

To appear in: *The New Cognitive Neurosciences*, 3rd edition
Editor: M. Gazzaniga. MIT Press, 2004.

Characterization of Neural Responses with Stochastic Stimuli

Eero P. Simoncelli, Liam Paninski, Jonathan Pillow, Odelia Schwartz*

Center for Neural Science, and
Courant Institute of Mathematical Sciences
New York University
New York, NY 10003

*The Salk Institute
for Biological Studies
La Jolla, CA 92037

August 6, 2003

A fundamental goal of sensory systems neuroscience is the characterization of the functional relationship between environmental stimuli and neural response. The purpose of such a characterization is to elucidate the *computation* being performed by the system. Qualitatively, this notion is exemplified by the concept of the “receptive field”, a quasi-linear description of a neuron’s response properties that has dominated sensory neuroscience for the past 50 years. Receptive field properties are typically determined by measuring responses to a highly restricted set of stimuli, parameterized by one or a few parameters. These stimuli are typically chosen both because they are known to produce strong responses, and because they are easy to generate using available technology.

While such experiments are responsible for much of what we know about the tuning properties of sensory neurons, they typically do not provide a complete characterization of neural response. In particular, the fact that a cell is tuned for a particular parameter, or selective for a particular input feature, does not necessarily tell us how it will respond to an arbitrary stimulus. Furthermore, we have no systematic method of knowing which particular stimulus parameters are likely to govern the response of a given cell, and thus it is difficult to design an experiment to probe neurons whose response properties are not at least partially known in advance.

This chapter provides an overview of some recently developed characterization methods. In general, the ingredients of the problem are: (a) the selection of a set of experimental stimuli; (b) selection of a model of response; (c) a procedure for fitting (estimation) of the model. We discuss solutions of this problem that combine stochastic stimuli with models based on an initial linear filtering stage that serves to reduce the dimensionality of the stimulus space. We begin by describing classical reverse correlation in this context, and then discuss several recent generalizations that increase the power and flexibility of this basic method.

Thanks to Brian Lau, Dario Ringach, Nicole Rust, and Brian Wandell for helpful comments on the manuscript. This work was funded by the Howard Hughes Medical Institute, and the Sloan-Swartz Center for Theoretical Visual Neuroscience at New York University.

1 Reverse correlation

More than thirty years ago, a number of authors applied techniques generally known as *white noise analysis*, to the characterization of neural systems (e.g., deBoer & Kuyper, 1968; Marmarelis & Naka, 1972). There has been a resurgence of interest in these techniques, partly due to the development of computer hardware and software capable of both real-time random stimulus generation and computationally intensive statistical analysis. In the most commonly used form of this analysis, known as *reverse correlation*, one computes the spike-triggered average (STA) by averaging stimulus blocks preceding a spike:

$$\hat{s} = \frac{1}{N} \sum_{i=1}^N \vec{s}_i$$

where the vector \vec{s}_i represents the stimulus block preceding the i th spike. The procedure is illustrated for discretized stimuli in figure 1. The STA is generally interpreted as a representation of the receptive field, in that it represents the “preferred” stimulus of the cell. White noise analysis has been widely used in studying auditory neurons (e.g., Eggermont *et al.*, 1983). In the visual system, spike-triggered averaging has been used to characterize retinal ganglion cells (e.g., Sakai & Naka, 1987; Meister *et al.*, 1994), lateral geniculate neurons (e.g., Reid & Alonso, 1995), and simple cells in primary visual cortex (V1) (e.g., Jones & Palmer, 1987; McLean & Palmer, 1989; DeAngelis *et al.*, 1993).

1.1 Model characterization with spike-triggered averaging

In order to interpret the STA more precisely, we can ask what model it can be used to characterize. The classical answer to this question comes from nonlinear systems analysis:¹ the STA provides an estimate of the first (linear) term in a polynomial series expansion of the system response function. If the system is truly linear, then the STA provides a complete characterization. It is well known, however, that neural responses are not linear. Even if one describes the neural response in terms of mean spike rate, this typically exhibits nonlinear behavior with respect to the input signal, such as thresholding and saturation. Thus, the first term of a Wiener/Volterra series, as estimated with the STA, will typically not provide a full description of neural response. One can of course include higher-order terms in the series expansion. But each successive term in the expansion requires a substantial increase in the amount of experimental data. And limiting the analysis only to the first and second order terms, for example, may still not be sufficient to characterize nonlinear behaviors common to neural responses.

Fortunately, it is possible to use the STA as a first step in fitting a model that can describe neural response more parsimoniously than a series expansion. Specifically, suppose that the

¹The formulation is due to Wiener (Wiener, 1958), based on earlier results by Volterra (Volterra, 1913). See (Rieke *et al.*, 1997) or (Dayan & Abbott, 2001) for reviews of application to neurons.

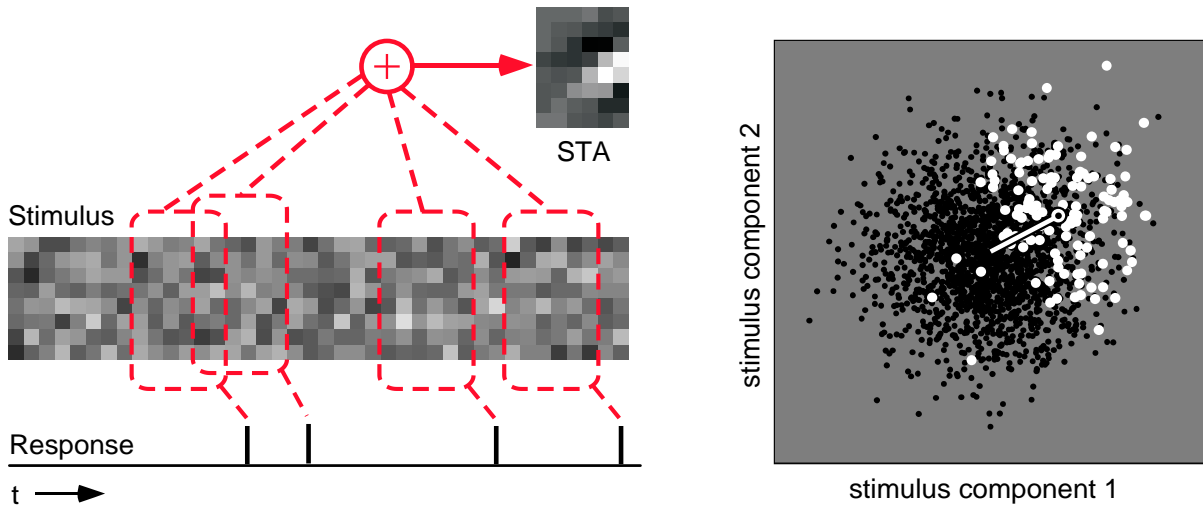


Figure 1. Two alternative illustrations of the reverse correlation procedure. **Left:** Discretized stimulus sequence and observed neural response (spike train). On each time step, the stimulus consists of an array of randomly chosen values (eight, for this example), corresponding to the intensities of a set of individual pixels, bars, or any other fixed spatial patterns. The neural response at any particular moment in time is assumed to be completely determined by the stimulus segment that occurred during a pre-specified interval in the past. In this figure, the segment covers six time steps, and lags three time steps behind the current time (to account for response latency). The spike-triggered ensemble consists of the set of segments associated with spikes. The spike-triggered average (STA) is constructed by averaging these stimulus segments (and subtracting off the average over the full set of stimulus segments). **Right:** Geometric (vector space) interpretation of the STA. Each stimulus segment corresponds to a point in a d -dimensional space (in this example, $d = 48$) whose axes correspond to stimulus values (e.g., pixel intensities) during the interval. For illustration purposes, the scatter plot shows only two of the 48 axes. The spike-triggered stimulus segments (white points) constitute a subset of all stimulus segments presented (black points). The STA, indicated by the line in the diagram, corresponds to the difference between the mean (center of mass) of the spike-triggered ensemble, and the mean of the raw stimulus ensemble. Note that the interpretation of this representation of the stimuli is only sensible under Poisson spike-generation - the scatter plot depiction implies that the probability of spiking depends only on the position in the stimulus space.

