Variability and information in a neural code of the cat lateral geniculate nucleus

Robert C. Liu[1,2], Svilen Tzonev[1,2],
Sergei Rebrik[1] and Kenneth D. Miller[1,2,3]

[1] Keck Center for Integrative Neuroscience
and Department of Physiology
[2] Sloan-Swartz Center for Theoretical Neurobiology
[3] Department of Otolaryngology
University of California
San Francisco, CA 94143-0444

Aug. 27, 2001

Final Draft of a manuscript that appeared as
Journal of Neurophysiology **86**, 2789-2806 (2001)

Address Correspondence To:
Robert C. Liu
UCSF Sloan-Swartz Center, Physiology Box 0444
513 Parnassus Avenue, HSE806
San Francisco, CA 94143-0444
Phone: 415-502-6682
Fax: 415-476-4929
Email: `liu@phy.ucsf.edu`

Abstract

A central theme in neural coding concerns the role of response variability and noise in determining the information transmission of neurons. This issue was investigated in single cells of the lateral geniculate nucleus of barbiturate anesthetized cats by quantifying the degree of precision in and the information transmission properties of individual spike train responses to full field, binary (bright or dark), flashing stimuli. We found that neuronal responses could be highly reproducible in their spike timing (about 1-2 ms standard deviation) and spike count (about 0.3 ratio of variance/mean, compared to 1.0 expected for a Poisson process). This degree of precision only became apparent when an adequate length of the stimulus sequence was specified to determine the neural response, emphasizing that the variables relevant to a cell's response must be controlled in order to observe the cell's intrinsic response precision. Responses could carry as much as 3.5 bits/spike of information about the stimulus, a rate that was within a factor of two of the limit the spike train can transmit. Moreover, there appeared to be little sign of redundancy in coding: on average, longer response sequences carried at least as much information about the stimulus as would be obtained by adding together the information carried by shorter response sequences considered independently. There also was no direct evidence found for synergy between response sequences. These results could largely, but not entirely, be explained by a simple model of the response in which one filters the stimulus by the cell's impulse response kernel, thresholds the result at a fairly high level, and incorporates a post-spike refractory period.

Introduction

To understand the coding of information by neurons, it is important to quantify the variability in their responses. When this variability is driven by changes in the stimulus, the neurons can use this to distinguish between stimuli. On the other hand, when this variability occurs in repeated responses to the same stimulus, it acts as noise that reduces the neurons' potential capacity to code information.

The study of neuronal variability has recently seen a rebirth of interest in association with the renewed use of information-theoretic techniques for analyzing neural coding (Bair 1999; Borst and Theunissen 1999; Buracas and Albright 1999; de Ruyter van Steveninck et al. 1997; Meister and Berry 1999; Rieke et al. 1997; Victor 1999). In the visual system, the precision of spike times and counts has been investigated in several neural areas, although only a few have looked at the lateral geniculate nucleus (LGN) (Guido and Sherman 1998; Hartveit and Heggelund 1994; Kara et al. 2000; Keat et al. 2001; Reich et al. 1997; Reinagel and Reid 2000; Sestokas and Lehmkuhle 1988). In this paper, we further explore the degree of precision found in LGN neurons of barbituate-anesthetized cat by examining both spike count and timing measures. We go on to quantify the amount of information transmitted by neurons about the stimulus, and to determine the degree to which models of response based on linear integration of inputs can account for the observed precision.

A unique feature of the present approach is that we closely examined the dependence of neuronal variability on the degree of specification of the stimulus. To do this, we employed a pseudo-random binary stimulus known as an M-sequence (Sutter 1992). We focused only on characterizing the neurons' response to temporally varying stimuli by showing full-field bright and dark frames, ignoring the center-surround spatial structure of LGN neurons. M-sequences provide a statistically efficient and convenient method for analyzing responses because they have the nice property that every sequence of bright and dark frames of a given length (up to some limit) is repeated the same number of times somewhere throughout the sequence (see Methods). This allowed us to simultaneously examine the responses – both the mean response and the variability in the response – to *every* sequence of a given length,

giving us a detailed characterization of the neural code for such sequences. By varying this length, we examined how much of the stimulus had to be specified to maximize the precision of a neuron's response: *e.g.*, if the neuron's response was influenced by the last ten frames and only five frames were specified, then the response would be averaged over the unspecified frames, causing the neuron's responses to appear more variable than they would be if the stimulus were fully specified. The variability remaining when the stimulus was fully specified reflected the neuron's intrinsic response variability.

It is common to characterize a cell's response by its linear temporal kernel which – as computed from an M-sequence stimulus, and neglecting normalization (see Methods) – is the difference between its mean response to a single bright frame and its mean response to a single dark frame. We found that average responses to a single bright or dark frame within a sequence showed Poisson-like spike count variability and temporal dispersion over tens of milliseconds, and the kernel was correspondingly temporally broad. But by specifying more of the stimulus – *e.g.* specifying eight consecutive frames – the response could become far more precise, with sub-Poisson spike count variability and temporal precision of 1-2 ms. The information conveyed by the neuron correspondingly increased, containing as much as 3.5 bits/spike about longer stimulus sequences. We found that this information depended on the specification of spike times down to 1 ms resolution, and that the information in consecutive spikes showed little redundancy or synergy. Finally, we determined that the precision obtained when multiple frames were specified could be largely, but not entirely, explained if the spike rate arose from a filtering of the stimulus by the cell's temporal kernel followed by thresholding, along with imposition of a post-spike refractory period.

Some of this work was previously presented in abstract form (Liu et al. 2000; Tzonev et al. 1997).

Methods

Experiments

We performed experiments on adult cats under a protocol approved by the University of California, San Francisco Committee on Animal Research. Cats were initially anesthetized with isoflurane (1-5%), and placed on a feedback-controlled heating pad to maintain body temperature at 37.5-38^{o}C. We established an iv line, and thereafter maintained anesthesia via thiopental sodium or sodium pentobarbital (the latter was given once anesthesia was stable). The heart rate, respiratory rate, core temperature, O_2 saturation, expiratory CO_2 and lung pressure were all continually monitored. After performing a tracheotomy, the animal was respirated with nitrous oxide in a 1:1 ratio with oxygen. We performed a craniotomy, and then paralyzed the animal by infusing gallamine (10mg/kg/hr in lactated dextrose ringers). The electroencephalogram (EEG) was subsequently monitored continuously. We reflected the optic disk onto a white background using a fiber optic light source, and inserted contact lenses to focus the eyes at a distance of 35-40 cm.

We recorded extracellularly using tetrodes (Gray et al. 1995) advanced through a guide tube inserted to within a few millimeters of the LGN. The LGN was recognized by the small (relative to surrounding structures) and monocular visual receptive fields, and by the match of topography across repeated penetrations to published accounts (Sanderson 1971). The electrodes were constructed from 13 μm diameter nickel chromium insulated wire (about 20 μm including the insulation). The tips were beveled and gold-plated, and the typical impedance was in the range of 0.8-1.5 MOhm. Tetrode signals were amplified and then digitized at 20 or 30 kHz with 12-bit resolution. The digitized data were continuously streamed to the disk. To separate signals from different neurons, we sorted based on the spike amplitudes measured at the four tetrode wires. Clustering was done manually using different two-dimensional projections of the four-dimensional space.

Stimulus

For visual stimulation, sequences of full-field bright and dark frames were presented on a computer monitor at the rate of 120 Hz, yielding a frame duration of $t_f \approx 8.3$ ms. Each frame varied randomly between bright or dark, with a photopic mean luminance; contrast (measured as $(L - D)/(L + D)$ where L and D were the luminances of bright and dark frames, respectively) for each full sequence was chosen from 6%, 14%, 20%, 40%, or 80%.

We generated random frames using a binary M-sequence, which is essentially a stream of pseudo-random bits having some special properties (see below). A bit value of 1 corresponded to a bright frame, and 0 corresponded to a dark frame.

An M-sequence of order n consists of $2^n - 1$ bits. The full sequence can be viewed as a collage of overlapping k-bit sequences, $k \leq n$, drawn from the list of all possible binary combinations of k bits. For example, for $k = 2$, the possible binary combinations are: (0) 00, (1) 01, (2) 10, and (3) 11. Thus a portion of the full sequence consisting of the bits 0110100 can be decomposed as the overlapping combination of the sequences (1), (3), (2), (1), (2), (0). The same decomposition procedure can be applied for any k. The M-sequence has the convenient property that all subsequences of length $k \leq n$ randomly appear within the full sequence the same number of times, namely 2^{n-k} occurrences (except that the all-zero sequence of length k appears $2^{n-k} - 1$ times). Because of this statistical regularity of the M-sequence, it is an excellent tool for the investigation of a cell's neural code.

Analysis

Cells were selected for analysis based on the following criteria. To ensure single cell isolation, we chose only cells with clearly isolated clusters in the various two-dimensional projections of the four-electrode amplitude space; clusters with clipped responses due to amplifier saturation were avoided. In order to achieve reasonable estimates of the information rates, at least 1000 spikes were required during the whole stimulus. Finally, only cells with ON or OFF linear temporal kernels (see below) were studied, since this formed the basis for the definition of response events. In total, 12 cells (4 ON, 8 OFF) in one cat were studied at five

contrast levels – 80% (9 cells), 40% (6 cells), 20% (3 cells), 14% (1 cell), and 6% (2 cells) – yielding a total of 21 trials.

Response events and precision analysis

To study the precision of spikes, we attempted to classify each individual spike as part of a spike event evoked in response to a specific sequence of k frames. This was done by applying the following algorithm, described here for an OFF cell. We determined the average stimulus before a spike, and defined the cell's mean conditional latency (conditioned on a spike) as the time to the zero-crossing between peak and trough in the spike-triggered-average stimulus (illustrated in Figure 1). Then, as shown in Figure 2, for each spike in the train, we looked back in time from the spike by the mean conditional latency and found the closest OFF transition (bright frame followed by dark frame) within a window of ± 1.5 frames; the spike was assigned to that transition. If there was no such transition, the spike was unclassified. We characterized sequences by their length k and the location t of the transition within the sequence (*e.g.* $k = 8$, $t = 3$ labeled an eight-frame sequence with a transition at the onset of the third frame – that is, between the fourth and third frames, where the first frame was the latest in time). For a given choice of k and t, a given transition was uniquely associated with a surrounding sequence, and the spike was assigned to that sequence. All spikes associated with the same sequence were labelled as part of the same event. The percentage of total spikes that were unclassified served as a measure of the level of "spontaneous" activity that was not driven by transitions.

Once the events were identified for a given choice of k and t, the probability that a specific sequence produced an event was computed by dividing the number of times some spike response (≥ 1 spike) was obtained for that sequence, by the total number of presentations of that sequence (i.e. $2 \times 2^{14-k}$ times). This quantity was called the event probability.

We assessed the timing precision of the first spike in an event for each sequence consisting of a specified number of frames, k, with transition location t. A distribution for the times to the first spike in an event (of one or more spikes) was obtained from the numerous

presentations of a particular k-frame sequence. A jacknife estimate of the standard deviation of this first-spike time was used as the index of the timing precision (Thomson and Chave 1991), and the error was taken as the square root of its variance. We approximated the overall first-spike timing jitter for a given k and t by the median standard deviation across all k-frame sequences with transition location t. The timing jitter was then studied as a function of k and t.

To determine whether the timing jitter was correlated with the event probability, we computed the Spearman rank-order correlation (Press et al. 1992, pp. 639-642) for eight-frame sequences that had $t = 3$, the transition position that generally resulted in the smallest timing jitter. In several cases, there were sequences with very small event probabilities, and hence very few event responses from which to estimate the timing jitter. This could result in particularly large or particularly small jitters. To test whether this may have biased our estimate of the correlation, we calculated the Spearman rank-order correlation under two conditions: (1) using all sequences, and (2) using only those sequences with event probabilities above a minimum probability. This minimum probability was arbitrarily taken to be $1/\sqrt{N}$, where N is the number of presentations per sequence ($N = 128$ for eight-frame sequences).

We also assessed the spike count precision of the events for each sequence of a specified k and t. In this case, we generated a histogram of the number of spikes in the event responses for each sequence, allowing for the possibility of no spikes. A jacknife estimate of the variance of that distribution was used as the index of that sequence's count precision. The error was again taken as the square root in the variance of this estimate. To summarize the results across all sequences of length k with a given t, the Fano factor (variance divided by the mean) for each sequence was also estimated by jacknife. The median spike count Fano factor was then used to show the dependence of spike count precision on k for a given t.

Information analysis

The information in the spike train about the stimulus was quantified using the "direct" method (de Ruyter van Steveninck et al. 1997; Strong et al. 1998a,b). This method estimates the mutual information between stimulus and response "directly" from the spike trains, without regard to the details of the stimulus/response relationship, and with very few assumptions about the coding strategy. This method relies on the fact that the mutual information between the stimulus and response can be written as the difference of two spike train entropies. First, the maximum amount of information that a spike train response $\mathcal{R}$ can provide about the stimulus is just given by the entropy of the spike train itself, $H(\mathcal{R})$. This is estimated from the probability distribution of spike responses over the course of the whole experiment, without specific knowledge of the stimulus. Second, the information the spike train carries about the stimulus is reduced from this maximum by the degree to which there is variability or noise $\mathcal{N}$ in the repeated responses to an identical stimulus, as measured by the spike train noise entropy, $H(\mathcal{N})$. This is estimated from the probability distribution of spike responses to multiple, identical presentations of the same stimulus, averaged over stimuli.

With the M-sequence, responses to the repeated presentations of each k-frame stimulus sequence were easily obtained. For each occurance of a specific k-frame sequence, the response beginning at a delay τ (ranging from 0 to 130 ms) relative to the onset of the initial frame of the sequence was divided into bins of size $\Delta\tau$ (usually 1 ms) containing the number of spikes in each bin. These bins were combined to form spike "words" of length $T = M\Delta\tau$, where M was an integer number of bins. For example, for $M = 3$, the joining of three bins containing 2, 0, and 1 spikes, respectively, would yield the word 201 (note that the absence of spikes in a bin can be informative, and its contribution was included).

We then computed the entropies for each choice of k, T, and $\Delta\tau$ by building the probability distribution of these words – across the whole experiment for $H_{k,\Delta\tau,T}(\mathcal{R})$, and across the multiple repeats of the i^{th} k-frame stimulus sequence ($i = 1, \ldots, 2^k$) at time-shift τ for $H_{i,\tau,k,\Delta\tau,T}(\mathcal{N})$. Note that the location of a transition, t, within the k-frame sequence was

now irrelevant and not specified; instead all k-frame sequences contributed equally to this analysis. Both T and $\Delta\tau$ were varied to obtain estimates of the entropy on different time scales. For a given T and $\Delta\tau$, the average information about the k-frame sequence that began at time τ before a response word was then given by $H_{k,\Delta\tau,T}(\mathcal{R}) - \langle H_{i,\tau,k,\Delta\tau,T}(\mathcal{N})\rangle_i$, where $\langle H(\mathcal{N})\rangle_i$ was the average noise entropy across all k-frame stimulus sequences (*i.e.*, average over i). We assigned the information about k-frame sequences, for the given T and $\Delta\tau$, as the maximum information across τ (see below).

First, though, for each combination of T, $\Delta\tau$, k, and τ, we corrected for finite-data errors. This was done by computing the mutual information for different partitions of the data: the whole data set, and the average over each half of the set, over each third, and each fourth. This average information was then plotted as a function of the number of partitions N, and fit to the functional form, $I = I_0 + I_1/N + I_2/N^2$ (Strong et al. 1998b). I_0 therefore represented the true information rate extracted from the limit of infinite data for a given T, $\Delta\tau$, k, and τ. Note however that when the amount of the data was too small, even this correction failed. Empirically, this occured when the ratio of I_2 to I_0 became large. We used a ratio of 2×10^{-3} as the border between sufficient and insufficient data, and show results only for cases in which data was sufficient by this criterion. In practice, the corrections for finite data were typically tiny, and the point of this procedure was primarily to screen out cases (*e.g.*, too-large k or too-large T) for which data was insufficient.

