Available online at www.sciencedirect.com

SciVerse ScienceDirect

Current Opinion in Neurobiology

Information theoretic approaches to understanding circuit function

Adrienne Fairhall[1], Eric Shea-Brown[2] and Andrea Barreiro[3]

The analysis of stimulus/response patterns using information theoretic approaches requires the full probability distribution of stimuli and response. Recent progress in using information-based tools to understand circuit function has advanced understanding of neural coding at the single cell and population level. In advances over traditional reverse correlation approaches, the determination of receptive fields using information as a metric has allowed novel insights into stimulus representation and transformation. The application of maximum entropy methods to population codes has opened a rich exploration of the internal structure of these codes, revealing stimulus-driven functional connectivity. We speculate about the prospects and limitations of information as a general tool for dissecting neural circuits and relating their structure and function.

Addresses
[1] Department of Physiology and Biophysics, University of Washington, Seattle, WA 98195-7290, USA
[2] Department of Applied Mathematics, University of Washington, Seattle, WA 98195, USA
[3] Department of Mathematics, Southern Methodist University, Dallas, TX 75275, USA

Corresponding author: Fairhall, Adrienne (fairhall@uw.edu, fairhall@u.washington.edu) and

Current Opinion in Neurobiology 2012, **22**:653–659

This review comes from a themed issue on **Microcircuits**

Edited by **Edward M Callaway** and **Eve Marder**

For a complete overview see the Issue and the Editorial

Available online 12th July 2012

0959-4388/$ – see front matter, © Elsevier Ltd. All rights reserved.

http://dx.doi.org/10.1016/j.conb.2012.06.005

Introduction

Over the past twenty years, information theory has become a central part of the arsenal of analysis tools in neuroscience. Mutual information is a measure of the correlation between two variables [1] and allows one to evaluate the quality of a proposed neuronal representation. It therefore plays a dual role in neuroscience. First, it is a method for discovering and interpreting correlational structure in inputs and outputs. While the examples we will consider here focus on sensory coding, the developments we consider should ultimately allow for generalizations of the notion of input and output to correlated firing in complex circuits. Second, it is a normative theory, providing a well-defined quantity that a system – via a coding strategy, a specific circuit dynamics, or even a behavior – might be expected to maximize.

Information measures

The uncertainty in a distribution of a random variable S is quantified by its **entropy**,

$$H[S] = -\int ds\, P(s) \log_2 P(s). \quad (1)$$

The **mutual information** quantifies the average gain in certainty about one variable S owing to knowledge of the other, R:

$$I(R;S) = H[S] - \int dr\, P(s) H[R|s] \quad (2)$$

$$I(R;S) = H[R] - \int ds\, P(r) H[S|r]. \quad (3)$$

When the response, R, is a binary variable, and the event is generally rare, the probability distribution of inputs associated with the non-event is very similar to the prior probability, and the information is approximated by the difference in entropy between the event-triggered stimuli and all stimuli. The MI between a spike and the stimulus can then be expressed as

$$I(S;\text{spike}) = \int ds\, P(s,\text{spike}) \log_2 \frac{P(s|\text{spike})}{P(s)}, \quad (4)$$

which is the **Kullback–Leibler divergence** between the joint distribution of inputs, $P(S)$, and those inputs that are associated with a spike, $P(S \mid \text{spike})$. When considering the output of more than one responder to an input, one can ask how much additional knowledge of the input is gained by measuring these responses simultaneously:

$$\Delta I(R_1,R_2;S) = I(R_1,R_2;S) - I(R_1;S) - I(R_2;S) \quad (5)$$

Synergy occurs when $\Delta I > 0$, and the two responses are more informative about the input when recorded together than when they are treated as independent. The responses are **redundant** when $\Delta I < 0$, which occurs when some of the information conveyed by the two responses is the same. There are several other related quantities such the **Fisher information** that are used to quantify the success of decoding; we will not discuss these here.

Here we will discuss recent developments in the applications of information methods to characterize neural coding at the single cell and population level. The evaluation of correlation using information depends on estimation of the entire probability distribution. Obtaining good estimates can be very challenging with limited data [2,3••], and progress will depend on intelligent simplifications of high-dimensional distributions. We will highlight advances in the use of approximation methods that are moving toward an increasingly complete description of probabilistic encoding of inputs by populations of neurons. A very important open question is the interpretation of observed statistical correlations in terms of the structure of underlying neuronal circuitry. We review recent work that has begun to address the relationship between the statistics of neural firing and structure.

