Available at
www.ComputerScienceWeb.com
POWERED BY SCIENCE DIRECT®

NEUROCOMPUTING

Neurocomputing 52–54 (2003) 31–37

www.elsevier.com/locate/neucom

# Non-parametric significance estimation of joint-spike events by shuffling and resampling

Gordon Pipa*, Sonja Grün[1]

*Department of Neurophysiology, Max-Planck-Institute for Brain Research, Deutschordenstr 46, Frankfurt/Main 60528 Germany*

**Abstract**

The 'unitary event' analysis method was designed to analyze multiple parallel spike trains for correlated activity. The null-hypothesis assumes Poissonian spike train statistics, however experimental data may fail to be consistent with this assumption. Here we present a non-parametrical significance test that considers the original spike train structure of experimental data. The significance of coincident events observed in simultaneously recorded spike trains is estimated on the basis of the same spike trains, however recombined to sets of non-corresponding trials ('trial shuffling'). Resampling from this set provides the distribution for coincidences reflecting the null-hypothesis of independence.


*Keywords:* Spike synchronization; Trial shuffling; Significance; Bootstrapping

## 1. Introduction

The assembly hypothesis postulates [5] that information processing in the cortex is mediated by groups of neurons by their coordinated spiking activity and is supported by a number of experimental studies (e.g. Ref. [11]). In order to test this hypothesis, the unitary event (UE) analysis was developed [3,4] and enabled to study the relation of spike synchronization to behavioral events [2,4,9,10]. In the UE-analysis,

* Corresponding author.
*E-mail address:* pipa@mpih-frankfurt.mpg.de (G. Pipa).
[1] Present address: Institute for Biology, Freie Universität Berlin, Königin-Luise Str. 28/30, 14195 Berlin, Germany.

0925-2312/03/$ - see front matter 
doi:10.1016/S0925-2312(02)00823-8

empirically observed coincidence counts are evaluated for their significance based on expectation by chance. The significance is estimated using a Poisson distribution parameterized by the expected number of coincidences, which is a function of the product of the firing rates of the contributing neurons. The implicit assumption is that the observed spike trains follow Poissonian statistics. However, experimental data often fail to be compatible with this assumption (e.g. Ref. [8]). In the work presented here we derive a non-parametric significance test, that is based on the experimental data, while respecting the original spike train structures. This is realized by combining shuffling of trials, and subsequent resampling from the set of coincidence counts of non-corresponding trials ('CSR'-method). Thus, first we introduce the procedure of trial shuffling and the generation of the set of data containing coincidence counts from individual trial combinations. Since we aim to compare the original found in $M$ trials, we derive in a second step the probability distribution of the sum of coincidence counts from $M$ trials by bootstrapping [1]. This probability distribution is representing the null-hypothesis $H_0$ of independence which allows us to estimate the significance of coincidences detected in corresponding trials. Requirements on the number of bootstrap samples with respect to the precision of the significance estimate are discussed.

## 2. Destruction of intrinsic joint-events by trial shuffling

To incorporate the null-hypothesis of independence, trials of the original data are shuffled to combine non-simultaneously recorded spike trains, and yield counts of chance coincident events. The underlying assumption is that the spike trains recorded in repetitive trials, i.e. under the same stimulus or behavioral condition, are realizations of the same neuronal process [6]. The latter is not required to be stationary in time.

Number of coincidence counts resulting from all different trial combinations compose the set $\Omega$. In case of $N = 2$ neurons, this set contains two classes of distinct subsets, i.e. the set of coincidence counts resulting from simultaneously observed trials $\Omega_s = \{\omega_s\}$, and the set of counts from non-simultaneous trial combinations $\Omega_0 = \{\omega_0\}$ (with $\Omega_0 = \Omega \setminus \Omega_s$). This can be illustrated in matrix form (Fig. 1B), where elements of $\Omega_s$ can be found on the diagonal, and elements of $\Omega_0$ as off-diagonal entries. In case of more than two neurons ($N > 2$), we define the set of counts from trial combinations with *all* trial indices being different ('completely shuffled trials') as

$$\Omega_0 = \{\omega_{j(l_1,\ldots,l_N)} | l_h \neq l_k, \text{ with } h \neq k; h,k \in 1 \ldots N \wedge l_h, l_k \in 1 \ldots M\} \tag{1}$$

and contains

$$S = \binom{M}{N} \cdot N! = \frac{M!}{(M-N)!} \tag{2}$$

($M$ the number of trials) elements. A non-empty set $\Omega_0$ respects the condition $M \geqslant N$, i.e. a minimum of $N$ trials needs to be observed to allow shuffling of trials. The number

