Fuzzy frequent pattern mining in spike trains

David Picado Muiño, Iván Castro León, and Christian Borgelt

European Centre for Soft Computing
Gonzalo Gutiérrez Quirós s/n, 33600 Mieres, Spain
{david.picado|ivan.castro|christian.borgelt}@softcomputing.es

Abstract. We present a framework for characterizing spike (and spike-train) synchrony in parallel neuronal spike trains that is based on identifying spikes with what we call *influence maps*: real-valued functions describing an influence region around the corresponding spike times within which possibly graded synchrony with other spikes is defined. We formalize two models of synchrony in this framework: the bin-based model (the almost exclusively applied model in the literature) and a novel, alternative model based on a continuous, graded notion of synchrony, aimed at overcoming the drawbacks of the bin-based model. We study the task of identifying frequent (and synchronous) neuronal patterns from parallel spike trains in our framework, formalized as an instance of what we call the fuzzy frequent pattern mining problem (a generalization of standard frequent pattern mining) and briefly evaluate our synchrony models on this task.

1 Introduction

The principles of *neural coding* (i.e., how information is represented in biological neural networks) are still not well understood and continue to be the topic of ongoing debate. Several coding schemes have been proposed in the neuroscience literature. Among them, one of the most prominent is the so-called *temporal coordination scheme* (see, e.g., [1]), first advocated by D.O. Hebb [2] and driven by more recent physiological and anatomical evidence, according to which the coordinated emission of *spikes* (i.e., electrical impulses, also called *action potentials*), in particular *synchronous* spiking (see, e.g., [3–5]) by groups of neurons plays a major role in neuronal information processing.

In order to understand the principles of coordinated neuronal activity and neural coding it is necessary to observe the activity of multiple neurons simultaneously. This is nowadays possible due to improvements in multiple-electrode recording (see, e.g., [6]), which allows to monitor the activity of hundreds of neurons. Such recordings are typically represented by sequences of points in time (i.e., point processes) and referred to as *parallel spike trains*.

It is not a trivial matter to determine what constitutes synchronous spiking in parallel spike trains. *Exact* spike-time coincidence cannot be expected and thus a characterization of spike synchrony in such terms is not suitable.

We present in Section 2 a very general, flexible framework for characterizing and quantifying synchrony among spikes (spike synchrony) and spike trains (spike-train synchrony), which is able to accommodate graded, continuous characterizations of synchrony. The framework is based on the identification of spikes with what we call *influence maps*: real-valued functions describing an *influence region* around the corresponding spike times within which synchrony with other spikes is defined. Together with our influence maps, a class of so-called *synchrony operators* (to quantify spike synchrony) and a class of *support operators* (to quantify spike-train synchrony) are defined, based on what we consider desirable criteria in the assessment of synchrony among spikes and spike trains.

Within this framework, two concrete models for characterizing and quantifying spike (and spike-train) synchrony are described in Section 3. The first model formalizes the so-called *bin-based* method, which is the almost exclusively applied one in the field and based on time-bin discretization: spikes are binned in time intervals of equal length. In this model, two or more spikes are considered to be synchronous if they lie in the same time bin. The second model characterizes synchrony in a graded fashion and aims at overcoming the main drawbacks of the bin-based model by avoiding time-bin discretization and bivalence in the amount of synchrony among spikes. Graded, non-bivalent (i.e., fuzzy, as in fuzzy set theory—see, e.g., [7]) synchrony itself is not an entirely new concept, though. Implicit notions of presumably non-bivalent, graded synchrony can be found in some papers on topics related to synchronous spiking, such as [8, 9]. However, in these approaches (at least in the two just mentioned) a formal derivation and treatment of such a concept is a non-trivial task.

In Section 4 we consider the problem of identifying frequent neuronal patterns in parallel spike trains. For the models presented in Section 3 the problem is formalized as an instance of what we call the *fuzzy frequent pattern mining problem*—a generalization of standard frequent pattern mining. The task of identifying frequent neuronal patterns is of particular relevance in itself but also in relation to tasks concerned with identifying other sets of neurons that are characterized by different synchrony requirements (such as, e.g., synchronous patterns, called *unitary events* in [10], or *neural assemblies*, as defined in e.g. [11, 12]).

In Section 5 we provide a summary of a first evaluation of our graded, continuous synchrony model (as presented in Section 3) for the identification of frequent neuronal patterns on artificially generated spike trains as an alternative to the commonly used bin-based model. Section 6 briefly summarizes our discussion.

