species i is denoted as $n(\Gamma(i))$.

$$\beta f_{bonding} = \sum_{i=1}^{n} x^{(i)} \left[\sum_{\alpha \in \Gamma(i)} \left(\ln X_\alpha^{(i)} - \frac{X_\alpha^{(i)}}{2} \right) + \frac{1}{2} n(\Gamma(i)) \right] \tag{3.6}$$

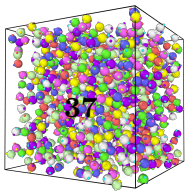
The quantity $X_\alpha^{(i)}$ is the probability that a patch α on species i is not bonded, and it is defined by the following mass balance equation

$$X_\alpha^{(i)} = \left[1 + \phi \sum_{j=1,n} x^{(j)} \sum_{\gamma \in \Gamma(j)} X_\gamma^{(j)} \Delta_{\alpha\gamma}^{(ij)} \right]^{-1} \tag{3.7}$$

where $\Delta_{\alpha\gamma}^{(ij)}$ does not depend on the species (i, j) if the diameter of all patchy particles is always the same, and it is given by

$$\Delta_{\alpha\gamma}^{(ij)} = \Delta_{\alpha\gamma} = \frac{1}{V_s} \int_{V_{\alpha\gamma}} g_{HS}(\mathbf{r}) (e^{\beta \epsilon_{\alpha\gamma}} - 1) d\mathbf{r} \tag{3.8}$$

where g_{HS} is the radial distribution function of hard spheres, $V_{\alpha\gamma}$ is the bonding volume and $\epsilon_{\alpha\gamma}$ is the bonding energy, both related to a bond between patches α and γ. As for any short-ranged potential, the static properties are controlled by the bonding volume [97], *i.e.* the volume in which a particle can move while being



bonded to another particle. Considering the Kern-Frenkel potential, introduced in Section 2.1, the bonding volume assumes the following simple expression

$$V_b = \frac{4\pi}{3}((\sigma+\delta)^3 - \sigma^3)\left[\frac{1-\cos(\theta_{max})}{2}\right]^2. \tag{3.9}$$

While patches are generally different and therefore interact via distinct Kern-Frenkel potentials, in the following, we consider that all bonds have the same bonding volume, *i.e.* $V_{\alpha\gamma} = V_b$. As a result, the dependence of $\Delta_{\alpha\gamma}$ on patch types arises solely from the bonding energy $\epsilon_{\alpha\gamma}$. Specifically, $\epsilon_{\alpha\gamma} = 0$ when two patches are not compatible, and $\epsilon_{\alpha\gamma} \neq 0$ otherwise with the exact value that can vary across different pairs of compatible patches. The parameter $\Delta_{\alpha\gamma}$ characterises the bond formed between patch α on a patchy particle of species i and patch β on a patchy particle of species j. It highlights that a bonding event results from a competition between the energetic gain in forming a bond, *i.e.* the Boltzmann factor, and the entropic cost in forming a bond, which is due to the reduction in translational entropy because of the confinement within the bonding volume. Since the bonding volume V_b is typically much smaller that the particle volume V_s, the ratio V_b/V_s is much smaller than one and the entropic loss is significant. As evident from equation. 3.9, the more the attraction is short-ranged, the more the bonding volume is small. Therefore, binding occurs only if the attraction is sufficiently strong, *i.e.* when $k_bT/\epsilon < 1$. According to the Arrhenius law, this ratio also governs the bond lifetime: an extended network forms when $k_bT/\epsilon < 1$, while a stable network emerges only when $k_bT/\epsilon \ll 1$. We can also approximate the radial distribution function with an expansion around its value at contact, as detailed in Refs. [98, 99]. With these approximations, affecting the results only quantitatively, but not qualitatively, equation 3.8 becomes:

$$\Delta_{\alpha\gamma} = \frac{1}{V_s} 4\pi\chi^2 \left\{ \left[\frac{(1+\delta)^3 - 1}{3} A_0\right] + \left[\frac{(1+\delta)^4 - 1}{4} A_1\right] \right\} (e^{\beta\epsilon_{\alpha\gamma}} - 1) \tag{3.10}$$

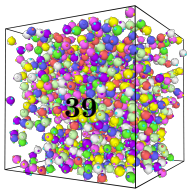
with

$$\begin{aligned} A_0 &= \tfrac{1-\frac{\phi}{2}+\frac{9\phi}{2}(1+\phi)}{(1-\phi)^3} \\ A_1 &= \tfrac{-\frac{9\phi}{2}(1+\phi)}{(1-\phi)^3} \\ \chi &= \tfrac{1-\cos\theta_{max}}{2} \end{aligned} \tag{3.11}$$

In summary, Wertheim's first order perturbation theory serves as the theoretical foundation for our predictions of the thermodynamic behaviour of patchy particle mixtures. Its strength lies in its ability to provide a good estimate of the Helmholtz free energy for multi-component systems of patchy particles by relying solely on the structural properties of the reference system and on the interaction potential characterising patchy particles. We underline that the Wertheim first order perturbation theory is a mean-field theory. The key element in the bonding contribution, $X_\alpha^{(i)}$, is a bonding probability that represents, on average, the fraction of unbonded sites of type α on particle i. Spatial correlations are neglected beyond the reference pair correlation function, sites cannot be engaged in more than one bond at the same time, loop formations are forbidden, and many-body bonding effects are not considered. Binding events are treated as pairwise, site-site interactions that are statistically independent and can only occur between particles belonging to different clusters. Nearby particles are accounted for only via the reference pair correlation function, and the theory does not consider the specific bonding arrangement, but rather how likely a site of type α is unbonded. This average probability is computed self-consistently, losing site-specific detail and retaining only site-type specificity. Although it operates under the same assumptions as a mean-field theory, it nonetheless captures the core thermodynamic behaviour of patchy particle mixtures, exhibiting strong qualitative agreement with the numerical simulations, whose practical details are introduced in the following section.

3.3 Monte Carlo simulations

Beyond theoretical analysis, we explore the SAT-design mixtures via numerical simulations. Specifically, we focus on studying their nucleation by performing Monte Carlo (MC) simulations, in preference to Molecular Dynamics (MD) simulations. MC simulations use a stochastic sampling approach, wherein the system eventually explores its equilibrium configuration through the acceptance or rejection of randomly proposed moves, while MD integrates Newton's equations of motion to evolve the system in time. As a result, nucleation events are typically observed in significantly less computational time with MC simulations. This is because nucleation is an activated process and, in MD simulations, the system must wait for sufficiently large and rare thermal fluctuations to overcome the nucleation free energy barrier. In contrast, a single MC move can, in principle, directly take the system across the barrier and sample the state corresponding to the free energy minimum. Thus, MC simulations allow for a more efficient exploration of the configurational space relevant for the thermodynamic characterisation of nucleation. Although MC simulations do not capture the dynamics of the process as MD simulations do, this limitation is



minor because nucleation outcome is mainly controlled by equilibrium properties.

3.3.1 Metropolis algorithm and roto-translational moves

In Monte Carlo simulations, physical observables are computed as ensemble averages over sampled configurations, weighted by their equilibrium probability density, which depends on the chosen ensemble. Since we are interested in equilibrium properties, such as energy and density, that depend only on particle positions, momenta can be factorised and integrated out. Thus, the sampling is performed in position space (and possibly volume or particle number, depending on the ensemble). To explore this space, trial moves modifying particle positions (as may be system volume or particle number) are proposed. Monte Carlo simulations take advantage from the Metropolis algorithm to assign acceptance probabilities to these moves. The Metropolis algorithm exploits the detailed balance sufficient condition to define these probabilities in such a way that the sampled configurations follow the desired equilibrium probability density. The method is statistically unbiased, *i.e.* the desired probability distribution will be the one asymptotically sampled regardless of the choice of the move. Nevertheless, the move definition influences the efficiency of the sampling. By efficiency, we refer to the computational effort required to sample configurations at equilibrium that, in turns, affects the statistical error. The best moves are those leading to the lowest statistical uncertainty in a fixed computational time.

When simulating patchy particle systems, the main trial moves attempted to achieve equilibrium are roto-translational moves. For each move, a random particle is displaced by a random radial displacement and rotated around a random axis of rotation by a random angular displacement. The move is then accepted according to the Metropolis acceptance probability, given by $min\{1, e^{-\beta\Delta U}\}$, where ΔU is the potential energy difference between final and initial states. If the move leads to the formation of a new bond, then $\Delta U = \epsilon < 0$, indicating an energy gain; in this case the move is always accepted. Conversely, if the move attempts to break an existing bond, then $\Delta U > 0$, and the acceptance probability decreases as the temperature decreases, since the energy gain in keeping a bond becomes more significant. When the displacement results in particles overlap, the potential energy diverges yielding an acceptance probability of zero, and the move is always rejected. Efficient sampling of the configurational space is achieved by applying roto-translations to individual particles. Displacing multiple particles simultaneously increases the likelihood of overlaps and consequent move rejections. This reduces the efficiency of the simulation, leading to a slower system evolution, increased autocorrelation, and higher statistical error. Roto-translational moves are well suited

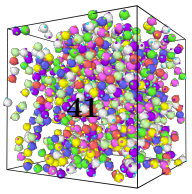
when working with anisotropic particles that require an accurate treatment of both rotational and translational degrees of freedom. They are sufficient to sample the canonical ensemble. However, for simulations in the isobaric-isothermal ensemble, volume change moves must also be included. In the case of the gran-canonical ensemble, particle insertion/deletion moves are additionally required.

3.3.2 AVB moves

When addressing association of patchy particles interacting *via* anisotropic and short-ranged interactions, simply roto-translation moves are often not the more efficient choice. Indeed, patchy particles self-assembly occurs when the thermal k_BT energy is much smaller than the bonding energy ϵ, which makes the Metropolis acceptance probability of an MC move that breaks a bond extremely low. Thus, almost all moves that try to break a bond are rejected, dramatically worsen the sampling of the phase space. To overcome this drawback, aggregation-volume-bias-moves (AVB) [100, 101] are introduced. They facilitate bond breaking by enhancing the associated acceptance probability. Specifically, there are two types of AVB moves (to satisfy the detailed balance): the AVB-B move and the AVB-U move. The AVB-B move (B stands for bonded) attempts to form a bond by moving a patchy particle into the bonding volume (V_b) of another patchy particle. This results in the creation of a new a bond between a previously unbonded pair of patchy particles. In contrast, the AVB-U move (U stands for unbonded) tries to break an existing bond by moving a bonded patchy particle outside the bonding volume ($V_o = 4\pi V - V_b$, where V is the volume of the simulation box) of its partner, thus eliminating the bond between the patchy particle pair. These moves are biased, and their acceptance probabilities are defined as

$$\begin{aligned} A_{AVB-B} &= \min\left\{1, \tfrac{(N-N_i-1)V_b}{(N_i+1)V_o} e^{-\beta\Delta E}\right\} \\ A_{AVB-U} &= \min\left\{1, \tfrac{N_i V_o}{(N-N_i)V_b} e^{-\beta\Delta E}\right\} \end{aligned} \tag{3.12}$$

where N_i is the number of particles that are bonded to particle i. Importantly, the acceptance probability for breaking a bond using an AVB-U moves is significantly higher than that of a simple roto-translation move. This is because the ratio V_o/V_b is much greater than one, given that the bonding volume V_b is much smaller than its complementary volume V_o. This effect is particularly pronounced in patchy particle system, where V_b is extremely small, making AVB moves highly effective for enhancing sampling.



In studying our patchy particle systems, we use MC simulations with AVB moves, but we checked with independent simulations that the same nucleation behaviour is also observed with unbiased simulations, albeit requiring longer simulation times.

Roto-translational moves enhanced with AVB scheme are enough to sample the stable phase of the system, but if we are interested in computing the free energy barrier of nucleation, more advanced techniques are needed. In the following section, we briefly report the Umbrella Sampling method.

3.3.3 Umbrella sampling

Nucleation phenomena are activated processes, i.e. in order to happen a free energy barrier must be overcome. We do not aim at fully reconstruct the free energy landscape, but rather we want to know the free energy cost $F(n)$ of forming a nucleus of size n in a metastable phase. We can obtain it from the probability $P(n)$ that the system is in a configuration where a crystalline cluster of size n is formed. Indeed:

$$F(n) = -k_B T \; logP(n) \tag{3.13}$$

Conceptually, the probability distribution along the reaction coordinate n is a quantity easily measurable in Monte Carlo simulations. However, when states of interest are rarely accessed, the statistical accuracy is poor and direct sampling is ineffective. This is exactly the case of nucleation phenomena. Indeed, in this scenario the system is stuck in a relative free energy minimum corresponding to the metastable phase (liquid), and eventually, when a rare fluctuation great enough to overcome the barrier (considerably larger than $k_B T$) occurs, it reaches the absolute minimum, and we start sampling configurations in the stable phase (solid). This means that with direct sampling we effectively probe only the configuration space around minima. However, to construct the nucleation free energy barrier, we need to sample the configuration space where crystalline clusters from small sizes to the critical one are formed, i.e we aim to sample configurations along and on top of the barrier. We can do it by modifying the Hamiltonian of the system when computing the acceptance probability of a move: an additional term, named biasing potential $V_{bias}(n)$, is added in order to ensure that the system visits regions that it might not otherwise explore sufficiently. The biasing potential depends only on the reaction coordinate n and $V_{bias}(n) = -F(n)$ would be the ideal choice since it guarantees a uniform sampling in the whole reaction coordinate interval. However, $F(n)$ is not known, and it is exactly what we want to compute. A solution is provided by the Umbrella Sampling (US) [102]: the range of interest of the reaction coordinate is

divided into small windows and a potential that constraints the system to explore regions close to the desired value of n is applied in each of them. The name derives from the biasing potential effect to bridge energetically separated regions. Often a harmonic potential is chosen such that:

$$H = H_{system} + V_{bias}(n) \quad \text{with} \quad V_{bias}(n) = \frac{1}{2}k(n - n_0)^2 \tag{3.14}$$

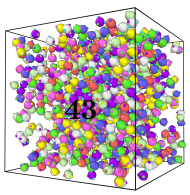
The system is forced to explore the phase space regions around n_0 with fluctuations depending on the value of k. By using a bias potential, we measure a bias distribution $P_{bias}(n)$ in each window, but we can easily retrieve the original unbiased one:

$$P(n) = P_{bias}(n)e^{\beta V_{bias}(n)} \tag{3.15}$$

The overall $P(n)$ is computed by combining the results from the different windows, therefore it is necessary that single-window distributions overlap with the neighbour ones. This means that the strength of the bias k must be carefully calibrated and may change for different windows (in our case, a constant value of $k = 0.005$ was sufficient). For instance a high k is required if the free energy is steep. However, we do not know *a priori* the slope of the free energy. To minimise the problem of selecting a bias potential for each window that ensures proper sampling, Virnau and Muller proposed the Successive Umbrella Sampling technique (SUS) [103], a kind of US in the limit of small windows size. In each window i the system is constrained to explore values of the reaction coordinate n belonging to a small interval of size ω ($n \in [i\omega, (i+1)\omega]$); Monte Carlo simulations are run in the grand-canonical ensemble where all moves bringing the system outside the window ranges are rejected. The windows size ω is so small, usually equal to 2 or 3, that the use of biasing potential is not necessary. Free energy difference between a state with reaction coordinate of n and the reference one $n = 0$ can be computed through the $P(n)/P(0)$ ratio. Here, this ratio is reconstructed by the product of the ratios between the boundary histograms of each window.

$$\frac{P(n)}{P(0)} = \frac{H_{0r}}{H_{0l}}\frac{H_{1r}}{H_{1l}} \cdots \frac{H_{nr}}{H_{nl}} \tag{3.16}$$

Indeed, during the simulation, for each independent window i, the number of times each configuration n is visited is saved into a histogram $H_i[n]$. It is important to underline that, in order to satisfy the detailed balance, the left boundary histogram



$H_{il} = H_i[i\omega]$ (and the right one $H_{ir} = H_i[(i+1)\omega]$) must be incremented by one even when we reject the moves overcoming the lower (or the upper) limit of the interval. By computing the non boundary histograms it is possible to define an adaptive weight if a biasing potential is still used. Both US and SUS are powerful techniques that enable to compute free energy projection along a reaction coordinate without knowing the free energy behaviour.

In brief, patchy particles are a valid coarse-grained model for all those systems characterised by anisotropic interactions such us DNA origami that can naturally mirrors the required bond geometry and specificity and whose self-assembly is hard to simulate with finer coarse-grained models like oxDNA. To verify the impact of combining the SAT-assembly methodology with DNA origami nanotechnology we study the thermodynamic and nucleation properties of patchy particle systems both theoretically, through the Wertheim first order perturbation theory and computationally via Monte Carlo simulations. In the next chapter, we focus on the determination of the thermodynamics behaviour of the N2c8 SAT-design binary mixture, introduced in Section 2.3, after a general overview on how phase diagram computation can be held.

Chapter 4

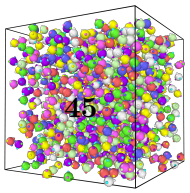
Reconstructing phase diagrams

This chapter outlines the methodology used to compute phase diagrams of mixtures and illustrates the phase behaviour of the SAT-designed N2c8 binary mixture introduced in Section 2.3. The presented results are extracted form papers *C. Beneduce et al., J. Chem. Phys., 2023* [55] and *C. Beneduce et al., ACS Nano, 2023* [33].

4.1 Isochoric thermodynamics

In single-component systems, the binodal curve is typically determined by integrating the Clausius-Clapeyron differential equation. This integration strategy can be extended to multi-component systems. To this end, it is convenient to adopt the isochoric thermodynamics framework [104, 105]. It allows the formulation of a set of differential equations that describe the thermodynamic behaviour at phase coexistence and can be integrated numerically.
In the canonical ensemble, the thermodynamic state of a n-component mixture is specified by temperature T, molar density ρ and mole fractions x_i, where i runs over the different n species. However, mole fractions have some disadvantages: they are not independent variables and, conversely to density, they are dimensionless causing the density mole fractions space to have an ill-defined metric. On the contrary, in the isochoric thermodynamics, the independent variables are molar densities ρ_i and the fundamental thermodynamic potential is the Helmholtz energy density Ψ. They are defined as:

$$\begin{aligned} \rho_i &= x_i \rho \\ \Psi(\boldsymbol{\rho}, T) &= \tfrac{A}{V} = a\rho \end{aligned} \tag{4.1}$$



where A is the Helmholtz energy and a is the molar Helmholtz energy, while ρ is the molar density of the n-component mixture $\rho = \sum_{i=1}^{n} \rho_i$ and $\boldsymbol{\rho}$ is the vector of molar densities $\boldsymbol{\rho} = (\rho_1, \rho_2, \dots \rho_n)$.
The local curvature of the Helmholtz energy density is encoded in the hessian matrix:

$$H = \begin{bmatrix} \left(\frac{\partial^2\Psi}{\partial{\rho_1}^2}\right)_T & \left(\frac{\partial^2\Psi}{\partial\rho_1\partial\rho_2}\right)_T & \cdots & \left(\frac{\partial^2\Psi}{\partial\rho_1\partial\rho_n}\right)_T \\ \left(\frac{\partial^2\Psi}{\partial\rho_2\partial\rho_1}\right)_T & \left(\frac{\partial^2\Psi}{\partial{\rho_2}^2}\right)_T & \cdots & \left(\frac{\partial^2\Psi}{\partial\rho_2\partial\rho_n}\right)_T \\ \vdots & \vdots & \ddots & \vdots \\ \left(\frac{\partial^2\Psi}{\partial\rho_n\partial\rho_1}\right)_T & \left(\frac{\partial^2\Psi}{\partial\rho_n\partial\rho_2}\right)_T & \cdots & \left(\frac{\partial^2\Psi}{\partial{\rho_n}^2}\right)_T \end{bmatrix} \tag{4.2}$$

If it is positive defined, then the state is a stable state. We know that two phases (labelled $'$ and $''$ in the following) coexist in equilibrium at constant temperature if, along the phase boundary, the pressure and the chemical potentials of each component are equal for both phases. This means that the variation of the pressure and of the chemical potentials along the phase boundary must be the same for both phases:

$$\begin{aligned} d\mu_i^{'} &= d\mu_i^{''} \quad \text{with } i = 1, 2, \dots n \\ dP^{'} &= dP^{''} \end{aligned} \tag{4.3}$$

having defined the chemical potentials and the pressure as $\mu_i = \partial\Psi/\partial\rho_i$ and $P = -\Psi + \sum_{i=1}^{n} \rho_i\mu_i$, respectively. By exploiting these conditions, we can identify the system of first order differential equations to be numerically integrated for evaluating the coexistence region. For the isothermal phase equilibrium of a binary mixture we must solve:

$$\begin{aligned} &\begin{bmatrix} H_{\Psi,1}^{'} \cdot \boldsymbol{\rho}^{''} & H_{\Psi,2}^{'} \cdot \boldsymbol{\rho}^{''} \\ H_{\Psi,1}^{'} \cdot \boldsymbol{\rho}^{'} & H_{\Psi,2}^{'} \cdot \boldsymbol{\rho}^{'} \end{bmatrix} \left(\frac{d\boldsymbol{\rho}}{dP}\right)_{T,\sigma}^{'} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \\ &H_{\Psi}^{''} \left(\frac{d\boldsymbol{\rho}}{dP}\right)_{T,\sigma}^{''} = H_{\Psi}^{'} \left(\frac{d\boldsymbol{\rho}}{dP}\right)_{T,\sigma}^{'} \end{aligned} \tag{4.4}$$

where the subscript σ indicates that derivatives are calculated along the phase boundary, and $H_{\Psi,i}$ indicates the i-th row of the Hessian matrix with $n = 2$ in Eq. 4.2: $H_{\Psi,1} = \left(\left(\frac{\partial^2\Psi}{\partial{\rho_1}^2}\right)_T, \left(\frac{\partial^2\Psi}{\partial\rho_1\partial\rho_2}\right)_T\right)$, $H_{\Psi,2} = \left(\left(\frac{\partial^2\Psi}{\partial\rho_2\partial\rho_1}\right)_T, \left(\frac{\partial^2\Psi}{\partial{\rho_2}^2}\right)_T\right)$. By starting from available accurate initial values, the integration of the derivatives of the molar densities in the coexisting phases over the desired range of pressure predicts how molar densities of vapour and liquid change with pressure. This enables the construction of the binary mixture pressure-concentration and density-concentration binodal curves. A brief discussion of complementary methods for investigating the role of fluctuations in these phase transitions is presented in Appendix C.

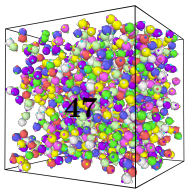
In summary, once a single pair of coexisting point is known, the entire coexistence region can be reconstructed by calculating how these points evolve along the binodal curve. This procedure relies on having a theoretical model, that is an analytical expression for the systems' free energy. Although effective, the integration gets stiff and does not proceed further close to critical points: the step-size of the adaptive step-size integrator [105] progressively decreases since the hessian determinant vanishes at critical points.

Isochoric thermodynamics and numerical integration are used to compute the coexistence region theoretically, while to estimate it computationally, Monte Carlo simulations in the Gibbs ensemble are performed.

4.2 Gibbs ensemble simulations

The computational study of coexistence between two phases at a certain temperature can be accomplished through Monte Carlo simulations in the Gibbs ensemble [106, 107]. Coexistence occurs between two simulation boxes that virtually interact with each other, without the disadvantage of sharing an explicit interface. The virtual interaction is set by specific Monte Carlo moves. In addition to roto-traslational moves presented in Section 3.3, the Gibbs ensemble technique incorporates volume moves and particle transfer moves. Volume moves alter the size of the two boxes keeping the total volume fixed, while in particle transfer moves a particle is displaced from one simulation box to the other. Figure 4.1 illustrates the Gibbs ensemble scheme, which enables the two boxes to achieve coexistence conditions characterised by internal, thermal, mechanical, and chemical equilibrium; as a result each box eventually represents a distinct phase. The two boxes constitute a system in the canonical ensemble where simulations are performed. Hence, thermal equilibrium is set up by definition. Internal equilibrium is achieved by roto-translations, mechanical one by volume moves, and chemical one by particle transfer moves.

It is important to note that Gibbs ensemble simulations are primarily suited for



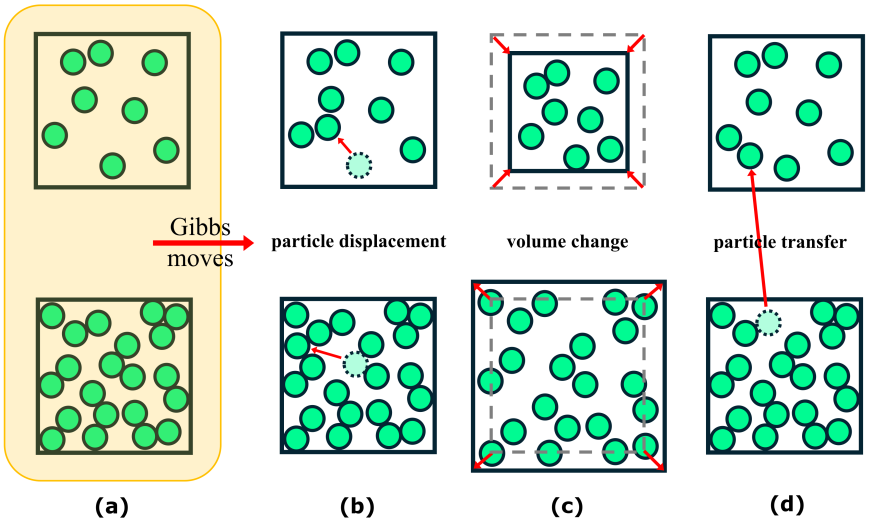


Figure 4.1. Schematic of Gibbs ensemble simulations. (a) Phase coexistence is simulated without a direct interface by employing two virtually interacting boxes. (b) As in standard MC simulations, particles in each box undergo roto-translational moves to achieve internal equilibrium. (c) Volumes change moves of size $\Delta V_{box_1} = -\Delta V_{box_2}$ are attempted to enforce mechanical equilibrium. (d) While the total number of particles in the combined system (yellow selection) is conserved, particle exchanges between the two boxes are allowed to reach chemical equilibrium. In MC Gibbs ensemble simulations, if phase coexistence occurs at a given NVT state point, two initially identical boxes will equilibrate at the characteristic densities of the coexisting phases.

studying vapour–liquid coexistence, which is the aspect of interest in this study. However, for completeness, we signal that their extension to solid phases is fundamentally limited by the nature of the required Monte Carlo moves. First, particle transfer moves become highly inefficient in dense phases. Inserting a particle into a high-density box often results in overlap with existing particles, leading to a zero acceptance probability. Additionally, the removal of a particle from a crystalline solid can distort or destroy the long-range structural order, further destabilising the system. Second, volume exchange moves also suffer from poor acceptance in the presence of solid phases. Shrinking the volume of a dense box can induce particle overlaps, while excessive volume expansion may disrupt the crystal lattice, leading to bond breaking and loss of crystallinity. Even moderate volume fluctuations are generally unfavourable in solids due to their low compressibility, making such moves statistically rare. For all these reasons, phase coexistence involving a solid phase cannot be reliably studied within the Gibbs ensemble framework.

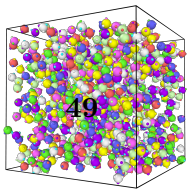
The frequency at which each move characterising the Gibbs ensemble approach is attempted is different since they have a different computation cost. While roto-translation and particle transfer moves involve a single particle at a time, volume

ones require the replacement of all particles of the system. This means that for volume moves, to compute the acceptance probability, the potential energy variations for all particles must be evaluated. This is why we set a frequency of 0.001 for these moves. Particle transfer moves are instead attempted 1 out of 10. Typically, the two boxes are initialised as identical. If the selected state point lies within the coexistence region, their densities will diverge until they reach their respective equilibrium values. Conversely, if densities keep fluctuating around their initial values, the state point is outside the coexistence region. To accelerate equilibration, prior knowledge of the approximate equilibrium densities of the vapour and the liquid phases can be leveraged by initialising each box near its expected equilibrium density. Finally, is worth nothing that, as it happens in thermodynamic integration, also with Gibbs ensemble simulations, coexistence close to critical points cannot be investigated. At these conditions, density fluctuations are important, and they cause the two boxes to frequently change their identity continuously switching from vapour to liquid and never equilibrating.

Having outlined the methodologies used to compute the liquid-vapour coexistence region in mixtures, in the following section we briefly describe its typical characteristics.

4.3 Phase behaviour of mixtures and azeotropy

In systems of two or more components, the pressure-concentration phase diagram exhibits a coexistence region. This is the main difference with respect to one-component systems where there is a single coexisting pressure for each temperature, *i.e.* there is a coexistence line. The presence of a coexistence region implies that the relative concentrations of the vapour and liquid coexisting phases are not the same. Figure 4.2 shows a qualitative pressure-concentration phase diagram for both an ideal (a) and a non-ideal (b) binary mixture. As expected, in both cases, the coexistence pressures reduce to a single value when the concentration of the first component is equal to zero or to one, *i.e.* when the binary mixture becomes a single-component system. However, if the mixture strongly deviates from ideal behaviour, it can exist another point, named azeotropic point [108, 109], at concentration different from zero and one, where the coexistence region reduces to a point. A mixture at the azeotropic point will separate into phases sharing the same azeotropic concentration, therefore behaving as a pure system that preserves the original ratio between components in phase separation.
In the context of distillation, the azeotrope is defined as the point where the fractional distillation process cannot modify the composition of the mixture since the distillate



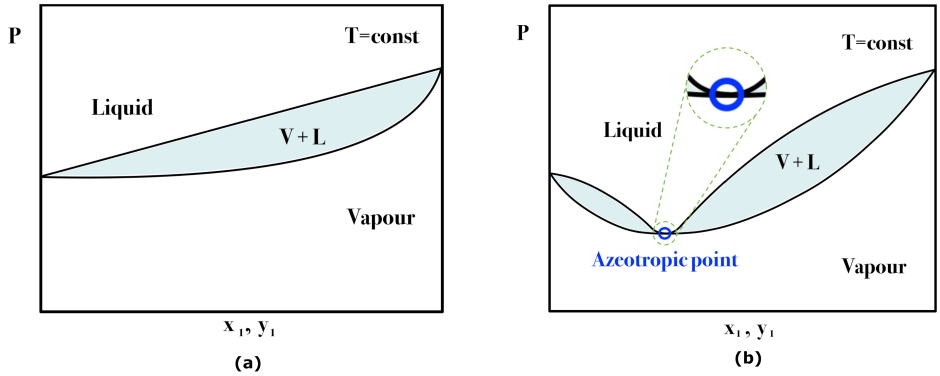


Figure 4.2. Comparison of the pressure-concentration phase diagram for an ideal binary mixture (a) and for a binary mixture with a negative azeotrope (b). (a) In binary mixtures the total vapour pressure varies in between the vapour pressure of the pure components (1 and 2), but its variation depends differently on the concentration of the liquid phase (x) and the vapour phase (y). Hence, unlike single-component systems, where the phase diagram features a coexistence line, mixtures exhibit a coexistence region. (b) Azeotropic mixtures still display a concentration where there is a unique coexisting pressure. This concentration, named azeotropic concentration, is located at the lowest (for a negative azeotrope) pressure at which the liquid and the vapour phases can coexist at the specific temperature T. The inset highlights that at the azeotropic point the bubble point curve and the dew point curve are tangent; the bubble point curve is the locus of points where vapour begin to coexist with the liquid phase as pressure is lowered starting from a point greater than the total vapour pressure, and the dew point curve is where liquid begins to coexist with the vapour as pressure is increased starting from a point in the vapour phase. Image reproduced from [33].

is again a solution with the same composition as the starting one. Indeed, to obtain the distillate, the mixture in a liquid phase is heated until the bubble point curve is reached and then the vapour resulting from the phase separation is collected and condensed. Since phase separation occurring at the azeotrope preserves the ratio between components, the distillate will have the same starting azeotropic concentration. Conversely, by preparing the mixture off the azeotropic concentration, the fractional distillation process can still be exploited, even if with some differences with respect to the ideal case where it allows purifying the more volatile component of the solution. In non-ideal pressure-concentration phase diagrams, the azeotrope can be located at the lowest or at the highest pressure at which the liquid and the vapour phases can coexist at a specific temperature T. The first case is referred to as negative azeotrope and the second one as positive azeotrope. In a negative azeotropic mixture it is possible to purify either the more volatile component or the less volatile one depending on if the system is prepared at a concentration smaller or greater than the azeotropic one. Whereas a positive azeotropic mixture cannot be purified from any of their components because fractional distillation eventually ends in the azeotrope. An example of a mixture having a negative azeotrope is the

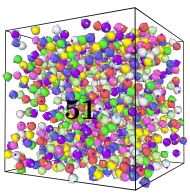
hydrochloric acid and water mixture, whereas for the positive case the most famous example is the ethanol–water mixture.

This discussion highlights how the thermodynamic behaviour becomes significantly more complex when moving from pure substances to mixtures, showing that even the addition of a single component introduces considerable challenges. Hence, in addition to the combinatorial problem, which we introduced in Chapter 2, when using multi-component systems we also face to a thermodynamic problem. According to Gibbs rule of phases [110], each component represents an additional thermodynamic degree of freedom of the system, extending the phase behaviour phenomenology. As a result, the composition of the liquid phase resulting from phase separation is not known a priori, while computing phase diagrams becomes increasingly difficult and time-consuming, and it is typically avoided in system with more three components. No general strategy to tackle the thermodynamic difficulties associated with the use of multi-component systems has so far been presented. We propose to overcome this problem by working with azeotropic mixtures. To this end, we develop design strategies to engineer azeotropicity at a specific concentration, as discussed in details in Chapter 7. Moreover, we anticipate that thermodynamic programming will prove crucial in self-assembly. Indeed, the transition from a dilute phase to a crystalline one can follow multiple nucleation pathways, including non-classical ones, adding complexity to the self-assembly process. A complex thermodynamic behaviour can interfere with the self-assembly pathway and, as argued and demonstrated in Chapter 5, self-assembly can be optimised by selecting nucleation pathways through azeotropic points.

4.4 Phase diagram of the N2c8 SAT-designed binary mixture

In this section, we present and analyse the phase behaviour of the N2c8 binary mixture, SAT-designed to self-assemble a cubic diamond crystal. We compute its phase diagrams both theoretically and computationally, using the techniques previously described in this chapter. We start by determining the binodal curve in the pressure-concentration and density-concentration phase diagrams through Wertheim's theory and numerical integration. Under conditions of constant pressure and temperature, the equilibrium state minimises the Gibbs free energy per particle g. g is the Laplace transform of the Helmholtz free energy per particle f, of which the Wertheim theory provides an analytical expression (Eq. 3.4), and it is defined as

$$g = \frac{P}{\rho} + f \tag{4.5}$$



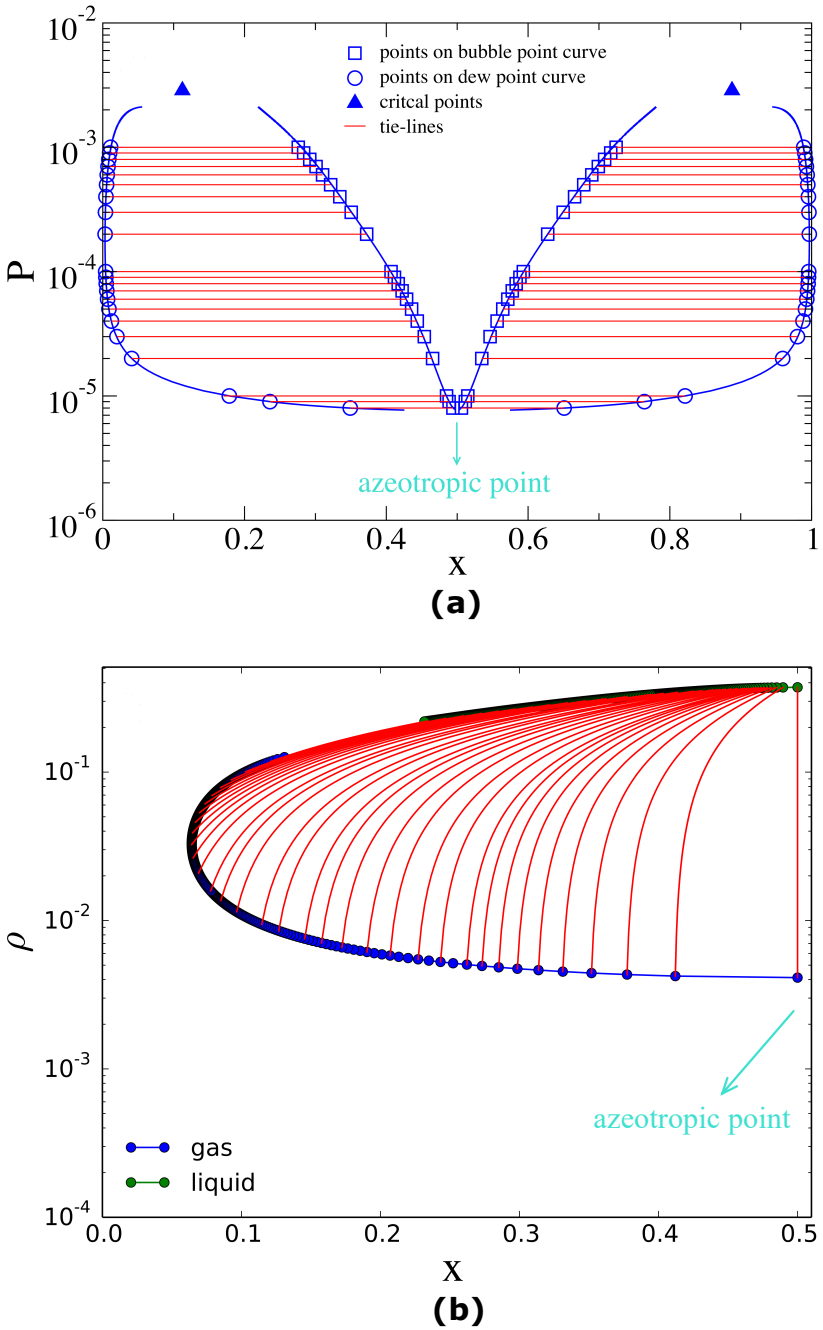


Figure 4.3. Wertheim pressure-concentration (a) and density-concentration (b) phase diagrams of the N2c8 solution. The (a) phase diagram is computed at temperature $T = 0.07$ while the (b) one at temperature $T = 0.08$, and x is the concentration of the first species. Symbols, connected by red tie-lines, represent the coexistence points obtained from the common tangent construction on the Gibbs free energy curve. In (b) tie-lines are not straight since the density axis is in logarithmic scale. Blue lines indicate the binodal curve computed by numerically integrating Eq. 4.4. Triangles are at the location of binary critical points. In (a) the coexistence region reduces to a point at $x = 0.5$ and in (b) the only vertical tie-line is the one at $x = 0.5$, confirming the azeotrope to be at equimolar concentration. Only when prepared as a homogeneous mixture with equal amounts of each species, the systems phase separates by preserving the original composition. Image reproduced from [33].

where P is the pressure and ρ is the total number density. Since the same total density can be achieved by mixing species at more than one pair of compositions $x_1 \equiv x$ and $x_2 = 1 - x_1$, firstly, we must minimise g for each fixed composition x with respect to the density ρ. In this way the Gibbs free energy becomes only a function of composition. Then, coexisting phases having the same temperature, pressure and chemical potential can be obtained by searching those points on $g(x)$ that are connected by a common tangent [90]. Finally, starting from a single pair of coexistence points found with the common tangent rule, we use the isochoric thermodynamics equations (Eq. 4.4) to trace the coexistence lines as a function of composition and pressure.

The pressure-concentration phase diagram obtained at $T = 0.07$ is shown in Figure 4.3 (a). Temperature is in units of ϵ/k_B, where ϵ is the potential energy gain from bond formation. Throughout the text, we consistently use these dimensionless units also expressing pressure in units of ϵ/σ^3 and density in units of $1/\sigma^3$, where σ is the diameter of patchy particles. In Figure 4.3 (b) we plot the coexistence region in the density-concentration phase diagram, at $T = 0.07$. From both, it is apparent the

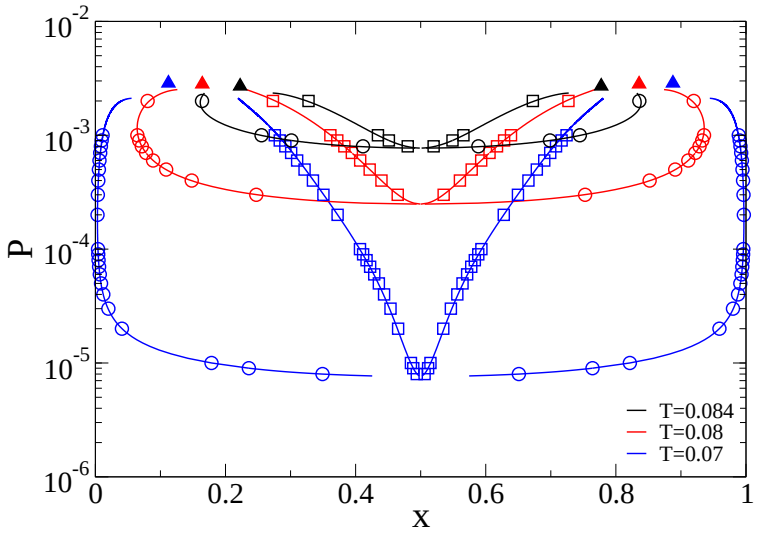
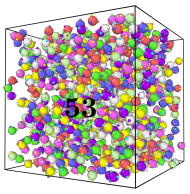
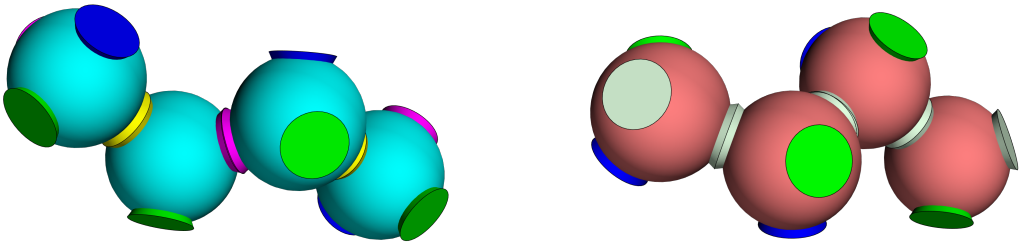


Figure 4.4. Wertheim pressure-concentration phase diagrams of the N2c8 solution at temperatures T = 0.07, T = 0.08 and T = 0.084. Circles represent points belonging to the dew point curve. Squares represent points belonging to the bubble point curve. Triangles indicate critical points. The binodal curves are computed by numerically integrating Eq. 4.4. With increasing temperature, the coexistence region shrinks: the azeotrope stays constant at $x = 0.5$, whereas the critical points shift towards the azeotropic concentration. Image reproduced from [33].



azeotropic nature of the N2c8 solution when the two species are mixed at equal ratio. In the pressure-concentration phase diagram it is exactly at $x = 0.5$ that the bubble point curve (where the vapour phase first appears in the liquid one when pressure is lowered starting from a point greater than the total vapour pressure [108]) and the dew point curve (where the liquid phase first originates in the vapour one when pressure in increased starting from a point in the vapour phase) are tangent and the coexistence region reduces to a point. In the density-concentration phase diagram, azeotropy at equimolar concentration is evident from the slope of the tie-lines, *i.e.* the lines connecting each pair of coexisting point, depicted in red in Figure 4.3. Only at $x = 0.5$ the tie-line is a vertical line connecting a high density and a low density phases at the same concentration. Since the azeotrope is at the lower extreme in the pressure-concentration phase diagram, the N2c8 binary mixture is a negative azeotropic binary mixture [108, 109].
The common tangent construction and the numerical integration of Eq. 4.4 are repeated for temperatures $T = 0.08$ and $T = 0.084$, and in Figure 4.4 we plot together the three resulting pressure-concentration phase diagrams. This makes evident the behaviour of the coexistence region as temperature varies. The N2c8 binary mixture belongs to distinctive type of binary phase diagram, in which the pure components ($x = 0$ and $x = 1$) do not have a liquid-vapour transition but their mixtures do. Indeed, Figure 4.4 shows that by decreasing temperature the coexistence region becomes larger without ever crossing the limit concentrations $x = 0$ and $x = 1$. This unconventional behaviour is intrinsic to the N2c8 design. It originates by the fact that patchy particles of the same species can bind to each other with no more than two bonds, as encoded in the interaction matrix (the one in Figure 2.5). Hence, at pure component conditions, particles can aggregate only into chains as depicted in Figure 4.5. Therefore, even if particles have four patches, when $x = 0$ or $x = 1$ they behave like bifunctional particles that have no liquid-vapour phase separation [91]. We note that the idea that systems with two-patches have a hidden critical point at $P = 0$ and $T = 0$ has been recently revisited in Ref. [111], and generalised to coloured patches in Ref. [112].

Going beyond Wertheim's theory, we study the phase behaviour of the N2c8 mixture computationally, *via* Monte Carlo simulations with AVB moves in the Gibbs ensemble introduced previously in this chapter. Simulations are performed at different temperatures ($T = 0.1, T = 0.097, T = 0.09, T = 0.08$) and, for each temperature, at different averaged (over the two boxes) densities and concentrations in order to compute the binodal curve in the density-concentration phase diagram. Results are shown in Figure 4.6 (a). System size is fixed at $N = 1000$ particles for temperatures $T = 0.097$ and $T = 0.1$ and at $N = 500$ particles for temperatures $T = 0.08$ and $T = 0.09$. Moreover, while for the two higher temperatures simulations

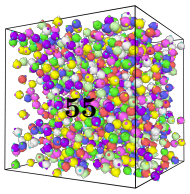


(a) (b)

Figure 4.5. In single-component systems obeying the N2c8 design, only chain aggregates can form. When the SAT-designed N2c8 binary mixture is reduced to a single-component system, composed solely of patchy particles of either first species (a) or second species (b), the particles can aggregate only to forming chains, effectively behaving as patchy particles with a valence of two.

are averaged over at least 8 independent runs, for the lowest ones we rely on single trajectories. This is because equilibration of patchy particle systems at low temperatures is particularly challenging [101]. This is reflected in the non-negligible error bars in Figure 4.6; however we argue that they are more significant than size-effect errors. The location of the binary critical point can not be estimated with Gibbs ensemble simulations [113]. Approximately it is located at the intersection of the binodal curve and the rectilinear diameter line (*i.e.* the line connecting the midpoint of the different tie-lines), which, in Figure 4.6, is represented for each temperature as a dashed line. Coexistence regions are plotted only for $x < 0.5$ since they are symmetric with respect to the azeotropic line $x = 0.5$. The binodal curve at each temperature is split in a liquid branch (square symbols), at high densities, and a vapour branch (circle symbols), at low densities. The trend of the binodal curves as well as the topology (as it is sometime called the shape behaviour of the coexistence region [108, 109]) of the density-concentration phase diagrams are the same as the Wertheim theoretical ones, as shown in Figure 4.6 (b). Moreover, as already noticed in pressure-concentration phase diagrams, the coexistence region shrinks by increasing temperature. Finally, we notice that although Monte Carlo simulations confirm the phase diagram topology and the presence of an azeotrope at concentration $x = 0.5$, Wertheim's theory tends to overestimate the size of the coexistence region [114]. In detail, the bubble point curves align almost perfectly, while Wertheim's prediction places the dew point curves at higher densities than the simulated ones.

