Statistical Identification of Synchronous Spiking

Matthew T. Harrison[1], Asohan Amarasingham[2], and Robert E. Kass[3]

Contents

[1] Division of Applied Mathematics, Brown University, Providence RI 02912 USA
[2] Department of Mathematics, City College of New York, New York NY 10031 USA
[3] Department of Statistics, Carnegie Mellon University, Pittsburgh PA 15213 USA

1 Introduction

In some parts of the nervous system, especially in the periphery, spike timing in response to a stimulus, or in production of muscle activity, is highly precise and reproducible. Elsewhere, neural spike trains may exhibit substantial variability when examined across repeated trials. There are many sources of the apparent variability in spike trains, ranging from subtle changes in experimental conditions to features of neural computation that are basic objects of scientific interest. When variation is large enough to cause potential confusion about apparent timing patterns, careful statistical analysis can be critically important. In this chapter we discuss statistical methods for analysis and interpretation of synchrony, by which we mean the approximate temporal alignment of spikes across two or more spike trains. Other kinds of timing patterns are also of interest [2, 57, 74, 69, 23], but synchrony plays a prominent role in the literature, and the principles that arise from consideration of synchrony can be applied in other contexts as well.

    The methods we describe all follow the general strategy of handling imperfect reproducibility by formalizing scientific questions in terms of statistical models, where relevant aspects of variation are described using probability. We aim not only to provide a succinct overview of useful techniques, but also to emphasize the importance of taking this fundamental first step, of connecting models with questions, which is sometimes overlooked by non-statisticians. More specifically, we emphasize that (i) detection of synchrony presumes a model of spiking without synchrony, in statistical jargon this is a null hypothesis, and (ii) quantification of the amount of synchrony requires a richer model of spiking that explicitly allows for synchrony, and in particular, permits statistical estimation of synchrony.

    Most simultaneously recorded spike trains will exhibit some *observed synchrony*. Figure 1 displays spike trains recorded across 120 trials from a pair of neurons in primary visual cortex, and circles indicate observed synchrony, which, in this case, is defined to be times at which both neurons fired within the same 5 ms time bin. Here, as is typical in cortex under many circumstances, the spike times are not highly reproducible across trials and are, instead, erratic. This variability introduces ambiguity in the neurophysiological interpretation of observed synchrony. On the one hand, observations of synchronous firing may reflect the existence of neuronal mechanisms, such as immediate common input or direct coupling, that induce precisely-timed, coordinated spiking. It is this potential relationship between observed synchrony and mechanisms of precise, coordinated spike timing that motivates this chapter. On the other hand, the apparent variability in spike timing could be consistent with mechanisms that induce only coarsely-timed spiking. For coarsely-timed spikes, with no mechanisms of precise timing, there will be time bins in which both neurons fire, leading to observed synchrony. In general, we might expect that spike timing is influenced by a variety of mechanisms across multiple time scales. Many of these influences, such as signals originating in the sensory system or changes in the extracellular environment, may affect multiple neurons without necessarily inducing precise, coordinated spiking. Each of these influences will affect the amount of observed

synchrony.

Observations of synchrony in the presence of apparent variability in spike timing thus raise natural questions. How much of the observed synchrony in Figure 1 is stimulus-related? How much comes from slow-wave oscillations? How much comes from mechanisms for creating precisely-timed, coordinated spiking? How much is from chance alignment of coarsely-timed spikes? Statistical approaches to such questions begin with the formulation of probabilistic models designed to separate the many influences on spike timing. Especially because of the scientific interest in mechanisms of precise spike timing, neurophysiologists often seek to separate the sources of observed synchrony into coarse temporal influences and fine temporal influences. Accordingly, many such statistical models are designed to create a distinction of time scale. In this chapter, we will discuss many of the modeling devices that have been used to create such distinctions.
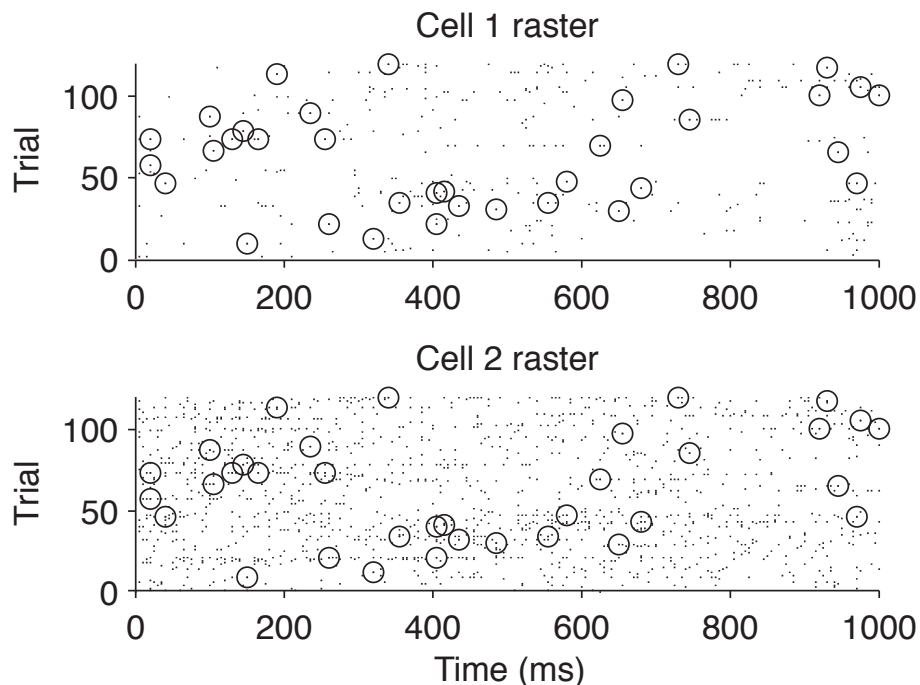
Figure 1: *Raster plots of spike trains on 120 trials from two simultaneously-recorded neurons, with synchronous spikes shown as circles. Here, observed synchrony is defined using time bins having 5 ms width. Spiking activity was recorded for 1 second from primary visual cortex in response to a drifting sinusoidal grating, as reported in [43] and [46].*

We wish to acknowledge that the basic statistical question of quantifying synchrony is only relevant to a subset of the many physiologically interesting questions about synchrony. The methods described here are designed to detect and measure precise synchrony created by the nervous system and can reveal much about the temporal precision of cortical dynamics, but they do not directly address the role of synchrony in cortical information processing. Cortical dynamics may create excess synchrony, above chance occurrence, even if this synchrony plays no apparent role in downstream processing. Conversely, downstream neurons might respond to

3

synchronous spikes as a proxy for simultaneously high firing rates even if the upstream neurons are not influenced by mechanisms that could create precise spike timing. Nevertheless, the existence of statistically detectable synchrony, beyond that explicable by otherwise apparent variation in firing patterns, not only reveals clues about cortical dynamics, but also forms a kind of *prima facie* evidence for the relevance of precise spike timing for cortical information processing.

2 Synchrony and time scale

In this section we discuss what we mean by *synchrony,* which we associate with transient changes in correlation on *sub-behavioral time scales.* Our main point is that the distinction between synchrony and other types of dependence relies primarily on distinctions of *time scale.*
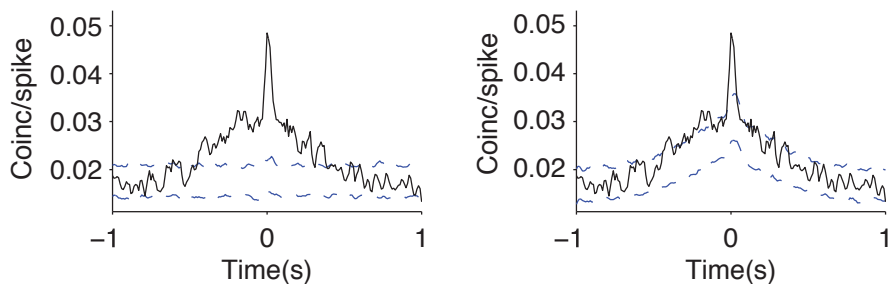

Figure 2: *Cross-correlation histograms from a pair of neurons recorded simultaneously from primary visual cortex, as described in [46]. The x-axis displays the time lag between the two spike trains. The CCH indicates correlations only at short lags, on the scale of the x-axis. (LEFT) Bands show acceptance region for a test of independence based on a point process model involving stimulus effects. (RIGHT) Bands show acceptance region for a test of independence based on a point process model involving both stimulus effects and slow-wave network effects, possibly due to anesthesia. (Figure courtesy of Ryan Kelly.)*

The cross-correlation histogram (CCH) is a popular graphical display of the way two neurons tend to fire in conjunction with each other. CCHs are described in more detail in Section 4.1, but, roughly speaking, the x-axis indicates the time lag (ℓ) of the synchrony and the y-axis indicates the number of observed lag-ℓ synchronous spikes. A peak (or trough) in the CCH provides evidence of correlation between the spike trains. Precisely-timed synchrony is usually identified with a narrow peak in the CCH. Figure 2 shows a CCH taken from a pair of neurons in primary visual cortex [46]. Its appearance, with a peak near 0, is typical of many CCHs that are used to demonstrate synchronous firing. While we presume readers to be familiar with CCHs, we wish to provide a few caveats and will review throughout Section 4 the ways CCHs have been modified to better identify synchrony.

The first thing to keep in mind when looking at a CCH is that its appearance, and therefore its biological interpretation, depends on the units of the x-axis. A 500 ms wide peak in the CCH carries a much different neurophysiological interpretation than a 5 ms wide peak, even though they would look the same if the x-axis were re-scaled. A broad peak on the scale of

4

hundreds of milliseconds can be explained by any number of processes operating on behavioral time scales, such as modulations in stimulus, response, metabolism, etc., many of which will be difficult to distinguish among. A peak of only a few milliseconds is indicative of processes operating on sub-behavioral time scales, perhaps revealing features of the micro-architecture of the nervous system. When we use the term "synchrony" we are referring to correlated spiking activity within these much shorter time windows.

To illustrate the concept of time scale, the left column of Figure 3 shows several artificial CCHs that are created by mathematically identical processes. The only difference is the temporal units assigned to a statistical coupling between two spike trains. The expected shape of each CCH can be made to look like the others simply by zooming in or out appropriately. If these were real CCHs, however, the biological explanations would be quite different. Only the fourth CCH (in the left column), with a narrow 1 ms wide peak would almost universally be described as an example of synchrony in the neurophysiology literature. The third CCH with a 10 ms wide peak is more ambiguous and investigators may differ on their willingness to identify such a peak as synchrony. The other two, with widths of 100 ms or more, would typically not be identified as synchrony.

The CCHs in Figure 3 are unrealistic only in their simplicity. The complexity of true spike trains only further complicates issues. A real CCH will show correlations from many processes operating on a variety of time scales. Some of these correlations may be synchrony. Statistical models that distinguish synchrony from other types of correlation will necessarily involve determinations of time scale.

3   Spike trains and firing rate

Statistical models of synchrony are embedded in statistical models of spike trains. In this section we review the basic probabilistic description of a spike train as a *point process*, and we draw analogies between the conditional intensities of a point process and the conceptual notion of a neuron's instantaneous firing rate. Our main point is that there are many different ways to define the conditional intensity of a point process and there are correspondingly many ways to define firing rates. As such, when discussing firing rates it is always important to be clear about terminology and modeling assumptions. We assume that readers are familiar with the basic objects of probability and statistics, but see the Appendix for a brief summary.

Since the work of Adrian [3] and Hartline [35], the most basic functional description of a neuron's activity has been its *observed firing rate*. A simple definition of observed firing rate at time t is the spike count per unit time in an interval containing t. The observed firing rate depends heavily on the length of the interval used in its definition. For long intervals, the observed firing rate is forced to be slowly-varying in time. For short intervals, the observed firing rate can be quickly-varying, but tends to fluctuate wildly.[4] This dependence on interval length complicates the use of observed firing rate, especially if one wishes to discuss quickly-varying firing rates.

For statistical purposes it is convenient to conceptualize spike trains as random, with the

---

[4]For tiny intervals of length δ, the observed firing rate jumps between zero (no spike in the interval) and $1/\delta$ (a single spike in the interval), depending more on the length of the interval than on any sensible notion of the rate of spiking.
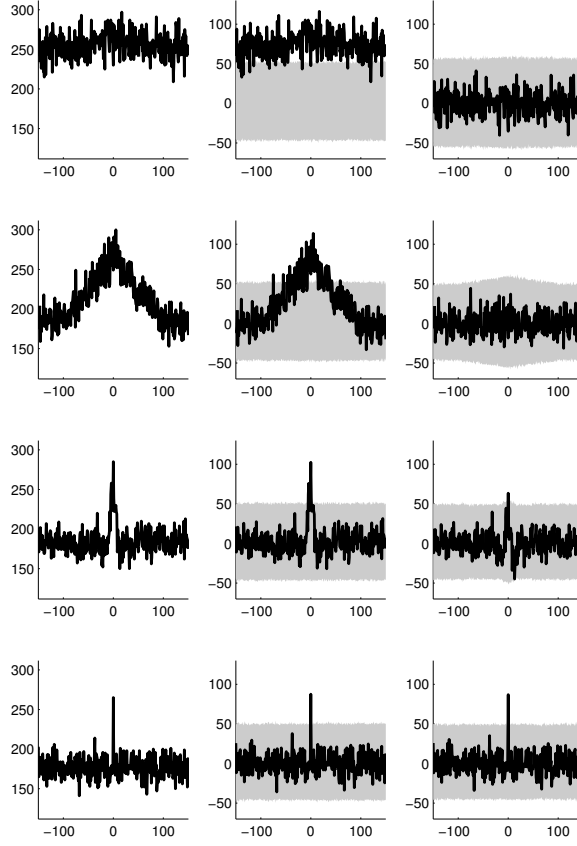
Figure 3: *CCHs for four simulations with correlations on different time scales. Each row corresponds to a different simulation. Each column corresponds to a different type of CCH. The x-axis is lag in milliseconds. The y-axis is observed synchrony count (1 ms bins). For each simulation, each of the two spike trains could be in either a high-firing-rate or low-firing-rate state at each instant in time. The two spike trains shared the same state and the state occasionally flipped. This shared (and changing) state creates a correlation between the two spike trains. The rate of flipping is what changes across these figures. In each case the state has a 50% chance of flipping every w ms, where $w = 1000, 100, 10, 1$ from top to bottom, respectively. (So $w = 1$ ms corresponds to "fine" temporal coupling, and $w = 1000$ ms corresponds to "coarse" temporal coupling. Of course, the distinction between "fine" and "coarse" is qualitative, not quantitative.)* Left column: *Raw CCHs.* Middle column: *Uniform-corrected CCHs (black line; see Section 4.3.6) and 95% simultaneous acceptance regions (gray region; see Section 4.3.4). Except for the units of the y-axis and negligible (in this case) edge effects, the uniform-corrected CCHs and their acceptance regions are essentially identical to using the correlation coefficient (Pearson's r between the lagged spike trains viewed as binary time series with 1 ms bins) and testing the null hypothesis of independence.* Right column: *20 ms jitter-corrected CCHs (see Section 4.5.2) with 95% simultaneous acceptance regions.*

observed rate of spiking fluctuating randomly around some theoretical time-varying *instantaneous firing rate*. This conceptualization decomposes the observed firing rate into a signal component, namely, the instantaneous firing rate, and a noise component, namely, the fluctuations in spike times. It is important to note that the statistical distinction between signal and noise is not necessarily a statement about physical randomness, but rather a useful modeling

device for decomposing fluctuations in the observed firing rate. Presumably, with sufficiently detailed measurements about the various sources of neural spiking, one could deterministically predict the timing of all spikes so that the noise becomes negligible and the instantaneous firing rate begins to coincide with the observed firing rate in tiny intervals. Since these detailed measurements are not available, it is convenient to model their effects with some notion of randomness. As this extreme case illustrates, instantaneous firing rates do not correspond to physical reality, but reflect a particular way of separating signal and noise into a form that is, hopefully, convenient for a particular scientific or statistical investigation.