2 Synchronous Spiking

Let N be a (finite) set of neurons. We are given parallel spike trains, one for each neuron in N, formalized as spike-time sequences of the form $\{t_1^a, ..., t_{k_a}^a\} \subset (0, T]$, for $a \in N$, where k_a is the number of times neuron a fires in the interval $(0, T]$. We denote the set of all these sequences by $\mathcal{S}$. Collections of sequences like $\mathcal{S}$ constitute the raw data on which we build our framework.

2.1 Influence Maps

The first step in setting our framework consists in our characterization of spikes. We do so by considering an *influence region* around each spike time in $\mathcal{S}$ together with a possibly varying *influence degree* along that region, which we formalize by introducing what we call *influence maps*.

Definition 1. *A function* $f : \mathbb{R} \to [0, \infty)$ *is called an* influence map *if*

$$\int_{-\infty}^{\infty} f(x)dx = 1.$$

The class of all possible influence maps is denoted by $\mathcal{F}$.

For any $f \in \mathcal{F}$ the set $\{x \in \mathbb{R} \mid f(x) > 0\}$ gives us the *influence region* of f (or, more precisely, of the spike represented by the influence map f). Intuitively, the influence region of a spike captures the times at which other spikes can be considered to be synchronous to some (positive) degree, as is detailed below.

It is worth noting that, although an influence map is formally a probability density function, we do *not* impose such an interpretation (i.e., an influence map is *not* meant to represent uncertainty about the actual location of a spike).

2.2 Spike Synchrony

In order to characterize spike synchrony based on our identification of spikes with influence maps in $\mathcal{F}$, we define a class of operators, called *synchrony operators*, aimed at quantifying the *degree of synchrony* of a collection of spikes.

We define synchrony operators on *multisets* (sometimes also called *bags*) over $\mathcal{F}$ (i.e., collections of elements of $\mathcal{F}$ that can contain multiple copies of the same element), which are formally defined as pairs $\langle \mathcal{F}, h \rangle$, with $h : \mathcal{F} \to \mathbb{N} \cup \{0\}$. Intuitively, the function h "counts" the occurrences of the elements of $\mathcal{F}$ in $\langle \mathcal{F}, h \rangle$. Throughout this paper, whenever convenient, we use set notation for multisets, for example, $\{f_1, f_1, f_2, f_2, f_3\}$ instead of $\langle \mathcal{F}, h \rangle$, with $h(f_1) = 2$, $h(f_2) = 2$, $h(f_3) = 1$ and $h(f) = 0$ for any $f \in \mathcal{F} \setminus \{f_1, f_2, f_3\}$. The collection of all possible multisets over $\mathcal{F}$ is denoted by $\mathscr{M}(\mathcal{F})$.

The use of multisets over $\mathcal{F}$ (instead of simple subsets of $\mathcal{F}$) is motivated by the fact that distinct spikes of different neurons in our recordings can be represented by the same influence maps, for example, if their respective spike times coincide exactly (but possibly also in other situations, see below).

In the definition of the class of synchrony operators and other classes of operators defined later we make use of the *sum operator* on multisets over $\mathcal{F}$, denoted by $\uplus$, and defined for two multisets $\langle \mathcal{F}, h_1 \rangle$ and $\langle \mathcal{F}, h_2 \rangle$ as follows:

$$\langle \mathcal{F}, h_1 \rangle \uplus \langle \mathcal{F}, h_2 \rangle = \langle \mathcal{F}, h_1 + h_2 \rangle = \langle \mathcal{F}, h \rangle,$$

where, for all $f \in \mathcal{F}$, $h(f) = h_1(f) + h_2(f)$.[1] It is worth noting that $\uplus$ is commutative and associative and thus its definition can be extended trivially to

[1] For a wider insight into basic multiset theory and, in particular, binary operations on multisets see e.g. [13].

more than two multisets. It is also straightforward to define the *cardinality* of a multiset $G = \langle \mathcal{F}, h \rangle$ as $|G| = |\langle \mathcal{F}, h \rangle| = \sum_{f \in \mathcal{F}} h(f)$.

Definition 2. *The function* $Sync : \mathscr{M}(\mathcal{F}) \to [0, 1]$ *is called a* synchrony operator *on* $\mathcal{F}$ *if* $Sync(G) \geq Sync(G \uplus H)$ *for* G, H *multisets in* $\mathscr{M}(\mathcal{F})$*.*

Definition 2 aims at formalizing the requirement that the degree of synchrony of a certain collection of spikes cannot increase if we add another spike to it.

