

Generation of Synthetic Spike Trains with Defined Pairwise Correlations

Ernst Niebur

niebur@jhu.edu

Krieger Mind/Brain Institute and Department of Neuroscience, Johns Hopkins University, Baltimore, MD 21218, U.S.A.

Recent technological advances as well as progress in theoretical understanding of neural systems have created a need for synthetic spike trains with controlled mean rate and pairwise cross-correlation. This report introduces and analyzes a novel algorithm for the generation of discretized spike trains with arbitrary mean rates and controlled cross correlation. Pairs of spike trains with any pairwise correlation can be generated, and higher-order correlations are compatible with common synaptic input. Relations between allowable mean rates and correlations within a population are discussed. The algorithm is highly efficient, its complexity increasing linearly with the number of spike trains generated and therefore inversely with the number of cross-correlated pairs.

1 Introduction ---

Understanding any natural or artificial system is considerably furthered by the possibility of not only being able to observe but also to manipulate it, that is, by modifying the system's state and observing the resulting changes in behavior. In neuroscience, recent technological advances have substantially increased the extent to which the microstate of nervous systems can be controlled. While electrical microstimulation has been one of the standard tools of neurophysiology for some time (for a recent review, see Cohen & Newsome, 2004), new technology allows the stimulation of substantially larger numbers of sites (see, e.g., Normann, Maynard, Rousche, & Warren, 1999; Sekirnjak et al., 2006). In addition, newly developed optical methods based on uncaging of neurotransmitters allow the activation of up to tens of thousands of locations per second with high spatial and temporal resolution (Callaway & Katz, 1993; Shepherd, Pologruto, & Svoboda, 2003; Bousein, Nawrot, Rotter, Aertsen, & Heck, 2005; Shoham, O'Connor, Sarkisov, & Wang, 2005; Miesenbock & Kevrekidis, 2005; Thompson et al., 2005). Realizing the full potential of these techniques will require the injection of sets of artificial spike trains with controlled rates and correlations between them. There is thus a need for efficient and systematic methods

that allow generating not just individual spike trains but large numbers of them and the possibility of varying their rates and correlation structure.

A similar need arises in theoretical neuroscience. While much of the earlier work in the field was based on the mean-rate coding assumption originating in the peripheral nervous system (Adrian & Zotterman, 1926), that is, the hypothesis that neural processing uses only the information contained in the mean number of spikes averaged over a few hundred milliseconds, this concept is being questioned in increasingly stronger terms (Abeles, 1982; Singer, 1999). A substantial number of current models explores the consequences of the temporal coding hypothesis, that is, that the nervous system does not discard all information in neural spike trains other than the mean activity level. It has been proposed that neural information may be represented by controlled changes in the autocorrelation function, as in the large number of models of oscillation coding (e.g., Gray & Singer, 1989; Niebur, Koch, & Rosin, 1993), or in cross-correlation functions, as in models studying synchrony (e.g., Niebur & Koch, 1994; Tiesinga & Sejnowski, 2004), or a combination of both.

Thus, both the development of quantitative neural models that rely on correlated activity and of experimental techniques that allow selectively exciting neural tissue in multiple sites require the availability of test data. The subject of this letter is the generation of sets of synthetic spike trains with controlled rates and cross-correlations.

A finite pairwise correlation within a population of N model neurons can be obtained by generating a set of N_2 random numbers with suitable distribution (e.g., Poisson) and assigning them randomly to the N spike trains. If $N_2 < N$, some states are necessarily identical, which introduces correlations between spike trains (Destexhe & Pare, 1999). While this method is efficient, the mean pairwise correlation is not parametrically controlled (only indirectly via the ratio N_2/N), and it is the same for all pairs of neurons. Other methods were based on manipulating stochastic processes by either subtracting events (Ogata, 1981) or adding events (Stroeve & Gielen, 2001). Different from these earlier methods, the algorithm developed here generates spike trains whose mean rates as well as cross-correlations between each pair of spike trains are free parameters that can be selected independently, subject to constraints discussed below.

The algorithm developed here is a generalization of our earlier work (Mikula & Niebur, 2003a, 2003b, 2004, 2005) in that it allows differing rates for the spike trains rather than requiring that all rates are identical. Furthermore, the cross-correlation between any two of these spike trains can be selected anywhere between the cases of independent spike trains (minimal correlation), on one hand, and having identical spike trains (maximal correlation), on the other, subject to limitations discussed below.

Rather than advancing to the general case immediately, we first introduce, in sections 2 and 3, the simple and intuitive case of considering two spike trains only. In sections 4 and 5, we generalize to an arbitrary number

of spike trains. Different from the earlier case of identical firing rates in all spike trains, not all firing rates and cross-correlations can be combined; it is impossible to have strong correlations in a set of spike trains with vastly different firing rates. The resulting constraints are studied in section 6.

2 Generating Two-Spike Trains

Our goal is to provide a constructive method for generating two stochastic processes (we use the terms *spike train* and *stochastic point process* interchangeably), x and y , each consisting of a series of binary events (0, or no spike in bin, and 1, or one spike in bin). We assume that spike trains are memoryless, that is, all their time bins are independent, and we can therefore analyze them separately (this assumption can be relaxed; see section 7). These are Bernoulli processes, with 1 occurring in process x with probability¹ $r_x = \langle x \rangle$, with $0 \leq r_x \leq 1$ (the expectation value of obtaining a 1; see equation 3.2) and 0 occurring with probability $1 - \langle x \rangle$. Likewise, in process y , 1 occurs with probability $r_y = \langle y \rangle$, with $0 \leq r_y \leq 1$, and 0 with probability $1 - \langle y \rangle$. We note that for most practical purposes, the firing probability in any given bin is low, $r_x \ll 1$, in which case the Bernoulli process can be approximated by a Poisson process. The pairwise correlation of the two processes is C_{xy} , the expectation value of the probability of obtaining coincident spikes with rate correction (see equation 3.4).

