This is a pre-copyedited, author-produced PDF of an article accepted for publication in *Advances in Complex Systems* following peer review. The version of record, Malte Harder and Daniel Polani, 'Self-organizing particle systems', *Advs. Complex Syst.* 16, 1250089, published October 22, 2012, is available online via doi:

https://doi.org/10.1142/S0219525912500890

Published by World Scientific Publishing.

SELF-ORGANIZING PARTICLE SYSTEMS

MALTE HARDER and DANIEL POLANI

Adaptive Systems Research Group, University of Hertfortshire
College Lane, Hatfield, AL10 9AB, United Kingdom
m.harder@herts.ac.uk, d.polani@herts.ac.uk

September 19, 2016

Abstract

The self-organization of cells into a living organism is a very intricate process. Under the surface of orchestrating regulatory networks there are physical processes which make the information processing possible, that is required to organize such a multitude of individual entities. We use a quantitative information theoretic approach to assess self-organization of a collective system. In particular we consider an interacting particle system, that roughly mimics biological cells by exhibiting differential adhesion behavior. Employing techniques related to shape analysis, we show that these systems in most cases exhibit self-organization. Moreover, we consider spatial constraints of interactions, and additionaly show that particle systems can self-organize without the emergence of pattern-like structures. However, we will see that regular pattern-like structures help to overcome limitations of self-organization that are imposed by the spatial structure of interactions.

Keywords: Self-organization; information theory; morphogenesis.

1 Introduction

The development of organisms is one of the most prominent examples of self-organization and the emergence of shapes. The process of forming shapes is usually an interplay between environmental dynamics (e.g. global physical rules), and agent actuations (e.g. a change of local properties) regulated through complex networks.

In the early stage of laying out body plans, morphological changes are induced mainly due to control of cell adhesion, cell motility and oriented cell division. In particular, differential cell adhesion prevents areas consisting of different tissues to mix and starts an automatic sorting process. This happens, if for example cells have been forced to mix in a solution [47]. Gastrulation, the

1

process of rearranging a ball of cells in the early stage of embryonic development into a more complex body structure, can be simulated by contractions in cell shape that then lead to an automatic rearrangement of cells forming an inner structure [26].

One important aspect of all these processes is that, in many cases, the information processing capabilities of the individual cells (i.e. agents) are severely limited, especially in scenarios that consider large collectives. In these cases the environmental dynamics dominate the process of organization while individual agents actively guide the process. Cells for example can change adhesion properties or partially contract. *Morphogenesis*, the formation of shapes, as we will see, can be achieved purely by environmental dynamics up to certain limits. The process of shape formation can be seen as a selection of a configuration which fulfills certain properties. Thus, the course of a given process typically leads to a reduction of entropy. In the context of this paper, we would like to reinterpret this as saying that there are information processing capabilities in the environment. This is justified by the view of the controlled dynamics of a system as an entropy reduction mechanism [43], more below.

Note that these capabilities are often rooted in the structure of the space and the physical laws that govern it. In [31] it has been shown that consistency in the embodiment of agents reduces cognitive load, lack of such consistency increases it. This implies that, consistency or homogeneity of the space also can increase the information processing capabilities of the environment, as a reduction of cognitive load means that the information needs to be processed elsewhere. Information processing/entropy reduction capabilities that a system provides can also be used by non-reactive systems (for example, we consider particles here instead of autonomous agents). In particular they can be a driving force of self-organization.

1.1 Self-organizing particle systems

In order to investigate the information processing capabilities of a morphogenetic process, we will use a model of particle collectives similar to the models in [10, 11, 35] that mimics features of cell adhesion and cell motility to a certain grade. A human observer easily detects organizational patterns in simulation runs of this model. In many cases the resulting particle configurations even resemble the morphology of biological structures, showing features that look like membranes or nuclei, see Fig. 1 for an example. However, a human observer is a quite subjective measure and not transferable. The question is: Is there an objective way to characterize whether a formation of particles is a result of mere chance, some driving force or is it due to the underlying dynamics that did let the system self-organize? Using quantitative methods it is possible to investigate such a formation process in greater detail. In particular information theory will serve us as a useful tool as it offers a universal and model agnostic way to investigate self-organization in arbitrary dynamical systems.

In self-organizing processes, individual parts of the whole system usually interact with each other, this is the case in particular in the particle model

considered here. Interactions have been the basis for information theoretic investigations before [19], and can be closely linked to information storage and transfer [44, 24, 5]. A requirement that organization can occur is the spread of information through the system, which in turn requires interaction between individual parts of the system [40].

2 Information Theory

Our studies will be based on information theory [38]. For self-containedness, we introduce here the basic notions of information theory, for a detailed account see Thomas and Cover [6]. A fundamental measure in information theory is the *entropy* of a probability distribution, measured in bits. It is defined by

$$H(X) := -\sum_{x \in \mathcal{X}} p(x) \log_2 p(x) \tag{1}$$

where X denotes a finite-valued random variable with values in $\mathcal{X}$ and $p(x)$ the probability that X takes on the value $x \in \mathcal{X}$. Entropy measures the uncertainty of the outcome of a random variable. Now *mutual information* is defined by $I(X;Y) := H(X) + H(Y) - H(X,Y)$. It is symmetric in X and Y and can be interpreted as the information the random variables mutually encode about each other. For continuous random variables, differential entropy is defined as

$$h(X) := \int_{\mathcal{X}} p(x) \log_2 p(x) dx. \tag{2}$$

Differential entropy differs in important aspects from its discrete counterpart and is, strictly spoken, not a true, but rather a "renormalized" entropy [16]. The analogous concept of differential mutual information, however as difference of entropies $I(X;Y) := h(X) + h(Y) - h(X,Y)$, still retains its character as a mutual information and discrete variables can be seamlessly replaced by continuous ones in the case of the latter.

2.1 Multi-information

Multi-information, also called total correlation [17] or integration [42], is one generalization of mutual information to the multivariate case. It is defined as

$$I(X_1, ..., X_n) := \sum_{i=1}^{n} H(X_i) - H(X_1, ..., X_n) \tag{3}$$

where the sum goes over the entropies of the marginal variables. The continuous case is defined analogously. Multi-information measures the statistical dependence between the individual parts of the whole system. Multi-information has been used as a complexity measure [1] and incorporates the idea that changes in one part of the system are reflected in the overall state.

There are several other multivariate generalizations of mutual information which all retain certain properties of the bivariate case. For a thorough discussion of most of these measures, especially in the context of analyzing time series, see [17]. Here we will use multi-information to measure the amount of self-organization that is present in the system.

3   Quantifying Self-Organization

Although it is quite easy for humans, from a visual standpoint, to point out whether we consider a system as self-organizing ("I know it when I see it"), there are surprisingly few quantitative characterizations of self-organization. Let us assume we observe some pattern that seems to self-organize over time, but we only see a single instance of this time series. Our intuition about self-organization implicitly assumes that there is a causal relation between the observed time-steps. However, this could be the improbable but possible observation of a time series of i.i.d random variables. In this case we do not want to quantify the system as self-organizing. By choosing a statistical or information theoretic formulation of self-organization we are able to account for this causal relationship.

Let us first consider what self-organization is not. Most importantly, there should not be an external or central force that controls the process in a "top-down" fashion. Also, if there is a system that is causally dependent [28, 2] on another self-organizing system, we do not want to consider the former as self-organizing. This means that the system should be autonomous, which is not a trivial task to determine [3], and external influences need to be separable, which can be investigated using the measure of causal information flow [2]. However, we will not consider these problems here, they are more apparent if the underlying model is unknown. Here, the model is designed to fulfill these criteria.

This still leaves the question of how to quantify organization. Organization means an increase of structure over time, whereby structure in this context usually is considered to be a spatial correlation between parts of the whole system. One account for a definition of self-organization can be found in [37], where self-organization is defined as the increase of statistical complexity of a system over time via ϵ-machines. Measuring self-organization via statistical complexity has the limitation and advantage that it assumes no structure on the space underlying the time-series. It is an advantage because the measure is very versatile and does not need to be changed for different spatial structures. On the other hand its a limitation, because the spatial structure will be implicitly encoded in the states of the ϵ-machine, which makes it less accessible for further analysis [37]. There is an extension of ϵ-machines called spatial ϵ-machines which have a structural assumption about the space they are defined on, but this has only been done for discrete spaces so far. Thus, we will describe a method which has been suggested as alternative to the measure of self-organization via statistical complexity (see [30] for a detailed survey on the foundation of self-

organization).

