Behavioral/Systems/Cognitive

Single Neuron Firing Properties Impact Correlation-Based Population Coding

Sungho Hong,[1] Stéphanie Ratté,[2] Steven A. Prescott,[2]* and Erik De Schutter[1,3]*

[1]Computational Neuroscience Unit, Okinawa Institute of Science and Technology, Okinawa 904-0411, Japan, [2]Department of Neurobiology and Pittsburgh Center for Pain Research, University of Pittsburgh, Pittsburgh, Pennsylvania 15213, and [3]Theoretical Neurobiology, University of Antwerp, 2610 Antwerp, Belgium

Correlated spiking has been widely observed, but its impact on neural coding remains controversial. Correlation arising from comodulation of rates across neurons has been shown to vary with the firing rates of individual neurons. This translates into rate and correlation being equivalently tuned to the stimulus; under those conditions, correlated spiking does not provide information beyond that already available from individual neuron firing rates. Such correlations are irrelevant and can reduce coding efficiency by introducing redundancy. Using simulations and experiments in rat hippocampal neurons, we show here that pairs of neurons receiving correlated input also exhibit correlations arising from precise spike-time synchronization. Contrary to rate comodulation, spike-time synchronization is unaffected by firing rate, thus enabling synchrony- and rate-based coding to operate independently. The type of output correlation depends on whether intrinsic neuron properties promote integration or coincidence detection: "ideal" integrators (with spike generation sensitive to stimulus mean) exhibit rate comodulation, whereas ideal coincidence detectors (with spike generation sensitive to stimulus variance) exhibit precise spike-time synchronization. Pyramidal neurons are sensitive to both stimulus mean and variance, and thus exhibit both types of output correlation proportioned according to which operating mode is dominant. Our results explain how different types of correlations arise based on how individual neurons generate spikes, and why spike-time synchronization and rate comodulation can encode different stimulus properties. Our results also highlight the importance of neuronal properties for population-level coding insofar as neural networks can employ different coding schemes depending on the dominant operating mode of their constituent neurons.

Introduction

Neurons in many brain areas exhibit correlated spiking but the role of those correlations remains controversial (Singer, 1993; Zohary et al., 1994; Engel et al., 1997; Gerstner et al., 1997; Shadlen and Movshon, 1999; Treisman, 1999; Salinas and Sejnowski, 2001; Palanca and DeAngelis, 2005; Averbeck et al., 2006; Schneidman et al., 2006; Wolfe et al., 2010). Noise correlations are generally thought to degrade coding efficiency (Averbeck et al., 2006) [with exceptions (Cafaro and Rieke, 2010)], but signal-dependent correlations could conceivably carry information. However, the feasibility of correlation-based coding has been called into question by the observation that output correlation varies with firing rate despite no change in input correlation (de la Rocha et al., 2007). If such a correlation–rate relationship always existed, input correlation could not be unambiguously decoded from output correlation, and transferred correlations would become meaningless (see Fig. 1). Importantly, correlations range from precise spike-time synchronization (on a millisecond timescale) to coarse rate comodulation (on a timescale up to seconds). We hypothesized that different types of correlation may differ fundamentally in how they are generated and what information they convey.

Propagation of correlated spiking depends on how individual neurons respond to correlated input (sensitivity to correlation) and whether groups of neurons respond with correlated output (transfer of correlation) such that postsynaptic neurons themselves receive correlated input (Abeles, 1991; Aertsen et al., 1996; Reyes, 2003). With respect to sensitivity to correlation, a critical factor is whether neurons operate as integrators or coincidence detectors: integrators respond to temporally dispersed inputs, whereas coincidence detectors respond selectively to rapid depolarization caused by temporally coincident (synchronous) inputs (Abeles, 1982; König et al., 1996). Operating mode (i.e., integration vs coincidence detection) reflects interplay between stimulus kinetics and spike threshold mechanism (see Results). With respect to the transfer of correlation, integrators spike repetitively at a rate proportional to their time-averaged input, whereas coincidence detectors respond to each suprathreshold input with an isolated spike—these spiking patterns are conducive to rate and temporal coding, respectively (Mainen and Sejnowski, 1995; König

Received July 20, 2011; revised Nov. 15, 2011; accepted Nov. 28, 2011.

Author contributions: S.H., S.A.P., and E.D.S. designed research; S.H., S.R., and S.A.P. performed research; S.H. and S.A.P. analyzed data; S.H., S.R., S.A.P., and E.D.S. wrote the paper.

This work was support by the Okinawa Institute of Science and Technology Promotion Corporation (S.H., E.D.S.) and NIH Grant NS076706 (S.A.P.). S.A.P. is also a Rita Allen Foundation Scholar in Pain and the 53rd Mallinckrodt Scholar.

The authors declare no competing financial interests.

*S.A.P. and E.D.S. contributed equally to this work.

Correspondence should be addressed to Sungho Hong, Computational Neuroscience Unit, Okinawa Institute of Science and Technology, 7542 Onna, Onna-son, Okinawa 904-0411, Japan. E-mail: shhong@oist.jp.

DOI:10.1523/JNEUROSCI.3735-11.2012

 0270-6474/12/321413-16$15.00/0

et al., 1996; Salinas and Sejnowski, 2001; Schreiber et al., 2004; Prescott et al., 2006; Prescott and Sejnowski, 2008; Tiesinga et al., 2008). Importantly, the temporal precision with which an individual neuron spikes should affect how synchronously a group of such neurons will spike to shared inputs, which bears directly on transfer of synchrony. Notably, neurons may be highly specialized for one operating mode, but most exhibit a context-dependent combination of modes (Maex et al., 2000; Rudolph and Destexhe, 2003; Prescott et al., 2006; Hong et al., 2008).

Using computer simulations and dynamic-clamp experiments, we identified different types of output correlation and investigated how transfer of each type of correlation depends on the intrinsic properties of cell pairs receiving correlated input. We show that pairs of "realistic" coincidence detectors exhibit rate comodulation and spike-time synchronization, whereas pairs of realistic integrators exhibit only rate comodulation. Synchrony, unlike rate comodulation, depends on spike-timing rather than rate. Rate- and synchrony-based coding are thus shown to operate independently.

Materials and Methods

Stimulus preparation

We constructed stimulus waveforms using the same procedure as de la Rocha et al. (2007). The stimulus $I(t)$ for a given trial was the linear summation of two Ornstein–Uhlenbeck (OU) processes (Uhlenbeck and Ornstein, 1930) described by the following:

$$I(t) = \mu + \sigma(\sqrt{c}\zeta_c(t) + \sqrt{1-c}\zeta(t)), \quad (1)$$

where μ and σ are the mean and SD of the stimulus, and c is the input correlation (i.e., the fraction of fluctuating input shared between neurons) (see Fig. 1*A*). The common component $\zeta_c(t)$ was instantiated once and applied to all trials, whereas the independent component $\zeta_i(t)$ was randomly updated for each trial. Each OU process was formed by the following:

$$d\zeta = -\frac{\zeta}{\tau}dt + N_\tau \xi\sqrt{dt}, \quad (2)$$

where $\xi(t)$ is Gaussian white noise with zero mean and unit variance. Sampling rates were 10 and 5 kHz for experiments and simulations, respectively. $N_\tau = (2/\tau)^{1/2}$ is a normalization constant that makes $\zeta(t)$ have unit variance. A correlation time $\tau = 5$ ms was used unless otherwise indicated.

Model neurons and simulation procedures

Two conductance-based neuron models were used. We modeled the integrator as a Morris–Lecar (ML) model with type 1 excitability (Prescott et al., 2008a) and the coincidence detector as a Hodgkin–Huxley low-sodium (HHLS) model with type 3 excitability (Lundstrom et al., 2008). Equations for the ML model are as follows:

$$C\frac{dV}{dt} = -g_L(V-E_L) - g_{Na}m_\infty(V)(V-E_{Na}) - g_K n(V-E_K) + I(t), \quad (3)$$

where

$$\frac{dn}{dt} = \frac{\phi}{\tau_n}(n_\infty - n),$$
$$m_\infty(V) = 0.5(1+\tanh(V-V_1)/V_2),$$
$$n_\infty(V) = 0.5(1+\tanh(V-V_3)/V_4),$$
$$\tau_n(V) = 1/\cosh(0.5(V-V_3)/V_4). \quad (4)$$

and $g_{Na} = 20$ mS/cm^2, $g_K = 20$ mS/cm^2, $g_L = 2$ mS/cm^2, $\phi = 0.15$, $V_1 = -1.2$ mV, $V_2 = 18$ mV, $V_3 = 0$ mV, and $V_4 = 10$ mV. The membrane capacitance per area C was 2 μF/cm^2 and the surface area was 100 μm^2.

Table 1. Stimulus conditions used for simulation of the integrator (ML), coincidence detector (HHLS), and FT models

Model	τ (ms)	Parameter	No. of samples	Minimum	Maximum
Integrator	5	c	5	0.1	0.5
		μ (pA)	12	345	375
		σ^2 (pA^2)	10	100	400
	50	c	5	0.1	0.5
		μ (pA)	10	330	400
		σ^2 (pA^2)	10	100	400
Coincidence detector	5	c	5	0.1	0.5
		μ (pA)	18	−150	200
		σ^2 (pA^2)	8	1000	2500
	50	c	5	0.1	0.5
		μ (pA)	25	−100	100
		σ^2 (pA^2)	10	400	4400
FT	5	c	5	0.1	0.5
		μ (pA)	1	0	0
		σ^2 (pA^2)	30	100	7000

Equations for the HHLS model are as follows:

$$C\frac{dV}{dt} = -g_L(V-E_L) - g_{Na}m^3h(V-E_{Na}) - g_K n^4(V-E_K) + I(t), \quad (5)$$

with activation variables m, n, and h governed by the following:

$$\tau(V)\frac{dz}{dt} = z_\infty(V) - z,$$
$$\tau_z = \frac{1}{\alpha+\beta}, \quad z_\infty = \frac{\alpha}{\alpha+\beta}, z = m, n, h. \quad (6)$$

where

$$\alpha_m = \frac{0.1(V+40)}{1-\exp[-0.1(V+40)]},$$
$$\beta_m = 4\exp[-0.556(V+65)],$$
$$\alpha_h = 0.07\exp[-0.05(V+65)],$$
$$\beta_h = \frac{1}{1+\exp[-0.1(V+35)]},$$
$$\alpha_n = \frac{0.01(V+55)}{1-\exp[-0.1(V+55)]},$$
$$\beta_n = 0.125\exp[-0.0125(V+65)]. \quad (7)$$

$g_{Na} = 41$ mS/cm^2, $g_K = 79$ mS/cm^2, $g_L = 0.3$ mS/cm^2, and the membrane capacitance $C = 1$ μF/cm^2 and the surface area was 100 μm^2.

The filter-and-threshold (FT) model consisted of three components: a linear filter to transform input to voltage, a voltage threshold, and an afterhyperpolarization (AHP). For the filter, we used the time derivative of a 15-ms-long Blackman filter, which was normalized to transform an input with variance 1 pA^2 to an output with a variance 0.1 mV^2. The threshold was 1 mV and the AHP inserted for each spike had −0.5 mV amplitude and 30 ms decay time.

All simulations in conductance-based models were performed in NEURON (Hines and Carnevale, 1997). Simulations in the FT model were performed using custom Python scripts. All code will be made available on ModelDB. Each stimulus condition (c, μ, σ^2) was repeated 2–10 times for 30 min of simulated time. All (c, μ, σ^2) combinations used are summarized in Table 1. For Figure 5, the 100–200 simulation runs conducted for each model resulted in a very large amount of data, making the calculation of correlations (from up to ~40,000 pairs) computationally challenging and the results difficult to present; therefore, we selected

only pairs having the same stimulus mean, which resulted in 1000 and 2500 pairs for the integrator and coincidence detector, respectively. This criterion was not applied to experimental data (see Fig. 7) since there were fewer trials.

Slice preparation and electrophysiology

Experimental protocols were approved by the University of Pittsburgh Institutional Animal Care and Use Committee and have been previously described (Prescott et al., 2006). Briefly, adult male Sprague Dawley rats were anesthetized with intraperitoneal injection of sodium pentobarbital (50–75 mg/kg) and perfused intracardially with ice-cold oxygenated (95% O_2 and 5% CO_2) sucrose-substituted artificial CSF (ACSF) containing the following (in mM): 252 sucrose, 2.5 KCl, 2 $CaCl_2$, 2 $MgCl_2$, 10 glucose, 26 $NaHCO_3$, 1.25 NaH_2PO_4, and 5 kynurenic acid. The brain was rapidly removed and sectioned coronally to give 300-μm-thick slices, which were kept in normal oxygenated ACSF (126 mM NaCl instead of sucrose and without kynurenic acid) at room temperature until recording.

Slices were transferred to a recording chamber constantly perfused with oxygenated (95% O_2 and 5% CO_2) ACSF heated to 31 ± 1°C. Pyramidal neurons in the CA1 region of hippocampus were recorded in the whole-cell configuration with >70% series resistance compensation using an Axopatch 200B amplifier (Molecular Devices). Membrane potential (after correction for the liquid junction potential of 9 mV) was adjusted to −70 mV through tonic current injection. Intracellular recording solution contained the following (in mM): 125 $KMeSO_4$, 5 KCl, 10 HEPES, and 2 $MgCl_2$, 4 ATP (Sigma-Aldrich), 0.4 GTP (Sigma-Aldrich), as well as 0.1% Lucifer yellow; pH was adjusted to 7.2 with KOH. Pyramidal morphology was confirmed with epifluorescence after recording. All experiments were performed in 10 μM bicuculline methiodide (Sigma-Aldrich), 10 μM CNQX (6-cyano-7-nitroquinoxaline-2,3-dione) (Sigma-Aldrich), and 40 μM D-AP-5 (D-2-amino-5-phosphonovaleric acid) (Ascent Scientific) to block background synaptic activity.

Stimuli (see above) were injected into the recorded neurons through the patch pipette. To manipulate spike threshold mechanism (see Results), an artificial "shunt" conductance ($E_{shunt} = -70$ mV, $g_{shunt} = 10$ nS) was applied via dynamic clamp implemented with a Digidata 1200A ADC/DAC board (Molecular Devices) and DYNCLAMP2 software (Pinto et al., 2001) running on a dedicated processor as previously described (Prescott et al., 2006; Prescott and De Koninck, 2009); update rate was 10 kHz. Traces were low-pass filtered at 2 kHz and digitized at 10 kHz using a CED 1401 computer interface (Cambridge Electronic Design).

Reverse correlation analysis

For simulation and experimental data, we calculated spike-triggered averages (STAs) and covariance of stimuli (STC). The STA is simply the average of the set of stimuli that led to spikes subtracted from the mean of the prior stimulus distribution (i.e., the distribution of all stimuli independent of spiking output). Here, to remove the ambiguities caused by the temporal correlations, we used the fluctuating part of the unfiltered stimulus, $I(t) = \sigma(\sqrt{c}\xi_c(t) + \sqrt{1-c}\xi_i(t))$; in other words, we used $\xi(t)$ instead of $\zeta(t)$ (see Eq. 2). Therefore, we have the following:

$$\mathrm{STA}(t) = \langle I(t_{spike} - t)\rangle_{spike} - \langle I\rangle_{prior}. \quad (8)$$

The time window for the STA was 200 ms before each spike, which captured most of the STA power. In a similar way, the STC and spike-triggered correlation of stimuli (STCor) $Q(t,t')$ are given by Bialek and de Ruyter Van Steveninck (1988) as follows:

$$\begin{aligned}\mathrm{STC}(t,t') &= \mathrm{Cov}_{spike} - \mathrm{Cov}_{prior}\\ &= \langle\{I(t_{spike} - t) - \mathrm{STA}(t)\}\\ &\qquad \{I(t_{spike} - t') - \mathrm{STA}(t')\}\rangle_{spike} - \mathrm{Cov}_{prior}\\ &= \langle I(t_{spike} - t)I(t_{spike} - t')\rangle_{spike} - \mathrm{STA}(t)\mathrm{STA}(t')\\ &\qquad - \langle I(-t)I(-t')\rangle_{prior'}\end{aligned}$$

$$Q(t,t') = \mathrm{STA}(t)\mathrm{STA}(t') + \mathrm{STC}(t,t'), \quad (9)$$

where Cov_{spike} and Cov_{prior} are the covariance matrices of the spike-triggered and prior stimuli, respectively. The STA and STCor were used for predicting a cross-correlogram (CCG) in the first- and second-order by Equations 29 and 31, respectively (see below, Predicting CCGs from reverse correlation analysis). Predicted spike train covariance and correlation were computed from the CCG in the same way as the measured CCGs (see below).

Calculation of the measured CCG and correlation

We computed the CCGs of each spike train as follows. We started by building a spike train from the spike times with a $\Delta t = 1$ ms time bin and computed the CCGs via the time-averaged unbiased empirical correlation function (Perkel et al., 1967): when a spike train for neuron i and kth repetition is $y_{i,k}(t)$, the cross-correlogram is given by the following:

$$\mathrm{CC}(\tau) = \mathrm{CCF}(\tau) - \mathrm{CCF}_{shuffle}(\tau),$$

$$\begin{aligned}\mathrm{CCF}(\tau) &= \langle y_1(0)y_2(\tau)\rangle\\ &= \frac{1}{N_{repeat}\Delta t(L - |\tau|)}\sum_{k=1}^{N_{repeat}}\int_0^{L-\tau} y_{1,k}(s)y_{2,k}(s+\tau)ds,\end{aligned}$$

$$\begin{aligned}\mathrm{CCF}_{shuffle}(\tau) &= \langle y_1(0)y_2(\tau)\rangle_{shuffle}\\ &= \frac{1}{N_{repeat}\Delta t(L - |\tau|)}\sum_{k=1}^{N_{repeat}}\int_0^{L-\tau} y_{1,k}(s)y_{2,k+1}(s+\tau)ds, \quad (10)\end{aligned}$$

where N_{repeat} is the number of repetitions, and L is the length of the spike trains and $y_{i,N\mathrm{repeat}+1} = y_{i,1}$.

From Equation 10, we computed the correlation of the spike-counts with the time window of size T as follows:

$$n_{i,k}(t) = \int_0^T y_{i,k}(t+\tau)d\tau = \int_{-\infty}^{\infty} y_{i,k}(t+\tau)\Theta_T(\tau)d\tau. \quad (11)$$

Θ_T is a window function giving $\Theta_T(t) = 1$ if $0 \le t \le T$ and $\Theta_T(t) = 0$ otherwise. The shift correlator computes the covariance as follows:

$$C_T = \langle n_1(t)n_2(t)\rangle - \langle n_1(t)n_2(t)\rangle_{shuffle}, \quad (12)$$

and the correlation coefficient is given by $\rho_T = C_T/(C_1C_2)^{1/2}$ where the auto-covariance of each neuron C_i ($i = 1,2$) was calculated in the same way. Equations 10 and 12 are related as follows:

$$\langle n_1(t)n_2(t)\rangle = \int_{-\infty}^{\infty}\mathrm{CCF}(\tau)\Phi_T(\tau)d\tau$$

$$\langle n_1(t)n_2(t)\rangle_{shuffle} = \int_{-\infty}^{\infty}\mathrm{CCF}(\tau)_{shuffle}\Phi_T(\tau)d\tau, \quad (13)$$

where $\Phi_T(t)$ is a triangular window function such that $\Phi_T(t) = |T - t|$ if $-T \le t \le T$ and $\Phi_T(t) = 0$ otherwise (Bair et al., 2001). Therefore, C_T as well as ρ_T were computed from CCGs via Equation 13.

In our analysis, we always computed the full CCGs and tried to analyze their behavior both at long and short timescales. The time window size T was not as important as in de la Rocha et al. (2007); therefore, in every case, we used $T = 200$ ms and dropped T from the notations.

Statistical analysis

We calculated the variance of the covariance, Var[C], by computing the bootstrap statistics of $\langle y_1(0)y_2(\tau)\rangle_{shuffle}$ in Equation 10; at each step, we constructed a resampled cross-correlogram $\langle y_1(0)y_2(\tau)\rangle_{resample}$ by random resampling from $\langle y_1(0)y_2(\tau)\rangle_{shuffle}$ and computed $\langle n_1n_2\rangle_{resample}$ by Equation 12. We collected 400 $\langle n_1n_2\rangle_{resample}$ with which the variance

became stable. The Var[C] is taken as this resample variance and C is also recalibrated with the resample mean as follows:

$$\mathrm{Var}[C] = \mathrm{Var}[\langle n_1 n_2 \rangle_{\mathrm{resample}}],$$
$$C \rightarrow C - E[\langle n_1 n_2 \rangle_{\mathrm{resample}}]. \quad (14)$$

For experimental data in Figure 7*B–D*, t scores were calculated from C and Var[C]. However, as for the simulation data, we could arbitrarily add more repetitions to suppress Var[C] down to sufficiently low level. For the analytic prediction, Var[C] could be computed in a similar way, but it was always insignificant (<1% of the covariance) for both experimental and simulation data.

In Figures 5 and 7, we evaluated the predictive power of the predicted C in terms of the coefficient of determination, R^2, defined by the following:

$$R^2 = 1 - \frac{\sum[\text{Measured } C/c) - (\text{Predicted } C/c)]^2}{\sum[(\text{Measured } C/c) - E[(\text{Measured } C/c)]]^2}. \quad (15)$$

By definition, $R^2 = 1$ signifies the perfect match, while the prediction has no correlation with the measured C when $R^2 = 0$. In some cases, $R^2 < 0$, and this signifies that the predictions in fact diverge away from the data with a different average.

Predicting CCGs from reverse correlation analysis

Here, we derive the first- and second-order prediction presented as Equations 29–32 in Results by using the reverse correlation analysis of the linear–nonlinear (LN) cascade model (Victor and Shapley, 1980; Meister and Berry, 1999). The LN model is composed of two stages: first, the stimulus $I(t)$ was linear filtered by the relevant features of the model $\{\varepsilon_\alpha\}$ ($\alpha = 1, 2, \ldots, D$), and the probability to spike at the given time bin $[t, t + \Delta t]$, $P(\text{spike}|\mathbf{x})$ where

$$x_\alpha = \int_0^\infty \varepsilon_\alpha(t') I(t - t') dt'. \quad (16)$$

Here, as above (see Reverse correlation analysis), we use the unfiltered stimulus $I(t) = \sigma(\sqrt{c}\xi_c(t) + \sqrt{1-c}\xi_i(t))$. Since the actual injected stimulus and $I(t)$ are linearly related, STA and STCor have the same temporal correlation as the actual stimulus, and, consequently, the temporal correlation naturally shows up in the predicted CCGs; for example, compare predicted CCGs for $\tau = 5$ and 50 ms in Figure 5*B*.

We follow the same derivation as in the study by de la Rocha et al. (2007): the common noise part $\delta I = \sqrt{c}\sigma\xi_c$ will be regarded as the perturbation on top of $I_0 = I(t)|_{c=0}$ and the comodulated part of the firing rate will be determined by Wiener series expansion in δI.

First-order prediction (see Results, Eqs. 29, 30). The firing rate change induced by a small perturbation $I_0 \rightarrow I_0 + \delta I$ can be approximated as follows:

$$P(\text{spike}) \approx P(\text{spike})_{\delta I = 0}\left(1 + \frac{1}{\sigma^2}\int_0^\infty \mathrm{STA}(t')\delta I(t - t')dt'\right), \quad (17)$$

where $\delta I(t)$ has Gaussian statistics with a variance σ^2 (Rieke et al., 1999; Hong et al., 2008). STA represents the spike-triggered average of stimuli. Then the correlation function of $q(t) = P(\text{spike at } t)$ is as follows:

$$\langle (q_1(0) - \langle q_1 \rangle)(q_2(\tau) - \langle q_2 \rangle)\rangle$$
$$= \frac{\langle q_1 \rangle \langle q_2 \rangle}{\sigma_1^2 \sigma_2^2}\int_0^\infty \mathrm{STA}_1(t')\mathrm{STA}_2(t'')\langle \delta I_1(-t')\delta I_2(\tau - t'')\rangle dt' dt''$$
$$= \frac{c\langle q_1 \rangle \langle q_2 \rangle}{\sigma_1 \sigma_2}\int_0^\infty \mathrm{STA}_1(t')\mathrm{STA}_2(t' + \tau)dt', \quad (18)$$

where $\langle \delta I_1(0)\delta I_2(\tau)\rangle = c\sigma_1\sigma_2\delta(\tau)$ if $\delta I_i = \sqrt{c}\delta\xi_c$. This is essentially the same derivation of the correlation–gain relationship based on the linear response theory (de la Rocha et al., 2007; Hong et al., 2008) since the STA is the linear kernel relating the stimulus and firing rate as in Equation 17.

We now define the firing rate $\nu(t) = q(t)/\Delta t$ and the predicted cross-correlogram for each pair as follows:

$$\mathrm{CCG}_{\text{first order}}(t) = c\frac{\nu_1 \nu_2}{\sigma_1 \sigma_2}\int \mathrm{STA}_1(t')\mathrm{STA}_2(t + t')dt', \quad (19)$$

which is Equation 29 in Results. Furthermore, if we use the identity (Chialvo et al., 1997; Hong et al., 2008),

$$\sigma\frac{\partial \nu}{\partial \mu} = \frac{\nu}{\sigma}\int \mathrm{STA}(t)dt, \quad (20)$$

the firing rate covariance is given by the following:

$$C_{\text{first order}} = \int d\tau \langle (\nu_1(0) - \langle \nu_1 \rangle)(\nu_2(\tau) - \langle \nu_2 \rangle)\rangle$$
$$= c\frac{\nu_1 \nu_2}{\sigma_1 \sigma_2}\int\int \mathrm{STA}_1(t)\mathrm{STA}_2(t + t')dtdt'$$
$$= c\frac{\nu_1 \nu_2}{\sigma_1 \sigma_2}\left(\int \mathrm{STA}_1(t)dt\right)\left(\int \mathrm{STA}_2(t'')dt''\right)$$
$$= c\sigma_1\sigma_2\frac{\partial \nu_1}{\partial \mu_1}\frac{\partial \nu_2}{\partial \mu_2}. \quad (21)$$

where we changed the variables in the third line as $t'' = t + t'$. The last line is equivalent to the original relationship between the correlation and gain (see Results, Eq. 30) (de la Rocha et al., 2007; Shea-Brown et al., 2008).

Second-order prediction (see Results, Eqs. 31, 32). When we include the second-order term in Equation 17, we obtain the following (Rieke et al., 1999; Hong et al., 2008):

$$P(\text{spike}) \approx P(\text{spike})_{\delta s = 0}\left(1 + \frac{1}{\sigma^2}\int_0^\infty \mathrm{STA}(t')\delta I(t - t')dt' + \frac{1}{2!\sigma^4}\int_0^\infty\int_0^\infty Q(t_1, t_2)\delta I(t - t_1)\delta I(t - t_2)dt_1 dt_2\right), \quad (22)$$

where $Q(t_1,t_2)$ is the STCor in Equation 9. The contribution to the CCG, Equation 31 in Results, is obtained in a straightforward way, as follows:

$$\delta\mathrm{CCG}_{\text{second order}}(t) = c^2\frac{\nu_1 \nu_2}{2\sigma_1^2\sigma_2^2}\int\int Q_1(t_1,t_2)Q_2(t_1 + t, t_2 + t)dt_1 dt_2. \quad (23)$$

Note that we excluded the self-contraction to generate a proper Wiener series (Rieke et al., 1999).

In particular, when we have the same input condition, $\mu_1 = \mu_2 = \mu$ and $\sigma_1 = \sigma_2 = \sigma$, for the same type of neuron as in Figures 3 and 6, the peak height at $t = 0$ is given by the following:

$$\mathrm{CCG}_{\text{first+second order}}(0) = c\frac{\nu^2}{\sigma^2}\int \mathrm{STA}(t)^2 dt + c^2\frac{\nu^2}{2\sigma^4}\int\int Q(t_1, t_2)^2 dt_1 dt_2. \quad (24)$$

Furthermore, we consider the case when the neuron is effectively well described by a single preferred spike-evoking stimulus feature (PSESF), $\varepsilon(t)$, such as when the neuron functions almost as a pure integrator or a pure coincidence detector. In this case, $Q(t_1,t_2)$ can be written as $Q(t_1,t_2) = Y(t_1)Y(t_2)$, where $Y(t) \propto \varepsilon(t)$, and therefore

$$\delta\mathrm{CCG}_{\text{second order}}(0) = c^2\frac{\nu^2}{2\sigma^4}\int\int Y(t_1)^2 Y(t_2)^2 dt_1 dt_2 = c^2\frac{\nu^2}{2\sigma^4}\left(\int Q(t_1, t_1)dt_1\right)\left(\int Q(t_2, t_2)dt_2\right). \quad (25)$$

From $\int Q(t_1, t_1)dt_1 = \sigma^3(\partial\nu/\partial\sigma)/\nu$ (Hong et al., 2008), the predicted peak height becomes

$$\mathrm{CCG}_{\text{first+second order}}(0) = \mathrm{CCG}_{\text{first order}}(0) + \frac{1}{2}c^2\sigma^2\left(\frac{\partial\nu}{\partial\sigma}\right)^2, \quad (26)$$

which is equivalent to Equation 32 in Results. Therefore, the firing rate gain with respect to stimulus variance contributes to the CCG peak height.

Note in the case of a single PSESF that the input mean and variance sensitivity of the firing rate are related as follows (Hong et al., 2008):

$$\frac{\partial\nu}{\partial\sigma} = \frac{\sigma}{\bar{\varepsilon}^2}\cdot\frac{\partial^2\nu}{\partial\mu^2}, \quad (27)$$

where

$$\bar{\varepsilon} = \int \varepsilon(t)dt \quad (28)$$

Therefore, when $\bar{\varepsilon} \neq 0$, as in the pure integrator, the contribution of the peak to the correlation is suppressed compared with the first-order prediction, Equation 30. In the pure coincidence detector, on the other hand, $\bar{\varepsilon} = 0$, but firing rate ν is also independent of μ (see Fig. 6*B*), which can still make $\partial\nu/\partial\sigma$ finite. Therefore, the predicted correlation in the pure coincidence detector [based on Eqs. 29, 30, and $\nu(\mu) = \text{constant}$] (Barreiro et al., 2010) will be profoundly modified by the quadratic-order approximation (see Results, Simulations in a phenomenological coincidence detector model).

Results

When two or more neurons receive correlated (i.e., shared) input, their output spike trains should exhibit some correlation despite the effects of independent input (Fig. 1*A*). Spike train covariance C and the correlation coefficient ρ (i.e., C normalized by spike train variance) should, therefore, carry information about the input correlation c. However, de la Rocha et al. (2007) showed that the relationship between input and output correlation (denoted correlation susceptibility $S = \rho/c$) depends on the mean μ and variance σ^2 of the input (Fig. 1*B*). This is important because stimulus-dependent changes in S prevent one-to-one mapping between input correlation and output correlation. Correlation-based coding is straightforward if S is independent of μ and σ^2 (Fig. 1*C*, left), but it is compromised or requires a more complicated decoding scheme if S varies with μ and σ^2 (Fig. 1*C*, right).

One way of understanding why this occurs is that, for a given input correlation, output correlation will vary depending on the sensitivity (i.e., gain) of the firing rate ν of each neuron with respect to the stimulus mean μ: if stimulus fluctuations occur within a steep region of the ν–μ curve, rate will fluctuate widely in each cell and the pair will exhibit large comodulated rate fluctuations, whereas stimulus fluctuations within a shallow region of the ν–μ curve will logically drive smaller comodulated rate fluctuations (Fig. 1*D*). Consequently, ρ "inherits" the same tuning as $\nu(\mu)$, even when c is fixed; under these conditions, c cannot be unambiguously decoded from ρ without knowledge of μ. This line of reasoning triggered three concerns: (1) it applies to rate comodulation but not necessarily to spike-time synchronization, and thus it neglects one component of output correlation; (2) it implicitly assumes that input fluctuations are "noise" rather than "signal"; and (3) output rate and synchronization (i.e., correlation based on spike-time synchronization as opposed to rate comodulation) are liable to be tuned to different stimulus properties. This led us to our overall hypothesis that some cell types may encode signal-dependent fluctuations with precise spike-time synchronization and can do so independently of rate-based coding of other stimulus features.

The dependence of output rate ν on input parameters μ and σ differs fundamentally between cell types, as shown in Figure 2 for our conductance-based models: ν is principally sensitive to μ in the case of integrators, whereas it is also very sensitive to σ in the case of coincidence detectors (Higgs et al., 2006; Arsiero et al., 2007; Lundstrom et al., 2008). For simulations reported here, input was treated as a continuous stream rather than as discrete synaptic inputs (Destexhe et al., 2001); nonetheless, σ reflects coordinated fluctuations in presynaptic activity (Fellous et al., 2003), the temporal structure of which is reflected in the autocorrelation time τ (shorter τ implies more precise synchrony), while c specifies the proportion of inputs shared between two postsynaptic neurons. The first two parameters, σ and τ, affect the temporal precision of spiking in each postsynaptic neuron, while c affects correlation across neurons—all three parameters ultimately affect output synchrony.

The greater sensitivity of coincidence detectors to input synchrony relative to integrators (Fig. 2*A*,*B*) is a direct consequence of active membrane properties: activation of outward current (or inactivation of inward current) at perithreshold potentials helps ensure spike generation selectively in response to fast stimulus fluctuations (i.e., synchronous inputs), whereas perithreshold-activating inward current (or inactivating outward current) encourages repetitive spiking in response to constant or slow-changing input (Fourcaud-Trocmé et al., 2003; Svirskis et al., 2004; Higgs et al., 2006; Arsiero et al., 2007; Lundstrom et al., 2008; Prescott et al., 2008a). Differential sensitivity to input synchrony can be demonstrated most succinctly by contrasting which stimulus features preferentially elicit spikes in each cell type. We estimated the preferred spike-eliciting stimulus feature as the spike-triggered-averaged stimulus (STA) of the response of each cell to noisy input (see examples in Fig. 2*B*). The integrator exhibits a relatively broad, monophasic STA (Fig. 2*C*, left), whereas the coincidence detector exhibits a biphasic STA with a positive phase that is remarkably narrow (Fig. 2*C*, right).

