Biol. Cybern. 73, 69–81 (1995)

Biological Cybernetics

Detecting higher-order interactions among the spiking events in a group of neurons

L. Martignon [1], H. Von Hasseln [1], S. Grün [2,3], A. Aertsen [2], G. Palm [1]

[1] Department of Neural Information Processing, University of Ulm, Oberer Eselsberg, D-89069 Ulm, Germany
[2] Center for Research of Higher Brain Functions, Department of Neurobiology, Weizmann Institute of Science, Rehovot 76100, Israel
[3] Institut für Neuroinformatik, Ruhr Universität, D-44780 Bochum, Germany

Received: 9 September 1994/Accepted in revised form: 13 January 1995

Abstract. We propose a formal framework for the description of interactions among groups of neurons. This framework is not restricted to the common case of pair interactions, but also incorporates higher-order interactions, which cannot be reduced to lower-order ones. We derive quantitative measures to detect the presence of such interactions in experimental data, by statistical analysis of the frequency distribution of higher-order correlations in multiple neuron spike train data. Our first step is to represent a frequency distribution as a Markov field on the minimal graph it induces. We then show the invariance of this graph with regard to changes of state. Clearly, only linear Markov fields can be adequately represented by graphs. Higher-order interdependencies, which are reflected by the energy expansion of the distribution, require more complex graphical schemes, like constellations or assembly diagrams, which we introduce and discuss. The coefficients of the energy expansion not only point to the interactions among neurons but are also a measure of their strength. We investigate the statistical meaning of detected interactions in an information theoretic sense and propose minimum relative entropy approximations as null hypotheses for significance tests. We demonstrate the various steps of our method in the situation of an empirical frequency distribution on six neurons, extracted from data on simultaneous multineuron recordings from the frontal cortex of a behaving monkey and close with a brief outlook on future work.

1 Introduction

There is a growing consensus that processing in the brain is organized in functional groups of neurons. Following Hebb (1949), these groups are commonly referred to as 'cell assemblies'. Over the years, a number of different, somewhat conflicting definitions of 'neuron assembly' have been proposed. Some of these were phrased in terms of anatomy, others in terms of shared function (e.g. motor), or shared stimulus response (for a review, see Gerstein et al. 1989). One operational definition for the cell assembly has been particularly influential: near-simultaneity or some other specific timing relation in the firing of the participating neurons. Such temporal coherence is at least in principle important to brain function: if two neurons converge on a third one, their synaptic influence is much larger for near-coincident firing, due to the spatiotemporal summation in the dendrite. Thus, synchrony of firing is directly available to the brain as a potential neural code (Abeles 1991).

In pursuit of experimental evidence for cell assembly activity in the brain, physiologists thus seek to observe the activities of many separate neurons simultaneously, preferably in awake, behaving animals. These 'multineuron activities' are then inspected for possible signs of interactions between the neurons. Results of such analyses may be used to draw inferences regarding the processes taking place within and between hypothetical cell assemblies. The conventional approach to study neuronal interactions is based on the use of cross-correlation techniques, usually applied to the activity of pairs (sometimes triplets) of neurons recorded under some appropriate stimulus conditions. The result is a time-averaged measure of the temporal correlation among the spiking events of the observed neurons under those conditions. Recent developments in analysis methodology have considerably expanded the scope of these studies. Thus, it is now possible to examine cooperativity in larger groups of neurons (Gerstein and Aertsen 1985; Aertsen et al. 1987), and to study the dynamic properties of the firing correlation between two neurons in fine detail (Aertsen et al. 1989). Application of these new measures has revealed interesting instances of time- and context-dependent synchronization dynamics in different cortical areas, particularly in awake, behaving animals (e.g. Aertsen and Gerstein 1991; Vaadia and Aertsen 1992; Vaadia et al. 1995).

Correspondence to: L. Martignon

Most of the above-described approaches were based on some form of averaging, either over time or over trials. In a real working brain, however, there is clearly no time for averaging: problems are solved as they come, without repetition and as singular events. Thus, recent investigations have focussed on the detection of individual instances of synchronized activity: 'unitary events', consisting of precise spike patterns in multiple-neuron activity, occurring more frequently than expected by chance (Abeles and Gerstein 1988; Grün et al. 1994). Indeed, recently many instances of such excessively repeating spike patterns were found in multineuron activity in the frontal cortex of behaving monkeys (Abeles et al. 1993a, b). These spatiotemporal patterns typically lasted up to a few hundreds of milliseconds, with an individual spike timing precision of 2-3 ms. Moreover, the composition, frequency and timing of these constellations of multiple spiking events were often associated with the occurrence of external (stimulus or behavioral) events. These findings indicate that precisely timed, higher-order interactions among groups of neurons may effectively be involved in the functional organization of assembly activity in the brain.

In view of these results, we set out to develop a conceptual framework that should enable us to capture the interactions among multiple neurons in a more formal sense. Evidently, this framework should not be restricted to the conventional case of interactions among pairs of cells, but also incorporate higher-order interactions, which cannot be reduced to lower-order ones. Finally, it should allow one to derive quantitative measures to detect the presence of such interactions in experimental data, by careful analysis of the statistical properties of various order correlations among multiple neuron spike train data. Here, we propose such a framework, based on the probabilistic formalism of a Markov random field, the geometric formulation of connectivity graphs, and the information theoretical measure of entropy. We will illustrate these various ideas by their application to data from physiological multiple neuron recordings in the cortex of the awake, behaving monkey.

1.1 Formulation of the problem: neuronal interactions and probability measures

It was Caianiello's idea (1975, 1986) to make systematic use of what he called the η- and χ-expansions of the signum function on binary artificial neurons, aiming at linearizing his neuronic equations. In his original notation the η-expansion corresponded to the choice of the bipolar states ± 1, whereas the χ-expansion was determined by the choice of the base-two states, 0, 1. He also studied the problem of the invariance of certain structural features expressed by the properties of these polynomial expansions with respect to changes of state.

In this paper we apply his approach to the non-deterministic version of the same situation, that is to say, we investigate the η- and χ-expansions of the negative natural logarithm of a strictly positive probability distribution on the configurations of a set of binary nodes.

More generally, we look at the polynomial expansion of the negative natural logarithm of such a distribution in terms of any pair of states $a \neq b$. It is an elementary but useful observation that its coefficients determine – unambiguously – the minimal graph on which the starting probability distribution satisfies the Markov property (i.e. the induced graph is the one with the least number of edges among those graphs, on which the distribution is a Markov field). The connectivity structure of the graph is not dependent on the particular choice of the states a, b. On this graph the distribution is a Markov field, and the polynomial expansion of its negative natural logarithm is its energy, with simple expressions for the potentials. The geometric and information theoretic properties of the manifold of these distributions have been explored in detail by Amari in a sequence of fundamental articles (Amari 1982, 1985, 1991, 1994; Amari and SunHan 1989; Amari et al. 1992), which form part of the theoretical basis of this work. In our approach, which is Markovian as well as information theoretic, we will work with the concept of energy, i.e. the natural logarithm of the distribution, and the concept of surprise, i.e. the logarithm in base 2 of the distribution, inspired by the terminology developed by Palm (1981).

In Sect. 2 we illustrate the construction of the minimal graph guided by the coefficients of the energy expansion. Once the graph is constructed, the theorem of Hammersley and Clifford (see Griffeath 1976; Grimmet 1973; Hammersley and Clifford 1968) or the polynomial version of this theorem given by Besag (1974) can be invoked to prove the Markov field property. Yet, in the case of binary neurons, things are so transparent that we present a short proof of the asserted property, for the sake of completeness. The invariance of the graph with respect to changes of states is also checked directly.

A straightforward interpretation of the energy coefficients as weights of the graph's edges only makes sense in the linear situation, i.e. when the energy expansion is of order 2. Thus, distributions which happen to have linear energies are in one-to-one correspondence with the weighted graphs that represent them. Clearly, in this case all weights are strictly dependent on the choice of the numerical values taken by the states (they could be 0, 1 in the binary case or -1, 1 in the bipolar case and, more generally, any $a, b \in \mathbb{R}$). In general, Markov graphs are not apt to reflect the amount of information contained in a density distribution, since they represent higher-order terms simply as cliques, which are unions of edges. In order to represent higher-order terms of non-linear energies, we use constellations, sets of edges and star-shapd connections of more than two nodes, and assembly diagrams, which are schemes conceived to represent interactions due to simultaneous activation and their intensities.

In specific situations, like modeling databases for classification tasks (see, for instance, Miller and Goodman 1993), it is appropriate to approximate the density distribution determined by empirical data by means of a linear Markov field. Obviously, linear and, in general, low-order approximations have the advantage of reducing the computational complexity of the energy expansion. But there are also contexts in which the information carried by higher-order terms is of intrinsic relevance, as is the case in the neurobiological scenario, where it is

important to detect significant simultaneous activity of groups of neurons (due, for example, to common input activation or some other form of activity synchronization).