[t]

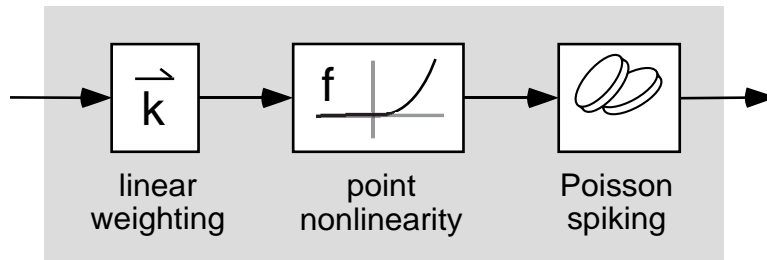


Figure 2. Block diagram of the linear-nonlinear-Poisson (LNP) model. On each time step, the components of the stimulus vector are linearly combined using a weight vector, \vec{k} . This response of this linear filter is then passed through a nonlinear function $f()$, whose output determines the instantaneous firing rate of a Poisson spike generator.

response is generated in a cascade of three stages: (1) a linear function of the stimulus over a recent period of time, (2) an instantaneous (also known as *static* or *memoryless*) nonlinear transformation, and (3) a Poisson spike generation process, whose instantaneous firing rate comes from the output of the previous stage. That is, the probability of observing a spike during any small time window is a nonlinear function of a linear-filtered version of the stimulus. This model is illustrated in figure 2, and we'll refer to it as a linear-nonlinear-Poisson (LNP) model. The third stage, which essentially amounts to an assumption that the generation of spikes depends only on the recent stimulus and not on the history of previous spike times, is often not stated explicitly but is critical to the analysis.

Under suitable conditions on the stimulus distribution and the nonlinearity, the spike-triggered average produces an estimate of the linear filter in the first stage of the LNP model (see (Chichilnisky, 2001) for overview and additional references). The result is most easily understood geometrically, as depicted in figure 1. Assume the stimulus is discretized, and that the response of the cell at any moment in time depends only on the values within a fixed-length time interval preceding that time. A typical stimulus would be the intensities of a set

of pixels covering some spatial region of a video display, for every temporal video frame over the time interval (see left panel of figure 1). In this case, the stimulus segment presented over the interval preceding a spike corresponds to a vector containing d components, one for the intensity of each pixel in each frame. The vectors of all stimulus segments presented during an experiment may be represented as a set of points in a d -dimensional stimulus space, as illustrated in figure 1. We'll refer to this as the *raw stimulus ensemble*. This ensemble is under the control of the experimenter, and the samples are typically chosen randomly according to some probability distribution. A statistically *white* ensemble corresponds to the situation where the components of the stimulus vector are uncorrelated. If in addition the density of each component is Gaussian, and all have the same variance, then the full d -dimensional distribution will be spherically symmetric.

In a model with Poisson spike generation, the probability of a spike occurring after a given stimulus block depends only on the content of that block, or equivalently, on the position of the corresponding vector in the d -dimensional space. From an experimental perspective, this means that the distribution of the spike-triggered stimulus ensemble indicates which regions of the stimulus space are more likely (or less likely) to elicit spikes. More specifically, for each region of the stimulus space, the ratio of the frequency of occurrence of spike-triggered stimuli to that of raw stimuli gives the instantaneous firing rate. From this description, it might seem that one could simply count the number of spikes and stimuli in each region (i.e., compute multi-dimensional histograms of the binned raw and spike-triggered stimuli), and take the quotient to compute the firing rate. But this is impractical due to the so-called "curse of dimensionality": the amount of data needed to sufficiently fill the histogram bins in a d -dimensional space grows exponentially with d . Thus, one cannot hope to compute such a firing rate function for a space of more than a few dimensions.

The assumption of an LNP model allows us to avoid this severe data requirement. In particular, the linear stage of the model effectively collapses the entire d -dimensional space onto a single axis, as illustrated in figure 3. The STA provides an estimate of this axis, under the assumption that the raw stimulus distribution is spherically symmetric² (e.g., Chichilnisky, 2001; Theunissen *et al.*, 2001). Once the linear filter has been estimated, we may compute its response and then examine the relationship between the histograms of the raw and spike-triggered ensembles within this one-dimensional space. Specifically, for each value of the linear response, the nonlinear function in the LNP model may be estimated as the quotient of the frequency of spike occurrences to that of stimulus occurrences (see figure 3). Because this quotient is taken between two one-dimensional histograms (as opposed to d -dimensional histograms), the data requirements for accurate estimation are greatly reduced. Note also that the nonlinearity can be arbitrarily complicated (even discontinuous). The only constraint is that it must produce a

²Technically, an elliptically symmetric distribution is also acceptable. The stimuli are first transformed to a spherical distribution using a whitening operation, the STA is computed, and the result is transformed back to the original stimulus space.

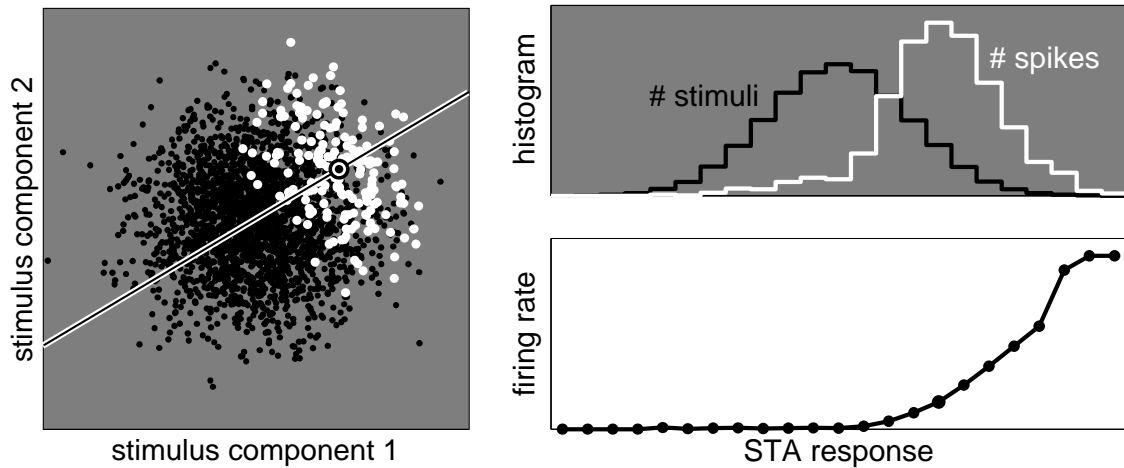


Figure 3. Simulated characterization of an LNP model using reverse correlation. The simulation is based on a sequence of 20,000 stimuli, with a response containing 950 spikes. **Left:** The STA (black and white “target”) provides an estimate of the linear weighting vector, \vec{k} (see also figure 1). The linear response to any particular stimulus corresponds to the position of that stimulus along the axis defined by \vec{k} (line). **Right, top:** raw (black) and spike-triggered (white) histograms of the linear (STA) responses. **Right, bottom:** The quotient of the spike-triggered and raw histograms gives an estimate of the nonlinearity that generates the firing rate.

change in the mean of the spike-triggered ensemble, as compared with the original stimulus ensemble. Thus, the interpretation of reverse correlation in the context of the LNP model is a significant departure from the Wiener/Volterra series expansion, in which even a simple sigmoidal nonlinearity would require the estimation of many terms for accurate characterization.

