LETTER Communicated by Ad Aertsen

Neural Coding: Higher-Order Temporal Patterns in the Neurostatistics of Cell Assemblies

Laura Martignon
Max Planck Institute for Human Development, Lentzeallee 94, 14195 Berlin, Germany

Gustavo Deco
Siemens AG, Corporate Technology ZT, 81739 Munich, Germany

Kathryn Laskey
Department of Systems Engineering, George Mason University, Fairfax, VA 22030, U.S.A.

Mathew Diamond
Cognitive Neuroscience Sector, International School of Advanced Studies, 34014 Trieste, Italy

Winrich Freiwald
Center for Cognitive Science, University of Bremen, 28359 Bremen, Germany

Eilon Vaadia
Department of Physiology, Hebrew University of Jerusalem, Jerusalem 91010, Israel

Recent advances in the technology of multiunit recordings make it possible to test Hebb's hypothesis that neurons do not function in isolation but are organized in assemblies. This has created the need for statistical approaches to detecting the presence of spatiotemporal patterns of more than two neurons in neuron spike train data. We mention three possible measures for the presence of higher-order patterns of neural activation—coef ficients of log-linear models, connected cumulants, and redundancies—and present arguments in favor of the coefficients of log-linear models. We present test statistics for detecting the presence of higher-order interactions in spike train data by parameterizing these interactions in terms of coefficients of log-linear models. We also present a Bayesian approach for inferring the existence or absence of interactions and estimating their strength. The two methods, the frequentist and the Bayesian one, are shown to be consistent in the sense that interactions that are detected by either method also tend to be detected by the other. A heuristic for the analysis of temporal patterns is also proposed. Finally, a Bayesian test is presented that establishes stochastic differences between recorded segments of data. The methods are applied to experimental data and synthetic data drawn from our statistical models. Our experimental data are drawn from multiunit recordings in the prefrontal cortex of

behaving monkeys, the somatosensory cortex of anesthetized rats, and multiunit recordings in the visual cortex of behaving monkeys.

1 Introduction

Hebb (1949) conjectured that information processing in the brain is achieved through the collective action of groups of neurons, which he called cell assemblies. One of the assumptions on which he based his argument was his other famous hypothesis: that excitatory synapses are strengthened when the involved neurons are frequently active in synchrony. Evidence in support of this so-called Hebbian learning hypothesis has been provided by a variety of experimental findings over the last three decades (e.g., Rauschecker, 1991). Evidence for collective phenomena confirming the cell assembly hypothesis has only recently begun to emerge as a result of the progress achieved by multiunit recording technology. Hebb's followers were left with a twofold challenge: to provide an unambiguous definition of cell assemblies and to conceive and carry out the experiments that demonstrate their existence.

Cell assemblies have been defined in terms of both anatomy and shared function. One persistent approach characterizes the cell assembly by near simultaneity or some other specific timing relation in the firing of the involved neurons. 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 (Abeles, 1991; Abeles, Bergman, Margalit & Vaadia, 1993). Thus syn-firing and other specific temporal relationships between active neurons have been posited as mechanisms by which the brain codes information (Gray, König, Engel, & Singer, 1989; Singer, 1994; Abeles & Gerstein, 1988; Abeles, Prut, Bergman, & Vaadia, 1994; Prut et al., 1998).

In pursuit of experimental evidence for cell assembly activity in the brain, physiologists thus seek to analyze the activation of many separate neurons simultaneously, preferably in awake, behaving animals. These multineuron activities are then inspected for possible signs of interactions among 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 is based on the use of cross-correlation techniques, usually applied to the activity of pairs (sometimes triplets) of neurons recorded under 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. Application of these measures has revealed interesting instances of time- and context-dependent synchronization dynamics in different cortical areas. For interactions among more than two neurons, Gerstein, Perkel, and Dayhoff (1985) devised the so-called gravity method. Their method views neurons as particles that attract each other and measures simultaneously the attractions

between all possible pairs of neurons in the set. The method, although it allows one to look at several neurons globally, does not account for nonlinear synapses or for synchronization of more than two neurons. Recent investigations have focused on the detection of individual instances of synchronized activity called unitary events (Grün, 1996; Grün, Aertsen, Vaadia, & Riehle, 1995; Riehle, Grün, Diesman, & Aertsen, 1997) between groups of two and more neurons. Of special interest are patterns involving three or more neurons, which cannot be described in terms of pair correlations. Such patterns are genuine higher order phenomena. The method developed by Prut et al. (1998) for detecting precise firing sequences of up to three units adopts this view, subtracting from three-way correlations all possible pair correlations. The models reported in this article were developed for the purpose of describing and detecting correlations of any order in a unified way.

In the data we analyze, the spiking events (in the 1 ms range) are encoded as sequences of 0s and 1s, and the activity of the whole group is described as a sequence of binary configurations. This article presents a family of statistical models for analyzing such data. In our models, the parameters represent spatiotemporal firing patterns. We generalize the spatial correlation models developed by Martignon, von Hasseln, Grün, Aertsen, and Palm (1995), to include a temporal dimension. We develop statistical tests for detecting the presence of a genuine order-n correlation and distinguishing it from an artifact that can be explained by lower-order interactions. The tests compare observed firing frequencies on the involved neurons with frequencies predicted by a distribution that maximizes entropy among all distributions consistent with observed information on synchrony of lower order. We also present a Bayesian approach in which hypotheses about interactions on a large number of subsets can be compared simultaneously. Furthermore, we introduce a Bayesian test to establish essential differences (i.e., differences that are not to be attributed to noise) between different phases of recording (e.g., prestimulus phase versus stimulus phase).

We compare our log-linear models with two other candidate approaches to measuring the presence of higher-order correlations. One of these candidates, drawn from statistical physics, is the connected cumulant. The other, drawn from information theory, is mutual information or redundancy. We argue that the coefficients of log-linear models, or effects, provide a more natural measure of higher-order phenomena than either of these alternate approaches.

We present the results of analyzing synthetic data generated from our models to test the performance of our statistical methods on data of known distribution. Results are presented from applying the models to multiunit recordings obtained by Vaadia from the frontal cortex of monkeys, to multiunit recordings obtained by Diamond from the somatosensory cortex of rats, and to multiunit recordings obtained by Freiwald from the visual cortex of behaving monkeys.

2 Measures for Higher-Order Synchronization

2.1 Effects of Log-Linear Models. The term *spatial correlation* has been used to denote synchronous firing of a group of neurons, while the term *temporal correlation* has been used to indicate chains of firing events at specific temporal intervals. Terms like *couple* and *triplet* have been used to denote spatiotemporal patterns of two or three neurons (Abeles et al., 1993; Grün, 1996) firing simultaneously or in sequence. Establishing the presence of such patterns is not straightforward. For example, three neurons may fire together more often than expected by chance[1] without exhibiting an authentic third-order interaction. For example, if a neuron participates in two couples, such that each pair fires together more often than by chance, then the three involved neurons will fire together more often than the independence hypothesis would predict. This is not, however, a genuine third-order phenomenon. Authentic triplets and, in general, authentic nth-order correlations must therefore be distinguished from correlations that can be explained in terms of lower-order correlations. We introduce log-linear models for representing firing frequencies on a set of neurons and show that nonzero coefficients or effects of these log-linear models are a natural measure for synchronous firing. We argue that the effects of log-linear models are superior to other candidate approaches drawn from statistical physics and information theory. In this section we consider only models for synchronous firing. Generalization to temporal and spatiotemporal effects is treated in section 3.

Consider a set of n neurons. Each neuron is modeled as a binary unit that can take on one of two states: 1 (firing) or 0 (silent). The state of the n units is represented by the vector $\underline{x} = (x_1, \ldots, x_n)$, where each x_i can take on the value zero or one. There are 2^n possible states for the n neurons. If all neurons fire independent of each other, the probability of configuration $\underline{x}$ is given by

$$p(x_1, \ldots, x_n) = p(x_1) \cdots p(x_n). \tag{2.1}$$

Methods for detecting correlations look for departures from this model of independence. Following a well-established tradition, we model neurons as Bernoulli random variables. Two neurons are said to be correlated if they do not fire independently. A correlation between two binary neurons, labeled 1 and 2, is expressed mathematically as

$$p(x_1, x_2) \neq p(x_1)p(x_2). \tag{2.2}$$

Extending this idea to larger sets of neurons introduces complications. It is not sufficient simply to compare the joint probability $p(x_1, x_2, x_3)$ with

[1] That is, more often than predicted by the hypothesis of independence.

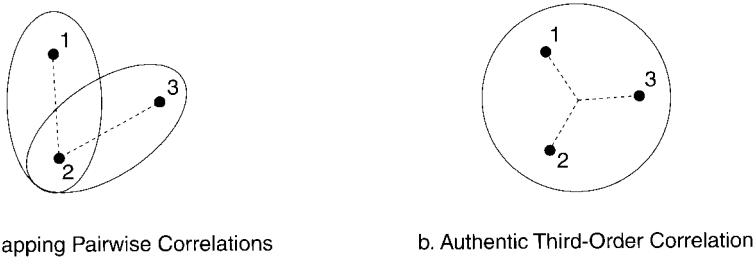

Figure 1: Overlapping doublets and authentic triplet.

the product $p(x_1)p(x_2)p(x_3)$ and declare the existence of a triplet when the two are not equal. This would confuse authentic triplets with overlapping doublets or combinations of doublets and singlets. Thus, neurons 1, 2, and 3 may fire together more often than the independence model, equation 2.1, would predict because neurons 1 and 2, and neurons 2 and 3, are each involved in a binary interaction (see Figure 1).

Equation 2.2 expresses in mathematical form the idea that two neurons are correlated if the joint probability distribution for their states cannot be determined from the two individual probability distributions $p(x_1)$ and $p(x_2)$. Now consider a set of three neurons. The probability distributions for the three pair configurations are $p(x_1, x_2)$, $p(x_2, x_3)$, and $p(x_1, x_3)$.[2] A genuine third-order interaction among the three neurons would occur if it were impossible to determine the joint distribution for the three-neuron configuration using only information on these pair distributions. A canonical way to construct a joint distribution on three neurons from the pair distributions is to maximize entropy subject to the constraint that the two-neuron marginal distributions are given by $p(x_1, x_2)$, $p(x_2, x_3)$, and $p(x_1, x_3)$. This is the distribution that adds the least information beyond the information contained in the pair distributions. It is well known[3] that the joint distribution thus obtained can be written as

$$p(x_1,x_2,x_3) = \exp\{\theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3 + \theta_{12} x_1 x_2 + \theta_{13} x_1 x_3 + \theta_{23} x_2 x_3\}, \quad (2.3)$$

where the θs are real-valued parameters and θ_0 is determined from the other θs and the constraint that the probabilities of all configurations sum to 1. In this model, there is a parameter for each individual neuron and each pair of neurons. Each of these parameters, which in the statistics literature are called *effects*, can be thought of as measuring the tendency to be active of the neuron(s) with which it is associated. Increasing the value of θ_i without

[2] Note that these pairwise distributions also determine the single-neuron firing frequencies $p(x_1)$, $p(x_2)$, and $p(x_3)$, which can be expressed as their linear combinations.

