INSTITUTE OF PHYSICS PUBLISHING NETWORK: COMPUTATION IN NEURAL SYSTEMS

Network: Comput. Neural Syst. **12** (2001) 199–213 www.iop.org/Journals/ne PII: S0954-898X(01)23295-3

A simple white noise analysis of neuronal light responses

E J Chichilnisky

Systems Neurobiology, The Salk Institute, 10010 North Torrey Pines Road, La Jolla, CA 92037-1099, USA

E-mail: ej@salk.edu

Received 20 November 2000

Abstract

A white noise technique is presented for estimating the response properties of spiking visual system neurons. The technique is simple, robust, efficient and well suited to simultaneous recordings from multiple neurons. It provides a complete and easily interpretable model of light responses even for neurons that display a common form of response nonlinearity that precludes classical linear systems analysis. A theoretical justification of the technique is presented that relies only on elementary linear algebra and statistics. Implementation is described with examples. The technique and the underlying model of neural responses are validated using recordings from retinal ganglion cells, and in principle are applicable to other neurons. Advantages and disadvantages of the technique relative to classical approaches are discussed.

1. Introduction

This paper describes a technique for quantifying the spatial, temporal and spectral response properties of spiking visual system neurons. Classically this has been accomplished by presenting pulsed, modulated or drifting stimuli such as spots, bars or gratings, measuring spike rate for a period during and after stimulation and assuming linearity to construct a model of neural response. The method presented here, based on *white noise* stimulation, has several advantages over this conventional paradigm. The stochastic, highly interleaved stimulus spans a wide range of visual inputs, is relatively robust to fluctuations in responsivity, avoids adaptation to strong or prolonged stimuli and is well suited to simultaneous measurements from multiple neurons. The analysis yields a full characterization of a neuron's spatial, temporal and chromatic sensitivity even when the response exhibits a common form of nonlinearity (see below) that precludes classical linear systems analysis.

White noise stimulation combined with Wiener kernel analysis can in principle be used to characterize neurons with arbitrarily complex nonlinear response properties (see Sakai 1992, Rieke *et al* 1997). However the practical benefits of this approach are limited because of the large data sets required and the difficulty of interpreting high-order kernels (but see Marmarelis

0954-898X/01/020199+15$30.00

and Naka (1972), Citron and Emerson (1983), Korenberg *et al* (1989); see Rieke *et al* (1997), Sakai (1992) for a fuller discussion). For this reason, white noise methods have often been used to probe stimulus selectivity qualitatively without an explicit model of neural response (e.g. Meister *et al* 1995, Brown *et al* 2000), or with an assumption of response linearity that greatly simplifies analysis (e.g. Reid and Alonso 1995, DeAngelis *et al* 1993). However many neurons display significant response nonlinearities, such as spike threshold and response saturation, that preclude linear analysis.

This paper focuses on a relatively simple form of white noise analysis that yields a quantitative model of neural response in spite of such nonlinearities (Marmarelis and Naka 1972, Korenberg and Hunter 1986, Hunter and Korenberg 1986, Korenberg *et al* 1989, Sakai 1992). It is assumed that the neuron's spike rate at any instant depends only on the value at the same instant of a *generator signal* which is a linear combination of visual inputs over space and recent time (e.g. a sum of synaptic currents proportional to local contrast). However, the instantaneous relationship between the generator signal and spike rate can be linear or nonlinear (e.g. spike generation). Under these conditions, the analysis technique presented here is useful (e.g. Emerson *et al* 1992, Sakai *et al* 1995, Anzai *et al* 1999, Chichilnisky and Baylor 1999) because it produces a full, quantitative model of neural response that is almost as simple as a strictly linear model but accounts for the response nonlinearity. This can be particularly useful in characterizing subthreshold responses of neurons with low maintained discharge (e.g. DeAngelis *et al* 1995) to explore subtle aspects of light response, or in analysing spatio-temporal sensitivity and detailed properties of visual inputs without confounds due to response nonlinearities (e.g. Chander and Chichilnisky 1999, Chichilnisky and Baylor 1999).

The purpose of this paper is to make this analysis technique clear and accessible to neuroscientists in the form of a self-contained guide. The procedure for building a complete quantitative input–output model of a visual system neuron using white noise stimulation is justified theoretically and described with examples. An empirical test of the model is formulated and applied to experimentally measured responses of retinal ganglion cells.

2. Definitions and assumptions

2.1. Stimulus and response

The period in which stimuli are presented and spikes are recorded is divided into bins of size Δt, typically a few milliseconds. The neural response f_t is defined as the number of spikes observed in time bin t. A successful model will predict the average number of spikes per time bin observed after the presentation of any stimulus.

Each stimulus is specified by a vector of dimension k. The entries of the stimulus vector $\boldsymbol{s}_t$ indicate intensity as a function of space and time immediately preceding t. The duration of each stimulus vector is assumed to exceed the *memory* of the neuron, that is, the period over which a stimulus can affect the response (empirically determined; see section 3.1). Thus the response f_t at time t depends only on $\boldsymbol{s}_t$.

Example. Suppose the stimulus is a spatially uniform grey field with variable intensity over time. The stimulus vector $\boldsymbol{s}_t$ indicates the intensity in the k time bins preceding t.

Example. Suppose the stimulus is generated using a colour computer display refreshing at 75 Hz ($\Delta t = \frac{1}{75}$ s), and the memory of the neuron is 200 ms ($15\Delta t$). Then each stimulus vector $\boldsymbol{s}_t$ denotes the red, green and blue phosphor intensity at each spatial location in each of 15 time bins preceding t. For a display with just 256 distinct pixels, $\boldsymbol{s}_t$ has dimension $k = 256 \times 3 \times 15 = 11\,520$.