Given these corrected informations, we assigned the information about k-frame sequences as follows. For the given k, T and $\Delta\tau$, we determined the τ that maximized the information. The information, I, was then assigned to be the average information over the bins within ± 4 ms around this maximum. (We chose this to correspond to about a frame width, so that averaging smoothed out any frame-related artifacts.) The information rate of the spike train, in units of bits/time, was $I/(M\Delta\tau)$. We converted this to units of bits/spike I_{sp} by dividing by the neuron's average spike rate, r, assessed over the entire two-M-sequence stimulus: $I_{sp} = I/(rM\Delta\tau)$.

This method worked well only for relatively short response words. Long response words

required long stimulus sequences to minimize the randomizing effect of different stimulus contexts on early or late portions of the response word. However, since each sequence repeated $2 \times 2^{14-k}$ times, as k increased, our estimate of the entropies degraded due to sampling problems. Thus, to consider very long response words, we employed a different strategy: we estimated a *lower bound* on the information carried by the spike train about the stimulus by applying the direct method to the two repeats of the full M-sequence. Assuming that the only thing in common between the two presentations of the M-sequence was the stimulus itself, and that therefore the noise in the two cases were uncorrelated, the information that one response $\mathcal{R}_1$ carried about the second response r_2, $I_{\Delta\tau,T}(\mathcal{R}_1, \mathcal{R}_2)$ should be a lower bound to the information between either response $\mathcal{R}$ and the stimulus $\mathcal{S}$, $I_{\Delta\tau,T}(\mathcal{S}, \mathcal{R})$ (Strong et al. 1998b). We took each response to be the spike train generated by each full M-sequence, minus the first and last 200 ms. We then computed each spike train's entropy, $H_{\Delta\tau,T}(\mathcal{R}_i)$, $i = 1, 2$, for words of length T, and the joint entropy, $H_{\Delta\tau,T}(\mathcal{R}_1, \mathcal{R}_2)$, for the co-occurence of words in the two spike trains. These were computed from the probability distributions for words by using overlapping intervals (incremented by $\Delta\tau$, to increase the effective number of samples). To correct for finite-data errors, data size scaling was applied in this case directly to the entropy estimations (rather than to the mutual information as in the data size scaling described above); an example is shown in Figure 3A. The mutual information between the two responses was then

$$I_{\Delta\tau,T}(\mathcal{R}_1, \mathcal{R}_2) = H_{\Delta\tau,T}(\mathcal{R}_1) + H_{\Delta\tau,T}(\mathcal{R}_2) - H_{\Delta\tau,T}(\mathcal{R}_1, \mathcal{R}_2). \quad (1)$$

In general, the dependence of the information on word length T for a given bin size $\Delta\tau$ was small. Hence, to summarize the dependence for a particular bin size, the infinite-word-length limit was taken by obtaining a linear fit to the plots of the (infinite data limit) entropies vs. $1/T$, and using the y-intercept as the (infinite word limit) entropy rates in the calculation of the information rate. The fit was performed only over the range of $1/T$ where sufficient data was available to accurately estimate the entropy rates, as illustrated in Figure 3b. In practice, T's ranged from 8 ms to 48 ms. Finally, the information per second from words of spikes was converted into the information per spike by dividing by the mean spike

rate across the whole experiment.

Models

We constructed quasi-linear threshold models of driven LGN spiking activity to investigate whether the observed precision could be explained by simple mechanisms. All models convolved the full M-sequence stimulus, binned at one-sixth the frame period, with the cell's temporal kernel to generate a firing function, $f(t)$ (linear part). These responses were thresholded and perhaps squared (nonlinear part) to generate firing rates $r(t)$, as follows. We defined $r(t) = \lambda_\theta([f(t) - \theta]^+)^p$, where $[x]^+ = x$, $x > 0$; $= 0$, otherwise; $p = 1$ for a linear function and $p = 2$ for a quadratic function; and λ_θ was chosen to make the mean of $r(t)$ equal to the observed mean firing rate. The value of the threshold θ was fit as described below. Finally, spikes were generated as a Poisson process from these rates, perhaps along with a refractory period, as will be described below.

The temporal kernel was determined as the spike-triggered-average stimulus, divided by the autocorrelation (or in Fourier space, the power spectrum) of the M-sequence stimulus (the power in the M-sequence at frequency f is proportional to $(\sin(f/r_f)/f)^2$, where $r_f = 120$ Hz is the frame rate). This division yields the linear filter that, applied to the stimulus, gives the best estimate of the response in the sense of least mean-square error (Rieke et al. 1997). The spike-triggered average and temporal kernel for one cell can be seen in Figure 1. The division is done in Fourier space, where it simplifies to a simple frequency-by-frequency division; otherwise it would involve multiplying one matrix by the inverse of another matrix. However, one does not want to continue dividing up to arbitrarily high frequencies where the power in the stimulus approaches zero, as this will just amplify high-frequency noise. We chose to do the division up to some cutoff frequency, and to set all power above that cutoff frequency to zero. To choose a cutoff frequency, we tried cutoffs from 75Hz to 100 Hz in 5 Hz steps. For each cutoff, we applied the corresponding filter to the M-sequence to obtain the output $f(t)$, converted this to a rate function $r(t)$ as described above using $p = 1$, and chose the threshold θ as that which minimized the mean-square error difference between

the predicted Poisson rate function and the eight-frame PSTH for the actual data. We then chose the cutoff frequency that gave the least mean-square error; this best cutoff was 90 Hz. This kernel was used subsequently in all models to draw actual spikes for PSTH comparison (see below).

The conversion from $r(t)$ to spikes was as follows. We interpolated $r(t)$ to achieve a temporal resolution of 1/60 of a frame (the spike-triggered average and temporal kernel had been computed in bins of 1/6 of a frame or about 1.39 ms). For the simple Poisson case, spikes were then generated in each time bin Δt with probability $r(t)\Delta t$, using $\Delta t = 139\mu$s. For the case of a Poisson process with a refractory period, a free firing rate, q, (Berry and Meister 1998) was generated assuming a specific refractory period, μ, by taking $q(t) = r(t)/(1 - r(t)\mu)$. Spikes were then drawn as in the Poisson case but using $q(t)$ rather than $r(t)$. In the case of only an absolute refractory period, the probability of a spike was set to zero for μ ms after each spike. We also tried adding an exponential recovery after the absolute refractory period, setting $\mu = \mu_{\text{abs}} + \mu_{\text{rel}}$, where μ_{abs} was the absolute refractory period and μ_{rel} was the exponential recovery of the probability from zero up to $q(t)$. This implementation for a relative refractory period is reasonable when μ_{rel} is smaller than the characteristic time over which the firing rate remains relatively constant.

For each of the models, an optimal threshold and refractory period(s) (if applicable) were selected simultaneously to minimize the mean-square error between the eight-frame PSTH's (taken over the 25 $\sim$1.39 ms bins after the onset of the sixth frame, where frame 1 is earliest in time) for the real data and the model. This was done by trying every threshold from 1 to 5 in steps of 0.2, (if applicable) absolute refractory periods from 1 to 4 ms and relative refractory periods from 0.5 to 4 ms in steps of 0.5 ms for which $q(t)$ remained positive, and then selecting the combination of threshold and refractory periods that gave the least mean-square error. These ranges seem reasonable, because in no case was the optimum parameter at an extreme of the range explored for that parameter. The mean firing rate over the whole stimulus in the model was typically matched to within a few percent of the data's mean.

Results

Full-frame, binary, 14-bit M-sequence stimuli were presented at different contrast levels. In general, this stimulus drove cells in the LGN well. Average spike rates across all cells and stimulus conditions ranged from 4.6 to 25.3 Hz. Neural responses were usually triggered by transitions from either bright to dark frames (OFF cell), or vice versa (ON cell); we refered to two-frame sequences of bright/dark or dark/bright as an OFF or ON transition, respectively. Each cell's polarity was determined by reverse correlating the spike train with the M-sequence stimulus. Figure 1 presents the spike-triggered-average stimulus for one of our good OFF cells (cell 4, 80% contrast) that had a strongly driven response producing nearly 7000 spikes. We use this cell to illustrate the main results of our analysis. A spike at time zero for this cell was generally preceded by a transition from bright to dark around 32 ms earlier. This time delay was referred to as the cell's mean conditional latency. Figure 1 also illustrates the cell's temporal kernel (see Methods), which represents the cell's temporal receptive field and has the same 32 ms mean conditional latency; we will return to this later.

An initial 1200 frames (ten seconds) from the beginning of the M-sequence were presented to adapt the cells to the stimulus ensemble before showing the M-sequences used in data analysis. After the conditioning, two repeats of the full M-sequence were displayed without delay. A total of $2 \times 2^{14-k}$ repetitions of each k-frame sequence ($k \leq 14$) occurred, *e.g.* 128 repeats of each eight-frame sequence. Because of this convenient property, it was natural to focus on responses to the set of k-frame sequences for different k.

Mean Response: The PSTH Matrix

The M-sequence stimulus presented frames of random stimuli in series rather than in isolation. To obtain an average response to a specific stimulus sequence, we extracted the individual spike responses to the multiple presentations of that sequence in the full M-sequence. Consider first the case of one-frame stimuli. The average response to single bright or dark frames of stimuli was generated in the form of a matrix of PSTH's (Figure 4). The shading

in each one-millisecond bin corresponds to the total number of spikes from all presentations of this sequence at that time relative to the frame onset. Note that there was a nonzero spike rate even at the time origin that was nearly the same for both bright and dark frames. This reflects the fact that at early times, the spikes were responses to earlier frames over which we had averaged. The response to the particular bright or dark frame was most clear around 32 ms, as expected from the cell's mean conditional latency.

One advantage of visualizing a PSTH matrix is in the ability to display the neuron's average responses to stimuli more complex than just a single frame, as shown in Figure 5 for two-frame sequences. This clearly shows that spikes tended to be generated near the mean conditional latency in response to an OFF transition (stimulus 2), whereas spiking was clearly suppressed near the mean conditional latency by an ON transition (stimulus 1). Note that the response to a dark frame (stimulus 0 in Figure 4) was now broken down according to whether the preceding frame was dark or bright (stimuli 0 and 2, respectively, in Figure 5).

Figure 6 displays the PSTH matrix (with 1 ms time bins) for the response to seven-frame sequences, sorted according to the rightmost two frames, f1 and f2 (we usually numbered frames in a k-frame sequence consecutively as fN, $N = 1, \ldots, k$, with f1 the latest in time and fk the earliest). This grouped together all responses to sequences with an OFF transition in the most recent two frames. As expected, a large vertical band of spikes centered approximately around 32 ms appeared in response to the OFF transition. One striking feature was the slight slant in time of the OFF response band near 32 ms. Qualitatively, for this cell, the time to the first spike was correlated with the amount of time the stimulus had been bright prior to the final transition to dark: the longer this time, the earlier the occurence of the first spike in the response.

Moreover, the spikes in this band were noticeably isolated in time on both sides by regions of virtually no spikes, suggesting that there was a high degree of temporal precision in the response when seven frames of the stimulus were specified. To examine this, each spike should ideally be classified as part of a response to a particular sequence. In the PSTH

matrix though, each spike occurred multiple times, each time associated with a different time frame and sequence. Hence, echos of the main OFF response appeared in the other quadrants of the PSTH matrix where an OFF transition occurred earlier in the sequence.

Event classification

To classify a spike to a unique sequence, a search was performed to find the OFF transition that was most likely to be responsible for a given spike. All spikes classified to the same transition were then grouped together as the spike "event" in response to the sequence containing that transition (see Methods). In practice, this algorithm reproduced the event structure quite well, as can be seen from the comparison of Figures 7 and 8. These show the PSTH matrix and the extracted unique spike events, respectively, for the 1/4 of eight-frame sequences having an OFF transition in their final two frames. The band of spikes near 32 ms was clearly reproduced in the spike events. Virtually all spikes in the train were accounted for by this technique; only 1.8% of the spikes were unclassified. (Note that spikes placed at random would show 5/16, or 31%, unclassified.)

In general, for the group data across all cells, 10 out of 21 trials had unclassified percentages less than 5%, while for the remaining 11 trials this was larger than 5%. Qualitatively, the unclassified percentage was correlated with the degree to which spikes were locked to the stimulus, as evidenced by visual isolation of spikes around the mean conditional latency in the PSTH matrix. When the spikes around the mean conditional latency could be visibly isolated (10 out of 21 trials), the algorithm appeared to yield fairly low unclassified percentages (9 out of those 10 trials). The one exception was a 40% contrast trial for an ON cell in which the events in response to an ON transition were fairly well isolated, yet the unclassified percentage was nevertheless high (26%), probably because spikes were also produced without a transition when the stimulus had been bright for several frames. In cases when locking was evident but poor (5 out of 21 trials had bands of increased spiking, but these were not well isolated), or when spiking was more indiscriminate (6 out of 21 trials had poorly distinguishable bands), the unclassified percentage tended to be larger (10 of these 11

trials had unclassified percentages above 5%). The one exception was a 6% contrast trial for an OFF cell with a weak linear kernel – its events were not well isolated, but its unclassified percentage was nevertheless low (3.5%).

For each sequence, we defined its *event probability* to be the percentage of its occurrences that evoked an event of one or more spikes.

Response Variability I: Spike Timing Precision

Using the binary k-frame sequences to characterize the stimulus, and the spike events to characterize the response, we turn to the next issue of this paper: a study of the reliability and precision of responses and their dependence on the stimulus. The timing precision of these events was examined by determining the jitter in the time of the first spike in the events associated with a particular sequence. This is shown in Figure 9a for the only possible two-frame sequence with an OFF transition. This sequence generated a spike response 49% of the time, and the time of the first spike had a standard deviation of 3.25 ± 0.04 ms. Since the responses to all possible combinations of stimulus frames before and after the two frames of the transition were averaged together, this standard deviation represented the precision achieved by the two frames of the OFF transition alone, when the other frames were unspecified. Its value was already less than the standard deviation expected (7.2 ms) if the first spike times were distributed uniformly over the three-frame search window that defined events.

When eight frames of the stimulus were specified, with an OFF transition occurring between frames f2 and f1 (which we denote as an "f2—f1 transition"), every sequence had less than 3 ms standard deviation in the time to the first spike, including sequences with both high and low event probability (Figure 9B). The median across sequences of this standard deviation was 1.28 ± 0.13 ms (median $\pm$ square root of the jacknife variance for the median sequence). These results suggested that a significant part of the timing jitter in response to two-frame stimuli was simply due to imprecise specification of the stimulus history. In particular, the average first spike times differed significantly for different eight frame sequences,

decreasing with increasing event probability, as plotted in 9c. This naturally broadened the width of the distribution of first spike times when the responses to different stimulus sequences were averaged together.

Since the timing precision clearly varied with the exact stimulus sequence, we wanted a more generic measure of the overall variability of the response for a given level of stimulus specification. We selected the median, over sequences, of the standard deviation of the time to the first spike in an event as a robust index for this purpose. Figure 10 plots this median standard deviation as a function of the number of frames specified. The upper and lower interquartile ranges are also depicted, showing that in some cases, the distribution of standard deviations is clearly asymmetric. Three curves are shown, corresponding to the location within the sequence of the OFF transition. The f2—f1 curve (OFF transition between f1 and f2) indicates that the median precision improved until three to four frames before the two frames encompassing the OFF transition were specified (five to six total frames). When four frames before the transition were specified, specifying additional frames after the transition (seven frames on the f3—f2 curve or eight frames on the f4—f3 curve) did not substantially alter the median precision, suggesting that these frames had little effect on the overall timing precision. This plateau in the precision likely reflected the intrinsic variability of the cell, since further stimulus specification did not further increase the precision. In the group data, the precision of all trials with less than 5% unclassified percentage (with the exception of a 6% contrast trial) improved with increased specification of the frames before the transition; the median standard deviation decreased on average by $41 \pm 13\%$ ($n = 9$) from the case where only two frames were specified to the case where eight frames were specified, with the ON or OFF transition between frames f4 and f3. When all trials were considered regardless of % unclassified, a decrease of $31 \pm 18\%$ ($n = 21$) was found on average.

