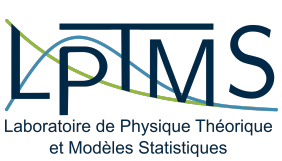
Frustrated Self Assembly with Multiple Particle Types

Pietro Caracciolo di Torella

Martin Lenz
Alessandro Pelizzola

July 27, 2024



Contents

1 Introduction

Self-assembly is a process where individual components spontaneously combine to form larger structures. In the case of protein aggregates the

Of particular interest are fibers. This specific type of aggregate appears to be the most common structure formed when the self-assembly process goes awry. For instance, pathological diseases such as Alzheimer's [6], Parkinson's Disease, and type II Diabetes [4] are believed to be caused by the formation of fibrillar aggregates instead of the biologically functional ones.

Self-assembly also holds significant importance as a construction technique for the development of novel materials [3]. Particularly when dealing with minute building blocks necessitating precise adjustments to attain specific material properties, self-assembly often is an interesting way to take. Achieving the desired structure mandates finely tuning the interactions among the constituents. A notable challenge in orchestrating self-assembled structures lies in ensuring they attain substantial yet finite dimensions. This objective is facilitated by a phenomenon known as self-limitation. Under attractive interactions, constituents continuously aggregate until exhaustion of available material. Conversely, repulsive interactions tend to maintain the constituents in a dispersed state. An intermediate regime exists wherein the structure self-limits its growth, reaching a critical size beyond which further aggregation becomes unfavorable, naturally ceasing further expansion.

In experiments it has been often observed that many different proteins, tend to aggregate in very similar aggregates. The process seem to be almost independent from the specifcities of each biological process. A systematic in vivo study of supramolecular self aggregation [1] has showed that across 73 different mutated proteins 30 of them aggregated in fibrils. This suggests that the principles governing self-assembly might be independent of the details of each biological system. For this reason it has been of large interest to describe the process of self-assembly using a physical model, where all details of the specific biological system have been washed away.

Proteins are objects with a complex shape, this complexity is reflected in the interactions that each protein has with other proteins. The location of different binding sites on the protein's surface can be highly anisotropic. Therefore the interaction between two proteins strongly depends on the relative orientations of the two. When multiple proteins come together there is no guarantee that all these complex objects with anisotropic interactions can fit perfectly together. This mismatch related to incompatible interactions is what is generally called geometrical frustration. We propose a model of self-assembly where geometrical frustration arises from the incompatible interactions of simple particles. L. Koehler, a former member of our group, has already implemented a 2 dimensional version of the model to study self-aggregation for identical sub-units. In her Ph.D. Thesis [2] she was able to observe many aggregate shapes reported before in experiments (crystals, sponges, micelles, fibers etc.). Different aggregates correspond to ground state configurations of particles with different interactions. In this work [2], the results were limited to systems with identical particles.

However, in the cell, many different biological complexes populate the environment. All of them in principle can come together and participate to the formation of multi components aggregate. Therefore the case where only identical sub-units participate to the assembly is a rough approximation of what actually happens in the self-assembly of real proteins complexes.

When multiple distinct species participate to the self-assembly process, they can take advantage of the combinatorial freedom related to have multiple components and co-assemble. In other words multiple species can relax the frustration arising from incompatible interactions by mixing. If the number of different species is large enough, co-assembly can enable the mixture to combine its sub-units in a way such that none of them is frustrated in the lowest energy state. How many distinct species are required for this to happen will depend on how complex are the single sub-units, and consequently their interactions.

The goal of this project is to study the interplay between complexity of the single sub-units, and number of species. When the particles are simple enough, with a relatively small number of distinct species we can expect that the system can relax all the frustration. The more complex the single sub-units become, the larger the number of species required to ease all the frustration.

In this project we will try to address this goal using the framework of [2] and adapting it to binary mixtures. This is a first generalization of the problem of identical particles and will be used as a benchmark for further studies.

First, we will look for those binary mixtures which use mixing to relax frustration. In order to find them we will simulate different mixtures and look for their ground state configuration. We will then try to classify these low energy states to understand which mixtures can ease all the frustration combining their interactions. Moreover, we will be also interested in those pairs of particles, that remain frustrated also with two components. The comparison between these two cases might then provide some insights on what characteristics makes the group of species that remain frustrated in binary mixtures more complex than the one which can relax all the frustration.

2 A model for mixtures of particles with complex interactions

In this Section we will describe the theoretical framework used in this report to model mixtures of two species of particles with complex interations.

The discussion is split in four main units. In Section 2.1 We describe in detail the model introduced by Koehler [2] and the framework of Local Energy Landscape described by Ronceray and Le Floch in [5]. The focus is on how we adapted the first to the case of mixtures of two species. Then, in Section 2.2 we will present the two different ensembles in which we are going to study the behaviour of binary mixture; the Canonical and the Grand Canonical ensemble. After that, in Section 2.3 we describe the method that we implemented to to sample the huge space of possible binary mixtures. Finally in Section 2.4 we will show how to use the symmetries of the model to reduce further the size of the problem, eliminating equivalent mixtures.

2.1 The lattice model

This section will be dedicated to the description of the model proposed in [2] to describe the behaviour of particles with complex interactions. To first get acquainted with the model we will describe it in its original version, where all particles are identical.

Particles are hexagons that sit on the vertices of a triangular lattice. The reason to use a triangular lattice is that on the triangular lattice frustration due to incompatible interaction can arise with only three sites. One famous example of this is the antiferromagnetic Ising model. Once we choose the lattice, the shape of the particle is the shape of its Wigner-Seitz cell. In this way, it will be possible to perfectly tile the the plane placing hexagonal particles on the vertices of the lattice. Each particle can be described by the set of its six faces which we will label as a, b, c... which are represented by different colors in Figure 2. When particles sit on nearest neighbouring sites of the lattice they interact putting in contact a pair of faces, three different examples are shown in Figure 2.

(a) pair of faces (a, b) (b) pair of faces (b, a) (c) pair of faces (c, a)

Figure 2: Three examples of particles interacting through a pair of faces. (a) and (b) correspond to the same local structure, instead (c) is a different one.

Each particle can have 6 distinct orientation while occupying the same lattice position. Therefore two particles can face 36 possible pair of faces (a, b) while sitting on neighbouring sites. However the energies of interaction between these 36 pair of faces are not all independent. The pair (a, b) and (b, a) will be associated with the same interaction energy. Using the nomenclature of the Local Energy Landscape framework introduced in [5] we can say that the pairs (a, b) and (b, a) are the same local structure.

In this model we can define a local structure as an unordered pair of faces. Only distinct local structures will be associated with different energies. This makes sense, for example we expect that

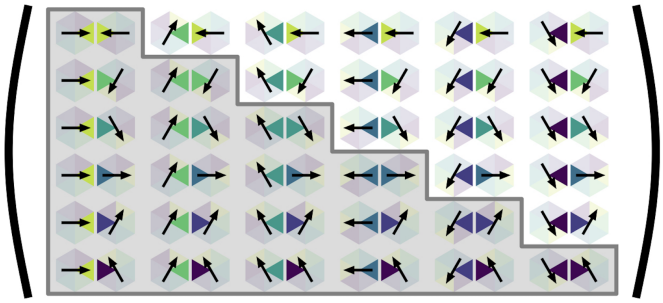


Figure 3: Interaction map for identical particles. Arrows indicate different orientations of each particle. In grey we highlighted all the distinct local structures.

the energies of interaction between two proteins will depend only on the nature of the binding sites that participate in the bond.

We can store all energies that define a system of identical particles in a symmetric 6×6 matrix. We will call this matrix interaction map (Figure 3).

We can use the energies of the single local structures s to construct the total energy of a specific configuration $\mathbf{s}$, which can be calculated as

$$\mathcal{H}(\mathbf{s}) = \sum_s N_s E_s, \tag{1}$$

where E_s are the energies of each local structure and N_s is how many local structures s participate to the configuration $\mathbf{s}$.

We want to generalize this framework to study binary mixtures. In particular we want to use this model to study the mixing mechanism that comes into play when multiple species are present. To achieve the latter, it is interesting to study the model in its dense limit, meaning the limit in which all lattice sites are occupied by a particle. In this way we forbid that particles minimize the energy of the system adapting the global shape of the aggregate. The remaining methods that the system can adopt to relax frustration are two: modify its internal structure by orienting the particles; and mixing, exploiting the combinatorial freedom associated with the presence of multiple components.

We will explain the procedure to extend the model to two different species. The same procedure generalize for any number N of species.

When we have two particle types A and B, we have 12 distinct faces type, a, b, etc. for species A and a', b', etc for species B, we can visualize this by the different face colors of the particles of Figure 4. There are two types of interaction: between particles of the same species ($A-A$ and $B-B$), and between particles of different species ($A-B$). Each of these can be fully parametrized by a 6×6 interaction map storing all possible pairs of faces that two particles can have in contact when sitting on neighbouring sites. In the case of interactions within the same species ($A-A$ and $B-B$) the interaction map is symmetric and only 21 local structures per species correspond to an independent values of the energy. Instead, faces belonging to different species are all distinct, which makes all 36 pairs of faces made by two particles of different species independent local structures. For example, the (a, b) and (b, a) are equivalent if the two particles are identical but (a, b') is different from (b, a').

We can use these 6×6 blocks to construct the interaction map for a binary mixture as a 12×12 matrix where the block on the n^{th} row and m^{th} column describes the interaction between particles of species n and m. In conclusion, to fully specify a binary mixture, we have to define 36 parameters for the interaction map that parametrize the $A-B$ interaction, and $2 \cdot 21$ parameters for the $A-A$ and $B-B$ interactions. In total we need 78 real parameters.

2.2 Using different ensembles answers different questions

The model defined in Section 2.1 can be studied both in the Canonical and in the Grand Canonical ensemble

Left: Two distinct species, A in green and B in purple.
Right: Interaction map for a binary mixture of particles. Each block is of the form of Figure 3. Blue indicates negative interaction energy for a given face pair while red stands for a positive one. The shade is proportional to the energy's magnitude.