In order to construct the two processes x and y , we follow the appendix in Mikula and Niebur (2003b). We start out with three independent Bernoulli processes that have either a 1 or a 0 in each bin and are binomially distributed. For one of the spike trains, referred to as reference spike train r in the following (because it is left unchanged), the probability of having a 1 in any given bin is p ; thus, a 0 is found with probability $(1 - p)$.

The other two spike trains, x and y , are constructed as follows. For spike train x , we start with a memoryless two-state stochastic process in each bin of which a 1 is found with probability p_x and a 0 with probability $(1 - p_x)$. We then switch, with a probability \sqrt{q} , the state in x to that in r . Note that in general (in all cases other than for $p = p_x$ or $q = 0$), the mean rate of the resulting spike train will be neither p nor p_x but an intermediate value to be computed later, in equation 3.2. The same procedure is applied to spike train y , with p_x replaced everywhere by p_y .

Table 1 shows the complete probability table for all combinations of a reference spike train (r , first column) and two spike trains x and y (second and third columns). Determining the probabilities P_{rxy} (fourth column) is straightforward. For the purpose of illustration, we illustrate computation

¹ We use the terms *firing probability* and *mean firing rate* interchangeably since they are proportional, the factor of proportionality being the bin width, a constant. See also section 7.

Table 1: Probability Table for All States of the Reference Spike Train (s_r) and Two Spike Trains s_x and s_y .

s_r	s_x	s_y	P_{rxy}
0	0	0	$(1 - p)[(1 - p_x) + p_x\sqrt{q}][(1 - p_y) + p_y\sqrt{q}]$
0	1	0	$(1 - p)p_x(1 - \sqrt{q})[(1 - p_y) + p_y\sqrt{q}]$
0	0	1	$(1 - p)[(1 - p_x) + p_x\sqrt{q}]p_y(1 - \sqrt{q})$
0	1	1	$(1 - p)p_xp_y(1 - \sqrt{q})^2$
1	0	0	$p(1 - p_x)(1 - p_y)(1 - \sqrt{q})^2$
1	1	0	$p[p_x + (1 - p_x)\sqrt{q}](1 - p_y)(1 - \sqrt{q})$
1	0	1	$p[(1 - p_x)(1 - \sqrt{q})][p_y + (1 - p_y)\sqrt{q}]$
1	1	1	$p(p_x + (1 - p_x)\sqrt{q})(p_y + (1 - p_y)\sqrt{q})$

Notes: The reference spike train has firing probability p , and the original spike trains x and y have firing probabilities p_x and p_y , respectively. In columns 1–3, 1 indicates a spike in the respective spike train, and 0 stands for no spike.

of the first probability, P_{000} . It is obtained by observing that the probability of having no spike in the reference spike train is $(1 - p)$ since the probability of obtaining a spike is p . This gives rise to the first factor, $(1 - p)$. Given the absence of a spike in the reference spike train, there are two possibilities that spike train x also does not contain a spike in this spike train.

The first is that it starts out with no spike. This happens with the probability $(1 - p_x)$, the first term in the square bracket. This is independent of the state of the reference spike train since whether a switch occurs or not, this spike train will always have no spike (switching would not change the state since, by assumption, the reference spike train does not contain a spike).

The second possibility occurs if the state in spike train x is switched to that in the reference spike train (0) even though originally (before switching) this state was 1. Since the latter occurs with the probability p_x and the switching with the probability \sqrt{q} , this additional probability is $p_x\sqrt{q}$, which completes the derivation of the first square bracket in P_{000} (note that these two events, having a spike occur in spike train x and switching the state to that of the reference spike train, are independent; therefore, their probabilities multiply to obtain the probability of the compound event). Finally, the same considerations apply to the second spike train (y), which therefore contributes a factor equal to that obtained for x but with all p_x replaced by p_y . The total probability is the product of those for the states of r , x , and y . The other cases in Table 1 are derived analogously.

3 Properties of Two Spike Trains

Having obtained Table 1, it is straightforward (although somewhat tedious) to verify the normalization condition:

$$\sum_{i=0,1} \sum_{j=0,1} \sum_{k=0,1} P_{ijk} = 1. \tag{3.1}$$

We now proceed to compute the mean firing rate of the spike trains. Direct substitution yields

$$r_x := \langle x \rangle := \sum_{i=0,1} \sum_{k=0,1} P_{i1k} = p_x + \sqrt{q}(p - p_x). \quad (3.2)$$

In this equation, the first and second equalities are the definition of the mean rate, and the last one is the result of direct substitution of terms from Table 1 and straightforward simplification. As anticipated, the mean rate $\langle x \rangle$ is influenced by p and p_x . Of course, for $q = 0$, we obtain $\langle x \rangle = p_x$, and for $q = 1$, we have $\langle x \rangle = p$. The analogous result is obtained for y :

$$r_y := \langle y \rangle := \sum_{i=0,1} \sum_{k=0,1} P_{i1k} = p_y + \sqrt{q}(p - p_y). \quad (3.3)$$

The rate-corrected cross-correlation (or covariance) between spike trains x and y is

$$\begin{aligned} C_{xy} &:= \langle xy \rangle - \langle x \rangle \langle y \rangle \\ &:= \sum_{i=0,1} P_{i11} - \left(\sum_{i=0,1} \sum_{k=0,1} P_{i1k} \right) \left(\sum_{i=0,1} \sum_{k=0,1} P_{i1k} \right) = (p - p^2)q, \end{aligned} \quad (3.4)$$

where the first and second equalities are definitions, and the third one is obtained after substitution from Table 1 and simplification. We observe that C_{xy} is independent of p_x and p_y . Furthermore, C_{xy} attains a maximum value of $q/4$ at $p = 1/2$ and decreases monotonically for larger and smaller p , to reach zero for both $p = 0$ and $p = 1$. Of course, C_{xy} is identically zero for all p for $q = 0$. For $q = 1$, all three spike trains are identical and have spike rate p and correlation $C_{xy} = p - p^2$, which is the autocorrelation of a binomial process train with probability p . The latter can also be computed directly from Table 1 as the variance of the reference spike train,

$$\text{var}_r := C_{rr} := \langle (r - \langle r \rangle)^2 \rangle = p - p^2, \quad (3.5)$$

where the same notation as that introduced in equation 3.4 was used.