3.1 Point processes, conditional intensities, and firing rates

The mathematical theory of point processes [18] is the natural formalism for the statistical modeling of spike trains. Suppose we observe n spikes (action potentials) on a single trial and we use $s_1, s_2, \ldots, s_n$ to denote the times at which the spikes occur (a spike train). Also, let us take the beginning of the observation period to be at time 0, and define $s_0 = 0$. Then $x_i = s_i - s_{i-1}$ is the ith interspike interval, for $i = 1, \ldots, n$. Given a set of positive interspike interval values $x_1, x_2, \ldots, x_n$ we may reverse the process and find $s_j = \sum_{i=1}^{j} x_i$ to be the jth spike time, for $j = 1, \ldots, n$. Similarly, in probability theory, if we have a sequence of positive random variables $X_1, X_2, \ldots$ the sequence of random variables $S_1, S_2, \ldots$ defined by $S_j = \sum_{i=1}^{j} X_i$ forms a *point process*. It is thus a natural theoretical counterpart of an observed spike train. Data are recorded with a fixed timing accuracy, often corresponding with a sampling rate such as 1 KHz. Spike trains recorded with an accuracy of 1 ms can be converted to a sequence of 0s and 1s, where a 1 is recorded whenever a spike occurs and a 0 is recorded otherwise. In other words, a spike train may be represented as a binary time series. Within the world of statistical theory, every point process may be considered, approximately, to be a binary time series.

Another way of representing both spike trains and their theoretical counterparts is in terms of counts. Let us use $N(t)$ to denote the total number of spikes occurring up to and including time t. Thus, for $s < t$, the spike count between time s and time t, up to and including time t, is $N(t) - N(s)$. The expected number of spikes per unit time is called the *intensity* of the point process. If the point process is time-invariant, it is called stationary or *homogeneous*, and the intensity is constant, say $\lambda \geq 0$. Otherwise, the point process is non-stationary or *inhomogeneous*, and the intensity can be a function of time, say $\lambda(t)$. The intensity can be conceptualized as the instantaneous probability (density) of observing a spike at time t, suggestively written as

$$P\big(\text{spike in}(t, t + dt]\big) = \lambda(t)dt, \tag{1}$$

where dt denotes an infinitesimal length of time.

At first glance, the intensity of a point process feels exactly like the notion of a theoretical time-varying instantaneous firing rate that we were looking for, and it is tempting to equate the two concepts. But in many cases the basic intensity does not match our intuitive notion of an instantaneous firing rate. Consider two experimental conditions in which a neuron has a consistently larger spike count in one condition than the other. It seems natural to allow the instantaneous firing rate to depend on experimental condition, so we need a notion of an intensity that can depend on experimental condition. Recall that biological spike trains have refractory periods so that a new spike cannot occur ("absolute refractory period") or is less likely

to occur ("relative refractory period") during a brief (millisecond-scale) interval immediately following a spike. If a spike just occurred, it might be useful in some contexts to think about the instantaneous firing rate as being reduced for a short time following a spike, so we need a notion of an intensity that can depend on the history of the process. Finally, we might imagine that the baseline excitability of a neuron is modulated by some mechanism, creating a kind of gain-modulation [61], that is not directly observable. If we want to model this mechanism as having an effect on the instantaneous firing rates, then we also need to consider intensities that depend on unobservable variables.

The intensity can be generalized to accommodate each of the above considerations by allowing it to depend on variables other than time. This more general notion of intensity is often called a *conditional intensity* because it is defined using conditional probabilities. It is the notion of a conditional intensity that corresponds most closely with the intuitive concept of a theoretical instantaneous firing rate. For example, if Z_t is some additional information, like experimental condition or attentional state (or both), then we could write the conditional intensity given Z_t as

$$P\big(\text{spike in}(t, t + dt]\big|Z_t\big) = \lambda(t|Z_t)dt. \tag{2}$$

Now the conditional intensity, and hence, the instantaneous firing rate, can depend on Z_t. Z_t can have both observed components, O_t, and unobserved components, U_t, so that $Z_t = (O_t, U_t)$. These components can be fixed or random, and they may or may not be time varying. When Z_t has random components, the conditional intensity given Z_t is random, too. Statistical models often draw distinctions between observed and unobserved variables and also between non-random and random variables. It is common to include in Z_t the internal history H_t of spike times prior to time t. Specifically, if prior to time t there are n spikes at times $s_1, s_2, \ldots, s_n$, we write $H_t = (s_1, s_2, \ldots, s_n)$. Including H_t as one of the components of Z_t lets the conditional intensity accommodate refractory periods, bursting, and more general spike-history effects. By also including other variables, the instantaneous firing rates can further accommodate stimulus effects, network effects, and various sources of observed or unobserved gain modulation.

The conditional intensity given only the internal history, i.e., $Z_t = H_t$, is special because it completely characterizes the entire distribution of spiking. There are no additional modeling assumptions once the conditional intensity is specified. Many authors reserve the term conditional intensity for the special case of conditioning exclusively on the internal history and use terms like the Z_t-intensity or the intensity with respect to Z_t for the more general case. We will use the term conditional intensity in the more general sense, because it emphasizes our main point: conditional intensities, and hence instantaneous firing rates, necessitate the specification of exactly what information is being conditioned upon.

The simplest of all point processes is the Poisson process. It has the properties that (i) for every pair of time values s and t, with $s < t$, the random variable $N(t) - N(s)$ is Poisson distributed and (ii) for all non-overlapping time intervals the corresponding spike count random variables are independent. Poisson processes have the special *memoryless* property that their conditional intensity given the internal history does not depend on the internal history, but is the same as the unconditional intensity: $\lambda(t|H_t) = \lambda(t)$. They cannot exhibit refractory periods, bursting, or any other spike history effects that do not come from the intensity. And the intensity cannot depend on unobserved, random variables. Homogeneous Poisson processes are time-invariant in the sense that the probability distribution of a count $N(t) - N(s)$ depends

only on the length of the time interval $t - s$. Its intensity is constant, say, $\lambda \geq 0$, and its entire distribution is summarized by this single parameter λ.

Observable departures from Poisson spiking behavior have been well documented (e.g., [70, 6, 63]). On the other hand, such departures sometimes have relatively little effect on statistical procedures, and Poisson processes play a prominent role in theoretical neuroscience and in the statistical analysis of spiking data. Certain types of non-Poisson processes can be built out Poisson processes by using random variables and the appropriate conditional intensities. For example, if U is a random variable and we model the conditional distribution given U as a Poisson process, i.e., $\lambda(t|H_t, U) = \lambda(t|U)$, then the resulting unconditional process is called a *Cox process* (or doubly stochastic Poisson process). It can be viewed as a Poisson process with a random instantaneous firing rate. Alternative analyses, not relying on Poisson processes at all, typically proceed by explicitly modeling $\lambda(t|H_t)$, thus capturing the (non-Poisson) history-dependence of a spike train.

3.2   Models of conditional intensities

For statistical purposes, it is not enough to specify which type of conditional intensity (i.e., which type of conditioning) we are equating with the notion of a firing rate, but we must also appropriately restrict the functional form of the conditional intensity (i.e., the allowable "shapes" of the function λ). If the functional form is fully general, we could fit any data set perfectly without disambiguating the potential contributions to observed synchrony.[5] There are three common ways to restrict the functional form of conditional intensities: identical trials, temporal smoothness, and temporal covariates.

For identical trials models, we assume that our conditional intensity is identical across trials. This relates the value of the conditional intensity across multiple points in time. Another way to think about identical trials is that time, i.e., the t in $\lambda(t)$, refers to trial-relative time, and that our data consists of many independent observations from a common point process. The identical trials perspective and the use of the (unconditional) intensity, or perhaps a conditional intensity given experimental condition, are what underlies the interpretation of the peri-stimulus time histogram (PSTH)[6] as an estimator of the firing rate. This perspective is also at the heart of the well-known shuffle-correction procedure for the CCH and for other graphical displays of dependence.

For temporal smoothness models, we assume that our conditional intensity is slowly-varying in time in some specific way. Severe restrictions on temporal smoothness, such as assumptions of stationarity or homogeneity, permit the aggregation of statistical information across large intervals of time. But even mild restrictions on temporal smoothness permit the aggregation of

---

[5]The problem is particularly easy to see for temporally discretized spike trains that can be viewed as binary time series. In this case, the intensities specify probabilities of observing a spike in each time bin and the probabilities can vary arbitrarily across time bins and across spike trains. We can choose an intensity that assigns probability one for each time bin with a spike and probability zero for each time bin without a spike. This intensity (i.e., the firing rate) perfectly accounts for the data, and we do not need additional notions, like synchrony, to explain its features. In statistical parlance we face a problem of *non-identifiability*: models with and without explicit notions of synchrony can lead to the same probability specification for the observed data.

[6]The PSTH is simply a histogram of all trial-relative spike times from a single neuron, perhaps normalized to some convenient units like spikes per unit time.

certain types of statistical information, particularly the signatures of precise timing [7]. This perspective underlies the jitter approach to synchrony identification.

For temporal covariates models, we assume that our conditional intensity depends on time only through certain time-varying, observable covariates, such as a time-varying stimulus or behavior, say Z_t. Models of hippocampal place cells, directionally tuned motor-cortical cells, and orientation-specific visual-cortical cells often take this form. Symbolically, we might write $\lambda(t|Z_t) = g(Z_t)$ for some function g. When Z_t repeats, we can accumulate statistical information about λ. Or, if g has some assumed parametric form, such as cosine tuning to an angular covariate, then every interval conveys statistical information about our conditional intensity.

In general, statistical models might combine these principles for modeling conditional intensities, or perhaps, incorporate restrictions that do not fit cleanly into these three categories. The generalized linear model approach that we discuss in Section 4.6 allows all of these modeling assumptions to be included in a common framework. We discuss several different types of models below from the perspective of synchrony detection. Many of these models involve point processes and instantaneous firing rates, i.e., conditional intensities, but they do not all use the same type of point process nor the same definition of firing rate. When interpreting a statistical model for spike trains it is crucial to understand how the model and the terminology map onto neurophysiological concepts. In particular, the term "firing rate" often refers to different things in different models.

4   Models for coarse temporal dependence

As we discussed in the Introduction, the statistical identification of synchrony requires a separation of the fine temporal influences on spike timing from the many coarse temporal influences on spike timing. In this section we focus primarily on models that only allow coarse temporal influences on spike timing. Besides being useful building blocks for more general models, these coarse temporal spiking models are actually quite useful from the statistical perspective of *hypothesis testing*. For hypothesis testing, we begin with a model, the *null hypothesis* or null model, that allows only coarse temporal spiking and ask if the observed data are consistent with the model. A negative conclusion, i.e., a rejection of the null hypothesis, is interpreted as evidence in favor of fine temporal spiking dynamics that create excess observed synchrony, although, strictly speaking, it is really only evidence that the null model is not an exact statistical description of the data. (The Appendix contains a brief overview of the principles of hypothesis testing.) Our main point in this section is that some models are more appropriate than others for describing the coarse temporal influences on spike timing.

In order to simplify the discussion and to provide easy visual displays of the key concepts we restrict our focus to the case of two spike trains and we concentrate primarily on the CCH as our way of measuring and visualizing raw dependence — for hypothesis testing, this means that *test statistics* are derived from the CCH. Section 4.7 discusses how to relax each of these restrictions.

4.1 Cross-correlation histogram (CCH)

We begin with a more precise description of the cross-correlation histogram (CCH). Suppose we have two spike trains with spike times $s_1, \ldots, s_m$ and $t_1, \ldots, t_n$, respectively. There are mn different *pairs* of spike times (s_i, t_j) with the first spike taken from spike train 1 and the second from spike train 2. Each of these pairs of spike times has a lag $\ell_{ij} = t_j - s_i$, which is just the difference in spike times between the second and first spike. These mn lags are the data that are used to create a CCH. Here, a CCH is simply a histogram of all of these lags.

There are many choices to make when constructing histograms. One of the most important is the choice of bin-width, or more generally, the choice of how to smooth the histogram. Figure 4 shows several CCHs constructed from a single collection of lags, but using different amounts of smoothing and showing different ranges of lags. By varying the degree of smoothing, one can choose to emphasize different aspects of the correlation structure. Only one of the smoothing choices reveals a narrow peak in the CCH that is suggestive of zero-lag synchrony (row 3, column 3). As there are entire subfields within statistics devoted to the art of smoothing histograms, we will not delve further into this topic, but caution that visual inspection of CCHs can be quite misleading, not in small part because of choices in smoothing.
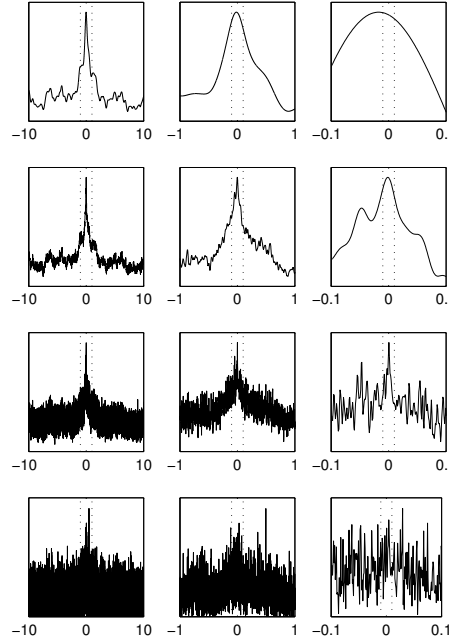
Figure 4: *CCHs for two simultaneously recorded neurons in monkey motor cortex (courtesy of Nicholas Hatsopoulos). Each column of CCHs shows a different range of lags, as indicated on the x-axis (in seconds). Each row of CCHs shows a different amount of smoothing. For the first two columns, the vertical dotted lines in each CCH indicate the region shown in the CCH to its immediate right. For the final column, the vertical dotted lines indicate ± 10 ms. Peaks within this final region correspond to the time scales usually associated with synchrony.*

Another practical detail when dealing with CCHs is how one deals with *edge effects*. Edge effects refer to the fact that, for a fixed and finite observation window, the amount of the observation window that can contain a spike-pair varies systematically with the lag of the

spike-pair. For example, if a spike-pair has a lag that is almost as long as the observation window, then one of the spikes must be near the beginning of the window and the other must be near the end. On the other hand, if the lag is near zero, then the spikes can occur anywhere in the observation window (as long as they occur close to each other). This creates the annoying feature that a raw CCH will typically taper to zero for very long lags, not because of any drop in correlation, but rather because of a lack of data.[7]

We view the CCH as a *descriptive statistic*, a simple summary of the data. Only later, when we begin to introduce various models will the CCH take on more nuanced interpretations. These modeling choices will affect such interpretations greatly. To illustrate the distinction, suppose we have finely discretized time so that the theoretical spike trains can be represented as binary time series, say $(U_k : k \in \mathbb{Z})$ and $(V_k : k \in \mathbb{Z})$, where, for example, U_k indicates spike or no spike in the first spike train in time bin k. ($\mathbb{Z}$ is the set of integers.) Suppose that we model each of the time series to be time-invariant or *stationary*.[8] Then the expected value of the product of U_k and $V_{k+\ell}$ depends only on the lag ℓ, and we can define the expected cross-correlation function (using signal processing terminology) as

$$\gamma(\ell) = E[U_k V_{k+\ell}] = E[U_0 V_\ell]$$

It is straightforward to verify that a properly normalized CCH[9] is an unbiased estimator of the expected cross-correlation function, γ. As such, the CCH takes on additional interpretations under the model of stationarity, namely, its interpretation as a (suitable) *estimator* of a theoretical model-based quantity. Note that the terminology involving cross-correlation functions is not standardized across fields. In statistics, for example, the cross-correlation function is usually defined as the Pearson correlation

$$\rho(\ell) = Cor(U_k, V_{k+\ell}) = \frac{\gamma(\ell) - E[U_k]E[V_{k+\ell}]}{\sqrt{Var[U_k]Var[V_{k+\ell}]}} = \frac{\gamma(\ell) - E[U_0]E[V_0]}{\sqrt{Var[U_0]Var[V_0]}}$$

which is a distribution-dependent shifting and scaling of γ (meaning that γ and ρ have the same shape as a function of ℓ). After appropriate standardization, the CCH can also be converted into a sensible estimator of ρ under the stationary model. In the following, however, we interpret the (raw) CCH only as a descriptive statistic without additional modeling assumptions, like stationarity.

