Generating surrogates for significance estimation of spatio-temporal spike patterns

Alessandra Stella[1,3*] ☯, Peter Bouss[1,3] ☯, Günther Palm[1,2], Sonja Grün[1,3]

1 Institute of Neuroscience and Medicine (INM-6) and Institute for Advanced Simulation (IAS-6) and JARA Institute Brain Structure-Function Relationships (INM-10), Jülich Research Centre, Jülich, Germany
2 Institute of Neural Information Processing, University of Ulm, Ulm, Germany
3 Theoretical Systems Neurobiology, RWTH Aachen University, Aachen, Germany

☯These authors contributed equally to this work.
* a.stella@fz-juelich.de

Abstract

Spatio-temporal spike patterns were suggested as indications of active cell assemblies. We developed the SPADE method to detect significant spatio-temporal patterns (STPs) with ms accuracy. STPs are defined as identically repeating spike patterns across neurons with temporal delays between the spikes. The significance of STPs is derived by comparison to the null-hypothesis of independence implemented by surrogate data. SPADE binarizes the spike trains and examines the data for STPs by counting repeated patterns using frequent itemset mining.

The significance of STPs is evaluated by comparison to pattern counts derived from surrogate data, i.e., modifications of the original data with destroyed potential spike correlations but under conservation of the firing rate profiles. To avoid erroneous results, surrogate data are required to retain the statistical properties of the original data as much as possible. A classically chosen surrogate technique is Uniform Dithering (UD), which displaces each spike independently according to a uniform distribution. We find that binarized UD surrogates of our experimental data (motor cortex) contain fewer spikes than the binarized original data. As a consequence, false positives occur. Here, we identify the reason for the spike reduction, which is the lack of conservation of short ISIs.

To overcome this problem, we study five alternative surrogate techniques and examine their statistical properties such as spike loss, ISI characteristics, effective movement of spikes, and arising false positives when applied to different ground truth data sets: first, on stationary point process models, and then on non-stationary point processes mimicking statistical properties of experimental data. We conclude that trial-shifting best preserves the features of the original data and has a low expected false-positive rate. Finally, the analysis of the experimental data provides consistent STPs across the alternative surrogates.

Author summary

The generation of surrogate data, i.e., the modification of original data to destroy a certain feature, is used for the implementation of a particular null-hypothesis if an analytical approach is not feasible. Thus, surrogate data generation has been

extensively used to assess significance of spike correlations, e.g., by displacing spikes locally and uniformly (uniform dithering, UD). One of the challenges of surrogate methods is to properly construct the desired null-hypothesis distribution, i.e., to not bias the null-hypothesis through altering the spike train statistics. Here we compare five surrogate techniques to UD (two newly introduced) in the context of spatio-temporal spike pattern detection for their performance, first on controlled spike trains based on point process models, and second on modeled artificial data serving as ground truth to assess the pattern detection in a more complex and realistic setting. UD fails as an appropriate surrogate because it leads to a loss of spikes and thus to a large number of false-positive patterns. The other surrogates achieve a better performance in detecting precisely timed higher-order correlations. Based on these insights, we analyze experimental data from pre-/motor cortex of macaque monkeys during a reaching-and-grasping task for spatio-temporal spike patterns.

1 Introduction

Surrogate generation is a standard tool for the significance evaluation of precise spike time correlation [1–4]. As in recent years the methods for identification of higher-order correlations in multiple parallel spike trains are being further developed [5], also surrogate techniques need to be understood in more depth [6]. Surrogate generation is an alternative to the analytical derivation of a null-hypothesis when the definition of an exact test distribution is difficult or even impossible [7, 8]. Such an approach is also referred to as a bootstrap test of hypothesis and is widely used and effective [9]. When the null-hypothesis is rejected, the results obtained are said to be statistically significant. In the context of temporal coding, the goal is to test whether the present temporal correlations (with millisecond precision) are given by chance or represent a feature of the network dynamics. This corresponds to formulating a null-hypothesis where the fine temporal correlations are not present ("independent" model).

We are interested in detecting significant spatio-temporal spike patterns as signatures of active assemblies. Precisely timed higher-order correlations have been studied by numerous authors [10–16]. However, these correlation-based studies were not well suited to find all possible spatio-temporal patterns in recordings of hundreds of neurons. Therefore, we developed a method (called SPADE) that detects and evaluates spatio-temporal spike patterns at a precision level of milliseconds [17–19]. In contrast to other methods aiming at the detection of spatio-temporal patterns, it does not build on pairwise analysis and enables to derive the significance of the patterns [5, 20]. SPADE detects patterns of action potentials between different neurons with particular time intervals between the spikes that repeat identically. It was originally conceived as a technique for detecting synchronous spike patterns [17, 21], but has been further developed for detecting delayed spike patterns [18, 19, 22] and for testing them statistically [19]. The null-hypothesis for the significance test is generated through surrogate data that implement independence between the spike trains given the varying firing rates.

In electro-physiological experiments, nowadays the activity of hundred(s) of neurons is recorded. Therefore, the number of pattern candidates can be very high - in an example data set at our hand we find up to 10^9 patterns. As a consequence, testing each pattern individually is not feasible [23–25], as this would result in massive multiple testing. Moreover, some methods for pattern detection rely on the assumption of Poisson spike trains (see examples in [5]). This assumption does not generally hold for biological spike trains, and thus could lead to false-positive detection [26, 27]. This is the reason why we resort to surrogates in order to generate the null-hypothesis distribution. We generate surrogate data sets such that putative precise time correlations in the

original data are destroyed, while the remaining statistical characteristics of the spike trains, e.g., firing rate (co-)modulations, are preserved. A classical approach regarding surrogate generation is uniform dithering, often also called jittering or teetering [1, 28–31], and was largely employed in various experimental studies [2, 29, 32, 33].

In former studies, uniform dithering was used for the significance evaluation of synchronous spike patterns [17, 21] or in the context of calibrating the SPADE method for the evaluation of spatio-temporal spike patterns using non-homogeneous Poisson spike data [19]. However, we noticed, when extensively applying SPADE for the first time on data from monkey motor cortex [34] searching for spatio-temporal patterns, that the surrogates contain fewer spikes than the original data. This could yield false-positive detection [4, 35].

Thus, this study comprises the evaluation of this effect, and a careful and detailed analysis of alternative surrogate methods to be used as a solution. So, after describing the SPADE method and the role of surrogates therein in more detail (in section 2), we analyze the origin of spike count reduction in UD surrogates (section 3.1) and identify the destruction of the original inter-spike interval (ISI) distribution by UD [27, 32] as the cause for it. We show by simulation of defined stochastic processes that the subsequent discretization (binning followed by clipping) of this surrogate data removes more spikes than of the original data.

To avoid this we describe alternative surrogates - mostly from the literature [32, 36], but also newly defined ones - that aim to preserve the ISI distribution, and explain their concrete implementation (section 3.2). In section 3.3, these alternative surrogates are compared in their statistical properties and compared to UD. This is done by applying the surrogates to well-defined ground truth data which do not contain patterns, i.e., stationary point processes with ISI distribution with dead-time (Poisson process with dead-time, PPD), or with a peaked ISI distribution (Gamma process) and for comparison a Poisson process. The statistical aspects that are explored are the effects on the potential loss of spikes, on the ISI distributions, auto-correlations, and the coefficient of variation (CV), but also the effective movement of the spikes in the surrogate as compared to the original data. As a summary, we provide a table with preserved and destroyed features and give an outline of their expected impact on the appropriate estimation of significance.

In a further step, we evaluate the performance of the surrogates in respect to false-positive patterns (FPs) when applying them to ground truth data that are generated to closely model the experimental data (section 3.4) which are later also analyzed (section 3.5). The data are modeled as independent PPD or Gamma processes with instantaneous firing rates estimated on a trial-by-trial and neuron-by-neuron level from the experimental data. The PPD includes the same dead-time as the experimental data through the spike sorting process, and the Gamma process is generated with a shape factor using the CV of the real data. The modeled data are then segmented in behaviorally relevant epochs (as the experimental data), which are analyzed separately by SPADE. As a result, we get the amount of FPs per data set (PPD and Gamma) and observe that UD generates in both cases a large number of FPs as compared to the alternative surrogates. The number of FPs increases with the firing rates.

Finally, in the last section 3.5 we analyze two experimental data sets, each from a different monkey, employing all explored surrogates. We indeed find STPs, for monkey N in almost all epochs, for monkey L in less. SPADE with UD finds many more STPs, which we mostly interpret as FPs, as concluded from the previous analyses. In contrast, the other analyses of SPADE with the alternative surrogates yield fewer STPs but more than expected by the null-hypothesis. In addition, we obtain almost the same results (identical patterns and pattern counts) in the various epochs across each monkey,

supporting the robustness of the analysis. In the discussion, we integrate the results of all analyses and conclude on trial-shifting as the surrogate of choice, since it keeps the spike trains intact, leads to the smallest change of the statistical properties, thus it is expected to neither over- nor underestimate the significance of the patterns.

The software and data resources are made available and in the supplementary information we provide analytical and other details of the analyses.

2 Materials and Methods

2.1 SPADE

The Spike PAttern Detection and Evaluation (SPADE) method consists of various steps, as summarized in Fig 1.

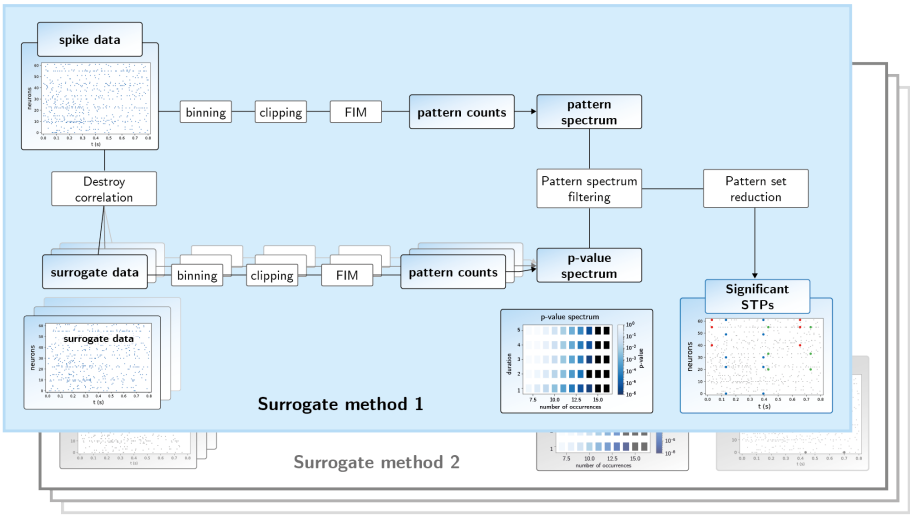


Figure 1. Workflow of the SPADE analysis. The top branch of the SPADE workflow shows the sequence of analysis steps of the original data until the pattern spectrum is derived. The bottom branch of the workflow starts with the generation of the surrogate data from the original data, followed by the same analysis steps as for the original data. The multiple overlapping panels in the lower branch indicate that this surrogate procedure is repeated many times, by which the p-value spectrum is built up. This then serves for the extraction of significant patterns through pattern spectrum filtering. After the application of the pattern set reduction, the significant STPs are provided as a result. 'Surrogate method 2' indicates that the part 'Destroy correlation' and 'surrogate data' are replaced by another surrogate method, but the other steps stay the same.

The spike train data are first discretized into exclusive time intervals (bins). Typically, the bin length consists of a few milliseconds, which at the same time defines the allowed temporal imprecision of the neuronal coordination. The procedure of discretization counts the number of spikes within each bin (binning, [3,17]), followed by reducing the bin content to 1 if a bin contains more than 1 spike (clipping). In the following we will call the combination of these two steps binarization. Candidate spatio-temporal patterns are then mined using the Frequent Itemset Mining (FIM) algorithm [23,37,38], which yields the number of occurrences of each non-trivial detected pattern, along with its occurrence times. A non-trivial pattern occurs at least a fixed number of times but is not explained as part of a larger pattern. The binarized data are the input to the mining algorithm. Pattern counts are then collected in the so-called *pattern spectrum*, i.e., the pattern counts are entered in a 3d-histogram according to

their number of spikes z, the number of pattern repetitions c, and the temporal extent from first to last spike d (see Fig 2C). We call the triplet $\{z, c, d\}$ a *signature*.

FIM efficiently collects and counts pattern candidates, nonetheless, the probability (thus the statistical significance) of each of the mined patterns has still to be evaluated. We aim at testing whether the patterns emerge as an effect of precisely timed neuronal coordination, or merely by the firing rate profile of independently firing neurons. For this reason, the null-hypothesis is that spike trains are mutually independent given their firing rate (co-)modulations and that the occurring patterns in the data are given by chance. All patterns that have a p-value lower than the threshold (after multiple testing correction) reject the null-hypothesis and are classified as significant.

To implement the null-hypothesis of independence of the spike trains, we resort to the generation of surrogate data. The goal is to generate data from the original spike trains in such a way that putative precise spike correlations in the original data are destroyed, while the remaining statistical characteristics of the spike trains, such as the single neuron firing rate modulations are preserved. The classical approach regarding surrogate generation is uniform dithering [1, 28, 29]: to obtain a surrogate data set from the original data, uniform dithering displaces each spike independently by a random amount δ from its original position. δ is sampled from a uniform distribution $U([-\Delta, +\Delta])$, and the maximum displacement Δ is usually a multiple of the bin size [36].

Each surrogate data set is subject to the same procedure as the original data: binarization, followed by FIM, which results in pattern counts per signature. To generate reliable statistics, we generate and analyze many surrogates (in the range of thousands: here 5000). From these 5000 sets of pattern counts, we compile a p-value spectrum, i.e., a 3-dimensional matrix containing for each signature the fraction of surrogates data sets containing patterns with that signature. This yields for each signature $\{z, c, d\}$ a p-value (Fig 2C). Due to the large amount of tests performed, we correct the significance level by the False Discovery Rate correction [39]. The number of tests considered is the number of occupied entries of the pattern spectrum having the highest number of occurrences per size and duration. We call this step of testing the significance *pattern spectrum filtering* (PSF). By comparing the pattern counts of the original data to the p-value spectrum, potentially significant patterns are extracted, and are further filtered by the *pattern set reduction* (PSR) [17]. Pattern set reduction consists of conditional tests of each pattern given any other pattern surviving the PSF, in order to remove spurious false positives resulting as a by-product of the overlap of true patterns and chance spikes.

When analyzing large-size experimental data, the search for all possible patterns can result into obtaining millions, if not billions, of putative patterns. Thus, the analysis can quickly become infeasible due to memory or time requirements. In order to optimize our analyses, we run the FIM algorithm separately per pattern size (from size 2 to size 10+ in steps of one). For each pattern size, we set a minimum number of occurrences min_{occ} a pattern has to occur to be further considered after the mining step. This minimum number is a rough estimate of the number of patterns expected by chance under independence of the spike trains, and we assume stationary Poisson processes with the average rate of each of the spike trains. Patterns with a lower number of occurrences are considered as spurious as they would be rejected anyway by the following statistical test. In addition, we fix the minimum number of occurrences of a pattern to be at least 10 *(*$min_{occ} \geq 10$ for all sizes; i.e., 30% of the number of trials of the considered experimental data). The FIM output is then aggregated for the pattern spectrum filtering and pattern set reduction.

Finally, as a result of the explained steps, we get significant patterns, i.e., STPs. For each pattern, we have the information about its number of occurrences, the involved

neurons, the lags between the pattern spikes, and the times of occurrences.

3 Results

3.1 Spike count reduction in surrogates generated by uniform dithering

The SPADE method is developed to be applied to experimental data. We started to analyze parallel spike trains recorded during a reach-to-grasp experiment in monkey pre-/motor cortex [34, 40]. The data were recorded with a 100-electrode Utah array and resulted in about 100-150 single units per session. The firing rates of the neurons and the regularity of their inter-spike intervals (ISIs) were shown to be highly variable and behavior-dependent [41, 42].

Hence, we checked if the statistical features of the spike trains are conserved after discretization and surrogate generation within our SPADE analysis. In particular, we tested the spike counts per neuron before and after the binarization steps for both the original and surrogate data. To our surprise, we noticed that for some neurons the total spike counts of the surrogates were much lower than those of the original data. Further analysis of this aspect showed, as a function of the mean firing rate of each neuron, that the higher the firing rate of the neuron the larger the spike count reduction and the corresponding mismatch. Fig 2A, shows that for two different data sets, each from a different monkey (data are published and freely accessible in [34]), we find a spike count mismatch of up to 10% between the surrogates and the original data (Fig 2A, bottom, gray). Such a difference in the spike count is troublesome, since it is expected to lead to a reduced pattern count of the surrogates as compared to the original data and, thus, is expected to yield an overestimation of the significance of the patterns.

3.1.1 Uniform dithering

The uniform dithering method consists in displacing each individual spike of each neuron by a small uniformly distributed random jitter $\sim U[-\Delta, +\Delta]$ around its original position. An example sketch is shown in Fig 4A. It is also known by the names: "jittering" or "teetering" and is a classical choice for surrogate generation and was employed in several experimental studies [2, 21, 29, 32, 33, 43]. It is widely used for its simplicity and computational speed for synchrony detection of pairwise (i.e., cross-correlogram significance estimation, [1, 4]), higher-order synchrony, and pattern detection [2, 44, 45]. In particular, it was chosen as the surrogate generation technique for synchrony and pattern detection using SPADE [18, 19, 21]. Nonetheless, in [35] the authors already demonstrated in the context of analyzing pairwise synchrony, that uniform dithering (UD) may lead to false positives in the case of regular firing properties (CV < 1).

The dither parameter of the method $\Delta > 0$ determines the maximal displacement of a spike from its original position. Typically, it is chosen to be a multiple of the bin size parameter (e.g., $15 - 25$ms), and must be chosen appropriately. If Δ is too small, it causes insufficient displacement of the correlated spikes and may lead to an underestimation of significance, whereas if Δ is too large, it yields a strongly smoothed firing rate profile [31] and thus to an inappropriate null-hypothesis. In how far it also affects other intrinsic statistical properties of the spike trains will be studied in this contribution.