The shape of the STA should, in theory, relate directly to the cross-correlation of output spiking given that the CCG is the overlap integral of the STA of each neuron, according to the first-order ap-

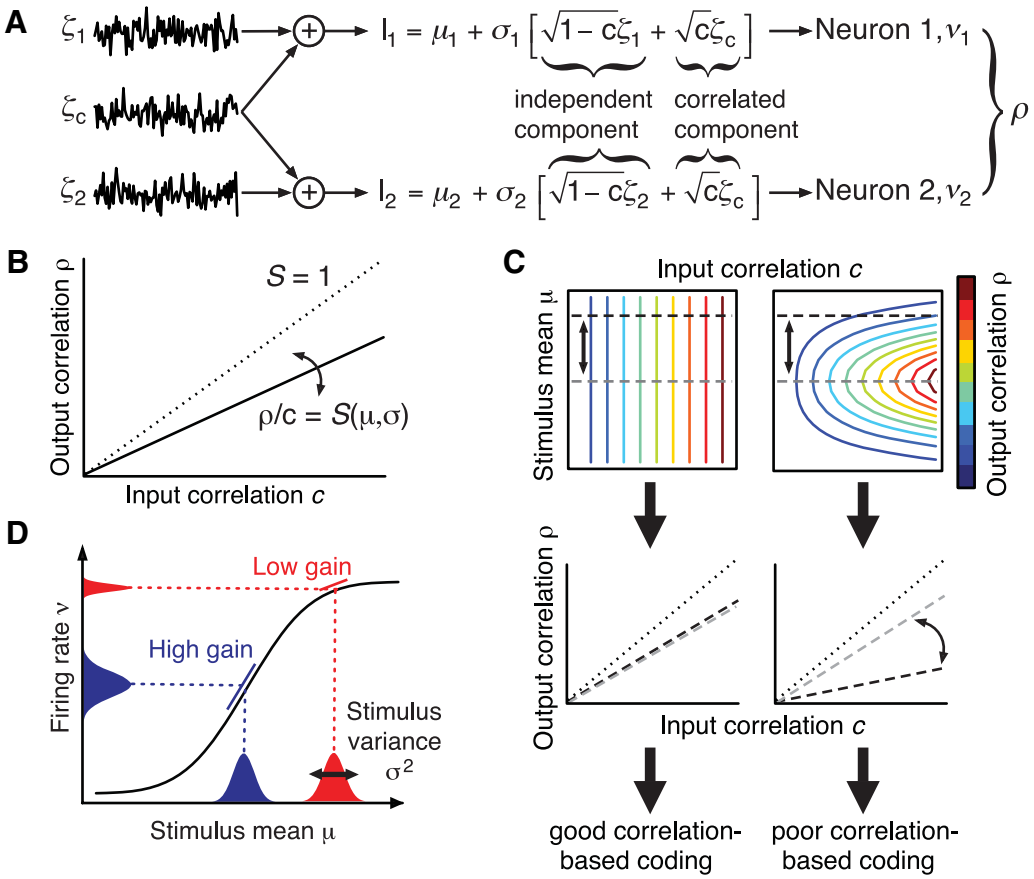


Figure 1. Relationship between input and output correlation. ***A***, Stimulation paradigm in which neurons 1 and 2 receive fluctuating input I_1, I_2, with mean μ_1, μ_2, and variance σ_1^2, σ_2^2. Fluctuating input was modeled as an Ornstein–Uhlenbeck process with $\tau = 5$ ms. Some fraction of that input is shared, or correlated, as defined by the input correlation c. Output firing rate ν_1, ν_2, and the output correlation coefficient ρ (=spike train covariance C normalized by variance) were measured. ***B***, Plotting output correlation ρ against input correlation c shows how much correlation is transferred by the pair of neurons. The slope of that curve, denoted correlation susceptibility S, is ≤ 1 but has been shown to depend on input parameters μ and σ (de la Rocha et al., 2007). ***C***, Input correlation c can only be unambiguously decoded from ρ (without knowledge of other input parameters) if S does not vary with other input parameters. The dashed curves on the bottom plots show horizontal cross-sections through 3-D plots (top) at different μ. An invariant ρ–c relationship (left) is conducive to good correlation-based coding, whereas a variable relationship (right) is not unless a more complicated decoding scheme is invoked. ***D***, If ν is tuned to μ, then fluctuations around μ will produce fluctuations in ν whose magnitude depends on $\partial\nu/\partial\mu$. If neurons 1 and 2 receive input with correlated fluctuations, ν_1 and ν_2 will be comodulated. Amplitude of ν comodulation naturally depends on $\partial\nu/\partial\mu$, rendering ρ and ν cotuned to μ. In that scenario, rate comodulation will not provide information about μ beyond that already provided by rates ν_1 and ν_2, but this does not rule out spike-time synchronization providing information about σ if input fluctuations are considered signal rather than noise.

proximation in c (see Materials and Methods) (de la Rocha et al., 2007), as follows:

$$\mathrm{CCG}_{\text{first order}}(t) \approx c\frac{\nu_1\nu_2}{\sigma_1\sigma_2}\int \mathrm{STA}_1(\tau)\cdot\mathrm{STA}_2(t+\tau)d\tau. \quad (29)$$

Hence, the broad monophasic integrator STA predicts a broad monophasic CCG for pairs of integrators (Fig. 2*D*, left), whereas the narrow biphasic coincidence detector STA predicts a narrow biphasic CCG for pairs of coincidence detectors (Fig. 2*D*, right). Using the relationship between the STA and firing rate, the covariance C of the output spike count is as follows (see Materials and Methods):

$$C = \int \mathrm{CCG}(t)dt \approx c\sigma_1\sigma_2\frac{\partial\nu_1}{\partial\mu_1}\frac{\partial\nu_2}{\partial\mu_2}, \quad (30)$$

which is consistent with the study by de la Rocha et al. (2007) because sensitivity (or gain) of ν with respect to μ is the dominant factor determining the degree of correlated spiking given a certain degree of input correlation (Fig. 1*D*). Strong dependence of the relationship between ρ and c (i.e., correlation susceptibility S) on μ is not conducive to correlation-based coding (Fig. 1*C*, right). However, the STA does not always accurately represent the preferred spike-eliciting stimulus feature (this occurs when spike generation is sensitive to higher-order stimulus statistics), which invalidates predictions based on Equation 29 in certain cases (see below).

We therefore set out to identify (1) whether and how correlation susceptibility $S(\mu,\sigma)$ differs between integrators and coincidence detectors, (2) what the consequences of such differences are for correlation-based coding, and (3) precisely why the stimulus dependence of S differs between cell types.

Simulations in conductance-based integrator and coincidence detector models

To compare the correlation susceptibility S ($=\rho/c$) of integrators and coincidence detectors, we conducted a series of numerical simulations using pairs of model neurons receiving a mix of correlated and independent input (Fig. 1*A*). For each neuron type, we varied mean μ under high- or low-variance σ^2 conditions, and evaluated the output rate ν and correlation ρ (Fig. 3). As

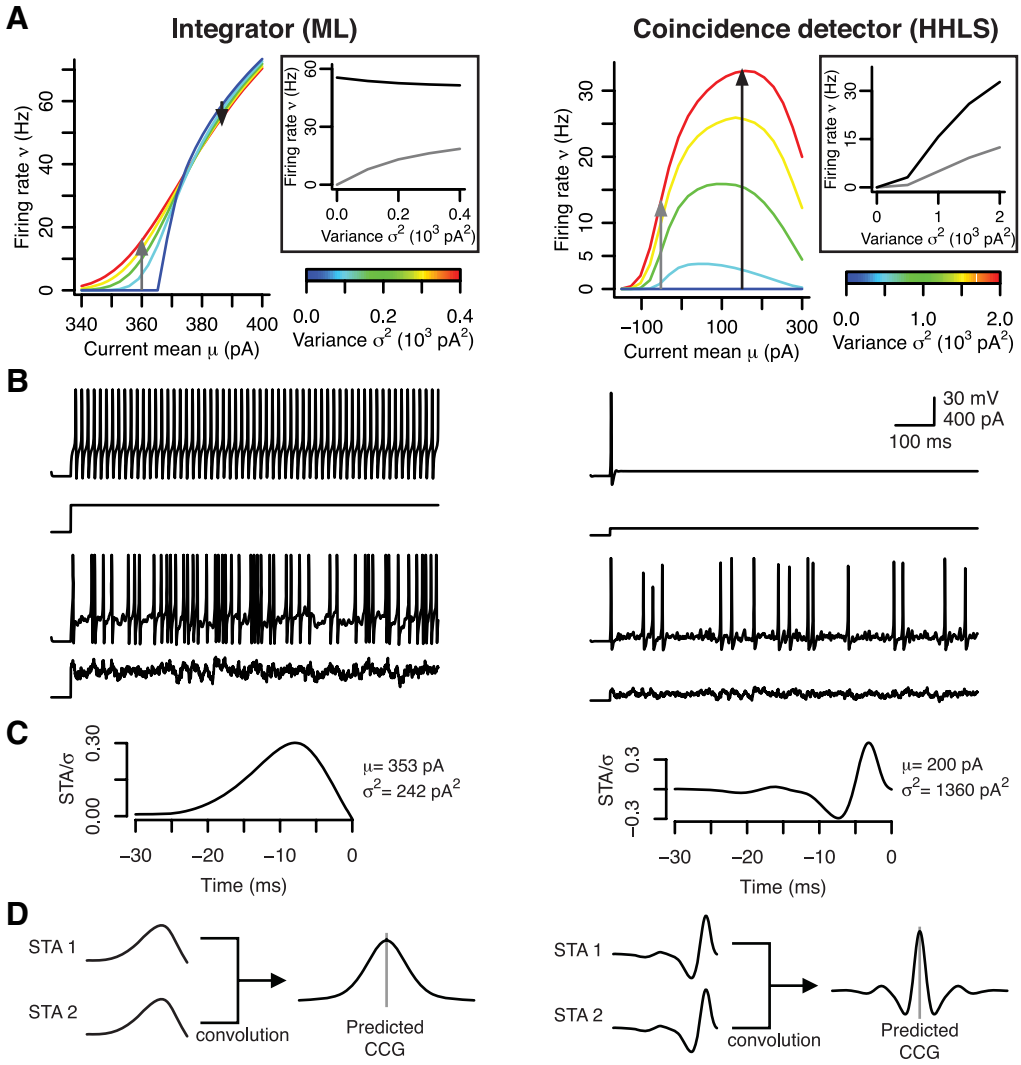


Figure 2. Integrators and coincidence detectors are sensitive to different stimulus statistics. Data here are based on simulations in conductance-based models (see Materials and Methods). ***A***, In the integrator (left), ν was sensitive to μ but was relatively insensitive to σ, whereas ν was sensitive to both μ and σ in the coincidence detector (right). The insets highlight the differential ability of each cell type to encode σ; black and gray curves correspond to arrows on ν–μ plots. Notably, firing rate variation with σ may reflect the rate of brief, suprathreshold input events rather than the (rate-encoded) magnitude of those events. ***B***, Sample traces show differential responsiveness of each cell type to constant and fluctuating input. The coincidence detector responds preferentially to fast stimulus transients because its voltage-dependent currents implement a high-pass filter. The integrator also responds to fast stimulus fluctuations, but its voltage-dependent currents encourage repetitive spiking even when input is constant. ***C***, The differential requirements for spike generation are evident from the spike-triggered-averaged response (STA) to fluctuating input. ***D***, The CCG corresponds to the overlap integral of the STA from each neuron within a pair. Therefore, pairs of integrators are predicted to exhibit broad, monophasic CCGs (left), whereas pairs of coincidence detectors are predicted to exhibit narrow, biphasic CCGs based on the typical shape of their STA (right).

illustrated in Figure 2*A*, maximal ν differs between cell types and between σ conditions, which leads to unavoidable differences in the range of ν on each panel of Figure 3. Nevertheless, the $\nu(\mu)$ curves (gray) are similarly shaped in all four conditions. If ρ simply inherited the same tuning as ν with respect to μ, then $\rho(\mu)$ curves should also be similarly shaped—clearly, they are not.

As expected for integrators, ρ/c increased to a maximum <1 as μ was increased in both the high and low σ cases (Fig. 3*A*, top). In coincidence detectors, however, a similar pattern was observed in the low σ case, but not in the high σ case, in which ρ/c decreased after reaching a peak (Fig. 3*A*, bottom). Figure 3*B* shows the same data replotted with ν on the x-axis. For coincidence detectors receiving high σ input, the correlation–rate relationship was strongly negative at high rates. The same coincidence detectors receiving low σ input exhibited little change in correlation across most of the (albeit narrow) range of firing rates, which implies that S is not significantly dependent on μ or ν in this cell type under these conditions. This property should be conducive to good correlation-based coding (see below).

To investigate how the stimulus dependence of S affects correlation-based coding by integrators and coincidence detectors, we measured ρ in response to different combinations of c and μ (Fig. 4*A*). As predicted, S was strongly dependent on μ in the case of integrators, which caused encoding of c by ρ to be ambiguous; in contrast, S was relatively unaffected by μ in the case of coincidence detectors, consistent with good correlation-based coding (compare Fig. 1*C*). Notably, Equation 29 failed to accurately predict ρ in the case of coincidence detectors (Fig. 4*A*, inset) despite having worked in the case of integrators (see below).

Small values of ρ observed among coincidence detectors stem from the shape of coincidence detector CCGs, which are narrow and biphasic, unlike integrator CCGs, which are broad and monophasic (Fig. 4*B*; compare prediction in Fig. 2*D*). Small val-

ues of ρ do not imply that synchrony transfer will fail, because downstream coincidence detectors prefer tall, narrow CCGs among their upstream neurons rather than short, broad CCGs, regardless of the absolute magnitude of ρ (S. Hong, S. A. Prescott, and E. De Schutter, unpublished observations). To summarize, correlation-based coding is viable despite small ρ as long as the relationship between ρ and c is insensitive to other stimulus features like μ. These data demonstrate the feasibility of correlation-based coding among coincidence detectors.

Next, we asked why $S(\mu,\sigma)$ differs between integrators and coincidence detectors. In all panels of Figure 3, ρ/c values predicted from Equation 29 were plotted for comparison with measured ρ/c values. The derivation of Equation 29 is based on linear response theory and requires several assumptions that are not strictly met in our simulations such as very small c. However, despite comparable stimulus parameters and identical methods used to calculate output correlation, Equation 29 predicted most of the output correlation observed for integrators but only a fraction of that observed for coincidence detectors. In the latter case, the degree of inaccuracy appeared to depend on σ. We reasoned that identifying why Equation 29 fails to accurately predict output correlation in coincidence detectors, especially under certain stimulus conditions, would help identify how correlated spiking in coincidence detectors differs from that in integrators.

To test the accuracy of the prediction, we plotted predicted covariance against measured covariance in Figure 5. If prediction by Equation 29 were consistently accurate (i.e., for all stimulus conditions), all data points would lie along the diagonal line. Figure 5*A* illustrates the consistent accuracy of the prediction for pairs of integrators versus its inconsistency for pairs of coincidence detectors. These data represent responses to a broad range of different μ and σ. The important observation is that the first-order prediction was reasonably accurate for all stimulus conditions in the case of integrators, whereas it was grossly inaccurate for many stimulus conditions in the case of coincidence detectors. Whereas Figure 3 illustrates stimulus conditions for which prediction by Equation 29 is good or bad, Figure 5 focuses on why Equation 29 sometimes fails to predict output correlation among coincidence detectors. By plotting covariance C rather than the correlation coefficient ρ, Figure 5*A* confirms that the prediction error seen for coincidence detectors is not attributable to autovariances in the spike train. Sample CCGs (Fig. 5*A*, right) reveal that the predicted CCG differs most from the measured CCG at the central peak; this is true for both integrators and coincidence detectors, but it translates into a larger discrepancy in predicting overall output correlation as the CCG gets narrower. Thus, these data qualitatively confirmed our starting predictions and identified a quantitative shortfall in the ability of Equation 29 to predict output correlation, especially in coincidence detectors.

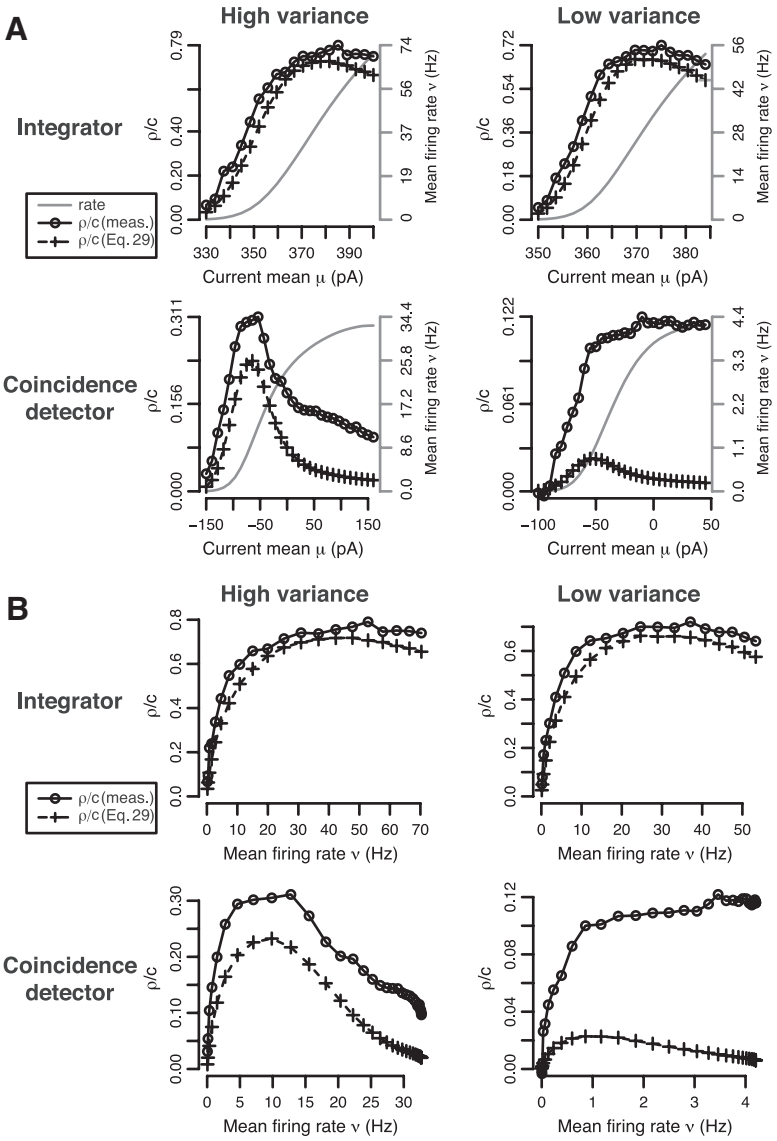


Figure 3. Output correlation is sensitive to stimulus variance in coincidence detectors, unlike in integrators, and has a variable relationship with output rate. ***A***, Using the same conductance-based models as in Figure 2, the integrator (top) and coincidence detector (bottom) were stimulated with high- and low-variance input: $\sigma_{high}^2 = 400\ pA^2$ and $\sigma_{low}^2 = 100\ pA^2$ for the integrator; $\sigma_{high}^2 = 2000\ pA^2$ and $\sigma_{low}^2 = 520\ pA^2$ for the coincidence detector. Output rate ν and correlation ρ measured from simulations (○) together with ρ predicted from Equation 29 (+) were plotted against input mean μ. In ***B***, ρ was replotted against ν to visualize the correlation–rate relationship. The range of ν (gray) varies between panels because of differences in maximal firing rate across cell types and stimulus conditions (Fig. 2*A*). For the integrator, ρ increased with ν and was quite accurately predicted by Equation 29 in both the high- and low-variance conditions. For the coincidence detector, ρ decreased for $\nu > 10$ Hz in the high-variance condition, which was only qualitatively predicted from Equation 29, and the prediction was even less accurate in the low-variance condition. These results demonstrate that output correlation in coincidence detectors is higher than predicted on the basis of rate comodulation, and is accentuated under stimulus conditions in which $\partial\nu/\partial\mu$ is low. Given that rate comodulation should be minimized under those conditions (Fig. 1*D*), correlation in excess of the prediction might be attributable to some mechanism other than rate comodulation. In every case, input correlation was $c = 0.3$.

In addition to reflecting the stimulus features to which each cell type is most sensitive, we reasoned that the width of the CCG also reflects the autocorrelation time of the correlated signal (which can be taken to reflect the temporal precision of spiking in presynaptic neurons) (see above). Therefore, we lengthened τ

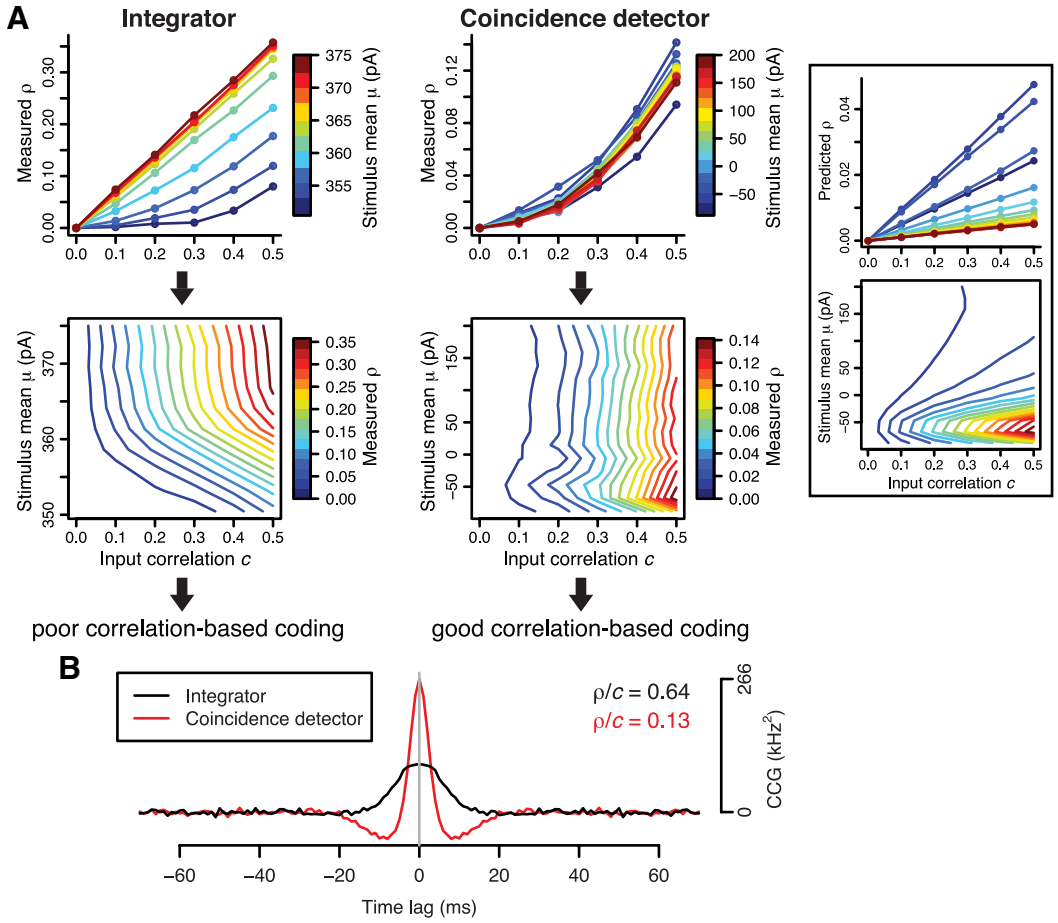


Figure 4. Correlation-based coding by integrators and coincidence detectors. ***A***, We measured output correlation ρ in response to different combinations of mean μ and input correlation c (top). Curves on 2-D plots correspond to horizontal cross-sections, at different μ, through 3-D plots (bottom; compare Fig. 1*C*). Tight clustering of ρ–c curves despite differences in μ is conducive to good correlation-based coding by coincidence detectors (right), which contrasts to the broad distribution of those curves for integrators (left). Equivalent plots for the coincidence detector based on prediction by Equation 29 (inset) shows that the first-order prediction clearly fails to account for output correlation among coincidence detectors. Noticeably, ρ is smaller among coincidence detectors than among integrators. This stems from differences in the CCG, typical examples of which are shown in ***B*** based on measurement from pairs of each cell type with the same mean firing rate of 12 Hz; $c = 0.3$. These measured CCGs confirmed CCG shapes predicted in Figure 2 *D*. Small values of ρ do not necessarily compromise synchrony transfer and correlation-based coding, whereas stimulus-dependent variability in S does (see text). In parts ***A*** and ***B***, $\sigma^2 = 100$ pA^2 for the integrator and $\sigma^2 = 1000$ pA^2 for the coincidence detector.

from 5 to 50 ms for simulations in Figure 5*B* based on the hypothesis that widening the CCG might reduce the prediction error. As expected, lengthening τ improved the prediction for both integrators and coincidence detectors but, whereas the CCG for the integrators was dramatically widened, the CCG for coincidence detectors remained narrow and the prediction still exhibited a sizeable error near the peak of the CCG. To determine whether prediction error near the peak of the CCG was sufficient to explain the discrepancy observed on the measured versus predicted C plots, we removed data within ± 2 ms of the peaks of all coincidence detector CCGs and replotted the data for $\tau = 5$ ms—this amounts to preferentially removing precisely synchronized spikes from the calculation of total output correlation. The result was a near-perfect prediction by Equation 29 (Fig. 5*C*), meaning the prediction error near the peak of narrow CCGs was sufficient to explain the failure of the prediction in coincidence detectors for all input conditions for which the original prediction was poor.

Next, we investigated (1) whether and how the original prediction might be improved so that the predicted CCG would more accurately match the measured CCG for coincidence detectors at the central peak, and (2) whether this would improve the measured versus predicted C plots (as expected given the results in Fig. 5*C*). We reasoned that if coincidence detectors are sensitive to stimulus variance (Fig. 2), then the prediction (Eq. 29) should take into account how stimulus variance affects spike generation. Therefore, we incorporated a second-order term into the prediction based on the following:

$$\delta\mathrm{CCG}_{\text{second order}}(t) = \frac{c^2 \nu_1 \nu_2}{2\sigma_1^2 \sigma_2^2} \iint Q_1(\tau_1, \tau_2) Q_2(\tau_1 + t, \tau_2 + t) d\tau_1 d\tau_2, \quad (31)$$

where $Q(t,t')$ is the spike-triggered correlation of the stimuli (STCor) (see Materials and Methods for derivation). Total output correlation should thus be predicted by the combination of Equations 29 and 31. For conditions in which input statistics and preferred spike-eliciting stimulus feature are equal, Equation 31 predicts the additional peak height as follows:

$$\delta\mathrm{CCG}_{\text{second order}}(0) = \frac{c^2}{2} \sigma^2 \left(\frac{\partial \nu}{\partial \sigma} \right)^2. \quad (32)$$

Using Equation 31, we find that the coincidence detector CCG was much better predicted at its central peak (Fig. 5*D*, right) and

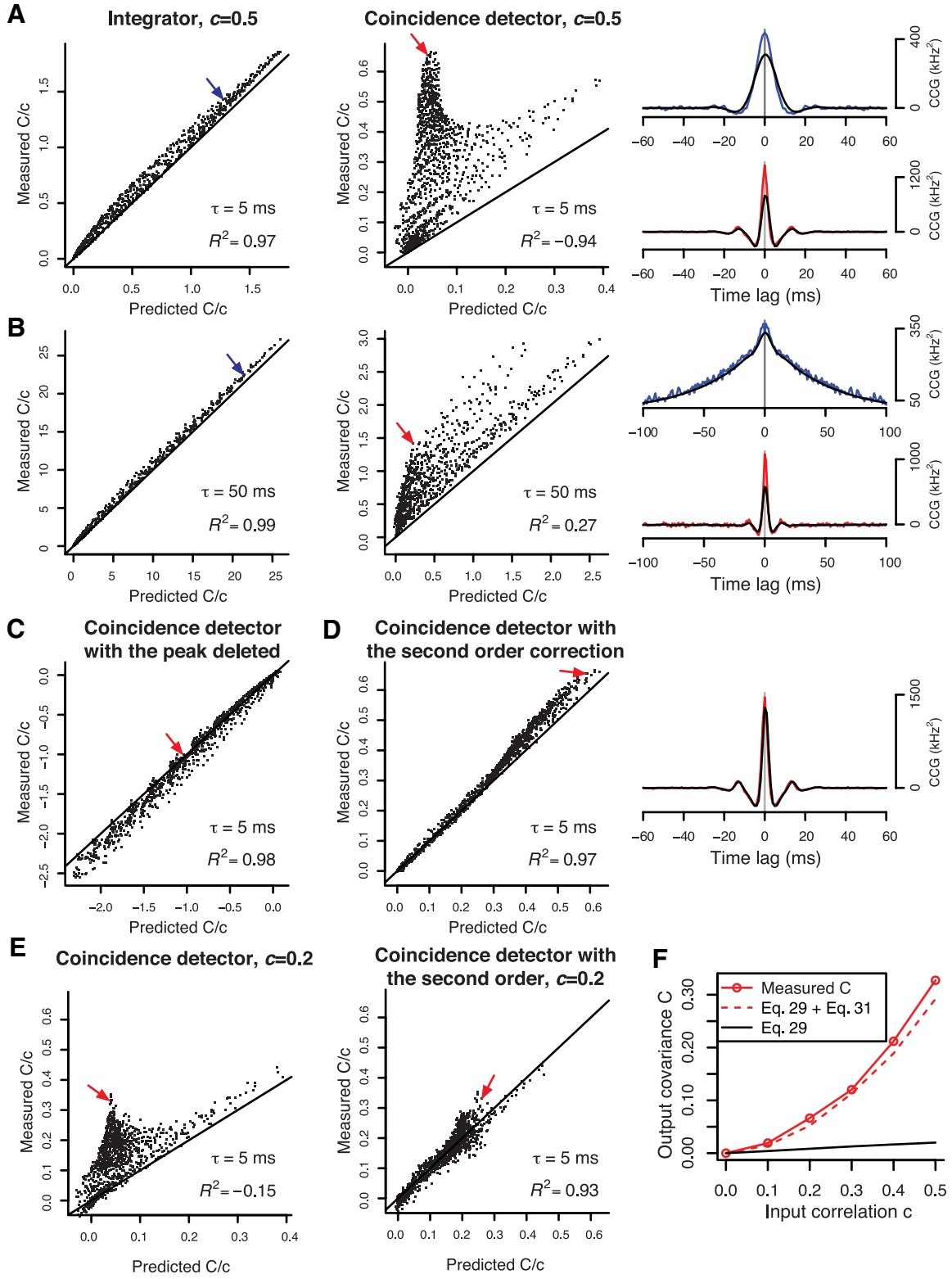


Figure 5. Sensitivity to second-order stimulus statistics explains why the first-order prediction cannot fully account for output correlation in coincidence detectors. ***A***, Measured output covariance normalized by input correlation (C/c) was plotted against predicted C/c for conductance-based integrator and coincidence detectors models receiving fluctuating input ($\tau = 5$ ms) for a broad range of μ and σ values. Prediction is based on Equation 29. The arrows point to data points for which sample CCGs are shown (right). Consistently accurate prediction (i.e., for all stimulus conditions) would be evident from points clustering around the line, which is seen for the integrator ($R^2 = 0.97$) but not for the coincidence detector ($R^2 = -0.94$). Predicted CCGs (black) deviated from measured CCGs (color) mostly around the central peak. This prediction error impacts the total output correlation for the coincidence detector more significantly than for the integrator because the CCG of the former is biphasic, which causes the predicted C to almost vanish. ***B***, Same plots as in ***A*** but with slowly fluctuating input ($\tau = 50$ ms). The integrator CCG broadened and resulted in improved prediction, but the coincidence detector CCG remained narrow and with still a sizable deviation from prediction. ***C***, Given that the first-order prediction deviated from measured correlation primarily near the central peak of the CCG, we removed data ± 2 ms around that peak and replotted measured versus predicted C/c based on data for $\tau = 5$ ms input. After removing the "excess" synchronization near the CCG peak, the first-order prediction was very accurate for coincidence detectors ($R^2 = 0.98$), meaning that Equation 29 fails specifically to explain (*Figure legend continues.*)

measured C was accurately predicted for a broad range of input conditions (Fig. 5*D*, left). Successfully modifying our quantitative prediction by incorporating Equation 31 argues that the difference in threshold mechanism is critical insofar as spike generation in integrators is preferentially sensitive to first-order stimulus statistics, whereas spike generation in coincidence detectors is also sensitive to second-order stimulus statistics (Fig. 2). On a more practical note, our ability to modify our quantitative prediction so that it works consistently (i.e., for all stimulus conditions) for integrators as well as for coincidence detectors confirms that we did not violate any fundamental assumption to the point of invalidating the prediction as a whole.

In principle, the second-order contribution (Eq. 31) can decrease much more quickly than the first-order prediction (Eq. 29) as the input correlation c becomes smaller. However, we found that the first-order prediction remained inconsistent for coincidence detectors even at relatively small values of c (Fig. 5*E*, left) and that including second-order terms significantly improved that prediction (Fig. 5*E*, right). In fact, measured C was closely matched by the second-order prediction across a broad range of c, and both were much larger than the first-order prediction except for very small c (Fig. 5*F*), proving that our result is not an artifact of large input correlations.

The nonlinear contribution to precise spike-time synchrony is not necessarily limited to second-order terms, and indeed the remaining systematic deviations from our second-order prediction in Figure 5, *D* and *E*, suggest that even higher-order terms contribute to synchrony, although that contribution is evidently quite small. Quantifying the nth order nonlinearity becomes more difficult for higher n, as the dimensionality of the nth moment of the spike-triggering stimuli rapidly increases, and was considered beyond the scope of the present study.

Simulations in a phenomenological coincidence detector model

For the conductance-based models described above, the difference in threshold mechanism between integrators and coincidence detectors derives from distinct spike initiation dynamics (Prescott et al., 2008a). Based on results from phenomenological models without "dynamics" (de la Rocha et al., 2007), one might suspect that the specific dynamics of the cell model are not vital for the correlation–rate relationship. The type of threshold and the dynamical mechanism responsible for "thresholding" are inextricably linked, but, nonetheless, to test whether the type of threshold is critical rather than the dynamics per se, we constructed a phenomenological coincidence detector model comprising a filter and threshold in which the filter is biphasic, like the STA in our conductance-based coincidence detector model (Fig. 6*A*; compare Fig. 2*C*). In this dynamics-free model, ν was completely independent of μ but varied with σ (Fig. 6*B*; compare Fig. 2*A*, right)—in effect, this model constitutes an "ideal" coincidence detector. Despite the first-order prediction (i.e., Eq. 29) that there should be no output correlation (given that $\partial\nu/\partial\mu = 0$) (Barreiro et al., 2010), pairs of ideal coincidence detectors receiving correlated input did indeed exhibit correlated spiking (Fig. 6*C*, left). Moreover, sample CCGs (Fig. 6*C*, right) were comparable in shape to those of the conductance-based coincidence detector model (compare Figs. 4, 5). Furthermore, the same prediction error was observed near the peak of the narrow CCGs when applying Equation 29, but our quantitative prediction was near-perfect when second-order terms (Eq. 31) were included. Overall, these results argue that the type of threshold used by a pair of neurons receiving correlated input will impact the correlation in their output.

To summarize results up to this point, ideal coincidence detectors (exemplified by our FT model) exhibit correlations arising solely from spike-time synchronization. In contrast, more realistic coincidence detectors (exemplified by our conductance-based HHLS model) exhibit a mixture of spike-time synchronization and rate comodulation, whereas realistic integrators (exemplified by our conductance-based ML model) exhibit mostly rate comodulation. The timescales of these two types of output correlation differ but nonetheless overlap. More importantly, the two types of output correlation exhibit fundamentally different sensitivities to firing rate: rate comodulation varies with firing rate, whereas spike-time synchronization does not. We predicted that real neurons should exhibit rate comodulation and spike-time synchronization if those neurons operate at least partially as coincidence detectors and, moreover, that the predominance of each type of output correlation will depend on the balance of operating modes.

Experiments in CA1 pyramidal neurons made to behave preferentially as integrators or coincidence detectors via manipulation of threshold mechanism by dynamic clamp

Integrators and coincidence detectors are found throughout the nervous system but tend to exhibit specializations beyond threshold mechanism. Because those additional specializations could confound our comparison of correlation susceptibility, we chose to compare correlation susceptibility in a single type of neuron whose threshold mechanism was experimentally manipulated such that the neuron behaves preferentially as an integrator or coincidence detector.