4.2   Statistical hypothesis testing

The data for the CCHs in Figure 4 came from two simultaneously recorded neurons on different electrodes in monkey motor cortex (courtesy of Nicholas Hatsopoulos; see [37] for additional

---

[7] For a contiguous observation window of length T, a common solution is to divide the CCH at lag ℓ by $T - |\ell|$, which is the amount of the observation window available to spike-pairs having lag ℓ. Many authors would include this in their definition of the CCH.

[8] $(U_k : k \in \mathbb{Z})$ is stationary if the distribution of the vector of random variables $(U_k, U_{k+1}, \ldots, U_{k+m})$ is the same as the distribution of $(U_j, U_{j+1}, \ldots, U_{j+m})$ for all j, k, m. In other words, the process looks statistically identical in every observation interval.

[9] For data collected in a continuous observation window with T time bins and defining the CCH as a standard histogram with bin widths that match the discretization of time, we need to divide the CCH at lag ℓ by $(T - |\ell|)(2T - 1)$. This is the edge correction of Footnote 7 with an additional term to match the units of γ.

details). Like many real CCHs, these would seem to show an amalgamation of correlations on different time scales, perhaps from different processes. The broad (± 500 ms) peak is easily attributable to movement-related correlations — like many neurons in motor cortex, these neurons have coarse temporal observed firing rates that modulate with movement. Smaller fluctuations on top of this broad correlation structure might be attributable to noise or perhaps finer time scale processes. The CCH alone can not resolve these possibilities. Careful statistical analysis is needed.

In order to test whether certain features of a CCH are consistent with coarse temporal spiking, we first need to create a null hypothesis, i.e., a statistical model that allows for coarse temporal influences on spike timing. Our emphasis in this chapter is on this first step: the choice of an appropriate null model. Next we need to characterize the typical behavior of the CCH under the chosen null model. This step involves a great many subtleties, because a null hypothesis rarely specifies a unique (null) distribution for the data, but rather a family of distributions. We only scratch the surface in our treatment of this second step, briefly discussing bootstrap approximation and conditional inference as possible techniques. Finally, we need to quantify how well the observed CCH reflects this typical behavior, often with a p-value. Here we use graphical displays showing bands of typical fluctuation around the observed CCH.

4.3 Independent homogeneous Poisson process (HPP) model

For pedagogical purposes, we begin with an overly simple model to illustrate the main ideas and introduce the terminology that we use throughout the chapter. Suppose we model two spike trains, say A and B, as independent homogeneous Poisson processes (HPP). In this case, it does not matter whether we think about repeated trials, or a single trial, because the intensities are constant and all disjoint intervals are independent. The distribution of each spike train is specified by a single parameter, the (constant) intensity, say λ^A and λ^B, respectively. Our model is that the spike trains only "interact" (with each other and with themselves) through the specification of these two constants. Once the constants are specified, there is no additional interaction because the spike trains are independent and Poisson.

Using the independent HPP model as a null hypothesis is perhaps the simplest way to formalize questions like, "Is there more synchrony than expected by chance?" or "Is there more synchrony than expected from the firing rates alone?" In order to be precise, these questions presume models that define "chance" and "firing rates" and "alone." The independent HPP model defines the firing rate to be the unconditional and constant intensity of a homogeneous Poisson process. Chance and alone refer to the assumption that the spike trains are statistically *independent*: there are no other shared influences on spike timing. In particular, there are no mechanisms to create precise spike timing.

We caution the reader that concepts like "chance" and "independence", which appear frequently throughout the chapter (and in the neurophysiology literature), can mean different things in the context of different models. For example, independence often means *conditional* independence given the instantaneous firing rates. But as we have repeatedly emphasized, there is great flexibility in how one chooses to define the instantaneous firing rates. This choice about firing rates then affects all subsequent concepts, like independence, that are defined in relation

to the firing rates.[10]

The independent HPP model fails to account for many biologically plausible influences on coarse temporal spiking. For example, it fails to account for systematic modulations in observed firing rate following the presentation of a repeated stimulus, which would require models with time-varying firing rates. If the independent HPP model were used to detect fast-temporal synchrony, then we could not be sure that a successful detection (i.e., a rejection of the independent HPP null hypothesis) was a result of synchrony or merely a result of coarse temporal spiking dynamics that are not captured by the independent HPP model. Despite these short-comings, the independent HPP model is pedagogically useful for understanding more complicated models.

4.3.1 Bootstrap approximation

The first problem that one encounters when trying to use the independent HPP model is that the firing rates λ^A and λ^B are unknown. The independent HPP model does not specify a unique distribution for the data, but rather a parameterized family of distributions. In particular, the model does not uniquely specify the typical variation in the CCH. The typical variation will depend on the unknown firing rate parameters λ^A and λ^B. In order to use the model, it is convenient to first remove this ambiguity about the unknown firing rates.

One possible approach to removing ambiguity is to replace the unknown λ^A and λ^B with approximations derived from the observed data, say $\lambda^A \approx \widehat{\lambda}^A$ and $\lambda^B \approx \widehat{\lambda}^B$, where each $\widehat{\lambda}$ is some sensible estimator of the corresponding parameter λ, such as the total spike count for that spike train divided by the duration of the experiment (the maximum likelihood estimate). Replacing unknown quantities with their estimates is the first and most important approximation in a technique called *bootstrap*; see Section 4.3.3. Consequently, we refer to this method of approximation as a *bootstrap approximation*. (The terminology associated with bootstrap is not used consistently in the literature.)

When using a bootstrap approximation, it is important to understand the method of estimation, its accuracy in the situation of interest, and the robustness of any statistical and scientific conclusions to estimation errors. For the independent HPP model, where we usually have a lot of data per parameter, this estimation step is unlikely to introduce severe errors. For more complicated models with large numbers of parameters, the estimation errors are likely to be more pronounced and using a bootstrap approximation is much more delicate and perhaps inappropriate.

Under the null model, after we have replaced the unknown parameters with their estimated values, we are left with a single (null) distribution for the data: independent homogeneous Poisson processes with known intensities $\widehat{\lambda}^A$ and $\widehat{\lambda}^B$, respectively.

---

[10]This ambiguity of terminology is not peculiar to spike trains. Probabilities (and associated concepts, like independence and mean) are always defined with respect to some frame of reference, which must be clearly communicated in order for probabilistic statements to make sense. Another way to say this is that all probabilities are *conditional* probabilities, and we must understand the conditioning event in order to understand the probabilities [40, p.15].

4.3.2 Monte Carlo approximation

After reducing the null hypothesis to a single null distribution, it is straightforward to describe the null variability in the CCH. Perhaps the simplest way to proceed, both conceptually and algorithmically, is to generate a sample of many independent Monte Carlo pseudo-datasets from the null distribution (over pairs of spike trains) and compute a CCH for each one. This collection of pseudo-data CCHs can be used to approximate the typical variability under the null distribution in a variety of ways. Some examples are given in Sections 4.3.4 and 4.3.7.

Monte Carlo approximation is ubiquitous in this review and in the literature. We want to emphasize, however, that Monte Carlo approximation is rarely an integral component of any method. It is simply a convenient method of approximation, and is becoming increasingly convenient as the cost of intensive computation steadily decreases. Even for prototypical resampling methods like bootstrap, trial-shuffling, permutation tests, jitter, and so on, the explicit act of resampling is merely Monte Carlo approximation. It is done for convenience and is not an important step for understanding whether the method is appropriate in a particular scientific or statistical context. In many simple cases, one can derive analytic expressions of null variability, avoiding Monte Carlo approximation entirely.

4.3.3 Bootstrap

A bootstrap approximation (Section 4.3.1) followed by Monte Carlo approximation (Section 4.3.2) is usually called *bootstrap* [20]. In the context here it is *bootstrap hypothesis testing*, because the bootstrap approximation respects the modeling assumptions of the null hypothesis. For the case of the independent HPP model described above, it is *parametric* bootstrap hypothesis testing, because the bootstrap approximation replaces a finite number of parameters with their estimates. For understanding the behavior of bootstrap it is the bootstrap approximation in Section 4.3.1 that is important. The Appendix contains a more detailed example describing bootstrap hypothesis testing.

4.3.4 Acceptance bands

We can summarize the relationship between the observed CCH and a Monte Carlo sample of pseudo-data CCHs in many ways. Eventually, one should reduce the comparison to an acceptable statistical format, such as a p-value, or perhaps a collection of p-values (maybe one for each lag in the CCH). See [7, 72] for some specific suggestions. For graphical displays, one can represent the null variability as *acceptance bands* around the observed CCH. If these bands are constructed so that 95% of the pseudo-data CCHs fall completely (for all lags) within the bands, then they are 95% *simultaneous* (or global) acceptance bands. If the observed CCH falls outside of the simultaneous bands at even a single lag, then we can reject the null hypothesis (technically, the null distribution, as identified by the estimated parameters from the bootstrap approximation) at level 5%. Edge effects and smoothing choices are automatically accommodated, since the pseudo-data CCHs have all of the same artifacts and smoothness. Alternatively, if these bands are constructed so that, at each time lag, 95% of the pseudo-data CCHs are within the bands at the time lag, then they are 95% *pointwise* acceptance bands. If we fix a specific lag, say lag zero, *before collecting the data*, and if the observed CCH falls outside of the pointwise bands at that specific lag, then we can reject the null at level

5%. The consideration of many lags, however, corresponds to testing the same null hypothesis using many different test-statistics, the test-statistics being different bins in the CCH. This is a multiple-testing problem and a proper statistical accounting involves a recognition of the fact that, even if the null hypothesis were true, we might expect to see lags where the CCH falls outside of the pointwise bands simply because we are examining so many time lags.

4.3.5 Conditional inference

For the independent HPP model, a bootstrap approximation is not the only tool available for removing the ambiguity of the unknown parameters λ^A and λ^B. If we condition on the pair of total observed spike counts for the two spike trains, say N^A and N^B, respectively, then the conditional distribution of the spike times given these spike counts no longer depends on λ^A and λ^B. In particular, the null conditional distribution is uniform, meaning that all possible arrangements of N^A spikes in spike train A and N^B spikes in spike train B are equally likely. We can use this uniquely specified null conditional distribution to describe the null variability in the CCH. This is called *conditional inference*.

As with bootstrap, once we have reduced the null hypothesis to a single distribution, Monte Carlo approximation is a particularly convenient computational tool. In particular, we can generate a Monte Carlo sample of pseudo-data (by independently and uniformly choosing the N^A and N^B spike times within their respective observation windows), construct from this a collection of pseudo-data CCHs, and then proceed exactly as before with p-values and acceptance bands. We call this *Monte Carlo conditional inference*. In this particular example, the only difference between the samples created by bootstrap and those created by Monte Carlo conditional inference is that the bootstrap samples have variable spike counts (and this variability is reflected in the resulting pseudo-data CCHs), whereas the conditional inference samples do not.

4.3.6 Uniform model and conditional modeling

Conditional inference avoids a bootstrap approximation, which can make a critical difference in settings where the approximation is likely to be poor, usually because of an unfavorable ratio of the number of parameters to the amount of data. It is also completely insensitive to the distribution of the conditioning statistic, in this case, the pair of total spike counts. Depending on the problem, this insensitivity may or may not be desirable. For the current example, the null hypothesis specifies that the joint distribution of total spike counts (N^A, N^B) is that of two independent Poisson random variables. By conditioning on these spike counts, this particular distributional assumption is effectively removed from the null hypothesis. In essence, we have enlarged the null hypothesis. The new null hypothesis is that the conditional distribution of spike times is uniform given the total spike counts. Independent homogeneous Poisson processes have this property, as do many other point processes. We call this enlarged model the *uniform model*. The uniform model is defined by restricting our modeling assumptions to certain conditional distributions, an approach that we call *conditional modeling*.

For the case of synchrony detection, enlarging the null hypothesis to include any distribution on total spike counts seems beneficial, because it allows for more general types of coarse temporal gain modulation in the null hypothesis without also including processes with fine temporal

modulations. If one was using the independent HPP model for purposes other than synchrony detection, this enlarging of the null hypothesis might be undesirable. Nevertheless, even using the enlarged null hypothesis afforded by conditional inference, the independent HPP model is rarely appropriate for the statistical detection of synchrony, because, as we mentioned before, it does not allow for any time-varying interactions, either within or across spike trains.

4.3.7   Model-based correction

Averaging all of the pseudo-data CCHs (whether from bootstrap or Monte Carlo conditional inference) gives (a Monte Carlo approximation of) the expected null CCH. For graphical displays of the CCH it can be visually helpful to subtract the expected null CCH from the observed CCH (and from any acceptance bands).[11] This new CCH is frequently called a *corrected CCH*. Since the correction depends on the null model and also the procedure (such as a bootstrap approximation or conditional inference) used to reduce the null model to a single null distribution, there are many types of corrected CCHs. Corrected CCHs are useful because, under the null hypothesis, they have mean zero for all lags. (Edge effects and smoothing procedures are automatically handled, similar to acceptance bands.) Any deviations from zero of the corrected CCH are a result of either noise or a violation of the assumptions of the null distribution.