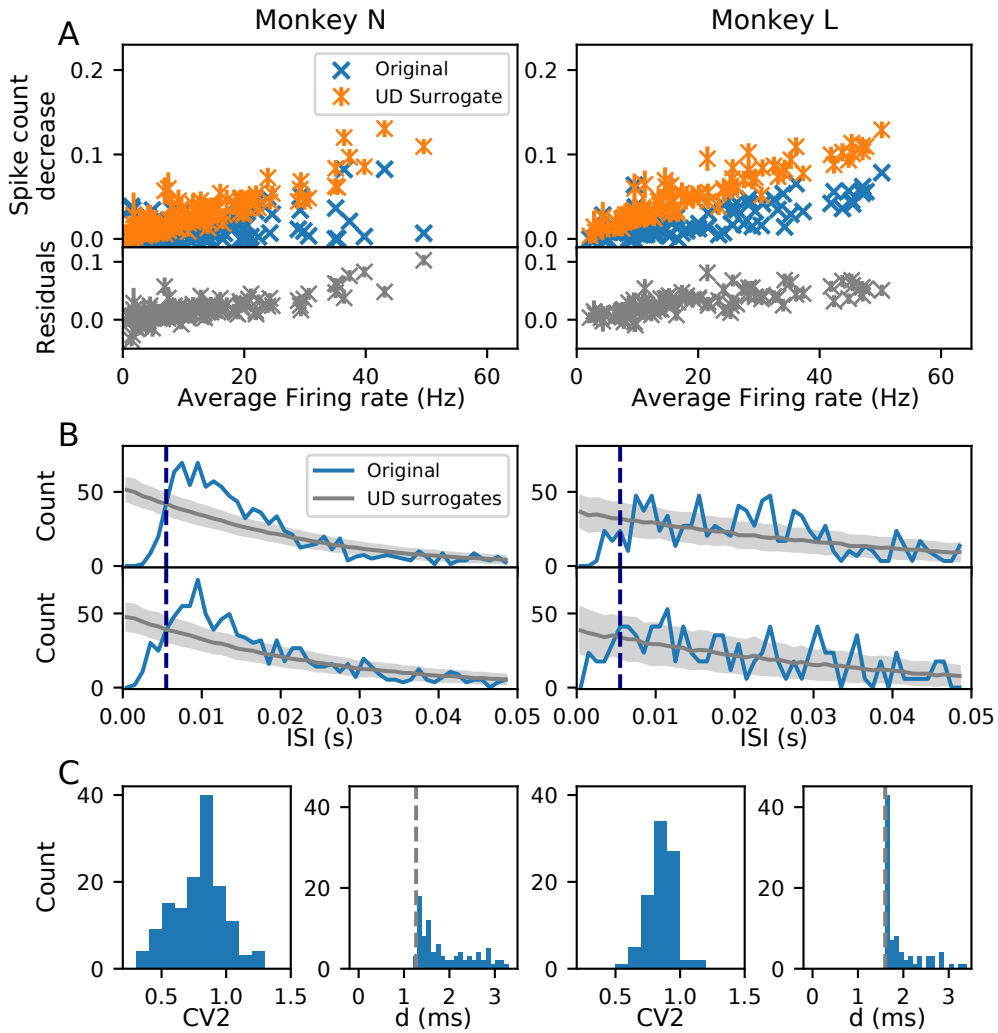


Figure 2. Modification of spike trains due to binarization. (A). Spike count reduction resulting from binarization and UD surrogate generation. Results from the analysis of two experimental data sets (sessions *i140703-001* and *l101210-001*) in the movement epoch of the trial type precision grip-high force (PGHF) of monkeys N (left) and L (right). Top panel: Spike count decrease as a function of the average firing rate. Blue crosses indicate the spike count reduction caused only by binarization of the original spike trains (one cross per neuron). Orange marks show the spike count reduction after surrogate generation by UD and binarization. The spike count reduction is normalized by the spike counts of the original continuous-time spike train. Orange bars indicate the standard deviation of the spike count reduction calculated across 100 surrogates. Bottom panel: residuals (gray) computed as the spike count difference between the original binarized spike trains (blue) and the UD binarized surrogates (orange), normalized as in the top panel. (B) and (C). Interval statistics of the data. B shows the ISI distribution of 2 neurons from monkey N (left) and 2 neurons of monkey L (right; in blue). Neurons represented are, with (channel-id, unit-id) notation, (7,1) and (16,1) for monkey N, (5,1) and (6,1) for monkey L. In gray are the ISIs of the respective UD surrogate distributions with the mean (dark gray) and the standard deviation of 500 surrogates in light gray. The bin size of the binning (here: 5ms) is shown by the dashed dark blue line. In C the CV2 distributions are shown for all neurons in each of the data sets (C, left subpanel). C, right subpanels, show the respective minimal ISI from the ISI distributions of all neurons. The dead-times assigned by the spike sorting algorithm are indicated by the dotted gray line (1.2ms for monkey N and 1.6ms for monkey L).

3.1.2 Origin of spike count reduction

The UD procedure as such is not deleting any spike. However, the SPADE workflow Fig 1 illustrates that the analysis steps of the original data and the surrogate data are identical from binning and clipping on. The only difference between the two data sets is the manipulation of the original data to generate the surrogate data set. Therefore, in the following, we will carefully analyze the reason for the spike count reduction.

Only the binarization step, i.e., binning followed by clipping, may reduce the spike count content in a bin. This step, applied to the original and to the surrogate data, leads to different results. A potential reason may rest on the inter-spike interval (ISI) statistics of the spike trains with and without dithering. Fig 2B, shows the ISI distribution for two example neurons of the original data (in blue; right for monkey N, left for monkey L) and for comparison, the ISI distributions of the uniform dithered surrogates (gray). In the original data, the ISI distribution is peaked at a certain ISI, here between 5 and 10ms, but the ISI distributions of the surrogate data are decaying exponentially, and thus also fill small ISIs. This indicates the fact that the original spike trains are more regular than the surrogates and small ISIs have a lower probability. This is confirmed by the measurements of the coefficient of variation, here CV2 which compensates for non-stationarity firing rates [46]. Indeed, the CV2 distribution of all neurons of each of the two data sets (Fig 2C, left subpanels) is rather below 1, i.e., more regular than Poisson. This raises the question if the surrogates lead indeed not only to different ISI distributions, but also to different CV2s, which will be investigated in section 3.3.

In addition, the distributions of the minimal ISI of each neuron per data set (Fig 2C, right subpanels) exhibit a minimal ISI of 1.3 ms for monkey N and of 1.6ms for monkey L. This corresponds to the dead-times of the spike sorting algorithm that cannot resolve overlapping spikes. The different dead-times for the two monkeys are due to a different number of sample data points considered for spike sorting [34]. The corresponding ISI distributions of the surrogate data (Fig 2B) show that there are ISIs smaller than the minimum ISI of the respective experimental data. Thus the dithering procedure generates shorter ISIs in the UD surrogates than in the original data.

To verify our interpretation that the combination of UD and binarization causes the spike count reduction, we perform a similar analysis on well-defined ground truth data, i.e., simulated spike trains of a well-defined point process model and parameters. We choose two types of stochastic point processes that model spike trains exhibiting similar ISI distributions as the experimental data. One is a Poisson process with a hard dead-time (PPD, [47]), and the other is a Gamma process. For simplicity, we now choose both of a constant rate but adapt the dead-times or shape factors, respectively, to account for the ISI features of the experimental data. The PPD is a variation of the classical Poisson process wherein no spike is generated within an interval to the previous spike smaller than the dead-time d. The Gamma process has a shape parameter γ which is related to the intrinsic regularity/irregularity of the spike train [48]. If $\gamma > 1$, the process is regular (i.e.,CV < 1); if instead $\gamma = 1$, it coincides with the classical Poisson process with an exponential ISI distribution, which here serves for reference. We do not consider here Gamma processes with CV > 1, i.e., with bursty spike trains.

Fig 3 shows the spike count reduction (expressed as $1 - N_{clip}/N$, where N_{clip} is the number of clipped spikes and N the total number of spikes) that results after binning (5ms) of the PPD (left) and Gamma process (right) with (dashed) and without (solid) application of uniform dithering, here obtained through analytical derivation (see Supplementary Information S1). For the PPD and Gamma models, we vary the dead-time parameter d and the shape factor γ, respectively. The graphs show the spike count reduction as a function of the firing rate of the processes. The PPD model shows a higher spike loss the higher the firing rate. The spike reduction is lower for larger

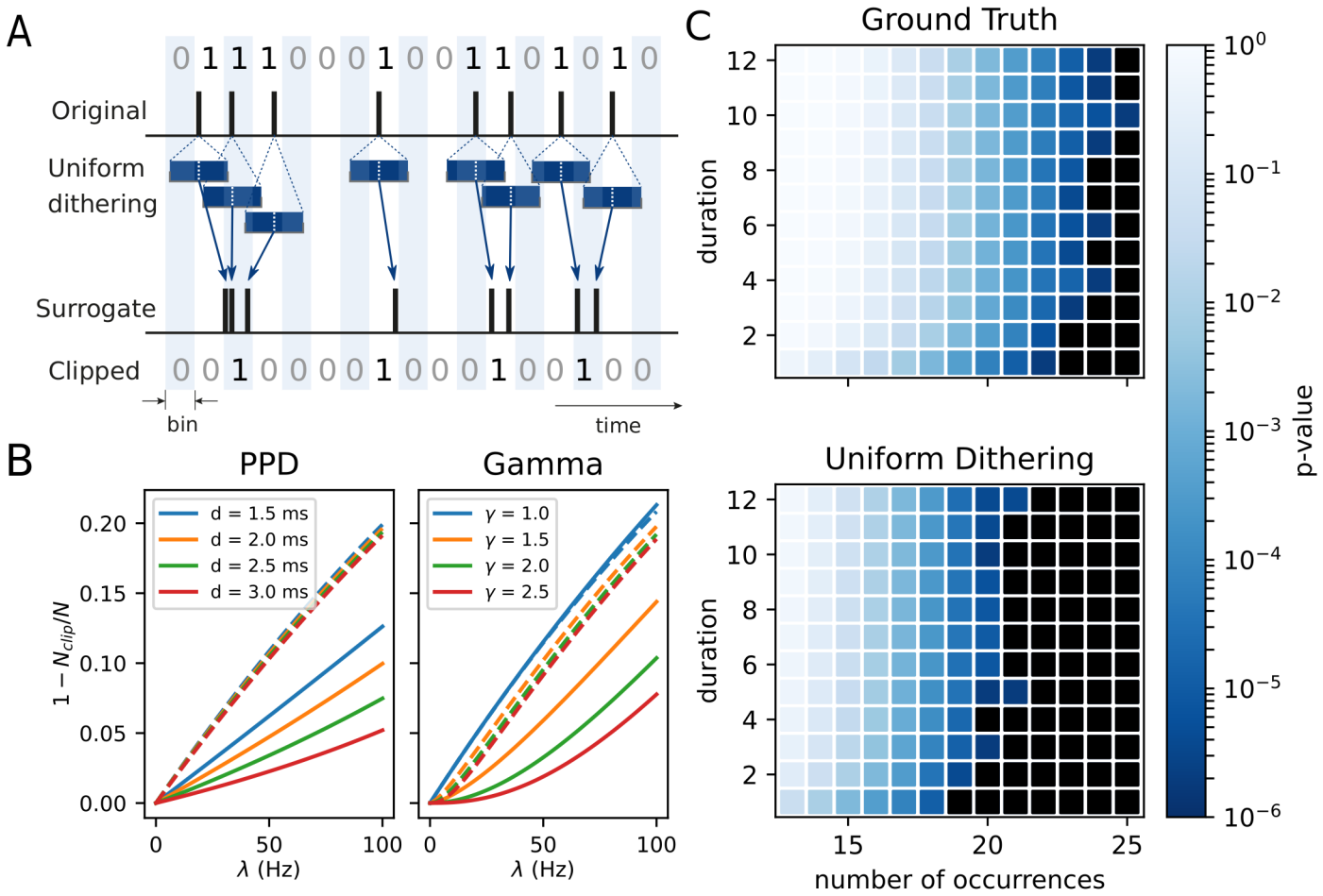


Figure 3. Origins of spike count reduction. (A). The sketch shows how a regular spike train is binarized. Below, it is illustrated how UD may change the spike times such that multiple spikes end up in single bins. The resulting binarized surrogate spike data are shown at the bottom. In contrast, due to the regular ISIs of the original process, its binned data are hardly losing spikes in comparison to the dithered version. (B). Spike count reduction (after binning in 5ms intervals and clipping) shown as analytical results (see Supplementary Information S1) for renewal point process models (PPD, left and Gamma process, right; solid lines, respectively), each with 4 different parameter sets (PPD: $d = 1.5, 2.0, 2.5, 3.0$ms, Gamma: $\gamma = 1, 1.5, 2, 2.5$, different colors). The dashed lines show the same quantity for their UD surrogates. The firing rate of the processes is also varied and shown along the x-axis. The spike count reduction is shown on the y-axis, expressed as $1 - N_{clip}/N$, where N_{clip} is the number of clipped spikes and N is the total number of spikes. (C). P-value spectrum of the original data (top) and of their UD surrogates (bottom) for mined patterns (of size 3 only) for a range of different pattern durations d (y-axis), and pattern counts (x-axis). The p-values are expressed by colors ranging from dark blue to light blue (see the color bar, identical for both spectra). The bin size is 5ms. The PPD data are of 100 realizations of $n = 20$ parallel independent spike trains with parameters $\lambda = 60$Hz, $d = 1.6$ms. The p-value spectrum for its uniformly dithered surrogates of the PPD data are derived from 5000 surrogates, dither parameter $\Delta = 25$ms.

dead-times. The uniform dithered PPD show for all dead-times an increase in the spike reduction with higher firing rates, but to a larger extent than for the original PPD processes. The Gamma process (right) for $\gamma > 1$ shows a similar result as for the PPD: an increase of spike count reduction for increasing firing rates, and the larger the shape factor, the lower the spike count reduction. The increase is rather parabolical compared to the PPD. The Poisson process (blue, $\gamma = 1$) shows a much larger and linear increase of spike reduction with rate, more strongly than for Gamma processes with $\gamma > 1$ (orange, green, red).

So, a) why does a Poisson-like process lose more spikes through binarization than a process with a non-exponential ISI distribution, and b) why does uniform dithering lead to a loss of spikes compared to the original experimental data? As shown above (Fig 2B), uniform dithering generates a more Poisson-like ISI distribution. Such a process contains spikes that follow each other in short intervals. Such a cluster of spikes may fall within a bin, and then the content of the bin is reduced to 1 by clipping (see for illustration Fig 3). A PPD process has a strict dead-time, leading to the fact that the probability of more than one spike in a bin is reduced, and thus there is a smaller spike loss. In fact, the goal of preserving the dead-time is to reduce the difference between the binned and clipped spike counts of the original spike train and its surrogates. The closer the dead-time to the bin size, the more unlikely it is that two spikes are dithered into the same bin (because of the dead-time constraint). Thus, the spike counts of the surrogates become closer to the spike count of the original spike trains.

3.1.3 Consequences of spike count reduction

Through the spike count reduction in the surrogate data also the expected occurrences of patterns may be reduced. This we prove by analyzing independent data comparing to UD surrogates. In Fig 3C, we represent the results of such a comparison. We generate 20 parallel artificial independent PPD spike trains with a stationary firing rate of $\lambda = 60$Hz, and with a dead-time of $d = 1.6$ms, and a total length of 1s. From these spike trains, we generate 5000 surrogates, each with uniform dithering, discretize them and count the number of patterns using FIM. Based on the generated pattern spectrum, we calculate the corresponding p-values. In addition, we aim to compare the p-value spectrum of the surrogate data (Fig 3C, bottom panel) with the p-value spectrum of the ground truth data. In order to do so, we generate a large number of the independent PPD processes and analyze these by FIM to extract the patterns and generate the p-value spectrum. In other words, knowing the ground truth, we can use new spike train realizations as surrogates. Fig 3C, top panel, shows the p-value spectrum of the PPD processes, for patterns of size 3, across different pattern durations. Since these data are independent, the resulting patterns are occurring by chance. The p-value spectrum of the surrogates of the PPD processes is shown below. The two p-value spectra illustrate that the p-values of the UD surrogates are smaller and UD surrogates have fewer pattern counts as compared to the ground truth data. In other words, pattern counts of the ground truth (PPD) data would become significant if compared to its UD surrogates although the ground truth data sets are independent. As a consequence, patterns in the independent original data would be classified as significant while being false positives.

From this analysis, we conclude that uniform dithering is not an appropriate surrogate method for spike data that either contain a hard dead-time or have a regular spiking behavior, as motor cortex data tend to have [42]. Therefore, we will next explore various surrogate techniques that are appropriate for such data.

3.2 Alternative surrogates

Here, we test five alternative surrogate methods, three known from the literature and two newly developed by us, and evaluate their applicability to our type of data. The goal is to find a surrogate type that preserves relevant features of the data, e.g., the spike count.

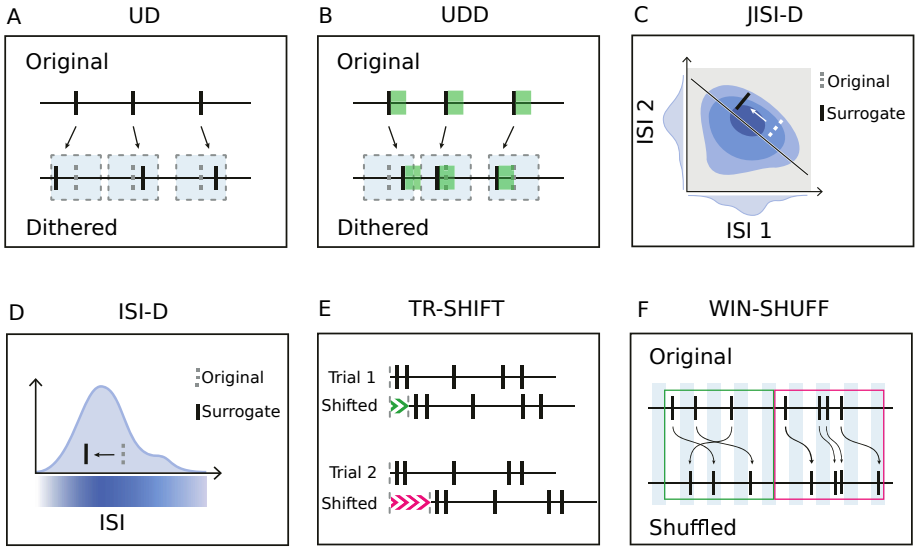


Figure 4. Sketches of the explored surrogate techniques. (A). Uniform Dithering (UD). Each spike is displaced according to a uniform distribution centered on the spike. Grey dotted rectangles represent the dithering window. (B). Uniform Dithering with dead-time (UDD). Similar to uniform dithering, but spikes are constrained not to be closer to each other than a given dead-time. Green shadows represent the dead-time after each spike. Grey dotted rectangles represent the dithering window. (C). Joint Inter-Spike Interval Dithering (JISI-D). Each spike is displaced according to the J-ISI distribution of the neuron, sampled from the data. J-ISI distribution, projected in a two-dimensional plane, is represented in blue. On the x and y-axes we represent the projections of the first and second ISI given three spikes. (D). Inter-Spike Interval Dithering (ISI-D). Each spike is displaced according to the ISI distribution of the neuron, sampled from the data. The ISI distribution is represented in blue, along with its intensity. (E). Trial Shifting (TR-SHIFT). Each trial is shifted by a randomly chosen amount from a uniform distribution (represented in green and pink), independently across trials and neurons. (F). Window Shuffling (WIN-SHUFF). Binned spike data are shuffled within exclusive windows (marked green and pink).

Different types of surrogates were already developed, however, in the context of different analysis methods. Some are concerned with regularity of spike data [1, 2, 21, 28, 29, 32, 33, 35, 36, 43]. The goal here is to find methods that are appropriate in the context of analyzing data for spatio-temporal patterns, i.e., spike patterns with allowed time intervals between the pattern spikes, and for different total durations of the patterns. An important property of surrogates is that some techniques preserve non-stationary firing rates as well. This was the focus of a series of papers [1, 4, 27, 35, 36]. Here, we aim to evaluate already existing techniques, and two new ones to be introduced, for their ability to preserve relevant features of the data, such as spike counts, non-stationarity of firing rates, ISI distribution, auto-correlation, joint-ISI distribution, etc.

In the following, we evaluate the surrogate methods which we think consider the requested aspects. Uniform dithering was explained in the former section and will be included in the upcoming tests for reference.

3.2.1 Uniform dithering with dead-time

We introduce uniform dithering with dead-time (UDD) as a variant of the classical uniform dithering [49] (Fig 4D). As the name states, the estimated dead-time d is conserved during the temporal displacement of each spike. This is done by limiting the window into which spikes can be dithered. The uniform displacement of each spike is restricted such that it is limited to the intervals to the neighboring spikes minus the dead-time d. Thus, it does not allow two dithered spikes to have a temporal distance smaller than the dead-time. Unlike for UD, the displacement of each spike is hence not independent of its neighbor. As described above in section 3.1.2, a dead-time may be introduced by spike sorting. Further, the biological absolute refractory period of neurons can yield minimal intervals larger than those inserted by the spike sorting. In order to estimate the dead-time, we fit it to the minimum inter-spike interval of the total spike trains. In case of low firing rate spike trains, it can happen that the minimum ISI is still in the range of hundreds of milliseconds, complicating the estimation of a biologically reasonable dead-time. For this reason, we define a parameter d_{max} such that, if the minimal inter-spike interval exceeds d_{max} (in our case 4ms), then we set $d = d_{max}$.

3.2.2 Joint-ISI dithering

Gerstein [32] suggested dithering spikes according to their joint-ISI distribution of adjacent ISI intervals. Joint inter-spike interval dithering (JISI-D; Fig 4E; [32, 35]) aims to keep the distribution of the preceding and following inter-spike intervals relative to a spike according to the joint-ISI probability distribution. This probability distribution is derived for each neuron from its spike train by calculating the corresponding joint-ISI histogram (with a default bin size of 1 ms). Dithering one spike according to such a two-dimensional histogram corresponds to moving the spike along the anti-diagonal of the joint-ISI distribution [32, 35]. The details of the surrogate procedure are described in the Supplementary Information S2.