[3] See, for example, Good (1963) or Bishop, Fienberg, and Holland (1975).

changing any other θs increases the probability that neuron i is in its "on" state. Increasing the value of θ_{ij} without changing any other θs increases the probability of simultaneous firing of neurons i and j. It is instructive to note that there is a second-order correlation between neurons i and j in the sense of equation 2.2 precisely when $\theta_{ij} \neq 0$.

Not all joint probability distributions on three neurons can be written in the form of equation 2.3. A general joint distribution for three neurons can be expressed by including one additional parameter:[4]

$$p(x_1, x_2, x_3) = \exp\{\theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3 + \theta_{12} x_1 x_2 + \theta_{13} x_1 x_3 + \theta_{23} x_2 x_3 + \theta_{123} x_1 x_2 x_3\}. \quad (2.4)$$

Holding the other parameters fixed and increasing the value of θ_{123} increases the probability that all three neurons fire simultaneously. Equation 2.3 corresponds to the special case of equation 2.4 in which θ_{123} is equal to zero, just as independence corresponds to the special case of equation 2.3 in which all second-order terms θ_{ij} are equal to zero. It seems natural, then, to define a genuine order-3 correlation as a joint probability distribution that cannot be expressed in the form of equation 2.3 because $\theta_{123} \neq 0$. As will be shown in section 3, this idea can be extended naturally to larger sets of neurons and temporal patterns.

Parameterizations of the general form equations 2.3 and 2.4 are called log-linear models because the logarithm of the probabilities can be expressed as a linear sum of functions of the configuration values. They correspond to the coefficients of the energy expansions of generalized Boltzmann machines in statistical mechanics.The parameters of the log-linear model are called effects. A joint probability distribution on n neurons is represented as a log-linear model in which effects are associated with subsets of neurons. A genuine nth-order correlation on a set A of n neurons is defined as the presence of a nonzero effect θ_A associate d with the set A in the log-linear model representing the joint distribution of configuration frequencies.

The information-theoretic justification for our proposed definition of higher-order interaction is theoretically satisfying. But if the idea is to have practical utility, we require a method for measuring the presence of genuine correlations and distinguishing them from noise due to finite sample size. Fortunately, there is a large statistical literature on estimation and hypothesis testing using log-linear models. In the following sections, we discuss ways to detect and quantify the presence of higher-order correlations. We also apply the methods to real and synthetic data.

[4] This can be seen by noting that $\log p(x_1, x_2, x_3)$ can be expressed as a set of linear equations relating the configuration probabilities to the θs. These equations give a unique set of θs for any given probability distribution.

2.2 Other Candidate Measures of Higher-Order Interactions. Before closing this section, we mention two further approaches to measuring the existence of higher-order correlations in a set of binary neurons. Connected cumulants are measures of higher order correlations used in statistical mechanics as well as in nonlinear system theory. They were introduced by Aertsen (1992) as a candidate parameter for spatiotemporal patterns of neural activation. Their introduction was motivated by their role in the Wiener-Volterra theory of nonlinear systems.

For three neurons, for instance, the connected cumulant is defined by

$$C_{123} = \langle x_1 x_2 x_3 \rangle - \langle x_1 \rangle \langle x_2 x_3 \rangle - \langle x_2 \rangle \langle x_1 x_3 \rangle - \langle x_3 \rangle \langle x_1 x_2 \rangle + 2 \langle x_1 \rangle \langle x_2 \rangle \langle x_3 \rangle . \tag{2.5}$$

In these expressions, the symbol $\langle \cdot \rangle$ stands for expectation. Thus, the symbol $\langle x_i \rangle$ denotes the expected firing frequency of the ith neuron. This formula can easily be generalized to any number of neurons.

Another approach to measuring higher-order correlations uses the concept of mutual information drawn from information theory. This approach was originally suggested by Grün (1996). For three neurons, the parameter defined by

$$I_3(x_1, x_2, x_3) = H(x_1) + H(x_2) + H(x_3) - H(x_1, x_2) - H(x_1, x_3) - H(x_2, x_3) + H(x_1, x_2, x_3), \tag{2.6}$$

where H denotes entropy (see Cover & Thomas, 1991, for details), is a measure of higher-order correlation. This formula can easily be generalized to any number n of neurons, for $n > 1$.

Defining second-order interactions as either $C_{12} = \langle x_1, x_2 \rangle - \langle x_1 \rangle \langle x_2 \rangle$ or $I_2 = H(x_1, x_2) - H(x_1) - H(x_2)$ is equivalent to the above definition as a nonzero effect of the appropriate log-linear model. However, extending these measures to correlations of order 3 or higher does not yield equivalent characterizations of the set of distributions exhibiting no interaction of the given order. A detailed discussion of both measures as well as a presentation of statistical methods for their detection is given in Deco, Martignon, and Laskey (1998).

In this article we argue that an interaction among a set of neurons should be modeled as a nonzero effect in a log-linear model for the joint distribution of spatiotemporal configurations of the involved neurons. There are several reasons for this choice. First, log-linear models are natural extensions of the Boltzmann machine. As we already mentioned, defining interactions as nonzero effects of log-linear models has a compelling information-theoretic justification. Second, methods exist for estimating effects and testing hypotheses about the degree of interaction using data from multiunit recordings. Third, the models can be used to examine behavioral corre-

lates of neuronal activation patterns by estimating, comparing, and testing hypotheses about similarities and differences in log-linear models for data segments recorded at different times under differing environmental stimuli.

3 A Mathematical Model for Spatiotemporal Firing Patterns

The previous section described several approaches to generalizing the concept of pairwise correlations to larger numbers of neurons. Parameters of log-linear models were proposed as a measure of higher-order correlations and justified by their ability to distinguish true higher-order correlations from apparent correlations that can be explained in terms of lower-order correlations. For clarity of exposition, the discussion was limited to synchronous firing on groups of two and three neurons. In this section, we generalize the approach to arbitrary numbers of neurons and temporal correlations.

Consider a set Λ of n binary neurons and denote by p the probability distribution on sequences of binary configurations of Λ. Assume that the sequence of configurations $\underline{x}_t = (x_{1t}, \ldots, x_{nt})$ of n neurons forms a Markov chain of order r. Let δ be the time step, and denote the conditional distribution for $\underline{x}_t$ given previous configurations by $p(\underline{x}_t \mid \underline{x}_{(t-\delta)}, \underline{x}_{((t-2\delta)}, \ldots, \underline{x}_{(t-r\delta)})$. We assume that all transition probabilities are strictly positive and expand the logarithm of the conditional distribution as

$$p(\underline{x}_t \mid \underline{x}_{(t-\delta)}, \underline{x}_{(t-2\delta)}, \ldots, \underline{x}_{(t-r\delta)}) = \exp\left\{\theta_0 + \sum_{A \in \Xi} \theta_A X_A\right\}. \tag{3.1}$$

In this expression, each A is a subset of pairs of nonnegative indices (i, m), where at least one value of m is equal to zero. The set A represents a possible spatiotemporal pattern. The index i indicates a neuron, and the index m indicates a time delay in units of δ. The variable $\underline{x}_A = \prod_{1 \le j \le k} x_{(i_j, m_j)}$ is equal to 1 in the event that all neurons represented by indices in A are active at the indicated time delays, and is equal to zero otherwise. For example, if there are 10 instances in which neuron 4 fires 30 ms after neuron 2, and if the time step is 10 ms, then $\underline{x}_A = 1$ for $A = \{(2, 0), (4, 3)\}$. The set Ξ for which $\theta_A \neq 0$ is called the *interaction structure* for the distribution p. The parameter θ_A is called the *interaction strength* for the interaction on subset A. Clearly, $\theta_A = 0$ means $A \notin \Xi$ and is taken to indicate absence of an order-$|A|$ interaction among neurons in A. We denote the structure-specific vector of nonzero interaction strengths by $\underline{\theta}_\Xi$.

Definition 1. *We say that neurons* $\{i_1, i_2, \ldots, i_k\}$ *exhibit a* spatiotemporal pattern *if there is a set of time intervals* $m_1\delta, m_2\delta, \ldots, m_k\delta$ *with* $r \ge m_i \ge 0$ *and at least one* $m_i = 0$, *such that* $\theta_A \neq 0$ *in (1), where* $A = \{(i_1, m_1), \ldots, (i_k, m_k)\}$.

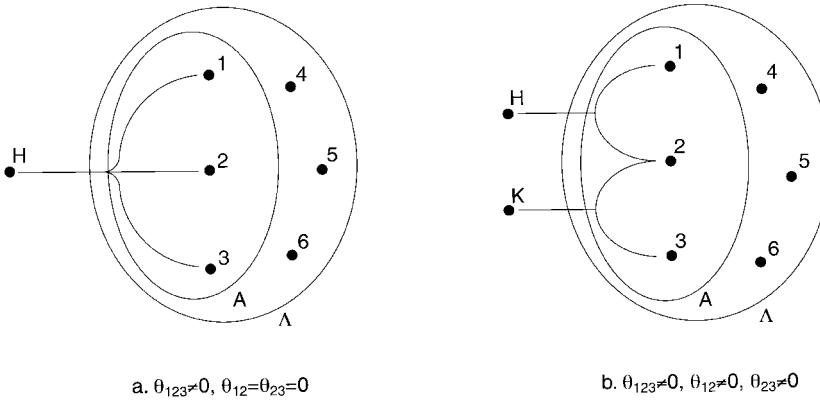

Figure 2: Nonzero θ_{123} due to common hidden inputs H and/or K.

Definition 2. *A subset* $\{i_1, i_2, \ldots, i_k\}$ *of neurons exhibits a* synchronization *or* spatial correlation *if* $\theta_A \neq 0$ *for* $A = \{(i_1, 0), \ldots, (i_k, 0)\}$.