The phase diagram analysis concludes with the temperature-density phase diagram at azeotropic concentration plotted in Figure 4.7 from the knowledge of the coexisting densities at different temperatures. As expected, it shows the peculiar signature of limited valence colloids: a coexistence region that does not extend up to high



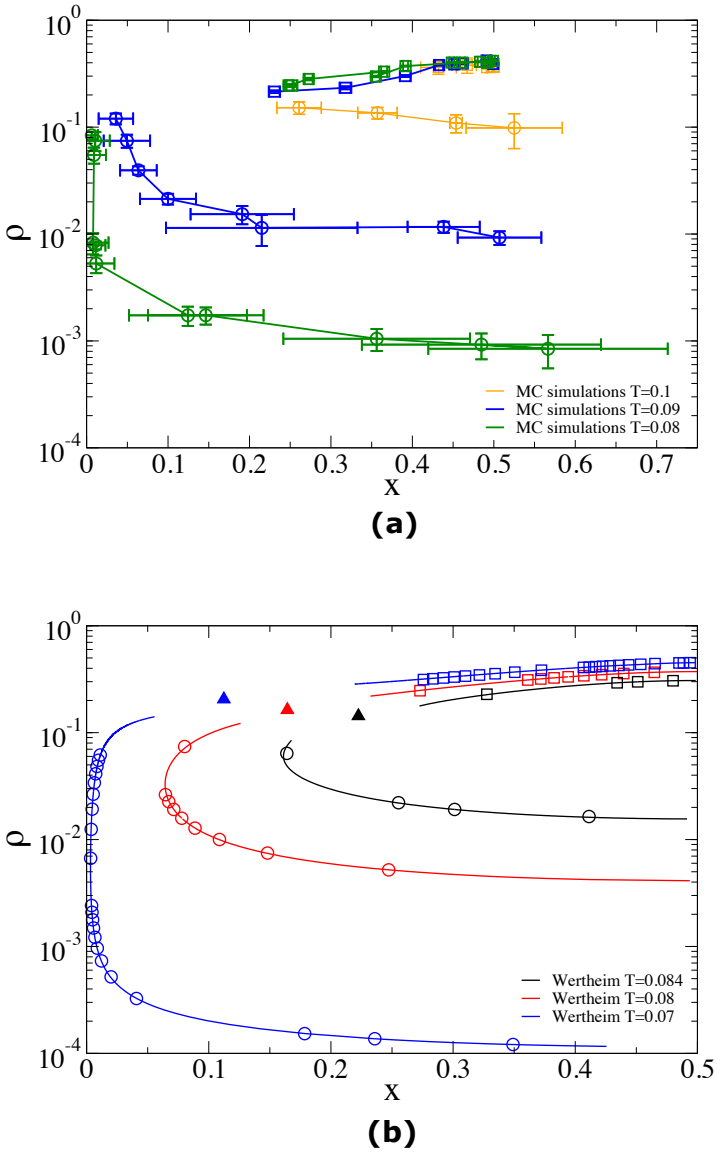


Figure 4.6. Comparison of numerical and theoretical density-concentration phase diagrams of the N2c8 solution. Binodal curves obtained from Monte Carlo simulations (a) and within the Wertheim first order perturbation theory (b) are plotted at different temperatures and they show good qualitative agreement. Circles represent points belonging to the dew point curve, while squares represent points belonging to the bubble point curve. The location of critical points is approximately indicated by triangles in (b). Image reproduced from [33].

densities, as it happens instead for systems characterised by an isotropic interaction potential [115, 116]. This is extremely important since it allows for the formation of equilibrium colloidal gels [114, 117], which do not age in time and do not require a kinetic arrest of the phase separation process to originate [118, 119]. This is because if there is a density interval where it is possible to cool down the system without

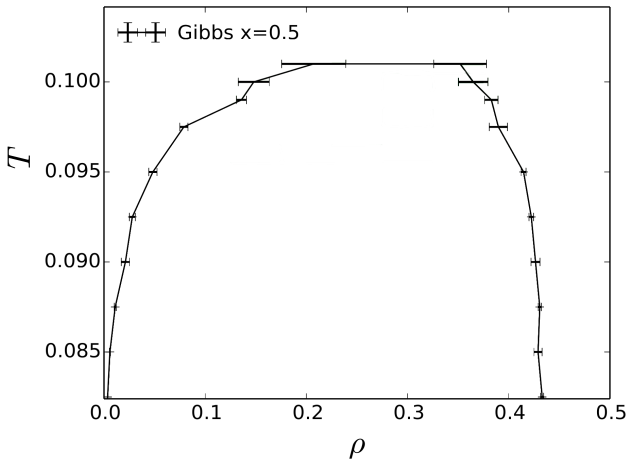
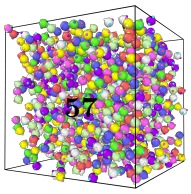


Figure 4.7. Temperature-density phase diagram of the N2c8 solution at azeotropic concentration. The black binodal curve, obtained with Monte Carlo simulations in the Gibbs ensemble, highlights the characteristic behaviour of reduced valence colloids: liquid-vapour phase coexistence occurs over a narrower density range compared to colloids with isotropic interactions.

encountering phase separation, particles are able to originate a percolating cluster where bond lifetime is much larger than the observation times. This is also a crucial example showing that certain thermodynamic conditions are pivotal for functionality. In the next chapter we will analyse the role of the azeotrope in self-assembly.

In conclusion, we computed the liquid-vapour coexistence regions of the SAT-designed N2c8 binary mixture both theoretically, by numerically integrating first order differential equations from isochoric thermodynamics and Wertheim's perturbation theory, and computationally, by performing Monte Carlo simulations in the Gibbs ensemble. The thermodynamic behaviour of the N2c8 binary mixture is summarised in Figure 4.8. The mixture has a re-entrant binary critical point line, which goes to zero temperature and pressure as the pure components are approached. We explained that this characteristic is immediately apparent from the colour design, where it can be seen than each species can form only two intra-species bonds, so when the system is made of only one component it can just form linear chains, and thus does not have a critical point at any finite temperature [120].The non-ideal nature of the mixture manifests itself in a line of azeotropic points at concentration equal to $x = 0.5$. Along such line the system always demixes at the same composition, *i.e.* the liquid and the vapour phases have an equimolar composition of both species. As we will see in the next chapter, this is particularly relevant for the nucleation



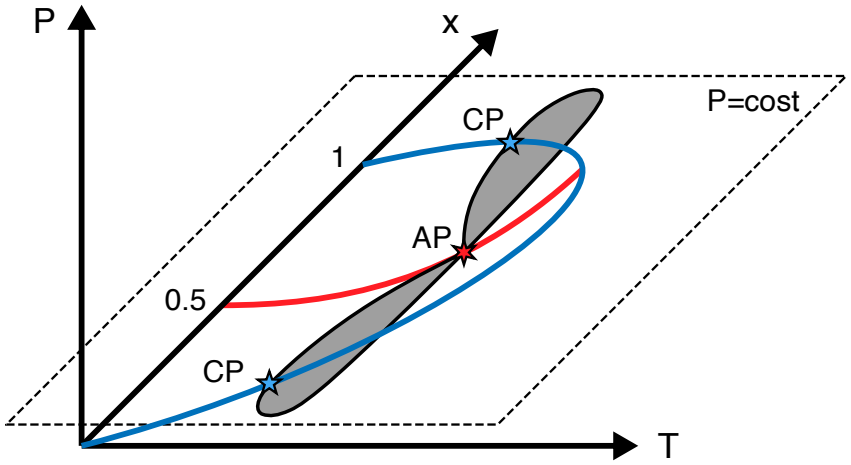


Figure 4.8. Schematic of pressure-temperature-concentration phase diagram of the N2c8 solution. The topology of the PTx phase diagram is equivalent to that of an ordinary azeotropic binary mixture, but in which the binary critical point line (the blue line) goes to $(P, T) \to 0$ for $x \to (0, 1)$. The red line, in the plane $x = 0.5$, represents the line of azeotropic points. The shaded area represents the coexistence region on a isobaric surface. Image reproduced from [55].

process, as the target cubic diamond crystal also has an equimolar composition.

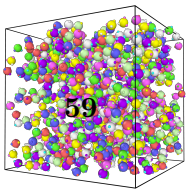
Chapter 5

Studying nucleation of multi-component systems

In this chapter we focus on studying the nucleation of the SAT-designed N2c8 binary mixture. We verify it is an effectively working design able to exclusively self-assemble into a cubic diamond crystal without defects and stacking faults, thus validating the SAT-assembly framework. Additionally, we deepen our understanding of nucleation process in mixtures. Nucleation in systems with a metastable liquid–vapour critical point is the prototypical example of a two-step nucleation process in which the appearance of the critical nucleus is preceded by the formation of a liquid-like density fluctuation. So far, the majority of studies on colloidal and protein crystallisation have focused on one-component systems, and we are lacking a clear description of two-step nucleation processes in multi-component systems, where critical fluctuations involve coupled density and compositional inhomogeneities. By combining Gibbs ensemble simulations and direct nucleation simulations over a wide range of thermodynamic conditions, we are able to pin down the role of the liquid–vapour metastable phase diagram on the nucleation process. The role of azeotropic points is also investigated. Given that the N2c8 solution is a multi-component system and also exhibit azeotropy, it represents an excellent case study for exploring these fundamental points. This chapter draws upon the findings published in papers *C. Beneduce et al., J. Chem. Phys., 2023* [55] and *C. Beneduce et al., ACS Nano, 2023* [33].

5.1 Order parameters

We run Monte Carlo simulations in the canonical ensemble for different state points and multiple replicas looking for nucleating trajectories. Simulation progression is measured in Monte Carlo (MC) sweeps. Each MC sweep represents a single



iteration of the simulation, during which multiple moves are attempted across the system. Over the simulation we track the potential energy that, representing the number of formed bonds, already give us information about a nucleation event. To support this information, we also track the number of crystalline particles. In order to classify particles to be in a liquid or crystalline phase, we use local bond order analysis [121]. For each particle i the $(2l+1)$ dimensional complex vector $q_{lm}(i)$ is computed as:

$$q_{lm}(i) = \frac{1}{N_b(i)} \sum_{j=1}^{N_b(i)} Y_{lm}(\mathbf{r_{ij}}) \tag{5.1}$$

where j runs over the $N_b(i)$ nearest neighbours of particle i, Y_{lm} are spherical harmonics, $\mathbf{r_{ij}}$ is the vector from particle i to particle j, m is an integer whose values range from $m = -l$ to $m = l$, and l is a free parameter (here set to 12). Two particles are connected if the coherence S_{ij} is larger than a certain threshold t_{ij}. S_{ij} is defined as the scalar product of Q_{lm}, and considering $l = 12$ we have:

$$S_{ij} = \sum_{m=-12}^{12} Q_{12m}(i) Q^*_{12m}(j) \tag{5.2}$$

where

$$Q_{lm}(i) = \frac{1}{N_b(i)} \sum_{k=0}^{N_b(i)} q_{lm}(k) \tag{5.3}$$

By using Q_{lm} we take into account also the contribution from the second shell of neighbours around particle i and not only the first shell as one would do by simply using q_{lm}. We set $t_{ij} = 0.75$, $N_b(i) = 16$ and the maximum distance to consider two particles as neighbours to 2.5. Usually a particle i is labelled as crystalline if its number of connections with its neighbours exceeds a certain value. In this work, we instead evaluate if the total coherence $\sum_j S_{ij}$ is larger than a threshold. The threshold is set by analysing the total coherence distribution of both pure liquid and thermalised perfect crystal configurations as well as mixed configurations taken from numerical simulations. In Fig 5.1 (a) we plot the distributions of the pure liquid and crystal phases. From these, we observe that any threshold approximately in between 13.5 and 15.5 would provide a reasonable distinction between the two phases. However, our primary interest lies in identifying the onset and growth of a

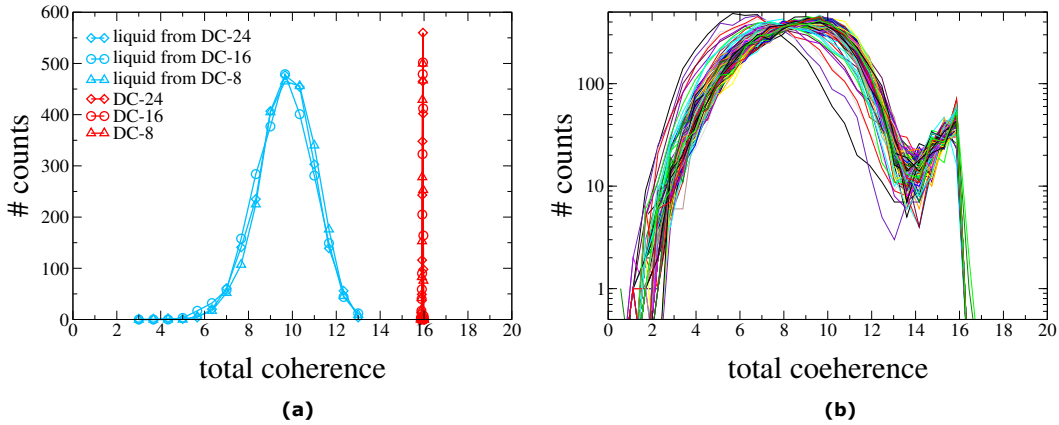


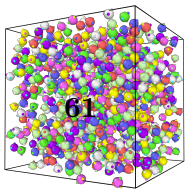
Figure 5.1. Total coherence distributions. (a) Distributions for pure liquid and pure crystalline configurations show distinct, and well separated peaks at different total coherence values (b). In mixed configurations, where a crystalline nucleus is embedded within a liquid, the distribution is broader, through the two picks remain discernible. A suitable threshold for distinguish phases using total coherence is identified to be at value equal to 14.2.

crystalline nucleus within a liquid phase, rather than distinguishing between fully liquid or fully crystalline configurations. For this reason, we must also consider the total coherence distribution of mixed phase configurations, shown in Fig 5.1 (b). As expected, these distributions still exhibit signatures of both liquid and crystal peaks, but they are significantly broader, and the viable threshold range is much narrower; 14.2 appears to be a suitable choice.

By tracking the evolution of this order parameter over MC sweeps, we can immediately identify nucleating trajectories. However, we also require additional information about the structural nature of the assemblies. In particular, since we want to find a solution that is able to exclusively self-assemble the cubic diamond crystal, we must be able to measure the fraction of particles in the cubic diamond phase and the one in the hexagonal diamond phase. We check the nucleating outcomes by means of additional order parameters always defined starting from spherical harmonics. The first one is

$$Q_l(i) = \sqrt{\frac{4\pi}{2l+1} \sum_{m=-l}^{l} |Q_{lm}(i)|^2} \tag{5.4}$$

while the second is



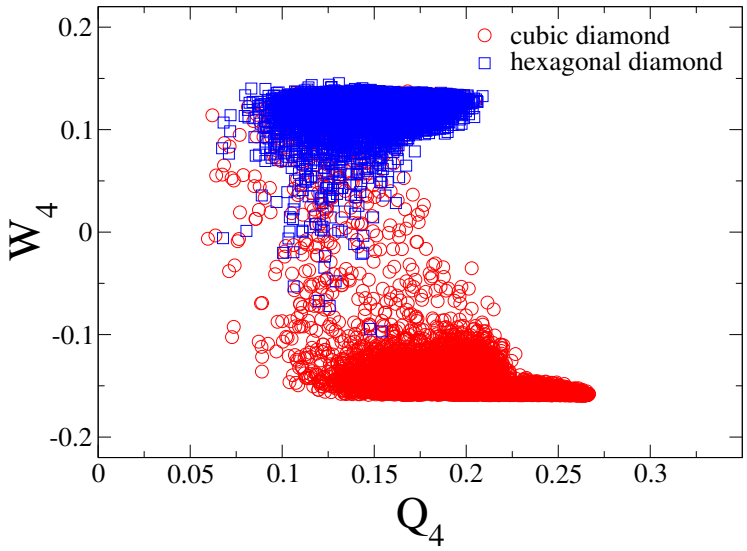


Figure 5.2. Averaged local bond order parameters. Each point represents the Q_4 and W_4 values of a particle from a given structure: blue squares correspond to the hexagonal diamond lattice, and red circles to the cubic diamond lattice. The dataset contains on the order of 10^4 points. The two distinct clusters demonstrate that Q_4 and W_4 together effectively distinguish between the two structures. The few points slightly deviating from the main clusters arise because the crystal structures are taken from nucleating simulations, where some particles may deviate from a perfect lattice arrangement.

$$W_l(i) = \frac{\sum_{m_1+m_2+m_3=0} \begin{pmatrix} l & l & l \\ m_1 & m_2 & m_3 \end{pmatrix} Q_{lm_1}(i) Q_{lm_2}(i) Q_{lm_3}(i)}{\left(\sum_{m=-l}^{l} |Q_{lm}(i)|^2 \right)^{\frac{3}{2}}} \tag{5.5}$$

where the term in parentheses is the Wigner 3-j symbol [122], m_1, m_2, m_3 are integers varying between $-l$ and l under the constraint that $m_1 + m_2 + m_3 = 0$, and l is chosen based on the specific structures we aim to detect; for our purpose, we set $l = 4$. We use both Q_4 and W_4, since only in a 2D plot of $Q_4 \sim W_4$ we can clusterise the two lattices, as illustrated in Fig 5.2. For further details we refer to [123].

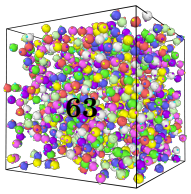
We are now prepared to analyse and understand simulation outcomes, gaining valuable insights into the self-assembly process of multi-component systems.

5.2 Density fluctuations and two step nucleation

In Section 3.1, we showed that DNA origami, of which patchy particles are a valid coarse-grained model, once properly designed, are able to self-assemble into ordered crystalline lattices even starting from quite dilute conditions [48, 73]

suggesting the exploitation of complex nucleation pathways that are still not fully understood. One such pathway is the so-called two-step nucleation pathway, in which a metastable vapour-liquid transition can spontaneously generate local regions with high particles' concentration, thus greatly enhancing the nucleation rate [124–135]. In the last decade, this process has been observed in simple systems that are either one-component systems, or ionic compounds: examples include hard particle fluids [136], colloids [129, 137, 138], salt solutions [139], calcium carbonate [140, 141]. Importantly, two-step processes play an important role in the crystallisation of globular proteins [137,142–145]. The formation of amyloid fibrils, which is implicated both in functional biological roles and in the emergence of diseases such as Alzheimer's disease and diabetes, arises from protein self-assembly via non-classical nucleation pathways [146, 147]. In these cases, fibrilisation is not the result of a single transition directly from the initial to the final state, but rather a multi-stage process. For colloidal particles, globular proteins and non-globular proteins, the short-range nature of the inter-particle interaction generates a metastable vapour-liquid coexistence which provides a mechanism for locally enhancing the density and triggering crystal formation. Spinodal decomposition and/or liquid-phase nucleation are equally active mechanisms. In addition, ten Wolde and Frenkel [137] have called attention on the possibility to exploit critical fluctuations as a mechanism for locally increasing density. In all these cases, a two-step nucleation process takes place, the first step being the local increase of density, and the second the formation of a crystal nucleus in one or several of such enhanced density regions.

We explore two-step nucleation mechanism in multi-component systems by analysing the N2c8 solution. We focus on this binary mixture since it is a meaningful example and since from Chapter 4 we already know its phase behaviour, but we highlight that our results are general. We perform MC simulations with AVB moves over a broad region of the phase diagram, looking for nucleation trajectories. Traditional molecular dynamics (MD) simulations are indeed impractical for extensively studying nucleation due to the slowness of particle diffusion. However, since the kinetic contribution to the nucleation rate is usually negligible compared to the thermodynamic contribution (*i.e.* the nucleation barrier) MC simulations provide a valid route for exploring nucleation behaviour. In Figure 5.3, we plot the results at temperature $T = 0.1$. Specifically, we run simulations for 500 particles over a density concentration grid $\rho \in [0.15, 0.4]$ with $\Delta\rho = 0.05$, and $x \in [0.15, 0.5]$ with $\Delta x = 0.05$. We analyse up to 10 (or 18 for state points where nucleation is more probable) independent runs for more than $3 \cdot 10^8$ steps each, and measure the fraction of trajectories that have nucleated, which is a quantity proportional to the nucleation rate. We define a nucleation event as a run in which at least 20% of the particles are found in a crystalline state. The fraction of particles in the cubic diamond phase are identified



with local bond-order analysis [121], as described in the previous section.

In Figure 5.3 this fraction is proportional to the colour intensity of a tile centred around the state point. To look at the interplay between phase separation and nucleation, we superimpose this colour map to the density-concentration phase diagram, already computed via Gibbs ensemble simulations in Chapter 4 (Figure 4.6(a)). It is apparent that nucleation occurs in correspondence of the vapour-liquid coexistence region, with the probability of nucleation being approximately the highest along the liquid binodal, signalising a two-step nucleation process. We note in fact that no nucleation events are observed at low concentrations, from $x = 0.15$ to $x = 0.25$, *i.e.* outside the phase coexistence region. To confirm the results of Figure 5.3, supporting the observations made for one-component systems, *i.e.* that density fluctuations promote nucleation [129, 133, 137, 148], we run additional simulations in a density-concentration grid at a lower temperature, $T = 0.097$, which makes the coexistence region wider. The obtained nucleation plot is presented in Figure 5.4. Compared to the case at $T = 0.1$, nucleation events occur over a much wider region

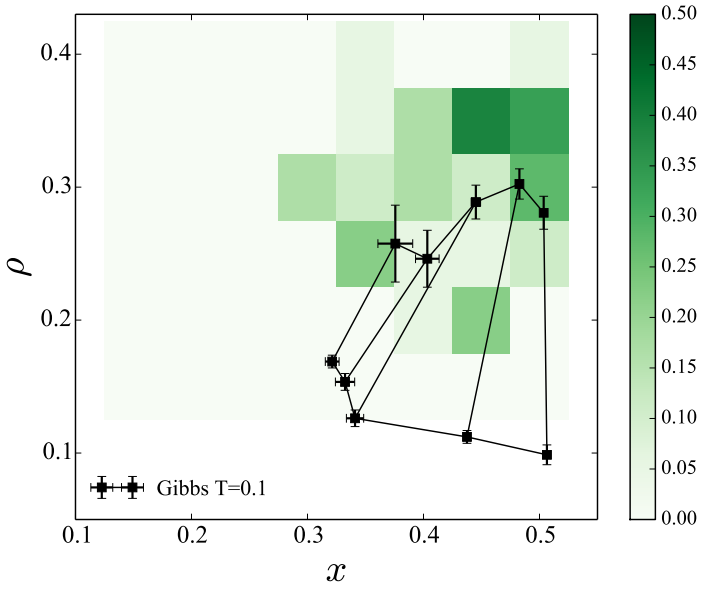


Figure 5.3. Nucleation plot for temperature $T = 0.1$. The binodal black curve obtained via Monte Carlo simulations in the Gibbs ensemble is superimposed to a green colour map reporting, for each considered state point, the ratio between the number of trajectories that nucleate over the total number of trajectories analysed. For each state point several Monte Carlo simulations with AVB dynamics in the NVT ensemble are performed, and we consider as successfully assembled the runs exhibiting a fraction of particles in the cubic diamond phase greater or equal to 0.2. Nucleation occurs in correspondence of the coexistence region, with higher probability along the liquid branch. Image reproduced from [55].

of the phase diagram, following the widening of the coexistence region, as expected. Size effects are also considered as the system size is increased to 1000 particles in this second analysis.

Visual inspection of the nucleation trajectories confirm that the nucleation event is preceded by the formation of a dense liquid phase, mostly via spinodal decomposition (*i.e.* liquid drops do not nucleate, but instead a coarsening liquid networks appears at the start of the simulations). In Figure 5.5, we show, as an example, the evolution of a nucleating trajectory at temperature $T = 0.097$, density $\rho = 0.3$, and azeotropic concentration. Specifically, in panel (a) we plot the potential energy per particle as a function of MC sweeps. It displays a two-step drop: initially the system is prepared in a metastable gas phase, which then forms a percolating liquid phase; in the second step (from $1.9 \cdot 10^8$ MC sweeps) nucleation occurs within the liquid drop and a cubic diamond crystal replaces all the liquid phase, finally coexisting with a low-density gas. In panel (b) we report the snapshot of a configuration from the final part of the first stage, where a liquid-vapour coexistence is established in the box.

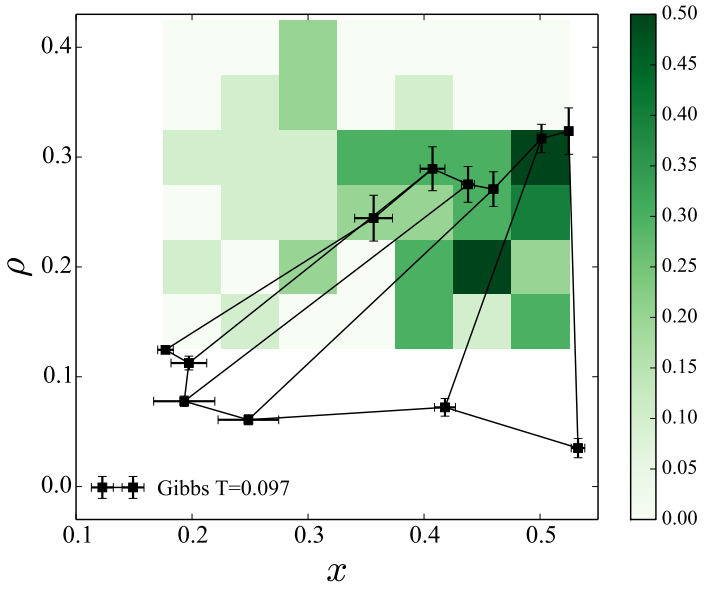
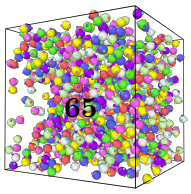


Figure 5.4. Nucleation plot for temperature $T = 0.097$**.** The binodal black curve obtained via Monte Carlo simulations in the Gibbs ensemble is superimposed to a green colour map reporting, for each considered state point, the ratio between the number of trajectories that nucleate over the total number of trajectories analysed. For each state point several Monte Carlo simulations with AVB dynamics in the NVT ensemble are performed, and we considered as successfully assembled the runs exhibiting a fraction of particles in the cubic diamond phase greater or equal to 0.2. Nucleation occurs in correspondence of the coexistence region, with higher probability along the liquid branch. Image reproduced from [55].



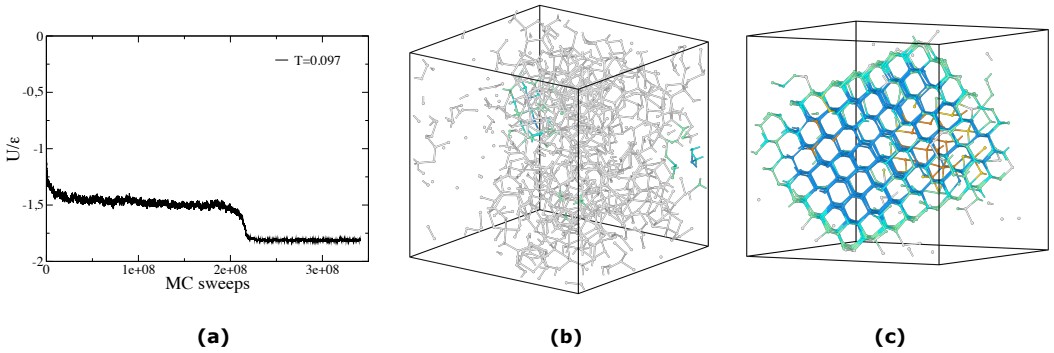


Figure 5.5. Overview of Monte Carlo simulations Plots are obtained from MC simulations with AVB moves in the NVT ensemble run at $T = 0.097$, $x = 0.5$ and $\rho = 0.3$. (a) Potential energy per particle as a function of MC sweeps: the energy drop signals nucleation event. (b) Snapshot of a metastable configuration before the energy drop showing that we deal with a two-step nucleation since crystalline nucleus arise in the denser liquid phase. (c) Snapshot of the last configuration showing that the N2c8 binary mixture eventually self-assembles into a cubic diamond crystal. Neither patches nor the two patchy particles species are shown, while blue/green colours indicate particles belonging to the cubic diamond phase. Snapshots and crystal identification are obtained with the Ovito software [149]. Image reproduced from [55].

Finally, panel (c) illustrates the last configuration of the trajectory, confirming the formation of a large crystal of the diamond cubic phase. We underline the absence of the hexagonal diamond phase and stacking fault defects. While notable, this result is not surprising, as the incompatibility with the hexagonal diamond structure is an inherent property encoded in the interaction matrix of the N2c8 mixture. Therefore, we can conclude that this design effectively targets the desired cubic diamond crystal, and that the SAT-assembly framework [9] provides a valid and robust framework for the inverse design of mixtures.

We have thus identified the two-step nucleation process as an effective pathway for promoting nucleation. Density fluctuations enhance the probability of overcoming the diffusional barriers by generating regions of higher local density. In the following section, we examine the role of composition fluctuations that are intrinsic of multi-component systems. We argue that they play a crucial role in self-assembly process, as we expect the formation of liquid like phase at compositions deviating from that of the target unit cell to hinder and slow down crystal growth.

5.3 Composition fluctuations and azeotropy

Compared to the one-component case, crystal formation and two-step pathways in multi-component systems have received much less attention. In these systems, composition fluctuations couple with density fluctuations, and it is not a priori

clear if the stoichiometric properties of the spontaneously generated liquid phase are consistent with the corresponding properties of the selected crystal. Even when the components are fully miscible in the solid phase (such as Pd and Ag), it was shown that the competition between demixing and crystallisation has a big impact on the growth process, causing large variations in the radial composition of the

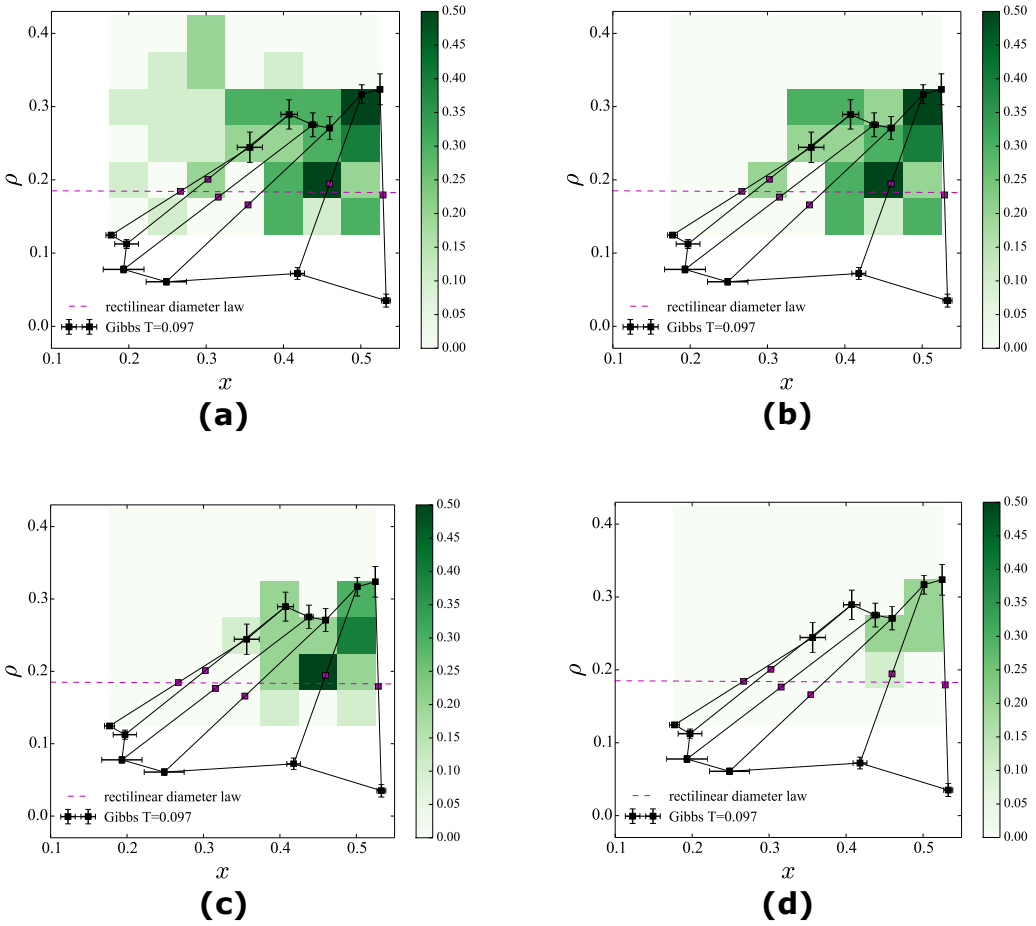
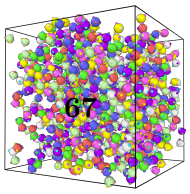


Figure 5.6. Nucleation plots for temperature $T = 0.097$ and different crystalline threshold. The binodal black curve obtained via Monte Carlo simulations in the Gibbs ensemble is superimposed to a green colour map reporting, for each considered state point, the ratio between the number of trajectories that nucleate over the total number of trajectories analysed. For each state point several Monte Carlo simulations with AVB dynamics in the NVT ensemble are performed, and we consider as successfully assembled the runs exhibiting a fraction of particles in the cubic diamond phase (t) greater or equal to 0.2 in (a), 0.5 in (b), 0.7 in (c) and 0.9 in (d). Purple squares represent averages of each pair of coexisting points and, according to the rectilinear diameter law, the best straight line passing through them (the dashed line) provides an indication of where the critical point is. Nucleation occurs for almost each considered state point, but extended cubic diamond lattice ($t \geq 0.7$) can appear only close to the liquid branch and to the azeotropic condition. Moreover, close to critical point crystallisation is not enhanced as it happens instead close to the azeotropic point. Image reproduced from [55].



nuclei [150]. Multi-component systems also display thermodynamic behaviour that is not found in one component systems, such as lines of binary critical points, azeotropic points, and re-entrant condensation. Experimental studies have already shown that phase separation in multi-component systems does not preserve the original ratio between the different species and that the phase behaviour strongly depends on the components' ratio [151–153], but it is yet unclear whether some of the peculiar mixture features can affect crystal nucleation. The results presented in the previous section clarify the role of density fluctuations in nucleation phenomena, explaining why the liquid-vapour metastability region can promote them. We here focus on composition fluctuations.

A target crystal has a fixed stoichiometric composition, while as explained in Section 4.3 and verified by phase diagram computation of the N2c8 solution, mixtures does not preserve the original ratio between the components upon phase separation and the composition of the resulting liquid phase typically differs from that of the target crystal. We illustrate the impact of this mismatch on nucleation efficiency in the four panels of Figure 5.6. In each panel, the density concentration grid of state points is coloured according to the fraction of independent trajectories in which crystalline particles include at least 20% (a), 50% (b), 70% (c), and 90% (d) of

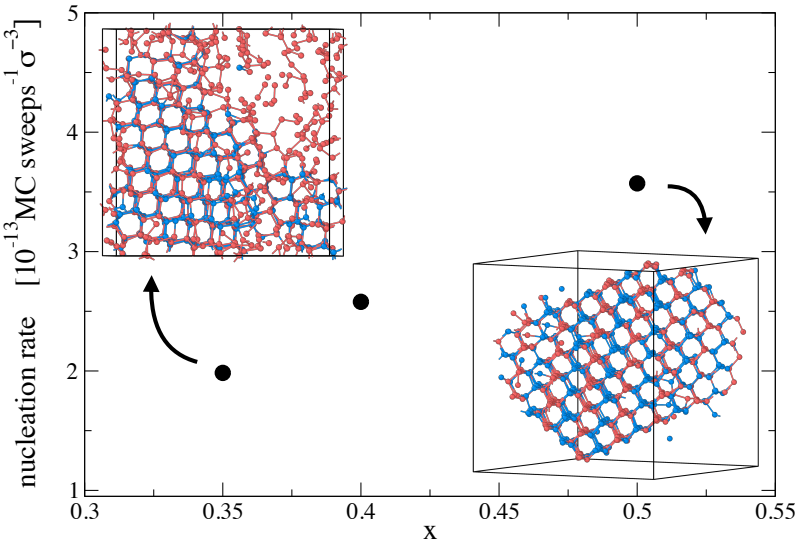


Figure 5.7. Nucleation rate and final configuration snapshots. Nucleation rate (black dots) as a function of concentration for systems of 1000 particles at temperature $T = 0.097$ and density $\rho = 0.3$. The two snapshots display the last configuration of a trajectory at $x = 0.35$ and at $x = 0.5$. Red and blue colours indicate the species to which a particle belongs: blue for the minority component, and red for the majority component. Self-assembly is promoted at the azeotrope. Image reproduced from [33].

the total system at the corresponding state point. We observe that, while small crystals can assemble along the coexistence region, the formation of larger crystals is progressively shifted towards the $x = 0.5$ concentration line, *i.e.* towards the azeotropic point.

Nucleation at the azeotropic point is advantageous as self-assembly can proceed ideally up to 100% without one component depleting before the other and an extended crystal can form. On the contrary, at off-azeotropic conditions, the self-assembled cubic diamond coexists with a gas phase composed of the majority component, that can only aggregate in chains. This behaviour is shown in the snapshots in Figure 5.7 depicting the last configuration of a mixture prepared at the azeotropic condition ($x = 0.5$) and one away from it ($x = 0.35$). A visual inspection of these snapshots highlights that crystal growth is limited when the concentration of the liquid phase is different from the stoichiometric ratio of the target crystal components. Whereas, going toward the azeotropic point, the density of the majority component diminishes until eventually all particles belong to the crystalline phase.

This becomes evident upon examining Figure 5.8, too. It illustrates that for a nucleating trajectory at temperature $T = 0.1$ and azeotropic concentration, nearly all particles transition into the crystalline phase by the end of the process. It also

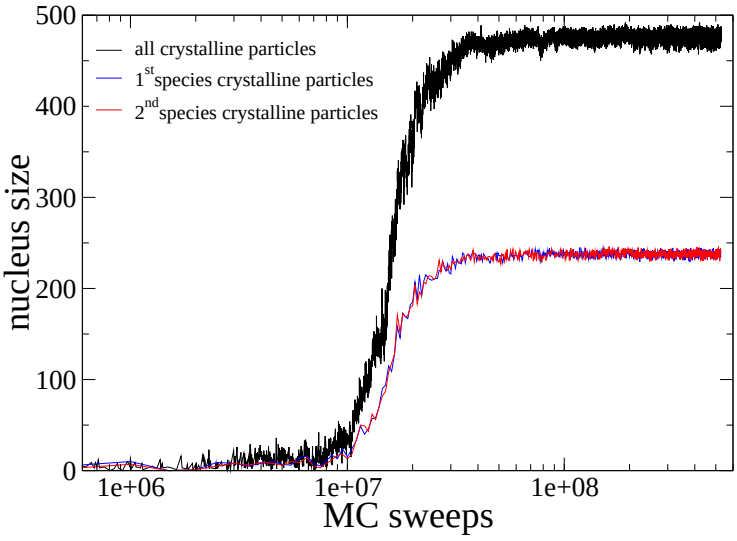
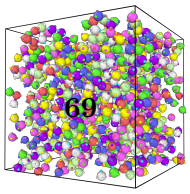


Figure 5.8. Progress of nucleus size for a nucleating trajectory at the azeotropic concentration x = 0.5, at temperature T = 0.1, density $\rho = 0.3$, and number of particles $N = 500$. The blue and the red lines represent the number of particles in the crystalline phase of the first and the second species, respectively. Nearly all particles are crystalline. Image reproduced from [33].



shows that the number of particles in the crystalline phase belonging to the first and second species is identical. These are direct consequences of the compositional match between the liquid and the crystal structure. This implies that we do not need to wait for the concentration of the local environment to match the one of the desired structure, and we can conclude that nucleation pathways along azeotropic points located at a proper concentration can enhance self-assembly. In Figure 5.7, we plot the nucleation rates computed at three different concentrations ($x = 0.35$, $x = 0.4$, and $x = 0.5$) to directly compare azeotropic and off-azeotropic trends. The nucleation rate is estimated as the number of trajectory that successfully nucleate within 3.5×10^8 MC sweeps, per unit of time and volume, via 56 Monte Carlo trajectories ran at temperature $T = 0.097$, density $\rho = 0.3$, and with $N = 1000$ particles. As expected, we observe that the nucleation rate increases towards the azeotropic concentration. Finally, it is worth nothing the quality of the assembled crystals: we observe nuclei free from defects. This confirms that the interaction matrix was successfully SAT-designed to avoid the hexagonal diamond phase, and this also forbids the formation of stacking faults that are the most common type of defects in diamond cubic crystals.

As demonstrated originally in Ref. [137], critical fluctuations are able to trigger crystal nucleation. However, we find here that in a binary mixture, it is instead the azeotropic point which more strongly promotes crystallisation. Indeed, from

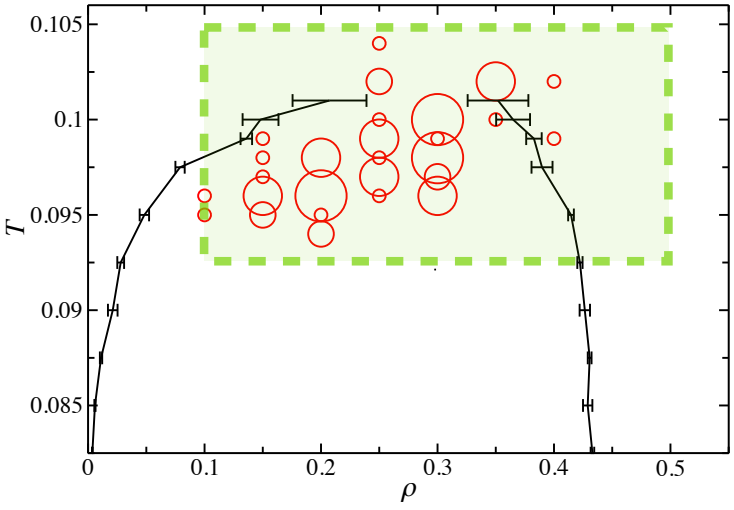
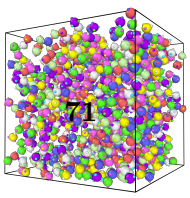


Figure 5.9. Nucleation plot at the azeotrope. $T - \rho$ phase diagram obtained from Gibbs ensemble simulations (black lines). The red circles are drawn in correspondence of the state points which nucleated. The radius of the red circles is proportional to the fraction of runs that successfully assembled within the simulation time of up to 5×10^8 MC sweeps. MC simulations are run over the state points belonging to the green area. We observe nucleation events aided by critical fluctuations. Image reproduced from [33].

Figure 5.6 we observe that nucleation do not occur in correspondence of the critical point that is approximately located at the intersection between the binodal curve with the purple rectilinear diameter line. The rectilinear diameter line is the best straight line that passes through the average of each pair of coexisting points.

We further investigate this apparent discrepancy. In Chapter 4, we observed that the azeotropic point and the critical point coincide when the two corresponding thermodynamic lines meet (Figure 4.8), which occurs at the maximum critical temperature at the azeotropic concentration $x = 0.5$. Hence, we study deeper the self-assembly process through the azeotropic point. In Figure 5.9, we illustrate the nucleation behaviour of the N2c8 design in the temperature-density plane at $x = 0.5$. We prepare disordered configurations at equimolar composition for different state points on a regular grid, with $\rho \in [0.1, 0.5]$ and $\Delta\rho = 0.05$, $T \in [0.920, 0.104]$ and $\Delta T = 0.002$. For each (ρ, T) state point we run 5 independent trajectories in the NVT ensemble with AVB biased moves [100]. The state points considered are enclosed in the green shaded area in Figure 5.9, and each trajectory is run for 5×10^8 MC sweeps or until crystallisation. We mark the successfully nucleating state points with red circles with a radius proportional to the fraction of trajectories that have nucleated. We superimpose (black line) the results from Gibbs ensemble simulations that have been initialised at equimolar conditions (Figure 4.7). Error bars are computed from 10 independent runs at each temperature, and the black lines connecting the points are guides to the eye, helping to identify the gas and liquid branches. We confirmed that once phase separation has occurred, both boxes (liquid and vapour) are still found at equimolar concentration for all temperatures, *i.e.* we are always at azeotropic conditions. From Figure 5.9, it is clear that the nucleation of the cubic diamond crystal occurs again in correspondence of the binodal curve, confirming that self-assembly is aided by the formation of dense liquid regions during the phase-separation process. Interestingly, some state points have nucleated outside the binodal boundaries, but close to the critical temperature. Thus, we also observe nucleation events at supercritical conditions. The system thus shows two-step nucleation aided both by spinodal decomposition at sub-critical temperatures, and by critical fluctuations at super-critical conditions, just as predicted by Ref. [137] for isotropic one-component systems.

To summarise, the self-assembly pathway at the azeotropic point is the following: an equimolar disordered solution first generates equimolar critical fluctuations or first demixes in an equimolar dense liquid, which then crystallises in an equimolar crystalline structure. Self-assembly at azeotropic conditions has the advantage of bypassing the difficulties associated with composition fluctuations, which could otherwise severely limit the nucleation rate. This behaviour confirms that at azeotropic



conditions a multi-component system effectively behaves as a single-component one, and it is a further evidence that azeotropy is an effective strategy to simplify the thermodynamics of mixtures. How to design azeotropic mixtures is a challenge we address in Chapter 7.

In conclusion, in this chapter, we examined the nucleation process of multi-component systems via the case study of the N2c8 binary mixture of patchy particles, SAT-designed to nucleate into a cubic diamond lattice. We verified that the desired crystal is the unique nucleation outcome, and we showed that the strongest enhancement of crystallisation occurs at an azeotropic point with the same stoichiometric composition of the crystal. Indeed, while the two-step nucleation pathway is an effective strategy to overcome diffusional barriers, azeotropic points are the key to lower compositional barriers. This evidence, combined with the understanding that azeotropy can help to overcome the thermodynamic problem of using multi-component systems (Chapter 4), drive us to explore a route for introducing and controlling azeotropy directly through interacting rules. We aim, therefore, to achieve a specific thermodynamic problem by design. This represents a step forward in inverse self-assembly, which has traditionally focused solely on addressing structural challenges, ensuring that the target structure assembles exclusively, while preventing the appearance of competing structures or kinetic traps. We instead demonstrated that the ability to control by design not only the final structure but also the intermediate steps leading to its formation are crucial for optimising self-assembly in mixtures. Before investigating this direction, the next chapter remains focused on nucleation, from a broader perspective. Specifically, we explore Classical Nucleation theory (CNT), which is the most widely used theory for comparing simulated and experimental nucleation results.

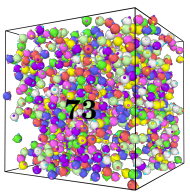
Chapter 6

A falsifiability test for Classical Nucleation Theory

In the previous chapter we examined the nucleation behaviour of the N2c8 SAT-designed mixture. If we aim to compare our computational findings with theoretical predictions, the most widely used framework is Classical nucleation theory (CNT). In this chapter, we focus on evaluating the validity of its core assumption, the capillarity approximation. This assumption posits that the nucleation properties can be inferred from the bulk properties of the melt and the crystal. Indeed, although CNT's simplicity and usefulness cannot be overstated, experiments and simulations regularly uncover significant deviations from its predictions, which are often reconciled through phenomenological extensions of the CNT, fuelling the debate over the general validity of the theory. By leveraging the polymorphs of the N2c8 binary mixture, we present a falsifiability test for any nucleation theory grounded in the capillarity approximation. We focus on cases where the theory predicts no differences in nucleation rates between different crystal polymorphs, and we rank the nucleation frequency of polymorphs that have the same free energy (both bulk and interfacial) across all state points. This chapter builds on the results reported in paper *C. Beneduce et al., PRL, 2025* [56].

6.1 CNT

Citing Baron Peters (who in turn paraphrases Popper [154]): "The most convincing test of a theory comes from special cases where the theory should fail if it is not true" [155]. Here, we show that the N2c8 system is ideally suited to test one of the basic assumptions of Classical Nucleation Theory (CNT), *i.e.* the so-called *capillarity approximation*. The capillarity approximation assumes that the critical nucleus has the same thermodynamic properties (surface tension, density, etc.) of



the bulk phase, which allows for a simple calculation of the free-energy barrier for the phase transition. As detailed in Appendix D, CNT defines the nucleation rate as $Ke^{-\Delta G(n_c)/k_BT}$, where K is the kinetic prefactor that accounts for the attachment rate of particles to the nucleus. $\Delta G(n_c)$ is the free-energy barrier for nucleation and is obtained by maximising $\Delta G(n) = -n|\Delta\mu| + \alpha n^{2/3}\gamma$, *i.e.* the Gibbs free energy cost to form a nucleus of n particles at constant pressure and temperature, where $\Delta\mu$ is the chemical potential difference between the crystal and the melt, γ is the interfacial free energy, and α is a proportionality constant accounting for the shape of the nucleus [156]. Due to its relative simplicity and predictive power, CNT has been perhaps the most used theoretical model to describe nucleation processes. Although it has been successful in many cases [157–160], there are also numerous cases where its predictions do not align with experimental and simulation results [125, 161–163]. In such instances, various phenomenological extensions to the theory [164–178] have been proposed to reconcile these discrepancies without abandoning the capillarity approximation, for example by introducing temperature-dependent interfacial free energies. Usual tests of CNT consists in comparing the measured nucleation rates with theoretical values. The problem with this approach is that the results depend strongly on quantities, notably γ, which are state-dependent and very difficult to measure accurately at conditions where homogeneous nucleation occurs. Citing Oxtoby, "Nucleation theory is one of the few areas of science in which agreement of predicted and measured rates to within several orders of magnitude is considered a major success" [179]. Here we propose a test that sidestep these challenges only requiring to rank the structures based on how frequently nucleation occurs.

6.2 The model system

We present a falsifiability test not only for CNT but also for any nucleation theory grounded in the capillarity approximation. Instead of accurately predicting nucleation rates, we focus on the polymorphic composition of the crystalline phase [180, 181], *i.e.* the ability of a material to exist in more than one crystalline structure [15, 182–187]. We leverage the N2c8 binary mixture that has three polymorphs with identical free energies (both bulk and interfacial) at all state points.