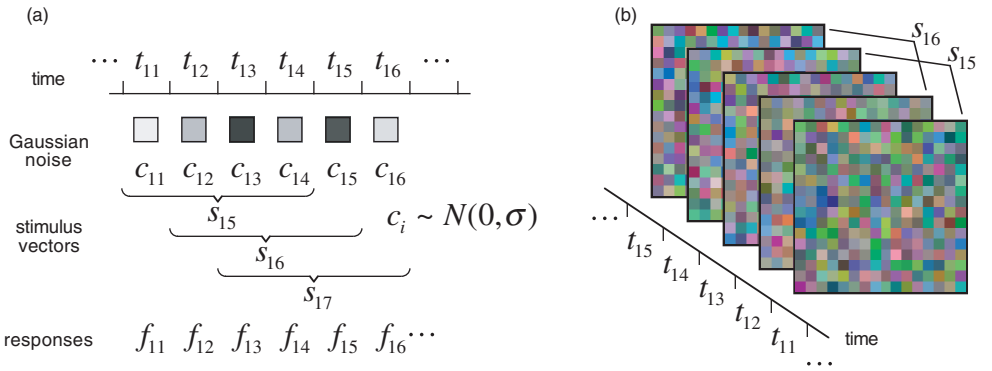

Figure 1. (*a*) Section of a spatially uniform, monochromatic Gaussian white noise sequence for a neuron with a memory of four time units. The contrast values in the stimulus vector s_{15}, for example, are presented in time bins 11–14. The response f_{15} to this stimulus vector is the number of spikes in time bin 15. Note that distinct stimulus vectors such as s_{15}, s_{16} and s_{17} overlap in time. (*b*) Gaussian white noise sequence with spatial, temporal and chromatic modulation for a neuron with a memory of four time units.

2.2. Stochastic stimuli: Gaussian white noise sequences

Each stimulus vector can be represented as a point in a k-dimensional *stimulus space* S; that is, a Cartesian coordinate system. Stimulus vectors are assumed to be drawn randomly from S. The probability of drawing a particular stimulus vector $\boldsymbol{s}$ is given by $P(\boldsymbol{s})$. This probability distribution is assumed to be radially symmetric about the origin in stimulus space. That is, any two stimulus vectors $\boldsymbol{s}, \boldsymbol{s}^* \in S$ with equal vector length have equal probability of being drawn:

$$|\boldsymbol{s}| = |\boldsymbol{s}^*| \Rightarrow P(\boldsymbol{s}) = P(\boldsymbol{s}^*). \tag{1}$$

Radial symmetry about the origin implies that some stimulus vectors must contain negative entries. A standard notation is to specify the *contrast*, or deviation from mean intensity, in the entries of stimulus vectors.

An efficient way to generate a set of stimulus vectors over time satisfying the above requirements is to use a *Gaussian white noise* sequence. Each stimulus vector $\boldsymbol{s}_t$ is created by filling the entries of a k-dimensional vector with independent draws from a Gaussian distribution. Thus the ensemble of stimulus vectors forms a multi-dimensional Gaussian distribution satisfying the radial symmetry assumption.

Example. Consider the case of a spatially uniform grey field of variable intensity. Suppose the memory of the neuron is four time units, and choose $k = 4$. At each time bin, a new contrast is selected and presented according to a random draw from a Gaussian distribution, as illustrated in figure 1(*a*).

Example. A more general Gaussian white noise sequence consists of a random, independent sequence of intensities for each spatial location, time bin and colour. This stimulus looks like a flickering coloured checkerboard pattern with no spatial, temporal or chromatic structure. Sample frames from such a stimulus are shown in figure 1(*b*).

Stimuli are presented sequentially without gaps in time, so each stimulus vector in a Gaussian white noise sequence overlaps in time with stimulus vectors that immediately precede and follow it (see figure 1). Thus stimulus vectors are not statistically independent.

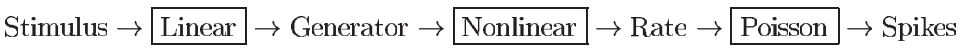

Figure 2. Model for neural response. First, a linear operation (inner product with $\boldsymbol{w}$) on the recent stimulus $\boldsymbol{s}$ produces a real-valued generator signal $\boldsymbol{w} \cdot \boldsymbol{s}$. Second, an instantaneous nonlinear function of the generator signal produces the spike rate $R(\boldsymbol{s})$. Third, a Poisson process determines the spike count f.

2.3. Model of neural response

The neural response to be modelled is the average value of the spike count f in the time bin immediately following a stimulus $\boldsymbol{s}$, or the *expected response* to the stimulus:

$$R(\boldsymbol{s}) = \langle f \mid \boldsymbol{s} \rangle = \sum_f f P(f \mid \boldsymbol{s}). \tag{2}$$

Here the notation $\langle\rangle$ denotes the expected value or average across repeated experiments. The function $R()$ can take on any non-negative real value (as opposed to the spike count, f, which is integer valued).

The goal is to estimate $R()$. To make this tractable a simplifying assumption is imposed, namely, that $R()$ is a *static nonlinear* function of a real-valued linear function of the stimulus:

$$R(\boldsymbol{s}) = N(\boldsymbol{w} \cdot \boldsymbol{s}). \tag{3}$$

Here $\boldsymbol{w}$ is a fixed vector of the same dimension as $\boldsymbol{s}$, $\boldsymbol{w} \cdot \boldsymbol{s}$ represents the dot (scalar, inner) product of vectors, and $N()$ is an arbitrary real-valued function of one variable. In the model of equation (3), stimulus intensities over space and time are weighted by the entries of the fixed vector $\boldsymbol{w}$, which defines the neuron's stimulus selectivity, and summed to give a single number $\boldsymbol{w} \cdot \boldsymbol{s}$, the *generator signal*. Most standard models of visual responses have just this linear architecture (e.g. Enroth-Cugell and Pinto 1970, Movshon *et al* 1978), but in equation (3) the generator signal controls firing rate via the (generally nonlinear) function $N()$. This architecture is useful because it allows for typical nonlinearities in neurons, such as spike threshold and response saturation, while remaining mathematically tractable: as will be shown below, $\boldsymbol{w}$ and $N()$ can be estimated easily and efficiently from neural responses. The model is depicted schematically in figure 2.

Note that equation (3) describes the dependence of the response on the stimulus, not on the recent response history. It is assumed that there is no dependence on the response history *per se*, that is, spikes are generated by a Poisson process[1] with a rate parameter equal to the expected response $R(\boldsymbol{s})$.

Example. Suppose that the total post-synaptic current in a neuron depends linearly on the stimulus, but that spike probability depends nonlinearly on current because of the biophysical properties of the ion channels that control spike generation. The static nonlinear model subsumes this case, a strictly linear model does not.

