Multi-information in the thermodynamic limit

Ionas Erb* Nihat Ay†

July 3, 2003

Abstract

From information theory, mutual information is known to measure stochastic interdependence of probability distributions with two subsystems. We use a generalised version of this measure: multi-information, the Kullback-Leibler distance of a distribution from its corresponding independent distribution, and give a definition within the framework of statistical mechanics. There, the theory of infinite-volume Gibbs measures allows for the description of phase coexistence: The interaction potential of a model can yield several Gibbs measures at the same time. We propose to take the least multi-information of all the translation-invariant Gibbs measures to define a quantity directly depending on the interaction potential. We show that it is enough to take this infimum over the pure, i.e. physically relevant states only. Our definition is applied to the two-dimensional Ising model and the main result is derived: In the Ising square lattice, multi-information as a function of temperature attains its isolated global maximum at the point of phase transition. There, the one-sided derivatives diverge. Finally, we also briefly discuss the behaviour for the one-dimensional Ising chain in a magnetic field.

*Interdisciplinary Centre for Bioinformatics, Kreuzstr. 7, 04103 Leipzig, Germany
†Max-Planck Institute for Mathematics, Inselstr. 22-26, 04103 Leipzig, Germany

1 Introduction

Shannon's mutual information compares the summed entropies H of two distributions $p_{\{1\}}$, $p_{\{2\}}$ with the entropy of their joint distribution $p_{\{1,2\}}$:

$$I(p_{\{1,2\}}) = H(p_{\{1\}}) + H(p_{\{2\}}) - H(p_{\{1,2\}}). \tag{1}$$

If more than two subsystems exist, one can keep the two-point property of I and let the quantity depend on the distance between the elementary subsystems now [Li]. There are also multivariate generalisations. Let Λ be a finite set. Co-information is an alternating sum of entropies of the marginals p_V of all subsystems $V \subset \Lambda$ [Be]. Being hard to calculate, this quantity also has the drawback of becoming negative in certain cases. In [TSE], a quantity is used where the mutual informations of all bipartitions of Λ are summed up. Another multivariate generalisation is called multi-information [SV]:

$$I(p_\Lambda) = \sum_{i \in \Lambda} H(p_{\{i\}}) - H(p_\Lambda) \tag{2}$$

This quantity allows for a nice information-geometric interpretation: It is the Kullback-Leibler distance [CT] of p_Λ from its factorized distribution $\otimes_{i\in\Lambda} p_{\{i\}}$ [Am, Ay1].

All these information-theoretic measures quantify stochastic interdependence in probability distributions. They are used for a variety of purposes: statements about information transmission in noisy channels [Sh], ICA (independent component analysis) in multivariate statistics [HKO], measuring dependencies amongst cells of a neural network [Be], measuring brain complexity [TSE], deriving learning rules for neural networks [L, Ay2], to mention only a few of the possible applications.

To give us an idea of the behaviour of a quantity like mutual information, let us consider a simple example: two units x_1, x_2 which can take values from $\{0, 1\}$. If we know the probabilities $p_{\{1,2\}}(x_1, x_2)$ of the four configurations, mutual information is given by (1). Let us introduce an additional parameter β by which we can adjust a generic $p_{\{1,2\}}$ from the equidistribution to one of the Dirac measures. We define

$$p_\beta(x_1, x_2) := \frac{(p_{\{1,2\}}(x_1, x_2))^\beta}{\sum_{x_1', x_2' \in \{0,1\}} (p_{\{1,2\}}(x_1', x_2'))^\beta}. \tag{3}$$

The denominator normalizes p_β. Mutual information is now a function of the parameter β. For a generic choice of $p_{\{1,2\}}$, the function $\beta \mapsto I(p_\beta)$, let us call it $I(\beta)$, is shown in figure (1). The trajectory of the curve $\beta \mapsto p_\beta$ within the simplex of all probability measures for the four configurations is shown in figure (2). It ranges from the barycentre to one of the corners of the simplex. The Kullback-Leibler distance of this curve from the surface of independent distributions is mutual information.

Looking at both these figures, the question about the maximum of I springs to mind. To give us an idea what such a maximization [Ay1, AK] can mean, we want to define multi-information in the context of statistical mechanics (a field that has benefited greatly from information theory [J1]). There we have a mathematical formalism that describes models of different structural richness. A simple example for a finite-volume state (Gibbs measure) is

$$p(x_i, i \in \Lambda) = e^{F + \sum_{\emptyset \neq V \subset \Lambda} \Theta_V \prod_{i \in V} x_i}, \tag{4}$$

with $x_i \in \{0, 1\}$. F is a normalization constant (the free energy). The coefficients $\Theta_V \in \mathbb{R}, V \subset \Lambda$ represent the strength of direct interaction between the units i. The set of all terms $\Theta_V \prod_{i \in V} x_i$ in the exponent is called an interaction potential. A parametrized family of such potentials is called a model. The parameters (c.f. the β in (3)) can be inverse temperature, magnetic field etc.

There are (infinite-volume) potentials whose infinite-volume state is not uniquely determined. Models showing this phenomenon for certain (critical) parameter values are said to exhibit phase coexistence. There are hints in the literature that measures of stochastic interdependence are maximized at critical parameter values [MKNYM, Ar], or at phase transitions in a less strict sense [CY, LF, GL]. One can go a step further and look at the structural phenomena occuring at the phase coexistence point in

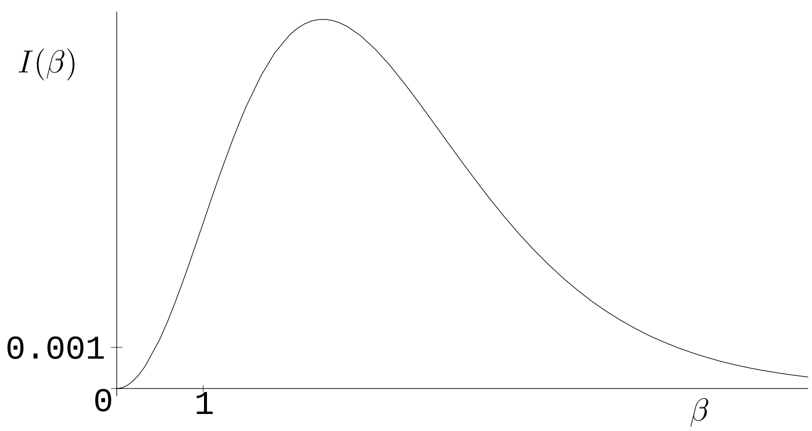

Figure 1: Plot of $I(\beta)$ for $p_1(0,0) = 0.1, p_1(0,1) = 0.2, p_1(1,0) = 0.3, p_1(1,1) = 0.4$. For $b = 0$ (p_β is the equidistribution) and $\beta \to \infty$ (p_β is Dirac measure) $I(\beta)$ vanishes, i.e. there are no stochastic dependencies for complete randomness and complete predicatability.

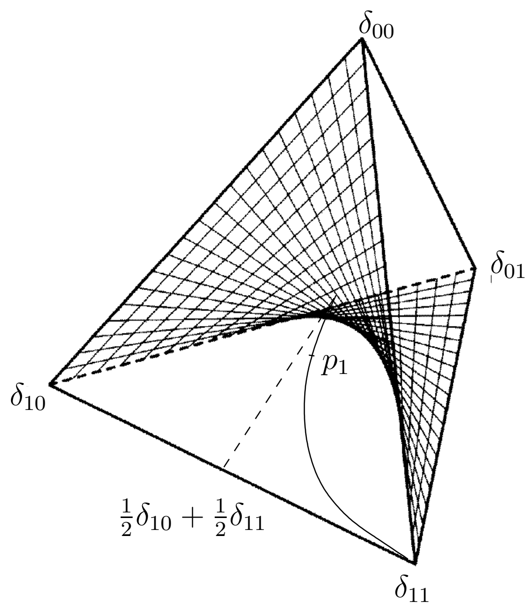

Figure 2: The set of probability distributions for the four configurations with the plane of factorizable p

standard models like the Ising square lattice: infinite-cluster formation, divergence of the correlation length. They can be seen as signs of "complex" behaviour. From this perspective it seems natural to assume that large stochastic interdependence is associated with high structural complexity. As stated already, mutual-information based quantities are also used as complexity measures, see also [FC].

Phase transitions thus seem to mark the "border of maximum complex behaviour" between complete randomness and absolute predictability. It is one of the objectives of the present work to give an example where this kind of statement can be made rigorous. To do so, we generalise multi-information to a quantity in the thermodynamic limit that takes into account a mathematical description of phase transitions, namely the non-uniqueness of infinite-volume Gibbs measures for certain interaction potentials.

The first part of this work will give a general definition of multi-information in the context of statistical mechanics. We propose a definition where the quantity depends directly on the interaction potential. In the second part, for the two-dimensional Ising model it is shown analytically that multi-information is maximized at the critical point.

2 Multi-information in statistical mechanics

2.1 Notation

Our systems take discrete values on the points of an infinite lattice. Let S be a finite set (the spin space), and let $x : \mathbb{Z}^d \mapsto S, i \mapsto x_i$ be configurations on the d-dimensional lattice of integers $\mathbb{Z}$. To make it a measurable space, the space of configurations $\Omega := S^{(\mathbb{Z}^d)}$ is equipped with the product sigma algebra $\mathcal{F}$, which contains the cylinder sets $\{x \in \Omega : X_\Lambda(x) = x_\Lambda\}$, where $x_\Lambda := (x_i)_{i\in\Lambda}$ is a configuration on the finite[1] set $\Lambda \subset\subset \mathbb{Z}^d$ and $X_\Lambda : \Omega \mapsto S^\Lambda, x \mapsto x_\Lambda$ the natural projection onto a finite configuration. Thus the projection X_Λ yields finite measurable spaces $(\Omega_\Lambda, \mathcal{F}_\Lambda)$, where $\Omega_\Lambda := S^\Lambda$ denotes the set of x_Λ and $\mathcal{F}_\Lambda$ its power set.

