Review

Conditional modeling and the jitter method of spike resampling

Asohan Amarasingham,[1]* Matthew T. Harrison,[2]* Nicholas G. Hatsopoulos,[3] and Stuart Geman[2]

[1]*Department of Mathematics, The City College of New York, and Program in Cognitive Neuroscience, The Graduate Center, City University of New York, New York, New York;* [2]*Division of Applied Mathematics, Brown University, Providence, Rhode Island; and* [3]*Department of Organismal Biology and Anatomy, Committees on Computational Neuroscience and Neurobiology, University of Chicago, Chicago, Illinois*

Amarasingham A, Harrison MT, Hatsopoulos NG, Geman S. Conditional modeling and the jitter method of spike resampling. *J Neurophysiol* 107: 517–531, 2012. First published October 26, 2011; doi:10.1152/jn.00633.2011.—The existence and role of fine-temporal structure in the spiking activity of central neurons is the subject of an enduring debate among physiologists. To a large extent, the problem is a statistical one: what inferences can be drawn from neurons monitored in the absence of full control over their presynaptic environments? In principle, properly crafted resampling methods can still produce statistically correct hypothesis tests. We focus on the approach to resampling known as jitter. We review a wide range of jitter techniques, illustrated by both simulation experiments and selected analyses of spike data from motor cortical neurons. We rely on an intuitive and rigorous statistical framework known as conditional modeling to reveal otherwise hidden assumptions and to support precise conclusions. Among other applications, we review statistical tests for exploring any proposed limit on the rate of change of spiking probabilities, exact tests for the significance of repeated fine-temporal patterns of spikes, and the construction of acceptance bands for testing any purported relationship between sensory or motor variables and synchrony or other fine-temporal events.

spike timing; neural coding; jitter; hypothesis testing; conditional inference; synchrony

1 Introduction

What is the temporal resolution of neuronal activity? What role does fine-temporal structure play in neuronal processing? Whereas millisecond resolution has been demonstrated in a multitude of special systems—the retina of several vertebrate species (Meister and Barry 1999), the mammalian medial superior olivary nucleus (Yin et al. 1990), the honeybee olfactory system (Wehr and Laurent 1996), the bat sonar system (Ferragamo et al. 1998), and the cat lateral geniculate nucleus (Alonso et al. 1996; Dan et al. 1998), to name a few—the significance of these observations to cortical processing in general is controversial (Gray 1999; Riesenhuber and Poggio 1999; Shadlen and Movshon 1999). Are they the exception or the rule? The advent of multielectrode technologies and the feasibility of stable simultaneous recordings from multiple neurons in awake and performing animals now offer hope for meaningfully addressing these fundamental issues.

Many authors have argued that the useful information transmitted from one neuron to another, in the general cortical circuitry, is well described by the number of spikes, as measured over a 10- or 20-ms interval, rather than the timing of the spikes, as might be measured at a 1- or 2-ms resolution. The issue is unresolved. On one hand, we are able to associate firing rates of select cortical neurons (as measured on a 10 or greater millisecond scale) with numerous perceptual and motor states, such as a face in the visual field or the direction of limb movement. On the other hand, there are indeed a multitude of systems, albeit highly specialized, which demonstrate (and almost certainly require) millisecond accuracy in spike timing. What is more, there is the often-repeated argument that neural representation must accommodate the relational structure among the parts of an action, a thought, a perceived object, or a scene, and that fine-temporal structure might offer a natural way to capture this kind of information (Abeles 1991; Bienenstock 1995; Geman 2006; Roelfsema et al. 1996; Singer 1999; von der Malsburg 1994).

The analysis of neuronal data in an awake animal is made especially challenging by inevitable changes, from instant to instant and trial to trial, of the statistical nature of the recorded signal (Amarasingham et al. 2006; Brody 1998; Churchland et al. 2010; Cohen and Kohn 2011). The status of tens of thousands of presynaptic neurons, not directly influenced by the experimental paradigm, is largely out of the control of the experimenter; the environments of individual neurons are in constant flux. Thus time-honored assumptions about "repeated" samples (having similar or identical distributions) become untenable in cortical recordings from awake and per-

---

*A. Amarasingham and M. T. Harrison contributed equally to this work.

Address for reprint requests and other correspondence: S. Geman, Division of Applied Mathematics, Brown Univ., 182 George St., Providence, RI 02912 (e-mail: stuart_geman@brown.edu).

forming animals. The unique character of variability in a neurophysiological data set raises, in statistical terms, concerns of robustness: the conclusion of a conventional statistical analysis of a spike train may be contaminated by the fact that trial-to-trial variability causes its underlying probabilistic model to be misspecified.

These observations highlight a fundamental challenge to gathering statistical evidence for a lack of precision in the cortical microcircuitry (Amarasingham et al. 2006). Imprecision could arise from nonrepeatability in the dynamics of individual neurons or, equally well, from a failure to reproduce a neuron's trial-to-trial input. On the other hand, these observations do not preclude collecting statistical evidence against any given limitation on precision or repeatability.

We will review a collection of statistical approaches to temporal resolution collectively known as jitter methods. We will formulate a definition of temporal resolution and review the application of jitter methods for testing the hypothesis that neurons operate at or below any specified resolution. These procedures are based on observations of precisely timed events in the spike record, such as counts of synchronous events across neurons or of more general spiking motifs, but their statistical significance is computed without estimating instant-to-instant firing rates or trial-to-trial variability of the participating neurons.

One way to motivate jitter is through a thought experiment, suggested by our colleague Prof. Elie Bienenstock of Brown University. Bienenstock imagines a pharmacological agent with the peculiar property that it modestly advances or retards every spike of every neuron, independently, for the duration of the drug's action. How large would the "jitter window" have to be to disrupt cortical processing? Would one feel the effect, or, indeed, would the brain even function, under a 5-ms random change in the timing of every spike? How about 10 ms? In addition to taking a step toward a definition of temporal resolution, the thought experiment suggests a statistical approach to testing any given hypothesized limit on resolution. In particular, beginning with an original data set of simultaneously recorded neurons, one could create a surrogate data set by randomly jittering each spike within a 10-ms window. How do statistics computed from the original data set compare with statistical features computed from an ensemble of such surrogate data sets, created by repeatedly jittering? This is the beginning of a statistical definition of temporal resolution and of a statistical test. Complications introduced by trial-to-trial variability are avoided because the individual trial structure is preserved; indeed, the approach applies just as sensibly to single-trial data and to cases where there is no clear notion of a trial at all.

As it stands, this type of test is crude in that it does not accommodate even slow variations in an underlying firing rate, much less bursting or absolute and relative refractory periods. We will formulate jitter more rigorously as an example of conditional modeling and exploit this framework to develop a suite of tools for measuring and testing temporal resolution, despite the time-varying nature of neuronal excitability, the nonrepeatability of trials, and the various phenomena, such as bursting and refractoriness, that statistically tie together individual spikes.

In various guises, jitter and related spike resampling methods are in common use (Abeles and Gat 2001; Amarasingham 2004; Baker and Lemon 2000; Butts et al. 2007; Date et al. 1999; Dragoi and Buzsáki 2006; Fujisawa et al. 2008; Furukawa and Middlebrooks 2002; Gerstein 2004; Grün 2009; Grün et al. 2009; Gütig et al. 2002; Harrison 2005; Harrison and Geman 2009; Hatsopoulos et al. 2003; Ikegaya et al. 2004; Jones et al. 2004; Lu and Wang 2003; Maldonado et al. 2009; Nádasdy et al. 1999; Oram et al. 1999; Pazienti et al. 2007, 2008; Pipa et al. 2007; Renart et al. 2010; Rieke et al. 1997; Rokem et al. 2006; Shmiel et al. 2005, 2006; Smith and Kohn 2008; Stark and Abeles 2009; see also Amarasingham et al. 2011).[1] The merits and limitations of these techniques, as well as their proper application and interpretation, are best viewed through rigorous statistical formulation. We provide an elementary overview of jitter-based resampling and a statistical framework that supports a precise statement of assumptions and a precise interpretation of conclusions. We provide a step-by-step development of jitter-based methods, for investigating temporal resolution, "excess" synchrony, and other measures of precision, and we will explore related formulas for making exact hypothesis tests about fine-temporal structure. Both simulated and real spike data are used for illustration. The discussion is largely nontechnical but includes frequent pointers to the mathematical appendix provided in a parallel technical report (Amarasingham et al. 2011), which also contains the details of the analyses and computer experiments used in the main text. Together, the text and appendix provide an essentially self-contained and complete formal statistical treatment.

In the following section (Section 2), taking the analysis of synchrony and the cross-correlogram as a pedagogical example, we introduce the basic idea of spike jitter and its utility for exploring temporal resolution. We make comparisons with trial shuffling and provide a framework, conditional modeling, for articulating and evaluating the statistical assumptions under which jitter and other resampling methods are valid. In Section 3, we demonstrate the scope of the conditional modeling framework by reviewing broader classes of jitter techniques, such as accommodate slow variations in underlying firing rate and structured firing patterns. Drawing from this development and in the way of further illustration, in Section 4 we analyze the evidence for fine-temporal structure in the recordings of three simultaneously recorded primate motor cortical neurons. We conclude in Section 5 with further discussion and summary.

2 Spike Re-Sampling, Spike Jitter, And Conditional Inference

2.1 Trial shuffles and interval jitter. Imagine making a 1-s recording from each of two neurons for each of 100 trials. Possibly, the two neurons have identical inhomogeneous (i.e.,

---

[1] Some authors prefer alternative terms for jittering, such as "dithering" (Gerstein 2004; Grün 2009; Pazienti et al. 2007, Pazienti et al. 2008), "teetering" (Shmiel et al. 2005), or "artificial jitter" (Rokem et al. 2006), etc., presumably to distinguish from another use of the word "jitter" as the intrinsic temporal variability in individual spikes. For example, in some situations involving highly reliable (Billimoria et al. 2006) or simulated or modeled spike trains (Pazienti et al. 2007), individual spikes can unambiguously be placed in correspondence with one another across trials. In that case, the temporal variability in a spike's timing, under this correspondence, can be quantified directly and is commonly called the "spike jitter" (Billimoria et al. 2006; Victor and Purpura 1996). In this article, we continue to use "jitter" in its resampling sense, leaving these two uses of the term to be disambiguated by context.

time varying) firing rates, and possibly, the firing rate profile changes from trial to trial. In such a circumstance, even slow variations in firing rate will induce a peak in the cross-correlation histogram at zero, and the peak will appear to be significant when measured, for example, against a corresponding cross-correlation histogram value produced by a "shuffle" (permutation) of trials of one or the other of the neurons. These effects have been widely discussed (Baker and Gerstein 2001; Brody 1998; Ventura et al. 2005b). A peak at and around zero in a shuffle-corrected cross-correlation histogram is not, in and of itself, an artifact; common variation in firing rates across two neurons induces a statistical correlation in spike timings. The issue is one of interpretation and, in particular, one of time scale. Statistically, these coupled but nonetheless slow temporal effects violate the assumption of independence, and violate it in the direction of increased synchrony, so significance with respect to the shuffling procedure is not surprising. It is not, however, due to finely timed spikes, per se. When the goal is to isolate effects at a particular time scale, independence across neurons is the wrong null hypothesis.

Is there a way to test for millisecond precision, per se, despite the evident correlation induced by shared trial-to-trial variation in firing rate? It is tempting to try to first correct for trial-to-trial variation and then to look for residual high-reso-

lution correlation, and in fact many methods along these lines have been proposed. As already mentioned in Section 1, we have some doubts about the feasibility of estimating parameters, much less rate functions, for individual trials. Here we focus instead on a jitter-based approach that avoids direct estimation of firing rates, whether common to trials or corrected for trial-to-trial variation.