The reverse correlation approach relies less on prior knowledge of the neural response properties by covering a wide range of visual input stimuli in a relatively short amount of time, and it can produce a complete characterization in the form of the LNP model (see (Chichilnisky, 2001) for further discussion). But clearly, the method can fail if the neural response does not fit the assumptions of the model. For example, if the neural nonlinearity and the stimulus distribution interact in such a way that the mean of the raw stimuli and mean of the spike-triggered stimuli do not differ, the STA will be zero, thus failing to provide an estimate of the linear stage of the model. Even if the reverse correlation procedure succeeds in estimating the model parameters, the model might not provide a good characterization. Specifically, the true neural response may not be restricted to a single direction in the stimulus space, or the spike generation may not be well-described as a Poisson process. In the following sections, we describe some extensions of reverse correlation to handle these types of model failure.

2 Extension to multiple dimensions with STC

The STA analysis relies on changes in the mean of the spike-triggered stimulus ensemble to estimate the linear stage of an LNP model. This linear stage corresponds to a single filter, which responds to a single direction in the stimulus space. But many neurons exhibit behaviors that are not well described by this model. For example, the “energy model” of complex cells in primary visual cortex posits the existence of two linear filters (an even- and odd-symmetric pair), whose rectified responses are then combined (Adelson & Bergen, 1985). Not only does this model use two linear filters, but the symmetry of the rectifying nonlinearity means that the STA will be zero, thus providing no information about the linear stage of the model. In this particular case, a variety of second-order interaction analyses have been developed to recover the two filters (e.g., Emerson *et al.*, 1987; Szulborski & Palmer, 1990; Emerson *et al.*, 1992).

We’d like to be able to characterize this type of multiple filter model. Specifically, one would like to recover the filters, as well as the nonlinear function by which their responses are combined. The classical nonlinear systems analysis approach to this problem (Marmarelis & Marmarelis, 1978; Korenberg *et al.*, 1989) proceeds by estimating the second-order term in the Wiener series expansion, which describes the response as a weighted sum over all pairwise products of components in the stimulus vector. The weights of this sum (the second-order Wiener kernel), may be estimated from the spike-triggered *covariance* (STC) matrix, computed as a sum of outer products of the spike-triggered stimulus vectors with the STA subtracted:³

$$C = \frac{1}{N-1} \sum_{i=1}^N (\vec{s}_i - \hat{s}) \cdot (\vec{s}_i - \hat{s})^T.$$

This second-order Wiener series gives a quadratic model for neural responses, and thus remains ill-equipped to accurately model sharply asymmetric or saturating nonlinearities. As in the case of the STA, however, the STC may be used as a starting point for estimation of another model that may be more relevant in describing some neural responses. In particular, one can assume that the neural response is again determined by an LNP cascade model (figure 2), but that the initial linear stage now is *multi-dimensional*. That is, the response comes from applying a small set of linear filters, followed by Poisson spike generation, with firing rate determined by some nonlinear combination of the filter outputs.

Under suitable conditions on the stimulus distribution and the nonlinear stage, the STC may be used to estimate the linear stage (de Ruyter van Steveninck & Bialek, 1988; Brenner *et al.*, 2000; Paninski, 2003). Again, the idea is most directly explained geometrically: we seek those directions in the stimulus space along which the *variance* of the spike-triggered ensemble differs from that of the raw ensemble. Loosely speaking, an increase in variance (with no change

³An alternative method is to project, rather than subtract, the STA from the stimulus set (Schwartz *et al.*, 2002; Rust *et al.*, 2004).

in the mean) indicates a stimulus dimension that is excitatory, and a decrease in variance indicates suppression. The advantage of this description is that variance analysis in multiple dimensions is very well-understood mathematically. The surface representing the variance (standard deviation) of the spike-triggered stimulus ensemble consists to those vectors \vec{v} satisfying $\vec{v}^T C^{-1} \vec{v} = 1$. This surface is an ellipsoid, and the principal axes of this ellipsoid may be recovered using standard eigenvector techniques (i.e., principal component analysis). Specifically, the eigenvectors of C represent the principal axes of the ellipsoid, and the corresponding eigenvalues represent the variances along each of these axes.⁴

Thus, by determining which variances are significantly different from those of the underlying raw stimulus ensemble, the STC may be used to estimate the set of axes (i.e., linear filters) from which the neural response is derived. As with the STA, the second nonlinear stage of the model may then be estimated by looking at the spiking response as a function of the responses of these linear filters. The correctness of the STC-based estimator can be guaranteed if (but only if) the stimuli are drawn from a Gaussian distribution (Paninski, 2003), a stronger condition than the spherical symmetry required for the STA. Spike-triggered covariance analysis has been used to determine both excitatory (de Ruyter van Steveninck & Bialek, 1988; Brenner *et al.*, 2000; Touryan *et al.*, 2002; Rust *et al.*, 2004) as well as suppressive (Schwartz *et al.*, 2002; Rust *et al.*, 2004) response properties of visual neurons. Here, we'll consider two simulation examples to illustrate the concept, and to provide some idea of the type of nonlinear behaviors that can be revealed using this analysis.

The first example, shown in figure 4, is a simulation of a standard V1 complex cell model (see also simulations in (Sakai & Tanaka, 2000)). The model is constructed from two space-time oriented linear receptive fields, one symmetric and the other antisymmetric (Adelson & Bergen, 1985). The linear responses of these two filters are squared and summed, and the resulting signal then determines the instantaneous firing rate:

$$g(\vec{s}) = r \left[(\vec{k}_1 \cdot \vec{s})^2 + (\vec{k}_2 \cdot \vec{s})^2 \right].$$

The recovered eigenvalues indicate that two directions within this space have substantially higher variance than the others. The eigenvectors associated with these two eigenvalues correspond to the two filters in the model.⁵ The raw- and spike-triggered stimulus ensembles may then be filtered with these two eigenvectors, and the two-dimensional nonlinear function that governs firing rate corresponds to the quotient of the two histograms, analogous to the one-dimensional example shown in figure 1. Similar pairs of excitatory axes have been ob-

⁴More precisely, the *relative* variance between the spike-triggered and raw stimulus ensembles can be computed either by performing the eigenvector decomposition on the difference of the two covariance matrices (de Ruyter van Steveninck & Bialek, 1988; Brenner *et al.*, 2000), or by applying an initial whitening transformation to the raw stimuli before computing the STC (Schwartz *et al.*, 2002; Rust *et al.*, 2004). The latter is equivalent to solving for principal axes of an ellipse that represents the ratio of spike-triggered and raw variances.

