Information in statistical physics

BEYOND SHANNON—The Structure & Meaning of Information

Institute of Advanced Studies, University of Amsterdam, May 6-8, 2019

Kristian Lindgren
Chalmers University of Technology
Gothenburg, Sweden

- Jaynes' information-theoretic formulation of statistical mechanics
- Two directions:
  - Macroscopic framework: Coarse-graining of patterns
  - Microscopic framework: Order/disorder in correlations

Information theory and statistical mechanics

(Jaynes, 1956)

Derivation of statistical physics distributions using the maximum entropy principle, under physical constraints, which gives the general form of the Gibbs distribution (as an equilibrium characterization).

Entropy: $$S = k_{\mathrm{B}} \sum_i p_i \log \frac{1}{p_i}$$

Constraints:

		Lagrangian variables
1) energy	$\sum_i p_i h(i) = U$	β
2) # of molecules	$\sum_i p_i f_k(i) = N_k \quad (k = 1, \ldots, M)$	λ_k
3) normalization	$\sum_i p_i = 1$	$(\mu - 1)$

Gibbs distribution: $$p_i = \exp\Big(- \mu - \beta h(i) - \sum_k \lambda_k f_k(i) \Big)$$

Lagrangian variables: $$\mu = \frac{pV}{k_{\mathrm{B}} T}, \quad \beta = \frac{1}{k_{\mathrm{B}} T}, \quad \lambda_k = -\frac{\mu_k}{k_{\mathrm{B}} T}$$

pressure p temperature T chemical potential μ_k

Information theory and statistical mechanics
(Jaynes, 1956)

Two frameworks, illustrating different directions:

1. The macroscopic direction, with coarse-graining, exemplified by patterns in chemical reaction-diffusion dynamics [1].

2. The microscopic direction, where we identify characteristic features of microstates that reflect thermodynamic properties; example given by reversible Ising dynamics in 1D [2].

[1] (Eriksson & Lindgren, 1987; Lindgren et al, 2004; Lindgren, 2015)
[2] (Lindgren & Olbrich, 2017)

The macroscopic direction — thermodynamics

The entropy of the Gibbs distribution results in Gibbs equation

$$S = k_{\mathrm{B}} \sum_i p_i \log \frac{1}{p_i} \quad \longrightarrow \quad U = TS - pV + \sum_k \mu_k N_k$$

Consider a system (V, U, N_k, S) in an environment characterized by (p_0, T_0, μ_{k0}). A relevant quantity is the relative information (or Kullback-Leibler divergence) between the corresponding Gibbs distributions, p and $p^{(0)}$:

$$K[p^{(0)}; p] = \sum_i p_i \log \frac{p_i}{p_i^{(0)}}$$

This results in the relation (dating back to Gibbs, 1873)

$$k_{\mathrm{B}} T_0 K[p^{(0)}; p] = U + p_0 V - T_0 S - \sum_k \mu_{k0} N_k \ = \ E$$

where E is the *exergy* or *available energy*, i.e., the maximum work obtainable when bringing the system into equilibrium with the environment $p^{(0)}$.

Information in a chemical pattern

For patterns in a chemical system we focus on μ_k, and we may assume constant temperature and pressure, $T = T_0$ and $p = p_0$, which with Gibbs equation results in

$$E = \sum_k N_k(\mu_k - \mu_{k0}) = k_{\mathrm{B}} T_0 \frac{N}{V} \int_V dx \sum_k c_k(x) \log \frac{c_k(x)}{c_{k0}}$$

where we have assumed ideal solutions, $\mu_k \sim \log c_k$, where c_k are concentrations, and also showing the generalization to spatial patterns $c_k(x)$ in the system.

$$K = \int_V dx \sum_k c_k(x) \log \frac{c_k(x)}{c_{k0}}$$

K is a relative information quantity that captures the physical information in a chemical pattern described by $c_k(x)$ with respect to an equilibrium reference c_{k0}.

Information in a chemical pattern

This information K can be decomposed with respect to position and length scales (using coarse-graining). First we note that

$$K = \int_V dx \sum_k c_k(x) \log \frac{c_k(x)}{c_{k0}} =$$

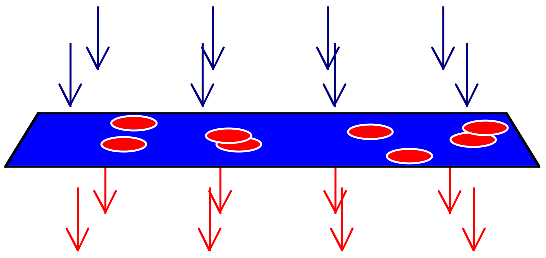
$$= \int_V dx \sum_k c_k(x) \log \frac{c_k(x)}{\bar{c}_k} \ + \ V \sum_k \bar{c}_k \log \frac{\bar{c}_k}{c_{k0}} \ = \ K_{\text{spatial}} \ + \ K_{\text{chem}}$$

where $\bar{c}_k$ is the average concentration in the system. K_{spatial} and K_{chem} quantifies deviation from spatial homogeneity and chemical equilibrium, respectively.

Gray-Scott model (self-replicating spots)

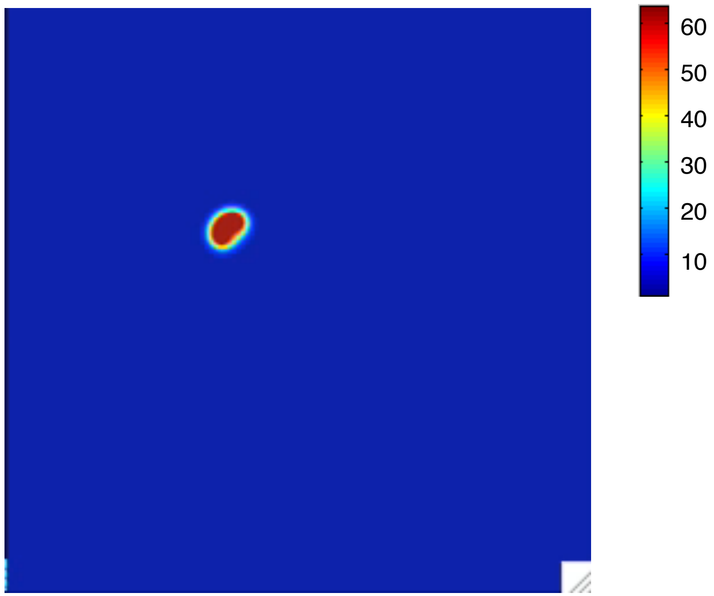
$$U + 2V \rightarrow 3V$$
$$V \rightarrow G$$

Gray & Scott, *Chem Eng Sci* (1984),
Pearson, *Science* (1993), and Lee et al, (1993).



Reaction-diffusion dynamics:

$$\frac{dc_{\mathrm{U}}}{dt} = D_{\mathrm{U}}\nabla^2 c_{\mathrm{U}} - (c_{\mathrm{U}} - k_{\mathrm{back}}c_{\mathrm{V}})c_{\mathrm{U}}^2 + F(1 - c_{\mathrm{U}})$$
$$\frac{dc_{\mathrm{V}}}{dt} = D_{\mathrm{V}}\nabla^2 c_{\mathrm{V}} + (c_{\mathrm{U}} - k_{\mathrm{back}}c_{\mathrm{V}})c_{\mathrm{U}}^2 - kc_{\mathrm{V}} - Fc_{\mathrm{V}})$$



