

# The Information Content of Spontaneous Retinal Waves

Daniel A. Butts and Daniel S. Rokhsar

*Physical Biosciences Division, Lawrence Berkeley National Laboratory and Department of Physics, University of California, Berkeley, Berkeley, California 94720-7300*

Spontaneous neural activity that is present in the mammalian retina before the onset of vision is required for the refinement of retinotopy in the lateral geniculate nucleus and superior colliculus. This paper explores the information content of this retinal activity, with the goal of determining constraints on the nature of the developmental mechanisms that use it. Through information-theoretic analysis of multielectrode and calcium-imaging experiments, we show that the spontaneous retinal activity present early in development provides information about the relative positions of retinal ganglion cells and can, in principle, be used at retinogeniculate and retinocollicular synapses to refine retinotopy. Remarkably, we find that most retinotopic information provided by retinal waves exists on relatively coarse time scales, suggesting that developmental

mechanisms must be sensitive to timing differences from 100 msec up to 2 sec to make optimal use of it. In fact, a simple Hebbian-type learning rule with a correlation window on the order of seconds is able to extract the bulk of the available information. These findings are consistent with bursts of action potentials (rather than single spikes) being the unit of information used during development and suggest new experimental approaches for studying developmental plasticity of the retinogeniculate and retinocollicular synapses. More generally, these results demonstrate how the properties of neuronal systems can be inferred from the statistics of their input.

*Key words: information theory; activity dependent; development; retinal waves; retinogeniculate; refinement; retinotopy*

Neuronal activity is required for the final stages of structural and functional maturation in many parts of the developing nervous system (Goodman and Shatz, 1993). Activity-dependent development in the CNS has been particularly well studied in the visual system, where there is a well defined mapping in connections between its various components. In mammals, for example, afferents from the retina connect “retinotopically” to both the lateral geniculate nucleus (LGN) and superior colliculus (SC): neighboring retinal ganglion cells (RGCs, the output layer of the retina) connect to neighboring cells in the LGN or SC. Retinotopy also exists in the connections between the LGN and visual cortex.

Although activity-independent cues are responsible for setting up an initial coarse retinotopy (Feldheim et al., 1998), the precise retinotopy present in the adult is not present at early stages of development (Sretavan and Shatz, 1987; Simon and O’Leary, 1992). Retinal arbors projecting into the LGN and SC initially occupy larger areas, and their axonal arbors are refined over the course of development via the elimination of incorrectly projecting afferents and the stabilization and elaboration of correctly projecting afferents. During this time, despite the absence of functional photoreceptors, neuronal activity is spontaneously generated within the retina (Galli and Maffei, 1988; Meister et al., 1991; for review, see Wong, 1999), and this activity has been

implicated in many aspects of axonal remodeling (Cramer and Sur, 1997; Penn et al., 1998) including refinement of retinotopy (Sretavan et al., 1988). Multielectrode (Meister et al., 1991; Wong et al., 1993) and imaging studies (Wong et al., 1995; Feller et al., 1996) have shown that this retinal activity is correlated between near neighbors such that the activity travels across the retina in waves. It remains to be determined whether the specific spatio-temporal patterning of these waves provides cues that instruct the refinement of retinotopy.

How could the firing patterns of RGCs be used by retinogeniculate and retinocollicular pathways to stabilize correctly projecting synapses while eliminating those that are misprojecting? It is thought that each synapse follows “learning rules,” by which feedback from local activity patterns is used by each synapse individually to determine whether it is correctly projecting and should be stabilized or it is misprojecting and should be eliminated instead. For example, because the activity of neighboring RGCs is correlated by the retinal waves, it has been proposed that a Hebbian-type learning rule (“cells that fire together wire together”) could be used to ensure that these cells connect to neighboring cells in the LGN and SC (Katz and Shatz, 1996; Wong, 1999). Several computational models using variations of Hebbian learning rules have successfully demonstrated that local learning rules of this nature can, in principle, produce retinotopic refinement (Haith and Heeger, 1998; Eglén, 1999; Elliott and Shadbolt, 1999). These models rely heavily on assumptions regarding the anatomy and physiology of the developing system, however, as well as on the nature of the learning rules themselves. Thus, although this theoretical work provides an “existence proof” of activity-dependent development, few constraints have been placed on the developmental mechanisms involved.

Here, we present a new approach to the study of activity-

Received June 3, 2000; revised Oct. 23, 2000; accepted Nov. 13, 2000.

This work was supported by the Department of Energy Grant LDRD-3668-27 to D.A.B. and D.S.R. We thank Carla Shatz, Marla Feller, Rachael Wong, and Markus Meister for providing the experimental data from the multielectrode and calcium-imaging experiments on which this work was based. In addition, this manuscript was greatly enhanced by comments from Marla Feller, Carla Shatz, Lisa Boulanger, and David Stellwagen and useful discussions with past and present members of the Shatz Lab and the Computational Neuroscience Course at Woodshole (1999).

Correspondence should be addressed to Dr. Daniel A. Butts, Department of Neurobiology, Goldenson 405, Harvard Medical School, 220 Longwood Avenue, Boston, MA 02115. E-mail: daniel\_butts@hms.harvard.edu.

Copyright © 2001 Society for Neuroscience 0270-6474/01/210961-13\$15.00/0

dependent mechanisms in the LGN and SC that uses only the statistical properties of the retinal activity and thus does not depend on assumptions regarding either learning rules or undiscovered experimental details of the LGN and SC. Instead, we rely on the tenet that if spontaneous activity does indeed instruct retinotopic refinement, then the signals comprising retinal waves must encode information about the relative positioning of RGCs. Through the application of a rigorous definition of “retinotopic information,” we can quantify the information produced by retinal activity. We find that this information is available over specific time scales and is conveyed by particular aspects of the retinal activity.

Specifically, we use experiments recording the simultaneous activity of retinal ganglion cells of the mammalian retina early in development, with multielectrode arrays [courtesy of Meister et al. (1991) and Wong et al. (1993)] and low-magnification calcium imaging [courtesy of Feller et al. (1996, 1997)], to assay the ability of retinal wave spike trains to convey information about the distance between retinal ganglion cells. Our analysis reveals that the retinotopic information is more robustly conveyed by bursts than by individual action potentials. Furthermore, information is available on time scales much longer than those considered previously as guiding synaptic plasticity in other developing systems (Zhang et al., 1998), which suggests that as-yet undiscovered mechanisms may govern activity-dependent development in the LGN and SC. We find that the bulk of retinotopic information at these time scales can be extracted by a simple coincidence-based Hebbian learning rule, in which pairs of bursts are either “coincident” or “not coincident,” and that the time window in which bursts are judged to be coincident is on the order of seconds.

Our methods demonstrate a new approach by which characteristics of a neuronal system can be deduced via the statistics of its input.

## MATERIALS AND METHODS

**Sampling error in probability distributions.** Throughout this paper, probability distributions must be estimated from a finite number of measurements. Consider the general problem of estimating the probability distribution  $p_i$  over a set with  $N$  categories ( $1 \leq i \leq N$ ). [For example, in many of the cases considered in this paper,  $p_i$  could represent  $p(\Delta t)$ , which corresponds to the set of time differences between 0 and some maximum  $T$ , divided up into bins of width  $\tau$ , so that  $N = T/\tau$ .] If, after  $M$  independent measurements, the number of times that  $i$  showed up was  $m_i$ , then the estimate of the probability distribution  $p_i$  is given by  $q_i = m_i/M$ .

This estimate  $q_i$  will approach  $p_i$  as  $M$  increases. For many different trials, each with  $M$  measurements,  $m_i$  will follow a binomial distribution with mean given by  $p_i M$  and variance given by  $p_i(1 - p_i)M$ . Thus, the estimated probability of a given bin has the following mean and SD:

$$q_i = p_i \pm \sqrt{p_i(1 - p_i)/M} \approx (m_i \pm \sqrt{m_i})/M, \quad (1)$$

where the above approximation holds for  $p_i \ll 1$ . Thus, to insure that the probability of a given bin  $i$  is adequately estimated, we need the number of samples in a given bin  $m_i \gg \sqrt{m_i}$  or equivalently  $m_i \gg 1$ .

**Calculating mutual information using a finite sample.** In this paper, we calculate the mutual information (MI) that burst onset time difference (BOTD)  $\Delta t$  encodes about the distance  $r$  between a pair of RGCs. Sampling errors in estimating the conditional probability distributions  $p(\Delta t|r)$  will typically bias the MI (Eq. 6) to a higher value. This occurs, in short, because errors in the estimated conditional distributions  $p(\Delta t|r)$  will be different (on average) for different  $r$  values, effectively making these distributions “more distinguishable,” although this distinguishability arises from sampling error. Because the statistics of the sampling error are known (see above), its effect on the calculated mutual information can be explicitly calculated (see Roulston, 1999). The MI calculated from estimates of the conditional probability distributions  $p(\Delta t|r)$  will be overestimated by a bias such that:

$$I_{\text{actual}}[r, \Delta t] \approx I_{\text{observed}}[r, \Delta t] - \frac{N_{\Delta t} N_r - N_{\Delta t} - N_r + 1}{2M \log 2}, \quad (2)$$

where  $N_{\Delta t}$  is the number of  $\Delta t$  bins and  $N_r$  is the number of  $r$  bins. The variance of  $I_{\text{observed}}[r, \Delta t]$  (from which error bars of Fig. 2A,B are calculated) is given by a more complicated formula:

$$V[I_{\text{observed}}[r, \Delta t]] = \frac{1}{M} \sum_r \sum_{\Delta t} p(r, \Delta t) [1 - p(r, \Delta t)] \times [\log_2 p(r) + \log_2 p(\Delta t) - \log_2 p(r, \Delta t) + I_{\text{observed}}[r, \Delta t]]^2. \quad (3)$$

Unfortunately, this estimate of both the bias and variance is not reliable when errors in the observed probability distribution are large (i.e.,  $m_i \approx 1$ ). Empirically,  $M/(N_{\Delta t} N_r) > 10$  is a good criterion for determining whether the bias can be accurately estimated (Roulston, 1999). In addition, the accuracy of the calculated MI can always be verified by artificially limiting the number of samples to verify that the estimate of mutual information is not changed.