A simple computer experiment illustrates the idea. An artificial data set has two simultaneously recorded neurons with a zero-lag peak in the shuffle-corrected cross-correlation histogram (Fig. 1). By construction, the peak comes from shared trial-to-trial variability in the Poisson intensity functions operating on relatively slow (50 ms) time scales (Fig. 1A), as evidenced in the cross-correlogram (Fig. 1C), and not from any additional correlations in spike timing. The peak is significant with respect to the shuffling procedure (Fig. 1D), but how are we to know and quantitatively report that it is the result of the relatively slow, 50-ms undulations in firing rate that are common between the two neurons in a given trial, but different from trial to trial? Suppose that instead of shuffling trials we were to generate surrogate spike trains by perturbing individual spikes, independently, within a small window, say 20 ms, mimicking the action of the fictional drug described in Section 1. Does the apparent synchrony survive? If so, then we could
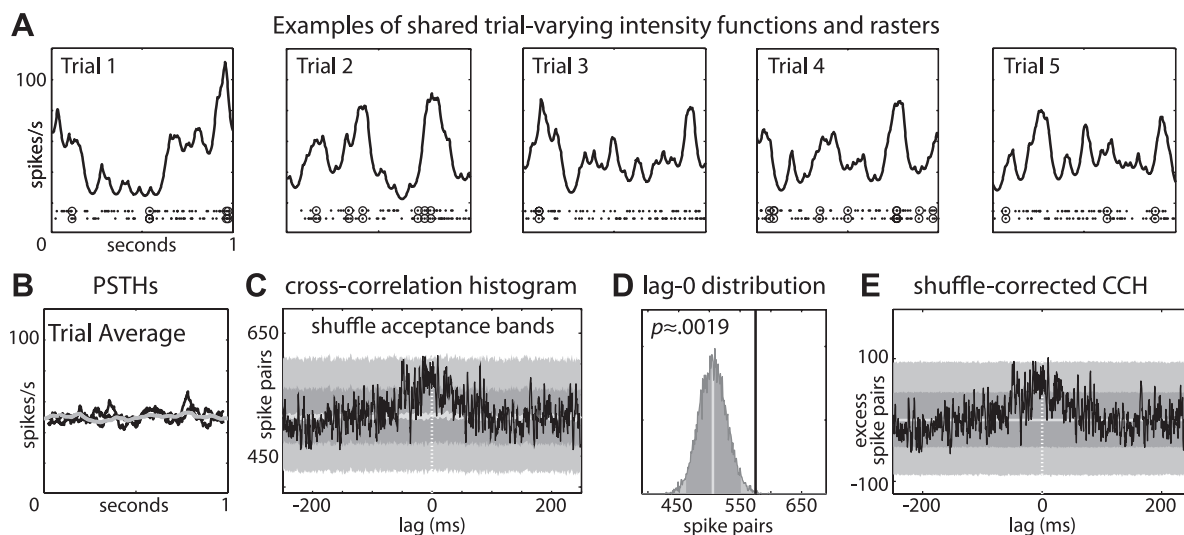

Fig. 1. Two identical, nonrepeating inhomogeneous Poisson "neurons." *A*: 5 of 100 randomly generated Poisson intensity functions with 50-ms-wide peaks. The intensity function for the *j*th trial is $f_j(t) = 10 + \Sigma_k g(t, \mu_{jk})$, where $k = 1, \ldots, 40$, the μ_{jk} values are independent, identically distributed uniformly on (0, 1), and $g(t, \mu)$ is a double exponential centered at μ with a standard deviation (bandwidth) of 50 ms (before wrapping) that wraps around (0, 1) so that it integrates to 1 over the unit interval. Thus the associated inhomogeneous Poisson process has a mean "firing rate" of exactly 50 Hz. (See Amarasingham et al. 2011.) In each trial, each of 2 "neurons" is an inhomogeneous Poisson process with identical intensity functions. Different intensity functions are used for different trials. Below the intensity functions are raster plots showing the times of each spike for each of the 2 neurons. Circles mark spikes, which occur within 1 ms of a spike in the other spike train. *B*: average of the 100 intensity functions across the 100 trials (gray line) and peristimulus time histogram (PSTH) for each spike train (2 black lines). Individual trial structure is lost in the PSTHs, which are approximately flat. *C*: cross-correlation histogram (CCH; black line) together with shuffle-derived 95% pointwise and simultaneous acceptance bands. The value of the CCH at lag *t* ms is the total number of spike pairs (across all trials) from the 2 neurons for which the 2nd neuron's spike occurs within $t \pm 1$ ms of the spike from the 1st neuron on the same trial. In particular, the value of the CCH at *lag 0* (dotted white line) corresponds to the number of observed ±1-ms synchronies (575, in this case). The shuffle acceptance bands are computed as follows. Trials for *neuron 1* are randomly permuted 10,000 times, generating 10,000 "surrogate" data sets. The CCH is computed for each surrogate, as well as for the original (observed) data, to create a large collection of CCHs. Dark gray "pointwise" acceptance bands are constructed so that for each lag, 95% of the CCHs in the collection are within the bands at that lag. Medium gray "simultaneous" acceptance bands are constructed so that 95% of the CCHs in that collection are completely within the bands for all lags and accommodate the statistical error that is introduced by the examination of multiple time lags within a single data set. (Although these acceptance bands correspond to acceptance regions for well-defined hypothesis tests, they are not, as is commonly mistaken, confidence intervals.) See Amarasingham et al. (2011) for these and additional details. The thick, light gray line in the center of the bands shows the ensemble mean of the 10,000 surrogate CCHs. *D*: histogram of surrogate CCH values at *lag 0* (±1-ms synchronies) with the mean (light gray vertical line), original CCH at *lag 0* (black vertical line), and acceptance bands (shaded as described above). The (1 sided) *P* value is the fraction of the histogram greater than or equal to the observed value. *E*: shuffle-corrected CCH together with acceptance bands. At each time lag, the original CCH and the acceptance bands are "corrected" by subtracting the ensemble mean of the surrogate CCHs. The correction itself has no statistical purpose but aids visualization, because the corrected CCH should fluctuate around zero (solid light gray line) for all lags if the null hypothesis is true.

reasonably interpret the peak in the cross-correlation histogram as arising from temporal effects acting on a time scale no finer than about 20 ms. In particular, we would conclude that the apparently significant zero-lag peak in the cross-correlation histogram is consistent with events operating on a slower time scale (in this case, the 50-ms-wide peaks in the shared intensity functions).

The perturbation experiment was performed and is illustrated in Fig. 2. The time axis was first partitioned, independently of the observed spikes, into disjoint and contiguous intervals of length 20 ms. Each spike of each neuron was then independently moved to a new location, chosen from the uniform distribution on the 20-ms interval to which it belonged in the original data. Repeating the process produces a sequence of surrogate spike trains (Fig. 2*A*). Application to the synthetic data from Fig. 1 produces a sequence of surrogate data sets, each consisting of 100 trials of a 1-s spike train for each of the two neurons. The number of synchronous spike pairs ("synchronies") across the two neurons, as defined by occurring within +1 or −1 ms of each other, was recorded for each of 10,000 surrogate data sets. Figure 2*B* shows the cross-correlation histogram with jitter-derived acceptance bands (compare to Fig. 1*C*), and Fig. 2*C* shows the distribution of synchronies (compare to Fig. 1*D*). More than 21% of the randomly generated synchrony counts exceed the unperturbed count: the interval jitter procedure suggests that the peak in the cross-correlation histogram in Fig. 1 is due to effects occurring on a time scale coarser than 20 ms.

2.2 Hypothesis testing and conditional inference. Interval jitter (Amarasingham 2004; Date et al. 1999; Harrison 2005; Harrison and Geman 2009) tests hypotheses about local (within interval) distributions on the placements of spikes. Given that there is a spike in a particular 10- or 20-ms interval, are all placements of the spike essentially equally probable? Temporal resolution is about precision in the placement of spikes. Questions about the probability of having a spike in a given interval, or having any given number of spikes in a given interval, are interesting in their own right, but are different and off the point. The focus is on timing and to the extent possible we would prefer a statistical approach that avoids assumptions about anything but the issue at hand.

Interval jitter is an example of conditional modeling. Statistical models are specified via conditional distributions on the data. Inference is conditioned on a temporal coarsening of the spiking processes, as specified by the numbers of spikes in the jitter intervals. No assumptions are made about the process, random or otherwise, that determines the existence and numbers of spikes found in the jitter intervals. The assumptions that are made, and therefore tested, address the question being asked: how precise is the timing of spikes? By conditioning on the numbers of spikes in each interval, we sidestep entirely the problem of specifying the process through which these numbers arise. If the intervals are, say, 10 ms wide, and if we were to assume that all arrangements in a given interval that preserve the observed number of spikes are equally likely, then, informally, we are testing the hypothesis that there is no temporal structure at a scale finer than 10 ms. Rejecting the null would provide evidence for faster inter- and intraneuronal mechanisms. More formally, we would be rejecting a compound null hypothesis that includes every point process for which the conditional placements of spikes within intervals are uniform,
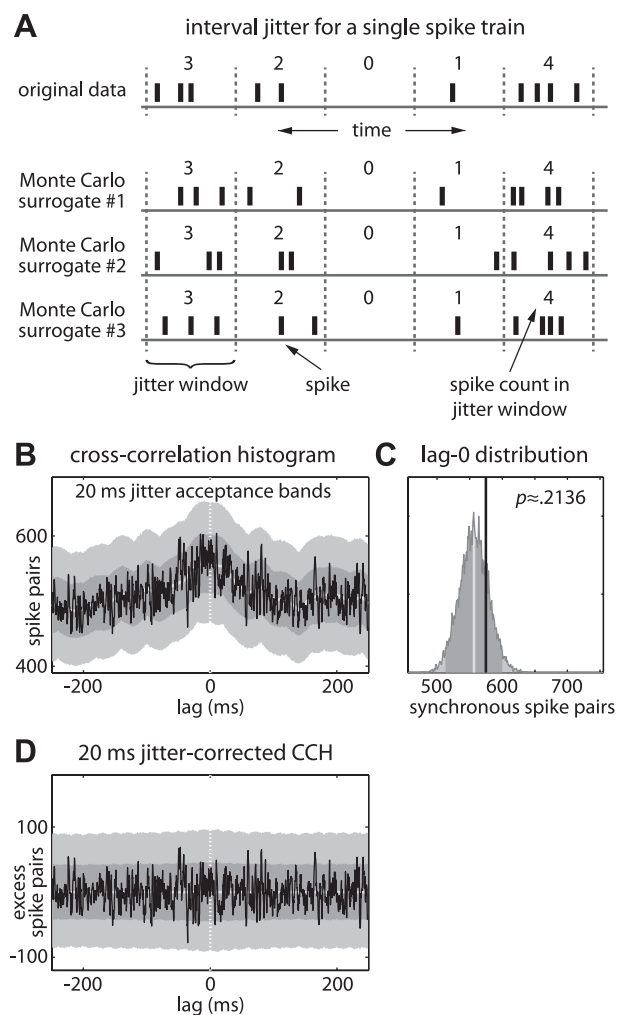

Fig. 2. Interval jitter. *A*: schematic illustration of Monte Carlo interval jitter for a single spike train. *Top* row shows a spike train with 10 spikes (bars). Time is partitioned into jitter windows (marked by dotted lines). The number of spikes in each jitter window is recorded (numbers above the spikes). A single Monte Carlo surrogate spike train (*surrogate 1*) is created by independently and uniformly choosing new spike times that preserve the spike counts in each jitter window. This is repeated independently for additional Monte Carlo surrogates (*surrogates 2* and *3*). For multiple spike trains, the interval jitter procedure is applied independently to each spike train. *B–D*: the interval jitter procedure with 20-ms jitter windows was applied to the original data set (100 trials and 2 spike trains) described in Fig. 1. This was repeated 10,000 times, creating 10,000 surrogate data sets. These jitter-based surrogate data sets were used exactly as the shuffle-based surrogate data sets in Fig. 1, *C–E*: to create jitter-derived (instead of shuffle-derived) acceptance bands and *P* values for the CCH and to "correct" the visual display of the CCH and acceptance bands so that the CCH has a mean of zero under the null hypothesis tested by 20-ms interval jitter. The CCH in *B* is identical to that in Fig. 1*C*, but the acceptance bands are derived from jittering, not shuffling. In contrast to the shuffle-derived tests in Fig. 1, the number of *lag 0* synchronies observed in the original, unperturbed spike trains (575, in this case) is not significant when measured against the jitter distribution (right-tail $P \approx 0.2136$). We would expect the jitter experiment to disrupt temporal processes operating at a time scale finer than 20 ms, in which case the number of synchronies in the unperturbed data would be large relative to the population of 10,000 synchrony counts produced by random perturbations. On the other hand, effects operating at slower time scales, as is the case in these experiments, are relatively well preserved; the number of synchronies in the unperturbed data is representative of the randomly generated population. The peak around zero in the CCH of Fig. 1*C* (and here in *B*) cannot be interpreted as evidence for fine-temporal synchrony, since slow and common fluctuations also produce similar zero-lag counts.

given the interval counts. And indeed, the proportion of surrogate spike trains for which synchrony counts are greater than that of the original train essentially provides an exact *P* value for this null hypothesis. (See Amarasingham et al. 2011 for a precise formal description of the test and a demonstration of its validity.)