Let us consider now a subset $A = \{a_1, ..., a_k\} \subseteq N$ of neurons, for some natural number $k \leq |N|$, and $G = \{f^{a_1}, ..., f^{a_k}\} \in \mathscr{M}(\mathcal{F})$, with f^{a_i} the influence map corresponding to a spike time of neuron a_i in $\mathcal{S}$, for all $i \in \{1, ..., k\}$. We refer to the joint-spike event G as an A-event (i.e., an event in $\mathcal{S}$ for the neuronal pattern or subset A) and denote the multiset of all such A-events generated from $\mathcal{S}$ (i.e. one for each combination of spike times in $\mathcal{S}$, one for each neuron in A) by E^A. Notice that E^A is a multiset, since distinct combinations of spike times in $\mathcal{S}$, one for each neuron in A, may generate the same multiset $G \in \mathscr{M}(\mathcal{F})$ above. We denote by $E^A_{Sync^+}$ the subset of A-events that have a strictly positive degree of synchrony with respect to the operator *Sync*.

2.3 Spike-train Synchrony

We now characterize and quantify spiking activity in spike trains and synchrony among them. We do so by means of the family of what we call *support* functions. We first define the set $\mathcal{H}^n$ as follows, for $n \in \mathbb{N}$:

$$\mathcal{H}^n = \{G \mid G \text{ is a multiset in } \mathscr{M}(\mathcal{F}) \text{ and } |G| = n\}.$$

That is, the set $\mathcal{H}^n$ comprises all representations of n spikes (of n neurons) by (multisets of) influence maps. It is meant to capture all possible representations of A-events for $|A| = n$ and thus all possible specific synchronous spiking events.

The collection of all possible multisets over $\mathcal{H}^n$ is denoted by $\mathscr{M}(\mathcal{H}^n)$. The use of $\mathscr{M}(\mathcal{H}^n)$ in the definition of our family of support functions is motivated by the possibility that different A-events in E^A, for some $A \subseteq N$, $|A| = n$, may be represented by the same multiset in $\mathscr{M}(\mathcal{F})$.

Definition 3. *A collection of maps* $Supp_n : \mathscr{M}(\mathcal{H}^n) \to \mathbb{R}^+ \cup \{0, \infty\}$*, for* $n \in \mathbb{N}$*, is a family of* support operators *on* $\mathcal{F}$ *if the following conditions hold:*

- ***Upward monotonicity***. *For* $\mathcal{H}, \mathcal{G} \in \mathscr{M}(\mathcal{H}^n)$*,* $Supp_n(\mathcal{H}) \leq Supp_n(\mathcal{H} \uplus \mathcal{G})$*.*[2]
- ***Downward monotonicity***. *For* $\mathcal{H} = \{H_1, ..., H_k\} \in \mathscr{M}(\mathcal{H}^n)$ *and* $\mathcal{H}' = \{H_1 \uplus \{f_1\}, ..., H_k \uplus \{f_k\}\} \in \mathscr{M}(\mathcal{H}^{n+1})$*, with* $f_i \in \mathcal{F}$ *for all* $i \in \{1, ..., k\}$ *and* $k \in \mathbb{N}$*, we have that* $Supp_n(\mathcal{H}) \geq Supp_{n+1}(\mathcal{H}')$*.*

Let $\mathcal{G} \in \mathscr{M}(\mathcal{H}^n)$ be a multiset of influence maps. We call $Supp_n(\mathcal{G})$ the *support* or *amount of synchrony* (if $n > 1$) of $\mathcal{G}$ with respect to the operator $Supp_n$.

[2] The ordering $\leq$ is defined on the extended positive reals $\mathbb{R}^+ \cup \{0, \infty\}$ as conventional.

Intuitively, a support operator $Supp_n$ is meant to aggregate the synchrony of all individual A-events (with $A \subseteq N$, $|A| = n$) over the (parallel) spike trains of the neurons in A.

Upward monotonicity encodes the intuitive idea that the spiking activity or, for $n > 1$, the amount of synchrony of a certain collection of multisets in $\mathcal{F}$ (e.g. a collection of A-events in E^A, for some $A \subseteq N$, $|A| = n$) should be no less than the spiking activity or amount of synchrony of smaller collections contained in it.

