Chalmers Publication Library

An Information-Theoretic Perspective on Coarse-Graining, Including the Transition from Micro to Macro

(article starts on next page)

Article

An Information-Theoretic Perspective on Coarse-Graining, Including the Transition from Micro to Macro

Kristian Lindgren

Complex systems group, Department of Energy and Environment, Chalmers University of Technology, SE-41296 Göteborg, Sweden; E-Mail: kristian.lindgren@chalmers.se; Tel.: +46-317723131

Academic Editor: Rick Quax

Abstract: An information-theoretic perspective on coarse-graining is presented. It starts with an information characterization of configurations at the micro-level using a local information quantity that has a spatial average equal to a microscopic entropy. With a reversible micro dynamics, this entropy is conserved. In the micro-macro transition, it is shown how this local information quantity is transformed into a macroscopic entropy, as the local states are aggregated into macroscopic concentration variables. The information loss in this transition is identified, and the connection to the irreversibility of the macro dynamics and the second law of thermodynamics is discussed. This is then connected to a process of further coarse-graining towards higher characteristic length scales in the context of chemical reaction-diffusion dynamics capable of pattern formation. On these higher levels of coarse-graining, information flows across length scales and across space are defined. These flows obey a continuity equation for information, and they are connected to the thermodynamic constraints of the system, via an outflow of information from macroscopic to microscopic levels in the form of entropy production, as well as an inflow of information, from an external free energy source, if a spatial chemical pattern is to be maintained.

Keywords: information theory; entropy; microscopic; macroscopic; second law; reaction-diffusion; pattern formation

---

1. Introduction

Entropy characterizes the disorder of a system. It takes very different forms depending on the level to which it is applied. In, for example, a chemical self-organizing system, the chemical pattern is

described by the macroscopic concentrations of the different molecules, and the entropy is a function of the concentration profile of the system. Such a description serves as a basis for a thermodynamic analysis, in which not only entropy, but also free energy has an important explanatory role. Free energy and entropy serve as opposite quantities, like the notions of order and disorder. The second law of thermodynamics tells us that for a closed system, the entropy increases until equilibrium is reached. The free energy – essentially the internal energy minus the entropy – evolves in the opposite direction.

In statistical mechanics, one defines macrostates, for example, the grand canonical ensemble, as the distribution over microstates that is consistent with certain expected values on internal energy and the number of molecules of different types. In that way, it is tightly connected to the thermodynamic description. Statistical mechanics entropy, though, is defined in a different way, as the entropy of the distribution over the microstates. However, the resulting entropy value is the same.

If entropy characterizes disorder, a relevant question to ask is whether a single microstate, *i.e.*, one microscopic realization from the distribution of the macrostate, can be said to have a certain disorder or "microscopic entropy". This would mean that by characterizing the configuration with all of its correlational structure, we should be able to determine an internal entropy of the single microstate. Under general assumptions in statistical mechanics, this is true for almost all microstates in the ensemble (in the thermodynamic limit), and it turns out that the information-theoretic way to quantify the randomness of a single configuration results in the macroscopic entropy [1].

In this paper, we present a perspective that shows how the microscopic entropy analysis connects the macroscopic entropy characteristics. A key question here is how the second law can be understood as we make the transition from the microscopic to the macroscopic level. At the microscopic level, the processes that govern the dynamics are reversible, and that implies a conservation of the disorder (defined in the information-theoretic way). This means that something is lost in the transition to the macro-level, since the second law of thermodynamics implies an increasing entropy.

In general, there are several ways in which one can look for structure in a system, and depending on what questions one is asking, there are different appropriate system descriptions one should choose. For an information characterization, this means that we need to decide on the probabilistic description of the system that we want to study.

The first choice deals with whether we are looking at spatial configurations, temporal patterns or a combination of those; or, to be more precise, whether the probabilistic description should be based on a distribution that stretches over space and/or time. From a dynamical systems perspective, a time series basis for probabilities is required, while for a statistical mechanics or thermodynamics description, it is sufficient with a series of snapshots and where probabilities are based only on the spatial extension of the system.

A different approach, also aiming for establishing connections between an information perspective and thermodynamic characteristics, is the one based on transfer entropy [2]; see, e.g., the work by Prokopenko and Lizier [3,4]. In this approach, the focus is on temporal aspects, and the change over time of the spatial configurations is not explicitly taken into account.

In the discussion here, the focus will be on an information characterization of the spatial configurations. We are, though, interested in how information quantities change over time as an effect of the laws of motion of the system. One aim is to clarify how the appropriate choice of description

changes when we move from microscopic to macroscopic levels and what this change in description means in terms of information. One point we want to make is how the second law of thermodynamics can be understood as an information loss related to the difference in descriptions of the microscopic and the macroscopic levels.

The perspective is also complemented with a review of previous work [5,6], where we continue the process of coarse-graining over increasing length scales at the macro-level. This serves as a basis for identifying the location of ordered spatial structure, not only at different positions in the system, but also on different length scales. This view connects to the thermodynamic constraints that, e.g., a chemical self-organizing system must obey, and it gives a consistent picture of how the formation of spatial pattern and entropy production is balanced by an inflow of free energy.

The paper is organized as follows. In Section 2, we present the information perspective that characterizes the microstate by identifying a local information quantity that has a spatial average that equals the entropy (per site) of the microstate configuration. This approach is applied to cellular automata, both irreversible and reversible, where it is illustrated how different spatial characteristics are captured. At the microscopic level, even though physically-realistic dynamics is reversible and has a conserved entropy, the increasing correlation lengths can be seen as a fundamental mechanisms for creating local disorder and, hence, an apparent increase in entropy. In Section 3, we present how the local information quantity defined at the micro-level is transformed into a macroscopically-defined information quantity. In the case of a chemical system, we get the macroscopic entropy based on concentration variables as the basis for the system description. In this transition, we discuss the character of the information that is lost when we move from the detailed microstate description to an aggregate concentration profile. The loss of information allows for the aggregate dynamics to be irreversible and the second law of thermodynamics to hold on this level. In Section 4, we present the review of the work on coarse-graining in chemical pattern-forming systems, and we show how such a coarse-graining perspective can be connected to the information loss in the micro-macro transition, as well as to the inflow of free energy that creates and maintains spatial chemical patterns. Finally, concluding remarks are given in Section 5.

2. Information Theory of the Microstate

We assume that the system is one-dimensional and discrete, so that the full state of the system can be described as a sequence of local states (this constraint can be relaxed, and the formalism can be extended to lattices in higher dimensions). Formally, we consider systems in the infinite length limit, where the microstates can be viewed as being generated by a stationary stochastic process.

Therefore, we can characterize our system by probability distributions P_n over symbol sequences of finite length n, ($n = 1, 2, ...$),

$$P_n = \{p(x_1, ..., x_n)\}_{(x_1, ..., x_n) \in \Lambda^n} \ , \tag{1}$$

where Λ denotes the set of possible local states. One way to make an information characterization of the microstate is to "read" the state from left to right, observing the state symbols one by one.

We assume that the microstate can be viewed as the outcome of an ergodic stochastic process. This means that, almost always, we can get the right statistics from a single (infinitely long) sequence of

symbols representing one microstate in the ensemble. The ergodic theorem (see, e.g., [7]) states that the calculation of an average $\langle f \rangle$ of a function f based on the probabilities $p(x_1...x_n)$ results in being the same as an average formed by following the individual symbol sequence $(s_1, s_2, s_3, ...)$, where s_k is the symbol at position k in the sequence,

$$\sum_{x_1...x_n \in \Lambda^n} p(x_1, ..., x_n) f(x_1, ..., x_n) = \lim_{T \to \infty} \frac{1}{T} \sum_{k=1}^{T} f(s_k, s_{k+1}, ..., s_{k+n-1}) . \tag{2}$$

Note that we are considering a spatial configuration characterized by the distributions P_n, which means that the averaging is not temporal, but spatial.

