

An exact method to quantify the information transmitted by different mechanisms of correlational coding

G Pola¹, A Thiele¹, K-P Hoffmann² and S Panzeri^{3,4}

¹ Department of Psychology, University of Newcastle upon Tyne, The Henry Wellcome Building for Neuroecology, Medical School, Framlington Place, Newcastle upon Tyne NE2 4HH, UK

² Allgemeine Zoologie und Neurobiologie, Ruhr-University Bochum, Germany

³ Department of Optometry and Neuroscience, UMIST, PO Box 88, Manchester M60 1QD, UK

E-mail: s.panzeri@umist.ac.uk

Received 13 March 2002, in final form 21 October 2002

Published 17 January 2003

Online at stacks.iop.org/Network/14/35

Abstract

We derive a new method to quantify the impact of correlated firing on the information transmitted by neuronal populations. This new method considers, in an exact way, the effects of high order spike train statistics, with no approximation involved, and it generalizes our previous work that was valid for short time windows and small populations. The new technique permits one to quantify the information transmitted if each cell were to convey fully independent information separately from the information available in the presence of synergy–redundancy effects. Synergy–redundancy effects are shown to arise from three possible contributions: a redundant contribution due to similarities in the mean response profiles of different cells; a synergistic stimulus-dependent correlational contribution quantifying the information content of changes of correlations with stimulus, and a stimulus-independent correlational contribution term that reflects interactions between the distribution of rates of individual cells and the average level of cross-correlation. We apply the new method to simultaneously recorded data from somatosensory and visual cortices. We demonstrate that it constitutes a reliable tool to determine the role of cross-correlated activity in stimulus coding even when high firing rate data (such as multi-unit recordings) are considered.

1. Introduction

The presence of correlations between the timing of spikes emitted by populations of neurons is a common experimental finding in nervous systems (Gawne and Richmond 1993, de Oliveira

⁴ Author to whom any correspondence should be addressed.

et al 1997, Lebedev *et al* 2000, Mastrorarde 1983). However, whether such correlations really contribute to encoding of sensory stimuli is highly debated. Evidence is controversial on whether stimulus modulations of correlation of firing among small groups of neurons are indeed coding for stimulus identity (Gray *et al* 1989, deCharms and Merzenich 1996, de Oliveira *et al* 1997, Nirenberg and Latham 1998, Villa *et al* 1999, Lamme and Spekreijse 1999, Shadlen and Movshon 1999, Oram *et al* 2001). In our view, some progress in resolving this controversy and in understanding the role of correlations in stimulus coding can be made by addressing questions such as the following: What is the total impact of correlations on information transmission? Does the population use correlation changes to convey information that is not available in firing rate variations of individual cells? Is the information conveyed by the population more than that conveyed by the single cells considered individually? Is correlation a reliable enough mechanism to support robust stimulus discrimination on the basis of a single observation of a neuronal population response?

It is clear that a direct application of information theory (Cover and Thomas 1991) to the analysis of simultaneously recorded neuronal responses can provide quantitative answers to some of the above questions (Nirenberg *et al* 2001, Petersen *et al* 2001, Reich *et al* 2001). However, information theoretic approaches are useful in this respect only if they allow a precise quantification of the modalities with which correlations contribute to neuronal information transmission, rather than just a quantification of the total information transmitted by the population. We have previously developed an information theoretic formalism (the series expansion formalism; see Panzeri *et al* (1999), Panzeri and Schultz (2001) and also DeWeese (1996)) that permits one to quantify separately the information available if each cell were to convey independent information from the deviations from independent information transmission (i.e. synergy–redundancy effects; we refer to *synergy* when there is more information in the simultaneous observation of the population responses than in the sum of information conveyed by each cell individually, and we refer to *redundancy* when the joint neuronal activity conveys less information than the sum of information conveyed by individual cells). These effects were shown to arise from three possible contributions:

- (a) a redundancy contribution due to similarities in the mean response profiles of different cells;
- (b) either a redundant or a synergistic stimulus independent correlation term that reflects the presence of interactions between the distribution of firing rates of individual cells and the average level of cross-correlated activity;
- (c) a synergistic stimulus-dependent correlational contribution that precisely quantified whether stimulus modulations of cross-correlation are used to carry information not present in the firing rates of individual cells alone.

The series expansion approach was based on the assumption that information is transmitted in post-stimulus time windows that are short compared to typical interspike-intervals, so that the average number of emitted spikes per stimulus presentation is low. One important practical advantage of this method is that, since in this limit only individual-cell mean firing rates and pair-wise correlations between spikes are important, its sampling properties are extremely good compared to a brute force evaluation of information directly from the full response probabilities (Schultz and Panzeri 2001). However, the obvious limitations of the series expansion are that it cannot be applied to long post-stimulus windows or large populations, and that it does not quantify the effect of higher order interactions between spikes or between large cell assemblies. This paper presents an investigation of how to overcome the limitations of the series expansion. We derive a new approach that breaks down the mutual information into its coding components

in the exact case, without stopping at a second-order approximation to describe the spike train statistics.

The paper is organized as follows. In section 2 the definition of mutual information is given. In section 3, we give and discuss the main result of this paper: the mathematical expressions for the information breakdown equations. In section 4, we illustrate the meaning of the terms of the information breakdown by applying it to simulated data. In section 5, we discuss the relation of our work to the recent paper of Nirenberg *et al* (2001). In section 6, we introduce (and test with computer simulations) corrections for finite sampling for each of the information components. In section 7 we apply the information breakdown to real data from somatosensory and visual cortices and discuss how the analysis can be applied on low- or high-firing rate data. Finally, in section 8 we discuss the consequences of our results.

2. The information carried by neuronal population responses

In this section we give the main definitions of information transmitted by the neuronal population activity about the stimuli. We consider a time period of duration T , associated with a dynamic or static sensory stimulus, during which the activity of C cells is observed. The neuronal population response to the stimulus in this post-stimulus time window is denoted by a vector \mathbf{r} , each element r_1, \dots, r_C of the vector describing the response of an individual cell. The response of each cell can be described in a number of ways depending on the experimental questions to be addressed. For example, the experimenter might be interested in a *spike count code*. In this case r_c would simply be the spike count of cell c measured in the post-stimulus time window $[0, T]$ on a given trial. Or else, the experimenter might wish to investigate a *spike timing code*. In this case the response r_c would be a sequence of spike arrival times $\{t_i^c\}$, t_i^c denoting the time of the i th spike emitted by the c th neuron in a given trial. Although in the applications presented in this paper we focus only on spike count codes, the equations we derive will be valid for any choice of neuronal code \mathbf{r} , including spike timing codes.

For a given choice of code, following Shannon (1948), we can write down the mutual information transmitted by the population response about the whole set of stimuli as

$$I(\mathcal{R}; \mathcal{S}) = \left\langle \sum_{\mathbf{r} \in \mathcal{R}} P(\mathbf{r}|s) \log_2 \frac{P(\mathbf{r}|s)}{P(\mathbf{r})} \right\rangle_s. \quad (1)$$

Mutual information quantifies how well an ideal observer of neuronal responses can discriminate between all the different stimuli, based on a single trial. In equation (1) the summation is over all possible population responses. Each different stimulus is denoted as s . The angular brackets indicate the average over different stimuli, $\langle A(s) \rangle_s \equiv \sum_{s \in \mathcal{S}} P(s) A(s)$. $P(\mathbf{r}|s)$ is the probability of observing a particular response \mathbf{r} conditional to stimulus s , and $P(\mathbf{r}) = \langle P(\mathbf{r}|s) \rangle_s$ is its average across all stimulus presentations. The probability $P(\mathbf{r}|s)$ is determined experimentally by repeating each stimulus in exactly the same way on many trials, while recording the neuronal responses.

3. The exact breakdown of information into its coding mechanisms

The Shannon's mutual information, as expressed in equation (1), quantifies the total information transmitted by the neuronal population activity. However, it tells us nothing about the specific contribution of cross-neuronal correlations to the total transmitted information. In addition, it does not tell us directly whether cross-correlations make the code redundant or

synergistic. In this paper we perform and present the mathematical work that permits one to quantify precisely the impact of cross-cell correlation on information transmission.

The first step needed in order to quantify the importance of cross-correlations is the construction of the stimulus-conditional *statistically independent response probability*, in addition to the simultaneous joint response probability $P(\mathbf{r}|s)$. By definition, two stochastic variables are statistically independent if their joint response probabilities equal the product of the individual probabilities⁵. Hence, the probability of independent population responses that ignores cross-cell correlations can be obtained by taking the product of the probability distributions of individual cells:

$$P_{ind}(\mathbf{r}|s) = \prod_{c=1}^C P(r_c|s). \quad (2)$$

The stimulus-unconditional independent response probability is computed as follows:

$$P_{ind}(\mathbf{r}) = \langle P_{ind}(\mathbf{r}|s) \rangle_s. \quad (3)$$

The independent response probability $P_{ind}(\mathbf{r}|s)$ provides, for each response \mathbf{r} , a predictor that can be used to test how much the real response probability $P(\mathbf{r}|s)$ deviates from the null hypothesis of no cross-correlation (i.e. $P(\cdot) = P_{ind}(\cdot)$). In this sense, $P_{ind}(\mathbf{r}|s)$ is conceptually similar to shuffle predictors used to test and quantify correlation in cross-correlogram (CCG) or joint-Peri-stimulus-time-histogram (JPSTH) analyses (Aertsen *et al* 1989). However, the mutual information in equation (1) depends on the full response probability distributions, and not only on a second-order quantification of neuronal statistics. Hence, to exactly describe the impact of correlation on information, it is necessary to define a response-related correlation measure that goes beyond second-order correlation and that takes naturally into account all the higher order statistics of the joint response probability distribution. In analogy with classic CCG analyses, the independent response probability can be used to quantify the *strength* of correlation between cells during a response \mathbf{r} . A natural definition for the normalized cross-correlation strength of population response \mathbf{r} is the following:

$$\gamma(\mathbf{r}|s) = \begin{cases} \frac{P(\mathbf{r}|s)}{P_{ind}(\mathbf{r}|s)} - 1 & \text{if } P_{ind}(\mathbf{r}|s) \neq 0 \\ 0 & \text{if } P_{ind}(\mathbf{r}|s) = 0 \end{cases} \quad (4)$$

where $\gamma(\mathbf{r}|s)$ quantifies how much the probability that neurons emit a response \mathbf{r} is higher than that expected in the uncorrelated case, normalized to the probability of event \mathbf{r} expected in the uncorrelated case. Positive values of this coefficient mean that the individual cell responses r_1, \dots, r_C , composing the population response \mathbf{r} , happen together during the same trial more frequently than if there was no cross-cell correlation. This correlation coefficient goes beyond second-order (pairwise) correlations and it takes into account all possible interaction orders between all neurons in the population. It is worth stressing that the correlation strengths $\gamma(\mathbf{r}|s)$ introduced above are response-dependent quantities. The $\gamma(\cdot)$ values provide a complete description of the deviations of neuronal response from independence because there is one $\gamma(\mathbf{r}|s)$ for each possible population response \mathbf{r} . Second-order quantifications of neuronal correlation, such as spike count Pearson correlation coefficient (Zohary *et al* 1994) or JPSTHs (Aertsen *et al* 1989), provide, instead, an average across trials of pair-wise only correlations. It is apparent that in some cases the second-order description of correlation might be incomplete,

⁵ As will be shown below, statistical independence of responses of a population does not necessarily imply that the population conveys information independently: there could be redundancy due to similarities in the response profiles of single cells. For this reason, in this paper we will make a distinction between statistical independence and informational independence.

for example, in the presence of some high order interactions between large groups of spikes or cells.

