Available online at www.sciencedirect.com
ScienceDirect
BioSystems 93 (2008) 90–100
Bio Systems
www.elsevier.com/locate/biosystems

# Decoding spike timing: The differential reverse-correlation method

Gašper Tkačik [a], Marcelo O. Magnasco [b,*]

[a] *Joseph Henry Laboratories of Physics, Lewis-Sigler Institute for Integrative Genomics, Princeton University, Princeton, NJ 08544, USA*
[b] *Laboratory of Mathematical Physics, The Rockefeller University, New York, NY 10075, USA*

Received 7 February 2008; received in revised form 21 April 2008; accepted 22 April 2008

**Abstract**

It is widely acknowledged that detailed timing of action potentials is used to encode information, for example, in auditory pathways; however, the computational tools required to analyze encoding through timing are still in their infancy. We present a simple example of encoding, based on a recent model of time-frequency analysis, in which units fire action potentials when a certain condition is met, but the *timing* of the action potential depends also on other features of the stimulus. We show that, as a result, spike-triggered averages are smoothed so much that they do not represent the true features of the encoding. Inspired by this example, we present a simple method, differential reverse correlations, that can separate an analysis of what causes a neuron to spike, and what controls its timing. We analyze with this method the leaky integrate-and-fire neuron and show the method accurately reconstructs the model's kernel.


*Keywords:* Neural networks; Spike timing; Reverse correlation

*PACS:* 87.10.+e; 05.20. −y; 89.70.+c

The receptive field of a sensory neuron is defined as an area in the stimulus parameter space in which the presence of a stimulus significantly affects the firing behavior of the neuron. For example, visual neurons may fire when a stimulus at a given position in space and with a given orientation or velocity is presented; in this case, the receptive field of that neuron will be a region in the (five-dimensional) space of position, orientation and velocity.

Methods to reconstruct receptive fields are many, and the details of the receptive field thus reconstructed are often dependent on which method is used, underscoring that the definition of a receptive field given above is rather loose. For example, in primary visual cortex there is a distinction between the "classical" receptive field (namely, the location and orientation where a single small line segment on a uniform background will, when presented by itself, elicit firing) and various "non-classical" receptive fields, i.e., regions of stimulus space where the presentation of a single segment can alter *subthreshold* electrical activity, or where the presentation of a second segment alters the firing rate elicited by a first segment located within the classical receptive field (Kapadia et al., 1995). This distinction dates back to Kuffler, who in 1953 observed the effect of lateral inhibition by illuminating the retina with two spots of light instead of only one, and consequently stated that "not only the areas from which responses can actually be set up by retinal illumination may be included in a definition of the receptive field but also all areas which show a functional connection, by an inhibitory or excitatory effect on a ganglion cell" (Gilbert et al., 2000; Kuffler, 1953).

In order to accelerate the characterization of receptive fields, various approaches are often employed in which stimuli are drawn at random from the stimulus parameter space and presented to a given neuron, whereupon all those that elicit firing are averaged together. This method is thus referred to as *spike-triggered averaging* of the stimuli (Simoncelli et al., 2004). While superficially resembling linear response theory, spike-triggered averages correlate a large input (a stimulus strong enough to elicit firing by itself) with a large output (spike/no spike), and hence the exact manner of implementation can change the reconstructed receptive field.

A particularly restrictive hidden assumption of spike-triggered methods is the existence of a link between the features of the stimulus that cause the spike to happen with the timing of the spike relative to the feature. For example, if neurons were performing linear filtering (with a continuous filter) of the stimulus followed by a spike-generating nonlinearity, then "more of" a feature that causes the neuron to spike will also cause it

* Corresponding author.
*E-mail address:* magnasco@rockefeller.edu (M.O. Magnasco).

0303-2647/$ – see front matter © 2008 Published by Elsevier Ireland Ltd.
doi:10.1016/j.biosystems.2008.04.011

to spike earlier, and spike-trigger averaging will give an appropriate reconstruction of the linear filter. At the other end of the spectrum, the neuron could perform template-matching in a discrete stimulus space – as is likely when the inputs are single spikes received from other neurons, and the spike-triggered averaging will yield interpretable results only if the output neuron responds after a fixed time lag relative to the time of a successful match. If the temporal placement of the spike depended on some *alternative* set of features in the stimulus, possibly different from the features that generated the spike in the first place, we would expect the spike-triggered averaging method to run into difficulties.

While these assumptions that link spike timing with spike generation may be plausible for primary sensory neurons, they are no longer tenable for neurons deeper in the network under any reasonable notion of a "timing code," as computing with spike timing evidently entails the transformation (computing) of the information-bearing units (spike times), and therefore, the ability to change the spike times depending upon conditions in other parts of the network. When the spike times are moved about by these other conditions, they may smear the spike-triggered averages so that they do no longer reflect accurately the conditions that trigger the spike.

In this paper we first review spike-triggered averaging and its relation to Wiener kernel analysis. We then present a simple model of an auditory neuron in which the "timing code" smearing described above really does occur. Motivated by this example we introduce an analysis method, which we call *differential reverse correlations* that could probe timing dependence on the stimuli directly. Finally, we explore in some detail a differential reverse-correlation analysis of the simplest dynamical spiking model, the leaky integrate-and-fire neuron.

## 1. Volterra And Wiener Expansions

Let us assume that an information processing system is adequately described as performing a transformation $\mathbb{F}$ that maps the input signal $x(t)$ to an output $y(t)$. If the functional $\mathbb{F}$ satisfies (rather stringent) smoothness conditions, then we can think of the following expression, called a *Volterra series*, as defining a Taylor series for $\mathbb{F}$

$$y(t) = \mathbb{F}[x](t) = H_0 + \int_0^\infty H_1(s)x(t-s)\,\mathrm{d}s + \int\int_0^\infty H_2(r,s)x(t-r)x(t-s)\,\mathrm{d}r\,\mathrm{d}s + \int\int\int_0^\infty H_3(q,r,s)x(t-q)x(t-r)x(t-s) \times \mathrm{d}q\,\mathrm{d}r\,\mathrm{d}s + \cdots \quad (1)$$

where we have assumed that the functional is time-invariant and causal. (For acausal functionals, the kernels are defined into the future as well as into the past.) All the usual problems of high-dimensional Taylor expansions afflict Volterra series and ultimately derive from our inability to estimate the radius of convergence of the series and thus to decide whether a particular function $x(t)$ is contained within it or not. A key problem here is not only to show that the remainder of the series should approach zero, but also that all the derivatives of this remainder should do so. Furthermore, it is frequently impossible to know if $\mathbb{F}$ has the smoothness conditions required for a proper Taylor expansion. Finally, estimation of the kernels is often difficult because of their non-orthogonality (Schetzen, 1989).