Figure 11a plots the dependence of the median standard deviation of the time to first spike on the percentage of unclassified spikes across the population of cells, for eight-frame sequences with f4—f3 transitions. The cluster of trials having unclassified percentages below 5% clearly exhibited high timing precision (with one exception for a 6% contrast trial)

– mean of 1.56 ± 0.39 ms (mean $\pm$ std, $n = 9$, excluding the outlier). Two trials (one cell at 40% contrast, another at 80% contrast) had high unclassified percentages (26% and 22%, respectively) but nevertheless had small median standard deviations (1.97 ± 0.06 and 2.28 ± 0.13 ms, respectively). The remaining nine trials that had more than 5% unclassified spikes clustered around 4.01 ± 0.58 ms (mean $\pm$ std). Many of these trials were less well driven, as evidenced by their generally lower firing rate, as shown in Figure 11c. Since this group of trials often responded more diffusely in time, making classification of spikes difficult, their poorer precision was not surprising. However, given that their precision was well below the 7.2 ms expected from random placement of spikes, it seems likely that this reflected a true property of the cells rather than an artifact of the classification method.

The timing precision showed only weak dependence on the event probability, that is, on the reliability with which a sequence evoked a response. Within the group data, the Spearman rank-order correlation was statistically significant ($p < .05$) when both all of the data and part of the data was analyzed (see Methods) in only 8 out of 21 trials. It was not significant for both conditions in another 7 out of 21 trials. In the remaining 6 trials, the significance level changed between the two conditions. The fact that 13 of 21 trials showed no clear correlation suggested that the dependence of first-spike-time standard deviation on event probability was not strong. That is, the temporal precision of response was not simply a result of a "strong" stimulus: even responses that were infrequently evoked could nonetheless be evoked at fairly precise times when they did occur. Hence, reliability and timing precision were not strongly coupled.

Response Variability II: Spike Count Precision

The timing precision analysis focused on how the stimulus affected the jitter of a single spike (namely the first spike in an event). To study the precision of the remaining spikes in an event, we analyzed the precision of the number of spikes in the events evoked by a stimulus. This spike count precision was characterized by examining the variance in the number of spikes per event vs. the mean number of spikes in an event. In the case of a Poisson process,

the variance is equal to the mean. At the other extreme, the minimum possible variance for a discrete counting process with a given mean m is obtained if the number of spikes in every event is either ceil(m) (the smallest integer $\geq m$) or floor(m) (the largest integer $\leq m$). This minimum variance varies periodically with the mean, dropping to zero at each integer and forming a scalloped curve between integers.

Figure 12a plots cell 4's spike count variance for the single two frame OFF sequence against its mean spike count. Also shown are the line expected for a Poisson process, and the scalloped curve representing the minimum possible variance. The variance for this sequence clearly fell close to the Poisson limit. When the stimulus history specification was expanded to eight frames, with the OFF transition between f2 and f1 (Figure 12B), most of the sequences remained Poisson-like, but a few began to have sub-Poisson responses. However, if we consider eight-frame sequences with the OFF transition between the f4 and f3 frames (Figure 12C), meaning that we specify two frames after the transition frames as well as four frames before, the variance for almost all sequences was significantly less than Poisson, falling in many cases close to the minimum-variance limit. These results were summarized by examining the median across sequences of the Fano factor, which is the ratio of the variance to the mean (Figure 13). For an f2—f1 transition, the Fano factor remained near the Poisson limit of one, regardless of the number of frames specified before the transition. However, if one or two frames were specified after the transition frames (f3—f2 transition or f4—f3 transition, respectively), then, the median Fano factor fell dramatically with increasing sequence length, reaching a value of 0.315 ± 0.037 (median $\pm$ square root of the jacknife variance for the median sequence) for eight-frame sequences with f4—f3 transition. This was a 72% reduction from the two-frame case.

The marked suppression of the noise by the specification of the frames after the transition can be straightforwardly understood. Figure 9b shows that when no frames were specified after the transition, there was a cluster of sequences that had event probabilities near 50%, regardless of how many frames were specified into the past. Once one frame was specified after the transition frames, however, the event probabilities diverged so that many sequences

produced events with nearly unit probability, while other sequences produced events with very small probability. This reflected the fact that, for this cell, the event produced by a transition from bright to dark (10) could be suppressed by a subsequent transition back to bright (101). On the other hand, if no subsequent transition occurred (100), an event was virtually always produced. This occurred almost irrespective of what happened before the OFF transition. The variance in spike count for each of the two cases (101 and 100) could be small. However, by not specifying the frame after the transition, as in the f2—f1 curve of Figure 13, the two cases were averaged together, producing a large variance, and a Fano factor close to unity. Thus, simply increasing the stimulus history was not always enough to obtain precise responses; enough frames both before and after the OFF transition had to be specified to maximize the precision (note that frames after the transition are still within the causal range where the linear kernel is sensitive to the stimulus).

In the group data, such large reductions in the spike count Fano factor were not very common. Comparing the median Fano factor for the two-frame case to the eight-frame, f4—f3 transition case, there was on average a $27 \pm 21\%$ ($n = 21$) reduction across all trials. Considering only those trials with unclassified percentages under 5% yielded a $38 \pm 24\%$ ($n = 10$) reduction; trials with unclassified percentages greater than 5%, $17 \pm 9\%$ ($n = 11$). The Fano factor itself was generally around or below 1 in nearly all cases, as shown in Figure 14. No strong dependence of the Fano factor on the unclassified percentage appeared in the data, except that Fano factors less than 0.5 occurred only in trials with less than 5% of spikes unclassified. Moreover, note that, among the low-unclassified-percentage trials, good timing precision did not necessarily imply good count precision (compare Figure 11a to Figure 14).

Information transmission

The spike timing and count variability measures discussed above gave some indication of the precision of LGN neurons. How much information did this level of precision allow the cells to transmit?

To address this, we changed our analysis method. The above analyses of variability

depended on defining events that associated each spike with a unique sequence that evoked it. This required specifying both sequence length and the location within the sequence of the transition (since spikes were associated with transitions and these two facts uniquely linked transitions to sequences). For the information analysis, we instead considered *all* sequences of a given length, without regard for the presence of a transition, and simply examined the response at some fixed time interval after the initiation of the sequence.

We computed information using the direct method (see Methods). We binned time into discrete units of size $\Delta\tau$, typically 1 ms, and defined the "letters" of the response "alphabet" as the number of spikes in a bin (0 or 1 for 1 ms bins). A string of M such letters formed a response "word" – for $M = 1$, the word was simply the number of spikes in a single bin. Ideally, the choice of bin size should reflect the degree of temporal resolution in the code, while the word size should reflect the longest time scale of temporal correlations in the code. The timing precision analysis suggested that a reasonable bin size was about 1 ms. Initially ignoring correlations between bins, we calculated the information about k-frame stimuli by considering only single-bin words at this resolution (Figure 15). The information grew with time from the onset of the stimulus sequence, provided that further stimulus frames continued to be specified, up to at least nine frames. At this point, the maximum information was about 3.5 bits/spike, and appeared to be nearing a plateau. The existence of a plateau was reasonable since a given response time bin should give little or no information about stimulus frames that occurred far in the past. For longer sequences ($k \geq 4$), the information began to drop from its peak at about 24-26 ms after the onset of the last frame in the sequence, or about 16-18 ms after the onset of the first unspecified frame. This suggests that 16-18 ms was the minimum delay for a frame to significantly influence the response. This was in rough agreement with our previous results that one and perhaps two frames after the transition frames can influence the spike count by vetoing or allowing spikes induced by the transition; if the response occurs 32 ms after the transition, then these frames would have onsets about 15 and 24 ms before the response that they influence.

We also compared the maximal observed information rate of 3.5 bits/spike to the cell's

maximum possible information rate, as measured by the entropy of its spike train. Achieving this maximum would imply that all of the cell's response variability (as measured in single 1 ms bins) was used to encode the stimulus. In fact, the coding efficiency, the ratio of the actual information coded to that which could possibly be encoded, was approximately 51% (for $k = 9$), so that the cell transmitted information in individual 1 ms bins at a level that was within a factor of two of its limit.

We next examined the role of time resolution on information encoding by varying the bin width. We considered 8 ms words of the spike train, and binned these words using either 1, 2, 4, or 8 ms resolution. If the precise timing of the spikes at these resolutions within the word were important for transmitting information, then we expected more information at smaller bins than larger bins. Finer resolution increases the possible information the spike train can code; if the actual information coded also grows, then the coding efficiency would not significantly change with increasing resolution. On the other hand, a falloff of the coding efficiency would indicate that the increased resolution is not being used to code information. We computed maximum information rates for eight-frame sequences, to ensure that there were sufficient repeats of each sequence to allow us to estimate the information for multiple-bin response words. The information rate increased from 2.4 bits per spike at 8 ms bins to 3.1 bits per spike at 2 ms bins, a 29% increase (Figure 16A), while the spike train entropy increased by 36% over the same range. That is, $\frac{0.29}{0.36} = 81\%$ of the increase in entropy associated with this increase in resolution was used to encode information. As a result, the coding efficiency stayed relatively flat, decreasing only about 5% from a bin size of 8 to 2 ms. Thus, the position of spikes at up to 2 ms resolution was significant for coding information. Improving the resolution by a factor of 2 from 2 to 1 ms yielded an additional 3% increase in information to 3.2 bits per spike, compared to an increase in entropy of 15%, suggesting that only 20% of the entropy change encoded information. Thus, while more information was encoded at this finer resolution, there was a diminishing return as the noise became a proportionately larger contributor to the cell's increased variability.

Redundancy or Synergy in Coding

Given that a temporal resolution down to 1 ms was useful, another important question to address is the manner in which patterns of spikes in these bins contributed to information transmission. Three possibilities exist: different 1 ms bins may code information independently; they may encode information redundantly, so that M-bin words code less information than M times the one-bin-word information; or they may interact synergetically, so that M-bin words code more than M times the one-bin-word information. Note that the degree of redundancy or synergy may change with M – for some word sizes, the responses may be more redundant, whereas for other word sizes, they may become synergetic.

We investigated the degree of synergy and redundancy in the LGN responses in three ways. First, we compared the information in 8 ms words with 1 ms bins to that found in 1 ms words. For our example cell, the 3.2 bits per spike for 8 ms words with 1 ms bins was close to the 3.3 bits per spike for 1 ms words found for eight-frame sequences in Figure 15, indicating only a little redundancy and no synergy between the responses of adjacent 1 ms bins. This near independence of spikes in 8 ms words was not simply due to Poisson firing since the distribution of the number of spikes within the 8 ms window showed a much larger probability for 2 spikes (0.21) than would be expected from the square of the 1 spike probability (0.008). This result suggests that the cell was bursting, although the bursts apparently did not lead to a large level of redundancy. This could happen because redundant patterns (such as bursts) might be used synergetically to code for the stimulus. It is important to point out that our measure looks at the *average* level of redundancy or synergy so that the combination of different groups of redundant and synergetic spikes could appear independent at this time scale.

Second, to determine whether this lack of significant redundancy or synergy survives at longer time scales, we examined the information in much longer words. Unfortunately, the direct method as applied to the repetitions of the k-frame sequences could not be used to study response word lengths longer than about 8 ms (for eight-frame stimuli), due to data insufficiency. Instead, we estimated a lower bound on the information in the entire response

to the full M-sequence, by using the two repeats of the full M-sequence (see Methods). Assuming that the stimulus was the only common drive for the two responses, then the information between the responses to the two repeats bounded from below the information either could carry about the stimulus. We compared this lower bound extracted in the limit of infinitely long response words to the exact information rates computed from eight-frame sequences and 8 ms responses (Figure 16). The lower bound came reasonably close to the information rates computed from 8 ms words across all bin sizes considered. This implies that there is little redundancy over times longer than 8 ms, but leaves open the possibility of synergy (if the true infinite-word information were much higher than our lower bound).

As a final test of the redundancy or synergy between spikes, we compared the exact information transmitted by individual 1 ms bins to the lower bound on the information transmitted by infinitely long words of 1 ms bins. Figure 17 plots these two measures for all cells and trials in the data, in terms of both bits per spike and per second. Nearly all of the trials fell close to the diagonal line where the information from single bins equaled the lower bound to the information from infinitely many bins. To the right of this diagonal line, coding is synergetic: more information is conveyed on average by combinations of spikes in 1 ms bins than by single spikes in 1 ms bins. To the left of this diagonal line, coding is redundant: less information is transmitted on average by the words of bins than by the single 1 ms bin. Hence, the fact that the trials aligned close to, but to the left of, the diagonal suggested that there was at most a slight amount of redundancy for real cells. Since the infinite word information is a lower bound, we can only be certain that the true information lay to the right of the plotted data – that is, there was little or no redundancy, and possibly some synergy.

Comparison to quasi-linear threshold models

The results reported here suggest that, while there was variation among the population, visual thalamic cells could exhibit very precise responses that conveyed considerable amounts of information per spike on average. Responses of these cells are often modelled as resulting

from the convolution of the cell's temporal kernel (shown in Figure 1) with the stimulus, followed by a nonlinear thresholding to generate a firing rate (see Methods). Is the degree of precision consistent with this picture?

We first chose the threshold that gave a best match of the model PSTH to the data, and assumed that spikes were generated randomly according to an inhomogenous Poisson process with the model PSTH (appropriately scaled to yield the same mean rate as the data). The best-matched model gave a broader and more symmetric PSTH than was observed in the data, suggesting that the precision of model spikes was significantly worse. Extracting spike events as described above allowed us to directly compare the precision of the model (Figure 18B) to the data (Figure 18A), in response to eight-frame sequences that had a bright to dark transition in the rightmost two frames. The model spike events were clearly more diffuse in time, and did not capture the details of the dependence of response onset times on stimulus sequence. We next considered models in which spikes were generated from a Poisson process with an absolute refractory period. We considered this for the case in which the firing rate was a linear function (Figure 18C) or a quadratic function (not shown) of the thresholded filter output. In each case, refractory period and threshold were chosen together to optimally match the data (least mean-square error in PSTH). The PSTH of the linear refractory model was slightly narrower than that derived without a refractory period, but continued to be wider than the data and to not show the temporal irregularity of the data. The model using a quadratic function gave results similar to, but slightly poorer than, those of the linear refractory model, so we do not consider it further. A relative refractory period in addition to an absolute refractory period also yielded quantitatively similar results as the case of an absolute refractory period alone.

The model's failure to capture the detailed structure of response onset times is specifically due to an underestimation of longer onset times, while shorter onset times were well reproduced by the model (Figure 19A). This discrepancy can be understood from an examination of the PSTH matrices (Figure 18): it appears that when two or more consecutive dark frames preceded the bright to dark transition, this lengthened first-spike times in the data, but this

effect was not picked up by any of the models. The models also reasonably reproduced the mean spike counts observed in the data, but showed a tendency to underestimate smaller mean counts and overestimate higher ones (Figure 19B).

The Poisson model did a poor job of reproducing the observed variability in spike timing or spike count (Figure 20, black dots). The inaccuracy in spike count precision is not surprising, because a Poisson model will always have a Fano factor of 1. However, the model incorporating a refractory period came much closer to reproducing the precision of the data (Figure 20, white triangles). This model tended to slightly overestimate smaller first-spike-time standard deviations and Fano factors and to underestimate larger ones, showing less overall diversity of first-spike-time standard deviations and Fano factors than the data.