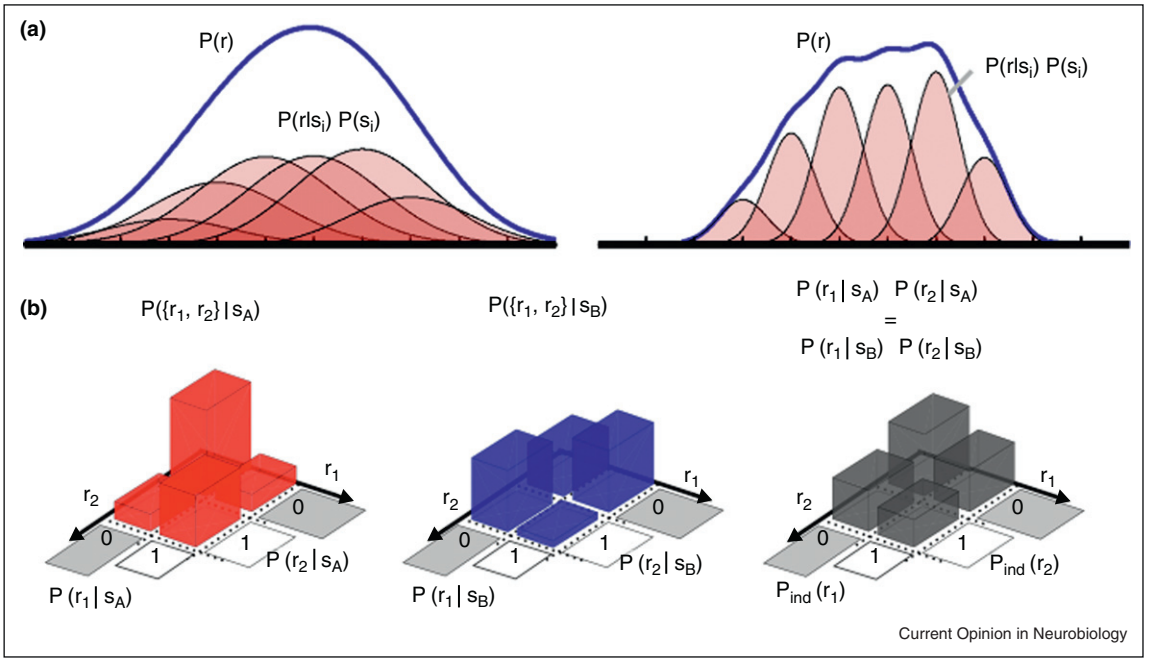
Information theory provides a formalism for quantifying the relationship between two (or more) random variables. A typical application in neuroscience is to quantify the codependence between an input **s** and a putative neural representation, **r**, where **r** might be the timing of a single spike, a spike pattern, the firing rate or a spatially averaged quantity such as the local field potential. The inputs and responses have distributions $P(\mathbf{s})$ and $P(\mathbf{r})$ respectively. If the response is predicted by **s**, knowledge about **s** should restrict the possible outputs **r**, $P(\mathbf{r}|\mathbf{s})$. This reduction in uncertainty is quantified by the mutual information, Eq. (2) and Figure 1.

Mutual information (MI) can be used to evaluate the ability of an output representation to convey stimulus information, without needing to invoke a decoding mechanism to extract that information, and in units that are stimulus-independent and response-independent [4•]. MI has been used to determine the temporal resolution of spike timing that carries maximal information [5,6], to test the role of complex spike patterns such as bursts in conveying stimulus information [7], and to compare the information content of features of neural activity on different timescales, such as spike timing and the local field potential [8]. Different neural 'symbols' may multiplex different components of stimulus information [9,10]. Definitions of the mutual information can be generalized to multiple variables [11].

Information and receptive fields

The concept that neuronal receptive fields should minimize redundancy, or shared information (see Box) [12], in neural coding [13,14] was used to predict that

Figure 1



Information encoded by the neural response can be quantified by the mutual information between response and stimulus $I(\mathbf{s};\mathbf{r})$. **(A)** The response distribution is given by $P(\mathbf{r}) = \int P(s_i)P(\mathbf{r}|s_i)$. When specifying a value of **s**, for example, s_i, significantly reduces uncertainty, or narrows the distribution, about **r**, the mutual information between **r** and **s** is large (left). The more precisely that **s** specifies **r**, the larger the information (right). **(B)** Information and correlations. Here, two stimuli s_A and s_B generate binary responses $\{r_1,r_2\}$ with identical marginal distributions ($p(r_1|s_A) = p(r_1|s_B)$ and $p(r_2|s_A) = p(r_2|s_B)$), yet differing joint responses ($p(\{r_1,r_2\}|s_A) \neq p(\{r_1,r_2\}|s_B)$). If correlations are ignored, $I(\mathbf{s},\mathbf{r}) = 0$; if correlations are maintained and the stimuli are equally likely, then $I(\mathbf{s};\mathbf{r}) \sim 0.23$ bits. From left to right: neural response to s_A, with $p(r_1 = 1|s_A) = p(r_2 = 1|s_A) = 0.4$ but response covariance $p(r_1r_2 = 1|s_A) - p(r_1 = 1|s_A)p(r_2 = 1|s_A) = 0.14$; neural response to s_B, with $p(r_1 = 1|s_B) = p(r_2 = 1|s_B) = 0.4$ but response covariance $p(r_1r_2 = 1|s_B) - p(r_1 = 1|s_B)p(r_2 = 1|s_B) = -0.12$; the neural response neglecting covariances is identical for each stimulus, carrying no information.