Pyramidal neurons, including those in the CA1 region of hippocampus, display the hallmarks of integrators when recorded in acute brain slices but behave more like coincidence detectors upon introduction of a virtual leak conductance by dynamic clamp (Fig. 7*A*), consistent with a predicted switch in threshold mechanism (Prescott et al., 2006, 2008b). Thus, by manipulating the membrane properties of CA1 pyramidal neurons, we were able to compare correlation susceptibility in a single cell type operating preferentially in one or the other mode—the shift in operating mode is quantitative, not absolute. In the interests of comparing experimental and simulation data [and to compare with past studies (de la Rocha et al. (2007)], we applied the same stimulation paradigm used in simulations (i.e., noisy current injection) together with the simplest dynamic-clamp manipulation capable of switching the threshold mechanism (i.e., constant leak conductance).

Compared with STAs in the integrator-mode (Fig. 7*A*, left), STAs in the coincidence detector-mode (Fig. 7*A*, right) had a

←

(*Figure legend continued.*) precisely synchronized spikes. ***D***, Equation 31 describes a second-order correction based on c^2. Prediction based on first- and second-order statistics (i.e., Eqs. 29, 31) gave dramatically improved accuracy near the peak of the coincidence detector CCG. Similar albeit smaller improvement was observed for the integrator (data not shown). According to the measured versus predicted C/c plot, inclusion of second-order terms could largely account for output correlation in the coincidence detector where the first-order prediction had failed. In ***A–D***, input correlation $c = 0.5$. ***E***, Same plots as in ***A*** and ***D*** for the coincidence detector but with input correlation $c = 0.2$. The red arrows indicate the same data marked in ***A*** and ***D***. ***F***, Success of the second-order prediction in accounting for measured output covariance across a broad range of input correlations, compared with failure of the first-order prediction, argues that output covariance among coincidence detectors is dominated by second-order terms except when input correlation c is extremely small ($\ll 0.1$). These show that the results in ***A–E*** qualitatively hold for a wide range of input correlation.

higher peak height, and were therefore steeper than integrator STAs, but the negative phase was far less prominent than in STAs from our coincidence detectors models (compare Fig. 2*C*). This is likely due to spike initiation in real neurons being influenced by more currents with gating variables spanning a broader range of timescales than were included in our minimal conductance-based models. In any case, as we would predict, the steep weakly biphasic STAs are consistent with the narrow weakly biphasic CCGs observed for pyramidal neurons in the coincidence detector-mode (see below). Most importantly for our purposes, firing rate gain with respect to μ, which is proportional to how biphasic the STA is (Eq. 20), was significantly lower in the coincidence detector-mode compared with the default integrator-mode (0.14 ± 0.041 vs 0.05 ± 0.016 Hz/pA; $p < 0.01$, t test), whereas firing rate gain with respect to σ was similar between modes (0.034 ± 0.022 vs 0.022 ± 0.011 Hz/pA), which means $\partial\nu/\partial\sigma$ was, relative to $\partial\nu/\partial\mu$, higher in the coincidence detector-mode (0.25 ± 0.16 vs 0.50 ± 0.29; $p < 0.01$, t test). This is consistent with the differential response properties reported for the conductance-based models in Figure 2*A*. Like for simulation data in Figure 5, we plotted our experimental data as measured versus predicted C based on the first-order prediction described in Equation 29 (Fig. 7*B*, top). As expected, the prediction error was much greater for the coincidence detector-mode than for the integrator-mode, which was also evident in the sample CCGs (Fig. 7*B*, bottom). Consistent with simulations (compare Figs. 4, 5, CCGs), the CCG for the coincidence detector-mode was much narrower than for the integrator-mode, and the first-order prediction deviated from the measured CCG primarily at its peak (Fig. 7*B*, bottom). As for simulation data, we hypothesized that removing a narrow region around the peak of the CCG (where the actual and predicted CCGs differ most) would dramatically improve the prediction as visualized on the measured versus predicted C plots, which indeed it did (Fig. 7*C*). Similarly, including the second-order terms described by Equation 31 improved our prediction, especially for the coincidence detector-mode (Fig. 7*D*). These experiments demonstrate that pyramidal neurons receiving shared input can exhibit output correlation in excess of that predicted from firing rate comodulation, and that this spike-time synchronization is greater when the neurons behave more like coincidence detectors.

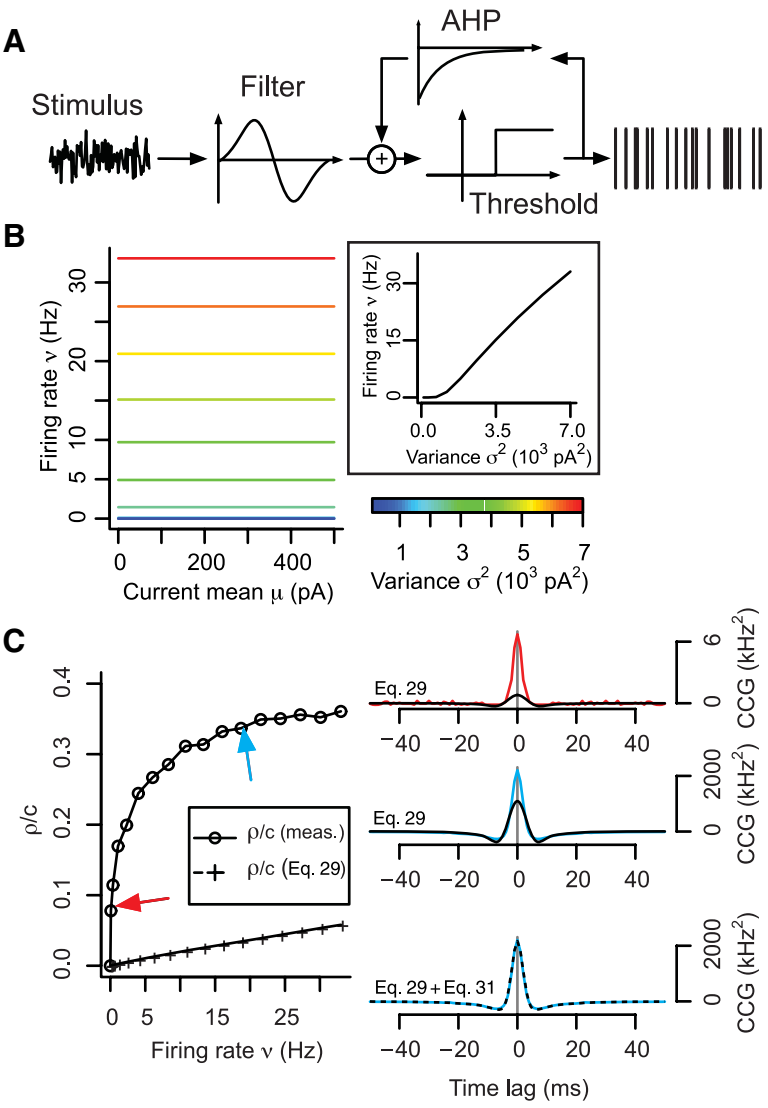


Figure 6. Phenomenological model sensitive only to stimulus variance exhibits correlated spiking due uniquely to spike-time synchronization. ***A***, FT model was constructed by coupling a biphasic filter (reminiscent of the coincidence detector STA; Fig. 2*C*) and a step-function representing threshold. ***B***, Because of its filter properties, the FT model is uniquely sensitive to stimulus fluctuations. Firing rate gain with respect to μ is zero but is positive with respect to σ. ***C***, Using the stimulation paradigm shown in Figure 1*A*, a pair of FT models were given correlated input with equal μ and σ; $c = 0.5$; $\tau = 5$ ms. Measured ρ/c values increased with ν, despite the first-order prediction (Eq. 29) giving a much smaller correlation, which vanishes as the time window size T increases (data not shown) on the basis of $\partial\nu/\partial\mu$ equaling 0. The arrows point to data for which sample CCGs are shown. Predicted CCGs (black) deviated from measured CCGs (color) in the same manner as when the first-order prediction (Eq. 29) was applied to the conductance-based coincidence detector model in Figure 5*A*. When second-order terms were included (Eq. 31), the predicted CCG was much more accurate (bottom).

Discussion

Through simulations and experiments, we have shown that pairs of neurons receiving correlated input can exhibit spiking that is correlated in different ways, and that the overall contribution of each type of output correlation depends on intrinsic cellular properties. Ideal integrators exhibit output correlations comprised entirely of rate comodulation. This makes sense given that individual integrators use rate encoding in which spiking is dictated by the mean (i.e., a first-order statistic) of the input; therefore, input can be reconstructed entirely by applying a first-order filter (i.e., the STA) to the spike train. Conversely, ideal coincidence detectors exhibit output correlations comprised entirely of spike-time synchronization (Fig. 6). However, realistic coincidence detectors exhibit output correlations comprised partly of spike-time synchronization and partly of rate comodulation (Fig. 5). This too makes sense given that individual coincidence detectors use temporal encoding in which spiking is sensitive to the

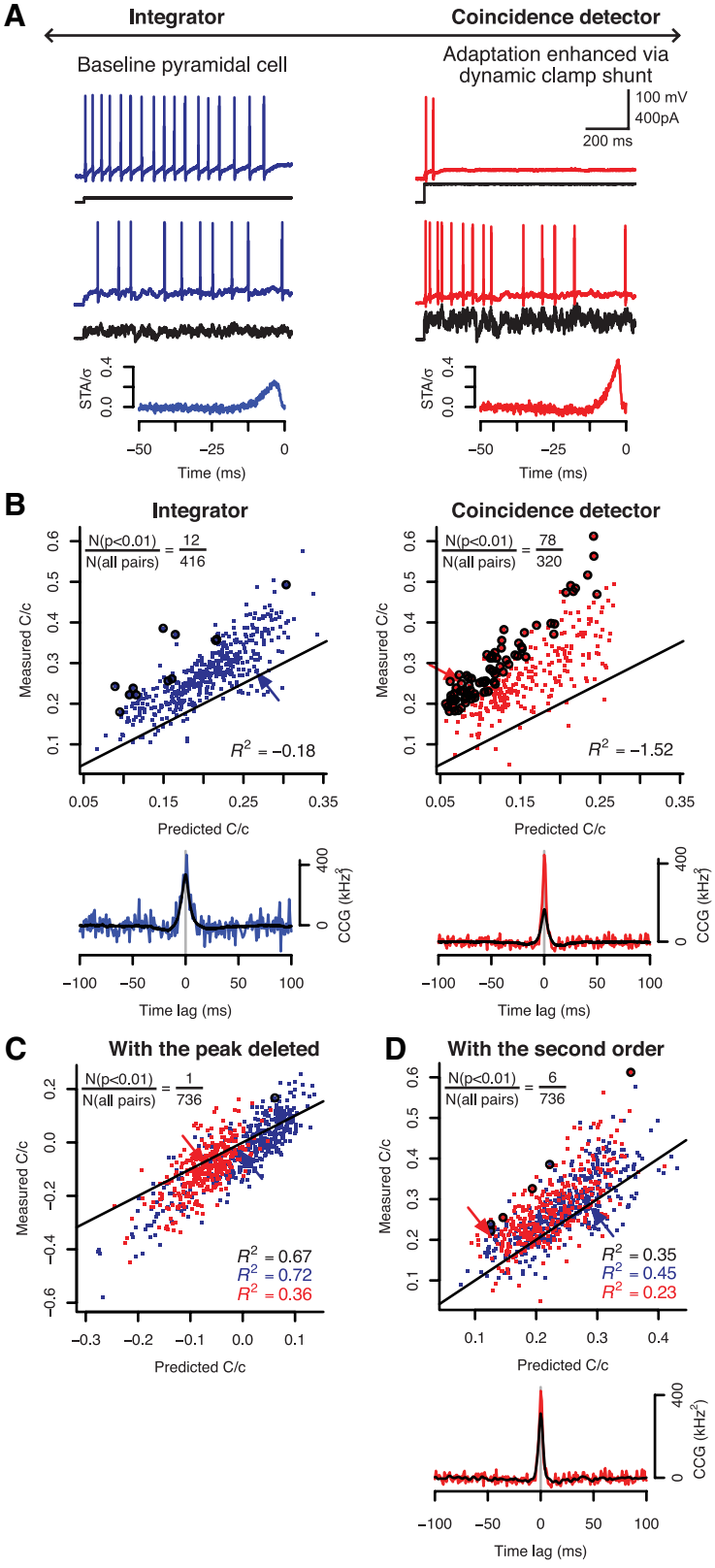


Figure 7. CA1 pyramidal neurons operating preferentially in a coincidence detector-mode exhibit greater spike-time synchronization than when operating in an integrator-mode. ***A***, Responses from a typical regular spiking pyramidal neuron to constant and fluctuating ($\tau = 5$ ms) input are shown without any manipulation (left) and with a virtual shunt (10 nS, −70 mV reversal potential) inserted by dynamic clamp (right). We used the shunt to enhance adaptation by depolarizing threshold, thereby shifting the neuron from acting preferentially as an integrator to acting preferentially as a coincidence detector, as shown previously (Prescott et al., 2006). Trials with and without dynamic clamp were interleaved to assess each neuron in both operating modes. The bottom panels show representative STA from each mode from one of the neurons comprising the pair used to sample CCGs (as indicated by arrows) in ***B–D***. ***B***, Following the same approach used for conductance-based models (Fig. 5), we plotted measured and predicted *C/c* for each "preferred" operating mode. The prediction here is based on Equation 29. The circles identify those points that deviated significantly from prediction ($p < 0.01$, *t* test). In the integrator-mode, only 12 of 416 points showed significant deviation of the measured *C/c* from prediction, while significantly more points (78 of 320) showed significant deviation in the coincidence detector-mode ($p < 0.001$, χ^2 test); in other words, the first-order prediction was less consistent for the coincidence detector-mode. The arrows point to data for which sample CCGs are shown. Predicted CCGs (black) deviated from measured CCGs (color) the same way as in Figure 5*A*. Data are from a total of 12 cells. ***C***, Measured and predicted *C/c* values for both operating modes were replotted after removing data ±2 ms around the peak in the CCGs. As in Figure 5*C*, excluding excess synchrony not predicted on the basis of rate comodulation gave a closer match between measured and predicted *C/c* values, leaving only one data point (circled) that deviated significantly from prediction ($p < 0.01$, *t* test). ***D***, Likewise, inclusion of second-order terms (i.e., Eq. 31) in our prediction gave a dramatically improved match, leaving only three data points (circled) for each operating mode with significant deviation from prediction ($p < 0.01$, *t* test). The improved prediction of spike-time synchrony afforded by inclusion of the second-order correction is also evident from the sample CCG (compare Fig. 5*D*).

variance (i.e., a second-order statistic) of the input; therefore, reconstructing the input from the spike train requires inclusion of first- and second-order filters (Theunissen and Miller, 1995). Real neurons and realistic conductance-based models are never pure integrators or coincidence detectors. We have juxtaposed the two operating modes for didactic purposes, but, ultimately, operating mode is a continuum representing the interplay between stimulus kinetics and neural dynamics. Neural dynamics—most notably spike generation—differ between neurons, can be modulated within a given neuron, and directly impact how a neuronal ensemble encodes and transmits information.

Some cell types are optimized for integration or coincidence detection; for example, many neurons early in the auditory pathway (e.g., cochlear nucleus and superior olive) are exquisite coincidence detectors (Manis and Marx, 1991; Cao et al., 2007; Mathews et al., 2010), whereas certain neurons in the entorhinal cortex are near-perfect integrators (Egorov et al., 2002). In addition to morphological specializations (Agmon-Snir et al., 1998), a broad range of ion currents can influence operating mode, the common feature being that such currents activate or inactivate at voltages near threshold, or that they impact the gating of other perithreshold currents, for example, by shifting voltage threshold (Prescott et al., 2006,

2008a). Whether pyramidal neurons in the neocortex and hippocampus function as integrators or coincidence detectors has been controversial (Abeles, 1982; Softky and Koch, 1993; Shadlen and Newsome, 1998). The answer depends not only on intrinsic cell properties but also on stimulus conditions (Rudolph and Destexhe, 2003) and other external factors like background synaptic activity that influence membrane conductance (Destexhe et al., 2003). Shunting and adaptation, especially in combination, can encourage coincidence detection in pyramidal neurons that might otherwise behave preferentially as integrators (Prescott et al., 2006, 2008b). The adaptation current I_M has been shown to encourage coincidence detection in hippocampal and neocortical pyramidal neurons (Hu et al., 2007; Guan et al., 2011) and is itself subject to endogenous modulators and to several drugs (for review, see Brown and Passmore, 2009). This supports the view that pyramidal neurons (and presumably other cell types through adaptation by I_M and via other mechanisms) can shift between operating modes. By not operating at one or the other extreme, pyramidal neurons can flexibly use rate- and/or synchrony-based coding (see below).

The coexistence of rate- and synchrony-based coding is consistent with recent modeling work showing that differently encoded information can be simultaneously propagated through feedforward networks depending on network properties (Kremkow et al., 2010; Kumar et al., 2010). Our results emphasize the importance of cellular properties for exactly the same issues. Network and cellular properties can be intimately related (see above regarding synaptic input and membrane conductance) and may interact nonlinearly such that forms of modulation, which individually have weak effects, combine to produce powerful gating mechanisms that switch a network between propagating synchronous or asynchronous activity (see below). But it is not necessarily the case that synchronous and asynchronous spikes act independently; for example, in spiny stellate cells receiving thalamocortical input via a limited number of weak synapses, asynchronous background cortical input may be crucial for setting the membrane potential of stellate cells close enough to spike threshold that synchronous thalamic inputs can elicit spikes (Douglas and Martin, 2007; da Costa and Martin, 2011). In that example, asynchronous input may facilitate the propagation of synchronous spiking and, furthermore, might act as a continuously variable gain-control mechanism rather than as a simple on–off switch.

An important related issue is correlation between excitatory and inhibitory input. This has been proposed as a mechanism to regulate information transmission (Salinas and Sejnowski, 2001; Kremkow et al., 2010), amplify external inputs (Murphy and Miller, 2009), and enhance response fidelity (Wehr and Zador, 2003; Cafaro and Rieke, 2010). Such correlations may exist with delays of only a few milliseconds, likely because of feedforward inhibition ensuring a narrow integration window (Wehr and Zador, 2003; Higley and Contreras, 2006). Under those conditions, presynaptic spikes synchronized with millisecond precision are required to reliably evoke responses in postsynaptic neurons.

Information encoded by rate or synchrony must be reliably transmitted (when allowed by gating mechanisms) and eventually decoded. Unambiguous decoding requires independent rate- and correlation-based coding (Fig. 1). Recent findings (de la Rocha et al., 2007) have cast doubt on this independence. We concur that a correlation–rate relationship compromises correlation-based coding based on rate comodulation, but we demonstrate here that output correlation and rate can be correlated without interfering with synchrony-based coding. Counting spikes neglects the information carried by spike timing. The absolute timing of spikes in a single presynaptic cell has little impact on postsynaptic activation, but the relative timing of spikes across multiple presynaptic cells is crucial if the postsynaptic cell operates as a coincidence detector; by extension, the absolute timing of synchronized volleys likely carries information about the original stimulus.

It follows that integrators exhibit the correlation–rate relationship described by de la Rocha et al. (2007), but that same relationship, although it can be observed in coincidence detectors (Barreiro et al., 2010), applies to only a fraction (i.e., to only one component) of their output correlation. Notably, de la Rocha et al. used computer models in which spike generation depended on mean input (i.e., integration) and the 1-s-long stimuli used in their experiments were sufficiently short that slow adaptation was minimal, thus favoring spike generation based on mean input (Prescott and Sejnowski, 2008). In addition to testing "integrator" models, we tested conductance-based and phenomenological models in which spike generation depends on input variance (i.e., coincidence detector models). Our experiments were conducted with 300-s-long stimuli to favor adaptation, our rationale being that neurons *in vivo* exist in a chronically depolarized and shunted state because of synaptic bombardment (Destexhe et al., 2003). By testing a single cell type under different "virtual network conditions" rather than testing different cell types representing each extreme of the continuum between integration and coincidence detection, our results show that the neuronal operating mode can be shifted along that continuum and that this shift in operating mode adjusts the balance of rate- and synchrony-based coding.

These results demonstrate the importance of cell-level properties for network-level coding. There has been a longstanding bias in the network modeling community to focus on synaptic weights and network architecture, with far less emphasis put on cellular properties. Here, we have shown what happens when neurons do not "integrate and fire." Similarly, Burak et al. (2009) and Barreiro et al. (2010) recently showed differences in output correlation depending on the type of model used. Effects of synaptic kinetics, background noise, and input spike train statistics on output correlation have also been documented (Maex et al., 2000; Tetzlaff et al., 2008; Ostojic et al., 2009). Tchumatchenko et al. (2010) have shown specifically that output correlations can be rate dependent or independent according to input conditions. Our results emphasize how both types of correlations can coexist and that this depends on input conditions and (how that input is encoded given) single cell firing properties. Indeed, whereas simplified models tend to favor pure integration or pure coincidence detection, real neurons and realistic conductance-based models exhibit a context-dependent mixture of operating modes. Neglecting the richness of cell-level coding properties will surely translate into underestimation of network-level coding possibilities.

In conclusion, different types of neurons or even the same neuron operating under different conditions can be differentially sensitive to first- and second-order stimulus statistics, namely mean and variance. If we do not assume that fluctuations are simply noise, and if a neuron is sensitive to those fluctuations (which is true of coincidence detectors), then output spiking should carry information about the signal variance. Two such neurons receiving correlated input may carry information about signal variance in their precisely synchronized spiking. Even if correlated with firing rate (indeed, such output correlations might arise from correlations between features of the original input), spike-time synchronization is separate from rate co-

modulation, which, by definition, is directly linked to the firing rate that is tuned to the signal mean. Both forms of correlation may coexist and operate independently. Thus, the richness of single-neuron coding abilities translates into even richer multineuron coding possibilities.

References

Abeles M (1982) Role of the cortical neuron: integrator or coincidence detector? Isr J Med Sci 18:83–92.

Abeles M (1991) Corticonics: neural circuits of the cerebral cortex. Cambridge, MA: Cambridge UP.

Aertsen A, Diesmann M, Gewaltig MO (1996) Propagation of synchronous spiking activity in feedforward neural networks. J Physiol Paris 90:243–247.

Agmon-Snir H, Carr CE, Rinzel J (1998) The role of dendrites in auditory coincidence detection. Nature 393:268–272.

Arsiero M, Lüscher HR, Lundstrom BN, Giugliano M (2007) The impact of input fluctuations on the frequency–current relationships of layer 5 pyramidal neurons in the rat medial prefrontal cortex. J Neurosci 27:3274–3284.

Averbeck BB, Latham PE, Pouget A (2006) Neural correlations, population coding and computation. Nat Rev Neurosci 7:358–366.

Bair W, Zohary E, Newsome WT (2001) Correlated firing in macaque visual area MT: time scales and relationship to behavior. J Neurosci 21:1676–1697.

Barreiro AK, Shea-Brown E, Thilo EL (2010) Time scales of spike-train correlation for neural oscillators with common drive. Phys Rev E Stat Nonlin Soft Matter Phys 81:011916.

Bialek W, de Ruyter Van Steveninck RR (1988) Real-time performance of a movement-sensitive neuron in the blowfly visual system: coding and information transfer in short spike sequences. Proc R Soc Lond B Biol Sci 234:379–414.

Brown DA, Passmore GM (2009) Neural KCNQ (Kv7) channels. Br J Pharmacol 156:1185–1195.

Burak Y, Lewallen S, Sompolinsky H (2009) Stimulus-dependent correlations in threshold-crossing spiking neurons. Neural Comput 21:2269–2308.

Cafaro J, Rieke F (2010) Noise correlations improve response fidelity and stimulus encoding. Nature 468:964–967.

Cao XJ, Shatadal S, Oertel D (2007) Voltage-sensitive conductances of bushy cells of the mammalian ventral cochlear nucleus. J Neurophysiol 97:3961–3975.

Chialvo DR, Longtin A, Müller-Gerking J (1997) Stochastic resonance in models of neuronal ensembles. Phys Rev E Stat Nonlin Soft Matter Phys 55:1798–1808.

da Costa NM, Martin KA (2011) How thalamus connects to spiny stellate cells in the cat's visual cortex. J Neurosci 31:2925–2937.

de la Rocha J, Doiron B, Shea-Brown E, Josić K, Reyes A (2007) Correlation between neural spike trains increases with firing rate. Nature 448:802–806.

Destexhe A, Rudolph M, Fellous JM, Sejnowski TJ (2001) Fluctuating synaptic conductances recreate in vivo-like activity in neocortical neurons. Neuroscience 107:13–24.

Destexhe A, Rudolph M, Paré D (2003) The high-conductance state of neocortical neurons in vivo. Nat Rev Neurosci 4:739–751.

Douglas RJ, Martin KA (2007) Mapping the matrix: the ways of neocortex. Neuron 56:226–238.

Egorov AV, Hamam BN, Fransén E, Hasselmo ME, Alonso AA (2002) Graded persistent activity in entorhinal cortex neurons. Nature 420:173–178.

Engel AK, Roelfsema PR, Fries P, Brecht M, Singer W (1997) Role of the temporal domain for response selection and perceptual binding. Cereb Cortex 7:571–582.

Fellous JM, Rudolph M, Destexhe A, Sejnowski TJ (2003) Synaptic background noise controls the input/output characteristics of single cells in an in vitro model of in vivo activity. Neuroscience 122:811–829.

Fourcaud-Trocmé N, Hansel D, van Vreeswijk C, Brunel N (2003) How spike generation mechanisms determine the neuronal response to fluctuating inputs. J Neurosci 23:11628–11640.

Gerstner W, Kreiter AK, Markram H, Herz AV (1997) Neural codes: firing rates and beyond. Proc Natl Acad Sci U S A 94:12740–12741.

Guan D, Higgs MH, Horton LR, Spain WJ, Foehring RC (2011) Contributions of Kv7-mediated potassium current to sub- and suprathreshold responses of rat layer II/III neocortical pyramidal neurons. J Neurophysiol 106:1722–1733.

Higgs MH, Slee SJ, Spain WJ (2006) Diversity of gain modulation by noise in neocortical neurons: regulation by the slow afterhyperpolarization conductance. J Neurosci 26:8787–8799.

Higley MJ, Contreras D (2006) Balanced excitation and inhibition determine spike timing during frequency adaptation. J Neurosci 26:448–457.

Hines ML, Carnevale NT (1997) The NEURON simulation environment. Neural Comput 9:1179–1209.

Hong S, Lundstrom BN, Fairhall AL (2008) Intrinsic gain modulation and adaptive neural coding. PLoS Comput Biol 4:e1000119.

Hu H, Vervaeke K, Storm JF (2007) M-channels (Kv7/KCNQ channels) that regulate synaptic integration, excitability, and spike pattern of CA1 pyramidal cells are located in the perisomatic region. J Neurosci 27:1853–1867.

König P, Engel AK, Singer W (1996) Integrator or coincidence detector? The role of the cortical neuron revisited. Trends Neurosci 19:130–137.

Kremkow J, Aertsen A, Kumar A (2010) Gating of signal propagation in spiking neural networks by balanced and correlated excitation and inhibition. J Neurosci 30:15760–15768.

Kumar A, Rotter S, Aertsen A (2010) Spiking activity propagation in neuronal networks: reconciling different perspectives on neural coding. Nat Rev Neurosci 11:615–627.

Lundstrom BN, Hong S, Higgs MH, Fairhall AL (2008) Two computational regimes of a single-compartment neuron separated by a planar boundary in conductance space. Neural Comput 20:1239–1260.

Maex R, Vos BP, De Schutter E (2000) Weak common parallel fibre synapses explain the loose synchrony observed between rat cerebellar Golgi cells. J Physiol 523:175–192.

Mainen ZF, Sejnowski TJ (1995) Reliability of spike timing in neocortical neurons. Science 268:1503–1506.

Manis PB, Marx SO (1991) Outward currents in isolated ventral cochlear nucleus neurons. J Neurosci 11:2865–2880.

Mathews PJ, Jercog PE, Rinzel J, Scott LL, Golding NL (2010) Control of submillisecond synaptic timing in binaural coincidence detectors by K_v1 channels. Nat Neurosci 13:601–609.

Meister M, Berry MJ (1999) The neural code of the retina. Neuron 22:435–450.

Murphy BK, Miller KD (2009) Balanced amplification: a new mechanism of selective amplification of neural activity patterns. Neuron 61:635–648.

Ostojic S, Brunel N, Hakim V (2009) How connectivity, background activity, and synaptic properties shape the cross-correlation between spike trains. J Neurosci 29:10234–10253.

Palanca BJ, DeAngelis GC (2005) Does neuronal synchrony underlie visual feature grouping? Neuron 46:333–346.

Perkel DH, Gerstein GL, Moore GP (1967) Neuronal spike trains and stochastic point processes. II. Simultaneous spike trains. Biophys J 7:419–440.

Pinto RD, Elson RC, Szücs A, Rabinovich MI, Selverston AI, Abarbanel HD (2001) Extended dynamic clamp: controlling up to four neurons using a single desktop computer and interface. J Neurosci Methods 108:39–48.

Prescott SA, De Koninck Y (2009) Impact of background synaptic activity on neuronal response properties revealed by stepwise replication of in vivo-like conditions in vitro In: The dynamic clamp: from principles to applications (Destexhe A, Bal T, eds), pp 89–114. New York: Springer.

Prescott SA, Sejnowski TJ (2008) Spike-rate coding and spike-time coding are affected oppositely by different adaptation mechanisms. J Neurosci 28:13649–13661.

Prescott SA, Ratté S, De Koninck Y, Sejnowski TJ (2006) Nonlinear interaction between shunting and adaptation controls a switch between integration and coincidence detection in pyramidal neurons. J Neurosci 26:9084–9097.

Prescott SA, De Koninck Y, Sejnowski TJ (2008a) Biophysical basis for three distinct dynamical mechanisms of action potential initiation. PLoS Comput Biol 4:e1000198.

Prescott SA, Ratté S, De Koninck Y, Sejnowski TJ (2008b) Pyramidal neurons switch from integrators in vitro to resonators under in vivo-like conditions. J Neurophysiol 100:3030–3042.

Reyes AD (2003) Synchrony-dependent propagation of firing rate in iteratively constructed networks in vitro. Nat Neurosci 6:593–599.

Rieke F, Warland D, Bialek W (1999) Spikes: exploring the neural code. Cambridge, MA: MIT.

Rudolph M, Destexhe A (2003) Tuning neocortical pyramidal neurons between integrators and coincidence detectors. J Comput Neurosci 14:239–251.

Salinas E, Sejnowski TJ (2001) Correlated neuronal activity and the flow of neural information. Nat Rev Neurosci 2:539–550.

Schneidman E, Berry MJ 2nd, Segev R, Bialek W (2006) Weak pairwise correlations imply strongly correlated network states in a neural population. Nature 440:1007–1012.

Schreiber S, Fellous JM, Tiesinga P, Sejnowski TJ (2004) Influence of ionic conductances on spike timing reliability of cortical neurons for suprathreshold rhythmic inputs. J Neurophysiol 91:194–205.

Shadlen MN, Movshon JA (1999) Synchrony unbound: a critical evaluation of the temporal binding hypothesis. Neuron 24:67–77, 111–125.

Shadlen MN, Newsome WT (1998) The variable discharge of cortical neurons: implications for connectivity, computation, and information coding. J Neurosci 18:3870–3896.

Shea-Brown E, Josić K, de la Rocha J, Doiron B (2008) Correlation and synchrony transfer in integrate-and-fire neurons: basic properties and consequences for coding. Phys Rev Lett 100:108102.

Singer W (1993) Synchronization of cortical activity and its putative role in information processing and learning. Annu Rev Physiol 55:349–374.

Softky WR, Koch C (1993) The highly irregular firing of cortical cells is inconsistent with temporal integration of random EPSPs. J Neurosci 13:334–350.

Svirskis G, Kotak V, Sanes DH, Rinzel J (2004) Sodium along with low-threshold potassium currents enhance coincidence detection of subthreshold noisy signals in MSO neurons. J Neurophysiol 91:2465–2473.

Tchumatchenko T, Malyshev A, Geisel T, Volgushev M, Wolf F (2010) Correlations and synchrony in threshold neuron models. Phys Rev Lett 104:058102.

Tetzlaff T, Rotter S, Stark E, Abeles M, Aertsen A, Diesmann M (2008) Dependence of neuronal correlations on filter characteristics and marginal spike train statistics. Neural Comput 20:2133–2184.

Theunissen F, Miller JP (1995) Temporal encoding in nervous systems: a rigorous definition. J Comput Neurosci 2:149–162.

Tiesinga P, Fellous JM, Sejnowski TJ (2008) Regulation of spike timing in visual cortical circuits. Nat Rev Neurosci 9:97–107.

Treisman A (1999) Solutions to the binding problem: progress through controversy and convergence. Neuron 24:105–110, 111–125.

Uhlenbeck GE, Ornstein LS (1930) On the theory of Brownian motion. Phys Rev 36:823–841.

Victor J, Shapley R (1980) A method of nonlinear analysis in the frequency domain. Biophys J 29:459–483.

Wehr M, Zador AM (2003) Balanced inhibition underlies tuning and sharpens spike timing in auditory cortex. Nature 426:442–446.

Wolfe J, Houweling AR, Brecht M (2010) Sparse and powerful cortical spikes. Curr Opin Neurobiol 20:306–312.

Zohary E, Shadlen MN, Newsome WT (1994) Correlated neuronal discharge rate and its implications for psychophysical performance. Nature 370:140–143.

OPEN ACCESS Freely available online

PLoS COMPUTATIONAL BIOLOGY

Intrinsic Gain Modulation and Adaptive Neural Coding

Sungho Hong¤*, **Brian Nils Lundstrom, Adrienne L. Fairhall**

Physiology and Biophysics Department, University of Washington, Seattle, Washington, United States of America

Abstract

In many cases, the computation of a neural system can be reduced to a receptive field, or a set of linear filters, and a thresholding function, or gain curve, which determines the firing probability; this is known as a linear/nonlinear model. In some forms of sensory adaptation, these linear filters and gain curve adjust very rapidly to changes in the variance of a randomly varying driving input. An apparently similar but previously unrelated issue is the observation of gain control by background noise in cortical neurons: the slope of the firing rate versus current (*f-I*) curve changes with the variance of background random input. Here, we show a direct correspondence between these two observations by relating variance-dependent changes in the gain of *f-I* curves to characteristics of the changing empirical linear/nonlinear model obtained by sampling. In the case that the underlying system is fixed, we derive relationships relating the change of the gain with respect to both mean and variance with the receptive fields derived from reverse correlation on a white noise stimulus. Using two conductance-based model neurons that display distinct gain modulation properties through a simple change in parameters, we show that coding properties of both these models quantitatively satisfy the predicted relationships. Our results describe how both variance-dependent gain modulation and adaptive neural computation result from intrinsic nonlinearity.

