

Focused review: Estimating the amount of information conveyed by a population of neurons

Marshall Crumiller², Bruce Knight¹, Yunguo Yu³ and Ehud Kaplan^{*2}

¹The Laboratory of Biophysics, The Rockefeller University, NY, NY, USA;

²The Neuroscience department, The Mount Sinai School of Medicine, NY, NY, USA;

³Department of Neuroscience, Weill Medical College, NY, NY, USA.

Correspondence:

Ehud Kaplan
Department of Neuroscience
The Mount Sinai School of Medicine
One Gustave Levy Place
Box 1065
New York, NY 10029
ehud.kaplan@mssm.edu

Running title:

Estimating information in neuronal populations

Key concepts

Entropy In information theory, entropy is a measure of the uncertainty associated with a random variable. It quantifies the disorder, or unpredictability of a collection of signals. It is related to the expected value of the information contained in a message, and is measured in bits. The concept was introduced in this context by C. Shannon's 1948 paper "A Mathematical Theory of Communication".

Mutual Information Mutual information between stimulus and response quantifies (in bits) the reduction in stimulus uncertainty gained from analyzing the response. It is calculated by subtracting from the total entropy the noise entropy, which is estimated from the variability of responses to repeated presentations of a stimulus.

Synergy/Redundancy If all the neurons in a population are independent, their group information equals the sum of the information each of them carries. If their group information is less than that sum, we have redundancy: some of the information delivered by some neurons is also delivered by others. If their group information has more than that sum, we have synergy, providing information that depends on the coordinated firing of some neurons, and cannot be extracted by examining individual neurons.

Fourier Analysis We are more interested a neuron's rate of transmitting information than we are in the development of the neuronal signal over time. Since Fourier analysis decomposes a set of neuronal outputs into a sum of Gaussian-distributed sine and cosine coefficients at various frequencies, it provides insight into the underlying processes that gave rise to the signal.

Abstract

Recent technological advances have made the simultaneous recording of the activity of many neurons common. However, estimating the amount of information conveyed by the discharge of a neural population remained a significant challenge. Here we describe our recently published analysis method that makes such estimates possible. We illustrate the use of the method with simulated neurons and from real data recorded from the lateral geniculate nucleus (LGN) of a monkey, and show its use in the calculation of redundancy and synergy among neuronal groups.

Keywords

Information, neural population, redundancy, frequency analysis

1. Introduction

The brain processes information by the coordinated activity of many neurons, and it is therefore natural to ask: *How much information does a neural population transmit?* In the past, several methods that estimated information rates from the firing pattern (Richmond et al., 1987; Richmond and Optican, 1987; Optican and Richmond, 1987; Bialek et al., 1991; Rieke et al., 1997; Strong et al., 1998; Brenner et al., 2000) or membrane potential (Borst and Theunissen, 1999; DiCaprio, 2004) of single neurons have been applied. The information contained in spike trains was estimated by calculating the *entropy* associated with the various temporal patterns of spike discharge, using Shannon’s formula (Shannon, 1949; Shannon and Weaver, 1949; Victor, 2006). Alternatively, knowledge about the stimulus confronting a sensory system was extracted by decoding the spike train of a single neuron (for a thorough review, see Quiroga and Panzeri, 2009).

Such calculations suffice when we are dealing with one neuron, but since all brain functions involve many interacting neurons, it is important to provide similar information estimates for a neuronal population. Simply adding up the information delivered by individual neurons in the population is valid only if the neurons act independently of one another, a situation that is rare in real brains (see, for example, Zohary et al., 1994; Bair et al., 2001; Latham and Nirenberg, 2005; Pillow et al., 2008). Methods that were adequate for single neuron data, such as the *Reconstruction method* (Bialek et al., 1991) or the *Direct Method* (Strong et al., 1998), are impractical for a population of neurons because of the “curse of dimensionality”: the huge multi-dimensional space inhabited by many diverse spike trains can only be sampled rather sparsely by most real-life neurophysiological experiments.

Calculating the information carried by a population of many neurons thus has remained a significant challenge (Brown et al., 2004; Quiroga and Panzeri, 2009), at the very time that the need for such estimates has become increasingly urgent: the technology of recording simultaneously from many neurons has become much more affordable and wide-spread, and data from such recordings are becoming common.

A quantitative measure of the information transmitted by a neural population will also make it possible to investigate *synergy* (population codes; for example, Gat and Tishby, 1999; Brenner et al., 2000; Latham and Nirenberg, 2005) and *redundancy* (less than additive combination of information) among interacting neurons, and could thus provide new insights into the computational principles employed by the brain. We now describe a method that estimates the amount of information carried by a population of spiking neurons, and demonstrate its use with both simulated data and data recorded from the *lateral geniculate nucleus* (LGN) of an anesthetized macaque monkey. To make the method accessible to a

wide audience of interested neuroscientists, we minimize here explicit use of mathematics. Interested readers can find details in the appendix of our original paper (Yu et al., 2010), where the details of the experimental procedures used to obtain the laboratory recordings can also be found.

2. Estimation of the information rate delivered by a neural ensemble

Here we describe in general terms a method, which we have devised and implemented, for the measurement of the rate of information flow from a set of neurons whose spiking activities are recorded simultaneously. The motivation for such a measurement is to help us gain insight into the manner in which a complex of interacting neurons assists in the processing of sensory data from the outside world into a form suitable for responsive action.