The entropy rate s of the process characterized by P_n is determined by the properties of the block entropy S_n,

$$S_n = \sum_{x_1...x_n \in \Lambda^n} p(x_1, ..., x_n) \log \frac{1}{p(x_1, ..., x_n)} , \tag{3}$$

and it is given by:

$$s = \lim_{n \to \infty} \frac{1}{n} S_n = \lim_{n \to \infty} S_n - S_{n-1} = \lim_{n \to \infty} \Delta S_n . \tag{4}$$

The entropy rate (or the entropy per symbol) s characterizes the remaining randomness of the system when all correlations have been taken into account [8]. Again, in our discussion, this characterizes the randomness of the spatial configuration.

2.1. Local Information

Now, we consider one realization or outcome (with infinite length) of the process characterized by P_m. This can be seen as a microstate or a specific configuration, which we will now analyze information-theoretically by reading it from, say, left to right. In order to characterize how much information we gain when we read a symbol, x_m, given that we know the $m-1$ symbols to the left $(x_1...x_{m-1})$, which we have already observed, we need the conditional probability:

$$p(x_m|x_1...x_{m-1}) = \frac{p(x_1...x_{m-1}x_m)}{p(x_1...x_{m-1})} . \tag{5}$$

This can then be used to define "left-sided" conditional local information:

$$I_{i,m}^{(\mathrm{L})} = \log \frac{1}{p(s_i|s_{i-m+1}...s_{i-1})} , \tag{6}$$

where s_i denotes the symbol at position i; see [9]. A corresponding information quantity $I_{i,m}^{(\mathrm{R})}$ conditioned on the $m-1$ cells to the right of the position i is similarly defined. A symmetric local information quantity can then be defined by:

$$I_{i,m} = \frac{1}{2} \big(I_{i,m}^{(\mathrm{L})} + I_{i,m}^{(\mathrm{R})} \big) . \tag{7}$$

Some illustrations of this local information quantity are shown in the next section.

As the quantity $I_{i,m}^{(\mathrm{L})}$ is defined, it reflects the information gained when one learns the local state at the position in question. On average, though, this will reflect the uncertainty about the next position given

the conditional block. Therefore, the local information gives the entropy of the system. As we sweep the system, we have:

$$\langle I_{i,m}^{(\mathrm{L})}\rangle = \lim_{N\to\infty,m\to\infty} \frac{1}{2N+1} \sum_{i=-N}^{N} \log \frac{1}{p(s_i|s_{i-m+1}...s_{i-1})} \ . \tag{8}$$

By using the ergodicity theorem, Equation (2), we find that:

$$\langle I_{i,m}^{(\mathrm{L})}\rangle = \lim_{m\to\infty} \sum_{x_1...x_m\in\Lambda^n} p(x_1...x_m) \log \frac{1}{p(x_m|x_1...x_{m-1})}$$

$$= \lim_{m\to\infty} \Delta S_m = s \ , \tag{9}$$

i.e., this quantity almost always results in the entropy rate of the process that defines the microstate properties or the entropy per site of the microstate.

2.2. Microscopic Dynamics

When the microstate changes according to some dynamics, the probabilistic description of the sequence is changed and so is its information properties. Mathematically, this is most conveniently dealt with by determining how the process given by P_m changes to a possibly different process P'_m under the dynamics.

We use cellular automata as a class of dynamical systems to illustrate changes in microstate properties under microscopic dynamics. In general, for deterministic dynamics, we have that the entropy per lattice site decreases in time,

$$\Delta_t s(t) \leq 0 \ , \tag{10}$$

see, e.g., [10]. Intuitively, this can be understood from the fact that, as s quantifies the disorder in the system, a deterministic transformation that is applied locally cannot introduce any new randomness, and hence, the entropy s cannot increase. Note, also, that most cellular automata rules are irreversible, and in that case, the entropy typically decreases as a result of spatial correlations being formed even from a pseudo-random initial state. Such a creation of spatial correlations can also be information-theoretically characterized, and one can identify information in correlations of different distances; see, e.g., [10,11] for illustrations of this in both one- and two-dimensional systems.

For reversible and deterministic dynamics, the entropy is conserved, since the total information is unchanged,

$$\Delta_t s(t) = 0 \ . \tag{11}$$

One simple example of this is found in 1D elementary cellular automata, where surjective rules (like the additive rules) are sufficiently reversible to conserve the microstate entropy. In this situation, one can show that there is a continuity equation describing how the local information is transported in the system [9].

By extracting the local information from the microstates in the time evolution of some cellular automata (CA) rules, we clearly see that the local information quantity works as a regularity filter as it dampens the information signal of all common patterns and emphasizes the rare occurrences; see

Figure 1 in which the irreversible rule, R18, is used. Particle like structures (pairs of 1's) move in a background characterized by a pattern where sequences of 0's of odd length are separated by single 1's. When colliding, the particles are annihilated, which reflects the irreversibility of rule R18. Regularity filters that highlight uncommon patterns in the dynamics of cellular automata are of interest also for characterizing information processing aspects of these systems; see, e.g., the work by Lizier *et al.* [12].
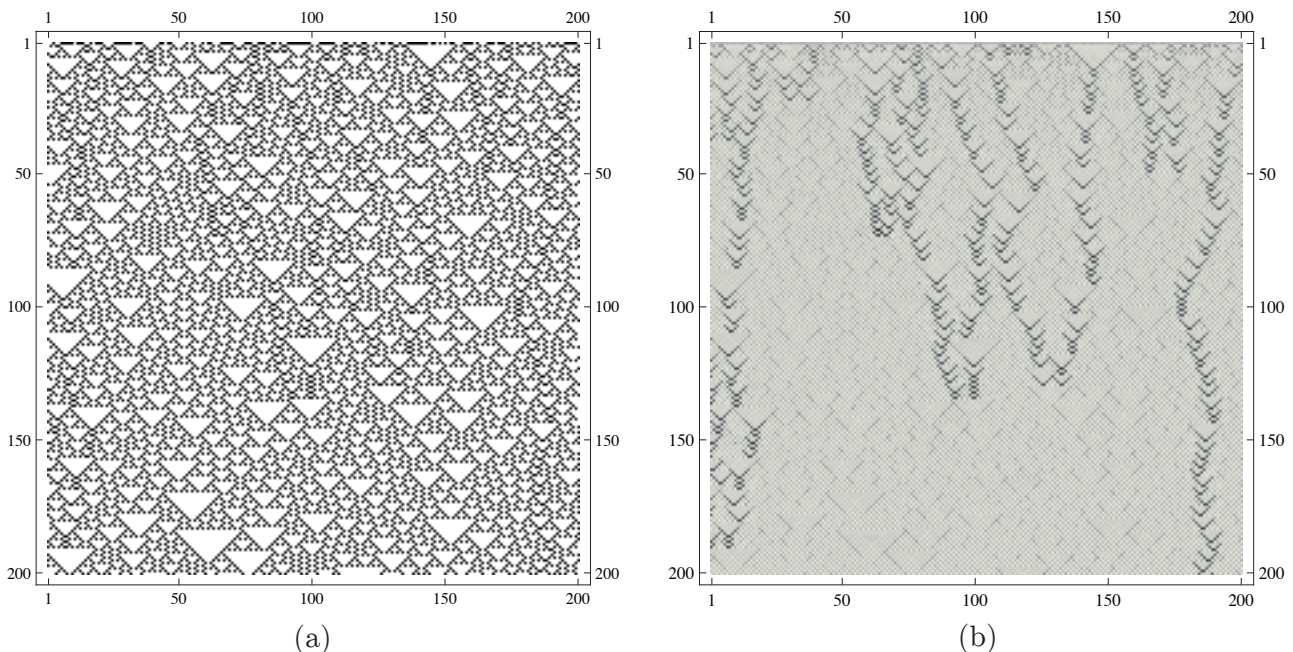

Figure 1. (**a**) The space-time pattern for cellular automaton R18. (**b**) A grey-scale representation of the corresponding local information for length eight of the conditional block (with higher information represented by darker grey). Particle-like information quanta move in a background with a simple probabilistic structure showing a lower information density. The annihilation of the information quanta is an indication of the irreversibility of rule R18.