Unfortunately, the recordings are often non-stationary and too short to comprise enough spikes to estimate the underlying joint-ISI probability distribution. Therefore, we apply on the joint-ISI histogram a 2d-Gaussian smoothing with variance σ^2, with σ of the order of milliseconds [35, 49]. We suggest using higher values for σ for sparser data sets. Nonetheless, there is no clear criterion for precisely representing the data without over-fitting them.

3.2.3 ISI dithering

ISI dithering (ISI-D; Fig 4F) is not considering the pair of current and subsequent ISI, instead, it dithers the individual spikes according to the ISI probability distribution only. However, for practical reasons, we implemented ISI-D as a special case of the joint-ISI dithering assuming that two consecutive ISIs are independent, i.e., that the joint-ISI histogram can be represented as the product of the ISI histogram with itself, ($p_{\text{j-ISI}}(\tau, \tau') = p_{\text{ISI}}(\tau) \cdot p_{\text{ISI}}(\tau')$; Supplementary Information S2). Thus, in comparison to the joint-ISI dithering, ISI-D does not take into account the correlations of subsequent ISI pairs, and is particularly useful when there are not enough data to estimate the joint-ISI distribution. As a result, not the distribution of pairs of subsequent ISIs are preserved, but the distribution of the single ISIs irrespective of their order.

3.2.4 Trial shifting

An alternative to single spike dithering is dithering the entire spike train. Trial shifting (TR-SHIFT; [1, 36]) consists of shifting all spike times identically by a random uniform amount $\sim U[-\Delta, +\Delta]$, independently neuron by neuron and trial by trial. The method

requires the time randomization to be different across neurons in the different trials, such that repeating identical patterns are shifted into different patterns from trial to trial. Therefore, the method is only applicable to data which can be segmented into different trials in a meaningful way. Alternatively, the data have to be segmented artificially, e.g., spontaneous/ongoing data.

TR-SHIFT has the benefit of keeping the entire spike train structures intact during each trial, i.e., firing rates, ISI distributions, and auto-correlation (1).

3.2.5 Window shuffling

We also introduce window shuffling (WIN-SHUFF), which divides the spike train into successive and exclusive small windows of predefined duration Δ_{WS}, and further divides the windows into bins of length b (Δ_{WS} should then chosen to be a multiple of b). The bins are then shuffled within each window, and spike times are randomized within each bin. This method conserves the number of spikes of the original data and the number of occupied bins in the original but already discretized data for each surrogate, i.e., there is no risk of spike count reduction (1). The bin size b is typically chosen to be equal to the bin size used in the SPADE analysis. The firing rate profile is modified by the local shuffling of the spikes, and its degree of resemblance to the original profile depends, similarly as for any other dithering technique, from the parameter Δ_{WS}, i.e., the window of the randomization. To facilitate the comparison to the other methods, we fix throughout the paper $\Delta_{\mathrm{WS}} = 2\Delta$.

3.3 Statistical comparison of surrogate methods

To better understand the statistical features preserved or changed by the suggested surrogate methods (Table 1), we compare the features numerically on stationary and independent data. For these data, we simulate point process models with well-defined properties: a Poisson process as a reference, Poisson process with dead-time (PPD), and a Gamma process. The latter two are chosen to mimic the ISI distributions of the experimental data (Fig 2B). Here the processes are chosen to be stationary, in order to be able to exclude yet another statistical aspect. In the next section (3.4), we will include non-stationary firing rates. We explore the effect of all the six surrogate methods onto the statistical properties of the stationary and independent data ('original data'), such as the loss of spikes, ISI properties, effective moving of spikes. The parameters of the data models are adjusted to be close to the experimental data at our hand and thus to enable the transfer to experimental data to be analyzed later (section 3.5).

Fig 5 summarizes all the results. The columns refer to the different spike train models (Poisson, PPD, and Gamma, from left to right, respectively) and in A the spike loss as a function of rate, in B the ISI distributions, in C the cross-correlation between the original and the surrogate data, and in D the auto-correlation of the original and the respective surrogate data. The 4th column displays a comparison of the CV of the original vs the surrogate data, the effectiveness of the displacement of the spikes through the surrogates and the change of a rate step through the surrogate (from top to bottom).

3.3.1 Spike Count Reduction related to Spike Train Statistics

Fig 5A illustrates the spike count reduction for all types of surrogates applied to the three different spike train models: the higher the rate the more spikes are lost. However, there are some differences in the degree of spike loss for different data models, which we now discuss separately.

Poisson process as original data. For Poisson data, the spike count loss increases approximately linearly with firing rate for all surrogates. This happens as well

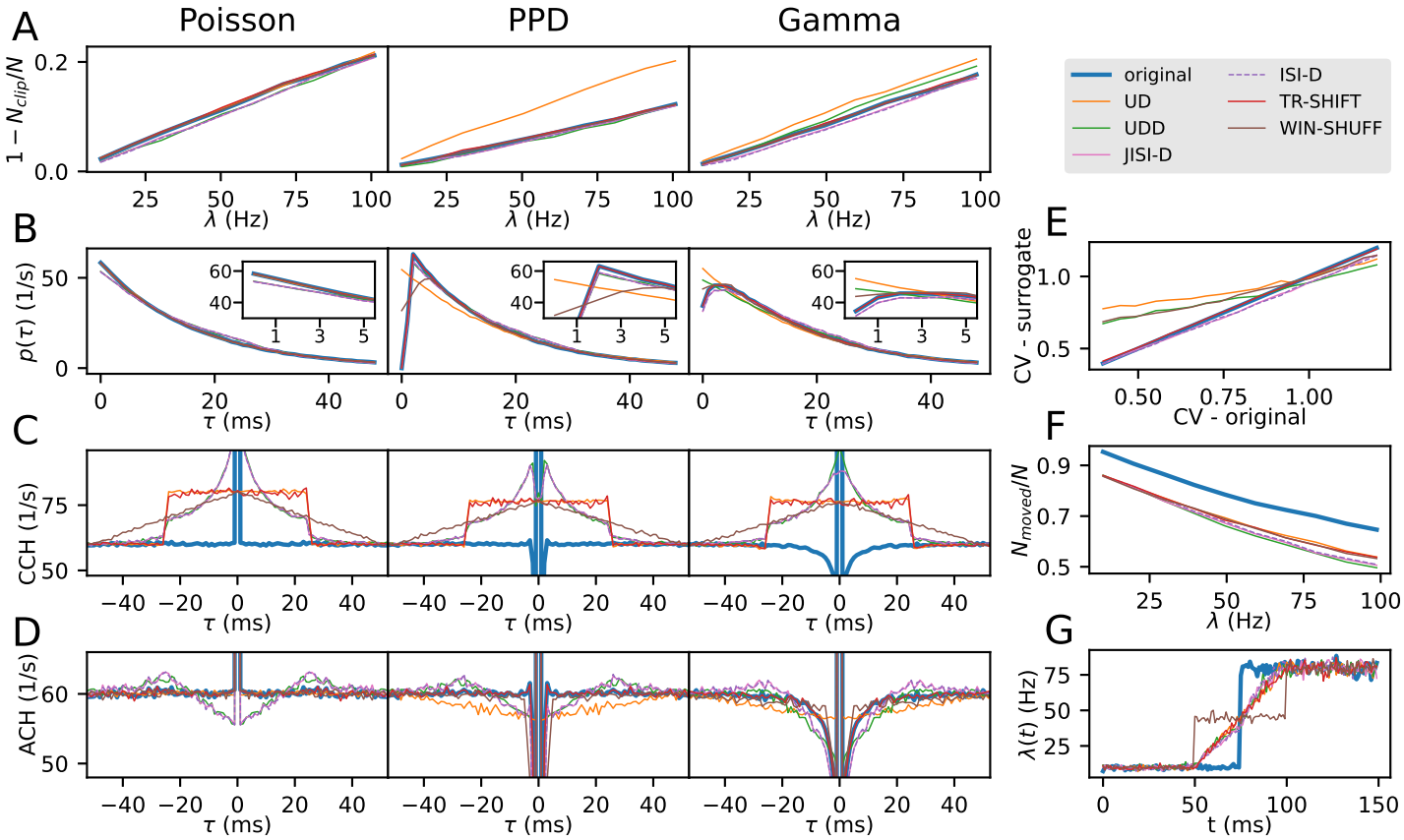


Figure 5. Overview of surrogate statistics. (A). The panel shows the spike count reduction of the artificially generated spike train data in blue (for Poisson, left; PPD $d = 1.6$ ms, middle; Gamma, right, $\gamma = 1.23$ - corresponding to a CV2 of 0.75) after binarization (bin width of 5ms) together with the corresponding surrogates in different colors (UD: orange, UDD: green, JISI-D: pink, ISI-D: violet, TR-SHIFT: red, WIN-SHUFF: brown). The spike count reduction is expressed as 1 minus the ratio of number of spikes in the spike train after clipping (N_{clip}) over the total number of spikes N. The firing rate is constant for each spike train and varies across realizations from 10 to 100Hz in steps of 10Hz (along the x-axis). The spike train durations are fixed such that, given the firing rate, all spike trains have an expected spike count of $10,000$ spikes. (B). The ISI distributions of original and surrogate spike trains are shown as a function of the time lag τ in milliseconds (resolution of 1ms). For each process, the corresponding spike trains have a firing rate of 60Hz and an average spike count of $500,000$ spikes. The ISI region smaller than 5ms, is shown in an inset at the upper right corner. C). The panel shows the cross-correlation between the original spike train (Poisson, PPD, and Gamma, in left, middle and right column, respectively) with each of the surrogates (same color code as in A and B), blue is the correlation with the original spike train with itself (i.e., the auto-correlation) as reference. (D). Auto-correlation histograms are shown before (solid blue) and after surrogate generation (colored lines). For C and D, the x-axis shows the time lag τ between the reference spikes and the surrogate spikes (C) and the other spikes in the spike train (D). For panels C and D, we use the same data as in panel B. In panels E, F, and G, we only examine Gamma spike trains. (E). The panel displays the relation of the original CV (x-axis) against the CV of the surrogates (y-axis). Parameters are the same as in Panels B, C and D (right), but we vary the CV ($CV = 1/\sqrt{\gamma}$) in steps of 0.05, ranging from 0.4 to 1.2 ($\lambda = 60$Hz). (F). The ratio of moved spikes (N_{moved}) over the spike count N is shown. We show it as a function of the firing rate from 10 to 100Hz in steps of 10Hz on the x-axis ($\gamma = 1.23$). (G). In the panel, we show the conservation of the rate profile of a Gamma spike train ($\gamma = 1.23$) and its corresponding surrogates. The firing rate change is a step function, going from 10Hz to 80Hz ($10,000$ realizations, spike train duration of 150ms), and is computed as a PSTH with a bin size of 1ms.

for the original Poisson process (blue), up to 20% for a rate of 100Hz. Only JISI-D, ISI-D, and UDD overall have a slightly lower loss. The ISI distributions (Fig 5B, left; shown for $\lambda = 60$Hz) indicate for all surrogate data an exponential decay, lower for short ISIs for JISI-D, ISI-D, and UDD. Thus, the different surrogates seem to generate again a Poisson-like process, since all surrogates dither the spikes around the original spike positions (Fig 5C, left): uniformly within the dither window for UD or TR-SHIFT, or in a triangular fashion for WIN-SHUFF, high probability around zero and exponential-like decay for JISI-D, ISI-D, and UDD up to the dither window of 25ms. In case of JISI-D, ISI-D and UDD the dithering is limited by construction to the interval between the former and the next spike. Thus, for Poisson data and in particular at high rates, the ISIs are often short, and thus the available dither window may not even be used completely, but only within the time interval between the two limiting spikes. Therefore spikes are on average less moved and stay relatively close to the former position.

The auto-correlations (Fig 5D, left) for UD, TR-SHIFT and WIN-SHUFF are flat besides the center peak, whereas JISI-D, ISI-D, and UDD show a decreased probability for very small ISIs, and then an increase up to the maximum dither width $\Delta = 25$ms with a peak above baseline. The reason for this difference is the limitation that spikes may not be exchanged in their order, as for the other methods. For example, if the reference spike is close to the following and not to the preceding, it will be more likely displaced backwards in time than forwards. This is also slightly visible in a difference of the ISI distributions (Fig 5B, left), compared to the other methods. Moreover, the increase towards Δ above baseline is due to the shift of dither probability to higher time intervals.

PPD process as original data. Next, we discuss the same aspects for the PPD process and its surrogates, which is a process that includes a dead-time (chosen as $d = 1.6$ms as in the experimental data). First of all, we observe that the loss of spikes (Fig 5A, middle) is considerably reduced for the PPD compared to the other models. This is true also for its surrogates, besides UD (orange), that loses more spikes. This may be explained by the ISI distributions (Fig 5B, middle): UD dithers many spikes into the interval of up to 5ms (see inset), corresponding to the bin width. The other surrogates, which preserve the dead-time, follow more closely the original ISI distribution. The distribution of the spike shift from its original position (Fig 5C, middle) is similar to the Poisson process case (same panel - left), but the probability that spikes stay close to their original position is slightly reduced (e.g., pink and dashed violet and green have a lower peak close to the center). For JISI-D, ISI-D, and UDD, the auto-correlations show - as compared to the corresponding surrogates of the Poisson process - also reduced short ISIs, but to a lesser extent. UD moves spikes into small ISIs, but in a smaller scale than the Poisson process (Fig 5A, left and middle, orange), and therefore shows a reduced probability when two spikes have a time difference of less than twice the dither parameter (here 50ms).

Gamma process as original data. A Gamma process does not contain a strict dead-time, but has a preferred ISI defined by the order of the process if $\gamma > 1$ (here $\gamma = 1.23$). The spike loss for the different surrogates (Fig 5A, right) is higher than for PPD, but lower than for Poisson. UD loses the most, JISI-D and ISI-D the least. The ISI distributions (Fig 5B, right) are most changed for UD and UDD, where spikes are shifted into short ISIs, and to a lesser degree for WIN-SHUFF. The other methods maintain the ISI distribution at small ISIs almost identical to the original one. Similarly, are the auto-correlations (Fig 5D, right): TR-SHIFT is identical to the original process; JISI-D and ISI-D are mostly preserving the auto-correlations, but still have a small bump above baseline at around Δ due to the dither restriction not to dither beyond the former and the next spike. WIN-SHUFF has a sharp reduction of spikes after the reference spike, and UD has a slight wide dip around 0. The shift of the surrogate spikes

from the original data (Fig 5C, right) is similar to the case of the Poisson and PPD models.

3.3.2 Effects on other Statistics

Coefficient of Variation of ISIs

We learned that the ISI distributions are affected for most of the surrogates. Fig 5E illustrates how the CV of the surrogates differs in contrast to the original Gamma process (rate fixed to 60Hz), where the CV ranges from 0.4 to 1.2 in steps of 0.05. Non-preservation of the CV in the surrogate data as compared to the original data can be a potential source of false positives, in particular for very small CVs or CVs > 1 [27]. To facilitate the comparison, we also show the diagonal (blue). UD - as expected - changes the CV the most, from original 0.4 to 0.75, i.e., losing strongly its high regularity, and increases even more, with a low slope, to a maximum slightly over 1.0 for the original CV of 1.25, so here burstiness is reduced. WIN-SHUFF and UDD behave similarly to UD, but start at a lower CV for CV $= 0.4$ of the original data; moreover, UDD stays below UD for all CVs. WIN-SHUFF has a slightly higher slope and reaches a maximum still below the original Gamma process.

JISI-D, ISI-D and TR-SHIFT start with identical CVs than Gamma, and TR-SHIFT keeps the same CV as the Gamma CV for all CVs. However, JISI-D and ISI-D have a lower slope as the Gamma process, but still reach high values about 0.05 less than the highest CV at 1.25.

In summary, although the ISI distributions in Panel B, right, seem not to be strongly affected, the effect on the CVs can be very strong. For UD, UDD, and WIN-SHUFF, the CV slightly changes (in both directions), and for JISI-D and ISI-D, the CV decreases. A strong change in the CV of the surrogates, can lead to false positives. Only for TR-SHIFT, the CV is unchanged.

Ratio of moved spikes Next, we study if spikes are indeed moved from their original bin in the surrogate bin. The reason of this interest is that if spikes are not sufficiently moved, correlations are not destroyed as intended, and thus may lead to false negatives. Therefore, we measure the ratio of the number of spikes that are displaced from their original bin position relative to the total number of spikes. We generate original data and its surrogate data and vary the firing rate (from 10 to 100Hz in steps of 10Hz, Fig 5E). If two spikes exchange their bin positions they are both considered as not moved. As a reference the spike ratio is also shown for two independent realizations of a Gamma spike train ($\gamma = 1.23$, blue line).

With increasing firing rates, the ratio of moved spikes decreases for the surrogates. Ideally, the surrogates should be similar to the effect attained on the original process, i.e., the colored lines in Fig 5E, should be as close as possible to the blue line. However, none of the surrogate techniques meets this ideal setting, and the difference to the independent case is almost constant around 10%. Nonetheless, we observe for all surrogates that the ratio of moved spikes decreases with increasing firing rates, which corresponds to the fact that increasingly more bins are already occupied and thus the resulting binned surrogate spike train is more similar to the binned original. UDD, ISI-D, and JISI-D displace less spikes, in particular for higher firing rates as compared to UD, TR-SHIFT, and WIN-SHUFF. The fewer the spikes that are not effectively displaced, the higher the peak at zero-delay of the cross-correlation of the original and the surrogate data (Fig 5C). Almost 50% of JISI-D, ISI-D, and UDD are not moved at 100Hz, and, for lower rates, they remain below the ratios of WIN-SHUFF, UD, and TR-SHIFT. As a consequence, we can expect that JISI-D, ISI-D, and UDD in general tend to yield more false negatives than WIN-SHUFF, UD, and TR-SHIFT.

Rate change in surrogates Changes in the firing rate profile of the surrogates compared to the original data may be a source for false positives [4]. An optimal

surrogate method should follow as closely as possible the original firing rate profile. Therefore, we test here an extreme case where the original data have a rate step (as in [35]), jumping from 10Hz to 80Hz (Fig 4G). We observe that for all surrogates but WIN-SHUFF the firing rate step is converted into a linear increase, which starts at 25ms (dither parameter Δ) before the step and ends at 25ms after the rate step. This behavior has already been derived analytically and observed in [35] for UD: it corresponds to the convolution of the original firing rate profile with the dither (boxcar) function. WIN-SHUFF introduces a second step in the firing rate profile, which is due to the fact that it generates a locally-stationary firing rate within the shuffling window (here 50ms). We conclude that all surrogate techniques smooth the original firing rate profile, whereas WIN-SHUFF creates an additional intermediate rate step. Thus, for steep increases in the firing rate profiles, we have to expect the occurrence of false-positive patterns due to this smoothing.

3.3.3 Summary of the effects of surrogates on the spike-train statistics

We explored different aspects of the statistics of the surrogate spike data as compared to its original process. In general, surrogate data are not identical to the original data, but change to a different degree. The effects for the three data models are summarized and not differentiated, since they anyway are similar. These are listed in Table 1. Features that are preserved are indicated by 'yes', approximately preserved ('approx.'), and not preserved ('no'). Given the insights obtained through the statistics considered, we expect that surrogates generated with UD, as they evidence a strong spike count reduction, lead to a large number of FPs. UDD surrogates might lead to FPs in case of regular data that do not exhibit a dead-time (e.g., Gamma spike trains). The study of the similarity of the surrogates to the original processes shows that JISI-D, ISI-D, and UDD might lead to fewer patterns detected, i.e., an underestimation of significance. Moreover, we expect FPs for WIN-SHUFF surrogates when the original firing rate profiles have a steep rate increase. Finally, the technique preserving the most statistics without showing any disadvantages is TR-SHIFT.

Feature/Method	UD	UDD	ISI-D	JISI-D	TR-SHIFT	WIN-SHUFF
Spike Count	no	approx.	approx.	approx.	yes	yes
ISI	no	no	approx.	approx.	yes	approx.
dead-time	no	yes	yes	yes	yes	no
Auto-correlation	no	no	no	no	yes	approx.
Firing rate modulation	approx.	approx.	approx.	approx.	approx.	approx.
Spike train regularity (CV $<$ 1; regular)	no	no	approx.	approx.	yes	no
Spike train regularity (CV $>$ 1; bursty)	no	no	approx.	approx.	yes	approx.