A final remark that should be made is that, since $\gamma(\mathbf{r}|s)$ are response-dependent quantities, they have to satisfy (for each stimulus s) a zero-sum constraint:

$$\sum_{\mathbf{r}} P_{ind}(\mathbf{r}|s) \gamma(\mathbf{r}|s) = 0. \quad (5)$$

This constraint stems directly from the normalization to one of response probabilities. We also need to introduce a coefficient that quantifies how similar the stimulus modulation of responses of individual cells is. This parameter is important for describing population coding: if all cells have similar stimulus selectivity, it is likely that the population code is redundant. To quantify the similarities of individual cell responses across stimuli, we hence introduce a ‘signal similarity’ coefficient⁶. In a way analogous to $\gamma(\cdot)$, the signal similarity coefficient is defined as follows:

$$v(\mathbf{r}) = \begin{cases} \frac{P_{ind}(\mathbf{r})}{\prod_c P(r_c)} - 1 & \text{if } (\prod_c P(r_c)) \neq 0 \\ 0 & \text{if } (\prod_c P(r_c)) = 0. \end{cases} \quad (6)$$

This coefficient is different from zero if ‘signals’ coming from individual neurons are either positively correlated (i.e. similar) or negatively correlated. It does depend only on response probabilities of individual cells and not on within-trial cross-correlation. It quantifies how similar across stimuli the responses probabilities of the individual cells that form the population are. We will show, in what follows, that this quantity is important to describe the information transmitted by the population. We note that the signal similarity coefficient is also response dependent, and it hence satisfies a normalization condition similar to equation (5), as follows:

$$\sum_{\mathbf{r}} \left(\prod_c P(r_c) \right) v(\mathbf{r}) = 0. \quad (7)$$

By rewriting the total information in terms of these quantities we were able to write the total information in components, each reflecting the contribution of a different coding mechanism:

$$I(\mathcal{R}; \mathcal{S}) = I_{lin} + I_{sig-sim} + I_{cor-ind} + I_{cor-dep}. \quad (8)$$

The meaning and mathematical expression of each of the components is described in the following. It is interesting to note that, as explained in what follows, those components have the same meaning as the corresponding components of the previously derived series expansion quantities (Panzeri *et al* 1999, Panzeri and Schultz 2001), but they are exact, and not an approximation of fixed order.

In the following we express the information components in a form that makes the understanding of the coding mechanisms and the comparison with the series expansion approximation simpler. However, we derived an equivalent and equally useful mathematical expression for these quantities as a difference of entropies and entropy-like quantities. These expressions are convenient for numerical implementation of the analysis, and are reported in appendix A. It is important to note that the sum of the components of this information breakdown *exactly* equals the total mutual information only if each individual term of the information breakdown is summed over all possible population responses. (see appendix A). In the series expansion approach, it was instead possible to truncate the expansion at a certain order.

⁶ We chose to name this coefficient ‘signal similarity’ to be consistent with previous studies (Gawne and Richmond 1993, Panzeri *et al* 1999).

3.1. The linear term—information transmitted by independent cells

The first term of the information breakdown I_{lin} is the information obtained if each cell were to convey independent information. In this case there would be no redundancy or synergy and the total information transmitted by the population would be a linear sum of the information conveyed by each individual cell:

$$I_{lin} = \sum_c \sum_{r_c} \left\langle P(r_c|s) \log_2 \frac{P(r_c|s)}{P(r_c)} \right\rangle_s. \quad (9)$$

Deviations from independent information transmission (i.e. synergy or redundancy effects) are expressed by the other three terms, considered in what follows.

3.2. Signal-similarity term

If there is any redundancy between cells, equation (9) can overestimate the information. Redundancy can be present even in the absence of cross-correlation if there are similarities in the distribution across stimuli of stimulus-conditional response probabilities of individual cells. The total impact of signal similarities on information transmission is a logarithmic function of the signal correlation coefficient, equation (6), and it is expressed as follows:

$$I_{sig-sim} = \frac{1}{\ln 2} \sum_r \left(\prod_c P(r_c) \right) \left\{ v(r) + (1 + v(r)) \ln \frac{1}{1 + v(r)} \right\}. \quad (10)$$

It is easy to see that signal similarity cannot lead to synergy: equation (10) is always less than or equal to zero, since $f(x) = x - (1 + x) \ln(1 + x)$ has a global maximum at $f(0) = 0$. It is equal to zero if and only if the signal similarity is precisely zero.

The first two terms of the information expansion, I_{lin} and $I_{sig-sim}$, depend only on the response properties of each cell evaluated individually, and not on cross-cell correlation. If the population responses are statistically independent ($P(\cdot) = P_{ind}(\cdot)$, i.e. $\gamma(\cdot) = 0$), these are the only non-zero contributions to the information breakdown. The sum of I_{lin} and $I_{sig-sim}$ quantifies how much information can be obtained from the neurons evaluated individually, without reference to the simultaneous activity of other neurons.

The next two terms in the information breakdown, given in what follows, are the correlational terms. They depend also on the correlation strength $\gamma(\cdot)$ of simultaneous response. They can be non-zero only if the correlation strength $\gamma(r|s)$ is different from zero for some response r or stimulus s . Hence they express any further effects that cross-cell correlations might have beyond that accounted for by individual cell properties. The sum of the two correlational terms $I_{cor-ind}$ and $I_{cor-dep}$ quantifies the amount of information genuinely available from the correlated activity of the whole population.

3.3. Stimulus-independent correlational component

Even if not stimulus-modulated, cross-correlations can still affect the neuronal code through an interaction between cross-cell correlation and signal similarity (Abbott and Dayan 1999, Oram *et al* 1998). The effect of stimulus-independent correlations to information transmission is

$$I_{cor-ind} = \sum_r \langle P_{ind}(r|s) \gamma(r|s) \rangle_s \log_2 \frac{1}{1 + v(r)}. \quad (11)$$

The first multiplicative factor reflects the effect of cross correlation, but these correlations are averaged across stimuli (weighted proportional to the probability of each response). The

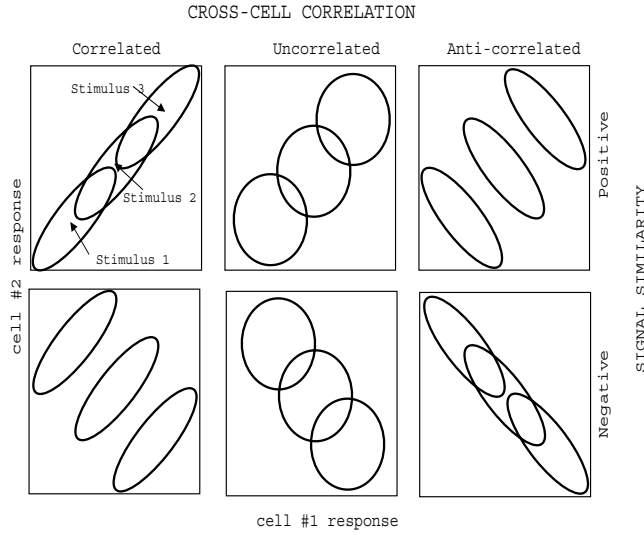


Figure 1. The effect of stimulus-independent correlations on information encoding. Each panel sketches joint distributions of responses of two hypothetical cells to three different stimuli. Each ellipse indicates the set of responses elicited by a given stimulus. In the upper panels, there is positive signal similarity (i.e. individual cell responses to each stimulus are positive correlated) whereas in the lower panels there is negative signal similarity. The sign of cross-correlations between the joint responses differs across columns of this figure. In general, if cross-cell correlation and signal similarities have opposite signs, the effect of stimulus-independent correlations increases the information about stimuli, because the joint response probabilities to each stimulus become more separated. If instead cross-cell correlation and signal similarities have the same sign, stimuli are less discriminable than in the other case. Redrawn from Oram *et al* (1998) and Petersen *et al* (2001).

logarithmic term depends instead on signal similarity. For a given response \mathbf{r} , the stimulus-independent correlational component is positive (synergistic) when signal similarity and cross-correlation have the opposite sign, and negative (redundant) otherwise (see figure 1 for an intuitive explanation). It is important to stress that, because of the zero-sum constraint in equations (5) and (7), the response-dependent coefficients γ and ν will change sign for different responses \mathbf{r} . Because of this, unlike for the other information components, it is not straightforward to understand whether the net effect of summing over all responses \mathbf{r} will result in a positive or a negative $I_{cor-ind}$. The series expansion formalism (Panzeri *et al* 1999) tells us that, if the response time window is short, the sign of $I_{cor-ind}$ will chiefly depend on that of second-order correlations. However, for long time windows, high order effects might be dominant in some cases.

3.4. Stimulus-dependent correlational component

Finally, the stimulus-dependent correlational component is

$$I_{cor-dep} = \sum_{\mathbf{r}} \left\langle P_{ind}(\mathbf{r}|s)(1 + \gamma(\mathbf{r}|s)) \log_2 \frac{\langle P_{ind}(\mathbf{r}|s') \rangle_{s'} (1 + \gamma(\mathbf{r}|s))}{\langle P_{ind}(\mathbf{r}|s') \rangle_{s'} (1 + \gamma(\mathbf{r}|s'))} \right\rangle_s. \quad (12)$$

It can be proved by means of basic information-theory inequalities that this term is non-negative, and it is zero if and only if, for any given response \mathbf{r} , the correlation strength $\gamma(\mathbf{r}|s)$ is stimulus independent. Therefore this term measures how well stimulus identity is ‘tagged’ in differences in trial-to-trial spike correlations across the stimuli.

It is interesting to note that the exact decomposition of neuronal population information into coding components is exactly analogous to the one previously performed in the short-time limit and expressing only up to pair-wise correlations. The exact breakdown formulae can be simply derived from the second-order series expansion ones by appropriately replacing the second-order γ and ν (defined in Panzeri and Schultz (2001)) with the exact ones, equations (4) and (6), and by replacing the instantaneous rates (which, in the short-time bin size limit are the probabilities of a spike from a cell in a time considered independently of other bins and cells) with the single cell response probabilities (i.e. the probability of each cell response considered independently of the other cells). Despite this analogy, there is, however, one important difference between the exact breakdown and the series expansion equations. In the exact case, the sum in the equation extends to all the possible population responses. In contrast, only first- and second-order moments of the response probability distributions enter the approximated series expansion equations, and the approximation can be made better and better by inserting successively higher terms.

Finally, we note that we verified that, when taking series expansion approximation of each individual term of the exact information breakdown, we correctly retrieved the expression of the corresponding series expansion component. This is an important consistency check and it means that, if the time window considered is short with respect to typical inter-spike intervals, the exact information breakdown and the series expansion will give similar numerical results. An explicit example of this equivalence is given in figure 5 by means of an application to real neuronal pairs with low-firing rate.