Following both the classical statistical approach of Kullback (1968) and the geometric-information theoretic approach of Csiszár (1975) and Amari (1991, 1994), Amari et al. (1992), we analyse in Sect. 3 the interaction degree of a distribution, which can be defined as the order of the polynomial expansion of the energy. Equivalent definitions can be formulated in terms of the marginals and of the frequencies of configurations. The distributions whose interaction degree is less than or equal to a fixed number also form a special submanifold; and the collection of these submanifolds satisfies interesting geometric properties. A fundamental result (see Csiszár 1975; Kullback 1968) is the following:

> given a distribution of interaction degree of at least i, there exists a unique distribution of interaction degree i, whose marginals or order i coincide with those of the given distribution and whose relative entropy from it is minimal. The two distributions coincide if among those with an energy expansion of order i they have the same degree of interaction.

An iterative algorithm that finds this ith degree best approximation with the same ith order marginals of a given distribution was provided in the late 1960 by Kullback et al. (Csiszár 1975; Gokhale and Kullback 1978; Ireland and Kullback 1968; Ku and Kullback 1969), dating back to an earlier work of Deming and Stephan (1940). A rigorous proof of its convergence was given by Csiszár (1975). Optimal low-degree approximations of high-degree distributions become adequate null hypotheses for the significance tests of detected higher-order interactions. Thus, the Markov field approach to the interpretation of the frequency distribution is combined with an information theoretic Ansatz in the application of Fisher's significance tests.

A natural context of applicability of our method is the neurobiological scenario, where interactions among neuronal spike trains have been analysed in a variety of settings and mathematical formalisms. Significant correlations, as treated by Palm, Aertsen and Gerstein (1988) and by Grün et al. (1994) and significant interactions, as we treat them here, describe strongly interrelated notions, as will become evident in the forthcoming sections. For reasons of consistency, we will use the term 'correlations' when referring to the phenomenon of time-related firing of spiking events and to 'interactions' as the underlying mechanism giving rise to these 'correlations'.

In Sect. 4 we provide a new methodology to analyse data on multiunit recordings in a Markov field frame, aiming at detecting interactions of all orders. We illustrate our method with empirical data on the frequencies of simultaneous activity of a set of neurons obtained by Grün et al. (1994) from the statistical analysis of experimental results on multiunit recordings by Vaadia et al. (1989, 1991).

2 The mathematical model

The model we use to represent distributions of frequencies consists of Markov fields as developed in Griffeath (1976). Since our aim is to model higher-order interactions, we will not restrict ourselves to Markov fields on graphs. After characterizing the class of distributions adequately modelled by graphs, we will go on to construct what we will call constellations and assembly diagrams.

2.1 Graphs

Assume that we are given a set of neurons (nodes, sites, predicates, . . .) labelled 1 through N, which at each time point can be in one of two possible states a or b, where a, b are real numbers and $a \neq b$. With Ω we denote the space of all configurations of a's and b's on the neurons, and with x_i, $1 \leqslant i \leqslant N$, the ith evaluation function that maps each configuration on its ith component, and by $\mathbf{x} = (x_1, \ldots, x_N)$ a vector in Ω.

Let π be any strictly probability distribution on Ω and denote with H the (well-defined) negative, natural logarithm of π. In terms of x_i, $1 \leqslant i \leqslant N$, $H = -\ln \pi$ admits a unique expansion

$$H(\mathbf{x}) = G_0 + \sum_{1 \leqslant i \leqslant N} G_i x_i + \sum_{1 \leqslant i < j \leqslant N} G_{ij} x_i x_j + \sum_{1 \leqslant i < j < k \leqslant N} \times G_{ijk} x_i x_j x_k + \ldots + G_{12\ldots N} x_1 \ldots x_N \qquad (1)$$

(Note that we have incorporated the normalization of π in G_0, which in statistical physics terms is the negative of the free energy.) This follows from the fact that H is a vector in the 2^N-dimensional vector space $\mathbb{R}^{\{a,b\}^N}$ of all functions of $\{a,b\}^N$ into $\mathbb{R}$ and that the 2^N functions

$$\{1, x_1, x_2, x_1 x_2, \ldots, x_1 x_2 \ldots x_N\}$$

form a basis of $\mathbb{R}^{\{a,b\}^N}$. Following Caianiello we use the terms η-expansion and χ-expansion for the special cases $a = 1$, $b = -1$ and $a = 0$, $b = 1$, respectively.[1]

We recall that a graph $\mathcal{G}$ on Λ is just any set of edges or pairs (i,j) of elements of Λ. If neurons i and j are connected by an edge they are said to be neighbours; the set of neighbours of k is denoted by $\mathcal{N}_k$ and is called the neighbourhood of k. A clique is a subset of Λ whose neurons are all neighbours of each other. A clique is maximal if it is not strictly contained in any other one. Let $k \in \Lambda$ and $\mathbf{x} \in \Omega$. We denote with $\bar{\mathbf{x}}^k$ the configuration that differs from $\mathbf{x}$ only in the kth component and thus attains on it the complementary value b if $x_k = a$ and vice versa. For any fixed $\mathbf{x} \in \Omega$ and $k \in \Lambda$, we denote with $V^k(\mathbf{x})$ the set of all $\mathbf{y} \in \Omega$ such that $x_j = y_j$ for all $j \in \mathcal{N}_k$. We say that $V^k(\mathbf{x})$ is the kth vicinity of $\mathbf{x}$.

[1] In order to determine the 2^N coefficients of H, we formally expand $H(\mathbf{x})$, for each configuration $\mathbf{x}$ according to (1), obtaining 2^N linear equations, whose variables are the coefficients. This system has a unique solution. If $a = 0$ and $b = 1$, the computations are quite simple. By replacing $\mathbf{x} = (0, 0, \ldots, 0)$, we obtain $-\ln \pi(0, 0, \ldots, 0) = G_0$. By replacing $\mathbf{x} = (1, 0, 0, \ldots, 0)$, we obtain $-\ln \pi(1, 0, 0, \ldots, 0) = G_0 + G_1$, and so on (compare with (7) in Sect. 3)

Given a graph $\mathcal{G}$ on a set of N nodes Λ with possible states a, b and a strictly positive distribution π on the space Ω of all configurations, we say that π satisfies the Markov property on $\mathcal{G}$ if

$$\mathcal{P}r(x_i|x_j, j \neq i, j \in \Lambda)$$
$$= \mathcal{P}r(x_i|x_j, j \neq i, j \text{ is a neighbour of } i). \quad (2)$$

for i, $1 \leqslant i \leqslant N$, $\mathbf{x}$ varying in Ω. Here, as usual $\mathcal{P}r(A|B)$ stand for the probability of event A given event B. If π satisfies the Markov property on $\mathcal{G}$, it is a Markov field on Λ.

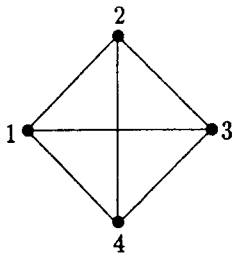
We now go back to the general situation where π is just any strictly positive distribution on Ω and give a simple rule for the construction of the graph induced by π. Observe that π obviously has the Markov property on the fully connected graph $\mathcal{F} = \{(i,j)\colon i \neq j,\ i,j \in \Lambda\}$.

The graph $\mathcal{G}$ induced by π is obtained by drawing every edge (i,j) such that the labels i and j appear as subindices of a non-zero coefficient of $H = -\ln \pi$.

Remark 1. We will not graphically represent the self-loops corresponding to terms $G_i\ x_i$, since we are mainly concerned with interactions between two or more neurons.

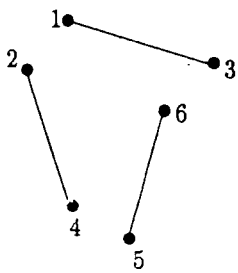
If $\mathcal{G}$ denotes the graph obtained according to the method given above, then π has the Markov property on $\mathcal{G}$. Furthermore, $\mathcal{G}$ is the graph with the least number of edges satisfying this property.

Example 1. (i) The distribution $\pi = e^{-\ln 40 + x_1 \ln 2 + x_2 \ln 2 + (x_1 x_2 x_3 x_4)\ln 2}$ induces the graph:



Observe that, in this case, $\mathcal{G}$ is connected, in fact, fully connected.

(ii) The distribution $\pi = e^{-(K + x_1x_3 + x_2x_4 + x_5x_6)}$, where K is the normalizing constant, induces the graph



Theorem 1. If $\mathcal{G}$ is the graph induced on Λ by a strictly positive distribution π on Ω, then

(i) π satisfies the Markov property on $\mathcal{G}$.

(ii) $\mathcal{G}$ is the smallest graph on the nodes of Λ on which π satisfies the Markov property.

(iii) The graph $\mathcal{G}$ induced by π is invariant with respect to changes of states, i.e. if a is replaced by a' and b by b', where $a' \neq b'$, the graph $\mathcal{G}'$ induced by π – redefined accordingly on the new configuration space Ω' – coincides with $\mathcal{G}$.