Citation: Hong S, Lundstrom BN, Fairhall AL (2008) Intrinsic Gain Modulation and Adaptive Neural Coding. PLoS Comput Biol 4(7): e1000119. doi:10.1371/journal.pcbi.1000119

Editor: Karl J. Friston, University College London, United Kingdom

Received January 30, 2008; **Accepted** June 9, 2008; **Published** July 18, 2008

Funding: This work was supported by a Burroughs-Wellcome Careers at the Scientific Interface grant, a Sloan Research Fellowship, and a McKnight Scholar Award to ALF. BNL was supported by grant number F30NS055650 from the National Institute of Neurological Disorders and Stroke, the Medical Scientist Training Program at the University of Washington supported by the National Institute of General Sciences, and an Achievement Rewards for College Scientists fellowship.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: shhong@oist.jp

¤ Current address: Computational Neuroscience Unit, Okinawa Institute of Science and Technology, Onna-son, Okinawa, Japan

Introduction

An *f-I* curve, defined as the mean firing rate in response to a stationary mean current input, is one of the simplest ways to characterize how a neuron transforms a stimulus into a spike train output as a function of the magnitude of a single stimulus parameter. Recently, the dependence of *f-I* curves on other input statistics such as the variance has been examined: the slope of the *f-I* curve, or gain, is modulated in diverse ways in response to different intensities of added noise [1–4]. This enables multiplicative control of the neuronal gain by the level of background synaptic activity [1]: changing the level of the background synaptic activity is equivalent to changing the variance of the noisy balanced excitatory and inhibitory input current to the soma, which modulates the gain of the *f-I* curve. It has been demonstrated that such somatic gain modulation, combined with saturation in the dendrites, can lead to multiplicative gain control in a single neuron by background inputs [5]. From a computational perspective, the sensitivity of the firing rate to mean or variance can be thought of as distinguishing the neuron's function as either an integrator (greater sensitivity to the mean) or a differentiator/coincidence detector (greater sensitivity to fluctuations, as quantified by the variance) [3,6,7].

An alternative method of characterizing a neuron's input-to-output transformation is through a linear/nonlinear (LN) cascade model [8,9]. These models comprise a set of linear filters or receptive field that selects particular features from the input; the filter output is transformed by a nonlinear threshold stage into a time-varying firing rate. Spike-triggered covariance analysis [10,11] reconstructs a model with multiple features from a neuron's input/output data. It has been widely employed to characterize both neural systems [12–15] and single neurons or neuron models subject to current or conductance inputs [16–19].

Generally, results of reverse correlation analysis may depend on the statistics of the stimulus used to sample the model [15,19–25]. While some of the dependence on stimulus statistics in the response of a neuron or neural system may reflect underlying plasticity, in some cases, the rapid timescale of the changes suggests the action of intrinsic nonlinearities in systems with *fixed* parameters [16,19,25–29], which changes the *effective* computation of a neuron.

Our goal here is to unify the *f-I* curve description of variance-dependent adaptive computation with that given by the LN model: we present analytical results showing that the variance-dependent modulation of the firing rate is closely related to adaptive changes in the *recovered* LN model if a fixed underlying model is assumed. When the model relies only on a single feature, we find that such a system can show only a single type of gain modulation, which accompanies an interesting asymptotic scaling behavior. With multiple features, the model can show more diverse adaptive behaviors, exemplified by two conductance-based models that we will study.

Results

Diverse Variance-Dependent Gain Modulations without Spike Rate Adaptation

Recently, Higgs et al. [3] and Arsiero et al. [4] identified different forms of variance-dependent change in the *f-I* curves of

Author Summary

Many neurons are known to achieve a wide dynamic range by adaptively changing their computational input/output function according to the input statistics. These adaptive changes can be very rapid, and it has been suggested that a component of this adaptation could be purely input-driven: even a fixed neural system can show apparent adaptive behavior since inputs with different statistics interact with the nonlinearity of the system in different ways. In this paper, we show how a single neuron's intrinsic computational function can dictate such input-driven changes in its response to varying input statistics, which begets a relationship between two different characterizations of neural function—in terms of mean firing rate and in terms of generating precise spike timing. We then apply our results to two biophysically defined model neurons, which have significantly different response patterns to inputs with various statistics. Our model of intrinsic adaptation explains their behaviors well. Contrary to the picture that neurons carry out a stereotyped computation on their inputs, our results show that even in the simplest cases they have simple yet effective mechanisms by which they can adapt to their input. Adaptation to stimulus statistics, therefore, is built into the most basic single neuron computations.

various neuron types in avian brainstem and in cortex. Depending on the type, neurons can have either increasing or decreasing gain in the *f-I* curve with increasing variance. These papers linked the phenomenon to mechanisms underlying spike rate adaptation, such as slow afterhyperpolarization (sAHP) currents and slow sodium channel inactivation. We recently showed [7] that a standard Hodgkin–Huxley (HH) neuron model, lacking spike rate adaptation, can show two different types of variance-dependent gain modulation simply by tuning the maximal conductance parameters of the model. These differences in gain modulation correspond to two different regimes in the space of conductance parameters. In one regime, which includes the standard parameters, a neuron periodically fires to a sufficiently large constant input current. In the other regime, a neuron never fires to a constant input regardless of its magnitude, but responds only to rapid fluctuations. This rarely discussed property has been termed *class 3 excitability* [30,31]. Higgs et al. [3] proposed that the type of gain modulation classifies the neuron as an integrator or differentiator.

Here, we examine two models that show these different forms of variance-dependent gain modulation without spike rate adaptation, and study the resulting LN models sampled with different stimulus statistics. We show that these *fixed* models generate variance-dependent gain modulation, and that this gain modulation is well predicted by aspects of the LN models derived from white noise stimulation. The two models are both based on the HH [32] active currents; one model is the standard HH model, and the other (HHLS) has lower Na^+ and higher K^+ conductances. The HHLS model is a class 3 neuron and responds only to a rapidly changing input. For this reason, the HHLS model can be thought of as behaving more like a differentiator than an integrator [3,7].

Figure 1 shows the different gain modulation behaviors of the HH and HHLS conductance-based models. For the HH model, Figure 1A, the *f-I* curves in the presence of noise are similar to the noiseless case except that they are increasingly smoothed at the threshold. In contrast, Figure 1C shows that the *f-I* curves of the HHLS model never converge toward each other as the noise level increases. This case resembles that of layer 5 pyramidal neurons in rat medial prefrontal cortex [4], as well as nucleus laminaris (NL) neurons in the chick auditory brainstem and some pyramidal neurons in layer 2/3 of rat neocortex [3]. While for these layer 2/3 neurons, there is evidence that this change in *f-I* curve slope may be related to the sAHP current [3], at steady state this effect can be obtained in general by tuning the maximal conductances without introducing any mechanism for spike rate adaptation [7].

Gain Modulation and Adaptation of Fixed Models

For a system described by an LN model with a single feature, we derive an equation relating the slopes of the firing rate with respect to stimulus mean and variance. We then consider gain modulation in a system with multiple relevant features and derive a series of equations relating gain change to properties of the spike-triggered average and spike-triggered covariance. Throughout, we assume that the underlying system is fixed, and that its parameter settings do not depend on stimulus statistics. For example, if the model has a single exponential filter with a time constant τ, we assume that τ does not change with the stimulus mean (I_0) or variance (σ^2). However, this does not mean that the model shows a single response pattern regardless of the statistical structure of stimuli. The sampled LN description of a nonlinear system with fixed parameters—even when the underlying model is an LN model [25]—can show interaction with the input statistics, leading to different LN model descriptions for different input parameters [19,25,27–29]. We refer to this as *intrinsic adaptation*.

One-Dimensional Model

An LN model is composed of its relevant features $\{\varepsilon_\mu(t)\}$ ($\mu = 1,2,\ldots,n$), which act as linear filters on an incoming stimulus, and a probability to spike given the filtered stimulus, $P(\text{spike}|\text{filtered stimulus})$. For a Gaussian white noise stimulus with mean I_0 and variance σ^2, the firing rate is

$$f(I_0,\sigma^2) = \int \mathrm{d}\mathbf{x}\; P(\text{spike}|I_0\bar{\boldsymbol{\varepsilon}}+\mathbf{x})p(\mathbf{x}) \tag{1}$$

where $\bar{\boldsymbol{\varepsilon}} = \int_0^\infty \boldsymbol{\varepsilon}(\tau)\mathrm{d}\tau$ is the time-integrated filter and $\mathbf{x}$ is the mean-subtracted noise stimulus filtered by the n relevant features. $p(\mathbf{x})$ is an n-dimensional Gaussian distribution with variance σ^2. We refer to the Materials and Methods section for a more detailed account of the model.

For a one-dimensional model $n = 1$, Equation 1 can be rewritten with change of variables

$$f(I_0,\sigma^2) = \int_{-\infty}^{\infty} \mathrm{d}x\; P(\text{spike}|x)p(x-I_0\bar{\varepsilon}) \tag{2}$$

Since $p(x)$ is Gaussian, it is also the kernel or Green's function of a diffusion equation in terms of (x,σ^2) and therefore so is $p(x-I_0\bar{\varepsilon})$ in terms of (I_0,σ^2). In other words, we have

$$\left(\frac{\partial}{\partial\sigma^2} - \frac{1}{2}\frac{\partial^2}{\partial x^2}\right)p(x-I_0\bar{\varepsilon})$$
$$= \left(\frac{\partial}{\partial\sigma^2} - \frac{1}{2\bar{\varepsilon}^2}\frac{\partial^2}{\partial I_0^2}\right)p(x-I_0\bar{\varepsilon}) = 0$$

Now operating with $\left(\frac{\partial}{\partial\sigma^2} - \frac{1}{2\bar{\varepsilon}^2}\frac{\partial^2}{\partial I_0^2}\right)$ on both sides of the equation, $p(x-I_0\bar{\varepsilon})$ is the only term on the left hand side of Equation 2 that

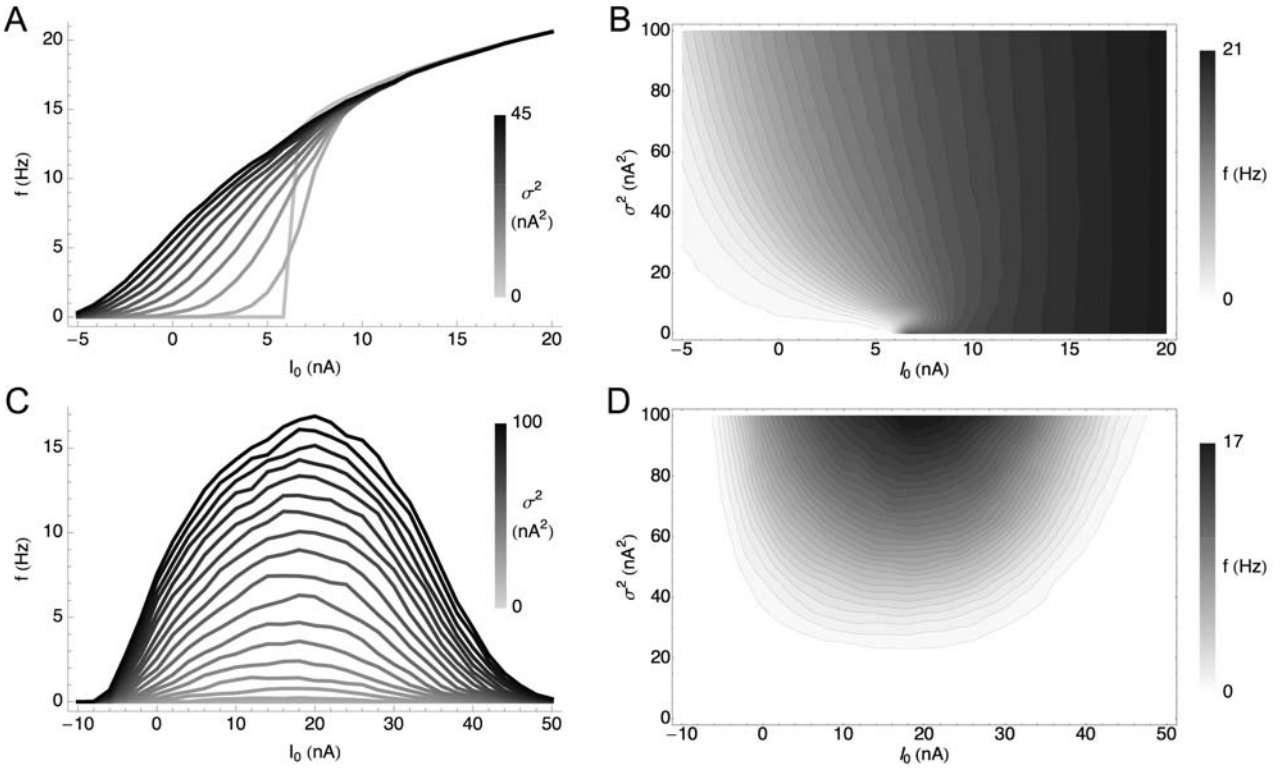


Figure 1. Variance-Dependent Gain Modulation of the HH and HHLS Model. Each model is simulated as described in the Materials and Methods section. (A) *f-I* curves of a standard HH model for differing 10 variances (σ^2) from 0 to 45 nA^2. The topmost trace is the response to the highest variance. Each curve is obtained with 31 mean values (I_0) ranging from −5 to 20 nA. (B) The same data as (A) plotted in the (mean, variance) plane. Lighter shades represent higher firing rates. We used cubic spline interpolation for points not included in the simulated data. (C,D) *f-I* curves of the HHLS model as in (A) and (B). 10 means from −10 to 50 nA and 21 variances from 0 to 100 nA^2 are used.
doi:10.1371/journal.pcbi.1000119.g001

depends on (I_0,σ^2) and therefore the right hand side of Equation 2 vanishes. Thus one finds

$$2\bar{\varepsilon}^2 \frac{\partial f}{\partial \sigma^2} = \frac{\partial^2 f}{\partial I_0^2} \tag{3}$$

The boundary condition is given by evaluating Equation 2 as $\sigma^2 \to 0$; here the Gaussian distribution becomes a delta function

$$\lim_{\sigma^2 \to 0} p(x - I_0\bar{\varepsilon}) = \delta(x - I_0\bar{\varepsilon})$$

and the boundary condition is given by the zero-noise *f-I* curve. Thus, when a model depends only on a single feature, $\varepsilon(t)$, the *f-I* curve with a noisy input is given by a simple diffusion-like equation, Equation 3, with a single parameter, the time integrated filter, $\bar{\varepsilon} = \int_0^\infty \varepsilon(\tau)\mathrm{d}\tau$, determining the diffusion constant $1/2\bar{\varepsilon}^2$.

Equation 3 states that the variance-dependent change in the firing rate is simply determined by the curvature of the *f-I* curve. Thus, a one-dimensional system displays only a single type of noise-induced gain modulation: as in diffusion, an *f-I* curve is gradually smoothed and flattened as the variance increases. Given a boundary condition, such as an *f-I* curve for a particular variance, the family of *f-I* relations can be reconstructed up to a scale factor by solving Equation 3. For example, one can predict how the neuron would respond to a noise stimulus based on its output in the absence of noise. Note that the solution of Equation 3 generalizes a classical result [33] based on a binary nonlinearity to a simple closed form which applies to any type of nonlinearity.

Figure 2A and 2B show a solution of Equation 3. While this one-dimensional model is based on the simplest and most general assumptions, it provides insights into the structure of variance-dependent gain modulation. The boundary condition is an *f-I* curve with no noise, $f = (I+0.1)^{1/2}$ for $I>0$ and $f = 0$ for $I \leq 0$, which imitates the general behavior of many dynamical neuron models around rheobase [34–36]. Compared with the HH conductance-based model, Equation 3 captures qualitative characteristics of the HH *f-I* curve despite differences due to the increased complexity of the HH model over a 1D LN model: in Figure 2A and 2B, there is a positive curvature (second derivative of firing rate with respect to current) of the *f-I* curve below rheobase related to the increase of the firing rate with increasing variance. In contrast, the behavior of the HHLS model cannot be described by Equation 3. Even though the *f-I* curves in Figure 1C mostly have negative curvature, the firing rate keeps increasing with variance, implying that the HHLS model cannot be described by a one-dimensional LN model.

We also compared Equation 3 with the *f-I* curves from two commonly used simple neuron models, the leaky integrate-and-fire (LIF) model (Figure 2C), and a similar model with minimal nonlinearity, the quadratic integrate-and-fire (QIF) model [37,38] (Figure 2D). The *f-I* curves of the two models are similar but have subtle differences: in the LIF model, firing rate never decreases with noise, even though parameters were chosen to induce a large negative curvature, as shown analytically in Text S1. The QIF

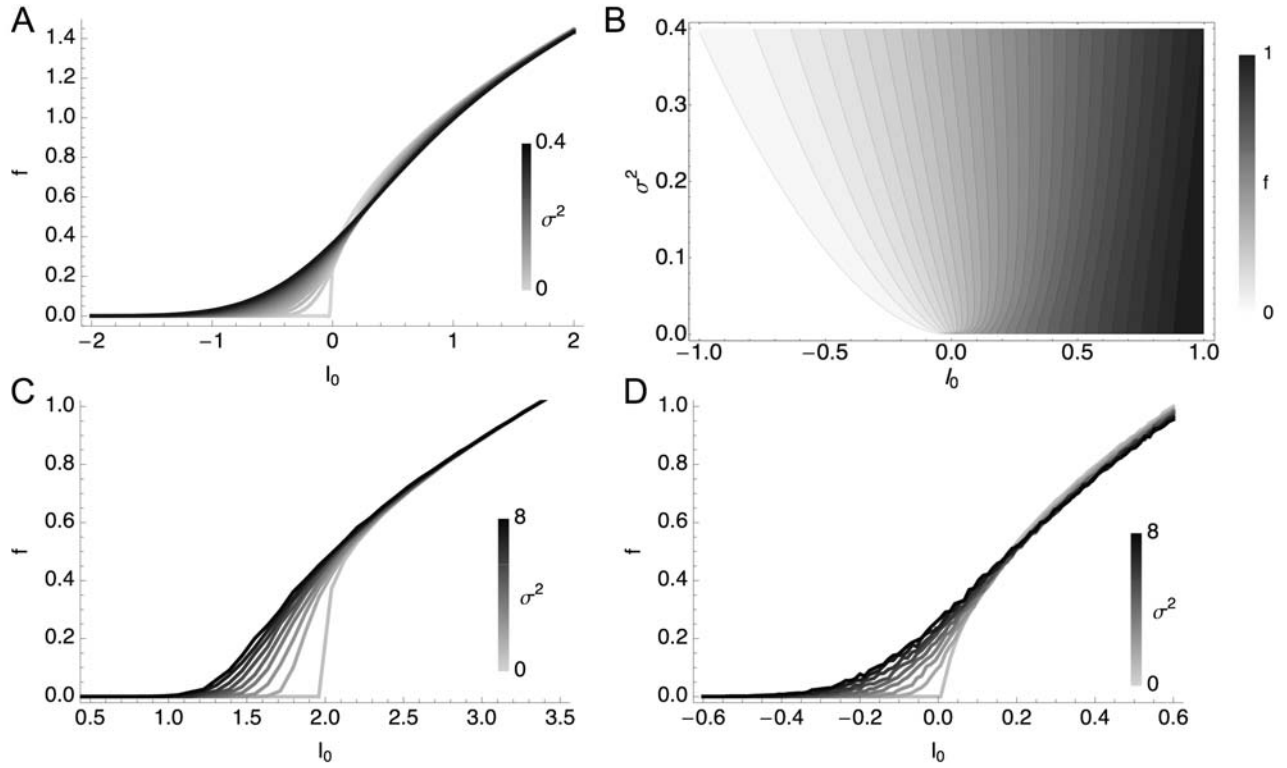


Figure 2. Variance-Dependent Gain Modulation of One-Dimensional Models. (A) Variance-dependent *f-I* curves of a one-dimensional model from the solution of Equation 3 with the boundary condition, $f=(I+0.1)^{1/2}$ for $I>0$ and $f=0$ for $I\leq 0$ at zero noise. (B) The firing rates of A in the (mean, variance) plane. (C) *f-I* curves of an LIF model. (D) *f-I* curves of a QIF model. The model parameters for the LIF and QIF are in the Materials and Methods section. We used 50 mean (I_0) values from 0 to 4 (LIF) and from -2 to 2 (QIF), and 8 variances (σ^2) from 0 to 8 for both models.
doi:10.1371/journal.pcbi.1000119.g002

model behavior is much more similar to the 1D LN model, marked by a slight decrease in firing rate at large I_0. From this perspective, the QIF is a *simpler* model in terms of the LN description despite the dynamical nonlinearity.

It is interesting to note that for one-dimensional models, the gain modulation given by Equation 3 depends only on the boundary condition, which implicitly describes how an input with a given mean samples the nonlinearity, but not explicitly on the details of filters or nonlinearity. An ideal differentiator, where firing rate is independent of the stimulus mean, is realized only when the filter has zero integral, $\bar{\varepsilon}=0$. This is also the criterion that would be satisfied if the filter itself were ideally differentiating. We will return to the relationship between the LN model functional description and that of the *f-I* curves in the Discussion.

Multidimensional Models

Here we examine gain modulation in the case of a system with multiple relevant features. In this case, one cannot derive a single simple equation such as Equation 3. Instead, we derive relationships between the characteristics of $f(I_0,\sigma)$ curves and quantities calculated using white noise analysis.

Fixed multidimensional models can display far more complex response patterns to different stimulus statistics than one-dimensional models, because linear components in the model can now interact nonlinearly [29]. For example, in white noise analysis, as the stimulus variance increases, the distribution of the filtered stimuli also expands and probes different regions of the nonlinear threshold structure of the model. This induces a variance-dependent rotation among the filters recovered through sampling by white noise analysis, and the corresponding changes in the spike-triggered average, spike-triggered covariance, and the sampled nonlinearity [19].

Here, we relate parameters of the changing spike-triggered average and spike-triggered covariance description to the form of the *f-I* curves. The relationships are derived by taking derivatives of each side of Equation 1 with respect to I_0 and σ^2 (see Materials and Methods section). The first order in I_0 establishes the relationship between the STA and the gain of the *f-I* curve with respect to the mean

$$\frac{\partial \log f}{\partial I_0}=\frac{1}{\sigma^2}\overline{\mathrm{STA}},\quad \overline{\mathrm{STA}}=\int_0^\infty \mathrm{d}\tau \mathrm{STA}(\tau) \tag{4}$$

The second order leads to a relationship between the second derivative of the *f-I* curve and the covariance matrix

$$\frac{\partial^2 \log f}{\partial I_0^2}=\frac{1}{\sigma^4}\overline{\Delta C},\quad \overline{\Delta C}=\int \mathrm{d}\tau\mathrm{d}\tau' \Delta C(\tau,\tau') \tag{5}$$

The gain with respect to the variance is

$$\frac{\partial \log f}{\partial \sigma^2}=\frac{1}{2\sigma^4}\left(\mathrm{Tr}\Delta C+\|\mathrm{STA}\|^2\right) \tag{6}$$

where

$$\mathrm{Tr}\Delta C=\int \mathrm{d}\tau \Delta C(\tau,\tau),\quad \|\mathrm{STA}\|^2=\int \mathrm{d}\tau \mathrm{STA}(\tau)^2$$

Equations 4–6 show how the nonlinear gain of an *f-I* curve with respect to input mean and variance is related to intrinsic adaptation as observed through changes in the STA and STC. Note that Equations 4–6 apply to one-dimensional LN models as well. In that case, the STA has the same shape as the feature in the model, and only its magnitude varies according to the overlap integral, Equation 1, between the nonlinearity of the model and the prior stimulus. This is the same for the STC, and thus Equations 4–6 are not independent. This leads to a single form of variance gain modulation, given by Equation 3. However, in a multidimensional model, changing the stimulus mean shifts the nonlinearity in a single direction, $\overline{\mathrm{STA}}$, while increasing the variance expands the prior in every direction in the stimulus space. Therefore, the overlap integral can show more diverse behaviors.

Conductance-Based Models

We now examine whether the gain modulation behaviors we have described can be captured by a multi-dimensional LN model. We tested this by computing *f-I* curves, spike-triggered averages and the spike-triggered covariance matrices for the noise-driven HH and HHLS models for a range of input statistics. Figure 3A, B, and C show the result of fitting simulation data from the HH (left) and HHLS (right) model to Equations 4, 5, and 6, respectively. The linear relationships are quite clear in Figure 3A

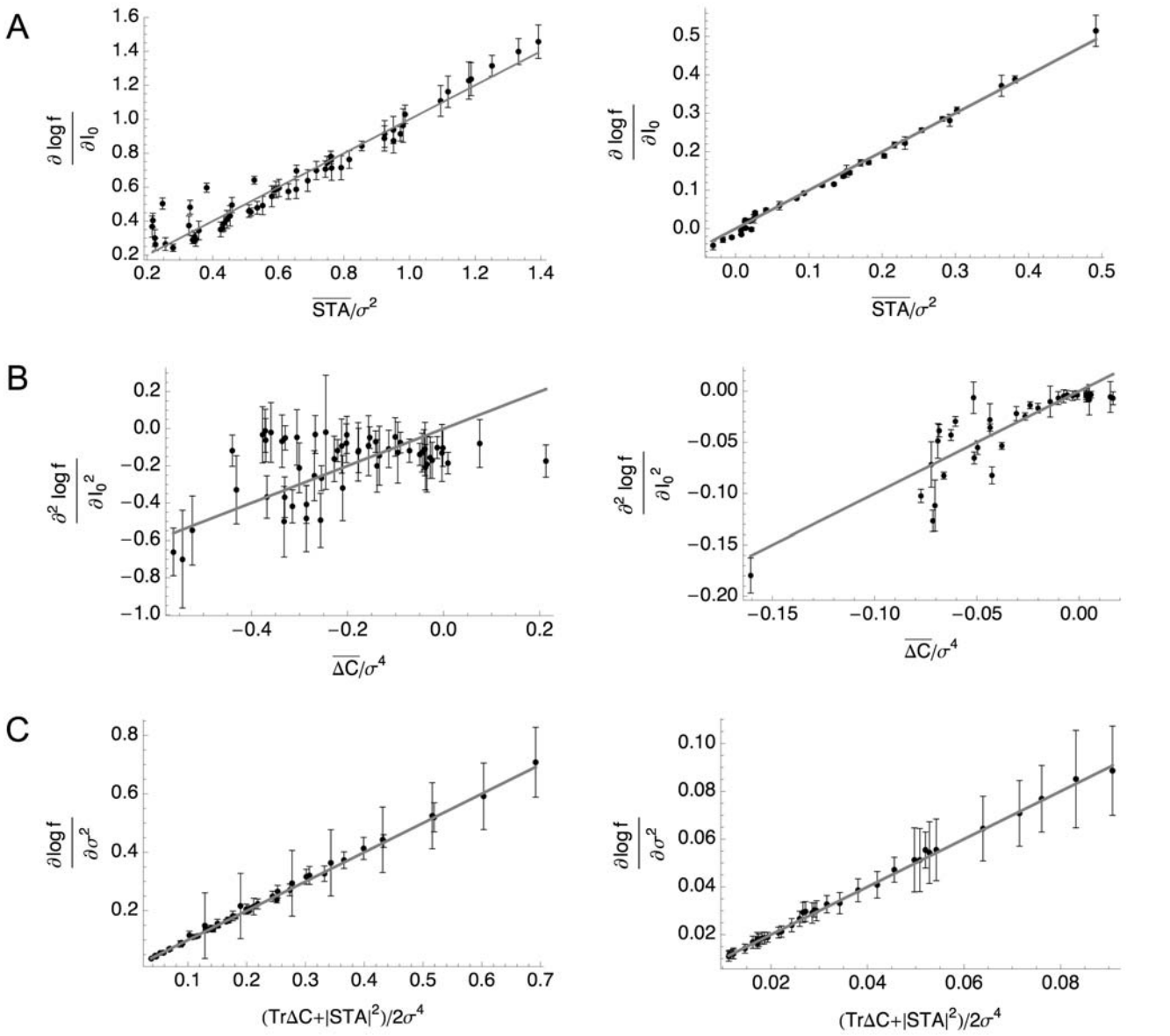


Figure 3. Derivatives of the Firing Rate Curves with Respect to Mean and Variance Related to Quantities Obtained by White Noise Analysis for the Standard HH (Left) and HHLS (Right) Models. Each point is calculated from the simulated data with a selected (mean, variance) input parameter pair, as described in the Materials and Methods section, and the gray lines represent our theoretical predictions, Equations 4–6, which hold when the variance dependent change in *f-I* curves is only due to intrinsic adaptation. (A) Gain versus the norm of the STA, as in Equation 4. (B) Gain change versus the spike-triggered covariance term of Equation 5. (C) Change of firing rate with respect to variance versus the function of the STA and spike-triggered covariance given in Equation 6.
doi:10.1371/journal.pcbi.1000119.g003

and 3C which show the gains with respect to mean and variance. Figure 3B involves the curvature of *f-I* curves, which is more difficult to calculate accurately, and shows larger errors. In every case, goodness of fit is $p<1.3\times10^{-6}$ and $p<5.8\times10^{-6}$ for the HH and HHLS where the upper bounds of p-values are given by the case of Equation 5, corresponding to Figure 3B. These results show that intrinsic adaptation of the LN model predicts the form of noise-induced gain modulation for these models.

Gain Rescaling of One-Dimensional Models

Here we discuss a consequence of intrinsic adaptation for neuronal encoding of mean and variance information for a one-dimensional model. In this case, Equation 3 completely specifies intrinsic adaptation, and therefore we will focus on this case.

Our first observation is that Equation 3 is invariant under the simultaneous rescaling of the mean and standard deviation, $I_0\rightarrow\alpha I_0$, $\sigma\rightarrow\alpha\sigma$, where α is an arbitrary positive number. This invariance is preserved if the solution is also a function of only a dimensionless variable I_0/σ, which would represent a signal-to-noise ratio if we describe the neuron's input/output function in terms of an *f-I* curve at a fixed noise level σ. Note that this situation is analogous to the Weber–Fechner [39,40] and Fitts' law [41], which states that perception tends to depend on only dimensionless variables that are invariant under scaling of the absolute magnitude of stimulus [42]. However, the invariance of Equation 3 under the scaling of a stimulus does not necessarily lead to the invariance of a firing rate solution. By rewriting Equation 2 in terms of the "rescaled" variables, $y=x/\sigma$ and $\mu=I_0/\sigma$, we get

$$f\left(\mu,\sigma^2\right)=\frac{1}{\sqrt{2\pi}}\int \mathrm{d}y\, \mathrm{e}^{-(y-\mu\bar{\varepsilon})^2/2} f_0\left(\frac{y\sigma}{\bar{\varepsilon}}\right) \tag{7}$$

where $f_0(I)=P(\text{spike}\,|\,I\bar{\varepsilon})$ is an *f-I* curve with no noise. Thus, the scaling of $f(I_0,\sigma^2)$ with standard deviation depends on the boundary condition, $f_0(I)$, which in principle can be any arbitrary function.

Nevertheless, in practice, the *f-I* curves of many dynamical neurons are not completely arbitrary but can share a simple scaling property, at least asymptotically. For example, in the QIF and many other neuron models, the *f-I* curve with no noise asymptotically follows a power law $f_0\sim(I_0-I_c)^{1/2}$ around the rheobase I_c [34–36]. In general, if $f_0(I)\propto I^\alpha$ asymptotically in such a regime, from Equation 7, the firing rate is asymptotically factorized into a σ dependent and $\mu=I_0/\sigma$ dependent part as

$$f\left(\mu,\sigma^2\right)\propto\sigma^\alpha F(\mu),\quad F(\mu)=\frac{1}{\sqrt{2\pi}}\int \mathrm{d}y\, \mathrm{e}^{-(y-\mu\bar{\varepsilon})^2/2} y^\alpha \tag{8}$$

In other words, I_0/σ becomes an *intermediate asymptotic* of the *f-I* curves [43].

To test to what extent this scaling relationship holds in the models we have considered, we calculated the *rescaled relative gain* of the *f-I* curves, which we define as $(\sigma/f)\,\partial f/\partial I_0=\sigma\,\partial\log f/\partial I_0$; the rescaled relative gain of Equation 8 depends only on $\mu=I_0/\sigma$, not on σ. Thus, if the rescaling strictly holds, this becomes a single-valued function of the signal-to-noise ratio, I_0/σ, regardless of the noise level σ.

We find evidence for this form of variance rescaling in the QIF, LIF, and HH models. Figure 4 shows the rescaled gains evaluated from the simulated data. The QIF and HH case, Figure 4B and 4D, match well with the solution of Equation 3, Figure 4A. In the LIF case, Figure 4C, the relative gain shows deviations at low variance, but it approaches a variance-independent limit at large σ. We also present an analytic account in Text S1. On the other hand, in Figure 4E, the HHLS model does not exhibit this form of asymptotic scaling at all. The role of the signal-to-noise ratio, I_0/σ, in the HHLS model appears to be quite distinct from the other models. In summary, Equation 3 predicts that one-dimensional LN models will have the tendency to decrease gain with increasing noise level. However, if the *f-I* curve of a neuron is power-law-like, the resulting gain modulation will be such that the neuron's sensitivity to mean stimulus change at various noise levels is governed only by the signal-to-noise ratio.

Discussion

In this paper, we have obtained analytical relationships between noise-dependent gain modulation of *f-I* curves and properties of the sampled linear/nonlinear model. We have shown that gain control arises as a simple consequence of the nonlinearity of the LN model, even with no changes in any underlying parameters.

For a system described by an LN model with only one relevant feature, a simple single-parameter diffusion relationship relates the *f-I* curves at different variances, where the role of the diffusion coefficient is taken by the integral of the STA. This form strictly limits the possible forms of gain modulation that may be manifested by such a system. The result qualitatively describes the variance dependent gain modulation of different neuron models such as the LIF, QIF, and standard HH neuron models. Models based on dynamical spike generation, such as QIF, showed better agreement with this result than the LIF model. The QIF model case is a good example of how a nonlinear dynamical system can be mapped onto an LN model description [19,44]. The QIF model has a single dynamical equation whose subthreshold dynamics are captured approximately by a linear kernel, which takes the role of the feature; one can then determine a threshold which acts as a binary decision boundary for spiking. Thus, it is reasonable that the QIF model and the one-dimensional LN model show a similar response pattern to a noisy input. When the system has multiple relevant features, we obtain equations relating the gain with respect to the input mean and the input variance to parameters of the STA and STC. We verified these results using HH neurons displaying two different forms of noise-induced gain control.

Previous work has related different gain control behaviors to a neuron's function as an integrator or a differentiator [3,7]. From an LN model perspective, the neuron's function is defined by specific properties of the filter or filters $\varepsilon(t)$. An integrating filter would consist of entirely positive weights; for leaky integrators these weights will decay at large negative times. A differentiating filter implements a local subtraction of the stimulus, and so should consist of a bimodal form where the positive weights approximately cancel the negative weights.

In general, characterizations of neural function by LN model and by *f-I* curves are quite distinct. The *f-I* approach we have discussed here describes the encoding of stationary statistical properties of the stimulus by time-averaged firing rate, while the LN model describes the encoding of specific input fluctuations by single spikes, generally under a particular choice of stimulus statistics. Indeed, the LN characterization can change with the driving stimulus distribution, even, in principle, from an integrator to a differentiator. Thus, a model may, for example, act as a differentiator on short timescales but as an integrator on longer timescales. For systems whose LN approximation varies with mean and variance, the neuron's effective computation changes with stimulus statistics, and so does the information that is represented.