Similarly, *downward monotonicity* is intended to capture the idea that synchronous spiking in a certain set of spike trains (the amount of synchrony of such a set) should be no less than the amount of synchrony of any of its proper subsets. This generalizes the monotonicity property in Definition 2 to the amount of synchrony among spike trains.

3 Synchrony Models

The most common method in tasks related to the identification of synchronous spiking in parallel spike trains—one may even say the almost exclusively applied method—is *bin-based* (i.e., it employs time discretization): spikes are binned in time intervals of equal length (time bins) that partition the recording-time interval $(0, T]$. In this approach, two or more spikes are considered to be synchronous if they lie in the same time bin[3] and thus what characterizes synchrony is the length and boundaries of time bins (although it is only the length of time bins that gives us the *intended* characterization of spike synchrony in this approach).

The bin-based characterization of synchrony can be easily formalized in our framework. For a time-bin length $d \in (0, T]$, each $a \in N$ and time t^a in $\mathcal{S}$ we can define, for example, the following influence map in $\mathcal{F}$:

$$f_{t^a}(x) = \begin{cases} \frac{1}{d} \text{ if } x \in \left(d \left\lfloor \frac{t^a}{d} \right\rfloor, d \left\lfloor \frac{t^a}{d} \right\rfloor + d\right] \\ 0 \text{ otherwise,} \end{cases}$$

where $\left\lfloor \frac{t^a}{d} \right\rfloor$ denotes the integer part of $\frac{t^a}{d}$. For $G \in \mathscr{M}(\mathcal{F})$ a multiset of influence maps, we define our synchrony operator as follows:

$$Sync(G) = \begin{cases} 1 \text{ if } \exists x \in \mathbb{R} : \min_{f_t \in G} f_t(x) > 0 \\ 0 \text{ otherwise.} \end{cases}$$

Synchronous spiking in the spike trains of neurons in $A \subseteq N$ can be quantified by $Supp_{|A|}(E^A)$, which we simply define here as follows:

$$Supp_{|A|}(E^A) = \sum_{G \in E^A} Sync(G).$$

The bin-based characterization of synchrony has two main drawbacks:

---

[3] *Clipping* is usually applied, that is, at most one spike per neuron is considered in each time bin. We follow this common practice here.

- **Boundary problem**. A collection of spikes separated by a time distance smaller than the bin length may be split into different bins and thus be regarded as non-synchronous.
- **Bivalence problem**. A collection of spikes can be either (fully) synchronous or non-synchronous. Small variations in the time distance between two spikes (possibly moving one of them over a time bin boundary) may cause a jump from (full) synchrony to non-synchrony and vice versa.

A graded, continuous characterization of synchrony, able to overcome these drawbacks, is certainly desirable. We present here a model within our framework that responds to this motivation.

The first step in setting our model is to decide the form of the influence regions around spike times within which positive synchrony degrees are defined. At least in the absence of any specific knowledge about noise or jitter in our recordings, symmetry of influence regions and influence maps around the points given by the corresponding spike times seems to be a reasonable choice. Likewise, in the absence of any evidence to the contrary, influence regions of the same length may be considered for all spikes of all neurons. We follow this view and consider for our model influence regions of a certain length, say $r \in \mathbb{R}^+$, that are symmetric around the corresponding spike times (i.e., for t a spike time in $\mathcal{S}$, we consider the influence region $[t - \frac{r}{2}, t + \frac{r}{2}]$).

We now characterize our influence maps. For our model we choose (possibly) the simplest subclass of $\mathcal{F}$, given by the maps of the following form, for t a spike time in $\mathcal{S}$ and $[t - \frac{r}{2}, t + \frac{r}{2}]$ the corresponding influence region:

$$f_t(x) = \begin{cases} \frac{1}{r} \text{ if } x \in [t - \frac{r}{2}, t + \frac{r}{2}], \\ 0 \text{ otherwise.} \end{cases}$$

For every $r \in \mathbb{R}^+$ we will denote the corresponding subclass of functions by $\mathcal{F}^r$.

We define our synchrony operator *Sync* as follows, for G a multiset in $\mathscr{M}(\mathcal{F}^r)$:

$$Sync(G) = \int_{-\infty}^{\infty} \min_{f \in G} f(x) dx. \tag{1}$$

We thus characterize synchrony among spikes by means of the *minimum* operator (a natural choice in the context: all influence maps must be positive for a point in time in order for this point in time to contribute to the synchrony) and measure their degree of synchrony by means of integration (that is, we aggregate over all points in time that contribute to the synchrony).