Discussion

We have found that LGN neurons can show great precision in their responses to M-sequence stimuli. For at least a subset of cells, spikes occur in discrete events triggered by an ON or OFF transition, with spike rates close to zero at other times. The time of the first spike in an event can be precise to 1-2 ms, and this precision can be maintained even for unreliable events (events that occur with low probability). The four frames before the transition frames influence the event timing, so these frames must be specified to discern the cell's spike timing precision. The number of spikes in an event can also show great precision, with Fano factor (ratio of variance to mean) approaching 0.3 (vs. a value of 1 expected for a Poisson process). The frames after a transition can "veto" or allow an event, so that two frames after the transition frames as well as the four before must be specified to discern the cell's spike count precision. This precision of response allows cells to carry up to 3.5 bits/spike of information about the stimulus. The coding efficiency of information transmitted in 1 ms bins can be within a factor of two of the limit set by the spike train's entropy – a limit that is achieved when all of the cell's variability is used to code information. The coding efficiency remains relatively constant as the temporal resolution for specifying spike times increases to at least 2

ms, and still more information is gained by decreasing resolution to 1 ms, indicating that the timing of spikes at these resolutions carries information about the stimulus. By comparing the information carried by 1 ms response words to that in 8 ms words and to the lower bound on the information transmitted by infinitely long response words, we find that there is at most only a modest amount of redundancy in the coding by successive spikes, and we find no evidence for synergy. Finally, this precision can be largely, but not entirely, accounted for by a model in which firing rate is generated by filtering the stimulus with the cell's temporal kernel and applying a threshold, followed by spike generation as a Poisson process with an absolute refractory period.

Previous work on spike timing and count precision

Our work adds to a growing body of work finding high response precision and high information rates in the LGN in response to full-field noise stimuli. Keat et al. (2001), in work contemporary with the present work, found 1-2 ms standard deviation for the time to the first spike in an event in response to full-field Gaussian white noise, in close agreement with the present results for binary white noise. Reinagel and Reid (2000) reported a particularly low width (standard deviation) of 0.6 ms for one PSTH peak in one cell's response to a full-field "naturalistic" noise stimulus, but did not more generally report on timing precision. Both of these papers and (Kara et al. 2000) demonstrated sub-Poissonian Fano factors in LGN responses to full-field Gaussian noise, in agreement with the present findings. Comparable precision of spike timing and count in response to full-field noise stimuli has been reported in the retina (Berry and Meister 1998; Berry et al. 1997; Kara et al. 2000; Keat et al. 2001).

Measures of response to other stimuli often do not show similar precision. Thus, Guido and Sherman (1998) measured the jitter in the time to first spike in responses to spots flashed in the center of the LGN cell receptive field, and reported standard deviations ranging from about 3 to 35 ms, depending on the mode of firing (burst vs. tonic). The greater variability seen in this case of a single flashed spot is akin to the spread of the PSTH seen when only a single frame is specified (Figure 4), and may reflect the lack of specification of the

cell's initial state. That is, when stimulated only by a blank screen (before stimulus onset), spontaneous activities may lead a cell to wander through a state space of comparable diversity to that created by the set of binary stimulus sequences that could precede a single frame in our experiments. Similar reasoning might also explain why statically flashed, spatially nonuniform stimuli have produced Fano factors larger than 1 in several LGN studies (Hartveit and Heggelund 1994; Levine et al. 1996; Sestokas and Lehmkuhle 1988). Reich et al. (1997) reported a PSTH standard deviation of 5 ms for one LGN cell in response to a slowly drifting sine grating, but at least some of this jitter was due to a slow drift in response phase across many trials, which may have represented a slow change in cell state; responses over a small set of adjacent trials showed considerably greater precision.

Specifying Neuronal State

The idea that responses to temporally modulated stimuli can show great precision, even while responses to more static stimuli may show greater variability, has already a long history (*e.g.* Buracas et al. 1998; de Ruyter van Steveninck et al. 1997; Mainen and Sejnowski 1995) and has stirred controversy (*e.g.* Egelhaaf and Warzecha 1999). Our findings add a focus on stimulus history, showing that sufficient specification of a temporally varying stimulus is key to revealing neural precision. By extension, this emphasizes the importance of control of neuronal state: noise may not be intrinsic to a neuron or a piece of neural tissue, but may instead simply represent variables that are not under the experimenter's control. While a dynamic stimulus may control neural firing and thus control a given cell's state, lack of a stimulus (a blank screen) yields spontaneous activities that are stochastic, being triggered at least in part by spontaneous quantal events in photoreceptors (Mastronarde 1989), and these in turn may lead a cell's state to wander in an uncontrolled way, presenting an uncontrolled initial condition at the moment of a flashed stimulus. A related argument was made by Buracas et al. (1998), who showed that whether or not a given stimulus evoked a spike in a cell of area MT was strongly correlated to the local field potential at the given time and place.

It is interesting that, at least for our binary stimuli, specification of 8 frames (67 ms) seems adequate to specify LGN state to sufficient precision to saturate spike timing and count precision, while 9-10 frames (75-83 ms) saturate the information coded by spikes. These numbers are in rough agreement with the width of the cell's temporal kernel (Figure 1), which differs from zero over a span of about 65-70 ms.

Previous work on neuronal information transmission in the LGN

The information rates we have found – 2-3.5 bits/spike, 20-90 bits/sec – are similar to those found by others in LGN who, like us, used "direct" methods (Eckhorn and Pöpel 1975; Reinagel and Reid 2000). These methods directly estimate the information carried by the spike train about the stimulus, without a requirement for explicit decoding, by assaying certain stimulus and response probability distributions (Eckhorn and Pöpel 1974; Strong et al. 1998a,b). Indirect methods, such as the stimulus reconstruction method (Rieke et al. 1997), rely on being able to "decode" the response. These methods provide only a lower bound to the information rates: any information that is successfully decoded was present, but there is no guarantee that all information that was present was successfully decoded. Rates found using indirect methods in LGN have generally been quite low – only about 2 bits/second (Dan et al. 1998; McClurkin et al. 1991; Reinagel et al. 1999) – suggesting that much information present in the LGN spike trains was missed by those methods.

Stimuli better matched to the receptive field may yield more information. Eckhorn and Pöpel (1975) found that spatially uniform stimuli yield lower LGN transmission rates (25-40 bits/second at the best flash rate) than spots isolated at the receptive field center (60-80 bits/second) (Eckhorn and Pöpel 1975). Their spot and full-field stimuli were only briefly flashed at slow, periodic intervals (up to 30 Hz). Full-field stimuli modulated randomly at higher rates drive relatively high LGN information rates, as shown both by our work and that of Reinagel and Reid (2000). The latter sees a range of information rates similar to what we have found, even though their naturalistic stimulus distribution contained much more entropy than our binary distribution (924 bits/second in their distribution, vs. 120

bits/second in ours). That suggests that we may be seeing the limits of what an LGN cell can code, at least to full-field stimuli. On the other hand, the results of Eckhorn and Pöpel (1975) suggest that both we and Reinagel and Reid (2000) might have seen even higher information rates if we had restricted flashes to the cells' receptive field centers.

High information rates like those reported here have also been observed in a variety of other systems, including retina, visual cortex, and insect motion-detecting neurons (*e.g.*, Berry et al. 1997; Buracas et al. 1998; de Ruyter van Steveninck et al. 1997; Reich et al. 2000; Strong et al. 1998a), suggesting that the precision found here may not be a special property of LGN or thalamic neurons.

Minimal redundancy

The issue of redundancy or synergy in the neural code has been addressed in numerous papers, but we are aware of only a few (Brenner et al. 2000; Reinagel and Reid 2000) that have looked at the issue in terms of temporal coding in a single neuron rather than population coding across multiple neurons. Reinagel and Reid (2000) found that LGN neurons can sometimes code more information on average in patterns of spikes than if those spikes were considered independently. The synergy they reported, however, was at most only about 20%, and many neurons were slightly redundant (up to 10%) or only very weakly synergetic. Our results are consistent with this in the sense that we also observe at most only mild redundancies in the coding by individual neurons. We cannot rule out synergies, but to the degree that our lower bound closely approximates the true information, the fact that none of our neurons lay to the right of the independence line in Figure 17 suggests that there are also no large synergies in the coding by individual neurons.

Models of response generation

We have found that a simple model of response generation, based on thresholding the output of the cell's temporal kernel applied to the stimulus and imposing a refractory period, can match much but not all of the precision of response that we observed. To achieve this result,

it was critical that the cell's temporal kernel be used, and not simply the spike-triggered average; use of the latter gave noticeably less precision in both timing and spike count (not shown).

The discrepancies between the precision of the model and the observed data most likely arise from the linear filter model rather than the specifics of the spike generation mechanism. The PSTH matrix generated by the model is somewhat wider and considerably more regular than that of the data. Most strikingly, the model fails to show the lengthening of first-spike times observed in the data when two or more consecutive dark frames preceded the bright/dark transition. This yields a less "jagged" left edge for the model PSTH compared to the data PSTH. This jagged edge is dominated by the first spikes in a response, which are unaffected by refractoriness. Accordingly, the error is unlikely to be in our model of spike generation and refractoriness, but rather in the model of PSTH generation by linear filtering. This is also suggested by the fact that the temporal kernel and the spike-triggered average both give similarly smooth leading edges (data not shown), so that it seems unlikely that a better filter would alter this result. It is further suggested by the results of Kara et al. (2000), who found that they could successfully model the spike count variability of LGN cells by beginning with the observed PSTH (rather than deriving the PSTH from a filter as we are doing) and adding both absolute ($\sim$ 1 ms) and relative ($\sim$ 20 ms) refractory periods extracted from the cell's interspike interval distribution.

Accounting for the observed PSTH presumably requires a more complex nonlinearity in our model of firing-rate generation than the thresholding used here; it would be interesting to determine whether contrast-gain-control mechanisms (Shapley and Victor 1978; Victor 1987) might be sufficient to reproduce the response onsets and improve the agreement between the model and data precision measures. Nonetheless, it should be noted that the model as it stands is significantly nonlinear. The optimal threshold value (optimal in the sense of least mean-square error in matching the data PSTH) was 80% of the root-mean-square of the output of the filtering of the stimulus by the temporal kernel (see legend to Figure 18), that is, it was necessary to set a significant fraction of positive filter outputs to zero. The

optimal absolute refractory period was 3 ms, long compared to probable biophysical absolute refractory periods of ≈ 1 ms. (When both an absolute and a relative refractory period were used, the optimum was similar, 2.5 ms absolute plus 0.5 ms relative refractory period).

An alternative approach to modeling the neural responses observed here is to dispense with a firing rate model altogether and instead directly model the spike generation process. Berry et al. (1997) found that responses of retinal neurons to full-field noise stimuli consisted of brief response events surrounded by substantial periods of zero spike rate, similar to the cases in our experiment in which most spikes could be accounted for by response events locked to ON or OFF stimulus transitions. This has led the same group more recently (Keat et al. 2001) to suggest that a rate description of such responses, in which spike probability is zero for extended periods interrupted by brief events, may be inadequate. Instead they proposed predicting the spikes themselves rather than a spike rate, by regarding the output of a cell's linear filter applied to the stimulus as a voltage-like variable rather than a rate, and counting upward-going threshold crossings of this voltage as spike times. Parametrizing the filter, adding a spike-induced "hyperpolarization" to represent refractoriness, and adding appropriate noise yielded a 20-parameter model (15 parameters describing the filter, and 5 additional parameters). They showed that such a model, fit individually to each cell by optimizing a cost function incorporating precision measures, could do a good job of replicating the cell's spiking events and their statistics for both retinal and LGN cells in response to Gaussian noise stimuli. We have no reason to doubt that the same models would well describe the responses to binary noise stimuli studied here.

Conclusion

LGN cells can show remarkable precision in their responses and code information at high rates and with high coding efficiency. Revealing this precision requires sufficient specification of the stimulus history. This points to the possibility that measurements of neuronal precision may be limited as much by the degree to which the experimenter controls the variables relevant to a cell's response as by the intrinsic precision of neural processing.

Acknowledgements

This work was supported by a grant from the Sloan Foundation (ST and RCL), the University of California President's Postdoctoral Fellowship (RCL), and NIH grant R01-NS33787 and gifts from the Swartz Foundation.

References

Bair, W. Spike timing in the mammalian visual system. *Curr. Opin. Neurobiol.* 9:447–453, 1999.

Berry, M. J., 2nd and Meister, M. Refractoriness and neural precision. *Journal of Neuroscience* 18(6):2200–11, 1998.

Berry, M. J., 2nd, Warland, D. K., and Meister, M. The structure and precision of retinal spike trains. *Proceedings of the National Academy of Sciences of the United States of America* 94(10):5411–6, 1997.

Borst, A. and Theunissen, F. E. Information theory and neural coding. *Nature Neuroscience* 2(11):947–57, 1999.

Brenner, N., Strong, S. P., Koberle, R., Bialek, W., and de Ruyter van Steveninck, R. R. Synergy in a neural code. *Neural Computation* 12(7):1531–52, 2000.

Buracas, G. T. and Albright, T. D. Gauging sensory representations in the brain. *Trends Neurosci.* 22:303–309, 1999.

Buracas, G. T., Zador, A. M., DeWeese, M. R., and Albright, T. D. Efficient discrimination of temporal patterns by motion-sensitive neurons in primate visual cortex. *Neuron* 20(5):959–69, 1998.

Dan, Y., Alonso, J. M., Usrey, W. M., and Reid, R. C. Coding of visual information by precisely correlated spikes in the lateral geniculate nucleus. *Nature Neuroscience* 1(6):501–7, 1998.

de Ruyter van Steveninck, R. R., Lewen, G. D., Strong, S. P., Koberle, R., and Bialek, W. Reproducibility and variability in neural spike trains. *Science* 275(5307):1805–8, 1997.

Eckhorn, R. and Pöpel, B. Rigorous and extended application of information theory to the afferent visual system of the cat. I. Basic concepts. *Kybernetik* 16:191–200, 1974.

Eckhorn, R. and Pöpel, B. Rigorous and extended application of information theory to the afferent visual system of the cat. II. Experimental tests. *Biol. Cybern.* 17:7–17, 1975.

Egelhaaf, M. and Warzecha, A. K. Encoding of motion in real time by the fly visual system. *Curr. Opin. Neurobiol.* 9:454–460, 1999.

Gray, C., Maldonado, P., Wilson, M., and McNaughton, B. Tetrodes markedly improve the reliability and yield of multiple single-unit isolation from multi-unit recordings in cat striate cortex. *J. Neur. Meth.* 63:43–54, 1995.

Guido, W. and Sherman, S. M. Response latencies of cells in the cat's lateral geniculate nucleus are less variable during burst than tonic firing. *Visual Neuroscience* 15(2):231–7, 1998.

Hartveit, E. and Heggelund, P. Response variability of single cells in the dorsal lateral geniculate nucleus of the cat. Comparison with retinal input and effect of brain stem stimulation. *J. Neurophysiol.* 72:1278–1289, 1994.

Kara, P., Reinagel, P., and Reid, R.C. Low response variability in simultaneously recorded retinal, thalamic, and cortical neurons. *Neuron* 27:635–646, 2000.

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

Levine, M. W., Cleland, B. G., Mukherjee, P., and Kaplan, E. Tailoring of variability in the lateral geniculate nucleus of the cat. *Biological Cybernetics* 75(3):219–27, 1996.

Liu, R. C., Tzonev, S., Rebrik, S., Kurgansky, A., and Miller, K. D. Spike precision and information in cat visual thalamus. *Society for Neuroscience Abstracts* 26:1196, 2000.

Mainen, Z. F. and Sejnowski, T. J. Reliability of spike timing in neocortical neurons. *Science* 268:1503–1506, 1995.

Mastronarde, D. N. Correlated firing of retinal ganglion cells. *Trends Neurosci.* 12:75–80, 1989.

McClurkin, J. W., Gawne, T. J., Richmond, B. J., Optican, L. M., and Robinson, D. L. Lateral geniculate neurons in behaving primates. I: Responses to two-dimensional stimuli. *J. Neurophysiol.* 66:777–793, 1991.