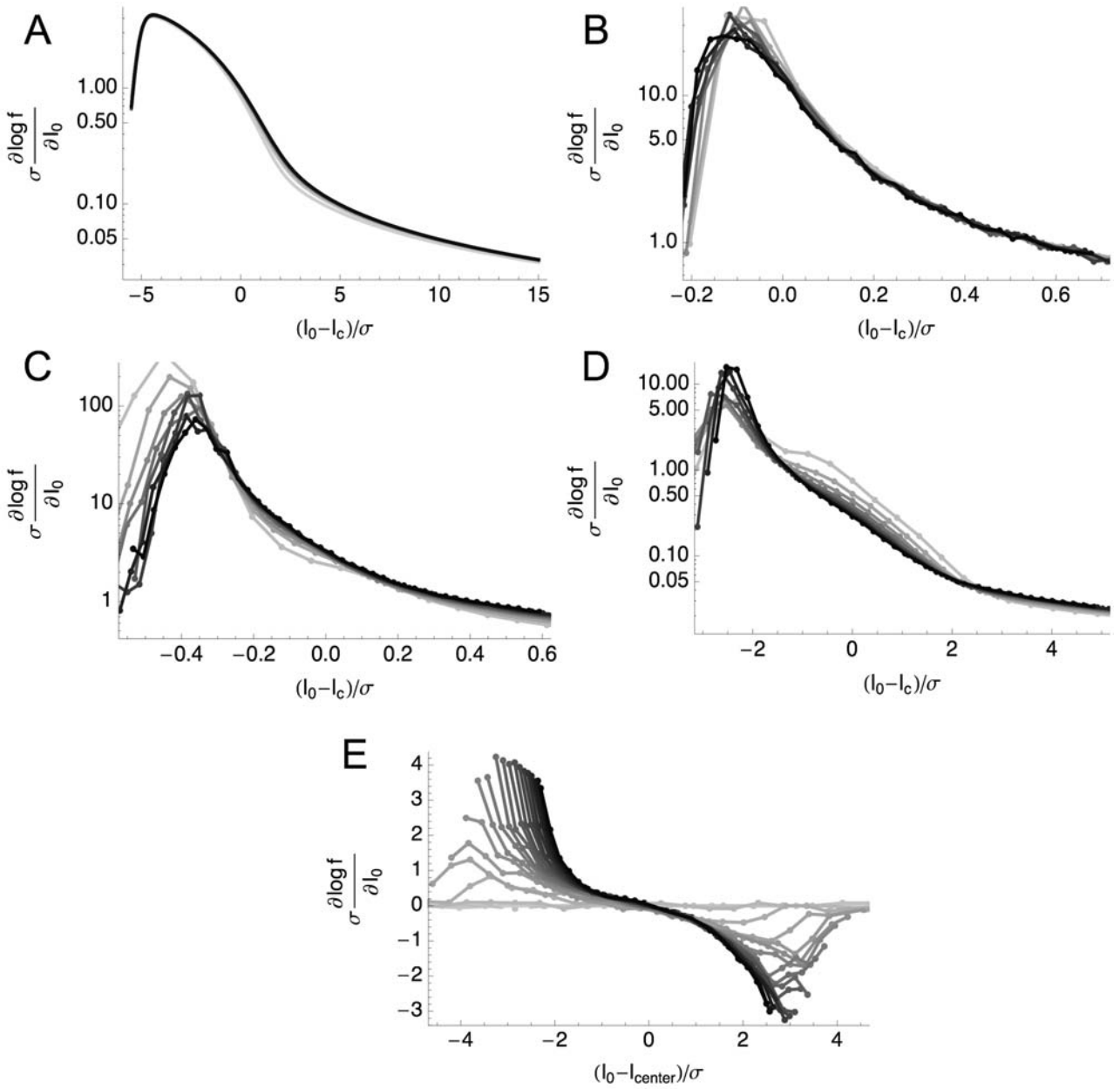


Figure 4. Rescaled Relative Gains of Variance-Dependent *f-I* Curves. (A) The one-dimensional LN, (B) QIF, and (C) LIF models. The same data as Figure 2 are used. (D) The standard HH model from Figure 1A and 1B. (E) The HHLS model from Figure 1C and 1D. Since the HHLS does not have a rheobase, we instead used $I_{center} = 20$ nA at which the variance-dependent firing rate increase is maximal.
doi:10.1371/journal.pcbi.1000119.g004

One might then ask how the system can decode the represented information. It has been proposed that statistics of the spike train might provide the information required to decode slower-varying stimulus parameters [22,45]. The possibility of distinguishing between responses to different stimulus statistics using the firing rate alone depends on the properties of the *f-I* curves.

The primary focus of this work is the restricted problem of single neurons responding to driving currents, where the integrated synaptic current in an in vivo-like condition is approximated to be a (filtered) Gaussian white noise [46–50]. However, our derivations can apply to arbitrary neural systems driven by white noise inputs, if *f-I* curves are interpreted as tuning functions with respect to the mean stimulus parameter. Given the generality of our results for neural systems, it would be interesting to test our results in cases where firing is driven by an external stimulus. A good candidate would be retinal ganglion cells, which are well-described by LN-type models [9,14,51–53], show adaptation to stimulus statistics on multiple timescales [23,54] and display a variety of dimensionalities in their feature space [14].

A limitation of the tests we have performed here is a restriction to the low firing rate regime where spike-triggered reverse correlation captures most of the dependence of firing probability on the stimulus. The effects of interspike interaction can be significant [16,17,55] and models with spike history feedback have

been developed to account for this [44,51,56,57]. We have not investigated how spike history effects would impact our results.

Although evidence suggests that gain modulation by noise may be enhanced by slow afterhyperpolarization currents underlying spike frequency adaptation [3], these slow currents are not required to generate gain enhancement in simple neuron models [7,19,25–29]. While one may generate diverse types of noise-induced gain modulation only by modifying the mechanism of generating a spike independent of spike history [7], in realistic situations, slow adaptation currents are present and will affect neural responses over many timescales [58–60]. In principle, it is possible to extend our result to include these effects: *f-I* curves under conditions of spike frequency adaptation have been already discussed [61–63] and can be compared to LN models with spike history feedback. However, our goal here was to demonstrate the effects that can occur independent of slow adaptation currents and before such currents have acted to shift neuronal coding properties.

The suggestive form of our result for one-dimensional LN models led us to look for a representation of neuronal output that is invariant under change in the input noise level. Our motivation is based on a simple principle of dimensional analysis: the gains of the *f-I* curves with noise may be asymptotically described by a signal-to-noise ratio, a dimensionless variable depending only on the stimulus itself. We showed that this may occur if the *f-I* curve with no noise obeys asymptotic power-law properties. Such a property has been determined to arise both from the bifurcation patterns of spike generation [31,34,35] and due to spike rate adaptation [61]. This relationship implies that the gain of the firing rate as a function of the mean should scale inversely with the standard deviation. Scaling of the gain of the nonlinear decision function with the stimulus standard deviation has been observed to some degree in a number of neural systems [10,15,22–25,29,64–67]. Such scaling guarantees maximal transmission of information [10,22]. As we and others have proposed, a static model might suffice to explain this phenomenon [25,27], although in some cases slow adaptation currents are known to contribute [65,66].

In summary, we have presented theoretically derived relationships between the variance-dependent gain modulation of *f-I* curves and intrinsic adaptation in neural coding. In real neural systems, any type of gain modulation likely results from many different mechanisms, possibly involving long-time scale dynamics. Our results show that observed forms of gain modulation may be a result of a pre-existing static nonlinearity that reacts to changes in the stimulus statistics robustly and almost instantaneously.

Materials and Methods

Biophysical Models

We used two single compartmental models with Hodgkin–Huxley (HH) active currents. The first one is an HH model with standard parameters while the second model (HHLS) has a lower Na^+ and higher K^+ maximal conductance. The voltage changes are described by [32]

$$C\frac{dV}{dt} = -\bar{g}_L(V-E_L) - \bar{g}_{Na}m^3h(V-E_{Na}) - \bar{g}_K n^4(V-E_K) + I(t)$$

and the activation variables m, n, and h behave according to

$$\tau_z(V)\frac{dz}{dt} = \bar{z}(V) - z, \quad \tau_z = \frac{1}{\alpha_z+\beta_z}, \quad \bar{z} = \frac{\alpha_z}{\alpha_z+\beta_z}, \quad z = m,n,h$$

where

$$\alpha_m = \frac{0.1(V+40)}{1-\exp[-0.1(V+40)]}, \quad \beta_m = 4\exp[-0.0556(V+65)],$$

$$\alpha_h = 0.07\exp[0.05(V+65)], \quad \beta_h = \frac{1}{1+\exp[-0.1(V+35)]},$$

$$\alpha_n = \frac{0.01(V+55)}{1-\exp[-0.1(V+55)]}, \quad \beta_n = 0.125\exp[-0.0125(V+65)]$$

The voltage V is in millivolts (mV).

For the HH model, the conductance parameters are $\bar{g}_K = 36$ mS/cm^2 and $\bar{g}_{Na} = 120$ mS/cm^2. The HHLS model has $\bar{g}_K = 41$ mS/cm^2 and $\bar{g}_{Na} = 79$ mS/cm^2. All other parameters are common to both models. The leak conductance is $\bar{g}_L = 0.3$ mS/cm^2 and the membrane capacitance per area C is 1 μF/cm^2. The reversal potentials are $E_L = -54.3$ mV, $E_{Na} = 50$ mV, and $E_K = -77$ mV. The membrane area is 10^{-3} cm^2, so that a current density of 1 μA/cm^2 corresponds to a current of 1 nA.

All simulations of these models were done with the NEURON simulation environment [68]. Gaussian white noise currents with various means and variances are generated with an update rate of 5 kHz ($dt = 0.2$ ms) and delivered into the model via current clamp. For the *f-I* curves, we simulated 4 min of input for each mean and variance pair. The whole procedure was repeated five times to estimate the variance of the *f-I* relationship, σ_{repeat}.

We ran another set of simulations for reverse correlation analysis and collected about 100,000 spikes for each stimulus condition. The means and variances of the Gaussian noisy stimuli were chosen such that the mean firing rate did not exceed 10 Hz, and we selected eight means and seven variances for the HH model, and nine means and four variances for the HHLS model.

Integrate-and-Fire-Type Models

In addition to the conductance-based model, we investigated the behavior of two heuristic model neurons driven by a noisy current input. Each model consists of a single dynamical equation describing voltage fluctuations of the form

$$C\frac{dV}{dt} = L(V) + I(t)$$

The first model is a leaky integrate-and-fire (LIF) model [69,70], for which $L(V) = -g_L(V-E_L)$. We used the parameters $g_L = 2$, $E_L = 0$, and $C = 1$. Since this choice of $L(V)$ cannot generate a spike, we additionally imposed a spiking threshold $V_{th} = 1$ and reset voltage $V_{\text{reset}} = -3$.

The second is a quadratic integrate-and-fire (QIF) model [31,37,38], for which $L(V) = g_L(V-E_L)(V-V_{th})/\Delta V$ where $\Delta V = V_{th} - E_L > 0$. We used $g_L = 0.5$, $E_L = 0$, $V_{th} = 0.1$, and $C = 1$. In this model, the voltage V can increase without bound; such a trajectory is defined to be a spike if it crosses $V_{\text{spike}} = 5$. After spiking, the system is reset to $V_{\text{reset}} = 0$.

These two models are simulated using a fourth-order Runge–Kutta integration method with an integration time step of $dt = 0.01$. The input current $I(t)$ was Gaussian white noise, updated at each time step, with a range of means and variances. The *f-I* curves were obtained from 1,000 s of stimulation for each (mean,variance) condition. We then compared the *f-I* curves from these models with the relationship derived in the Results section, Equation 5. A numerical solution of the partial differential equation was obtained using a PDE solver in Mathematica (Wolfram Research, Inc.).

Linear/Nonlinear Model

We use the linear/nonlinear (LN) cascade model framework to describe a neuron's input/output relation. We will focus on the dependence of the firing rate of a fixed LN model on the mean and variance of a Gaussian white noise input.

We will take the driving input to be $I(t) = I_0 + \xi(t)$ where I_0 is the mean and $\xi(t)$ is a Gaussian white noise with variance σ^2 and zero mean. The linear part of the model selects, by linear filtering, a subset of the possible stimuli probed by $I(t)$. That subset is expressed as n relevant features $\{\varepsilon_\mu(t)\}$, $(\mu = 1,2,\ldots,n)$. Interpreted as vectors, the components of any stimulus that are relevant to changing the firing rate can be expressed in terms of projections onto these features. The firing rate of the model for a given temporal sequence $I(t)$ depends only on $\mathbf{s}$, the input filtered by the n relevant features. Thus the firing rate from the given stimulus depends on the convolution of the input with all n features and can be written as $P(\text{spike}\,|\,\mathbf{s} = I_0\bar{\boldsymbol{\varepsilon}} + \mathbf{x})$ where

$$\bar{\varepsilon}_\mu = \int_0^\infty \mathrm{d}\tau\ \varepsilon_\mu(\tau), \quad x_\mu = \int_0^\infty \mathrm{d}\tau\, \varepsilon_\mu(\tau)\xi(t-\tau)$$

Since $I(t)$ is white noise with stationary statistics, the projections x_μ can be taken to be stationary random variables chosen from a Gaussian distribution at each t.

Given the filtered stimulus, a nonlinear decision function $P(\text{spike}\,|\,I_0\bar{\boldsymbol{\varepsilon}} + \mathbf{x})$ generates the instantaneous time-varying firing rate. For a specified model and stimulus statistics, the mean firing rate $f(I_0,\sigma^2) = P(\text{spike})$ is simply

$$f\left(I_0,\sigma^2\right) = \int \mathrm{d}\mathbf{s}\ P(\text{spike}|\mathbf{s})P(\mathbf{s}) = \int \mathrm{d}\mathbf{x}\ P(\text{spike}|I_0\bar{\boldsymbol{\varepsilon}} + \mathbf{x})p(\mathbf{x}) \quad (9)$$

where

$$p(\mathbf{x}) = \frac{1}{(2\pi\sigma^2)^{\pi/2}} \cdot \exp\left(-\frac{1}{2\sigma^2}\|\mathbf{x}\|^2\right)$$

Equation 9 describes an *f-I* curve of the model in the presence of added noise with variance σ^2. The slope or *gain* of the firing rate with respect to mean or variance can be computed if $P(\text{spike}\,|\,I_0\bar{\boldsymbol{\varepsilon}} + \mathbf{x})$ is known. However, the gains can be also obtained in terms of the first and second moments of $P(\text{spike}\,|\,I_0\bar{\boldsymbol{\varepsilon}} + \mathbf{x})$, which can be measured *directly* by reverse correlation analysis.

Reverse Correlation Analysis

We used spike-triggered reverse correlation to probe the computation of the model neurons through an LN model. We collected about 100,000 spikes and corresponding ensembles of spike triggered stimulus histories in a 30 ms long time window preceding each spike.

From the spike-triggered input ensembles, we calculated spike-triggered averages (STAs) and spike-triggered covariances (STCs). The STA is simply the average of the set of stimuli that led to spikes subtracted from the mean of the "prior" stimulus distribution, the distribution of all stimuli independent of spiking output

$$\text{STA}(t) = \langle I\left(t_{\text{spike}} - t\right)\rangle_{\text{spike}} - \langle I\rangle_{\text{prior}} = \langle \xi\left(t_{\text{spike}} - t\right)\rangle_{\text{spike}} \quad (10)$$

Therefore, one may consider only the noise part of the zero mean stimulus.

When computing the STC, the prior's covariance is subtracted

$$\begin{aligned}\Delta C(t,t') &= C_{\text{spike}} - C_{\text{prior}} \\ &= \langle \left\{\xi\left(t_{\text{spike}} - t\right) - \text{STA}(t)\right\} \\ &\quad \left\{\xi\left(t_{\text{spike}} - t'\right) - \text{STA}(t')\right\}\rangle_{\text{spike}} - C_{\text{prior}}\end{aligned} \quad (11)$$

Statistical Analysis

In calculating the slope and curvature of the *f-I* curves, we used 6–10 degree polynomial fitting of the *f-I* curves, where in any single case, the lowest degree was used which provided both a good fit and smoothness. From the fitting procedure, we obtained the standard deviation of the residuals, σ_{fit}. This was repeated five times for *f-I* curves computed using different noise samples, and from this we computed σ_{repeat}, the standard deviation of each computed slope and curvature. We estimated the total error of our calculation as $\sigma_{\text{total}} = (\sigma_{\text{repeat}}{}^2 + \sigma_{\text{fit}}{}^2)^{1/2}$. In practice, σ_{repeat} was always greater than σ_{fit} by an order of magnitude. This σ_{total} was used for the error bars in Figure 3.

To evaluate the goodness of fit in Figure 3, we used the Pearson χ^2 test by using the reduced χ^2 statistic

$$\chi^2 = \sum \frac{(O-E)^2}{\sigma_{\text{total}}^2}$$

where O and E represent the right and left hand sides of Equations 4–6, respectively. From this, the p-values are estimated from the cumulative density function of the χ^2 distribution, $Q(\chi^2/k,k)$. The degree of freedom is $k = 54$ and $k = 34$ for the HH and HHLS, respectively.

Derivation of Equations 4–6

We first present two key identities: the first one, which depends on the form of $\mathbf{s}$ having additive mean and noise components, is a change of variables for the gradient of $P(\text{spike}\,|\,\mathbf{x} + I_0\bar{\boldsymbol{\varepsilon}})$

$$\frac{\partial P(\text{spike}|\mathbf{x} + I_0\bar{\boldsymbol{\varepsilon}})}{\partial I_0} = \sum_\mu \bar{\varepsilon}_\mu \frac{\partial P(\text{spike}|\mathbf{x} + I_0\bar{\boldsymbol{\varepsilon}})}{\partial x_\mu} \quad (12)$$

Secondly, when x is a Gaussian random variable with zero mean and variance σ^2, by using integration by parts in can be seen that any function $F(x)$ satisfies

$$\langle F'(x)\rangle = \frac{1}{\sigma^2}\langle xF(x)\rangle \quad (13)$$

$$\begin{aligned}\langle F''(x)\rangle &= \frac{1}{\sigma^2}\langle [xF(x)]'\rangle - \frac{1}{\sigma^2}\langle F(x)\rangle \\ &= \frac{1}{\sigma^4}\langle x^2F(x)\rangle - \frac{1}{\sigma^2}\langle F(x)\rangle\end{aligned}$$

Then, we first take derivatives of both sides of Equation 9 (or equivalently Equation 1), by I_0 and σ^2, and apply Equations 12 and 13. The first order in I_0 is

$$\frac{\partial \log f}{\partial I_0} = \frac{1}{f}\frac{\partial f}{\partial I_0} = \frac{1}{f}\sum_{\mu}\bar{\varepsilon}_{\mu}\langle\frac{\partial}{\partial x_{\mu}}P(\text{spike}|\mathbf{x}+I_0\bar{\boldsymbol{\varepsilon}})\rangle_{\mathbf{x}} = \frac{1}{\sigma^2}\cdot\frac{1}{f}\sum_{\mu}\bar{\varepsilon}_{\mu}\langle x_{\mu}P(\text{spike}|\mathbf{x}+I_0\bar{\boldsymbol{\varepsilon}})\rangle_{\mathbf{x}} \quad (14)$$

The second order is given by

$$\frac{\partial^2 \log f}{\partial I_0^2} = \frac{1}{f}\frac{\partial^2 f}{\partial I_0^2} - \frac{1}{f^2}\left(\frac{\partial f}{\partial I_0}\right)^2,$$
$$\frac{\partial^2 f}{\partial I_0^2} = \sum_{\mu,\nu}\bar{\varepsilon}_{\mu}\bar{\varepsilon}_{\nu}\langle\frac{\partial}{\partial x_{\mu}}\frac{\partial}{\partial x_{\nu}}P(\text{spike}|\mathbf{x}+I_0\bar{\boldsymbol{\varepsilon}})\rangle_{\mathbf{x}} = \frac{1}{\sigma^4}\sum_{\mu,\nu}\bar{\varepsilon}_{\mu}\bar{\varepsilon}_{\nu}\langle\left(x_{\mu}x_{\nu}-\sigma^2\delta_{\mu\nu}\right)P(\text{spike}|\mathbf{x}+I_0\bar{\boldsymbol{\varepsilon}})\rangle_{\mathbf{x}}, \quad (15)$$

where $\delta_{\mu\nu}$ is a Kronecker delta symbol. The gain with respect to variance is

$$\frac{\partial f}{\partial \sigma^2} = -\frac{n}{2\sigma^2}f + \frac{1}{2\sigma^4}\sum_{\mu}\langle x_{\mu}^2 P(\text{spike}|\mathbf{x}+I_0\bar{\boldsymbol{\varepsilon}})\rangle_x = \frac{1}{2\sigma^4}\sum_{\mu}\langle\left(x_{\mu}^2-\sigma^2\right)P(\text{spike}|\mathbf{x}+I_0\bar{\boldsymbol{\varepsilon}})\rangle_{\mathbf{x}} \quad (16)$$

Now, we show how the right hand sides of Equations 14–16 correspond to the STA and the STC. Given a Gaussian white noise signal $\xi(t)$, we can split it as $\xi = \xi_{\parallel}+\xi_{\perp}$, where $\xi_{\parallel}$ belongs to the space spanned by our basis features $\{\varepsilon_{\mu}\}$, and therefore relevant to spiking. $\xi_{\perp}$ is the orthogonal or irrelevant part. $\xi_{\parallel}$ can be written as

$$\xi_{\parallel}(t) = \boldsymbol{\varepsilon}\cdot\mathbf{x} = \sum_{\mu}x_{\mu}\varepsilon_{\mu}, \quad x_{\mu} = \int_0^{\infty}\mathrm{d}\tau\,\varepsilon_{\mu}(\tau)\xi(t-\tau)$$

Again, $\mathbf{x}$ is a Gaussian variable from a distribution Equation 9.

The STA is

$$\text{STA} = \langle\xi\rangle_{\text{spike}} = \langle\xi_{\parallel}\rangle_{\text{spike}} = \int \mathrm{d}^n\mathbf{x}(\boldsymbol{\varepsilon}\cdot\mathbf{x})P(\mathbf{x}+I_0\bar{\boldsymbol{\varepsilon}}|\text{spike})$$

since $\xi_{\perp}$ is irrelevant and does not make any contribution. Here we use Bayes theorem

$$\frac{P(\text{spike}|\mathbf{x}+I_0\bar{\boldsymbol{\varepsilon}})}{P(\text{spike})} = \frac{P(\mathbf{x}+I_0\bar{\boldsymbol{\varepsilon}}|\text{spike})}{P(\mathbf{x}+I_0\bar{\boldsymbol{\varepsilon}})}$$

As in Equation 9, $P(\mathbf{s}=\mathbf{x}+I_0\bar{\boldsymbol{\varepsilon}}) = p(\mathbf{x})$, and therefore the STA becomes

$$\text{STA} = \int \mathrm{d}^n\mathbf{x}(\boldsymbol{\varepsilon}\cdot\mathbf{x})\frac{P(\text{spike}|\mathbf{x}+I_0\bar{\boldsymbol{\varepsilon}})}{P(\text{spike})}p(\mathbf{x}) = \frac{1}{f}\sum_{\mu}\varepsilon_{\mu}\langle x_{\mu}P(\text{spike}|\mathbf{x}+I_0\bar{\boldsymbol{\varepsilon}})\rangle_{\mathbf{x}}$$

Comparing this result with Equation 14, we obtain Equation 4.

A similar calculation for the second order [19] shows

$$\Delta C(t,t') = \frac{1}{f}\sum_{\mu,\nu}\varepsilon_{\mu}(t)\varepsilon_{\nu}(t')\langle\left(x_{\mu}x_{\nu}-\sigma^2\delta_{\mu\nu}\right)P(\text{spike}|\mathbf{x}+I_0\bar{\boldsymbol{\varepsilon}})\rangle_{\mathbf{x}} - \text{STA}(t)\cdot\text{STA}(t')$$

This result, combined with Equations 15 and 16, leads to Equations 5 and 6, respectively.

Supporting Information

Text S1. Firing Rate of the LIF Model with Noisy Stimuli. Found at: doi:10.1371/journal.pcbi.1000119.s001 (0.09 MB DOC)

Author Contributions

Conceived and designed the experiments: SH. Analyzed the data: SH. Contributed reagents/materials/analysis tools: BL. Wrote the paper: SH BL AF. Derived the equations: SH.

References

1. Chance FS, Abbott LF, Reyes AD (2002) Gain modulation from background synaptic input. Neuron 35: 773–782.
2. Fellous JM, Rudolph M, Destexhe A, Sejnowski TJ (2003) Synaptic background noise controls the input/output characteristics of single cells in an in vitro model of in vivo activity. Neuroscience 122: 811–829.
3. Higgs MH, Slee SJ, Spain WJ (2006) Diversity of gain modulation by noise in neocortical neurons: regulation by the slow afterhyperpolarization conductance. J Neurosci 26: 8787–8799.
4. Arsiero M, Lüscher HR, Lundstrom BN, Giugliano M (2007) The impact of input fluctuations on the frequency–current relationships of layer 5 pyramidal neurons in the rat medial prefrontal cortex. J Neurosci 27: 3274–3284.
5. Prescott SA, Koninck YD (2003) Gain control of firing rate by shunting inhibition: roles of synaptic noise and dendritic saturation. Proc Natl Acad Sci USA 100: 2076–2081.
6. Prescott SA, Ratté S, De Koninck Y, Sejnowski TJ (2006) Nonlinear interaction between shunting and adaptation controls a switch between integration and coincidence detection in pyramidal neurons. J Neurosci 26: 9084–9097.
7. Lundstrom BN, Hong S, Fairhall AL (2008) Two computational regimes of a single-compartment neuron separated by a planar boundary in conductance space. Neural Comput 20: 1239–1260.
8. Victor J, Shapley R (1980) A method of nonlinear analysis in the frequency domain. Biophys J 29: 459–483.
9. Meister M, Berry II MJ (1999) The neural code of the retina. Neuron 22: 435–450.
10. Brenner N, Bialek W, de Ruyter van Steveninck R (2000) Adaptive rescaling maximizes information transmission. Neuron 26: 695–702.
11. Simoncelli EP, Paninski L, Pillow J, Schwartz O (2004) Characterization of neural responses with stochastic stimuli. In: Gazzaniga M, ed. The Cognitive Neurosciences. 3rd edition. Cambridge (Massachusetts): MIT Press.
12. Rust NC, Schwartz O, Movshon JA, Simoncelli EP (2005) Spatiotemporal elements of macaque V1 receptive fields. Neuron 46: 945–956.
13. Stanley GB, Lei FF, Dan Y (1999) Reconstruction of natural scenes from ensemble responses in the lateral geniculate nucleus. J Neurosci 19: 8036–8042.
14. Fairhall AL, Burlingame CA, Narasimhan R, Harris RA, Puchalla JL, et al. (2006) Selectivity for multiple stimulus features in retinal ganglion cells. J Neurophysiol 96: 2724–2738.
15. Maravall M, Petersen RS, Fairhall AL, Arabzadeh E, Diamond ME (2007) Shifts in coding properties and maintenance of information transmission during adaptation in barrel cortex. PLoS Biol 5: e19. doi:10.1371/journal.pbio.0050019.
16. Agüera y Arcas B, Fairhall AL (2003) What causes a neuron to spike? Neural Comput 15: 1715–1749.
17. Agüera y Arcas B, Fairhall AL, Bialek W (2003) Computation in a single neuron: Hodgkin and Huxley revisited. Neural Comput 15: 1789–1807.

18. Slee SJ, Higgs MH, Fairhall AL, Spain WJ (2005) Two-dimensional time coding in the auditory brainstem. J Neurosci 25: 9978–9988.
19. Hong S, Agüera y Arcas B, Fairhall AL (2007) Single neuron computation: from dynamical system to feature detector. Neural Comput 19: 3133–3172.
20. Atick JJ (1992) Could information theory provide an ecological theory of sensory processing? Network (Bristol, England) 3: 213–251.
21. Theunissen FE, Sen K, Doupe AJ (2000) Spectral-temporal receptive fields of nonlinear auditory neurons obtained using natural sounds. J Neurosci 20: 2315–2331.
22. Fairhall A, Lewen G, Bialek W, de Ruyter van Steveninck RR (2001) Efficiency and ambiguity in an adaptive neural code. Nature 412: 787–792.
23. Baccus SA, Meister M (2002) Fast and slow contrast adaptation in retinal circuitry. Neuron 36: 909–919.
24. Nagel KI, Doupe AJ (2006) Temporal processing and adaptation in the songbird auditory forebrain. Neuron 51: 845–859.
25. Gaudry KS, Reinagel P (2007) Benefits of contrast normalization demonstrated in neurons and model cells. J Neurosci 27: 8071–8079.
26. Rudd ME, Brown LG (1997) Noise adaptation in integrate-and-fire neurons. Neural Comput 9: 1047–1069.
27. Paninski L, Lau B, Reyes AD (2003) Noise-driven adaptation: in vitro and mathematical analysis. Neurocomputing 52: 877–883.
28. Yu Y, Lee TS (2003) Dynamical mechanisms underlying contrast gain control in single neurons. Phys Rev E 68: 011901.
29. Borst A, Flanagin VL, Sompolinsky H (2005) Adaptation without parameter change: dynamic gain control in motion detection. Proc Natl Acad Sci USA 102: 6172–6176.
30. Hodgkin AL (1948) The local electric changes associated with repetitive action in a non-medullated axon. J Physiol 107: 165–181.
31. Izhikevich EM (2006) Dynamical Systems in Neuroscience: The Geometry of Excitability and Bursting. Cambridge (Massachusetts): MIT Press.
32. Hodgkin AL, Huxley AF (1952) A quantitative description of membrane current and its application to conduction and excitation in nerve. J Physiol 463: 391–407.
33. Spekreijse H, Reits D (1982) Sequential analysis of the visual evoked potential system in man: nonlinear analysis of a sandwich system. Ann N Y Acad Sci 388: 72–97.
34. Ermentrout GB (1994) Reduction of conductance-based models with slow synapses to neural nets. Neural Comput 6: 679–695.
35. Rinzel JM, Ermentrout GB (1989) Analysis of neuronal excitability. In: Koch C, Segev I, eds. Methods in Neuronal Modeling: From Synapses to Networks. Cambridge (Massachusetts): MIT Press. pp 135–170.
36. Hoppensteadt F, Izhikevich EM (1997) Weakly Connected Neural Nets. Berlin: Springer-Verlag.
37. Ermentrout GB, Kopell N (1986) Parabolic bursting in an excitable system coupled with a slow oscillation. SIAM J Appl Math 4: 233–253.
38. Ermentrout B (1996) Type I membranes, phase resetting curves, and synchrony. Neural Comput 8: 979–1001.
39. Weber EH (1834) De Pulsu, Resorptione, Auditu et Tactu. Annotiones Anatomicae et Physiologicae. Lipsiae: Koehler.
40. Fechner G (1966) Elements of Psychophysics. New York: Holt, Rinehart and Winston.
41. Fitts PM (1954) The information capacity of the human motor system in controlling the amplitude of movement. J Exp Psychol 47: 381–391.
42. Stevens SS (1986) Psychophysics: Introduction to Its Perceptual, Neural, and Social Prospects. Piscataway (New Jersey): Transaction Publishers.
43. Barenblatt GI (2003) Scaling. Cambridge, UK: Cambridge University Press.
44. Gerstner W, Kistler W (2002) Spiking Neuron Models: Single Neurons, Populations, Plasticity. Cambridge, UK: Cambridge University Press.
45. Lundstrom BN, Fairhall AL (2006) Decoding stimulus variance from a distributional neural code of interspike intervals. J Neurosci 26: 9030–9037.
46. Gerstein GL, Mandelbrot B (1964) Random walk models for the spike activity of a single neuron. Biophys J 4: 41–68.
47. Bryant HL, Segundo JP (1976) Spike initiation by transmembrane current: a white-noise analysis. J Physiol 260: 279–314.
48. Mainen ZF, Sejnowski TJ (1995) Reliability of spike timing in neocortical neurons. Science 268: 1503–1506.
49. Destexhe A, Paré D (1999) Impact of network activity on the integrative properties of neocortical pyramidal neurons in vivo. J Neurophysiol 81: 1531–1547.
50. Rudolph M, Destexhe A (2003) Characterization of subthreshold voltage fluctuations in neuronal membranes. Neural Comput 15: 2577–2618.
51. Keat J, Reinagel P, Reid RC, Meister M (2001) Predicting every spike: a model for the responses of visual neurons. Neuron 30: 803–817.
52. Chichilnisky EJ (2001) A simple white noise analysis of neuronal light responses. Network (Bristol, England) 12: 199–213.
53. Pillow JW, Paninski L, Uzzell VJ, Simoncelli EP, Chichilnisky EJ (2005) Prediction and decoding of retinal ganglion cell responses with a probabilistic spiking model. J Neurosci 25: 11003–11013.
54. Smirnakis SM, Berry MJ, Warland DK, Bialek W, Meister M (1997) Adaptation of retinal processing to image contrast and spatial scale. Nature 386: 69–73.
55. Pillow JW, Simoncelli EP (2003) Biases in white noise analysis due to non-Poisson spike generation. Neurocomputing 52–54: 109–115.
56. Truccolo W, Eden UT, Fellows MR, Donoghue JP, Brown EN (2005) A point process framework for relating neural spiking activity to spiking history, neural ensemble, and extrinsic covariate effects. J Neurophysiol 93: 1074–1089.
57. Paninski L, Pillow J, Lewi J (2006) Statistical models for neural encoding, decoding, and optimal stimulus design. Prog Brain Res 165: 493–507.
58. Schwindt PC, Spain WJ, Foehring RC, Stafstrom CE, Chubb MC, et al. (1988) Multiple potassium conductances and their functions in neurons from cat sensorimotor cortex in vitro. J Neurophysiol 59: 424–449.
59. Spain WJ, Schwindt PC, Crill WE (1991) Two transient potassium currents in layer V pyramidal neurones from cat sensorimotor cortex. J Physiol 434: 591–607.
60. La Camera G, Rauch A, Thurbon D, Lüscher HR, Senn W, et al. (2006) Multiple time scales of temporal response in pyramidal and fast spiking cortical neurons. J Neurophysiol 96: 3448–3464.
61. Ermentrout B (1998) Linearization of F-I curves by adaptation. Neural Comput 10: 1721–1729.
62. Benda J, Herz AVM (2003) A universal model for spike-frequency adaptation. Neural Comput 15: 2523–2564.
63. La Camera G, Rauch A, Lüscher HR, Senn W, Fusi S (2004) Minimal models of adapted neuronal response to in vivo-like input currents. Neural Comput 16: 2101–2124.
64. Kim KJ, Rieke F (2001) Temporal contrast adaptation in the input and output signals of salamander retinal ganglion cells. J Neurosci 21: 287–299.
65. Arganda S, Guantes R, de Polavieja GG (2007) Sodium pumps adapt spike bursting to stimulus statistics. Nat Neurosci 10: 1467–1473.
66. Diaz-Quesada M, Maravall M (2008) Intrinsic mechanisms for adaptive gain rescaling in barrel cortex. J Neurosci 28: 696–710.
67. Ringach DL, Malone BJ (2007) The operating point of the cortex: neurons as large deviation detectors. J Neurosci 27: 7673–7683.
68. Hines ML, Carnevale NT (1997) The NEURON simulation environment. Neural Comput 9: 1179–1209.
69. Knight BW (1972) Dynamics of encoding in a population of neurons. J Gen Physiol 59: 734.
70. Tuckwell HC (1988) Introduction to Theoretical Neurobiology. Cambridge, UK: Cambridge University Press.

ARTICLE Communicated by Tatyana Sharpee

Single Neuron Computation: From Dynamical System to Feature Detector

Sungho Hong
shhong@u.washington.edu
Department of Physiology and Biophysics, University of Washington, Seattle, WA 98195, U.S.A.

Blaise Agüera y Arcas
blaisea@microsoft.com
Program in Applied and Computational Mathematics, Princeton University, Princeton, NJ 08544

Adrienne L. Fairhall
fairhall@u.washington.edu
Department of Physiology and Biophysics, University of Washington, Seattle, WA 98195, U.S.A.