In our calculation of the mutual information of the multielectrode array data, the limitation of  $M/(N_{\Delta t} N_r) > 10$  sets a constraint on the number of  $\Delta t$  bins that we can use for BOTD and restricts the time resolution of the conditional probability distributions  $p(\Delta t|r)$ . Because  $N_r = 9$  and the postnatal day 4 (P4) experiment that we use in our analysis provides 42,000 burst comparisons (for all  $\Delta t < 4$  sec), controlling the sampling error requires no more than 400 bins, corresponding to a minimum bin size of 10 msec. In calculating the information of spike-timing differences (see Fig. 2B), however, there are 100 times as many measurements of the spike time difference  $\Delta t_{i,j}$ , allowing the probability distributions  $p(\Delta t_{i,j}|r)$  to be sampled at a submillisecond time resolution.

**Analysis of the multielectrode experiments.** This paper uses multielectrode data recorded and analyzed previously by Meister et al. (1991) and Wong et al. (1993). We used data that were spike-sorted and assigned to electrode positions previously (as described in these papers). Spike times were specified to a time resolution of 50  $\mu\text{sec}$ . We defined a *burst* to be any cluster of spikes that were separated from each other by  $< 2$  sec, although changing this criterion (i.e., ranging anywhere from 1 to 5 sec) had no appreciable effect on our analysis. In this analysis, even a single isolated spike is considered a burst, although we also look at the effects of restricting the number of spikes in bursts (see Fig. 3).

The multielectrode array consists of a triangular lattice of 61 electrodes, with an electrode spacing of 70  $\mu\text{m}$  (see Meister et al., 1991). In many cases, a given electrode might record spikes from more than one neuron. To be true to the spatial resolution of the multielectrode array, distances between electrodes were classified in 70  $\mu\text{m}$  bins: 0–35  $\mu\text{m}$  (same electrode), 35–105  $\mu\text{m}$  (neighboring electrode), 105–175  $\mu\text{m}$  (two electrode spacings), and so on.

**Analysis of the calcium-imaging experiments.** Data from the calcium-imaging experiments provided by Feller et al. (1996, 1997) are in the form of low-magnification (6 $\times$ ) movies stored on videotape (see these papers for the details of the experiment) and were analyzed for this paper using NIH Image and programs written in C++. To calculate the information content of this activity, we needed to extract the “activity onset times” of areas of the retina sampled from a triangular lattice with 35  $\mu\text{m}$  spacing. In these experiments, wave activity in a given area causes a rise in the fluorescence signal of that area (see Fig. 4) for several seconds. The precise onset time of this activity, however, is often masked by fluctuations in the fluorescence signal. As a result, the following methods were developed, in large part by trial and error, and were found to be most effective at distinguishing the onset times of waves.

For each of these points, the fluorescence signal is averaged over a 33  $\mu\text{m}$  square for every frame of the movie (30 frames/sec) using NIH Image. Let the signal at a given point be represented by  $f(t)$ . Examples of the time course of  $f(t)$  are shown (see Fig. 4; a darkening of the fluorescence signal is shown as an increase in this figure). First, we perform initial wave discrimination and a coarse determination of its onset time by looking at the function:

$$g(t) = \int_t^{t+2\text{sec}} f(t') dt' - \int_{t-2\text{sec}}^t f(t') dt'. \quad (4)$$

At approximately the time of a wave, this function rises from zero and peaks at or near the onset time of the wave before returning back to zero.

Because these peaks are significantly larger than other peaks produced by natural fluctuations in the fluorescence signal, legitimate peaks (corresponding to wave activity) are distinguished from this noise by calculating the area under each peak and using a simple threshold. The coarse onset time  $t_i$  of each wave  $i$  is then assigned to each local maximum of  $g(t)$ .

This timing estimate  $t_i$  is then refined. First, the average fluorescence signal before the wave  $f_{av}$  is calculated by averaging  $f(t)$  from 5 to 1 sec before  $t_i$ . The maximum fluorescence signal  $f_{max}$  in the 2 sec after wave onset is also determined. Then, the revised estimate of the onset time is calculated using the best linear fit to  $f(t)$  (smallest chi-square) between the heights of  $f_{av} + 0.15 \times (f_{max} - f_{av})$  and  $f_{av} + 0.85 \times (f_{max} - f_{av})$ . The revised timing is given by the intersection of the best linear fit with a horizontal line at  $f_{av}$ .

These methods are able to distinguish the onset time of wave activity to sufficient accuracy, because the time resolution of the resulting wave onset times is at least as good as that measured from the burst onset times in the multielectrode experiments (see Fig. 5).

Activity in the calcium-imaging experiments was sampled in a triangular lattice with a spacing of 35  $\mu\text{m}$ . As a result, we classified distances with a 35  $\mu\text{m}$  resolution: 17.5–52.5  $\mu\text{m}$  (neighboring points), 52.5–87.5  $\mu\text{m}$  (two lattice spacings), and so on up to 1.33 mm. Note that there is no bin for cells separated from 0–17.5  $\mu\text{m}$  like the multielectrode array has, because only one calcium signal could be recorded from a given point.

*Comparisons between the multielectrode and calcium-imaging experiments.* The prior probability distributions of the multielectrode experiment  $p_{me}(r)$  and the calcium-imaging experiment  $p_{ci}(r)$  are implicitly different because of increases in the spatial resolution and extent that the imaging experiment affords. This will result in implicitly different values of MI, although they are describing the same phenomena.

To compare the information content of these experiments, it is necessary to scale the two prior distributions to agree with each other. To do so simulates the situation in which the imaging experiment actually samples from the same set of cells that the multielectrode array does but otherwise does not change anything about what the imaging experiment observes. For example, the conditional distributions  $p(\Delta t|r)$  are independent of the prior distribution, and changing the prior does not affect them.

First, the spatial resolution of the imaging data (35  $\mu\text{m}$ ) is collapsed to the spatial resolution of the multielectrode data (70  $\mu\text{m}$ ). Then, the new marginal distribution  $p_{ci}(\Delta t)$  was calculated using the following formula:

$$p_{ci,new}(\Delta t) \equiv \sum_r p_{me}(r) p_{ci}(\Delta t|r). \quad (5)$$

Finally, the mutual information of the imaging experiment is calculated (Eq. 6) using the multielectrode prior  $p_{me}(r)$  and the marginal distribution calculated in Equation 5.

## RESULTS

### Spatial information is encoded in the temporal properties of retinal waves

Multielectrode recordings from RGCs of the ferret just after birth (P0–P5) show that retinal ganglion cells undergo spontaneous episodes of activity approximately once every 2 min. Episodes are composed of bursts of action potentials that contain between 1 and 100 spikes and last an average of 1.1 sec. An example of such an episode is shown in Figure 1A, reproduced from data provided by M. Meister, R. Wong, and C. J. Shatz (Meister et al., 1991; Wong et al., 1993). The *left side* of Figure 1A shows the positions of electrodes in this array that recorded from cells in this experiment (P4 ferret retina), and the *shading of circles* in the array represents the timing of the burst onset of cells recorded at that position during the particular episode. Bursts among neighboring cells are often correlated in time, such that near-neighbors fire close together in time relative to RGCs that are more distant. Along the direction of propagation, the activity spreads sequentially across the retina, spanning the length of the multielectrode array (Fig. 1A, *right*, see *traces 1–8*). These action potentials are carried through the optic nerve and are known to

evoke action potentials in LGN neurons that in turn relay activity through to the visual cortex (Mooney et al., 1996).

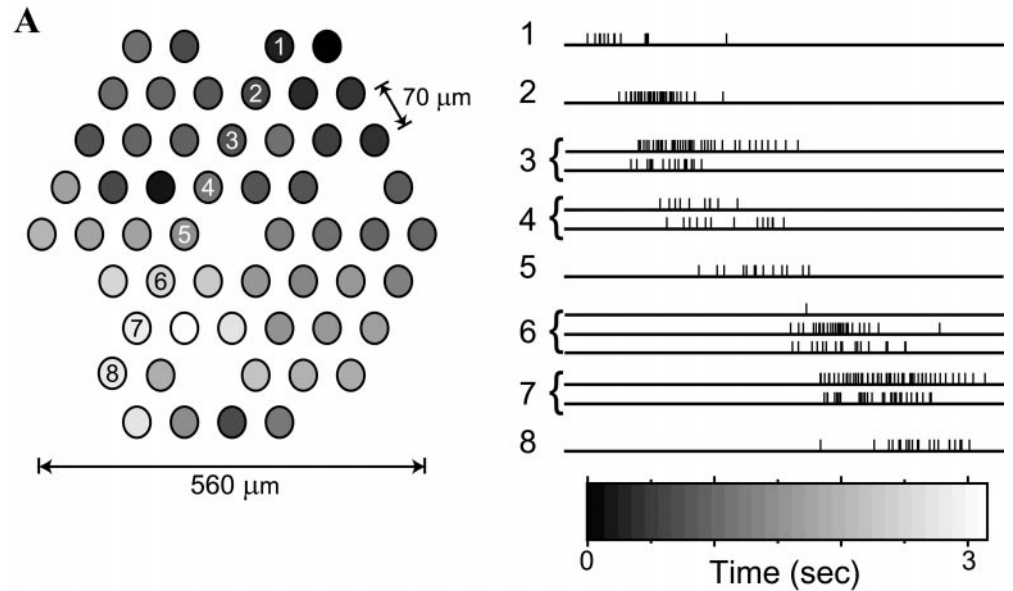
If action potential activity of RGCs is useful for refining retinotopy, the temporal structure of these spike trains must encode information about the relative position of RGC afferents. How can the amount of such information be assessed and quantified? Consider a pair of retinal ganglion cells separated by a distance  $r$  in the retina. If retinal wave activity did not encode information about the retinotopic separation of the pair, then relationships between the spike trains of the pair of neurons would be the same whether the pair was close together ( $r$  small) or far apart ( $r$  large). On the other hand, if information about retinotopic separation is encoded in the retinal waves, then there must exist temporal comparisons between the spike trains of each RGC that change as a function of  $r$ . The analysis of these comparisons is the focus of this work.