Meister, M. and Berry, M. J. The neural code of the retina. *Neuron* 22:435–450, 1999.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P. *Numerical Recipes in C.* Cambridge University Press Cambridge second edition, 1992.

Reich, D. S., Mechler, F., Purpura, K. P., and Victor, J. D. Interspike intervals, receptive fields, and information encoding in primary visual cortex. *Journal of Neuroscience* 20(5):1964–74, 2000.

Reich, D. S., Victor, J. D., Knight, B. W., Ozaki, T., and Kaplan, E. Response variability and timing precision of neuronal spike trains in vivo. *J. Neurophysiol.* 77:2836–2841, 1997.

Reinagel, P., Godwin, D., Sherman, S. M., and Koch, C. Encoding of visual information by lgn bursts. *Journal of Neurophysiology* 81(5):2558–69, 1999.

Reinagel, P. and Reid, R. C. Temporal coding of visual information in the thalamus. *Journal of Neuroscience* 20(14):5392–400, 2000.

Rieke, F., Warland, D., de Ruyter van Steveninck, R., and Bialek, W. B. *Spikes: Exploring the Neural Code.* MIT Press Cambridge, MA, 1997.

Sanderson, K. J. The projection of the visual field to the lateral geniculate and medial interlaminar nuclei in the cat. *J. Comp. Neurol.* 143:101–118, 1971.

Sestokas, A. K. and Lehmkuhle, S. Response variability of x- and y-cells in the dorsal lateral geniculate nucleus of the cat. *Journal of Neurophysiology* 59(2):317–25, 1988.

Shapley, R. M. and Victor, J. D. The effect of contrast on the transfer properties of cat retinal ganglion cells. *J. Physiol. (London)* 285:275–298, 1978.

Strong, S. P., de Ruyter van Steveninck, R. R., Bialek, W., and Koberle, R. On the application of information theory to neural spike trains. In: *Pacific Symposium on Biocomputing '98*, edited by R. B. Altman, A. K. Dunker, L. Hunter, and T. E. Klein Maui, HI, USA. Singapore World Scientific, 1998a pp. 621–32.

Strong, S. P., Koberle, R., De Ruyter Van Steveninck, R. R., and Bialek, W. Entropy and information in neural spike trains. *Physical Review Letters* 80(1):197–200, 1998b.

Sutter, E. E. A deterministic approach to nonlinear systems analysis. In: *Nonlinear Vision*, edited by R. B. Pinter and B. Nabet. CRC Press Boca Raton, 1992, pp. 171–220.

Thomson, D. J. and Chave, A. D. Jacknifed error estimates for spectra, coherences, and transfer functions. In: *Advances in Spectrum Analysis and Array Processing*, edited by S. Haykin volume 1. Prentice Hall, 1991, pp. 58–113.

Tzonev, S., Rebrik, S., and Miller, K. D. Response specificity of lateral geniculate nucleus neurons. *Soc. Neurosci. Abstr.* 23, 1997. Poster available as http://mccoy.ucsf.edu/Papers/Sfn97/SFN97_poster.html.

Victor, J. D. The dynamics of the cat retinal x-cell centre. *J. Physiol.* 386:219–246, 1987.

Victor, J. D. Temporal aspects of neural coding in the retina and lateral geniculate. *Network* 10:R1–R66, 1999.

Figure Legends

Figure 1: The spike-triggered-average stimulus and the temporal kernel for an OFF cell. The vertical axis represents the stimulus luminance on a linear scale, normalized and shifted so that $+1$ represents the bright frame luminance L, -1 represents the dark frame luminance D, and 0 represents the mean luminance $(L+D)/2$. The solid line depicts the spike-triggered average; the dotted line shows the temporal kernel, obtained by normalizing (in the frequency domain) the spike-triggered average by the stimulus spectrum, up to a cutoff of 90 Hz (see Methods). The temporal kernel represents the cell's temporal receptive field: it is the linear filter that, when applied to the stimulus, best predicts the cell's response in the sense of least mean-square error. Both functions show a strong bright-to-dark transition in the stimulus approximately 32 ms before the spike occurred. This was defined as the cell's mean conditional latency.

Figure 2: Algorithm for assigning a spike to an OFF transition. For each spike in the spike train, look back in time by the mean conditional latency, L, and find the closest OFF transition with a ± 1.5 frame window. In this case the indicated spike is assigned to the OFF transition indicated by the arrow. If more than one OFF transition is present within the window, the one closest to L is taken. Note that multiple spikes falling within a $L \pm 1.5$ frame window of a transition can be assigned to the same transition, and thus be part of the same spike event.

Figure 3: Estimate of the lower bound to the average information per spike between the stimulus and words of spikes. A: The entropy of 48 ms words within the spike train, binned in 1 ms bins, scaled with data fraction in a controllable fashion. B: As the inverse length of the spike words decreased, the amount of data (for fixed recording length) decreased, and the single spike train and joint spike train entropy estimations began to fail for words longer than approximately 50 ms. An infinite word length extrapolation for the entropy rates was obtained by fitting to the region where the data was sufficient.

Figure 4 : PSTH matrix of cell 4's spike responses to one-frame sequences (*i.e.*, to bright frames or dark frames). Responses were histogrammed in 1 ms bins relative to time 0, defined as the time of onset of the stimulus frame, f1. The stimulus itself is illustrated to the left of 0 time (gray represents a bright frame, black represents a dark frame). For one-frame sequences, virtually no spikes were observed in response to a bright frame (stimulus 1) at approximately a mean conditional latency (32 ms) from its onset, while a large number of spikes were seen at a similar time after a dark stimulus (stimulus 0). Responses to stimulus 0 around 32 msec largely represent responses to the 1/2 of cases in which the frame preceding it was a bright frame, creating an OFF transition. Similarly, many spikes are seen in response to stimulus 1 about 1 frame later (around 40 ms), representing responses to the 1/2 of cases in which the bright frame was followed by a dark frame.
Conventions for this and future PSTH-matrix figures: Time 0 is defined relative to the sequence as shown at the top of the matrix: *e.g.* here time 0 is the time of onset of the single frame of the stimulus, f1; for multiple-frame stimuli, the last frame in time would be f1, the preceeding frame f2, etc., so that fk would be the k^{th} frame in reverse temporal order. The stimulus frame sequence is shown to the left of time 0 in temporal sequence from left to right (left being earlier in time), with dark representing a dark frame and gray representing a bright frame.

Figure 5: PSTH matrix of cell 4's spike responses to two-frame sequences, histogrammed in 1 ms bins relative to the onset of frame f1. For two-frame sequences, increased spiking was observed around the mean conditional latency from an OFF transition (stimulus sequence 2). Note that averaging the two-frame responses over the left frame (f2) for a given value of the right frame (f1) yields the one-frame responses for f1: *e.g.*, the average of responses to stimuli 1 and 3 gives the response to stimulus 1 in Figure 4.

Figure 6: PSTH matrix of cell 4's spike responses to seven-frame sequences, histogrammed in 1 ms bins relative to the onset of frame f1 (maximum spike rate of 535 Hz). The stimuli were sorted beginning with f2, f1, f3, f4, ...f7, revealing an isolated band of spikes approximately a mean conditional latency (32 ms) from an OFF transition in f2—f1 (second quadrant). Similar bands of spikes were observed at earlier times in response to different stimulus sequences; these "spike echos" were simply responses to earlier OFF transitions.

Figure 7: PSTH matrix of cell 4's spike responses to eight-frame sequences containing an OFF transition between frames f2—f1, histogrammed in 1 ms bins relative to the onset of stimulus frame f1 (maximum spike rate of 688 Hz).

Figure 8: Cell 4's extracted spike event responses to eight-frame sequences containing an OFF transition between frames f2—f1, histogrammed in 1 ms bins relative to the onset of frame f1. The quality of the algorithm for associating spikes with a particular sequence can be judged by comparing this against the PSTH matrix for eight-frame sequences in Figure 7 (or more objectively by determining the percentage of unclassified spikes, see text).

Figure 9: Standard deviation of the distribution of times to the first spike in an event for a given sequence, plotted against the probability that an event was evoked for that sequence. A: When only the two frames of the OFF transition were specified, the single two-frame sequence had a standard deviation of $3.25 \pm .04$ ms, and an event probability of 49%. B: When a total of eight frames were specified, including the OFF transition between f2—f1, all sequences had sub-3 ms standard deviations, with a median standard deviation of 1.28 ± 0.03 ms. The median sequence is shown as a filled circle. C: Mean first-spike times plotted against the event probability, for case B.

Figure 10: Median standard deviation, across all $k-$frame sequences, of the distribution of times to the first spike, plotted against the number of frames in the stimulus window,

k. The three curves correspond to different locations of the OFF transition, *e.g.* f2—f1 indicated the OFF transition was between frames f1 (the latest in time) and f2. The upper and lower interquartile ranges for each point indicate the width of the distribution of standard deviations across sequences. When no frames were specified after the transition (curve f2—f1), the median standard deviation leveled off at about 1.3 ms, once five to six total frames were specified. When one or two frames were specified after the transition, the curves simply shifted to the right, indicating that the three to four frames before the two frames encompassing the transition were primarily responsible for improving the timing precision. Note that the f3—f2 curve is offset to the right by 0.1 frame for clarity.

Figure 11: Group data (across all cells and contrasts) showing the behavior of the spike timing precision for trials having different unclassified spike percentages. A: The median standard deviation of the time to the first spike in an event tended to be lower for trials where fewer spikes were unclassified. Note however that in two trials, the standard deviation was still around 2 ms despite a high unclassified percentage. B: There was no strong dependence of the firing rate on the unclassified percentage. All analyses were for eight-frame sequences with ON or OFF transition f4—f3.
Conventions for this and future group data figures: black points indicate cells having unclassified percentages less than 5%; grey points correspond to cells having greater than 5% of spikes unclassified. Lines connect points corresponding to the same cell. Different contrast levels are distinguished by differently shaped points, as assigned in the key.

Figure 12: Variance of the distribution of the number of spikes in each event in response to a stimulus sequence with an OFF transition, plotted against the mean number of spikes in an event for that sequence. A: When only the two frames of the OFF transition were specified, the single two frame sequence had a variance nearly equal to its mean, consistent with a Poisson process for spiking (straight line). B: When a total of eight frames were specified, including the OFF transition between f2—f1, most of the sequences still exhibited

variances close to their means, although a few sequences that evoked larger numbers of spikes on average showed significantly sub-Poissonian variances. C: When two frames were specified after the OFF transition (eight-frame sequences with OFF transition between f4—f3), however, most sequences displayed sub-Poisson variances, with several sequences approaching the minimum-variance limit imposed by the discreteness of the spikes (scalloped line).

Figure 13: Median Fano factor (variance over mean) of the spike count, taken across all $k-$frame sequences, plotted against the number of frames in the stimulus window, k. The three curves correspond to different specifications of the frames surrounding the OFF transition; labelling is the same as in Figure 10. The upper and lower interquartile ranges for each point indicate the width of the distribution of Fano factors across sequences. When no frames were specified after the transition (f2—f1 curve), the median Fano factor stayed close to unity, the value expected for a Poisson process. By specifying one or two frames after the transition, the Fano factor was dramatically reduced with increasing stimulus window size, indicating that the frames after the two frames encompassing the transition were critical for improving the spike count precision. This was consistent with a spike vetoing effect by the frames occuring after the transition. Note that the f3—f2 curve is offset to the right by 0.1 frame for clarity.

Figure 14: Group data showing the behavior of the spike count precision for trials having different unclassified spike percentages. The median Fano factor could be less than one regardless of the percentage of unclassified spikes, but the lowest factors (< 0.5) were seen only for small unclassified percentages.

Figure 15: Average per spike information between k-frame sequences and a single 1 ms bin at a time relative to the onset of frame f1. Different curves correspond to different numbers of frames in the stimulus. The shape of the one-frame case can be compared to the PSTH matrix in Figure 4, although note the difference in notation and placement of the frames. Approximately 20 ms after frame f1 was shown, the information in the one-frame case in-

creased, reaching a peak of one bit per spike at about 30 ms. This corresponded to the time at which a spike only occurred in response to a dark frame, and virtually never to a bright frame. In contrast, around 40 ms after the onset of f1, a spike was much more likely to be associated with a bright frame rather than a dark frame, yielding a second peak in the information. For larger k, the (smoothed) maximum stimulus information (see Methods) began to level off near 3.5-3.6 bits/spike (for nine- and ten-frame sequences). At early times, ten-frame sequences showed spurious information due to data insufficiency; this did not affect the peak informations for ten-frame sequences because the probability distribution for words at those times was completely different from that near the peak.

Figure 16: Information transmission at different time resolutions. A: The information calculated by the direct entropy method is shown for two cases: solid circles were computed from the multiple repeats of eight-frame stimuli, focusing on 8 ms spike words at 1, 2, 4, and 8 ms bin sizes; open circles were computed from the two full repeats of the M-sequence, and represent only a lower bound on the information from infinitely long words of spikes. Both measures showed an increase in information at finer time resolutions, and the lower bound was close to the information observed with 8 ms words. B: The corresponding coding efficiencies at the different time resolutions showed only a little change with bin size, indicating that the increased capacity gained at higher temporal resolution was being used by the neuron to code information about the stimulus.

Figure 17: Group data on the information in 1 ms bins plotted against the lower bound for the information in infinitely long words of 1 ms bins. The former was computed from the multiple repeats of nine-frame sequences, while the latter was estimated from the two repeats of the full M-sequence. A: Data plotted as information per spike. B: Data plotted as information per second. In both representations, the data fell close to the diagonal line where the 1 ms bin information equals the infinitely long word information. This suggests that there was at most a minimal amount of redundancy, and possibly synergy, in the coding

by successive spikes.

Figure 18: Comparison of the data to three quasi-linear models: spike event PSTH matrices. Parameters used for each model gave the best match between the data and models' *full* PSTH. A: Data. B: Linear gain, Poisson spikes. C: Linear gain, Poisson with 3 ms absolute refractory period. The stimulus was convolved with the temporal kernel of the real cell (Figure 1), thresholded, and scaled to approximately match the mean (and peak, for refractory model) firing rates of the real cell. Threshold and, in c and d, refractory period were selected to minimize the mean square error between the model PSTH and the data PSTH. The general location of responses was roughly accurate, but the widths of the responses and the asymmetry in the peaks of the data were not fully reproduced by the models. For both linear Poisson and linear refractory models, the optimal threshold was 3.4 in units of output of the convolution of the temporal kernel with the stimulus, where the kernel was normalized as in Figure 1 and bright and dark stimulus frames were represented as ± 1. For comparison, the output of the convolution had a peak value of 10.9 and an rms of 4.3 (and mean 0).

Figure 19: Comparison between the mean first-spike times and spike counts for the data and the models, plotted for all eight-frame sequences with an OFF transition between f4 and f3. A: The first-spike times in the data were well approximated by all of the models, except that larger times were underestimated. B: The spike counts were generally well predicted by the models, although there was a small tendency to underestimate small counts and overestimate high counts.

Figure 20: Comparison between (A) the spike timing standard deviations and (B) spike count Fano factors for the data and the models, plotted for all eight-frame sequences with an OFF transition between f4 and f3. The Poisson model poorly matches the precision of the data. The model with refractoriness does considerably better, but shows some tendency to overestimate smaller first-spike times/Fano factors and underestimate larger ones. The

points with zero standard deviation and Fano factor showed only one nonzero spike response across the 128 trials.