⁵Technically, the recovered eigenvectors represent two orthogonal axes that span a subspace containing the two filters.

tained from STC analysis of V1 cells in cat (Touryan *et al.*, 2002) as well as monkey (Rust *et al.*, 2004).

As a second example, we choose a simplified version of a divisive gain control model, as have been used to describe nonlinear properties of neurons in primary visual cortex (Albrecht & Geisler, 1991; Heeger, 1992). Specifically, our model neuron’s instantaneous firing rate is governed by one excitatory filter and one divisively suppressive filter:

$$g(\vec{s}) = r \frac{1 + (\vec{k}_1 \cdot \vec{s})^2}{1 + (\vec{k}_1 \cdot \vec{s})^2/2 + (\vec{k}_2 \cdot \vec{s})^2}.$$

The simulation results are shown in figure 5. The recovered eigenvalue distribution reveals one large-variance axis and one small-variance axis, corresponding to the two filters, \vec{k}_1 and \vec{k}_2 respectively. After projecting the stimuli onto these two axes, the two-dimensional nonlinearity is estimated, and reveals an approximately saddle-shaped function, indicating the interaction between the excitatory and suppressive signals. Similar suppressive filters have been obtained from STC analysis of retinal ganglion cells (both salamander and monkey) (Schwartz *et al.*, 2002) and simple and complex cells in monkey V1 (Rust *et al.*, 2004). In these cases, a combined STA/STC analysis was used to recover multiple linear filters. The number of recovered filters was typically large enough that the direct estimation of the nonlinearity (i.e., dividing the spike-triggered histogram by the raw histogram) was not feasible. As such, the nonlinear stage was estimated by fitting specific parametric models on top of the output of the linear filters.

3 Experimental caveats

In addition to the limitations of the LNP model, it is important to understand the tradeoffs and potential problems that may arise in using STA/STC characterization procedures. We provide a brief overview of these issues, which can be quite complex. See (Rieke *et al.*, 1997; Chichilnisky, 2001; Paninski, 2003) for further description.

The accuracy of STA/STC filter estimates depends on three elements: (1) the dimensionality of the stimulus space, (2) the number of spikes collected, and (3) the strength of the response signal, relative to the standard deviation of the raw stimulus ensemble.⁶ The first two of these interact in a simple way: the quality of estimates increases as a function of the ratio of the number spikes to the number of stimulus dimensions. Thus, the pursuit of more accurate estimates leads to a simultaneous demand for more spikes and reduced stimulus dimensionality.

These demands must be balanced against several opposing constraints. The collection of a

⁶Technically, the response signal strength is defined as the STA magnitude, or in the case of STC to the square root of the difference between the eigenvalue and σ^2 , the variance of the raw stimuli (Paninski, 2003).

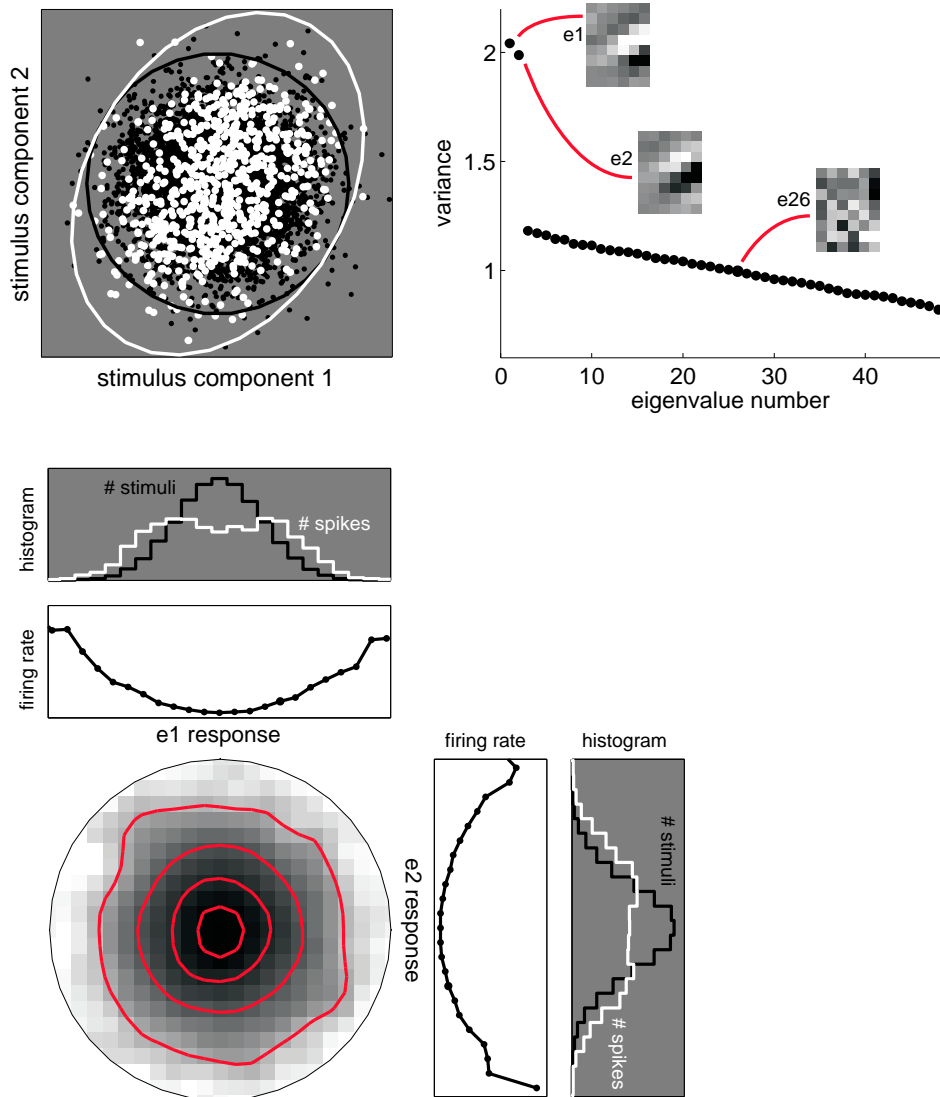


Figure 4. Simulated characterization of a particular LNP model using spike-triggered covariance (STC). In this model, the Poisson spike generator is driven by the sum of squares of two oriented linear filter responses. As in figure 1, filters are 6×8 , and thus live in a 48-dimensional space. The simulation is based on a sequence of 50,000 raw stimuli, with a response containing 4,500 spikes. **Top, left:** simulated raw and spike-triggered stimulus ensembles, viewed in a two-dimensional subspace that illustrates the model behavior. The covariance of these ensembles within this two-dimensional space is represented geometrically by an ellipse that is three standard deviations from the origin in all directions. The raw stimulus ensemble has equal variance in all directions, as indicated by the black circle. The spike-triggered ensemble is elongated in one direction, as represented by the white ellipse. **Top, right:** Eigenvalue analysis of the simulated data. The principle axes of the covariance ellipse correspond to the eigenvectors of the spike-triggered covariance matrix, and the associated eigenvalues indicate the variance of the spike-triggered stimulus ensemble along each of these axes. The plot shows the full set of 48 eigenvalues, sorted in descending order. Two of these are substantially larger than the others, and indicate the presences of two axes in the stimulus space along which the model responds. The others correspond to stimulus directions that the model ignores. Also shown are three example eigenvectors (6×8 linear filters). **Bottom, one-dimensional plots:** Spike-triggered and raw histograms of responses of the two high-variance linear filters, along with the nonlinear firing rate functions estimated from their quotient (see figure 3). **Bottom, two-dimensional plot:** the quotient of the two-dimensional spike-triggered and raw histograms provides an estimate of the two-dimensional nonlinear firing rate function. This is shown as a circular-cropped grayscale image, where intensity is proportional to firing rate. Superimposed contours indicate four different response levels.

