TECHNISCHE
UNIVERSITÄT
WIEN
Vienna University of Technology

DIPLOMARBEIT

Machine learning assisted exploration of patchy triangle systems

zur Erlangung des akademischen Grades

Diplom-Ingenieur

im Rahmen des Studiums

Masterstudium Technische Physik

eingereicht von

Felix Müller
Matrikelnummer 01426879

ausgeführt am Institut für Theoretische Physik
der Fakultät für Physik der Technischen Universität Wien

Betreuung
Betreuer/in: Assistant Prof. Emanuela Bianchi
Mitwirkung: Dr. Carina Karner

Wien, 18.09.2024

(Unterschrift Verfasser/in) (Unterschrift Betreuer/in)

Abstract

The field of two-dimensional patchy particles opens up research for a wide range of technical applications such as, but not limited to, membranes and surface layer applications. These patchy particles can be engineered at different scales using DNA folding, manufacturing of nano-particles or serve as a model system for molecules with similar geometrical shapes or bonding behavior. The aim of this work was to gain more understanding of the self-assembly behavior of triangular patchy particles for different particle and system parameters. To achieve this, a Monte Carlo simulation was performed to study the self-assembly of different types patchy triangles in the canonical ensemble. Visual examination of snapshots allows for a generation of state diagrams that serve as a reference for the automated classification on the basis of order parameters. Among the observed structures are chains, gels, porous, and close packed crystals which require the use of a set of different order parameters for a proper classification. The results of the self-assembly depend on both system and particle parameters, such as patch arrangements on the particle edge, referred to as patch topologies. A phenomena occurring when trying to identify clusters in high dimensional spaces is the so called Curse of Dimensionality, which can create different issues when the number of data points is too small for the data's dimensions. To deal with this issue, the space spanned by multiple order parameters was reduced to two dimensions using the conventional Principal Component Analysis (PCA). Further utilization of state-of-the-art machine learning methods suggested the use of Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN) to classify the now two-dimensional dataset into different groups of similar systems. This clustering method successfully categorized the systems into different observable phases, without a visual analysis of all singular snapshots. These results contribute to a refined understanding of self-organization processes for patchy triangles.

Contents

1 Introduction **5**
- 1.1 From single particles to clusters . . . 5
  - 1.1.1 Self-assembly and phase transitions . . . 5
  - 1.1.2 Self-assembly for materials design . . . 6
  - 1.1.3 Patchy Particles . . . 7
  - 1.1.4 Patchy triangles and their applications . . . 7
  - 1.1.5 Self-assembly simulations . . . 8
- 1.2 Gaining a quantitative understanding of large particle assemblies . . . 9
  - 1.2.1 Order Parameters . . . 9
  - 1.2.2 Automated classification . . . 10
  - 1.2.3 Data comprehension for large data sets . . . 11

2 Simulation **12**
- 2.1 Monte Carlo Method . . . 12
- 2.2 Particle Parameters . . . 14
- 2.3 Canonical Ensemble . . . 17
- 2.4 Runs and Data . . . 18
- 2.5 Expected Results . . . 19

3 System Classification **24**
- 3.1 Cluster Recognition . . . 24
- 3.2 Order Parameters . . . 25
  - 3.2.1 Bonds . . . 25
  - 3.2.2 Positional Order . . . 26
  - 3.2.3 Orientational Order . . . 26
  - 3.2.4 Porosity . . . 27
- 3.3 Topology study . . . 31
  - 3.3.1 2asym1c . . . 31
  - 3.3.2 3asym . . . 39
  - 3.3.3 Mouse . . . 46
  - 3.3.4 6patch . . . 53
  - 3.3.5 VN . . . 60
  - 3.3.6 VO . . . 65

4 Machine Learning **73**
- 4.1 Dimensionality reduction and order parameters . . . 73
- 4.2 PCA results by topology . . . 79
- 4.3 Clustering and automated classification . . . 94

5 Summary **104**

1 Introduction

1.1 From single particles to clusters

1.1.1 Self-assembly and phase transitions

Self-assembly is a process responsible for the spontaneous formation of structures from disordered smaller components [1]. The broadness of this definition is reflected in the wide range of possible scenarios, since the concept is so general that it can be observed at different length scales: it may be a formation of atomic solids from atoms, of polymer chains from molecules or of micro-structured crystals from colloidal particles. This holds true as long as there exists appropriately strong interaction forces scaled alongside. Nature itself has been taking advantage of self-organization since the dawn of time, as it serves as the guiding principle for the assembly of living cells [2, 3]. Even the process of evolution appears to be influenced by self-organization more than Darwin believed [4]. Humanity quickly discovered the practicality of self-assembly when casting metal and to this day it still serves as an excellent tool for modern material synthesis. On a molecular scale this process is evident, but research has led to the exploration and creation of more complex matter and applications [5, 6].
Under suitable circumstances, this process can be utilized for the designed creation of a wide variety of structures due to those properties' scalability.
Self-assembly can occur via different processes through clustering, chaining, cross linking, and liquid-liquid, liquid-gas phase separation [7–9]. One of the self-assembly processes observed in this work is nucleation, a typically stochastic process consisting in the formation of aggregates of bonded components, followed by a growth phase where the crystalline structures grow from these primary nuclei. Such a process is a phase transition, i.e. a transition of the system from one state to another. Structures will proceed to grow from these primary nuclei, which is an organization of matter representing a phase transition, defined as a medium's or substance's change of state [10, 11]. This change may involve a transition from one state of matter to another, structural, or a change of physical properties like magnetic ordering. A correctly defined order parameter is one that is able to represent such a phase transition in numerical form, typically by passing or reaching a characteristic value as the transition completes. Classical nucleation theory formally describes the kinetics of nucleation and proposes a prediction for the nucleation rate R. Relevant to the nucleation rate are the number of nucleation sites N_s, the attachment rate j describing the average rate of new connections with the nucleus, the Boltzmann constant k_B, the temperature T, the free energy cost of the nucleus at the top of the nucleation barrier ΔG^*, and the probability that a new phase will be formed by a nucleus at the top of the barrier, the Zeldovich

factor Z [12, 13]. This nucleation rate can be thus expressed as

$$R = N_s Z j \exp\left(-\frac{\Delta G^*}{k_B T}\right). \tag{1}$$

As seen above, the nucleation barrier plays part in the nucleation rate, as it relates to the probability of forming a nucleus. With surface tension σ of the spherically assumed nucleus with the size r and the difference in free energy per unit of volume between the two different nucleation phases Δg_v, the nucleation barrier can be written as

$$\Delta G^* = \frac{16\pi\sigma^3}{3|\Delta g_v|^2}. \tag{2}$$

This the Gibbs free energy of a nucleus with the so-called critical radius. Attachment of new particles to nuclei larger than this critical radius leads to a decrease in free energy, thus encourages self-assembly. Once the free energy has reached its minimum, the self-assembly process has concluded [14].

1.1.2 Self-assembly for materials design

For the creation of new materials, colloidal particles have excelled due to their versatility. Colloidal particles are suspended in a solution and vary in size and shape, although spherical specimens are more commonly used. Examples of such spherical colloids are made out of polymethyl methacrylate (PMMA), polystyrene, and nano crystals. A colloidal suspension of PMMA may result in glass, gel and cluster states, but it can also be combined with other materials to diversify the possible phases, for example with ferromagnetic properties [15, 16]. Polystyrene has been observed in gel, glass, and close-packed crystal states. Clever engineering allows for other possibilities, such as two-dimensional non-closely packed arrays of polystyrene, by controlling their size [17].

The use of nano crystals opens up a field of complex particles, that can also be engineered into crystal, gel, and glass states while retaining the properties of the used nano crystals [18]. Additionally their use gives way to new phases such as superlattice structures and materials that can be harder than atomic solids such as graphite [19, 20]. The programmability of such systems opens up a wide variety of new synthetic materials, and their creation methods go beyond the molecular scale [21]. Some of the recently popular methods on a nanometer scale take advantage of colloidal self-assembly. The generation of such colloidal nano-particles can be accomplished via bottom-up approaches, such as the utilization of nucleation of atoms into small particles, or so-called DNA origami [22–26]. DNA origami requires a long DNA strand which can be manipulated into various shapes and

fixed through the use of shorter DNA strands, which tie the long strand to itself. The folding of DNA therefore presents itself as a bottom-up approach to synthesize colloids of any desired shape. Colloidal particles synthesized by any of these methods can be used for all sorts of surface layer preparations in order to modify their substrate's physical properties. Examples include, but are not limited to, the modifications of electromagnetic, photonic, or even fluorescent properties [27–29]. Modern synthesis methods allow for the generation of a wide range of shapes and sizes of colloids, which can be applied using wet-chemical synthesis. To induce and control self-assembly in different ways, these colloidal particles are altered through surface modification creating topological interaction sites. Such particles are nowadays labeled patchy particles.

1.1.3 Patchy Particles

A hard shell particle, a particle that does not allow for any overlaps with other particles nor deformation, decorated with at least one anisotropic interaction area on its surface, is called a patchy particle. These anisotropic modifications introduce directional forces between all patchy particles and are a core component of the self-assembly process in these systems. The most utilized of such particles would be the Janus particle, which is a round particle that shows a difference in some property between its two semi-spheres. These differences in chemical, electrical, or physical properties can lead to strong interaction forces between pairs of patchy particles [30, 31]. One possibility to produce such particles is masking areas of particles using waxes, gels, or elastomers during treatments, followed by demasking to reveal the original area which is left unaffected during any surface treatment. Another one would be the deposition of a small metal layer onto particles, for which polystyrene particles are commonly used. Patchy particles bond with each other through these attracting patches, forming larger connected networks [32–35]. Those patches can vary in number, size, shape, placement, and their interaction potential. A smaller patch size creates more directionality in the interaction between two particles, and with an intentional selection of their placement provides a reliable way to reproduce replicable structures. Any slight change to the selected particle's or patches' parameters can drastically influence the assembled structures' properties [36]. This sensitivity to patch placement during self-assembly results in highly adaptable structures with properties heavily depending on the parameters of their building units.

1.1.4 Patchy triangles and their applications

Through the use of the previously mentioned methods, it is possible to create patchy particles of any shape, allowing for the generation of all sorts of structures

through self-assembly. Experiments on the generation of triangular DNA origami have already been conducted and show promising results for the future [37]. The focus of this work is on two-dimensional triangular patchy particles and the study of correlations between various patch placements as well as the properties of their two-dimensional aggregates. Significantly different patch placements will be referred to as different patch topologies. Once generated, these particles can model any system made of units that closely resemble these patchy triangles in their connectivity properties. An example would be to view the triangles as a bounding box or a simplified stencil for molecules, such as the trimesic acid aggregates, which are open for bonds in just the same positions as the attractive patches on triangles [38–40]. A similar approach has been conducted through the use of patchy rhombi with four patches for TPTC (p-terphenyl-3,5,3',5'-tetracarboxylic acid), which has four functional carboxyl groups at the ends of the molecule, resembling the connective behavior that can be simulated [41]. With an intentional selection of building blocks, two-dimensional geometric particles could be used for membranes or for surface layer applications with a low number of defects while utilizing the self-assembly process during manufacturing [42, 43]. Another application is the engineering of self-cleaning surfaces or the modification of their wettability [44–46]. The third popular use case for patchy particles is in optics, for example, as photonic crystals, since they allow for modification of the photonic band gap, or for structural coloring and iridescence [47–51].

1.1.5 Self-assembly simulations

It is possible to skip the manual process of setting up experiments for all the systems of interest by taking advantage of the current state of computer science. One approach to simulate the process of self-assembly for multiple patch topologies and their variations is the utilization of Monte Carlo simulations [52]. A Monte Carlo simulation is a method of gaining an understanding of a problem by studying an ensemble of systems generated with probabilistic behavior, thus within their physical limitations. The nature of this simulation approach creates particle behavior similar to Brownian motion and a probabilistic self-assembly via the attractive patches [53, 54]. At its core, the Monte Carlo simulation consists of random moves, that need to pass an acceptance criterion. This criterion considers the probability density of finding the system in a suggested configuration and then uses a random number generator to accept or discard a chosen move. This process makes it possible to study the equilibrium phases of randomly self-assembled systems of each patch topology. The simulation slowly iterates through time in a discrete manner, while calculating the probability of random actions (movements) at each step.
At this point it is necessary to introduce the boundary condition for hard particles:

an infinite potential preventing any overlap. Attractive and repulsive forces are all taken into consideration to discard any cases of physically impossible behavior. Any new configuration needs to adhere to these interactions by passing the acceptance criterion. If a new configuration is accepted, it is stored to serve as the starting point for the next time step. Multiple simulations based on different random starting configurations, where each configuration is an ensemble of particles with a given position and orientation. Within a relatively short time frame it becomes possible to study a large parameter space on the Vienna Scientific cluster. With strong attraction forces, the standard Monte Carlo moves (translational and rotational) show low acceptance rates, which also lead to larger clusters getting stuck in place. Through the introduction of virtual cluster moves it is possible to avoid these so-called kinetic traps [55].
An alternative to the Monte Carlo simulation is performed by the use of molecular dynamics simulations. In this simulation approach, forces are calculated for the initial conditions and then, after integrating Newton's equations of motion for the selected time frame, one can calculate system averages for observable quantities [52]. For hard particles a molecular dynamics simulation increases in computational intensity. To calculate these systems, pair collisions have to be calculated iteratively, and the collision times must be stored. Following this, the algorithm has to search for all particles that are affected by this collision before calculating the next collision [56].
In order to gain more samples of the patchy triangles' phase space, the ensemble of each patch topology is further divided into systems that have assembled under different temperatures and slight variations in their characteristic patch placements. Due to technical reasons, the scenario chosen for this simulation was the one of a canonical ensemble, in which the number of particles and system temperature within a simulation box of certain size all remain static. Thus, it simulates the behavior of a system of a set number of particles of a specific patch configuration within a static volume under a static temperature. Once its internal energy has converged to a minimum, a system has reached a stable or metastable state. At this point, the simulations are concluded and enable the study of resulting phases [57].

1.2 Gaining a quantitative understanding of large particle assemblies

1.2.1 Order Parameters

A typical tool for the study of an ensemble of systems is the utilization of order parameters for a quantitative classification. Order parameters are scalar or vector values that show a characteristic behavior for different phases. This per-

mits to understand the state of a system, and therefore classifying it, without visual inspection of a snapshot. Order parameters also typically allow to spot a phase transition, since they should return noticeably different values for different phases, if the right parameters are selected. A common example would be the bond-orientational order parameters Q4 and Q6 [58]. The Q4 order parameter represents cubic symmetry and converges towards 1 for a crystal with a cubic lattice structure. With a decrease in cubic positional order, Q4 converges towards 0, which represents a system with no cubic positional order. In this work, order parameters are selected to account for the relative position, relative rotation, number of bonds per particle, as well as the porosity of a system. This enables the identification of the different phases.

1.2.2 Automated classification

Once the phase space is mapped to an order parameter space, it becomes possible to find clusters in the order parameter space. With correctly chosen order parameters, these clusters represent all systems in similar states and will also reflect differences. Application of modern clustering algorithms to this large dataset speeds up the process of classification. Clustering algorithms are sensitive methods based on very different mathematical principles resulting in entirely different solutions, even if applied to the same problems. It is therefore essential to select a suitable algorithm for any problem to receive a reasonable result [59]. Factors to take into consideration when selecting the right method are, for example, the dimensions of the problem or the shape in space, knowledge over the number of clusters, and the relevance of noise recognition. The selected clustering algorithm is called Hierarchical Density-Based Spatial Clustering of Applications with Noise, or HDBSCAN [60]. This density-based approach allows for a classification independent of cluster shapes without introducing a bias by limiting the number of classes one is looking for. This is desired because it allows the algorithm find any patterns within the ensemble of assembled systems, without creating an error due to initial assumptions concerning the number or shape of clusters. Additionally, the hierarchical component of this method allows the user to retrieve the probability of a point assigned to any cluster. This can be applied as a way to detect phase transitions where the systems are a mixture of two phases, thus returning conclusive order parameters. Once applied, the method's parameters are adjusted to find the desired outcome, HDBSCAN automatically assigns any system to a cluster and thus to a class or phase the system is in. This approach enables the detection of patterns in the combinations of order parameters for each phase, without the need to identify singular specific order parameters for each one.

1.2.3 Data comprehension for large data sets

The space spanned by order parameters increases in its number of dimensions with each additional order parameter needed for classification. The difficulty of analyzing higher-dimensional spaces creates a trade off between selecting as many parameters as needed for correct classification and as few as possible for analysis, also known as the curse of dimensionality [61]. Simply adding more order parameters is therefore not desirable as they increase computation times and additionally might not even be completely independent of each other. Once a selection of suitable order parameters is found, methods of dimensionality reduction can be applied to reduce a potentially complex multi dimensional space into something more easily comprehensible, thus avoiding the curse of dimensionality. Nowadays, multiple options for dimensionality reduction processes are in use and have been comprehensively studied [62]. Out of those, the Principal Component Analysis (PCA) was selected [63]. This method is an unsupervised, linear transformation algorithm, creating a new linear combination of order parameters that spans a new subspace. Those order parameters are selected to explain most of the variance within the system, thus finding a combination of the most relevant order parameters. Depending on the amount of principal components used, one can scale the order parameter space to different dimensions. PCA performs worse for non-linear data and requires data standardization to deliver the best results. It has earned its use in the space of neural network models and machine learning for its quick calculation speed [64, 65].

2 Simulation

2.1 Monte Carlo Method

The self-assembly through nucleation of anisotropic colloids is simulated using the Monte Carlo method. This simulation approach describes the fluctuation-induced first order phase transition by iterating through all particles and creating movements similar to Brownian motion. Each single particle is assigned a random translation and rotation, followed by a subsequent collision detection. Once this is done collectively for all particles the system has gone through a Monte Carlo sweep. The probability of accepting a move is correlated to the proposed move's system energy, where a higher energy will lead to a lesser probability. Using a uniform random number generator, some moves pass the acceptance criterion, while others are discarded. The particles used are hard particles, described by the pair-potential V, as well as a distance d and rotation Ω between two particles (i, j). Since the pair-potential $V(d_{ij}, \Omega_{ij})$ is infinite in the case of an overlap, the probability of accepting moves with any overlap between two hard particles is 0. This is due to the factor $\exp(-\beta V)$ in the Metropolis algorithm. A visualization of the pair-potential V can be found in figure 1.

$$V(d_{ij}, \Omega_{ij}) = \begin{cases} \infty & \text{if overlapped} \\ 0 & \text{if not overlapped} \end{cases} \tag{3}$$

Attractive circular patches are placed on the surface of the two-dimensional particles, enabling the particles to form attractive bonds. These patches are not considered as a part of the geometry of the hard particles and only act as interaction points. The pair-interaction of those patches is described by a finite square well potential W. For any two patches, the interaction range is equal to the diameter δ of one patch. The interaction is modeled using the pair-potential W shown in figure 1 for any distance p_{ij} from the patch center.

$$W(p_{ij}) = \begin{cases} -\epsilon_{ij} & \text{if } p_{ij} < \delta \\ 0 & \text{if } p_{ij} \geq \delta \end{cases} \tag{4}$$

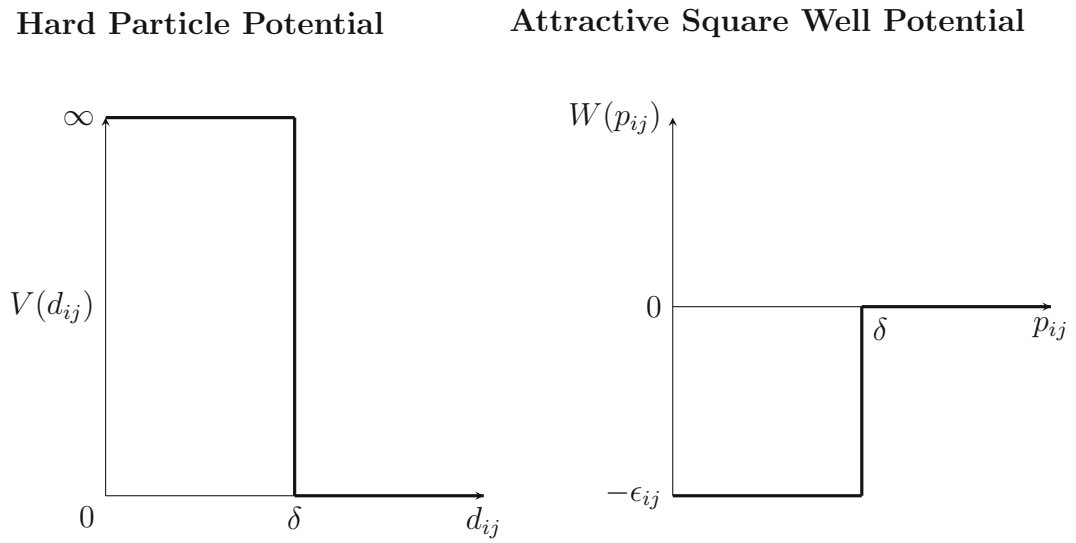


Figure 1: Left: The infinite step pair-potential of hard shell circles (i, j) with the diameter δ, allowing for no penetration of any two particles.
Right: The attractive square well potential used to define the pair-interaction between two circular patches (i, j) with the diameter δ.

Furthermore, the formation of clusters requires the introduction of virtual cluster moves in addition to the single particle moves. Otherwise, particle clusters can get stuck in so-called kinetic traps, mentioned in section 1.1.5. These cluster moves introduce translations and rotations to groups of particles.

The method chosen to detect a collision or overlap between two particles is based on the separating axis theorem shown in figure 2 [66]. By checking the vertices' projections onto the face normals of one triangle, it becomes possible to spot any overlap between particles. If the overall distance D between the projections of the outermost ends of both particles is larger than the sum of each particle's own projection P_1 and P_2 onto a face normal, there is no overlap along this axis. Therefore it is necessary to check if

$$D > P_1 + P_2, \tag{5}$$

is true for at least one of the face normals of a triangle. Through the application of this criterion, it becomes possible to detect overlaps between the bodies of any two particles.

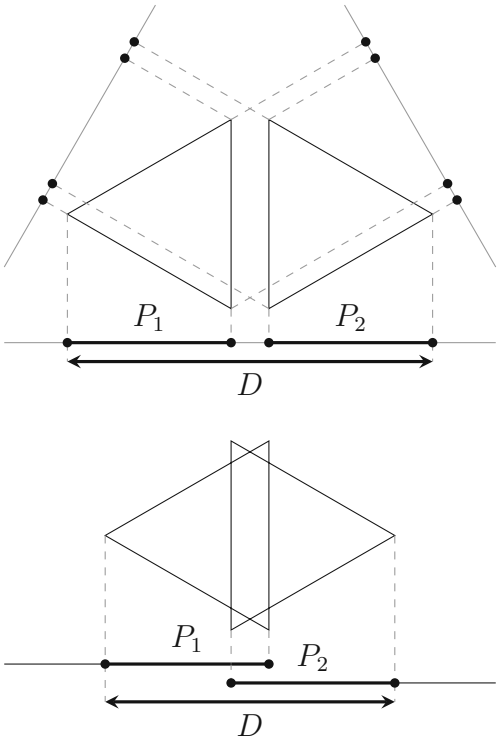


Figure 2: Collision detection through application of the separating axis theorem. Upper: No collision detected, since the condition in equation (5): $D > P_1 + P_2$, is true for one of the face normals, thus a separating axis is found. Lower: Collision detected because $D \leq P_1 + P_2$ holds true for the projections onto all of the face normals.

The framework for the Monte Carlo Simulation was written by C. Karner for the investigation of patchy rhombi [41]. Thus it only had to be extended to allow for the simulation of triangular particles, which included defining a new particle type, the collision detection using the separating axis theorem, and periodic boundary conditions.

2.2 Particle Parameters

The particles used in our Monte Carlo simulations are patchy hard shell equilateral triangles. Circular interaction patches are placed on the edges or vertices of these triangles. Different placement positions and number of attached patches lead to

a number of different possible variations, so-called patch topologies. All of those topologies are subject to the same set of parameters: the number of patches n_p and their placement along the vertices Δ, the patch radius r, and the side length of the equilateral triangle L, shown in figure 3. For this work, the side length will be constant at $L = 1$ and the patch radius at $r = 0.05$, whereas the patch placement and number will be varied to explore a broad parameter space. Due to previous work on patchy rhombi, the choice of $r = 0.05$ appears as a good starting point to study the self-assembly of patch triangles, since it serves as a good compromise between interaction range and inter-particle distance [36].

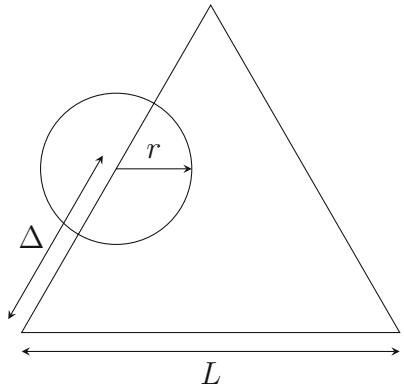


Figure 3: A depiction of all the particle parameters, where $n_\mathrm{p} = 1$, $L = 1$, $\Delta = 0.5$ and $r = 0.2$. The patch radius in the simulation is $r = 0.05$, but is set to a larger size in this graphic for better readability.

After manually searching for patch topologies that have the possibility to form into different small structures, six are selected, which are described below and are shown in figure 4. The results of the manual small cluster analysis will be discussed in section 2.5. Some of the patches will remain at the same position for all simulations of their topology, while others are static but placed at different positions Δ for different runs. The reference vertices for Δ are defined also defined in figure 4 for each topology and explained in the list of topologies below. Some of the chosen patch topologies include the asymmetrically placed patches, which references to patches that are placed on two different edges and a patch placement of Δ for one patch and $1 - \Delta$ for the other with respect to their connecting vertex.

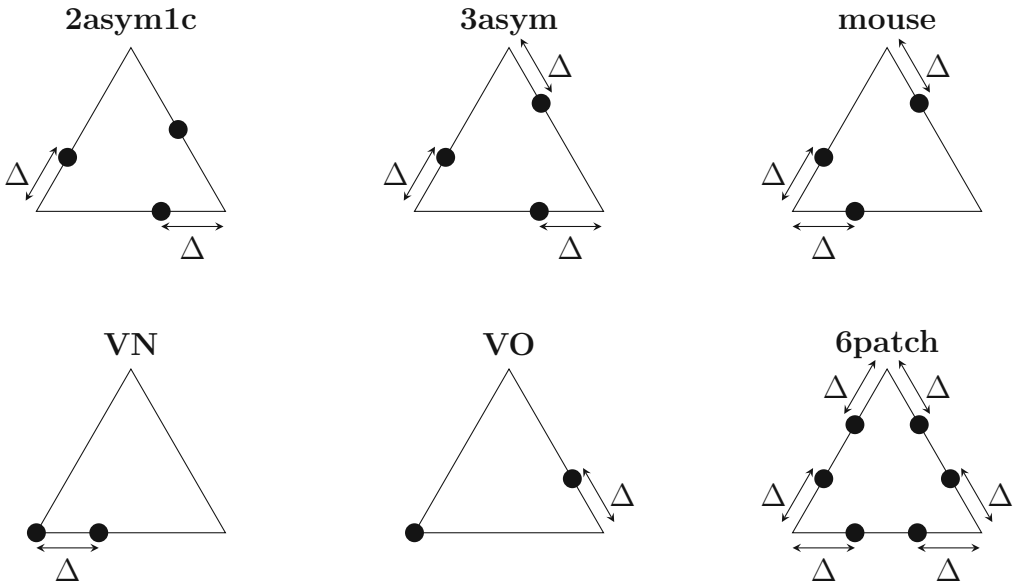


Figure 4: List of selected patch topologies. The patches drawn as black circles sit on the border of the hard triangles. Patches at the edges of a triangle (**VN**, **VO**) or defined as centered along a vertex (**2asym1c**) will remain there for all iterations of their respective topology. In this figure the position is set to $\Delta = 0.33$ and the patch radius to $r = 0.05$ for all topologies.

The list of the selected patch topologies is the following:

2asym1c: Two asymmetrically placed patches along two of the edges and the last one centered on the last edge with $\Delta \in \{0.05k | k \in \mathbb{N} : 2 \leq k \leq 10\}$.

3asym: Three asymmetrically placed patches, one on each edge of the triangle with $\Delta \in \{0.05k | k \in \mathbb{N} : 2 \leq k \leq 10 \vee k \in \{0\}\}$.

mouse: Two asymmetrically placed patches and one symmetric to either one of those with $\Delta \in \{0.05k | k \in \mathbb{N} : 2 \leq k \leq 10\}$.

VN: One patch placed on a vertex and one at a neighboring edge with $\Delta \in \{0.05k | k \in \mathbb{N} : 2 \leq k \leq 20\}$.

VO: One patch placed on a vertex and one at its opposite edge with $\Delta \in \{0.05k | k \in \mathbb{N} : 2 \leq k \leq 10\}$.

6patch: Six patches, two on each edge with $\Delta \in \{0.05k | k \in \mathbb{N} : 2 \leq k \leq 9\}$.

For the topologies 2asym1c, 3asym, mouse and 6patch, a patch placement of $\Delta > 0.5$ creates particles, which are mirror-symmetrical to particles of the same

topology with patches placed at $1-\Delta$. These topologies are only simulated for $0.1 \leq \Delta \leq 0.5$ in order to avoid redundant calculations for the mirror-symmetrical high Δ variations ($0.1 \leq \Delta \leq 0.45$ in the case of the 6patch topology, to avoid stacking of patches at $\Delta = 0.5$). The case $\Delta = 0$ is only studied for the 3asym topology, since $\Delta = 0$ would be redundant for other. For example, VN with $\Delta = 1$ is the same particle as VO with $\Delta = 0$. Setting $0 < \Delta \cdot \sin\left(\frac{\pi}{3}\right) < r$ would lead to some undesired bonds due to the patch unintentionally showing on a second side of the triangle. Additionally, patches placed within certain Δ ranges allow for potential patch to patch overlaps occurring within the area of a triangle. Both of the cases mentioned above are depicted in figure 5 below. The possibility

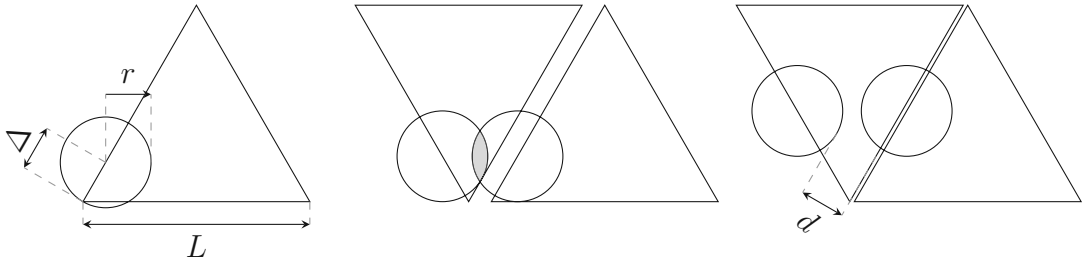


Figure 5: Left: Patch interaction spills over to another side of the triangle if $0 < \Delta \cdot \sin\left(\frac{\pi}{3}\right) \leq r$. Center: Interaction between two particles are possible, even if the patches aren't placed along the sides facing each other. Right: Bonds between two patches inside the area of a triangle can be avoided if $d > r$.
$n_\mathrm{p} = 1$, $L = 1$, $\Delta = 0.2$ and $r = 0.2$ for a better visualization.

of overlaps within the area of a triangle was initially missed when deciding the space of simulated parameters, thus some variations are simulated which show significantly different behavior. The self-assembly of all systems with $\Delta = 0.1$ is influenced by this phenomena. To avoid this issue the distance d between a patch and any other edge should be at least larger then r, as seen in figure 5. ThusΔ should be set within $2r/\sin\left(\frac{\pi}{3}\right) < \Delta < 1 - 2r/\sin\left(\frac{\pi}{3}\right)$ when working with equilateral triangles of side length $L = 1$. For a patch radius $r = 0.05$, that means $0.11547 < \Delta < 1 - 0.11547$ avoids those undesired overlaps.

2.3 Canonical Ensemble

The principal thermodynamic variables (N, V, T) of the canonical ensemble are explored in all combinations of the options shown in table 1 [52]. Instead of setting the volume V to a fixed value, it is calculated for selected particle densities φ and absolute number of particles N with a given volume of a single particle using $V = \frac{N}{\varphi}$. The self-assembly is mediated by controlling the temperature.

Higher temperatures generally decrease the possibility of bonds being energetically favorable, making them more likely to break apart. With lower temperatures the self-assembly proceeds faster, but with a higher likelihood of retaining defects in the structure. Initially, the temperature $T = 0.1$ was chosen as it has been used in similar simulations for rhombi [36]. Preliminary runs showed self-assembly with too many defects, and therefore higher temperatures T are needed for the simulation of self-assembling triangles. The effects of temperature on the self-assembly was scanned for a range of $T = 0.1$ to $T = 0.25$ and the most promising candidates are selected, shown in table 1.

variable	selected values
N	$\{1200\}$
φ	$\{0.1, 0.25, 0.3\}$
T	$\{0.115, 0.13, 0.145, 0.16, 0.175, 0.19\}$

Table 1: All selected values for the principal thermodynamic variables of the canonical ensemble. V is defined by the particle density φ.

2.4 Runs and Data

The final number of runs N_{runs} is quickly compiled using the amount of different densities $N_\varphi = 3$, temperatures $N_\text{T} = 6$, and patch positions N_Δ. Since N_Δ varies for each topology, it is advisable to refer to table 2 for a summary of patch topologies and their individual number of patch positions $N_{\Delta,\text{top}}$.

topology (top)	2asym1c	3asym	mouse	VN	VO	6patch
$N_{\Delta,\text{top}}$	9	10	9	19	9	8

Table 2: Some topologies allowed for a wider range of Δ, when observing the possible unique patch configurations mentioned above.

Furthermore, each unique combination is simulated $N_\text{ID} = 16$ times, resulting in a final number of runs of:

$$N_{\text{runs}} = N_\varphi \times N_\text{T} \times \sum_{\text{top}} N_{\Delta,\text{top}} \times N_\text{ID} = 18432. \tag{6}$$

A little trick one can employ to artificially expand the number of final data points for statistical analysis is to consider a set of final states $N_{\text{states}} = 5$ from one system such that

$$N_{\text{data}} = N_{\text{runs}} \times N_{\text{states}} = 92160. \tag{7}$$

After initial testing to find the suitable simulation parameters, the calculations were started on the VSC3. The calculations later had to be moved to VSC4 due to the shutdown of VSC3. Runs continued until system energy had converged to a minimum for all systems. At that point, they were stopped and ready for analysis.

2.5 Expected Results

Due to the nature of the Monte Carlo method the system of colloids will self-assemble, if given enough time under the right thermodynamic conditions. The topologies shown in figure 4 are handpicked to be a part of a selection of possible small assemblies, created under the criteria of satisfying all possible bonds without creating overlaps. It is expected that some of the resulting systems will show bigger instances of the examples shown in figures 6 to 12 below. Some will combine into crystals, while others form gels, depending on the system variables for each individual run. With at least two patches of the same Δ at different edges, the systems can show porous or closed crystal structures as seen below for the 2asym1c and 3asym topologies figures 6 to 8 and for also the 6patch topology in figure 9. Centered patches typically lead to more closed structures, where off-centered patches create flexible gels or porous crystals.