In practice, rather than specifying the rates p_x and p_y of the uncorrelated stochastic processes and the correlation \sqrt{q} with the reference spike train, we are interested in generating spike trains with specified mean rates and cross-correlation. That is, $\langle x \rangle$, $\langle y \rangle$, and C_{xy} are given quantities, and we have to determine parameters p_x , p_y , and q given these values. By solving equations 3.2 to 3.4 for p_x , p_y , and q , we find that a pair of spike trains with means $\langle x \rangle$, $\langle y \rangle$ and cross-correlation C_{xy} is obtained by choosing p_x , p_y ,

and q as

$$p_x = \frac{\langle x \rangle - p\sqrt{q}}{1 - \sqrt{q}} \quad (3.6)$$

$$p_y = \frac{\langle y \rangle - p\sqrt{q}}{1 - \sqrt{q}} \quad (3.7)$$

$$q = \frac{C_{xy}}{p - p^2}, \quad (3.8)$$

where p , the rate of the reference spike train, is formally a free parameter. There are limitations on the choice of this parameter (e.g., $p - p^2 \geq C_{xy}$ to ensure $q \leq 1$ in equation 3.8), which will be discussed in the more general case of arbitrary numbers of spike trains in section 6. Note that these limitations do not apply if all firing rates are identical ($\langle x \rangle = \langle y \rangle = p$) since p_x and p_y do not depend on q in this case (Mikula & Niebur, 2003a).

4 Arbitrary Number of Spike Trains

For N spike trains, there are 2^N possible states in each bin. For nonzero q , all these states are entangled because each randomly chosen pair has to have the same mean covariance. Despite the apparent complexity of the problem, we can compute the probability of each of the states as the sum of two terms, as follows. The trick is that we specify only the correlation with the reference spike train r and that the pairwise correlation is then obtained from expressions like equation 3.4. For instance, Table 1 shows that the state without a spike in either x and y has a total probability of

$$\{(1 - p)[(1 - p_x) + p_x\sqrt{q}][(1 - p_y) + p_y\sqrt{q}] + \{p(1 - p_x)(1 - p_y)(1 - \sqrt{q})^2\},$$

where the terms in the first and second parentheses correspond to the cases without and with a spike in r , respectively.

The algorithm to generate N correlated spike trains is thus as follows. We start by generating N uncorrelated, memoryless stochastic point processes (i.e., they have independent time bins) and binomial firing statistic in each bin. As before, each is a Bernoulli process, and process i , with $i \in \{1, \dots, N\}$, has a mean firing rate (probability of having state 1) of p_i ; $0 \leq p_i \leq 1$. One additional process with the same statistics and with mean rate p is generated and referred to, as before, as the reference spike train r . To generate the set of N correlated spike trains, we switch, with a probability \sqrt{q} , the state of uncorrelated spike train i to that of spike train r . Using the notation from the previous paragraph, let (s_1, s_2, \dots, s_N) be the state of the N spike trains

Table 2: Probability of Obtaining State s_i in Process i Depends on the State of the Reference Process.

s_r	s_i	$P_{s_r s_i}^{(i)}$	$P^{(i)}(s_r, s_i)$
0	0	$(1 - p_i) + p_i \sqrt{q}$	$(1 - p)[(1 - p_i) + p_i \sqrt{q}]$
1	0	$(1 - p_i)(1 - \sqrt{q})$	$p(1 - p_i)(1 - \sqrt{q})$
0	1	$p_i(1 - \sqrt{q})$	$(1 - p)p_i(1 - \sqrt{q})$
1	1	$p_i + (1 - p_i)\sqrt{q}$	$p[p_i + (1 - p_i)\sqrt{q}]$

Notes: Column 1: state of reference process. Column 2: state of process i . The last column shows the probability of obtaining the combination of state s_r in the reference process and state s_i in process i , and the third column shows only that component of this probability that is independent of p . The first factor in $P^{(i)}(s_r, s_i)$ stems from the reference spike train r , the second factor comes from spike train i and depends on both its own state and that of r . Note that in order to avoid proliferation of symbols, we distinguish the symbols for the probabilities in column 4 from the expressions in column 3 (which are not probabilities but terms used in equation 4.1 and beyond) by using a subscript notation in column 3 and parentheses in column 4 (top row).

in a given time bin. The probability of finding the system in this state is then

$$P(s_1, s_2, \dots, s_N) = (1 - p) \prod_{i=1}^N P_{0s_i}^{(i)} + p \prod_{i=1}^N P_{1s_i}^{(i)}, \tag{4.1}$$

where $P_{0s_i}^{(i)}$ and $P_{1s_i}^{(i)}$ are the terms from the third column of Table 2.

We will show now that equations 3.6 to 3.8 generalize immediately to the case of arbitrary N . The mean rate of process i is defined as the expectation value of this process having state 1, that is,

$$r_i := P(s_i = 1) = (1 - p)(P_{00}^{(1)} + P_{01}^{(1)})(P_{00}^{(2)} + P_{01}^{(2)}) \dots P_{01}^{(i)} \dots (P_{00}^{(N)} + P_{01}^{(N)}) \\ + p(P_{10}^{(1)} + P_{11}^{(1)})(P_{10}^{(2)} + P_{11}^{(2)}) \dots P_{11}^{(i)} \dots (P_{10}^{(N)} + P_{11}^{(N)}).$$

The sums within each parenthesis in this equation are again unity, and we obtain

$$r_i = (1 - p)P_{01}^{(i)} + pP_{11}^{(i)} = p_i + \sqrt{q}(p - p_i), \tag{4.2}$$

where the second equation is obtained using column 3 in Table 2. Equation 4.2 is the generalization of equation 3.2 for spike train i .