In everyday usage, the word “information” implies knowledge that enables us to choose among alternatives. The quantitative technical term “information”, measured in bits and calculated on the basis of our given knowledge, tells us how many alternatives our knowledge enables us to choose among. A prototypical example is an idealized information channel which, as time progresses, delivers a coded message in the form of sequential dots or dashes. Upon the next symbol’s occurrence, since it might have been of either type, the number of potential alternatives to our explicit message doubles. Since that happens at every step, the total number of alternatives is 2 raised to a power that is the number of steps. Stated in a more technical way: the number of steps is the logarithm (to the base 2) of the number of alternatives. A practical definition of the information in a message should have the property that when a long message is composed of two shorter ones, the information content of the total should be the sum of the information contents of its parts. In the example above, clearly the log (base 2) of the number of alternatives respects this demand. We identify that logarithm as the information in its technical and quantitative sense. An increment of this information by one unit (which corresponds to a doubling of alternatives) is called one “bit”.

We may extend the example above to the computation of information rate from a collection of neurons through a series of generalizations. A first generalization is to situations where the different symbols are constrained to appear with different likelihoods. A second generalization is to the use of more symbols. A third is to situations where the probability of a symbol’s occurrence depends on the past history of the message. (Information theory as applied to text can be included at this point.)

2.1. Noise

A fourth generalization, toward the multi-neuron example, is the inclusion of noise. In detail this involves the specification of ways in which errors can produce incorrect symbols. How might this situation be dealt with in the laboratory? In all of the above situations, a long enough sample of signal would be sufficient to deduce the process of symbol selection; the present challenge requires more. But we may ask further that the source repeatedly generate the same intended long signal. In the absence of noise, such a sequence of symbols generates no further information after the first repeat, but the presence of noise presents new alternatives for each new repeat, and the information rate for the string of repeats can be calculated by the methodology already devised. This is the *noise information*. It can be subtracted from the total information (which was calculable without repeats) to yield the information rate of the intended uncontaminated signal. To distinguish the two calculated quantities whose difference yielded the signal information, they are referred to as “entropies”, in analogy to similar expressions that arise in statistical mechanics.

2.2. Continuous signals

A fifth generalization is an extension of the theory to continuous signals. One may proceed as in elementary calculus, by considering a sequence of approximate problems in which, as one proceeds along the sequence, shorter time steps are taken but at each level, within a time step, the signal is approximated as constant and at one of a discrete set of values. The challenge, however, proves more subtle than in elementary calculus. Neither total entropy nor noise entropy approaches a finite value, as finer divisions into time-steps always create further additional alternatives. However, for small time-divisions the two entropies increase in the same way, and their difference, the signal information, approaches a well-defined limit. With this generalization information theory as applied to electronic signals can be included.

The information theory of continuous signals exploits something that was implicit in our prototypical example. Given a message, there are many ways to encode it in a two-symbol string of the same length. For example, any permutation of order will serve. (It is decoded by the inverse permutation.) Considering the way we calculated the information value, clearly that information value is not changed by the recoding. This carries over to the generalization with continuous signals.

It was famously observed by Fourier that any smooth continuous signal of finite duration can be approximated indefinitely well by a constant plus a sum of weighted sines and cosines that oscillate with frequencies that are integer multiples of a single fundamental frequency. A list of the weightings of all the oscillating sines and cosines enables us to reconstruct, across time, the signal that gave rise to them. Thus the list of weightings serves as an encoded form of the signal in terms of a discrete set of values, and as in the encoded prototype example discussed above, the evaluation of the quantity of signal information may be calculated directly from the list of weightings. It is familiar that in practical electronic communication, where noise is unavoidable, a sender may be assigned a limited range of frequencies within which to encode messages; and that the bit-rate that the sender may achieve is proportional to that “band width” of frequencies.

2.3. Information in a single neuron spike train

A sixth generalization, in practical terms a simplification, positions us to calculate the information rate of a neuron’s spike train. Technically, following Fourier, each weighting coefficient of a general continuous signal is to be found by first multiplying that signal by the relevant sine or cosine wave, and then by integrating that product across the time span of the signal. One might approximately treat a spike-train similarly by manufacturing a function of time that has at each spike-time a unit-area impulse shaped like the neuron action potentials one sees on an oscilloscope. One might also use a briefer and taller rescaling of that shape. By including enough sine and cosine waves, the impulse-shape at each occurrence time still will be recovered by the weighted summation. At low frequencies, where the period of the sine or cosine is long compared to the duration of the unit-area impulse, the deduced weighting coefficients will not depend on the shape of the impulse. We may take the brief-impulse limit. In that limit the integration of the product of signal and sine (or cosine) becomes the sum of the evaluation of the sine (or cosine) at all spike times, from which the weighting coefficient follows.

Information theory applies most naturally to signals that eventually become independent of their more distant past, and gives reliable numerical results when signal samples are available that are long compared to their forgetting time. For a signal to “become independent” implies that, at least gradually, it accumulates unpredictable changes. Technically, such a signal is termed “stochastic”. In the Fourier context, and for such a stochastic signal, we may use a feature that (to our knowledge) has

not previously been exploited in the context of information content of spike trains. A stochastic signal gives a stochastic Fourier weighting coefficient, which is characterized by a probability distribution over possible values. An example of that weighting coefficient is found by performing a specific integral (involving sine or cosine, mentioned above) over the time-span. That time-span may be broken up into a uniform set of sub-spans, and that breaks the integral up into a sum of integrals over the sub-spans. Each of those stochastic sub-span integrals has the same probability distribution, and since the sub-spans are still long compared to the memory time, the probability distributions are independent. We can show that those distributions are not heavy-tailed (do not feature outliers frequently enough to create computational pathologies). When one sums numerous stochastic variables that are independent and have the same light-tailed probability distribution, the distribution of that sum is Gaussian; this statement is the central limit theorem. Thus the overall integral, and the Fourier weighting coefficient, is Gaussian distributed (which we have empirically verified for our data; see Yu et al., 2010). A Gaussian distribution is fully characterized by its mean and variance. In the information-theoretic sense it has a well-defined entropy, expressed in terms of its variance.

