

# Theoretical analysis of reverse-time correlation for idealized orientation tuning dynamics

Gregor Kovačič · Louis Tao · David Cai · Michael J. Shelley

Received: 21 January 2007 / Revised: 21 January 2008 / Accepted: 19 February 2008  
© Springer Science + Business Media, LLC 2008

**Abstract** A theoretical analysis is presented of a reverse-time correlation method used in experimentally investigating orientation tuning dynamics of neurons in the primary visual cortex. An exact mathematical characterization of the method is developed, and its connection with the Volterra–Wiener nonlinear systems theory is described. Various mathematical consequences and possible physiological implications of this analysis are illustrated using exactly solvable idealized models of orientation tuning.

**Keywords** Random walks · Mexican hat · Primary visual cortex · Orientation tuning dynamics

---

**Action Editor: Jonathan D. Victor**

---

G. Kovačič (✉)  
Rensselaer Polytechnic Institute,  
Troy, NY, USA  
e-mail: kovacg@rpi.edu

L. Tao  
New Jersey Institute of Technology, Newark, NJ, USA  
e-mail: tao@njit.edu

*Present Address:*  
L. Tao  
Center for Bioinformatics, School of Life Sciences,  
Peking University, Beijing, China

D. Cai · M. J. Shelley  
Courant Institute of Mathematical Sciences,  
New York University, New York, NY, USA

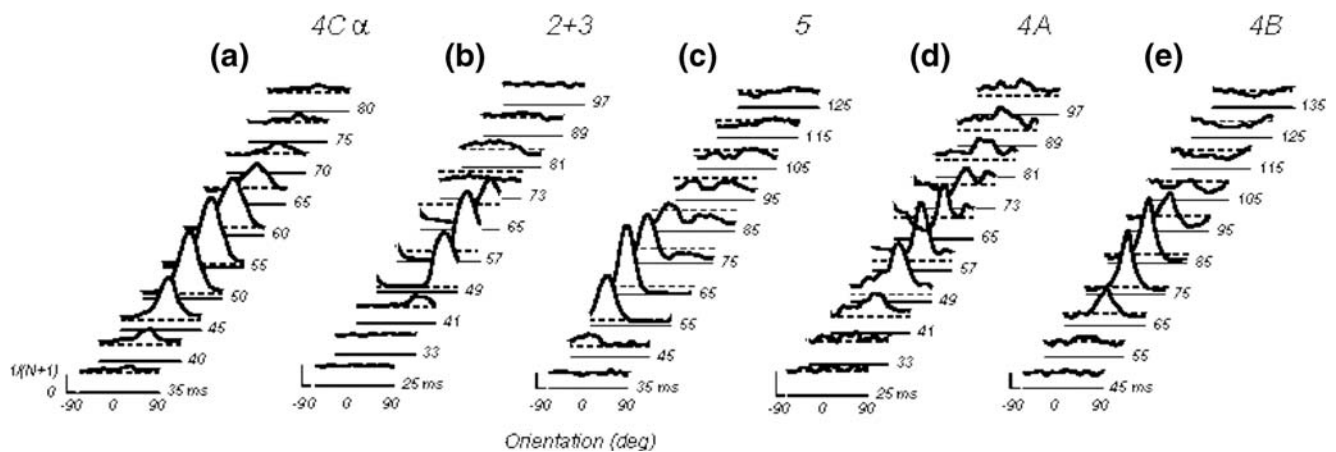
D. Cai  
e-mail: cai@cims.nyu.edu

M. J. Shelley  
e-mail: shelley@cims.nyu.edu

## 1 Introduction

The precise mechanism responsible for the orientation selectivity of neurons in the mammalian primary visual cortex (V1) remains one of the central mysteries in modern neuroscience. Experimentally, orientation tuning curves are elicited from neurons most often by using bar-like or grating-like stimuli. For example, drifting or contrast-reversing gratings yield stationary tuning curves and the corresponding orientation preference of the probed neurons. To find the time-course of the given neuron's orientation tuning, more sophisticated grating sequences are used, such as the quasi-random sequences used in the experiments of Ringach et al. (1997a, 2003), Ringach and Shapley (2004), Xing et al. (2005), which are combined with the reverse-time correlation (RTC) analysis of the measured neuronal spike trains. In these works, correlating the spike train of a neuron against a time series of randomly presented stimulus angles constructs  $\mathcal{P}(\tau, \theta)$ , the probability that  $\tau$  milliseconds before a spike was produced, a grating with angle  $\theta$  was presented. This probability is closely related to the typical tuning response of a cortical neuron to a specific orientation embedded within a long, arbitrary stimulus. In this second interpretation of  $\mathcal{P}(\tau, \theta)$ ,  $\tau$  is simply the response time. Figure 1 shows a series of time slices through experimentally measured  $\mathcal{P}(\tau, \theta)$  for a number of neurons in various V1 layers in macaque. In the evolution of  $\mathcal{P}(\tau, \theta)$  with  $\tau$  is seen the dynamics of tuning by the cortex, with time-scales for the onset of selectivity, the development of sharpened response, and, in some cases, suppression of certain angle responses in the later stages of the tuning process.

RTC analysis is closely related to the Wiener series representation of the input-output relation in nonlinear



**Fig. 1** Orientation tuning dynamics of neurons in various V1 layers measured experimentally in Ringach et al. (1997a) [reprinted from Ringach et al. (1997a) by permission from Macmillan Publishers Ltd., Copyright 1997]. Suppression of later-time response to the optimal orientation is visible in neurons (b) and (e).

Mexican-hat tuning curves are present in the neurons (c), (d), and (e). This time evolution of the tuning curve shape is believed to be a result of the dynamic competition between cortical excitation and inhibition (Nelson and Frost 1978; Sillito et al. 1980; Shapley et al. 2003; Ringach and Shapley 2004; Xing et al. 2005)

systems (Wiener 1958). The Wiener series is an orthogonalized counterpart of the Volterra series (Volterra 1930), a representation of the output signal in a given nonlinear system as a series whose terms are convolutions of time-delayed copies of the input signal with multi-dimensional kernels. The Wiener representation became a practical tool after the discovery by Lee and Schetzen (1965) that the leading-order Wiener kernels can be obtained by an appropriate reverse-time correlation of the output signal against a product of multiple time-delayed copies of the Gaussian white-noise input signal. Comprehensive theoretical accounts and reviews of the Volterra–Wiener theory are presented in Marmarelis and Marmarelis (1978), Schetzen (1980), Rugh (1981), Victor (1992), Matthews and Sicuranza (2000), with Marmarelis and Marmarelis (1978), Victor (1992) concentrating on physiological systems. An extensive bibliography of books and articles addressing and using the Volterra–Wiener theory in both engineering and the biosciences is given in Giannakis and Serpedin (2001).

The Wiener series representation and RTC analysis have been used to study sensory systems in a range of neurophysiological systems, including, for example, cochlear resolution and encoding (de Boer 1967; de Boer and de Jongh 1978), visual systems of insects (Marmarelis and McCann 1973; McCann 1974; de Ruyter van Steveninck and Bialek 1988; Rieke et al. 1996; Juusola and French 1997; van Kleef et al. 2005) and horseshoe crabs (Brodie et al. 1978a, b), retinal neurons of catfish (Naka et al. 1975) and cats (Victor and Shapley 1978), and contrast response (Ohzawa et al. 1985) and directional selectivity (Citron

and Emerson 1983) of cat cortical cells. A collection of papers on the use of RTC analysis in visual systems is provided in Pinter and Nabet (1992). Some more recent works include using RTC to study the dynamics of color tuning in macaque V1 (Cottaris and DeValois 1998), velocity and direction selectivity in V1 and MT cortical areas of cats and macaques (Borghuis et al. 2003; Pack et al. 2006; Vajda et al. 2006), spatio-temporal and spectro-temporal receptive fields of neurons in the primary auditory cortex (Brugge et al. 2001; Rutkowski et al. 2002; Miller et al. 2002; Klein et al. 2006), tuning characteristics of auditory nerve fibers and the ventral cochlear nuclei of frogs (Yamada and Lewis 1999) and mammals (Lewis et al. 2002; Recio-Spinoso et al. 2005; Recio-Spinoso and Van Dijk 2006), the enhancing of detection of rapid and coincident auditory stimuli in auditory neurons (Svirskis et al. 2003), temporal filter properties of neurons in the zebrafish optic tectum (Ramdya et al. 2006), spike-timing dependent receptive field plasticity of tadpole tectal neurons (Vislay-Meltzer et al. 2006), the relationship between physiological receptive fields and feature selectivity on the one hand and psychophysical perceptive fields and feature detection on the other (Ringach 1998; Neri 2004a, b; Neri and Levi 2006), the average relative dependence of neuronal spiking on the injected currents and prior spiking history (Powers et al. 2006), and the relationship between dimensionality reduction in the stimulus space and feature selectivity (Aguera y Arcas et al. 2003).

RTC and Wiener kernel analysis have also been successfully used to map out receptive fields of LGN (Reid and Shapley 1992; Reid et al. 1997; Wolfe and Palmer

1998; Reid and Shapley 2002; Alitto et al. 2005) and V1 (Jones and Palmer 1987; DeAngelis et al. 1993; Pei et al. 1994; DeAngelis et al. 1995; Ringach 2002; Borghuis et al. 2003) neurons in cats and monkeys. Such mappings have by now also become integral parts of more general studies, including of the shrinking of V1 neuronal receptive field sizes during long stimuli (Suder et al. 2002), overlap of “on” and “off” regions in V1 neuronal receptive fields (Mata and Ringach 2004), the distinctions between the receptive fields of “bursting” and “tonic” LGN spikes (Alitto et al. 2005), the effects of GABA inhibition blocking on the receptive field components in cat V1 neurons (Pernberg et al. 1998), and the effect of drifting grating stimuli in the surround on the classical receptive field properties of cat V1 neurons (Palmer and Nafziger 2002). Visual stimuli used in the mapping procedure are usually white-noise-like in space and time, however, using these if fine spatial resolution is required becomes a very complex task in practice. A simplified, approximate RTC procedure was devised in Ringach et al. (1997b), in which the white-noise-in-space-and-time stimuli are replaced by suitable spatially orthonormal systems of functions, whose shapes are chosen so as to best capture the salient features of the receptive fields in question. Using random sequences of such functions as stimuli and cross-correlating the resulting spike trains against them reduces the complexity of the receptive field reconstruction at the linear order via RTC without significantly affecting its accuracy. The results of Ringach (2002) were obtained by the method of Ringach et al. (1997b), and approximate methods similar to it were also developed and used in Cottaris and DeValois (1998), Borghuis et al. (2003).

The approximate RTC method of Ringach et al. (1997b) was also used in Ringach et al. (1997a, 2003) to study the dynamics of neuronal orientation tuning in macaque primary visual cortex. Related works (Celebrini et al. 1993; Pei et al. 1994; Volgushev et al. 1995; Gillespie et al. 2001; Müller et al. 2001; Ringach et al. 2002; Bredfeldt and Ringach 2002; Mazer et al. 2002; Nishimoto et al. 2005) studied orientation and also spatial-frequency tuning dynamics in the V1 of cats and macaques by using similar methods. The RTC method of Ringach et al. (1997b) was also used to probe the dependence of the orientation tuning dynamics on the stimulus size (Xing et al. 2005; Chen et al. 2005), and the dependence of orientation selectivity on adaptation to the orientation of prior drifting-grating stimuli (Dragoi et al. 2002; Felsen et al. 2002).

In this paper, we develop a precise mathematical description related to the experimental protocol used in Ringach et al. (1997a, 2003), which includes a general

interpretation of the stimuli used in Ringach et al. (1997a, 2003) as a collection of correlated binary noise stimuli. We find expressions for the measured probability function  $\mathcal{P}(\tau, \theta)$  in terms of both time-averaged and ensemble-averaged firing rates. Our description allows us to show how the experimentally measured quantities relate to the corresponding first-order Wiener kernels. (We note that in this paper, we reserve the term Wiener kernel only for the kernels associated with the Gaussian white noise input.) In particular, we use the probability  $\mathcal{P}(\tau, \theta)$  and (a correlated version of) the Gaussian white noise limit of binary noise stimuli to compute the first-order Wiener kernels exactly for a set of simple leaky feed-forward and ring (Ben-Yishai et al. 1995; Pugh et al. 2000) models of orientation tuning dynamics. Surprisingly, we show that they vanish identically due to the averaging over the spatial phase used in the experiment. In this way we conclude that, in the models we study, the variations of the measured  $\mathcal{P}(\tau, \theta)$  in  $\tau$  and  $\theta$  do not arise from the (vanishing) linear Wiener kernels, but rather from the (nonlinear) effects of the higher-order terms in the Wiener series. This important observation leads, in the remainder of the paper, to the discussion of how the natural interpretation of  $\mathcal{P}(\tau, \theta)$ , as the average dynamical response of a neuron to a finite-duration flash of a specific orientation embedded in a long random stimulus, yields useful information about the structure of cortico-cortical connections. Along the way we also show, in amplification of the result in Pugh et al. (2000), that a class of purely feed-forward V1 neuronal models with no leakage exhibits no orientation tuning dynamics. This is consistent with the view that V1 neuronal orientation tuning may not be a purely feed-forward process (Nelson and Frost 1978; Sillito et al. 1980; Sompolinsky and Shapley 1997; Shapley et al. 2003).

As mentioned in the first paragraph and demonstrated in Fig. 1, the time evolution of the probability  $\mathcal{P}(\tau, \theta)$  reveals various temporal and spatial scales involved in the V1 neuronal orientation tuning dynamics. Questions that experimentalists have been addressing include what is the underlying mechanism responsible for the dynamic change of the shape of orientation tuning curves (Pei et al. 1994; Volgushev et al. 1995; Ringach et al. 1997a, 2002; Bredfeldt and Ringach 2002; Ringach et al. 2003; Sompolinsky and Shapley 1997; Shapley et al. 2003; Ringach and Shapley 2004), and whether these curves are separable in space and time and thus just scale in time (Celebrini et al. 1993; Gillespie et al. 2001; Müller et al. 2001; Mazer et al. 2002; Sharon and Grinvald 2002). The idealized models in this paper assume explicitly that the LGN filter processing the stimulus and feeding it to a V1

neuron is separable; in other words, the integration kernel used in our LGN model is the product of a spatial and a temporal function. Thus, in the absence of any additional cortical processing, the tuning curves of a V1 neuron should simply scale in time. This property is clearly retained in our feedforward models in the limit of short presentation times even though a nonlinear operation, i.e., thresholding of the stimuli arriving from the LGN, is used. On the other hand, we present a simple, idealized, coupled, linear V1 firing-rate model, which accounts for the interaction among cortical neurons, and even this simple model already imparts nontrivial dynamical shape changes to the V1 neuronal orientation tuning curves. A particularly striking case of such a shape change is the “Mexican-hat” tuning curve, which likely emerges as a consequence of the interaction between cortical excitation and inhibition in the real cortex (Nelson and Frost 1978; Sillito et al. 1980; Shapley et al. 2003; Ringach and Shapley 2004; Xing et al. 2005). Our idealized coupled model also exhibits Mexican-hat orientation tuning curves and accounts for many of their expected properties. For example, this model predicts the existence of Mexican-hat tuning curves in some cases when the extent of the local cortical inhibition is half of the extent of the local cortical excitation, which is one possible interpretation of physiological data in V1 and is used in the large computational model of McLaughlin et al. (2000).

The paper is organized as follows. In Section 2, we present a description of the probability function  $\mathcal{P}(\tau, \theta)$  in terms of time- and ensemble-averaged neuronal firing rates. In Section 3, we present the description of the experimental stimuli as a collection of correlated binary noise processes, and develop a Volterra–Wiener-type description of the experimental procedure of Ringach et al. (1997a, 2003). In Section 4, we present an idealized mean-field model, derived from the mean-field model of Shelley and McLaughlin (2002), with Gabor-like LGN stimulus filtering, and discuss the orientation tuning properties and the first-order Wiener kernel for the leaky-feedforward version of the idealized model. Finally, in Section 5, we use a linearized, coupled, excitatory-inhibitory version of the idealized model to discuss properties of Mexican-hat tuning curves. In Appendix 1, we present a probabilistic lemma used throughout the paper. In Appendix 2, we discuss the errors introduced in the reconstruction of the Wiener kernels by approximating Gaussian white noise stimuli by a collection of binary noise processes with finite presentation times. Detailed calculations used to obtain the results presented in Sections 3, 4, and 5 are rendered in Appendices 3, 4, and 7, respectively. In Appendix 5, we present the calculation showing that

the pure feedforward version of the idealized model with no leakage exhibits no orientation tuning at all. Appendix 6 contains the description of a numerical integrate-and-fire ring model, which we use to verify some assertions of Section 4.

## 2 Mathematical description of the experimental procedure

The experimental procedure employed in Ringach et al. (1997a, 2003) uses as the stimulus a sequence of randomly-flashed gratings with a single spatial frequency but different orientations  $\theta_i$  and spatial phases  $\phi_j$ . In particular, in each such sequence, the grating orientation angles  $\theta_i$  and spatial phases  $\phi_j$  are chosen randomly and independently from some sufficiently fine uniform distributions. A new grating is presented every  $\nu$  milliseconds, with the presentation interval  $\nu$  corresponding to the screen refresh rates of 60 Hz in Ringach et al. (1997a) and 50 Hz in the later works (Ringach et al. 2002, 2003; Bredfeldt and Ringach 2002).

In the course of the experimental procedure, the resulting spike train from the V1 neuron under investigation is recorded. For any given time  $\tau$ , a histogram of orientation angles  $\theta_i$  and spatial phases  $\phi_j$  is assembled as follows: If at the time  $t$  the neuron fired a spike, one finds the angle  $\theta_{i_0}$  and the phase  $\phi_{j_0}$  which were presented at the time  $t - \tau$ , and increments the corresponding box in the histogram. After normalizing by the total number of spikes, one obtains the function  $\mathcal{Q}(\tau, \theta_i, \phi_j)$ , which gives the probability that,  $\tau$  milliseconds before a spike, a grating with the angle-phase pair  $(\theta_i, \phi_j)$  was presented. Hence,  $\tau$  is the reverse correlation time.

Since the studies of Ringach et al. (1997a) were only concerned with the dynamics of orientation tuning, the histogram that resulted from the experiment was further averaged over the spatial phases  $\phi_j$ . This was also deemed necessary because minute random eye movements during the experiment prevent accurate measurement of the preferred neuronal spatial phase. Through this averaging and again after the normalization by the total number of spikes, the marginal probability distribution function  $\mathcal{P}(\tau, \theta_i)$  was generated. This function gives the probability that,  $\tau$  milliseconds before a spike, a grating with the orientation  $\theta_i$  was presented regardless of the spatial phase. A series of time-slices through such probability density functions for various V1 layers, obtained experimentally in Ringach et al. (1997a), is reproduced in Fig. 1. These time-slices are interpreted as dynamic tuning curves, a point to which we will return later in this section.



We now describe a number of mathematical quantities that may serve as continuous-time equivalents of the experimental histograms. In the subsequent sections, we discuss the relationship of these quantities to concepts in nonlinear systems theory as well as to various neuronal network mechanisms and cortical architectures.

We begin with a mathematical description of the stimulus used in the experiments of Ringach et al. (1997a). If  $\mathbf{X} = (X, Y)$  is the location on the screen and  $t$  is time, the light intensity  $\mathcal{I}(\mathbf{X}, t)$  of the stimulus on the screen is described by the formula

$$\mathcal{I}(\mathbf{X}, t) = \mathcal{I}_0 v^{-1/2} \{1 + \varepsilon \sin[\mathbf{k}(\theta^{(n)}) \cdot \mathbf{X} - \phi^{(n)}]\}, \quad nv < t < (n + 1)v, \tag{1}$$

with  $n = \dots, -2, -1, 0, 1, 2, \dots$ . Here  $\mathcal{I}_0 v^{-1/2}$  and  $\varepsilon$  are the intensity and the contrast of the stimulus, respectively,  $\mathbf{k}(\theta^{(n)}) = k(-\sin \theta^{(n)}, \cos \theta^{(n)})$ , and  $k/2\pi$  is the spatial frequency of the stimulus. The orientation angles  $\theta^{(n)}$  are drawn randomly and independently every  $v$  milliseconds from the discrete uniform distribution of angles  $\theta_i = \pi i/I, i = 0, \dots, I - 1$ , on  $[0, \pi)$ , and the spatial phases  $\phi^{(n)}$  from the discrete uniform distribution of phases  $\phi_j = 2\pi j/J, j = 0, \dots, J - 1$  on  $[0, 2\pi)$ . Here,  $I$  and  $J$  are some sufficiently large integers. In the experiments of Ringach et al. (1997a),  $I$  varied between 15 and 60, and  $J = 4$ ; the majority of the neurons investigated in the experiments of Ringach et al. (2003) were probed by stimuli with  $I = 18$  and  $J = 8$  (the experiments of Ringach et al. (2003) also included occasional “blanks,” i.e. gratings with no contrast. These can easily be incorporated in the theory developed here).

We view the orientations  $\theta^{(n)}$  and spatial phases  $\phi^{(n)}$  as forming two random, bi-infinite sequences

$$\vec{\theta} = (\dots, \theta^{(-2)}, \theta^{(-1)}, \theta^{(0)}, \theta^{(1)}, \theta^{(2)}, \dots) \quad \text{and} \\ \vec{\phi} = (\dots, \phi^{(-2)}, \phi^{(-1)}, \phi^{(0)}, \phi^{(1)}, \phi^{(2)}, \dots), \tag{2}$$

specific to each experimental run. We have scaled the intensity of the stimulus with the factor  $v^{-1/2}$  in Eq. (1) to facilitate the discussion of the “white noise” limit,  $v \rightarrow 0$  (while this scaling does not reflect any realistic physiological case, it elucidates the relationship between the particular RTC method discussed in this paper and the general Volterra–Wiener theory in an especially clear fashion). Note that we are using subscripts, as in  $\theta_i$ , to denote the values of the orientation angles (and spatial phases), and superscripts in parentheses, as in  $\theta^{(k)}$ , to denote the time slots in which they are presented, that is, their positions in the sequence  $\vec{\theta}$  (and  $\vec{\phi}$ ).

For any given pair of sequences  $\vec{\theta}$  and  $\vec{\phi}$ , as in Eq. (2), of the presented grating orientation angles and spatial phases, the correlation functions

$$\mathcal{N}(\tau, \theta_i, \phi_j) = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T \chi_{\theta_i, \phi_j}(t) m(t + \tau, \vec{\theta}, \vec{\phi}) dt, \tag{3a}$$

$$\mathcal{M}(\tau, \theta_i) = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T \chi_{\theta_i}(t) m(t + \tau, \vec{\theta}, \vec{\phi}) dt, \tag{3b}$$

are convenient theoretical counterparts to the histograms used to compute the probability distributions  $\mathcal{Q}(\tau, \theta_i, \phi_j)$  and  $\mathcal{P}(\tau, \theta_i)$  introduced at the beginning of this section. In formulas (3),  $m(t, \vec{\theta}, \vec{\phi})$  is the firing rate generated by the neuron in question under the stimulus (1), and  $\chi_{\theta_i, \phi_j}(t)$  and  $\chi_{\theta_i}(t)$  are the indicator functions of the unions of the time intervals (of the type  $nv < t < (n + 1)v$ , where  $n$  is an integer), during which the orientation-phase pair  $(\theta_i, \phi_j)$  and the angle  $\theta_i$  were presented, respectively. In other words, the correlation functions  $\mathcal{N}(\tau, \theta_i, \phi_j)$  and  $\mathcal{M}(\tau, \theta_i)$  are just the respective raw (unnormalized) histograms with the spike counts computed using the firing rate and normalized by the duration of the experiment.

Let us remark that, for simplicity and definiteness, we consider in what is to follow the continuous firing rate,  $m(t, \vec{\theta}, \vec{\phi})$ . In principle, this rate is computed in some appropriately short time interval around the time  $t$  as the ratio between the number of spikes the neuron in question fires in this interval and the length of this interval. However, we only use  $m(t, \vec{\theta}, \vec{\phi})$  for theoretical purposes, and the practical calculations of the probability distributions  $\mathcal{Q}(\tau, \theta_i, \phi_j)$  and  $\mathcal{P}(\tau, \theta_i)$  in an experiment should proceed as described in Ringach et al. (1997a, 2003) and at the beginning of this section.

Clearly, the continuous counterparts of the two probability distribution functions  $\mathcal{Q}(\tau, \theta_i, \phi_j)$  and  $\mathcal{P}(\tau, \theta_i)$  are given by the expressions

$$\mathcal{Q}_{\text{cont}}(\tau, \theta_i, \phi_j) = \frac{\mathcal{N}(\tau, \theta_i, \phi_j)}{\sum_{k=1}^I \sum_{l=1}^J \mathcal{N}(\tau, \theta_k, \phi_l)} \quad \text{and} \\ \mathcal{P}_{\text{cont}}(\tau, \theta_i) = \frac{\mathcal{M}(\tau, \theta_i)}{\sum_{k=1}^I \mathcal{M}(\tau, \theta_k)}, \tag{4}$$

respectively.

Two quantities closely related to the correlation functions  $\mathcal{N}(\tau, \theta_i, \phi_j)$  and  $\mathcal{M}(\tau, \theta_i)$ , given by Eq. (3), are the “interval-specific firing rate averages”

$$N(\tau, \theta_i, \phi_j) = \lim_{K \rightarrow \infty} \frac{1}{2K} \sum_{k=-K}^K m(n_{i,j}^k v + \tau, \vec{\theta}, \vec{\phi}), \tag{5a}$$

$$M(\tau, \theta_i) = \lim_{K \rightarrow \infty} \frac{1}{2K} \sum_{l=-K}^K m(n_l^i v + \tau, \vec{\theta}, \vec{\phi}), \tag{5b}$$

where the time  $t = n_{i,j}^k \nu$  is the beginning of the  $k$ -th time interval during which the grating with the orientation-phase pair  $(\theta_i, \phi_j)$  was flashed, and the time  $t = n_l^i \nu$  is the beginning of the  $l$ -th time interval during which the grating with the orientation  $\theta_i$  was flashed, respectively. In other words,  $N(\tau, \theta_i, \phi_j)$  is the average, over all the time intervals in which only the grating with the orientation  $\theta_i$  and spatial phase  $\phi_j$  was shown, of the firing rate  $\tau$  time units after the beginning of every such time interval. Likewise, the corresponding phase-averaged quantity, i.e.,  $M(\tau, \theta_i)$  is the analogous average over all time intervals in which only the gratings with the orientation  $\theta_i$  were flashed, regardless of their spatial phases.

Equations (3) and (5) imply

$$\begin{aligned} \mathcal{N}(\tau, \theta_i, \phi_j) &= \frac{1}{IJ\nu} \int_0^\nu N(s + \tau, \theta_i, \phi_j) ds \\ &= \frac{N(\tau, \theta_i, \phi_j)}{IJ} + \mathcal{O}(\nu), \end{aligned} \tag{6a}$$

$$\mathcal{M}(\tau, \theta_i) = \frac{1}{I\nu} \int_0^\nu M(s + \tau, \theta_i) ds = \frac{M(\tau, \theta_i)}{I} + \mathcal{O}(\nu), \tag{6b}$$

where, as we recall,  $I$  is the total number of different orientations and  $J$  is the total number of different spatial phases contained in the collection of the gratings presented in the experiment. This is because, on each interval  $n_{i,j}^k \nu < t < (n_{i,j}^k + 1)\nu$  in which the orientation-phase pair  $(\theta_i, \phi_j)$  is presented jointly, we have

$$\begin{aligned} &\int_{n_{i,j}^k \nu}^{(n_{i,j}^k + 1)\nu} \chi_{\theta_i, \phi_j}(t) m(t + \tau, \vec{\theta}, \vec{\phi}) dt \\ &= \int_0^\nu m(n_{i,j}^k \nu + s + \tau, \vec{\theta}, \vec{\phi}) ds. \end{aligned}$$

The difference in normalization comes about as follows. The correlation functions in Eq. (3) and on the left-hand side of Eq. (6) are normalized by the total duration of the stimulus. On the other hand, the interval-specific firing rate averages in Eq. (5) and on the right-hand side of Eq. (6) are normalized over the total number of times the grating with the specific orientation and phase  $(\theta_i, \phi_j)$ , or the specific phase  $\theta_i$  is presented. In the limit of an infinitely long stimulus, the two sets of normalizations differ by factors of  $IJ$  and  $I$ , respectively.