3.1 Self-organization via observers

We define self-organization as an increase of multi-information between observer variables $I(X_1, ..., X_n)$ during the run of the process [29]. A collection of random variables $X_1, ..., X_n$ are called *observers* of a system X if the system is fully determined by the collection of random variables, that means $H(X|X_1, ..., X_n)$ vanishes, and all variables only depend on X, meaning $H(X_1, ..., X_n|X)$ vanishes as well. The observers impose a "coordinate system" on the observed random variable and can be chosen freely. The additional specification is not very problematic in practice, as there are often natural choices for observer variables, as many systems are just a collection of individual random variables.

For a completely random process this measure never detects any self-organization, as there is no increase in correlation between observer variables. Note that there could be correlation between observers because they observe the same part of X, but because they exclusively depend on X, there is no hidden common cause between observers that could increase the correlation. The other extreme case is that the entropy of the whole system vanishes, in which case there also cannot be an increase of multi-information between the observers. So, to achieve self-organization, the system requires some remaining, degree of freedom.

Interestingly, this definition also gives the opportunity to build hierarchies by considering coarse to fine grained observers, which then leads to a decomposition of self-organization. If k-observers are grouped to one coarse-grained observer variable $\tilde{X}_i$ we get that the following multi-information term

$$I(\underbrace{X_1, ..., X_{i_1}}_{\tilde{X}_1}, \underbrace{X_{i_1+1}, ..., X_{i_2}}_{\tilde{X}_2}, ..., \underbrace{X_{i_{k-1}+1}, ..., X_{i_k}}_{\tilde{X}_k}) \tag{4}$$

can be decomposed into $k+1$ multi-information terms (see [30] for a derivation)

$$= I(\tilde{X}_1, ..., \tilde{X}_k) + \sum_{j=1}^{k} I(X_{i_{j-1}}, ..., X_{i_j}). \tag{5}$$

The decomposition now allows the separation of organization that is apparent within individual parts of the system, where each part is a coarse-grained observer, and organization, that can only be explained as an interaction between coarse-grained observers.

This allows us to see whether there are parts of the system that dominate the process of self-organization. For example by grouping individual observers by common properties of the observed variables, it is possible to see whether a specific property has a higher contribution to the self-organization or whether most of the self-organization is a result of interaction between different coarse-grained observers. We will now use this formulation of self-organization to investigate spatial self-organization in particle collectives.

4 Particle collectives & self-organisation

4.1 The Particle Model

There are numerous models of morphogenesis and pattern formation including reaction-diffusion models [25], cellular automata [46], diffussion-limited aggregation [45], L-systems [33] and agent based models [4]. The particle model we will describe is based on the model by [10, 11], and shares some similarities with the Swarm Chemistry model [35]. It mimics the way biological cells stick together by cell adhesion, allowing different types to recognize each other.

In our model, each particle interacts with all particles within a certain cut-off radius r_c. For reasons of simplicity, as well as to be able to have long range interactions, we ignore a cell-like tessellation (as opposed to [10]), where interactions can only take place between direct neighbors of the tessellation. The equation of motion for each particle is given by

$$\dot{z}_i = \sum_{j \in N_{r_c}(i)} -F_{\alpha\beta}(\|\Delta z_{ij}\|_2)\Delta z_{ij} + w \tag{6}$$

where $\Delta z_{ij} = z_i - z_j$, $N_{r_c}(i)$ denotes the set of indices of particles within radius r_c of particle i and $F_{\alpha\beta}$ is a force-scaling function, α the type of particle i, β the type of particle j and w an additive white Gaussian noise term, where $w \sim \mathcal{N}(0, 0.05)$ throughout all experiments. Note that the velocity is proportional to the force applied and thus the dynamics are studied in the strong limit of friction. This assumption holds for example for the motion of insects and cellular motility, in contrast to the movement of larger animals and humans which can build up momentum.

Now, the equation of motion can be solved using Euler-Maruyama integration [20, 32]. We used the following two force-scaling functions, see Fig. 2 for a plot of both functions.

$$F_{\alpha\beta}^1(x) = k_{\alpha\beta}\left(1 - \frac{r_{\alpha\beta}}{x}\right) \tag{7}$$

and

$$F_{\alpha\beta}^2(x) = k_{\alpha\beta}\left(\frac{1}{\sigma_{\alpha\beta}^2}e^{-\frac{x^2}{2\sigma_{\alpha\beta}}} - e^{-\frac{x^2}{2\tau_{\alpha\beta}}}\right). \tag{8}$$

The matrices $k_{\alpha\beta}, r_{\alpha\beta}$ as well as $\sigma_{\alpha\beta}$ and $\tau_{\alpha\beta}$ define the interactions between the particles and have a strong impact on the dynamics of the experiment. Values for the parameters were chosen from the following ranges: $k_{\alpha\beta} \in [1.0, 10.0]$, $r_{\alpha\beta} \in [0.0, 1.0]$ and $\tau_{\alpha\beta} \in [1.0, 10.0]$ with $\sigma_{\alpha\beta} = 1$ throughout all experiments. Note that choosing a non-symmetric matrix often leads to unstable dynamics or cycling patterns as the preferred distance is mutually different, we therefore only consider symmetric matrices in what follows. The force-scaling function defines how much attraction or repulsion the particles show among each other depending on the type and distance between particles. The first force-scaling function F^1 shows stronger attraction compared to repulsion than F^2. For each type, the force-scaling functions result in a preferred distance of particles

of other types, denoted $r_{\alpha\beta}$. By using smaller diagonal values than the off-diagonal elements in $k_{\alpha\beta}$ or $r_{\alpha\beta}$ it is possible to force clustering of particles of the same type.

In Fig. 3 are three examples of equilibrium states of particle collectives. For the particle collective with only one type, a simple disc shaped pattern can be seen. The collective is considered to be in equilibrium, if for several time steps the sum of the L^2 norm of the sum of all forces acting on each particle is below a specific threshold.

4.2   Measuring organization in particle collectives

To measure self-organization within a particle collective using multi-information as introduced in Sec. 3.1 we need to define observer variables. A natural choice would simply be the collection of variables denoting the positions of each individual particle. However, we have to consider that certain transformations of the configuration leave the shape of the particle collective invariant. So, if we do not consider these invariants, the measured multi-information can be different from what we do want to consider as organization towards a shape. But even if, in the stochastic limit, rotations and translations are equidistributed, by factoring them out we reduce the sparsity of samples in the space of possible configuration of particles.

There are several accounts on spatial statistics and stochastics [36, 23], however in these references interacting particle systems are defined as (continuous time) Markov processes on discrete domains while our experiments are in the continuous domain. In the area of geo-information systems and medical image processing, there is a large interest in statistical models of shapes, and there is a large body of literature on shape models [9, 39, 12]. One particular problem, the alignment of overlapping images or shapes, is similar to the problem of reducing our experiment samples (i.e. the simulations) to an invariant representation.

Rotation, translation as well as permutation of particles of the same type leave the observable shape, as well as the dynamics involved, invariant. Let $ISO^+(2)$ denote the group of direct isometries (rotation, translation and identity) of the euclidean plane. This group now acts on the space of particle configurations $\mathcal{Z}$ by rigid body motions:

$$\mathcal{Z} \times ISO^+(2) \rightarrow \mathcal{Z}. \tag{9}$$

To account for permutations, let S_n denote the permutation group of n-elements, which also naturally acts on the space of samples by permuting the particle vectors for all time steps. Now we consider the subgroup $S_n^* \subset S_n$ that permutes only particles of the same type. The direct product $F = ISO^+(2) \times S_n^*$ now classifies all shape invariant transformations.