In the case of absence of any temporal dependencies, the configurations at different times are independent, and we drop the time index:

$$p(\underline{x}) = \exp\left\{\theta_\phi + \sum_{A\in\Xi} \theta_A X_A\right\} \tag{3.2}$$

where each A is a nonempty subset of Λ and $\underline{x}_A = \prod_{i\in A} x_i$.

Of course, we expect temporal correlation of some kind to be present, one such example being the refractory period after firing. Nevertheless, equation 3.2 may be adequate in cases of weak temporal correlation and generalizes naturally to the situation of temporal correlation. Increasing the bin width can result in temporal correlations of short time intervals manifesting as synchronization.

It is clear that equations 2.3 and 2.4 are special cases of equation 3.2 for the case of three neurons. The interaction structure Ξ for equation 2.4 consists of all nonempty subsets of neurons. For equation 2.3, the interaction structure consists of all single-neuron and two-neuron subsets.

Although the models, equations 3.1 and 3.2, are statistical and not physiological, one would naturally expect synaptic connection between two neurons to manifest as nonzero θ_A for the set A composed of these neurons with the appropriate delays. One example leading to nonzero θ_A in equation 3.2 would be simultaneous activation of the neurons in A due to common input (see Figure 2a). Another example is simultaneous activation of overlapping subsets covering A, each by a different unobserved input (see Figure 2b).

An attractive feature of our models is that the two cases of Figure 2 can be distinguished when the neurons in A exhibit no true lower-order interactions. Suppose a model such as depicted in Figure 2 is constructed for all neurons, both observed and unobserved. The model of Figure 2a corresponds to a log-linear expansion with only individual neuron coefficients and the fourth-order effect θ_{H123}, where H is a hidden unit, as in Figure 2. The model of Figure 2b contains individual neuron effects and the triplet effects θ_{H12} and θ_{K23}, where K is another hidden unit different from H. When the probabilities are summed over the unobserved neurons to obtain the distribution over the three observed neurons, both models will contain third-order coefficients θ_{123}. However, the model of Figure 2a will contain no nonzero second-order coefficients, and the model of Figure 2b will contain nonzero coefficients of all orders up to and including order 3. Thus log-linear models have the desirable feature that the existence of a nonzero θ_A coupled with $\theta_B = 0$ for $B \subset A$ indicates a correlation among the neurons in A possibly involving unobserved neurons. On the other hand, a nonzero θ_A together with $\theta_B \neq 0$ for $B \subset A$ may indicate an interaction that can be explained away by interactions involving subsets of A and unobserved neurons.

4 Estimation and Testing Based on Maximum Likelihood Theory

4.1 Construction of a Test Statistic for the Presence of Interactions. We are interested in the problem of detecting nonzero θ_A for subsets A of neurons in a set of n neurons whose firing frequencies are governed by model 3.1 or 3.2. That is, for a given set A, we are interested in testing which of the two hypotheses,

$$H_0: \quad \theta_A = 0$$
$$H_1: \quad \theta_A \neq 0$$

is the case. We wish to declare the existence of a higher-order interaction only when the observed correlations are unlikely to be a chance result of a model in which $\theta_A = 0$.

We apply the theory of statistical hypothesis testing (Neyman & Pearson, 1928) to construct a test statistic whose distribution is known under the null hypothesis. Then we choose a critical range of values of the test statistic that are highly unlikely under the null hypothesis but are likely to be observed under plausible deviations from the null hypothesis. If the observed value of the test statistic falls into the critical range, we reject the null hypothesis in favor of the alternative hypothesis.

Consider the problem of testing $\theta_A \neq 0$ against $\theta_A = 0$ in the log-linear model, equation 3.2 for synchronous firing in observations with no temporal correlation. A reasonable test statistic is the likelihood ratio statistic, denoted by G^2. To construct the G^2 test statistic, we first obtain the maximum

likelihood estimate of the parameters $\theta_{\underline{=}}$ under the null and alternative hypotheses. We construct our test statistic by conditioning on the silence of neurons outside A. We compute estimates $p_1(\underline{x})$ and $p_0(\underline{x})$ for the vector of configuration probabilities under the alternative (H_1) and null (H_0) hypotheses, respectively. Using these probability estimates, we compute the likelihood ratio test statistic,

$$G^2 = 2N \sum_{\underline{x}} p_1(\underline{x}) \left(\ln p_1(\underline{x}) - \ln p_0(\underline{x}) \right), \tag{4.1}$$

where N is the sample size for the test (the number of configurations of 0s and 1s in which neurons outside of A are silent). As the sample size for the test tends to infinity, the distribution for this test statistic under H_0 tends to a chi-squared distribution, where the number of degrees of freedom is the difference in the number of parameters in the two models being compared. In this case, the larger model has a single extra parameter, so the reference distribution has one degree of freedom. The interpretation of test results is as follows. If the value of the test statistic is highly unusual for a chi-squared distribution with one degree of freedom, this casts doubt on the null hypothesis H_0 and is therefore regarded as evidence that the additional parameter θ_A is nonzero.

4.2 A Caveat on Statistical Power. An explicit expression for θ_A in terms of the probabilities of configurations is given by

$$\theta_A = \sum_{B \subset A} (-1)^{|A-B|} \ln p(\chi_B) \tag{4.2}$$

where χ_B represents the configuration of 1s on all elements of B and 0s elsewhere. Thus, θ_A is fully determined by the probabilities of configurations that have 0s outside A. Conditioning on the silence of neurons outside A and thus discarding observations for which neurons outside A are active simplifies estimation of θ_A and construction of a test statistic for H_1 against H_0. The distribution of the test statistic does not depend on whether there are nonzero coefficients associated with sets containing neurons outside A.

Ignoring all observations for which neurons outside A are firing, as the above test does, reduces statistical power as compared with a test based on all the observations.[5] This is not a major difficulty when overall firing frequencies are low and the number of neurons is small, but can make the approach infeasible for large numbers of neurons. An alternative approach would be to construct a test using all the data. To construct such a test,

[5] A statistical hypothesis test has a high power if the probability of correctly rejecting a false null hypothesis is high and high significance if the probability of incorrectly rejecting a true null hypothesis is low.

it is necessary to estimate parameters for an interaction structure over all neurons, which allows for $\theta_B \neq 0$ for all sets B for which it is reasonable to expect that interactions may be present (typically, all subsets of neurons less than or equal to a given order). Unfortunately, neither of these approaches scales to large numbers of neurons. As the number of neurons grows, the probability that some neuron outside the set A is firing also grows, increasing the fraction of the data that needs to be discarded for the first approach. The second approach suffers from the problem that the number of simultaneously estimated parameters grows as an order }A{ polynomial in the number of neurons, resulting in unstable parameter estimates when the number of neurons is large.

In contrast, as we will argue, the natural Occam's razor (Smith and Spiegelhalter, 1980) associated with the Bayesian mixture model that will be introduced in section 5 permits that method to be applied usefully even with very large sets of neurons. Although we do regard statistical significance of a coefficient as a useful indicator of existence of correlation of the associated neurons, the Bayesian approach described below makes use of all the data and is well suited to problems involving large numbers of parameters.

4.3 Estimating Parameters Under Null and Alternative Hypotheses. The maximum likelihood estimate of $\underline{\theta}_{\Xi}$ under the alternative hypothesis H_1: $\Xi = 2^A$ is easily obtained in closed form by setting the frequency distribution $p_1(\underline{x})$ over configurations of neurons in the set A to the sample frequencies and solving for $\underline{\theta}_{2^A}$. The maximum likelihood estimate for $\underline{\theta}_{\Xi}$ under the null hypothesis H_0: $\Xi = 2^A - A$ may be obtained by a standard likelihood maximization procedure such as iterative proportional fitting (IPF) (see, for instance, Bishop et al., 1975). As an alternative to IPF, we have developed a simple one-dimensional optimization procedure, described in the appendix, which we call the constrained perturbation procedure (CPP). We have found computation time for CPP to be much faster than for IPF. Once the estimate $\underline{\theta}_{\Xi}$ has been obtained, one can solve for the estimated configuration probabilities $p_0(\underline{x})$ for the null hypothesis.

5 Statistical Tests Based on Bayesian Model Comparison

5.1 Detecting Synchronizations Using the Bayesian Approach. An alternative approach to the Neyman and Pearson hypothesis testing described in section 4 is Bayesian estimation. Standard frequentist methods may run into difficulties on high-dimensional problems such as the one treated in this article. In contrast, high-dimensional problems pose no essential difficulty in the Bayesian approach as long as prior distributions are chosen appropriately. For this reason, and also because of advances in computational methods, the Bayesian approach has been gaining in favor as a theoretically sound and practical way to treat high-dimensional problems.

In this section we focus on the problem of determining synchronization structure in the absence of temporal correlation, as represented by equation 3.2. A heuristic treatment of temporal correlations is discussed in section 5.3, and a full treatment of spatiotemporal interactions will appear in a future article.

To apply the Bayesian approach, we assign a prior probability to each interaction structure Ξ and a continuous prior distribution for the nonzero parameters $\underline{\theta}_{\Xi}$. A sample of observations from multiunit recordings is used to update the prior distribution and obtain a posterior distribution. The posterior distribution also assigns a probability to each interaction structure and a continuous probability distribution to the nonzero θ_A given the interaction structure.

The prior distribution can be thought of as a bias that keeps the model from jumping to inappropriately strong conclusions. In our problem, it is natural to assign a prior distribution that introduces a bias toward models in which most θ_A are zero. We use a prior distribution in which all θ_A are independent of each other and are probably equal to zero. We then assume a continuous probability density for θ_A conditional on its being nonzero, in which values very different from zero are less likely than values near zero. Because of the bias introduced by the prior distribution, a large posterior probability that $\theta_A \neq 0$ indicates reasonably strong observational evidence for a nonzero parameter. A formal specification for the prior distribution is given in the appendix.

We estimate posterior distributions for both structure and parameters in a unified Bayesian framework. A Markov chain Monte Carlo model composition algorithm (MC^3) is used to search over structures (Madigan & York, 1993). Laplace's method (Tierney & Kadane, 1986; Kass & Raftery, 1995) is used to estimate the posterior probability of structures. For a given interaction structure, θ_A is estimated by the mode of the posterior distribution for that structure, and its standard deviation is estimated by the appropriate diagonal element of the inverse Fisher information matrix. The posterior probability that $\theta_A \neq 0$ is the sum of the probability of all interaction structures in which $\theta_A \neq 0$. An overall estimate for θ_A given that it is nonzero is obtained as a weighted average of structure-specific estimates, where each structure is weighted by its posterior probability and the result is normalized over structures in which $\theta_A \neq 0$. Formulas for the estimates may be found in the appendix.