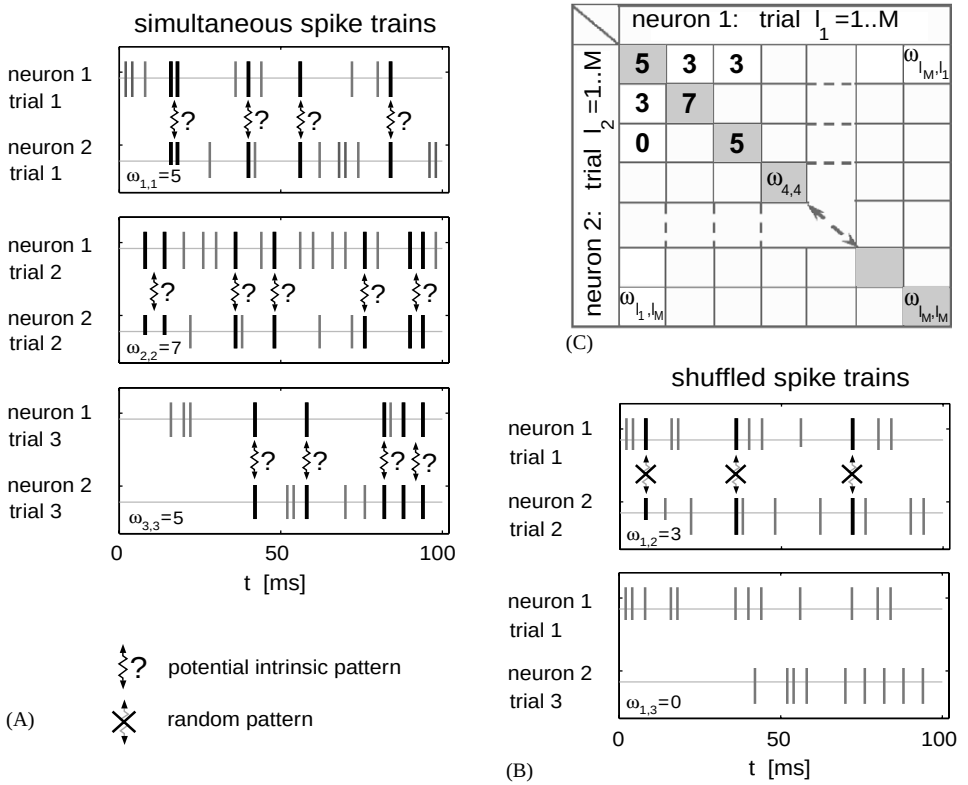



Fig. 1. Destruction of intrinsic patterns by trial shuffling illustrated for two neurons. (A) Sketch of a set of three trials (out of $M$) of two simultaneously recorded neurons. Coincident spike events are marked as potential intrinsic joint-spike patterns; (B) Examples of two pairs of shuffled spike trains from (A). Coincident events (marked) occur by chance only; (C) Matrix containing the number of coincident events per trial $\omega_{l_1,l_2}$ detected in all possible trial combinations of $M$ trials. $l_1, l_2 \in 1, 2, \ldots, M$ indicate trial indices of neuron 1 (horizontal) and 2 (vertical), respectively. Coincidence counts detected in simultaneous trials enter the matrix on the diagonal ($l_1 = l_2$), off-diagonal ($l_1 \neq l_2$) elements result from shuffled trials.

of elements in $\Omega_0$ increases rapidly with increasing number of trials and neurons (Fig. 2A).

## 3. Generating the distribution of coincidence counts

To estimate the significance of the number of coincidences $\bar{\omega}_s$ observed in the $M$ experimental trials, the probability distribution of the number of coincident events expected in $M$ trials has to be estimated. For doing that we resample from the set of non-simultaneous (single) trial combinations $\Omega_0$ all possible combinations of $M$ elements and compute their sum $\bar{\omega}_0^*$. The total set $\Gamma$ of sums of all possible combinations contains $B$ elements (Eq. 3). The ideal estimate of the probability distribution for coincidence counts $p(\bar{\omega}_0^*)$ (see Ref. [7] for a full mathematical derivation) can be

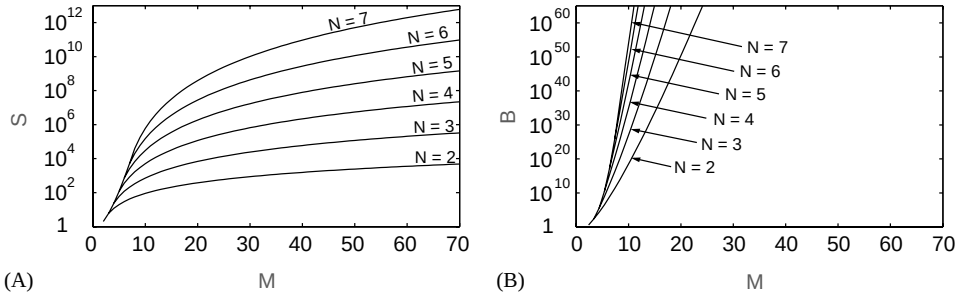



Fig. 2. (A) Number of shuffle elements ($S$) as a function of the number of simultaneously observed neurons ($N$) and the number of trials ($M$). (B) Number of possible bootstrap samples ($B$) of $M$ non-corresponding trial combinations as a function of the number of neurons and trials (note: logarithmic scaling of the $y$ axes).

constructed by forming the histogram (normalized) of all elements of $\Gamma$ (see Fig. 3A, top):

$$p(\bar{\omega}_0^*) = \frac{(\#v \in \Gamma | v = \bar{\omega}_0^*)}{B} \quad \text{with} \quad B = |\Gamma| = S^M. \tag{3}$$

Fig. 2B illustrates that the size $B$ of the set $\Gamma$ is extremely large even for a small number of trials ($M$), and increases even further for a larger number of trials and neurons. Obviously, the construction of the probability distribution $p(\bar{\omega}_0)$ based on all elements of $\Gamma$ is not practical in general (but see also Ref. [7]). Thus, as an alternative, we derive an estimate of (Eq. 3) on a random subset $\Gamma_\xi$ of $\Gamma$ consisting of $\xi < B$ elements:

$$\hat{p}(\bar{\omega}_0^*) = \frac{(\#v \in \Gamma_\xi | v = \bar{\omega}_0^*)}{\xi}. \tag{4}$$

As expected, the accuracy of (Eq. 4) depends on the sample size $\xi$ of the chosen subset $\Gamma_\xi$ (compare Fig. 3A, top to Fig. 3A, middle and bottom). The influence of the size of the subsample $\Gamma_\xi$ on the significance estimation is discussed in the following.

### 3.1. Significance estimation

The significance of the empirical coincidences $\bar{\omega}_s$ detected in the simultaneously observed trials may be expressed by the probability of finding $\bar{\omega}_s$ or even more coincidence counts in case of independent spike trains ('joint-$p$-value', [3]). Thus, instead of using a Poisson distribution parameterized by the number of expected events [3], we now estimate the significance on the basis of the probability distribution derived by shuffling and resampling (Eq. 4):

$$\alpha^*(\bar{\omega}_s, \Gamma_\xi) = \int_{\bar{\omega}_s}^{\infty} \hat{p}(\bar{\omega}_0^*)\, \mathrm{d}\bar{\omega}_0^* \tag{5}$$

in the following abbreviated as $\alpha^*$ (indicated by the area of gray bars of $\hat{p}(\bar{\omega}_0^*)$ in Fig. 3A, bottom). Since the set $\Gamma_\xi$ underlying the significance estimate is a random

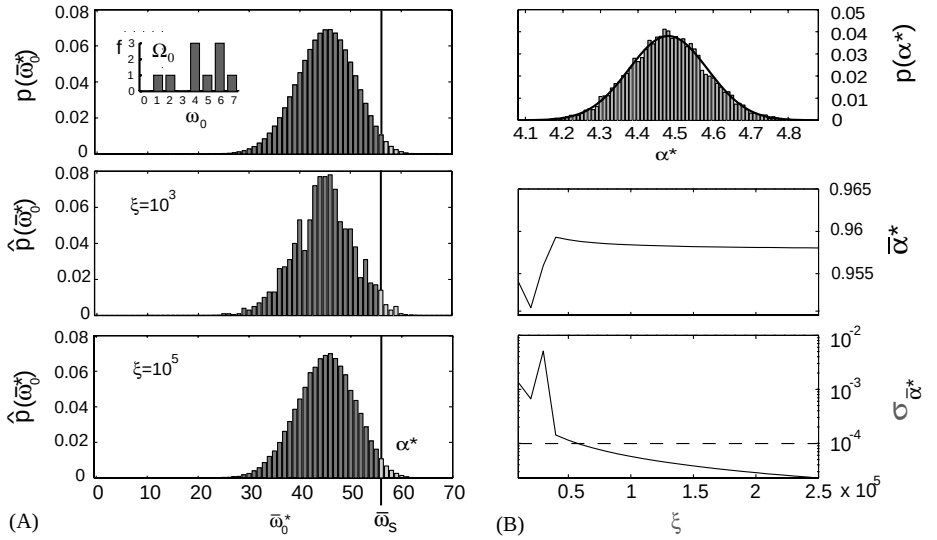



Fig. 3. (A) Probability distribution of coincidence counts from non-simultaneously recorded trials ($N = 2$, $M = 10$). Top: ideal estimate of the probability distribution $p(\bar{\omega}_0^*)$ for coincidences computed from all elements of $\Gamma$. Middle, bottom: estimates $\hat{p}(\bar{\omega}_0^*)$ of $p(\bar{\omega}_0^*)$ each based on a subset $\Gamma_\xi$ of $\Gamma$ consisting of $\xi$ elements (middle: $\xi = 10^3$; bottom: $\xi = 10^5$). (B) Top: distribution of significance estimates $\alpha^*$ of $\bar{\omega}_s$ (area of gray bars of $\hat{p}(\bar{\omega}_0^*)$ for $\bar{\omega}_0^* \geqslant \bar{\omega}_s$ in (A)) based on $10^4$ Monte–Carlo simulations. Each simulation is based on independently drawn subsets $\Gamma_\xi$, each consisting of $\xi = 10^4$ elements. Normal approximation of $p(\alpha^*)$ (solid line) with parameters $\bar{\alpha}^* = 0.0448$ and $\sigma_{\bar{\alpha}^*} = 0.0010$. Mean ((B), middle) and standard deviation (B, bottom) of five estimates of $\alpha^*$ as a function of $\xi$. The standard deviation decreases with $\xi$, while the mean converges to the expectation value of $\alpha^*$. In this example the required precision for $\bar{\alpha}^* \pm 0.01\%$ (dashed line) is reached at $\xi = 0.6 \times 10^5$.

sample of $\Gamma$, the significance $\alpha^*$ itself is a random variable. Fig. 3B (top) illustrates the distribution of significance estimates $\alpha^*$ of $\bar{\omega}_s$, derived on the basis of random sets $\Gamma_\xi$ from $\Gamma$ for fix $\xi$. The distribution can be approximated by a normal distribution (solid line in Fig. 3B, top). Thus, the best estimate of the significance is given by its mean $\bar{\alpha}^*$, and its standard deviation $\sigma_{\bar{\alpha}^*}$ yields a measure for the confidence interval. It can be shown that $\sigma_{\bar{\alpha}^*}$ decreases with the sample size $\xi$ of $\Gamma_\xi$ (Fig. 3B, bottom). By setting a threshold on $\sigma_{\bar{\alpha}^*}$ at the required precision for the significance estimate (e.g. 0.01% as in Fig. 3B, bottom), we derive a truncation criteria for the minimal necessary size $\xi$ of the set $\Gamma_\xi$.

## 4. Discussion

We have shown, that the approach of combined trial shuffling and resampling (CSR-method) allows us to estimate the significance of joint-spike events. Coincidence counts from simultaneously recorded spike trains are compared to coincidence counts resulting from shuffled trial combinations of the same data set. Since the number of possible

combinations increases considerably with the number of trials (and neurons), the full set $\Gamma$ of combinations can in general not be considered due to computation time and memory requirements. In Ref. [7] we discussed an approach which uses efficient combinatorial methods to estimate the significance based on $\Gamma$, which, however, is restricted to limited number of trials and low number of coincidence counts.