White noise methods are a powerful tool for characterizing the computation performed by neural systems. These methods allow one to identify the feature or features that a neural system extracts from a complex input and to determine how these features are combined to drive the system's spiking response. These methods have also been applied to characterize the input-output relations of single neurons driven by synaptic inputs, simulated by direct current injection. To interpret the results of white noise analysis of single neurons, we would like to understand how the obtained feature space of a single neuron maps onto the biophysical properties of the membrane, in particular, the dynamics of ion channels. Here, through analysis of a simple dynamical model neuron, we draw explicit connections between the output of a white noise analysis and the underlying dynamical system. We find that under certain assumptions, the form of the relevant features is well defined by the parameters of the dynamical system. Further, we show that under some conditions, the feature space is spanned by the spike-triggered average and its successive order time derivatives.

1 Introduction

A primary goal of sensory neurophysiology is to understand how stimuli from the natural environment are encoded in the spiking output of

neurons. A useful tool for performing this characterization is white noise analysis (Marmarelis & Marmarelis, 1974; Wiener, 1958; Rieke, Warland, Bialek, & de Ruyter van Steveninck, 1997), whereby a system is stimulated with a randomly varying broadband signal and the relevant features driving the system are determined by correlating output spikes with the signal. Extensions of this analysis to second order have been used to find multiple stimulus features to which the system is sensitive (de Ruyter van Steveninck & Bialek, 1988; Brenner, Bialek, & de Ruyter van Steveninck, 2000; Bialek & de Ruyter van Steveninck, 2005). Recently, these methods have been applied to characterize the computation performed by a single neuron on its current inputs (Agüera y Arcas, Fairhall, & Bialek, 2003; Agüera y Arcas & Fairhall, 2003; Slee, Higgs, Fairhall, & Spain, 2005).

Due to detailed knowledge of the dynamics of ion channels, it is possible to build dynamical models of current flow in neurons that accurately reproduce the voltage response of single neurons to current (or conductance) inputs. White noise analysis provides us with the capacity to reduce this complex set of dynamical equations to a simple functional model with concrete components: linear feature selection followed by a nonlinear decision function generating spikes. This procedure is compelling as it provides a clear intuition about the form of the changes in the stimulus to which the system is sensitive and how the system combines these features in the decision to fire. Second order white noise methods have provided insight into coding properties at the systems level, in visual (Brenner et al., 2000; Fairhall et al., 2006; Touryan, Lau, & Dan, 2002; Rust, Schwartz, Movshon, & Simoncelli, 2005; Horwitz, Chichilnisky, & Albright, 2005) and somatosensory (Petersen, 2003; Maravall, Peterson, Fairhall, Arabzadeh, & Diamond, 2007) cortex. More recently, these methods have also been applied to characterize neural coding in single auditory (Slee et al. 2005), central (Powers, Dai, Bell, Percival, & Binder, 2005), and model (Agüera y Arcas, et al. 2003; Agüera y Arcas & Fairhall, 2003) neurons. However, mostly missing from these analyses is a clear link between the obtained stimulus features and the biophysical properties of the circuit or neuron. This issue is particularly well posed for single neurons under stimulation by direct somatic current injection, where the biophysical parameters of the soma must dominate the neuron's computation. Thus, given the power of white noise methods, the questions that we wish to address here are: How does the dynamical system governing neural behavior map to the neuron's functional characterization derived using white noise analysis? How are the features determined by the neuron's biophysical properties?

1.1 Minimal Spiking Models. Neural dynamics are typically described by conductance-based models that describe the temporal evolution of the

voltage V due to an input current I and the variable ionic conductances of ion channels:

$$C dV/dt = -\sum_i g_i(V)(V - E_i) + I, \tag{1.1}$$

where C is the membrane capacitance. Each conductance type i has a reversal potential E_i, where the conductances g_i corresponding to different ion channels may be voltage dependent through the dynamics of the corresponding activation and inactivation gating variables ξ_i and ϕ_i:

$$g_i = \bar{g}_i \xi_i^{p_i} \phi_i^{q_i} \tag{1.2}$$

$$d\xi_i/dt = f_{\xi_i}(\xi, V) \tag{1.3}$$

$$d\phi_i/dt = f_{\phi_i}(\phi, V), \tag{1.4}$$

where $\bar{g}_i$ are constants and the functions f_ξ and f_ϕ, are affine in ξ and ϕ, respectively, but not in V. The exponents p_i and q_i are usually taken to be integers and describe the cooperativity of molecules required to gate the ion channel. This set of equations is highly nonlinear due to this cooperativity and the voltage dependence of the conductances. While ultimately we would like to understand systems of many dynamical variables, we will begin by analyzing a simplified spiking system: the FitzHugh-Nagumo (FN) neural model (FitzHugh, 1961; Nagumo, Arimoto, & Yoshikawa, 1962). The FitzHugh-Nagumo system is cubic in V, the lowest-order nonlinearity supporting spike-like behavior, and is defined by

$$\psi\, dV/dt = V(1 - V)(a + V) - W + c + I, \tag{1.5}$$

$$dW/dt = V - bW, \tag{1.6}$$

where V is voltage, W is a generic inactivation variable with linear recovery dynamics, I is the input current, and a, b, c, and ψ are parameters, $\psi \ll 1$. The nullclines of the system are the curves along which $dV/dt = 0$ (the V nullcline, a cubic) and $dW/dt = 0$ (the W nullcline, a straight line). Figure 1a shows the nullclines for zero input I on the phase plane. Trajectories have been traced forward and backward in time through a randomly chosen set of initial conditions. The unique intersection of the nullclines is the stable fixed point. Elsewhere in the plane, trajectories always spiral counterclockwise, but qualitatively, there are two kinds of orbits, ultimately both converging to the stable fixed point. Subthreshold orbits make only small excursions; however, for trajectories starting with either low inactivation (roughly $W < -.45$) or high voltage (for example, $V > 0$), the system makes a large excursion around the right branch of the V nullcline, corresponding to a spike, before settling back to the fixed point. Although

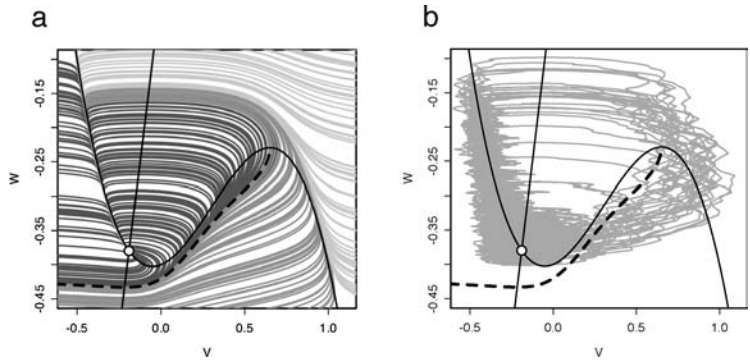


Figure 1: (a) Phase plane of the FitzHugh-Nagumo neuron model with trajectories and threshold (dashed line). Parameters are $a = 0.1$, $b = 0.5$, $c = -0.4$, and $\psi = 0.015$. (b) A single extended trajectory on the FN phase plane driven by white noise input, showing multiple spikes. The threshold (dashed line) still makes an approximate boundary of the spiking and nonspiking portions of the trajectory (gray).

we will generally be considering a time-varying input $I(t)$, we will make heavy use of the $I = 0$ phase plane; one can regard this as representing the instantaneous flow if the input current is switched off. For $I = 0$, one can define a separation of spiking initial conditions from subthreshold conditions, marked in light versus dark gray in Figure 1a. This separation of behavior can be thought of as a threshold. Typically the threshold for spiking is taken to be a threshold in voltage only, but this phase-plane picture demonstrates that one should really consider the threshold to depend on both V and the hidden inactivation variable(s) (Izhikevich, 2006).

One can approximate this "dynamical" threshold by considering a minimal spiking trajectory. There is no absolute distinction between spiking and subspiking trajectories in type II neurons (Gerstner & Kistler, 2002) such as Hodgkin-Huxley and FitzHugh-Nagumo, so this choice is necessarily somewhat arbitrary. Starting with a point on the phase plane at the end of the spike plateau, one can eliminate time from the dynamical equations, solving parametrically for the curve passing through this point. Here, we obtained one such curve by numerically solving the equations from the local maximum of the cubic nullcline, given by $\frac{d}{dV} V(1 - V)(a + V) = 0$.

One of the issues we explore in this article is the relationship between the geometry of the threshold in the multidimensional space of the dynamical variables and the measurements one makes in white noise analysis.

1.2 White Noise Analysis. Our approach is based on the functional perturbation expansion method, which we briefly review here. The simplest such expansion is known as the Volterra series (Volterra, 1930; Marmarelis,

& Marmarelis, 1974; Marmarelis, 2004), where the functional $y(t) = y[I(t)]$ is expanded as

$$y(t) = h_0 + \int dt_1\, h_1(t_1) I(t - t_1) + \frac{1}{2!} \int dt_1 dt_2\, h_2(t_1, t_2) I(t - t_1) I(t - t_2) + \cdots, \tag{1.7}$$

where the kernels h_i are called Volterra kernels. One problem with Volterra analysis is that the successive order kernels are not independent. For example, h_2 contains a constant component $\sim \int dt_1\, h_2(t_1, t_1)$.

Wiener analysis (Wiener, 1958; Rieke et al., 1997) avoids this issue. In this framework, $I(t)$ is assumed to be zero mean gaussian white noise satisfying

$$\langle I(t) I(t') \rangle = \sigma^2 \delta(t - t').$$

Under Wiener analysis, $y[I(t)]$ is expanded as

$$y(t) = g_0 + \int dt_1\, g_1(t_1) I(t - t_1) + \frac{1}{2!} \left\{ \int dt_1 dt_2\, g_2(t_1, t_2) I(t - t_1) I(t - t_2) - \sigma^2 \int dt'\, g_2(t', t') \right\} + \cdots. \tag{1.8}$$

In the Wiener expansion, each kernel is independent of the others, since at each order, one subtracts out the lower-order components. Therefore, the kernels g_i can be easily obtained by measuring the correlation functions between $y(t)$ and $I(t)$—for example,

$$\begin{aligned} \langle y(t) \rangle &= g_0 + \int dt_1\, g_1(t_1) \langle I(t - t_1) \rangle \\ &\quad + \frac{1}{2!} \left\{ \int dt_1 dt_2\, g_2(t_1, t_2) \langle I(t - t_1) I(t - t_2) \rangle \right. \\ &\quad \left. - \sigma^2 \int dt'\, g_2(t', t') \right\} + \cdots \\ &= g_0, \\ \langle y(t) I(t - t_1) \rangle &= g_1(t_1)/\sigma^2, \\ \langle y(t) I(t - t_1) I(t - t_2) \rangle &= g_2(t_1, t_2)/\sigma^4, \end{aligned}$$

and so on. The correlation method can be used to determine the relationship between Wiener and Volterra kernels. From equation 1.8,

$$\begin{aligned} g_0 &= h_0 + \int dt_1\, h_1(t) \left\langle I(t-t_1) \right\rangle \\ &\quad + \frac{1}{2!} \int dt_1 dt_2\, h_2(t_1, t_2) \left\langle I(t-t_1) I(t-t_2) \right\rangle + \cdots \\ &= h_0 + \frac{\sigma^2}{2!} \int dt'\, h_2(t', t') + \cdots, \\ g_1(t) &= h_1(t) + \frac{\sigma^2}{2!} \int dt' h_3(t', t', t) + \cdots, \end{aligned}$$

and so on.

Wiener analysis has been successfully used to determine the relevant feature space for neural systems by reverse correlation of the spiking output with white noise input. For this application, the output is taken to be the sequence of spike times, $\rho(t) = \sum_i \delta(t - t_i)$. Let us define the stimulus $\mathbf{s}$ as a linear transformation of the current input $I(t)$,

$$s_i = \int dt'\, f_i(t') I(t - t'), \tag{1.9}$$

where f_i are some set of independent linear filters. The goal is to determine how a single spike is generated by I, reduced to a vector $\mathbf{s}$ comprising projections along the relevant dimensions f_i.

We need to compute the probability of spiking as a function of the input, $P\left(\text{spike}|\mathbf{s}\right)$,

$$P\left(\text{spike}|\mathbf{s}\right) = P(\text{spike}) G(\mathbf{s}), \tag{1.10}$$

where $G(\mathbf{s})$ is the input-output function relating $\rho(t)$ to the time-varying stimulus $\mathbf{s}(t)$, and we have separated out the scale factor $P(\text{spike})$ proportional to the mean firing rate. In principle, since the equations of motion are deterministic, $P\left(\text{spike}|\mathbf{s}\right)$ is a Boolean function. However, as we will be approximating $\mathbf{s}$ by a finite-dimensional truncation and will be solving for P itself in successive orders, the result will no longer be Boolean.

We can compute $P\left(\text{spike}|\mathbf{s}\right)$ using Bayes' theorem (Rieke et al., 1997),

$$G(\mathbf{s}) = \frac{P\left(\text{spike}|\mathbf{s}\right)}{P(\text{spike})} = \frac{P\left(\mathbf{s}|\text{spike}\right)}{P(\mathbf{s})}. \tag{1.11}$$

This shows that the Wiener kernels of $G(\mathbf{s})$ are the moments of $P\left(\mathbf{s}|\text{spike}\right)$, which can be easily measured. In this article, we focus on the first two moments of this distribution.[1] The first moment is the spike-triggered average (STA), $\bar{s}(t)$,

$$\bar{s}(t) = \left\langle I(t_i - t)\right\rangle_i = \int d\mathbf{s}\ \mathbf{s} P\left(\mathbf{s}|\text{spike}\right) = \int d\mathbf{s}\ \mathbf{s} P(\mathbf{s}) G(\mathbf{s}), \tag{1.12}$$

where the average $\langle\cdot\rangle_i$ over the spike occurrence times $\{t_i\}$ is replaced with an average over the spike-triggering ensemble. When the neuron has multiple characteristic features, the STA may contain little information about them since it points in only a single direction, which is a certain linear combination of relevant features. The second-order kernel, however, can give such information. This kernel is related to the spike-triggered covariance (STC) matrix C, which we will define as

$$\begin{aligned} C(t,t') &= \left\langle I(t_i - t) - \bar{s}(t),\, I(t_i - t') - \bar{s}(t')\right\rangle_i - \left\langle I(\tau - t) I(\tau - t')\right\rangle_\tau \\ &= \sigma^2 g_0 \delta(t,t') + \sigma^4 g_2(t,t') - \bar{s}(t)\bar{s}(t') - \sigma^2\delta(t,t') \\ &= \sigma^4 g_2(t,t') - \bar{s}(t)\bar{s}(t'). \end{aligned} \tag{1.13}$$

$g_2(t,t')$ is the second-order Wiener kernel of $G(\mathbf{s})$. When a functional $F[\mathbf{s}]$ depends on gaussian noise $\mathbf{s}$, $\left\langle s_i F[\mathbf{s}]\right\rangle = \sum_j \left\langle s_i s_j\right\rangle \left\langle \partial_j F[\mathbf{s}]\right\rangle$. Thus, we can expand g_2 (Bialek & de Ruyter van Steveninck, 2005),

$$g_2(t,t') = \sum_{i,j} \left\langle \frac{\partial^2 G}{\partial s_i \partial s_j}\right\rangle f_i(t) f_j(t'), \tag{1.14}$$

where equation 1.9 introduces the linear filters that define the components of $\mathbf{s}$.

Therefore, unless the modified Hessian $\hat{H}_{ij} = \langle \frac{\partial^2}{\partial s_i \partial s_j} G\rangle - \langle \frac{\partial}{\partial s_i} G\rangle \langle \frac{\partial}{\partial s_j} G\rangle$ has null directions,[2] the STC is spanned by the relevant feature space $\{f_i\}$. Diagonalizing the STC and extracting the eigenvectors with the largest absolute eigenvalue will determine the leading dimensions spanning the feature space (Brenner et al., 2000; Agüera y Arcas et al., 2003; Bialek & de Ruyter van Steveninck, 2005). Eigenmodes appearing with negative eigenvalue correspond to stimulus directions in which the variance of the

[1] Note that the zeroth-order kernel is simply $g_0 = 1$.

[2] Since $\bar{s}(t)\bar{s}(t')$ is only rank one, there are only two possibilities leading to null directions. The first is when $\langle \partial_i \partial_j G\rangle$ has null directions, in which case it is excluded (for our purposes) by definition. The second case is when $\langle \partial_i \partial_j G\rangle$ has an eigenvector $\bar{s}(t)$ with an eigenvalue $||\bar{s}\,(t)||^2$, which does not happen generically.

spike-conditional distribution is reduced relative to the stimulus prior distribution; eigenmodes with positive eigenvalues are directions with increased variance. Such cases may arise if the spike-conditional distribution is bimodal in some direction or if it forms a ring around the origin; phase-invariant (complex) cells are an example (Simoncelli, Pillow, Paninski, & Schwartz, 2004; Rust et al., 2005; Fairhall et al., 2006; Petersen, 2003; Maravall et al., 2007).

In this article, we work with a multidimensional space of dynamical variables, as in equations 1.1 to 1.6, for which we obtain the Volterra series. When we denote them with y_i, the Wiener kernels for $G(y_i)$ can be expanded as

$$g_1[G] = \sum_i \frac{\partial G}{\partial y_i} g_1[y_i] + \cdots,$$
$$g_2[G] = \sum_i \frac{\partial G}{\partial y_i} g_2[y_i] + \sum_{i,j} \frac{\partial^2 G}{\partial y_i \partial y_j} g_1[y_i] g_1[y_j] + \cdots$$
$$\cdots$$

where $g_n[y_i]$ is an nth Wiener kernel of y_i and dots represent higher-order terms than quadratic. Here we note that $g_n[y_i]$ is generated by the infinite series of Volterra kernels. Moreover, we will see later that the Volterra kernels at each order can be constructed from the set of first-order kernels.

2 Linear Systems

In section 1, we discussed the application of white noise methods to experimental data where the output of the system is measured as the firing times of spikes. With access to the entire dynamical system, one could attempt to model the relationship between the input $I(t)$ and the output voltage $V(t)$, including both subthreshold and spiking behavior. As we have discussed, neurons are highly nonlinear. Perturbative expansions such as Wiener and Volterra series generally are unable to capture the nonlinearities of spike generation with few terms. However, a very successful and highly influential approach to the modeling of neural systems has been to separate a linear filtering stage from an explicit nonlinearity, capturing the very sharp voltage increase of the spike (Victor & Shapley, 1980; Gerstner & Kistler, 2002; Kistler, Gerstner, & van Hemmen, 1997; Berry & Meister, 1999). This is known as a cascade model, and in various forms is the theoretical basis for many of the simple neuron models in use, including integrate-and-fire neuron (Gerstner & Kistler, 2002; Agüera y Arcas & Fairhall, 2003), the spike response model (Gerstner & Kistler, 2002), generalized integrate-and-fire models (Paninski, Pillow, & Simoncelli, 2004) and cascade models of retinal ganglion cells (Victor & Shapley, 1979, 1980; Keat, Reinagel, Reid, & Meister, 2001; Pillow, Paninski, Uzzell, Simoncelli, & Chichilnisky, 2005).

The success of these models suggests that a breakdown of the dynamical system into its linear component and a threshold is a useful approximation to examine. In our analysis of the dynamical system, we will take the approach of experimental work, where the output is reduced to a sequence of spike times. Our goal here is to derive the form of the multidimensional linear stage that arises directly from the equations of motion and examine the consequences of thresholds of various forms.

We begin by walking through some simple linearized systems analysis and clarify the connection with the cascade model picture. Rewriting equations 1.1 to 1.4 in the following simple form,

$$\partial_t y_k = f_k(y_1, y_2, \ldots, y_N) + I(t)\delta_{k0}, \quad k = 0, 1, \cdots, n-1, \tag{2.1}$$

we assume that they have a fixed-point solution $y_k(t) = y_k^{(0)}$ when $I(t) = 0$. Now the linear approximation of the system around the fixed point is given by[3]

$$\mathcal{D}_{km} y_m^{(1)} = I(t)\delta_{k0}, \tag{2.2}$$

where

$$\mathcal{D}_{km} = \delta_{km}\partial_t - J_{km}, \quad J_{km} = \left.\frac{\partial f_k}{\partial y_m}\right|_{y^{(0)}}. \tag{2.3}$$

The linear filters are the kernels of equation 2.3, $\epsilon^{(1)} = \mathcal{D}^{-1}$. They satisfy

$$\mathcal{D}_{km}\epsilon_{ml}^{(1)}(t) = \delta_{kl}\delta(t), \tag{2.4}$$

giving the linear filter form for the system's evolution:

$$y_k^{(1)}(t) = \int_{-\infty}^{t} dt'\, \epsilon_k^{(1)}(t - t')I(t'), \tag{2.5}$$

where $\epsilon_k^{(1)} = \epsilon_{k0}^{(1)}$. Equation 2.4 can be solved by diagonalizing the Jacobian matrix J_{km}. Letus consider the following diagonalization with eigenvalues λ_α:

$$\lambda_\alpha \delta_{\alpha\beta} = U_{\alpha k} J_{km} U^{-1}_{m\beta}. \tag{2.6}$$

[3] From here on, we use the Einstein summation convention, for example, $C_{km}d_m \equiv \sum_m C_{km}d_m$.

Here we employ the notation that indices in the diagonalized basis are Greek letters. Therefore, equation 2.4 can be written in the new basis as

$$\mathcal{D}_\alpha \tilde{\epsilon}^{(1)}_\alpha = U_{\alpha 0}\delta(t), \quad \mathcal{D}_\alpha = \partial_t - \lambda_\alpha, \tag{2.7}$$

where the first-order kernels $\tilde{\epsilon}^{(1)}$ in this basis are

$$\tilde{\epsilon}^{(1)}_\alpha(t) = e^{\lambda_\alpha t} H(t),$$

where $H(t)$ is the Heaviside step function and arises from causality. The eigenvalues λ_α can be real or complex. The sign of the real part of the eigenvalues is determined by the stability properties of the fixed point at $y^{(0)}$. For a neural system, this fixed point should be attracting, so all eigenvalues have negative real part, and thus eigenvectors decay as $t \to \infty$. This underscores that the linear filters can take only a limited number of forms: decaying exponentials or damped oscillations.

We note the following facts. First, in general, an n-dimensional system will have n independent first-order kernels, except when the Jacobian has a degenerate spectrum. Further, the linear kernels can be generated from a "master kernel,"

$$\varphi(t) = \sum_\alpha e^{\lambda_\alpha t} H(t),$$

and its time derivatives up to nth order. To show this, from the expression of the derivative of a master kernel,

$$\partial_t{}^k \varphi(t) = \sum_\alpha \lambda_\alpha^k e^{\lambda_\alpha t} = T_{k\alpha} e^{\lambda_\alpha t}, \quad T_{k\alpha} = \lambda_\alpha^k, \quad (k, \alpha = 0, \ldots, n-1),$$

up to additive singular terms coming from the derivatives of $H(t)$. $T_{k\alpha}$ is a Wronskian matrix of exponentials at $t = 0$. The determinant of $T_{k\alpha}$ is the Vandermonde determinant $\det(T_{k\alpha}) = \prod_{\alpha>\beta}(\lambda_\alpha - \lambda_\beta)$, which cannot vanish by definition (Gradshteyn & Ryzhik, 2000). This means that the master kernel and its derivatives can be transformed to $e^{\lambda_\alpha t}$'s in a nonsingular way.

Therefore, we conclude that all first-order kernels can be expressed in terms of a master kernel and its time derivatives. Frequently the first-order kernel is nonzero at $t = 0$, so that the time derivative of the filter also includes a delta function at $t = 0$. We will see the appearance of this singular component in the white noise analysis.

To recover the filters corresponding to the dynamical variables, one must invert the diagonalizing operation, equation 2.6, to obtain linear combinations of the eigenmodes:

$$\epsilon_i^{(1)}(t) = U_{i\alpha}^{-1}\tilde{\epsilon}_\alpha^{(1)}(t) = \sum_\alpha c_\alpha e^{\lambda_\alpha t} H(t), \tag{2.8}$$

where the coefficients c_α derive from the components of the Jacobian. Thus, the filters of the linear system are sums of exponentials—either purely real or with an imaginary component. To the extent that subthreshold dynamics are well approximated by the linearized system, this shows clearly why one might expect to find both integrate-and-fire-like neurons, with filters having purely real associated eigenvalues, and oscillate-and-fire neurons (Izhekevich, 2001; Hutcheon & Yarom, 2000) with a corresponding eigenvalue having a nonzero imaginary part. In terms of the linear system, the possible set of filters or feature space of the system is dual with the dynamical variables that define the system. The subset of features that are relevant to a spike occurring are those that contribute to crossing the internal state over "threshold."

The properties outlined above have an interesting implication for white noise analysis in the case that the system is well described by subthreshold linearity. As we will see in the next section, the STA is typically a linear combination over the relevant features and their derivatives. If the components of the STA belong to the true feature space, as they typically will, one can derive the basis for the feature space from the STA and its successive order time derivatives, as has been empirically observed (Rieke, 1991).

2.1 Higher-Order Series Terms. The higher-order approximations to the system can also be expressed in terms of first-order kernels. For example, the second-order approximation is given by the following equation:

$$\mathcal{D}_{km} y_m^{(2)} = \frac{1}{2!}\, H_{kmn} y_m^{(1)} y_n^{(1)}, \quad H_{kmn} = \left.\frac{\partial^2 f_k}{\partial y_m \partial y_n}\right|_{y^{(0)}}, \tag{2.9}$$

whose solution is

$$\begin{aligned} y_k^{(2)} &= \frac{1}{2!}\int ds_1 ds_2\, \epsilon_k^{(2)}(t - s_1, t - s_2) I(s_1) I(s_2), \\ \epsilon_k^{(2)}(t_1, t_2) &= \int dt'\, H_{lmn} \epsilon_{kl}^{(1)}(t') \epsilon_m^{(1)}(t_1 - t') \epsilon_n^{(1)}(t_2 - t'). \end{aligned} \tag{2.10}$$

In this way, all of the higher-order kernels can be obtained by outer products of first-order kernels. The detailed computation is summarized in appendix A.

2.2 White Noise Analysis of Threshold-Crossing Linear Neurons. In this section, we assume that the higher-order nonlinearity of the system is well captured by a threshold. Since voltage is generally the only observed variable, "the threshold" is often taken to be a threshold on the voltage. The imposition of a threshold on the voltage alone has been applied even to neural models such as FN and the Hodgkin-Huxley model system, where there is no difficulty in defining a threshold in V and the "hidden" dynamical variables of W (FN) or m, n, and h (HH). Here we consider the implications of considering the threshold to apply in multiple dimensions. Before we discuss the validity of the threshold approximation for an FN neuron, we illustrate the results from applying white noise analysis to models with a variety of threshold structures acting on subthreshold linear dynamics given by the linearized FN system.

Previous work has treated simple one-dimensional linear and nonlinear models, in which a spike is defined when the output of a single linear filter f on a random current input crosses a threshold. One can show (DeWeese, 1995; M. Meister, private communication, November 2004) that covariance analysis on this model finds two modes: the filter f and its time derivative, f'. The derivative mode is a result of the criterion defining the spike only on the upward crossing of the threshold:

$$\frac{d}{dt}\int^t d\tau f(t-\tau)I(\tau) > 0 \quad \rightarrow \quad \int^t d\tau f'(t-\tau)I(\tau) > 0. \tag{2.11}$$

Thus, the variance of projections onto f' is also reduced with respect to the prior. Covariance analysis has also been performed (Fairhall et al., 2006) on more realistic models in which an exponentially decaying afterhyperpolarization (AHP) is added to the voltage following a spike (Gerstner & Kistler, 2002; Keat et al., 2001). This makes the threshold-crossing model more realistic, but the manipulation does not alter the (f, f') eigenmode structure unless the timescale of the AHP is short enough that spiking occurs many times during the superthreshold fluctuation, which destroys the correlation with positive f'.

Here we assume that the system evolves according to linear subthreshold dynamics. We treat five cases. In the first three, the threshold is linear, but in different components of the 2D space; in the last two, the threshold is a nontrivial two-dimensional structure. The thresholds used are:

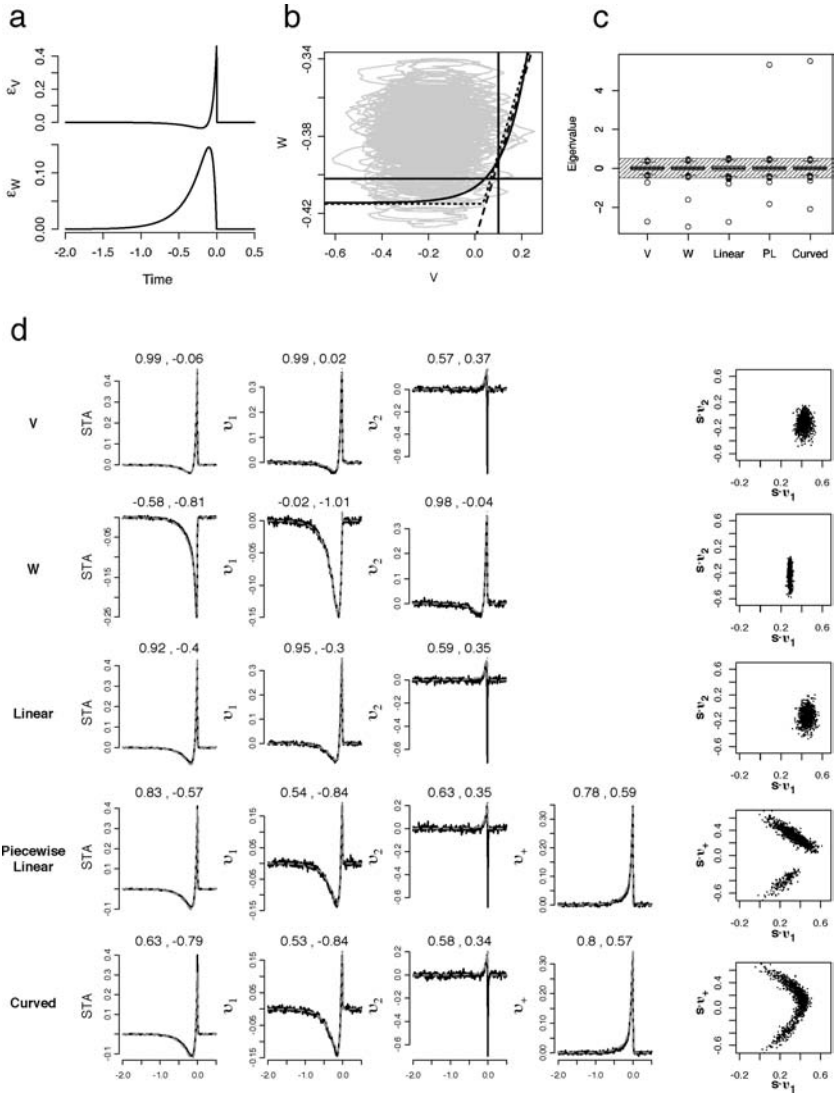
1. A traditional threshold in V only
2. A threshold in W only
3. A threshold on a linear combination of V and W
4. The union of two piecewise linear segments in V and W, that is, a logical "or" on 1 and 2
5. A curved threshold on a smooth, nonlinear function of V and W

From the discussion in Section 2, the linear space in which the n-dimensional system operates is defined by the n filter inversions of the dynamical system, (see equation 2.5). The effect of the threshold is to select as relevant from those n dimensions the subspace defined by the filter directions that have a nonzero projection onto the threshold. Thus, we would predict that for the linear threshold, we will find two relevant modes: the primary direction f, which is normal to the direction of the threshold, and f', its time derivative. For a threshold in a single dimension, V or W only, f should recover the V or W filter, respectively, and the corresponding time derivative. For a threshold that is a linear combination of V and W, f should be a linear combination of the V and W filters. The distribution of trajectories is simply a linear transformation of the original gaussian stimulus distribution. Any linear threshold in this plane will produce negative eigenmodes, as the set of points selected by the threshold must have decreased variance with respect to this prior.

When the threshold is not linear in V and W, we expect to find two primary filters, f_1 and f_2, which will be some linear combination of the V and W filters, and, in principle, the time derivatives. However, the number of modes will be less than or equal to $n + 1$ because, as we have shown, the time derivatives span the same space as the linear variables themselves. (Note that the $+1$ accounts for the singular component of the time derivative, $\delta(t)$, arising from discontinuity of the filters at $t = 0$.) Furthermore, a threshold that is not linear may produce a spike-conditional distribution with a direction in which the variance is actually increased with respect to the prior; such a direction will appear with a positive eigenvalue.

Figure 2b shows the thresholds that we applied to the FN first-order linear dynamics. The results of the covariance analysis are shown in Figures 2c and 2d. Figure 2c shows the covariance eigenvalue spectra for all five cases. The significant eigenvalues are empty circles beyond the error level denoted by the shaded box.[4] In the cases with thresholds on V only, on W only, and linear in both V and W, only two modes are obtained. With 2D thresholds, cases 4 and 5, three eigenmodes appear. Figure 2d shows the corresponding modes. We see that the V threshold picks out $\epsilon_V^{(1)}$, the W threshold selects $\epsilon_W^{(1)}$, and the (V, W)-linear threshold leads to a linear

[4] Significance of eigenvalues can be evaluated as follows. Let N be the number of spikes and d the dimension of each sampled stimulus. To estimate the finite size or dimension error, we choose N d-dimensional stimuli at random (Rust et al. 2005). In this case, the covariance matrix C in equation 1.14 is filled with $d(d+1)/2$ random numbers drawn from a normal distribution $\mathcal{N}(0, \sigma^2\sqrt{2/N})$ in the large N limit. When $d \gg 1$, the eigenvalues of this matrix follow Wigner's semicircle law, and their distribution is bounded by $\pm 2\sqrt{2}\sigma^2\sqrt{d/N}$, which is the error level (Mehta, 1991). When the stimulus has small correlation, the semicircle is not a good approximation of the eigenvalue distribution, but change in the upper and lower bounds rarely exceeds an order of magnitude, which can be checked by numerical simulation.


a
b
c
d
Time
V
W
Eigenvalue
V
W
Linear
PL
Curved
STA
0.99 , -0.06
0.99 , 0.02
0.57 , 0.37
-0.58 , -0.81
-0.02 , -1.01
0.98 , -0.04
0.92 , -0.4
0.95 , -0.3
0.59 , 0.35
0.83 , -0.57
0.54 , -0.84
0.63 , 0.35
0.78 , 0.59
0.63 , -0.79
0.53 , -0.84
0.58 , 0.34
0.8 , 0.57
Piecewise Linear
Curved

combination of these filters. In the latter two cases, the eigenmodes are a linear combination of $\epsilon_V^{(1)}$, $\epsilon_W^{(1)}$, and $\partial_t \epsilon_V^{(1)}$, itself a linear combination of the V and W kernels and $\delta(t)$.

Figure 2d also shows the projection of spike-triggered stimuli onto two of the distinguished eigenvectors. In the first three cases, the linearity of the threshold is manifested in the 2D plane as a single elliptic cluster, while the nonlinear threshold cases show richer structure.

Recall that the filters are generated by linear combinations of a master filter and its time derivatives. This implies that our multidimensional threshold structure has a natural interpretation as a dynamical threshold, depending not only on the voltage V but also on $\partial_t V$, $\partial_t{}^2 V$, and so on (Azouz & Gray, 2000). We will show how well this approximation fits for the Hodgkin-Huxley model in the final section.