retinal receptive fields should act to cancel out common, or predictable, inputs [15,16]. However, the underlying assumption of redundancy reduction has recently been challenged by the observation of considerable redundancy in retinal ganglion cells arising from the overlapping of their receptive fields [17]. Furthermore, while the filtering properties of RGC receptive fields do significantly decorrelate responses to natural images, neuronal threshold nonlinearities turn out to play an even more significant role [18].

Information can be used directly to evaluate how well receptive fields capture the true feature selectivity of neural systems. It thus also provides a method for searching for receptive fields. Eq. (4) can be used to quantify the information captured by a reduced model: that is, one that replaces the stimulus by its similarity to, or projection into, a low-dimensional feature space **f** [19,20]. Maximizing Eq. (4) amounts to searching for a set of features **f**, Figure 2, that maximizes the difference between the entire set of stimuli, viewed in that basis, and the set of stimuli that are associated with spikes. These **maximally informative dimensions (MID)** then maximize the mutual information between the reduced stimulus and spike over the explored stimulus ensemble [21,22].

This powerful method liberates reverse correlation techniques for determining receptive fields from the use of Gaussian white noise stimuli. This allows one to estimate the feature selectivity of systems using more natural stimuli [23] – in some cases driving neurons that may not even be well-stimulated using Gaussian white noise – and can expose changes in processing as a result of adaptation to the stimulus ensemble. For example, Sharpee *et al.* [24] found that the filtering properties of neurons in V1 adjust when processing natural images compared to with white noise.

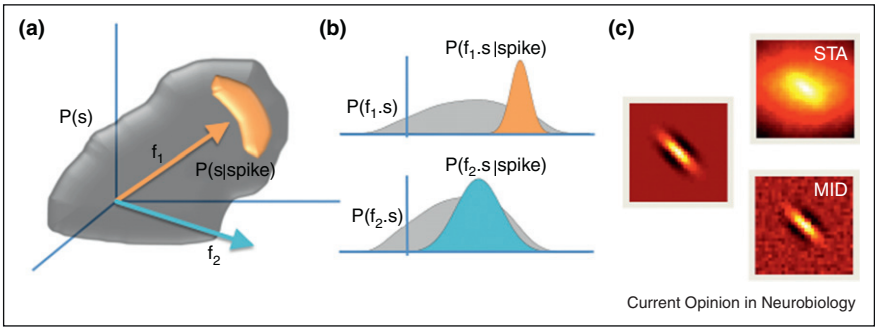
The MID method can be applied to understand the transformations that occur between subsequent processing stages. According to the data processing inequality, a transformation of data cannot increase its information content [1]; however, the retina–LGN synapse was shown to preserve information almost perfectly, despite a reduction in firing rate [25]. Furthermore, this method can allow one to track changes in the structure or dimensionality of the feature set at different processing stages [26•]. Identification of multi-component receptive fields motivates and guides the search for specific circuit mechanisms of stimulus filtering. The MID method has two downsides, however, that may limit some practical applications. Estimating the required probability distributions, particularly for multiple stimulus dimensions, can require daunting amounts of data. Moreover, this optimization has no global solution and the procedure may get caught in local minima.

Population coding

The advent of high-throughput imaging and recording techniques allows simultaneous recording of hundreds of well-isolated single units. Information allows one to determine the statistical relationships between patterns of neural response and an input, whether that input is sensory or expressed in terms of the activity of another set of neurons. However, to apply these methods to neuronal populations, one must estimate the probability of occurrence of all combinations of stimuli and the responses $\{r_j\}_{j=1,N}$ of a population of N neurons – $P(r_1, r_2, \ldots, r_N;$ stimulus). Any direct approach requires sampling an enormous number of states that increases exponentially with N and rapidly outstrips the capacity of any possible experiment [27].