4. Application of the exact information breakdown to simulated neuronal pairs

In order to illustrate and check our new analysis method, we extensively applied the method to simulated responses of neuronal pairs reflecting different ways of encoding information through correlations. In this section we present some of these applications on simulated data with the aim of facilitating the understanding of the various terms of the information breakdown. In the following examples we considered pairs of cells and we considered the information carried by a joint *spike count* code, because the following real data examples also computed information carried by spike counts. However, it is important to stress again that our formalism is general and can be applied, as it is, to spike timing population codes as well. The information content of each component was computed on the basis of 256 simulated trials per stimulus, and corrections for finite sampling (described in the next section) were applied.

4.1. Application to Poisson spike trains

We started by applying the method to simulated uncorrelated data. In this case all the correlational components are expected to carry zero information. We simulated the spike trains of two cells responding to two different stimuli. In each trial, we generated, independently for each cell, 1 second of simulated data according to a stationary Poisson process with mean rates reported in figure 2(a) (left panel).

When applying the information breakdown, we found that both correlational components conveyed zero information. This correctly reflects the fact that the two simulated neurons were uncorrelated (for illustration, CCGs are plotted in the central panel of figure 2(a)). Since the mean response profile of the two neurons was similar across the stimulus set, we expected to find some negative signal-similarity contribution leading to redundant information. Indeed, we found that, although the dominant information term in this case was the linear one, equation (9),

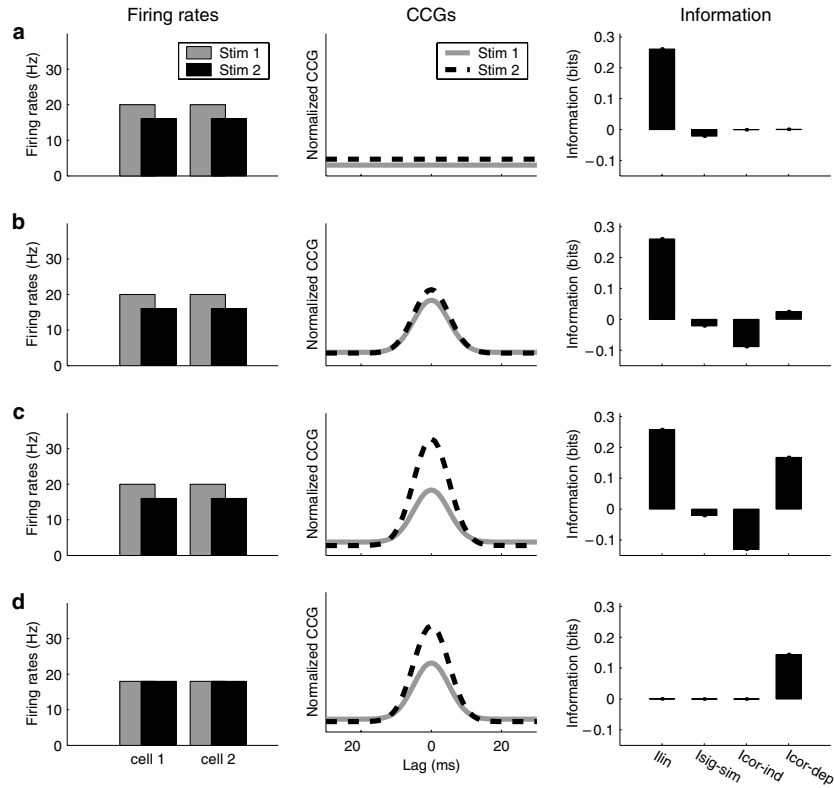


Figure 2. Components of the information breakdown. The meaning of the information components is illustrated by applying the exact information breakdown to simulated neuronal pairs responding to two stimuli and reflecting different ways of encoding information through correlations. Data were generated as follows. We first created, independently for each cell, spikes from a Poisson process. We then generated spikes from a third Poisson process, and these spikes were then added to both cells in order to create cross-correlation. In order to avoid synchronization with infinite time precision, the shared spike times added to the second cell were shifted together in time by a random amount chosen anew for each trial from a zero-mean Gaussian distribution with standard deviation of 5 ms. In all cases, a joint spike count code over a 1 s long window was considered. The left panels plot the mean firing rate of the two cells; the central panels plot the CCG (computed analytically from knowledge of the simulated processes); the right panels report the values of each information component. The terms I_{lin} , $I_{sig-sim}$, $I_{cor-ind}$ and $I_{cor-dep}$ refer to the various components of the information breakdown as described in section 3. (a) Uncorrelated spike trains; a Poisson process is used for each cell and no shared spikes are added. (b) Correlated data with weak stimulus modulation of correlation; in this case the ratio of independent versus shared spikes was approximately the same for both stimuli and hence the correlation strength $\gamma(\cdot)$ was only weakly stimulus modulated. The simulation parameters were as follows. For both cells, the mean rate of the independently generated spikes was 10 Hz for the first stimulus and 8 Hz for the second stimulus. The mean rate of the shared spikes was 10 Hz to the first stimulus and 8 Hz to the second one. (c) Correlated data with strong stimulus modulation of correlation; in this case the ratio of independent versus shared spikes was much higher for the first stimulus than for the second. As a consequence, the $\gamma(\cdot)$ coefficients were stimulus dependent. The data were generated as follows. For both cells, the mean rate of the independently generated spikes was 10 Hz for the first stimulus and 1 Hz for the second stimulus. The mean rate of the shared spikes was 10 Hz to the first stimulus and 15 Hz to the second one. (d) Correlated data with strong stimulus modulation of correlation and no rate modulation; the ratio of independent versus shared spikes was much higher for the first stimulus than for the second, but the mean firing rate of each cell did not change across cell or stimulus. For both cells, the mean rate of the independently generated spikes was 9 Hz for the first stimulus and 1 Hz for the second stimulus. The mean rate of the shared spikes was 9 Hz to the first stimulus and 17 Hz to the second one.

there was a negative contribution from the signal-similarity information term. The signal similarity redundancy was relatively small, as predicted by equations (6) and (10).

4.2. Application to correlated data

We then applied the method to correlated data and checked whether it is able to disambiguate correctly between the two possible stimulus-dependent and stimulus-independent correlational mechanisms.

For each simulated stimulus, we generated correlated neuronal response pairs according to the following simple procedure (Brody 1999). We first created, independently for each cell, spikes from a Poisson process with a certain mean rate (in principle different for each stimulus and cell). We then generated spikes from a third Poisson process (characterized by a mean rate in general different from that of the first two simulated processes). The spikes generated from this third process were then added to both cells, and this led to cross-correlation.

The amount of correlation was modulated across stimuli by varying the number of simulated shared spikes with respect to the spikes simulated independently for each cell.

4.2.1. The case of weak stimulus modulation of correlation. We first generated correlated data where the fraction of correlated spikes was held approximately constant across stimuli. This is reported in figure 2(b). For illustration, CCGs to both stimuli are plotted in the central panel of figure 2(b).

The overall mean responses of the two simulated neurons were similar across stimuli (figure 2(b); left panel). As a result of the procedure used in the simulation, stimulus modulations of correlation density were very small; see the CCGs plotted in the central panel of figure 2(b). The information analysis was able to describe well the type of correlational encoding used by the simulated data. We found that the stimulus-independent correlational component was of appreciable size. The stimulus-independent correlational component was negative because it originated from correlated spike trains from neurons with similar tuning to stimuli (see figure 1 and equation (11)). The stimulus-dependent correlational component was very small, correctly reflecting the nature of the simulated data.

4.2.2. The case of strong stimulus modulation of correlation. We then generated correlated data with a fraction of shared spikes that was strongly stimulus modulated (figure 2(c); see CCGs in the central panel). When performing the information analysis, we found that the values of I_{lin} , $I_{sig-sim}$ were similar to those of the previous simulation. This is consistent because the mean firing rates were equal to the previous case. The stimulus-independent correlational component $I_{cor-ind}$ was again negative, reflecting an average positive cross-cell correlation strength, and single cell responses presenting similarities across stimuli. The difference with the previous simulation is that now the stimulus-dependent correlational component conveys a large amount of information (61% of the total mutual information).

4.2.3. The case of purely correlational information (no rate modulation). Finally, we wanted to show that the two information components I_{lin} and $I_{sig-sim}$ do not incorrectly contain some information that is conveyed by the cross-cell correlated activity. To clarify this point, we applied the information breakdown to simulated data which presented no mean response variation across stimuli or cell, but that modulated the cross-correlation strength to different stimuli (figure 2(d)). Results of the information analysis are reported in figure 2(d). In this case only the stimulus-dependent correlational component conveys information. Hence, also in this case the simulated coding mechanism is correctly revealed by the information breakdown.

In summary, for all simulated cases the information breakdown method was able to correctly describe the nature of the underlying cross-correlation coding mechanism.

5. Quantifying the information that can only be decoded considering the stimulus variations of correlated firing

Equation (1) quantifies the information transmitted (or *encoded*) by the neuronal responses, and it does not specify any form or model of the downstream neuronal decoder that is going to read the output of the considered population. The data-processing inequality (Cover and Thomas 1991) shows that the information that any decoder can extract from the neuronal responses is less than or equal to that encoded. Thus equation (1) provides an upper bound that can be used to test the accuracy of any decoding model.

An interesting question is whether a decoder that ignores the response correlations and uses an uncorrelated probability model to decode the spike train would be able to extract all the information present in the population responses (Wu *et al* 2001). A new and very interesting way to quantify the effect of using the no-correlation hypothesis in interpreting correlated data has been proposed by Nirenberg *et al* (2001). The idea is as follows. Information about a stimulus can be thought of as the average number of binary digits (or yes/no questions) necessary to describe the stimulus minus the average number of binary digits needed to identify the stimulus given the observation of a response (assuming optimal coding strategy). When treating the cells as independent, we use the approximate probability $P_{ind}(r|s)$ instead of the true one, $P(r|s)$. The incorrect statement provided by the approximated model increases the complexity of the description of the stimulus given the response (Cover and Thomas 1991). The idea of Nirenberg *et al* (2001) is to measure the information loss due to the use of the uncorrelated model as the average increase in the length of the binary digit code needed to characterize the stimulus given the response, averaged over responses (Nirenberg *et al* 2001, Cover and Thomas 1991):

$$\Delta I \equiv D(P(s|r) \| P_{ind}(s|r)) \equiv \sum_r P(r) \sum_s P(s|r) \log_2 \frac{P(s|r)}{P_{ind}(s|r)} \quad (13)$$

where D is conditional relative entropy (see Cover and Thomas 1991, p 22, equation (2.61)). In the context of this work on the role of correlation in encoding, we were interested in investigating the relationship between ΔI , the information lost by the uncorrelated ‘yes/no’ decoder, and the correlational components of the information breakdown. By using the Bayes rule and performing some algebra, one can verify that ΔI exactly equals the information carried by the stimulus modulations of correlation, equation (12). This means that ΔI , the information neglected by the ‘yes/no’ decoder based on the uncorrelated model, is the information specifically carried by the stimulus-modulation of the correlation strength.

6. Sampling properties and bias corrections

In practice, the information components must be estimated from experimental probabilities obtained from a limited number N_s of repeated presentations of each stimulus s . This leads to a systematic error (or bias) in the estimate of the mutual information and of its components; the size of the bias being inversely related to the number of trials (Panzeri and Treves 1996). It is possible to largely alleviate this problem by computing analytic expressions for the limited sampling biases and subtracting it from the raw estimates obtained directly from the experimental probabilities. These corrections have been computed previously only for the total information (Panzeri and Treves 1996, Victor 2000).

