

Generating Spike Trains with Specified Correlation Coefficients

Jakob H. Macke*

jakob@tuebingen.mpg.de

Max Planck Institute for Biological Cybernetics, 72076 Tübingen, Germany

Philipp Berens

berens@tuebingen.mpg.de

Alexander S. Ecker

alexander.ecker@tuebingen.mpg.de

*Max Planck Institute for Biological Cybernetics, 72076 Tübingen, Germany, and
Baylor College of Medicine, Department of Neuroscience, Houston, TX 77030, U.S.A.*

Andreas S. Tolias

atolias@cns.bcm.edu

Baylor College of Medicine, Department of Neuroscience, Houston, TX 77030, U.S.A.

Matthias Bethge

mbethge@tuebingen.mpg.de

Max Planck Institute for Biological Cybernetics, 72076 Tübingen, Germany

Spike trains recorded from populations of neurons can exhibit substantial pairwise correlations between neurons and rich temporal structure. Thus, for the realistic simulation and analysis of neural systems, it is essential to have efficient methods for generating artificial spike trains with specified correlation structure. Here we show how correlated binary spike trains can be simulated by means of a latent multivariate gaussian model. Sampling from the model is computationally very efficient and, in particular, feasible even for large populations of neurons. The entropy of the model is close to the theoretical maximum for a wide range of parameters. In addition, this framework naturally extends to correlations over time and offers an elegant way to model correlated neural spike counts with arbitrary marginal distributions.

1 Introduction ---

Neurons primarily communicate with each other by generating sequences of action potentials. In order to understand how neural systems perform

*Jakob Macke and Philipp Berens contributed equally to this work.

computations and process sensory information, we need to understand the structure of firing patterns in large populations of neurons. These spike trains can exhibit substantial correlations both in time and across different neurons. In the visual pathway, for example, various types of correlations have been observed at many stages, including the retina (Mastrorade, 1983; Meister, Lagnado, & Baylor, 1995), lateral geniculate cortex (Alonso, Usrey, & Reid, 1996), striate (Reich, Mechler, & Victor, 2001; Kohn & Smith, 2005; Montani, Kohn, Smith, & Schultz, 2007), and extrastriate cortical areas (Gawne & Richmond, 1993; Zohary, Shadlen, & Newsome, 1994; Bair, Zohary, & Newsome, 2001).

Assessment of their functional implications has been complicated by the fact that correlations in neural recordings can occur for very different reasons. For instance, if a cell has a tendency to fire bursts of multiple spikes, this can be due to an intrinsic mechanism that makes the cell more sensitive for a brief period of time after the generation of an action potential. Alternatively, bursts can be caused by temporal correlations in the input to the cell. Without our knowing all the inputs, these two different possibilities cannot be distinguished. It is even more difficult to determine whether correlated firing between two neurons is caused by common input to the population or direct synaptic couplings. Here, we seek to provide a flexible framework for efficient simulation and modeling of correlated spike trains with specified pairwise correlations even if their origin and source are unknown. Our framework can be used to study the impact of correlations on various timescales. In particular, it can be used to test and extend insights of previous analytical work (Abbott & Dayan, 1999), derived under simplifying assumptions, to parameter regimes much closer to experimentally measured data (Ecker, Berens, Bethge, Logothetis, & Tolias, 2007).

On short timescales up to several milliseconds, neurons can either fire a spike or not, and thus the activity of a group of neurons can be described as a binary pattern (see Figure 1, top). Apart from the mean activity of each cell, correlations between neurons influence the frequency of each of the possible patterns occurring. Maximum entropy modeling can be used to investigate whether the knowledge of the means and pairwise correlations is sufficient to capture the statistical structure of population activity. Schneidman, Berry, Segev, and Bialek (2006) and Shlens et al. (2006) showed that the maximum entropy distribution for given pairwise interactions, the Ising model, can accurately predict the frequency of many spike patterns in a population. However, the Ising model is limited by its poor scalability to high dimensions (Schneidman et al., 2006; Shlens et al., 2006; Tang et al., 2008) and thus is of limited use when dealing with large populations of neurons. We model binary neural population patterns using the dichotomized gaussian (DG) distribution (Pearson, 1909; Emrich & Piedmonte, 1991; Cox & Wermuth, 2002), which is fully specified by the mean and covariance of the data. This makes it possible to model binary neural population patterns even for previously intractably large populations. While the DG does not

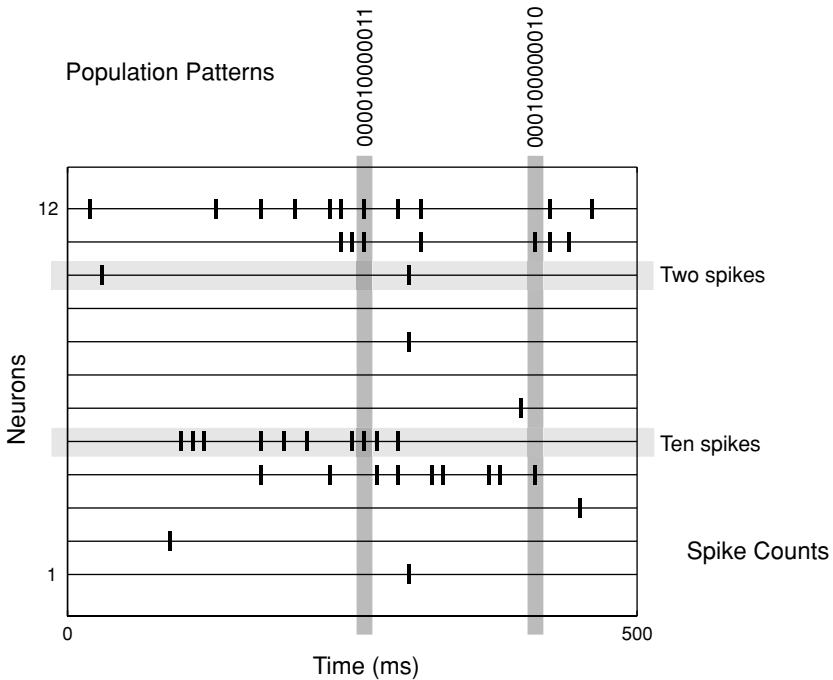


Figure 1: At any given point in time, neurons can either fire a spike or not. Thus, spiking patterns across neurons (dark gray, vertical) can be described as binary vectors (top), and their statistics can be captured by multivariate binary probability distributions. Spiking patterns of one neuron across time (light gray, horizontal) can also exhibit substantial structure. For example, spikes of the lower highlighted neurons seem to occur more often in short bursts of three or four spikes. Correlations between the spikes of two neurons can lead to correlations in the total number of spikes emitted in a given time interval, the spike count (right). We describe a flexible binary population model, which can be used to capture correlations between spikes across both time and neurons, as well as correlations in their spike counts.

have maximum entropy, its entropy is close to the theoretical maximum for a wide range of parameters (Bethge & Berens, 2008).

We also show that essentially the same model can be used to study effects of temporal correlations in spike trains as they arise as a consequence of bursting activity or refractory periods (Sherman, 2001). To this end, we demonstrate that taking temporal correlations into account improves the similarity between simulated spike trains and spike trains recorded from a neuron in the primary visual cortex of a macaque.

Alternative algorithms for generating correlated binary random variables can be applied only to low-dimensional problems (Bahadur, 1961;

Gange, 1995; Lee, 1993) or to correlation matrices that have a specific structure (Lunn & Davies, 1998; Park, Park, & Shin, 1996; Gütig, Aharonov, Rotter, & Sompolinsky, 2003), such as constant correlation between all pairs of neurons. Qaqish (2003) presents a systematic review of these techniques and suggests an alternative method based on conditional linear distributions. Niebur (2007) has proposed a method for the generation of spike trains by specifically assuming common input, that is, inserting spikes from a latent process into independent spike trains. Given the means and pairwise correlations, this algorithm requires the specification of additional parameters that influence the higher-order properties of the distribution. In section 2.2, we include a comparison of the Ising model, the DG, and Niebur's algorithm.

Frequently the activity of neurons is not investigated with fine temporal precision but summarized as the number of spikes in a trial of a given length (see Figure 1, right). In modeling studies, these spike counts have often been assumed to be Poisson distributed and neurons to be independent. However, correlations between spike counts have been reported in various visual areas (Bair et al., 2001; Kohn & Smith, 2005). In this situation, it is desirable to have models that can be fit to both the measured interneuron correlations and the observed single cell spike count histograms. Unfortunately, fitting maximum entropy models with these constraints is computationally difficult.

Starting with a binary model, we describe a technique to sample spike counts from a correlated neural population by summing over binary spike trains. In this way, we gain insight into how correlations in spike counts arise from correlations in the underlying binary processes and how the Fano factor of the spike count is related to temporal correlations. Because experimentally measured spike counts are often not Poisson distributed, we extend this technique to arbitrary marginal statistics with specified correlations. We demonstrate its capabilities by modeling a population of simultaneously recorded neurons from the primary visual cortex of a macaque and show how a model with correlations accounts for the data far better than a model assuming independence.

Implementations of our algorithms for generating spike trains using the dichotomized gaussian distribution and generating correlated Poisson samples in Matlab are available from our Web site: www.kyb.mpg.de/bethgegroup/code/efficientsampling.