[1] While Poisson spike generation in real neurons is impossible because of refractoriness caused by inactivation of sodium channels, if the refractory period is short compared to Δt, Poisson spike generation may be a reasonable approximation. For non-Poisson spiking (such as gamma or integrate-and-fire processes) the average spike rate is an incomplete but still useful measure of response, and can be predicted using the same model (Victor, personal communication). For simplicity, only the Poisson case is considered here. See Rieke *et al* (1997) for a fuller discussion of the statistics of neural responses.

3. Estimating the stimulus–response relationship

3.1. Spike-triggered average and relation to R()

The first step in characterizing $R()$, which defines the relation between the stimulus and the expected number of spikes, is achieved by examining the *spike-triggered average* stimulus or *STA*. This is defined as the average stimulus preceding a spike in the cell, i.e. the sum of the stimuli that preceded each spike divided by the total number of spikes:

$$\boldsymbol{a} = \frac{\sum_{t=1}^{T} \boldsymbol{s}_t f_t}{\sum_{t=1}^{T} f_t}. \tag{4}$$

Here T represents the length of the recording period. The STA, $\boldsymbol{a}$, is a vector with the same dimension as each stimulus vector. It indicates what stimulus, on average, caused the cell to spike. More importantly the STA is related in a simple way to $\boldsymbol{w}$ (Hunter and Korenberg 1986, Korenberg *et al* 1989). To establish this relation, first divide the numerator and denominator in equation (4) by T:

$$\boldsymbol{a} = \frac{\frac{1}{T}\sum_{t=1}^{T} \boldsymbol{s}_t f_t}{\frac{1}{T}\sum_{t=1}^{T} f_t}. \tag{5}$$

Assume recording happens for a long time, i.e. $T \to \infty$. The denominator in equation (5) is the total number of spikes divided by the recording time—the average firing rate. As $T \to \infty$ it is assumed that this value approaches a nonzero limit $\langle f \rangle$. Similarly it is assumed that the numerator approaches a limit $\langle \boldsymbol{s} f \rangle$. This expectation can be expressed equivalently as the sum over all stimulus–response pairs weighted by the probability of observing that stimulus–response pair:

$$\langle \boldsymbol{s} f \rangle = \sum_{\boldsymbol{s}} \sum_{f} \boldsymbol{s} f P(\boldsymbol{s} \,\&\, f).$$

This can be simplified using the identity $P(\boldsymbol{s} \,\&\, f) = P(\boldsymbol{s})P(f \mid \boldsymbol{s})$, re-arranging terms and substituting from equation (2):

$$\begin{aligned}\langle \boldsymbol{s} f \rangle &= \sum_{\boldsymbol{s}} \sum_{f} \boldsymbol{s} f P(\boldsymbol{s}) P(f \mid \boldsymbol{s}) \\ &= \sum_{\boldsymbol{s}} \boldsymbol{s} P(\boldsymbol{s}) \sum_{f} f P(f \mid \boldsymbol{s}) \\ &= \sum_{\boldsymbol{s}} \boldsymbol{s} P(\boldsymbol{s}) R(\boldsymbol{s}).\end{aligned}$$

Hence for large values of T,

$$\boldsymbol{a} = \frac{1}{\langle f \rangle} \sum_{\boldsymbol{s}} \boldsymbol{s} P(\boldsymbol{s}) R(\boldsymbol{s}). \tag{6}$$

Thus the STA approaches a sum of stimulus vectors, each weighted by its probability of being drawn and the average response it induces, normalized by the average firing rate, as might be expected intuitively from equation (4).

A key simplification of equation (6) (Meister, personal communication) is achieved using the two assumptions above, namely the static nonlinear form of the response (equation (3)) and the radial symmetry of the stimulus distribution (equation (1)). For each stimulus $\boldsymbol{s} \in S$, radial symmetry implies that there is another stimulus $\boldsymbol{s}^* \in S$ in a location symmetric about the linear weighting vector $\boldsymbol{w}$ (see figure 3) with equal probability of being drawn: $P(\boldsymbol{s}) = P(\boldsymbol{s}^*)$.

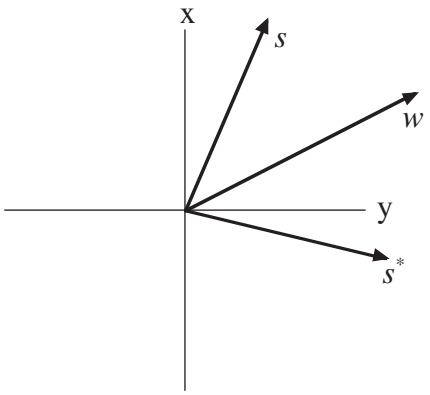

Figure 3. Symmetry of the stimulus distribution. The coordinate axes represent a two-dimensional stimulus space S, for example contrast at two spatial locations or two time points prior to the response. By assumption (equation (1)), for each stimulus $\boldsymbol{s} \in S$ there is another stimulus $\boldsymbol{s}^*$ symmetric about $\boldsymbol{w}$ with equal probability of being drawn.

Substituting $N(\boldsymbol{w} \cdot \boldsymbol{s})$ for $R(\boldsymbol{s})$ from equation (3), and grouping the terms in the sum of equation (6) into pairs,

$$\boldsymbol{a} = \frac{1}{\langle f \rangle} \sum_{\boldsymbol{s},\boldsymbol{s}^*} [\boldsymbol{s} P(\boldsymbol{s}) N(\boldsymbol{w} \cdot \boldsymbol{s}) + \boldsymbol{s}^* P(\boldsymbol{s}^*) N(\boldsymbol{w} \cdot \boldsymbol{s}^*)].$$

The pairs can be combined because $P(\boldsymbol{s}) = P(\boldsymbol{s}^*)$, and because $\boldsymbol{s}$ and $\boldsymbol{s}^*$ are symmetric about $\boldsymbol{w}$ so $\boldsymbol{w} \cdot \boldsymbol{s} = \boldsymbol{w} \cdot \boldsymbol{s}^*$:

$$\boldsymbol{a} = \frac{1}{\langle f \rangle} \sum_{\boldsymbol{s},\boldsymbol{s}^*} (\boldsymbol{s} + \boldsymbol{s}^*) N(\boldsymbol{w} \cdot \boldsymbol{s}) P(\boldsymbol{s}).$$

Finally, since $\boldsymbol{s}$ and $\boldsymbol{s}^*$ are symmetric about $\boldsymbol{w}$, their sum is proportional to $\boldsymbol{w}$. The remaining terms in the sum are all scalar quantities. Therefore each term in the sum is proportional to $\boldsymbol{w}$, and so is the sum:

$$\boldsymbol{a} \propto \boldsymbol{w}.$$

This is the main result: the linear part of the model, $\boldsymbol{w}$, is directly proportional to the STA, which is easily estimated directly from the spike train (Hunter and Korenberg 1986, Korenberg *et al* 1989).