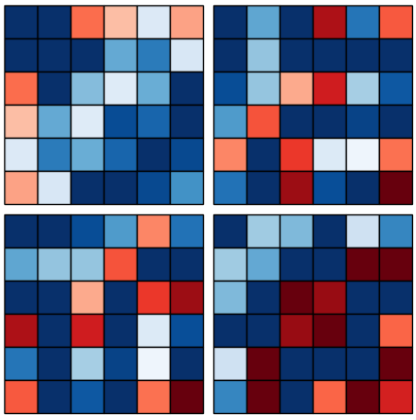


Figure 4: Binary mixture.

The equilibrium configuration of the mixtures in the Canonical ensemble can display phase separation. Indeed in the Canonical ensemble the number of particles of each species is fixed. If the relative density imposed on the two species does not correspond to any stable phase we will observe the formation of two or more phases. Therefore performing simulations in the Canonical ensemble can provide insights on which mixtures phase separate and which don't at different values of the relative densities. There is no strong argument to say a priori which value of the density would be the most insightful. We can suppose that the probability of observing the formation of multiple phases will be higher if we fix the relative density close to one with respect to having one of the two species which is very rare. In this project we only considered mixtures in the Canonical ensemble where the number of particles for each species is the same.

In the Grand Canonical ensemble we will allow particles of one species to mutate into particles of the other species. In this case the relative density of the two particle types is allowed to change in time. The equilibrium configuration in the Grand Canonical ensemble will display only the most stable phase. Grand Canonical results will allow us to understand which mixtures use the combinatorial freedom that arises from having more than one components to ease the frustration caused by incompatible interactions. In this ensemble we can have different values for the chemical potential of two species. Although there is no strong argument that we can use to fix the difference in chemical potential of the two species. Therefore we will start by setting it to zero.

Once we made a choice for the relative density in the Canonical ensemble and for the difference of the chemical potential in the Grand Canonical one, a mixture is completely specified defining the two species that compose it. Therefore in the following mixtures of species A and B will be considered the same system as mixture of B and A.

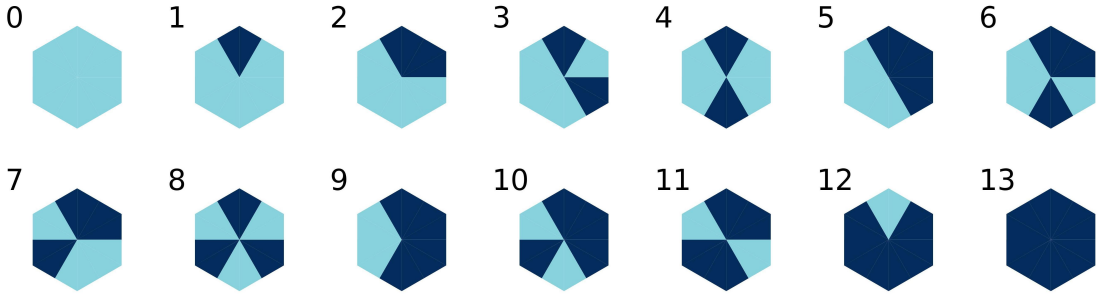


Figure 5: The fourteen distinct hexagonal patchy particles with two patch colors. Dark and light patches are represented as dark and light slices.

2.3 Patchy particles

The space of all possible binary mixtures is $\mathbb{R}^{78}$, one real parameter for each local structure. This space is computationally impossible to explore exhaustively. In this Section we will tackle this problem by focusing on a specific subset of interaction maps: interaction maps associated to patchy particles.

Patchy particles is the name given to a large class of particles characterized by a discrete number of short-range and highly directional interaction sites. Interactions sites are represented as patches attached on the particles' surface. In the general case the patches can be of any given number of colors. Each of them can be thought as a different interaction that the units of a biological system can implement to interact with each other. Each patch can be positioned in any point of the surface of the particle.

Here we decided to study the simplest case, only two possible interaction sites; represented with dark or light patches. We can imagine them as representing hydrophobic and hydrophilic residues in a protein for example.

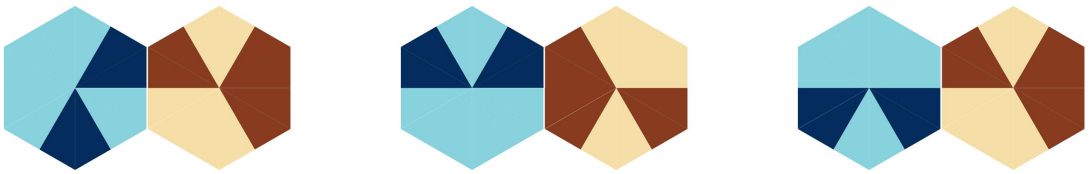
Moreover, we decided to position patches on the vertices of the hexagons. This is only one of the possible choices, one other possibility is to locate the interaction sites on the faces of the particles. Within these constraints all possible hexagonal patchy particles that we can construct are showed in Figure 5.

From now on, we will refer to binary mixtures as (N_1, N_2), where N_1 and N_2 will always refer to Figure 5. We can use patchy particles to build up an interaction map if we define how the patches interact between each other. One of the possible choices is to build the interaction energies such that the patches of the same colors will stick while mismatching ones will repel. If all lattice sites are occupied, the magnitudes of the interactions between patches does not matter to determine the ground state configuration. Instead what matters is only their relative values. The smallest energy will be the favourable interaction and the other will be unfavourable. Therefore we can achieve attraction between matching patches if we define the energy of each local structure as

$$E_s = \alpha n_s \tag{2}$$

where n_s is the number of mismatching patches in the local structure s. We fix $\alpha = 10$. In the dense limit the magnitude of α does not influence the behaviour of the system. The only constraint if we want to have attraction between patches of the same color it that α must be positive.

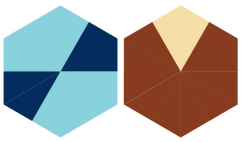
With this definition for the energy of a local structures there are only three possible energies which are showed in Figure 6.

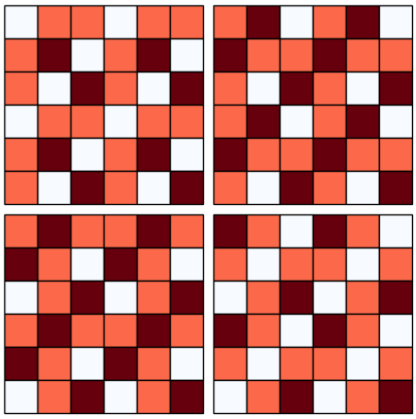


(a) Bond of energy 0 (b) Bond of energy α (c) Bond of energy 2α

Figure 6: Examples of the three different possible energy costs of a bond in the patchy particles framework.

The patchy particles framework makes the problem more tractable reducing the space of all possible binary mixtures. Now the entries of the interaction map are completely determined by the choice of the particle. To fully characterize a mixture, we do not need anymore to choose 78 real parameters. We need to only choose the two particles out the ones of Figure 5. The total number of mixtures that we can define in this way is 196 (14^2). If we consider only mixtures in which the two particles are different we have 182 possible mixtures. Furthermore, for the reasons explained in Section 2.2 we do not need to consider as distinct the mixtures in which the two species are taken in a different order. Therefore we are left with 91 possible mixtures.





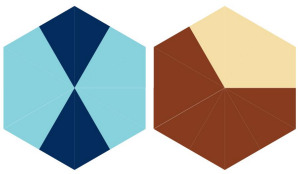
Left: Two patchy particles
Right: Interaction map for a binary mixture of patchy particles. Now only three colors are present. White for 6a, red for 6b, and dark red for 6c.

Figure 7: Binary mixture of patchy particles .

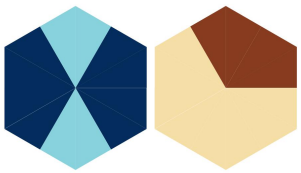
2.4 The symmetries of the system reduce the number of non-equivalent mixtures

The 91 mixtures that we found using the patchy particles framework are not all distinct.

Consider the transformation that exchanges all dark patches of a particle with light ones and viceversa. We will refer to this transformation as color exchange. Now imagine to apply color exchange simultaneusly to both particles of the mixture. An example is showed in Figure 8. This



(a) Mixture (4,9)

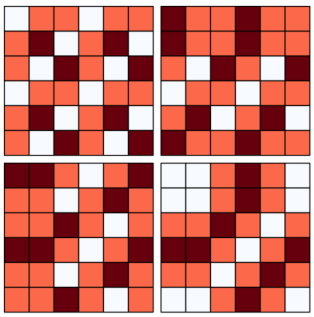


(b) Mixture (11,2)

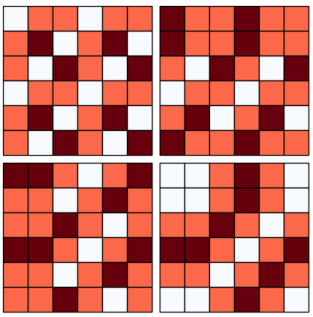
Figure 8: Examples of two equivalent mixtures under color exchange

mapping leaves the interaction map invariant (Figure 9). Therefore two mixtures A and B that are mapped one into the other by color exchange share the same Hamiltonian. This means that for each configuration of A there exists one of B with the same Boltzmann weight which can be obtained color exchanging each particle in the system. Pairs of mixtures that are linked by color exchange will be referred as equivalent for the rest of this Section.

For the consideration made above it is redundant to consider in our simulations both mixtures in a pair of equivalent ones. Therefore we will only simulate one representative mixture for each pair. The total number of non-equivalent mixtures can be calculated using the following considerations. There are 7 mixtures which remain unchanged under color exchange. This can happen for two reasons, both species are invariant under color exchange as in Figure 10a, or color exchange is equivalent to transform the first species into the other and vice-versa as it happens in Figure 10b. The remaining 84 species form 42 pairs of equivalent mixtures. Adding up the 7 invariant to the count we remain with 49 distinct mixtures.