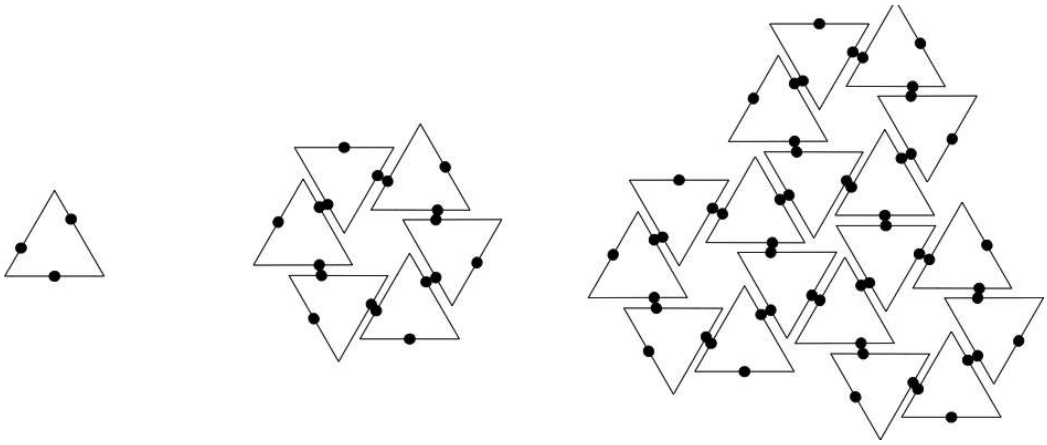


Figure 6: Small cluster analysis of the 2asym1c topology with $\Delta = 1/3$. Left: A single particle. Center: A ring of six particles creates a pore at its center, whose size changed with Delta. Right: These compact clusters form larger systems with multiple pores and different pore sizes. The small pore in the center when combining three rings (right) will remain the same as it's a result of the centered patch. Gels and crystals will become increasingly more compact when Δ approaches 0.5

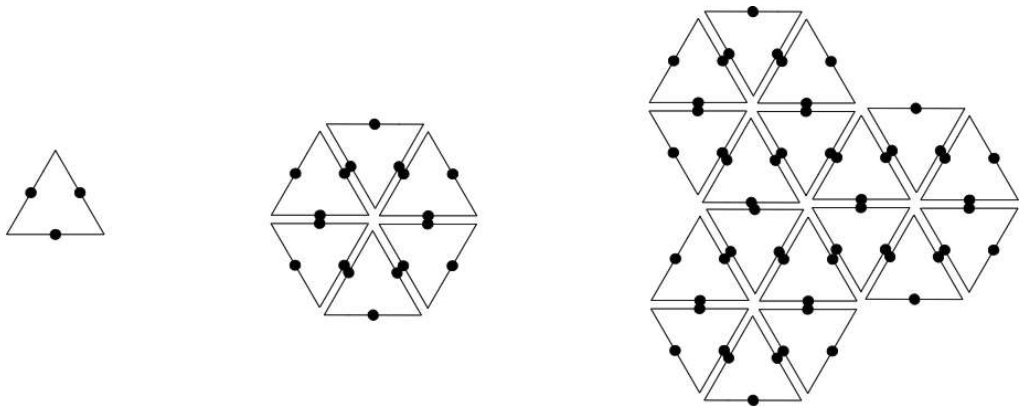


Figure 7: Small cluster analysis of the 3asym topology with $\Delta = 1/2$. Left: A single particle. Center: Particles only consisting of centered edges form close-packed clusters. Right: These small close-packed clusters self-assemble into larger pore free structures.

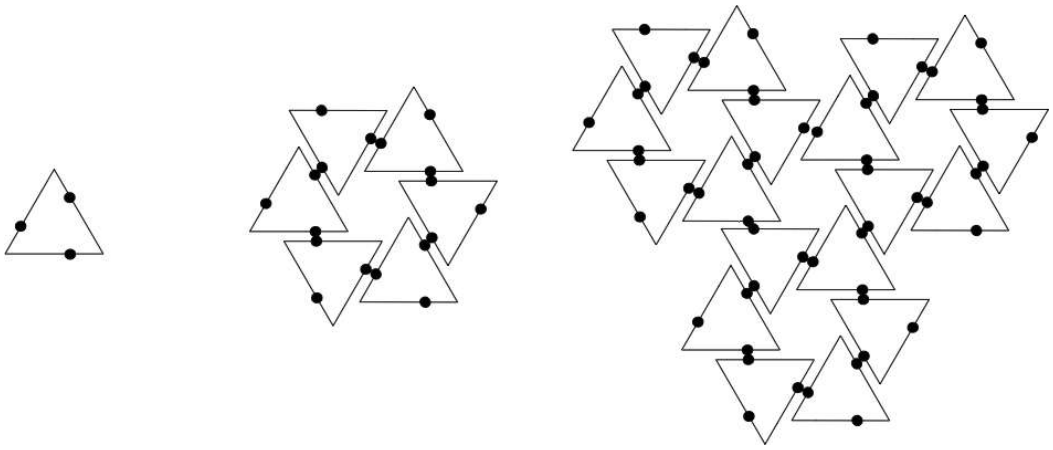


Figure 8: Small cluster analysis of the 3asym topology with $\Delta = 1/3$. Left: A single particle. Center: As patches are placed closer towards the edges, pores are created with increasingly larger pore sizes. Right: Selection of the patch placement for the 3asym patch topology allows for very precise control over the pore size. This combination of topology and Δ is expected to deliver very porous gels and crystals.

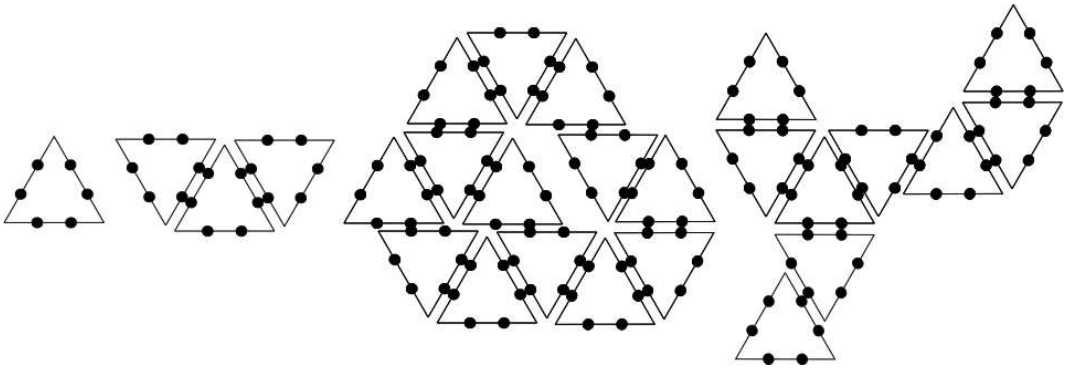


Figure 9: Small cluster analysis of the 6patch topology with $\Delta = 1/3$. First: A single particle. Second: For this topology, very compact clusters are expected. Third: These can form crystals that either are completely closed or potentially characteristic pores. Fourth: Gels based on this geometry are very likely to form close-packed connections.

For the mouse topology, the potential pores in crystals would not arise at the center of ring-like clusters, but rather only between hexagonal arrangements figure 10.

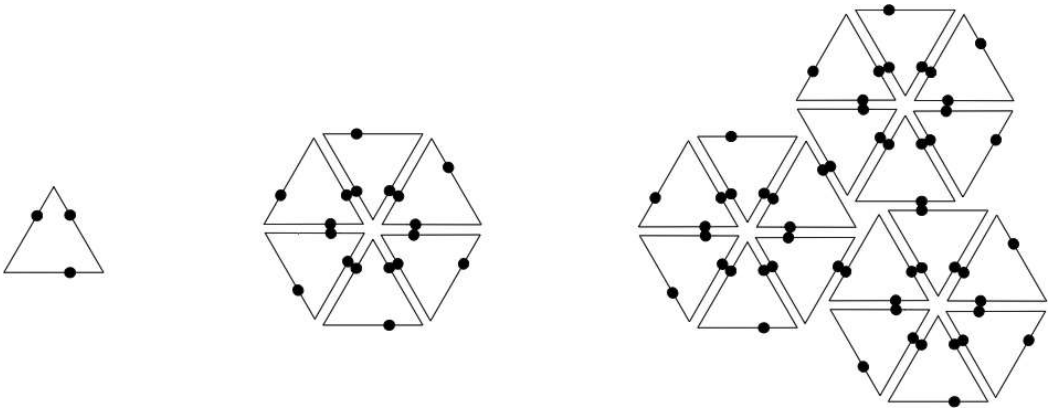


Figure 10: Small cluster analysis of the mouse topology with $\Delta = 1/3$. Left: A single particle. Center: For this topology, very characteristic closed substructures are expected. Right: Here are here are no pores within a hexagonal substructure, but instead only between the hexagonal groups.

When studying the options for two-patch topologies, two options arise. If the second patch is on the edge opposite the vertex patch, the vertex patch can again lead to the formation of hexagonal shapes with connective patches as seen in figure 11. These connective patches can create closed crystals, or gels. Alternatively, they can lead to chain structures instead of gels with multiple branches.

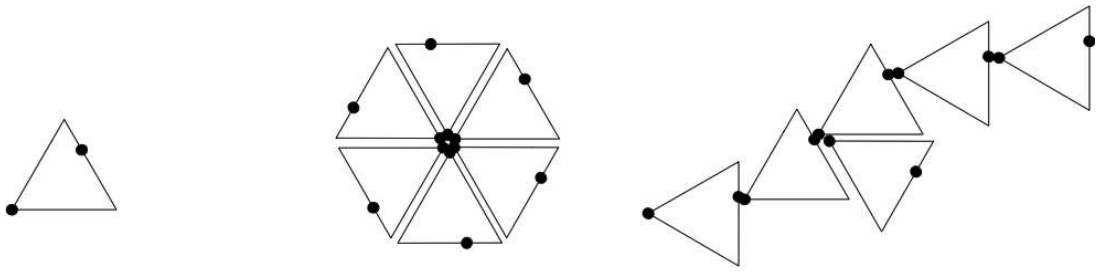


Figure 11: Small cluster analysis of the VO topology with $\Delta = 2/3$. Left: A single particle. Center: The hexagonal shape could perform similar to the mouse patch topology. Right: Using this patch topology, chains of triangles are now also possible.

Once the second patch is placed on an edge connected to the patch vertex, self-assembly is expected to create multiple small and saturated clusters seen in figure 12. These small saturated clusters cannot form larger structures, remaining in gel or liquid state.

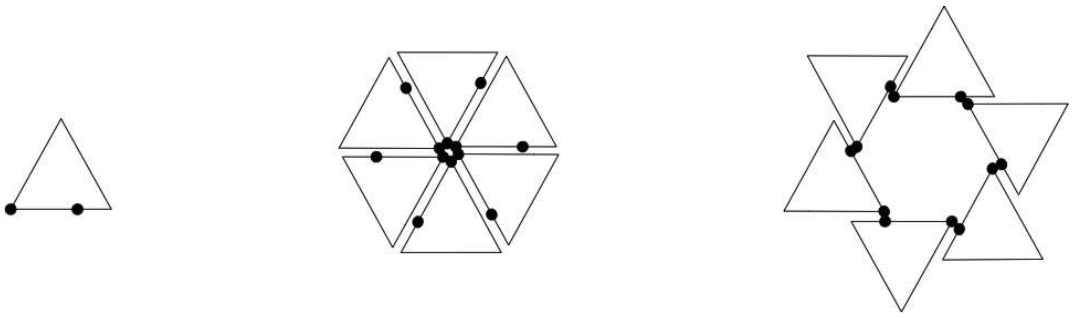


Figure 12: Small cluster analysis of the VN topology with $\Delta = 2/3$. Left: A single particle. Besides gels and chains, two options seem reasonable. The hexagonal shape (center) or closed off rings (right) will form liquids under the right circumstances.

Due to the number of variables and their ranges, there is a large amount of potential combinations to be explored (seen in section 2.4) and the options shown above are not an exhaustive list. Machine learning algorithms will be used in section 4 to analyze the resulting space of self-assembling triangular anisotropic colloids under different conditions.

3 System Classification

Different phases show different kinds of structures in the particle clusters. In order to quantify all those different systems, a set of order parameters is introduced. These order parameters span a space, which can be used to group systems into different categories such as fluid, gel-like crystalline, and porous states. The aim here is to create an overview of the possible states and state diagrams for each topology. Some of the states will be split into different subgroups, which will be described in more detail. Order parameters are selected to describe the particles relative orientation, position, number of bond, and also the porosity of clusters. They are then used to classify different states by identifying patterns in the order parameter space.

3.1 Cluster Recognition

The initial step necessary to analyze those systems is to find out which particles are grouped as clusters via their patches. First, the distance between two close patches p_{ij} is compared to their patch diameter δ. If this distance is smaller, or more explicitly, if $p_{ij} \leq \delta$, then the patches overlap and therefore have formed a bond. The information about which patches have bonded is gathered to define groups of bonded particles as clusters. The Voronoi decomposition from the freud package is implemented with periodic boundary conditions for a calculation of order parameters, but can be utilized already for particle cluster recognition, as seen in figure 13 [67–69]. Using the Delaunay-Triangulation the Voronoi method separates the space into regions defined with the input points at their center. Each of those cells contains all the points closest to its own center compared to all other centers, using the Euclidean metric. At first, just the triangles' center positions are used as the input points with periodic boundary conditions. The method returns information about which cells are neighbors to each other. This reduces the number of patch-to-patch distance checks necessary to identify clusters to just taking the particles' information of a handful of neighbor cells.

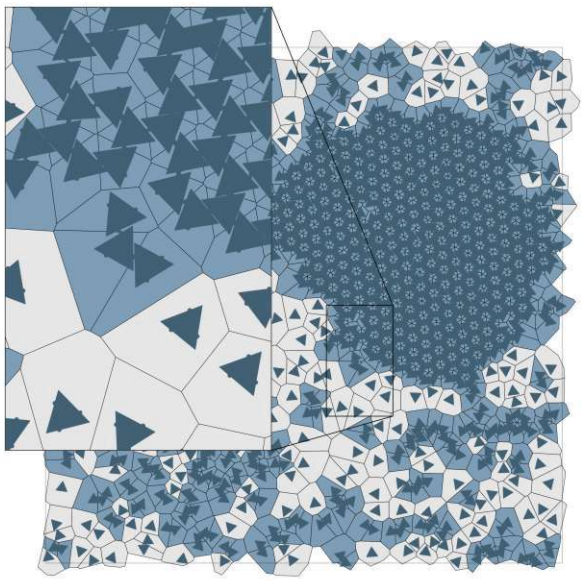


Figure 13: Example of a crystallizing system, with different large porous clusters with hexagonal pore symmetry. The Voronoi decomposition is applied to the set of triangle center points. Particles are colored in orange, Voronoi cells of bonded particles drawn in green, Voronoi cells of particles without any bonds in grey. Topology = 3asym; $\Delta = 0.2$; $\phi = 0.1$; $T = 0.13$

3.2 Order Parameters

To investigate the large number of systems it is helpful to quantify them by using order parameters. The goal of the selected order parameters is to identify different types of states. By doing so it will be possible to find groups of clusters representing systems of similar states in the order parameter space.

3.2.1 Bonds

The first order parameter $\Phi_b(t)$ chosen to quantify the system describes the bonding behavior of the system in terms of the number of bonding partners. This parameter helps to distinguish more accurately between gels, crystals, chains, and other possible phases. It is based on the weighted arithmetic mean of the number bonded patches b_i per particle i over the total number of particles, which is

denoted as N This order parameter is normalized by putting the bonded patches per particle b_i relative to six, the maximum number of patches per particle and is expressed as

$$\Phi_{\mathrm{b}} = \frac{1}{N} \sum_{i=1}^{N} \frac{b_i}{6}. \tag{8}$$

The results of the bond order parameter are within a range from 0 to 1. A system with no bonds would return 0 and a system where six patches on all particles are bonded will return a value closer to 1. The latter won't occur in the scope of this work which is calculated using the canonical ensemble, where the assemblies usually have a boundary layer of particles in which not all bonds are satisfied, even with periodic boundary conditions,

3.2.2 Positional Order

To determine positional order, the Steinhardt Order parameters Φ_4 and Φ_6 are chosen [58]. These will help distinguish especially between square (in the case of Φ_4) or hexagonal lattice structures, using Φ_6. First Φ_n^{c} is calculated for each cluster c with a size N_{c}

$$\Phi_k^{\mathrm{c}} = \frac{1}{n_c} \sum_{i=1}^{n_c} \left| \frac{1}{N_n} \sum_{j=1}^{N_n} \exp(ik\theta_{ij}) \right| \tag{9}$$

and then given their relative cluster size as a weight when summed up for Φ_k. The weighted arithmetic mean is used to reduce the influence of smaller clusters on the total order parameter Φ_k. Here n_{c} denotes the number of particles in a cluster with N_n neighbors each and N the total number of bonded particles in the system. The angle between the ij-particle bond vector and an arbitrary fixed direction in space is written as θ_{ij}. And lastly, k is set to 4 or 6 depending on the desired order parameter, thus allowing the positional order parameter to be expressed as

$$\Phi_k = \sum_{c=1}^{N_{\mathrm{c}}} \frac{n_c}{N} \Phi_k^c. \tag{10}$$

The parameter is defined to return a value between 0 and 1, where 0 represents a cluster with no square ($k = 4$) or hexagonal ($k = 6$) lattice structure and 1 represents a perfectly ordered one.

3.2.3 Orientational Order

With the use of Φ_θ it will be possible to gain an understanding of dominant relative orientations in the different systems. For simplicity only bonded neighbors and

relative angle of their orientation θ_i are considered. The factor s takes rotational symmetry of triangles into account by setting $s = 3$. The number of bonds per cluster c is written as B_c and is used to determine the orientational order parameter Φ_θ^{c} per cluster with a size N_{c}. Again, n_{c} denotes the number of particles in a cluster with N_n neighbors each and N the total number of bonded particles in the system. Once the order parameter is calculated for each cluster

$$\Phi_\theta^{\mathrm{c}} = \frac{1}{2}\left(\frac{1}{B_c}\sum_{i=1}^{B_c}\cos\left(s\theta_i\right) + 1\right), \tag{11}$$

it is followed by a calculation of Φ_θ, the weighted arithmetic mean. The orientational order parameter can therefore be expressed as

$$\Phi_\theta = \sum_{c=1}^{N_{\mathrm{C}}} \frac{n_c}{N}\Phi_\theta^{\mathrm{c}}. \tag{12}$$

The cosine will result in 1 for parallel bonds and -1 for antiparallel, while taking the order of rotational symmetry $s = 3$ for triangles into account. In the end, the order parameter is rescaled to range from 0 to 1, such that it matches the scale of the other order parameters.

3.2.4 Porosity

Another selected criterion for classification of such structures is their porosity. Initially, the intention was to gain this information from the Voronoi decomposition by additionally incorporating the positions of vertices and patches as points for the Delaunay-Triangulation.

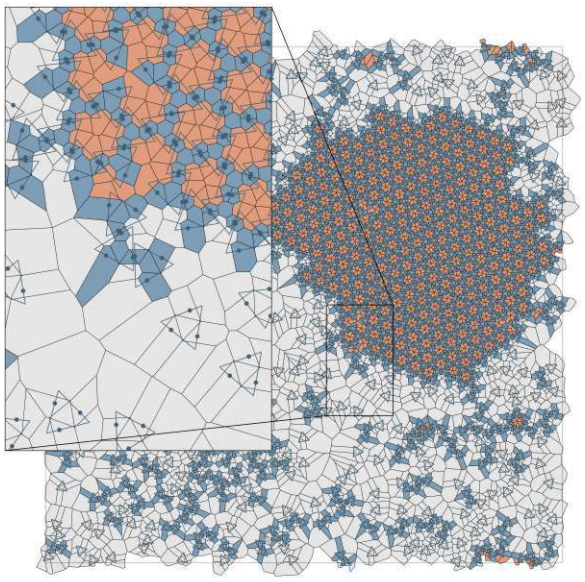


Figure 14: For this Voronoi decomposition the particles' vertices and patch positions are also considered as input points. For better readability, the particles are reduced to their outlines and patches, drawn in blue. The cells corresponding to the center points of bonded particles are also drawn in blue. If a cell originates from a patch that has formed a bond then it is also counted as particle cells. All other cells are void space, but if they are enclosed by particle cells then they are detected as a pore cell and colored orange instead of grey. This shows a consistent approximation of the clusters' pore sizes for some ensembles.
Topology = 3asym; $\Delta = 0.5$; $T = 0.16$; $\phi = 0.25$

The results for including more points are shown in figure 14. With the increased number of cells, a more detailed analysis would be possible, after classifying them correctly. Cells that originate from the positions or connected patches are counted as particle cells. The remaining cells are labeled as grey void or blue pore cells. While it is possible to gain information for larger pores using this technique, it shows to be inaccurate for small pores as shown in figure 15. The cells originating from vertex points overestimate those small pores since their boundary is always halfway to other close Delauney-Triangulation points.

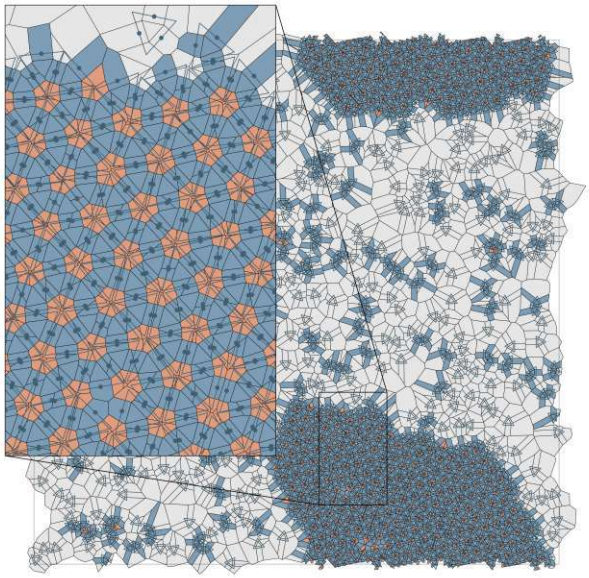
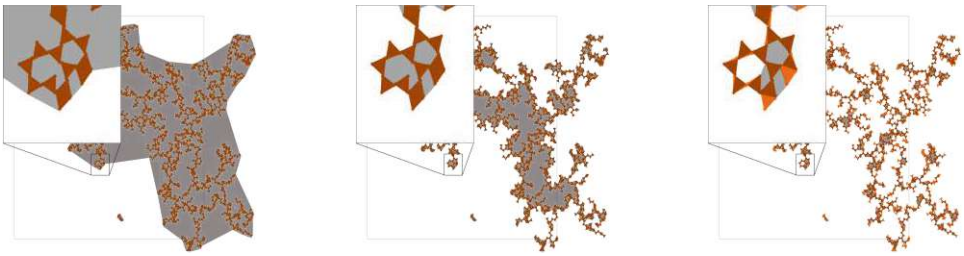


Figure 15: Once the pores get smaller, the Voronoi approach proves to be unreliable in measuring the porosity of the clusters. The particles' outlines are drawn in blue and the Voronoi cells corresponding to the center position of bonded particles drawn in blue. Cells created from bonded patches are also drawn in blue. All the blue cells are labeled particle cells. Any other cell surrounded by those particle cells is colored orange to represent a pore and the rest as gray if they are void. Topology = 3asym; $\Delta = 0.5$; $T = 0.16$; $\phi = 0.25$

Due to the limitations of the Voronoi approach, alpha shapes are used to calculate the porosity of a cluster instead [70]. Alpha shapes are piecewise linear geometries that create a bounding area of a set of points and are a generalization of the concept of convex hulls. Instead of purely convex shapes like the convex hull, alpha shapes allow for concave sections, allowing them to determine the shape generated by a set of points. How tight those shapes fit the set of points can be configured using the α parameter, where $\alpha = 0$ is equal to the convex hull of the set of points. More specifically, α is used to define the border points such that the perimeter of a circle with radius $1/\alpha$ can be drawn between any two points on the border of the set without containing any other point of the set within it. With

alpha shapes, these arcs are simplified to linear connections between those edge points. If α is low then the bounding area does not fit tightly around the set of points, whereas a too-high alpha value leads to a too-tight fit and and therefore unrealistic representation of the clusters. This is showcased in figure 16. Manual iteration through α values in randomly chosen snapshots of different phases led to the selection of $\alpha = 1.73$. It would be possible to automate the search for the optimal α for each individual system.



(a) $\alpha = 0.1$ **(b)** $\alpha = 1.73$ **(c)** $\alpha = 1.9$

Figure 16: Examples of how the α parameter influences how tight the alpha complex fits the set of input points, which in this case are the particles' vertices. The particles are colored in orange and the different alpha shapes are overlayed in gray. Low alpha values lead to an overestimate of the pore area as seen in **(a)**, whereas using too large alpha values underestimates the pore area which is displayed in **(c)**. The manually chosen alpha value of $\alpha = 1.73$ serves as an approximate solution and is showcased in **(b)**. Topology = 6patch; $\Delta = 0.1$; $T = 0.145$; $\phi = 0.25$

Using the area of a cluster's alpha shape $A_{\alpha,c}$, it is possible to determine the total space enclosed by particles. To determine the area of unoccupied space, the area of each particle is subtracted from the alpha shapes, which is always the area of an equilateral triangle A_{T} with a side length of 1. Again, the number of particles that are part of a cluster c is written as n_c and N_{c} is the total number of clusters. Next, the largest unoccupied pore area found in all of the systems A_{max} is found, and used to normalize the order parameter. Thus the order parameter

Φ_ϕ is defined as

$$\Phi_\phi = \frac{1}{A_{\max}} \sum_{c=1}^{N_c} \left[A_{\alpha,c} - A_T n_c \right]. \tag{13}$$

A value of 0 would represent a system where no area is enclosed by a group of particles and a value of 1 represents the case where the pore area is equal to the size of the largest pore area found within the data, which excludes the area covered by particles.

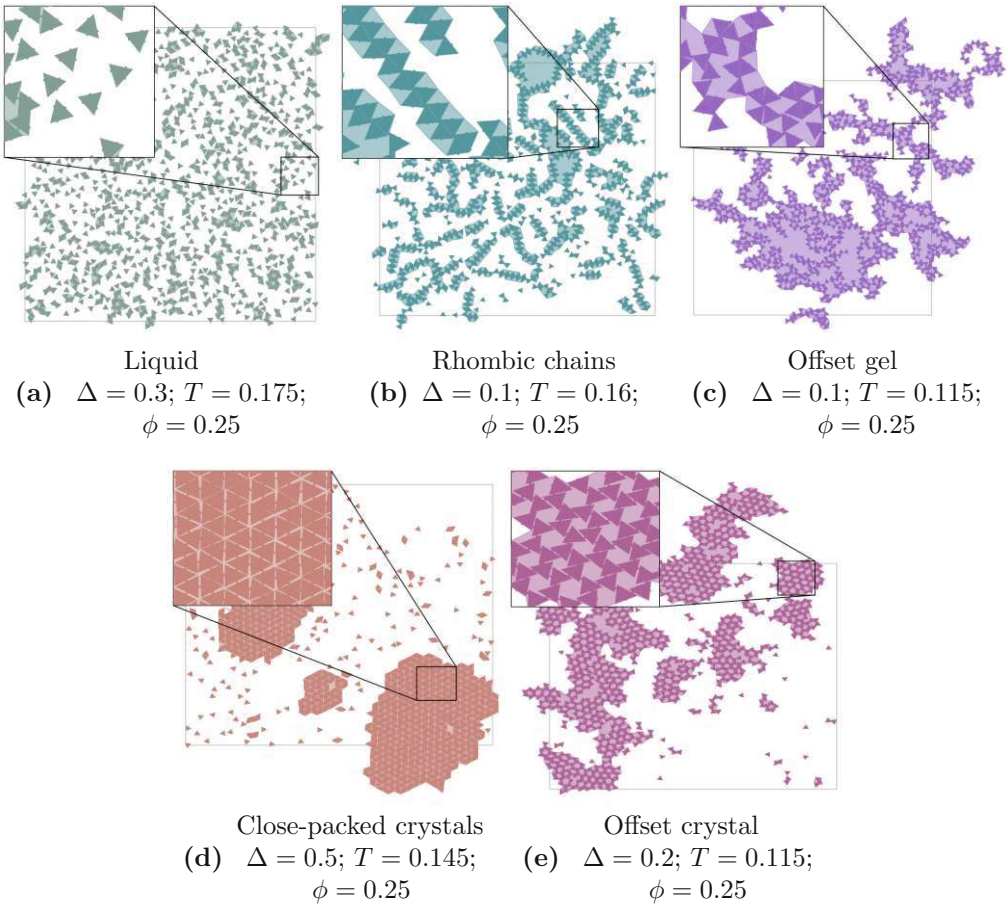
3.3 Topology study

The study of each patch topology reveals different phases depending on the system temperature and patch placements. These are subsequently discussed for each particle type. Variation in temperature and patch placement reveals a state diagram unique to each patch topology. These phases are classified by visual examination of snapshots for each T-Δ combination. Some of the patch topologies show significantly different behavior for $\Delta = 0.1$. The reason for this is the patch spillover explained in figure 5. Additionally the chosen order parameters for bonding behavior Φ_b, positional order Φ_4 and Φ_6, orientational order Φ_θ, and porosity Φ_ϕ are reviewed for each topology individually. The chosen approach is to take the average of each order parameter for each Δ-T combination. This approach assumes that all systems of the same T and Δ show similar behavior. The physical states occurring within the scope of this thesis can be broadly categorized into a handful of options. Liquids, where either most particles are not part of a cluster, or the average cluster size is close to 1. Crystals, which show a regular lattice structure and long-range order. These crystals can be porous, close-packed, and can even appear with a periodic substructure. Systems are categorized as chains, when the clusters are of a linear structure with two ends on average. Lastly, there are the gel states. Gels are disordered networks without long-range order with varying levels of porosity. Some show localized short range order in the form of different substructures. In the following section, physical states are labelled in accordance with these guidelines

3.3.1 2asym1c

A large variety of structures are observed for the 2asym1c topology, of which characteristic snapshots are shown in figure 17 and a quantitative representation of where the different structures appear in the temperature versus Δ plane is reported in figure 18 for $\phi = 0.25$. Examination of this diagram shows a liquid phase occurring at temperatures $T \geq 0.16$, while temperatures below this threshold allow the formation of various cluster states. At these relatively high temperatures,

bonds are not stable, which is why cluster sizes remain small. Only in the $\Delta = 0.1$ regime are larger clusters still found at the high end of the chosen temperature spectrum $0.115 \leq T \leq 0.19$. The behavior here is noticeably different due to the patch spillover explained in figure 5 and leads to the assembly of rhombic chains. These rhombic chains consist of pairs of particles that are connected with other pairs. For the same Δ at lower temperatures, the systems transition into a gel phase, where edges of bonded triangles are not perfectly aligned, but offset, creating a very porous gel. This freedom results in very porous and disordered structures in terms of position as well as orientation. Since the particles in these systems are not aligned along their edges and vertices, they are labeled as offset gels in the following. For all other patch placements this patch topology assembles into crystals with different pore sizes, depending on the location of the patches. More centered patches lead to smaller pores, leading to the generation of close-packed crystals. Placing the patches closer to the edge increases the pore size, as their center-to-center distance increases with Δ. These states are labeled as offset crystals.



Liquid
(a) $\Delta = 0.3$; $T = 0.175$; $\phi = 0.25$

Rhombic chains
(b) $\Delta = 0.1$; $T = 0.16$; $\phi = 0.25$

Offset gel
(c) $\Delta = 0.1$; $T = 0.115$; $\phi = 0.25$

Close-packed crystals
(d) $\Delta = 0.5$; $T = 0.145$; $\phi = 0.25$

Offset crystal
(e) $\Delta = 0.2$; $T = 0.115$; $\phi = 0.25$

Figure 17: The liquid phase **(a)** of the 2asym1c topology does not form stable bonds and therefore the clusters remain relatively small in size. The centered patch allows two particles to form rhombic pairs which then further assemble into chains for **(b)** via the patch spillover. Offset gels as seen in **(c)** appear to be the most disordered assembly of the cluster phases. While they also show rhombic pairs, the overall structure is less positionally and orientationally ordered. Crystals can be classified into two types, depending on their pore size: close-packed crystals **(d)**, which are almost perfectly aligned at their edges and vertices, and offset crystals **(e)**, which generally form larger pores. This difference is a result of how centered the patches are placed. Off-centered patches lead to porous structures, whereas centered patches result in compact assemblies.

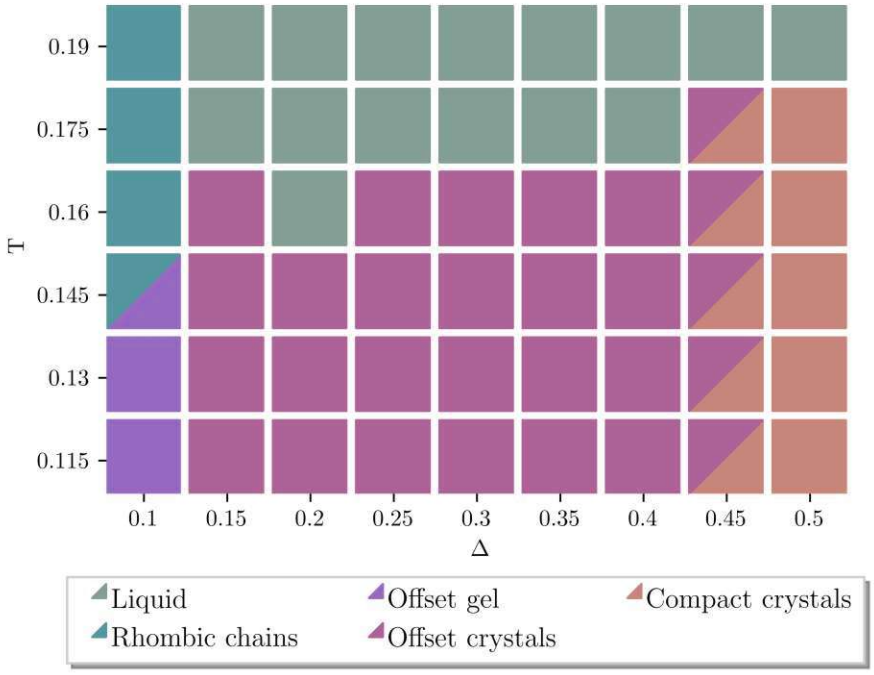


Figure 18: Visually classified Δ-T state diagram of the 2asym1c patch topology for $\phi = 0.3$. The patchy triangles show significantly different behavior at $\Delta = 0.1$ due to the patch spillover explained in figure 5. For high temperatures of $T \leq 0.145$ they form more stable rhombic chains, while below this threshold Δ results in the self assembly of offset gels. All other Δ possibilities can be found in 1 of two states: liquid or crystallized. Depending on Δ, these crystals form pores between the particles, which requires the classification of two types of crystals: close-packed crystals without pores, and offset crystals with pore sizes depending on Δ. When visual classification as one specific type of state is not obvious, systems are colored with both of the potential state colors found above.

Furthermore, it appears that Φ_4, the square positional order parameter, works best at identifying a difference between liquid and crystallized states. The average for $\phi = 0.3$ of this order parameter per T and Δ is shown in figure 19. Particle pairs or triplets occasionally found in liquids can be misleading in the interpretation of the diagram, since the positional order is only evaluated per cluster and not globally. This suggests square lattice structures where pairs of particles are found. In the realm of close-packed crystals the partially found square lattice structure disappears ($T = [0.115, 0.145]$ and $\Delta = [0.4, 0.5]$).

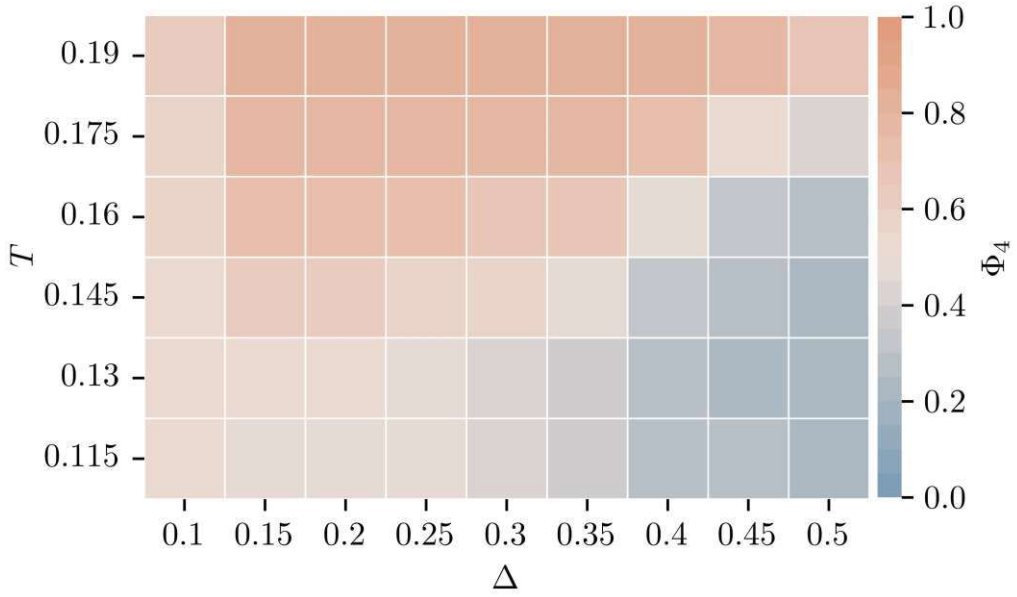


Figure 19: Average of the positional order parameter Φ_4 for each Δ-T combination of the 2asym1c patch topology with a particle density of $\phi = 0.3$. For systems that show no square lattice structure, this order parameter reaches 0 (blue). With an increasing appearance of a square lattice structure Φ_4 increases to 1 (orange). A more detailed description of the positional order parameters can be found in section 3.2.2.