Proof. (i) We set

$$N_k = \{j\colon\ j \text{ and } k \text{ are sub-indices of a coefficient } G_{i_1 \ldots i_s} \neq 0\}$$

and

$$V^k(\mathbf{x}) = \{\mathbf{y}\colon\ x_i = y_i, i \in N_k\}$$

Given the form of $H(\mathbf{x})$, we have

$$H(\mathbf{x}) - H(\bar{\mathbf{x}}^k) = H(\mathbf{y}) - H(\bar{\mathbf{y}}^k) \quad \forall\, \mathbf{y} \in V^k(\mathbf{x}) \quad (3)$$

Therefore

$$\mathcal{P}r(x_k|x_i,\ i \in N_k)$$
$$= \frac{\sum_{\mathbf{y} \in V^k(\mathbf{x})} \exp - H(\mathbf{y})}{\sum_{\mathbf{y} \in V^k(\mathbf{x})} \exp - H(\mathbf{y}) + \sum_{\mathbf{y} \in V^k(\mathbf{x})} \exp - H(\bar{\mathbf{y}}^k)}$$
$$= \frac{\sum_{\mathbf{y} \in V^k(\mathbf{x})} \exp - H(\mathbf{y})}{[1 + \exp(H(\mathbf{y}) - H(\bar{\mathbf{y}}^k))] \sum_{\mathbf{y} \in V^k(\mathbf{x})} \exp - H(\mathbf{y})}$$
$$= \frac{1}{1 + \exp(H(\mathbf{y}) - H(\bar{\mathbf{y}}^k))} = \mathcal{P}r(x_k \,|\, x_i, i \neq k) \quad (4)$$

where we used (3) in passing from the second to the third line. Thus, π has the Markov property on $\mathcal{G}$, and N_k coincides with the set $\mathcal{N}_k$ of neighbours of k, as defined above. In order to prove (ii), we observe that (3) ceases to hold if we eliminate an edge (i,j) from $\mathcal{G}$.

(iii) We begin by remarking that there exist $\alpha, \beta \in \mathbb{R}$ such that $\alpha a + \beta = a'$ and $\alpha b + \beta = b'$. If $\mathbf{x}$ is any vector in Ω, we write $\mathbf{x}'$ for $\alpha\mathbf{x} + \beta$. On Ω' we define $\pi'(\mathbf{x}') = \pi(\mathbf{x})$ and write $H' = -\ln \pi'$. The coefficients of H' are given by

$$G'_{i_1 i_2 \ldots i_s} := \alpha^s \Bigg(G_{i_1 i_2 \ldots i_s} + \beta \sum_{1 \leqslant i_{s+1} \leqslant N} G_{i_1 i_2 \ldots i_s i_{s+1}}$$
$$+ \beta^2 \sum_{1 \leqslant i_{s+1} < i_{s+2} \leqslant N} G_{i_1 i_2 \ldots i_s i_{s+1} i_{s+2}}$$
$$+ \beta^3 \sum_{i \leqslant i_{s+1} < i_{s+2} < i_{s+3} \leqslant N} G_{i_1 i_2 \ldots i_s i_{s+1} i_{s+2} i_{s+3}}$$
$$+ \ldots + \beta^N G_{i_1 i_2 \ldots i_N} \Bigg) \quad (5)$$

Clearly, if $G_{i_1 \ldots i_r}$ is the non-zero coefficient of highest order in the expansion of H we necessarily have $G'_{i_1 \ldots i_r} \neq 0$. For a lower-order non-zero coefficient $G'_{i_1 \ldots i_r}$, we may have $G'_{i_1 \ldots i_r} = 0$, yet this happens only if some other coefficient $G'_{i_1 \ldots i_s}$ of H, such that $\{i_1, \ldots, i_r\} \subset \{i_1, \ldots, i_s\}$ is non-zero. We deduce that

some coefficient of H' whose sub-indices include those of $G'_{i_1 \ldots i_r}$ is necessarily non-zero. This proves our assertion. □

Remark 2. Observe that the χ-expansion of $-\ln \pi$ represents the energy in terms of the canonical potential of this Markov field (see Griffeath 1976).

The simplest way to determine the graph induced by π is to make use of the η-expansion of $-\ln \pi$. In fact, in the case of the bipolar states $\{-1, 1\}$, the functions

$$\left\{\frac{1}{2^N} f_A : A \subseteq \Lambda\right\}, \quad \text{where } f_A = \prod_{i \in A} x_i$$

form an orthonormal basis of $\mathbb{R}^{\{-1,1\}^N}$. Thus, the coefficients of the η-expansion H are given by

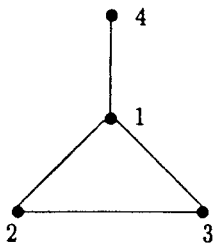
$$G_A = \left\langle \frac{1}{2^N} f_A, H \right\rangle = -\frac{1}{2^N} \langle f_A, \ln \pi \rangle$$

$$= -\frac{1}{2^N} \sum_{\mathbf{x}} f_A(\mathbf{x}) \ln \pi(\mathbf{x})$$

and $H(\mathbf{x}) = \sum_{A \subseteq \Lambda} G_A \cdot f_A(\mathbf{x})$, where $\langle \ldots \rangle$ denotes the usual scalar product in $\mathbb{R}^{\{-1,1\}^N}$.

2.2 Constellations and assembly diagrams

We saw that the minimal graph induced by a strictly positive distribution π on Ω is uniquely determined by the distribution and invariant with respect to changes of state. The converse is not true; take for instance $\pi_1 = e^{-(K_1 + x_1x_4 + x_1x_2x_3)}$ and $\pi_2 = e^{-(K + x_1x_4 + x_1x_2 + x_2x_3 + x_1x_3)}$. Both induce the same graph:



In certain contexts this can be a drawback of the graphs constructed above. Another problem is that, in general, there is no $1-1$ identification of the coefficients of H with the edges of $\mathcal{G}$. However, for distributions with linear energies, i.e. energies of the form

$$H(\mathbf{x}) = G_0 + \sum_{1 \leqslant i \leqslant N} G_i x_i + \sum_{1 \leqslant i < j \leqslant N} G_{ij} x_i x_j \qquad (6)$$

the situation is simple. In this case there is a natural $1-1$ correspondence between weights of edges and second-order coefficients of H. This leads to the following definition.

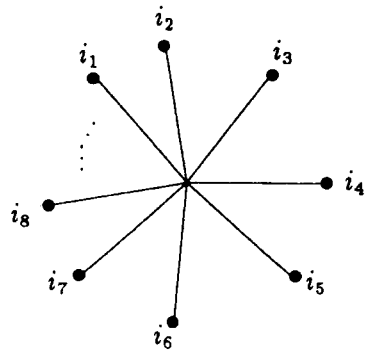
Definition 1. Let π be a strictly positive distribution on Ω. We say that π is a linear Markov field if $H = -\ln \pi$ has a polynomial expansion of degree two.

If π is a linear Markov field on Λ and $\{a, b\}$ is the set of states considered, each coefficient $G_{i,j} \neq 0$ of $H = -\ln \pi$ will be identified with the weight of the edge (i, j). Obviously, the set of weights is strictly dependent on the choice of a, b.

Linear Markov fields include e.g. discrete Hopfield models and Boltzmann machines. In both cases the third term of (6) is written as $\mathbf{x}^T J \mathbf{x}$, where J is an $n \times n$ matrix; our G_{ij} is then $\frac{1}{2}(J_{ij} + J_{ji})$, and $J_{ij} = J_{ji}$.

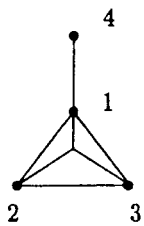
Our purpose is now to discuss more complex graphical schemes that represent higher-order interactions. The first one we propose is a mathematical generalization of graph, called constellation.

Definition 2. Let π be a strictly positive distribution on Ω and $H = -\ln \pi$. For every non-zero coefficient $G_{i_1 \ldots i_r}$ in H, we say that the r-tuple $(i_1, \ldots, i_r)$ is a star. In the case of two nodes only, we represent the corresponding star by an edge. If a star has more than two elements, or vertices, we represent it graphically by connecting the vertices $i_1, \ldots, i_r$ with an additional point $c_{i_1 \ldots i_r}$ which is not in Λ, as below:



We denote by $\mathcal{S}$ the set of all stars corresponding to maximal cliques in the graph determined by H and call the union of $\mathcal{G}$ and $\mathcal{S}$ the constellation induced by π.

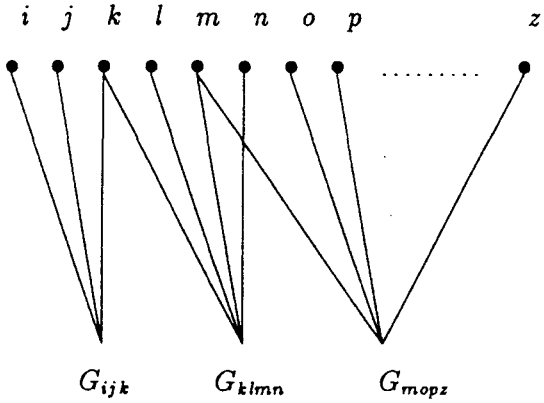
Example 2. Take $e^{-(K + x_1x_4 + x_1x_2x_3)}$ where K is the normalizing constant. The constellation $\mathcal{C}$ induced by π is



A moment's reflection proves the following theorem.

Theorem 2. The constellation induced by a strictly positive distribution on a binary configuration space is invariant with respect to changes of state.