The correlation function is computed analogously. By definition, the correlation function of processes i and k is the expectation value of both processes having state 1 corrected by the product of their rates:

$$\begin{aligned}
 C_{ik} &:= P(s_i = 1, s_k = 1) - r_i r_k \\
 &= (1 - p)(P_{00}^{(1)} + P_{01}^{(1)}) \dots P_{01}^{(i)} \dots P_{01}^{(k)} \dots (P_{00}^{(N)} + P_{01}^{(N)}) \\
 &\quad + p(P_{10}^{(1)} + P_{11}^{(1)}) \dots P_{11}^{(i)} \dots P_{11}^{(k)} \dots (P_{10}^{(N)} + P_{11}^{(N)}) - r_i r_k. \tag{4.3}
 \end{aligned}$$

Sums in the parentheses being unity, we are left with

$$C_{ik} = (1 - p)P_{01}^{(i)} P_{01}^{(k)} + pP_{11}^{(i)} P_{11}^{(k)} - r_i r_k. \tag{4.4}$$

Using again $P_{01}^{(i)}$ and $P_{11}^{(i)}$ from Table 2 and equation 4.2, we obtain, after simplification,

$$C_{ik} = (p - p^2)q, \tag{4.5}$$

which is the generalization of equation 3.4.

In practice, the mean rates r_i for $i \in \{1, \dots, N\}$ of the correlated processes and their pairwise correlation C are given, and we need to determine the parameters p_i and q of the uncorrelated stochastic processes. We therefore solve equations 4.2 and 4.5 for these parameters to obtain

$$p_i = \frac{r_i - p\sqrt{q}}{1 - \sqrt{q}} \tag{4.6}$$

$$q = \frac{C}{p - p^2}, \tag{4.7}$$

which are the generalizations of equations 3.2 to 3.4. Section 6 discusses constraints on the choice of these parameters.

5 Arbitrary Cross-Correlations

So far we have assumed that all cross-correlations are identical. To generalize to arbitrary covariances, rather than changing the state in all stochastic processes with the same probability \sqrt{q} , the state in process i is changed to that of the reference process with probability $\sqrt{q_i}$, which can be different for each i . The table of probabilities is then just as in Table 2 but with \sqrt{q} replaced by $\sqrt{q_i}$ everywhere. Straightforward algebra along the lines of equations 4.1 to 4.5 shows that the mean firing rate of process i is unchanged and given as in equation 4.2, and that the mean cross-correlation function

between processes i and k , as defined in equation 4.3, is

$$C_{ik} = (p - p^2)\sqrt{q_i q_k}. \quad (5.1)$$

Of course, in the special case of $q_i = q_k$, this reduces to equation 4.5.

The mean rates are obtained as in equation 4.6, with \sqrt{q} replaced by $\sqrt{q_i}$ in the numerator and denominator. Likewise, given a set of cross-correlations $C_{ij} = C_{ji}$; $i, j = 1, \dots, N$, the corresponding parameters q_{ij} are computed from equation 5.1. In the most general case, when all C_{ij} are different, half of the N parameters q_i can be chosen freely (subject to constraints similar to those discussed in section 6), and the other $N/2$ are determined by equation 5.1. We note in passing that another way to add structure to the correlation at the population level is to add additional reference spike trains.

6 Constraints on Rates and Correlations

While arbitrary values for $q \in [0, 1]$ and $p \in [0, 1]$ can be chosen if the mean rates of all spike trains are identical (Mikula & Niebur, 2003a), a set of conditions between p and q needs to be satisfied when mean rates differ (for simplicity, we discuss the case of equal pairwise cross-correlation only; the generalization discussed in section 5 does not pose any particular difficulties but complicates the notation). This is intuitively understood by the fact that too large a spread of firing rates does not allow enforcing pairwise strong correlation on all pairs. Formally, the constraint manifests itself by the requirement that q in equation 4.7 and p_i in equation 4.6 need to be well-defined probabilities with $0 \leq q, p_i \leq 1$. Equation 4.7 requires²

$$C \leq p - p^2. \quad (6.1)$$

We solve for p to obtain

$$\frac{1}{2} - \sqrt{\frac{1}{4} - C} \leq p \leq \frac{1}{2} + \sqrt{\frac{1}{4} - C}. \quad (6.2)$$

Equation 4.6 yields

$$r_i - p\sqrt{q} \geq 0 \quad (6.3)$$

$$r_i - p\sqrt{q} \leq 1 - \sqrt{q}. \quad (6.4)$$

² As noted at the end of section 3, these constraints do not apply if all mean rates are identical, $p_i \equiv p \forall i = 1, \dots, N$, since in that case $p_i = r_i = p$ in equation 4.6, and the equation becomes an identity. The same is true for the trivial cases of $q = 0$ or $q = 1$.

Inequalities 6.3 and 6.4 can be combined as

$$1 - \frac{1 - r_i}{\sqrt{q}} \leq p \leq \frac{r_i}{\sqrt{q}}. \quad (6.5)$$

For a given p and q , the right inequality 6.5 determines the frequency of the spike train with the smallest possible r_i as

$$r_{\min} = p\sqrt{q}, \quad (6.6)$$

and the left inequality determines the frequency of the spike train with the largest r_i ,

$$r_{\max} = 1 - (1 - p)\sqrt{q}. \quad (6.7)$$

In the following, we consider the practically more important reverse question: For a given set of r_i , can we find (at least) one p such that the algorithm can be applied, and if so, what are the allowable values for p ?

For any given p_i and $q < 1$, a solution of inequality 6.5 exists because the left inequality can be transformed into $\frac{r_i}{\sqrt{q}} - (\frac{1}{\sqrt{q}} - 1)$, which is smaller than the right-hand side since by assumption, $\sqrt{q} < 1$ (the case $q = 1$ is trivial, leading to N identical spike trains with rate p). The condition does, however, establish a constraint on a set of mean rates r_i that can be generated with this procedure because if their spread becomes too large, it is not possible to find a p that satisfies both inequalities 6.5 for all i .