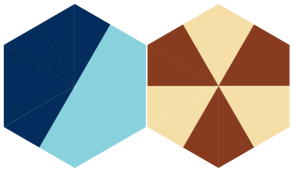


(a) Original interaction map for the (4,9) mixture

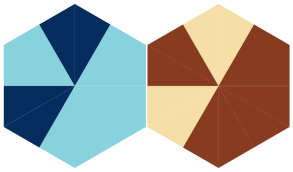


(b) Interaction when the mixture is color exchanged into (11,2)

Figure 9: Symmetry of the interaction map under color exchange



(a) Particles invariant under color exchange



(b) Particles mapped int one another by color exchange

Figure 10: Examples of invariant mixtures under color exchange

3 Optimization of Monte Carlo simulated annealing to find low energy configurations

The number of possible configurations of our system increases exponentially with the system size. To search the equilibrium configuration in this vast space we use the Monte Carlo simulated annealing algorithm. In this Section we will describe the algorithm, and the methods we implemented to optimize its parameters to look for the ground state configurations.

In Section 3.1 we will describe how to implement the algorithm to simulate binary mixtures. Then we describe how we searched for the optimal values of the algorithm parameters; the initial temperature in Section 3.2 and the number of Monte Carlo steps in Section 3.3.

3.1 The algorithm and its implementation

The core idea of the algorithm is to slowly cool down the system from a high temperature T_{start} to a lower one T_{end}. At each temperature, the system updates its configuration for a fixed number of Monte Carlo steps N_{steps}. The temperature affects the transition rates between different configurations which are calculated using the Metropolis rule. Namely at each step a new configuration is proposed and the probability to accept it is

$$min\left[1, e^{-\frac{\Delta E}{T}}\right].$$

With ΔE the difference between the energies of the proposed, and the current configuration.

At high temperatures the algorithm is basically rejection free and the system can explore extensively the space of configurations. At lower temperatures the probability to move from a given configuration to another of higher energy decreases. When the temperature is equal to zero the system cannot increase its energy anymore and falls in the closest local minimum.

The temperatures at which these regimes take place are set by the energy scale that we impose on the system. Indeed the relevant quantity that appears in the transition rates is $-\frac{\Delta E}{T}$. In our model this scale of energy is set by the value of α fixed in Section 2.3.

All the mixtures found in Section 2.3 have been simulated using this algorithm. The simulations have been performed for 24×24 lattice sites systems with periodic boundary conditions. This size allows, while remaining computationally accessible, to accommodate repetitive patterns with a wide range of different periodicities (2, 3, 4, 6, 8, and 12). Allowing the highes number of possible periodicities increases to observe periodic patterns. For example, if a mixture is able to relax all frustration in a pattern of period p, and the size of the system is not a multiple of p, we might not observe the pattern because of the defects caused by this mismatch. We carried out the simulations both in the Canonical and in the Grand Canonical ensemble, with the values of relative density and chemical potential discussed in 2.2.

In the case of the Canonical ensemble, at each step we can either choose a particle at random and change its orientation, or we choose two particles and exchange their lattice positions. At each step one of these two operations is performed with equal probability. In the case of the Grand Canonical ensemble, in addition to these moves, particles can also change type. At each step one of the three moves is executed with probability $\frac{1}{3}$.

It is important to check that each allowed move leaves the Hamiltonian, and as a consequence the whole energy landscape, invariant. This requirement is met if the operations that are performed during the move derive from symmetries of the Hamiltonian itself. This is the case for all the moves described above. Changing particles' orientations reflects the invariance of the interaction map under 6-fold discrete rotations. The exchange of two particles' positions instead is allowed because the interaction map is position independent. Lastly, the possibility to switch type depends on the fact that the energies of the local structures do not specifically depend on the relative densities of particles in the system.

The moves that we chose are delocalized, meaning that the position and orientation of each particle at a given step is completely uncorrelated with the one at the next step. This allows a faster exploration of the phase space. However, this type of move prevents us from inferring anything on the dynamics and on the possible effects of kinetics on the final configuration.

Another thing that we should take into account is that for the system is easier to find the ground state in the Grand Canonical ensemble. This can be under stood in terms of acceptance probability of Monte Carlo moves. More precisely looking at the two moves that change the species of the particle sitting on a given lattice site. In the Grand Canonical case a particle can mutate, meaning that it can change species without affecting the configuration at any other lattice site. In this case the maximum energy difference ΔE between final and initial configuration is 120 (using $\alpha = 10$). Conversely in the Canonical ensemble we have to exchange two particles, meaning that two sites will contribute to ΔE. This doubles the maximum value that the energy difference can take; as a result the acceptance rate is way smaller. To summarize, for this model local changes in the configuration result into higher acceptance rates with respect to non local ones. For this reason we will use Canonical simulations to look for optimal parameters. The optimized values found in this ensemble will also work in the Grand Canonical one.

3.2 Choosing the starting temperature

Very often the free energy landscape on the possible configurations is an extremely complicated function. More precisely, the number of local minima is very large. Finding the global minimum in such a landscape can become a computationally hard task. One necessary requirement that our algorithm must satisfy, is that the final result must be as independent as possible from the initial condition. Therefore, it is very important to be sure that in the first steps of the simulation the system is allowed to reach any configuration, independently form where it starts. If this condition is not satisfied, the simulation trajectory will remain confined in a subset of all the possible configurations and, if the global minimum does not belong to this subset, we will never be able to observe it. At infinite temperature, the transition rates between each pair of configuration are always one, meaning that all configurations are accessible. Practically, we will never achieve infinite temperature, but the idea is to find the value of T_{start} for which the energy difference between states has a negligible effect on the transition rates; in this way we will ensure that the final configuration won't be determined by the location of the initial configuration.

The next two Sections will describe the two methods that we implemented to choose a value of a starting temperature.

In Section 3.2.1 we will use the entropy as a measure of how much the simulation is free to move across all configurations. Then in Section 3.2.2 we will understand why, in the case the system has a crossover at a given temperature, it is important to start at a larger one.

3.2.1 The entropy as a measure of mobility in phase space

At infinite temperature, because of how transition rates between different configurations are calculated, each state is equally accessible. Therefore we expect that the probability to observe a given pair of faces in the system, averaged over an ensemble of configurations, will be the same. In other words, the infinite temperature probability distribution over the possible ordered pair of faces will be uniform. Considering that the total number of ordered pair of faces is 144, one for each entry of the interaction map, each of them will have a probability of $\frac{1}{144}$. Using this information we can calculate the infinite temperature probability distribution over the local structures. Local structures, as we defined them in 2.1, are pairs of unordered faces. Therefore not all local structures will be equiprobable. The local structures that correspond to diagonal entries of the interaction map in Figure 4 will have half the probability of the off diagonal ones, which correspond to two distinct ordered pairs of faces. Then the infinite temperature distribution for the local structures will be

$$p_\infty(s=(a,b)) = \begin{cases} \frac{1}{144} & \text{if } a=b \\ \frac{1}{72} & \text{if } a\neq b \end{cases}. \tag{3}$$

With (a,b) an unordered pair of faces and s a given local structure. Increasing the temperature we expect that the probability distribution over the local structures $p_T(s)$ approaches $p_\infty(s)$ up to the temperature T^* where the two distribution become indistinguishable.

The specific value of the temperature for which this happens is of course pair dependent, although the energy scale of the interactions is the same for every mixtures. As we discussed the transition rate depends only on the ratio between energy and temperature thus we can expect that there won't be large differences in the value of T^* for different mixtures. Therefore, in this Section we will get a qualitative idea of the value of T^* performing our analysis on one specific mixture, (mixture (2,8)). In the next Section we will instead consider all mixtures to confirm the conclusions obtained here.

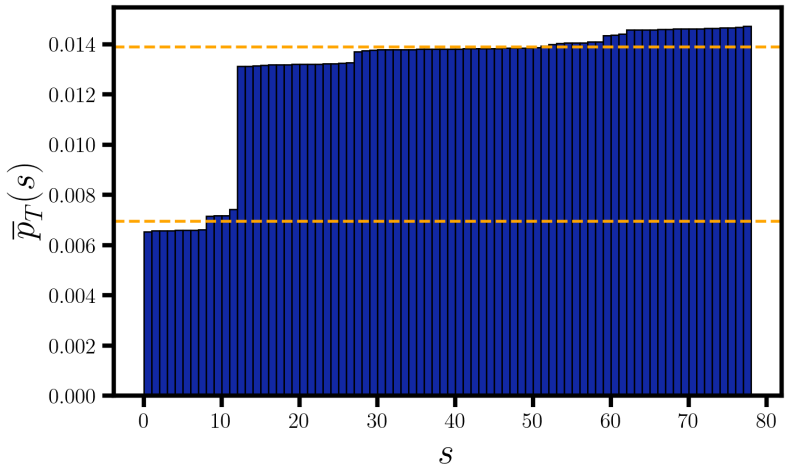


Figure 11: Empirical distribution over local structures at T=200. Bins have been sorted in increasing order. The dashed lines correspond to $\frac{1}{144}$ (the lower) and $\frac{1}{72}$ (the higher)

We performed numerical simulations at a fixed temperature for values of T equal to 10, 20,50,100 and 200. At each temperature we sampled 1000 configurations. Between each sample we performed a number of Monte Carlo steps equal to the size of the system (24×24). At each temperature we calculated the empirical distribution $\overline{p}_T(s)$ over local structures. By visual inspection we confirmed

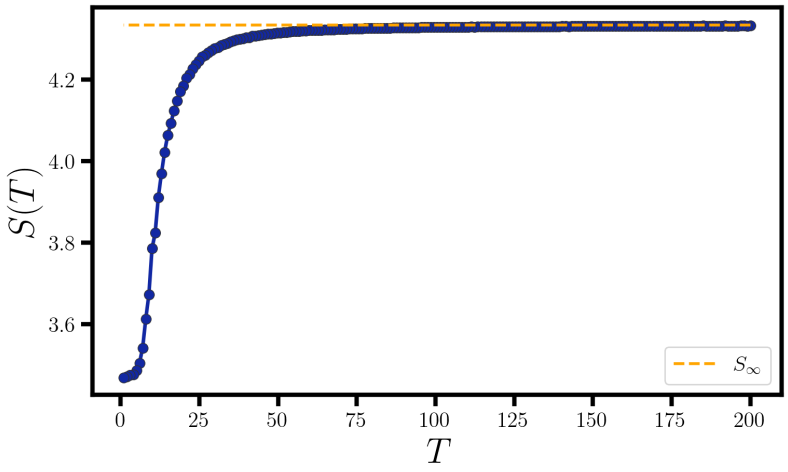


Figure 12: Entropy as a function of temperature. In yellow the value of S_∞.

that $\overline{p}_T(s)$ approaches $p_\infty(s)$ when we increase the temperature. In Figure 11 we show the empirical distribution calculated for T=200.