Table 1. Table summarizing the statistical properties of the six discussed surrogate techniques conserve (yes/ approx./ no). The dead-time conservation is evaluated based only on results of PPD process, otherwise on the results for all data models (Poisson, PPD, Gamma).

3.4 Performance of the surrogate methods

Next, we apply the six surrogate techniques to artificial data sets that are generated based on two experimental data sets [34] to study the effect of the various surrogate methods on the occurrence of false positives [35]. The experimental data are two sets of recordings from approximately 100 parallel spike trains from macaque monkey motor and premotor cortex during performance of a reach-to-grasp behavior [40]. We model the data with the same firing rate profiles as the experimental data, and use as point process models a) the PPD to mimic the dead-time of the data (due to spike sorting), and also b) Gamma processes to account for their CVs. Note that the modeled spike trains exhibit non-stationary firing rate profiles. The spike trains are generated independently and hence all observed spike patterns occur by-chance and can be considered as false positives (FPs).

3.4.1 Description of experimental data and preprocessing

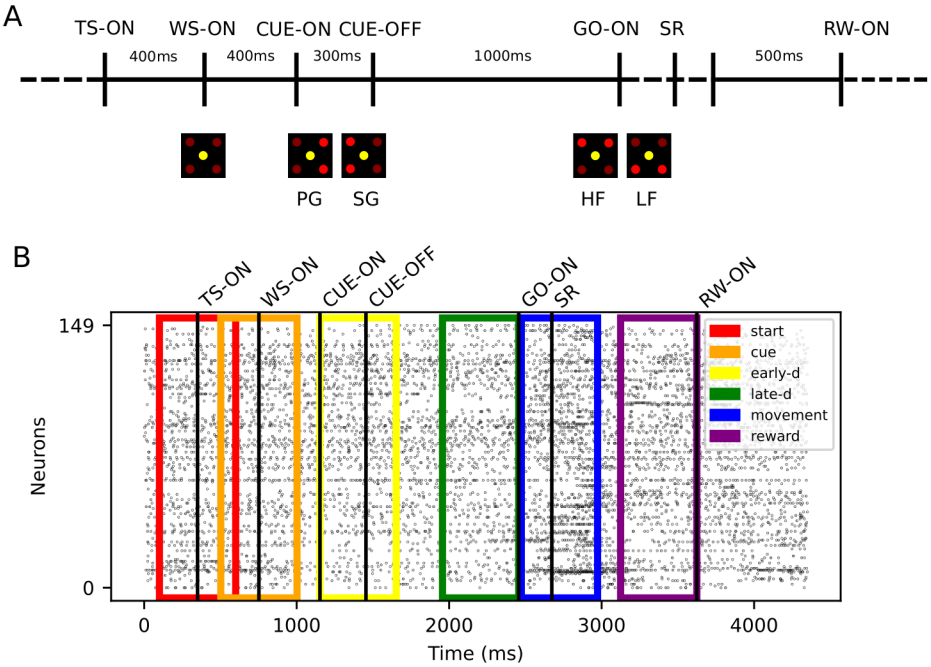


Figure 6. Experimental protocol and data preprocessing. (A). The trial start (TS) is self-initiated by the monkey. A waiting signal (WS-ON) prepares the monkey for the visual cue presented at CUE-ON, providing the grip type instruction (PG/SG). After 1000ms, the monkey is presented a second visual cue (GO-ON), specifying the force needed to pull the object (HF or LF) and the GO signal. The switch release (SR) marks the beginning of the movement. The monkey touches the object and maintains the grip for 500ms until the reward (RW-ON). The timing of the behavioral events SR and RW are variable and depend reaction time and movement speed. (B). The panel shows the simultaneous spiking activity of all neurons (y-axis) over time (x-axis) for one example trial (first successful trial of session *i140703-001*) of trial type PGHF. Each dot indicates one spike. The trials are aligned to trial start (TS-ON). The six colors rectangles represent the position of the six trial epochs (see legend).

Experiment and data The experimental data were recorded during a reaching and grasping task from the pre-/motor cortex of two macaque monkeys. Both monkeys were chronically implanted with a 100-electrode Utah array (Blackrock Microsystems). The experimental protocol is schematized in Fig 6A and was also published in [34] and

in [40]: monkeys N and L were trained to self-initiate a trial by pressing a start button (registered as trial start, TS). Then, after a fixed time of 400ms, a visual signal (yellow LED) was shown, to attract the attention of the monkey (waiting signal, WS). After another 400ms-long waiting period, a first visual cue (2 LEDs on) was presented to the monkey for a period of 300ms (from CUE-ON to CUE-OFF) indicating the grip type: full-hand side grip (SG) or two-fingers precision grip (PG). Followed by another waiting period of 1000ms, the GO-signal was presented, containing also the information of the expected grip force (high, HF, or low, LF, by LEDs). The behavioral conditions were selected in a randomized fashion. The start of the monkey movement is recorded as the release of a switch (SR). Subsequently, the object touch (OT) and the beginning of the holding period (HS) are indicated. After 500ms of holding the object in place , the reward (RW) was given to the monkey and the trial finished.

The experimental data sets consist of two sessions (*i140703-001* and *l101210-001*) of 15 min of electrophysiological recordings containing around 35 trials per trial type, i.e., combinations of grip and force type. Each session is spike sorted using the Plexon Offline Spike Sorter (version 3.3). The spike data were extracted and are available on `https://gin.g-node.org/INT/multielectrode_grasp`.

Preprocessing for analysis We only consider neurons satisfying the following constraints: SNR < 2.5 (signal-to-noise ratio of spike shapes), average firing rate across trials < 70Hz. Hypersynchronous (artifact) spikes across electrodes arising at the sampling resolution are detected automatically, classified as artifacts, and finally removed as in [21]. Only successful trials are retained. Both experimental sessions are analyzed separately, further differentiating between the different trial types. For the analysis, trials are segmented into six 500ms-long separate epochs, to account for the behaviorally relevant events (as in [21], and represented in Fig 6B). Segments of the same epochs in the same trial type are concatenated and yield 24 (4 trial types $\times 6$ epochs) data sets per session.

3.4.2 Simulation of experimental data and SPADE analysis

We create the artificial data by generating as many spike trains as in the concatenated data sets, and use the original firing rate profiles of the individual neurons, estimated with an optimized kernel density estimation as designed in [50, 51] on a single trial-by-trial basis. To account for dead-time and regularity of the data, we model separately the spike trains with a PPD and a Gamma process. For the PPD data, we estimate the dead-time for each neuron and each combination of epoch and trial type by taking their minimum ISI (fourth inset, panel A of Fig 7). For the data modeled as a Gamma process, we instead fix the shape factor for each neuron and each combination of epoch and trial type by estimating the CV of the process in operational time [48] and then transform the CV into the shape factor ($\gamma = \frac{1}{\mathrm{CV}^2}$, [52]). The process is generated in operational time and then transformed back into real time. The resulting CV2 distribution of all neurons of the data, simulated as a Gamma process, is close to the one of the experimental data (third inset, panel A of Fig 7). Note that the Gamma process does not have an absolute dead-time but for $\gamma < 1$, it has a low probability for small ISIs and can be regarded as containing a relative dead-time [48]. The resulting firing rate of the artificial data is, for both processes, close to the one of the original spike train (first inset, panel A of Fig 7).

For the two experimental sessions, each of the 24 sets of concatenated data is modeled using the two point process models, resulting in a total of $2 \times 24 \times 2 = 96$ data sets. The SPADE analysis is performed on all data sets, separately for each of the six surrogate techniques. We set the bin size to 5 ms, the maximum pattern duration to 60ms, the significance level to $\alpha = 0.05$, and use 5000 surrogates. For all surrogate

techniques, we set the dither parameter to $\Delta = 25$ms. Each resulting pattern is counted as a false positive.

3.4.3 False positive analysis

Fig 7, panel B shows the false-positive results for the PPD (left) and the Gamma modeled data sets (right). For each data model, we show the number of FPs per data set and the total number of FPs (in text) across the six surrogate methods. Results show that we retrieve for all surrogate techniques, except for uniform dithering, a small number of FPs. For UD, we get for the PPD process in Fig 7B, left: UD (522; 88.1%), UDD (13; 2.2%), JISI-D (14; 2.3%), ISI-D (14; 2.3%), TR-Shift (15; 2.5%), and WIN-SHUFF (14; 2.4%). Contrarily, we see in Fig 7B (right) that the analysis of the Gamma data leads to the following number of FPs: UD (302; 77.6%), UDD (52; 13.4%), JISI-D (8; 2%), ISI-D (8; 2%), TR-Shift (9; 2.3%), and WIN-SHUFF (10; 2.6%). Considering all results, we conclude that UD leads to a very high number of false positives compared to the other surrogate methods, as we expected from our results of the previous sections. We observe a relatively high number of FPs for the application of UDD on Gamma data, expected from our observations in section 3.3.1. The remaining surrogate techniques exhibit a similar number of FPs.For further analyses, given these observations, we form four groups of FPs (and, thus, of neurons), depending on which surrogate techniques they are expressed in. The first and predominant group are the FPs present only in the SPADE analysis performed with UD surrogates. Second, we group together FPs present in all surrogate techniques. Third, in the case of Gamma data, we distinguish a subset of FPs found with both UD and UDD surrogates. Finally, we pool all FPs present in any other combination of surrogate analysis.

To get an understanding of the rate properties of neurons that contribute to the FPs, we consider their average firing rates (over time and trials) and the group that they belong to (Fig 7C and D). In general, we find FPs in all analyzed data sets, but one (monkey L, movement, Gamma, all conditions). We observe that almost all neurons involved in FPs have an average firing rate higher than 20Hz. Neurons belonging to the first group are the largest set, and are present for both monkeys and data models, and in almost all data sets. The second group is present for both monkeys and models, but is more represented for PPD. The third group is present for both monkeys only for the Gamma model. This was already expected, given the higher spike count reduction, for UD and UDD in the case of Gamma spike trains (section 3.3.1).

We also inspect the CV2, averaged over trials, of units involved in FPs (S1 Fig). FPs occur in neurons with a relatively low CV2 (CV2 < 1), but, importantly, this is not the case for neurons with very low CV2s (especially for monkey N). In fact, given the results of [27], the most regular spike trains are expected to be involved in false-positive patterns, which is not the case here. In almost all cases, bursty neurons (CV2 > 1) are not involved in FPs.

In summary, we observe that the surrogate technique leading to most FPs is UD, followed by UDD (only in the case of Gamma data). Neurons exhibiting an average firing rate higher than 20Hz, and having a CV2 < 1 are predominantly involved in FPs. Moreover, there is a small amount of FPs detected using all surrogate techniques, which is expected given a certain significance threshold. Nonetheless, regular and high-rate neurons are more prone to raise the false-positive rate [53].

3.5 Application to experimental data

As a last step, we apply SPADE with the six surrogate techniques to the two sessions of experimental data introduced in section 3.4.1. Here our goal is to analyze with SPADE experimental data, for which we do do not know the ground truth (i.e., the presence

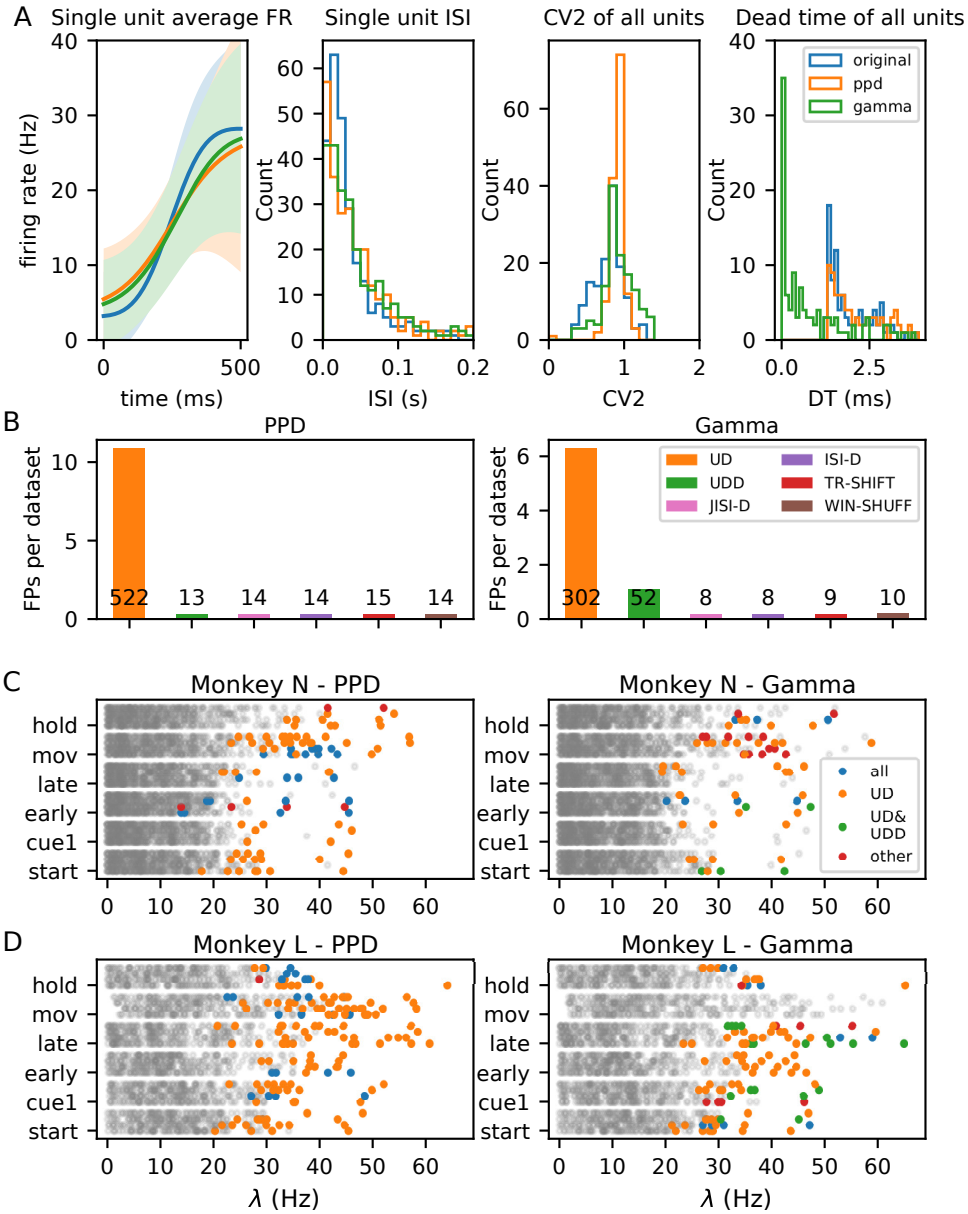


Figure 7. Evaluation and analysis of false positives across surrogate techniques for pattern detection with SPADE. (A). Comparison of statistics of the original to the generated artificial data during the movement epoch (PGHF trial type). In blue, orange and green we represent original, PPD and Gamma data spike trains respectively. Left graph: average firing rate of a single unit across one epoch of 500ms; second from left: ISI distribution of a single unit; third from left: average CV2 estimated trial wise for all neurons; fourth from left: dead-time as minimum ISI for all neurons. (B). number of false positives (FPs) detected across surrogate techniques (color coded) normalized over the 48 (2 sessions $\times$ 6 epochs $\times$ 4 trial types) data sets analyzed, left for PPD and right for Gamma process data analyses. Numbers in text represent the total number of FPs over all data sets per surrogate technique. (C) and (D). Average firing rate of neurons for each monkey (N in panel C and L in panel D), epoch (y axis) and behavioral type (for each epoch ordered as PGHF, PGLF, SGHF, SGLF). Left PPD data, and right Gamma data. Colored dots represent individual units involved in FPs: blue dots indicate average firing rate of units involved in FP patterns found for all surrogate techniques, orange dots for UD surrogates, green dots for UD and UDD only, and red dots for other combinations of different surrogate techniques. Grey dots represent the average firing rate of individual units not involved in any FP.

and amount of significant patterns) and show the differences resulting from the application of the surrogate techniques in the significance testing. In Fig 8, we present the found number of significant patterns for each epoch and trial type (different colors). The results are shown for each monkey separately, since their data differ in terms of CV2, dead-time and firing rates (Fig 7 and S1 Fig). The exact number of patterns found is: UD (N:203, L:121), UDD (N:14, L:14), JISI-D (N:10, L:10), ISI-D (N:10, L:10), TR-SHIFT (N:7, L:14), WIN-SHUFF (N:11, L:11). Crucially, we detect more patterns (almost double the amount) in the analysis of experimental data than in the independently generated artificial data, except for UD and UDD (on Gamma data) surrogates.

A first observation is that the amount of significant patterns found using UD is much higher (note different y-axis scale!) for both monkeys as compared to the other five surrogate techniques. Patterns occur mostly during the movement (mov) epoch where the firing rates are highest. Thus, given the calibration results from the former section, a large amount of those is likely to be FPs.

Taking from now on into consideration all surrogate techniques but UD, for monkey N (Fig 8, left column) we find patterns across all epochs, almost for all surrogates. The pattern numbers are relatively similar within and across epochs. Very specific is the pattern occurrence during the movement epoch (Fig 8, left column, pink): in fact, the same pattern occurs for SGLF behavioral context in all surrogates, but TR-SHIFT. During the start epoch, all surrogates show patterns in relation to SGLF, but some (UDD and TR-SHIFT) also in relation to SGHF, and others (JISI-D, ISI-D and WIN-SHUFF) in relation to PGHF. In cue epoch, all surrogates find patterns in relation to PGLF trials (Fig 8, left column, light blue). During early delay (earl-d) all surrogate techniques find patterns for PGHF trials, in addition, for UDD, a pattern for SGHF, and one in PGLF trials for WIN-SHUFF. During the late waiting epoch (late-d) patterns occur only in PGHF and PGLF trials (Fig 8, left column, blue and light blue), but we also observe patterns in SGHF trials for UDD (green, second row). Finally, during hold, we find patterns in PGLF trials for all surrogate techniques. In addition, we find a pattern in SGHF trials for UDD and a pattern in PGHF trials for JISI-D, ISI-D and WIN-SHUFF.

For monkey L (Fig 8, right column) most patterns occur during the movement epoch for PGHF, PGLF, and SGHF, however in slightly different combinations. During the late phase of the waiting period (late-d) four out of the five surrogates find the same patterns (1 for SGLF and 1 for SGHF). During the hold epoch only for UDD and TR-SHIFT we find the same patterns, 1 for PGHF and 1 for PGLF. We do not detect any significant patterns in start, cue, and early delay epochs.

Interestingly, for each monkey, the significant patterns are specific to the epochs, i.e., identical patterns do not repeat in different epochs, but the patterns are different in the temporal delays of the spikes and are mostly composed of different neurons. Previous studies on this experiment ([41], Figure 2) showed that monkey L has on average a shorter reaction time than monkey N, and has shorter and more pronounced rate increases within the movement epoch, which may explain explain the high amount of patterns during that epoch. In contrast, monkey N shows patterns already in the start epoch, and the number of detected patterns remains almost constant throughout epochs.