The order parameter Φ_6 still successfully detects the difference between the gel states found at $\Delta = 0.1$ and all others, which is shown in figure 20. Additionally there is a slight difference found at the lower end of the Δ-T spectrum ($T = [0.115, 0.145]$ and $\Delta = [0.15, 0.25]$), where porous crystals can be found. As a positional order parameter it suffers from the same flaw as Φ_4, where it detects systems of small clusters as ones with hexagonal lattice structure.

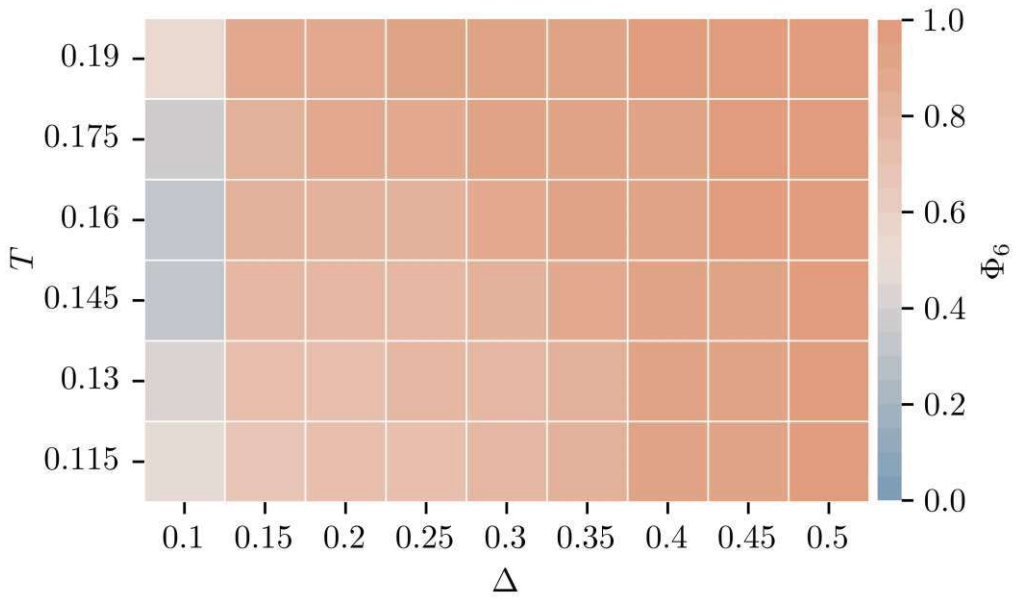


Figure 20: Average of the positional order parameter Φ_6 for each Δ-T combination of the 2asym1c patch topology with a particle density of $\phi = 0.3$. For systems that show no hexagonal lattice structure, this order parameter reaches 0 (blue). With an increasing appearance of a hexagonal lattice structure, Φ_6 increases to 1 (orange). A more detailed description of the positional order parameters can be found in section 3.2.2.

The orientational order parameter Φ_θ reacts to the higher degree of bond angle flexibility found at $\Delta = 0.1$, as seen in figure 21. This is a result of more than three patches forming a single bond combined with the patch spillover, leading to the formation of rhombic chains and offset gels, which show a higher variety of bond angles. The two particles connected via a central particle show a higher relative orientation towards each other, leading to triangles with orientations between anti-parallel and parallel orientations. Due to steric effects, this difference is not as significant at $\Delta = 0.1$ compared to crystals or liquids. Once the patches are more centered at the edges, the triangles of the 2asym1c patch topology connect exclusively in an anti-parallel manner. Therefore, Φ_θ is close to 0 for all other patch placements Δ in the 2asym1c patch topology.

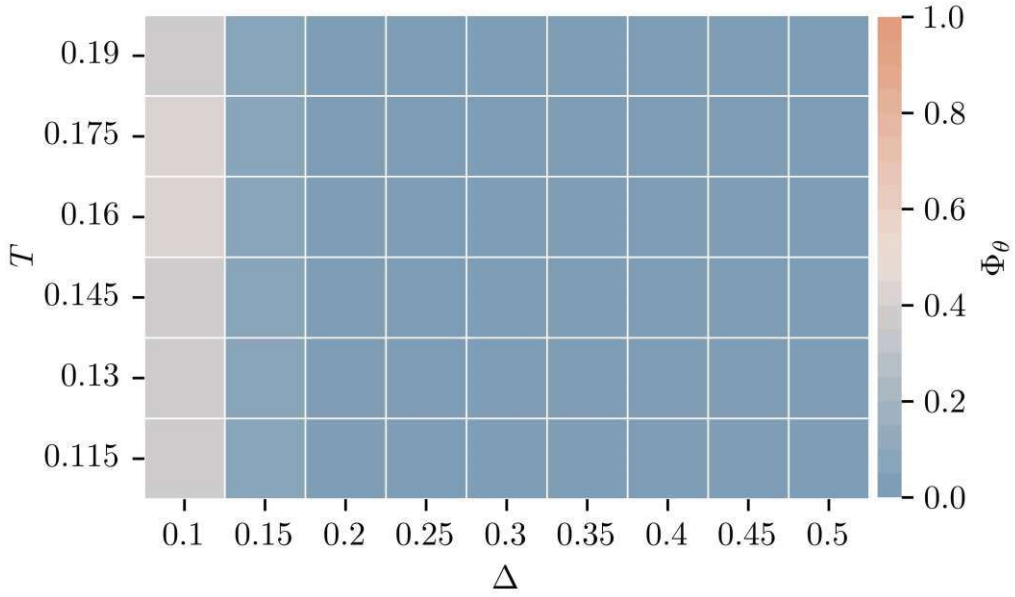


Figure 21: Average of the orientation order parameter Φ_θ for each Δ-T combination of the 2asym1c patch topology with a particle density of $\phi = 0.3$. For systems with mostly anti-parallel alignment of triangles,, this order parameter converges to 0 (blue). If the assemblies show more triangles facing the same direction, Φ_θ increases up to its limit of 1 (orange). A more detailed description of the orientational order parameter can be found in section 3.2.3.

To describe the systems found for the 2asym1c patch topology, it helps to study the overview of the porosity order parameter shown in figure 22. The 2asym1c patch topology shows porosity mostly in the offset gel state or in offset crystals. Thus, the highest porosity is found at $\Delta = 0.1$ and $T = 0.115$. With an increase in Δ, the porosity decreases as crystals become more compact. An increase in temperature leads to the formation of rhombic gels or a liquid state, which shows less porosity.

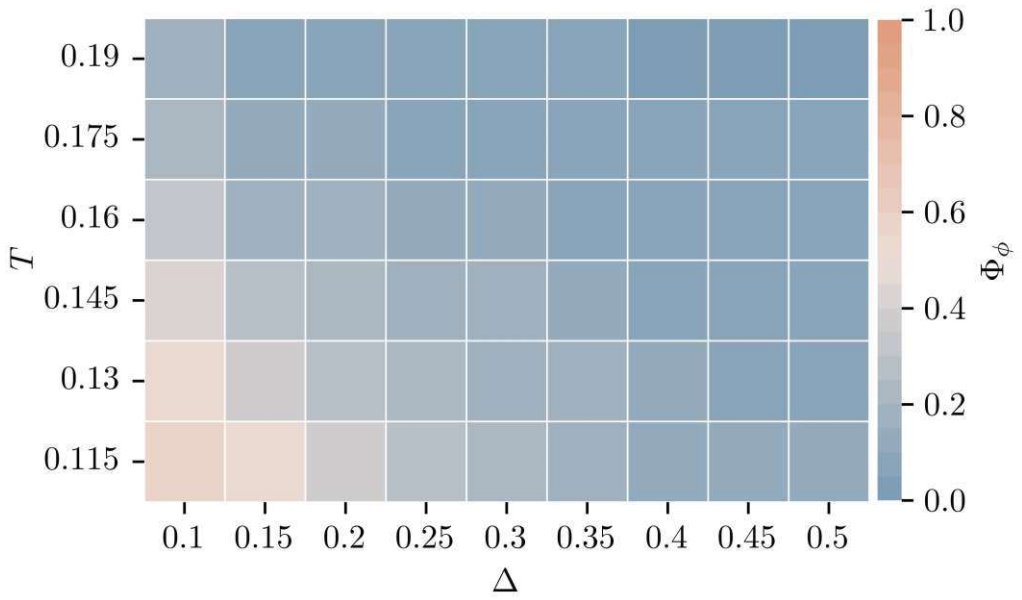


Figure 22: Average of the porosity order parameter Φ_ϕ for each Δ-T combination of the 2asym1c patch topology with a particle density of $\phi = 0.3$. This order parameter converges to 0 (blue) when there is no unoccupied space enclosed by connected triangles. The porosity order parameter is related to the largest pore space (space enclosed by connected particles) found within all systems. If the order parameter reaches 1 (orange), the system has the largest pore space out of all systems in this analysis. A more detailed description of the porosity order parameter can be found in section 3.2.4.

The last order parameter to be looked at for the 2asym1c patch topology is Φ_b, which describes the average amount of bonding partners per particle and is shown in figure 23. The lowest average number of bonds per particle is found for systems in a liquid state ($\Delta = [0.15, 0.5]$ and $T = [0.175, 0.19]$). As expected, the offset gels and rhombic chains at $\Delta = 0.1$ show more patches that are engaged in bonds. The highest number of bonded patches per particle is found for the crystallized systems at $\Delta = [0.15, 0.5]$ and $T = [0.115, 0.145]$.

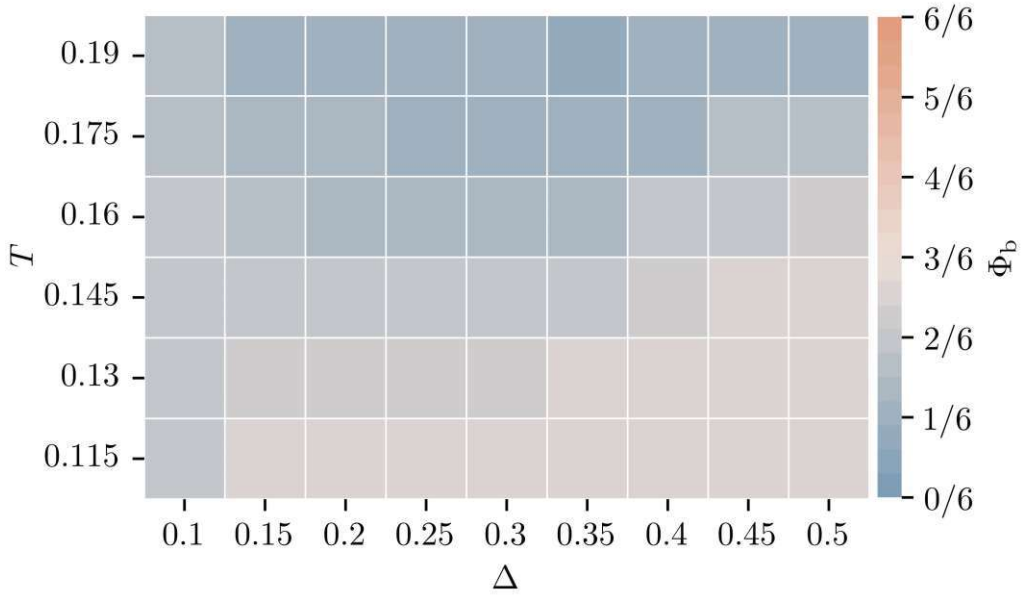
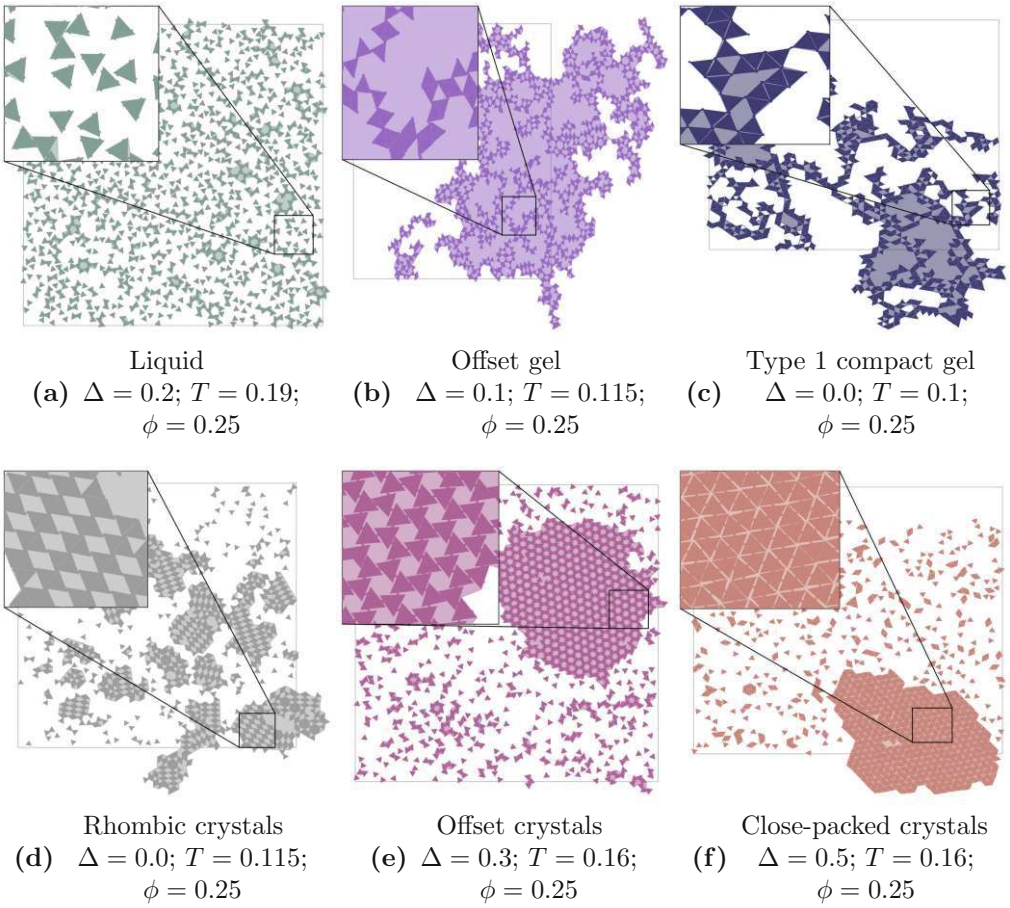


Figure 23: Average of the bond order parameter Φ_b for each Δ-T combination of the 2asym1c patch topology with a particle density of $\phi = 0.3$. The order parameter is 0 (blue) if there are no bonded patches and 1 (orange) if all particles contain six patches that are engaged in bonds. In the case of the 2asym1c patch topology, $\Phi_b = 3/6$ once all bonds are satisfied. The details for this choice are explained in section 3.2.1.

3.3.2 3asym

The different phases observed for the 3asym patch topology are shown in figure 24 and the corresponding state diagram for $\phi = 0.3$ reported in figure 25 shows similarities to the one of the 2asym1c patch topology. Similar to previously seen, liquids are found at the high end of the temperature interval at $T \geq 0.175$, as at relatively high temperatures bonds are not long-lived. For the 3asym patch topology, one can distinguish between offset and compact gels, where compact is used for gels where particles share most of their edges with adjacent triangles. The compact gel is the only phase that appears when all patches are placed at the vertices and is labeled type 1 compact gel, to remain distinguishable from the other compact gel types seen later on. The numbers in this naming scheme are chosen at random and carry no additional descriptive or hierarchical information. This geometry with $\Delta = 0$ allows for very compact growth and at this patch position these gels are energetically very favorable. The rhombic crystals and the offset gel observed at $\Delta = 0.1$ are similar to the ones observed in figure 17 as for this patch topology, the patch spillover occurs. With $T = 0.145$, there is a similar threshold

between rhombic structures and the offset gel found at temperatures below the threshold, while the former is found above. The 3asym patch topology also forms offset and close-packed crystals in a similar manner to the 2asym1c state diagram. Offset crystals are found for $0.15 \leq \Delta \leq 0.45$, which is the point where the size of the patch radius allows for the formation of close-packed crystals without any offset between their edges.



Liquid
(a) $\Delta = 0.2$; $T = 0.19$; $\phi = 0.25$

Offset gel
(b) $\Delta = 0.1$; $T = 0.115$; $\phi = 0.25$

Type 1 compact gel
(c) $\Delta = 0.0$; $T = 0.1$; $\phi = 0.25$

Rhombic crystals
(d) $\Delta = 0.0$; $T = 0.115$; $\phi = 0.25$

Offset crystals
(e) $\Delta = 0.3$; $T = 0.16$; $\phi = 0.25$

Close-packed crystals
(f) $\Delta = 0.5$; $T = 0.16$; $\phi = 0.25$

Figure 24: Observed phases of the 3asym patch topology. Liquids **(a)** only appear at higher temperatures where bonds are not long-living. One can distinguish between open **(b)** and compact **(c)** gels, depending on the relative orientation and distance of the particles. Crystals appear in three different possible structures: rhombic **(d)** where triangle pairs form rhombic particles which assemble into larger crystals, porous **(e)** and compact **(f)**.

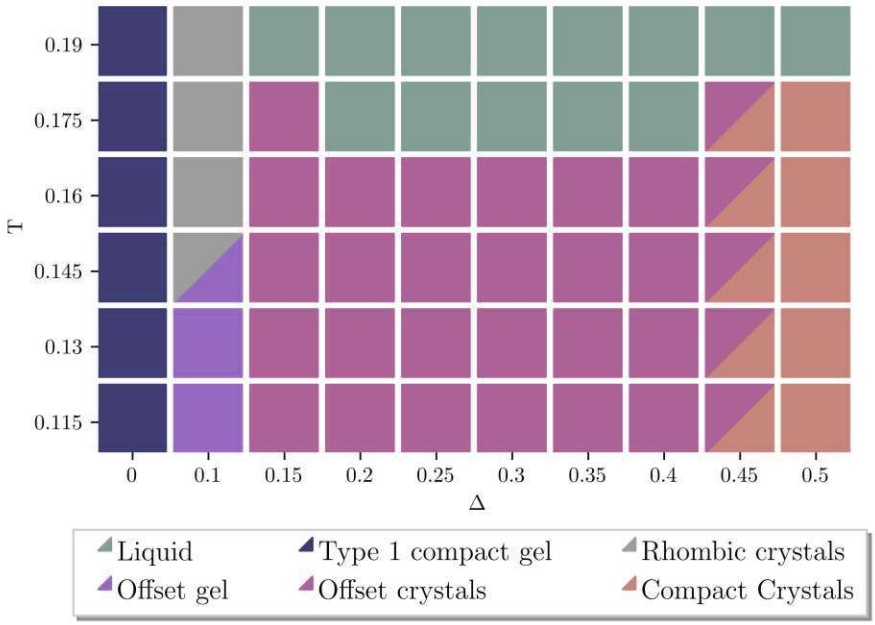


Figure 25: Visually classified Δ-T state diagram of the 3asym patch topology for $\phi = 0.3$. For $\Delta = 0$, triangles form a type 1 compact gel for all of the chosen temperatures. The patchy triangles show significantly different behavior at $\Delta = 0.1$ due to the patch spillover explained in figure 5. Rhombic crystals can be found from $T = 145$ and above, whereas offset gels are found below that temperature. All other possible Δ variations of the 3asym patch topology either self-assemble into crystals or remain in a liquid state due to the temperature. Close-packed crystals can only be found for $\Delta = 0.45$ and $\Delta = 0.5$, where the patch placement allows for the self-assembly of crystals without pores. Offset crystals can be found under suitable temperatures ($T \leq 0.175$) and Δ ($0.15 \leq \Delta \leq 0.45$). When visual classification as one specific type of state is not obvious, systems are colored with both of the potential state colors found above.

The Φ_4 order parameter seen in figure 26 detects square lattice order in the liquid regime of the 3asym patch topology. With a decrease in temperature, clusters start forming from the particle pairs responsible for the results of this order parameter. These larger structures don't adhere to a square lattice structure, leading to a decrease in Φ_4. Also, the crystals and gels found at $\Delta = 0.1$ show some square lattice structure. For compact structures such as the compact gel at $\Delta = 0$, close-packed crystals, as well as regular porous crystals, the positional order is far away from resembling a square lattice structure.

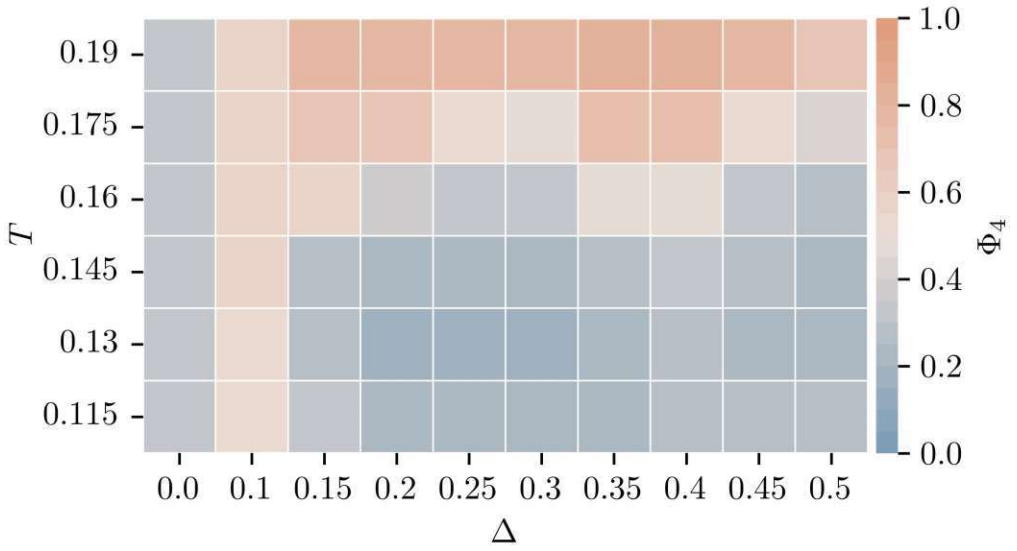


Figure 26: Average of the positional order parameter Φ_4 for each Δ-T combination of the 3asym patch topology with a particle density of $\phi = 0.3$. For systems that show no square lattice structure, this order parameter reaches 0 (blue). With an increasing appearance of a square lattice structure Φ_4 increases to 1 (orange). A more detailed description of the positional order parameters can be found in section 3.2.2.

The particle pairs found in liquids, as well as the crystal structures which self-assembled using the 3asym topology, all show hexagonal lattice structure, as seen in figure 27. The only exceptions are the rhombic crystals which can be found at $\Delta = 0.1$. These, as well as gel types of the 3asym patch topology, show little hexagonal lattice structure.

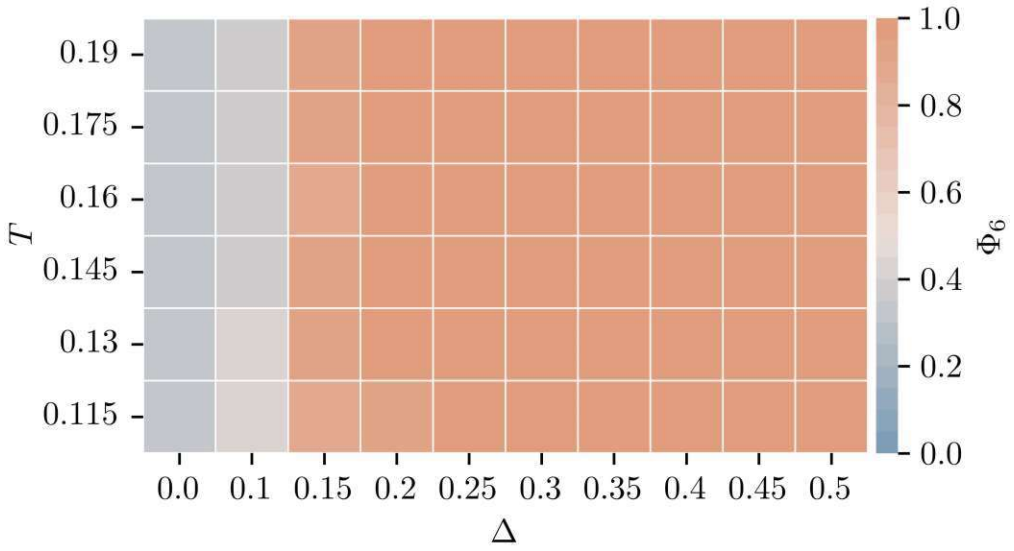


Figure 27: Average of the positional order parameter Φ_6 for each Δ-T combination of the 3asym patch topology with a particle density of $\phi = 0.3$. For systems that show no hexagonal lattice structure, this order parameter reaches 0 (blue). With an increasing appearance of a hexagonal lattice structure Φ_6 increases to 1 (orange). A more detailed description of the positional order parameters can be found in section 3.2.2.

In figure 28 the orientational order is shown for the gels and rhombic crystals found at $\Delta = 0$ and $\Delta = 0.1$. The small clusters found in liquids, as well as the crystal structures, show anti-parallel orientational order due to steric effects.

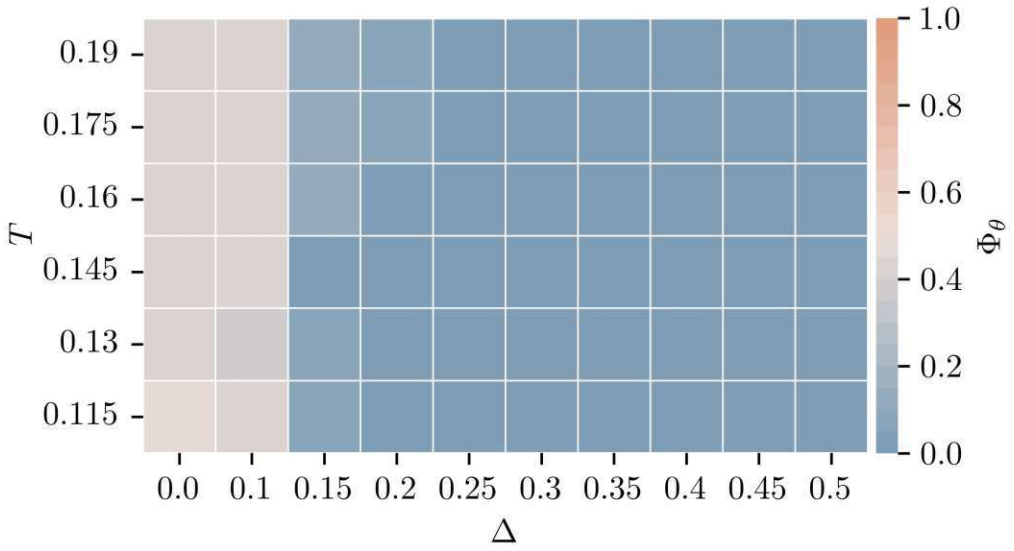


Figure 28: Average of the orientation order parameter Φ_θ for each Δ-T combination of the 3asym patch topology with a particle density of $\phi = 0.3$. For systems with mostly anti-parallel alignment of triangles, this order parameter converges to 0 (blue). If the assemblies show more triangles facing the same direction, Φ_θ increases up to its limit of 1 (orange). A more detailed description of the orientational order parameter can be found in section 3.2.3.

The porosity order parameter for the 3asym patch topology is shown in figure 29. The gels show a higher porosity at lower temperatures $T = [0.115, 0.145]$. In crystal structures, the order parameter correctly reflects the correlation between porosity and patch placement. The offset crystals with $\Delta = 0.2$ are recognized as more porous than the close-packed crystals found at $\Delta = 0.5$.

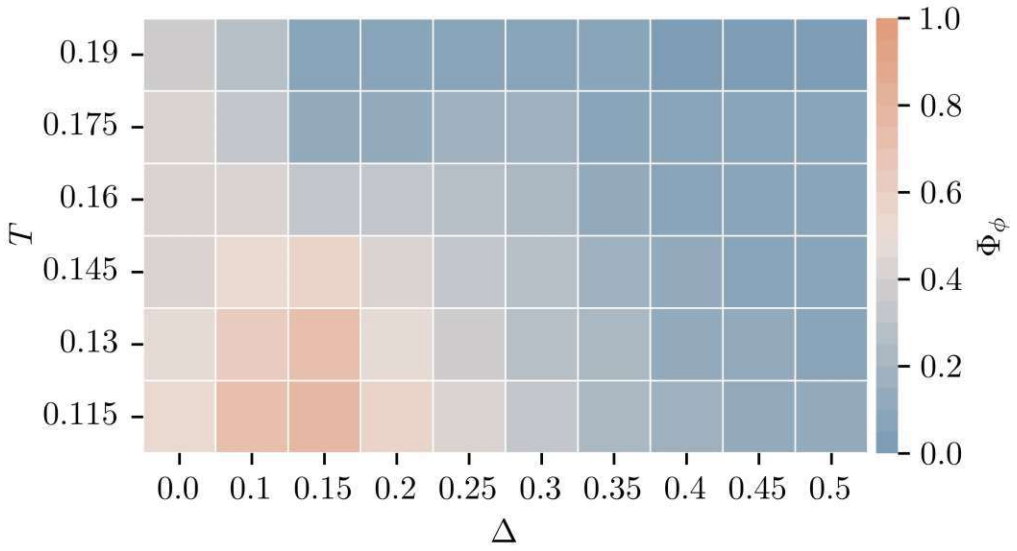


Figure 29: Average of the porosity order parameter Φ_ϕ for each Δ-T combination of the 3asym patch topology with a particle density of $\phi = 0.3$. This order parameter converges to 0 (blue) when there is no unoccupied space enclosed by connected triangles. The porosity order parameter is put into relation to the largest pore space (space enclosed by connected particles) found within all systems. If the order parameter reaches 1 (orange), the system has the largest pore space out of all systems in this analysis. A more detailed description of the orientational order parameter can be found in section 3.2.4.

In figure 30 provides insight into the bonding behavior of the 3asym patch topology for different Δ-T combinations. As expected, the systems found in the liquid state at $\Delta = [0.15, 0.5]$ and $T = [0.175, 0.19]$, contain fewer patches in bonds per particle. The type 1 compact gel found at $\Delta = 0$ contains more bonded patches per particle than the offset gel and rhombic chains found at $\Delta = 0.1$. An interesting anomaly is the higher number of bonded patches per particle found at $\Delta = [0.15, 0.3]$ and $T = [0.115, 0.13]$ compared to crystals of higher Δ. A short investigation led to the conclusion that this is the result of a higher surface to boundary size found for the porous crystals.

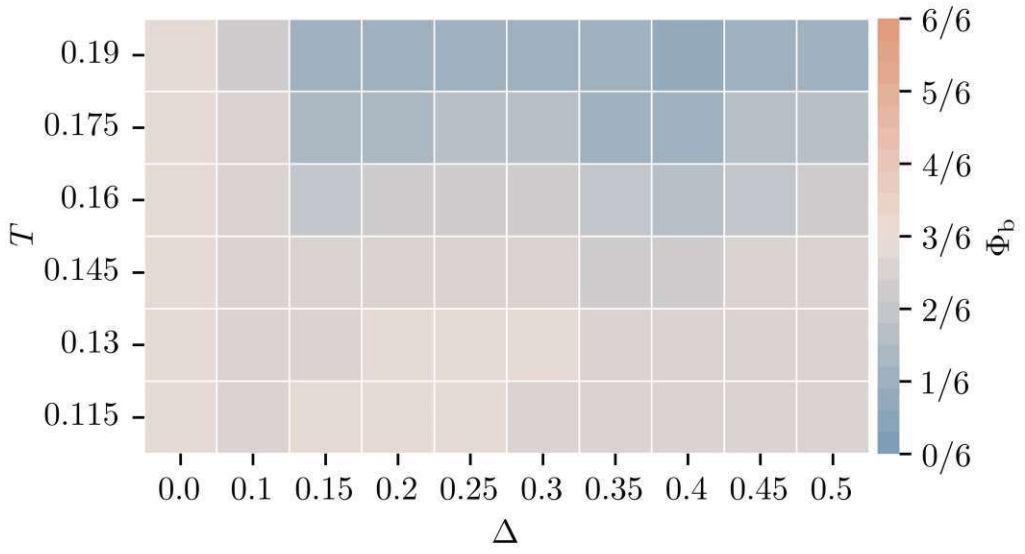
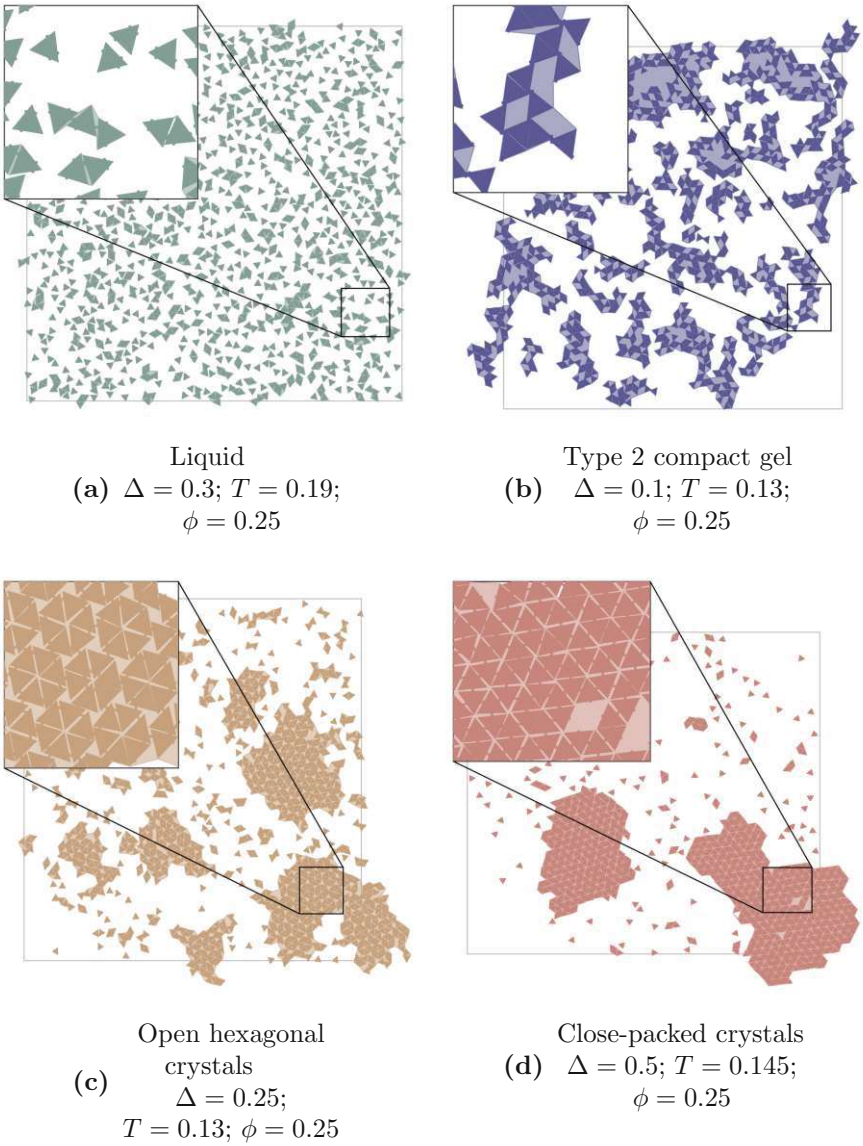


Figure 30: Average of the bond order parameter Φ_b for each Δ-T combination of the 3asym patch topology with a particle density of $\phi = 0.3$. The order parameter is 0 (blue) if there are no bonded patches and 1 (orange) if all particles contain six patches that are engaged in bonds. In the case of the 3asym patch topology, $\Phi_b = 3/6$ once all bonds are satisfied. The details for this choice are explained in section 3.2.1

3.3.3 Mouse

The mouse patch topology shows some slight variations of previously discussed phases in figure 31. Visually different from previously seen gels, the type 2 compact gel is expected to behave similarly in the order parameter analysis compared to other types of compact gels. Furthermore, the offset crystals show a unique hexagonal substructure and are therefore labeled open hexagonal crystals. When studying the state diagram in figure 32 for $\phi = 0.3$, it becomes clearer why the comparisons are made to other patch topologies and their state diagrams. Even with slight variations in behavior compared to other topologies, the state diagram closely resembles the ones for the 2asym1c and 3asym patch state diagrams, figure 17 and figure 24. Also in this patch topology, the effects of the patch spillover can be observed at $\Delta = 0.1$.