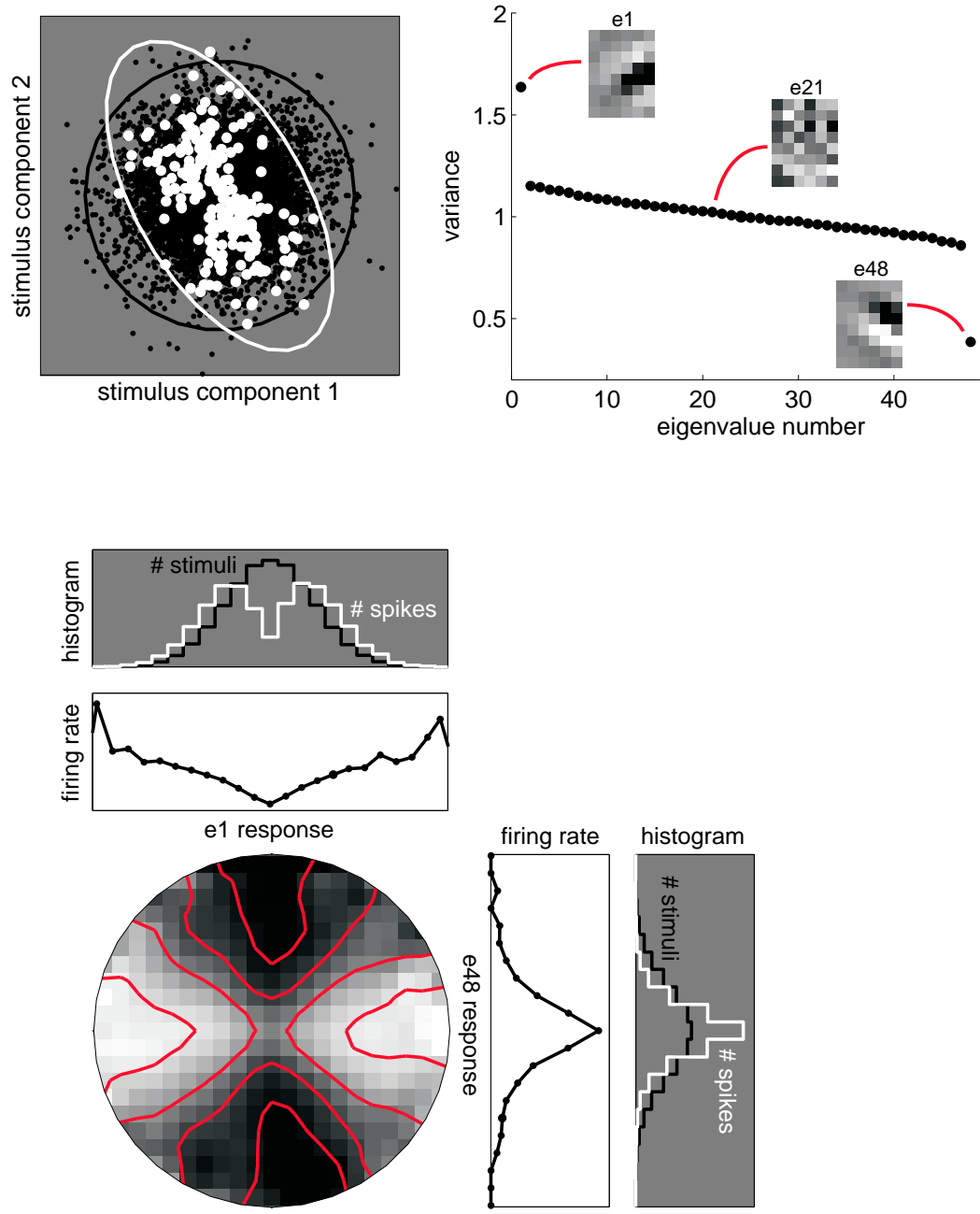


Figure 5. Characterization of a simulated LNP model, constructed from the squared response of one linear filter divided by the sum of squares of its own response and the response of another filter. The simulation is based on a sequence of 200,000 raw stimuli, with a response containing 8,000 spikes. See text and caption of figure 4 for details.

large number of spikes is limited by the realities of single-cell electrophysiology. Experimental recordings are restricted in duration, especially since the response properties need to remain stable and consistent throughout the recording. On the other hand, reducing the stimulus dimensionality is also problematic. One of the most widely touted advantages of white noise characterization over traditional experiments is that the stimuli can cover a broader range of visual input stimuli, and that these randomly selected stimuli are less likely to induce artifacts or experimental bias than a set that is hand-selected by the experimenter.

In practice, however, white noise characterization still requires the experimenter to place restrictions on the stimulus set. For visual neurons, even with stimuli composed in the typical fashion from individual pixels, one must choose the spatial size and temporal duration of these pixels. If the pixels are too small, then not only will the stimulus dimensionality be large (in order to fully cover the spatial and temporal “receptive field”), but the effective stimulus contrast that reaches the neuron will be quite low, resulting in a low spike rate. Both of these effects will reduce the accuracy of the estimated filters. On the other hand, if the pixels are too large, then the recovered linear filters will be quite coarse (since they are constructed from blocks that are the size of the pixels). More generally, one can use stimuli that provide a better basis for receptive field description, and that are more likely to elicit strong neural responses, by defining them in terms of parameters that are more relevant than pixel intensities. Examples include stimuli restricted in spatial frequency (Ringach *et al.*, 1997) and stimuli defined in terms of velocity (de Ruyter van Steveninck & Bialek, 1988; Bair *et al.*, 1997; Brenner *et al.*, 2000).

While the choice of stimuli plays a critical role in controlling the accuracy (variance) of the filter estimates, the probability distribution from which the stimuli are drawn must be chosen carefully to avoid bias in the estimates. For example, with the single-filter LNP model, the stimulus distribution must be spherically symmetric in order to guarantee that the STA gives an unbiased estimate of the linear filter (e.g., Chichilnisky, 2001). Figure 6 shows two simulations of an LNP model with a simple sigmoidal nonlinearity, each demonstrating that the use of non-spherical stimulus distributions can lead to poor estimates of the linear stage. The first example shows a “sparse noise” experiment, in which the stimulus at each time step lies along one of the axes. For example, many authors have characterized visual neurons using images with only a single white/black pixel amongst a background of gray pixels in each frame (e.g., Jones & Palmer, 1987). As shown in the figure, even a simple nonlinearity (in this case, a sigmoid) can result in an STA that is heavily biased.⁷ The second example uses stimuli in which each component is drawn from a uniform distribution, which produces an estimate biased toward the “corner” of the space. The use of non-Gaussian distributions (e.g., uniform or binary) for white noise stimuli is quite common, as the samples are easy to generate and the resulting stimuli can have higher contrast and thus produce higher average spike rates. In

⁷Note, however, that the estimate will be unbiased in the case of a purely linear neuron, or of a halfwave-rectified linear neuron (Ringach *et al.*, 1997).