As we show in the next paragraph, the interval-specific firing rate average  $N(\tau, \theta_i, \phi_j)$  can be recast in the form

$$N(\tau, \theta_i, \phi_j) = \langle m \rangle_{\vec{\theta}, \vec{\phi}, \theta^{(0)} = \theta_i, \phi^{(0)} = \phi_j}(\tau, \theta_i, \phi_j), \tag{7a}$$

which is the firing rate at the time  $\tau$ , ensemble-averaged (as defined by an appropriate limit and denoted by  $\langle \cdot \rangle$ ) over all possible bi-infinite presentation sequences  $\vec{\theta}$  and  $\vec{\phi}$  of grating orientation angles  $\theta_k$  and spatial phases  $\phi_l$  such that the grating presented in the time interval  $[0, \nu]$  has the fixed orientation angle  $\theta^{(0)} = \theta_i$  and spatial phase  $\phi^{(0)} = \phi_j$ . Likewise, the interval-specific firing rate average  $M(\tau, \theta_i)$  can be recast in the form

$$M(\tau, \theta_i) = \langle m \rangle_{\vec{\theta}, \vec{\phi}, \theta^{(0)} = \theta_i}(\tau, \theta_i), \tag{7b}$$

which is the firing rate at the time  $\tau$ , ensemble-averaged over all possible bi-infinite presentation sequences  $\vec{\theta}$  and  $\vec{\phi}$  of angles  $\theta_k$  and phases  $\phi_l$  such that the grating presented in the time interval  $[0, \nu]$  has the fixed orientation angle  $\theta^{(0)} = \theta_i$  regardless of the spatial phase.

We here briefly outline the reason for the equality in Eq. (7b); the equality (7a) is true for almost the same reason. First, each term  $m(n_{i,j}^k \nu + \tau, \vec{\theta}, \vec{\phi})$  in Eq. (5b) can be time-shifted to become  $m(\tau, P^{n_i^k} \vec{\theta}, P^{n_i^k} \vec{\phi})$ , where  $P^{n_i^k} \vec{\theta}$  and  $P^{n_i^k} \vec{\phi}$  are the sequences  $\vec{\theta}$  and  $\vec{\phi}$  with all their indices decreased by  $n_i^k$ . In other words,  $P^{n_i^k} \vec{\theta}$  and  $P^{n_i^k} \vec{\phi}$  are the presentation sequences that maintain the exact order of  $\vec{\theta}$  and  $\vec{\phi}$ , but are shifted  $n_i^k \nu$  time units into the past. Invoking the ‘‘ergodicity’’ assumption, we treat  $(P^{n_i^k} \vec{\theta}, P^{n_i^k} \vec{\phi})$ ,  $k = \dots, -2, -1, 0, 1, 2, \dots$ , as independent pairs of sequences with the only common property of showing the angle  $\theta_i$  in the time interval  $0 \leq t < \nu$ . In this way, we transform the time averaging in Eq. (5b) into the ensemble averaging in Eq. (7b).

The quantities (7a) and (7b) can be used to represent the experimental data from a somewhat different perspective. In particular, the interval-specific firing rate average  $N(\tau, \theta_i, \phi_j)$  can be interpreted as the typical tuning response of a neuron in a continuously stimulated cortex to the presentation of a specific orientation and phase of duration  $\nu$ , embedded within a much longer, arbitrary stimulus. The interval-specific firing rate average  $M(\tau, \theta_i)$  represents the analogous response without regard to the grating’s spatial phase. In the forthcoming sections we will see why this particular point of view may be particularly fruitful in using RTC as a tool for studying orientation tuning in V1.

Finally, let us remark that, comparing Eqs. (3) and (5), we see that for finite presentation times  $\nu$ ,

the correlation functions  $\mathcal{N}(t, \theta_i, \phi_j)$  and  $\mathcal{M}(t, \theta_i)$  appear to give a better approximation to the actual spike count than the interval-specific firing rate averages  $N(t, \theta_i, \phi_j)$  and  $M(t, \theta_i)$ , respectively. This is because, as Eq. (3) shows,  $\mathcal{N}(t, \theta_i, \phi_j)$  and  $\mathcal{M}(t, \theta_i)$  sample the instantaneous firing rate continuously via an integral to produce the spike count, while  $N(t, \theta_i, \phi_j)$  and  $M(t, \theta_i)$  in Eq. (5) replace the integral by crude Riemann sums over intervals of width  $\nu$ , whose accuracy thus worsens with increasing  $\nu$ . This observation will be important in Section 5. However, in the next few sections, we will be interested in the limit of small  $\nu$ , in which case Eqs. (3) and (5) are equally accurate due to Eq. (6).

### 3 Binary noise characterization of the stimulus and the Volterra–Wiener theory

The Volterra–Wiener theory of nonlinear systems (Volterra 1930; Wiener 1958; Marmarelis and Marmarelis 1978; Schetzen 1980; Victor 1992) concerns the stimulus-response relation of a nonlinear system, such as a network of neurons in V1. In particular, this theory seeks an explicit representation of the system’s output response, say  $r(t)$ , in terms of its input stimulus, say  $s(t)$ , expressed as a Taylor-like functional series, referred to by the name Volterra series. This Volterra series has the form

$$r[s](t) = L_0 + \int_0^\infty L_1(\tau_1)s(t - \tau_1) d\tau_1 + \frac{1}{2!} \int_0^\infty \int_0^\infty L_2(\tau_1, \tau_2)s(t - \tau_1)s(t - \tau_2) d\tau_1 d\tau_2 + \dots + \frac{1}{n!} \int_0^\infty \dots \int_0^\infty L_n(\tau_1, \dots, \tau_n)s(t - \tau_1) \dots s(t - \tau_n) d\tau_1 \dots d\tau_n + \dots, \tag{8}$$

where the notation  $r[s](t)$  is used to emphasize the functional dependence of the response on the entire history of the particular input stimulus. The Volterra kernels,  $L_n(\tau_1, \dots, \tau_n)$ , give the appropriate weights to the contributions of all possible products of the stimulus values at all possible times in the past to the current response value. With no loss of generality, these kernels are assumed to be symmetric functions of their arguments. An orthogonalized version of the Volterra series, due to Wiener (1958), can be used to represent a broader set of systems, for which the Volterra series may not converge. The relationship between the two series, including the types of convergence, is analogous to that between the ordinary Taylor series and a series of orthogonal functions. Detailed expressions relating Volterra and Wiener series can be found in Schetzen (1980).

The first natural question that arises in the nonlinear systems theory is how to measure or compute the Volterra or Wiener kernels. For the latter, the answer is given by the surprisingly simple Lee-Schetzen theory: to find its Wiener kernels, the system must be “probed” by Gaussian white-noise stimuli (Lee and Schetzen 1965; Schetzen 1980). In what is to follow we only consider the first kernels, in which case the Wiener kernel can be computed from the formula

$$W_1(\tau_1) = \frac{1}{A^2} \langle r[s_0](t)s_0(t - \tau_1) \rangle_{s_0} = \lim_{T \rightarrow \infty} \frac{1}{2A^2T} \int_{-T}^T r[s_0](t)s_0(t - \tau_1) dt. \tag{9}$$

Here,  $s_0(t)$  is a zero-mean Gaussian white-noise stimulus,  $A^2$  its power, and  $\langle \cdot \rangle_{s_0}$  denotes the ensemble averaging over all realizations of  $s_0(t)$ . The last equality is true provided we assume ergodicity.

The task of connecting the quantities related to orientation tuning in V1 under the stimulus (1), used in the experiments of Ringach et al. (1997a, 2003) and discussed in the previous section, and the Volterra–Wiener theory of nonlinear systems is complicated by the fact that the experimental stimulus (1) depends on both space and time. The standard Lee–Schetzen method for kernel identification in such a case consists of probing the system with (approximate) spatio-temporal Gaussian white-noise stimuli. This method was used, for example, to measure the receptive-field shape and dynamics of LGN cells in Reid and Shapley (1992), Reid et al. (1997), Wolfe and Palmer (1998), Reid and Shapley (2002) and V1 cells in Jones and Palmer (1987), DeAngelis et al. (1993, 1995). However, as in experiments this method becomes increasingly complex with increasing resolution, the alternative stimulus (1) is used in Ringach et al. (1997a, 2003), which is not Gaussian white noise in space, but reduces the complexity of the subsequent data analysis. The use of (1) as the stimulus is motivated by the conjecture that the processing performed by a spiking LGN or V1 neuron consists of a linear filter followed by a static nonlinearity (Ringach et al. 1997b). The stimulus (1) essentially consists of drawing at each time  $n\nu$ , with uniform probability, an element of the spatially orthogonal set of gratings  $\sin[\mathbf{k}(\theta_i) \cdot \mathbf{X} - \phi_j]$  and their

negatives (provided both  $I$  and  $J$  are even). In this case, the results of Ringach et al. (1997b) show that RTC reproduces, up to a factor, the projection of the linear filter on this orthogonal set at the delay time  $\tau$ . If all the orientations  $\theta$  and phases  $\phi$ , as well as all the lengths of the vector  $\mathbf{k}(\theta)$  were allowed in the orthogonal set, this procedure would simply reconstitute the original linear filter via the Fourier transform, and might possibly be extended to a mixed time-domain/spatial-frequency-domain version of the Volterra–Wiener theory (cf. the works (Victor 1979; Victor and Knight 1979; Victor and Shapley 1980) on the Volterra–Wiener theory in the frequency domain). When probing a V1 neuron with a string of gratings whose lengths of  $\mathbf{k}(\theta)$  are tightly distributed around the neuron’s preferred spatial frequency, the procedure of Ringach et al. (1997b) provides for a much quicker (approximate) reconstruction of the linear filter, and thus the receptive field, than the classical procedure using spatio-temporal white noise stimuli. The results of Nykamp and Ringach (2002) extend the procedure of Ringach et al. (1997b) to include the identification of the static nonlinearity.

We mention a novel polynomial kernel regression approach to discrete Volterra–Wiener expansions (Franz and Schölkopf 2006) that may be relevant here. This method assumes that the input is a finite, discrete sample of signals. In this approach, the  $n$ -th order partial sums of the Volterra and Wiener series are computed directly via a regression procedure without first explicitly evaluating the corresponding kernels. In addition, one does not need to restrict the stimuli to Gaussian white noise to identify these partial sums. The approach of Franz and Schölkopf (2006) can easily be generalized to space-and-time dependent setups such as that addressed in Ringach et al. (1997a, b, 2003) and its higher-order Volterra–Wiener counterparts. In particular, a sufficiently large set of randomly-flashed gratings (1) could be used as stimuli in conjunction with the procedure of Franz and Schölkopf (2006) to compute a discrete Volterra–Wiener approximation to a neuronal receptive field to a high degree of precision without the costly cross-correlation with spatio-temporal white-noise stimuli.

In this paper, we address the theoretical issue of how to connect the interval-specific firing rate averages  $N(t, \theta_i, \phi_j)$  and  $M(t, \theta_i)$  in Eq. (5)—or equivalently, the correlation functions  $\mathcal{N}(t, \theta_i, \phi_j)$  and  $\mathcal{M}(t, \theta_i)$  in Eq. (3)—motivated by the experimental procedure of Ringach et al. (1997a, 2003), to appropriate Volterra–Wiener kernels associated with continuous, Gaussian white noise, temporal inputs (in Appendix 2, we discuss the errors in the Wiener kernel construction introduced by the binary white noise discretization

of the continuous-time Gaussian white noise). First, we must determine the mathematical characterization of the random experimental stimulus (1) in the limit as  $\nu \rightarrow 0$ . Since, together with (Ringach et al. 1997a, 2003), we are not interested in the spatial shape of the neuronal receptive field but only in the orientation (and spatial-phase) tuning dynamics, and since the stimulus (1) only has noise-like characteristics in time, we focus on this aspect of Eq. (1). We first observe that both time series,  $(\theta(t), \phi(t))$  and  $\theta(t)$  in Eq. (3), are discrete noise-like processes. However, in order to find the corresponding Wiener kernels, one would have to correlate the firing rate against  $\theta(t)$  and/or  $\phi(t)$ , which is not done in Ringach et al. (1997a, 2003), and so also not in Eqs. (5) and (3). Alternatively, to analyze the procedure of Ringach et al. (1997a, 2003), we observe that the process of the orientation-phase pair  $(\theta_i, \phi_j)$  being presented versus not being presented is a binary noise process, whose value can switch every time  $n\nu$ . After proper scaling and normalization, let us denote this process by  $s_{vij}(t)$ ,  $i = 1, \dots, I$ ,  $j = 1, \dots, J$  (its precise properties are derived and described in Appendix 3). Now, since precisely one grating, i.e., a particular orientation-phase pair  $(\theta_i, \phi_j)$ , is presented at any given time, the noises  $s_{vij}(t)$ ,  $i = 1, \dots, I$ ,  $j = 1, \dots, J$ , must necessarily be correlated. In fact, we show in Appendix 3 that we can take their sum over  $i$  and  $j$  to vanish. In the limit as  $\nu \rightarrow 0$ , we show that the noises  $s_{vij}(t)$  tend, in the statistical sense, to a collection of  $IJ$  correlated Gaussian white noises  $s_{ij}(t)$ . We emphasize that, while these noises are correlated with each other, each  $s_{ij}(t)$ , with fixed  $i$  and  $j$ , is a (temporal) Gaussian white noise process, since it has no temporal autocorrelations, as shown in Appendix 3. We emphasize that the Volterra–Wiener kernels we obtain in this way are not the Volterra–Wiener kernels corresponding to the Gaussian white noise in time and space.

With the aid of the correlated noise sets

$$\{s_{vij}(t) \mid i = 1, \dots, I, j = 1, \dots, J\} \quad \text{and} \\ \{s_{ij}(t) \mid i = 1, \dots, I, j = 1, \dots, J\},$$

we can extend the existing Volterra–Wiener approach to cover the case of correlated Gaussian white noises and use this extension for calculating the corresponding first-order Wiener kernels in terms of the correlation functions (3) and interval-specific firing rate averages (5) defined in the previous section. This is done in Appendix 3, and the results are presented below. A brief comparison of our results with those of Ringach



et al. (1997b), Nykamp and Ringach (2002) is also presented there. In Appendix 2, we discuss the error committed in the determination of the Wiener kernels by the time-discretization we employ, that is, by using a binary noise process whose steps are taken every  $\nu$

milliseconds. As expected, we show that this error is of order  $\mathcal{O}(\nu^{1/2})$ .

We write the firing rate  $m(t, \Theta, \Phi)$  of a given V1 neuron with the preferred orientation  $\Theta$  and preferred spatial phase  $\Phi$  as the Volterra series in the form

$$\begin{aligned}
 m[\vec{s}] (t, \Theta, \Phi) &= L_0(\Theta, \Phi) + \sum_{i=1}^I \sum_{j=1}^J \int_0^\infty L_1^{i,j}(\tau_1, \Theta, \Phi) S_{ij}(t - \tau_1) d\tau_1 \\
 &+ \frac{1}{2!} \sum_{i,k=1}^I \sum_{j,l=1}^J \int_0^\infty \int_0^\infty L_2^{ik,jl}(\tau_1, \tau_2, \Theta, \Phi) S_{ij}(t - \tau_1) S_{kl}(t - \tau_2) d\tau_1 d\tau_2 + \dots,
 \end{aligned}
 \tag{10}$$

where  $S_{ij}(t)$  can be any of the binary noise stimuli  $s_{vij}(t)$ , or, in the limit as  $\nu \rightarrow 0$ , the correlated Gaussian white noise stimulus  $s_{ij}(t)$ . In order for Eq. (10) to be a true Volterra expansion, the corresponding kernels must be independent of the particular choice of the stimulus  $s_{vij}(t)$ , that is, of the presentation time  $\nu$ . As usual, the Wiener series associated with this Volterra expansion is obtained via an orthogonalization involving the correlated Gaussian white noise stimuli  $s_{ij}(t)$ , and approximates the firing rate in the least-squares sense. This orthogonalization is presented in Appendix 3.

In Appendix 3, we extend the Lee–Schetzen formula (9) to our correlated-noise case. This extension gives the first-order Wiener kernels as

$$\begin{aligned}
 W_1^{i,j}(\tau, \Theta, \Phi) &= IJ \langle m[\vec{s}] (t, \Theta, \Phi) s_{ij}(t - \tau) \rangle_{\vec{s}}
 \end{aligned}
 \tag{11a}$$

$$= IJ \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T m[\vec{s}] (t, \Theta, \Phi) s_{ij}(t - \tau) dt.
 \tag{11b}$$

Here, in analogy with Eq. (9),  $\langle \cdot \rangle_{\vec{s}}$  denotes ensemble averaging over all possible realizations of the random noise collection  $\vec{s}$ . The second equality holds for any given noise collection  $\vec{s}$  with the  $ij$ -th component  $s_{ij}$ , and is true because of ergodicity.

We now discuss the relationship between the Volterra–Wiener representations of the firing rate on the one hand and the correlation functions (3) and the “interval-specific firing rate averages” (7) on the other. Since the expressions (3) and (7) involve the stimulus in a linear fashion, we are indeed allowed to focus on the first-order, that is linear, kernels, given by

formula (11a). After some algebra, presented in Appendix 3, we find

$$\begin{aligned}
 W_1^{i,j}(\tau, \Theta, \Phi) &= \lim_{\nu \rightarrow 0} \frac{IJ}{\sqrt{\nu}} [\mathcal{N}(\tau, \theta_i, \phi_j) - \langle \mathcal{N} \rangle_{\tau, \theta^{(0)}, \phi^{(0)}}] \\
 &= \lim_{\nu \rightarrow 0} \frac{1}{\sqrt{\nu}} [N(\tau, \theta_i, \phi_j) - \langle N \rangle_{\tau, \theta^{(0)}, \phi^{(0)}}],
 \end{aligned}
 \tag{12}$$

where  $\mathcal{N}(\tau, \theta_i, \phi_j)$  is defined in Eq. (3a) and  $N(\tau, \theta_i, \phi_j)$  in Eq. (5a) or (7a), and  $\langle \cdot \rangle_{\tau, \theta^{(0)}, \phi^{(0)}}$  denotes averaging over all the arguments in the subscript.

Note that, since from Eq. (12),

$$\begin{aligned}
 N(\tau, \theta_i, \phi_j) &= \langle N \rangle_{\tau, \theta^{(0)}, \phi^{(0)}} + \nu^{1/2} W_1^{i,j}(\tau, \Theta, \Phi) + \mathcal{O}(\nu), \\
 &\text{as } \nu \rightarrow 0,
 \end{aligned}$$

the term containing the first Wiener kernel,  $\nu^{1/2} W_1^{i,j}(\tau, \Theta, \Phi)$ , is small for short presentation duration  $\nu$ , and so it may not be easy to extract the first kernel from the measurements in practice. In particular, it may also be overshadowed by noise, as will be discussed below.

The experimental results of Ringach et al. (1997a, 2003) take into account the fact that minute eye movements of the experimental animal blur the spatial phase of the stimulus, or equivalently, the preferred spatial phase of the V1 neuron in question. These results can therefore be interpreted mathematically in two ways. One is as the average of measurements performed on an ensemble of neurons which are all identical except in their preferred spatial phases. The other is as the ensemble average over all the stimulus phases  $\phi_j$  for a particular V1 neuron, which is what is done in the experiment by averaging the histograms on the phases  $\phi_j$ . Thus, from the definitions (3), and in accordance with the second interpretation that in the experiment the phases  $\phi_j$  are averaged out, we see that

the correlation functions  $\mathcal{N}(\tau, \theta_i, \phi_j)$  and  $\mathcal{M}(\tau, \theta_i)$  are simply related by the equation

$$\mathcal{M}(\tau, \theta_i) = \sum_{j=1}^J \mathcal{N}(\tau, \theta_i, \phi_j).$$

Moreover, the first Wiener kernel corresponding to  $\mathcal{M}(\tau, \theta_i)$  is

$$\begin{aligned} W_1^i(\tau, \Theta) &= \frac{1}{J} \sum_{j=1}^J W_1^{i,j}(\tau, \Theta, \Phi) \\ &= \lim_{\nu \rightarrow 0} \frac{I}{\sqrt{\nu}} [\mathcal{M}(\tau, \theta_i) - \langle \mathcal{M} \rangle_{\tau, \theta^{(0)}}] \\ &= \lim_{\nu \rightarrow 0} \frac{1}{\sqrt{\nu}} [M(\tau, \theta_i) - \langle M \rangle_{\tau, \theta^{(0)}}], \end{aligned} \tag{13}$$

where  $\mathcal{M}(\tau, \theta_i)$  is the correlation function (3b) and  $M(\tau, \theta_i)$  is the interval-specific firing rate average (5b).

In the next section, we use the results obtained here to analyze the relationships between the correlation functions (3) and first-order Wiener kernels in a series of neuronal models of increasing complexity. We should note already here, however, the crucial difference in the ways in which the first-order Wiener kernels  $W_1^{i,j}(\tau, \Theta, \Phi)$  and  $W_1^i(\tau, \Theta)$  and, say, the interval-specific firing rate averages  $N(\tau, \theta_i, \phi_j)$  and  $M(\tau, \theta_i)$ , are defined. Namely, each  $W_1^{i,j}(\tau, \Theta, \Phi)$  or  $W_1^i(\tau, \Theta)$  is defined independently of any particular presentation time  $\nu$ . On the other hand, as seen from Eq. (5) or (7),  $N(\tau, \theta_i, \phi_j)$  and  $M(\tau, \theta_i)$  are defined as ensemble averages over certain presentations of a particular binary noise  $s_{vij}$ , and thus are functions of  $\nu$ .

#### 4 Orientation tuning and Wiener kernels in idealized models

In this section we illustrate the concepts described in the previous two sections by computing the interval-specific firing rate averages  $N(t, \theta^{(0)}, \phi^{(0)})$  and  $M(t, \theta^{(0)})$  in Eq. (5), and the corresponding first-order Wiener kernels, for a leaky feed-forward model.

The model discussed in this section, as well as the model discussed in Section 5, are simplifications of the thresholded-linear, mean-field, simple-cell model (Shelley and McLaughlin 2002),

$$\begin{aligned} m_P(\mathbf{x}, t, \Phi) &= \left\{ f_{lgn}(t, \Theta, \Phi) \right. \\ &\quad \left. + \sum_{P'} C_{PP'} K_{PP'} * G_{P'} * \langle m_{P'} \rangle_{\Phi}(\mathbf{x}, t) \right\}^+, \\ P, P' &= E, I, \end{aligned} \tag{14}$$

with

$$f_{lgn}(t, \Theta, \Phi) = -g_L + \left( \frac{V_E}{V_T} - 1 \right) g_{lgn}(t, \Theta, \Phi) \tag{15}$$

and

$$C_{PP'} = S_{PP'} \left( \frac{V_{P'}}{V_T} - 1 \right). \tag{16}$$

Here,  $\Theta$  and  $\Phi$  denote the neuron’s preferred orientation angle and phase, respectively,  $\mathbf{x}$  is the position of the neuron in the cortex which also determines the neuron’s preferred orientation,  $\Theta$  (see below),  $E$  and  $I$  stand for excitatory and inhibitory,  $m_P(\mathbf{x}, t, \Phi)$ , with  $P = E, I$ , are the respective firing rates,  $V_E, V_I$  and  $V_T$  are the (non-dimensionalized) excitatory and inhibitory reversal and threshold potentials, respectively,  $g_L$  is the leakage conductance,  $g_{lgn}(t, \Theta, \Phi)$  is the LGN drive (to be described below),  $K_{PP'}(\mathbf{x})$  and  $G_P(t)$  are the spatial and temporal cortical kernels, respectively, the asterisk denotes convolution, the symbol  $\langle \cdot \rangle_{\Phi}$  denotes averaging over the preferred spatial phase, and  $\{\cdot\}^+ = \max\{\cdot, 0\}$  denotes “thresholding.” (In all our numerical computations we use  $V_E = 14/3$ ,  $V_I = -2/3$  and  $V_T = 1$ .) Recall that convergent feed-forward input from several LGN neurons establishes the orientation preference of a given V1 neuron (Reid and Alonso 1995). Across V1, this preference is laid out in pinwheel patterns, each pinwheel having an orientation preference singularity at its center (Bonhoeffer and Grinvald 1991; Blasdel 1992a, b; Maldonado et al. 1997). In the local polar coordinates on any given orientation pinwheel in V1,  $\mathbf{x} = r(\cos(2\Theta + \Delta), \sin(2\Theta + \Delta))$ , that is, the angle of the neuron position  $\mathbf{x}$  around the pinwheel center corresponds to twice its preferred orientation angle on the screen. Within the “pinwheel lattice” that tessellates V1, we choose  $\Delta = 0$  and  $\Delta = \pi/2$ , respectively, in any pair of neighboring pinwheels, to provide for a smooth transition in the orientation preference from one pinwheel to the next. The constants  $S_{PP'}$  are the cortical coupling constants in the model. The form of the factors  $V_P/V_T - 1$  in the LGN drive  $f_{lgn}(t, \Theta, \Phi)$  in Eq. (15) and in the coupling coefficients  $C_{PP'}$  in Eq. (16) stems from the derivation of the Eq. (14) from a logarithmic mean-field model in the linear limit. Details are given in Shelley and McLaughlin (2002).

The most drastic simplification of the model (14), which we will use in some of the discussion to follow, is the leaky feed-forward model, obtained by setting all the cortical coupling constants  $S_{PP'} = 0$ . In this case, there is no more distinction between the forms of the equations for the inhibitory and excitatory firing rates, so we suppress the indices  $E$  and  $I$  (or  $P$  and  $P'$ ),

respectively. The thresholded linear model (14) now becomes

$$m(t, \Theta, \Phi) = \left\{ -g_L + \left( \frac{V_E}{V_T} - 1 \right) g_{lgn}(t, \Theta, \Phi) \right\}^+, \quad (17)$$

with the firing rate depending on the position  $\mathbf{x}$  in the pinwheel only through the neuron’s preferred orientation  $\Theta$ , and also depending on the neuron’s preferred phase  $\Phi$ . A yet more special case is represented by the purely feed-forward model with no leakage, i.e.,  $g_L = 0$ . In this case, since the LGN drive must be non-negative, thresholding can be dropped in Eq. (17).

It will be also advantageous to consider a modification of the leaky feed-forward model (17) with added uniform global cortical inhibition.

$$m(t, \Theta, \Phi) = \left\{ -g_L + \left( \frac{V_E}{V_T} - 1 \right) g_{lgn}(t, \Theta, \Phi) - C_{PI} g_I \right\}^+. \quad (18)$$

Here,  $g_I$  is the strength of the inhibition and  $C_{PI}$  is the strength of the coupling. Such global cortical inhibition

$$g_{lgn}(t) = R_B + \frac{1}{\pi \lambda \mu} \int_{\mathbb{R}^2} dX dY \int_{-\infty}^t ds e^{-\left(\frac{X^2}{\lambda^2} + \frac{Y^2}{\mu^2}\right)} \sin(\mathbf{K} \cdot \mathbf{X} - \Phi) G_{lgn}(t-s) \mathcal{I}(\mathbf{X}, s), \quad (19)$$

where  $\mathbf{X} = (X, Y)$  is the position on the screen,  $R_B$  is the background LGN firing rate,  $\mathcal{I}(\mathbf{X}, t)$  is the randomly flashed stimulus (1),  $\mathbf{K} = K(-\sin \Theta, \cos \Theta) = K(0, 1)$ , with  $K/2\pi$  being the preferred spatial frequency, and  $\Phi$  is both the preferred phase of the neuron and also the phase that determines the shape of the on- and off-regions in the neuron’s receptive field. In the Gaussian kernel,  $\lambda$  and  $\mu$  are the semi-axes of the neuron’s elliptically-shaped receptive field. Customarily,  $\lambda > \mu$ , signifying that the receptive field is partitioned into strip-shaped on- and off- regions, parallel to the major axis of the ellipse. Also,  $G_{lgn}(t)$  is the LGN time-response kernel, which we choose to be (Benardete and Kaplan 1999; Wielaard et al. 2001)

$$G_{lgn}(t) = \begin{cases} 0, & t \leq 0, \\ \left( \frac{t}{\tau_{lgn}} \right)^\zeta \left( \alpha^{\zeta+1} e^{-\alpha \frac{t}{\tau_{lgn}}} - \beta^{\zeta+1} e^{-\beta \frac{t}{\tau_{lgn}}} \right), & t > 0, \end{cases} \quad (20)$$

with integer  $\zeta$  and  $\alpha > \beta$ . This kernel produces a response that has a finite rise time after  $t = 0$ . Note that

$$\int_{-\infty}^{\infty} G_{lgn}(t) dt = 0, \quad (21)$$

in agreement with Gielen et al. (1981), Bernadete and Kaplan (1999).

may arise, for example, near a pinwheel center, due to the fact that the spatial convolution on the right-hand side of Eq. (14) becomes equal to the averaging over the preferred angle there (Shelley and McLaughlin 2002). Rather than pursuing the solution of Eq. (14) in this case, we simplify the problem by assuming  $g_I$  to be approximately constant, which is sufficient for the purposes of illustration aimed at in this section.

We use the simplest possible filter to model the effective LGN input, a Gabor function (Marcelja 1980; Daugman 1985). We should note that this choice of the LGN model lacks any description of the LGN rectification and saturation; a more realistic LGN model that includes rectification and saturation is discussed in Appendix 5 (a general LGN model, which includes both rectification and saturation, can be found, for example, in Tao et al. (2004)). For simplicity and with no loss of generality, we restrict the display of the Gabor function to the case in which  $\Theta = 0$  is the preferred orientation of the V1 neuron in question. The general case can be described by using an appropriate rotation. The LGN input into the V1 neuron is given by

After some straightforward algebra, using Eq. (21), we compute the LGN input (19) for a fixed presentation  $\bar{\theta}$  and  $\bar{\phi}$  of the stimulus (1) to be the series

$$g_{lgn}(t) = R_B + \frac{\mathcal{I}_0 \varepsilon v^{-1/2}}{2} e^{-\frac{1}{4} K^2 \mu^2} \times \sum_{n=-\infty}^{\infty} f(\theta^{(n)}) \mathcal{G}(t-nv) \cos[\phi^{(n)} - \Delta(\theta^{(n)})], \quad (22)$$

where,

$$f(\theta^{(n)}) = e^{-\frac{1}{4} [k^2 \lambda^2 \sin^2 \theta^{(n)} + k^2 \mu^2 \cos^2 \theta^{(n)}]} \times \sqrt{2 [\cosh(\mu^2 K k \cos \theta^{(n)}) - \cos 2\Phi]}, \quad (23)$$

and  $\Delta(\theta^{(n)})$  is the phase given by the formula

$$\tan \Delta(\theta^{(n)}) = \tan \Phi \coth \left( \frac{1}{2} \mu^2 K k \cos \theta^{(n)} \right). \quad (24)$$

Moreover,

$$\mathcal{G}(t) = \int_{-\infty}^t G_{lgn}(t-s) \chi_{[0,v]}(s) ds = \mathcal{F}(t) - \mathcal{F}(t-v), \quad (25)$$

$$\mathcal{F}(t) = \int_{-\infty}^t G_{lgn}(s) ds,$$

and  $\chi_{[0,v]}(t)$  is the indicator function of the interval  $[0, v]$ . We emphasize here that, in order to keep the

LGN drive (22) positive, we should also scale the background firing rate  $R_B$  in Eq. (22) as  $\nu^{-1/2}$ . We will return to this issue below, but let us here remark that this scaling is an artificial device to keep our model valid in the  $\nu \rightarrow 0$  limit rather than a reflection of the experimentally measured background firing rate, which is independent of  $\nu$ .

For the leaky-feedforward model (17), the interval-specific firing rate averages,  $N(t, \theta^{(0)}, \phi^{(0)})$  and  $M(t, \theta^{(0)})$ , as well as the corresponding linear Volterra–Wiener kernels, can be computed explicitly. To do this, we treat the LGN input (22) as a sum of random variables, which are themselves functions of the random orientation angles  $\theta^{(n)}$  and spatial phases  $\phi^{(n)}$ . To compute  $N(t, \theta^{(0)}, \phi^{(0)})$ , we take all  $\theta^{(n)}$  and  $\phi^{(n)}$  to be random, except  $\theta^{(0)}$  and  $\phi^{(0)}$  presented in the time slot  $[0, \nu]$ , which are fixed. Likewise, to compute  $M(t, \theta^{(0)})$ , we take all the spatial phases  $\phi^{(n)}$  to be random, as well as all the orientation angles  $\theta^{(n)}$  except  $\theta^{(0)}$ , which is again fixed.

When both the orientation  $\theta^{(0)}$  and the spatial phase  $\phi^{(0)}$  are fixed, the random part of  $g_{lgn}(t)$  in Eq. (22) is the composite random variable

$$R = \nu^{-1/2} \sum_{\substack{n=-\infty \\ n \neq 0}}^{\infty} f(\theta^{(n)}) \mathcal{G}(t - n\nu) \cos \phi^{(n)}, \quad (26)$$

in which all the phases  $\phi^{(n)}$  and all the angles  $\theta^{(n)}$  are random (note that we have omitted the phase  $\Delta(\theta^{(n)})$ , since we will be averaging over the spatial phases  $\phi^{(n)}$ , and adding an additional phase does not change the outcome of this averaging due to the Lemma presented in Appendix 1). From the probability density function,  $p_R(r, t)$ , of the variable  $R$  in Eq. (26), which will be computed below, via formulas (7a), (17), and (22), we can compute the interval-specific firing rate average (5a) as

$$N(t, \theta^{(0)}, \phi^{(0)}) = \int_{\frac{c}{D} - \nu^{1/2} \tilde{\Psi}(t, \theta^{(0)}, \phi^{(0)})}^{\infty} [-C + \nu^{1/2} D \tilde{\Psi}(t, \theta^{(0)}, \phi^{(0)}) + Dr] p_R(r, t) dr, \quad (27)$$

where

$$C = -\left(\frac{V_E}{V_T} - 1\right) R_B + g_L, \quad D = \frac{\mathcal{I}_0 \varepsilon}{2} \left(\frac{V_E}{V_T} - 1\right) e^{-\frac{1}{4} K^2 \mu^2}, \quad (28)$$

and

$$\tilde{\Psi}(t, \theta^{(0)}, \phi^{(0)}) = \nu^{-1} f(\theta^{(0)}) \mathcal{G}(t) \cos[\phi^{(0)} - \Delta(\theta^{(0)})], \quad (29)$$

with  $f(\theta^{(0)})$ ,  $\mathcal{G}(t)$  and  $\Delta(\theta^{(0)})$  given in Eqs. (23), (25), and (24), respectively. Note that the expression  $\tilde{\Psi}(t, \theta^{(0)}, \phi^{(0)})$  in Eqs. (27) and (29) is deterministic because both the orientation angle  $\theta^{(0)}$  and the spatial phase  $\phi^{(0)}$  in the presentation time slot  $[0, \nu]$  are assumed to be fixed. Note also that in Eq. (27), the part of the integrand in the square brackets is non-negative. In fact, because of the thresholding in the model (17), the lower limit of the integral in Eq. (27) is chosen to be precisely where the expression in the square brackets crosses from negative to positive. The same remark will hold for the analogous formula (38) below. Finally, note that  $\tilde{\Psi}(t, \theta^{(0)}, \phi^{(0)}) = \Psi(t, \theta^{(0)}, \phi^{(0)}) + \mathcal{O}(\nu)$ , where

$$\Psi(t, \theta^{(0)}, \phi^{(0)}) = f(\theta^{(0)}) G_{lgn}(t) \cos[\phi^{(0)} - \Delta(\theta^{(0)})]. \quad (30)$$

This is because  $\mathcal{G}(t) = \nu G_{lgn}(t) + \mathcal{O}(\nu^2)$  by Eq. (25).

For small presentation times  $\nu$ , using Eq. (26) and asymptotics, we find for the probability density function  $p_R(r, t)$  of the random variable  $R$  in Eq. (26) the expression

$$p_R(r, t) = \frac{1}{\sqrt{2\pi} V_0} e^{-\frac{r^2}{2V_0}} + \mathcal{O}(\nu), \quad (31)$$

with

$$V_0 = \frac{f_2}{2} \int_0^\infty G_{lgn}^2(\xi) d\xi, \quad f_2 = \frac{1}{\pi} \int_0^\pi f^2(\theta) d\theta. \quad (32)$$

Of course, the  $\mathcal{O}(\nu)$  term in Eq. (31) is integrable and retains its order after integration. The details of the calculation leading to formula (31) are similar to, but simpler than, those given in Appendix 4 that lead to formula (39). From formula (27) we thus compute, as  $\nu \rightarrow 0$ ,

$$\begin{aligned} N(t, \theta^{(0)}, \phi^{(0)}) &= -\frac{C}{2} \operatorname{erfc}\left(\frac{C}{D\sqrt{2V_0}}\right) + D\sqrt{\frac{V_0}{2\pi}} e^{-C^2/2V_0 D^2} \\ &\quad + \frac{D\nu^{1/2}\Psi(t, \theta^{(0)}, \phi^{(0)})}{2} \operatorname{erfc}\left(\frac{C}{D\sqrt{2V_0}}\right) + \mathcal{O}(\nu), \end{aligned} \quad (33)$$

where “erfc” denotes the complementary error function,

$$\operatorname{erfc}(z) = \frac{2}{\sqrt{\pi}} \int_z^\infty e^{-u^2} du,$$

and  $\Psi(t, \theta^{(0)}, \phi^{(0)})$  is defined in formula (30). This calculation can be carried out provided  $\zeta \geq 1$  in the LGN temporal kernel (20).



We note that:

1. For a purely leaky feed-forward model with no global inhibition, the constant  $C$  in formulas (33) and (28) must be scaled as  $C = -Av^{-1/2} - B$ , that is, we must have  $R_B = \mathcal{O}(v^{-1/2})$ , so that the LGN input (22) does not become negative, as emphasized after Eq. (25). In this case, formula (33) simplifies to

$$N(t, \theta^{(0)}, \phi^{(0)}) = Av^{-1/2} + B + Dv^{1/2}\Psi(t, \theta^{(0)}, \phi^{(0)}) + \mathcal{O}(v). \tag{34}$$

2. The constant  $C$  in formula (33) may remain of order  $\mathcal{O}(1)$  if the neuron is strongly inhibited by uniform global inhibition, such as described by the modified model (18), in which case

$$C = -\left(\frac{V_E}{V_T} - 1\right) R_B + g_L + C_{PIGI}, \tag{35}$$

where  $C_{PI}$  is some appropriate cortical coupling strength. If  $C_{PIGI}$  is the  $\mathcal{O}(v^{-1/2})$ -size uniform global cortical inhibition, this global inhibition may compensate the  $\mathcal{O}(v^{-1/2})$ -size background firing rate  $R_B$  so that  $C = \mathcal{O}(1)$ . As mentioned before, such global cortical inhibition may arise, for example, near a pinwheel center, induced by the  $\mathcal{O}(v^{-1/2})$ -size LGN drive and/or strong coupling  $C_{PI}$ .

From Eqs. (33) and (34) we note that, to  $\mathcal{O}(v^{1/2})$ , the only tuned part of the interval-specific firing rate average  $N(t, \theta^{(0)}, \phi^{(0)})$  is the term containing the expression  $\Psi(t, \theta^{(0)}, \phi^{(0)})$ . Formulae (33) and (34) clearly show that term becomes untuned as  $v \rightarrow 0$ . In this limit, the convergence rate of  $N(t, \theta^{(0)}, \phi^{(0)})$  to the untuned response is  $v^{1/2}$ . The signal-to-noise ratio scales as  $v$  for the purely leaky feed-forward case, and  $v^{1/2}$  for the strongly, globally inhibited case. We here remark that these scaling results are only valid in the low-contrast regime, since the Gabor representation (19) of the LGN output is only appropriate in this regime. This is because in the high-contrast regime, thresholding and saturation dominate the LGN processing and, as mentioned before, these effects are not included in our LGN model.

Finally, from Eqs. (12) and (33), for the leaky feed-forward case, we compute the first-order Wiener kernels for this model to be

$$W_1^{i,j}(t, \Theta, \Phi) = Df(\theta_i)G_{lgn}(t) \cos[\phi_j - \Delta(\theta_i)]. \tag{36}$$

(In contrast to the previous discussion, we here use the subscripts in the orientation angle  $\theta_i$  and phase  $\phi_j$  because we emphasize their specific values, which determine the choice of the kernel  $W_1^{i,j}(t, \Theta, \Phi)$ .) From Eq. (36), it follows that the first Wiener kernel for this model essentially reproduces the  $n = 0$  term in the LGN drive (22), with  $\theta^{(0)} = \theta_i$  and  $\phi^{(0)} = \phi_j$ , in the limit as  $v \rightarrow 0$ . This is because, in this limit,  $\mathcal{G}(t) \approx vG_{lgn}(t)$ , as can be seen from formula (25) [an interpretation of  $W_1^{i,j}(t, \Theta, \Phi)$  as proportional to the projection of the LGN filter in Eq. (19) on the particular grating in (1) with orientation  $\theta_i$  and spatial phase  $\phi_j$  is given at the end of Appendix 3]. In the strongly, globally inhibited case, the constant  $D$  in Eq. (36) must be replaced by  $D \operatorname{erfc}(C/D\sqrt{2V_0})/2$ .

Likewise, when only the orientation  $\theta^{(0)}$  is fixed, the random part of  $g_{lgn}(t)$  in Eq. (22) is the composite random variable

$$S = v^{-1/2} \sum_{n=-\infty}^{\infty} f(\theta^{(n)}) \mathcal{G}(t - nv) \cos \phi^{(n)}, \tag{37}$$

in which all the phases  $\phi^{(n)}$  and all the angles  $\theta^{(n)}$  save one,  $\theta^{(0)}$ , are random. From the probability density function,  $p_S(s, t, \theta^{(0)})$ , for the random variable  $S$ , again to be computed below, via formulae (7b), (17), (22), and (37), we can compute the interval-specific firing rate average (5b) as

$$M(t, \theta^{(0)}) = \int_{C/D}^{\infty} [-C + Ds] p_S(s, t, \theta^{(0)}) ds. \tag{38}$$

For small presentation times  $v$ , we compute the probability density function  $p_S(s, t, \theta^{(0)})$  of the random variable  $S$  in Eq. (37) as

$$p_S(s, t, \theta^{(0)}) = \frac{1}{\sqrt{2\pi V(t, \theta^{(0)})}} \left( e^{-\frac{s^2}{2V(t, \theta^{(0)})}} + vV_1 \frac{\partial^4}{\partial s^4} e^{-\frac{s^2}{2V(t, \theta^{(0)})}} \right) + \mathcal{O}(v^2), \tag{39}$$

where

$$V(t, \theta^{(0)}) = V_0 + v \frac{[f^2(\theta^{(0)}) - f_2] G_{lgn}^2(t)}{2}, \tag{40}$$

$V_0$  and  $f_2$  are defined as in Eq. (32), and

$$V_1 = \frac{f_4 - 2f_2^2}{64} \int_0^{\infty} G_{lgn}^4(\xi) d\xi, \quad f_4 = \frac{1}{\pi} \int_0^{\pi} f^4(\theta) d\theta. \tag{41}$$

The details of the derivation leading to Eq. (39) are given in Appendix 4. The interval-specific firing rate average  $M(t, \theta^{(0)})$  is then calculated from Eq. (38) to be

$$\begin{aligned}
 M(t, \theta^{(0)}) = & -\frac{C}{2} \operatorname{erfc}\left(\frac{C}{D\sqrt{2V_0}}\right) + D\sqrt{\frac{V_0}{2\pi}} e^{-C^2/2V_0D^2} \\
 & + \nu \frac{D[f^2(\theta^{(0)}) - f_2] G_{lgn}^2(t)}{4\sqrt{2\pi} V_0} e^{-C^2/2V_0D^2} \\
 & + \nu V_1 \sqrt{\frac{2V_0}{\pi}} D \left(\frac{C^2}{2V_0D^2} - 1\right) e^{-C^2/2V_0D^2} + \mathcal{O}(\nu^2).
 \end{aligned}
 \tag{42}$$

Again, we note that:

1. For the case of sufficient global inhibition, as explained above, the constant  $C$  in Eq. (42) may remain of  $\mathcal{O}(1)$ . The interval-specific firing rate average  $M(t, \theta^{(0)})$  in Eq. (42) becomes untuned as  $\nu \rightarrow 0$  just as  $N(t, \theta^{(0)}, \phi^{(0)})$ , except faster. In particular, for finite  $C$ , the convergence rate of  $M(t, \theta^{(0)})$  to the untuned response in Eq. (42) is  $\nu$ , and the signal-to-noise ratio scales as  $\nu$  for this strongly, globally inhibited case.
2. For a purely leaky feed-forward model, in order for the LGN input to remain positive, we must set  $C = -A\nu^{-1/2} - B$ . Together with Eq. (42), this implies

$$M(t, \theta^{(0)}) = A\nu^{-1/2} + B + \mathcal{O}(\nu^2). \tag{43}$$

Equation (43) shows that the yet more subtle tuning in the pure leaky feed-forward case, when  $C = -A\nu^{-1/2} - B$ , cannot be captured by the present calculation, and only occurs at the order  $\mathcal{O}(\nu^2)$  or later.

From Eqs. (36), (42), and (13), namely,

$$M(\tau, \theta_i) = \langle M \rangle_{\tau, \theta^{(0)}} + \nu^{1/2} \frac{1}{J} \sum_{j=1}^J W_1^{i,j}(\tau, \Theta, \Phi) + \mathcal{O}(\nu),$$

(with  $\langle \cdot \rangle_{\tau, \theta^{(0)}}$  denoting averaging on the indicated arguments) since there is no  $\mathcal{O}(\nu^{1/2})$  contribution in Eq. (42), we find for the corresponding Wiener kernel  $W_1^i(\tau, \Theta)$  in Eq. (13) that

$$W_1^i(\tau, \Theta) = \frac{1}{J} \sum_{j=1}^J W_1^{i,j}(\tau, \Theta, \Phi) = 0. \tag{44}$$

In other words, the first-order Wiener kernel corresponding to the phase-averaged firing rate  $\langle m(t, \Theta, \Phi) \rangle_\Phi$  vanishes identically. (Recall from the remark at the end of Section 3 that the Wiener kernels  $W_1^{i,j}(\tau, \Theta, \Phi)$  in Eqs. (12) and (36), and  $W_1^i(\tau, \Theta)$  in Eqs. (13) and

(44), are *independent* of the presentation duration  $\nu$ .) At first, this result may appear paradoxical, however, one needs to recall the statement made at the end of Section 3 that  $\langle m(t, \Theta, \Phi) \rangle_\Phi$  can be thought of as arising from averaging over a set of V1 neurons subjected to the same stimuli and with identical receptive fields in all aspects except the preferred spatial phase. It thus appears that the averaging over the spatial phase will make the contributions of the receptive fields' on- and off-regions cancel each-other out, at least at the linear order, and will thus also wipe out the orientation tuning (cf. the end of Appendix 3).

In view of the vanishing of the first-order Wiener kernel  $W_1^i(\tau, \Theta)$  obtained in the previous paragraph, it is thus important to observe that, in the globally inhibited case in which  $C = \mathcal{O}(1)$ , the  $\mathcal{O}(\nu)$ -size orientation-tuned term in Eq. (42) is a contribution to the interval-specific firing rate average  $M(t, \theta_i)$  of the higher-order terms in the Wiener series computed with, and correlated against, the binary noises  $s_{vij}(t)$ . Likewise, since Eqs. (13) and (42) imply

$$\begin{aligned}
 0 = W_1^i(\tau, \Theta) & \\
 & = \frac{1}{J} \sum_{j=1}^J W_1^{i,j}(\tau, \Theta, \Phi) \\
 & = \nu^{-1/2} [M(\tau, \theta_i) - \langle M \rangle_{\tau, \theta^{(0)}}] + \mathcal{O}(\nu^{1/2}) \\
 & = \nu^{1/2} \frac{D[f^2(\theta^{(0)}) - f_2] G_{lgn}^2(\tau)}{4\sqrt{2\pi} V_0} e^{-C^2/2V_0D^2} + \mathcal{O}(\nu^{1/2}),
 \end{aligned}$$

this same  $\mathcal{O}(\nu)$ -size orientation-tuned term in Eq. (42) gives the  $\mathcal{O}(\nu^{1/2})$ -size error in the computation of the Wiener kernel  $W_1^i(\tau, \Theta)$  in Eq. (44) by using cross correlation with the binary noises  $s_{vij}(t)$  for small but finite  $\nu$  instead of the limiting correlated Gaussian white noises  $s_{ij}(t)$  (cf. Appendix 2). Moreover, we should notice that the factors in this tuned term are *quadratic* in the corresponding components  $f(\theta)$  and  $G_{lgn}(t)$  of the LGN kernel (22) in contrast to the linear powers of the same quantities present in the nonzero first-order Wiener kernels  $W_1^{i,j}(\tau, \Theta, \Phi)$  in Eq. (36). Finally, let us point out that the vanishing of the first-order Wiener kernel  $W_1^i(\tau, \Theta)$  forces the tuned portion of the interval-specific firing rate average  $M(t, \theta_i)$  to vanish at a faster rate in  $\nu$ , namely at  $\mathcal{O}(\nu)$ , than it would vanish otherwise. We will return to this issue below.

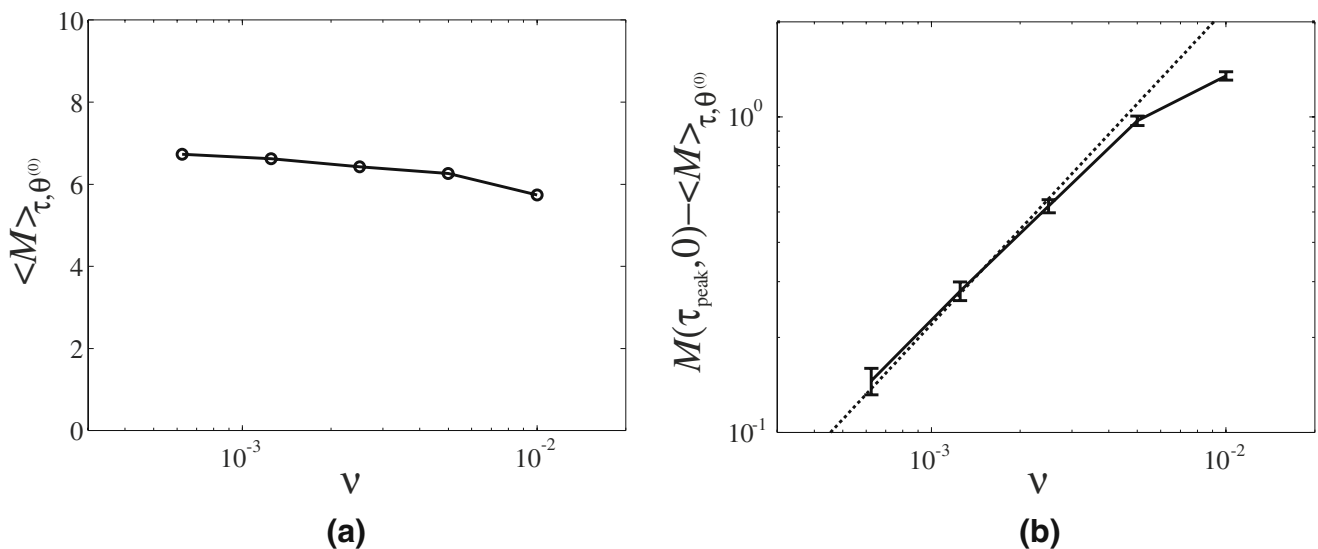
From Eqs. (33), (30), and (42) we can also conclude that the leaky feed-forward model (17) produces “separable” tuning curves in the limit of vanishing presentation time,  $\nu \rightarrow 0$ . In other words, at the leading order in  $\nu$ , the non-constant terms in the interval-specific firing

rate averages,  $N(\tau, \theta_i, \phi_j)$  in Eq. (33) and  $M(\tau, \theta_i)$  in Eq. (42), are products of spatial and temporal functions. This simply reflects the “separable” nature of the LGN Gabor kernel in Eq. (19), built into our model and preserved due to the model’s thresholded-linear nature. In Section 5, we show how the addition of cortico-cortical interactions even to a simplified linear model may destroy this “separability” of the V1 neuronal tuning curves even for a “separable” LGN kernel.

We further remark that formula (38) implies a completely untuned interval-specific firing rate average  $M(t, \theta^{(0)})$  in the case of a purely feed-forward model with no leakage or global inhibition. In other words, averaging over the spatial phase wipes out the orientation tuning of a purely feed-forward V1 neuron with no leakage. The proof is given in Appendix 5, and generalized to more realistic models that include descriptions of feed-forward convergence of inputs from individual on- and off-center LGN neurons onto the given V1 neuron, as well as *finite* presentation times  $\nu$ . This result gives an indication that, for a purely feed-forward V1 neuron with no leakage, the reverse-time correlation method of Ringach et al. (1997a, 2003) should yield untuned histograms, amplifying the similar conclusion reached analytically for a family of *instantaneous* LGN time response kernels in Pugh et al. (2000).

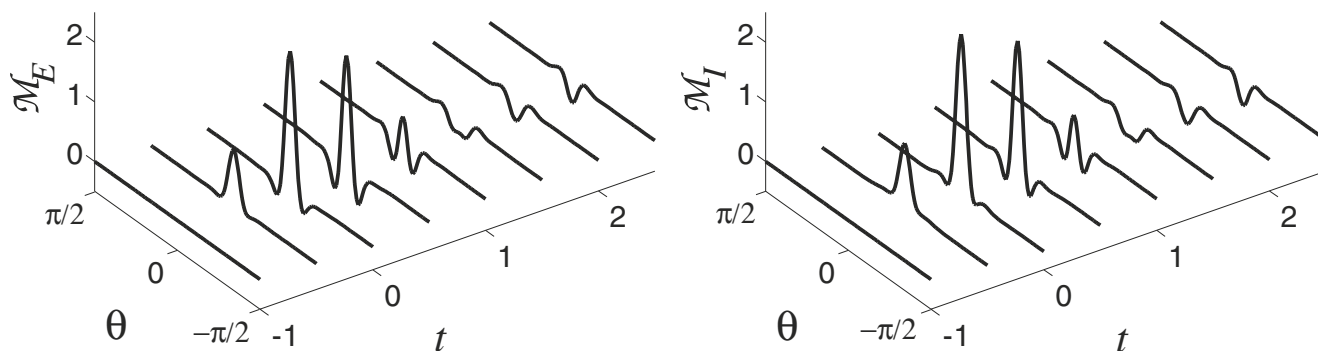
Finally, we comment on the relevance of the asymptotic formulas (42) and (43), expressing the interval-specific firing rate average  $M(\tau, \theta^{(0)})$ , for more

general models. In this regard, we show the results of numerical simulations of an integrate-and-fire ring model with a more realistic LGN input. The details of this model are described in Appendix 6, and the details of the corresponding LGN drive in Appendix 5. In particular, in this model, neither the artificial  $\mathcal{O}(\nu^{1/2})$  scaling of the background firing rate nor the added global inhibition is used. The results are presented in Fig. 2. The figure shows the dependence of the average  $\langle M \rangle_{\tau, \theta^{(0)}}$  and the difference  $M(\tau_{\text{peak}}, 0) - \langle M \rangle_{\tau, \theta^{(0)}}$  on the presentation duration  $\nu$  (here,  $\tau_{\text{peak}}$  is the value of  $\tau$  at the highest point of the interval-specific firing rate average  $M(\tau, \theta^{(0)})$ ). It can be seen that these quantities exhibit the same scaling as formula (42). In particular, the tuned portion of  $M(\tau, \theta^{(0)})$  scales as  $\mathcal{O}(\nu)$ . This scaling lets us conclude that the first-order Wiener kernel  $W_1^i(\tau, \Theta)$  of the phase-averaged firing rate  $\langle m(t, \Theta, \Phi) \rangle_{\Phi}$  must again vanish, or else the tuned portion of  $M(\tau, \theta^{(0)})$  would appear at  $\mathcal{O}(\nu^{1/2})$ . On the other hand, the absence of any tuning at  $\mathcal{O}(\nu)$  in formula (43) appears to be related to the complete absence of tuning in the case with no leakage described in the previous paragraph and further elaborated in Appendix 5. Namely, for relatively weak leakage of  $\mathcal{O}(1)$ , as compared to the intensity of the stimulus and the background firing rate of  $\mathcal{O}(\nu^{-1/2})$ , the relative strength of the tuned versus the untuned portions of the interval-specific firing rate average  $M(\tau, \theta^{(0)})$  should decrease, to reach zero with no leakage present.



**Fig. 2** RTC for an integrate-and-fire ring model of orientation tuning. Scaling of the background,  $\langle M \rangle_{\tau, \theta^{(0)}}$  in (a), and the peak tuned,  $M(\tau_{\text{peak}}, 0) - \langle M \rangle_{\tau, \theta^{(0)}}$  in (b), portions of the interval-specific firing rate average,  $M(\tau, \theta^{(0)})$ , for decreasing presentation duration  $\nu$ . The value  $M(\tau_{\text{peak}}, 0)$  is calculated at the peak

of  $M(\tau, \theta^{(0)})$ . The presentation duration  $\nu$  is given in seconds, the interval-specific firing rate average in spikes per second. The horizontal scales are logarithmic, as is the vertical scale in (b). The dotted line in (b) indicates slope one. The preferred orientation of the neuron under study is  $\Theta = 0$



**Fig. 3** Mexican-hat tuning curves generated by the linear ring model (45). *Left* excitatory neuron. *Right* inhibitory neuron. The time unit is  $\nu = 10$  ms, the orientation angle unit is radian. The LGN parameters are  $\zeta = 0, \alpha = 4, \beta = 0.5, \tau_{LGN} = 30$  ms in Eqs. (20) and (25), and  $\sigma_{LGN} = 0.15$  rad in Eq. (110). The cortical

parameters are the time constants,  $\tau_E = 2$  ms and  $\tau_I = 8$  ms in Eq. (115), the extents of the excitatory and inhibitory cortical connections in radians,  $\sigma_E = 0.4$  and  $\sigma_I = 0.2$  in Eq. (114), and the coupling constants  $S_{EE} = 0.8, S_{EI} = 7.6, S_{IE} = 1.5,$  and  $S_{II} = 7.6$  in Eqs. (45) and (16)

### 5 Mexican hat tuning curves

The shapes of the orientation tuning curves of V1 neurons, revealed by the reverse-time correlation procedure in the experiments of Ringach et al. (1997a), appear to be closely related to the architecture of the cortical synaptic connections impinging on the neuron under investigation. Important properties that these tuning curves may shed some light on are the extent of, and relationship between, the cortico-cortical excitation and inhibition. In particular, it is hypothesized that tuned inhibition suppresses intermediate orientations between the optimal and orthogonal, resulting in Mexican-hat tuning curves (Nelson and Frost 1978; Sillito et al. 1980; Ringach et al. 2003; Ringach and Shapley 2004; Xing et al. 2005). In this section, to gain theoretical insight, we discuss some simple network mechanisms that appear responsible for the emergence of such Mexican hats in the context of a simple idealized linear ring model.

First, we exhibit a simple idealized model of a coupled linear network that produces Mexican-hat shaped profiles. To this end, we again drastically simplify the thresholded linear model (14). In particular, we drop the thresholding, set the preferred phase of the neuron in the LGN drive (19) to  $\Phi = 0$ , fix all the phases in the stimulus at  $\phi^{(n)} \equiv 0$ , and drop the phase averaging. [It is important to comment here that if we retained phase averaging in this linear model, the combination of linearity and phase averaging would erase any orientation tuning. This would be a linear version of the vanishing of the first-order Wiener kernel  $W_1^i(\tau, \Theta)$  in Eq. (44)]. Finally, we consider only the ring architecture (Ben-Yishai et al. 1995; Pugh et al. 2000), that is, the position  $\mathbf{x}$  in the cortex is replaced by the preferred angle  $\Theta$ , so that the spatial convolution in Eq. (14) becomes a

convolution in the angle variable only. After a rescaling of variables and ensemble averaging over the presented angles  $\theta^{(n)}$  with  $n \neq 0$ , in Appendix 7, we derive the equations for the (rescaled) interval-specific firing rate average  $M_P(t, \theta^{(0)})$  to be

$$M_P(t, \theta) = [f(\theta) - \langle f \rangle_\theta] \mathcal{G}(t) + \sum_{P'} C_{PP'} K_{PP'} * G_{P'} * M_{P'}(t, \theta), \quad P, P' = E, I. \tag{45}$$

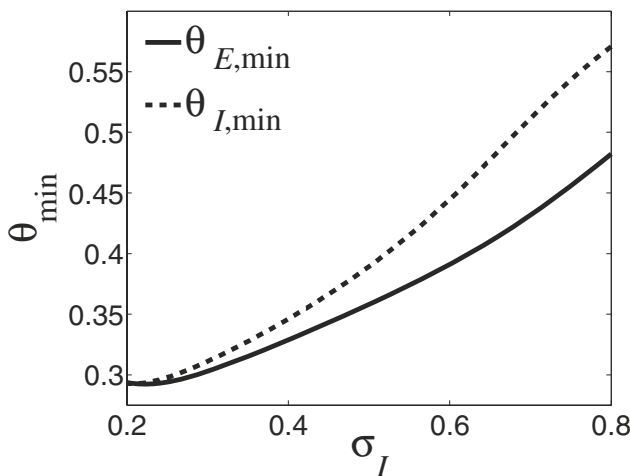
In this equation,  $f(\theta)$  is the spatial profile of the LGN input described by (some simplified version of) formula (23),  $\mathcal{G}(t)$  is the LGN time course in Eq. (25),  $C_{PP'}$  are the cortical coupling coefficients in Eq. (16), and  $K_{PP'}$  and  $G_{P'}$  are some appropriate angular and temporal cortical kernels, whose choice is described in Appendix 7. Equations (45) form a pair of linear convolution equations, which we solve in Appendix 7 by a combination of the Laplace transform in  $t$  and Fourier series in  $\theta$ . Here, we proceed directly to the results.

Instead of the (rescaled) interval-specific firing rate average  $M_P(t, \theta)$  in Eq. (5), we choose to present a similarly rescaled correlation function,  $\mathcal{M}_P(t, \theta)$  in Eq. (3) (recall that the two are connected via the formula (6b), and that for finite  $\nu$ ,  $\mathcal{M}_P(t, \theta)$  is a better approximation to the relevant spike count than  $M_P(t, \theta)$ , as explained at the end of Section 2). A pair of tuning curves  $\mathcal{M}_P(t, \theta), P = E, I$ , is plotted in Fig. 3. The time scale is measured in units of the presentation time  $\nu$ . The values of the LGN and cortical time constants, the cortical coupling constants, and the ratios between the angular extents of the excitatory and inhibitory cortical connections are displayed in the figure captions. The neuron's position within the corresponding orientation

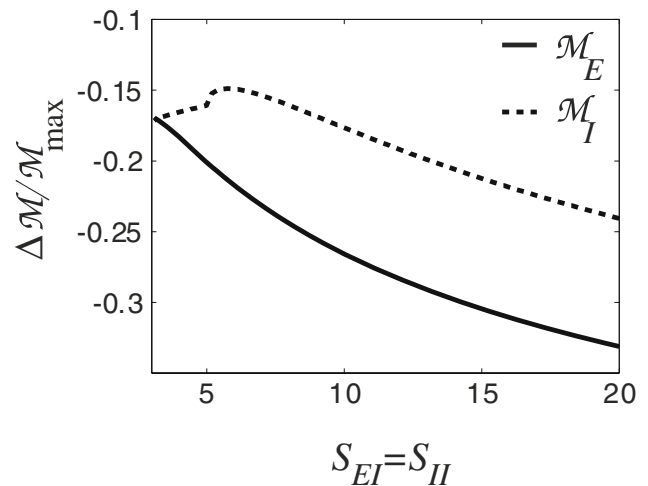


pinwheel is taken to be such that the  $\theta$ -width  $\sigma_{LGN}$  of the LGN input roughly equals the angular extent of the inhibitory cortical connections  $\sigma_I$ , which are both half the angular extent of the excitatory cortical connections  $\sigma_E$ . (See also the explanation below.) The connection between the sets of constants  $C_{PP}$  in the text and  $S_{PP}$  in the figures is given by formula (16).

From Fig. 3, we see that a typical tuning curve rises from zero shortly after  $t = -v$ , steepens into a bell-shaped curve, acquires two minima at two symmetric orientation angle values away from both the optimal and orthogonal that give it the shape of a Mexican hat, then develops an inversion at the preferred angle, and finally levels off to zero at large time values. The (at first rather confusing) property that the tuning curve starts rising at  $t = -v$  instead of the more expected  $t = 0$  is due to the averaging in Eq. (6b) and the finite length of the presentation,  $v$ . In particular, the firing rate of the neuron in question near the beginning of the time interval in which  $\theta = \theta_i$  is presented is clearly correlated with the orientation angle presented up to time  $v$  in the future, since it is the same angle  $\theta_i$ . This apparent artifact of the method is not present in the experiments because of the delay time that the stimulus needs to reach V1 after illuminating the retina. Also, the long inversion “trough” in Fig. 3 appears to be at least partly an artifact of the linearity of the model,



**Fig. 4** Increasing the extent of the inhibitory cortico-cortical connections increases the width of the Mexican hat. The values of the orientation angles  $\theta_{P,\min}$  at the minima of the correlation functions  $\mathcal{M}_P(t, \theta)$  are plotted versus the extent  $\sigma_I$  of the inhibitory cortical connections. The orientation angle unit is radian. The LGN parameters are  $\zeta = 0, \alpha = 4, \beta = 0.5, \tau_{LGN} = 30$  ms in Eqs. (20) and (25), and  $\sigma_{LGN} = 0.15$  rad in Eq. (110). The cortical parameters are the time constants,  $\tau_E = 2$  ms and  $\tau_I = 8$  ms in Eq. (115), the extent of the excitatory connections in radians,  $\sigma_E = 0.4$  in Eq. (114), and the coupling constants  $S_{EE} = 0.8, S_{EI} = 7.6, S_{IE} = 1.5,$  and  $S_{II} = 7.6$  in Eqs. (45) and (16)

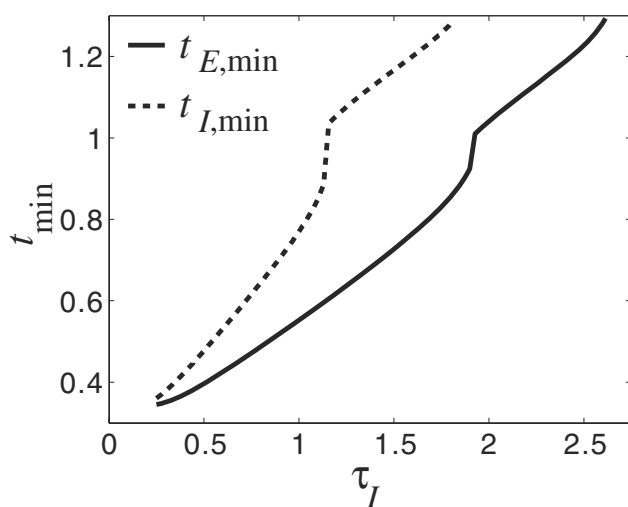


**Fig. 5** Increasing the strengths of the inhibitory cortico-cortical connections,  $S_{EI}$  and  $S_{II}$ , increases the depth of the Mexican hat. The dependent variable is depth =  $\Delta \mathcal{M}_P / \mathcal{M}_{P,\max}$ , with  $\Delta \mathcal{M}_P = \mathcal{M}_P(t_{\min}, \theta_{\min}) - \mathcal{M}_P(t_{\min}, \pi/2)$ , where  $(t_{\min}, \theta_{\min})$  is one of the two the lowest points, and  $\mathcal{M}_{P,\max}$  is the highest point of the Mexican-hat tuning curve. The LGN parameters are  $\zeta = 0, \alpha = 4, \beta = 0.5, \tau_{LGN} = 30$  ms in Eqs. (20) and (25), and  $\sigma_{LGN} = 0.15$  rad in Eq. (110). The cortical parameters are the time constants,  $\tau_E = 2$  ms and  $\tau_I = 8$  ms in Eq. (115), the extents of the excitatory and inhibitory cortical connections in radians,  $\sigma_E = 0.4$  and  $\sigma_I = 0.2$  in Eq. (114), and the coupling constants  $S_{EE} = 0.8$  and  $S_{IE} = 1.5$  in Eqs. (45) and (16)

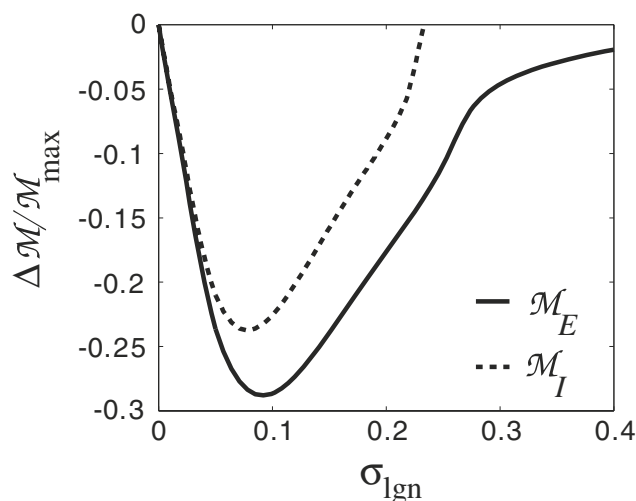
as neither formula (42) nor the experimental tuning curves in Fig. 1 show any indication of its presence.

The model’s oversimplifications notwithstanding, Fig. 3 clearly shows the presence of Mexican-hat tuning curves in this model. Moreover, these tuning curves exist when the *angular* extent of the cortical inhibition is one half the extent of the cortical excitation, roughly mimicking some of the local properties of V1 networks (McLaughlin et al. 2000) (in fact, in realistic V1 networks, this ratio is between the spatial extents of cortical inhibition and excitation. The two ratios, in angle and space, are almost identical for cells in iso-orientation domains far from pinwheel centers, while for cells near pinwheel centers the relationship between the two ratios is more complicated). This modeling detail stands in contrast to many other models, which use longer (angular or spatial) extent of the inhibition rather than the excitation (Ben-Yishai et al. 1995; Douglas et al. 1995; Somers et al. 1995; Ernst et al. 2001). In other respects, the tuning curves  $\mathcal{M}_P(t, \theta)$  exhibit the expected properties, as shown in Figs. 4, 5, and 6. In particular, Fig. 4 shows that the width of the Mexican hat is proportional to the extent of the cortico-cortical inhibitory connections, Fig. 5 shows that its depth is proportional to the strength of the inhibitory cortico-cortical synapses, and Fig. 6 shows that

its onset time is proportional to the inhibitory cortical time scale. Note that we define the position of the Mexican hat to be at one of the two symmetric local minima of the tuning curve over all times, see Fig. 3 (Fig. 6 shows that these minima tend to exist for moderate values of the inhibitory time scale  $\tau_I$  until they merge with the inversion “trough” at  $\theta = 0$  for larger values of  $\tau_I$ ). We define the depth of the Mexican hat to be the difference between the local minimum value of  $\mathcal{M}_P(t, \theta)$  and  $\mathcal{M}_P(t, \pi/2)$  at the same value of time,  $t$ , normalized by the maximal value of the tuning curve  $\mathcal{M}_P(t, \theta)$  over all times  $t$  and orientations  $\theta$ , i.e.,  $\text{depth} = \Delta\mathcal{M}_P/\mathcal{M}_{P,\max}$ , with  $\Delta\mathcal{M}_P = \mathcal{M}_P(t_{\min}, \theta_{\min}) - \mathcal{M}_P(t_{\min}, \pi/2)$ , where  $(t_{\min}, \theta_{\min})$  is one of the two lowest points, and  $\mathcal{M}_{P,\max}$  is the highest point of the Mexican-hat tuning curve. Note from Figs. 4, 5, and 6 that the Mexican hats are more pronounced in the tuning curves of excitatory than inhibitory neurons, and that the Mexican hats of the inhibitory neurons appear broader and have longer onset times. In addition, in Fig. 7, we present the dependence of the Mexican hat size on the position of the neuron in question within the orientation pinwheel, modeled by the ratio between the  $\theta$ -width of the LGN input and the extent of the cortical connections. The rationale for this modeling is that, while the angular width of the LGN input stays largely constant from one V1 neuron to the next, the effective



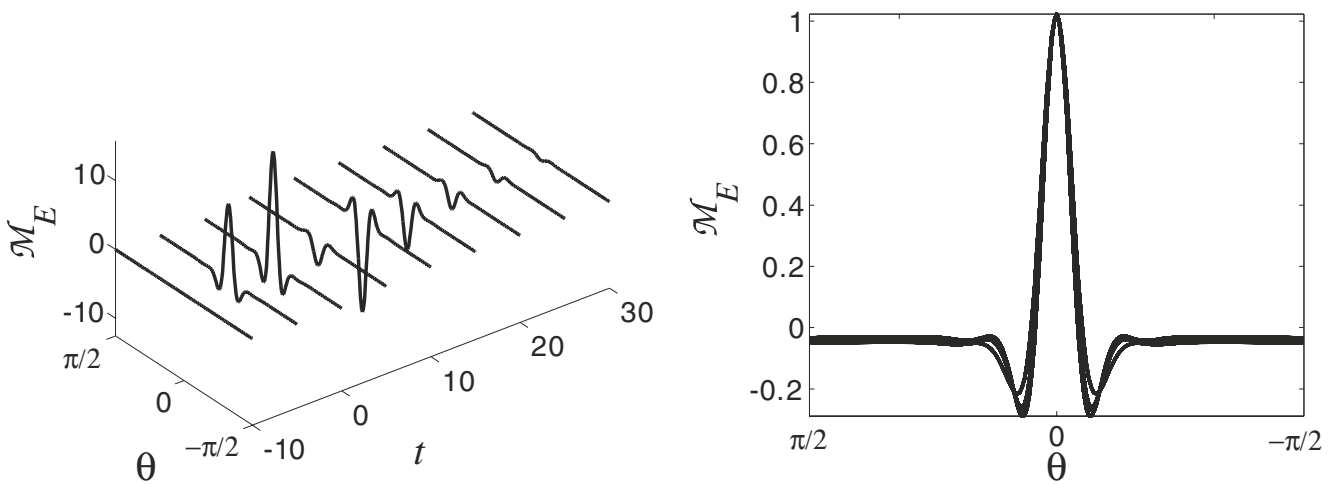
**Fig. 6** Increasing the inhibitory cortico-cortical decay time  $\tau_I$  increases the onset time  $t_{\min}$  of the Mexican hat. The kinks at  $t_{\min} = \nu$  in the graphs are artifacts of the insufficient smoothness of the model at  $t = \nu$ , as explained in Appendix 7 after formula (129). The time unit is  $\nu = 10$  ms. The LGN parameters are  $\zeta = 0$ ,  $\alpha = 4$ ,  $\beta = 0.5$ ,  $\tau_{\text{LGN}} = 30$  ms in Eqs. (20) and (25), and  $\sigma_{\text{LGN}} = 0.15$  rad in Eq. (110). The cortical parameters are the time constant  $\tau_E = 2$  ms in Eq. (115), the extents of the excitatory and inhibitory cortical connections in radians,  $\sigma_E = 0.4$  and  $\sigma_I = 0.2$  in Eq. (114), and the coupling constants  $S_{EE} = 0.8$ ,  $S_{EI} = 7.6$ ,  $S_{IE} = 1.5$ , and  $S_{II} = 7.6$  in Eqs. (45) and (16)



**Fig. 7** The depth of the Mexican hat first increases and then decreases with the distance from the pinwheel center. The relative  $\theta$ -width of the LGN input,  $\sigma_{\text{lgn}}$ , is inversely proportional to the distance of the neuron from the pinwheel center. The dependent variable is  $\text{depth} = \Delta\mathcal{M}_P/\mathcal{M}_{P,\max}$ , with  $\Delta\mathcal{M}_P = \mathcal{M}_P(t_{\min}, \theta_{\min}) - \mathcal{M}_P(t_{\min}, \pi/2)$ , where  $(t_{\min}, \theta_{\min})$  is one of the two lowest points, and  $\mathcal{M}_{P,\max}$  is the highest point of the Mexican-hat tuning curve. The orientation angle unit is radian. The LGN parameters are  $\zeta = 0$ ,  $\alpha = 4$ ,  $\beta = 0.5$ ,  $\tau_{\text{LGN}} = 30$  ms in Eqs. (20) and (25). The cortical parameters are the time constants,  $\tau_E = 2$  ms and  $\tau_I = 8$  ms in Eq. (115), the extents of the excitatory and inhibitory cortical connections in radians,  $\sigma_E = 0.4$  and  $\sigma_I = 0.2$  in Eq. (114), and the coupling constants  $S_{EE} = 0.8$ ,  $S_{EI} = 7.6$ ,  $S_{IE} = 1.5$ , and  $S_{II} = 7.6$  in Eqs. (45) and (16)

radius of the ring model decreases with the proximity to the pinwheel center, and thus the angular extents of the cortical connections increase in inverse proportion to it. In order to avoid periodicity artifacts, as an approximation, we rather scale the angular width of the LGN input instead as proportional to the neuron’s distance from the pinwheel center. As can be seen from Fig. 7, Mexican hats are the most pronounced at moderate distances from the pinwheel center, where the angular width of the LGN input is about half that of the inhibitory cortical connections, and vanish near the pinwheel center as well as in iso-orientation domains far from it.

Finally, directly from Eq. (45), we see that the feedforward version of the model produces separable orientation tuning curves. In other words, when all the cortico-cortical coupling constants  $C_{PP'} = 0$ , the interval-specific firing rate averages  $M_P(t, \theta)$  [and the correlation functions  $\mathcal{M}_P(t, \theta)$ ] are products of spatial and temporal terms. Clearly, for nonzero coupling constants  $C_{PP'}$ , the temporal onset of Mexican-hat tuning curves introduces dynamic changes in the tuning-curve shapes. In the cases when such a temporal onset occurs, such as that shown in Fig. 3, this onset excludes tuning-curve separability in space and time.



**Fig. 8** “Separability” of tuning curves generated by the linear ring model (45) for a long presentation duration,  $\nu = 100$  ms. *Left* tuning curves for an excitatory neuron. *Right* tuning curves from the left scaled by their height at  $\theta = 0$ . The *near-overlap of the scaled tuning curves* indicates approximate “separability.” The time unit is 10 ms, the orientation angle unit is radian. The LGN parameters are  $\zeta = 0$ ,  $\alpha = 4$ ,  $\beta = 0.5$ ,  $\tau_{LGN} = 30$  ms in Eqs. (20)

and (25), and  $\sigma_{LGN} = 0.15$  rad in Eq. (110). The cortical parameters are the time constants,  $\tau_E = 2$  ms and  $\tau_I = 8$  ms in Eq. (115), the extents of the excitatory and inhibitory cortical connections in radians,  $\sigma_E = 0.4$  and  $\sigma_I = 0.2$  in Eq. (114), and the coupling constants  $S_{EE} = 0.8$ ,  $S_{EI} = 7.6$ ,  $S_{IE} = 1.5$ , and  $S_{II} = 7.6$  in Eqs. (45) and (16). The tuning curves for an inhibitory neuron behave in an almost identical fashion

A remark concerning the loss of separability discussed in the previous paragraph, however, is appropriate here. In particular, one would intuitively expect that the dynamic changes in the tuning-curve shapes are produced by the interplay among the time scales of the LGN, cortico-cortical interactions, and the stimulus duration  $\nu$ . Therefore, these dynamic shape changes should be most readily observable when  $\nu$  is of the order comparable to the physiological time scales or shorter. On the other hand, for sufficiently long  $\nu$ , one would expect the memory of the short time-scale dynamics to be lost in the RTC procedure, and the tuning curves to be separable on the long,  $\mathcal{O}(\nu)$ , time scale. Such a scenario clearly takes place in our model, as can be seen in Fig. 8. This observation about the approximate separability for long  $\nu$  may also suggest a possible explanation for why tuning-curve separability was observed in Gillespie et al. (2001), where  $\nu = 40$ -100 ms, which appears to be somewhat longer than the typical time-scales involved in V1 dynamics.

### 6 Discussion

We have addressed theoretically the increasingly popular experimental method using RTC to study orientation tuning dynamics of V1 neurons (Ringach et al. 1997a, 2002, 2003; Gillespie et al. 2001; Müller et al. 2001; Bredfeldt and Ringach 2002; Dragoi et al. 2002;

Felsen et al. 2002; Mazer et al. 2002; Chen et al. 2005; Nishimoto et al. 2005; Xing et al. 2005). We have provided an exact mathematical description of this method, both of the stimuli used as a collection of correlated binary noise processes in Eqs. (66) and (67), and the collected delay-time-orientation-spike histograms as correlation functions and interval-specific firing rate averages in Eqs. (3), (5), and (7). This description complements the subspace projection interpretation of Ringach et al. (1997a, b), where each flashed grating within the stimulus is interpreted as lying in the particular subspace of the plane-wave Fourier basis of the stimuli on the screen which has the given V1 neuron’s preferred spatial frequency.

We have only considered flashed stimuli with randomly varying orientation angles and spatial phases. To bring out yet better the typical tuning response of a cortical neuron to a specific orientation embedded within a long, arbitrary stimulus, it appears reasonable to also add random variations in the spatial frequency. This was done in Ringach et al. (2003). As a result, the flashed gratings run through the entire (discretization of the) plane-wave Fourier basis, which makes the subspace projection approximation of Ringach et al. (1997a, b, 2003) more accurate. Adding random variations in the spatial frequency to the flashed gratings can also easily be incorporated into our exact description, but will not change any of our conclusions.

One advantage of our description appears to be the purely temporal interpretation (10) of orientation

tuning dynamics in terms of the nonlinear Volterra–Wiener systems theory. As a result, it is straight forward to extract the first-order Wiener kernels describing the tuning of a V1 neuron to the given orientation angle  $\theta_i$  and spatial phase  $\phi_j$  from the measured spike histograms, as displayed in Eqs. (12) and (13). In the experimentally most interesting case of orientation tuning alone when the histograms are averaged over the spatial phase, calculations on simple models (17) and (18) as well as the results of the more sophisticated integrate-and-fire ring model simulations shown in Fig. 2 indicate that the V1 neuron’s corresponding first-order Wiener kernel vanishes. This appears to happen because the spatial phase averaging is equivalent to adding firing rates from V1 neurons which are identical in all their properties except spatial phase, so that the contributions from the on- and off-regions of their receptive fields add to a constant at the linear order, which also erases the orientation tuning at that order. Thus, the measured histograms instead consist of contributions from the higher-order terms in the Wiener series. This does not happen in the absence of phase averaging: the first-order Wiener kernels (12) corresponding to the orientation-spatial phase tuning are definitely nonzero.

In practical calculations, such as estimations of receptive fields, one alternative to the Volterra–Wiener approach is to assume explicitly that a V1 neuron acts as a combination of a linear filter followed by a threshold nonlinearity. RTC is only used to determine the characteristics of the linear filter (Chichilnisky 2001; Nykamp and Ringach 2002; Nykamp 2003). The characteristics of the (assumed form of the) nonlinearity are computed from these and the experimental data. Another alternative is the kernel regression approach of Franz and Schölkopf (2006), discussed in Section 3.

The dynamic interplay between various types of cortical excitation and inhibition is believed to be brought out by the RTC, including in the form of Mexican-hat tuning profiles (Shapley et al. 2003; Ringach and Shapley 2004). Our explicitly solvable idealized model appears consistent with this belief. Previous theoretical works were divided on whether to model the (spatial or angular) extent of the local cortical inhibition in V1 as broader (Ben-Yishai et al. 1995; Douglas et al. 1995; Somers et al. 1995; Ernst et al. 2001) or narrower (McLaughlin et al. 2000; Kang et al. 2003) than that of the local cortical excitation, with broader inhibition thought of as being more likely to give rise to Mexican-hat orientation tuning profiles. The present work somewhat reconciles these two viewpoints. In our model, on the one hand, we have found Mexican-hat profiles when the angular extent of the cortical inhibi-

tion is twice that of the cortical excitation. On the other hand, though, what we find more relevant than the ratio of these two extents is the ratio of the effective angular width of the LGN drive of a given V1 neuron to the angular extent of its cortical inhibition. The Mexican-hat profiles appear the most pronounced when this ratio is about one half. This ratio may depend on the neuron’s position within an orientation pinwheel. While the angular width of the LGN drive can be reasonably assumed to be independent of the neuron’s position in the pinwheel, the effective angular extent of its inhibition appears to vary from covering all the angles near a pinwheel singularity to just one angle in iso-orientation domains far from it (Shelley and McLaughlin 2002). On the other hand, recent experimental (Marino et al. 2005) and theoretical (Tao et al. 2006) evidence points to the conclusion that, due to the sparsity of synapses, V1 neuronal discharges behave as if these neurons felt approximately the same angular extents of cortical excitation and inhibition, respectively, regardless of their position in the pinwheel.

Finally, in recent experiments (Xing et al. 2005; Chen et al. 2005), the dependence of the tuning-curve shapes on the size of the stimulus was studied. The appearance of Mexican hats was observed for large-size stimuli and linked to tuned inhibition evoked by long-range, orientation-specific, excitatory cortical connections to neurons outside of the give neuron’s orientation pinwheel. Deriving an idealized model to describe this situation is beyond the scope of the current paper and may be the topic of a future publication.

**Acknowledgements** The authors would like to thank J. Biello, P. Kramer, D. McLaughlin, D. Nykamp, A. Rangan, R. Shapley, and E. Vanden Eijnden for helpful discussions. G.K. was supported by NSF grants IGMS-0308943 and DMS-0506287, and gratefully acknowledges the hospitality of the Courant Institute of Mathematical Sciences and Center for Neural Science during his visits at New York University in 2000/01 and 2003/04. L.T. was supported by the NSF grant DMS-0506257. D.C. and M.J.S. were partly supported by the NSF grant DMS-0506396.

## Appendix

### 1 Probabilistic lemma

**Lemma** *Let  $x$  be drawn from the uniform distribution on  $[0, 1)$ , and  $y$  taken independently from another arbitrary distribution. Consider the random variable  $\tilde{z} = x + y + k$ , where  $k$  is that unique integer for which  $0 \leq x + y + k < 1$ . Then, the distribution of  $\tilde{z}$  is the uniform distribution on  $[0, 1)$ .*



(That is,  $\tilde{z}$  is the random variable  $z = x + y$  shifted into the unit interval. This shifting operation is how we understand the equivalence of the random variables  $\psi$  and  $\psi + 2\pi$  in studying  $\sin \psi$ .)

*Proof* Let  $P(y)$  be the density of the random variable  $y$ . The density of the random variable  $z = x + y$  is given then by the relation

$$R(z) = \int_0^1 P(z - x) dx,$$

and hence the distribution of  $\tilde{z}$  is given by

$$\tilde{R}(\tilde{z}) = \sum_{k=-\infty}^{\infty} R(\tilde{z} + k) = \int_{-\infty}^{+\infty} P(x) dx \equiv 1,$$

i.e., the uniform distribution. □

## 2 Discretization effects in the Volterra–Wiener theory

In this appendix, we discuss the error introduced in the reconstruction (58) of the Wiener kernels by using approximate binary random noise with finite presentation time  $\nu$  instead of the Gaussian white noise as the stimulus. For simplicity, we assume that the system in question can be described by a convergent Volterra series of the form (8). In particular, we show that the error is of order  $\mathcal{O}(\nu^{1/2})$ .

We do not address the related question of how to represent the responses to bandwidth-limited stimuli, such as orientation angles that can only change every  $\nu$  milliseconds, by a discrete version of the Volterra–Wiener theory. This type of a discrete version is developed, for instance in Klein and Yatsui (1979).

**Approximation of the Gaussian white-noise stimulus by binary white noise** As already discussed in Section 3, in order to find the Wiener kernels of a given nonlinear system, we “probe” this system with Gaussian white-noise stimuli. Such stimuli, say  $s_0(t)$ , have the correlation properties

$$\langle s_0(t_1) s_0(t_2) \rangle = A^2 \delta(t_1 - t_2), \tag{46a}$$

$$\langle s_0(t_1) \dots s_0(t_{2n+1}) \rangle = 0, \tag{46b}$$

$$\langle s_0(t_1) \dots s_0(t_{2n}) \rangle = \sum \prod \langle s_0(t_i) s_0(t_j) \rangle. \tag{46c}$$

Here,  $\langle \cdot \rangle$  denotes the ensemble averaging over all realizations of the white noise  $s_0(t)$ ,  $A^2$  is the power of the

stimulus,  $\delta(t)$  is the Dirac function, and the sum in the last equation runs over all distinct products of  $n$  two-point correlations.

In practice, we can replace the probing Gaussian white-noise stimulus  $s_0(t)$  by the binary white noise stimulus with finite presentation time

$$s_\nu(t) = \frac{A}{\sqrt{\nu}} \begin{cases} p, & \text{with probability } q, \\ -q, & \text{with probability } p, \end{cases} \tag{47}$$

$$k\nu < t < (k+1)\nu, \quad k = \dots, -1, 0, 1, \dots,$$

where  $p + q = 1$ . The values of  $s_\nu(t)$  in different intervals  $k\nu < t < (k+1)\nu$  are assumed to be independent. The correlation properties of this binary white noise are as follows: First,

$$\langle s_\nu(t) \rangle = 0. \tag{48a}$$

Second, if  $t_1, \dots, t_m$  all lie in the same interval  $k\nu < t < (k+1)\nu$ , then

$$\langle s_\nu(t_1) \dots s_\nu(t_m) \rangle = A^m \nu^{-\frac{m}{2}} [p^m q + p(-q)^m]. \tag{48b}$$

Finally,

$$\langle s_\nu(t_1) \dots s_\nu(t_n) \rangle = \sum \prod \langle s_\nu(t_i) \dots s_\nu(t_j) \rangle, \tag{48c}$$

where within each factor  $\langle s_\nu(t_i) \dots s_\nu(t_j) \rangle$ , all the times  $t_l$  lie in the same interval  $k\nu < t < (k+1)\nu$ , every  $s_\nu(t_l)$  is represented in each term once and only once, and the sum is performed over all distinct such terms.

Let us now show that the correlation function relations (48c) converge to the correlation function relations (46) of the Gaussian white noise  $s_0(t)$  as  $\nu \rightarrow 0$ . To this end, let us compute the integral

$$\int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} L(\tau_1, \dots, \tau_n) \langle s(t - \tau_1) \dots s(t - \tau_n) \rangle d\tau_1 \dots d\tau_n, \tag{49}$$

with an arbitrary function  $L(\tau_1, \dots, \tau_n)$ , which we can assume to be symmetric in all its arguments. Formula (48c) and the symmetry of the function  $L$  show that the integral in Eq. (49) breaks up into a linear combination of integrals of the form

$$\int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} L(\tau_1, \dots, \tau_n) \prod_{j=1}^l \langle s(t - \tau_{p_{j-1}+1}) \dots s(t - \tau_{p_j}) \rangle d\tau_1 \dots d\tau_n \tag{50}$$

where  $0 = p_0 < p_1 < \dots < p_{l-1} < p_l = n$ , and all the indices in the product, regardless of the angle brackets, are ordered as  $1, \dots, n$ . In formula (50), all the

times  $\tau_k$ , with  $p_{j-1} < k < p_j$ , are contained in the same intervals  $m_j v < t < (m_j + 1)v$ . Formula (48b)

shows that the integral in Eq. (50) is proportional to the expression

$$\begin{aligned}
 & A^n v^{-\frac{n}{2}} \sum_{m_1=-\infty}^{\infty} \dots \sum_{m_l=-\infty}^{\infty} \int_0^v \dots \int_0^v L(t - m_1 v - \sigma_1, \dots, t - m_1 v - \sigma_{p_1}, \\
 & \quad \dots, t - m_l v - \sigma_{p_{l-1}+1}, \dots, t - m_l v - \sigma_{p_l}) d\sigma_1 \dots d\sigma_n \\
 &= A^n v^{-\frac{n}{2}} v^{n-l} \left[ \sum_{m_1=-\infty}^{\infty} \dots \sum_{m_l=-\infty}^{\infty} L(t - m_1 v, \dots, t - m_1 v, \dots, t - m_l v, \dots, t - m_l v) v^l + \mathcal{O}(v) \right] \\
 &= A^n v^{\frac{n}{2}-l} \left[ \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} L(t_1, \dots, t_1, \dots, t_l, \dots, t_l) dt_1 \dots dt_l + \mathcal{O}(v) \right]. \tag{51}
 \end{aligned}$$

In the last two lines,  $m_j$  and  $t_j$ , respectively, appear as arguments of the function  $L$  precisely  $p_j - p_{j-1}$  times. The estimate showing that the error inside the square brackets is really  $\mathcal{O}(v)$  will be presented later.

First, however, let us discuss the scaling in the formula (51). By Eq. (48a), it should be clear that the smallest number of factors within any pair of angle brackets in formula (50) can be two. Therefore,  $l \leq n/2$ . Moreover,  $l = n/2$  precisely when  $n$  is even, and each pair of angle brackets in Eq. (50) contains precisely two factors. If  $n$  is odd, the largest terms are of the order  $v^{\frac{1}{2}}$ . In other words, as  $v \rightarrow 0$ , in the sense of distributions, the correlation function relation (48c) limits onto the correlation function relations (46) with an error of order  $v^{\frac{1}{2}}$ .

We now make the  $\mathcal{O}(v)$  estimate in formula (51) precise. In fact, we only show that the expression on the first two lines equals the expression on the last line. The equality with the middle line is shown in a similar fashion. We need to estimate the difference

$$\begin{aligned}
 & v^{l-n} \sum_{m_1=-\infty}^{\infty} \dots \sum_{m_l=-\infty}^{\infty} \int_0^v \dots \int_0^v L(t - m_1 v - \sigma_1, \dots, t - m_1 v - \sigma_{p_1}, \\
 & \quad \dots, t - m_l v - \sigma_{p_{l-1}+1}, \dots, t - m_l v - \sigma_{p_l}) d\sigma_1 \dots d\sigma_n \\
 & - \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} L(t_1, \dots, t_1, \dots, t_l, \dots, t_l) dt_1 \dots dt_l, \tag{52}
 \end{aligned}$$

in which  $t_j$  appears in precisely those slots in  $L$  in which  $m_j$  appears. This difference can be rewritten as

$$\begin{aligned}
 & v^{l-n} \sum_{m_1=-\infty}^{\infty} \dots \sum_{m_l=-\infty}^{\infty} \int_0^v \dots \int_0^v \left[ L(t - m_1 v - \sigma_1, \dots, t - m_1 v - \sigma_{p_1}, \right. \\
 & \quad \dots, t - m_l v - \sigma_{p_{l-1}+1}, \dots, t - m_l v - \sigma_{p_l}) \\
 & - L(t - m_1 v - \sigma_{p_1}, \dots, t - m_1 v - \sigma_{p_1}, \\
 & \quad \dots, t - m_l v - \sigma_{p_l}, \dots, t - m_l v - \sigma_{p_l}) \left. \right] d\sigma_1 \dots d\sigma_n. \tag{53}
 \end{aligned}$$

The absolute value of the expression in the square brackets can be estimated as smaller than

$$\begin{aligned}
 & R(t - m_1 v - \sigma_{p_1}, t - m_2 v - \sigma_{p_2}, \dots, t - m_l v - \sigma_{p_l}) \\
 & \quad \times (|\sigma_1 - \sigma_{p_1}| + \dots + |\sigma_{p_{l-1}} - \sigma_{p_1}| + |\sigma_{p_{l+1}} - \sigma_{p_2}| + \\
 & \quad \dots + |\sigma_{p_{2-1}} - \sigma_{p_2}| + \dots + |\sigma_{p_{l-1}+1} - \sigma_{p_l}| + \\
 & \quad \dots + |\sigma_{p_{l-1}} - \sigma_{p_l}|), \tag{54}
 \end{aligned}$$

where  $R$  is some non-negative integrable function of  $l$  arguments. In particular, the suprema of the partial derivatives of  $L$  in some fixed-size neighborhood of the point

$$\begin{aligned}
 & (t - m_1 v - \sigma_{p_1}, \dots, t - m_1 v - \sigma_{p_1}, \\
 & \quad \dots, t - m_l v - \sigma_{p_l}, \dots, t - m_l v - \sigma_{p_l})
 \end{aligned}$$

will be the ingredients of  $R$ . The integral  $\int_0^v |\sigma_i - \sigma_j| d\sigma_i$  can easily be evaluated to be

$$\int_0^v |\sigma_i - \sigma_j| d\sigma_i = \frac{v^2}{4} + \left(\frac{v}{2} - \sigma_j\right)^2, \tag{55}$$

which is a parabola, whose maximal value in the interval  $0 \leq \sigma_j \leq v$  is  $v^2/2$ . Therefore, the difference in Eq. (52) can be bounded in absolute value by the expression

$$\begin{aligned}
 & v^{l-n} (n-l) \frac{v^{n+1-l}}{2} \sum_{m_1=-\infty}^{\infty} \dots \sum_{m_l=-\infty}^{\infty} \int_0^v \dots \\
 & \quad \int_0^v R(t - m_1 v - u_1, \dots, t - m_l v - u_l) du_1 \dots du_l \\
 & = \frac{(n-l)v}{2} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} R(v_1, \dots, v_l) dv_1 \dots dv_l \\
 & = \mathcal{O}(v), \tag{56}
 \end{aligned}$$

as promised.

**Reconstruction of the Wiener kernels by using binary white noise with finite presentation time**

We proceed to study the error incurred by using only the binary white noise  $s_v(t)$  in Eq. (47) instead of all the realizations of the Gaussian white noise  $s_0(t)$  to probe the system (8). We first recall that the Wiener representation of Eq. (8) consists of a series

$$r[s](t) = P_0[s] + P_1[s](t) + P_2[s](t) + \dots + P_n[s](t) + \dots, \tag{57}$$

in which the terms  $P_n[s](t)$  are called the  $n$ -th order Wiener functionals. Each is an  $n$ -th order Volterra polynomial in the stimulus  $s(t)$  with the leading-order Volterra kernel  $W_n(T_1, \dots, T_n)$ , which we will call the  $n$ -th leading-order Wiener kernel (cf. Schetzen 1980). The functionals  $P_n[s](t)$  are pairwise orthogonal with respect to the inner product  $\langle P_m[s_0](t)P_n[s_0](t) \rangle$ , with  $\langle \cdot \rangle$  denoting ensemble averaging over all possible realizations of the Gaussian white noise stimulus,  $s_0(t)$ . It is known that the  $n$ -th leading-order Wiener kernel uniquely determines the entire Wiener functional  $P_n[s](t)$ , that is, all the lower-order Volterra kernels in it, see Wiener (1958), Schetzen (1980), Victor (1992).

In practice, one probes for the Wiener kernels,  $W_m(T_1, \dots, T_m)$  by using the Lee–Schetzen cross correlation algorithm (Schetzen 1980). In particular, for pairwise distinct values  $T_1, \dots, T_m$ , the  $m$ -th leading-order Wiener kernel  $W_m(T_1, \dots, T_m)$  can be computed by cross correlation as

$$W_m(T_1, \dots, T_m) = \frac{1}{A^{2^m m!}} \langle r[s_0](t)s_0(t - T_1) \dots s_0(t - T_m) \rangle, \tag{58}$$

where  $s_0(t)$  is zero-mean Gaussian white noise [it therefore has the correlation properties (46)]. Assuming ergodicity, one can write Eq. (58) in the form

$$W_m(T_1, \dots, T_m) = \frac{1}{A^{2^m m!}} \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T r[s_0](t)s_0(t - T_1) \dots s_0(t - T_m) ds \tag{59}$$

for any particular realization of the white noise  $s_0(t)$ . For arbitrary  $T_1, \dots, T_m$ , in order to avoid singularities when some of the  $T_j$ 's are equal, we must use the formula

$$W_m(T_1, \dots, T_m) = \frac{1}{A^{2^m m!}} \langle \{r[s_0](t) - r_{m-1}[s_0](t)\} s_0(t - T_1) \dots s_0(t - T_m) \rangle, \tag{60}$$

where  $r_{m-1}[s_0](t)$  is the  $(m - 1)$ -st order Wiener approximation of the response  $r[s_0](t)$ . We will not use this last result here, but will instead handle the singularities explicitly. In any given system, the leading-order Wiener kernels completely determine the dependence of the response  $r[s](t)$  on any stimulus  $s(t)$  via the Wiener series (57). For the important case  $m = 1$ , the corresponding Wiener functional contains just one term and equals

$$P_1[s](t) = \int_0^\infty W_1(T_1)s(t - T_1) dT_1, \tag{61}$$

with the first-order Wiener kernel  $W_1(T_1)$  computed from Eq. (9).

We now take a closer look at the expression (58). Let us use the notation  $\tilde{W}_m$  for the right-hand side of Eq. (58) with  $s_0(t)$  replaced by  $s_v(t)$ , regardless of the mutual position of  $T_1, \dots, T_m$ . Using Eq. (8), we have

$$\begin{aligned} \tilde{W}_m(T_1, \dots, T_m) &= \langle r[s_v](t)s_v(t - T_1) \dots s_v(t - T_m) \rangle \\ &= L_0 \langle s_v(t - T_1) \dots s_v(t - T_m) \rangle \\ &\quad + \int_0^\infty L_1(\tau_1) \langle s_v(t - T_1) \dots s_v(t - T_m)s_v(t - \tau_1) \rangle d\tau_1 \\ &\quad + \frac{1}{2!} \int_0^\infty \int_0^\infty L_2(\tau_1, \tau_2) \langle s_v(t - T_1) \dots s_v(t - T_m) \\ &\quad \times s_v(t - \tau_1)s_v(t - \tau_2) \rangle d\tau_1 d\tau_2 + \dots \\ &\quad + \frac{1}{n!} \int_0^\infty \dots \int_0^\infty L_n(\tau_1, \dots, \tau_n) \langle s_v(t - T_1) \dots s_v(t - T_m) \\ &\quad \times s_v(t - \tau_1) \dots s_v(t - \tau_n) \rangle d\tau_1 \dots d\tau_n + \dots \end{aligned} \tag{62}$$

(Here and from now on, we abuse the notation  $\langle \cdot \rangle$  to now denote ensemble averaging only over all the realizations of the noise  $s_v(t)$ , with fixed  $v$ .) Extending all the integrals to the entire real axis by defining the kernels to vanish if any of the arguments is negative and using Eq. (48c), we find, as in showing Eq. (50), that the general term in formula (62) is equal to a linear combination of terms of the form

$$\begin{aligned} &\langle s_v(t - T_1) \dots s_v(t - T_{q_1}) \rangle \dots \langle s_v(t - T_{q_{i-1}+1}) \dots s_v(t - T_{q_i}) \rangle \\ &\int_{-\infty}^\infty \dots \int_{-\infty}^\infty L_n(\tau_1, \dots, \tau_n) \langle s_v(t - \tau_1) \dots s_v(t - \tau_{p_1}) \rangle \dots \\ &\quad \times \langle s_v(t - \tau_{p_{j+1}+1}) \dots s_v(t - \tau_{p_j}) \rangle \langle s_v(t - T_{q_{i+1}}) \dots \\ &\quad \times s_v(t - T_{q_{i+1}})s_v(t - \tau_{p_{j+1}}) \dots s_v(t - \tau_{p_{j+1}}) \rangle \dots \\ &\quad \times \langle s_v(t - T_{q_{n-1}+1}) \dots s_v(t - T_{q_n})s_v(t - \tau_{p_{k-1}+1}) \dots \\ &\quad \times s_v(t - \tau_{p_k}) \rangle d\tau_1 \dots d\tau_n. \end{aligned} \tag{63}$$

Due to the symmetry of the kernels  $L_n$  and  $\tilde{W}_m$  in all their arguments, all the indices in  $T_i$ 's and  $\tau_j$ 's can be considered ordered, so that  $q_h = m$  and  $p_k = n$ . Since, again, the smallest number of factors inside a pair of angle brackets is two, we must have  $i \leq m/2$ ,  $j \leq n/2$  and  $i + j + l \leq (m + n)/2$ , where  $l$  is the number of correlations in which both  $s_\nu(t - T_\alpha)$ 's and  $s_\nu(t - \tau_\beta)$ 's appear.

Let us define the functions  $B(t_1, \dots, t_p)$  as follows. First,  $B(t) = 1$ . Second, for  $p > 1$ ,  $B(t_1, \dots, t_p) = 1$  if  $\kappa\nu < t_1, \dots, t_p < (\kappa + 1)\nu$  for some integer  $\kappa$ , and  $B(t_1, \dots, t_p) = 0$  otherwise.

Returning to Eq. (63), using Eq. (48b), we find that Eq. (63) is proportional to

$$\begin{aligned}
 & \nu^{-\frac{n+m}{2}} B(T_1, \dots, T_{q_1}) \dots B(T_{q_{i-1}+1}, \dots, T_{q_i}) B(T_{q_i+1}, \dots, T_{q_{i+1}}) \dots B(T_{q_{h-1}+1}, \dots, T_{q_h}) \\
 & \times \sum_{m_1=-\infty}^{\infty} \dots \sum_{m_i=-\infty}^{\infty} \int_0^\nu \dots \int_0^\nu L_n(t - m_1\nu - \sigma_1, \dots, t - m_1\nu - \sigma_{p_1}, \dots, \\
 & t - m_i\nu - \sigma_{p_{i-1}+1}, \dots, t - m_i\nu - \sigma_{p_i}, t - k_1\nu - \sigma_{p_i+1}, \dots, t - k_1\nu - \sigma_{p_{i+1}}, \dots, \\
 & t - k_l\nu - \sigma_{p_{k-1}+1}, \dots, t - k_l\nu - \sigma_{p_k}) d\sigma_1 \dots d\sigma_n \\
 & = \nu^{-\frac{n+m}{2}} B(T_1, \dots, T_{q_1}) \dots B(T_{q_{i-1}+1}, \dots, T_{q_i}) B(T_{q_i+1}, \dots, T_{q_{i+1}}) \dots B(T_{q_{h-1}+1}, \dots, T_{q_h}) \\
 & \quad \times \nu^{p_j-j} \nu^{n-p_j} \left[ \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} L_n(u_1, \dots, u_1, \dots, u_j, \dots, u_j, t - k_1\nu, \dots, t - k_1\nu, \dots, \right. \\
 & \quad \left. \dots, t - k_l\nu, \dots, t - k_l\nu) du_1 \dots du_j + \mathcal{O}(\nu) \right] \\
 & = \nu^{-\frac{n-m}{2}-j} B(T_1, \dots, T_{q_1}) \dots B(T_{q_{i-1}+1}, \dots, T_{q_i}) B(T_{q_i+1}, \dots, T_{q_{i+1}}) \dots B(T_{q_{h-1}+1}, \dots, T_{q_h}) \left[ \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \right. \\
 & \quad \left. \times L_n(u_1, \dots, u_1, \dots, u_j, \dots, u_j, t - k_1\nu, \dots, t - k_1\nu, \dots, t - k_l\nu, \dots, t - k_l\nu) du_1 \dots du_j + \mathcal{O}(\nu) \right]. \tag{64}
 \end{aligned}$$

Here,

$$k_1\nu < T_{q_i+1}, \dots, T_{q_{i+1}} < (k_1 + 1)\nu.$$

...

$$k_l\nu < T_{q_{h-1}+1}, \dots, T_{q_h} < (k_l + 1)\nu.$$

Integrating expression (64) against a function of  $T_1, \dots, T_m$ , we see that the result scales as  $\nu^{\frac{n+m}{2}-i-j-l}$ . This result can never become singular, and  $\mathcal{O}(1)$  terms arise precisely when  $i + j + l = (n + m)/2$ , that is, when all the correlations are among pairs, which gives the correct integral against the leading-order Wiener kernel,  $W_m(T_1, \dots, T_m)$ , in the limit as  $\nu \rightarrow 0$ . The first-order error is  $\mathcal{O}(\nu^{\frac{1}{2}})$ .

Expression (64) clearly can have singularities as  $\nu \rightarrow 0$ . Since  $j \leq n/2$ , the worst possible singularity is of order  $\nu^{-\frac{m}{2}}$ . In this case, we must also have  $l = 0$ . As explained in the previous paragraph, this singularity may or may not contribute to an integral against a function of  $T_1, \dots, T_m$ , depending on the structure of the product of the  $B$ 's. The singularities whose integrals

are nonzero must be eventually subtracted from the stimulus in order to determine the  $n$ -th leading-order Wiener kernel as described by formula (62). Lower-order singularities pollute the discrete leading-order Wiener kernel pointwise, but are harmless in the actual Wiener representation of the systems, and can therefore simply be discarded.

Let us finally concentrate more specifically on the first Wiener kernel,  $\tilde{W}(T_1)$ . For this kernel,  $m = 1$ , and therefore necessarily also  $l = 1$ . Terms of  $\mathcal{O}(1)$  arise for odd  $n$  and  $j = (n - 1)/2$ . These sum up to the first-order Wiener kernel  $W_1(T_1)$ . The  $\mathcal{O}(\nu^{\frac{1}{2}})$  terms arise for even  $n$ , and comprise the lowest-order error. There are no singular terms in this case.

### 3 Volterra–Wiener theory for correlated white noise stimuli

In this section we present the detailed derivations of the results presented in Section 3 that connect the

quantities related to orientation tuning in V1 under the stimulus (1), used in the experiments of Ringach et al. (1997a, 2003), and the Volterra–Wiener theory of nonlinear systems.

**Experimental stimulus as a collection of correlated binary noise and its Gaussian white noise limit** First, we establish in what manner the stimulus (1) can be mapped into a set of  $IJ$  binary random noise stimuli. We begin by rewriting Eq. (1) in the form

$$\begin{aligned} \mathcal{I}(\mathbf{X}, t) &= \mathcal{I}_0 v^{-1/2} \left\{ 1 + \varepsilon \sum_{i=1}^I \sum_{j=1}^J \chi_{\theta_i, \phi_j}(t) \sin[\mathbf{k}(\theta_i) \cdot \mathbf{X} - \phi_j] \right\}, \end{aligned} \tag{65}$$

where each  $\chi_{\theta_i, \phi_j}(t)$  is the indicator function of the set of intervals in which the grating with the orientation angle  $\theta_i$  and the spatial phase  $\phi_j$  was presented. In other words, the value of  $\chi_{\theta_i, \phi_j}(t)$  equals 1 during those time intervals in which the grating with the orientation-phase pair  $(\theta_i, \phi_j)$  was presented, and 0 during the time intervals when any other grating was presented.

We now further rewrite Eq. (65) as

$$\mathcal{I}(\mathbf{X}, t) = \mathcal{I}_0 \left\{ v^{-1/2} + \varepsilon \sum_{i=1}^I \sum_{j=1}^J \left[ s_{vij}(t) + \frac{1}{\sqrt{v}IJ} \right] \sin[\mathbf{k}(\theta_i) \cdot \mathbf{X} - \phi_j] \right\}, \tag{66}$$

where

$$s_{vij}(t) = v^{-1/2} \left[ \chi_{\theta_i, \phi_j}(t) - \frac{1}{IJ} \right], \quad i=1, \dots, I, \quad j=1, \dots, J,$$

is now indeed a collection of  $IJ$  binary noise processes. The process  $s_{vij}$  corresponds to the flashing of the grating with the orientation  $\theta_i$  and spatial phase  $\phi_j$ . Since, within any given time interval  $kv < t < (k+1)v$ , a grating with the orientation-phase pair  $(\theta_i, \phi_j)$  is presented with probability  $1/IJ$  and thus not presented with probability  $1 - 1/IJ$ , each process  $s_{vij}$  assumes the values

$$s_{vij}(t) = \frac{1}{\sqrt{v}} \begin{cases} p, & \text{with probability } q, \\ -q, & \text{with probability } p, \end{cases} \quad kv < t < (k+1)v, \quad k = \dots, -1, 0, 1, \dots, \tag{67}$$

where

$$q = \frac{1}{IJ}, \quad p = 1 - \frac{1}{IJ}.$$

The binary noise processes  $s_{vij}$  in Eqs. (66) and (67) must necessarily be correlated because only one grating is presented at any given time. In particular, two processes,  $s_{vij}(t)$  and  $s_{vmn}(t)$ , are correlated through the following conditional probabilities: At any given time  $t$ , if  $s_{vij}(t) = p/\sqrt{v}$ , then  $s_{vmn}(t) = -q/\sqrt{v}$  with probability 1, that is, if the grating with the orientation  $\theta_i$  and spatial phase  $\phi_j$  is presented, no other grating can be presented at the same time. If  $s_{vij}(t) = -q/\sqrt{v}$ , then

$$s_{vmn}(t) = \frac{1}{\sqrt{v}} \begin{cases} p, & \text{with probability } \frac{1}{IJ-1}, \\ -q, & \text{with probability } 1 - \frac{1}{IJ-1}, \end{cases} \tag{68}$$

which means that if the grating with the orientation  $\theta_i$  and spatial phase  $\phi_j$  is not presented, any other grating may be presented with probability  $1 - 1/(IJ - 1)$ . The values of  $s_{vij}(t_1)$  and  $s_{vmn}(t_2)$  in different intervals,  $k_1v < t_1 < (k_1 + 1)v$ ,  $k_2v < t_2 < (k_2 + 1)v$ , are independent, as gratings in such intervals are presented randomly and independently.

From the above, noting that  $s_{vij}$  is scaled as  $1/\sqrt{v}$ , in the limit as  $v \rightarrow 0$ , we can show that the binary processes  $s_{vij}(t)$  tend to correlated Gaussian white noise signals  $s_{ij}(t)$ , with

$$\langle s_{ij}(t) \rangle_{\vec{s}} = 0, \tag{69a}$$

and correlation properties

$$\langle s_{ij}(t_1) s_{ij}(t_2) \rangle_{\vec{s}} = \frac{IJ-1}{I^2 J^2} \delta(t_1 - t_2), \tag{69b}$$

$$\langle s_{ij}(t_1) s_{kl}(t_2) \rangle_{\vec{s}} = -\frac{1}{I^2 J^2} \delta(t_1 - t_2), \quad i \neq k \text{ or } j \neq l, \tag{69c}$$

$$\langle s_{i_1 j_1}(t_1) \dots s_{i_{2n+1} j_{2n+1}}(t_{2n+1}) \rangle_{\vec{s}} = 0, \tag{69d}$$

$$\langle s_{i_1 j_1}(t_1) \dots s_{i_{2n} j_{2n}}(t_{2n}) \rangle_{\vec{s}} = \sum \prod \langle s_{i_l j_l}(t_l) s_{i_m j_m}(t_m) \rangle_{\vec{s}}. \tag{69e}$$

where  $\delta(\cdot)$  denotes the Dirac delta function,  $\langle \cdot \rangle_{\vec{s}}$  denotes ensemble averaging over all realizations of the collection  $\vec{s}$  of the correlated noise stimuli,  $\vec{s} = \{s_{kl} \mid k = 1, \dots, I, l = 1, \dots, J\}$ , and the sum in the last equation runs over all distinct products of  $n$  two-point correlations. Observe that the correlated noise signals  $s_{ij}(t)$  decorrelate if  $I \rightarrow \infty$  or  $J \rightarrow \infty$ . We do not present the derivation of Eq. (69) here, but only note that it is similar to the derivation of formulas (48) in Appendix 2



in the  $\nu \rightarrow 0$  limit. We do notice, however, another important property of the noises  $s_{ij}(t)$ , namely

$$\sum_{i=1}^I \sum_{j=1}^J s_{ij}(t) = 0. \tag{70}$$

**Kernel orthogonalization for a collection of correlated white noise stimuli** We indicate the kernel orthogonalization procedure with respect to the set of correlated white noises,  $s_{ij}(t)$ . This procedure leads to the mutually-orthogonal Wiener functionals that can be used for the Wiener-series analog of formula (10).

First, any zeroth-order Wiener functional  $p_0[\vec{S}](t)$  is just a constant,

$$p_0[\vec{S}](t) = w_0,$$

for any collection  $\vec{S}$  of stimuli  $s_{vij}(t)$  or  $s_{ij}(t)$ . Any first-order Wiener functional

$$p_1[\vec{S}](t) = \sum_{i=1}^I \sum_{j=1}^J \int_0^\infty w_1^{ij}(\tau_1) S_{ij}(t - \tau_1) d\tau_1 + w_{1,0}$$

must satisfy the orthogonality condition  $\langle p_1[\vec{S}](t) p_0[\vec{S}](t) \rangle_{\vec{S}} = 0$  for any zeroth-order Wiener functional  $p_0[\vec{S}](t)$ , with  $\langle \cdot \rangle_{\vec{S}}$  again denoting the ensemble averaging over all realizations of the collection  $\vec{S}$  of the correlated Gaussian white noise stimuli  $s_{ij}(t)$ . Equation (69a) immediately implies  $w_{1,0} = 0$ , and so  $p_1[\vec{S}](t)$  must have the form

$$p_1[\vec{S}](t) = \sum_{i=1}^I \sum_{j=1}^J \int_0^\infty w_1^{ij}(\tau_1) S_{ij}(t - \tau_1) d\tau_1. \tag{71}$$

Moreover, by Eq. (70), which holds for the binary noises  $s_{vij}(t)$  as well as for the Gaussian noises  $s_{ij}(t)$ , and so all stimuli  $\vec{S}$  under our consideration, we see that we can subtract any function of  $t$  from all the kernels in Eq. (71) without changing the result. Therefore, we can assume with no loss of generality that

$$\sum_{i=1}^I \sum_{j=1}^J w_1^{ij}(\tau_1) = 0. \tag{72}$$

Likewise, any second-order Wiener functional  $p_2[\vec{S}](t)$  must be orthogonal to all zeroth- and first-order Wiener functionals and therefore must obey the equations  $\langle p_2[\vec{S}](t) p_0[\vec{S}](t) \rangle_{\vec{S}} = 0$  and  $\langle p_2[\vec{S}](t) p_1[\vec{S}](t) \rangle_{\vec{S}} = 0$  for all the realizations of the correlated Gaussian white

noises  $\vec{s}(t) = \{s_{ij}(t)\}$ . The correlation properties (69) and an argument similar to that leading to Eq. (72) imply the form

$$\begin{aligned} p_2[\vec{S}](t) = & \sum_{i,k=1}^I \sum_{j,l=1}^J \int_0^\infty \int_0^\infty w_2^{ik,jl}(\tau_1, \tau_2) S_{ij}(t - \tau_1) S_{kl}(t - \tau_2) d\tau_1 d\tau_2 \\ & - \frac{1}{IJ} \sum_{i=1}^I \sum_{j=1}^J \int_0^\infty w_2^{ii,jj}(\tau_1, \tau_1) d\tau_1 \\ & + \frac{1}{I^2 J^2} \sum_{i,k=1}^I \sum_{j,l=1}^J \int_0^\infty w_2^{ik,jl}(\tau_1, \tau_1) d\tau_1. \end{aligned}$$

Formula (70) now implies

$$\sum_{i=1}^I \sum_{j=1}^J w_2^{ik,jl}(\tau_1, \tau_2) = \sum_{k=1}^I \sum_{l=1}^J w_2^{ik,jl}(\tau_1, \tau_2) = 0,$$

so that finally

$$\begin{aligned} p_2[\vec{S}](t) = & \sum_{i,k=1}^I \sum_{j,l=1}^J \int_0^\infty \int_0^\infty w_2^{ik,jl}(\tau_1, \tau_2) S_{ij}(t - \tau_1) S_{kl}(t - \tau_2) d\tau_1 d\tau_2 \\ & - \frac{1}{IJ} \sum_{i=1}^I \sum_{j=1}^J \int_0^\infty w_2^{ii,jj}(\tau_1, \tau_1) d\tau_1. \end{aligned}$$

We proceed in the same vein at higher orders.

**First-order Lee–Schetzen formula for a collection of correlated white noise stimuli**

Here, we develop an analog of the Lee–Schetzen formula (9) for the firing rate in our correlated-noise case. The orthogonalization developed in the previous subsection is used to ensure that terms of order two and higher become orthogonal to the lower-order terms. Thus, the terms containing higher-order Wiener kernels will vanish when we correlate the firing rate  $m[\vec{s}](t, \Theta, \Phi)$  against a single copy of the binary noise,  $s_{ij}(t - \tau)$ . Using formulas (69), we compute

$$\begin{aligned} \langle m[\vec{s}](t, \Theta, \Phi) s_{ij}(t - \tau) \rangle_{\vec{S}} & = \sum_{k=1}^I \sum_{l=1}^J \int_0^\infty W_1^{k,l}(\tau_1, \Theta, \Phi) s_{kl}(t - \tau_1) s_{ij}(t - \tau) d\tau_1 \\ & = \frac{1}{IJ} W_1^{i,j}(\tau, \Theta, \Phi) - \frac{1}{(IJ)^2} \sum_{k=1}^I \sum_{l=1}^J W_1^{k,l}(\tau, \Theta, \Phi). \end{aligned} \tag{73}$$

This is a consistent singular linear system, since summing both sides on  $i$  and  $j$  yields the identity  $0 = 0$ , which follows from Eq. (70) for the left-hand side of the first line, and by simple algebra for the last line. The system’s singular nature implies that the solution

of Eq. (73) is not unique, and we can postulate a linear equation connecting the kernels  $W_1^{i,j}(\tau, \Theta, \Phi)$ . As in the previous subsection, with no loss of generality, we can use Eq. (72), that is,

$$\sum_{k=1}^I \sum_{l=1}^J W_1^{k,l}(\tau, \Theta, \Phi) = 0,$$

and we find for the first-order Wiener kernels the formulas

$$W_1^{i,j}(\tau, \Theta, \Phi) = IJ \langle m[\vec{s}](t, \Theta, \Phi) s_{ij}(t - \tau) \rangle_{\vec{s}}, \tag{74}$$

which is the same as Eq. (11a). Here, as in Eq. (69),  $\langle \cdot \rangle_{\vec{s}}$  denotes ensemble averaging over all possible realizations of the random noise collection  $\vec{s}$ .

**Relationship between the experimental data and Wiener kernels** We now discuss the relationship between the Volterra-Wiener representations of the firing rate on the one hand and the correlation functions (3) and the “interval-specific firing rate averages” (7) on the other. Since the expressions (3) and (7) involve the stimulus in a linear fashion, we are indeed allowed to focus on the first-order, that is linear, kernels, given by formula (74). By ergodicity, we have

$$W_1^{i,j}(\tau, \Theta, \Phi) = IJ \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T m[\vec{s}](t, \Theta, \Phi) \times s_{ij}(t - \tau) dt \tag{75}$$

for any given noise collection  $\vec{s}$  with the  $ij$ -th component  $s_{ij}$  [this is Eq. (11b)]. We can rewrite this equation in the form

$$W_1^{i,j}(\tau, \Theta, \Phi) = IJ \lim_{T \rightarrow \infty} \lim_{\nu \rightarrow 0} \frac{1}{2T\sqrt{\nu}} \int_{-T}^T m(t, \Theta, \Phi, \vec{\theta}, \vec{\phi}) \times \left[ \chi_{\theta_i, \phi_j}(t - \tau) - \frac{1}{IJ} \right] dt, \tag{76}$$

where  $\chi_{\theta_i, \phi_j}(t)$  is the indicator function of the set of intervals on which the angle  $\theta_i$  and the phase  $\phi_j$  were presented simultaneously. If we rewrite Eq. (3a) as

$$\mathcal{N}(\tau, \theta_i, \phi_j) = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T \chi_{\theta_i, \phi_j}(t - \tau) m(t, \Theta, \Phi, \vec{\theta}, \vec{\phi}) dt,$$

we find

$$\begin{aligned} W_1^{i,j}(\tau, \Theta, \Phi) &= \lim_{\nu \rightarrow 0} \frac{1}{\sqrt{\nu}} \left[ IJ \mathcal{N}(\tau, \theta_i, \phi_j) - \bar{m}(\Theta, \Phi) \right] \\ &= \lim_{\nu \rightarrow 0} \frac{1}{\sqrt{\nu}} \left[ N(\tau, \theta_i, \phi_j) - \langle N \rangle_{\tau, \theta^{(0)}, \phi^{(0)}} \right], \end{aligned} \tag{77}$$

with the “interval-specific firing rate average,”  $N(\tau, \theta_i, \phi_j)$ , defined in Eq. (7a). Formula (77) is the same as Eq. (12). Here,  $\bar{m}$  denotes the time-averaged firing rate, and  $\langle \cdot \rangle_{\tau, \theta^{(0)}, \phi^{(0)}}$  denotes averaging over all the arguments of the function  $N(\tau, \theta^{(0)}, \phi^{(0)})$ . The last line in Eq. (77) is true because, due to ergodicity, the time-averaged firing rate  $\bar{m}$  is independent of the presentation sequences  $\vec{\theta}$  and  $\vec{\phi}$ . Therefore,

$$\bar{m} = \overline{\langle m \rangle_{\vec{\theta}, \vec{\phi}}}, \tag{78}$$

where the overbar represents time averaging and  $\langle \cdot \rangle_{\vec{\theta}, \vec{\phi}}$  ensemble averaging over all the components of  $\vec{\theta}$  and  $\vec{\phi}$ . But, by Eq. (7a), the right-hand side of Eq. (78) clearly equals  $\langle N \rangle_{\tau, \theta^{(0)}, \phi^{(0)}}$ .

Likewise, we consider the correlation function  $\mathcal{M}(\tau, \theta_i)$  in Eq. (3b), corresponding to the process of finding the “preferred-phase averaged” firing rate  $\langle m(t, \Theta, \Phi) \rangle_{\Phi}$ . As explained at the end of Section 3, from the definitions (3), and in accordance with the fact that in the experiment the phases  $\phi_j$  are averaged out, we see that the correlation functions  $\mathcal{N}(\tau, \theta_i, \phi_j)$  and  $\mathcal{M}(\tau, \theta_i)$  are simply related by the equation

$$\mathcal{M}(\tau, \theta_i) = \sum_{j=1}^J \mathcal{N}(\tau, \theta_i, \phi_j).$$

Therefore, the first Wiener kernel corresponding to  $\mathcal{M}(\tau, \theta_i)$  is

$$\begin{aligned} W_1^i(\tau, \Theta) &= \frac{1}{J} \sum_{j=1}^J W_1^{i,j}(\tau, \Theta, \Phi) \\ &= \lim_{\nu \rightarrow 0} \frac{1}{\sqrt{\nu}} \left[ I \mathcal{M}(\tau, \theta_i) - \bar{m}(\Theta, \Phi) \right] \\ &= \lim_{\nu \rightarrow 0} \frac{1}{\sqrt{\nu}} \left[ M(\tau, \theta_i) - \bar{m}(\Theta, \Phi) \right] \\ &= \lim_{\nu \rightarrow 0} \frac{1}{\sqrt{\nu}} \left[ M(\tau, \theta_i) - \langle M \rangle_{\tau, \theta^{(0)}} \right], \end{aligned} \tag{79}$$

where  $M(\tau, \theta_i)$  is the interval-specific firing rate average (5b). This is Eq. (13).

**Comparison with previous results** We present a very brief comparison with the results of Ringach et al. (1997b), Nykamp and Ringach (2002), and restrict ourselves to the limit as  $\nu \rightarrow 0$ . The result of Ringach et al. (1997b) holds for neuronal models which consist of a linear filter followed by a static nonlinearity. This is indeed the case for the leaky feed-forward models (17) and (18). Since  $\int_{-\infty}^{\infty} G_{lgn}(t) dt = 0$ , the effect of the linear filter in Eq. (19) on the stimulus (1) is the same as if Eq. (1) consisted of the spatially mutually-orthogonal

gratings  $\mathcal{I}_0 \varepsilon v^{-1/2} \sin[\mathbf{k}(\theta_i) \cdot \mathbf{X} - \phi_j]$ . These are drawn at each time  $nv$  with uniform probability. The result of Ringach et al. (1997b) states that RTC reproduces, up to a factor, the (spatial) projection of the linear filter in Eq. (19) on this orthogonal set at the time  $\tau$ . We have already computed this projection when we calculated formula (19), and the projection coefficients equal  $(\mathcal{I}_0 \varepsilon v^{-1/2}/2) e^{-\frac{1}{4} K^2 \mu^2} f(\theta_i) G_{lgn}(\tau) \cos[\phi_j - \Delta(\theta_i)]$ , which is proportional to the first Wiener kernel (36). Thus, the first Wiener kernels in the theory presented in this paper are proportional to the projection coefficients in Ringach et al. (1997b). Averaging these coefficients over the phases  $\phi_j$  yields zero in agreement with the result presented at the end of our Section 4. The physiological interpretation is also the same as that at the end of that Section: spatial-phase averaging of receptive fields which are identical in every aspect except their preferred spatial phase erases their sensitivity (at least in the first approximation) due to the cancellations of the on- and off-regions. Finally, our formula (27) is in agreement with formulas (2) through (5) in Nykamp and Ringach (2002).

#### 4 Calculation of a probability density function

In this appendix, we compute the probability density function  $p_S(s, t, \theta^{(0)})$  (in formula (39)) of the random variable  $S$  in Eq. (37),

$$S = v^{-1/2} \sum_{n=-\infty}^{\infty} f(\theta^{(n)}) \mathcal{G}(t - nv) \cos \phi^{(n)}, \tag{80}$$

in which all the phases  $\phi^{(n)}$  and all the angles  $\theta^{(n)}$  save one,  $\theta^{(0)}$ , are random. Here,  $f(\theta^{(n)})$  is defined by formula (23), and  $\mathcal{G}(t)$  by Eq. (25). Clearly,  $S$  is an infinite sum

$$S = \sum_{n=-\infty}^{\infty} S^{(n)},$$

of independent individual random variables

$$S^{(n)} = v^{-1/2} f(\theta^{(n)}) \mathcal{G}(t - nv) \cos \phi^{(n)}, \tag{81}$$

in which both  $\theta^{(n)}$  and  $\phi^{(n)}$  are independent random variables distributed uniformly on the intervals  $[0, \pi]$  and  $[0, 2\pi]$ , respectively, if  $n \neq 0$ . For  $n = 0$ , in  $S^{(0)}$ , only  $\phi^{(0)}$  is an independent random variable distributed uniformly on the interval  $[0, 2\pi]$  while  $\theta^{(0)}$  is fixed.

**An exact formula for  $p_S(s, t, \theta^{(0)})$**  Let us first note that the probability density of the random variable  $X^{(n)} = \cos \phi^{(n)}$  is given by the formula

$$p_{X^{(n)}}(x) = \begin{cases} \frac{1}{\pi \sqrt{1-x^2}}, & x^2 < 1, \\ 0, & x^2 \geq 1. \end{cases} \tag{82}$$

The probability density of the random variable  $Y^{(n)} = f(\theta^{(n)})$  unfortunately cannot be computed explicitly, however, this does not matter here. Instead, we use the formula

$$p_{AB}(z) = \int_0^\infty p_A\left(\frac{z}{y}\right) p_B(y) \frac{dy}{y}, \tag{83}$$

which is valid if the variable  $B$  is nonnegative, to compute the probability density of the random variable  $S^{(n)} = v^{-1/2} f(\theta^{(n)}) \mathcal{G}(t - nv) \cos \phi^{(n)}$ , for  $n \neq 0$ , which is

$$p_{S^{(n)}}(z) = \frac{1}{\pi} \int_0^\pi p_{X^{(n)}}\left(\frac{z}{v^{-1/2} f(\theta^{(n)}) \mathcal{G}(t - nv)}\right) \times \frac{d\theta^{(n)}}{v^{-1/2} f(\theta^{(n)}) |\mathcal{G}(t - nv)|}. \tag{84}$$

In this calculation, we also used the facts that

$$p_{\theta^{(n)}}(\theta) = \begin{cases} \frac{1}{\pi}, & 0 \leq \theta < \pi, \\ 0, & \text{otherwise,} \end{cases}$$

and that the probability density of  $S^{(n)}$  is independent of whether it contains  $\mathcal{G}(t - nv)$  or  $|\mathcal{G}(t - nv)|$  as a factor.

The characteristic function for the variable  $S^{(n)}$  is the Fourier transform of the probability density (84), which, after a scaling of the integration variable and interchanging of the order of integration, becomes

$$\Phi_n(\kappa) = \frac{1}{\pi} \int_0^\pi d\theta^{(n)} \int_{-\infty}^\infty du e^{i\kappa u v^{-1/2} f(\theta^{(n)}) \mathcal{G}(t - nv)} p_{X^{(n)}}(u). \tag{85}$$

The inner integral can be written in the form

$$\frac{2}{\pi} \int_0^1 \frac{\cos \alpha u}{\sqrt{1-u^2}} du = J_0(\alpha),$$

with  $\alpha = v^{-1/2} f(\theta^{(n)}) \mathcal{G}(t - nv)$ , and  $J_0$  being the Bessel function of order zero, so that

$$\Phi_n(\kappa) = \frac{1}{\pi} \int_0^\pi J_0\left(\kappa v^{-1/2} f(\theta^{(n)}) \mathcal{G}(t - nv)\right) d\theta^{(n)}, \quad n \neq 0. \tag{86}$$

In a similar fashion, since  $\theta^{(0)}$  is fixed, the characteristic function of the zeroth term in the sum (37) can be computed simply as

$$\Phi_0(\kappa) = J_0\left(\kappa v^{-1/2} f(\theta^{(0)})\mathcal{G}(t)\right). \tag{87}$$

The characteristic function for the variable  $S$  in formula (37) is written as the infinite product

$$\begin{aligned} \Phi_S(\kappa) &= J_0\left(\kappa v^{-1/2} f(\theta^{(0)})\mathcal{G}(t)\right) \\ &\times \prod_{\substack{n=-\infty \\ n \neq 0}}^{\infty} \left\{ \frac{1}{\pi} \int_0^\pi J_0\left(\kappa v^{-1/2} f(\theta^{(n)})\mathcal{G}(t-nv)\right) d\theta^{(n)} \right\}. \end{aligned} \tag{88}$$

Note that, since, for all large positive  $n$ ,  $\mathcal{G}(t-nv) = 0$ , this product is only infinite and not bi-infinite. The probability density function for the variable  $S$  in formula (37) is given as

$$\begin{aligned} p_S(s, t, \theta^{(0)}) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} d\kappa e^{-i\kappa s} J_0\left(\kappa v^{-1/2} f(\theta^{(0)})\mathcal{G}(t)\right) \\ &\times \prod_{\substack{n=-\infty \\ n \neq 0}}^{\infty} \left\{ \frac{1}{\pi} \int_0^\pi J_0\left(\kappa v^{-1/2} f(\theta^{(n)})\mathcal{G}(t-nv)\right) d\theta^{(n)} \right\}. \end{aligned} \tag{89}$$

**Some elementary properties of the probability density  $p_S(s, t, \theta^{(0)})$**  First, note that, since  $|J_0(x)| \leq 1$  for all values of  $x$ , every characteristic function  $\Phi_n(\kappa)$ , given by Eq. (86), satisfies the same inequality,  $|\Phi_n(\kappa)| < 1$ . This implies that the  $\kappa$ -integral in the probability density (89) converges no worse than the Fourier transform of the Bessel function alone.

For the random variable

$$\Sigma = v^{-1/2} \sum_{n=-\infty}^{\infty} f(\theta^{(n)})\mathcal{G}(t-nv) \cos \phi^{(n)}, \tag{90}$$

in which now all the angles  $\theta^{(n)}$  and all the phases  $\phi^{(n)}$  are random, the probability density function is

$$\begin{aligned} p_\Sigma(s, t) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} d\kappa e^{-i\kappa s} \\ &\times \prod_{n=-\infty}^{\infty} \left\{ \frac{1}{\pi} \int_0^\pi J_0\left(\kappa v^{-1/2} f(\theta^{(n)})\mathcal{G}(t-nv)\right) d\theta^{(n)} \right\}, \end{aligned} \tag{91}$$

which is periodic in  $t$  with period  $v$ . Therefore, by Eq. (38), the average  $\langle M(t, \theta^{(0)}) \rangle_{\theta^{(0)}}$  of the interval-specific firing rate average  $M(t, \theta^{(0)})$  in Eq. (7b) is also periodic

in  $t$  with the same period. (This can also be shown directly.)

Also, for  $t < 0$ , we have  $\mathcal{G}(t) = 0$  so that  $J_0\left(\kappa v^{-1/2} f(\theta^{(0)})\mathcal{G}(t)\right) = 1$ . If  $t$  is sufficiently negative, then other terms around it also equal 1. Therefore, near such a  $t$ , the probability density (89) and the interval-specific firing rate average (7b) repeat themselves with the period  $v$ . The same is, approximately, true for large  $t \gg 0$ .

The above two properties of Eq. (7b) stand in contrast to the analogous properties of the correlation function  $\mathcal{M}(t, \theta^{(0)})$  in Eq. (3b). In particular, by Eq. (6b), the average  $\langle \mathcal{M}(t, \theta^{(0)}) \rangle_{\theta^{(0)}}$  is constant in time, as are  $\mathcal{M}(t < 0, \theta^{(0)})$  and  $\mathcal{M}(t \rightarrow \infty, \theta^{(0)})$ . Thus, even though Eq. (6b) shows that the interval-specific firing rate average  $M(t, \theta^{(0)})$  and an integer multiple of the correlation function  $\mathcal{M}(t, \theta^{(0)})$  differ by an  $\mathcal{O}(v)$  quantity, the correlation function  $\mathcal{M}(t, \theta^{(0)})$  appears more suitable to capture the average tuning dynamics of a V1 neuron. This is in agreement with a similar observation at the end of Section 2.