In this section we focus on how to compute effective bias corrections also for the individual components. This step is necessary to obtain reliable estimates of the information carried by each component in practical applications of spike train analysis. We will present some analytical evaluation of the bias of each component (with more details reported in appendix B). We will also use computer simulations to evaluate the accuracy of the bias correction procedures, and to illustrate rules of thumb that guide the evaluation of the amount of trials necessary to ensure that each component is computed with reasonable accuracy. For validation of the method, we simulated both correlated and uncorrelated neuronal pairs, and we studied the precision of the information estimates as a function of the number of simulated trials, both before and after applying the bias corrections derived here. The results for uncorrelated data (generated as in section 4.1) are reported in figure 3, and the results for correlated data, simulated as in section 4.2.1), are plotted in figure 4. We considered the information carried by the joint spike count of the simulated neuronal pairs in a 1 s long window. For the information calculation, we compressed the number of possible responses by grouping the response firing rates into four classes per neuron. The boundaries of each class were chosen so as to obtain equipopulated responses across bins. (This procedure is computationally simple and is effective at preserving information because it maximizes the response entropy for a given number of bins. For more accurate procedures for minimizing the information loss due to response quantization, see Dimitrov and Miller (2001).)

It is useful to start by briefly considering the results for the analytic bias corrections for the mutual information and the rules of thumb for its effective correction. The expression for the upward bias for the transmitted information has been calculated in several previous studies (Panzeri and Treves 1996, Victor 2000), and it is as follows:

$$\text{Bias}[I] = \frac{1}{2N \ln 2} \left\{ \sum_s (\tilde{R}_s - 1) - (\tilde{R} - 1) \right\} \quad (14)$$

where N is the total number of trials (across all stimuli) and \tilde{R} is the number of relevant response classes across all stimuli (i.e. the number of different responses r with non-zero probability $P(r)$ of being observed). \tilde{R}_s is the number of relevant response classes to stimulus s (i.e. the number of different responses r with non-zero stimulus-conditional probability $P(r|s)$). In the simulations presented in figures 3 and 4, there were a total of $4 \times 4 = 16$ population response classes. It can be seen that the information estimates require at least ≈ 64 trials per stimulus in order to become reasonably accurate after bias correction⁷.

We repeated the simulations by systematically varying the total number of response classes. We found, consistently across different simulations, that the information was reasonably well bias-corrected if N_s was at least $\approx 2-4$ times bigger than the number of population response classes R . This provides a stronger constraint on the size of the population than can be analysed given a certain number of trials per stimulus N_s . If there are C cells, each providing \tilde{R}_c possible responses ($c = 1, \dots, C$), then the total number of possible responses of the population scales as the product of the number of possible responses of individual cells:

$$\tilde{R} \simeq \prod_c \tilde{R}_c. \quad (15)$$

⁷ We computed bias corrections from equation (14) by using the Panzeri and Treves (1996) procedure to count the number of relevant bins. This may not be the most effective procedure to estimate the bias. It is possible that future application of recent advanced techniques (Victor 2002) will further reduce the number of trials needed to compute the information components. However, the analytical expressions computed here for the averaged bias of each component are useful to obtain an estimate of the relative magnitude and behaviour of the sampling errors of each component, whatever the procedure chosen to control for sampling problems.

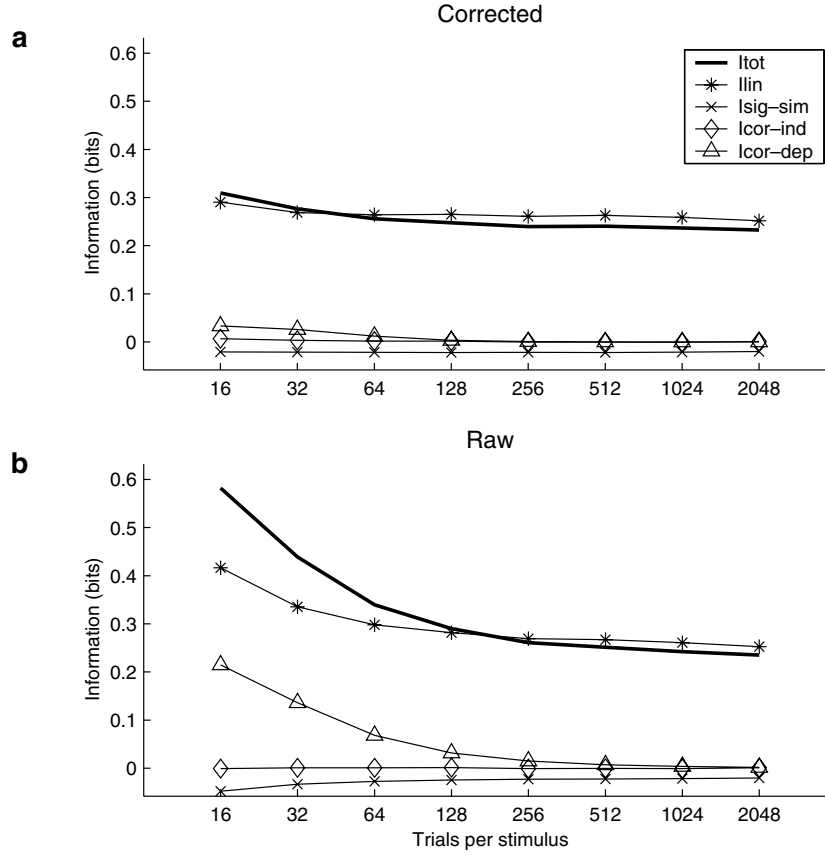


Figure 3. Sampling behaviour of the information components. We generated uncorrelated simulated data (see section 4.1) and we tested how the estimates depend on the data size. The value of each information component is plotted as a function of the number of trials per stimulus. The full line (denoted as I_{tot} in the legend) is the total mutual information, equation (1). The other lines represent the values of each information component (see legend). (a) Values of information components obtained after subtracting the bias corrections described in the text. (b) ‘Raw’ values of information components obtained without using any bias corrections. Results were averaged over 100 random repetitions of the simulation.

We now consider the bias of each individual component. We start this study by considering the linear term in equation (9). It is the sum of single cell information; hence its bias correction is easily obtained from Panzeri and Treves (1996) as a sum of single cell contributions:

$$\text{Bias}[I_{lin}] = \frac{1}{2N \ln 2} \left\{ \sum_s \left[\sum_c (\tilde{R}_{c;s} - 1) \right] - \left[\sum_c (\tilde{R}_c - 1) \right] \right\} \quad (16)$$

where $\tilde{R}_{c;s}^s$ are the relevant bins corresponding to the probability distribution $P(r_c|s)$. The magnitude of the bias of this term is dictated by the sum of possible responses $\sum_c \tilde{R}_c$. This term is one of the most two biased components, but it is not as biased as the whole information. In fact, the bias of total information depends instead on its product, equation (15). In the simulations considered in figures 3 and 4, there were eight classes in total in the sum (four per neuron), and I_{lin} was well corrected with at least 32 trials per stimulus.

The signal-similarity component and the stimulus-independent correlation component are very weakly biased. We computed bias correction terms for $I_{sig-sim}$ and $I_{cor-ind}$. They are

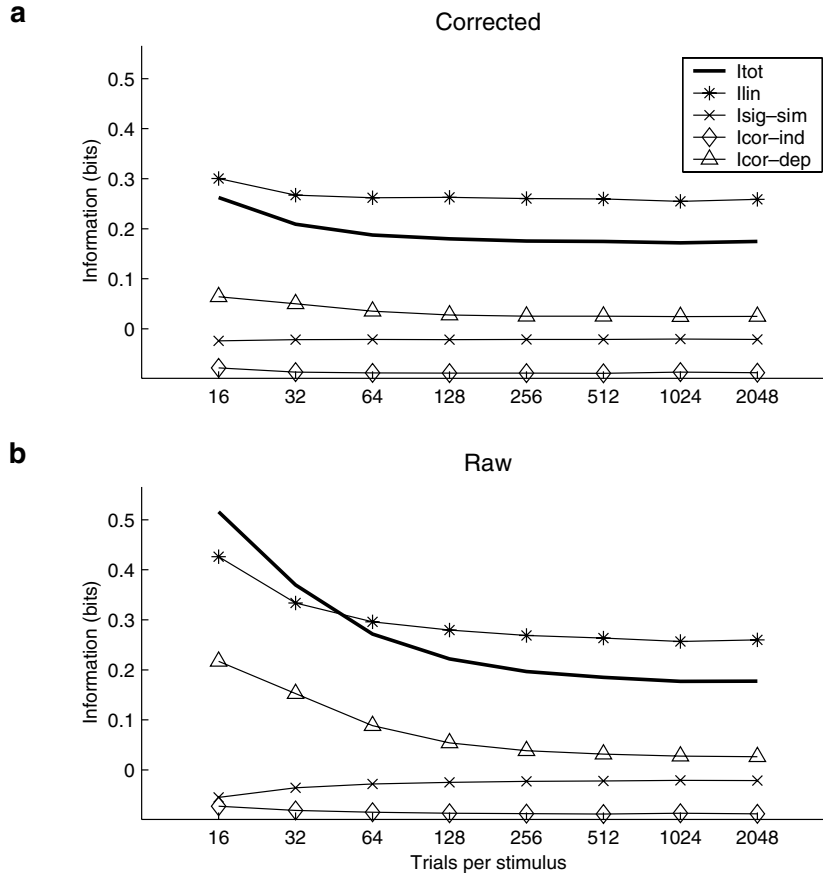


Figure 4. The sampling behaviour of the information components. Conventions are as presented in figure 3. However, this time the sampling behaviour was studied using a set of correlated simulated data (see section 4.2.1).

reported in appendix B. Computer simulations presented in figures 3 and 4 show that, if the number of trials per stimulus is at least 16–32, the bias of $I_{sig-sim}$ and $I_{cor-ind}$ is very little. The application of the correction term permits one to obtain very precise estimates of both $I_{sig-sim}$ and $I_{cor-ind}$. The main reason why $I_{sig-sim}$ and $I_{cor-ind}$ are much less biased than I_{lin} and $I_{cor-dep}$ is that, unlike the other two components, they depend only on the unconditional response probabilities $P(r)$, and not on the (less well sampled) stimulus conditional response probabilities $P(r|s)$.

We consider finally the stimulus-dependent correlational term $I_{cor-dep}$. This is the most biased of all components (figures 3 and 4). In fact its bias can be expressed approximately as a difference between the bias of the total information I and the bias of the linear information I_{lin} . This is because $I_{cor-dep}$ is expressed as a difference between the total information I and the linear information I_{lin} and of other terms of negligible bias (see appendices). Hence, correcting $I_{cor-dep}$ for limited sampling requires roughly the same number of trials needed to correct for the total information (otherwise the I term is not under statistical control). However, the overall magnitude of its bias is less than that of the total information (because of the subtraction of the bias of I_{lin}). In the simulations reported in figures 3 and 4, $I_{cor-dep}$ required some ≈ 32 –64 trials per stimulus to be well corrected, similarly to the total mutual information.

