

**Supplementary Information for the manuscript**

***“Synchrony makes neurons fire in sequence—and  
stimulus properties determine who is ahead”***

**Martha N. Havenith, Shan Yu, Julia Biederlack, Nan-Hui Chen, Wolf Singer, and  
Danko Nikolić**

## CONTENTS

### **Supplemental Methods**

#### **Supplemental Figure S1**

Spike waveforms and independence between CCH delays and activation delays

#### **Supplemental Figure S2**

Consistency of firing delays across individual coincident events

#### **Supplemental Figure S3**

Average peak widths of CCHs

#### **Supplemental Figure S4**

Event-to-event variability of firing times and firing rates

#### **Supplemental Figure S5**

Robustness of firing sequences to assembly size

#### **Supplemental Figure S6**

Firing sequences in multi-unit activity

#### **Supplemental Figure S7**

Stimulus-dependent changes in firing times for neurons from different cortical depths

#### **Supplemental Figure S8**

Firing sequences for neurons classified as fast-spiking and regular-spiking

#### **Supplemental Figure S9**

Trial-to-trial variability of firing times and firing rates

#### **Supplemental Figure S10**

Oscillation strength and frequency for all data sets

**Supplemental Figure S11**

Theoretical proposal for the generation and readout of firing sequences

**Supplemental Figure S12**

Example validation of the model shown in Supplementary Figure 9

**Supplemental Table S1**

Summary of statistical results testing stimulus-dependent changes of firing sequences

**Supplemental Discussion 1**

The origin of firing sequences - Sequential firing with parallel wiring

**Supplemental Discussion 2**

The readout of firing sequences - Sensing input timing

**Supplemental References**

## **Supplemental Methods**

### **1. Surgical procedures**

In five adult cats, anaesthesia was induced with a mixture of Ketamine (Ketanest, Parke-Davis, 10 mg/kg, intramuscular) and Xylazine (Rompun, Bayer, 2 mg/kg, intramuscular). Following tracheotomy, anaesthesia was maintained with 70% N<sub>2</sub>O and 30% O<sub>2</sub>, supplemented with ~1.0% Halothane. After craniotomy, the level of Halothane was reduced to ~0.5%. When anaesthesia was stable and sufficiently strong to prevent any vegetative reactions, the animal was paralyzed with Pancuronium bromide applied intravenously (Pancuronium, Organon, 0.15 mg/(kg,h)). The respiration rate was adjusted manually to keep end-tidal CO<sub>2</sub> at 3-4%. Rectal temperature was kept at 37-38 °C. Glucose and electrolytes were supplemented intravenously. All procedures complied with the German law for the protection of animals and were overseen by a veterinarian.

### **2. Data acquisition**

Multi-unit (MU) activity was recorded using silicon-based 16-channel probes supplied by the Center for Neural Communication Technology at the University of Michigan (Michigan probes). The electrode contacts (recording sites) had an impedance of 0.3–0.5 MΩ at 1000 Hz and were organized in a 4×4 matrix on four shanks, with a distance of 0.2 mm between the neighboring contacts. Thus, the recording area of one probe spanned ~0.6×0.6 mm. The distance between neighboring electrode contacts was sufficient to prevent them from recording activity of the same neuron—in no case did cross-correlation

histograms (CCHs) show the high, narrow peaks in the very centre of the CCH characteristic of auto-correlations.

We always inserted two Michigan probes simultaneously into area 17 of the same hemisphere (illustrated in supplemental Fig. S7), thus recording from 32 electrode contacts in parallel. In one probe, good signals were available only on nine recording channels. Thus, in total, we recorded from 153 sites.

Signals were amplified 1000 $\times$ , filtered at 500 Hz to 3.5 kHz, and sampled with a frequency of 32 kHz. Action potentials (spikes) were detected with a two-sided threshold discriminator adjusted manually to yield a signal-to-noise ratio of approximately 2:1. For each detected action potential, the time of the event was recorded together with the spike waveform over a duration of 1.2 ms.

To extract single unit (SU) activity, spike waveforms were sorted offline using principal component analysis (PCA), according to the following two criteria: 1) Waveform separation: The shape of the spike waveforms should be clearly defined and different from all other waveforms in the MU, as indicated by the first three PCA components. 2) Refractory Period: Not more than 0.5% of spikes should occur with an inter-spike-interval of  $< 1$  ms. The 126 isolated SUs that satisfied these criteria (7-20 neurons per recording probe) had an average firing rate of  $18 \pm 15$  spikes/second (mean  $\pm 2$  SD) in the most optimal stimulation condition.

### **3. Visual stimulation**

Visual stimuli were presented on a 21-inch Hitachi CM813ETPlus monitor with a refresh rate of 100 Hz, positioned at a distance of 57 cm from the eyes. The pupils were dilated with atropine, the nictitating membrane retracted with neosynephrine and the cornea protected with contact lenses containing artificial pupils of 2 mm diameter. After refraction, the eyes were focused on the monitor with correcting lenses. The optical axes of the two eyes were aligned by mapping the receptive fields (RF) for each eye separately and placing a prism in front of one eye to achieve binocular fusion of the RFs. All RFs were determined by manual mapping with high-contrast white bars on a black background. Due to the spatial proximity between electrode contacts, the RFs recorded from one probe always overlapped, producing clusters that spanned up to  $\sim 10^\circ$  of visual angle. The stimuli were always positioned such that their centers matched the centre of the RF cluster.

Stimuli were always presented binocularly, using the presentation software ActiveSTIM ([www.ActiveSTIM.com](http://www.ActiveSTIM.com)). In the main set of recordings, for which the largest part of the reported analyses was carried out, we presented high-contrast drifting sinusoidal gratings of circular shape, and varied the orientation and drifting direction in twelve  $30^\circ$ -steps, thereby covering the entire circle of  $360^\circ$ . In three cats, the gratings had a spatial frequency of  $2.4^\circ/\text{cycle}$  and covered  $12^\circ$  of visual angle, in two cases moving with a speed of  $2.2^\circ/\text{second}$ , and in one with  $6.6^\circ/\text{second}$ . In the remaining two cats, gratings had a spatial frequency of 1 and  $2.5^\circ/\text{cycle}$ , respectively, covering 7 and  $18^\circ$  of visual angle and moving with a speed of 1.5 and  $2^\circ/\text{second}$ . To establish the internal origin of the time delays in CCHs (Fig. 1), in three animals we made an additional recording in which high-

contrast moving bars were presented. The bars spanned  $0.5^\circ \times 6^\circ$  of visual angle, appeared on the screen at an eccentricity of  $2^\circ$  from the center of the RF cluster and moved across the centre of the cluster with a speed of  $1^\circ/\text{second}$ , disappearing after four seconds (see methods section for the choices of bar orientations and movement directions). Stimuli were always presented in a randomized order. For each stimulus condition, we recorded responses to 40-120 stimulus repetitions (trials). This number of trials was needed to determine the time delays from CCHs accurately. The recordings were made in 2–6 blocks, each block consisting of 20–60 trials per stimulus condition. Each block lasted 20 – 100 minutes, and often several hours long breaks were made between different blocks. This enabled us to investigate the stability of the relative firing times over longer periods of time. The responses obtained from the two simultaneously inserted Michigan probes were analyzed separately. Thus, no timing relations between units recorded from different probes were investigated in the present study.

#### **4. Pair-wise extraction of preferred delays**

In the past, time delays have been measured from CCHs by fitting either a damped Gaussian (i.e. Gabor) function (Konig, 1994; Konig et al., 1995b; Roelfsema et al., 1997; Nikolić, 2007) or a cosine function (Schneider and Nikolić, 2006; Schneider et al., 2006). These two methods yield almost identical results ( $r = 0.98$ ; see Schneider and Nikolić, 2006) and the average measurement error was  $<1$  ms. Here, we fitted a Gaussian function to the central part of the CCH ( $\pm 15$  ms, see Havenith et al., 2009). Our tests indicated that this function produced the same results as the other two ( $r > 0.95$ ; 91 CCHs tested), but was least sensitive to the initial conditions and least likely to enter local minima during the fitting process.

For this procedure, 1 ms binning is optimal for estimating the time delays with maximal accuracy—estimates reaching even sub-millisecond precision (Havenith et al., 2009). The sub-millisecond precision of the fitting procedure could be achieved only if the CCH peak consisted of at least three consecutive bins (1 ms width each) with at least  $N = \sim 8$  entries (coincident events). With fewer entries, precision decreased rapidly due to fitting errors (see Havenith et al., 2009, for more details). This requirement had several implications for the selection of the data included in the analyses:

i) Not all neurons in all stimulation conditions generated enough spikes to produce a sufficient number of entries for the CCHs in order to enable a reliable assessment of the preferred time delays. This was usually a problem with stimuli of non-optimal orientation, which did not drive the neurons well. If a neuron had overall low firing rates and participated in CCHs with insufficient entries, it was completely excluded from the analysis. However, we did retain neurons that had ‘poor’ CCHs only in some stimulation conditions. Thus, for the sake of completeness of the pair-wise estimates, we accepted enlarged measurement errors in these conditions. The reason is that our analysis required that the networks from which we reconstructed the firing sequences did not contain missing values (i.e., the networks needed to be complete). The criterion for the inclusion of a unit was that the Gaussian functions fitted the CCHs with  $r^2 \geq 0.5$  in more than 50% of the cases. This criterion provided sufficiently accurate estimates of the relative firing times because each value was computed on the basis of multiple CCHs. This procedure eliminated 48 of the 126 isolated SUs (38%), leaving 78 neurons for further analysis (5-12 per recording probe; median: 7 neurons). This criterion could not be applied for the computations shown in Figure 5B, and for the analysis of time-resolved changes in firing



times shown in Figure 5E-F. These analyses needed to include also responses with very low spike counts (e.g. for non-preferred grating orientations). Thus, all neurons with non-empty CCHs were included in these analyses, irrespective of the goodness of fit of Gaussian functions.

ii) As mentioned, some grating orientations evoked particularly low firing rates in the recorded neurons, producing very sparse, or sometimes even completely empty, CCHs. Therefore, in some analyses we had to exclude the responses to certain grating orientations. A stimulation condition was removed from the analysis if it was not possible to compute a firing sequence for five or more neurons after applying the criterion explained in i). In one data set, neurons responded vigorously enough to compute firing sequences for all twelve stimulation conditions. The smallest number of stimulation conditions that entered the analysis was two, and the median was 3.5. As mentioned above, we did not use these criteria for the computation of the orientation tuning of the relative firing times.

iii) Finally, it was not feasible to estimate the inter-trial variability of firing sequences directly by determining time delays for single trials and to make at the same time a reasonable comparison to firing rates. This is because, due to low firing rates, insufficient entries were available to estimate CCHs from single trials, which lead to a disproportionate increase in the error of time delay estimates compared to estimates of firing rates (for an illustration of the trial-to-trial variability in estimated time delays, see Nikolić, 2007, and supplemental Fig. S9). This lack of distributivity of errors in the estimates of time delays is not an indication of inherent imprecision of time delays but

occurs due to the fact that the measure of delay is extracted indirectly, by fitting a Gaussian function. The fitting errors of each parameter do not reduce with equal speed. For example, the baseline of the Gaussian can be estimated with accuracy that is proportional to  $1/\sqrt{n}$ , where  $n$  would denote the amount of available data, e.g. the number of trials. However, other parameters (e.g. the mean and the width) can be estimated with a similar rate only when the baseline is estimated accurately, and the estimate of the Gaussian's peak delay is in turn reliant on all three of these other parameters. Consequently, the error with which the mean delay of the Gaussian is measured, is for small numbers of data points, much larger than would be expected from the widths of the centre peaks. In other words, with a small number of trials (that is, CCHs containing only a small number of coincident events), the best fit is likely to position the Gaussian at values far outside the region within which the centre peaks of CCHs vary (e.g. an estimated delay of 30 ms can be easily encountered). Consequently, an average of  $k$  Gaussian fits, each of which is applied to  $1/k$  of the total available amount of data, will be much more inaccurate than a single Gaussian fit calculated for a CCH including all data. This is referred to as the lacking distributivity of errors. Thus, in order to obtain accurate estimates in single trials, each trial would be required to contain an unrealistically high number of entries (e.g. rich MU instead of SU). Therefore, time delays had to be estimated always from CCHs calculated across multiple trials. Due to these limitations, to estimate the variability of firing times across repeated measurements we randomly split the data sets into two halves 100 times and calculated CCHs from each half of the data. This restriction does not apply to rate responses, as they can be calculated for  $n$  individual trials and subsequently averaged to obtain an estimate with the error decreasing with the usual factor of  $\sim 1/\sqrt{n}$ .

To ensure that CCHs did not include rate co-variation (Brody, 1999), the responses showing the strongest rate variation, i.e. those immediately following the onset and offset of the stimuli, were not included in the CCH analysis. Hence, CCHs were computed only for the sustained responses over a duration of 3.2 or 1.8 seconds for grating and bar stimuli, respectively. Stimulus-locked rate co-variation can be detected by computing CCHs between responses with shuffled trial orders, i.e., shift predictors (Perkel et al., 1967). We computed shift predictors for a random sample of four datasets and most of these 423 shifted CCHs were flat, indicating that rate co-variation had not contributed to the peaks in the CCHs. More importantly, irrespective of the shapes of the shift predictors, the estimates of phase shifts were identical with and without subtraction of shift predictors ( $r = 0.99$ ;  $n = 423$ ).