5.2 Detecting Changes in Synchronization Patterns. The Bayesian approach can also be used to infer whether there are systematic differences in firing rates and interactions during different time segments. A fully Bayesian treatment of this problem would embed the model 3.3 in a larger model, including all time segments under consideration, and would explicitly model hypotheses in which the θ_A are identical or different in different time segments. We instead apply a simpler approach, in which we choose a single

interaction structure Ξ rich enough to provide a reasonable fit to both segments of data. We compute separate posterior distributions $\underline{\theta}_{\Xi 1}$ and $\underline{\theta}_{\Xi 2}$ under the hypothesis that the two segments are governed by different distributions and a combined posterior distribution for $\underline{\theta}_{\Xi c}$ that treats both segments as a single data set. Using a method described in detail in the appendix, we use these estimates to compute a posterior probability for the hypothesis that the two segments are governed by the same distribution.

5.3 A Heuristic for Detecting Temporal Patterns. The correct approach to modeling temporal structure is the Markov chain defined in equation 3.2. As currently implemented, our methods are unable to handle temporal correlation for even very small sets of neurons. In this section we describe a heuristic that can be useful as an approximation. We see a temporal pattern as a shifted synchrony. Suppose we observe that neuron frequently 1 fires exactly two bins after neuron 2. When do we declare that that this phenomenon is significant? As a heuristic, we shift the data of neuron 2 two bins ahead of neuron 1 and neglect the two first bins of neuron 1 and the last two of neuron 2. We now have an artificially created parallel process for these two neurons such that the temporal pattern has become a synchrony. We will declare that a temporal pattern is significant if the synchrony obtained by performing the corresponding shifts is significant. For clusters of three neurons, three-dimensional displays are easily obtained. For clusters with more than three neurons, graphical displays can still be provided by using projection techniques (Martignon, 1999).

Figure 3 illustrates the histogram of the correlations between two neurons

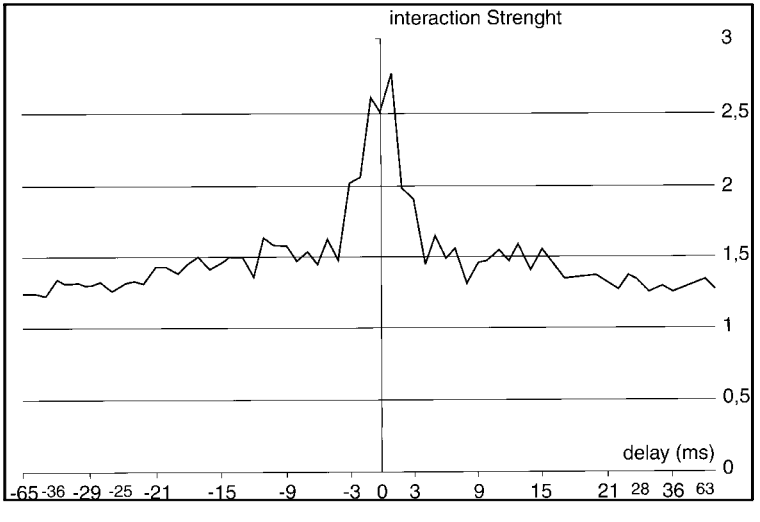

Figure 3: Histogram of estimated interactions of two neurons (1,2 of Table 4).

recorded by Freiwald in the visual cortex of a macaque (see section 7.2). This histogram, as one would expect, is equivalent to the one obtained by means of the method proposed by Palm, Aertsen, and Gerstein (1988) based on the concept of surprise.

5.4 Discussion of the Bayesian Approach. The frequentist procedures described in section 4 are designed to address the question of whether a particular set A of neurons, fixed in advance of performing the test, exhibits an interaction, as evidenced by $\theta_A \neq 0$. The problem we face is different: we wish to discover which sets of neurons are involved in spatiotemporal patterns. The Bayesian approach is designed for just such problems. Our Bayesian procedure automatically tests all hypotheses in which $\theta_A \neq 0$ against all hypotheses in which $\theta_A = 0$, for every A we consider. A major advantage for our application is that we have no need to choose a specific null hypothesis against which to test the alternative hypothesis each time.

The Bayesian approach is often more conservative than the frequentist approach. We have sometimes obtained a low posterior probability that $\theta_A \neq 0$ even when a frequentist test rejects $\theta_A = 0$. We consider this conservatism an advantage.

Another advantage of the Bayesian approach is its ability to treat situations in which the data do not clearly distinguish between alternate models. Suppose each of two hypotheses, $\theta_A \neq 0$ and $\theta_B \neq 0$, can be rejected individually by a standard frequentist test, but neither can be rejected when the other nonzero effect is in the model. This might occur when the sets A and B overlap, the model containing only $\theta_{A \cap B}$ is inadequate to explain the data, but the data are not sufficient to determine which neurons outside $A \cap B$ are involved in the interaction. Because the two models $\langle \theta_A = 0, \theta_B \neq 0 \rangle$ and $\langle \theta_A \neq 0, \theta_B = 0 \rangle$ are nonnested, constructing a frequentist test to determine which effect to include in the model is not straightforward. The Bayesian mixture approach handles this situation naturally. A mixture model will assign a moderately high probability to the hypothesis that each effect individually is nonzero and a very small probability to the hypothesis that both are nonzero.

The theoretical arguments in favor of the Bayesian approach amount to little unless the method is practical to apply. Until recently, computational constraints limited the application of Bayesian methods to very simple problems. Naive application of the model described is clearly intractable. With n neurons and considering only synchronicity with no temporal correlation, there are 2^{2^n} interaction structures to consider. Clearly it is infeasible to enumerate all of these for even moderately sized sets of neurons. We performed explicit enumeration for problems of four or fewer neurons. For larger sets, we sampled interaction structures using MC^3. We have found that our MC^3 approach discovers good structures after a few hundred iterations and that a few thousand iterations are sufficient to obtain accurate parameter esti-

mates. However, the algorithm we currently use estimates the entire set of probabilities, one for each configuration, for each structure considered. The algorithm thus scales as 2^{nr} and becomes intractable for large numbers of neurons or long temporal windows.[6] Our current implementation in Lisp-Stat (Tierney, 1990) on a 133 MHz Macintosh Power PC enumerates the 2048 interaction structures for four neurons in about 1 hour and samples 5000 structures on six neurons in about 8 hours. We reimplemented our sampling algorithm in C++ on a 300 MHz Pentium computer and achieved speedup of more than an order of magnitude. Nevertheless, it is clearly necessary to develop faster estimation procedures. We are working on approximation methods that scale as a low-order polynomial in the number of neurons and the length of the time window. With such approximation methods, the Bayesian approach will become tractable for more complex problems. In this article our objective is to demonstrate the value of the methods on data sets for which our current approach is feasible. Future work will extend the reach of our methods to larger data sets and problems of temporal correlation.

In summary, the Bayesian approach is naturally suited to simultaneous testing of multiple hypotheses. With appropriate choice of prior distribution, it handles simultaneous estimation of many parameters with less danger of overfitting than frequentist methods because of its bias toward models with a small number of adjustable parameters. It treats nonnested hypotheses with no special difficulty. It provides a natural way to account for situations in which the data indicate the existence of higher-order effects but are ambiguous about the specific neurons involved in the higher-order effects. The approach described here is thus quite attractive from a theoretical perspective, but is currently tractable only for small sets of neurons exhibiting no temporal correlation. Further work is needed to develop computationally tractable estimation procedures for larger sets of neurons exhibiting temporal correlation.

6 Applying the Models to Simulated Data

The first simulation we present describes four neurons modeled as integrate-and-fire units. Three neurons labeled 1,2,3 are excited by a fourth one, labeled 0, which receives an input of 2. The resolution is of 1 msec and the simulation is of 200,000 msec. The units' thresholds are of 10 MV (reset of 0 MV), $\tau = 25$ ms, and the refractory time is 1 msec. A gaussian noise is added to the potential ($\sigma = 0.5$ MV). Figure 4 describes three different situations.

[6] This intractability also plagues naive application of the frequentist tests described in section 4 above, which also rely on computing the entire vector of configuration probabilities for the null and alternative hypotheses

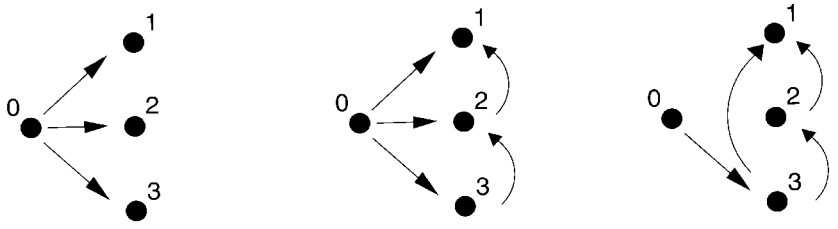

Figure 4: Three connectivity situations

In the first two graphs of Figure 4 (from left to right) we depict situations in which neuron 0 excites neurons 1, 2, and 3. In the second graph there are two additional interactions. In the third graph we depict a situation where neuron 0 excites only one of the remaining three neurons and there are other interactions between neurons 1, 2, and 3. Data were generated from simulations for the three situations. Both the frequentist and the Bayesian methods were used with the three sets of data. The triplet (1,2,3) was detected as significant at the 0.0001 level by the frequentist method and a probability of 0.99 by the Bayesian method for data corresponding to the situation described by the first graph. The triplet was also significant at the 0.0001 level and had a probability of 0.97 for data corresponding to the second situation. In data corresponding to the third situation, none of the methods detected the triplet.

To test the ability of our models to identify known interaction structures, we applied our models to synthetically generated data. We generated data randomly from a four-neuron model of the form 3.3, where we specified Ξ and θ_A. Table 1 shows the θ_A for $|A| \geq 2$ in the model we used to generate the data.