In summary, a rule of thumb for the amount of data required to compute the information breakdown is as follows. When there are more trials (at least 2–4 times) per stimulus than population response classes, then the total mutual information can be reasonably well corrected for finite sampling. In this case, all other components are also well enough sampled. The two most biased components are I_{lin} and especially $I_{cor-dep}$. When there are enough data to correct for the bias of I_{lin} and $I_{cor-dep}$, $I_{sig-sim}$ and $I_{cor-ind}$ are also well corrected and cause no problem.

It is important to note that the above bias equations and behaviours are specific to the case where the independent probability distribution, equation (3), is obtained taking the product of probabilities of individual cells. Another numerical approach used to compute $P_{ind}(\cdot)$ is to combine responses of different cells from trials to the same stimulus, but shuffled or shifted-by-one. This is the approach taken, for example, by Nirenberg *et al* (2001). We chose to compute $P_{ind}(\cdot)$ as a product of individual cell probabilities because it gave a slightly better performance of the numerical evaluation. In particular, computing $P_{ind}(\cdot)$ as a product ensures that, for a given r and s , $P_{ind}(r|s) = 0$ implies that $P(r|s) = 0$. This is an important consistency property for the information breakdown formulae, and it might occasionally not be fulfilled by e.g. the shifting-by-one procedure if the number of data available is low.

7. Application to real neurophysiological recordings

In this section we apply the new method to real simultaneous neuronal recordings in sensory cortices. With this application we wish to demonstrate the use of the new exact information breakdown method, and outline some of the neurophysiological questions that can be addressed with it.

We previously introduced a series expansion method that performs the same type of decomposition of information into coding components as the exact information breakdown presented here. The series expansion method has very good sampling properties (Schultz and Panzeri 2001), but can only be applied to systems with relatively low firing rates, with only few spike emitted in a typical trial. For this reason, we find it interesting to apply the new method (and compared it to the series expansion) on two data sets with different mean response properties. We first present an application of the exact information breakdown to neuronal pairs in rat somatosensory cortex. This dataset presents low firing rates. In the second example, we present an application of the information breakdown to multi-unit activity (MUA) simultaneously recorded from two sites in visual area MT of awake monkeys.

7.1. Analysis of neuronal pairs in rat somatosensory cortex

We first apply the new method to neuronal pairs recorded in rat somatosensory cortex to address whether cross-cell correlation contributes to the coding of the location of a stimulus applied to a single whisker on the snout. Since typical spike counts for barrel cortical neurons are low, the mutual information was well approximated by a series expansion method (Panzeri *et al* 2001, Petersen *et al* 2001). Since both the exact breakdown and the series expansion method should be applicable in this case, we will illustrate the relative advantage of each procedure. We will also discuss which questions can be addressed by the joint application of both the exact information breakdown and the series expansion method.

We analysed 52 pairs of neurons recorded from the barrel field of somatosensory cortex of urethane-anaesthetized rats. These data were kindly made available to us by M E Diamond and M Lebedev (Lebedev *et al* 2000). Both neurons were located in barrel-column D_2 , and each neuron was recorded from a different electrode. Vibrissae C_{1-3} , D_{1-3} and E_{1-3} were

stimulated one at a time, using a piezoelectric wafer controlled by a voltage generator. The stimulus was an up–down step function of $80\ \mu\text{m}$ amplitude and 100 ms duration, delivered once per second, 50 times for each vibrissa, 3 mm from their base (see Lebedev *et al* (2000) for full details). The analysis was based on a post-stimulus cumulative time window that was increased in 10 ms steps from 0 to 40 ms. Since we did not have enough trials per stimulus to compute the joint spike timing information with the exact information breakdown, we considered the information transmitted by the joint spike count code. The ‘response’ on each trial was, for each cell, simply the number of spikes occurring in the time window. We applied the finite sampling corrections described in the previous section.

We explored the nature of the population code employed by these neurons by examining the contributions from the separate terms of the exact information breakdown and of the series expansion. The results of information analysis are reported in figure 5. The exact information breakdown and the series expansion method give very similar results, as expected by the fact that firing rates are low. In both cases most of information is conveyed by the linear term (the sum of single cell information). There is appreciable redundancy in both the signal-similarity information term, and in the stimulus-independent correlational component. This is because the mean response profiles of neurons located in the same barrel are similar to each other, and because there is some positive cross-correlation. The values of I_{lin} , $I_{sig-sim}$ and $I_{cor-ind}$ computed with the two methods coincided within 2%. The stimulus-dependent correlational component was relatively small for both methods, although it was bigger when computed in long windows with the exact information breakdown. At 40 ms post-stimulus, it was 0.01 bits bigger than when computed with the series expansion. This amount is of the order of the residual bias that is expected in this data range for this component (figures 3 and 4); hence it is likely that part of this difference (which is, however, small and present only for longer windows) is due to sampling errors. Hence, we conclude that both the series expansion and the exact information breakdown give equivalent results on this dataset.

What is the neurophysiological conclusion that we can get from this comparison? The series expansion quantifies the contribution to information processing of only the second-order statistic among spikes, whereas the exact method quantifies the contribution of all-order statistics. Thus, the conclusion that we derive from this comparison is that second-order statistics is sufficient to describe this population code. The example we presented suggests that the joint use of both the series expansion method and the exact information breakdown is one of the tools that can be used to address the role of high order statistics in population coding, a subject of current investigation (Laubach *et al* 1999, Nakahara and Amari 2002).

7.2. Analysis of information encoded in cross-correlation of multi-unit activity in visual cortex of awake monkeys

The salient feature of the new information breakdown proposed here is that, if sufficient data are collected, it can be used to assess the information content of cross-cell correlation for any type of neuronal recording, independently of the mean firing rate level. This is a most useful property because several studies reporting evidence on the role of correlation in the cortex (Gray *et al* 1989, Fries *et al* 1997, Lamme and Spekreijse 1999) are based on recording of multi-unit activity (MUA). MUA involves typically high firing rates and collection of spikes from several cells, and hence the assumptions of the series expansion method are certainly violated in this case.

In this section we apply the information breakdown to multi-unit recordings collected from the MT visual area of a behaving monkey.

Monkeys were trained on a direction discrimination task. They were fixating a screen centrally (fixation window $\pm 0.5^\circ$ – 1°). A structured background was back-projected onto the

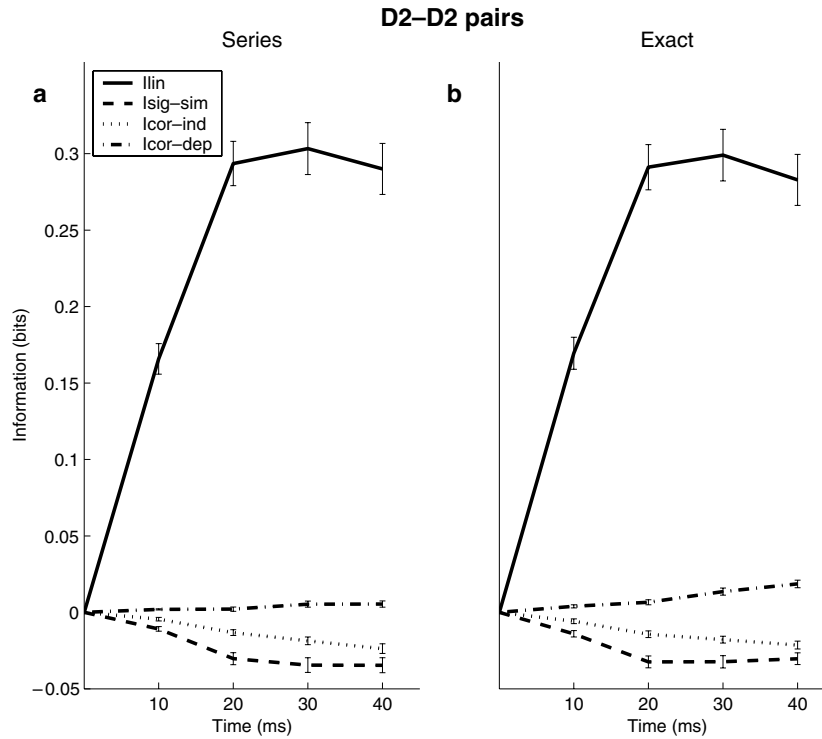


Figure 5. A comparison of series expansion method and the exact information breakdown method on an application to neuronal pairs simultaneously recorded from barrel-column D2 in rat barrel cortex. (a) Series expansion results; (b) exact information breakdown results.

screen. After a randomized period of time a grating was projected onto the screen moving in one of four possible directions along the cardinal axes, positioned over the receptive fields of the neurons under study. Luminance contrast, i.e. visibility, of the stimulus was randomized (0, 2, 4, 17%). Monkeys performed a reaction time task and indicated the perceived direction of motion by a hand movement to one of four touch bars located in front of the chest. MUA was recorded through seven electrodes independently placed in area MT and subjected to threshold discrimination. Neural data analysed in this paper was from high luminance stimulus conditions. For more detailed information see Thiele *et al* (1999).

We wish to illustrate how the information breakdown can be used to obtain a credible evaluation of the information content of cross-correlated firing in this case. We considered the information transmitted by the joint spike count code. We either computed the spike counts in cumulative post-stimulus time windows increased in steps of 50 ms ((b) in figures 6 and 7, called cumulative information), or in sliding time windows 50 ms long ((c) of figures 6 and 7, called instantaneous information). For the information calculation, we compressed the number of possible responses by grouping the response firing rates into four equi-populated classes per neuron. Binning the responses into subclasses, of course, decreases the information. However, it is preferable to obtain an information value that is approximated because of binning than to get an estimate that is out of statistical control because of under-sampling and biases. Given that 50–60 trials per stimulus were available, the binned information values should be free from bias problems (see figures 3 and 4). In addition to the new analysis method, we used conventional peri-stimulus-time histogram (PSTH) and CCG analysis. To evaluate the strength

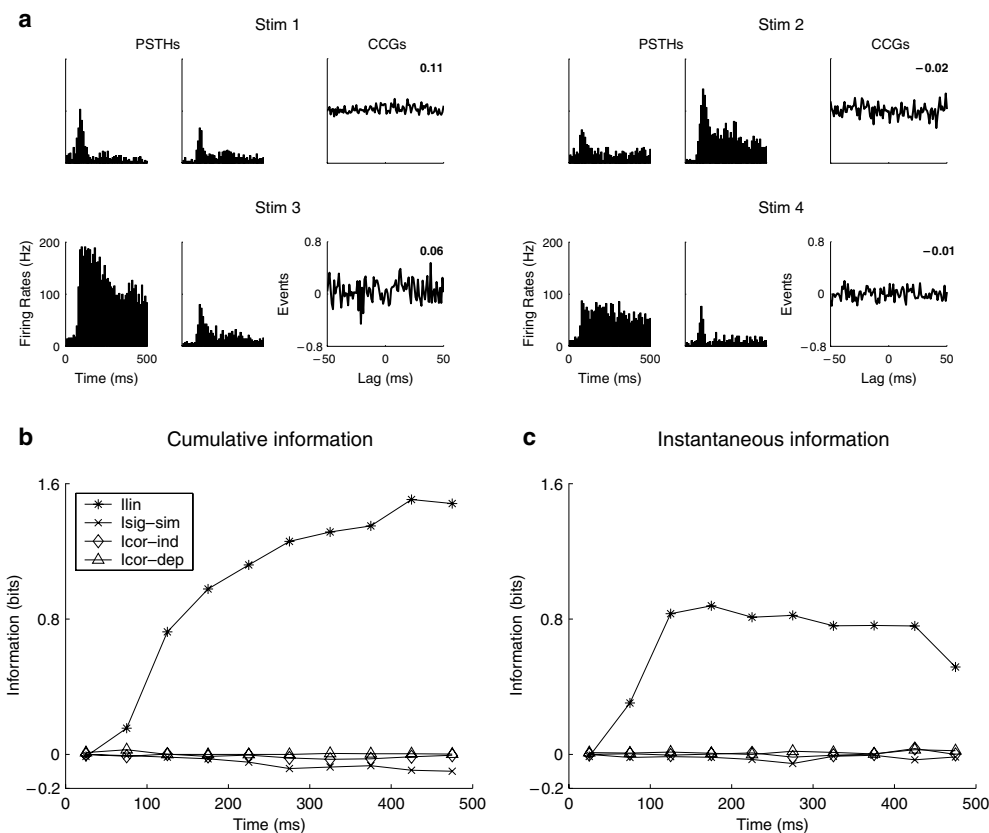


Figure 6. Information breakdown of multi-unit recordings from MT paired recording sites. In this example, two uncorrelated recording sites were analysed. (a) PSTHs and CCGs across the set of four stimuli (values of CCG correlation coefficients reported in the CCGs panels); (b) cumulative information breakdown; (c) instantaneous information breakdown.