The middle column of Figure 3 shows *uniform-corrected CCHs*, i.e., model corrected CCHs using the uniform model with conditional inference to specify a null distribution, along with the associated simultaneous acceptance bands. (Using the independent HPP model with bootstrap creates almost identical figures — the only difference being slightly wider acceptance bands that account for the additional variability in spike counts.) Uniform-correction essentially subtracts a constant from the CCH,[12] so it cannot isolate synchrony time-scale dependencies. From Figure 3 we see that the CCH exceeds the acceptance bands, signaling a rejection of the null hypothesis of independence, regardless of the width of the CCH peak.

4.4   Identical trials models

Many neurophysiological experiments with repeated trials exhibit spike trains whose observed firing rates vary systematically during the trials, as evidenced by peri-stimulus time histograms (PSTHs) with clearly non-constant shapes. As we mentioned in Section 3.2, the PSTH shows the trial-averaged observed firing rate as a function of (trial-relative) time for a single spike train. The essential idea behind identical trials models is to account for the structure of each spike train's PSTH.

When used as a null hypothesis for detecting synchrony, identical trials models attempt to decompose the influences on observed synchrony into those that result from a time-varying PSTH and those that do not. This decomposition is useful in some contexts, but it does not directly address our focus here, which is the decomposition into fine temporal influences and

---

[11]In many cases, one can compute the expected null CCH without explicit generation of pseudo-datasets. This can greatly accelerate computation, although, proper hypothesis testing and the construction of acceptance bands, especially simultaneous acceptance bands, typically do require explicit generation of pseudo-datasets.

[12]Technically, it may not be a constant depending on edge effects and the method of smoothing, but ignoring these details this constant corresponds to an estimate of the $E[U_0]E[V_0]$ term in the definition of the Pearson cross-correlation function $\rho(\ell)$ from Section 4.1.

coarse temporal influences. In Section 4.4.4 we discuss trial-to-trial variability models, which preserve much of the familiar structure and terminology of identical trials models, but attempt to better distinguish between fine and coarse temporal influences on observed synchrony.

4.4.1 Joint peri-stimulus time histogram (JPSTH)

We have been using the CCH as our descriptive measure of interactions between spike trains. For experiments with repeated trials, the joint peri-stimulus time histogram (JSPTH) is another important graphical display of the joint interactions between spike trains [4]. The JPSTH is especially convenient because, unlike the CCH, it is much simpler to anticipate visually the effects of each spike train's PSTH. For two simultaneously recorded spike trains, the construction of a JPSTH begins with all pairs of trial-relative spike times (s_i, t_j) that happen to occur during the same trial, where the first spike time comes from spike train one and the second from spike train two. These pairs are similar to those that defined the CCH histogram, except that we now use trial-relative time and we only consider pairs from the same trial. The JPSTH is simply a two-dimensional histogram of these spike pairs.

The JPSTH decomposes the CCH according to spike time. If, when constructing the CCH, we only consider spike pairs that occurred during a trial and during the same trial, then this *trial-restricted CCH* can be derived directly from the JPSTH by summing (or integrating) along diagonals. The main diagonal (where $s_i = t_j$) gives lag zero in the CCH and the off-diagonals give the other lags. In principle, then, the JPSTH shows more than the CCH. In practice, the JPSTH requires much more data to reliably distinguish signal from noise. This is especially true for synchrony-like correlations, which can easily be obscured by noise or over smoothing.

The first column of Figure 5 shows some PSTH, JPSTH, and trial-restricted CCH examples. Rows A and B show data corresponding to the second row of Figure 3, but with trials defined differently in each case. Only in the first case are the trials aligned with the shared state of the spike trains (see Figure 3 caption) revealing the underlying block structure of the simulation in the JPSTH. Row C corresponds to the final row of Figure 3. With more data and very little smoothing, a narrow ridge along the diagonal would appear in this case, but for the given data we were unable to choose the smoothing parameters in a way that made the ridge visible. (It is visible in the trial-restricted CCH.) Rows D–E show real data recorded from anesthetized monkey primary visual cortex for a sinusoidal drifting grating stimulus (courtesy of Matt Smith; see [66] for additional details). The periodicity of the spiking response for the individual spike trains is clearly apparent in the PSTHs, and this is strongly reflected in the JPSTH.

From the hypothesis testing perspective, the JPSTH is simply another test statistic, just like the CCH. We can use any null hypothesis and proceed exactly as with the CCH. Because the JPSTH is two dimensional, it is difficult to show acceptance bands, and we only show model-corrected JPSTHs in this chapter. In practice, of course, we strongly recommend drawing proper statistical conclusions, whether or not they are easy to visualize.

4.4.2 Independent inhomogeneous Poisson process (IPP) model

One of the most basic identical trials models for a single spike train is the inhomogeneous Poisson process (IPP) model. The spike train on a single trial is modeled as an IPP with intensity $\lambda(t)$, where t refers to trial-relative time, and the observations across trials are modeled
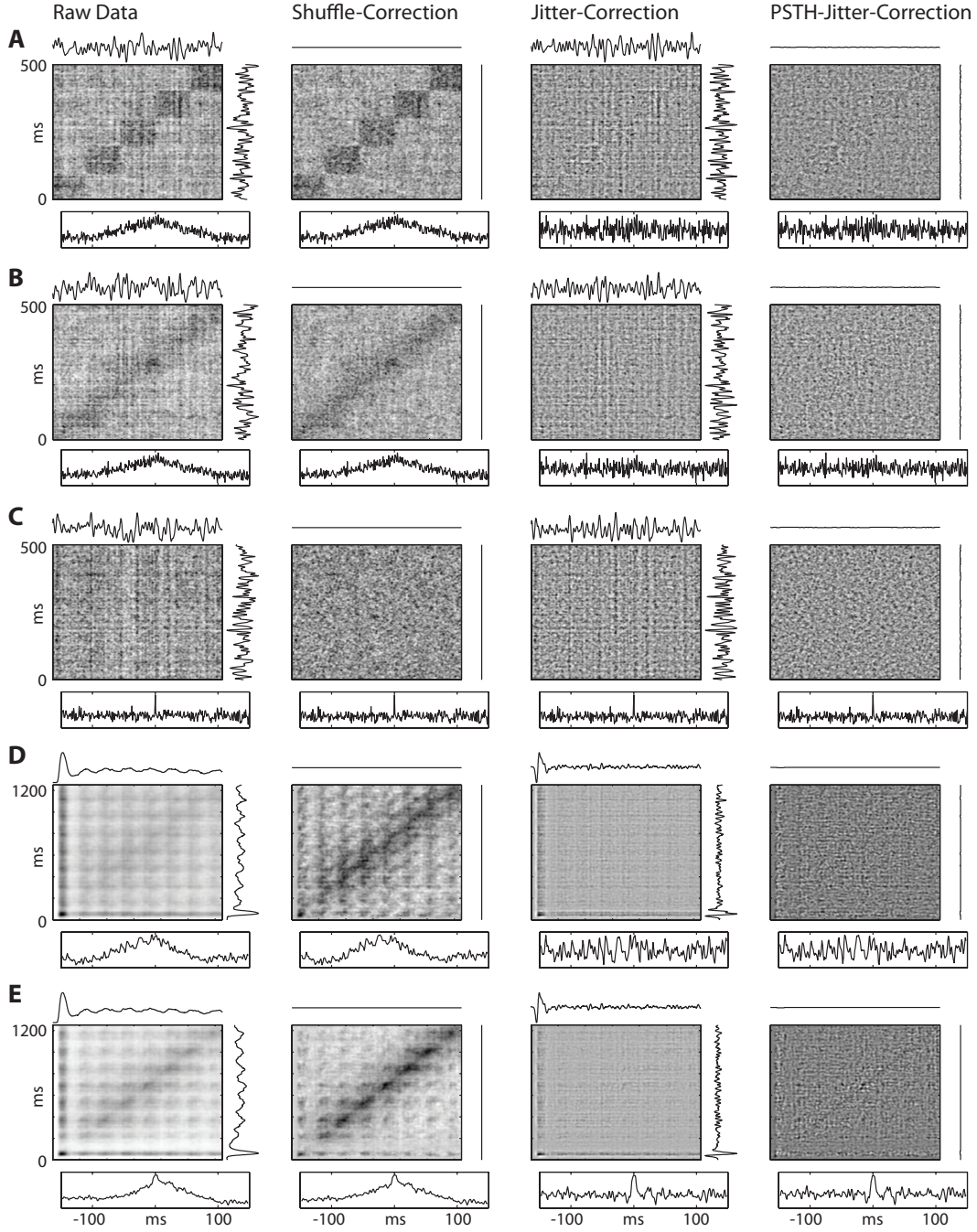
Figure 5: *Examples of model-correction on different datasets. Each group of plots shows a model-corrected JPSTH (central image), two model-corrected PSTHs (lines above and to the right of each JPSTH, corresponding to the two spike trains), and a model-corrected, trial-restricted CCH (plots below the JPSTH). Left-to-right, (1) raw data (no model-correction), (2) shuffle-correction, (3) 20 ms jitter-correction, (4) 20 ms PSTH-jitter-correction (Section 4.5.2). Row A: Data from Figure 3 row 2 with 500 ms trials aligned to switching of hidden coupling variable. Row B: Same as row A, but with misaligned trials. Row C: Data from Figure 3 row 4. Rows D–E: Real data recorded from anesthetized monkey primary visual cortex for a sinusoidal drifting grating stimulus (courtesy of Matt Smith).*

as independent and identically distributed (iid) observations from this same IPP.[13] The PSTH (normalized to the appropriate units) can be interpreted as an estimator of the intensity under this model. For two spike trains, say A and B, we will additionally assume that they are mutually independent, with respective intensities $\lambda^A(t)$ and $\lambda^B(t)$. We call this null model the independent IPP model. Note that there are two different independence assumptions: (i) independence across trials (even for a single spike train) and (ii) independence across spike trains (even for a single trial). The independent IPP model is the natural generalization of the independent HPP model to allow for time-varying firing rates. It formalizes questions like, "Is there more synchrony than expected from the firing rates alone?", with the same probabilistic concepts as in the independent HPP model, except now the intensities are allowed to be time-varying.

Now that we have fixed a null hypothesis, the abstract steps of hypothesis testing proceed exactly as in Sections 4.3.1–4.3.7. We must find some way to handle the ambiguity of the unknown intensities, often with a bootstrap approximation or with conditional inference. Then we can compute acceptance bands and/or p-values, often with Monte Carlo samples. And we can create model-corrected graphical displays, like CCHs or JPSTHs, to better visualize the conclusions.

There are, however, several additional subtleties as compared with Sections 4.3.1–4.3.7, especially if we want to use a bootstrap approximation, because the unknown intensities are now functions. Estimating a function always requires care, and typically involves, either explicitly or implicitly, additional smoothness assumptions that should be communicated along with the model. Appropriate choice and understanding of these smoothness assumptions are especially important in the context of synchrony, because synchrony requires a fine discrimination of time scale that can be strongly affected by temporal smoothing. The special case here of estimating the intensity of an IPP is closely related to the statistical topic of non-parametric density estimation, and various methods of this kind may be used to smooth the peri-stimulus time histogram (PSTH) and thereby obtain good estimates of the two firing-rate functions $\lambda^A(t)$ and $\lambda^B(t)$. For example, [45] develop firing rate estimators using Bayesian adaptive regression splines (BARS) [21]. The bands shown in the left panel of Figure 2 come from this type of bootstrap. The bands indicate the middle 95% of CCH values under the null hypothesis of independence. A similar method may be applied in the non-Poisson case, using estimated conditional intensity functions based on the internal history [72].

When our primary emphasis is on testing the independence-across-spike-trains assumption in the null independent IPP model, we can avoid a bootstrap approximation by using a particular type of conditional inference procedure called the *permutation test*, or trial shuffling. The effective null hypothesis for the permutation test is much larger than the independent IPP model (but includes independent IPP as a special case), and we discuss it in the next section.

4.4.3   Exchangeable model and trial shuffling

Consider the following nonparametric model for identical trials. For a single spike train, the trials are independent and identically distributed (iid), and multiple spike trains are mutually

---

[13]If the experiment involves multiple conditions, then we typically restrict this model to a single condition. Another way to write this is to use the conditional intensity $\lambda(t|Z)$ given the experimental condition Z, and model the spike train as a conditionally IPP given Z.

independent. These are the same two independence assumptions as in the independent IPP model, and also the same identically distributed assumption, but we avoid any additional modeling assumptions about the spike trains. We refer to this as the *exchangeable model*, for reasons we explain below. The exchangeable model includes the independent IPP model, but it also includes all types of non-Poisson processes with refractory periods, bursting, or any other auto-correlation structure. The notion of firing rate is associated with the conditional intensity given the internal history, namely, $\lambda^A(t|H_t^A)$ for spike train A (and similarly for other spike trains), where t refers to trial-relative time. (Recall from Section 3.1 that this notion of firing rate completely specifies the distribution of the spike train.)

Suppose that, for each spike train, we condition on the exact sequence of trial-relative spike times for each trial *but not the trial order*. The only remaining uncertainty in the data is how to *jointly* order the trials for each spike train. Under the exchangeable model, this conditional distribution is uniquely specified: all trial orderings are equally likely[14] — they are *exchangeable*. Conditioning removes all ambiguity about the firing rates $\lambda(t|H_t)$ of each spike train. For the special case of IPP spike trains, conditioning removes ambiguity about the intensities $\lambda(t)$. Furthermore, Monte Carlo sampling under this null conditional distribution is easy: uniformly and independently shuffle the trial order for each spike train to create pseudo-datasets (from which one can create pseudo-data CCHs, JPSTHs, or other test statistics). Model-based correction under the exchangeable model with conditional inference is usually called *shuffle-correction*. It is one of the most popular methods of correcting CCHs and JPSTHs so that they better reveal dependencies between spike trains that are not explained by the PSTHs. (Notice that pseudo-datasets created by trial shuffling exactly preserve the PSTH of each spike train, so shuffle-correction necessarily removes features of graphical displays that directly result from the PSTHs.) Proper hypothesis tests under this model are called *permutation tests*.

It is important to understand the scientific questions addressed by the exchangeable model (with conditional inference). By conditioning on so much of the data, the exchangeable model cannot address most questions about individual spike trains. It does, however, focus attention on the dependence between spike trains. The conditional distribution of trial orderings given the spike times precisely captures the scientific question of whether the spike trains have dependence above that which can be accounted for by identical trials, because it is the trial order of each spike train that controls which trials from spike train A are paired with which from spike train B. If there is no additional dependence, then the spike times from spike train A on a given trial, say trial i, should have no predictive power for choosing among the possible trials from spike train B to be paired with A's trial i. In this case, all trial orderings are equally likely. Alternatively, if there is additional dependence, then the spike times on A's trial i should provide some information about which of B's trials are paired with A's trial i. In this case, trial orderings are not equally likely. Hypothesis testing using the exchangeable model can, in principle, distinguish among these two cases. But they cannot make further distinctions of time scale.