In order to better visualize the fact that the empirical distribution approaches the theoretical one we decided to use the Shannon's entropy, which given the probability distribution, can be calculated using

$$S(p) = \sum_s p(s) ln[p(s)] \tag{4}$$

For the infinite temperature distribution the entropy is $S_\infty = 4.33$nats. We performed numerical simulations at a fixed temperature for values of T in the range $(1 - 200)$ and at each temperature we calculated the entropy of the empirical probability distribution. The entropy as a function of temperature is plotted in Figure 12.

As we could have expected from Figure 11, the entropy measured for the empirical distribution at T=200 and for the theoretical one is practically the same. This holds down to a much lower temperature. By visual inspection we can suggest that at T=40 the system is still able to move across all configurations of phase space. We are going to test this value for T_{start} in the next section.

3.2.2 Locating the crossover temperature

The discussion made in Section 3.2.1 seems to suggest that increasing the value of T_{start} is always advantageus. However, an excessively large value of the starting temperature has its drawbacks. When the temperature is high, each configuration is equivalent, consequently, the system wanders in phase space without focusing on any specific region. As a result, if the time spent at high temperature is too long, the simulation will spend a lot of computational resources to ineffectively visit high energy regions of the phase space. The goal is to find a temperature high enough to not remain confined, but also small enough to waste as little computations as possible.

We can understand these two regimes as two different states of the system. The high temperature state, in which the thermal energy will allow the system to attain every possible configuration, irrespective of its energy. Here the entropic contribution will dominate and the system will most likely be disordered. Instead, in the low temperature state, the energy of the configuration will replace the entropy as the main contribution of the configuration's free energy. Therefore, at low temperature most configurations will be inaccessible to the system. In the low temperature state most particles will be frozen and we won't expect qualitative changes in the configuration. In between these two states, there will be some cross-over temperature at which the simulation switches from the high-T regime to the low-T one.

These considerations suggest that the optimal value for the starting temperature must be above the crossover, to navigate the full phase space, but not much higher, in order to not waste several

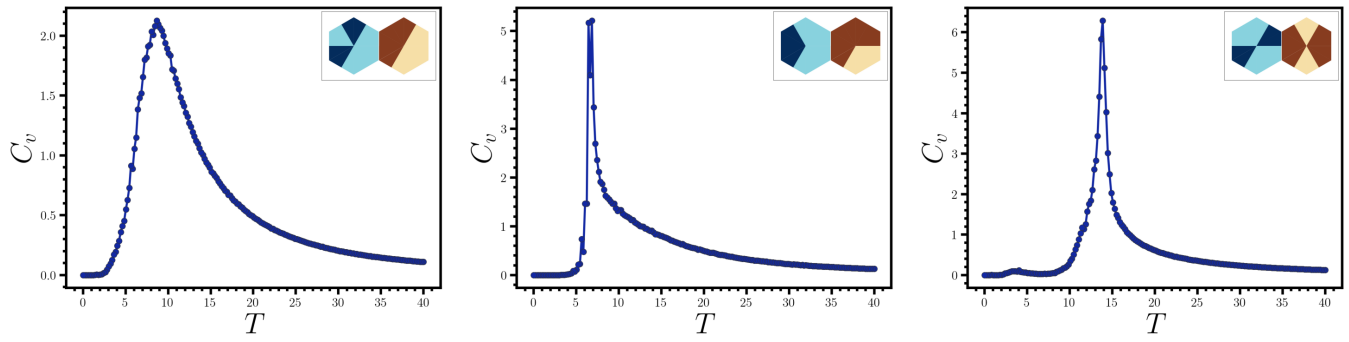


Figure 13: Heat capacity against temperature for mixtures (3,5), (2,9) and (4,11).

computations.

In order to spot the crossover temperature, we calculated the heat capacity using the fluctuation dissipation theorem

$$C_v = \frac{1}{L^2 T^2}(\langle E^2 \rangle - \langle E \rangle^2). \tag{5}$$

Where L^2 is the number of lattice sites.

At the crossover temperature the fluctuations in the energy are known to diverge, therefore we expect the specific heat to have a peak at the crossover.

To confirm this we performed numerical simulations for all 49 mixtures annealing them from $T_{start} = 40$ to $T_{end} = 0$. For each mixture we performed 50 independent simulations. In each of them we calculated the specific heat as a function of temperature using Equation 5. We averaged the curves of the 50 simulations for each of the mixtures and plotted the resulting curve. Three examples are presented in Figure 13.

By visual inspection of the specific heat curves for all mixtures we see that the crossover temperature happens below T=40 for all of them.

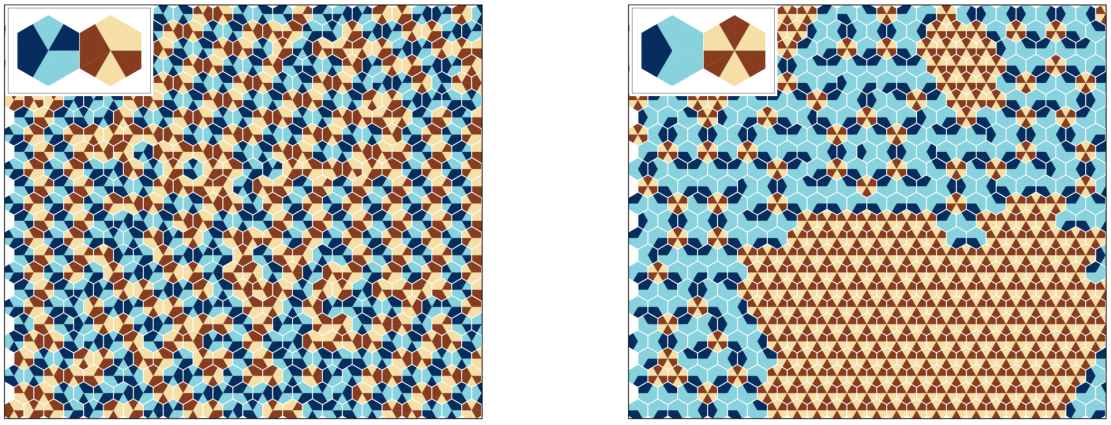
Therefore we decided to use $T_{start} = 40$ for all the simulations.

3.3 The number of Monte Carlo steps

In this section we will discuss how, fixed the starting temperature, we can still improve the algorithm's ability to look for equilibrium equilibrium configurations. In particular in this section we are going to optimize over the number of Monte Carlo steps that the simulation performs at each temperature, and on the number of temperatures steps performed to cool down the system (N_T). As in the case of the starting temperature we need to balance performances and computational costs. Having enough steps at every temperature is essential to equilibrate the system at each of them. If we cool down too fast, the simulation will not have the time to explore all the configurations accessible to each single temperature. On the other hand spending too many steps at a given temperature is just a waste of computational power. The same reasoning holds for the number of temperature steps.

The optimal number of steps for each temperature will depend on the system' size, for this reason we will define 1 Monte Carlo update as a number of Monte Carlo steps equal to the system size. We will look for an optimal value of $N_{updates}$.

The optimal value for $N_{updates}$ will also depend on the mixture. To understand which mixtures requires more updates we simulated one system for all 49 mixtures using $N_{updates} = 1000$ and $N_T = 100$. These values have been chosen to be able to perform all simulations in a reasonable computational time. Looking at the snapshots of the final configurations that the algorithm found for all different mixtures we can make the following considerations. Some mixtures have reached their ground state, for example the mixtures (0,13), (0,8) or (0,1). In these cases the single species could tile the plane with zero energy, therefore the ground state, which is the situation that we observe in the snapshots, is a complete phase separation.



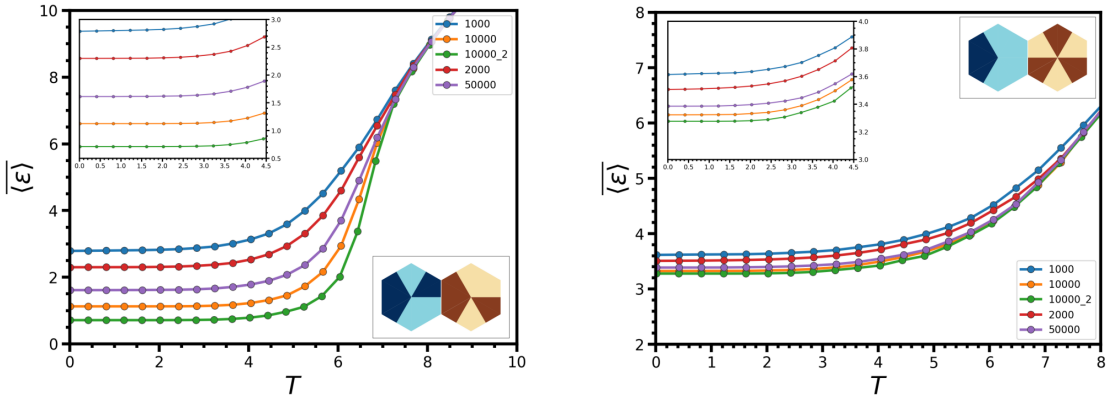
(a) (6,7) (b) (2,8)

Figure 14: Configuration for two representative mixtures obtained with $N_{updates} = 1000$ and $N_T = 100$.