It is important to note that these transformation also have to property that they leave the dynamics of the system invariant. Let $\mathbf{z}^{(t)} \in \mathcal{Z}$ denote the configuration of the particle collective at time t, then

$$p(\mathbf{z}^{(t)}|\mathbf{z}^{(t-1)}) = p(f\mathbf{z}^{(t)}|f\mathbf{z}^{(t-1)}) \text{ for all } f \in F \text{ and all } \mathbf{z}^{(t)}, \mathbf{z}^{(t-1)}. \tag{10}$$

7

This means that a configuration that is transformed will lead to a distribution of configurations in the future, that is equivalent to the distribution of the transformed future states of the original configuration.

In the case that additionally the initial state is invariant under the action of this transformation group, we also have $p(\mathbf{z}^{(t)}) = p(f\mathbf{z}^{(t)})$ for all t. Thus we can easily factor out the transformation group, and get random variables over the space of shapes (transformation invariant particle configurations). Factoring out all symmetries F from $\mathcal{Z}$ then leads to a reduced space of particle configurations $\mathcal{W}$ which then allows us to define a random variable $\mathbf{W}^{(t)}$ (the whole collective at time t) and corresponding observer variables $\mathbf{W}_1^{(t)}, ..., \mathbf{W}_n^{(t)}$ for a collective with n particles. The indistinguishableness of particles of the same type means that no observer $\mathbf{W}_i^{(t)}$ can be used to make predictions about the future of a specific particle. Measuring multi-information on these derived random variables $\mathbf{W}_1^{(t)}, ..., \mathbf{W}_n^{(t)}$ ignores certain degrees of freedom, i.e., rotation, permutations of particles of the same type and translation. Now we can express every configuration of particles $\mathbf{z}$ as a permutation, translation and rotation of invariant coordinates $\mathbf{w}$, i.e. for all $\mathbf{z}$ there exists $\mathbf{w}$ and $f \in F$ such that $\mathbf{z} = f\mathbf{w}$. Due to the group structure of F and the invariance of the states (at all times) under transformations of F we have

$$
\begin{aligned}
I(\mathbf{Z}_1, ..., \mathbf{Z}_n) &= \int_{\mathcal{Z}} p(\mathbf{z}_1, ..., \mathbf{z}_n) \log \frac{p(\mathbf{z}_1, ..., \mathbf{z}_n)}{p(\mathbf{z}_1) \cdots p(\mathbf{z}_n)} d\mathbf{z} & (11)\\
&= \int_F \int_{\mathcal{W}} p(f(\mathbf{w}_1, ..., \mathbf{w}_n)) \log \frac{p(f(\mathbf{w}_1, ..., \mathbf{w}_n))}{p(f\mathbf{w}_1) \cdots p(f\mathbf{w}_n)} d\mathbf{w} df & (12)\\
&= \int_F \int_{\mathcal{W}} p(\mathbf{w}_1, ..., \mathbf{w}_n) \log \frac{p(\mathbf{w}_1, ..., \mathbf{w}_n)}{p(\mathbf{w}_1) \cdots p(\mathbf{w}_n)} d\mathbf{w} df & (13)\\
&= I(\mathbf{W}_1, ..., \mathbf{W}_n). & (14)
\end{aligned}
$$

Therefore, factoring out the transformation group F does not change the multi-information of the observers, in the case of an invariant system. However, that means that in practice we would have to sample the initial state with full support on $\mathbb{R}^2$ which would lead to a very sparse sampling and because of the finite cut-off radius to almost no interactions between particles. To avoid this, we use an initial distribution of particles, which is uniform within a certain radius around the origin, so that particles are initially placed uniformly on a centered disc. This initial distribution is still invariant with respect to rotation and permutation, but not translation invariant. Even though the equality from above thus does not hold anymore, we argue, that due to the symmetry of the forces ($r_{\alpha\beta}, k_{\alpha\beta}, \sigma_{\alpha\beta}, \tau_{\alpha\beta}$ are symmetric in all cases) and the independence of the noise, there is no information coming from translation in the original coordinates $\mathbf{z}^{(t)}$ that is relevant for the organization of the collective, so that we can simply consider the factored coordinates $\mathbf{w}^{(t)}$ to measure multi-information and successively self-organization.

4.2.1 Indistinguishable particles

If we make predictive statements about a particle we need to be able to identify particles through time, otherwise the statistics about the future of a particular particle are skewed. That the interchangeability of variables has an impact on information processing and measurements has been considered before in terms of recoding equivalence [7]. By reordering the particles, we lose the information to identify the same particle over time and they become indistinguishable. To measure self-organization of shapes we actually want indistinguishable particles (if they have the same type) and therefore we introduced the permutation group S_n^* as one set of shape invariant transformations. Distinguishing them would mean that there can be an event that increases the measurement of self-organization, but is not reflected in the shape and structure of the particle configuration. For example there could be a permutation of two particles of the same type that is always reflected by a permutation of two particles of same type elsewhere in the system. This would then be taken into account by the multi-information, but has no impact on the shape that is formed.

On the other hand we also do not want to equate particles which have a different type, and show different interactions. Particles of different types should be distinguishable as permutations of particles of different type actually change the shape of the configuration. Additionally, if they would be indistinguishable this would contradict the assumption made in (10) that the dynamics are invariant to permutations of indistinguishable particles.

The problem of indistinguishable particles and the related change of entropy is also a problem in thermodynamics where it is known as Gibbs phenomenon and Mixing paradox [13, 18]. By only making a distinction between particles that show observably different behavior we are in line with the solution to this problem in physics [18].

5   Methods

This section describes how we derive estimates of multi-information from simulation samples of particle dynamics. The reader not interested in the computational details may skip this section to Sec. 6 without loss of understanding.

5.1   Particle and sample space

Each simulation runs with a fixed number of n particles, l different types and each particle gets a fixed type assigned at the start of the simulation. The particles are located in the infinite two-dimensional plane $\mathbb{R}^2$ and are initialized with a uniform distribution on a disc of fixed radius. Each particle is of a specific type. The types can vary between different experiments, but the properties ($r_{\alpha\beta}$, etc.) of each type are fixed for all simulation samples of one experiment. The assignment of a type to a particle is fixed over the time of the simulation run.

9

Each simulation run is a sample and is denoted by

$$\bar{\mathbf{z}} = (\mathbf{z}^{(1)}, ..., \mathbf{z}^{(t_{\max})}) \tag{15}$$

where each time step is a vector of particle coordinates

$$\mathbf{z}^{(t)} = (z_1^{(t)}, ..., z_n^{(t)}). \tag{16}$$

To gather statistics for an experiment, we need to run the simulation multiple times. The collection of all m samples is denoted

$$\mathfrak{z} = (\bar{\mathbf{z}}_1, ..., \bar{\mathbf{z}}_m)^\top = (\mathfrak{z}^{(1)}, ..., \mathfrak{z}^{(t_{\max})}). \tag{17}$$

Now, let the space of all particle vectors $\mathbf{z} = (z_1, ..., z_n)$ be denoted $\mathcal{Z}$ and $\mathbf{Z}^{(t)}$ the random variable over $\mathcal{Z}$ at time step t, so all $\mathbf{z}^{(t)} \in \mathfrak{z}^{(t)}$ are samples of $\mathbf{Z}^{(t)}$.

5.2 Factoring out symmetries

Now we factor out the symmetries for each time step as introduced in Sec. 4.2. The samples $\mathbf{z}^{(t)} \in \mathfrak{z}^{(t)}$ for each time step t, the raw output of the simulations, are still with respect to a common coordinate system. We will now proceed to factor out translations, rotations and permutations resulting in processed samples $\mathbf{w}^{(t)} \in \mathfrak{w}^{(t)}$ for each time step t. In practice this is done by expressing all particle configuration samples $\mathbf{z}^{(t)} \in \mathfrak{z}^{(t)}$ with respect to its centroid. This is followed by aligning all configuration samples $\mathbf{z}^{(t)}$ for each time step using an ICP (Iterative Closest Point) algorithm [48, 34]. For the application of the alignment the particle configuration is transferred to a three dimensional representation where the third coordinate of each particle is represented by its type, where the type coordinates are scaled by a factor a magnitude larger than the diameter of the collective. Thus the alignment respects the type of the particles. After the alignment the coordinates of all particle are reordered by types and correspondences. Correspondences between particles of different samples, but of the same type, are found using a nearest neighbor search within the ICP algorithm (implementation from the point cloud library [34]). This means that particles close to each other in different samples at the same time are considered to represent the same particle. Note that the notion of same particle establishes a correspondence between different samples at a specific time step. The correspondence between particles of the same sample, but different time steps is, however, lost in this process.