At the risk of belaboring the point, there is another ("generative") way to understand the null hypothesis: imagine that you are in charge of generating the spike data that are recorded in an experiment. You first partition time into intervals (say 10 ms). You then assign spike counts to every interval of every neuron in every trial. These assignments can follow any rule whatsoever, deterministic or stochastic, with arbitrary dependencies between intervals and across neurons and trials, or no dependencies at all, or anything in between. Hidden variables, like the states of 10 billion unobserved neurons, can be determining or irrelevant to your assignments, as you wish. You then assign the actual locations of the spikes within intervals according to a distribution that includes the hypothesized limits on the precision of spike timing. (Here this limit is that all intrainterval locations are equally likely; in Sections 3.1 and 3.2, we describe generalizations of such limitations. The statistical formulation, via conditioning, remains the same.)

Trial shuffling is another, familiar, example of conditional inference. Given two neurons observed over *N* trials, inference is conditioned on the two sets of spike-time recordings, each set with *N* records representing the respective observations from the two neurons. The assumption, which again defines a compound null hypothesis, is on the conditional distribution: every one of the *N!* pairings of records is equally likely (as would be the case, for example, if the trials were statistically repeatable and the two neurons' responses statistically independent). Notice that the null hypothesis says nothing about temporal resolution. There is nothing wrong with using a tally of 1-ms synchronies to test the null hypothesis, so long as we keep in mind that an excess of synchronies in the original pairing, relative to the trial-shuffled pairings, rejects the hypothesis that all pairings are equally likely, not a hypothesized lack of precision of spike timing.

2.3 Synchrony as a test statistic. Jitter methods compare an original data set with an artificially generated ensemble of jittered data sets, but this comparison must be based on some feature of the data, such as the synchrony count, the shape of the cross-correlation histogram (Section 2.7), the number of repeating spiking motifs (Section 4.1), or an estimate of the mutual information between spiking and stimulus (see also Section 4.2). In the terminology of hypothesis testing, this feature of the data is called the test statistic. Jitter methods can be applied with any test statistic. Our emphasis here in Section 2 is jitter, not the test statistic. We use synchrony count as the test statistic primarily for pedagogical purposes. In Section 4, we illustrate jitter with other test statistics.

2.4 Basic jitter. In the experiment of Section 2.1, the intervals across which spikes were (uniformly) perturbed were chosen independently of the neurons' observed spike times. Here, by way of motivation, we compare this former procedure with the intuitive and common approach of simply jittering individual spikes using jitter intervals that are centered at the original spike locations. We refer to this familiar, heuristic procedure as "basic jitter": to form surrogate spike trains, simply jitter the spikes in an interval centered around their

original locations. Whereas the "*P* values" that result from such perturbations are indeed (one sided) tail probabilities under the jitter distribution on synchrony counts, they are not, strictly speaking, rejection probabilities in a well-formulated hypothesis test. No "null hypothesis" has been articulated, and in point of fact, none could be. The implicit assumption, when comparing observed synchrony counts with a distribution of counts derived from a spike resampling algorithm, is that the original locations of spikes are "statistically indistinguishable" from the resampled locations, provided that the null hypothesis is true. Statistically indistinguishable is called "exchangeable." A formal definition is included in the appendix of Amarasingham et al. (2011), but the idea is simple enough: an observer who believed in the null hypothesis could not pick out the real spike train from a lineup that also included the simulated spike trains. The trouble with basic jitter is that the original spikes stand out; they can always be found at or near the center of their resampled surrogates.

In typical practice and for simple test statistics, like synchrony counts, there may not always be a big difference between basic jitter and interval jitter. Indeed, the two procedures are quite similar, and, unsurprisingly, the results of the experiments in Sections 2.1 are qualitatively unchanged by using basic jitter, rather than interval jitter, with 20-ms jitter windows. However, an important distinction is that, as described in Section 2.2, interval jitter does make for an exact hypothesis test of a large class of null hypotheses: those spike-generating processes for which the locations of spikes, conditioned on the interval spike counts (i.e., the numbers of spikes found in each interval), are all equally likely. (See further discussion in Amarasingham et al. 2011.) Basic jitter is only one example of many heuristic spike resampling methods that have appeared in the literature. Although we recognize that these heuristic methods are useful exploratory tools, we recommend some caution in their use, and we recommend against their misuse as hypothesis tests. Whenever possible, heuristic methods, like basic jitter, should be converted into proper statistical methods, like interval jitter, which use well-formulated statistical models that facilitate communication and interpretation of results.

2.5 Accidental synchrony. One way to experiment with fine-temporal structure is to modify the simulated spike trains depicted in Fig. 1 by injecting randomly placed pairs of highly synchronous spikes, as might occur, for example, through common inputs or some other physiological process. To what extent, if any, will the presence of these pairs alter the conclusion that the fine-temporal structure is adequately explained by a slow modulation in firing rates? After all, each injection contributes real fine-temporal structure, but how is this to be separated from the "accidental" synchrony that results from correlated but slow changes in firing rate?

Three additional data sets were produced with progressively more injected synchronous spikes (Fig. 3). For each of the three data sets, a distribution of synchrony was generated by jittering spikes, exactly as in the experiment reported in Fig. 2. Figure 3B shows the resulting jitter distributions, along with the number of observed (±1 ms) synchronies and the *P* value of the observed synchronies relative to the jitter distribution. Notice that a total of 55 injected synchronies (Fig. 3B, *third* row, injection rate 0.5 Hz) in 100 s of recording containing
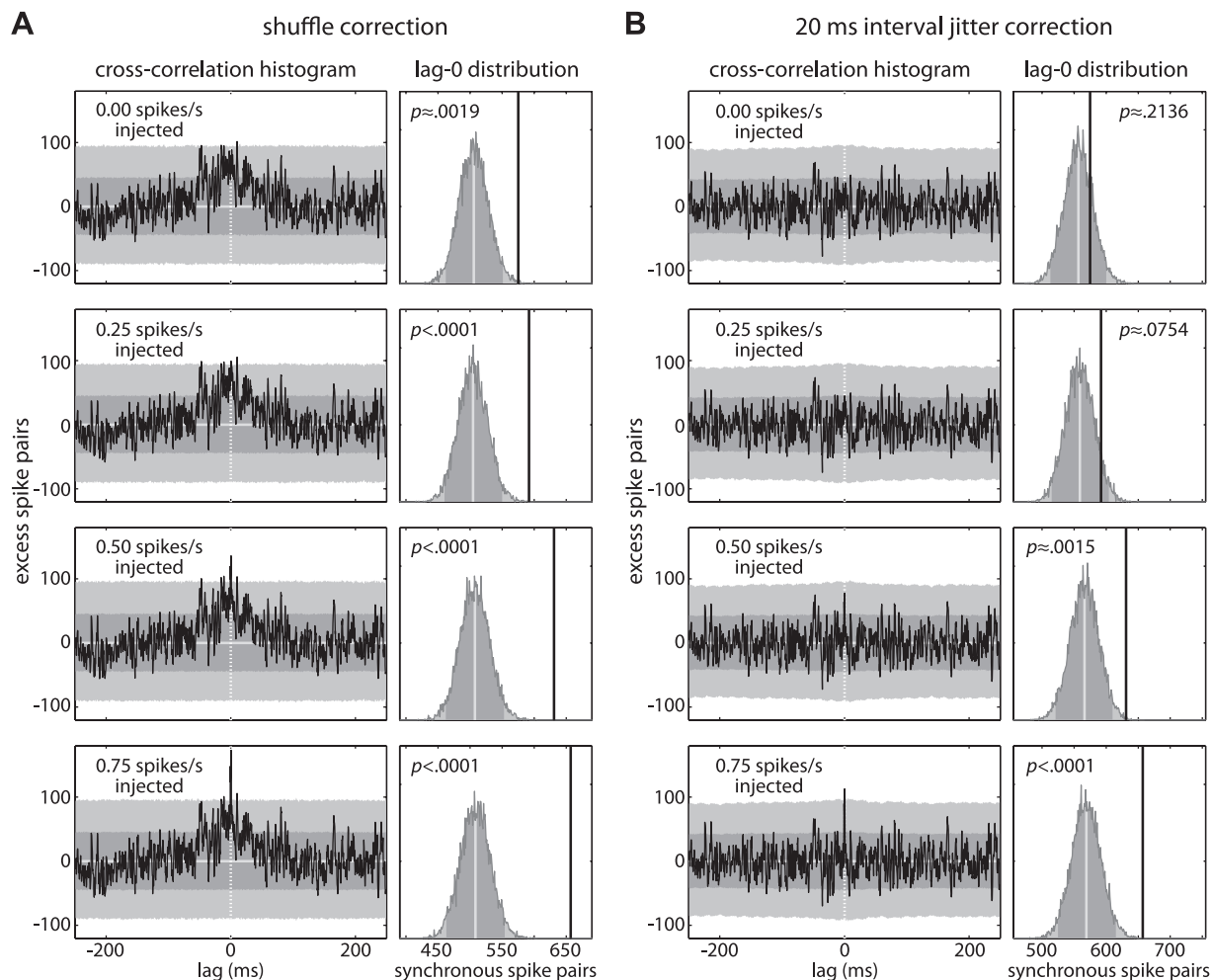
Fig. 3. Injected synchrony. Data in *top* rows correspond exactly to the data from Fig. 1: the shuffle-corrected CCH and shuffle-derived *lag 0* distribution in *A* are identical to those in Fig. 1, *E* and *D*, respectively, and the 20-ms interval jitter-corrected CCH and *lag 0* distribution in *B* are identical to those in Fig. 2, *D* and *C*, respectively. Data in the 2nd–4th rows correspond to new data sets created by injecting synchronous spike pairs into the original data set at the indicated rates of 0.25, 0.50, and 0.75 spikes/s, corresponding to mean values of 25, 50, and 75 total injected spike pairs, respectively, over all 100 trials (see Amarasingham et al. 2011). The actual (total) numbers of injected synchronous pairs are Poisson random variables and turned out to be 20, 55, and 88. *A*: shuffle-corrected CCHs and *lag 0* distributions. Shuffle-correction flags the zero-lag peak in the CCH as significant regardless of whether spikes are injected or not, i.e., because of slow, common rate fluctuations alone. *B*: 20-ms jitter-corrected CCHs and *lag 0* distributions. Small numbers of injected synchronies (relative to the total number of "accidental synchronies") are easily detected as statistically significant fine-temporal structure, whereas accidental synchronies arising only from slow, common rate fluctuations are not significant. A rough estimate for the number of injected synchronies is obtained by subtracting the mean of the histogram distribution from the observed synchrony count. This estimate works well here: the differences, corresponding to 0, 20, 55, and 88 injected synchronies, are 18.3, 32.9, 64.9, and 87.9, respectively. (The estimate is "rough" because it replaces the unobservable number of accidental synchronies by the mean of the jitter distribution. The spread of the jitter distribution gives some notion of the variability in this estimate.)