When the species composing the mixture become less symmetric, it is not intuitive what the ground state will look like. There are two different hint that we can use, when looking at a snapshot, to suppose that the ground state configuration has not been reached. These two situations are exemplified in Figure 14.

In the mixture of Figure 14a, it seems that the system could tile the plane with a repetitive ordered pattern. Although, because of a too small $N_{updates}$ we observe many distinct domains separated by defects. Maybe increasing the annealing time we could observe a perfectly ordered structure.

In the case of Figure 14b, we observe phase separation. There are two signs that the simulation might have been carried out too fast. First, there are multiple clumps of the brown phase; what we expect is to see one single cluster. Second, the blue phase seems to be rich of frustrated particles; namely, particles that cannot match all their patches.



(a) (6,7) (b) (2,8)

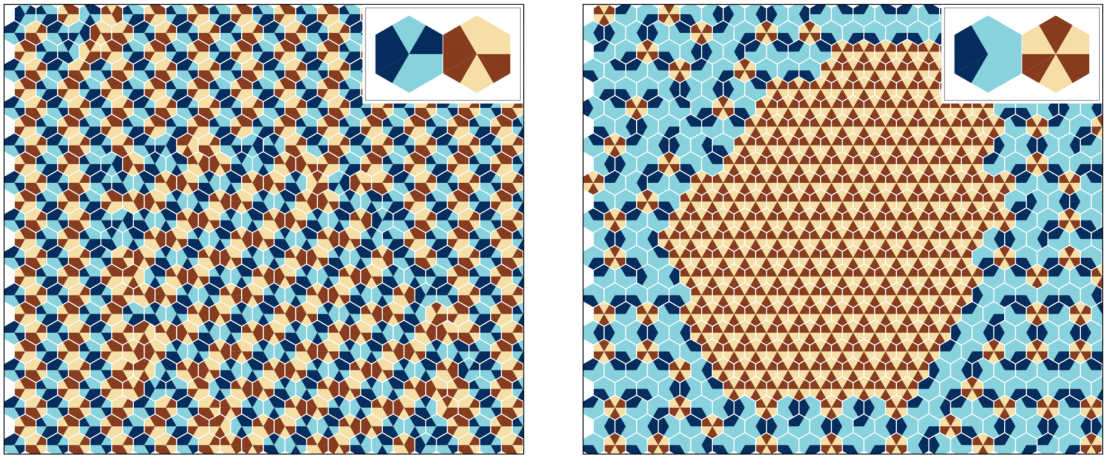
Figure 15: Average energy per particle during the annealing for mixtures (6,7) and (2,8). The overline indicates that the curves are obtained averaging over the ensemble of simulations. All simulations use $N_T = 100$ apart from the green curves which have $N_T = 200$.

For frustrated systems, where the degeneracy of the energy levels becomes large, it is hard to find the exact global minimum because of the several local ones.

We are going to use these mixtures to optimize the value of $N_{updates}$. To quantitatively evaluate the performance of the algorithm, we used the average energy per particle $\langle \varepsilon \rangle$ of the final configuration. For both mixtures, we simulated an ensemble of 50 systems, and for each of them,

we kept track of $\langle\varepsilon\rangle(T)$. The results for different values of $N_{updates}$ are showed in Figure 15.

The curves show that the mixture $(6,7)$ consistently reduces the final energy per particle when the value of $N_{updates}$ is increased; furthermore, also doubling the value of N_T seems to improve the algorithm performances. On the other hand, for mixture $(2,8)$, the value of $\langle\varepsilon\rangle$ at zero temperature seems to scarcely improve when $N_{updates}$ increases above 5000. To have an idea of how the final configurations changed, we can again look at a snapshot obtained with the largest values of the parameters: $N_{updates} = 10000$ and $N_T = 200$ (Figure 16).



(a) (6,7) (b) (2,8)

Figure 16: Representative final configurations obtained for the test mixtures using 200 temperatures and $N_{updates} = 10000$.

In opposition to what we observed in Figure 14 mixture $(6,7)$ creates two large crystalline domains and $(2,8)$ forms a single cluster of brown particles.

The formation of crystalline domains in mixture $(6,7)$ is related to the existence of three equivalent ground states with different orientations of the "stripes". If at the crossover, the system starts nucleating domains with different orientations in different locations, it will be hard for the algorithm to rotate all of them in the same orientation. This difficulty could be overcome if instead of using single particle moves, one introduces cluster moves, namely Monte Carlo moves that change the position of many particles at the same time.

At this stage, the mixed phase observed in $(2,8)$ does not seem to order. The reason might be related to finite size effects. Indeed, once the brown cluster is formed, it imposes some specific boundary conditions on the blue phase. These conditions might not match the periodicity of the presumed ordered phase, which consequently will remain disordered. One idea could be to use algorithms such as the Gibbs Ensemble algorithm, which are specifically tuned to study phase separation.

The snapshots of Figure 16 might not be the ground states of the mixtures. However they clarify what is the behaviour with respect to mixing for of both of them. Mixture $(6,7)$ tends to mix in an ordered pattern, instead mixture $(2,8)$ clearly wants to phase separate and we don't expect an abrupt change in the behaviour in the ground state.

Therefore we can consider the values $N_{updates} = 1000$ and $N_T = 200$ good enough for the purposes of this work. We will use these values in all the following discussion.

4 The Phenomenology of the Patchy Particle Model

We simulated all the different mixtures in both the Canonical and Grand Canonical ensemble. We simulated 50 independent realizations for each mixture to have a sufficient number of systems. In this way we try to eliminate the effects of configuration-to-configuration fluctuations. In Figures 17 and 18, we present one representative snapshot for each mixture.

This Section will be dedicated to comment all the different scenarios observed in the simulation results. In Section 4.1, we will address the results obtained in the Canonical ensemble. Instead in Section 4.2 we will discuss the Grand Canonical ones.

From now on, in the whole Section 4, we are going to refer to the mixtures of Figure 17 as (C,N_1,N_2) and of Figure 18 as (G,N_1,N_2) where C and G stand for Canonical and Grand Canonical, and N_1 and N_2 are row and column numbers, respectively. The mixture (C/G,0,0) will then be the top left corner.

4.1 Results in the Canonical Ensemble

In the Canonical ensemble, some mixtures tend to mix in a uniform phase while others form two distinct phases. We are going to refer to these two stereotypical situations as "mixed" and "phase separated" systems.

Phase separated systems can be classified into two broad categories. First, the ones that separate into two pure phases, meaning phases composed of only one species. For example, the mixtures (C,0,1) and (C,2,3). Second, the ones that form mixed phases, where both species participate. This is the case of (C,1,1) and (C,1,4). These mixtures seem to suggest that the reason behind phase separation is the fact that the proportion of the two species in the mixed phase is not the same as the one imposed on the system, in this case 1:1.

One clear example of this is system (C,3,6). In this case, the mixed phase is ordered, thus we can calculate the relative proportion of particles using the unit cell of the pattern. Out of four particles, only one must be brown to perfectly tile the plane with the mixed phase; therefore, the remaining ones end up forming a second pure brown phase. Another example is (C,4,3). In this case, the particles in excess (particle 4) do not fit well together as particles 8 did in (C,3,6), therefore they do not come together in a single clump. Instead, the superfluous particles become "star-shaped" defects of the crystalline mixed phase. What we can expect from the Grand Canonical simulations of this system is to find an ordered phase with a value of the relative density different from 1.

If a given system admits a stable mixed phase where both particles are equally present, then we can observe it in the Canonical ensemble. This is the case, for example, for (C,2,4) or for (C,0,5). The former presents a crystalline ground state. For the same reasons mentioned in Section 3.2 for the mixture (6,7), there are multiple crystalline domains corresponding to equivalent equilibrium configurations. Conversely, (C,0,5) does not display any sign of ordering, and multiple particles remain frustrated.

4.2 Results in the Grand Canonical Ensemble

In the Grand Canonical ensemble, the relative density of the two species is not fixed. For this reason, the equilibrium configurations will display only one phase of the system, more precisely the lowest energy one. Many mixtures end up being in a single pure phase. This is often the case when one of the two particles can tile the plane without any mismatching patch; as for example particle 0 or particle 8. Among the mixtures whose equilibrium state comprises only one species, (G,3,6) is particularly interesting. If we look at its canonical counterpart we see that both phases, the pure and the mixed, could tile the plane with zero total energy. Although in all Grand Canonical simulations of the mixture (5,8) we only observe the pure brown phase. This behavior can be understood by the following entropic argument. Particle 8 has only two distinct configurations, each of them corresponding to three different orientations. Instead, for particle 5 all six orientations correspond to distinct state. This results in a higher degeneracy of the pure brown phase with respect to the mixed one. Consequently, it becomes improbable to observe the mixed phase for (G,3,6) even if it is energetically equivalent to the pure one.

Beyond pure phases, we observe that a non-negligible subset of systems mixes at equilibrium. We can distinguish two categories of mixed phases. Crystals, in which the particles tile the whole plane with a periodic pattern; and amorphous phases where there is no evidence of ordering. The results for mixture (G,4,3) confirm the hypothesis we made using its Canonical equilibrium state. The system periodically tiles the lattice with an ordered phase where the relative densities of blue and brown particles are 2:1 relaxing all defects.

Crystalline phases do not present multiple domains in the Grand Canonical simulations; in contrast with what happens instead for their Canonical equivalents. This supports the suppositions that we made in Section 3.1.