Why is this useful? The linear part of the model indicates how the neuron weights different spatial, temporal and spectral components of the stimulus. For example, the spatial structure of $\boldsymbol{w}$ describes the neuron's spatial receptive field. The temporal structure of $\boldsymbol{w}$ describes the impulse response of the underlying linear summation. The memory of the neuron can be assessed by examining the duration of the impulse response, i.e. the time offset from the spike at which the STA converges to zero. Finally, the chromatic structure of $\boldsymbol{w}$ reflects the weighting of inputs from different spectral classes of photoreceptor. Thus the linear weighting vector $\boldsymbol{w}$ contains essential information about how the neuron integrates visual inputs.

Example. The linear component of visual responses was estimated for macaque monkey retinal ganglion cells recorded *in vitro* at low photopic light levels (experimental methods given by Chichilnisky and Baylor (1999)). The intensities of the red, green and blue phosphors of a colour computer display were selected randomly and independently from a Gaussian distribution at each spatial location and each point in time (see figure 1(*b*)). For each cell

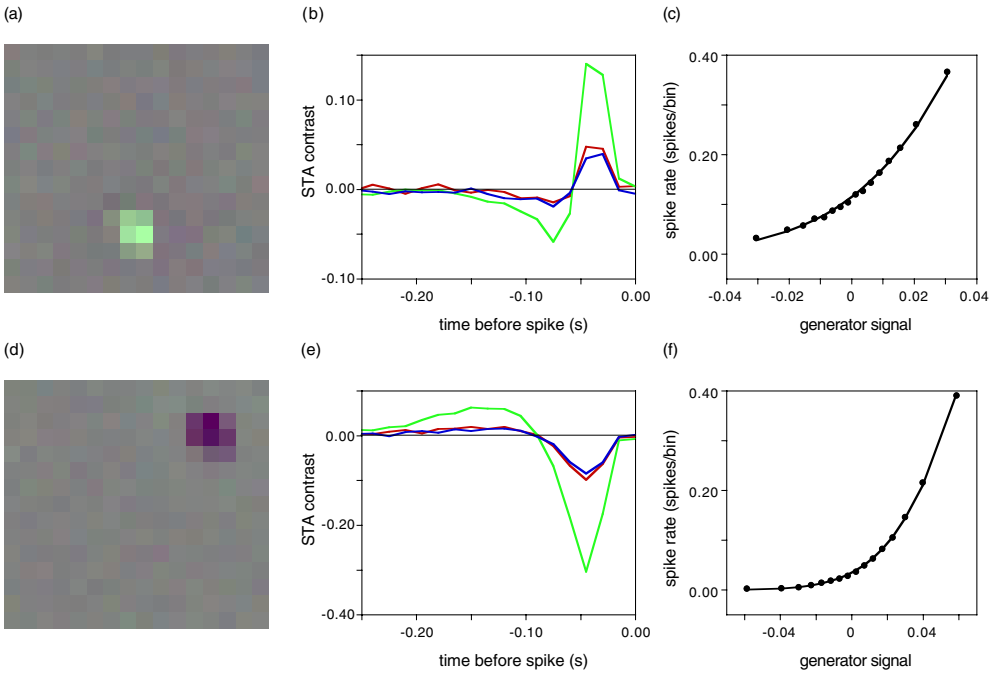

Figure 4. (*a*) Spatial profile of the STA from a macaque ON-centre retinal ganglion cell 45 ms before a spike. Each square stimulus element was 60 μm on a side. (*b*) Time course of the STA at the centre of the receptive field. Red, green and blue phosphors are shown in corresponding colours. The mean of the Gaussian intensity distribution for each gun was 0.5 and the standard deviation was 0.16. Contrast refers to the difference from the mean intensity. (*c*) Static nonlinearity $N(g)$ estimated from this cell. The smooth curve is a parametrized form of the cumulative normal function fitted to the data. See the text for details. (*d*) Same as (*a*), for a simultaneously recorded OFF cell. (*e*) Same as (*b*), for the cell in (*d*). (*f*) Same as (*c*), for the cell in (*d*). $\Delta t = 15$ ms.

the STA was computed. The STA was a movie, containing contrast values (deviations from mean intensity of 0.5) for each of the three guns at each spatial location in a sequence of time bins preceding a spike. Figure 4(*a*) shows a frame of this movie 45 ms prior to a spike. This ON-centre cell spiked, on average, after a brightening in the region of the display where the STA had entries significantly greater (brighter) than zero (grey)—the centre of the neuron's spatial receptive field. A weak antagonistic surround (darker than grey) encircles the centre. The STA tapered to near zero far from the centre indicating that stimulus perturbations at these locations were not correlated with spikes. Figure 4(*b*) shows the red, green and blue phosphor intensities in the STA summed over the centre of the receptive field plotted as a function of time prior to the spike. The tapering to zero at about 200 ms prior to the spike indicates that stimulus perturbations before this time were not correlated with spikes and thus defines the memory of the cell. Since the STA is proportional to the linear weighting of visual inputs $\boldsymbol{w}$, these time courses represent the time-reversed impulse responses to red, green and blue phosphor modulation in the centre of the receptive field. The biphasic form of these impulse responses indicates temporal bandpass filtering. The relative amplitude of the red, green and blue traces reflects the relative strength of inputs to the cell from the three types of cone photoreceptor. The STA time course in the surround was weaker and of opposite polarity (not shown). Similar results from a simultaneously recorded OFF-centre ganglion cell are shown in figures 4(*d*) and (*e*).

3.2. Estimating the nonlinearity

Since the STA is proportional to $\boldsymbol{w}$, completing the model (equation (3)) involves only estimating the constant of proportionality and the (generally nonlinear) function $N()$.