An example of a temporal comparison that might convey information about the retinotopic separation is the timing difference between the onset of RGC bursting. We define a burst to be any cluster of one or more spikes fired by a single cell that occur within 2 sec of each other and are separated from other spikes fired by that cell by at least 2 sec before and after. Burst onset time difference (BOTD) between two bursts is then simply the difference in time between the first spike of each burst. This is only one possible example of a temporal comparison between spike trains; there are many other comparisons that might carry retinotopic information, including those that use measures of correlation or the timing of individual action potentials. As we shall see, BOTD is a fundamental temporal comparison for the type of activity present in the developing retina, and other measures can be directly related to it. As a result, we will use BOTD to illustrate the methods of this paper in detail in this section and the next.

As seen in the spike trains of Figure 1A, *right*, the BOTD  $\Delta t$  is usually smaller for RGCs that are close together (compare *adjacent rows*), whereas cells that are further apart typically have longer delays. Such sequential firing occurs along the direction of propagation for a given wave (as in Fig. 1A, *traces*). In contrast, cells aligned along the wave front [i.e., perpendicular to the direction of propagation (Fig. 1A, from the *top left* to the *bottom right*)] often fire with small time differences despite having large spatial separations. Large variations in wave-front velocity and direction (Feller et al., 1997) further confuse any strict relationship between burst onset time difference and retinotopic separation. Thus, a given BOTD  $\Delta t$  occurs for a range of retinotopic separations, and conversely a particular  $r$  will produce a range of BOTDs.

To address this issue quantitatively, we calculate the probability that a pair of cells will have a BOTD of  $\Delta t$  given that they are separated by a distance  $r$  (Fig. 1B). This defines the conditional probability distribution  $p(\Delta t|r)$ , representing the probability of observing a BOTD of  $\Delta t$  between a pair of cells separated by a distance  $r$ . The multielectrode array (see Fig. 1A) forms a triangular lattice with a 70  $\mu\text{m}$  spacing and a diameter of 560  $\mu\text{m}$ . To be true to the spatial resolution of the array, we classify distances between any two electrodes into nine bins: 0–35  $\mu\text{m}$  (same electrode), 35–105  $\mu\text{m}$  (neighboring electrodes), 105–175  $\mu\text{m}$  (two lattice spacings), and so on up to 525–560  $\mu\text{m}$ . We calculate the nine possible conditional probability distributions that can be measured with this experiment; Figure 1B shows four of them.

The temporal resolution (bin size) that was chosen in creating the distributions shown in Figure 1B is limited by the number of



**Figure 1.** The relationship between burst onset time difference and retinotopic separation. *A*, Spontaneous bursting activity travels across the retinal ganglion cell layer of the developing mammalian retina. *Left*, The spatial locations of electrodes (numbered 1–8) that recorded from RGCs during a retinal wave are shown, with the gray scale corresponding to the relative time of burst onset (gray-scale bar, at right). Data are from P4 ferret retina (Wong et al., 1993). *Right*, The spike trains recorded from eight electrodes along a line (in *A*) are shown; burst onsets are mostly sequential. Note that an electrode often recorded from two or more cells. *B*, Conditional probability distributions  $p(\Delta t|r)$  demonstrate the likelihood that pairs of RGCs separated by a given distance will have burst onset time differences at different  $\Delta t$ . Data are shown for four different separations ( $r$  values). Typical error bars that result from sampling are shown, because each distribution is estimated from  $M$  total measurements divided between  $N$  total bins.

temporal comparisons between RGC pairs that could be made over the duration of the experiment (20 min), because there must be a minimum amount of data per bin to distinguish real variations from sampling error (see Materials and Methods). With enough data, we could in principle construct probability distributions up to the temporal precision of the experiment itself (50  $\mu$ sec). For the distributions shown (Fig. 1*B*), the temporal resolution is 10 msec, and the total number of samples  $M$  divided by the number of bins  $N$  is labeled on each panel. Error bars of typical magnitude are shown and are calculated using the standard sampling error of  $\sqrt{m}$  for a bin with  $m = M/N$  counts.

The fact that these distributions change as a function of retinotopic separation  $r$  means that BOTD encodes spatial information. As  $r$  increases, the most probable BOTD shifts away from zero, and the distribution of probable  $\Delta t$  significantly broadens, such that by  $385 \mu\text{m} < r < 455 \mu\text{m}$  (Fig. 1*B*, bottom right), there is a nearly uniform probability that any  $\Delta t$  will be observed.

The degree to which these distributions change with  $r$  is related to the average amount of information gained by a single BOTD observation and is likewise related to the number of waves needed to distinguish the distributions over time. In this paper, we will quantify this dependence using the Shannon Mutual Information (MI), a quantitative measure of the interdependence of retinotopic separation  $r$  and BOTD  $\Delta t$ . By use of the conditional distributions  $p(\Delta t|r)$  described above:

$$I[r, \Delta t] = \sum_r p(r) \sum_{\Delta t} p(\Delta t|r) \log_2 \left[ \frac{p(\Delta t|r)}{p(\Delta t)} \right], \quad (6)$$

where  $p(r)$  is the *prior distribution*, representing the probability that two recorded neurons chosen at random are a distance  $r$  apart. The prior distribution is determined by the physical positions of the neurons recorded in the experiment. After the prior  $p(r)$  is determined, the remaining term in Equation 6 can be

computed:  $p(\Delta t) = \sum_r p(r) p(\Delta t|r)$ . Because these distributions must be estimated from limited experimental data, an additional term that corrects for the resulting bias is added (Treves and Panzeri, 1995; Roulston, 1999), as described in Materials and Methods.

The MI has been studied in great detail both as a mathematical entity (Shannon and Weaver, 1949; Cover and Thomas, 1991) and in specific relation to neuroscience (Rieke et al., 1997; Borst and Theunissen, 1999). Notice that if the two variables  $r$  and  $\Delta t$  are independent, then the distribution of  $\Delta t$  will not depend on  $r$ , i.e.,  $p(\Delta t|r) = p(\Delta t)$ , and the term inside the logarithm becomes unity making the MI between  $r$  and  $\Delta t$  zero. The MI is always non-negative and grows as the conditional distributions  $p(\Delta t|r)$  become more distinct from each other and hence also more distinct from their weighted average  $p(\Delta t)$ .

Using data from multielectrode recordings performed by Meister et al. (1991) on P0–P5 ferret retinas and a time resolution of 10 msec (as in Fig. 1B), we found the mutual information between retinotopic separation and BOTD to be  $I[r, \Delta t] = 0.128 \pm 0.003$  bits, where the uncertainty is an estimate of the sampling error  $p(\Delta t|r)$  caused by the limited amount of experimental data (see Materials and Methods). Although this number has specific meaning with regard to the average reduction in amount of the uncertainty of  $r$  from a single measurement of BOTD (see Shannon and Weaver, 1949; Rieke et al., 1997), we do not rely on a direct interpretation of the absolute value of MI in this paper. Such an interpretation is complicated by several factors, including that a given LGN neuron receives input from an unknown number and distribution of RGCs that furthermore change as a function of age, affecting  $I[r, \Delta t]$  via the prior distribution  $p(r)$ . Additional complications include the difficulty in accounting for the information encoded by more than pairs of RGCs and the accumulation of information over the weeks that retinal waves are present.

As a result, we use MI as a *relative* measure through which different types of measurement can be quantitatively compared. As described above, MI represents the amount of change in the conditional distributions  $p(\Delta t|r)$  as a function of retinotopic separation  $r$  (Fig. 1B) and is able to capture nonlinear relationships within these probability distributions (Roulston, 1997). Measurements that are more effective at extracting retinotopic information will have larger differences in their conditional distributions and a higher MI.

### **BOTD conveys retinotopic information at coarse time scales**

We first analyze the structure of the information present in burst onset time difference before looking at other possible temporal comparisons that might contain information about retinotopic separation. As discussed above, MI is meaningful as a basis for quantitative comparisons between different possible ways of extracting retinotopic information from spike trains. The first set of comparisons that we make is between mutual information between retinotopic separation  $r$  and BOTD  $\Delta t$  at different time resolutions. We add random time offsets to all of the burst onset times in each experiment and recalculate mutual information  $I[r, \Delta t]$ . The time offsets are chosen randomly from a normal distribution with a zero mean and SD  $\sigma$ .

If the addition of temporal noise of magnitude  $\sigma$  decreases the MI, then we infer that the resolution of time differences on a scale smaller than  $\sigma$  is useful for distinguishing different retinotopic separations. In this case, if retinogeniculate synapses were unable

to resolve such timing differences, then they would be unable to take full advantage of the retinotopic information present in the wave activity.

Conversely, if the addition of temporal noise does not affect the MI, then the retinotopic information would be robust to timing errors on the order of  $\sigma$ , and the retinogeniculate synapse would not gain any information by being able to resolve such small timing differences. Thus, by investigating the dependence of mutual information on time resolution, we can discover the temporal scale on which developmental mechanisms responsible for retinotopy should act to make optimal use of the available information.

The dependence of mutual information on the temporal noise magnitude  $\sigma$  is shown in Figure 2A for two multielectrode experiments. When very small timing errors are introduced (*left*), the full information content of BOTD,  $0.128 \pm 0.003$  bits, is present. As expected, large timing errors (*right*) can completely eliminate the information present in BOTD. Notably, the full information is present up until  $\sigma \approx 100$  msec. This leads us to the important conclusion that a finer time resolution is not necessary to extract the retinotopic information available from this source.

As shown in the next section, this 100 msec time resolution is not particular to burst onset time difference but applies for a host of other temporal comparisons between RGC spike trains. One hundred milliseconds is a natural time scale of the retinal waves, because the average RGC spacing in the P0–P5 ferret retina is  $\sim 20 \mu\text{m}$ , and retinal waves propagate with an average speed of  $200 \mu\text{m}/\text{sec}$  (Wong et al., 1993; Feller et al., 1997), meaning that, on average, neighboring RGCs will fire 100 msec apart. In total, these results suggest that if BOTDs are significant in refining retinotopy at the retinogeniculate synapse during the period of our study, the mechanisms that are responsible for activity-dependent refinement of retinotopy could not gain additional information by distinguishing time resolutions finer than 100 msec. This is one of our principal results.