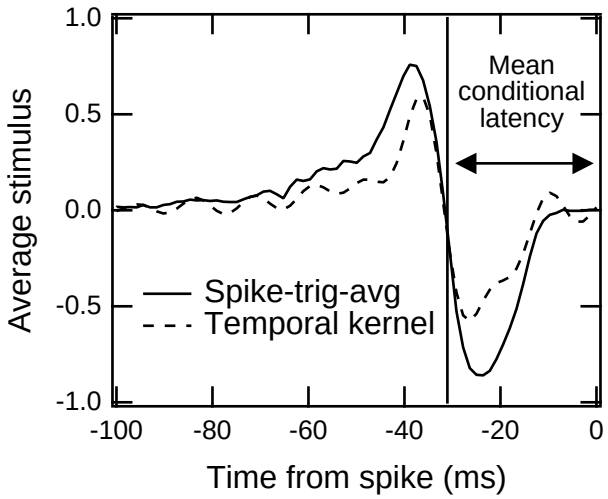
1.0
0.5
0.0
-0.5
-1.0
Average stimulus
Mean
conditional
latency
Spike-trig-avg
Temporal kernel
-100
-80
-60
-40
-20
0
Time from spike (ms)

Figure 1:

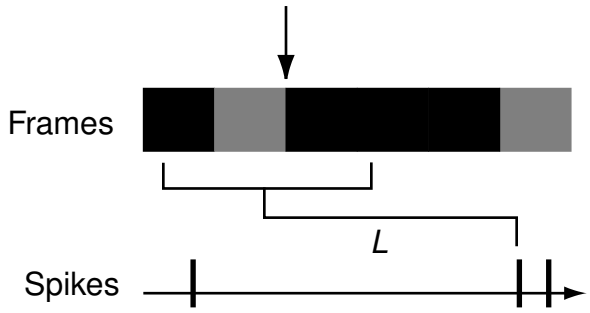

Figure 2:

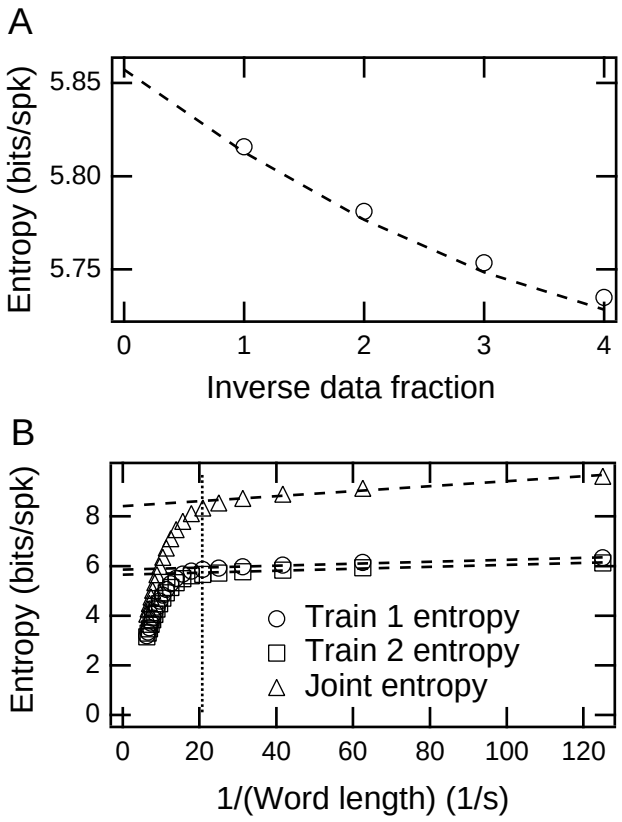
A
Entropy (bits/spk)
5.85
5.80
5.75
0
1
2
3
4
Inverse data fraction
B
Entropy (bits/spk)
8
6
4
2
0
0
20
40
60
80
100
120
1/(Word length) (1/s)
Train 1 entropy
Train 2 entropy
Joint entropy

Figure 3:

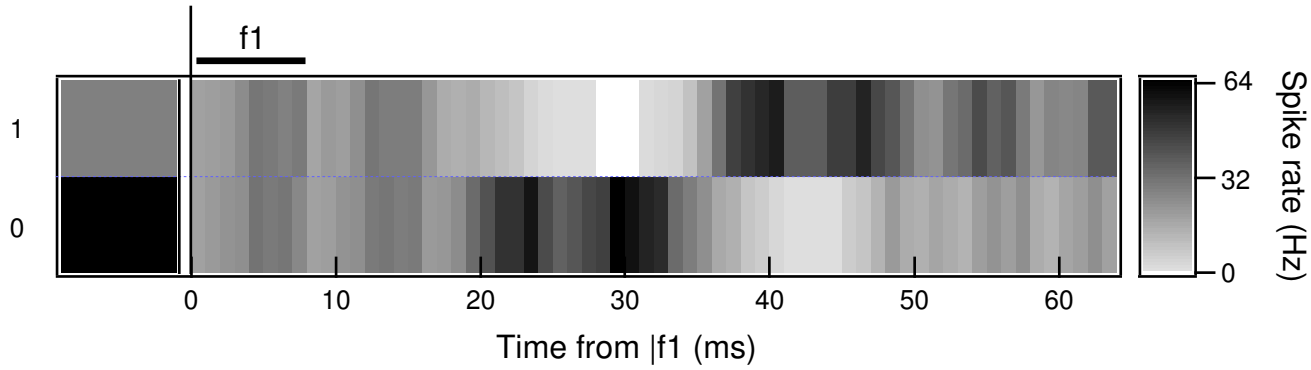

Figure 4:

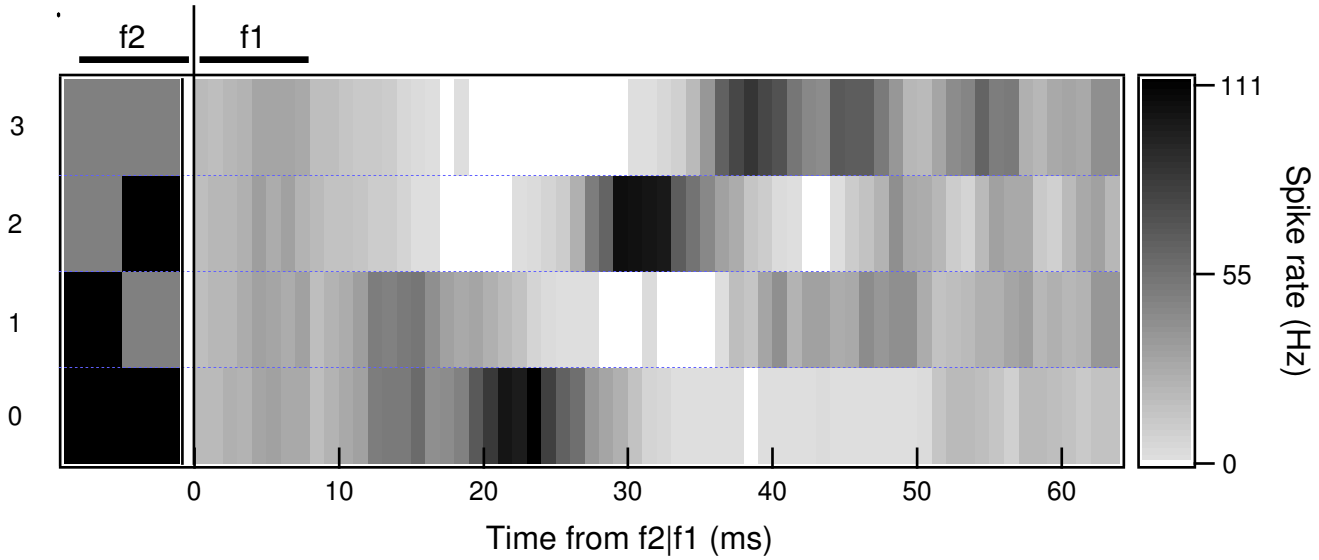
f2
f1
3
2
1
0
0
10
20
30
40
50
60
Time from f2|f1 (ms)
111
55
0
Spike rate (Hz)

Figure 5:

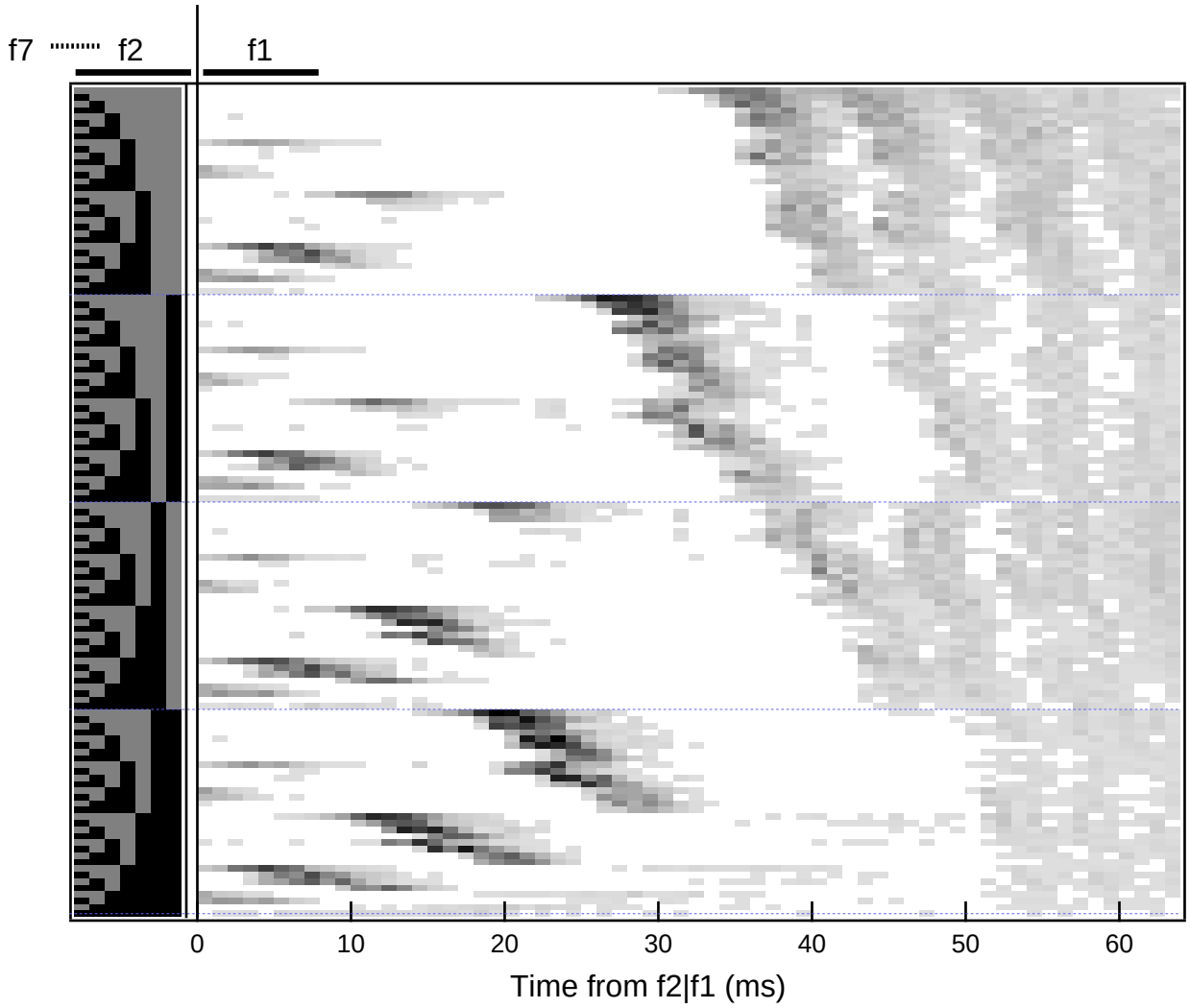

Figure 6:

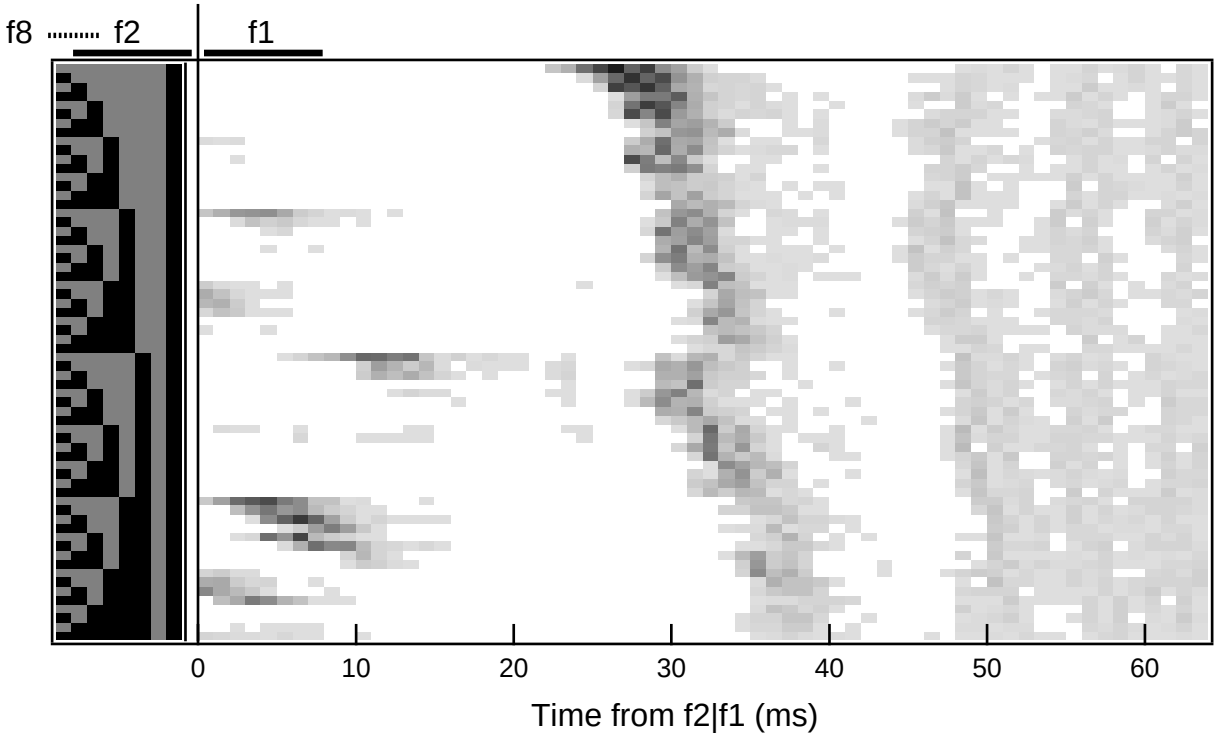

Figure 7:

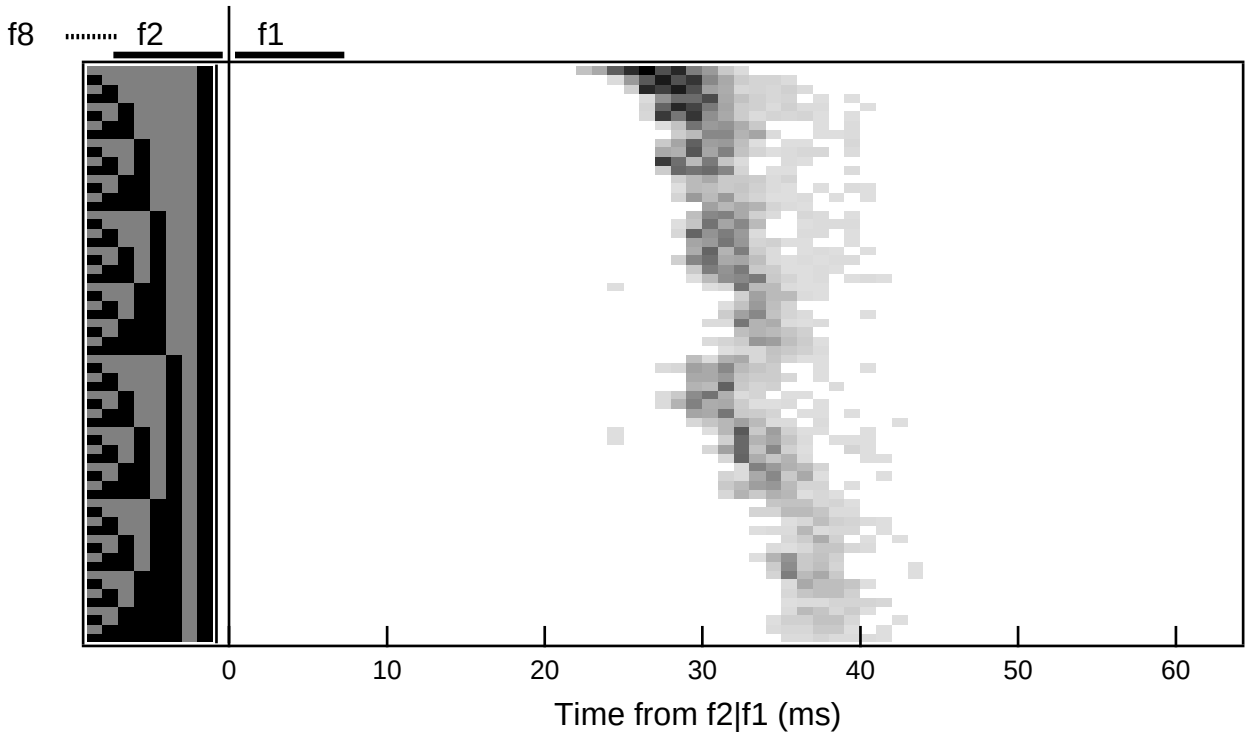

Figure 8:

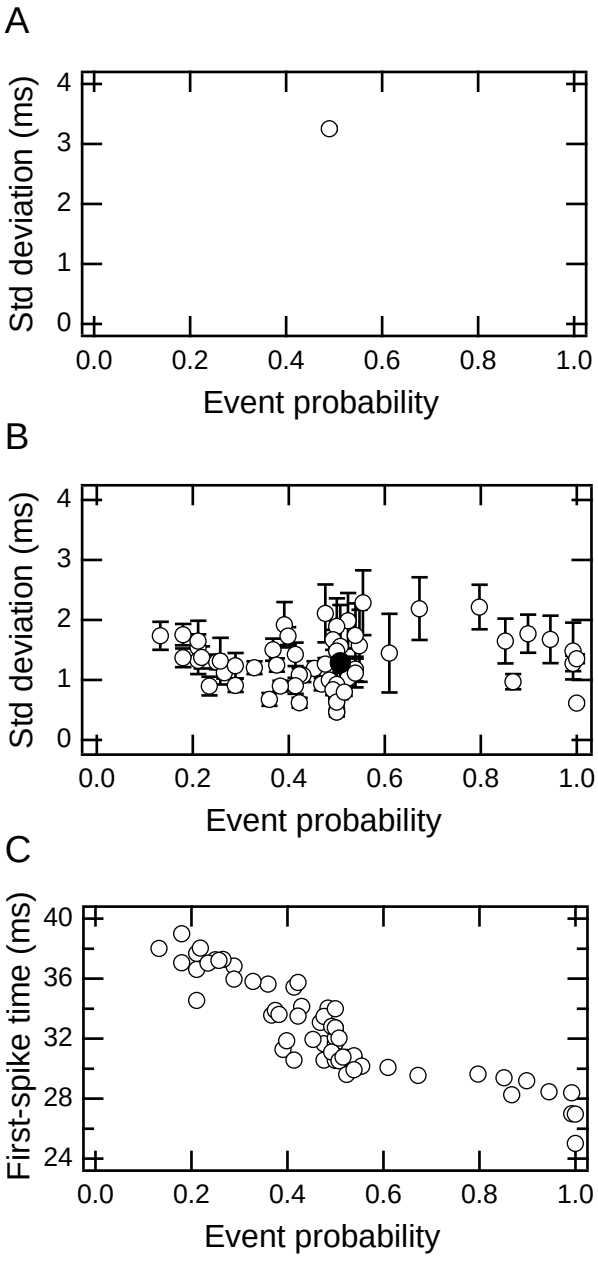
A
Std deviation (ms)
0
1
2
3
4
0.0
0.2
0.4
0.6
0.8
1.0
Event probability
B
Std deviation (ms)
0
1
2
3
4
0.0
0.2
0.4
0.6
0.8
1.0
Event probability
C
First-spike time (ms)
24
28
32
36
40
0.0
0.2
0.4
0.6
0.8
1.0
Event probability

Figure 9:

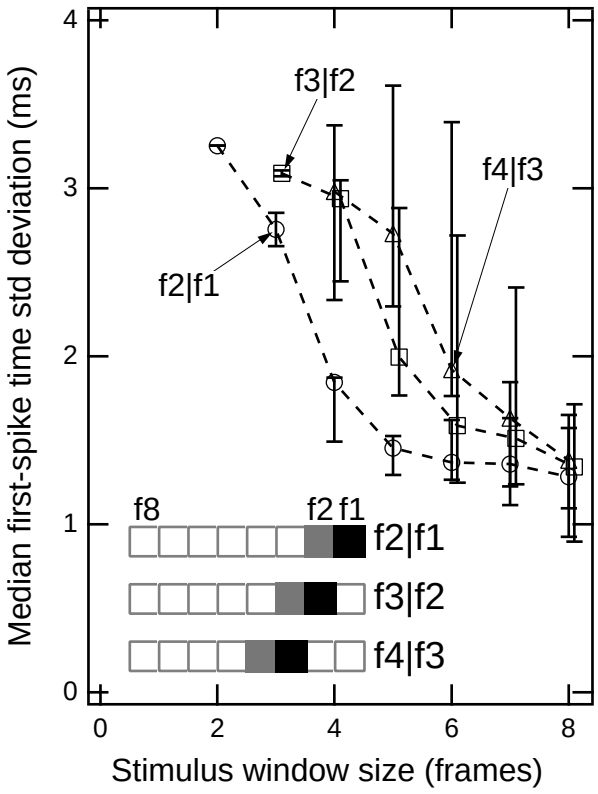
Median first-spike time std deviation (ms)
Stimulus window size (frames)
f3|f2
f4|f3
f2|f1
f8
f2 f1
f2|f1
f3|f2
f4|f3
0
1
2
3
4
0
2
4
6
8

Figure 10:

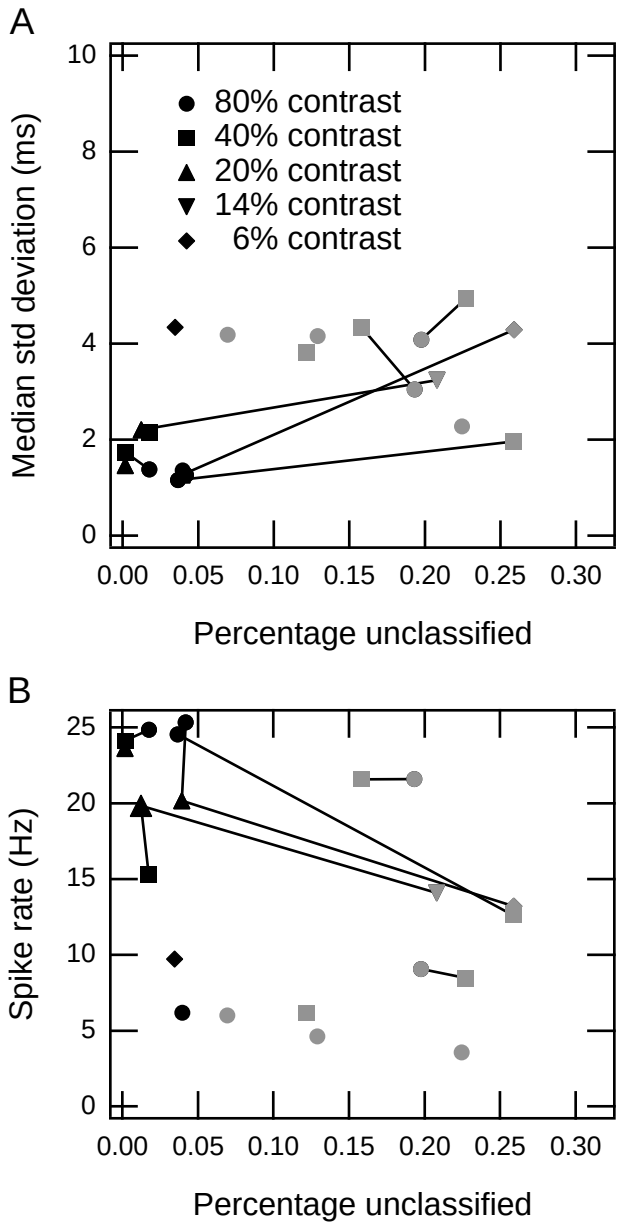
A
Median std deviation (ms)
80% contrast
40% contrast
20% contrast
14% contrast
6% contrast
Percentage unclassified
B
Spike rate (Hz)
Percentage unclassified

Figure 11:

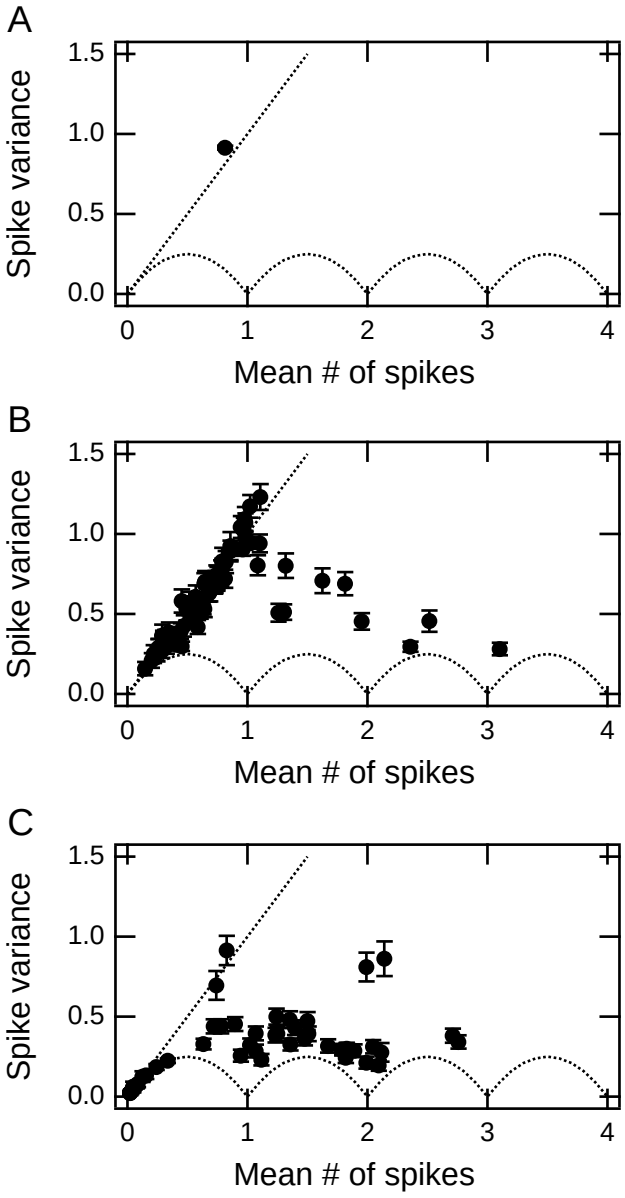

Figure 12:

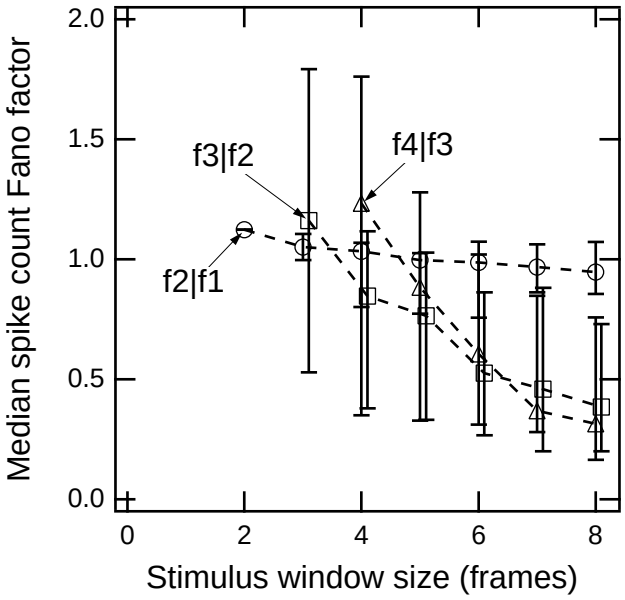
Median spike count Fano factor
2.0
1.5
1.0
0.5
0.0
0
2
4
6
8
Stimulus window size (frames)
f3|f2
f4|f3
f2|f1

Figure 13:

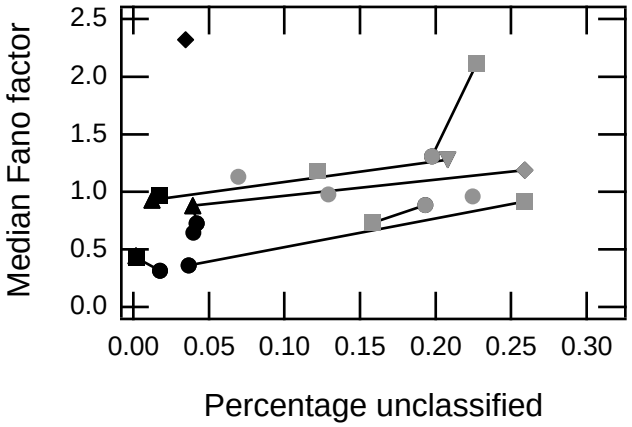
Median Fano factor
2.5
2.0
1.5
1.0
0.5
0.0
0.00
0.05
0.10
0.15
0.20
0.25
0.30
Percentage unclassified

Figure 14:

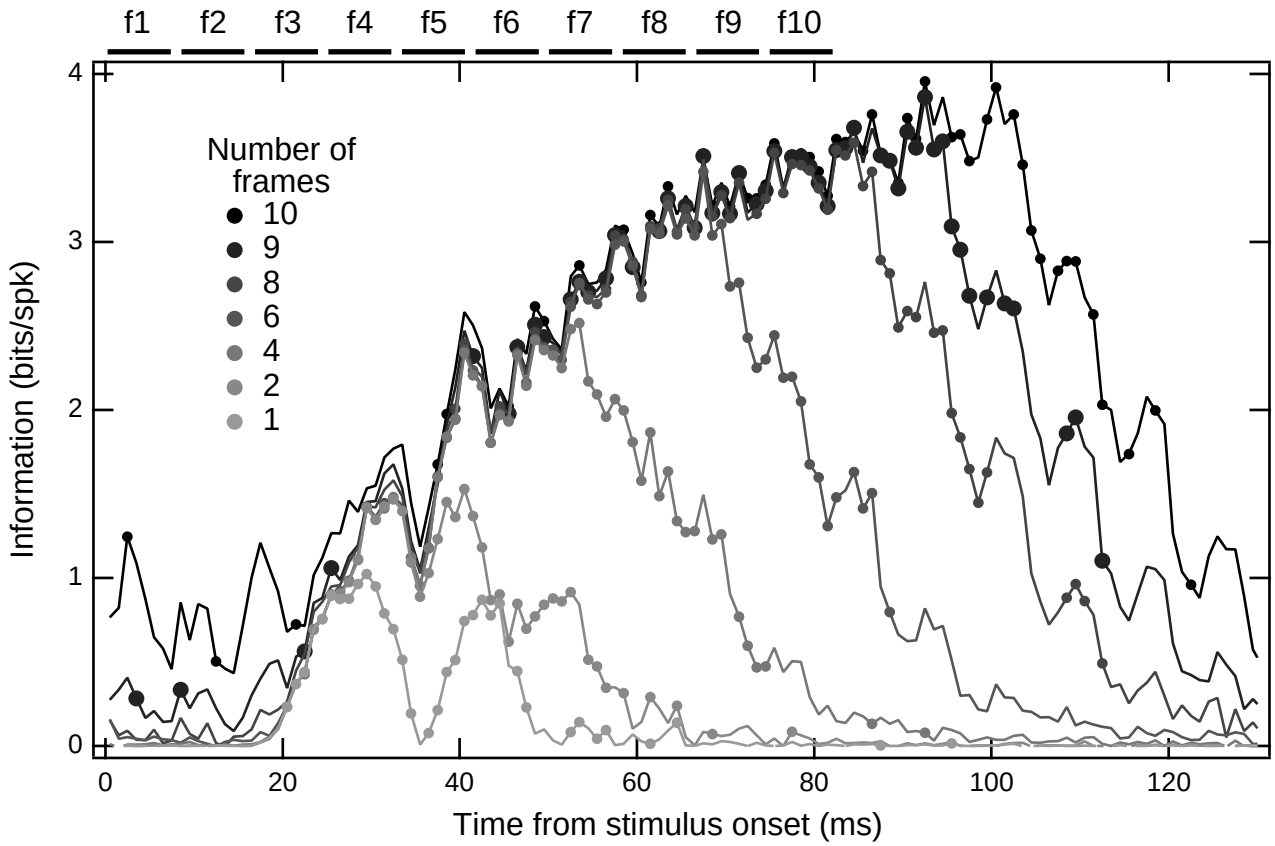
f1
f2
f3
f4
f5
f6
f7
f8
f9
f10
Number of frames
10
9
8
6
4
2
1
Information (bits/spk)
Time from stimulus onset (ms)
0
1
2
3
4
0
20
40
60
80
100
120