The constant of proportionality is indeterminate because the units of the inferred generator signal are arbitrary and are confounded with $N()$. For example, in equation (3), $\boldsymbol{w}$ could be doubled and the input sensitivity of $N()$ halved with no effect on the predictions of the model, so without loss of generality one may assume that the constant of proportionality is unity and $\boldsymbol{a} = \boldsymbol{w}$.

Estimation of $N()$ is simple because it is a real-valued function of one variable. Since $\boldsymbol{a} = \boldsymbol{w}$, an estimate for the generator signal at time t is $g_t = \boldsymbol{a} \cdot \boldsymbol{s}_t$. Since the spike count f_t at time t is known one might visualize $N()$ by plotting the spike count f_t as a function of g_t for each time point $1 \leqslant t \leqslant T$. Unfortunately f_t is discrete, noisy and typically small. Hence this plot would contain many superimposed and indistinguishable points with roughly the same range of f spanned for each value of g, and so would be difficult to visualize.

A better approach is to examine the *average* spike count in time bins with nearly equal generator signals. Since many different stimuli were presented over the recording interval, roughly the same generator signal will be observed at many times in the course of stimulation. (The range of values taken by the generator signal during the recording period depends on $\boldsymbol{w}$ and on the standard deviation σ of the Gaussian distribution used to generate the stimulus sequence.) Suppose H is a collection of time bins that all have similar generator signal values, $\boldsymbol{a} \cdot \boldsymbol{s}_t \approx \bar{g}$ for all $t \in H$. Averaging the spike counts observed in these time bins yields an estimate of the expected spike count given a value $\bar{g}$ of the generator signal, $\langle f \,|\, \boldsymbol{w} \cdot \boldsymbol{s} \approx \bar{g} \rangle$, but from equations (2) and (3) this is $N(\bar{g})$. To estimate the entire function $N()$ this process is repeated for many different values of the generator signal g achieved at some point during the stimulation period. (In practice, to avoid estimation biases it is sometimes important to use separate periods of recording to estimate $\boldsymbol{w}$ and the function $N()$.)

Examples are shown for two macaque retinal ganglion cells in figures 4(c) and (f). In these examples $N()$ is an accelerating function of its inputs, a common finding. With stronger stimuli (larger σ), $N()$ often begins to saturate and assumes a sigmoidal shape (see figure 5(b)).

While the points in figure 4(c) provide estimates of $N()$ for particular generator signal values, generating predictions from the model typically requires a continuous-valued approximation to $N()$. For retinal ganglion cells, a parametrized form of the *cumulative normal* density (the indefinite integral of the normal distribution) usually provides an excellent fit. Specifically, if $C()$ is the cumulative normal density, $N(x) \approx \alpha C(\beta x + \gamma)$ where the parameters α, β and γ are selected to fit the data in the graph of $N()$ with least squared error. An example is shown by the smooth curves in figures 4(c) and (f). The parameters have the following interpretation:

- α—maximum firing rate of the neuron;
- β—sensitivity of the nonlinearity to the generator signal;
- γ—maintained drive to the cell that determines the spike rate in the absence of net visual stimulation (e.g. $\boldsymbol{s} = 0$). Negative values of γ indicate a drive that must be overcome to produce spikes, as in the case of a spike threshold.

While the cumulative normal density may not describe the form of $N()$ for other cell types, a similar procedure can be used to create a smooth functional approximation with interpretable parameters. The characterization of $N()$ completes the model (equation (3)) of firing rate as a function of the stimulus.

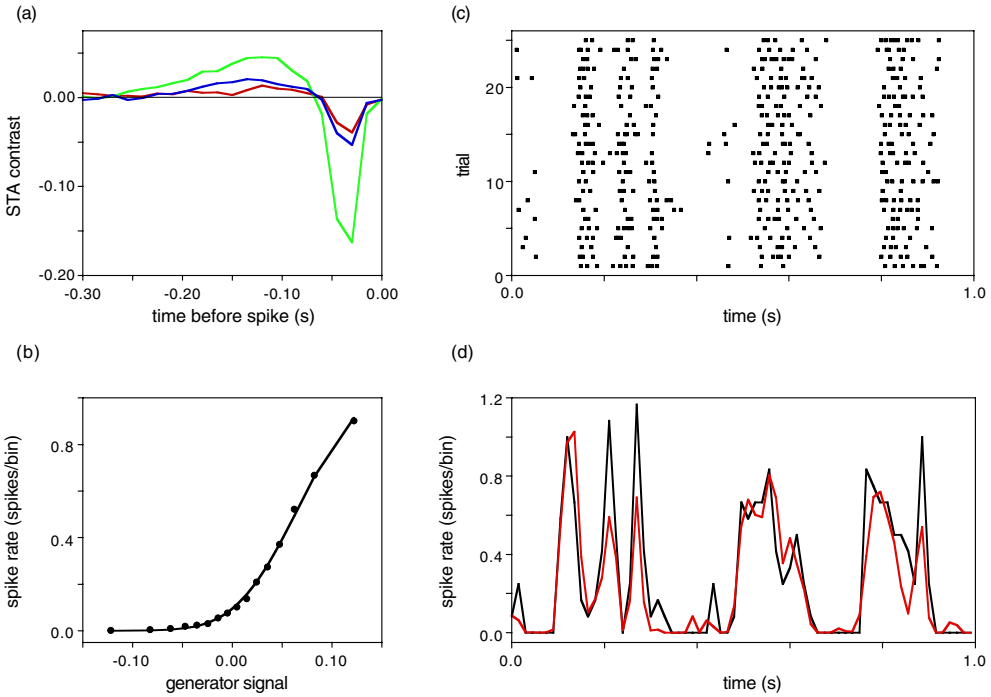

Figure 5. (*a*) Time course of STA from a single macaque retinal ganglion cell stimulated with a spatially uniform Gaussian white noise sequence. The negative polarity of the main lobe indicates this is an OFF cell. (*b*) Static nonlinear function $N()$ estimated from the same cell. (*c*) Rasters of responses to 25 repeated presentations of a 20 s Gaussian white noise sequence different from that used in (*a*). 1 s of the recording is shown. (*d*) Average spike rate as a function of time estimated from spikes shown in (*c*) (black trace) and model predictions (red trace). $\Delta t = 15$ ms.