Resolution – length scale

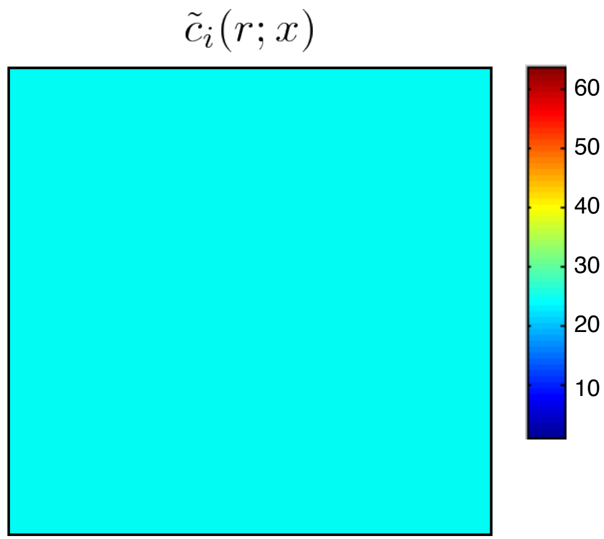
- We define the pattern of a certain component i at resolution r by "kernel smoothing" (convolution) of $c_i(x)$ with a Gaussian of width r

$$\tilde{c}_i(r;x) = \frac{1}{\sqrt{2\pi}r} \int_{-\infty}^{\infty} e^{-w^2/2r^2} c_i(x-w)\,dw$$

- with the properties

$$\tilde{c}_i(0;x) = c_i(x)$$

$$\tilde{c}_i(\infty;x) = \bar{c}_i(x)$$



Worst resolution ($r \to \infty$)

Resolution – length scale

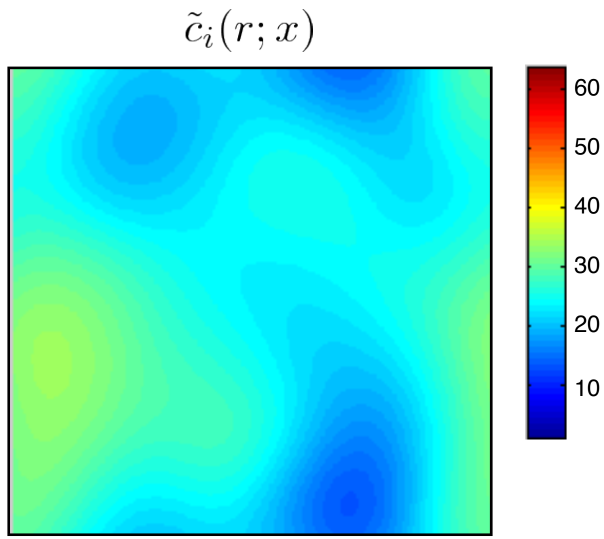
- We define the pattern of a certain component i at resolution r by "kernel smoothing" (convolution) of $c_i(x)$ with a Gaussian of width r

$$\tilde{c}_i(r;x) = \frac{1}{\sqrt{2\pi}r} \int_{-\infty}^{\infty} e^{-w^2/2r^2} c_i(x-w)\, dw$$

- with the properties

$$\tilde{c}_i(0;x) = c_i(x)$$

$$\tilde{c}_i(\infty;x) = \bar{c}_i(x)$$



Bad resolution (high r)

Resolution – length scale

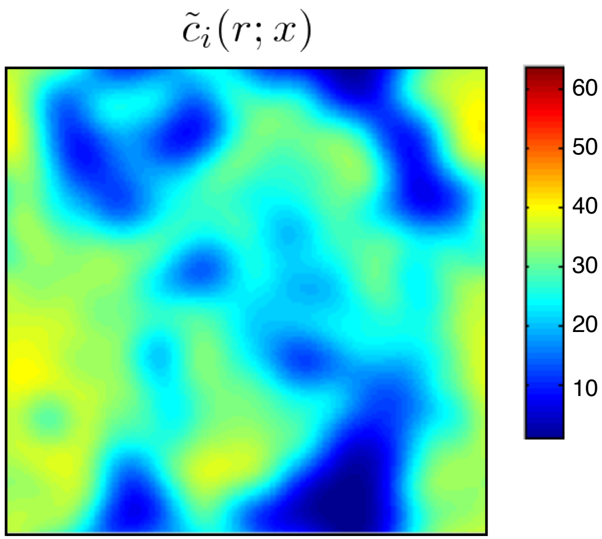
- We define the pattern of a certain component i at resolution r by "kernel smoothing" (convolution) of $c_i(x)$ with a Gaussian of width r

$$\tilde{c}_i(r;x) = \frac{1}{\sqrt{2\pi}r}\int_{-\infty}^{\infty} e^{-w^2/2r^2}\, c_i(x-w)\, dw$$

- with the properties

$$\tilde{c}_i(0;x) = c_i(x)$$

$$\tilde{c}_i(\infty;x) = \bar{c}_i(x)$$



Better resolution (decreasing r)

Resolution – length scale

- We define the pattern of a certain component i at resolution r by "kernel smoothing" (convolution) of $c_i(x)$ with a Gaussian of width r

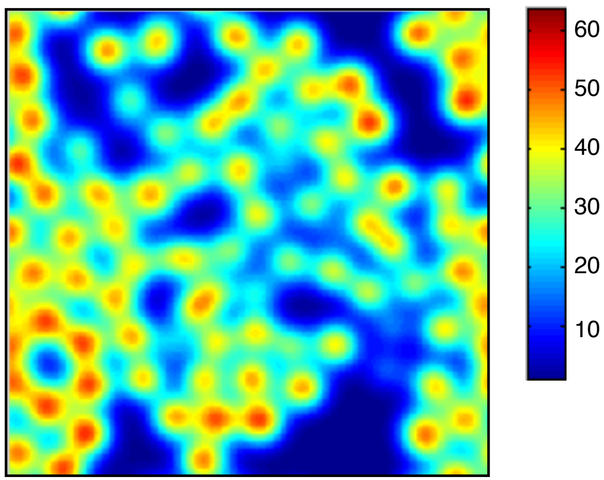
$$\tilde{c}_i(r; x) = \frac{1}{\sqrt{2\pi} r} \int_{-\infty}^{\infty} e^{-w^2/2r^2} c_i(x - w)\, dw$$

- with the properties

$$\tilde{c}_i(0; x) = c_i(x)$$

$$\tilde{c}_i(\infty; x) = \bar{c}_i(x)$$

$\tilde{c}_i(r; x)$



Better resolution (decreasing r)

Resolution – length scale

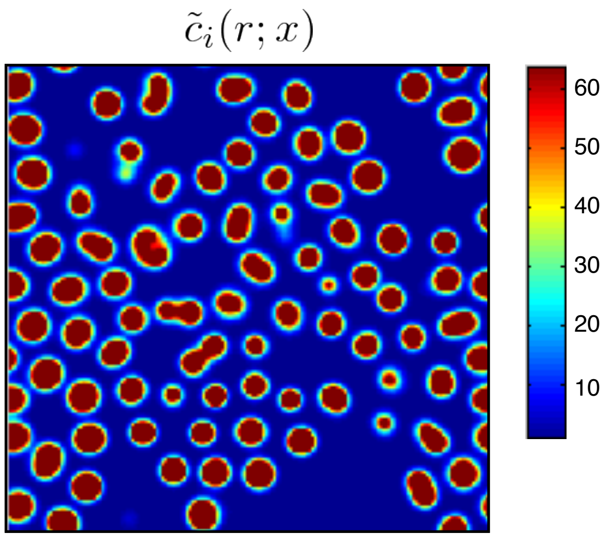
- We define the pattern of a certain component i at resolution r by "kernel smoothing" (convolution) of $c_i(x)$ with a Gaussian of width r

$$\tilde{c}_i(r;x) = \frac{1}{\sqrt{2\pi}r}\int_{-\infty}^{\infty} e^{-w^2/2r^2}\, c_i(x-w)\, dw$$