We will first define multi-information on these finite spaces, for p_Λ a probability measure on $(\Omega_\Lambda, \mathcal{F}_\Lambda)$. At this point, the form of the measure is of no importance.

Definition 2.1: Let p_Λ be a probability measure on $(\Omega_\Lambda, \mathcal{F}_\Lambda)$, where Λ is a finite set. The *multi-information* of p_Λ is defined by

$$I(p_\Lambda) := \sum_{i\in\Lambda} H(p_{\{i\}}) - H(p_\Lambda). \tag{5}$$

Here, $H := -\sum_{x_\Lambda \in \Omega_\Lambda} p_\Lambda(x_\Lambda) \ln p_\Lambda(x_\Lambda)$ denotes the Shannon entropy and $p_{\{i\}}(x_i) := \sum_{x_{\Lambda\setminus\{i\}}} p(x_{\Lambda\setminus\{i\}}, x_i)$ are the marginal distributions of the elementary subsystems in Λ. ♦

2.2 Thermodynamic limit

To define multi-information for distributions on the infinite measurable space $(\Omega, \mathcal{F})$, our starting point are measures p_Λ on finite spaces $(\Omega_\Lambda, \mathcal{F}_\Lambda)$, $\Lambda \subset\subset \mathbb{Z}^d$. These we consider as being obtained from a translation invari-

[1] We denote finiteness of subsets by $\subset\subset$.

ant measure p on $(\Omega, \mathcal{F})$ by defining its marginal distributions $p_\Lambda(x_\Lambda) := p(X_\Lambda = x_\Lambda)$. Translation invariance of p is defined by

$$p(\{(x_{i+j})_{j\in\mathbb{Z}^d} | (x_j)_{j\in\mathbb{Z}^d} \in A\}) = p(A) \quad \forall A \in \mathcal{F},\ \forall i \in \mathbb{Z}^d. \tag{6}$$

Existence and properties of the van-Hove limit [R] of multi-information follow in straightforward fashion from well-known results for entropy (see the appendix for a proof). Notice that the set of translation invariant measures is a simplex [S, Ge].

Theorem and Definition 2.2: *Let p be a translation invariant probability measure on $(\Omega, \mathcal{F})$. Then the van-Hove limit $\lim_{\Lambda \nearrow \mathbb{Z}^d} \frac{1}{|\Lambda|} I(p_\Lambda) =: I(p)$ exists and $I(p) \in [0, \ln |S|]$. The function $p \mapsto I(p)$ is concave and lower-semicontinuous (w.r.t. the weak* topology).*

The quantity $I(p)$ depends on the *state* of a system. In statistical mechanics, however, models are defined via the *interaction* between their constituents (spins, particles). In the following, we want to obtain a definition which directly depends on the interaction potential.

2.3 Phase coexistence

The construction of measures in infinite volume [Do] [LR] can yield non-uniqueness for a given interaction, so the description of phase coexistence becomes possible. For an interaction-dependent definition of multi-information we have to choose from a set of possible measures now. To introduce the necessary notation and to make our point clear, we give a brief description of the standard construction of infinite-volume Gibbs measures. All the results stated in this paragraph can be found in this or a similar form in [Ge, S], for short descriptions of the subject see also [Pe, Gr].

From finite-volume statistical mechanics one knows the form of the conditional probabilities for a finite configuration given an exterior configuration that the measure p on $(\Omega, \mathcal{F})$ should have [Wi]. Specifying these, one obtains a condition that possible infinite-volume Gibbs measures should fulfill. For this, we need to define interaction potentials in infinite volume.

Definition 2.3: A *potential* Φ on $\mathbb{Z}^d$ is a family of functions $\{\Phi_V\}_{V \subset\subset \mathbb{Z}^d}$ from Ω to $\mathbb{R}$ with

(i) Φ_V is X_V-measurable for all $V \subset\subset \mathbb{Z}^d$

(ii) The series $E_\Lambda^\Phi(x_\Lambda, y_{\Lambda^c}) := \sum_{\substack{V \subset\subset \mathbb{Z}^d: \\ V \cap \Lambda \neq \emptyset}} \Phi_V(x_\Lambda, y_{\Lambda^c})$ converges for all $\Lambda \subset\subset \mathbb{Z}^d$ and for all $(x_\Lambda, y_{\Lambda^c}) := x \in \Omega$ (where Λ^c denotes the complement of Λ in $\mathbb{Z}^d$.).

$E_\Lambda^\Phi(x_\Lambda, y_{\Lambda^c})$ is the *energy* of x_Λ with boundary condition y_{Λ^c}. ♦

This definition enables us to specify the Gibbsian conditional probabilities for the desired measures. Since we no nothing about the existence of these measures, we can only fix probability kernels (i.e. loosely speaking, conditional probabilities "waiting for a measure"). Let $\Omega_{\Lambda^c} = S^{\mathbb{Z}^d \setminus \Lambda}$. Using a definition from [Ge], a *specification* for finite S is given by a family $\{k_\Lambda^\Phi\}_{\Lambda \subset\subset \mathbb{Z}^d}$ of probability kernels from $(\Omega_{\Lambda^c}, \mathcal{F}_{\Lambda^c})$ to $(\Omega, \mathcal{F})$ where

$$A \mapsto k_\Lambda^\Phi(A | y_{\Lambda^c}) := \sum_{\substack{x_\Lambda: \\ (x_\Lambda, y_{\Lambda^c}) \in \mathcal{A}}} \frac{e^{-E_\Lambda^\Phi(x_\Lambda, y_{\Lambda^c})}}{\sum_{x'_\Lambda \in \Omega_\Lambda} e^{-E_\Lambda^\Phi(x'_\Lambda, y_{\Lambda^c})}}. \tag{7}$$

Here, Φ is a potential, $\Lambda \subset\subset \mathbb{Z}^d$, $A \in \mathcal{F}$ and $y_{\Lambda^c} \in \Omega_{\Lambda^c}$. Such specifications fulfill consistency conditions analogous to those of conditional probabilities. The set of DLR measures is now defined as the solution set of $p(A|\mathcal{F}_{\Lambda^c}) = k_\Lambda^\Phi(A|\cdot)$ p-a.s. for all finite volumes Λ and events A. Here, $p(A|\mathcal{F}_{\Lambda^c})$ is the conditional expectation of 1_A, i.e. of the indicator function for an event A, given the sigma algebra of events outside Λ. For the definition of conditional expectations given a sub-sigma algebra, see e.g. [Ba]. The properties of the set of DLR measures are well known.

Proposition and Definition 2.4: *Given a potential Φ, the set of infinite-volume Gibbs states (DLR measures) is given by*

$$\mathcal{G}(\Phi) = \left\{ p \text{ on } (\Omega, \mathcal{F}) : p(A|\mathcal{F}_{\Lambda^c}) = k_\Lambda^\Phi(A|\cdot) \;\; p\text{-}a.s.\ \forall A \in \mathcal{F}, \Lambda \subset\subset \mathbb{Z}^d \right\}.$$

$\mathcal{G}(\Phi)$ is a compact, convex set (more precisely: a simplex). Depending on the potential Φ, there are the following possibilities for its cardinality:

$$|\mathcal{G}(\Phi)| = 0, \tag{8}$$

$$|\mathcal{G}(\Phi)| = 1, \tag{9}$$

$$|\mathcal{G}(\Phi)| = \infty. \tag{10}$$

The set of Gibbs measures is always non-empty if the potential is translation invariant, i.e. if it fulfills

$$\Phi_{V+i}((x_{i-j})_{j\in\mathbb{Z}^d}) = \Phi_V(x) \quad \forall x \in \Omega,\ \forall V \subset\subset \mathbb{Z}^d,\ \forall i \in \mathbb{Z}^d. \tag{11}$$

Notice that even translation-invariant potentials need not have only translation-invariant states. Since the thermodynamic limit of multi-information was obtained for translation-invariant states, we will actually need the set of translation-invariant Gibbs measures $\mathcal{G}_I(\Phi)$, i.e. the intersection of all translation-invariant measures on $(\Omega, \mathcal{F})$ with the set of Gibbs measures of translation invariant potentials. The set $\mathcal{G}_I(\Phi)$ is also compact and convex and its cardinality can be 1 or infinity.

2.4 Multi-information of a potential

We are now in the position to define multi-information as a function of the interaction potential of a statistical-mechanics model. Our aim is to extract the minimum stochastic complexity of a model, so we define

Definition 2.5: Multi-information given a translation invariant potential Φ is defined by

$$I(\Phi) := \inf_{p\in\mathcal{G}_I(\Phi)} I(p), \tag{12}$$

where $I(p)$ is given by proposition and definition 2.2. ♦

Remark 2.6: Because of lower-semicontinuity of $I(p)$ (Theorem 2.2) and compactness of $\mathcal{G}_I$ it follows that the infimum is indeed attained. This follows from general statements about extrema of semicontinuous functions over compact sets, c.f. Theorem 25.9 in [Ch]. ★

The non-uniqueness expressed by (10) is called *phase coexistence.* Phases are the extreme points of the simplex $\mathcal{G}(\Phi)$, which are also just the physically realised states[2]. These so-called pure states have fluctuation-free macroscopic quantities. On the other hand, we can construct convex combinations of them, which do not stand for physically realised states but which express our uncertainty about the state we are in [Ge]. That is why the following proposition helps motivating our choice of defining $I(\Phi)$.