We move now to the definition of the family of support operators $Supp_n$ for characterizing and quantifying spiking activity of neurons and synchronous spiking of groups of neurons in our model. Let us start by assuming that the spike train of any single neuron in N is represented by a collection of influence maps in $\mathcal{F}^r$ with non-overlapping influence regions. (For reasonably small r this can be expected, due to the sparseness of spike times in real neuronal recordings and to the *refractory period* of neurons, that is, a certain period of time needs to elapse between the emission of two spikes—see, e.g., [14]). Under such an assumption the *naïve* characterization of $Supp_n$ as

$$Supp_n(\mathcal{H}) = \sum_{G \in \mathcal{H}} Sync(G), \tag{2}$$

for $\mathcal{H} \in \mathscr{M}(\mathcal{H}^n)$ and $n \in \mathbb{N}$, and the corresponding quantification of synchrony for $A \subseteq N$ given by $Supp_{|A|}(E^A)$ would constitute a reasonable choice. However, if the aforementioned assumption does not hold (i.e., if it is not the case that the influence regions of maps representing spikes emitted by the same neuron do not overlap) the above family of operators $Supp_n$ is not a family of support operators as set in Definition 3. An alternative characterization of $Supp_n$ in agreement with Definition 3 is thus needed for our model and we opt for

$$Supp_n(\mathcal{H}) = \int_{-\infty}^{\infty} \max_{G \in \mathcal{H}} (\min_{f \in G} f(x))\, dx, \tag{3}$$

for $\mathcal{H} \in \mathscr{M}(\mathcal{H}^n)$ and $n \in \mathbb{N}$. In words, we merge every collection of functions $G \in \mathcal{H}$ (e.g., every A-event in E^A for the neuronal pattern $A \subseteq N$) into the new function $\min_{f \in G} f(x)$, which characterizes synchrony in G, and integrate the maximum of all such functions over the real line. This removes the problems resulting from overlapping influence maps: overlapping influence maps can lead to overlaps of the functions $\min_{f \in G} f(x)$ that represent different A-events and thus certain time regions may be "counted" multiple times if we simply sum the synchrony of A-events. Taking the maximum eliminates this potential multiplicity and thus ensures downward monotonicity of our family of support operators.

Note that we can equivalently express $Supp_{|A|}(E^A)$, for $A \subseteq N$, as

$$Supp_{|A|}(E^A) = \int_{-\infty}^{\infty} \min_{a \in A} (\max_{f \in G_a} f(x))\, dx, \tag{4}$$

with $G_a = \{f_1^a, ..., f_{k_a}^a\} \in \mathscr{M}(\mathcal{F})$ the multiset representing the spike train in $\mathcal{S}$ corresponding to neuron $a \in A$ within our model.

Note that, given our choice of $\mathcal{F}^r$ as the class of influence maps, Equation (4) can be computed easily. This motivated us to prefer Equation (3) over alternative definitions of $Supp_n$ that are (arguably) better from a conceptual point of view. For example, one may consider to choose at most one A-event in E^A for each spike time in $\mathcal{S}$ in the computation of $Supp_n$ (e.g. those events which would maximize the support given to the corresponding collection of spike trains), so that no spike contributes to more than one synchronous event.[4]

4 Fuzzy Frequent Pattern Mining

We now move to the task of identifying frequent neuronal patterns, by which we mean sets of neurons in N whose corresponding spike trains in $\mathcal{S}$ have an amount

[4] Some of these alternative characterizations are being considered in current research.

of synchrony or support greater than or equal to a certain user-defined frequency threshold. We start by presenting a generalization of the standard frequent pattern mining problem based on what we call fuzzy transactions, which we refer to as the *fuzzy frequent pattern mining* problem—which however differs from other known fuzzy generalizations in the literature that receive the same name, such as e.g. that introduced in [15], based on the consideration of transactions with potentially missing items, or in [16], where fuzzy membership (to a transaction) is assigned to items.

In order to describe and formalize our problem we follow here [17] for standard frequent pattern mining terminology and notation.

Let $\mathcal{I} = \{a_1, ..., a_n\}$ be an item base, for $n \in \mathbb{N}$. A *fuzzy transaction* T (over $\mathcal{I}$) is a pair $\langle id, \eta \rangle$, where id is a unique *transaction identifier* and $\eta : \mathcal{P}(\mathcal{I}) \to [0, 1]$ is a function that assigns *degrees of membership* to all subsets of $\mathcal{I}$ (i.e., the pair $\langle \mathcal{P}(I), \eta \rangle$ is a *fuzzy set*, with η its definitory membership function—see, e.g., [7]).