With four neurons and constraining all the single-neuron effects to be in the model, there are 11 effects to be estimated. Thus, there are $2^{11} = 2048$ different interaction structures to be considered. We were able to enumerate all interaction structures for the Bayesian model average, eliminating the need for the MCMC search. Table 1 shows the results of the Bayesian analysis for simulated segments of length 10,000 up to 640,000. It is interesting to note the increasing ability to detect interactions as the segment length increases, as well as the increasing accuracy of the estimates of the interaction strengths. It is also interesting to note that the third data set assigns a 45% probability to an effect for $A = \{1, 3, 4\}$, a set for which $\theta_A = 0$, but there is considerable overlap with sets for which there are true interactions. In general, it may be possible to determine that certain neurons are involved in interactions with certain other neurons, even when it is not possible to determine precisely which other neurons are also involved in specific interactions.

Table 1: Results from Simulated Synchronization Patterns.

		10,000 Samples		40,000 Samples		160,000 Samples		640,000 Samples	
Cluster A	θ_A	$P(\theta_A \neq 0)$	MAP Est. θ_A	$P(\theta_A \neq 0)$	MAP Est. θ_A	$P(\theta_A \neq 0)$	MAP Est. θ_A	$P(\theta_A \neq 0)$	MAP Est. θ_A
1,2	0	0.01	0.02	0.01	0.06	0.00	0.00	0.00	0.01
1,3	0.05	0.02	0.12	0.00	0.03	0.27	0.06	1.00	0.05
1,4	0.10	0.01	0.05	0.72	0.13	1.00	0.12	1.00	0.11
2,3	0	0.01	−0.03	0.00	−0.02	0.00	−0.01	0.00	−0.01
2,4	0.30	0.52	0.21	1.00	0.33	1.00	0.30	1.00	0.30
3,4	0.50	1.00	0.57	1.00	0.52	1.00	0.51	1.00	0.50
1,2,3	0.30	0.03	0.20	1.00	0.39	1.00	0.36	1.00	0.30
1,2,4	0	0.01	0.10	0.21	0.20	0.00	0.04	0.00	0.00
1,3,4	0	0.04	0.20	0.01	0.09	0.45	0.11	0.00	0.01
2,3,4	0	0.03	0.19	0.00	−0.03	0.00	0.12	0.00	0.01
1,2,3,4	0.20	0.10	0.43	0.02	0.18	0.03	0.12	1.00	0.21

7 Detecting Synchronization in Experimental Data

7.1 Recordings from the Frontal Cortex of Rhesus Monkeys. In this section we present results from fitting our model to data obtained by Vaadia and collegues. The recording and behavioral methodologies were described in detail in Prut et al. (1998). Briefly, spike trains of several neurons were recorded in frontal cortical areas of rhesus monkeys. The monkeys were trained to localize a source of light and, after a delay, to touch the target from which the light was presented. At the beginning of each trial, the monkeys touched a "ready-key," and then a central red light was turned on (*ready signal*). Three to 6 seconds later, a visual cue was given in the form of a 200 ms light blink coming from the left or the right. After a delay of 1 to 32 seconds, the color of the ready signal changed from red to yellow and the monkeys had to release the ready key and touch the target from which the cue was given.

The spikes of each neuron were encoded as discrete events by a sequence of zeros and ones with time resolution of 1 millisecond. The activity of the whole group of simultaneously recorded six neurons was described as a sequence of configurations or vectors of these binary states. Since the method presented in this article for detecting synchronizations does not take into account nonstationarities, the data were selected by cutting segments of time during which the cells did not show significant modulations of their firing rate. This selection was performed by the experimenter by means of eyeballing heuristics. The segments were taken from 1000 ms before the cue onset to 1000 ms thereafter. The duration of each segment was 2 seconds. We used 94 such segments (total time of 188 seconds) for the analysis presented

Table 2: Results for Preready Signal Data, Coefficients with Posterior Probability > 0.20.

Cluster A	G^2	Posterior Probability of A	MAP Estimate of $\tilde{\theta}_A$	Standard Deviation of θ_A
4,6	.001	0.99	0.49	0.11
3,4	.232	0.18	0.59	0.25
3,4,6	.257	0.38	0.78	0.29
2,3,4,5	0.13	0.48	2.34	0.82
1,4,5,6	.98	0.29	−1.85	1.20

below. The data were then binned in time windows of 40 milliseconds. The choice of the bin length was determined in discussion with the experimenter (see also Vaadia et al., 1995, for the rationale). The frequencies of configurations of zeros and ones in these windows are the data used for analysis in this article. We analyzed recordings prior to the ready signal separately from data recorded after the ready signal. Each of these data sets is assumed to consist of independent trials from a model of the form 2.20. Tables 2 and 3 present results from applying our method to these data sets. Likelihood ratio statistics (G^2), posterior probabilities of nonzero effects, point estimates of effect magnitudes, and standard deviations are shown for all interactions with posterior probability at least 20%.

There was a high-probability second-order interaction in each data set: (4, 6) in the preready data and (3, 4) in the postready data. In the preready data, a fourth-order interaction (2, 3, 4, 5) had posterior probability near 50% (this represents nearly five times the prior probability of 0.1).

Several data sets from this type of experiment were analyzed by means of the methods presented in this article under the guidance of the sixth author, who conducted the experiments. In certain cases we compared our results on pairwise correlations with results that the experimenter and coworkers obtained by traditional methods, and, as we expected, our detected interactions coincided with theirs. With rare exceptions interactions of order 2 or more were all characterized by positive effects.

Table 3: Results for Postready Signal Data: Effects with Posterior Probability > 0.20.

Cluster A	G^2	Posterior Probability of A	MAP Estimate of $\tilde{\theta}_A$	Standard Deviation of θ_A
3,4	0.03	0.95	1.00	0.27
2,5	0.091	0.44	1.06	0.36
1,4,6	0.752	0.26	0.38	0.15
2,4,5,6	1.454	0.22	−1.53	1.33
1,4,5,6	0.945	0.35	1.12	0.46

Comparison of Tables 2 and 3 reveals some similarities and some important differences. The preready data show strong evidence for the (4,6) interaction. Although the posterior probability of a pair interaction between neurons 4 and 6 in the postready data was very small (the value was 0.03), there are several higher-order terms with moderately high posterior probability that involve these two neurons. Overall, we estimate a 99.9% probability that neurons 4 and 6 are involved in some interaction in the preready data and a 62.0% probability in the postready data. The postready data show strong evidence for the (3,4) interaction; this two-way interaction has probability only 18% in the preready data. Again, there are several moderately probable interactions in the preready data involving these two neurons. We estimate a 91.2% chance that neurons 3 and 4 are involved in some interaction in the preready data and a 99.0% chance in the postready data.

It thus appears plausible that the pairs (4,6) and (3,4) are involved in interactions, possibly involving other neurons in the set, in both pre- and postready segments. We might speculate that a model involving pair interactions (4,6) and (3,4) is a plausible structure for both pre- and postready data. We might then ask whether the differences between the estimates in Tables 2 and 3 are merely artifacts or whether there are statistically detectable differences between pre- and postready data. To answer this question, we applied the test described in sections 5.2 and A.1.2. We fit a modelcontaining all pair interactions to both pre- and postready data and t o the combined data sets. We computed a posterior probability of only 1.7×10^{-7} that both pre- and postready segments come from the same distribution, using a 50% prior probability that they came from the same distribution. Thus, there is extremely strong evidence that there are detectable differences between the pre- and postready segments.

Both of the second-order interactions we detected are positive. Results of this type—that is, of varying second-order structure across the phases of the experiment—have been obtained previously by several groups of researchers (Grün, 1996; Abeles et al., 1993; Riehle et al., 1997) who have used a frequentist method based on rejecting the null hypothesis that the distribution of the joint process is binomial. Both this method and that of Palm et al., (1988) are essentially consistent with the method presented here for the second-order case. A difference between the frequentist approach introduced by Palm et al. and ours is the test used for determining significance. We use G^2, while Palm et al. use a Fisher significance test comparing the data with the binomial distribution determined by the success probabilities fixed by the first-order model.

7.2 Higher-Order Synchronization in the Visual Cortex of the Behaving Macaque. Precise patterns of neural activity of groups of neurons in the visual cortex of behaving macaques were detected by means of the analysis of data on multiunit recordings obtained by Freiwald. The experiment consisted of a simple fixation task: the monkey was supposed to learn to

detect a change in the intensity of a fixated light point on a screen and react with a motoric answer. The fixation task paradigm is the following: A fixation point appears on the screen and stays on the screen for 3 seconds. The monkey fixates this point and then presses a bar. After the bar is pressed, an interval of approximately 5 seconds begins, during which the monkey fixates the point and continues to press the bar. After 5 seconds, the point darkens and disappears from the screen, while the monkey releases the bar. During the described interval, a short visual stimulus is presented on the screen, which is not relevant for behavior. Its function is that the monkey learns that the disappearance of the fixation point can occur only after the stimulus presentation. Thus, the monkey is not forced to total concentration on the fixation point. The data presented here were obtained by recordings of the activity of four neurons in the inferotemporal cortex (IT) on repeated trials of 5.975 ms each. We decomposed each trial into a first chunk of 1,000 ms, considered stationary, a nonstationary interval of 800 ms, and then four segments of 1,000 ms each, followed by a short last segment of 175 ms. We chose a bin width of 10 ms. We present a comparison between the activity during the first fixation segment and the activity of the segments after stimulus. We performed tests of synchronization with both the frequentist and the Bayesian methods and obtained consistent results. The results of the Bayesian method are displayed in Tables 4 and 5. We first performed pairwise comparisons of fully saturated models for the four poststimulus segments as described in sections 5.2 and A.1.2 and determined that none of these four segments was statistically distinguishable from the others. We therefore combined these segments into a single data set. We then compared the fixation phase with the four poststimulus segments and found that it differed from them. We therefore show the results from the fixation phase in Table 4 and the combined poststimulus phase in Table 5.

We found that Table 4 exhibits extremely strong excitatory couples for all pairs of neurons and a few triplets. The comparison of these interactions with those of the poststimulus segments (see Table 5) is quite interesting. All pairs of neurons exhibit positive pair interactions in both data sets. The magnitudes of most of the interaction strengths are similar. However, the (1, 4) interaction is nearly twice as strong in the fixation phase as in the poststimulus phase. The (1, 2) interaction is also larger in the fixation phase by a factor of about 1.5. The threeway interaction found in the fixation phase also appears in the poststimulus phase. The poststimulus data show two additional three-way interactions that are improbable in the fixation phase.