6.2.1 Three isotypic polymorphs

The SAT-assembly algorithm, introduced in Chapter 3, can also be used to list all possible ways to fill the lattice positions of the target cell by particles belonging to a selected design. We can indeed insert additional clauses fixing the patch colouring and the colours interactions variable. In this way the SAT solutions represent all

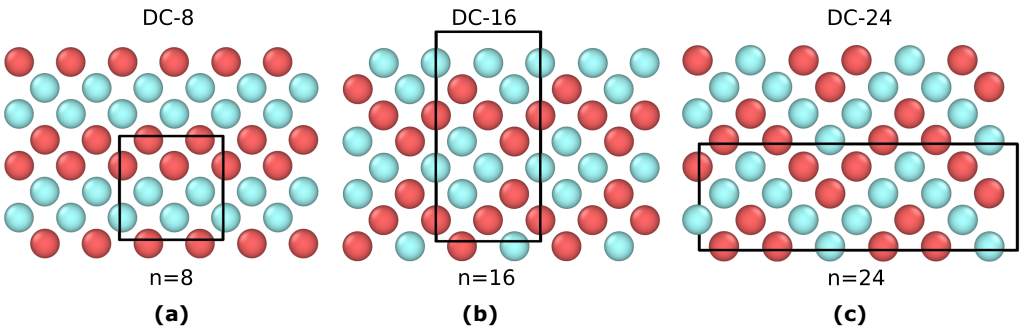
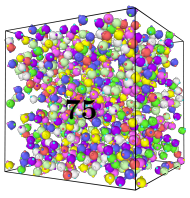


Figure 6.1. Overview of cubic diamond polymorphs. The SAT-assembly framework allows us to enumerate all the possible arrangements of the N2c8 particles in the 48 particle cubic diamond lattice. Three periodic patterns are identified: DC-8 (a), DC-16 (b), and DC-24 (c). They are labelled and displayed in increasing order of unit cell size (black box). Image adapted from [56].

the different possible ways these fixed design patchy particles can be arranged in the target unit cell. This is a great accomplishment since we can automatically identify all the existing polymorphs that, regardless of their nucleation abilities, are compatible with a target lattice of a certain size (defying McCrone's law [1]). These polymorphs are isotypic forms of crystal, *i.e.* they share the same kinetic positions and symmetry of the crystal lattice, but differ in the way the different species are arranged on the lattice sites. Crucially they are fully degenerate structures sharing the same bulk and interfacial free energies. For this reason, they constitute an ideal system to test CNT since, given that there are no free-energy differences between the different polymorphs and that they all form from the same liquid phase, all these structures should have identical nucleation properties, according to the capillarity approximation. Such isotypes whose specificity is necessary to meet the desired requirements to perform our general test, can be generated via SAT-assembly framework for any desired lattice. Here we choose a 48-particle cubic diamond cell populated by particles belonging to the N2c8 design. Within the SAT-assembly framework, by keeping the interaction matrix fixed, we find that there are three possible periodically repeated patterns of N2c8 patchy particles within the lattice positions. All these arrangements are illustrated in Figure 6.1 where we display the [001] plane of the polymorphs. The three structures belong to the P1 space group, meaning that there is no symmetry other than the translational one. Yet, they differ in the unit cell size (black box in Figure 6.1), which we use for labelling: DC-X where $X \in [8, 16, 24]$ refers to the size of the unit cell of each polymorph. We use here the word unit cell to indicate the smallest repeating unit of the two species,

[1] the number of forms known for a given compound is proportional to the time and money spent in research on that compound



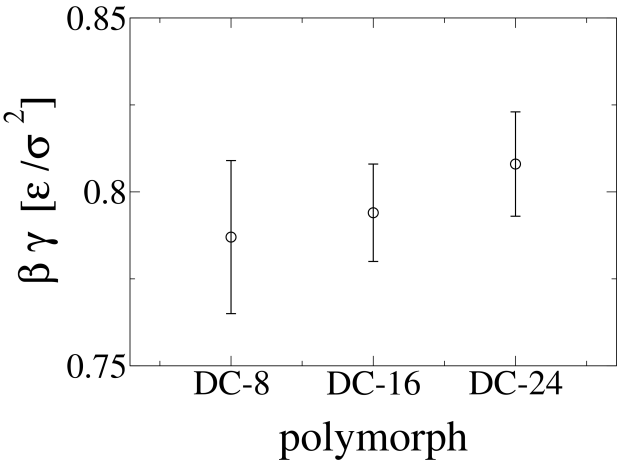


Figure 6.2. Surface tension of the three polymorphs. Each estimate comes from an average of four independent Monte Carlo simulations in the grand canonical ensemble. All values are considered equal within an error of 3%. Image adapted from [56].

regardless of patch colouring, that, when stacked together, creates the crystal lattice, as shown in Figure 6.1. It is worth noting that if the patch arrangement in the unit cell is also considered (not shown in Figure 6.1), the DC-8, DC-16, DC-24 have unit cells of 16, 16, and 48 atoms, respectively. In the following we will indicate the size of the unit cell only referring to the species occupation, regardless of the patch arrangement. Although other polymorphs with larger unit cell sizes exist, they do not appear in our simulations because nucleation involves critical nuclei smaller than 30 particles under the conditions investigated. We speculate that crystals with unit cells larger than the critical nucleus size n_c, which is temperature-dependent, are prevented from nucleating. The three polymorphs investigated here are isotypic, their composition is always equimolar and they have only translational symmetry. In each of the crystals the same number and type of bonds are established, and lattice vibrations are controlled by the geometry of the patches, which is the same for both species of the mixture. Therefore, all polymorphs have the same bulk free energy. Moreover, we verify the equality of their (solid/liquid) interfacial free energies. With successive umbrella sampling simulations [103], as explain in the next paragraph, we have computed the free energy cost of forming an interface between each polymorph and the same liquid phase at coexistence conditions (determined via direct coexistence simulations). The resulting average values of surface free energy are reported in Figure 6.2 for the [100] square plane: the surface tensions of the three polymorphs are the same within the error (approximately 3%). The relevant macroscopic properties on which CNT is based are therefore the same in the DC-8, the DC-16 and the DC-24 structures.

6.2.2 Liquid-solid interfacial tension with BSUS

The surface tensions of the three polymorphs is computed by using the bias successive umbrella sampling technique (BSUS) testing this approach for the first time on liquid-solid phases of a system of patchy particles. Monte Carlo simulations are run in parallel in distinct but overlapping density windows in the grand-canonical ensemble (μVT); the first window consists of a perfect crystal box of $N = 1536$ particles and the successive ones are initialised by removing particles from the crystal, two by two until $N = 1290$, in the direction perpendicular to the selected interface. We choose a window size of three particles. In each window the number of times each possible state ($\{N_{i-1}, N_i, N_{i+1}\}$) is visited is counted in order to eventually reconstruct the probability distribution $log(P(N))$. The density at which the crystal is prepared is the coexisting one $\rho_{coex} = 0.52$, as evaluated by direct coexistence simulations in NPT ensemble at $T = 0.133$ and $P = 0.2$. The activity z is fixed at 2.4 and the number of crystalline particles in each window is also calculated. Its behaviour through the 124 windows is shown in red in Figure 6.3 (a): the coexistence and the solid regions are easily identified. In the coexistence region the number of crystalline particles grows linearly, while at $N = 1300$ it straightly goes to zero indicating the sudden melt of the crystal due to interface fluctuations. A snapshot of a configuration at coexistence is reported in Figure 6.3 (b). Identifying the coexistence region allows us to reweight the probability distribution $log(P(N))$ (green line in Figure 6.3 (a)) in order to obtain the equilibrium one $log(P_{coex}(N))$ (black line in Figure 6.3 (a)) that is plotted in black in Figure 6.3 (a). Indeed, the equilibrium

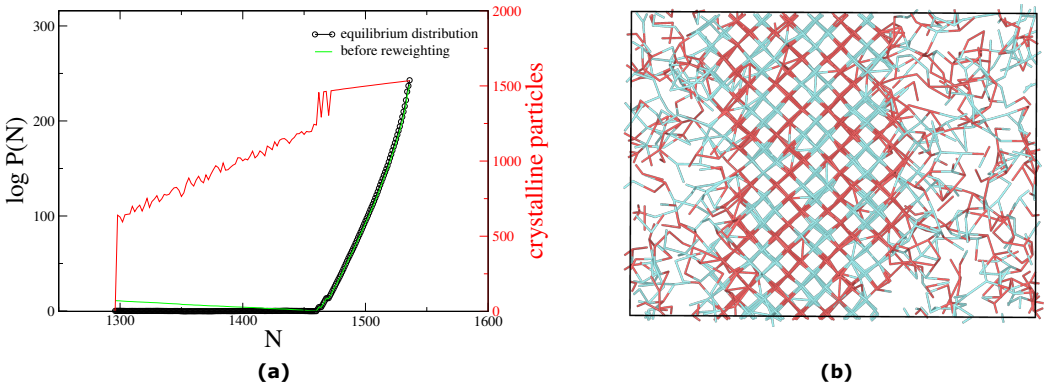
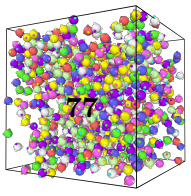


Figure 6.3. Surface tension computation. (a) Not normalised probability distribution function $logP(N)$ from BSUS simulations of DC-16 cubic diamond crystal and liquid coexistence at $T = 0.133$ and $z = 2.25$. The distribution has been reweighted (from green line to black line) to have a flat profile at coexistence, as highlighted by black circles. The number of crystalline particles (red line) emphasises the coexistence and the crystalline regions. (b) Snapshot of a configuration from BSUS window of 1350 particles showing the two coexisting phases; only bonds are represented. Image adapted from [56].



distribution must be flat in the coexisting region since, at coexistence, there is no free energy cost in moving the interface along the direction perpendicular to it. Since $P_{coex}(N) = P(N)e^{\beta(\mu_{coex}-\mu)N}$, in a logarithmic y-axis plot we can straightforwardly determine the coexistence activity $z_{coex} = e^{\beta\mu_{coex}}$ by linear fitting the coexistence region of $log(P(N))$ vs. N. Then we can scale the distribution such that its slope is effectively zero at coexistence. Having now the distribution at equilibrium we compute the interfacial tension by measuring the height of the crystal peak and dividing it by two times the area of the interface. Indeed, as visible in Figure 6.3 (b), the planar interfaces are two since periodic boundary conditions are active. This procedure is repeated few times for each polymorph and the resulting average values of interfacial tension are reported in Figure 6.2. We can assert that they are all the same within an error of 3%, thus confirming that the three polymorphs can be considered equal in their interfacial properties.

6.3 The test

Having characterised the bulk properties of the three polymorphs, we now consider their nucleation properties to perform our test and assess whether the CNT prediction of complete equality holds false. To this end we select two state points, identified in the previous chapter as favourable for crystallisation, and we run extensive Monte Carlo simulations in the canonical ensemble to collect nucleation events. Each simulation trajectory must then be classified to determine whether the system is in the liquid phase or has crystallised, and, in this latter case, to identify the specific polymorph formed. To distinguish between liquid and crystalline particles, we can use local bond order analysis based on spherical harmonics, as explained in the previous chapter. However, this method is not sufficient to differentiate between the three polymorphs. Therefore, we introduce a new order parameter, which is described in detail in the following paragraph and takes into account the bond geometry and the particle orientations.

6.3.1 A new order parameter

Although total coherence allows us effectively label particles as liquid or crystalline, it fails to distinguish between the three polymorphs as their total coherence histograms perfectly overlap. Other methods based on the Steinhardt order parameter [188], such as the one introduced by Frenkel [189] or the one using a two-dimensional analysis, whether averaged [123] or not [190], are equally ineffective. Indeed, all three structures are cubic diamond crystals with the same lattice positions and the same tetrahedral bond geometry. They only differ in the species located

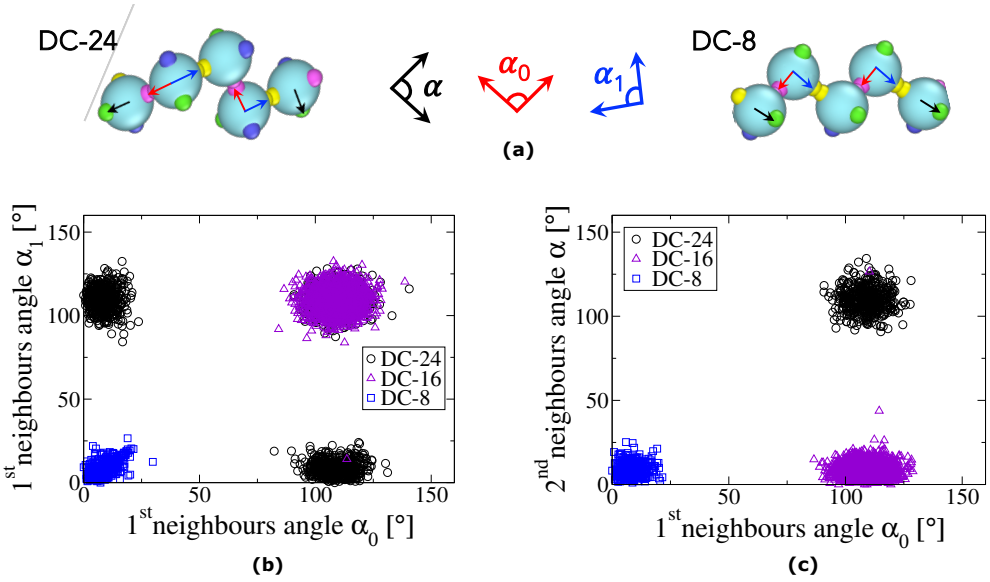
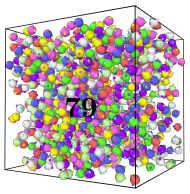


Figure 6.4. Orientational order parameter to distinguish between the three polymorphs. We are able to assign a polymorph label to each particle by analysing the orientation of first (b) and second (c) neighbours of the same species: angle α_0 and α_1 of first neighbours and α of second neighbours are evaluated. (a) DC-8 arrangement of particles of the first species that serves as an example to illustrate the three analysed angles. For perfect crystals, if the first neighbours have the same orientation (α_0=α_1=0), we deal with a DC-8, whereas if α_0=0 and α_1=109 (or α_0=109 and α_1=0) it is a DC-24. If α_0=α_1=109 it could be either a DC-24 and a DC-16 and we need second neighbours angle α: if α=0 it is a DC-16 while if α=109 it is a DC-24. In simulations we do not have perfect crystals so angles widen around the previous values as showed in (b) and (c) using black circles for DC-24, purple triangles for DC-16, blue squares for DC-8 and green circles for liquid. Image adapted from [56].

at each lattice point. Hence, we develop a system specific order parameter that incorporates the species information. By looking at the three perfect crystals we observe that it is possible to classify them as DC-8, DC-16 and DC-24 with a criterion that only relies on the distance between a particle and its third neighbour of the same species. By counting how many times each distance occurs we can identify the crystal. However, this method does not work with thermalised crystals whose distances are not sharply defined. Moreover, it is a global criterion that does not provide a label to each particle. If we are interested in classify each particle we have to extract information from the interaction matrix too; in this way we can determine neighbours through the Kern-Frenkel potential definition and keep track of the orientation of each patchy particle. Indeed, a particle can be labelled as DC-8, DC-16 or DC-24 by looking at the relative orientation between its first and second neighbours of the same species. In details, we measure the angles α_0 and α_1 for the first neighbours and the angle α for the second neighbours, as shown in Figure 6.4



(a) taking as an example the first species of the DC-8 and the DC-24 polymorphs, respectively.

In the perfect DC-8 crystal, particles have the same orientation, *i.e.* $\alpha_0 = \alpha_1 = \alpha = 0$. Conversely, the perfect DC-16 is characterised by $\alpha_0 = \alpha_1 = 109$ and the perfect DC-24 either by $\alpha_0 = 0$ and $\alpha_1 = 109$ (or vice versa by $\alpha_0 = 109$ and $\alpha_1 = 0$) or by $\alpha_0 = \alpha_1 = 109$. This is why we need to go to second neighbours; we can distinguish the DC-24 structure from the DC-16 one since for the first $\alpha = 109$, whereas for the second $\alpha = 0$. In simulations, we do not have perfect crystals so the angles are not sharply defined, and their distribution spreads out around the previous values. Measured α_0, α_1 (and α_0, α) angles in simulations are shown in Figure 6.4 (b) (and Figure 6.4 (c)). Purple triangles are used to mark DC-16 angles and black circles represent the DC-24 ones. Data come from nucleating trajectories ending in an extended crystal. Having defined an effective order parameter, we can analyse the nucleation properties of the three isotypes.

6.3.2 Nucleation properties of the isotypes

To inspect if the three polymorphs have the same nucleation ability, as predicted by CNT, we rank them according to their nucleation frequency. We remind that temperature T is in unit of ϵ/k_B, density ρ is in unit of $1/\sigma^3$ and pressure P in unit of ϵ/σ^3, where σ is the patchy particle diameter, ϵ is the square-well potential depth and $k_B = 1$. In particular, we simulate $N = 500$ patchy particles [101] at equimolar concentration in the canonical ensemble at $T = 0.1$, $\rho = 0.35$ (600 trajectories, both with and without aggregation-volume bias (AVB) moves [100, 101]) and at $T = 0.104$, $\rho = 0.4$ (300 trajectories with AVB dynamics). We label successfully nucleated those trajectories having a fraction of particles in the cubic diamond phase

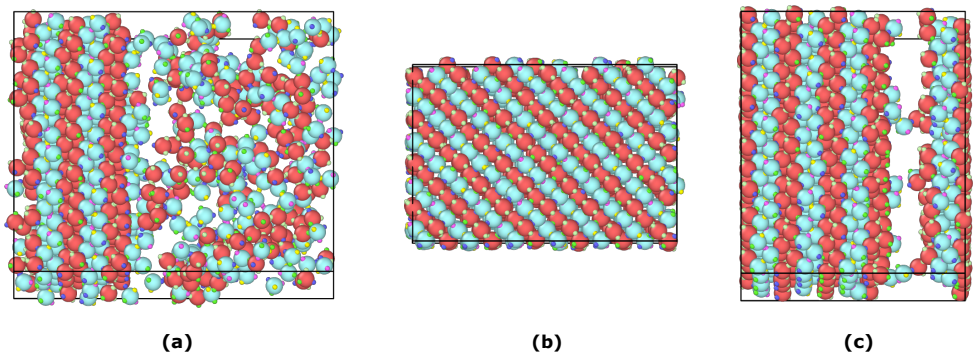
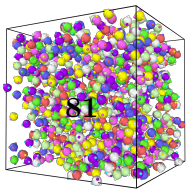


Figure 6.5. DC-8 growth. Snapshots of DC-8/liquid coexistence from NPT simulation at T=0.096 and P=0.02. An initial slab of crystal is added to the box with liquid particles (a) with the plane (b) serving as the interface. The crystal rapidly grows eventually spreading throughout the entire box (c). Image reproduced from [56].

greater than 0.5 and classify the obtained crystals. As shown in Figure 6.6 (a), and conversely to CNT predictions, the three polymorphs have different nucleation rates: the majority of nucleating trajectories form the DC-24 polymorph, the one with the largest unit cell size, while only a single nucleation event is observed for the one with the smallest unit cell size, the DC-8 polymorph. Specifically, for the $T = 0.1$, $\rho = 0.35$ state point, out of 600 trajectories, 43 form the DC-24, 15 the DC-16 and none the DC-8 with AVB dynamics and 30 form the DC-24, 11 the DC-16 and 1 the DC-8 with no AVB moves. Similar results are observed for the $T = 0.104$, $\rho = 0.4$ state point (35, 6 and 0, respectively, out of 300 trajectories). In Figure 6.6 (b) we report the progress in time of the nucleus size for two typical trajectories: the blue one spontaneously self-assembling a DC-24 polymorph, and the green one crystallising into a DC-16 structure. Snapshots of the critical nucleus are also displayed, showing that the different polymorphs are already distinguishable at critical sizes. Additionally, we analyse the growth of polymorphs to explicitly exclude the presence of cross nucleation of one polymorph on top of another. In particular, we perform additional simulations to test the growth of a crystalline slab of each polymorph through the [100] cubic plane. Our purpose is to explicitly verify if one polymorph can grow on a surface of a different polymorph. Simulations are done in the NPT ensemble. As illustrated in Figure 6.5 (a) for the DC-8 structure, the initial configuration involves placing a crystal slab in direct coexistence with liquid particles. The crystal is oriented such that the plane we want to analyse constitutes the interface with the liquid phase (Figure 6.5 (b)). The final snapshot in Figure 6.5 (c) shows how the initial crystalline slab expands to occupy the entire box, without the emergence of any other polymorph. Hence, we can conclude that the lack of trajectories self-assembling the smallest unit cell size polymorph cannot be attributed to the DC-8 initially nucleating and then changing identity as it grows. Others crystalline planes were not analysed. Nevertheless, by tracking the time evolution of the number of crystalline particles of each polymorph and by visual inspection of initial configurations of nucleating trajectories, we confirm that each polymorph originates and grows from a nucleus of the same polymorph.

The different nucleation rates of the three polymorphs are reflected by the height of their free energy barriers, which we compute with the Umbrella Sampling technique, described in Chapter 3. We run NPT Monte Carlo simulations at $T = 0.104$ and $P = 0.018$ with a harmonic bias potential. Each simulation is prepared by inserting in the liquid phase a crystalline seed of size n_0, where n_0 increases by 5 particles in successive simulations. The initial seeds have a roughly spherical shape and the same density for each polymorph. We carefully verify that the chosen bias potential and thermodynamic conditions ensure a good sampling, *i.e.* that there is an appreciable overlap between simulations with successive n_0, that no spontaneous nucleation



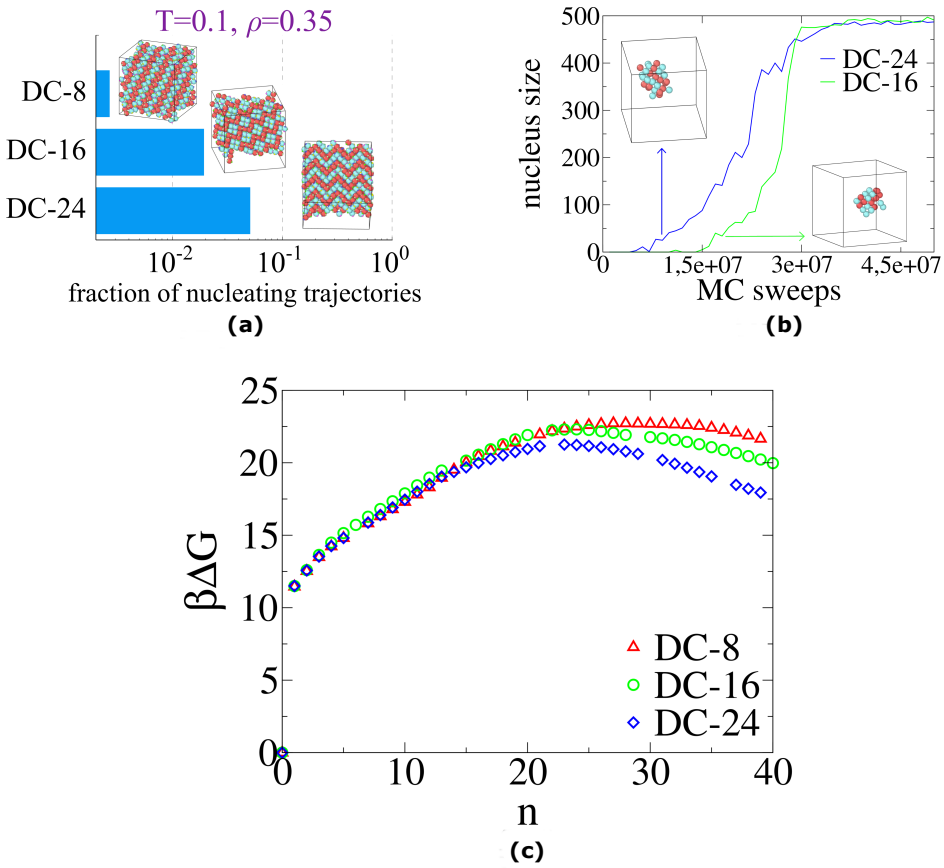


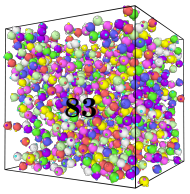
Figure 6.6. Direct nucleation simulations and nucleation barriers (a) Fraction of trajectories ending up into each polymorph out of 600 with snapshots of final configurations. Simulations are run in the canonical ensemble at $T = 0.1$, $\rho = 0.35$, $N = 500$. Interestingly, the polymorph with the largest unit cell size is the one that nucleates most frequently. (b) Time evolution of the number of crystalline particles with a snapshot of the critical nucleus for the DC-24 (blue) and the DC-16 (green) structures. (c) Gibbs free energy for the formation of crystal nucleus of size n for the three polymorphs computed by means of umbrella sampling technique at $T = 0.104$ and $P = 0.018$. The different trend in nucleation simulations is confirmed: the smallest unit cell size polymorph has the highest barrier. Image reproduced from [56].

occurs, and that there is no change in polymorph identity during the simulation. The resulting barriers are reported in Figure 6.6 (c). In order of increasing barrier height, we find DC-24, then DC-16, and finally DC-8, confirming that the polymorph with the lowest barrier is the one with the largest unit cell. The height difference between the DC-24 and DC-16 barrier is approximately $1k_BT$, a value that aligns well with the fractions of trajectories nucleating into the two polymorphs. The same comparison cannot be made for the DC-8 case as only a single nucleation event is observed for this polymorph. The critical size n_c for all nuclei is $n_c \lesssim 30$, *i.e.* smaller than the largest unit cell (48 particles) used in our exhaustive search of competing

polymorphs.

6.4 CNT failure

We have observed that the three polymorphs exhibit different nucleation properties, specifically in terms of nucleation rates and barrier heights, and that these differences cannot be attributed to variations in the bulk properties of the crystals. We ruled out the capillarity approximation, but we still have to explain why the DC-24 structure is the more frequently nucleation outcome. In the following, we investigate whether the different nucleation properties of the polymorphs can be traced back to the structural fluctuations within the melt. We follow the idea that if the liquid phase from which the nucleus arises already exhibits some degree of order in particle orientation, this will favour the nucleation of the polymorph that is structurally closer to the melt [191]. The three polymorphs, DC-8, DC-16, and DC-24, differ in fact by the orientation between neighbours of the same species. In particular, for each patchy particle, we can consider the angle α formed between the patch orientations of its two second-nearest neighbour of the same species (see inset of Figure 6.7 (b)): as already shown in Figure 6.4 (c), for the bulk polymorphs we have $\alpha_{\text{DC-24}} \sim 109°$, $\alpha_{\text{DC-16}} \sim 9°$, and $\alpha_{\text{DC-8}} \sim 9°$. In Figure 6.7 (b) we plot the radial profile of α from the centre of mass of nuclei of DC-24 (blue diamonds), DC-16 (green circles) and DC-8 (red triangles) polymorphs. We use configurations and trajectories from the umbrella sampling window of size $n \sim 50$ and, in order to compute α, we consider only particles having two second-nearest neighbours of the same species. To show the crystalline profile, in Figure 6.7 (a), we show the radial distribution of the total coherence which has high values (around 14) for the bulk diamond crystal, and small values (around 9.5) for the melt, irrespective of the polymorph. The figure shows the transition from the core of the nuclei at short distances to the melt at large distances. In the core region, α assumes the value of the corresponding bulk polymorph ($\alpha_{\text{DC-24}} \sim 109°$, $\alpha_{\text{DC-16}} \sim 9°$, and $\alpha_{\text{DC-8}} \sim 9°$). In the interfacial region, instead, the value of α is always that of DC-24 polymorph, irrespectively of the nucleus type. This shows that the orientational order in the melt resembles more closely that of the DC-24 polymorph, *i.e.*, the polymorph which nucleates more frequently and has the lowest nucleation barrier. A similar argument can be made to rationalise why the DC-8 is the less frequently nucleating polymorph, as the difference in the nucleation barrier heights of the DC-8 and DC-16 polymorphs can still be traced back to the liquid structure by looking at the distribution of the angle α_0. α_0 defines the relative orientation between first neighbours of the same species, and it is illustrated in the inset of Figure 6.8 (b) for patchy particles of the first species belonging to the DC-24 structure.



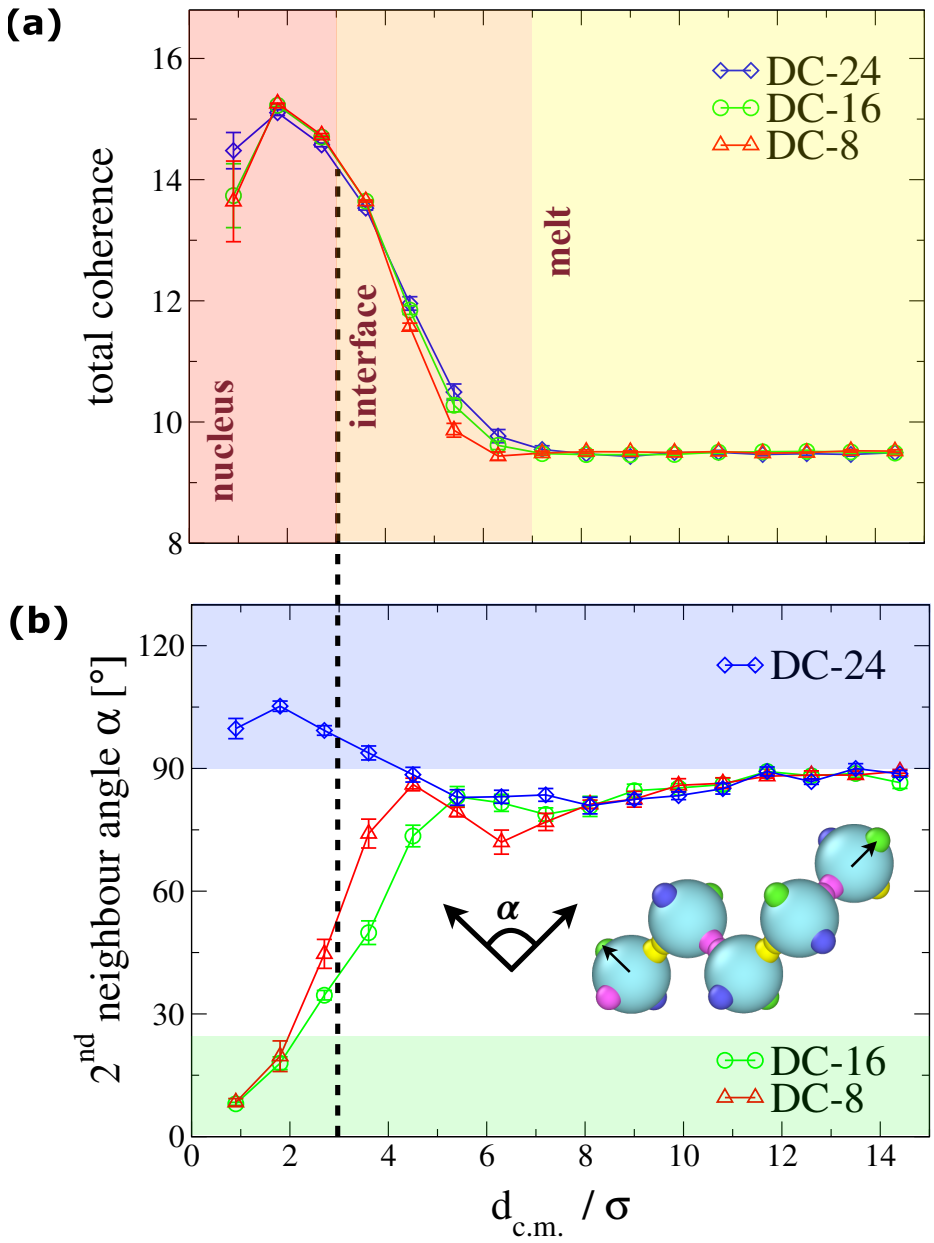


Figure 6.7. Bond orientational order in the liquid phase. Total coherence (a) and α angle (b) as a function of the distance of each particle from the centre of mass $d_{c.m.}$ of the largest crystalline cluster. Points are computed by averaging on different configurations and trajectories characterised by a crystalline nucleus of 50 particles: a DC-24 (blue diamonds), a DC-16 (green circles), and a DC-8 (red triangles) nucleus. α defines the relative orientation between second neighbours of the same species as illustrated for patchy particles of the first species of the DC-24 polymorph in the inset of figure (b). The coloured bands in (a) helps to locate the different regions: the nucleus, the interface and the melt. The coloured bands in (b) define the range of α values typical of the bulk DC-24 crystal (blue band) and of the bulk of both DC-16 and DC-8 (green band) polymorphs; particles in the interfacial region as well as the ones in the melt have α angles characteristic of the DC-24 structure. Image reproduced from [56].

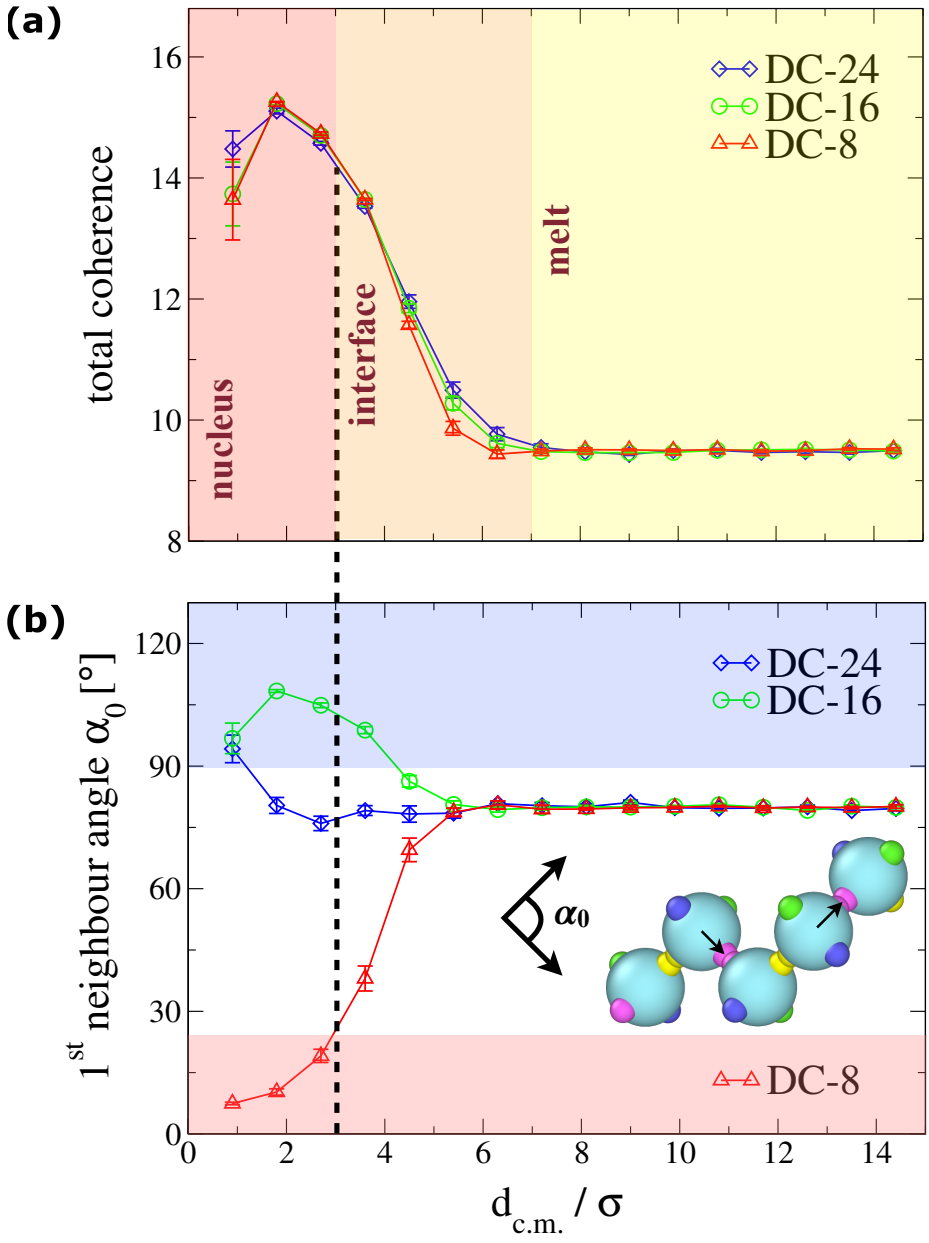
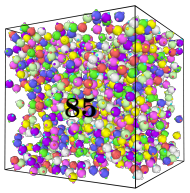


Figure 6.8. First neighbours radial profiles. Average radial profile of total coherence (a) and α_0 angle (b) from the centre of mass $d_{c.m.}$ of a 50 particles crystalline nucleus of DC-24 (blue diamonds), DC-16 (green circles) and DC-8 (red triangles) polymorphs. (a) α_0 defines the relative orientation between first neighbours of the same species, as illustrated for patchy particles of the first species of the DC-24 polymorph in the inset of figure (b). The coloured bands in (a) helps to locate the crystalline, the interfacial and the melt regions. The coloured bands in (b) define the range of α_0 values common to the bulk DC-24 and the bulk DC-16 (blue band) polymorphs and the one characteristic of the bulk DC-8 (red band) structure. Red band deviates from the typical values of the melt which are more closely approached by the blue band. Image reproduced from [56].



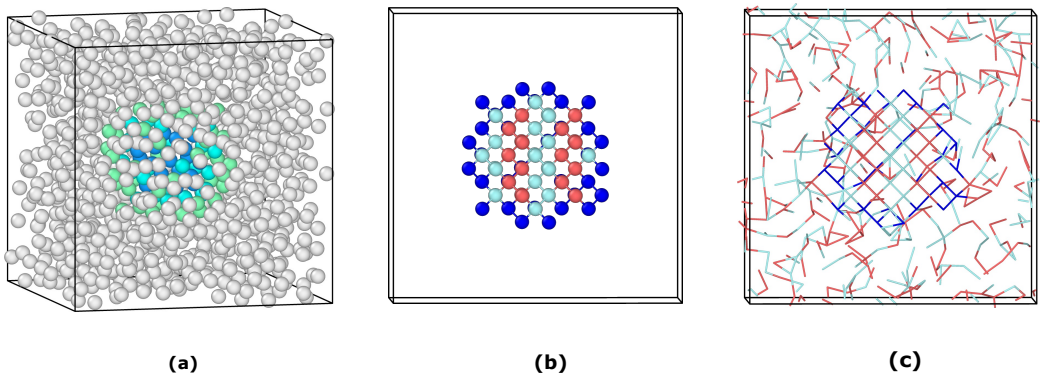


Figure 6.9. Initial and final snapshots with a frozen DC-8 crystalline seed. Snapshot of a typical initial configuration where a crystalline seed of 100 (coloured) particles is surrounded by 1000 liquid (white) particles. (b) The nucleus is frozen and the particles coloured blue are surface particles whose energy in vacuum is U_0. During the simulation, these surface particles can establish bonds with melt particles (c). This allows us to evaluate how many new nucleus-melt bonds are formed in each crystalline polymorph and determine if there is a structure where bond formation is more favourable. Image reproduced from [56].

The DC-8 and DC-16 crystals share common α values that do not align with the typical melt values. However, by plotting the radial profile of α_0 angle for the three polymorphs (Figure 6.8 (b)), we observe that their bulk values are different. In particular, the DC-8 angle is the farthest from the melt values. This explains why the DC-8 polymorph has the highest nucleation barrier and the DC-24 polymorph the lowest one. Moreover, the similarity between the DC-16 and DC-24 bulk α_0 angles highlights the need to extend the analysis to second neighbours to fully understand the distinct nucleation behaviours of the polymorphs.

Furthermore, a thorough analysis confirms that the observed difference in the nucleation frequency between the polymorphs cannot be attributed to the presence of favourable or unfavourable bonding sites on their surfaces. We verified that the three polymorphs do not differ in the density of attachment on surface sites. We ran additional Monte Carlo simulations in the canonical ensemble with 1000 particles. In particular, we prepared the system by inserting in the liquid phase a crystalline supercritical seed of size 100 (Figure 6.9 (a)). To prevent the nucleus from growing, the state point is chosen to be unstable for the crystal phase ($T = 0.16, \rho = 0.35$), while to prevent it from disappearing, the particles of the seed are kept fixed during the simulations. We evaluated the potential energy difference ($U - U_0$) of surface particles, relative to the configuration where the nucleus is in vacuum (Figure 6.9 (b)). In this way, we measure the new bonds formed between surface particles and melt ones (Figure 6.9 (c)). We underline that U_0 and the orientation of the patches are the same for all three frozen nuclei.

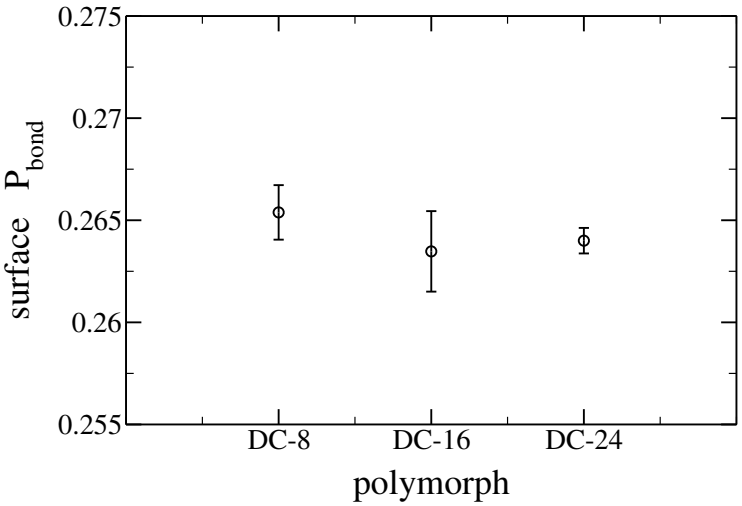
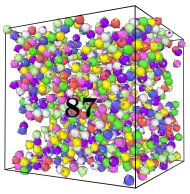


Figure 6.10. Surface-melt particles bonds formation. The average surface-melt bonding probability is the same (within the error) for the three polymorphs. This indicates that the surface particles of the three nucleus share the same bonding ability. Image reproduced from [56].

In Figure 6.10 we plot, for each polymorph, the average surface-melt bonding probability. It is obtained by normalising the $(U - U_0)$ difference with respect to the maximum number of bonds that surface particles can establish with the melt. Values are obtained by averaging over different configurations and trajectories, and they appear to be the same for the three structures. Hence, each nucleus has the same ability to make bonds with melt particles.

We can assert that the differences in the heights of the nucleation barriers cannot be ascribed to the presence of unfavourable surface particles that would lead to an entropic penalty in the less frequent structures. This provides additional evidence that the three polymorphs share the same surface tension. Hence, they are completely degenerate and, according to CNT predictions, should not display any differences in their nucleation frequency.

In conclusion, through extensive molecular simulations, our test shows that fully degenerate polymorphs exhibit remarkably different nucleation properties, directly contradicting CNT's predictions. Unlike previous tests of classical nucleation theory (CNT), which focus on accurately measuring nucleation rates and comparing them to theoretical predictions, leading to a pass or fail outcome depending on the system and/or state point, in this work we propose an alternative approach. We examine cases where CNT predicts no difference in nucleation rates between different polymorphs. This shift in focus sidesteps the common reliance on precise free-energy measurements and their dependence on specific state points. Instead, it only requires ranking the polymorphs according to their nucleation frequency.



Crucially, while discrepancies between CNT and measured nucleation rates can often be accounted for through *ad hoc* extensions to the theory, our test cannot be satisfied by any modification of CNT built upon the capillarity approximation. To run the falsifiability test we use the N2c8 binary mixture of patchy particles where three different polymorphs, despite having identical bulk and interfacial free energies, exhibit significantly different nucleation rates. We argue that CNT's primary limitation lies in its neglect of structural fluctuations within the liquid phase. Indeed, the polymorph that nucleates more frequently is the one whose orientational order more closely resembles that of the liquid phase where the nucleus arises; we can attribute the difference in nucleation rate to structural fluctuations within the melt. This test support all alternative theories beyond CNT that incorporate the structural properties of the liquid phase [5,132,133,182,191–198], rather than relying only on macroscopic quantities.

This analysis concludes our investigation of nucleation behaviour in multi-component systems. We found that while density and composition fluctuations can promote nucleation, by helping to overcome diffusional and compositional barriers, respectively, structural fluctuations appear to play a key role in selecting the resulting polymorph.
In the next chapter, we introduce the concept of inverse thermodynamics to explore possible ways to control by design the thermodynamic behaviour of multi-component systems. Achieving optimal self-assembly requires controlling not only the target structure, but also the intermediate steps leading to its formation. Our goal is to gain a more comprehensive control over the entire self-assembly pathway. Once integrated in the SAT-assembly framework we will have an optimal tool for inverse designing that allows to insert thermodynamic requirements in addition to the topological ones. In particular, having demonstrated why nucleation pathways along azeotropic points at a proper concentration can enhance self-assembly and claimed why azeotropic mixtures are an effective strategy to also simplify the thermodynamics of multi-component systems, we focus on how to originate and displace the azeotropic point by design. We choose this thermodynamic condition because it is relevant for self-assembly optimisation. However, we emphasise that this kind of control holds broader theoretical significance, as it demonstrate a strategy for designing interaction potentials with a targeted thermodynamic behaviour.

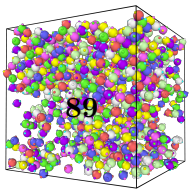
Chapter 7

Inverse thermodynamics

The traditional goal of inverse self-assembly is to design interactions that drive particles toward a desired target structure. However, achieving successful self-assembly also requires tuning the thermodynamic conditions under which the structure is stable. In this chapter, we extend the inverse design paradigm to explicitly address this challenge by developing a framework for inverse thermodynamics, *i.e.* the design of interaction potentials that realise specific thermodynamic behaviour. We demonstrate how precise control over both bonding topology and bond energetics of patchy particle mixtures enables the programming of targeted phase behaviour. In particular, having demonstrated that the inclusion of an azeotrope into artificial designs reveals effective pathways for the self-assembly of complex structures and provide a way to control the self-assembly of an arbitrary number of components, we establish design principles for azeotropic demixing and show how to create mixtures that exhibit azeotropy at any prescribed concentration. The results presented in this chapter originate from papers *C. Beneduce et al., ACS Nano, 2023* [33] and *C. Beneduce et al., J. Phys. Chem. B, 2025* [199].

7.1 Designing interactions for targeted phase behaviour

In atomic and molecular systems, interactions are fixed by quantum mechanics, limiting the ability to engineer specific thermodynamic behaviours. Soft matter, by contrast, thrives on tunability [31, 200–204]: interactions can be deliberately designed, whether through colloidal functionalisation [75, 205–208], DNA-mediated binding [65, 77, 209, 210], or anisotropic shapes [205, 211–213]. This flexibility has enabled the realisation of exotic states of matter, such as equilibrium gels with vanishing coexistence regions [115, 116], re-entrant phase transitions [214–216], and empty liquids stabilised by limited valence [217]. But beyond forward design (predicting properties from interactions), a deeper challenge emerges: the inverse thermody-



namics problem. That is, how should interactions be tailored to achieve a desired thermodynamic outcome? Significant strides have been made toward solving this problem [13, 218–226]. For instance, reducing the range of attractive potentials systematically depresses the critical temperature, eventually rendering the liquid-gas transition metastable [227, 228]. Similarly, finely tuned short-range repulsions can induce solid-solid transitions [229], while competing interactions generate re-entrant phenomena like melting upon cooling or condensation upon dilution [230–233]. These examples reveal general principles, but they often focus on qualitative trends (*e.g.*, shifting phase boundaries) rather than quantitative control (*e.g.*, prescribing exact coexistence conditions). Certain thermodynamic conditions are pivotal for functionality. In equilibrium gelation, for example, the narrowing of the coexistence region (a hallmark of "empty liquids" [114]) suppresses phase separation, enabling gelation without interference of a glass transition [114, 116, 117]. In self-assembly, azeotropy, a condition where liquid and vapour phases share identical composition, plays an analogous role. Azeotropes act as thermodynamic attractors during nucleation: by fixing the composition of coexisting phases, they maximise the yield of target structures when the azeotropic point matches the stoichiometry of those structures. In Chapter 5, we demonstrated that azeotropicity accelerates self-assembly by matching liquid and crystal composition. Here, we focus on the inverse problem: how can interactions be designed to place an azeotropic point at any desired concentration?

7.2 Inverse design of azeotropic points

The effective assembly of complex structures often requires the use of multiple components, each new component increasing the thermodynamic degrees of freedom of the system and hence the complexity of the self-assembly pathway. In Chapter 4, we demonstrated the possibility to use azeotropy as an attractive strategy to tame the complex phase behaviour of multi-component systems. At the azeotrope, the mixture behaves effectively as a single-component system and phase separation results into coexisting phases having the same concentration as the parent homogeneous phase. In Chapter 5, we also showed the advantages of combining azeotropic behaviour with self-assembly by explicitly including azeotropic points along the nucleation pathways of mixtures. They are summarised below:

- The ability to considerably enhance the reaction rates of the self-assembly process by quenching the system in a region of liquid-vapour metastability: in fact, it is well-established that for one-component systems nucleation rates increase in proximity of density fluctuations like the ones found near liquid-vapour critical points [234] and spinodal loci [148];

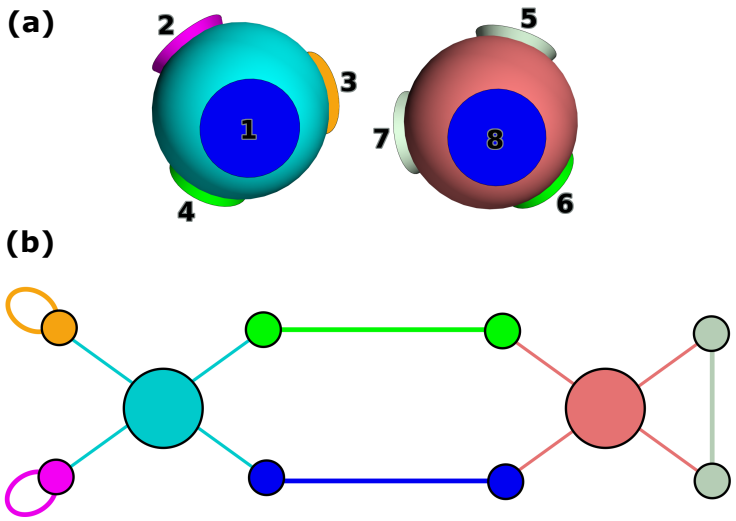
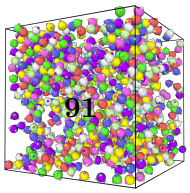


Figure 7.1. The N2c8 design. (a) 3D representation of the N2c8 design. It is a binary mixture of patchy particles, each featuring four patches tetrahedrally arranged. The first species, whose type of patches are labelled $1-4$, is coloured cyan, while the second species, with type of patches labelled $5-8$, is coloured red. The notation N2 refers to the two species, and c8 represents the eight distinct types of patches. Each pair of patches capable of forming bonds is represented by the same colour. (b) Graph of the N2c8 design. The two central nodes represent the two species, with each species connected to other nodes corresponding to its patches. Patches of same colour indicate interacting patches. All the links are coloured accordingly: links from each species to its patches are coloured to match the species, while links between pair of interacting patches are coloured to correspond with the patches. Image reproduced from [199].