A more general expansion is the Wiener series (Wiener, 1958), in which smoothness is not assumed, and convergence of the series is only expected in the mean square sense, with respect to a specific family of inputs $x$, namely, Gaussian white-noise signals $\eta(t)$ with correlator $\langle\eta(t)\eta(s)\rangle = 2\epsilon^2\delta(t-s)$:

$$\mathbb{F}[\eta](t) = K_0 + \int_0^\infty K_1(s)\eta(t-s)\,\mathrm{d}s + \int\int_0^\infty K_2(r,s)\eta(t-r)\eta(t-s)\,\mathrm{d}r\,\mathrm{d}s + \int\int\int_0^\infty K_3(q,r,s)\eta(t-q)\eta(t-r)\eta(t-s) \times \mathrm{d}q\,\mathrm{d}r\,\mathrm{d}s + \cdots. \quad (2)$$

The series is now almost perfectly orthogonal, save a minor detail: the white-noise correlators $\langle\eta(t)\eta(s)\rangle = 2\epsilon^2\delta(t-s)$ cause the $n$ th order Wiener functional to contain lower order Volterra kernels which are obtained by integration of the higher order kernel with respect to pairs of lag times (Franz and Scholkopf, 2003). Because of this almost orthogonal property, estimation of the kernels becomes straightforward. The kernels are given by the time averages of products of the response with the input at various lags; for example,

$$K_3(q,r,s) = \frac{1}{\sigma^3 3!}\langle y(t)\eta(t-q)\eta(t-r)\eta(t-s)\rangle$$

where $\sigma = \mathrm{Var}\eta$ and the angular brackets represent averaging.

While the problems typical of Taylor expansions are alleviated, they have only been traded for the problems of least-squares polynomial fitting. First and foremost, just like polynomial-fit coefficients change if the range over which a function is fitted changes, the Wiener kernels may change upon changing $\sigma$, the variance of $\eta$. For sufficiently smooth functionals $\mathbb{F}$ this won't happen, such as those having only quadratic nonlinearities; a typical example for which the kernels would depend on $\sigma$ would be a functional involving thresholds through discontinuities or discontinuous derivatives.

The Wiener expansion allows a larger number of functionals to be analyzed, namely, all systems whose response to white noise has (a) finite variance and (b) finite memory (Schetzen, 1989). The latter condition excludes, for example, multistable and chaotic systems: in multistable systems, switching between stable attractors can be effected by the input, permitting arbitrarily long memory; and in chaotic systems the memory of the initial condition increases, rather than decreases, with time.

An alternative way to characterize the set of systems which can be expanded by the Wiener method was given by Victor and Knight (1979), who observed that if the input $x(t)$ is a sum of

sinusoids whose frequencies are irrationally related, the kernels provide the expansion of all combination tones. Therefore, if the Fourier spectrum of the response of the system to a finite set of sinusoids is broad-band (i.e., finite support) then the expansion does not converge and does not adequately represent the system. For example, to connect with the previous characterization, when a nonlinear oscillator is forced with a single frequency and becomes chaotic, its response becomes broadband; therefore, such a nonlinear oscillator should not be characterized by Wiener kernels.

## 2. Spike-Triggered Methods

A method in widespread use to characterize the responses of neurons is the *spike-triggered average* (Simoncelli et al., 2004; Dayan and Abbott, 2001; de Boer and Kuyper, 1968; Bryant and Segundo, 1976; Marmarelis and Naka, 1972), also widely known as the "reverse-correlation" method. In this method, the snippet of the stimulus preceding each spike is time-shifted (so that the spike corresponds to time 0) and then averaged. Any feature of the stimulus that causes the neuron to spike will then stand out as long as the spike-triggered average is different from unconditioned stimulus average. Another version, in wide use in visual neuroscience, is to present a series of images—spatial (rather than temporal) stimuli—and then average those that cause spikes; the average image causing spikes may contain information about the features that particular neuron responds to. Spike-triggered averaging is usually employed in conjunction with "white-noise" stimuli, such as temporal white noise, or random checkerboards in the spatial variant. In this case the method is transparently a variant of Wiener kernel methods.

For example, consider the spatial pattern obtained in a neurophysiology assay through the following procedure: a set of random black-and-white checkerboards is presented to a cell in the visual pathway, the checkerboards averaged with a weight representing the response of the unit, and the unconditioned average checkerboard subtracted from this. If the unit is a primary photoreceptor, then the spatial pattern thus obtained will be positive definite and reflect the point-spread function of the optical system of the eye. If the unit is a retinal ganglion cell, the spatial average will have a negative ring around a positive core, implicating inhibitory interactions. If the cell is a "simple cell" in primary visual cortex, the average will not be circularly symmetric, but will typically contain three elongated parallel and equally spaced bars, the central one being positive and the two side ones negative, representing the sensitivity of the cell to oriented line segments. Notice that there is a minuscule probability that a random checkerboard actually contains a bitmap of an oriented line segment – the method is, therefore, reconstructing the best stimulus without ever presenting one.

Because of the successful application in neuroscience of the spike-triggered average as an approximation to the abstract notion of a "receptive field," there is a widespread tendency to confuse or identify the two. As we demonstrate below, this can be a source of problems both in real and simulated neuronal systems.

In the auditory case the object being STA'ed varies. Considering the spikes in the auditory nerve, for instance, one may average the sound stimulus itself; for white-noise stimulus, this is a classic Wiener kernel analysis (Recio-Spinoso et al., 2005). The first (linear) kernel then represents an "analyzing wavelet," and for auditory nerve fibers tuned to frequencies below 3 kHz this will be an enveloped oscillation at the central frequency of the cell, with the temporal extent of the envelope being inversely proportional to the bandwidth of the fiber. This kernel shows that the fiber is phase-locked to that central frequency at the corresponding bandwidth. The second (quadratic) Wiener kernel, or the "spike-triggered covariance," is a function of two time delays – it will look like a checkerboard if the fiber is phase-locked, or contain ripples if the fiber is not phase-locked (e.g., above 5 kHz). In this latter regime, the linear kernel typically vanishes.

Deeper in the auditory pathway, however, phase-locking is either lost or hard to demonstrate. In this case researchers typically resort to characterizing the time-frequency characteristics of the neuron by computing the *spike-triggered average sonogram* elicited by white noise. What is presented in this case is indeed temporal white noise, but for the spike-triggered analysis the input is first separated in frequency and time by some standard time-frequency analysis method (such as a sonogram), and the result, not the waveform itself, is averaged through spike-triggering. This introduces the added complication that the object being averaged is not the signal, which is played as a stimulus, but rather a *nonlinear functional* of it that contains several arbitrary parameters. As we shall show below, this procedure introduces serious artifacts, in addition to the problem that one now has to specify which "kind" of receptive field one is discussing.