- with the properties

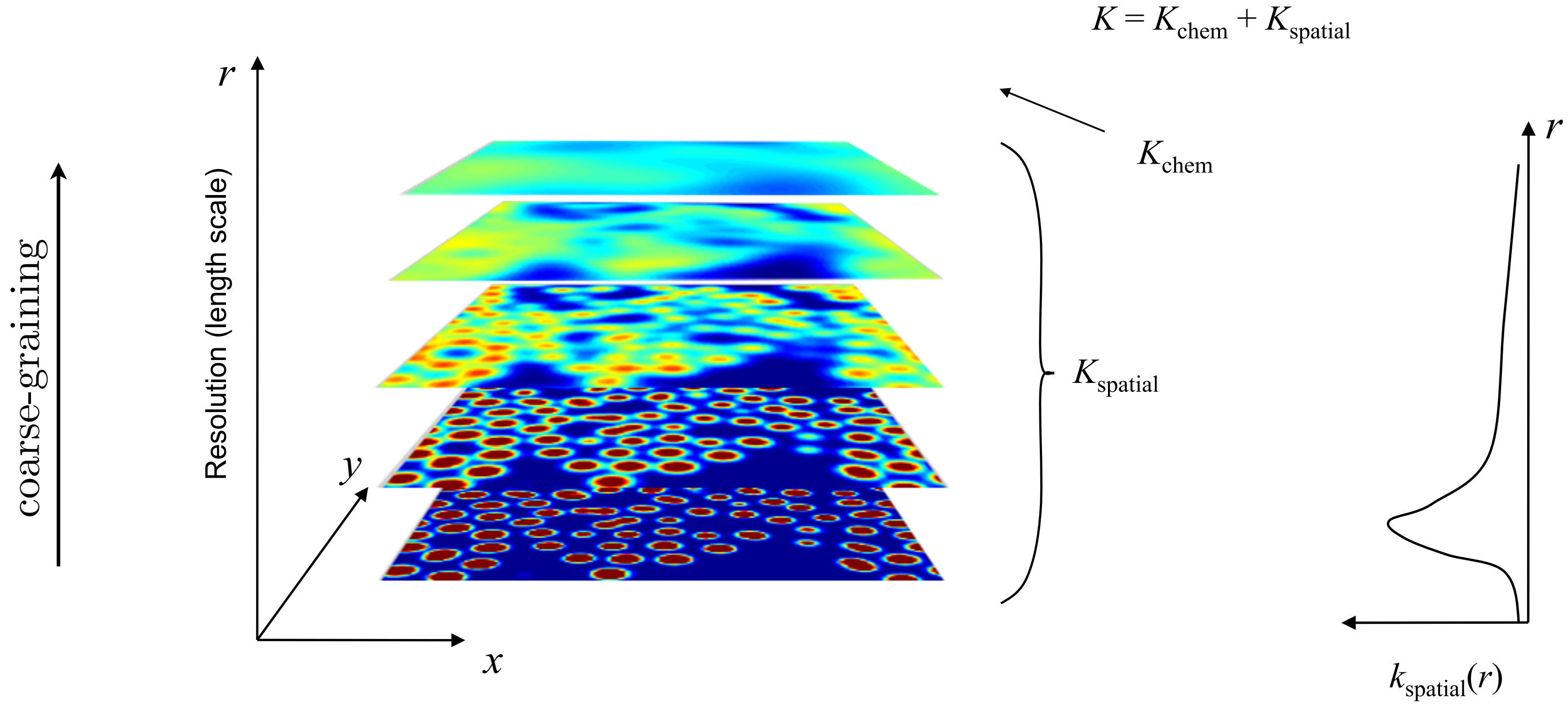
$$\tilde{c}_i(0;x) = c_i(x)$$

$$\tilde{c}_i(\infty;x) = \bar{c}_i(x)$$



High resolution ($r \approx 0$)

Space spanned by resolution and position



$$K_{\text{spatial}} = \int_0^\infty \frac{dr}{r} \int_V dx\, k(r,x) \qquad \text{where} \qquad k(r,x) = \sum_i \tilde{c}_i(r,x)\left(r\, \nabla \ln \tilde{c}_i(r,x)\right)^2 \geq 0$$

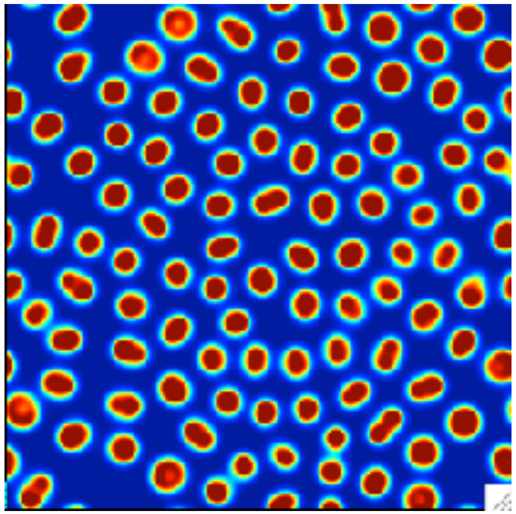
information density

$$k_{\text{spatial}}(r) = \int dx\, k(r,x)$$

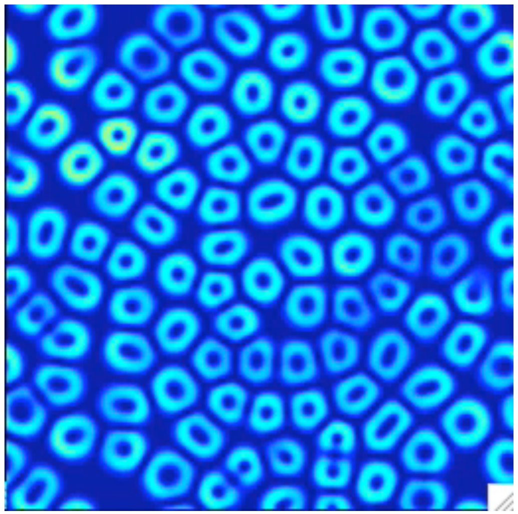
Information density in the Gray-Scott model

- The information density for two resolution levels r illustrates the presence of spatial structure at different length scales.