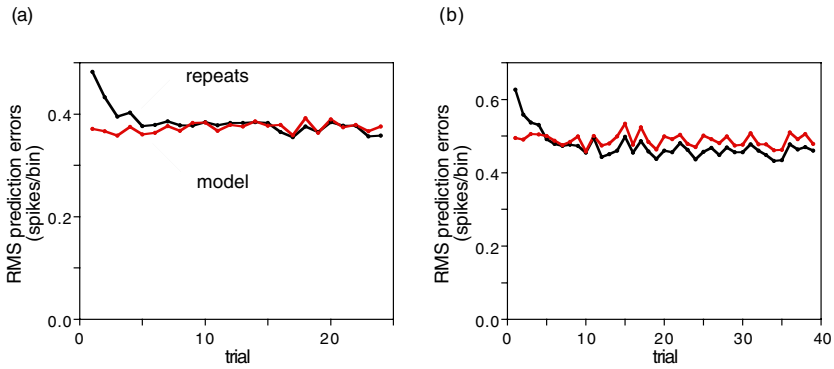

Figure 6. (*a*) RMS error of predictions of spike counts over time as a function of number of repeated trials for the macaque retinal ganglion cell shown in figure 5. The red trace shows the RMS difference between observed spike counts and model predictions. The black trace shows the RMS difference between observed spike counts and predictions from repeated trials. (*b*) The same plot for a second macaque ganglion cell. $\Delta t = 15$ ms.

4. Test of the model

The cells in figure 4 display significant response nonlinearities: if their responses were linear, the points in figures 4(*c*) and (*f*) would fall on a straight line. The great majority of several

hundred macaque and salamander retinal ganglion cells from several dozen retinas analysed this way display similar nonlinearities (data not shown), so using classical linear methods to characterize their light responses would fail. How well does the static nonlinear model describe their light responses?

What follows is a test of the accuracy of the model. Predictions of the model were compared to the firing rate measured directly by repeatedly presenting the same stimulus and averaging the number of spikes recorded in each time bin. The latter is the maximum-likelihood, unbiased estimate of firing rate—a perfect model of firing rate would match it exactly.

The STA and static nonlinearity for one macaque monkey OFF retinal ganglion cell obtained with spatially uniform white noise stimulation are shown in figures 5(*a*) and (*b*) (experimental procedures are given by Chichilnisky and Baylor (1999)). The model based on these estimates of $\boldsymbol{w}$ and $N()$ was then used to predict firing rate over time in response to a different spatially uniform Gaussian white noise sequence presented repeatedly. Rasters of spikes recorded in repeated trials are shown in figure 5(*c*). The vertical structure in the plot results from repeated presentation of the same stimulus. The firing rate as a function of time obtained by averaging spike counts across trials is shown by the black trace in figure 5(*d*). The firing rate as a function of time predicted by the model is shown by the red trace. Qualitatively, the model captures the time course of firing rate modulations, though significant deviations are sometimes observed.

How well does the model predict light responses in individual trials? The root mean square (RMS) difference between the spike rate predicted by the model and the observed spike counts in the tenth trial of the repeat sequence was 0.384 spikes per bin for the cell in figure 5. In comparison, the RMS difference between the spike rate over time obtained by averaging counts from the nine previous repeat trails and the spike counts in the tenth trial was 0.382 spikes per bin. That is, the model prediction of spike counts over time was nearly as accurate as the prediction obtained by averaging the results from prior repeats of the same experiment.

The robustness of this finding to the number of repeat trials used is examined in figure 6(*a*), for the same cell as figure 5. The red curve shows the RMS difference between the model prediction of firing rate over time and the observed spike counts over time on the nth trial. This curve is fairly flat, indicating stable recordings and light responses. The black curve shows the RMS difference between the spike rate over time estimated by averaging the first $n-1$ trials and the observed spike counts over time on the nth trial. Since estimates of firing rate obtained from a finite number of repeated trials are variable and quantized, more trials provide better predictions up to a point and thus the black curve initially declines. Since spike counts in each trial are variable and quantized, the RMS error of even a perfect model of spike rate is bounded from below and thus the black curve asymptotes to a nonzero value. However, model accuracy closely matches the asymptotic accuracy of predictions from averaging previous trials.

The same analysis on a second cell is shown in figure 6(*b*). For this cell the accuracy of predictions from averaging previous trials exceeds that of model predictions at about ten trials. That is, the model provides predictions about as accurate as ten repeated experiments, but more repeats provide better spike count predictions (asymptotic RMS error 0.465 spikes/bin) than the model does (average RMS error 0.499 spikes/bin).

The average model error is plotted as a function of the asymptotic error of the repeat trial predictions for 25 cells from three macaque retinas in figure 7(*a*). Model predictions generally approach the ideal accuracy of repeated trials. Results from 12 cells from one tiger salamander retina are shown in figure 7(*b*). Firing rates are on average much lower in this species, but the quality of model predictions is similar.

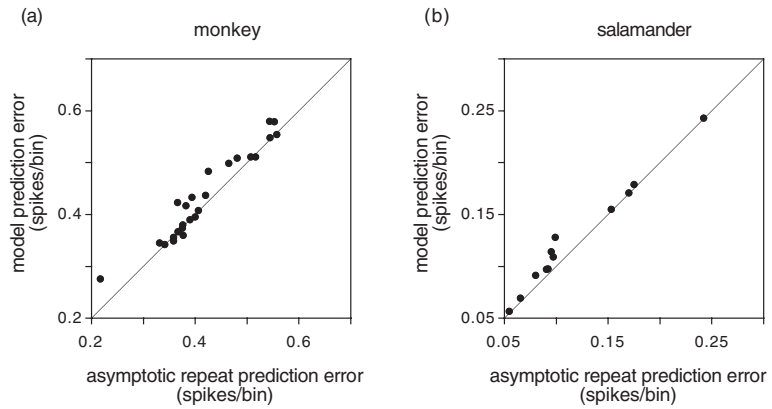

Figure 7. (*a*) RMS model prediction error as a function of asymptotic RMS repeat prediction error for 25 cells from three macaque monkey retinas. Included in the sample are ON and OFF non-opponent cells with large and small receptive fields (possibly parasol and midget), and blue–yellow colour opponent cells. (*b*) The same plot for 12 cells from one salamander retina, of both ON and OFF types. Cells were selected only for stability of responses to repeated trials. $\Delta t = 15$ ms.