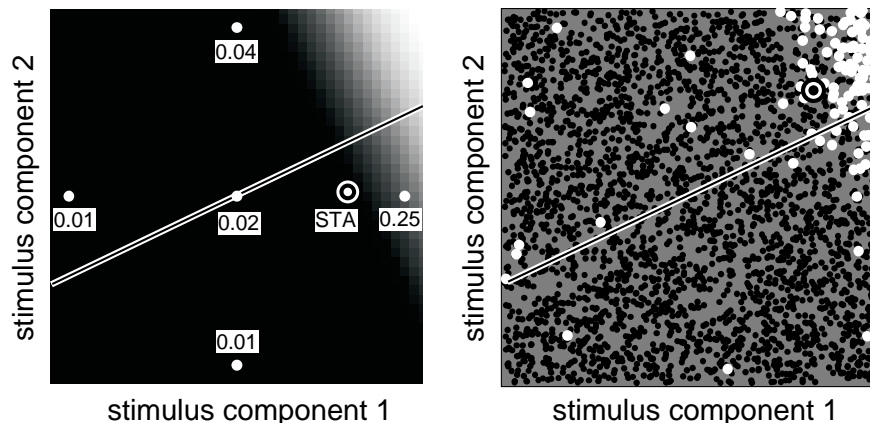


Figure 6. Simulations of an LNP model demonstrating bias in the STA for two different non-spherical stimulus distributions. The linear stage of the model neuron corresponds to an oblique axis (line in both panels), and the firing rate function is a sigmoidal nonlinearity (firing rate corresponds to intensity of the underlying grayscale image in the left panel). In both panels, the black and white “target” indicates the recovered STA. **Left:** Simulated response to sparse noise. The plot shows a two-dimensional subspace of a 10-dimensional stimulus space. Each stimulus vector contains a single element with a value of ± 1 , while all other elements are zero. Numbers indicate the firing rate for each of the possible stimulus vectors. The STA is strongly biased toward the horizontal axis. **Right:** Simulated response of the same model to uniformly distributed noise. The STA is now biased toward the corner. Note that in both examples, the estimate will not converge to the correct answer, regardless of the amount of data collected.

practice, their use has been justified by assuming that the linear filter is smooth relative to the pixel size/duration (e.g., Chichilnisky, 2001).

While the generalization of the LNP model to the multidimensional case substantially increases its power and flexibility, the STC method can fail in a manner analogous to that described for the STA. Specifically, if the neural response varies in a particular direction within the stimulus space, but the variance of the spike-triggered ensemble does not differ from the raw ensemble in that direction, then the method will not be able to recover that direction. In addition, the STC method is more susceptible to biases caused by statistical idiosyncrasies in the stimulus distribution than is the STA. These concerns have motivated the development of estimators that are guaranteed to converge to the correct linear axes under much more general conditions (Sharpee *et al.*, 2003; Paninski, 2003). The basic idea is quite simple: instead of relying on a particular statistical moment (e.g., mean or variance) for comparison of the spike-triggered and raw stimulus distributions, one can use a more general comparison function that can identify virtually any difference between the two distributions. A natural choice for such a function is information-theoretic: one can compare the *mutual information* between a set of filter responses and the probability of a spike occurring. The resulting estimator is more computationally expensive, but has been shown to be more accurate in several different simulation examples (Sharpee *et al.*, 2003; Paninski, 2003).

4 Non-Poisson spike generation

The LNP models described above provide an alternative to the classical Wiener series expansion, but they still assume that the information a neuron carries about the stimulus is contained in its instantaneous firing rate. These models thus ignore any history dependence in the spike train that might result from the dynamics underlying spike generation, such as the refractory period. A number of authors have demonstrated that these Poisson assumptions do not accurately capture the statistics of neural spike trains (Berry & Meister, 1998; Reich *et al.*, 1998; Keat *et al.*, 2001). It is therefore important to ask: (1) How do realistic spiking mechanisms affect the LNP characterization of a neuron? and (2) Is it possible to extend the characterization methods described above to incorporate more realistic spiking dynamics?

The first of these questions has been addressed using simulations and mathematical analysis of neural models with both Hodgkin-Huxley and leaky integrate-and-fire spike generation (Agüera y Arcas *et al.*, 2001; Pillow & Simoncelli, 2003; Paninski *et al.*, 2003; Agüera y Arcas & Fairhall, 2003). In these cases, spike generation nonlinearities can interfere with the temporal properties of the linear filters estimated with STA or STC analysis. Figure 7 shows an example, using a model in which a single linear filter drives a non-Poisson spike generator. In this case, the STA provides a biased estimate of the true linear filter. Moreover, the history-dependent effects of spike generation are not confined to a single direction of the stimulus space. Even though the model response is generated from the output of a single linear filter, STC analysis reveals additional relevant directions in the stimulus space. Thus, describing non-Poisson responses with an LNP model results in a high-dimensional characterization, when a low-dimensional model with a more appropriate spike generator would suffice.

A recently proposed approach to the problem of spike-history dependence is to perform STA/STC analysis using only isolated spikes, or those that are widely separated in time from other spikes (Agüera y Arcas *et al.*, 2001; Agüera y Arcas & Fairhall, 2003). This has the advantageous effect of eliminating refractory and other short-term effects from the responses being analyzed, but as a consequence does not characterize the history-dependence of the spikes. Furthermore, the discarded spikes, which may constitute a substantial proportion of the total, correspond to periods of rapid firing and thus seem likely to carry potent information about a neuron's selectivity.

An alternative is to modify the LNP description to incorporate more realistic spike generation effects, and develop characterization procedures for this model. One proposed technique incorporates a "recovery function" that modulates the spike probability following the occurrence of each spike (Miller, 1985; Berry & Meister, 1998; Kass & Ventura, 2001). Specifically, the instantaneous Poisson firing rate is set by the product of the output of an LN stage and this recovery function. The resulting model can produce both absolute and relative refractory effects.

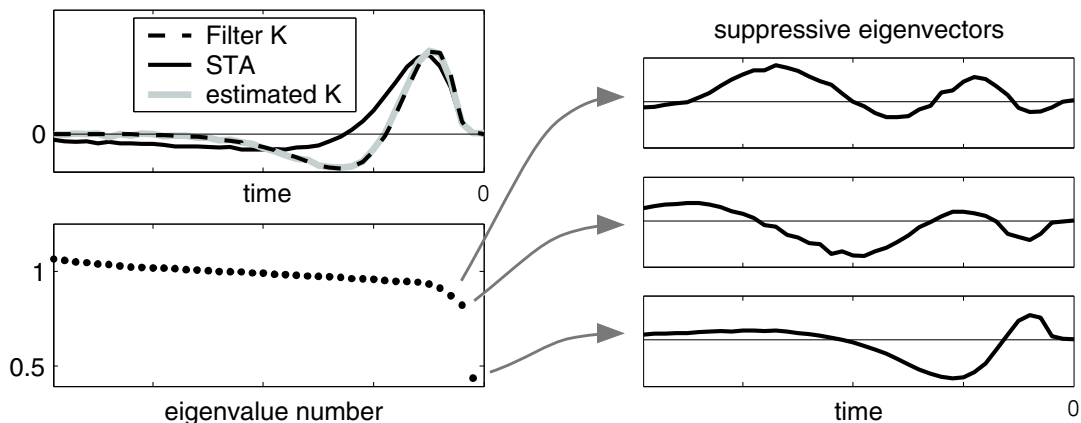


Figure 7. Simulated spike-triggered analysis of a neuron with noisy, leaky integrate-and-fire (NLIF) spike generation. Input is filtered with a single linear filter (K), followed by NLIF spiking. A purely temporal filter was selected because the effects of non-Poisson spike generation manifest themselves in the temporal domain. **Upper left:** Linear filter of the model (dashed line), a 32-sample function chosen to resemble the temporal impulse response of a macaque retinal ganglion cell. Also shown are the STA computed from a simulated spike train (solid), and the linear filter estimated using maximum likelihood (Pillow *et al.*, 2004). **Lower left:** Eigenvalues computed from the spike-triggered covariance of the simulated spike train. **Right:** Linear filters (eigenvectors) associated with the three smallest eigenvalues.