Equipped with this preprocessing, we reach an isometry- and permutation-reduced representation of the particle collective in terms of processed samples $\mathbf{w}^{(t)} \in \mathfrak{w}^{(t)}$. We can now use the statistics of these samples to calculate the multi-information $I(\mathbf{W}_1^{(t)}, ..., \mathbf{W}_n^{(t)})$.

The invariant representation also has the advantage that the samples are much denser in the space of possible configurations which improves the quality of the estimates. It is important to note, that for statistics that need to track particles over time, we cannot use the permutation reduced representation because we would then lose any correspondence of particles over time, e.g. [21].

5.3 Estimation of multi-information

To estimate the multi-information we used the Kraskov-Stögbauer-Grassberger Estimator [22]. The estimate is based on a k-nearest-neighbor search. We chose $k = 5$ for all experiments. The estimate is not very sensitive for changes of k and we get similar results with $k = 2$ or $k = 10$. If k is chosen too large the resulting estimate vanishes for almost all samples. The estimator for m samples and n variables is given by

$$I(\mathbf{W}_1^{(t)}, ..., \mathbf{W}_n^{(t)}) \triangleq \psi(k) + (n-1)\psi(m) - \langle \psi(c_1) + \psi(c_2) + ... + \psi(c_n) \rangle_{\mathbf{w}^{(t)}} \quad (18)$$

where ψ is the digamma function and the brackets denote the average taken over all samples. The c_i are dependent of $\mathbf{w}^{(t)}$ and defined as follows: let $N_k(\mathbf{w}^{(t)})$ denote the k-th neighbor of the sample $\mathbf{w}^{(t)}$ using the following metric

$$\|\mathbf{w}' - \mathbf{w}\| := \max_{i \in \{1, ..., n\}} \|\mathbf{w}'_i - \mathbf{w}_i\|_2 \quad (19)$$

then c_i is defined as

$$c_i = \left| \left\{ \mathbf{w}'^{(t)} \in \mathfrak{w}^{(t)} : \|\mathbf{w}'^{(t)}_i - \mathbf{w}^{(t)}_i\|_2 < \|N_k(\mathbf{w}^{(t)})_i - \mathbf{w}^{(t)}_i\|_2 \right\} \right| - 1. \quad (20)$$

The idea is that a high correlation between the variables leads on average to a low count of samples per variable that, for each sample, are closer to the sample itself than the k-th neighbor over all variables, thus maximizing the estimator.

In our tests this approximation shows less variance and is much faster to calculate than other methods. This holds especially in the higher dimensional setting with more than ten particles (20 dimensional) and only few samples (500 to 1000). We compared the method to a kernel based approach which was multiple orders of magnitudes slower and showed a larger variance in higher dimensions [41]. We also compared it to a shrinkage type binning estimator [15], which overestimated the multi-information in higher dimension due to the sparse sampling, so much that almost no change in information could be seen.

5.3.1 A further approximation

For large collectives, the alignment of samples and the estimation of the multi-information can still be a computationally expensive task. Therefore, we can reduce the dimensionality of the problem by introducing mean random variables. We perform a k-means clustering on the particles of each type and thus recover $l \cdot k$ mean variables $\hat{\mathbf{W}}_1^{(t)}, ..., \hat{\mathbf{W}}_{lk}^{(t)}$, where l is the number of types. Now we use $I(\hat{\mathbf{W}}_1^{(t)}, ..., \hat{\mathbf{W}}_{lk}^{(t)})$ as an approximation measure for the multi-information $I(\mathbf{W}_1^{(t)}, ..., \mathbf{W}_n^{(t)})$.

This must be done carefully, because the clustering process itself can introduce structure into the collective of particles, and thus can lead to a higher measurement of multi-information than actually is present. On the other hand, the clustering ignores all small scale self-organization processes, and hence the measured multi-information is less than the exact value.

6 Results

In the following sections all experiments were run with a sample size of $m = 500$ to $m = 1000$, $t_{\max} = 100$ to $t_{\max} = 250$ and $k = 4$ for the knn-selection of the multi-information estimator. For systems with more than 60 particles the k-means clustering approximation was used to reduce the number of dimensions. The first observation is that this method can be used in practice to detect self-organization and there is a visual correlation between the formation process and the increase of the multi-information estimate as depicted in Fig. 4. In the beginning the sum of the marginal entropies $H(\mathbf{W}_i^{(t)})$ is as large as the overall entropy of the system because there is no correlation between particles at all. Over time, the marginal entropies decrease, however the overall entropy decreases even faster as the variations of individual particles are correlated. This then leads to an increase of multi-information over time. In Fig. 6 we see snapshots of different samples of one specific experiment run. The final shapes show a certain variety, and there are two visually distinguishable categories of shapes. One has a triangular cluster of dark particles in the center whereas the other one has a cluster of light and outlined particles, where the light particles are sandwiched between the outlined particles. So even though there are several final states, they can be grouped into several shapes with specific distinctive visual features.

If the types are restricted to a single type, which means each particle interacts with each other in the same way, the resulting equilibrium configuration is always a regular grid for force-scaling function F^2 (see rightmost panel in Fig. 3), and the self-organization is very low. This is due to two effects: Firstly, the regular grid is also always in the form of a disc, there is no variety in shapes, so the entropy for each particle is already very low in general (after alignment), and secondly, small perturbations in the grid structure are local and do not spread through the grid.

Interestingly, this is not always the case when we used the force-scaling function $F^{(1)}$ which show a different behavior. For example, if the cut-off radius r_c is larger than $2r_{\alpha\alpha}$ (we consider only one type) and there are 20 particles, then the particles configure into two concentric regular polygons where the rotation of the inner polygon with regard to the outer polygon shows one degree of freedom (see Fig. 7). This already leads to a relatively high amount (compared to all other experiments we analyzed) of self-organization with just one type, as shown in Fig. 5. This already foreshadows the insight from Sec. 6.1 that the amount of self-organization a system can exhibit depends strongly on how individual particles interact with each other. In the next section we will explore this relationship in terms of interaction radius, number of types and self-organization in greater detail.

It can be seen in Fig. 5 that the multi-information is still increasing at time step 250, even though a visual inspection of the simulation samples shows that there is hardly any movement except a slight expansion of the configuration. We constrained our simulations in general to 250 time steps because of the limitation of computational capacity. In most of our experiments the equilibrium was

attained well before the maximum of 250 time steps was reached. When this was not the case the system was either still slowly expanding, but already formed most of its final shape, or it reached a limit cycle with a periodic dynamic. In the case of the periodic limit cycle, the multi-information entered a plateau before the maximum of 250 steps was reached, however, the equilibrium stopping criterion which requires nearly vanishing forces for several time steps was not fulfilled at any time.

6.1 Comparison of interaction types

Simulations with $l = 3$ to 5 types and $n = 20$ to 120 particles almost always show quantifiable self-organization reflected in multi-information (see Fig. 4 for a typical example). Observing an increase in multi-information is independent of the employed force-scaling function, and only vanishes in cases where the interactions between the different types are very similar (however as mentioned earlier, this is not a sufficient condition). Initially we hypothesized that increasing the number of types above a certain ratio of distinct types to particles will generally lead to a decrease of self-organization. It can be seen in Fig. 8 that there is a decrease in self-organization with increasing number of types (for a fixed number of particles), when $F^{(2)}$-scaling is used with randomly generated type matrices. However, the assumption does not hold for $F^{(1)}$ scaling.

In Fig. 9 it can be seen that with increasing cut-off radius r_c, the self-organization increases even in the case where there are the same number of types as there are particles $l = m$. This result seems counter intuitive at first: the resulting particle configurations look arbitrary and unstructured, though the multi-information is increasing over time. But note: we have to consider that spatial regularities are not a necessary condition for self-organization, but that the mutual interactions define possible attractors to which the particles then organize. Because of the large number of different types compared to particles the structure is not (and cannot expected to be) regular. On the other hand, if the interactions are locally limited, either because of a small cut-off radius or because of a decreasing force-scaling function like $F^{(2)}$, the self-organization is limited as well (Fig. 9, $r_c \leq 7.5$). If we compare this to the self-organization exhibited by systems with the same amount of particles, same local limitations on interactions, but considerably fewer different types, we can make the following observation: The increase of multi-information over time in these systems is much higher than in those being local and having as many types as particles (see Fig. 10).