**The small- $v$  limit of the probability density  $p_S(s, t, \theta^{(0)})$**

This limit can be derived in a manner reminiscent of the proof of the central limit theorem and the related Edgeworth expansion; see for instance Chapter 16 of Feller (1966). We begin with the observation that  $v^{-1/2}\mathcal{G}(t) \approx v^{1/2}G_{\text{Ign}}(t)$  for small  $v$ , which follows from Eq. (25). We then use the formula

$$J_0(x) = \sum_{l=0}^{\infty} \frac{(-1)^l x^{2l}}{2^{2l} (l!)^2}$$

to expand

$$\begin{aligned} &J_0\left(\kappa v^{-1/2} f(\theta^{(0)})\mathcal{G}(t)\right) \\ &= 1 - \frac{\kappa^2 f^2(\theta^{(0)}) \mathcal{G}^2(t)}{4v} + \frac{\kappa^4 f^4(\theta^{(0)}) \mathcal{G}^4(t)}{64v^2} + \mathcal{O}(v^3) \end{aligned}$$

and

$$\begin{aligned} &\frac{1}{\pi} \int_0^\pi J_0\left(\kappa v^{-1/2} f(\theta^{(n)})\mathcal{G}(t-nv)\right) d\theta^{(n)} \\ &= 1 - \frac{\kappa^2 f_2 \mathcal{G}^2(t-nv)}{4v} + \frac{\kappa^4 f_4 \mathcal{G}^4(t-nv)}{64v^2} + \mathcal{O}(v^3), \end{aligned}$$

with  $f_2$  and  $f_4$  defined in Eqs. (32) and (41), respectively. These two expansions, together with Eq. (88), imply

$$\begin{aligned} \log \Phi_S(\kappa) &= \log \left( 1 - \frac{\kappa^2 f^2(\theta^{(0)}) \mathcal{G}^2(t)}{4v} \right. \\ &\quad \left. + \frac{\kappa^4 f^4(\theta^{(0)}) \mathcal{G}^4(t)}{64v^2} + \mathcal{O}(v^3) \right) \\ &\quad + \sum_{\substack{n=-\infty \\ n \neq 0}}^{\infty} \log \left( 1 - \frac{\kappa^2 f_2 \mathcal{G}^2(t-nv)}{4v} \right. \\ &\quad \left. + \frac{\kappa^4 f_4 \mathcal{G}^4(t-nv)}{64v^2} + \mathcal{O}(v^3) \right) \\ &= -\frac{\kappa^2 f^2(\theta^{(0)}) \mathcal{G}^2(t)}{4v} - \frac{\kappa^4 f^4(\theta^{(0)}) \mathcal{G}^4(t)}{64v^2} \\ &\quad - \frac{\kappa^2 f_2}{4v} \sum_{\substack{n=-\infty \\ n \neq 0}}^{\infty} \mathcal{G}^2(t-nv) + \frac{\kappa^4 (f_4 - 2f_2^2)}{64v^2} \\ &\quad \times \sum_{\substack{n=-\infty \\ n \neq 0}}^{\infty} \mathcal{G}^4(t-nv) + \mathcal{O}(v^3). \end{aligned} \tag{92}$$

We now express the time-dependent coefficients in Eq. (92) in terms of  $G_{lgn}(\cdot)$ . First, using Eq. (25), we approximate

$$\mathcal{G}(t) = vG_{lgn}(t) + \mathcal{O}(v^2). \tag{93}$$

This makes the first term in the fifth line of Eq. (92) of  $\mathcal{O}(v)$ , and the second of  $\mathcal{O}(v^2)$ , which we will neglect. We then compute that the sum

$$\mathcal{G}_2(t) = \sum_{n=-\infty}^{\infty} \mathcal{G}^2(t-nv) \tag{94}$$

satisfies the equation

$$\mathcal{G}_2(t) = 2 \sum_{n=0}^{\infty} \mathcal{F}(t+nv) \left[ \mathcal{F}(t+nv) - \mathcal{F}(t+(n+1)v) \right], \tag{95}$$

for all  $t$  satisfying  $0 \leq t \leq v$ . Here,  $\mathcal{F}(t)$  is defined as in Eq. (25). The functional form of  $\mathcal{F}(t)$ , as it follows from Eqs. (25) and (20), lets us Taylor expand Eq. (95) in the form

$$\begin{aligned} \mathcal{G}_2(t) &= -2v \sum_{n=0}^{\infty} \mathcal{F}(t+nv) \left[ \mathcal{F}'(t+nv) + \frac{v}{2} \mathcal{F}''(t+nv) \right. \\ &\quad \left. + \frac{v^2}{6} \mathcal{F}'''(t+nv) + \dots \right]. \end{aligned} \tag{96}$$

(This is because all the series involving the derivatives of  $\mathcal{F}$  converge and the formal  $\mathcal{O}(v^k)$  terms are indeed of that order.) We assume  $\zeta \geq 1$  in Eq. (20), or, more generally, at worst  $G_{lgn}(t) = \mathcal{O}(t)$  and so  $\mathcal{F}(t) = \mathcal{O}(t^2)$

for small  $t$ . We use integration by parts and push the lower integration limit from  $t$  to 0 to compute from Eq. (96) that

$$\mathcal{G}_2(t) = v \int_0^{\infty} G_{lgn}^2(\xi) d\xi + \mathcal{O}(v^3). \tag{97}$$

We conclude that this formula is true for all  $t$  since  $\mathcal{G}_2(t)$  is periodic in  $t$ .

Using Eqs. (93) and (97), we thus derive the first term in the sixth line of Eq. (92) to be

$$\begin{aligned} & -\frac{\kappa^2 f_2}{4v} [\mathcal{G}_2(t) - \mathcal{G}^2(t)] \\ &= -\frac{\kappa^2 f_2 \mathcal{G}_2(t)}{4v} + v \frac{\kappa^2 f_2 G_{lgn}^2(t)}{4} + \mathcal{O}(v^2) \\ &= -\frac{\kappa^2 f_2}{4} \int_0^{\infty} G_{lgn}^2(\xi) d\xi + v \frac{\kappa^2 f_2 G_{lgn}^2(t)}{4} + \mathcal{O}(v^2) \\ &= -\frac{\kappa^2}{2} \left( V_0 - v \frac{f_2 G_{lgn}^2(t)}{2} \right) + \mathcal{O}(v^2), \end{aligned}$$

where  $V_0$  is defined as in Eq. (32). Thus, the sum of the fifth line and the first term in the sixth line of Eq. (92) is

$$\begin{aligned} & -\frac{\kappa^2}{2} \left( V_0 + v \frac{[f^2(\theta^{(0)}) - f_2] G_{lgn}^2(t)}{2} \right) \\ & + \mathcal{O}(v^2) = -\frac{\kappa^2 V(t, \theta^{(0)})}{2} + \mathcal{O}(v^2), \end{aligned} \tag{98}$$

where  $V(t, \theta^{(0)})$  is the expression in Eq. (40). Finally

$$\begin{aligned} & \sum_{\substack{n=-\infty \\ n \neq 0}}^{\infty} \mathcal{G}^4(t-nv) \\ &= \sum_{n=-\infty}^{\infty} \mathcal{G}^4(t-nv) - \mathcal{G}^4(t) \\ &= v^3 \int_0^{\infty} G_{lgn}^4(\xi) d\xi + \mathcal{O}(v^4) - v^4 G_{lgn}^4(t) + \mathcal{O}(v^5) \\ &= v^3 \int_0^{\infty} G_{lgn}^4(\xi) d\xi + \mathcal{O}(v^4). \end{aligned} \tag{99}$$

Therefore, using Eqs. (92), (98), and (99), the characteristic function  $\Phi_S(\kappa)$  can be expressed as

$$\begin{aligned} \Phi_S(\kappa) &= \exp \left[ -\frac{\kappa^2 V(t, \theta^{(0)})}{2} + \kappa^4 v V_1 + \mathcal{O}(v^2) \right] \\ &= \exp \left[ -\frac{\kappa^2 V(t, \theta^{(0)})}{2} \right] [1 + \kappa^4 v V_1 + \mathcal{O}(v^2)], \end{aligned} \tag{100}$$



where  $V_1$  is defined as in Eq. (41), and the second line is valid for all fixed finite  $\kappa$ .

Using the formula

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-i\alpha x - \alpha^2/2} d\alpha = e^{-x^2/2},$$

the differentiation rule for Fourier transforms, Eqs. (100), (89), and the fact that large values of  $\kappa$  contribute negligible amounts to the integration in Eq. (89), we finally deduce Eq. (39).

### 5 Orientation tuning dynamics for a purely feed-forward model with no leakage

In this appendix, we concentrate on purely feed-forward models with no leakage, i.e.,  $g_L = 0$  in Eq. (15), and on *finite* presentation duration  $\nu$ . We show that, in a fairly large class of such models, averaging over the spatial phase wipes out the orientation tuning of the corresponding V1 neuron.

First, in the limit as  $\nu \rightarrow 0$ , we observe that formula (38) yields a completely untuned interval-specific firing rate average  $M(t, \theta^{(0)})$  for a purely feed-forward model (17) with no leakage or global inhibition when the LGN drive is modeled by the simple Gabor function (19). To show this fact, we first recall that in Eq. (38), the part of the integrand in the square brackets is non-negative. Furthermore, as discussed above, in order for the LGN input not to become negative, if there is no leakage, i.e.  $g_L = 0$ , or inhibition, the magnitude of the constant  $C$  must be chosen sufficiently large so that the values of the variable  $s$  for which the probability density function  $p_S(s, t, \theta^{(0)})$  is non-zero must all be contained in the interval  $C/D \leq s < \infty$ . Therefore, we can extend the integral in Eq. (38) to  $\int_{-\infty}^{\infty}$ . Replacing  $\phi^{(n)}$  by  $\pi - \phi^{(n)}$  in the random variable  $S$  in Eq. (37) shows that this variable has a symmetric range, and thus the function  $p_S(s, t, \theta^{(0)})$  is even in  $s$ . The implication from Eq. (38) is then that in the case of a purely feed-forward model without leakage,  $M(t, \theta^{(0)}) = -C$ , independent of the orientation angle  $\theta^{(0)}$ .

Since the result of the previous paragraph could be a consequence of the very specific and simple LGN drive model and the  $\nu \rightarrow 0$  limit, we now briefly show that it is instead a much more general phenomenon. In particular, we here assume a much more realistic LGN drive in the feed-forward model (17), namely, a feed-

forward convergence of inputs from individual LGN neurons as used in our integrate-and-fire simulations described above, as well as the large-scale V1 network computations of McLaughlin et al. (2000), Tao (2004). Such an LGN drive is given by the formula

$$g_{lgn}(t) = \sum_{q=1}^Q g_{lgn}^q(t) = \sum_{q=1}^Q \left\{ R_B \pm \int_{\mathbb{R}^2} dX dY \int_{-\infty}^t ds \times K_{lgn}(|\mathbf{X} - \mathbf{X}_q|) G_{lgn}(t-s) \mathcal{I}(\mathbf{X}, s) \right\}^+, \quad (101)$$

where  $\mathbf{X} = (X, Y)$ ,  $R_B$  is the LGN neurons' background firing rate,  $G_{lgn}(t)$  is the LGN time-response kernel [for instance, as given by Eq. (20)],  $\mathcal{I}(\mathbf{X}, s)$  is the stimulus (1),  $K_{lgn}$  is the *isotropic* spatial kernel for the LGN neurons,  $\mathbf{X}_q$  is the receptive field center of the  $q$ -th LGN neuron that feeds into the V1 neuron in question, and  $\{a\}^+ = \max\{a, 0\}$ . The sign  $\pm$  corresponds to neurons with "on" and "off" receptive fields, whose firing is triggered by either the presence or absence of a bright stimulus in their receptive field center, respectively. Each term in the sum on the right-hand side of Eq. (101) represents the output of a single LGN neuron. Note that the number  $Q$  of LGN neurons feeding into it generally varies from one V1 neuron to the next (Tanaka 1985; Reid and Alonso 1995; Alonso et al. 2001; Ringach 2002; Tao et al. 2004). Moreover, the argument to be presented here will not change if we assume that the shapes and strengths of the LGN kernels, or the background firing rates, vary from one LGN neuron to the next, i.e., we can assume  $R_B$ ,  $K_{lgn}$ , and  $G_{lgn}$  are  $q$ -dependent.

Since  $g_{lgn}(t) \geq 0$  and  $g_L = 0$  in the case at hand, we can omit the thresholding in Eq. (17) and write

$$m(t) = \left( \frac{V_E}{V_T} - 1 \right) g_{lgn}(t) = \left( \frac{V_E}{V_T} - 1 \right) \sum_{q=1}^Q g_{lgn}^q(t), \quad (102)$$

where  $g_{lgn}^q(t)$  are the individual bracketed terms in the sum (101). Using Eq. (7b), we compute the interval-specific firing rate average  $M(t, \theta_i)$  to be

$$M(t, \theta_i) = \left( \frac{V_E}{V_T} - 1 \right) \sum_{q=1}^Q \left\langle g_{lgn}^q \right\rangle_{\theta, \bar{\phi}, \theta^{(0)} = \theta_i} (t, \theta_i), \quad (103)$$

where the averaging is as in Eq. (7b).

Let us now consider any individual term on the right-hand side of Eq. (103). First, for any arbitrary radially-symmetric kernel  $K_{lgn}$ , we compute

$$\begin{aligned}
 g_{lgn}^q(t) &= \left\{ R_B \pm I_0 \varepsilon v^{-\frac{1}{2}} \sum_{n=-\infty}^{\infty} \mathcal{G}(t-nv) \right. \\
 &\quad \times \left. \int_{\mathbb{R}^2} dXdY K_{lgn}(|\mathbf{X}-\mathbf{X}_q|) \sin[\mathbf{k}(\theta^{(n)}) \cdot \mathbf{X}-\phi^{(n)}] \right\}^+ \\
 &= \left\{ R_B \pm I_0 \varepsilon v^{-\frac{1}{2}} \sum_{n=-\infty}^{\infty} \mathcal{G}(t-nv) \sin[\mathbf{k}(\theta^{(n)}) \cdot \mathbf{X}_q-\phi^{(n)}] \right. \\
 &\quad \times \left. \int_{\mathbb{R}^2} dU dV K_{lgn}(|\mathbf{Y}|) \cos[\mathbf{k}(\theta^{(n)}) \cdot \mathbf{Y}] \right\}^+ \\
 &= \left\{ R_B \pm I_0 \varepsilon v^{-\frac{1}{2}} \kappa_{lgn}(k) \sum_{n=-\infty}^{\infty} \mathcal{G}(t-nv) \right. \\
 &\quad \times \left. \sin[\mathbf{k}(\theta^{(n)}) \cdot \mathbf{X}_q-\phi^{(n)}] \right\}^+. \tag{104}
 \end{aligned}$$

Here,  $(U, V) = \mathbf{Y} = \mathbf{X} - \mathbf{X}_q$ ,  $\mathcal{G}(t)$  is as in Eq. (25), and

$$\kappa_{lgn}(k) = \int_{\mathbb{R}^2} d\xi d\eta K_{lgn}(\sqrt{\xi^2 + \eta^2}) \cos k\xi \tag{105a}$$

$$= \int_{\mathbb{R}^2} dU dV K_{lgn}(|\mathbf{Y}|) \cos[\mathbf{k}(\theta^{(n)}) \cdot \mathbf{Y}] \tag{105b}$$

is seen to be independent of the orientation angle  $\theta^{(n)}$  by rotating the coordinate system in Eq. (105a) so that the  $\xi$ -axis is parallel to the vector  $\mathbf{k}(\theta^{(n)})$ .

To find  $\langle g_{lgn}^q \rangle_{\vec{\theta}, \vec{\phi}, \theta^{(0)}=\theta_i}$ , we must average the last line in the expression (104) over all phases  $\phi^{(n)}$  and over all orientations  $\theta^{(n)}$  except  $\theta^{(0)}$ . However, due to the Lemma in Appendix 1, the distribution of the variable  $\mathbf{k}(\theta^{(0)}) \cdot \mathbf{X}_q - \phi^{(0)}$ , shifted into the interval  $[0, 2\pi)$ , is the same as the distribution of  $\phi^{(0)}$ , i.e., uniform. Thus, the averaging over the phase  $\phi^{(0)}$  alone erases the dependence of  $\langle g_{lgn}^q \rangle_{\vec{\theta}, \vec{\phi}, \theta^{(0)}=\theta_i}$  on the value of the orientation  $\theta^{(0)}$ . By Eq. (103), the interval-specific firing rate average  $M(t, \theta^{(0)})$  is therefore also untuned in this case, and by Eq. (6b), so is the correlation function  $\mathcal{M}(t, \theta^{(0)})$ .

### 6 Integrate-and-fire ring model

For the simulations presented in Fig. 2, we use a network model of  $n = 1024$  conductance-based, integrate-and-fire, point neurons, laid out on a ring with 32 different preferred orientations  $\Theta_i$  and 8 preferred spatial

phases  $\Phi_j$ . These are conferred by convergent LGN inputs, which are modeled by Poisson spike trains, with each train modulated by a firing-rate function of the form (101). There are three excitatory neurons and one inhibitory neuron for each combination of a preferred orientation and spatial phase. The lateral, cortico-cortical connections are all-to-all, with the strengths of the connections depending on the difference  $\Theta_i - \Theta_j$  between the preferred orientations, and independent of the preferred spatial phases of the neurons. The details of the model are as follows.

The membrane potential of the  $i$ -th model neuron follows the time evolution

$$\begin{aligned}
 \frac{d}{dt} v_P^{(i)} &= -g_L (v_P^{(i)} - V_r) - g_{PE}^{(i)}(t) (v_P^{(i)} - V_E) \\
 &\quad - g_{PI}^{(i)}(t) (v_P^{(i)} - V_I), \tag{106}
 \end{aligned}$$

where, as in the rest of this paper,  $P = E, I$  indicates whether the neuron is excitatory or inhibitory. The voltages  $V_E > V_r > V_I$  are the excitatory, rest, and inhibitory reversal potentials, and as in the main text, their values are set to  $V_E = 14/3$ ,  $V_r = 0$ , and  $V_I = -2/3$ , respectively. The firing threshold is set to  $V_T = 1$ . The time scale of the leakage conductance  $g_L$  is  $g_L^{-1} = 20$  ms.

The time-dependent conductances  $g_{PP'}^{(i)}(t)$  in the model (106) arise from the convergent LGN input and from network activity and have the form

$$g_{PE}^{(i)}(t) = S_{PE} \sum_{j=1}^n K_{PE}(\Theta_i - \Theta_j) \sum_k G_E(t - t_k^j) + g_{lgn}^{(i)}(t), \tag{107a}$$

$$g_{PI}^{(i)}(t) = S_{PI} \sum_{j=1}^n K_{PI}(\Theta_i - \Theta_j) \sum_k G_I(t - T_k^j) + g_{inh}(t). \tag{107b}$$

Here, the coupling coefficients  $S_{PP'}$  have the same meaning as in Eq. (16). In Fig. 2, they are set to  $S_{EE} = 1.0$ ,  $S_{EI} = 2.0$ ,  $S_{IE} = 6.0$ , and  $S_{II} = 2.0$ . The spatial coupling kernels  $K_{PE}(\theta)$  and  $K_{PI}(\theta)$  are taken to be Gaussians of the form (114) with both  $\sigma_P = \pi/8$ . The temporal kernels  $G_P$  have the form

$$G_P(t) = \frac{1}{\tau_{Pd} - \tau_{Pr}} \left[ \exp\left(-\frac{t}{\tau_{Pd}}\right) - \exp\left(-\frac{t}{\tau_{Pr}}\right) \right] \text{ for } t \geq 0,$$

and vanish for  $t < 0$ . Here,  $\tau_{Ed} = 5$  ms,  $\tau_{Id} = 10$  ms, and both  $\tau_{Pr} = 1$  ms. The symbol  $t_k^j$  ( $T_k^j$ ) indicates the time of the  $k$ -th spike of the  $j$ -th excitatory (inhibitory) neuron.

The LGN driving term in Eq. (107a),  $g_{lgn}(t)$ , is a Poisson spike train with spike strengths 0.01 and the

spiking rate modulated as the sum (101) of convergent inputs from a distribution of 30 identical “on” and “off” LGN cells, described in Appendix 5. The forms of the spatial and temporal kernels,  $K_{lgn}$  and  $G_{lgn}$ , in Eq. (101) are the same as those used in McLaughlin et al. (2000), Wielaard et al. (2001). In particular,

$$K_{lgn}(x) = \frac{a}{\pi \sigma_a^2} \exp\left(-\frac{x^2}{\sigma_a^2}\right) - \frac{b}{\pi \sigma_b^2} \exp\left(-\frac{x^2}{\sigma_b^2}\right),$$

with  $\sigma_a = 0.066^\circ$ ,  $\sigma_b = 0.093^\circ$ ,  $a = 1$ , and  $b = 0.74$ . The temporal kernel  $G_{lgn}(t)$  is given by a constant multiple of formula (20) with  $\zeta = 5$ ,  $\tau_{lgn} = 3 \text{ ms}$ ,  $\alpha = 1$ , and  $\beta = 5/3$ . The overall constant in  $G_{lgn}(t)$  is determined so that the peak firing rate for each LGN cell driven with a full-contrast drifting grating stimulus equals 100 Hz. The term  $g_{inh}(t)$  in Eq. (107b) models global random inhibitory inputs, which are taken to be Poisson trains with rate 500 Hz and spike strengths 0.1. The input stimulus used in the simulation is the sequence of gratings (1) with randomly selected orientation angles and spatial phases. The integrate-and-fire system governed by Eqs. (106) and (107) is integrated numerically by the method of Shelley and Tao (2001).

In our simulations with this spiking integrate-and-fire model, the spike histograms  $\mathcal{M}(\tau, \theta_i)$  of delay times  $\tau$  and orientations  $\theta_i$ , and the corresponding probability density functions  $\mathcal{P}(\tau, \theta_i)$ , are constructed in the same way as from the experimental data, which is described at the beginning of Section 2.

### 7 Solution of the model of Mexican-hat tuning curves

In this appendix, we present the details of the calculations outlined in Section 5. As described there, we first drastically simplify the thresholded linear model (14) by dropping the thresholding, setting the preferred phase of the neuron in the LGN drive (19) to  $\Phi = 0$ , fixing all the phases in the stimulus at  $\phi^{(n)} \equiv 0$ , and dropping the phase averaging. By considering only the ring architecture, that is, replacing the position  $\mathbf{x}$  in the cortex by the preferred angle  $\Theta$ , the spatial convolution in Eq. (14) becomes a convolution in the angle variable only,

$$K * F(\theta) = \frac{1}{\pi} \int_{-\pi/2}^{\pi/2} K(\theta - \theta') F(\theta') d\theta'.$$

Thus, the model (14) reduces to the coupled linear ring model

$$m_P(t, \Theta) = f_{lgn}(t, \Theta) + \sum_{P'} C_{PP'} K_{PP'} * G_{P'} * m_{P'}(t, \Theta),$$

$$P, P' = E, I. \tag{108}$$

The simplified LGN drive (22) becomes

$$f_{lgn}(t, \Theta) = A + B \sum_{n=-\infty}^{\infty} f(\theta^{(n)} - \Theta) \mathcal{G}(t - nv). \tag{109}$$

Here,  $A = -g_L + (V_E/V_T - 1)R_B$ ,  $B = (V_E/V_T - 1) \mathcal{I}_{0E} v^{-1/2} e^{-K^2 \mu^2/4} / 2$ , with  $K$  and  $\mu$  described after formula (19). (Since we are about to scale  $A$  and  $B$  out of the model, we do not specify the values of any constants involved in them.) The orientation-angle profile  $f(\theta) \equiv f(\theta, 0)$  is given by Eq. (23), and the time course  $\mathcal{G}(t)$  by Eq. (25). We choose  $\zeta = 0$  in Eqs. (20) and (25) for the simplest form of  $\mathcal{G}(t)$ . With no loss of generality, we can take the preferred angle  $\Theta$  of the neuron under study to be  $\Theta = 0$ . Moreover, in this idealized model,  $f(\theta)$  can be simplified even further without affecting the qualitative nature of the results; we choose it to be the periodic version of the Gaussian of the form

$$f(\theta) = \frac{1}{\sqrt{\pi} \sigma_{lgn}} \exp\left(-\frac{\theta^2}{\sigma_{lgn}^2}\right). \tag{110}$$

After ensemble averaging Eqs. (108) over the orientation angles  $\theta^{(n)}$ 's, except  $\theta^{(0)}$ , using formula (7b), we find the equations for the excitatory and inhibitory interval-specific firing rate averages  $M_P(t, \theta^{(0)})$ ,  $P = E, I$ , [see Eqs. (7b) and (5b)] to be

$$M_P(t, \theta^{(0)}) = A + B[f(\theta^{(0)}) - \langle f \rangle_\theta] \mathcal{G}(t) + \sum_{P'} C_{PP'} K_{PP'} * G_{P'} * M_{P'}(t, \theta^{(0)}). \tag{111}$$

Normalizing so that  $K_{PP'} * G_{P'} * 1 = 1$  and dropping the superscript 0 from  $\theta^{(0)}$ , we linearly shift and rescale  $M_P(t, \theta) = \alpha_P + \beta_P \tilde{M}_P(t, \theta)$ . Choosing

$$\alpha_P = \frac{A(1 - C_{PP} + C_{PQ})}{(1 - C_{EE})(1 - C_{II}) - C_{EI}C_{IE}},$$

$$\beta_P = B, \quad P, Q = E, I, \quad P \neq Q, \tag{112}$$

and dropping the tilde in  $\tilde{M}_P(t, \theta)$ , we derive the equations for the (rescaled) interval-specific firing rate average  $M_P(t, \theta^{(0)})$  to be

$$M_P(t, \theta) = [f(\theta) - \langle f \rangle_\theta] \mathcal{G}(t) + \sum_{P'} C_{PP'} K_{PP'} * G_{P'} * M_{P'}(t, \theta),$$

$$P, P' = E, I, \tag{113}$$

which is exactly Eq. (45).

For this model, we choose the angular cortical kernel to be a periodic version of the Gaussian of the form

$$K_{PP'}(\theta) = \frac{1}{\sqrt{\pi} \sigma_{P'}} \exp\left(-\frac{\theta^2}{\sigma_{P'}^2}\right), \tag{114}$$

where  $\sigma_P$ ,  $P = E, I$ , denotes the angular extents of the excitatory and inhibitory cortical connections, respectively. For the temporal cortical kernel we choose the function

$$G_P(t) = \begin{cases} 0, & t \leq 0, \\ \frac{1}{\tau_P} \exp\left(-\frac{t}{\tau_P}\right), & t > 0. \end{cases} \tag{115}$$

Equations (113) form a pair of linear convolution equations, which we solve by a combination of the Laplace transform in  $t$  and Fourier series in  $\theta$ . In particular, for the  $n$ -th Fourier coefficient of the Laplace-transformed (rescaled) interval-specific firing rate averages,  $\hat{M}_{P,n}(\lambda)$ , we find a simple set of two equations

$$\hat{M}_{P,n}(\lambda) = f_n \hat{G}(\lambda) + \sum_{P'} C_{PP'} K_{PP',n} \hat{G}_{P'}(\lambda) \hat{M}_{P',n}(\lambda), \tag{116}$$

$P, P' = E, I,$

where  $K_{PP',n}$  and  $f_n$  are the  $n$ -th Fourier coefficients of the spatial cortical kernel  $K_{PP'}(\theta)$  and the spatial profile  $f(\theta) - \langle f \rangle_\theta$ , and  $\hat{G}_{P'}(\lambda)$  and  $\hat{G}(\lambda)$  are the Laplace transforms of the temporal cortical kernel  $G_{P'}(t)$  and the LGN time course  $G(t)$ , respectively (note that  $f_0 = 0$ ).

For the spatial kernel  $K_{PP'}(\theta)$ , the periodic version of the Gaussian of the form (114), we find that its  $n$ -th Fourier coefficient is given by

$$K_{PP',n} = \frac{1}{\pi} \exp\left(-\frac{n^2 \sigma_P^2}{4}\right). \tag{117}$$

For the temporal cortical kernel  $G_P(t)$  in Eq. (115) we find

$$\hat{G}_P(\lambda) = \frac{1}{\lambda + 1/\tau_P}. \tag{118}$$

As mentioned in Section 5, the temporal LGN kernel is (20) with  $\zeta = 0$ . This implies

$$\hat{G}(\lambda) = \tau_{lgn}(1 - e^{-\nu\lambda}) \left( \frac{1}{\lambda + \beta/\tau_{lgn}} - \frac{1}{\lambda + \alpha/\tau_{lgn}} \right). \tag{119}$$

Finally, for the spatial profile  $f(\theta)$  in Eq. (110), we compute

$$f_n = \exp\left(-\frac{n^2 \sigma_{lgn}^2}{4}\right), \quad n \neq 0. \tag{120}$$

The solution of Eq. (116) is

$$\hat{M}_{P,n}(\lambda) = \frac{(\lambda + 1/\tau_Q) [\lambda + (1 + C_{PQ} K_{PQ,n} - C_{QQ} K_{QQ,n})/\tau_P]}{(\lambda - \lambda_{1,n})(\lambda - \lambda_{2,n})} f_n \hat{G}(\lambda), \tag{121}$$

$Q \neq P$

where  $\lambda_{1,n}$  and  $\lambda_{2,n}$  are given by

$$\lambda_{1,n}, \lambda_{2,n} = \frac{\lambda_{E,n} + \lambda_{I,n}}{2} \pm \sqrt{\frac{(\lambda_{E,n} - \lambda_{I,n})^2}{4} + \frac{C_{IE} K_{IE,n} C_{EI} K_{EI,n}}{\tau_E \tau_I}} \tag{122}$$

with

$$\lambda_{P,n} = \frac{C_{PP} K_{PP,n} - 1}{\tau_P}, \quad P = E, I. \tag{123}$$

Note that  $C_{EE}, C_{IE} \geq 0$  and  $C_{II}, C_{EI} \leq 0$ , so that  $C_{IE} K_{IE,n} C_{EI} K_{EI,n} / \tau_E \tau_I \leq 0$ . Note also that  $\hat{M}_{P,0}(\lambda) = 0$  because  $f_0 = 0$ .