Theorem 2.7: *Let* $\mathrm{ex}(\mathcal{G}_I)$ *be the set of extreme points of* $\mathcal{G}_I$*. We have*

$$I(\Phi) = \inf_{p \in \mathcal{G}_I(\Phi)} I(p) = \inf_{p \in \mathrm{ex}(\mathcal{G}_I)(\Phi)} I(p). \tag{13}$$

So to take the infimum in definition 2.5 is not only justified by extracting the least model complexity but also by the fact that the infimum is attained in a physically relevant state. To illustrate definition 2.5 and proposition 2.7, figure 3 shows $I(p)$ over the set of infinite-volume Gibbs measures in the case of the two-dimensional Ising model. This example will also be the topic of the next section.

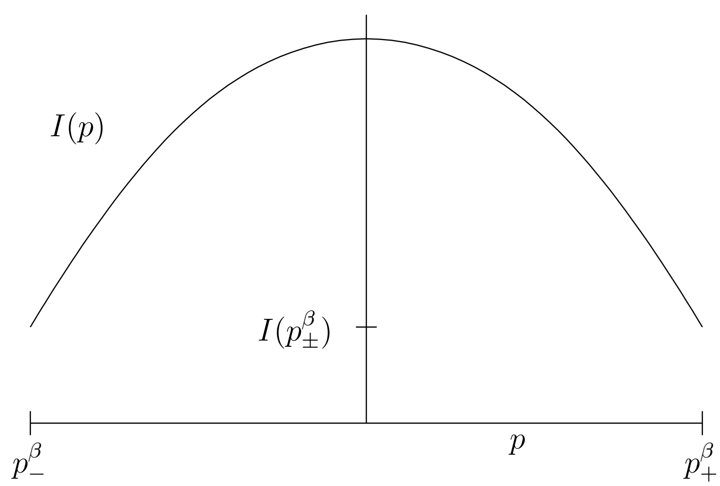

Figure 3: Schematic view of multi-information depending on p in the $2d$ Ising model

[2]Also, the property of ergodicity is equivalent with being an extreme point of the simplex of translation-invariant probability measures.

3 Ising square lattice

3.1 Multi-information for the model

Usually it is hard to calculate complexity measures for non-trivial systems. Here, we take advantage of the wealth of exact results for the standard example of statistical mechanics, the two-dimensional Ising model. Definition 2.5 can now be applied to the Ising potential, we have

$$\Phi_V^\beta(x) = -\beta x_i x_j \qquad \text{if } V = \{i,j\} \subset \mathbb{Z}^2 \text{ where } |i-j| = 1, \tag{14}$$

and $\Phi_V^\beta(x) = 0$ for all other sets V, the spin space $S = \{\pm 1\} \ni x_i$ and $\beta \in \mathbb{R}^+$. The parameter β is the inverse temperature and stands for the strength of interaction between spins. We use the results existing for this model, i.e. the explicit expressions for free energy and magnetization, critical temperature and the known set of Gibbs measures, for a list of references see [Ge].

Let us first present a visualization of the main result of this paper: A plot of multi-information of the potential (14) as a function of inverse temperature (see figure 4). What one can see is a sharp isolated global maximum at the point of phase transition. The analytic result will be given in the next section.

To derive a formula for multi-information, we take advantage of the existing exact results, which we will cite in the following. It is well known that below a critical temperature the set of infinite-volume Gibbs measures is the convex hull of two extreme probability measures:

$$\mathcal{G}(\Phi^\beta) = \{tp_-^\beta + (1-t)p_+^\beta : t \in [0,1]\} \tag{15}$$

where the two extreme points $p_\pm^\beta$ are connected by a spin-flip symmetry that can be written as

$$p_+^\beta\,(X_\Lambda = x_\Lambda) = p_-^\beta\,(X_\Lambda = -x_\Lambda) \quad \forall \Lambda \subset\subset \mathbb{Z}^d. \tag{16}$$

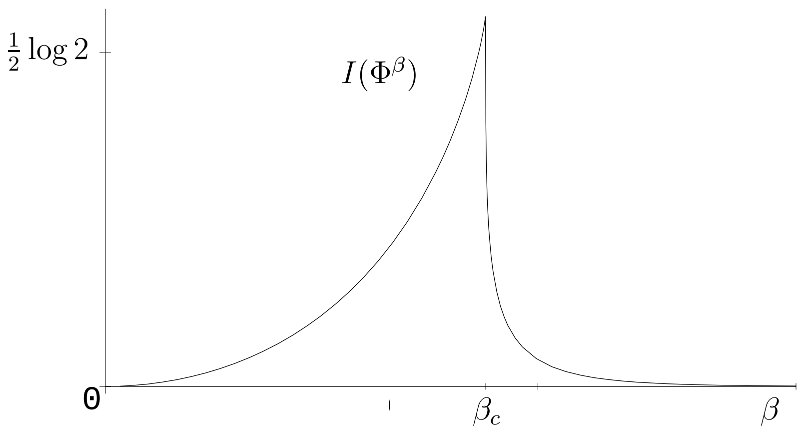

Figure 4: Multi-information of the Ising square lattice

Moreover, for the single-spin expectations (the magnetization) we have $p_-^\beta(X_0) = -p_+^\beta(X_0)$. It is essential that these *order parameters* are non-zero for $\beta > \beta_c$. The Yang formula (a rigorous result, see [S], p. 153) is given by

$$m_\beta := p_+^\beta(X_0) = \begin{cases} (1 - \sinh^{-4} 2\beta)^{\frac{1}{8}} & \text{if } \beta > \beta_c, \\ 0 & \text{otherwise.} \end{cases} \quad (17)$$

The essential feature of the model is a continuous phase transition at a critical temperature β_c:

$$\text{sinh} 2\beta_c = 1, \qquad \text{i.e.} \qquad \beta_c = \frac{1}{2}\ln(1+\sqrt{2}). \quad (18)$$

We will also need the entropy (per unit volume)

$$h(\beta) := h(p_\pm^\beta) = \ln(\sqrt{2}\ \cosh 2\beta) + \frac{1}{\pi}\int_0^{\frac{\pi}{2}} \ln\left\{1 + \sqrt{1-\kappa_\beta^2 \sin^2\omega}\right\} d\omega$$
$$- 2\beta\ \tanh 2\beta - \beta\frac{\sinh^2 2\beta - 1}{\sinh 2\beta \cosh 2\beta}\left[\frac{2}{\pi}\int_0^{\frac{\pi}{2}} \frac{d\omega}{\sqrt{1-\kappa_\beta^2 \sin^2\omega}} - 1\right], \quad (19)$$

where

$$\kappa_\beta = \frac{2\sinh 2\beta}{\cosh^2 2\beta}. \qquad (20)$$

This expression for the entropy can be found using the results for free energy $f(\beta)$ and energy $e(\beta)$, see e.g. [Wa], because of

$$h(\beta) = \beta e(\beta) - \beta f(\beta). \qquad (21)$$

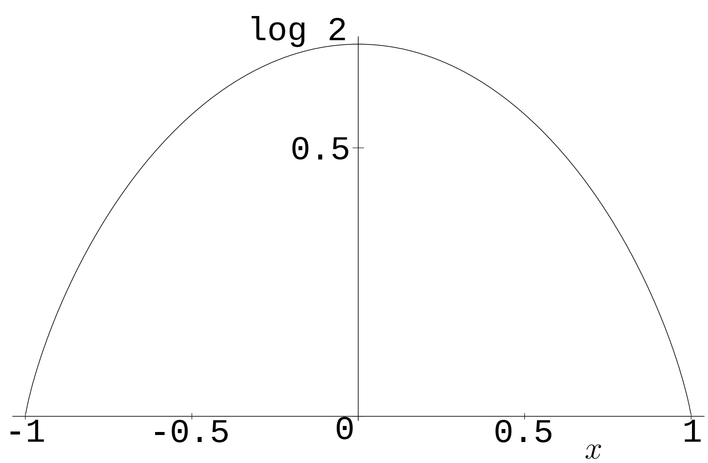

Figure 5: The function $s(x)$

Theorem 3.1: *Let m_β and $h(\beta)$ be defined by (17) and (19). Also, let*

$$s(x) = -\frac{1+x}{2}\ln\frac{1+x}{2} - \frac{1-x}{2}\ln\frac{1-x}{2}, \qquad x \in [-1,1], \qquad (22)$$

(see figure 5)[3]*. Multi-information of the Ising square lattice is given by*

$$I(\Phi^\beta) = s(m_\beta) - h(\beta). \qquad (23)$$

[3] $0\ln 0 := 0$.