- Increase the kinetics of the self-assembly reaction: if the concentration of the azeotropic point is the same as the crystal composition, one can avoid slow diffusion-limited process, where the crystal nucleus has to wait for the composition of the local environment to match the one of the target structure [235];

- The yield of the self-assembly process can proceed theoretically until all components are exhausted (to 100%), as the liquid phase will form at the same composition of the target crystalline structure.

Having established why it is important to design azeotropic mixtures, we now show how to select patchy particle systems that exhibit azeotropic points at a desired concentration. We rely on patchy particles which, as presented in Chapter 2, are a paradigm for programmable matter. They offer unparalleled control over both bonding topology (via patch number and geometry) and interaction energetics (via patch strength and specificity [236]) and are a valid coarse-grained model for all associating systems characterised by anisotropic interactions. Among them there are also DNA-origami, introduced in Chapter 3 as one of the most promising techniques



for experimentally realise multi-component systems with such high binding specificity.

In a mixture of patchy particles, the design is determined by how the patches and how they are distributed across the different components. As explained in Chapter 2, this can be effectively represented using an interaction matrix. Here, we introduce an alternative approach: a graph-based representation of designs. Larger nodes correspond to species, smaller nodes to patches, and edges between patches indicate complementary binding interactions. Self-loops represent self-complementary patches, *i.e.* patches that can bind to identical patches on other particles of the same species. The weight of each edge measures the interaction energy of the bond. A graph is unweighted if all interactions have the same energy, or weighted otherwise. As an example, in Figure 7.1 we plot the graph for the N2c8 mixture. In this framework, the inverse thermodynamics problem consists determining how to modify the graph topology to achieve target thermodynamic properties. This can be done by altering either the edge configuration or the edge weights, thereby directly acting on the key properties that define patchy particles. We aim to inverse design an azeotropic line (extending from $P = 0$, $T = 0$ to a binary critical point) at a predetermined concentration. This target concentration could correspond, for instance, to the stoichiometric ratio of the crystalline phase, where nucleation rate is maximised [55]. In the following, we present both strategies that tune the bond connectivity (the list of edges) or the interaction energies (the edge weights) to achieve control over the location of the azeotropic point.

7.3 Azeotropy from bond connectivity rules

The effects of bond connectivity on the thermodynamics can be quantified within the Wertheim perturbation theory, which, as described in Section 3.2, provides a closed-form description of phase equilibria in associating fluids. In particular, in deriving the azeotropy conditions we will make use of the law of mass action 3.7 [89,92,95], which quantifies the probability for a patch α of species i to be non-bonded. This probability is denoted X^i_α and we rewrite here its expression:

$$X^{(i)}_\alpha = \left[1 + \phi \sum_{j=1,N_s} x^{(j)} \sum_{\gamma\in\Gamma(j)} X^{(j)}_\gamma \Delta_{\alpha\gamma}\right]^{-1}$$

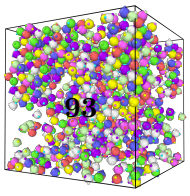
where $\Gamma(j)$ is the set of all patches of species j, ϕ is the total packing fraction, x_j is the molar fraction of species j, and $\Delta_{\alpha\gamma}$, being proportional to $e^{\epsilon_{\alpha\gamma}}$, accounts for the bond strength between patch α and γ. A detailed expression for $\Delta_{\alpha\gamma}$ is reported in Eq. 3.10. Since we are acting exclusively on the bond connectivity, the following simplification is considered: any pair of interacting patches forms bonds of the same

type (bonding volume V_b and energy ϵ), so that $\Delta_{\alpha\gamma} = \Delta$ if α and γ interact, or $\Delta_{\alpha\gamma} = 0$ if they don't. The interaction matrix $\boldsymbol{\Upsilon}$ has elements $\Upsilon_{\alpha\gamma} = \Delta_{\alpha\gamma}/\Delta$ that are ones if patches α and γ interact, and zeros if they don't. In Chapter 2, we showed that one possible strategy to compute $\boldsymbol{\Upsilon}$, *i.e.* and define which pair of patches should interact such that the particles self-assemble into a desired structure is the SAT-assembly framework [41]. Now we focus on general conditions one needs to impose on $\Upsilon_{\alpha\gamma}$ in order to obtain azeotropic mixtures, regardless of the desired target structure. Here, our goal in no longer strictly tied to self-assembly optimisation, but is broader: we aim to control thermodynamic behaviour through interactions design. While we chose to focus on azeotropy as a case study due to its importance in self-assembly, our free-energy framework is general. It depends explicitly on the interaction graph, and although the specific constraints will vary depending on the thermodynamic property of interest, the methodological foundation we introduce provides a flexible route toward inverse design.

We know that at the azeotropic point the mixture behaves as a single-component system. Hence, we can identify the condition where all patches α of each species i have the same probability of being bonded (and equivalently of not being bonded) as a sufficient condition for the appearance of an azeotrope. In the Wertheim formalism this means that the mass balance Eq. 3.7 must be the same for all patches of the system: $X_\alpha^{(i)} = X$. We notice that for azeotropic solutions $X_\alpha^{(i)} = X$ the Helmholtz free energy (Eq. 3.4) formally reduces to that of a single-component system, confirming the thermodynamic simplification occurring at this special point. Considering as an example a binary mixture with four patches, we indeed have:

$$
\begin{aligned}
\beta f_{bonding} &= \sum_{i=1}^{2} x^{(i)} \left[\sum_{\alpha \in \Gamma(i)} \left(\ln X_\alpha^{(i)} - \frac{X_\alpha^{(i)}}{2} \right) + \frac{1}{2} n(\Gamma(i)) \right] = \\
&= x^{(1)} \left[\left(\ln X_1 - \frac{X_1}{2} \right) + \left(\ln X_2 - \frac{X_2}{2} \right) + \left(\ln X_3 - \frac{X_3}{2} \right) + \left(\ln X_4 - \frac{X_4}{2} \right) + \left(\frac{M}{2} \right) \right] + \\
&+ x^{(2)} \left[\left(\ln X_5 - \frac{X_5}{2} \right) + \left(\ln X_6 - \frac{X_6}{2} \right) + \left(\ln X_7 - \frac{X_7}{2} \right) + \left(\ln X_8 - \frac{X_4}{8} \right) + \left(\frac{M}{2} \right) \right] = \\
&= x^{(1)} \left[4 \left(\ln X - \frac{X}{2} \right) + \left(\frac{M}{2} \right) \right] + x^{(2)} \left[4 \left(\ln X - \frac{X}{2} \right) + \left(\frac{M}{2} \right) \right] = \\
&= \left(x^{(1)} + x^{(2)} \right) \left[4 \left(\ln X - \frac{X}{2} \right) + \left(\frac{M}{2} \right) \right] = \\
&= 4 \left(\ln X - \frac{X}{2} \right) + \left(\frac{M}{2} \right)
\end{aligned}
\tag{7.1}
$$

since for a binary mixture with both species having four patches $\sum_{i=1}^{2} x^{(i)} = 1$ and



$n(\Gamma(i)) = M = 4$.

In summary, we look for bonding conditions under which all the $N_s \times N_p$ (number of species × number of colours) mass balance equations 3.7 become equivalent. This provides a sufficient condition for the appearance of azeotropy in a multi-component system. Following this idea, we identify three families of rules:

- the *bond exclusivity* condition for generating azeotropic points at equimolar conditions;
- the *bond multiplicity* condition for locating azeotropic points at non-equimolar conditions;
- the *fully-connected bond* condition for creating an ideal azeotrope, *i.e.* mixtures that are azeotropic at any concentration.

A detailed analysis of each rule is reported in the following subsections, where we confirm their validity not only within Wertheim's theory, but also explicitly through Monte Carlo simulations. We provide examples for each condition using specific systems; however the rules are generally valid for an N_s-component mixture where all species have the same diameter σ, the same number and placements of patches, and differ only in the patches type that, while sharing the same binding energy, has a proper binding specificity.

7.3.1 Bond exclusivity condition

One condition ensuring azeotropy is the bond exclusivity constraint requiring that each patch has a unique bonding partner (that can be itself in case of self-complementarity) among all patches of all species in the mixture [33]. This implies that all patches are distinct and the graph consists of nodes uniquely connected in one-to-one relationship; each node connects to exactly one other node, and no node is targeted by more than one source. Equivalently, it means that the interaction matrix $\boldsymbol{\Upsilon}$ has a single one for each row, located at a different column for different rows. Eq. 3.7 simplifies as follows: the sum over the patches ($\sum_{\gamma\in\Gamma(j)}$) and the sum over the species ($\sum_{j=1,N_s}$) reduce to a single contribution since the patch α belonging to species i can interact only with its partner patch γ belonging to species j (j can be also equal to i as well as α can be equal to γ). Therefore, the $N_s \times N_p$ mass balance equations for $X_\alpha^{(i)}$ reduce all to equations of the form

$$X_\alpha^{(i)} = \left[1 + \phi x^{(j)} X_\gamma^{(j)} \Delta_{\alpha\gamma}\right]^{-1} \tag{7.2}$$

which couple only $X_\alpha^{(i)}$ with $X_\gamma^{(j)}$. Moreover, since all bonds have the same strength, $\Delta_{\alpha\gamma} \equiv \Delta$ for all patches α and γ. By substituting the equations for $X_\gamma^{(j)}$ in the equation for $X_\alpha^{(i)}$, one obtains, even without knowing the exact patchy particles design, that under the bond exclusivity condition all the $N_s \times N_p$ mass balance equations become all equivalent to

$$X_\alpha^{(i)} + \phi\, x^{(i)}[X_\alpha^{(i)}]^2\Delta + \phi(x^{(j)} - x^{(i)})X_\alpha^{(i)}\Delta - 1 = 0 \tag{7.3}$$

We observe that this equation admits the azeotropic solution $X_\alpha^{(i)} = X$ for all patches α and species i of an N_s-component mixture, if $x^{(i)} = 1/N_s$, *i.e.* at equimolar conditions, where X is the solution of

$$X + \frac{\phi}{N_s}X^2\Delta - 1 = 0 \tag{7.4}$$

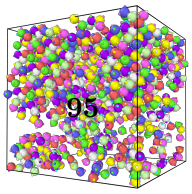
Thus the bond exclusivity condition generates an azeotrope at equimolar conditions: a N_s-component mixture will be azeotropic if it is prepared by mixing all the N_s components at the equimolar composition $1/N_s$. This rule can be exploited for self-assembling designs where the target structure is composed of an equal number of all species. Experimentally, it can be realised by functionalising particles surface by complementary DNA strands [237, 238], as described in Section 3.1.
The bond exclusivity condition is easily generalised to cases where multiple-bonding is allowed, *i.e.* when one patch is capable of bonding to more than one patch (also this case can be realised with single DNA strands functionalisation as explained in Subsection 3.1.2) and/or when the patches are not distinct (when the interaction matrix has repeated columns or rows, *i.e.* when its determinant is zero). An azeotrope an equimolar conditions can be obtained by ensuring that every patch has the same total number (m) of bonding partners (distributed over one or more species). In this case, the mass-balance equation admits the solution $X_\alpha^{(i)} = X$, with X satisfying the following equation

$$X + \phi\frac{m}{N_s}X^2\Delta - 1 = 0 \tag{7.5}$$

The multiple-bonding between patches can also be leveraged to move azeotropy away from equimolar concentrations, as explained in the following subsections.

7.3.2 Bond multiplicity condition

A simple generalisation of the bond exclusivity condition allows moving the azeotropic point to off-equimolar conditions. Considering a binary mixture where the ratio between the two species (denoted as (1) and (2)) is $1 : n$, in order to have an



azeotrope at $x^{(2)} = nx^{(1)}$ (*i.e.* $x^{(1)} = 1/(n+1)$ and $x^{(2)} = n/(n+1)$) it is sufficient to enforce:

- bond exclusivity to all patches bonding to species (2), *i.e.* each patch has a unique bonding partner with species (2);

- n-bond multiplicity to all patches bonding to species (1), *i.e.* each patch has n bonding partners with species (1).

In this case, in the mass balance equations for $X_\alpha^{(i)}$ the sum over the species ($\sum_{j=1,2}$) still drops out while the sum over the patches ($\sum_{\gamma\in\Gamma(j)}$) reduces to one or to two terms depending on if patch α satisfies the bond exclusivity condition or the bond multiplicity condition, respectively. With these conditions, we indeed observe that all mass balance equations (Eq. 3.7), admit the azeotropic solution $X_\alpha^{(i)} = X$ with X satisfying the following equation

$$X + \frac{n}{1+n}\phi X^2 \Delta - 1 = 0 \tag{7.6}$$

From a graph perspective the bond multiplicity condition constraints different nodes to have different edge valences. Specifically, nodes connected to species (2) have a single edge, while those connected to species (1) have n edges. Analogously, the interaction matrix exhibits a block structure. The implications of this bond multiplicity condition on both the form of the mass balance equations and the topology of the graph and interaction matrix are examined in more detail in the dedicated section through an example. We finally notice from Eq. 7.6 that this rule reduces to the previous bond exclusivity case when $n = 1$.

This recipe is generalisable to multi-component systems with more than two species: in this case, the bond multiplicity condition requires establishing a bond with m patches belonging to certain species, where m is the least common multiple between component ratios. While bond multiplicity provides a feasible way to shift the azeotropic point away from equimolar conditions, once the number of species and of patches is set, it is not possible to design a mixture exhibiting azeotropy at an arbitrary concentration. For instance for a binary mixture with four patches tetrahedrally arranged, there is no design satisfying the bond multiplicity condition for the ratio 1 : 3. These cases can be addressed by a special condition, called *ideal azeotropy*, for which azeotropy occurs at all concentrations, or by lifting the requirement that all bonds have the same energy, $\Delta_{\alpha\gamma} \neq \Delta$. We see both methods, in this order, in the following.

7.3.3 Fully-connected bond condition

In this subsection, we present a special set of bonding constraints ensuring the mixture to exhibit azeotropic demixing across all concentrations, *i.e.* the composition of the two coexisting phases always remains constant during phase separation. We name such rule fully-connected bond condition and the resulting system ideal azeotrope. For a general N_s-component mixture of patchy particles with N_p patches, the fully-connected bond condition is achieved when each patch can bind with N_s patches, each located on a different species. In this case the sum $\sum_{\gamma\in\Gamma(j)}$ in the mass balance equation 3.7 drops out, as there is only one bonding partner on each species, becoming

$$X_\alpha^{(i)} = \left[1 + \phi \sum_{j=1,N_s} x^{(j)} X_\beta^{(j)} \Delta_{\alpha\beta_j}\right]^{-1} \tag{7.7}$$

where the patch β on particle j is the unique bonding partner of patch α on species i. Now, assuming that all bonds are of the same type, $\Delta_{\alpha\gamma} = \Delta$, and remembering that $\sum_j x^{(j)} = 1$, we see that the mass balance equation admits azeotropic solutions $X_\alpha^{(i)} = X$, where X satisfies

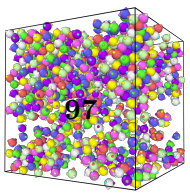
$$X + X^2\phi\Delta - 1 = 0 \tag{7.8}$$

Similar to a graph that satisfies the bond exclusivity condition, a graph obeying the fully-bonded rule also features an equal number of edges per node. However, unlike to the bond exclusivity condition, each node in the fully-bonded case exhibits a N_s degeneracy instead of a single exclusive connection.

Specific designs satisfying each presented rule are reported in the Example section. Importantly, these conditions can be inserted as additional clauses in the SAT-assembly framework. In fact, acting over bond connectivity constraints the number of bonds each patch can establish and the type of species these patches belong to. This implies that we can write boolean clauses ensuring azeotropicity by handling the colour interaction variables and the patch colouring variables. Taking the bond exclusivity condition, for example, we must add the following clauses for each patch of the system:

$$\neg x^{pcol}_{s,p,c_i} \vee \neg x^{pcol}_{s',p',c_i} \tag{7.9}$$

where s' runs over the N_s species and p' over the patches of each species different from p. In this way each patch will have a different colour; this, combined with the clause $C^{int}_{c_i,c_j,c_k} = \neg x^{int}_{c_i,c_j} \vee \neg x^{int}_{c_i,c_k}$ ensuring each colour has a single complementary



colour, guarantees the onset of a design featuring an azeotrope at concentration $1/N_s$.

Before looking at some examples, we focus on the effect of differentiate the bonding energy between the patches.

7.4 Azeotropy from bond energy control

While in the previous section we propose a precise control over azeotropic composition through bond topology modifications, an alternative approach becomes essential when the bonding topology must remain constrained. In such cases, we can instead exploit the tunability of interaction energies to engineer the desired azeotropic behaviour.

Two distinct phases (denoted as phase $'$ and phase $''$) coexist at a certain temperature and pressure when there is chemical equilibrium, *i.e.*:

$$\mu_i' = \mu_i''$$

In multi-component systems, the chemical potential μ_i of each component depends on the system's composition. Since the mole fractions sum to unity ($x_1 + \cdots + x_n = 1$), only $n-1$ composition variables are independent. For binary mixtures, we select one independent composition variable (denoted x). The equilibrium condition requires:

$$\begin{aligned} d(\mu_i' - \mu_i'') &= 0 \\ \frac{\partial \mu_i'}{\partial x'} dx' - \frac{\partial \mu_i''}{\partial x''} dx'' &= 0 \end{aligned}$$

Along the azeotropic line, the compositions coincide ($x' = x''$ and $dx' = dx''$), simplifying the condition to:

$$\left.\frac{\partial \mu_i'}{\partial x}\right|_{x'} = \left.\frac{\partial \mu_i''}{\partial x}\right|_{x''}$$

This equality of chemical potential derivatives, in addition to the equality of chemical potentials themselves, defines the special thermodynamic state of an azeotrope. Our algorithm exploits this relationship to control azeotropic composition.

The azeotropic point can be systematically shifted through controlled modifications of interaction parameters. For each temperature and pressure, we theoretically construct the Gibbs free energy per particle $g(x)$ by minimising Wertheim's free energy (Eq. 3.5) with respect to density for each composition x. Under coexistence conditions, $g(x)$ forms a two-branched curve representing the competing phases, with

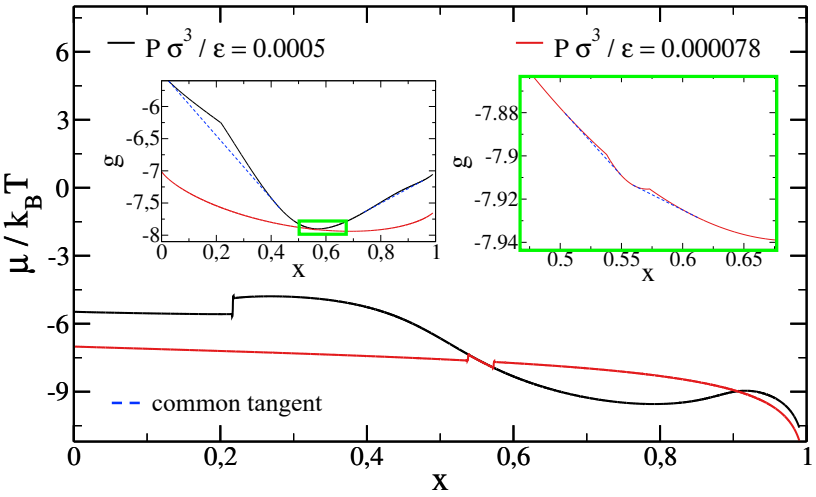
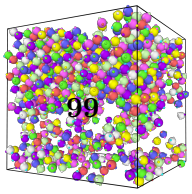


Figure 7.2. Chemical potential behaviours near and far from azeotropic conditions. The iterative process for shifting the azeotrope to a desired concentration follows a trial and error approach. In the graph we illustrate a step of the algorithm where setting $\epsilon^{'} = 1.2$ ultimately lead to an azeotrope at $x \sim 0.55$. By changing the energy from ϵ to $\epsilon^{'}$ starting from a known azeotropic state $(T_{azeo}, P_{azeo}, x_{azeo})$, the system moves away from the azeotropic conditions. The chemical potential as a function of the composition, shown in black (or the Gibbs free energy curve (in black in the left inset) satisfying a common tangent construction) is tracked to identify when a new coexistence condition is met by modifying the temperature and the pressure. Next, by fine tuning the pressure, the thermodynamic conditions where the chemical potential becomes continuous and with a single concavity (corresponding to a tangency in the Gibbs free energy curves of the two phases) is progressively met, as illustrated by the red curves. The azeotropic concentration x_{azeo}^{iter} at these new conditions $(T_{azeo}^{iter}, P_{azeo}^{iter})$ is found and used as a feedback for the next iteration of $\epsilon^{'}$. Imaged reproduced from [199].

their intersection determining the equilibrium concentrations through the common tangent construction (illustrated by dashed blue lines in the insets of Figure 7.2). At the azeotropic point, these branches become tangent at a single concentration where both phases share identical composition. From $g(x)$, we compute the chemical potentials $\mu_i(x)$, which exhibit distinct signatures of phase behaviour. In coexistence regions, $\mu_i(x)$ becomes multivalued (with multiple concentrations yielding identical chemical potential) and develops discontinuities. The width of these coexistence regions is pressure-dependent: it decreases with decreasing (increasing) pressure for mixtures exhibiting negative (positive) azeotropes [108, 109], and vanishes at the azeotropic point, as illustrated in Figure 7.2. Our algorithm begins with known azeotropic conditions $(P_{\text{azeo}}, T_{\text{azeo}}, x_{\text{azeo}})$. We apply small perturbations to the target bond energy ϵ, then compute the updated $\mu_i(x)$. By adjusting pressure and temperature, we track the evolution of the coexistence region until $\mu_i(x)$ becomes single-valued, identifying the new azeotropic point $(P_{\text{azeo}}^{\text{iter}}, T_{\text{azeo}}^{\text{iter}}, x_{\text{azeo}}^{\text{iter}})$. While these adjustments can be performed manually for simple cases, the procedure could be



automated through iterative schemes such as gradient descent. This process iterates until convergence to the desired concentration $x_{\text{azeo}}^{\text{goal}}$ is achieved.

In summary, we identified two approaches to designing azeotropic mixtures. The first involves the equality of all bonding probabilities for all patches of the system, while the second requires the simultaneous equality of chemical potentials and of their derivatives across coexisting phases. In the following section, we put these concepts in practice.

7.5 Examples

In this section we apply the presented azeotropic rules to specific cases, and we verify their effectiveness.

7.5.1 Bond exclusivity

Our case study design, the N2c8 solution, which we explored in the previous chapters, is able to exclusively self-assemble into a cubic diamond and satisfies the bond exclusivity condition. Its graph is shown in Figure 7.1 and its interaction matrix is

$$\boldsymbol{\Upsilon}_{\text{N2c8}} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \tag{7.10}$$

We can observe that each small node of the graph is exclusively connected to another node and that the interaction matrix has a single one for each row located at different columns for different rows. Hence, as required by the bond exclusivity condition, each path has a unique bonding partner. Since this rule guarantees the appearance of an azeotrope at equimolar conditions, and being the N2c8 solution a binary mixture, we expect this design to feature an azeotropic line at the concentration $x^{(1)} = x^{(2)} = 1/2$. The pressure-concentration and density-concentration phase diagrams we have computed in Chapter 4, both theoretically within the Wertheim perturbation theory and via Monte Carlo simulations in the Gibbs ensemble, confirm the expected azeotropic nature of this design. In the pressure-concentration phase

diagram 4.3 (a) the binodal curve reduces to a point at concentration equal to x=0.5 and in the density-concentration phase diagrams 4.3 (b) it is only at x=0.5 that the tie-line is a straight line connecting two phases at different densities at the same concentration.
For completeness, we report also the mass balance equations 3.7 for this specific design:

$$\begin{aligned}
&\text{if } \alpha = 2,3 \\
&X_\alpha + \phi x^{(1)} {X_\alpha}^2 \Delta - 1 = 0 \\
&\\
&\text{if } \alpha = 5,7 \\
&X_\alpha + 2\phi x^{(2)} {X_\alpha}^2 \Delta - 1 = 0 \\
&\\
&\text{if } \alpha = 1,4 \\
&X_\alpha + \phi x^{(1)} {X_\alpha}^2 \Delta + \phi X_\alpha \Delta (x^{(2)} - x^{(1)}) - 1 = 0 \\
&\\
&\text{if } \alpha = 6,8 \\
&X_\alpha + \phi x^{(2)} {X_\alpha}^2 \Delta + \phi X_\alpha \Delta (x^{(1)} - x^{(2)}) - 1 = 0
\end{aligned} \tag{7.11}$$

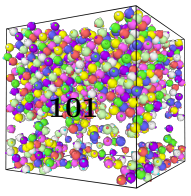
We notice, as expected, that if $x^{(1)} = x^{(2)}$ all the equations in Eq. 7.11 become equal to:

$$X + \frac{\phi}{2} X^2 \Delta - 1 = 0 \tag{7.12}$$

7.5.2 Bond multiplicity in a binary mixture

If we want to move to off azeotropic condition the previous binary mixture, we can redefine its binding rules according to the following graph (Figure 7.3) and interaction matrix (7.13).

$$\boldsymbol{\Upsilon}_{\text{N2c6}} = \begin{pmatrix}
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\
1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1
\end{pmatrix} \tag{7.13}$$



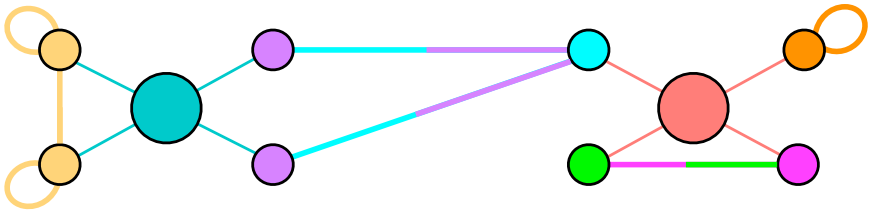


Figure 7.3. Binary mixture obeying to the bond multiplicity condition. The graph represents a design that ensures a binary mixture of patchy particles, each with four patches, exhibits an azeotrope at $1:2$ ratio. The two species are represented by the bigger nodes, coloured red, and cyan. Each node is connected by links of the same colour to its respective patches, indicated as smaller nodes. We have six different patches, *i.e.* colours. The desired $1:2$ ratio is reflected in the bond connectivity: each smaller node connected to the second red species is incident to two links (one for the species and one for the compatible patches), while nodes connecting to the remaining species are incident to three links (one for the species and two for the two compatible patches).

Unlike the bond exclusivity case, not all patches are different and the number of distinct mass balance equations is the smaller than $N_s \times N_p$. Indeed, the interaction matrix features two pairs of equal rows, and the interaction energies are the same for all patches of the system. This is why we name it $\boldsymbol{\Upsilon}_{\mathrm{N2c6}}$: there are two different species, and six independent patches(colours). In particular, $X_1^{(1)} = X_2^{(1)}$ and $X_3^{(1)} = X_4^{(1)}$, and with simple algebraic steps, the mass balance equations 3.7 reduce to one of the following four forms

$$
\begin{aligned}
&\text{if } \alpha = 1,2 \\
&X_\alpha + 2\phi x^{(1)} {X_\alpha}^2 \Delta + \phi X_\alpha \Delta (x^{(2)} - 2x^{(1)}) - 1 = 0 \\
&\\
&\text{if } \alpha = 3,4 \\
&X_\alpha + 2\phi x^{(1)} {X_\alpha}^2 \Delta - 1 = 0 \\
&\\
&\text{if } \alpha = 5 \\
&X_\alpha + \phi x^{(2)} {X_\alpha}^2 \Delta + \phi X_\alpha \Delta (2x^{(1)} - x^{(2)}) - 1 = 0 \\
&\\
&\text{if } \alpha = 6,7,8 \\
&X_\alpha + \phi x^{(2)} {X_\alpha}^2 \Delta - 1 = 0
\end{aligned}
\tag{7.14}
$$

They further simplify to Eq. 7.15 if $x^{(2)} = 2x^{(1)}$.

$$
X + \frac{2}{3}\phi X^2 \Delta - 1 = 0 \tag{7.15}
$$

Since $x^{(1)} + x^{(2)} = 1$, then $x^{(1)} = 1/3$ and $x^{(2)} = 2/3$: this binary mixture displays an azeotrope at the specific non equimolar concentration of $x^{(1)} = 1/3$ and $x^{(2)} = 2/3$. This is exactly what we expected, since the graph and the interaction matrix satisfy the bond multiplicity condition for $n = 2$. In fact, we have that patches connecting to the second species have a single one for each row (a single edge connection in the graph), whereas patches connecting to the first species have two ones for each row (a double edge connection in the graph). Hence, we have patches either under bond exclusivity constraint or establishing n-bond multiplicity with $n = 2$, as required by the bond multiplicity rule to design an azeotrope at $1 : 2$ ratio.

A computational verification of the bond multiplicity rule is performed in the following subsection, applied to a more representative case targeting the cubic diamond assembly.

7.5.3 Bond multiplicity in a ternary mixture

In this subsection, we target a three component system capable of spontaneously assembling into a cubic diamond at off-equimolar conditions. Specifically, the desired

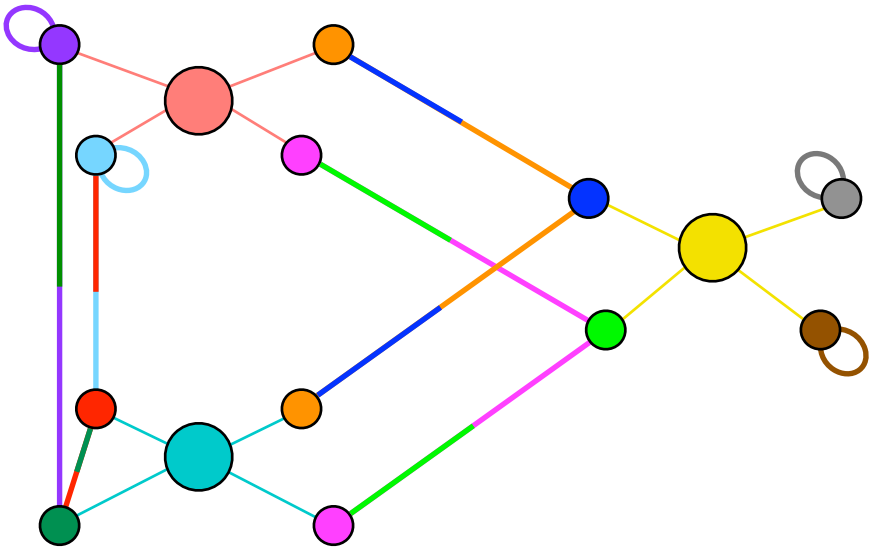
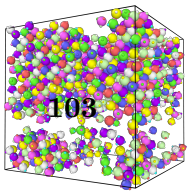


Figure 7.4. Ternary mixture obeying to the bond multiplicity condition. The graph represents a design that ensures a three component system of patchy particles, each with four patches, to have an azeotrope at $1 : 1 : 2$ ratio. The three species are represented by the bigger nodes, coloured red, cyan and yellow. Each bigger node is connected by links of the same colour to its respective patches, illustrated as smaller nodes. We have ten different patches, accordingly represented with distinct colours. Each smaller node connected to the third yellow species is incident to two links, while nodes connecting to the remaining species are incident to three links, satisfying the bond multiplicity condition for the $1 : 1 : 2$ ratio.



16 particle cubic diamond unit cell is made by tetravalent patchy particles that are 4 of the first species, 4 of the second species and 8 of the third species. Hence, in order to achieve optimal self-assembly we aim to select a nucleation pathway through an azeotrope at the same 1:1:2 stoichiometry ratio as the species appear in the target structure, as demonstrated in Chapter 5. To this end we can leverage the bond multiplicity condition. The connectivity we must satisfy dictates that each patch making a bond with the majority component (the third species) must obey the bond exclusivity condition, whereas all other patches should have two bonding partners among patches of the first and second species. A graph obeying this bond connectivity is shown in figure 7.4 and its interaction matrix is reported in 7.16.

$$\boldsymbol{\Upsilon}_{\text{N3c10}} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \tag{7.16}$$

Each row has either a single one among the last four columns, or two one distributed among the first eight columns. Equivalently, in the graph each node connected to the third species features a single edge, while nodes connecting to the two other species have two edges. We also highlight the presence of two identical pairs of rows, meaning that two pairs of patches are equal. In the graph, this is illustrated by using the same colours for those patches. This design is found by means of the SAT-assembly framework, where in addition to clauses ensuring the exclusive compatibility with the cubic diamond topology, we added clauses satisfying the bond multiplicity condition. In particular, we opened to the possibility of forming two bonds with the same patches while still maintaining bond exclusivity for all patches interacting with the third species. This guarantees the appearance of an azeotrope at concentration equal to $x^{(3)} = 2x^{(1)} = 2x^{(2)}$. To confirm the effectiveness of the design, we run several Monte Carlo simulations with 500 particles for different temperatures and densities at the expected azeotropic concentration. In Figure 7.5 (a) we show a snapshot of a typical configuration after the metastable liquid has formed but

before nucleation. Since vapour-liquid separation occurs at azeotropic conditions, the vapour slab does not show excess particles for any component. Figure 7.5 (b) shows a snapshot of a correctly assembled cubic diamond crystal that formed at the target $1:1:2$ composition. The last (c) panel shows the number of crystalline particles of each species as a function of Monte Carlo sweeps, demonstrating that the particles belonging to the first and second species (blue and red line, respectively) always have the same concentration, which is half as that of the third species (green line). We also note that the final nucleus comprises almost all particles, demonstrating that nucleation is indeed occurring at the same starting composition that matches the one of the target crystal. Phase separation is occurring without changing the

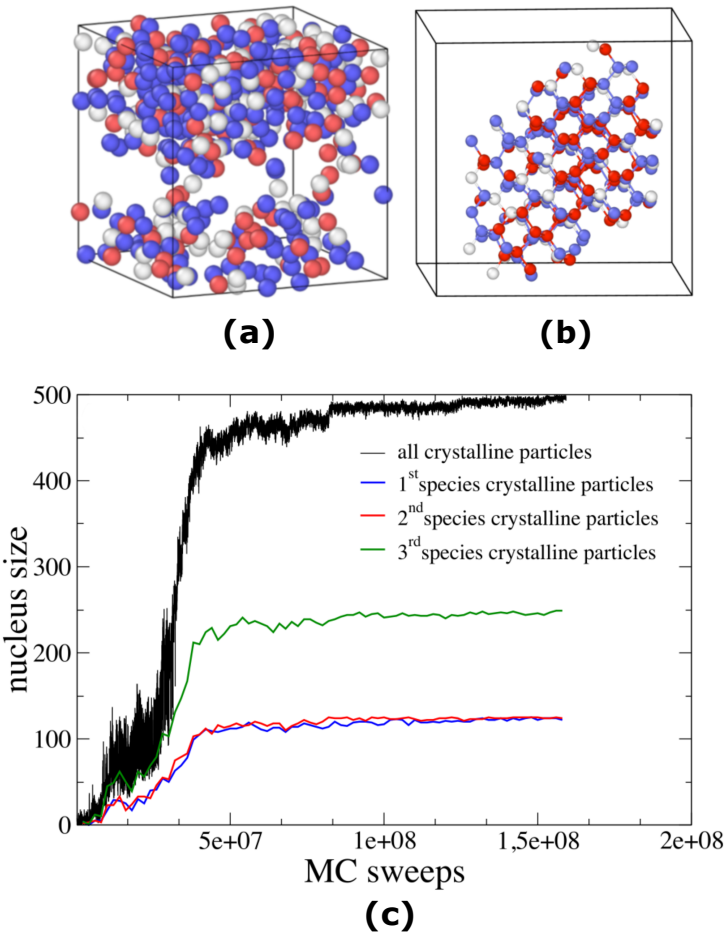
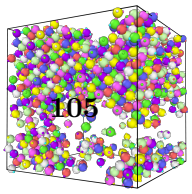


Figure 7.5. Self-assembly of a cubic diamond from a ternary mixture prepared at the azeotropic concentration $\mathbf{x^{(3)} = 2x^{(1)} = 2x^{(2)}}$. (a) Coexistence between the vapour and the metastable liquid phase. (b) Largest crystalline nucleus in the final MC configuration. (c) Size of the largest crystalline cluster as a function of MC sweeps (black line). The number of crystalline particles belonging to the third species (green line) is exactly double that of the first and the second species (blue and red lines, respectively) during the whole nucleation process. Image reproduced from [33].



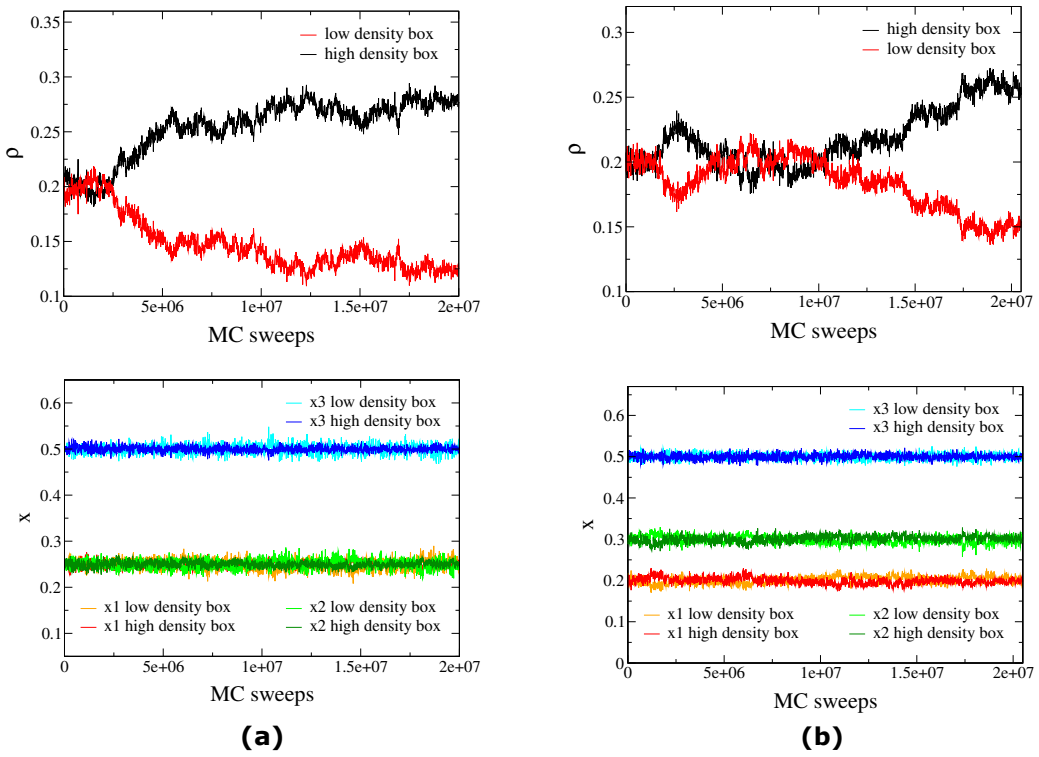


Figure 7.6. Gibbs ensemble simulations at azeotropic conditions. We track the evolution of the density and the concentration in two equally initialised boxes at temperature $T = 0.098$, density $\rho = 0.2$, and number of particles $N = 1000$, with different component ratios: $x_1 = x_2 = 0.25$, $x_3 = 0.5$ in (a), and $x_1 = 0.2$, $x_2 = 0.3$, $x_3 = 0.5$ in (b). In both cases, phase separation occurred, as indicated by the densities of the boxes equilibrating to different values, while preserving the initial component ratio. Both (a) and (b) correspond to azeotropic conditions.

initialised component ratio, demonstrating we are at the azeotropic concentration. Here, all particles in the liquid phase can participate in the formation of the crystal. This confirm that we successfully designed a ternary mixture with an azeotrope at $1 : 1 : 2$ through the bond multiplicity rule and that it correctly assembles in the target cubic diamond crystal at these controlled and optimal conditions.

Furthermore, we draw attention to the fact that this system consisting of three species and ten colours, shows multiple azeotropes. This becomes evident by directly analysing its mass balance equations. Writing down Eq. 3.7 for each patch of the system, and performing the required substitutions, we obtain four types of equations:

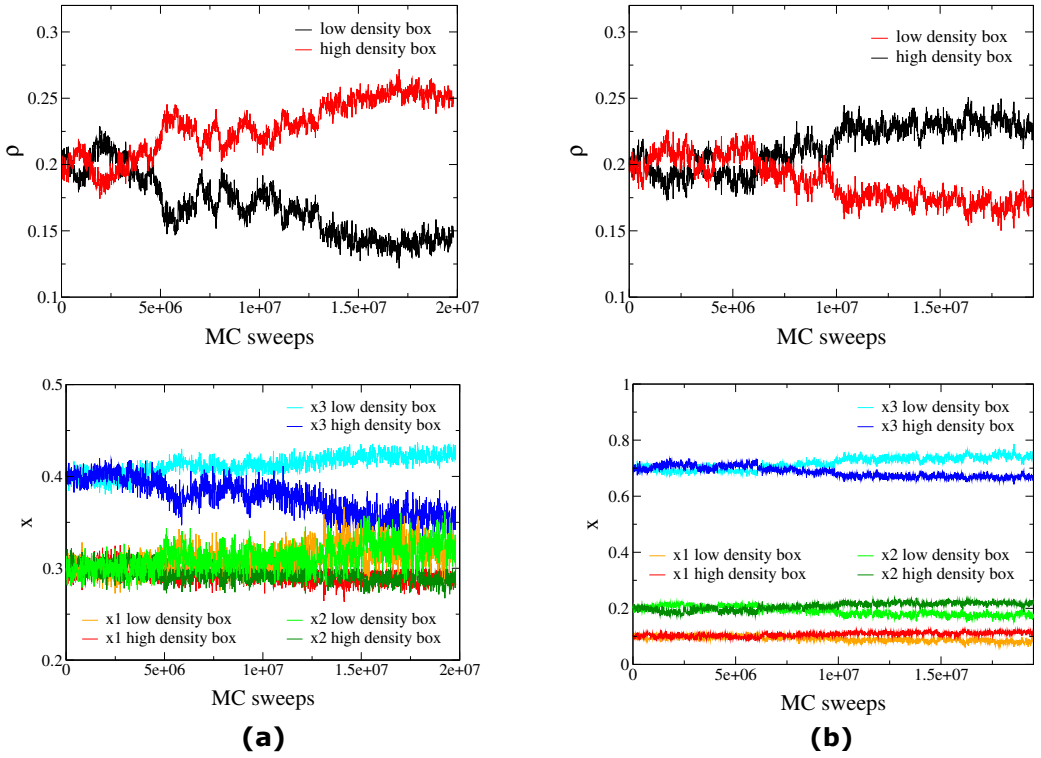
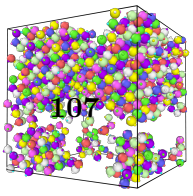


Figure 7.7. Gibbs ensemble simulations at off-azeotropic conditions. We track the evolution of the density and the concentration in two equally initialised boxes at temperature $T = 0.098$, density $\rho = 0.2$, and number of particles $N = 1000$, with different component ratios: $x_1 = x_2 = 0.3$, $x_3 = 0.4$ in (a), and $x_1 = 0.1$, $x_2 = 0.2$, $x_3 = 0.7$ in (b). In both cases, phase separation occurred, as indicated by the densities of the boxes equilibrating to different values, at off-azeotropic conditions. When the condition $x_1 + x_2 = x_3$ is no longer satisfied, the system enters the coexistence region, and vapour and liquid phases coexist at different concentrations.

$$
\begin{aligned}
&\text{if } \alpha = 8, 11 \\
&X_\alpha + \phi x^{(3)} {X_\alpha}^2 \Delta - 1 = 0 \\
&\\
&\text{if } \alpha = 1, 3, 5, 7 \\
&X_\alpha + \phi (x^{(1)} + x^{(2)}) {X_\alpha}^2 \Delta - 1 = 0 \\
&\\
&\text{if } \alpha = 0, 2, 4, 6 \\
&X_\alpha + \phi (x^{(1)} + x^{(2)}) {X_\alpha}^2 \Delta + \phi X_\alpha \Delta (x^{(3)} - (x^{(1)} + x^{(2)}) - 1 = 0 \\
&\\
&\text{if } \alpha = 9, 10 \\
&X_\alpha + \phi x^{(3)} {X_\alpha}^2 \Delta + \phi X_\alpha \Delta ((x^{(1)} + x^{(2)}) - x^{(3)}) - 1 = 0
\end{aligned}
\tag{7.17}
$$



In particular, the second type of equations comes from the equality of $X_1 = X_3$ (and $X_5 = X_7$) that holds for invariance under patches exchange. All the equations become equal if $x^{(3)} = x^{(1)} + x^{(2)}$. Considering that $x^{(1)} + x^{(2)} + x^{(3)} = 1$, then $x^{(3)} = 0.5$. Hence, we have a free parameter and the presented connectivity generates an azeotrope at whatever $x^{(1)}$ and $x^{(2)}$ values whose sum is equal to 0.5, including the targeted $x^{(1)} = x^{(2)} = 0.25$ composition. This is shown in Figure 7.6 through Gibbs ensemble simulations run at temperature $T = 0.098$, and density $\rho = 0.2$, a state point located inside the vapour-liquid coexistence region. We observe that the concentration of the two boxes stays constant if we prepare them at: $x^{(1)} = x^{(2)} = 0.25, \quad x^{(3)} = 0.5$; $x^{(1)} = 0.3, \quad x^{(2)} = 0.2, \quad x^{(3)} = 0.5$; $x^{(1)} = 0.35, \quad x^{(2)} = 0.15, \quad x^{(3)} = 0.5$. Conversely, at $x^{(1)} = x^{(2)} = 0.3, \quad x^{(3)} = 0.4$, and $x^{(1)} = 0.1, x^{(2)} = 0.2, \quad x^{(3)} = 0.7$ phase separation does not preserve the original ratio between the components: we are at off azeotropic conditions. In the following subsection we show an example of a mixture exhibiting azeotropicity at any concentration.

7.5.4 Fully connected

Another possible approach to shift the azeotrope to off-equimolar concentration is by altering its bond topology to fulfil the fully connected bond requirements. This originates azeotropic demixing across all compositions, where dew and bubble curves collapse into a single line, a condition we term an ideal azeotropy. Here we show the striking example of a binary mixture of patchy particles with four patches. A possible interaction matrix satisfying the fully-connected bond recipe is:

$$\boldsymbol{\Upsilon}^{\text{ideal}}_{\text{N2c6}} = \begin{pmatrix} 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \end{pmatrix} \tag{7.18}$$

Differently from the bond exclusivity interaction matrix that exhibits a single one for each row, this matrix has two ones for each row: the first is located among the first four columns (first species) and the other among the last four columns (second species). This ensures that, as required, each patch has two interacting patches, one on each specie as highlighted by the graph in Figure 7.8.
We now demonstrate that azeotropy is achieved without requiring equimolarity.

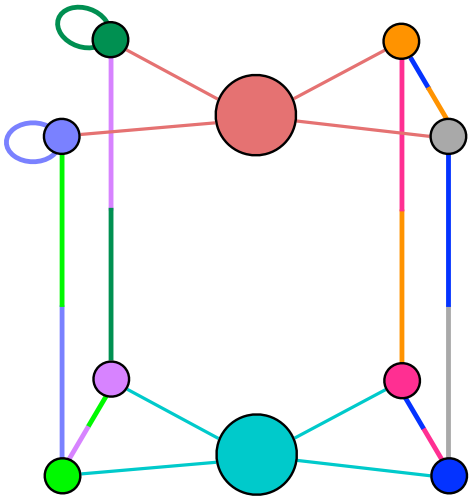


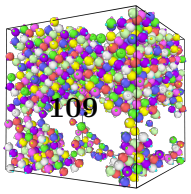
Figure 7.8. Ideal azeotropic binary mixture. The graph features two central nodes, representing the two species, coloured cyan and red. Each node is connected by links of the same colour to its respective patches, indicated as smaller nodes. Each patch can connect to two other patches, one from the first species and the other from the second species. Each smaller node is incident to three links: two coloured according to the patches it interacts with and the third one (cyan or red) indicating the species it belongs to. Image reproduced from [199].