From the second of the inequalities 6.5 we have $p \leq \frac{r_i}{\sqrt{q}}$ and thus $p^2 \leq \frac{r_i^2}{q}$, bearing in mind that the positive solution of this equation will be the correct one since p is a probability. Replacing q in this inequality by using equation 3.8, we obtain $p^2 \leq \frac{r_i^2(p-p^2)}{C}$ and after division by p ($\neq 0$),

$$p \leq \frac{r_i^2}{C + r_i^2}. \quad (6.8)$$

Thus, p must be smaller than the expression on the right for all r_i . For $C = 0$, the condition is fulfilled for all probabilities p , but for increasing C , it becomes more and more stringent.

Using again equation 3.8, we rewrite the first inequality in equation 6.5 as

$$1 - \frac{1 - r_i}{\sqrt{C}} \sqrt{p - p^2} \leq p \quad (6.9)$$

or

$$p + A\sqrt{p - p^2} - 1 \geq 0, \quad (6.10)$$

where $A := \frac{1-r_i}{\sqrt{C}}$ is a positive constant.

At equality, this is a nonlinear equation of the type

$$p = 1 - A\sqrt{p - p^2}, \quad (6.11)$$

whose solution depends, through A , on r_i and C . In the relevant interval $p \in]0, 1[$, the right-hand side of this equation is a U-shaped function with exactly one crossing with the identity. These fixed points are given by $p = C[(1 - r_i)^2 + C]^{-1}$.

Although a set of p can always be found for any given pair of values (r_i, C) , it is possible that the intervals given by conditions equations 6.8 and 6.9 are nonoverlapping for a given C and a range of values $[r_{i,\min}, r_{i,\max}]$. No solution with the specified set of mean rates and pairwise correlation C between all pairs exists in this case.

Only if at least one p can be found that simultaneously fulfills equations 6.8 and 6.9 can the algorithm be applied. Figure 1 shows the solutions of inequalities 6.5 as a function of r_i . Only those values of p that are between the two lines corresponding to the left inequality (solid line) and the right inequality (dashed line) fulfill both inequalities and are therefore acceptable solutions. If the pairwise correlation is relatively small, as in Figure 1a with $C = 0.02$, a suitable p value can be chosen for all but the most extreme values of r_i , and in this case there is usually a large range of suitable p values. For large correlation, for example, for $C = 0.2$ as shown in Figure 1b (note that the maximum of completely correlated, i.e., identical, spike trains is attained for $C = 0.25$), the constraints become more stringent and the spread of allowable values for p is zero for a substantial subset of r_i (those below and above the crossing points). It is impossible to generate sets of correlated spike trains with unequal mean rates; only sets with $r_i = p\forall i$ are allowed.

The shaded part of the abscissa (in Figures 1a and 1b) is an example for a range of firing rates $[r_{i,\min}, r_{i,\max}]$. The projection onto the ordinate shows that for small correlation (see Figure 1a, $C = 0.02$), a large range of p values can be chosen (shaded part of the ordinate), while for strong correlation ($C = 0.2$, in Figure 1b), the range of possible p values is much smaller. For an even larger range, for example, if $[r_{i,\min}, r_{i,\max}] = [0.4, 0.6]$, it is impossible to generate spike trains with cross-correlation $C = 0.2$ between each pair of the set.

Figure 2 shows that spike trains computed with the described algorithm have the desired properties. The upper panels of Figure 2 show examples of 300 bins each of 100 spike trains, with, respectively, low correlation

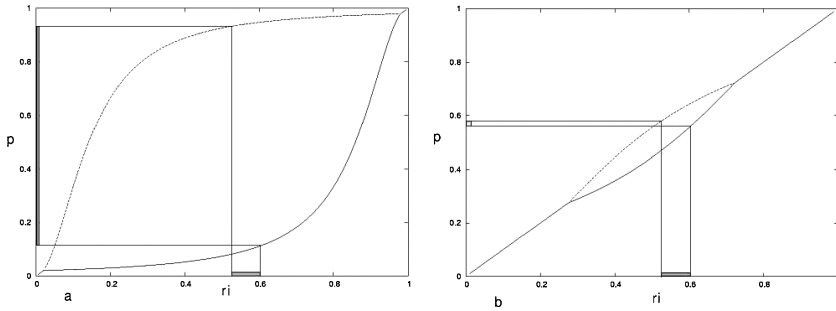


Figure 1: Solutions of the inequalities 6.5 as a function of r_i for $C = 0.02$ (a) and $C = 0.2$ (b). The solid lines show the limits of the solution (i.e., at equality) for the left inequality; the dashed lines are the equivalent for the right inequality. Acceptable p values are all those between the two curves, that is, above the full lines and below the dashed lines. For a chosen range of mean rates (shaded on the abscissa), the range of acceptable values (shaded on the ordinate) is significantly smaller for $C = 0.2$ than for $C = 0.02$. The spread of mean rates is zero for rates r_i that are outside the range where the curves cross.

($C = 0.02$) in the upper left panel and high correlation ($C = 0.2$) in the upper right. Even at $C = 0.02$, spike trains are clearly correlated and, as expected, correlation is very strong at $C = 0.2$. The lower left panel of Figure 2 demonstrates that the actual firing rates, computed as the expectation value of the number of events in equation 3.2 and plotted on the ordinate, scatter closely around the selected values (on the abscissa). For illustration, we generated 20 spike trains with 1000 bins each with firing rates varying in equidistant steps, with $p_x \in [0.3, 0.4]$ for $C = 0.02$ and $p_x \in [0.5, 0.6]$ for $C = 0.2$ (these intervals were chosen simply so we could show both data sets in one plot). In both cases, $p = 0.58$ was chosen, which is from the allowable range in Figure 1. The bottom right panel of Figure 2 is a plot of the cross-correlation values between all these 20 spike trains (that is, the expectation value computed in equation 3.4) plotted as a function of the parameter C . Again, it is seen that the cross-correlation of the simulated spike trains scatters around the chosen value.