Examples of shuffle-corrected JPSTHs and CCHs can be seen in the second column of Figure 5. Rows D–E show how shuffle-correction reduces the effects of the PSTHs on the JPSTH,

---

[14]If we have n trials, labeled $1, \ldots, n$, for each of m spike trains, then there are $n!$ permutations of the trial labels for each spike train, giving a total of $(n!)^m$ different choices for arranging the trials of all spike trains. The conditional distribution of these choices under the exchangeable model is uniform, that is, each choice has probability $(n!)^{-m}$.

more clearly revealing additional correlations. One can also see from the figures that shuffle-correction and the exchangeable null model (including the independent IPP null model) are agnostic about the time scale of the interactions between spike trains. There are many plausible neuronal mechanisms, such as slow-wave oscillations or slow-temporal gain modulations, and experimental artifacts, such as imprecisely aligned trials or drifts in signal-to-noise ratio, that (i) could influence observed spike timing in multiple neurons, (ii) do not reflect processes that induce precise, coordinate spiking, and (iii) are not accounted for by the basic identical trials models because, for example, they may not be precisely aligned to the trials. Consequently, hypothesis tests based on these basic models may not isolate the fine temporal influences of observed synchrony. Different models are needed.

4.4.4 Trial-to-trial variability (TTV) models

For investigations of the fine temporal influences on observed synchrony, the utility of identical trials models is diminished by the possible existence of shared, coarse temporal influences on observed spike timing that cannot be modeled using identical trials [10, 9, 11, 12, 13, 17, 30, 54, 73, 53]. Examples of such influences might include slow-temporal gain modulations, slow-wave oscillations, metabolic changes, adaptation over trials, reaction time variability, imprecisely aligned trials, or drifts in the recording signal. Ideally, one would like to formulate a null hypothesis that allowed for these coarse temporal sources of correlation. The resulting hypothesis tests and model-corrected JPSTHs and CCHs would then target precise spike timing. In this section we discuss a particular type of model formulation, often referred to as *trial-to-trial variability (TTV)* models, that broaden identical trials models in order to better account for shared, coarse temporal influences on spike timing. TTV models are not pure identical trials models, but include elements of both identical trials models and temporal smoothness models (Section 4.5). TTV models cause a great deal of confusion in the analysis of synchrony, and their use and interpretation requires care.

The idea behind TTV models is to preserve the basic intuitive structure of identical trials models, but to allow the firing rates to change on each trial. Coarse temporal interactions that differ across trials can now be modeled as affecting the firing rates. This allows the firing rates of different spike trains to co-modulate with each other on a trial-by-trial basis, greatly increasing the scope of possible interactions that can be modeled with the concept of firing rates. We can now test for additional interactions above those accounted for by the (trial-varying) firing rates in order to better isolate additional fine temporal mechanisms of precise spike timing. Notice the distinction between time scales in this line of reasoning. Coarse temporal interactions get absorbed into the definition of firing rates, but fine temporal interactions do not. It is this distinction of time scales that make TTV models appropriate for the distinction between coarse and fine temporal influences on observed synchrony. And, crucially, it is this distinction of time scale that must be explicitly understood when interpreting TTV models in the context of synchrony or more general investigations of precise spike timing.

There are many ways to formulate TTV models mathematically. One approach is to ignore the trial structure all together; basically, to model the data as a single trial. As this abandons the concept of trials (even if the experimental protocol involves repeated trials), such models are typically not called TTV models and we discuss them in Sections 4.4 and 4.5 below. A more common approach, and the one we take in this section, begins with the introduction of

an unobserved process, say $U(k)$, for each trial k, that is common to all spike trains. Firing rates are defined as conditional intensities given U (and anything else of interest). For example, the firing rate of spike train A on trial k could be the conditional intensity $\lambda^A(t|H_t^A, U(k))$ and similarly for spike train B, where t is trial relative time. Since the firing rates of both A and B depend on the common value of $U(k)$, they can co-modulate across trials. In the context of IPP models, we could further assume that the spike trains are conditionally Poisson given $U(k)$, namely, that $\lambda^A(t|H_t^A, U(k)) = \lambda^A(t|U(k))$. If $U(1), U(2), \ldots$ are iid, then this conditionally Poisson model is called a *Cox process* or a *doubly stochastic Poisson process*.[15] If the functional form of $\lambda(t|U)$ is fully general, meaning that changes in U across trials can completely alter all aspects of the conditional intensity, then we can fit any data set perfectly without disambiguating the potential contributions to observed synchrony (c.f., Footnote 5). TTV models must restrict the types of variability that are allowed across trials. And for our purposes, we would expect these restrictions to involve some distinction of time scale.

With an appropriate TTV model for individual spike trains, we can formulate a null hypothesis that the spike trains are conditionally independent given the firing rates, i.e., given each $\lambda(t|H_t, U(k))$. Notice that our definition of firing rate depends on the hidden variable U. This is a different definition of firing rate than in previous models we have considered. There have been a few attempts to use TTV models in this way for isolating synchrony-like correlations. For example, [13] introduces a Cox process where $\lambda^A(t|U(k)) = c^A(U(k))g^A(t) + d^A(U(k))$ for scalars $c^A(U(k)), d^A(U(k))$ and a function $g^A(t)$, and similarly for spike train B. In words, the time-varying shape of the intensity, $g^A(t)$, cannot vary across trials, but the total magnitude of the intensity, $c^A(U(k))$, and the baseline activity, $d^A(U(k))$, can vary across trials and can co-vary across spike trains. This allows the firing rates to model very slow gain modulations and is often called *amplitude variability* or *excitability variability*. A bootstrap approximation can be used to remove the ambiguity of the c's, d's, and λ's (with all the caveats from Section 4.4.2). There are many other examples of models that allow U to introduce highly specific changes in some baseline intensity $\lambda(t)$. One of the more recent is [73], which extends the amplitude variability approach by allowing the c's to also vary slowly with time according to certain specific parametric assumptions. Another recent approach to TTV is the PSTH-jitter method described in Section 4.5.2.

4.5 Temporal smoothness models

Pure temporal smoothness models do not make use of repeated trials, even if the experiment involves repeated trials. They are essentially single trial models and place restrictions on how quickly the firing rates (suitably defined) can change. By allowing the firing rates of multiple spike trains to co-modulate, but limiting how quickly they can co-modulate, temporal

---

[15]The Cox process is instructive for understanding the confusing terminology that surrounds TTV models. Each trial of a Cox process begins with the iid selection of a trial-specific intensity function from some distribution over possible intensities. Then the spike times (for that trial only) are generated according to an inhomogeneous Poisson process with that trial-specific intensity. For a Cox process, the trials are independent and identically distributed, and the unconditional intensity $\lambda(t)$ does not vary across trials. There would seem to be no trial-to-trial variability. In many neuroscience contexts, however, it is not the unconditional intensity that maps onto the intuitive notion of a firing rate, but rather, the trial-specific intensities, which are conditional intensities, say $\lambda(t|U)$, given the trial-specific selection, U. It is these conditional intensities that vary across trials (because U varies across trials); hence the trial-to-trial variability terminology.

smoothness models are designed to explicitly address the issue of precise spike timing. Any coarse temporal interactions can be absorbed into the definition of the firing rates, whereas, all precise spike timing must be explained by additional mechanisms. Temporal smoothness models can often be combined with identical trials models to create TTV models. This is especially convenient for situations where the firing rates could co-modulate quickly in response to a precisely timed stimulus, but where this type of stimulus-locked, fine temporal, coordinated spiking is not reflective of the types of fine temporal mechanisms an investigator wishes to study.

4.5.1 Independent inhomogeneous slowly-varying Poisson model

Consider the independent inhomogeneous Poisson process (independent IPP) null model (Section 4.4.2) *for a single trial*. As discussed in Section 3.2 (c.f., Footnote 5), this model is too general to be useful without additional constraints. For the case of two spike trains, say A and B, suppose that we further assume that the intensities $\lambda^A(t)$ and $\lambda^B(t)$ are piecewise constant over the disjoint Δ-length intervals $(k\Delta - \Delta, k\Delta]$ for integers $k \in \mathbb{Z}$. If we let λ_k^A denote the common value of $\lambda^A(t)$ for $k\Delta - \Delta < t \leq k\Delta$, and similarly for λ_k^B, then the null hypothesis is completely characterized by the λ_k^A's and λ_k^B's. We can think about this model as a sequence of small independent *homogeneous* Poisson process models (Section 4.3), one for each Δ-length interval. For shorthand we refer to this null model as the independent Δ-IPP model.

The piecewise constant assumption in the independent Δ-IPP model is designed to loosely capture the intuition of an intensity function that is slowly varying on time scales qualitatively larger than Δ. Certainly one could model slowly varying intensities in any number of ways. The key point is that for synchrony detection one must chose a sensible way to quantify what "slowly varying" means. Here this is quantified by Δ, which is often taken to be something on the order of 5–50 ms, but could be any value depending on the time scales of the phenomena an investigator wishes to disambiguate. The basic intuition for the independent Δ-IPP model is the same as for previous models we have considered: "Is there more synchrony than expected from the firing rates alone?" The difference is that now firing rates are only allowed to vary on time scales larger than Δ. Any other probabilistic structure in the joint spiking must be accounted for by mechanisms that we are not including in our definition of firing rates. Note that Δ, which controls time scale of the null hypothesis of no excess synchrony, is usually chosen to be somewhat larger than the time scale used in the definition of observed synchrony, the latter of which plays the role of a test statistic, and, loosely speaking, controls the time scale of the alternative hypothesis of excess synchrony.

There are two rationales for choosing the piecewise constant characterization of slowly varying intensities in the independent Δ-IPP model. First, for essentially any sensible characterization of intensities that are slowly varying on time scales qualitatively larger than Δ, we would expect these intensities to be approximately constant on Δ-length intervals. To a first approximation then, this particular null model includes all independent inhomogeneous Poisson processes with appropriately slowly varying intensities. Second, the piecewise constant nature of the intensities permits a conditional inference approach to removing ambiguity about the intensities, called *jitter*, which we discuss below. This allows the model to be used successfully in practice.

In order to use the independent Δ-IPP model, we need a method for handling the ambiguity

of the unknown λ_k's. A bootstrap approximation is one potential solution. In each Δ-length interval and for each neuron, we can approximate the constant intensity in that interval with an estimate based on the data (usually the number of spikes in the interval divided by Δ). Δ will typically be small, with each interval having only a few, often zero, spikes. Consequently, this is not a convincing approximation. Great care must be taken when using a bootstrap approximation in situations where the resulting approximation is poor, because then the sensibility of the method largely rests upon whether its eventual use, such as accessing the null variability of a CCH, is robust to large approximation errors. Assessing this robustness can be quite challenging.

A less dubious procedure, in this case, is to condition on the *sequences* of spike counts in each Δ-length interval for each spike train, that is, on the sequences $\left(N^A(k\Delta)-N^A(k\Delta-\Delta):k\in\mathbb{Z}\right)$ and $\left(N^B(k\Delta)-N^B(k\Delta-\Delta):k\in\mathbb{Z}\right)$. For the independent Δ-IPP null model, the conditional distribution of spike times given the sequence of spike counts is uniform in each Δ-length interval, regardless of the actual values of the intensities. Generating Monte Carlo pseudo-data from this null conditional distribution is trivial: independently and uniformly perturb (or "jitter") each spike in the Δ-length interval in which it occurred.

This jitter conditional inference procedure permits exact statistical testing of the independent Δ-IPP model, and, following the reasoning above, it permits approximate statistical testing of essentially any independent IPP model (whether single trial, identical trials, or TTV) with intensities that are slowly varying on time scales qualitatively larger than Δ [7]. These Poisson models, however, make precise assumptions about the joint distribution of spike counts (namely, independent Poisson random variables), whereas the conditional inference procedure is insensitive to these assumptions. Conditional inference, in this case, is effectively testing a much larger null hypothesis, and it is important to understand if this larger null model is still appropriate for the scientific task of distinguishing between fine temporal and coarse temporal influences on observed synchrony.

4.5.2 Δ-uniform model and jitter

Here we more carefully inspect the effective null hypothesis tested by the jitter conditional inference approach described in the previous section. Conditional inference is useful because, under the independent Δ-IPP model, the conditional distribution of spike times given the sequence of spike counts in Δ-length intervals is uniform; there is no remaining ambiguity. For convenience, we will call this the Δ-uniform conditional distribution. Distributions within the independent Δ-IPP model have the Δ-uniform conditional distribution, but there are many other distributions with this property, such as Cox processes with trial-specific intensities that are Δ-piecewise constant. Consider the class of distributions over pairs of spike trains that have the Δ-uniform conditional distribution. We refer to this class of distributions as the Δ-*uniform model*.[16] *The Δ-uniform model can be tested by the same jitter conditional inference procedure*

---

[16]Here is a more formal description of the Δ-uniform model using notation that also extends to more than two spike trains. Let $s = (s_1, s_2, \ldots, s_m)$ be an increasing sequence of spike times and define $C_k(s) = \#\{i : k\Delta - \Delta < s_i \leq k\Delta\}$ to be the number of spikes in s that occur during the kth Δ-length jitter window. Let $C(s) = (C_1(s), C_2(s), \ldots)$ be the sequence of spike counts in jitter windows. For each $j = 1, \ldots, n$, let $S^j = (S_1^j, S_2^j, \ldots, S_{M_j}^j)$ be the increasing sequence of spikes from the jth spike train. The Δ-uniform model for discrete time is that $P\left(S^1 = s^1, \ldots, S^n = s^n \big| C(S^1) = c^1, \ldots, C(S^n) = c^n\right)$ is constant on the set of allowable

used for the Δ-IPP model in the previous section.

The Δ-uniform model is a very large class of spike train models. The sequences of spike counts in Δ-length intervals, namely, $\left(N^A(k\Delta) - N^A(k\Delta - \Delta) : k \in \mathbb{Z}\right)$ and $\left(N^B(k\Delta) - N^B(k\Delta - \Delta) : k \in \mathbb{Z}\right)$, can be viewed as temporally coarsened versions of the original spike trains. The Δ-uniform model allows for any joint distribution over these temporally coarsened, simultaneously recorded spike trains, including any type of dependence within or across the temporally coarsened spike trains. To a first approximation, it allows for all coarse temporal interactions between spike trains, regardless of their source, where "coarse" means qualitatively longer time-scales than Δ. On the other hand, since the Δ-uniform model specifies that the precise spike times are uniform subject to the temporally coarsened spike trains, it does not allow for any precise temporal interactions among spikes, including mechanisms that create precise synchrony. When used as a null hypothesis, the Δ-uniform model is designed to detect precise spike timing of any source [7].

As with the independent Δ-IPP model, conditioning on the sequences of spike counts in Δ-length intervals (i.e., conditioning on the temporally coarsened spike trains) removes all ambiguity about the unknown coarse distribution, and uniquely specifies the Δ-uniform conditional distribution. Exact quantitative inferences, such as p-values, for the Δ-uniform conditional distribution are computationally feasible in special cases [5, 7, 32, 33], but inference usually proceeds via Monte Carlo approximation. Monte Carlo pseudo-data from the Δ-uniform conditional distribution are easily generated by a procedure called *jitter*: independently and uniformly perturb (or "jitter") each spike in the Δ-length interval in which it occurred. Hypothesis tests, acceptance bands, and model-based corrections can be based on this pseudo-data exactly as in Sections 4.3.2–4.3.5. The third column of Figure 3 shows $\Delta = 20$ ms jitter-corrected CCHs for the corresponding raw CCHs in the left column. Only the final two processes show significant synchrony, consistent with our intuition that these processes involve correlations on time scales shorter than 20 ms. 20 ms jitter-corrected JPSTHs, PSTHs, and CCHs can be see in the third column of Figure 5. Rows C and E are the only rows that suggest precise spike timing.