Even better, we show that this binary mixture exhibits azeotropy not only if the system is at a particular concentration, but whatever ratio the two species are mixed together. We begin by verifying that the interaction matrix 7.18 generates a law of mass action (Eq. 3.7) admitting an azeotropic solution $X_\alpha^{(i)} = X$ of the form

$$X + \phi X^2 \Delta - 1 = 0 \tag{7.19}$$

that is independent of molar fractions x_i. This demonstrates that the system has an azeotrope at every concentration $x_A \in [0, 1]$. In the mass balance equations for $X_\alpha^{(i)}$ of a design obeying the interaction matrix in 7.18, the sum over the species $\sum_{j=1,2}$ does not drop out while the sum $\sum_{\gamma \in \Gamma(j)} X_\gamma^{(j)} \Delta_{\alpha\gamma}$ still reduces to one term as for the case where each patch can make a bond only with another patch. Indeed now, even if each patch makes a bond with two other patches, the patches involved in the bonds are located one on the first species and the other on the second species. Hence, for each patch α, $\Delta_{\alpha\gamma}$ is different from zero only for two patches, γ and δ, not belonging to the same patchy particle species. Therefore, $X_\alpha^{(i)}$ is recast as

$$X_\alpha^{(i)} = \frac{1}{1 + \phi \left[x^{(i)} X_\gamma^{(i)} \Delta_{\alpha\gamma} + x^{(j)} X_\delta^{(j)} \Delta_{\alpha\delta} \right]} \tag{7.20}$$



Now we impose the equal bonding energy condition that allows to set $\Delta_{\alpha\gamma}$, for whatever α and γ, at the same value denoted as Δ. In this way, for each patch α, $X_\alpha^{(i)}$ becomes of the form

$$X_\alpha^{(i)} = \frac{1}{1 + \phi\left[x^{(i)} X_\gamma^{(i)} + x^{(j)} X_\delta^{(j)}\right]\Delta} \tag{7.21}$$

In particular, we have eight equations. For example, the ones for the patches 1 and 2 are:

$$\begin{aligned} X_1^{(1)} &= \frac{1}{1+\phi\left[x^{(1)} X_2^{(1)} + x^{(2)} X_5^{(2)}\right]\Delta} \\ X_2^{(1)} &= \frac{1}{1+\phi\left[x^{(1)} X_1^{(1)} + x^{(2)} X_6^{(2)}\right]\Delta} \end{aligned} \tag{7.22}$$

We notice that $X_1^{(1)} = X_6^{(2)} \equiv X$ and that $X_2^{(1)} = X_5^{(2)} \equiv X'$. This implies that

$$\begin{aligned} X &= \frac{1}{1+\phi[x^{(1)}+x^{(2)}]X'\Delta} = \frac{1}{1+\phi X'\Delta} \\ X' &= \frac{1}{1+\phi[x^{(1)}+x^{(2)}]X\Delta} = \frac{1}{1+\phi X\Delta} \end{aligned} \tag{7.23}$$

By replacing the expression for $X^{'}$ in the equation for X and vice versa we obtain the two equal equations:

$$\begin{aligned} X + X^2\phi\Delta - 1 = 0 \\ X^{'} + X^{'2}\phi\Delta - 1 = 0 \end{aligned} \tag{7.24}$$

Therefore, satisfying the same equations, $X_1^{(1)} = X_6^{(2)} = X_2^{(1)} = X_5^{(2)}$. We are left to demonstrate that also $X_3^{(1)}, X_4^{(1)}, X_7^{(2)}, X_8^{(2)}$ are defined by equations equal to the ones in Eq. 7.24. Firstly we notice that if $X_7^{(2)} = X_8^{(2)}$ then $X_3^{(1)} = X_4^{(1)}$ and this would imply that $X_7^{(2)} = X_8^{(2)} = X_3^{(1)} = X_4^{(1)} \equiv X^{''}$. Hence, we can write

$$X^{''} = \frac{1}{1 + \phi[x^{(1)} + x^{(2)}]X^{''}\Delta} = \frac{1}{1 + \phi X^{''}\Delta} \tag{7.25}$$

which can be rewritten as

$$X^{''} + X^{''2}\phi\Delta - 1 = 0 \tag{7.26}$$

i.e. the same equation as the ones reported in Eq. 7.24. Therefore, if $X_7^{(2)} = X_8^{(2)}$

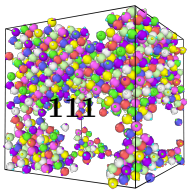
then all the X_α, for whatever patch α, are equal. The equalities of all the X_α are valid for whatever value $x^{(1)}$ (and so $x^{(2)}$) takes. This means that this binary mixture is always an azeotropic binary mixture. Finally, the equality $X_7^{(2)} = X_8^{(2)}$ holds because, since the physics does not change if patch 7 is replaced by patch 8 and patch 3 is replaced by patch 4, then the equations must be invariant under these exchanges satisfying equalities $X_7^{(2)} = X_8^{(2)}$ and $X_3^{(1)} = X_4^{(1)}$. In conclusion the bonding Helmholtz free energy is

$$\begin{aligned}\beta f_{bonding} =& x^{(1)}\left[4\left(\ln X - \frac{X}{2}\right) + \left(\frac{M}{2}\right)\right] + \\ &+ x^{(2)}\left[4\left(\ln X - \frac{X}{2}\right) + \left(\frac{M}{2}\right)\right] = \\ =& \left(x^{(1)} + x^{(2)}\right)\left[4\left(\ln X - \frac{X}{2}\right) + \left(\frac{M}{2}\right)\right] = \\ =& 4\left(\ln X - \frac{X}{2}\right) + \left(\frac{M}{2}\right)\end{aligned} \tag{7.27}$$

where $n(\Gamma(i)) = M = 4$ (with $i = 1, 2$) since we deal with patchy particles species having both four patches. As expected, we notice that, for whatever concentration, $\beta f_{bonding}$ is equal to the free energy of a single-component system.

In order to verify this design, not only theoretically, but also computationally, we construct its phase diagrams with numerical simulations. To study its demixing behaviour, we perform Gibbs ensemble simulations [106, 107]. We initialise two boxes at density $\rho = 0.2$ by randomly placing 500 particles in each. Specifically, the two boxes were prepared by equally and randomly populating them with $x \times 500$ particles of species A and $(1 - x) \times 500$ particles of species B. We explore the concentration range $x \in [0, 1]$ with a step size of $\Delta x = 0.2$ (where x denotes the concentration of the first species A), and we run simulations over a temperature range from $T = 0.125$ to $T = 0.138$. Within this temperature range, coexistence between the two phases is observed. The density profiles for the case $T = 0.138$ and $x = 0.8$ are shown in Figure 7.9 (a), as an example: the two initially identical boxes equilibrate to different values, each corresponding to the density of the liquid and vapour phases. Moreover, as shown in Figure 7.9 (b) for the case $T = 0.138$, each concentration remains equal to its initial value in both boxes, resulting in a density-concentration phase diagram characterised by all tie-lines parallel to the density axis.

For each simulated temperature T and each initial concentration x, we compute the coexistence densities and concentrations by averaging the values assumed by each box at equilibrium. The equilibrium densities become independent of concentration.



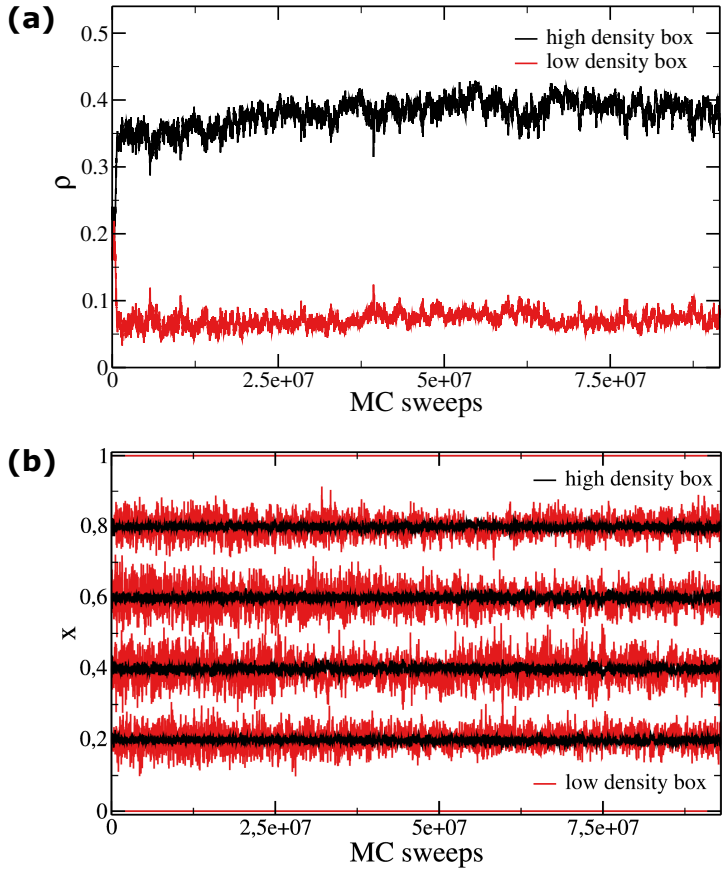


Figure 7.9. Gibbs ensemble simulations of the always azeotropic mixture. (a) The two starting equal boxes with density $\rho = 0.2$, N=500 particles, and temperature $T = 0.138$ rapidly reach an equilibrium configuration with one box (liquid phase) denser than the other one (vapour phases). The figure shows just the $x = 0.8$ case for clearness. (b) Unlike density, the concentration values in both boxes remains constant throughout the simulation. The concentration of the vapour phase (depicted in red) exhibits greater fluctuations than that of the liquid phase (depicted in black) due to the smaller number of particles in the vapour box, but in both cases the initial concentration is preserved. This holds true regardless of the starting concentration, proving that the mixture is always azeotropic. Image adapted from [199].

Consequently, the density-concentration phase diagram, shown in Figure 7.10 (a), is characterised by a bubble point curves and a dew point curve both parallel to each other and to the concentration axis. The system's behaviour as a single-component system is evident also from the temperature-concentration phase diagrams in Figure 7.10 (b), which is plotted for each initial concentration of the two boxes. The overlapping curves demonstrate the absence of concentration dependence.

The concentration independence properties of ideal azeotropes also reflects in their crystallisation behaviour. To check this we run NVT simulations at $T = 0.109$ and $\rho = 0.3$. As illustrated by nucleation snapshots in Figure 7.11, crystallisation, which

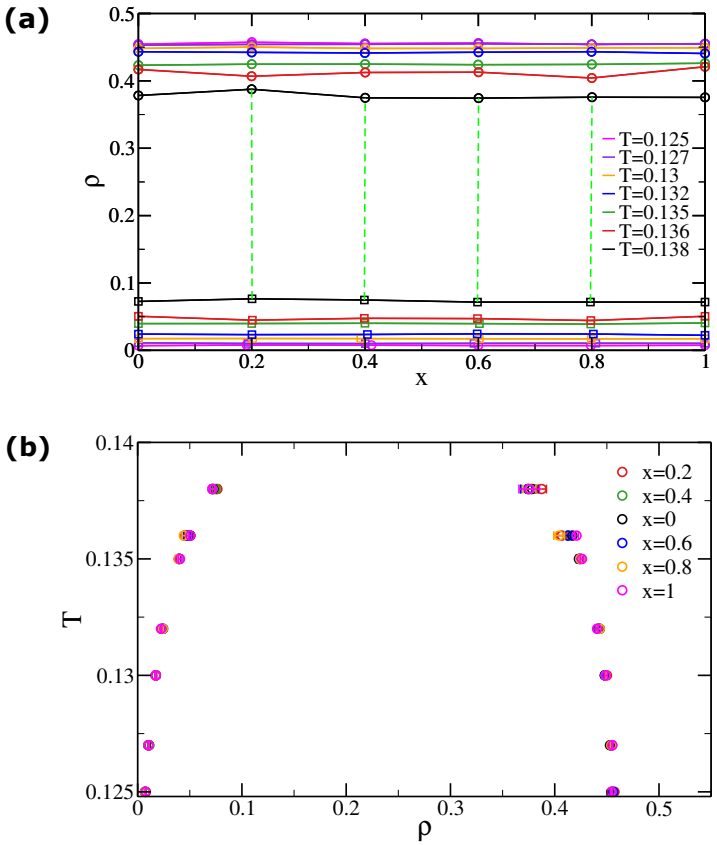
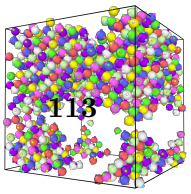


Figure 7.10. Density-concentration and temperature-density phase diagrams of the always azeotropic binary mixture. Monte Carlo simulations with AVB moves in the Gibbs ensemble are performed for each concentration x of the first species, ranging from 0 to 1 in steps of 0.2. The starting configuration consists of two equal boxes with density $\rho = 0.2$ and N=500 particles. We study in parallel temperatures ranging from $T = 0.125$ to $T = 0.138$. Once equilibration is achieved and one box results in a liquid phase while the other in a vapour phase, equilibrium density and concentration values are computed. (a) Density-concentration phase diagrams. Although error bars are omitted for clarity, both the concentration and density errors are less than 10^{-3}. The dew and the bubble point curve are parallel to the concentration axis and the tie lines (green dashed lines) are straight. Regardless of the starting concentration, the two coexisting phases maintain the same concentration value: the mixture is always azeotropic. (b) Temperature-density phase diagrams. The coexistence region is independent of the concentration, confirming that the designed mixture always behaves as a single-component system. Image reproduced from [199].

occurs in the denser liquid phase after phase separation, involves substitutional solids where each lattice site occupied by one species can be replaced by the other. This explains why extended crystals can form at any concentration. These observations support the conclusion that the two species have equivalent bonding probabilities in the mean-field approximation and that the mixture behaves effectively as a single-component system.



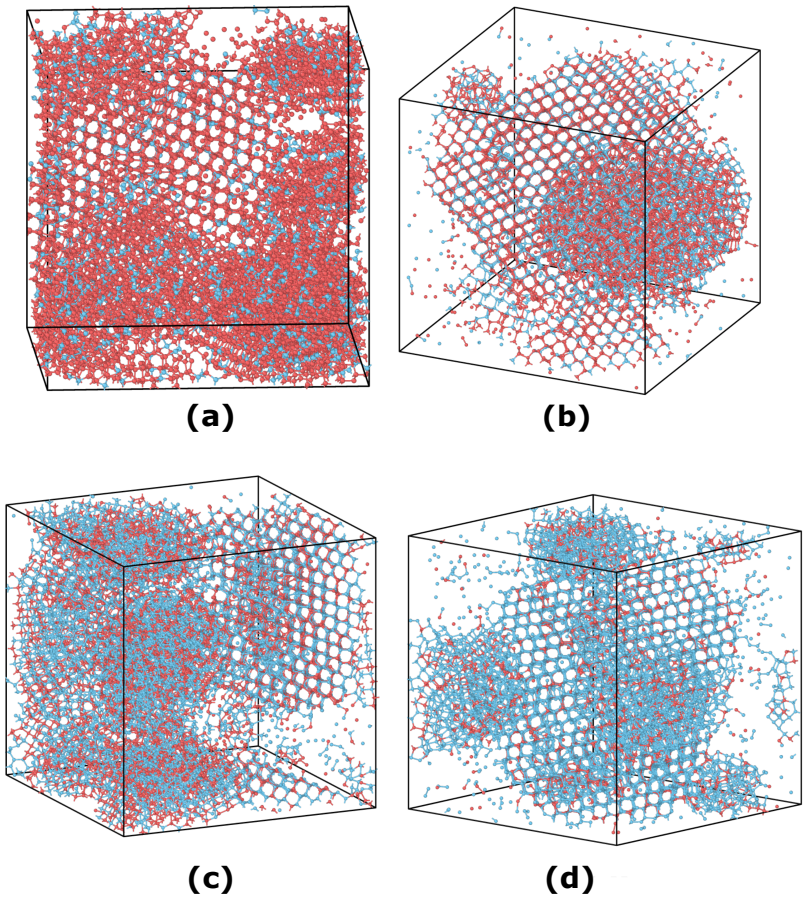


Figure 7.11. Nucleation of the always azeotropic binary mixture. Snapshots of successfully nucleated trajectories at different concentrations x of the first species: $x = 0.2$ in (a), $x = 0.4$ in (b), $x = 0.6$ in (c), $x = 0.8$ in (d). Regardless of concentration, large crystal formation is observed, as each lattice position can be occupied equally by either a particle of the first species (blue coloured) or of the second species (red coloured). Monte Carlo simulations are performed with AVB moves in the canonical ensemble at temperature $T = 0.109$, density $\rho = 0.3$, and number of particles $N = 1000$. Image adapted from [199].

To further confirm the ideal character of the solution, we measure the chemical potential at different compositions with the S0 method. Introduced in Ref. [239], it provides a way of computing chemical potentials of mixtures from structure factors. While leveraging the Kirkwood-Buff [240] theory in relating composition fluctuations in the grand canonical ensemble with chemical potentials derivatives, it departs from it in evaluating such equilibrium fluctuations. In fact, instead of using volume integrals over the radial distribution function, static structure factor are employed. This avoids boundary effects and the need of finite size corrections. An overview of the Kirkwood-Buff theory and the S0 method is provided in Appendix E. The following equation for the excess chemical potential of component A in a binary

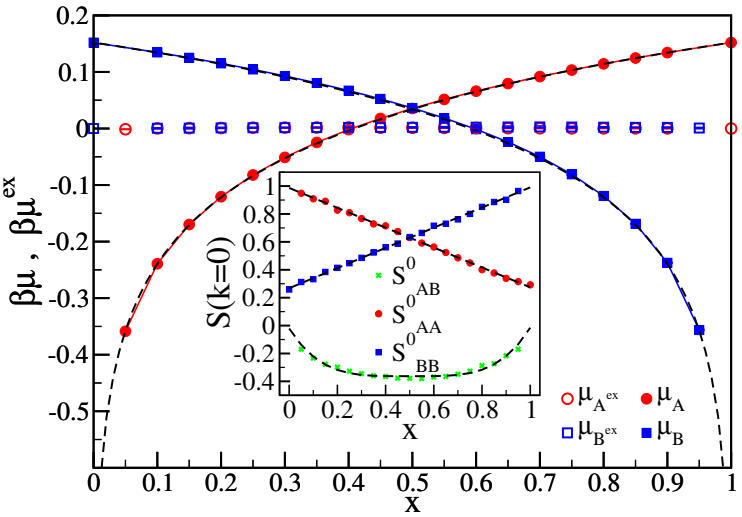


Figure 7.12. Chemical potentials from the S0 method. The total chemical potential μ (full symbols) and the excess chemical potential μ^{ex} (empty symbols) of both species highlight the near-ideal behaviour of the always azeotropic mixture. In the inset, the zero-wave number ($\mathbf{k}$=0) limit of the three partial structure factors S^0_{AA}, S^0_{BB} and S^0_{AB} are plotted. The dashed lines indicate the analytic expression fitted to the data and later on used to solve Eq. 7.28 for μ^{ex}. Image reproduced from [199].

mixture is obtained.

$$\mu_A^{ex}(x_A) = k_B T \int_1^{x_A} \frac{dx_A}{x_A} \left[\frac{1}{S^0_{AA} - S^0_{AB}\sqrt{x_A/x_B}} \right] \tag{7.28}$$

Then, the whole chemical potential is obtained according to:

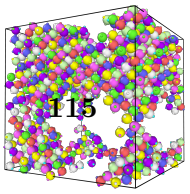
$$\mu_A(x_A) = \mu_A^0 + k_B T \ln\left(\frac{x_A}{x_A^0}\right) + \mu_A^{ex}(x_A) \tag{7.29}$$

where x_A (x_B) denotes the molar fraction of species A (B), and μ_A^0 is the chemical potential at a standard molar fraction. We selected this reference to be the pure state $x_A^0 = 1$. Simulations in the NPT ensemble are first carried out for different molar fractions to compute the structure factors S_{AA}, S_{BB}, and S_{AB}. Then, the values of S^0_{AA}, S^0_{BB}, and S^0_{AB} in $\mathbf{k} = 0$ are extrapolated by fitting the Ornstein-Zernicke relation [241], given by:

$$S_{AB}(\mathbf{k}) = \frac{S^0_{AB}}{1 + k^2 \xi_A \xi_B} \tag{7.30}$$

where ξ_i is the correlation length of species i. Finally, by numerically evaluating the integral in Eq. 7.28, the excess chemical potential is calculated.

Simulations are performed in the NPT ensemble for species concentrations in the



range $x \in [0.05, 1]$. The simulation parameters used are: $N = 4000$ particles, pressure $P = 0.125$, and temperature $T = 0.17$. The system density for these parameters is found to be $\rho \sim 0.352$. The $\mathbf{k} = 0$ values of the determined structure factors $S_{AA}(\mathbf{k})$, $S_{BB}(\mathbf{k})$, and $S_{AB}(\mathbf{k})$ are shown in the inset of Figure 7.12, where the black dashed line represents the fit performed on the obtained data according to Eq. 7.30. The excess chemical potential is computed using Eq. 7.28, with uncertainties estimated from variations in the fitted partial structure factors shown in the inset of Figure 7.12. To determine the standard chemical potential for pure components, we perform grand canonical ensemble simulations at temperature $T = 0.17$, systematically varying the chemical potential until the measured density matches that obtained from NPT simulations. This yield a standard chemical potential of $\mu^0 = 0.152$. As shown in the main panel of Figure 7.12, the total chemical potentials closely follow the ideal value (dashed black lines). This near-ideal behaviour confirms the system's ideal azeotropic character, where composition changes minimally affect the excess chemical potential that constantly remains close to zero across all concentrations.

The discussion of a design that generates an always azeotropic mixture concludes our set of examples illustrating azeotropic bonding rules based solely on bond connectivity. The next subsection presents a case where bond energy control comes into play.

7.5.5 Bonding energy control

Modifying the bond topology of a design has proven effective for precisely tuning azeotropic concentrations. However, when the bond connectivity cannot be changed, in order to achieve the targeted azeotropic concentration, another route becomes necessary. We propose to manipulate the bonding energies. This situation occurs, for instance, when we have already designed a mixture to exclusively self-assemble a specific structure, and we want to change its azeotropic behaviour. Because lattice selectivity is inherently defined by the bond topology, altering the connectivity would

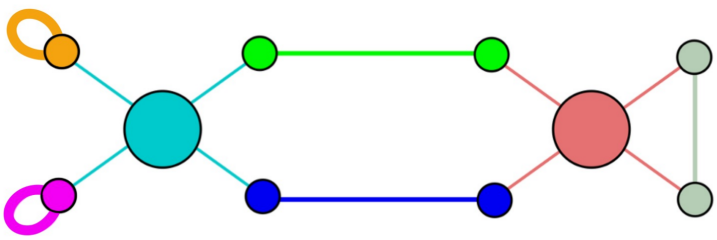


Figure 7.13. The N2c8 design with differentiated bonding energies. Compared to the original N2c8 design illustrated in Figure 7.1, the edge strength varies across nodes. Specifically, self-interacting connections are emphasised in bold to indicate that the energies of these self-interacting patches are higher relative to the others.

compromise the system's ability to nucleate the targeted structure. For example, how can we change the bonding energies of the N2c8 binary mixture (Figure 7.1) to shift its azeotrope from concentration equal to $x = 0.5$ to concentration equal to $x = 0.6$?

By applying the methodology described in Section 7.4 we find it is possible by modifying the interaction parameter of the self-interacting patches 2 and 3 from $\epsilon_{22} = \epsilon_{33} = 1$ to $\epsilon' = 1.35$. This prediction is verified through the following theoretical and numerical analysis. First, we compute the pressure-concentration and density-concentration phase diagrams within Wertheim's theory. They are plotted in Figure 7.14 (a) and Figure 7.14 (b), respectively. For comparison, these diagrams are superimposed with those of the original N2c8 mixture computed in Chapter 4. The theoretical results demonstrate that while the original system shows

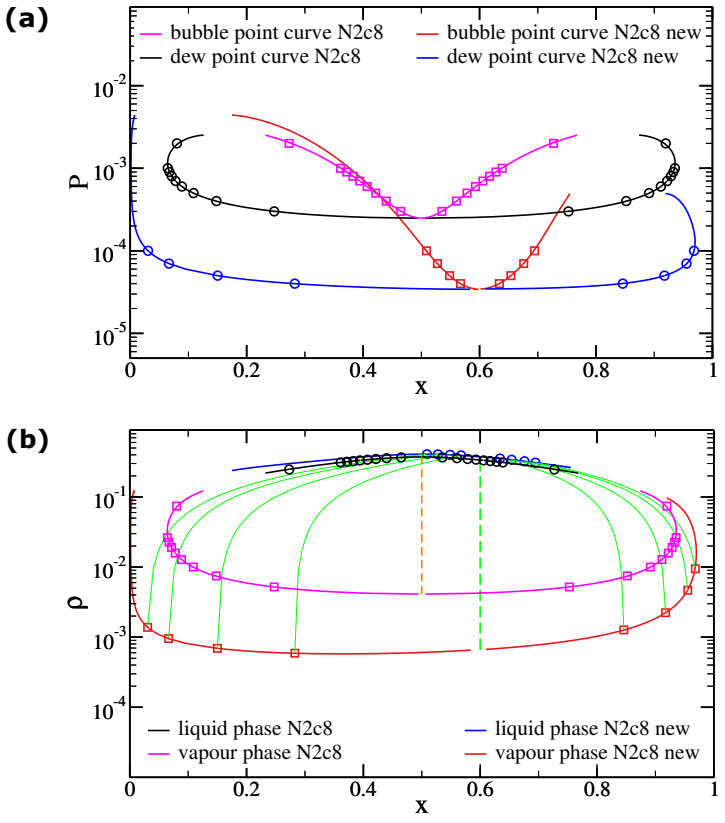
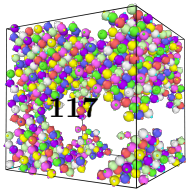


Figure 7.14. Phase diagrams of N2c8 mixtures for different bonding energies. Phase diagrams comparing two interaction schemes at $T = 0.08$: (1) uniform $\epsilon_i = 1$ (black-magenta) and (2) differentiated $\epsilon_i = 1$, $\epsilon_j = 1.35$ (blue-red) for self-complementary interactions. The azeotropic point shifts from $x = 0.5$ to $x = 0.6$ through selective bond energy modification while preserving the topology. (a) Pressure-concentration diagram shows the azeotrope as the point where coexistence region collapses. (b) Density-concentration diagram identifies the azeotrope via vertical tie-lines (dashed lines). Image reproduced from [199].



azeotropic behaviour at concentration equal to $x = 0.5$ (manifested as a single point where the coexistence region collapses), the modified system with $\epsilon' = 1.35$ exhibits this azeotropic point at $x = 0.6$. This shift is equally evident in the density-concentration diagram, where the characteristic straight tie-line (dashed) moves accordingly from $x = 0.5$ to $x = 0.6$.

To validate these theoretical predictions, we also perform Gibbs ensemble Monte Carlo simulations at temperature $T = 0.098$ and density $\rho = 0.1$. Systems are prepared with varying initial concentrations ($x = 0.2, 0.4, 0.6, 0.8$) in two equal simulation boxes. As shown in Figure 7.15, only the system with $x = 0.6$ maintains constant concentration in both boxes throughout the simulation, confirming this concentration as the new azeotropic one. This agreement between theoretical predictions and numerical simulations provides robust verification of our design strategy. It is worth noting that Gibbs ensemble simulations, are a method ideally suited for mapping phase diagrams of complex fluids with minimal computational cost and for rigorously validate our predictions without determining the full phase diagram. In fact, unlike brute-force free energy calculations, the Gibbs ensemble directly samples coexistence conditions, bypassing the need for thermodynamic

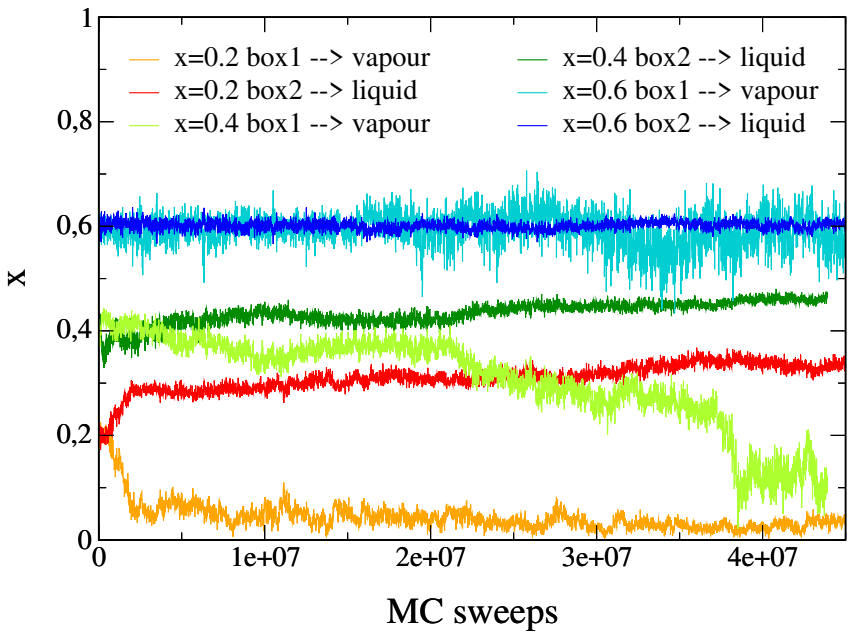
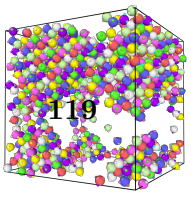


Figure 7.15. Gibbs ensemble simulations for the new N2c8 binary mixture. For different concentration of the first species ($x = 0.2$, $x = 0.4$, $x = 0.6$), we initialised two equal boxes with density $\rho = 0.1$, temperature $T = 0.098$ and $N = 500$ particles. By tracking the concentration of the first species throughout the simulations, we observe that the equilibrium value remains the same as the initial one when the two boxes are prepared at $x = 0.6$. The new azeotrope is at concentration equal to$x = 0.6$. Image reproduced from [199].

integration, a critical advantage for multi-component [90, 242].

In this chapter, we introduced the idea of inverse thermodynamics. We explained how rules acting on bond connectivity and bonding energies can effectively lead to a precise control over desired thermodynamic properties of multi-component systems. This, in turns, facilitates a more effective solution to the inverse self-assembly challenge [243–245]. We focused on azeotropy, a special thermodynamic condition where the system behaves as a single-component systems and self-assembly is promoted if the azeotrope appears at concentrations matching the ones of the target lattice. Importantly, these rules can be integrated, as additional clauses, directly in the SAT-assembly framework, which therefore can be used to insert also thermodynamic requirements in addition to the structural ones. In the last chapter, we explore how to control also the photonic properties of the assembly.



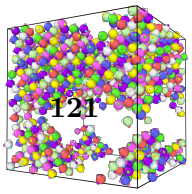
Chapter 8

Controlling optical activity of multi-component assemblies

The SAT-assembly framework has proved effective in designing interactions of multi-component systems to achieve specific structural and thermodynamic properties. It can be leveraged to approach experiments with an optimal and refined design that ensures the target structure is the unique outcome and forms through a pathway that promotes the self-assembly process. In this final chapter, we aim to further tailor the assembly by establishing, always by design, the onset of desired photonic properties.

8.1 The importance of isotypes

Colloidal structures can exhibit remarkable photonic properties when there is a periodic spatial modulation of the refractive index [246]. Among these, the cubic diamond structure, the prototypical example in this work, can function as a photonic crystal when realised at colloidal scales using DNA-origami scaffolded particles or surface functionalised nanoparticles [49, 50, 247]. The photonic band gap, the range of frequencies forbidden from propagating through the crystal, can be tuned by adjusting the size of the origami or nanoparticles and by carefully selecting the materials of both the particles and the filling medium, thereby controlling the refractive index contrast. Importantly, this tuning can be achieved even in single-component systems. But what about mixtures? Can they give rise to additional or enhanced photonic properties relevant to technological applications?
Mixtures are a central focus of this work. They are at the core of the SAT-assembly design strategy and are responsible for the emergence of isotypes, thermodynamically degenerate polymorphs that have been fundamental to build our falsifiability test for Classical Nucleation Theory, as detailed in Chapter 6. These isotypes arise from



different arrangements of the same set of components on identical lattice sites. This suggests that more tailored optical functionalities can be achieved by appropriately filling a crystal lattice with different particle species. For example, by intentionally designing chiral motifs, we can induce intrinsic optical activity within the cubic diamond structure, effectively transforming it into a polarising material. In this chapter, we focus on this interesting case, and we present a chiral cubic diamond crystal that can be realised with two different approaches.

8.2 A chiral isotype of the cubic diamond crystal

The importance and potential of different species arrangements from a photonic perspective become evident in our leading example of the SAT-designed N2c8 binary mixture, summarised in Figure 8.1. Through Monte Carlo simulations in the canonical ensemble, we confirmed its ability to exclusively self-assemble into a cubic diamond crystal (Chapter 5) and we observed the emergence of three isotypes, illustrated in Figure 6.1. Among these polymorphs we recognise the presence of chiral patterns in the DC-16 structure. As detailed in Figure 8.1, the red species exhibit helical arrangements. The same is true for the blue species if we set a different cubic plane as the observation plane. This suggests the formation of a chiral cubic diamond crystal.

To assess whether the structure exhibits the expected chiroptical properties, we compute its band structure through the MIT Photonic-Bands software MPB [248]. We assign a refractive index of $n = 2.6$ to particles of the first species and $n = 1.5$

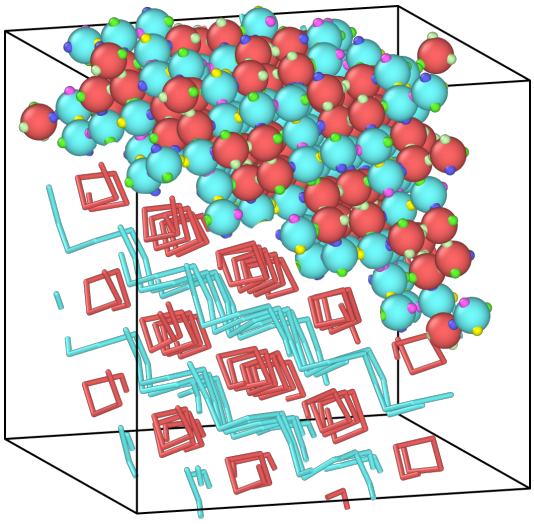


Figure 8.1. Chiral cubic diamond crystal. Within the cubic diamond unit cell, the two species (red and cyan coloured) adopt distinctive helical arrangements that induce chirality in the crystal structure.

to those of the second species.

The band structure is computed by solving Maxwell's equations in a periodic medium, considering that the system is in linear regime and that the material is isotropic, transparent, and non-magnetic. Since equations are linear, we can separate the spatial and temporal dependence, and assume the electric field varies harmonically in time: $E(\mathbf{r}, \mathbf{t}) = \mathbf{E}(\mathbf{r})\mathbf{e}^{-\mathbf{i}\omega\mathbf{t}}$. This holds because any field can be expanded as a combination of harmonic modes. Thus, we can focus on solving for the spatial dependence. These assumptions lead to a master equation in the form of an eigenvalue problem, where the eigenmodes describe the spatial variation of the electric field, and the eigenvalues correspond to the allowed mode frequencies. The dispersion relations are then obtained by sampling the Brillouin zone at various wave vectors $\mathbf{k}$ and evaluating the eigenvalues at those points. In particular, due to the periodicity of the structure, Bloch's theorem is applied: the solutions to the wave equation take the form $E(\mathbf{r}) = \mathbf{e}^{\mathbf{ik}\cdot\mathbf{r}}\mathbf{u}(\mathbf{r})$, where $u(\mathbf{r})$ is a periodic function with the same periodicity as the lattice. The resulting band structure along the $\Gamma \to Y$ direction of the conventional unit cell is shown on the left side of Figure 8.2. We observe a band gap around the normalised frequency $\frac{fa}{c} \sim 0.85$, where a is the lattice constant,

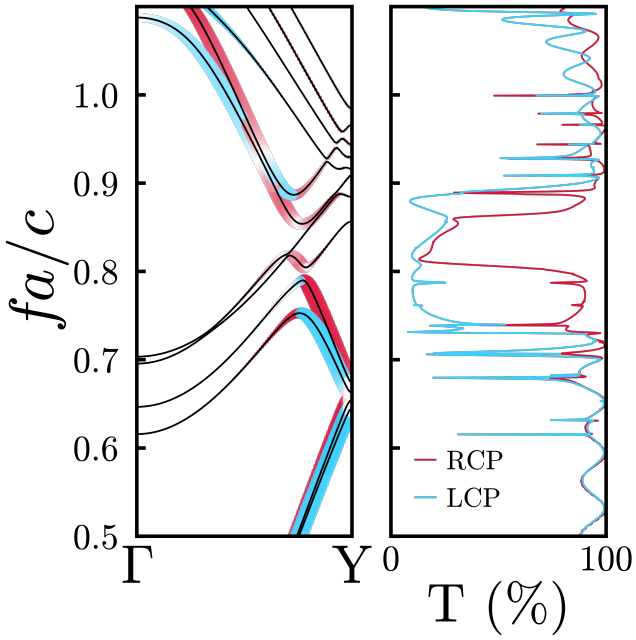
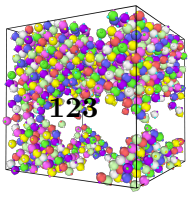


Figure 8.2. Chiroptical properties of chiral cubic diamond crystal. On the left is the partial photonic band structure of the conventional unit cell of the cubic diamond crystal in Figure 8.1 with cyan particles having a refractive index of $n = 2.6$ and red particles of $n = 1.5$. The colours of the bands reflect the chirality of the eigenmodes. The right panel shows the transmission spectra for right- and left-circularly polarised light along this direction, in red and cyan colours, respectively.



and c is the speed of light in vacuum. The photonic band structure computation is accompanied by a polarisation analysis. The handedness of the bands is quantified using the circular dichroism index, which is defined from the degree of right/left circular polarisation of each eigenmode by using the overlap integral between the Bloch mode under analysis and a right or left circularly polarised plane wave incident perpendicularly to the mode. Hence, this index measures the preference of a mode to couple with radiation of specific circular polarisation. For all details we refer to papers [249–251]. Results are reported on the left side of Figure 8.2 assigning a colour to the bands: red for right-handed and cyan for left-handed. Based on this analysis, two chiral stop bands can also be identified in the frequency ranges $\frac{fa}{c} \sim 0.75 - 0.8$ and $\frac{fa}{c} \sim 0.87 - 0.9$. On the left side of Figure 8.2, we plot the transmittance spectra for right circular polarised (RCP), in red, and left circular polarised (LCP), in cyan, plane waves hitting a finite crystal slab at normal incidence. They are computed using the finite difference time domain package MEEP [252]. It simulates the time response of a material to an incident electric field by numerically solving Maxwell's equations in the time domain. Typically, a short broadband pulse (often Gaussian) is introduced towards the structure. In this way, by Fourier transforming the resulting time domain signal recorded at a transmission monitor, the spectrum of transmitted radiation can be obtained over a broad frequency range in a single simulation run. These spectra confirm the behaviour inferred from the polarisation resolved band structure: at the $\frac{fa}{c} \sim 0.85$ band gap, transmission of both RCP and LCP is suppressed, whereas within the chiral stop bands, only RCP is transmitted through the crystal.

The chiroptical properties of the DC-16 chiral cubic diamond are then verified, and in the following section we illustrate a strategy to ensure its exclusive assembly.

8.3 A heterogeneous nucleation approach

The N2c8 design, while capable of assembling into the DC-16 chiral cubic diamond, can also give rise to the DC-8 and the DC-24 polymorphs (Figure 6.1), whose species arrangements do not introduce any additional photonic property beyond those due to the cubic diamond geometry. As a result, this binary mixture has the drawback of nucleating also in these two other, non-chiral structures. In general, to reliably assemble a specific polymorph, the DC-16 one in the presented case, the only viable strategy for a given design is to employ a heterogeneous nucleation approach. Specifically, we must determine whether it exists a small seed of particles, distinct for each polymorph, that can exclusively assemble into the desired structure without transitioning into a different arrangement during the

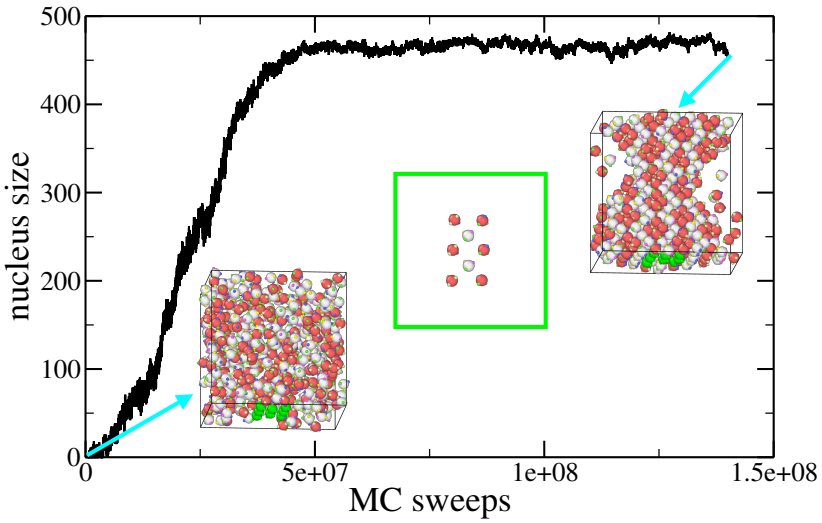
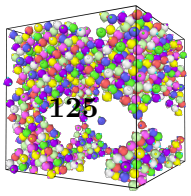


Figure 8.3. Nucleation of chiral cubic diamond from seeding. Time evolution of the number of crystalline particle for a seeded trajectory in the canonical ensemble at temperature $T = 0.104$, density $\rho = 0.23$, and number of particles $N = 500$. The system is seeded with a plane of 8 particles (green box), as shown in the snapshot of the initial configuration. The final configuration confirms that the chiral cubic diamond can grow readily and selectively from the chosen seed.

growth. By examining, for simplicity, the cubic plane of the three polymorphs, we identify a planar seed composed of eight particles, with a unique configuration for each case. Figure 8.3 illustrates the seed corresponding to the chiral DC-16 polymorph, oriented along the cubic $[1, 0, 0]$ plane, in the green framed inset. Figure 8.4 does the same for the DC-8 and the DC-24 structures, in (a) and (b), respectively. To assess whether these seeds are valuable candidates, we perform Monte Carlo simulations in the canonical ensemble at a state point favourable to nucleation (temperature $T = 0.104$, density $\rho = 0.23$, with 500 particles). Initial configurations are prepared by placing the seed at the bottom of a cubic simulation box with periodic boundary conditions, while the remaining particles are randomly distributed in position and orientation. A snapshot of such initial configuration is displayed in the bottom left corner of Figure 8.3 for the DC-16 case and in the second panels of Figure 8.4 for the DC-8 and DC-24 polymorphs. During the simulation, the seeded particles are constrained to prevent their unbinding. To mimic experimental conditions as closely as possible, the seed is not kept completely fixed; instead, its particles are allowed to move within harmonic constraints in both translations and rotations. In practice, roto-translation moves are attempted for all particles, but, for the seeded particles, additional harmonic potential terms are introduced in the potential energy.



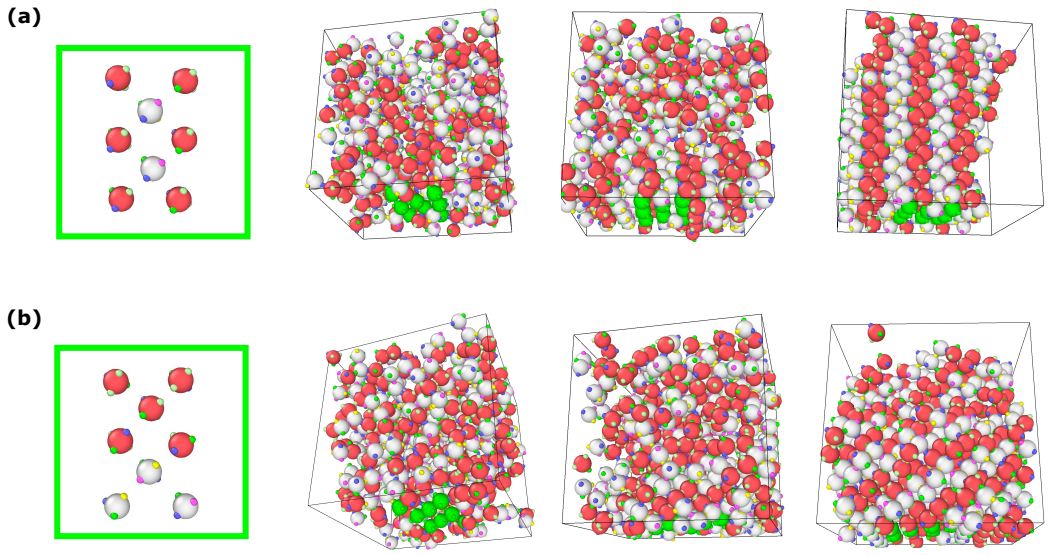


Figure 8.4. Nucleation of DC-8 and DC-24 from seeding. Snapshots across time of configurations from seeding simulations in the canonical ensemble at temperature $T = 0.104$, density $\rho = 0.23$, and number of particles $N = 500$. The DC-8 polymorph in (a) and the DC-24 polymorph in (b) effectively grow from each initial seed consisting of 8 particles in a plane (green boxes).

Specifically, harmonic oscillators are applied to control displacements along the three Cartesian axes, and rotations defined by Euler angles. The translational and rotational stiffness constants are distinct and can be tuned to accurately reflect experimental behaviour.
In Figure 8.3, we track the evolution of the number of crystalline particles for a single simulated trajectory. As desired, the planar seed of eight particles readily initiates the nucleation of the chiral cubic diamond, whose assembled structure is depicted in the snapshot in the top right corner [253]. For completeness, in Figure 8.4 we display a series of snapshots from seeded trajectories resulting in the DC-8 and DC-24 polymorphs, respectively. It is worth noting that alternative seeds with fewer particles are also possible, potentially simplifying the experimental realisation. For example, we identified a viable seed consisting of only four particles. However, its nucleation frequency is significantly lower compared to the eight particle seed. Therefore, the optimal seed size represents a trade-off between feasibility and performance.

In summary, the N2c8 design can selectively assemble the chiral cubic diamond crystal only indirectly, by seeding the nucleation. In the next section, we explore how to nucleate the DC-16 structure by design, looking for a new interaction matrix that is not compatible with the other non-chiral arrangements, through the SAT-assembly framework.

8.4 An homogeneous nucleation approach

The direct formation, *i.e.* via homogeneous nucleation, of a cubic diamond crystal with helical patterns requires a solution different from the N2c8 one and that can suppress competing polymorphs by design. However, excluding the other two possible arrangements is not straightforward. To avoid the formation of alternative crystals, such as the hexagonal diamond, we can reformulate the SAT-assembly problem to exclude a particular solution if it is also compatible with the undesired topology, as already explained in Chapter 2. However, our two adverse polymorphs of the cubic diamond, being isotypes, do not differ in their topology but only in the arrangements of the species on the same lattice sites. As a result, the traditional approach is ineffective in this case. Additionally, another challenge arises: the N2c8 design is the minimal solution to the inverse self-assembly problem for cubic diamond crystal, using the smallest number of distinct species and patch types. Seeking an alternative design thus requires increasing the number of building units. This, in turns, necessitates verifying whether all possible arrangements, for each subgroup choice that reduces the design to two species, display helical patterns.

We adopt a brute-force approach: we constrain the lattice positions that a specific

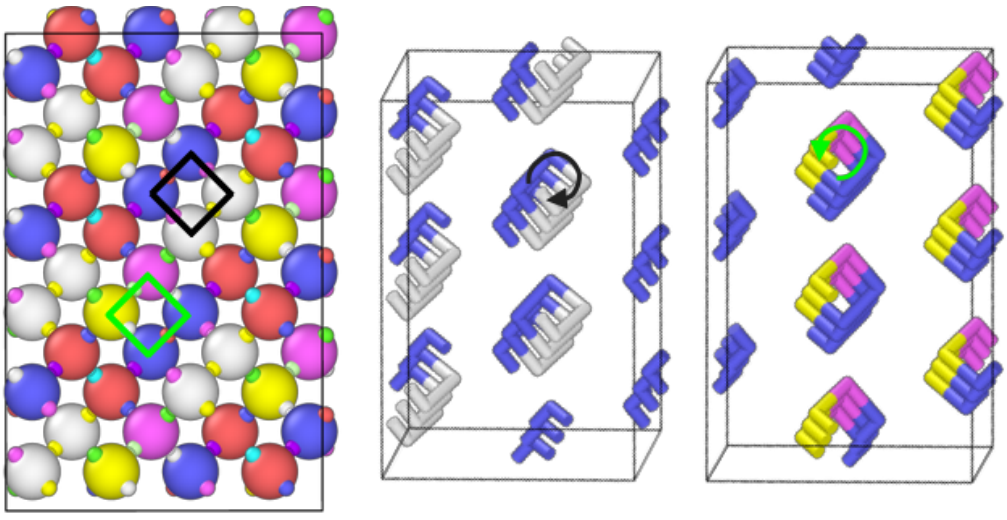
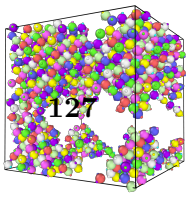


Figure 8.5. Selective design for chiral cubic diamond. A design exclusively compatible with the chiral cubic diamond lattice is obtained by increasing the number of distinct species to 5 and assigning 15 different colours. Interestingly, a different grouping of particles species gives rise to opposite handedness, as indicated by the black and green squares in panel (a) and detailed in panels (b) and (c) showing the bond arrangement of each subgroup.