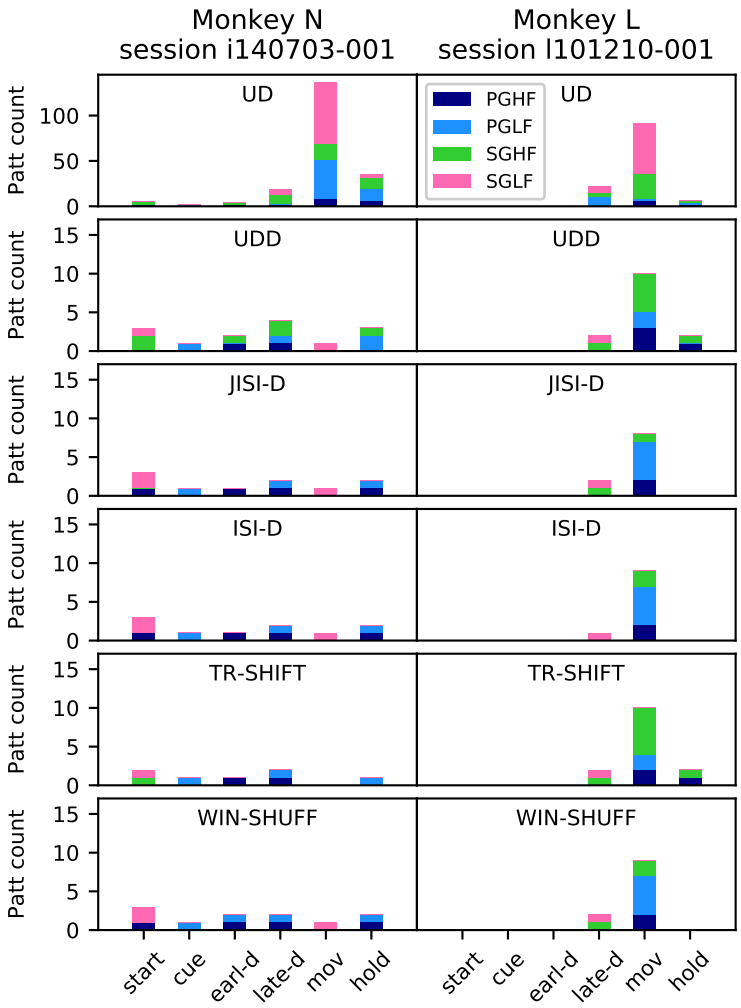


Figure 8. Analysis of experimental data. SPADE results for two sessions of experimental data (data described in3.4.1): session *i140703-001* (left) and session *l101210-001* (right). Histograms represent the number of significant patterns detected by SPADE in each epoch (start, cue, early-delay, late-delay, movement and hold), color coded according to the grip type (precision/side grip -PG/SG- and low/high force -LF/HF). Each row corresponds to one surrogate technique (note the different y-axis scale for UD).

4 Discussion

Here we have shown that the usage of uniform dithering (UD) as a surrogate technique, followed by binarization (binning and clipping) of the spike train, leads to a mismatch in the spike counts between the original and the surrogate data (Fig 2). The spike count reduction is significant and increases monotonically with the firing rate, which we verified analytically and through simulations (Fig 3B, Fig 5A). Moreover, we showed that two factors play a large role in the spike count reduction: the neuronal dead-time and the firing regularity (Fig 3B). These two components are contained in the auto-structure of the experimental data from pre-/motor cortex of macaque monkey we aimed to analyze here [34, 41].

In the context of spatio-temporal spike pattern detection using SPADE, we observed that the spike reduction in the UD surrogates in combination with binarization of the spike data leads to an underestimation of patterns in the surrogate data compared to the original data, which in turn leads to an overestimation of pattern significance in the original data (Fig 3C). The ultimate consequence of this problem is the occurrence of false positives. Fortunately, SPADE is a modular method: different types of surrogates can be used, while the mining algorithm and testing steps stay identical. For this reason, we were able to analyze the same data sets by using different surrogates, to then compare the results.

In order to verify and solve the issue of the spike count mismatch, we compared five alternative surrogate techniques to UD: uniform dithering with dead-time (UDD), window shuffling (WIN-SHUFF) -both newly introduced-, joint inter-spike interval dithering (JISI-D; [32]), inter-spike interval dithering (modified from [32]), and trial shifting [36] (Fig 4). All surrogate approaches have the goal of destroying the exact spike timing relations across the neurons. We quantified which statistical features of the original spike trains are conserved in a surrogate, examining the spike counts, the ISI distribution, the auto-correlation, the cross-correlation, the firing rate modulations, the ratio of moved spikes, and the regularity (Table 1). These were evaluated on artificial data (Poisson, PPD, and Gamma spike trains) that contained partly the main features of the data (dead-time and regularity) (Fig 5). We observed that UD does not preserve the spike count, and leads to a very strong spike count reduction for the PPD model, not for the Poisson model, and not (or very small) for the Gamma model. As experimental data typically exhibit a dead-time or it is introduced by spike sorting, we concluded that UD is not an adequate method to estimate correlations within our context, especially as the effect of clipping increases with the firing rate.

Further we tested the surrogate alternatives in respect to false positives on ground truth data, that included in addition realistic firing rate modulations as observed in the experimental data. These were based on two experimental sessions of motor cortex of macaque monkey involved in a reaching and grasping task [34, 40]. We stress the importance of generating test data that are very similar to experimental data, in order to closely model all features that typically lead to complications when trying to robustly estimate the null-hypothesis of independence [4]. Such realistic data serve as ground truth to identify the strengths and weaknesses of the tested surrogate techniques. In this case, we modeled both experimental sessions with PPD and with Gamma processes, with firing rate modulations, dead-times, and regularities estimated for each neuron from the original data (Fig 7A). The analysis of these data with SPADE led to a large number of FPs when employing UD. However, all other surrogate techniques showed a considerably low number of FPs. A minimal number of false positives is to be expected, as it is inherent to any statistical test. Due to this result, we conclude that UD is not appropriate for its employment within the context of the SPADE analysis, whereas all other techniques can be considered valid.

Finally, having validated the five alternative surrogate techniques against UD, we analyzed experimental data from [34]. UD in this context leads to a large number of detected patterns (Fig 8) in contrast to the other surrogates. Still, the number of patterns detected by the other surrogate methods is larger than for the analysis of artificial data, that were independently generated, but otherwise very similar. Given the results obtained from the previous sections, analytically and through simulations, we consider the patterns detected by UD as putative false positives, taking into consideration that in the case of experimental data we have no ground truth at hand. Moreover, we consider the patterns detected in the experimental data by the other surrogate methods as significant, i.e., not resulting from any overestimation of the significance, since we also found that the patterns retrieved for UDD, JISI-D, ISI-D,

WIN-SHUFF, and TR-SHIFT are highly overlapping, i.e., they show the almost identical participating neurons, lags, and occurrence numbers. Hence, we conclude that the different surrogates, even though they move the spikes in different ways, lead to a similar number of significant patterns. In other words, the five surrogate techniques lead to an almost identical significance level.

From that we can conclude that all five surrogate techniques presented as alternatives to UD are appropriate for the detection of spatio-temporal spike patterns in the SPADE analysis. Nonetheless, we suggest TR-SHIFT as the surrogate of choice for the SPADE analysis, because it is a technique that 1) is easy to explain and to implement, 2) reflects more closely the hypothesis of temporal coding, 3) reproduces exactly the most relevant statistical features of a spike train (Table 1 and Fig 5) is as conservative as the other methods that we propose, and, 5) employs fewer parameters than the other techniques with the same performance.

We stress that the surrogate techniques do not have to be restricted to the particular context of the SPADE method, and are actually used in other studies for the evaluation of correlation [4, 29, 31, 33, 35, 45, 53–60]. Still the particular surrogate needs to be chosen appropriately and cautiously case by case. Not only because the statistical test might produce false positives (or false negatives), but also because the concrete null-hypothesis distribution represents the model assumed for brain functioning. The degree of how conservative or liberal the statistical analysis can be, through the choice of the surrogate technique, becomes then not only a feature of the test but more a scientific question per se regarding neural coding.

Several studies have already investigated the impact of different surrogate techniques in the context of spike time correlations. For example, in [1], the authors evaluated the influence of surrogate techniques on cross-correlation analysis of two parallel spike trains; in [4, 35], the focus was on the effect of surrogate techniques on synchronous events in the context of the Unitary Events analysis [3, 27, 55, 56, 58, 61]. Due to the results of these studies, we have concentrated on surrogate techniques that preserve the firing rate profile of the original neurons, but methods such as spike train randomization (within single trials, [26]), spike exchange (across neurons or trials, [59, 62]), ISI shuffling (within and across trials, [63–66]), spike shuffling across neurons (within-trial, [63, 65]) do not fulfill our requirements [4]. Other methods are designed to preserve the auto-correlation of a spike train, with the assumption of stationarity and Markovianity of a process [67, 68]. Some studies also already evidenced problems arising from the application of uniform dithering, such as the non-preservation of the ISI distribution [35], in particular in the case of the Poisson process [69], but not in the context of multiple parallel spike trains, or in the context of binarization. We extended here studies of comparisons of surrogate techniques to the context of spatio-temporal patterns, i.e., including delays between spikes of multiple neurons. A similar comparison of statistical methods have been done only at the level of pairwise correlations [35]. The relevance of this further step is sensible, as it has been argued that the processing of information may be reflected in the presence of delayed higher-order correlations in parallel spike trains [70–74].

Other methods discuss the employment of surrogates as a liability more than an asset [75, 76], pointing in particular to the computational burden of the approach, and motivating closed-form testing (assuming a point process model for the spike train) as a lighter alternative. Nonetheless, the estimation of an analytical form of the null-hypothesis often involves assumptions or approximations that can also provoke the detection of false positives [4, 20, 27, 77]. In addition, treating non-stationary processes analytically is cumbersome and mostly there is no possibility of an analytical treatment, for example in the case with time-varying firing rates. Moreover, regarding the computational costs, several improvements have been made for the performance of both

SPADE and the surrogate implementations (`https://elephant.readthedocs.io/en/latest/release_notes.html`). Nonetheless, depending on the size of the data set and the number of surrogates employed, large analyses can still take up to several hours on a computer cluster.

The implementation of all surrogate techniques and of SPADE is included in the Elephant python package `http://python-elephant.org`. For implementations of other surrogate techniques available in the neuroscience literature in other programming languages, together with the explanation of their algorithms, we refer to SPIKY (matlab, [78]), TISEAN (fortran, [79]), and SpiSeMe (C++, matlab and python, [68]).

Our study concentrated on the problems arising from the application of UD together with binarization, and false-positive patterns resulting as a consequence. Another aspect to study would be to concentrate on false negatives (FNs), i.e., true positive patterns present in data and not detected by SPADE. Some steps have been already done in this direction [49], where results showed that the employment of UDD and JISI-D do not result in a large number of FNs. Having showed that the proposed surrogate techniques lead to almost the same significance level, we expect that the results obtained for UDD and JISI-D can extend to ISI-D, WIN-SHUFF, and TR-SHIFT. In addition, artificial data with inserted patterns testing the specificity of SPADE had already been studied in [17–19], albeit only for the Poisson data model with stationary firing rates.

A number of sessions of the experiment considered in this paper have already been analyzed in [21]. The authors found many patterns with SPADE when evaluating only for synchronous spikes, using UD as the surrogate technique of choice. Thus, one might question the findings of [21], given the results of this article. However, in [21] the authors employed a version of the FIM algorithm called CoCoNAD [23–25] that mined patterns in continuous time, without resorting to clipping. As we see in S2 Fig, we find more synchronous spike events using CoCoNAD as compared to using FIM with all surrogate methods even with UD (although there is a small reduction of detected patterns for UD). Unfortunately, we were not (yet) able to transform CoCoNAD such that it is also applicable for the detection of spatio-temporal patterns, but had to stick to time discretized data.

Currently, we are applying SPADE to a large set of experimental data to investigate the presence of spatio-temporal patterns in relation to behavior. We aim to study their statistics and features, to test for their behavioral relevance.

Software and Data resources

The code to perform and reproduce the analyses presented in this study can be found at (`https://github.com/INM-6/SPADE_surrogates`), along with the code to reproduce Fig 2, Fig 3, Fig 5, Fig 6B, Fig 7, Fig 8 and Supplementary Figures 1 and 2. Figures Fig 1, Fig 4, Fig 6A are sketches created manually with vector graphics editors. The experimental data analyzed in section 3.5 can be found at (`https://gin.g-node.org/INT/multielectrode_grasp`). The artificial data is generated from the experimental data within the *SPADE_surrogates* repository. The SPADE method and all implementations of the surrogate techniques are included in the Elephant Python package `http://python-elephant.org`.

5 Supporting Information

5.1 Supporting Information 1: Derivation of spike count after binarization

The spike count after binarization (N_{clip}) for renewal processes can be obtained directly from the inter-spike interval (ISI) distribution $\rho(\tau), \tau > 0$. The probability of a single spike to be discarded depends on the interval to the previous one as: $P_{\text{clip}}(\tau) = 1 - \frac{\tau}{b}$ for $\tau < b$, where b is the bin size [49]. Thus, the probability of a single spike to be clipped away decreases linearly with increasing interval to the previous spike. In the extreme case of $\tau = 0$ (if an ISI is equal to zero), the next spike is discarded, and if $\tau = b$, the next spike is not discarded.

Averaging over the ISI distribution, the firing rate after binarization λ_{clip} becomes

$$\frac{\lambda_{\text{clip}}}{\lambda} = \langle P_{\text{clip}}(\tau) \rangle = \int_0^b d\tau \, \left(1 - \frac{\tau}{b}\right) \, \rho(\tau).$$

For a PPD process with dead-time d, the ISI distribution is

$$\rho^d(\tau) = \tilde{\lambda} e^{-\tilde{\lambda}(\tau - d)} \Theta(\tau - d)$$

with the effective rate $\tilde{\lambda} = \frac{\lambda}{1-\lambda d}$ [47, 49], and $\Theta(x)$ being the Heaviside function. Thus, the firing rate after binarization (λ_{clip}) follows as

$$\frac{\lambda^d_{\text{clip}}}{\lambda} = 1 + \left(-\frac{1}{\lambda b} + \frac{1}{\tilde{\lambda} b} e^{-\tilde{\lambda}(b-d)} \right) \Theta(b - d).$$

For a Gamma process of shape γ, we have

$$\rho^\gamma(\tau) = \gamma\lambda \frac{(\gamma\lambda\tau)^{\gamma-1}}{\Gamma(\gamma)} e^{-\gamma\lambda\tau} \Theta(\tau).$$

Thus, the corresponding firing rate after binarization results to be:

$$\frac{\lambda^\gamma_{\text{clip}}}{\lambda} = \frac{1}{\Gamma(\gamma)} (\bar{\Gamma}(\gamma, \gamma\lambda b) - \frac{1}{\gamma\lambda b} \bar{\Gamma}(\gamma + 1, \gamma\lambda b))$$

with $\bar{\Gamma}(\gamma, x)$ being the upper incomplete Gamma function. Further, we denote the lower incomplete function as $\underline{\Gamma}(\gamma, x)$ [80].

These clipped firing rates are shown in Fig 3. There, we also show the clipped firing rates after the application of uniform dithering. These are only analytically feasible in the range of some approximations. In detail, readers can find them in [49]. The results are:

$$\frac{\lambda^{\text{UD}}_{\text{clip}}}{\lambda} = 1 - \frac{1}{2}\lambda_{\text{UD}} b + \frac{1}{6}(\lambda_{\text{UD}} b)^2 + \mathcal{O}((\lambda b)^3),$$

with λ_{UD} depending on the process model and the firing rate. In our case, these are:

$$\lambda^d_{\text{UD}} = \frac{1}{2\Delta^2} \sum_{n=1}^{\lfloor \frac{2\Delta}{d} \rfloor} \left((2\Delta - nd) \frac{\underline{\Gamma}(n, \tilde{\lambda}(2\Delta - nd))}{\Gamma(n)} - \frac{n}{\tilde{\lambda}} \frac{\underline{\Gamma}(n+1, \tilde{\lambda}(2\Delta - nd))}{\Gamma(n+1)} \right),$$

$$\lambda^\gamma_{\text{UD}} = \frac{1}{2\Delta^2} \sum_{n=1}^{\infty} \left(2\Delta \frac{\underline{\Gamma}(n\gamma, 2\gamma\lambda\Delta)}{\Gamma(n\gamma)} - \frac{n}{\lambda} \frac{\underline{\Gamma}(n\gamma + 1, 2\gamma\lambda\Delta)}{\Gamma(n\gamma + 1)} \right).$$

5.2 Supporting Information 2: Algorithm of joint-ISI and ISI dithering

Here, we explain the algorithm of the joint-ISI dithering in more detail. We start by neglecting the problems of discretization. Thus, we consider a continuous distribution. Given a spike train's joint-ISI distribution $p(\tau, \tau')$, the displacement of the spike i at time t_i by the amount of time δ will depend on the neighboring spikes $i-1$ and $i+1$, according to

$$u^i(\delta) \propto p((t_i + \delta) - t_{i-1}, t_{i+1} - (t_i + \delta))\Theta(\Delta - |\delta|)$$

with Δ denoting the maximal displacement of each spike and $\Theta(x) = 1$ if $x \geq 0, 0$ otherwise. For a graphical explanation of this refer either to [35] or [49]. In our algorithm, we first displace the even-numbered spikes, and then the odd-numbered ones. An alternative option would be to do it spike-by-spike, as it was done in [35]. Since the joint-ISI distribution is zero for inter-spike intervals smaller than the dead-time, the dead-time is conserved in this procedure.

The maximal displacement Δ of each spike should be sufficiently large such that the variance [49] and the kurtosis [32, 35] of the joint-ISI distribution along the anti-diagonal do not change significantly (an effect resulting from a higher probability to dither a spike into the middle bulk of the joint-ISI p.d.f. due to the limited displacement range). A further point related to the kurtosis is the introduction of a square-root term on the j-ISI histogram [32], which counterbalances the change in kurtosis. In our case, we use $\Delta = 25$ ms instead $\Delta = 8$ ms as used in [32], which makes this correction unnecessary.

In practice, we cannot estimate a continuous joint-ISI distribution from the data but only a discretized one by making a histogram. To avoid a sparse joint-ISI histogram (with multiple entries being zero) a Gaussian smoothing is applied. The joint-ISI histogram is thus written as $p^{\sigma}_{n,n'}$, where σ is the standard deviation of the Gaussian convolution kernel. For the conservation of the dead-time, entries in the domain of the joint-ISI distribution where intervals are smaller than the dead-time are (optionally) set back to zero. The dead-time is estimated for each spike train, bounded to a maximal value, that we fix to 4 ms. For the analyses shown in Fig 5, we do not use a Gaussian smoothing, while for the other applications, we employ a σ of 2 ms (i.e., twice the bin size) which is chosen to be small enough not to change the ISI distribution significantly. In the latter case, we further employ the cutoff in order to preserve the dead-time.

Further, the joint-ISI histogram is truncated on a triangular domain below an anti-diagonal of the joint-ISI distribution, where we choose that the triangular border is at the sum of two neighboring ISIs of 100 ms. Inside the truncated area, the probability to displace spike i by n bins with a bin size b is

$$u^i_n \propto p^{\sigma}_{\lfloor \frac{t_i - t_{i-1}}{b} \rfloor - n, \lfloor \frac{t_{i+1} - t_i}{b} \rfloor + n} \Theta(\lfloor \frac{\Delta}{b} \rfloor - n).$$

Outside this truncation, we assume the distribution on each anti-diagonal to be constant. Thus, we dither spikes uniformly there.

For the ISI dithering, the procedure is the same with the only difference that the joint-ISI distribution $p(\tau, \tau')$ is not calculated from the data but only the ISI distribution $p(\tau)$, or in the binned and smoothed fashion p^{σ}_n. The dithering probability for spike i thus becomes

$$u^i_n \propto p^{\sigma}_{\lfloor \frac{t_i - t_{i-1}}{b} \rfloor - n} \cdot p^{\sigma}_{\lfloor \frac{t_{i+1} - t_i}{b} \rfloor + n} \Theta(\lfloor \frac{\Delta}{b} \rfloor - n).$$

Therefore, in comparison to the joint-ISI dithering, this does not consider the first order of serial ISI correlations. In the case of renewal processes, the joint-ISI

distribution is symmetric with respect to the diagonal, thus coincides to the outer product of the ISI distribution with itself [49]. As we analyze in Fig 5, the Poisson Process, PPD, and Gamma are all renewal processes, thus, the results coincide for joint-ISI and ISI dithering.