The Δ-uniform model was originally conceived as a new modeling approach for addressing questions of precise spike timing based on the perspective that scientific questions about spike timing were largely questions about the conditional distribution of the precise placement of spikes given the coarse temporal structure of the spike trains [1, 7, 19, 26, 27, 34, 36, 55, 56, 67]. Various applications to neurophysiological questions can be found in [58, 16, 22, 24, 25, 41, 47, 52, 60, 64, 65, 66, 59, 48, 33]. It is one of the simplest and most effective models for isolating and visualizing synchrony-like correlations.

The Δ-uniform model requires an appropriately chosen test-statistic, such as the observed synchrony count or the CCH, so that the hypothesis test has power for detecting the type of precise timing that is being investigated, such as synchrony. An appropriate test-statistic is also important for helping to mitigate the very real possibility that the null hypothesis is rejected because of the wrong type of precise spike timing, such as refractory periods, bursting, or precise stimulus locking. A better solution is to change the null hypothesis so that it also includes those types of precise spike timing that are not being investigated. For example, [34] modify the Δ-

---

$s^1, \ldots, s^n$, namely, the set for which each s^j is an increasing sequence of spike times over the appropriate observation interval and for which each $C(s^j) = c^j$ $(j = 1, \ldots, n)$. The probability is zero elsewhere. For continuous time, the characterization of the model is the same, except we use the probability density instead of the probability. The model allows for any joint distribution on the sequences of spike counts $C(S^1), \ldots, C(S^n)$.

uniform model so that individual spike trains can also have precise temporal auto-correlation structure, such as refractory periods and bursting, and they describe an algorithm, called *pattern jitter*, for exact Monte Carlo sampling from an appropriate null conditional distribution. Replacing jitter with pattern jitter ensures that statistical conclusions about precise synchrony are not artifacts of precise auto-correlation structure. For another example, [8, 66] incorporate the Δ-uniform model into a TTV model by also conditioning on the observed PSTH. Monte Carlo sampling uses an algorithm called PSTH-jitter that does not jitter spikes uniformly, but rather jitters according to the trial-averaged PSTH. Replacing jitter with PSTH-jitter ensures that statistical conclusions about precise synchrony are not artifacts of stimulus-locked firing rates (at least if the trials are correctly aligned). The fourth column of Figure 5 shows $\Delta = 20$ ms PSTH-jitter-corrected CCHs. The utility of PSTH-jitter-correction for visualizing the JPSTH is especially apparent in rows D–E where the PSTH modulates strongly during the trial.

4.5.3 Heuristic spike resampling methods

The jitter Monte Carlo sampling algorithm is closely related to a heuristic spike-resampling approach that we call *spike-centered jitter* (various terms are used in the literature such as jitter, dither, or teeter), which creates pseudo-datasets by jittering each observed spike in a window centered around its original location. Spike-centered jitter is motivated by the observations that (i) the resulting pseudo-data essentially preserve all coarse temporal correlations of the original spike trains, (ii) the resulting pseudo-data essentially destroy any fine temporal correlations that may have been present in the original spike trains, and consequently (iii) comparisons between the original and the pseudo-data can reveal the presence of fine temporal correlations. Most heuristic procedures are designed according to similar criteria.

Unlike the jitter procedure described in the previous section, spike-centered jitter does not correspond to a well-defined null hypothesis and cannot be used to generate meaningful p-values, hypothesis tests, or other statistical conclusions [7, 34]. Like spike-centered jitter, there are many other spike resampling techniques in the literature that are not based on a true underlying model (e.g., [27, 38, 51, 53]). Heuristic procedures are invaluable data exploration tools, but cannot be used to draw statistical conclusions and should not be discussed using statistical terminology. Proper statistical conclusions require a proper statistical model. Just as spike-centered jitter is closely related to jitter and the Δ-uniform model, we suspect that most heuristic spike resampling approaches are closely related to some model. Whenever possible, we prefer replacing heuristic procedures with more careful model-based approaches.

4.6 Generalized regression models

The point process conception described in Section 3 was used in Section 4 to provide a rigorous statistical framework for separating synchronous spiking from correlated firing that occurs at broader time scales. The framework in Sections 4.2–4.5 was aimed at testing specific null hypothetical (non-synchronous) models, which progressed from the simplest null, the independent homogeneous Poisson process model, to much more general null models that allow complicated spike patterns while assuming independent uniform conditional spiking distributions within intervals of length Δ. In this section we describe an alternative approach, based on generalized regression models, that uses measured variables, or covariates, to describe correlation at coarse

times, and provides estimates of the magnitude of excess synchronous spiking beyond that produced under independence—the null case corresponding to independence, and thus zero excess synchronous spiking. Generalized regression models parameterize the effects of synchrony, and thereby characterize the excess probability of spiking beyond that predicted by independence. In order to provide such information, which can not be obtained from the hypothesis testing framework in Sections 4.2–4.5, the regression approach introduces additional modeling assumptions.

Let us return to the definition of instantaneous firing rate in terms of the conditional intensity function, as in Equation (2), and re-express the conditional intensity in the form

$$\log \lambda(t|H_t) = \text{ stimulus effects } + \text{ history effects}, \tag{3}$$

where we use "stimulus effects" and "history effects" to refer to some linear combination of measured variables that represent, respectively, the response to stimulus and the spiking history prior to time t. When applied to binary data in discrete time bins, assumed to arise from a point process, Equation (3) defines what is often called a *generalized linear model (GLM)* [44]. In ordinary linear regression the response values may range over a wide scale. Here, instead, the model is "generalized" to accommodate the representation of spike trains as a sequence of 0s and 1s.

One of the purposes of regression methods is to adjust effects of interest, such as stimulus effects, by removing the effects of covariates that are considered irrelevant, such as refractory effects represented as spiking history. In addition to spike history effects, generalized regression models may incorporate additional covariates that represent slow fluctuations in firing rate that occur across trials [73]. Furthermore, measured network effects related to neural inputs can be included in the generalized regression model and they, too, can be effectively removed from the analysis [46, 43, 42]. If, for example, each of two neurons were to have highly rhythmic spiking in conjunction with an oscillatory potential produced by some large network of cells, then the two neurons would both tend to spike near the times of the peaks of the oscillatory potential. For a high-frequency oscillation (in the mid-to-high gamma range), the peaks would be concentrated within a range of a few milliseconds, and this could be an important source of synchrony [71]. For slower oscillations, the extra spiking would occur across broader time lags (the CCH peak would be broader), but the extra spikes could mask any possible synchronous spikes that might occur—and that might carry information about the stimulus. In either case, it would be helpful to identify the oscillatory effects, and to look for additional synchronous spiking not associated with the oscillations. An illustration of the slow-wave spiking activity across an array of neural spike trains is given in Figure 6.

Details of regression-based synchrony identification using point process representations and generalized linear models may be found in [42]. In outline, given multiple trials of spike trains for neurons A and B, we consider them to follow point processes with conditional intensities $\lambda^A(t|H_t^A)$ and $\lambda^B(t|H_t^B)$ and we also consider the synchronous spikes to follow point processes[17] with conditional intensity $\lambda^{AB}(t|H_t)$. Here H_t^A and H_t^B are the spiking histories prior to time t for each neuron separately, within each trial, and H_t is the combined spiking history prior to

---

[17]Technically, we assume a family of point processes indexed by bin size δ such that as $\delta \to 0$ $\lambda_\delta^{AB}(t|H_t)$ vanishes at the same rate as δ.
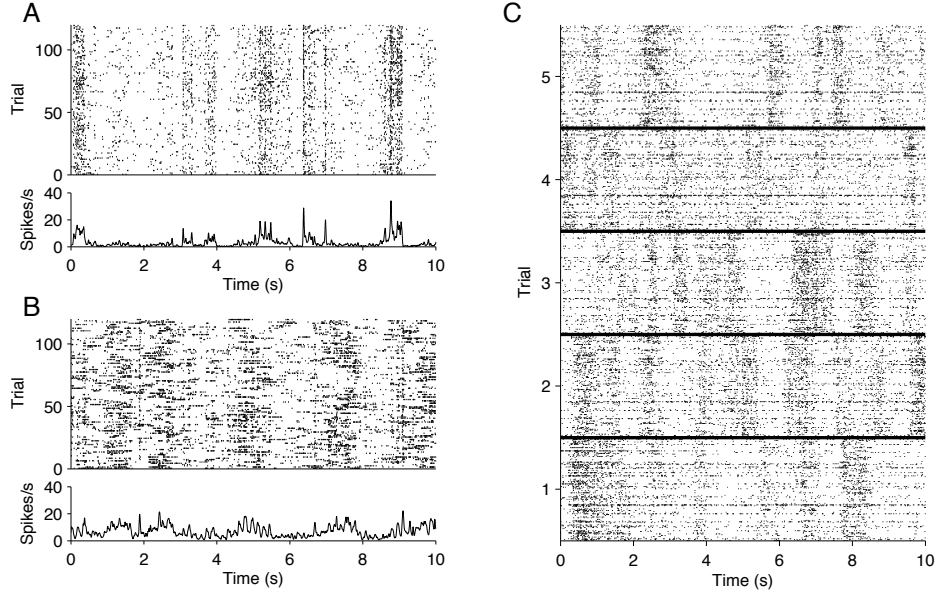
Figure 6: *Neural spike train raster plots for repeated presentations of a drifting sine wave grating stimulus. (A): Single cell responses to 120 repeats of a 10 second movie. At the top is a raster corresponding to the spike times, and below is a peri-stimulus time histogram (PSTH) for the same data. Portions of the stimulus eliciting firing are apparent. (B): The same plots as in (A), for a different cell. (C): Population responses to the same stimulus, for 5 repeats. Each block, corresponding to a single trial, is the population raster for $\nu = 128$ units. On each trial there are several dark bands, which constitute bursts of network activity sometimes called "up states." Up state epochs vary across trials, indicating they are not locked to the stimulus. (Figure reprinted from [43].)*

time t. We define $\zeta(t)$ by

$$\zeta(t) = \frac{\lambda^{AB}(t|H_t)}{\lambda^A(t|H_t^A)\lambda^B(t|H_t^B)} \tag{4}$$

so that under independence we have the null hypothesis

$$H_0 : \zeta(t) = 1 \tag{5}$$

for all t, and $\zeta(t) - 1$ becomes the excess proportion of spiking above that which would occur under independence. We may also include additional variables in these conditional intensity functions. [43, 42] included a variable, which we here label x, based on the joint spiking of 118 other neurons recorded from the same array. This was meant to capture slowly oscillating network effects. Thus, the individual conditional intensities would be written $\lambda^A(t|H_t^A, x)$ and $\lambda^B(t|H_t^B, x)$, the point being that neurons A and B would be more likely to fire during network bursts due to slow oscillations, as displayed in Figure 6. When the right-hand side of (4) is modified to include the variable x, we change the notation on the left-hand side to become $\zeta_x(t)$ and write the independence null hypothesis as

$$H_0 : \zeta_x(t) = 1 \tag{6}$$

In words, (5) is the hypothesis that synchronous spiking is due entirely to the individual-neuron fluctuations in firing rate, and spike times are otherwise independent. Hypothesis (6) allows synchronous spiking to be due, in addition, to slow bursting activity shared by both neurons, but precise spike times are otherwise independent.

[43] illustrated the regression-based point process formulation by analyzing two different pairs of neurons under the assumption that $\zeta = \zeta(t)$ was time-invariant. For both pairs, a bootstrap test of the null hypothesis in (5) gave very small p-values, indicating that there was excess synchrony beyond that produced by chance from the point processes that were assumed to have generated the individual spike trains. To check whether the excess synchrony might have been due to the slow bursting activity, [43] estimated the parameter ζ_x for each pair. For one pair, bootstrap methods produced an estimate of excess spiking $\log_e \hat{\zeta}_x = .06$ with standard error $SE = .15$, so that the test of the null hypothesis in (6) produced a large p-value, and excess synchrony was, apparently, produced by the slow network bursts like those visible in Figure 6. For the other pair, however, they obtained $\log_e \hat{\zeta}_x = .82$ with standard error $SE = .23$ giving a very small p-value for the null hypothesis in (6) and indicating that there remained additional substantial excess synchrony, more than double the number of spikes produced under independence[18], even after the bursting had been accounted for. For this pair, the excess synchrony may well have been providing stimulus-related information.

4.7   Multiple spike trains and more general precise spike timing

We have focused of synchrony analysis, but with respect to hypothesis testing, observed synchrony merely plays the role of a *test statistic*. Our main focus has been on models for the null hypothesis of no excess synchrony, which we have primarily framed as null hypotheses for no precise spike timing of any kind. For this reason, many of the models we have been discussing are also appropriate for testing for the presence of other types of precise spike timing, such as recurring motifs of precisely timed spikes. Only the test statistic needs to change in order to increase the statistical power for detecting the phenomenon of interest. Changing the test statistic is particularly straightforward in the case of Monte Carlo approximation: simply compute the new test statistic on the original dataset and the Monte Carlo pseudo-datasets, and then proceed as before. We have focused on precise pairwise synchrony, with the implication that a relevant test-statistic, such as observed synchrony or a related feature of the CCH or JPSTH, would be used. But for other types a precise spike timing, a different test-statistic might be more appropriate.

For example, in [28] Grün et al. define a *unitary event* (UE) as the event of spike coincidence across a constellation of neurons (in practice consisting of more than two neurons). This is the natural generalization of the observed synchrony statistic to higher-order interactions. The test statistic is then the number of occurrences of a particular UE, or a global statistic that combines multiple UE's. These statistics can then be calibrated in the sense of model-correction, or, more formally, used as the test statistic in a hypothesis test. The interpretation is exactly the same: the focal point of the analysis is the model, and unitary events play a role exactly analogous to observed synchrony or to the CCH.

UE statistics have been applied in combination with different models. In [28] the null

---

[18]Exponentiating .82 gives 2.27, with an approximate 95% confidence interval of (1.4,3.6).

hypothesis is that spike trains are independent homogeneous Poisson processes[19] (Section 4.3) and a bootstrap approximation is used to remove ambiguity about the unknown intensities. That is, the Poisson rate is estimated from the spike trains, and the estimated Poisson rates are used to derive the distribution of the UE statistic under the (Poisson) model. The observed UE statistic is then compared with the derived distribution. A UE statistic is also used in [31], but using the uniform model with conditional inference (Section 4.3.6). The assumption of stationarity is problematic. The Δ-uniform null model (Section 4.5.2), tested with a UE statistic, is a natural alternative approach. As another approach to nonstationarity, [29] consider a slowly varying independent IPP model (Section 4.5.1) and propose a bootstrap approximation using a kernel-based (uniform) smoother to estimate the unknown intensities.