2 Generating Spike Trains with Defined Pairwise Correlation Structure

2.1 Sampling from a Multivariate Binary Distribution. A population of N neurons, each of which either does or does not spike at any given time, can be in any of 2^N different states. Estimating the full joint distribution over these 2^N patterns becomes quickly infeasible for increasing numbers

of neurons N . However, it is often possible to measure the firing rates and pairwise correlations—the first and second moments. Our goal is to generate spike trains $X_i(t)$ from a population of N neurons where first and second moments have been specified.

We assume the spike trains to be discretized into sufficiently small time bins, such that each time bin contains at most one spike, and set $X_i(t) = 1$ if neuron i fires a spike in bin t and $X_i(t) = 0$ otherwise. Thus, we want to draw samples from the joint distribution of a set of N correlated Bernoulli random variables with mean $r_i = \langle X_i(t) \rangle_t$ and pairwise covariances $\Sigma_{ij} = \langle X_i(t)X_j(t) \rangle_t - \langle X_i(t) \rangle_t \langle X_j(t) \rangle_t$. For the moment, we assume that there are no correlations across time (i.e., that $X(t)$ and $X(s)$ are independent for $s \neq t$), but we will generalize our method to incorporate these in section 2.3.

We now describe the dichotomized gaussian distribution (DG) that is well suited for this task of generating binary random vectors with specified correlations. A sample from the DG distribution is obtained by first drawing a sample from a N -dimensional gaussian random variable U and then thresholding it into 0 and 1:

$$X_i = 1 \quad \text{iff } U_i > 0 \quad \text{where } U \sim \mathcal{N}(\gamma, \Lambda) \quad (2.1)$$

The thresholding operation will change the moments, so X will, in general, not have the same mean and covariance as U . However, the effect of the truncation can be calculated (Block & Fang, 1988) and corrected for: we can choose the mean γ and covariance Λ of U such that after truncation, X has the desired moments \mathbf{r} and Σ .

Assuming (without loss of generality) unit variances for U , that is, $\Lambda_{ii} = 1$, the mean spiking probabilities r and covariance Σ of X are given by

$$\begin{aligned} r_i &= \Phi(\gamma_i) \\ \Sigma_{ii} &= \Phi(\gamma_i)\Phi(-\gamma_i) \\ \Sigma_{ij} &= \Psi(\gamma_i, \gamma_j, \Lambda_{ij}), \end{aligned} \quad (2.2)$$

where $i \neq j$ and $\Psi(x, y, \lambda) = \Phi_2(x, y, \lambda) - \Phi(x)\Phi(y)$. Here, Φ is the cumulative distribution of a univariate gaussian with mean 0 and variance 1, and $\Phi_2(x, y, \lambda)$ is its bivariate counterpart with correlation λ .

The mean values γ_i can be found by inverting equation 2.2: $\gamma_i = \Phi^{-1}(r_i)$. Determining Λ_{ij} generally requires finding a suitable value such that $\Sigma_{ij} - \Psi(\gamma_i, \gamma_j, \Lambda_{ij}) = 0$. This equation can be solved for any Σ_{ij} , which is the covariance between two binary random variables, as shown in appendix B. The inversion can be efficiently done by numerical computations, since the function is monotonic in Λ_{ij} and has a unique zero crossing, and we know that the solution is in the interval $[-1, 1]$. For example, if one wants to model two neurons with firing rates $r_1 = 0.5$, $r_2 = 0.25$ and covariance $\Sigma_{12} = 0.1$, the parameters of the DG are given by $\gamma_1 = 0$, $\gamma_2 = -0.67$, and $\Lambda_{12} = 0.39$.

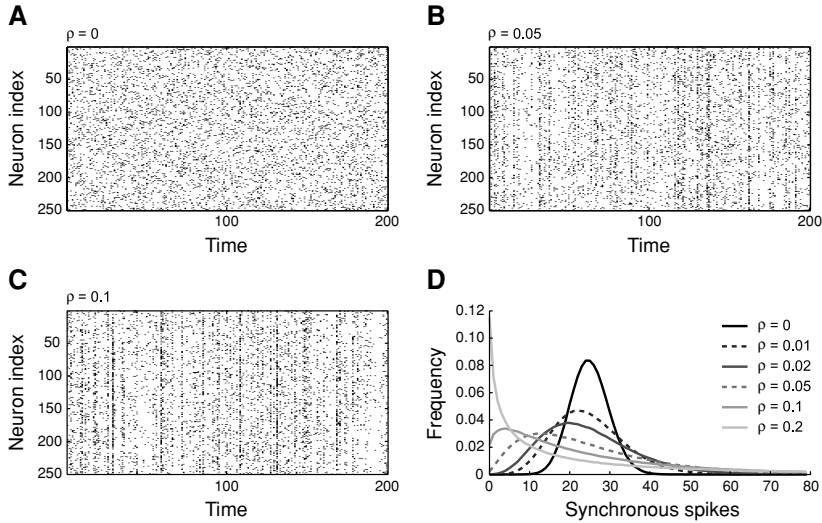


Figure 2: Raster plots of synthetically sampled multineuron firing patterns. The neurons depicted in panels A–C all have the same constant firing rate of 0.1 spike per bin and vary only in their correlation structure. (A) All neurons are independent. (B, C) The pairwise correlation between any pair of neurons is 0.05 and 0.1, respectively. Patterns in which many neurons fire simultaneously occur more frequently with increasing correlation strength. (D) Shows how the probability of observing k out of the 250 neurons to spike simultaneously varies with correlation.

The parameter Λ_{ij} can be interpreted as an alternative characterization of the strength of correlation of neurons i and j . It is -1 if the correlation is as negative as possible and 1 if the correlation is maximal. The (Pearson) correlation coefficient between two neurons modeled as binary random variables does not have this property, as its range is constrained by their mean firing probabilities (see below).

Once γ and Λ are determined, sampling from the DG distribution is as efficient as sampling from a multivariate normal. Importantly, the DG has no free parameters for a given mean and covariance. As the latent variable is a multivariate normal, conditioning and marginalizing on the latent variable can be performed conveniently.

To demonstrate the effect of increasing the correlation strength, we generated synthetic spike trains from a population of 250 neurons (see Figure 2). We simulated a segment of length 250 bins and mean firing rate 0.1 spike per bin. We simulated populations with independent neurons and with pairwise correlations $\rho = 0.05$ and $\rho = 0.1$. With increasing correlation strength, we visually find more patterns where many neurons are active

in any given bin (compare Figure 2A to 2C). This is also true quantitatively: patterns with many synchronously active neurons are found dramatically more often in strongly correlated populations than in the independent, for which only patterns with about 15 to 35 spikes occur frequently (see Figure 2D).

Care is needed when simulating random variables with covariance matrices that are not directly estimated from experimental data but constructed by other considerations: not every positive definite symmetric matrix can be used as the covariance matrix of a multivariate binary distribution. For example, for two binary random variables X and Y with means p and q , the covariance is bounded by

$$\max\{-pq, -(1-p)(1-q)\} \leq \text{Cov}(X, Y) \leq \min\{(1-q)p, (1-p)q\}.$$

Furthermore, in three or more dimensions, this condition is only necessary, not sufficient. It is not hard to find symmetric, positive definite matrices that satisfy it (for any pair of dimensions) but for which no suitable binary random variable exists (see appendix B). This is a general property of binary distributions, not a shortcoming of specific models. In general, no easy rule for determining the feasibility of a covariance matrix is known.

However, the DG model provides a convenient sufficient condition; it can be used to show that a given matrix Σ is a valid binary covariance matrix. We can calculate the covariance matrix Λ of the underlying gaussian for r and Σ using equation 2.2, and if Λ is positive definite, we have shown the existence of a suitable binary random variable by explicitly constructing it. A limitation of the DG model is that it does not exist for all feasible covariance matrices Σ . While equation 2.2 always has a solution, the resulting Λ might not be positive definite and thus cannot be the covariance matrix of the latent gaussian (Chaganty & Joe, 2006). This can occur when there are strong negative correlations. However, the correlations found in studies of binary activation patterns so far are mostly weak and positive (Schneidman et al., 2006; Shlens et al., 2006), such that they can often be modeled by a DG without further effort. If a desired correlation structure results in a matrix Λ that is nonpositive definite, it is possible to find the closest valid correlation matrix $\bar{\Lambda}$ using an efficient iterative procedure described by Higham (2002).

2.2 Higher-Order Correlations and Maximum Entropy. In the previous section, we introduced a model with given first and second moments. In order to specify a multivariate binary distribution of more than two dimensions uniquely, however, the information provided by the first and second moments is not sufficient. The degrees of freedom of an N -dimensional distribution grow exponentially like $2^N - 1$, whereas the first and second moments can constrain only $N(N+1)/2$ of them. As measuring all higher-order correlations becomes rapidly prohibitive with increasing N , it

is important to address the question of what model should be chosen if the higher-order correlations are unknown.