Remark 3.2: Notice that similar expressions can be found for all translation-invariant models with binary spin space. ★

3.2 The maximum of multi-information

Putting some effort into bounding the terms in (23), one can obtain analytic results connecting the phase transition with maximum multi-information:

Theorem 3.3: *In the two-dimensional Ising model, multi-information as a function $\beta \mapsto I(\Phi^\beta)$ of inverse temperature attains its isolated global maximum at the point of phase transition $\beta = \beta_c$. At this point, the left-sided derivative goes to $+\infty$, the right-sided one to $-\infty$.*

The rest of this chapter is devoted to the proof of the above theorem. Two technical lemmas are needed. Using the shorthand notation

$$\Theta(\beta) := \frac{\sinh^2 2\beta - 1}{\sinh 2\beta \cosh 2\beta} \left[\frac{2}{\pi} \int_0^{\frac{\pi}{2}} \frac{d\omega}{\sqrt{1 - \kappa_\beta^2 \sin^2 \omega}} - 1 \right], \tag{24}$$

we have the following bounds:

Lemma 3.4: *Let $\beta \geq \beta_c$. Then*

$$\beta\Theta(\beta) \leq \min \left\{ \frac{\sinh 2\beta - 1}{2} \ln \frac{\sinh 2\beta + 1}{\sinh 2\beta - 1}, \frac{\beta}{\sinh 2\beta \cosh 2\beta} \right\}. \tag{25}$$

Moreover, $\Theta(\beta_c) = 0$.

$$\begin{aligned} & - \ln \left(\sqrt{2} \cosh 2\beta \right) + 2\beta \tanh 2\beta \\ & \qquad \leq \min \left\{ 2\beta_c(\beta - \beta_c) + \sqrt{2}\beta_c - \ln 2, \frac{-\beta}{\sinh 2\beta \cosh 2\beta} + \ln \sqrt{2} \right\}, \end{aligned} \tag{26}$$

$$s(m_\beta) \leq \ln 2 - \frac{(1 - \sinh^{-4} 2\beta)^{\frac{1}{4}}}{2} - \frac{(1 - \sinh^{-4} 2\beta)^{\frac{1}{2}}}{12}. \tag{27}$$

Lemma 3.5: *For $0 \leq y \leq 1/2$ we have*

$$\frac{(1-(1+y^2)^{-4})^{\frac{1}{4}}}{2} > \frac{y^2}{2}\ln\frac{2+y^2}{y^2}. \tag{28}$$

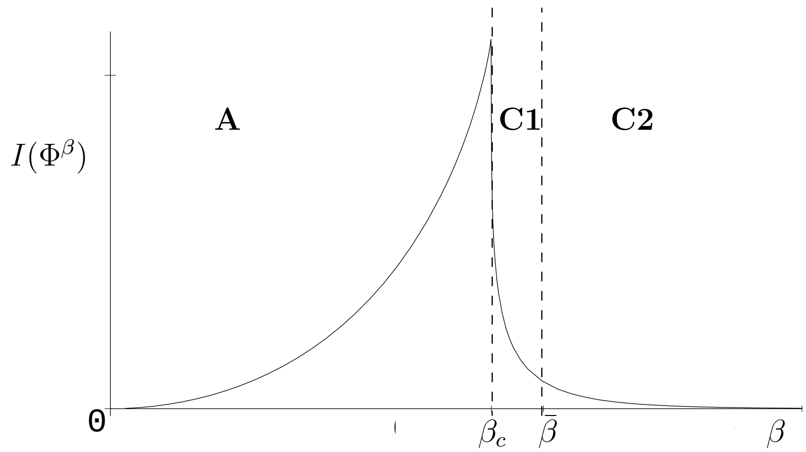

Figure 6: Partitioning the β domain for the proof of Theorem 3.3.

Proof of Theorem 3.3: Multi-information is considered in four different regimens (cf. fig. 6). Let us start with the high-temperature case:

(A) $\beta \leq \beta_c$: We consider the monotonicity of $I(\Phi^\beta)$. Here, the order parameter m_β vanishes. Thus, as an immediate consequence of Theorem 3.1 we obtain

$$I(\Phi^\beta) = \ln 2 - h(\beta), \qquad \beta \leq \beta_c. \tag{29}$$

Using (21), the β derivative is

$$\begin{aligned}\frac{dI(\Phi^\beta)}{d\beta} = -\frac{dh(\beta)}{d\beta} = \frac{d}{d\beta}[\beta f(\beta) - \beta e(\beta)] \\ = \frac{d(\beta f(\beta))}{d\beta} - e(\beta) - \beta\frac{de(\beta)}{d\beta} = -\beta\frac{de(\beta)}{d\beta}.\end{aligned} \tag{30}$$

We obtained the last equality because of

$$\frac{d(\beta f(\beta))}{d\beta} = e(\beta). \tag{31}$$

(For this sort of equations, see e.g. [R], p. 56.) Applying (31) to the left-over term in (30) gives

$$\frac{dI(\Phi^\beta)}{d\beta} = -\beta \frac{d^2(\beta f(\beta))}{d\beta^2} \geq 0. \tag{32}$$

This relation follows from the convexity of $-\beta f(\beta)$, see again [R], p. 54. Hence, the monotonicity up to the critical point is known.

(B) $\beta = \beta_c$: Let us now consider the non-analyticity at the critical temperature. Below β_c (for the left-sided derivative) we use (30). The divergence of the specific heat is known from the literature, we use

$$\frac{de(\beta)}{d\beta} = \frac{8}{\pi} \ln|\beta - \beta_c| + \text{bounded terms}, \tag{33}$$

see [S] (p. 152). This expression goes to $-\infty$ as $\beta \nearrow \beta_c$. Thus, the left-sided derivative of $I(\Phi^\beta)$ goes to $+\infty$.

Above β_c, the derivative of the first term in (23) comes into play. We have

$$\frac{ds(m_\beta)}{d\beta} = \frac{\sinh^{-4} 2\beta}{\tanh 2\beta} m_\beta^{-7} \frac{1}{2} \ln \frac{1 - m_\beta}{1 + m_\beta}. \tag{34}$$

Expanding the logarithm into a series (see e.g. [BS]), we obtain the following bound:

$$m_\beta^{-1} \frac{1}{2} \ln \frac{1 - m_\beta}{1 + m_\beta} = -1 - \sum_{n=1}^{\infty} \frac{m_\beta^{2n}}{2n+1} \leq -1. \tag{35}$$

Hence,

$$\frac{ds(m_\beta)}{d\beta} \leq -\frac{\sinh^{-4} 2\beta}{\tanh 2\beta} m_\beta^{-6}. \tag{36}$$

In order to write (33) such that it depends on m_β, we use

$$(1 - \sinh^{-4} 2\beta) \leq 8\sqrt{2}(\beta - \beta_c), \tag{37}$$

which follows from the concavity of the left-hand side (which thus remains underneath its tangent through β_c). Together with the formula (17) for m_β we have

$$\ln(\beta - \beta_c) \geq 8 \ln m_\beta - \ln(8\sqrt{2}). \tag{38}$$

Put all this together for

$$\lim_{m_\beta \searrow 0} \frac{dI(\Phi^\beta)}{d\beta} \leq \lim_{m_\beta \searrow 0} \left[-\frac{\sinh^{-4} 2\beta}{\tanh 2\beta} m_\beta^{-6} - \frac{8}{\pi} 8 \ln m_\beta + \text{bounded terms} \right]$$

$$= \lim_{y \to \infty} y^6 \left[-\sqrt{2} + \frac{64 \ln y}{\pi y^6} + \frac{\text{b. T.}}{y^6} \right] = -\infty, \quad (39)$$

where we used the substitution $y := 1/m_\beta$ and $\sinh^{-4} 2\beta_c / \tanh 2\beta_c = \sqrt{2}$.

(C) $\beta_c < \beta$: For the remaining β domain we only show

$$I(\Phi^\beta) < I(\Phi^{\beta_c}), \qquad \beta > \beta_c. \quad (40)$$

Together with Theorem 3.1 this becomes

$$s(m_\beta) - h(\beta) < \ln 2 - h(\beta_c), \qquad \beta > \beta_c. \quad (41)$$

With (18), (19), the entropy at β_c is found to be

$$h(\beta_c) = \ln 2 - \sqrt{2}\beta_c + \frac{1}{\pi} \int_0^{\frac{\pi}{2}} \ln \left[1 + \cos\omega\right] d\omega. \quad (42)$$

(We used $\Theta(\beta_c) = 0$ from Lemma 3.4 and $\cosh 2\beta_c = \sqrt{2}$.) The relation to be shown, (41), thus becomes

$$s(m_\beta) - \ln(\sqrt{2}\cosh 2\beta) + \frac{1}{\pi} \int_0^{\frac{\pi}{2}} \ln \frac{1 + \cos\omega}{1 + \sqrt{1 - \kappa_\beta^2 \sin^2 \omega}} d\omega$$

$$+ 2\beta \tanh 2\beta + \beta\Theta(\beta) - \sqrt{2}\beta_c < 0, \qquad \beta > \beta_c. \quad (43)$$

We now start with the observation that the integral in (43) is smaller or equal to zero since $\cos\omega \leq \sqrt{\ldots}$ and thus the argument of the logarithm is smaller or equal to 1. It thus suffices to no longer consider this term in the following. For an additional partitioning of the domain above β_c we use

$$\bar{\beta} := \frac{1}{2} \operatorname{arsinh} (1 - K^4)^{-\frac{1}{4}}, \qquad K := 2(\sqrt{2}\beta_c - \frac{3}{2} \ln 2). \quad (44)$$

(C1) $\beta_c < \beta \leq \bar{\beta}$: If we feed the corresponding terms from Lemma 3.4 into (43), we obtain the following inequality whose proof suffices to prove (43):

$$-\frac{(1 - \sinh^{-4} 2\beta)^{\frac{1}{4}}}{2} - \frac{(1 - \sinh^{-4} 2\beta)^{\frac{1}{2}}}{12}$$

$$+ 2\beta_c(\beta - \beta_c) + \frac{\sinh 2\beta - 1}{2} \log \frac{\sinh 2\beta + 1}{\sinh 2\beta - 1} < 0 \quad (45)$$

We now show that the sum of the first and last terms of the LHS, as well as the sum of the two terms in-between them are negative in the required range β_c up to $\bar{\beta}$. For first and last term we need Lemma 3.5. Also using

$$\sinh 2\beta =: 1 + y^2 \tag{46}$$

as well as

$$\sinh 2\bar{\beta} < 1 + (1/2)^2 \tag{47}$$

(which can be checked using (44), $\sinh 2\bar{\beta} \approx 1.18$) it is clear that the sum of first and last terms of the LHS of (45) are smaller zero in the required range. For the middle terms, by squaring the corresponding inequality we obtain

$$4\beta_c^2(\beta - \beta_c)^2 < \frac{1 - \sinh^{-4} 2\beta}{144}, \qquad \beta_c < \beta \leq \bar{\beta}. \tag{48}$$

Here, we do the following: At $\beta = \beta_c$, both sides are equal to zero. Taking the second derivatives shows that the LHS is a concave function, the RHS a convex one. If the inequality holds for the point $\bar{\beta}$, it also holds for the entire interval $(\beta_c, \bar{\beta}]$. Using (44) one calculates for $\beta = \bar{\beta}$

$$4\beta_c^2(\bar{\beta} - \beta_c)^2 < \frac{K^4}{144}, \tag{49}$$

from which we obtain (taking the square root, shifting terms and applying the hyperbolic sine)

$$\sinh 2\bar{\beta} < \sinh \left[\frac{K^2}{12\beta_c} + 2\beta_c\right] = 1.19471... \tag{50}$$

Putting in the value of $\sinh \bar{\beta}$ using (44) shows that this relation indeed holds. Hence (48) holds in the required β range including $\bar{\beta}$, and thus (43) holds.