7 Discussion

A thorough understanding of biological neural systems requires the development of quantitative models, both experimental and theoretical. This is of particular importance for understanding neural microcircuitry, one of the most complex and at the same time practically and philosophically important structures known. Both theoretical (computational) and experimental models of neural tissue need a principled way of stimulating the model

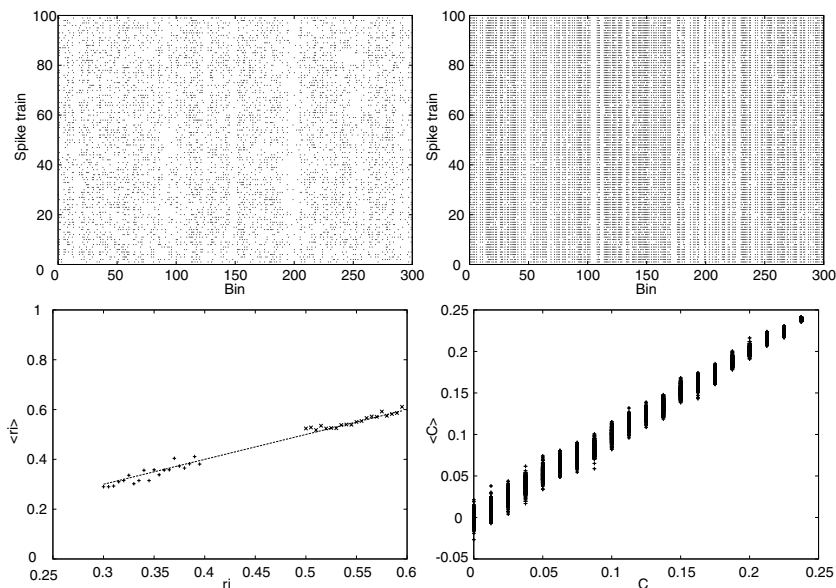


Figure 2: Properties of simulated spike trains. (Upper left) Raster plot of 100 spike trains with $p = 0.2$ and $C = 0.02$. (Upper right) Same, but $C = 0.2$. (Bottom left) Measured rates of 40 simulated spike trains plotted against nominal rates. For illustration, rates are shown for 20 spike trains with equidistant rates in the range $[0.3, 0.4]$ for $C = 0.02$ (+ signs) and from the range $[0.5, 0.6]$ for $C = 0.2$ (x signs). (Lower right) Average cross-correlation between spike trains plotted as a function of C . It can be seen that the observed cross-correlation, computed from equation 3.4, scatters around the nominal value.

substrate, in vitro or in silico. Recent reports demonstrate substantial progress in overcoming the technical difficulties of delivering stimuli with high spatial and temporal resolution into the tissue. While the simplest stimulation patterns might be independently distributed random processes, the next stages of investigation will surely demand more structured patterns.

This letter introduces an efficient and straightforward algorithm for constructing excitation patterns consisting of stochastic processes with controlled mean activity and cross-correlations between all pairs of processes. The algorithm is highly efficient and requires on the order of $N \times B$ operations where N is the number of spike trains and B the number of time bins per spike train. Remarkably, its complexity is thus linear with N , even though the number of cross-correlations generated increases as $o(N^2)$. Furthermore, the algorithm lends itself to parallelization up to the finest grain (individual spike trains and individual bins).

An attractive alternative approach for generating correlated sets of model spike trains is based on maximum entropy principles. The basic idea (perhaps most clearly described by Jaynes, 1957, but going back to Gibbs and Laplace) is that the most rational course of action is to avoid making assumptions in the absence of information. Applied to the question of spike train statistics, it has been argued that spike trains with controlled second-order correlations should be generated with maximum entropy in all higher orders. This means that all but second-order correlations must be minimal since higher-order correlations would introduce structure and therefore lower-than-maximum entropy.

Algorithms for spike train generation and analysis based on maximum entropy principles (e.g., Bohte, Spekreijse, & Roelfsema, 2000; Nakahara & Amari, 2002) are powerful methods for characterizing ensembles of spike trains if higher-order correlations are either known to vanish or if the structure of the higher-order correlations is known (and can thus be taken into account), or if absolutely no information about them is available. However, none of these three cases is the situation considered here, namely, when we want to generate spike trains with correlational structure similar to that found in biological neural systems. Although we assume that only second-order correlations are explicitly known, as is usually the case (some exceptions in which higher-order correlations are observed and analyzed can be found are Gerstein & Clark, 1964; Abeles & Goldstein, 1977; Abeles & Gerstein, 1988; Abeles, 1991; Martignon, von Hasseln, Grün, Aertsen, & Palm, 1995; Riehle, Grün, Diesmann, & Aertsen, 1997), and that we have no detailed information about higher-order correlations, we do know that the spike trains are generated in a highly connected neural network. For this reason, it is unlikely that no correlations exist between them at any higher than second order, which is the assumption made by maximum entropy methods. Applying maximum entropy methods with vanishing higher-order terms in this situation would then violate the basic rationale for using these methods. Paraphrasing from Jaynes (1957), the goal is to generate a probability distribution that correctly represents our state of knowledge about the system, or finding a probability assignment that avoids bias, while agreeing with whatever information is given.

It is always possible to use a maximum entropy approach, enforcing vanishing higher-order correlations, in a case in which insufficient information about higher-order correlations is available. In this study, we adopt a different strategy, with the explicit goal of taking into account all available information, including the likely presence of higher-order correlations. Although we have no detailed knowledge about those correlations in the absence of sufficient experimental data, we at least know that such correlations are likely to exist in a hierarchically connected complex network.

The solution adopted in this study is to generate spike trains whose higher-order correlations correspond to those found in neural systems

whose correlations are due to common input. Although this is by no means the only possible source of correlations (examples for others being direct and indirect interactions between neurons, as was observed more than a generation ago; Moore, Segundo, Perkel, & Gerstein, 1970), common input is widely recognized to be an important source of correlation in spike trains. The spike trains generated by the methods developed here are thus particularly suited as models of neural activity where neurons are influenced by spiking input from a common source.