~5,000 spikes in each neuron and almost 600 accidental synchronies is already detected as significant ($P = 0.015$).

Returning to the issue of time scale, and keeping in mind that the jitter window in these experiments is 20 ms wide, it is not unreasonable to think of the jitter distribution on synchronies as a distribution on accidental synchronies. These are synchronies that occur by chance, given the numbers and approximate locations of spikes and given that rate changes occur on a coarse time scale, greater than 20 or so milliseconds. Correlations at finer time scales would be disrupted, and indeed the observed numbers of synchronies in the experiments with injected pairs move further into the tails of the jitter distributions as more pairs are injected. In fact, as shown in Fig. 3*B*, we can make a very rough estimate of the number of injected synchronies by subtracting the mean of the jitter distribution from the observed number of synchronies. It is always a good

idea to examine physiologically interpretable measures, like this rough estimate of injected synchrony rate, to verify the scientific, in addition to the statistical, significance of *P* values. Here we caution that although the hypothesis testing conclusions drawn from interval jitter are precise statistical conclusions, this rough "estimate" of injected synchrony rate is heuristic and should be interpreted as such. Proper statistical estimation of injected synchrony would require the development of precise models of injected synchrony, which are outside of the scope of this article.

2.6 Temporal resolution. What if repeated trials really are "repeated," in the statistical sense, and the spike trains are independent of one another? For example, suppose that the response of two neurons is characterized by firing rates that do not vary from trial to trial. There might then appear little to argue against trial shuffling to explore the significance of fine-temporal structure,

such as synchrony, above and beyond what might be explained by firing rates. However, if the two fixed firing rates for the neurons were themselves rapidly changing and correlated in time, then this might represent a possible mechanism, or at least a model, for fine-temporal structure. Yet, this kind of fine-temporal structure would be invisible to trial shuffling exactly because the trials are truly repeatable and the spike processes are dictated solely by their respective firing rates. By contrast, whether or not the assumption of statistical repeatability is valid, temporal processes operating at fast time scales relative to a given jitter window would be disrupted by jitter and thus in principle detectable by comparing statistics of the unperturbed spike train with those of the jittered spike train.

An extreme example is illustrated in Fig. 4. Pairs of model neurons are assumed to share inhomogeneous Poisson firing rates, and the same rate functions are repeated from trial to trial. By varying the bandwidths of the inhomogeneous rate functions, the sensitivity of jitter can be tested as a function of bandwidth and its relation to the size of the jitter window (20 ms in these experiments). As expected, jitter disrupted temporal structure at scales near or smaller than the jitter window. A similar experiment (not shown) fixing bandwidth while systematically increasing the size
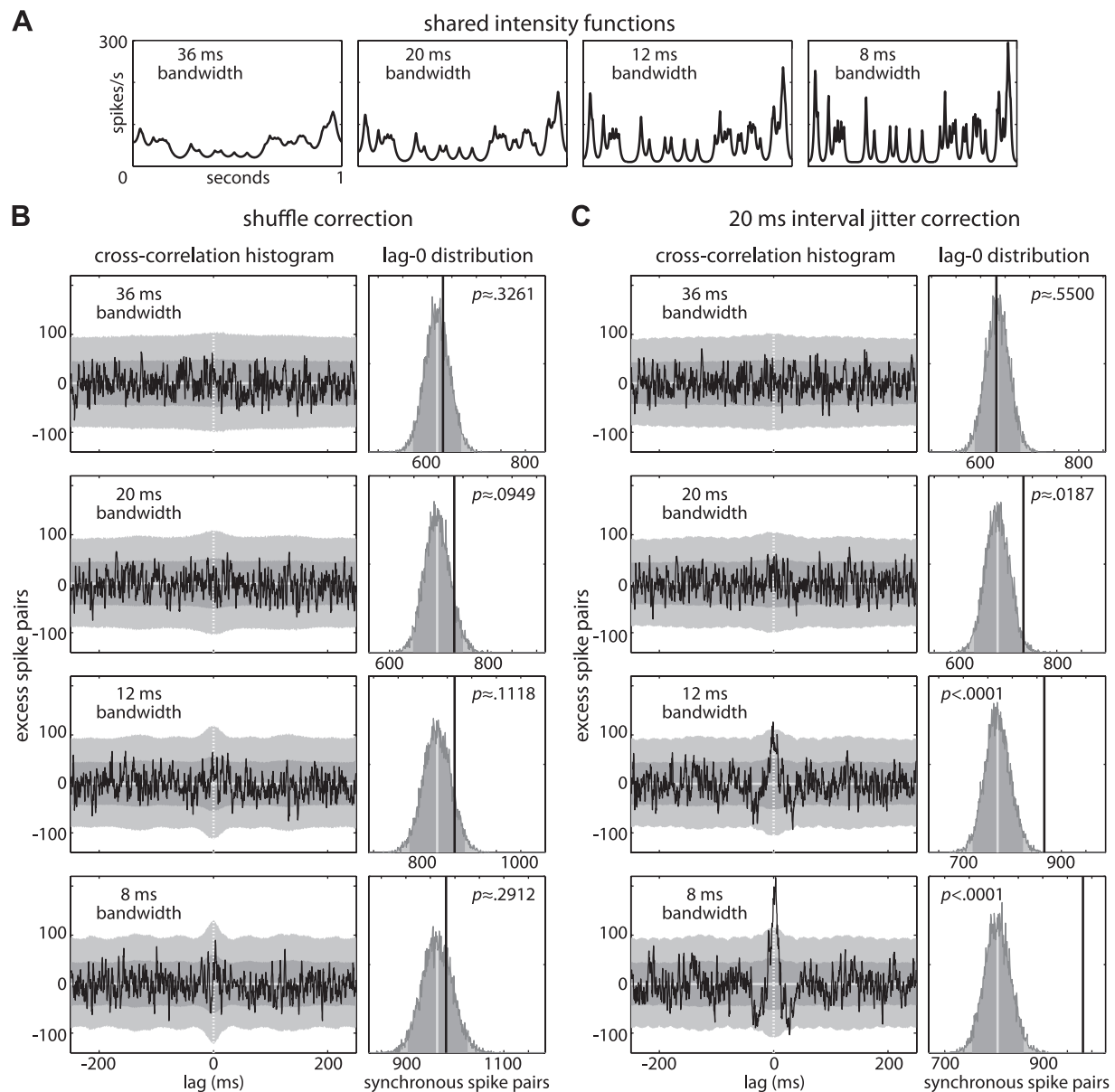

Fig. 4. Temporal resolution. *A*: rate functions, generated as in Fig. 1 but now with bandwidths of 36, 20, 12, and 8 ms (*left* to *right*). Each rate function is shared by *neuron 1* and *neuron 2* and is used to produce 100 Poisson spike processes for each neuron, simulating 100 independent and identically distributed trials of duration 1 s. (Unlike Fig. 1, there is no trial-to-trial variability.) *B*: for each rate function, the shuffle-corrected CCH from all 100 trials is shown, along with the *lag 0* distribution (as in Fig. 1, *E* and *D*). The CCHs do not reveal anything unusual, since the spike trains are mutually independent and there is no trial-to-trial variability. *C*: same as *B*, except using 20-ms interval jitter instead of trial shuffling. The 20-ms jitter disrupts spike correlations that arise from shared rate fluctuations with bandwidth of 8 ms but not with bandwidth of 36 ms, with a general progression from more disrupted to less disrupted through intermediate values. Comparing *B* and *C*, excess synchrony arising from fast, shared rate fluctuations is missed in the shuffle-corrected CCH but flagged in the jitter-corrected CCH. (The fact that by using 20-ms jitter we obtain statistical significance precisely for 20-ms bandwidths and higher frequency is, in this case, a coincidence. The time scales involved should only be qualitatively compared.)

of the jitter window gives similar results: a more or less monotonic rise in significance (drop in *P* value).

2.7 The jitter-corrected cross-correlation histogram. A traditional way to look for synchrony is to look for peaks in the cross-correlation histogram in and around *lag* 0. The other bins, corresponding to other spike-to-spike intervals, provide analogous evidence for other fine-temporal relationships, such as those that might arise from direct (monosynaptic) connections between two recorded neurons. As described in Fig. 1, *C* and *E*, and further illustrated in Figs. 3*A* and 4*B*, one way to explore peaks in the cross-correlation histogram is through surrogate data generated by trial shuffle. An alternative approach, suggested by the apparent superiority of jitter methods in differentiating the temporal scales underlying observed coincidences, is to use jitter in place of shuffle to generate the surrogate data. The construction, then, is the same as the one described in Fig. 1*E*, except that 10,000 jitter-generated pairs of spike trains are used instead of 10,000 shuffle-generated pairs of spike trains. Figures 2*D*, 3*B*, and 4*C* revisit the experiments of Section 2, comparing shuffle-corrected with jitter-corrected cross-correlation histograms. The results are consistent with the observations already made about the potential hazards of trial shuffling and the potential benefits of jitter-based resampling. In that neither approach imposes a computational barrier, we recommend jitter over trial shuffling for producing corrected cross-correlation histograms and acceptance bands when the goal is to isolate fine-temporal correlations.

3 Variations on the Jitter Theme

Jitter methods are an instance of conditional modeling. Null hypotheses of temporal resolution are specified via conditional distributions on the data. The key idea is to condition on the spike counts within intervals. The temporal extent of the intervals then specifies the temporal resolution that is hypothesized. Once the conditioning framework is in place, many extensions and generalizations follow naturally. In this section, we review a selection of examples.

3.1 Rate of change of intensity functions. Hypothesis tests, perhaps because of their simplicity as statistical decisions, can make extremely efficient use of data. This phenomenon is already apparent in Section 2.5, where very small quantities of injected synchronies (e.g., 55 out of ~5,000 spikes and 600 accidental synchronies) can be picked out by the jitter procedure. Such sensitivity, a virtue, has a subtle corollary that calls for caution: a hypothesis test is equally capable of picking out a very fine misspecification of the null hypothesis. Examining this idea closely, the jitter null hypothesis, that having fixed the number of spikes in a jitter window every configuration of those spikes has precisely the same probability, might be criticized as an extreme interpretation of the absence of temporal precision. In particular, it is important to understand how firing rate variations within a jitter window affect jitter-derived conclusions. If, for example, we were to assume that the spike sequences in two neurons were generated by Poisson processes with time-varying rate functions (as in the artificially generated data of Section 2), how might we test the hypothesis that firing rates do not vary by more than, say, 25% within any 20-ms window? Interval jitter can accommodate exact tests for this and related hypotheses (Amarasingham 2004; Harrison 2005)

through the use of nonuniform jitter distributions, chosen, for example, to maximally promote synchronies. See Amarasingham et al. (2011) and its appendix for details.

How does the *P* value, as computed using the interval jitter procedure introduced in Section 2.1, hold up under the more conservative and exact test provided by interval jitter with nonuniform relocation probabilities? Referring to the experiments in Section 2.5, we found that in the case of 0.75-Hz injected synchrony (Fig. 3, *bottom* row) we could allow for as much as an 81% change in firing rate (i.e., maximum rate/minimum rate $\leq$ 1.81) over each 20-ms jitter interval and still achieve a *P* value below 0.1. (Of course, it is easier to reject stricter limits on the change in firing rate; smaller *P* values go with smaller percentages.) Staying with the Poisson example (but the null hypothesis here is far more general; see Section 2.2 and Amarasingham et al. 2011), "we reject (at *P* = 0.1) the hypothesis that the data was generated from an inhomogeneous Poisson process with rate function piecewise linear on 20-ms jitter intervals and changing no faster than 81% in any one interval." In plain words, the evidence favors a fast (finetemporal) mechanism (as well it should given that we injected pairs of millisecond-synchronous spikes). For 0.5-Hz injected synchrony (Fig. 3, *third* row), *P* remains below 0.1 with up to 41% change in each 20-ms interval. In short, interval jitter continues to flag the presence of small numbers of synchronous events even when recast into the framework of a well-specified, exact, and conservative hypothesis test.