(C2) $\bar{\beta} < \beta$: Lemma 3.4 again makes (43) an inequality whose proof suffices to prove (43):

$$\frac{3}{2}\ln 2 - \sqrt{2}\beta_c - \frac{(1 - \sinh^{-4} 2\beta)^{\frac{1}{4}}}{2} < 0, \qquad \bar{\beta} < \beta. \tag{51}$$

the corresponding equality is just solved by $\bar{\beta}$, cf. (44). As the LHS is monotonically decreasing, the inequality holds above $\bar{\beta}$. ■

4 Discussion; The Ising chain in a field

What does the maximization of multi-information mean? In our example we can view multi-information as the difference of two terms: The information about the single spin that stems from the whole system, $\ln 2 - h(p)$ (information per site) and the information about the single spin that comes from one single spin only, $\ln 2 - H_0(p_0)$. This information difference can be seen as the uncertainty about the internal mechanisms of the system in the sense that information about the subsystems does not imply complete information about the whole system, see figure 7.

Another measure for internal dependencies is $p(X_i X_j) - p(X_i)p(X_j)$, the covariance. In the Ising square lattice its behaviour is very similar to that of multi-information, see figure 8, see also [Li]. Similarly to entropy, the plotted correlation functions of nearest and next-nearest neighbours also diverge logarithmically at the citical temperature [MW]. So the left- and right-sided derivatives of covariance behave qualitatively the same way as the ones of multi-information.

Figure 4 fits nicely into the universal picture discussed in [Ar]. Clearly, the singularity there is due to the phase transition, which is also characterized by non-analyticity of the free energy. Let us now discuss an example where no phase transition takes place. There is a complete solution for the Ising chain in a magnetic field b. Our interaction potential depends on this additional parameter now. In analogy to (23) we can write down a formula for multi-information (for entropy and magnetization see e.g. [Pa]) and plot it for different values of b (figure 9). Without magnetic field the ground state is described by a unique Gibbs measure that gives equal weight to both Dirac measures (c.f. also the dotted line in figure 2). The magnetization stays zero through all temperatures and multi-information is completely determined by the entropy of the whole system. For $b > 0$ the symmetry is broken, magnetization is positive for finite β. Now there is competition between the order creating and destroying influences of field

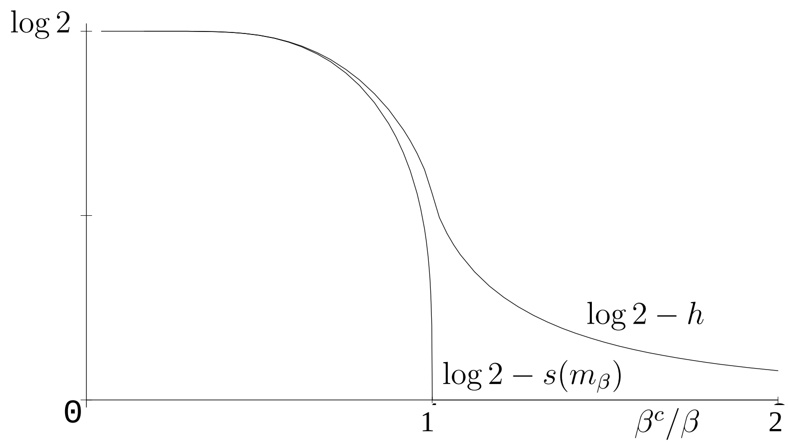

Figure 7: 2d Ising: $\ln 2 - h$ and $\ln 2 - s(m_\beta)$ depending on β_c/β. Their difference is multi-information.

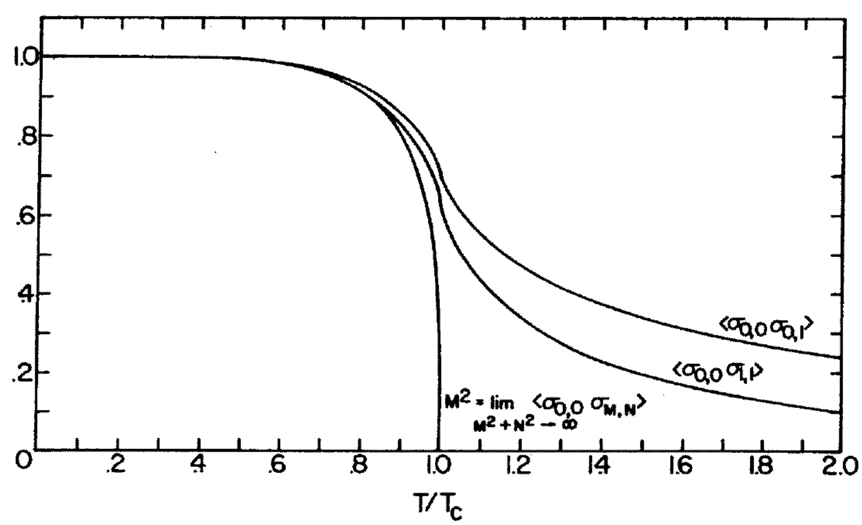

Figure 8: 2d Ising: correlation functions of near neighbours and m_β^2 (lower curve) depending on β_c/β. The difference between correlation and m_β^2 is covariance. (figure taken from [MW].)

and temperature. Interdependencies in the system are highest if these influences balance out.

Finally, let us consider the behaviour in the parameter b (figure 10). In the two-dimensional case a phase transition is known to take place at $b = 0$. Together with our ("numerically") observed maximum in the one-dimensional case, we may conjecture that the multi-information of the square lattice will show a maximum with a singularity at this point.

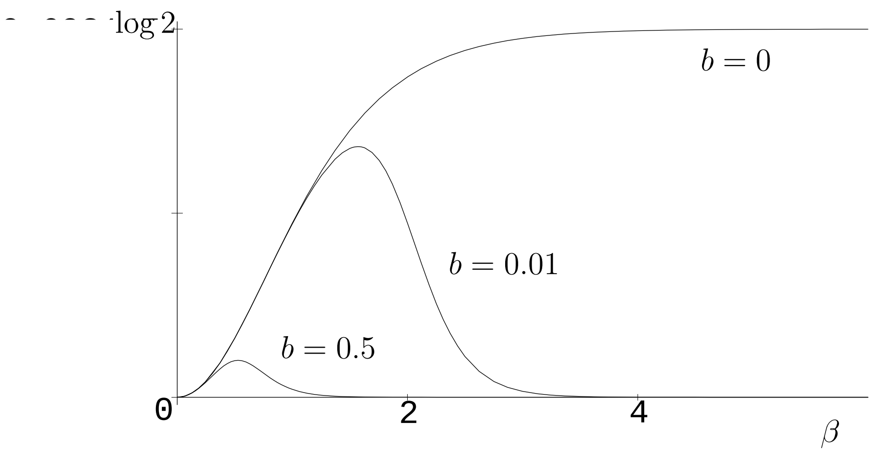

Figure 9: Graphs of multi-information in the Ising chain for different magnetic fields

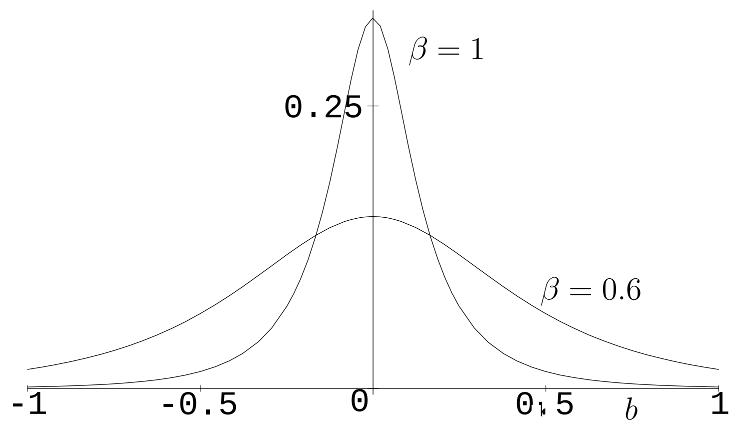

Figure 10: Multi-information of the Ising chain depending on the magnetic field for different temperatures

5 Proofs of lemmas and theorems

Let us first state a lemma which will be needed for the proof of Theorem 2.2:

Lemma 5.1: *Let p, q be probability measures on $(\Omega, \mathcal{F})$. We have*

(i) $0 \le H(p_{\{0\}}) \le \ln |S|$,

(ii) $H\left((tp + (1-t)q)_{\{0\}}\right) \ge tH(p_{\{0\}}) + (1-t)H(q_{\{0\}}) \quad \forall t \in [0,1]$,

(iii) $H(p_{\{0\}})$ *is continuous for the weak* topology.*

Proof: For (i), (ii) see [CT]. In our case the measures are marginals of p, q, but (ii) follows immediately from the affinity of the projection of p onto $p_{\{0\}}$, i.e. $(tp + (1-t)q)_{\{0\}} = tp_{\{0\}} + (1-t)q_{\{0\}}$.
(iii) follows from the continuity of entropy w.r.t. $p_{\{0\}}$, see [Ca]. It remains to show that the projection π_0 of p onto $p_{\{0\}}$ is continuous, i.e. that from

$p_n \to p$ follows

$$\pi_0(p_n) \to \pi_0(p). \tag{52}$$

Continuity for the weak* topology on our topological space Ω means that $p_n \to p$ is equivalent to $p_n(f) \to p(f) \forall f \in C(\Omega)$ (being the space of continuous functions for the product topology). For f we choose the indicator function $1_{\{X_0=x_0\}}$ (which is continuous since the inverse images of 1 and 0 are open sets). Now we have

$$\pi_0(p_n)(x_0) = p_n(X_0 = x_0) = p_n(1_{\{X_0=x_0\}}) \to p(1_{\{X_0=x_0\}}) = \pi_0(p)(x_0)\ \forall x_0 \in S.$$

Hence (52) holds. ■

Proof of Theorem 2.2: We use the existence of the van-Hove limit, upper-semicontinuity and affinity of the entropy $\lim_{\Lambda \nearrow \mathbb{Z}^d} \frac{1}{|\Lambda|} H(p_\Lambda) =: h(p) \in [0, \ln |S|]$ (cf. [Is], these properties follow immediately from the proof for a more generally defined entropy not requiring finite S) and Lemma 5.1. Similarly to $I(p_\Lambda)$, we can split $I(p)$ into marginal entropy and entropy:

$$\begin{aligned} I(p) = \lim_{\Lambda \nearrow \mathbb{Z}^d} \frac{1}{|\Lambda|} \left[\sum_{i \in \Lambda} H(p_{\{i\}}) - H(p_\Lambda) \right] &= \lim_{\Lambda \nearrow \mathbb{Z}^d} \frac{1}{|\Lambda|} \left[|\Lambda| H(p_{\{0\}}) - H(p_\Lambda) \right] \\ &= H(p_{\{0\}}) - \lim_{\Lambda \nearrow \mathbb{Z}^d} \frac{1}{|\Lambda|} H(p_\Lambda) = H(p_{\{0\}}) - h(p), \end{aligned} \tag{53}$$

where the second equality follows from translation invariance. Since we have all the desired properties for $h(p)$ and $H(p_{\{0\}})$, the theorem follows immediately. ■

Proof of Theorem 2.7: We have to show that the Infimum is always attained in an extreme point of $\mathcal{G}_I$. This follows from compactness and convexity of $\mathcal{G}_I$ as well as lower-semicontinuity and concavity of $I(p)$ according to Theorem 25.9 in vol. 2 of [Ch]. ■

Proof of Theorem 3.1: We show:

(i) $I(\Phi^\beta) = I(p^\beta_\pm)$,

(ii) $I(p^\beta_\pm) = s(m_\beta) - h(\beta)$.

Let us begin with (i): According to (15), the $p^\beta_\pm$ are the only extreme Gibbs measures. First we show that $I(p)$ is symmetric around $(p^\beta_- + p^\beta_+)/2$. For this we use the measures $p = (1-t)p_-^\beta + tp_+^\beta$ and $p' = tp_-^\beta + (1-t)p_+^\beta$ for $t \in [0,1]$. Because of the spin-flip symmetry (16), for $\Lambda \subset\subset \mathbb{Z}^d$ we have

$$\begin{aligned} H(p_\Lambda) &= -\sum_{x_\Lambda \in \Omega_\Lambda} [tp_-^\beta(x_\Lambda) + (1-t)p_+^\beta(x_\Lambda)] \ln [tp_-^\beta(x_\Lambda) + (1-t)p_+^\beta(x_\Lambda)] \\ &= -\sum_{x_\Lambda \in \Omega_\Lambda} [tp_+^\beta(-x_\Lambda) + (1-t)p_-^\beta(-x_\Lambda)] \ln [tp_+^\beta(-x_\Lambda) + (1-t)p_-^\beta(-x_\Lambda)] \\ &= -\sum_{x_\Lambda \in \Omega_\Lambda} [tp_+^\beta(x_\Lambda) + (1-t)p_-^\beta(x_\Lambda)] \ln [tp_+^\beta(x_\Lambda) + (1-t)p_-^\beta(x_\Lambda)] = H(p'_\Lambda). \end{aligned} \tag{54}$$

Taking the limit does not change the above argumentation, so $h(p) = h(p')$ and together with (53) we also have $I(p) = I(p')$. By theorem 2.7 $I(\Phi^\beta) = \inf_{p \in \{p_-^\beta, p_+^\beta\}} I(p)$, because of the above symmetry we have $I(p_-^\beta) = I(p_+^\beta)$ (see figure 3).

(ii): A general relation between expectation and probability of the single spin is

$$p(X_0) = \sum_{x_0 = \pm 1} p(X_0 = x_0)x_0 = p(X_0 = 1) - p(X_0 = -1). \tag{55}$$

Also using $\sum_{x_0=\pm 1} p(x_0) = 1$, we obtain for the single-spin probability

$$p(X_0 = x_0) = \frac{1 + x_0 p(X_0)}{2}. \tag{56}$$

Hence

$$H_0(p_\pm^\beta) = -\sum_{x_0=\pm 1} \frac{1 + x_0 p_\pm^\beta(X_0)}{2} \ln \frac{1 + x_0 p_\pm^\beta(X_0)}{2} = s\left(p_+^\beta(X_0)\right). \tag{57}$$

Since s is an even function, both expectation values lead to the same result. From (53) follows (ii). ■

Proof of Lemma 3.4:

Equation (25): $\Theta(\beta)$ is shorthand for

$$\Theta(\beta) := \frac{\sinh^2 2\beta - 1}{\sinh 2\beta \cosh 2\beta} \left[\frac{2}{\pi} \int_0^{\frac{\pi}{2}} \frac{d\omega}{\sqrt{1 - \kappa_\beta^2 \sin^2\omega}} - 1 \right]. \tag{58}$$

1.) $\beta\Theta(\beta) \le \frac{\sinh 2\beta - 1}{2} \log \frac{\sinh 2\beta + 1}{\sinh 2\beta - 1}$, $\Theta(\beta_c) = 0$

We start with the contents of the square brackets in (58). Taking the -1 into the integral, the resulting expression in the integral can be modified like this:

$$\frac{1-\sqrt{1-\kappa_\beta^2\sin^2\omega}}{\sqrt{1-\kappa_\beta^2\sin^2\omega}} \le \frac{1-\left(1-\kappa_\beta^2\sin^2\omega\right)}{\sqrt{1-\kappa_\beta^2\sin^2\omega}} \le \frac{\kappa_\beta^2\sin\omega}{\sqrt{1-\kappa_\beta^2\sin^2\omega}}. \tag{59}$$

We can integrate over this (cf. e.g. [BS], bestimmtes Integral #21). We have

$$\int_0^{\frac{\pi}{2}} \frac{\sin\omega}{\sqrt{1-\kappa^2\sin^2\omega}}d\omega = \frac{1}{2\kappa}\ln\frac{1+\kappa}{1-\kappa}, \qquad |\kappa|<1. \tag{60}$$

Also using the definition of κ_β, cf. (20), we obtain

$$\Theta(\beta) \le \frac{\sinh^2 2\beta - 1}{\sinh 2\beta\cosh 2\beta}\frac{2}{\pi}\left[\frac{\sinh 2\beta}{\cosh^2 2\beta}\ln\frac{1+\kappa_\beta}{1-\kappa_\beta}\right]. \tag{61}$$

Once again we use (20) for

$$\frac{1+\kappa_\beta}{1-\kappa_\beta} = \frac{\cosh^2 2\beta + 2\sinh 2\beta}{\cosh^2 2\beta - 2\sinh 2\beta} = \frac{1+\sinh^2 2\beta + 2\sinh 2\beta}{1+\sinh^2 2\beta - 2\sinh 2\beta} = \left(\frac{\sinh 2\beta + 1}{\sinh 2\beta - 1}\right)^2. \tag{62}$$

Together with (61) this gives

$$\Theta(\beta) \le \frac{\sinh^2 2\beta - 1}{\cosh^3 2\beta}\frac{4}{\pi}\ln\frac{\sinh 2\beta + 1}{\sinh 2\beta - 1}. \tag{63}$$

For the first statement of the lemma we still have to show that

$$\beta\frac{\sinh^2 2\beta - 1}{\cosh^3 2\beta}\frac{4}{\pi} \le \frac{\sinh 2\beta - 1}{2}. \tag{64}$$

Using $\sinh^2 2\beta - 1 = (\sinh 2\beta + 1)(\sinh 2\beta - 1)$, the inequality becomes

$$\frac{\sinh 2\beta + 1}{\cosh^3 2\beta} \le \frac{\pi}{8\beta}. \tag{65}$$

The LHS can be bounded as follows (note that for $\beta > \beta_c$ we have $\sinh 2\beta > 1$):

$$\frac{\sinh 2\beta + 1}{\cosh^3 2\beta} \le \frac{\sinh^2 2\beta + 1}{\cosh^3 2\beta} = \frac{\cosh^2 2\beta}{\cosh^3 2\beta} \le \frac{1}{1+(2\beta)^2/2}. \tag{66}$$

The last relation follows from the series expansion of the hyperbolic cosine. Whenever this last expression is smaller than the RHS of (65), the inequality holds. Thus we have to show

$$\frac{1}{1+(2\beta)^2/2} \le \frac{\pi}{8\beta} \tag{67}$$

or

$$0 \leq \beta^2 - \frac{4}{\pi}\beta + \frac{1}{2}, \tag{68}$$

which is fulfilled for $\beta = 0$. Since the corresponding equation has no real zeroes, the relation also holds for all the other β, and thus (64) is proven. We still have to show that $\Theta(\beta_c) = 0$. Clearly, $\Theta(\beta) \geq 0$ for $\beta \geq \beta_c$. In part (B1) of the proof of Theorem 3.3 we have moreover shown that $\beta\Theta(\beta) \leq (1 - \sinh^{-4} 2\beta)^{\frac{1}{4}}/2$, given the just proven first statement of the lemma. Hence we have

$$0 \leq \Theta(\beta_c) \leq \frac{(1 - \sinh^{-4} 2\beta_c)^{\frac{1}{4}}}{2\beta_c} = 0. \tag{69}$$