In the case of the additive rule, R60, producing the space-time pattern shown in Figure 2a, the corresponding illustration indicates how local information is transferred in the system; see Figure 2b–d. Cellular automaton rule R60 adds the states of the present and left neighbor cells modulo two in parallel over the whole lattice or, equivalently, applies the XOR operation to pairs of cells. In this simulation, we have started with a low entropy initial microstate, characterized by a few ones in a background of zeros, given by the Bernoulli distribution.

In the special case of rule R60, one can analytically solve the local information (in the limit of an infinite conditional block of symbols) as a function of time [9], showing that the high local information quantities associated with ones in the original microstate is split into left-hand and right-hand local information being separated as time goes on; see Figure 3. When the conditional block has a limited size, as is illustrated in Figure 2, the full local information becomes increasingly more difficult to detect as time goes on. This phenomenon depends on the fact that correlation information is extending over larger and larger distances. Note, though, that there are still certain points in time when there is more

local information that can be detected already for limited sizes of the conditional block. This change in correlation information can be characterized by a complexity quantity, as is shown in the next section.
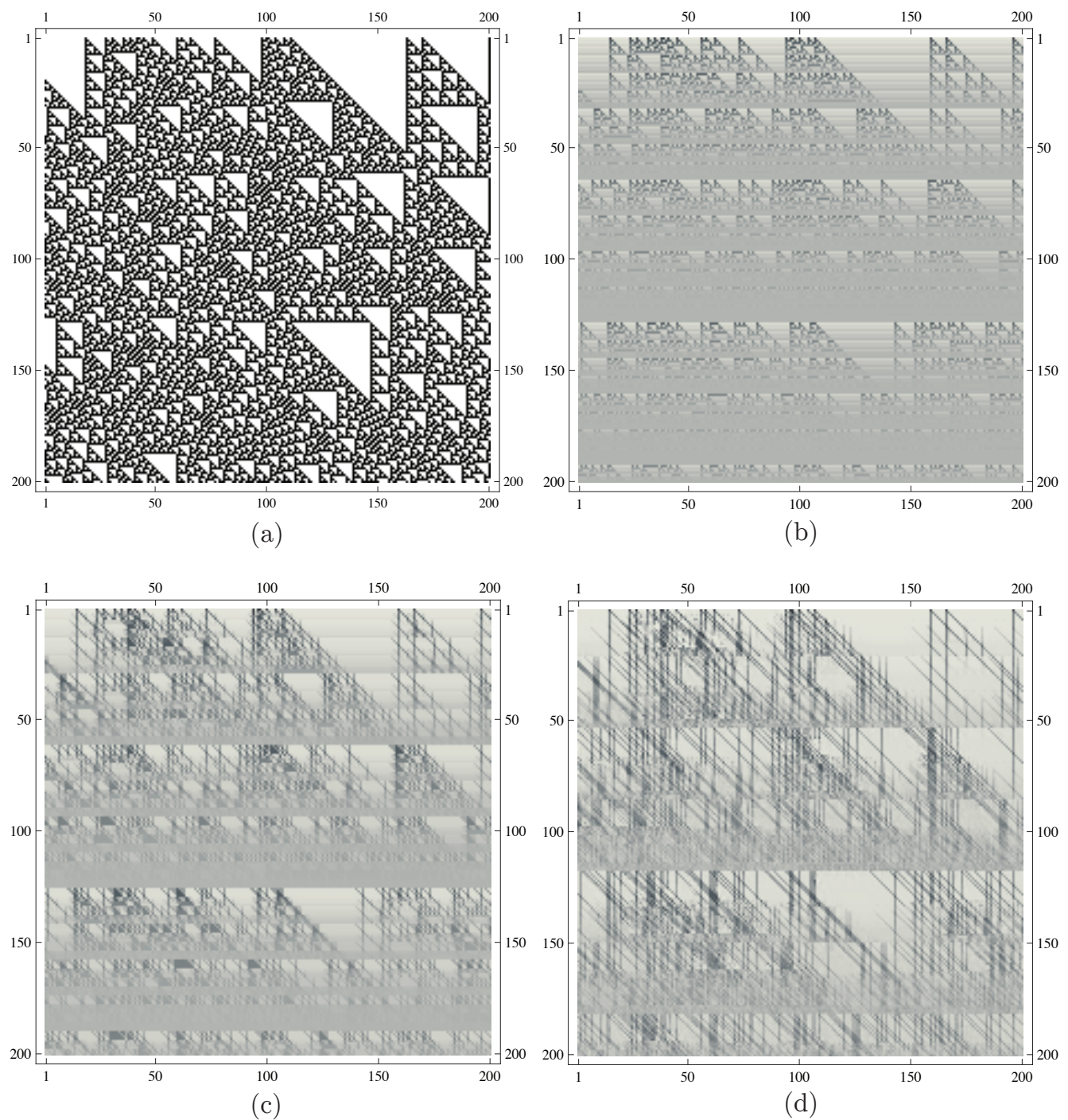
Figure 2. (**a**) The space-time pattern of elementary Rule 60, with time going downwards. (**b**)–(**d**) The local information for lengths 0, 4 and 12 of the conditional block (for the information characterization, the darker the pixel, the higher is the local information).
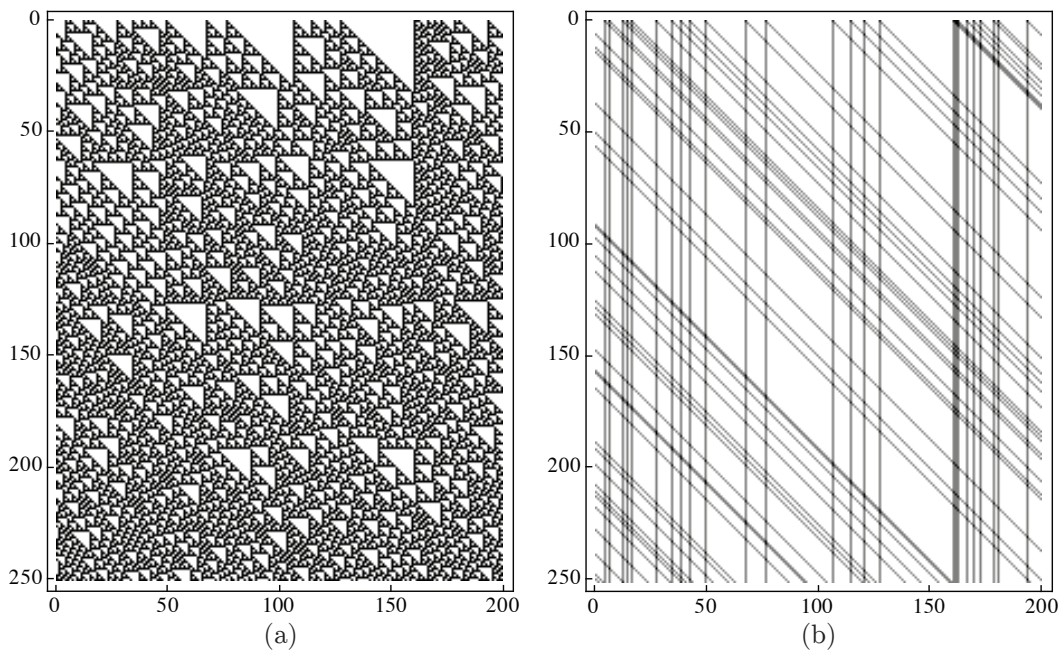
Figure 3. (**a**) The space-time pattern of Rule 60 is shown; (**b**) the corresponding information density in the limit of the infinite size of the conditional block. In this case, the local information can be analytically solved, and we see that each original local state "1" corresponds to high local information. In the time evolution, this quantity is split into two parts, one staying in the position, while the other one moves to the right.

2.3. Correlations and Complexity in Cellular Automata

By gradually extending the conditional block when determining the local information, we improve the probabilistic description about what to expect at the current position. More of the correlation information is brought in, and on average, this leads to a decreasing information gain as a function of conditional block length.

The information in correlations, k_m, over blocks of length m can be defined as:

$$k_m = \langle I_{i,m}^{(\mathrm{L})} - I_{i,m-1}^{(\mathrm{L})} \rangle \tag{12}$$
$$= \sum_{x_1...x_{m-1}} p(x_1...x_{m-1}) \sum_{x_m} p(x_m|x_1...x_{m-1}) \log \frac{p(x_m|x_1...x_{m-1})}{p(x_m|x_2...x_{m-1})} \ .$$

The last line shows that k_m is an average over a Kullback–Leibler divergence [13], where $p(x_m|x_1...x_{m-1})$ is compared to an "*a priori*" distribution $p(x_m|x_2...x_{m-1})$ with a shorter conditional block, which ensures that the correlation information is non-negative, $k_m \geq 0$. One can also define a "density information" by $k_1 = \log N - S_1$ (where $N = |\Lambda|$ is the number of local states), which characterizes information in the knowledge about frequencies of states compared to a uniform *a priori* distribution. Then, the total information of the spatial configuration can be fully decomposed into contributions from the different correlation lengths and the entropy s, $\log N = s + \sum_{m=1}^{\infty} k_m$; see, e.g., [10] (note that there is a more general scheme in which information in lattice configurations can be detected, in which one does not need to rely on a one-sided sample in the conditional part [11]).

The most well-known information-theoretic complexity quantity, the excess entropy [14] or the effective measure complexity [15], can be expressed in terms of the correlation information, even though this form is less commonly used,

$$\eta = \sum_{m=2}^{\infty} (m-1)k_m \ . \tag{13}$$

From this perspective, the excess entropy is a weighted sum of all correlation information quantities, using the length of the conditional block as a weight. This means that the excess entropy is related to the average correlation length (determined by the information characteristics) of the system.

For the elementary rule, R60, discussed in the previous section, one can show [16] that starting (at $t=0$) from a Bernoulli state with entropy below the maximum, $s < \log 2$, the excess entropy increases linearly in time,

$$\eta(t) = \begin{cases} ts \ , & \text{if } s < \log 2 \ , \\[2mm] 0 \ , & \text{if } s = \log 2 \ . \end{cases} \tag{14}$$

This clearly shows that even if the entropy is conserved, there may be complex dynamics that lead to correlation information being distributed over ever increasing distances in the system. In the long run, for rule R60, even if the entropy is as low as it was initiated at $t=0$, it may be intractable to detect the correlations in order to find this. If one would only look at the system locally, it may appear as if the entropy has increased. In fact, in the infinite time limit, it will almost always appear as if the system is in a microstate characterized by maximum entropy.

3. Transition from Micro to Macro

As a thought experiment, let us consider a deterministic lattice gas. Take the hexagonal lattice with up to six particles at each lattice site, with two possible particle types (A and B), pointing in the corresponding lattice directions (this means we have 3^6 possible states). The time step in this micro dynamics is composed of two parts: (i) collision rules conserve momentum and can be designed so that they are deterministic; and (ii) particles move to adjacent cells according to the direction of their velocity. A snapshot of a microstate after collisions and before translation is illustrated in Figure 4. This means that we have a perfectly reversible microscopic dynamics, and the entropy per lattice site is conserved in the time evolution. This is a two-component version of the FHP lattice gas, by Frisch, Hasslacher, and Pomeau [17].

In order to discuss the transition from micro to macro, let us now consider a microstate of a large system composed of many particles, where the length scales of concentration variations over space are large. We also assume that the density is low, *i.e.*, most of the lattice sites are empty. Now, if the starting point is randomly distributed concentration peaks of particle A in a background of B particles, we do have a state with lower entropy s than we would have had if all particles would be evenly distributed. When the micro-dynamics unfolds, longer and longer correlations are formed of increasing order.
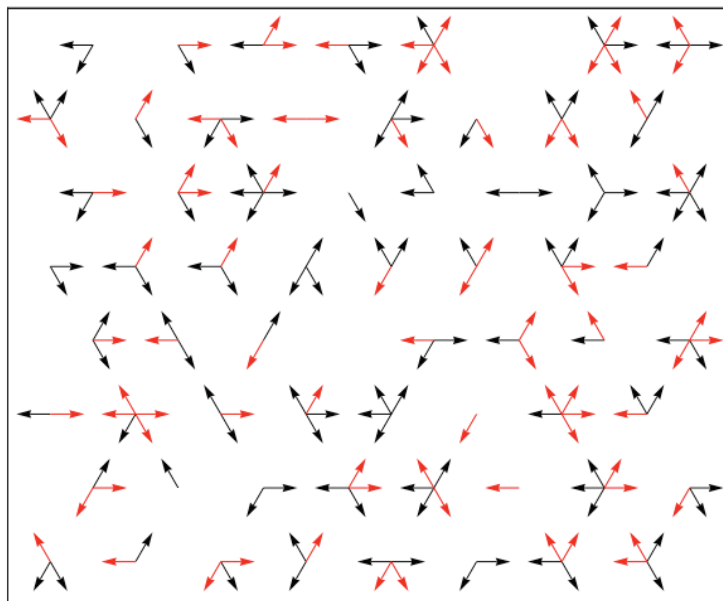
Figure 4. An illustration of the state in a two-component lattice gas on a hexagonal lattice (with black and red particles), in which particles have collided and are about to take the translational step moving to adjacent cells.

Initially, relatively few conditional lattice sites are needed for a good approximation of the local information, since the local concentration is reflected by the states in the conditional block and since there are not yet any long-range correlations in the system. For example, if there are many Type A molecules in the conditional block, we do expect that there is a high probability of finding an A molecule also in the next lattice site; or, even stronger, the relative frequencies of the molecular species in the conditional block serve as a very good guess for the probability distribution over the molecules to be found in the next lattice site.

When micro-level correlation lengths increase in the system, we lose information about the local state. The reasoning above still applies, as a first approximation, but it neglects the information present in the long-range high-order correlations. However, this information becomes increasingly more difficult to detect. Even in a simulation model, where we keep track of all particles, it is a computationally difficult task to extract this information.

The information that is, in general, available at the macroscopic level, as a result of a measurement of the system, is the information equivalent to what we get from the local conditional block, *i.e.*, the local concentration. This reasoning thus suggests that we have a transition in the description from the microscopic to the macroscopic level. At the micro-level, the conditional probabilities give the appropriate representation capturing the full physical description of the system, even including long-range correlations, which would imply constant entropy whenever the micro dynamics is reversible. At the macro-level, these probabilities are replaced by the local probabilities of finding molecules of a certain type, *i.e.*, the concentration variables of the system. Note that the micro-level conditional block gives us the information of the local concentration, *i.e.*, information equivalent to the locally-observed concentration at the macro-level.

3.1. From Conditional Probabilities to Local Concentrations

At the microscopic level, the local information contained in a lattice state i at position x is determined by, e.g., the left-sided conditional probability $p(i|i_{x-(m-1)}...i_{x-1})$, where (in this case) the local states to the left of x are possibly holding information about the lattice state in position x. If we extend the size of the block to infinity ($m \to \infty$), we will have the full conditional probability of the state in lattice site x, and the resulting local information, $-\log p(i|i_{x-(m-1)}...i_{x-1})$, then has an average that equals the conserved microscopic entropy: the quantity that cannot increase as all high-order long-range correlations are included, *cf.* Equations (9) and (11). By limiting the size m of the conditional block, we lose the information in those long-range correlations. The conditional block still contains some information about the state of site x as it provides us with a microscopic sample of locally-present particle types, which determines the "concentrations" of the different particle types at position x.

In the transition from micro to macro, we thus replace the conditional probabilities with concentration variables. Schematically, putting this in a one-dimensional lattice, for a molecule i at position x, we would have,

$$p(i|i_{x-(m-1)}...i_{x-1}) \to c_i(x) \ . \tag{15}$$

If the block size m is unbounded, the microscopic representation can potentially capture all correlations, and the resulting entropy measure is then conserved under a microscopic reversible dynamics. However, the macroscopic representation, $c_i(x)$, only make use of the information more locally. This can be expressed as $c_i(x) \approx p(i|i_{x-(m_0-1)}...i_{x-1})$, where m_0 is relatively small in the sense that concentration variations over m_0 can be neglected. The macroscopic representation is therefore not capturing any information that may be present in the long-range correlations that may have an origin in a low entropy initial state of the system. Therefore, an entropy calculation based on the macroscopic concentration $c_i(x)$ may indeed increase in the time evolution. (Note that the "empty" state is missing in the macro representation. This needs to be treated more carefully if the density is high, *i.e.*, if the molecular state probabilities are of the same order as the one for the empty states.)

If one would not extend the conditional block as time evolves, the uncertainty of the state one is observing, in the cell at position x, increases and so does, on average, the local information. It would in that case appear as if the entropy s of the microstate increases with time.

Therefore, if the size of the conditional block is kept constant, which would be the case if we let the local conditional block determine the probability of the next molecule and, in that way, use the local concentration as the description variable, we loose the information contained in long-range correlations.

3.2. The Resulting Macroscopic Entropy

The thought experiment we consider corresponds to the the mixing of ideal gases under the process of diffusion. We should therefore expect that the replacement of the description suggested by the micro-macro transition, Equation (15), also gives us the appropriate entropy characterization of the macroscopic system.

The information density of the microstate, as expressed in Equation (9), then by the use of Equation (15), gives us the entropy expressed in the macroscopic variable,

$$
\begin{aligned}
s &\approx \left\langle \ \log \frac{1}{p(i|i_{x-m+1}...i_{x-1})} \ \right\rangle \\
&\approx \frac{1}{L} \sum_x \sum_i c_i(x) \log \frac{1}{c_i(x)} + C \ ,
\end{aligned}
\tag{16}
$$

where L is the length of the system and C is a constant coming from the fact that we do not distinguish between the different particle directions. This expression, apart from a constant, a factor of Boltzmann's constant and the number of particles in the system, is the entropy of mixing used to describe ideal gases. Again, at the macro-level, the empty state is excluded from the sum. When empty positions in the microstate dominate, holding a fraction c of the lattice sites (with c close to one), the contribution to the entropy from the term $c \log c$ from the empty positions vanishes. That this approach, involving a discretisation of space, works for deriving the Sackur–Tetrode equation for the ideal gas entropy has been demonstrated before [1].

Note that if we would like to include chemical reactions in the model, while still keeping the reversibility for microscopic dynamics, then we would have to include also a heat bath that has the capability of absorbing or releasing the energy necessary for the corresponding microscopic reactions. An example of such a model, with a microscopic reversible mechanism for the energy dissipating phenomenon of diffusion-limited aggregation, has been constructed by D'Souza and Margolus [18].

If we consider the continuum limit, where $x \in \mathbb{R}$, we can easily derive the change of this entropy as a result of diffusion, where we now also express diffusion as a process at the macroscopic level,

$$
\frac{dc_i(x)}{dt} = D_i \nabla^2 c_i(x) \ ,
\tag{17}
$$

where D_i is the diffusion constant for species i (in this very restricted thought experiment, where the particles are moving with unit velocities on a lattice, the diffusion constant is the same for all species).

The change in entropy S, where system entropy $S = L s$ (considering a system in the thermodynamic limit $L \to \infty$), can then be obtained by:

$$
\begin{aligned}
\frac{dS}{dt} &= \frac{d}{dt} \int dx \ \sum_i c_i(x,t) \log \frac{1}{c_i(x,t)} \\
&= \int dx \ \sum_i D_i \frac{(\nabla c_i(x,t))^2}{c_i(x,t)} \geq 0 \ ,
\end{aligned}
\tag{18}
$$

where we used a partial integration assuming either periodic or non-flow boundary conditions. This is just the standard illustration that the total entropy of a closed system may not decrease under diffusion, even though we may have a decrease in the local entropy term $\sum_i c_i(x) \log(1/c_i(x))$.

The increase in the entropy based on the macroscopic description thus depends on the fact that we lose information due to increasing long-range high-order correlations.

4. Coarse-Graining of the Macrostate

In the transition from micro to macro, we leave the detailed microscopic descriptions that involve long-range and high-order correlations between particles or molecules. At the microscopic level, the different choices of conditional block length served as a basis for the decomposition of information, a decomposition that gave us the characterization of the microstate, both in terms of correlation information and how it is distributed over the system in the form of the excess entropy.

At the macro-level of a chemical system, when we use the concentration as the variables describing the system, there is another perspective that can be taken in order to make a decomposition of the information. One may ask the question of at what length scale do we find information and where in the system is information located. This question is possible to address by using the process of coarse-graining of the macroscopic concentration variables and using that as a basis for the decomposition of information [5,6,19]. The following is a brief review of that approach.

Since we are considering chemical systems in which reactions may occur, a proper information-theoretic characterization of a chemical pattern should include both the concentration distribution in the system, *i.e.*, $c_i(x)$, and the corresponding equilibrium concentrations, $c_{i,0}$. For simplicity, we will now use a concentration definition, which is normalized under the summation over molecular species, $\sum_i c_i(x) = 1$.

Instead of the entropy characterization of Equation (16), we now use the Kullback–Leibler divergence between $c_{i,0}$ as an *a priori* description and the actual concentration distribution $c_i(x)$. This information quantity is then integrated over the whole system to give the total information, K, of the chemical pattern,

$$K = \int dx\, K[c_0; c(x)] = \int dx \sum_i c_i(x) \log \frac{c_i(x)}{c_{i,0}} \ . \tag{19}$$

The information captured by the Kullback–Leibler divergence is quantifying order, and in that way, it is opposite of the entropy quantity. Entropy production typically means that the total information, K, is decaying and that the entropy of the system increases.