3.2 Resampling patterns. Relative and absolute refractory periods, and the patterned sequences of neuronal spikes known as bursting, in and of themselves constitute a form of finetemporal structure. A neuron that fires in highly prototypical bursts, with well-preserved interspike intervals, is indeed operating with high precision. In that these phenomena restrict the placements of spikes relative to neighboring spikes, they already violate the hypothesis under which interval jitter calibrates the likelihood of an observed count of synchrony. How then is the practitioner to interpret observations of "excess" synchrony? Are the alignments an epiphenomenon, secondary to the structuring effects of refractory periods (Baker and Lemon 2000; Oram et al. 1999) and bursting (Abeles and Gat 2001), or are they due, instead, to less local properties such as common input, network oscillations, or poststimulus transients?

Pattern jitter is a method for exploring fine-temporal structure in spike trains that cannot be explained by short-range temporal dependencies among spikes of single neurons. The idea is to construct a randomization technique that not only preserves spike counts but also preserves the local statistics of neighboring spikes within single neurons. To this end, pattern jitter is a spike resampling method that rigidly moves local patterns of successive spikes. The short-range statistics within individual neurons are preserved, exactly, through restrictions on the allowable translations. Like interval jitter, pattern jitter corresponds to a proper statistical hypothesis test, as will be transparent from the conditional modeling perspective.

A "pattern" is a sequence of successive spikes from a single neuron satisfying two constraints: on one hand, the distances between neighboring spikes in the pattern are small, but on the other hand, the pattern is well separated from surrounding patterns. Precisely, given a "history" parameter, *R*, successive spikes in a pattern are separated by no more than *R* ms, and

there are no surrounding spikes from the same neuron that are within R ms of the pattern spikes. The actual number of spikes in a pattern is variable. The smallest pattern is made of a single spike, but larger patterns will be typical when R is large. For the synchrony analysis at the beginning of Section 4, we used $R = 100$ ms; for the triplet analysis in Section 4.1, we explored a variety of values from 1 to 16 ms. Notice that the parameter R defines a unique partitioning of a spike train into patterns.

Rather than jittering individual spikes, we jitter patterns. If we designate the location (time) of a pattern to be the location of its first spike, then pattern jitter is just a generalization of interval jitter: $R = 0$, which generates single-spike patterns, is replaced by $R > 0$. Windows are defined a priori, as with interval jitter (Section 2.1), but patterns (rather than spikes) are randomly relocated in such a way that the first spike of each pattern (i.e., the pattern location) remains in its original interval. The only complication is that to preserve the statistics of local spike arrangements (as defined by the history parameter R), we must prevent the appearance of new patterns in the new (jittered) spike sequence. Hence, patterns must remain separated by more than R ms. Fortunately, an ensemble of pattern-preserving spike trains can be efficiently generated (Harrison and Geman 2009; see http://www.dam.brown.edu/ptg/jitter/ to download a library of pattern jitter tools).

Figure 5 illustrates pattern jitter using two jitter intervals (20 and 64 ms) at three different values of R (0, 25, and 100 ms). The point is to produce a set of surrogate records that are statistically indistinguishable (exchangeable) from each other and from the recorded spike train, under a null hypothesis that is consistent with physiological interactions between spikes having a range of up to R ms within a neuron. The null hypothesis focuses on the issue of temporal resolution. There are no unnecessary assumptions. One way to appreciate this is to interpret the null hypothesis from the generative point of view, as in Section 2.2: one produces spike records by choosing sequences of spike patterns that accommodate absolute and relative refractory periods, bursting, or any other interspike structure up to the specified history length of R ms (i.e., every

2 successive spikes in a pattern are separated by no more than R ms). There are no constraints on the choice of these patterns; they can be large or small, and they can depend on each other or not. The only constraint is on the locations of the patterns, and it is the one implied by their definition: they cannot be positioned so closely as to destroy the chosen pattern structure, i.e., they need to be spaced by more than R ms (distance between last spike of one pattern and first spike of the following pattern exceeds R ms). If we think of a pattern as located in the interval that contains its first spike, then the rest of the story is the same as for spike jitter. Hypotheses about temporal resolution are translated into intrainterval location likelihoods. A variety of statistics (e.g., synchrony counts, as in Section 2.1, or repetitions of precisely timed spike sequences, as in Section 4.1) can then be used to construct exact tests for the validity of the null hypothesis. (See Amarasingham et al. 2011 for further simulation experiments illustrating the uses of pattern jitter and the role of the choice of R).

4 Jitter Analysis of Three Motorcortical Neurons

Three neurons (designated N1, N2, and N3) were selected from simultaneous multineuronal records extracted from the measurements of a chronically implanted, 100-electrode array (1.0-mm electrode length; 400-μm interelectrode separation) in primary motor cortex (MI) of the monkey (Hatsopoulos et al. 2007). The monkey was operantly trained to move a cursor to targets projected onto a horizontal, reflective surface in front of the monkey. At any one time, a single target appeared at a random location in the workspace, and the monkey was required to move to it. As soon as the cursor reached the target, the target disappeared and a new target appeared in a new, random location (see Amarasingham et al. 2011 for more details on spike sorting and experimental protocol).

There were 391 successfully completed trials, each with a sequence of several targets. Spike times were discretized to 1/30 ms and were analyzed for the duration of the first second for each neuron of each successful trial. The three neurons, N1,
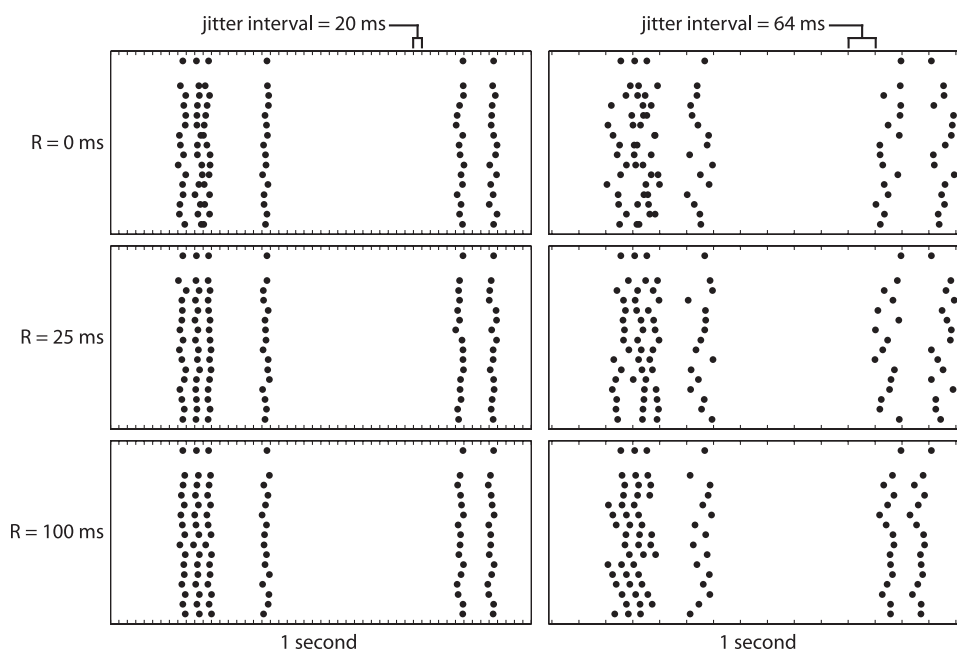
Fig. 5. Pattern jitter. The top horizontal line of dots for each panel shows an identical sequence of observed spike times (dots) during a 1-s interval. Each of the remaining 15 lines of dots shows an independent spike sequence generated by pattern jitter. Jitter intervals in *left* panels are 20 ms; those in *right* panels are 64 ms. Notice that coarse-temporal modulations in firing rate are preserved, more so with 20-ms jitter than with 64-ms jitter. The *top*, *middle*, and *bottom* rows have $R = 0$-, 25-, and 100-ms histories, respectively. The only patterns in the *top* and *middle* rows are single spikes. Nevertheless, because the *middle* row has a larger R, the sampled spike sequences are more regular: spikes cannot be jittered too close to one another or new patterns will be created. The *bottom* row has 3 patterns. Notice that the patterns are preserved in the sampled sequences and that only the first spike of each pattern is constrained to remain in its original jitter interval. (Note also that in the case of $R = 0$, pattern jitter preserves "histories of length zero," meaning that it prevents 2 spikes from landing in the same bin.)

N2, and N3, were recorded from separate electrodes. All three had relatively low firing rates: 5.39, 5.12, and 4.28 Hz, respectively. Spike-time rasters and poststimulus time histograms (PSTH) are shown in Fig. 6A. The three shuffle-corrected cross-correlation histograms (N1/N2, N1/N3, and N2/N3), along with shuffle-derived 95% acceptance bands (as described in Fig. 1), are shown in Fig. 6B, *left*.

There is evidence in the cross-correlation histograms for fine-temporal correlations between N1 and N2 as well as between N1 and N3. Analyses with interval and pattern jitter essentially rule out explanations for the excess synchrony that are based on slow but correlated fluctuations in firing rates and trial-to-trial variation (Fig. 6B, *middle*), or "history effects,"

such as relative or absolute refractory periods, or bursting (Fig. 6B, *right*). (Although these fine-temporal correlations cannot be reasonably explained as artifacts of trial-to-trial variation or statistical dependencies among spikes within single neurons, our purpose here is to demonstrate jitter techniques rather than to argue for the existence or role of fine-temporal structure. Thus we want to be clear that these three neurons were not chosen randomly from a library of recordings. To the contrary, they were selected for the specific purpose of illustration.)