(a) Liquid $\Delta = 0.3$; $T = 0.19$; $\phi = 0.25$

(b) Type 2 compact gel $\Delta = 0.1$; $T = 0.13$; $\phi = 0.25$

(c) Open hexagonal crystals $\Delta = 0.25$; $T = 0.13$; $\phi = 0.25$

(d) Close-packed crystals $\Delta = 0.5$; $T = 0.145$; $\phi = 0.25$

Figure 31: Some phases of the mouse patch topology appear as slight variations of previously discussed phases. A liquid phase **(a)** can be found at higher temperatures, where the bonds are not stable. The type 2 compact gel **(b)** and open hexagonal crystals **(c)** are new variations of possible gel or crystal phases. For $\Delta \leq 0.45$, the hexagonal substructure disappears in crystals and the triangles self-assemble into a close-packed crystal phase.

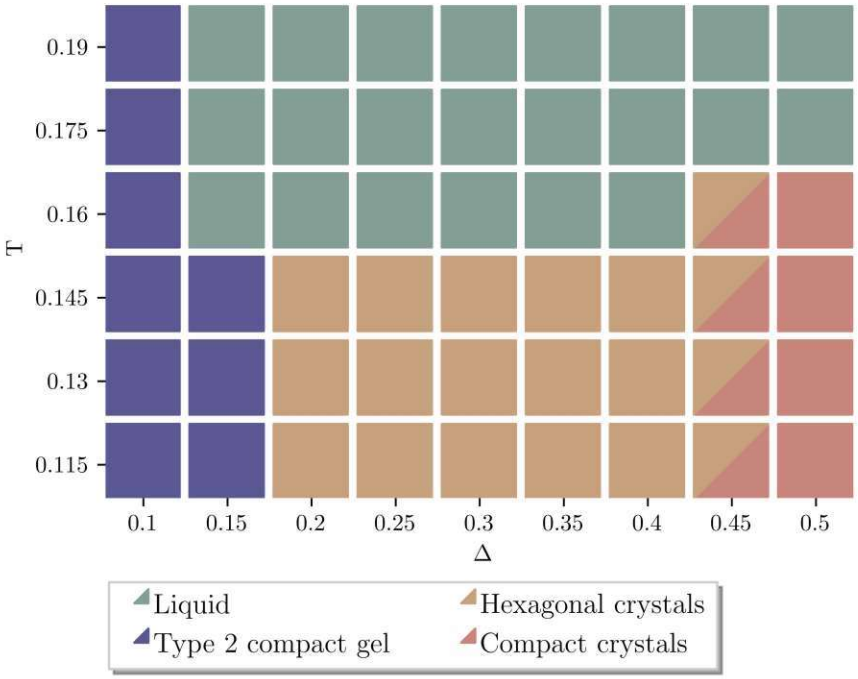


Figure 32: The visually classified Δ-T state diagram for $\phi = 0.25$ of the mouse patch topology is similar to the 2asym1c or 3asym state diagram, but shows a less rich collection of phases in the chosen Δ-T range. For $\Delta = 0.1$ and $\Delta = 0.15$ systems are found as type 2 compact gels. From $\Delta = 0.15$ on, temperatures $T \leq 0.16$ result in systems that remain in their liquid state. When possible crystals are of the close-packed crystal type ($\Delta = 0.45$ and $\Delta = 0.5$), otherwise they appear with hexagonal substructure and are categorized as open hexagonal crystals.

The occasional particle pairs in liquids lead to higher values of Φ_4 which describes the appearance of cubic lattice structures. This structure only disappears in the close-packed crystal regime ($T = [0.115, 0.145]$ and $\Delta = [0.45, 0.5]$). For the gels at $\Delta = 0.1$ it appears even less than in liquid or open hexagonal crystal states.

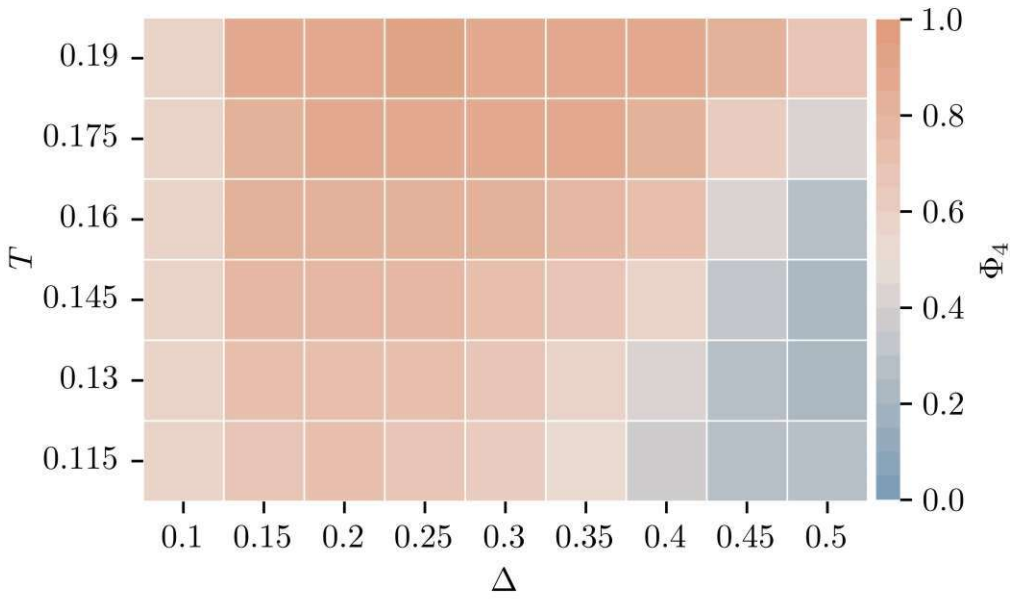


Figure 33: Average of the positional order parameter Φ_4 for each Δ-T combination of the mouse patch topology with a particle density of $\phi = 0.3$. For systems that show no square lattice structure, this order parameter reaches 0 (blue). With an increasing appearance of a square lattice structure Φ_4 increases to 1 (orange). A more detailed description of the positional order parameters can be found in section 3.2.2.

The order parameter Φ_6 reveals the appearance of open hexagonal crystal substructures at $T = [0.115, 0.145]$ and $\Delta = [0.25, 0.35]$ in figure 34. These small and local hexagonal substructures result in defects in terms of the global hexagonal structure, which leads to a decrease in Φ_6. At $\Delta = 0.1$ the 3asym patch topology assembles into type 2 compact gels, which also show less hexagonal lattice structure. The transition at $\Delta = [0.15, 0.2]$ and $T = [0.115, 0.145]$ from type 2 compact gels to hexagonal clusters results in mixed systems with higher hexagonal lattice structure. Once the hexagonal substructure starts disappearing with the increase in $\Delta > 0.35$, the global hexagonal lattice structure described by Φ_6 increases. As discussed previously, the few small clusters found in liquid systems at $T = [0.16, 0.19]$ are misleading for this order parameter, since they show hexagonal lattice structure from a very small scale.

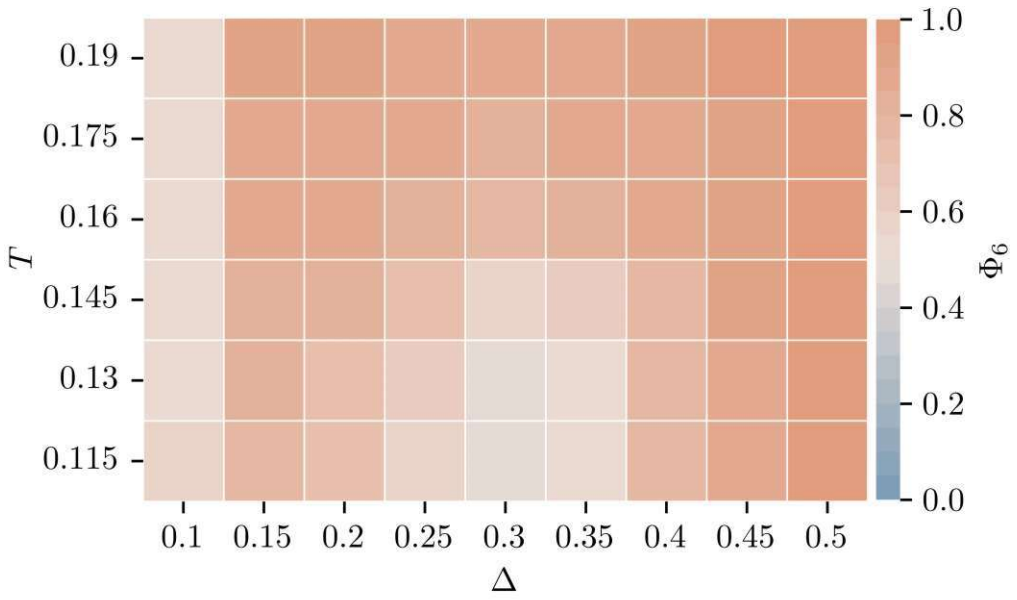


Figure 34: Average of the positional order parameter Φ_6 for each Δ-T combination of the mouse patch topology with a particle density of $\phi = 0.3$. For systems that show no hexagonal lattice structure, this order parameter reaches 0 (blue). With an increasing appearance of a hexagonal lattice structure Φ_6 increases to 1 (orange). A more detailed description of the positional order parameters can be found in section 3.2.2.

The orientation order parameter shown in figure 35 behaves similarly to the other patch topologies containing 3 patches. The type 2 compact gel found at $\Delta = 0.1$ allows for different bond angles due to the patch spillover. All other patch placements Δ result in anti-parallel orientation due to steric effects. This is true even for the small clusters found in the liquid states at $T = 0.16$ and above.

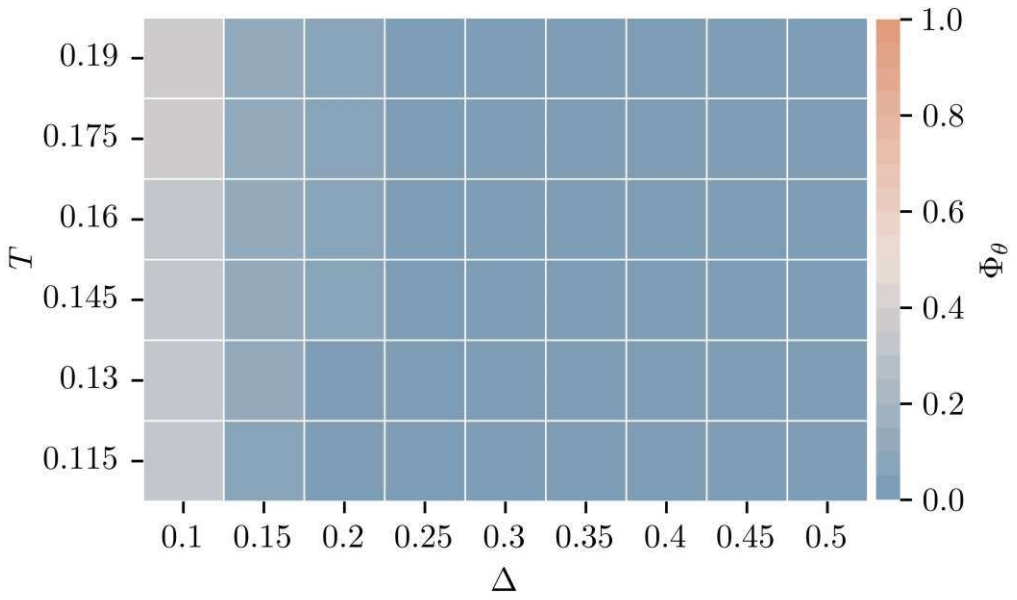


Figure 35: Average of the orientation order parameter Φ_θ for each Δ-T combination of the mouse patch topology with a particle density of $\phi = 0.3$. For systems with mostly anti-parallel alignment of triangles, this order parameter converges to 0 (blue). If the assemblies show more triangles facing the same direction, Φ_θ increases up to its limit of 1 (orange). A more detailed description of the orientational order parameter can be found in section 3.2.3.

The hexagonal substructures which can be found in the mouse topology show smaller pore sizes than the offset crystals seen before. This is shown in figure 36, by the lack of a significant increase in Φ_ϕ at $\Delta = [0.15, 0.4]$ and $T = [0.115, 0.145]$, which is observable in the previous patch topologies for offset crystals. At $\Delta = 0.1$ there is a slight increase in porosity, but the type 2 compact gels are not the most porous gel variation.

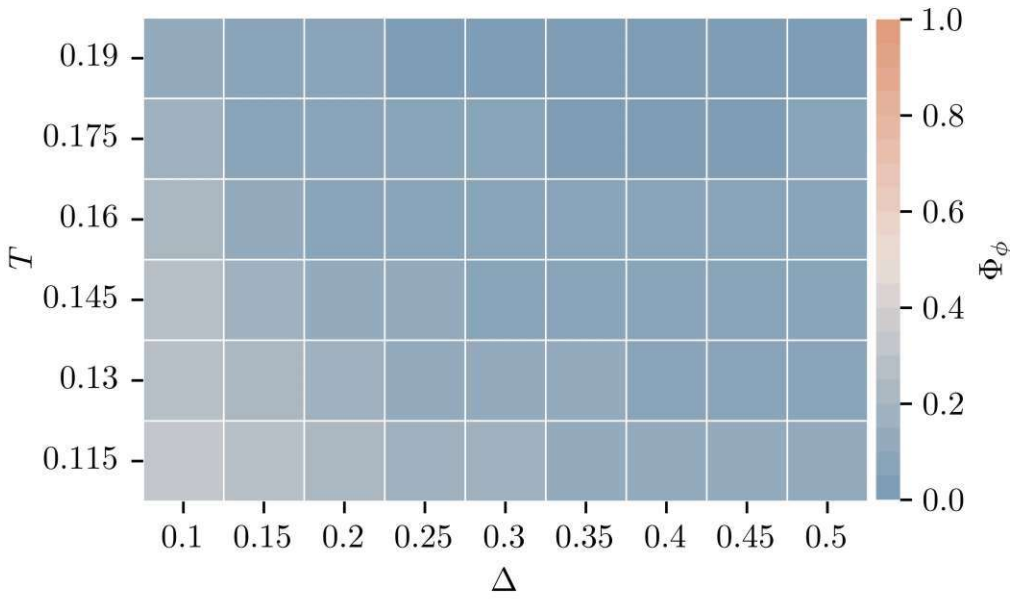


Figure 36: Average of the porosity order parameter Φ_ϕ for each Δ-T combination of the mouse patch topology with a particle density of $\phi = 0.3$. This order parameter converges to 0 (blue) when there is no unoccupied space enclosed by connected triangles. The porosity order parameter is put into relation to the largest pore space (space enclosed by connected particles) found within all systems. If the order parameter reaches 1 (orange), the system has the largest pore space out of all systems in this analysis. A more detailed description of the orientational order parameter can be found in section 3.2.4.

The number of bonded patches per particle is the last order parameter discussed for the mouse patch topology and shown in figure 37. The type 2 compact gels found at $\Delta = 0.1$ show a similar number of bonded patches per particle as the crystals found at $\Delta = [0.4, 0.5]$ and $T = [0.115, 0.13]$. At the transition boundary from gel to crystal around $\Delta = [0.15, 0.2]$ and $T = [0.115, 0.145]$ there is a slight drop in bonded patches per particle. With an increase in temperature, this number is further decreased until systems reach a liquid state at $\Delta = [0.15, 0.5]$ and $T = [0.16, 0.19]$.

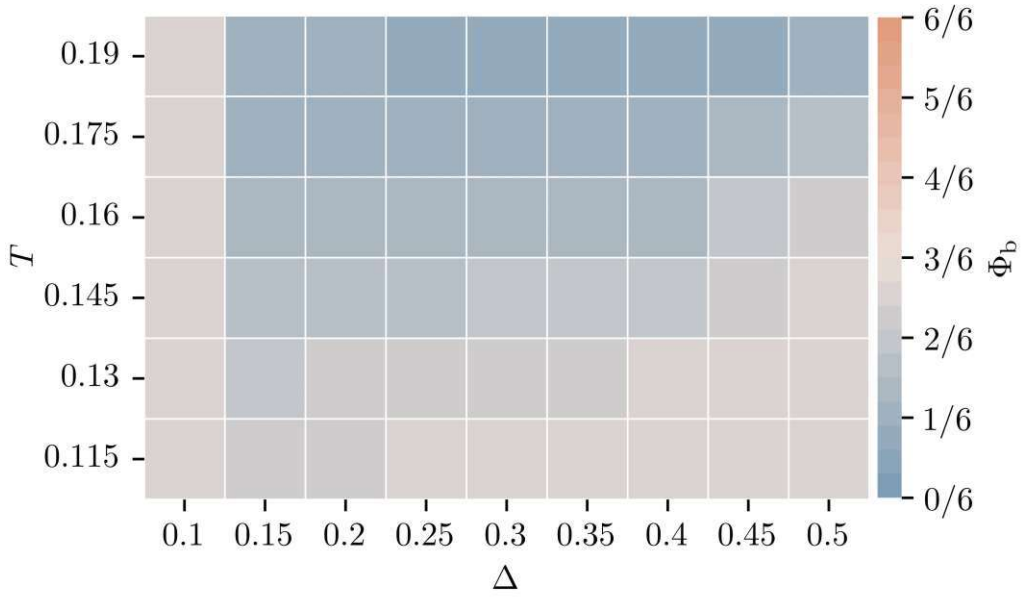
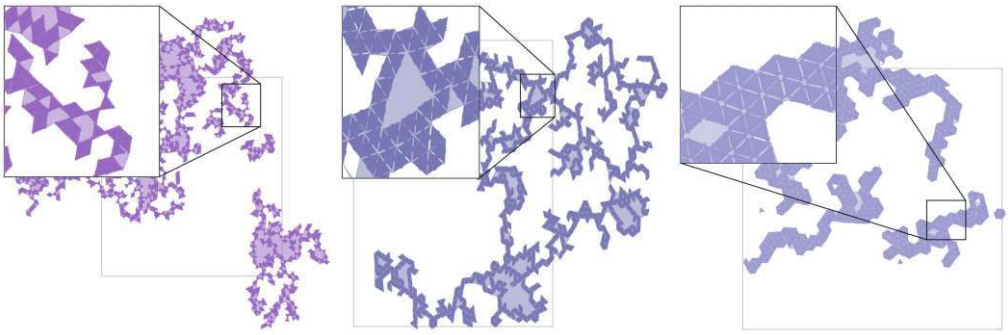


Figure 37: Average of the bond order parameter Φ_b for each Δ-T combination of the mouse patch topology with a particle density of $\phi = 0.3$. The order parameter is 0 (blue) if there are no bonded patches and 1 (orange) if all particles contain six patches that are engaged in bonds. In the case of the mouse patch topology, $\Phi_b = 3/6$ once all bonds are satisfied. The details for this choice are explained in section 3.2.1

3.3.4 6patch

In the selected temperature range the 6patch patch topology only appears in 3 phases: offset gel, type 3 compact gel and type 4 compact gel, which are all displayed in figure 38. Figure 39 displays the state diagram of this patch topology for $\phi = 0.3$. A temperature above $T = 0.145$ can lead to a more compact self-assembly and therefore one can differentiate between the resulting type 3 and type 4 compact gels. The offset gel only appears for $\Delta = 0.1$, due to the patch spill over explained in figure 5. This time both patches close to a vertex can bond with the same patch of a different particle. Offset and type 3 compact gels show a low number of average bonds per particle compared to type 4 compact gels. Due to the symmetry of the 6patch topology, particles are more inclined to form compact connections instead of offset ones.



Offset gel
(a) $\Delta = 0.1$; $T = 0.19$; $\phi = 0.25$

Type 3 compact gel
(b) $\Delta = 0.45$; $T = 0.19$; $\phi = 0.25$

Type 4 compact gel
(c) $\Delta = 0.15$; $T = 0.19$; $\phi = 0.25$

Figure 38: The three different phases found in the 6patch patch topology. Offset gel **(a)**, type 3 compact gel **(b)** and type 4 compact gel **(c)**. The offset gels show a less ordered structure since patches are placed closer to a vertex. This allows the patches along the same edge to bond with patches of different particles, which results in gaps being formed. The type 3 compact gel forms compact branches of particles with two neighbors. Type 4 compact gels are a mix between gel and crystals states. They show multiple branched out close-packed clusters.

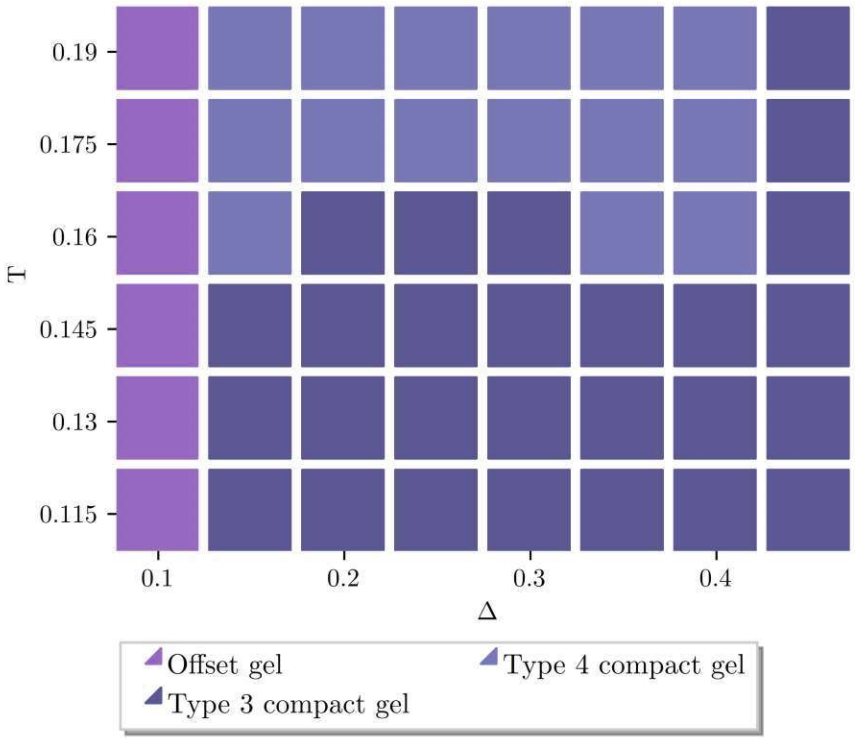


Figure 39: The visually classified Δ-T state diagram of the 6patch patch topology for $\phi = 0.3$. Offset gels only appear at $\Delta = 0.1$ and are a result of the patch spillover explained in figure 5. The temperature T appears to be the relevant parameter to separate type 3 from type 4 compact gels at this temperature range.

The average positional order parameter Φ_4 shown in figure 39 shows unexpected results. The lattice structure for the compact gels at $\Delta = 0.2$ and $\Delta = 0.25$ appears to be significantly different, especially at $T < 0.145$. A closer visual examination of those systems revealed a higher occurrence of triangles that have bonded with two different triangles at the same edge due to the patch position. These irregularities disappear for $\Delta = 0.1$ where a more compact structure appears to be more favorable. The type 4 compact gels found at $\Delta = [0.15, 0.4]$ and $T = [0.16, 0.19]$ have a slight increase in thickness compared to the type 3 compact gels. This comes with a decrease in square lattice structure, since those close-packed areas show hexagonal lattice structure. At $\Delta = 0.1$ the 6patch patch topology assembles into offset gels, which are recognized to have a higher square lattice structure. This is presumably a result of the disappearance of larger close-packed sections and a rather porous structure.

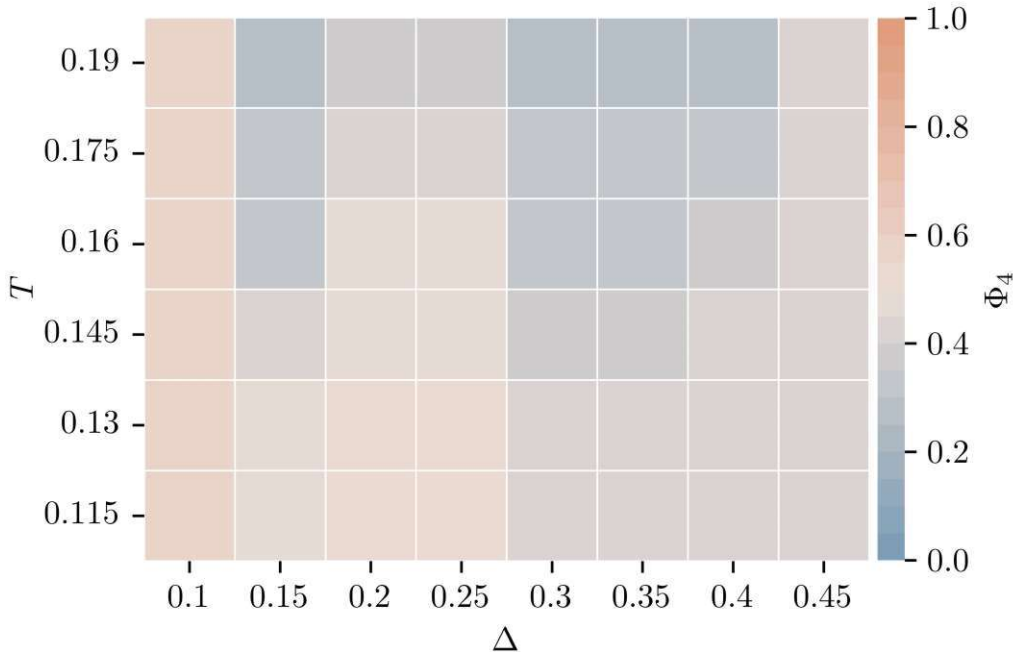


Figure 40: Average of the positional order parameter Φ_4 for each Δ-T combination of the 6patch patch topology with a particle density of $\phi = 0.3$. For systems that show no square lattice structure, this order parameter reaches 0 (blue). With an increasing appearance of a square lattice structure Φ_4 increases to 1 (orange). A more detailed description of the positional order parameters can be found in section 3.2.2.

With the increase in square lattice structure observed for the offset gels found at $\Delta = 0.1$ comes a decrease in hexagonal lattice structure, as seen in figure 41. This correlation does not translate into the observed differences in hexagonal lattice structure described by Φ_6 between the type 3 and type 4 compact gels. Only for $\Delta = 0.2$ and $\Delta = 0.25$ a decrease is observed Φ_6 for all temperatures. As explained above, these systems show a higher occurrence of triangles with two different bonded neighbors along the same edge. These defects could result in a decrease in Φ_6.

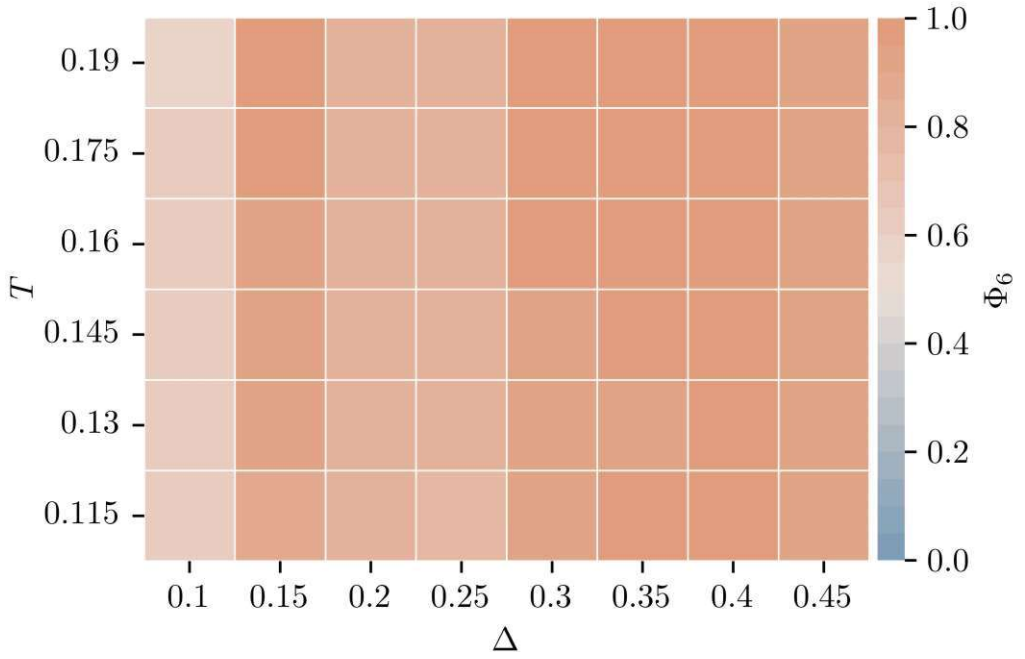


Figure 41: Average of the positional order parameter Φ_6 for each Δ-T combination of the 6patch patch topology with a particle density of $\phi = 0.3$. For systems that show no hexagonal lattice structure, this order parameter reaches 0 (blue). With an increasing appearance of a hexagonal lattice structure Φ_6 increases to 1 (orange). A more detailed description of the positional order parameters can be found in section 3.2.2.

The orientational order parameter Φ_θ responds to the bond flexibility, which occurs with a patch placement of $\Delta = 0.1$ and can be seen in figure 42. While this patch placement allows a wider range of bond angles, all other Δ result in anti-parallel orientation, as expected for close-packed triangles.

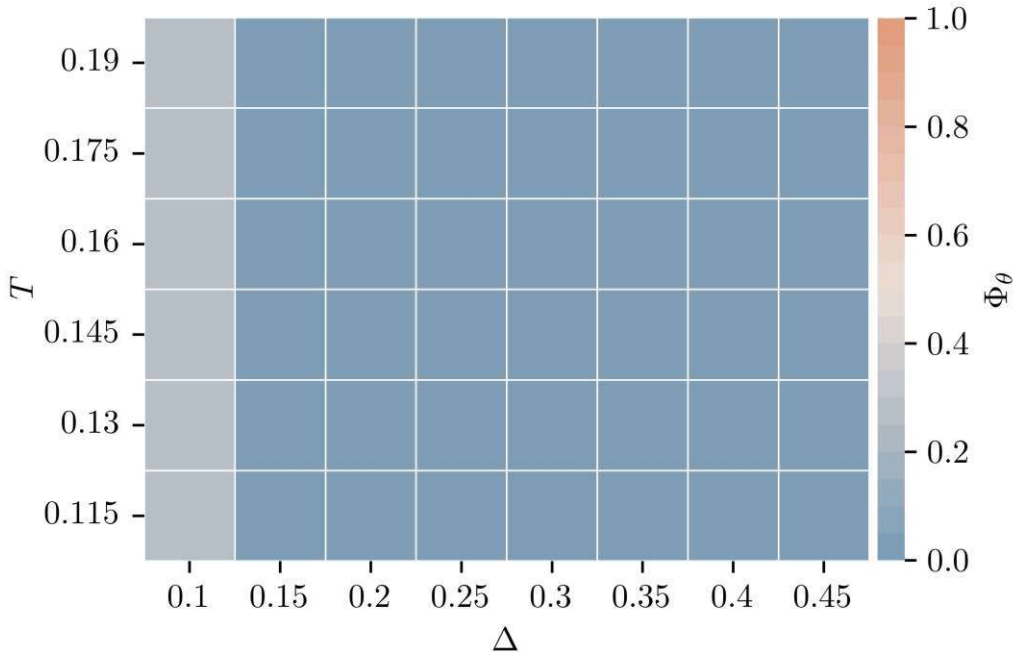


Figure 42: Average of the orientation order parameter Φ_θ for each Δ-T combination of the 6patch patch topology with a particle density of $\phi = 0.3$. For systems with mostly anti-parallel alignment of triangles, this order parameter converges to 0 (blue). If the assemblies show more triangles facing the same direction, Φ_θ increases up to its limit of 1 (orange). A more detailed description of the orientational order parameter can be found in section 3.2.3.

The porosity order parameter Φ_ϕ is shown in figure 43 and describes the space enclosed by connected particles. This chosen definition of Φ_ϕ does not properly detect the behavior at $\Delta = 0.2$ and $\Delta = 0.25$ described above. These small and still somewhat rare pores created by those defects do not carry enough weight when the space enclosed by connected particles is used as the only contributor in the porosity order parameter. Otherwise it successfully detects the differences between the type 4 compact gels found at $\Delta = [0.15, 0.4]$ and $T = [0.16, 0.19]$ and the type 3 compact gels otherwise found for $\Delta > 0.1$. The offset gels at $\Delta = 0.1$ are also recognized as more porous than the other structures.

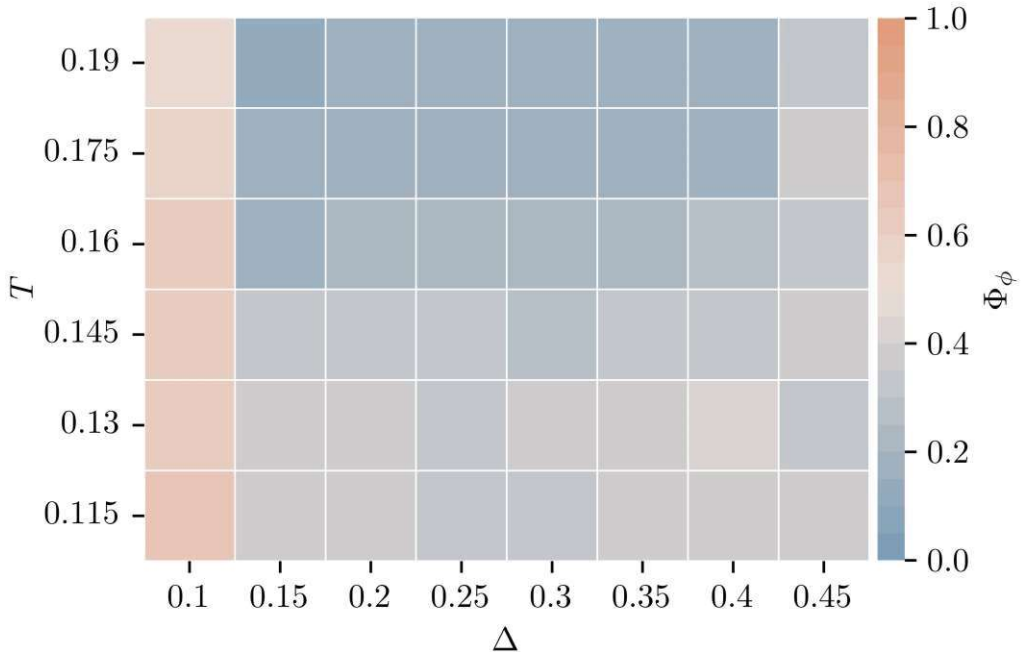


Figure 43: Average of the porosity order parameter Φ_ϕ for each Δ-T combination of the 6patch patch topology with a particle density of $\phi = 0.3$. This order parameter converges to 0 (blue) when there is no unoccupied space enclosed by connected triangles. The porosity order parameter is put into related to the largest pore space (space enclosed by connected particles) found within all systems. If the order parameter reaches 1 (orange), the system has the largest pore space out of all systems in this analysis. A more detailed description of the porosity order parameter can be found in section 3.2.4.

The bond order parameter Φ_b seen in figure 44 appears to be the weakest order parameter to classify the structures found within the 6patch patch topology. For the type 4 compact gels found at $\Delta = [0.15, 0.4]$ and $T = [0.16, 0.19]$ it detects the highest number of bonded patches per particle. Between the offset gel and type 3 compact gels the difference seems to be negligible.