In addition to these problems, "receptive fields" characterized through reverse-correlation methods inherit all mathematical problems from Wiener kernels. In particular, because the Wiener expansion is not a proper Taylor expansion of the functional, but rather a least-squares fit, the kernels depend (like polynomial-fit coefficients) on the range over which they are fitted, and thus, for example, on the amplitude of the noise. It becomes operationally impossible then to distinguish an *adaptive* response of a neuron, e.g., a unit with gain-control, from such kernel variation.

A particularly egregious example was analyzed in detail by Agüera y Arcas and Fairhall (2003), who discuss one of the simplest dynamical models of action-potential generation, the leaky integrate-and-fire model (LIF). This model assumes as the most important dynamical contribution to spike generation the voltage $V$ across the neuronal membrane and as the most important parameters the capacitance $C$ of the membrane and its effective conductance $1/R$ (the "leak"), abstracting all action-potential generation as being both fast and deterministically stereotyped. The model contains a single equation for membrane voltage $V$, which is charged through a current representing the input, and discharged through the effective conductance. When the membrane potential reaches a threshold, a spike is "generated," and

the potential is reset to its resting value regardless of the previous history:

$$C\dot{V} = -\frac{V}{R} + I(t). \tag{3}$$

The LIF model is thus controlled by the time constant of the membrane $\tau = RC$; $I(t)$, the current into the cell, is the input. One would naively expect that reverse-correlation methods would be able to immediately pick up the "receptive field" of this model, i.e., an exponentially decaying kernel which, when convolved with the past history of $I(t)$, gives the current voltage. Moreover, a spike-triggered analysis should be able to reliably and directly estimate $\tau$, the timescale of the decay. As shown in Agüera y Arcas and Fairhall (2003) this expectation is wrong, and the reason illustrates well the problems with STAs and overly naive interpretations of them. Consider $I(t)$ to be a realization of white noise and label the times of the spikes as $t_i$. After the neuron spiked at a particular $t_{i-1}$, the voltage $V$ started at zero and evolved as follows up until the spike at $t_i$:

$$CV(t) = \int_{t_{n-1}}^{t} \mathrm{e}^{-s/\tau} I(t-s)\,\mathrm{d}s \qquad t_{n-1} < t < t_n$$

from where the kernel for this particular spike is indeed $\exp(-s/\tau)$, but only for $t_n - s > t_{n-1}$; it is **zero** for $t_n - s < t_{n-1}$. The spike-triggered average then will be an *average* of the functions

$$f(t,T) = \begin{cases} \exp\left(\dfrac{-t}{\tau}\right) & t < T \\ 0 & t > T \end{cases}$$

where $T$ is the *inter-spike interval* (ISI), weighted by the probability distribution of said intervals $P(T)$. The latter is evidently a function of the parameters of the input, so that the variance of the noise, a potential DC offset that may change the neuronal firing regime, etc., all enter into the distribution. As a result, the value of $\tau$ is irretrievably obscured. Furthermore, as the STA changes with changing input parameters, one could be tempted to attribute these changes to neuronal adaptation processes, which are self-evidently absent in Eq. (3).

As a solution to this problem Agüera y Arcas and Fairhall proposed the following method. Only spikes which are well-separated from the previous spike will be used for averaging; for these, instead of a simple STA, they perform a spike-triggered

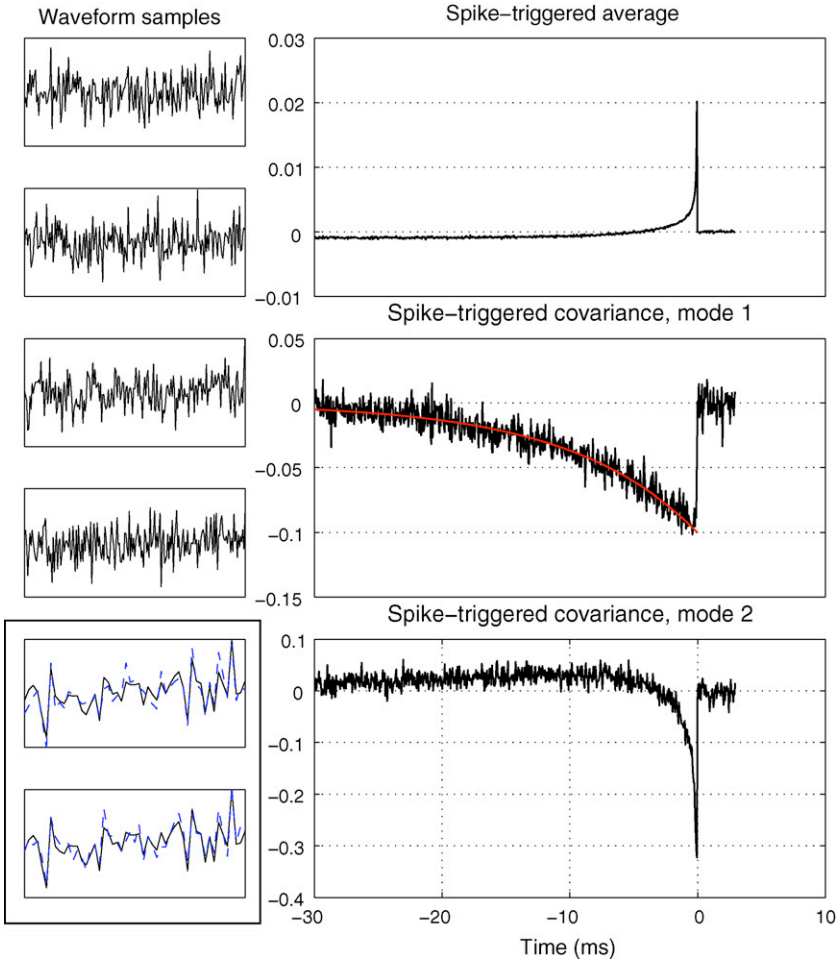


Fig. 1. Schematic diagram of the reverse-correlation methods. On the left, waveform samples that elicited the spike at $t = 0$. On the right, STA and first two eigenvectors of the spike-triggered covariance. Notice that the first eigenvector reproduces well the theoretical expectation of exponential decay with $\tau = 10$ ms.

PCA analysis, i.e., they compute the spike-triggered covariance matrix and find its eigenvalues and eigenvectors (Fig. 1). A finite number of eigenvalues are expected to contribute if the "firing criterion" is finite dimensional (for a similar analysis of a Hodgkin-Huxley model neuron, see Agüera y Arcas et al., 2003). The leading eigenvalue is, in this case, the exponential kernel with a decay time of $\tau$. There is, interestingly, a *second* nontrivial eigenvector, which represents the constraint that the voltage is supposed to reach the threshold *from below* (otherwise, the neuron would already have fired in the immediate past).

To summarize, it is risky to confuse the abstract notion of a "receptive field" with the concrete operational definition of a spike-triggered average, as the latter depends on implementation parameters unrelated to the neuron. In particular, the simplest definition of a STA applied to the simplest dynamical model of spike generation, the LIF neuron, utterly fails to measure $\tau$, the only internal parameter of the LIF model. A possible way to correct these issues is provided by Agüera y Arcas and Fairhall (2003), although it is somewhat subtle and requires collecting a large number of spikes.