2.3 Analysis of Threshold-Crossing Models in Multiple Dimensions. An advantage of the simple models introduced in the previous section is that they can be treated analytically. From equation 1.11, we have

$$G(\mathbf{s}) = \frac{P\left(\text{spike}|\mathbf{s}\right) P(\mathbf{s})}{P(\text{spike})} = \frac{P\left(\text{spike}|\mathbf{s}\right) P(\mathbf{s})}{\int D\mathbf{s}\, P\left(\text{spike}|\mathbf{s}\right) P(\mathbf{s})},$$

where $\int D\mathbf{s} = \int ds_1 ds_2 \cdots$. Instead of directly computing $G(\mathbf{s})$, we characterize it by a moment generating function:

$$W[\mathbf{j}] = \log \int D\mathbf{s} P(\text{spike}|\mathbf{s}) P(\mathbf{s}) e^{\mathbf{j}\cdot\mathbf{s}}. \tag{2.12}$$

We define an n-dimensional first-order system, given by dynamical variables y_k and the corresponding linear filters ϵ_k as in equation 2.2. We assume

---

Figure 2: (a) First-order kernels of the FN neuron model, drawn as filters, that is, in $-t$. (b) Various thresholds for the linear system. V (vertical), W (horizontal), linear (dashed), piecewise linear (dotted), and curved (solid curve) thresholds are used. (c) Spectra of covariance matrices for the thresholds. The shaded box represents the level of error from finite sample size and dimension. (d) STAs and covariance modes of the linear system with various thresholds. The nonsingular part of each mode is fitted using least squares to a linear combination of the (orthogonalized) first-order kernels as $v_{fit} = c_V \epsilon_V^{(1)} + c_W \epsilon_W^{(1)}$. The kernels are normalized, excluding the δ-function component. The coefficients (c_V, c_W) are displayed above each plot, and the gray line is a fitted function. For each case, we show the projection of spike-triggered current histories onto either the leading two negative modes, v_1 and v_2, or v_1 and v_+, depending on the threshold.

that the stimulus is a gaussian white noise current $I(t)$ with zero mean and variance σ^2. Thus, as before,

$$y_k(t) = \int_0^\infty d\tau \epsilon_k(\tau) I(t-\tau).$$

We denote a random segment of the current stimulus $I(\tau)$, $\tau \leq t$ as an infinite-dimensional vector $\mathbf{s}$.[5] Any sample of $\mathbf{y}$ is therefore a functional of the random variable $\mathbf{s}$, $\mathbf{y}[\mathbf{s}]$; for simplicity, we will write $\mathbf{y}$.

Spiking is determined by crossing a threshold $\theta(\mathbf{y}) = 0$ in the phase space from below. Then,

$$P\left(\text{spike}|\mathbf{s}\right) = \delta(\theta(\mathbf{y}))H(\dot{\mathbf{y}} \cdot \nabla_{\mathbf{y}}\theta(\mathbf{y}))\dot{\mathbf{y}} \cdot \nabla_{\mathbf{y}}\theta(\mathbf{y}).$$

The Heaviside function $H(\cdot)$ ensures that spiking occurs only on a threshold crossing from below, and the weight factor $\dot{\mathbf{y}} \cdot \nabla_{\mathbf{y}}\theta(\mathbf{y})$ is a geometric factor accounting for the flux at the threshold. The filters ϵ_k are not necessarily normalized or orthogonal to each other, and it is convenient to define the orthonormal basis $\{f_\mu\}$, which spans the same space as $\{\epsilon_k\}$. Then there will be a linear transformation $T_{\mu k}$,

$$f_\mu = T_{\mu k}\epsilon_k,$$

and so we define a new coordinate system:

$$z_\mu \equiv T_{\mu k} y_k.$$

As $\mathbf{s}$ is gaussian, the orthonormally transformed variable $\mathbf{z}$ is also gaussian with variance σ^2.

Now we can separate the stimulus into two components: its projection into the subspace spanned by the $\{f_\mu\}$ and the orthogonal component. We denote the corresponding directions of $\mathbf{j}$ as $\mathbf{j} = \mathbf{j}_\parallel + \mathbf{j}_\perp$, where

$$j_{\parallel\mu}(t) = \int_0^\infty d\tau\ j(t-\tau) f_\mu(\tau).$$

[5] In practice, any application to data requires discretization in time. We use the convention that a function $f(t)$ is discretized as $\hat{f}_t = f(t)\sqrt{\Delta t}$, where Δt is the time step. This implies $\int f^2 dt \approx \sum_i f(i\Delta t)^2 \Delta t = \sum_i \hat{f}_i^2 = \hat{\mathbf{f}} \cdot \hat{\mathbf{f}}$, and thus the vector $\hat{\mathbf{f}}$ obtained by discretizing $f(t)$ has a vector norm corresponding to the L^2 norm.

The moment generating function can then be separated as $W[\mathbf{j}] = W_{\parallel}[\mathbf{j}_{\parallel}] + W_{\perp}[\mathbf{j}_{\perp}]$, where

$$W_{\perp}[\mathbf{j}_{\perp}] = \log \int D\mathbf{s}_{\perp}\, e^{-s_{\perp}^2/2\sigma^2 + \mathbf{j}_{\perp}\cdot\mathbf{s}_{\perp}} = \frac{\sigma^2}{2} j_{\perp}(t)^2 + \text{const}, \tag{2.13}$$

where $\mathbf{s}_{\perp}$ are the components of $\mathbf{s}$ orthogonal to the plane spanned by $\mathbf{z}$, and

$$\begin{aligned} W_{\parallel}[\mathbf{j}_{\parallel}] &= \log \int d^n\mathbf{z}\, \det(T^{-1}) e^{-z^2/2\sigma^2} \delta(\theta(\mathbf{z})) H(\dot{\mathbf{z}}\cdot\nabla_{\mathbf{z}}\theta(\mathbf{z}))\dot{\mathbf{z}}\cdot\nabla_{\mathbf{z}}\theta(\mathbf{z}) e^{\mathbf{j}_{\parallel}\cdot\mathbf{z}} \\ &= \log \int d^n\mathbf{z}\, e^{-z^2/2\sigma^2} \delta(\theta(\mathbf{z})) w(\mathbf{z}) e^{\mathbf{j}_{\parallel}\cdot\mathbf{z}} + \text{const}, \end{aligned} \tag{2.14}$$

with

$$w(\mathbf{z}) = H((R\mathbf{z})\cdot\nabla_{\mathbf{z}}\theta(\mathbf{z}))(R\mathbf{z})\cdot\nabla_{\mathbf{z}}\theta(\mathbf{z}),$$

where $R = T J T^{-1}$ is the Jacobian matrix of the linear system, equation 2.2, in the $\{f_\mu\}$ basis, and we use

$$\dot{z}_\mu = T_{\mu k}\dot{y}_k = T_{\mu k} J_{kl} y_l = (T J T^{-1})_{\mu\nu} z_\nu.$$

As previously discussed, the closure under time differentiation of the space spanned by the first-order kernels is assured by $\dot{\mathbf{z}}$, as in equation (A.5) in appendix A. Note that our separation of $W[\mathbf{j}]$ depends on this particular property. For a fixed spike time t, restricting ourselves to the region where $\tau < t$, $W_{\perp}[\mathbf{j}]$ becomes a static integral. Thus, equation (2.14) shows clearly how the properties of this model emerge. $P\left(\text{spike}|\mathbf{s}\right)$ is given by a threshold with a weight function $w(\mathbf{z})$ up to some linear transformations. Therefore, for example, the linear threshold case reduces to the one-dimensional case. Since $\nabla_{\mathbf{z}}\theta(\mathbf{z})$ is a constant vector, every computation reduces to a one-dimensional integral in this direction, and the corresponding filter is a linear combination of f_μs, as we have observed in Figure 2d. Furthermore, a variety of structures may be generated depending on how trajectories cross the threshold.

From equations 2.14 and 2.13, we can derive the analytic forms of the first- and second-order moments:

$$\begin{aligned} \text{STA}(\tau) &= \sum_\mu c_\mu f_\mu(\tau), \qquad c_\mu = \frac{1}{N}\int d^n\mathbf{z}\, z_\mu e^{-z^2/2\sigma^2} w(\mathbf{z}), \\ N &= \int d^n\mathbf{z}\, e^{-z^2/2\sigma^2} w(\mathbf{z}), \end{aligned} \tag{2.15}$$

and

$$C(\tau,\tau') = \sum_{\mu,\nu} \left(c_{\mu\nu} - c_\mu c_\nu - \sigma^2 \delta_{\mu\nu}\right) f_\mu(\tau) f_\nu(\tau'),$$

$$c_{\mu\nu} = \frac{1}{N} \int d^n\mathbf{z}\, z_\mu z_\nu e^{-z^2/2\sigma^2} w(\mathbf{z}). \tag{2.16}$$

Note that we have assumed a prior based on the distribution of randomly driven trajectories of the linear system. This takes no account of perturbations of this distribution due to flux from the system's return from a spike. This assumption is equivalent to assuming that the last spike is in the distant past, so that memory of that perturbation has vanished. In this article, we consider only this "isolated spike" case. We return to this point in the discussion.

2.4 Variance Dependence. Through equations 2.15 and 2.16, this model captures an explicit dependence of the white noise outcome on the stimulus statistics, in particular, the variance σ^2. While the subthreshold dynamics are linear, the dependence on σ is nonlinear due to the threshold shape and the weight function $w(z)$. We discuss two examples below.

We first consider a linear threshold. Through a suitable linear transformation, this case simply reduces to a filter-and-fire model with a single filter, say, f_0, and a fixed threshold $z_0 = \theta_f$. Now the distribution of threshold crossing points is constrained by $z_0 = \theta_f$ and $\dot{z}_0 > 0$. As we mentioned in the previous section, $\dot{z}_0$ lies in the originally defined feature space and is given by another single filter, which we denote by f_1. In other words, when the normalized $\dot{f}_0$ is denoted by f_0', we can choose $f_1 = f_0'$ by a suitable orthogonal linear transformation. Thus, our system depends on two filters, which are depicted schematically in Figure 3a.

The STA is the centroid of the distribution of threshold crossing points. Equation 2.15 reproduces a previously known result (M. Meister, private communication November 2004; Fairhall et al., 2006),

$$\mathrm{STA}(\tau) = \theta_f f_0(\tau) + \frac{\sigma}{\sqrt{\pi/2}} f_1(\tau). \tag{2.17}$$

With high-variance $\sigma \gg \theta_f$, the STA is dominated by f_1. In the feature space, with increasing variance, the threshold stays the same, but a larger portion of it is crossed by trajectories driven by the larger variance ensemble, as can be seen in Figure 3a.

When the threshold is curved, the σ dependence is considerably more complicated. We will consider an extreme but analytically tractable version

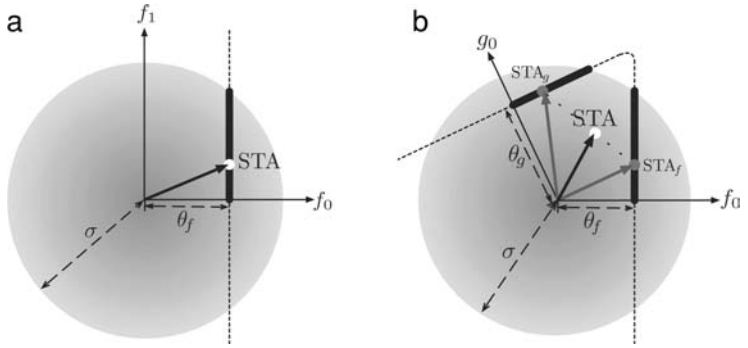


Figure 3: (a) Diagram of a single filter model. The disc of radius σ represents the prior gaussian distribution. The dotted line is the model threshold imposed at $z_0 = \theta_f$. The thick line is a distribution of spike-triggered stimuli, and the white dot represents its average, the STA. (b) Diagram of a model whose threshold is the union of two lines, each imposed at a distance $\theta_{f,g}$ from the center. Two gray dots denote the STAs for each segment.

of this case to illustrate the point: let the curved threshold be approximated by two linear ones imposed at θ_f and θ_g in the f_0 and g_0 directions, respectively, as in Figure 3b. In this case, one segment of the threshold imposes a dependence on the filters f_0 and its (normalized) derivative, f_1, while the other selects g_0 and $g_1 = \dot{g}_0/\|\dot{g}_0\|$. The space of relevant features is still two-dimensional or three-dimensional, including the δ-function, since g_0 and g_1 are linear combinations of f_0, f_1, and possibly also $\delta(t)$.

Now the STA of this system is

$$\begin{aligned} \text{STA}(\tau) &= \cos^2\varphi \cdot \text{STA}_f(\tau) + \sin^2\varphi \cdot \text{STA}_g(\tau), \\ \varphi &= \tan^{-1} e^{(\theta_f^2-\theta_g^2)/4\sigma^2}, \end{aligned} \tag{2.18}$$

where $\text{STA}_\epsilon(\tau) = \theta_\epsilon \epsilon_0(\tau) + \frac{\sigma}{\sqrt{\pi/2}}\epsilon_1(\tau)$ as in equation 2.17. Note that as in Figure 3b, the STA does not lie on the threshold. As for a variety of experimental examples such as complex cells (Touryan et al., 2002), neurons of rat barrel cortex (Petersen, 2003), and some retinal ganglion cells (Fairhall et al., 2006), the spike-triggered stimuli are poorly represented by their first-order statistics, the STA. Also, the coefficients of $\text{STA}_{f,g}$ depend exponentially on $\theta_{f,g}$ and σ. Thus, the system shows a nonlinear dependence on the stimulus variance.

Figure 4 shows an example. The W threshold model does not show any significant change, except sharpening of the STA due to the increased component of $\dot{\epsilon}_W$, Figure 4b, and linear broadening of the spike-triggered stimulus distribution. The piecewise linear threshold case is more dramatic: while

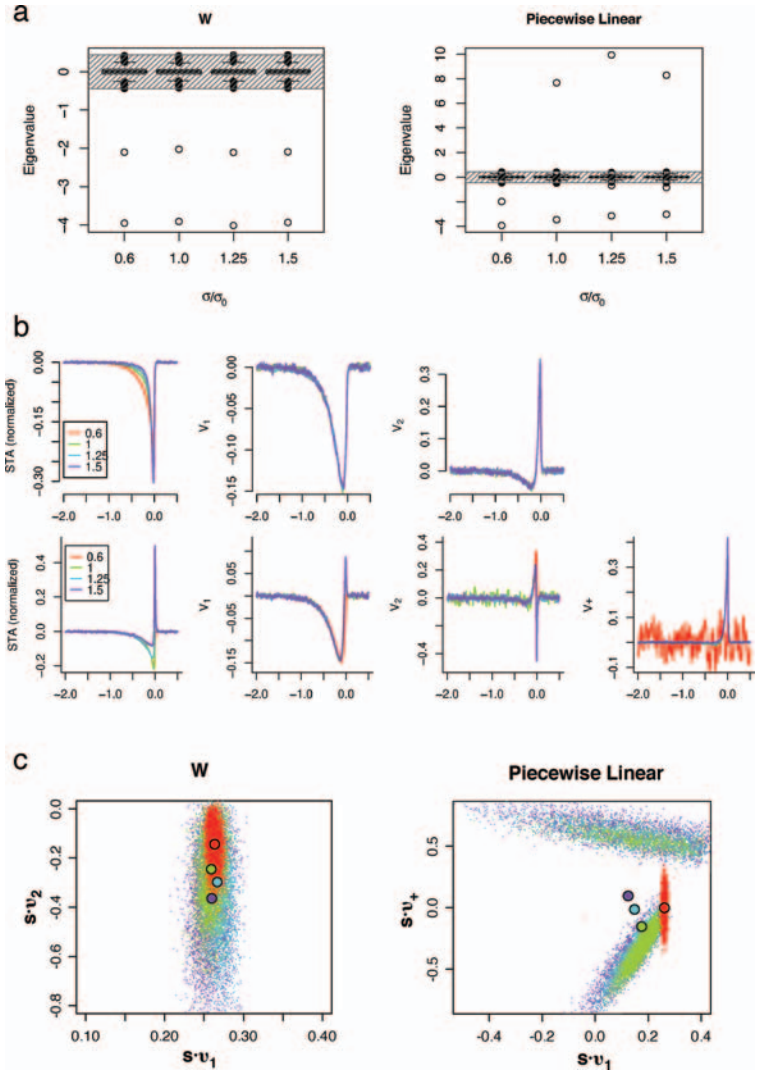


Figure 4: Two examples of variance dependence in the linear system model. (a) Spectra of the covariance matrix for several different variance stimuli. (b) Significant eigenmodes. (c) Projection of spike-triggering stimuli onto the filter subspace defined by $v_{1,2}$ for the linear and $v_{1,+}$ for the piecewise linear threshold, respectively, colored according to the stimulus variance as in the legend of b. The circled points are the centers of the distributions, corresponding to the respective STAs, colored according to the respective distribution.

the smallest variance does not drive the system hard enough to produce a positive eigenmode (hence, the one-dimensional distribution of projections seen in Figures 4b and 4c, red), a new significant mode emerges as the variance increases. The STA changes beyond sharpening and almost looks like a different model at high variance compared with low variance. Each of the significant modes changes as more trajectories cross the threshold from the other side and the principal axis of the distribution rotates, as seen in Figure 4c.

3 Dynamical Threshold

In the previous section, we considered neuron models composed of linear filters derived from the FN neuron, and some choices of imposed thresholds. However, when we consider the full FN neuron, the threshold arises from the structure of the FN equations, 1.5 and 1.6. Here we discuss the importance of the threshold identification for reverse correlation analysis.

The problem of the identification of a threshold arises immediately on attempting a reverse correlation analysis; spike times are often defined by the threshold crossing of the voltage. Here also we can impose an arbitrary threshold in V to identify each spike. The STA obtained using this scheme is displayed in Figure 5a. It is clear that it cannot be well described by the first-order kernels. However, this does not mean breakdown of the analysis; rather, it underscores the point that the STA or any other single spike quantities should be computed using the "correct" threshold that we take to be the dynamical threshold discussed in section 1.

Figure 5b compares the spike-triggered stimuli in the fixed V and dynamical threshold cases. While the peaks of the stimuli are spread out in time in the fixed V threshold case, for the dynamical threshold, the peaks lie in a narrow band around the spike time. As we can see from Figure 1, the spiking trajectories cross the dynamical threshold before crossing the V threshold, and this timing difference, Δt, depends on W. Additionally, since the system is driven by gaussian white noise, the variability increases with the timing difference, inducing a point-spread function on the STA. Hence, each spike-triggered stimulus is contaminated by this temporal jitter, and estimated filters are distorted. As mentioned in section 1.1, the FN neuron does not have a clear-cut dynamical threshold; the threshold was chosen to some degree arbitrarily, and this will induce some error in the estimation of filters. However, Figure 6a shows the improvement in the STA computed using the dynamical threshold. Figures 5c and 5d show the point-spread distribution—the distribution of values of Δt—and how it is correlated with W. This situation is similar to that discussed in Aldworth, Miller, Gedeon, Cummins, & Dimitrov, (2005), where it was noted that temporal or spatial jitter can blur the estimation of filters and receptor fields. There, a blind deconvolution algorithm was used to "dejitter" and realign

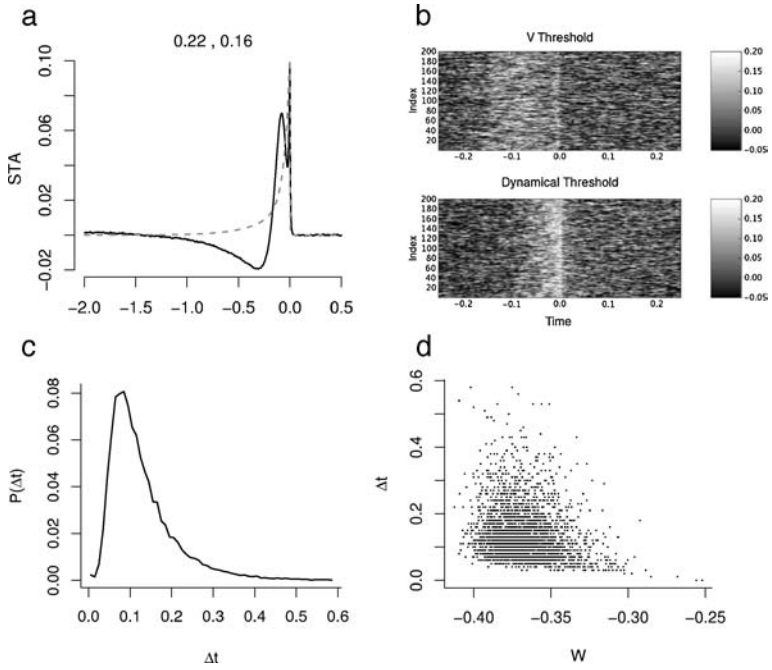


Figure 5: Effect of a dynamical threshold in the FN model. (a) STA of the full FN system with a constant V threshold. As before, a linear fit by the kernels is also shown. (b) Comparison of spike-triggered stimuli with different threshold choices in the FN model. (c) Temporal point-spread function induced by the choice of threshold. Δt is the time delay from crossing the dynamical threshold to the fixed V threshold. (d) Time delay of each spike-triggered stimulus plotted against the value of W at its threshold crossing point.

the spike-triggered stimuli, which dramatically sharpened the estimate of the spike-triggering stimulus. In a real system, jitter may indeed be due at least partly to noise, but it is also possible that a component of such jitter is deterministic and due to variability in the point of dynamical threshold crossing, as in the FN case. Blind deconvolution may then be viewed as an empirical approach to recovering a dynamical threshold based on spike times originally recorded using a voltage threshold. It is interesting to note that estimated jitter in such a case is correlated with projection onto the STA derivative.[6]

[6] This correlation arises because the STA derivative is the linear approximation to time translation of the STA. Hence, if a current history aligns with the STA better under a small time translation or jitter, it will have a projection onto the STA derivative proportional

4 Fully Nonlinear Systems

In this section, we discuss the covariance analysis of the full FN system and compare the result with the same analysis of the first- and second-order approximation.

We identified approximately 2×10^6 spikes first using a voltage threshold and then backtraced each trajectory to the point where it crosses the dynamical threshold shown in Figure 1. The trajectory might cross the threshold multiple times before spiking due to the noisy input. In this case, we used the first crossing point after the trajectory diverges from that of the second-order approximation. This is based on the assumption that the second order is a sufficiently good approximation of the system in the subthreshold regime. For both the first- and second-order system, crossing of the dynamical threshold is used to identify spikes. Right after a spike, we imposed a postspike inhibitory period equal to about a "spike width," as empirically determined from the full system.

The results are shown in Figures 6 and 7. Figure 6a shows that the STA of the first-order approximation is quite similar to that of the full system, and the second-order STA is even closer. From the covariance analysis, we see that the FN neuron, like the HH model (Agüera y Arcas et al., 2003), has one dominant negative and one dominant positive eigenvalue. However, the results from the covariance analysis of the three systems differ considerably. Both the second-order and the full system show a relatively large number of significant eigenvalues. This is due to the contribution of the second order $g_2[V]$ and $g_2[W]$ (and higher-order) kernels in equation 2.1. However, these are relatively suppressed compared to the first-order modes. Further, the spectrum derived from the second-order system, Figure 6b, has no positive significant eigenvalue.

Figure 6c provides more detailed information. The first-order case is just as we have seen previously, with three modes that are well described by linear combinations of the linear kernels. In the second-order case, v_1 and v_2 are comparable to those of the first-order. v_3, which is not approximated by the first-order kernels, must arise from the second-order kernels. In this regard, the full system is well matched with the second-order approximation. There also, v_3 is comparable to that in the second order, and $v_{1,2}$ are derived from the linear kernels. However, the full system also has a positive mode, v_+, from the linear kernels, which resembles the first order rather than the second.

to that jitter. In section 2, we demonstrated that the STA and its derivatives are, up to a linear transformation, the filters associated with the dynamical variables of the linearized system. Therefore, there is a relationship between blind deconvolution to optimize fit to an STA and estimation of a dynamical threshold based on the "hidden" (non-V) dynamical variables.

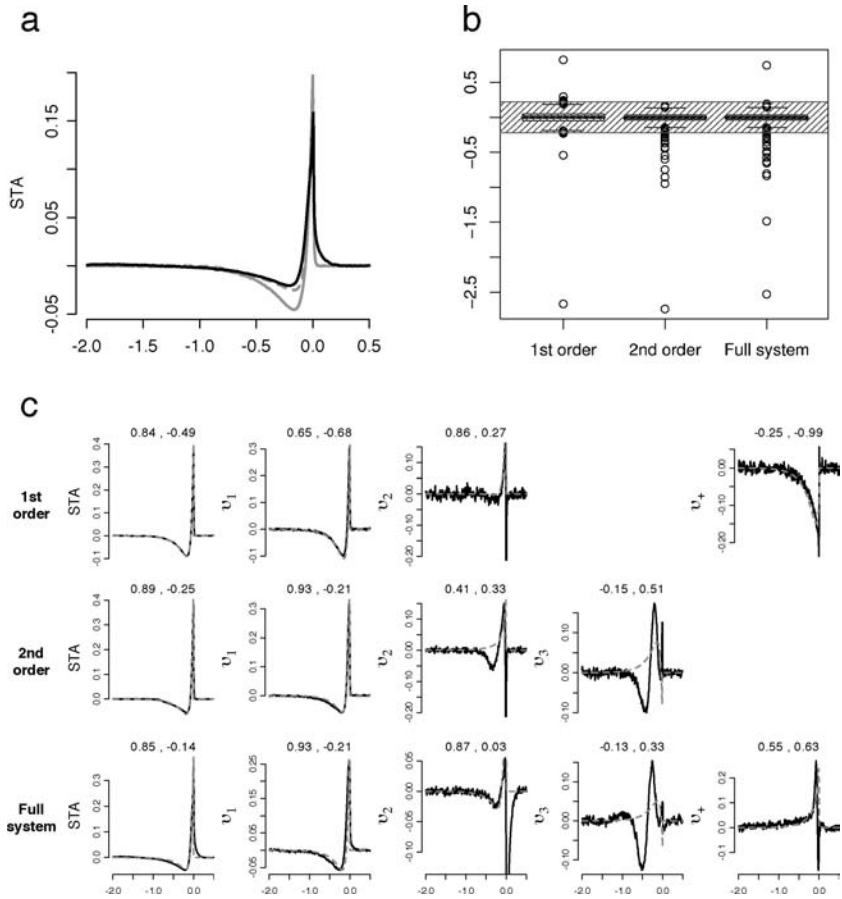


Figure 6: (a) STAs of an FN model neuron, and the first (gray) and second (dashed) order approximation. (b) Spectra of covariance matrices of the FN model and approximations. (c) STAs and covariance modes for the first- and second-order approximations and the full FN system.

The geometry of the spike-triggering stimulus projections is shown in Figure 7a. The first order, not surprisingly, recapitulates the curved threshold case in Figure 2d. However, the second order is more like the (V, W)-linear threshold case, while again the full system resembles the curve in Figure 2d. A possible explanation is that each system probes different parts of the dynamical threshold. Figure 7b marks the density of threshold crossing points for each model. In contrast to the first-order and the full system, which access a large section of the threshold with nontrivial curvature, the second order probes only a small and almost linear section. This may be the reason for the lack of a positive eigenvalue. As in the toy models

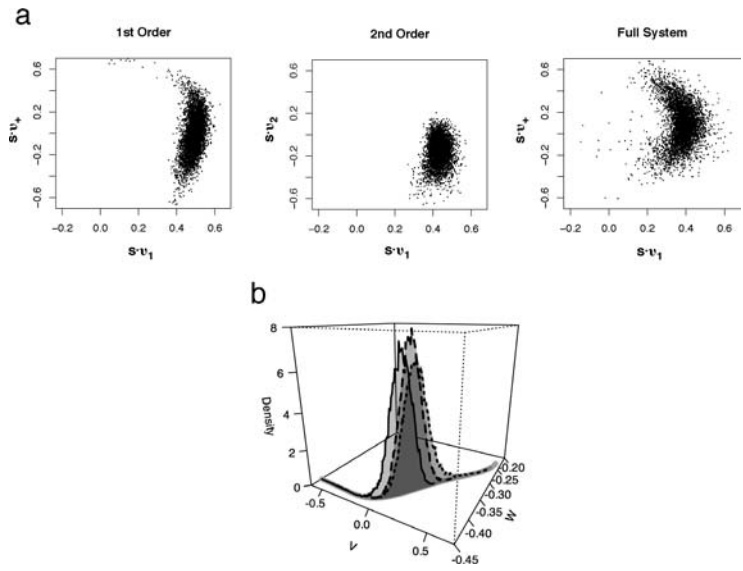


Figure 7: (a) Spike-triggered stimuli projected into a feature space defined by $v_{1,+}$ for the first-order case and full system and $v_{1,2}$ for the second-order system. (b) The density of threshold crossing points in the first-order (solid), second-order (dashed), and full (dotted) systems, plotted along the threshold curves in the $V - W$ plane.

in section 2, the contributions of the relevant dimensions are determined not only by local information (filters) but also by the global structure of a multidimensional or dynamical threshold.

5 Abbott-Kepler Model

In this section, we apply the same analysis to a two-dimensional model which is more nonlinear and more realistic than the FN model.

Abbott and Kepler (1990) developed a two-dimensional reduction of the Hodgkin-Huxley model, based on the observation that there is a separation of timescales between the faster m and the slower n and h variables. m is then replaced with its asymptotic value at the membrane voltage V, while n and h are controlled by another voltage variable U. The equations of the Abbott-Kepler (AK) model are of the form

$$C\frac{dV}{dt} = f(V, U) + I(t), \tag{5.1}$$

$$\frac{dU}{dt} = g(V, U), \tag{5.2}$$

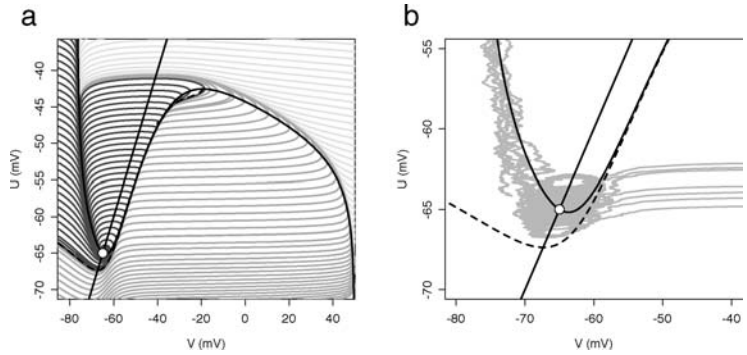


Figure 8: (a) Phase plane of the Abbott-Kepler model with trajectories, nullclines, and a threshold, in the absence of input. (b) AK phase plane with the injected noisy input.

where $f(V, U)$ and $g(V, U)$ are nonlinear functions in V and U. Their derivation is briefly sketched in appendix C, and we refer to the original paper (Abbott & Kepler, 1990) for further detail. The nullcline for U is given by $g(V, U) = 0$, which is satisfied by $V = U$. The V nullcline, $f(V, U) = 0$, is more complicated and obtained numerically. The two nullclines intersect at a fixed point $V = U = -65\,\mathrm{mV}$.

Figure 8a shows the phase plane of this model with zero input current. Like Figure 1, the threshold structure, which can be obtained numerically, is visible. Due to the strong nonlinearity, spiking trajectories are well defined on the phase plane, and the threshold has less ambiguity than for the FN model. Again, we try to identify the dynamics of the system in the subthreshold regime with the first-order approximation. Unlike an FN neuron, the Jacobian of this system has complex eigenvalues $\lambda_{\pm} = -0.2118 \pm i0.4035\,\mathrm{ms}^{-1}$, and therefore the first-order kernels $\epsilon^{(1)}_{V,U}$ oscillate, Figure 10a. This is consistent with the oscillatory linearized behavior associated with the full Hodgkin-Huxley model near equilibrium.

Before we carry out covariance analysis on this model, we examine the effect of a dynamical threshold, as in section 3. Figure 9a shows the spike-triggering stimuli aligned with both a fixed threshold in V, chosen as $V = -40\,\mathrm{mV}$ to unambiguously select spiking trajectories, and the dynamical threshold. The typical time shift is of order of 1 ms or less; the overall STA suffers from only a slight time or phase shift when the V threshold is used. However, in the small timescale of 1 ms or less, the V-threshold STA differs from that in the dynamical threshold case, which is characterized by a large delta-function component at the spike time. Both STAs are well approximated by linear combinations of $\epsilon^{(1)}_{V,U}$, up to a small deformation around -15 ms due to the effects of multiple spikes.

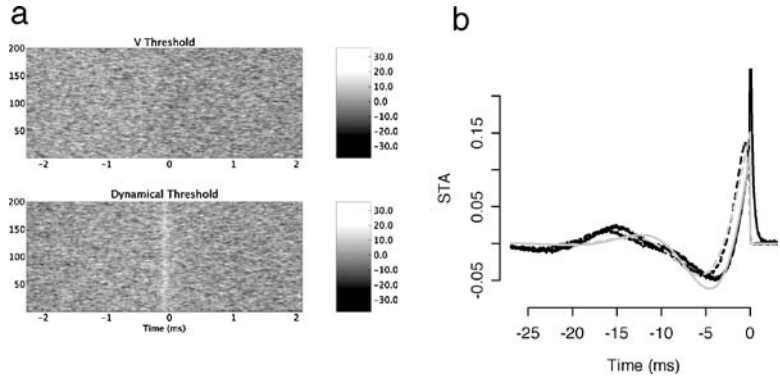


Figure 9: (a) Spike-triggering current histories with two different threshold schemes in the AK model. (b) STAs of the AK model with the fixed $V = -40\,\mathrm{mV}$ threshold (dashed) and dynamical threshold (solid). Also, they are least squares fitted by linear combinations of $\epsilon_{V,U}$ (gray).

Figure 10 shows the results from covariance analysis carried out with approximately 2×10^6 spikes for gaussian white noise stimuli with various variances. For comparison with previous results, we selected three eigenmodes corresponding to the leading two negative and the largest positive eigenvalues. We find that they are reasonably well approximated by linear combinations of $\epsilon^{(1)}_{V,U}$ in most cases, although they are sometimes affected by a δ-function at $t = 0$ and a large multispike effect at high variances. The multispike effect can be eliminated by considering only the isolated spikes when the spike rate is low (Agüera y Arcas & Fairhall, 2003). At higher variances, isolated spikes are rare, and there is a stronger influence of oscillating "silence modes" (Agüera y Arcas et al., 2003). As in the FN neuron case, we identify modes other than those in Figure 10c as "nonlinear modes."