Another alternative is to use an explicit parametric model of spike generation and to develop estimation techniques for front-end stimulus selectivity in conjunction with the parameters of the spike generator (e.g., Keat *et al.*, 2001; Pillow *et al.*, 2004). As an example, consider the estimation of a two-stage model consisting of a linear filter followed by a noisy, leaky integrate-and-fire (NLIF) spike generator. The stimulus dependence of this model is determined by the linear filter, but the N and P stages of the LNP model are replaced by the NLIF mechanism. Although direct STA analysis cannot recover the linear filter in this model, it is possible to use a maximum likelihood estimator to recover both the linear filter and the parameters of the spike generator (threshold voltage, reset voltage, leak conductance, and noise variance) (Pillow *et al.*, 2004). The estimation procedure can start from the STA as an initial guess for the true filter, and ascend the likelihood function to obtain optimal estimates of the filter and the NLIF parameters. This procedure is computationally efficient and is guaranteed to converge to the correct answer. A simulated example is shown in figure 7. The method provides a characterization of both the spatio-temporal filter that drives neural response and the nonlinear biophysical response properties that transform this drive into spike trains.

5 Discussion

We've described a set of techniques for characterizing the functional response properties of neurons using stochastic stimuli. We've relied throughout on an assumption that the response

of the neuron is governed by an initial linear stage that serves to reduce the dimensionality of the stimulus space. While this assumption may seem overly restrictive, it is important to realize that it is the fundamental ingredient that allows one to infer general response properties from measured responses to a relatively small number of stimuli. The linear stage is followed by a nonlinearity upon which we place fairly minimal constraints. We described two moment-based methods of estimating the linear stage – STA and STC – which are both conceptually elegant and efficient to calculate.

In addition to the assumption of an initial low-dimensional linear stage, there are two well-known potential drawbacks of these approaches. First, the techniques place fairly strong constraints on the set of stimuli that must be used in an experiment. There has been an increased interest in recent years in presenting naturalistic stimuli to neurons, so as to assess their behavior under normal operating conditions (e.g., Dan *et al.*, 1996; Baddeley *et al.*, 1998; Theunissen *et al.*, 2001; Ringach *et al.*, 2002; Smyth *et al.*, 2003). Analysis of such data is tricky, since naturalistic images are highly non-Gaussian (Field, 1987; Daugman, 1989), and (as described earlier) the basic STA/STC technique relies on a Gaussian stimulus distribution. Estimators based on information-theoretic measures, as described in section 3, seem promising in this context since they place essentially no restriction on the stimulus ensemble.

A second drawback is that the assumption of Poisson spike generation provides a poor account of the spiking behavior of many neurons (Berry & Meister, 1998; Reich *et al.*, 1998; Keat *et al.*, 2001). As discussed in section 4, STA/STC analysis of an LN model driving a more realistic spiking mechanism (e.g., integrate-and-fire or Hodgkin-Huxley) can lead to significant biases in the estimate of the linear stage. A number of techniques currently under development are attempting to address these issues.

Finally, we mention two interesting directions for future research. First, the techniques described here can be adapted for the analysis of multi-neuronal interactions (e.g., Nykamp, 2003). Such methods have been applied, for example, in visual cortex (Tsodyks *et al.*, 1999), motor cortex (Paninski *et al.*, 2004) and hippocampus (Harris *et al.*, 2003). Second, it would be desirable to develop techniques that can be applied to a cascaded series of LNP stages. This will be essential for modeling responses in higher-order sensory areas, which are presumably constructed from more peripheral responses.

References

- Adelson, E. H. & Bergen, J. R. (1985). Spatiotemporal energy models for the perception of motion. *Journal of the Optical Society of America A*, 2, 284–299.
- Agüera y Arcas, B. & Fairhall, A. L. (2003). What causes a neuron to spike? *Neural Computation*, 15, 1789–1807.
- Agüera y Arcas, B., Fairhall, A. L., & Bialek, W. (2001). What can a single neuron compute? In *Adv. Neural Information Processing Systems (NIPS*00)*, volume 13, pages 75–81.
- Albrecht, D. G. & Geisler, W. S. (1991). Motion sensitivity and the contrast-response function of simple cells in the visual cortex. *Visual Neuroscience*, 7, 531–546.
- Baddeley, R., Abbott, L. F., Booth, M. C., Sengpiel, F., Freeman, T., Wakeman, E. A., & Rolls, E. T. (1998). Responses of neurons in primary and inferior temporal visual cortices to natural scenes. *Proc. Roy. Soc. (Lond.)*, B264, 1775–1783.
- Bair, W., Cavanaugh, J. R., & Movshon, J. A. (1997). Reconstructing stimulus velocity from neuronal responses in area MT. In Mozer, M. C., Jordan, M. I., & Petsche, T., editors, *Adv. Neural Information Processing Systems (NIPS*96)*, volume 9, pages 34–40. MIT Press.
- Berry, M. & Meister, M. (1998). Refractoriness and neural precision. *Journal of Neuroscience*, 18, 2200–2211.
- Brenner, N., Bialek, W., & de Ruyter van Steveninck, R. R. (2000). Adaptive rescaling maximizes information transmission. *Neuron*, 26, 695–702.
- Chichilnisky, E. J. (2001). A simple white noise analysis of neuronal light responses. *Network: Computation in Neural Systems*, 12(2), 199–213.
- Dan, Y., Atick, J. J., & Reid, R. C. (1996). Efficient coding of natural scenes in the lateral geniculate nucleus: Experimental test of a computational theory. *J. Neuroscience*, 16, 3351–3362.
- Daugman, J. G. (1989). Entropy reduction and decorrelation in visual coding by oriented neural receptive fields. *IEEE Trans. Biomedical Engineering*, 36(1), 107–114.
- Dayan, P. & Abbott, L. F. (2001). *Theoretical Neuroscience*. MIT Press, Cambridge, MA.
- de Ruyter van Steveninck, R. & Bialek, W. (1988). Coding and information transfer in short spike sequences. In *Proc. Soc. Lond. B. Biol. Sci.*, volume 234, pages 379–414.
- DeAngelis, G. C., Ohzawa, I., & Freeman, R. D. (1993). The spatiotemporal organization of simple cell receptive fields in the cat's striate cortex. II. Linearity of temporal and spatial summation. *Journal of Neurophysiology*, 69, 1118–1135.