5. Discussion

5.1. Summary of the procedure

White noise stimulation and the analysis presented here can be used to obtain a full characterization of the spatial, temporal and chromatic sensitivity of spiking neurons whose responses conform to the static nonlinear model. The essential steps in the procedure are:

(i) choose a stimulus space and present a sequence of randomly selected stimuli which are distributed radially symmetrically in this space;
(ii) compute the spike-triggered average to estimate the linear weights and generator signal;
(iii) compute the mean spike rate as a function of the generator signal to estimate the nonlinearity.

Simple tests based on repeated trials (see section 4) may be used to empirically evaluate the model's accuracy.

5.2. Advantages and disadvantages

What are the advantages and disadvantages of the method presented here compared to conventional methods for estimating signaling properties of visual neurons?

Advantages

Nonlinearities. The method deals naturally and easily with instantaneous nonlinearities that are common in neurons, such as spike threshold and saturation. Evidence of such nonlinearities is given in figures 4(*c*), (*f*) and 5(*b*). These nonlinearities preclude standard linear systems analysis, such as spatial and temporal frequency sensitivity characterizations with sinusoidal contrast modulations (Fourier analysis). However, the weights in the linear stage of the static nonlinear model are easily interpretable, just as they are in a purely linear model: they define the relative sensitivity of the neuron to different aspects of the stimulus. Also, unlike Fourier

analysis, results are expressed directly in the domain of space and time, not spatio-temporal frequency. This avoids additional steps commonly taken to estimate the spatial receptive field and temporal impulse response from responses to sinusoids, such as extracting the fundamental response component and fitting models to the spatio-temporal frequency spectrum (e.g. Croner and Kaplan 1995), which imposes additional assumptions and may introduce systematic errors. Note that in some cases frequency-domain methods are desirable and efficient (Victor 1979, Ringach *et al* 1997).

Interleaving. Because white noise stimuli are interleaved at a fine timescale, estimates of the relative sensitivity of the neuron to different stimuli are fairly robust to non-stationarity in neural responses or unanticipated interruption or termination of the experiment. Also, since the memory of the cell is assessed *post hoc* from a continuous stimulus sequence (section 3.1), there is no need to impose a long delay between successive stimulus presentations to prevent interaction between stimuli. This contrasts with recording the time course of responses to flashed stimuli, where a conservatively long inter-stimulus interval is important.

Adaptation. Because a Gaussian white noise sequence has approximately constant mean and standard deviation over short and long timescales, and because strong and weak stimuli are highly interleaved, effects of adaptation are minimized. This contrasts with stimuli such as pulsed spots and bars that drive the cell strongly for a brief period preceded and followed by no drive.

Multi-neuron recording. The method is easily extended to simultaneous recordings from multiple neurons by stimulating a larger area of the visual field and applying the analysis described here to each cell separately. By contrast, methods that involve centring a stimulus on the receptive field force the experimenter to characterize cells serially. Parallel stimulation and recording is more efficient, provides information about correlated activity and provides internal controls in experiments where global fluctuations in sensitivity can complicate comparisons between cells analysed sequentially.

Disadvantages

On-line interpretation. Since white noise stimuli are random and rapidly interleaved, it can be difficult for the experimenter to immediately assess specific aspects of the neural response (e.g. is it a colour-opponent cell?) by directly observing spike trains over time during stimulation. Instead, computation of the STA is usually required as a first step.

Response variability. Since a Gaussian white noise sequence essentially never repeats, the method provides no direct information about response variability. This must be inferred by grouping responses to stimuli of similar strength or by performing repeat experiments (see section 4).

5.3. Additional issues

Can the above analysis be performed using a stimulus generated by a non-Gaussian white noise process? Many experimenters have used a binary pseudo-random sequence in which the contrast at each point in space and time takes one of two values, for example black or white (e.g. Emerson *et al* 1992, Reid and Alonso 1995). Such stimuli can achieve higher contrasts as well as precisely, rather than asymptotically, balanced statistics (see Sutter 1987, Reid *et al* 1997). Such stimulus vectors are not distributed radially symmetrically in stimulus space, and the STA is not in general proportional to the neural weighting of visual inputs. However proportionality does hold in some cases. First, proportionality is preserved in neurons with precisely linear or half-wave rectified responses. Second, if the integration time of the elementary inputs to

the neuron being studied (e.g. photoreceptors) is long compared to Δt, then at each point in time the inputs effectively respond to a sum of many draws from the binary distribution. The central limit theorem implies that this sum approaches a Gaussian distribution in the limit of long integration times. Thus the applicability of non-Gaussian sequences depends on the properties of the cell and its inputs, while Gaussian sequences are more generally applicable.

Does the dependence on a radially symmetric stimulus distribution constrain the coordinate frame (e.g. colour space) in which stimuli are generated? The STA is guaranteed to be proportional to the linear weighting vector only in a coordinate frame in which the stimulus distribution is radially symmetric, and the resulting weights are valid only for stimuli represented in this coordinate frame. However, if there exists a linear transformation M that symmetrizes the stimulus distribution, the above analysis can be applied to the transformed stimuli to obtain weights valid for the transformed stimuli. It is easily shown that these weights can be transformed to be valid for stimuli in the original coordinate frame using the transpose (not inverse) of M.

Can the technique be applied to neurons outside the visual system? This paper focused on examples in the visual system for concreteness, but none of the theory or application depended on using visual stimuli. Application to another system would rely on defining the entries of stimulus vectors to be quantities that are represented approximately linearly in the inputs to the neuron being studied, such as contrast in the case of retinal ganglion cells. While the method was formulated for spiking neurons here, it is readily applied to continuous-time signals as well (see Sakai 1992).