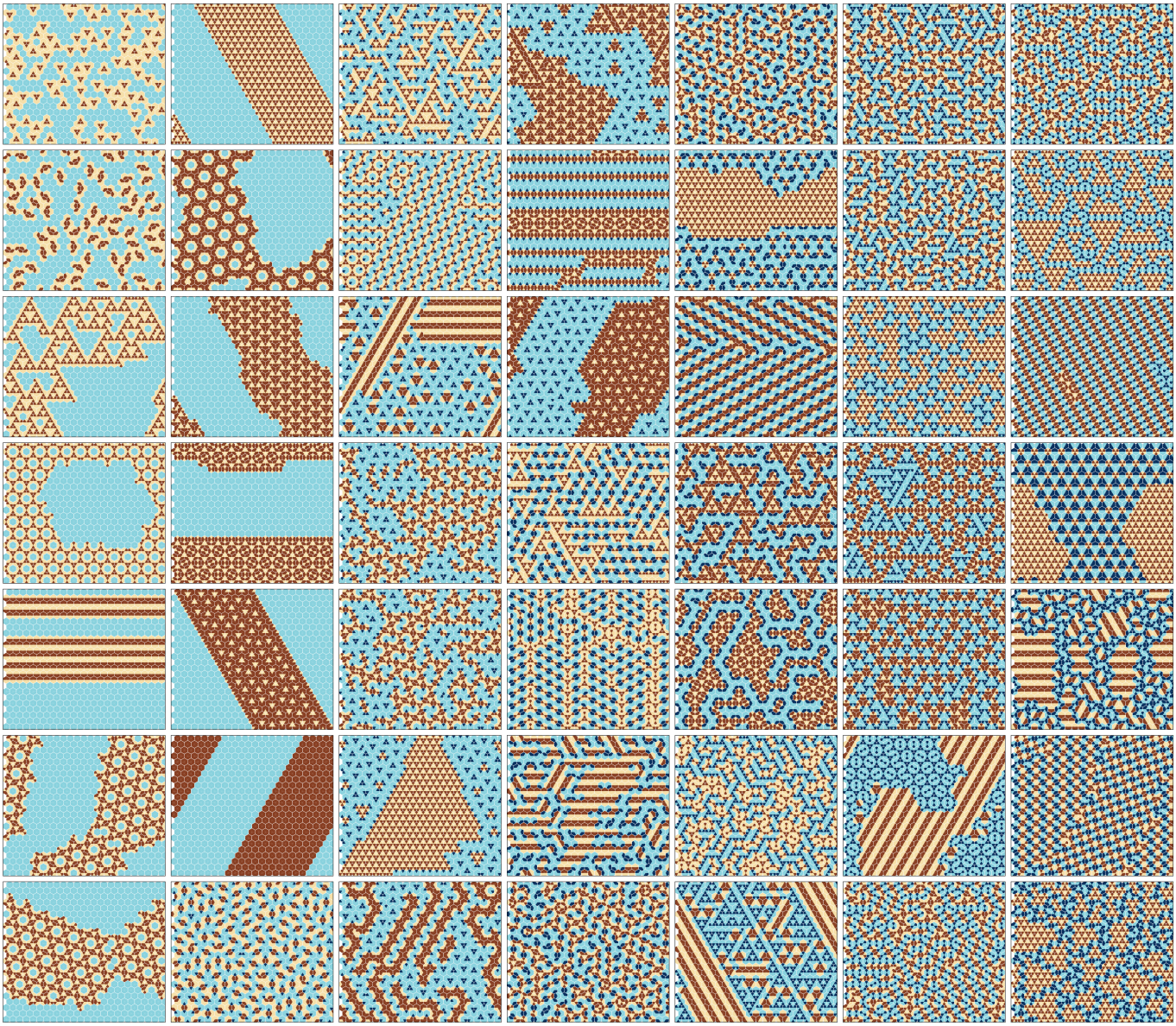


Figure 17: Representative configuration for each mixture after the annealing in the Canonical ensemble

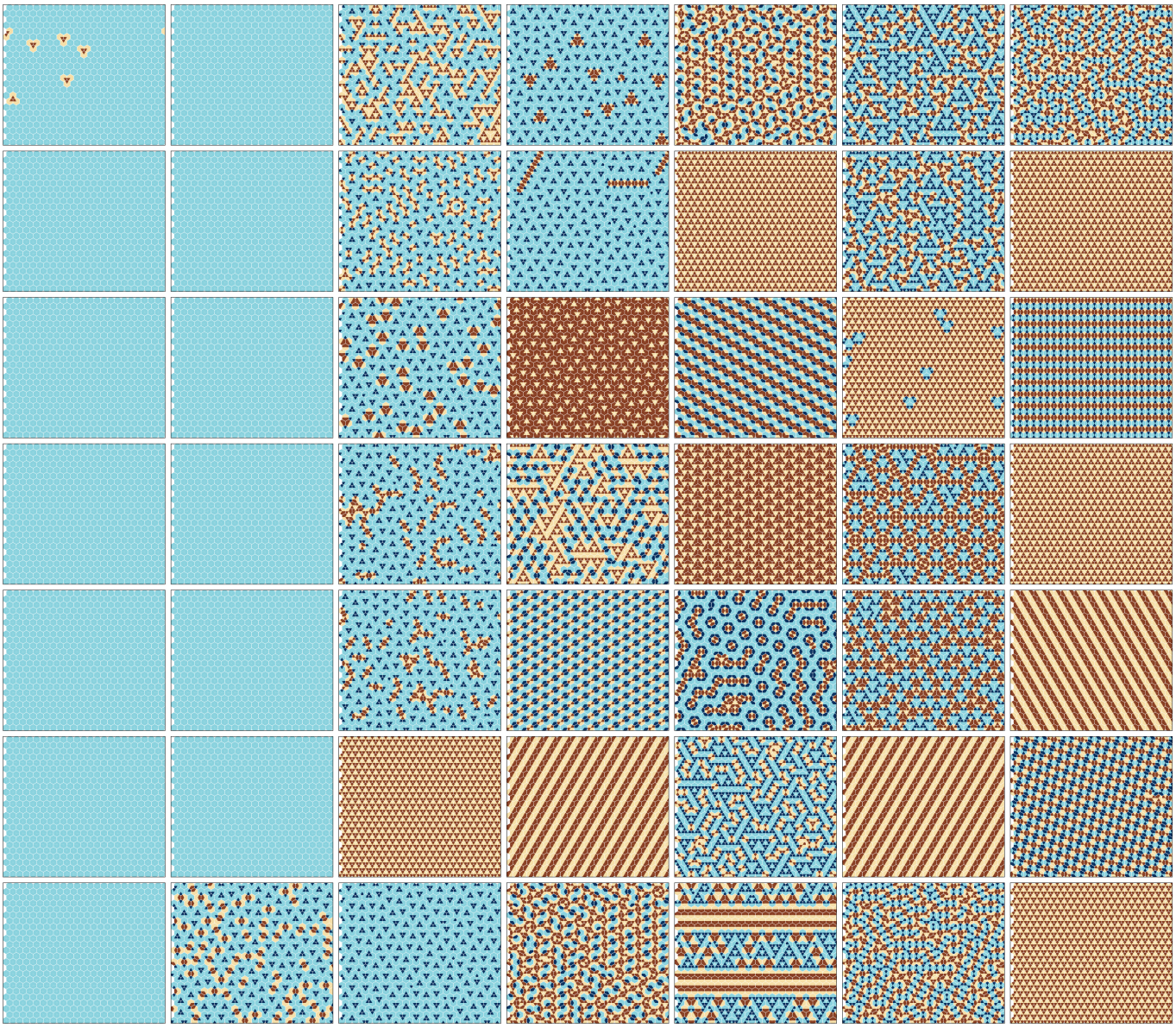


Figure 18: Representative configuration for each mixture after the annealing in the Grand Canonical ensemble

5 Classification of different phases

The discussion made in Section 4 showed how the different equilibrium configurations can form two qualitatively different phases; pure and mixed. Furthermore, we observed two distinct behaviors of mixing: crystalline patterns and amorphous ones.

The main goal of this section will be to make a quantitative classification of the mixtures based on their equilibrium phases obtained in the Grand Canonical ensemble. In Section 5.1 we distinguish between mixed and pure phases. After that, in Section 5.2 we find a scalar parameter to discern between amorphous and crystalline configurations.

5.1 Pure or Mixed?

In pure phases, one of the two species has completely disappeared from the lattice. On the contrary, in mixed phases, both particles participate in the equilibrium configuration. Therefore, one natural

parameter that allows us to distinguish between these two scenarios is the density of the sparsest particle ρ_s. The strength of this indicator is that it does not distinguish which species disappears, thus removing the symmetry between pure brown phases and pure blue phases.

The classification of each system has been done using the ensemble average of ρ_s over fifty independent simulations. One phase is classified as pure if $\overline{\rho_s} \leq 0.01$. We set the threshold just above zero to take care of rare fluctuations which could cause a misclassification of some mixtures.

In Figure 19, we can observe all the systems ordered by their value of $\overline{\rho_s}$.

Pure and mixed phases are correctly distinguished. This can be understood by looking at the different snapshots in Figure 19. On the left, for values of $\overline{\rho_s}$ close to zero, we have phases composed by only one species, which we classify as pure. On the right instead, we have the phases in which both species are present.

Among mixed phases, there seems to be a lack of a preferred value for the equilibrium density and the value of $\overline{\rho_s}$ spans uniformly the range form 0.01 to 0.5.