### **Diverse temporal comparisons between spike trains convey the same amount of retinotopic information**

As noted above, burst onset time difference is just one possible temporal comparison that can be made between the spike trains of two cells. The structure of RGC spike trains, which consist of short episodes (average of 1.1 sec) of a relatively high firing rate (average of 12 Hz) surrounded by large stretches lasting an average of 2 min with no firing (Wong et al., 1993), suggests that the bursts themselves might represent a single timing signal without regard to the timing of spikes within each burst.

There are a variety of single timing signals that can be derived from bursts that might be used to make alternative temporal comparisons, such as the time of the  $n$ th spike of a burst, the time of the maximal firing rate, the average time of the first five spikes, etc. Of the variety of other burst timings that we tested, few had as much information as BOTD, although most had an MI within a factor of two of that contained in BOTD (data not shown). This is not surprising because the MI of BOTD does not decrease significantly when individual burst timings are offset on the order of 100 msec (as shown in Fig. 2A). Other timing signals that arise from a burst will typically be delayed from the burst onset time by approximately the same amount, plus or minus a couple hundred milliseconds. As a result, the comparisons between such an alternative burst-timing signal can be viewed as the BOTD offset by a random time delay of average magnitude  $\sigma$ , and the resulting MI can be read off of Figure 2A. For example, the second spike in a

burst occurs with an average latency of 80 msec from the first spike, although this varies from burst to burst. If the retinogeniculate synapse were to miss the first spike in a particular burst (and therefore misjudge the onset), it would have negligible effect on the information provided, because information content does not decrease with timing error until 100 msec (Fig. 2A). Thus, although we could not test all possible burst-timing schemes, our findings are consistent with a model in which each burst conveys a timing signal and burst onset is a fair estimate of that timing signal.

There still remains the possibility that the bulk of the information provided by the retinal waves is encoded by temporal com-

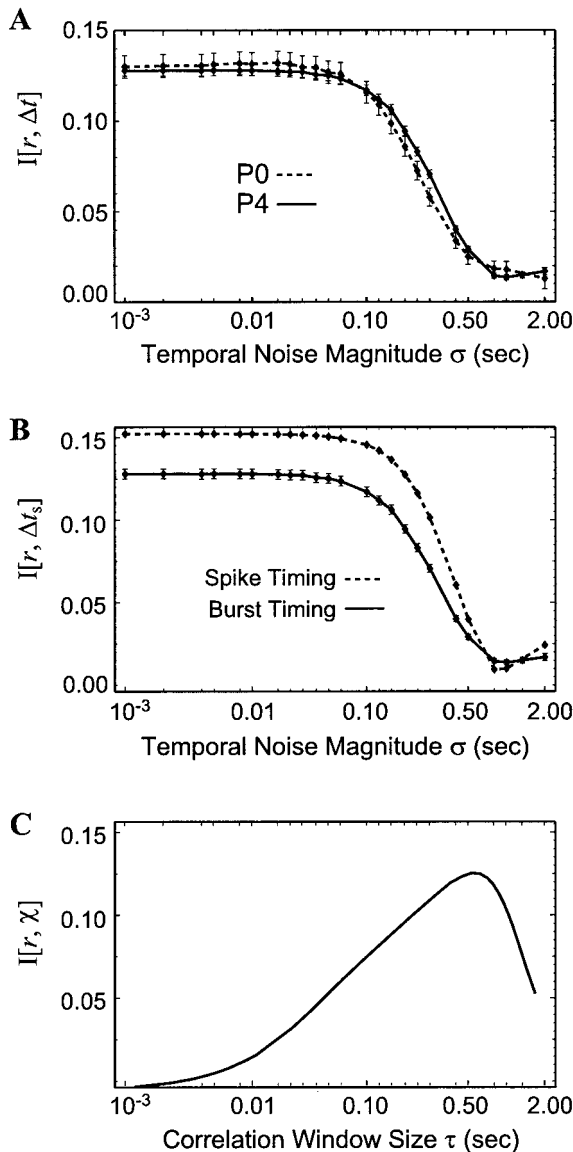
parisons that are not explicitly dependent on the burst structure. We therefore consider the possibility that individual action potentials convey separate timing signals, regardless of where in the burst they fall. We calculate the time difference  $\Delta t_s$  from between each spike of one cell relative to every spike of a second cell and use exactly the same methods of calculating the mutual information of BOTD: for every pair of cells separated by a distance  $r$ , we tabulate the time difference of each pair of spikes between the two cells and calculate the conditional probability distribution  $p(\Delta t_s | r)$  and the mutual information  $I[r, \Delta t_s]$ .

Figure 2B shows this mutual information with different magnitudes of temporal noise added. The nature of the spike time differences is very different; for example the average burst consists of 15 spikes, so there are  $15^2 (= 225)$  times more interspike measurements than BOTD observations. Yet, spike time difference contains almost the same amount of information about retinotopic separation (0.15 bits) as does BOTD (0.13 bits). Furthermore, the 15% more information encoded in spike timings can be accounted for by considering the number of spikes in bursts, as demonstrated in the next section.

Most notable, however, is that, although individual spike times are known to a precision of 50  $\mu$ sec, spike time differences have the same temporal resolution that burst time differences have; MI is essentially constant for temporal resolutions more precise than 100 msec.

A given pair of bursts will yield an average of 225 spike time comparisons, while providing only one BOTD. Does each spike time comparison give the same information as the single BOTD, meaning that a given pair of bursts will convey 225 times the information in spike timings? In fact, successive spike time differences carry redundant information, meaning that a BOTD between two bursts of a given size would yield a predictable distribution of spike time differences with no additional information in the individual spike time differences. The similarity between the MI of spike time differences and the MI of BOTD suggests that the structure of this distribution of spike time differences conveys little additional information, and the bulk of the information is conveyed by the mean, which is often very close to the BOTD. This is consistent with the fact that the time resolution of the information in individual spike times (100 msec) is slightly more than the average time between spikes during a burst (80 msec), meaning that information is not tied to the timing of particular spikes. We conclude that mechanisms at the retinogeniculate synapse may use either spike timing or burst timing to extract retinotopic information, because the same information conveyed by spikes is represented reliably at the burst level.

Another information-containing comparison of RGC spike trains is the number of coincident spikes between pairs of cells. This idea was originally proposed in Wong et al. (1993) and arises from an expectation that functional changes occurring at the retinogeniculate synapse might be governed by correlation-detecting mechanisms similar to those responsible for the synaptic modification observed in the hippocampus [i.e., long-term potentiation (LTP) and long-term depression (LTD) (Bear and Malenka, 1994)]. We define the *per-spike correlation index* ( $\chi$ ) between two cells A and B as the number of spikes that cell B fires in a  $\tau$  msec window centered around each spike of cell A. The correlation index presented in Wong et al. (1993) is our  $\chi$  averaged over the experiment and normalized by the firing rate of cell B. They found a clear (but not strict) dependence of correlation



**Figure 2.** The time resolution of different measures of retinotopic information. *A*, Gaussian-distributed noise with SD  $\sigma$  was added to burst onset times of two multielectrode experiments (P0 and P4), and the mutual information between these times and retinotopic separation was calculated. *B*, The mutual information from the P4 experiment (solid line, also in *A*) was compared with the MI between retinotopic separation and spike time difference, using the same techniques of adding different magnitudes  $\sigma$  of Gaussian-distributed noise. *C*, The mutual information between retinotopic separation  $r$  and per-spike correlation index  $\chi$  is shown as a function of window size  $\tau$ .

index on retinotopic separation: neighboring cells have an order of magnitude higher index than distant cells have.

The per-spike correlation index  $\chi$  can be compared with other per-event measurements of spike times and burst times that are made here. Instead of generating a single  $\chi$  for each pair of cells, we find the conditional probability distribution of  $\chi$  values for each separation  $p(\chi|r)$ , where the mean of this distribution matches the value found by Wong et al. (1993). As for BOTD, we can quantify this dependence on retinotopic separation by calculating the conditional probability distributions  $p(\chi|r)$  of correlation index  $\chi$  and retinotopic separation  $r$ . The resulting mutual information  $I[r, \chi]$  is shown as a function of coincidence window size  $\tau$  in Figure 2C. Although Wong et al. (1993) suggested a window of 50 msec (in analogy to LTP and LTD in the hippocampus), we see that greater amounts of retinotopic information exist for larger window sizes, because the MI peaks at 600 msec.

The coarse time resolution seen in Figure 2C is consistent with that seen in the MI of spike time differences and BOTD. Together, these calculations demonstrate that fine temporal features of retinal waves do not play a role in providing retinotopic information and that the information conveyed by BOTD is at least equivalent to that of other comparisons made between RGC spike trains.

### Bursts with many spikes are more significant than are bursts with fewer spikes

Although burst-timing differences and spike-timing differences convey approximately the same magnitude of mutual information about retinotopic separation, there is somewhat more information conveyed by spike-timing difference (15% more, see Fig. 2B). This discrepancy can be accounted for by considering the burst size in addition to BOTD in the calculation of retinotopic information. We will see below that bursts with fewer spikes actually carry less information than do bursts with many spikes. Although this has a significant effect on the information of bursts, the effect of small bursts on the information in spike time differences is naturally attenuated because relatively few spike time comparisons result from a burst with few spikes. On the other hand, in calculating the mutual information of BOTD, a burst consisting of a single spike has a weight equal to that of one with 50 spikes. By simultaneously considering burst length and burst timing, small bursts can be discriminated from larger bursts.

To consider the burst size at the same time as BOTD, we need to extend our definition of mutual information to include more than two variables. We introduce  $X$  to represent an additional observable(s) associated with each BOTD measurement (such as

the burst size). The modified mutual information that BOTD and burst size ( $\Delta t$  and  $X$ ) provide about retinotopic separation  $r$  is given by:

$$I[r, \{\Delta t, X\}] = \sum_r p(r) \sum_{\Delta t} \sum_X p(\Delta t, X|r) \log_2 \left[ \frac{p(\Delta t, X|r)}{p(\Delta t, X)} \right]. \quad (7)$$

In the case in which  $X$  is not related to  $\Delta t$ , then  $p(X, \Delta t|r) = p(\Delta t|r) p(X|r)$ , and  $p(\Delta t, X) = p(\Delta t) p(X)$ , so that the total amount of retinotopic information is just a sum of their separate information:  $I[r, \{\Delta t, X\}] = I[r, \Delta t] + I[r, X]$ .

Here, we use  $X$  to parameterize the sizes of the two bursts; categories based on the sizes of each burst are shown in Figure 3A. (These categories are chosen so that each  $X$  has a sufficient amount of data for an accurate calculation of the MI.) Because of the limited amount of data, it is only possible to make eight categories for  $X$  (see Materials and Methods), but this is enough to make a sufficient distinction between the information content of large and small bursts.

Calculating the conditional probability distributions  $p(\Delta t, X|r)$  and the resulting MI (from Eq. 7) gives  $I[r, \{\Delta t, X\}] = 0.147 \pm 0.03$  bits, which is very close to the  $0.151 \pm 0.01$  bits contained in the spike time difference. To determine why considering burst size increases the information content of BOTD, we introduce the *conditional mutual information*  $I[r, \Delta t|X]$ ; the information between  $r$  and  $\Delta t$  for a particular value of  $X$  (i.e., a particular pair of burst sizes) is given by:

$$I[r, \Delta t|X] = \sum_r p(r) \sum_{\Delta t} p(\Delta t|r, X) \log_2 \left[ \frac{p(\Delta t|r, X)}{p(\Delta t|X)} \right]. \quad (8)$$

Time comparisons involving bursts consisting of only one spike ( $X = 1$ ) contain only 0.044 bits of information about retinotopic separation, compared with time comparisons in which both bursts consist of 20 spikes or more ( $X = 8$ ), which contain 0.30 bits. The full range of conditional information  $I[r, \Delta t|X]$  is shown for each  $X$  in Figure 3B.

The conditional information  $I[r, \Delta t|X]$  is related to the total information expressed in Equation 7 (Cover and Thomas, 1991):

$$I[r, \{\Delta t, X\}] = I[r, X] + \sum_X p(X) I[r, \Delta t|X]. \quad (9)$$

Because burst size alone provides no information about retinotopic separation ( $I[r, X] = 0$ ), the total retinotopic information of

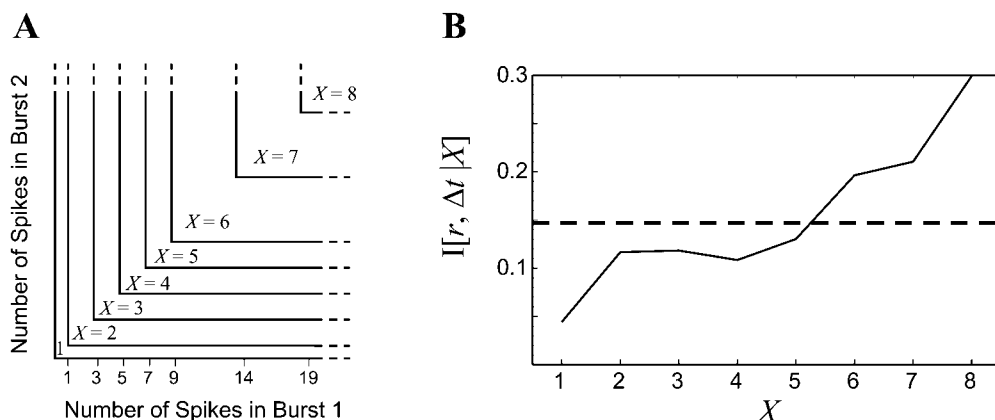


Figure 3. The mutual information considering burst size. A, Pairs of bursts were classified into categories  $X$  based on the number of spikes in each burst. These categories were chosen so that approximately the same amount of pairs falls into each. B, The mutual information between retinotopic separation  $r$  and burst onset time difference  $\Delta t$  conditional on burst size  $X$  (solid line) is shown. The total mutual information  $I[r, \{\Delta t, X\}]$  is shown as a dashed line.

a simultaneous consideration of BOTD and burst size is simply the weighted average of the conditional information.

This decomposition of the total MI into conditional information (Fig. 3*B*) shows that the amount of information conveyed by burst timing depends on the size of the burst. Bursts with many spikes carry a more reliable timing signal with respect to providing retinotopic information, whereas single-spike events convey almost no information. Together with our previous results, these analyses of the multielectrode array data suggest that bursts are the relevant units of information at this stage of visual system development.

### Low-magnification calcium-imaging experiments contain the same information as multielectrode experiments but filter out small bursts

The above analysis is based on experiments that used a multielectrode array to record the spike trains of retinal ganglion cells (Meister et al., 1991; Wong et al., 1993). Although such experiments are able to distinguish the individual action potentials of up to 100 recorded cells, such a method only samples the activity over 560  $\mu\text{m}$  for relatively short periods of time ( $\sim 20$  min). To gain further insight into the information content of retinal waves, we now analyze a second type of experiment that visualizes the spontaneous activity of the retina over much larger spatial scales and over longer times. The bursting of RGCs is accompanied by a large influx of calcium, which can be directly detected using the calcium-sensitive fluorescent dye fura-2 AM. In particular, spontaneous retinal activity can be monitored over an area of 2  $\text{mm}^2$  for periods of time up to 100 min (Feller et al., 1996, 1997).

Although this approach significantly increases the spatial extent over which retinal activity can be monitored, there are two potential drawbacks (see Wong, 1998). First, calcium imaging is not able to resolve individual spikes; changes in fluorescence correspond to the cumulative calcium signal. Second, the timing of the calcium signal onset (i.e., burst onset) can only be estimated with a timing precision on the order of 100 msec (see Materials and Methods). Fortunately, our multielectrode experiment analysis suggests that these two concerns are not significant in our study of calcium signals, because the burst onset timing provides

the full scope of retinotopic information that is present in the retinal activity. As a result, individual spike timings do not need to be known. Furthermore, we have seen that the time resolution of the retinotopic information is on the order of 100 msec, suggesting that the lack of temporal precision afforded by the imaging experiment should not affect our analysis.

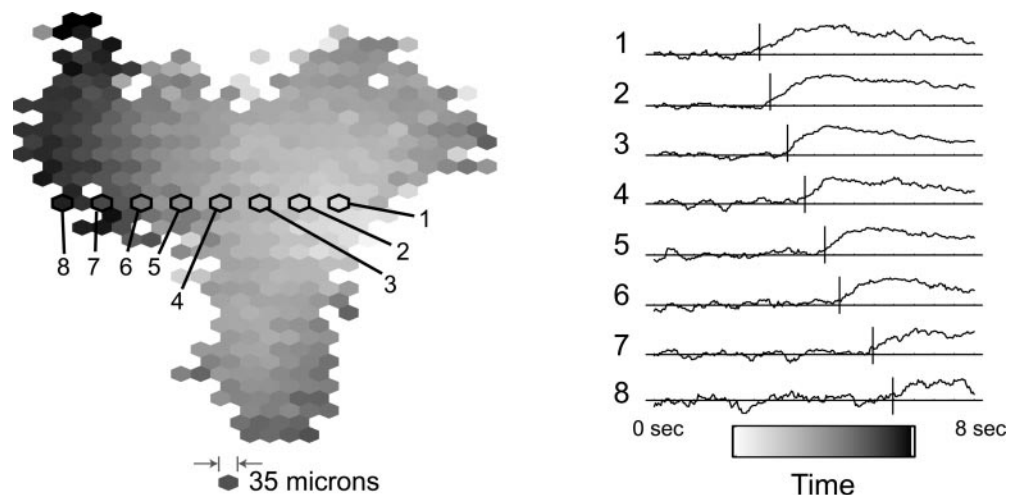
Figure 4 shows the time evolution of a single retinal wave, visualized with the imaging experiment. The timing of wave activity is determined at each point in a triangular lattice with 35  $\mu\text{m}$  spacing and dimensions of  $1.4 \times 1.2$  mm. The shading in Figure 4 represents the relative timing of wave activity at each point, with the corresponding timing bar shown at the *bottom right*. Along one of the directions of wave-front propagation (Fig. 4, *areas labeled 1–8*), the fluorescence changes occur sequentially (Fig. 4, *right*), in an analogous way to the data shown in Figure 1. The imaging experiment allows the full extent of wave propagation to be visualized (Fig. 4).

Fluorescence *traces* from eight points during the wave are shown in Figure 4, *right*. Without wave activity, the fluorescence level fluctuates around the average (*horizontal line*). During wave activity, the fluorescence level gradually rises as calcium enters bursting RGCs and the area darkens (Feller et al., 1996; Wong, 1998). The timing signal (*vertical bar*) derived from this fluorescence change is not trivial to extract, because the initial onset is often masked by fluctuations in fluorescence. The timing signal is given by the intersection of the previous average fluorescence (*horizontal line*) with the slope of the fluorescence rise (see Materials and Methods).

After the timing of activity in each area is determined, the conditional probability distributions can be estimated, and the MI can be calculated. We used data from three imaging experiments on P0–P4 ferret retina, lasting a cumulative total of 50 min. Because all three experiments had approximately the same probability distributions (data not shown), these data are combined to get better statistics.

We first demonstrate that the MI is equivalent for both types of experiment. The multielectrode array has the ability to observe only a small fraction of the area that is observed via calcium

**Figure 4.** The time evolution of a retinal wave visualized over 2  $\text{mm}^2$  using calcium imaging. Activity occurs sequentially along the path of the wave, but the large spatial scale of the imaging experiment allows the full-wave evolution to be visualized. *Left*, Using low-magnification calcium imaging of a P4 ferret retina, a timing signal of the fluorescence change during a wave is determined at each point in a *triangular* lattice. The onset time is represented by the *gray-scale level*; the corresponding timing shown in a *bar below the traces at right*. *Right*, The individual fluorescence *traces* from eight points (with the same spacing as the multielectrode array in Fig. 1) is shown. The fluorescence level fluctuates around the average (*horizontal line*) until the area undergoes wave activity, leading to a higher (i.e., *darker*) fluorescence value. The *vertical bar* shows the derived timing signal of the fluorescence change, determined using techniques described in Materials and Methods. The *vertical scale* is in arbitrary units representing the brightness of the pixels of the image on videotape.



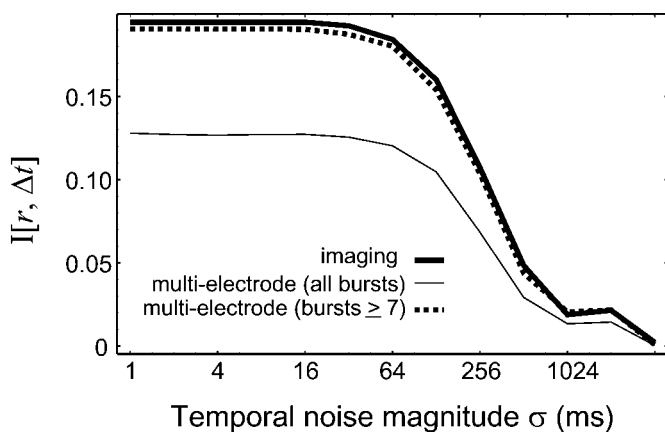


imaging; the imaging experiment is able to sample dimensions that are double the size of the array and at twice the spatial resolution. A naïve comparison neglecting the possibility that information is conveyed outside of the dimensions of the multi-electrode array would find a discrepancy between the MI of the two experiments. To make a direct comparison, we therefore scale the prior distribution  $p(r)$  of the imaging experiment to match that of the multi-electrode experiment (see Materials and Methods), such that the MI of each experiment reflects an equivalent distribution of cell pairs.

Figure 5 shows the scaled MI between burst time onset difference  $\Delta t$  and retinotopic separation  $r$  as a function of noise magnitude, calculated from the imaging experiment (*thick solid line*) and multi-electrode experiment (*thin solid line*; the same as in Fig. 2A). We see that the information measured from the two experiments has the same time resolution, but there is 50% more retinotopic information contained in the imaging experiment.

This difference can be attributed to the inability of the imaging experiment to resolve small bursts. As discussed in the last section, small bursts contain relatively little information, and as a result, the combined MI of both small and large bursts leads to a value of MI at their average (Fig. 3B, *dashed line*). By omitting the relatively imprecise contribution of the small bursts, the average measurement of BOTD would contain more information, resulting in a larger observed MI.

To see whether this explains the discrepancy between the observed MIs of the multi-electrode and imaging experiments, we recalculate the MI of the multi-electrode experiment using only bursts with larger numbers of spikes. In the case in which only bursts with seven or more spikes are included in our calculation, the MI of the multi-electrode experiment matches that of the imaging experiments (Fig. 5, *thick dashed line*). This interpretation can be verified by comparing the interburst intervals in both experiments. In the imaging experiment, the average interburst interval (at a given location) is 126 sec. Although the average interburst interval of the multi-electrode experiment is 68 sec when all bursts are considered, it rises to 132 sec when only bursts with greater than or equal to seven spikes are considered, con-



**Figure 5.** Comparison between the information content measured in two types of experiments. The information content of the calcium-imaging experiment (*thick solid line*) has the same time resolution as that of the multi-electrode experiment (*thin solid line*). The mutual information between retinotopic separation  $r$  and burst onset time difference  $\Delta t$  is plotted for different temporal noise magnitudes. The difference in the magnitude of MI between the experiments disappears when bursts with less than seven spikes are ignored when calculating  $I[r, \Delta t]$  from the multi-electrode data (*thick dashed line*).

sistent with the interpretation that the coarse filtering inherent in the imaging technique is responsible for the apparent difference in MI between the two experiments.

### Time delays on the order of seconds encode the bulk of the information content

The large area visualized in imaging experiments allows many more pairs of RGCs to be sampled in a given amount of time (compared with the multi-electrode experiments), resulting in a significant increase in the number of pairs of cells separated by a distance  $r$  that can be observed. The larger amount of data allows us to measure the information content of BOTDs up to 120 sec, the average interwave interval in the retina (Feller et al., 1997). Here we explore whether longer BOTDs have the capacity to carry retinotopic information.

To investigate this issue, we must use an MI that explicitly takes the maximum BOTD that we use into account. Up to this point we have used a cutoff BOTD of 4 sec in the calculation of MI (Figs. 2, 3, 5). We refer to the maximum BOTD  $T$  as the *observation window*, because BOTDs outside this window (i.e.,  $\Delta t > T$ ) have not yet been considered in our calculation of  $I[r, \Delta t]$ .

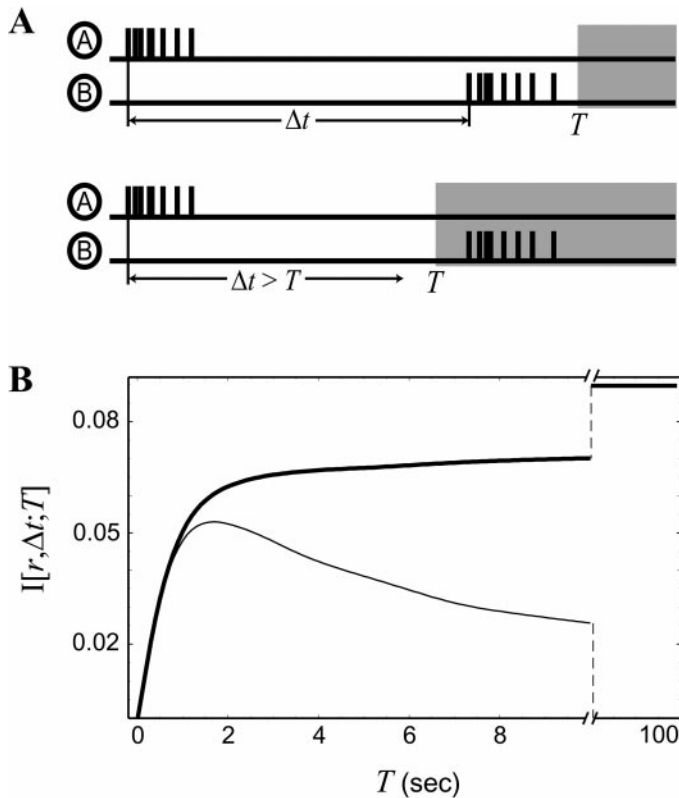
Unfortunately, the MI calculated with a certain observation window  $T$  cannot be directly compared with an MI with a different  $T$ . In particular, increasing the size of the observation window allows more measurements to be made, but at the same time, the average information gained per measurement decreases.

To avoid these complications, we must keep the number of measurements the same as we vary  $T$ . To accomplish this, we consider a given measurement of BOTD between two RGCs (cell A and cell B) as a two-step process. First, when cell A fires a burst, either cell B bursts within the observation window  $\Delta t \leq T$ , or it does not. Second, if cell A and cell B burst with  $\Delta t \leq T$ , then  $\Delta t$  is observed. As we will discuss in the next section, each stage of this two-step measurement provides information about retinotopic separation between cells A and B.

To include the first stage of the measurement in a calculation of MI, we create a slightly different set of conditional probability distributions of  $p'(\Delta t|r)$ . For a given pair of RGCs separated by a distance  $r$ , the conditional probability distribution  $p'(\Delta t|r)$  includes BOTDs shorter than  $T$ , as calculated in previous sections (Fig. 6A, *top*). Now, in addition to measurements within the observation window, we consider the extra possibility that  $\Delta t > T$ . The last time bin of the conditional distribution of  $p'(\Delta t|r)$  records the probability that this occurs (Fig. 6B, *bottom*). In this way, each burst of cell A either gets classified with a  $\Delta t \leq T$  or gets put into the last bin of the conditional distribution  $p'(\Delta t|r)$ , meaning  $\Delta t > T$ .

As the observation window  $T$  is decreased from 120 sec (Fig. 6A), more measurements are grouped into the last bin ( $\Delta t > T$ ), and their specific BOTD is neglected. Thus, each calculation of  $I[r, \Delta t; T]$  is derived from the same number of measurements, and the total number of measurements is now independent of the size of the observation window  $T$ . The resulting mutual information  $I[r, \Delta t; T]$  is calculated as before (Eq. 6) but with the new conditional probability distributions  $p'(\Delta t|r)$ . It is now directly comparable with the other MIs calculated with different  $T$ .

The new MI is shown in Figure 6B as the *thick line* for a range of  $T$  up to 10 sec. The information  $I[r, \Delta t; T]$  saturates at 0.09 bits as  $T$  nears the average interwave interval, because increasing  $T$  further does not include additional measurements. Note that this number is not directly comparable with the MI calculated in



**Figure 6.** The retinotopic information exists at course time scales. *A*, The observation window  $T$  is varied. For each pair of cells (cells *A*, *B*), a measurement is made for each burst of cell *A*. *Top*, If cell *B* bursts with its burst onset time difference  $\Delta t$  within  $T$  of the burst onset time of cell *A*, its BOTD is recorded normally. Otherwise, it is classified as  $\Delta t > T$  and added together with all other such measurements. *Bottom*, As the observation window is decreased, more and more measurements are given this classification, and information about a specific BOTD is discarded. *B*, The mutual information between  $r$  and  $\Delta t$  is calculated as a function of the observation window size  $T$  (*thick line*). This information  $I[r, \Delta t; T]$  decreases as more temporal information is discarded. The mutual information between retinotopic separation and simply whether  $\Delta t \leq T$  (coincident) or  $\Delta t > T$  (noncoincident), ignoring the specific value of  $\Delta t$ , is calculated as a function of the observation window  $T$  (*thin line*).

previous sections because of our altered probability distributions (below we present a formula relating the two).

Less than one-third of the total information is accounted for by measurements of  $\Delta t > 2.5$  sec, demonstrating that there is negligible retinotopic information to be gained by incrementally pushing the observation window out  $>2.5$  sec.

More strikingly, because measurements with  $\Delta t$  up to 2 sec each contain a significant amount of the information, a further restriction of BOTD (i.e., making  $T < 2$  sec) forfeits a significant amount of the available information. For example, if the retinogeniculate synapse were only sensitive to  $\Delta t \leq 100$  msec, it would receive  $<5\%$  of the available retinotopic information. These results demonstrate that an “optimal” learning rule would make use of burst onset time differences on the order of seconds.

#### A simple coincidence-based learning rule makes use of the bulk of retinotopic information

As described in the previous section, the new MI reflects a two-step observation: first it is determined whether cell *B* fires a burst within the observation window after cell *A* fires, and then, if  $\Delta t \leq T$ , the BOTD is measured. How does the new two-step

mutual information  $I'[r, \Delta t; T]$  compare with the original mutual information  $I[r, \Delta t]$  that only considered the second measurement? Let  $Y(T)$  represent the first step of this measurement, so that  $Y(T) = 1$  when  $\Delta t \leq T$  and  $Y(T) = 0$  when cell *B* does not burst within the observation window after cell *A*. The new MI  $I'[r, \Delta t; T]$  is the information that two observations provide about retinotopic separation; by use of the notation in Equation 9 in which BOTD  $\Delta t$  and burst size  $X$  were simultaneously considered,  $I'[r, \Delta t; T]$  is equivalent to  $I[r, \{\Delta t, Y(T)\}]$ . Adapting Equation 9 to this situation yields the relationship between the new MI and the old MI:

$$I'[r, \Delta t; T] = I[r, Y(T)] + f_{\text{tot}} I[r, \Delta t], \quad (10)$$

where  $f_{\text{tot}} = \sum_r p(r, Y = 1)$  is the total fraction of times that one cell burst within the observation window  $T$  of a second cell. Thus, retinotopic information is encoded each time cell *A* bursts, whether or not cell *B* fires a burst within the observation window. This information has two components, the information of firing within the observation window itself ( $I[r, Y]$ ) and the information of measuring BOTD ( $I[r, \Delta t]$ ) (both of which implicitly depend on  $T$ ). The latter is the same information calculated in Figure 2*A* (where  $T = 4$  sec). Because the information of  $\Delta t$  is only gained when cells *A* and *B* fire bursts within the observation window, this information is attenuated by  $f_{\text{tot}}$  (the fraction of times this occurs) in its contribution to the total information.

How much information is contained in the observation that two bursts were coincident within the observation window  $I[r, Y(T)]$ ? Such a measurement ignores any knowledge gained through the observation of BOTD within the observation window. Remarkably, we find that this MI represents a significant fraction of the total information expressed in Equation 10. This MI is shown in Figure 6*B* as a *thin line*. For smaller observation windows ( $T < 1$  sec),  $I[r, Y]$  closely matches the total information  $I'[r, \Delta t; T]$  (Fig. 6*B*, *thick line*). From Equation 10, we see that the difference between these two curves is the information encoded in specific BOTDs, equal to  $f_{\text{tot}} I[r, \Delta t]$ .

For larger observation windows, more bursts become classified as coincident, and less can be learned about spatial separation from this categorization; as a result,  $I[r, Y(T)]$  decreases for  $T > 2$  sec. At the same time, as more pairs of bursts occur within the observation window, the information of BOTD  $I[r, \Delta t]$  is able to significantly contribute to the total MI.

Thus, the bulk of the available retinotopic information, whether in BOTD or a simple coincidence-based mechanism, exists at coarse time scales (on the order of seconds). A significant fraction of the information can be extracted by a simple coincident/not coincident learning rule, but such a learning rule must have this coarse time resolution.

## DISCUSSION

Results from this study demonstrate that spontaneous retinal waves carry retinotopic information to primary targets in the developing visual system. How can spontaneous (i.e., unstimulated) activity convey information? Because retinal activity at this stage of development (P0–P5 in ferret) has distinctive spatiotemporal properties (Feller et al., 1997), the timing of activity of any two RGCs is related to the distance between them. In this way, spontaneous activity can provide information about the spatial structure of the system that produces it. We measured the relationship between the timing of activity and the position of the cells that generate it using conditional probability distributions

and quantified their interdependence using the formalism of information theory (Shannon and Weaver, 1949; Borst and Theunissen, 1999).

The idea that the timing of spontaneous activity across the retina might be used to refine retinotopy was proposed when it was discovered that the activity of neighboring cells is correlated in time (Meister et al., 1991). Here we directly analyze the information content of retinal waves to determine the time scales over which the information exists and the ways in which information may be most effectively extracted from features of the RGC spike trains. Under the assumption that retinal waves drive retinotopic refinement in RGC targets (discussed below), it follows that the mechanisms in these targets that are responsible for activity-dependent development are likely to be tuned to the information content that we have measured. In this way, our results probe developmental mechanisms in the LGN and SC by using only the statistics of their input.

Although the initial connections between RGCs and their primary targets are established via activity-independent means such as molecular guidance cues (Goodman and Shatz, 1993; Feldheim et al., 1998), refinement of these connections requires activity (Penn et al., 1998; Wong, 1999). Although the abolition of sodium action potentials prevents retinotopic refinement in the LGN (Sretavan et al., 1988), with similar results in the SC (Kobayashi et al., 1990; Simon et al., 1992), it is not clear that spatial patterning of retinal activity is specifically necessary for such refinement to occur (Crair, 1999).

Similar developmental processes that occur in cold-blooded vertebrates have been more extensively studied (Udin and Fawcett, 1988). In particular, during optic nerve regeneration in goldfish, refinement of retinal axonal arbors in the tectum is prevented by rearing in strobe light, which presumably destroys any spatial patterning that visual experience might confer to RGC firing (Schmidt and Eisele, 1985; Cook and Becker, 1990). Early retinotopic refinement in mammals, considered in this work, occurs before the onset of visual experience (Shatz, 1996; Wong, 1999), making such manipulations of early retinal activity in mammals more difficult (but see Stellwagen et al., 1999). Evidence of instructive retinotopic refinement in cold-blooded vertebrates, combined with evidence that activity-dependent retinotopic refinement occurs in mammals, suggests that it is likely that retinotopic refinement is instructive in mammals and needs information provided by the retinal waves.

Of course, both retinal waves and mechanisms at the retinogeniculate synapse are subject to biological constraints and may also have developed to receive other nonretinotopic information from the retina such as that driving eye segregation (Penn et al., 1998) and on/off segregation (Cramer and Sur, 1997). Although many of these developmental processes may not require specific spatial patterning, it is possible that retinal waves play multiple roles and might be tuned to provide more than just retinotopic information. Thus, although our results are not direct evidence of the existence of particular mechanisms in the LGN or SC, they suggest constraints that can direct future experiments studying the retinogeniculate and retinocollicular systems that are appropriate for the properties of the information content presented here.

### **The information content places constraints on possible activity-dependent learning rules**

The activity-dependent refinement observed at the system level is thought to arise from activity-mediated decisions on a synapse-

by-synapse level (Goodman and Shatz, 1993). Synapses may follow a set of learning rules by which the timing of presynaptic and postsynaptic activity leads to the stabilization or elimination of that synapse (Katz and Shatz, 1996). Experiments addressing this issue are beginning to help explain the functional changes that may occur on a relatively short time scale (such as LTP and LTD). Using an LGN slice preparation, Mooney et al. (1993) demonstrated that the developing RG synapse is capable of long-lasting synaptic enhancement by the eliciting bursts of action potentials in the optic tract that resulted in simultaneous bursting of LGN neurons. Unfortunately, because Mooney et al. only measured changes in the bulk synaptic current of the RGC axons that were stimulated in the optic nerve, the conditions through which individual RG synapses are modified were not addressed (i.e., when afferents are not all synchronously active).

These conditions were specifically studied in the *Xenopus* retinotectal system (Zhang et al., 1998), when single RGC spikes were elicited at different latencies relative to postsynaptic depolarizations in the tectum. Zhang et al. found a remarkable learning rule: the synapse became potentiated when the RGC spike preceded the postsynaptic spike by <20 msec but became depressed if it trailed the postsynaptic spike within 20 msec. Latencies of >20 msec had no effect. Although such results may reflect the conditions under which retinotopy is refined by visual experience (as occurs in the frog), in view of our results, we believe that these conditions may not be directly applicable to development in the mammalian RG system, where refinement occurs before the onset of visual experience.

First, our results strongly suggest that bursts, not individual isolated spikes, are the unit of information during early development at the mammalian retinogeniculate synapse. Individual spikes within the burst do not convey any additional information relative to the burst as a whole, and isolated spikes in fact convey a relatively unreliable timing signal. This finding is consistent with results found in other developing systems (Lisman, 1997).

Second, we find that the relevant time scale of the retinotopic information carried to the RG synapse is between 100 msec and 2 sec. Although some information is conveyed in spike correlations over smaller time windows (see Fig. 2C), it is only a fraction of the information that could be extracted by using a large time window. To make optimal use of the available retinotopic information, mechanisms should be able to discern time differences on the order of seconds (Fig. 6).

Finally, we found that the bulk of information available at these time scales could be extracted by a simple learning rule, where bursts that occur within an observation window on the order of seconds are classified as coincident and otherwise are classified as not coincident. Such a rule is very similar to the basic Hebbian principle that is often used to describe NMDA-mediated LTP and LTD (Bear and Malenka, 1994). The time scales of our proposed rule, however, are significantly larger than the 50 msec observation window usually suggested for NMDA-mediated LTP. The presence of other non-NMDA mechanisms that might act locally in an activity-dependent manner has been implied by numerous studies (Katz and Shatz, 1996). One candidate is neurotrophins (Bonhoeffer, 1996; Schuman, 1999), which may be released by active postsynaptic neurons and absorbed by presynaptic neurons (or their synapses) that had been active previously within an unknown (but probably larger) time window. Recently, a model of RG development has explored such a possibility (Elliott and Shadbolt, 1999).