With this theoretical machinery available, and with a neuron on hand in the laboratory, from which we can record a train of nerve impulses that are responding to an external stimulus that has finite memory time, the rate at which the neuron delivers signal information may be found by the use of the following experimental design. To create the laboratory stimulus we first choose, from the above external stimulus, a segment that is long enough to enable statistically reliable entropy measurements. This we will call the “repeat” segment. We then divide the much longer total stimulus into a large number of segments equal in duration to the repeat segment, and call these “unique” segments. The laboratory stimulus is constructed by alternating repeat and unique segments into a single long stimulus. The first step of data analysis is to re-separate the responses to the repeat segments from those to the unique segments. Processing the unique segments will yield the total entropy and processing the repeat segments will yield the noise entropy. Their difference will yield the cell’s signal information and, finally, division by the common duration of the response spike trains gives the cell’s response information bit-rate in response to the original stimulus.

In the case of the laboratory data reported here, in which monkeys watched a flickering video screen, the fastest time-scale was established by the stimulus frame-rate, which was $1/160$ of a second. The segment duration was 8 seconds, which established a Fourier fundamental frequency of $1/8$ hertz. There were 128 repeats and as many unique segments, so that the total time of an experimental run was $2 \times 128 \times 8 = 2,048$ sec = 34.133 min.

Since the probability density of any particular Fourier weighting coefficient is Gaussian, its entropy contribution follows from an explicit formula once the variance of that distribution is determined from the data. For clear visualization, it is convenient to subtract the noise entropy from the total entropy separately for each Fourier component, and so obtain that component’s contribution to the signal information rate. Figure 1 (panel 3) illustrates this across the range of important frequencies. It shows a moving average, which has been cosine-bell weighted across a 4 Hz frequency window. We note that the total entropy drops to the level of the noise entropy shortly before the frame-rate frequency of 160 hertz is reached. There is no need to evaluate Fourier coefficients beyond this cut-off frequency.

2.4. Information in multiple neurons spike trains

A seventh and final generalization includes the situation where simultaneous spike-time records are obtained from a number of cells. Our procedure is a straightforward generalization from the case of spike-times from a single cell. For each of the several cells, any particular Fourier coefficient is a stochas-