Figure 15:

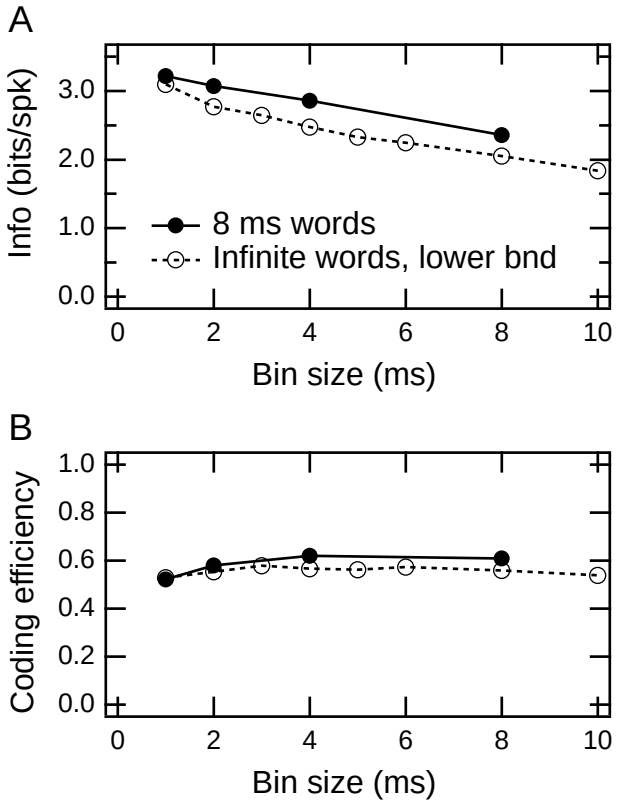
A
Info (bits/spk)
3.0
2.0
1.0
0.0
8 ms words
Infinite words, lower bnd
0
2
4
6
8
10
Bin size (ms)
B
Coding efficiency
1.0
0.8
0.6
0.4
0.2
0.0
0
2
4
6
8
10
Bin size (ms)

Figure 16:

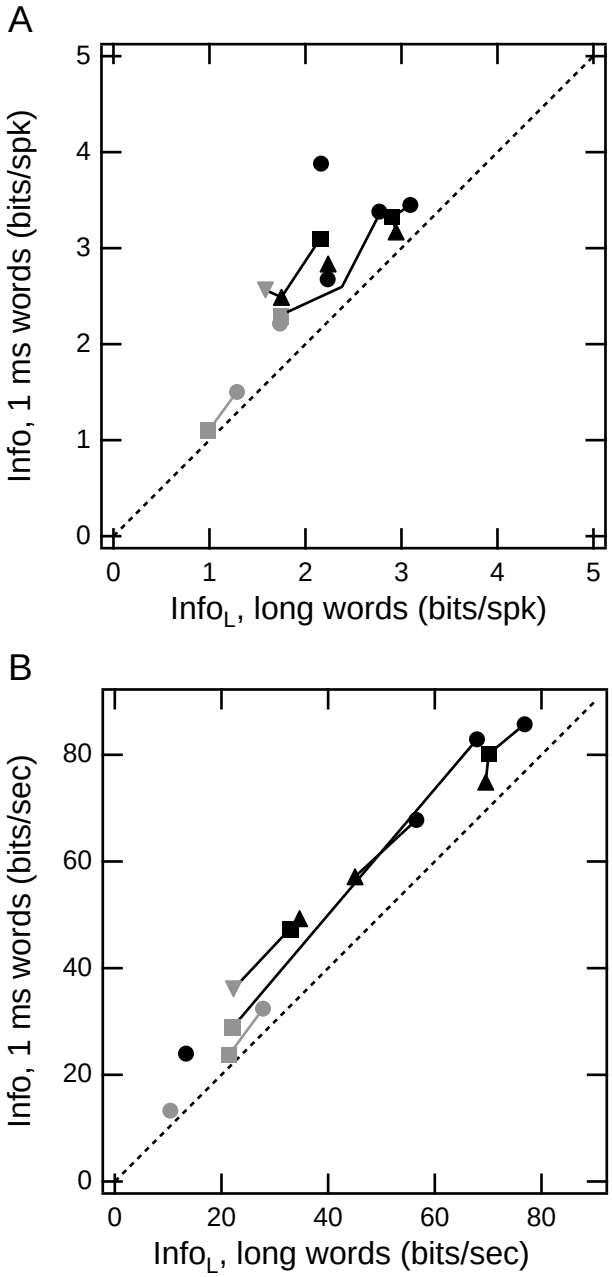
A
Info, 1 ms words (bits/spk)
Info$_L$, long words (bits/spk)
0
1
2
3
4
5
B
Info, 1 ms words (bits/sec)
Info$_L$, long words (bits/sec)
0
20
40
60
80

Figure 17:

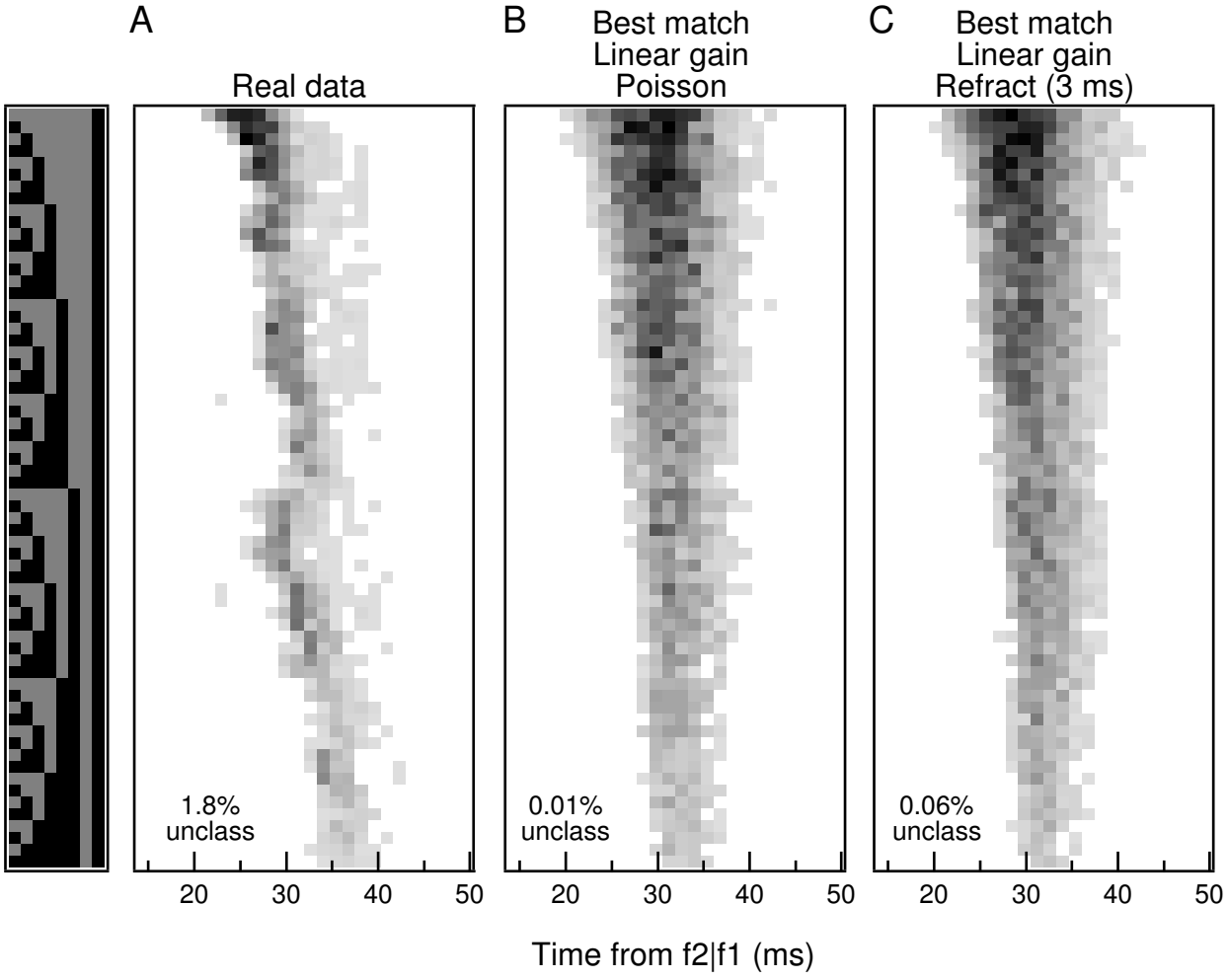
A
Real data
B
Best match
Linear gain
Poisson
C
Best match
Linear gain
Refract (3 ms)
1.8%
unclass
0.01%
unclass
0.06%
unclass
20
30
40
50
Time from f2|f1 (ms)

Figure 18:

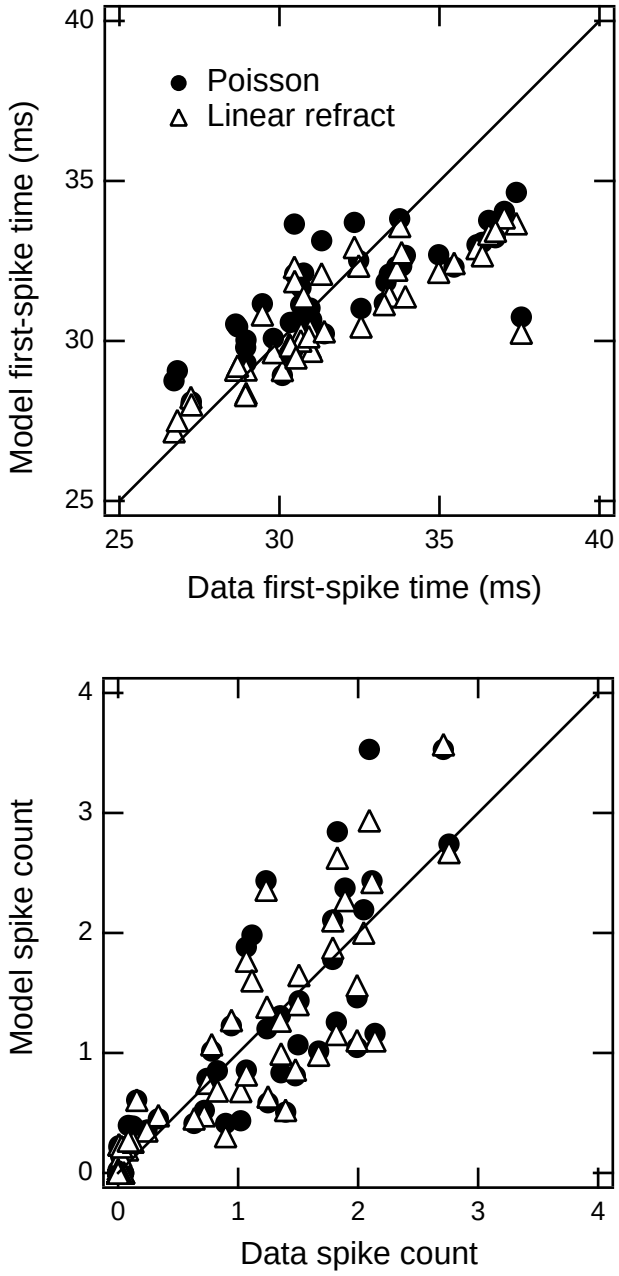
Poisson
Linear refract
Model first-spike time (ms)
Data first-spike time (ms)
40
35
30
25
25
30
35
40
Model spike count
Data spike count
4
3
2
1
0
0
1
2
3
4

Figure 19:

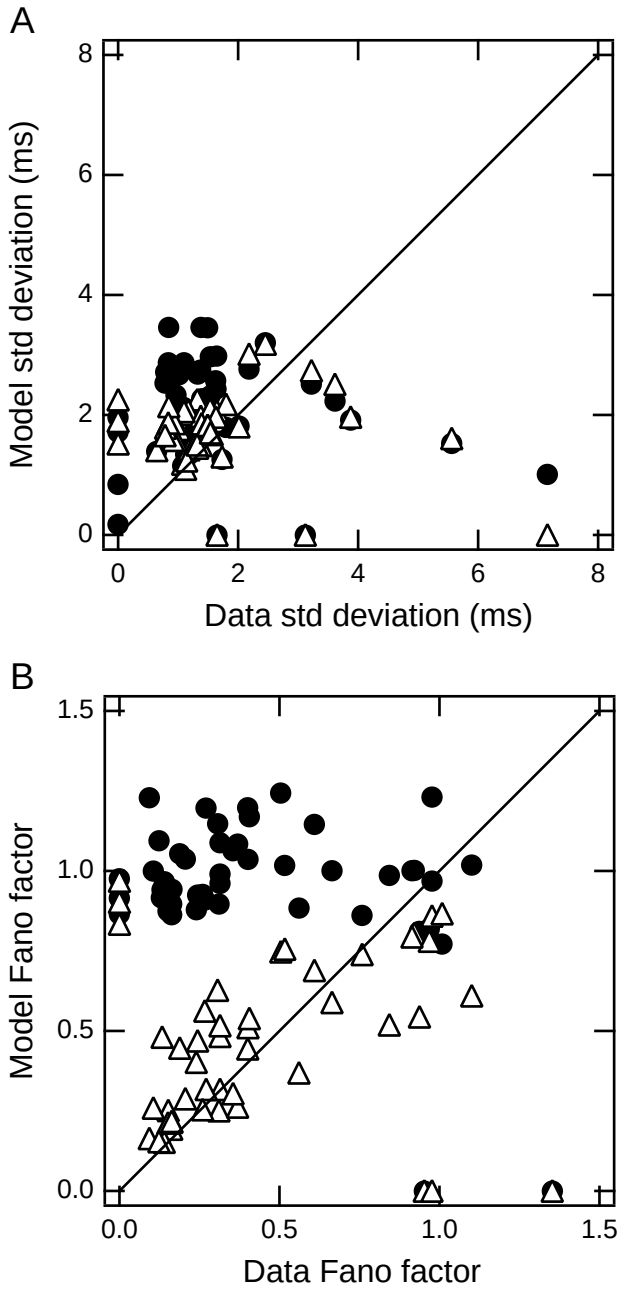
A
Model std deviation (ms)
Data std deviation (ms)
0
2
4
6
8
B
Model Fano factor
Data Fano factor
0.0
0.5
1.0
1.5

Figure 20: