Model Based Information Calculations
for Neuronal Encoding

Daniel A. Butts[1], Adrien E. Desjardins[2], and Garrett B. Stanley[3]

Abstract— **Information theory provides assumption-free measures of the encoding properties of neurons. However, the amount of data required for such calculations is often prohibitive, and the relationship between the neuron's information rate and its encoding properties is often obscured. We use a common neural model to derive the instantaneous information rate as a function of easily measured parameters of the neuron, such as its receptive field. This method allows for information rates to be calculated using a fraction of the data that other methods require, and explicitly relates properties of the neuron's encoding to its ability to transmit information.**

I. INTRODUCTION

INFORMATION theory provides a framework for quantifying what and how much information about the outside world is being encoded by neurons in the sensory pathways [1]. Methods for estimating rates of information transmission attempt to make as few assumptions about the underlying encoding mechanism as possible, but as a result require an often prohibitory amount of data in order to estimate the underlying probability distributions between stimulus and response. Furthermore, it is often difficult to directly relate information measures to particular functional properties of the encoding process.

Since the early work of Adrian and Barlow, the receptive field (RF) has served as the fundamental building block for our current understanding of the early sensory pathways [2], [3]. More generally, the spatiotemporal receptive field (STRF) is inclusive of the temporal encoding properties in addition to spatial. It is thus desirable to derive information measures that are directly related to these spatial and temporal characteristics that are reflective of the underlying neuronal circuitry.

Here, we develop an approach to measure information transmission in sensory systems that directly utilizes a functional model of the STRF for spiking neurons, which we denote model-based information calculation (MBIC). We derive formulae for instantaneous rates of information transmission based on a two-stage model of neuronal encoding that has been utilized in a number of experimental contexts, and apply it in examples drawn from the early visual pathway.

We find that the MBIC approach allows for the calculation of instantaneous information rates using a fraction of the data compared with direct methods of information calculation [4], suggesting that information calculations might be performed in a variety of experimental paradigms currently inaccessible to such analysis. More significantly, the analyses presented here allow us to directly assess to what extent subtle features of encoding that are captured by traditional information measures are

[1] Department of Neurobiology, Harvard Medical School, 220 Longwood Ave., Boston, MA 02115. Email: dbutts@hms.harvard.edu

[2] Biophysics Program, Harvard University, 240 Longwood Avenue, Boston, MA 02115. Email: adesjard@mit.edu

[3] Division of Engineering & Applied Sciences, Harvard University, 29 Oxford Street, Cambridge, MA 02138. Email: gstanley@deas.harvard.edu

also captured by the encoding models and therefore by the corresponding MBIC.

II. MODELS OF ENCODING

Since the earliest experimental explorations of the visual system, the spatiotemporal receptive field (STRF) has served as a useful concept for characterizing how the system responds to stimuli [5], [6]. The STRF describes both the functional nature of the spatial integration across the receptive field and the dynamics of the temporal integration. The STRF, however, does not fully define the input-output relationship of a neuron. A neuron's output consists of spikes, which, over many trials can be pooled to estimate a firing rate as a result of the stimulus. Due to the fact that firing rates must be non-negative, and that these firing rates result in a probabilistic response from trial to trial, a full description of the encoding of a neuron must incorporate two further steps: a (1) non-linearity that maps the output of the STRF onto non-negative values, and (2) a probabilistic generator of spike times. The block diagram for a general nonlinear encoding model for spiking visual system neurons is shown in Figure 1.
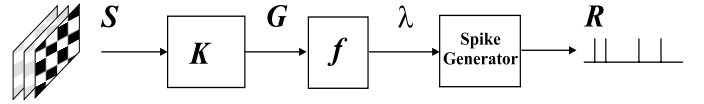

Fig. 1. Block diagram for the general nonlinear encoding model.