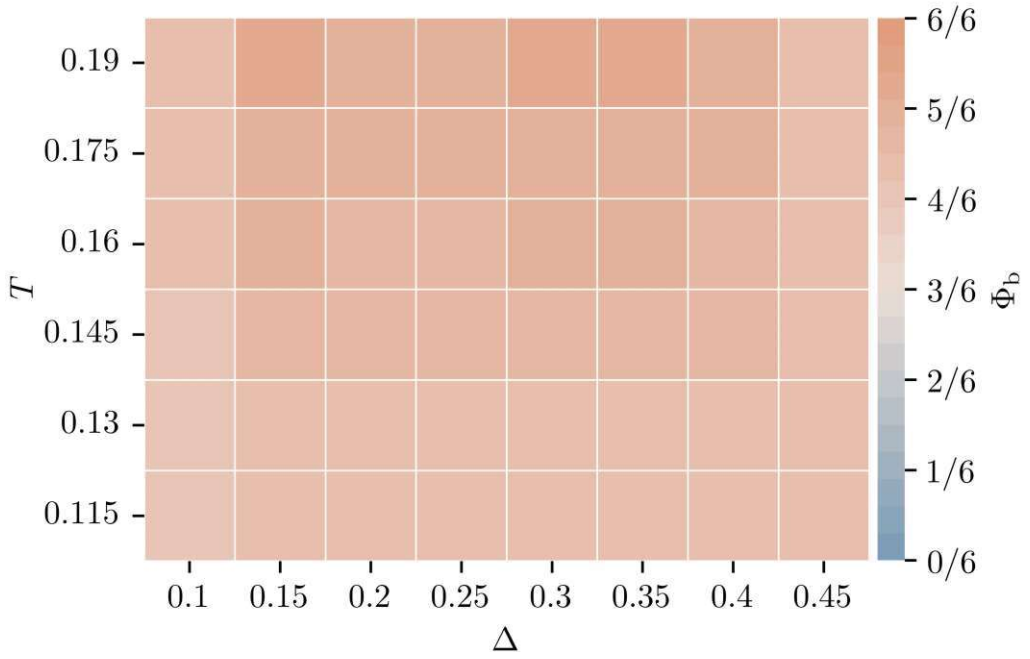
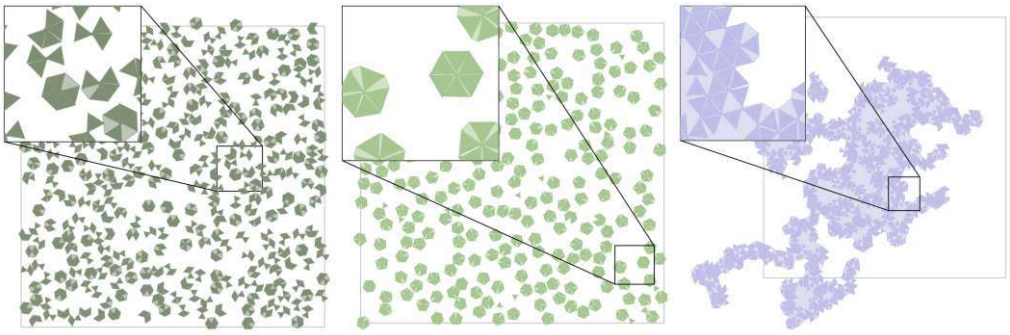


Figure 44: Average of the bond order parameter Φ_b for each Δ-T combination of the 6patch patch topology with a particle density of $\phi = 0.3$. The order parameter is 0 (blue) if there are no bonded patches and 1 (orange) if all particles contain six patches that are engaged in bonds. The details for this choice are explained in section 3.2.1

3.3.5 VN

The VN patch topology introduces liquids consisting of small clusters, which is seen in figure 45. These liquids are present across most of the investigated parameter space. At $\Delta = 0.1$ and $\Delta = 0.15$ these small clusters consist mostly of triangle triplets and hexagonal clusters. The triplets appear when triangles bond at the edge patch, whereas the hexagons are a collection of triangles that have bonded with their vertex patch. For liquids with $\Delta \geq 0.2$, the triangle triplets disappear and the liquids consist mostly of hexagonal clusters. To separate it from this second liquid type, the first one is labeled small-cluster liquid and the second hexagonal liquid. As soon as $\Delta \geq 0.9$, the hexagonal liquid transitions into a type 5 compact gel phase, as seen in figure 46, since the patch position enables the formation of larger structures.



(a) Small-cluster liquid $\Delta = 0.1$; $T = 0.19$; $\phi = 0.25$

(b) Hexagonal liquid $\Delta = 0.9$; $T = 0.19$; $\phi = 0.25$

(c) Type 5 compact gel $\Delta = 1.0$; $T = 0.19$; $\phi = 0.25$

Figure 45: The VN patch topology results in three different phases: small-cluster liquids **(a)**, hexagonal liquids **(b)**, and type 5 compact gels **(c)**. When six triangles are connected via their vertex patch at their center, they form isolated hexagons. Most clusters found in hexagonal liquids only consist of five to six particles, whereas in small-cluster liquids clusters are as small as three triangles. Unlike the small clusters of size five or six, those triplets are connected not just via their vertex patch, but also their edge patch. Type 5 compact gels are a new variety of gels with a mix of highly flexible bonds between two particles and compact structures. This is enabled by the vertex patch of the VN topology.

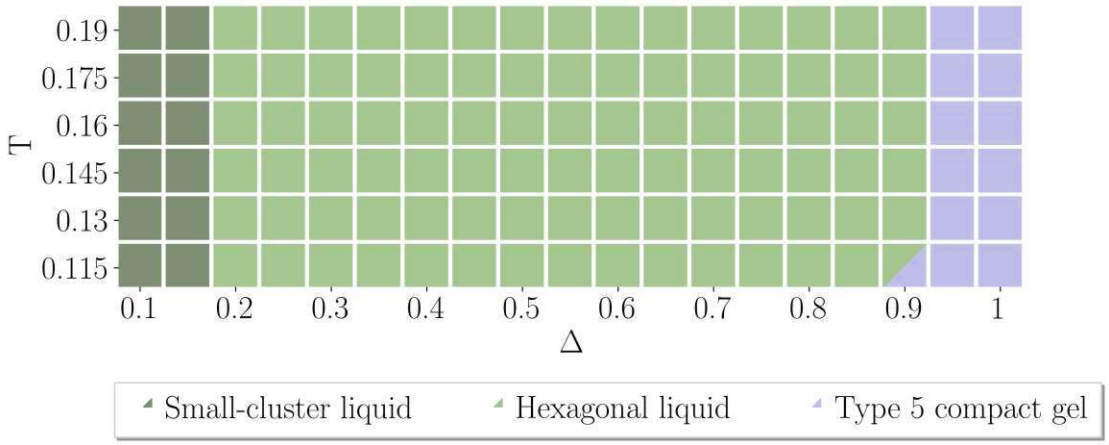


Figure 46: The state diagram for the VN patch topology at $\phi = 0.3$ shows that gels only become energetically favorable once the edge-patch is placed close to the second vertex. Otherwise the triangles form small clusters. At $\Delta = 0.1$ and $\Delta = 0.15$ these clusters are of size 3 and above. The hexagonal liquids found between $\Delta = 0.2$ and $\Delta = 0.9$ show cluster sizes of five or six particles.

The positional order parameter Φ_4 responsible for detecting square lattice structures is shown in figure 47. Within the small clusters found in the hexagonal liquid, Φ_4 detects some square lattice structure. This is probably a result of the small sample size within each cluster. For the large clusters of the type 5 compact gels found at $\Delta = 0.95$ and $\Delta = 1.0$, Φ_4 decreases substantially. Furthermore, it indicates the small-cluster liquid found at $\Delta = 0.1$ and $\Delta = 0.15$.

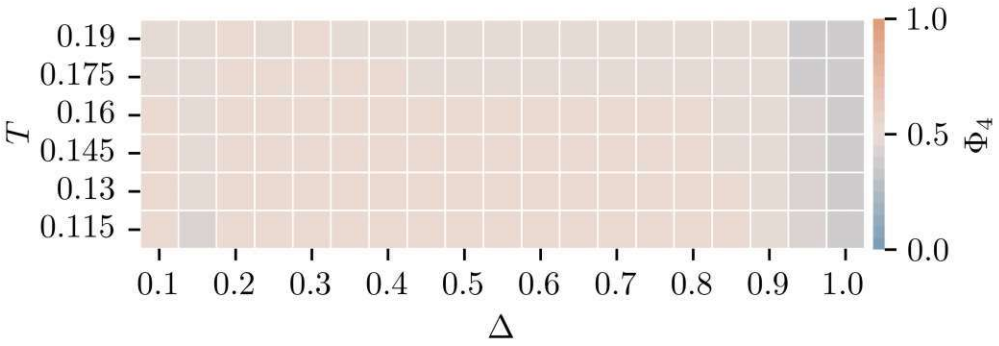


Figure 47: Average of the positional order parameter Φ_4 for each Δ-T combination of the VN patch topology with a particle density of $\phi = 0.3$. For systems that show no square lattice structure, this order parameter reaches 0 (blue). With an increasing appearance of a square lattice structure Φ_4 increases to 1 (orange). A more detailed description of the positional order parameters can be found in section 3.2.2.

The small-cluster liquid found at $\Delta = 0.1$ and $\Delta = 0.15$ is more distinctly identified through Φ_6 as seen in figure 48. The small clusters found in this liquid next to the hexagonal cluster lower Φ_6 noticeably. For the hexagonal liquid found from $\Delta = 0.2$ to $\Delta = 0.85$, Φ_6 increases due to the small local hexagonal structures. As soon as type 5 compact gels emerge at $\Delta = 0.85$ and above, the positional order parameter Φ_6 decreases.

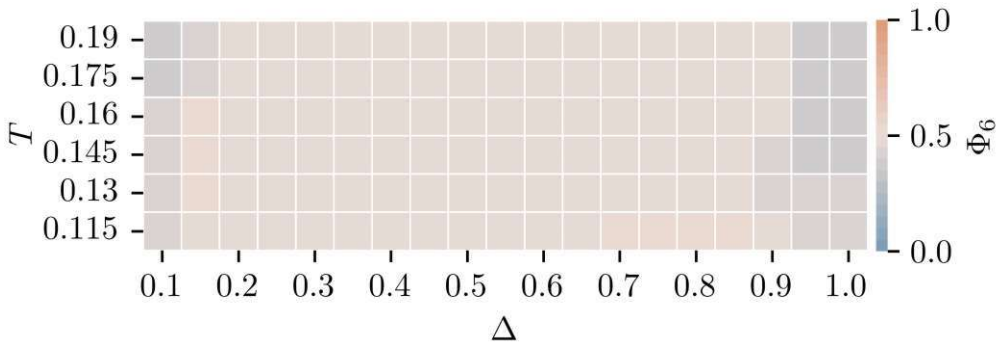


Figure 48: Average of the positional order parameter Φ_6 for each Δ-T combination of the VN patch topology with a particle density of $\phi = 0.3$. For systems that show no hexagonal lattice structure, this order parameter reaches 0 (blue). With an increasing appearance of a hexagonal lattice structure Φ_6 increases to 1 (orange). A more detailed description of the positional order parameters can be found in section 3.2.2.

The orientation order parameter Φ_θ seen in figure 49 appears to be flawed for the case of VN. Since ideal small hexagonal clusters are typically close-packed, all triangles are anti-parallel to their neighbors. For the calculation of Φ_θ all bonds of a triangle are taken into account, thus all other five triangles of a hexagon. The decrease in Φ_θ observed at $\Delta = [0.2, 0.9]$ and $T = [0.145, 0.19]$ correlates to a higher average cluster size. The hexagons found at these temperatures are either fully satisfied, thus consisting of six triangles, or of five, For the temperatures $T < 0.145$, a higher number of clusters with a size of three our four particles are observed.

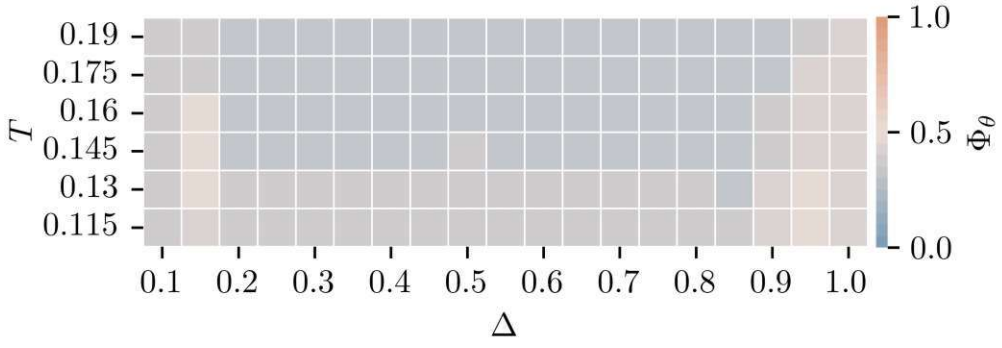


Figure 49: Average of the orientation order parameter Φ_θ for each Δ-T combination of the VN patch topology with a particle density of $\phi = 0.3$. For systems with mostly anti-parallel alignment of triangles, this order parameter converges to 0 (blue). If the assemblies show more triangles facing the same direction, Φ_θ increases up to its limit of 1 (orange). A more detailed description of the orientational order parameter can be found in section 3.2.3.

Since the VN patch topology mostly assembled into liquids consisting of small clusters or individual close-packed hexagons, the porosity order parameter Φ_ϕ returned mostly values close to 0, which is shown in figure 50. Only for larger assemblies, in this case the type 5 compact gel found at $\Delta = 0.85$ and above, does Φ_ϕ increase significantly.

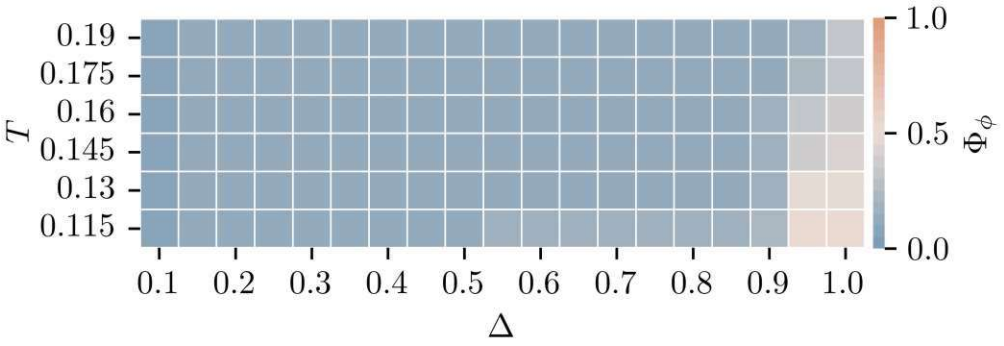


Figure 50: Average of the porosity order parameter Φ_ϕ for each Δ-T combination of the VN patch topology with a particle density of $\phi = 0.3$. This order parameter converges to 0 (blue) when there is no unoccupied space enclosed by connected triangles. The porosity order parameter is put into relation to the largest pore space (space enclosed by connected particles) found within all systems. If the order parameter reaches 1 (orange), the system has the largest pore space out of all systems in this analysis. A more detailed description of the orientational order parameter can be found in section 3.2.4.

The bond order parameter Φ_b works great at identifying the hexagonal liquids, which can be seen in figure 51. From $\Delta = 0.2$ to $\Delta = 0.9$, most particle only consist of one patch that has engaged in any bond, the vertex patch. At patch positions below $\Delta = 0.2$, the edge patch is utilized to form small clusters, such as the triangle triplets. For $\Delta \geq 0.9$, the edge patch is also engaged in bonds to form the type 5 compact gel.

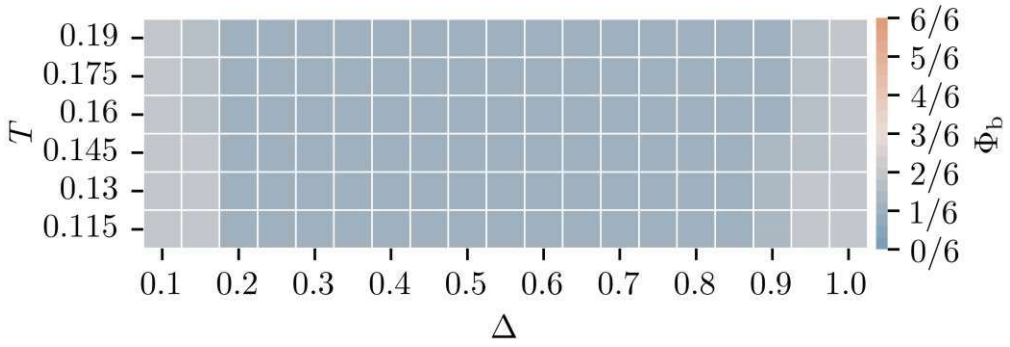
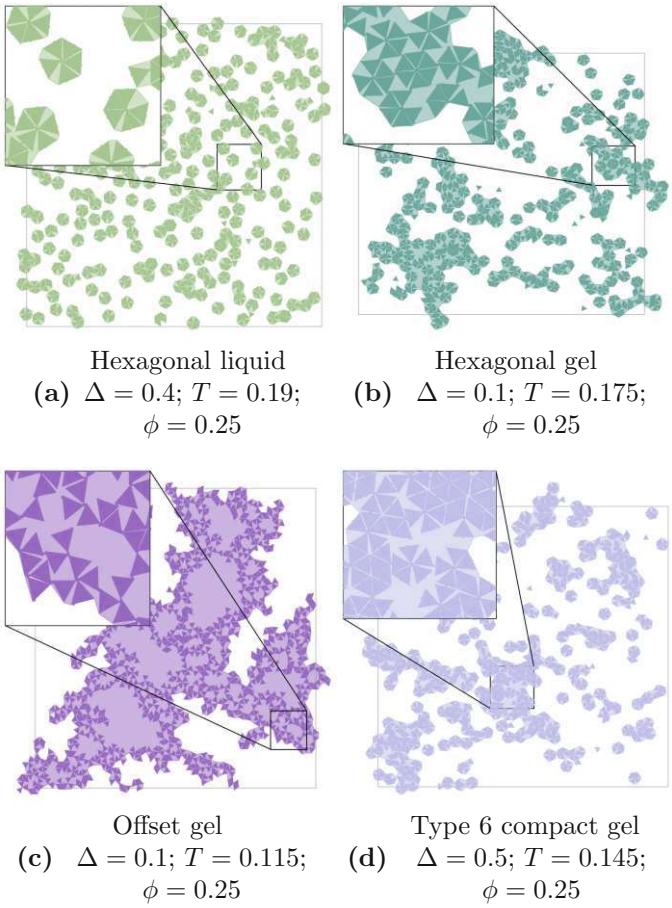


Figure 51: Average of the bond order parameter Φ_b for each Δ-T combination of the VN patch topology with a particle density of $\phi = 0.3$. The order parameter is 0 (blue) if there are no bonded patches and 1 (orange) if all particles contain six patches that are engaged in bonds. In the case of the VN patch topology, $\Phi_b = 2/6$ once all bonds are satisfied. The details for this choice are explained in section 3.2.1

3.3.6 VO

As a result of the patch placed along the edge opposite to the vertex patch, the VO patch topology assembles into structures containing local hexagonal structures which are linked together via the edge patch. The found assemblies are shown in figure 52. For $T \leq 0.145$, offset gel phases are favorable, which is shown in the state diagram figure 53 for $\phi = 0.3$. From $\Delta = 0.1$ to $\Delta = 0.4$, the assemblies appear in the form of hexagonal gels or even hexagonal liquids. The centered edge patch at $\Delta = 0.45$ or $\Delta = 0.5$ results in type 6 compact gels. All of the states found for the VO patch topology gain a lot of bond flexibility due to the vertex patch.



Hexagonal liquid
(a) $\Delta = 0.4$; $T = 0.19$; $\phi = 0.25$

Hexagonal gel
(b) $\Delta = 0.1$; $T = 0.175$; $\phi = 0.25$

Offset gel
(c) $\Delta = 0.1$; $T = 0.115$; $\phi = 0.25$

Type 6 compact gel
(d) $\Delta = 0.5$; $T = 0.145$; $\phi = 0.25$

Figure 52: Possible states found for the VO patch topology. Due to the vertex patch, triangles can form liquids consisting mostly of hexagonal clusters **(a)**. Similar to the hexagonal liquid found in the VN topology, most sub-clusters are not fully satisfied but only consist of five particles, leaving room for one more. The small hexagonal sub-clusters can connect via the edge patches to each other to form hexagonal gels **(b)**. These substructures disappear at lower temperatures and where offset gels can be found **(c)**. The flexibility given by the vertex patch creates highly disordered gel structures. With a centered edge patch, these gel structures show an increase in orientational and positional order, due to close-packed regions. This gel type is the type 6 compact gel **(d)**.

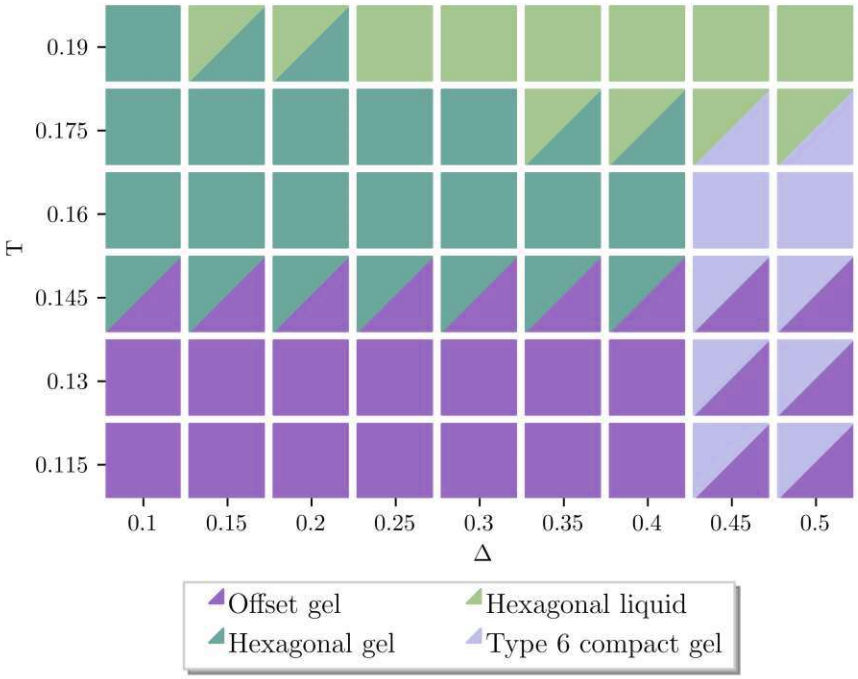


Figure 53: The visually classified Δ-T of the VO patch topology for $\phi = 0.3$. Offset gels are more likely to form at the low temperatures of $T = [0.115, 0.145]$. An increase in temperature gives rise to the appearance of hexagonal sub clusters for $\Delta \leq 0.4$. For $\Delta \geq 0.45$, compact regions become more common with the temperature increase, thus leading to the assembly of type 6 compact gels. At $T = 0.19$, and in some cases even $T = 0.175$, large structures become less favorable than hexagonal liquids.

The assemblies found for the VO patch topology don't show any distinctive behavior in Φ_4, which is shown in figure 54. The most significant results are the ones of the type 6 compact gel found at $\Delta = [0.45, 0.5]$ and $T = [0.115, 0.175]$, where the square lattice structure is the weakest. All other assemblies do not show distinguishable differences in terms of their square lattice structure and thus Φ_4. One possible explanation could be the bond flexibility, leading to irregular positional structure.

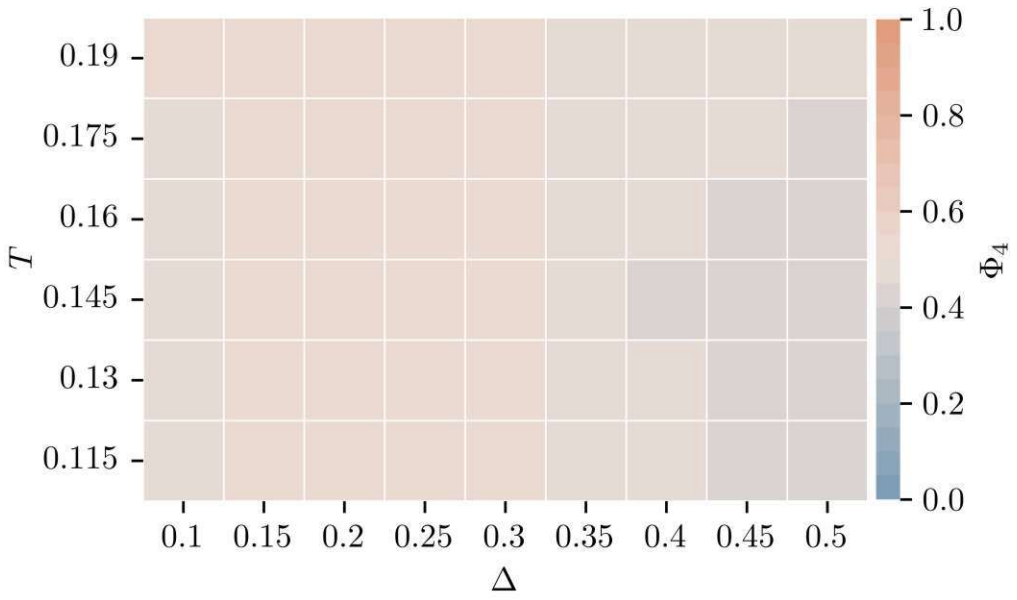


Figure 54: Average of the positional order parameter Φ_4 for each Δ-T combination of the VO patch topology with a particle density of $\phi = 0.3$. For systems that show no square lattice structure, this order parameter reaches 0 (blue). With an increasing appearance of a square lattice structure Φ_4 increases to 1 (orange). A more detailed description of the positional order parameters can be found in section 3.2.2.

In figure 55, the results of Φ_6 are shown for the VO patch topology. The type 6 compact gel found at $\Delta = [0.45, 0.5]$ and $T = [0.115, 0.175]$ show the strongest results at $T = [0.115, 0.13]$. The bond angle flexibility due to the vertex patch, leads to weaker positional structure in other assemblies. This is reflected in Φ_6, since those systems therefore show little hexagonal lattice structure.

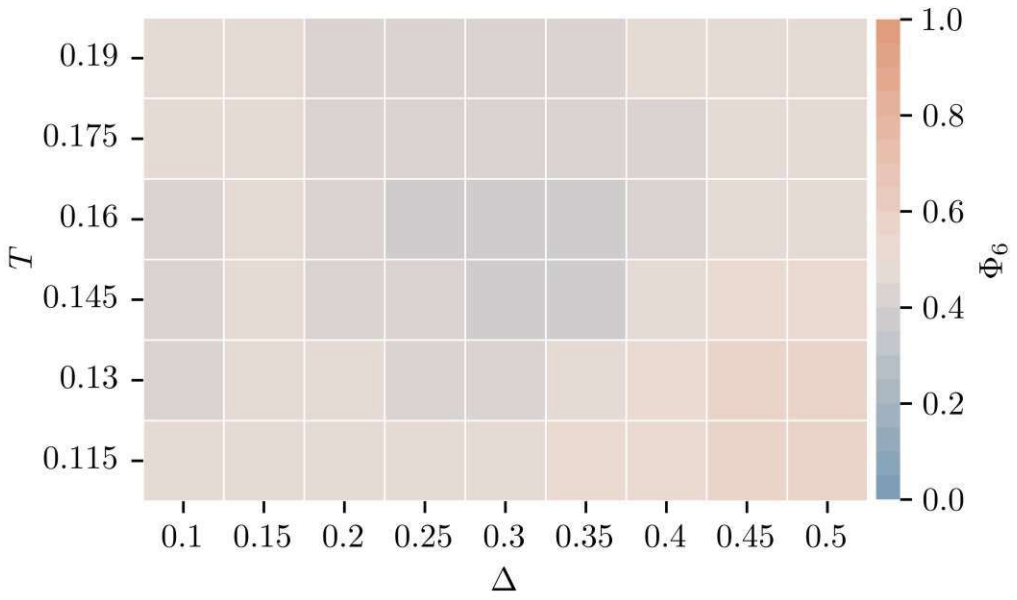


Figure 55: Average of the positional order parameter Φ_6 for each Δ-T combination of the VO patch topology with a particle density of $\phi = 0.3$. For systems that show no hexagonal lattice structure, this order parameter reaches 0 (blue). With an increasing appearance of a hexagonal lattice structure Φ_6 increases to 1 (orange). A more detailed description of the positional order parameters can be found in section 3.2.2.

The results of the orientational order parameter Φ_θ for the VO patch topology suffer from the same issue as the VN patch topology and are shown in figure 56. The calculation of Φ_θ relies on the bond orientation between connected triangles, which can lead to mixed results when multiple triangles connect at the same patch. While there is a trend towards anti-parallel alignment, the multiple triangles connected via a single patch in different orientation lead to higher values of Φ_θ.

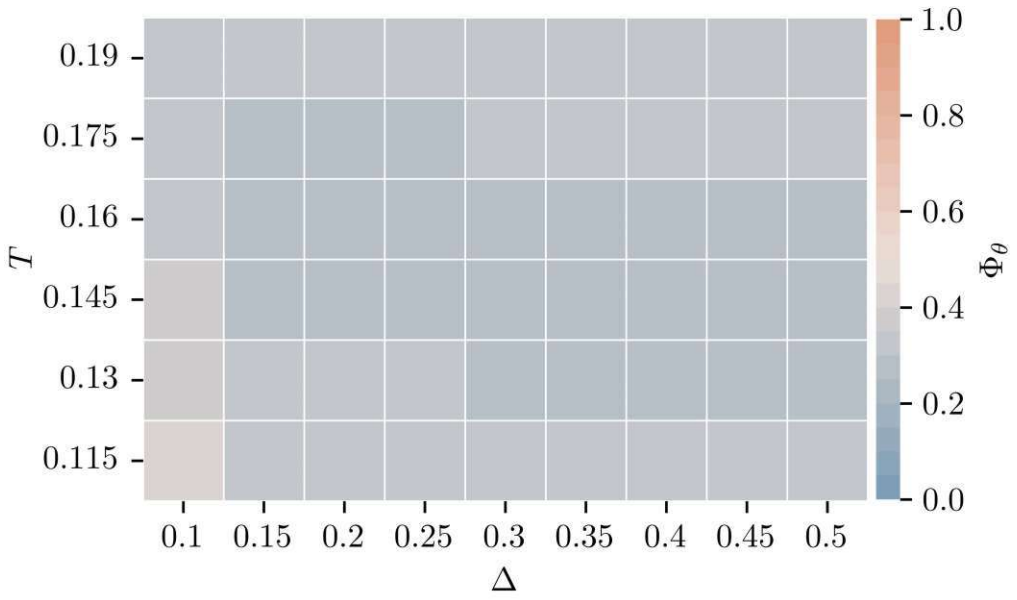


Figure 56: Average of the orientation order parameter Φ_θ for each Δ-T combination of the VO patch topology with a particle density of $\phi = 0.3$. For systems with mostly anti-parallel alignment of triangles, this order parameter converges to 0 (blue). If the assemblies show more triangles facing the same direction, Φ_θ increases up to its limit of 1 (orange). A more detailed description of the orientational order parameter can be found in section 3.2.3.

When it comes to porosity within the VO systems, Φ_ϕ works well at recognizing the appearance of offset gels, as seen in figure 57. These gels can be found at temperatures $T = 0.115$ and $T = 0.13$ and appear as the most porous. The other gel types, the hexagonal gel found at $\Delta = [0.1, 0.4]$ and $T = [0.145, 0.19]$, and the type 6 compact gel found at $\Delta = [0.45, 0.5]$ and $T = [0.115, 0.175]$, show a significantly lower amount of unoccupied area enclosed by clusters of triangles. The hexagonal liquids appearing at $\Delta = [0.15, 0.5]$ and $T = [0.175, 0.19]$ also show low values of Φ_ϕ.

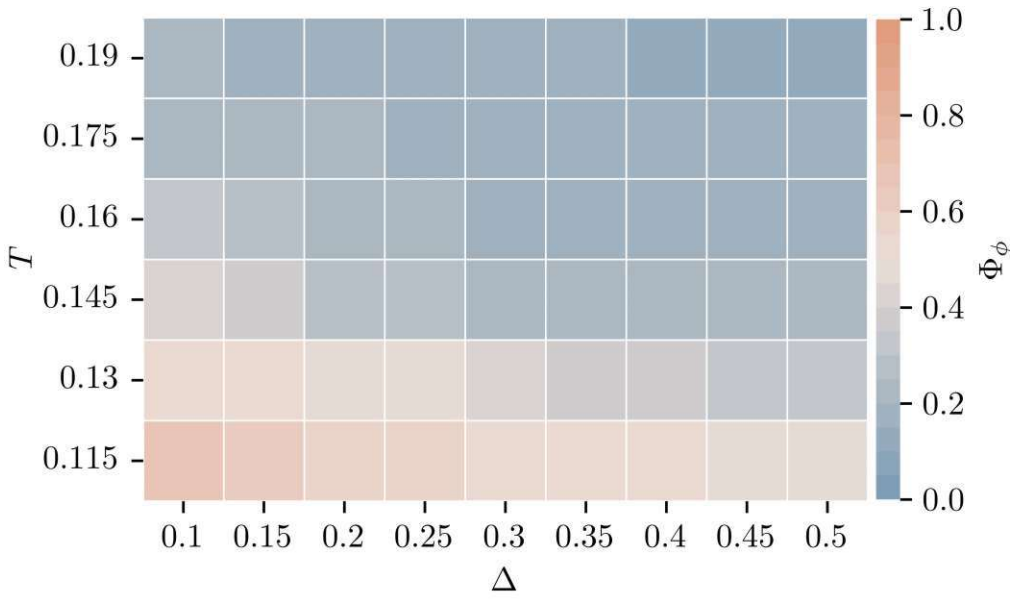


Figure 57: Average of the porosity order parameter Φ_ϕ for each Δ-T combination of the VO patch topology with a particle density of $\phi = 0.3$. This order parameter converges to 0 (blue) when there is no unoccupied space enclosed by connected triangles. The porosity order parameter is put into relation to the largest pore space (space enclosed by connected particles) found within all systems. If the order parameter reaches 1 (orange), the system has the largest pore space out of all systems in this analysis. A more detailed description of the orientational order parameter can be found in section 3.2.4.

The last order parameter discussed for the VO patch topology is Φ_b, the bond order parameter, which can be seen in figure 58. Systems with a higher number of small hexagonal clusters, such as the hexagonal liquid found at $\Delta = [0.15, 0.5]$ and $T = [0.175, 0.19]$ or any of its mixed phases, show a lower number of bonded patches per particle. Otherwise, Φ_b does not recognize any differences in bonding behavior between the different gel types.