The fundamental strategy to cope with this problem is regularization (Chen & Haykin, 2002), which can ultimately be traced back to Occam's razor. The maximum entropy principle (Jaynes, 1978) provides an elegant way of regularizing the estimate of a distribution, as it minimizes the worst-case risk with respect to the log loss (Grünwald & Dawid, 2004). To what extent risk minimization of the log loss can be seen as the most favorable strategy is often not clear, and other regularization properties might be preferable in specific cases. For instance, the DG model has the nice regularization property that the marginal statistics of any subspace can be estimated independent of the complementary subspace. This is not the case for the Ising model. Nevertheless, good regularization methods usually yield solutions with large entropy.

Recently, the maximum entropy approach has gained much interest in descriptive spike train analysis (Schneidman et al., 2006; Shlens et al., 2006; Nirenberg & Victor, 2007; Tang et al., 2008). However, use of the second-order maximum entropy (MaxEnt) distribution, the Ising model, is computationally challenging in large populations. Exact parameter fitting is possible only for very small populations of neurons. For larger populations, approximate algorithms are computationally costly or even infeasible when big populations of neurons are investigated (for algorithmic advances, see Broderick, Dudik, Tkacik, Schapire, & Bialek, 2007). Even after the parameters are identified, one has to use Markov chain Monte Carlo (MCMC) techniques to draw samples from an Ising model. MCMC sampling is exact in the limit of infinitely many samples, but only approximate for chains of practical sample size. If the sample size is not big enough, samples in the chain are correlated with each other, and the entropy of the obtained samples is smaller than it should be. This makes sampling computationally costly in situations where specialized algorithms (Swendsen & Wang, 1987; Wolff, 1989) are not applicable. In contrast, the parameters of a DG model are easy to fit using equation 2.2, and exact samples can be directly generated by thresholding a multivariate gaussian.

To study the higher-order correlations of the DG distribution, we compare it in this section against the maximum entropy distribution with the same second-order moments and an algorithm proposed by Niebur (2007) for generating correlated spike trains (which we refer to as the latent spike train algorithm). While the maximum entropy distribution effectively minimizes the magnitude of higher-order correlations (Watanabe, 1960; Amari, 2001), the latent spike train algorithm is designed to explicitly model common input between groups of neurons. Therefore, it tends to generate samples with higher-order correlations, for example, with synchronous events of concerted firing across many neurons. Correlations between binary spike trains are induced by generating additional hidden spike trains and randomly inserting their spikes into the actual spike

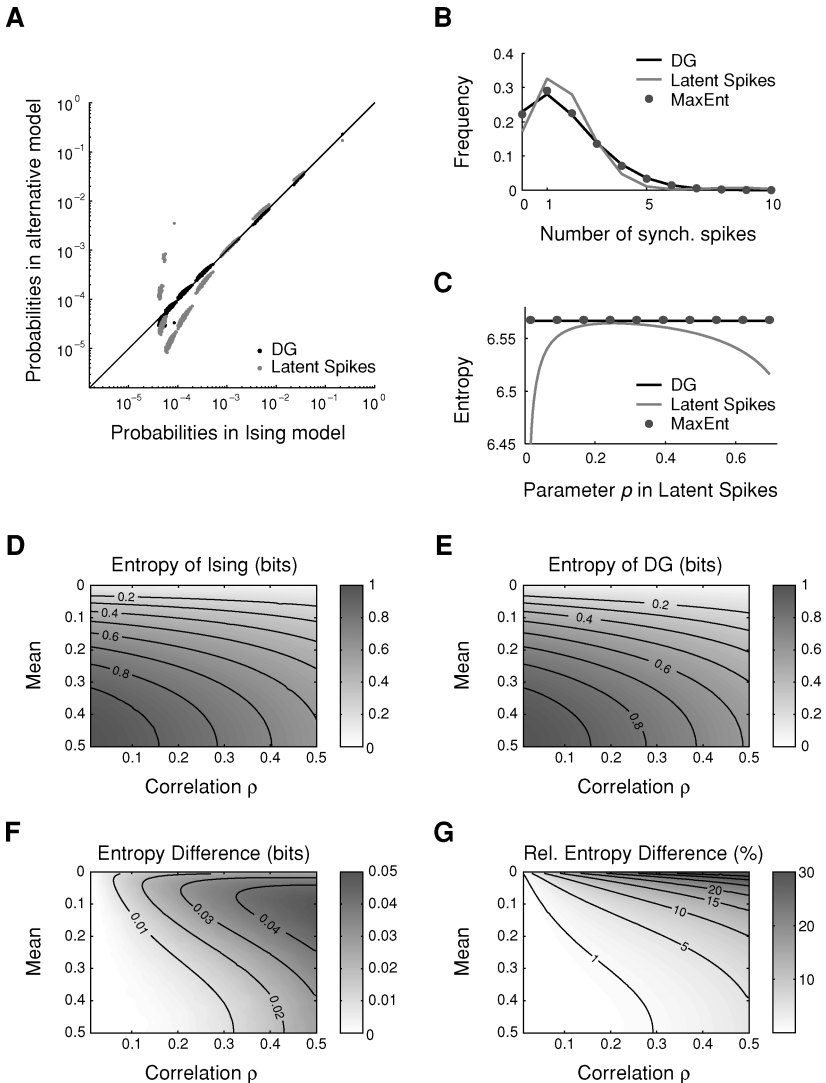
trains. In this sense, it can be seen as complementary to maximum entropy approaches.

Figure 3 shows a comparison of the maximum entropy distribution, the DG, and Niebur's latent spike train algorithm with parameter $p = 0.022$ (the firing probability of the latent spike train). We considered a population of 10 neurons with firing rates equally spaced on $[0.15, 0.20]$ and constant pairwise covariance 0.01 between each pair of neurons. We calculated the probabilities of each of the possible 2^{10} binary words of spikes and no spikes that the population could exhibit. While the probabilities of the MaxEnt model and the DG model are very similar, as quantified by the Jensen-Shannon divergence (Schneidman et al., 2006), which is 3.3×10^{-4} (see Figure 3A), there are noticeable differences between the MaxEnt distribution and the distribution induced by the algorithm with latent spike trains (Jensen-Shannon divergence 2.0×10^{-2}) in the population considered here. For example, the probability of all neurons being silent simultaneously is 0.222 under the MaxEnt model and 0.230 in the DG model but 0.171 in the latent spike train algorithm (with this choice of parameters). If neurons were independent, the probability of silence would be 0.146.

We explored how these differences change with the free parameter p in the latent spike model and found that for some choices of p , the entropy of the distribution of patterns generated by the latent spike algorithm is close to that of the DG model, while it is substantially smaller for others (see Figure 3C). Thus, even a model that explicitly models common input can lead to data with close to maximal entropy. However, the algorithm is not merely consistent with correlations arising from common input; it actually requires the user to choose a specific form of common input by specifying up to $\frac{N}{2}$ parameters. The choice of parameters can substantially change the entropy of the generated data. In the case considered here (constant pairwise covariance), it might be argued that one could identify values of the parameter p for which the entropy is reasonably big (see Figure 3C). In the general case, however, p will be high-dimensional and identification of suitable parameters nontrivial. In contrast, the DG model is fully specified by the mean and covariance, provides an efficient alternative to the Ising model, and allows one to generate spike trains that have close to maximum entropy. As the normal distribution is the MaxEnt distribution for a given mean and covariance on the real numbers, this might come as no surprise.

To see whether the close match between the DG and the Ising model also persists in high-dimensional cases, we compared the Ising model and the DG in a simple example of a population of 100 neurons with constant firing rate r and pairwise correlation coefficient ρ between any pair of neurons. In this simple model, all firing patterns with the same number of spikes are equally probable, so entropies can be calculated without having to sum over all binary patterns. We varied r and ρ from 0 to 0.5 and compared the entropies of the resulting models. For this population, the

entropy of the two models is very similar across a wide range of moments (see Figures 3D–3G), and the difference in entropy never exceeds 0.05 bits per dimension. It is most pronounced for models with very small firing rates but extremely strong correlations. The difference in entropy is also equal to the Kullback-Leibler divergence $KL(P_{DG}, P_{\text{Ising}})$ between the DG and the Ising model (Grünwald & Dawid, 2004). However, it is not guaranteed that the probabilities that the two models assign to each of the different patterns are very similar. The exact relationship between the Ising model and the



DG is beyond the scope of this letter. For further comparisons, see Bethge and Berens, 2008.

The DG, the Ising model, and the latent spike train algorithm can be seen as three complementary means of modeling multineuron firing patterns that differ in both their higher-order statistics and their practical usability. The question of which model provides the best description of neural activity is ultimately an empirical one.

2.3 Spike Trains with Temporal Correlations. Until now, we have assumed there are no correlations across time. Thus, the population activities $X(s)$ and $X(t)$ at different time points s and t are independent random variables. Here we extend our algorithm to capture temporal correlations, such as induced by the refractory period, bursting behavior, or other firing patterns of neurons. This can be done in a straightforward manner. Rather than basing our model on temporally independent gaussians $U(t)$, we obtain binary samples $X(t)$ by thresholding elements of a gaussian time series $V(t)$ that has temporal correlations:

$$X_i = 1 \quad \text{iff } V_i > 0 \quad \text{where} \quad V(t) \sim \mathcal{N}(\gamma(t), \Lambda(s, t)),$$

where $\gamma(t)$ is the mean function and $\Lambda(s, t)$ the covariance function. It should be noted that if temporal correlations are strongly negative, Λ might not be positive definite, in which case an approximation technique might be necessary (Higham, 2002; Ghosh & Henderson, 2003).