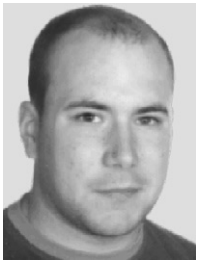
Here we presented a reliable estimate of the distribution of coincident events for significance testing based on random samples of subsets of trial recombinations. The precision of the significance estimate can be adjusted to a requested level by successively increasing the size of the random subset. The method integrates well in the scheme of the unitary event analysis [3,4] and provides an improved method for the significance estimation. In contrast to the approach in Ref. [3,4], (1) the assumption of a specific underlying spike train model is not required, since the null-hypothesis of independence is incorporated on the basis of the original experimental spike trains including their temporal structure. (2) Assumptions regarding stationarity are relaxed. A prerequisite of the method is that stationarity across trials is fulfilled, which actually is the definition of a 'trial' [6]. However, here the assumption of stationarity in time, implicit to most correlation analysis techniques, is not required.

## Acknowledgements

We thank Wolf Singer for stimulating discussions and his support. This work was in part funded by the Volkswagen Foundation.

## References

[1] B. Efron, R.J. Tibshirani, An Introduction to the Bootstrap, Chapmann Hall, London, CRC Press, Boca Raton, 1998.
[2] F. Grammont, A. Riehle, Precise spike synchronization in monkey motor cortex involved in preparation for movement, Exp. Brain Res. 128 (1999) 118–122.
[3] S. Grün, M. Diesmann, A. Aertsen, 'Unitary Events' in multiple single-neuron activity. I. Detection and significance, Neural Comput. 14 (2002) 43–80.
[4] S. Grün, M. Diesmann, A. Aertsen, 'Unitary Events' in multiple single-neuron activity. II. Non-Stationary data, Neural Comput. 14 (2002) 81–119.
[5] D.O. Hebb, Organization of Behavior. A Neurophysiological Theory, Wiley, New York, 1949.
[6] D.H. Perkel, G.L. Gerstein, G.P. Moore, Neuronal spike trains and stochastic point processes. II. Simultaneous spike trains, Biophys. J. 7 (1967) 419–440.
[7] G. Pipa, M. Diesmann, S. Grün, Significance of joint-spike events based on trial-shuffling by efficient combinatorial methods, submitted for publication.
[8] D.S. Reich, F. Mechler, K.P. Purpura, J.D. Victor, Interspike intervals, receptive fields, and information encoding in primary visual cortex, J. Neurosci. 20 (2000) 1964–1974.
[9] A. Riehle, F. Grammont, M. Diesmann, S. Grün, Dynamical changes and temporal precision of synchronized spiking activity in motor cortex during movement preparation, J. Physiol. (Paris) 94 (2000) 569–582.
[10] A. Riehle, S. Grün, M. Diesmann, A. Aertsen, Spike synchronization and rate modulation differentially involved in motor cortical function, Science 278 (1997) 1950–1953.
[11] W. Singer, A.K. Engel, A.K. Kreiter, M.H.J. Munk, S. Neuenschwander, P.R. Roelfsema, Neuronal assemblies: necessity, signature and detectability, Trends Cog. Sci. 1 (1997) 252–261.



**Gordon Pipa** was born in 1974. With his development of a high speed image processor he finished 3rd in the German federal competition for young scientists in 1996 (Patents: US6231185, WO99/03066, PCT01454, EP0993657, DE9701454). He is holder of the honorary medal of the Friedrich-Bläse-Foundation and of a federal scholarship. He studied Physics and Electrical Engineering at the Technical University of North Rhine–Westphalia in Aachen and received his M.Sc. in Physics in 2002 from the J.W. Goethe University (Frankfurt a.M., Germany). His Masters thesis was carried out in the Department of Neurophysiology at the Max-Planck-Institute for Brain Research (Frankfurt a.M., Germany).

**Sonja Grün** was born in 1960 in Germany. She received her M.Sc. in Physics from the Eberhard-Karls University in Tübingen in 1991. After associations with the Ruhr-University in Bochum and the Weizmann Institute of Science (Rehovot, Israel), she received her Ph.D. in Physics in 1996 (Ruhr-University, Bochum). From 1996 to 1997 she was a post-doctoral fellow at the Hebrew University in Jerusalem, Israel. From 1998 to 2002 she led the group for data analysis and modeling the Max-Planck-Institute for Brain Research in Frankfurt/Main, Germany. Since Dec. 2002 she is Assistant Professor for Neuroinformatics at the Freie Universität in Berlin, Germany. Her research interests are in the field of theoretical neuroscience with focus on analyis of data from electrophysiological recordings and identification of cooperative processes in biological neural networks.