While we have focused on pairwise cross-correlations in this work, it is possible to make the mean rates of all processes time variant (by adjusting the parameters p_i) and thus also to vary their autocorrelations. This allows, for instance, modeling absolute and refractory periods by temporarily lowering the spike rate of a process after a spike has occurred in this process. Figure 3 shows an example of correlation functions for a process with an absolute refractory period of two bins. Relative refractory periods are implemented in the same way but with a finite lowered probability of firing after each spike. The formulas for the computation of rates and correlations, equations 4.6 and 4.7, remain unchanged, with N reduced by the number of neurons that are in the absolute refractory period, and with r_i in equation 4.6 replaced by a lowered rate (subject to the constraints in section 6) for neurons in the relative refractory period. Figure 3 shows, as expected, vanishing entries in the autocorrelation function (squares) during the absolute refractory period, which is reflected in smaller cross-correlation (crosses) because of the finite correlation between the spike trains. The figure also shows that refractoriness decreases the mean firing rate (from 0.2 to about 0.14 in this case) and leads to an increase of the autocorrelation function immediately following the refractory period, with a corresponding peak in the power spectrum (Bair, Koch, Newsome, & Britten, 1994; Franklin & Bair, 1995). This peak, more noticeable for higher firing rates where higher-order peaks also can be observed (data not shown), is again reflected in attenuated form in the cross-correlation function. We note that the occurrence of these structures is a matter of principle (Bair et al., 1994; Franklin & Bair, 1995) and not due to the specifics of our implementation.

The possibility of varying the firing rates independently (rather than for all spike trains in unison) and the resulting requirement that bins are independent alleviates limitations of our older algorithm (Mikula & Niebur, 2003a). Experimentally observed cross-correlograms frequently have broad maxima rather than a sharp peak only in the center bin (as in, for example, Mikula & Niebur, 2003a; Destexhe & Pare, 1999). Spike jitter can be introduced by allowing transitions into a given bin of a spike train not only from the corresponding bin in the reference spike train but also (with smaller probability) from the neighboring bins. The probabilities for the states are computed as in equation 4.1 and, for a stationary situation (i.e., p_i and jitter distribution independent of time), the rates (see equation 4.2) are unchanged. The width of the jitter distribution will determine the width of the

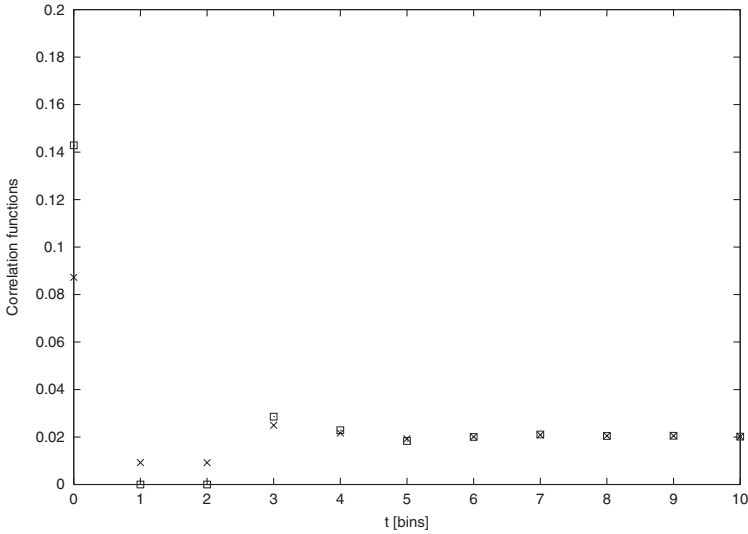


Figure 3: Correlation functions with refractory period. Shown are the average autocorrelation functions (squares) of 20 spike trains and the average cross-correlation functions (crosses) between them, with an absolute refractory period of 2 bins. All processes have mean rates of $p = 0.2$ and correlation between spike trains is $C = 0.1$. The refractory period introduces a depression not only in the autocorrelation but, because of the correlation between spike trains, also in the cross-correlation. Also note the increase in correlation immediately following the refractory period in both functions.

resulting correlation function. Some kind of integral over the peak width is usually employed as a measure of synchrony, and the one-bin definition of synchrony (see equation 4.3) is replaced by a more complex expression that involves this measure.

As a general observation, we established in section 6 that there are combinations of rates and pairwise cross-correlations that cannot be achieved, namely, high cross-correlations and, simultaneously, a large spread in the rates of the individual spike trains. Within the limits described in sections 5 and 6, pairwise cross-correlations and mean firing rates can be chosen freely. A free parameter in the procedure is the bin width, which determines the relation between the probability of the 1 event in the stochastic process and the mean firing rate of the simulated spike train generated: mean firing rate = bin width \times probability (see also note 1). The only formal requirement in the choice of the bin width is that there cannot be more than one event in any given bin, and the actual choice will depend on constraints provided by the system under study. For computational models, a suitable bin size could be the inverse maximal rate at which the neurons are known to fire;

for the generation of artificial spike trains to be used in microstimulation, it could be the inverse maximal stimulation frequency.

The spike trains introduced here can be used in a variety of situations that we have explored previously with spike trains with identical rates, for example, to study the effect of spike-timing-dependent plasticity (Mikula & Niebur, 2003b), the interplay of excitation and inhibition (Mikula & Niebur, 2004), and in feedforward networks of coincidence detectors (Mikula & Niebur, 2005). The ease of generating large numbers of synthetic spike trains in a highly efficient algorithm should prove to be of use in a large number of applications. On request, the author will make available sets of spike trains with the described statistics.

Acknowledgments

I thank Emery Brown for discussions. The work was supported by NIH grants NS43188-01A1, 5R01EY016281-02, and R01-NS40596.