Figure 3: (A–C) Empirical comparison of three binary distributions with identical second-order structure. We considered the responses of 10 neurons with firing rates on the interval $[0.15, 0.20]$ and pairwise covariance 0.01. (A) A scatter plot of the log probabilities of each of the 2^{10} binary words when drawn from a MaxEnt distribution, DG, and Niebur’s latent spike train algorithm with parameter $p = 0.022$. In this population, the DG is very similar to the MaxEnt model, as can be seen from the fact that all of its values are on or near the diagonal. (B) The probability of observing n simultaneous spikes in each of the three models from A. The MaxEnt and the DG models are almost indistinguishable. (C) The entropy of the MaxEnt model ($H = 6.568$), the DG model ($H = 6.567$), and the latent spike train algorithm for p ranging from 0.16 to 0.70 (H in $[6.443, 6.564]$). The entropy of the latent spike model depends strongly on the choice of parameters. (D–G) Comparison of two population models with 100 neurons and identical second-order structure. Each population has constant firing rates r and constant pairwise correlation ρ . We varied r and ρ from 0 to 0.5 and calculated the entropy (per neuron) of both a DG and an Ising model with these moments (D, E). We also calculated the difference in entropy ($H_{\text{ising}} - H_{\text{DG}}$, F) and the relative difference in entropy ($\frac{H_{\text{ising}} - H_{\text{DG}}}{H_{\text{ising}}}$, G).

In the general case, modeling temporal correlations over M time bins and N neurons requires estimation of about $\frac{1}{2}M^2N^2$ parameters. Since this becomes intractable relatively quickly on limited data, we here consider a special case and assume $V(t)$ to be stationary. Therefore, the mean firing rate $\gamma(t) = \gamma$ is assumed to be constant, and the temporal covariances $\Lambda(s, t) = \Lambda(s - t)$ are only a function of the distance τ between two time bins (the auto- and cross-covariance functions of the N neurons). For a given τ , this is an $N \times N$ matrix. We assume that there are no correlations beyond a delay of τ_{K-1} , in which case temporal covariances need only to be estimated at K different time lags $\tau_0 = 0, \dots, \tau_{K-1}$. In this special case, there are only about KN^2 parameters to be estimated.

We demonstrate the benefit of taking temporal correlations into account by modeling the spike trains of a neuron recorded from the primary visual cortex of an awake behaving monkey using tetrodes (see Figure 4A and appendix A). The neuron was repeatedly stimulated with an oriented grating at eight different orientations in random order, and 15 trials at the neuron's preferred orientation are shown. Spikes were binned at 6 ms resolution, and the autocovariance was computed for up to $\tau_{10} = 60$ ms. While the independent model (see Figure 4B) fails to reproduce the regular spiking patterns of this neuron, taking into account temporal correlations makes the sampled data (see Figure 4C) and the real data nearly indistinguishable. This is also illustrated by the close match between the interspike interval distributions shown in Figure 4D.

Sometimes it is useful to generate spike trains with a given temporal correlation structure on the fly, for example, for online simulations or if the dimensionality MN is too large to sample directly. Since the conditional distribution of the latent variable $P(V(t) | V(t - \tau_1), \dots, V(t - \tau_{K-1}))$ is again a gaussian distribution that can be derived in closed form, samples can be generated iteratively one at a time by conditioning on the last K samples drawn from the hidden gaussian:

$$V(t) \sim \mathcal{N}(\gamma + CB^{-1}[V(t - \tau_1), \dots, V(t - \tau_{K-1})] - \gamma), A - CB^{-1}C^T),$$

where $A = \Lambda(0)$, B is a block Toeplitz matrix of size $(K - 1)N$ formed by $\Lambda(0), \dots, \Lambda(\tau_{K-2})$ with $\Lambda(0)$ along the main diagonal, and $C = [\Lambda(\tau_1), \dots, \Lambda(\tau_{K-1})]$, obtained by concatenating $K - 1$ covariance matrices of size $N \times N$. Since the matrices A , B , and C are constant and can be pre-computed, each step in this iterative sampling procedure has a complexity of $O(N^2K)$, eliminating superlinear terms in the desired number of time bins. Note that we obtain an especially simple case when only one neuron is modeled. Then $\Lambda(\tau_i)$ is no longer a matrix but a scalar, and the previous equation reduces to a very simple form.

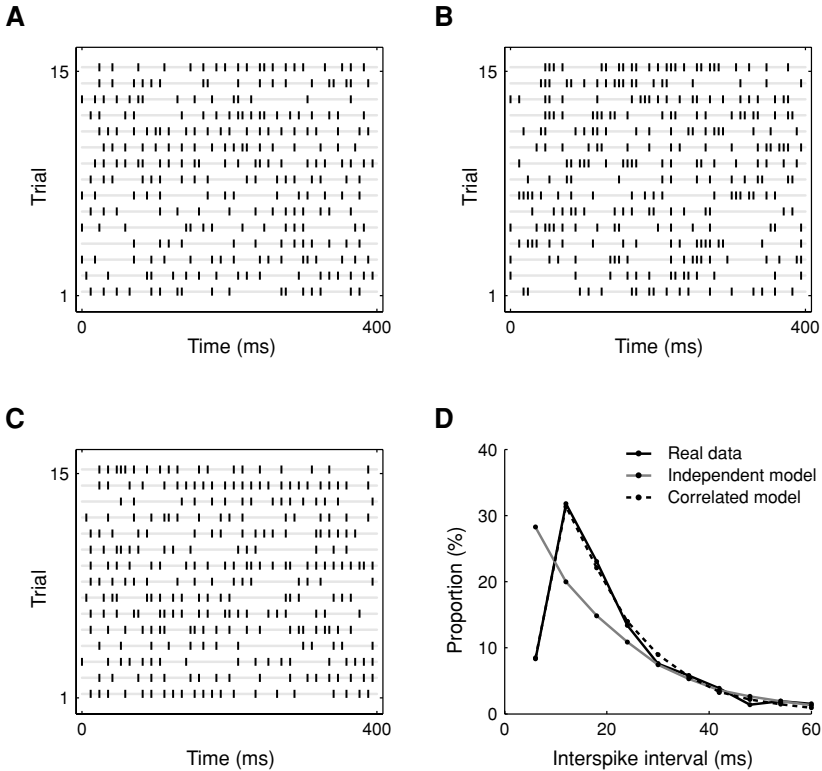


Figure 4: Spike train rasters and interspike interval distributions. (A) Sustained response of a single neuron during 15 trials of visual stimulation with an oriented bar. (B) Simulated spike trains using a constant firing rate matching that in A and no temporal correlations between time bins. (C) Simulated spike train matching both firing rate and autocorrelation of the real neuron. Note how this reproduces the cell’s regular firing pattern quite well compared to B. (D) Interspike interval distribution for real and sampled data. While the independent model (B) overestimates the spike doublets in successive bins, the model taking into account temporal correlations is able to reproduce the correct interspike interval distribution.

3 Generation of Correlated Spike Counts

3.1 Poisson Spike Counts. Often the exact timing of individual spikes is not of interest, and only the total number of spikes in a given interval, the spike count, is analyzed. For example, classical studies have described the response properties of neurons in primary visual cortex and area MT by plotting the average number of spikes emitted during stimulation as

a function of a single variable, such as orientation or direction of motion (Hubel & Wiesel, 1968). Accordingly, many theoretical studies on population coding take only spike counts into account (Pouget, Dayan, & Zemel, 2003, for review), and any information about the exact timing of spikes is discarded. Correlations in spike counts can, for example, arise when the firing activity of multiple neurons is modulated by internal states of the system, such as up and down states. In this case, there will be trials on which the firing rates of many neurons are enhanced. Once responses are averaged over trials, such a modulation will manifest itself in (positive) trial-by-trial correlations across neurons (Kass & Ventura, 2006).