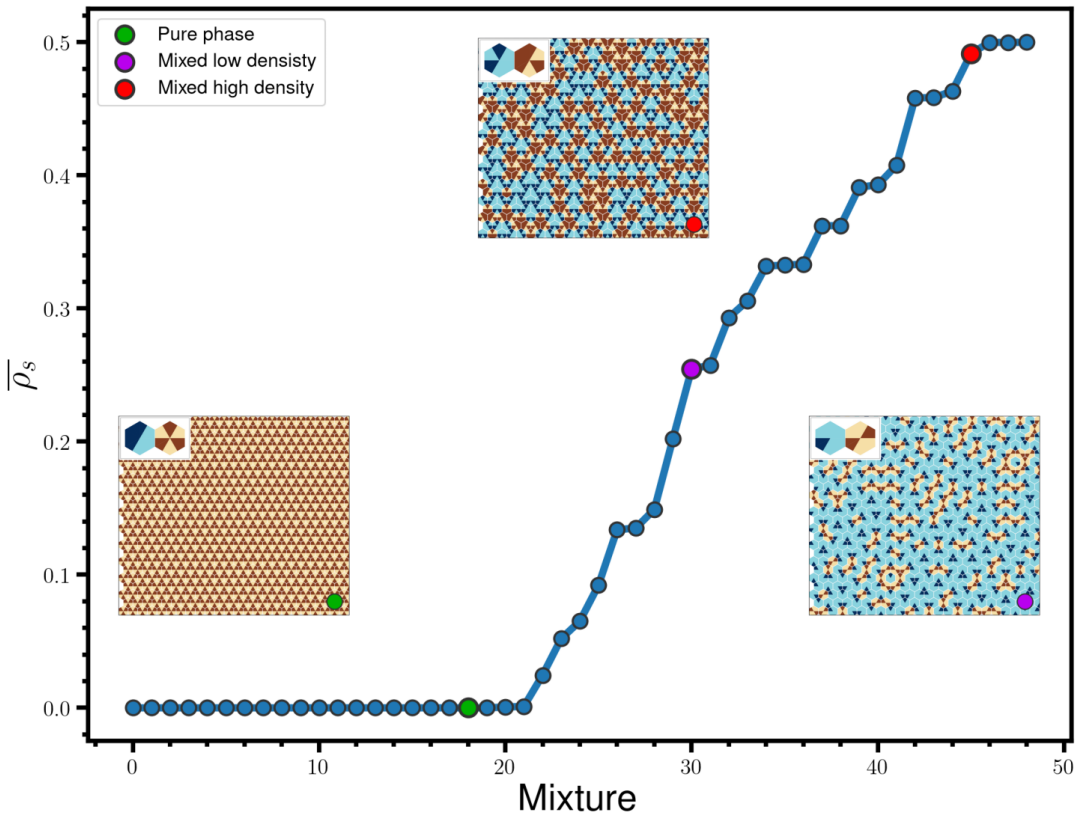


Figure 19: Density of the sparsest particle in increasing order for all the 49 mixtures. Overline means average over all the equilibrium configurations obtained for each system in the Grand Canonical ensemble.

5.2 Crystalline or Amorphous?

This section will be dedicated to the search for a parameter able to discriminate between crystalline and amorphous phases. Crystals present a high degree of ordering and the pattern that tiles the plane is periodic. Instead, amorphous states seem to lack of any degree of correlation between local environments in different locations of the lattice. We can quantitatively rephrase these two statements using the real space correlation function. To this purpose, it is useful to define

$$\Theta_1(i) = \begin{cases} 1 & \text{if site } i \text{ is occupied by species one} \\ 0 & \text{otherwise} \end{cases} . \tag{6}$$

Then the real space correlation function for a given mixture is defined as

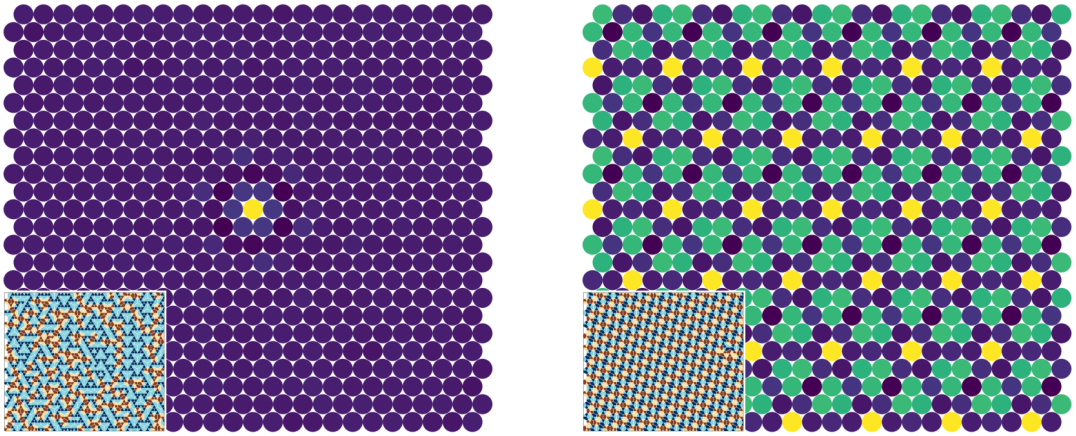
$$G(\mathbf{r}_i - \mathbf{r}_j) = \langle \Theta_1(i)\Theta_1(j)\rangle. \tag{7}$$

However, the system has discrete translational invariance, thus Equation 7 can be simplified to $G(\mathbf{r}) = \langle \Theta_1(0)\Theta_1(r)\rangle$.

This function expresses the probability of finding a particle of species one in position $\mathbf{r}$ given that a particle of the same species is in the origin. $G(\mathbf{r})$ is the same for both species up to a constant shift. Indeed, we have that $\Theta_2(i) = 1 - \Theta_1(i)$ and

$$\langle \Theta_2(i)\Theta_2(j)\rangle = 1 - \langle \Theta_1(i)\rangle - \langle \Theta_1(j)\rangle + \langle \Theta_1(i)\Theta_1(j)\rangle = \langle \Theta_1(i)\Theta_1(j)\rangle + \text{const.} \tag{8}$$

Therefore it makes no difference to talk about the correlation function of one species or of the other. For this reason we can refer to $G(\mathbf{r})$ as the real space correlation function of the mixture. In Figure 20 we plot $G(\mathbf{r})$ for two mixtures, (3,7) and (6,7), for which the equilibrium configuration in the Grand Canonical ensemble are an amorphous phase and a crystalline one respectively.



(a) (3,7) (b) (6,7)

Figure 20: Real space correlation function $G(\mathbf{r})$ for an amorphous and a crystalline phase. The origin is in the center of the lattice. Bright sites correspond to higher probability compared to the darker ones. The color coding is normalized with respect to the highest peak.

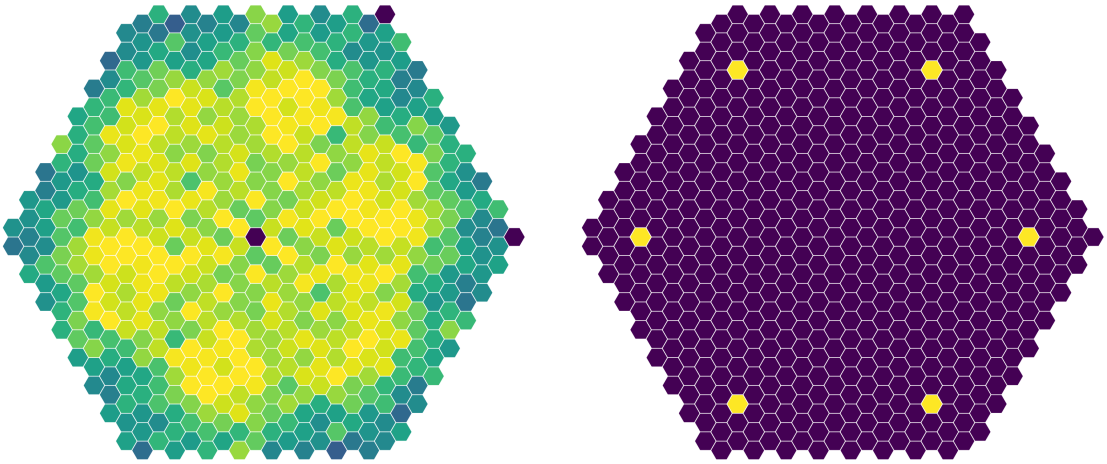
Figure 7 shows that in amorphous phases the correlation function decays as the distance from the origin increases. Instead for crystalline phases, even far away from the origin there are some sites which the probability of having the same species as the one sitting in the origin is still high (bright spots). This comes from the fact that correlation function is calculated over the 50 configurations. Therefore in amorphous states, where there is not a well defined pattern the fact that a particle sits in the origin does not increase the probability that, in a random equilibrium configuration, one of the same species sits on any other specific lattice site. Instead when there is crystalline order, there are a finite number of crystalline configurations that repeat. These are for example the 3 equivalent ground states for mixture (6,7). In those specific configuration each particle has well defined position. Therefore, in a random equilibrium configuration, some sites have a higher probability to have the same species as the one sitting in the origin.

Now we want to use the information contained in the correlations to quantitatively distinguish between the two phases. Employing $G(\mathbf{r})$ for this purpose turned out to be challenging. Instead, we found that correlations in reciprocal space are the most suitable for our goal. The equivalent of the function $G(\mathbf{r})$ in reciprocal space is the structure factor

$$S(\mathbf{q}) = \langle \widehat{\Theta}(\mathbf{q})\widehat{\Theta}(\mathbf{q})^*\rangle \tag{9}$$

where $\hat{}$ denotes the Fourier transform and * complex conjugation.