## 5. Hypothesis testing

### 5.1 Positioning errors of units on the time axis

The extraction of firing sequences is described in the Methods section. Pair-wise time delays are not guaranteed to be additive. In principle, spike trains can also produce non-additive time delays (Schneider et al., 2008). Therefore, for each data set, we needed to ensure that the time delays were sufficiently additive to warrant the application of our analyses. The degree to which the preferred firing sequence correctly represents the measured pair-wise time delays is expressed in the average positioning error per unit,  $\sigma_{Add}$ , defined elsewhere (Schneider et al., 2006) as  $\sigma_{Add} \equiv \hat{\sigma}_x$ . To obtain  $\sigma_{Add}$ , first the differences between the distance  $\delta_{ij}$  separating two units  $i$  and  $j$  on the time axis and their measured time delay  $\varphi_{ij}$  are summed up as follows:

$$Q_{Add} = \sum_{i < j} (\varphi_{ij} - \delta_{ij})^2. \quad (\text{Equation S1})$$

Next,  $Q_{Add}$  is normalized by the number of time delays that enter its computation (adapted from Schneider et al., 2006; Equation 4):

$$\sigma_{Add} = \sqrt{Q_{Add} \frac{2}{(n-2)n}}. \quad (\text{Equation S2})$$

$\sigma_{Add}$  is also referred to as the additivity error because the measured time delay between two units and their distance on the time axis correspond only if a data set is additive (see Figs. 2C-E).

In addition to this average error across all units, we computed an individual positioning error  $\sigma_{Add(k)}$  for each unit  $k$ . Similar to the average error for all units,  $\sigma_{Add(k)}$  was also computed as a normalized sum of errors, but only for the time delays  $\varphi_{jk}$  of the unit  $k$  with all other units  $j$  ( $j \neq k$ ).

$$\sigma_{Add(k)} = \sqrt{\sum_{j \neq k} (\varphi_{jk} - \delta_{jk})^2 \frac{1}{(n-2)n}}. \quad (\text{Equation S3})$$

It can be shown that the mean of  $\sigma_{Add(k)}^2$  across all units  $k$  equals the overall additivity error

$$\sigma_{Add}^2.$$

## 5.2 Multiple comparisons of firing sequences

The ANOVA-*Add* allows only for comparisons between two firing sequences. Thus, when comparing changes in firing sequences for more than two stimulus conditions, we conducted multiple pair-wise ANOVA-*Adds* and controlled for the type I error by correcting  $p$ -values using the Dunn-Sidak correction for multiple comparisons (Dunn, 1961). A change in a firing sequence was considered significant if the smallest corrected  $p$ -value did not exceed  $p = 0.05$ .

To correct for multiple comparisons in the transitivity test, we used an alpha level of 0.01.

## 5.3 Bootstrap error

In order to estimate the error variance of the preferred firing times across repeated measurements, we used a bootstrapping procedure, whereby we computed firing sequences from a randomly chosen subsample of trials. Each subsample contained half of the total number of available trials, which were sampled with replacement. For each data set, 100 such samples were made and the standard deviation of the obtained relative firing times was used as an estimate of the measurement error, referred to as ‘Bootstrap error’, and is shown in Figures 3E and 4C.

## 5.4 Stability over time

To estimate the stability of the firing sequences over time, we split the data in two halves as they were recorded in time, i.e., early vs. late recording blocks (unlike the separation

into odd and even trials described earlier). For the number of recordings used in each analysis and the other details see supplemental Table S1.

## **6 Delays in RF activation**

To estimate the delays in activation caused by the RF properties of different units, we used two different techniques. For the data shown in Figure 1D, we applied the double sliding window technique for the measurement of onset latencies, based on PSTHs (Berenyi et al., 2007). We computed PSTHs with a resolution of 5 ms, and the width of the sliding window was 400 ms. The method rarely entered local minima, and the estimated latencies of the response onset corresponded well to those obtained by visual inspection of the PSTHs. For the data shown in supplemental Figures S1B-C, we computed PSTHs with a bin size of 10 ms, then generated cross-correlation functions of those PSTHs for delays of  $\pm 1000$  ms. The delays in PSTH peaks were then estimated by the same procedures based on fitting a Gaussian as used for the regular CCHs.

## **7 Sequence-triggered average of the local field potential**

The sequence-triggered average (SQTA) of the local-field potential (LFP) displayed in Figure 3B is computed by first averaging the LFP signals recorded across all electrodes with which we recorded also spiking activity. Next, the classical spike-triggered average (STA) of the LFP is computed for each of the neurons in the sequence. Finally, to obtain the SQTA, the relative positions of the individual STAs are offset according to the relative firing times of the corresponding neurons, and then averaged. For example, if a sequence includes two neurons with the preferred firing times of +2 and -2 ms, the two corresponding STAs would be averaged with an offset of 4 ms.

## 8 Orientation and direction tuning of firing times

In each of the polar plots in Figure 5B, a constant time,  $t_c$ , needed to be subtracted from the unit's firing times relative to all the other units for all 12 stimulus directions. The reason was that, unlike rate responses, relative firing times do not have an absolute zero. Hence, the choice of  $t_c$  was arbitrary and in turn, affected strongly the degree to which firing times appeared direction selective (i.e., the 'sharpness' of the tuning curve). Thus, in Figures 5B,C, the sharpness of orientation and direction selectivity should not be compared across firing rates and firing times. We chose values of  $t_c$  that roughly matched the sharpness of the corresponding rate tuning curves in order to facilitate visual comparisons of their orientation and direction preferences. Importantly, the subtraction of  $t_c$  neither affects the estimates of the preferred stimulus direction nor the Pearson's correlation coefficient,  $r$ , calculated between firing rates and firing times across the 12 stimulation conditions.

## 9 Time-resolved analysis

The time-resolved analysis of firing times and firing rates was applied to three data sets of 5 to 7 SUs ( $n = 17$  units in total) that were taken from the three oscillatory SU recordings. Firing sequences and mean firing rates were extracted for analysis windows of 450 to 550 ms duration that were moved in 10 ms steps. All the units in a group had overlapping RFs. To compare the modulation amplitudes of firing rates and firing times, we fitted both types of responses with sine functions of the same temporal frequency as the presented grating stimuli. For firing rates, the modulation amplitude was then determined by comparing the amplitude of the sinusoidal modulation of the rate response by the grating

(F1) to the mean change in response amplitude from spontaneous to stimulus-evoked activity (F0), known as F1/F0 ratio (Movshon et al., 1978a, 1978b; Skottun et al., 1991). Rate responses with an F1/F0 ratio  $>1$  were classified as simple, whereas rate responses with an F1/F0 ratio  $<1$  were classified as complex. For firing times, the modulation amplitude of the fitted sine was not normalized, but simply expressed in milliseconds.

## **10 Single-trial estimates of preferred firing times**

When estimating the preferred firing times from single trials, as shown in supplemental Figure S9, it was not possible to test, at the same time, the degree of additivity between pair-wise time delays. Instead, the procedure described in sections 4 and 5 needed to be simplified, and additivity needed to be assumed rather than tested. As mentioned, single trials do not have enough data (i.e. spike counts) to complete a full pair-wise analysis of preferred spike delays. Hence, we simply computed a CCH between each unit on the one hand and, on the other hand, a large conglomerated multi- unit consisting of all the remaining units in that data set. In theory, the delay of the resulting CCH will correspond to the unit's preferred firing time extracted from pair-wise delay measurements under the assumptions that i) all pair-wise delays are additive and ii) all units have identical firing rates. This method has the advantage of being able to deal with low spike counts: No problems of incomplete delay networks exist and, in total, fewer parameters need to be estimated ( $n$  CCHs instead of  $n(n-1)/2$  CCHs per data set). Thus, the preferred firing times extracted by this simplified method were sufficiently accurate to warrant the analysis of the delays across individual trials. However, the advantages of this approach are traded off against the following shortcomings: i) Additivity of pair-wise delays is assumed, not tested. ii) Neurons with low rate responses are underrepresented in the resulting



conglomerated CCH. iii) By lumping many neurons with different relative firing times into the same CCH, the CCH peak becomes broader than those between individual units. Hence, the estimates of the time delays suffer. Therefore, the present analysis should not be considered a replacement for the full analysis described in sections 4 and 5.

To quantify the trial-to-trial variability of the measurements, for each neuron a coefficient of variation (CV) was calculated from the standard deviation (SD) of the trial-by-trial estimates, and from their mean, so that  $CV = SD / mean$ . Note that computing the CV based on the mean firing time was not possible because firing times have an arbitrary zero and hence, the mean firing time of a neuron is meaningful only in relation to other firing times. Therefore, trial-to-trial variability of the relative firing times was compared to that of firing rates by their respective ability to distinguish stimuli. Similarly to the analysis of mutual information (MI) in Figures 6A-C, we calculated the stimulus-dependent differences in the respective estimates of rates and times and plotted those as a function of the trial number (supplemental Figs. S9B-G). We used the same data as in the MI analysis but limited the comparisons to the stimuli with maximum differences in the orientation (or direction in case that orientations were identical). In all comparisons between time and rate codes, identical data segments were always used and hence, rate and time estimates were made always based on the same set of action potentials.

## **11 Consistency of pair-wise firing delays across individual coincident events**

The distribution of the time delays between correlated discharges is reflected by the width of the CCH centre peaks: the narrower the peak, the smaller the variability. We calculated the likelihood that two action potentials, fired in short succession, occurred in the

preferred rather than in the opposite temporal order. The scatter plots in supplemental Figure S2C,D, which display the proportion of consistent firing orders as a function of the size of the measured preferred delay, were fitted with the upper half of a sigmoid function, using a least-squared errors method for nonlinear regression. The fitted function was of the form

$$f(b, m, t, x) = 100 \times \left[ b + \frac{1}{m + e^{-tx}} \right], \quad (\text{Equation S4})$$

The sigmoid was chosen for theoretical reasons because it represented the cumulative distribution of a Gaussian function, which fitted well the centre peaks of CCHs. Hence, the proportion of consistent firing orders corresponded to the area beneath a Gaussian function (the CCH peak), as calculated from the CCH center to its corresponding end. For the oscillatory CCHs, the values for the parameters  $b$ ,  $m$  and  $t$  were 0.20, 2.24 and 0.49 without shift predictor correction and 0.01, 0.84 and 0.20 with shift predictors subtracted. For the non-oscillatory CCHs, these values were 0.52, 22.10 and 6.29 without shift predictor correction and 0.76, 1.35 and 0.44 with shift predictors subtracted. The fraction of consistent delays was computed for all 1351 CCHs obtained from the 78 investigated SUs (724 CCHs from oscillatory, and 627 CCHs from non-oscillatory data sets).

When studying the consistency of firing orders, one needs to take into account the occurrence of spurious correlations, i.e., coincidences that occur by chance and do not result from specific timing mechanisms such as neuronal synchronization. Evidence suggests that cortical neurons acting as read-outs of time-structured inputs are able to ignore spurious coincidences as they adapt their sensitivity to the level of background activity, and thus also to the level of spurious coincident events (Azouz and Gray, 2000,

2003). To remove spurious coincidences from the CCHs, shift predictors were calculated and subtracted from the original CCHs. Shift-predictors are CCHs computed between pairs of spike trains recorded in different experimental trials. Thus, they do not detect internal coordination of firing but preserve the auto-correlation structure of the spike trains and hence, allow for the assessment of spurious correlations (see the example in supplemental Fig. S2A,B) (Perkel et al., 1967).

## **12 Robustness to assembly size**

To investigate how the estimated firing times change if the network is sub-sampled, for each neuron we successively decreased the number of additional neurons used to compute its firing time, from the original number, down to two neurons (i.e. a sequence of three neurons in total). For each neuron the number of other neurons with which relative firing times were computed (network size) was reduced gradually from the maximum number down to two neurons. This step-wise reduction was repeated ten times per neuron and per stimulus condition, and the standard deviation of the estimated firing times was averaged across all network sizes and stimulus conditions.

## **13 Classification of spike waveforms**

To classify neurons as fast-spiking (FS) or regular-spiking (RS), we adapted the approach of Mitchell et al. (2007). We first measured the time delay between the peak (maximum) and trough (minimum) of the spike waveform. Spike waveforms were inverted where necessary to enter first the peak then the trough. To characterize the time course of the repolarization for each waveform, we made the following measurement: We first computed the difference between consecutive voltage values of the waveform after the

trough, and then did a linear regression on these difference values to calculate their slope. If the repolarization decelerated (FS neurons), the differences between consecutive data points became smaller, leading to a negative slope. If the repolarization accelerated (RS neurons), the difference values would become larger, leading to a positive slope.

#### **14 Oscillation score**

The oscillation strength in a data set was estimated by computing its oscillation score (OS; Muresan et al., 2008). To this end, an auto-correlation histogram (ACH) was computed with 1 ms resolution for each unit and for each investigated stimulus condition. The OS is defined as the ratio between the power at the strongest frequency within a band of interest, which covers in our case the upper beta/lower gamma band (25-50 Hz), and the average power of all other frequencies in that band. The OS computes a corrected oscillation power by removing the artifacts induced by the presence of the centre peak in the ACH and by possible bursting behavior of the cells. An OS of 5 for a preferred frequency of 40 Hz indicates that the oscillatory power at 40 Hz is five times larger than the average power in all the other investigated frequencies.

Within a data set, the units tended to have similar OS, indicating a collective fluctuation in the level of oscillatory activity. Thus, the overall oscillation strength for a data set was estimated simply as the average OS for all individual ACHs. In accordance with the recommendations given in Muresan et al. (2008), data sets with an average  $OS \geq 5$  were classified as oscillatory, and data sets with an  $OS < 5$  as non-oscillatory. For the responses to gratings drifting in different directions, seven data sets were classified as non-

oscillatory (OS: 1.5 – 4.8; mean: 3.2; median: 3.5) and three as oscillatory (OS: 6.0, 6.3 and 7.2).