A transaction $T = \langle id, \eta \rangle$ is said to *support* a set $J \subseteq \mathcal{I}$ to the degree ζ if $\eta(J) = \zeta$. For $\mathcal{D}$ a fuzzy transaction database, the *cover* of J in $\mathcal{D}$ consists of the set of transactions in $\mathcal{D}$ that support J to a degree strictly greater than 0. The *support* of J in $\mathcal{D}$ is given by a function of the membership degrees assigned to J by the transactions in the cover of J that is anti-monotone on the partially ordered set $\mathcal{P}(\mathcal{I})$ with respect to set inclusion.

A subset $J \subseteq \mathcal{I}$ is called *frequent* in $\mathcal{D}$ if its support in $\mathcal{D}$ is no less than some user-specified *minimum support* $\sigma_{\min}$. Our problem thus consists in finding the collection of all frequent subsets in $\mathcal{I}$, which we denote by $F(\mathcal{D}, \sigma_{\min})$.

Note that standard frequent pattern mining can be formalized in terms of fuzzy transactions by means of membership functions that assign value 1 to all subsets of, say, J—for some $J \subseteq \mathcal{I}$—and value 0 to all other subsets of $\mathcal{I}$ (which actually allows us to dispense with η altogether and simply specify the set J). The corresponding support would, in this case, be given by $\sum_{\langle id, \eta \rangle} \eta(J)$.

4.1 Frequent Neuronal Patterns

In our context, the set of neurons N plays the role of $\mathcal{I}$, the item base. In general, fuzzy transactions are formally obtained as follows: each point in time $t \in (0, T]$ gives rise to a transaction $\langle t, \eta_t \rangle$ (and thus we work with a continuous, i.e., uncountable database), where η_t is given by a function on the values taken at t by the influence maps representing spikes in $\mathcal{S}$.

In particular, in relation to our model based on a continuous characterization of synchrony given in Section 3, we have

$$\eta_t(A) = \min_{a \in A}(\max_{f \in G_a} f(t)) \qquad \text{and} \qquad Supp_{|A|}(E^A) = \int_{-\infty}^{\infty} \eta_t(A) dt,$$

for A and G_a as in Equation (4).

Frequent neuronal patterns in N will be those subsets $A \subseteq N$ whose corresponding spike trains (in $\mathcal{S}$) show an amount of synchrony greater than or equal to a certain support threshold $\sigma_{\min}$ (i.e., those subsets $A \subseteq N$ with $Supp_{|A|}(E^A) \geq \sigma_{\min}$).

4.2 Algorithms for Fuzzy Frequent Pattern Mining

In general, the main difference compared to standard frequent pattern mining algorithms (see, e.g., [17]) consists in how the support for sets $A \subseteq N$ is calculated. For the bin-based model of synchrony, the assessment of the support for sets $A \subseteq N$ is as in standard frequent pattern mining. Our alternative graded model of synchrony, based on the identification of spikes with influence maps in the class $\mathcal{F}^r$, favours algorithms suitable for a *vertical layout* of the database, similar in spirit to the well-known Eclat algorithm. In particular, the collection of sequences $\{f^{a_1}, ..., f^{a_k}\} \subset \mathcal{F}^r$ corresponding to the spike times for neuron a, for each $a \in N$, would constitute our primary database. In fact, due to the shape of influence maps in $\mathcal{F}^r$, the sequence of the corresponding influence regions suffices. Thus $Supp_{|A|}(E^A)$—or, equivalently, $Supp_{|A|}(E^A_{Sync^+})$—can be computed as follows: in a first step we form the union of all influence regions corresponding to neuron $a \in A$, for all $a \in A$ (this corresponds to taking the maximum of the influence maps over $(0, T]$ and finding the intervals in which it is not 0). In a second step, we intersect these unions in order to obtain the time intervals with $\eta_t(A) > 0$ (this corresponds to taking the minimum of the maxima of the influence maps and finding the region in which the result is not 0). Finally $Supp_{|A|}(E^A_{Sync^+})$ is computed by summing the lengths of the time intervals obtained and dividing by r.