7.3 Neuronal Interactions in Anesthetized Rat Somatosensory Cortex. We also applied the models to data collected from the vibrissal region of the rat cortex. The vibrissae are arranged in five rows on the snout (named A, B, C, D, and E) with four to seven vibrissae in each row. Neurons in the vibrissal area of the somatosensory cortex are organized into columns; layer

Table 4: Synchronies with Probability Larger Than 0.1 in the First 1000 ms in the Fixation Task.

Cluster A	Posterior Probability of A	MAP Estimate $\tilde{\theta}_\xi$	Standard Deviation of θ_A
3,4	0.99	0.69	0.15
2,4	1.00	0.56	0.11
2,3	1.00	1.54	0.15
1,4	1.00	0.70	0.09
1,3	0.99	1.72	0.12
1,2	0.99	1.45	0.09
2,3,4	0.22	0.66	0.28
1,3,4	0.03	−0.34	0.25
1,2,3	0.99	−1.26	0.23

IV of the column is composed of a tightly packed cluster of neurons called a barrel (Woolsey & Van der Loos, 1970). Each barrel column is topologically related to a single mystacial vibrissa on the contralateral snout; thus vibrissa D2 (second whisker in row D) is linked to column D2 (Welker, 1971). Our wish was to find out whether interacting neurons are distributed across cortical columns or restricted to single columns. Furthermore we examined whether the spatial distribution of interacting neurons varies as a function of the stimulus site.

The data analyzed here were recorded from the cortex of a rat lightly anesthetized with urethane (1.5 g/kg) and held in a stereotaxic apparatus. Cortical neuronal activity was recorded through an array of six tungsten microelectrodes inserted in cortical columns D1, D2, D3,and C3, as well as in the septum located between columns D1 and D2 (see Figure 5). Using spike sorting methods, events from 15 separate neurons were discriminated.

Table 5: Synchronies with Probability Larger Than 0.1 in the 4000 ms After the Visual Stimulus.

Cluster A	Posterior Probability of A (Frequency)	MAP Estimate $\tilde{\theta}_\xi$	Standard Deviation of θ_A
3,4	1.00	0.70	0.10
2,4	1.00	0.53	0.05
2,3	1.00	1.51	0.06
1,4	1.00	0.39	0.06
1,3	1.00	1.45	0.06
1,2	1.00	0.94	0.04
2,3,4	0.99	−0.59	0.12
1,3,4	0.85	−0.43	0.12
1,2,3	1.00	−0.68	0.09

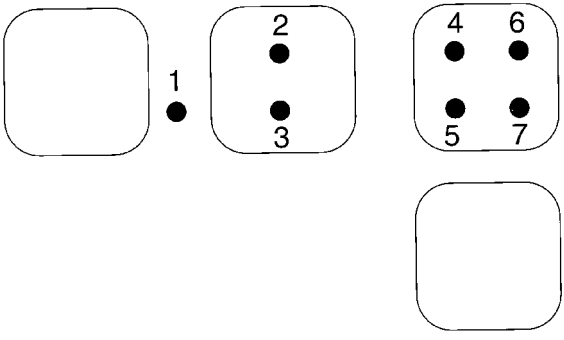

Figure 5: Seven neurons located between and in four barrels corresponding to four vibrissae.

We analyzed data for 7 of these 15 neurons. Locations of these 7 neurons are shown in Figure 5. During the recording session, individual whiskers were deflected across 50 to 55 trials by a computer-controlled piezoelectric wafer that delivered an up-down movement lasting 100 ms; the stimulus was repeated once per second (Diamond, Armstrong-Jones, & Ebner, 1993). We have restricted the analysis to data collected during stimulation of vibrissae D1, D2, and D3 and only during the 200 ms peristimulus period (100 ms after the up movement and 100 ms after the down movement). For each stimulus site, then, there were 10,000 to 11,000 ms of data. As in the previous section, the spiking events (in the 1 ms range) of each neuron were encoded as a sequence of zeros and ones, and the activity of the whole group was described as a sequence of configurations or vectors of these binary states.

The interactions illustrated in Table 6 are those for which the likelihood test statistic is significant at the 0.0001 level. We report the probabilities of these interactions obtained by means of the Bayesian approach and the estimated Θs . Table 6 shows the interactions occurring during stimulation of whiskers D1, D2, and D3.

From analysis of this limited data set, several interesting observations can be made: (1) only positive interactions were detected; (2) interactions occurred both within a cortical columnar unit and across columns; (3) interactions were stimulus dependent (neurons were not continuously correlated with one another but became grouped together as a function of the stimulus parameter); and (4) during different stimulus conditions, a single neuron may take part in different groups.

8 Discussion

This article proposes a parameterization of synchronization and, in general, of temporal correlation of neural activation of more than two neurons. This

Table 6: Interactions Significant at the 0.0001 Level and High Probability of Occurring.

Cluster	D 1		D 2		D 3	
	$P(\theta_A \neq 0)$	Est. θ_A	$P(\theta_A \neq 0)$	Est. θ_A	$P(\theta_A \neq 0)$	Est. θ_A
4,5,6	0.63	3.52	—	—	—	—
4,7	—	—	0.68	2.19	—	—
3,4,6	—	—	0.75	3.50	—	—
2,7	—	—	0.76	2.22	—	—
2,5	—	—	0.52	1.55	—	—
1,4,5	—	—	0.54	3.01	—	—
3,7	—	—	—	—	0.58	1.55

Note: Blanks correspond to low probabilities and high significance levels.

parameterization has a sound information-theoretical justification. Recent work by Amari (1999) on information geometry on hierarchical decomposition of stochastic interactions provides mathematical support of our thesis that the effects of log-linear models are the adequate parameterization of pure higher-order interactions. If the problem is to analyze a group of neurons and establish whether they fire in synchrony, the statistical problems of estimating the significance of the corresponding effect being nonzero have a straightforward, reliable treatment. If the problem is to detect the whole interaction structure in a group of neurons (i.e., which subsets have interactions and how strong), complexity becomes an inevitable feature of any candidate treatment. We present both a frequentist and a Bayesian approach to the problem and argue that the Bayesian approach has major advantages over the frequentist approach. The approach is naturally suited to simultaneous estimation of many parameters and simultaneous testing of multiple hypotheses. Yet the Bayesian approach is of great complexity when the temporal interaction structure of, say, eight neurons is to be assessed. Even the heuristic shift method briefly discussed in section 5.3 becomes very complex for the Bayesian approach in the form described here. New techniques for speeding up Bayesian search across structures are presented in a forthcoming article.

An important underlying assumption in our investigations has been that the data are basically stationary. This is seldom the case. As we know from the important work by Abeles et al. (1995), cortical activity flips from one quasistationary state to the next. These authors developed a hidden Markov model technique to determine these flips. A combination of their methods with the ones presented here would allow the analysis of structure changes through quasistationary segments.

Another advancement in global structure analysis of spike activity has been obtained by de Sa', de Charms, and Merzenich (1997) by means of a Helmholtz machine that determines highly probable patterns of activation

based on spike train data. Here again a combination of methods would allow analysis of the correlation structure in probable patterns.

The reason for our presentation of both the frequentist and the Bayesian approach to detecting correlation structure is twofold. Our wish is, on the one hand, to reinforce the use of the frequentist method, in spite of all its shortcomings, because it is quick and can be used for large numbers of neurons and long time delays. Our experience with the systematic comparison of the two methods has enhanced the confidence that a careful frequentist strategy is reliable. On the other hand, we promote the Bayesian approach, which, from the point of view of experimenters, may seem cumbersome or too complex, but which, in the long run, will probably become an accepted tool.

As a final remark let us observe that higher-order interactions in the empirical data analyzed in this article did not seem to be frequent.

Appendix

A.1 The Constrained Perturbation Procedure. In this appendix we construct the null hypotheses for the tests proposed in section **X**. Assume that p is a probability distribution defined on the configurations of a set of neurons. Consider the following problem: Find a distribution p* such that the marginals of p^* coincide with those of p and p^* maximizes entropy among all distributions whose marginals of order less than $|A|$ coincide with those of p.

It can be proved (Whittaker, 1989) that there is a unique distribution p^{**} on the configurations of A whose marginals of order less than $|A|$ coincide with those of p and such that $\theta^{**}{}_A = 0$, where $\theta^{**}{}_A$ is the coefficient corresponding to A in the log expansion of p^{**}. Let us combine this fact with Good's (1963) famous theorem, which states that the distribution p^* maximizing entropy in the manifold of distributions with the same marginals of order less than $|A|$ of p necessarily satisfies $\theta^*{}_A = 0$. We conclude that the distribution p^{**} has to coincide with p^*. We sketch the construction of such a distribution p* for the simple case of three neurons. Suppose that $A = \{1, 2, 3\}$ and that p is a strictly positive distribution on A. We want to construct the distribution p^* that maximizes entropy among all those consistent with the marginals of order 2, denoted by Marg.({{1, 2}, {1, 3}, {1, 3}}). Define

$$
\begin{aligned}
p*(1,1,1) &= p(1,1,1) - \Delta \\
p*(1,1,0) &= p(1,1,0) + \Delta \\
p*(1,0,1) &= p(1,0,1) + \Delta \\
p*(1,0,0) &= p(1,0,0) - \Delta \\
p*(0,1,1) &= p(0,1,1) + \Delta
\end{aligned}
$$

$$p*(0,1,0) = p(0,1,0) - \Delta$$
$$p*(0,0,1) = p(0,0,1) - \Delta$$
$$p*(0,0,0) = 1 - \sum_{B\subset\{1,2,3\}} p(\chi_B).$$

Let us compute the marginal $p^*(x_1 = 1, x_2 = 1)$. We have $p^*(x_1 = 1, x_2 = 1) = p(1,1,1) - \Delta + p(1,1,0) + \Delta = p(1,1,1) + p(1,1,0) = p(x_1 = 1, x_2 = 1)$.

Now we have to solve for Δ in the equations ${\theta^*}_A = 0$. Observe that our construction contains only one step where a numerical approximation becomes necessary: solving for ${\theta^*}_A = 0$ in terms of Δ. This can be done by using any one-dimensional optimization procedure such as Newton's approximation method. Here Newton's approximation method has to be used in combination with the condition that the perturbation Δ is such that all candidate solutions are probability distributions, and thus their values are between 0 and 1. Let us now sketch the construction in the general case of N neurons. If B is a nonempty subset of A, denote by χ_B the configuration that has a component **1** for every index in B and 0 elsewhere. For each nonempty subset B, define $p^*(\chi_B) = [(\chi_B) + (-1)]^{|B|}\Delta$, where, again, Δ is to be determined by solving for ${\theta^*}_A \equiv 0$. Every marginal of order $N-1$ is the sum of the probabilities of two configurations that coincide on $N-1$ entries and differ on the remaining one. Thus, the sign of Δ will be different in the two summands, and Δ will cancel out. Therefore, the marginal of p^* will coincide with the correspondent marginal of p, as condition 1 requires.