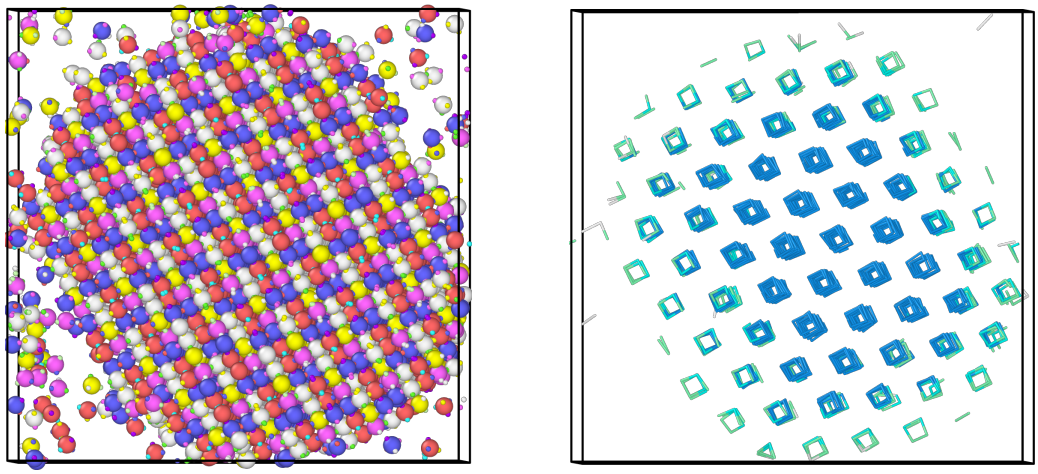
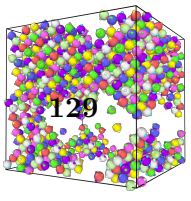


Figure 8.6. Self-assembly of chiral cubic diamond. (a) The snapshot on the left shows the successful growth of a cubic diamond crystal obeying to the N5c15 design, SAT-determined to exclusively self-assemble a chiral structure. (b) The snapshot on the right clearly highlights the helical motifs that impart chiroptical properties to the structure. Particles coloured in blue are identified as having cubic symmetry with Ovito software [149]. Simulations are performed in the canonical ensemble at temperature $T = 0.102$, density $\rho = 0.25$.

species can fill, thereby directly drawing the desired pattern. Specifically, in the SAT-assembly framework, we progressively increase the number of species and patch types, while also imposing that the first two species cannot occupy helical positions. This can be done straightforwardly by acting on the unit cell placement variables $x^L_{l,s,o}$ appearing in the boolean clauses. If L_{chi} is the set of lattice positions where we aim to observe helices, then the additional clauses are: $\neg x^L_{l,s,o}$ for all $l \in L_{chi}$, $s \in [0,1]$, $o \in R$, and $\neg x^L_{l,s,o}$ for all $l \notin L_{chi}$, $s \notin [0,1]$, $o \in R$. We always target a 16 particle cubic diamond unit cell, while checking designs against a 32 particle hexagonal diamond unit cell. We stop when we find a solution in which all possible unconstrained rearrangements of species over the lattice sites always result in our desired chiral pattern. Following this procedure, we identify the N5c15 solution depicted in Figure 8.5 (a), which uses five species and fifteen different patch types. This solution can be reduced to the original binary mixture by mapping species 0 and 1 (white and blue coloured, respectively) to the original species 1 (in red in Figure 8.1), and species 2,3, and 4 (blue, yellow and magenta coloured, respectively) to the original species 0 (in cyan in Figure 8.1). This last subgroup (green square in Figure 8.5 (a)) defines the target chiral pattern, as shown in Figure 8.5 (c). Interestingly, the first subgroup (black square in Figure 8.5 (a)) also defines a chiral pattern as illustrated in Figure 8.5 (b), while it has the opposite twist direction with

respect to the original one. This opens to a further tunability of photonic properties since by selecting different subgroup of species, we can change the handedness of the chiral crystal. Figure 8.6 (a) and Figure 8.6 (b) illustrate two final snapshots confirming the large-scale formation of a chiral cubic diamond crystal with the N5c15 design. Monte Carlo simulations are performed in the canonical ensemble at temperature $T = 0.102$, density $\rho = 0.25$.

In conclusion, we illustrated how multi-component systems can originate assemblies with captivating optical properties. In particular, we investigated the design of chiral patterns by controlling the particle arrangement over lattice sites. As for binding rules for azeotropy appearance, the restrictions on the positions that different species can occupy are easily incorporated into the SAT-assembly framework by adding clauses to the unit cell placement variables $x^L_{l,s,o}$. This final chapter, by providing photonic consideration, completes our analysis of inverse self-assembly in multi-component systems. We have shown the tunability of soft matter systems and the efficiency of the SAT-assembly framework in designing building unit to achieve desired structures with precise thermodynamic and optical properties. To summarise, our framework allows to: automatically identify the minimal solution able to self-assemble exclusively into a target structure, while avoiding alternative assemblies; exhaustively enumerate all polymorphs of a given design, regardless of their nucleation abilities; insert additional clauses for specific thermodynamic and photonic requirements, such as the presence of an azeotrope or of chiral structures. Together, these capabilities underscore the effectiveness of SAT-assembly as a general tool for inverse designing complex materials.

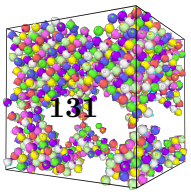


Chapter 9

Conclusions

To fulfil the potential of nanotechnology in engineering self-assembled materials with desired mechanical, optical and thermodynamic properties, we sought general principles underlying the crystallisation of multi-component systems, which can be leveraged to optimise the process. Obtaining a complex target structure, such as a photonic crystal, requires either carefully optimised interaction potentials that make the target a free-energy minimum, or a mixture of simpler building blocks that render the desired structure accessible by reducing its overall symmetry. Moreover, it is essential to avoid competing local minima that could kinetically frustrate the assembly process. In recent years, it has become increasingly clear that using multi-component systems can shift the challenge from accurately designing the shapes of the building units and the interaction potentials, *e.g.* by introducing torsional interactions, to optimising a generic interaction matrix between the different components. Recent advances in DNA-nanotechnology, which enables the design of building blocks with arbitrary interaction rules, are promising, opening the door to this latter approach.

Several strategies have been proposed to design interactions in multi-component systems. One such strategy is the **SAT-assembly framework**, which translates the design problem for **patchy particles** into a Boolean satisfiability problem. This effectively allows handling the **combinatorial problem** inherent in exploring all possible designs, as the interactions between the different components are amenable to numerical solutions. However, enlarging the number of species also opens up a **thermodynamic challenge**. Each additional component increases the thermodynamic degrees of freedom, which considerably complicates the phase behaviour and the assembly pathways. Hence, the ability to design interaction matrices exclusively compatible with a target structure topology is a necessary but not sufficient condition for observing the nucleation of such structures under experimental conditions. These often involve preparing the system in very dilute phase, from which the classical

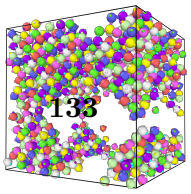


nucleation rate is severely suppressed by both **diffusional and compositional barrier**.

Regarding the diffusional barrier, probably the most promising strategy to overcome it is **two-step nucleation**, a non-classical nucleation pathway where crystallisation is aided by the formation of dense liquid drops. This mechanism has been the subject of intense research, but especially in the context of one-component systems. So far not much work has been devoted to mixture and particularly to the second barrier, the compositional barrier, which controls the rate at which a particle of the correct species enters in contact with the incipient nucleus. In this work, we have presented a first study in this direction by using patchy particles, which represent a convenient model for all systems whose interactions can be described by an isotropic repulsion and strong directional attractions. We have examined the two-step nucleation process in a binary mixture of patchy particles designed to assemble exclusively into a cubic diamond crystal. In particular, we have investigated the role that the metastable liquid-vapour phase diagram plays in nucleation. Comparing Gibbs ensemble simulations with direct nucleation runs, we have established that nucleating state points coincide with the locus of metastability of the mixture, confirming the relevant role played by the thermodynamic instabilities, generating local environments with a density enhanced in comparison to the average value. Moreover, we have shown that compositional effects in multi-component systems play a crucial role in the nucleation and growth of crystalline nuclei: crystallisation is enhanced at state points where the emerging liquid phase has a composition close to that of the crystal. In fact, we have observed that, because the crystal has a fixed stoichiometric composition, the formation of a liquid phase at other compositions slows down and limits the growth of the crystal. This presents a major challenge in mixtures where the composition of the liquid phase resulting from phase separation is not known a priori and typically differs from that of the target structure. Unlike single-component systems, which are characterised by a coexistence line, we have highlighted that mixtures exhibit a coexistence region. As a result, phase separation does not preserve the original ratio between the components and determining these coexisting compositions requires computing the phase diagram, which is often extremely challenging. We have demonstrated, within Wertheim's perturbation theory and through numerical simulations, that many of these thermodynamic difficulties can be removed by employing **azeotropic mixtures** and by selecting a self-assembly pathway along an azeotrope that matches the stoichiometry of the target crystal. We have illustrated that two-step nucleation pathways in systems prepared under azeotropic conditions are analogous to those of one-component systems and lead to optimal self-assembly. The coexisting liquid phase

naturally adopts the ideal composition for templating the crystal, thereby overcoming compositional barriers and significantly boosting assembly rates. Consequently, we have revealed the importance of explicitly including azeotropic points in artificial designs as a crucial step toward establishing of a fully consistent framework for the self-assembly of arbitrary structures. We have proven that two-step nucleation, augmented with azeotropic conditions, serves as a key guiding principle for achieving optimal inverse self-assembly. Interestingly, we have also observed that nucleation in mixtures, conversely to single-component systems, not occur close to critical points. Nevertheless, we have found evidence of critical-like enhanced nucleation along the binary critical line when it intersects the azeotropic line, confirming the unique thermodynamic character of the azeotrope. In this condition, the system, regardless of the number of components, behaves as a single-component system, with the liquid and the vapour phases coexisting at the same initial composition.

We have then extended the inverse design paradigm beyond structure alone to include selected thermodynamic properties. We have presented a framework for **inverse thermodynamics**, *i.e.* the design of interaction potentials that realise specific thermodynamic behaviours. We have demonstrated how to program azeotropy in multi-component systems through precise manipulation of microscopic interactions. Methodologically, we have combined Wertheim's thermodynamic perturbation theory with Gibbs ensemble simulations. This approach has avoided computationally expensive full phase diagram calculations; instead, just a handful of targeted simulations have sufficed to validate our azeotropic rules by confirming concentration stability only under coexistence at the azeotrope. Rules for including azeotropy directly into the design have been identified by exploiting the laws of mass-action within the Wertheim perturbation theory. Specifically, we have presented design principles for azeotropic demixing which involve complementary approaches acting both on **bond topology** design, which alters the connectivity rules between components, and on **bond energy** modulation, which selectively tunes interaction strengths. The simplest condition on bond connectivity is the *bond exclusivity* condition, asserting that an equimolar azeotropic point can be obtained by imposing that each patch has a unique interaction partner. The equimolar condition can be relaxed, and the azeotropic point can be located at a desired concentration vector $\mathbf{x}$, by considering the *bond multiplicity* condition, which requires some patches to have more than one possible interaction partner. Finally, the *fully-connected bond* condition, where each patch has one interaction partner on each of the species of the system, corresponds to an *ideal azeotrope* occurrence, *i.e.*where demixing occurs azeotropically at any concentrations. Our work aimed to bridge the gap between forward and inverse thermodynamics. By establishing design rules for azeotropicity, we have provided a



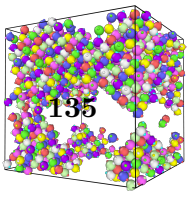
blueprint for tailoring phase behaviour in self-assembling systems. More broadly, our approach has exemplified how soft matter's tunability can be harnessed to solve thermodynamic challenges that we face when dealing with multi-component systems. The principles of inverse thermodynamics, which enable precise control of phase behaviour, can synergistically enhance solutions to the inverse self-assembly problem. Crucially, we have underlined that both inverse thermodynamics and inverse self-assembly can be formulated through Boolean satisfiability constraints on the building blocks' interactions. This unified computational framework, termed SAT-assembly, has already demonstrated remarkable success in engineering diverse nanostructures, including novel crystals, viral capsids, quasi crystals, and amorphous materials. Together, these advances establish a paradigm for **materials by design**, where thermodynamic pathways and structural outcomes can be programmed in tandem to unlock new functional materials.

We have also shown that the SAT-assembly framework can be leveraged to program specific photonic properties. In particular, we have focused on **chirality** design, and we have proposed a way to guarantee the appearance of helical structures by requiring specific lattice positions to be occupied by specific species. This strategy is readily implemented in SAT-assembly by introducing supplementary clauses that constraint the placement of the different species.

The idea of leveraging the different arrangements that the same species can adopt on identical lattice sites, an aspect unique to multi-component systems, has also been crucial in designing a **falsifiability test for Classical Nucleation Theory** (CNT) and for any extension of the theory that do not discard the capillarity approximation. We have analysed a system in which all polymorphs originate from the same liquid phase and have the same bulk and interfacial free energies to investigate a scenario in which CNT predicts no differences in the nucleation rates of the polymorphs. Conversely to predictions, through extensive molecular simulations, we have observed that they exhibit remarkably different nucleation frequencies. Crucially, unlike traditional tests that compare measured nucleation rates to theoretical predictions, we did not rely on hard to measure and state dependant quantities, but we only needed to rank the structures based on how frequently their nucleation occurred. Our observations have revealed that the melt exhibits a local orientational order typical of the polymorph with the highest nucleation rate. This has suggested that the structural fluctuations in the melt, both in terms of their size and orientational order, rather than the bulk properties of the infinitely large crystals, determine which polymorph will nucleate. Hence, we have identified that one of the shortcomings of CNT is its failure in accounting for the short-range order possessed by the super cooled liquid state and that nucleation properties cannot be inferred solely from the

bulk properties of the crystal and of the melt, as is assumed to be true within the capillarity approximation.

In conclusion, by a deep study on how nucleation occurs in multi-component systems, this works demonstrates that the receipt of *mix, match, and assembly* is an effective strategy to inverse design tailored materials. Our analysis employed a combination of theoretical, numerical, and simulation-based approach, focusing on patchy particle model systems. Recent experimental advances, however, hold the promise to fully unlock the potential of such systems, whose strength lies in achieving specificity through directional and selective interactions. In particular, **DNA nanotechnology** already allows the fabrication of patchy particles in the form of wireframe DNA origami and gold nanoparticles with programmed patterning. In fact, DNA strands naturally encode sequence complementary and thus allow the assembly of structures with desired interactions. This versatile control makes DNA-based systems uniquely suited to implement the highly specific, short-range and anisotropic bonding required for inverse thermodynamics and inverse self-assembly.



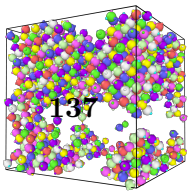
Appendix A

Boolean clauses in the SAT-assembly framework

This appendix is intended to further explain the boolean clauses introduced in Chapter 2. The SAT-assembly problem is defined by the building units and the target structure. Given the topology of the target structure, the number of different species of patchy particles N_s and the number of different patch types N_c, do a patchy particle design and interaction rules exist such that each bond in the target structure is formed?

Each patchy particle features a set of patches whose number and spatial arrangement mirror the bond pattern of the target structure, with each patch is assigned a type (colour). Moreover, as outlined in Figure A.1, each particle has a species, defined by its set of patch types, and an orientation. The target structure, on the other hand, is defined by sites and slots: each lattice site has slots that connect it to neighbouring sites. The goal is to determine whether there are colour rules and colour assignements for the patches of each species such that every site can be occupied by a patchy particle, oriented so that its patches overlap the site's slots, while ensuring that neighbouring slots are filled with interacting patches (patches of compatible colours). This information is enclosed in four boolean variables: $x^{pcol}_{s,p,c}$, $x^{L}_{l,s,o}$, $x^{A}_{l,k,c}$, $x^{int}_{c_i,c_j}$. When they are true, it means that species s has patch p coloured c, site l has species s with orientation o, site l has slot k coloured c, and colour c_i is compatible with colour c_j, respectively.

Requiring that a target structure be satisfied means controlling the bonds that each patchy particle occupying a lattice site can establish and, in turns, the orientation of the particles and the types of their patches, as well as the interacting patches. For this reason, the SAT-assembly formulation requires clauses that constraint the values the binary variables can take. These clauses are Boolean equations, and in the following we show how they naturally arise.



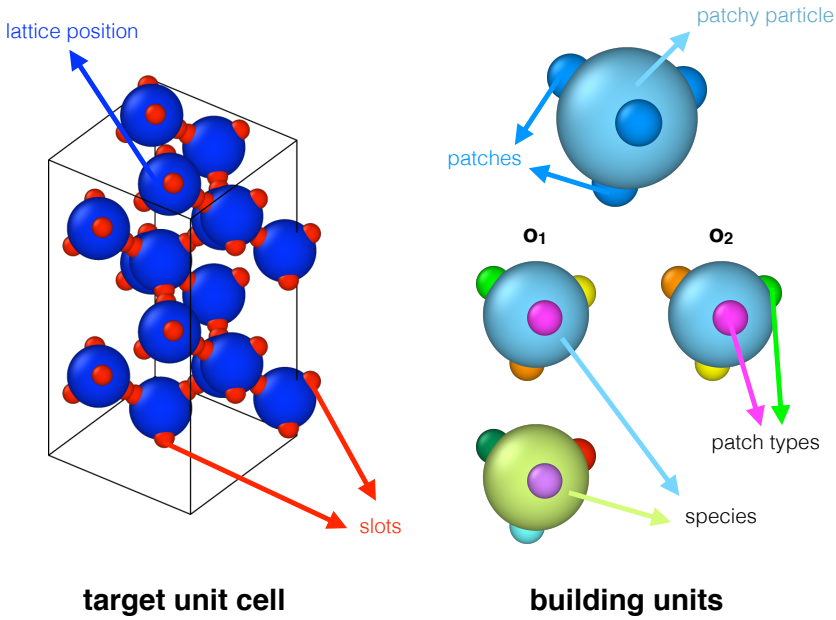


Figure A.1. SAT-assembly dictionary. Visual overview of the fundamental elements of the target unit cell and the building units that enter the SAT-assembly framework. **Left:** the unit cell contains n *lattice positions* (in blue), where each position corresponds to a specific location for a particle. The red *slots* indicate the direction in which bonding between neighbouring particles occur. **Right:** *patchy particles* serve as the building units. They are hard-core spherical particles with attractive sites, named *patches*, shown in light blue. Patchy particles have various *orientations*, such as o_1 and o_2. Additionally, there are different *patch types* which are represented by different colours. Two patchy particles with different patch types constitute two distinct *species*; to aid in the identification, the core regions are coloured differently.

The input of the colouring problem is the topology of the target structure, which defines each bond in the lattice as, for example, "Site 0 slot 3 with Site 5 slot 1". This means that lattice site 0 has lattice site 5 as a neighbour, and they interact through slots 3 and 1, respectively. Hence, these two slots must have complementary colours. It is then clear that the topology identifies neighbours sites and the interacting slots to which complementary colours must be assigned; this is enforced with the following clause:

$$C^{lint}_{l_i,k_i,l_j,k_j,c_i,c_j} = (x^A_{l_i,k_i,c_i} \wedge x^A_{l_j,k_j,c_j}) \implies x^{int}_{c_i,c_j} \tag{A.1}$$

Indeed, if lattice sites l_i is connected to lattice site l_j via slots k_i and k_j, which are coloured c_i and c_j respectively, then c_i and c_j must be complementary colours. However, clauses in the SAT-assembly frameworks must be defined only in terms of OR ($\vee$) and NOT ($\neg$) logical operator. Consequently, the previous more intuitive form with the AND ($\wedge$) and IMPLICATION ($\implies$) operators is substituted with

the equivalent expression

$$C^{lint}_{l_i,k_i,l_j,k_j,c_i,c_j} = \neg x^{A}_{l_i,k_i,c_i} \vee \neg x^{A}_{l_j,k_j,c_j} \vee x^{int}_{c_i,c_j} \tag{A.2}$$

which is the one reported in the main text.

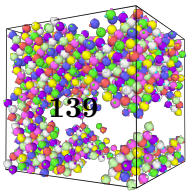
Orientation o	Mapping ϕ_o
1	(1,2,3,4)
2	(1,4,2,3)
3	(1,3,4,2)
4	(2,4,3,1)
5	(2,1,4,3)
6	(2,3,1,4)
7	(4,1,3,2)
8	(4,2,1,3)
9	(4,3,2,1)
10	(3,1,2,4)
11	(3,4,1,2)
12	(3,2,4,1)

Table A.1. Mapping between patches and slots $\phi_o(k) = p$ for a tetrahedral patchy particle. The tetrahedral symmetry comes with 12 possible orientations. For example, the first one defines the identity mapping $\phi_0 : (1, 2, 3, 4) \to (1, 2, 3, 4)$ while the fourth one the mapping $\phi_4 : (1, 2, 3, 4) \to (2, 1, 4, 3)$ meaning that slot number 1 is occupied by patch number 2 $\phi_4(1) = 2$, slot number 2 is occupied by patch number 1 $\phi_4(2) = 1$, slot number 3 is occupied by patch number 4 $\phi_4(3) = 4$, and slot number 4 is occupied by patch number 3 $\phi_4(4) = 3$.

Next, each lattice site must be occupied by a specific species. Since patchy particles admit multiple orientations, there are different ways to arrange patches over the fixed slots of the lattice. The number of such possible arrangements is given by the orientational symmetry of the building block, and each physical rotation of the patchy particle can be expressed as a mapping between patches and slots, as summarised in Table A.1 for a tetrahedral patchy particle.
Hence, for each lattice site, in addition to the species, the orientation must also be specified, subject to the constraint that patches overlap with slots and have the same colour as the slots, so that the existing bonds with neighbouring sites are satisfied. Therefore, the clause C^{LS} is required:

$$C^{LS}_{l,s,o,c,k} = \neg x^{L}_{l,s,o} \implies (x^{A}_{l,k,c} \iff x^{pcol}_{s,\phi_o(k),c}) \tag{A.3}$$

if the lattice site l is filled by species s,then s must be oriented with orientation o such that the patch $\phi_o(k)$, occupying the slot k coloured c, also has colour c. As for the previous clause, the following extended form, which includes only the NOT and



OR operators, is used.

$$C^{LS}_{l,s,o,c,k} = (\neg x^{L}_{l,s,o} \vee \neg x^{A}_{l,k,c} \vee x^{pcol}_{s,\phi_o(k),c}) \wedge (\neg x^{L}_{l,s,o} \vee \neg x^{A}_{l,p,c} \vee x^{pcol}_{s,\phi_o(k),c}) \tag{A.4}$$

As a matter of course, each site can be occupied by a single species with a specific orientation, as well as each patch is assigned a single colour. This is guaranteed by two additional clauses C^{pcol} and C^{L}, defined as

$$\begin{aligned} C^{pcol}_{s,p,c_k,c_l} &= \neg x^{pcol}_{s,p,c_k} \vee \neg x^{pcol}_{s,p,c_l} \\ C^{L}_{l,s_i,o_i,s_j,o_j} &= \neg x^{L}_{l,s_i,o_i} \vee \neg x^{L}_{l,s_j,o_j} \end{aligned} \tag{A.5}$$

Indeed, by considering the NOT and OR truth tables, C^{pcol}_{s,p,c_k,c_l} is false if both negation of variables ($\neg x^{pcol}_{s,p,c_k}$ and $\neg x^{pcol}_{s,p,c_l}$) are false, hence if both variables (x^{pcol}_{s,p,c_k} and x^{pcol}_{s,p,c_l}) are true, *i.e.* if two different colours are assigned to the same patch. In the same way, $C^{L}_{l,s_i,o_i,s_j,o_j}$ is false if two distinct particles (differing either in species, orientation, or both) are assigned to the same lattice position. If one of such cases occur, the problem is unsatisfiable since one of these clauses would not be satisfied. We indeed remind that a solution exists if there is a set of variables that makes all the clauses true simultaneously, as guaranteed by connecting them with AND operators. The exact same form characterises the C^{int} clause,

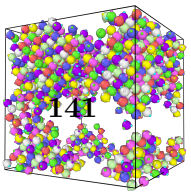
$$C^{int}_{c_i,c_j,c_k} = \neg x^{int}_{c_i,c_j} \vee \neg x^{int}_{c_i,c_k} \tag{A.6}$$

It enforces colour complementarity to be exclusive: *i.e.* if colour c_i is complementary to colour c_j, then it cannot also be complementary to colour c_k.
Finally, the colouring problem is defined for a specific number of species N_s and colours N_c; we aim to know if it exists a solution for those fixed values. Hence, clauses $C^{alls.}$ and $C^{allc.}$ are fundamental to avoid that only a subset of species and colours is used. They are defined as

$$\begin{aligned} C^{alls.} &= \bigvee_{\forall l,o} x^{L}_{l,s,o} \\ C^{allc.} &= \bigvee_{\forall s,p} x^{pcol}_{s,p,c} \end{aligned} \tag{A.7}$$

and they guarantee that each species (and each colour) is used at least once by connecting all $x^{L}_{l,s,o}$ variables over all possible lattice sites, orientation and species (and all $x^{pcol}_{s,p,c}$ variables over all possible species, patches and colours) with OR

operators.



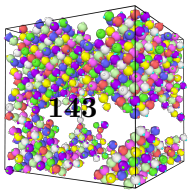
Appendix B

Overview of Wertheim perturbation theory

Perturbation theories aim to predict the thermodynamic behaviour of a system by using its interparticle potential and a reference system with a well-known Helmholtz free energy and correlation functions. Wertheim targeted fluids with short-ranged and highly directional interactions, thereby moving beyond the isotropic case. This appendix reviews the fundamental principles of the theory Wertheim developed across a series of four papers [88,254–256] to describe associating fluids. While perturbation theories are typically formulated in terms of the Helmholtz free energy difference $\Delta A = A - A_R$ between the system under analysis and the reference one (denoted with subscript R), the Wertheim formalism is grounded in graph theory. Consequently, it is more convenient to employ the grand canonical ensemble (V, T, μ), which facilitates the derivation of relations between the microscopic structure of the fluid and its thermodynamic properties, as discussed in Appendix E. The characteristic potential is the grand canonical potential $\Omega = -K_B T ln\Xi$, where Ξ is the grand canonical partition function defined as

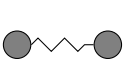
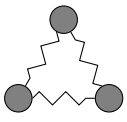
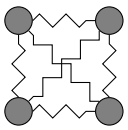
$$\Xi = \sum_{N \geq 0} \frac{e^{\beta N \mu}}{N! \Lambda^{3N}} \int e^{-\beta [U_N(\mathbf{r_1}, \ldots, \mathbf{r_N}) + \sum_{\mathbf{m=1}}^{\mathbf{N}} \phi(\mathbf{r_m})]} d\mathbf{r_1} \ldots \mathbf{dr_N} \tag{B.1}$$

where N is the number of particle, $\beta = k_B T$ with k_B being the Boltzmann constant, μ is the chemical potential, Λ is the de Broglie thermal wavelength, $\phi(\mathbf{r})$ encodes the interaction between an external field and a particle at configuration $\mathbf{r}$ (specifying both position and orientation), and $U_N(\mathbf{r_1} \ldots \mathbf{r_N}) = \sum_{\mathbf{m>n}}^{\mathbf{N}} \mathbf{u}(\mathbf{r_{mn}})$ with $u(\mathbf{r_{mn}})$ being the pair potential between particle m and n. It is useful to recast the grand canonical partition function as a functional of the activity $z(\mathbf{r}) = \frac{\mathbf{e}^{\beta(\mu - \phi(\mathbf{r}))}}{\mathbf{\Lambda^3}}$; Eq. B.1 thus becomes



$$\begin{aligned}\Xi[z] &= \sum_{N\geq 0}\int\left(\frac{1}{N!}\prod_{m=1}^{N} z(\mathbf{r_m})\right)\mathbf{e}^{-\beta\sum_{\mathbf{m=1}}^{\mathbf{N}}\phi(\mathbf{r_m})}\mathbf{dr_1}\dots\mathbf{dr_N} = \\ &= 1 + \int z(\mathbf{r})\mathbf{dr} + \frac{\mathbf{1}}{\mathbf{2!}}\int \mathbf{z(r_1)z(r_2)e^{-\beta u(r_{12})}dr_1dr_2} + \\ &+ \int z(\mathbf{r_1})\mathbf{z(r_2)z(r_3)e^{-\beta u(r_{12})}e^{-\beta u(r_{13})}e^{-\beta u(r_{23})}dr_1dr_2dr_3} + \dots\end{aligned} \tag{B.2}$$

Making the addends of $\Xi[z]$ explicit gets rapidly complex, and a graph representation is more concise:

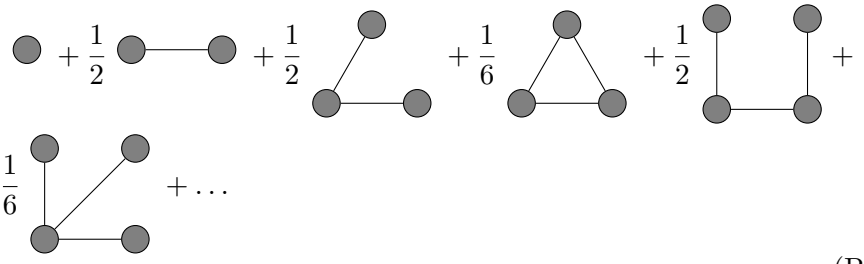
$\Xi[z] = 1 +$  $+$  $+$  $+\dots$ (B.3)

In graph theory the activity of a particle is represented as a node, named *root point*, which becomes a *field point* when activity is integrated over the particle configuration. Conventionally, root points are represented as empty circles ○, while field points as full circles ●. Edges in the graph, represented as zigzag lines, correspond to the Boltzmann factor $e^{-\beta u(\mathbf{r_{mn}})}$, which is referred to as the *e-bond*. To further develop Eq. B.3 through graph manipulation, it is convenient to introduce the Mayer function $f(\mathbf{r_{mn}}) = \mathbf{e}^{-\beta\mathbf{u}(\mathbf{r_{mn}})} - \mathbf{1}$. In graph formalism then

$$e^{-\beta u(\mathbf{r_{mn}})} = 1 + f(\mathbf{r_{mn}})$$

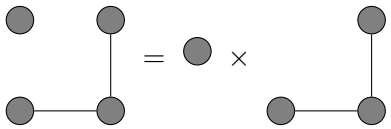
 (B.4)

having represented the Mayer function as a straight line, referred to as an *f-bond*. This is crucial since in this way the logarithm of the partition function can be expressed as an infinite sum over *connected graphs*:

$\ln\Xi[z] =$  (B.5)

Each graph represents a multiple-particle integral involving Mayer functions. The sum of all such integrals involving the same number l of nodes is defined as the

l particle cluster integral and this approach is known as cluster expansion. The introduction of Mayer functions allow for the elimination of *disconnected graphs*, *i.e.* graphs in which not all nodes are connected by f-bonds. In fact, these disconnected graphs can be factored into products of connected subgraphs (this means, as shown below, that the integration can be performed independently), and by taking the logarithm, these products are converted into a sum.



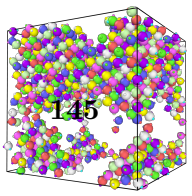
(B.6)

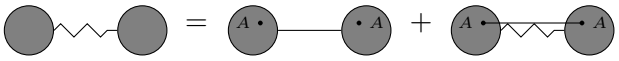
$$\begin{aligned} &\int z(\mathbf{r_1})\mathbf{z}(\mathbf{r_2})\mathbf{z}(\mathbf{r_3})\mathbf{z}(\mathbf{r_4})\mathbf{f}(\mathbf{r_{23}})\mathbf{f}(\mathbf{r_{34}})\mathbf{dr_1}\mathbf{dr_2}\mathbf{dr_3}\mathbf{dr_4} = \\ &= \int z(\mathbf{r_1})\mathbf{dr_1} \times \int \mathbf{z}(\mathbf{r_2})\mathbf{z}(\mathbf{r_3})\mathbf{z}(\mathbf{r_4})\mathbf{f}(\mathbf{r_{23}})\mathbf{f}(\mathbf{r_{34}})\mathbf{dr_2}\mathbf{dr_3}\mathbf{dr_4} \end{aligned} \tag{B.7}$$

Wertheim exploits graph reduction in Eq. B.5 to derive an analytical expression for the Helmholtz free energy. In perturbation theory, the starting point is the decomposition of the pair potential into a reference contribution and a residual perturbation. Typically, the reference potential represents a non-bonding interaction accounting for particle repulsion, whereas the residual term describes the associative contribution. Accordingly, the Mayer function can be expressed as a sum of two contributions: a reference Mayer function, $f_R(\mathbf{r_{mn}}) = \mathbf{e}^{-\beta \mathbf{u_R}(\mathbf{r_{mn}})} - \mathbf{1}$, and a perturbation Mayer function, $f_P(\mathbf{r_{mn}}) = \mathbf{e}^{-\beta \mathbf{u_P}(\mathbf{r_{mn}})} - \mathbf{1}$. The total Mayer function then reads $f(\mathbf{r_{mn}}) = \mathbf{f_R}(\mathbf{r_{mn}}) + \mathbf{e}^{-\beta \mathbf{u_R}(\mathbf{r_{mn}})}\mathbf{f_P}(\mathbf{r_{mn}})$. Wertheim chooses hard spheres as the reference system and a pairwise attractive interaction that depends on both the distance and the relative orientation of the particles. Indeed, to account for anisotropic interactions, each particle is endowed with a set Γ of attractive sites distributed over its surface. Hence,

$$\begin{aligned} u(\mathbf{r_{mn}}) &= \mathbf{u_R}(\mathbf{r_{mn}}) + \sum_{\alpha \in \mathbf{\Gamma}} \sum_{\gamma \in \mathbf{\Gamma}} \mathbf{u}_{\alpha\gamma}(\mathbf{r_{m\alpha n\gamma}}) \\ f(\mathbf{r_{mn}}) &= \mathbf{f_R}(\mathbf{r_{mn}}) + \mathbf{e}^{-\beta \mathbf{u_R}(\mathbf{r_{mn}})} \left[\prod_{\alpha \in \mathbf{\Gamma}} \prod_{\gamma \in \mathbf{\Gamma}} (\mathbf{f}_{\alpha\gamma}(\mathbf{r_{mn}}) + \mathbf{1}) - \mathbf{1} \right] \end{aligned} \tag{B.8}$$

where $|\mathbf{r_{m\alpha n\gamma}}|$ is the distance between attractive sites α of particle A (oriented Ω_m) and γ of particle B (oriented Ω_n).This type of interacting potential leads to a new diagrammatic representation. Each node is a particle and carries labels identifying attractive sites; edges exist both between particles (e-bonds and repulsive f-bonds) and sites (attractive f-bonds). For example, considering a single species with a single site, Eq. B.8 becomes $f(\mathbf{r_{mn}}) = \mathbf{f_R}(\mathbf{r_{mn}}) + \mathbf{e}^{-\beta \mathbf{u_R}(\mathbf{r_{mn}})}\mathbf{f_{AA}}(\mathbf{r_{mn}})$ and its graph representation is:





(B.9)

This allows for grouping the several graphs according to the bonding state of its nodes, *i.e.* according to the number and the type of attractive sites that are bonded. A **multiple-density formalism** is introduced; the number density $\rho(\mathbf{r})$ is defined as

$$\rho(\mathbf{r}) = \sum_{\gamma \subset \mathbf{\Gamma}} \rho_\gamma(\mathbf{r}) \tag{B.10}$$

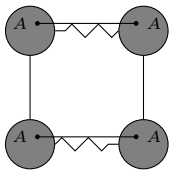
where γ runs over all possible bonding states Γ. For example, for a system of N identical particles with two attractive sites A and B, the possible bonding states for a particle are: none of the sites is bonded, only site A or B is bonded, both sites A and B are bonded. Thus, $\rho(\mathbf{r}) = \rho_{\mathbf{0}}(\mathbf{r}) + \rho_{\mathbf{A}}(\mathbf{r}) + \rho_{\mathbf{B}}(\mathbf{r}) + \rho_{\mathbf{AB}}(\mathbf{r})$. Each $\rho_\gamma(\mathbf{r})$ is proportional to the probability that a particle at configuration $\mathbf{r}$ has γ sites bonded. Consequently, $\rho(\mathbf{r})$ represents the probability of finding a particle at $\mathbf{r}$ regardless of its bonding state. This framework effectively describes associating fluids where particles are not equivalent, as they exist in distinct bonding states depending on which attractive sites are bonded. From a statistical perspective, each bonding state is treated as a distinct species with its own number density. Wertheim also introduced the density function $\sigma_\alpha(\mathbf{r})$:

$$\sigma_\alpha(\mathbf{r}) = \sum_{\gamma \subseteq \alpha} \rho_\gamma(\mathbf{r}) \tag{B.11}$$

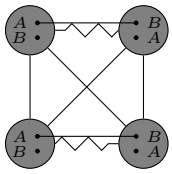
where the summation is over all bonding states γ that contain the set of attractive sites α. It represents the probability that a particle at $\mathbf{r}$ has set of sites α bonded. Following the previous example, we have $\sigma_A(\mathbf{r}) = \rho_{\mathbf{A}}(\mathbf{r}) + \rho_{\mathbf{AB}}(\mathbf{r})$, $\sigma_B(\mathbf{r}) = \rho_{\mathbf{B}}(\mathbf{r}) + \rho_{\mathbf{AB}}(\mathbf{r})$, and $\sigma_{AB}(\mathbf{r}) = \rho_{\mathbf{AB}}(\mathbf{r})$. We also notice that $\sigma_0(\mathbf{r}) = \rho_{\mathbf{0}}(\mathbf{r})$ and $\sigma_\Gamma(\mathbf{r}) = \rho(\mathbf{r})$. This formalism is significant because the Helmholtz free energy difference between the target and the reference system can be expressed via the logarithm of the gran canonical partition function in terms of these densities. Specifically, the Helmholtz free energy functional is related to the grand canonical potential through: $A[\rho] = \int \rho(\mathbf{r})(\mu - \phi(\mathbf{r}))\mathbf{dr} + \mathbf{\Omega}[\rho]$. While the diagrammatic cluster expansion obtained for the Helmholtz free energy difference is exact, it is intractable as it involves the sum of an infinite number of diagrams.

To resolve this, Wertheim introduces an **early consideration of the geometry of interactions** (Ⓢ$_1$). The considered interaction potential leads to significant cancellations of graphs due to steric incompatibilities, *i.e.* hard core repulsion and excluded volume effects. As the attractive potential becomes short-ranged, the

number of particles that can bind to a given interaction site decreases. Wertheim considers the limiting case of the *one bond per site* condition, in which each site can participate in only one association bond. This constraint can be viewed as a restriction to a specific class of systems rather than as an approximation. He also introduces the **single-chain approximation** ($\textcircled{S}_2$): graphs, such as the following ones, containing more than one single dimer are excluded.



(B.12)



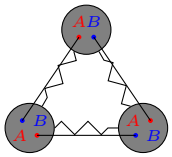
(B.13)

The first case Wertheim takes into account consists of particles with a single association site, and he demonstrates that ($\textcircled{S}_1 + \textcircled{S}_2$) allow the attractive f_{AA}-bond to be factorised and the resulting graphs to be summed into the pair correlation function of the reference fluid $g_R^{(2)}(\mathbf{r}, \mathbf{r}')$. This yields a self-consistent density relation for the monomer fraction $\frac{\rho_0(\mathbf{r})}{\rho(\mathbf{r})}$:

$$\frac{\rho_0(\mathbf{r})}{\rho(\mathbf{r})} = \left[1 + \int f_{AA}(\mathbf{r}, \mathbf{r}') \mathbf{g_R^{(2)}}(\mathbf{r}, \mathbf{r}') \boldsymbol{\rho_0}(\mathbf{r}') \mathbf{dr}'\right]^{-\mathbf{1}} \tag{B.14}$$

Crucially, the fraction of unbonded particles in the system is obtained from a pair correlation integral combining a well-characterised structural quantity of the reference system with the associative Mayer function. Therefore, the Helmholtz free energy difference admits an analytical evaluation for a given number density profile $\rho(\mathbf{r})$.

For two attractive sites, in contrast to the previous case where associated structures (attractive f-bonds) are restricted to dimers, clusters with an arbitrary value of attractive f-bonds may now form, including loops (see Eq. B.15 for a visual example). Consequently, Wertheim introduces an additional approximation by **neglecting loops** ($\textcircled{S}_3$) and limiting the analysis to chain associates.



(B.15)

Under this further assumption, the diagrammatic cluster expansion can be expressed

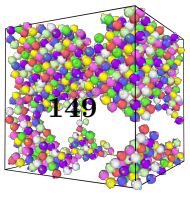
as a sum over subsets of graphs corresponding to s-mers. Consider all such graphs would entail infinitely many contributions, which formally require up to ∞-order particle distribution functions $g_R^{(2)}(\mathbf{r}, \mathbf{r}') \dots \mathbf{g}_{\mathbf{R}}^{(\infty)}(\mathbf{r}, \mathbf{r}', \dots, \mathbf{r}^{\infty})$. To obtain an analytic expression and self-consistent density expressions for the probability that each site is bonded, the diagrammatic expansion must be closed perturbatively by truncating it at a finite order, *i.e.* by neglecting contributions from higher order graphs. Specifically, Wertheim truncates the expansion at orders $n = 1$ and $n = 2$, giving rise to thermodynamic perturbation theory TPT-1 and TPT-2, respectively. In TPT-1, only subgraphs up to dimers are retained, and these can be exactly resummed into the two-particle distribution function of the reference fluid $g_R^{(2)}(\mathbf{r}, \mathbf{r}')$. As a result, the Helmholtz free energy is expressed analytically in terms of the bonding probabilities of the associative site A and B. These probabilities are determined via association integrals, defined as the volume integrals of the product of the site-specific Mayer function and the reference two-particle distribution function. In TPT-2, diagrams containing up to one trimer are included. In this case the bonding probabilities depend on both the two-particle and three-particle distribution function of the reference system: $g_R^{(2)}(\mathbf{r}, \mathbf{r}')$ and $g_R^{(3)}(\mathbf{r}, \mathbf{r}', \mathbf{r}'')$, respectively. More generally, a TPT-n theory can be formulated by retaining diagrams containing up to $n + 1$-mers. For further details we refer to Ref. [257].

In conclusion, Wertheim transforms the formally exact multi-density description into a probabilistic theory characterised solely by the fractions of unbonded sites which obey mass-action equations and determine the associating free energy. The expression for the Helmholtz free energy of patchy particle systems, provided in Chapter 3, is derived by a further development of the ideas of Wertheim performed by Champan and coworkers which extended the theory to describe particles with an arbitrary number of attractive sites, allowing for different valences among particles [89,258–261]. In this case, results have been derived only in the first order approximation, *i.e.* by effectively considering only two-body correlations.

Appendix C

Stability analysis of mixtures

For completeness about theoretical tools for mixture phase behaviour analysis, we note that Chen and Forstmann introduced a scheme to characterize the thermodynamic instabilities of binary mixtures [262]. By linking macroscopic instability to local compositional and density order, their approach identifies whether a phase transition is driven by concentration and/or density fluctuations [263]. Stability is determined via the second functional derivative of the grand canonical potential, which is related to the direct correlation functions (see Appendix E). These functions account for the non-local nature of particle interactions, *i.e.* how a fluctuation at position $\mathbf{r}$ affects the system in $\mathbf{r}'$. While a real space analysis requires solving an eigenvalue problem for an integral equation, applying a Fourier transformation converts spatial convolutions into algebraic products of wave vector dependent fluctuations. The resulting stability matrix $M(k)$ reduces to the macroscopic thermodynamic Hessian in the limit of $k \to 0$. By analyzing $M(k)$ across different wave vectors, it is possible to identify the wavelength driving the instability. The spinodal line is located where the smallest eigenvalue of the stability matrix vanishes. The associated eigenvector can be expressed as a linear combination of two basis vectors representing total density and concentration fluctuations and its direction determines which physical process dominates the phase transition. The angle of this eigenvector relative to the concentration axis quantifies the coupling between demixing and condensation when the system becomes unstable. Angles of 0° and 90° represent the limiting cases of pure demixing (only concentration fluctuations occur) and pure condensation (only density fluctuations occur), respectively.



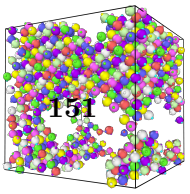
Appendix D

Overview of Classical Nucleation Theory

This appendix is devoted to briefly summarise Classical Nucleation Theory (CNT) of which a falsifiability test is presented in Chapter 6. CNT describes first order phase transitions by considering nucleation as a homogeneous process in which a spherical droplet of a more stable phase originates and grows within a metastable parent phase. Particles can be either unbound or part of a cluster. A key assumption of CNT is the *capillarity approximation*, which treats clusters formed in the parent phase as having the same macroscopic properties as the bulk phase. Each cluster is assumed to possess a uniform density identical to that of the stable bulk phase and a surface tension equivalent to that of a flat interface. Therefore, the theory assumes clusters evolve solely by changing size, maintaining a fixed structure and composition. This allows the nucleation process to be described using a single reaction coordinate which accounts for the cluster size. This is why CNT provides a simple expression for the Gibbs free energy change associated with forming a nucleus of n particles of the new phase, under constant temperature and pressure:

$$\Delta G(n) = -n\,|\Delta\mu| + \alpha\gamma n^{\frac{2}{3}} \tag{D.1}$$

This expression is the sum of two competing contributions. A negative and favourable volume term which reflects the free energy gain from transitioning to a more stable phase and it is proportional to n and to the chemical potential difference $\Delta\mu$ between the two phases. A positive and unfavourable surface term which represents the free energy cost of sustaining a phase boundary and it is proportional to $n^{\frac{2}{3}}$ and to the surface tension γ. For small nuclei, the interfacial positive term (depicted with a cyan line in Figure D.1) dominates, while for large nuclei, the negative term (shown in blue in Figure D.1) becomes dominant. This results in a maximum in the free



energy curve, which defines the critical nucleus size n_C and the corresponding free energy barrier $\Delta G(n_C)$.

$$\frac{d\Delta G(n)}{dn} = 0 \implies n = n_C \tag{D.2}$$

$$n_C = \left(\frac{2\alpha\gamma}{3\left|\Delta\mu\right|}\right)^3 \qquad \text{and} \qquad \Delta G(n_C) = \frac{4(\alpha\gamma)^3}{27\left|\Delta\mu\right|^2} = \frac{1}{2} n_C \left|\Delta\mu\right| \tag{D.3}$$

In summary, according to CNT the nucleation barrier is given by the critical nucleus size. This simple expression hides one of the shortcoming of CNT: the barrier height depends on the choice of order parameter used to distinguish between liquid and crystalline particles. This is why more rigorous techniques, such as Umbrella Sampling, are used to evaluate the free energy profile and the nucleation barrier. We also notice that another order parameter suitable for describing the size of spherical clusters is their radius. By using the radius of the crystalline nucleus instead of the number of crystalline particles, the same expression for the nucleation barrier can be obtained.

The probability that a spontaneous fluctuation can overcome this barrier is proportional to the Boltzmann factor. The CNT expression for the nucleation rate is obtained by multiplying this probability by a kinetic prefactor related to the rate at which particles attach to the nucleus:

$$J = Ke^{-\beta\Delta G(n_C)} \tag{D.4}$$

Since, as expected, when the two phases are at coexistence ($|\Delta\mu| = 0$) the nucleation barrier diverges, homogeneous nucleation can happen only at supercooling, *i.e.* when the liquid phase is below its melting temperature T_m (the temperature at which it coexists with the solid phase). The further below T_m we go, the higher the thermodynamic driving force toward the stable solid phase. However, this is compensated by the decrease in kinetics, accounted for by the pre-exponential factor K in the expression of the nucleation rate J, which represents the number of critical nuclei originated per unit time and per unit volume. In fact, K can be written as

$$K = \rho_l w^+ Z \tag{D.5}$$

where ρ_l is the number density of nucleation sites, w^+ is the attachment rate (*i.e.* the number of particles in unit time that attach to the new nucleus and depends on the diffusion coefficient and the diffusion distance), and Z is the Zeldovich factor, which is related to the curvature of the free energy barrier and provides the fraction of

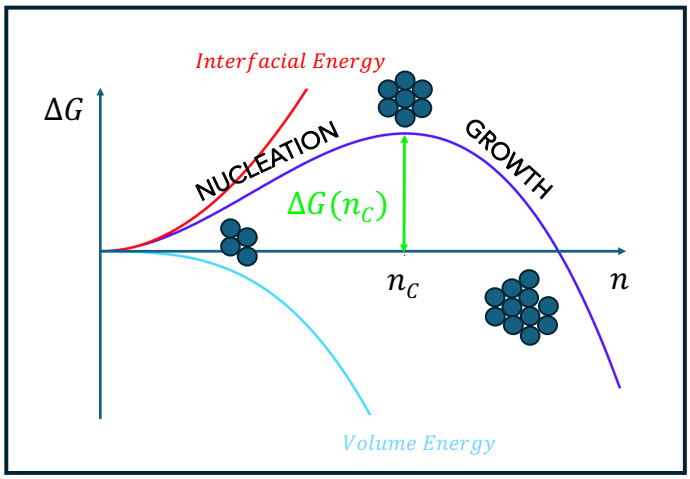
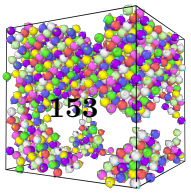


Figure D.1. CNT overview. The formation of a new stable phase from a metastable one is an activated process that proceeds via nucleation and growth of nuclei of the new phase. The associated free energy barrier ΔG is a function of the cluster size n and its maximum is determined by the critical nucleus size n_c. The critical nucleus corresponds to the size at which a cluster has equal probability to grow or dissolve. Only for $n > n_c$ the system enters the growth regime because the gain in bulk free energy (cyan line) outweighs the free energy penalty of creating an interface (red line).

critical nucleus that successfully overcome the barrier rather than dissolve [264, 265]. In conclusion, as illustrated in Figure D.1, the formation of a new stable phase from a metastable one occurs in two stages: nucleation and growth. For small nuclei the surface to volume ratio is high, leading to a large free energy cost associated with maintaining the interface. As a result, these small clusters are transient and tend to dissolve. However, if the nucleus exceeds the critical size n_C, it becomes energetically favourable for it to grow into the new stable phase. The two key quantities that govern this process are the chemical potential difference $\Delta\mu$ between the two phases and the surface tension γ. This highlights the core capillarity approximation of CNT: the idea that emerging nuclei have the same thermodynamic properties as the bulk phase.

CNT is the most widely used theoretical model to describe nucleation process, mainly because of its relative simplicity and its ability to make useful prediction. In many cases it works well, but in others, it does not. To fix the mismatches, various modification to the original theory have been proposed. Most of these maintain the core capillarity approximation, attempting to adjust the details without changing the foundation: correcting terms are typically included into the nucleation rate expression, either in the exponent (*i.e.* in the free energy cost) or as a multiplicative factor. However, the success or lack thereof of such procedures are system dependent and state-point dependent, and only provide a phenomenological assessment of CNT. This has led to a situation where the reliability of the theory is unclear, and whether



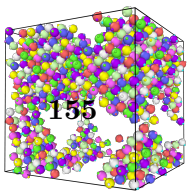
CNT validity is still a matter of debate. Our goal in Chapter 6 is to put to the test the capillarity approximation directly by checking where the theory should fail if it is not true.