- deBoer, E. & Kuyper, P. (1968). Triggered correlation. In *IEEE Trans Biomed Eng*, volume 15, pages 169–179.
- Eggermont, J. J., Johannesma, P. I. M., & Aertsen, A. M. H. J. (1983). Reverse-correlation methods in auditory research. *Q. Rev. Biophysics*, 16, 341–414.
- Emerson, R. C., Bergen, J. R., & Adelson, E. H. (1992). Directionally selective complex cells and the computation of motion energy in cat visual cortex. *Vision Research*, 32, 203–218.
- Emerson, R. C., Citron, M. C., Vaughn, W. J., & Klein, S. A. (1987). Nonlinear directionally selective subunits in complex cells of cat striate cortex. *J. Neurophysiol.*, 58, 33–65.
- Field, D. J. (1987). Relations between the statistics of natural images and the response properties of cortical cells. *J. Opt. Soc. Am. A*, 4(12), 2379–2394.
- Harris, K. D., Csicsvari, J., Hirase, H., Dragoi, G., & Buzsáki, G. (2003). Organization of cell assemblies in the hippocampus. *Nature*, 424, 552–556.
- Heeger, D. J. (1992). Normalization of cell responses in cat striate cortex. *Visual Neuroscience*, 9, 181–198.
- Jones, J. P. & Palmer, L. A. (1987). The two-dimensional spatial structure of simple receptive fields in cat striate cortex. *Journal of Neurophysiology*, 58(6), 1187–1211.
- Kass, R. E. & Ventura, V. (2001). A spike-train probability model. *Neural Computation*, 13, 1713–1720.
- Keat, J., Reinagel, P., Reid, R. C., & Meister, M. (2001). Predicting every spike: A model for the responses of visual neurons. *Neuron*, 30, 803–817.
- Korenberg, M. J., Sakai, H. M., & Naka, K. I. (1989). Dissection of neuron network in the catfish inner retina. III: Interpretation of spike kernels. *J Neurophysiol*, 61, 1110–1120.
- Marmarelis, P. Z. & Marmarelis, V. Z. (1978). *Analysis of physiological systems: The white noise approach*. Plenum Press, London.
- Marmarelis, P. Z. & Naka, K. (1972). White-noise analysis of a neuron chain: An application of the Wiener theory. *Science*, 175, 1276–1278.
- McLean, J. & Palmer, L. A. (1989). Contribution of linear spatiotemporal receptive field structure to velocity selectivity of simple cells in area 17 of cat. *Vision Research*, 29, 675–679.
- Meister, M., Pine, J., & Baylor, D. A. (1994). Multi-neuronal signals from the retina: acquisition and analysis. *J. Neurosci. Meth.*, 51, 95–106.
- Miller, M. I. (1985). Algorithms for removing recovery-related distortion from auditory-nerve discharge patterns. *J. Acoust. Soc. Am.*, 77, 1452–1464.

- Nykamp, D. (2003). White noise analysis of coupled linear-nonlinear systems. *SIAM J. on Applied Mathematics*, 63, 1208–1230.
- Paninski, L., Fellows, M., Shoham, S., Hatsopoulos, N., & Donoghue, J. (2004). Nonlinear population models for the encoding of dynamic hand position signals in MI. *Neurocomputing*. Presented at: Computational Neuroscience, Alicante Spain, July 2003.
- Paninski, L. (2003). Convergence properties of some spike-triggered analysis techniques. *Network: Computation in Neural Systems*, 14, 437–464.
- Paninski, L., Lau, B., & Reyes, A. (2003). Noise-driven adaptation: in vitro and mathematical analysis. *Neurocomputing*, 52, 877–883. Presented at Computational Neuroscience, July 2002.
- Pillow, J. W., Paninski, L., & Simoncelli, E. P. (2004). Maximum likelihood estimation of a stochastic integrate-and-fire neural model. In *Neurocomputing*. Elsevier. Presented at Computational Neuroscience, Alicante Spain, July 2003.
- Pillow, J. W. & Simoncelli, E. P. (2003). Biases in white noise analysis due to non-Poisson spike generation. In *Neurocomputing*, volume 52-54, pages 109–115. Elsevier. Presented at Computational Neuroscience, Chicago, 21-25 July 2002.
- Reich, D., Victor, J., & Knight, B. (1998). The power ratio and the interval map: Spiking models and extracellular recordings. *The Journal of Neuroscience*, 18, 10090–10104.
- Reid, R. C. & Alonso, J. M. (1995). Specificity of monosynaptic connections from thalamus to visual cortex. *Nature*, 378(6554), 281–284.
- Rieke, F., Warland, D., de Ruyter van Steveninck, R. R., & Bialek, W. (1997). *Spikes: Exploring the Neural Code*. MIT Press, Cambridge, MA.
- Ringach, D., Hawken, M., & Shapley, R. (2002). Receptive field structure of neurons in monkey primary visual cortex revealed by stimulation with natural image sequences. *Journal of Vision*, 2, 12–24.
- Ringach, D. L., Sapiro, G., & Shapley, R. (1997). A subspace reverse-correlation technique for the study of visual neurons. *Vision Research*, 37, 2455–2464.
- Rust, N. C., Schwartz, O., Movshon, J. A., & Simoncelli, E. P. (2004). Spike-triggered characterization of excitatory and suppressive stimulus dimensions in monkey V1 directionally selective neurons. In *Neurocomputing*. Elsevier. Presented at Computational Neuroscience, Alicante Spain, July 2003.
- Sakai, H. M. & Naka, K. (1987). Signal transmission in the catfish retina. v. sensitivity and circuit. *J. Neurophysiol.*, 58, 1329–1350.

- Sakai, K. & Tanaka, S. (2000). Spatial pooling in the second-order spatial structure of cortical complex cells. *Vision Research*, 40, 855–871.
- Schwartz, O., Chichilnisky, E. J., & Simoncelli, E. P. (2002). Characterizing neural gain control using spike-triggered covariance. In Dietterich, T. G., Becker, S., & Ghahramani, Z., editors, *Adv. Neural Information Processing Systems (NIPS*01)*, volume 14, pages 269–276, Cambridge, MA. MIT Press.
- Sharpee, T., Rust, N. C., & Bialek, W. (2003). Maximizing informative dimensions: Analyzing neural responses to natural signals. In Becker, S., Thrun, S., & Obermayer, K., editors, *Adv. Neural Information Processing Systems (NIPS*02)*, volume 15, Cambridge, MA. MIT Press.
- Smyth, D., Willmore, B., Baker, G. E., Thompson, I. D., & Tolhurst, D. J. (2003). The receptive-field organization of simple cells in primary visual cortex of ferrets under natural scene stimulation. *J. Neurosci.*, 23, 4746–4759.
- Szulborski, R. G. & Palmer, L. A. (1990). The two-dimensional spatial structure of nonlinear subunits in the receptive fields of complex cells. *Vision Research*, 30, 249–254.
- Theunissen, F. E., David, S. V., Singh, N. C., Hsu, A., Vinje, W. E., & Gallant, J. L. (2001). Estimating spatio-temporal receptive fields of auditory and visual neurons from their responses to natural stimuli. *Network*, 12(3), 289–316.
- Touryan, J., Lau, B., & Dan, Y. (2002). Isolation of relevant visual features from random stimuli for cortical complex cells. *J Neurosci*, 22(24), 10811–10818.
- Tsodyks, M., Kenet, T., Grinvald, A., & Arieli, A. (1999). Linking spontaneous activity of single cortical neurons and the underlying functional architecture. *Science*, 286, 1943–1946.
- Volterra, V. (1913). *Lecons sur les Fonctions de Lignes*. Gauthier-Villars, Paris.
- Wiener, N. (1958). *Nonlinear Problems in Random Theory*. Wiley, New York.