Note that the total information quantity, K, can also be derived from the grand canonical ensemble, as a quantity being proportional to the work extractable from the system described by $c_i(x)$ in an environment characterized by $c_{i,0}$, under the assumption that pressure and temperature are kept constant and that the Gibbs free energy of molecular species is in the form of ideal solutions. The total information (multiplied by particle density, temperature and Boltzmann's constant) can conversely be seen as the work needed to construct the system out of an equilibrium environment characterized by $c_{i,0}$.

4.1. Coarse-Graining of a Chemical Pattern

The coarse-graining of the macroscopic description is obtained by applying the convolution with a Gaussian of width r to the concentration distributions $c_i(x)$. This results in a coarse-grained concentration distribution, $\tilde{c}_i(r; x)$, which depends on a resolution parameter r representing the degree of coarse-graining applied,

$$\tilde{c}_i(r, x) = \frac{1}{(2\pi)^{d/2}} \int dz\, e^{-z^2/(2r^2)} c_i(x - z) \ , \tag{20}$$

where d is the dimension of the system (in the following, we assume periodic boundary conditions). For $r = 0$, we have perfect resolution and $\tilde{c}_i(0; x) = c_i(x)$, while in the limit $r \to \infty$, we get the average concentrations in the system, $\tilde{c}_i(r \to \infty; x) = \bar{c}_i$. The coarse-graining of a concentration pattern can be schematically represented as a series of pictures with different levels of resolution, as shown in Figure 5.
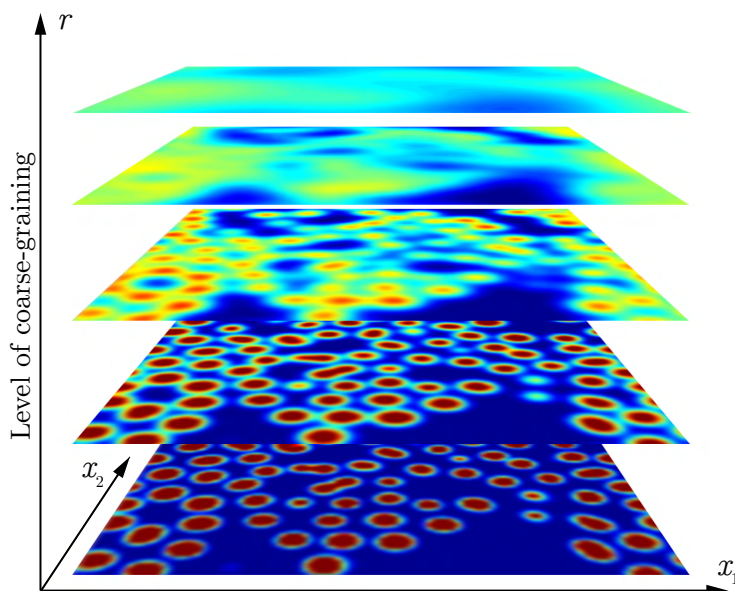
Figure 5. An illustration of a two-dimensional chemical pattern at different levels of coarse-graining, represented by the resolution length scale r, along with a schematic illustration of how the spatial information may be decomposed between different length scales (the color representation ranging from blue to red corresponds to concentration levels of the catalytic component from low to high concentrations).

4.2. Information Decomposition over Space and Levels of Coarse-Graining

The total information K of the chemical pattern can first be decomposed into two information quantities in a Kullback–Leibler form, representing the deviation from homogeneity, K_{spatial}, and deviation of the average concentrations from chemical equilibrium, K_{chem}, respectively. The spatial and the chemical information are therefore defined as:

$$K_{\text{spatial}} = \int dx \sum_i c_i(x) \log \frac{c_i(x)}{\bar{c}_i} \,, \tag{21}$$

$$K_{\text{chem}} = V \sum_i \bar{c}_i \log \frac{\bar{c}_i}{c_{i,0}} \,, \tag{22}$$

where V is the volume of the system, so that:

$$K = K_{\text{spatial}} + K_{\text{chem}} \,. \tag{23}$$

The spatial information is the term that can be further decomposed into contributions from different positions and different levels of coarse-graining. This decomposition was fully developed in [5] and results in an information density $k(r, x)$ spanning position x and length scale r,

$$k(r, x) = \sum_i \tilde{c}_i(r; x) \big(r \nabla \log \tilde{c}_i(r; x) \big)^2 \geq 0 \ . \tag{24}$$

This quantity locates information at the edges of the patterns in the spatial configuration. It is a complete decomposition of the spatial information,

$$K_{\text{spatial}} = \int \frac{dr}{r} \int dx \ k(r, x) \ . \tag{25}$$

The information density as function of position for two different levels of coarse graining is illustrated in Figure 6b,c, from the concentration pattern, Figure 6a, and in the Gray–Scott model of self-replicating spots [20,21]. For better resolution or less coarse-graining, the information density detects the separate concentration peaks, and the information density is high at the edges of the pattern. For the higher coarse-graining level, the concentration peaks are no longer detectable, but here, the areas that are empty can be seen instead, as they are characterized by a larger length scale. For more details, see [6].

(a) Concentration pattern   (b) $k(0.01, x)$ (c) $k(0.05, x)$
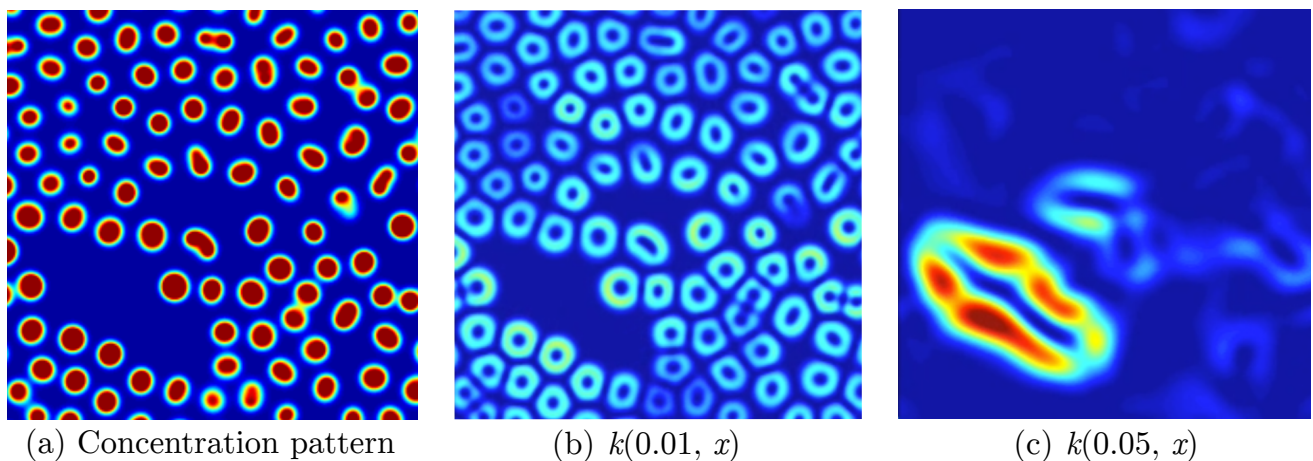
Figure 6. (**a**) A snapshot of the concentration pattern in the Gray–Scott model of self-replicating spots. The corresponding information density $k(r, x)$ at two different levels of coarse-graining, $r = 0.01$ and $r = 0.05$, is shown in (**b**) and (**c**), respectively (the concentration level, here of the self-catalytic component, as well as the information density are represented by colors from blue to red corresponding to low to high levels).

4.3. Flows of Information Driven by Reaction Diffusion Dynamics

The total information K in a chemical pattern changes in time due to chemical reactions and diffusion, and in addition to this, if the system is open, there is an information change due to chemical flows across the system boundary. The reaction-diffusion dynamics can be written,

$$\dot{c}_i(x, t) = D_i \nabla^2 c_i(x, t) + F_i(c(x, t)) + B_i(c_i(x, t)) \ , \tag{26}$$

where F_i is the reaction term and B_i is the diffusion flow across the system boundary (here, assumed to be perpendicular to the extension of the system) that depends on the contact with a reservoir (the reservoir is here assumed to be characterized by constant concentrations $c_i^{(\text{res})}$ and boundary diffusion constants b_i for the different species i).

First, we note that for a closed system, $B_i = 0$, the total information K decays, as a consequence of Equations (19) and (26), due to the entropy production in the chemical reactions and the diffusion,

$$\frac{d}{dt}K = \int dx \left(-\sum_i D_i \frac{(\nabla c_i(x))^2}{c_i(x)} + \sum_i F_i(c(x)) \log \frac{c_i(x)}{c_{i,0}} \right) \leq 0 . \tag{27}$$

The two terms in the integrand (with an added minus sign) correspond to the local entropy production due to diffusion and due to chemical reactions, respectively. Note that the also the second sum is negative, which follows from the relations between the equilibrium concentrations $c_{i,0}$ and the reaction functions F_i. The information loss resulting from the dynamics, when $dK/dt < 0$, can be seen as a loss of information from the macroscopic description. From the system perspective, this is due to the underlying reversible microscopic dynamics leading to correlation information being distributed over increasing distances and a higher degree of freedom, in the same way as we discussed about the initial diffusion example in Section 3.2. The information loss expressed by Equation (27) thus comes from the fact that the system is described by the macroscopic variables $c_i(x)$. If we would have a complementary description in terms of microscopic configurations along with a reversible microscopic dynamics, with all molecular correlations kept, then the full entropy of the system would not increase, but the long-range high-order correlations at the micro-level would ensure that the entropy would stay constant. Therefore, the loss of information at the macro-level due to entropy production in Equation (27) can thus be viewed as a "flow of information" from macro to micro, and we therefore define an outflow of information from the macroscopic level into the microscopic level, $j_r(r, x)$ at the coarse-graining level $r = 0$, as the entropy production,

$$j_r(r = 0, x) = \sum_i D_i \frac{(\nabla c_i(x))^2}{c_i(x)} - \sum_i F_i(c(x)) \log \frac{c_i(x)}{c_{i,0}} \geq 0 . \tag{28}$$

This means that as long as the system is not in (internal) equilibrium, there will be a loss of information until a uniform equilibrium state is reached with $K = 0$.

For a system to be capable of forming and maintaining a chemical pattern with $K > 0$, it is then clear that the system needs to be open. Since information, just like free energy, cannot be created, there must be an "inflow" of information, typically in the form of free energy, to the system. If there is a chemical species i serving as a "fuel" for the reaction processes and if the system is in diffusion contact with a reservoir containing this species with a high concentration, then there is an inflow of information that may maintain a high level of chemical information K_{chem}. Similarly, if there is a "waste" product of the reaction scheme, the reservoir should hold that component in low concentration, allowing for a diffusive outflow of the waste. Again, this process will contribute to keeping up the chemical information of the system.

The spatial information K_{spatial}, though, with its characteristic length scales r and positions in the system, needs to be formed by the reaction-diffusion processes that drive the changes in information density $k(r, x, t)$ in between the inflow and the outflow of information. This means that we want to

identify flows across length scales $j_r(r, x, t)$ and across space $j_x(r, x, t)$ that connect the inflow and the outflow, but that may lead to aggregation of information at certain positions when patterns are formed. The flow across length scales j_r is defined to be positive in the direction towards smaller length scales, in accordance with the direction of the flow leaving the system, as defined in Equation (28). We define the flow $j_r(r, x, t)$ by generalizing the flow defined by the entropy production at level $r = 0$ to an arbitrary level of coarse-graining,

$$j_r(r, x, t) = \sum_i \left(D_i \frac{(\nabla \tilde{c}_i)^2}{\tilde{c}_i} - \tilde{F}_i(c) \log \frac{\tilde{c}_i}{c_{i0}} \right) , \tag{29}$$

where we have replaced c_i and F_i in Equation (28) with their coarse-grained counterparts. Now, one can derive the flow j_r at the fully coarse-grained level, as $j_r(r \to \infty) = -dK_{\text{chem}}/dt$, showing that the chemical information is transformed into the flow j_r. This is consistent with the view that the chemical information, which does not reflect any spatial order, but only chemical disequilibrium, serves as a possible source for information at finite levels of coarse-graining. Again, only in the open system, the inflow of free energy may sustain a certain level of chemical information, which would be necessary for maintaining spatial chemical patterns.

This means that the information density $k(r, x, t)$, which characterizes the spatial patterns in the system, should obey a continuity equation for information,

$$\dot{k}(r, x, t) = r \frac{\partial}{\partial r} j_r(r, x, t) - \nabla \cdot j_x(r, x, t) + J(r, x, t) . \tag{30}$$

We have also introduced an additional term J, which is a result of the fact that a system that is open for diffusion flows across the system boundary will experience a decay in spatial structure due to this. This decay is generally small compared to the pattern forming processes and only plays a minor role in the information dynamics picture.

The continuity equation for information, Equation (30), then leads to a definition of the spatial flow j_x. For the details of the derivation of the terms in the continuity equation, we refer the reader to [19]. These terms are all based on the coarse-grained level r concentration distributions $\tilde{c}_i(r, x)$. We end up with the following definition for the spatial flow j_x,

$$j_x(r, x, t) = -r^2 \nabla \sum_i \tilde{F}_i(c) \log \frac{\tilde{c}_i}{c_{i,0}} , \tag{31}$$

and for the sink term J, we have,

$$J(r, x, t) = -\sum_i b_i \big(\tilde{c}_i + c_i^{(\text{res})} \big) \big(r \nabla \log \tilde{c}_i \big)^2 \leq 0 . \tag{32}$$

Here, b_i is the diffusion constant across the system boundary, and $c_i^{(\text{res})}$ is the reservoir concentration of the chemical component i.

In this picture of how information is participating in the process of pattern formation, we see that the spatial flow j_x only depends on the reaction dynamics. Diffusion only enters in the flow j_r across coarse-graining levels. This is due to the fact that the coarse-graining, for each molecular component, is equivalent to a diffusion process. Thus, the decay of a pattern only due to diffusion is characterized by a vertical flow of information towards finer length scales.

During pattern formation, there is an information flow originating at the largest length scales, in the form of chemical information, K_{chem}, which leads to information flowing "down" towards finer levels of resolution. The flow may be redirected and accumulated at certain length scales and positions when spatial patterns are formed. If the system is closed, the initial chemical information will eventually be depleted, and the built-up spatial structure will decay as the information disappears into microscopic degrees of freedom via the flow j_{r} and the system approaches a homogenous equilibrium. Only if the system is open, e.g., by being in contact with a reservoir that supplies a fuel and removes reaction products, the chemical information can be maintained, and the pattern forming processes can be sustained.

5. Conclusions

In this paper, we have presented an information perspective on coarse-graining from the microscopic level via the micro-macro transition towards higher characteristic length scales at the macroscopic levels. The starting point is a local information quantity, applicable for microstate characterization. The advantage with this local information is that its spatial average equals a microscopic entropy—the standard entropy defined for stationary stochastic processes that generate typical microstate configurations—which, in general, also equals the statistical mechanics entropy of the system.

The novel contribution here is that we show how this local information quantity can be followed in the transition from micro to macro, when local states in the micro configuration are aggregated into macroscopic concentration variables. In this transition, it becomes clear what information is lost. At the micro-level, when microscopic dynamics is reversible, the correlational information that in the time evolution becomes increasingly more difficult to detect is conserved and implies that the microstate entropy is a conserved quantity. If the long-range high-order correlations cannot be detected, the system will appear, and behave, as if the microstate entropy increases. Therefore, in the micro-macro transition, when local states are aggregated into concentration variables, the correlational information cannot be kept. It is then demonstrated that the local information quantity is transformed into a local macroscopic entropy. The simple example of diffusion then directly illustrated that this macro entropy, as is well known, is increasing under diffusion until a homogenous equilibrium is reached.

The perspective presented connects also to a further coarse-graining process, previously developed in [6,19], applicable to chemical reaction-diffusion dynamics. In this framework, we identify a decomposition of information into contributions from different positions and different length scales. The reaction-diffusion dynamics imposes flows of information from larger length scales (higher levels of coarse-graining) towards finer levels of resolution; see Figure 7. These flows obey a continuity equation of information, and we identify the loss of information from the macro-level, *i.e.*, the entropy production due to diffusion and chemical reactions, as a flow of information leaving the macro-level disappearing to the micro-level, where it is distributed over the microscopic degrees of freedom. In the micro dynamics, this is the information previously discussed that is spread out over high-order and long-range correlations over increasing distances.
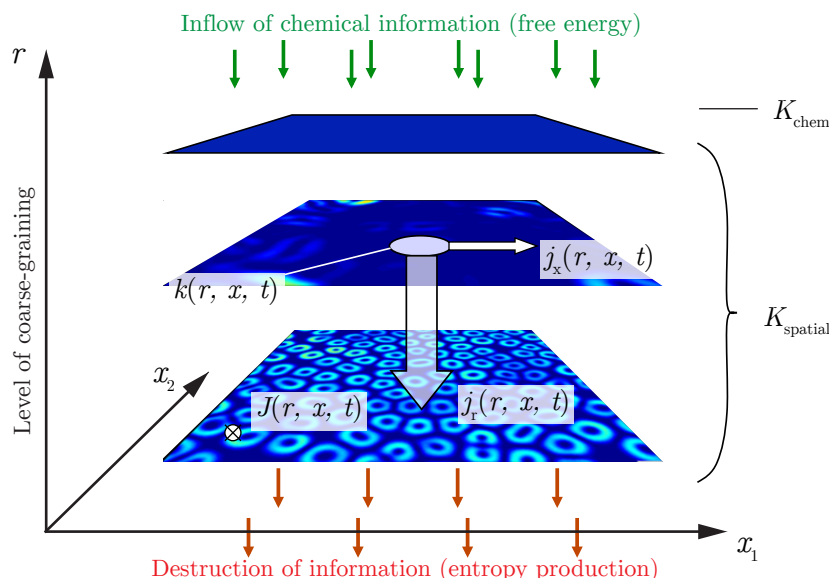
Figure 7. Schematic representations of the flows characterizing the information dynamics in an open reaction-diffusion system capable of pattern formation. When a spatial pattern is formed, we see an increase in the information density $k(r, x)$ at certain positions and at certain levels of coarse-graining corresponding to the characteristic length scales of the pattern. This increase comes from an accumulation of the information flows, originating at the top level in the form of chemical information, K_{chem}, flowing downwards across scales (j_{r}) and being redirected by the spatial flow (j_{x}). Information is continuously leaving the system as entropy production at the finest level of resolution, when information is distributed over microscopic degrees of freedom not represented in the macroscopic description. Only if the system is open, the information loss can be balanced by an inflow of chemical information (free energy), and the pattern formation process can be sustained.

Acknowledgments

Conflicts of Interest

The author declares no conflict of interest.

References

1. Lindgren, K. Microscopic and macroscopic entropy. *Phys. Rev. A* **1988**, *38*, 4794–4798.
2. Schreiber, T. Measuring information transfer. *Phys. Rev. Lett.* **2000**, *85*, 461–464.
3. Prokopenko, M.; Lizier, J.T.; Price, D.C. On thermodynamic interpretation of transfer entropy. *Entropy* **2013**, *15*, 524–543.

4. Prokopenko, M.; Lizier, J.T. Transfer entropy and transient limits of computation. *Sci. Rep.* **2014**, *4*, doi:10.1038/srep05394.

5. Eriksson, K.E.; Lindgren, K. Structural information in self-organizing systems. *Phys. Scr.* **1987**, *35*, 388–397.

6. Lindgren, K.; Eriksson, A.; Eriksson, K.E. Flows of information in spatially extended chemical dynamics. In *Artificial Life IX: Proceedings of the Ninth International Conference on the Simulation and Synthesis of Living Systems*; Pollack, J., Bedau, M.A., Husbands, P., Watson, R.A., Ikegami, T., Eds.; The MIT Press: Cambridge, MA, USA, 2004; pp. 456–460.

7. Gray, R.M. *Probability, Random Processes, and Ergodic Properties*; Springer: Berlin/Heidelberg, Germany, 2009.

8. Shannon, C.E. A mathematical theory of communication. *Bell Syst. Tech. J.* **1948**, *27*, 379–423, 623–656.

9. Helvik, T.; Lindgren, K.; Nordahl, M. Continuity of information transport in surjective cellular automata. *Commun. Math. Phys.* **2007**, *272*, 53–74.

10. Lindgren, K. Correlations and random information in cellular automata. *Complex Syst.* **1987**, *1*, 529–543.

11. Helvik, T.; Lindgren, K. Expressing the entropy of lattice systems as sums of conditional entropies. *J. Stat. Phys.* **2014**, *155*, 687–702.

12. Lizier, J.T.; Prokopenko, M.; Zomaya, A.Y. Information modification and particle collisions in distributed computation. *Chaos: Interdiscip. J. Nonlinear Sci.* **2010**, *20*, 037109.

13. Kullback, S.; Leibler, R.A. On Information and Sufficiency. *Ann. Math. Stat.* **1951**, *22*, 79–86.

14. Feldman, D.P.; Crutchfield, J.P. Structural information in two-dimensional patterns: Entropy convergence and excess entropy. *Phys. Rev. E* **2003**, *67*, 051104.

15. Grassberger, P. Toward a quantitative theory of self-generated complexity. *Int. J. Theor. Phys.* **1986**, *25*, 907–938.

16. Lindgren, K.; Nordahl, M.G. Complexity measures and cellular automata. *Complex Syst.* **1988**, *2*, 409–440.

17. Pomeau, B.; Hasslacher, Y.; Frisch, U. Lattice-gas automata for the Navier-Stokes equation. *Phys. Rev. Lett.* **1986**, *56*, 1505–1508.

18. D'Souza, R.M.; Margolus, N.H. A thermodynamically reversible generalization of Diffusion Limited Aggregation. *Phys. Rev. E* **1999**, *60*, 264–274.

19. Lindgren, K. *Information Theory for Complex Systems—An information Perspective on Complexity in Dynamical Systems, Physics, and Chemistry*; Lecture Note for Chalmers; Chalmers University of Technolgy: Göteborg, Sweden, 2014.

20. Lee, K.J.; McCormick, W.D.; Ouyang, Q.; Swinney, H.L. Pattern formation by interacting chemical fronts. *Science* **1993**, *261*, 192–194.

21. Pearson, J.E. Complex Patterns in a Simple System. *Science* **1993**, *261*, 189–192.