4.1 Beyond synchrony: other measures of precision. In principle, an ensemble of jittered spike trains, constrained to reproduce length-R local spike statistics and coarse (trial dependent) time-varying firing rates, can be used to explore any
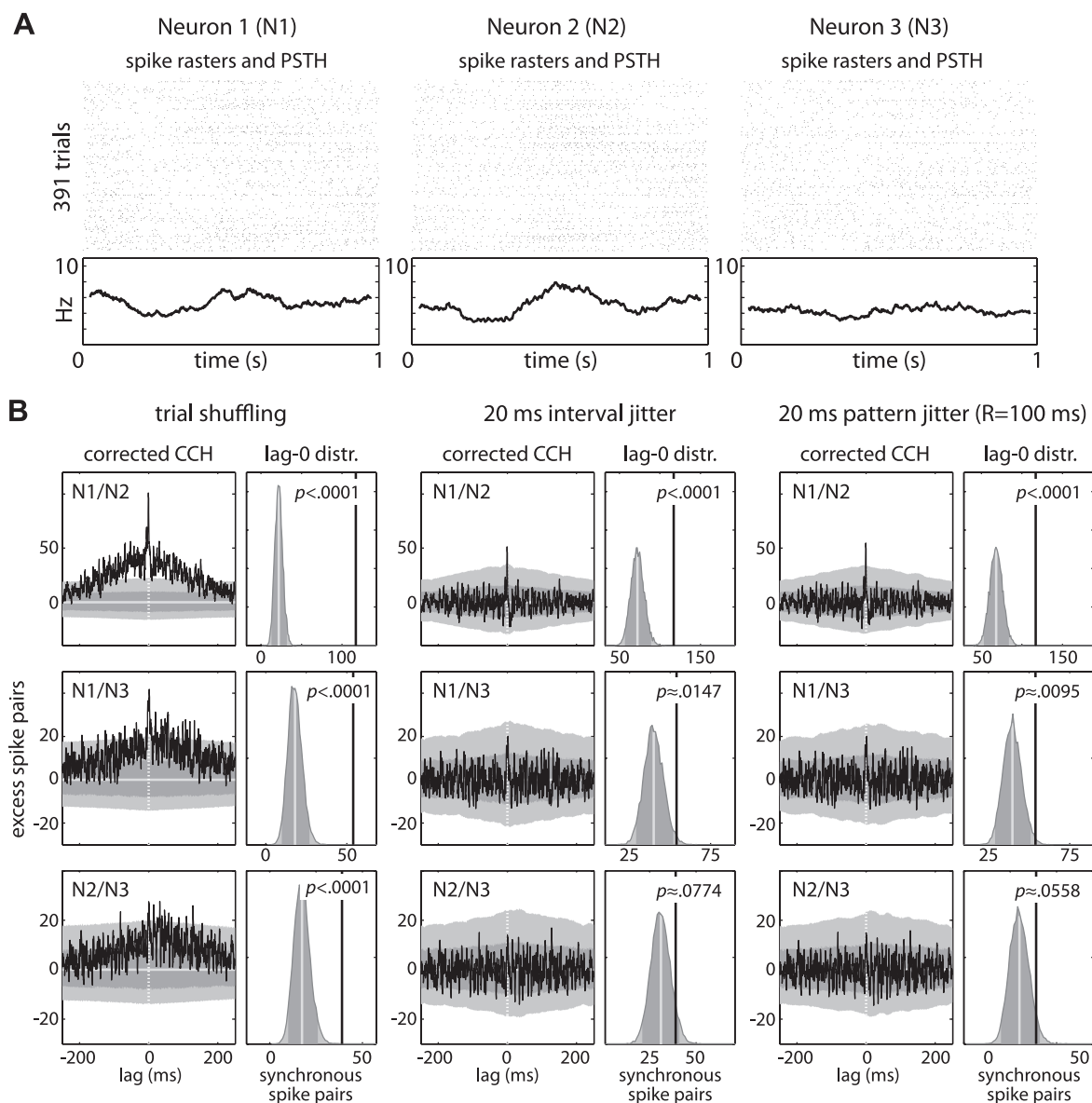

Fig. 6. Simultaneous recordings from monkey motor cortex. *A*: raster plot and poststimulus time histogram (PSTH) for each of 3 simultaneously recorded neurons. Data are from the first 1,000 ms of each of 391 trials. *B, left*: CCHs with 95% shuffle-derived acceptance bands and shuffle-derived distributions of synchrony counts ($\pm$ 1 ms), for each of the three pairs of analyzed neurons. *Middle*: single-spike interval jitter. Jitter interval is 20 ms. Low P values for the N1/N2 and N1/N3 neuron pairs rule out slow and correlated variations in firing rate as a likely explanation for the excess synchrony. In fact, even allowing for up to 50% variation in firing rate within each 20-ms window (see Section 3.4), the P value for the N1/N2 pair remains below 0.001, and the P value for the N1/N3 pair remains below 0.05. *Right*: pattern jitter, also using 20-ms intervals. Spikes are jittered while exactly preserving all interspike intervals of duration less than or equal to $R = 100$ ms. The N1/N2 and N1/N3 pairs remain significant at $P = 0.0001$ and $P = 0.01$, respectively, effectively ruling out absolute or relative refractory periods, or bursting phenomena, as explanations for the excess synchrony.

measure of temporal precision, not just spike synchronies. To illustrate, we looked at the recurrences of specific, precisely timed sequences of three spike events ("triplets") within neuron N2, throughout the 391 successfully completed trials (a total of 1,752 s of data containing 8,455 spikes). Abeles and colleagues offered repetitions of precisely timed spike sequences (Abeles and Gerstein 1988; Abeles et al. 1993) as evidence for a hypothesized collective neuronal dynamic known as a synfire chain. They reported that the numbers of repetitions of such spike sequences were excessive (Abeles et al. 1993), measured against what would be expected from Poisson-like statistics. On the other hand, Oram et al. (1999) and later Baker and Lemon (2000), both using spike resampling methods, found that the number of repetitions in their data, of triplets as well as other precisely timed sequences, was adequately explained by taking into account the distribution of nearest-neighbor interspike intervals and inhomogeneous firing rates.

There are many statistics involving repeating triplets that could be examined for evidence of fine-temporal structure (Abeles and Gat 2001; Baker and Lemon 2000; Oram et al. 1999) We chose to look at the maximum number of repetitions of any triplet (Date et al. 1999), reasoning that an extremal statistic is likely to be particularly sensitive to repeatable and precise influences on spike timings. The N2 recordings contained 15 triplets (t_1, t_2, t_3) with interspike times $t_2 - t_1 = 23$ ms and $t_3 - t_2 = 27$ ms, and this was the maximum number of repetitions of any triplet in the record. Pattern jitter points to significant rate changes occurring on a scale of about 64 ms (or finer), or significant interspike interactions that extend at the least 8 ms, or both (see Fig. 7).

Obviously, triplet repetitions could be substituted with other measures of precision, dictated by the experiment and the experimental focus. Spike sequences involving more than three spikes, or more than one neuron, and statistics such as the total number of repeating sequences (not just the maximum number of repeated triplets), are some of the choices, but really any function of the recorded data (i.e., any "statistic") can be used to test for fine-temporal structure with the same resampling method. Naturally, different statistics have different power, depending on the hypothesis being tested. Repetitions of multispike sequences make a sensible statistic when looking for synfire chains; millisecond synchrony is sensible when looking for common input. In any case, the usual caution applies: it is unwise and poor practice to use a statistic suggested by the data itself, at least for any purpose beyond exploratory analysis. In fact, it is arguably good practice to use simple or at least clearly motivated statistics; an uninitiated reader may be more likely to suspect that an exotic, unmotivated statistic has been "fished out" of the data, possibly for good reason.

4.2 Temporal resolution and the neural representation of behavior. When do precisely timed events, such as millisecond synchronies or specific patterns of interspike intervals, carry information about sensory stimuli or motor actions? What role, if any, can jitter methods play in assessing an apparent relationship between a fine-temporal event (e.g., millisecond synchronies) and a response or stimulus variable (e.g., the direction of movement or the identity of a target object)? Although jitter-based hypothesis testing is not, per se, a tool for measuring information or quantifying dependence, it can still be useful in these contexts for evaluating whether a relationship is
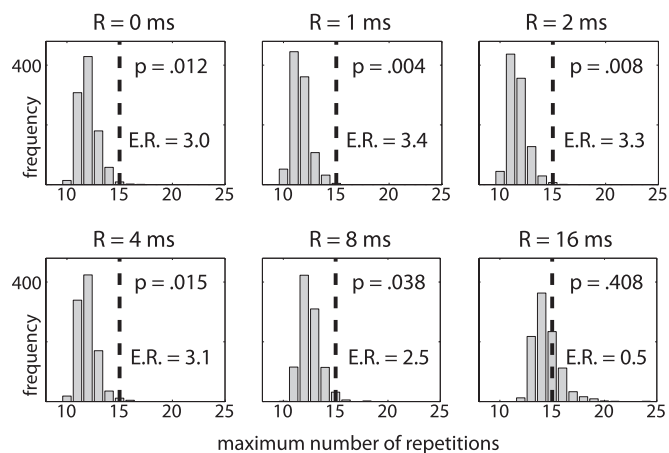

Fig. 7. Maximum number of triplet repetitions. A triplet is a sequence of 3 increasing intratrial times (t_1, t_2, t_3) rounded to the nearest ms, each occupied by a spike, such that $t_2 - t_1$ and $t_3 - t_2$ are no greater than 1 s. There may be multiple spikes, in addition to the spike at t_2, in the interval between t_1 and t_3. Triplets with $t_2 - t_1 = 23$ ms and $t_3 - t_2 = 27$ ms occurred 15 times in N2 during 391 successful trials; no other triplet repeated as often. The 6 histograms show the maximum numbers of repetitions of triplets in 1,000 resampled spike trains produced by pattern jitter. The jitter interval in all of these experiments is 64 ms, and the pattern histories are, as indicated, 0 (jittering individual spikes), 1, 2, 4, 8, and 16 ms. (Notice that history values of 27.5 ms or larger would preserve exactly all 15 of the maximally repeating triplets in the original recording.) The displayed *P* values are the respective probabilities of finding 15 or more repetitions of a triplet in the ensemble of resampled spike trains. [With $R = 0$ and jitter interval 32 ms (not shown), the corresponding *P* value is 0.055.] Excess repetition (E.R.) is the difference between 15 (the observed maximum repetition count) and the mean of the resampling distribution. Small *P* values are evidence for additional structure in the spike-time recordings, beyond what could be reasonably accounted for by changes in firing rate on a time scale coarser than the jitter interval or by spike-history effects of durations below *R*.

indeed the result of fine-temporal structure. To choose a few prominent examples, Jones et al. (2004) and Butts et al. (2007) used jitter techniques to assess the contribution of spike timing to the ability of a reconstruction model to predict stimuli from spike trains in the rat trigeminal system and cat lateral geniculate nucleus, respectively. Fujisawa et al. (2008) employed a jitter principle to identify and subsequently study the dynamics of putatively monosynaptic activity in a large population of simultaneously recorded rat medial prefrontal cortex neurons.

Here we reanalyze the data from neurons N1, N2, and N3 to illustrate the application of pattern jitter for exploring the relationship between synchrony events and direction of movement. For a given pair of neurons, let $S = 1$ indicate the occurrence of a synchrony and $D = d$ indicate the instantaneous movement in the direction d. One way to quantify the relationship between synchrony and movement direction is through the ratio $\theta(d) = P(D = d, S = 1)/[P(D = d)P(S = 1)]$. If $\theta(d)$ does not vary with d, then $\theta(d)$ is identically one and movement direction and synchrony are statistically independent. If $\theta(d)$ varies with d, then there is a statistical relationship between synchrony and movement direction such that directions d with large values of $\theta(d)$ are associated with high probabilities of synchrony. In this sense, $\theta(d)$ can be viewed as a synchrony-based tuning curve. [Amarasingham et al. 2011 contains further discussion of the interpretation of $\theta(d)$, which arises as a fundamental quantity in both statistics and information theory. Our point, however, is not to argue for $\theta(d)$ as a uniquely well-suited measure of the relationship; there are

many others, some more model based, that explicitly attempt to separate, or identify, the relative contributions from firing rates and synchrony in the prediction of direction.]

Figure 8 shows estimates of $\theta(d)$ for synchronies defined by each of the three neuron pairs. In each case, the estimated $\theta(d)$ modulates strongly with d, indicating a relationship between precise millisecond synchrony and movement direction. Synchronous spikes between neurons N1 and N2 are more common when the monkey is moving in the downward and rightward direction, and more common between neurons N1 and N3 with movement in the upward direction. There are other notable features, and in fact each pair of neurons produces several peaks in the directional tuning curve. What should be made of these relationships? Are they evidence for a tight statistical, and perhaps even mechanistic, connection between the fine-temporal structure of spike patterns and movement direction? If so, then we would expect these relationships to be largely altered by jittering the original spikes. On the other hand, if the same relationships are largely preserved after jittering, then it seems likely that they are primarily a result of the coarse temporal dynamics of spiking and not additional precise spike timing: we are merely viewing these coarse dynamics through their effects on precise millisecond synchrony.