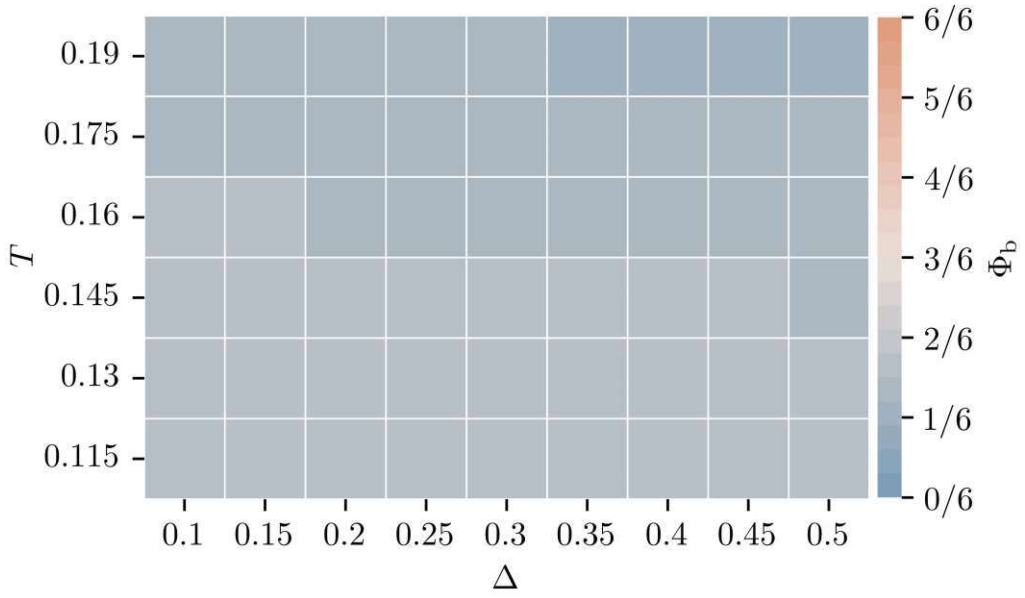


Figure 58: Average of the bond order parameter Φ_b for each Δ-T combination of the VO patch topology with a particle density of $\phi = 0.3$. The order parameter is 0 (blue) if there are no bonded patches and 1 (orange) if all particles contain six patches that are engaged in bonds. In the case of the VO patch topology, $\Phi_b = 2/6$ once all bonds are satisfied. The details for this choice are explained in section 3.2.1

4 Machine Learning

4.1 Dimensionality reduction and order parameters

To summarize, the systems considered are all possible combinations of the thermodynamic parameters reported in table 1 for each Δ of each of the six patch topologies. Dimensionality reduction of a given data set can speed up computation times and simplifies the accuracy analysis of the clustering algorithms later on. Additionally, the "curse of dimensionality" can be mitigated. A common algorithm for dimensionality reduction is the Principal Component Analysis (PCA). The principal component analysis searches for the parameters that explain the highest amount of variance by computing the covariance matrix to identify correlations. A helpful way to gain a deeper understanding of the PCA's results is the component-wise study of the percentage of explained variance. This is shown in figure 59 as blue bars, which provide insight into the ratio of information retained in each component during dimensionality reduction. Additionally, the plot shows the cumulative sum of this percentage in orange. This sum reveals the necessary principal components needed to describe the data in the parameter space with a desired amount of explained variance. Using the first three principal components, the sum of the explained variance is equal to 94.04%.

The composition of each principal component is shown in figure 60, by displaying the square of the explained variance per component. By studying this plot, it becomes possible to better understand each order parameters' overall success at classifying the different assemblies. When studying the parameter contribution to each principal component, the positional order parameter Φ_6 which recognizes hexagonal lattice structures, seems to be the most relevant out of the chosen order parameters. Next to it is the orientational order parameter Φ_θ. The combination of mostly these two parameters already explains 86.15% of the variance.

The second principal component is made up of mostly the average amount of bonded patches per particle Φ_b and the positional order parameter Φ_4, which recognizes square lattice structures. The orientation order parameter Φ_θ responded mostly to the difference in systems at $\Delta \leq 0.1$ to larger Δ. This suggests that if systems with the patch spillover explained in figure 5 were to be removed, it would not be as crucial for explaining the variance. Regardless, in combination with Φ_6, Φ_4, and Φ_b, the system are well described. Since close-packed triangles assemble into hexagonal lattice structures, it makes sense that Φ_6 is effective at describing the variety of different systems. It successfully recognizes any close-packed areas, independently of the overall physical state and thus even recognizes differences in some of the variations of gels, crystals, or liquids. Since the small clusters occasionally found within liquids can mislead the order parameter Φ_6 into recognizing hexagonal lattice structure despite it only being short range order.

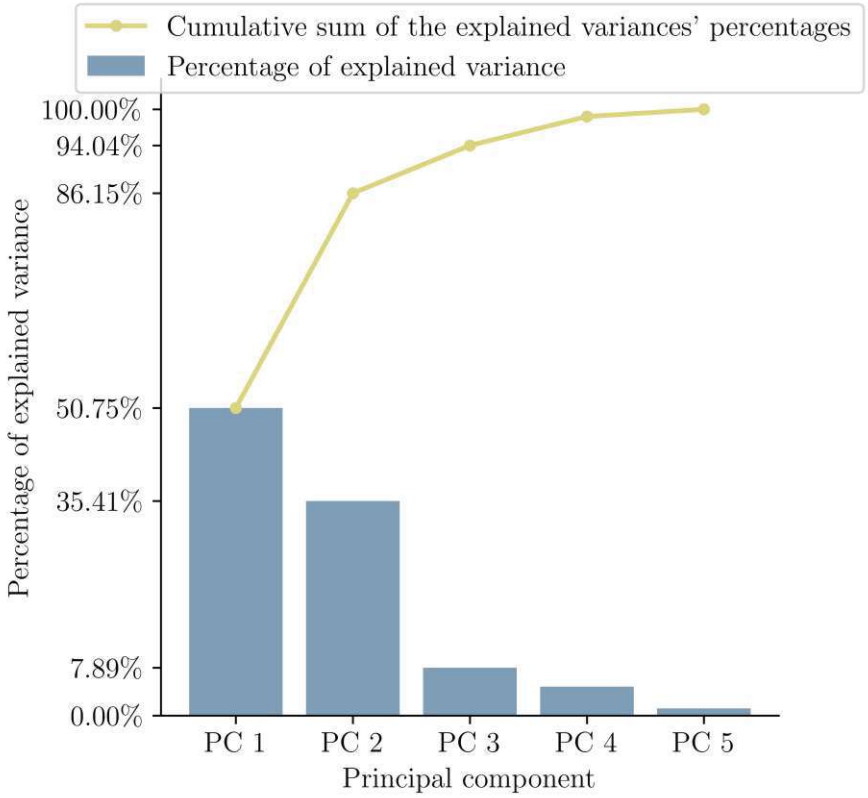


Figure 59: The percentage of explained variance for each principal component (blue) and their cumulative sum (orange). By using the first three principal components, one can explain 94.04% of the variance.

This flaw is mitigated through Φ_b, which served well at recognizing systems in a liquid state. The contribution of Φ_4 could be traced back to two different origins. With the appearance of close-packed hexagonal lattice structures, Φ_4 tends to decrease. Additionally, a sparsely-packed periodic structure of triangles resembles a square lattice structure where each triangle connects with two others at each of its vertices.
The porosity order parameter Φ_ϕ shows some contribution to PC 1 and PC 2, but becomes relevant for PC 3 and PC 4. It is decent at describing the transition from gel or porous crystal states to close-packed crystals, compact gels, or liquids, but fails at more nuanced recognition of different gel variants.

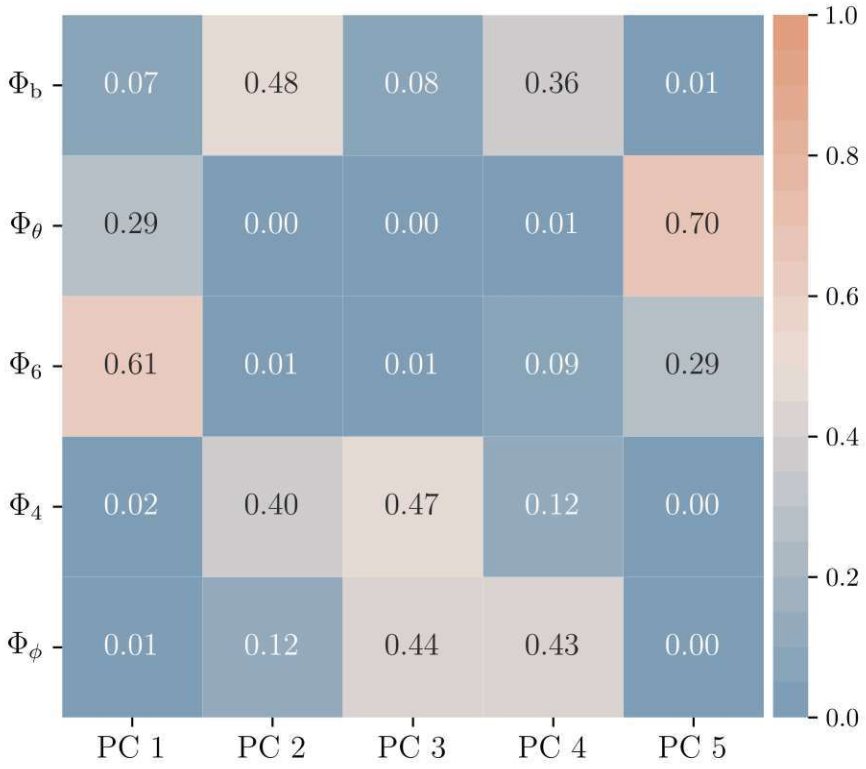


Figure 60: Contribution of each order parameter in our dimensionality reduction displayed by showing their contribution to each principal component. The explained variance of each contributor to a principal component is squared to reflect a scale where 0 (blue) contributes nothing to a principal component and 1 (orange) in case a principal component is described only by this order parameter.

The results of the PCA, specifically the values of the first two principal components, are shown in figure 61. This plot shows the data after it has been reduced to two dimensions, thus delivering the first comprehensive visualization of the chosen order parameters' results.
Around (0.3,0.1), there is a cluster consisting of all patch topologies, except the 6patch patch topology. Contained in this cluster are all systems of the VN (dark blue) and VO (light blue) patch topology and a small set of the patch topologies consisting of three patches: 3asym (green), 2asym1c (red), and mouse (ocher). This suggests that these systems are liquids made up of small clusters or hexagons found in the two patch patch topologies VN and VO, as well as the systems of $\Delta \leq 0.1$ from the patch topologies 3asym, 2asym1c, and mouse, due to the clusters size.
At (0.4,0.5) is an area where all the three-patch patch topologies converge. A

physical state in which all of them appear, would be the liquids with a low amount of small clusters. With a decrease in temperature, these patch topologies start to show different behaviors in their assembly, which is why they start to branch out with an increase in PC 1 and PC 2.
For the 6patch patch topology (purple), two clusters are found: A larger one at (-0.4,0.4) and a smaller one at (0,0.4). Within the 6patch patch topology, there is one type of state, the offset gels found at $\Delta = 0.1$, which appeared less often than the compact gels found for all other deltas. The gel types from the 6patch patch topology appear to deliver significant different results after the PCA, than all the other gels which are found in the first mentioned cluster at (0.3,0.1). The

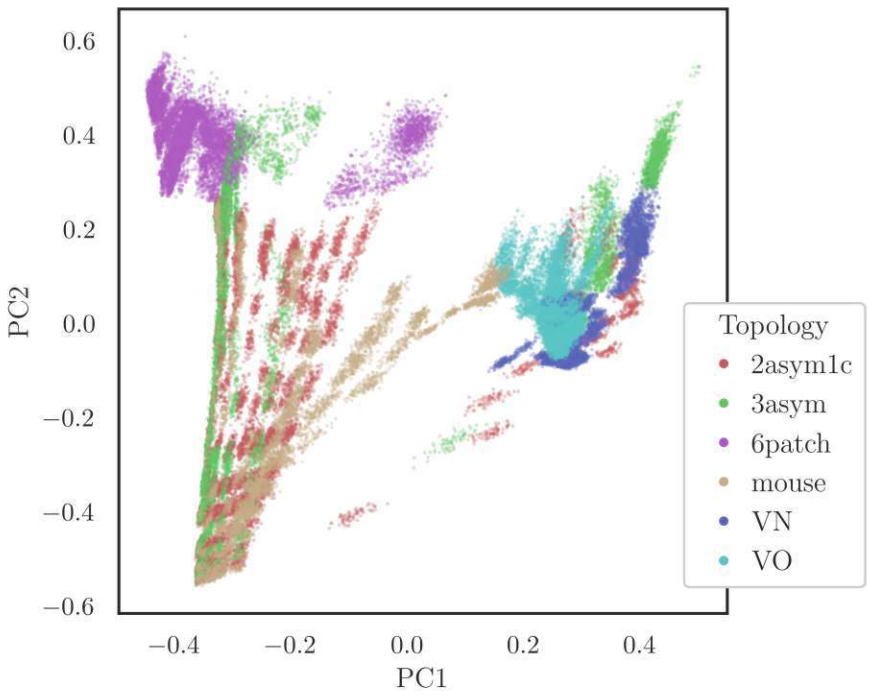


Figure 61: Result of reducing the order parameter space dimensions by using PCA. In this case, it is reduced to two dimensions, the first two principal components, PC 1 and PC 2. This combination explains 86.15% of the variance, which is seen in figure 59. Each patch topology is shown in a different color.

order parameter that behaves significantly differently for the 6patch systems is Φ_b, the bond order parameter. The reason for this behavior is the norm that is used for this order parameter, the maximum number of patches on any single particle, which is the highest for the 6patch topology. In figure 62 the correlation of each order parameter and the principal components is shown, which also reveals the order parameter values for each data point. The image for Φ_b shows a significant

increase for all systems of the 6patch topology. Additionally, Φ_b contributes more to PC 2 and therefore shows an increase with the increase in PC 2, thus systems with a higher number of bonds per particle can be found at higher values of PC 2. The other two order parameters correlated to the second principal component are Φ_4 and a lesser contribution of Φ_ϕ. Despite its minor contribution to PC 2, the porosity order parameter Φ_ϕ shows a correlation with the increase in PC 2. Systems with a higher porosity can be found at higher values of PC 2. Lastly, the positional order Φ_4 which indicates the occurrence of square lattice structures, reveals a higher amount of square lattice order towards the lower end of PC 2. Here, assemblies in a liquid state are found. These liquids contain a small amount of small clusters, which Φ_4 recognizes as clusters with a high amount of square lattice order. As larger clusters appear and therefore PC 2 increases, Φ_4 starts to decrease.

PC 1 mostly consists of the positional order parameter Φ_6 and the orientational order parameter Φ_θ. The orientational order parameter decreases with a reduction of PC 1, thus systems with primarily anti-parallel orientation order are found at low PC 1 values. At higher values of PC 1 (0.35 and above), Φ_θ mostly detects systems with random orientations. The last order parameter, Φ_6, identifies hexagonal lattice structure. The clusters occurring in systems with high values of Φ_6 show close-packed areas, leading to low values of PC 1.

To summarize, systems with low values of PC 1 exhibit anti-parallel order and hexagonal lattice structure, while systems with high PC 1 exhibit random orientational order or even parallel orientations and less hexagonal lattice structure. If a system has low PC 2 values, it has square lattice structure, a small sum of area of unoccupied space enclosed by particles and a low number of patches engaged in bonds per particle.

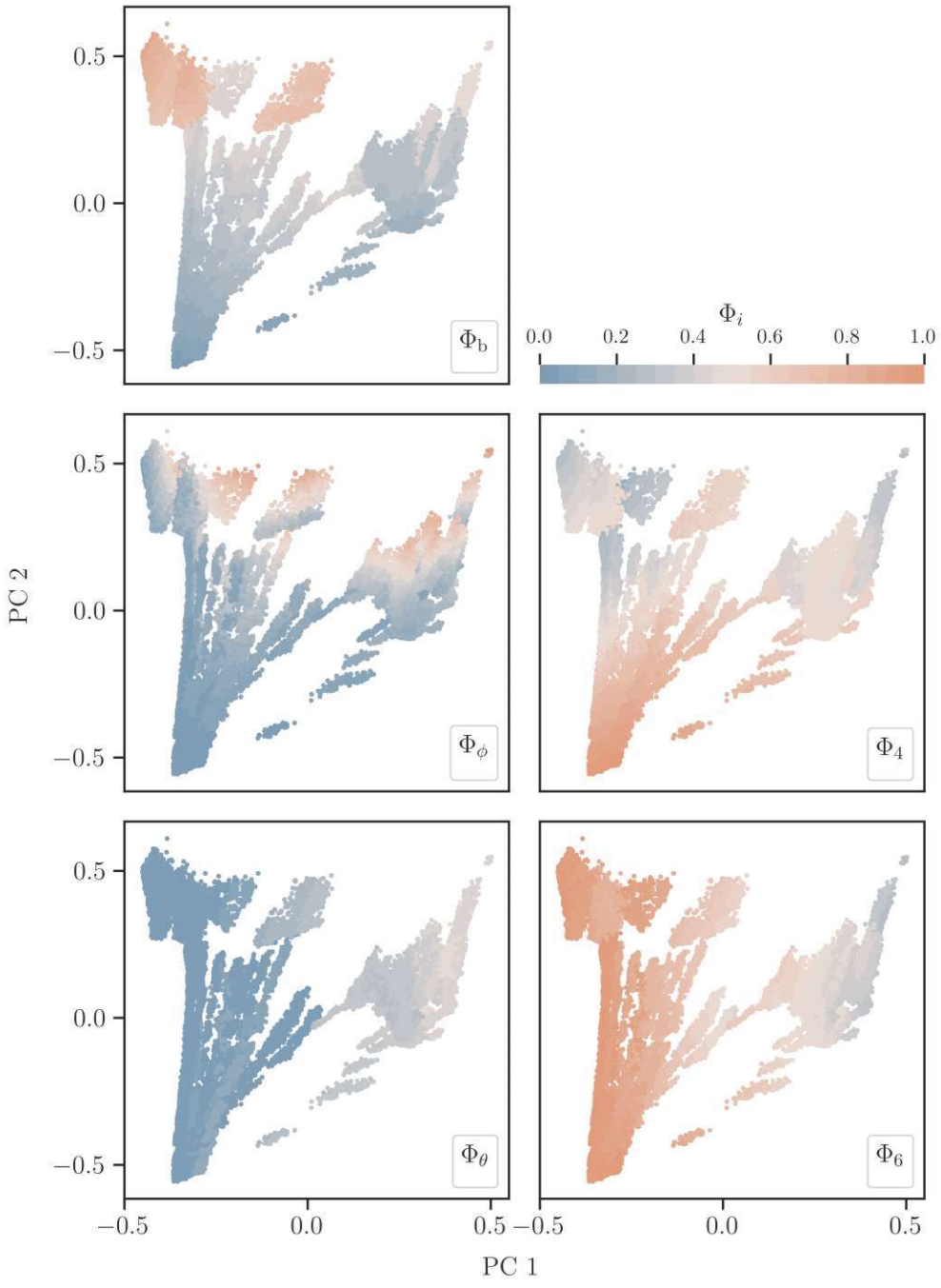
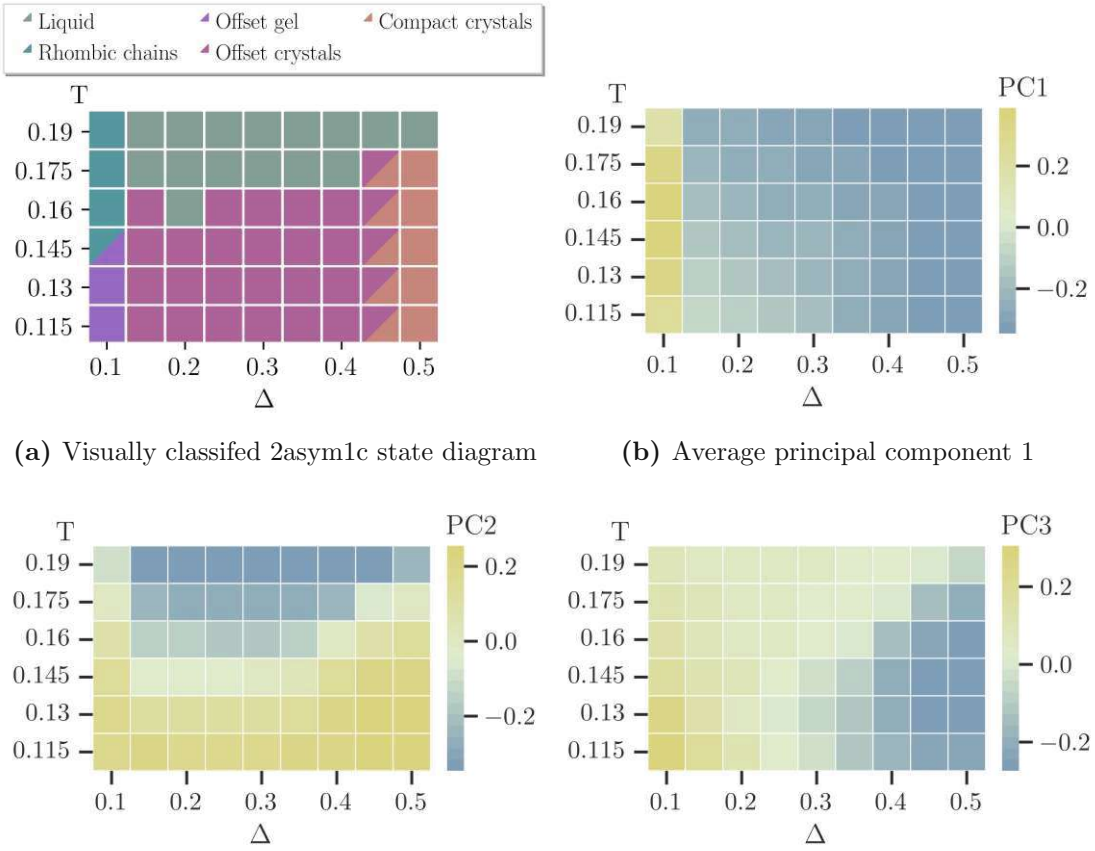


Figure 62: Each order parameter Φ_i where $i = \{\mathrm{b}, \phi, \theta, 4, 6\}$ of each system after application of the principal component analysis and thus a reduction to two dimensions. Some order parameters, such as the orientation order parameter Φ_θ and the porosity order parameter Φ_ϕ never reach their potential maximum value of 1. On the other side, the positional order parameters Φ_4 and Φ_6 return rarely reach 0. The bond order parameter Φ_b responds strongly to the 6patch topology due to the definition of its norm.

4.2 PCA results by topology

The combination of the different principal components ideally results in a complete description of a system's physical state. For the 2asym1c patch topology, this is shown in figure 63, by comparing the principal components to the visually classified state diagram. The first principal component succeeds in identifying the differences between offset gels and rhombic chains found at $\Delta = 0.1$ versus all other possible states. There is a slight increase in PC 1 at $\Delta = [0.15, 0.25]$ and $T = [0.115, 0.145]$, which describes the appearance of offset crystals. In case of the 2asym1c patch topology, the second principal component works well at identifying systems in a liquid state. By using the third principal component, separation between different crystals types becomes even easier.



(a) Visually classifed 2asym1c state diagram

(b) Average principal component 1

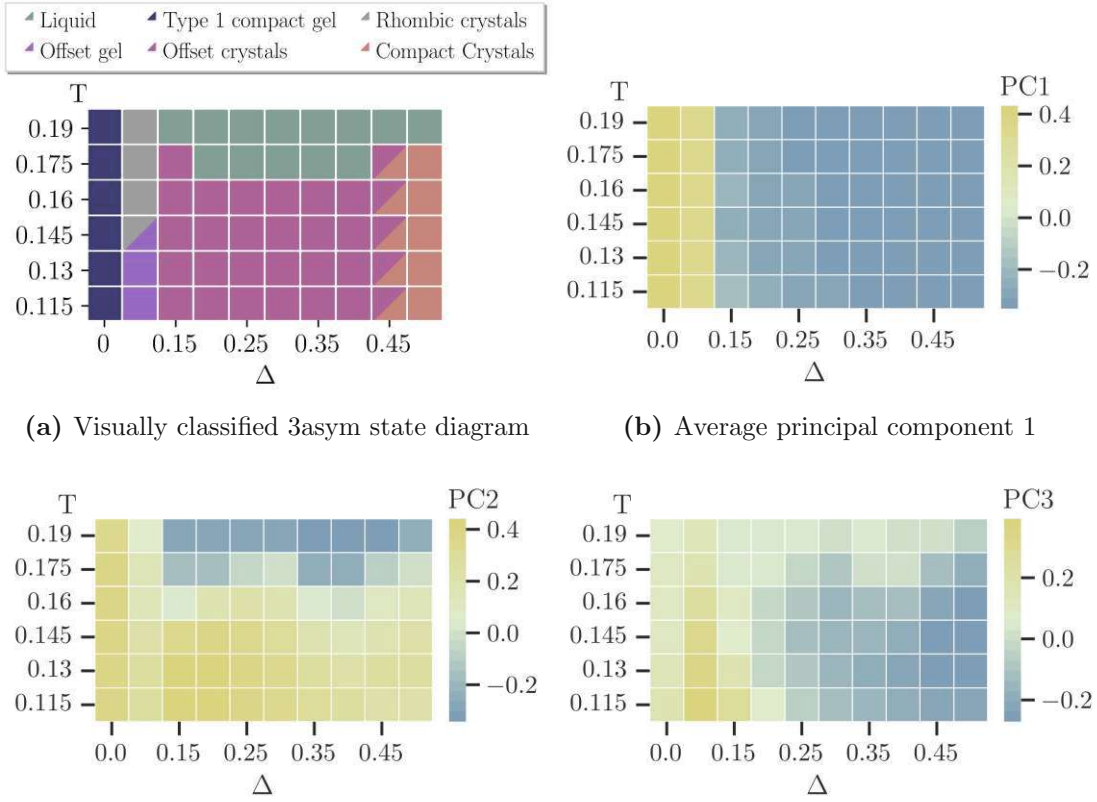
(c) Average principal component 2

(d) Average principal component 3

Figure 63: Comparison of the visual classification results **(a)** with the first three principal components of the 2asym1c patch topology with $\phi = 0.3$. The different physical states of this patch topology are shown in different colors and are manually classified. Each of the principal components reveals a different trend in the chosen set of order parameters. The first principal component **(b)** is a combination of the positional order parameter Φ_6, revealing a high amount hexagonal lattice structure for low PC 1 values, and the orientational order parameter Φ_θ, which is 0 for systems with anti-parallel orientations and 1 for parallel ones. A low value of Φ_θ is reflected as a higher value in PC 1 and a high value of Φ_θ as a lower value in PC 1. The second principal component **(c)** decreases with Φ_b, the average of bonded patches per particle, and is inversely related to Φ_4, which describes the square lattice structure found in the systems. The third principal component **(d)** increases with the porosity Φ_ϕ and the amount of square lattice structure Φ_4.

In case of the 3asym patch topology, the first principal components delivers

the same results as seen above, which is shown in figure 64 PC 1 successfully identifies the appearance of different structures at $\Delta \leq 0.1$. Also similar to before is the behavior of the second principal component, describing the appearance of systems in a liquid state. The third principal component appears to be weaker at recognizing the appearance of close-packed crystals, but captures a difference between the gels found at $\Delta = 0$ and $\Delta = 0.1$.



(a) Visually classified 3asym state diagram

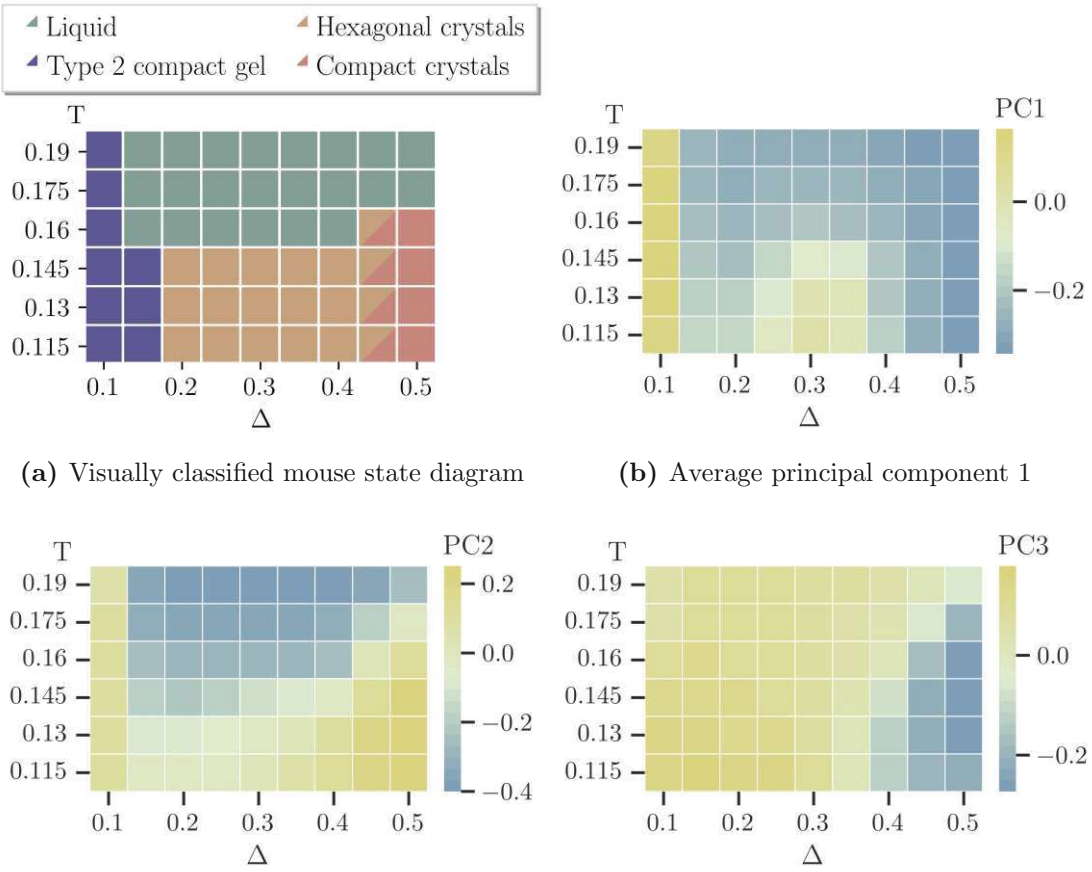
(b) Average principal component 1

(c) Average principal component 2

(d) Average principal component 3

Figure 64: Comparison of the visual classification results **(a)** with the first three principal components of the 3asym patch topology with $\phi = 0.3$. The different physical states of this patch topology are shown in different colors and are manually classified. Each of the principal components reveals a different trend in the chosen set of order parameters. The first principal component **(b)** is a combination of the positional order parameter Φ_6, revealing a high amount hexagonal lattice structure for low PC 1 values, and the orientational order parameter Φ_θ, which is 0 for systems with anti-parallel orientations and 1 for parallel ones. A low value of Φ_θ is reflected as a higher value in PC 1 and a high value of Φ_θ as a lower value in PC 1. The second principal component **(c)** decreases with Φ_b, the average of bonded patches per particle, and inversely relates to Φ_4, which describes the square lattice structure found in the systems. The third principal component **(d)** increases with the porosity Φ_ϕ and the amount of square lattice structure Φ_4.

When it comes to the mouse patch topology, PC 1 displays slightly different results, which can be seen in figure 65. The hexagonal substructures found in the open hexagonal crystals lead to an increase in PC 1 at $\Delta = [0.25, 0.45]$ and $T = [0.115, 0.145]$. Otherwise, it also behaves similar to the PC 1 of the other three-patch patch topologies and shows the appearance of gels at $\Delta = 0.1$. The second principal component describes the appearance of systems in a liquid state, similar to what was previously observed. For the mouse patch topology, the third principal component successfully identifies the close-packed crystals as a different physical state.



(a) Visually classified mouse state diagram

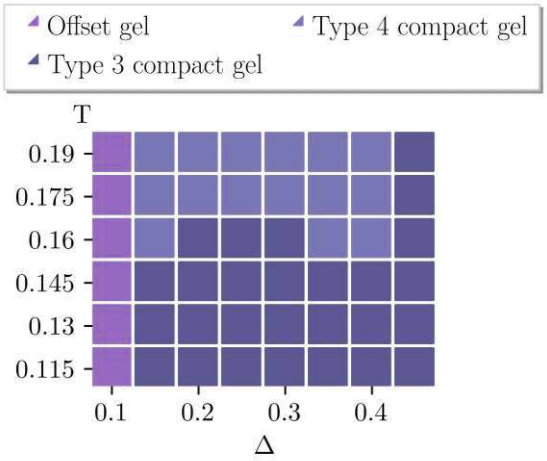
(b) Average principal component 1

(c) Average principal component 2

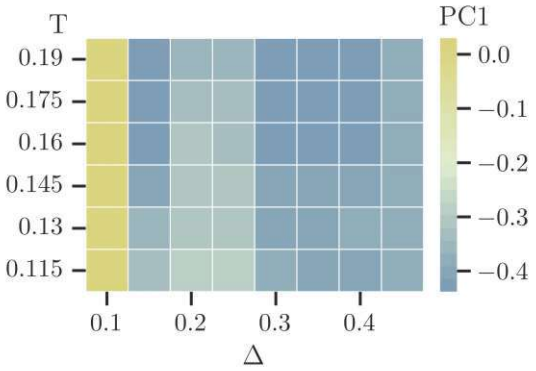
(d) Average principal component 3

Figure 65: Comparison of the visual classification results **(a)** with the first three principal components of the mouse patch topology with $\phi = 0.3$. The different physical states of this patch topology are shown in different colors and are manually classified. Each of the principal components reveals a different trend in the chosen set of order parameters. The first principal component **(b)** is a combination of the positional order parameter Φ_6, revealing a high amount hexagonal lattice structure for low PC 1 values, and the orientational order parameter Φ_θ, which is 0 for systems with anti-parallel orientations and 1 for parallel ones. A low value of Φ_θ is reflected as a higher value in PC 1 and a high value of Φ_θ as a lower value in PC 1. The second principal component **(c)** decreases with Φ_b, the average of bonded patches per particle, and inversely relates to Φ_4, which describes the square lattice structure found in the systems. The third principal component **(d)** increases with the porosity Φ_ϕ and the amount of square lattice structure Φ_4.

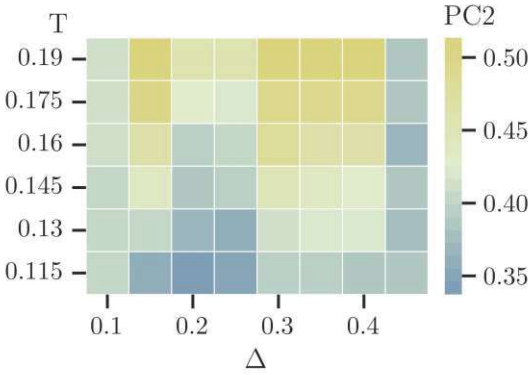
Also in the case of the 6patch patch-topology, the first principal components main information is the appearance of a gel state at $\Delta = 0.1$, which is shown in figure 66. Additionally, it also captures the slightly different behavior at $\Delta = 0.2$ and $\Delta = 0.25$, which is explained in section 3.3.4. With the 6patch patch topology, PC 2 delivers different results than before, since all of the studied systems are in a gel state. It captures the slight difference in gel types, identifying a different gel type at $T \geq 0.175$ than below. Lastly, the third principal component appears to work exceptionally well for this patch topology. It identifies the appearance of the offset gel at $\Delta = 0.1$ and a difference between type 3 and type 4 compact gels.



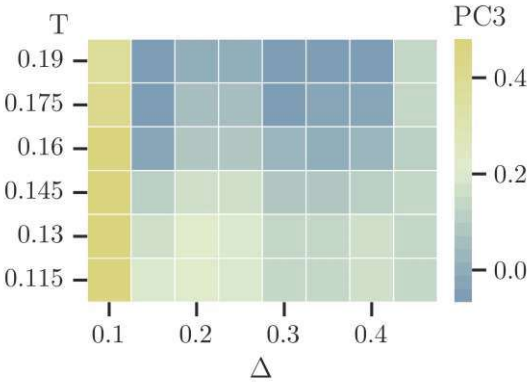
(a) Visually classified 6patch state diagram



(b) Average principal component 1



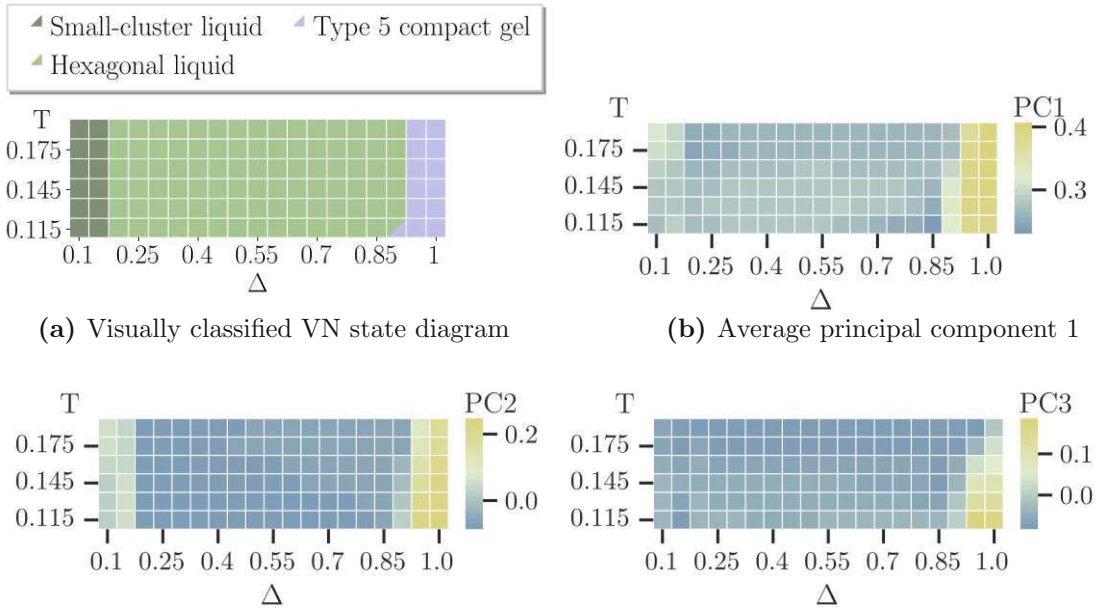
(c) Average principal component 2



(d) Average principal component 3

Figure 66: Comparison of the visual classification results **(a)** with the first three principal components of the 6patch patch topology with $\phi = 0.3$. The different physical states of this patch topology are shown in different colors and are manually classified. Each of the principal components reveals a different trend in the chosen set of order parameters. The first principal component **(b)** is a combination of the positional order parameter Φ_6, revealing a high amount hexagonal lattice structure for low PC 1 values, and the orientational order parameter Φ_θ, which is 0 for systems with anti-parallel orientations and 1 for parallel ones. A low value of Φ_θ is reflected as a higher value in PC 1 and a high value of Φ_θ as a lower value in PC 1. The second principal component **(c)** decreases with Φ_b, the average of bonded patches per particle, and inversely relates to Φ_4, which describes the square lattice structure found in the systems. The third principal component **(d)** increases with the porosity Φ_ϕ and the amount of square lattice structure Φ_4.