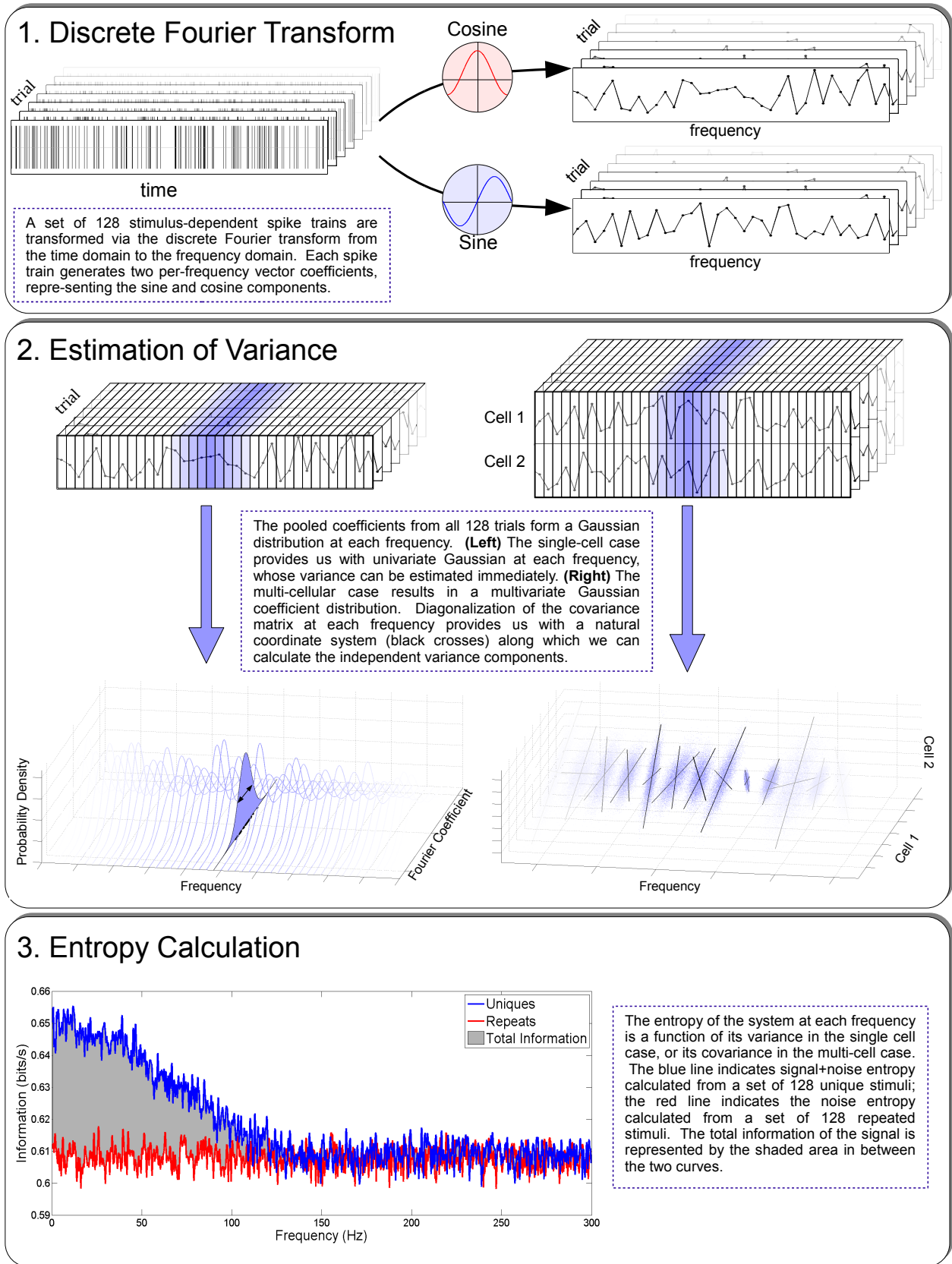


Figure 1: *The three steps that are required for calculating the information carried by a neural population: Fourier transform of each spike train; variance estimation, and entropy-information calculation.*

tic variable, and because the neurons recorded are all engaged in a common information-processing endeavor, typically these stochastic variables are not independent. This lack of independence of Fourier coefficients at a given frequency even goes a bit further, as we may anticipate correlations between the Fourier sine coefficient of one cell and the Fourier cosine coefficient of another. Among themselves, therefore, the Fourier sine and cosine coefficients at a given frequency present us with a multivariate probability density distribution, over twice as many variables as the number of cells in the spike-train record.

However, with a record that is long compared to the memory-time of the common stimulus, we can again observe that the stochastic vector that generates the multivariate probability distribution is the sum of stochastic vectors generated independently over shorter sub-epochs, from identical multivariate distributions. So, as in the previous case, the multivariate central limit theorem holds, and the overall multivariate distribution is a multivariate Gaussian. (Again, in our application, we have verified that this holds for our experimental data; Yu et al., 2010.) As before, in our application there will be two multivariate Gaussians, one from the unique segments from which we will calculate the total entropy, and one from the repeat segments, which will yield the noise entropy.

In the one-cell analysis we noted that a univariate Gaussian was fully characterized by the values of its mean and of its variance. Also, its entropy was given by an expression that depended only on its variance, which could be determined from laboratory data. A natural generalization holds for a multivariate Gaussian. It is fully characterized once we have obtained the mean values of all its variables and have also obtained its covariance matrix, which is a matrix composed of the mean values of all products of paired departures of the variables from their mean values. The covariance matrix may be calculated directly from the laboratory data. The entropy of a multivariate Gaussian is a single number that depends on the covariance matrix in a very specific way.

2.4.1. A bit of theoretical detail

A multivariate Gaussian distribution has some simplifying properties, which follow from the definition of the covariance matrix to which it is related. In particular, the covariance matrix shows us how we may pick, for our distribution, an alternative set of coordinates, whose origin lies at the distribution's maximum, and with axis directions so chosen that the distribution becomes the product of 1-dimensional (univariate) distributions, with variances specified by the data-determined covariance matrix. A simple calculation in information theory shows that if a higher dimensional distribution is the product of several lower dimensional distributions, then its entropy is the sum of the entropies of its components (Cover and Thomas, 2006, section 9.4). With the use of this observation, the information rate for the set of cells may be calculated by the same means that were used in the calculation for a single cell.

Some technical comments are in order, regarding what will be seen when these methods are used to examine actual experimental data. We have found the method to be robust when used in a thoughtful way. Some features may be anticipated in the multi-cell data themselves.

Because the covariance matrix depends on products of numbers derived from choice of axis pairs that can be made in either order, it is a symmetric matrix. Because of this property it defines a new set of orthogonal directions that have the property that, for new coordinates chosen in those directions, the density distribution would have no correlations between different directions; in terms of pairing those axes, a new correlation matrix would have only diagonal terms. Those diagonal terms state the variances of the univariate Gaussian distributions whose product gives back the multivariate Gaussian. The diagonalization of a symmetric matrix is a standard exercise in linear algebra. The component variances are the so-called "characteristic values" or "eigenvalues" of the matrix. The chosen directions

are the “principal component” directions. When dealing with correlations among stochastic variables, it is common to find that only a few characteristic values are large, and the rest are much smaller. This is a likely outcome in the processing of simultaneous spike data.

Further special structure may be anticipated within the correlation matrix, and this follows from the observation that a good experiment yields spike trains that are stochastically stationary in time. This is to say that the values that appear in the correlation matrix do not depend on the choice of one’s time-origin, and this has consequences even for a single cell whose correlation matrix, at a given frequency, has only the 4 entries of the coefficients of sine and cosine correlated with themselves and with each other. If one shifts the time-origin by $\frac{1}{4}$ of the cycle period, sine becomes cosine while cosine becomes the negative of sine. We can conclude from this that, in the limit of a long data span, the self-correlations of both sine and cosine should be equal. A slightly more detailed use of the quarter-cycle sine-cosine relationship leads to the conclusion that the cross-correlation must be its own negative and hence zero. Thus the correlation matrix is diagonal, and has 2 identical characteristic values.