It is worth re-emphasizing here the distinction between a statistic and the null hypothesis. In UE analysis, the statistic is the number of unitary events, which involve higher-order constellations of activity. When a UE statistic is used to reject a null hypothesis, the justifiable statistical conclusion is simply that there is evidence against that null hypothesis. While this may be regarded as a kind of *prima facie* suggestion of higher-order interaction, that conclusion requires some additional assumption or analysis. For example, it may be possible to introduce a framework, such as the generalized regression framework of Section 4.6, according to which a large value of some point process quantity (analogous to $\zeta(t)$ in Equation (4)) will correspond to higher-order correlation, and will make it likely that the UE statistic will reject the null hypothesis. Such a quantity could then be estimated from the data. The strict interpretation of a rejection of the Δ-uniform model is, similarly, that coarse spike timing models (quantified by Δ) are poorly-fit to the spike trains. In general, the choice of statistic affects the power of the test (cf., the discussion of statistical significance in Section A.2) and, in the case of rejection of the null hypothesis, some quantification of the magnitude of the effect—via additional modeling assumptions—is likely to be helpful. [43] showed how the generalized regression framework can accommodate higher-order synchrony among multiple spike trains by including higher-order quantities analogous to $\zeta(t)$ defined in Equation (4). For example, in the case of three-way interaction among three neurons, the analogous quantity represents the excess probability of three-way synchrony above that predicted by a model involving all three pairs of two-way interactions. When the point processes are assumed to be stationary, the methods of [43] reduce to those of [50, 62].

4.8   Comparisons across conditions

Hypothesis testing is designed to detect unusual synchrony, and the related model-based correction perspective is useful for preserving the familiar formats of graphical displays, like the CCH and JPSTH, but tailoring them to reveal specific types of correlations, like synchrony. Hypothesis testing and model-based correction are much less useful, however, for drawing comparisons across conditions. Hypothesis testing focuses entirely on null models capturing the absence of the target phenomenon, like synchrony. Comparing synchrony across different conditions requires explicit models of synchrony, something that is lacking in the hypothesis testing perspective.

---

[19]Depending on whether spike trains are modeled as discrete or continuous. In the discrete case (spike or no spike in time bins of finite extent), the model is stationary Bernoulli. In the continuous case, it is stationary Poisson. The two models are closely related and practically equivalent for small time bins.

Consider the following hypothetical example that illustrates the ambiguities of drawing comparisons across conditions based solely upon hypothesis testing. The point of the example is to determine which of two experimental conditions (I or II) exhibits more unusual synchrony between two simultaneously recorded spike trains (A or B). The spike trains are homogeneous Poisson (but not independent) in each condition. Spike train A has rates of 30 and 10 spikes/s in conditions I and II, respectively. Spike train B also has rates of 30 and 10 spikes/s in conditions I and II, respectively. On average 10% of the spikes (i.e., 3 spikes/s) in condition I are forced to be precisely synchronous. On average 20% of the spikes (i.e., 2 spikes/s) in condition II are forced to be precisely synchronous. The spike times are otherwise independent. (In each case, additional precise alignments between spikes can arise by chance.)

The absolute rate of (unusual) synchronous firing is higher in condition I (3 Hz) than in condition II (2 Hz), but the percentage of spikes that are synchronous is higher in condition II (20%) than in condition I (10%). Hypothesis testing and model-based correction are useful for revealing the unusual synchrony in each condition, but they do not speak to the scientific issue of which condition has more synchrony.

As [39] and others have noted (see also the final discussion in [68]), there is no uniquely compelling answer as to which condition has more synchrony in an example like this. The scientific question is too vague. The answer will depend on how one chooses to *model* synchrony. For example, in the injected synchrony model, whereby synchronous spikes are simultaneously injected into each neuron, then condition I has more synchrony than condition II, since it has more injected spikes. Alternatively, using log-linear models [49], for which synchrony is quantified by an odds ratio related to how much a spike in one spike train increases the chances of a nearby spike in the other spike train, then condition II has more synchrony than condition I, since it has a higher percentage of synchronous spikes. Section 4.6 provides additional details about explicit statistical models of synchrony that can be used to quantify differences across conditions within a common statistical framework.

5   Discussion

We have focused on basic or prominent methods for the statistical identification of precisely-timed synchrony. Figure 7 contains a summary of the main methods that we covered. More generally, any survey of the neurophysiology literature will uncover an enormous diversity of methods employed to analyze spike timing patterns, many of which we have not discussed here. (For example, we have made no mention of either spectral methods or information theory; our general remarks about the importance of proper statistical modeling apply equally to these and other approaches.) Coupled with the experimental complexity of spike data sets, this diversity can present a challenge to investigators and readers.

In reviewing the statistical thinking that underlies different approaches, our major conclusion is that a great deal of attention ought to be focused on the process by which a neuroscientific question is translated into the terms of a precise statistical model. Once a model is specified, inferential techniques can be evaluated or compared using relatively common statistical criteria. On the other hand, many fundamental concepts in neurophysiology and in the neural coding literature remain vaguely conceived in terms of many alternative precise definitions. Here we have examined "synchrony," "chance," and "firing rate" but more generally we also have in

Model	Section	temporal modulation of firing rates	trial-to-trial variability	applicable to single trials	history dependent spiking	explicit separation of time scales	invariant to unknown parameters	specialized synchrony detection	synchrony quantification
Independent HPP	4.3			✔					
Uniform	4.3.6			✔			✔		
Independent IPP*	4.4.2	✔				?			
Exchangeable	4.4.3	✔			✔		✔		
Cox Process*	4.4.4	✔	✔			?		✔	
Independent Δ-IPP	4.5.1	✔	✔	✔		✔		✔	
Δ-Uniform	4.5.2	✔	✔	✔		✔	✔	✔	
Δ-Uniform Extensions*	4.5.2	✔	✔	✔	✔	✔	✔	✔	
Generalized Regression*	4.6	✔	✔	✔	✔	?		✔	✔

Figure 7: *Summary of the main models discussed in this paper. Each row corresponds to a different model, which is described in the indicated section of the text. Each column corresponds to a desirable feature from the perspective of scientifically and statistically appropriate investigations of synchrony and more general precise timing. Models that allow for* temporal modulation of firing rates *can better account for the correlations between spike counts and temporal covariates that are observed in many experiments (cf. Sections 3, 4.4ff). Models allowing for* trial-to-trial variability *additionally account for temporal modulation on time scales spanning experimentally controlled trials (cf. Section 4.4.4). Some models are* applicable to single trials *(cf. Sections 4.3, 4.5, 4.6) and some allow for* history dependent spiking, *such as refractory period and bursting (cf. Sections 3, 4.4.3, 4.5.2, 4.6). Models that accommodate* explicit separation of time scales *allow for some degree of quantification of the temporal precision of spiking (cf. Sections 2, 4.5). The category* invariant to unknown parameters *signifies conditional inference for reducing a large null hypothesis to a single null distribution (as opposed to a bootstrap approximation), which increases robustness and avoids approximation errors, particularly in models with large numbers of unknown parameters (cf. Sections 4.3.5, 4.3.6, 4.4.3, 4.5.2). Some models can explicitly disambiguate dependences on fine time scales from those on coarse time scales, permitting* specialized synchrony detection *(cf. Sections 2, 4.5). Finally, explicit models of synchrony permit statistical* synchrony quantification, *as opposed to just synchrony detection (cf. Sections 4.6). While these categories are somewhat imprecise, we hope that they are more or less identifiable with the key concepts repeated throughout the text. Our assignment of a feature to a given model is based on the observation that many simultaneously recorded spike trains appear to have correlations on a spectrum of time scales ranging from milliseconds to hours. Models with an asterisk (*) are large, flexible model classes that require careful formulation and often custom software in order to have all of the indicated features. Within these models, a question mark (?) means that the feature should be possible, but is not fully present in existing formulations.*

mind "information," "population coding," and "neuronal ensembles," all of which have multiple mathematical definitions in the literature. Consideration of the meaning of these terms is tied intimately to the construction and application of statistical models, the precise specification of which is a pre-requisite for quantitative data analysis. Statisticians are fond of saying that the most important part of an analysis is the mapping between the scientific problem and the statistical model (see [14] together with the published discussion and rejoinder).

Returning to the focus of this chapter, a take-home message is that precisely-timed synchrony, as it is commonly understood in neurophysiology, enters into statistical models of spike trains through a consideration of time scale. This might be formalized in several ways. For example, in the Δ-uniform model (jitter) and its variants, the time scale is modeled explicitly. In many trial-to-trial variability approaches, a time scale is implicitly specified through the time scale at which the conditional intensity function is assumed to vary across trials. Furthermore, it is rarely helpful to use models with no notion of time scale, such as the exchangeable-model (trial-shuffling), to identify precise synchrony. A second message is that hypothesis testing and estimation are distinct statistical enterprises. As we mentioned in Sections 4.7 and 4.8 hypothesis tests are designed to reveal whether the data are inconsistent with models that do not involve precise synchrony dependence. They are not designed to quantify synchrony. Instead, by introducing further assumptions that explicitly model the synchrony time-scale dependencies (as in Section 4.6), we can relate various parameters associated with synchrony to other covariates, such as stimulus or behavior.

One of the current challenges in the statistical modeling of multiple spike trains is the development of scientifically-appropriate and tractable models that can account for a variety of sources of dependence across a spectrum of time scales, including precise synchrony and other higher-order spatio-temporal spiking patterns. Progress is likely to be gradual, with current models informing future neurophysiological experiments, which will then lead to modeling refinements and statistical innovations.

Acknowledgment

Appendix

A   Probability and random variables

Random variables are used as theoretical counterparts to data. A useful mathematical convention is to represent data with lower case letters like x and to identify the corresponding random variables with capital letters like X. Thus, for example, x could be a measured spike count and X the corresponding random variable, which our statistical model might assume follows a Poisson distribution. We write the probability of X taking a value between a and b as $P(a < X < b)$. Two random variables X and Y are said to be independent if for all numbers

a, b, c, d we have $P(a < X < b$ and $c < Y < d) = P(a < X < b)P(c < Y < d)$. Equivalently,[20] X and Y are independent if, having observed that the value of Y is y, the conditional probability of X taking a value between a and b, which we write as $P(a < X < b|Y = y)$, satisfies $P(a < X < b) = P(a < X < b|Y = y)$ for every a, b, y, that is, X and Y are independent if knowing the value of Y does not change any of the probabilities associated with the outcomes of X (and, conversely, knowing X does not change the probabilities associated with Y). When data values occur sequentially across time, the value at time t is often written x_t and the corresponding random variable would be written X_t. Such sequential data and random variables are usually called *time series*. For theoretical purposes, the time index is allowed to take all integer values, both negative and positive. To emphasize this we write the theoretical time series as $(X_t : t \in \mathbb{Z})$, where $\mathbb{Z}$ represents all the integers.

A.1   Statistical models and scientific questions

To assess variation in spiking that could lead to synchronous events "by chance" we must begin with a statistical model. Statistical models describe variation in terms of probability. For example, the simplest probabilistic description of the variation in the spike counts in a particular time interval, across trials, would be to assume they follow a Poisson distribution. A more general specification would be to assume that, on every trial, the timing of spike trains follows a Poisson process. Even more generally, we might assume that on every trial the timing of spike trains follows a point process. The main idea is that the statistical model replaces data with hypothetical, random quantities; the real-world variation in the data is modeled using the mathematics of probability theory. Then, using the statistical model, we can state precisely what we mean when we say that synchrony could arise by chance.

Here we are using "model" in the widest statistical sense. A model could be restricted to involve a small number of free parameters (such as a time-invariant firing rate) or it could be allowed to be very flexible, with a possibly infinite number of parameters (the firing rate could be allowed to vary in mildly constrained ways across both time and experimental conditions). We stress this point because certain statistical methods are occasionally called "model free." Such jargon is misleading. Rigorous statistical assessments always rely on statistical models.

It might be nice to be able to validate a statistical model compellingly, so that it could be considered a "correct" description of measured variation. This is essentially never possible. An important aspect of data analysis involves assessment of how well a model fits the data (e.g., [15]). In applying such goodness-of-fit methods, however, the question, "Does the model fit the data?" is really a short-hand for the more accurate question, "How well does the model fit for our specified scientific purpose?" For some purposes it may be entirely reasonable to assume a Poisson model for spike counts, while for other purposes—even with the same data—it could be grossly misleading.

In particular, one of the key ideas behind the introduction of statistical models is to incorporate within them some notion of regularity, or signal, which is distinguished from noise. Thus, a data-based "sample" mean is distinguished from a theoretical or "population" mean; a data-based firing rate is distinguished from a theoretical firing rate; and a data-based cross-correlation function is distinguished from a theoretical cross-correlation function. Each of these

---

[20]We are ignoring mathematical subtleties treated by measure theory.

theoretical constructs has a definition, which is embedded in the statistical model.

A.2   Hypothesis tests and Monte Carlo approximation

To evaluate whether synchronous firing, indicated by a peak in the CCH, could be due to chance, we introduce a baseline statistical model according to which (1) the spiking activity of each of the two neurons is represented by a point process, and (2) the two point processes are statistically (conditionally) independent (under a specified conditioning). The part of the baseline model that specifies independence is called the *null hypothesis* and is labelled H_0. A statistical test of H_0 is based on some *test statistic*, which becomes a random variable Z, and when applied to a particular set of data the test statistic takes a value z_{obs}. The p-value is then the probability $p = P(Z > z_{obs}|H_0)$ (or, sometimes, $p = P(|Z| > |z_{obs}||H_0)$) where the notation indicates that the probability is based on the assumption that H_0 is correct. The p-value is therefore an indication of the likelihood of a large value of the test statistic "due to chance alone," in the absence of genuine statistical dependence between the neuronal spike trains. When the p-value is very small (e.g., much less than .01) we would conclude that if H_0 is correct, a very rare event would have occurred and, rare events being rare, therefore H_0 is probably not correct: there is evidence against independence, and in favor of some mechanism that produces extra synchrony.

As a simple pedagogical example, suppose we model two neurons, A and B, as 5 Hz homogeneous Poisson processes, and we observe their spike trains repeatedly over 100 1-second intervals. As a test statistic we may compute the number of instances of a spike from neuron A and spike from neuron B occurring within 5 ms of each other. We could take the null hypothesis to be that the two processes are independent, and we can then compute the probabilities associated with each possible number z_{obs} of spike co-occurrences, within a 5 ms window, for 100 sets of 1-second observation times. This would allow us to find the p-value. If the p-value were small, then we would say there is evidence of synchrony *under the assumption of homogeneous Poisson firing at 5 Hz.*

This logic of hypothesis testing is very widely accepted. Its key features are that a null hypothesis specifies a statistical model, which is used to compute the p-value for a particular test statistic. The statistical subtleties and concerns center especially on the choice of model. In the example we just mentioned, the last italicized phrase is worrisome: what if the spiking activity of either neuron is not well-modelled as a Poisson process at 5 hz? In addition, in some contexts the test statistic may be problematic, potentially, if it is sensitive to departures from the null hypothesis other than those of central scientific interest (for then the p-value might be small for some reason other than the presence of excess synchrony). Finally, a further issue is that many procedures do not produce an exact p-value, but only an approximate one, though in many cases the approximation may be sufficiently accurate for practical purposes.