An increase of multi-information can either be reached by the decrease of the overall entropy of the system, the increase of the marginal observer entropies or a mixture of both. In the case of a decrease of overall entropy and constant observer entropies, the system looks the same for each observer and it is quite clear that the correlation between the particles must have increased. In the opposite case, the degree of freedom each observer sees is increased while the overall entropy of the system stays constant, therefore there must exist some correlation between the particles that removes the individual degrees of freedom

13

when looking at the system as a whole.

To reach an increase in correlation (i.e. multi-information) among the particles information needs to spread through the collective [40]. And hence, it is not surprising that long-range interactions lead to a lot of self-organization. What is, however, quite interesting is, that this is also possible if the interactions are local but limited in variety. In these cases, where interactions are local, there are almost always smaller clusters interacting with each other. Each cluster shows a very regular structure and consists of particles of one type. We will further discuss this observation in Sec. 7.2.

6.1.1 Localization of organization

If we have a cluster structure with spatially confined subsystems, there is a natural question: Is it possible to locate where the largest contribution to the organization is made? In Sec. 3.1, we showed that it is possible to decompose the multi-information of the observer variables into the several multi-information terms, that each measure the multi-information of a subset of the observer variables, and one term that measures the multi-information between these coarse-grained joint observer variables. We now consider the joint random variable of all observers of each type of particles as coarse-grained observers $\tilde{\mathbf{W}}_1, ..., \tilde{\mathbf{W}}_l$ (see Sec. 3.1), and calculate the multi-information individually. The results using this decomposition do not draw a clear picture: A general observation is that in every experiment we were able see organization on all levels. If we normalize the decomposition with respect to the total multi-information for each time step, it can be seen that in the beginning of the experiment that the relative contribution of each decomposition term still varies, while after a few steps the relative contribution values will settle, even though the whole multi-information is still increasing and the system is still organizing (see Fig. 11). However, there is no common pattern in the decomposition that we were able to relate to the dynamics of the system. Almost all experiments show this pattern of an early phase where the decomposition varies by large amounts, and then a phase where the decomposition settles down. But, at this stage we were not able to link the dominance of a specific type or the inter-type multi-information, to specific observable behaviors.

7 Discussion

We used multi-information as a measure for self-organization and applied it to experiments of interacting particle systems. Estimations of multi-information were obtained using the Kraskov-Stögbauer-Grassberger estimator [22]. The approach seems to be in general transferable to other discrete-time dynamical systems that share the same invariants (or the same method can be adapted to other invariant transformations).

As mentioned earlier, defining a measure for self-organization is not a straightforward task. With our definition one has to be careful in the choice of the ob-

server variables. However, the results show that particle/type-based observers are a practicable approach to measure self-organization in spatial systems. This measure also has the advantage that we do not need to assume stationarity of the system we investigate.

Regarding the specific model we chose, a core observation is that the choice of the force-scaling function, and therefore the form of interaction between particles, has a strong influence on the resulting dynamics. We have studied particular aspects of collective organization; the lessons drawn from may have implications for the understanding of biological collective self-organization in space.

7.1 Uniform collectives

Our first observation was, that a uniform collective (only one type) when forming regular grids only shows a small amount of measurable self-organization. Recalling that the particles are initially uniformly distributed in a disc of certain radius, and then form an almost unique equilibrium configuration, we see that the measured self-organization is in agreement with the statistical complexity notion of self-organization. The time-dependent statistical complexity [8] of the initial state vanishes as it is completely random, but also once the system reaches the equilibrium state there was no spatial or temporal variance, and so the statistical complexity vanishes as well. On the other hand, the process from randomly distributed states to a regular grid structure is similar to crystals or paracrystallines which often are considered as self-organized [14]. So, it could be that there is a small increase in multi-information in the beginning that then should vanish again, when the system settled on the equilibrium state. However, we were not able to show an early increase of multi-information with the estimator. It is possible that the amount of early self-organization is smaller than the bias of the estimator.

These discussions lead to an interesting related question. Suppose we observe two processes X_t, Y_t, both starting with a uniform distribution over all states, going into a phase of self-organization where the multi-information of some observer variables increases. But then, the first process X_t goes back to a uniform distribution, whereas Y_t cools down to a deterministic periodic process. In both cases there is no difference between the multi-information at the initial-state as well as the multi-information after the organization phase. Though there are qualitative differences in the process and we would consider X_t as temporarily organized while Y_t should be considered as self-organized even after the cooling down phase. Therefore it is helpful to look at the evolution of entropy over time as well, which helps here to distinguish both cases, even though it does not serve as a measure for self-organization in itself.

7.2 Long range interactions

Our main observation of the self-organization of particle systems concerned the variation of the cut-off radius r_c and the number of types in the particles.

Here we were able to see that given unconstrained interactions ($r_c = \infty$) the self-organization can be very high even if the particles all have distinct types with different mutual interactions. This was surprising insofar that the particle configuration in these settings do not show much spatial structure, and there is generally no emergent description in terms of clusters interacting with each other. However, the configurations show a lot of statistical structure, i.e. correlations, that the multi-information is able to detect. This can be related to the retrieval of spatial configuration of sensors using information-distance [27]. The distances are in this case represented by the $r_{\alpha\beta}$ radii and the experiment with $r_c = \infty$ is equivalent to the relaxation procedure that was used in [27] for the reconstruction of spatial structure. Another interesting point here is, that self-organization can occur without exhibiting a visually emergent spatial structure, this could support the idea put forward in [37] that self-organization and emergence are separate concepts.

Now, decreasing the cut-off radius r_c also decreases the observable self-organization if the number of types is held constant. This supports another assumption about self-organization: Information spread through the system is a crucial property of self-organizing systems. By limiting the cut-off radius, we are constraining the particles ability to transfer information through the system and therefore its ability to organize.

Surprisingly however, if the number of types is decreased (with fixed small value of r_c), the self-organization increases and we can observe emergent structures like balls enclosed in circles, layers of different types (see Fig. 12). It seems that the common co-occurence of self-organization and emergence of clustered structures is a result of the way a system can achieve higher overall complexity when interactions are locally constrained. Even with limited r_c, the homogeneity of the space as well as the homogeneity of local structures allow long-range structural interactions between groups of particles, which in turn allows to produce to a higher amount of self-organization of the whole system.

7.3   Future work

The methods developed in [24] promise to furnish tools to investigate the information dynamics between individual particles over time. We tried to measure the information transfer between particles, but so far the results are still inconclusive and this is part of ongoing research.

Our investigations showed the multi-information is decomposed without clear preference in the initial phase, but settle to a fixed amount of contribution to the total amount of multi-information. So it is already possible to distinguish qualitatively different phases of the self-organization process. Having a measure like the interaction measure devised by [19], it might be possible to use a continuous version of this measure to investigate the interaction dynamics and autonomy of sub-processes in the particle collectives, relate it to the amount of self-organization that is exhibited, and such get even more insight to process of self-organization.

So far we only considered systems where particles have fixed dynamics. The

system simply self-organizes according to fixed "physics". Already with this, we can reach quite a variation of patterns and configurations by merely changing the parameters of the dynamics. However these changes, if seen in the context of biological systems, will be limited by evolutionary speed and quite inflexible. If we want to get systems exhibiting higher complexity there needs to be some "informed" local guidance, either through external top-down intervention in the organization process, or by considering reactive agents instead of particles that just follow rules. As part of the rich toolbox of information-theoretic concepts, the present methods plug in immediately into the existing arsenal of information-theoretic tools to model agents in their environment and this scenario opens many doors for further investigations of guided self organization.

Acknowledgments

References

[1] Ay, N., Olbrich, E., and Bertschinger, N., A Unifying Framework for Complexity Measures of Finite Systems, *Complexity* (2006) 1–15.

[2] Ay, N. and Polani, D., Information Flows in Causal Networks, *Advances in Complex Systems* **11** (2008) 17–41.