## **15 Oscillation frequency**

To determine the dominant frequency of oscillation for each firing sequence, we first selected the pairs of neurons whose CCHs had a goodness of fit of  $r^2 > 0.5$ . For those pairs, we computed a Fast Fourier Transform over a time window of  $\pm 128$  ms to obtain the power spectrum. The power spectra showed a well-defined single peak in the range of 20–60 Hz in almost all cases. The frequency of each CCH was defined as the frequency with the highest power within this frequency range. The CCHs belonging to the same firing sequence usually displayed very similar oscillation frequencies, the differences rarely exceeding 5 Hz. Thus, the dominant frequency of a firing sequence was simply determined by averaging the oscillation frequencies displayed by the individual CCHs.

## **16 Spike train thinning**

To test the robustness of the Oscillation Score to small spike count (shown in supplemental Figs. S10E-G), all spike trains were thinned by randomly removing spikes in steps of 100. At each step a new Oscillation Score was computed. The smallest number of spikes used was always between 200 and 300.

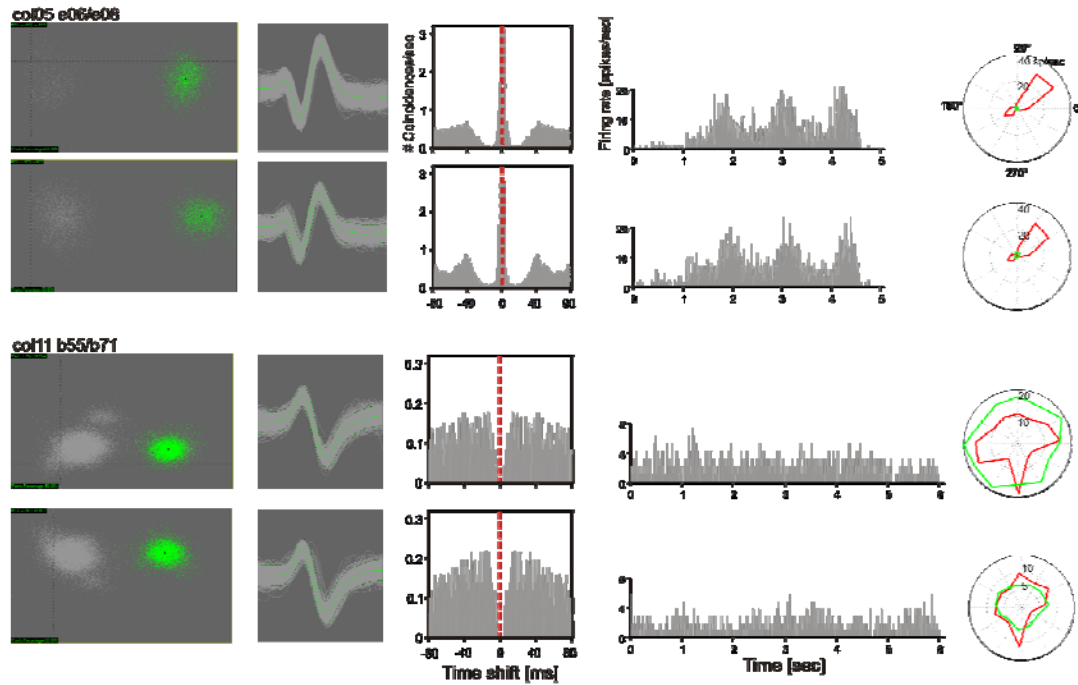
## **17 Non-stationary Poisson process**

In the simulation shown in supplemental Figure S12, we used a two-sided exponential decay as a basic function for the firing probability,  $p$ , by calculating the firing probabilities in each oscillation cycle as  $p = \exp(-|t - t_0| / \tau)$ , where  $t$  is time,  $t_0$  is the time at which the

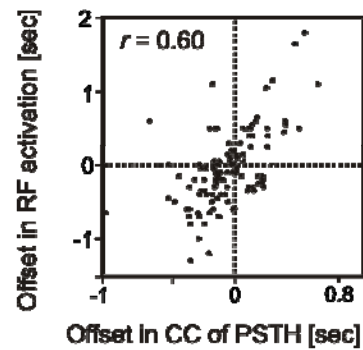
firing probability reaches its local maximum, and  $\tau$  is the decay constant. This model is illustrated in supplemental Figure S11. The functions describing oscillation cycles (fast component) are next concatenated in a quasi-periodic way (with variations in the oscillation periods) into a prototype of  $P$ , which was then scaled for each neuron and stimulation condition individually based on the corresponding PSTH (slow component). The scaling procedure adjusted the firing rates of the model to those of the data. Hence, the simulated spike trains reproduced accurately the PSTHs of the data (not shown). This function produces probability profiles similar to those shown in supplemental Figure S11C. Other functions, including one-sided exponential decays produce qualitatively similar results. At the onset of each simulated trial, the initial phase of  $P$  was randomized such that the phase difference between a given neuronal pair was kept constant, the value of which was determined by the average time delay extracted from the real CCH. The model did not have “free” parameters as all estimates came from the data: oscillation frequency, the decay constant  $\tau$  (determined from the width of centre peak) and the variation of the oscillation periods (determined by the heights of the satellite peaks in the CCH relative to the centre peak).  $\tau$  was kept constant. For each calculation of a CCH 10,000 trials were simulated. Note that this model does not make any assumptions about the origin of the delays in synchronization.

**Supplemental Figure S1: Spike waveforms and independence between CCH delays and activation delays**

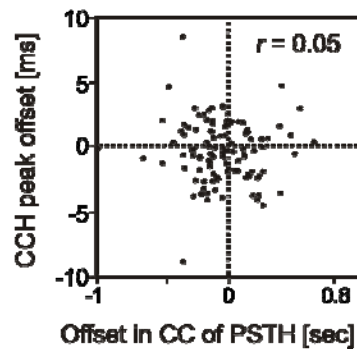
**A**



**B**



**C**



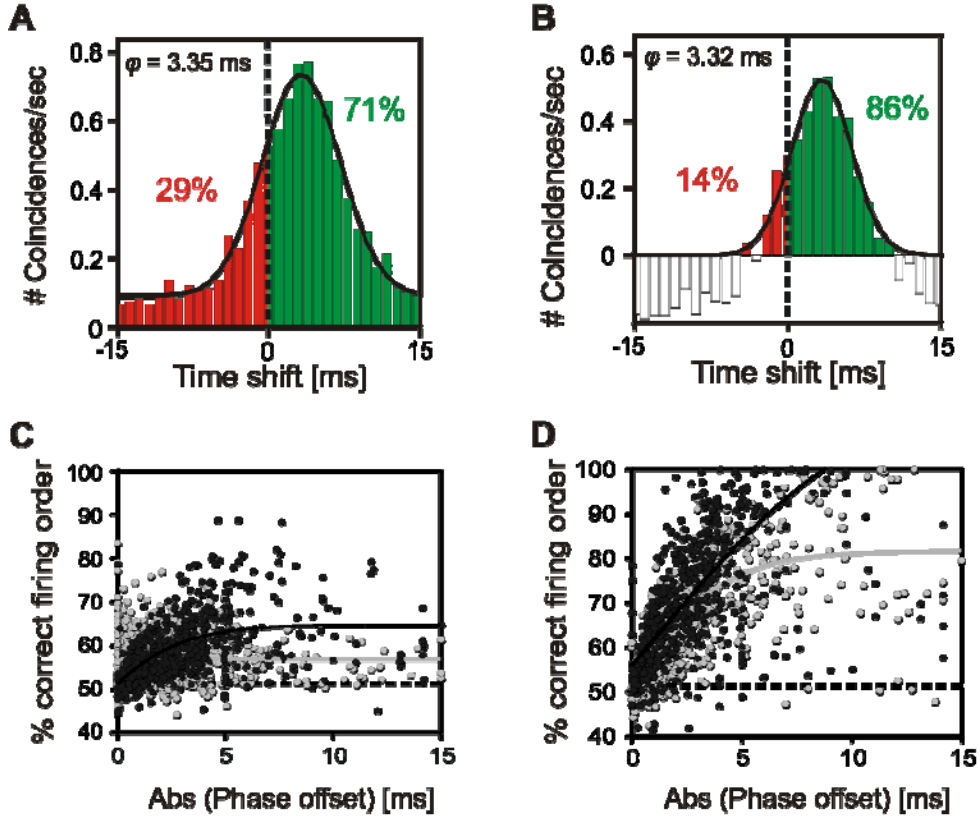
(A) Two examples of spike waveforms classified as single units. Panels from left to right: Two-dimensional representation of clusters formed by the three-dimensional principal component analysis of multi-unit signals; Spike wave forms of the cluster in green; Auto-correlation histograms; PSTHs for responses to one stimulus (the stimulus was presented

from 1 to 4.5 seconds after the start of the recording for the data shown in the upper panels, and between 1 and 5 seconds in the lower panels); Direction tuning to grating stimuli moving in 12 directions. Green: spontaneous activity. Red: Stimulus response. To indicate the stability of the recording, two recordings are made for each unit, 3 hours apart (unit above) and 7 hours apart (unit below).

**(B)-(C)** Analysis of the delays in peak activation, supplementing the results on the delays in the onset of receptive field (RF) activation shown in Figure 1D. The delays in the PSTH peaks are estimated from the cross-correlations of the PSTHs, as described in section 6 of supplemental Methods. **(B)** The delays in PSTH peaks (offset in the cross-correlation of the PSTHs) correlate positively with the delays in the RF activation. **(C)** The delays of PSTH peaks between pairs of neurons do not correlate with the synchronization delays for the same neurons extracted from their CCH.



**Supplemental Figure S2: Consistency of pair-wise firing delays across individual coincident events**



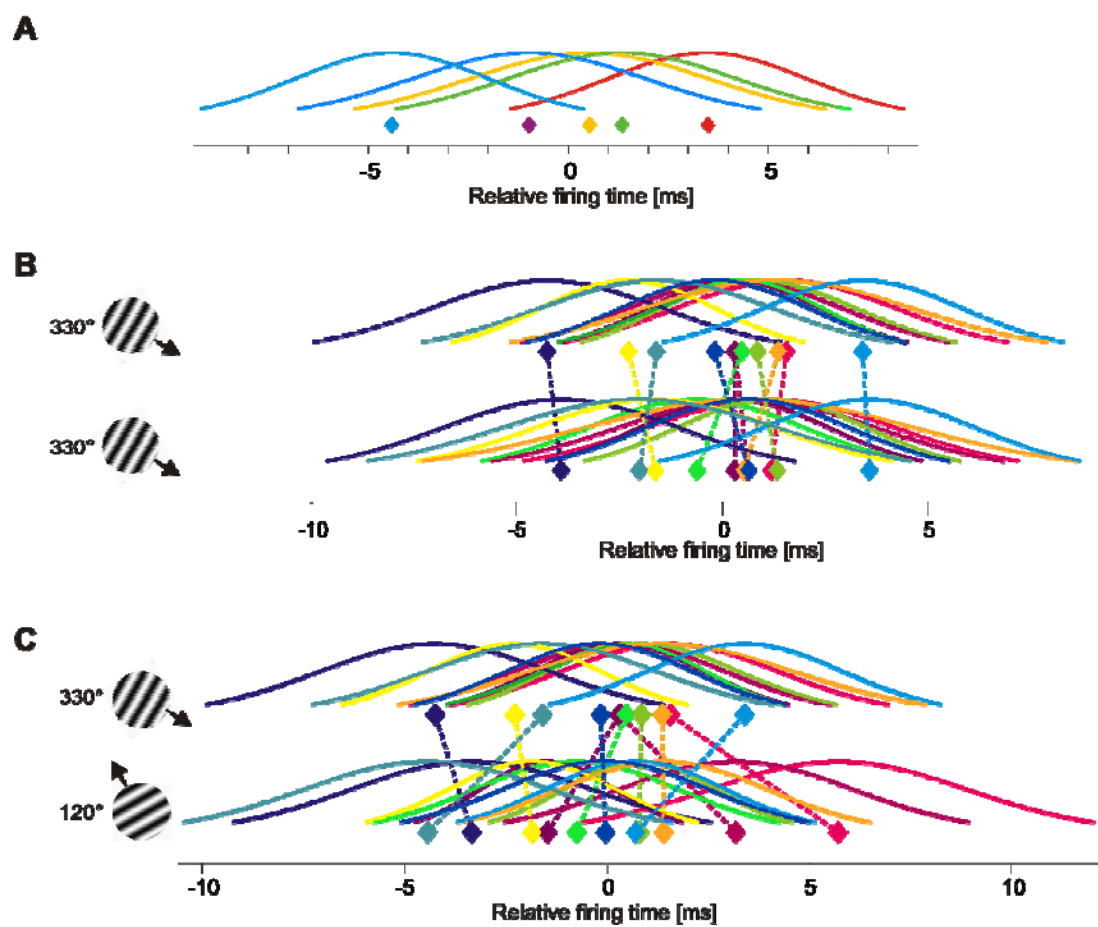
(A) Example CCH between two SUs in response to visual stimulation, shown for time shifts of  $\pm 15$  ms. Black solid line: fitted Gaussian curve to estimate the preferred time delay. Vertical dashed line: zero delay. The area at the right side of the zero delay (green color) represents the coincidences that occurred in the temporal order consistent with the estimated preferred delay,  $\varphi$ . The coincidences on the left side (red) occurred in the opposite, inconsistent, order. The respective proportion of the events that occurred in a given order is indicated as percentage.