2.) $\Theta(\beta) \leq 1/\sinh 2\beta \cosh 2\beta$

We start with the observation that for $\beta \geq \beta_c$ the following equation holds:

$$\begin{aligned}\sqrt{1-\kappa_\beta^2} &= \sqrt{\frac{\cosh^4 2\beta - 4\sinh^2 2\beta}{\cosh^4 2\beta}} \\ &= \frac{\sqrt{\left(\sinh^2 2\beta + 1\right)^2 - 4\sinh^2 2\beta}}{\cosh^2 2\beta} = \frac{\sinh^2 2\beta - 1}{\cosh^2 2\beta}. \end{aligned} \tag{70}$$

Using this equation, (58) becomes

$$\Theta(\beta) = \coth 2\beta \sqrt{1-\kappa_\beta^2} \left[\frac{2}{\pi} \int_0^{\frac{\pi}{2}} \frac{d\omega}{\sqrt{1-\kappa_\beta^2 \sin^2\omega}} - 1 \right]. \tag{71}$$

We shift the root into the square brackets and obtain the integrand

$$\sqrt{\frac{1-\kappa^2}{1-\kappa^2\sin^2\omega}} - \sqrt{1-\kappa^2}. \tag{72}$$

This is a continuous function in ω which has the value zero at $\omega = 0$ and the value $1 - \sqrt{1-\kappa^2}$ at $\omega = \pi/2$. Connecting these two points, one obtains the diagonal of a rectangle with area $(1 - \sqrt{1-\kappa^2})\frac{\pi}{2}$. In the following we want to show that the integral can be bounded from above by half of the area of the described rectangle. For this, we have to show that the part of the area A of the rectangle above the integrand is greater or equal to the part B of the area below the integrand (the integral itself). If this holds, we have $B \leq (A+B)/2$. Instead of comparing A and B, we compare their respective integrands. We obtain the integrand of A by twice reflecting the integrand of B: once in the vertical line through $\pi/4$, once in the horizontal line through $\frac{1-\sqrt{1-\kappa^2}}{2}$. The resulting inequality is

$$\sqrt{\frac{1-\kappa^2}{1-\kappa^2\sin^2\omega}}-\sqrt{1-\kappa^2} \leq 1-\sqrt{1-\kappa^2}-\left[\sqrt{\frac{1-\kappa^2}{1-\kappa^2\sin^2\left(\frac{\pi}{2}-\omega\right)}}-\sqrt{1-\kappa^2}\right]$$
$$= 1-\sqrt{\frac{1-\kappa^2}{1-\kappa^2\cos^2\omega}}, \quad (73)$$

or put differently,

$$\left\{\sqrt{\frac{1-\kappa^2}{1-\kappa^2\sin^2\omega}}+\sqrt{\frac{1-\kappa^2}{1-\kappa^2\cos^2\omega}}\right\} \leq 1+\sqrt{1-\kappa^2}. \quad (74)$$

The expression in curly brackets is symmetric around $\pi/4$ because of $\cos^2\omega = \sin^2(\pi/2-\omega)$. In order to prove the inequality, we just have to show that the expression is monotonic decreasing up to $\pi/4$ (for $\omega = 0$ we have equality). For this we take the derivative of the LHS w.r.t. ω:

$$\frac{d}{d\omega}\{\} = \frac{\kappa^2}{2}\sin 2\omega\left[\left(1-\kappa^2\sin^2\omega\right)^{-\frac{3}{2}}-\left(1-\kappa^2\cos^2\omega\right)^{-\frac{3}{2}}\right] \leq 0. \quad (75)$$

The expression is smaller or equal to zero, since because of $\cos^2\omega \geq \sin^2\omega$ for $\omega \leq \pi/4$ the contents of the square brackets is smaller or equal to zero, while the prefactor stays positive because of $\sin 2\omega \geq 0$ for $\omega \leq \pi/4$. Thus we have shown that $B \leq A$, or

$$\int_0^{\frac{\pi}{2}}\left[\sqrt{\frac{1-\kappa^2}{1-\kappa^2\sin^2\omega}}-\sqrt{1-\kappa^2}\right]d\omega \leq \frac{1}{2}\left(1-\sqrt{1-\kappa^2}\right)\frac{\pi}{2}. \quad (76)$$

We continue with (71). With the bound of the integral we obtain

$$\Theta(\beta) = \coth 2\beta\frac{2}{\pi}\int_0^{\frac{\pi}{2}}\left[\sqrt{\frac{1-\kappa_\beta^2}{1-\kappa_\beta^2\sin^2\omega}}-\sqrt{1-\kappa_\beta^2}\right]d\omega$$
$$\leq \coth 2\beta\frac{2}{\pi}\frac{1}{2}\left(1-\sqrt{1-\kappa_\beta^2}\right)\frac{\pi}{2}. \quad (77)$$

Again using (70), the last expression becomes

$$\frac{1}{2}\coth 2\beta\left(1-\frac{\sinh^2 2\beta-1}{\cosh^2 2\beta}\right)=$$
$$\frac{1}{2}\,\frac{\cosh 2\beta}{\sinh 2\beta}\,\frac{\cosh^2 2\beta-\sinh^2 2\beta+1}{\cosh^2 2\beta}=\frac{1}{\sinh 2\beta\cosh 2\beta}. \quad (78)$$

Equation (26):

1.) $-\ln\left(\sqrt{2}\cosh 2\beta\right)+2\beta\tanh 2\beta\leq 2\beta_c(\beta-\beta_c)+\sqrt(2)\beta_c-\ln 2$

We expand the LHS into a Taylor series around β_c:

$$\Big[2\beta\tanh 2\beta-\ln(\sqrt{2}\cosh 2\beta)\Big]_{\beta=\beta_c}=$$
$$2\beta_c\tanh 2\beta_c-\ln(\sqrt{2}\cosh 2\beta_c)+\frac{4\beta_c}{\cosh^2 2\beta_c}(\beta-\beta_c)$$
$$+\frac{4}{\cosh^2 2\beta_c}\left[1-4\beta_c\tanh 2\beta_c\right]\frac{(\beta-\beta_c)^2}{2}+O\left((\beta-\beta_c)^3\right). \quad (79)$$

As one can see, the second derivative of the LHS is smaller zero for $4\beta\tanh 2\beta>1$ (which holds for $\beta>\beta_c$), so the function stays below its tangent in β_c, and for an upper bound the series can be truncated after the first term:

$$2\beta\tanh 2\beta-\ln(\sqrt{2}\cosh 2\beta)\leq\sqrt{2}\beta_c-\ln 2+2\beta_c(\beta-\beta_c). \quad (80)$$

(Notice that $\cosh 2\beta_c=\sqrt{2}$.)

2.) $-\ln\left(\sqrt{2}\cosh 2\beta\right)+2\beta\tanh 2\beta\leq\frac{-\beta}{\sinh 2\beta\cosh 2\beta}+\ln\sqrt{2}$

We have the equality

$$-\ln(\sqrt{2}\ \cosh 2\beta)+2\beta\ \tanh 2\beta=\beta\left(2\tanh 2\beta-2\right)+\ln\frac{\sqrt{2}}{1+e^{-4\beta}}. \quad (81)$$

(From the definition of $\cosh 2\beta$ we factored out $e^{2\beta}$.) Moreover

$$2\tanh 2\beta-2+\frac{1}{\sinh 2\beta\cosh 2\beta}=\frac{(\sinh 2\beta-\cosh 2\beta)^2}{\sinh 2\beta\cosh 2\beta}=\frac{e^{-4\beta}}{\sinh 2\beta\cosh 2\beta}. \quad (82)$$

With these it follows that

$$-\ln(\sqrt{2}\cosh 2\beta)+2\beta\tanh 2\beta-\frac{\beta}{\sinh 2\beta\cosh 2\beta}\leq$$
$$\beta\frac{e^{-4\beta}}{\sinh 2\beta\cosh 2\beta}-\ln\left[1+e^{-4\beta}\right]+\ln\sqrt{2}\leq\ln\sqrt{2}. \quad (83)$$

The last relation was obtained using the fact that the sum of the first two terms does not exceed zero. To show this, we modify the first term as follows:

$$\beta \frac{e^{-4\beta}}{\sinh 2\beta \cosh 2\beta} = \frac{2\beta e^{-4\beta}}{\sinh 4\beta} \leq \frac{2\beta e^{-4\beta}}{4\beta} = \frac{e^{-4\beta}}{2}. \tag{84}$$

Now we have the inequality

$$\begin{aligned} \frac{e^{-4\beta}}{2} &\leq \ln\left[1 + e^{-4\beta}\right] \\ &= \frac{e^{-4\beta}}{2} + \frac{e^{-4\beta}}{2} - \frac{e^{-8\beta}}{2} + \sum_{n=2}^{\infty}\left[\frac{1}{e^{4\beta(2n-1)}(2n-1)} - \frac{1}{e^{4\beta(2n)}2n}\right], \end{aligned} \tag{85}$$

since on the RHS the terms behind the first $\exp(-4\beta)/2$ are pairwise greater 0 (we expanded $\ln(1+x)$, cf. [BS]). Thus (83) holds.