From a statistical perspective, when an investigator proposes a method of testing for synchrony, we want to know what statistical model is used for the null hypothesis, and whether the p-value is valid in the sense of equalling, at least approximately, the putative probability under H_0.

We have not yet specified how the p-value would be computed, but a very general approach is to use computer simulation to generate sets of pseudo-data spike trains repeatedly, under the assumptions specified by H_0, and then to compute the value of the test statistic for each set

of pseudo-data. For instance, one might generate 10,000 sets of pseudo-data, compute the test statistic for each, and then examine the proportion of sets of pseudo-data for which the test statistic exceeds z_{obs}. This proportion would be very nearly equal to the p-value (it is typically accurate to at least 2 digits). In practice, even if the Poisson assumption is applied, one does not know the firing rate and instead "plugs in" the value of firing rate observed in the data. This kind of plug-in estimation is known as a bootstrap approximation and the entire procedure is known as *bootstrap* hypothesis testing. Plug-in estimation produces accurate p-values when there is enough data to ensure that, so long as the null hypothesis is true, the estimate is reliable.

References

[1] M Abeles and I Gat, *Detecting precise firing sequences in experimental data*, Journal of Neuroscience Methods **107** (2001), 141–154.

[2] M Abeles and G L Gerstein, *Detecting spatiotemporal firing patterns among simultaneously recorded single neurons*, Journal of Neurophysiology **60** (1988), no. 3, 909–924.

[3] E.D. Adrian, *The impulses produced by sensory nerve endings: Part i*, J. Physiol. (London) **61** (1926), 49–72.

[4] AM Aertsen, GL Gerstein, MK Habib, and G. Palm, *Dynamics of neuronal firing correlation: modulation of "effective connectivity"*, Journal of Neurophysiology **61** (1989), no. 5, 900–917.

[5] A Amarasingham, *Statistical methods for the assessment of temporal structure in the activity of the nervous system*, Ph.D. thesis, Brown University Division of Applied Mathematics, 2004.

[6] A Amarasingham, T-L Chen, S Geman, MT Harrison, and D Sheinberg, *Spike count reliability and the poisson hypothesis*, Journal of Neuroscience **26** (2006), 801–809.

[7] A. Amarasingham, M. T. Harrison, N. Hatsopolous, and S. Geman, *Conditional modeling and the jitter method of spike re-sampling*, Journal of Neurophysiology **107** (2012), 517–531.

[8] A Amarasingham, MT Harrison, and S Geman, *Jitter methods for investigating spike train dependencies.*, Computational and Systems Neuroscience (COSYNE), 2007.

[9] W Bair, E Zohary, and WT Newsome, *Correlated firing in macaque visual area mt: time scales and relationship to behavior*, Journal of Neuroscience **21** (2001), no. 5, 1676–1697.

[10] SN Baker and GL Gerstein, *Determination of response latency and its application to normalization of cross-correlation measures*, Neural Computation **13** (2001), 1351–1377.

[11] Y Ben-Shaul, H Bergman, Y Ritov, and M Abeles, *Trial to trial variability in either stimulus or action causes apparent correlation and synchrony in neuronal activity*, Journal of Neuroscience Methods **111** (2001), no. 2, 99–110.

[12] C D Brody, *Slow variations in neuronal resting potentials can lead to artefactually fast cross-correlations in the spike trains*, Journal of Neurophysiology **80** (1998), 3345–3351.

[13] _____, *Disambiguating different covariation types*, Neural Computation **11** (1999), 1527–1535.

[14] E N Brown and R E Kass, *What is statistics? (with discussion)*, American Statistician **63** (2009), no. 2, 105–123.

[15] E.N. Brown, R. Barbieri, V. Ventura, R.E. Kass, and L.M. Frank, *The time-rescaling theorem and its application to neural spike train data analysis*, Neural Computation **14** (2002), no. 2, 325–346.

[16] D A Butts, C Weng, J Jin, C-I Yeh, N A Lesica, J-M Alonso, and G B Stanley, *Temporal precision in the neural code and the timescales of natural vision*, Nature **449** (2007), 92–96.

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

[18] D.J. Daley and D.D. Vere-Jones, *An introduction to the theory of point processes*, vol. 2, Springer, 2007.

[19] A Date, E Bienenstock, and S Geman, *On the temporal resolution of neural activity*, Tech. report, Brown University Division of Applied Mathematics, 1998.

[20] A.C. Davison and D.V. Hinkley, *Bootstrap methods and their application*, Cambridge Univ Pr, 1997.

[21] I. Dimatteo, C.R. Genovese, and R.E. Kass, *Bayesian curve-fitting with free-knot splines*, Biometrika **88** (2001), no. 4, 1055–1071.

[22] G Dragoi and G Buzsáki, *Temporal encoding of place cells by hippocampal cell assemblies*, Neuron **50** (2006), 145–157.

[23] J-M Fellous, PHE Tiesinga, PJ Thomas, and TJ Sejnowski, *Discovering spike patterns in neuronal responses'*, Journal of Neuroscience **24** (2004), no. 12, 2989–3001.

[24] S Fujisawa, A Amarasingham, M T Harrison, and G Buzsáki, *Behavior-dependent short-term assembly dynamics in the medial prefrontal cortex*, Nature Neuroscience **11** (2008), 823–833.

[25] S Furukawa and J C Middlebrooks, *Cortical representation of auditory space: information-bearing features of spike patterns*, Journal of Neurophysiology **87** (2002), 1749–1762.

[26] G L Gerstein, *Searching for significance in spatio-temporal firing patterns*, Acta Neurobiologiae Experimentalis **64** (2004), 203–207.

[27] S Grün, *Data-driven significance estimation for precise spike correlation*, Journal of Neurophysiology **101** (2009), 1126–1140.

[28] S Grün, M Diesmann, and A Aertsen, *Unitary events in multiple single-neuron spiking activity: I. detection & significance*, Neural Computation **14** (2002), no. 1, 43–80.

[29] ――――, *Unitary events in multiple single-neuron spiking activity: II. nonstationary data*, Neural Computation **14** (2002), no. 1, 81–119.

[30] S Grün, A Riehle, and M Diesmann, *Effects of cross-trial nonstationarity on joint-spike events*, Biological Cybernetics **88** (2003), no. 5, 335–351.

[31] R Gütig, A Aertsen, and S Rotter, *Statistical significance of coincident spikes: Count-based versus rate-based statistics*, Neural Computation **14** (2002), no. 1, 121–153.

[32] MT Harrison, *Discovering compositional structures*, Ph.D. thesis, Brown University Division of Applied Mathematics, 2004.

[33] ――――, *Accelerated spike resampling for accurate multiple testing controls*, Neural Computation (in press).

[34] MT Harrison and S Geman, *A rate and history-preserving algorithm for neural spike trains*, Neural Computation **21** (2009), no. 5, 1244–1258.

[35] H.K. Hartline, *The receptive fields of optic nerve fibers*, Am. J. Physiol. **130** (1940), 690–699.

[36] N Hatsopoulos, S Geman, A Amarasingham, and E Bienenstock, *At what time scale does the nervous system operate?*, Neurocomputing **52** (2003), 25–29.

[37] N.G. Hatsopoulos, Q. Xu, and Y. Amit, *Encoding of movement fragments in the motor cortex*, Journal of Neuroscience **27** (2007), no. 19, 5105–5114.

[38] Y Ikegaya, G Aaron, R Cossart, D Aronov, I Lamp, D Ferster, and R Yuste, *Synfire chains and cortical songs: Temporal modules of cortical activity*, Science **304** (2004), no. 5670, 559–564.

[39] H Ito and S Tsuji, *Model dependence in quantification of spike interdependence by joint peri-stimulus time histogram*, Neural Computation **12** (2000), 195–217.

[40] H. Jeffreys, *Scientific inference*, Cambridge Univ Pr, 1931.

[41] L M Jones, D A Depireux, D J Simons, and A Keller, *Robust temporal coding in the trigeminal system*, Science **304** (2004), 1986–1989.

[42] R.E. Kass and R.C. Kelly, *A framework for evaluating pairwise and multiway synchrony among stimulus-driven neurons*, Neural Computation (2012), to appear.

[43] R.E. Kass, R.C. Kelly, and W.-L. Loh, *Assessment of synchrony in multiple neural spike trains using loglinear point process models*, Annals of Applied Statistics **5** (2011), 1262–1292.

[44] RE Kass, V Ventura, and EN Brown, *Statistical issues in the analysis of neural data*, Journal of Neurophysiology **94** (2004), 8–25.

[45] R.E. Kass, V. Ventura, and C. Cai, *Statistical smoothing of neuronal data*, Network-Computation in Neural Systems **14** (2003), no. 1, 5–16.

[46] R.C. Kelly, M.A. Smith, R.E. Kass, and T.-S. Lee, *Local field potentials indicate network state and account for neuronal response variability*, Journal of Computational Neuroscience **29** (2010), 567–579.

[47] T Lu and X Wang, *Information content of auditory cortical responses to time-varying acoustic stimuli*, Journal of Neurophysiology **91** (2003), 301–313.

[48] P Maldonado, C Babul, W Singer, E Rodriguez, D Berger, and S Grün, *Synchronization of neuronal responses in primary visual cortex of monkeys viewing natural images*, Journal of Neurophysiology **100** (2008), 1523–1532.

[49] L. Martignon, G. Deco, K. Laskey, M. Diamond, W. Freiwald, and E. Vaadia, *Neural coding: higher-order temporal patterns in the neurostatistics of cell assemblies*, Neural Computation **12** (2000), no. 11, 2621–2653.

[50] L Martignon, H von Hasseln, S Grün, A Aertsen, and G Palm, *Detecting higher-order interactions among the spiking events in a group of neurons*, Biological Cybernetics (1995), 69–81.

[51] A Moekeichev, M Okun, O Barak, Y Katz, O Ben-Shahar, and I Lampl, *Stochastic emergence of repeating cortical motifs in spontaneous membrane potential fluctuations in vivo*, Neuron **53** (2007), no. 3, 413–425.

[52] Z Nádasdy, H Hirase, A Czurkó, J Csiscvari, and G Buzsáki, *Information content of auditory cortical responses to time-varying acoustic stimuli*, Journal of Neurophysiology **91** (2003), 301–313.

[53] M W Oram, M C Wiener, R Lestienne, and B J Richmond, *Stochastic nature of precisely timed spike patterns in visual system neural responses*, Journal of Neurophysiology **81** (1999), 3021–3033.

[54] Q Pauluis and S N Baker, *An accurate measure of the instantaneous discharge probability, with application to unitary joint-event analysis*, Neural Computation **12** (200), no. 3, 647–669.

[55] A Pazienti, M Diesmann, and S Grün, *Lecture notes in computer science: Advances in brain, vision, and artificial intelligence*, ch. Bounds on the ability to destroy precise coincidences by spike dithering, pp. 428–437, Springer Berlin, Heidelberg, Germany, 2007.

[56] A Pazienti, P E Maldonado, M Diesmann, and S Grün, *Effectiveness of systematic spike dithering depends on the precision of cortical synchronization*, Brain Research **1225** (2008), 39–46.

[57] Y Prut, E Vaadia, H Bergman, I Haalman, H Slovin, and M Abeles, *Spatiotemporal structure of cortical activity: properties and behavioral relevance*, Journal of Neurophysiology **79** (1998), 2857–2874.

[58] A Renart, J de la Rocha, P Bartho, L Hollander, N Parga, A Reyes, and K D Harris, *The asynchronous state in cortical circuits*, Science **327** (2010), 587–590.

[59] F Rieke, D Warland, R van Steveninck, and W Bialek, *Spikes: Exploring the neural code*, MIT Press, Cambridge, Massachusetts, 1997.

[60] A Rokem, S Watzl, T Gollisch, M Stemmler, A V M Herz, and I Samengo, *Spike-timing precision underlies the coding efficiency of auditory receptor neurons*, Journal of Neurophysiology **95** (2006), 2541–2552.

[61] E. Salinas and T.J. Sejnowski, *Gain modulation in the central nervous system: where behavior, neurophysiology, and computation meet*, The Neuroscientist **7** (2001), no. 5, 430.

[62] E. Schneidman, M.J. Berry, R. Segev, and W. Bialek, *Weak pairwise correlations imply strongly correlated network states in a neural population*, Nature **440** (2006), no. 7087, 1007–1012.

[63] S. Shinomoto, H. Kim, T. Shimokawa, N. Matsuno, S. Funahashi, K. Shima, I. Fujita, H. Tamura, T. Doi, K. Kawano, et al., *Relating neuronal firing patterns to functional differentiation of cerebral cortex*, PLoS computational biology **5** (2009), no. 7, e1000433.

[64] T Shmiel, R Drori, O Shmiel, Y Ben-Shaul, Z Nadasdy, M Shemesh, M Teicher, and M Abeles, *Neurons of the cerebral cortex exhibit precise interspike timing in correspondence to behavior*, Proceedings of the National Academy of Sciences **102** (2005), no. 51, 18655–18657.

[65] T Shmiel, R Drori, O shmiel, Y Ben-Shaul, Z Nadasdy, M Shemesh, M Teicher, and M Abeles, *Temporally precise cortical firing patterns are associated with distinct action segments*, Journal of Neurophysiology **96** (2006), 2645–2652.

[66] MA Smith and A Kohn, *Spatial and temporal scales of correlation in primary visual cortex*, Journal of Neuroscience **28** (2008), no. 48, 12591–12603.

[67] E Stark and M Abeles, *Unbiased estimation of precise temporal correlations between spike trains*, Journal of Neuroscience Methods **179** (2009), 90–100.

[68] B Staude, S Grün, and S Rotter, *Higher-order correlations in non-stationary parallel spike trains: statistical modeling and inference*, Frontiers in Computational Neuroscience **4** (2010), no. 16, 1–17.

[69] P Tiesinga, JM Dellous, and TJ Sejnowski, *Regulation of spike timing in visual cortical circuits*, Nature Reviews Neuroscience **9** (2008), 97–107.

[70] D.J. Tolhurst, J.A. Movshon, and A.F. Dean, *The statistical reliability of signals in single neurons in cat and monkey visual cortex*, Vision Res. **23** (1983), 775–785.

[71] P.J. Uhlhaas, G. Pipa, B. Lima, L. Melloni, S. Neuenschwander, D. Nikolić, and W. Singer, *Neural synchrony in cortical networks: history, concept and current status*, Frontiers in Integrative Neuroscience **3** (2009).

[72] V Ventura, C Cai, and RE Kass, *Statistical assessment of time-varying dependence between two neurons*, Journal of Neurophysiology **94** (2005), 2940–2947.

[73] _____, *Trial-to-trial variability and its effect on time-varying dependence between two neurons*, Journal of Neurophysiology **94** (2005), 2928–2939.

[74] AEP Villa, IV Tetko, B Hyland, and A Najem, *Spatiotemporal activity patterns of rat cortical neurons predict responses in a conditioned task*, Proc Natl Acad Sci USA **96** (1999), 1106–1111.