Spike count statistics can be modeled by splitting up a time interval $[0, T]$ into several small intervals, where their length $\Delta t = \frac{T}{M}$ is chosen so small that each of these intervals contains at most one spike. We set X_{ik} to 1 if the i th neuron fired a spike in bin k . Then the spike count vector Y in the window $[0, T]$ is calculated by summing up the M random variables: $Y_i^{(M)} = \sum_{k=1}^M X_{ik}$. If there are no temporal correlations and each X_{ik} has the same firing probability $p_i = \mu_i \Delta t$, then $Y_i^{(M)}$ has a multivariate binomial distribution with parameters p_i and M . It is convenient to consider the limit of infinitesimally small bins. Then $Y = \lim_{M \rightarrow \infty} Y^{(M)}$ has a multivariate Poisson distribution with mean $\mu_i T$ (Kawamura, 1979). Thus, we can obtain (approximate) samples from a multivariate Poisson distribution by simulating the underlying binary process using the DG distribution (see section 2.1) and summing over it. In the absence of temporal correlations, the covariance of two Poisson random variables can be computed from the covariance of the binary processes from which they are constructed:

$$\begin{aligned} \text{Cov}(Y_i^{(M)}, Y_j^{(M)}) &= \text{Cov}\left(\sum_{l=1}^M X_{il}, \sum_{k=1}^M X_{jk}\right) \\ &= \sum_{k=1}^M \text{Cov}(X_{ik}, X_{jk}). \end{aligned}$$

Interestingly, this view of spike counts implies that the resulting Poisson random variables have pairwise correlations that are nonnegative. The covariance of two Bernoulli random variables with means q and w is bounded below by $-qw$. In the limit, the firing probability of each individual Bernoulli goes to zero, and the lower bound on the covariance between the sums approaches zero:

$$\begin{aligned} \text{Cov}(Y_i^{(M)}, Y_j^{(M)}) &\geq \sum_{k=1}^M (-\mu_i \mu_j \Delta t^2) \\ &= -M \frac{T^2 \mu_i \mu_j}{M^2} = -\frac{T^2 \mu_i \mu_j}{M}. \end{aligned}$$

For large M , this lower bound converges to 0 from below, implying that negative correlations between Poisson spike counts are inconsistent with a construction that assumes underlying spike trains with no temporal structure. We will provide an alternative method for sampling from a distribution with Poisson marginal statistics in section 3.3 that also allows specification of negative correlations.

3.2 The Effect of Temporal Correlations. Experimentally observed spike counts can deviate markedly from a Poisson distribution. For Poisson random variables, the ratio between mean and variance (the Fano factor) is 1, while for cortical neurons, it is often different from 1 (Gur, Beylin, & Snodderly, 1997). Clearly the model of an underlying binary process is valid for spike trains, so these deviations have to originate from our assumption that the binary random variables X_{ik} and X_{il} are independent for different time bins k and l . By relaxing this assumption, we can use the underlying binary spike process to sample spike counts with specified Fano factors.

Suppose that the covariance between any two time bins of the same neuron is given by $v\Delta t^2$ —that there is some nonzero correlation across time. In this case, the variance of the spike count for single neurons is given by

$$\text{Var}(Y_i) = \mu_i T + T^2 v.$$

The Fano factor is $F(Y_i) = 1 + \frac{v}{\mu_i T}$. Thus, positive temporal correlations between single spikes lead to spike counts that are overdispersed in comparison to Poisson distributions.

The assumption that the covariance between any two time bins is the same for each pair of bins is somewhat unrealistic. For example, points that are nearby in time will show stronger correlations than ones that are farther apart. Similarly, it is conceivable that the structure of correlations changes over the time interval, for example, the responses to the onset of the stimulus could be less correlated than the sustained response. In this case, equation 3.1 has to be modified. The firing rate μ_i is replaced by the average of the time-varying rate $\mu_i(t)$ and the covariance v by the temporal average of the covariance function $v(s, t)$:¹

$$\text{Var}(Y_i) = \int_0^T \mu_i(t) dt + \int \int_{s \neq t} v(s, t) ds dt.$$

In this case, the Fano factor of the spike count is determined by the average covariance across time, that is, by the integral of the covariance $v(s, t)$ across all pairs of time points s and t (Ventura, Cai, & Kass, 2005). If the value of the integral is positive, the Fano factor is larger than 1.

¹As before, the rate and covariance functions have to be chosen such that a binary process with these moments exists.

In summary, we can use binary spike trains to simulate spike counts with specified means and covariances, both with Poisson marginal statistics and variants with different Fano factors. In the next section, we generalize this further and show how spike counts with arbitrary marginal distributions can be obtained. In this case, it is easier not to model the underlying spike process explicitly, but rather to draw samples directly from the spike count distribution.

3.3 Spike Counts with Arbitrary Marginal Distributions. In experiments with many trials, it is possible to estimate not only the mean firing rate of each neuron but the complete histogram of its spike counts. In other words, it will be possible to estimate the probability that each neuron fires exactly 1, 2, 3, ... spikes up to a maximal firing rate of M_i spikes. These histograms will not necessarily be well described by a Poisson distribution or some other functional form of a probability distribution. In addition, one can estimate correlations by measuring the covariance of the spike counts, Σ_{ij} . Here, we show how one can simulate synthetic spike counts consistent with the empirical histograms p_{ik} and covariance Σ without any assumptions about the histograms.

We generalize the construction from section 2 (Nelsen, 1987) by first generating samples from a normal random variable with covariance Λ . Whereas in section 2, the gaussian random variable was thresholded to yield binary spike events of 0 or 1, they will now be discretized into M_i different states to yield spike counts of 0, 1, 2, ... M_i . Therefore, we refer to this model as the discretized gaussian (DG) distribution of which the dichotomized gaussian is a special case. As before, each entry of Λ is determined by finding the unique zero crossing of a nonlinear function. Importantly, each Λ_{ij} depends on only the marginal distributions along dimensions i and j and the corresponding entry of the covariance matrix, Σ_{ij} , and is independent of all the other random variables. Therefore, one can efficiently calculate Λ even for populations with many neurons.

Concretely, we want to generate spike counts Y_i for a set of N neurons that have spike count probabilities $P(Y_i = k) = p_{ik}$. We suppose that neuron Y_i will fire no more than M_i spikes. Samples are generated by discretizing an N -dimensional normal random variable U by setting

$$Y_i = k \text{ exactly if } \gamma_{i,k} < U_i \leq \gamma_{i,k+1}.$$

We say that neuron i fired k spikes if the i th coordinate of U is between the two limits $\gamma_{i,k}$ and $\gamma_{i,k+1}$. The limits γ are such that $P(U_i \leq \gamma_{i,k}) = P(Y_i < k)$, that is, $\gamma_{i,k} = \Phi^{-1}(P(Y_i < k))$, for each $k \in 1, 2, \dots, M$. Then the second

moments of Y are given by

$$E(Y_i \cdot Y_j; \lambda) = \sum_{k \leq M_i, l \leq M_j} k l P(Y_i = k, Y_j = l) \tag{3.1}$$

$$= \sum_{j \leq M_i, l \leq M_j} k l \Phi_2(\gamma_{i,k}, \gamma_{j,l}, \gamma_{i,k+1}, \gamma_{j,l+1}, \lambda), \tag{3.2}$$

where $\Phi_2(x_1, y_1, x_2, y_2, \lambda)$ denotes the probability mass of a standard bivariate gaussian with correlation λ in the rectangle with corners (x_1, y_1) and (x_2, y_2) . To ensure that Y has the desired second-order statistics, we have to choose Λ_{ij} such that $\text{Cov}(Y_i, Y_j; \Lambda_{ij}) = \Sigma_{ij}$. This equation determines Λ_{ij} uniquely and can be solved by standard numerical procedures. In principle, the optimization is not harder than in the binary case but requires roughly $M_i M_j$ evaluations of gaussian cumulative distribution functions at each iteration. Thus, the complexity of fitting the DG in this setting is only $(MN)^2$, whereas the state space grows as M^N . As before, the construction can be used only if the resulting matrix Λ_{ij} is indeed positive definite. This approach is related to work based on copulas (Jenison & Reale, 2004) in that we specify the marginal distributions and their desired second-order correlation coefficients. Inference for copula parameters is not straightforward in the case of discrete random variables (Genest & Neslehova, 2007).

Once the initial computations are performed, sampling from the DG is very efficient. Therefore, if a large number of samples is desired or spike counts with many spikes should be simulated, this approach is attractive even for the generation of Poisson random variables. In addition, one can use it to obtain spike counts with Poisson marginals and negative correlations, as well as spike counts with arbitrary marginals.

Figure 5 shows the joint distribution of a pair of Poisson random variables with mean $\mu = 10$ spikes and correlation coefficients $\rho = -0.5, 0,$ and 0.5 . One can see how the correlations influence the total number of spikes emitted by a population (see Figures 5D and 5F). With positive correlations, the distribution has a wider peak, as firing patterns consisting of many spikes across the population are more likely than in the independent case (see Figure 5E). In extracellular recordings, it is often the case that a single electrode picks up spikes from several nearby neurons. If these neurons are positively correlated, the total spike count (i. e., the multiunit activity) will have a higher variability than would be the case for independent neurons. With negative correlations, the distribution of the total number of spikes is more peaked compared to the independent case (see Figure 5F).