Concentration of V:
$c_V(\mathbf{x}, t)$

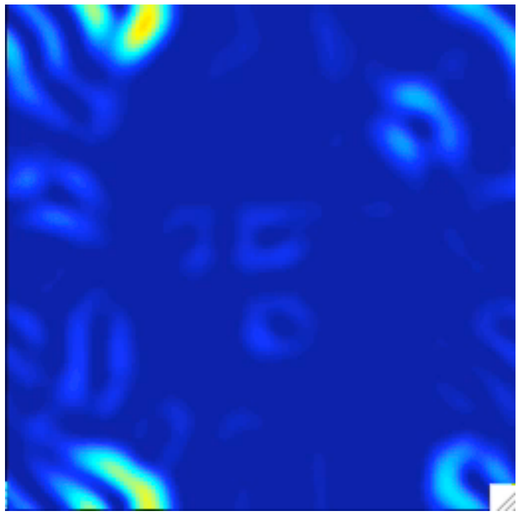


Information density:
$k(r{=}0.01, \mathbf{x}, t)$



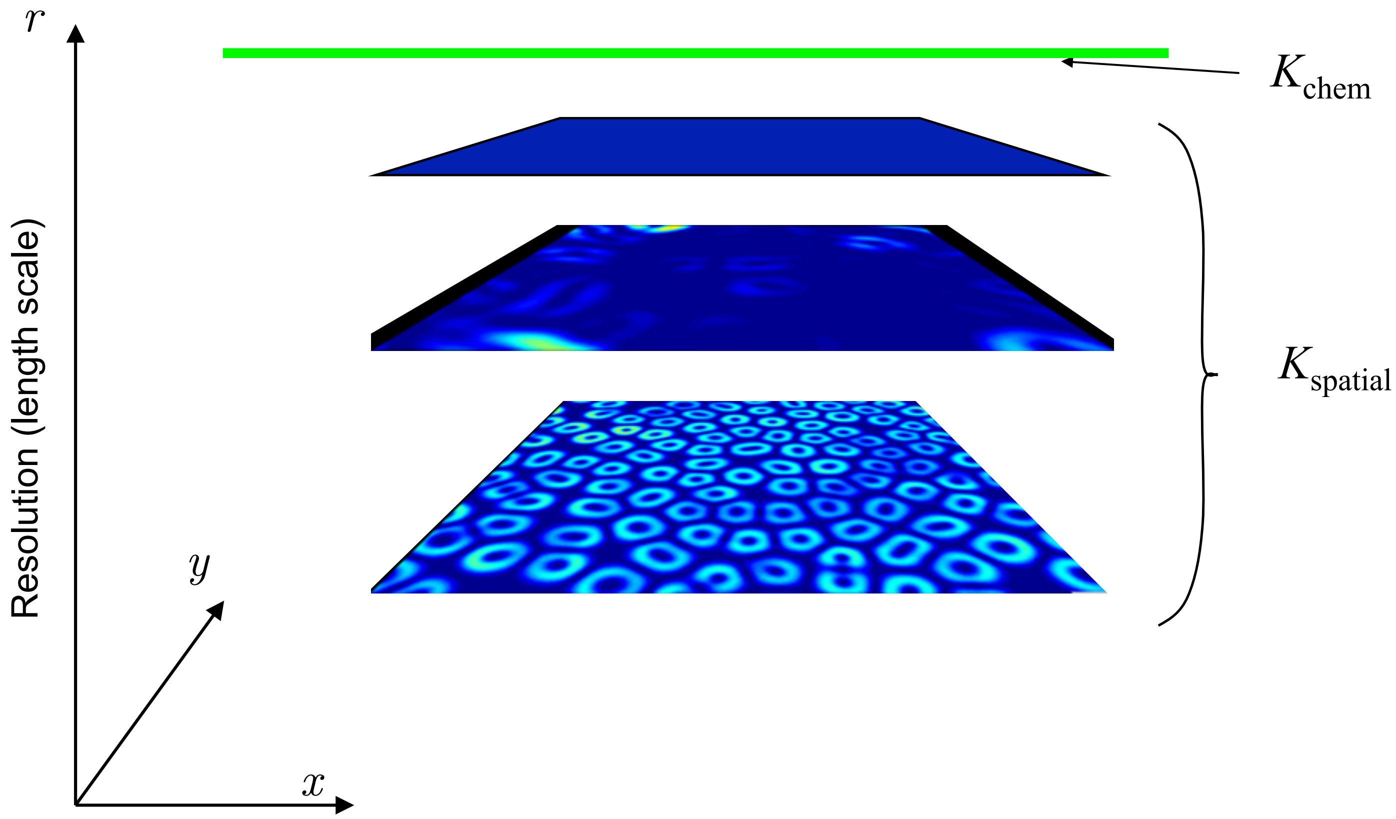
Good resolution

$k(r{=}0.05, \mathbf{x}, t)$



Worse resolution

Continuity equation for information



Information change driven by reaction-diffusion dynamics

Dynamics of $c_k(x, t)$ given by reaction-diffusion equations (with possible flow B_k across system boundary in case of an open system),

$$\frac{dc_k}{dt} = D_k \nabla^2 c_k + F_k(c) + B_k(c_k)$$

Information in the chemical pattern is destroyed by the processes of diffusion and reactions, due to entropy production σ

$$\frac{dK}{dt} = -\int_V dx \Big(\underbrace{\sum_k D_k \frac{(\nabla c_k)^2}{c_k}}_{\sigma_{\text{diffusion}} \geq 0} - \underbrace{\sum_k F_k(c) \log \frac{c_k}{c_{k0}}}_{\sigma_{\text{reactions}} \geq 0} \Big)$$

Information change driven by reaction-diffusion dynamics

Dynamics of $c_k(x,\ t)$ given by reaction-diffusion equations (with possible flow B_k across system boundary in case of an open system),

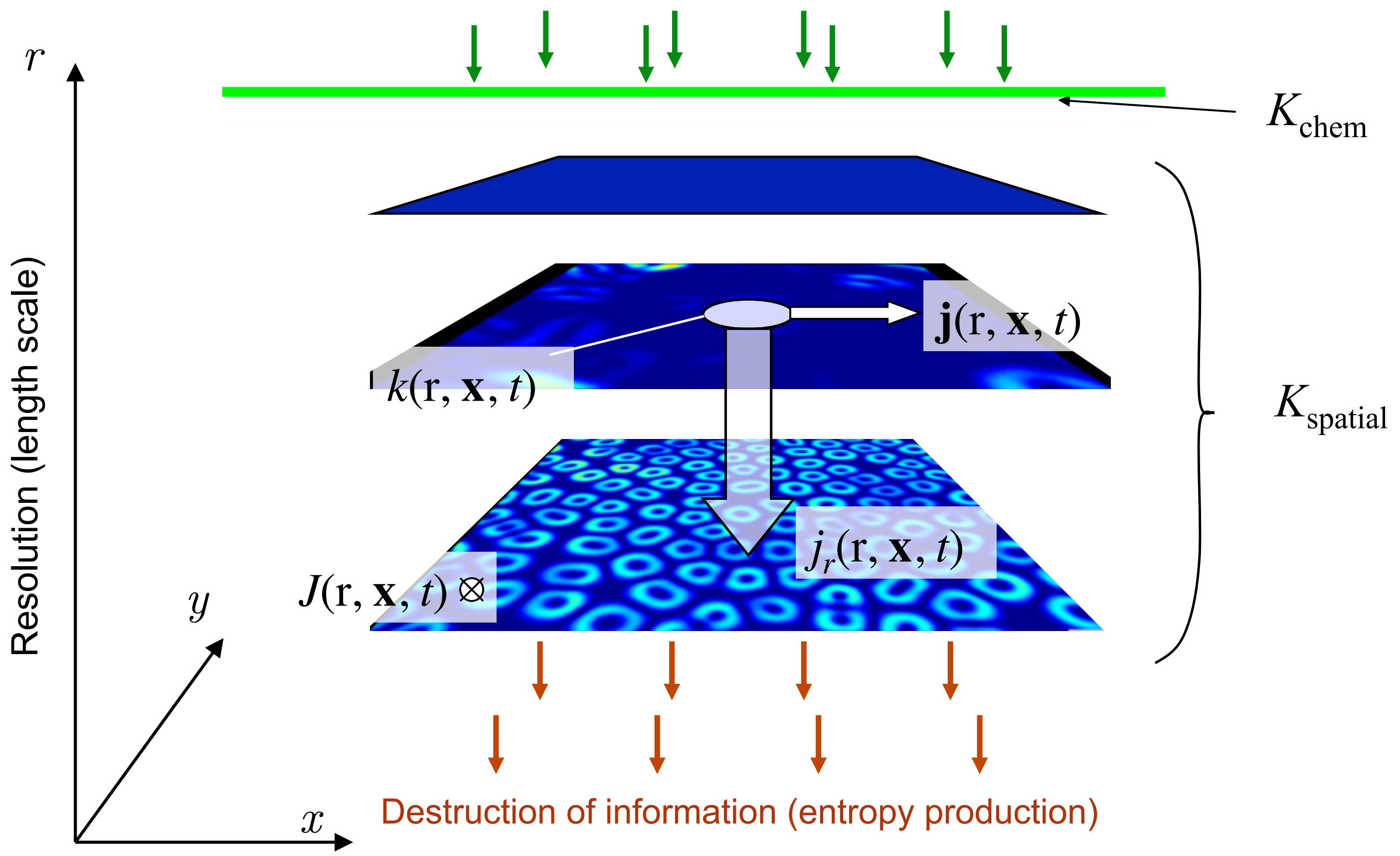
$$\frac{dc_k}{dt} = D_k \nabla^2 c_k + F_k(c) + B_k(c_k)$$

Information in the chemical pattern is destroyed by the processes of diffusion and reactions, due to entropy production σ

$$\frac{dK}{dt} = -\int_V dx \Big(\underbrace{\sum_k D_k \frac{(\nabla c_k)^2}{c_k}}_{\sigma_{\text{diffusion}} \geq 0} - \underbrace{\sum_k F_k(c) \log \frac{c_k}{c_{k0}}}_{\sigma_{\text{reactions}} \geq 0} \Big)$$

Flows of chemicals across the system boundaries may lead to sustaining a certain level of chemical information K_{chem}, which may drive a pattern formation (or maintenance) process.