We compare results at different variances with our discussion in section 2.4. Some features of variance dependence are shared with the toy model in section 2.4: the eigenvalue spectrum drifts, and the corresponding modes rotate among themselves. However, the Abbott-Kepler modes also exhibit more complicated behavior. Figure 11 shows projections of spike-triggered stimuli onto two distinguished eigenvectors $v_{1,+}$ and corresponding threshold crossing points. At low variance, the system crosses mostly one side of the threshold while the projections trace out the curvature of the threshold segment. As the variance increases, some crossing points begin to appear on the left side of the threshold. However, this does not overcome the expansion of a crossing point distribution on the right side, and the modes corresponding to this direction dominate. At high variance, there are many crossing points on both sides, reflected in the bimodal distribution of the

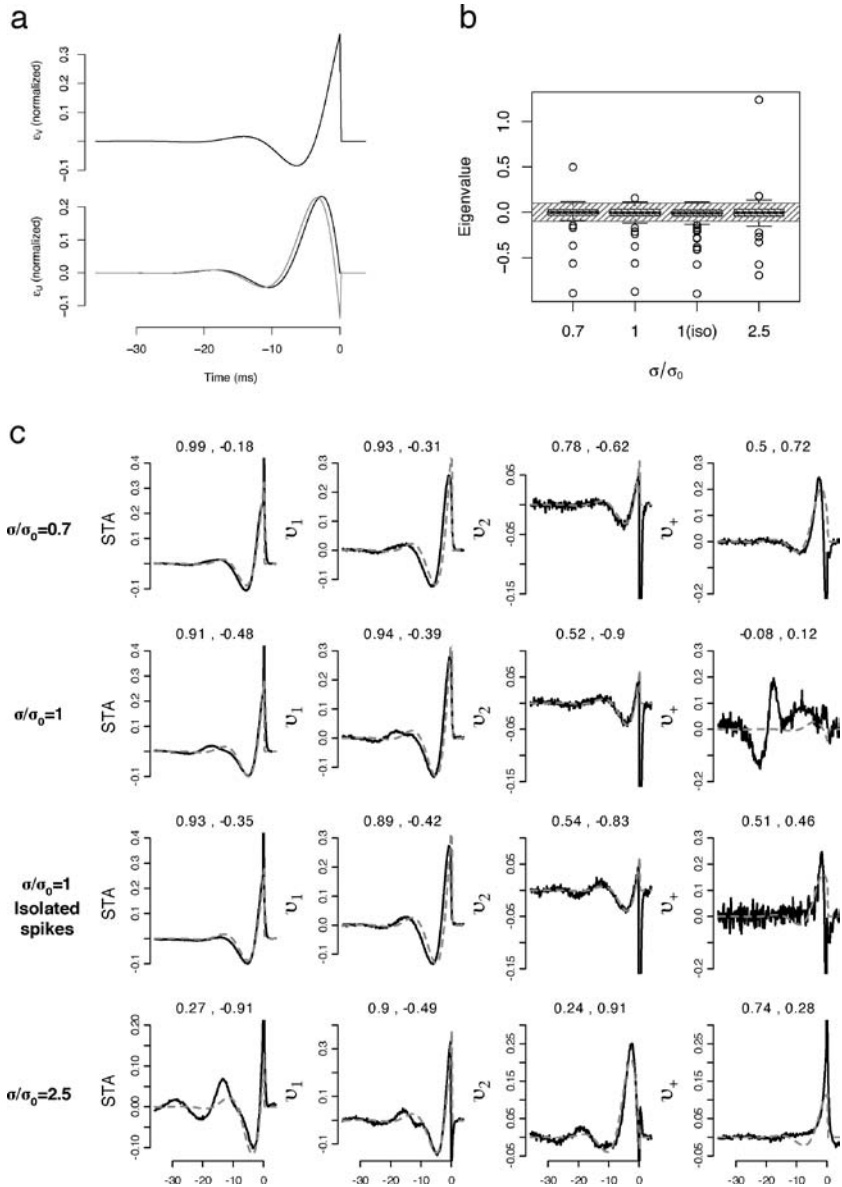


Figure 10: (a) Normalized first-order kernels $\bar{\epsilon}_{V,U}$ of the AK model and the component of ϵ_U orthogonal to ϵ_V, normalized (gray). (b) Eigenvalue spectrum of the covariance matrix of the AK model with each variance. $\sigma_0 = 13$pA. (c) Eigenmodes of the covariance matrices. As previously, $v_{1,2}$ and v_+ correspond, respectively, to the two smallest negative eigenvalues and the largest positive one.

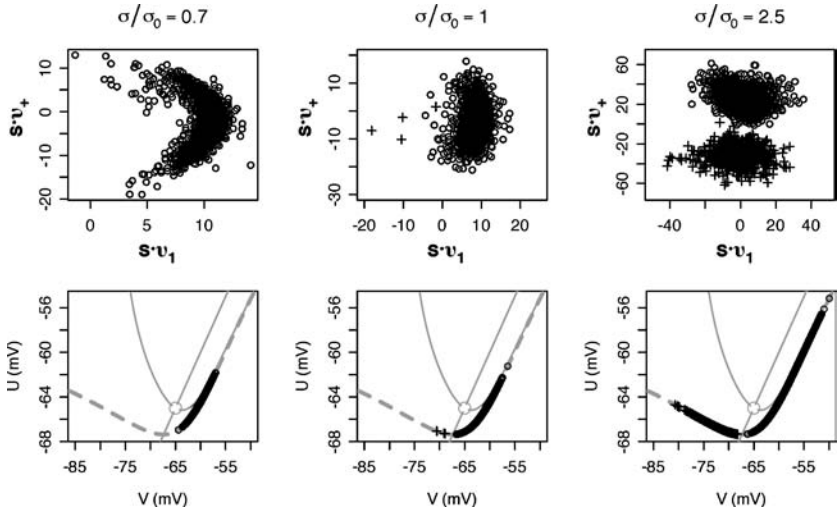


Figure 11: Projection of spike-triggering stimuli onto v_1 and v_+ in the AK model. Below are threshold crossing points in the phase plane of each case, marked by crosses, $V \leq -67\,\mathrm{mV}$, and circles, when $V \geq -67\,\mathrm{mV}$, over the dynamical threshold (dashed). The other gray curves are the nullclines, with their intersection, the fixed point, marked by a circle.

projections, previously seen in the toy models. Thus, this is another example of how the results of reverse correlation analysis are affected by both the filter properties and the interaction of the stimulus ensemble with the threshold geometry.

So far, we have discussed only two-dimensional dynamical models. We began with some artificial toy models with purely linear subthreshold dynamics and proceeded to the minimal FitzHugh-Nagumo spiking model. We applied the lessons learned there to the more nonlinear Abbott-Kepler model. We will conclude with a partial analysis of the higher-dimensional Hodgkin-Huxley model.

6 Hodgkin-Huxley Model

Higher-dimensional systems require nontrivial extensions of the methodology we have used with two-dimensional systems. First, it is much harder to find a dynamical threshold, which could now be a multidimensional hypersurface rather than a curve. If we do not align the spike-triggered stimuli according to the dynamical threshold, the obtained filters may include a distribution of time delays. For example, the reverse correlation analysis may result in broadened filters,

$$\tilde{f}_\mu(t) = \int d\tau\; f_\mu(\tau) p_\sigma(t-\tau), \tag{6.1}$$

where $p_\sigma(\tau)$ is a point-spread function, depending on the input variance σ^2, as in Figure 5c.[7]

However, we can still gain some insights from $\tilde{f}_\mu$. If $p(t)$ has a narrow support, say less than a millisecond, it has negligible effects for larger time windows. It is also possible that some consequences of our analysis will still apply to $\tilde{f}_\mu$ with only a few reasonable assumptions. For example, if p_σ is of limited temporal extent, then the derivatives of $\tilde{f}_\mu$ can be simply

$$\tilde{f}_\mu^{(n)} \approx \int d\tau \; f_\mu^{(n)}(\tau) p_\sigma(t-\tau). \tag{6.2}$$

Therefore, as we discussed in section 2, the linear modes among $\{\tilde{f}_\mu\}$ should still be (approximately) closed under time differentiation.

To demonstrate this, we use the four-dimensional Hodgkin-Huxley model, and the following strategy: we select the linear modes out of the significant filters from covariance analysis. Now, since the STA is approximately their linear combination, the time derivatives of the STA should be written in terms of the linear combinations as well if they are closed under time differentiation. Figure 12 shows that this holds. In the low-variance case, Figure 12a, the three linear modes chosen provide good fits to the time derivatives of the STA. The high-variance case, Figure 12b, is affected by multispike effects and a filtering artifact, but it is also well fitted. Therefore, we can conclude that the time derivatives of the STA span the same space as three linear modes. One might expect a fourth linear mode due to the dimensionality of the model, but this is not as significant as the others; this agrees with previous covariance analysis that the HH model can be well described as a quasi-three-dimensional system (Agüera y Arcas et al., 2003).

We note that it is not clear whether the linear modes in the two cases span the same feature space. This is difficult to ascertain since the point-spread function can in principle depend on the stimulus variance. In Figure 12c, we see that while v_1 of the high-variance case can be fitted by the linear modes of the low variance, the other modes show a small deviation even around 5 ms. This might indicate an interesting variance dependence as in section 2.4, but we will not pursue this issue in this article.

7 Summary

Here we have investigated the meaning or interpretation of the features derived from the spike-triggered covariance method applied to two simple

[7] In fact, it can be much more complicated than this. Since the time delay depends on the location of threshold crossing in a phase space, additional temporal correlations are introduced. Therefore, each filter may have a different point-spread function, or filters may depend on the time delay structure in a complicated way. Here will discuss only the simplest case.

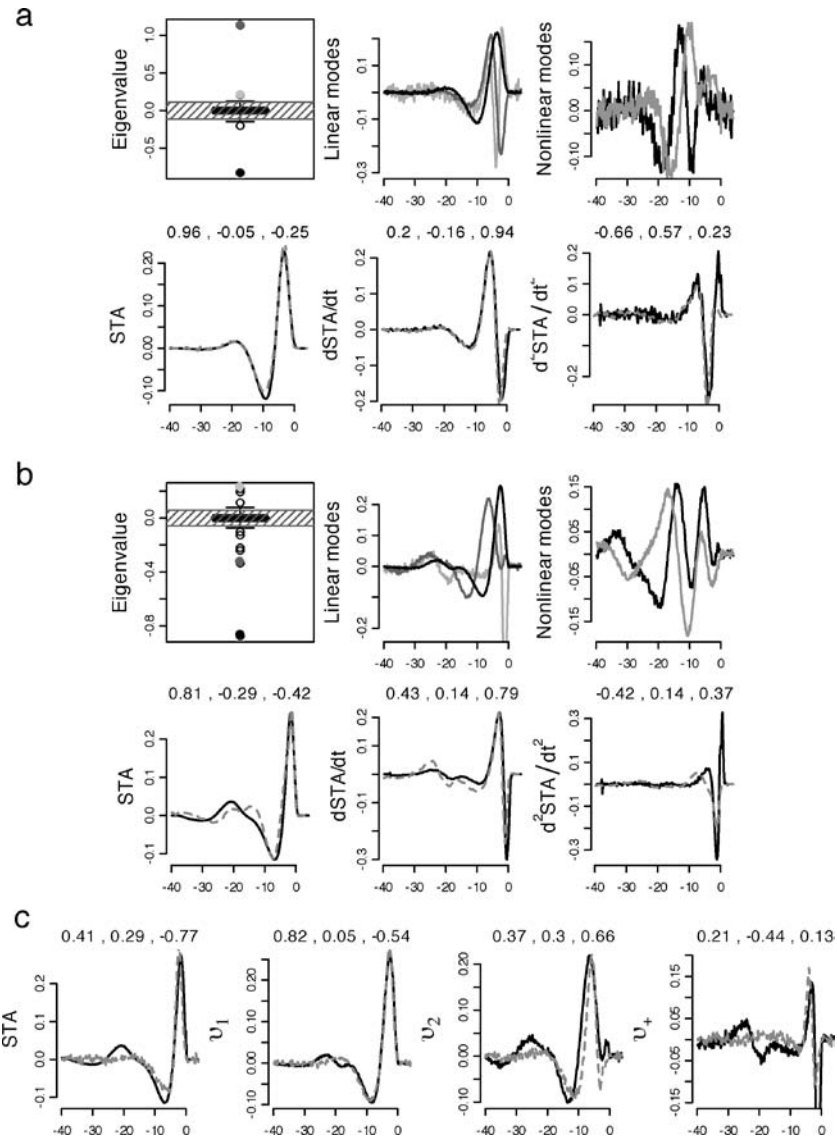


Figure 12: Covariance bases of the linear feature space of the Hodgkin-Huxley model recovered at (a) low and (b) high variance. The three most significant eigenvalues are denoted with the same shades as the corresponding linear eigenmodes. (c) The STA and linear modes of the high-variance case and their fits using the linear modes of the low-variance case in the HH model.

neuron models: the FitzHugh-Nagumo model, the minimal spiking neuron model, and the Abbott-Kepler model, a more faithful two-dimensional reduction of the Hodgkin-Huxley system. The power of white noise analysis is that it provides a data-driven method to reduce a high-dimensional dynamical system to a functional model that captures the essential computation of the system in terms familiar to systems neuroscience: a receptive field that filters the stimulus and a threshold function over the filtered stimulus. Our goal here was to analyze the output of a white noise analysis in terms of what it can reveal and how it depends on the underlying dynamical system. In this, our approach is distinct from the elegant work of Huys, Ahrens, & Paninski (2006), where responses to white noise stimuli are used to fit the parameters of a conductance-based model.

Dealing with simple two-dimensional systems, our observations of spiking dynamics in the phase plane motivated the following reduced model: dynamics in the phase plane are approximated by the perturbative expansion, in particular the linear approximation, and the system's nonlinearity is captured by a spiking threshold, determined for zero input, that extends through multiple dimensions. This model is a generalization of a basic filter-and-fire model to multiple dimensions and extends it in two significant ways. One is the identification of the filters with the dynamical variables of the system; the other is the generalization of the concept of the threshold.

The simplified model with linear dynamics and a multidimensional curved threshold was treated both analytically and phenomenologically, using numerical simulation and reverse correlation (covariance) analysis of spike-triggered stimuli. This led to several insights, some of which are related to existing observations. First, for this case, the feature space derived from covariance analysis is spanned by the linear kernels. Since the set of kernels is closed under time differentiation, the same feature space can also be spanned by a generic linear combination of the kernels, such as the spike-triggered average and its n time derivatives. However, not every kernel contributes as a relevant feature. The threshold structure plays a role of selecting the relevant ones from the kernels: for example, a linear threshold selects a single filter and its time derivative. In general, a threshold spanning a d-dimensional subspace will select $d + 1$ features from among linear combinations of the kernels. We show further for this model that the distribution of spike-triggered stimuli in the covariance feature space corresponds to that of threshold crossing points, up to a suitable linear transformation.

We also used this model to illustrate the effects of the interaction of the threshold nonlinearity with the stimulus ensemble. We show that a complex threshold geometry leads to a nontrivial variance dependence of the eigenmodes and eigenvalues of covariance analysis and, even more so, the spike-triggered average. In the linearized subthreshold model, the subspace of eigenmodes is not changed, but the spike-triggering ensemble

may rotate through this subspace, leading to a variance-dependent spike-triggered average. This is one example of stimulus-variance dependence in a nonlinear, nonadapting system (Yu & Lee, 2003; Borst, Flanagin, & Sompolinsky, 2005). We show this effect in the analysis of the Abbott-Kepler model neuron.

The identification of a curved or dynamical threshold raises an issue in reverse correlation analysis regarding the determination of spike time. For the FitzHugh-Nagumo and Abbott-Kepler models, we compared two cases: when spike times are identified using a threshold in voltage and when the spike times are found using the crossing of an identified curve that leads to spiking for zero current input. Surprisingly, the spike-triggered averages showed remarkable differences. For a voltage threshold, the time delay from the earlier threshold crossing point blurs the estimated filters. It is shown that using the dynamical threshold leads to a better fitting of the estimated filters by the system's linear kernels, as well as improving the sharpness and substantially altering the shape of the STA.

In all of our discussion here, we have concentrated on the subthreshold dynamics bringing the system to the point of threshold. We do not analyze, and our simplified models do not accommodate, the dynamics of the system immediately after a spike. In the phase-plane picture, spiking reinjects the system into the subthreshold regime in a nonrandom way, affecting the subsequent probability flux to threshold, analogous to the one-dimensional reset of the integrate-and-fire neuron treated in Paninski, Lau, & Reyes, (2003). In our approach, we have assumed that the system has remained below threshold for long enough that its location in the subthreshold space has been randomized by the driving current. This corresponds to an analysis of isolated spikes only, a simplification we have used before (Agüera y Arcas et al., 2003) and employed here for the covariance analysis. A number of works have treated this issue explicitly with a variety of methods: treating the interspike interval as the primary symbol (de Ruyter van Steveninck & Bialek, 1988; Rieke et al., 1997), solving for the interspike interaction using the interspike intervals (Pillow & Simoncelli, 2003), simultaneously solving for the linear kernel over stimulus history and spike history using the autoregressive moving average (Powers et al., 2005; Truccolo, Eden, Fellows, Donoghue, & Brown, 2005), and fitting parameters of an explicit model for the effective postspike current (Kistler et al., 1997; Keat et al., 2001; Paninski et al., 2004; Pillow et al. 2005). While the simplification we have used here allows us to find direct connections between the covariance modes without the confound of the interspike interaction (Agüera y Arcas et al., 2003; Agüera y Arcas, & Fairhall, 2003) it is clearly not a complete model for spiking responses. A first step toward a more complete spiking model may be to consider the perturbed subthreshold distributions induced by the influx of trajectories following spikes. This is in effect a mean field approximation, taking into account the overall spike rate for a given stimulus ensemble. Further steps could be taken by introducing a return map

deterministically relating the point of threshold crossing to a point of return into the subthreshold domain. The integrate-and-fire model is the most trivial implementation of such a return map; an equivalent map in multiple dimensions would reintroduce all spike trajectories into the subthreshold domain at an identified point (this is $V = 0$ for integrate-and-fire). Such a many-to-one map implies that the neuron's state is completely reset by a spike, which is incorrect for neurons with slow conductances that modulate spike afterpotentials. A less degenerate map seems more appropriate for such cases. Another important step is the addition of these slow conductances. Using our formalism, such conductances may be representable simply as additional dimensions of threshold curvature with corresponding longer-timescale stimulus filters.

White noise analysis allows the derivation of intuitive functional models for neural computation: What does the neuron compute? In this article, we have drawn concrete correspondences between the components of these functional models and parameters of the underlying dynamical system.

Appendix A: Volterra Expansion of a Dynamical System

Let us consider an n-dimensional dynamical system, perturbed by an external input $I(t)$ as equation 2.1. For convenience, we introduce a dimensionless expansion parameter η, with which we will expand the equation. More precisely, equation 2.1 becomes

$$\partial_t y_k = f_k(y_1, y_2, \ldots, y_N) + \eta \delta_{k0} I(t). \tag{A.1}$$

We also have the perturbation expansion of y_k as

$$y_k = y_k^{(0)} + \eta y_k^{(1)} + \eta^2 y_k^{(2)} + \cdots. \tag{A.2}$$

By plugging equation A.2 into equation A.1, we obtain

$$\begin{aligned}\partial_t \left(\eta y_k^{(1)} + \eta^2 y_k^{(2)} + \cdots\right) = \eta I(t)\delta_{k0} &+ \frac{\partial f_k(y^{(0)})}{\partial y_m} \eta y_m^{(1)} \\ &+ \frac{1}{2!} \frac{\partial^2 f_k(y^{(0)})}{\partial y_m \partial y_n} \eta^2 y_m^{(1)} y_n^{(1)} \\ &+ \frac{\partial f_k(y^{(0)})}{\partial y_m} \eta^2 y_m^{(2)} + \cdots.\end{aligned}$$

By comparing two sides order by order, we obtain a series of equations satisfied by the perturbative expansion at each order as

$$\mathcal{D}_{km} y_m^{(1)} = I\delta_{k0}, \quad \mathcal{D}_{km} = \delta_{km}\partial_t - \left.\frac{\partial f_k(x^{(0)})}{\partial x_m}\right|_{y^{(0)}},$$

which is equation 2.2,

$$\mathcal{D}_{km} y_m^{(2)} = \frac{1}{2!} H_{kmn} y_m^{(1)} y_n^{(1)}, \quad H_{kmn} = \frac{\partial^2 f_k(y^{(0)})}{\partial y_m \partial y_n}, \tag{A.3}$$

which is equation 2.9, and so on.

These equations can be solved recursively by using a kernel $\epsilon^{(1)} = \mathcal{D}^{-1}$, which therefore satisfies

$$\mathcal{D}_{km}\epsilon_{ml}^{(1)}(t, t') = \delta_{kl}\delta(t - t').$$

Now the solution of the first order can be written as

$$y_k^{(1)}(t) = \int dt'\, \epsilon_k^{(1)}(t - t') I(t'),$$

where $\epsilon_k^{(1)} = \epsilon_{k0}^{(1)}$. Also from equation A.3,

$$\begin{aligned} y_k^{(2)} &= \frac{1}{2!}\int dt'\, \epsilon_{kl}^{(1)}(t - t') \\ &\quad \times \left(H_{lmn} \int ds_1 ds_2\, \epsilon_m^{(1)}(t' - s_1)\epsilon_n^{(1)}(t' - s_2) I(s_1) I(s_2) \right) \\ &= \frac{1}{2!}\int ds_1 ds_2\, \epsilon_k^{(2)}(t - s_1, t - s_2) I(s_1) I(s_2), \end{aligned}$$

$$\epsilon_k^{(2)}(s_1, s_2) = \int dt'\, H_{lmn}\epsilon_{kl}^{(1)}(t')\epsilon_m^{(1)}(s_1 - t')\epsilon_n^{(1)}(s_2 - t'),$$

which is equation 2.10. Since this procedure can be carried out to higher orders, the higher-order kernels are outer products of the linear kernels.

In addition, the first-order kernels are related to each other by simple time differentiation. To show this, we rewrite equation 2.4 as follows:

$$\epsilon_m^{(1)}(t - t') = \mathcal{D}_{mn}^{-1}\delta_{n0}\delta(t - t'). \tag{A.4}$$

Now the derivatives can be computed as

$$\begin{aligned}\partial_t \epsilon_m^{(1)} &= \partial_t \mathcal{D}_{mn}^{-1} \delta_{n0} \delta(t-t') \\ &= (\partial_t \delta_{ml} - J_{ml} + J_{ml}) \mathcal{D}_{ln}^{-1} \delta_{n0} \delta(t-t') \\ &= \mathcal{D}_{ml} \mathcal{D}_{ln}^{-1} \delta_{n0} \delta(t-t') + J_{ml} \epsilon_l^{(1)} \\ &= \delta_{m0} \delta(t-t') + J_{ml} \epsilon_l^{(1)}. \end{aligned} \tag{A.5}$$

This shows that the nonsingular part of the first-order kernels can be obtained by the linear combination of the derivatives of other kernels.

Appendix B: A Linear Model from the FitzHugh-Nagumo System

Here we derive the linearization of the FitzHugh-Nagumo model, beginning with equations 1.5 and 1.6. The fixed point can be obtained by simultaneously solving the nullcline equations. We denote it by (V_0, W_0) and expand the system around this point.

First, the Jacobian is

$$J = \begin{pmatrix} F'(V_0)/\psi - 1/\psi \\ 1 - b \end{pmatrix},$$

where $F(V) = V(1-V)(a+V)$, and this defines a linear system:

$$\begin{aligned}\partial_t V &= \frac{F'(V_0)}{\psi} V - \frac{1}{\psi} W + I(t) \\ \partial_t W &= V - bW.\end{aligned}$$

J has eigenvalues $\lambda_\pm$:

$$\begin{aligned}\lambda_\pm &= \frac{1}{\psi}\left[f_-(V_0) \pm \sqrt{f_+(V_0)^2 - \psi} \right] \\ &= -b + \left[f_+(V_0) \pm \sqrt{f_+(V_0)^2 - \psi} \right] \\ &= -b + \kappa_\pm, \end{aligned} \tag{B.1}$$

where $f_\pm(V) = (F'(V) \pm b\psi)/2$. As in equation 2.6, J is diagonalized by a matrix U,

$$U = \frac{1}{\lambda_+ - \lambda_-} \begin{pmatrix} 1 & \lambda_- + b \\ -1 & -(\lambda_+ + b) \end{pmatrix}.$$

From equation 2.8, we obtain the first-order kernels, which solve the linear system

$$\epsilon_V^{(1)}(t) = e^{-bt}\partial_t S(t)H(t),$$
$$\epsilon_W^{(1)}(t) = e^{-bt}S(t)H(t), \tag{B.2}$$

where $S(t) = (e^{\kappa_+ t} - e^{\kappa_- t})/(\kappa_+ - \kappa_-)$. $\epsilon_{V,W}^{(1)}(t)$ with our choice of parameters is drawn in Figure 2a.

Appendix C: Derivation of the Abbott-Kepler Model

In this section, we show the derivation of a two-dimensional neuron model used in section 5. For further details, we refer to the original paper (Abbott & Kepler, 1990).

We begin with a Hodgkin-Huxley equation, which is defined by equation 1.1 and the following parameters:

$$g_L = \bar{g}_L, \qquad g_K = \bar{g}_K n^4, \qquad g_{Na} = \bar{g}_{Na} m^3 h. \tag{C.1}$$

$$\tau_z(V)\frac{dz}{dt} = \bar{z}(V) - z, \qquad \tau_z = \frac{1}{\alpha_z + \beta_z},$$
$$\bar{z} = \frac{\alpha_z}{\alpha_z + \beta_z}, \qquad z = m, n, h \tag{C.2}$$

$$\alpha_m = \frac{.1(V+40)}{1 - \exp[-.1(V+40)]}, \qquad \beta_m = 4\exp[-.0556(V+65)],$$
$$\alpha_h = .07\exp[-.05(V+65)], \qquad \beta_h = \frac{1}{1+\exp[-.1(V+35)]}, \tag{C.3}$$
$$\alpha_n = \frac{.01(V+55)}{1 - \exp[-.1(V+55)]}, \qquad \beta_n = .125\exp[-.0125(V+65)].$$

The key observation is that τ_m is much smaller than $\tau_{n,h}$, while τ_h and τ_n are mutually comparable. Therefore, m can be approximated by its value at equilibrium, $\bar{m}(V)$, and h and n can be represented by the same equilibrium voltage, U. In other words,

$$m \approx \bar{m}(V), \qquad h \approx \bar{h}(U), \qquad n \approx \bar{n}(U).$$

Using this, we obtain equation 5.1 as

$$\begin{aligned} F &= \sum_{i=L,K,Na} g_i(V - E_i) \\ &\approx \bar{g}_L(V - E_L) + \bar{g}_K \bar{n}(U)^4(V - E_K) + \bar{g}_{Na}\bar{m}(V)^3\bar{h}(U)(V - E_{Na}) \\ &= -f(V, U). \end{aligned}$$

An equation for U, equation 5.2, is obtained by requiring that time dependence of the active current F due to h and n in the Hodgkin-Huxley model is mimicked by $\bar{h}(U)$ and $\bar{n}(U)$. This implies

$$\frac{dF}{dh}\frac{dh}{dt} + \frac{dF}{dn}\frac{dn}{dt} \approx \left(\frac{\partial f}{\partial \bar{h}}\frac{d\bar{h}}{dU} + \frac{\partial f}{\partial \bar{n}}\frac{d\bar{n}}{dU}\right)\frac{dU}{dt}. \tag{C.4}$$

Again, we approximate equation C.2 in the same way,

$$\frac{dz}{dt} \approx \frac{1}{\tau_z(V)}\left(\bar{z}(V) - \bar{z}(U)\right), \qquad z = h, n.$$

Plugging this in equation C.4, we can solve for dU/dt as a function of V and U, which we denoted by $g(V, U)$ in equation 5.2 as

$$g(V, U) = \frac{\bar{g}_{Na}(V - E_{Na})\bar{m}(V)^3(\bar{h}(V) - \bar{h}(U))/\tau_h(V) + 4\bar{g}_K(V - E_K)\bar{n}(U)^3(\bar{n}(V) - \bar{n}(U))/\tau_n(V)}{\bar{g}_{Na}(V - E_{Na})\bar{m}(V)^3\bar{h}'(U) + 4\bar{g}_K(V - E_K)\bar{n}(U)^3\bar{n}'(U)}.$$

References

Abbott, L. F., & Kepler, T. (1990). Model neurons: From Hodgkin-Huxley to Hopfield. In L. Garrido (Ed.), *Statistical mechanics of neural networks* (pp. 5–18). Berlin: Springer-Verlag.

Agüera y Arcas, B., & Fairhall, A. (2003). What causes a neuron to spike? *Neural Computation, 15*, 1715–1749.

Agüera y Arcas, B., Fairhall, A., & Bialek, W. (2003). Computation in a single neuron: Hodgkin and Huxley revisited. *Neural Computation, 15*, 1789–1807.

Aldworth, Z., Miller, J., Gedeon, T., Cummins, G., & Dimitrov, A. (2005). Dejittered spike-conditioned stimulus waveforms yield improved estimates of neuronal feature selectivity and spike timing precision of sensory interneurons. *J. Neurosci., 25*, 5323–5332.

Azouz, R., & Gray, C. (2000). Dynamic spike threshold reveals a mechanism for synaptic coincidence detection in cortical neurons in vivo. *Proc. Natl. Acad. Sci. USA, 97*, 8110–8115.

Berry II, M. J., & Meister, M. (1999). The neural code of the retina. *Neuron, 22*, 435–450.

Bialek, W., & de Ruyter van Steveninck, R. R. (2005). *Features and dimensions: Motion estimation in fly vision.* Preprint, arXiv:q-bio/0505003 [q-bio.NC].

Borst, A., Flanagin, V. L., & Sompolinsky, H. (2005). Adaptation without parameter change: Dynamic gain control in motion detection. *Proc. Natl. Acad. Sci. USA, 102*(17), 6172–6176.

Brenner, N., Bialek, W., & de Ruyter van Steveninck, R. R. (2000). Adaptive rescaling maximizes information transmission. *Neuron, 26*, 695–702.

de Ruyter van Steveninck, R. R., & Bialek, W. (1988). Real-time performance of a movement sensitive in the blowfly visual system: Information transfer in short spike sequences. *Proc. Roy. Soc. Lond. B, 234*, 379–414.

DeWeese, M. (1995). *Optimization principles for the neural code.* Unpublished doctoral dissertation, Princeton University.

Fairhall, A. L., Burlingame, C. A., Narasimhan, R., Harris, R. A., Puchalla, J. L., & Berry II, M. J. (2006). Selectivity for multiple stimulus features in retinal ganglion cells. *J. Neurophysiol., 96*, 2724–2738.

FitzHugh, R. (1961). Impulse & physiological states in models of nerve membrane. *Biophysics J., 1*, 445–466.

Gerstner, W., & Kistler, W. (2002). *Spiking neuron models: Single neurons, populations, plasticity*. Cambridge: Cambridge Press.

Gradshteyn, I. S., & Ryzhik, I. M. (2000). *Table of integrals, series, and products* (6th ed.). San Diego: Academic Press.

Horwitz, G., Chichilnisky, E., & Albright, T. (2005). Blue-yellow signals are enhanced by spatiotemporal luminance contrast in macaque V1. *J. Neurophysiol., 93*(4), 2263–2278.

Hutcheon, B., & Yarom, Y. (2000). Resonance, oscillation and the intrinsic frequency preferences of neurons. *Trends Neurosci., 23*(5), 216–222.

Huys, Q. J., Ahrens, M. B., & Paninski, L. (2006). Efficient estimation of detailed single-neuron models. *J Neurophysiol., 96*, 872–890.

Izhikevich, E. (2001). Resonate-and-fire neurons. *Neural Networks, 14*, 883–894.

Izhikevich, E. M. (2006). *Dynamical systems in neuroscience: The geometry of excitability and bursting.* Cambridge, MA: MIT Press.

Keat, J., Reinagel, P., Reid, R. C., & Meister, M. (2001). Predicting every spike: A model for the responses of visual neurons. *Neuron, 30(3)*, 803–817.

Kistler, W., Gerstner, W., & van Hemmen, J. L. (1997). Reduction of the Hodgkin-Huxley equations to a single-variable threshold model. *Neural Computation, 9*, 1015–1045.

Maravall, M., Petersen, R. S., Fairhall, A. L., Arabzadeh, E., & Diamond, M. E. (2007). Shift in coding properties and maintenance of information transmission during adaptation in barrel cortex. *PLoS Biol., 5*, e19.

Marmarelis, P. Z., & Marmarelis, V. Z. (1974). Measurement of the Wiener kernels of a non-linear system by cross-correlation. *International Journal of Control, 2*, 234–254.

Marmarelis, V. (2004). *Nonlinear dynamic modeling of physiological systems*. Piscataway, NJ: IEEE Press.

Mehta, M. L. (1991). *Random matrices* (3rd ed.). New York: Academic Press.

Nagumo, J., Arimoto, S., & Yoshikawa, Z. (1962). An active pulse transmission line simulating nerve axon. *Proc. IRE, 50*, 2061–2071.

Paninski, L., Lau, B., & Reyes, A. (2003). Noise-driven adaptation: in vitro and mathematical analysis. *Neurocomputing, 52*, 877–883.

Paninski, L., Pillow, J., & Simoncelli, E. (2004). Maximum likelihood estimation of a stochastic integrate-and-fire neural encoding model. *Neural Comp., 16*, 2533–2561.

Petersen, R. (2003). Coding of dynamical whisker stimulation in the rat somatosensory cortex: A spike-triggered covariance analysis. *Society for Neuroscience*, 57.1.

Pillow, J. W., Paninski, L., Uzzell, V. J., Simoncelli, E. P., & Chichilnisky, E. J. (2005). Prediction and decoding of retinal ganglion cell responses with a probabilistic spiking model. *J. Neurosci., 25*(47), 11003–11013.

Pillow, J. W., & Simoncelli, E. P. (2003). Biases in white noise analysis due to non-Poisson spike generation. *Neurocomputing, 52–54*, 109–115.

Powers, R., Dai, Y., Bell, B., Percival, D., & Binder, M. (2005). Contributions of the input signal and prior activation history on the discharge behavior of rat motoneurones. *J. Physiol., 562*, 707–724.

Rieke, F. (1991). *Physical principles underlying sensory processing and computation*. Unpublished doctoral dissertation, University of California, Berkeley.

Rieke, F., Warland, D., Bialek, W., & de Ruyter van Steveninck, R. R. (1997). *Spikes: Exploring the neural code*. Cambridge, MA: MIT Press.

Rust, N., Schwartz, O., Movshon, J., & Simoncelli, E. (2005). Spatiotemporal elements of macaque V1 receptive fields. *Neuron, 46*, 945–956.

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

Slee, S. J., Higgs, M. H., Fairhall, A. L., & Spain, W. J. (2005). Two-dimensional time coding in the auditory brainstem. *J. Neurosci., 25*(43), 9978–9988.

Touryan, J., Lau, B., & Dan, Y. (2002). Isolation of relevant visual features from random stimuli for cortical complex cells. *J. Neurosci., 22*, 10811–10818.

Truccolo, W., Eden, U. T., Fellows, M. R., Donoghue, J. P., & Brown, E. N. (2005). A point process framework for relating neural spiking activity to spiking history, neural ensemble, and extrinsic covariate effects. *J. Neurophysiol., 93*(2), 1074–1089.

Victor, J. D., & Shapley, R. M. (1979). The nonlinear pathway of Y ganglion cells in the cat retina. *J. Gen. Physiol., 74*(6), 671–689.

Victor, J., & Shapley, R. (1980). A method of nonlinear analysis in the frequency domain. *Biophys. J., 29*(3), 459–483.

Volterra, V. (1930). *Theory of functions and of integral and integrodifferential equations*. New York: Dover.

Wiener, N. (1958). *Nonlinear problems in random theory.* Cambridge, MA: MIT Press.

Yu, Y., & Lee, T. S. (2003). Dynamical mechanisms underlying contrast gain control in single neurons. *Phys. Rev. E, 68*(1 Pt. 1), 011901–011901.

Received May 31, 2006; accepted November 10, 2006.