5.3 Supplementary Figure 1

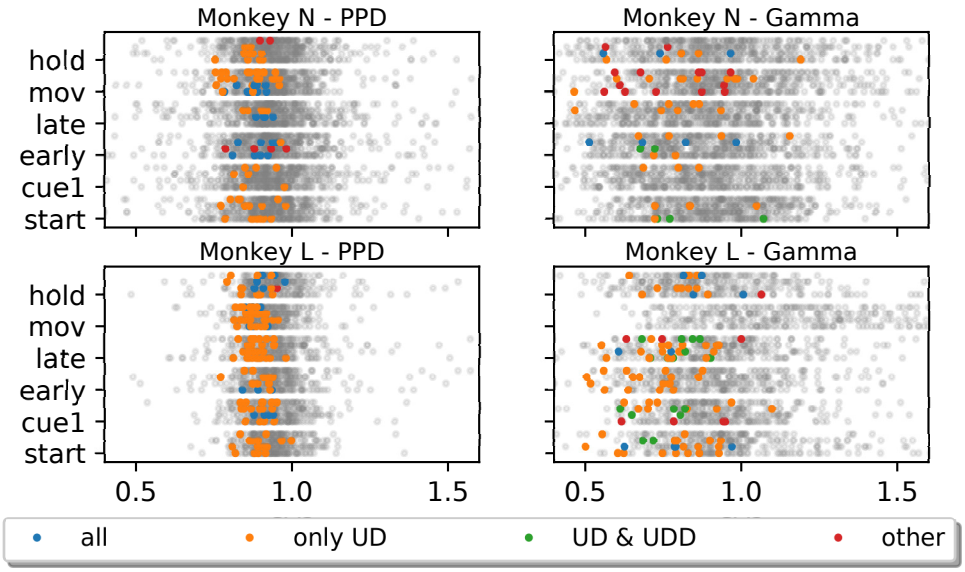


Figure 9. Average CV2 for neurons participating in FP patterns against all neurons. Average CV2 of neurons for each monkey (N in panel C and L in panel D), epoch (y axis) and behavioral type (for each epoch ordered as PGHF, PGLF, SGHF, SGLF). Left PPD data, and right Gamma data. Colored dots represent individual units involved in FPs: blue dots indicate CV2 of units involved in FP patterns found for all surrogate techniques, orange dots for UD surrogates, green dots for UD and UDD only, and red dots for other combinations of different surrogate techniques. Grey dots represent the CV2 of individual units not involved in any FP.

5.4 Supplementary Figure 2

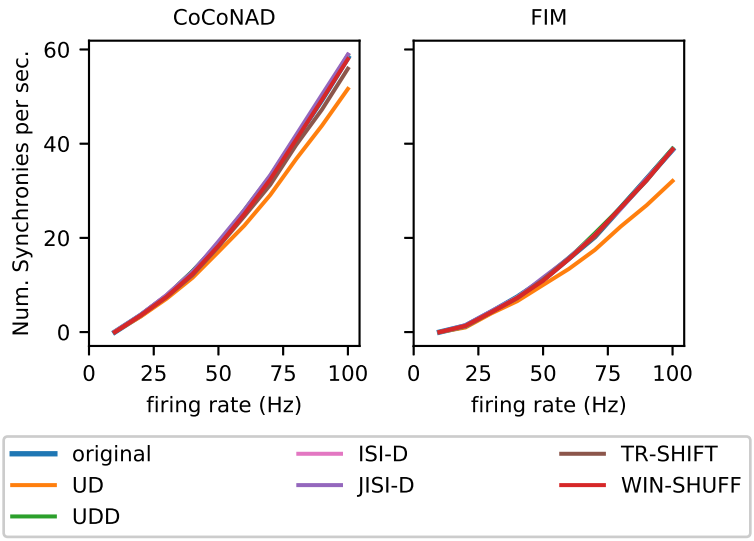


Figure 10. Comparison of number of synchronies detected by CoCoNAD and FIM in function of firing rate. The original data, for each firing rate, consists of 100 realizations of two independent 5s-long PPD spike trains ($d = 1.6$ms). Firing rate ranges from 10 to 100Hz in steps of 10Hz. Synchronies are measured with a temporal precision of 5ms. Spike trains are analyzed with CoCoNAD (left) and FIM (right), and with different surrogate techniques (colored). One surrogate instance is created for each spike train realization.

Author Contributions

Conceptualization, A.S., P.B., G.P., S.G.; Data generation, A.S., P.B.; Data analysis, A.S., P.B.; Software implementation, A.S., P.B.; Visualization, A.S., P.B.; Writing, A.S., P.B., G.P., S.G.; Project administration, S.G.; Funding acquisition, S.G. All authors have read and agreed to the published version of the manuscript.

Acknowledgments

This project has received funding from the European Union's Horizon 2020 Framework Programme for Research and Innovation under Specific Grant Agreement No. 785907 (Human Brain Project SGA2), No. 945539 (Human Brain Project SGA3), from the Helmholtz Association Initiative and Networking Fund under project number ZT-I-0003, and from the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - 368482240/GRK2416. We thank Sebastian Lehmann (SBC) for the help in the design and development of the graphics.

References

1. Louis S, Borgelt C, Grün S. Generation and Selection of Surrogate Methods for Correlation Analysis. In: Rotter S, Grün S, editors. Analysis of Parallel Spike Trains. Berlin: Springer; 2010. p. 359–382.

2. Abeles M, Gat I. Detecting precise firing sequences in experimental data. Journal of Neuroscience Methods. 2001;107(1–2):141–154.

3. Grün S, Diesmann M, Aertsen A. Unitary events in multiple single-neuron spiking activity: I. Detection and significance. Neural Computation. 2002;14(1):43–80.

4. Grün S. Data-driven significance estimation of precise spike correlation. J Neurophysiol. 2009;101:1126–1140.

5. Quaglio P, Rostami V, Torre E, Grün S. Methods for identification of spike patterns in massively parallel spike trains. Biol Cybern. 2018; p. 1–24. doi:10.1007/s00422-018-0755-0.

6. Louis S, Borgelt C, Grün S. Complexity distribution as a measure for assembly size and temporal precision. Neural Networks. 2010;23(6):705–712. doi:10.1016/j.neunet.2010.05.004.

7. Kass RE, Ventura V, Brown EN. Statistical issues in the analysis of neuronal data. Journal of Neurophysiology. 2005;1(94):8–25.

8. Ventura V. Bootstrap Tests of Hypotheses. In: Rotter S, Grün S, editors. Analysis of Parallel Spike Trains. Berlin: Springer; 2010. p. 383–308.

9. Efron B, Tibshirani RJ. An Introduction to the Bootstrap. vol. 57 of Monographs on Statistics and Applied Probability. Boca Raton, London, New York, Washington D.C.: Chapmann and Hall/CRC; 1993.

10. Villa AE, Abeles M. Evidence for spatiotemporal firing patterns within the auditory thalamus of the cat. 1990;509(2):325–327.

11. Martignon L, Von Hassein H, Grün S, Aertsen A, Palm G. Detecting higher-order interactions among the spiking events in a group of neurons. Biological cybernetics. 1995;73(1):69–81.

12. Riehle A, Grün S, Diesmann M, Aertsen A. Spike synchronization and rate modulation differentially involved in motor cortical function. Science. 1997;278(5345):1950–1953. doi:10.1126/science.278.5345.1950.

13. Prut Y, Vaadia E, Bergman H, Haalman I, Slovin H, Abeles M. Spatiotemporal structure of cortical activity: properties and behavioral relevance. J Neurophysiol. 1998;79(6):2857–2874.

14. Villa AE, Tetko IV, Hyland B, Najem A. Spatiotemporal activity patterns of rat cortical neurons predict responses in a conditioned task. Proceedings of the National Academy of Sciences. 1999;96(3):1106–1111.

15. Kilavik BE, Roux S, Ponce-Alvarez A, Confais J, Grün S, Riehle A. Long-term modifications in motor cortical dynamics induced by intensive practice. J Neurosci. 2009;29(40):12653–12663.

16. Shimazaki H, Amari Si, Brown EN, Grün S. State-space analysis of time-varying higher-order spike correlation for multiple neural spike train data. PLOS Comput Biol. 2012;8(3):e1002385.

17. Torre E, Picado-Muiño D, Denker M, Borgelt C, Grün S. Statistical evaluation of synchronous spike patterns extracted by frequent item set mining. Frontiers in computational neuroscience. 2013;7:132.

18. Quaglio P, Yegenoglu A, Torre E, Endres DM, Grün S. Detection and evaluation of spatio-temporal spike patterns in massively parallel spike train data with SPADE. Frontiers in computational neuroscience. 2017;11:41.

19. Stella A, Quaglio P, Torre E, Grün S. 3d-SPADE: Significance evaluation of spatio-temporal patterns of various temporal extents. Biosystems. 2019;185:104022. doi:10.1016/j.biosystems.2019.104022.

20. Stella A. Comparison of statistical methods for spatio-temporal patterns detection in multivariate point processes: an application to neuroscience (Master's thesis); 2017.

21. Torre E, Quaglio P, Denker M, Brochier T, Riehle A, Grün S. Synchronous spike patterns in macaque motor cortex during an instructed-delay reach-to-grasp task. Journal of Neuroscience. 2016;36(32):8329–8340.

22. Yegenoglu A, Quaglio P, Torre E, Grün S, Endres D. Exploring the usefulness of formal concept analysis for robust detection of spatio-temporal spike patterns in massively parallel spike trains. In: International Conference on Conceptual Structures. Springer; 2016. p. 3–16.

23. Picado-Muiño D, Borgelt C, Berger D, Gerstein GL, Grün S. Finding neural assemblies with frequent item set mining. Front Neuroinformatics. 2013;7:9.

24. Borgelt C, Picado-Muiño D. Simple Pattern Spectrum Estimation for Fast Pattern Filtering with CoCoNAD. International Symposium on Intelligent Data Analysis. 2013;.

25. Borgelt C, Picado-Muiño D. Finding Frequent Patterns in Parallel Point Processes. International Symposium on Intelligent Data Analysis. 2013;.

26. Grün S, Riehle A, Diesmann M. Effect of cross-trial nonstationarity on joint-spike events. Biological cybernetics. 2003;88(5):335–351.

27. Pipa G, Grün S, van Vreeswijk C. Impact of Spike Train Autostructure on Probability Distribution of Joint Spike Events. Neural Comput. 2013;25. doi:10.1162/NECO_a_00432.

28. Date A, Bienenstock E, Geman S. On the temporal resolution of neural activity. Division of Applied Mathematics, Brown University; 1998.

29. Hatsopoulos N, Geman S, Amarasingham A, Bienenstock E. At what time scale does the nervous system operate? Neurocomputing. 2003;52–54:25–29.

30. Stark E, Abeles M. Unbiased estimation of precise temporal correlations between spike trains. Journal of Neuroscience Methods. 2009;179(1):90–100. doi:https://doi.org/10.1016/j.jneumeth.2008.12.029.

31. Pazienti A, Maldonado PE, Diesmann M, Grün S. Effectiveness of systematic spike dithering depends on the precision of cortical synchronization. Brain research. 2008;1225:39–46.

32. Gerstein GL. Searching for significance in spatio-temporal firing patterns. Acta Neurobiol Exp. 2004;64:203–207.

33. Maldonado P, Babul C, Singer W, Rodriguez E, Berger D, Grün S. Synchronization of Neuronal Responses in Primary Visual Cortex of Monkeys Viewing Natural Images. J Neurophysiol. 2008;100(3):1523–1532.

34. Brochier T, Zehl L, Hao Y, Duret M, Sprenger J, Denker M, et al. Massively parallel recordings in macaque motor cortex during an instructed delayed reach-to-grasp task. Scientific Data. 2018;5:180055. doi:10.1038/sdata.2018.55.

35. Louis S, Gerstein GL, Grün S, Diesmann M. Surrogate spike train generation through dithering in operational time. Front Comput Neurosci. 2010;4(127).

36. Pipa G, Wheeler DW, Singer W, Nikolic D. NeuroXidence: reliable and efficient analysis of an excess or deficiency of joint-spike events. Journal of Neuroscience Methods. 2008;25:64–88.

37. Zaki MJ. Mining Non-Redundant Association Rules. Data Mining and Knowledge Discovery. 2004;9(3):223–248. doi:10.1023/B:DAMI.0000040429.96086.c7.

38. Borgelt C. Frequent Item Set Mining. In: Wiley Interdisciplinary Reviews (WIREs): Data Mining and Knowledge Discovery. vol. 2. J. Wiley & Sons, Chichester, United Kingdom; 2012. p. 437–456.

39. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society Series B (Methodological). 1995;57(1):289–300.

40. Riehle A, Wirtssohn S, Grün S, Brochier T. Mapping the spatio-temporal structure of motor cortical LFP and spiking activities during reach-to-grasp movements. Frontiers in Neural Circuits. 2013;7. doi:10.3389/fncir.2013.00048.

41. Riehle A, Brochier T, Nawrot M, Grün S. Behavioral context determines network state and variability dynamics in monkey motor cortex. Front Neural Circuits. 2018;12(52). doi:10.3389/fncir.2018.00052.

42. Mochizuki Y, Onaga T, Shimazaki H, Shimokawa T, Tsubo Y, Kimura R, et al. Similarity in neuronal firing regimes across mammalian species. Journal of Neuroscience. 2016;36(21):5736–5747. doi:10.1523/jneurosci.0230-16.2016.

43. Shmiel T, Drori R, Shmiel O, Ben-Shaul Y, Nadasdy Z, Shemesh M, et al. Temporally precise cortical firing patterns are associated with distinct action segments. J Neurophysiol. 2006;96(5):2645–2652.

44. Gansel KS, Singer W. Detecting multineuronal temporal patterns in parallel spike trains. Frontiers in neuroinformatics. 2012;6.

45. Torre E, Canova C, Denker M, Gerstein G, Helias M, Grün S. ASSET: analysis of sequences of synchronous events in massively parallel spike trains. PLOS Comput Biol. 2016;12(7):e1004939. doi:10.1371/journal.pcbi.1004939.

46. Holt GR, Softky WR, Koch C, Douglas RJ. Comparison of discharge variability *in vitro* and in *in vivo* in cat visual cortex neurons. J Neurophysiol. 1996;75(5):1806–1814.

47. Deger M, Helias M, Boucsein C, Rotter S. Statistical properties of superimposed stationary spike trains. Journal of Computational Neuroscience. 2012;32(3):443–463.

48. Nawrot MP, Boucsein C, Rodriguez Molina V, Riehle A, Aertsen A, Rotter S. Measurement of variability dynamics in cortical spike trains. Journal of Neuroscience Methods. 2008;169:374–390.

49. Bouss P. Statistical Evaluation of Dithering Methods for Pattern Detection (Master's thesis); 2020.

50. Shinomoto S. Estimating the Firing Rate. In: Rotter S, Grün S, editors. Analysis of Parallel Spike Trains. Berlin: Springer; 2010. p. 21–35.

51. Shimazaki H, Shinomoto S. Kernel bandwidth optimization in spike rate estimation. Journal of computational neuroscience. 2010;29(1):171–182.

52. van Vreeswijk C. Stochastic Models of Spike Trains. In: Rotter S, Grün S, editors. Analysis of Parallel Spike Trains. Berlin: Springer; 2010. p. 3–20.

53. Harrison MT, Geman S. A rate and history-preserving resampling algorithm for neural spike trains. Neural Computation. 2009;21(5):1244–1258.

54. Gerstein GL, Perkel DH, Subramanian K. Identification of functionally related neural assemblies. Brain research. 1978;140(1):43–62.

55. Pipa G, Grün S. Non-Parametric Significance Estimation of Joint-Spike Events by Shuffling and Resampling. Neurocomputing. 2003;52–54:31–37.

56. Pipa G, Diesmann M, Grün S. Significance of Joint-Spike Events Based on Trial-Shuffling by Efficient Combinatorial Methods. Complexity. 2003;8(4):79–86.

57. Pazienti A, Diesmann M, Grün S. Bounds of the ability to destroy precise coincidences by spike dithering. In: International Symposium on Brain, Vision, and Artificial Intelligence. Springer; 2007. p. 428–437.

58. Pipa G, Riehle A, Grün S. Validation of task-related excess of spike coincidences based on NeuroXidence. Neurocomputing. 2007;70:2064–2068.

59. Smith MA, Kohn A. Spatial and temporal scales of neuronal correlation in primary visual cortex. Journal of Neuroscience. 2008;28(48):12591–12603.

60. Dann B, Michaels JA, Schaffelhofer S, Scherberger H. Uniting functional network topology and oscillations in the fronto-parietal single unit network of behaving primates. Elife. 2016;5:e15719.

61. Grün S, Diesmann M, Aertsen A. Unitary events in multiple single-neuron spiking activity: II. Nonstationary data. Neural computation. 2002;14(1):81–119.

62. Harrison M, Amarasingham A, Geman S. Jitter methods for investigating spike train dependencies. Computational and Systems Neuroscience Abstracts III-17. 2007;.

63. Ikegaya Y, Aaron G, Cossart R, Aronov D, Lampl I, Ferster D, et al. Synfire chains and cortical songs: temporal modules of cortical activity. Science. 2004;304(5670):559–564.

64. Masuda N, Aihara K. Duality of rate coding and temporal coding in multilayered feedforward networks. Neural Computation. 2003;15(1):103–125.

65. Nádasdy Z, Hirase H, Czurkó A, Csicsvari J, Buzsáki G. Replay and time compression of recurring spike sequences in the hippocampus. Journal of Neuroscience. 1999;19(21):9497–9507.

66. Rivlin-Etzion M, Ritov Y, Heimer G, Bergman H, Bar-Gad I. Local shuffling of spike trains boosts the accuracy of spike train spectral analysis. Journal of neurophysiology. 2006;95(5):3245–3256.

67. Ricci L, Castelluzzo M, Minati L, Perinelli A. Generation of surrogate event sequences via joint distribution of successive inter-event intervals. Chaos: An Interdisciplinary Journal of Nonlinear Science. 2019;29(12):121102.

68. Perinelli A, Castelluzzo M, Minati L, Ricci L. SpiSeMe: A multi-language package for spike train surrogate generation. Chaos: An Interdisciplinary Journal of Nonlinear Science. 2020;30(7):073120.

69. Platkiewicz J, Stark E, Amarasingham A. Spike-centered jitter can mistake temporal structure. Neural computation. 2017;29(3):783–803.

70. Abeles M. Corticonics: Neural Circuits of the Cerebral Cortex. 1st ed. Cambridge: Cambridge University Press; 1991.

71. Diesmann M, Gewaltig MO, Aertsen A. Stable propagation of synchronous spiking in cortical neural networks. Nature. 1999;402(6761):529–533.

72. Izhikevich EM. Polychronization: computation with spikes. Neural computation. 2006;18(2):245–282.

73. Bienenstock E. A model of neocortex. Network: Computation in neural systems. 1995;6(2):179–224.

74. Oettl LL, Scheller M, Filosa C, Wieland S, Haag F, Loeb C, et al. Phasic dopamine reinforces distinct striatal stimulus encoding in the olfactory tubercle driving dopaminergic reward prediction. Nature communications. 2020;11(1):1–14.

75. Staude B, Grün S, Rotter S. Higher-order correlations in non-stationary parallel spike trains: statistical modeling and inference. Frontiers in Computational Neuroscience. 2010;4. doi:10.3389/fncom.2010.00016.

76. Russo E, Durstewitz D. Cell assemblies at multiple time scales with arbitrary lag constellations. eLife. 2017;6:e19428. doi:10.7554/eLife.19428.

77. Quaglio P. Comparative correlation analyses of high-dimensional point processes: applications to neuroscience (Master's thesis); 2015.

78. Kreuz T, Mulansky M, Bozanic N. SPIKY: a graphical user interface for monitoring spike train synchrony. Journal of neurophysiology. 2015;113(9):3432–3445.

79. Hegger R, Kantz H, Schreiber T. Practical implementation of nonlinear time series methods: The TISEAN package. Chaos: An Interdisciplinary Journal of Nonlinear Science. 1999;9(2):413–435.

80. Davis P. Gamma Function and Related Functions. In: Abramowitz M, Stegun I, editors. Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables. 10th ed. Applied Mathematics Series 55. Washington D.C.: National Bureau of Standards; 1972. p. 253–266.