of correlation from the CCG, we used the method of Bair *et al* (2001). In brief, we computed shuffle-predictor subtracted CCGs and auto-correlograms (ACG). The strength of correlation was computed as the area of the subtracted CCG divided by the square root of the product of the areas of the subtracted ACGs of each recording sites. The areas of all correlograms were computed in the central 20 ms, in order to evaluate only the strength of correlation of near-coincident spikes. As demonstrated by Bair *et al* (2001), this CCG correlation coefficient is equivalent to a Pearson-normalized spike count correlation coefficient based only on near-synchronous spikes.

An analysis of the first example of joint MUA recordings is shown in figure 6. The PSTHs of the two recording sites showed very different stimulus dependencies, and the CCG were flat and with very small correlation strength (CCG strength values are reported in the CCG panels of figures 6 and 7). The information analysis gave results very consistent with PSTH and CCG analysis. All information components leading to possible redundancy or synergy were very small, and the two sites conveyed nearly independent information, both for cumulative and instantaneous information.

In figure 7 we report the results of the analysis of two strongly correlated recording sites. PSTHs of MUA on each site showed a clear stimulus selectivity, and were also very similar

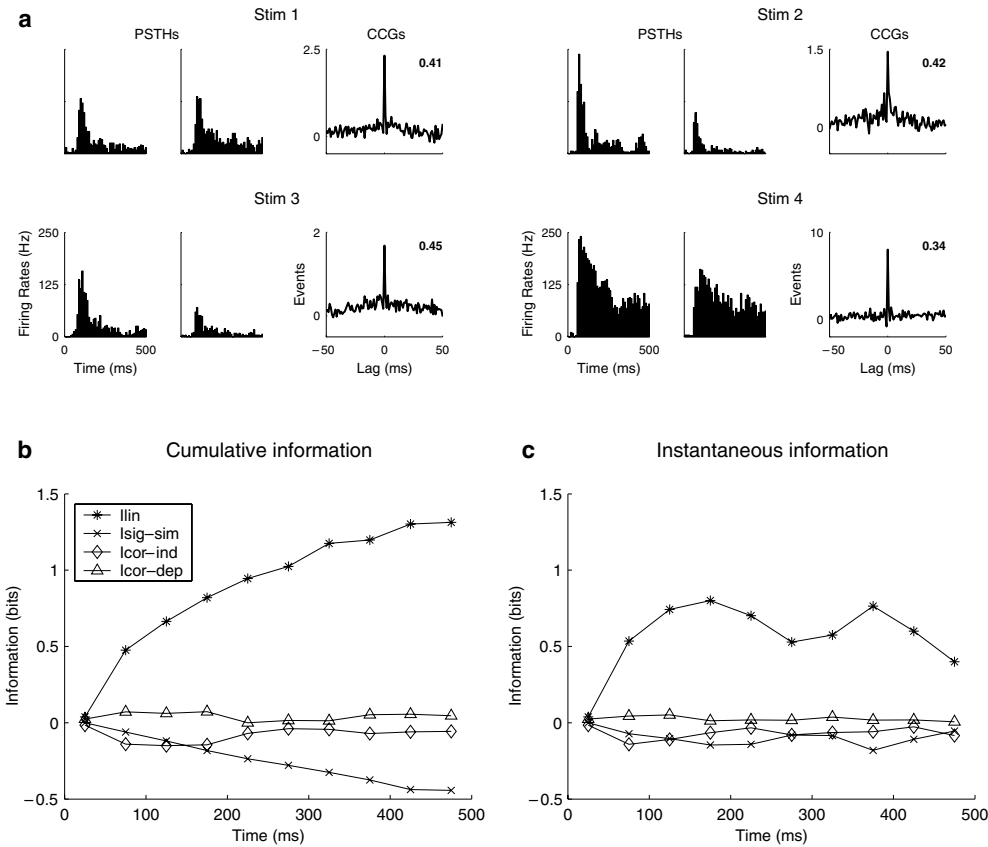


Figure 7. Information breakdown of multi-unit recordings from MT paired recording sites. In this second example, two strongly correlated recording sites were analysed. Conventions are as in figure 6.

between each other. CCGs presented a strong narrow central peak, with high correlation coefficient. The correlation coefficient of the CCG appeared to be much less stimulus-modulated than the PSTHs and mean rates. Also in this case, the exact information breakdown analysis gave results very consistent with PSTH and CCG analysis. The similarity of PSTHs of the two sites led to an appreciable signal-similarity information redundancy (-54% of the total at 500 ms post-stimulus). The stimulus-independent correlational component also decreased the information (this could be explained by the simultaneous presence of similarity of PSTHs and of positive cross-cell correlation). There was some information in the stimulus modulations of cross-correlation strength, but this information was only a small fraction of the total information carried by the pair (figure 7, parts (b) and (c)).

It is worth pointing out that the series expansion analysis gave non-credible results in both examples. Although it is evident from PSTH differences that in both examples there was stimulus-related information in the individual cells responses to different stimuli, the series expansion approach clearly failed to provide sensible results after approximately 80–100 ms post-stimulus, leading to zero or even negative values for the total mutual information (not shown). This confirms the usefulness of the new technique for analysing high firing rates systems and/or long post-stimulus times.

8. Discussion

A systematic quantification of how cross-correlation contributes to neuronal information transmission and to coding of stimulus features is an important step towards resolving some of the controversies on the role of correlation in neuronal coding (Gawne and Richmond 1993, Nirenberg *et al* 2001, Petersen *et al* 2001, Oram *et al* 2001). Information theoretic formalisms of the type presented allow a direct and exhaustive investigation of the contribution of neuronal correlation to stimulus representation in the brain. Thus, in our view, they constitute useful means to increase our understanding of the function of correlated firing in cortical processing.

This work permits one to obtain a detailed and exact quantification of the information contribution of several specific coding mechanisms. We previously developed an approximated approach (the series expansion method (Panzeri *et al* 1999, Panzeri and Schultz 2001)) that performed the same separation of information into coding components. However, the series expansion approach takes into account only the second-order statistics of neuronal interactions, and it is valid only when just a few spikes per stimulus presentation are typically emitted. The range of applicability of the series expansion method can be extended significantly by the inclusion of the contribution of higher order statistics (S R Schultz, personal communication). However, the series expansion cannot be applied to high firing rate data, such as MUA. MUA is an important tool to investigate the role of correlated firing. One important and unique feature of the new exact information breakdown is that it can be applied to high firing rate systems and multi-unit data.

The exact information breakdown, like any direct quantification of information from the stimulus-response probabilities (Strong *et al* 1998), requires relatively large amounts of data in order to obtain unbiased information estimates. In general, it requires more data than the series expansion method (Schultz and Panzeri 2001). Hence, although both the exact and approximated methods are designed to address the very same neurophysiological questions, they have complementary ranges of applicability. Of particular interests are experimental conditions in which firing rates are low enough for the series expansion to be applicable, and data are abundant enough for the new exact method to be also applied. In this case, comparing the information obtained by the second-order series expansion to the information obtained with the exact method can lead to determine whether higher order correlations convey any information. We performed this comparison on pairs of neurons recorded in rat somatosensory cortex, and we found that using only second-order statistics led to an estimation of information that was, component by component, almost identical to the information conveyed by the full response probability distribution (see figure 5). This allowed us to demonstrate that in this case the population of somatosensory neurons did not convey any information by means of higher order response statistics.

A limitation of studies such as the present one is that they only reveal what information is available (or ‘encoded’) in the neuronal responses. How the nervous system might make use of the available information is a separate issue. One specific question is relevant in the context of correlated coding. If there is information encoded in correlated firing, could this information be decoded by a downstream neuronal system that ignores correlation? Nirenberg *et al* (2001) recently tackled this question in the retina by explicitly quantifying how much information is lost in trying to decode a correlated population response by using an uncorrelated model. In this paper we found that the information neglected by the uncorrelated model equals the information specifically carried by the stimulus-modulation of the correlation strength. We find this result interesting for at least two reasons. First, it adds further significance to the information breakdown equations derived in this paper. It shows that they might be used not only to infer how correlations *encode* information, but also how a downstream system could

decode this information. In particular, if the stimulus-dependent correlational term is zero, the decoder can ignore correlation with no information loss, even if there is positive information in the stimulus-independent correlational mechanism. Second, this result might help further clarifying the significance and the implications of recent work on correlations in the retina (Nirenberg *et al* 2001).

Acknowledgments

We are indebted to P Latham, R S Petersen and S R Schultz for extremely useful discussions of concepts related to this work. We thank M Lebedev and M E Diamond for kindly providing us the data used for the example analysis in figure 5. This research was supported by an MRC Research Fellowship (SP), by BBSRC, HFSP and EU (GP, SP), by Wellcome Trust Grant 066372/Z/01/Z (SP) and by DFG SFB 509 Grant (AT, KPH).

Appendix A. Verification of the information breakdown formulae

In this appendix we give a concise explanation of how to verify that the sum of the components of the information breakdown equals the mutual information, equation (1). The idea is to express equations (9)–(12) in terms of entropies and entropy-like quantities, the sum of which apparently equals the mutual information.