A.2 Bayesian Estimation and Hypothesis Testing Methods.

A.2.1 Inferring Structure and Parameters from Observations. We are concerned with the problem of inferring the interaction structure Ξ and interaction strengths $\underline{\theta}_\Xi$ from a sample $\underline{x}_1, \ldots, \underline{x}_N$ of N independent observations from the distribution $p(x \mid \underline{\theta}_\Xi, \Xi)$ given by equation 3.3. Initial information about $\underline{\theta}_\Xi$ prior to obtaining the data is expressed as a prior probability distribution over structures and interaction strengths. We use a mixed discrete-continuous prior distribution denoted by $g(\underline{\theta}_\Xi \mid \Xi)\pi_\Xi$, where π_Ξ is the prior probability of structure Ξ and $g(\underline{\theta}_\Xi \mid \Xi)$ is a continuous density function for $\underline{\theta}_\Xi$.

We chose a prior distribution that encodes the assumptions that all singleton effects are nonzero, most effects of order 2 or greater are nonzero, and most nonzero effects are not too large. Specifically, the prior distribution assumes all singleton effects are included with certainty, and each interaction A of order 2 or greater is included with probability 0.1 independent of the other interactions.[7] Thus, an interaction structure Ξ that includes all

[7] We varied the prior probability between 0.025 and 0.40. As expected, extreme poste-

singleton sets and k interactions of order 2 or higher has prior probability $\pi_\Xi = 10^{-k}$.

Conditional on an interaction structure Ξ, the interaction strength θ_A corresponding to any absent interaction $A \notin \Xi$ is assumed to be identically zero. We assume that the nonzero θ_A are independent of each other and are normally distributed with zero mean and standard deviation $\sigma = 2$. This prior distribution encodes a prior assumption that the θ_A are symmetrically distributed about zero, that is, excitatory and inhibitory interactions are equally likely. The standard deviation $\sigma = 2$ (that is, most θ_A lie between -4 and 4) is based on previous experience applying this class of models (Martignon et al., 1995) to other data, as well as discussions with neuroscientists. We initially regarded the symmetry of the normal distribution as unrealistic because we expected most interaction parameters to be positive. However, we felt that the convenience of the normal distribution outweighed this disadvantage. We found in our data analyses that many estimated interactions, especially those of order higher than 2, were negative (we attribute these negative higher-order interactions to overlapping pair interactions with external neurons).

The joint probability of the observed set of configurations for a fixed interaction structure, viewed as a function of $\underline{\theta}_\Xi$, is called the *likelihood function* for $\underline{\theta}_\Xi$ given Ξ. From equation 3.3, it follows that the likelihood function can be written

$$L(\underline{\theta}_\Xi) = p(\underline{x}_1 \mid \underline{\theta}_\Xi, \Xi) \cdots p(\underline{x}_N \mid \underline{\theta}_\Xi, \Xi) = \exp\left\{ N \sum_{A \in \Xi} \theta_A \bar{t}_A \right\}, \qquad \text{(A.1)}$$

where $\bar{t}_A = \frac{1}{N} \sum t_A$ is the frequency of simultaneous activation of the nodes in the set A and is referred to as the marginal frequency for A. The random vector[8] $\underline{\overline{T}}$ of marginal frequencies is called a *sufficient statistic* for $\underline{\theta}_\Xi$ because the likelihood function, and therefore the posterior distribution of $\underline{\theta}_\Xi$, depends on the data only through $\underline{\overline{T}}$.[9] The expected value of $\overline{T}_A$ is the probability that $T_A = 1$ and is simply the marginal function for the set A.

The posterior distribution is also a mixed discrete-continuous distribution $g^*(\underline{\theta}_\Xi \mid \Xi)\pi^*_\Xi$. The posterior density for θ_Ξ conditional on structure Ξ is given by

$$g(\underline{\theta}_\Xi \mid \Xi, \overline{x}_1, \ldots, \underline{x}_N) = Kp(\underline{x}_1 \mid \underline{\theta}_\Xi, \Xi) \cdots p(\underline{x}_N \mid \underline{\theta}_\Xi, \Xi) g(\underline{\theta}_\Xi \mid \Xi), \quad \text{(A.2)}$$

rior probabilities remained extreme. Manipulation of the prior probabilities had the most effect when the evidence for or against the presence of an interaction was not overwhelming (i.e., posterior probabilities roughly in the range between 0.1 and 0.9).

[8] We use the standard convention of denoting random variables by uppercase letters and their observed values by lowercase letters. We denote vectors by underscores.

[9] The sufficient statistic $\underline{\overline{T}}$ contains T_A only for those clusters $A \in \Xi$. For notational simplicity, the dependence of $\underline{\overline{T}}$ on the structure Ξ is suppressed.

where the constant of proportionality K is chosen so that equation A.2 integrates to 1:

$$K = \left(\int_{\underline{\theta}_\Xi} p(\underline{x}_1 \mid \underline{\theta}_\Xi, \Xi) \cdots p(\underline{x}_N \mid \underline{\theta}_\Xi, \Xi) g(\underline{\theta}_\Xi \mid \Xi)\, d\underline{\theta}_\Xi \right)^{-1}. \tag{A.3}$$

In this expression, the range of integration is over $\underline{\theta}_\Xi$ where $\Xi' = \{A\colon A \in \Xi, A \neq \varnothing\}$. That is, the joint density is integrated only over those parameters that vary independently. Because the probabilities are constrained to sum to 1, $\theta_\varnothing$ is a function of the other θ_A:

$$\theta_\varnothing = -\log \left(\sum_{\underline{x}} \exp \left\{ \sum_{\substack{A \in \Xi \\ A \neq \varnothing}} \theta_A t_A(\underline{x}) \right\} \right). \tag{A.4}$$

A.2.2 Approximating the Posterior Distribution. Using data to infer the distribution involves two subproblems: (1) given an interaction structure Ξ, infer the parameters θ_A for $A \in \Xi$, and (2) determine the posterior probability of an interaction structure Ξ. Neither of these subproblems has a tractable closed-form solution. We discuss approximation methods for each in turn.

Options for approximating the structure-specific posterior density function $g^*(\underline{\theta}_\Xi \mid \Xi) = g(\underline{\theta}_\Xi \mid \Xi, \underline{x}_1, \ldots, \underline{x}_N)$ include analytical or sampling-based methods. Because we must estimate posterior distributions for a large number of structures, we expected Monte Carlo methods to be infeasibly slow. We therefore chose to use the standard large-sample normal approximation.

The approximate posterior mean is given by the mode $\tilde{\underline{\theta}}_\Xi$ of the posterior distribution. We found the posterior mode by using Newton's method to maximize the logarithm of the joint mass density function:

$$\tilde{\underline{\theta}}_\Xi - \text{argmax}_{\underline{\theta}_\Xi} \left\{ \log \left(p(\underline{x}_1 \mid \underline{\theta}_\Xi, \Xi) \cdots p(\underline{x}_N \mid \underline{\theta}_\Xi, \Xi) g(\underline{\theta}_\Xi \mid \Xi) \right) \right\}. \tag{A.5}$$

When we used arbitrary starting values, we encountered occasional instabilities. This problem was alleviated by running an iterative proportional fitting algorithm on a "pseudosample" consisting of a small, positive constant added to the sample counts (the positive constant bounded the starting probabilities away from zero for configurations with no observations). We then iterated using Newton's method to find the maximum a posteriori estimate.

The posterior covariance is approximated by the inverse of the Hessian matrix of second partial derivatives evaluated at the maximum a posteriori estimate. This matrix can be obtained in closed form:

$$\tilde{\Sigma}_\Xi = \left[D^2_\theta \log \left(p(\underline{x}_1 \mid \underline{\theta}_\Xi, \Xi) \cdots p(\underline{x}_N \mid \underline{\theta}_\Xi, \Xi) g(\underline{\theta}_\Xi \mid \Xi) \right) \right]^{-1}. \tag{A.6}$$

The posterior probability of a structure Ξ is given by

$$\pi^*_{\Xi} = p(\Xi \mid \underline{x}) \propto p(\underline{x} \mid \Xi)\pi_{\Xi}, \tag{A.7}$$

where the first term on the right is obtained by integrating $\underline{\theta}_{\Xi}$ out of the joint parameter/data density function:

$$p(\underline{x} \mid \Xi) = \int_{\underline{\theta}_{\Xi}} p(\underline{x}_1 \mid \underline{\theta}_{\Xi}, \Xi) \cdots p(\underline{x}_N \mid \underline{\theta}_{\Xi}, \Xi) g(\underline{\theta}_{\Xi} \mid \Xi)\, d\underline{\theta}_{\Xi}. \tag{A.8}$$

There is no closed-form solution for this integral. We use Laplace's method (Tierney & Kadane, 1986; Kass & Raftery, 1995), which again relies on the normal approximation to the posterior distribution. Assuming that $p(\underline{x} \mid \underline{\theta}_{\Xi}, \Xi)$ is approximately normal with mean A.5 and covariance A.6, and integrating the resulting normal density as in A.8, we obtain:

$$p(\underline{x} \mid \Xi) \approx (2\pi)^{d/2} \left| \tilde{\Sigma}_{\Xi} \right|^{1/2} p(\underline{x}_1 \mid \tilde{\underline{\theta}}_{\Xi}, \Xi) \cdots p(\underline{x}_N \mid \tilde{\underline{\theta}}_{\Xi}, \Xi) g(\tilde{\underline{\theta}}_{\Xi} \mid \Xi). \tag{A.9}$$

The normal approximation is accurate for large sample sizes (De Groot, 1970}). The posterior density function is always unimodal, and we have observed that the conditional density for θ_A given the other θ's is not too asymmetric even when the corresponding marginal T_A is small. We have also found that very small frequencies typically give rise to very small posterior probabilities of the associated effects, in which case the accuracy of the approximation is not crucial. Nevertheless, the quality of the Laplace approximation is a question worthy of further investigation.