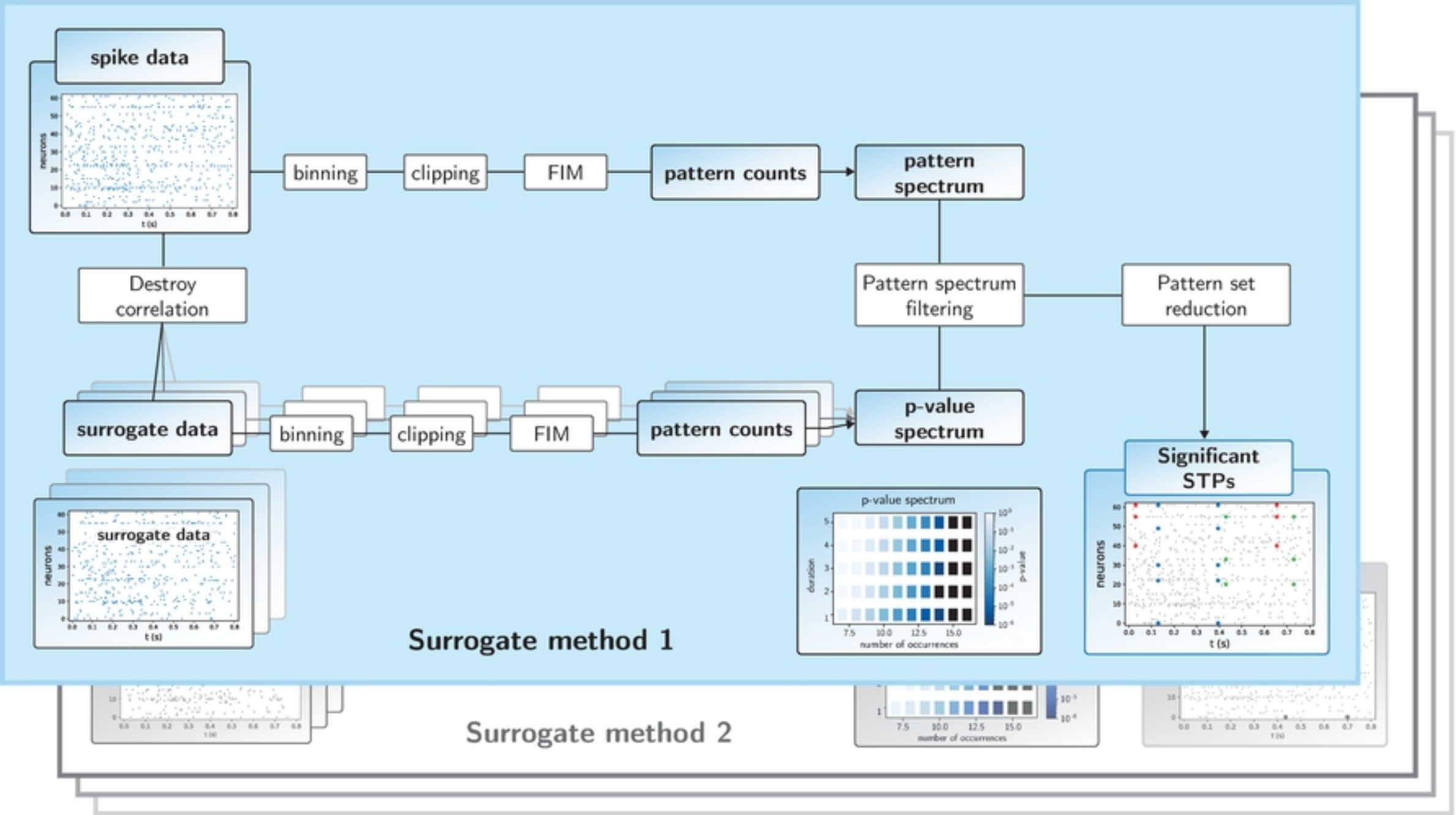

spike data
binning
clipping
FIM
pattern counts
pattern spectrum
Destroy correlation
Pattern spectrum filtering
Pattern set reduction
surrogate data
binning
clipping
FIM
pattern counts
p-value spectrum
Significant STPs
surrogate data
p-value spectrum
Surrogate method 1
Surrogate method 2

FIgure 1

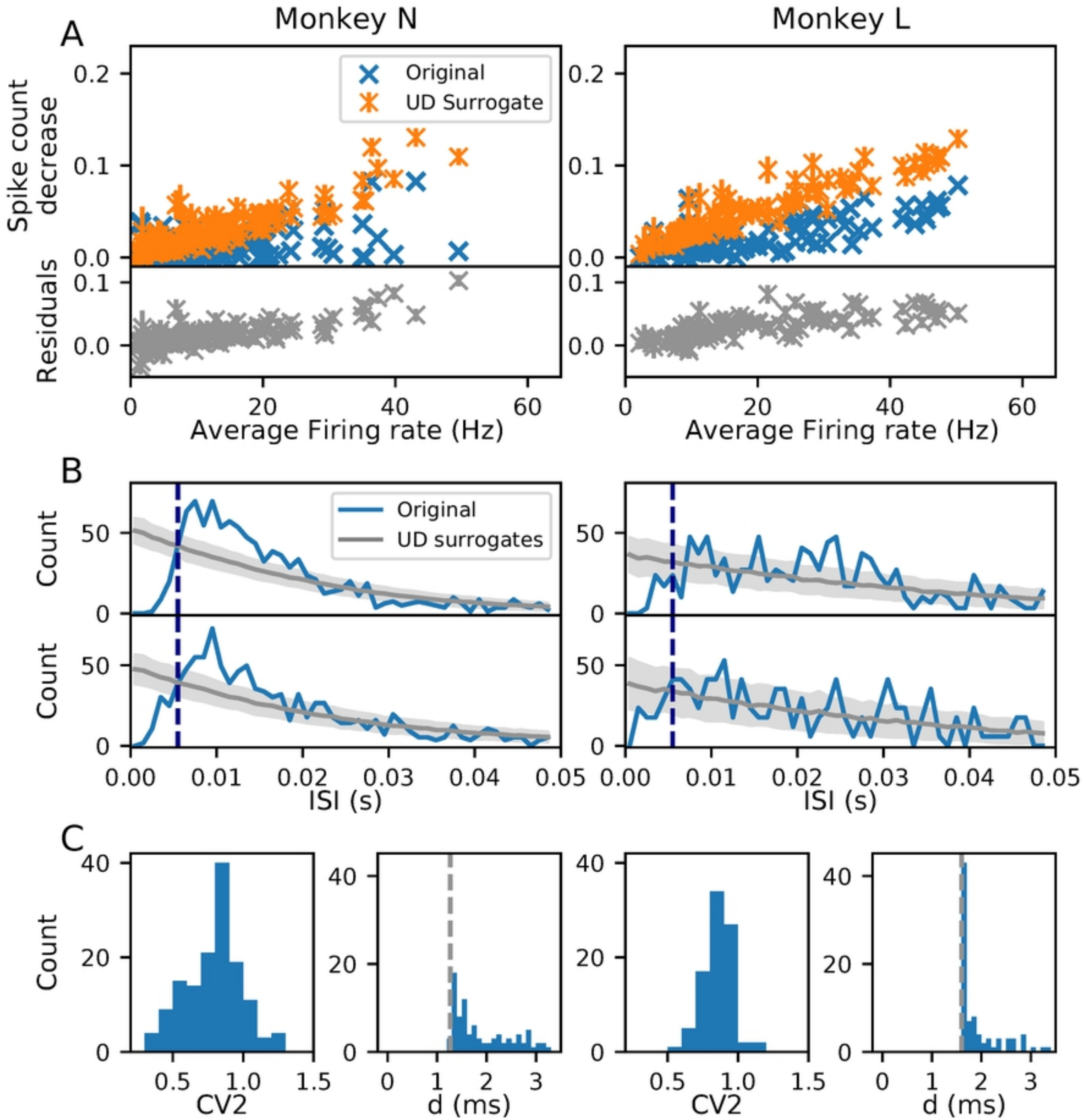

A
Monkey N
Monkey L
Original
UD Surrogate
Spike count decrease
Residuals
Average Firing rate (Hz)
B
Original
UD surrogates
Count
ISI (s)
C
Count
CV2
d (ms)

Figure 2

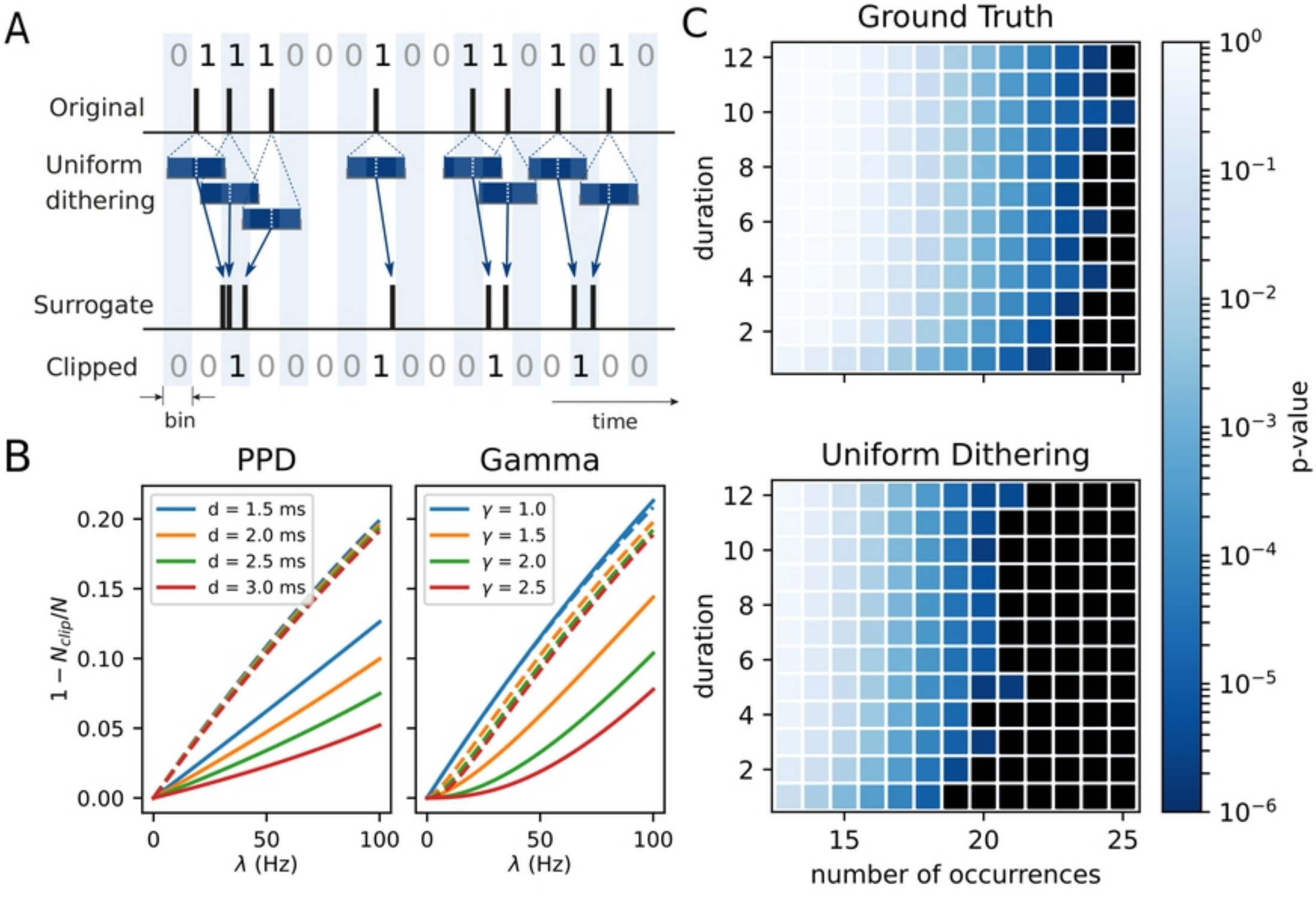

A
0 1 1 1 0 0 0 1 0 0 1 1 0 1 0 1 0
Original
Uniform dithering
Surrogate
Clipped
0 0 1 0 0 0 0 1 0 0 0 1 0 0 1 0 0
bin
time
B
PPD
Gamma
d = 1.5 ms
d = 2.0 ms
d = 2.5 ms
d = 3.0 ms
γ = 1.0
γ = 1.5
γ = 2.0
γ = 2.5
$1 - N_{clip}/N$
λ (Hz)
C
Ground Truth
Uniform Dithering
duration
number of occurrences
p-value

Figure 3

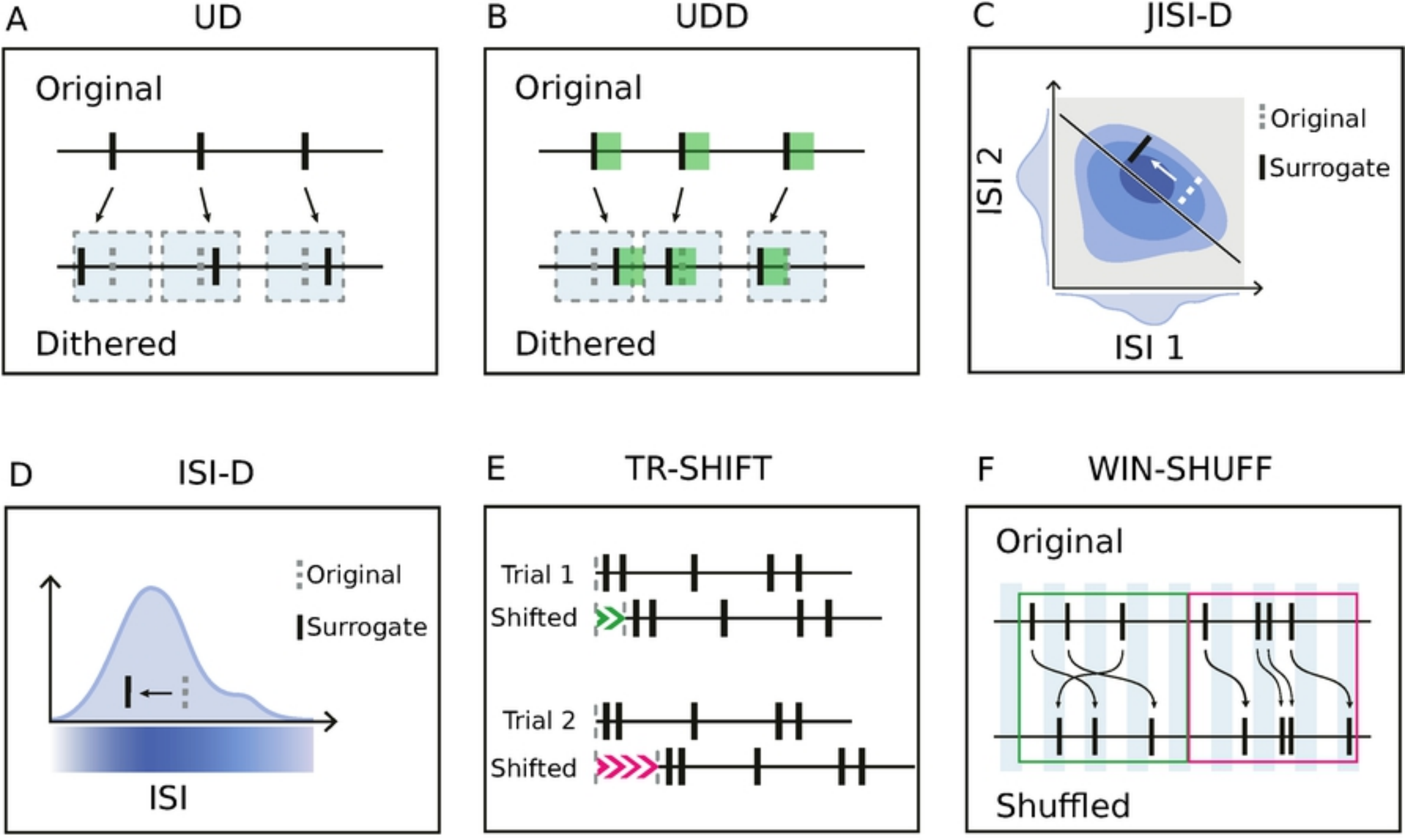

A
UD
Original
Dithered
B
UDD
Original
Dithered
C
JISI-D
Original
Surrogate
ISI 2
ISI 1
D
ISI-D
Original
Surrogate
ISI
E
TR-SHIFT
Trial 1
Shifted
Trial 2
Shifted
F
WIN-SHUFF
Original
Shuffled

Figure 4

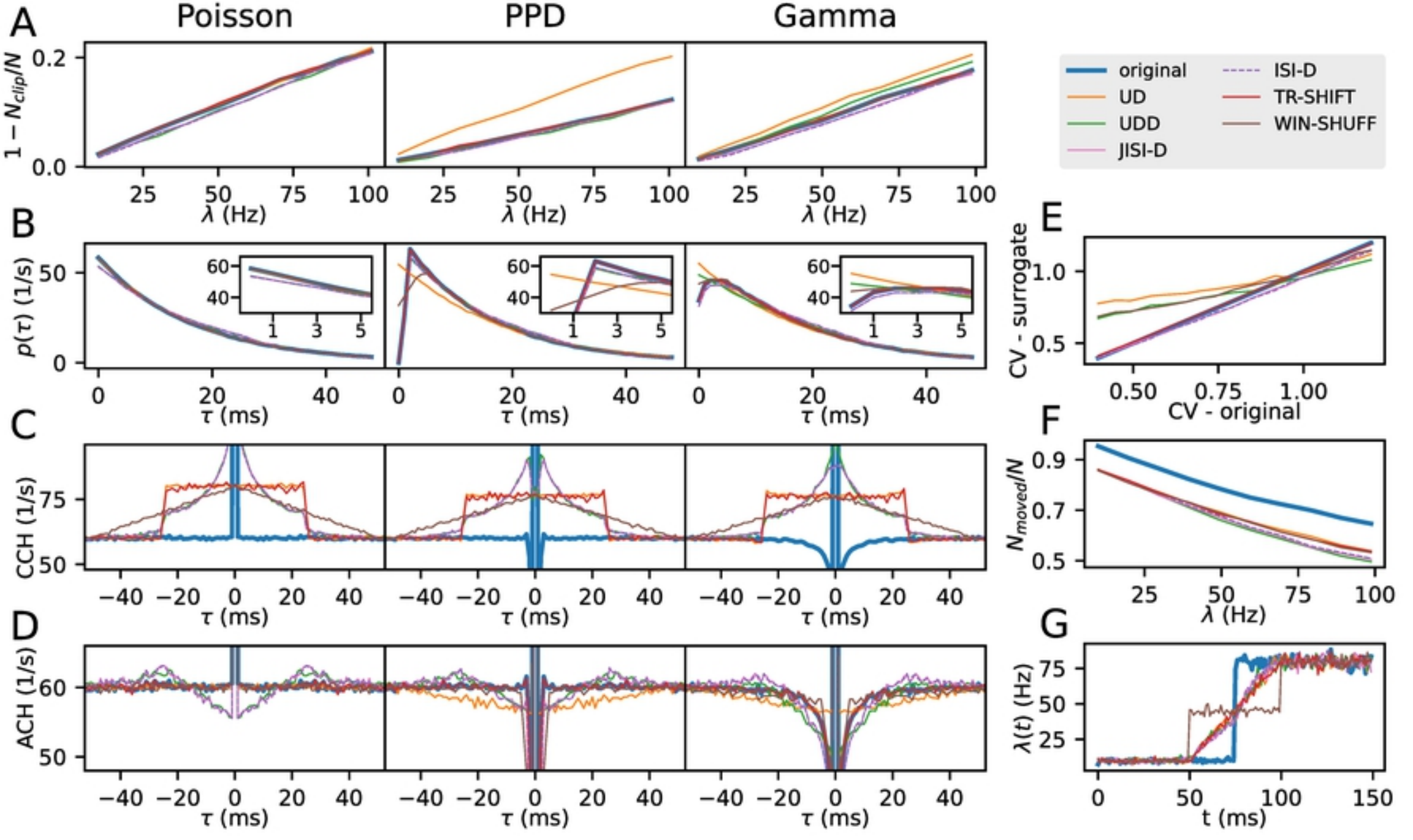

A
Poisson
PPD
Gamma
$1 - N_{clip}/N$
λ (Hz)
original
UD
UDD
JISI-D
ISI-D
TR-SHIFT
WIN-SHUFF
B
p(τ) (1/s)
τ (ms)
C
CCH (1/s)
D
ACH (1/s)
E
CV - surrogate
CV - original
F
N_{moved}/N
G
λ(t) (Hz)
t (ms)