First, it is easy to show by standard information-theoretic calculations (Cover and Thomas 1991), that the linear information breakdown term, equation (9), can be written as a difference of entropies:

$$I_{lin} = \sum_c (H(\mathcal{R}_c) - H(\mathcal{R}_c|S)) \quad (17)$$

where $H(\mathcal{R}_c)$ and $H(\mathcal{R}_c|S)$ are, respectively, the entropy and the conditional entropy of the probability distributions of the c th cell:

$$H(\mathcal{R}_c) = - \sum_{r_c} P(r_c) \log_2 P(r_c) \quad (18)$$

$$H(\mathcal{R}_c|S) = - \left\langle \sum_{r_c} P(r_c|s) \log_2 P(r_c|s) \right\rangle_s. \quad (19)$$

Let us now consider the signal similarity information component, equation (10). Using the definition of the signal similarity coefficient, equation (6), and the zero-sum constraint, equation (7), it can be proved that

$$I_{sig-sim} = H_{ind}(\mathcal{R}) + \sum_{\mathbf{r}} P_{ind}(\mathbf{r}) \log_2 \left(\prod_c P(r_c) \right) \quad (20)$$

where

$$H_{ind}(\mathcal{R}) = - \sum_{\mathbf{r}} P_{ind}(\mathbf{r}) \log_2 P_{ind}(\mathbf{r}). \quad (21)$$

Then, using the following identity:

$$\sum_{\mathbf{r}} P_{ind}(\mathbf{r}) \log_2 P(r_c) = -H(\mathcal{R}_c) \quad (22)$$

valid for each cell $c = 1, \dots, C$, we can finally express the signal similarity term as

$$I_{sig-sim} = H_{ind}(\mathcal{R}) - \sum_c H(\mathcal{R}_c). \quad (23)$$

Let us now consider the stimulus-independent correlational component, equation (11). Using the definition of ν , equation (6), the zero-sum constraint, equation (5) and the identity (22), equation (11) can be re-expressed as follows:

$$I_{cor-ind} = -H_{ind}(\mathcal{R}) + \chi[P, P_{ind}] \quad (24)$$

where χ is an entropy-like term, defined as

$$\chi[P, P_{ind}] = - \sum_{\mathbf{r}} P(\mathbf{r}) \log_2 P_{ind}(\mathbf{r}). \quad (25)$$

Finally, using similar manipulations it can be shown that the stimulus dependent correlational component, equation (12), can be rearranged in the following equivalent form:

$$I_{cor-dep} = I - \chi[P, P_{ind}] + \sum_c H(\mathcal{R}_c|S). \quad (26)$$

In summary, the information breakdown can be re-expressed in the following form:

$$I_{lin} = \sum_c (H(\mathcal{R}_c) - H(\mathcal{R}_c|S)) \quad (27)$$

$$I_{sig-sim} = H_{ind}(\mathcal{R}) - \sum_c H(\mathcal{R}_c) \quad (28)$$

$$I_{cor-ind} = -H_{ind}(\mathcal{R}) + \chi[P, P_{ind}] \quad (29)$$

$$I_{cor-dep} = I - \chi[P, P_{ind}] + \sum_c H(\mathcal{R}_c|S). \quad (30)$$

It is now straightforward to verify that the sum of equations (27)–(30) exactly equals the total information I . Together with the properties of the individual terms discussed in section 3, this constitutes a proof that the equations we derived indeed form an exact and meaningful breakdown of information into coding components. Expressing the information components as in equations (27)–(30) is convenient for numerical implementation of the analysis and for calculation of the bias of each component (see appendix B).

Appendix B. Computation of bias corrections terms

In this appendix we describe the calculation of the bias correction for the components of the information breakdown. As shown in appendix A these components can be expressed in terms of $H(\mathcal{R}_c)$, $H(\mathcal{R}_c|S)$, $H_{ind}(\mathcal{R})$, $\chi[P, P_{ind}]$ and I , see equations (27)–(30). Hence, it is sufficient to compute the bias corrections only for these entropies and entropy-like quantities. We assume that there are N_s trials to each stimulus s (N is the total number of trials across stimuli). We also assume that the outcome of each trial is an independent realization of the same stochastic process. We define the bias of a given functional of the probabilities as the difference between the average value of the functional when the probabilities are computed from N trials only and the value of the functional computed with the true probabilities (obtained from an infinite number of observations).

We calculate the bias of each probability functional using a simple procedure, namely by using a Taylor series expansion of these quantities around the true value and then averaging over all possible outcomes of the N trials. We considered only up to the first two terms in the Taylor series expansion. This approximation is good if there are enough experimental trials N to make small the fluctuations of the estimated probabilities around the asymptotic value.

To illustrate this kind of calculation, let us start from the simplest case, i.e. the calculation of the bias correction for the entropy of the response of a single cell c , equation (18). We denote by a subscript N the calculation of the probability from N trials instead that from the true

underlying probability distributions. The average systematic error (bias) can be approximately computed by taking a Taylor series expansion for $H_N(\mathcal{R}_c)$ around the true probability $P(r_c)$, and then averaging over all possible outcomes with N trials:

$$\begin{aligned} \text{Bias}[H(\mathcal{R}_c)] &= \langle H_N(\mathcal{R}_c) \rangle_N - H(\mathcal{R}_c) \\ &\approx \sum_{\tilde{r}_c} \frac{\delta H}{\delta P(\tilde{r}_c)} \langle P_N(\tilde{r}_c) - P(\tilde{r}_c) \rangle_N \\ &\quad + \frac{1}{2} \sum_{\tilde{r}_c} \frac{\delta^2 H}{\delta P(\tilde{r}_c)^2} \langle (P_N(\tilde{r}_c) - P(\tilde{r}_c))^2 \rangle_N + \dots \end{aligned} \quad (31)$$

where $\frac{\delta H}{\delta P}$ stands for the functional derivative of the entropy with respect to response probability distribution, computed in the asymptotic value $P(\tilde{r}_c)$. We introduce the symbol \tilde{r}_c to distinguish it from r_c which is a running variable used to define $H(\mathcal{R}_c)$ (see equation (18)). Multinomial combinatorics allow one to obtain the following expression for the bias and variance⁸ of $P(\tilde{r}_c)$:

$$B_N[P(\tilde{r}_c)] = \langle P_N(\tilde{r}_c) - P(\tilde{r}_c) \rangle_N = 0, \quad (32)$$

$$\sigma_N^2[P(\tilde{r}_c), P(\tilde{r}_c)] = \langle (P_N(\tilde{r}_c) - P(\tilde{r}_c))^2 \rangle_N = \frac{P(\tilde{r}_c)(1 - P(\tilde{r}_c))}{N} + o\left(\frac{1}{N}\right). \quad (33)$$

Hence

$$\text{Bias}[H(\mathcal{R}_c)] \approx -\frac{1}{2N \ln 2} (\tilde{R}_c - 1), \quad (34)$$

where \tilde{R}_c denotes the number of ‘relevant’ bins, i.e. the response bins in which the occupancy probability $P(\tilde{r}_c)$ is not zero (see Panzeri and Treves 1996). Analogously it is possible to compute the bias correction for the conditional entropy. Hence the bias correction for the single cell terms reads

$$\text{Bias}[H(\mathcal{R}_c|\mathcal{S})] \approx -\frac{1}{2N \ln 2} \sum_s (\tilde{R}_{c;s} - 1), \quad (35)$$

where $\tilde{R}_{c;s}$ are the relevant bins corresponding to the probability distribution $P(\tilde{r}_c|s)$. The bias correction for the total mutual information I is reported in equation (14) and it can be computed simply as the bias difference of unconditional and conditional response entropies.

Let us now calculate the bias correction for $H_{ind}(\mathcal{R})$ by using the approach outlined above. In this case the calculation is more complex because $H_{ind}(\mathcal{R})$ is a functional of the probability distributions $P(s)$ and $P(r_c|s)$ through the independent stimulus-unconditional probability distribution $P_{ind}(r)$. The leading term of the bias correction for $H_{ind}(\mathcal{R})$ is given by

$$\begin{aligned} \text{Bias}[H_{ind}] &\approx \frac{1}{2} \sum_s \frac{\delta^2 H_{ind}}{\delta P(s)^2} \sigma_N^2[P(s), P(s)] + \frac{1}{2} \sum_{s, s', s \neq s'} \frac{\delta^2 H_{ind}}{\delta P(s) \delta P(s')} \sigma_N^2[P(s), P(s')] \\ &\quad + \frac{1}{2} \sum_s \sum_c \sum_{\tilde{r}_c} \frac{\delta^2 H_{ind}}{\delta P(\tilde{r}_c|s)^2} \sigma_N^2[P(\tilde{r}_c|s), P(\tilde{r}_c|s)] \\ &\quad + \frac{1}{2} \sum_s \sum_{b, c, b \neq c} \sum_{\tilde{r}_b, \tilde{r}_c} \frac{\delta^2 H_{ind}}{\delta P(\tilde{r}_b|s) \delta P(\tilde{r}_c|s)} \sigma_N^2[P(\tilde{r}_b|s), P(\tilde{r}_c|s)], \end{aligned} \quad (36)$$

⁸ The variance of a probability distribution P_X and the co-variance of two probability distributions P_X and P_Y are defined respectively as

$$\begin{aligned} \sigma_N^2[P_X, P_X] &\equiv \langle (P_{X,N} - P_X)^2 \rangle_N, \\ \sigma_N^2[P_X, P_Y] &\equiv \langle (P_{X,N} - P_X)(P_{Y,N} - P_Y) \rangle_N, \end{aligned}$$

where we remind that $\langle \dots \rangle_N$ stands for averaging over all possible outcomes with N trials.

where $\frac{\delta H_{ind}}{\delta P}$ stands for the functional derivative of the independent entropy with respect to response probability distributions (computed in the asymptotic value), and $\sigma_N^2[\cdot, \cdot]$ are the variances and co-variances of the probability distributions $P(s)$ and $P(\tilde{r}_c|s)$. In equation (36) we do not report the contributions of the linear derivatives because, like in equation (31), they do not contribute to the bias of the independent entropy. In the above we also omitted the quadratic terms which are zero and thus do not contribute to the bias of $H_{ind}(\mathcal{R})$. As in the derivation of the bias correction for single cell entropy we consider the Taylor expansion up to second order. After performing the functional derivatives, the equation for the bias correction (36) becomes

$$\begin{aligned} \text{Bias}[H_{ind}] &\approx -\frac{1}{2 \ln 2} \sum_s \widehat{\sum}_{\mathbf{r}} \frac{P_{ind}^2(\mathbf{r}|s)}{P_{ind}(\mathbf{r})} \sigma_N^2[P(s), P(s)] \\ &\quad - \frac{1}{2 \ln 2} \sum_{s, s', s \neq s'} \widehat{\sum}_{\mathbf{r}} \frac{P_{ind}(\mathbf{r}|s) P_{ind}(\mathbf{r}|s')}{P_{ind}(\mathbf{r})} \sigma_N^2[P(s), P(s')] \\ &\quad - \frac{1}{2 \ln 2} \sum_s \sum_c \sum_{\tilde{r}_c} \widehat{\sum}_{\mathbf{r}} \frac{P^2(s)}{P_{ind}(\mathbf{r})} \frac{P_{ind}^2(\mathbf{r}|s)}{P^2(\tilde{r}_c|s)} \delta_{[r_c, \tilde{r}_c]} \sigma_N^2[P(\tilde{r}_c|s), P(\tilde{r}_c|s)] \\ &\quad - \frac{1}{2 \ln 2} \sum_s \sum_{b, c, b \neq c} \sum_{\tilde{r}_b, \tilde{r}_c} \widehat{\sum}_{\mathbf{r}} P(s) \delta_{[r_b, \tilde{r}_b]} \delta_{[r_c, \tilde{r}_c]} \frac{P_{ind}(\mathbf{r}|s)}{P(\tilde{r}_b|s) P(\tilde{r}_c|s)} \\ &\quad \times \left(\frac{P(s) P_{ind}(\mathbf{r}|s)}{P_{ind}(\mathbf{r})} + \ln P_{ind}(\mathbf{r}) \right) \sigma_N^2[P(\tilde{r}_b|s), P(\tilde{r}_c|s)], \end{aligned} \quad (37)$$

where $\widehat{\sum}_{\mathbf{r}}$ is a summation restricted to the response variables \mathbf{r} such that $P_{ind}(\mathbf{r}) \neq 0$, and $\delta_{[r_c, \tilde{r}_c]}$ is a Kronecker delta (i.e. $\delta_{[r_c, \tilde{r}_c]} = 1$ if $r_c = \tilde{r}_c$ and $\delta_{[r_c, \tilde{r}_c]} = 0$ if $r_c \neq \tilde{r}_c$). The values of the variances and co-variances are as follows:

$$\sigma_N^2[P(s), P(s)] = \frac{P(s)(1 - P(s))}{N} + o\left(\frac{1}{N}\right), \quad (38)$$

$$\sigma_N^2[P(s), P(s')] = -\frac{P(s)P(s')}{N} + o\left(\frac{1}{N}\right), \quad (39)$$

$$\sigma_N^2[P(\tilde{r}_c|s), P(\tilde{r}_c|s)] = \frac{P(\tilde{r}_c)(1 - P(\tilde{r}_c))}{N_s} + o\left(\frac{1}{N_s}\right), \quad (40)$$

$$\sigma_N^2[P(\tilde{r}_b|s), P(\tilde{r}_c|s)] = -\frac{P(\tilde{r}_b)P(\tilde{r}_c)}{N_s} + \frac{P(\tilde{r}_b, \tilde{r}_c)}{N_s} + o\left(\frac{1}{N_s}\right). \quad (41)$$