To demonstrate the ability of our algorithm to generate artificial spike counts with realistic, possibly non-Poisson distributions, we used a simultaneous recording of 16 well-isolated neurons in primary visual cortex of an awake, behaving macaque monkey (see appendix A). The number of

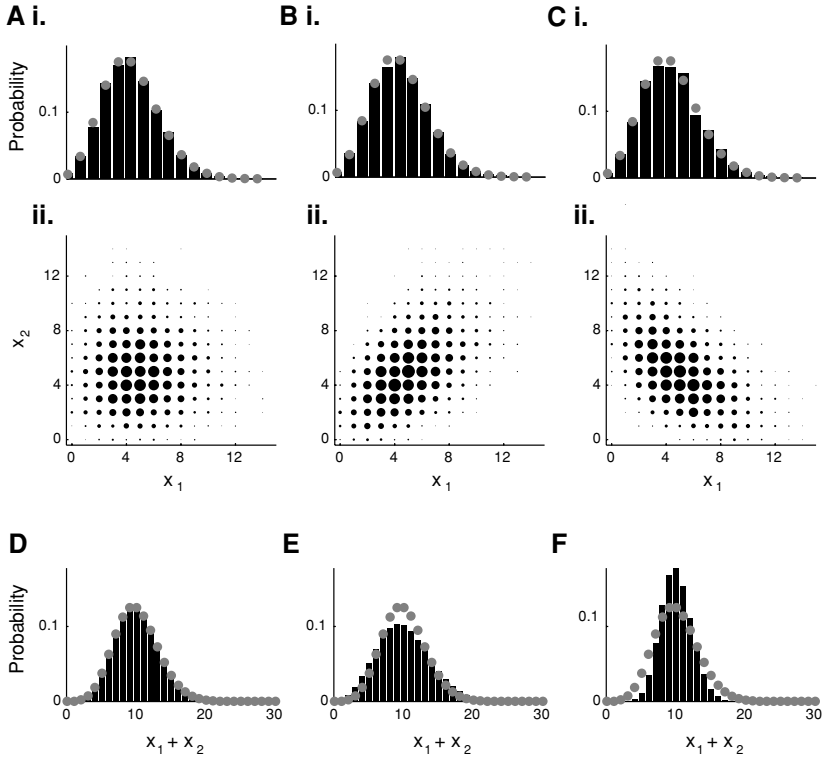


Figure 5: (A–C) Marginals (i) and joint distributions (ii) of a pair of Poisson-distributed random variables with means $r = 5$ and correlation coefficients $\rho = 0, -0.5$, and 0.5 , respectively. Although all marginal distributions look identical, the joint distributions are clearly shaped by the correlation. (Ai–Ci) Black bars are used to show the empirical histograms of the marginal distributions. Overlaid, the probability mass function of a Poisson distribution with mean 5 is shown with gray circles. (Aii–Cii) The joint probability mass of the two random variables is shown. The size of the circles indicates the probability mass. (D–F) Distribution of the sum of the two random variables as in A to C. Again, we overlay the Poisson distribution with mean 10 indicated by the gray circles. While the sum of the two independent variables is again Poisson, positive correlations lead to overdispersion and negative correlations to underdispersion (E, F).

spikes fired by each neuron in the population during each fixation (1306 fixations) was recorded while the monkey was viewing natural images. We estimated the spike count distributions of each neuron, which were far away from Poisson distributions in all cases (as demonstrated in Figure 6A), and the pairwise correlations between neurons. It should be noted that both correlations in the stimulus and variations in fixation time contributed to

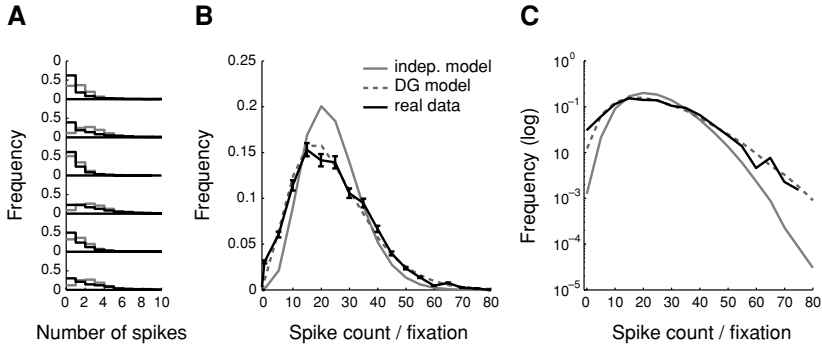


Figure 6: Spike counts in a population of 16 cells in V1. (A) Histograms of the number of spikes fired by six representative cells in each fixation period. Most cells have spike count distributions that are far from a Poisson distribution (best-fitting Poisson plotted in gray). All Fano factors are larger than 1 (mean 3.7, SD 1.3). Therefore, we need to be able to generate synthetic spike counts with non-Poisson marginals to closely match the statistics of the recorded data. (B) Histogram of the total number of spikes fired by the population in each fixation period (black line), bin size 5 spikes. Destroying correlations by shuffling trials leads to a histogram that is substantially different, indicating that correlations have a strong influence on the total spike count statistics (gray line). The histogram of samples from a DG with matching marginals and correlations is very similar to the real data (black dotted line). (C) Same plot as B but on a log scale. Here, the “independent” histogram underestimates spike counts with a very low or very large number of the spikes and overestimates spike counts near the mean (mean spike count = 26.06). The standard deviations are 13.54 for real data, 10.14 for the independent model, and 13.6 for the DG.

the correlations (mean correlation 0.050, SD 0.0523). Although pairwise correlations were nevertheless weak, ignoring the correlations leads to misestimation of the total number of spikes fired across the population during a fixation (see Figures 6B and 6C). The histogram of spike counts is substantially different if correlations are destroyed by shuffling trials. For example, the probability of the total spike count being between 15 and 30 is 0.43 in the data set, but it would be 0.55 if neurons were independent. In contrast, if spike counts are sampled from a discretized gaussian distribution with matching marginal distributions and matching correlations, the histogram can be reproduced well. In this case, the probability of spike counts being between 15 and 30 is 0.45.

4 Discussion

Whether a neuron does or does not fire a spike at any given point in time depends in many cases on not only the stimulus but also the activity of other

neurons in the system and its own activity in the recent past. Even if pairwise correlations are seemingly small, ignoring them can lead to dramatic errors in estimating the occurrence of firing patterns across many neurons. Therefore, to model the activity of populations of neurons in a realistic way, correlations across neurons and across time have to be taken into account. However, the functional role and implications of different types of neural correlations are still subject to debate (Nirenberg & Latham, 2003; Schneidman, Bialek, & Berry, 2003; Averbeck, Latham, & Pouget, 2006).

We have presented a flexible and conceptually simple framework for generating synthetic spike trains with a wide range of different firing rates and covariance structures. Generation of spike trains by binarizing (or discretizing) a multivariate gaussian random variable is possible even for large populations of neurons and for spike count statistics with arbitrary marginal distributions. In these situations, exact maximum entropy methods are not feasible, and techniques based on MCMC sampling are cumbersome and of limited accuracy. Our algorithms offer a practical and efficient alternative and make it possible to generate spike trains with specified correlation structure with little additional effort compared to independent ones.

Recently Tang et al. (2008) showed that maximum entropy models that do not take into account the correlations across time fail to predict the evolution of correlated states in a variety of cortical cultures. The sequences of active states in the data were substantially longer than in data sampled from their model. As there were sizable temporal correlations, they hypothesize that this could be resolved by a model that incorporates temporal correlations but did not propose a concrete model. The DG is a promising candidate, as it can be built entirely from pairwise correlations but can easily be fit to high-dimensional data. Also, the point-process framework can be used to model spike trains with spatial and temporal correlations (see, e.g., Gerwinn, Macke, Seeger, & Bethge, 2008) if spikes are considered to be events in continuous time. It is an interesting question how the discrete time models discussed in this letter can be related to these point-process models.

An experimental application of our technique could include naturalistic microstimulation of populations of neurons. These techniques allow the direct manipulation of neural population activity in various brain areas (e.g., Salzman, Britten, & Newsome, 1990). Advances in electrical (Brecht et al., 2004; Pezaris & Reid, 2007) and optical (Boyden, Zhang, Bamberg, Nagel, & Deisseroth, 2005; Huber et al., 2008) techniques make it possible to probe neural ensembles with increased temporal and spatial precision. This opens up the possibility of stimulations with different correlation structures to study their effect in driving neural responses.

In summary, the techniques introduced in this letter allow one to study correlated neural systems under more naturalistic assumptions about their statistics. While analytical insights in the role of correlations in neural systems (Abbott & Dayan, 1999) often require making simplifying

assumptions, the efficient simulation of correlated spike trains can help us to understand and verify to what extent these insights can be extrapolated to the regimes we encounter in experimental recordings (Ecker et al., 2007). In this way, our techniques provide an important link between computational studies of neural population coding and the fast-growing body of experimental work in this field where simultaneous recordings from dozens of neurons in parallel are becoming increasingly available (Buzsaki, 2004, for review).

Appendix A: Experimental Methods

Neural data were recorded from one awake, behaving monkey (*Macaca mulatta*) weighing 16 kg. The studies were approved by the local authorities (Regierungspräsidium) and were in full compliance with the guidelines of the European Community (EUVD 86/609/EEC) for the care and use of laboratory animals. Surgical, electrophysiological, and postprocessing methods have been described in detail in Tolias et al. (2007). Briefly, form-specific recording chambers built out of titanium were positioned stereotactically with the aid of high-resolution magnetic resonance anatomical images over the left hemisphere operculum in area V1. Inside the chambers, a custom-built array of tetrodes was chronically implanted. Neural activity was recorded at 32 kHz, digitized, and stored using the Cheetah data acquisition system (Neuralynx, Tucson, AZ), and spikes were extracted whenever the recorded voltage exceeded a predefined threshold on any tetrode channel. Single units were isolated using a custom-built offline clustering system working on features extracted from the recorded waveforms, and only well-isolated neurons were used in our analysis. The animal was implanted with a scleral search coil. Eye movements were monitored online and recorded for offline analysis at 200 Hz.