The small clusters found using the VN patch topology are recognized as significantly different from the type 5 compact gel by PC 1, which can be seen in figure 67. By using PC 2, even the small-cluster liquid and hexagonal liquid states can be separated from another. With a slight temperature gradient at $\Delta \geq 0.9$, PC 3 carries a little more information about potential variations of the type 5 compact gel state.



(a) Visually classified VN state diagram

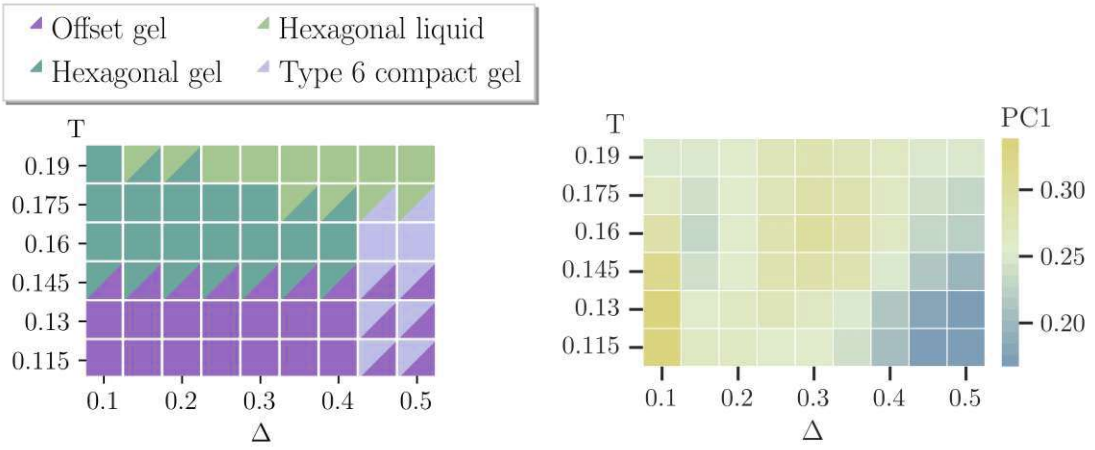
(b) Average principal component 1

(c) Average principal component 2

(d) Average principal component 3

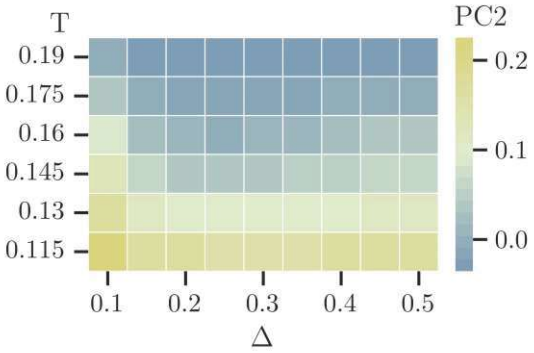
Figure 67: Comparison of the visual classification results **(a)** with the first three principal components of the VN patch topology with $\phi = 0.3$. The different physical states of this patch topology are shown in different colors and are manually classified. Each of the principal components reveals a different trend in the chosen set of order parameters. The first principal component **(b)** is a combination of the positional order parameter Φ_6, revealing a high amount hexagonal lattice structure for low PC 1 values, and the orientational order parameter Φ_θ, which is 0 for systems with anti-parallel orientations and 1 for parallel ones. A low value of Φ_θ is reflected as a higher value in PC 1 and a high value of Φ_θ as a lower value in PC 1. The second principal component **(c)** decreases with Φ_b, the average of bonded patches per particle, and inversely relates to Φ_4, which describes the square lattice structure found in the systems. The third principal component **(d)** increases with the porosity Φ_ϕ and the amount of square lattice structure Φ_4.

The state diagram of the VO patch topology appears as the most complex and the states' descriptions in terms of their principal components are displayed in figure 68. The first principal component shows significantly different behavior at $\Delta = 0.1$ as well as $\Delta = [0.4, 0.5]$ and $T = [0.115, 0.145]$. To further analyze this patch topology, the second principal component becomes necessary, which introduces a temperature gradient for PC 2, recognizing the appearance of hexagonal gels and hexagonal liquids at higher temperatures and offset gels at lower temperatures. The focus of the third principal component appears to be similar to the second, but it is suggesting the differences between hexagonal gels and liquids as well as a difference between compact and offset gels in their order paramters.

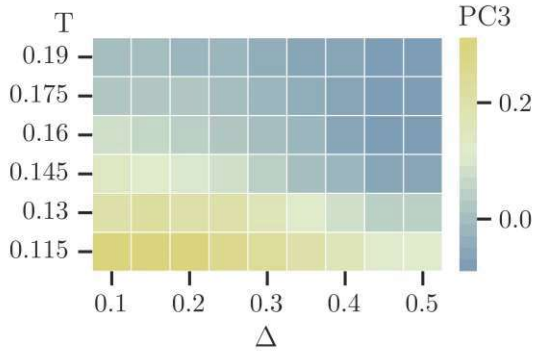


(a) Visually classified VO state diagram

(b) Average principal component 1



(c) Average principal component 2



(d) Average principal component 3

Figure 68: Comparison of the visual classification results **(a)** with the first three principal components of the VO patch topology with $\phi = 0.3$. The different physical states of this patch topology are shown in different colors and are manually classified. Each of the principal components reveals a different trend in the chosen set of order parameters. The first principal component **(b)** is a combination of the positional order parameter Φ_6, revealing a high amount hexagonal lattice structure for low PC 1 values, and the orientational order parameter Φ_θ, which is 0 for systems with anti-parallel orientations and 1 for parallel ones. A low value of Φ_θ is reflected as a higher value in PC 1 and a high value of Φ_θ as a lower value in PC 1. The second principal component **(c)** decreases with Φ_b, the average of bonded patches per particle, and inversely relates to Φ_4, which describes the square lattice structure found in the systems. The third principal component **(d)** increases with the porosity Φ_ϕ and the amount of square lattice structure Φ_4.

The third principal component can be visualized as a third dimension, which is represented by a color gradient in figure 69. It reveals not just differences between the selected patch topologies, but also a gradient within each of those. This is particularly visible for the 6patch clusters found at (-0.4,0.4) and (0,0.4).

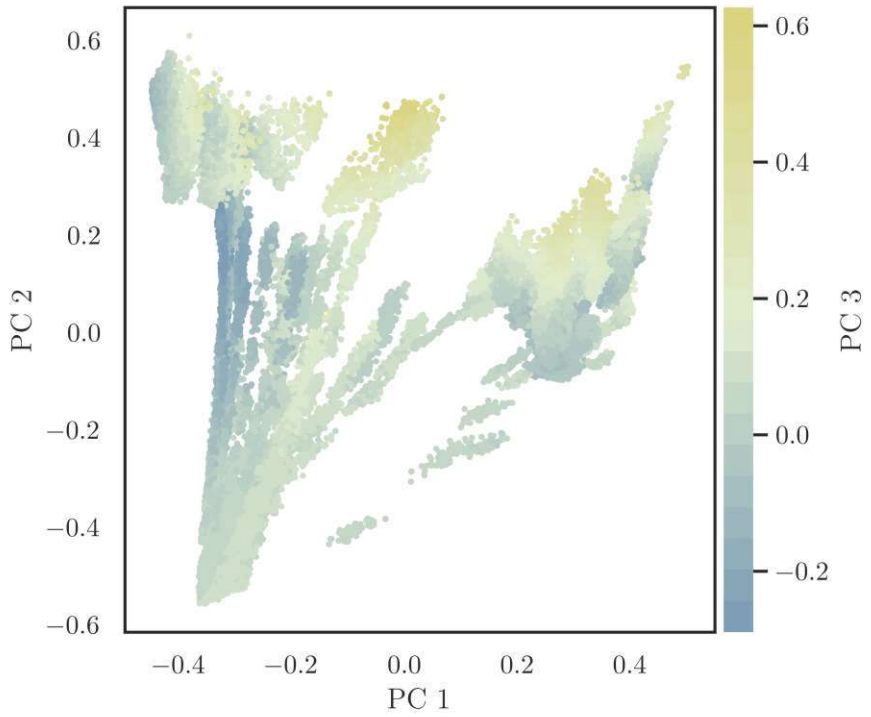


Figure 69: All three principal components visualized which allow the explanation of 94.04% of the variance. The third principal component is introduced to the two-dimensional plot through the use of a color gradient. PC 3 increases with the porosity Φ_ϕ and square lattice structure Φ_4.

To show the similarities between all patch topologies, PC 1 and PC 2 are plotted against each other and shown in figure 70. The three-patch patch topologies 3asym (green), 2asym1c (red), and mouse (ocher) appear as a large cluster of systems at (-0.3,0.4), which are in a liquid state at $T \geq 0.175$. There are some outliers: the systems with $\Delta \leq 0.1$ which show a different behavior from the rest of systems of their patch topology at all temperatures. At $T = 0.175$, the first systems of these topologies show an increase in PC 2, which continues for all temperatures. This suggests the disappearance of square lattice structure, an increasing porosity and an increase in the number of patches engaged in bonds per

particle, thus a transition from the liquid phase to larger clusters. At $T = 0.145$, most systems of these topologies were manually classified as different varieties of gels or crystals. These variations can be seen as the different branches of the fanned out cluster. Branches to the left show higher hexagonal lattice structure and greater anti-parallel orientational order, as indicated by PC 1. This means close-packed variations are found more to the left, whereas open structures at the right-hand branches of the fanned out cluster.
The VO patch topology (light blue) converges to a denser cluster with increasing temperature, accompanied by an average decrease in PC 2. This is where the systems of the VO patch topology are in a hexagonal liquid state, thus correlating with the decrease in Φ_ϕ, the porosity order parameter, which detects a smaller area enclosed by the clusters in hexagonal liquids compared to the larger clusters found for the VO patch topology.

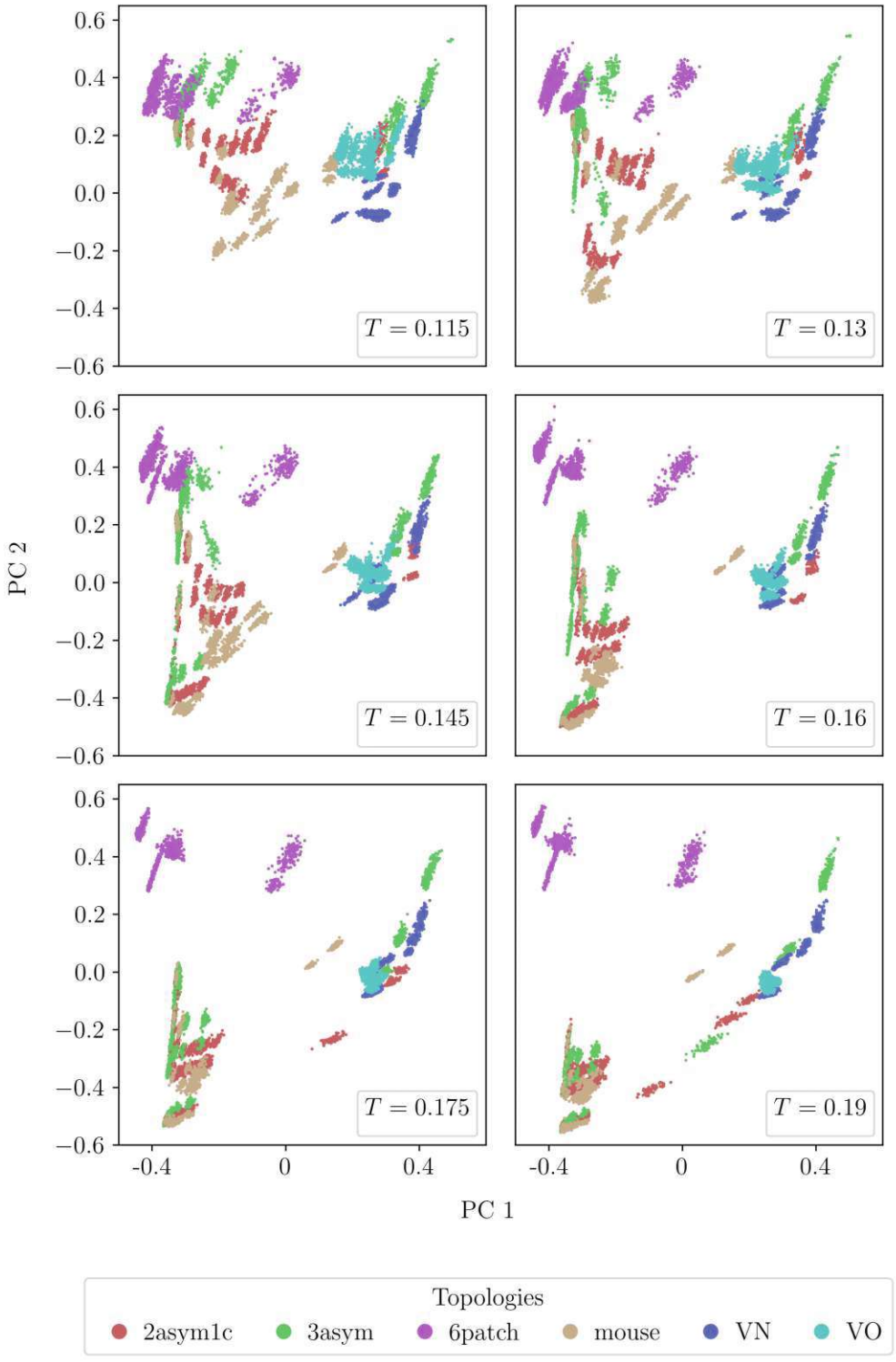


Figure 70: The first two principal components of every system divided by temperatures, where each color reflects a different topology. Overlapping clusters suggest similarities between assembles of different patch topologies. The three-patch patch topologies 3asym (green), 2asym1c (red), and mouse (ocher) show similar behavior. The gels found at $\Delta \leq 0.1$ perform distinctively different from the rest of their family, yet show similarities to the two-patch patch topologies VN (dark blue) and VO (light blue) in the results of the PCA. This can be seen at all temperatures, as the clusters of systems where PC1 ≥ 0. The 6patch patch topology (purple) shows a few different clusters, two located around (0.4,0.4) and one at (0,0.4). The second one is also a result of the anomaly at $\Delta = 0.1$

In the case of the 6patch patch topology (purple), there is little difference between the systems of different temperatures. They are roughly separable into two to three clusters, which will be further discussed with the help of figure 71. Also the VN (dark blue) patch topology shows little variation in its physical states at different temperatures. Most of the chosen patch topologies show significantly different behavior at $\Delta \leq 0.1$, which results in a general shift to higher values of PC 1, due to the loss of hexagonal lattice structure and orientational order.
When the VN patch topology is studied for different Δ, it shows consistent behavior from $\Delta = 0.2$ to $\Delta = 0.55$, where hexagonal liquids are found. At $\Delta \leq 0.15$ small cluster liquids can be found, recognizable as a small increase in PC 2 and decrease in PC 1. With the appearance of gel structures found at $\Delta \geq 0.9$, VN shifts to higher values of PC 1.
The most significant behavior in terms of the principal component that can be observed for different Δ of the VO patch topology, is the direction in which the top end of the linear cluster points. When keeping Δ fixed, this patch topology appeared as variations of offset gels at low temperatures, then transitioning to hexagonal or compact gel structures before turning into hexagonal liquids for some Δ at high temperatures. At $\Delta \geq 0.45$ systems transitioned into a compact gel state before reaching the hexagonal liquid state, resulting in a substantial decrease in orientational, explaining the tilt, since this correlates with a decrease in PC 1 as explained above. As mentioned previously, the orientational order parameter calculates the average pair-orientation of particles connected via one patch. The fact that these hexagonal clusters are connecting to the same patch of each particle leads to lower average pair-orientation for each particle, when compared to the compact gels which show larger close-packed areas due to the edge patch. This decrease in Φ_θ is reflected as an increase in PC 1. Since the hexagonal substructure does not increase long range hexagonal lattice structure, systems of $\Delta \leq 0.4$ are shown as vertical linear clusters. The different Δ of the VO patch topology converge to the lower point of the linear clusters, which is where the hexagonal

liquid state systems are found. The elongation of the clusters of VO for each Δ is explained by the different values of Φ_ϕ, the porosity order parameter. A higher sum of area enclosed by connected particles leads to higher values of Φ_ϕ and thus PC 2.
When it comes to the elongated shape of the three-patch patch topologies 3asym (green), 2asym1c (red), and mouse (ocher), there is a different origin. Here, the appearance of a square lattice structure, identified by Φ_4 within the rare small clusters in liquids, results in low values of PC 2. As the cluster size increases, the square lattice structure disappears, thus leading to higher values of PC 2. The lower the value of PC 1 is for a branch, the more hexagonal lattice structure is observed, even at lower temperatures (larger clusters), which can be recognized as increasingly close-packed assemblies.
The last patch topology reviewed for each Δ in terms of its order parameters after dimension reduction, is the 6patch topology (purple). At $\Delta \geq 0.15$ it remains relatively consistent up to $\Delta = 0.4$. Since the systems of the 6patch patch topology for $\Delta = 0.45$ all assembled into one kind of compact gel, the cluster is denser compared to those at lower Δ.

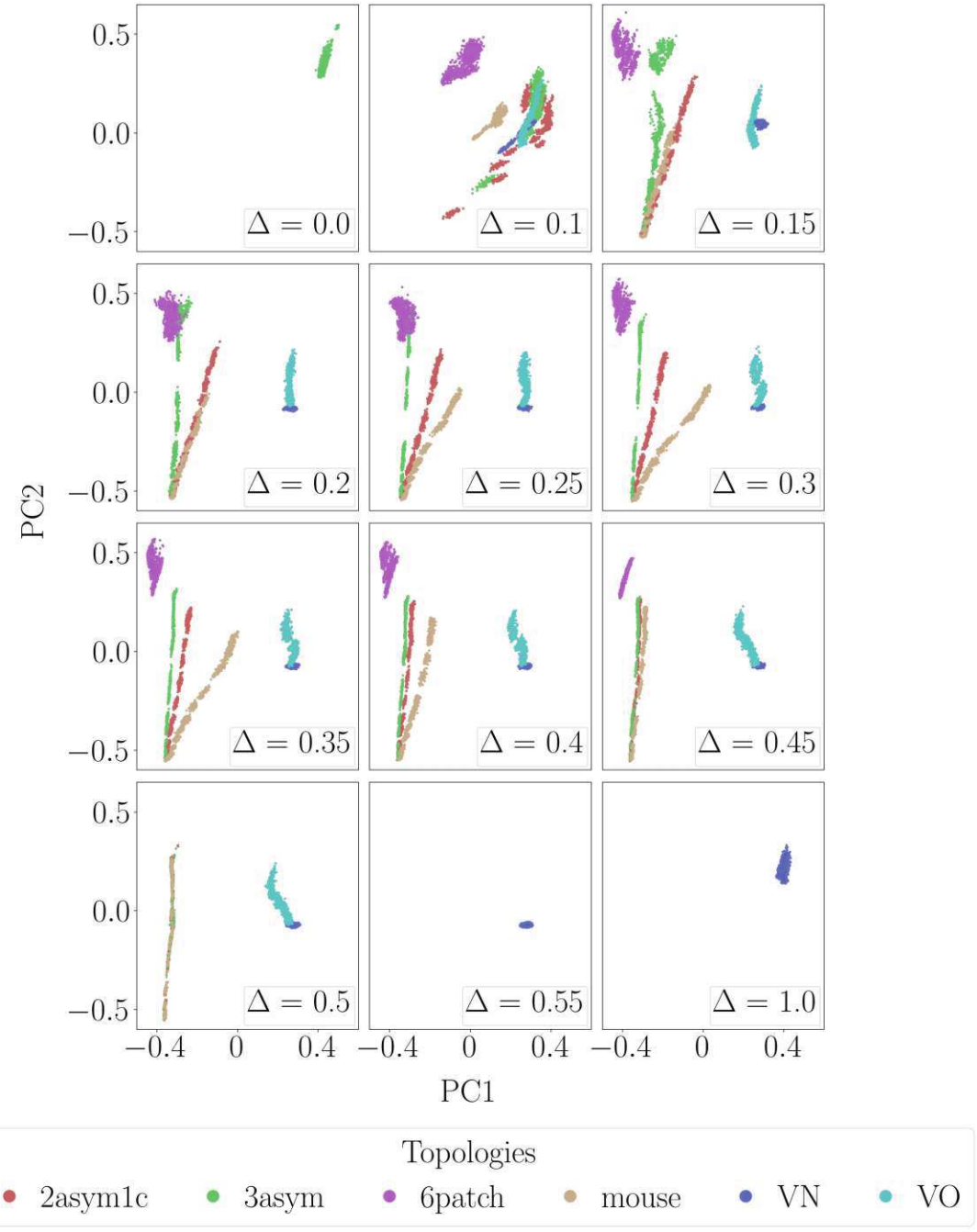


Figure 71: The first two principal components of every system divided by temperatures, where each color reflects a different topology. Overlapping clusters suggest similarities between assembles of different patch topologies. The three-patch patch topologies 3asym (green), 2asym1c (red), and mouse (ocher) show similar behavior. The gels found at $\Delta \leq 0.1$ perform distinctively different from the rest of their family, yet show similarities to the two-patch patch topologies VN (dark blue) and VO (light blue) in the results of the PCA. Elongated shapes appear when there are transitions into different states occurring with a change in temperature while Δ remains constant. The VN topology behaves consistently from $\Delta = 0.15$ to $\Delta = 0.85$. For patch positions $\Delta \geq 0.9$, the VN systems jump to higher values of PC 1 and PC 2, where they remains constant up to $\Delta = 0.1$ The 6patch patch topology (purple) appears at one cluster at (0,0.4) for $\Delta = 0.1$ and at (0.4,0.4) for all other deltas.

4.3 Clustering and automated classification

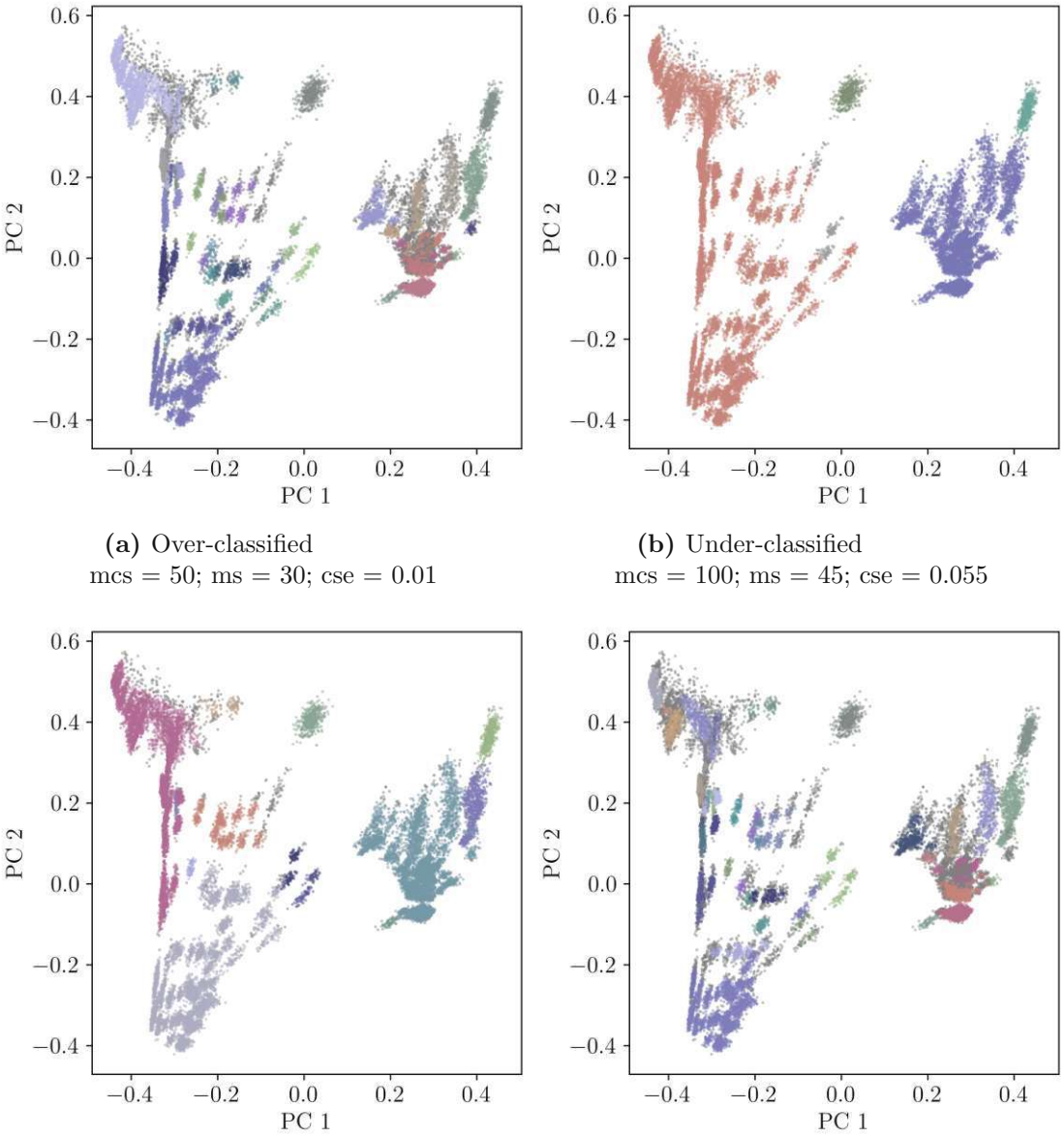
The final step to classify the state points using machine learning is the application of clustering algorithms on the data set. Multiple options are available nowadays, which are suitable for different problems. Based on the given data set and desired outcome, HDBSCAN is selected. As a density-based clustering algorithm, it is robust to any differences in densities or shapes between clusters. Additionally, it requires no prior knowledge of the number of clusters, therefore reduces any possible bias or limitation due to human error. A probability is assigned to each data point whether it is part of any cluster. There are three parameters to adjust for the available dataset to improve clustering using the scikit-learn algorithm [71]. The parameter min_cluster_size (mcs) filters all groups by their size, with any clusters smaller than its value are considered noise. For this density-based algorithm, min_samples (ms) determines whether points are categorized as core points or part of the outer region of a cluster. This choice is based on the minimum number of points in a neighborhood, where min_samples includes the point in question itself. The last parameter adjusted is cluster_selection_epsilon (cse), which serves as a distance threshold for grouping close clusters together. Other parameters are left untouched at their default setting. The HDBSCAN algorithm is applied on the three-dimensional data set spanned by PC 1, PC 2, as well as PC 3. The results are tested for a range of different parameter values expressed in table 3.

parameter	values
mcs	$\{30, 50, 80, 100, 130, 150, 250, 350, 450\}$
ms	$\{1, 5, 10, 15, 20, 30, 45, 60, 80, 100, 200, 300, 400\}$
cse	$\{0.01, 0.015, 0.02, 0.03, 0.035, 0.04, 0.045, 0.05, 0.055, 0.065, 0.07, 0.1\}$

Table 3: The ranges of parameters which were utilized when using the HDBSCAN algorithm on the dataset. The parameter min_cluster_size (mcs) sets a minimum size threshold for clusters, min_samples (ms) the number of neighbors a point must contain to be part of a cluster, and cluster_selection_epsilon (cse) a distance threshold, responsible for grouping clusters together.

The number of unique states which are found manually by visual classification is 16. A minimum cluster size of 350 data points led to the identification of fewer than nine clusters on average, thereby not providing the desired insights within each topology. With the given dataset, the goal is to assign as many data point clouds as possible with at least one cluster. Using the clusters recognized by HDBSCAN, one can create state diagrams. One approach when creating the automatically classified state diagrams, is to ignore points which are recognized as noise (not assigned to a cluster) for the creation of the state diagram. The mode of each parameter combination is calculated to reveal which cluster the systems with a specified Δ-T combination are most commonly assigned to. The results of different HDBSCAN parameters can be broadly categorized into five different groups, all of which are shown in figure 72.
First, there are those HDBSCAN results with a smaller number of clusters than number of visually classified physical states. Their state diagrams for each topology are monotonous, and mostly don't show any differences between states. The resulting state diagrams for each topology are shown in figure 74.
The next group of results shows a much higher number of clusters than expected. Results within this group split the systems into more groups than expected, resulting in potentially over-classified state diagrams with too many different options, which can be seen in figure 73.
The third type of result delivers state diagrams for the three-patch patch topologies (3asym, 2asym1c, and mouse) that are close to what was predicted. This type of clustering fails at the nuanced classification of different gel types found within the VO and 6patch topology, as well as recognizing a difference between the small-cluster liquids and hexagonal liquids of the VN patch topology. This is shown in the state diagrams in figure 75.
Attempting to find a parameter combination that performs better for the other patch topologies has some drawbacks. The three-patch patch topologies are typically either over-classified (figure 76) or a much higher number of data-points are considered noise (figure 77).

Ideally, there would be at least one recognized cluster for each visually identified physical state. To get this result, it appears that some points have to be accepted as noise or to accept over-classification for some physical states. A result closely resembling the manually created state diagrams is achieved using the parameter combination mcs = 250, ms = 100, cse = 0.03. The resulting state diagrams are shown in figure 78.

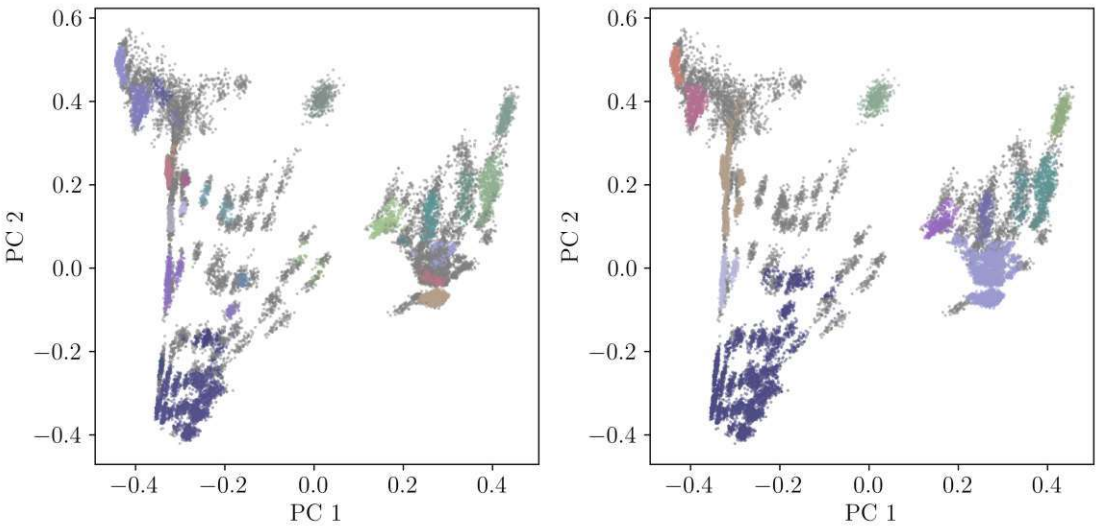


(a) Over-classified
mcs = 50; ms = 30; cse = 0.01

(b) Under-classified
mcs = 100; ms = 45; cse = 0.055

(c) Under-classfied 6patch, VO, VN
mcs = 50; ms = 45; cse = 0.045

(d) Over-classified 3asym, 2asym1c, mouse
mcs = 50; ms = 45; cse = 0.01

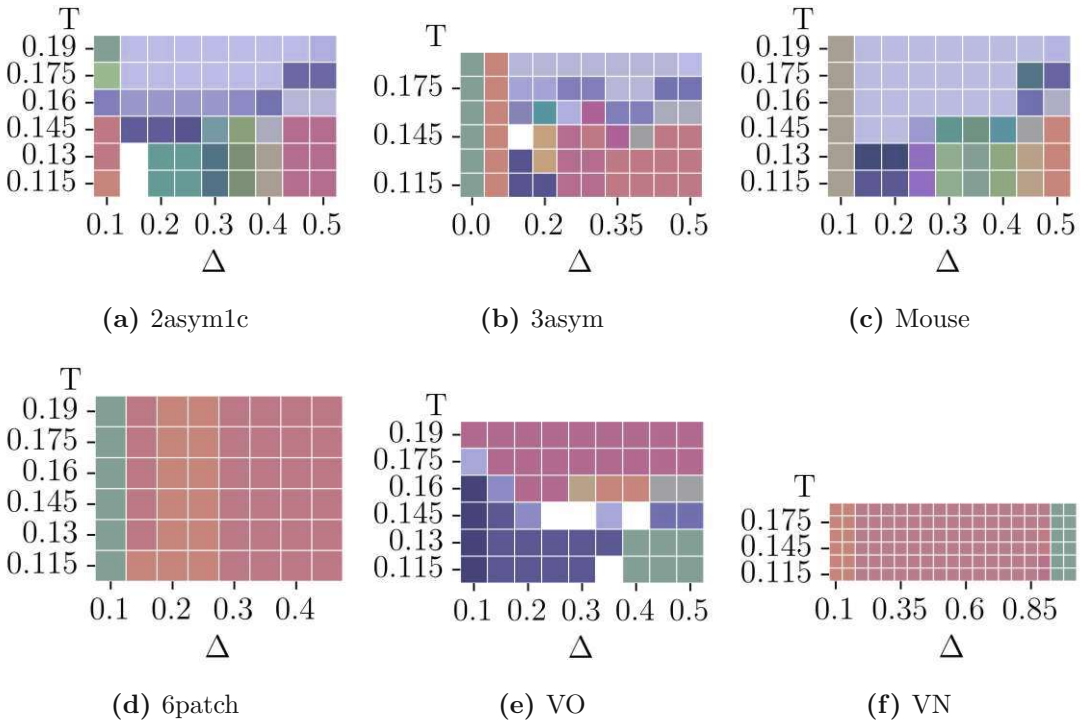


(e) Increased noise
mcs = 30; ms = 100; cse = 0.01

(f) Decent classification
mcs = 250; ms = 100; cse = 0.03

Figure 72: Different HDBSCAN results for different min_cluster_size (mcs), min_samples categorizes (ms), and cluster_selection_epsilon (cse) for $\phi = 0.3$. Each cluster identified by HDBSCAN is drawn in a different color, while noise is drawn in gray.