5 Evaluation and Results

We provide some results[5] concerning a first evaluation of the two models presented in Section 3 on the identification of synchronous patterns from $\mathcal{S}$-like samples: in our settings, frequent patterns $A \subseteq N$ (with respect to some threshold $\sigma_{\min}$) whose support in $\mathcal{S}$ is significantly greater than that expected by chance under the assumption of independence in the spiking activity of neurons in A (found by means of a statistical significance test on the set of frequent neuronal patterns in N).[6] Synchronous patterns are the object of the *unitary events* analysis methodology, based on time-bin discretization (see [10]).

The results provided here aim at emphasizing some of the drawbacks of the bin-based model of synchrony for tasks such as the aforementioned, thus motivating alternative characterizations of synchrony within our framework, like the one introduced earlier in Section 3. For our evaluation we employed an Eclat-based routine in keeping with the indications given in Section 4.2—for the graded model of synchrony.

We generated two types of $\mathcal{S}$-like samples: samples of independent spike trains and of correlated spike trains.

[5] A full account of our evaluation methodology and results—mostly out of the scope of this paper—will be reported elsewhere, in the form of a journal publication.

[6] The statistical significance test is often redundant for this task. In practice, it is possible to determine a minimum support threshold and minimum pattern cardinality beyond which statistical significance of any neuronal pattern is granted. Research in this direction is being carried out at present.

Independent Spike Trains. Spike trains generated as Poisson point processes with properties similar to those of some experimental data. Two types of processes with typical average rates were considered: stationary and non-stationary (based on a phasic-tonic rate responses—see, e.g., [18]).

A small number of neurons was considered ($|N| = 10$) and the time duration was set to 1000 milliseconds (i.e., $T = 1000$) for all generated spike trains. Two data sets, each with 1000 trials for each neuron, were generated: one data set with independent stationary Poisson processes and the other with non-stationary Poisson processes.

Correlated Spike Trains. In order to generate data sets with varying amounts of synchrony among spike trains we essentially adopted the basic features of the SIP (Single Interaction Process) model—described in [19]—along with some modifications aimed at generating non-exact spike coincidences.

We generated two data sets like the ones above to represent the background activity of neurons in N. The size of spike coincidences was set to 10 (i.e., $|N|$). In order to determine the time and number of such coincidences a random choice of $n \in \{3, 4, 5, 6, 7\}$ points in the interval $[0, T]$ was considered for each trial. Each spike time generated this way was added to the background spiking activity of neurons in N by taking into account some random time deviation in order to produce *non-exact* spike coincidences, modeled by means of a uniform random variable on the interval $[\frac{-1}{2}, \frac{1}{2}]$.

The assessment of significant synchrony was done by manipulation of our original trials (i.e., generation of surrugate data). The method employed is *trial shuffling*, a well-established method for destroying spike synchrony across instances of possibly non-stationary point processes (see, e.g., [20]). From the surrogate trials obtained this way we assess the critical amounts of synchrony corresponding to each neuronal pattern $A \subseteq N$ beyond which, given a significance level of 0.01, we declare synchrony to be significant.

The two models presented in Section 3 were tested for influence-region and time-bin lengths $r, d \in \{1, 2, 3, 4\}$. Values for the minimum support $\sigma_{\min}$ were taken from the set $\{0, 1, 2\}$.

Our tests made some of the problems of the bin-based model evident. In particular, results on correlated spike trains show the negative effects of the boundary problem in the characterization of synchrony based on time-bin discretization. Even though (except for $d = 4$ in neuronal patterns of cardinality three) the critical amounts of synchrony estimated from surrogate data for every neuronal pattern $A \subseteq N$ of cardinality at least three were all smaller than three (recall that the number of injected coincidences in trials is at least three), the number of undetected joint-spike coincidences was in general high, even for reasonably large time bins (for $d = 3$ and even $d = 4$).

With respect to the two models, amounts of synchrony strictly greater than 0 in neuronal patterns beyond a certain cardinality are rare and thus, effectively, the significance level goes often far below 1% (since the critical amount of synchrony may be 0 and the condition for significance in our testing procedure was chosen to be an amount of synchrony *strictly greater than* the critical one). In the bin-based model this problem is also observed in relation to other amounts of synchrony. For example, the critical amount of synchrony (with respect to $d = 1$) corresponding to 4-neuron patterns is 1, being an amount of synchrony greater than or equal to 2 for these patterns very rare. Therefore, since we require an amount of synchrony strictly greater than the corresponding critical amount for a neuronal pattern to be considered synchronous, we will have as a result a very small number of synchronous patterns of this cardinality in these trials. Overall, discretization of the sample space in the bin-based approach greatly undermines our testing procedure.