Visual stimuli were displayed using a dedicated graphics workstation (TDZ 2000; Intergraph Systems, Huntsville, AL) with a resolution of 1280×1024 pixels and refresh rate of 85 Hz, running an OpenGL-based stimulation program. The behavioral aspects of the experiment were controlled using the QNX real-time operating system (QSSL, Ontario, Canada). In the orientation experiment, the monkey acquired fixation on a colored square target (0.2°) for 300 ms. Then a sine wave grating stimulus was presented on the screen for 500 ms (size of the grating: 5° in diameter, spatial frequency: 4 cycles/ $^\circ$, 100% contrast). Eight equally spaced orientations were randomly interleaved. During the stimulation period, the animal was required to maintain fixation within a window of $\pm 1^\circ$. In the free-viewing experiment, the monkey was shown a sample from a collection of 1000 natural scenes for 2 seconds. During this time, the animal could freely look around the scene. For the analysis presented here, we extracted fixation periods from the recorded eye movement data. To compute fixation periods, we detected points where the norm of the derivative of the eye trace crossed

a predefined threshold. We regarded times between the saccade offset of the previous saccade and the next onset as fixation periods. Extremely short fixations (below 30 ms) and fixations where the eye trace was significantly changing over time (drift) were discarded. To obtain robust estimates of the moments, we discarded a small number of fixations in which at least one neuron fired more than 20 spikes. Fixation periods were 278.0 ± 104.6 ms long (mean \pm SD). The firing rate of a neuron was obtained by summing up all spikes fired by the respective neuron between the start and the beginning of a fixation period.

Appendix B: Constraints on the Covariance of a Binary Distribution and Existence of a Solution for DG

We want to determine the range of possible covariances between two binary random variables X and Y with means $P(X = 1) = p$ and $P(Y = 1) = q$. The covariance $C = \text{Cov}(X, Y) = P(X = 1, Y = 1) - pq$. As probabilities are nonnegative, this yields $C \geq -pq$. Also, $P(X = 1, Y = 1) \leq \min(p, q)$, which leads to $C \leq \min(p(1 - q), q(1 - p))$. Finally, $P(X = 0, Y = 0)$ has to be nonnegative. But $P(X = 0, Y = 0) = 1 - p - q + P(X = 1, Y = 1) = 1 - p - q + C - pq$. This implies that $C \geq -(1 - p)(1 - q)$. Moreover, these conditions are sufficient for the existence of a two-dimensional binary random variable with means p and q and covariance C .

To show that equation 2.2 always has a solution for any p, q , and $\Sigma_{ij} = C$ that satisfy the conditions above, we have to show that we can find a λ such that $\Phi_2(\gamma_1, \gamma_2, \lambda) = C + pq =: P(11)$, with $\Phi(\gamma_1) = p$ and $\Phi(\gamma_2) = q$. As Φ_2 is continuous in λ , it is sufficient to show that $\Phi_2(\gamma_1, \gamma_2, -1) \leq P(11)$ and $\Phi_2(\gamma_1, \gamma_2, 1) \geq P(11)$. Let V_λ be a bivariate gaussian random variable with mean 0, variances 1, and correlation λ ; then $\Phi_2(\gamma_1, \gamma_2, -1) = P(V_1 \leq \gamma_1, V_2 \leq \gamma_2) = P(-\gamma_2 \leq V_1 \leq \gamma_1) = p + q - 1 = P(11) - P(X = 0, Y = 0) \leq P(11)$. Similarly, $\Phi_2(\gamma_1, \gamma_2, 1) = P(V_1 \leq \gamma_1, V_2 \leq \gamma_2) = P(V_1 \leq \min(\gamma_1, \gamma_2)) = \min(p, q) \geq P(11)$.

To see that there are matrices that satisfy the constraints above but are not the covariance matrix of any binary distribution, consider the matrix

$$\begin{pmatrix} 1/4 & -1/8 & -1/8 \\ -1/8 & 1/4 & -1/8 \\ -1/8 & -1/8 & 1/4 \end{pmatrix}.$$

It is quickly verified that this matrix is positive definite, and the conditions above hold for all subsets of two of the three neurons. However, there is no binary distribution with this matrix as covariance. If it existed, each neuron X, Y, Z would have mean 0.5, and the probability of at least two out of the three neurons spiking simultaneously would be $P(X + Y + Z = 2) = -1/8 + 1/4 = 1/8$. However, the probability of three

simultaneous spikes is

$$\begin{aligned}
 P(X + Y + Z = 3) &= P(X + Y + Z > 0) \\
 &\quad - P(X = 1) - P(Y = 1) - P(Z = 1) \\
 &\quad + P(X + Y = 2) + P(X + Z = 2) + P(Y + Z = 2) \\
 &\leq 1 - 1/2 - 1/2 - 1/2 + 1/8 + 1/8 + 1/8 \\
 &= -1/8.
 \end{aligned}$$

Thus, the constraints that this covariance imposes on the probability of observing one or two spikes in the population would force the probability of three simultaneous spikes to be a negative number. As this is clearly not possible, such a matrix cannot be the covariance of a three-dimensional binary distribution.

Appendix C: Fitting the Dichotomized Gaussian for Firing Rates

$\mu = 0.5$

Equations 2.2 have a closed-form solution if the firing rate of each neuron i is $\mu_i = 0.5$. In this case, $\Lambda_{ij} = \sin(2\pi \Sigma_{ij})$. This solution can be obtained by calculating the probability mass of a bivariate gaussian in one quadrant of the 2D plane. We need to find λ such that

$$A = \frac{1}{2\pi\sqrt{1-\lambda^2}} \int \int_{\mathbf{x} \geq 0} e^{-\frac{1}{2}\mathbf{x}^T \Lambda^{-1} \mathbf{x}} d\mathbf{x} = \Sigma_{ij} + 1/4,$$

where $\mathbf{x} = (x, y)^T$. We write $\Lambda = UDU^T$, where $u = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$, and $D = \frac{1}{2} \begin{pmatrix} 1+\lambda & 0 \\ 0 & 1-\lambda \end{pmatrix}$. Changing variables from \mathbf{x} to $\mathbf{u} = D^{-1/2}U^T \mathbf{x}$, we get that

$$\begin{aligned}
 A &= \frac{1}{2\pi} \int \int_{U^T D^{1/2} \mathbf{u} \geq 0} e^{-\frac{1}{2}\mathbf{u}^T \mathbf{u}} d\mathbf{u} \\
 &= \frac{\alpha}{2\pi},
 \end{aligned}$$

where $\alpha = \arccos(-\lambda)$.

$$\begin{aligned}
 \text{Hence, } \lambda &= -\cos\left(2\pi\left(\Sigma_{ij} + \frac{1}{4}\right)\right) \\
 &= \sin(2\pi \Sigma_{ij}).
 \end{aligned}$$

Acknowledgments

We thank Sebastian Gerwinn for helpful discussions and Fabian Sinz, R. James Cotton, and David Greenberg for comments on the manuscript.

This work is supported by the German Ministry of Education, Science, Research and Technology through the Bernstein award to M.B. (BMBF; FKZ: 01GQ0601), a National Eye Institute National Research Service Award to A.S.T. and the Max Planck Society.

References

- Abbott, L. F., & Dayan, P. (1999). The effect of correlated variability on the accuracy of a population code. *Neural Computation*, *11*, 91–101.
- Alonso, J. M., Usrey, W. M., & Reid, R. C. (1996). Precisely correlated firing in cells of the lateral geniculate nucleus. *Nature*, *383*(6603), 815–819.
- Amari, S. I. (2001). Information geometry on hierarchy of probability distributions. *IEEE Transactions on Information Theory*, *47*, 1701–1711.
- Averbeck, B. B., Latham, P. E., & Pouget, A. (2006). Neural correlations, population coding and computation. *Nature Reviews Neuroscience*, *7*, 358–366.
- Bahadur, R. R. (1961). A representation of the joint distribution of responses to n dichotomous items. *Studies in Item Analysis and Prediction*, *6*, 158–168.
- Bair, W., Zohary, E., & Newsome, W. T. (2001). Correlated firing in macaque visual area MT: Time scales and relationship to behavior. *Journal of Neuroscience*, *21*, 1676–1697.
- Bethge, M., & Berens, P. (2008). Near-maximum entropy models for binary neural representations of natural images. In J. C. Platt, D. Koller, Y. Singer, & S. Roweis (Eds.), *Advances in neural information processing systems*, *20* (pp. 97–104). Cambridge, MA: MIT Press.
- Block, H., & Fang, Z. (1988). A multivariate extension of Hoeffding's lemma. *Annals of Probability*, *16*(4), 1803–1820.
- Boyden, E. S., Zhang, F., Bamberg, E., Nagel, G., & Deisseroth, K. (2005). Millisecond-timescale, genetically targeted optical control of neural activity. *Nature Neuroscience*, *8*, 1263–1268.
- Brecht, M., Schneider, M., Sakmann, B., & Margrie, T. W. (2004). Whisker movements evoked by stimulation of single pyramidal cells in rat motor cortex. *Nature*, *427*(6976), 704–710.
- Broderick, T., Dudik, M., Tkacik, G., Schapire, R. E., & Bialek, W. (2007). Faster solutions of the inverse pairwise Ising problem. Available online at <http://arxiv.org/abs/0712.2437v2>.
- Buzsaki, G. (2004). Large-scale recording of neuronal ensembles. *Nature Neuroscience*, *7*, 446–451.
- Chaganty, N. R., & Joe, H. (2006). Range of correlation matrices for dependent Bernoulli random variables. *Biometrika*, *93*(10), 197–206.
- Chen, Z., & Haykin, S. (2002). On different facets of regularization theory. *Neural Computation*, *14*, 2791–2846.
- Cox, D. R., & Wermuth, N. (2002). On some models for multivariate binary variables parallel in complexity with the multivariate gaussian distribution. *Biometrika*, *89*, 462–469.
- Ecker, A. S., Berens, P., Bethge, M., Logothetis, N. K., & Tolias, A. S. (2007). *Studying the effects of noise correlations on population coding using a sampling method*. Abstract