Continuity equation for information



$$\frac{d}{dt} k(r,\mathbf{x},t) = r \frac{\partial}{\partial r} j_r(r,\mathbf{x},t) - \nabla \cdot \mathbf{j}(r,\mathbf{x},t) + J(r,\mathbf{x},t)$$

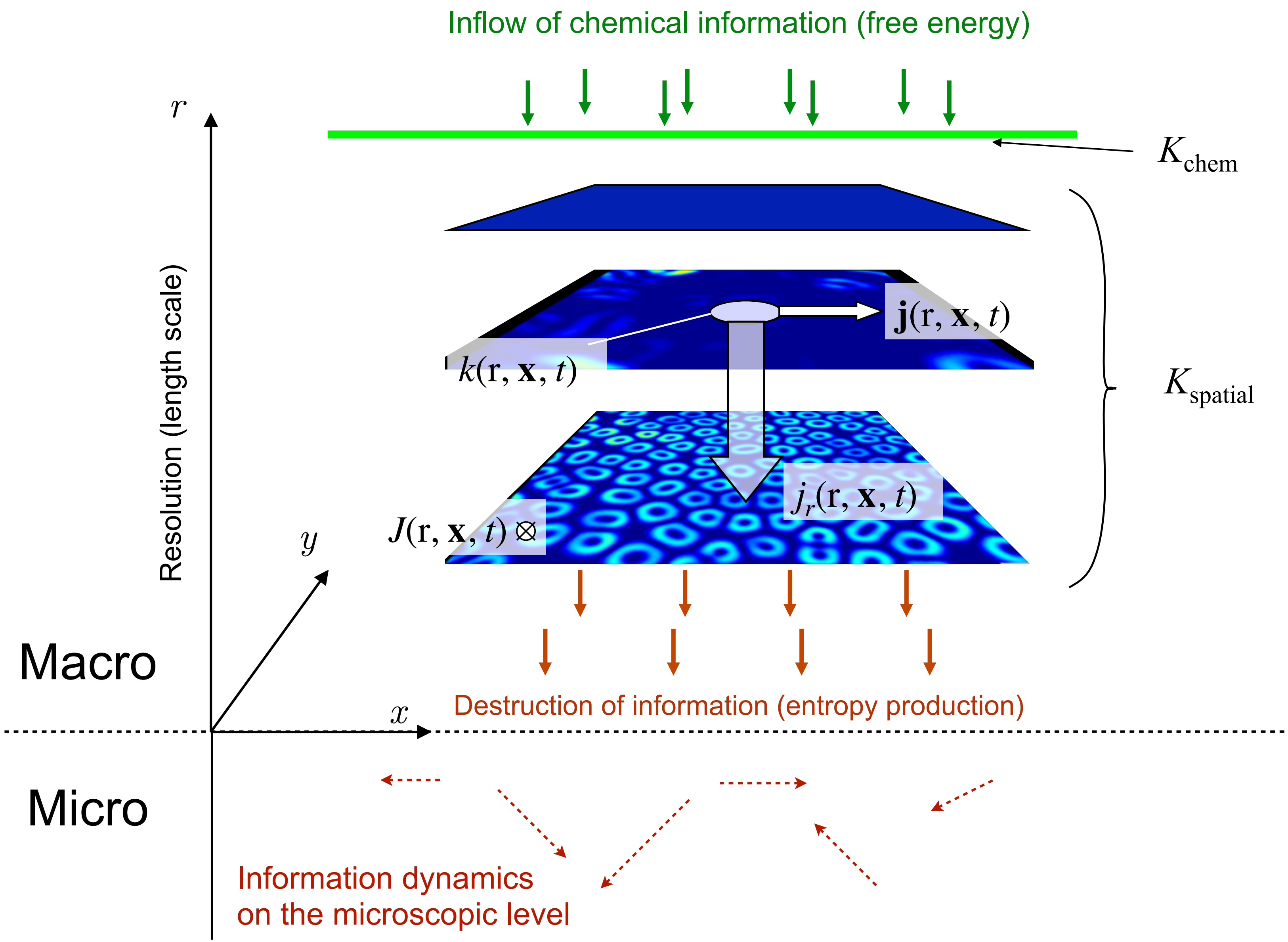
Flow in scale

Sinks (open system)

Information density

Flow in space

Continuity equation for information



The microscopic direction

Decomposition of information in entropy (disorder) and redundancy (order) provides us with a framework for investigating how (apparent) disorder may increase in reversible dynamics.

We will assume the system to be a lattice system in 1 dimension with n possible states per site. The total information per site of $I_{\text{tot}} = \log\, n$ is then

$$I_{\text{tot}} = s \; + \; k_{\text{corr}} \qquad (\text{ entropy density } + \text{ redundant information })$$

where

$$s = \lim_{m \to \infty} \frac{1}{m} S_m \qquad (\; S_m \text{ is the entropy for } m\text{-length sequences })$$

$$k_{\text{corr}} = \sum_{m=1}^{\infty} k_m$$

The terms k_m adding up to the total redundant information can be interpreted as correlation information over sequences of length m

$$k_m = -\Delta_2 S_m \geq 0$$

The microscopic direction

One key characteristic of such a system is its *excess entropy* η which is proportional to the average correlation length

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

The microscopic direction

In a closed physical system obeying a reversible time dynamics, the entropy (and hence the redundancy) is a conserved quantity.

The entropy density as we defined it, can in the thermodynamic limit (infinite system size) be associated with a single microstate. (Assuming spatial ergodicity.)

It is also the case that in equilibrium, for a spatially discrete system, the entropy density of a such a microstate equals the statistical mechanics entropy of the macrostate.

The question is then: How can a reversible microdynamics with a conserved entropy be understood for a system that approaches (or appears to approach) an equilibrium state with a higher entropy?

The microscopic direction

If the system is closed, the answer must include how redundant information is spread out over increasing correlation lengths.

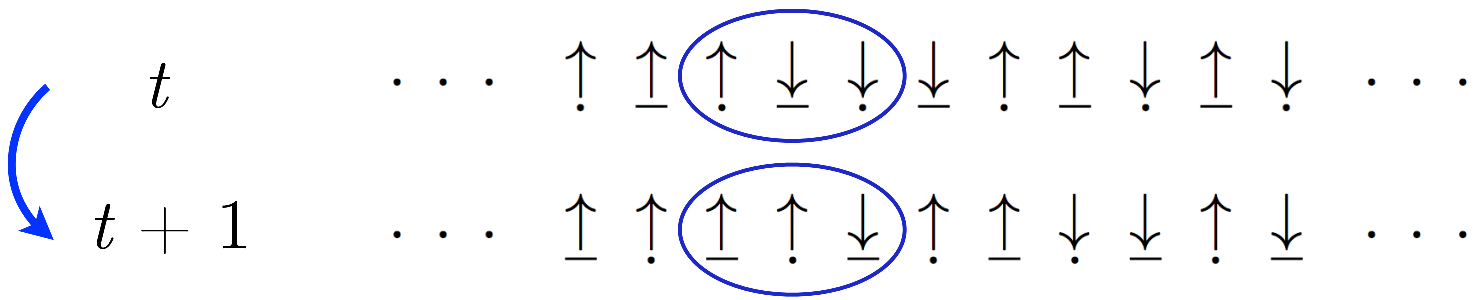
We illustrate how this happens in detail by the Q2R model (cellular automaton) that with local energy conservation flips spins in a one-dimensional Ising system[1].

[1] (Lindgren & Olbrich, 2017)

A reversible and energy conserving Ising dynamics

<u>The Q2R rule[1] in one dimension:</u>

- Update spins by alternating between the two sub-lattices, of odd and even positions respectively.
- Interaction energy u given by: $u(\uparrow\uparrow) = -1$, $u(\uparrow\downarrow) = +1$.
- A spin in an updating state is flipped if energy is not changed.