Appendix E

The Kirkwood-Buff theory and the evaluation of chemical potentials in simulations

The Kirkwood-Buff theory serves as a fundamental bridge in statistical mechanics between the microscopic structure of a multi-component system and its macroscopic thermodynamic properties. At the heart of this connection is the response of the system to changes in particle number which is enclosed in the chemical potentials. In a mixture, the chemical potential μ_i of the i-component is defined as the change in free energy resulting from a variation in the number of particle of that component N_i. Kirkwood and Buff demonstrated that chemical potentials derivatives can be mapped into equilibrium fluctuations of particle number, which are ultimately expressed as volume integrals of radial distribution functions [240]. This appendix outlines the derivation of these **Kirkwood-Buff integrals**, originally defined for infinite, open, isotropic, and homogeneous systems, beginning with the statistical mechanics of the grand canonical ensemble. This ensemble describes an open system of volume V in thermodynamic equilibrium with a reservoir. Unlike closed systems, which only exchange energy, permeable boundaries allow for the exchange of both energy and particles. The reservoir acts as an infinite source of particles and energy, maintaining the system at constant chemical potential μ and temperature T, *i.e.* in a state of thermal and chemical equilibrium, while the system's particle numbers and energy fluctuate. Hence, the thermodynamic state of the system is defined by the variables (μ, V, T) and the characteristic thermodynamic potential is the grand potential defined as

$$\Omega = -k_B T ln \Xi = F - \sum_{i=1}^{m} N_i \mu_i \tag{E.1}$$



where m is the number of different species, F is the Helmholtz free energy, and Ξ is the grand canonical partition function:

$$\Xi = \sum_{N_A=0}^{\infty} \sum_{N_B}^{\infty} \dots \sum_{N_m}^{\infty} e^{\sum_{i=1}^{m} \beta \mu_i N_i} \frac{1}{\prod_{i=1}^{m} h^{3N_i} N_i!} \int e^{-\beta H} d\mathbf{r^N} \mathbf{dp^N} \tag{E.2}$$

where h is the Planck constant. By noting that the Hamiltonian $H(\mathbf{r^N}, \mathbf{p^N}) = \mathbf{K}(\mathbf{p^N}) + \mathbf{U}(\mathbf{r^N})$ depends on the momenta only through the kinetic energy K, the momenta can be integrated out. By defining the activities $z_i = \frac{e^{\beta \mu_i}}{\Lambda_i^3}$ and the canonical configurational integral $Z_{\{N_i\}} = \int e^{-\beta U_N(\mathbf{r_1}, \dots, \mathbf{r_N})} d\mathbf{r_1} \dots \mathbf{dr_N}$ for a specific set of $\{N_A, N_B, \dots, N_m\}$ Eq. E.2 becomes

$$\Xi = \sum_{\{N_i\}} \left(\prod_{i=1}^{m} \frac{z_i^{N_i}}{N_i!} \right) Z_{\{N_i\}} \tag{E.3}$$

where Λ_i are the de Broglie thermal wavelength and $U_N(\mathbf{r_1}, \dots \mathbf{r_N})$ is the interaction potential, a multi-body term depending on the configurational degrees of freedom (positions and orientations). Thus, $\mathbf{r} = (\mathbf{x}, \mathbf{\Omega})$ generally denotes a one particle configuation specifying both its position $\mathbf{x}$ and orientation Ω; however, in the original derivation by Kirkwood and Buff, which considered isotropic fluids, $\mathbf{r}$ corresponds simply to the particle position. The partition function is the normalisation of the equilibrium probability density f^0 which is a fundamental quantity entering in ensemble averages. Given an observable O, its ensemble average is indeed defined as:

$$\begin{aligned} \langle O \rangle &= \int O(\mathbf{r_1}, \dots, \mathbf{r_N}) \mathbf{f^0}(\mathbf{r_1}, \dots, \mathbf{r_N}) \mathbf{dr_1} \dots \mathbf{dr_N} \\ &= \frac{1}{\Xi} \sum_{\{N_i\}} \left(\prod_{i=1}^{m} \frac{z_i^{N_i}}{N_i!} \right) \int O(\mathbf{r_1}, \dots, \mathbf{r_N}) \mathbf{e}^{-\beta \mathbf{U_N}(\mathbf{r_1}, \dots, \mathbf{r_N})} \mathbf{dr_1} \dots \mathbf{dr_N} \end{aligned} \tag{E.4}$$

having considered the case where the observable does not depends on momenta. We can define the instantaneous microscopic density of particles of species α as

$$\rho_\alpha(\mathbf{r}) = \sum_{\mathbf{i=1}}^{\mathbf{N}_\alpha} \delta(\mathbf{r} - \mathbf{r_i}) \tag{E.5}$$

where $\mathbf{r}$ is the spatial coordinate at which the density is evaluated, $\mathbf{r_i}$ is the position of particle i and the sum runs over all N_α particles of species α in a given microstate. By integrating $\rho_\alpha(\mathbf{r})$ over $\mathbf{r}$ in the volume V, the total number of particle of species α in the system in that specific microstate is obtained:

$$\int_V \rho_\alpha(\mathbf{r}) \mathbf{dr} = \mathbf{N}_\alpha \tag{E.6}$$

Averaging the microscopic density over all possible microstates yields the one particle density distribution $\rho_\alpha^{(1)}(\mathbf{r})$. This quantity represents the probability density of finding a particle of species α at position $\mathbf{r}$:

$$\begin{aligned}
\langle \rho_\alpha(\mathbf{r}) \rangle &= \langle \sum_{i=1}^{N_\alpha} \delta(\mathbf{r} - \mathbf{r_i}) \rangle = \\
&= \frac{1}{\Xi} \sum_{\{N_k\}} \left(\prod_{k=1}^{m} \frac{z_k^{N_k}}{N_k!} \right) \int \left[\sum_{i=1}^{N_\alpha} \delta(\mathbf{r} - \mathbf{r_i}) \right] \mathbf{e}^{-\beta \mathbf{U_N}(\mathbf{r_1}, \ldots, \mathbf{r_N})} \mathbf{dr_1} \ldots \mathbf{dr_N} = \\
&= \frac{1}{\Xi} \sum_{\{N_k\}} \left(\prod_{k=1}^{m} \frac{z_k^{N_k}}{N_k!} \right) N_\alpha \int \delta(\mathbf{r} - \mathbf{r_1}) \mathbf{e}^{-\beta \mathbf{U_N}(\mathbf{r_1}, \ldots, \mathbf{r_N})} \mathbf{dr_1} \ldots \mathbf{dr_N} = \\
&= \frac{1}{\Xi} \sum_{\{N_k\}} \left(\prod_{k=1}^{m} \frac{z_k^{N_k}}{N_k!} \right) N_\alpha \int e^{-\beta U_N(\mathbf{r}, \ldots, \mathbf{r_N})} d\mathbf{r_2} \ldots \mathbf{dr_N} = \\
&= \rho_\alpha^{(1)}(\mathbf{r})
\end{aligned} \tag{E.7}$$

where $\mathbf{r_1}$ denotes the position of an arbitrary particle of species α and we have considered that, for fixed N_α, $\langle \sum_{i=1}^{N_\alpha} \delta(\mathbf{r} - \mathbf{r_i}) \rangle$ is equal to $N_\alpha \langle \delta(\mathbf{r} - \mathbf{r_1}) \rangle$ due to permutation symmetry; each particle of species α contributes identically to the ensemble average. Finally, by integrating $\rho_\alpha^{(1)}(\mathbf{r})$ over $\mathbf{r}$ in the volume, we obtain the average number of particles α in the system:

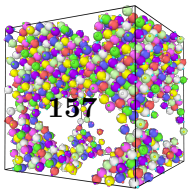
$$\int_V \rho_\alpha^{(1)}(\mathbf{r}) \mathbf{dr} = \langle \mathbf{N}_\alpha \rangle \tag{E.8}$$

We note that if the system is homogeneous, then $\rho_\alpha^{(1)}(\mathbf{r})$ is a constant, denoted c_α and the integral becomes $\int_V \rho_\alpha^{(1)}(\mathbf{r}) \mathbf{dr} = \mathbf{c}_\alpha \mathbf{V}$. Thus, $c_\alpha = \frac{\langle N_\alpha \rangle}{V}$ represents the concentration of species α, following the convention of Kirkwood and Buff, this quantity denotes the bulk number density (typically expressed in mol/L rather than $particle/m^3$).

The instantaneous two-particle density $\rho_{\alpha\beta}(\mathbf{r}', \mathbf{r}'')$ is defined as

$$\rho_{\alpha\beta}(\mathbf{r}', \mathbf{r}'') = \sum_{\mathbf{i=1}}^{\mathbf{N}_\alpha} \sum_{\substack{j=1 \\ j \neq i}}^{\mathbf{N}_\beta} \delta(\mathbf{r}' - \mathbf{r_i}) \delta(\mathbf{r}'' - \mathbf{r_j}) \tag{E.9}$$

Similar to the one particle case, we can consider its ensemble average



$$\begin{aligned}
\langle \rho_{\alpha\beta}(\mathbf{r}', \mathbf{r}'') \rangle &= \langle \sum_{i=1}^{N_\alpha} \sum_{\substack{j=1 \\ j\neq i}}^{N_\beta} \delta(\mathbf{r}' - \mathbf{r_i}) \delta(\mathbf{r}'' - \mathbf{r_j}) \rangle = \\
&= \frac{1}{\Xi} \sum_{\{N_k\}} \left(\prod_{k=1}^{m} \frac{z_k^{N_k}}{N_k!} \right) \int \left[\sum_{i=1}^{N_\alpha} \sum_{\substack{j=1 \\ j\neq i}}^{N_\beta} \delta(\mathbf{r}' - \mathbf{r_i}) \delta(\mathbf{r}'' - \mathbf{r_j}) \right] \mathbf{e}^{-\beta \mathbf{U_N}(\mathbf{r_1}, \ldots, \mathbf{r_N})} \mathbf{dr_1} \ldots \mathbf{dr_N} = \\
&= \frac{1}{\Xi} \sum_{\{N_k\}} \left(\prod_{k=1}^{m} \frac{z_k^{N_k}}{N_k!} \right) N_\alpha N_\beta \int \delta(\mathbf{r}' - \mathbf{r_1}) \delta(\mathbf{r}'' - \mathbf{r_2}) \mathbf{e}^{-\beta \mathbf{U_N}(\mathbf{r_1}, \ldots, \mathbf{r_N})} \mathbf{dr_1} \ldots \mathbf{dr_N} = \\
&= \frac{1}{\Xi} \sum_{\{N_k\}} \left(\prod_{k=1}^{m} \frac{z_k^{N_k}}{N_k!} \right) N_\alpha N_\beta \int e^{-\beta U_N(\mathbf{r}', \mathbf{r}'', \ldots, \mathbf{r_N})} d\mathbf{r_3} \ldots \mathbf{dr_N} = \\
&= \rho^{(2)}_{\alpha\beta}(\mathbf{r}', \mathbf{r}'')
\end{aligned} \tag{E.10}$$

which is the probability density of finding a particle of species α at position $\mathbf{r}'$ and a particle of species β at position $\mathbf{r}''$. By integrating $\rho^{(2)}_{\alpha\beta}(\mathbf{r}', \mathbf{r}'')$ over the volume for both $\mathbf{r}'$ and $\mathbf{r}''$, we obtain

$$\int_V \rho^{(2)}_{\alpha\beta}(\mathbf{r}', \mathbf{r}'') \mathbf{dr}' \mathbf{dr}'' =$$

if $\alpha \neq \beta$

$$\begin{aligned}
&= \frac{1}{\Xi} \sum_{\{N_i\}} \left(\prod_{i=1}^{m} \frac{z_i^{N_i}}{N_i!} \right) N_\alpha N_\beta \int e^{-\beta U_N(\mathbf{r}', \mathbf{r}'', \ldots, \mathbf{r_N})} d\mathbf{r}' \mathbf{dr}'' \mathbf{dr_3} \ldots \mathbf{dr_N} = \\
&= \langle N_\alpha N_\beta \rangle
\end{aligned}$$

if $\alpha = \beta$

$$\begin{aligned}
&= \frac{1}{\Xi} \sum_{\{N_i\}} \left(\prod_{i=1}^{m} \frac{z_i^{N_i}}{N_i!} \right) N_\alpha (N_\alpha - 1) \int e^{-\beta U_N(\mathbf{r}', \mathbf{r}'', \ldots, \mathbf{r_N})} d\mathbf{r}' \mathbf{dr}'' \mathbf{dr_3} \ldots \mathbf{dr_N} = \\
&= \langle N_\alpha^2 \rangle - \langle N_\alpha \rangle
\end{aligned}$$

$$\implies \int_V \rho^{(2)}_{\alpha\beta}(\mathbf{r}', \mathbf{r}'') \mathbf{dr}' \mathbf{dr}'' = \langle \mathbf{N}_\alpha \mathbf{N}_\beta \rangle - \delta_{\alpha\beta} \langle \mathbf{N}_\alpha \rangle \tag{E.11}$$

where $\delta_{\alpha\beta}$ is the Kronecker delta, which equals 0 if $\alpha \neq \beta$ and 1 otherwise. Indeed, since autocorrelation (the correlation between a particle and itself) is not considered

being the sum over j limited to $j \neq i$, when α and β belong to the same species, we have $N_\alpha(N_\alpha - 1) = N_\alpha^2 - N_\alpha$ (and not N_α^2) distinct pairs of particles. The Kirkwood-Buff integrals are defined starting from the following expression

$$\int_V (\rho^{(2)}_{\alpha\beta}(\mathbf{r}', \mathbf{r}'') - \rho^{(1)}_\alpha(\mathbf{r}')\rho^{(1)}_\beta(\mathbf{r}''))\mathbf{dr}'\mathbf{dr}'' =$$

$$= \int_V \rho^{(2)}_{\alpha\beta}(\mathbf{r}', \mathbf{r}'')\mathbf{dr}'\mathbf{dr}'' - \int_{\mathbf{V}} \rho^{(1)}_\alpha(\mathbf{r}')\mathbf{dr}' \int_{\mathbf{V}} \rho^{(1)}_\beta(\mathbf{r}'')\mathbf{dr}'' = \tag{E.12}$$

$$= \langle N_\alpha N_\beta \rangle - \delta_{\alpha\beta}\langle N_\alpha \rangle - \langle N_\alpha \rangle \langle N_\beta \rangle$$

where the last equality holds because of Eq. E.11 and Eq. E.8. The two particle radial distribution function $g^{(2)}_{\alpha\beta}(\mathbf{r}', \mathbf{r}'')$ is then introduced. Its expression is:

$$g^{(2)}_{\alpha\beta}(\mathbf{r}', \mathbf{r}'') = \frac{\rho^{(\mathbf{2})}_{\alpha\beta}(\mathbf{r}', \mathbf{r}'')}{\rho^{(\mathbf{1})}_\alpha(\mathbf{r}')\rho^{(\mathbf{1})}_\beta(\mathbf{r}'')} \tag{E.13}$$

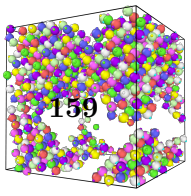
When the system is homogeneous, $\rho^{(1)}_\alpha(\mathbf{r}') = \mathbf{c}_\alpha$, $\rho^{(1)}_\beta(\mathbf{r}'') = \mathbf{c}_\beta$, and $g^{(2)}_{\alpha\beta}(\mathbf{r}', \mathbf{r}'')$ depends only on the distance $r = |\mathbf{r}'' - \mathbf{r}'|$ between the two particles. Hence, $\rho^{(2)}_{\alpha\beta}(\mathbf{r}', \mathbf{r}'') = \mathbf{c}_\alpha\ \mathbf{c}_\beta\ \mathbf{g}^{(\mathbf{2})}_{\alpha\beta}(\mathbf{r})$ and Eq. E.12 can be written as

$$\begin{aligned} \int_V (c_\alpha\ c_\beta\ g^{(2)}_{\alpha\beta}(r) - c_\alpha\ c_\beta)\ d\mathbf{r}'\mathbf{dr} &= V c_\alpha c_\beta \int_V \left[g^{(2)}_{\alpha\beta}(r) - 1\right] d\mathbf{r} = \\ &= \frac{\langle N_\alpha \rangle \langle N_\beta \rangle}{V} \int_V \left[g^{(2)}_{\alpha\beta}(r) - 1\right] d\mathbf{r} \\ &= \langle N_\alpha N_\beta \rangle - \delta_{\alpha\beta}\langle N_\alpha \rangle - \langle N_\alpha \rangle \langle N_\beta \rangle \end{aligned} \tag{E.14}$$

Finally, Kirkwood-Buff integrals $G_{\alpha\beta}$ are defined as

$$\begin{aligned} G_{\alpha\beta} &= \int_0^\infty \left[g^{(2)}_{\alpha\beta}(r) - 1\right] d\mathbf{r} \\ &= V \frac{\langle N_\alpha N_\beta \rangle - \langle N_\alpha \rangle \langle N_\beta \rangle}{\langle N_\alpha \rangle \langle N_\beta \rangle} - \frac{\delta_{\alpha\beta}}{c_\alpha} \end{aligned} \tag{E.15}$$

being $\frac{V\delta_{\alpha\beta}}{\langle N_\beta \rangle} = \frac{\delta_{\alpha\beta}}{c_\alpha}$ as $\delta_{\alpha\beta} \neq 0$ only if $\alpha = \beta$, *i.e.* $N_\alpha = N_\beta$. Hence, the Kirkwood-Buff integrals express the structure of the system in terms of equilibrium particle number fluctuations, which can be used to evaluate the derivatives of chemical potentials with respect to concentration. For a binary AB mixture, at constant pressure P and temperature T, the differential of the concentration of component A is given by



$$\begin{aligned} dc_A &= \frac{\partial c_A}{\partial \ln z_A} d\ln z_A + \frac{\partial c_A}{\partial \ln z_B} d\ln z_B = \\ &= \left[\left(\frac{\partial c_A}{\partial \mu_A} \right)_{\mu_B} d\mu_A + \left(\frac{\partial c_A}{\partial \mu_B} \right)_{\mu_A} d\mu_B \right] = \\ &= \left[\left(\frac{\partial c_A}{\partial \mu_A} \right)_{\mu_B} - \frac{c_A}{c_B} \left(\frac{\partial c_A}{\partial \mu_B} \right)_{\mu_A} \right] d\mu_A \end{aligned} \tag{E.16}$$

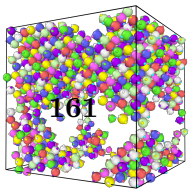
where the subscripts indicate the thermodynamic variables held constant during differentiation. The final equality is obtained by applying the Gibbs-Duhem relation, which takes the form $c_A d\mu_A + c_B d\mu_B = 0$ at constant P and T [266]. Consequently,

$$\left(\frac{\partial c_A}{\partial \mu_A} \right)_{T,P} = \left(\frac{\partial c_A}{\partial \mu_A} \right)_{\mu_B} - \frac{c_A}{c_B} \left(\frac{\partial c_A}{\partial \mu_B} \right)_{\mu_A} \tag{E.17}$$

Let's evaluate each addend of Eq. E.17 by considering that $\frac{\partial}{\partial \ln z_A} = z_A \frac{\partial}{\partial z_A}$ and that $\frac{\partial}{\partial z_A}\left(\frac{1}{\Xi}\right) = -\frac{\partial \Xi}{\partial z_A}\frac{1}{\Xi^2}$

$$
\begin{aligned}
k_B T V \frac{\partial c_A}{\partial \mu_A} &= k_B T V \frac{\partial \langle N_A \rangle}{\partial \mu_A} \frac{1}{V} = \frac{\partial \langle N_A \rangle}{\partial \ln z_A} = \\
&= \frac{\partial}{\partial \ln z_A} \left(\frac{1}{\Xi} \sum_{\{N_A, N_B\}} \frac{z_A^{N_A} z_B^{N_B}}{N_A! N_B!} Z_{\{N_A, N_B\}} N_A \right) = \\
&= \frac{1}{\Xi} \frac{\partial}{\partial \ln z_A} \left(\sum_{\{N_A, N_B\}} \frac{z_A^{N_A} z_B^{N_B}}{N_A! N_B!} Z_{\{N_A, N_B\}} N_A \right) + \\
&+ \frac{\partial}{\partial \ln z_A} \left(\frac{1}{\Xi} \right) \sum_{\{N_A, N_B\}} \frac{z_A^{N_A} z_B^{N_B}}{N_A! N_B!} Z_{\{N_A, N_B\}} N_A = \\
&= z_A \frac{1}{\Xi} \sum_{\{N_A, N_B\}} N_A \frac{z_A^{N_A - 1} z_B^{N_B}}{N_A! N_B!} Z_{\{N_A, N_B\}} N_A - \\
&- z_A \left(\sum_{\{N_A, N_B\}} \frac{z_A^{N_A} z_B^{N_B}}{N_A! N_B!} Z_{\{N_A, N_B\}} N_A \right) * \\
&* \left(\frac{1}{\Xi^2} \sum_{\{N_A, N_B\}} N_A \frac{z_A^{N_A - 1} z_B^{N_B}}{N_A! N_B!} Z_{\{N_A, N_B\}} \right) = \\
&= \frac{1}{\Xi} \sum_{\{N_A, N_B\}} \frac{z_A^{N_A} z_B^{N_B}}{N_A! N_B!} Z_{\{N_A, N_B\}} N_A^2 - \\
&- \left(\frac{1}{\Xi} \sum_{\{N_A, N_B\}} \frac{z_A^{N_A} z_B^{N_B}}{N_A! N_B!} Z_{\{N_A, N_B\}} N_A \right)^2 = \\
&= \langle N_A^2 \rangle - \langle N_A \rangle^2 = \langle \Delta N_A^2 \rangle
\end{aligned}
\tag{E.18}
$$

$$
\implies \frac{\partial c_A}{\partial \mu_A} = \frac{\langle \Delta N_A^2 \rangle}{k_B T V} = \frac{\langle \Delta N_A^2 \rangle}{\langle N_A \rangle} \frac{c_A}{k_B T}
$$



$$
\begin{aligned}
k_BTV\frac{\partial c_A}{\partial \mu_B} &= k_BTV\frac{\partial\langle N_A\rangle}{\partial\mu_B}\frac{1}{V} = \frac{\partial\langle N_A\rangle}{\partial \ln z_B} = \\
&= \frac{\partial}{\partial \ln z_B}\left(\frac{1}{\Xi}\sum_{\{N_A,N_B\}}\frac{z_A^{N_A}z_B^{N_B}}{N_A!N_B!}Z_{\{N_A,N_B\}}N_A\right) = \\
&= \frac{1}{\Xi}\frac{\partial}{\partial \ln z_B}\left(\sum_{\{N_A,N_B\}}\frac{z_A^{N_A}z_B^{N_B}}{N_A!N_B!}Z_{\{N_A,N_B\}}N_A\right)+ \\
&+ \frac{\partial}{\partial \ln z_B}\left(\frac{1}{\Xi}\right)\sum_{\{N_A,N_B\}}\frac{z_A^{N_A}z_B^{N_B}}{N_A!N_B!}Z_{\{N_A,N_B\}}N_A = \\
&= z_B\frac{1}{\Xi}\sum_{\{N_A,N_B\}}N_B\frac{z_A^{N_A}z_B^{N_B-1}}{N_A!N_B!}Z_{\{N_A,N_B\}}N_A- \\
&- z_B\left(\sum_{\{N_A,N_B\}}\frac{z_A^{N_A}z_B^{N_B}}{N_A!N_B!}Z_{\{N_A,N_B\}}N_A\right)* \\
&* \left(\frac{1}{\Xi^2}\sum_{\{N_A,N_B\}}N_B\frac{z_A^{N_A}z_B^{N_B-1}}{N_A!N_B!}Z_{\{N_A,N_B\}}\right) = \\
&= \frac{1}{\Xi}\sum_{\{N_A,N_B\}}\frac{z_A^{N_A}z_B^{N_B}}{N_A!N_B!}Z_{\{N_A,N_B\}}N_AN_B- \\
&- \left(\frac{1}{\Xi}\sum_{\{N_A,N_B\}}\frac{z_A^{N_A}z_B^{N_B}}{N_A!N_B!}Z_{\{N_A,N_B\}}N_A\right)* \\
&* \left(\frac{1}{\Xi}\sum_{\{N_A,N_B\}}\frac{z_A^{N_A}z_B^{N_B}}{N_A!N_B!}Z_{\{N_A,N_B\}}N_B\right) = \\
&= \langle N_AN_B\rangle - \langle N_A\rangle\langle N_B\rangle = \langle \Delta N_A\Delta N_B\rangle
\end{aligned}
\tag{E.19}
$$

$$
\Longrightarrow \frac{\partial c_A}{\partial \mu_B} = \frac{\langle \Delta N_A\Delta N_B\rangle}{k_BTV} = \frac{\langle \Delta N_A\Delta N_B\rangle}{\langle N_A\rangle}\frac{c_A}{k_BT}
$$

since $\langle \Delta N_A\Delta N_B\rangle = \langle (N_A - \langle N_A\rangle)(N_B - \langle N_B\rangle)\rangle = \langle N_AN_B\rangle - \langle N_A\rangle\langle N_B\rangle$ This simplification is valid due to the linearity of the ensemble average operation and to the fact that average values are constant. The expression for the derivative $\left(\frac{\partial \mu_A}{\partial c_A}\right)_{T,P}$ is the reciprocal of $\left(\frac{\partial c_A}{\partial \mu_A}\right)_{T,P}$ because both partial derivatives operate in the same thermodynamic space under the identical constraints of constant temperature and pressure. Thus,

$$\left(\frac{\partial \mu_A}{\partial c_A}\right)_{T,P} = \left[\left(\frac{\partial c_A}{\partial \mu_A}\right)_{\mu_B} - \frac{c_A}{c_B}\left(\frac{\partial c_A}{\partial \mu_B}\right)_{\mu_A}\right]^{-1} =$$
$$= \left[\frac{c_A}{k_B T}\frac{\langle \Delta N_A^2\rangle}{\langle N_A\rangle} - \frac{c_A}{c_B}\frac{c_A}{k_B T}\frac{\langle \Delta N_A \Delta N_B\rangle}{\langle N_A\rangle}\right]^{-1} = \tag{E.20}$$
$$= \frac{k_B T}{c_A}\left[\frac{\langle \Delta N_A^2\rangle}{\langle N_A\rangle} - \frac{\langle \Delta N_A \Delta N_B\rangle}{\langle N_B\rangle}\right]^{-1}$$

having considered that $\frac{c_A}{c_B}\frac{1}{\langle N_A\rangle} = \frac{\langle N_A\rangle}{V}\frac{V}{\langle N_B\rangle}\frac{1}{\langle N_A\rangle} = \frac{1}{\langle N_B\rangle}$. Inserting the Kirkwood-Buff integrals G_{AA} and G_{AB}

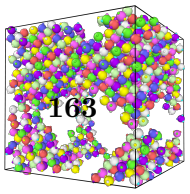
$$G_{AA} = V\frac{\langle N_A^2\rangle - \langle N_A\rangle^2}{\langle N_A\rangle^2} - \frac{1}{c_A} \qquad G_{AB} = V\frac{\langle N_A N_B\rangle - \langle N_A\rangle\langle N_B\rangle}{\langle N_A\rangle\langle N_B\rangle} \tag{E.21}$$

Eq. E.20 may be rewritten after some algebraic rearrangement as

$$\left(\frac{\partial \mu_A}{\partial c_A}\right)_{T,P} = \frac{1}{1 + c_A(G_{AA} - G_{AB})} \tag{E.22}$$

In conclusion, in the thermodynamic limit, *i.e.* for an infinite large system, Kirkwood-Buff integrals provide a rigorous connection between the microscopic structure (the radial distribution functions) and the macroscopic thermodynamic properties (chemical potentials) which describe the global response of the system to changes in the concentration of its components.

Eq. E.22 is formally exact in the thermodynamic limit, and its application to evaluate chemical potentials in molecular simulations, where the system is finite, presents several difficulties. The primary issue stems from the definition of Kirkwood-Buff integrals via radial distribution functions, which converge to unity $g^{(2)}(r) \to 1$ only as $r \to \infty$. In finite systems, the thermodynamic limit is never reached. Instead, the integration of $g^{(2)}(r)$ is performed over a fluctuating tail; particle correlations and particle number fluctuations are truncated. Moreover, since the surface to volume ratio is large, a non-negligible fraction of particles is affected by boundary effects. While periodic boundary conditions (PBC) effectively eliminate explicit surfaces, they do not resolve the intrinsic finiteness of the system and the evaluation of particle correlations and particle number fluctuations are distorted. In a cubic box of side length L, correlations at distances exceeding $L/2$ are suppressed, and spurious autocorrelations are introduced as particles interact with their own periodic images. This prevents density fluctuations from dissipating naturally, leading to an incorrect estimation of the variance. Furthermore, grand canonical Monte Carlo



(μVT) simulations suffer from poor sampling efficiency due to the low acceptance rate of particle insertion moves caused by frequent overlap. For these reasons, μVT simulations are avoided in favour of NPT simulations, where observables are computed within small sub-volumes v that effectively act as open μvT ensembles. However, this approach is still constrained by the closed nature of the global system. As a consequence, Kirkwood-Buff integrals without finite-size corrections can be reliably used only if simulations in the μVT ensemble of an infinite and open system were feasible. To address these limitations, recent work by Cheng [239] proposes reaching the thermodynamic limit in simulations by operating in reciprocal space: structure factors $S(\mathbf{k})$ are directly computed and their values are then extrapolated to the limit $\mathbf{k} \to \mathbf{0}$. In a binary AB mixture the structure factor S_{AB} is defined as

$$S_{AB}(\mathbf{k}) = \frac{\mathbf{1}}{\sqrt{\mathbf{N_A N_B}}} \langle \rho_{\mathbf{A}}(\mathbf{k}) \rho^*_{\mathbf{B}}(\mathbf{k}) \rangle \tag{E.23}$$

where $\mathbf{k}$ is the wave vector and $\rho_\alpha(\mathbf{k})$ is the Fourier transform of the instantaneous microscopic density $\rho_\alpha(\mathbf{r})$, previously defined in Eq. E.5. $\rho^*_\alpha(\mathbf{k})$ is the complex conjugate of $\rho_\alpha(\mathbf{k})$; since the density field is a real function $\rho^*_\alpha(\mathbf{k}) = \rho_\alpha(-\mathbf{k})$. $\rho_\alpha^(\mathbf{k})$ takes the following expression

$$\begin{aligned} \rho_\alpha(\mathbf{k}) &= \mathcal{F}\{\rho_\alpha(\mathbf{r})\} = \int \mathbf{e}^{-\mathbf{i}\mathbf{k}\cdot\mathbf{r}} \rho_\alpha(\mathbf{r}) \mathbf{dr} \\ &= \int e^{-i\mathbf{k}\cdot\mathbf{r}} \sum_{i=1}^{N_\alpha} \delta(\mathbf{r} - \mathbf{r_i}) \mathbf{dr} = \\ &= \sum_{i=1}^{N_\alpha} e^{-i\mathbf{k}\cdot\mathbf{r_i}} \end{aligned} \tag{E.24}$$

Hence, the structure factor $S_{AB}(\mathbf{k})$ in Eq. E.23, becomes

$$S_{AB}(\mathbf{k}) = \Big\langle \sum_{\mathbf{i=1}}^{\mathbf{N_A}} \sum_{\mathbf{j=1}}^{\mathbf{N_B}} \mathbf{e}^{-\mathbf{i}\mathbf{k}\cdot(\mathbf{r_{iA}} - \mathbf{r_{jB}})} \Big\rangle = \delta_{\mathbf{AB}} + \frac{\mathbf{1}}{\sqrt{\mathbf{N_A N_B}}} \Big\langle \sum_{\mathbf{i=1}}^{\mathbf{N_A}} \sum_{\substack{j=1 \\ j \neq i}}^{\mathbf{N_B}} \mathbf{e}^{-\mathbf{i}\mathbf{k}\cdot(\mathbf{r_{iA}} - \mathbf{r_{jB}})} \Big\rangle \tag{E.25}$$

The Kronecker delta δ_{AB} arises from the self-correlation terms ($i = j$), which exists only when $A = B$. In that case, $\sum_{i=1}^{N_A} e^0 = N_A$ and $\frac{N_A}{\sqrt{N_A N_B}} = \frac{N_A}{\sqrt{N_A^2}} = 1$. The second term, representing distinct particle correlations, can be rewritten using the two particle density $\rho^{(2)}_{AB}(\mathbf{r}', \mathbf{r}'')$ (Eq. E.10):

$$
\begin{aligned}
&\frac{1}{\sqrt{N_A N_B}}\Big\langle \sum_{i=1}^{N_A}\sum_{\substack{j=1\\ j\neq i}}^{N_B} e^{-i\mathbf{k}\cdot(\mathbf{r_{iA}}-\mathbf{r_{jB}})}\Big\rangle = \\
&= \frac{1}{\sqrt{N_A N_B}}\int\Big(\Big\langle \sum_{i=1}^{N_A}\sum_{\substack{j=1\\ j\neq i}}^{N_B}\delta(\mathbf{r_{iA}}-\mathbf{r}')\delta(\mathbf{r_{jB}}-\mathbf{r}'')\Big\rangle \mathbf{e}^{-\mathbf{ik}\cdot(\mathbf{r}'-\mathbf{r}'')}\Big)\mathbf{dr}'\mathbf{dr}'' = \\
&= \frac{1}{\sqrt{N_A N_B}}\int \rho^{(2)}_{AB}(\mathbf{r}',\mathbf{r}'')\mathbf{e}^{-\mathbf{ik}\cdot(\mathbf{r}'-\mathbf{r}'')}\mathbf{dr}'\mathbf{dr}'' \\
&= \frac{c_A c_B}{\sqrt{N_A N_B}} V\int g^{(2)}_{AB}(r)e^{-i\mathbf{k}\cdot\mathbf{r}}d\mathbf{r} = \\
&= \sqrt{c_A c_B}\int g^{(2)}_{AB}(r)e^{-i\mathbf{k}\cdot\mathbf{r}}d\mathbf{r}
\end{aligned}
\tag{E.26}
$$

where we have used the definition of the two particle radial distribution function $g^{(2)}_{AB}(\mathbf{r})$ (Eq. E.13) and the assumption of translational invariance. Since $g^{(2)}(r) \to 1$ as $r \to \infty$, the integral diverges. Its Fourier transform of $g^{(2)}_{AB}(r)$ contains a singular contribution: a Dirac delta function $\delta(\mathbf{k})$. This term contributes only at $\mathbf{k} = \mathbf{0}$, while it vanishes for all $\mathbf{k} \neq \mathbf{0}$. To obtain a well-behaved structure factor, we introduce the edge function $h(r) = g^{(2)}(r) - 1$. This subtraction removes the $\delta(\mathbf{k})$ contribution, which corresponds to the uniform background density. In scattering experiments, $\mathbf{k} = \mathbf{0}$ coincides with the beam direction and is thus not physically accessible. Consequently, this procedure isolates the relevant density-density correlation fluctuations, yielding:

$$
S_{AB}(\mathbf{k}) = \delta_{\mathbf{AB}} + \sqrt{\mathbf{c_A c_B}}\int[\mathbf{g^{(2)}_{AB}}(\mathbf{r}) - \mathbf{1}]\mathbf{e}^{-\mathbf{ik}\cdot\mathbf{r}}\mathbf{dr} \tag{E.27}
$$

By considering the thermodynamic limit, we observe that the structure factor is defined through Kirkwood-Buff integrals

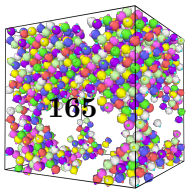
$$
S^0_{AB} = \lim_{\mathbf{k}\to\mathbf{0}} S_{AB}(\mathbf{k}) = \delta_{\mathbf{AB}} + \sqrt{\mathbf{c_A c_B}}\mathbf{G_{AB}} \tag{E.28}
$$

Consequently, the chemical potential variation with respect to the concentration in Eq. E.22 can be computed through structure factors at $\mathbf{k} = \mathbf{0}$. It can be readily verified that

$$
\left(\frac{\partial \mu_A}{\partial c_A}\right)_{T,P} = \frac{k_B T}{c_A}\left[\frac{1}{S^0_{AA} - S^0_{AB}\sqrt{\frac{c_A}{c_B}}}\right] \tag{E.29}
$$

is equal to Eq. E.22.

In conclusion, the thermodynamic limit in molecular simulations can be reliably accessed by computing structure factors in the NPT ensemble and extrapolating



their values to $\mathbf{k} = \mathbf{0}$. Although particle number fluctuations are rigorously related to thermodynamic quantities via the limit $\mathbf{k} \to \mathbf{0}$ in the μVT ensemble, the NPT ensemble can be regarded as an effective reservoir. For isotropic and homogeneous systems, the local structure remains identical across ensembles, even though global fluctuations vary between μVT and NPT. This interpretation allows the validity of the obtained relations to be extended to any sub-volume v of the larger NPT ensemble, including its whole volume. When using NPT simulations, the error introduced by the closed nature of the ensemble, primarily affects the $\mathbf{k} = \mathbf{0}$ point (corresponding to infinite distances), which is precisely the value that is extrapolated. Since it is not directly measured, the finite-size effects can be effectively regulated. In summary, by shifting the analysis to reciprocal space, Cheng's approach, named S0 method, overcomes the challenge of integrating non converging real space functions over finite size volumes. By extrapolating structure factors to the $\mathbf{k} = \mathbf{0}$ limit, the thermodynamic limit becomes accessible, enabling the computation of chemical potentials in mixture from molecular simulations of finite, closed systems.

Bibliography

[1] P. K. Chaffey, X. Guan, X. Wang, Y. Ruan, Y. Li, S. G. Miller, A. H. Tran, T. N. Koelsch, L. F. Pass, and Z. Tan, "Quantitative effects of O-linked glycans on protein folding," *Biochemistry*, vol. 56, no. 34, pp. 4539–4548, 2017.

[2] J. D. Perlmutter, M. R. Perkett, and M. F. Hagan, "Pathways for virus assembly around nucleic acids," *Journal of molecular biology*, vol. 426, no. 18, pp. 3148–3165, 2014.

[3] T. Christoff-Tempesta, Y. Cho, D.-Y. Kim, M. Geri, G. Lamour, A. J. Lew, X. Zuo, W. R. Lindemann, and J. H. Ortony, "Self-assembly of aramid amphiphiles into ultra-stable nanoribbons and aligned nanoribbon threads," *Nature Nanotechnology*, vol. 16, no. 4, pp. 447–454, 2021.

[4] S. soft matter group, "SATa CLAUSE: Assembling a special Christmas star." `https://www.youtube.com/watch?v=U7p5m9Lu1AU`, 2023.

[5] S. Whitelam and R. L. Jack, "The statistical mechanics of dynamic pathways to self-assembly," *Annual review of physical chemistry*, vol. 66, no. 1, pp. 143–163, 2015.

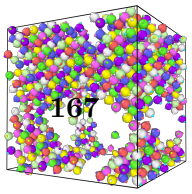
[6] S. K. Kumar, G. Kumaraswamy, B. L. Prasad, R. Bandyopadhyaya, S. Granick, O. Gang, V. N. Manoharan, D. Frenkel, and N. A. Kotov, "Nanoparticle assembly: a perspective and some unanswered questions," *Current Science*, pp. 1635–1641, 2017.

[7] L. Filion, M. Marechal, B. van Oorschot, D. Pelt, F. Smallenburg, and M. Dijkstra, "Efficient method for predicting crystal structures at finite temperature: variable box shape simulations," *Physical review letters*, vol. 103, no. 18, p. 188302, 2009.

[8] L. Filion and M. Dijkstra, "Prediction of binary hard-sphere crystal structures," *Physical Review E*, vol. 79, no. 4, p. 046714, 2009.

[9] L. Rovigatti, J. Russo, F. Romano, M. Matthies, L. Kroc, and P. Šulc, "A simple solution to the problem of self-assembling cubic diamond crystals," *Nanoscale*, vol. 14, no. 38, pp. 14268–14275, 2022.

[10] D. E. Pinto, P. Šulc, F. Sciortino, and J. Russo, "Design strategies for the self-assembly of polyhedral shells," *Proceedings of the National Academy of Sciences*, vol. 120, no. 16, p. e2219458120, 2023.



[11] D. E. Pinto, P. Šulc, F. Sciortino, and J. Russo, "Automating Blueprints for the Assembly of Colloidal Quasicrystal Clusters," *ACS nano*, vol. 19, no. 1, pp. 512–519, 2024.

[12] A.-Y. Jee, K. Lou, H.-S. Jang, K. H. Nagamanasa, and S. Granick, "Nanoparticle puzzles and research opportunities that go beyond state of the art," *Faraday discussions*, vol. 186, pp. 11–15, 2016.

[13] M. Dijkstra and E. Luijten, "From predictive modelling to machine learning and reverse engineering of colloidal self-assembly," *Nature materials*, vol. 20, no. 6, pp. 762–773, 2021.

[14] R. P. Feynman, "There's Plenty of Room at the Bottom," in *Annual Meeting of the American Physical Society*, (Pasadena, California), 1959. Caltech, December 29, 1959. Reprinted in *Caltech Engineering and Science*, vol. 23, no. 5, pp. 22–36 (1960).

[15] A. Bupathy, D. Frenkel, and S. Sastry, "Temperature protocols to guide selective self-assembly of competing structures," *Proceedings of the National Academy of Sciences*, vol. 119, no. 8, p. e2119315119, 2022.

[16] É. Ducrot, M. He, G.-R. Yi, and D. J. Pine, "Colloidal alloys with preassembled clusters and spheres," *Nature materials*, vol. 16, no. 6, pp. 652–657, 2017.

[17] D. R. Nelson, "Toward a tetravalent chemistry of colloids," *Nano Letters*, vol. 2, no. 10, pp. 1125–1129, 2002.

[18] V. N. Manoharan, M. T. Elsesser, and D. J. Pine, "Dense packing and symmetry in small clusters of microspheres," *Science*, vol. 301, no. 5632, pp. 483–487, 2003.

[19] Z. Zhang, A. S. Keys, T. Chen, and S. C. Glotzer, "Self-assembly of patchy particles into diamond structures through molecular mimicry," *Langmuir*, vol. 21, no. 25, pp. 11547–11551, 2005.

[20] F. Romano, J. Russo, and H. Tanaka, "Influence of patch-size variability on the crystallization of tetrahedral patchy particles," *Physical review letters*, vol. 113, no. 13, p. 138303, 2014.

[21] J. D. Halverson and A. V. Tkachenko, "DNA-programmed mesoscopic architecture," *Physical Review E*, vol. 87, no. 6, p. 062310, 2013.

[22] F. Romano and F. Sciortino, "Patterning symmetry in the rational design of colloidal crystals," *Nature communications*, vol. 3, p. 975, 2012.

[23] D. F. Tracey, E. G. Noya, and J. P. K. Doye, "Programming patchy particles to form complex periodic structures," *The Journal of Chemical Physics*, vol. 151, no. 22, p. 224506, 2019.

[24] M. C. Rechtsman, F. H. Stillinger, and S. Torquato, "Optimized interactions for targeted self-assembly: application to a honeycomb lattice," *Physical review letters*, vol. 95, no. 22, p. 228301, 2005.

[25] E. Marcotte, F. H. Stillinger, and S. Torquato, "Optimized monotonic convex pair potentials stabilize low-coordinated crystals," *Soft Matter*, vol. 7, no. 6, pp. 2332–2335, 2011.

[26] E. Marcotte, F. H. Stillinger, and S. Torquato, "Communication: Designed diamond ground state via optimized isotropic monotonic pair potentials," *The Journal of Chemical Physics*, vol. 138, no. 6, p. 061101, 2013.

[27] G. Zhang, F. Stillinger, and S. Torquato, "Probing the limitations of isotropic pair potentials to produce ground-state structural extremes via inverse statistical mechanics," *Physical Review E*, vol. 88, no. 4, p. 042309, 2013.

[28] M. Z. Miskin, G. Khaira, J. J. de Pablo, and H. M. Jaeger, "Turning statistical physics models into materials design engines," *Proceedings of the National Academy of Sciences*, vol. 113, no. 1, pp. 34–39, 2016.

[29] B. A. Lindquist, R. B. Jadrich, and T. M. Truskett, "Communication: Inverse design for self-assembly via on-the-fly optimization," *The Journal of Chemical Physics*, vol. 145, no. 11, p. 111101, 2016.

[30] D. Chen, G. Zhang, and S. Torquato, "Inverse design of colloidal crystals via optimized patchy interactions," *The Journal of Physical Chemistry B*, vol. 122, no. 35, pp. 8462–8468, 2018.

[31] R. Kumar, G. M. Coli, M. Dijkstra, and S. Sastry, "Inverse design of charged colloidal particle interactions for self assembly into specified crystal structures," *The Journal of chemical physics*, vol. 151, no. 8, 2019.

[32] S. Whitelam and I. Tamblyn, "Neuroevolutionary learning of particles and protocols for self-assembly," *Physical review letters*, vol. 127, no. 1, p. 018003, 2021.

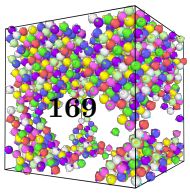
[33] C. Beneduce, F. Sciortino, P. Šulc, and J. Russo, "Engineering Azeotropy to Optimize the Self-Assembly of Colloidal Mixtures," *ACS nano*, vol. 17, no. 24, pp. 24841–24853, 2023.

[34] W. Bol, "Monte Carlo simulations of fluid systems of waterlike molecules," *Molecular Physics*, vol. 45, no. 3, pp. 605–616, 1982.

[35] N. Kern and D. Frenkel, "Fluid–fluid coexistence in colloidal systems with short-ranged strongly directional attraction," *The Journal of chemical physics*, vol. 118, no. 21, pp. 9882–9889, 2003.

[36] F. Smallenburg and F. Sciortino, "Liquids more stable than crystals in particles with limited valence and flexible bonds," *Nature Physics*, vol. 9, no. 9, pp. 554–558, 2013.

[37] E. G. Noya, I. Zubieta, D. J. Pine, and F. Sciortino, "Assembly of clathrates from tetrahedral patchy colloids with narrow patches," *The Journal of chemical physics*, vol. 151, no. 9, 2019.



[38] S. Whitelam and I. Tamblyn, "Learning to grow: Control of material self-assembly using evolutionary reinforcement learning," *Physical Review E*, vol. 101, no. 5, p. 052604, 2020.

[39] A. Bupathy, D. Frenkel, and S. Sastry, "Temperature Protocols to Guide Selective Self-Assembly of Competing Structures," *arXiv preprint arXiv:2110.11274*, 2021.

[40] R. Mao, E. Pretti, and J. Mittal, "Temperature-controlled reconfigurable nanoparticle binary superlattices," *ACS nano*, vol. 15, no. 5, pp. 8466–8473, 2021.

[41] J. Russo, F. Romano, L. Kroc, F. Sciortino, L. Rovigatti, and P. Šulc, "SAT-assembly: A new approach for designing self-assembling systems," *Journal of Physics: Condensed Matter*, vol. 34, no. 35, p. 354002, 2022.

[42] A. Murugan, J. Zou, and M. P. Brenner, "Undesired usage and the robust self-assembly of heterogeneous structures," *Nature communications*, vol. 6, no. 1, p. 6203, 2015.

[43] N. A. Mahynski, R. Mao, E. Pretti, V. K. Shen, and J. Mittal, "Grand canonical inverse design of multicomponent colloidal crystals," *Soft Matter*, vol. 16, no. 13, pp. 3187–3194, 2020.

[44] N. A. Mahynski, E. Pretti, V. K. Shen, and J. Mittal, "Using symmetry to elucidate the importance of stoichiometry in colloidal crystal assembly," *Nature Communications*, vol. 10, no. 1, p. 2028, 2019.

[45] E. Pretti, H. Zerze, M. Song, Y. Ding, R. Mao, and J. Mittal, "Size-dependent thermodynamic structural selection in colloidal crystallization," *Science advances*, vol. 5, no. 9, p. eaaw5912, 2019.

[46] F. Romano, J. Russo, L. Kroc, and P. Šulc, "Designing patchy interactions to self-assemble arbitrary structures," *Physical Review Letters*, vol. 125, no. 11, p. 118003, 2020.

[47] N. Een, "MiniSat: A SAT solver with conflict-clause minimization," in *Proc. SAT-05: 8th Int. Conf. on Theory and Applications of Satisfiability Testing*, pp. 502–518, 2005.

[48] W. Liu, M. Tagawa, H. L. Xin, T. Wang, H. Emamy, H. Li, K. G. Yager, F. W. Starr, A. V. Tkachenko, and O. Gang, "Diamond family of nanoparticle superlattices," *Science*, vol. 351, no. 6273, pp. 582–586, 2016.

[49] M. He, J. P. Gales, É. Ducrot, Z. Gong, G.-R. Yi, S. Sacanna, and D. J. Pine, "Colloidal diamond," *Nature*, vol. 585, no. 7826, pp. 524–529, 2020.

[50] K. Ho, C. T. Chan, and C. M. Soukoulis, "Existence of a photonic gap in periodic dielectric structures," *Physical Review Letters*, vol. 65, no. 25, p. 3152, 1990.

[51] C. M. Soukoulis and M. Wegener, "Optical metamaterials—more bulky and less lossy," *Science*, vol. 330, no. 6011, pp. 1633–1634, 2010.

[52] T. Ngo, C. Liddell, M. Ghebrebrhan, and J. Joannopoulos, "Tetrastack: Colloidal diamond-inspired structure with omnidirectional photonic band gap for low refractive index contrast," *Applied physics letters*, vol. 88, no. 24, p. 241920, 2006.