This structure can be thought of as a two-stage process: the first stage of processing being a spatiotemporal filter, the second stage of processing being the spike generation mechanism. The spatiotemporal visual stimulus S is acted upon by a functional K to produce a scalar generating signal $G = K(S)$. By means of the filter K, the temporal history and spatial extent of the stimulus are mapped onto the generating signal G, as illustrated for a one-dimensional stimulus in Figure 2.
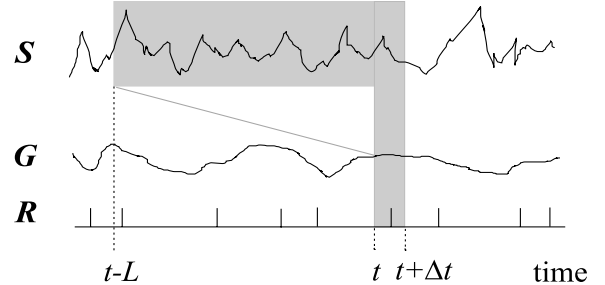

Fig. 2. Projection of the stimulus onto the generating signal and the spike train.

The generating signal is subsequently nonlinearly scaled by the function $f(\cdot)$ to produce the intensity λ, which is the time-

varying rate of an inhomogeneous Poisson process. The function $f(\cdot)$ captures the static nonlinear properties of the stimulus/response relationship, namely those of rectification to produce a non-negative spiking rate and saturation that represents an upper bound on the firing rate. Finally, we define R to represent the number of spikes in a small time interval $[t, t + \Delta t]$.

III. MODEL BASED INFORMATION CALCULATIONS

The mutual information between the stimulus S and response R ensembles is defined as [7]:

$$I[S;R] = \sum_s p_S(s) \sum_r p_{R|S}(r|s) \log_2 \frac{p_{R|S}(r|s)}{p_R(r)}$$

Here, we limit our calculation to encompass instantaneous information rates only, meaning the ensemble of possible responses encompasses only the number of spikes in individual bins of width Δt. In the limit as $\Delta t \to 0$, the information rate $I[S,R]/\Delta t$ is only non-zero for the single-spike response. As a result, the imposed model structure becomes:

$$\frac{dI[S;R]}{dt} \approx \frac{1}{\Delta t} \sum_s p_S(s) Pr\{\text{spike}|s\} \log_2 \frac{Pr\{\text{spike}|s\}}{Pr\{\text{spike}\}}$$

The probability of a spike given a particular stimulus in the context of our model structure is given by $Pr\{\text{spike}|s\} = f(K(s))\Delta t = f(g)\Delta t$, giving

$$\frac{dI[S;R]}{dt} = \sum_s p_S(s) f(K(s)) \log_2 \frac{f(K(s))}{\sum_{s'} p_S(s') f(K(s'))} \quad (1)$$

The calculation described in (1) involves integrating over the stimulus space, which for spatiotemporal stimuli, can be prohibitive. However, since the relationship between S and G is purely deterministic, we can express the information rate as:

$$\frac{dI[S;R]}{dt} = \sum_g p_G(g) f(g) \log_2 \frac{f(g)}{\sum_{g'} p_G(g') f(g')}$$

If the stimulus S is zero mean with variance σ_s^2, and independent across space and time, then the corresponding generating signal G is approximately Gaussian with zero mean and variance $\sigma_g^2 = ||K||^2 \sigma_s^2$.

IV. MBIC OF VISUAL NEURONS

We apply our MBIC analysis to the early visual system, where extensive information calculations have already been performed [8], [9], [10]. Specifically, we use data generated by simple models of LGN neurons that employ the coding characteristics shown in Figure 1. Using a realistic linear kernel K and nonlinear function f, we use a simulation to generate spike trains in the context of a full-field flicker stimulus (with a temporal component, but no spatial component), and calculate the instantaneous information rate using both direct methods and the MBIC.

A direct calculation of information in this context [4], [9] requires enough data to properly estimate the spike train entropy $H[R]$, but more importantly, requires multiple repeats of the same stimulus sequence to estimate the noise entropy $H[R|S]$.