Consider the prominent peak at about $d = 120°$ in the direction-tuning curve based on N1/N2 synchronies (Fig. 8). Our estimator of $\theta(120)$ is a function of the data, i.e., a statistic, and in this regard no different from, say, a synchrony count or maximum occurrence of precisely timed triplets. In fact, our estimate of the curve $\theta(d)$ across varying d is no different from a cross-correlation histogram across varying lags. We used these estimators as test statistics to evaluate the pattern-jitter null hypothesis ($R = 100$ ms with 20-ms jitter intervals) in exactly the same way that we used the cross-correlation histogram and the *lag 0* synchrony counts in Fig. 6B. [Strictly speaking, we are conditioning not only on the observed patterns and their interval locations but also on the sequence of directions $D(t)$, because these too remain fixed in the subsequent jitter experiments.] Under this null hypothesis, we would expect the observed estimate of $\theta(d)$ to be typical of the

ensemble of identical estimators derived from the ensemble of surrogate spike trains generated by pattern jitter. This is indeed the case: there is nothing notable about the original record, vis-à-vis the estimate of $\theta(d)$, compared with the ensemble of estimators derived under the null hypothesis, as indicated by the acceptance bands in Fig. 8. Jitter analysis indicates that the relationship between fine-temporal structure and movement direction is secondary to coarsely timed differences in the locations of spikes.

In Fig. 6B we used the observed number of precise millisecond synchronies as a test statistic to reject the null hypothesis of no fine-temporal structure. In Fig. 8, we used the estimated tuning curve, $\theta(d)$, as a test statistic for the same null hypothesis, this time failing to reject. In these data, there is strong evidence for fine-temporal structure that is not explained by the coarse placement of spikes, but there is little to suggest that fine-temporal structure adds to coarse structure for the purpose of predicting movement direction. Quantitative conclusions about why the data reveal fine-temporal structure using one test statistic but not another require statistical models incorporating varying degrees of precise timing. More generally, what is needed to go further is a formulation not only of the null hypothesis but also of its many alternatives.

5 Summary and Discussion

It has by now been widely observed that the classical distinction between rate and temporal coding lies on a continuum and is more generally an issue of time scale (Dayan and Abbott 2001; Rieke et al. 1997). In vivo evidence regarding timing (Tiesinga et al. 2008) is commonly drawn from statistical analyses of correlations between environmental or behavioral signals and spike counts measured on various time scales. For example, many investigators model neurons with stationary or time-varying firing rates for each trial and assume that neurons are firing independently, given their firing rates (Frostig et al. 1990; Martignon et al. 1995; Prut et al. 1998; Villa et al. 1999). Estimated firing rates are then incorporated into a bootstrap hypothesis test of such models (Aertsen et al. 1989; Grün et al. 2002a, 2002b; Samonds et al. 2006; Ventura
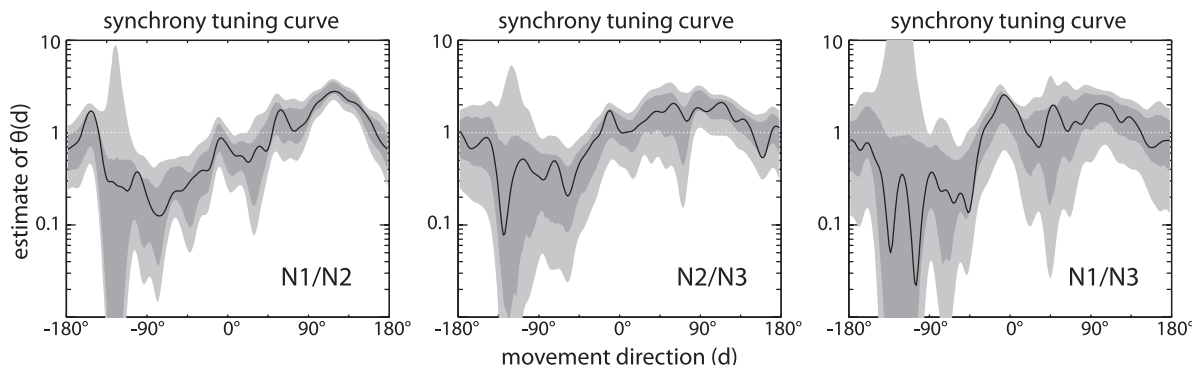

Fig. 8. Jitter-calibrated tuning curves. For a given pair of neurons, Ni/Nj, and a given time $t = 1, 2, \ldots, T$ (binned to 1 ms), let $S(t)$ be the 0–1 "synchrony process," where $S(t) = 1$ signals a spike in neuron Ni during the tth bin that occurred within ± 1 ms of a spike in neuron Nj. Let $D(t)$ be the direction, in degrees, of movement of the manipulator (and hence of the monkey's hand) in time bin $t + s$, where s is chosen as a typical delay between motor cortex and observed kinematics, 100 ms in this analysis. The distribution on direction, $P(D = d)$, and the conditional distribution, $P(D = d|S = 1)$, given a pair of synchronous spikes, were estimated using a simple kernel density estimator with Gaussian kernel and bandwidth of 6°. The ratio of these estimates gives an estimate of the likelihood-ratio tuning curve, $\theta(d) = P(D = d, S = 1)/[P(D = d)P(S = 1)] = P(D = d|S = 1)/P(D = d)$, for each direction d. Estimates of $\theta(d)$ are plotted for each of the 3 pairs of neurons (log-scale). One thousand surrogate spike trains generated by pattern jitter (20-ms jitter window, 100-ms preserved history) were used to produce 1,000 additional synchrony-based tuning curves and 95% pointwise (dark gray) and simultaneous (light gray) acceptance bands (see Amarasingham et al. 2011).

et al. 2005a; see also Ito and Tsuji 2000) wherein fine-temporal statistics such as synchrony counts (see also Abeles and Gerstein 1988; Abeles et al. 1993) are utilized as the test statistic. Another approach along similar lines is to compare observed synchrony counts with those expected if all trial pairings were equally likely, as in the shuffle predictor (Constantinidis et al. 2002; Hirabayashi and Miyashita 2005; Perkel et al. 1967). The key underlying assumption (null hypothesis) of these techniques is that neurons act independently of each other, given a probabilistic structure that is common to all trials.

However, many authors (Bair et al. 2001; Ben-Shaul et al. 2001; Brody 1998) have pointed out that the relevance of such an analysis to the issue of time scale is unclear. The basic, and quite plausible, counterexamples are forms of trial-to-trial variability, such as latency variations (Ventura 2004) or slowly varying common input, such as that due to internal variables not controlled by the experiment (Baker and Gerstein 2001; Czanner et al. 2008; Grün et al. 2003; Kass and Ventura 2006). One proposal for dealing with such concerns is to incorporate trial-varying firing rates into parametric models of the spiking process, which can again be tested, for example, by estimated trial-varying firing rates incorporated into bootstrap tests (Bair et al. 2001; Ben-Shaul et al. 2001; Brody 1999; Pauluis and Baker 2000; Ventura et al. 2005b). Of course, in any such model the number of parameters that must be estimated from the data is proportional to the number of trials, and for this reason, at least, standard statistical issues concerning model complexity become very delicate. Implicitly or explicitly, stronger modeling assumptions are required. This raises a question of interpretation. Does the rejection of such a model, using synchrony counts as a test statistic, for example, indicate that there is fine-temporal structure, or does it reflect, instead, the particulars of the model?

Jitter methods are a statistically rigorous alternative approach. In place of estimating firing rates, an ensemble of spike processes is defined through random perturbations of observed spike timings, with the intention of preserving coarse characteristics of the spiking process (e.g., fluctuations in firing rate) while disrupting fine-temporal structure. Conclusions about the existence of fine-temporal structure are then derived from a comparison of the original spiking process with those in the randomly generated ensemble.

Jitter is an application of conditional modeling. The task of modeling the structure and mechanisms of the spike-generating process are largely sidestepped by formulating hypotheses in terms of the conditional distribution on the locations of spikes or spike patterns, given their observed counts in coarsely defined bins (the jitter window). Local firing rates are preserved to a specified resolution, determined by the width of the window, and local spike dependencies are preserved to a specified range, determined by the history parameter R. How surprising are apparent coincidences of spike timing, such as a large number of synchronous spike pairs or multiple repetitions of precisely timed spike sequences? An ensemble of jittered spike trains produces a corresponding ensemble of coincidences. By comparing the original observation with the ensemble, hypotheses about precision are put to the test: do the observations in the recorded spike train depart from the range of values that could be reasonably expected under the hypothesis that spike times are unconstrained beyond slowly varying firing rates and local interspike interactions?

Jitter is a tool for exploring hypotheses about temporal resolution. It should not be expected to address questions about mechanism that go beyond, or are unrelated to, issues of timing. Thus, for example, statistical or physiological models for spike generation, such as Poisson-process models or more general point-process models (e.g., Ventura et al. 2005b), serve a different purpose. In our view, it is best not to confuse this type of model building and parameter estimation with testing hypotheses about temporal resolution or with exploring data for evidence of excess synchrony or surprising repetitions of spike sequences. Indeed, the two are nearly orthogonal and can be complementary. In Section 4.2 we estimated synchrony-based direction-tuning curves by using standard and time-honored statistical methods. Jitter played no role, except in the interpretation of these curves in terms of temporal resolution. In that it focuses only on the timing issue, jitter is a good way to evaluate the statistical significance of the estimated tuning curves or, for that matter, any other estimated relationship between fine-temporal structure and behavior.

As a final remark on conditional modeling, we note that this point of view serves to clarify the relationship between spike or pattern jitter and bootstrap methods. The two are superficially similar and easily confused. Bootstrap samples are generally interpreted as samples from the full (i.e., unconditional) distribution on the data, whereas jitter gives samples from a hypothesized conditional distribution. The interpretations are completely different. Bootstrap samples, for example, might be used to investigate the variability in observed firing rates. It would make no sense to use jitter samples for this purpose, since by design, slowly varying firing rates are preserved, almost exactly, from one surrogate spike train to another. (See Amarasingham et al. 2011 for additional remarks on the distinction between conditional and unconditional inference.) Although the primary difference is one of interpretation, we also note that proper use of bootstrap requires good estimates of the full data distribution. As we have pointed out, this is a strong requirement, especially in neurophysiology. Jitter avoids the estimation problem by narrowly focusing on scientifically relevant questions about conditional distributions.

There are many variations on the theme. The jitter distribution can be uniform in a window, or tilted by as much as a prespecified amount, to probe for lower bounds on the intrinsic rate of change of spike probabilities. A model that purports to predict a stimulus or a motor response from fine-temporal neuronal events, such as synchronies or precisely timed spike sequences, can be evaluated for its sensitivity to temporal precision. We have reviewed these and many other applications of the jitter method. In each instance, the statistical assumptions are revealed and the statistical correctness is established through the conditional modeling perspective.

REFERENCES

Abeles M. *Corticonics: Neuronal Circuits of the Cerebral Cortex.* Cambridge, UK: Cambridge University Press, 1991.

Abeles M, Bergman H, Margalit E, Vaadia E. Spatiotemporal firing patterns in the frontal cortex of behaving monkeys. *J Neurophysiol* 70: 1629–1638, 1993.

Abeles M, Gat I. Detecting precise firing sequences in experimental data. *J Neurosci Methods* 107: 141–154, 2001.

Abeles M, Gerstein G. Detecting spatiotemporal firing patterns among simultaneously recorded single neurons. *J Neurophysiol* 60: 909–924, 1988.

Aertsen A, Gerstein G, Habib M, Palm G. Dynamics of neuronal firing correlation: modulation of "effective connectivity". *J Neurophysiol* 61: 900–917, 1989.

Alonso J, Usrey W, Reid R. Precisely correlated firing in cells of the lateral geniculate nucleus. *Nature* 383: 815–819, 1996.

Amarasingham A. *Statistical Methods for the Assessment of Temporal Structure in the Activity of the Nervous System* (dissertation). Providence, RI: Brown University, 2004.

Amarasingham A, Chen TL, Geman S, Harrison MT, Sheinberg D. Spike count reliability and the Poisson hypothesis. *J Neurosci* 26: 801–809, 2006.

Amarasingham A, Harrison MT, Hatsopoulos NG, Geman S. Conditional modeling and the jitter method of spike resampling: supplement. arXiv: 1111.4296 [stat.ME], 2011.

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

Baker S, Gerstein G. Determination of response latency and its application to normalization of cross-correlation measures. *Neural Comput* 13: 1351–1377, 2001.