Constellations share with graphs the invariance property and are useful in contexts where changes of state are necessary. But there are situations in which only a fixed pair of states makes sense, and the important issue is to represent all interactions, and only the interactions. In this case an assembly diagram, like the following, seems appropriate:



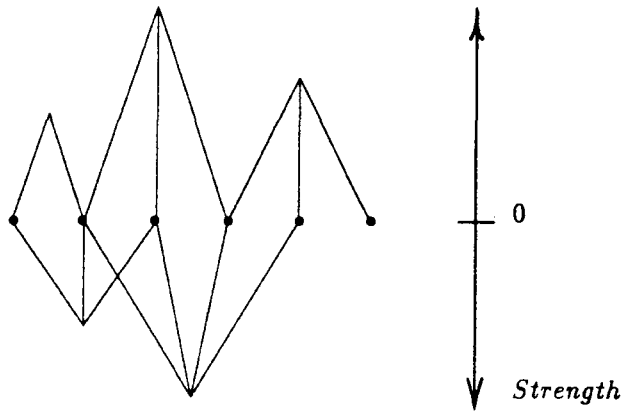
Here we arrange the neurons in a horizontal array and connect the subsets corresponding to non-zero coefficients of H to external points. It is possible to introduce a vertical dimension in the scheme, in order to represent the numerical values of the coefficients, as we will do in Sect. 3. Obviously, interactions between two neurons are also representable as connected by an external point. Thus, assembly diagrams will be useful in the neurobiological scenario in those situations in which simultaneous activation, even of two neurons, is the phenomenon being investigated, rather than mono-synaptical connections.

Remarks. Whereas in specific situations in the neurobiological context higher-order coefficients may carry fundamental information, in other scenarios the tendency is to linearize the Markov field, i.e. to replace it by another one with a second-order energy. This mainly reflects the wish of reducing the computational complexity. Such a replacement inevitably causes losses, and the effort is to minimize them. If the purpose is to model databases in order to perform classification tasks, truncating the energy expansion after the second-order coefficients and renormalizing provide an adequate linear approximation. Another way of reducing higher-order to second-order energies was proposed and used by Pinkas (1991) in the context of propositional calculus. In his approach propositions are embedded in an energy expansion. The advantage of this embedding is that energy minimization is equivalent to satisfiability. In this context, what has to be preserved when replacing a higher-order energy by a second-order one is the set of global minima. Pinka's strategy is to adjoin hidden units to the original set of nodes and determine the weights of the resulting edges in such a way that the new energy has the same set of minima as the original one.

In the information theoretic approach, as developed by Kullback et al. (Gokhale and Kullback 1978; Ireland and Kullback 1968; Ku and Kullback 1969) and Csiszár (1975), the effort is to linearize with a minimal loss of relative information. Kullback et al. found algorithms that approximate higher-order energies by means of lower-order ones, minimizing the loss of relative entropy. We will discuss their results in the next section. Amari (Amari 1982, 1985, 1991, 1994; Amari and SunHan 1989; Amari et al. 1992) unified the statistical, the information theoretic and the geometric approches of the exponential family of probability distributions by conceiving a mathematical model whose structural aspects reflect the properties investigated in each approach. In this framework, Amari also discussed the approximation of distributions by Boltzmann machines (Hinton and Sejnowski 1986), minimizing relative entropy.

3 The meaning of the energy coefficients in the χ-expansion

As we have seen, weighted graphs are adequate representations of linear Markov fields. In this section we propose assembly diagrams as adequate models for small sets of neurons embedded in larger nets, aiming at representing all possible simultaneous activations due to common inputs. Thus, we start with a list of (strictly positive) frequencies of configurations of zeros and ones on a set of N neurons and calculate the energy expansion of this distribution. We proceed by drawing an assembly diagram like the following:



where the joining points above the neuron-array represent positive coefficients and joining points below the neuron-array represent negative ones. The heights (and depths) of the joining points correspond to the absolute numerical values of the coefficients.

3.1 Positive and negative interactions

Our wish is now to extract from the energy coefficients all the information they contain on possible interactions. Amari et al. (1989) investigated the information on stochastic interactions contained in the second-order coefficients of a linear Markov field on binary (0, 1)-neurons. Their assertion is that G_{ij} represents

the degree of conditional dependence of the firing of the ith and jth neurons. In what follows, we generalize this statement. Assume that we have a strictly positive probability distribution π on N neurons, with possible states 0 and 1, and that the negative logarithm H of π is expanded as in Sect. 2. An easy computation shows that if A is any subset of $\{1, \ldots N\}$, then

$$-G_A = \sum_{a \subseteq A} (-1)^{|A-a|} \ln \pi(\chi_a) \tag{7}$$

where χ_a denotes the vector having ones at the components belonging to $a \subseteq A$ and zeros else where. Suppose, for instance, that we are dealing with six neurons and $A = \{1, 2, 3\}$. Solving for $\pi(1, 1, 1, 0, 0, 0)$ in the equation above, we get

$$\pi(1,1,1,0,0,0) = (\exp -G_{123}) \left[\frac{\pi(1,1,0,0,0,0)\pi(1,0,1,0,0,0)\pi(0,1,1,0,0,0)\pi(0,0,0,0,0,0)}{\pi(1,0,0,0,0,0)\pi(0,1,0,0,0,0)\pi(0,0,1,0,0,0)} \right] \tag{8}$$

The ratio on the right-hand side will be called interaction threshold of neurons 1, 2, 3. If $G_{123} = 0$, then the probability of the first three neurons firing simultaneously can be expressed exactly in terms of the probabilities of the simultaneous firings of their strict subsets. Loosely speaking, the interaction of these three neurons is not real since it is made up of the interactions of their subsets of two and one neurons. Thus, we declare that our data do not represent an interaction of this triplet of neurons. If $-G_{123} \gg 0$, the interaction threshold of the three neurons is dominated by the probability of their simultaneous firing, which is adequately interpreted as triple correlation or positive interaction of the triplet. Consequently, a negative interaction or an anticorrelation will correspond to a positive coefficient G_{123}.

These considerations are easily generalized to the case of any finite number of neurons. They lead to the following definitions.

Definition 3. Let π be a strictly positive distribution on the 0, 1-configuration of a set of N neurons. Furthermore, assume that the energy $H = -\ln \pi$ has been expanded as in (1). A subset A of neurons is

(i) correlated if $-G_A \gg 0$
(ii) uncorrelated if $G_A = 0$
(iii) anticorrelated if $-G_A \ll 0$

Inspired by the ideas developed in Palm et al. (1988), $-G_A$ will be called the interaction surprise of the set of neurons A.

Remark 3.[2] The G_A's are quantities which express the 'direct' interaction of neurons in A, which cannot be reduced to those of partial interaction (so that A is what we call a 'star'). From the geometrical point of view, the direction in which G_A changes but no marginals π_B (where $B \subset A$, properly) change is orthogonal in the manifold of probability distributions (in the sense of the Fisher information metric, or equivalently in the sense of correlations of the corresponding score functions) to any changes of the marginal distributions π_B, $B \subset A$ $(B \neq A)$ (see Amari and SunHan 1989).

[2] This fact was kindly pointed out by the referee

3.2 The interaction degree of a distribution

Assume that six coins are tossed together over and over again and that we are informed about the outcomes of each trial, but we cannot see the coins. These coins might be all connected to each other by threads we are not able to see, or perhaps an influence on the outcome. This is, at least, our conjecture. If after a large number of trials we observe that the outcomes are uniformly distributed, our tendency will be to imagine that there are no connections between the coins and that all coins are fair coins. Clearly, the energy expansion of the uniform distribution is reduced to the constant term G_0. If the coins are not interconnected but some of them are 'unfair', we will eventually have an energy expansion of polynomial degree 1.

In order to discuss interactions of degree higher than 1 in a formal way, we must make use of the elementary concept of marginals of a distribution. Let us briefly recall some basic facts on probability distributions in the restricted context of configuration spaces on binary nodes.

Let π be a strictly positive distribution on the set of 0, 1-configurations of a set Λ of N neurons. The marginals of the distribution are defined as the probabilities of configurations taking fixed pre-established values at some fixed components. The marginal functions are defined as follows:

Definition 4. *Let $A \subseteq \Lambda$. If $A = \emptyset$,* we define $\pi_\emptyset = 1$, For $A \neq \emptyset$, we define

$$\pi_A(\mathbf{x}) = \text{Prob}(x_i, i \in A)$$

for every $\mathbf{x} \in \Omega$.