The correlation matrix for two spike-trains shows the remaining features of the general case. It is natural to view that 4×4 correlation matrix as composed of four 2×2 sub-matrices. By the argument above, we know the form of the two on-diagonal sub-matrices: each has off-diagonal zeroes and identical on-diagonal entries. (The on-diagonal value is larger for the cell whose average firing rate is faster.) From the earlier general argument we know that the two off-diagonal sub-matrices are transposes of each other. Again we can argue that for each of these sub-matrices the on-diagonal entries must be the same. But because the cells are different, an argument similar to that for the one-cell case shows that one of these sub-matrices has off-diagonal entries that are not equal but are indeed the negatives of each other. Put together, these pieces of special structure constrain the 4×4 correlation matrix to have two pairs of identical characteristic values. And this carries over to the general case. With N cells, the $2 \times N$ column correlation matrix yields characteristic values which occur as N pairs of identical values.

Laboratory data are of limited duration and contain residual noise. A correlation matrix obtained from the laboratory will have only an approximation of the structure above, and (as we observe) will yield not identical but closely spaced pairs of characteristic values. Such matrices can be given an evident improvement: For each 2×2 sub-matrix that is off-diagonal, average its two on-diagonal terms and also average each of its off-diagonal elements with the negative of its transpose; for the diagonal sub-matrices set their off-diagonal entry pair to zero. When this simplification is performed one can at once determine numerically that the characteristic values occur in pairs that have equal value. That this is to be expected theoretically follows from a deeper observation. By simple direct computation one checks that the subset of 2×2 matrices, with equal diagonals and skew off-diagonals, under both matrix addition and matrix multiplication, compose exactly as do corresponding complex numbers under ordinary addition and multiplication. Technically, the two systems are isomorphic. Consequently the eigenvalue problem must solve in the same way if the 2×2 sub-matrices are replaced by their isomorphic complex numbers, and this may be done in two different complex-conjugate ways. In fact this has a practical usefulness, as there are Matlab programs which deal directly with the characteristic value problem for the complex-entry matrix, which is of a substantially reduced size.

2.5. Estimating redundancy and synergy

Redundancy refers to the situation where the information carried by a collection of neurons is less than the summed information of the individual members. Synergy refers to the opposite situation, where the information carried by the collection is greater. The topic of synergy or redundancy in neural codes has been the subject of much interest and theoretical discussion (Gawne and Richmond, 1993; Gat

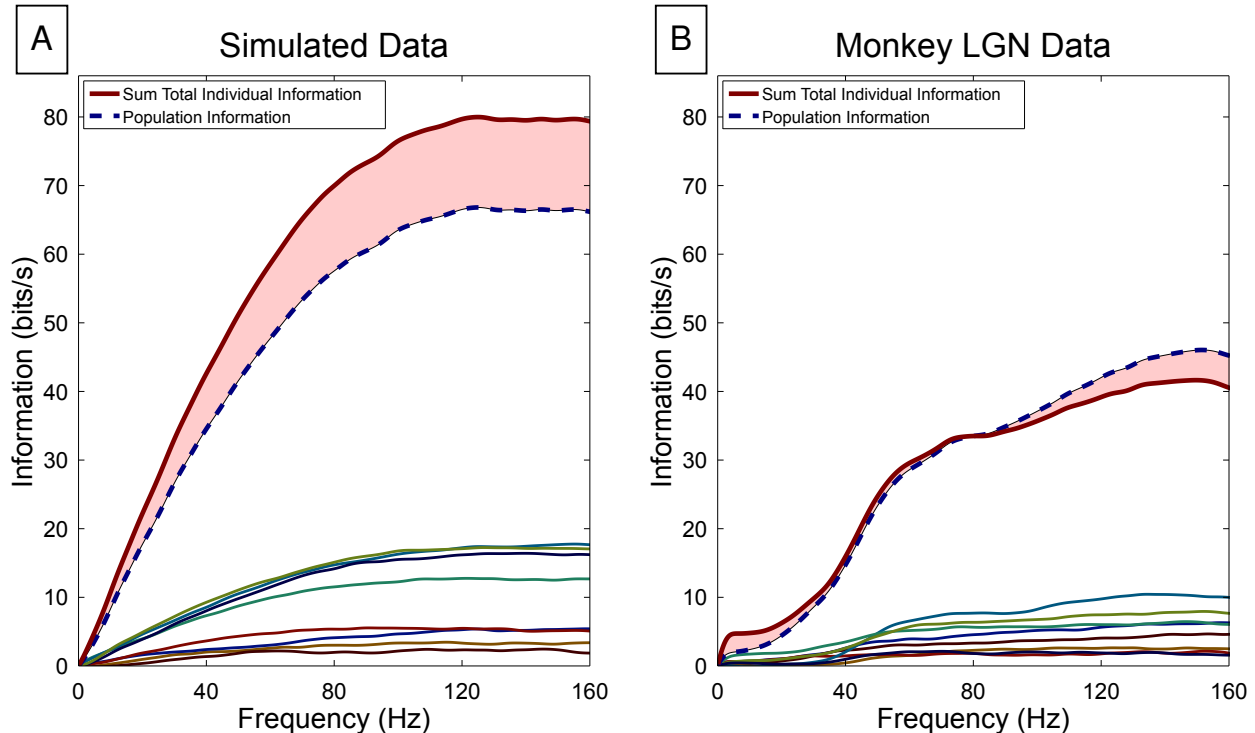


Figure 2: *Cumulative information rates as a function of frequency for single cells and for the population.* Colored lines near the bottom indicate the cumulative individual cell information for eight simulated cells (A) and eight LGN cells (B). In surrogate cells, the sum total information of all individual cells exceeds the information conveyed by the group together, indicating redundancy. In the monkey LGN cells, the sum total exceeds the group information until approximately half the stimulus frequency, after which synergy dominates.

and Tishby, 1999; Panzeri et al., 1999; Brenner et al., 2000; Petersen et al., 2001; Panzeri and Schultz, 2001; Bezzi et al., 2002; Pola et al., 2003; Schneidman et al., 2003; Latham and Nirenberg, 2005; Montani et al., 2007). The method presented here provides the means for quantitatively evaluating redundancy or synergy either among individual neurons or among collections of neurons merged into a larger collection. At the heart of the matter is the role of inherent “noise correlations” (those not induced by a common driving stimulus) within the nervous system, and their role in helping the brain (or the neuroscientist) determine what stimulus elicited the recorded responses. Our method does not explicitly measure correlations among the recorded neurons, but since the method relies on analyzing the *covariance* matrix of the Fourier-transformed spike trains, the correlation effects are included implicitly in the calculation.

2.5.1. Examples

Adding information from individual neurons Figure 2 shows, in its left frame, simulated results from 8 model neurons. These were of the currently much-used Poisson type: each produced an inhomogeneous Poisson point process driven at a time-dependent rate that was the same as that used as a stimulus in the laboratory. The mean rates of these 8 neurons were set at the mean rates of 8 actual neurons recorded in our experiment in the monkey LGN. The lower lines show the cumulative information, with advancing frequency, of those individual simulated neurons. The upper line in red

shows the sum of those 8 cumulative information plots. The dashed blue line below it shows the cumulative information calculated for the merged group of 8. The difference between the two curves is a measure of the redundancy of the information that those neurons individually carry.

The right frame of Figure 2 shows the corresponding calculation for the neurons recorded in our laboratory from the monkey LGN. We see several new features. The individual neurons, not surprisingly, show a low-frequency cutoff. The details of that cutoff are somewhat different for different neurons. When we compare the cumulative information of the merged group to the summed information of the individuals, we see that redundancy at lower frequencies crosses over to synergy at higher frequencies. This recurring observation merits further study.

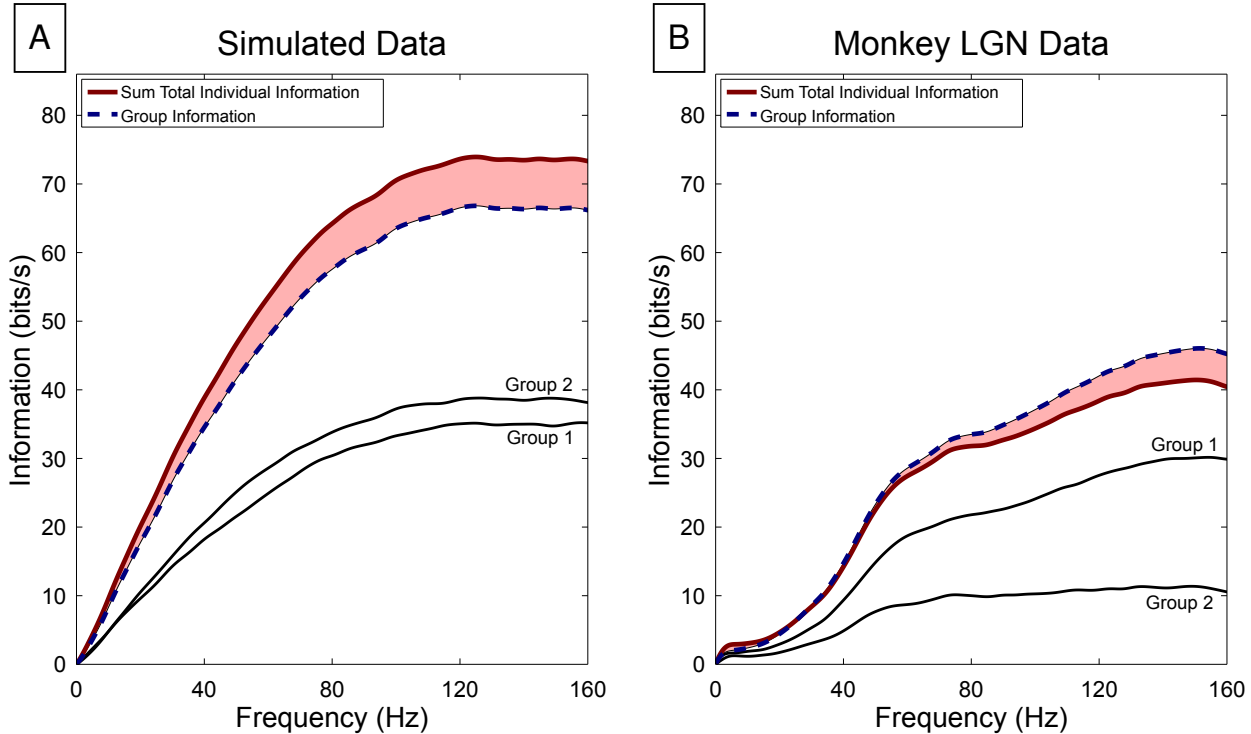


Figure 3: *Redundancy and Synergy among groups of neurons in a population.* Information rates from simulated Poisson spike trains (A), created to match the firing rates of monkey LGN neurons (B). In both panels the 8 cells were split into two groups, matched approximately for firing rates. The total group information is shown in a dashed blue line, and the summed information from two groups is shown in a solid red line. For the LGN cells, but not for the simulated cells, the curves cross around 30 Hz: below 30 Hz we see redundancy, while above it we see synergy.

Merging information from groups of neurons In Figure 3 we examine the effect of merging groups of neurons into a single larger group. For both the simulated neurons and the LGN neurons we have divided the neurons into two groups. We have ranked them in order of increasing mean firing rate, and have put the even and odd numbered neurons in separate groups. From the previous figure for simulated neurons we recall that that merger removed redundant information, so the information of the two groups should already be reduced from the total single neuron information, which is confirmed in the left frame. Similarly for the laboratory data (right frame) the gap is reduced from what the previous figure showed. Again, for the two groups of LGN cells we see that as frequency increases

there is a transition from redundancy to synergy.

3. DISCUSSION

We have described a new method (Yu et al., 2010) for the estimation of the amount of information delivered by the discharges of a neuronal population. The method fills a gap in the armamentarium of the neuroscientist who is interested in the information processing aspects of the brain, and is timely in view of the abundance of multi-neuron recordings that is appearing in the literature. We now mention a few caveats, and comments on two other recent approaches.

3.1. Caveats and challenges for the new method

Repeated stimuli Our method requires repeated presentations of a stimulus in order to calculate the noise entropy. This requirement is shared by other methods, such as the *Direct Method* (Strong et al., 1998). It would be beneficial to have a method that did not require repeated presentations of a stimulus, and which offered some other way of estimating the noise entropy.

Differences between responses to unique and repeat stimuli The methodology presented here confronts the “curse of dimensionality” head-on. In our application, every 8 second spike train is represented as a point in a space of $2 \times 8 \times 160 = 2,560$ dimensions: sine and cosine Fourier coefficients at evenly spaced frequencies from 1/8 Hz to 160 Hz. The challenge becomes tractable when we note that the central limit theorem tells us a great deal about how these points must be distributed in that large space, and, in particular, that correlations across dimensions are confined to two-dimensional sub-spaces defined by a sine-cosine pair at each frequency. The needed computations may be performed one pair at a time, and the central limit theorem further tells us that the final result may be derived from a simple analytic expression.

However, the laboratory provides us with only finite samples, while the theory addresses an ensemble of indefinite size. For example, the third panel in Figure 1 illustrates how an atypical sample may lead to a challenge in data analysis. In that figure we see that total entropy and noise entropy converge at high frequency, which furnishes a cut-off for the sum in eq. 26 of Yu et al. (2010). This convergence is predicted by the theory. But in the laboratory, the noise entropy is estimated from responses to repetitions of a sampled “repeat” stimulus, and if that sample is atypical, the computed noise entropy may converge to a slightly different value than the corresponding total entropy. Knowing the origin of the problem, one can apply a small common sense adjustment to remove it. But currently there is no overall theory to point the way that such small-sample adjustments should be made. We look forward to the input of others who have the kind of experience needed to help bring this approach to a more mature usefulness.

Robustness against errors in spike sorting In our experience, the method is reasonably robust against errors in spike sorting, such as missed spikes, mis-assigned spikes, etc. However, its robustness has limits: if many spikes are erroneously assigned to more than one neuron, this is bound to affect the redundancy/synergy calculation.

Computational efficiency The information calculations illustrated here may be performed on a desk-top computer in a few minutes. Computing time scales roughly with the number of spikes fired by the neuronal population.

Available software The software used in the analysis discussed in this review is freely available at <http://camelot.mssm.edu/~kaplane/ekhomepage.html>.

3.2. Other approaches

We mention here only two recent alternative approaches. For a comprehensive but concise analysis of several other methods, see Victor (2006).

CSSR In addition to calculating the amount of information in the firing of a neural population, there are various other ways of quantifying the richness and complexity of the population’s discharge. For example, the *Causal State Splitting Reconstruction* (CSSR) algorithm (Shalizi et al., 2002; Haslinger et al., 2010) finds the simplest system of hidden Markov states that is consistent with a time series (from one or more sources). If this method proves robust in a multi-neuron context, it offers a means of measuring information without the need for repeated stimulation.

Generalized Linear Model A recent method (Pillow et al., 2011) uses a model-based approach to compute the mutual information between a stimulus and the discharge of a neural population. Unlike our method, which is model-free, their method relies on a computational encoding model whose properties guarantee a unique solution to the decoding problem (“*which stimulus was most likely to produce the population’s response?*”). The model’s parameters are optimized in a training session before the model is used for decoding purposes. We note that decoding the stimulus based on the population’s responses is a somewhat different challenge from calculating how much information about the stimulus the multiple spike trains carry.

Acknowledgments This research was supported by NIH grants EY016371, EY12867, GM71558 and core grant EY01867. We thank Youping Xiao for his help with this project.

References

- Bair, W., Zohary, E., and Newsome, W. T. (2001). Correlated firing in macaque visual area MT: time scales and relationship to behavior. *J Neurosci*, 21(5):1676–1697.
- Bezzi, M., Diamond, M. E., and Treves, A. (2002). Redundancy and synergy arising from pairwise correlations in neuronal ensembles. *J Comput Neurosci*, 12(3):165–174.
- Bialek, W., Rieke, F., de Ruyter van Steveninck, R. R., and Warland, D. (1991). Reading a neural code. *Science*, 252:1854–1857.
- Borst, A. and Theunissen, F. (1999). Information theory and neural coding. *Nature Neurosci.*, 2(11):947–957.
- Brenner, N., Strong, S. P., Koberle, R., Bialek, W., and de Ruyter van Steveninck, R. R. (2000). Synergy in a neural code. *Neural Comput*, 12(7):1531–1552.
- Brown, E. N., Kass, R. E., and Mitra, P. P. (2004). Multiple neural spike train data analysis: state-of-the-art and future challenges. *Nat Neurosci*, 7(5):456–461.
- Cover, T. and Thomas, J. (2006). *Elements of information theory*. Wiley Interscience, NY, second edition.
- DiCaprio, R. A. (2004). Information transfer rate of nonspiking afferent neurons in the crab. *J Neurophysiol*, 92(1):302–310.
- Gat, I. and Tishby, N. (1999). Synergy and redundancy among brain cells of behaving monkeys. In Kearns, M., Solla, S., and Cohn, D., editors, *Advances in neural processing systems*, volume 11, pages 465–471. MIT press.
- Gawne, T. J. and Richmond, B. J. (1993). How independent are the messages carried by adjacent inferior temporal cortical neurons? *J Neurosci*, 13(7):2758–2771.
- Haslinger, R., Klinkner, K. L., and Shalizi, C. R. (2010). The computational structure of spike trains. *Neural Comput*, 22(1):121–157.
- Latham, P. E. and Nirenberg, S. (2005). Synergy, redundancy, and independence in population codes, revisited. *J Neurosci*, 25(21):5195–5206.
- Montani, F., Kohn, A., Smith, M. A., and Schultz, S. R. (2007). The role of correlations in direction and contrast coding in the primary visual cortex. *J Neurosci*, 27(9):2338–2348.
- Optican, L. M. and Richmond, B. J. (1987). Temporal encoding of two-dimensional patterns by single units in primate inferior temporal cortex. III. information theoretic analysis. *J. Neurophysiol.*, 57:162–178.
- Panzeri, S. and Schultz, S. R. (2001). A unified approach to the study of temporal, correlational, and rate coding. *Neural Comput*, 13(6):1311–1349.
- Panzeri, S., Schultz, S. R., Treves, A., and Rolls, E. T. (1999). Correlations and the encoding of information in the nervous system. *Proc Biol Sci*, 266(1423):1001–1012.
- Petersen, R. S., Panzeri, S., and Diamond, M. E. (2001). Population coding of stimulus location in rat somatosensory cortex. *Neuron*, 32(3):503–514.

- Pillow, J. W., Ahmadian, Y., and Paninski, L. (2011). Model-based decoding, information estimation, and change-point detection techniques for multineuron spike trains. *Neural Comput*, 23(1):1–45.
- Pillow, J. W., Shlens, J., Paninski, L., Sher, A., Litke, A. M., Chichilnisky, E. J., and Simoncelli, E. P. (2008). Spatio-temporal correlations and visual signalling in a complete neuronal population. *Nature*, 454(7207):995–999.
- Pola, G., Thiele, A., Hoffmann, K. P., and Panzeri, S. (2003). An exact method to quantify the information transmitted by different mechanisms of correlational coding. *Network*, 14(1):35–60.
- Quiroga, R. Q. and Panzeri, S. (2009). Extracting information from neuronal populations: information theory and decoding approaches. *Nat Rev Neurosci*, 10(3):173–185.
- Richmond, B. J. and Optican, L. M. (1987). Temporal encoding of two-dimensional patterns by single units in primate inferior temporal cortex. II. quantification of response waveform. *J. Neurophysiol.*, 57:147–161.
- Richmond, B. J., Optican, L. M., Podell, M., and Spitzer, H. (1987). Temporal encoding of two-dimensional patterns by single units in primate inferior temporal cortex. I. response characteristics. *J. Neurophysiol.*, 57:132–146.
- Rieke, F., Warland, D., Steveninck, R. d., and Bialek, W. (1997). *Spikes: Exploring the Neural Code*. MIT Press, Cambridge, MA.
- Schneidman, E., Bialek, W., and Berry, M. J. (2003). Synergy, redundancy, and independence in population codes. *J Neurosci*, 23(37):11539–11553.
- Shalizi, C. R., Shalizi, K. L., and Crutchfield, J. P. (2002). An algorithm for pattern discovery in time series. *Technical Report, Santa Fe Institute*, pages 1–26. <http://arxiv.org/abs/cs.LG/0210025>.
- Shannon, C. and Weaver, W. (1949). *A Mathematical theory of communication*. Univ. of Illinois Press, Chicago, IL.
- Shannon, C. E. (1949). Communication in the presence of noise. *Proceedings of the IEEE*, 37(1):10–21.
- Strong, S. P., Koberle, R., de Ruyter van Steveninck, R. R., and Bialek, W. (1998). Entropy and information in neural spike trains. *Phys Rev Lett.*, 80:197–200.
- Victor, J. D. (2006). Approaches to information-theoretic analysis of neural activity. *Biological Theory*, 1(3):302–316.
- Yu, Y., Crumiller, M., Knight, B., and Kaplan, E. (2010). Estimating the amount of information carried by a neuronal population. *Frontiers in Computational Neuroscience*, 4:10.
- Zohary, E., Shadlen, M. N., and Newsome, W. T. (1994). Correlated neuronal discharge rate and its implications for psychophysical performance. *Nature*, 370(6485):140–143.