The roots  $\lambda_{j,n}$ ,  $j = 1, 2$ , are either real and equidistant from their average  $(\lambda_{E,n} + \lambda_{I,n})/2$ , or complex with the real part  $(\lambda_{E,n} + \lambda_{I,n})/2$ . In the former case, the farthest  $\lambda_{1,n}$  and  $\lambda_{2,n}$  can be away from  $(\lambda_{E,n} + \lambda_{I,n})/2$  is  $(\lambda_{I,n} - \lambda_{E,n})/2$ . In the latter case, there is an additional frequency in the system. If one of the roots  $\lambda_{j,n}$ ,  $j = 1, 2$ , has a positive real part, there will be an instability in the solutions  $M_P(t, \theta)$ . This happens for both real and complex  $\lambda_{j,n}$  if  $(\lambda_{E,n} + \lambda_{I,n})/2 > 0$ , and, additionally for real ones, if  $(\lambda_{E,n} + \lambda_{I,n})/2 < 0$  but  $\lambda_{E,n} \lambda_{I,n} < C_{IE} K_{IE,n} C_{EI} K_{EI,n} / \tau_E \tau_I$ . This exhausts the possible bifurcation structure of the solutions that is due to the values of the cortical coupling constants, spatial kernels, and time scales.

The  $n$ -th Fourier coefficient,  $M_{P,n}(t)$ , of the tuning function  $M_P(t, \theta)$  can now be evaluated by using residue calculus. In order to do so, we define the function

$$\hat{\Gamma}_{P,n}(\lambda) = f_n \tau_{lgn} \frac{(\lambda + 1/\tau_Q) [\lambda + (1 + C_{PQ} K_{PQ,n} - C_{QQ} K_{QQ,n})/\tau_P]}{(\lambda - \lambda_{1,n})(\lambda - \lambda_{2,n})} \times \left( \frac{1}{\lambda + \beta/\tau_{lgn}} - \frac{1}{\lambda + \alpha/\tau_{lgn}} \right) \tag{124}$$

from the appropriate parts of formulas (121) and (119). Then,

$$M_{P,n}(t) = \Gamma_{P,n}(t) - \Gamma_{P,n}(t - \nu), \tag{125}$$

where  $\Gamma_{P,n}(t)$  is the inverse Laplace transform of the function  $\hat{\Gamma}_{P,n}(\lambda)$ . This is due to the well known

property of Laplace transforms,  $\widehat{F(t-v)}(\lambda) = e^{\lambda v} \widehat{F}(\lambda)$ . For  $t \geq 0$ , we compute  
 Moreover

$$\Gamma_{P,n}(t) = \begin{cases} 0, & t < 0, \\ \text{sum of the residues of } \widehat{\Gamma}_{P,n}(\lambda)e^{\lambda t}, & t \geq 0. \end{cases} \tag{126}$$

$$\Gamma_{P,n}(t) = \frac{f_n \tau_{lgn}}{\tau_E \tau_I} \left\{ \sum_{\substack{j=1 \\ i \neq j}}^2 \frac{\tau_{lgn}(\alpha - \beta)(1 + \tau_P \lambda_j) (1 + C_{PQ} K_{PQ,n} - C_{QQ} K_{QQ,n} + \tau_Q \lambda_j)}{(\lambda_j - \lambda_i)(\alpha + \tau_{lgn} \lambda_j)(\beta + \tau_{lgn} \lambda_j)} e^{\lambda_j t} \right. \\ \left. - \frac{(\tau_{lgn} - \tau_P \alpha) [\tau_{lgn}(1 + C_{PQ} K_{PQ,n} - C_{QQ} K_{QQ,n}) - \tau_Q \alpha]}{(\alpha + \lambda_1 \tau_{lgn})(\alpha + \lambda_2 \tau_{lgn})} e^{-\alpha t / \tau_{lgn}} \right. \\ \left. + \frac{(\tau_{lgn} - \tau_P \beta) [\tau_{lgn}(1 + C_{PQ} K_{PQ,n} - C_{QQ} K_{QQ,n}) - \tau_Q \beta]}{(\beta + \lambda_1 \tau_{lgn})(\beta + \lambda_2 \tau_{lgn})} e^{-\beta t / \tau_{lgn}} \right\}, \tag{127}$$

with  $P, Q = E, I$  and  $Q \neq P$ . The interval-specific firing rate average  $M_P(t, \theta)$  can now be computed from formulas (125) and (127) as

$$M_P(t, \theta) = \sum_{n=-\infty}^{\infty} M_{P,n}(t) e^{2in\theta}. \tag{128}$$

From Eq. (128), we can easily calculate the correlation function  $\mathcal{M}_P(t, \theta)$  via the formula (6b). After first rescaling  $\mathcal{M}_P(t, \theta) \rightarrow \alpha_P + \beta_P I \tilde{\mathcal{M}}_P(t, \theta)$ , with the constants  $\alpha_P$  and  $\beta_P$  as in Eq. (112), this calculation replaces all the exponential factors of the form  $e^{\gamma t}$  in Eq. (127) by the functions

$$\Lambda(t, \gamma, v) = \begin{cases} 0, & t < -v, \\ (e^{\gamma(t+v)} - 1) / \gamma, & -v \leq t < 0, \\ e^{\gamma t} (e^{\gamma v} - 1) / \gamma, & t \geq 0. \end{cases} \tag{129}$$

After dropping the tilde, the rescaled, orientation-tuned part of the correlation function,  $\mathcal{M}_P(t, \theta)$ , is then described by a Fourier series analogous to Eq. (128).

Note that formula (129) yields only a continuous solution  $\mathcal{M}_P(t, \theta)$ , not a continuously differentiable one. This can be traced back to the choice of  $\zeta = 0$  in the temporal LGN kernel given by Eqs. (20) and (25), which also makes this kernel only continuous. This lack of smoothness appears to be the source of the kinks at  $t_{\min} = v$  in Fig. 6.

## References

- Aguera y Arcas, B., Fairhall, A. L., & Bialek, W. (2003). Computation in a single neuron: Hodgkin and Huxley revisited. *Neural Computation*, 15, 1715–1794.
- Alitto, H. J., Weyand, T. G., & Usrey, W. M. (2005). Distinct properties of stimulus-evoked bursts in the lateral geniculate nucleus. *Journal of Neuroscience*, 25(2), 514–523.
- Alonso, J.-M., Usrey, W. M., & Reid, R. C. (2001). Rules of connectivity between geniculate cells and simple cells in cat primary visual cortex. *Journal of Neuroscience*, 21, 4002–4015.
- Ben-Yishai, R., Bar-Or, R., & Sompolinski, H. (1995). Theory of orientation tuning in the visual cortex. *Proceedings of the National Academy of Science USA*, 92, 3844–3848.
- Benardete, E., & Kaplan, E. (1999). The dynamics of primate M retinal ganglion cells. *Visual Neuroscience*, 16, 355–368.
- Blasdel, G. (1992a). Differential imaging of ocular dominance and orientation selectivity in monkey striate cortex. *Journal of Neuroscience*, 12, 3115–3138.
- Blasdel, G. (1992b). Orientation selectivity, preference, and continuity in the monkey striate cortex. *Journal of Neuroscience*, 12, 3139–3161.
- Bonhoeffer, T., & Grinvald, A. (1991). Iso-orientation domains in cat visual cortex are arranged in pinwheel like patterns. *Nature*, 353, 429–431.
- Borghuis, B. G., Perge, J. A., Vajda, I., van Wezel, R. J. A., van de Grind, W. A., & Lankheet, M. J. M. (2003). The motion reverse correlation (MRC) method: A linear systems approach in the motion domain. *Journal of Neuroscience Methods*, 123, 153–166.
- Bredfeldt, C. E., & Ringach, D. L. (2002). Dynamics of spatial frequency tuning in macaque V1. *Journal of Neuroscience*, 22(5), 1976–1984.
- Brodie, S. E., Knight, B. W., & Ratliff, F. (1978a). The response of the limulus retina to moving stimuli: A prediction by fourier synthesis. *Journal of General Physiology*, 72, 129–166.



- Brodie, S. E., Knight, B. W., & Ratliff, F. (1978b). The spatiotemporal transfer function of the limulus lateral eye. *Journal of General Physiology*, *72*, 167–202.
- Brugge, J. F., Reale, R. A., Jenison, R. L., & Schnupp, J. (2001). Auditory cortical spatial receptive fields. *Audiology and Neurootology*, *6*, 173–177.
- Celebrini, S., Thorpe, S., Trotter, Y., & Imbert, M. (1993). Dynamics of orientation coding in area V1 of the awake primate. *Visual Neuroscience*, *10*, 811–825.
- Chen, G., Dan, Y., & Li, C.-Y. (2005). Stimulation of non-classical receptive field enhances orientation selectivity in cat. *Journal of Physiology*, *564*(1), 233–243.
- Chichilnisky, E. J. (2001). A simple white noise analysis of neuronal light responses. *Network*, *12*, 199–213.
- Citron, M. C., & Emerson, R. C. (1983). White noise analysis of cortical directional selectivity in cat. *Brain Research*, *279*, 271–277.
- Cottaris, N. P., & DeValois, R. L. (1998). Temporal dynamics of chromatic tuning in macaque primary visual cortex. *Nature*, *395*(6705), 896–900.
- Daugman, J. G. (1985). Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters. *Journal of the Optical Society of America A*, *2*(7), 1160–1169.
- DeAngelis, G., Ohzawa, I., & Freeman, R. (1993). Spatiotemporal organization of simple-cell receptive fields in the cat's striate cortex. I. general characteristics and postnatal development. *Journal of Neurophysiology*, *69*, 1091–1117.
- DeAngelis, G., Ohzawa, I., & Freeman, R. (1995). Receptive-field dynamics in the central visual pathways. *Trends in Neurosciences*, *18*(10), 451–458.
- de Boer, E. (1967). Correlation studies applied to frequency resolution of the cochlea. *Journal of Auditory Research*, *7*, 209–217.
- de Boer, E., & de Jongh, H. R. (1978). On cochlear encoding: Potentialities and limitations of the reverse-correlation technique. *Journal of the Acoustical Society of America*, *63*(1):115–135.
- de Ruyter van Steveninck, R., & Bialek, W. (1988). Real-time performance of a movement-sensitive neuron in the blowfly visual system: Coding and information transfer in short spike sequences. *Proceedings of the Royal Society of London B*, *234*, 379–414.
- Douglas, R., Koch, C., Mahowald, M., Martin, K., & Suarez, H. (1995). Recurrent excitation in neocortical circuits. *Science*, *269*, 981–985.
- Dragoi, V., Sharma, J., Miller, E. K., & Sur, M. (2002). Dynamics of neuronal sensitivity in visual cortex and local feature discrimination. *Nature Neuroscience*, *5*(9), 883–891.
- Ernst, U. A., Pawelzik, K. R., Sahar-Pikielny, C., & Tsodyks, M. V. (2001). Intracortical origin of visual maps. *Nature Neuroscience*, *4*(4), 431–436.
- Feller, W. (1966). *An introduction to probability theory and its applications*, (Vol. 2). New York: Wiley.
- Felsen, G., Shen, Y.-S., Yao, H., Spor, G., Li, C., & Dan, Y. (2002). Dynamic modification of cortical orientation tuning mediated by recurrent connections. *Neuron*, *35*, 945–954.
- Franz, M. O., & Schölkopf, B. (2006). A unifying view of Wiener and Volterra theory and polynomial kernel regression. *Neural Computation*, *18*, 3097–3118.
- Giannakis, G. B., & Serpedin, E. (2001). A bibliography on nonlinear system identification. *Signal Processing*, *81*, 533–580.
- Gielen, C., van Gisbergen, J., & Ventic, A. (1981). Characterization of spatial and temporal properties of monkey LGN Y-cells. *Biological Cybernetics*, *40*, 157–170.
- Gillespie, D. C., Lampl, I., Anderson, J. S., & Ferster, D. (2001). Dynamics of the orientation-tuned membrane potential response in cat primary visual cortex. *Nature Neuroscience*, *4*, 1014–1019.
- Jones, J. P., & Palmer, L. A. (1987). The two-dimensional spatial structure of simple receptive fields in cat striate cortex. *Journal of Neurophysiology*, *58*, 1187–1211.
- Juusola, M., & French, A. S. (1997). Visual acuity for moving objects in first- and second-order neurons of the fly compound eye. *Journal of Neurophysiology*, *77*, 1487–1495.
- Kang, K., Shelley, M. J., & Sompolinsky, H. (2003). Mexican hats and pinwheels in visual cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *100*, 2848–2853.
- Klein, D. J., Simon, J. Z., Depireux, D. A., & Shamma, S. A. (2006). Stimulus-invariant processing and spectrotemporal reverse correlation in primary auditory cortex. *Journal of Computational Neuroscience*, *20*, 111–136.
- Klein, S. A., & Yatsui, S. (1979). Nonlinear systems analysis with non-gaussian white stimuli: General basis functionals and kernels. *IEEE Transactions on Information Theory*, *25*, 495–500.
- Lee, Y. N., & Schetzen, M. (1965). Measurement of the wiener kernels of a nonlinear system by cross correlation. *International Journal of Control*, *2*, 237–254.
- Lewis, E. R., Henry, K. R., & Yamada, W. M. (2002). Tuning and timing in the gerbil ear: Wiener-kernel analysis. *Heart Research*, *174*, 206–221.
- Maldonado, P., Godecke, I., Gray, C., & Bonhoeffer, T. (1997). Orientation selectivity in pinwheel centers in cat striate cortex. *Science*, *276*, 1551–1555.
- Marcelja, S. (1980). Mathematical description of responses of simple cortical cells. *Journal of the Optical Society of America*, *70*(11), 1297–1300.
- Marino, J., Schummers, J., Lyon, D. C., Schwabe, L., Beck, O., Wiesing, P., et al. (2005). Invariant computations in local cortical networks with balanced excitation and inhibition. *Nature Neuroscience*, *8*, 194–201.
- Marmarelis, P. Z., & McCann, G. D. (1973). Development and application of white noise modeling techniques for studies of insect visual nervous system. *Kybernetik*, *12*, 74–89.
- Marmarelis, P. Z., & Marmarelis, V. Z. (1978). *Analysis of physiological systems*. New York: Plenum.
- Mata, M. L., & Ringach, D. L. (2004). Spatial overlap of ON and OFF subregions and its relation to response modulation ratio in macaque primary visual cortex. *Journal of Neurophysiology*, *93*, 919–928.
- Matthews, V. J., & Sicuranza, G. L. (2002). *Polynomial signal processing*. New York: Wiley.
- Mazer, J., Vinje, W., McDermott, J., Schiller, P., & Gallant, J. (2002). Spatial frequency and orientation tuning dynamics in area V1. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(3), 1645–1650.
- Miller, L. M., Escabi, M. A., Read, H. L., & Schreiner, C. E. (2002). Spectrotemporal receptive fields in the lemniscal auditory thalamus and cortex. *Journal of Neurophysiology*, *87*, 516–527.
- Müller, J. R., Metha, A. B., Krauskopf, J., & Lennie, P. (2001). Information conveyed by onset transients in responses of striate cortical neurons. *Journal of Neuroscience*, *21*, 6978–6990.
- McCann, G. D. (1974). Nonlinear identification theory models for successive stages of visual nervous systems of flies. *Journal of Neurophysiology*, *37*, 869–895.

- McLaughlin, D., Shapley, R., Shelley, M., & Wielaard, J. (2000). A neuronal network model of macaque primary visual cortex (V1): Orientation selectivity and dynamics in the input layer 4C $\alpha$ . *Proceedings of the National Academy of Sciences of the United States of America*, *97*, 8087–8092.
- Naka, K.-I., Marmarelis, P. Z., & Chan, R. Y. (1975). Morphological and functional identifications of catfish retinal neurons. III. functional identification. *Journal of Neurophysiology*, *38*, 92–131.
- Nelson, J. I., & Frost, B. (1978). Orientation selective inhibition from beyond the classical receptive field. *Brain Research*, *139*, 359–365.
- Neri, P. (2004a). Estimation of nonlinear psychophysical kernels. *Journal of Vision*, *4*, 82–91.
- Neri, P. (2004b). Attentional effects on sensory tuning for single-feature detection and double-feature conjunction. *Vision Research*, *44*, 3053–3064.
- Neri, P., & Levi, D. M. (2006). Receptive versus perceptive fields from the reverse-correlation viewpoint. *Vision Research*, *46*, 2465–2474.
- Nishimoto, S., Arai, M., & Ohzawa, I. (2005). Accuracy of subspace mapping of spatiotemporal frequency domain visual receptive fields. *Journal of Neurophysiology*, *93*, 3524–3536.
- Nykamp, D. Q. (2003). Measuring linear and quadratic contributions to neuronal response. *Network*, *14*, 673–702.
- Nykamp, D. Q., & Ringach, D. L. (2002). Full identification of a linear-nonlinear system via cross-correlation analysis. *Journal of Vision*, *2*, 1–11.
- Ohzawa, I., Sclar, G., & Freeman, R. D. (1985). Contrast gain control in the cats visual system. *Journal of Neurophysiology*, *3*, 651–667.
- Palmer, L. A., & Nafziger, J. S. (2002). Effects of surround motion on receptive-field gain and structure in area 17 of the cat. *Visual Neuroscience*, *19*, 335–353.
- Pack, C. C., Conway, B. R., Born, R. T., & Livingstone, M. S. (2006). Spatiotemporal structure of nonlinear subunits in macaque visual cortex. *Journal of Neuroscience*, *26*(3), 893–907.
- Pei X. X., Vidyasagar, T. R., Volgushev, M., & Creutzfeldt, O. D. (1994). Receptive field analysis and orientation selectivity of postsynaptic potentials of simple cells in cat visual cortex. *Journal of Neuroscience*, *14*, 7130–7140.
- Pernberg, J., Jirrmann, K.-U., & Eysel, U. T. (1998). Structure and dynamics of receptive fields in the visual cortex of the cat (area 18) and the influence of GABAergic inhibition. *European Journal of Neuroscience*, *10*, 3596–3606.
- Pinter, R. B., & Nabet, B. (1992). *Nonlinear vision: Determination of neural receptive fields, function, and networks*. Boca Raton: CRC Press.
- Powers, R. K., Dai, Y., Bell, B. M., Percival, D. B., & Binder, M. D. (2006). Contributions to the input signal and prior activation history to the discharge behaviour of rat motoneurons. *Journal of Physiology*, *562*(3), 707–724.
- Pugh, M., Ringach, D., Shapley, R., & Shelley, M. (2000). Computational modeling of orientation tuning dynamics in V1 neurons. *Journal of Computational Neuroscience*, *8*, 143–159.
- Ramdyia, P., Reiter, B., & Engert, F. (2006). Reverse correlation of rapid calcium signals in the zebrafish optic tectum *in vivo*. *Journal of Neuroscience Methods*, *157*, 230–237.
- Recio-Spinoso, A., Temchin, A. N., Van Dijk, P., Fan, Y.-H., & Ruggero, M. A. (2005). Wiener-kernel analysis of responses to noise of chinchilla auditory-nerve fibers. *Journal of Neurophysiology*, *93*, 3615–3634.
- Recio-Spinoso, A., & Van Dijk, P. (2006). Analysis of responses to noise in the ventral cochlear nucleus using Wiener kernels. *Heart Research*, *216–217*, 7–18.
- Reid, R. C., & Alonso, J.-M. (1995). Specificity of monosynaptic connections from thalamus to visual cortex. *Nature*, *378*, 281–284.
- Reid, R. C., & Shapley, R. (1992). Spatial structure of cone inputs to receptive fields in primate lateral geniculate nucleus. *Nature*, *356*, 716–718.
- Reid, R. C., & Shapley, R. (2002). Space and time maps of cone photoreceptor signals in macaque lateral geniculate nucleus. *Journal of Neuroscience*, *22*(14), 6158–6175.
- Reid, R. C., Victor, J. D., & Shapley, R. M. (1997). The use of m-sequences in the analysis of visual neurons: Linear receptive field properties. *Visual Neuroscience*, *14*(6), 1015–1027.
- Rieke, F., Warland, D., de Ruyter van Steveninck, R., & Bialek, W. (1996). *Spikes: exploring the neural code*. Computational Neuroscience. Cambridge: MIT Press.
- Ringach, D. L. (1998). Tuning of orientation detectors in human vision. *Vision Research*, *38*, 963–972.
- Ringach, D. L. (2002). Spatial structure and symmetry of simple-cell receptive fields in macaque primary visual cortex. *Journal of Neurophysiology*, *88*, 455–463.
- Ringach, D., & Shapley, R. (2004). Reverse correlation in neurophysiology. *Cognitive Science*, *28*, 147–166.
- Ringach, D., Hawken, M., & Shapley, R. (1997a). Dynamics of orientation tuning in macaque primary visual cortex. *Nature*, *387*, 281–284.
- Ringach, D., Sapiro, G., & Shapley, R. (1997b). A subspace reverse correlation technique for the study of visual neurons. *Vision Research*, *37*, 2455–2464.
- Ringach, D. L., Bredfeldt, C. E., Shapley, R. M., & Hawken, M. J. (2002). Suppression of neural responses to nonoptimal stimuli correlates with tuning selectivity in macaque V1. *Journal of Neurophysiology*, *87*(14), 1018–1027.
- Ringach, D., Hawken, M., & Shapley, R. (2003). Dynamics of orientation tuning in macaque V1: The role of global and tuned suppression. *Journal of Neurophysiology*, *90*, 342–352.
- Rugh, W. J. (1981). *Nonlinear system theory: The Volterra/Wiener approach*. Baltimore: Johns Hopkins University Press.
- Rutkowski, R. G., Shackleton, T. M., Schnupp, J. W. H., Wallace, M. N., & Palmer, A. R. (2002). Spectrotemporal receptive field properties of single units in the primary, dorsocaudal and ventrorostral auditory cortex of the guinea pig. *Audiology and Neurotology*, *7*, 214–227.
- Schetzen, M. (1980). *The Volterra and Wiener theories of nonlinear systems*. New York: Wiley.
- Shapley, R., Hawken, M., & Ringach, D. L. (2003). Dynamics of orientation selectivity in the primary visual cortex and the importance of cortical inhibition. *Neuron*, *38*, 689–699.
- Sharon, D., & Grinvald, A. (2002). Dynamics and constancy in cortical spatiotemporal patterns of orientation processing. *Science*, *295*, 512–515.
- Shelley, M., & McLaughlin, D. (2002). Coarse-grained reduction and analysis of a network model of cortical response. I. drifting grating stimuli. *Journal of Computational Neuroscience*, *12*(2), 97–122.
- Shelley, M., & Tao, L. (2001). Efficient and accurate time-stepping schemes for integrate-and-fire neuronal networks. *Journal of Computational Neuroscience*, *11*, 111–119.
- Sillito, A. M., Kemp, J. A., Milson, J. A., & Berardi, A. (1980). A re-evaluation of the mechanisms underlying simple cell orientation selectivity. *Brain Research*, *194*, 517–520.

- Somers, D., Nelson, S., & Sur, M. (1995). An emergent model of orientation selectivity in cat visual cortical simple cells. *Journal of Neuroscience*, *15*, 5448–5465.
- Sompolinsky, H., & Shapley, R. M. (1997). New perspectives on the mechanisms for orientation selectivity. *Current Opinion in Neurobiology*, *7*(4), 514–522.
- Suder, K., Funke, K., Zhao, Y., Kerscher, N., Wennekers, T., & Wörgötter, F. (2002). Spatial dynamics of receptive fields in cat primary visual cortex related to the temporal structure of thalamocortical feedforward activity. *Experimental Brain Research*, *144*, 430–444.
- Svirskis, G., Dodla, R., & Rinzel, J. (2003). Subthreshold outward currents enhance temporal integration in auditory neurons. *Biological Cybernetics*, *89*, 333–340.
- Tanaka, K. (1985). Organization of geniculate inputs to visual cortical cells in the cat. *Vision Research*, *25*, 357–364.
- Tao, L., Cai, D., McLaughlin, D. W., Shelley, M. J., & Shapley, R. (2006). Orientation selectivity in visual cortex by fluctuation-controlled criticality. *Proceedings of the National Academy of Sciences of the United States of America*, *103*, 12911–12916.
- Tao, L., Shelley, M., McLaughlin, D., & Shapley, R. (2004). An egalitarian network model for the emergence of simple and complex cells in visual cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(1), 366–371.
- Vajda, I., Borghuis, B. G., van de Grind, W. A., & Lankheet, M. J. M. (2006). Temporal interactions in direction-selective complex cells of area 18 and the posteromedial lateral suprasylvian cortex (PMLS) of the cat. *Visual Neuroscience*, *23*, 233–246.
- van Kleef, J., James, A. C., & Stange, G. (2005). A spatiotemporal white noise analysis of photoreceptor responses to UV and green light in the dragonfly median ocellus. *Journal of General Physiology*, *126*(5), 481–497.
- Victor, J. D. (1979). Nonlinear systems analysis: Comparison of white noise and sum of sinusoids in a biological system. *Proceedings of the National Academy of Sciences of the United States of America*, *76*, 996–998.
- Victor, J. D. (1992). Nonlinear systems analysis in vision: Overview of kernel methods. In B. Nabet & R. B. Pinter (Eds.), *Nonlinear vision: Determination of neural receptive fields, function, and networks* (pp. 1–37). Boca Raton: CRC Press.
- Victor, J. D., & Knight, B. W. (1979). Nonlinear analysis with an arbitrary stimulus ensemble. *Quarterly of Applied Mathematics*, *37*, 113–136.
- Victor, J. D., & Shapley, R. M. (1978). The nonlinear pathway of Y ganglion cells in the cat retina. *Journal of General Physiology*, *74*, 671–689.
- Victor, J. D., & Shapley, R. M. (1980). Method of nonlinear analysis in the frequency-domain. *Biophysical Journal*, *29*(3), 459–483.
- Vislay-Meltzer, R. L., Kampff, A. R., & Engert, F. (2006). Spatiotemporal specificity of neuronal activity directs the modification of receptive fields in the developing retinotectal system. *Neuron*, *50*, 101–114.
- Volgushev, M., Vidyasagar, T. R., & Pei, X. (1995). Dynamics of the orientation tuning of postsynaptic potentials in the cat visual cortex. *Visual Neuroscience*, *12*, 621–628.
- Volterra, V. (1930). *Theory of functionals and of integral and integro-differential equations*. London: Blackey.
- Wielaard, J., Shelley, M., McLaughlin, D., & Shapley, R. (2001). How Simple cells are made in a nonlinear network model of the visual cortex. *Journal of Neuroscience*, *21*(14), 5203–5211.
- Wiener, N. (1958). *Nonlinear problems in random theory*. Technology Press Research Monographs. Cambridge: The Technology Press of Massachusetts Institute of Technology.
- Wolfe, J., & Palmer, L. A. (1998). Temporal diversity in the lateral geniculate nucleus of cat. *Visual Neuroscience*, *15*, 653–675.
- Xing, D., Shapley, R. M., Hawken, M. J., & Ringach, D. L. (2005). Effect of stimulus size on the dynamics of orientation selectivity in macaque V1. *Journal of Neurophysiology*, *94*, 799–812.
- Yamada, W. M., & Lewis, E. R. (1999). Predicting the temporal responses of nonphase-locking bullfrog auditory units to complex acoustic wave-forms. *Heart Research*, *130*, 155–170.