The marginal functions have a dual character with regard to the products of components in the χ-expansion of the energy. Amari investigated this duality in an information geometric framework. The interesting aspect in our context is that the marginal functions allow us to construct another important potential, which again reflects the interactions of sets of neurons. For every $A \subseteq \Lambda$ define

$$J_A(\mathbf{x}) = \sum_{b \subseteq A} (-1)^{|A-b|} \ln \pi_b(\mathbf{x})$$

Combining the usual terminology of information theory (see e.g. Cover and Thomas 1991) and the concept of surprise (Palm 1981), we call J_A the mutual surprise of the subset A of Λ. We set $J_\emptyset = 1$ and observe that the family J_A, $A \subseteq \Lambda$ forms a potential of our Markov field (Griffeath 1976), since the sum of all mutual surprises is the energy of π. The canonical potential is related to the

mutual surprise potential by means of the following equation:

$$G_A \prod_{i \in A} x_i = \sum_{b \subseteq A \subseteq d \subseteq \Lambda} (-1)^{|A-b|} J_d(\mathbf{x}^b) \tag{9}$$

where $A \neq \emptyset$ and $\mathbf{x}^b$ is the configuration that coincides with $\mathbf{x}$ at the components in b and takes the value 0 elsewhere. The proof of this fact can be found in Griffeath (1976), Prop. 12.14.[3] We are now ready to define the degree of interaction of a distribution in terms of marginals. The definition we present here follows the classical approach of Ku and Kullback (1968):

Definition 5. For a fixed k with $1 \leqslant k \leqslant N$, let $\mathcal{M}_k$ be the set of all ith order marginals of π for $1 \leqslant i \leqslant k$. Assume that k is the first number between 1 and N for which $\mathcal{M}_k$ determines the distribution, in the sense that it represents sufficient statistics for π. Then k is called the interaction degree of π or the dependence degree of π. It will be denoted with $\deg(\pi)$.

Clearly, $\deg(\pi)$ is well and uniquely defined. The following theorem is from Kullback and Csiszár.

Theorem 3. Given a distribution π with $\deg(\pi) \geqslant i$, there exists a unique distribution of interaction degree i, whose marginals of order i coincide with those of the given distribution and whose relative entropy from it is minimal, among those with an energy expansion of order i. This distribution is the one with maximal entropy among those of degree i and same i-marginals as π.

An iterative algorithm that finds this ith degree best approximation of a given distribution maintaining its ith order marginals was provided in the late 1960s by Kullback et al. (Csiszár 1975; Gokhale and Kullback 1978; Ireland and Kullback 1968; Ku and Kullback 1969; see also Bishop et al. 1989). A rigorous proof of its convergence was given by Csiszár (1975). The following theorem combines the concepts of 'sufficient statistics' and 'log-linear models' in classical multivariate analysis (see Bishop et al. 1989; Martignon and Laskey 1995).

Theorem 4. For every strictly positive distribution π, $\deg(\pi)$ coincides with the order of the polynomial expansion of the energy.

Although $\deg(\pi)$ does not depend upon the choice of states, we usually refer to the χ-expansion of the energy.

We are now ready to tackle the problem of establishing the significance of the coefficients of the χ-expansion of the energy, which, as we have seen, reflect the interactions implicit in the frequencies. Of course, these detected interactions are reliable only if the data are reliable. The experimenter who provides us with the data may always fear that the frequencies he or she counted were the result of pure chance. This is one of the classical problems of data analysis or hypothesis testing in statistics, and there are classical methodologies to handle them. Fisher's methodology, which is perhaps the most popular, consists of a comparison of the frequency data with those that would arise under a so-called null hypothesis. In Palm et al. (1988) the significance of correlations of two neurons was established by using the null hypotheses of independence. In other words a null hypothesis of degree 1 with the same first-order marginals of the data was used to test the significance of second-order correlations.

If we want to test the significance of triplets instead of couples, we have to go one step further and search for the most adequate null hypothesis of degree 2. The natural approach is to use the minimum relative entropy approximation of degree 2 given by the iterative proportional fitting procedure (IPFP, see Appendix A). This procedure can obviously be used for every degree of interaction. The method we just described can be called the IPFP method for significance testing of interactions detected from empirical data. It is based on the principle of minimum discrimination information or minimum relative entropy.

We recall that if p is any strictly positive distribution on Ω, the relative entropy, or relative information, or even, in Kullback's wording, the discrimination information of p with respect to π is given by

$$\mathcal{I}(p; \pi) = \sum_{\mathbf{x}} p(\mathbf{x}) \ln \frac{p(\mathbf{x})}{\pi(\mathbf{x})}$$

4 Detecting interactions among six neurons

The method we developed was tailored for a specific task in data analysis, which arises in experimental situations, where the inner structure of a system can only be 'guessed' from outcomes of phenomena that, according to our hypothesis, reflect this structure. In the preceding section we discussed the paradigmatic situation of coins being tossed in such a way that we do not see them, but we know the outcome of each trial.

Another example is the following: suppose we want to detect the associations existing among six people, by observing, through a long period of time, the frequency with which subsets of them have lunch together. If we observe that three of them meet for lunch often, we will tend to think that they are associated, even more so if the subsets of order two of this triplet of persons meet for lunch seldom. Every meeting for lunch may be encoded as a vector of 0's and 1's, where each 0 represents an absent person and each 1 a present person. The coefficients of the energy expansion of the frequency distribution of these vectors will give us an idea of all possible associations.

The concrete application we will discuss here concerns data analysis of 'real neurons'. For an analysis of neurobiological scenarios, the nodes and connections of

[3] Let $\mathbb{1}$ denote the configuration whose components are all equal to 1. The numbers $J_A(\mathbb{1})$, for $A \subseteq \Lambda$, have interesting properties, which are an analogon – at the level of expectations – of the numbers G_A. But if we want to detect the interaction of the neurons A, G_A is more exactly a measure than $J_A(\mathbb{1})$, since the latter takes into account information on the simultaneous firing of neurons in sets larger than A

our abstract model must be adequately specified. The straightforward correspondence attributes nodes to neurons and the states 0, 1 to the neurons outputs. This means that a neuron is in state 1 whenever it generates an action potential and in state 0 if it is silent.

The specific experimental situation we are investigating here is the following. While a monkey was performing a delayed sensory-motor association task (data from Vaadia et al. 1989, 1991, 1995), the simultaneous spiking activity of several neurons was recorded. The monkey had to localize stimuli (visual or auditory, data are pooled) by putting his hand on a touch bar after a GO signal. The data for analysis (see Appendix B) were taken from time sections around the GO signal in trials selected for one particular spatial location. We normalized for stationarities of firing rates as follows: for each of the neurons, we estimated the instantaneous firing rates by evaluating the time-dependent probability of spike occurrences, averaged over trials (so-called PSTH). Observing these firing rates for all neurons in parallel and at discrete time steps (bins), we obtain a series of vectors, each one containing the instantaneous firing rates of the several neurons at the particular instant in time. These vectors are clustered by use of the k-means cluster algorithm. Vectors within a cluster define the time segment, in which the rates of the contributing neurons are (quasi) stationary in parallel, and thus represent a 'joint-stationary' regime (S. Grün, manuscript in preparation).

The simultaneous observation of the spiking events of N, e.g. six, neurons can be mathematically described as N-parallel point processes (Grün et al. 1994). By appropriate binning (here 3 ms), this can be transformed to an N-fold (0, 1) process. We can describe the joint activity of the N neurons for each time step (bin) as an N-vector, containing the 0's and 1's across the neurons. For N neurons, this vector can take any of 2^N possible values; coincident activity of any subgroup of neurons is represented by those vectors which contain multiple occurrences of 1's in their respective coordinates. By counting the number of times such configurations occur during the recording, we can experimentally determine the frequency of occurrence of such coincident events. In Appendix B.1 we have listed in a table all 64 theoretically possible configurations for the six neurons examined (first column), together with their computed frequencies of occurrence during the selected time interval of 930 time steps (second column). Observe that only 16 of 64 possible configurations did indeed occur at least once during this time period; for the remaining 48 configurations the empirical frequency of occurrence equals zero. Since our approach is based on expanding the negative logarithm of the distribution, these zero frequencies evidently pose a serious problem. One could argue that the zero frequencies are, in fact, due to the finite duration of observation, and that for a long enough measurement each possible configuration would eventually occur at least once. This has been our approach, and in the third column of Appendix B.1 we have listed the values of a strictly positive approximation of our original frequency distribution. Modifying the experimental data is a delicate matter. We have chosen the naive attitude of introducing a small quantity ε (here $\varepsilon = 1.0 \times 10^{-11}$) as the frequency of all 'silent' configurations. We are aware that equal treatment of all silent configurations is against the experimental evidence. In fact, from the data we have, we see that frequencies of configurations dramatically decrease as the number of 1's in them increases. A sounder way to produce an adequate strictly positive approximation of the empirical distribution would be to set ε^2 for silent configurations with two 1's, ε^3 for silent configurations with three 1's, and so on. There are, of course, even better statistical strategies based on the analysis of available data. Yet for the moment, since we are essentially illustrating a methodology, we will work with this rough approximation.

The interaction surprises of couples and triplets for the strictly positive approximation of the frequency distribution are listed in Appendix B.2. These numbers will be helpful for the analysis of those couples and triplets that are not silent, i.e. those whose coincident firing has a strictly positive 'real' frequency. For those which were silent during the observation interval, we have to declare that we have insufficient data. We begin by examing, for example, triplet $\{2,4,5\}$, which is non-silent (in fact, it occurred twice during the 930 time steps). We observe that one of its subsets of order two is silent, namely $\{2,5\}$. All other subsets are non-silent. If we were to calculate G_{245} by (7) for the empirical distribution π, we would get infinity. Writing down (8) for the triplet $\{2,4,5\}$ for the empirical distribution π, we obtain the formal expression

$$\pi(0,1,0,1,1,0) = (\exp - G_{245}) \left[\frac{\pi(0,1,0,1,0,0)\pi(0,0,0,1,1,0)\pi(0,1,0,0,1,0)\pi(0,0,0,0,0,0)}{\pi(0,1,0,0,0,0)\pi(0,0,0,1,0,0)\pi(0,0,0,0,1,0)} \right] \tag{10}$$

Since $\pi(0,1,0,0,1,0)$ is zero and no other frequencies are zero, we see that $\pi(0,1,0,1,1,0)$ definitely dominates the interaction threshold of this triplet. If we replace $\pi(0,1,0,0,1,0)$ by a small quantity ε and let ε tend to zero, then the interaction threshold tends to zero. This asymptotic study based on the empirical distribution indicates that our triplet $\{2,4,5\}$ has a relevant interaction. A look at G_{245} in the energy expansion of the strictly positive approximation p of π shows us that, as expected, the interaction surprise of $\{2,4,5\}$ is high. For the other non-silent triplet $\{1,2,5\}$, we can perform a similar analysis. There are two silent subsets, which are the pairs $\{1,2\}$ and $\{2,5\}$, and the approximated interaction threshold, obtained by replacing those two zero-frequencies by ε, tend to zero as ε tend to zero. It is important to note that this would be so even if we had chosen different small quantities for the two silent configurations. The coefficient G_{125} in Appendix B.2 denotes a very high interaction, even higher than that of triplet $\{2,4,5\}$. This is plausible since $\{1,2,5\}$ has more silent sub-pairs.

The next step is to estimate the significance of these triplet correlations. An adequate measure for the global

presence of triple correlation in our empirical frequency distribution is the relative entropy of its strictly positive approximation p from the distribution of interaction degree 2 which has its same second-order information, that is, its same second-order marginals. The IPFP approximation of degree 2 of the strictly positive approximation p, which maintains its same second-order marginals, is given by p^* in the fourth column of Appendix B.1. The relative entropy of p from p^* is

$$\mathcal{I}(p; p^*) = 0.00805694$$

This number indicates the presence of triple interaction [the relative entropy test is equivalent to the G-test (see Bishop et al. 1989) and approximately equivalent to the χ^2 test]. The Fisher significance test shows that triplet $\{2,4,5\}$ has a significant triple interaction at the 0.05 level. In fact, if p^* denotes the model obtained by performing IPFP on the set

$$\{p_i\colon 1 \leqslant i \leqslant 6\} \cup \{p_{ij}\colon 1 \leqslant i < j \leqslant 6\} \cup \{p_{245}\}$$

it can be calculated that p^* cannot be rejected at the 0.05 level.

If we are interested in a local analysis of the interaction of triplet $\{2,4,5\}$, we can again make use of the second-degree approximation p^* of p. The probability of obtaining the frequency $\frac{2}{930} = 0.00215054$ of configuration $(0,1,0,1,1,0)$ in independent trials, assuming that the chances of success are $p^*(0,1,0,1,1,0) = 0.000379707$, is

$$C_2^{930} \times (0.00037907)^2 \times (1 - 0.00037907)^{928} \tag{11}$$

This probability is low (0.05 is the usual threshold for significance). If our empirical frequencies for triplets were higher ($\geqslant 0.01$), we would feel able to declare that the simultaneous firing of neurons $\{2,4,5\}$ is significant. But we are in a quite different situation since more than half of all configurations are silent. What we can do is to compare the probability obtained in (11) with the maximal possible probability of obtaining configuration $(0,1,0,1,1,0)$ twice in 930 independent trials. This maximal probability is attained for the case that the chances of success are exactly $\frac{2}{930}$ and is given by

$$C_2^{930} \times \left(\frac{2}{930}\right)^2 \times \left(\frac{928}{930}\right)^{928} \tag{12}$$

The ratio of (11) over (12) is approximately 0.25. The analysis of non-silent couples is more rewarding. From the data we see, for instance, that $\{3,5\}$ has a frequency of 0.00752688 and that its subsets (i.e. the single neurons) in this case are obviously also non-silent. In this case we can easily compute the empirical interaction surprise from the data since here there are no silent conclusions involved. We determine the distribution of degree 1 with the same first-order marginals of the strictly positive approximation p (which is simply the product of the first-order marginals of p) and perform the same analysis that was performed for the second-degree interactions. We see that the relative entropy of p from the first-degree approximation is ~ 0.7, and that there are quite a few significantly interacting couples. An interesting couple is $\{3,4\}$, since its interaction surprise is almost zero, which means that these two neurons are nearly independent. Independence, i.e. lack of a positive or negative interaction, is also an important phenomenon in the study of coincident firing in biological neural nets.

As a result of this analysis, we feel that for the study of empirical data we need a more subtle concept of 'interaction degree', which will take into account the information theoretical aspects discussed above. The following definition seems natural:

Definition 6. Let p be a strictly positive approximation. We will say that the interaction character of p at the ε-level is k, if k is the maximal number of simultaneously non-silent with the following property: the relative entropy of p from the IPFP approximation of degree $(k-1)$ of p is greater or equal to ε.

If we are dealing with an empirical distribution with silent configurations, we will have to refer to the interaction character of its strictly positive approximations.

5 Discussion and outlook

After a formal treatment of graphical schemes associated with the energy expansion of the frequency distribution, we investigated the meaning of the coefficients of this expansion in detail. We propose what might be a methodology for analysing interactions indicated by the coefficients of the energy expansion. We are aware that other methodologies with a more subtle significance test are possible. We only claim that IPFP approximations provide the adequate substitute of the degree 1 null hypothesis which has been used so far to estimate the significance of second-order correlations. There are essentially two cases:

(I) The frequency distribution is strictly positive.
(II) Some configurations are silent.

In the first we perform the following steps:

(i) We expand the negative logarithm of the distribution.
(ii) We draw the assembly diagram corresponding to the expansion.
(iii) For every non-zero coefficient of order $k \geqslant 2$ of the energy expansion, we calculate the IPFP approximation of order $k-1$ of the distribution.
(iv) We calculate the interaction character of p.
(v) We perform significance tests using the models obtained through the application of IPFP as null-hypotheses.

In the second case we begin by declaring that for all silent configurations we have 'insufficient data'. Then we proceed as follows:

(i) We approximate the empirical frequency distribution by means of a strictly positive one, obtained by adding a quantity ε to every zero-frequency and renormalizing.
(ii) We expand the negative logarithm of this approximation.

(iii) We proceed as in case (I) and calculate the interactions character of the approximation.
(iv) For every non-silent configuration we look at the interaction threshold [see (8) in the text] . If both numerator and denominator contain ε's, we are unable to continue our analysis. If not, we check whether the frequency of the configuration dominates or is dominated by the threshold. The associated coefficient will be an indicator of correlation or anti-correlation of the corresponding set of neurons.
(v) Of those subsets of neurons which admitted a complete interaction analysis (i.e. the experimental frequency of their simultaneous firing was strictly positive, and the interaction threshold calculated in terms of the strictly positive approximation did not contain ε's in both numerator and denominator), we calculate the significance using the corresponding IPFP approximation.

As already mentioned, our attitude of choosing an equal quantity ε as the frequency of all silent configurations in the strictly positive approximation of the experimental frequency distribution is very naive. Yet, in the analysis of our data [see (iii)], we have been careful enough to not make a statement concerning the interaction of those groups of neurons whose interaction threshold does not tend to a defined limit, as ε tends to 0. There are methods to decide how to choose ε based on a statistical or even a purely Bayesian analysis of the data. This detailed analysis of the optimal ways to approximate the experimental frequency distribution is the object of future work.

Also, the significance tests we proposed can be replaced by more sophisticated ones. The criteria for establishing sound significance tests in this empirical situation are far from being clear. Ours is just the effort to begin a formal discussion on the subject. In our set of data, non-silent configurations with two or more 1's were either silent or had very low frequencies. To declare them as 'significant' seems somewhat risky. The problem of establishing which should be the lowest frequency to be taken seriously is unsolved. There are both strictly statistical and strictly bayesian efforts towards a solution. In the meantime, before a consolidated strategy is generally agreed upon, one might decide to introduce a threshold of relevance in purely absolute terms and consider as candidates for significance testings only 'relevant' interactions. In our example, $\frac{2}{930}$ or even $\frac{3}{930}$ could be the threshold of relevance. In this case we would be forced to admit that our data do not indicate significant interactions of triplets. This type of discussion is still at a speculative level.

In the case of strictly positive frequency distributions, though, we have shown that Markov fields and assembly diagrams provide a theoretically justified, conceptually meaningful and computationally tractable framework for data analysis of coincident firing. The next steps are to perform this type of analysis on larger sets of data on the one hand and to check its efficiency in the case of simulated spiking neurons on the other. We are currently elaborating such a more detailed study of the significance of correlations and fully bayesian treatment of neural interactions.

Acknowledgements. We acknowledge J. G. Taylor, who motivated our study of the invariance of graphs with regard to changes of state. We thank Ulla Mitzdorf and Herbie Glünder for useful suggestions. Friedhelm Schwenker provided mathematical advice. Physiological data were kindly made available by Eilon Vaadia, Department of Physiology, Hadassah Medical School, Hebrew University, Jerusalem, Israel. This work develops part of the post-doctoral project ('Anwendungen von Bayes und Markov Netzwerken Nr. 1544) of the Deutsche Forschungs Gemeinschaft.[4] Additional support was provided by the German Ministry of Science and Technology (BMFT: Verbundprojekt WINA, HvH; Neurobiologie-Program, AA and SG) and the Human Frontier Science Program (AA). We thank the referee for his valuable suggestions.

Appendix

A The iterative proportional fitting procedure (IPFP)

We sketch the iterative procedure we use in order to obtain the minimum relative entropy estimate of interaction degree 2 of a distribution, keeping all its second-order marginals fixed. This is a special case of an approximation algorithm proposed by Kullback et al. (Gokhale and Kullback 1978; Ireland and Kullback 1968; Ku and Kullback 1969; Kullback 1948) for the purpose of hypothesis testing in multidimensional contingency tables. A detailed treatment of this kind of algorithm can be found in Bishop et al. (1989).

Assuming that we have a distribution π on the space Ω of all configurations of 0's and 1's on N neurons and maintaining all notations introduced through the article, we state the problem of constrained probability estimation as follows: Let k be an integer with $1 \leqslant k \leqslant N$ and let $\mathscr{P}^{(k)}$ be the manifold of all distributions with an energy expansion of order k or less. For any fixed set of marginals of order less than or equal to k, find the (unique!) distribution $p^* \in \mathscr{P}^{(k)}$ that minimizes the discrimination information from π defined by

$$\mathscr{I}(p;\pi) = \sum_{\mathbf{x}} p(\mathbf{x}) \ln \frac{p(\mathbf{x})}{\pi(\mathbf{x})}$$

and maintains the fixed set of marginals. That this problem is well posed and solvable is the statement of a fundamental result with a long history (see Bishop et al. 1989, Sects. 10.2 and 10.2-1). We present the iterative solution given in Ku and Kullback (1968, 1969) and Csiszár (1975), concentrating on the case $k = 2$, which is of interest for us. With all second-order marginals fixed, p^* is obtained as the limit of the iterates described by the equations:

$$p^{KS+1}(\mathbf{x}) = \frac{p_{\{1,2\}}(\mathbf{x})}{p^{KS}_{\{1,2\}}(\mathbf{x})} p^{KS}(\mathbf{x})$$

$$p^{KS+2}(\mathbf{x}) = \frac{p_{\{1,3\}}(\mathbf{x})}{p^{KS+1}_{\{1,3\}}(\mathbf{x})} p^{KS+1}(\mathbf{x})$$

$$p^{(KS+1)S}(\mathbf{x}) = \frac{p_{\{N-1,N\}}(\mathbf{x})}{p^{(K+1)S-1}_{\{N-1,N\}}(\mathbf{x})} p^{(K+1)S-1}(\mathbf{x})$$

where $S = N(N-1)/2$, $K = 0, 1, 2, \ldots$ (the cycle index), p^r is the rth probability distribution in the iteration and p^0 is the uniform distribution. The convergence of this sequence of iterates was first rigorously shown by Csiszár (1975).

B Example

B.1 The experimental frequency distribution, the strictly positive approximation, and its IPFP approximation of degree 2, which fixes second-order marginals

[4] Preliminary versions of this work were published in the proceedings of the Workshop 'Wissensverarbeitung mit neuronalen Netzen', 17. Fachtagung für künstliche Intelligenz (see Martignon et al. 1993) and in the proceedings of the Biocybernetics Seminar of the Istituto di Cibernetica del CNR, Naples, September 1993 (see Martignon et al. 1994)

Note: The zeros of the original distribution are approximated by a small ε, which yields the (normalized) distribution p. The distribution p^* denotes the minimum discrimination information estimate of p. After 10 cycles there were no changes at the 8th digit of the value of relative entropy. The low value of the relative entropy is also an indicator of the weakness of triple interaction.

Number of cycles: 10
rel. Entropy $I(p;p^*) = 0.00805694$

Configurations	Original distribution	p	$\rightarrow$ p^*
0 0 0 0 0 0	0.744086	0.744086	0.74569
0 0 0 0 0 1	0.00322581	0.00322581	0.00322581
0 0 0 0 1 0	0.0462366	0.0462366	0.0445135
0 0 0 0 1 1	0	$1*10(-11)$	$1.6*10(-10)$
0 0 0 1 0 0	0.0698925	0.0698925	0.0687409
0 0 0 1 0 1	0	$1*10(-11)$	$1.6*10(-10)$
0 0 0 1 1 0	0.00107527	0.00107527	0.0023463
0 0 0 1 1 1	0	$1*10(-11)$	$4.5*10(-18)$
0 0 1 0 0 0	0.0354839	0.0354839	0.0361773
0 0 1 0 0 1	0	$1*10(-11)$	$1.6*10(-10)$
0 0 1 0 1 0	0.00752688	0.00752688	0.00686222
0 0 1 0 1 1	0	$1*10(-11)$	$2.5*10(-17)$
0 0 1 1 0 0	0.00322581	0.00322581	0.00288421
0 0 1 1 0 1	0	$1*10(-11)$	$6.8*10(-18)$
0 0 1 1 1 0	0	$1*10(-11)$	0.000312817
0 0 1 1 1 1	0	$1*10(-11)$	$6.1*10(-25)$
0 1 0 0 0 0	0.0215054	0.0215054	0.0190213
0 1 0 0 0 1	0	$1*10(-11)$	$1.6*10(-10)$
0 1 0 0 1 0	0	$1*10(-11)$	0.00257471
0 1 0 0 1 1	0	$1*10(-11)$	$1.7*10(-17)$
0 1 0 1 0 0	0.00322581	0.00322581	0.00490597
0 1 0 1 0 1	0	$1*10(-11)$	$2.2*10(-17)$
0 1 0 1 1 0	0.00215054	0.00215054	0.000379707
0 1 0 1 1 1	0	$1*10(-11)$	$1.4*10(-24)$
0 1 1 0 0 0	0	$1*10(-11)$	$9.2*10(-11)$
0 1 1 0 0 1	0	$1*10(-11)$	$7.9*10(-25)$
0 1 1 0 1 0	0	$1*10(-11)$	$3.9*10(-11)$
0 1 1 0 1 1	0	$1*10(-11)$	$2.8*10(-31)$
0 1 1 1 0 0	0	$1*10(-11)$	$2.0*10(-11)$
0 1 1 1 0 1	0	$1*10(-11)$	$9.5*10(-32)$
0 1 1 1 1 0	0	$1*10(-11)$	$5.0*10(-12)$
0 1 1 1 1 1	0	$1*10(-11)$	$1.9*10(-38)$
1 0 0 0 0 0	0.050376	0.0505376	0.0503705
1 0 0 0 0 1	0	$1*10(-11)$	$1.6*10(-10)$
1 0 0 0 1 0	0.00752688	0.00752688	0.00781345
1 0 0 0 1 1	0	$1*10(-11)$	$2.0*10(-17)$
1 0 0 1 0 0	0.00215054	0.00215054	0.00186562
1 0 0 1 0 1	0	$1*10(-11)$	$3.1*10(-18)$
1 0 0 1 1 0	0	$1*10(-11)$	0.000165472
1 0 0 1 1 1	0	$1*10(-11)$	$2.3*10(-25)$
1 0 1 0 0 0	0.00107527	0.00107527	0.000700975
1 0 1 0 0 1	0	$1*10(-11)$	$2.2*10(-18)$
1 0 1 0 1 0	0	$1*10(-11)$	0.000345513
1 0 1 0 1 1	0	$1*10(-11)$	$9.3*10(-25)$
1 0 1 1 0 0	0	$1*10(-11)$	0.0000224534
1 0 1 1 0 1	0	$1*10(-11)$	$3.9*10(-26)$
1 0 1 1 1 0	0	$1*10(-11)$	$6.3*10(-6)$
1 0 1 1 1 1	0	$1*10(-11)$	$9.1*10(-33)$
1 1 0 0 0 0	0	$1*10(-11)$	0.000728399
1 1 0 0 0 1	0	$1*10(-11)$	$4.4*10(-18)$
1 1 0 0 1 0	0.00107527	0.00107527	0.000256207
1 1 0 0 1 1	0	$1*10(-11)$	$1.3*10(-24)$
1 1 0 1 0 0	0	$1*10(-11)$	0.0000754821
1 1 0 1 0 1	0	$1*10(-11)$	$2.5*10(-25)$
1 1 0 1 1 0	0	$1*10(-11)$	0.000015181
1 1 0 1 1 1	0	$1*10(-11)$	$4.1*10(-32)$
1 1 1 0 0 0	0	$1*10(-11)$	$1.0*10(-12)$
1 1 1 0 0 1	0	$1*10(-11)$	$6.4*10(-33)$
1 1 1 0 1 0	0	$1*10(-11)$	$1.1*10(-12)$
1 1 1 0 1 1	0	$1*10(-11)$	$5.9*10(-39)$
1 1 1 1 0 0	0	$1*10(-11)$	$9.0*10(-14)$
1 1 1 1 0 1	0	$1*10(-11)$	$3.0*10(-40)$
1 1 1 1 1 0	0	$1*10(-11)$	$5.8*10(-14)$
1 1 1 1 1 1	0	$1*10(-11)$	$1.6*10(-46)$

B.2 Coefficients of the energy expansion of p (the strictly positive approximation of the experimental frequency distribution)
Note that we have only listed the coefficients up to order 3.