Equation (27): The function $s(x)$ can be rewritten as follows:

$$\begin{aligned} s(x) &= -\frac{1+x}{2}\ln\frac{1+x}{2} - \frac{1-x}{2}\ln\frac{1-x}{2} \\ &= \ln 2 - \frac{1}{2}\left[(1+x)\ln(1+x) + (1-x)\ln(1-x)\right]. \end{aligned} \tag{86}$$

The expression in square brackets is expanded (see again [BS]) and bounded below:

$$\begin{aligned} [\,] &= (1+x)\sum_{n=1}^{\infty}(-1)^{n+1}\frac{x^n}{n} - (1-x)\sum_{n=1}^{\infty}\frac{x^n}{n} \\ &= \sum_{n=1}^{\infty}\left[(-1)^{n+1}\frac{x^n}{n} - \frac{x^n}{n}\right] + x\sum_{n=1}^{\infty}\left[(-1)^{n+1}\frac{x^n}{n} + \frac{x^n}{n}\right] \\ &= -\sum_{n=1}^{\infty}\frac{x^{2n}}{n} + 2\sum_{n=1}^{\infty}\frac{x^{2n}}{2n-1} = \sum_{n=1}^{\infty}\frac{x^{2n}}{2n^2-n} \geq x^2 + \frac{x^4}{6}. \end{aligned} \tag{87}$$

This bound is possible since all the coefficients in the sum are positive. Together with (17) for m_β we thus obtain

$$s(m_\beta) \leq \ln 2 - \frac{(1-\sinh^{-4}2\beta)^{\frac{1}{4}}}{2} - \frac{(1-\sinh^{-4}2\beta)^{\frac{1}{2}}}{12}. \tag{88}$$

■

Proof of Lemma 3.5: In order to show that

$$\frac{(1-(1+y^2)^{-4})^{\frac{1}{4}}}{2} > \frac{y^2}{2}\ln\frac{2+y^2}{y^2}, \qquad 0 \leq y \leq \frac{1}{2} \tag{89}$$

we show that the LHS is greater than $\frac{3}{5}y$, the RHS is smaller than $\frac{3}{5}y$. So for the RHS we have to show

$$\frac{5}{6}y \ln \frac{2+y^2}{y^2} < 1. \tag{90}$$

We only need $y \leq 1/2$. In this case $2 + y^2 \leq 9/4$, and thus we also have

$$\frac{5}{6}y \ln \frac{2+y^2}{y^2} \leq \frac{5}{6}y \ln \frac{9}{4y^2} = -\frac{5}{3}y \ln \frac{2}{3}y. \tag{91}$$

By equating the first derivative to zero we obtain the maximum of the function $-y \ln(2y/3)$ at $3/(2\mathrm{e})$. Hence

$$-\frac{5}{3}y \ln \frac{2}{3}y \leq \frac{5}{3}\,\frac{3}{2\mathrm{e}} < 1. \tag{92}$$

With this, (90) is shown for $y \leq 1/2$. For the LHS of (89) one has to prove:

$$\frac{(1-(1+y^2)^{-4})^{\frac{1}{4}}}{y} > \frac{6}{5}. \tag{93}$$

The LHS is modified as follows:

$$= \sqrt[4]{\frac{(1+y^2)^4 - 1}{y^4(1+y^2)^4}} = \frac{1}{(1+y^2)}\sqrt[4]{\frac{y^8+4y^6+6y^4+4y^2}{y^4}} \geq \frac{\sqrt[4]{4y^{-2}}}{1+y^2}. \tag{94}$$

Since in the last expression the denominator is monotonically decreasing, the numerator increasing, for a lower bound it suffices to evaluate the expression for the greatest y (we again choose $y \leq 1/2$):

$$\frac{\sqrt[4]{4y^{-2}}}{1+y^2} \geq \frac{\sqrt[4]{4(\frac{1}{2})^{-2}}}{1+\frac{1}{4}} = \frac{8}{5} > \frac{6}{5}, \qquad y \leq \frac{1}{2}. \tag{95}$$

■

Acknowledgements

I.E. thanks Jürgen Jost for general support and Ulrich Steinmetz for discussions.

References

[Am] Amari, Shun-ichi: *Information geometry on hierarchy of probability distributions*, IEEE Trans. Inform. Theory **47** (2001) 1701-1711.

[Ar] Arnold, Dirk V.: *Information-theoretic analysis of phase transitions*, Complex Systems **10** (1996) 143-155.

[Ay1] Ay, Nihat: *An information geometric approach to a theory of pragmatic structuring*, The Annals of Probability **30** 1 (2002) 416-436.

[Ay2] Ay, Nihat: *Locality of global stochastic interaction in directed acyclic networks*, Neural Computation **14** 12 (2002) 2959 – 2980.

[AK] Ay, Nihat; Knauf, Andreas: *Maximizing Multi-Information*, Max-Planck-Institute of Mathematics in the Sciences Preprint no. 42 (2003).

[Ba] Bauer, Heinz: *Wahrscheinlichkeitstheorie und Grundzüge der Maßtheorie*, Walter de Gruyter & Co, Berlin, 1968.

[Be] Bell, A.J.: *The co-information lattice*, Preprint Redwood Neuroscience Institute

[BS] Bronstein, I.N.; Semendjajew, K.A.: *Taschenbuch der Mathematik*, BSB B.G. Teubner Verlagsgesellschaft, Leipzig, 1989.

[Ca] Catlin, D.E.: *Estimation, Control, and the Discrete Kalman Filter*, Springer-Verlag, New York 1989.

[Ch] Choquet, Gustave: *Lectures on Analysis, Vol. I-III*, W.A. Benjamin, London 1969.

[CT] Cover, Thomas M.; Thomas, Joy A.: *Elements of Information Theory*, John Wiley & Sons, Inc., New York 1991.

[CY] Crutchfield, J.P.; Young, K.: *Computation at the onset of chaos*, in: Complexity, Entropy and the Physics of Information, Hrsg.: Zurek, W.H., Addison-Wesley Publishing Company, Reading Mass. 1990.

[De] Deco, Gustavo; Obradovic, Dragan: *An Information-Theoretic Approach to Neural Computing*, Springer-Verlag, New York 1996.

[Do] Dobrushin, R.L.: *The description of a random field by means of conditional probabilities and condition of its regularities*, Th. Prob. Appl. **13** (1968) 458-486.

[FC] Feldman, David P.; Crutchfield, James P.: *Discovering noncritical organization: statistical mechanical, information theoretic, and computational views of patterns in one-dimensional spin systems*, Santa Fe Institute Working Paper 98-04-026.

[Ge] Georgii, Hans-Otto: *Gibbs Measures and Phase Transitions*, Walter de Gruyter, Berlin, New York 1988.

[GL] Greenfield, Elliot; Lecar, Harold: *Mutual information in a dilute, asymmetric neural network model*, Physical Review E **63** (2001) 041905.

[Gr] Griffeath, D: *Introduction to random fields*, in: Kemeney, J.G; Snell, J.L.; Knapp, A.W.: *Denumerable markov chains*, Springer New York, Heidelberg, Berlin 1976.

[HKO] Hyvärinen, A; Karhunen, J; Oja, E: *Independent Component Analysis*, John Wiley 2001.

[Is] Israel, Robert B.: *Convexity in the Theory of Lattice Gases*, Princeton University Press, Princeton 1979.

[J1] Jaynes, E.T.: *Information theory and statistical mechanics*, Physical Review **106** (1957) 620-630.

[Li] Li, Wentian: *Mutual information versus correlation functions*, Journal of Statistical Physics **60** 5/6 (1990) 823-837.

[L] Linsker, R: *Self-organization in a perceptual network*, IEEE Computer **21** (1988) 105-117.

[LR] Lanford, O.E.; Ruelle, D.: *Observables at infinity and states with short range correlations in statistical mechanics*, Comm. Math. Phys. **13** (1969) 194-215.

[LF] Luque, Bartolo; Ferrera, Antonio: *Measuring mutual information in random Boolean networks*, Complex Systems **12** (2001) 241-246.

[MKNYM] Matsuda, H; Kudo, K.; Nakamura, R.; Yamakawa, O.; Murata, T.: *Mutual information of Ising systems*, International Journal of Theoretical Physics **35** 4 (1996) 839-845.

[MW] McCoy, Barry M.; Wu, Tai Tsun: *The Two-Dimensional Ising Model*, Harvard University Press, Cambridge Massachusetts 1973.

[Pa] Pathria, R.K.: *Statistical Mechanics*, Butterworth-Heinemann, Oxford 1996.

[Pe] Petritis, Dimitri: *Thermodynamic formalism of neural computing*, in: Dynamics of Complex Interacting Systems, Hrsg.: Goles, E.; Servet, M; Kluwer Academic Publishers, Dordrecht, Boston, London 1996.

[R] Ruelle, David: *Statistical Mechanics - Rigorous Results*, World Scientific, Singapore / Imperial College Press, London 1999.

[ST] Samengo, Inés; Treves, Alessandro: *Representational capacity of a set of independent neurons*, Physical Review E, **63** (2000) 011910.

[Sh] Shannon, C.E.: *A mathematical theory of communication*, Bell System Tech. J. **27** (1948) 379-423, 623-656.

[S] Simon, Barry: *The Statistical Mechanics of Lattice Gases*, Princeton University Press, Princeton 1993.

[SV] Studenty, M.; Vejnarova, J: *The multiinformation function as a tool for measuring stochastic dependence* in Jordan, M.I. (ed.) *Learning in Graphical models*, Dordrecht, Kluwer 1998.

[TSE] Tononi, Giulio; Sporns, Olaf; Edelman, Gerald: *A measure for brain complexity: Relating functional segregation and integration in the nervous system*, Proc. Natl. Acad. Sci. USA **91** (1994) 5033-5037.

[Wa] Wannier, Gregory H.: *Statistical Physics*, Dover Publications, Inc., New York 1987.

[Wi] Winkler, Gerhard: *Image Analysis, Random Fields and Dynamic Monte Carlo Methods*, Springer-Verlag Berlin Heidelberg New York 1995.