6 Conclusion and Future Work

We have presented a general, flexible framework for the characterization and quantification of spike and spike-train synchrony—based on some general, desirable criteria such notions should meet—able to accommodate (possibly) graded notions of synchrony, like the one presented in Section 3 as an alternative to the bin-based characterization, aimed at overcoming some of its main drawbacks.

Motivated by the task of identifying frequent (and synchronous) neuronal patterns from parallel spike trains, we formalized the so-called fuzzy frequent pattern mining problem—a generalization of standard frequent pattern mining.

Alternative synchrony models are currently being tested on the basis of methodological and/or conceptual improvements. In particular, as pointed out earlier, quantification of spike-train synchrony based on the choice of at most one joint-spike event per spike in the computation of the amount of synchrony (i.e., $Supp_n$) is being considered.

References

1. Abeles, M.: Local Cortical Circuits: An Electrophysiological Study. Springer Verlag, Berlin/Heidelberg, Germany (1982)
2. Hebb, D.O.: The Organization of Behavior. J. Wiley & Sons, New York, USA (1949)
3. Abeles, M.: Role of the cortical neuron: integrator or coincidence detector? Israel Journal of Medical Science **18** (1982) 83–92
4. Goedeke, S., Diesmann, M.: The mechanism of synchronization in feed-forward neuronal networks. New Journal of Physics **10** (2008) 015007
5. König, P., Engel, A.K., Singer, W.: Integrator or coincidence detector? the role of the cortical neuron revisited. Trends in Neuroscience **19** (1996) 130–137
6. Buzsáki, G.: Large-scale recording of neuronal ensembles. Nature Neuroscience **7** (2004) 446–461
7. Zadeh, L.A.: Fuzzy sets. Information and Control **8** (1965) 338–353
8. Gerstein, G.: Gravitational clustering. In: Analysis of Parallel Spike Trains. Springer, Berlin/Heidelberg, Germany (2010) 157–172
9. Grün, S., Diesmann, M., Grammont, F., Riehle, A., Aertsen, A.: Detecting unitary events without discretization of time. Journal of Neuroscience Methods **93** (1999) 67–79

10. Grün, S., Diesmann, M., Aertsen, A.: Unitary event analysis. In: Analysis of Parallel Spike Trains. Springer, Berlin/Heidelberg, Germany (2010) 191–218
11. Feldt, S., Waddell, J., Hetrick, V.L., Berke, J.D., Zochowski, M.: A functional clustering algorithm for the analysis of dynamic network data. Physical Review E: Statistical, Nonlinear and Soft Matter Physics **79** (2009)
12. Gerstein, G.L., Perkel, D.H., Subramanian, K.: Identification of functionally related neural assemblies. Brain Research **140** (1978) 43–62
13. Syropoulos, A.: Mathematics of multisets. In: Multiset Processing: Mathematical, Computer Science and Molecular Computing Points of View. Springer, Berlin/Heidelberg, Germany (2001) LNCS 2235:347–358
14. Tuckwell, H.C.: Introduction to Theoretical Neurobiology, Volumes I and II. Cambridge University Press, Cambridge, MA, USA (1988)
15. Wang, X., Borgelt, C., Kruse, R.: Mining fuzzy frequent item sets. In: Proceedings of the 11th International Fuzzy Systems Association World Congress (IFSA05), Tsinghua University Press and Springer-Verlag (2005) 533
16. Delgado, M., Marín, N., Sánchez, D., Vila, M.A.: Fuzzy association rules: general model and applications. IEEE Transactions on Fuzzy Systems **11** (2003) 214–225
17. Goethals, B.: Frequent set mining. In: Data Mining and Knowledge Discovery Handbook. Springer, Berlin/Heidelberg, Germany (2005) 377–397
18. Louis, S., Borgelt, C., Grün, S.: Generation and selection of surrogate methods for correlation analysis. In: Analysis of Parallel Spike Trains. (2010) 359–382
19. Kuhn, A., Aertsen, A., Rotter, S.: Higher-order statistics of input ensembles and the response of simple model neurons. Neural Computation **15** (2003) 67–101
20. Grün, S.: Data-driven significance estimation of precise spike correlation. Journal of Neurophysiology **101** (2009) 1126–1140