References

- Abeles, M. (1982). *Local cortical circuits*. New York: Springer-Verlag.
- Abeles, M. (1991). *Corticonics—Neural circuits of the cerebral cortex*. Cambridge: Cambridge University Press.
- Abeles, M., & Gerstein, G. (1988). Detecting spatiotemporal firing patterns among simultaneously recorded single neurons. *J. Neurophysiol.*, *60*, 909–924.
- Abeles, M., & Goldstein, M. (1977). Multispikes train analysis. *Proc. IEEE*, *65*, 762–773.
- Adrian, E. D., & Zotterman, Y. (1926). The impulses produced by sensory nerve endings. Part 2: The response of a single end organ. *J. Physiol.*, *61*, 151–171.
- Bair, W., Koch, C., Newsome, W., & Britten, K. (1994). Power spectrum analysis of bursting cells in area MT in the behaving monkey. *J. Neuroscience*, *14*(5), 2870–2892.
- Bohte, S. M., Spekreijse, H., & Roelfsema, P. R. (2000). The effects of pairwise and higher-order correlations on the firing rate of a postsynaptic neuron. *Neural Computation*, *12*, 153–179.
- Boucsein, C., Nawrot, M., Rotter, S., Aertsen, A., & Heck, D. (2005). Controlling synaptic input patterns in vitro by dynamic photo stimulation. *J. Neurophysiology*, *94*(4), 2948–2958.
- Callaway, E. M., & Katz, L. C. (1993). Photostimulation using caged glutamate reveals functional circuitry in living brain slice. *Proc. Nat. Acad. Sci., USA*, *90*, 7661–7665.
- Cohen, M. R., & Newsome, W. T. (2004). What electrical microstimulation has revealed about the neural basis of cognition. *Curr. Opin. Neurobiol.*, *14*(2), 169–177.
- Destexhe, A., & Pare, D. (1999). Impact of network activity on the integrative properties of neocortical pyramidal neurons in vivo. *J. Neurophysiology*, *81*, 1531–1547.
- Franklin, J., & Bair, W. (1995). The effect of a refractory period on the power spectrum of neuronal discharge. *SIAM J. Appl. Math.*, *55*(4), 1074–1093.
- Gerstein, G., & Clark, W. (1964). Simultaneous studies of firing patterns in several neurons. *Science*, *143*, 1325–1327.

- Gray, C., & Singer, W. (1989). Stimulus-specific neuronal oscillations in orientation columns of cat visual cortex. *Proc. Nat. Acad. Sci., USA*, *86*, 1698–1702.
- Jaynes, E. T. (1957). Information theory and statistical mechanics. *Physical Review*, *106*(4), 620–630.
- Martignon, L., von Hasseln, H., Grün, S., Aertsen, A., & Palm, G. (1995). Detecting higher-order interactions among the spiking events in a group of neurons. *Biological Cybernetics*, *73*(1), 69–81.
- Miesenbock, G., & Kevrekidis, I. G. (2005). Optical imaging and control of genetically designated neurons in functioning circuit. *Annu. Rev. Neurosci.*, *28*, 533–563.
- Mikula, S., & Niebur, E. (2003a). The effects of input rate and synchrony on a coincidence detector: Analytical solution. *Neural Computation*, *15*, 539–547.
- Mikula, S., & Niebur, E. (2003b). Synaptic depression leads to nonmonotonic frequency dependence in the coincidence detector. *Neural Computation*, *15*(10), 2339–2358.
- Mikula, S., & Niebur, E. (2004). Correlated inhibitory and excitatory inputs to the coincidence detector: Analytical solution. *IEEE Transactions on Neural Networks*, *15*(5), 957–962.
- Mikula, S., & Niebur, E. (2005). Rate and synchrony in feedforward networks of coincidence detectors: Analytical solution. *Neural Computation*, *14*(4), 881–902.
- Moore, G., Segundo, J., Perkel, D., & Gerstein, G. (1970). Statistical signs of synaptic interactions in neurons. *Biophysical Journal*, *10*, 876–900.
- Nakahara, H., & Amari, S. (2002). Information-geometric measure for neural spikes. *Neural Comput.*, *14*(10), 2269–2316.
- Niebur, E., & Koch, C. (1994). A model for the neuronal implementation of selective visual attention based on temporal correlation among neurons. *Journal of Computational Neuroscience*, *1*(1), 141–158.
- Niebur, E., Koch, C., & Rosin, C. (1993). An oscillation-based model for the neural basis of attention. *Vision Research*, *33*, 2789–2802.
- Normann, R. A., Maynard, E. M., Rousche, P. J., & Warren, D. J. (1999). A neural interface for a cortical vision prosthesis. *Vision Research*, *39*(15), 2577–2587.
- Ogata, Y. (1981). On Lewis simulation method for point processes. *IEEE Transactions on Information Theory*, *27*, 23–31.
- Riehle, A., Grün, S., Diesmann, M., & Aertsen, A. (1997). Spike synchronization and rate modulation differentially involved in motor cortical function. *Science*, *278*, 1950–1953.
- Sekirnjak, C., Hottowy, P., Sher, A., Dabrowski, W., Litke, A. M., & Chichilnisky, E. J. (2006). Electrical stimulation of mammalian retinal cells with multielectrode arrays. *J. Neurophysiology*, *95*(6), 3311–3327.
- Shepherd, G. M., Pologruito, T. A., & Svoboda, K. (2003). Circuit analysis of experience-dependent plasticity in the developing rat barrel cortex. *Neuron*, *38*(2), 277–290.
- Shoham, A., O'Connor, D. H., Sarkisov, D. V., & Wang, S. S. (2005). Rapid neurotransmitter uncaging in spatially defined patterns. *Nature Methods*, *2*(11), 837–843.
- Singer, W. (1999). Neuronal synchrony: A versatile code for the definition of relations? *Neuron*, *24*, 49–65.

- Stroeve, S., & Gielen, S. (2001). Correlation between uncoupled conductance-based integrate-and-fire neurons due to common and synchronous presynaptic firing. *Neural Computation, 13*, 2005–2029.
- Thompson, S. M., Kao, J. P., Kramer, R. H., Poskanzer, K. E., Silver, R. A., Digregorio, D., & Wang, S. S. (2005). Flashy science: Controlling neural function with light. *J. Neurosci., 25*(45), 10358–10365.
- Tiesinga, P. H. E., & Sejnowski, T. J. (2004). Rapid temporal modulation of synchrony by competition in cortical interneuron networks. *Neural Computation, 16*, 251–275.

Received February 2, 2006; accepted November 29, 2006.