How accurately does the static nonlinear model describe light responses? The test presented in section 4 showed that retinal ganglion cell spike counts over time were predicted approximately as accurately by the spike rate predictions of the model as by spike rates obtained from repeated stimulus presentations. In other words the model provided a good description of spike rate over time for the stimulus and the temporal resolution tested. As with classical models, this leaves open how accurately it can describe responses to very different stimuli as well as spike train structure at fine temporal resolution (Berry *et al* 1997). For example, important phenomena such as adaptation and refractory periods are not accounted for in the model. Extensions of the model to account for such phenomena may be useful in specific cases. For example, if adaptation is roughly constant during white noise stimulation (since the mean and variance of stimulus intensity are constant), the model may provide a good approximation to responses obtained in that state of adaptation (Sakai *et al* 1995). Changes in the linear weighting function with light level or contrast might capture the effects of adaptation (Chander and Chichilnisky 1999).

How accurately does the static nonlinear model describe the responses of more central neurons? Arguably, the model is most likely to be applicable in the early visual pathways (e.g. retina) while additional linear, static nonlinear or more complex transformations in the chain of neurons conveying signals to more central neurons (e.g. cortex) may make the model inapplicable. Perhaps surprisingly, the model can describe light responses of some cortical cells (DeAngelis *et al* 1993, Anzai *et al* 1999), perhaps because collections of nonlinear inputs spanning different contrast ranges can combine to form a roughly linear spatio-temporal sensitivity profile. More complex nonlinearities observed in some cortical neurons might also be approached with extensions of the technique. For example, consider a direction-selective neuron that does not linearly integrate contrast but sums local optic flow signals computed by its inputs. In this case a moving stimulus with direction and speed modulated according to a Gaussian white noise process may be useful to probe spatial and temporal sensitivity while accounting for the neuron's instantaneous spiking nonlinearity.

Apart from such extensions of the technique, the static nonlinear model is perhaps best viewed as a very tractable and more accurate substitute for the classical and commonly used strictly linear model of spike rate (Enroth-Cugell and Pinto 1970, Movshon *et al* 1978) that fails to capture obvious and nearly ubiquitous response nonlinearities in neurons.

Acknowledgments

I thank David Heeger, Fred Rieke and Denis Baylor for valuable discussions, and Jonathan Demb, Greg Horwitz, Jonathan Victor and Valerie Uzzell for comments on the manuscript. This work was supported by a Helen Hay Whitney Postdoctoral Fellowship and an Alfred P Sloan Foundation Research Fellowship.

References

Anzai A, Ohzawa I and Freeman R D 1999 Neural mechanisms for processing binocular information I. Simple cells *J. Neurophysiol.* **82** 891–908

Berry M J, Warland D K and Meister M 1997 The structure and precision of retinal spike trains *Proc. Natl Acad. Sci. USA* **94** 5411–6

Brown S P, He S and Masland R H 2000 Receptive field microstructure and dendritic geometry of retinal ganglion cells *Neuron* **27** 371–83

Chander D and Chichilnisky E J 1999 Contrast adaptation and gain changes in salamander and monkey retina *Soc. Neurosci. Abstr.* **25** 1431

Chichilnisky E J and Baylor D A 1999 Receptive-field microstructure of blue-yellow ganglion cells in primate retina *Nature Neurosci.* **2** 889–93

Citron M C and Emerson R C 1983 White noise analysis of cortical directional selectivity in cat *Brain Res.* **279** 271–7

Croner L J and Kaplan E 1995 Receptive fields of P and M ganglion cells across the primate retina *Vis. Res.* **35** 7–24

DeAngelis G C, Anzai A, Ohzawa I and Freeman R D 1995 Receptive field structure in the visual cortex: does selective stimulation induce plasticity? *Proc. Natl Acad. Sci. USA* **92** 9682–6

DeAngelis G C, Ohzawa I and Freeman R D 1993 Spatiotemporal organization of simple-cell receptive fields in the cat's striate cortex. II. Linearity of temporal and spatial summation *J. Neurophysiol.* **69** 1118–35

Emerson R C, Korenberg M J and Citron M C 1992 Identification of complex-cell intensive nonlinearities in a cascade model of cat visual cortex *Biol. Cybernet.* **66** 291–300

Enroth-Cugell C and Pinto L 1970 Algebraic summation of centre and surround inputs to retinal ganglion cells of the cat *Nature* **226** 458–9

Hunter I W and Korenberg M J 1986 The identification of nonlinear biological systems: Wiener and Hammerstein cascade models *Biol. Cybern.* **55** 135–44

Korenberg M J and Hunter I W 1986 The identification of nonlinear biological systems: LNL cascade models *Biol. Cybern.* **55** 125–34

Korenberg M J, Sakai H M and Naka K 1989 Dissection of the neuron network in the catfish inner retina. III. Interpretation of spike kernels *J. Neurophysiol.* **61** 1110–20

Marmarelis P Z and Naka K 1972 White-noise analysis of a neuron chain: an application of the Wiener theory *Science* **175** 1276–8

Meister M, Lagnado L and Baylor D A 1995 Concerted signaling by retinal ganglion cells *Science* **270** 1207–10

Movshon J A, Thompson I D and Tolhurst D J 1978 Spatial summation in the receptive fields of simple cells in the cat's striate cortex *J. Physiol.* **283** 53–77

Reid R C and Alonso J M 1995 Specificity of monosynaptic connections from thalamus to visual cortex *Nature* **378** 281–4

Reid R C, Victor J D and Shapley R M 1997 The use of M-sequences in the analysis of visual neurons: linear receptive field properties *Vis. Neurosci.* **14** 1015–27

Rieke F M, Warland D, de Ruyter van Steveninck R R and Bialek W 1997 *Spikes: Exploring the Neural Code* 1st edn (Cambridge, MA: MIT Press)

Ringach D L, Sapiro G and Shapley R 1997 A subspace reverse-correlation technique for the study of visual neurons *Vis. Res.* **37** 2455–64

Sakai H M 1992 White-noise analysis in neurophysiology *Physiol. Rev.* **72** 491–505

Sakai H M, Wang J L and Naka K 1995 Contrast gain control in the lower vertebrate retinas *J. Gen. Physiol.* **105** 815–35

Sutter E E 1987 A practical non-stochastic approach to non-linear time domain analysis *Advanced Methods of Physiological System Modeling* vol 1 (Los Angeles, CA: University of Southern California)

Victor J D 1979 Nonlinear systems analysis: comparison of white noise and sum of sinusoids in a biological system *Proc. Natl Acad. Sci. USA* **76** 996–8