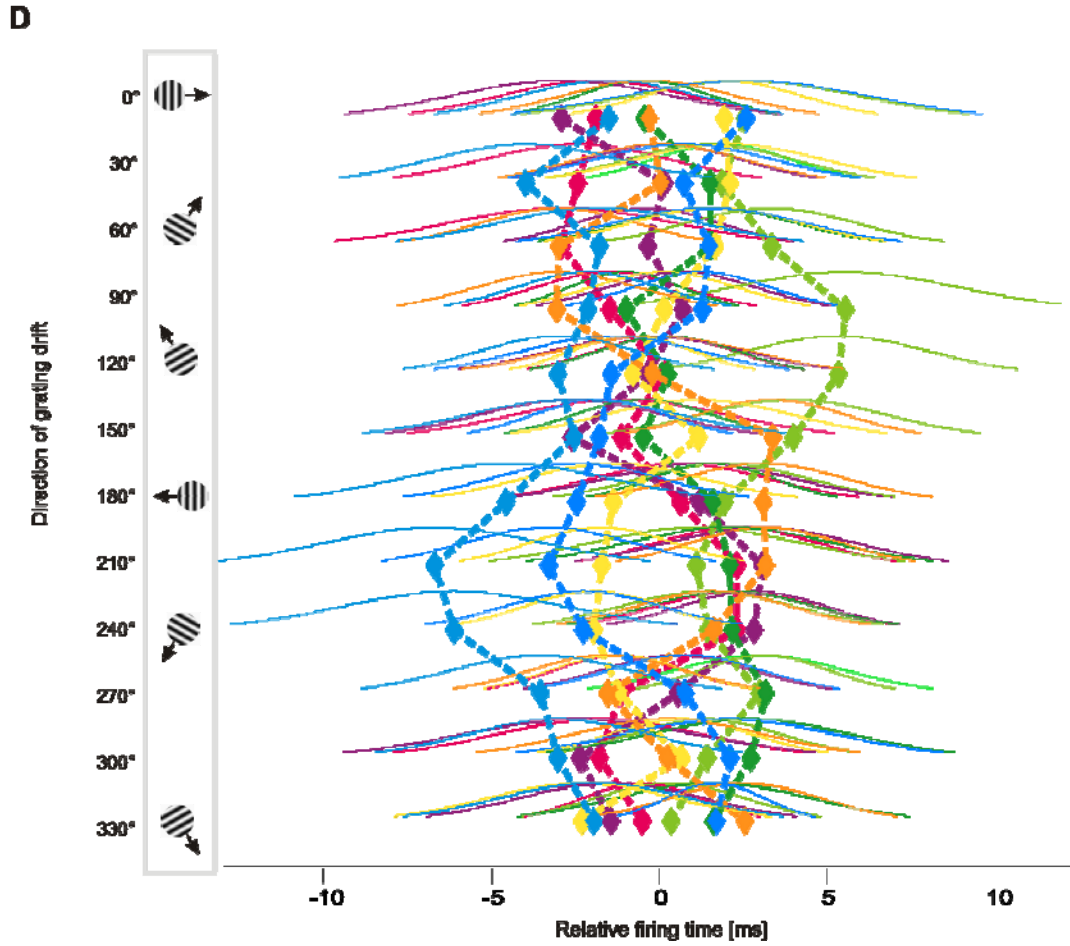
**(B)** The same CCH as in **(A)** but with the shift-predictor subtracted to correct for spurious coincidences. All notations are the same as in **(A)**. Only the positive coincidence counts were used to calculate the proportions of events that occurred with correct and incorrect firing orders (green and red colors).

**(C)** The proportion of consistent coincident events (y-axis) shown as a function of the size of the measured preferred delay between a pair of neurons (shown as the absolute value; x-axis). Black dots: CCHs from the data sets with strong oscillations. Gray dots: data sets with weak or absent oscillations. Solid lines: fitted sigmoid function. Dashed line: the chance level of 0.5.

**(D)** Same scatter plot as in **(C)**, but for CCHs with shift-predictors subtracted. Without shift predictor correction, the proportion of correct firing orders was generally moderate and showed only a small advantage of data sets with gamma oscillations (black dots) over data sets lacking oscillations (grey dots). In contrast, with shift predictor correction, proportions were considerably larger and depended strongly both on the size of the preferred delay and the oscillation strength. The methods used for these analyses are detailed in supplemental Methods, section 11.

Supplemental Figure S3. Average peak widths of CCHs





Average peak widths of CCHs, indicating the variability of the delays in individual coincident events, shown for the same data as in

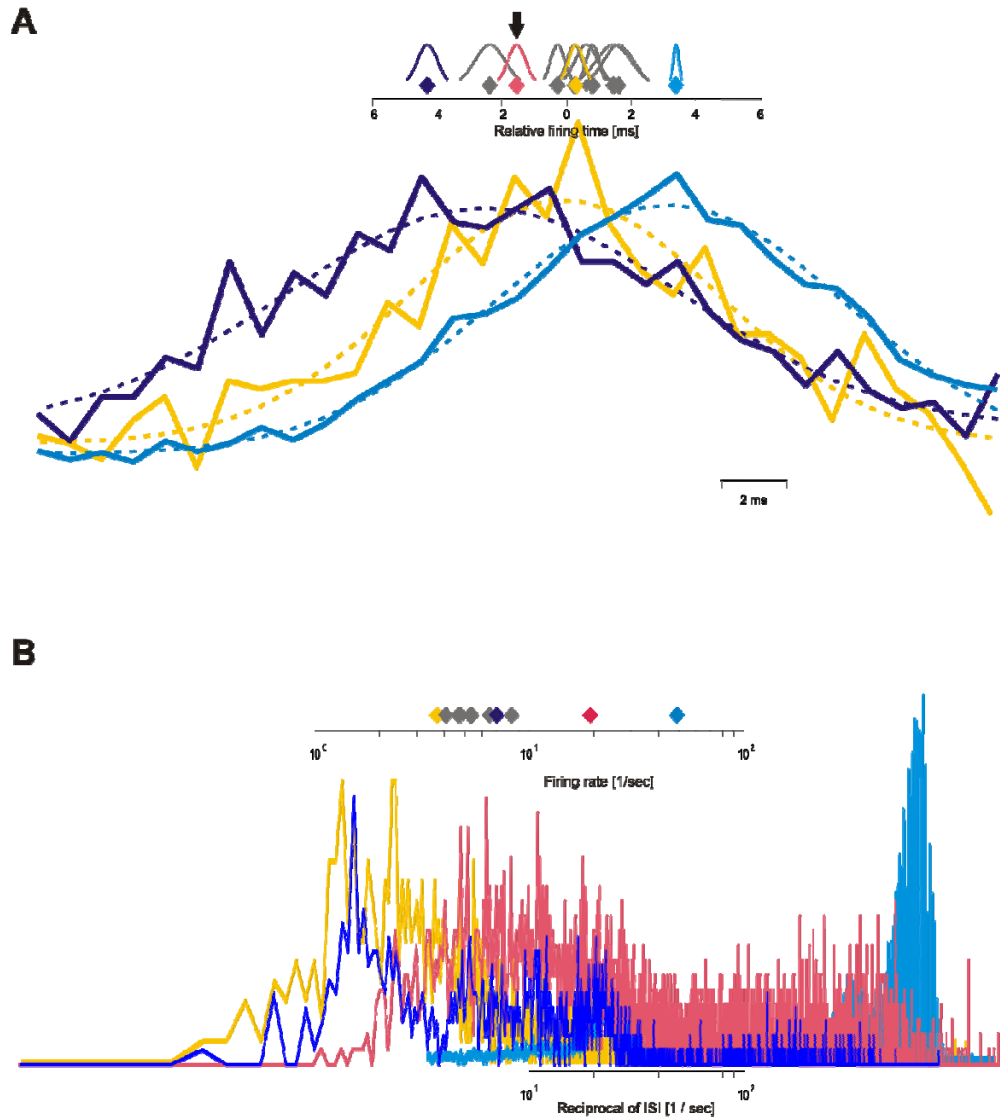
(A) Figure 3A in the main text

(B) Figure 3G in the main text.

(C) Figure 4A in the main text

(D) Figure 5A in the main text.

**Supplemental Figure S4: Event-to-event variability of firing times and firing rates**

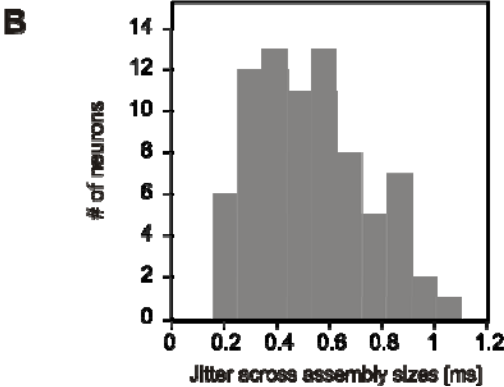
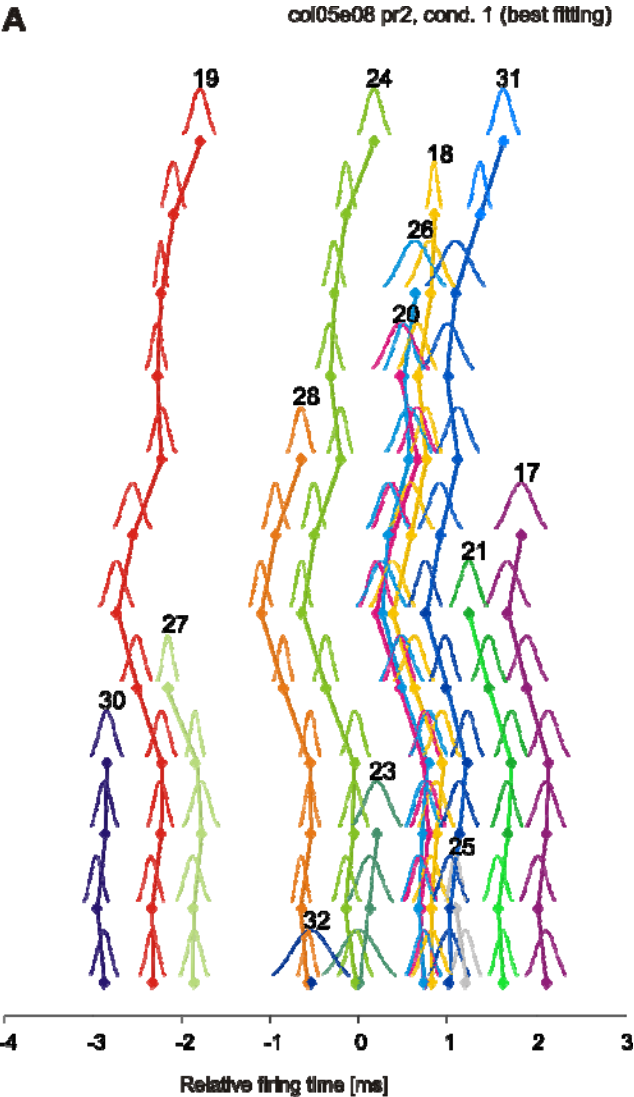


Estimates of firing times and firing rates from the smallest possible amount of data—i.e. from individual coincident events in case of firing times and from individual inter-spike-intervals (ISI) in case of firing rates. The distributions of the single-event estimates (bottom of each panel) are shown in relation to the corresponding average estimates computed from all data available (120 stimulus presentations; top of each panel).

(A) The firing sequence from Figure 3G with three of the corresponding CCH centre peaks, all data shown on the same time scale. The three CCHs show relations between four neurons, one chosen as the reference in all cases (vertical arrow). The zero of each CCH is aligned with the position of the reference. The plot illustrates the magnitude of the average delay (top) in comparison to the width of the distribution of individual delays (bottom). Gray: Other neurons in the sequence, the CCHs of which are not shown. Small Gaussians: Additivity errors. Dashed lines: Gaussians fitted to the CCH peak to estimate the preferred time delays between units. The four example neurons were chosen to cover the maximum spans of both firing times and firing rates.

(B) Top panel: Average firing rates of the neurons shown in (A). Bottom panel: Corresponding distributions of the instantaneous firing rates (computed as  $1/\text{ISI}$ ; inter-spike-interval) of the same neurons. Due to highly skewed distributions of the instantaneous firing rates, the plots are shown in a logarithmic scale. These results are consistent with previous conclusions: The variability and errors of firing times are comparable to those of firing rates.