[3] Bertschinger, N., Olbrich, E., Ay, N., and Jost, J., Autonomy: An information theoretic perspective., *Biosystems* **91** (2008) 331–345.

[4] Bonabeau, E., From classical models of morphogenesis to agent-based models of pattern formation, *Artificial Life* **3** (1997) 191–211.

[5] Ceguerra, R., Lizier, J., and Zomaya, A., Information storage and transfer in the synchronization process in locally-connected networks, in *Artificial Life (ALIFE), 2011 IEEE Symposium on* (IEEE, 2011), pp. 54–61.

[6] Cover, T. M. and Thomas, J. A., *Elements of Information Theory 2nd Edition*, Wiley Series in Telecommunications and Signal Processing (Wiley-Interscience, 2006).

[7] Crutchfield, J., Information and its metric, *Nonlinear Structures in Physical Systems–Pattern Formation, Chaos and Waves* (1990) 119–130.

[8] Crutchfield, J., Semantics and thermodynamics, in *Santa Fe Institue Studies in the Sciences of Complexity*, Vol. 12 (Addison-Wesley Publishing Co, 1992), pp. 317–317.

[9] Davies, R., Twining, C., and Taylor, C., *Statistical models of shape: optimisation and evaluation* (Springer-Verlag New York Inc, 2008).

17

[10] Doursat, R., Organically grown architectures: Creating decentralized, autonomous systems by embryomorphic engineering, *Organic Computing* (2008) 167–199.

[11] Doursat, R., Programmable Architectures That Are Complex and Self-Organized : From Morphogenesis to Engineering, *Artificial Life* (2008) 181–188.

[12] Dryden, I. and Mardia, K., *Statistical shape analysis*, Vol. 4 (John Wiley & Sons New York, 1998).

[13] Gibbs, J., *On the equilibrium of heterogeneous substances* (Transactions of the Connecticut Academy, 1874).

[14] Grom, G., Lockwood, D., McCaffrey, J., Labbe, H., Fauchet, P., White Jr, B., Diener, J., Kovalev, D., Koch, F., Tsybeskov, L., *et al.*, Ordering and self-organization in nanocrystalline silicon, *Nature* **407** (2000) 358–361.

[15] Hausser, J. and Strimmer, K., Entropy inference and the James-Stein estimator , with application to nonlinear gene association networks, *Most* (2009) 1–18.

[16] Haykin, S., *Communication systems* (Wiley-India, 2008).

[17] James, R., Ellison, C., and Crutchfield, J., Anatomy of a bit: Information in a time series observation, *Arxiv preprint arXiv:1105.2988* (2011).

[18] Jaynes, E., *The Gibbs paradox* (Kluwer Academic: Dordrecht, 1992).

[19] Kahle, T., Olbrich, E., Jost, J., and Ay, N., Complexity measures from interaction structures, *Physical Review E* **79** (2009) 11.

[20] Kloeden, P., Platen, E., and Schurz, H., Stochastic differential equations, *Numerical Solution of SDE Through Computer Experiments* (1994) 63–90.

[21] Kondor, I., *Group theoretical methods in machine learning* (Columbia University, 2008).

[22] Kraskov, A., Stögbauer, H., and Grassberger, P., Estimating mutual information, *Phys. Rev. E* **69** (2004) 066138.

[23] Liggett, T., *Interacting particle systems*, Vol. 276 (Springer Verlag, 1985).

[24] Lizier, J., Prokopenko, M., and Zomaya, A., Local information transfer as a spatiotemporal filter for complex systems, *Physical Review E* **77** (2008) 026110.

[25] Meinhardt, H., *Models of biological pattern formation* (Academic Press London, 1982).

[26] Odell, G., Oster, G., Burnside, B., and Alberch, P., A mechanical model for epithelial morphogenesis, *Journal of Mathematical Biology* **9** (1980) 291–295.

[27] Olsson, L., Nehaniv, C., and Polani, D., Sensory channel grouping and structure from uninterpreted sensor data, in *Evolvable Hardware, 2004. Proceedings. 2004 NASA/DoD Conference on* (IEEE, 2004), pp. 153–160.

[28] Pearl, J., *Causality: models, reasoning and inference* (Cambridge Univ Press, 2000).

[29] Polani, D., Measures for the organization of self-organizing maps, *Studies in Fuzziness and Soft Computing* **78** (2002) 13–44.

[30] Polani, D., Foundations and formalizations of self-organization, *Advances in applied self-organizing systems* (2008) 19–37.

[31] Polani, D., An informational perspective on how the embodiment can relieve cognitive burden, in *Artificial Life (ALIFE), 2011 IEEE Symposium on* (IEEE, 2011), pp. 78–85.

[32] Press, W., Flannery, B., Teukolsky, S., Vetterling, W., *et al.*, *Numerical recipes*, Vol. 547 (Cambridge Univ Press, 1986).

[33] Prusinkiewicz, P., Visual models of morphogenesis, *Artificial life* **1** (1993) 61–74.

[34] Rusu, R. and Cousins, S., 3d is here: Point cloud library (pcl), in *Robotics and Automation (ICRA), 2011 IEEE International Conference on* (IEEE, 2011), pp. 1–4.

[35] Sayama, H., Swarm chemistry., *Artificial life* **15** (2009) 105–14.

[36] Schinazi, R., *Classical and spatial stochastic processes* (Springer, 1999).

[37] Shalizi, C. R., *Causal Architecture, Complexity and Self-Organization in Time Series and Cellular Automata*, Ph.D. thesis (2001).

[38] Shannon, C. E., A Mathematical Theory of Communication, *Bell System Technical Journal* **27** (1948) 379–423.

[39] Small, C., *The statistical theory of shape* (Springer Verlag, 1996).

[40] Steudel, B. and Ay, N., Information-theoretic inference of common ancestors, *Arxiv preprint arXiv:1010.5720* (2010).

[41] Suzuki, T., Sugiyama, M., Sese, J., and Kanamori, T., Approximating mutual information by maximum likelihood density ratio estimation, in *JMLR workshop and conference proceedings*, Vol. 4 (Citeseer, 2008), pp. 5–20.

[42] Tononi, G., Sporns, O., and Edelman, G., A measure for brain complexity: relating functional segregation and integration in the nervous system, *Proceedings of the National Academy of Sciences* **91** (1994) 5033.

[43] Touchette, H. and Lloyd, S., Information-theoretic approach to the study of control systems, *Physica A: Statistical Mechanics and its Applications* **331** (2004) 140–172.

[44] Wang, X. R., Miller, J. M., Lizier, J. T., Prokopenko, M., and Rossi, L. F., Measuring Information Storage and Transfer in Swarms, *Physica* (2003).

[45] Witten Jr, T. and Sander, L., Diffusion-limited aggregation, a kinetic critical phenomenon, *Physical Review Letters* **47** (1981) 1400–1403.

[46] Wolfram, S., Theory and applications of cellular automata (1986).

[47] Wolpert, L., Beddington, R., Jessell, T., Lawrence, P., Meyerowitz, E., and Smith, J., *Principles of development*, Vol. 3 (Oxford University Press New York.:, 2002).

[48] Zhang, Z., Iterative point matching for registration of free-form curves and surfaces, *International journal of computer vision* **13** (1992) 119–152.
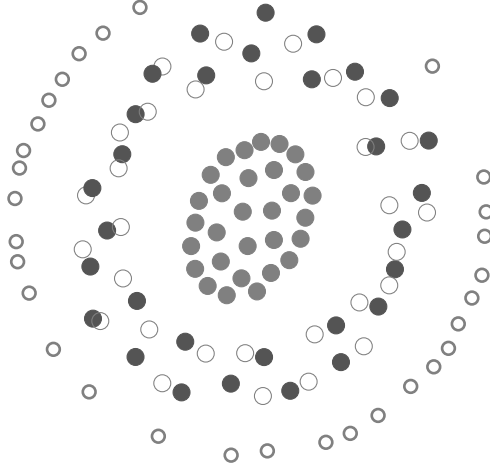
Figure 1: Example of a particle configuration.
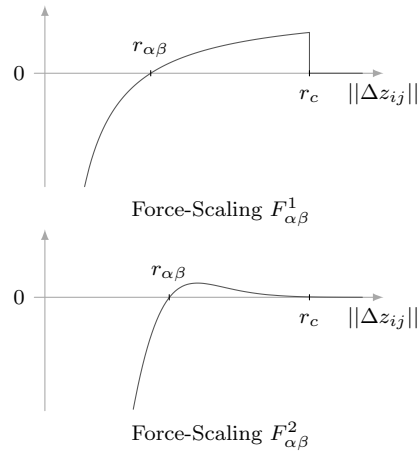
Figure 2: Comparison of both force-scaling functions, $r_{\alpha\beta}$ denotes the prefered distance between particles of type α and β. For $F^{(1)}$ this radius can be directly specified. The long range attraction of $F^{(1)}$ is only cut off by the radius r_c.