## The teleology of retinal waves

In addition to setting spatial and temporal constraints on developmental mechanisms in the LGN and SC, our work gives insight into the evolutionary design of retinal waves. Retinal waves have many interesting properties, such as their slow rate of propagation (Meister et al., 1991; Wong et al., 1993), the limitation of their propagation to “domains” (Feller et al., 1996, 1997), and the fact that the activity consists of bursts, both in RGCs (Meister et al., 1991) and in LGN neurons (Mooney et al., 1996). The functional role of each of these aspects of retinal waves may be understood in the context of our results.

First, our results suggest that waves that propagate over larger areas of the retina would have less information content. The information contained in retinal waves predominantly expresses neighbor relationships, conveyed by burst onset times of 2 sec or less. Such a finding recapitulates one of the fundamental paradigms of activity-dependent development: the Hebbian learning rule that “cells that fire together wire together” (Katz and Shatz, 1996). Were retinal waves to propagate over larger regions of the retina, synchronously bursting cells would be increasingly far apart as the wave expanded, and less information overall would be provided about interneuron separations.

Another interesting aspect of the retinal waves is their slow rate of propagation, with an average speed of 100–300  $\mu\text{m}/\text{sec}$  (Meister et al., 1991). At this speed, neighboring RGCs (spaced  $\sim 20 \mu\text{m}$  apart) will burst an average of 100 msec apart. In our study, we found that this sets the time scale over which retinotopic information is available, suggesting that this might be a significant time scale for developmental mechanisms in the LGN and SC. In view of this, although many aspects of retinal waves are different among mammalian species, wave velocities are mostly conserved (Wong, 1999) and might be fixed by the needs of similar developmental mechanisms.

Finally, bursts are important in conveying information across developing systems (Lisman, 1997). Our results show that large bursts tend to carry much more reliable timing signals (Fig. 3) compared with those of small bursts and especially single spikes. Because of the possibility that individual spikes and small bursts are generated accidentally and are thus “noise,” burst size is a clear marker that the signal was actually part of the wave-generating machinery. Evidence suggests that RGC bursts are caused by large, featureless synaptic currents (Feller et al., 1996; Butts et al., 1999), and as a result spike timing within them may be arbitrary.

Bursting may have an additional role. We found that the bulk of retinotopic information conveyed by BOTD can be extracted by using an observation window on the order of 1 sec (Fig. 6). Because RGC bursts last an average of 1 sec, burst latencies on coarse time scales might be derived from spike latencies with much tighter temporal precision. In other words, the effect of individual spike coincidences, like Zhang et al. (1998) address, might combine in a nonlinear manner within the context of a burst, leading to an effective learning rule on the burst level that is derived from spike coincidences. In any case, an understanding of how retinal waves might drive retinogeniculate development requires the exploration of learning rules based on much larger time scales.

Mechanisms of activity-dependent synaptic modifications are used in a variety of situations both during development and adult life. Because endogenously generated retinal activity plays a role in the refinement of retinotopy in mammals (Sre-

tavan et al., 1988), it is conceivable that it evolved to provide the necessary activity during the long mammalian maturation that occurs before the onset of visual experience. If this were the case, then the spatiotemporal patterning of retinal waves, unlike activity arising from the visual scene, would be tuned to provide the appropriate stimulus for mechanisms of activity-dependent refinement in the LGN and SC. In this way, observable spontaneous activity present throughout the developing brain (neocortex, spinal cord, etc.) might provide a window through which to study the activity-dependent mechanisms responsible for brain development.

## REFERENCES

- Bear MF, Malenka RC (1994) Synaptic plasticity: LTP and LTD. *Curr Opin Neurobiol* 4:389–399.
- Bonhoeffer T (1996) Neurotrophins and activity-dependent development of the neocortex. *Curr Opin Neurobiol* 6:119–126.
- Borst A, Theunissen FE (1999) Information theory and neural coding. *Nat Neurosci* 2:947–957.
- Butts DA, Feller MB, Shatz CJ, Rokhsar DS (1999) Retinal waves are governed by collective network properties. *J Neurosci* 19:5380–5390.
- Cook JE, Becker DL (1990) Spontaneous activity as a determinant of axonal connections. *Eur J Neurosci* 2:162–169.
- Cover TM, Thomas JA (1991) *Elements of information theory*. New York: Wiley.
- Crair MC (1999) Neuronal activity during development: permissive or instructive? *Curr Opin Neurobiol* 9:88–93.
- Cramer KS, Sur M (1997) Blockade of afferent impulse activity disrupts on/off sublamination in the ferret lateral geniculate nucleus. *Dev Brain Res* 98:287–290.
- Eglen SJ (1999) The role of retinal waves and synaptic normalization in retinogeniculate development. *Philos Trans R Soc Lond [Biol]* 354:497–506.
- Elliott T, Shadbolt NR (1999) A neurotrophic model of the development of the retinogeniculocortical pathway induced by spontaneous retinal waves. *J Neurosci* 19:7951–7970.
- Feldheim DA, Vanderhaeghen P, Hansen MJ, Frisén J, Lu Q, Barbacid M, Flanagan JG (1998) Topographic guidance labels in a sensory projection to the forebrain. *Neuron* 21:1303–1313.
- Feller MB, Wellis DP, Stellwagen D, Werblin FS, Shatz CJ (1996) Requirement for cholinergic synaptic transmission in the propagation of spontaneous retinal waves. *Science* 272:1182–1187.
- Feller MB, Butts DA, Aaron HL, Rokhsar DS, Shatz CJ (1997) Dynamic processes shape spatiotemporal properties of retinal waves. *Neuron* 19:293–306.
- Galli L, Maffei L (1988) Spontaneous impulse activity of rat retinal ganglion cells in prenatal life. *Science* 242:90–91.
- Goodman CS, Shatz CJ (1993) Developmental mechanisms that generate precise patterns of neuronal connectivity. *Cell* 72:77–98.
- Haith GL, Heeger D (1998) A computational model of retinogeniculate development. *Computational neuroscience: trends in research, 1998* (Bower JM, ed), pp 35–40. New York: Plenum.
- Katz LC, Shatz CJ (1996) Synaptic activity and the construction of cortical circuits. *Science* 274:1133–1138.
- Kobayashi T, Nakamura H, Yasuda M (1990) Disturbance of refinement of retinotectal projection in chick embryos by tetrodotoxin and gray-anotoxin. *Dev Brain Res* 57:29–36.
- Lisman JE (1997) Bursts as a unit of neural information: making unreliable synapses reliable. *Trends Neurosci* 20:38–43.
- Meister M, Wong RO, Baylor DA, Shatz CJ (1991) Synchronous bursts of action potentials in ganglion cells of the developing mammalian retina. *Science* 252:939–943.
- Mooney R, Madison DV, Shatz CJ (1993) Enhancement of transmission at the developing retinogeniculate synapse. *Neuron* 10:815–825.
- Mooney R, Penn AA, Gallego R, Shatz CJ (1996) Thalamic relay of spontaneous retinal activity prior to vision. *Neuron* 17:863–874.
- Penn AA, Riquelme PA, Feller MB, Shatz CJ (1998) Competition in retinogeniculate patterning driven by spontaneous activity. *Science* 279:2108–2112.
- Rieke F, Warland D, de Ruyter van Steveninck RR, Bialek W (1997) *Spikes: exploring the neural code*. Cambridge, MA: MIT.
- Roulston MS (1997) Significance testing of information theoretic functionals. *Physica D* 110:62–66.
- Roulston MS (1999) Estimating the errors on measured entropy and mutual information. *Physica D* 125:285–294.
- Schmidt JT, Eisele LE (1985) Stroboscopic illumination and dark rearing block the sharpening of the regenerated retinotectal map in goldfish. *Neuroscience* 14:535–546.
- Schuman EM (1999) Neurotrophin regulation of synaptic transmission. *Curr Opin Neurobiol* 9:105–109.

- Shannon CE, Weaver W (1949) The mathematical theory of communication. Urbana, IL: University of Illinois.
- Shatz CJ (1996) Emergence of order in visual system development. *Proc Natl Acad Sci USA* 93:602–608.
- Simon DK, O'Leary DDM (1992) Development of topographic order in the mammalian retinocollicular projection. *J Neurosci* 12:1212–1232.
- Simon DK, Prusky GT, O'Leary DDM, Constantine-Paton M (1992) *N*-Methyl-D-aspartate receptor antagonists disrupt the formation of a mammalian neural map. *Proc Natl Acad Sci USA* 89:10593–10597.
- Sretavan DW, Shatz CJ (1987) Axon trajectories and pattern of terminal arborization during the prenatal development of the cat's retinogeniculate pathway. *J Comp Neurol* 255:386–400.
- Sretavan DW, Shatz CJ, Stryker MP (1988) Modification of retinal ganglion cell axon morphology by prenatal infusion of tetrodotoxin. *Nature* 336:468–471.
- Stellwagen D, Shatz CJ, Feller MB (1999) Dynamics of retinal waves are controlled by cyclic AMP. *Neuron* 24:673–685.
- Treves A, Panzeri S (1995) The upward bias in measures of information derived from limited data samples. *Neural Comput* 7:399–407.
- Udin SB, Fawcett (1988) Formation of topographic maps. *Annu Rev Neurosci* 11:289–327.
- Wong ROL (1998) Calcium imaging and multi-electrode recordings of global patterns of activity in the developing nervous system. *Histochem J* 30:217–229.
- Wong ROL (1999) Retinal waves and visual system development. *Annu Rev Neurosci* 22:29–47.
- Wong RO, Meister M, Shatz CJ (1993) Transient period of correlated bursting activity during development of the mammalian retina. *Neuron* 11:923–938.
- Wong ROL, Chernjavsky A, Smith SJ, Shatz CJ (1995) Early functional neural networks in the developing retina. *Nature* 374:716–718.
- Zhang LI, Tao HW, Holt CE, Harris WA, Poo MM (1998) A critical window for cooperation and competition among developing retinotectal synapses. *Nature* 395:37–44.