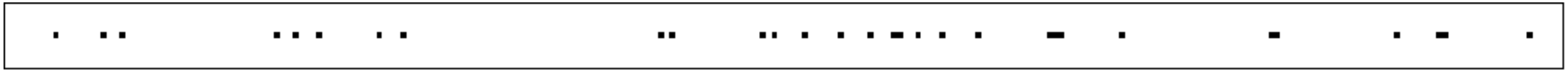


Updating state: _

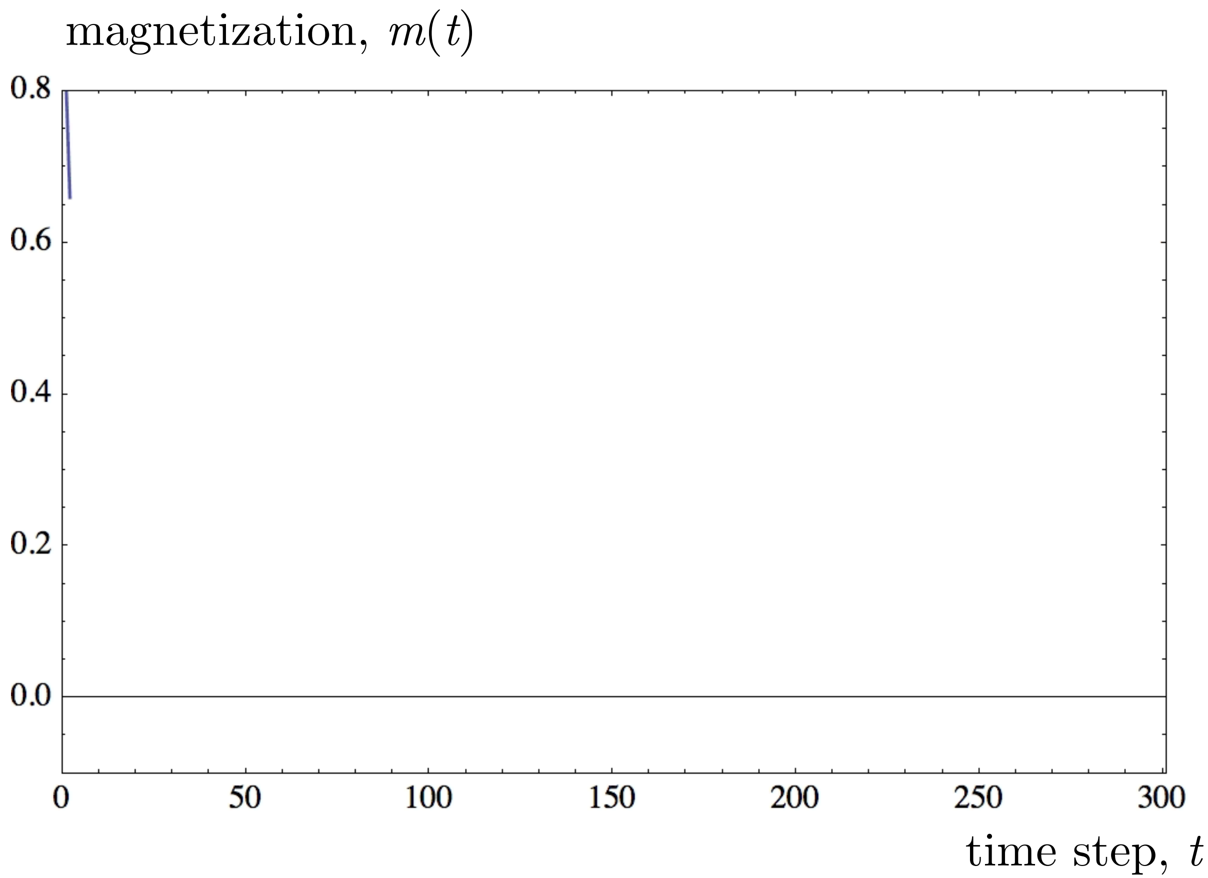
Quiescent state: .

[1] Vichniac, 1984

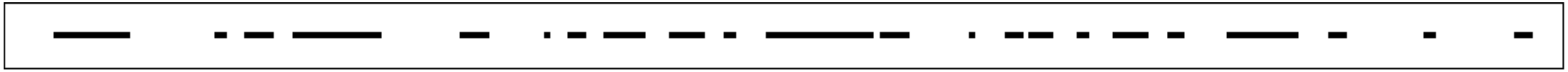
Reversible Ising dynamics in 1D



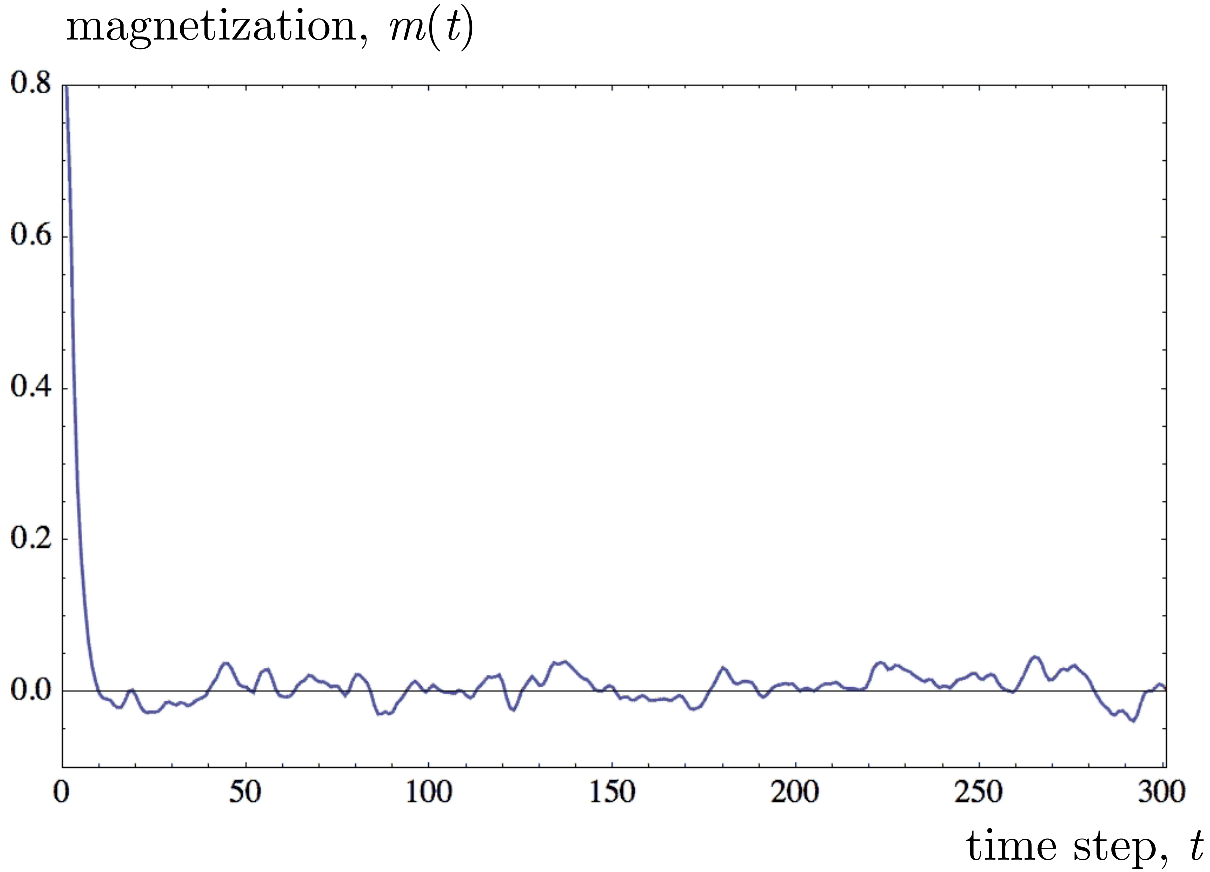
Showing 250 sites from a system of length 2^{14}.