Supplemental Figure S5. Robustness of firing sequences to assembly size

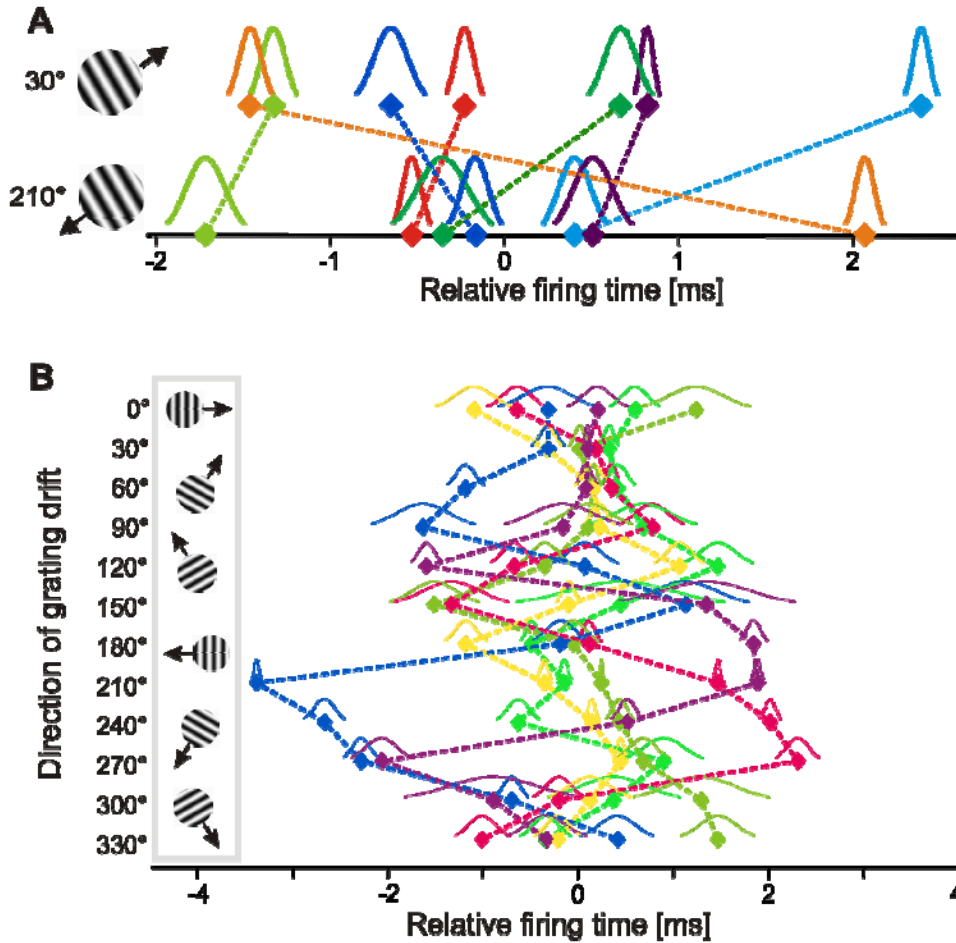


**(A)** Firing sequences extracted for different numbers of neurons. The full firing sequence of 13 neurons is shown on the bottom. At each step one neuron is removed from the analysis and a new firing sequence is calculated.

**(B)** Distribution of the variability of firing time estimates across networks of different sizes for all 78 single-units investigated. SD: standard deviation. For details see section 12 of supplemental Methods.



**Supplemental Figure S6: Firing sequences in multi-unit activity**

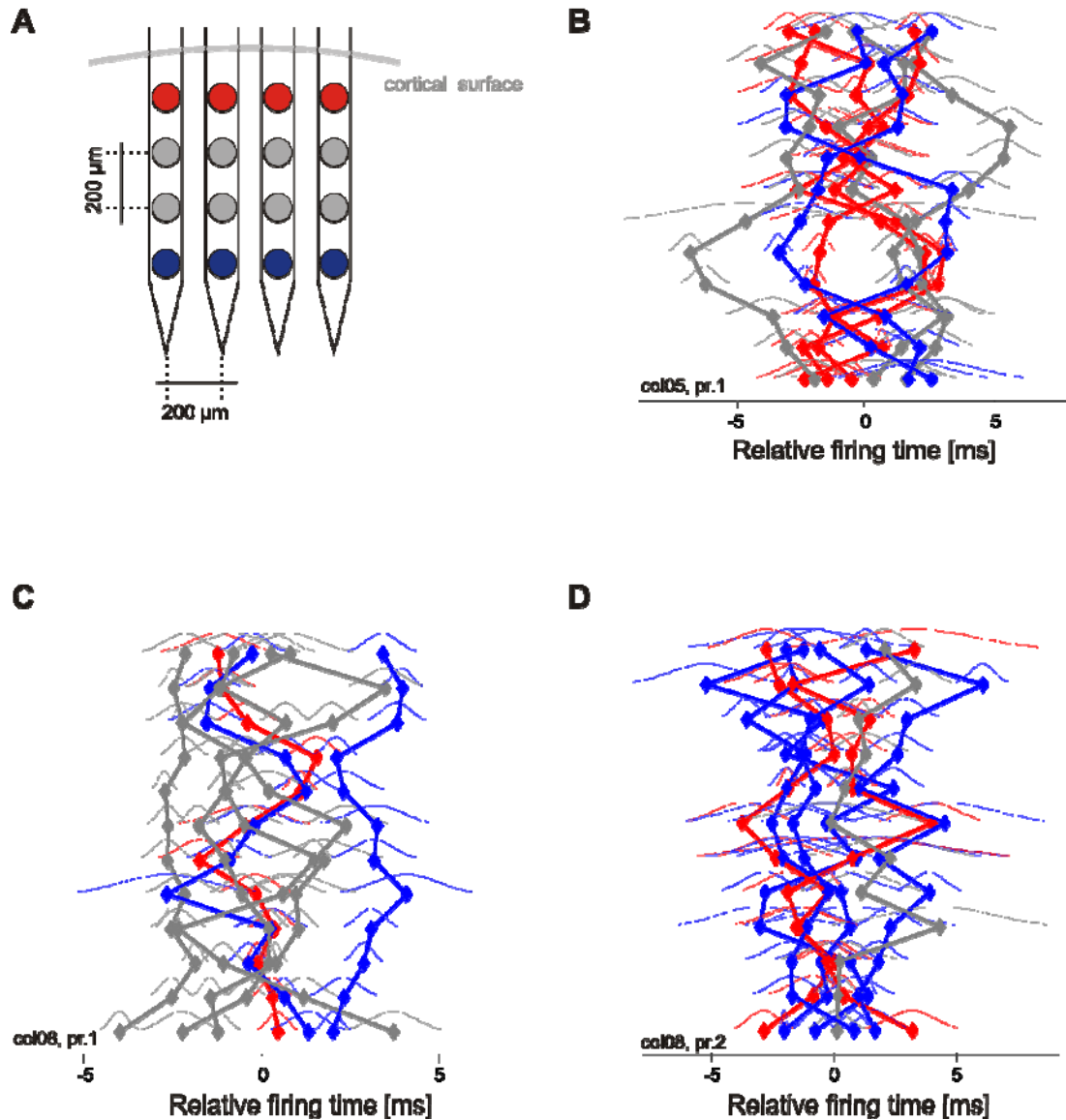


(A) Example of a change in a firing sequence computed for a group of multi-units (MU) when the drifting direction, but not the orientation, of a grating stimulus changed. MU firing sequences were computed for three of five experiments (i.e. six of ten data sets) by analyzing the non-sorted spiking signals prior to the application of the spike sorting procedures. Otherwise, the methods were identical to those applied to single units.

(B) Changes in firing sequences shown for 12 directions of grating drift, indicating that the firing times of MUs exhibit also orientation and direction tuning. As expected, the spans of firing sequences of MUs were smaller than those of SUs (on average,  $3.62 \pm 4.53$

for MU vs.  $6.60 \pm 2.99$  for SU). The statistical disadvantages coming from the smaller magnitudes of delays were associated with several advantages that resulted from the overall smoother CCHs and thus, more accurate fits of Gaussian functions (i.e., more accurate estimates of time delays), which also led to smaller measurement errors. Thus, overall, a smaller number of trials was needed to estimate a firing sequence between MUs reliably (often, 10 trials were sufficient).

**Supplemental Figure S7: Stimulus-dependent changes in firing times for neurons from different cortical depths**

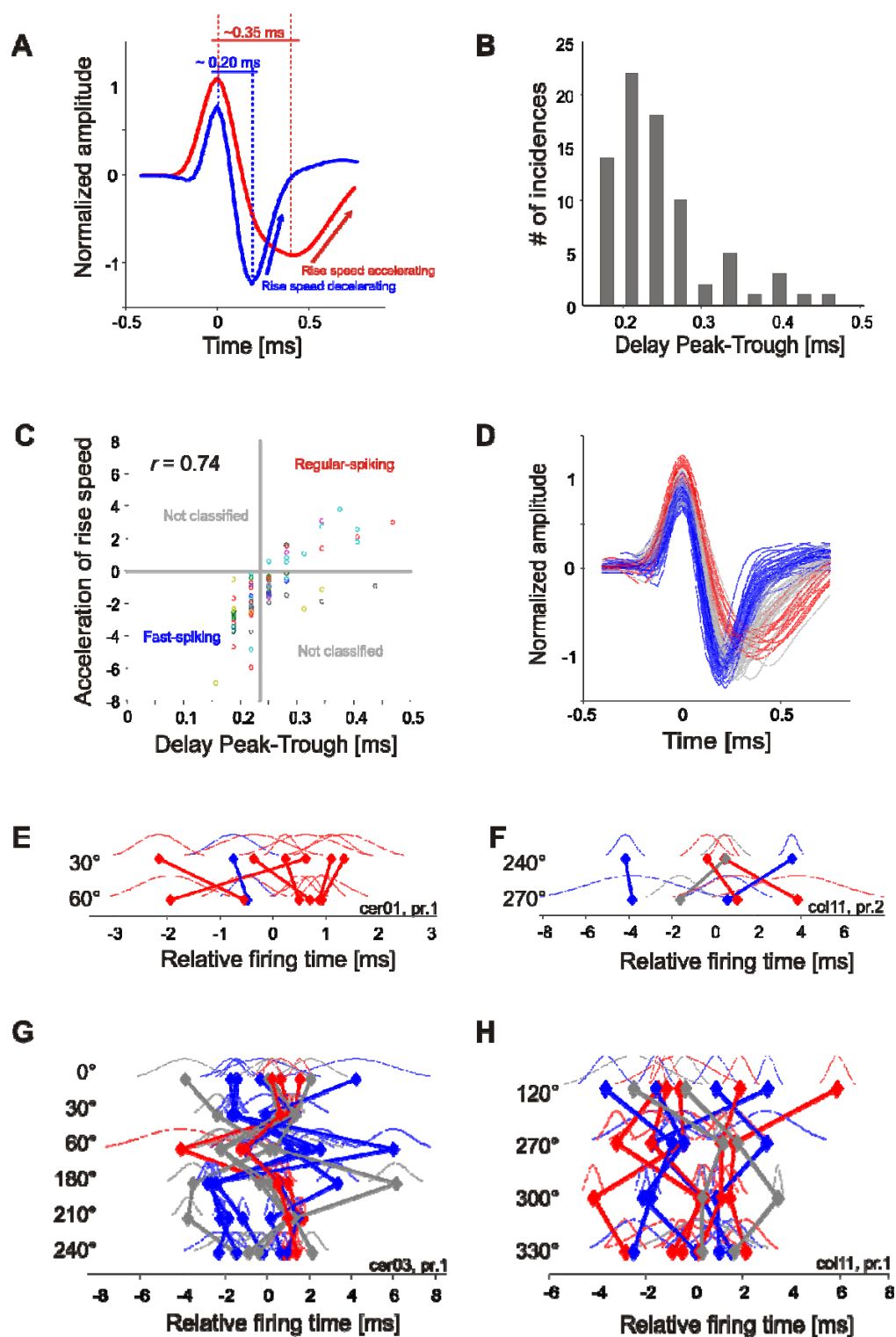


(A) Schematic illustration of the spatial configuration of recording sites.

(B)-(D) Firing sequences in response to 12 grating orientations (30° steps). Red: Neurons recorded by the top row of recording sites. Blue: Neurons recorded by the bottom row of recording sites. Grey: Neurons recorded in the two intermediate rows of recordings sites.

The relative firing times of neurons from different recording depths are clearly interwoven.

Supplemental Figure S8: Firing sequences for neurons classified as fast-spiking and regular-spiking



**(A)** Two example spike waveforms of neurons classified as fast-spiking (FS; shown in blue) and regular-spiking (RS, shown in red). Neurons were classified according to the delay between peak and trough, and the time course of the repolarization: In FS neurons, repolarization was fast directly after the trough and leveled out subsequently (steeper slope of the arrow; rise speed decelerating), in RS neurons, repolarization was initially slow, speeding up gradually (shallower slope of the arrow; rise speed accelerating).