[53] F. Romano, E. Sanz, and F. Sciortino, "Crystallization of Tetrahedral Patchy Particles in Silico," *J. Chem. Phys.*, vol. 134, no. 17, p. 174502, 2011.

[54] A. Neophytou, D. Chakrabarti, and F. Sciortino, "Facile self-assembly of colloidal diamond from tetrahedral patchy particles via ring selection," *Proceedings of the National Academy of Sciences*, vol. 118, no. 48, 2021.

[55] C. Beneduce, D. EP Pinto, P. Šulc, F. Sciortino, and J. Russo, "Two-step nucleation in a binary mixture of patchy particles," *The Journal of Chemical Physics*, vol. 158, no. 15, p. 154502, 2023.

[56] C. Beneduce, D. E. Pinto, L. Rovigatti, F. Romano, P. Šulc, F. Sciortino, and J. Russo, "Falsifiability test for classical nucleation theory," *Physical Review Letters*, vol. 134, no. 14, p. 148201, 2025.

[57] G. van Anders, N. K. Ahmed, R. Smith, M. Engel, and S. C. Glotzer, "Entropically patchy particles: engineering valence through shape entropy," *Acs Nano*, vol. 8, no. 1, pp. 931–940, 2013.

[58] Z. Zhang and S. C. Glotzer, "Self-assembly of patchy particles," *Nano Letters*, vol. 4, no. 8, pp. 1407–1413, 2004.

[59] A. B. Pawar and I. Kretzschmar, "Fabrication, assembly, and application of patchy particles," *Macromolecular rapid communications*, vol. 31, no. 2, pp. 150–168, 2010.

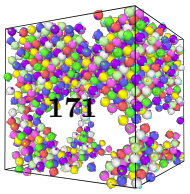
[60] E. Bianchi, R. Blaak, and C. N. Likos, "Patchy colloids: state of the art and perspectives," *Physical Chemistry Chemical Physics*, vol. 13, no. 14, pp. 6397–6410, 2011.

[61] F. Romano and F. Sciortino, "Colloidal self-assembly: patchy from the bottom up," *Nature materials*, vol. 10, no. 3, p. 171, 2011.

[62] K. Suzuki, K. Hosokawa, and M. Maeda, "Controlling the number and positions of oligonucleotides on gold nanoparticle surfaces," *Journal of the American Chemical Society*, vol. 131, no. 22, pp. 7518–7519, 2009.

[63] A. Jaekel, P. Lill, S. Whitelam, and B. Saccà, "Insights into the Structure and Energy of DNA Nanoassemblies," *Molecules*, vol. 25, no. 23, p. 5466, 2020.

[64] Z. Li, Y. Lim, I. Tanriover, W. Zhou, Y. Li, Y. Zhang, K. Aydin, S. C. Glotzer, and C. A. Mirkin, "DNA-mediated assembly of Au bipyramids into anisotropic light emitting kagome superlattices," *Science Advances*, vol. 10, no. 29, p. eadp3756, 2024.



[65] W. Zhou, Y. Lim, H. Lin, S. Lee, Y. Li, Z. Huang, J. S. Du, B. Lee, S. Wang, A. Sánchez-Iglesias, *et al.*, "Colloidal quasicrystals engineered with DNA," *Nature materials*, vol. 23, no. 3, pp. 424–428, 2024.

[66] J.-W. Kim, J.-H. Kim, and R. Deaton, "DNA-Linked Nanoparticle Building Blocks for Programmable Matter," *Angewandte Chemie International Edition*, vol. 50, no. 39, pp. 9185–9190, 2011.

[67] Y. Wang, Y. Wang, D. R. Breed, V. N. Manoharan, L. Feng, A. D. Hollingsworth, M. Weck, and D. J. Pine, "Colloids with valence and specific directional bonding," *Nature*, vol. 491, no. 7422, p. 51, 2012.

[68] L. Feng, R. Dreyfus, R. Sha, N. C. Seeman, and P. M. Chaikin, "DNA patchy particles," *Advanced Materials*, vol. 25, no. 20, pp. 2779–2783, 2013.

[69] P. W. K. Rothemund, "Folding DNA to create nanoscale shapes and patterns," *Nature*, vol. 440, no. 7082, pp. 297–302, 2006.

[70] Y. Tian, J. R. Lhermitte, L. Bai, T. Vo, H. L. Xin, H. Li, R. Li, M. Fukuto, K. G. Yager, J. S. Kahn, *et al.*, "Ordered three-dimensional nanomaterials using DNA-prescribed and valence-controlled material voxels," *Nature materials*, vol. 19, no. 7, pp. 789–796, 2020.

[71] Y. Xiong, S. Yang, Y. Tian, A. Michelson, S. Xiang, H. Xin, and O. Gang, "Three-Dimensional Patterning of Nanoparticles by Molecular Stamping," *ACS nano*, vol. 14, no. 6, pp. 6823–6833, 2020.

[72] N. C. Seeman and H. F. Sleiman, "DNA nanotechnology," *Nature Reviews Materials*, vol. 3, no. 1, pp. 1–23, 2017.

[73] T. Zhang, C. Hartl, K. Frank, A. Heuer-Jungemann, S. Fischer, P. C. Nickels, B. Nickel, and T. Liedl, "3D DNA Origami Crystals," *Adv. Mater.*, vol. 30, no. 28, p. 1800273, 2018.

[74] A. Cumberworth, D. Frenkel, and A. Reinhardt, "Simulations of DNA-origami self-assembly reveal design-dependent nucleation barriers," *Nano Letters*, vol. 22, no. 17, pp. 6916–6922, 2022.

[75] I. Chakraborty, D. J. Pearce, R. W. Verweij, S. C. Matysik, L. Giomi, and D. J. Kraft, "Self-assembly dynamics of reconfigurable colloidal molecules," *ACS nano*, vol. 16, no. 2, pp. 2471–2480, 2022.

[76] J. P. Doye, H. Fowler, D. Prešern, J. Bohlin, L. Rovigatti, F. Romano, P. Šulc, C. K. Wong, A. A. Louis, J. S. Schreck, *et al.*, "The oxDNA coarse-grained model as a tool to simulate DNA origami," in *DNA and RNA Origami: Methods and Protocols*, pp. 93–112, Springer, 2023.

[77] H. Liu, M. Matthies, J. Russo, L. Rovigatti, R. P. Narayanan, T. Diep, D. McKeen, O. Gang, N. Stephanopoulos, F. Sciortino, *et al.*, "Inverse design of a pyrochlore lattice of DNA origami through model-driven experiments," *Science*, vol. 384, no. 6697, pp. 776–781, 2024.

[78] S. M. Douglas, A. H. Marblestone, S. Teerapittayanon, A. Vazquez, G. M. Church, and W. M. Shih, "Rapid prototyping of 3D DNA-origami shapes with caDNAno," *Nucleic acids research*, vol. 37, no. 15, pp. 5001–5006, 2009.

[79] J. Bush, S. Singh, M. Vargas, E. Oktay, C.-H. Hu, and R. Veneziano, "Synthesis of DNA origami scaffolds: Current and emerging strategies," *Molecules*, vol. 25, no. 15, p. 3386, 2020.

[80] S. Dey, C. Fan, K. V. Gothelf, J. Li, C. Lin, L. Liu, N. Liu, M. A. Nijenhuis, B. Saccà, F. C. Simmel, *et al.*, "DNA origami," *Nature Reviews Methods Primers*, vol. 1, no. 1, p. 13, 2021.

[81] O. A. Vasilyev, B. A. Klumov, and A. V. Tkachenko, "Chromatic patchy particles: Effects of specific interactions on liquid structure," *Physical Review E*, vol. 92, no. 1, p. 012308, 2015.

[82] Z. Gong, T. Hueckel, G.-R. Yi, and S. Sacanna, "Patchy particles made by colloidal fusion," *Nature*, vol. 550, no. 7675, pp. 234–238, 2017.

[83] L. Rovigatti, P. Šulc, I. Z. Reguly, and F. Romano, "A comparison between parallelization approaches in molecular dynamics simulations on GPUs," *Journal of computational chemistry*, vol. 36, no. 1, pp. 1–8, 2015.

[84] E. Poppleton, R. Romero, A. Mallya, L. Rovigatti, and P. Šulc, "OxDNA. org: a public webserver for coarse-grained simulations of DNA and RNA nanostructures," *Nucleic acids research*, vol. 49, no. W1, pp. W491–W498, 2021.

[85] E. Poppleton, M. Matthies, D. Mandal, F. Romano, P. Šulc, and L. Rovigatti, "oxDNA: coarse-grained simulations of nucleic acids made simple," *Journal of Open Source Software*, vol. 8, no. 81, p. 4693, 2023.

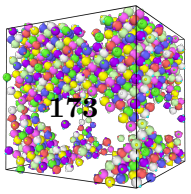
[86] J. SantaLucia Jr, "A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics," *Proceedings of the National Academy of Sciences*, vol. 95, no. 4, pp. 1460–1465, 1998.

[87] M. E. Fornace, J. Huang, C. T. Newman, N. J. Porubsky, M. B. Pierce, and N. A. Pierce, "NUPACK: Analysis and design of nucleic acid structures, devices, and systems," 2022.

[88] M. Wertheim, "Fluids with Highly Directional Attractive Forces. I. Statistical thermodynamics," *J. Stat. Phys.*, vol. 35, no. 1, pp. 19–34, 1984.

[89] W. G. Chapman, G. Jackson, and K. E. Gubbins, "Phase Equilibria of Associating Fluids: Spherical Molecules with Multiple Bonding Sites," *Mol. Phys.*, vol. 65, pp. 1–31, 1988.

[90] D. de Las Heras, J. M. Tavares, and M. M. T. da Gama, "Phase diagrams of binary mixtures of patchy colloids with distinct numbers of patches: the network fluid regime," *Soft Matter*, vol. 7, no. 12, pp. 5615–5626, 2011.



[91] E. Bianchi, J. Largo, P. Tartaglia, E. Zaccarelli, and F. Sciortino, "Phase Diagram of Patchy Colloids: Towards Empty Liquids," *Physical Review Letters*, vol. 97, p. 168301, 2006.

[92] D. d. l. Heras, J. M. Tavares, and M. M. T. da Gama, "Phase diagrams of binary mixtures of patchy colloids with distinct numbers and types of patches: The empty fluid regime," *The Journal of chemical physics*, vol. 134, no. 10, p. 104904, 2011.

[93] L. Rovigatti, D. de las Heras, J. M. Tavares, M. M. Telo da Gama, and F. Sciortino, "Computing the phase diagram of binary mixtures: A patchy particle case study," *The Journal of chemical physics*, vol. 138, no. 16, p. 164904, 2013.

[94] F. Seiferling, D. de Las Heras, and M. M. Telo da Gama, "Percolation in binary and ternary mixtures of patchy colloids," *The Journal of Chemical Physics*, vol. 145, no. 7, p. 074903, 2016.

[95] P. Teixeira and J. Tavares, "Phase behaviour of pure and mixed patchy colloids—Theory and simulation," *Current Opinion in Colloid & Interface Science*, vol. 30, pp. 16–24, 2017.

[96] R. Braz Teixeira, D. de Las Heras, J. M. Tavares, and M. M. Telo da Gama, "Phase behavior of a binary mixture of patchy colloids: Effect of particle size and gravity," *The Journal of Chemical Physics*, vol. 155, no. 4, p. 044903, 2021.

[97] J. Russo, F. Leoni, F. Martelli, and F. Sciortino, "The physics of Empty Liquids: from Patchy particles to Water," *Reports on Progress in Physics*, 2021.

[98] I. Nezbeda, J. Kolafa, and Y. V. Kalyuzhnyi, "Primitive model of water: II. Theoretical results for the structure and thermodynamic properties," *Molecular physics*, vol. 68, no. 1, pp. 143–160, 1989.

[99] F. Sciortino, E. Bianchi, J. F. Douglas, and P. Tartaglia, "Self-assembly of patchy particles into polymer chains: A parameter-free comparison between Wertheim theory and Monte Carlo simulation," *The Journal of chemical physics*, vol. 126, no. 19, p. 194903, 2007.

[100] B. Chen and J. I. Siepmann, "A novel Monte Carlo algorithm for simulating strongly associating fluids: Applications to water, hydrogen fluoride, and acetic acid," *The Journal of Physical Chemistry B*, vol. 104, no. 36, pp. 8725–8734, 2000.

[101] L. Rovigatti, J. Russo, and F. Romano, "How to simulate patchy particles," *The European Physical Journal E*, vol. 41, no. 5, p. 59, 2018.

[102] G. M. Torrie and J. P. Valleau, "Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling," *Journal of computational physics*, vol. 23, no. 2, pp. 187–199, 1977.

[103] P. Virnau and M. Müller, "Calculation of free energy through successive umbrella sampling," *The Journal of chemical physics*, vol. 120, no. 23, pp. 10925–10930, 2004.

[104] U. K. Deiters, "Differential Equations for the Calculation of Isopleths of Multicomponent Fluid Mixtures," *Fluid Phase Equilib.*, vol. 447, pp. 72–83, 2017.

[105] I. H. Bell and U. K. Deiters, "On the Construction of Binary Mixture p-x and T-x Diagrams from Isochoric Thermodynamics," *AIChE J.*, vol. 64, no. 7, pp. 2745–2757, 2018.

[106] A. Z. Panagiotopoulos, "Direct Determination of Phase Coexistence Properties of Fluids by Monte Carlo Simulation in a New Ensemble," *Mol. Phys.*, vol. 61, no. 4, pp. 813–826, 1987.

[107] A. Z. Panagiotopoulos, N. Quirke, M. Stapleton, and D. Tildesley, "Phase Equilibria by Simulation in the Gibbs Ensemble: Alternative Derivation, Generalization and Application to Mixture and Membrane Equilibria," *Mol. Phys.*, vol. 63, no. 4, pp. 527–545, 1988.

[108] J. M. Smith, H. C. Van Ness, M. M. Abbott, and M. T. Swihart, *Introduction to chemical engineering thermodynamics*. McGraw-Hill Singapore, 1949.

[109] W. J. Moore, *Physical Chemistry*. Prentice-Hall, 1962.

[110] K. Akahane, J. Russo, and H. Tanaka, "A possible four-phase coexistence in a single-component system," *Nature communications*, vol. 7, no. 1, pp. 1–8, 2016.

[111] D. Stopper, H. Hansen-Goos, R. Roth, and R. Evans, "Remnants of the disappearing critical point in chain-forming patchy fluids," *The Journal of Chemical Physics*, vol. 152, no. 11, p. 111101, 2020.

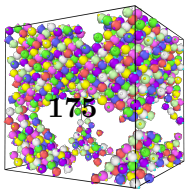
[112] J. Tavares and P. Teixeira, "Remnants of the disappearing critical point (s) in patchy fluids with distinct interaction patches," *The Journal of Chemical Physics*, vol. 153, no. 8, p. 086101, 2020.

[113] D. Frenkel and B. Smit, *Understanding molecular simulation: from algorithms to applications*, vol. 1. Elsevier, 2001.

[114] J. Russo, F. Leoni, F. Martelli, and F. Sciortino, "The physics of empty liquids: From patchy particles to water," *Reports on Progress in Physics*, vol. 85, no. 1, p. 016601, 2022.

[115] J. Russo, P. Tartaglia, and F. Sciortino, "Reversible gels of patchy particles: role of the valence," *The Journal of chemical physics*, vol. 131, no. 1, 2009.

[116] P. Swinkels, R. Sinaasappel, Z. Gong, S. Sacanna, W. Meyer, F. Sciortino, and P. Schall, "Networks of limited-valency patchy particles," *Physical Review Letters*, vol. 132, no. 7, p. 078203, 2024.



[117] F. Sciortino and E. Zaccarelli, "Equilibrium gels of limited valence colloids," *Current opinion in colloid & interface science*, vol. 30, pp. 90–96, 2017.

[118] P. J. Lu, E. Zaccarelli, F. Ciulla, A. B. Schofield, F. Sciortino, and D. A. Weitz, "Gelation of particles with short-range attraction," *Nature*, vol. 453, no. 7194, pp. 499–503, 2008.

[119] S. Buzzaccaro, R. Rusconi, and R. Piazza, ""Sticky" Hard Spheres: Equation of State, Phase Diagram, and Metastable Gels," *Physical Review Letters*, vol. 99, p. 098301, 2007.

[120] M. Rubinstein, R. H. Colby, *et al.*, *Polymer physics*, vol. 23. Oxford university press New York, 2003.

[121] H. Tanaka, H. Tong, R. Shi, and J. Russo, "Revealing key structural features hidden in liquids and glasses," *Nature Reviews Physics*, vol. 1, no. 5, pp. 333–348, 2019.

[122] L. D. Landau and E. M. Lifshitz, *Quantum mechanics: non-relativistic theory*, vol. 3. Elsevier, 2013.

[123] W. Lechner and C. Dellago, "Accurate determination of crystal structures based on averaged local bond order parameters," *The Journal of chemical physics*, vol. 129, no. 11, 2008.

[124] D. Kashchiev, P. G. Vekilov, and A. B. Kolomeisky, "Kinetics of two-step nucleation of crystals," *The Journal of chemical physics*, vol. 122, no. 24, p. 244706, 2005.

[125] D. Erdemir, A. Y. Lee, and A. S. Myerson, "Nucleation of crystals from solution: classical and two-step models," *Accounts of chemical research*, vol. 42, no. 5, pp. 621–629, 2009.

[126] P. G. Vekilov, "The two-step mechanism of nucleation of crystals in solution," *Nanoscale*, vol. 2, no. 11, pp. 2346–2357, 2010.

[127] G. I. Tóth, T. Pusztai, G. Tegze, G. Tóth, and L. Gránásy, "Amorphous nucleation precursor in highly nonequilibrium fluids," *Phys. Rev. Lett.*, vol. 107, no. 17, p. 175702, 2011.

[128] R. P. Sear, "The non-classical nucleation of crystals: microscopic mechanisms and applications to molecular crystals, ice and calcium carbonate," *International Materials Reviews*, vol. 57, no. 6, pp. 328–356, 2012.

[129] T. K. Haxton, L. O. Hedges, and S. Whitelam, "Crystallization and arrest mechanisms of model colloids," *Soft matter*, vol. 11, no. 48, pp. 9307–9320, 2015.

[130] J. Russo and H. Tanaka, "Nonclassical pathways of crystallization in colloidal systems," *MRS Bulletin*, vol. 41, no. 5, pp. 369–374, 2016.

[131] G. C. Sosso, J. Chen, S. J. Cox, M. Fitzner, P. Pedevilla, A. Zen, and A. Michaelides, "Crystal nucleation in liquids: Open questions and future challenges in molecular dynamics simulations," *Chemical reviews*, vol. 116, no. 12, pp. 7078–7116, 2016.

[132] J. F. Lutsko, "How crystals form: A theory of nucleation pathways," *Science advances*, vol. 5, no. 4, p. eaav7399, 2019.

[133] D. James, S. Beairsto, C. Hartt, O. Zavalov, I. Saika-Voivod, R. K. Bowles, and P. H. Poole, "Phase transitions in fluctuations and their role in two-step nucleation," *The Journal of chemical physics*, vol. 150, no. 7, 2019.

[134] C. Desgranges and J. Delhommelle, "Can Ordered Precursors Promote the Nucleation of Solid Solutions?," *Phys. Rev. Lett.*, vol. 123, no. 19, p. 195701, 2019.

[135] D. Kashchiev, "Classical nucleation theory approach to two-step nucleation of crystals," *Journal of Crystal Growth*, vol. 530, p. 125300, 2020.

[136] S. Lee, E. G. Teich, M. Engel, and S. C. Glotzer, "Entropic colloidal crystallization pathways via fluid–fluid transitions and multidimensional prenucleation motifs," *Proceedings of the National Academy of Sciences*, vol. 116, no. 30, pp. 14843–14851, 2019.

[137] P. R. ten Wolde and D. Frenkel, "Enhancement of protein crystal nucleation by critical density fluctuations," *Science*, vol. 277, no. 5334, pp. 1975–1978, 1997.

[138] P. Tan, N. Xu, and L. Xu, "Visualizing kinetic pathways of homogeneous nucleation in colloidal crystallization," *Nature Physics*, vol. 10, no. 1, pp. 73–79, 2014.

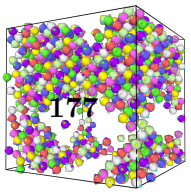
[139] H. Jiang, P. G. Debenedetti, and A. Z. Panagiotopoulos, "Nucleation in aqueous NaCl solutions shifts from 1-step to 2-step mechanism on crossing the spinodal," *The Journal of chemical physics*, vol. 150, no. 12, p. 124502, 2019.

[140] D. Gebauer, A. Völkel, and H. Cölfen, "Stable Prenucleation Calcium Carbonate Clusters," *Science*, vol. 322, no. 5909, pp. 1819–1822, 2008.

[141] E. M. Pouget, P. H. H. Bomans, J. A. C. M. Goos, P. M. Frederik, G. de With, and N. A. J. M. Sommerdijk, "The Initial Stages of Template-Controlled CaCO3 Formation Revealed by Cryo-TEM," *Science*, vol. 323, no. 5920, pp. 1455–1458, 2009.

[142] D. Fusco and P. Charbonneau, "Crystallization of asymmetric patchy models for globular proteins in solution," *Physical Review E*, vol. 88, no. 1, p. 012721, 2013.

[143] S. James, M. K. Quinn, and J. J. McManus, "The self assembly of proteins; probing patchy protein interactions," *Physical Chemistry Chemical Physics*, vol. 17, no. 7, pp. 5413–5420, 2015.



[144] J. J. McManus, P. Charbonneau, E. Zaccarelli, and N. Asherie, "The physics of protein self-assembly," *Current opinion in colloid & interface science*, vol. 22, pp. 73–79, 2016.

[145] D. Fusco and P. Charbonneau, "Soft matter perspective on protein crystal assembly," *Colloids and Surfaces B: Biointerfaces*, vol. 137, pp. 22–31, 2016.

[146] T. P. Knowles, M. Vendruscolo, and C. M. Dobson, "The amyloid state and its association with protein misfolding diseases," *Nature reviews Molecular cell biology*, vol. 15, no. 6, pp. 384–396, 2014.

[147] T. C. Michaels, D. Qian, A. Šarić, M. Vendruscolo, S. Linse, and T. P. Knowles, "Amyloid formation as a protein phase transition," *Nature Reviews Physics*, vol. 5, no. 7, pp. 379–397, 2023.

[148] L. Xu, S. V. Buldyrev, H. E. Stanley, and G. Franzese, "Homogeneous crystal nucleation near a metastable fluid-fluid phase transition," *Physical Review Letters*, vol. 109, no. 9, p. 095702, 2012.

[149] A. Stukowski, "Visualization and analysis of atomistic simulation data with OVITO–the Open Visualization Tool," *Modelling and simulation in materials science and engineering*, vol. 18, no. 1, p. 015012, 2009.

[150] C. Desgranges and J. Delhommelle, "Unraveling the coupling between demixing and crystallization in mixtures," *Journal of the American Chemical Society*, vol. 136, no. 23, pp. 8145–8148, 2014.

[151] Y. Wang, A. Lomakin, J. J. McManus, O. Ogun, and G. B. Benedek, "Phase behavior of mixtures of human lens proteins Gamma D and Beta B1," *Proceedings of the National Academy of Sciences*, vol. 107, no. 30, pp. 13282–13287, 2010.

[152] Y. Wang, A. Lomakin, R. F. Latypov, and G. B. Benedek, "Phase separation in solutions of monoclonal antibodies and the effect of human serum albumin," *Proceedings of the National Academy of Sciences*, vol. 108, no. 40, pp. 16606–16611, 2011.

[153] M. Heidenreich, J. M. Georgeson, E. Locatelli, L. Rovigatti, S. K. Nandi, A. Steinberg, Y. Nadav, E. Shimoni, S. A. Safran, J. P. Doye, *et al.*, "Designer protein assemblies with tunable phase diagrams in living cells," *Nature Chemical Biology*, vol. 16, no. 9, pp. 939–945, 2020.

[154] K. Popper, "Conjectures and Refutations," *New York: HarperandRow*, 1963.

[155] B. Peters, "Crystal nucleation: Rare made common and captured by Raman," *Proceedings of the National Academy of Sciences*, vol. 119, no. 21, p. e2204971119, 2022.

[156] K. Kelton and A. L. Greer, *Nucleation in condensed matter: applications in materials and biology*. Elsevier, 2010.

[157] R. C. Miller, R. J. Anderson, J. Kassner Jr, and D. E. Hagen, "Homogeneous nucleation rate measurements for water over a wide range of temperature and nucleation rate," *The Journal of Chemical Physics*, vol. 78, no. 6, pp. 3204–3211, 1983.

[158] R. Strey, P. Wagner, and Y. Viisanen, "The problem of measuring homogeneous nucleation rates and the molecular contents of nuclei: progress in the form of nucleation pulse measurements," *The Journal of Physical Chemistry*, vol. 98, no. 32, pp. 7748–7758, 1994.

[159] A. A. Manka, D. Brus, A.-P. Hyvärinen, H. Lihavainen, J. Wölk, and R. Strey, "Homogeneous water nucleation in a laminar flow diffusion chamber," *The Journal of chemical physics*, vol. 132, no. 24, 2010.

[160] S. C. C. Prado, J. P. Rino, and E. D. Zanotto, "Successful test of the classical nucleation theory by molecular dynamic simulations of BaS," *Computational Materials Science*, vol. 161, pp. 99–106, 2019.

[161] H. Lihavainen, Y. Viisanen, and M. Kulmala, "Homogeneous nucleation of n-pentanol in a laminar flow diffusion chamber," *The Journal of Chemical Physics*, vol. 114, no. 22, pp. 10031–10038, 2001.

[162] T. H. Zhang and X. Y. Liu, "How does a transient amorphous precursor template crystallization," *Journal of the American Chemical Society*, vol. 129, no. 44, pp. 13520–13526, 2007.

[163] K. Iland, J. Wölk, R. Strey, and D. Kashchiev, "Argon nucleation in a cryogenic nucleation pulse chamber," *The Journal of chemical physics*, vol. 127, no. 15, 2007.

[164] R. C. Tolman, "The effect of droplet size on surface tension," *The journal of chemical physics*, vol. 17, no. 3, pp. 333–337, 1949.

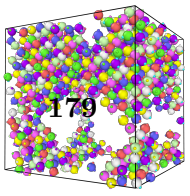
[165] J. Langer, "Theory of nucleation rates," *Physical Review Letters*, vol. 21, no. 14, p. 973, 1968.

[166] M. Plischke and Z. Rácz, "Active zone of growing clusters: Diffusion-limited aggregation and the Eden model," *Physical review letters*, vol. 53, no. 5, p. 415, 1984.

[167] P. A. Rikvold, H. Tomita, S. Miyashita, and S. W. Sides, "Metastable lifetimes in a kinetic Ising model: dependence on field and system size," *Physical Review E*, vol. 49, no. 6, p. 5080, 1994.

[168] I. Ford, "Nucleation theorems, the statistical mechanics of molecular clusters, and a revision of classical nucleation theory," *Physical Review E*, vol. 56, no. 5, p. 5615, 1997.

[169] R. A. Ramos, P. A. Rikvold, and M. Novotny, "Test of the Kolmogorov-Johnson-Mehl-Avrami picture of metastable decay in a model with microscopic dynamics," *Physical Review B*, vol. 59, no. 14, p. 9053, 1999.



[170] S. Sides, P. A. Rikvold, and M. Novotny, "Kinetic Ising model in an oscillating field: Avrami theory for the hysteretic response and finite-size scaling for the dynamic phase transition," *Physical Review E*, vol. 59, no. 3, p. 2710, 1999.

[171] D. Reguera and H. Reiss, "Extended Modified Liquid Drop- Dynamical Nucleation Theory (EMLD- DNT) Approach to Nucleation: A New Theory," *The Journal of Physical Chemistry B*, vol. 108, no. 51, pp. 19831–19842, 2004.

[172] J. Merikanto, E. Zapadinsky, A. Lauri, and H. Vehkamäki, "Origin of the Failure of Classical Nucleation Theory: Incorrect Description of the Smallest Clusters," *Physical review letters*, vol. 98, no. 14, p. 145702, 2007.

[173] D. Gebauer and H. Cölfen, "Prenucleation clusters and non-classical nucleation," *Nano Today*, vol. 6, no. 6, pp. 564–584, 2011.

[174] S. Prestipino, A. Laio, and E. Tosatti, "Systematic improvement of classical nucleation theory," *Physical review letters*, vol. 108, no. 22, p. 225701, 2012.

[175] J. R. Espinosa, C. Vega, C. Valeriani, and E. Sanz, "Seeding approach to crystal nucleation," *The Journal of chemical physics*, vol. 144, no. 3, 2016.

[176] G. M. Coli and M. Dijkstra, "An artificial neural network reveals the nucleation mechanism of a binary colloidal AB13 crystal," *ACS nano*, vol. 15, no. 3, pp. 4335–4346, 2021.

[177] W. Gispen and M. Dijkstra, "Brute-force nucleation rates of hard spheres compared with rare-event methods and classical nucleation theory," *The Journal of chemical physics*, vol. 159, no. 8, 2023.

[178] W. Gispen and M. Dijkstra, "Finding the differences: Classical nucleation perspective on homogeneous melting and freezing of hard spheres," *The Journal of Chemical Physics*, vol. 160, no. 14, 2024.

[179] D. W. Oxtoby, "Nucleation of first-order phase transitions," *Accounts of chemical research*, vol. 31, no. 2, pp. 91–97, 1998.

[180] P. R. Ten Wolde and D. Frenkel, "Homogeneous nucleation and the Ostwald step rule," *Physical Chemistry Chemical Physics*, vol. 1, no. 9, pp. 2191–2196, 1999.

[181] P. M. Piaggi and M. Parrinello, "kinetcting polymorphism in molecular crystals using orientational entropy," *Proceedings of the National Academy of Sciences*, vol. 115, no. 41, pp. 10251–10256, 2018.

[182] A. E. Van Driessche, N. Van Gerven, P. H. Bomans, R. R. Joosten, H. Friedrich, D. Gil-Carton, N. A. Sommerdijk, and M. Sleutel, "Molecular nucleation mechanisms and control strategies for crystal polymorph selection," *Nature*, vol. 556, no. 7699, pp. 89–94, 2018.

[183] C. L. Burcham, M. F. Doherty, B. G. Peters, S. L. Price, M. Salvalaglio, S. M. Reutzel-Edens, L. S. Price, R. K. R. Addula, N. Francia, V. Khanna, *et al.*, "Pharmaceutical Digital Design: From Chemical Structure through Crystal Polymorph to Conceptual Crystallization Process," *Crystal Growth & Design*, vol. 24, no. 13, pp. 5417–5438, 2024.

[184] W. Ostwald, "Studien über die Bildung und Umwandlung fester Körper: 1. Abhandlung: Übersättigung und überkaltung," *Zeitschrift für physikalische Chemie*, vol. 22, no. 1, pp. 289–330, 1897.

[185] J. De Yoreo and S. Whitelam, "Nucleation in atomic, molecular, and colloidal systems," *MRS Bulletin*, vol. 41, pp. 357–360, 2016.

[186] B. Sadigh, L. Zepeda-Ruiz, and J. L. Belof, "Metastable–solid phase diagrams derived from polymorphic solidification kinetics," *Proceedings of the National Academy of Sciences*, vol. 118, no. 9, p. e2017809118, 2021.

[187] M. Sleutel, J. Lutsko, A. E. Van Driessche, M. A. Durán-Olivencia, and D. Maes, "Observing classical nucleation theory at work by monitoring phase transitions with molecular precision," *Nature communications*, vol. 5, no. 1, p. 5598, 2014.

[188] P. J. Steinhardt, D. R. Nelson, and M. Ronchetti, "Bond-orientational order in liquids and glasses," *Physical Review B*, vol. 28, no. 2, p. 784, 1983.

[189] P. Rein ten Wolde, M. J. Ruiz-Montero, and D. Frenkel, "Numerical calculation of the rate of crystal nucleation in a Lennard-Jones system at moderate undercooling," *The Journal of chemical physics*, vol. 104, no. 24, pp. 9932–9947, 1996.

[190] C. Desgranges and J. Delhommelle, "Crystallization mechanisms for supercooled liquid Xe at high pressure and temperature: Hybrid Monte Carlo molecular simulations," *Physical Review B—Condensed Matter and Materials Physics*, vol. 77, no. 5, p. 054201, 2008.

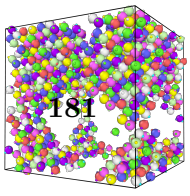
[191] J. Russo and H. Tanaka, "The microscopic pathway to crystallization in supercooled liquids," *Scientific reports*, vol. 2, p. 505, 2012.

[192] J. Russo and H. Tanaka, "Selection mechanism of polymorphs in the crystal nucleation of the Gaussian core model," *Soft Matter*, vol. 8, no. 15, pp. 4206–4215, 2012.

[193] J. F. Lutsko and C. Schoonen, "A microscopic approach to crystallization: challenging the classical/non-classical dichotomy," *arXiv preprint arXiv:2406.17543*, 2024.

[194] F. Zhang, "Nonclassical nucleation pathways in protein crystallization," *Journal of Physics: Condensed Matter*, vol. 29, no. 44, p. 443002, 2017.

[195] J. Russo and H. Tanaka, "Crystal nucleation as the ordering of multiple order parameters," *The Journal of Chemical Physics*, vol. 145, no. 21, 2016.



[196] S. Becker, E. Devijver, R. Molinier, and N. Jakse, "Unsupervised topological learning approach of crystal nucleation," *Scientific Reports*, vol. 12, no. 1, p. 3195, 2022.

[197] C. Schoonen and J. F. Lutsko, "Crystal polymorphism induced by surface tension," *Physical review letters*, vol. 129, no. 24, p. 246101, 2022.

[198] W. Gispen, G. M. Coli, R. van Damme, C. P. Royall, and M. Dijkstra, "Crystal polymorph selection mechanism of hard spheres hidden in the fluid," *ACS nano*, vol. 17, no. 9, pp. 8807–8814, 2023.

[199] C. Beneduce, G. Mastriani, P. Šulc, F. Sciortino, and J. Russo, "Inverse Thermodynamics: Designing Interactions for Targeted Phase Behavior," *The Journal of Physical Chemistry B*, vol. 129, no. 39, pp. 10153–10161, 2025.

[200] Z. Preisler, T. Vissers, F. Smallenburg, and F. Sciortino, "Crystals of Janus colloids at various interaction ranges," *The Journal of Chemical Physics*, vol. 145, no. 6, 2016.

[201] T. A. Nguyen, A. Newton, D. J. Kraft, P. G. Bolhuis, and P. Schall, "Tuning patchy bonds induced by critical Casimir forces," *Materials*, vol. 10, no. 11, p. 1265, 2017.

[202] T. Dasgupta, G. M. Coli, and M. Dijkstra, "Tuning the glass transition: enhanced crystallization of the laves phases in nearly hard spheres," *ACS nano*, vol. 14, no. 4, pp. 3957–3968, 2020.

[203] T. Dwyer, T. C. Moore, J. A. Anderson, and S. C. Glotzer, "Tunable assembly of host–guest colloidal crystals," *Soft Matter*, vol. 19, no. 36, pp. 7011–7019, 2023.

[204] W. Van Saarloos, V. Vitelli, and Z. Zeravcic, *Soft Matter: Concepts, Phenomena, and Applications*. Princeton University Press, 2024.

[205] S. Sacanna and D. J. Pine, "Shape-anisotropic colloids: Building blocks for complex assemblies," *Current opinion in colloid & interface science*, vol. 16, no. 2, pp. 96–105, 2011.

[206] I. Chakraborty, V. Meester, C. van der Wel, and D. J. Kraft, "Colloidal joints with designed motion range and tunable joint flexibility," *Nanoscale*, vol. 9, no. 23, pp. 7814–7821, 2017.

[207] D. Chintha, S. K. Veesam, E. Boattini, L. Filion, and S. N. Punnathanam, "Modeling of effective interactions between ligand coated nanoparticles through symmetry functions," *The Journal of Chemical Physics*, vol. 155, no. 24, 2021.

[208] Y. Zhang, G. Giunta, H. Liang, and M. Dijkstra, "Shape-induced crystallization of binary DNA-functionalized nanocubes," *The Journal of Chemical Physics*, vol. 158, no. 18, 2023.

[209] J. A. Diaz A, J. S. Oh, G.-R. Yi, and D. J. Pine, "Photo-printing of faceted DNA patchy particles," *Proceedings of the National Academy of Sciences*, vol. 117, no. 20, pp. 10645–10653, 2020.

[210] A. McMullen, S. Hilgenfeldt, and J. Brujic, "DNA self-organization controls valence in programmable colloid design," *Proceedings of the National Academy of Sciences*, vol. 118, no. 46, p. e2112604118, 2021.

[211] D. J. Kraft, J. Hilhorst, M. A. Heinen, M. J. Hoogenraad, B. Luigjes, and W. K. Kegel, "Patchy polymer colloids with tunable anisotropy dimensions," *The Journal of Physical Chemistry B*, vol. 115, no. 22, pp. 7175–7181, 2011.

[212] D. J. Kraft, R. Ni, F. Smallenburg, M. Hermes, K. Yoon, D. A. Weitz, A. van Blaaderen, J. Groenewold, M. Dijkstra, and W. K. Kegel, "Surface roughness directed self-assembly of patchy particles into colloidal micelles," *Proceedings of the National Academy of Sciences*, vol. 109, no. 27, pp. 10787–10792, 2012.

[213] W. Zhou, Y. Li, K. Je, T. Vo, H. Lin, B. E. Partridge, Z. Huang, S. C. Glotzer, and C. A. Mirkin, "Space-tiled colloidal crystals from DNA-forced shape-complementary polyhedra pairing," *Science*, vol. 383, no. 6680, pp. 312–319, 2024.

[214] F. Sciortino, A. Giacometti, and G. Pastore, "Phase diagram of Janus particles," *Physical review letters*, vol. 103, no. 23, p. 237801, 2009.

[215] J. Russo, J. Tavares, P. Teixeira, M. Telo da Gama, and F. Sciortino, "Reentrant phase diagram of network fluids," *Physical review letters*, vol. 106, no. 8, p. 085703, 2011.

[216] S. Roldán-Vargas, F. Smallenburg, W. Kob, and F. Sciortino, "Phase diagram of a reentrant gel of patchy particles," *The Journal of chemical physics*, vol. 139, no. 24, 2013.

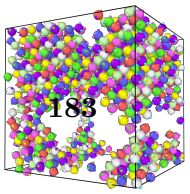
[217] E. Bianchi, J. Largo, P. Tartaglia, E. Zaccarelli, and F. Sciortino, "Phase diagram of patchy colloids: Towards empty liquids," *Physical review letters*, vol. 97, no. 16, p. 168301, 2006.

[218] M. F. Hagan and G. M. Grason, "Equilibrium mechanisms of self-limiting assembly," *Reviews of modern physics*, vol. 93, no. 2, p. 025008, 2021.

[219] S. Semwal, C. Clowe-Coish, I. Saika-Voivod, and A. Yethiraj, "Tunable colloids with dipolar and depletion interactions: Toward field-switchable crystals and gels," *Physical Review X*, vol. 12, no. 4, p. 041021, 2022.

[220] C. L. Kennedy, D. Sayasilpi, P. Schall, and J.-M. Meijer, "Self-assembly of colloidal cube superstructures with critical Casimir attractions," *Journal of Physics: Condensed Matter*, vol. 34, no. 21, p. 214005, 2022.

[221] E. Marino, R. A. LaCour, T. C. Moore, S. W. van Dongen, A. W. Keller, D. An, S. Yang, D. J. Rosen, G. Gouget, E. H. Tsai, *et al.*, "Crystallization of binary nanocrystal superlattices and the relevance of short-range attraction," *Nature Synthesis*, vol. 3, no. 1, pp. 111–122, 2024.



[222] T. E. Videbæk, D. Hayakawa, G. M. Grason, M. F. Hagan, S. Fraden, and W. B. Rogers, "Economical routes to size-specific assembly of self-closing structures," *Science advances*, vol. 10, no. 27, p. eado5979, 2024.

[223] Y. Shelke, D. J. Pearce, and D. J. Kraft, "Optimal self-assembly pathways towards colloidal lattices with tunable flexibility," *arXiv preprint arXiv:2406.19260*, 2024.

[224] F. Lu, Y. Zhang, T. Dwyer, A. Michelson, T. C. Moore, H. Yan, K. Kisslinger, H. Zhang, X. Chen, S. C. Glotzer, *et al.*, "Octo-diamond crystal of nanoscale tetrahedra with interchanging chiral motifs," *Nature materials*, pp. 1–9, 2025.

[225] J. De Yoreo, S. Y. Schmid, B. Helfrecht, A. Stegmann, B. Legg, H. Pyles, J. Chen, J. R. Edison, M. Ziatdinov, Z. Preisler, *et al.*, "Impact of Colloidal Forces and Broken Symmetry on the Design of a Protein-Material Interface," 2025.

[226] L. B. Frechette, N. Sundararajan, F. Caballero, A. Trubiano, and M. F. Hagan, "Computer simulations show that liquid-liquid phase separation enhances self-assembly," *arXiv preprint arXiv:2505.10400*, 2025.

[227] H. N. Lekkerkerker, W.-K. Poon, P. N. Pusey, A. Stroobants, and P. . Warren, "Phase behaviour of colloid+ polymer mixtures," *Europhysics Letters*, vol. 20, no. 6, p. 559, 1992.

[228] E. J. Meijer and D. Frenkel, "Colloids dispersed in polymer solutions. A computer simulation study," *The Journal of chemical physics*, vol. 100, no. 9, pp. 6873–6887, 1994.

[229] P. Hemmer, E. Velasco, L. Mederos, G. Navascués, and G. Stell, "Solid-solid transitions induced by repulsive interactions," *The Journal of Chemical Physics*, vol. 114, no. 5, pp. 2268–2275, 2001.

[230] S. Roldán-Vargas, F. Smallenburg, W. Kob, and F. Sciortino, "Gelling by heating," *Scientific reports*, vol. 3, no. 1, p. 2451, 2013.

[231] W. B. Rogers and V. N. Manoharan, "Programming colloidal phase transitions with DNA strand displacement," *Science*, vol. 347, no. 6222, pp. 639–642, 2015.

[232] F. Bomboi, F. Romano, M. Leo, J. Fernandez-Castanon, R. Cerbino, T. Bellini, F. Bordi, P. Filetici, and F. Sciortino, "Re-entrant DNA gels," *Nature communications*, vol. 7, no. 1, p. 13191, 2016.

[233] T. Hvozd, Y. V. Kalyuzhnyi, V. Vlachy, and P. T. Cummings, "Empty liquid state and re-entrant phase behavior of the patchy colloids confined in porous media," *The Journal of Chemical Physics*, vol. 156, no. 16, 2022.

[234] P. R. t. Wolde and D. Frenkel, "Enhancement of protein crystal nucleation by critical density fluctuations," *Science*, vol. 277, no. 5334, pp. 1975–1978, 1997.

[235] J. Russo, F. Romano, and H. Tanaka, "Glass forming ability in systems with competing orderings," *Physical Review X*, vol. 8, no. 2, p. 021040, 2018.

[236] A. Neophytou, F. Sciortino, and J. Russo, "Designing the self-assembly of disordered materials via color frustration," *Advanced Materials*, 2025 in pressing.

[237] D. Nykypanchuk, M. M. Maye, D. Van Der Lelie, and O. Gang, "DNA-guided crystallization of colloidal nanoparticles," *Nature*, vol. 451, no. 7178, pp. 549–552, 2008.

[238] S. Y. Park, A. K. Lytton-Jean, B. Lee, S. Weigand, G. C. Schatz, and C. A. Mirkin, "DNA-programmable nanoparticle crystallization," *Nature*, vol. 451, no. 7178, pp. 553–556, 2008.

[239] B. Cheng, "Computing chemical potentials of solutions from structure factors," *The Journal of Chemical Physics*, vol. 157, no. 12, 2022.

[240] J. G. Kirkwood and F. P. Buff, "The statistical mechanical theory of solutions. I," *The Journal of chemical physics*, vol. 19, no. 6, pp. 774–777, 1951.

[241] J.-L. Barrat and J.-P. Hansen, *Basic concepts for simple and complex liquids*. Cambridge University Press, 2003.

[242] C. Vega, E. Sanz, J. Abascal, and E. Noya, "Determination of phase diagrams via computer simulation: methodology and applicationsto water, electrolytes and proteins," *Journal of Physics: Condensed Matter*, vol. 20, no. 15, p. 153101, 2008.

[243] C. L. Bassani, G. Van Anders, U. Banin, D. Baranov, Q. Chen, M. Dijkstra, M. S. Dimitriyev, E. Efrati, J. Faraudo, O. Gang, *et al.*, "Nanocrystal assemblies: Current advances and open problems," *ACS nano*, vol. 18, no. 23, pp. 14791–14840, 2024.

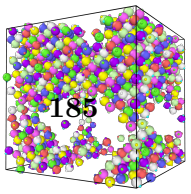
[244] D. Hayakawa, T. E. Videbæk, G. M. Grason, and W. B. Rogers, "Symmetry-guided inverse design of self-assembling multiscale DNA origami tilings," *ACS nano*, vol. 18, no. 29, pp. 19169–19178, 2024.

[245] A. Michelson, L. Shani, J. S. Kahn, D. C. Redeker, W.-I. Lee, K. R. DeOlivares, K. Kisslinger, N. Tiwale, H. Yan, A. Pattammattel, *et al.*, "Scalable fabrication of Chip-integrated 3D-nanostructured electronic devices via DNA-programmable assembly," *Science Advances*, vol. 11, no. 13, p. eadt5620, 2025.

[246] J. D. Joannopoulos, S. G. Johnson, J. N. Winn, and R. D. Meade, *Photonic Crystals: Molding the Flow of Light*. Princeton University Press, 2nd ed., 2008.

[247] M. Maldovan and E. L. Thomas, "Diamond-structured photonic crystals," *Nature materials*, vol. 3, no. 9, pp. 593–600, 2004.

[248] S. G. Johnson and J. D. Joannopoulos, "Block-iterative frequency-domain methods for Maxwell's equations in a planewave basis," *Optics express*, vol. 8, no. 3, pp. 173–190, 2001.



[249] J. C. W. Lee and C. Chan, "Polarization gaps in spiral photonic crystals," *Optics express*, vol. 13, no. 20, pp. 8083–8088, 2005.

[250] N. D. Lemke, J. Von Stecher, J. Sherman, A. Rey, C. Oates, and A. Ludlow, "p-Wave cold collisions in an optical lattice clock," *Physical Review Letters*, vol. 107, no. 10, p. 103902, 2011.

[251] W. Flavell, A. Neophytou, A. Demetriadou, T. Albrecht, and D. Chakrabarti, "Programmed Self-Assembly of Single Colloidal Gyroids for Chiral Photonic Crystals," *Advanced Materials*, vol. 35, no. 23, p. 2211197, 2023.

[252] A. F. Oskooi, D. Roundy, M. Ibanescu, P. Bermel, J. D. Joannopoulos, and S. G. Johnson, "MEEP: A flexible free-software package for electromagnetic simulations by the FDTD method," *Computer Physics Communications*, vol. 181, no. 3, pp. 687–702, 2010.

[253] S. soft matter group, "SATa CLAUSE: When magic snow falls, a enchanted meadow blossoms." `https://www.youtube.com/watch?v=0A7xmMguwjg`, 2024.

[254] M. S. Wertheim, "Fluids with highly directional attractive forces. II. Thermodynamic perturbation theory and integral equations," *Journal of statistical physics*, vol. 35, no. 1, pp. 35–47, 1984.

[255] M. S. Wertheim, "Fluids with highly directional attractive forces. III. Multiple attraction sites," *Journal of statistical physics*, vol. 42, no. 3, pp. 459–476, 1986.

[256] M. S. Wertheim, "Fluids with highly directional attractive forces. IV. Equilibrium polymerization," *Journal of statistical physics*, vol. 42, no. 3, pp. 477–492, 1986.

[257] W. Zmpitas and J. Gross, "Detailed pedagogical review and analysis of Wertheim's thermodynamic perturbation theory," *Fluid Phase Equilibria*, vol. 428, pp. 121–152, 2016.

[258] W. G. Chapman, "Ph.D. Thesis," 1988.

[259] W. G. Chapman, G. Jackson, and K. E. Gubbins, "Phase equilibria of associating fluids: chain molecules with multiple bonding sites," *Molecular Physics*, vol. 65, no. 5, pp. 1057–1079, 1988.

[260] W. G. Chapman, K. E. Gubbins, G. Jackson, and M. Radosz, "SAFT: Equation-of-state solution model for associating fluids," *Fluid Phase Equilibria*, vol. 52, pp. 31–38, 1989.

[261] W. G. Chapman, K. E. Gubbins, G. Jackson, and M. Radosz, "New reference equation of state for associating liquids," *Industrial & engineering chemistry research*, vol. 29, no. 8, pp. 1709–1721, 1990.

[262] X. Chen and F. Forstmann, "The demixing and gas–liquid instability of a binary Yukawa fluid," *The Journal of chemical physics*, vol. 97, no. 5, pp. 3696–3703, 1992.

[263] R. Fantoni, D. Gazzillo, and A. Giacometti, "Thermodynamic instabilities of a binary mixture of sticky hard spheres," *Physical Review E—Statistical, Nonlinear, and Soft Matter Physics*, vol. 72, no. 1, p. 011503, 2005.

[264] Y. B. Zeldovich, "On the theory of new phase formation: cavitation," *Acta Physicochem., USSR*, vol. 18, p. 1, 1943.

[265] J. Frenkel, *Kinetic Theory of Liquids*. Oxford: Oxford University Press, 1946.

[266] J.-P. Hansen and I. R. McDonald, *Theory of simple liquids: with applications to soft matter*. Academic press, 2013.