Reversible Ising dynamics in 1D



Showing 250 sites from a system of length 2^{14}.

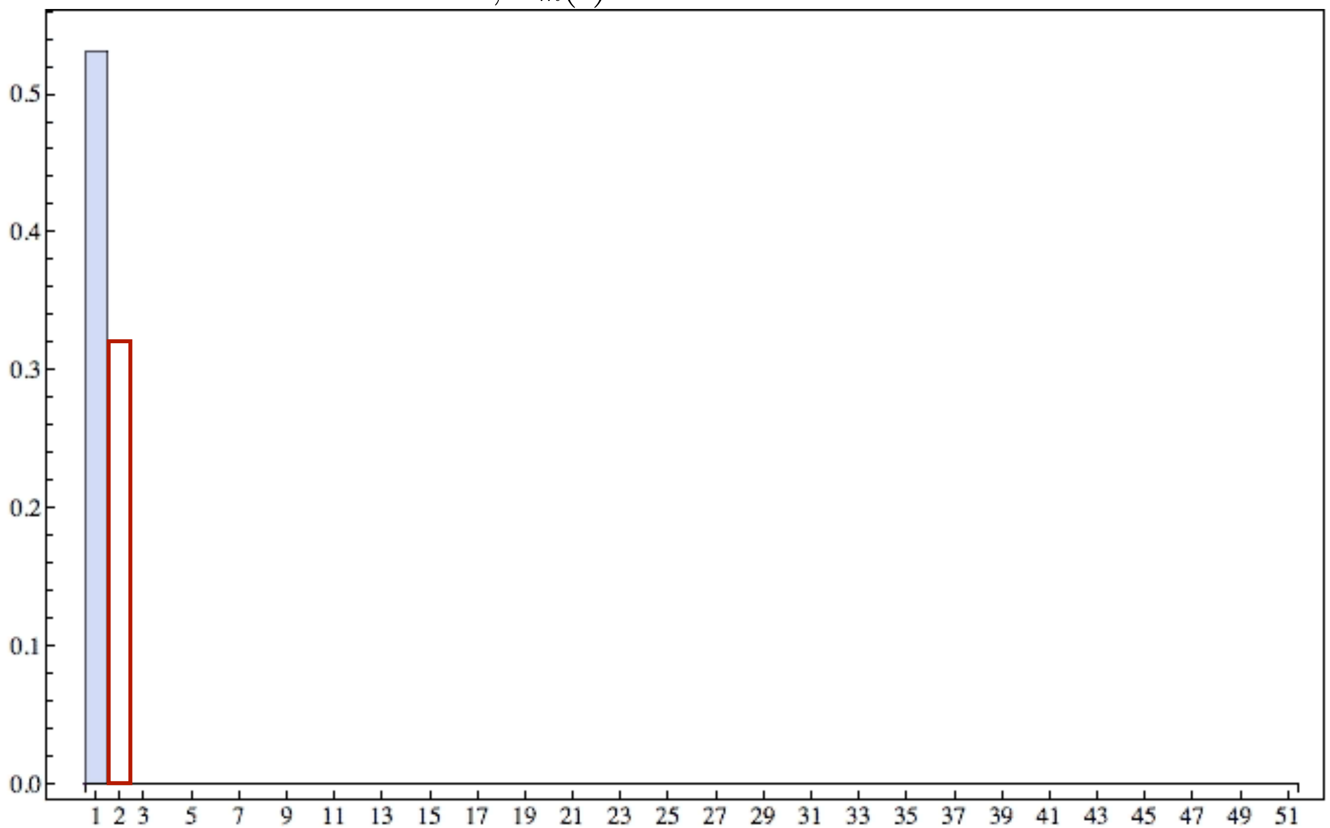


Evolution of correlation information

Showing the correlation information k_m, including the density information k_1, up length $m = 51$, over time ($t = 0, 1, 2, ..., 60$).

Contributions to correlation information — ordered information

correlation information, $k_m(t)$



correlation lengths, m

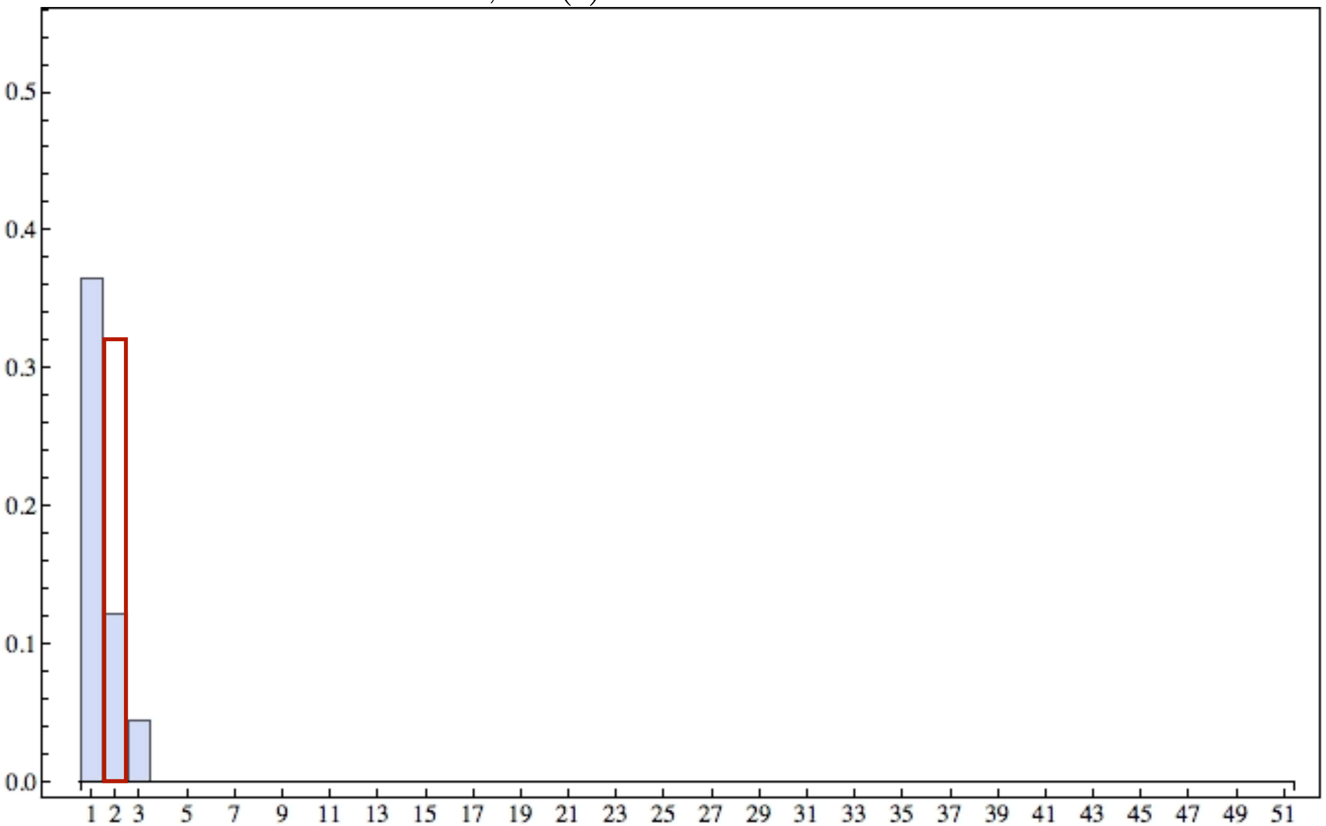
□ — correlation information at equilibrium

Evolution of correlation information

Showing the correlation information k_m, including the density information k_1, up length $m = 51$, over time ($t = 0, 1, 2, ..., 60$).