Figure 5

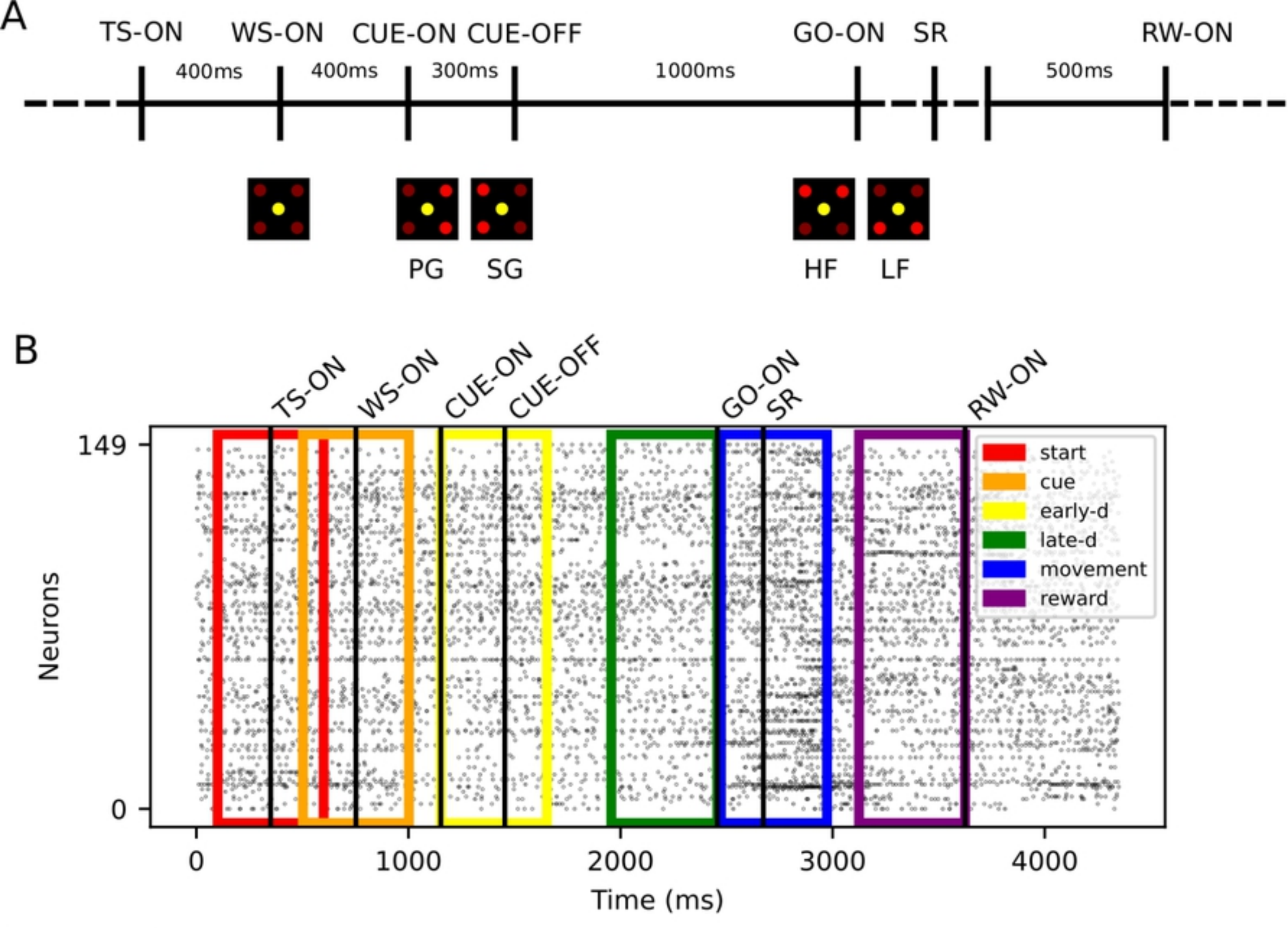

A
TS-ON
WS-ON
CUE-ON
CUE-OFF
GO-ON
SR
RW-ON
400ms
400ms
300ms
1000ms
500ms
PG
SG
HF
LF
B
TS-ON
WS-ON
CUE-ON
CUE-OFF
GO-ON
SR
RW-ON
149
0
Neurons
start
cue
early-d
late-d
movement
reward
0
1000
2000
3000
4000
Time (ms)

FIgure 6

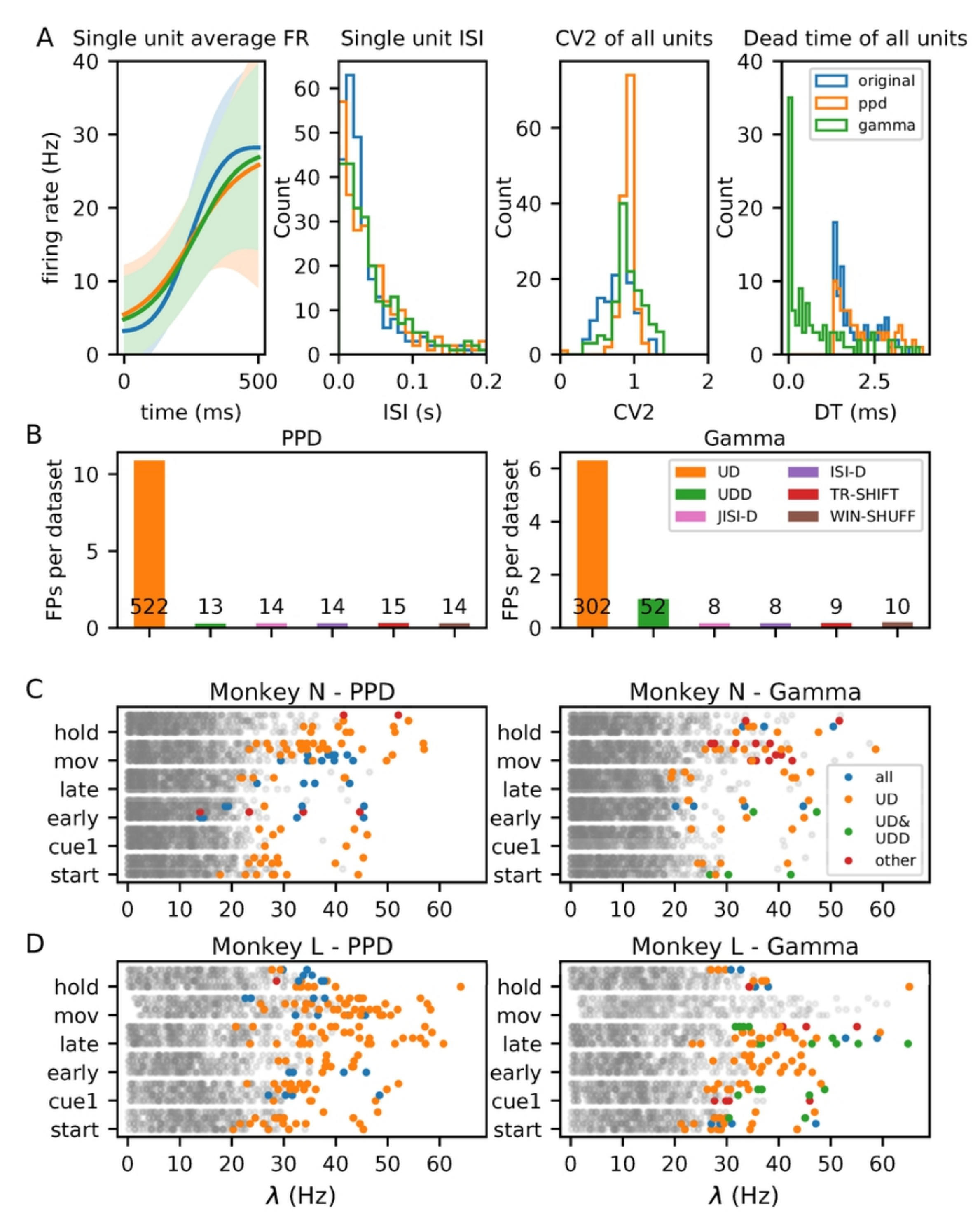

A
Single unit average FR
Single unit ISI
CV2 of all units
Dead time of all units
firing rate (Hz)
time (ms)
Count
ISI (s)
CV2
DT (ms)
original
ppd
gamma
B
PPD
Gamma
FPs per dataset
522
13
14
14
15
14
302
52
8
8
9
10
UD
UDD
JISI-D
ISI-D
TR-SHIFT
WIN-SHUFF
C
Monkey N - PPD
Monkey N - Gamma
hold
mov
late
early
cue1
start
all
UD
UD&
UDD
other
D
Monkey L - PPD
Monkey L - Gamma
λ (Hz)

Figure 7

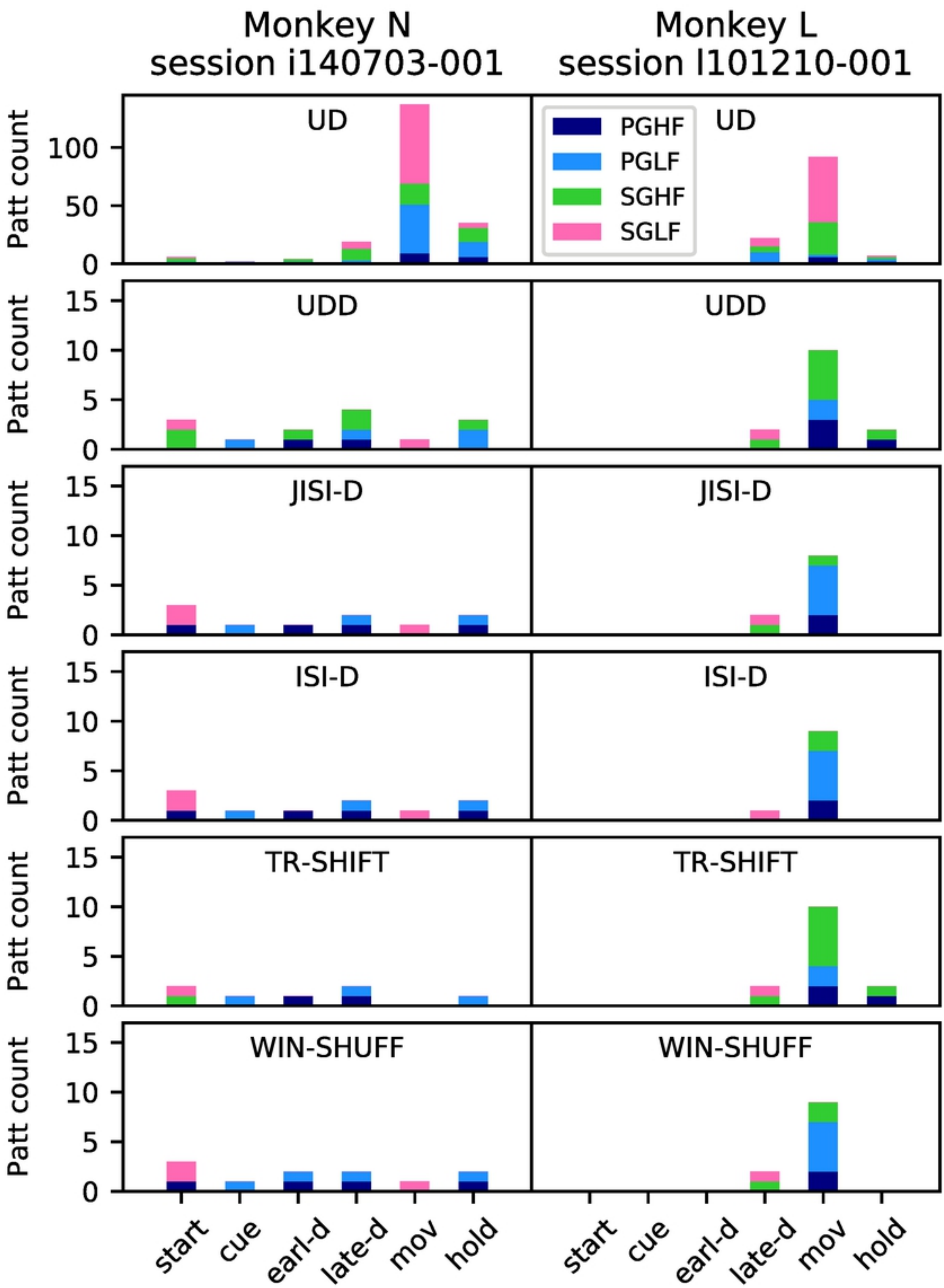
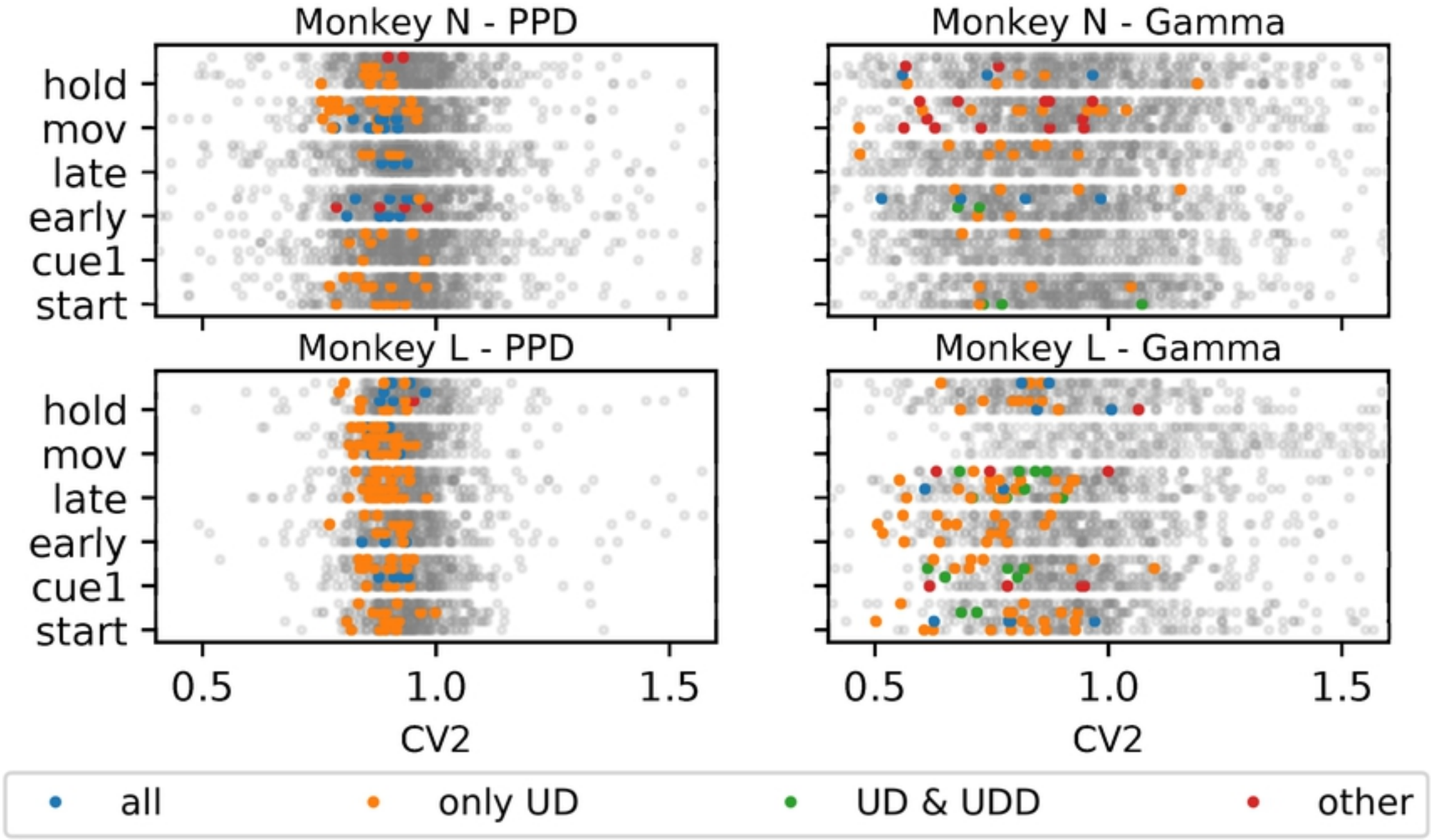
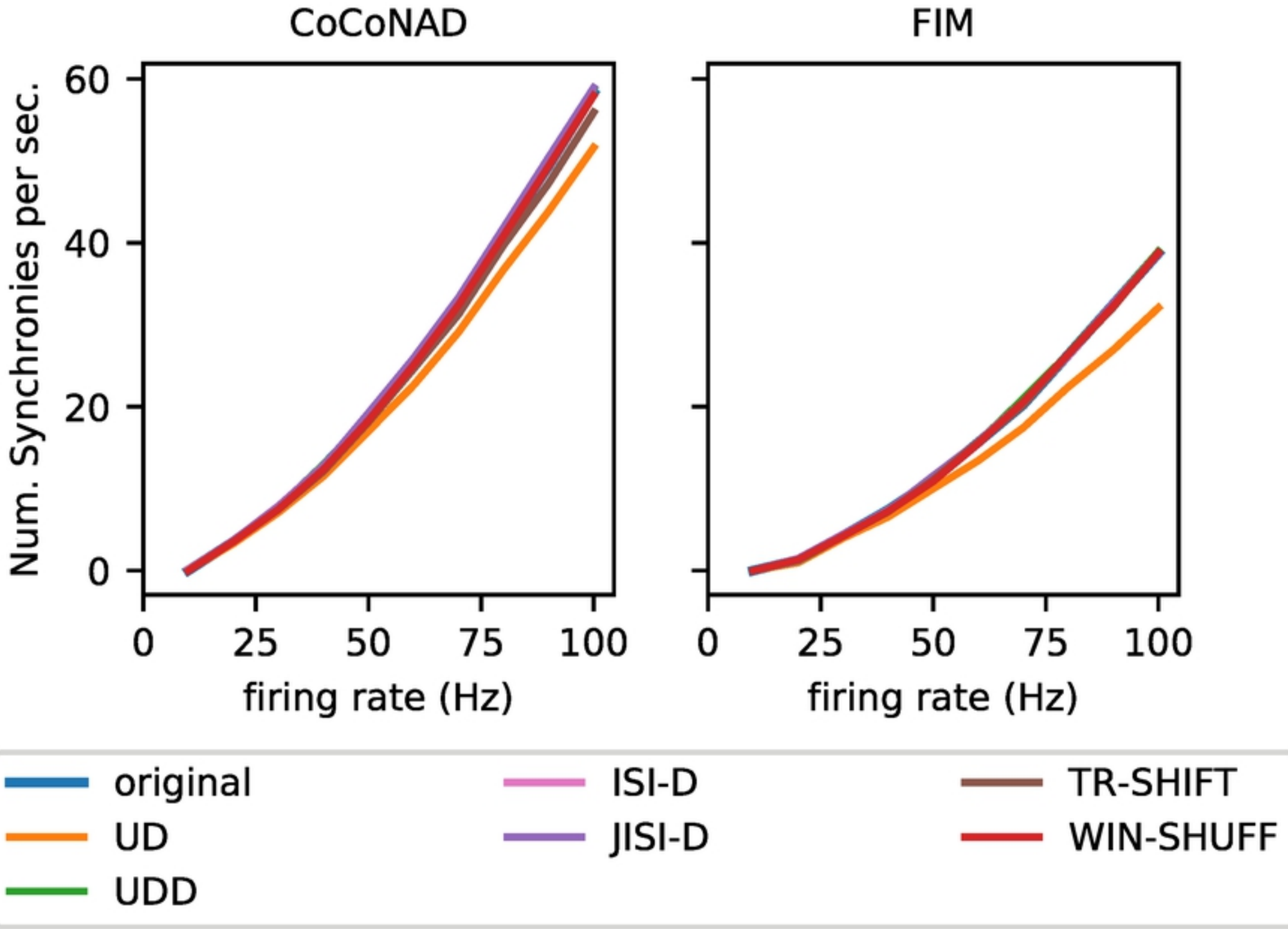

Monkey N
session i140703-001
Monkey L
session l101210-001
PGHF
PGLF
SGHF
SGLF
UD
UDD
JISI-D
ISI-D
TR-SHIFT
WIN-SHUFF
Patt count
start
cue
earl-d
late-d
mov
hold

Figure 8



Supplementary Figure 1



Supplementary Figure 2