21
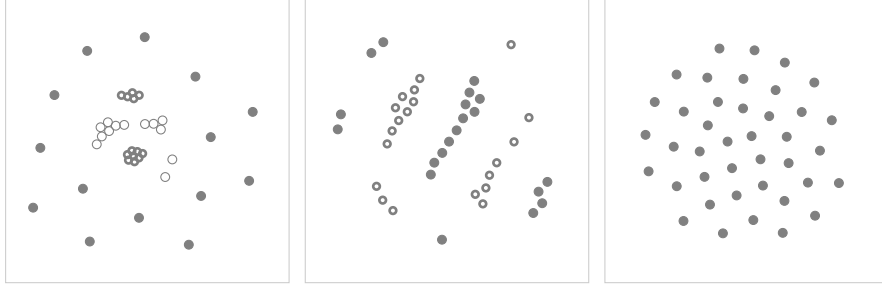
Figure 3: Examples of equilibrium states of particle collectives with different number of types.
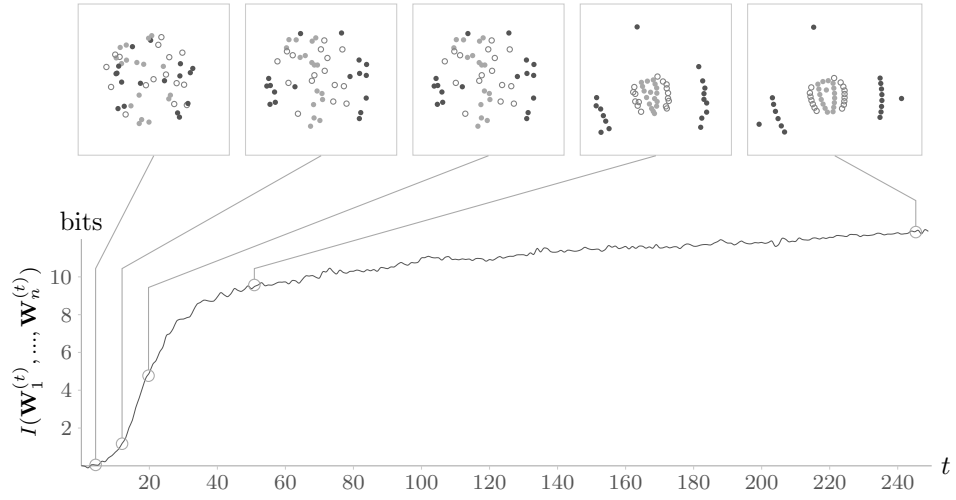

Figure 4: Multi-information between particles plotted against time with $n = 50$, $l = 3$, $r_c = 5.0$ and $r_{\alpha\beta} = \{\{2.5, 5.0, 4.0\}, \{5.0, 2.5, 2.0\}, \{4.0, 2.0, 3.5\}\}$. The increase of multi-information correlates with the visual organization shown by snapshots of one particular sample at different times $t = \{0, 10, 20, 50, 249\}$.
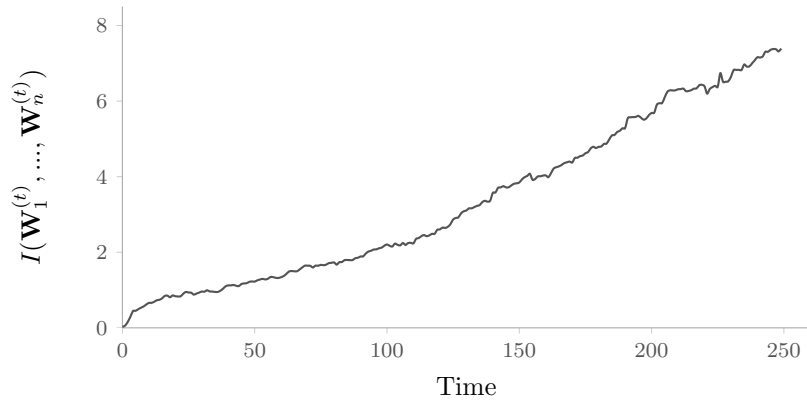
Figure 5: Multi-information plotted against time using force-scaling function $F^{(1)}$, 20 particles of one type, 500 samples.
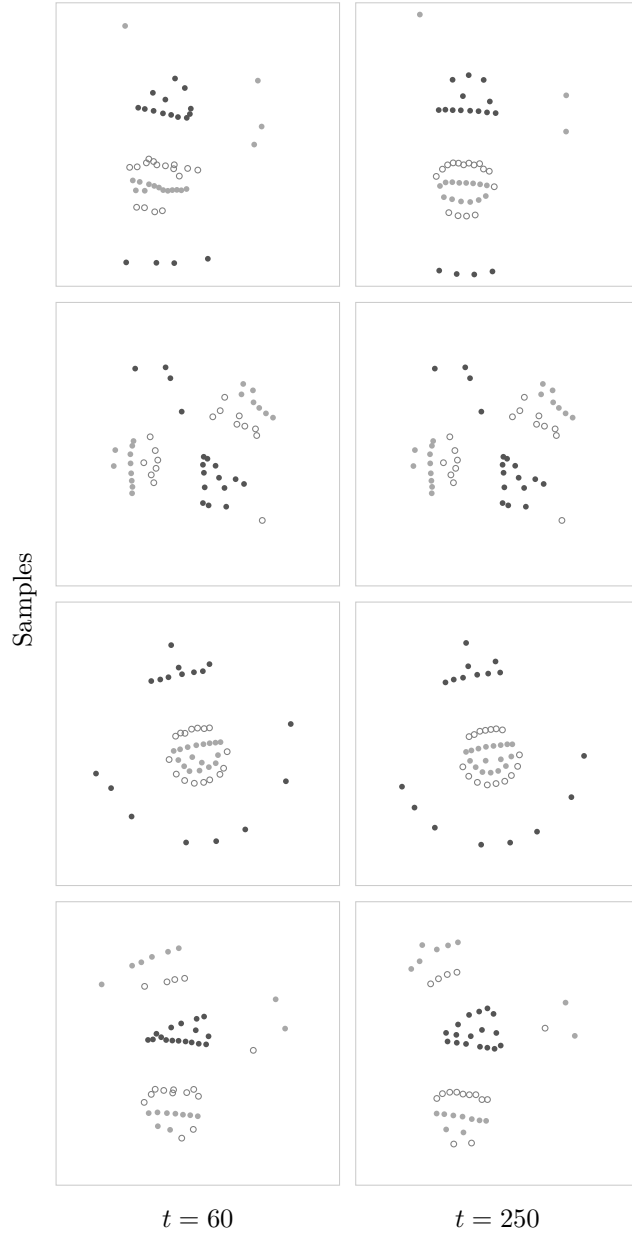
Figure 6: Snapshots of different samples of the experiment shown in Fig. 4 at $t = 60$ (left) and $t = 250$ (right).
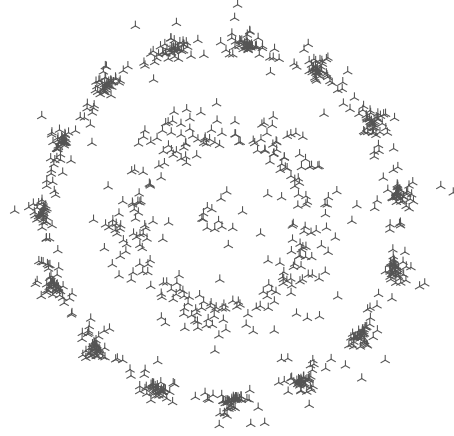
Figure 7: Plot of all particles of all samples at time $t = 250$ using force-scaling function $F^{(1)}$, 20 particles of one type per sample. It can be seen that the outer ring has been much better aligned so that for each particle samples match more closely (denser clusters), while this is not possible for the inner ring of particles as their alignment related to the outer ring is a degree of freedom.
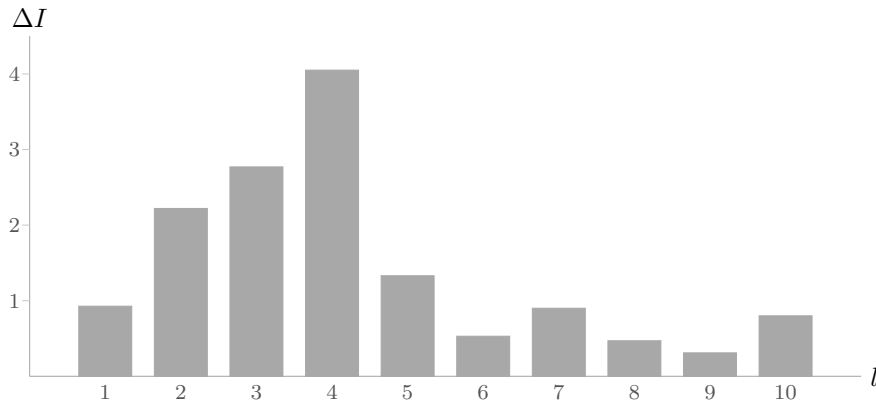
Figure 8: Increase of multi-information between $t = 0$ and $t = 250$ using force-scaling function $F^{(2)}$, for different numbers of types. Average over 10 randomly generated types with mutual preferred distance radii $r_{\alpha\beta}$ between 1.0 and 5.0.