Equation A.7 gives the posterior probability of a structure only up to a proportionality constant. The posterior probability is obtained by computing equation A.7 for all structures and then normalizing. The total number of structures grows as , where k is the number of neurons. For more than four neurons, this number is much too large to enumerate explicitly. For data sets with more than four neurons, we used an MC^3 algorithm to search over structures (Madigan & York, 1993). The algorithm works as follows:

1. Begin at initial structure Ξ. Compute $p(\underline{x} \mid \Xi)\pi_{\Xi}$.[10]
2. Nominate a candidate structure A to add to or delete from Ξ to obtain a new structure Ξ'. The candidate structure is nominated with probability $\rho(A \mid \Xi)$.[11]

[10] To avoid numeric underflow, we compute the logarithm of the required probability. However, for clarity of exposition, we present the algorithm in terms of probabilities.

[11] To improve acceptance probabilities, we modified the distribution $\rho(A' \mid \Xi)$ as sampling progressed. The sampling probabilities were bounded away from zero to ensure ergodicity.

3. Compute $p(\underline{x} \mid \Xi)\pi_{\Xi}$.
4. Accept the candidate structure Ξ' with probability

$$p_{Accept} = \min\left\{1, \frac{p(\underline{x} \mid \Xi')\pi_{\Xi'}\rho(A \mid \Xi')}{p(\underline{x} \mid \Xi')\pi_{\Xi'}\rho(A \mid \Xi)}\right\}. \tag{A.10}$$

5. Unless the stopping criterion is met, return to step 2; else stop.

We discarded a burn-in run of 500 samples (quite sufficient to move the sampler into a range of high-probability models on the data sets we tried) and based our estimates on a run of 5000 samples. We implemented our algorithm in Lisp-Stat running on a 133 MHz Macintosh PowerBook. For the six-neuron data sets described in section 7, a run of 5000 samples took about 8 hours. This was reduced to about 2 hours on a 266 MHz PowerBook G3.

We computed posterior probabilities for each effect as the frequency of occurrence of that effect in the sample run. We estimate the conditional expected value,

$$\tilde{\theta}_A - \frac{q}{\#[A \in \Xi]} \sum_{A \in \Xi} \tilde{\theta}_{A|\Xi}, \tag{A.11}$$

where the structure-specific estimate is $\tilde{\theta}_{A|\Xi}$ obtained from equation A.5. Similarly, we compute a variance estimate as the sum of within- and between-structure components. The within-structure variance component,

$$\tilde{\sigma}^2_{A,w} = \frac{1}{\#[A \in \Xi]} \sum_{A \in \Xi} \tilde{\sigma}^2_{A|\Xi}, \tag{A.12}$$

where the summands of equation A.12 are the appropriate diagonal elements from equation A.6. The between-structure component is given by

$$\tilde{\sigma}^2_{A,b} = \frac{1}{\#[A \in \Xi]} \sum_{A \in \Xi} \left(\tilde{\theta}_{A|\Xi} - \tilde{\theta}_A\right)^2. \tag{A.13}$$

A.2.3 Detecting Changes Between Different Time Segments. Suppose we are concerned with detecting whether an environmental stimulus produces detectable changes in neuronal activity. We might examine this question by comparing models of the form 3.3 for segments recorded before and after occurrence of the stimulus. Of course, estimates for the two segments will differ. The question of interest is whether these differences can be explained by random noise or whether they represent real differences in activation patterns that can be associated with the stimulus.

One way to address this question would be to construct a super model encompassing both pre- and postready segments and extend the methods

described to include hypothesis tests for which interactions were the same as or different for the two data sets. We took a simpler approach, which could be performed using the models we have described.

Suppose we have two segments of data, labeled x_1 and x_2, of the same neurons recorded at different times. We begin by fitting models of the form 3.3 for each data set individually and for the combined data set. We then selected a model structure Ξ that included effects for any set A that had high probability in either an individual model or the combined model. We fit this single structure for the individual and combined data sets. We then computed predictive probabilities as in equation A.9 for each of the three data sets. We denote these by $p_C(\underline{x}_1, \underline{x}_2 \mid \Xi)$ for the combined model, $p_1(\underline{x}_1 \mid \Xi)$ for the first segment, and $p_2(\underline{x}_2 \mid \Xi)$, for the second segment. Assuming that the individual and combined models are equally likely a priori, the posterior probabilities were computed using the formula:

$$\frac{p_{Combined}}{p_{Separate}} = \frac{p_C(\underline{x}_1, \underline{x}_2 \mid \Xi)}{p_1(\underline{x}_1 \mid \Xi) + p_2(\underline{x}_2 \mid \Xi)}. \tag{A.14}$$

Acknowledgments

We gratefully acknowledge the anonymous reviewers of an earlier version of this article for useful comments that have greatly improved the article. K. L. was supported by a Career Development Fellowship with the Krasnow Institute for the study of cognitive science at George Mason University. The expriments by W. F. were conducted at the Max-Planck-Institute for Brain Research in Frankfurt/Main together with Wolf Singer and Andreas K. Kreiter. This work was supported by HFSP grant RG-20/95 B and SFB 517.

References

Abeles, M. (1991). *Corticonics: Neural circuits of the cerebral cortex*. Cambridge: Cambridge University Press.

Abeles, M., Bergman, H., Gat, I., Meilijson, I., Seidenmann, E., Tishby, N., & Vaadia, E. (1995). Cortical activity flips among quasi-stationary states. *Proc. Natl. Acad. Sci., 92*, 8616–8620.

Abeles, M., Bergman, H., Margalit, E. and Vaadia, E. (1993). Spatiotemporal firing patterns in the frontal cortex of behaving monkeys. *Journal of Neurophysiology, 70*, 1629–1638.

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

Abeles, M., Prut, Y., Bergman, H., & Vaadia, E. (1994). Synchronization in neural transmission and its importance for information processing. *Prog. Brain. Res., 102*, 395–404.

Aertsen, A. (1992). *Detecting triple activity*. Preprint.

Amari, S. (1999). *Information geometry on hierarchical decomposition of stochastic interactions*. Preprint.

Bishop, Y., Fienberg, S., & Holland, P. (1975). *Discrete multivariate analysis*. Cambridge, MA: MIT Press.

Cover, T., & Thomas, J. (1991).*Elements of information theory*. New York: Wiley.

Deco, G., Martignon, L., & Laskey, K. (1998). Neural coding: Measures for synfire detection. In Wulfram Gestner (Ed.) *Advances in artificial neural networks* (pp. 78–83) New York: Springer-Verlag.

De Groot, M. H. (1970). *Optimal statistical decisions*. New York: McGraw-Hill.

de Sa', R., de Charms, R., & Merzenich, M. (1997). Using Helmholtz machines to analyze multi-channel neuronal recordings. *Neural Information Processing System, 10*, 131–136.

Diamond, M. E., Armstrong-Jones, M. A., & Ebner, F. F.(1993). Experience-dedpendent plasticity in the barrel cortex of adult rats. *Proceedings of the National Academy of Science, 90*, 2602–2606.

Gerstein, G., Perkel, D., & Dayhoff, J. (1985). Cooperative firing activity in simultaneously recorded populations of neurons: Detection and measurement. *Journal of Neuroscience, 5*, 881–889.

Good, I. (1963). On the population frequencies of species and the estimation of population parameters. *Biometrika, 40*, 237–264.

Gray, C. M., König, P., Engel, A. K., & Singer, W. (1989). Oscillatory responses in cat visual cortex exhibit inter-columnar synchronization which reflects global stimulus properties. *Nature*, 388, 334–337.

Grün, S. (1996). *Unitary joint-events in multiple-neuron spiking activity-detection, significance and interpretation*. Frankfurt: Verlag Harry Deutsch.

Grün, S., Aertsen, A., Vaadia, E., & Riehle, A. (1995). *Behavior-related unitary events in cortical activity*. Poster presentation at Learning and Memory, 23rd Gröttingen Neurobiology Conference.

Hebb, D. (1949). *The organization of behavior*. New York: Wiley.

Kass, R., & Raftery, A. (1995). Bayes factors. *Journal of the American Statistical Association, 90*, 773–795.

Madigan, D., & York, J. (1993). *Bayesian graphical models for discrete data* (Tech. Rep. No. 239). Pullman: Department of Statistics, University of Washington.

Martignon, L. (1999). *Visual representation of spatio-temporal patterns in neural activation.* (Tech. Rep.) Berlin: Max Planck Institute for Human Development.

Martignon, L., von Hasseln, H., Grün, S., Aertsen, A., & Palm, G. (1995). Detecting higher order interactions among the spiking events of a group of neurons. *Biol. Cyb., 73*, 69–81.

Neyman, J., & Pearson, E. (1928). On the use and the interpretation of certain test criteria for purposes of statistical inference. *BiometriKa, 20*, 175–240, 263–294.

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

Prut, Y., Vaadia, E., Bergman, H., Haalman, I., Slovin, H., & Abeles, M. (1998). Spatiotemporal structure of cortical activity: Properties and behavioral relevance. *Journal of Neurophysiology, 79*, 2857–2874.

Rauschecker, J. (1991) Mechanisms of visual plasticity: Hebb synapses, NMDA reception and beyond. *Phys. Rev., 71*, 587–615.

Riehle, A., Grün, S., Diesmann, M., & Aertsen, A. (1997). Spike synchronization and rate modulation differentially involved in motor critical function. *Science, 281*, 34–42.

Singer, W. (1994). Time as coding space in neocortical processing: A hypothesis. In Buzsaki, G. et al. (Eds). *Temporal coding in the brain*. Berlin: Springer-Verlag.

Smith, A. F. M., & Spiegelhalter, D. J. (1980). Bayes factors and choice criteria for linear models. *Journal of the Royal Statistical Society, series B, 42*, 213–220.

Tierney, L. (1990). *Lisp-Stat*. New York: Wiley.

Tierney, L., & Kadane, J. (1986). Accurate approximations of posterior moments and marginal densities. *Journal of the American Statistical Association, 81*, 82–86.

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

Welker, C. (1971). Microelectrode delineation of fine grain somatotopic organization of SmI cerebral neocortex in albino rat. *Brain Res., 26*, 259–275.

Whittaker, J. (1989). *Graphical models in applied multivariate statistics*. New York: Wiley.

Woolsey, T. A., & Van der Loos, H. (1970). The structural organization of layer IV in the somatosensory region, SI, of mouse cerebral cortex. *Brain Res., 17*, 205–242.

Received August 7, 1997; accepted February 16, 2000.