After replacing the explicit values of variances and covariances in equation (37) and performing some algebra, we obtained the following final expression for the bias of $H_{ind}(\mathcal{R})$:

$$\begin{aligned} \text{Bias}[H_{ind}] &= \frac{1}{2N \ln 2} \widehat{\sum}_{\mathbf{r}} \frac{1}{P_{ind}(\mathbf{r})} \left(\left(C^2 - 1 - \frac{\alpha(\mathbf{r}|s)}{P_{ind}(\mathbf{r}|s)} - \beta(\mathbf{r}|s) \right) P_{ind}^2(\mathbf{r}|s) \right) \\ &\quad + \frac{1}{2N \ln 2} \left[1 + \widehat{\sum}_{\mathbf{r}} \sum_s (C^2 - C - \beta(\mathbf{r}|s)) P_{ind}(\mathbf{r}|s) \ln P_{ind}(\mathbf{r}) \right] \end{aligned} \quad (42)$$

where C is the number of cells and $\alpha(\mathbf{r}|s)$ and $\beta(\mathbf{r}|s)$ are defined as follows:

$$\alpha(\mathbf{r}|s) = \sum_c \frac{P_{ind}(\mathbf{r}|s)}{P(r_c|s)}, \quad (43)$$

$$\beta(\mathbf{r}|s) = \sum_{b, c, b \neq c} \frac{P(r_b, r_c|s)}{P(r_b|s) P(r_c|s)}. \quad (44)$$

It is worth stressing that both $\alpha(\mathbf{r}|s)$ and $\beta(\mathbf{r}|s)$ are regular and finite quantities in the range allowed for the sum over \mathbf{r} .

Finally, let us consider the bias correction for $\chi[P, P_{ind}]$. This term can be considered as a functional of $P(s)$, $P(r_c|s)$, and $P(\mathbf{r}|s)$. The bias correction can be computed, as above, calculating the Taylor series expansion of $\chi[P, P_{ind}]$ around the true values of the probability distributions and, then, averaging over all possible outcomes with N trials. The required variances and co-variances are those given by equations (38)–(41) and the following:

$$\sigma_N^2[P(\tilde{r}_c|s), P(\mathbf{r}|s)] = -\frac{P(\tilde{r}_c|s)P(\mathbf{r}|s)}{N_s} + \delta_{[r_c, \tilde{r}_c]} \frac{P(\mathbf{r}|s)}{N_s} + o\left(\frac{1}{N_s}\right). \quad (45)$$

The result of the calculation for the bias correction for χ is

$$\begin{aligned} \text{Bias}[\chi] = & -\frac{1}{2N \ln 2} \sum_{\mathbf{r}} \frac{P(\mathbf{r})}{P_{ind}^2(\mathbf{r})} \left\langle \left(C^2 - 1 - \frac{\alpha(\mathbf{r}|s)}{P_{ind}(\mathbf{r}|s)} - \beta(\mathbf{r}|s) \right) P_{ind}^2(\mathbf{r}|s) \right\rangle_s \\ & + \frac{1}{2N \ln 2} \left[1 + \sum_{\mathbf{r}} \frac{P(\mathbf{r})}{P_{ind}(\mathbf{r})} \sum_s P_{ind}(\mathbf{r}|s) (C^2 - C - \beta(\mathbf{r}|s)) \right] \\ & + \frac{1}{2N \ln 2} \sum_{\mathbf{r}} \frac{1}{P_{ind}(\mathbf{r})} \left\langle P(\mathbf{r}|s) P_{ind}(\mathbf{r}|s) \left(2C - 2 - 2 \frac{\alpha(\mathbf{r}|s)}{P_{ind}(\mathbf{r}|s)} \right) \right\rangle_s. \end{aligned} \quad (46)$$

In the absence of correlation ($P = P_{ind}$) the total bias for the stimulus-independent correlational component (obtained by subtracting $H_{ind}(\mathcal{R})$ to $\chi[P, P_{ind}]$, see equation (29)) is zero, no matter how big N is. This is because the stimulus-independent correlational component does not contain any correlation term in the argument of the logarithm, see equation (11).

References

- Abbott L F and Dayan P 1999 The effect of correlated variability on the accuracy of a population code *Neural Comput.* **11** 91–101
- Aertsen A M H J, Gerstein G L, Habib M K and Palm G 1989 Dynamics of neuronal firing correlation: modulation of ‘effective connectivity’ *J. Neurophysiol.* **61** 900–17
- Bair W, Zohary E and Newsome W T 2001 Correlated firing in macaque visual area mt: timescales and relationship to behaviour *J. Neurosci.* **21** 1676–97
- Brody C D 1999 Correlations without synchrony *Neural Comput.* **11** 1537–51
- Cover T M and Thomas J A 1991 *Elements of Information Theory* (New York: Wiley)
- deCharms R C and Merzenich M M 1996 Primary cortical representation of sounds by the coordination of action potentials *Nature* **381** 610–13
- de Oliveira S C, Thiele A and Hoffman K-P 1997 Synchronization of neuronal activity during stimulus expectation in a direction discrimination task *J. Neurosci.* **17** 9248–60
- DeWeese M 1996 Optimization principles for the neural code *Network: Comput. Neural Syst.* **7** 325–31
- Dimitrov A G and Miller J P 2001 Neural coding and decoding: communication channels and quantization *Network: Comput. Neural Syst.* **12** 441–72
- Fries P, Roelfsema P R, Engel A K, Konig P and Singer W 1997 Synchronization of oscillatory responses in visual cortex correlates with perception in interocular rivalry *Proc. Natl Acad. Sci. USA* **94** 12699–704
- Gawne T J and Richmond B J 1993 How independent are the messages carried by adjacent inferior temporal cortical neurons? *J. Neurosci.* **13** 2758–71
- Gray C M, Konig P, Engel A K and Singer W 1989 Oscillatory responses in cat visual cortex exhibit inter-columnar synchronization which reflects global stimulus properties *Nature* **338** 334–7
- Lamme V A F and Spekreijse H 1999 Neuronal synchrony does not represent texture segregation *Nature* **396** 362–6
- Laubach M, Shuler M and Nicolelis M A L 1999 Independent component analyses for quantifying neuronal ensemble interactions *J. Neurosci. Methods* **94** 141–54
- Lebedev M A, Mirabella G, Erchova I and Diamond M E 2000 Experience-dependent plasticity of rat barrel cortex: redistribution of activity across barrel-columns *Cereb. Cortex* **10** 23–31

- Mastrorade D N 1983 Correlated firing of cat retinal ganglion cells: I. Spontaneously active inputs to x- and y-cells *J. Neurophysiol.* **49** 303–24
- Nakahara H and Amari S 2002 Information-geometric measure for neural spikes *Neural Comput.* **14** 2269–316
- Nirenberg S, Carcieri S M, Jacobs A and Latham P E 2001 Retinal ganglion cells act largely as independent encoders *Nature* **411** 698–701
- Nirenberg S and Latham P E 1998 Population coding in the retina *Curr. Opin. Neurobiol.* **8** 488–93
- Oram M W, Hatsopoulos N, Richmond B and Donoghue J 2001 Excess synchrony in motor cortical neurons provides redundant direction information with that from coarse temporal measures *J. Neurophysiol.* **86** 1700–16
- Oram M W, Földiák P, Perrett D I and Sengpiel F 1998 The ‘Ideal Homunculus’: decoding neural population signals *Trends Neurosci.* **21** 259–65
- Panzeri S, Petersen R, Schultz S, Lebedev M and Diamond M 2001 The role of spike timing in the coding of stimulus location in rat somatosensory cortex *Neuron* **29** 769–77
- Panzeri S and Schultz S 2001 A unified approach to the study of temporal, correlational and rate coding *Neural Comput.* **13** 1311–49
- Panzeri S, Schultz S R, Treves A and Rolls E T 1999 Correlations and the encoding of information in the nervous system *Proc. R. Soc. B* **266** 1001–12
- Panzeri S and Treves A 1996 Analytical estimates of limited sampling biases in different information measures *Network: Comput. Neural Syst.* **7** 87–107
- Petersen R S, Panzeri S and Diamond M 2001 Population coding of stimulus location in rat somatosensory cortex *Neuron* **32** 503–14
- Reich D S, Mechler F and Victor J D 2001 Independent and redundant information in nearby cortical neurons *Science* **294** 2566–8
- Schultz S and Panzeri S 2001 Temporal correlations and neural spike train entropy *Phys. Rev. Lett.* **86** 5823–6
- Shadlen M N and Movshon J A 1999 Synchrony unbound: a critical evaluation of the temporal binding hypothesis *Neuron* **24** 67–77
- Shannon C E 1948 A mathematical theory of communication *AT&T Bell Labs. Tech. J.* **27** 379–423
- Strong S, Koberle R, de Ruyter van Steveninck R and Bialek W 1998 Entropy and information in neural spike trains *Phys. Rev. Lett.* **80** 197–200
- Thiele A, Distler C and Hoffmann K P 1999 Decision-related activity in the macaque dorsal visual pathway *Eur. J. Neurosci.* **11** 2044–58
- Victor J D 2000 Asymptotic bias in information estimates and the exponential (bell) polynomials *Neural Comput.* **12** 2797–804
- Victor J D 2002 Binless strategies for estimation of information from neuronal data *Phys. Rev. E* at press
- Villa A E, Tetko I, Hyland B and Najem A 1999 Spatiotemporal activity patterns of rat cortical neurons predict responses in a conditioned task *Proc. Natl Acad. Sci. USA* **96** 1106–11
- Wu S, Nakahara H and Amari S 2001 Population coding with correlation and an unfaithful model *Neural Comput.* **13** 775–97
- Zohary E, Shadlen M N and Newsome W T 1994 Correlated neuronal discharge rate and its implication for psychophysical performance *Nature* **370** 140–3