25

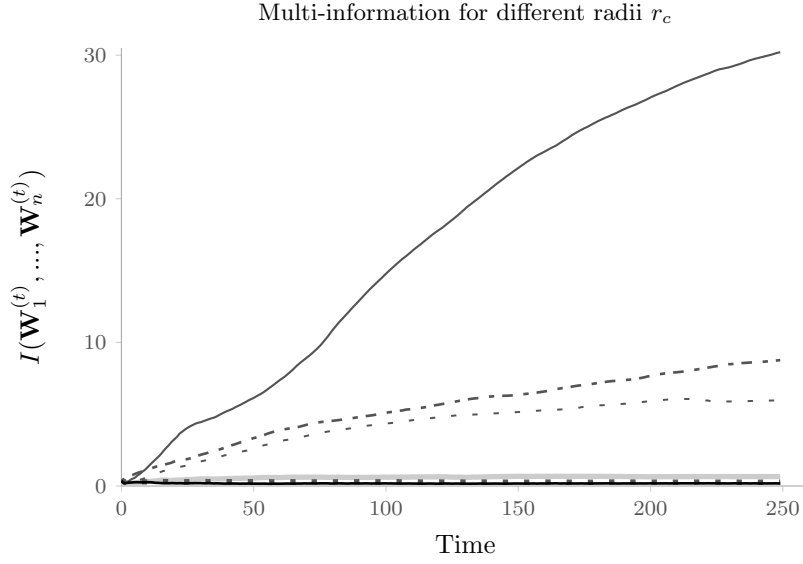Multi-information for different radii r_c

Figure 9: Multi-information plotted against time using force-scaling function $F^{(1)}$, 20 particles, 20 types, averaged over 10 samples of random types where $r_{\alpha\beta} \in [2.0, 8.0]$, $k_{\alpha\beta} = 1$ for different cut-off radii r_c: ▪ ▪ ▪ 2.5, —— 5.0, —— 7.5, - - · 10.0, - ▪ - 15.0, —— ∞.

Multi-information for different number of types and r_c
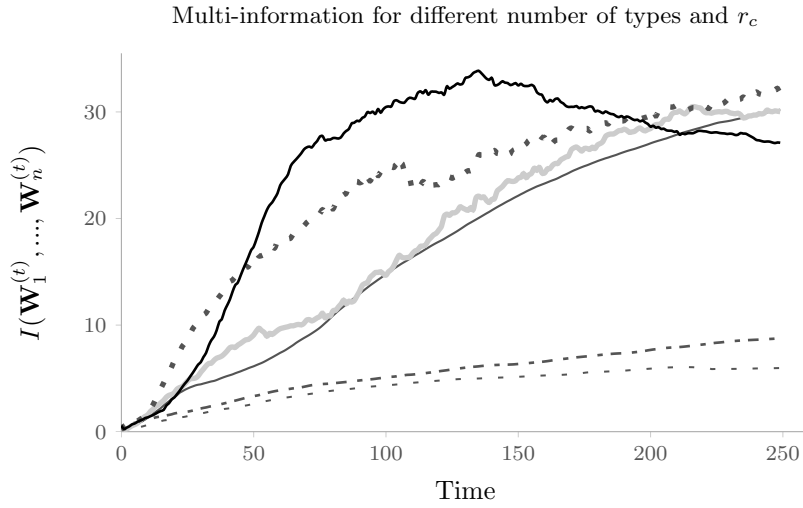
Figure 10: Multi-information of 20 particles plotted against time using force-scaling function $F^{(1)}$, averaged over 10 samples of random types where $r_{\alpha\beta} \in [2.0, 8.0]$, $k_{\alpha\beta} = 1$ for different cut-off radii r_c and different number of types: - - · $l = 20$, $r_c = 10$, - ▪ - $l = 20$, $r_c = 15$, —— $l = 20$, $r_c = \infty$, —— $l = 5$, $r_c = 10$, ▪ ▪ ▪ $l = 5$, $r_c = 15$, —— $l = 5$, $r_c = \infty$.
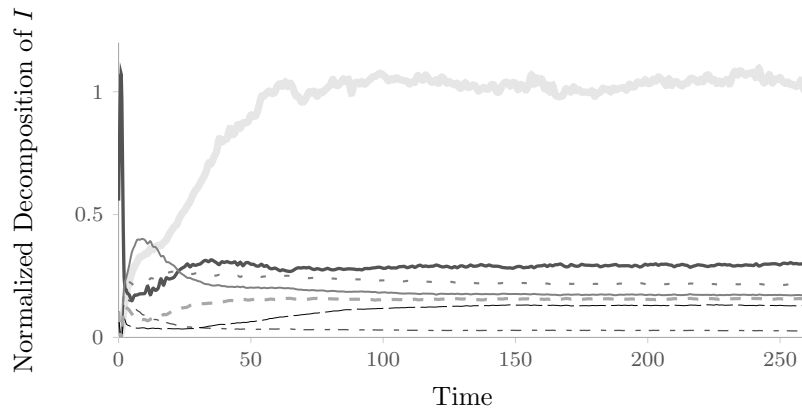
Figure 11: Contribution of the different terms of the decomposition normalized with the multi-information in each time-step (▬ is a normalized plot of the multi-information between all particles). ▬ $I(\tilde{\mathbf{W}}_1, ..., \tilde{\mathbf{W}}_l)$, multi-information between all particles of ▬ type 1, - - - type 2, - · - type 3, ── type 4, ▬ ▬ · type 5. This is a decomposition of one of the sampled types with $l = 5, r_c = 15$ from Fig. 10.
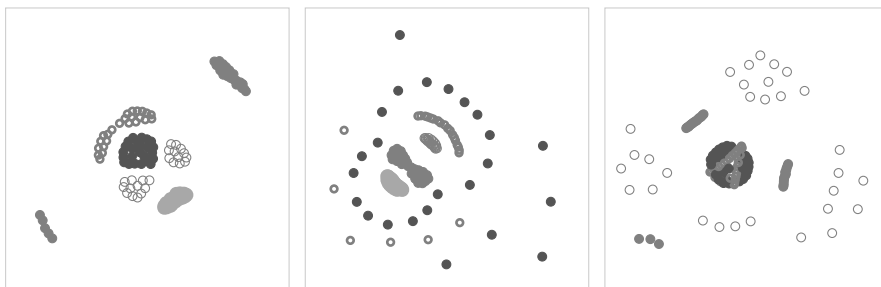

Figure 12: Examples of emergent structures in particle collectives.