Using the parameter combination mcs = 50, ms = 30, cse = 0.01, HDBSCAN identifies 54 different clusters.. This is twice the number expected by the preliminary analysis. Especially for the three patch topologies this is particularly evident, as seen in figure 73. If the goal is to detect more nuanced variations, this combination seems promising. In addition to the detailed cluster detection for $\Delta \geq 0.1$ in the three-patch topologies, the chosen parameters effectively separated gel, liquid, and crystal states. For the two-patch patch topologies, this parameter combination performs very well, identifying the variations within VO and VN systems, without running into the problem of over-classification. In the 6patch patch topology, two different gel types are observed at opposite ends of the temperature range. However, these variations are not captured using this parameter combination.

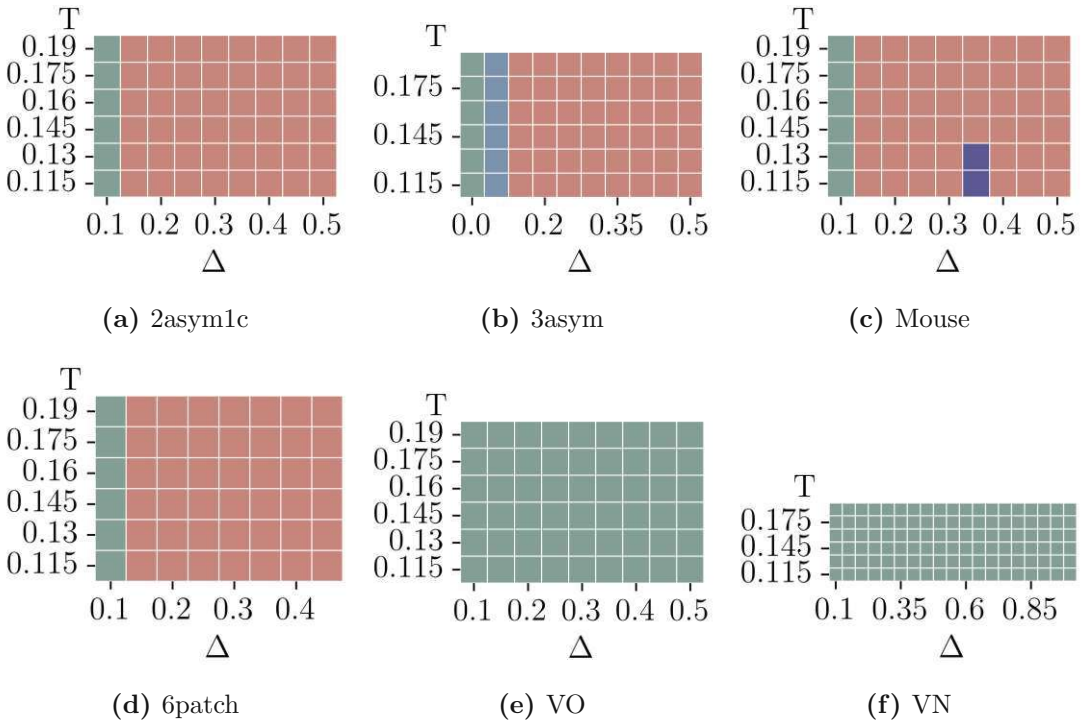


(a) 2asym1c **(b)** 3asym **(c)** Mouse

(d) 6patch **(e)** VO **(f)** VN

Figure 73: HDBSCAN clustering results of PC1, PC2, and PC3 of each system, depicted as a Δ-T state diagram for $\phi = 0.3$. The mode of each Δ-T combination is calculated to determine the most commonly assigned cluster after noise points are removed. Every different shade represents a unique cluster and squares are left empty where none of the systems for this Δ-T combination is assigned to a cluster. In this case, HDBSCAN identified more clusters than expected, especially for the three-patch topologies **(a)-(c)**, leading to an over-classification of the data. The chosen HDBSCAN parameters are mcs = 50, ms = 30, and cse = 0.01.

An opposite result is observed when using the parameters mcs = 100, ms = 45, and cse = 0.055, which can be seen in figure 74. In this case, HDBSCAN only distinguishes between physical states where the patch spillover occurs at $\Delta \leq 0.1$ and the rest. HDBSCAN detects six clusters, which is far fewer than the number of visually determined systems."

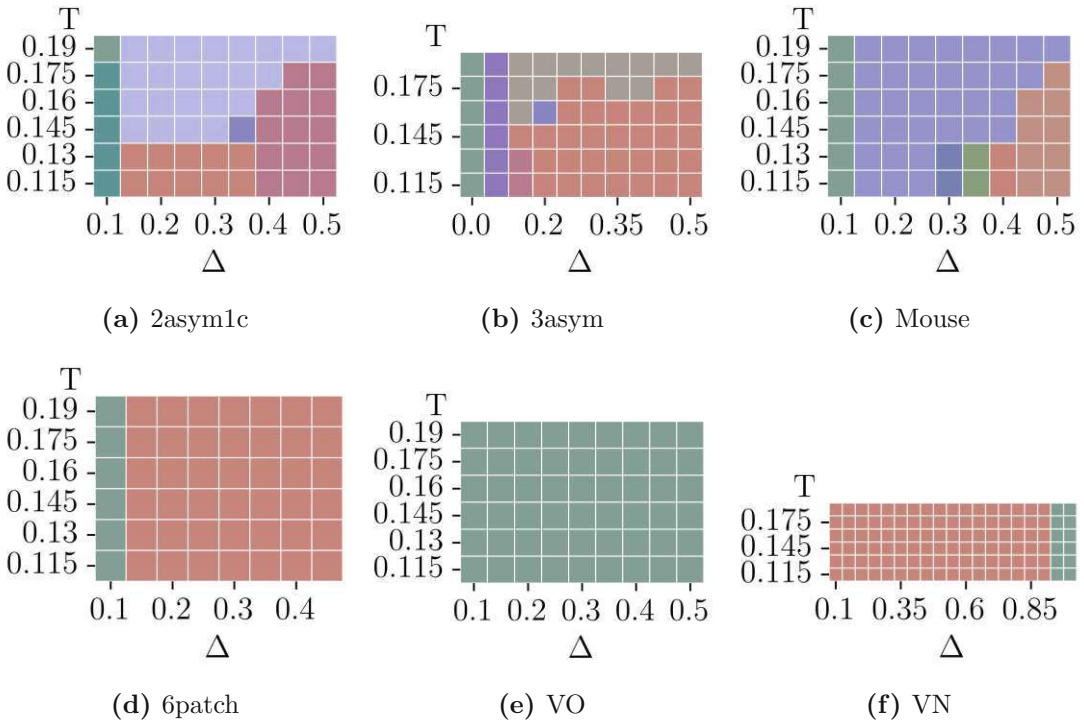


(a) 2asym1c **(b)** 3asym **(c)** Mouse

(d) 6patch **(e)** VO **(f)** VN

Figure 74: HDBSCAN clustering results of PC1, PC2, and PC3 of each system, depicted as a Δ-T state diagram for $\phi = 0.3$. The mode of each Δ-T combination is calculated to determine the most commonly assigned cluster after noise points are removed. Every different shade represents a unique cluster and squares are left empty where none of the systems for this Δ-T combination is assigned to a cluster. In this case, HDBSCAN identified fewer clusters than expected, indicating that this parameter selection is not suitable for investigating this data. The chosen HDBSCAN parameters are mcs = 100, ms = 45, and cse = 0.055.

An intermediate solution is achieved using the parameters mcs = 50, ms = 45, and cse = 0.045, as shown in figure 75. In this case, each of the three patch topologies is only assigned to a few clusters. There is a reasonable detection of the liquid state, the different states at $\Delta \leq 0.1$, and some distinction between porous and close-packed crystals. The other topologies remain under-classified, meaning the clusters found by HDBSCAN contain almost information about the different state variations. This time, HDBSCAN identified 14 different clusters, which is close to ideal, but it is not fully successful for all patch topologies.

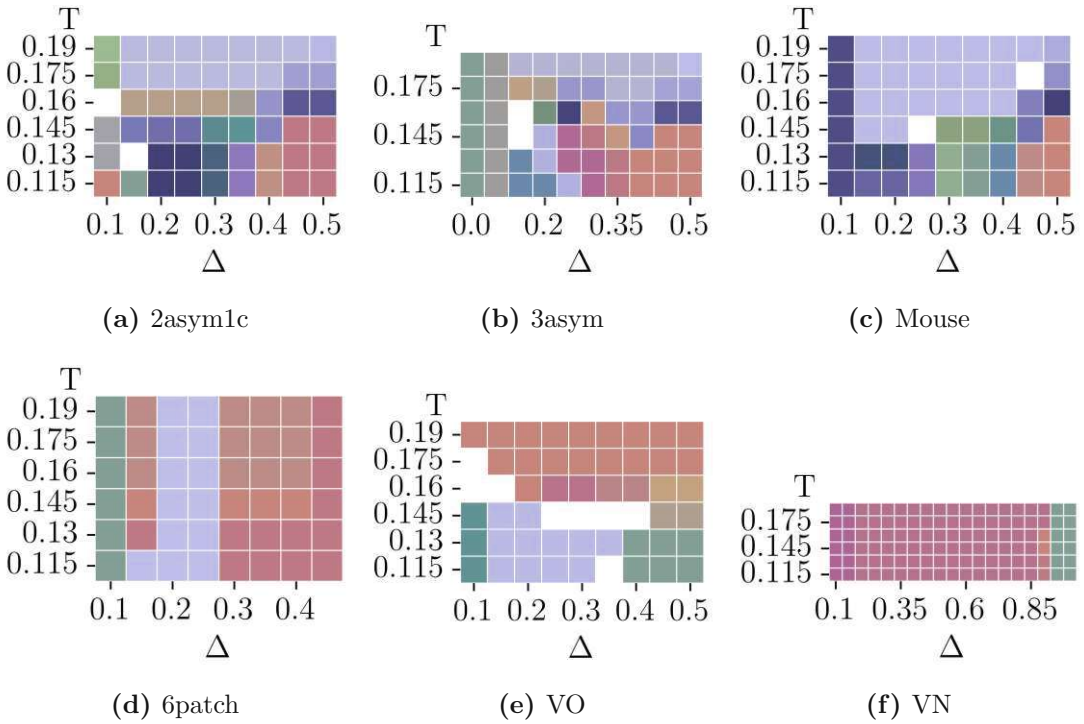


(a) 2asym1c **(b)** 3asym **(c)** Mouse

(d) 6patch **(e)** VO **(f)** VN

Figure 75: HDBSCAN clustering results of PC1, PC2, and PC3 of each system, depicted as a Δ-T state diagram for $\phi = 0.3$. The mode of each Δ-T combination is calculated to determine the most commonly assigned cluster after noise points are removed. Every different shade represents a unique cluster and squares are left empty where none of the systems for this Δ-T combination is assigned to a cluster. In this case, HDBSCAN worked well for the three-patch patch topologies, but failed to identify more than two physical states for the other patch topologies. The chosen HDBSCAN parameters are mcs = 50, ms = 45, and cse = 0.045.

Revealing more information within the two-patch patch topologies and the 6patch topology comes with some difficulties. The combination mcs = 50, ms = 45, and cse = 0.01 seen in figure 76 works better for these patch topologies. Nevertheless, it is not a suitable parameter combination, since the issue of over-classification increases in the case of the three-patch patch topologies, which is reflected by the number of clusters found by HDBSCAN: 50.

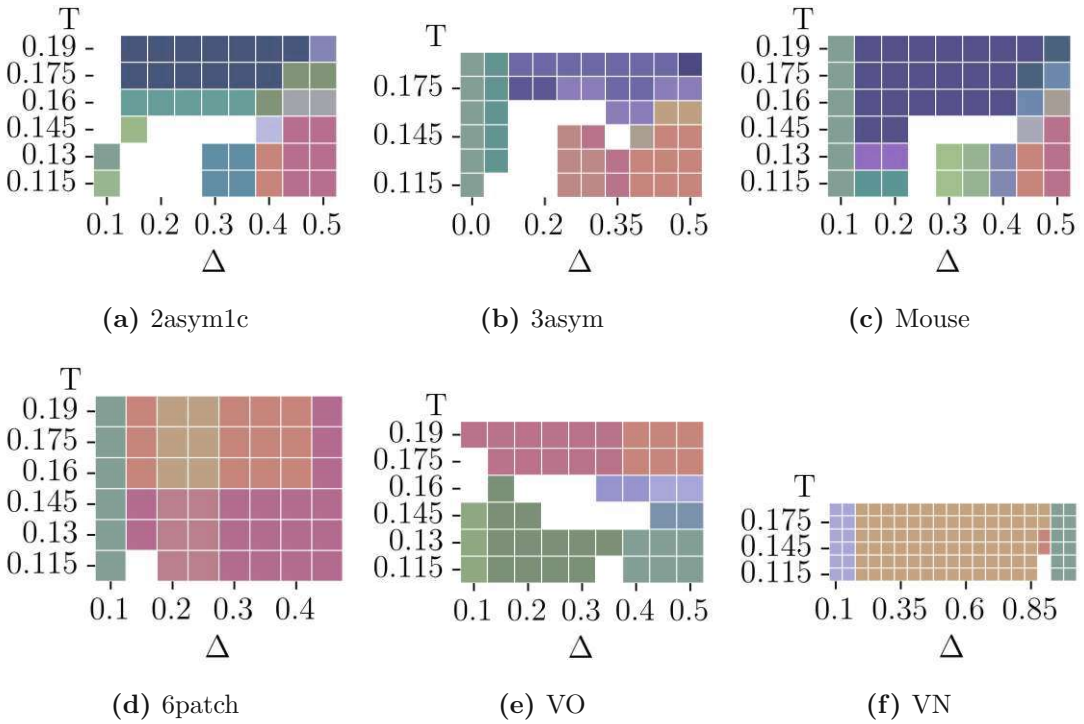


(a) 2asym1c **(b)** 3asym **(c)** Mouse

(d) 6patch **(e)** VO **(f)** VN

Figure 76: HDBSCAN clustering results of PC1, PC2, and PC3 of each system, depicted as a Δ-T state diagram for $\phi = 0.3$. The mode of each Δ-T combination is calculated to determine the most commonly assigned cluster after noise points are removed. Every different shade represents a unique cluster and squares are left empty where none of the systems for this Δ-T combination is assigned to a cluster. In this case, the number of clusters identified by HDBSCAN is very high for the three-patch topologies, but for all others revealed a very well fitting state diagram. The chosen HDBSCAN parameters are mcs = 50, ms = 45, and cse = 0.01.

Tuning the parameters to mcs = 30, ms = 100, and cse = 0.01 results in a clustering with more noise and The results can be seen in figure 77. While some noise is acceptable, this combination of HDBSCAN parameters loses more information to noise than desired, while still over-classifying in some areas. The effect of over-classification has been reduced to 32 clusters found by HDBSCAN.

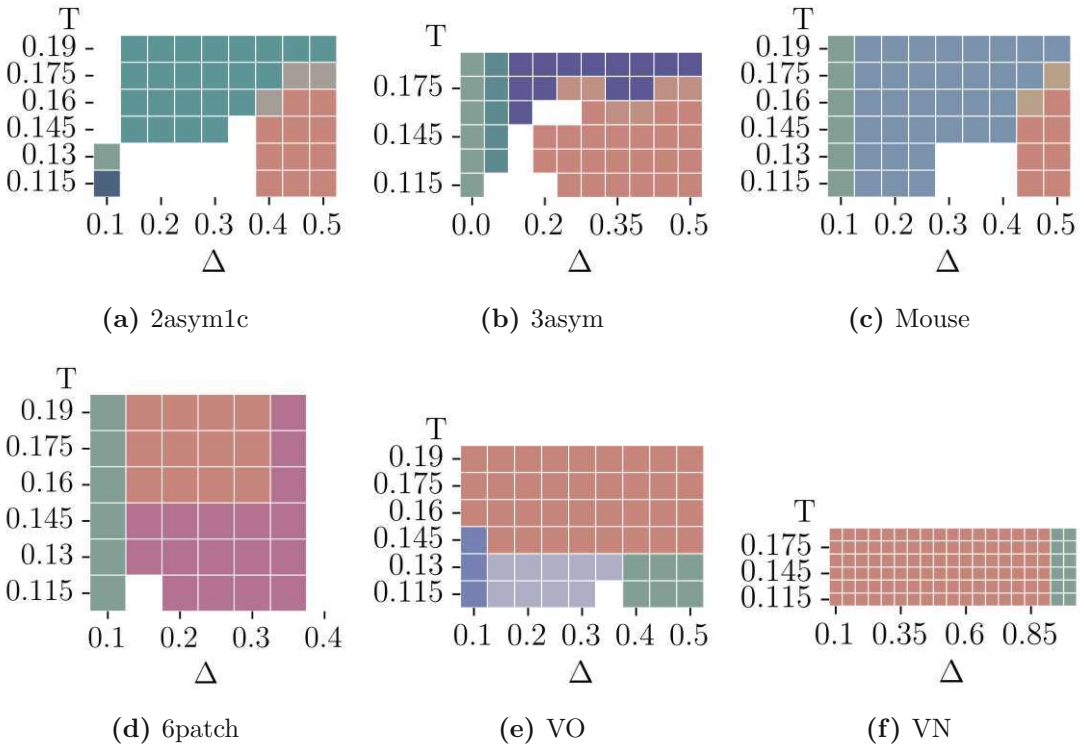


(a) 2asym1c **(b)** 3asym **(c)** Mouse

(d) 6patch **(e)** VO **(f)** VN

Figure 77: HDBSCAN clustering results of PC1, PC2, and PC3 of each system, depicted as a Δ-T state diagram for $\phi = 0.3$. The mode of each Δ-T combination is calculated to determine the most commonly assigned cluster after noise points are removed. Every different shade represents a unique cluster and squares are left empty where none of the systems for this Δ-T combination is assigned to a cluster. In this case, the number of clusters identified by HDBSCAN is very high for the three-patch topologies, but for all others revealed a very well-fitting state diagram. The chosen HDBSCAN parameters are mcs = 50, ms = 45, and cse = 0.01.

The final set of parameters reveals that a broad classification using HDBSCAN is possible on this data set. For each patch topology, a handful of systems are assigned to a small number of clusters when using mcs = 250, ms = 100, and cse = 0.03. The number of detected clusters is equal to thirteen and the collection of state diagrams is shown in figure 78.



(a) 2asym1c **(b)** 3asym **(c)** Mouse

(d) 6patch **(e)** VO **(f)** VN

Figure 78: HDBSCAN clustering results of PC1, PC2, and PC3 of each system, depicted as a Δ-T state diagram for $\phi = 0.3$. The mode of each Δ-T combination is calculated to determine the most commonly assigned cluster after noise points are removed. Every different shade represents a unique cluster and squares are left empty where none of the systems for this Δ-T combination is assigned to a cluster. In this case, HDBSCAN reveals a state diagram close to what is expected, besides the large number of data points which are not classified. The chosen HDBSCAN parameters are mcs = 50, ms = 45, and cse = 0.01.

5 Summary

The goal of this work was to gain an understanding of the self-assembly of patchy triangles while testing the usability of machine learning algorithms for an automated large scale classification. To do so, a collection of different patch topologies was simulated using a Monte Carlo simulation. The self-assembly occurred in a canonical ensemble for different patch positions and temperatures. A handful of order parameters were chosen to quantitatively describe the resulting systems. This data was then investigated using modern machine learning algorithms, which were able to identify clusters of systems within the order parameter space.

Although the selection of patch topologies was not exhaustive for the description of triangular patchy particles, it resulted in a good overview of these systems. The topologies 2asym1c, 3asymc, and mouse behave similarly in the studied parameter ranges: a liquid phase for the higher half of the temperature range and aggregates (gels or compact clusters) at low temperatures. These clusters turn into gels or porous crystals for the peripheral patch placement variations and into increasingly compact aggregates for centered patches. Similar to the previously studied patchy rhombi, it is possible to control the pore size through variation of the distance from patches to a vertex.

The two-patch topologies VO and VN both generated hexagonal patterns. VN systems mostly remained in a liquid state with a large number of hexagon-like clusters without any peripheral patches. The VO patch topology allowed for the self-assembly of a gel type that contained these hexagonal sub-clusters. With a high delta, bonds become possible in the gaps in the hexagonal shape, creating its own type of gel. The hexagonal clusters of the VO topology kept their shape and bonded with others at low temperatures, forming a unique gel. At higher temperatures, this short-range order disappeared and self-assembled into yet another gel type. Lastly, the 6patch topology showed more porous gel types at higher temperatures and close-packed ones at lower temperatures. Due to the rotational symmetry of triangles, there appeared only one type of close-packed tiling, unlike previously studied patchy rhomboids.

Modern machine learning algorithms proved themselves as a useful tool for the classification of such a large variety of systems. With an intentional selection of order parameters, it is possible to quantitatively distinguish between gels, crystals and liquids. In this case the order parameters could be improved to differentiate better between porous and close-packed crystals as well as different gel types.

The positional order parameters suffered from the appearance of small clusters in liquids, which make it seem like ordered systems, as the free particles are not considered in Φ_4 and Φ_6. One solution to deal with this would be the introduction of a cluster size threshold for the calculation of these order parameters. One has to be careful with the implementation of this idea, since there could be states such

as the hexagonal liquid found in the VN topologies, which consists only of small clusters. Another way to improve the understanding of positional order would be the inclusion of the local bond order parameters proposed by H. Eslami [72]. They could reveal a more intricate understanding of different assemblies that go beyond the appearance of square or hexagonal lattice structures. Another idea that came up during this work was to define a position order parameter based on the information contained within the radial distribution function: a linear combination of the relative maximum peak height and the mean squared deviation from 1. This would reveal information on short as well as long range positional order.
The bond order parameter Φ_b also has room for improvement. Crystals simulated in the canonical ensemble can be left with border particles that show a lower number of bonding partners than the bulk of the crystal. The same applies for the end of gels, although the influence on the bond order parameter is most likely smaller here with sufficiently long gels. To deal with this issue the simulation could be done in a grand canonical ensemble, which potentially allows the growth of clusters to a point where there are no more border or dangling particles. Alternatively one could try to identify such particles and skip them during the order parameter calculation.
The orientation order parameter Φ_θ is calculated as the average orientation between connected particles. As long as multiple bonds per patch are allowed, a hexagon connected via vertex patches at its center will therefore deliver different results than a hexagon which is composed of triangle pairs connected via patches along their edges.
It is best to perform a pre-analysis of possible results in order to find suitable order parameters, such that the machine learning algorithms work as desired. The combination of PCA and HDBSCAN found dense groups of systems with similar order parameters and successfully recognized systems of the same type, even if they originated from different patch topologies. Different HDBSCAN parameters result in various levels of noise and classification sensitivity, which can be adjusted to reach the desired clustering. By taking the mode of each Δ-T combination after removing noise, it is possible to assign each combination to a cluster, as long as HDBSCAN assigned at least one point of any given combination to a cluster. Overall, the application of dimensionality reduction in combination with a clustering algorithm has proven itself as a good approach to investigate the large amount of systems.

References

[1] G. M. Whitesides and B. Grzybowski. "Self-Assembly at All Scales". In: *Science* 295.5564 (2002), pp. 2418–2421. URL: https://www.science.org/doi/abs/10.1126/science.1070821.

[2] P. Ball. *The Self-made Tapestry: Pattern Formation in Nature*. Oxford University Press, 2001. ISBN: 9780198502432. URL: https://books.google.at/books?id=Fz1RAAAAMAAJ.

[3] E. Karsenti. "Self-organization in cell biology: a brief history". In: *Nature reviews Molecular cell biology* 9.3 (2008), pp. 255–262.

[4] S. A. Kauffman. *The origins of order: Self-organization and selection in evolution*. Oxford University Press, USA, 1993.

[5] J. M. Lehn. "Toward self-organization and complex matter". In: *Science* 295.5564 (2002), pp. 2400–2403.

[6] J. C. Huie. "Guided molecular self-assembly: a review of recent efforts". In: *Smart Materials and Structures* 12.2 (2003), p. 264.

[7] J. B. Clarke et al. In: *Pure and Applied Chemistry* 66.3 (1994), pp. 577–594. URL: https://doi.org/10.1351/pac199466030577.

[8] A. D. Jenkins et al. In: *Pure and Applied Chemistry* 68.12 (1996), pp. 2287–2311. URL: https://doi.org/10.1351/pac199668122287.

[9] Z. Xu et al. "Liquid-liquid phase separation: Fundamental physical principles, biological implications, and applications in supramolecular materials engineering". In: *Supramolecular Materials* 2 (2023), p. 100049. ISSN: 2667-2405. URL: https://www.sciencedirect.com/science/article/pii/S2667240523000193.

[10] D. Olander. *General thermodynamics*. CRC Press, 2007.

[11] D. W. Oxtoby, H. P. Gillis, and L. J. Butler. *Principles of modern chemistry*. Cengage AU, 2016.

[12] R. P. Sear. "Nucleation: theory and applications to protein solutions and colloidal suspensions". In: *Journal of Physics: Condensed Matter* 19.3 (2007), p. 033101.

[13] D. W. Oxtoby. "Homogeneous nucleation: theory and experiment". In: *Journal of Physics: Condensed Matter* 4.38 (1992), p. 7627.

[14] D. W. Oxtoby. "Homogeneous nucleation: theory and experiment". In: *Journal of Physics: Condensed Matter* 4.38 (Sept. 1992), p. 7627. URL: https://dx.doi.org/10.1088/0953-8984/4/38/001.

[15] W. C. K. Poon. "The physics of a model colloid–polymer mixture". In: *Journal of Physics: Condensed Matter* 14.33 (2002), R859.

[16] M. Li et al. "Ferroelectric phase diagram of PVDF: PMMA". In: *Macromolecules* 45.18 (2012), pp. 7477–7485.

[17] D. Kim et al. "Two-dimensional non-close-packed arrays of polystyrene microspheres prepared by controlling the size of polystyrene microspheres". In: *Polymer* 185 (2019), p. 121938.

[18] Z. M. Sherman et al. "Colloidal nanocrystal gels from thermodynamic principles". In: *Accounts of Chemical Research* 54.4 (2021), pp. 798–807.

[19] M. A. Boles, M. Engel, and D.V. Talapin. "Self-assembly of colloidal nanocrystals: From intricate structures to functional materials". In: *Chemical reviews* 116.18 (2016), pp. 11220–11289.

[20] D. Vanmaekelbergh. "Self-assembly of colloidal nanocrystals as route to novel classes of nanostructured materials". In: *Nano Today* 6.4 (2011), pp. 419–437.

[21] M. Mastrangeli et al. "Self-assembly from milli-to nanoscales: methods and applications". In: *Journal of micromechanics and microengineering* 19.8 (2009), p. 083001.

[22] N. T. K. Thanh, N. Maclean, and S. Mahiddine. "Mechanisms of Nucleation and Growth of Nanoparticles in Solution". In: *Chemical Reviews* 114.15 (Aug. 2014), pp. 7610–7630. URL: https://doi.org/10.1021/cr400544s.

[23] P. W. K. Rothemund. "Folding DNA to create nanoscale shapes and patterns". In: *Nature* 440.7082 (2006), pp. 297–302.

[24] S. Dey et al. "DNA origami". In: *Nature Reviews Methods Primers* 1.1 (2021), p. 13. URL: https://doi.org/10.1038/s43586-020-00009-8.

[25] M. R. Jones, N. C. Seeman, and C. A. Mirkin. "Programmable materials and the nature of the DNA bond". In: *Science* 347.6224 (2015), p. 1260901.

[26] N. C. Seeman. "Nucleic acid junctions and lattices". In: *Journal of Theoretical Biology* 99.2 (1982), pp. 237–247. ISSN: 0022-5193. URL: https://www.sciencedirect.com/science/article/pii/0022519382900029.

[27] M. Karg et al. "Colloidal self-assembly concepts for light management in photovoltaics". In: *Materials Today* 18.4 (2015), pp. 185–205.

[28] A. C. Stelson, W. A. Britton, and C. M. Liddell Watson. "Photonic crystal properties of self-assembled Archimedean tilings". In: *Journal of Applied Physics* 121.2 (2017).

[29] S. Ithurria et al. “Colloidal nanoplatelets with two-dimensional electronic structure”. In: *Nature materials* 10.12 (2011), pp. 936–941.

[30] A. Walther and A. H. E. Muller. “Janus particles: synthesis, self-assembly, physical properties, and applications”. In: *Chemical reviews* 113.7 (2013), pp. 5194–5261.

[31] P. G. De Gennes. “Soft matter”. In: *Science* 256.5056 (1992), pp. 495–497.

[32] Z. Zhang and Sharon C S. C. Glotzer. “Self-assembly of patchy particles”. In: *Nano letters* 4.8 (2004), pp. 1407–1413.

[33] A. B. Pawar and I. Kretzschmar. “Fabrication, assembly, and application of patchy particles”. In: *Macromolecular rapid communications* 31.2 (2010), pp. 150–168.

[34] D. J. Kraft et al. “Self-assembly of colloids with liquid protrusions”. In: *Journal of the American Chemical Society* 131.3 (2009), pp. 1182–1186.

[35] W. Li et al. “Colloidal molecules and patchy particles: Complementary concepts, synthesis and self-assembly”. In: *Chemical Society Reviews* 49.6 (2020), pp. 1955–1976.

[36] C. Karner, F. Müller, and E. Bianchi. “A Matter of Size and Placement: Varying the Patch Size of Anisotropic Patchy Colloids”. In: *International Journal of Molecular Sciences* 21.22 (2020), p. 8621.

[37] G. Tikhomirov, P. Petersen, and L. Qian. “Triangular DNA origami tilings”. In: *Journal of the American Chemical Society* 140.50 (2018), pp. 17361–17364.

[38] D. J. Duchamp and R. E. Marsh. “The crystal structure of trimesic acid (benzene-1, 3, 5-tricarboxylic acid)”. In: *Acta Crystallographica Section B: Structural Crystallography and Crystal Chemistry* 25.1 (1969), pp. 5–19.

[39] S. V. Kolotuchin et al. “Self-assembly of 1, 3, 5-benzenetricarboxylic (Trimesic) acid and its analogues”. In: *Chemistry–A European Journal* 5.9 (1999), pp. 2537–2547.

[40] J. MacLeod. “Design and construction of on-surface molecular nanoarchitectures: lessons and trends from trimesic acid and other small carboxlyated building blocks”. In: *Journal of Physics D: Applied Physics* 53.4 (2019), p. 043002.

[41] C. Karner, C. Dellago, and E. Bianchi. “Design of patchy rhombi: From close-packed tilings to open lattices”. In: *Nano Letters* 19.11 (2019), pp. 7806–7815.

[42] J. M. Dickhout et al. “Produced water treatment by membranes: A review from a colloidal perspective”. In: *Journal of colloid and interface science* 487 (2017), pp. 523–534.

[43] M. Zeng et al. “Hierarchical, self-healing and superhydrophobic zirconium phosphate hybrid membrane based on the interfacial crystal growth of lyotropic two-dimensional nanoplatelets”. In: *ACS applied materials & interfaces* 10.26 (2018), pp. 22793–22800.

[44] A. M. Telford et al. “Mimicking the wettability of the rose petal using self-assembly of waterborne polymer particles”. In: *Chemistry of Materials* 25.17 (2013), pp. 3472–3479.

[45] H. J. Tsai and Y. L. Lee. “Facile method to fabricate raspberry-like particulate films for superhydrophobic surfaces”. In: *Langmuir* 23.25 (2007), pp. 12687–12692.

[46] Z. Cheng et al. “Superhydrophobic shape memory polymer arrays with switchable isotropic/anisotropic wetting”. In: *Advanced Functional Materials* 28.7 (2018), p. 1705002.

[47] H. Cho et al. “Self-assembled stretchable photonic crystal for a tunable color filter”. In: *Optics letters* 43.15 (2018), pp. 3501–3504.

[48] J. Ge and Y. Yin. “Responsive photonic crystals”. In: *Angewandte Chemie International Edition* 50.7 (2011), pp. 1492–1522.

[49] J.B. Kim et al. “Designing structural-color patterns composed of colloidal arrays”. In: *ACS applied materials & interfaces* 11.16 (2019), pp. 14485–14509.

[50] S. H. Kim et al. “Self-assembled colloidal structures for photonics”. In: *NPG Asia Materials* 3.1 (2011), pp. 25–33.

[51] J. Hou, M. Li, and Y. Song. “Patterned colloidal photonic crystals”. In: *Angewandte Chemie International Edition* 57.10 (2018), pp. 2544–2553.

[52] D. Frenkel and B. Smit. *Understanding molecular simulation: from algorithms to applications*. Elsevier, 2023.

[53] K. Kikuchi et al. “Metropolis Monte Carlo method as a numerical technique to solve the Fokker—Planck equation”. In: *Chemical Physics Letters* 185.3-4 (1991), pp. 335–338.

[54] E. Sanz and D. Marenduzzo. “Dynamic Monte Carlo versus Brownian dynamics: A comparison for self-diffusion and crystallization in colloidal fluids”. In: *The Journal of chemical physics* 132.19 (2010).

[55] S. Whitelam and P. L. Geissler. "Avoiding unphysical kinetic traps in Monte Carlo simulations of strongly attractive particles". In: *The Journal of chemical physics* 127.15 (2007).

[56] M. P. Allen, D. Frenkel, and J. Talbot. "Molecular dynamics simulation using hard particles". In: *Computer physics reports* 9.6 (1989), pp. 301–353.

[57] H. B. Callen. *Thermodynamics and an Introduction to Thermostatistics*. John wiley & sons, 1991.

[58] P. J. Steinhardt, D. R. Nelson, and M. Ronchetti. "Bond-orientational order in liquids and glasses". In: *Physical Review B* 28.2 (1983), p. 784.

[59] R. Xu and D. Wunsch. "Survey of clustering algorithms". In: *IEEE Transactions on neural networks* 16.3 (2005), pp. 645–678.

[60] L. McInnes, J. Healy, and S. Astels. "hdbscan: Hierarchical density based clustering." In: *J. Open Source Softw.* 2.11 (2017), p. 205.

[61] R. Bellman and R. Kalaba. "On adaptive control processes". In: *IRE Transactions on Automatic Control* 4.2 (1959), pp. 1–9.

[62] F. Anowar, S. Sadaoui, and B. Selim. "Conceptual and empirical comparison of dimensionality reduction algorithms (pca, kpca, lda, mds, svd, lle, isomap, le, ica, t-sne)". In: *Computer Science Review* 40 (2021), p. 100378.

[63] S. Wold, K. Esbensen, and P. Geladi. "Principal component analysis". In: *Chemometrics and intelligent laboratory systems* 2.1-3 (1987), pp. 37–52.

[64] J. P. Bharadiya. "A Tutorial on Principal Component Analysis for Dimensionality Reduction in Machine Learning". In: *International Journal of Innovative Science and Research Technology* 8.5 (2023), pp. 2028–2032.

[65] S. Surono et al. "Developing an optimized recurrent neural network model for air quality prediction using K-means clustering and PCA dimension reduction". In: *International Journal of Innovative Research and Scientific Studies* 6.2 (2023), pp. 330–343.

[66] E. G. Golshtein and N. V. Tretyakov. "Modified Lagrangians and monotone maps in optimization". In: (1996).

[67] A. Okabe et al. "Spatial tessellations: concepts and applications of Voronoi diagrams". In: (2009).

[68] B. D. Dice et al. "Analyzing Particle Systems for Machine Learning and Data Visualization with freud". In: *Proceedings of the 18th Python in Science Conference*. Ed. by C. Calloway et al. 2019, pp. 27–33. URL: `http://conference.scipy.org/proceedings/scipy2019/bradley_dice.html`.

[69] V. Ramasubramani et al. “freud: A Software Suite for High Throughput Analysis of Particle Simulation Data”. In: *Computer Physics Communications* 254 (2020), p. 107275. ISSN: 0010-4655. URL: http://www.sciencedirect.com/science/article/pii/S0010465520300916.

[70] N. Akkiraju et al. “Alpha shapes: definition and software”. In: *Proceedings of the 1st international computational geometry software workshop*. Vol. 63. 66. 1995.

[71] F. Pedregosa et al. “Scikit-learn: Machine Learning in Python”. In: *Journal of Machine Learning Research* 12 (2011), pp. 2825–2830.

[72] Hossein Eslami, Parvin Sedaghat, and Florian Müller-Plathe. “Local bond order parameters for accurate determination of crystal structures in two and three dimensions”. In: *Physical Chemistry Chemical Physics* 20.42 (2018), pp. 27059–27068.