In Figure 3, we show how the estimate of the instantaneous information rate converges to its final value of 39.3 bits/sec as a function of the number of one-minute stimulus repeats.
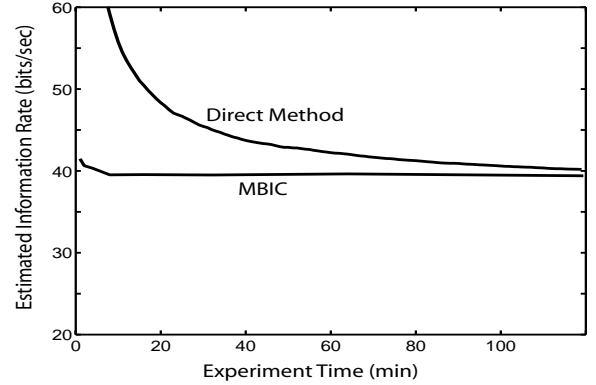

Fig. 3. Convergence of the information rate using the direct method and the MBIC method.

To apply an MBIC to this data, we use previously established methods of reverse correlation to estimate the linear kernel K and the non-linear function f of the neuron [11]. Since these functions are directly incorporated into the MBIC of eq. 1, the dependence of this calculation on the number of repeats reduces to the ability to estimate these linear parameters. Not surprisingly, we find that the MBIC has an initial error after a single one-minute trial of only 2 bits/sec, and converges to within a bit/sec of its final value with four stimulus repeats.

The ability of the MBIC to accurately estimate the information rate of this neuron with little data strongly depends on the degree to which the neuron under study conforms to the model employed in the MBIC, as was explicitly the case here. However, such a model for visual neurons is widely applied, and features of real neurons that are not captured in this particular model might be incorporated in future applications.

In this example we have demonstrated that the calculation of instantaneous information rates requires a relatively small amount of data, and thus might be applied to a variety of systems and experimental paradigms that are currently inaccessible to information analyses. Furthermore, the explicit dependence of the MBIC on the receptive field properties gives insight into how the parameters of the receptive field of neurons relative to its information coding capacities.

REFERENCES

[1] F. Rieke, D. Warland, R. van Stevenfck, and W. Bialek, *Spikes: Exploring the Neural Code*, MIT Press, Cambridge, MA, 1997.

[2] E. D. Adrian, "The impulses produced by sensory nerve endings. part i.," *J. Physiol (Lond)*, vol. 61, pp. 49–72, 1926.

[3] H. B. Barlow, "Possible principles underlying the transformations of sensory messages," in *Sensory Communication*, W. A. Rosenblith, Ed., pp. 217–234. MIT Press, 1961.

[4] S. P. Strong, R. Koberle, R. R. de Ruyter van Stevenfck, and W. Bialek, "Entropy and information in neural spike trains," *Physical Review Letters*, vol. 80, pp. 197–200, 1998.

[5] D. J. Pinto, J. C. Brumberg, D. J. Simons, and G. B. Ermentrout, "A quantitative population model of whisker barrels: re-examining the wilson-cowan equations," *J. Comput. Neurosci*, vol. 3, pp. 247–264, 1996.

[6] H. T. Kyriazi and D. J. Simons, "Thalamocortical response transformations in simulated whisker barrels," *J. Neurosci.*, vol. 13, pp. 1601–1615, 1993.

[7] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, Wiley-Interscienc, 1991.

[8] R. C. Liu, S. Tzonev, S. Rebrik, and K. D. Miller, "Variability and information in a neural code of the cat lateral geniculate nucleus," *J. Neurophysiol.*, vol. 86, pp. 2789–2806, 2001.

[9] P. Reinagel and R. C. Reid, "Temporal coding of visual information in the thalamus," *J. Neurosci.*, vol. 20, pp. 5392–5400, 2000.

[10] M. Meister and M.J. Berry, "The neural code of the retina," *Neuron*, vol. 22, pp. 435–450, 1999.

[11] P. Dayan and L. F. Abbott, *Theoretical Neuroscience*, MIT Press, Cambridge, 2001.