**(B)** The distribution of peak-trough delays for all 78 neurons investigated.

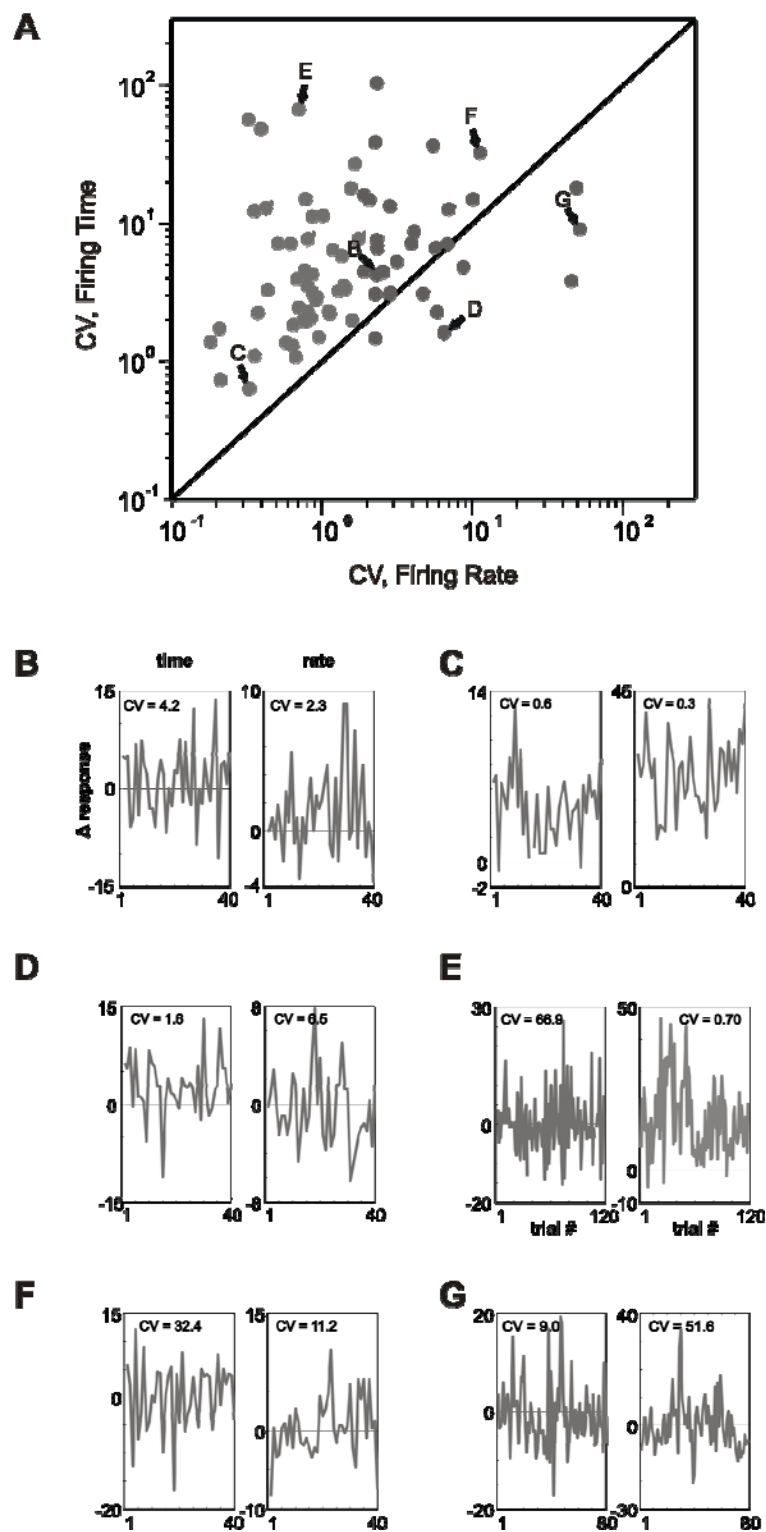
**(C)** Relation between peak-trough delay and repolarization, as described in **(A)**. For the calculation of the acceleration in rise speed, see section 13 of supplemental Methods.

Colors denote different data sets. Neurons were classified as FS if their waveforms had a peak-trough delay  $< 0.23$  ms and a decelerating rise speed; and classified as RS if their spike waveforms had a peak-trough delay  $> 0.23$  and an accelerating rise speed. Other spike waveforms were treated as unclassified.

**(D)** Average spike waveforms for all 78 neurons, grouped into FS (blue), RS (red) or unclassified (grey).

**(E)-(H)** Examples of firing sequences containing both FS and RS neurons in response to different grating orientations (marked on the left-hand side). The relative firing times of the two neuron-types were not separated, but strongly interwoven.

Supplemental Figure S9: Trial-to-trial variability of firing times and firing rates.

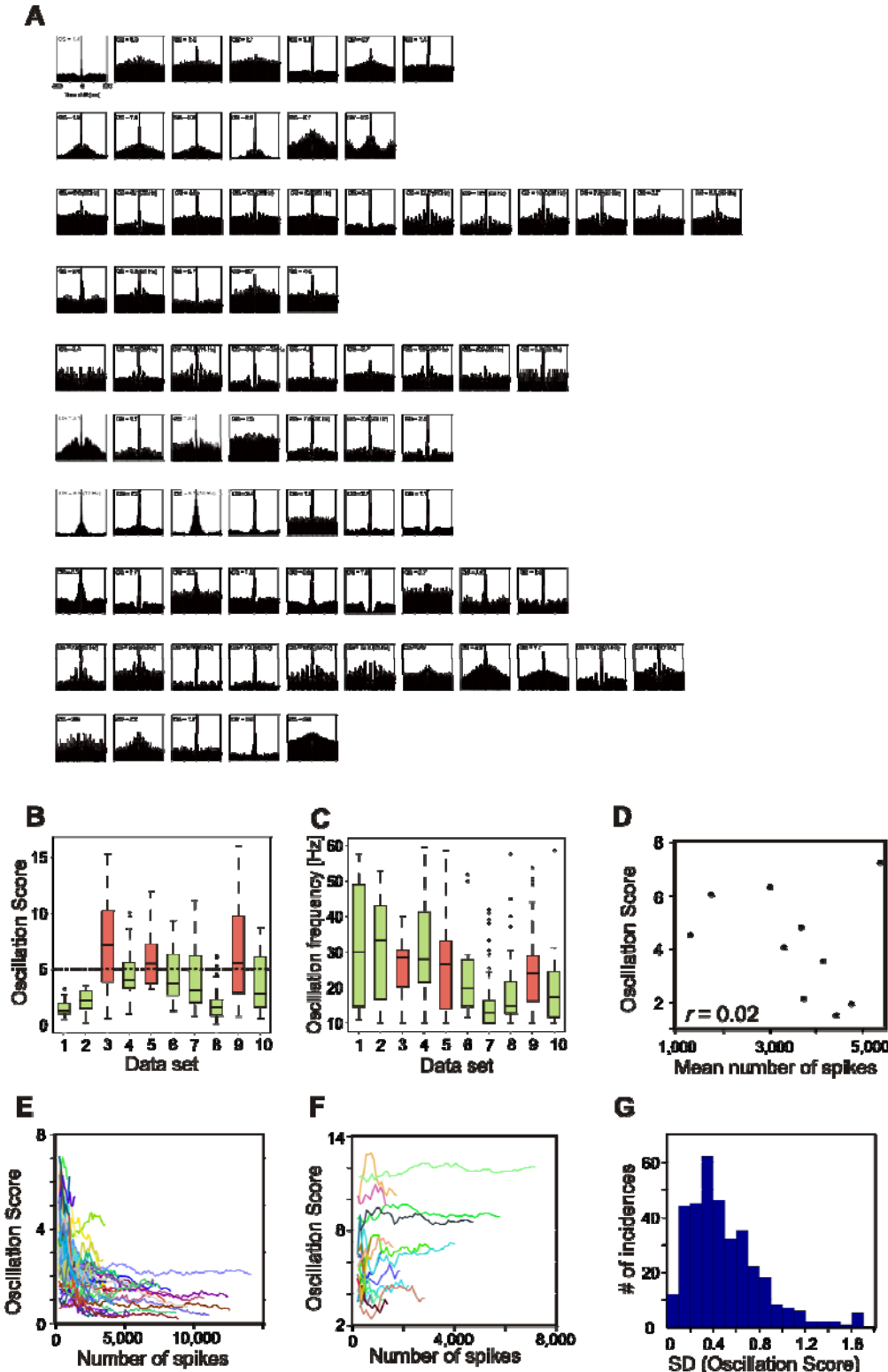


(A) Trial-to-trial variability of firing rates (x-axis) and relative firing times (y-axis) for all investigated neurons, computed as coefficient of variation (CV, for details see supplemental Methods, section 10). For each stimulus presentation (trial) the change of the response (i.e., firing rate or firing time) was calculated for the most optimal grating and 90 degree offset orientation. Arrows point to neurons shown in detail in (B)–(G). Due to highly skewed distributions of CV values, the plot is shown in log-log coordinates. The scatter of CV values confirms the conclusions drawn from the analysis of Mutual Information shown in Figure 6 - stimulus-dependent changes across individual stimulus presentations were somewhat less variable for firing rates than firing times but, firing times nevertheless still operated at a comparable level of reliability.

(B) – (G) Trial-to-trial differences ( $\Delta$  response) in the relative firing times (left panel) and firing rates (right panel) in response to the same two gratings as in (A) shown for the six representative neurons indicated by arrows in (A). Note that CV values were large when  $\Delta$  response was  $\sim 0$ , but small for large  $\Delta$  response.



Supplemental Figure S10: Oscillation strength and frequency for all data sets



**(A)** Auto-correlation histograms (ACH) of spike signals. For each neuron in this study, one ACH is shown. Rows: Each of the ten investigated data sets. The examples are chosen to represent a stimulus orientation medium optimal for evoking oscillations for a given neuron, so that the shown ACHs reflect well the overall oscillation score assigned to each data set. Upper left corner in each ACH: Oscillation Score (OS) for that ACH. Parentheses: the extracted oscillation frequency, provided if  $OS > 5$ .

**(B)** Distribution of Oscillation Scores for each data set. Horizontal dashed line: Threshold for classifying data sets into strong or weak oscillations. Red boxes: data sets with strong oscillations. Green boxes: data sets with weak oscillations. Thick black horizontal lines: median. Box ends: 25th percentile. Whiskers: the largest and smallest data points not exceeding the box end by  $1.5 \times$  the box size. Crosses: outliers. X-axis: Nominal ordering of data sets matching the order in **(A)**. For details on the computation of the OS, see section 14 of supplemental Methods.

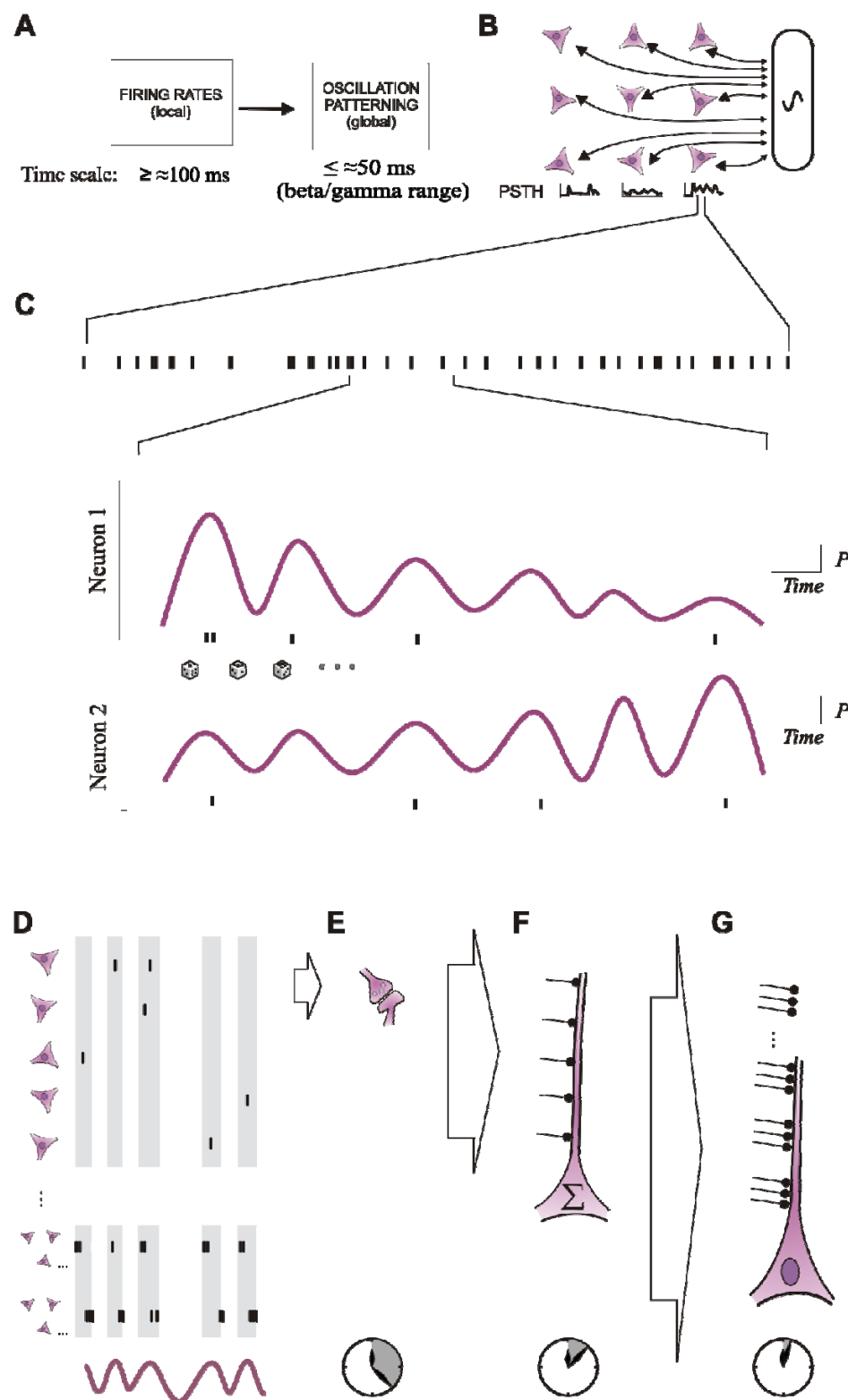
**(C)** Same as **(B)** but for oscillation frequencies. For details on the computation of oscillation frequencies, see supplemental Methods, section 15.

**(D)** The OS in a given data set as a function of the average number of spikes per auto-correlation histogram (ACH) in the data set.

**(E) - (F)** The effect of spike train thinning on the estimated OS. Spikes are randomly removed from the spike trains (see section 16 of supplemental Methods). Shown are changes of OS for multiple ACHs (color coded) belonging to the two data sets in which spiking thinning produced respectively the strongest **(E)** and the weakest effects **(F)**. Typically, small spike counts lead to an overestimate of the oscillation strength—a result consistent with previous reports (Muresan et al., 2008).

(G) Distribution of the variability of OS, expressed as standard deviation (SD) across all thinned spike trains. The number of action potentials had only a small effect on the estimate of the OS (dominant SD = 0.4), indicating that the lower coding capability of firing rates and firing times in the absence of oscillations cannot be explained by a smaller number of action potentials in these recordings.

**Supplemental Figure S11: Theoretical proposal for the generation and readout of firing sequences**



**(A)** Illustration of a non-stationary Poisson process that generates synchronized oscillatory activity. The probability of generating a spike at any moment is the function of a slow component denoted firing rate and a fast component denoted oscillatory patterning. The two components are adjusted independently as neurons can generate the same firing rate with stronger and weaker oscillatory patterning.