- presented at the Neural Coding, Computation and Dynamics Conference, Hossegor, France.
- Emrich, L. J., & Piedmonte, M. R. (1991). A method for generating high-dimensional multivariate binary variates. *American Statistician*, *45*, 302–304.
- Gange, S. J. (1995). Generating multivariate categorical variates using the iterative proportional fitting algorithm. *American Statistician*, *49*, 134–138.
- Gawne, T. J., & Richmond, B. J. (1993). How independent are the messages carried by adjacent inferior temporal cortical neurons? *Journal of Neuroscience*, *13*, 2758–2771.
- Genest, C., & Neslehova, J. (2007). A primer on discrete copulas. *ASTIN Bulletin*, *37*, 475–515.
- Gerwinn, S., Macke, J. H., Seeger, M., & Bethge, M. (2008). Bayesian inference for spiking neuron models with a sparsity prior. In J. C. Platt, D. Koller, Y. Singer, & S. Roweis (Eds.), *Advances in neural information processing systems*, *20* (pp. 529–536). Cambridge, MA: MIT Press.
- Ghosh, S., & Henderson, S. G. (2003). Behavior of the Norta method for correlated random vector generation as the dimension increases. *ACM Transactions on Modeling and Computer Simulation*, *13*, 276–294.
- Grünwald, P. D., & Dawid, A. P. (2004). Game theory, maximum entropy, minimum discrepancy, and robust Bayesian decision theory. *Annals of Statistics*, *32*, 1367–1433.
- Gur, M., Beylin, A., & Snodderly, D. M. (1997). Response variability of neurons in primary visual cortex (V1) of alert monkeys. *Journal of Neuroscience*, *18*, 2914–2920.
- Gütig, R., Aharonov, R., Rotter, S., & Sompolinsky, H. (2003). Learning input correlations through nonlinear temporally asymmetric Hebbian plasticity. *Journal of Neuroscience*, *23*, 3697–3714.
- Higham, N. J. (2002). Computing the nearest correlation matrix—a problem from finance. *IMA Journal of Numerical Analysis*, *22*(3), 329–343.
- Hubel, D. H., & Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *Journal of Physiology*, *195*, 215–243.
- Huber, D., Petreanu, L., Ghitani, N., Ranade, S., Hromadka, T., Mainen, Z., et al. (2008). Sparse optical microstimulation in barrel cortex drives learned behaviour in freely moving mice. *Nature*, *451*(7174), 61–64.
- Jaynes, E. T. (1978). Where do we stand on maximum entropy inference. In R. D. Levine & M. Tribus (Eds.), *The maximum entropy formalism*. Cambridge, MA: MIT Press.
- Jenison, R. L., & Reale, R. A. (2004). The shape of neural dependence. *Neural Computation*, *16*, 665–672.
- Kass, R. E., & Ventura, V. (2006). Spike count correlation increases with length of time interval in the presence of trial-to-trial variation. *Neural Computation*, *18*(11), 2583–2591.
- Kawamura, K. (1979). The structure of multivariate Poisson distribution. *Kodai Mathematical Journal*, *2*, 337–345.
- Kohn, A., & Smith, M. A. (2005). Stimulus dependence of neuronal correlation in primary visual cortex of the macaque. *Journal of Neuroscience*, *25*, 3661–3673.
- Lee, A. J. (1993). Generating random binary deviates having fixed marginal distributions and specified degrees of association. *American Statistician*, *47*(3), 209–215.

- Lunn, A. D., & Davies, S. J. (1998). A note on generating correlated binary variables. *Biometrika*, *85*(2), 487–490.
- Mastrorarde, D. N. (1983). Correlated firing of cat retinal ganglion cells. I. Spontaneously active inputs to *x*- and *y*-cells. *Journal of Neurophysiology*, *49*, 303–324.
- Meister, M., Lagnado, L., & Baylor, D. A. (1995). Concerted signaling by retinal ganglion cells. *Science*, *270*(5239), 1207–1210.
- Montani, F., Kohn, A., Smith, M. A., & Schultz, S. R. (2007). The role of correlations in direction and contrast coding in the primary visual cortex. *Journal of Neuroscience*, *27*, 2338–2348.
- Nelsen, R. B. (1987). Discrete bivariate distributions with given marginals and correlation. *Communications in Statistics-Simulation and Computation*, *16*(1), 199–208.
- Niebur, E. (2007). Generation of synthetic spike trains with defined pairwise correlations. *Neural Computation*, *19*, 1720–1738.
- Nirenberg, S., & Latham, P. E. (2003). Decoding neuronal spike trains: How important are correlations? *Proceedings of the National Academy of Sciences*, *100*, 7348–7353.
- Nirenberg, S. H., & Victor, J. D. (2007). Analyzing the activity of large populations of neurons: How tractable is the problem? *Current Opinion in Neurobiology*, *17*, 397–400.
- Park, C. G., Park, T., & Shin, D. W. (1996). A simple method for generating correlated binary variates. *American Statistician*, *50*(4), 306–310.
- Pearson, K. (1909). On a new method of determining correlation between a measured character *a*, and a character *b*, of which only the percentage of cases wherein *b* exceeds (or falls short of) a given intensity is recorded for each grade of *a*. *Biometrika*, *7*, 96–105.
- Pezaris, J. S., & Reid, R. C. (2007). Demonstration of artificial visual percepts generated through thalamic microstimulation. *Proceedings of the National Academy of Sciences*, *104*, 7670–7675.
- Pouget, A., Dayan, P., & Zemel, R. S. (2003). Inference and computation with population codes. *Annual Reviews of Neuroscience*, *26*, 381–410.
- Qaqish, B. (2003). A family of multivariate binary distributions for simulating correlated binary variables with specified marginal means and correlations. *Biometrika*, *90*, 455–463.
- Reich, D. S., Mechler, F., & Victor, J. D. (2001). Independent and redundant information in nearby cortical neurons. *Science*, *294*, 2566–2568.
- Salzman, C. D., Britten, K. H., & Newsome, W. T. (1990). Cortical microstimulation influences perceptual judgements of motion direction. *Nature*, *346*, 174–177.
- Schneidman, E., Berry, M. J., Segev, R., & Bialek, W. (2006). Weak pairwise correlations imply strongly correlated network states in a neural population. *Nature*, *440*(7087), 1007–1012.
- Schneidman, E., Bialek, W., & Berry, M. (2003). Synergy, redundancy, and independence in population codes. *Journal of Neuroscience*, *23*(37), 11539–11553.
- Sherman, S. M. (2001). Tonic and burst firing: Dual modes of thalamocortical relay. *Trends in Neuroscience*, *24*(2), 122–126.
- Shlens, J., Field, G. D., Gauthier, J. L., Grivich, M. I., Petrusca, D., Sher, A., et al. (2006). The structure of multi-neuron firing patterns in primate retina. *Journal of Neuroscience*, *26*(32), 8254–8266.

- Swendsen, R. H., & Wang, J. S. (1987). Nonuniversal critical dynamics in Monte Carlo simulations. *Physical Review Letters*, *58*, 86–88.
- Tang, A., Jackson, D., Hobbs, J., Chen, W., Smith, J. L., Patel, H., et al. (2008). A maximum entropy model applied to spatial and temporal correlations from cortical networks in vitro. *Journal of Neuroscience*, *28*, 505–518.
- Tolias, A. S., Ecker, A. S., Siapas, A. G., Hoenselaar, A., Keliris, G. A., & Logothetis, N. K. (2007). Recording chronically from the same neurons in awake, behaving primates. *Journal of Neurophysiology*, *98*, 3780–3790.
- Ventura, V., Cai, C., & Kass, R. E. (2005). Statistical assessment of time-varying dependency between two neurons. *Journal of Neurophysiology*, *94*, 2940–2947.
- Watanabe, S. (1960). Information theoretical analysis of multivariate correlation. *IBM Journal of Research and Development*, *4*(1), 66–82.
- Wolff, U. (1989). Collective Monte Carlo updating for spin systems. *Physical Review Letters*, *62*, 361–364.
- Zohary, E., Shadlen, M. N., & Newsome, W. T. (1994). Correlated neuronal discharge rate and its implications for psychophysical performance. *Nature*, *370*, 140–143.

Received February 18, 2008; accepted May 9, 2008.