Contributions to correlation information — ordered information

correlation information, $k_m(t)$



correlation lengths, m

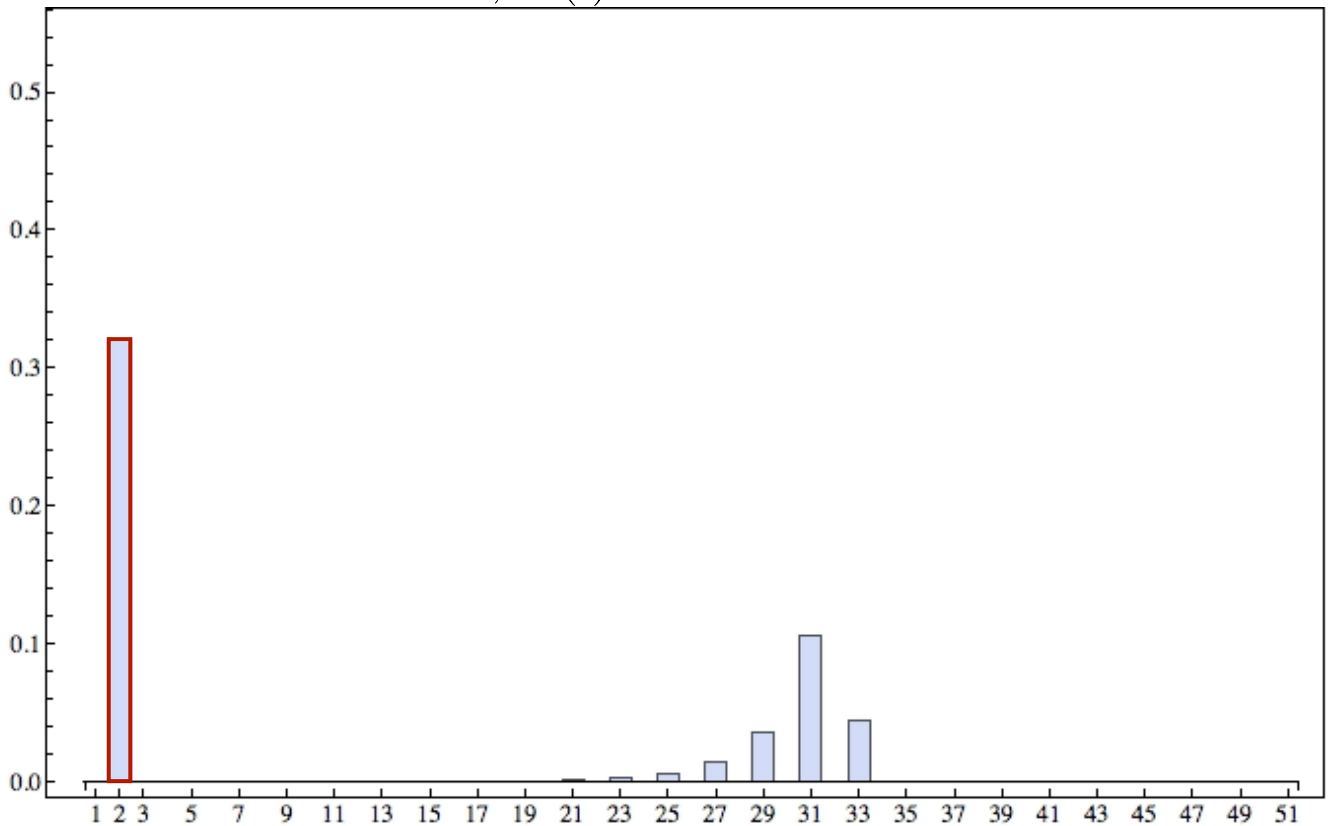
□ — correlation information at equilibrium

Evolution of correlation information

Showing the correlation information k_m, including the density information k_1, up length $m = 51$, over time ($t = 0, 1, 2, ..., 60$).

Contributions to correlation information — ordered information

correlation information, $k_m(t)$



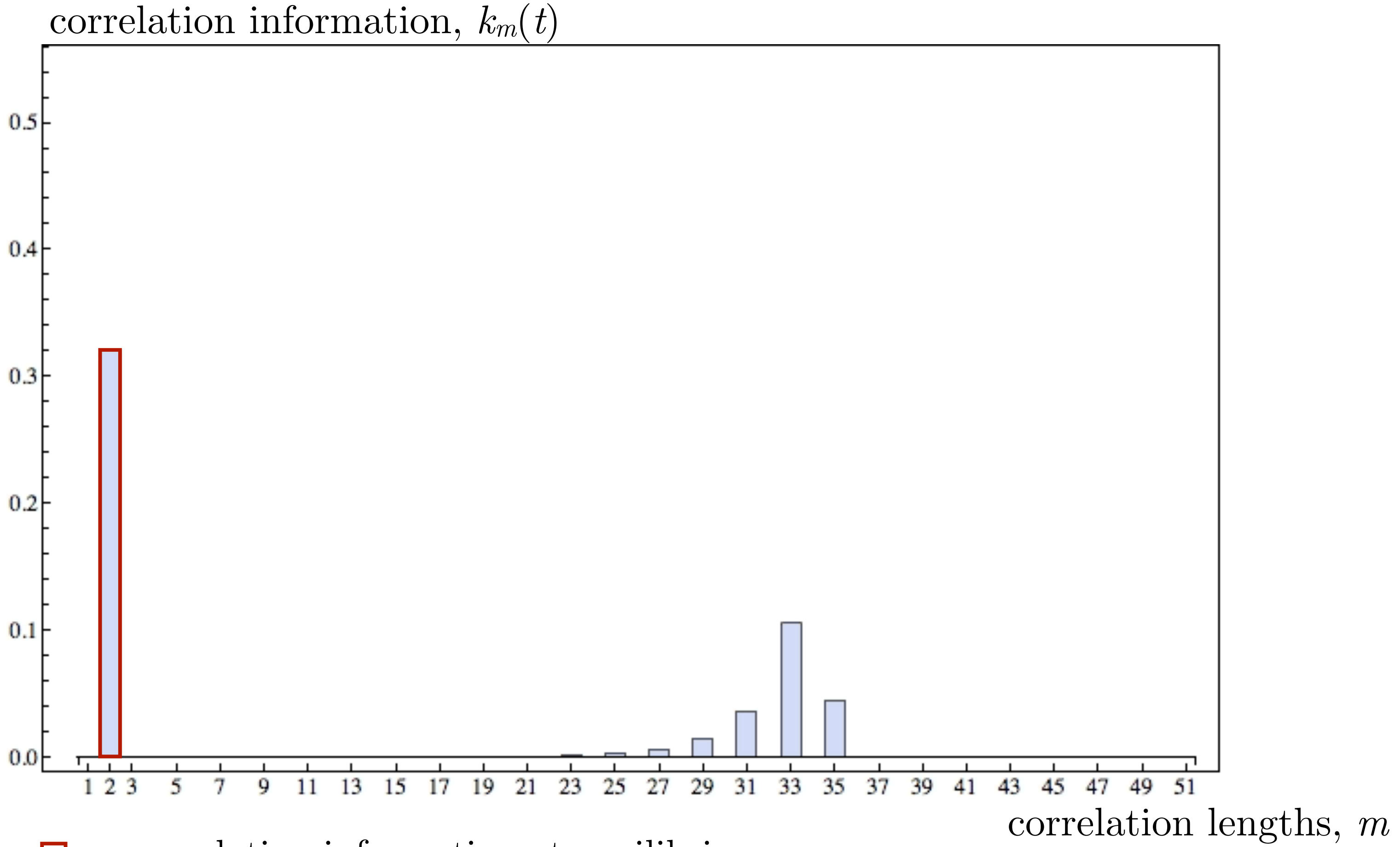
correlation lengths, m

□ — correlation information at equilibrium

Evolution of correlation information

Showing the correlation information k_m, including the density information k_1, up length $m = 51$, over time ($t = 0, 1, 2, ..., 60$).

Contributions to correlation information — ordered information



□ — correlation information at equilibrium