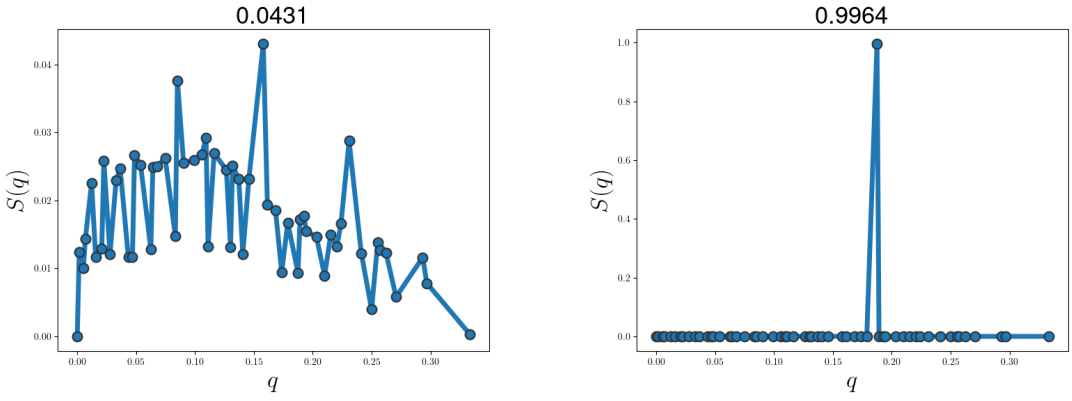
The structure factor takes values on the wave vectors $\mathbf{q}$ of the triangular lattice's Brillouin zone. Its peaks correspond to the characteristic wave vectors of the pattern that tiles the lattice. Therefore, we expect that crystalline phases will display well-defined peaks for specific values of $\mathbf{q}$, the so-called Bragg peaks; instead, amorphous phases will show a much broader range of wave vectors with a non zero value of $S(\mathbf{q})$. In Figure 21, the structure factors are presented for the same mixtures as in Figure 20.



(a) (3,7) (b) (6,7)

Figure 21: Structure factor $S(\mathbf{q})$ for an amorphous and a crystalline phase. Color coding is the same as employed in Figure 20. The color coding is normalized with respect to the highest peak. We set the $\mathbf{q} = 0$ mode to 0.

We want now to extract from this correlation function, a scalar parameter, that is able to distinguish between crystalline and amorphous configurations. We will first rewrite $S(\mathbf{q})$ as a function q, the absolute value of $\mathbf{q}$. This can be simply done by summing over all the values of $\mathbf{q}$ which are at the same distance from the origin. Then in order to compare structure factors of different mixtures we will normalize the function $S(q)$ such as its integral is equal to one.



(a) (3,7) (b) (6,7)

Figure 22: Structure factor $S(q)$ for an amorphous and a crystalline phase. The title of the figure is the height of the tallest peak.

In crystalline phases, $S(q)$ will be sharply peaked on its Bragg peaks. Instead, in amorphous states, the structure factor will be non-zero over a wider range of wavevectors. Also in amorphous

states, different degrees of ordering are reflected in their structure factors; the more the structure factor is peaked on some specific values of $\mathbf{q}$ the more ordered we expect the configuration to be. Ordering is not the only aspect we can perceive in the structure factor. For example, chiral $S(\mathbf{q})$ corresponds to chiral phases, meaning phases in which one of the two species exhibits a defined chirality (species 6 and 7). Additionally, peaks at small wavevectors indicate the formation of single species' clusters.

In crystal phases, the mass of the function $S(q)$ will be all concentrated on the Bragg peaks, whose height approaches 1; instead in amorphous phases it will shared between different values of q. Therefore we can expect that the maximum of $S(q)$ in amorphous states is much smaller than one. These considerations are supported by the curves in Figure 22. We then try to use the height of the tallest peak in the structure factor as an indicator to discern between crystals and amorphous phases. In Figure 23, we plotted the cumulative distribution of mixtures against the height of the tallest peak. In other words, the curve in Figure 23 indicates, for any given height, what fraction of systems has a tallest peak shorter than that.

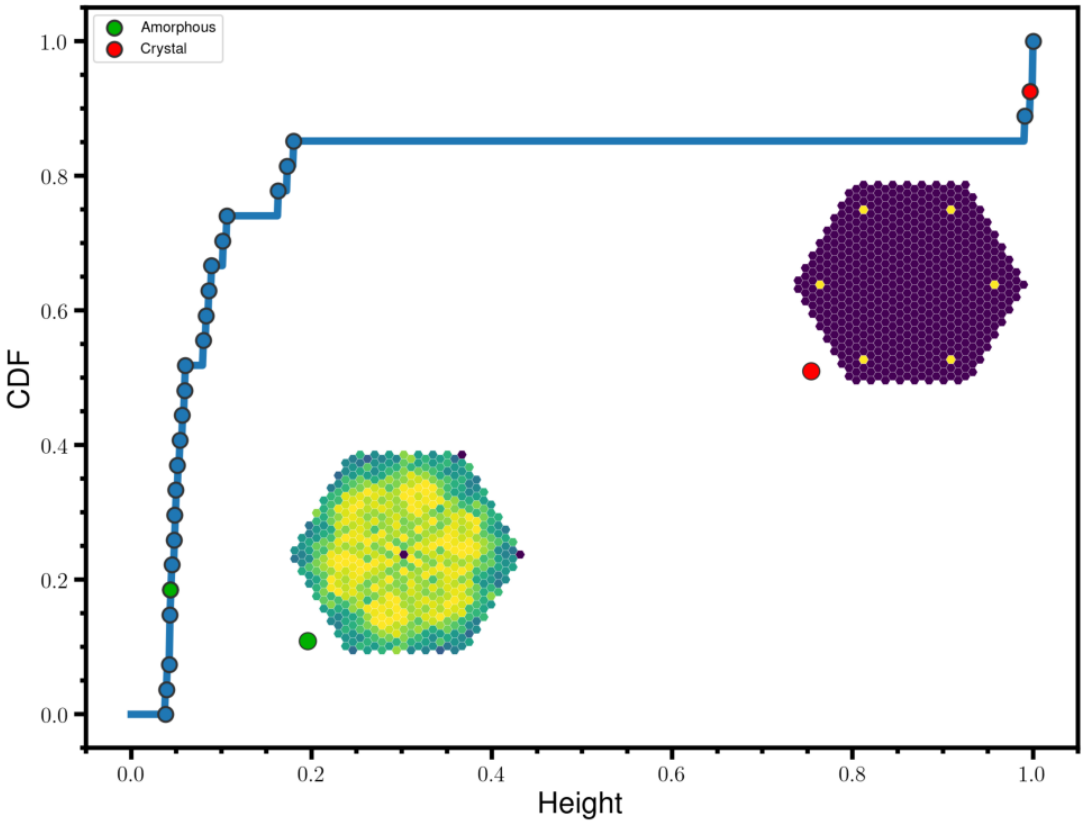


Figure 23: Height of the tallest peak in the structure factor $S(|\mathbf{q}|)$.

We can clearly distinguish between two groups of mixtures. On the right the crystalline phases, whose maximum is almost the only nonzero value of $S(q)$. On the left instead the amorphous phases, where the support of the structure factor spans a wider range. The maximum of the structure factor can be successfully used as a classification parameter to distinguish between amorphous and crystalline phases.

6 Discussion

In the Introduction, we stated that mixtures of multiple species could ease frustration by exploiting the combinatorial freedom that derives form having various component. In this work we tried to understand this mechanism in the framework of binary mixtures of patchy particles. Simulating the system in the Grand Canonical ensemble we were able to distinguish three types of ground state configurations (Figure 18).

Some systems were able to fully relax frustration, finding a zero energy ground state, using only one species. We named pure phases the ground states of those systems. Some examples are the ground states formed by particle 0 and particle 8. In these cases one of the two species participating to the mixture is able to tile the plane in a configuration of zero energy.

In other cases instead, we found that the ground state configuration is a co-participation of both species. We were able to use the information contained in the correlation functions to distinguish two classes of these mixed phases; crystalline and amorphous (Section 5.2).

In the crystalline phase, the two species cooperate to form a periodic pattern. Interestingly, particles forming this kind of phases cannot tile the plane in an unfrustrated ground state when taken independently. Instead when taken in pairs of particles whose interactions are somehow compatible they can relax all the frustration and find a zero energy ground state. One further consideration that we can make on these phases is that almost all crystalline phases, with exception of mixture (2,4), are mixtures of two species which transform into each other by color exchange (Figure 8). This information might be relevant to further investigate what makes the interaction between two sub units compatible, and therefore able to optimally relax frustration.

The last category of phases, are the amorphous ones. Conversely to what happens for crystalline phases, in these cases the frustration is not fully relaxed and some particles remain with unsatisfied interactions. These phases suggests that, for those mixtures, the combinatorial freedom gained by adding one particle type, is not enough to fully balance the complexity of the two species. We might expect is that some of them, increasing the number of species in the mixture will be able to fully relax frustration.

In future works, we will investigate further these three scenarios to understand which characteristics make a species more or less complex. These question can be addressed by studying mixtures of a larger number of species. An interesting question could be, for given species, what is the minimum number of particles that we need to add to completely relax frustration. This question could be helpful to understand what makes some particles more complex than others.

Computationally speaking, simulating mixtures with multiple particles will become more and more expensive. Having some analytical framework would be of great help to at least make some educated choices about what species might be interesting among all the ones present in the vast space of all possible mixtures. At the moment, in our group, we are trying to develop a mean field theory of the model that we presented. The simulations carried out in this project will serve as a validation set for the predictions made with such model.

References

[1] Héctor García Seisdedos, Charly Empereur-mot, Nadav Elad, and Emmanuel Levy. Proteins evolve on the edge of supramolecular self-assembly. *Nature*, 548:244, 08 2017.

[2] Lara Koehler. *Principles of self-assembly for particles with simple geometries and complex interactions*. Thesis, Université Paris-Saclay, Jul 2023.

[3] Na Liu and Tim Liedl. Dna-assembled advanced plasmonic architectures, 05 2021.

[4] Roma Rambaran and Louise Serpell. Amyloid fibrils abnormal protein assembly. *Prion*, 2:112–117, 07 2008.

[5] Pierre Ronceray and Bruno Le Floch. Range of geometrical frustration in lattice spin models. *Phys. Rev. E*, 100:052150, Nov 2019.

[6] Janice L. White, Marvin L. Hackert, Manfred Buehner, Margaret J. Adams, Geoffrey C. Ford, Paul J. Lentz, Ira E. Smiley, Steven J. Steindel, and Michael G. Rossmann. A comparison of the structures of apo dogfish m4 lactate dehydrogenase and its ternary complexes. *Journal of Molecular Biology*, 102(4):759–779, 1976.