Baker S, Lemon R. Precise spatiotemporal repeating patterns in monkey primary and supplementary motor areas occur at chance levels. *J Neurophysiol* 84: 1770–1780, 2000.

Ben-Shaul Y, Bergman H, Ritov Y, Abeles M. Trial to trial variability in either stimulus or action causes apparent correlation and synchrony in neuronal activity. *J Neurosci Methods* 111: 99–110, 2001.

Bienenstock E. A model of neocortex. *Network* 6: 179–224, 1995.

Billimoria CP, DiCaprio RA, Birmingham JT, Abbott LF, Marder E. Neuromodulation of spike-timing precision in sensory neurons. *J Neurosci* 26: 5910–5919, 2006.

Brody C. Slow variations in neuronal resting potentials can lead to artefactually fast cross-correlations in the spike trains. *J Neurophysiol* 80: 3345–3351, 1998.

Brody C. Disambiguating different covariation types. *Neural Comput* 11: 1527–1535, 1999.

Butts D, Weng C, Jin J, Yeh CI, Lesica N, Alonso JM, Stanley GB. Temporal precision in the neural code and the timescales of natural vision. *Nature* 449: 92–96, 2007.

Churchland MM, Yu BM, Cunningham JP, Sugrue LP, Cohen MR, Corrado GS, Newsome WT, Clark AM, Hosseini P, Scott BB, Bradley DC, Smith MA, Kohn A, Movshon JA, Armstrong KM, Moore T, Chang SW, Snyder LH, Lisberger SG, Priebe NJ, Finn IM, Ferster D, Ryu SI, Santhanam G, Sahani M, Shenoy KV. Stimulus onset quenches neural variability: a widespread cortical phenomenon. *Nat Neurosci* 13: 369–378, 2010.

Cohen MR, Kohn A. Measuring and interpreting neuronal correlations. Measuring and interpreting neuronal correlations. *Nat Neurosci* 14: 811–819, 2011.

Constantinidis C, Williams G, Goldman-Rakic P. A role for inhibition in shaping the temporal flow of information in prefrontal cortex. *Nat Neurosci* 5: 175–180, 2002.

Czanner G, Eden U, Wirth S, Yanike M, Suzuki W, Brown E. Analysis of between-trial and within-trial neural spiking dynamics. *J Neurophysiol* 99: 2672–2693, 2008.

Dan Y, Alonso J, Usrey W, Reid R. Coding of visual information by precisely correlated spikes in the lateral geniculate nucleus. *Nat Neurosci* 1: 501–507, 1998.

Date A, Bienenstock E, Geman S. On the temporal resolution of neural activity. *Soc Neurosci Abstr* 25: 1411, 1999.

Dayan P, Abbott L. *Theoretical Neuroscience.* Cambridge, MA: MIT Press, 2001.

Dragoi G, Buzsáki G. Temporal encoding of place cells by hippocampal cell assemblies. *Neuron* 50: 145–157, 2006.

Ferragamo M, Haresign T, Simmons J. Frequency tuning, latencies, and responses to frequency-modulated sweeps in the inferior colliculus of the echolocating bat, *Eptesicus fuscus. J Comp Physiol A* 182: 65–79, 1998.

Frostig R, Frostig Z, Harper RM. Recurring discharge patterns in multiple spike trains: I detection. *Biol Cybern* 62: 487–493, 1990.

Fujisawa S, Amarasingham A, Harrison MT, Buzsáki G. Behavior-dependent short-term assembly dynamics in the medial prefrontal cortex. *Nat Neurosci* 11: 823–834, 2008.

Furukawa S, Middlebrooks J. Cortical representation of auditory space: information-bearing features of spike patterns. *J Neurophysiol* 87: 1749–1762, 2002.

Geman S. Invariance and selectivity in the ventral visual pathway. *J Physiol (Paris)* 100: 212–224, 2006.

Gerstein G. Searching for significance in spatio-temporal firing patterns. *Acta Neurobiol Exp (Warsz)* 64: 203–207, 2004.

Gray CM. The temporal correlation hypothesis of visual feature integration: still alive and well. *Neuron* 24: 31–47, 1999.

Grün S. Data-driven significance estimation for precise spike correlation. *J Neurophysiol* 101: 1126–1140, 2009.

Grün S, Diesmann M, Aertsen A. Unitary events in multiple single-neuron spiking activity: I. detection and significance. *Neural Comput* 14: 43–80, 2002a.

Grün S, Diesmann M, Aertsen A. Unitary events in multiple single- neuron spiking activity: II. nonstationary data. *Neural Comput* 14: 81–119, 2002b.

Grün S, Riehle A, Diesmann M. Effects of cross-trial nonstationarity on joint-spike events. *Biol Cybern* 88: 335–351, 2003.

Gütig R, Aertsen A, Rotter S. Statistical significance of coincident spikes: count-based versus rate-based statistics. *Neural Comput* 14: 121–153, 2002.

Harrison MT. *Discovering Compositional Structures* (dissertation). Providence, RI: Brown University, 2005.

Harrison MT, Geman S. A rate and history-preserving re-sampling algorithm for neural spike trains. *Neural Comput* 21: 1244–1258, 2009.

Hatsopoulos NG, Geman S, Amarasingham A, Bienenstock E. At what time scale does the nervous system operate? *Neurocomputing* 52: 25–29, 2003.

Hatsopoulos NG, Xu Q, Amit Y. Encoding of movement fragments in the motor cortex. *J Neurosci* 27: 5105–5114, 2007.

Hirabayashi T, Miyashita Y. Dynamically modulated spike correlation in monkey inferior temporal cortex depending on the feature configuration within a whole object. *J Neurosci* 25: 10299–10307, 2005.

Ikegaya Y, Aaron G, Cossart R, Aronov D, Lampl I, Ferster D, Yuste R. Synfire chains and cortical songs: temporal modules of cortical activity. *Science* 304: 559–564, 2004.

Ito H, Tsuji S. Model dependence in quantification of spike interdependence by joint peri-stimulus time histogram. *Neural Comput* 12: 195–217, 2000.

Jones L, Depireux D, Simons D, Keller A. Robust temporal coding in the trigeminal system. *Science* 304: 1986–1989, 2004.

Kass RE, Ventura V. Spike count correlation increases with length of time interval in the presence of trial-to-trial variation. *Neural Comput* 18: 2583–2591, 2006.

Lu T, Wang X. Information content of auditory cortical responses to time-varying acoustic stimuli. *J Neurophysiol* 91: 301–313, 2003.

Maldonado P, Babul C, Singer W, Rodriguez E, Berger D, Grün S. Synchronization of neuronal responses in primary visual cortex of monkeys viewing natural images. *J Neurophysiol* 100: 1523–1532, 2008.

Martignon L, von Hasseln H, Grün S, Aertsen A, Palm G. Detecting higher-order interactions among the spiking events in a group of neurons. *Biol Cybern* 73: 69–81, 1995.

Meister M, Barry MJ. The neural code of the retina. *Neuron* 22: 435–440, 1999.

Nádasdy Z, Hirase H, Czurkó A, Csicsvari J, Buzsáki G. Replay and time compression of recurring spike sequences in the hippocampus. *J Neurosci* 19: 9497–9507, 1999.

Oram M, Wiener M, Lestienne R, Richmond B. Stochastic nature of precisely timed spike patterns in visual system neural responses. *J Neurophysiol* 81: 3021–3033, 1999.

Pauluis Q, Baker SN. An accurate measure of the instantaneous discharge probability, with application to unitary joint-event analysis. *Neural Comput* 12: 647–669, 2000.

Pazienti A, Diesmann M, Grün S. Bounds on the ability to destroy precise coincidences by spike dithering. In: *Lecture Notes in Computer Science: Advances in Brain, Vision, and Artificial Intelligence.* Heidelberg, Germany: Springer, 2007, p. 428–437.

Pazienti A, Maldonado PE, Diesmann M, Grün S. Effectiveness of systematic spike dithering depends on the precision of cortical synchronization. *Brain Res* 1225: 39–46, 2008.

Perkel D, Gerstein G, Moore G. Neuronal spike trains and stochastic point processes. II. Simultaneous spike trains. *Biophys J* 7: 419–440, 1967.

Pipa G, Riehle A, Grün S. Validation of task-related excess of spike coincidences based on NeuroXidence. *Neurocomputing* 70: 2064–2068, 2007.

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

Renart A, de la Rocha J, Bartho P, Hollender L, Parga N, Reyes A, Harris KD. The asynchronous state in cortical circuits. *Science* 327: 587–590, 2010.

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

Riesenhuber M, Poggio T. Are cortical models really bound by the binding problem? *Neuron* 24: 87–93, 1999.

Roelfsema PR, Engel AK, Konig P, Singer W. The role of neuronal synchronization in response selection: a biologically plausible theory of structured representations in the visual cortex. *J Cogn Neurosci* 6: 603–625, 1996.

Rokem A, Watzl S, Gollisch T, Stemmler M, Herz A, Samengo I. Spike-timing precision underlies the coding efficiency of auditory receptor neurons. *J Neurophysiol* 95: 2541–2552, 2006.

Samonds JM, Zhou Z, Bernard MR, Bonds AB. Synchronous activity in cat visual cortex encodes collinear and cocircular contours. *J Neurophysiol* 95: 2602–2616, 2006.

Shadlen M, Movshon J. Synchrony unbound: a critical evaluation of the temporal binding hypothesis. *Neuron* 24: 67–77, 1999.

Shmiel T, Drori R, Shmiel O, Ben-Shaul Y, Nádasdy Z, Shemesh M, Teicher M, Abeles M. Neurons of the cerebral cortex exhibit precise interspike timing in correspondence to behavior. *Proc Natl Acad Sci USA* 102: 18655–18657, 2005.

Shmiel T, Drori R, Shmiel O, Ben-Shaul Y, Nádasdy Z, Shemesh M, Teicher M, Abeles M. Temporally precise cortical firing patterns are associated with distinct action segments. *J Neurophysiol* 96: 2645–2652, 2006.

Singer W. Neuronal synchrony: a versatile code for the definition of relations? *Neuron* 24: 49–65, 1999.

Smith MA, Kohn A. Spatial and temporal scales of correlation in primary visual cortex. *J Neurosci* 48: 12591–12603, 2008.

Stark E, Abeles M. Unbiased estimation of precise temporal correlations between spike trains. *J Neurosci Methods* 179: 90–100, 2009.

Tiesinga P, Fellous JM, Sejnowski T. Regulation of spike timing in visual cortical circuits. *Nat Rev Neurosci* 9: 97–109, 2008.

Ventura V. Testing for and estimating latency effects for Poisson and non-Poisson spike trains. *Neural Comput* 16: 2323–2349, 2004.

Ventura V, Cai C, Kass R. Statistical assessment of time-varying dependence between two neurons. *J Neurophysiol* 94: 2940–2947, 2005a.

Ventura V, Cai C, Kass R. Trial-to-trial variability and its effect on time-varying dependence between two neurons. *J Neurophysiol* 94: 2928–2939, 2005b.

Victor JD, Purpura JP. Nature and precision of temporal coding in visual cortex: a metric-space analysis. *J Neurophysiol* 76: 1310–1326, 1996.

Villa AE, Tetko IV, Hyland B, Najem A. Spatiotemporal activity patterns of rat cortical neurons predict responses in a conditioned task. *Proc Natl Acad Sci USA* 96: 1106–1111, 1999.

von der Malsburg C. The correlation theory of brain function. Reprinted in: *Models of Neural Networks II*, edited by Domany E, van Hemmen JL Schulten K. Berlin: Springer, 1994.

Wehr M, Laurent G. Odour encoding by temporal sequences of firing in oscillating neural assemblies. *Nature* 384: 162–166, 1996.

Yin TC, Chan JC. Interaural time sensitivity in medial superior olive of cat. *J Neurophysiol* 64: 465–488, 1990.