**(B)** Oscillatory patterning is a global rhythm shared across neurons. In contrast, neuronal firing rates change independently of each other, as indicated by different PSTH shapes (bottom).

**(C)** Top: An example spike train generated by a non-stationary Poisson process. Bottom: Probabilities of generating action potentials,  $P$ , resulting from the two components in **(A)**—shown for two neurons with different firing rate dynamics but synchronized oscillation. The probability functions are synchronized globally but the occurrence of individual spikes is determined locally, i.e. independently (illustrated by dice throws). Therefore, the firing probabilities of different neurons correlate only on the fast but not on the slow time scale.

**(D)–(G)** Illustration of resulting firing patterns and potential readout mechanisms. Clock hands: Relative speeds of each of the readouts.

**(D)** Top: Spike trains of neurons responding at a preferred delay relative to the oscillatory cycle, represented by the trace at the bottom of the figure. Bottom: Combined spike trains of groups of neurons with similar preferred delays. In this case timing information can be extracted in each oscillation cycle.

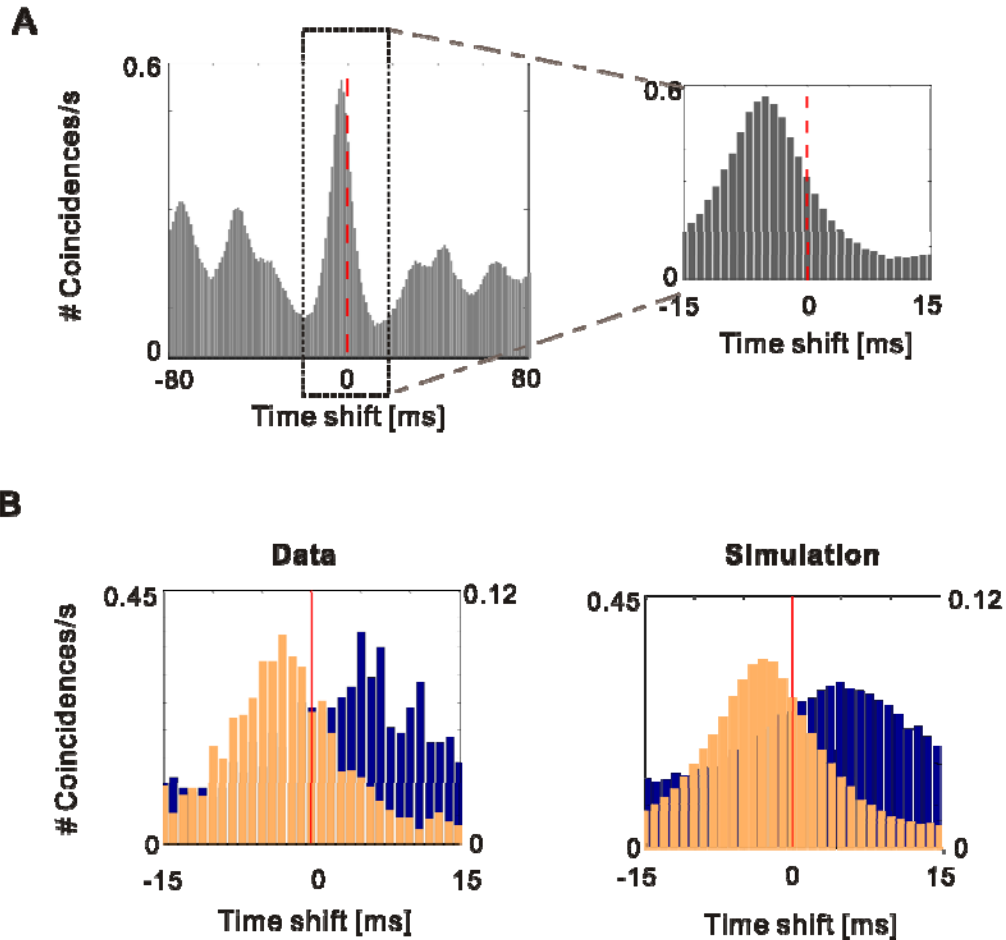
**(E)** Temporal integration at a single synapse selective to the temporal order of spike arrivals. Spike-timing-dependent plasticity predicts that a synapse will strengthen if the

pre-synaptic neuron fires shortly before the post-synaptic neuron, and will weaken if the order is reversed.

**(F)** A combination of spatial and temporal integration through coincidence integration: Fast dendritic summation (e.g. ~15 ms time constant) may be most efficient when distal inputs arrive earlier than those proximal to the soma. Slow integration at the soma (e.g. ~100 ms time constant) may average out the errors in the input sequences detected at dendrites.

**(G)** An increase in the total number of simultaneous inputs may reduce or even eliminate the need for slow integration at soma.

**Supplemental Figure S12. Example validating the model shown in supplemental Figure S11**



Example CCHs created by a non-stationary Poisson process described in supplemental Figures S11A-C.

(A) CCH from simulated spike trains matching the PSTHs of neurons shown in Figures 1A-C. The firing probabilities changed according to a combination of the the neurons' individual PSTHs and a common oscillatory process, which was independent of the neurons' firing rates. For more details, see supplemental Methods, section 17.

(B) A calculation similar to that in (A) but for the neuronal pair shown in Figure 4B (the

second pair from the left). Left: real data (120 stimulus repetitions). Right: simulated data (10,000 stimulus repetitions). In all cases, simulations produce the same average number of coincident events as observed empirically.



**Supplemental Table S1: Summary of statistical results testing stimulus-dependent changes of firing sequences**

<b>Cat</b>	<b>RC</b>	<b>Probe</b>	<b>AS</b>	<b>MC</b>	<b>CR</b>	<b>T</b>	<b>UR</b>	<b>UA</b>	<b>OS</b>	<b>df (Add)</b>	<b>F(Add)</b>	<b>p(Add)</b>	<b>N(Tns)</b>	<b>C(Tns)</b>	<b>p(Tns)</b>	<b>df(RM)</b>	<b>F(RM)</b>	<b>p(RM)</b>	<b>df(UE)</b>	<b>F(UE)</b>
cer01	a42	A	2	30	2/5	80	20	7	1.5	6, 30	4.0	<b>0.004</b>	35	4	> 0.10	6, 14	3.3	<b>0.032</b>	7,16	1.1
cer01	a42	B	2	30	2/5	80	11	6	2.1	5, 20	1.3	0.320	20	0	<b>&lt;0.05</b>	5, 12	1.4	0.311	6,14	2.8
cer03	a55	A	6	<b>60</b>	4/8	110	12	12	<b>7.2</b>	11, 110	22.0	<b>0.000</b>	220	8	<b>&lt;0.001</b>	55, 72	25.3	<b>0.000</b>	8,90	3.2
cer03	a55	B	4	30	4/8	110	7	5	4.8	4, 12	1.8	0.736	10	1	> 0.10	12, 20	2.3	<b>0.049</b>	4,30	0.6
col05	e06	A	3	30	2/3	40	13	9	<b>6.0</b>	8, 56	17.6	<b>0.000</b>	84	4	<b>&lt;0.001</b>	16, 27	37.6	<b>0.000</b>	5,24	13.0
col05	e06	B	3	30	2/3	40	11	7	4.5	6, 30	11.7	<b>0.000</b>	35	4	> 0.10	12, 21	24.9	<b>0.000</b>	3,16	3.6
col08	e18	A	12	<b>90</b>	4/9	80	7	7	4.0	6,30	27.4	<b>0.000</b>	35	0	<b>&lt; 0.01</b>	66, 84	7.4	<b>0.000</b>	9,220	1.9
col08	e18	B	5	<b>90</b>	4/9	80	15	9	1.9	8,56	16.3	<b>0.000</b>	84	1	<b>&lt;0.001</b>	32, 45	3.3	<b>0.000</b>	7,64	9.1
col11	b50	A	4	<b>60</b>	6/16	120	18	11	<b>6.3</b>	10,90	12.1	<b>0.000</b>	165	10	<b>&lt;0.001</b>	30, 40	11.6	<b>0.000</b>	5,36	6.0
col11	b50	B	2	30	6/16	120	12	5	3.5	4, 12	2	0.157	10	1	>0.10	4, 10	2.2	0.148	4,10	1.8

**Supplemental Table S1.** Information on the significance tests and the basic properties for the 10 data sets of responses to changing grating orientations. **Cat**, code for each experiment made with an individual cat; **RC**, code for the recording session; **Probe**, the identity of the Michigan probe from which the data were analyzed; **AS**, The number of stimulus conditions that were analyzed (Maximum: 12); **MC**, Maximum change in the grating orientation across the analyzed stimulus conditions, expressed in degrees. As the orientation, not the movement direction, of the grating is considered, the maximum possible value is 90°; **CR**, number of concatenated blocks of recordings , and number of hours over which the recordings were spread; **T**, the number of stimulus presentations; **UR**, the total number of single units that could be isolated in the recording; **UA**, the number of units selected for the analysis; **OS**, mean oscillation score of the data set. Significance is reported for four statistical tests: additivity-ANOVA (*Add*), transitivity test (*Tns*), the classical repeated-measurement ANOVA (*RM*) and the ANOVA of unitary spike events (*UE*).  $p(Add)$  values are corrected for multiple comparisons (see Experimental Procedures). Significant  $p$ -values are printed in bold font. The significance level for the transitivity test was set to  $p = 0.01$  to control for multiple comparisons. **df**, degrees of freedom; **N**, the total number of triples in the network; **C**, the count of non-transitive triples. For  $UA = 5$  units ( $N = 10$  triples), due to the small sample size the maximum possible significance of the transitivity test is relatively low ( $p \sim 0.11$ ) even when all triples are transitive. Numbers printed in bold face indicate large changes in drift direction of the grating stimuli (MC), high oscillation score (OS, values  $> 5$ ), and statistically significant stimulus-dependent changes in firing sequences [categories  $p(Add)$ ,  $p(Tns)$  and  $p(RM)$ ].

## **Supplemental Discussion 1**

### **The origin of firing sequences - Sequential firing with parallel wiring**

To fully understand the mechanisms by which delays in spiking activity are generated, and hence, the principles by which neurons form and adjust the firing sequences, future studies will be necessary. The present evidence suggests that the sequential activation of connected cells (Abeles, 1991; Prut et al., 1998; Beggs and Plenz, 2003; Izhikevich, 2006)—i.e. each preceding cell activates the next one—plays only a partial role. First, the CCHs reported here do not have the shape characteristic of direct synaptic connections (Alonso and Martinez, 1998; Alonso et al., 2001; Schneider and Nikolić, 2006). Second, given the anatomy of the visual cortex, and the spatial layout of the recording probes that cover only a small cortical volume, it is unlikely that the five to twelve neurons included in a firing sequence would form a chain of direct synaptic activation. Third, our non-selective inclusion of neurons for the analyses is unlikely to bring the focus on the strongly connected ones. Hence, inevitably, most of the investigated neuron pairs will exhibit cross-correlations based on mechanisms other than chained activations.

Common input is the alternative. In the visual cortex, a special form of common input—that based on shared oscillatory rhythms in the gamma (and beta) range—has been consistently associated with the formation of centre peaks in the CCHs (Engel et al., 1991; Konig et al., 1995b, Roelfsema et al., 1997). Therefore, the question how delays emerge and change, begins with the question: Which constraints and flexibilities do gamma oscillations impose on spike timing? Recent progress in the stochastics of oscillations indicates an unexpected degree of flexibility, which allows neurons to make “decisions” about generating an action potential independently of the oscillatory rhythm—even when they are strongly engaged in that rhythm. Empirically observed CCHs can be reproduced

well by stochastic models in which the firing probability of all neurons is modulated by a common gamma rhythm, yet the actual firing pattern of each neuron is modulated independently, e.g. by local excitatory and inhibitory inputs (this is known also as non-stationary Poisson process and is illustrated in supplemental Fig. S11A-C). Such stochastic processes can generate oscillations (Nikolić et al., 2008), synchrony (Schneider and Nikolić, 2006; Markowitz et al., 2008), locking of spike signals to LFP (Zeitler et al., 2008) and also spiking delays and additivity (Schneider, 2008), matching those found in experimental data. To illustrate the applicability of these stochastic processes to our data, we used a similar model to reproduce the example CCHs shown in Figures 1B and 4B (supplemental Fig. S12). The good correspondence suggests that even strong oscillations and synchrony leave enough freedom for each neuron to conduct internal computations and eventually decide whether to fire within a cycle of oscillation and if so, when to fire within that cycle. Any network model (as opposed to a stochastic model) that attempts to explain these computations should not be studied in isolation but should integrate a number of other known facts about brain oscillation and synchrony (Nikolić, 2009).

An indication that the specific properties of spike delays (magnitude, direction etc.) are determined independently of oscillations—despite neurons obeying the rhythm—is our finding that, when oscillation frequency changes, the magnitude of the delay does not follow (Fig. 3C). As a consequence, a delay will be measured more consistently when expressed in milliseconds than in the phase of the oscillation.

The internal computations that determine a neuron's spike delay relative to the population may take into account the strength of input, more optimally stimulated cells firing earlier than the less optimally stimulated ones (Alonso and Martinez, 1998; Harris et al., 2002; O'Keefe and Burgess, 2005; Fries et al., 2007; Harvey et al., 2009). Note that this does not

necessarily imply a strong correlation between firing times and firing rates, as firing rates are also shaped by other processes like excitatory-inhibitory coupling (Dragoi et al., 2002; Kapfer et al., 2007; Attallah and Scanziani, 2009) and adaptation (Harris et al., 2002; Gutnisky and Dragoi, 2008). Currently, it is not known how these processes affect firing times. Another factor may be the cell's susceptibility for generating repolarizing currents. In cells with strong repolarizing currents, inhibitory inputs may advance the firing times, and excitatory inputs may delay them (Tsubo et al., 2007; Kwag and Paulsen, 2009a, 2009b). This is another potential explanation for the presently discovered low correlations between the firing rates and firing times.