## 3. A Timing-Based Auditory Model

It has recently been proposed that the auditory system could analyze rapid frequency modulations using a method in the class of reassigned spectrograms (Gardner and Magnasco, 2006). The central idea in this proposal is as follows. Any given nerve fiber in the auditory nerve carries (below central frequencies of 3 kHz in mammals and 9 kHz in birds) nervous impulses which phase-lock to frequencies close to the central frequency of the fiber. Even though a fair range of fibers will be excited by a single sine wave at moderate volumes, each fiber will be phase-locked to the stimulus and thus information about the frequency of the stimulus can be reconstructed from temporal information in any given fiber. This temporal information can be used in two ways: first, the time interval between consecutive action potentials in a single fiber gives information about the instantaneous frequency of the stimulus; and second, the time difference between "adjacent" action potentials in nearby fibers (i.e., nerve fibers with similar CFs) gives information about salient times in the stimulus.

More specifically, the mapping is

$$\omega_{\mathrm{ins}}(\omega, t) = \frac{\partial \phi}{\partial t} \tag{4}$$

$$t_{\mathrm{ins}}(\omega, t) = t - \frac{\partial \phi}{\partial \omega}, \tag{5}$$

where the phase $\phi$ is given from the Gabor transform of the input $x(t)$:

$$\phi(\omega, t) = \Im \log \int \mathrm{e}^{i\omega(t-t')}\, \mathrm{e}^{-(t-t')^2/2\sigma^2} x(t')\, \mathrm{d}t'. \tag{6}$$

We shall abstract a neuronal unit to follow these rules: our model auditory neuron will fire whenever the $\omega_{\mathrm{ins}}$ estimate is within a certain range $\omega_{\mathrm{CF}} \pm \Delta$ for some $\omega$ in the range $\omega_{\mathrm{CF}} \pm$

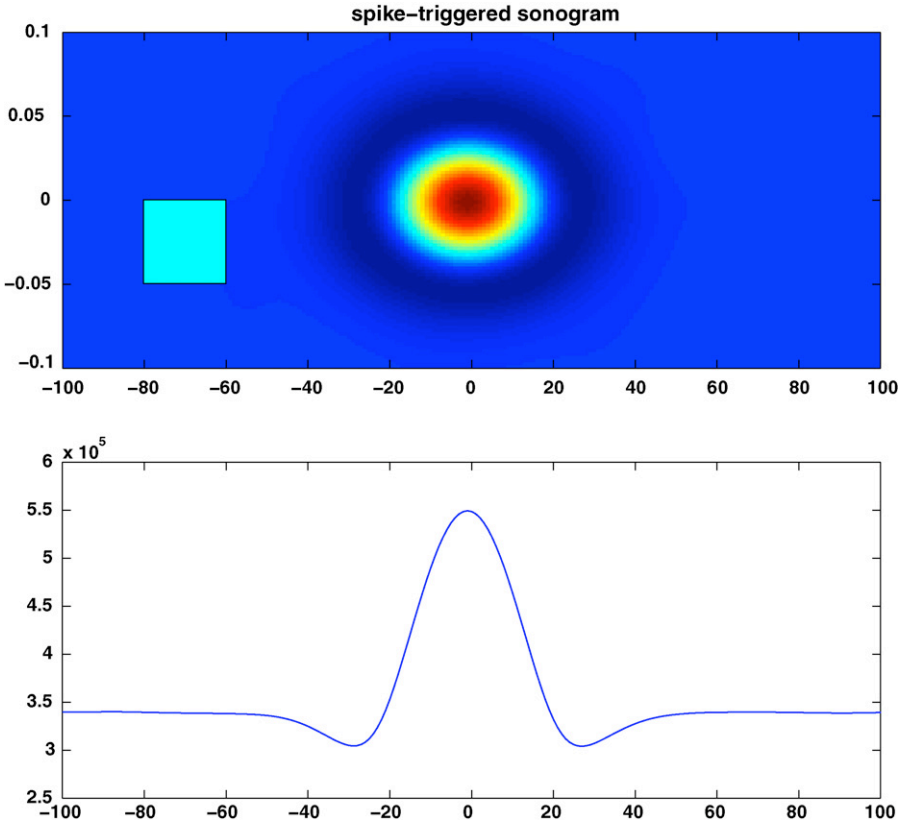


Fig. 2. Spike-triggered sonogram of the model neuron that responds by spiking whenever the instantaneous frequency estimate of the signal is within a small frequency interval around the neuron's central frequency $\omega_{\mathrm{CF}}$ (see text). Time on $x$-axis, frequency on $y$-axis in arbitrary units; the extent of the analyzing wavelet is shown as a square in the lower left corner; a cross-section at the central frequency is shown in the lower panel.

$1/\sigma$. When this happens, our neuron fires at time $t_{\text{ins}}$, i.e., the time of firing is corrected by the local group delay of the stimulus.

It was shown by Gardner and Magnasco (2006) that the above estimators lead to a time-frequency representation which is sparse, meaning that rather than appearing bloomed by the uncertainty principle, the representation displays sharp lines and curves even for spectrally dense sounds such as white noise. In view of such a result, one would expect the "receptive field" of our putative neuron to be quite sharp. Such expectation would be met with a big surprise.

First, even though the original model is phase-locked and synchronous, because the time of firing is continuously adjusted due to the time derivative of the local instantaneous frequency, the lock to the underlying oscillation is destroyed – the spike-triggered average of the stimulus simply vanishes.

Second, we could then try to reconstruct the receptive field as usual in auditory physiology, i.e., as the spike-triggered sonogram. The outcome of this calculation is shown in Fig. 2, where one may see that this calculation only succeeded in reconstructing the analyzing wavelet of the sonogram itself, rather than displaying the sharp features shown in the reassigned spectrograms computed by Gardner and Magnasco (2006). Furthermore it is to be noted that there is no inhibition in our model, yet the spike-triggered sonogram has a central "on" feature surrounded by an inhibitory "off" halo. This fake inhibitory halo is formed because the sonograms of white noise, rather than being uncorrelated like white noise itself, are correlated objects. Only if we compute the spike-triggered *reassigned sonogram*, or *instagram*, shown in Fig. 3, do we see a sharp feature comparable to the resolution observed in the reassigned spectrograms, which still retains some of the fake "inhibitory" surround.

## 4. Differential Reverse Correlations

In neural pathways where the responses are adequately described by rates, in particular in systems well approximated by the LNP model (linear–nonlinear–Poisson) (Simoncelli et al., 2004), the STA can adequately reconstruct useful features. We shall now focus on systems where the responses are "reliable," in the sense that repetition of the same stimulus elicits similar spike patterns, so that a given spike occurs during some range of time with a fair probability. In other words, individual spikes can be identified and tracked through a rasterplot. An attempt to characterize such systems through Poisson spike rates would end up with extremely high rates for the spike epochs and zero rates away from them.

Many of the caveats ascribed above to Wiener kernel methodology would simply go away if we had a means of appropriately expanding a small deviation from a given outcome due to a small deviation from a given input. To illustrate how this program can be carried out, let's consider the following situation. We have a neuronal system $\mathbb{F}$, and we have a continuous-time input $I(t)$ into this system, and the output of the system are a set of spike

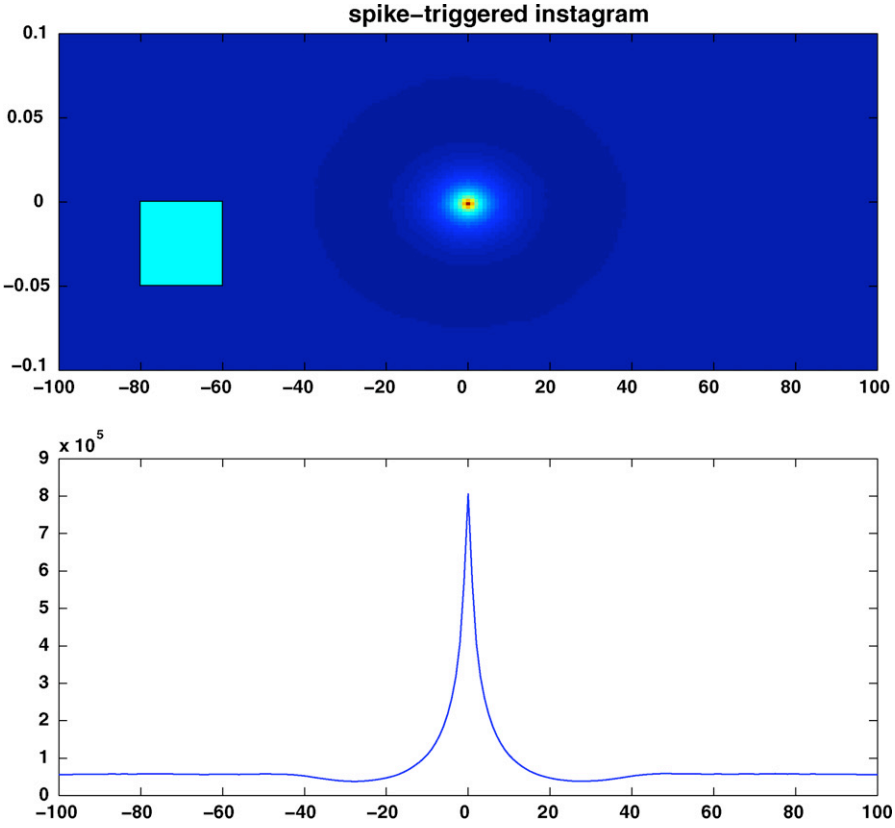


Fig. 3. Spike-triggered instagram of the model neuron descried in caption to Fig. 2. Here, spike-triggered averaging was performed in the "instantaneous" plane ($t_{\text{ins}}$, $\omega_{\text{ins}}$), resulting in a much sharper feature.

times $t_i$:

$$\{t_i\} = \mathbb{F}[I(t)]. \tag{7}$$

The input $I(t)$ is not assumed to have any structure other than the ability to elicit spikes from this neuronal system. Let us now repeatedly perturb these responses, by slightly corrupting the input $I(t)$ with noise $\eta^j(t)$:

$$\{t'_i\} = \mathbb{F}[I(t) + \eta^j(t)]. \tag{8}$$

If, as assumed, our system $\mathbb{F}$ is such that – at least for some small window around $t_i$ and across input perturbations $j$– we can identify and track the spikes homologous to the original spike at $t_i$, then we can define a time displacement $\Delta t^j$ of a spike due to the corrupting noise $\eta^j$:

$$\Delta\tau^j = t_i^j - t_i$$

and correlate these $\Delta\tau^j$ with the $\eta^j$ that caused the time displacement.

Recently, Dimitrov and Gedeon have studied the effects of "temporal spike jitter" on the estimates of the statistical properties of the response-conditional ensemble (Dimitrov and Gedeon, 2006); if this jitter really is just noise, uncorrelated with the stimulus, then the biases that it induces in, e.g., spike-triggered average and covariance should be studied and removed, as shown in that paper. In contrast, here we *purposefully* induce the jitter with the perturbation $\eta$, with the hope that any correlation between the jitter and the perturbation would reveal details about neuronal processing.

An example calculation is shown in Fig. 4 for the LIF model. The stimulus was repeated with different realizations of the small corrupting noise 10,000 times, and a rasterplot generated. Sets of homologous spikes were isolated. The figure presents the reconstructed Wiener kernel for a single one of these sets of spikes. The exponential decay of the kernel was computed to be 0.997, only 0.3% away from the true value of 1. While 10,000 repetitions of the stimulus is, admittedly, utterly impractical in an experimental context, the precision in the determination of this constant is similarly utterly unneeded. We will expand upon this calculation in more detail in the next section.

This method presents considerable potential advantages over other methods in existence. First, it is mathematically much better defined than other methods, being a Taylor expansion of a small change in response, in powers of a small change in the input. Second, we obtain a different Wiener kernel for every spike, allowing us to verify if different spikes are really always paying attention to the same input parameters or not. Third, this expansion can be done around any stimulus $I(t)$ which elicits reliable responses; these could be natural sounds or conspecific vocalizations. Finally, this method can deal with a situation (like our auditory example above) where the spike time is affected by features of the stimulus, which do not enter into the decision to spike or not.

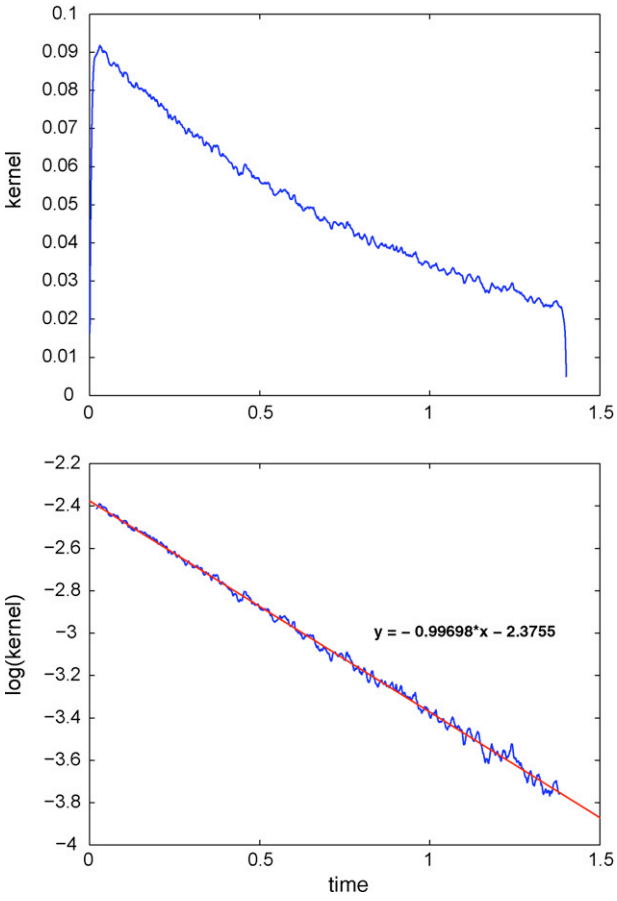


Fig. 4. Kernel reconstruction using differential reverse-correlation method, lower figure shows the reconstructed kernel on a logarithmic axis with the fitted linear function. Model neuron is leaky integrate-and-fire with $\tau = 1$ (arbitrary units); 10,000 presentations of stimulus perturbed with different instantiations of white noise, tracking changes in spike timing for a single spike across presentations. In differential reverse-correlation method the (linear) kernel is reconstructed as the linear filter that, when convolved with stimulus perturbations $\eta^j$, best predicts the changes in timing for the spike originally at $t_i$.

## 5. DRC Analysis Of The Leaky Integrate-And-Fire Neuron

We shall now carry out our analysis in more detail. Our system is, as above, the leaky integrate-and-fire neuron; we now set the membrane time constant $\tau = 10$ ms, threshold $V_c = 10$ mV; we use Euler integration with time step $\Delta t = 0.05$ ms; the system is driven with white-noise current with noise spectral density of $\sigma = \sqrt{200}\,\mu$A per time step. Fig. 5 shows that we can obtain stable reconstructions of the exponential kernel across a range of method parameters (such as spike isolation times and perturbing noise amplitudes), and that the inferred values for the decay constant $\tau$ agree with the underlying true value, as seen from Fig. 6.

The noise magnitude $\epsilon$ with which the spike-evoking waveform is perturbed would, in a real neuron, have three regimes: the regime in which the perturbing noise is too small to change the firing deterministically (because the effect is swamped by the noise in the spike generator); the regime in which the expansion (for the shift in spike timing) is linear in the noise magnitude

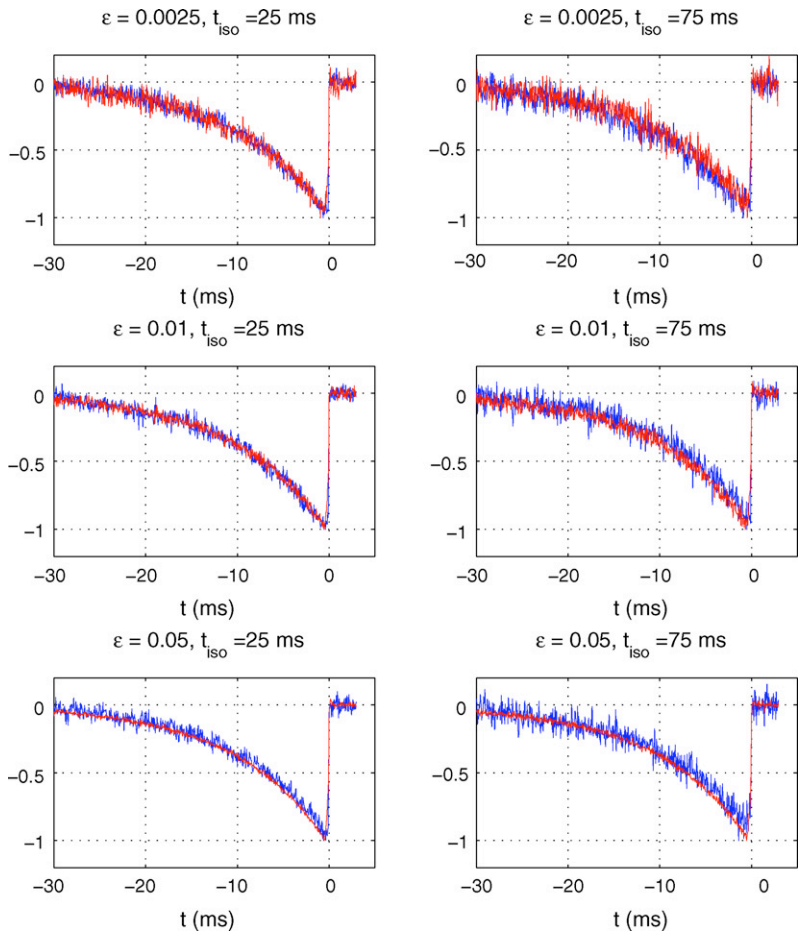


Fig. 5. Kernels recovered with the spike-triggered covariance (blue) and our method (red). This is plotted for a variety of noise perturbation strengths ($\epsilon$) and isolation times $t_{\mathrm{iso}}$. For each set of parameters, we repeatedly (500 times) generate on the order of 1000 isolated spikes (i.e., spikes that have not been preceded by another spike for $t_{\mathrm{iso}}$), by driving the system with white noise $I(t)$. For each isolated spike, we find the spike-triggering waveforms, i.e., snippets $\xi(t)$ of the stimulus $I(t)$, which we keep "frozen" and to which we repeatedly (100 times) add different realizations of white noise to get a perturbed waveform, $\xi'(t) = \xi(t) + \epsilon\eta(t)$; $\epsilon$ is the relative standard deviation of the added noise perturbation $\eta(t)$ compared to the standard deviation of white-noise stimulus used to generate spikes, $I(t)$. These perturbed waveforms result in the original spike being moved by an amount $\Delta\tau$, measured in units of integration time bins (0.05 ms). Our kernel is then the best linear filter that explains the observed $\Delta\tau$ when multiplied by the perturbations $\eta$. We plot the kernels averaged over all trials. For a variety of parameters, the spike-triggered covariance and our method agree well. Note that our kernels are less noisy, but this is because they are computed with more data (not just from the original spike triggering waveforms, but including also all the of the perturbations). Note also that for very small noise, the reconstructed kernels are more noisy because the spikes do not shift detectably (i.e., by less than 1 integration time bin). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

(Fig. 8, left panel, see also Fig. 7); and the regime where the perturbation is large enough so that a significant number of spikes are "destroyed" or "created" (Fig. 8, right panel). When we compute the kernels from the observed shifts in spike times, $\Delta\tau$, the trials in which no shift can be found, presumably because the spike disappeared, are ignored. It seems, however, that this introduces no biases. For isolation times that are too short, also, the identification of 'homologous' spikes can be problematic due to potential confusion with other nearby spikes. Note that even in the numerical simulations performed here we find an equivalent of a small- $\epsilon$ regime, because the integration is numerical and the spike shifts are truncated to integer multiples of the integration time bin, $\Delta t$ Fig. 7; this can be seen by the large apparent noise for kernel reconstructions for small $\epsilon$.

To conclude the section, we would like to draw attention to the perturbations to the original spike-triggering waveform which cause the perturbed waveform not to elicit a spike any longer. Since spiking/not-spiking are discrete events, there is a boundary in the high-dimensional space of stimuli around the spike-evoking stimulus, such that the neuron responds to stimuli within the region enclosed by such boundary with a spike and to stimuli outside the boundary with silence. This 'geometry' of decision boundaries (recently explored in the context of information transmission by Sharpee and Bialek (2007)) is probably an interesting description of neuronal behavior, and is perhaps accessible with the differential reverse-correlation method. We shall briefly discuss this issue in the next section, but leave a more thorough analysis for future work.

## 6. Revealing Internal States

In the previous section, we assumed that the stimulus eliciting the spike causes the membrane voltage to cross the threshold

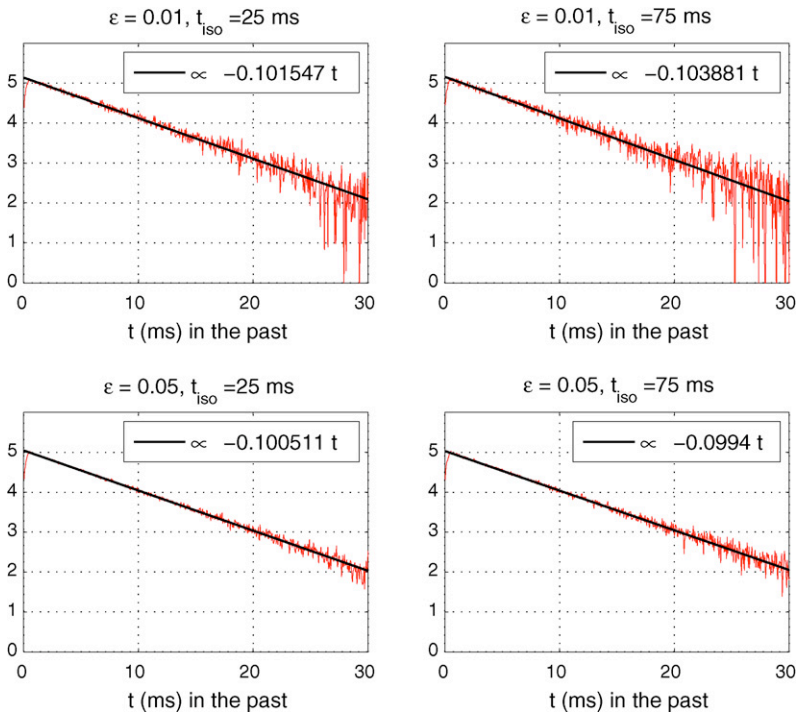


Fig. 6. Fitting the exponentials to kernels obtained with the DRC method, across the range of parameters. We expect decay with constant $1/\tau = 0.1\,\mathrm{ms}^{-1}$ and see values within 1 percent of this. The straight line (black) is fitted on all time-points except for the last 5 ms.

smoothly, in which case the result of a small amount of corrupting noise $\eta$ will be to change voltages slightly, and thus to change the time of firing inversely proportionally to the slope at this the voltage crosses the threshold.

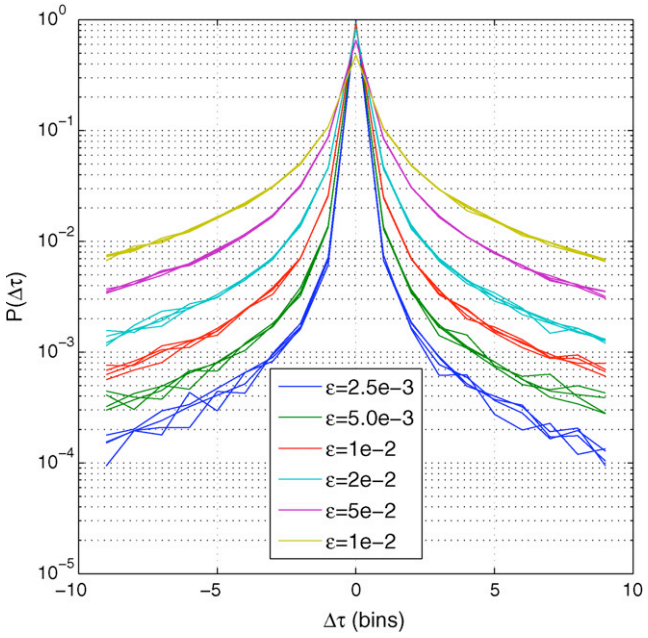


Fig. 7. Distribution of shifts in the spike timing caused by the added noise perturbation $\epsilon$ of different magnitudes (legend). Multiple lines of the same color represent the results at different isolation times, $t_{\mathrm{iso}}$, which has no noticeable effect on the plotted distribution. As expected, the fraction of spikes that shift by more than 1 integration time bin increases with the perturbation.

Some stimuli, like white noise, may not cause a smooth evolution of membrane voltage. In the leaky integrate-and-fire case, the voltage evolves as an Orstein–Uhlenbeck process, a process resembling a random walk at the small scales. Therefore, if a small fluctuating voltage is added to the membrane voltage, the resulting $\Delta\tau$ will not be distributed smoothly. This is easier to see by noting that adding a small fluctuating voltage to the voltage caused by the input alone is equivalent to keeping the latter unchanged while subtracting the fluctuating voltage from the threshold.

If the $\Delta\tau$ are small, the fluctuating voltage does not change appreciably on the scale of $\Delta\tau$ and can be considered to be just a constant, proportional to the integral of $\eta$ via the corresponding kernel. This causes a fluctuation in the threshold level, with a spike generated at the first crossing of the voltage with this threshold. Imagine the voltage describing the shape of a "mountain." An observer on the far right looks at this mountain and only segments of the mountain reachable as first crossings are visible by the observer; any descending portion, or any portion occluded behind an earlier valley is hidden. Spikes can then only occur at the visible regions.

The function describing this visibility is the cumulative maximum, i.e.,

$$\mathrm{cum}\max_x f = \max_{y \le x} f(y).$$

A given function is smaller or equal than its own cumulative maximum. The portions where it is equal to it are the "visible" portions and the regions where it is lower the hidden portions.

Given the voltage caused by the input alone, $V(t)$, its cumulative maximum $W(t)$, and the probability density $P(z)$ that the

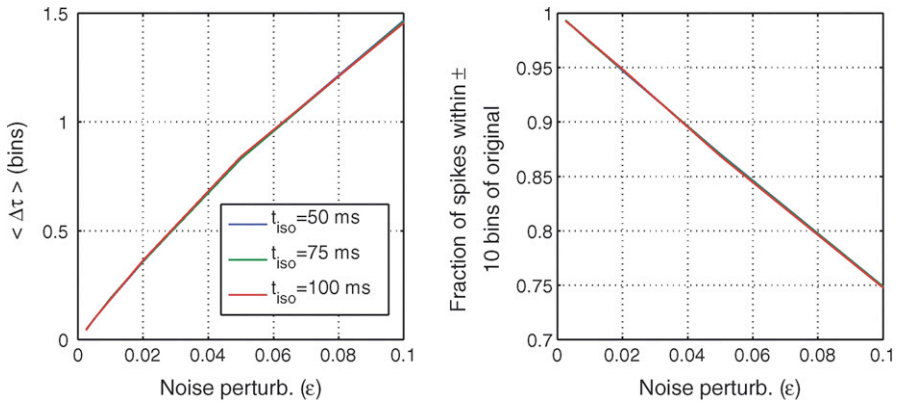


Fig. 8. Left: the average of the absolute shift caused by noise perturbation of magnitude $\epsilon$. We observe an approximately linear effect independent of the isolation time. Right: when we look for the shift in spike timing, the new spike should be located within some arbitrary (here 10 integration bins) window around the original spike. If not, the spike has either shifted by more than 10 bins (which is quite unlikely, see the distribution of time shifts $P(\Delta\tau)$ in Fig. 7), or the spike has been "destroyed" by the perturbation. Here we plot the fraction of spikes that can be located after the perturbation, as a function of the perturbing noise magnitude.

fluctuating threshold is at voltage $z$, the probability distribution of the $\Delta\tau$ is then given by

$$P(\Delta\tau) = W^{-1}(z)P(z), \tag{9}$$

from where it is seen that a determination of the probability density of $\Delta\tau$ directly allows the computation of the cumulative maximum of the transmembrane voltage in the vicinity of the threshold crossing, and thus access to important internal information of the neuronal response. This detailed information can then be more carefully correlated to the input to characterize what features of the input cause this neuron to spike.

## 7. Conclusions

The new method for receptive field analysis that we propose here, called the differential reverse-correlation method, is based upon applying small perturbations to the spike-triggering stimulus (e.g., a small amount of added white noise) and correlating small changes in spike timing with such small perturbations.

This method is applicable to neural systems and stimuli where individual spikes are elicited by the stimulus in a reliable fashion, and permits receptive field analysis on a spike-by-spike basis. Indeed, in seems reasonable to examine the causes of spiking on a single spike basis *before* such causes are hastily lumped together into a spike-triggered average, to ensure that such averaging makes sense at all.

Instead of asking "What causes a given neuron to spike," in this paper we, therefore, ask about the causes and determinants of spike timing for each spike separately. In principle, the kernels reconstructed by using the differential reverse-correlation method could differ substantially from spike to spike, just like coefficients in the Taylor expansion of a function generally differ depending on the point about which the expansion is taken. While this (potential) dependence of kernels on the underlying spike-triggering stimulus can be seen as problematic in terms of dimensionality reduction of the stimulus space, it is not necessarily a problem of the method itself; rather, it might be a reflection of the complexity of neuronal processing, which the DRC method could offer novel experimental access to. On the upside, it could turn out that even when probed on a spike-by-spike basis, there is a simplicity and consistency in the DRC kernels, offering great hopes for deeper understanding of neural computation. Last but not least, the method permits analysis of all spike-eliciting stimuli, among others especially stimuli from the natural ensemble.

While in its current version this method is still data-hungry, we expect that acceleration methods, such as templating or m-sequences, which have been applied to general Wiener kernels, can also be successfully implemented.

## Acknowledgements

This work was supported in part by NIDCD under grant R01DC007294. GT wishes to thank the Center for Studies in Physics and Biology at The Rockefeler University for its hospitality.

## References

Agüera y Arcas, B., Fairhall, A.L., 2003. What causes a Neuron to spike? Neural. Comput. 15, 1789–1807.

Agüera y Arcas, B., Fairhall, A.L., Bialek, W., 2003. Computation in a single neuron: Hodgkin and Huxley revisited. Neural. Comput. 15, 1715–1749.

Bryant, H.L., Segundo, J.P., 1976. Spike initiation by transmembrane current: a white-noise analysis. J. Physiol. 260, 279–314.

Dayan, P., Abbott, L.F., 2001. Neural Encoding I, firing rate and spike statistics. In: Theoretical Neuroscience. MIT Press, Cambridge, MA.

de Boer, R., Kuyper, P., 1968. Triggered correlation. IEEE Trans. Biomed. Eng. 15, 169–179.

Dimitrov, A.G., Gedeon, T., 2006. Effects of stimulus transformations on estimates of sensory neuron selectivity. J. Comp. Neurosci. 20, 265–283.

Franz, M.O., Scholkopf, B., Implicit Wiener Series, Technical Report No. TR-114, Max Planck Institute for Biological Cybernetics (2003).

Gardner, T.J., Magnasco, M.O., 2006. Sparse time-frequency representations. Proc. Natl. Acad. Sci. U.S.A. 103, 6094–6099.

Gilbert, C., Ito, M., Kapadia, M., Westheimer, G., 2000. Interactions between attention, context and learning in primary visual cortex. Vision Res. 40, 1217–1226.

Kapadia, M.K., Ito, M., Gilbert, C.D., Westheimer, G., 1995. Improvement in visual sensitivity by changes in local context: parallel studies in human observers and in V1 of alert monkeys. Neuron 15, 843–856.

Kuffler, S.W., 1953. Discharge patterns and functional organization of mammalian retina. J. Neurophys. 16, 37–68.

Marmarelis, P.Z., Naka, K., 1972. White-noise analysis of a neuron chain: sn application of the Wiener theory. Science 175, 1276–1278.

Recio-Spinoso, A., Temchin, A.N., van Dijk, P., Fan, Y.H., Ruggero, M.A., 2005. Wiener-kernel analysis of responses to noise of chinchilla auditory-nerve fibers. J. Neurophys. 93, 3615–3634.

Schetzen, M., 1989. The Volterra and Wiener theories of nonlinear systems. Krieger, Malabar.

Sharpee, T., Bialek, W., 2007. Neural decision boundaries for maximal information transmission. PLoS ONE 25, e646.

Simoncelli, E.P., Paninski, L., Pillow, J., Schwartz, O., 2004. Characterization of neural responses with stochastic stimuli. In: Gazzaniga, M. (Ed.), The cognitive neurosciences, 3rd Ed. MIT Press, Cambridge, MA.

Victor, J.D., Knight, B.W., 1979. Nonlinear analysis with an arbitrary stimulus ensemble. Quart. Appl. Math. 2, 113–136.

Wiener, N., 1958. Nonlinear problems in random theory. Wiley, New York.