G[1,2,3] = −22.296053	G[1,2] = 18.799545	G[1] = 2.689438
G[1,2,4] = −2.688922	G[1,3] = 0.807069	G[2] = 3.543854
G[1,2,5] = −39.108087	G[1,4] = 0.791802	G[3] = 3.043078
G[1,2,6] = −38.391409	G[1,5] = −0.874148	G[4] = 2.365199
G[1,3,4] = 15.303554	G[1,6] = 16.902425	G[5] = 2.778386
G[1,3,5] = 17.816802	G[2,3] = 18.445905	G[6] = 5.440974
G[1,3,6] = −20.398933	G[2,4] = −0.468079	G[0] = 0.295599
G[1,4,5] = 15.886159	G[2,5] = 18.710598	
G[1,4,6] = −20.383665	G[2,6] = 16.048010	
G[1,5,6] = −18.717715	G[3,4] = 0.032697	
G[2,3,4] = −1.929817	G[3,5] = −1.227788	
G[2,3,5] = −20.261195	G[3,6] = 16.548785	
G[2,3,6] = −38.037769	G[4,5] = 1.396001	
G[2,4,5] = −22.479520	G[4,6] = 17.226665	
G[2,4,6] = −19.123785	G[5,6] = 16.813478	
G[2,5,6] = −38.302462		
G[3,4,5] = 16.645265		
G[3,4,6] = −19.624560		
G[3,5,6] = −18.364075		
G[4,5,6] = −20.987865		

References

Abeles M (1991) Corticonis. Cambridge University Press, Cambridge, UK

Abeles M, Bergman H, Margalit, E, Vaadia E (1993a) Spatiotemporal firing patterns in the frontal cortex of behaving monkeys. Neurophysiol 70:1629–1643

Abeles M, Prut Y, Bergman H, Vaadia E, Aertsen A (1993b) Integration, synchronicity and periodicity. In: Aertsen A (ed) Brain theory: spatio-temporal aspects of brain function. Elsevier, Amsterdam, pp 149–181

Abeles M, Gerstein GL (1988) Detecting spatiotemporal firing patterns among simultaneously recorded single neurons. J Neurophysiol 60:909–924

Aertsen A, Gerstein GL (1991) Dynamic aspects of neuronal cooperativity:fast stimulus-locked modulations of 'effective connectivity'. In: Krüger J (ed) Neuronal cooperativity. Springer, Berlin Heidelberg New York, pp 52–67

Aertsen A, Bonhoeffer T, Krüger J (1987) Coherent activity in neuronal populations:analysis and interpretation. In: Caianiello ER (ed) Physics of cognitive processes. World Scientific Publishing, Singapore, pp 1–34

Aertsen A, Gerstein GL, Habib MK, Palm G (1989) Dynamics of neuronal firing correlation: modulation of 'effective connectivity'. J Neurophysiol 61:900–917

Amari S (1982) Differential geometry of curved exponential families – curvatures and information loss. Ann Stat 10:357–385

Amari S (1985) Differential-geometrical methods in statistics. Springer Lecture Notes in Statistics, Vol 28. Springer, Berlin Heidelberg New York

Amari S (1991) Dualistic geometry of the manifold of higher-order neurons. Neural Networks 4:443–451

Amari S (1994) Information geometry of the EM and em algorithms for neural networks. Tech report, Department of Mathematical Engineering and Information Physics, Faculty of Engineering, University of Tokyo

Amari S, Kurata K, Nagaoka H (1992) Information geometry of Boltzmann machines. IEEE Trans Neural Networks 3:260–271

Amari S, SunHan T (1989) Statistical inference under multiterminal rate restrictions: A differential geometric approach. IEEE Trans Inf Theory 35:217–227

Besag J (1974) Spatial interaction and the statistical analysis of lattice systems. J Stat Soc B 34:75–83

Bishop Y, Fienberg S, Holland P (1989) Discrete multivariate analysis, 10th edn. MIT Press, Cambridge, Mass.

Caianiello E (1975) Synthesis of boolean nets and time behaviour of a general mathematical neuron. Biol Cybern 18:111

Caianiello E (1986) Neuronic equations revisited and completely solved. In: Palm G, Aertsen A (eds) Brain theory. Springer, Berlin Heidelberg New York

Cover TM, Thomas JA (1991) Elements of information theory. Wiley, New York

Csiszár I (1975) I-divergence geometry of probability distributions and minimization problems. Ann Probab 3:146–158

Deming WE, Stephan FF (1940) On a least squares adjustment of a sampled frequency table when the expected marginals totals are known. Ann Math Stat 11:427–444

Gerstein G, Aertsen A (1985) Representation of cooperative firing activity among simultaneously recorded neurons. J Neurophysiol 54:1513–1527

Gerstein GL, Bedenbaugh P, Aertsen A (1989) Neuronal assemblies. IEEE Trans Biomed Eng 36:4–14

Gokhale DV, Kullback S (1978) The information in contingency tables. Dekker, New York

Griffeath D (1976) Introdction to random fields. Appendix in Knapp A, Kemeny J, Snell J (eds) Denumerable Markov chains. Springer, Berlin Heidelberg New York

Grimmett GR (1973) A theorem about random fields. Bull Lond Math Soc 5:81–84

Grün S, Aertsen A, Abeles M, Gerstein G, Palm G (1994) Behavior-related neuron group activity in the cortex. Proc 17th Ann Meeting of the European Neurosci Association. Oxford University Press In:ENA, Oxford

Grün S, (1994) Aertsen A, Abeles M, Gerstein G, Palm G. On the significance of coincident firing in neuron group activity. In: Elsner N, Breer H (eds) Sensory transduction. Stuttgart, Thieme, p 558

Hammersley JM, Clifford P (1968) Markov fields on finite graphs and lattices. University of California, Berkeley

Hebb D (1949) The organization of behavior, a neurophysiological theory. Wiley, New York

Hinton GE, Sejnowski TJ (1986) Learning and relearning in Boltzmann machines. (Parallel distributed processing, Vol 1) MIT Press, Cambridge, Mass pp 282–317

Ireland CT, Kullback SS (1968) Contingency tables with given marginals. Biometrika 55:179–188

Ku HH, Kullback S (1968) Interaction in multidimensional contingency tables: An information theoretic approach. J Res NBS Math 72B:159–199

Ku HH, Kullback S (1969) Approximating discrete probability distributions. IEEE Trans Inf Theory IT-15:444–447

Kullback S (1968) Information theory and statistics. Dover, New York

Martignon L, Laskey BK (1995) Statistical inference methods for classifying higher order neural correlations. In: Hermann H (ed) Proceedings of the International Workshop on Supercomputers and the Brain. World Scientific, Singapore (in press)

Martignon L, Hasseln H von, Palm G (1993) Modelling stochastic networks: from data to the connectivity structure. (Informatik aktuell, Subreihe Künstliche Intelligenz) Gesamtdarstellung des Workshops auf der KI-Jahrestagung, Berlin 1993. Springer, Berlin Heidelberg New York pp 50–58

Martignon L, Hasseln H von, Grün S, Palm G (1994) Modelling the interaction in a set of neurons implicit in their frequency distribution: a possible approach to neural assemblies. In: Taddei C et al. (eds) Collected lectures of the seminar on biocybernetics. Ist di Cibernetica, Naples. Rosenberg-Sellier, Torino

Miller JW, Goodman RM (1993) Probability estimation from a database. In: Cowan JD, Hanson SJ, Giles CL (eds) Advances in Neural Information Processing Systems, Vol 5. Morgan Kaufman, San Mateo, pp 531–538

Palm G (1981) Evidence, information and surprise. Biol Cybern 42:57–68

Palm G, Aertsen A, Gerstein G (1988) On the significance of correlations among neuronal spike trains. Biol Cybern 59: 1–11

Pinkas G (1991) Energy minimization and the satisfiability of propositional logic. In: Sejnowsky T, Touretzky D, Ellman H, Hinton G, (eds) Proc. of the 1990 Connectionist Models Summer School. Morgan Kaufman, San Mateo, pp 23–31

Vaadia E, Aertsen A (1992) Coding and computing in the cortex:single neuron activity and cooperative phenomena. In: Aertsen A, Braitenberg V (eds) Information processing in the cortex. Springer, Berlin Heidelberg New York.

Vaadia E, Bergman H, Abeles M (1989) Neuronal activities related to higher brain functions – theoretical and experimental implications. IEEE Trans Biomed Eng BME-36. 25–35

Vaadia E, Ahissar E, Bergman H, Lavner Y (1991) Correlated activity of neurons: a neural code for higher brain functions? In: Krüger J (ed) Neuronal cooperativity. Springer, Berlin Heidelberg New York, pp 249–279

Vaadia E, Haalman I, Abeles M, Bergman H, Prut Y, Slovin H, Aertsen A (1995) Dynamics of neuronal interactions in the monkey cortex to behavioral events. Nature 373:515–518