Similar principles have been considered to govern the phase precession in the hippocampus. Here, each neuron adjusts its spike timing relative to a common reference signal (the theta oscillation), resulting in sequences of spikes that encode information about e.g. the location of the animal (O'Keefe and Recce, 1993; Skaggs et al., 1996; Harris et al., 2002; Dragoi and Buzsaki, 2006). A cautionary note is necessary: Although the present findings are reminiscent of the hippocampal phase precession, this does not necessarily mean that the two phenomena, or their underlying mechanisms, are the same. For example, in the hippocampus, the sequence of spiking within a cycle of theta oscillation mirrors the sequence in which the corresponding place fields were activated. In contrast, we found no relation between the sequence in which a bar activates visual receptive fields and the sequence in which the corresponding neurons fire (see Fig. 1 and supplemental Fig. S1B-C). Also, in the hippocampus, spike times progress unidirectionally (i.e. precess to ever earlier spike phases), whereas in the present data, the way in which spike times change over the time course of visual stimulation varies across

neurons (see Fig. 5). Further comparisons between the two phenomena are open for future studies.

With regard to the possible origin of firing sequences it is important to consider the temporal structure of the input to the primary visual cortex. Neural responses in the lateral geniculate nucleus (LGN) are locked remarkably well to the timing of stimuli of high contrast and fast temporal dynamics (up to  $< 1$  ms error; see Reinagel and Reid, 2000, 2002; Liu et al., 2001). However, this precision appears to be lost when stimuli with less driving force are processed (Reich et al., 1997; Butts et al., 2007; Lesica et al., 2007; Desbordes et al., 2008). Recent findings suggest that, in this case, spikes are nevertheless precisely timed internally (Desbordes et al., 2008), particularly relative to oscillations in the  $\sim 60$  Hz range (Koepsell et al., 2009). The precision of this internal timing has been reported to remain high across different types of stimuli (Desbordes et al., 2008). This precision seems to increase in the presence of oscillations (see Koepsell et al., 2009), and operates on a time scale similar to that observed in the present study (CCH peak widths of  $\sim 10$  ms, see Desbordes et al., 2008). Thus, the principle of internal spike timing in the visual cortex may also hold for subcortical structures such as the LGN. This opens up the possibility that during cortico-thalamic synchronization, internally referenced spike times play a role in the communication between LGN and V1 (Castelo-Branco et al., 1998; Neuenschwander et al., 2002; Briggs and Usrey, 2007). At present, little evidence is available regarding the existence of similar, internally timed, firing sequences in the retina (but see Meister et al., 1995; Pillow et al., 2008). Nevertheless, as the retina also generates high-frequency oscillations (Neuenschwander and Singer, 1996; Neuenschwander et al., 1999), this possibility remains open and should be investigated. Note that the firing

sequences presented here are unlikely to be a direct product of spike timing generated in the LGN. The reason is that none of our data contained the fast oscillations ( $>\sim 60$  Hz) typical for LGN. Instead, neurons were synchronized at slow gamma and beta frequencies of up to 40 Hz, a frequency thought to arise from intracortical synchronization (Castelo-Branco et al., 1998; Munk and Neuenschwander, 2000; Sillito and Jones, 2002).

## **Supplemental Discussion 2**

### **The readout of firing sequences - Sensing input timing**

Variations in spike timing can only be exploited for coding if downstream neurons are capable of distinguishing the respective changes in the temporal order of synaptic inputs. Several experimental studies have indicated that neuronal networks are sensitive to small variations in input timing. In the auditory system, small delays between different inputs are used for sound localization, where the time differences are detectable in the sub-millisecond range (Jeffress, 1948; Grothe, 2003). In the retina, the sequence of activation in arrays of converging inputs along a dendrite is responsible for the generation of direction selectivity in starburst amacrine cells (Hausselt et al., 2007). Cortical neurons are also sensitive to the precise temporal structure of inputs, responding preferentially to inputs that arrive within a short time window ( $\sim 15$  ms) (Mainen and Sejnowski, 1995; Usrey et al., 2000, Azouz and Gray, 2000, 2003). However, this does not mean that perfect zero-lag synchronization of inputs is always optimal. Instead, a specific temporal order of inputs may be necessary to counteract differing axonal and dendritic conduction delays, which then achieves maximum synchronization of inputs at the point of their integration—e.g., at the cell's soma (Rall, 1962; London and Hausser, 2005; Hausselt et al., 2007; Stiefel and Sejnowski, 2007; Branco et al., 2010). Thus, the summation of delayed (or asynchronous) inputs has been shown to be more effective when the sequential activation proceeds from the dendritic periphery towards the soma rather than the other way around (Rall, 1962; London and Hausser, 2005; Branco et al., 2010). Moreover, a neuron may implement short integration constants in dendrites (coincidence detection) and long integration constants at the soma (accumulation illustrated in supplemental Fig. S11F) (Williams and Atkinson, 2008). In addition, the strength and timing of post-synaptic



responses is likely to be modulated by other time-sensitive mechanisms such as fast feed-forward inhibition (Delorme and Thorpe, 2001; Van Rullen et al., 2005; Hirsch et al., 2008) and the phase relation between pre- and postsynaptic neurons (Tsubo et al. 2007; Fries et al., 2007; Cardin et al., 2009). Even faster decoding, i.e. requiring much shorter integration constants at the soma, may be implemented when the number of presynaptic neurons is considerably larger than the number of neurons we sampled in parallel in the present study (supplemental Fig. S11D, bottom; supplemental Fig. S11G). Our result that the extraction of firing sequences became more reliable when based on multi-unit activity suggests that such a decoding would benefit from the inputs of neurons with similar orientation preferences that arrive at similar locations along the dendrite (e.g., inputs from one orientation being distal and from others more proximal to the soma). Thus, the presently observed small variations in the relative timing and changes in the order of firing are likely to have functionally relevant effects on the responses of postsynaptic neurons in the cortex.

### **Supplemental References**

Abeles M (1991) *Corticonics : neural circuits of the cerebral cortex*. Cambridge ; New York: Cambridge University Press.

Alonso JM, Usrey WM, Reid RC (2001) Rules of connectivity between geniculate cells and simple cells in cat primary visual cortex. *J Neurosci* 21:4002-4015.

Atallah BV, Scanziani M (2009) Instantaneous modulation of gamma oscillation frequency by balancing excitation with inhibition. *Neuron* 62:566-577.

Azouz R, Gray CM (2000) Dynamic spike threshold reveals a mechanism for synaptic coincidence detection in cortical neurons in vivo. *Proc Natl Acad Sci U S A* 97:8110-8115.

Beggs JM, Plenz D (2003) Neuronal avalanches in neocortical circuits. *J Neurosci* 23:11167-11177.

Berenyi A, Benedek G, Nagy A (2007) Double sliding-window technique: a new method to calculate the neuronal response onset latency. *Brain Res* 1178:141-148.

Briggs F, Usrey WM (2007) Cortical activity influences geniculocortical spike efficacy in the macaque monkey. *Front Integr Neurosci* 1:3.

Brody CD (1999) Correlations without synchrony. *Neural Comput* 11:1537-1551.

Butts DA, Weng C, Jin J, Yeh CI, Lesica NA, Alonso JM, Stanley GB (2007) Temporal precision in the neural code and the timescales of natural vision. *Nature* 449:92-95.

Castelo-Branco M, Neuenschwander S, Singer W (1998) Synchronization of visual responses between the cortex, lateral geniculate nucleus, and retina in the anesthetized cat. *J Neurosci* 18:6395-6410.

Delorme A, Thorpe SJ (2001) Face identification using one spike per neuron: resistance to image degradations. *Neural Netw* 14:795-803.

Dragoi V, Sharma J, Miller EK, Sur M (2002) Dynamics of neuronal sensitivity in visual cortex and local feature discrimination. *Nat Neurosci* 5:883-891.

Dunn OJ (1961) Multiple Comparisons among Means. *Journal of the American Statistical Association* 56:52-&.

Grothe B (2003) New roles for synaptic inhibition in sound localization. *Nat Rev Neurosci* 4:540-550.

Gutnisky DA, Dragoi V (2008) Adaptive coding of visual information in neural populations. *Nature* 452:220-224.

Hirsch JA, Martinez LM, Pillai C, Alonso JM, Wang Q, Sommer FT (2003) Functionally

distinct inhibitory neurons at the first stage of visual cortical processing. *Nat Neurosci* 6:1300-1308.

Izhikevich EM (2006) Polychronization: computation with spikes. *Neural Comput* 18:245-282.

Jeffress LA (1948) A place theory of sound localization. *J Comp Physiol Psychol* 41:35-39.

Kapfer C, Glickfeld LL, Atallah BV, Scanziani M (2007) Supralinear increase of recurrent inhibition during sparse activity in the somatosensory cortex. *Nat Neurosci* 10:743-753.

Kwag J, Paulsen O (2009) The timing of external input controls the sign of plasticity at local synapses. *Nat Neurosci* 12:1219-1221.

Kwag J, Paulsen O (2009) Bidirectional control of spike timing by GABA(A) receptor-mediated inhibition during theta oscillation in CA1 pyramidal neurons. *Neuroreport* 20:1209-1213.

Lesica NA, Jin J, Weng C, Yeh CI, Butts DA, Stanley GB, Alonso JM (2007) Adaptation to stimulus contrast and correlations during natural visual stimulation. *Neuron* 55:479-491.

Liu RC, Tzonev S, Rebrik S, Miller KD (2001) Variability and information in a neural code of the cat lateral geniculate nucleus. *J Neurophysiol* 86:2789-2806.

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

Markowitz DA, Collman F, Brody CD, Hopfield JJ, Tank DW (2008) Rate-specific synchrony: using noisy oscillations to detect equally active neurons. *Proc Natl Acad Sci U S A* 105:8422-8427.

Meister M, Lagnado L, Baylor DA (1995) Concerted signaling by retinal ganglion cells. *Science* 270:1207-1210.

Mitchell JF, Sundberg KA, Reynolds JH (2007) Differential attention-dependent response modulation across cell classes in macaque visual area V4. *Neuron* 55:131-141.

Movshon JA, Thompson ID, Tolhurst DJ (1978) Spatial summation in the receptive fields of simple cells in the cat's striate cortex. *J Physiol* 283:53-77.

Movshon JA, Thompson ID, Tolhurst DJ (1978) Spatial and temporal contrast sensitivity of neurones in areas 17 and 18 of the cat's visual cortex. *J Physiol* 283:101-120.

Munk MH, Neuenschwander S (2000) High-frequency oscillations (20 to 120 Hz) and their role in visual processing. *J Clin Neurophysiol* 17:341-360.

Neuenschwander S, Castelo-Branco M, Baron J, Singer W (2002) Feed-forward

synchronization: propagation of temporal patterns along the retinothalamocortical pathway. *Philos Trans R Soc Lond B Biol Sci* 357:1869-1876.

Neuenschwander S, Castelo-Branco M, Singer W (1999) Synchronous oscillations in the cat retina. *Vision Res* 39:2485-2497.

Neuenschwander S, Singer W (1996) Long-range synchronization of oscillatory light responses in the cat retina and lateral geniculate nucleus. *Nature* 379:728-732.

Nikolić D, Moca VV, Singer W, Muresan RC (2008) Properties of multivariate data investigated by fractal dimensionality. *J Neurosci Methods* 172:27-33.

O'Keefe J, Burgess N (2005) Dual phase and rate coding in hippocampal place cells: theoretical significance and relationship to entorhinal grid cells. *Hippocampus* 15:853-866.

Pillow JW, Shlens J, Paninski L, Sher A, Litke AM, Chichilnisky EJ, Simoncelli EP (2008) Spatio-temporal correlations and visual signalling in a complete neuronal population. *Nature* 454:995-999.

Prut Y, Vaadia E, Bergman H, Haalman I, Slovin H, Abeles M (1998) Spatiotemporal structure of cortical activity: properties and behavioral relevance. *J Neurophysiol* 79:2857-2874.

Rall W (1962) Theory of physiological properties of dendrites. *Ann N Y Acad Sci* 96:1071-1092.

Reich DS, Victor JD, Knight BW, Ozaki T, Kaplan E (1997) Response variability and timing precision of neuronal spike trains in vivo. *J Neurophysiol* 77:2836-2841.

Reinagel P, Reid RC (2000) Temporal coding of visual information in the thalamus. *J Neurosci* 20:5392-5400.

Reinagel P, Reid RC (2002) Precise firing events are conserved across neurons. *J Neurosci* 22:6837-6841.

Schneider G (2008) Messages of oscillatory correlograms: a spike train model. *Neural Comput* 20:1211-1238.

Sillito AM, Jones HE (2002) Corticothalamic interactions in the transfer of visual information. *Philos Trans R Soc Lond B Biol Sci* 357:1739-1752.

Stiefel KM, Sejnowski TJ (2007) Mapping function onto neuronal morphology. *J Neurophysiol* 98:513-526.

Williams SR, Atkinson SE (2008) Dendritic synaptic integration in central neurons. *Curr Biol* 18:R1045-1047.

Zeitler M, Fries P, Gielen S (2008) Biased competition through variations in amplitude of gamma-oscillations. *J Comput Neurosci* 25:89-107.