A possible resolution comes from the use of lower order approximations to the full population response. At the

Figure 2



Finding stimulus dimensions with information maximization. **(A)** The MID method searches for a direction in stimulus space, that is, a filter **f**. The stimulus **s** is high-dimensional and is described by some distribution, the gray cloud $P(\mathbf{s})$. One then takes the component of **s** along a direction **f**. **(B)** This projected stimulus, $\mathbf{f} \cdot \mathbf{s}$, has a (one-dimensional) distribution $P(\mathbf{f} \cdot \mathbf{s})$. This is known as the prior distribution. The spike-triggered stimuli (denoted in orange), have distribution $P(\mathbf{f} \cdot \mathbf{s}|$ spike). The goal is to determine **f** such that the Kullback–Leibler distance (see Box) between these two distributions is maximized; here, direction $\mathbf{f}_1$ improves the separation of the distributions compared with $\mathbf{f}_2$. Finding the direction **f** that maximizes the Kullback–Leibler distance is equivalent to maximizing the mutual information between the projected stimulus and the occurrence of a spike. **(C)** For a model cell with the filter shown on the left and a sigmoidal nonlinearity, driven with natural images, the MID method recovers the true filter considerably better than the spike-triggered average (with thanks to T. Sharpee).

extreme, one might assume that each neuron responds independently: $P(r_1, r_2, \ldots, r_N) \sim \Pi_{i=1}^{N} P(r_i)$. In general, of course, this assumption is too strong: neurons are driven by common stimuli owing to overlapping receptive fields. Thus, a weaker assumption is that neurons respond independently, once taking into account possibly overlapping stimulus sensitivities, or **signal correlation**: $P(r_1, r_2, \ldots, r_N|\text{stimulus}) \sim \Pi_{i=1}^{N} P(r_i|\text{stimulus})$.

However, neurons may share input variability that is not stimulus-related, and thus show correlations beyond those predicted by the input. This is known as **noise correlation** [12]. Noise correlations can have a significant impact on the efficacy of decoding information from observations [28,29••]; depending on the form of this correlation, it may help or hinder decoding [30,31] (PE Latham and Y Roudi, 2011, http://arxiv.org/abs/1109.6524v1).

Maximum entropy approaches

Thus, the next step in developing an improved approximation for the response distribution is to take into account noise correlations, but in direct attempts to capture $P(r_1, r_2, \ldots, r_N \mid \text{stimulus})$ we are still faced with a problem of unmanageable scale. One approach is to specify only pairwise noise correlations – dependencies among two cells at a time – and to extrapolate from here.

Temporarily putting aside the question of stimulus dependence, let us first consider the response space, $P(\mathbf{r})$. One would like to pose an approximate form for this potentially very complex distribution of responses that will take into account the observations one can reasonably make, while making minimal assumptions. The natural model that does this is known as the **maximum entropy distribution**, which is the distribution that has the most entropy and therefore makes the fewest assumptions about the structure of the distribution. It has the form:

$$P(\mathbf{r}) = Z^{-1} \exp\left(\sum_i h_i r_i + \sum_{i,j} J_{ij} r_i r_j + \cdots\right), \qquad (6)$$

where Z is a normalization factor, and the parameters h_i, J_{ij},..., are fixed to reproduce the observed moments. For example, knowing the mean (first-order) responses $\langle r_i \rangle$ will fix the parameters h_i. Including the terms with J_{ij} allows one to match the second order moments, or the correlations between the firing of any neuron pair, $\langle r_i r_j \rangle$. One can then compare the observed frequency of specific patterns with the probability computed from this model. Such second-order models have been shown to improve the prediction of network patterns significantly [27,28,32,33] over those of independent models.

Despite this improvement, the second-order maximum entropy description fails to completely capture the distribution of patterns in large networks, implying the need to incorporate higher-order structure. To capture these 'beyond-pairwise' interactions, recent approaches retain selected higher-order terms in the probability model, Eq. (5). For example, the distribution of summed population spike count in somatosensory cortex – that is, the total number of simultaneously active units out of the 24 recorded – could be described by including third order terms [34•]. Encouragingly, this suggests that a relatively low-order correction may achieve the correct structure for significantly larger populations. Ganmor *et al.* [29••] aimed to capture not just the summed population output, but the entire distribution of spike patterns, in recordings of ~100 retinal ganglion cells. Their insight was to only include in Eq. (5) selected terms corresponding to frequently occurring responses. This 'reliable interaction' approach cleverly avoids the need to include every term at a given order. Equating the model probability of patterns that occur frequently to their well-sampled empirical frequencies produces a set of linear equations for the model parameters. This fit is thus rapid, easily learnable and is guaranteed by construction to reproduce the best-known observed probabilities.

Stimulus dependence

Ultimately, one needs to resolve response distributions as a function of the stimulus. Returning to the characterization of single neuron responses, one can again use the idea of the maximum entropy distribution to simplify stimulus dependence. Unlike maximizing information (Figure 2), maximizing entropy is a convex and therefore relatively easily solved problem. One can then instead apply a maximum entropy approach to the response distributions, conditioned on the stimulus, also known as the noise entropy. The constraints will now apply not to the correlations between responses, but to correlations between stimulus and response [35•,36••,37]. Constraining only the first order correlation, $\langle \mathbf{r} \mid \mathbf{s} \rangle$ – or the spike-triggered average – results in a logistic response function that depends on a single stimulus dimension. Including the next order terms means that the spike-triggered covariance [20] also appears in the logistic function, which allows for some of the complex multidimensional response functions that are observed in data [38]. Given this flexibility, such 'minimal models' are likely to give good fits to a wide variety of sensory responses.

The correlated population models described previously have so far only captured fairly coarse-grained stimulus dependence. Third-order interactions were necessary to capture information about whisker vibrations in rat somatosensory cortex [34•]. In retina, spike-triggered average stimuli for the significantly correlated cell triplets or quadruplets differ from the spike-triggered averages of the constituent pairs, indicating that these higher-order interactions do convey distinct stimulus information [29••]. To date, these methods applied to populations do not fully take into account evolution in time.

Ultimately, we expect that a maximum entropy approach that jointly constrains stimulus and response correlation and incorporates their temporal dynamics is a difficult but likely path to a complete population coding model [39].

One may also ask about the ability of a maximum entropy decoder to extract information from the observed responses (cf [40,31], Latham and Roudi (2011)). In [29••], the use of selected higher order terms translated into a three-fold speed increase in classifying new stimulus segments, compared with the pairwise model. On the contrary, Oizumi *et al.* [41] find that despite significant pairwise correlations in spike responses in RGCs, an independent decoder could extract 95% of the information present about the identity of a natural scene movie. Understanding and building intuition for how and why these cases differ is an important next step for information-based methods in neuroscience.

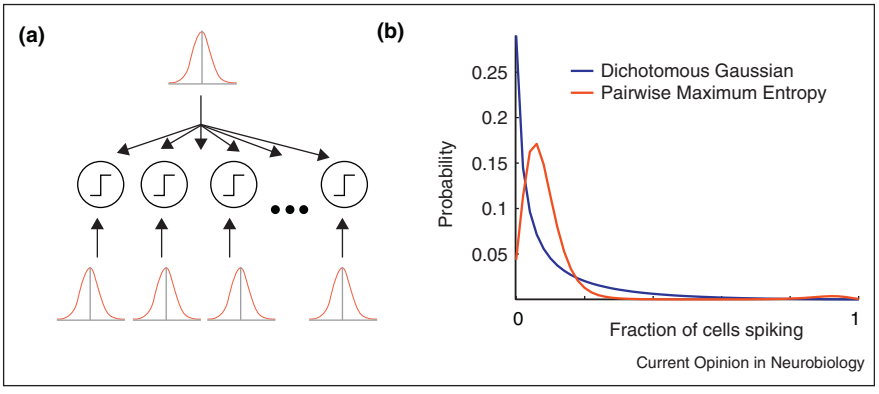
What do correlation measures reveal about circuit structure and mechanism?

Does the existence of higher-order interactions in Eq. (5) predict the existence of corresponding features of the physical circuit? Answers to this question so far are mixed, and point to fruitful opportunities for future research. For example, a pairwise maximum entropy model produces extremely accurate descriptions of population activity in the retina [42], and the extent of statistical interactions agreed with general expectations from retinal anatomy. However, it has been shown that pairwise models can always be expected to fit population activity provided the firing rate is sufficiently low, regardless of the underlying circuit [43]. Barreiro and colleagues (2012, http://arxiv.org/abs/1011.2797v3) find examples in which purely pairwise inputs to triplets of cells can yield very strong third-order interaction terms, and Roudi *et al.* [44] find that in large, sparse network simulations, pairwise statistical interactions in maximum-entropy models do not predict direct synaptic connections.

We anticipate that further understanding of the relationship between structure and statistics will come from probabilistic models that adopt more mechanistic features than the statistical description of Eq. (5) – but without becoming so complex that they are impossible to fit (cf. [44]). One recent example is the dichotomous Gaussian method [45••], which has been fit to *in vivo* cortical slice data [46•]. Here, spikes are generated by passing a correlated Gaussian signal through a threshold – which may be viewed as a crude 'binary' model of neural spiking in response to overlapping, common inputs (see Figure 3). Surprisingly, this model contains the same number of parameters ($\sim N^2$) as the pairwise maximum entropy model (Eq. (5), truncated at second order), but is considerably more accurate in capturing population-wide spiking patterns (for up to 56 cells) and multichannel LFP patterns (approximately 100 channels). In related work, generalized linear models [47] using $\sim N^2$ parameters have been successfully used to model population-wide activity in retina – and have been able to recover qualitative features of circuit connectivity, such as the relative contributions of common input vs. reciprocal coupling [28,48•].

We anticipate further advances in the ability of these and related spiking models to uncover circuit connectivity, together with circuit mechanisms such as distributions of synaptic inputs (Barreiro *et al.*, 2012, http://arxiv.org/abs/1011.2797v3). An important challenge for these models will be to identify the circuit mechanisms behind aspects of the population response that contribute most to encoding

Figure 3



A schematic of the dichotomous Gaussian model of population activity (simplified version). **(A)** *N* cells receive a common Gaussian input, as well as independent inputs; the summed input is compared to a threshold to either generate a spike or silence at each timestep. **(B)** Intriguingly, the output of this idealized model of the thresholding mechanism of spike generation produces highly different population statistics than the corresponding pairwise maximum entropy model [45••,46•]. The red and blue lines show the fraction out of $N = 50$ model cells that spike simultaneously in a given timestep; both models are fixed to have the same firing rate and pairwise correlation [45••].

of stimuli. Models provide a rapid testbed for this question by allowing us to compare different circuit mechanisms that conserve basic features of the population response (such as firing rates and pairwise correlations) but differ in more subtle aspects – such as higher order statistics of the population response (cf. [48•,49] and Figure 3).

Normative models

Aside from stimulus/response characterization, one of the most intriguing if speculative and sometimes controversial uses of information is as an objective function to explain coding and behavioral strategies. The ability of many neural systems to encode stimuli in units that are rapidly normalized by a changing input range allows them to maintain high rates of information transmission [50]. Patterns in behavioral search strategies may be qualitatively predicted as a search for information; the zigzag flight pattern of a moth searching for the source of a plume of pheromone has been proposed to driven by 'infotaxis', an effort to reduce uncertainty about the stimulus [51,52]. Similarly, the trajectories of eye movements are consistent with a theory in which they are driven by entropy reduction [53,54].

Conclusions

Information theoretic tools are generating new methods to determine rich models for neural stimulus representation. The use of maximum entropy distributions has had tremendous recent impact, driving rapid progress in building complete population response models. As yet, these methods are limited in their ability to capture complex stimulus dependence and temporal dynamics. An interplay between these statistical methods and alternative generative models of network activity will be needed to interpret the observations of higher-order correlations and their significance in terms of underlying circuit motifs. However, such studies are likely to provide a framework for linking future connectomic data with functional models of stimulus representation.

Acknowledgements

ALF acknowledges support from NSF 0928251 and NIH 1R21NS072691, and ESB from the Burroughs Wellcome Fund Scientific Interfaces Program and NSF Grants 1056125, 0818153, and 1122106. We thank Tatyana Sharpee for providing Figure 2C and David Leen (UW Applied Mathematics) for the algorithm and MATLAB coding that produced Figure 3B.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Cover TM, Thomas JA: *Information Theory*. New York: John Wiley & Sons, Inc.; 1991.

2. Nemenman I, Bialek W, de Ruyter van Steveninck R: **Entropy and information in neural spike trains: progress on the sampling problem**. *Phys Rev E Stat Nonlinear Soft Matter Phys* 2004, **69**:56111.

3. •• Ince RA, Senatore R, Arabzadeh E, Montani F, Diamond ME, Panzeri S: **Information-theoretic methods for studying population codes**. *Neural Netw* 2010, **23**:713-727.
This is an extensive review of recent literature applying information to characterize neural systems, particularly helpful for its discussion of multiple recent approaches to finite size corrections.

4. • Quian Quiroga R, Panzeri S: **Extracting information from neural populations: information theory and decoding approaches**. *Nat Rev Neurosci* 2009, **10**:173-185.
This pedagogical review of information theoretic approaches focuses on the relationship between information and decoding methods.

5. Reinagel P, Reid RC: **Temporal coding of visual information in the thalamus**. *J Neurosci* 2000, **20**:5392-5400.

6. Nemenman I, Lewen GD, Bialek W, de Ruyter van Steveninck RR: **Neural coding of natural stimuli: information at sub-millisecond resolution**. *PLoS Comput Biol* 2008, **4**:e1000025.

7. Arabzadeh E, Panzeri S, Diamond ME: **Whisker vibration information carried by rat barrel cortex neurons**. *J Neurosci* 2004, **24**:6011-6020.

8. Belitski A, Panzeri S, Magri C, Logothetis NK, Kayser C: **Sensory information in local field potentials and spikes from visual and auditory cortices: time scales and frequency bands**. *J Comput Neurosci* 2010, **29**:533-545.

9. Fairhall AL, Lewen GD, Bialek W, de Ruyter Van Steveninck RR: **Efficiency and ambiguity in an adaptive neural code**. *Nature* 2001, **412**:787-792.

10. Panzeri S, Brunel N, Logothetis NK, Kayser C: **Sensory neural codes using multiplexed temporal scales**. *Trends Neurosci* 2010, **33**:111-120.

11. Schneidman E, Still S, Berry M, Bialek W: **Network information and connected correlations**. *Phys Rev Lett* 2003, **91**:238701.

12. Schneidman E, Bialek W, Berry MJ II: **Synergy, redundancy, and independence in population codes**. *J Neurosci* 2003, **23**:11539-11553.

13. Barlow HB: **Possible principles underlying the transformation of sensory messages**. In *Sensory Communication.* Edited by Rosenblith W. Cambridge, MA: MIT Press; 1961.

14. Attneave F: **Some informational aspects of visual perception**. *Psychol Rev* 1954, **61**:183-193.

15. Srinivasan MV, Laughlin SB, Dubs A: **Predictive coding: a fresh view of inhibition in the retina**. *Proc R Soc Lond B Biol Sci* 1982, **216**:427-459.

16. Atick: **Could information theory provide an ecological theory of sensory processing?** *Network* 1992, **3**:213-252.

17. Puchalla JL, Schneidman E, Harris RA, Berry MJ: **Redundancy in the population code of the retina**. *Neuron* 2005, **46**:493-504.

18. Pitkow X, Meister M: **Decorrelation and efficient coding by retinal ganglion cells**. *Nat Neurosci* 2012, **15**:628-635.

19. Fairhall AL, Burlingame CA, Narasimhan R, Harris RA, Puchalla JL, Berry MJ 2nd: **Selectivity for multiple stimulus features in retinal ganglion cells**. *J Neurophysiol* 2006, **96**:2724-2738.

20. Schwartz O, Pillow JW, Rust NC, Simoncelli EP: **Spike-triggered neural characterization**. *J Vis* 2006, **6**:484-507.

21. Sharpee T, Rust NC, Bialek W: **Analyzing neural responses to natural signals: maximally informative dimensions**. *Neural Comput* 2004, **16**:223-250.

22. Pillow JW, Simoncelli EP: **Dimensionality reduction in neural models: an information-theoretic generalization of spike-triggered average and covariance analysis**. *J Vis* 2006, **6**:414-428.

23. Simoncelli EP, Olshausen BA: **Natural image statistics and neural representation**. *Annu Rev Neurosci* 2001, **24**:1193-1216.

24. Sharpee TO, Sugihara H, Kurgansky AV, Rebrik SP, Stryker MP, Miller KD: **Adaptive filtering enhances information transmission in visual cortex**. *Nature* 2006, **439**:936-942.

25. Sincich LC, Horton JC, Sharpee TO: **Preserving information in neural transmission**. *J Neurosci* 2009, **29**:6207-6216.

26. • Atencio CA, Sharpee TO, Schreiner CE: **Receptive field dimensionality increases from the auditory midbrain to cortex**. *J Neurophysiol* 2012, **107**:2594-2603.
A recent example of the application of maximally information dimension methods to evaluate transformation in coding, finding an increase in the number of relevant stimulus dimensions.

27. Schneidman E, Berry MJ II, Segev R, Bialek W: **Weak pairwise correlations imply strongly correlated network states in a neural population**. *Nature* 2006, **440**:1007-1012 http://dx.doi.org/10.1038/nature04701.

28. Pillow J, Shlens J, Paninski L, Sher A, Litke A, Chichilnisky EJ, Simoncelli EP: **Spatiotemporal correlations and visual signaling in a complete neuronal population**. *Nature* 2008, **454**:995-999.

29. •• Ganmor E, Segev R, Schneidman E: **Sparse low-order interaction network underlies a highly correlated and learnable neural population code**. *Proc Natl Acad Sci USA* 2011, **108**:9679-9684.
Demonstrates a data-driven method for incorporating selected higher order interactions into a pairwise-maximum entropy fit in retina.

30. Abbott L, Dayan P: **The effect of correlated variability on the accuracy of a population code**. *Neural Comput* 1999, **11**:91-101.

31. Averbeck BB, Lee D: **Effects of noise correlations on information encoding and decoding**. *J Neurophysiol* 2006, **95**:3633-3644 http://dx.doi.org/10.1152/jn.00919.2005.

32. Shlens J, Field G, Gauthier J, Grivich M, Petrusca D, Sher A, Litke AM, Chichilnisky EJ: **The structure of multi-neuron firing patterns in primate retina**. *J Neurosci* 2006, **26**:8254.

33. Tang A, Jackson D, Hobbs J, Chen W, Smith JL, Patel H, Prieto A, Petrusca D, Grivich MI, Sher A *et al.*: **A maximum entropy model applied to spatial and temporal correlations from cortical networks in vitro**. *J Neurosci* 2008, **28**:505-518.

34. • Montani F, Ince RA, Senatore R, Arabzadeh E, Diamond ME, Panzeri S: **The impact of high-order interactions on the rate of synchronous discharge and information transmission in somatosensory cortex**. *Philos Trans A Math Phys Eng Sci* 2009, **367**:3297-3310.
Computes mutual information between whisker vibration velocity and population level response in somatosensory cortex, showing that third-order interactions are sufficient to capture the information content in the response of a group of roughly 20 neural 'clusters' (of 2–5 cells each).

35. • Globerson A, Stark E, Vaadia E, Tishby N: **The minimum information principle and its application to neural code analysis**. *Proc Natl Acad Sci USA* 2009, **106**:3490-3495.
Introduces the concept of the minimum information principle, which determines the minimum information that a given set of observations contains about the stimulus.

36. •• Fitzgerald JD, Sincich LC, Sharpee TO: **Minimal models of multidimensional computations**. *PLoS Comput Biol* 2011, **7**:e1001111.
Derives a minimal model of responses to neural stimuli by constructing maximum noise entropy (or minimum information) solutions that are constrained by successive order moments of the stimulus–response distribution.

37. Fitzgerald JD, Rowekamp RJ, Sincich LC, Sharpee TO: **Second order dimensionality reduction using minimum and maximum mutual information models**. *PLoS Comput Biol* 2011, **7**:e1002249.

38. Fairhall AL, Burlingame CA, Puchalla J, Harris R, Berry II MJ: **Feature selectivity in retinal ganglion cells**. *J Neurophysiol* 2006, **96**:2724-2738.

39. Tishby N, Pereira FC, Bialek W: **The information bottleneck method**. In *Proceedings of the 37-th Annual Allerton Conference on Communication, Control and Computing*. 1999:368-377.

40. Latham P, Nirenberg S: **Synergy, redundancy, and independence in population codes, revisited**. *J Neurosci* 2005, **25**:5195.

41. Oizumi M, Ishii T, Ishibashi K, Okada M: **Mismatched decoding in the brain**. *J Neurosci* 2010, **30**:4815-4826.

42. Shlens J, Field GD, Gauthier JL, Greschner M, Sher A, Litke AM, Chichilnisky EJ: **The structure of large-scale synchronized firing in primate retina**. *J Neurosci* 2009, **29**:5022-5031.

43. Roudi Y, Nirenberg S, Latham PE: **Pairwise maximum entropy models for studying large biological systems: when they can work and when they can't**. *PLoS Comput Biol* 2009, **5**:e1000380.

44. Roudi Y, Tyrcha J, Hertz J: **Ising model for neural data: model quality and approximate methods for extracting functional connectivity**. *Phys Rev E* 2009, **79**:051915.

45. •• Macke JH, Opper M, Bethge M: **Common input explains higher-order correlations and entropy in a simple model of neural population activity**. *Phys Rev Lett* 2011, **106**:208102.
Shows how beyond-pairwise correlations emerge in models with 'pair-wise' parameterizations from the operation of thresholding.

46. • Yu Y, Yang H, Nakahara H, Sanstos GS, Nikolić D, Plenz D: **Higher-order interactions characterized in cortical activity**. *J Neurosci* 2011, **31**:17514-17526.
Applies the model of Macke *et al.* (2011) to array recordings in cortex, finding that it captures the underlying population structure at pairwise and higher orders of statistical interaction.

47. Truccolo W, Eden UT, Fellows MR, Donoghue JP, Brown EN: **A point process framework for relating neural spiking activity to spiking history, neural ensemble, and extrinsic covariate effects**. *J Neurophysiol* 2005, **93**:1074-1089.

48. • Vidne M, Ahmadian Y, Shlens J, Pillow JW, Kulkarni J, Litke AM, Chichilnisky EJ, Simoncelli E, Paninski L: **Modeling the impact of common noise inputs on the network activity of retinal ganglion cells**. *J Comp Neurosci* 2011 http://dx.doi.org/10.1007/s10827-011-0376-2. [online only at present].
An extension of the generalized linear model methods to allow for common noise inputs. Approaches like this will be important to combine with information theoretic measures to assess the importance of unobserved inputs.

49. Series P, Latham P, Pouget A: **Tuning curve sharpening for orientation selectivity: coding efficiency and the impact of correlations**. *Nat Neurosci* 2004, **7**:1129-1135.

50. Wark B, Lundstrom BN, Fairhall A: **Sensory adaptation**. *Curr Opin Neurobiol* 2007, **17**:423-429.

51. Vergassola M, Villermaux E, Shraiman BI: **'Infotaxis' as a strategy for searching without gradients**. *Nature* 2007, **445**:406-409.

52. Agarwala Chiel H, Thomas P: **Pursuit of food versus pursuit of information in a Markovian perception–action loop model of foraging**. *J Theor Biol* 2012, **304**:235-272.

53. Renninger LW, Verghese P, Coughlan J: **Where to look next? Eye movements reduce local uncertainty**. *J Vis* 2007, **7**:6.

54. Najemnik J, Geisler WS: **Eye movement statistics in humans are consistent with an optimal search strategy**. *J Vis* 2008, **8** 4.1–14.