# Detecting rhythmic spiking through the power spectra of point process model residuals

Karin M. Cox[1,2*], Daisuke Kase[2,3,4], Taieb Znati[1], Robert S. Turner[2,3,4,5*]

[1] Department of Computer Science, University of Pittsburgh, Pittsburgh, Pennsylvania, 15260, United States of America
[2] Aligning Science Across Parkinson's (ASAP) Collaborative Research Network, Chevy Chase, Maryland, 20815, United States of America
[3] Department of Neurobiology, University of Pittsburgh, Pittsburgh, Pennsylvania, 15261, United States of America
[4] Systems Neuroscience Center, University of Pittsburgh, Pittsburgh, Pennsylvania, 15261, United States of America
[5] Center for Neuroscience, University of Pittsburgh, Pittsburgh, Pennsylvania, 15261, United States of America

[*] Corresponding authors

Email: kmc51@pitt.edu (KMC), rturner@pitt.edu (RST)

ORCID identifiers: 0000-0002-5418-0508 (KMC), 0000-0001-6787-5862 (DK), 0000-0002-9389-3033 (TZ), 0000-0002-6074-4365 (RST)

## Abstract

Oscillations figure prominently in models of brain function and dysfunction. It would be informative to analyze oscillations at the level of individual neurons, by estimating the power spectral density (PSD) of spike trains. Unfortunately, spike train spectra exhibit a global distortion generated by the neuronal recovery period ("RP", the post-spike drop in spike probability, which can extend beyond the refractory period). This distortion can increase false negative and false positive rates in statistical tests for oscillatory effects. An existing "shuffling" procedure corrects for RP distortion by removing the spectral component explained by the inter-spike interval (ISI) distribution. However, this procedure sacrifices any oscillation-related information in the ISIs, and power at the corresponding frequencies in the PSD. Here, we ask whether a three-step "residuals" method can improve upon the shuffling method's performance. First, we estimate the RP duration ($n_r$) from the ISI distribution. Second, we fit the spike train with a point process model that predicts spike likelihood based on the time elapsed since the most recent of any spikes falling within the preceding $n_r$ milliseconds. Third, we compute the PSD of the model's residuals. We initially compared the residuals and shuffling methods' performance over diverse synthetic spike trains. The residuals method generally yielded the most accurate classification of true- versus false-positive oscillatory power, with this result principally driven by enhanced sensitivity to oscillations in sparse spike trains. Subsequent evaluations used single-unit data from the internal globus pallidus (GPi) and ventrolateral anterior thalamus (VLa) of a parkinsonian monkey, in which pathological alpha-beta oscillations (8-30 Hz) were anticipated. Over these units, the residuals method reported the greatest incidence of significant alpha-beta power, with low firing rates predicting residuals-selective oscillation detection. Overall, these results encourage further development of the residuals approach, including expansion to capture additional aperiodic spectral features.

## Author Summary

Models of brain function propose that some neurons fire action potentials ("spikes") rhythmically. That is, the spikes arise from a latent rate function that oscillates at a consistent frequency. We might attempt to detect such oscillations by seeking peaks against a flat baseline in a spike train's power spectrum. Yet the baseline is not flat: The transient post-spike suppression in spike rate ("recovery period") introduces a distortion that can both obscure true oscillatory peaks and generate false positives. An established method aims to address the distortion by removing the spectral content that can be recreated with shuffled versions of the spike train. However, this procedure can remove some information about spike rhythmicity, thereby hindering sensitivity to true oscillations. We developed an alternative method that uses regression to remove recovery period effects from the spike train, and computes the spectrum of the residuals. Evaluations on synthetic data demonstrated that this "residuals" method improved upon the established method's sensitivity to true oscillations. In experimentally-acquired data from a primate model of Parkinson's Disease, the residuals method also returned increased detections of the expected pathological 8-30 Hz oscillations. These results suggest that the residuals method may support future discoveries of otherwise undetected spike rhythms.

## Introduction

To estimate oscillations in neural activity, it is conventional to compute the power spectral density (PSD) of measures of extracellular fields (e.g., local field potentials (LFPs), or electrocorticograms (ECoG)). Such signals reflect a complex mixture of contributions from neuronal populations [1], and are therefore not optimal when the objective is to investigate oscillations at the level of individual neurons. In these situations, we can attempt to apply spectral analysis to single-unit spike trains. Ideally, the resulting PSD output – along with post-processing and statistics -- should enable the inference of oscillations in the rate function that generated the observed spikes [2]. We adopt the working definition of an oscillation as a rhythmic modulation in action potential ("spike") occurrence within a narrow frequency range (expanding upon the usage from [3], in the context of continuous signals).

To detect true periodic components in a spike-derived PSD function accurately, we must first address a challenge: These spectra exhibit a distortion pattern that is common to phenomena subject to measurement "dead time" (i.e., an interval after an event, when the occurrence or detection of a subsequent event is impeded [4]). In the context of single-unit spike trains, this dead time is typically attributed to the neuronal refractory period, which may be absolute ($p$(spike) = 0) or relative (gradually recovering $p$(spike)). The simplest example of this distortion occurs in the context of a hypothetical spike train arising from a Poisson process interrupted by an absolute refractory period. In this case, the resulting PSD departs from the flat spectrum anticipated for an unmodified Poisson process, and is instead characterized by a trough over a range of low frequencies, followed by elevated power (often accompanied by visible ringing) over the high frequencies [5, 6]. The precise form of the distortion differs modestly for relative refractory periods (see Fig. 1 for a synthetic example), and is exacerbated by high firing rates (FR) and long refractory period durations [7].

The present research examines two methods – including a commonly-used approach (the "shuffling method"), and a novel alternative (the "residuals method") – that aim to remove this dead time-induced distortion from spike train spectra, and ultimately facilitate more sensitive and selective detection of the periodic effects of interest. Beyond these two options (and an analytical variant on the shuffling approach [8]), we are aware of no additional methods specifically designed to correct distortion in neuronal spike train spectra. Therefore, we limit our comparison to these two methods.

Before describing the correction methods, it is necessary to clarify a terminology choice: Instead of attributing the dead time to the refractory period, we will use a mechanism-neutral term: "recovery period" (see [9] for similar usage). In data from our lab (including the empirically-sampled units we present here), we have observed several examples of units that exhibit a dead time (i.e., sharp post-spike drops in the hazard function) that far exceeds the typical duration of the channel dynamics responsible for refractoriness (5-10 ms [10]). Such long post-spike pauses might be generated by either circuit-level or intrinsic mechanisms (such as those underlying pacemaker patterns [11]). Since neither of the methods that we discuss attempt to selectively correct for the biological refractory period *per se*, we adopt the generic recovery period (RP) terminology.

When a spike train features an oscillation, the corresponding peak in the PSD is superimposed on the recovery period-associated distortion pattern (see Fig. 2A for synthetic examples). The distortion complicates standard statistical procedures that aim to distinguish such oscillatory PSD peaks from baseline by using a null hypothesis model that is estimated from the spectrum itself. In particular, detection problems arise when the baseline does not account for the RP contribution.

An especially detailed treatment of RP-associated distortion [7] -- which also introduced the established shuffling method -- illustrated the oscillation detection challenges with a straightforward statistical test that assumes a Poisson process (in the time domain) and corresponding white noise (in the frequency domain) as the baseline. These authors observed that, over the highest frequencies (beyond the frequency bands typically prioritized by neuroscientists), the PSD functions of neuronal

spike trains tended to approach those of matched-FR Poisson processes, thereby motivating the construction of confidence intervals (CI) based on the mean and variance over these high-frequency points. Points in the PSD function that cross the CI bound (see Fig. 2A, dashed lines, for CI examples) are labeled as significant, with the possible added requirement that the points also fall within an *a priori* search range (set at (0, 100] Hz in the figures). Because this statistical test does not account for the RP distortion, it is prone to miss oscillatory peaks that fall within the low frequency trough (see the two examples in Fig. 2A), and incorrectly label as significant points at higher frequencies, beyond the trough, which exhibit elevated power as a consequence of the distortion effect alone [7]. Although the examples in Fig. 2A do not show false positives within the (0, 100] Hz search range, the PSD in the left panel does include a peak that crosses the CI threshold at ≈112 Hz. Note that this general pattern of false negatives and positives also appears when the CIs for the Poisson null hypothesis are estimated using an alternative analytical procedure (see [7] and [12] for details).

A switch to a different, but still RP-neglecting baseline model is insufficient to resolve these issues. For example, consider the common practice of fitting a "1/f-like" function to the spectra of field potential measures, in order to separate an assumed aperiodic baseline from periodic effects [3]. Although this practice can be effective for such population-level signals – on which individual units' RPs will not have a discernable impact – it is unclear how a 1/f-like function could be successfully fit to a raw spike train spectrum.

To account for the RP distortion, the correction method proposed by [7] builds upon the null hypothesis that the spike train emerges from a renewal process with a matched inter-spike interval (ISI) distribution. Recall that a renewal process is a sequence of events (here, spikes) for which the waiting times (the ISIs) are independently sampled from an identical distribution [13]. Since we define the recovery period as an influence solely on the waiting time between two successive events, and not on any higher order structure (e.g., correlations between successive waiting times), the renewal null hypothesis does capture the temporal structure attributable to the RP.

The ISI shuffling method described by [7] estimates the renewal equivalent spectrum for a spike train. According to the simplest, global version of this scheme, several control spike trains are generated by randomly permuting the ISIs of the original. The PSDs computed from these control spike trains are then averaged to estimate the renewal equivalent spectrum. An analytical alternative to this Monte Carlo procedure does also exist, which leverages an established method to compute the renewal process spectrum from an ISI distribution [8, 14]. Although this analytical approach is fast and non-stochastic, it does require the use of rectangular spectral analysis windows. For the present work, we focus on the shuffling approach, given its flexibility to tapered windows (e.g., Hamming windows, or the Slepian sequences used by the multitaper method [15]).

Regardless of whether shuffling or the analytical method is used, the resulting control spectrum is then divided out from the original spectrum to obtain the corrected PSD. Given a corrected spectrum, it is customary to proceed with the aforementioned procedure of constructing CIs based on the high frequencies, under the assumption that the null hypothesis of a flat spectrum is now appropriate.

Shuffling correction continues to see common use in single-unit studies, especially in research on the motor control circuitry (e.g., [16-20]), but also in other domains [21, 22]. The method is a well-justified option when researchers are willing to forgo the temporal information present in the ISI distribution. However, as noted by [7], the ISIs can carry information about the rhythmic occurrence of individual spikes, which the shuffling method will remove. A simple, rigid example of this occurs in the case of precisely timed pacemaking activity – that is, single spikes separated by a relatively consistent period $P$, plus or minus a modest amount of jitter (i.e., an approximation of a Dirac comb [23]). In such cases, the pacemaking interval is indistinguishable from a recovery period, implying that such patterns should be detected through means other than RP-corrected spectra (e.g., by seeking instances of extremely low ISI variance). A more subtle example occurs when rhythmic spiking appears to be driven by an

approximately sinusoidal function, but the mean FR is so low that spikes often occur only once per oscillation cycle, and may at times skip cycles altogether. To detect and statistically validate such cases, spectral analysis may prove more helpful, but how to effectively address RP distortion while preserving the sparse information regarding the underlying rhythmicity remains unresolved.

Here we explore a new "residuals" method which aims to selectively remove RP-associated variance from a spike train, while leaving intact the information present in the long-lag ISIs. The specific method we describe should be seen as one instantiation of a general, mixed time- and frequency-domain framework that can be adapted flexibly to accommodate analysis goals that are more complex than those prioritized here. We also note that the method represents the combination of a number of existing ideas and techniques from prior work that has sought to estimate spike rhythms exclusively in the time domain (e.g., [24-26]).

The residuals method consists of three stages. First, similar to previous works [24, 27], we use the ISI distribution to inform an estimate of the RP duration ($\hat{n}_r$) for the input spike train. Second, we construct a point process model (PPM) that predicts the time-varying spike probability ($p_{spk}(t)$) as a function of the time elapsed since the most recent spike, but only when a spike was observed within the last $\hat{n}_r$ ms; outside of this RP interval, the prediction relies on the default constant term alone. In other words, this "bounded last-spike" model estimates the component of the data that can be explained by the base firing rate and the first $\hat{n}_r$ ms of the estimated post-spike hazard function.

The fitted PPM is assumed to account for the variance associated with the RP, thereby implying that the raw residuals (actual - predicted values) should represent a "corrected" time series. Therefore, as the third step, we submit the residuals to spectral analysis. The resulting PSD is subjected to the same statistical procedures as were described for the shuffling-corrected spectra.

To our knowledge, no prior work has applied such an approach to the analysis of spike train oscillations. However, we do note some conceptual relatedness to a two-step PPM fitting procedure reported by [24], which was used to account for slow fluctuations in spike rate prior to seeking oscillations in the time domain.

We present a comparison of the shuffling and residuals methods on both synthetic and real neuronal spike trains. In the synthetic data – for which ground truth is certain -- we report a general advantage for the residuals method in the accurate classification of true and false oscillatory points in the corrected spectra. In follow-up analyses, we find that the residuals method is particularly advantageous in the detection of true oscillations, especially when these are of a moderately strong amplitude, and arise from units that fire at rates only modestly greater than the frequency of the oscillation to be detected.

The neuronal data were acquired from the internal globus pallidus (GPi) and ventrolateral anterior thalamus (VLa) of a parkinsonian monkey, thereby presenting a dataset for which the *a priori* expectation of pathological oscillations is high (specifically in the 12-30 Hz beta- and 8-12 Hz alpha-frequency ranges [28, 29]), and for which spiking may tend to be sparse (particularly within the VLa, [30]). As expected for biological data, the findings are complex. We do find several instances of candidate alpha-beta oscillations that are detected by the residuals method exclusively, especially amongst the slowest-spiking units. At the same time, we also find a small number of alpha-beta oscillations flagged exclusively by the shuffling method. We additionally observe aperiodic spectral components (1/f-like trends and probable bursting) that residuals but not shuffling correction leave intact, and propose options for expansion upon the residuals method that could account for these additional phenomena.

## Materials and Methods

### Ethics Statement

The single nonhuman primate (NHP) dataset analyzed in this study (see "Experimental Data Methods") was acquired under a protocol approved by the Institutional Animal Care and Use Committee

of the University of Pittsburgh (#18093682). Experimental procedures and animal care complied with the National Institutes of Health Guide for the Care and Use of Laboratory Animals, the PHS Policy on the Humane Care and Use of Laboratory Animals, and the recommendations of the Weatherall report. The animal was housed with a single cage mate in a climate-controlled room, received regular enrichments, and consumed a diet that included fresh fruits and vegetables daily. Surgical procedures were conducted under general anesthesia, with analgesics administered post-operatively.

## Methods Overview

In the subsections that follow, we first describe the procedures for comparing the two correction methods (shuffling and residuals) on synthetic spike trains. These include the steps for generating the spike trains, calculating the original and corrected PSDs, and evaluating the methods' accuracy in detecting known oscillations. Subsequently, we describe the experimental procedures used in acquiring the NHP data, and also the few details of PSD analysis that differed between the NHP data and the synthetic spike trains.

All simulations and analyses were implemented in Matlab (R2022a, The Mathworks, Inc., Natwick, MA; RRID:SCR_001622). Matlab code sufficient to reproduce the reported results is available on Github (https://github.com/kc13/residuals_spectral_analysis; DOI: 10.5281/zenodo.8323322). We make use of conventional Matlab notation for referring to the ranges spanned by / increments separating vector elements ([min:step:max]).

Data files are available in both a Zenodo repository (DOI: 10.5281/zenodo.8313070) and the Github repository (see also the Data Availability statement). The original synthetic datasets and NHP spike train data are available on Zenodo. The Github repository stores copies of the NHP spike train data, and also synthetic datasets that are represented in a postprocessed state (the subsampling and pROC output, see "Comparison of ROC curves for the Correction Methods"; note the original synthetic datasets exceed Github file size limits).

## Synthetic Data Methods

### *Spike Train Simulation*

We generated synthetic spike trains through the procedure described in [7]. For each time bin $t$ ($\Delta t$ = 1 ms), the spike probability $p_{spk}(t)$ was determined according to the discrete-time approximation of an inhomogeneous Poisson process:

$$p_{spk}(t) = \begin{cases} p_{SS}(t) = p_{base} + p_{osc} \cdot sin(2\pi f_{osc} t \cdot 0.001)\,, & n > n_r \\ p_{RP}(t) = p_{SS}(t) \cdot k^{n_r+1-n}, & n \leq n_r \end{cases} \tag{1}$$

Here, $n$ indexes the millisecond latency since the last spike occurrence, and $n_r$ denotes the duration of the recovery period (RP). Therefore, the spike probability at $t$ depends on whether the simulated unit is in the recovery period (RP) or has recovered the steady state (SS) firing pattern. During the steady state, $p_{spk}(t)$ reflects the sum of two components: (1) $p_{\text{base}}$, corresponding to the mean steady state firing rate (FR), and (2) an oscillation term, in which $0 \leq p_{osc} \leq p_{\text{base}}$ governs the oscillation amplitude, and $f_{osc}$ (specified in Hz) indicates the oscillation frequency. Following from [2], we set $p_{osc}$ indirectly through variation of the modulation index, $m = p_{osc}/p_{base}$; i.e., the ratio of the peak and mean steady state FRs, respectively. The above $p_{\text{osc}}$ constraint implies that $0 \leq m \leq 1$. For all simulations, $p_{base} << .5$, thereby implying that $p_{spk}(t)$ remained bounded in [0,1].

During the recovery period, $p_{spk}(t)$ is set equal to the steady state probability following modulation by an exponential recovery term, where $0 \leq k < 1$. Following from [7], our simulations adopted default RP parameter settings of $k = 0.7$ and $n_r = 9$; a set of follow-up tests considered the use of other values (see "Evaluation of the PSD Correction Methods").

The duration of the synthetic spike trains ($T$) was constrained to multiples of 1024 ms (the segment size used for spectral analysis). The settings used for $T$ and three other free parameters ($p_{base}$, $m$, $f_{osc}$) are described in further detail in "Evaluation of the PSD Correction Methods". The output of each simulation run was a $T$ x 1 binary vector (a "delta vector"), indicating time bins of spike occurrence at a resolution of 1 kHz.

### *Uncorrected PSD*

To obtain the initial, uncorrected PSD estimates, we submitted the delta vectors (demeaned segment-wise) to Welch's method (Matlab *pwelch*, Hamming window length = NFFT = 1024 ms, Fs = 1 kHz, zero segment overlap). PSD functions were represented as 513-element vectors spanning the frequency range [0, 500] Hz, implying a frequency resolution of $\Delta f \approx 0.9766$ Hz.

Within each PSD function, we sought points of statistically significant power in the range (0,100] Hz. Significance was determined using the CI-construction procedure from [7] which was described in the Introduction. For our analyses, points in the (0, 100] Hz range of a spectrum were labeled as significant if they exceeded the upper limit of a one-sided, $[100*(1-\alpha_c)]$% confidence interval formed using the mean and SD over the [250, 500] Hz range of the same spectrum ($\alpha_c$ = Bonferroni-corrected significance level, with respect to the 102 points in the (0,100] range). All figures of uncorrected PSDs were generated with $\alpha_c = .05$.

For the shuffling and residuals methods detailed below, all spectral analysis and statistical testing steps made use of the same routines as were applied to the uncorrected PSDs, and the same default $\alpha_c = .05$ for figure generation; $\alpha_c$ was varied away from this default for the ROC analyses described in "Evaluation of the PSD Correction Methods".

### *Shuffling-corrected PSD*

To obtain the shuffling-corrected PSD for a given spike train, we first generated 100 surrogate, ISI-shuffled versions of the original delta vector (Fig. 2B). We implemented the global shuffling method described by [7] (i.e., random permutation of all ISIs across the entire vector). We chose this approach over local shuffling (permutation of ISIs within nonoverlapping segments) since our simulations emphasized relatively low firing rates (necessitating a broad shuffling scope to admit a sufficient number of ISIs) and did not use time-varying simulation parameters (eliminating the need to capture local variations). Additionally, using the same primary evaluation that we describe for the shuffling and residuals methods (see "Evaluation of the PSD Correction Methods"), we verified superior performance for global as compared to local shuffling. Code sufficient to reproduce this observation is available in the Github repository.

Each of the 100 globally shuffled spike trains was submitted to spectral analysis, and the mean of the resulting PSDs was divided from the original uncorrected PSD, producing the final shuffling-corrected PSD.

### *Residuals-corrected PSD*

The residuals method entails three steps: (1) estimation of the recovery period duration, (2) fitting a point process model (PPM) to the spike train, which estimates the effect of the time elapsed since the most recent spike, up to a history bound corresponding to the estimated RP duration, and (3) applying spectral analysis to the residuals of the PPM fit.

### Estimation of Recovery Period Duration

We define the recovery period as a post-spike reduction in a unit's spike probability, terminating upon return to a steady state firing rate. For our synthetic spike trains (excluding a series of follow-up datasets), the ground truth duration of the recovery period was $n_r$ = 9 ms.

Our estimation approach builds upon a general logic that was described in [27], although note that these authors did not propose an explicit estimation algorithm. We expand upon these authors' ideas by detailing a specific algorithm that operates on the probability density function (PDF) of the inter-spike intervals (ISI). To introduce our approach, we use a simple synthetic spike train for which $m = 0$ (i.e., no oscillation present, such that firing is shaped by only $p_{base}$ and the RP); however, subsequent analyses will show that the algorithm performs reasonably when $m > 0$.

The ISI PDF for an example spike train generated with $m = 0$ is shown in Fig. S1A. Recall that, in the case of Poisson firing (i.e., $n_r = 0$, with an intensity $\lambda = p_{base}$), the observed ISIs are sampled from an exponential PDF [31]: $P(ISI = x) = \lambda e^{-\lambda x}$. Because a spike train with an RP is not Poisson, an exponential curve fit to such a spike train's ISI PDF will overestimate the probability mass assigned to the ISI lags $\leq n_r$. Now, consider what we should observe if we progressively shift the exponential curve fit rightward. Fig S1A illustrates this process for a series of example lags $L$ assigned to the $x = 1$ position of the exponential fit. When $L \leq n_r$, the exponential curve continues to overestimate the probability mass for the lowest lags. However, when $L = n_r + 1$, the exponential model of this left-cropped ISI distribution is appropriate, as firing outside of the RP is Poisson. Our strategy for identifying the boundary between the end of the RP and the start of the steady state relies on this expected transition. We additionally leverage the observation that statistical evidence in favor of the exponential fit (i.e., the extent to which the addition of an exponential curve explains variance beyond that explained by a constant term alone) tends to be stronger when $L = n_r + 1$ than when it is incremented to the immediately neighboring lags to the right.

The estimation algorithm proceeded as follows:

1. Given a delta vector, compute the ISI histogram, with bins spanning the ms lags in [1, max(ISI)].
2. For each $L$ in *lags*:
    a. Extract the histogram values over the range [L, max(ISI)].
    b. Normalize the extracted histogram to unit area to form a PDF.
    c. Fit a single scaled exponential to the PDF using the generalized linear model (GLM) of the form $PDF(x) = e^{B_0} e^{B_1 x} \rightarrow \log\big(PDF(x)\big) = B_0 + B_1 x$. See the next subsection for more detail on this form of GLM.
    d. Compute the deviance difference, which is a standard statistic for assessing the improvement in goodness-of-fit contributed by a full GLM ("model 1", here set to the constant term plus the exponential lag effect), compared to a GLM restricted to the constant $B_0$ term ("model 0"). The difference statistic is $\Delta D = D_0 - D_1 = -2\big[\log \hat{L}_0 - \log \hat{L}_1\big]$; i.e., the log likelihood ratio scaled by -2 [31, 32].
    e. Upon confirming the first instance of a $\Delta D$ value that is a local maximum ($\Delta D(L-1) < \Delta D(L) > \Delta D(L+1)$), break out of the loop iterating over *lags*.
3. The L associated with the local $\Delta D$ maximum marks the start of the steady state region; the estimated recovery period duration, $\hat{n}_r$, is set to L-1 ms.

For the synthetic unit with $m = 0$, this process of finding the lag $L$ associated with the local $\Delta D$ maximum is illustrated in Fig. S1B. Note that in this case, the duration of the recovery period is accurately estimated as L-1 = 9 ms.

Figures 3A-B illustrate the same process as applied to the two synthetic oscillatory spike trains introduced in Fig. 2. Although the presence of non-zero oscillation (i.e., $m > 0$) implies that the steady-state firing is not Poisson, we observe that the curve-fitting algorithm is robust to this departure and again correctly identifies the ground truth RP duration of 9 ms. Evaluation of the RP estimation accuracy over a full dataset of synthetic spike trains is presented in the Results.

### Bounded Last-Spike Point Process Model

Once the RP duration is estimated, the next step is to estimate the RP shape. For this purpose, we fit the spike train with a point process model (PPM).

With reference to [9, 24, 31, 33, 34], we will briefly review the fundamental concepts and terminology for point process models in general, and for our implemented Poisson GLM in particular. A PPM seeks to predict the likelihood that a discrete event (e.g., a spike) will occur at time *t*, given a set of covariates. For a stochastic point process represented in continuous time (e.g., a spike train expressed as a series of spike timestamps), the PPM models the conditional intensity function (CIF):

$$\lambda[t|\theta, H(t)] = \lim_{\Delta t \to 0} \frac{P[N(t+\Delta t) - N(t) = 1 \mid \theta, H(t)]}{\Delta t} \tag{2}$$

where $P[N(t+\Delta t) - N(t) = 1 \mid \theta, H(t)]$ is the conditional probability of a discrete event instance (e.g., a spike) occurring in the time interval $(t, t+\Delta t]$, given a covariate history $H(t)$ and model parameters $\theta$. For our discrete-time spike train representation (the delta vectors), this continuous formulation may be approximated using the assumptions that for $\Delta t \leq 1$ ms, $P$(spike in $(t, t+\Delta t]$) ≈ $\lambda[\theta, H(t)]\Delta t$. For the remainder of this text, as a shorthand, we will follow the convention of using *t* to index the ms-width bins of the discrete-time representation (as opposed to continuous time).

Given this discrete-time representation, a Poisson GLM (i.e., a generalized linear model with a log link function, also known as Poisson regression) is one example of a PPM that may be used to predict spike probability. We implemented the bounded last-spike model as a Poisson GLM of the form

$$log\lambda(\mathrm{t}|\mathrm{H_t}, \beta) = \beta_0 + \sum_{j=1}^{\hat{n}_r} \beta_j \, \delta_{t-j,s_*(t)} \tag{3}$$

where *s*(t)* denotes the time of the most recent spike [33], and $\delta_{t-j,s_*(t)}$ is the Kronecker delta function, returning 1 when *t-j* equals *s*(t)*, and 0 otherwise. In other words, the model estimates the variance in the logarithm of spike intensity that is explained by either the constant term alone (outside of the RP) or the constant plus the weighted contribution of the millisecond interval transpired since the last spike. Consequently, inside the RP, λ is modeled as the multiplicative contribution of the exponentiated constant and last-spike factors ($e^{\beta_0} e^{\beta_j}$, for $j = t - s_*(t)$), and as $e^{\beta_0}$ outside of the RP.

The model was fit to the spike train data using Matlab *fitglm,* which applies the conventional iteratively re-weighted least squares (IRLS) method to seek the log likelihood-maximizing parameters. Note that the log likelihood function for a model such as (3) is known to be concave [9], thereby implying that if a maximum exists, it must be unique.

### Spectral Analysis of Model Residuals

Given a set of parameter estimates $\hat{\beta}$, the raw residuals of the modeled spike train (the observed - fitted values) are calculated as

$$r(t) = I(t) - e^{\left[\hat{\beta}_0 + \sum_{j=1}^{\hat{n}_r} \hat{\beta}_j \, \delta_{t-j,s*(t)}\right]} \tag{4}$$

where *I(t)* represents the binary spike train input to the PPM. The final step of the residuals method is submission of *r(t)* to spectral analysis. Note that (4) only defines *r(t)* for $t \geq \hat{n}_r + 1$ ms. To maintain alignment with the original time series (and the divisibility by the 1024 ms segment size), the initial 1024 ms segment submitted to *pwelch* consisted of the initial 1024-$\hat{n}_r$ ms of computed residuals, demeaned and prepended with zero-padding of $\hat{n}_r$ ms.

### Clarifications and Relationship to Pre-existing PPM Approaches

We conclude this description of the residuals method with two points concerning the "bounded last-spike" PPM in particular. The first point concerns the use of the term "bounded": Here, we intend to emphasize the bounded, $\hat{n}_r$-millisecond historical window within which the most recent spike must fall, to be eligible to contribute to the spike intensity prediction at time $t$. We make this point to prevent confusion with another common usage of "bounded": For a discrete-time point process, the count of events that may occur in each sampled time bin is expected to be bounded in [0,1], thereby satisfying the conditions for "orderliness" [35]. Although our model was not named for this type of boundedness, we note that the orderliness expectation is met by the synthetic time series we are analyzing (given the constraints of our simulation framework) and is very likely to be met by our sample of empirically-acquired spike trains (given the typical assumption of a 1 ms lower bound on the neuronal refractory period [31]).

As a second point, we note that the bounded last-spike PPM draws upon multiple features of pre-existing PPMs in the spike modeling literature, although it is also important to clarify how our current implementation differs from these earlier approaches. For this purpose, two categories of pre-existing PPMs are relevant. The models in both categories aimed to account for spike activity relatively comprehensively (in contrast to our model, which aims to selectively capture the RP effect).

A principal representative of the first category is the Inhomogeneous Markov Interval (IMI) model of [33]. The IMI model predicts spike intensity as a function of the time elapsed since the most recent spike (*t-s*(t)*, with no bound imposed) as further modulated by a function the current clock time (*t*); the two functions' forms are estimated with regression splines. Therefore, our current model shares the IMI model's incorporation of a *t-s*(t)* term, but differs in its exclusion of the clock time term, and also of the *t-s*(t)* information when *t-s*(t)* > $\hat{n}_r$. Our PPM also differs in its modeling of the *t-s*(t)* effect with a series of delta functions, as opposed to the splines used by [33]. We did test a version of the residuals method that replaced the delta functions with cubic splines (specifically, the modified cardinal splines from [34]), and that adjusted the number and location of the knots according to rules that depended on the magnitude of $\hat{n}_r$. The spline-based model was found to perform slightly worse than the delta function version, as demonstrated through the same evaluation procedures as were used to compare global and local shuffling. See the Github repository for code sufficient to reproduce this result.

The second category consists of autoregressive PPMs which predict spike intensity as a function of the spike counts recorded for each of a series of historical time bins (e.g., [25, 26]). Specifically, these models consider a long historical time interval (up to *t* - 150 ms in [25]), divide this interval into time bins of interest, and use the counts of spikes recorded in each bin as individual predictor variables in the PPM. Typically, a subset of narrow (e.g., 1 ms) bins positioned over the most recent historical timepoints (e.g,. -10 through -1 ms) are assumed to capture the combined effects of both the recovery period and any other short-lag history effects (e.g., bursting). Our present approach draws upon these PPMs' use of indicator functions to model history effects limited within bounded historical intervals, but differs in its exclusive focus on the most recent $\hat{n}_r$ millseconds, and on the timing of the most recent spike that occurred within that historical window.

### *Evaluation of the PSD Correction Methods*

Comparison of the residuals and shuffling methods entailed (1) simulation of synthetic spike trains over a range of physiologically-relevant parameters, (2) generation of the two methods' corrected PSDs, with tabulation of hit and false alarm rates for these PSDs across a range of $\alpha_c$ levels, and (3) comparison of the methods' classification of true and false oscillations, by way of ROC (Receiver Operating Characteristic) analysis (collapsed across parameter settings) and regression analysis of the residuals-shuffling differences in the hit and false alarm rates (to examine parameter effects).

## Synthetic Spike Train Test Sets

Method evaluation was first carried out on a primary synthetic test set; subsequently, we generated a number of additional test sets to address follow-up questions. For each test set, and for each of the unique combinations resulting from the crossing of the varied simulation parameters, we randomly generated one hundred spike trains.

For the primary test set, 100 spike trains were simulated for each of the 540 unique combinations that resulted from crossing the following four varied parameters:

(1) Spike train duration ($T$): $\{30,60,120\} \times 1024$ ms windows
(2) Oscillation frequency ($f_{osc}$): $\{7,9,12,20,32\}$ Hz
(3) Base firing rate offset ($p_{base_offset}$, such that $p_{base} = f_{osc} + p_{base_offset}$): $+2^{[0:1:5]}$ Hz
(4) Modulation index ($m = p_{osc}/p_{base}$): [0:0.2:1]

These parameters emphasize relatively low duration and firing rate settings, with the short durations consistent with the limitations that can be posed by real experimental data, and the low FRs in line with the motivation for developing the residuals method. At the same time, the setting of $p_{\text{base}}$ as a positive offset from $f_{\text{osc}}$ ensured that, in most cases, the synthetic unit averaged > 1 spike per oscillation cycle, although the recovery period implied that the average dropped to slightly < 1 in some of the $f_{osc}$ + 1 cases. For $m$, note that the inclusion of 0 allowed for tracking of false positive detections of significant power when no oscillatory effect was introduced into the simulation.

Three kinds of variants on this primary dataset were created to address follow-up questions. First, we generated one dataset in which $p_{base}$ and $p_{base_offset}$ were directly manipulated, leaving $f_{osc} = p_{base} - p_{base_offset}$ as the indirectly manipulated variable. This dataset allowed for tests to determine whether the methods' relative performance was significantly influenced by the magnitude of the $p_{base} - f_{osc}$ difference, when controlling for $p_{base}$ itself. For this dataset, the simulations used the same $T$ and $m$ settings as were used for the primary dataset. We crossed these factors with two levels of $p_{base}$ (15 Hz, 20 Hz) and 4 levels of $p_{base_offset}$ ([2:2:8] Hz), resulting in 144 unique parameter combinations.

The second and third variations involved the use of higher FRs, and alternatives to the default RP parameters, respectively. For the high-FR simulations, the set of sampled $p_{base_offset}$ values was changed to [35:10:85] Hz; these were added to the $f_{osc}$ values used for the primary dataset ((2) in the list above) to determine the $p_{base}$ settings. For the altered-RP simulations, we tested both shorter and longer RP durations ($n_r$) and varying exponential steepness terms ($k$); full details are described in the Results.

## Labeling of Hits and False Alarms

For each of the 540 x 100 (or 144 x 100) synthetic spike trains in a test set, the shuffling and residuals methods were both applied to produce corrected PSDs. Given a corrected PSD and an $\alpha_c$ setting, statistical testing returned a binary vector indicating whether oscillatory power exceeded the CI bound for each of the 102 frequency bins of width $\Delta f$ = 0.9766 Hz in (0,100] Hz. Frequency bins are labeled by their lower bound (0.9766, 1.9531, 2.9297…). Let **f** denote the frequency labels and s(f) denote the binary significance output. We classified a PSD as containing a "hit" (i.e., a true positive detection of oscillatory power) if (1) m > 0 and (2) any s(f) = 1 for those labels $f_i \in$ **f** that were the 3 nearest neighbors of $f_{\text{osc}}$ (e.g., $f_i \in \{10.7422,11.7188,12.6953\}$ for $f_{\text{osc}}$ = 12 Hz). We classified a PSD as containing a false alarm (FA) if either of two cases held true: (1) $m$ = 0, s(f) = 1 for any label $f_i \in$ **f**, or (2) $m$ > 0, s(f) = 1 for $f_i \in |f_i - f_{\text{osc}}| > 5$ Hz.

Note that the output of this classification procedure was a pair of binary labels for each PSD, indicating whether each contained ≥ 1 hit, and ≥ 1 FA, respectively. For the ROC analysis described below, counts of hits and FAs refer to the counts of these PSD-level labels (as opposed to the total counts of individual hit- or FA-classified frequency bins).

### Comparison of ROC curves for the Correction Methods

ROC analysis of the shuffling and residuals methods consisted of three steps: (1) random selection of subsamples of the PSDs, to be used for a bootstrapping procedure for statistical testing, (2) calculation of partial ROC curves and corresponding partial area under the curve (pAUC) values for the subsampled PSDs, and (3) statistical comparison of the pAUC values yielded by the shuffling and residuals methods.

The subsampling procedure was iterated 1000 times. For each subsampling iteration, we randomly sampled, without replacement, 20 of the 100 synthetic spike trains generated within each of the 540 (or 144) parameter combinations. The full resulting subsample of 540 x 20 = 10800 spike trains (or 144 x 20 = 2880 spike trains), and the corresponding hit and FA classifications for the shuffling- and residuals-corrected PSDs, were used to generate separate shuffling and residuals ROC curves for that iteration.

For a given subsampling iteration and correction method, we formed an ROC curve as the function relating hits to FAs over a range of $\alpha_c$ values. The full set of $\alpha_c$ values consisted of the sorted vector of the values 1 and 5 multiplied by a series of negative powers of 10 ($[1\ 5]^T * [10^{[-8:1:-1]}]$) and 1. The curves were formed by collapsing across the simulation parameter combinations, such that a point on an ROC curve represented the fraction of the 10800 corrected PSDs that were associated with ≥ 1 hit and ≥ 1 FA, respectively. (Note that the inclusion of the $m$ = 0 case implies that the hit rate would fall short of an upper asymptote of 1 for an ideal observer.)

For each analyzed dataset, the above procedure resulted in 1000 pairs of shuffling- and residuals-generated ROC curves. As the ranges of spanned hit and FA rates varied slightly across the individual ROC curves, we analyzed the classification performance represented by these curves with a partial area under the curve (pAUC) approach [36]. Specifically, the raw curves were cropped (using linear interpolation as needed) to span the intersection of all 2000 curves' FA ranges. For each cropped curve, we used the trapezoidal rule (Matlab *trapz*) to compute the area under this bounded curve region. Due to the bootstrapping process, the paired differences in the shuffling and residuals methods' pAUC estimates ($D_{pAUC}(i) = pAUC_{res}(i) – pAUC_{shuf}(i)$) were assumed to approximate a normal distribution; therefore, a two-tailed paired *t*-test was used to assess the null hypothesis of $D_{pAUC} = 0$.

### Analysis of Simulation Parameter Effects on Relative Hit and False Alarm Rates

To examine the simulation parameter effects, we focused on a fixed $\alpha_c$ = .05 threshold, and applied regression analyses to the residuals-shuffling differences in HRs ($D_{HR}$) and FAs ($D_{FA}$) separately. To avoid specific distributional assumptions regarding $D_{HR}$ and $D_{FA}$, we conducted these analyses using the same 1000 subsamples as were used for the pROC tests. We considered all simulation parameter combinations, with the exception of those for which $m$ = 0 (to omit cases when $D_{HR}$ = 0 - 0 by definition, and with this same exception applied to the $D_{FA}$ analyses for consistency). Consequently, 450 x 1000 (or 120 x 1000) unique $D_{HR}$ and $D_{FA}$ estimates, respectively, contributed to the two analyses, for each dataset on which they were performed.

Separate general linear models for the two dependent variables (DVs), $D_{HR}$ and $D_{FA}$, were constructed using five independent variables (IVs): the mean-centered equivalents of the four directly manipulated simulation parameters ($T$, $f_{osc}$, $p_{base_offset}$, and $m$ for most datasets, with $f_{osc}$ replaced with $p_{base}$ when $p_{base}$ was directly manipulated), and also the square of the mean-centered modulation strength parameter ($(m-\overline{m})^2$), given trends in the data visualizations that were suggestive of quadratic effects of the $m$ factor (Fig. 4A-B, Fig. S2). For each DV, the full GLM consisted of the intercept term plus the sum of the main effect, 2-way interaction, and 3-way interaction terms formed from the five IVs.

### Post-Hoc Analyses of False Alarms in Spike Trains with High Firing Rates

In addition to the above ROC and regression analyses, we ran an additional set of post-hoc analyses to the synthetic dataset that had been generated with high FR settings (see "Synthetic Spike Train Test Sets" above). These follow-up assessments were motivated by the observation of unexpectedly high FA rates, especially for spike trains generated with long $T$ and high $m$, and either low $f_{osc}$ (residuals method)

or high $f_{osc}$ (shuffling method; see Fig. S8-S10). We report three analyses, all of which summarized over the 1000 spike train subsamples, and focused on the $T$ = 120 x 1024 ms and $m$ = 1.0 conditions.

The first analysis was a linear regression consisting of two IVs ($f_{osc}$, $p_{base_offset}$) and the DV (the subsampled $D_{FA}$ scores). Consequently, 30 (= number of $f_{osc}$ x $p_{base_offset}$ combinations) x 1000 (= $n$ subsamples) unique $D_{FA}$ estimates contributed to this analysis. The regression model included both main effects and the $f_{osc}$ x $p_{base_offset}$ interaction term.

The second and third analyses entailed the calculation of descriptive statistics. These statistics were computed separately for the two correction methods. For one set of descriptive statistics, we report the mean absolute FA rate for each of the two methods and the two most extreme $f_{osc}$ cases (7 Hz, 32 Hz); this value simply corresponded to the grand average of the FA rates, collapsed over the 1000 subsamples and six $p_{base_offset}$ settings. For the other set, we report summaries of the peak power (and associated frequencies) for the PSDs' false alarm points, expressed as a ratio over the peak power for the hit points. For each correction method, we iterated over the 1000 subsamples of spike trains, and for each subsample, over the 30 unique $f_{osc}$ x $p_{base_offset}$ setting combinations. Within each subsample, each unique $f_{osc}$ x $p_{base_offset}$ setting combination is represented by 20 spike trains (given the constraint that $T$ = 120 x 1024 ms and $m$ = 1.0). For the correction method under consideration, we identified the FA and hit points present within the 20 respective corrected PSDs. For each PSD, the peak (i.e., maximum) power estimates were identified over any identified FA and hit points, respectively. For a PSD that included both FA and hit points (as was common in the $T$ = 120 x 1024 ms, $m$ = 1.0 cases), we computed the $peak_{FA}$ / $peak_{Hit}$ power ratio. These peak ratios and the associated peak FA frequencies were then averaged over the PSDs (for which they were defined), and then averaged over the 30 parameter setting combinations and 1000 subsamples.

## Experimental Data Methods

### *Subject*

We analyzed data from a single rhesus monkey (monkey G, 7.1 kg, female) which was one of two animals studied in an investigation of MPTP (1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine)-induced parkinsonism. Here, we selectively analyzed data from the post-MPTP state. A prior report based on the pre-MPTP data [30] provides detailed explanations of several components of the experimental protocol that are also applicable to the post-MPTP dataset; for these components, we briefly recap the main points below.

### *Task*

Single-unit extracellular signals were collected during performance of a visually-cued reaching task. on which monkey G had been trained to a high degree of proficiency. Food rewards were delivered for successfully performed trials. Full task details are available in [30]. Here, the most relevant task component is the pre-reach delay period that spanned a variable duration interval (1.01 - 3.50 s for the trials included in our analyses) from trial start to go-cue onset. During this delay period, the left hand rested at a start position (a metal bar) fixed at waist height to the animal's left. The onsets/offsets of hand contact with this start position were detected with an infrared sensor.

### *Surgery*

Devices were implanted surgically to allow access to the right thalamus and globus pallidus via a parasagittal approach. The surgery was performed under sterile conditions with ketamine induction followed by Isoflurane anesthesia. Vital signs (i.e., pulse rate, blood pressure, respiration, end-tidal pCO2, and EKG) were monitored throughout the surgery to ensure a proper level of anesthesia. Analgesics and antibiotics were administered prophylactically.

On protocols.io, we have posted two step-by-step protocols relevant to the surgical procedures (craniotomy https://www.protocols.io/private/8791490F352811EEAFC70A58A9FEAC02 and head

fixation post and recording chamber implantation https://www.protocols.io/private/A317A58E352711EEAFC70A58A9FEAC02).

### *MPTP Administration*

The animal was rendered hemiparkinsonian by injection of MPTP into the right internal carotid artery (0.5-mg/kg, [37]). This model of parkinsonism was chosen to facilitate care of the animal and to increase the likelihood that the animal would perform the behavioral task following intoxication [38]. The MPTP administration procedure was performed under general anesthesia (1-3% Isoflurane) and prophylactic antibiotics and analgesics were administered post-surgically. The animal developed stable signs of parkinsonism contralateral to the infusion (i.e., on the left side of the body).

On protocols.io, we have posted a protocol that includes a description of the procedures for internal carotid artery administration of MPTP (https://www.protocols.io/private/7BED34B9490A11EEACE60A58A9FEAC02).

### *Data Acquisition and Spike Sorting*

Regions of GPi and VLa to be targeted for single-unit recording were located using standard neurophysiologic mapping methods [30]. The extracellular spiking activity of neurons in GPi and VLa was collected using multiple glass-insulated tungsten microelectrodes (0.5–1.5MΩ, Alpha Omega Co.) or 16-contact linear probes (0.5–1.0MΩ, V-probe, Plexon Inc.). Data were amplified (4×, 2Hz–7.5kHz), digitized at 24 kHz (16-bit resolution; Tucker Davis Technologies), and saved to disk as continuous data. The stored neuronal data were high-pass filtered (Fpass: 300Hz, Matlab *firpm*) and thresholded, and candidate action potentials were sorted into clusters in principal components space (Off-line Sorter, Plexon Inc.; RRID:SCR_000012). Clusters were accepted as well-isolated single-units only if the unit's action potentials were of a consistent shape and could be separated reliably from the waveforms of other neurons as well as from background noise throughout the period of recording. Times of spike occurrence were saved at millisecond accuracy.

On protocols.io, we have posted protocols relevant to electrophysiological recording (https://www.protocols.io/private/7CF970AD3BE111EE882A0A58A9FEAC02) and recording chamber maintenance (https://www.protocols.io/private/F56F59C63BDD11EE882A0A58A9FEAC02).

### *Task-Aligned Analysis Windows*

To focus on intervals of the monkey G spike trains when oscillatory activity was expected to be high, we restricted our analyses to task windows spanning the first second of each trial. These windows sampled data from the pre-reach delay periods; such intervals of enforced hand stillness have been historically associated with elevated beta power [39]. The 1000 ms window was chosen instead of the PSD segment size (1024 ms) because a small fraction of the go-cue onsets occurred at post-start latencies in (1000, 1024] ms. This selection step jointly served a second purpose of reducing potential sources of substantial non-stationarity in the analyzed spike data (along with a firing rate screening step, described further below).

### *Unit and Trial Selection Criteria*

The process of selecting units and trials to include in analysis proceeded in three stages. As a first step, we excluded units recorded from the GPi and VLa when these failed to meet basic quality checks (e.g., poor isolation during spike sorting, or highly anomalous spike waveform features) or exhibited outlier FRs (as measured over the full recording period) of > 106.48 Hz (GPi) or > 39.50 Hz (VLa). Units with FRs < 1 Hz were also excluded.

A second, trial selection step was prompted by the observation of units for which the "hold" (i.e., pre-reach delay) period-windowed data exhibited visible inhomogeneity in firing rates over trials, often appearing as blocks of trials during which spiking was either absent or very sparse, preceded or followed

by blocks for which spiking was more regular. As this pattern was suggestive of considerable non-stationarity remaining in the hold period-restricted data, we implemented a procedure to detect and remove windows associated with trials with aberrant firing rates. This procedure itself entailed multiple steps:

1. Prepare an *nW* x 1 vector **S** indicating the number of spikes in each 1000 s window.
2. Remove those rows from **S** reporting < 2 spikes.
3. Use a Gaussian Mixture Model (GMM) to cluster the remaining entries of **S**, using a forward search procedure to set the number of Gaussians (*k*) parameter:
    a. For each **k** in 1 to length(**S**):
        i. Fit *k* Gaussians to the data in **S** (Matlab *fitgmdist*) using initial centers ($\mu$), variance ($\sigma^2$), and mixing proportion (*p*) settings as determined by the Variance Partitioning method [40].
        ii. Record the Bayesian Information Criterion (BIC) value for this *k*. The BIC is a negative log-likelihood-based measure that is penalized by the magnitude of *k*; lower BIC values are preferable.
        iii. Break out of the loop when either:
            1. BIC(k) > BIC(k-1)
            2. A failure to converge is reported for iteration *k*.
    b. Upon exiting the loop, return the GMM with the number of Gaussians set to *k-1*.
4. Assign each row in **S** to one of the *k* Gaussians (hard clustering with Matlab *cluster*, which assigns the cluster labels that maximize the probability of observing the data, given the GMM parameters).
5. Find the Gaussian **G** with the maximum number of assigned windows. Retain only those windows from **S** that were assigned to **G**.
6. From the windows assigned to **G,** remove those windows for which the number of spikes *nSpk* is an outlier, given the mean and standard deviation parameters estimated for **G**. A window is defined as an outlier when $|nSpk-\mu_g| > 3 * \sigma_g$).

For each unit, the output of this trial (i.e., window) selection step was assumed to represent a set of 1000 ms segments for which the firing rates were relatively homogeneous. As a final step, we discarded those units for which < 30 windows survived the trial selection procedure, or for which the mean firing rate over the surviving windows was < 1 Hz.

### *Windowed Data: Uncorrected PSD*

Note that Welch's method processes time series segments independently (i.e., the final estimated PSDs are the average of the individual PSD estimates for the Hamming-windowed 1024 ms data segments). Prior to calling *pwelch*, we prepared the hold period data by demeaning and adding 24 ms of zero padding to each 1 s hold period vector, and then concatenating these 1 s vectors into a single time series. This vector was then submitted to identical PSD analysis procedures as were described for the synthetic data.

### *Windowed Data: Shuffling-corrected PSD*

Similar to the synthetic dataset, many of the recorded units exhibited low FRs, such that local shuffling was not a practical option. We extended the global shuffling option from [7] (which was only defined for unbroken time series) to temporally separated data segments with the following algorithm (all indices assume indexing starts at 1):

1. Input: **dMat** ("delta matrix"), an *nD* x *T* matrix, consisting of the original *nD* delta vectors of length *T* msec (here, *T* = 1000; *nD* = number of trials).

2. Create a list, **sISI**, consisting of the concatenation of all ISIs encountered in **dMat**, sorted in descending order of size. (Note that **sISI** will effectively function like a max-oriented priority queue.)
3. Initialize an *nD* x *T* matrix **wMat** ("working matrix") with zero entries.
4. Initialize an *nD* x 1 vector **timeLeft** with entries of the value *T*-1 (here, 999), corresponding to the size of the largest ISI that might be inserted into each row of **wMat**. (Note that 1 ms must be reserved for insertion of the first spike for the first ISI).
5. While **sISI** is not empty:
    a. **isi** ← top (max duration) entry, removed from **sISI**
    b. **openRows** ← indices of rows in which **timeLeft** ≥ **isi**
    c. **targetRow** ← randomly chosen row from **openRows**
    d. Insert **isi** into **wMat[targetRow]**:
        i. If this is the first ISI entered into targetRow, place "spikes" (1 entries) at **wMat[targetRow][1]** and **wMat[targetRow][1+isi]**
        ii. Otherwise, place a spike at **wMat[targetRow][i+isi]**, where **i** is the index of the rightmost spike present in the row
    e. **timeLeft[targetRow] ← timeLeft[targetRow] - isi**

6. For each row in **wMat:**
    a. If > 2 spikes are present in a row, randomly re-sequence the ISIs (as in the standard shuffling procedure)
    b. If **timeLeft[row]** > 0: Right shift the spikes (insert a leading series of 0s) by a random integer drawn from [0,**timeLeft[row]**]
7. **shufMat** ("shuffled matrix") ← **wMat**

To summarize, the algorithm implements global ISI shuffling by (1) initializing an empty data matrix, and a pool of ISIs to be inserted into it, (2) randomly inserting ISIs into the new matrix, proceeding from the longest to shortest ISIs, and packing the rows in a left-aligned fashion, and (3) randomly re-sequencing and right-shifting the ISIs in each row, to eliminate the structure introduced by step (2).

Each matrix of shuffled spikes was submitted to PSD analysis through an identical manner as was described for the uncorrected PSDs, and the standard division of the original PSD by the mean of the shuffling-generated PSDs was then used to produce the corrected PSD.

### *Windowed Data: Residuals-corrected PSD*

Recall that, in the case of an unbroken time series, the residuals output, *r(t)*, is not defined for the first $\hat{n}_r$ ms. In the case of the windowed data, the initial $\hat{n}_r$ ms must be dropped for the residuals output corresponding to each trial. Therefore, the delta vectors corresponding to each individual trial were zero-padded by 24+$\hat{n}_r$ ms, to both account for the initial RP-dependent lag, and to match the expected 1024 ms segment size. All other aspects of the residuals method remain as described in the context of the unbroken synthetic data.

### *Reference Synthetic Data for Comparison with Bursting Units*

We conducted follow-up examination of two VLa spike trains, which exhibited spiking timing patterns and corrected PSDs that were distinct from those generally observed in our synthetic datasets. To better characterize the distinctive features of these spike trains, which included suspected burst-firing behavior, we generated additional sets of reference spike trains, which did not include a bursting component, but did roughly match the two VLa units with respect to other key parameters. All simulations used the same framework as described above. For the synthetic data matched to the first VLa unit, $T$ = 118 x 1024 ms, $f_{osc}$ = 13 Hz, and $p_{base}$ = 15 Hz. For the simulations matched to the second

VLa unit, $T$ = 51 x 1024 ms, $f_{osc}$ = 14 Hz, and $p_{base}$ = 21 Hz. For both sets of simulations, $m$ = 0.6, $n_r$ = 1, and $k$ = 0, and $n$ = 100 synthetic spike trains were generated.

### *Population-Level Comparisons of Oscillation Detection Rates*

We used three analyses to compare the counts of significant oscillations reported by the two methods, for specific frequency bands of interest. All three analyses were performed over the pooled GPi and VLa data. The first analysis compared the proportions of PSDs for which the residuals and shuffling methods reported at least one point of significant power within the 8-30 Hz alpha-beta band. Proportions were compared with McNemar's $\chi^2$ test, with continuity correction (R *mcnemar.test*, R Core Team, 2021; RRID:SCR_001905). The second analysis focused specifically on those PSDs for which at least one method reported an alpha-beta oscillation (as defined above), and used logistic regression (Matlab *glmfit*, logit link) to ask whether the corresponding unit's mean FR inversely predicted residuals-specific identification of alpha-beta incidence. The dependent measure was a binary coding of the PSDs (1 = alpha-beta detected by residuals method alone; 0 = detection by either both methods or shuffling only). The third analysis used McNemar's $\chi^2$ to compare the proportions of PSDs for which the residuals and shuffling methods reported at least two consecutive points of significant power in the (0,4] Hz delta range. This delta band measure was intended to serve as a substitute for more traditional forms of 1/f-like trend assessment (which involve curve fits that do not accommodate possible RP distortion [3]).

# Results

## Results Overview

In the following, we first consider the shuffling and residuals methods in the context of synthetic spike trains. We will highlight two example cases that aid in illustrating the procedures and outputs of the two correction methods, and then present the results of a systematic comparison of the methods' performance over simulations that encompass diverse parameter settings. Subsequently, we present the shuffling- and residuals-corrected output for the empirical single-unit data collected from the parkinsonian monkey, and highlight cases in which the divergence of the two methods' output is particularly notable.

## Synthetic Data Results

### *Shuffling Method: Examples of Detected and Non-Detected Oscillations*

Stochastic spike trains were generated according to the framework described in [2, 7]; full details are provided in the Materials and Methods. To briefly recap, spike arrivals were sampled from 1 kHz rate functions, $p_{\mathrm{spk}}(t)$, which could be varied with respect to the parameters that governed the exponentially rising recovery period (steepness $k$; duration $n_r$), the oscillation (frequency $f_{osc}$; modulation strength $m$), the baseline FR ($p_{base}$), and the simulation duration ($T$). We considered a variety of parameter settings, although with an initial priority on relatively low $p_{base}$ values, given the challenges posed by cases in which oscillation detection may be particularly reliant on the information present in the ISI distribution.

Figure 2 presents two illustrative cases generated from these simulation procedures (spike train duration $T$ = 60 $x$ 1024 ms, $f_{\mathrm{osc}}$ = 9 Hz; for all synthetic data, $n_r$ = 9 and $k$ = 0.7, unless otherwise specified). The two cases are similar in that standard statistical assessment of the uncorrected spike trains fails to detect the underlying 9 Hz oscillatory component, but differ in the factors that likely contribute to this failure, and in the shuffling method's ability to recover this spectral feature.

Figure 2A shows the uncorrected spectra. As compared to the second (Fig 2A *right*) example, the first spike train was generated with a relatively high base FR (41 Hz), but with a somewhat low amplitude modulation factor ($m$ = 40%). As noted in the Introduction, high FRs lead to more dramatic RP-related distortions of the associated spectra [7]. Given these combined limitations (low $m$, greater distortion), the 9 Hz peak fails to cross the upper bound of the 95% corrected CI. The second spike train was

generated with a stronger modulation setting ($m$ = 60%), and a lower base FR (11 Hz), which results in less dramatic RP distortion; however, these advantages are offset by the reduced oscillation signal-to-noise ratio (SNR) that is also a known consequence of low FRs [2].

To both of these spike trains, we applied the standard global shuffling procedure (Fig 2B), yielding estimates of the renewal-equivalent power spectra (Fig 2A, red lines) and the shuffling-corrected PSDs. For the high FR, low $m$ example, the mean shuffled data PSD follows the RP distortion trend while avoiding the implanted oscillation. Consequently, division by this shuffling-generated PSD achieves its intended outcome: The corrected spectrum (Fig 2C *left*) is generally flat, apart from the 9 Hz peak, which does surpass the upper frequency-derived CI threshold.

However, such successful peak recovery is not observed for the low FR, high $m$ example (Fig 2C *right*). In this case, the shuffled PSD appears to capture both the distortion trend, and, to a modest extent, the 9 Hz peak. Correction reduces the relative height of this spectral feature, thereby compounding the challenges to detecting the spectral peak that were already present due to the low SNR. Even with the RP distortion removed, the 9 Hz peak falls just short of the significance threshold.

Figure 2D, which plots the ISI PDFs for the two cases, provides a time domain perspective on these different shuffling outcomes, by highlighting the spike timing information that shuffling correction removes. In the higher FR, lower $m$ case (Fig 2D *left*), > 50% of the ISIs were ≤ 23 ms, thereby limiting the opportunity for the somewhat modest amplitude, 9 Hz oscillatory trend (period ≈ 111.11 ms) to substantially influence the shape of the ISI distribution. Qualitatively, the PDF resembles the form expected for a Poisson process modulated by a recovery period alone (i.e., an initial dip followed by an exponential distribution). For the lower FR, higher $m$ case (Fig 2D *right*), 50% of the ISIs exceeded 73 ms (≈65.7% of the 9 Hz cycle length), and > 31% were 112 ms or longer (i.e., the spiking skipped entire cycles). These long survival times afforded greater opportunity for the moderately strong oscillation term to more visibly shape the ISI distribution, with relative peaks and valleys appearing in rough correspondence to that expected for the 9 Hz rhythm.

To summarize, the lower FR/higher $m$ example illustrates an instance relevant to the general caveat raised by [7]: If much of the rhythmicity of interest is present in the ISI distribution, then caution is warranted when applying the shuffling method (or the analytical alternative), as such correction will remove any oscillatory information present in the first-order ISI statistics. What our example here helps to clarify is that such scenarios emerge not only from precise pacemaker-like patterns, but also in spike trains in which a known sinusoidal drive is expressed through relatively sparse spiking. This observation motivated our interest in examining how the residuals method, which models and removes spike history effects restricted to a short history window, might perform in such scenarios.

### *Residuals Method: Procedure and Example Case Performance*

Figure 3 illustrates the steps of the residuals method, and the outcomes for the two example synthetic spike trains introduced above. To reiterate, the main steps include (1) estimation of the RP duration, (2) fitting the spike train with the bounded last-spike point process model (PPM), with the bound set to the RP duration, and (3) applying spectral analysis to the residuals of the fitted PPM.

#### RP Estimation Algorithm

The RP duration estimation method is illustrated in Fig. 3A-B (for the two example cases) and also Fig. S1 (for a spike train with no oscillation). To recap (see Materials and Methods for full details), we sought to translate the general strategy outlined by [27] into a specific algorithm that might perform with reasonable effectiveness on limited, sparse spiking data.

We specifically implemented a procedure that compares the variance explained by a series of right-shifted exponential curves fit to the ISI PDF. We pursued this approach after observing, on cases such as the Fig. 2-3 examples, that the curve goodness-of-fit often reaches its first local maximum when the

exponential is anchored to the first post-RP time bin (i.e., when a unit with *m* = 0% should first exhibit ISIs consistent with Poisson spiking), even when noise and oscillatory trends create visible departures of the post-RP region from a clean exponential form. In Fig. 3A we highlight select iterations of the right-shifted curve fits, for the two examples. When the curves are left-anchored to the first possible start lag (1 ms), one observes visibly poor fits over the initial, RP-spanning time bins. As shown in Fig. 3B, these mismatches correspond to relatively low values for our goodness of fit metric (the deviance difference, see Materials and Methods). Iterating through the subsequent RP-containing bins, the fit measure gradually improves, reaching local peaks (in both examples) at the first post-RP time bin of 10 ms. After detecting subsequent decreases in the exponential term's explanatory power at 11 ms, we confirm the presence of initial local maxima, and return, for both spike trains, RP duration estimates ($\hat{n}_r$) that match the ground truth $n_r$ of 9 ms.

Note that more thorough evaluation of RP estimation performance is provided in the Methods Evaluation section, where we report on the systematic comparison of the shuffling and residuals methods.

### Bounded Last-Spike Point Process Model

The next step of the residuals method entails the estimation of the shape of the RP (i.e., the $t\text{-}s_*(t)$ effect; see Materials and Methods) bounded within a historical interval of $\hat{n}_r$ millseconds. Fig. 3C illustrates the inputs and outputs of our current estimation procedure on the two example cases and with $\hat{n}_r$ = 9 ms. As in global shuffling, the input (leftmost panels, zoomed in on the first 209 ms) was the entire binary spike train vector.

As stated in the Introduction, we chose to use the point process modeling (PPM) framework to estimate the $t\text{-}s_*(t)$ effect. In general, PPMs may operate on binary incidence vectors to estimate discrete-time approximations of conditional intensity functions (CIFs; see [9, 31] for additional background). When PPMs are used to predict future spikes on the basis of spike history, the CIF ($\lambda[t|\theta, H(t)]$) represents the time varying probability that a spike will occur in time bin $t$, given the spiking history $H(t)$ and the model parameters (Θ). PPMs can be implemented in a variety of frameworks; we chose the well-established route of utilizing the Poisson Generalized Linear Model framework (i.e., linear regression with a log link function). The GLM took on the form $log\lambda(\mathrm{t}|\mathrm{H_t}, \beta) = \beta_0 + \sum_{j=1}^{\hat{n}_r} \beta_j\, \delta_{t-j,s_*(t)}$, where $\beta_0$ is the intercept term, the $\delta_{t-j,s_*(t)}$ terms indicate whether the most recent spike fell within the time bin corresponding to *j* ms in the past, and the $\beta_j$ weight the additive contributions of these last-spike lags to the predicted logλ(t). Once the vector $\hat{\beta}$ of coefficient estimates is found, the linear expression is exponentiated to obtain the fit of the model to the original spike train (excluding the first $\hat{n}_r$ time bins for which no fit is defined). Due to the exponentiation step, the constant and the last-spike factors combine multiplicatively in the estimation of λ(t) [9].

Qualitatively, the GLM fits to the two example spike trains (Fig. 3C, center panels) do appear to approximate the forms anticipated on the basis of the ground truth RP and $p_{base}$ parameters. When no spikes occurred in the prior $\hat{n}_r$ ms, the predicted λ(t) were approximately 0.041 and 0.011 for examples #1 and #2, respectively. When spikes did occur, the predicted λ(t) approximated the gradually rising RP trend.

### Spectral Analysis

Given a GLM fit, the residuals time series is generated as the actual-predicted time series (Fig. 3C, rightmost panels; note the initial $\hat{n}_r$ time bins are filled with zero padding). The residuals are submitted to identical spectral analysis steps as are applied to the original time series. For the two example cases (Fig. 3D), the known 9 Hz peaks are successfully recovered, including for the lower FR spike train (Fig 3D, *right*) for which the shuffling method did not detect significant oscillation. This sensitivity gain – in tandem with the desired flattening of the RP distortion – is the result that we aimed to achieve by

selectively removing bounded $t\text{-}s_*(t)$ effects, and leaving intact the temporal structure expressed by the longer lags of the ISI distribution.

### *Methods Evaluation*

#### Oscillation Detection Performance: Low-to-Moderate Spike Rates

So far, we have shown a pair of example spike trains, for which the residuals method either matched or exceeded the shuffling method's detection of a known oscillation. To compare the two methods comprehensively, we ran a series of simulations and analyses that took into account a broad set of factors that may influence method performance, including the simulation parameter settings, and the corrected alpha level ($\alpha_c$) of the significance test. Evaluation began with a primary synthetic dataset that focused on low-to-moderate base firing rates (range 8-64 Hz), oscillations in the theta-low gamma ranges (7-32 Hz), and the default recovery period parameters; we also generated a series of secondary datasets to address follow-up questions.

For the primary dataset, we directly manipulated four simulation parameters ($T$, $f_{osc}$, $m$, and $p_{base_offset} = p_{base} - f_{osc}$); the tables in Fig. 4A-B list the full set of parameter settings. For each unique setting combination, we generated 100 spike trains, computed the shuffling- and residuals-corrected PSDs, and tabulated the true and false positive oscillation detection rates for each (see the Fig. 4 caption for the hit rate (HR) and false alarm (FA) rate definitions).

To compare the classification accuracy of the two methods over a range of $\alpha_c$ thresholds, we conducted a pROC (partial Receiver Operating Characteristic) analysis (Fig. 4D). This analysis was implemented over a pooled dataset (i.e., collapsing over the parameter settings in Fig. 4A-B, which are further considered in the additional analyses below). As in traditional ROC analysis, we plotted hit versus false alarm rates as the decision threshold ($\alpha_c$) was varied, with shuffling and residuals curves generated for each of 1000 subsamples of the full spike train dataset (with the subsampling performed to obtain bootstrapped estimates of variability). Since the original curves spanned variable FA and HR ranges, which did not fully span [0,1] (even with the inclusion of extreme significance thresholds), we truncated each curve to a partial ROC estimate that covered a restricted FA range (see Fig. 4 caption for details). The above steps enabled confirmation that the partial area under the ROC curve (pAUC) was reliably greater for the residuals method as compared against the shuffling method ($t$-test of $pAUC_{res}$-$pAUC_{shuf}$ : $t(999) = 1.3048e+03$, $p << .001$). This result indicates that the residuals method generally improves oscillation detection performance on this collapsed-parameter, primary synthetic dataset.

We have described a rationale for anticipating that the residuals method may be especially advantageous for the lowest-FR spike trains, and in particular for sensitivity to the detection of oscillations when the base FR is low relative to the frequency of the oscillation to be detected. This reasoning leads to the prediction of an inverse relationship between $p_{base_offset}$ and the magnitude of the residuals-associated improvement in hit rates. A qualitative trend consistent with this prediction is visible in the top right table of Fig. 4A (representing $D_{HR}$ = the residuals-shuffling differences in HR, for the original, non-subsampled data). For the select simulation parameters reported in the table, the predicted inverse relationship between $D_{HR}$ and $p_{base_offset}$ is apparent as a general tendency for higher $D_{HR}$ values (and warmer heatmap colors) towards the lower rows (corresponding to lower $p_{base_offset}$ settings) of the table. The Fig. 4A $D_{HR}$ values also point to a general inverted U-shaped trend relating $m$ to $D_{HR}$, such that the residuals' advantage appears to be strongest for moderate modulation strengths, as compared to the lowest $m$ levels (when the oscillation is either non-existent or too weak for either method to detect) and highest $m$ levels (when the oscillation is readily detected by both methods).

To quantify and statistically test these trends, and to assess the parametric effects on the methods' relative sensitivity and specificity more broadly, we conducted separate regression analyses of $D_{HR}$ and $D_{FA}$ ($FA_{res}$-$FA_{shuf}$), with the significance threshold held fixed ($\alpha_c = .05$), and the manipulated simulation

parameters taken into consideration. As the regression analyses assume normally distributed noise in $D_{HR}$ and $D_{FA}$, we conducted these regressions using $D_{HR}$ and $D_{FA}$ values derived from the same subsampled HR and FA output as informed the pROC analysis. Fig. S2-S3 show $D_{HR}$ and $D_{FA}$ (summarized as the mean +/- standard error over subsamples) for all 540 unique combinations of the manipulated parameters. The $D_{HR}$- and $D_{FA}$-specific regression models included predictors corresponding to all four directly manipulated parameters (T,$f_{osc}$, $p_{base_offset}$, and $m$, with all $m$ = 0 cases excluded), and a quadratic term for $m$.

The full regression results for the primary synthetic dataset are provided in Tables S1-S2; we highlight a subset of the significant effects here. The $D_{HR}$ results (Table S1) statistically confirmed the predicted inverse relationship between $D_{HR}$ and $p_{base_offset}$ ($p_{base_offset}$ term: $t$ = -292.25, $p$ << .001) and the observed inverted U-shaped relationship between $D_{HR}$ and $m$ ($m^2$ term: $t$ = -296.41, $p$ << .001). As an additional test of the $p_{base_offset}$ effect – which was correlated with $p_{base}$ in these simulations, thereby creating ambiguity in the interpretation of the effect as reflecting either base FR - $f_{osc}$ or the absolute base FR – we created a follow-up synthetic dataset in which a directly manipulated $p_{base}$ factor (levels = 15 or 20 Hz) was crossed with an independently varied $p_{base_offset}$ factor (levels = [2:2:8] Hz) and the same $T$ and $m$ parameter sets as were described above. Regression analyses of the $D_{HR}$ estimates from these simulations (Table S3; $D_{HR}$ values are plotted in Fig. S4) replicated the expected negative $p_{base_offset}$ effect ($t$ = -59.81, $p$ << .001). This result supports the interpretation of the margin between the firing rate and to-be-detected oscillation frequency as independently influential on the residuals method's enhanced sensitivity (beyond the effect of the absolute $p_{base}$).

No specific predictions were made regarding the two methods' relative false alarm rates, and for many of the parameter combinations represented in Fig. S3 (and Fig. S5 for the dataset with the direct $p_{base}$ manipulation), the $D_{FA}$ values clustered near zero. Nevertheless, the very high $n$ of subsamples considered resulted in high sensitivity to parametric effects, and significant main effects and interactions were found (Table S2; Table S4); however, their modest scale leaves uncertain the practical implications of these trends. As discussed below, more noticeable trends were found upon introducing higher FRs to the simulations.

### Oscillation Detection Performance: High Firing Rates

The residuals method was motivated by a problem presented by sparse spike trains, but we might also ask how effectively it performs with more vigorously spiking units. To this end, we created a secondary dataset that differed from the primary dataset in its use of higher $p_{base_offset}$ values (Fig. S6-S8).

Three patterns stand out in the results. First, as anticipated, high FRs reduced the residuals method's overall performance advantage: $pAUC_{res}$-$pAUC_{shuf}$ remained reliably positive (Fig. S6D), but with only a narrow margin of performance improvement. Second, as also anticipated, this narrowed performance gap was mirrored by a similar narrowing in the residuals' sensitivity advantage. $D_{HR}$ values remained mostly positive but very modest (Fig. S7; Table S5).

Third, we observed a complex interaction pattern in the methods' relative false alarm rates. Complete illustrations and statistical results are available in Fig. S8-S11 and Table S6. Here, we will highlight four principal observations (all based on quantities summarized over the 1000 subsamples; see Materials and Methods). (1) As duration $T$ increased, and especially as the modulation strength $m$ approached 100%, $D_{FA}$ scores grew either very positive (for low oscillation frequencies) or negative (for high frequencies; Fig. S8). In other words, in the high $T$, high $m$ regime, the residuals method switched from producing either more or fewer FAs than shuffling did, as $f_{osc}$ increased ($T$ x $f_{osc}$ x $m$ interaction: $t$ = -178.70, $p$ << .001). (2) Although $f_{osc}$ is correlated with the $p_{base}$ parameter, the sign flip in $D_{FA}$ scores cannot be easily attributed to increasing FRs. In the $T$ = 120 x 1024 ms, $m$ = 100% cases, the effect of $p_{base_offset}$ on $D_{FA}$ negatively interacted with $f_{osc}$ ($t_{f_osc\ x\ pbase_offset}$ = -68.23; $p$ << .001), reflecting a tendency for the sign of the FR offset effect to likewise flip. (3) Plots of the average absolute FAs for the lowest $f_{osc}$

setting (7 Hz, Fig. S9) show that in the high *T*, high *m* cases, the shuffling FA rates remained mostly low (mean over all subsample x $p_{base}$ combinations = 10.16%) in contrast to the elevated residuals FA rates (overall mean = 47.55%). For the highest $f_{osc}$ setting (32 Hz, Fig. S10), the high *T*, high *m* cases resulted in very elevated FA rates for shuffling (overall mean = 76.09%) and less dramatically elevated FA rates for the residuals method (overall mean = 31.73%). (4) Summarizing over all *T* = 120 x 1024 ms, *m* = 1.0 cases, both methods' false detections were typically characterized by very low peak power, relative to that observed for hit points (mean $power_{FA}/power_{Hit}$ = 0.15 for residuals and 0.22 for shuffling) and high, gamma range frequencies (means = 83.10 Hz for residuals and 57.86 Hz for shuffling). Fig. S11 presents representative PSDs for each method, in addition to samples of the associated synthetic spike trains. Although an ideal solution would prevent the false detections, their consistent features could aid in the development of post-hoc rules for flagging suspected FA points in either method's corrected PSDs.

### Evaluation of Recovery Period Duration Estimation

Overall, the preceding analyses indicate that the residuals method achieves its primary objective of enhancing oscillation detection in sparse spiking scenarios, while maintaining performance comparable to that of the shuffling method when spiking is more vigorous (excluding the limited high-FA cases described above). Correspondingly, the first step of the method – the estimation of the RP duration – was also reasonably accurate. In the primary synthetic dataset, for the RP estimates collapsed across the full set of 540 x 100 spike trains (Fig. 4C), 51.07% identified the 9 ms $n_r$ setting precisely, 85.13% erred by ≤ 1 ms, and 94.78% erred by ≤ 2 ms. The follow-up, high FR dataset also demonstrated generally low error rates (Fig. S6C: % of absolute errors, $|err|$, ≤ 0, 1, or 2 ms: 66.22%, 93.53%, 99.87%).

To determine whether this accurate performance might generalize to other RP durations and forms, we generated three additional secondary datasets, with the original RP parameters modified to produce either a long relative RP ($n_r$ = 18 ms, steepness parameter *k* = 0.7), a short, sharply-rising relative RP ($n_r$ = 4 ms, *k* = 0.4), or a short absolute RP ($n_r$ = 3 ms, *k* = 0). Estimates clustered closely around the ground truth $n_r$ for both the 3 ms absolute RP (Fig. S12C: % $|err|$ ≤ 0, 1, or 2 ms: 87.28%, 96.62%, 98.90%) and the 4 ms relative RP (Fig. S13C: % $|err|$ ≤ 0, 1, or 2 ms: 84.03%, 96.58%, 99.01%). When the relative RP with the default *k* = 0.7 steepness is doubled to 18 ms, absolute error rates are comparable to those for the original 9 ms duration if the deviations under consideration are doubled accordingly (Fig. S14C: % $|err|$ ≤ 0, 2, or 4 ms: 44.79%, 86.73%, 92.72%).

These datasets also afford the opportunity to ask whether the residuals method retains its performance advantage when confronted with variable RP features. As reported above for the default RP datasets, panels A, B, and D of Fig. S12-S14 show, for each RP setting and correction method, highlighted HR and FA outcomes for select parameter combinations, and the overall pROC and pAUC output. For all three RP configurations, residuals remains the better-performing oscillation detector (*t*-tests of $pAUC_{res}$-$pAUC_{shuf}$ : *t*s(999) > 1.0608e+03, *p*s << .001).

## Experimental Data Results

### *Motivation and Approach*

Further comparison of the residuals and shuffling methods, conducted on real, biological single-unit data, served two purposes. First, any qualitative discrepancies between the neuronal and synthetic spectra provide useful information that can help inform the development of more accurate simulation frameworks in the future. Second, and most importantly for the present purposes, these evaluations can help experimentalists anticipate what might be encountered when applying either PSD correction method to their own data, and determine what further development, if any, might be needed to tailor the methods to the specific needs of real-world research questions.

In empirically-observed, neuronal spike trains, the ground truth is unknowable. That said, we prioritized sampling neural data acquired during circumstances for which the prior expectation of oscillations is high, based on evidence from sources that are distinct from spike trains. Towards that end, we evaluated the two correction methods' output on single-unit data recorded from the internal globus pallidus (GPi) and GPi-targeted ventrolateral anterior thalamus (VLa) of a monkey with MPTP-induced parkinsonism, during the pre-reach delay (i.e., "hold") periods of a cued reaching task. Across multiple species and motor-related brain regions, pathologically exaggerated oscillations in neural activity are a hallmark of parkinsonism, especially within the beta (12-30 Hz) and neighboring alpha (8-12 Hz) frequency bands (see [28, 29] for recent general reviews, and [41, 42] for examples specific to the thalamus and GPi, respectively). Moreover, alpha-beta oscillations are especially pronounced during periods of postural maintenance, such as the hold periods under consideration here [39]. Since the bulk of the evidence concerning these oscillations has been obtained from recordings of local field potentials (LFPs), our expectation of their occurrence in spike trains is based on a source that, while ultimately derived from the same underlying electrophysiological signal, is not affected by RP-related (and general point process-related) challenges to oscillation measurement.

The shuffling and residuals methods' output were compared for 92 GPi units (analyzed recording duration min, median, max = 30, 80.5, 262 s) and 78 VLa units (31, 74.5, 191 s). As anticipated based on previous reports [30], firing rates for the analyzed rest periods were generally slower for the VLa than for the GPi ($p << .001$ by the Wilcoxon rank sum test; min, median, max FR for VLa = 3.09, 12.24, 41.02 Hz; for GPi = 3.13, 39.90, 114.96 Hz). Our simulation results allow us to predict the likely sensitivity of the two methods for detecting oscillations in data with durations and firing rates typical of this empirical dataset. The synthetic data hit rate results for a representative low beta oscillatory frequency (12 Hz), low base firing rate (13 Hz), and short spike train duration (30 x 1024 ms; Fig. S15) predict that, for an oscillation modulation strength of at least 60%, the residuals method at minimum will detect the oscillation > 65% of the time.

Application of the two methods to the empirical dataset revealed considerable variability in the resulting corrected PSDs, but some recurring patterns do stand out. Here, we will highlight two of these patterns with four representative units (Fig. 5, Fig S16). The full population results are summarized in Fig. 6. To highlight the (0,100] Hz region over which we sought significant power, all neural spectral plots are confined to the initial 100 Hz; however, note that full spectra may be generated using code from the Github repository that accompanies this report.

### *Single Unit Examples*

Fig. 5 presents two units that illustrate the first recurring pattern, which is characterized by its consistency with what the simulations would predict: beta (or alpha) peaks that were recovered by the residuals method exclusively, from sparse spike trains that share qualitative features with our synthetic data. Fig. 5A shows such an example from the VLa (mean FR = 11.90 Hz, analyzed duration = 63 s). The uncorrected PSD matches the canonical RP distortion form, with a candidate beta oscillation (peak = 13.67 Hz) sitting atop the depressed low-frequency region. The oscillation is flagged as significant in the residuals PSD, but not the uncorrected or shuffling PSD. In this case, the behavior of the shuffling method is similar to that illustrated in the right-most column of Fig. 2A: The shuffling approximation of the renewal equivalent PSD captures not only the global distortion-consistent trend, but also much of the candidate beta peak, resulting in partial removal of this feature from the shuffling-corrected PSD. The 11.90 Hz FR indicates that, when considered relative to the candidate beta oscillation's frequency, a spike occurred on average just .87 times per cycle. Fig. 5B shows a second example of this pattern, in an unusually slowly-spiking GPi unit (mean FR = 7.65 Hz, analyzed duration = 96 s, peak of alpha-beta oscillation = 15.62 Hz).

Fig. S16 presents two VLa units that illustrate the second pattern, broadly categorized as suspected alpha-beta oscillations accompanied by temporal structure that was not incorporated into our simulations. Such additional structure included both burst firing (that is, transient intervals of elevated spike rates, see both panels A and B) and prominent 1/f-like trends (panel B). The two correction methods' behavior could differ considerably in these scenarios, with variable outcomes for their relative sensitivity to likely oscillations.

In the first VLa example (Fig. S16A), a likely low beta oscillation (peak alpha-beta frequency = 12.70 Hz) appears in tandem with a sporadic tendency for the unit to fire in a highly stereotyped pattern of two-spike bursts. The joint presence of these two patterns is most apparent in the autocorrelation plot (top left panel). Key features of the intra- and extra-burst spiking can be discerned from the ISI PDF (Fig. S17A) and joint [$ISI_n$,$ISI_{n+1}$] PDF (Fig. S17C, [43]): ISIs of 2-3 ms occur exclusively between pairs of spikes that are bracketed by much longer ISIs (≥ 14 ms), and all remaining spikes are separated by ISIs that last at least 7 ms (and often much longer). As such, the ISI statistics for this unit are consistent with a description of the spike train as switching between two different states, for which the effective recovery period varies from very short (intra-burst) to much longer (extra-burst).

This two-state characterization lends insight into the PSD outcomes. In the uncorrected spectrum, the beta peak is sunken into a typical distortion-related trough, and the shuffling method achieves satisfactory flattening of the spectrum and raising of the peak. In contrast, the residuals method performs very little distortion correction, due to the 1 ms estimated RP duration (see Fig. 17A-B, *left*) and resulting 1 ms bound on the last-spike effects estimated by the PPM. Such a model should adequately account for RP effects within the two-spike bursts: On simulated, non-bursty spike trains with a 1 ms recovery period and parameters that otherwise roughly matched those for this VLa unit (see Materials and Methods), the residuals method exhibited accurate oscillation detection (for $f_{osc}$ = 12 Hz; $m$ = 0.6: hit rate = 100%, FA rate = 7%). The apparent under-correction in this VLa example is likely due to the insufficient removal of the long RP effects that occur outside of the bursts. In the Discussion we propose potential extensions upon the residuals method that may address the need to accommodate burst and non-burst states with differing RP dynamics.

In the second VLa example (Fig. S16B), a likely low beta oscillation (peak alpha-beta frequency = 13.67 Hz) appears alongside a somewhat different pattern of spiking, relative to the first example. The ISIs are not split across disconnected zones of the histograms (Fig. 17A, *right*), but the presence of bursting episodes is suggested by the high density of very short ISI incidence (28.08% of all ISIs ∈ [2,3] ms, compared to a mean of 4.83% for matched, non-bursty synthetic units; see Materials and Methods) and sequential co-occurrence (6.02% of all $ISI_n$, $ISI_{n+1}$ pairs ∈ [2,3] ms, compared to a mean of 0.25% for the matched synthetic units). Therefore, it is conceivable that this unit's spiking behavior might also be appropriately described as switching between states of higher FRs (intra-burst) and lower FRs (extra-burst). If so, whether the RP would also be best characterized as varying between the two states is unclear.

These spike pattern observations may again aid interpretation of the PSD results, to some extent. In the uncorrected spectrum, the beta peak sits atop a prominent 1/f-like trend, to which aperiodic burst-associated FR fluctuations could be a contributor [44]. The 1/f-like feature stands out more strongly than any RP distortion that may be present, consistent with the possibility that a very short RP characterizes much of the spike train. The shuffling method does remove the 1/f-like curve, but also removes the beta peak. The residuals method again performs little correction – owing to the 1 ms estimated RP – and therefore retains both the beta peak and 1/f-like trend (with many of the elevated low-frequency points of the PSD also labeled as significant). As alluded to in the Introduction, such 1/f-like features should ideally be modeled and considered an additional component of the baseline against which the statistical significance of candidate oscillatory peaks is determined. Possible strategies for extracting the 1/f-like component – which include both the aforementioned extensions to the residuals method to model

bursts, and post-hoc function fits to the corrected PSD – will also be considered further in the Discussion.

### *GPi and VLa Population Results*

Fig. 6 illustrates results from the two correction methods for the whole neuronal dataset. The three columns show simplified versions of the spectra (again zoomed in to 0-100 Hz on the *x* axis), in which any points that are both significant and labeled as a local spectral peak are labeled in magenta, and all other significant points are marked as blue. Spectra are plotted according to the order of the neuron's observed firing rate, with the lowest-FR units at the top.

The population-level patterns mirror those that we highlight with the individual unit examples. First, we found trends consistent with those that our simulations would predict, with respect to overall incidence of alpha-beta oscillation detections, and the variation of these detections with firing rates. Collapsing across the GPi and VLa units, the residuals method reported that a greater proportion of the total units exhibited significant alpha-beta power, compared to the shuffling method (66/170 versus 37/170 units, McNemar's $\chi^2$ = 19.12, $p$ < 1.3e-5). Moreover, for those units for which alpha-beta band oscillations were reported by at least one method, lower firing rates were predictive of an increased likelihood of residuals-only detection (as compared to shuffling-only or joint method detection; $\beta_{FR}$ = -0.089, p < 1.5e-5). Therefore, to the extent that the exclusively residuals-identified power did indeed correspond to true underlying alpha-beta oscillations, the method did fulfill the expectation of enhanced sensitivity, especially in the presence of low firing rates. Note also that the residuals method did identify alpha-beta oscillations over units with a variety of estimated recovery periods (median RP = 3.5 ms, iqr = 6 ms), consistent with the high hit rates seen across the varied RP parameters in the simulations (Fig. S12-S14).

At the population level, the second pattern – non-oscillatory contributors to spike timing – is most visible as a regular occurrence of 1/f-like effects. In Fig. 6, PSDs with strong 1/f-like trends tend to appear as significant points in the lowest frequency bands, starting with the (0,4] Hz delta band, and at times extending well beyond it. Using a simple rule to label likely 1/f-like trends (≥ 2 consecutive significant points within (0,4] Hz; see Materials and Methods), we identified a greater incidence of these trends in the residuals- versus shuffling-corrected PSDs (85/170 versus 25/170 units, McNemar's $\chi^2$ = 54.40, $p$ < 1.7e-13). This result is consistent with the shuffling method's aggressive removal of the 1/f-like trend in Fig. S16B. Also consistent with the Fig. S16B unit, we can identify additional cases in Fig. 6 for which the residuals method retains a 1/f-like trend that extends into a series of significant alpha-beta points. Therefore, with respect to the above observations of residuals-only alpha-beta detections, an open question does remain regarding the extent to which putative periodic power in these bands would remain significant after removal of the 1/f-like, presumably aperiodic spectral component.

Again consistent with the Fig. S16B example, we did observe additional units for which probable bursting may have contributed to 1/f-like trends. For example, for four additional VLa units that met the above delta band rule, > 25% of the ISIs fell within 2-3 ms (range: 35.76% - 42.91%), in spite of overall mean FRs of 5.35 Hz - 8.26 Hz.

## Discussion

### Discussion Overview

To recap, our principal objective was to evaluate a novel approach for removal of the recovery period's distortion of spike train power spectra, with the ultimate aim of improving the accuracy of oscillation detection. An established ISI shuffling method [7] approaches this problem by extracting the spectral component explained by the spike train's ISI distribution alone. Because the ISI distribution contains information about not only the recovery period, but also oscillations, the shuffling method may

hinder detection of relatively subtle oscillatory spiking, especially in slowly spiking units. As an alternative, we developed a regression-based "residuals" method that also models and removes temporal structure related to the ISIs, but only those ISIs that fall within a historical bound determined by the estimated duration of the recovery period (RP) function. We compared residuals-corrected versus shuffling-corrected PSDs to determine whether the former enhanced sensitivity to oscillations, with this effect predicted to be greatest for low FR spike trains.

The following sections summarize the predicted and less anticipated aspects of our findings, propose future directions for method development, and situate the residuals method in the context of the broader set of tools available for rhythmic spiking analysis.

## Predicted Findings

The two methods were initially evaluated on synthetic spike trains, generated with varying settings for firing rate, oscillatory frequency and modulation strength, and recording duration, and with the RP represented as an exponentially-recovering modulation of spike probability. Collapsing over these variable settings, we found that residuals correction enabled more accurate classification of true versus false positive points in the power spectra. More detailed examination of the varied parameters, and of their impact on the methods' relative hit and false alarm rates, revealed three general trends. First, and as anticipated, residuals correction primarily benefited from enhanced sensitivity to the ground truth oscillations, especially when firing rates were low, relative to the oscillation frequency. Second, two other predictors of strong relative residuals performance (moderate oscillation modulation ($m$); short spike train duration ($T$)) suggest that this method may be most useful for identifying subtle oscillatory effects that reside close to a threshold level of detectability. Third, the largely sensitivity-driven benefit was gained, for most parameter settings, without exaggeration of false positive oscillation detections. On a limited subset of high FR, strongly oscillating spike trains, both the residuals and shuffling methods did tend to produce PSDs with very low-power FA peaks; we propose strategies for addressing these FAs farther below.

The largely favorable results for the residuals method generalized across variable recovery period durations (3, 4, 9, and 18 ms), and shapes (i.e., absolute or exponentially-rising, with variable steepness). This persistence of successful detection across RP characteristics was accompanied by relatively accurate estimation of the ground truth RP durations by the ISI distribution-fitting algorithm, with estimation errors rarely exceeding 1-2 ms.

In the real GPi and VLa data – keeping in mind the caveat concerning the unknowability of the ground truth – we did observe additional evidence consistent with our expectations. Focusing on the alpha-beta frequency range that was anticipated in the parkinsonian NHP under consideration, we found that the residuals method reported significant oscillations for a greater fraction of the units than the shuffling method did. This result was anticipated on the basis of the prevalence of low-to-moderate FRs in this dataset (especially amongst the VLa units), and follow-up tests confirmed the relevance of low spike rates for predicting residuals-only oscillation detection. Although we cannot presently determine whether the residuals-flagged oscillations are "real", it is conceivable that future empirical work could aid in providing corroborating evidence. For example, confidence that an oscillatory drive governs a unit's spiking could be bolstered by the discovery of matching oscillatory trends in the unit's synaptic input dynamics (e.g., as tracked by high resolution biosensors of the relevant neurotransmitters [45]). One might also anticipate that any significant spike oscillations may correspond to similar oscillations in a unit's membrane voltage. Consistent with this expectation, a recent study in mice [46] observed that striatal cholinergic interneurons that exhibited robust delta rhythms in their spiking (as estimated in the time domain with an ISI-based heuristic) also showed strong delta rhythms in subthreshold membrane voltage (as estimated with wavelets).

A subset of the GPi and VLa spike trains exhibited aperiodic structure that the current implementation of the residuals method was not designed to address. Consequently, the corresponding residuals PSDs should be interpreted with care. Below, we discuss potential strategies for addressing these additional patterns that may arise in empirical spike trains.

## Areas for Future Development

In our synthetic and empirical spike trains, three special cases gave rise to unanticipated outcomes in the corrected PSDs. Below, we briefly review these cases, and the attendant directions for future development.

### *False alarms with fast spiking and strong modulation*

When operating on synthetic inputs generated with high base FRs ($p_{base}$) and strong oscillatory modulation ($m$), both shuffling and residuals correction produced spectra with elevated false alarm rates (Fig. S8). This trend worsened with increasing spike train duration ($T$), and with variation in the "true" oscillation frequency ($f_{osc}$). The nature of the $f_{osc}$ effect depended on the method in question. For the lowest frequencies (e.g., 7 Hz; Fig. S9), residuals correction commonly yielded false positives, but the shuffling FA rate remained mostly low. As $f_{osc}$ approached the highest tested settings (e.g., 32 Hz; Fig. S10), shuffling FAs become extremely frequent, while the residuals FA rate moderated somewhat, such that the relative specificity of the two methods flipped. Under the FA-promoting conditions, both methods reported false peaks that tended to fall within the gamma band, and were characterized by very low power that barely exceeded the statistical significance threshold (Fig. S11A). The distinctive features of these false alarms suggest that a simple PSD postprocessing step (e.g., imposition of a more stringent significance and/or effect size threshold) may help minimize their appearance in future analyses.

Future development of the residuals method should seek to prevent these elevated false alarms. In the fast spiking, strong modulation cases, a key challenge concerns the heightened correlation between the effect that we are attempting to model (the recovery period) and the unmodeled oscillation effect. This correlation engenders an "omitted variable bias" [47] in RP function estimation. Visualization of representative spike trains (Fig. S11B) illustrates the core difficulty: For a strongly oscillating spike train, the spikes and the recovery period regressors will principally cluster under the positive phases of the oscillation cycle. Consequently, the RP regressors are associated not only with the recent occurrence of single spikes, but also with an elevation of the background firing rate above its overall mean. The result is an underestimation of the RP's suppressive effect, which in turn leads to incomplete removal of the RP distortion effect from the corrected PSD.

Stevenson [47] noted that such biases might be reduced through introduction of variables that may capture suspected unmodeled effects (here, oscillations of unknown frequency), but also acknowledged that such strategies can be assumption-intensive and nontrivial. Such variables are present in those PPMs that estimate spike rhythms entirely in the time domain (e.g., [24-26]). We recap the pros and cons of these time domain methods in the concluding section below ("Conclusions: Tools for Spike Oscillation Analysis.").

### *Burst firing impacts on the corrected spectra*

Spike "bursts" consist of transient intervals of rapid spiking, occurring against a background of a slower-rate steady state [48-50]. In the empirical data, we observed multiple units that displayed likely bursting, as inferred from their ISI statistics (e.g., as in Fig. S16-S17). Previous investigations have reported bursting in the parkinsonian GPi and VLa (see [51] for a review).

Burst firing can introduce two variations in the spike train that may influence the form of the residuals-corrected PSD. First, by definition, bursting creates increases in the background FR, which may occur aperiodically. Our current residuals implementation does not remove such variation, which may in

turn contribute to 1/f-like trends in the resulting PSD (consistent with Fig. S17B; see also [44]). Second, bursting could alter the recovery period, as suggested by the Fig. S17A VLa example. Our current model cannot accommodate distinct burst and non-burst recovery periods, thereby preventing the accurate removal of all RP-associated variance. In the VLa example, the model estimated and removed the short, within-burst RP, resulting in insufficient correction for the long, between-bursts RP.

Future development could conceivably expand the current residuals method to capture burst- and non-burst states. Previous work has proposed discrete state-space models designed to infer burst states in point process time series [52, 53] and could inform such an approach. Armed with an algorithm that labels spike train timepoints with their likelihood of pertaining to burst or non-burst intervals, one could attempt to model the two states' base FRs and RPs with separate sets of delta functions, with these functions weighted by the inferred probabilities that either state is active at each timepoint (i.e., $log\lambda(\mathrm{t}|\mathrm{H_t},\beta) = \sum_{k=1}^{2} \alpha_k(t) \left[\sum_{j=1}^{\hat{n}_r(k)} \beta_{jk}\, \delta_{t-j,s_*(t)}\right]$, where *k* indexes over the states, and the $\alpha_k(t)$ denote the complementary state probabilities). Note that adoption of this strategy would require *a priori* estimates of the two states' RP durations ($\hat{n}_r(k)$), and therefore the development of a dual-RP estimation algorithm.

### *Aperiodic spectral components*

Several GPi and VLa units exhibited prominent 1/f-like trends in both the uncorrected and residuals spectra. "1/f-like" refers to spectral components that are well-approximated by a $1/f^{\chi}$ function, to which researchers commonly add a "knee" parameter to accommodate a potential bend in the curve [54]. Strong 1/f-like effects are ubiquitous in neural activity spectra (and especially EEG/LFP PSDs) and represent the frequency domain manifestation of aperiodic trends. Removal of these trends will be necessary for rigorous interpretation of the suprathreshold points (e.g., the green points in Fig. S16B) in the residuals-corrected PSDs.

Two general strategies may help achieve this removal. First, as noted above, aperiodic burst spiking may contribute to the 1/f-like component [44]. Therefore, effective removal of burst-related variance at the regression stage could ultimately attenuate some 1/f-like effects. However, bursts represent just one of the many diverse underlying dynamics that may give rise to 1/f-like trends [55]. Therefore, complementary procedures will still be needed to remove any remaining aperiodic effects.

The second strategy entails the standard postprocessing step of removing a fitted $1/f^{\chi}$ function (with a possible knee component) from the PSD. This step would assume that residuals correction has removed sufficient RP-related distortion from the spectrum. The routine introduced by [3] offers one option for achieving a robust curve fit to the PSD. This method additionally fits a series of Gaussians to any narrowband spectral peaks, and those Gaussians that sufficiently exceed the aperiodic baseline are inferred to represent likely oscillations.

## Adaptation to Time-Frequency Analysis

In its present form, the residuals method assumes the presence of stable oscillatory components. Adaptation of the method to the analysis of time-varying oscillations may be feasible, depending on two factors: the unit's spike rate, and the stability of its RP function. Vigorous spiking and a stable RP present the optimal scenario. In this case, one could reasonably apply the PPM strategy that we described here, and submit the residuals time series to the preferred time-frequency analysis method (e.g., short-time Fourier Transform (STFT), Wavelet Transform, or bandpass filtering followed by Hilbert Transform).

If spiking is slow, then insufficient data may be available to detect transient oscillations, even with adequate RP correction. Simulations may aid in estimating the minimal spike rate required for sensitivity to any hypothesized effects.

If the RP function is likely to vary over time, the pattern of the suspected variation should inform the adaptation approach. If one assumes that the RP varies across transitions between latent discrete

states, but remains stable within states, then it may be reasonable to first apply a multi-state PPM (as we have described for bursting) and then submit the residuals to time-frequency analysis. Alternatively, if one assumes both fast spiking and RP stability within each segment of an STFT analysis, one might conceivably execute the full residuals routine (RP estimation/PPM/PSD) on each segment independently. More challenging data features (e.g., complicated RP variation patterns) will require the development of methods that go beyond the basic approaches that we have laid out here.

If time-frequency analyses are attempted, appropriate postprocessing should be applied to extract aperiodic trends. For example, Wilson et al. [56] recently proposed a time-resolved method for aperiodic and periodic trend decomposition, which extends upon the stationarity-assuming method of Donoghue et al. [3].

## Conclusions: Tools for Spike Oscillation Analysis

Throughout this report, we have referenced a set of available tools that both estimate spike oscillations and control for recovery period effects. Here, we briefly summarize the major features of each method, with an eye to the tradeoffs to consider when deciding which approach to adopt for a given project. We divide these tools into those that operate entirely in the time domain, and those that return outputs in the frequency domain.

### *Time domain methods*

We further split the time domain methods into two subcategories. The first subcategory encompasses all Poisson GLM-based PPMs that may capture both recovery period and oscillation effects (e.g., [25, 26]). As described in the Materials and Methods ("Clarifications and Relationship to Pre-existing PPM Approaches"), these PPMs typically model spike history effects with a series of indicator functions of varying width. A set of narrow, short-lag regressors estimate recovery period and bursting effects, and wider, longer-lag regressors are positioned to estimate periodic effects. For example, spike counts summarized over the -50 to -40 ms bin may be used to estimate 20-25 Hz rhythms [25]. Note that the fundamental differences between such methods and the current residuals implementation are relatively modest. The principal distinctions concern the choice of indicator regressors to include, and the estimation of oscillatory trends through either strategically spaced regressors or export of the residuals to spectral analysis.

Depending on a researcher's goals, the replacement of the PSD with binned history effects may stand out as a limitation of these alternative GLM methods. The bin-based oscillation estimates sacrifice the high frequency resolution of a standard DFT-produced spectrum, and represent a departure from the typical basis set of sinusoids. These characteristics may complicate projects that aim to draw comparisons with previous spectral data, or derive precise oscillation frequency information. However, when these limitations are acceptable, these GLM methods can offer some advantages. For example, by incorporating oscillation terms alongside the short-lag regressors, these models may mitigate the omitted variable bias in RP estimation.

The second subcategory consists of the Latent Oscillatory Spike Train (LOST) model of [24]. LOST is a distinct, Bayesian PPM, which shares some key features with the regression-based models, but also differs in significant ways. The shared features include short-lag history terms (implemented through splines) and other covariate regressors (e.g., related to task events) that could be easily incorporated into a standard GLM. The core difference entails the modeling of oscillations. LOST represents underlying oscillatory drives as continuous latent states, which are constrained by priors, but may vary over time with respect to their precise modulation strengths and center frequencies. These latent oscillatory components in turn shape the ongoing spike probability. This approach therefore offers a number of benefits, especially for those users who seek a principled accounting for modest fluctuations in the properties of the rhythmic drive. Moreover, as with the above-mentioned GLM methods, LOST's

simultaneous estimation of short-lag and oscillatory trends may reduce any omitted variable bias effects. Note that LOST does constrain the number of distinct oscillatory effects that the user can simultaneously estimate, and may be best suited for small datasets (due to both runtime demands and manual fine-tuning requirements).

### *Frequency domain methods*

The shuffling and residuals methods form the frequency domain category. These methods offer the advantage of a corrected power spectrum estimate.

We reiterate that the shuffling method remains a reasonable approach for RP distortion correction in moderate-to-high-FR spike trains. In addition, shuffling appeared to handle some instances of aperiodic structure (e.g., 1/f-like trends, and the Fig. S16A bursting example) more effectively than the current residuals implementation did. The principal tradeoff for these strengths is reduced sensitivity to oscillations in sparse spike trains. Note also that the shuffling method's run time may grow very long as the spike train duration and count of shuffling iterations increases, and that this method shares the residuals method's vulnerability to biases introduced by strong oscillations (as indicated by Fig. S10-S11).

The residuals method is an especially useful option when a power spectrum is required, spike rates are low, and when other factors, such as modest modulation strength and short recording duration, challenge oscillation detection. In the implementation we present here, the method draws upon an accessible and flexible GLM framework. Moreover, the generation of a corrected time series output – and not only corrected oscillation estimates – expands the set of analyses that the underlying framework might ultimately support, including common time-frequency analyses.

## Acknowledgments

We thank Lisa Nieman-Vento for her contributions to animal care.

## References

1. Buzsaki G, Anastassiou CA, Koch C. The origin of extracellular fields and currents--EEG, ECoG, LFP and spikes. Nat Rev Neurosci. 2012;13(6):407-20. Epub 20120518. doi: 10.1038/nrn3241. PubMed PMID: 22595786; PubMed Central PMCID: PMCPMC4907333.
2. Matzner A, Bar-Gad I. Quantifying spike train oscillations: biases, distortions and solutions. PLoS Comput Biol. 2015;11(4):e1004252. Epub 20150424. doi: 10.1371/journal.pcbi.1004252. PubMed PMID: 25909328; PubMed Central PMCID: PMCPMC4409360.
3. Donoghue T, Haller M, Peterson EJ, Varma P, Sebastian P, Gao R, et al. Parameterizing neural power spectra into periodic and aperiodic components. Nat Neurosci. 2020;23(12):1655-65. Epub 20201123. doi: 10.1038/s41593-020-00744-x. PubMed PMID: 33230329; PubMed Central PMCID: PMCPMC8106550.
4. Bachetti M, Huppenkothen D. No time for dead time: use the Fourier amplitude differences to normalize dead-time-affected periodograms. The Astrophysical Journal Letters. 2018;853(2):L21.
5. Bair W, Koch C, Newsome W, Britten K. Power spectrum analysis of bursting cells in area MT in the behaving monkey. J Neurosci. 1994;14(5 Pt 1):2870-92. PubMed PMID: 8182445; PubMed Central PMCID: PMCPMC6577471.
6. Edwards BW, Wakefield GH, Powers NL. The spectral shaping of neural discharges by refractory effects. J Acoust Soc Am. 1993;93(6):3353-64. doi: 10.1121/1.405718. PubMed PMID: 8326062.
7. 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. J Neurophysiol. 2006;95(5):3245-56. Epub 20060111. doi: 10.1152/jn.00055.2005. PubMed PMID: 16407432.

8. Whalen TC, Willard AM, Rubin JE, Gittis AH. Delta oscillations are a robust biomarker of dopamine depletion severity and motor dysfunction in awake mice. J Neurophysiol. 2020;124(2):312-29. Epub 20200624. doi: 10.1152/jn.00158.2020. PubMed PMID: 32579421; PubMed Central PMCID: PMCPMC7500379.
9. Truccolo W, Eden UT, Fellows MR, Donoghue JP, Brown EN. A point process framework for relating neural spiking activity to spiking history, neural ensemble, and extrinsic covariate effects. Journal of neurophysiology. 2005;93(2):1074-89.
10. Kandel ER, Koester JD, Mack SH, Siegelbaum SA. Principles of Neural Science, Sixth Edition: McGraw-Hill Education; 2021.
11. Surmeier DJ, Mercer JN, Chan CS. Autonomous pacemakers in the basal ganglia: who needs excitatory synapses anyway? Curr Opin Neurobiol. 2005;15(3):312-8. doi: 10.1016/j.conb.2005.05.007. PubMed PMID: 15916893.
12. Halliday DM, Rosenberg JR, Amjad AM, Breeze P, Conway BA, Farmer SF. A framework for the analysis of mixed time series/point process data--theory and application to the study of physiological tremor, single motor unit discharges and electromyograms. Prog Biophys Mol Biol. 1995;64(2-3):237-78. doi: 10.1016/s0079-6107(96)00009-0. PubMed PMID: 8987386.
13. Durrett R. Essentials of Stochastic Processes: Springer International Publishing; 2016.
14. Gerstner W, Kistler WM, Naud R, Paninski L. Neuronal dynamics: From single neurons to networks and models of cognition: Cambridge University Press; 2014.
15. Thomson DJ. Spectrum estimation and harmonic analysis. Proceedings of the IEEE. 1982;70(9):1055-96.
16. Baaske MK, Kormann E, Holt AB, Gulberti A, McNamara CG, Potter-Nerger M, et al. Parkinson's disease uncovers an underlying sensitivity of subthalamic nucleus neurons to beta-frequency cortical input in vivo. Neurobiol Dis. 2020;146:105119. Epub 20200928. doi: 10.1016/j.nbd.2020.105119. PubMed PMID: 32991998; PubMed Central PMCID: PMCPMC7710979.
17. Hasegawa T, Chiken S, Kobayashi K, Nambu A. Subthalamic nucleus stabilizes movements by reducing neural spike variability in monkey basal ganglia. Nat Commun. 2022;13(1):2233. Epub 20220425. doi: 10.1038/s41467-022-29750-2. PubMed PMID: 35468893; PubMed Central PMCID: PMCPMC9038919.
18. Hashimoto T, Muralidharan A, Yoshida K, Goto T, Yako T, Baker KB, et al. Neuronal activity and outcomes from thalamic surgery for spinocerebellar ataxia. Ann Clin Transl Neurol. 2018;5(1):52-63. Epub 20171210. doi: 10.1002/acn3.508. PubMed PMID: 29376092; PubMed Central PMCID: PMCPMC5771317.
19. Pasquereau B, Turner RS. Primary motor cortex of the parkinsonian monkey: differential effects on the spontaneous activity of pyramidal tract-type neurons. Cereb Cortex. 2011;21(6):1362-78. Epub 20101102. doi: 10.1093/cercor/bhq217. PubMed PMID: 21045003; PubMed Central PMCID: PMCPMC3097989.
20. Tachibana Y, Iwamuro H, Kita H, Takada M, Nambu A. Subthalamo-pallidal interactions underlying parkinsonian neuronal oscillations in the primate basal ganglia. Eur J Neurosci. 2011;34(9):1470-84. doi: 10.1111/j.1460-9568.2011.07865.x. PubMed PMID: 22034978.
21. Sani S, Shimamoto S, Turner RS, Levesque N, Starr PA. Microelectrode recording in the posterior hypothalamic region in humans. Neurosurgery. 2009;64(3 Suppl):ons161-7; discussion ons7-9. doi: 10.1227/01.NEU.0000334051.91501.E3. PubMed PMID: 19240565; PubMed Central PMCID: PMCPMC3777657.
22. Schmidt SL, Chew EY, Bennett DV, Hammad MA, Frohlich F. Differential effects of cholinergic and noradrenergic neuromodulation on spontaneous cortical network dynamics. Neuropharmacology. 2013;72:259-73. Epub 20130503. doi: 10.1016/j.neuropharm.2013.04.045. PubMed PMID: 23643755.
23. Van Drongelen W. Signal processing for neuroscientists: Academic press; 2018.

24. Arai K, Kass RE. Inferring oscillatory modulation in neural spike trains. PLoS Comput Biol. 2017;13(10):e1005596. Epub 20171006. doi: 10.1371/journal.pcbi.1005596. PubMed PMID: 28985231; PubMed Central PMCID: PMCPMC5646905.
25. Sarma SV, Cheng ML, Eden U, Williams Z, Brown EN, Eskandar E. The effects of cues on neurons in the basal ganglia in Parkinson's disease. Front Integr Neurosci. 2012;6:40. Epub 20120726. doi: 10.3389/fnint.2012.00040. PubMed PMID: 22855673; PubMed Central PMCID: PMCPMC3405280.
26. Saxena S, Sarma SV, Patel SR, Santaniello S, Eskandar EN, Gale JT. Modulations in Oscillatory Activity of Globus Pallidus Internus Neurons During a Directed Hand Movement Task-A Primary Mechanism for Motor Planning. Front Syst Neurosci. 2019;13:15. Epub 20190430. doi: 10.3389/fnsys.2019.00015. PubMed PMID: 31133824; PubMed Central PMCID: PMCPMC6524693.
27. Bar-Gad I, Ritov Y, Bergman H. The neuronal refractory period causes a short-term peak in the autocorrelation function. J Neurosci Methods. 2001;104(2):155-63. doi: 10.1016/s0165-0270(00)00335-6. PubMed PMID: 11164241.
28. Deffains M, Bergman H. Parkinsonism-related beta oscillations in the primate basal ganglia networks - Recent advances and clinical implications. Parkinsonism Relat Disord. 2019;59:2-8. Epub 20181214. doi: 10.1016/j.parkreldis.2018.12.015. PubMed PMID: 30578010.
29. Singh A. Oscillatory activity in the cortico-basal ganglia-thalamic neural circuits in Parkinson's disease. Eur J Neurosci. 2018;48(8):2869-78. Epub 20180208. doi: 10.1111/ejn.13853. PubMed PMID: 29381817.
30. Schwab BC, Kase D, Zimnik A, Rosenbaum R, Codianni MG, Rubin JE, et al. Neural activity during a simple reaching task in macaques is counter to gating and rebound in basal ganglia-thalamic communication. PLoS Biol. 2020;18(10):e3000829. Epub 20201013. doi: 10.1371/journal.pbio.3000829. PubMed PMID: 33048920; PubMed Central PMCID: PMCPMC7584254.
31. Kass RE, Eden UT, Brown EN. Analysis of Neural Data: Springer New York; 2014.
32. Kim S, Putrino D, Ghosh S, Brown EN. A Granger causality measure for point process models of ensemble neural spiking activity. PLoS Comput Biol. 2011;7(3):e1001110. Epub 20110324. doi: 10.1371/journal.pcbi.1001110. PubMed PMID: 21455283; PubMed Central PMCID: PMCPMC3063721.
33. Kass RE, Ventura V. A spike-train probability model. Neural Comput. 2001;13(8):1713-20. doi: 10.1162/08997660152469314. PubMed PMID: 11506667.
34. Sarmashghi M, Jadhav SP, Eden U. Efficient spline regression for neural spiking data. PLoS One. 2021;16(10):e0258321. Epub 20211013. doi: 10.1371/journal.pone.0258321. PubMed PMID: 34644315; PubMed Central PMCID: PMCPMC8513896.
35. Orcioni S, Paffi A, Apollonio F, Liberti M. Revealing spectrum features of stochastic neuron spike trains. Mathematics. 2020;8(6):1011.
36. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez JC, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics. 2011;12:77. Epub 20110317. doi: 10.1186/1471-2105-12-77. PubMed PMID: 21414208; PubMed Central PMCID: PMCPMC3068975.
37. Bankiewicz K, Oldfield E, Chiueh C, Doppman J, Jacobowitz D, Kopin I. Hemiparkinsonism in monkeys after unilateral internal carotid artery artery infusion of 1-methyl-4-phenyl-1, 2, 3, 6-tetrahydropyridine (MPTP). Life sciences. 1986;39(1):7-16.
38. Bankiewicz KS, Sanchez-Pernaute R, Oiwa Y, Kohutnicka M, Cummins A, Eberling J. Preclinical models of Parkinson's disease. Curr Protoc Neurosci. 2001;Chapter 9:Unit9 4. doi: 10.1002/0471142301.ns0904s09. PubMed PMID: 18428552.
39. Engel AK, Fries P. Beta-band oscillations--signalling the status quo? Curr Opin Neurobiol. 2010;20(2):156-65. Epub 20100330. doi: 10.1016/j.conb.2010.02.015. PubMed PMID: 20359884.
40. Su T, Dy JG. In search of deterministic methods for initializing K-means and Gaussian mixture clustering. Intelligent Data Analysis. 2007;11(4):319-38.

41. Sarnthein J, Jeanmonod D. High thalamocortical theta coherence in patients with Parkinson's disease. J Neurosci. 2007;27(1):124-31. doi: 10.1523/JNEUROSCI.2411-06.2007. PubMed PMID: 17202479; PubMed Central PMCID: PMCPMC6672280.
42. Eisinger RS, Cagle JN, Opri E, Alcantara J, Cernera S, Foote KD, et al. Parkinsonian Beta Dynamics during Rest and Movement in the Dorsal Pallidum and Subthalamic Nucleus. J Neurosci. 2020;40(14):2859-67. Epub 20200227. doi: 10.1523/JNEUROSCI.2113-19.2020. PubMed PMID: 32107277; PubMed Central PMCID: PMCPMC7117906.
43. Rodieck RW, Kiang NY, Gerstein GL. Some quantitative methods for the study of spontaneous activity of single neurons. Biophys J. 1962;2(4):351-68. doi: 10.1016/s0006-3495(62)86860-x. PubMed PMID: 14493108; PubMed Central PMCID: PMCPMC1366397.
44. Grüneis F, Nakao M, Yamamoto M. Counting statistics of 1/f fluctuations in neuronal spike trains. Biological cybernetics. 1990;62(5):407-13.
45. Leopold AV, Shcherbakova DM, Verkhusha VV. Fluorescent Biosensors for Neurotransmission and Neuromodulation: Engineering and Applications. Front Cell Neurosci. 2019;13:474. Epub 20191023. doi: 10.3389/fncel.2019.00474. PubMed PMID: 31708747; PubMed Central PMCID: PMCPMC6819510.
46. Shroff SN, Lowet E, Sridhar S, Gritton HJ, Abumuaileq M, Tseng HA, et al. Striatal cholinergic interneuron membrane voltage tracks locomotor rhythms in mice. Nat Commun. 2023;14(1):3802. Epub 20230626. doi: 10.1038/s41467-023-39497-z. PubMed PMID: 37365189; PubMed Central PMCID: PMCPMC10293266.
47. Stevenson IH. Omitted Variable Bias in GLMs of Neural Spiking Activity. Neural Comput. 2018;30(12):3227-58. Epub 20181012. doi: 10.1162/neco_a_01138. PubMed PMID: 30314428.
48. Cotterill E, Charlesworth P, Thomas CW, Paulsen O, Eglen SJ. A comparison of computational methods for detecting bursts in neuronal spike trains and their application to human stem cell-derived neuronal networks. J Neurophysiol. 2016;116(2):306-21. Epub 20160420. doi: 10.1152/jn.00093.2016. PubMed PMID: 27098024; PubMed Central PMCID: PMCPMC4969396.
49. Ermentrout B, Terman DH. Mathematical foundations of neuroscience: Springer; 2010.
50. Izhikevich EM. Neural excitability, spiking and bursting. International journal of bifurcation and chaos. 2000;10(06):1171-266.
51. Rubin JE, McIntyre CC, Turner RS, Wichmann T. Basal ganglia activity patterns in parkinsonism and computational modeling of their downstream effects. Eur J Neurosci. 2012;36(2):2213-28. doi: 10.1111/j.1460-9568.2012.08108.x. PubMed PMID: 22805066; PubMed Central PMCID: PMCPMC3400124.
52. Kleinberg J, editor Bursty and hierarchical structure in streams. Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining; 2002.
53. Tokdar S, Xi P, Kelly RC, Kass RE. Detection of bursts in extracellular spike trains using hidden semi-Markov point process models. J Comput Neurosci. 2010;29(1-2):203-12. Epub 20090821. doi: 10.1007/s10827-009-0182-2. PubMed PMID: 19697116.
54. Miller KJ, Sorensen LB, Ojemann JG, den Nijs M. Power-law scaling in the brain surface electric potential. PLoS Comput Biol. 2009;5(12):e1000609. Epub 20091218. doi: 10.1371/journal.pcbi.1000609. PubMed PMID: 20019800; PubMed Central PMCID: PMCPMC2787015.
55. Kramer MA, Chu CJ. The 1/f-like behavior of neural field spectra are a natural consequence of noise driven brain dynamics. bioRxiv. 2023:2023.03. 10.532077.
56. Wilson LE, da Silva Castanheira J, Baillet S. Time-resolved parameterization of aperiodic and periodic brain activity. Elife. 2022;11. Epub 20220912. doi: 10.7554/eLife.77348. PubMed PMID: 36094163; PubMed Central PMCID: PMCPMC9467511.

## Figures

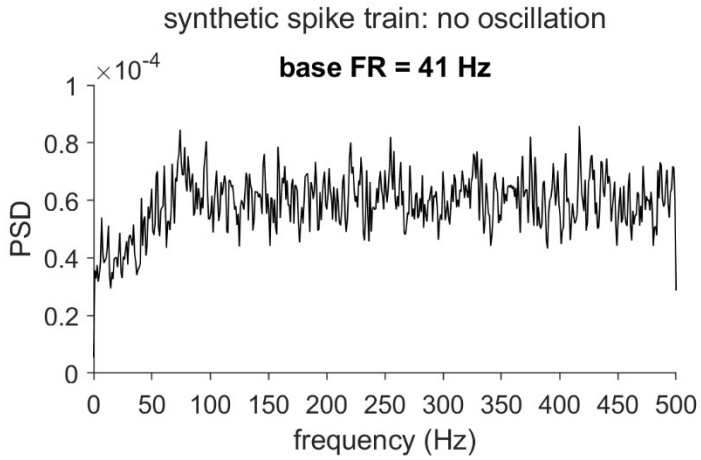



**Fig. 1. Recovery period distortion of the power spectrum for a synthetic spike train with no oscillation.** Power spectral density (PSD) of a spike train simulated under the assumption of Poisson spiking interrupted by a roughly exponential recovery period (RP, the post-spike reduction in spike probability, which can extend beyond the refractory period). Relative to the flat spectrum of a Poisson process, this RP-distorted PSD exhibits a characteristic trough over the lowest frequencies, and relative elevation and ringing over the higher frequencies. The PSD was generated according to the simulation framework from [7] (Materials and Methods; baseline firing rate (FR) = 41 Hz; RP duration $n_r$ = 9 ms and steepness $k$ = 0.7; simulation duration $T$ = 60 x 1024 ms).

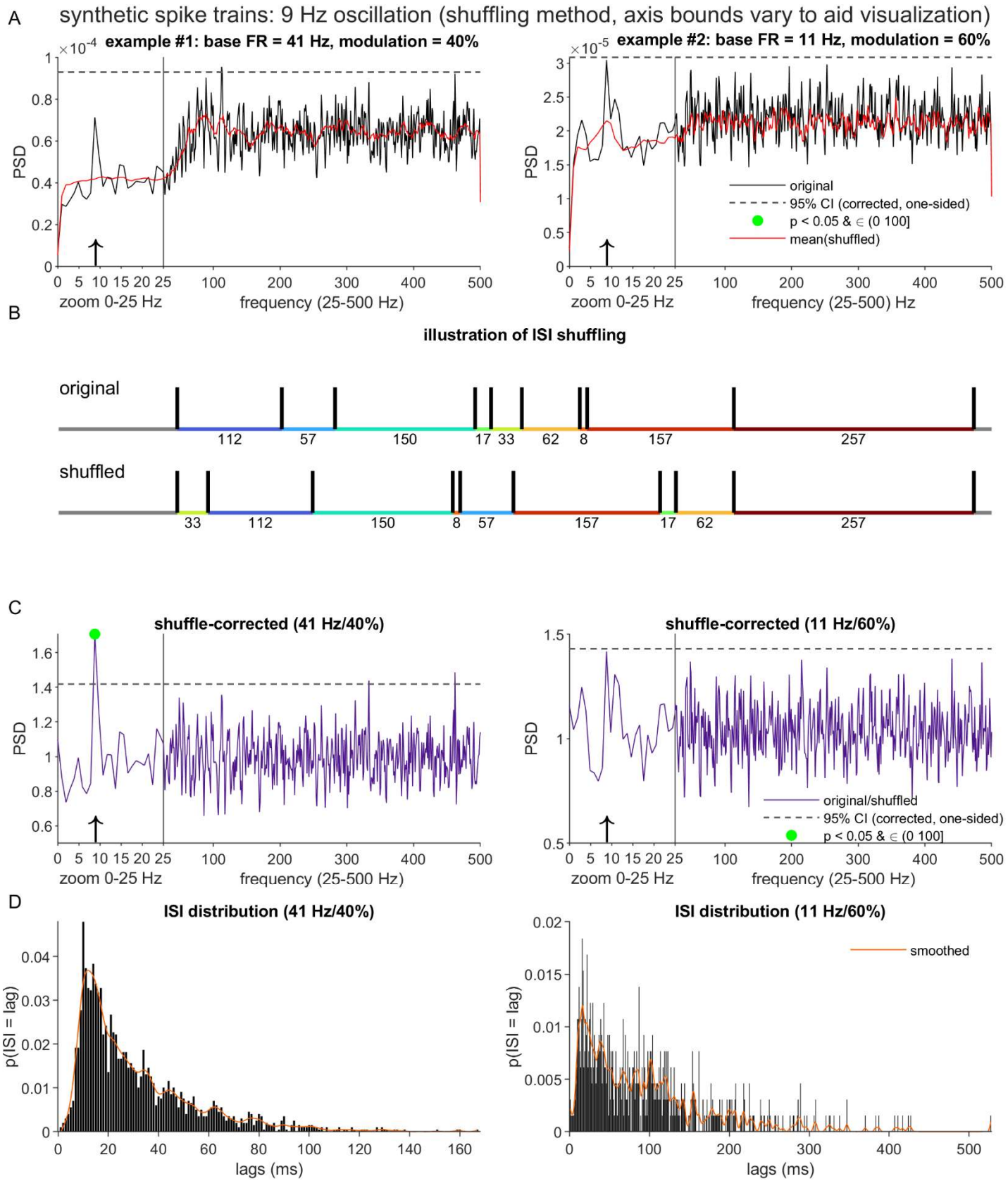

A
synthetic spike trains: 9 Hz oscillation (shuffling method, axis bounds vary to aid visualization)
example #1: base FR = 41 Hz, modulation = 40%
example #2: base FR = 11 Hz, modulation = 60%
PSD
zoom 0-25 Hz
frequency (25-500 Hz)
frequency (25-500) Hz
original
95% CI (corrected, one-sided)
p < 0.05 & ∈ (0 100]
mean(shuffled)
B
illustration of ISI shuffling
original
shuffled
C
shuffle-corrected (41 Hz/40%)
shuffle-corrected (11 Hz/60%)
original/shuffled
D
ISI distribution (41 Hz/40%)
ISI distribution (11 Hz/60%)
smoothed
p(ISI = lag)
lags (ms)

**Fig. 2. Recovery period distortion of the PSDs for synthetic spike trains with oscillations, and the limitations of inter-spike interval-based correction.** (A) PSDs (black lines) for two synthetic spike trains, generated using the same RP parameters and simulation durations as Fig. 1, and added oscillation terms of matching oscillation frequency ($f_{osc}$ = 9 Hz). The spike trains differed with respect to the oscillation modulation strength (*m*) and the baseline firing rate (base FR; see column titles). Dashed lines mark the corrected 95% one-sided confidence interval (CI) constructed under an approximation of a null model of Poisson spiking (see Materials and Methods). Green dots (absent in (A)) indicate PSD points that both cross the CI bound and fall within an a priori search range of (0, 100] Hz. Red lines plot the spectral component that the ISI shuffling algorithm [7] estimates to be attributable to the structure present in the inter-spike intervals (ISIs). (B) Illustration of one iteration of ISI shuffling on a truncated 1000 ms dataset (actual implementation applied global shuffling to the full spike train duration). The average PSD of several shuffled spike trains forms the estimated ISI-attributable spectrum. (C) The shuffling-corrected PSDs, equal to the ratio of the original and shuffling-estimated PSDs. (D) ISI probability distributions for the two example spike trains. Smoothing spline fits (orange lines) highlight the more visible shaping of the ISI distribution by the 9 Hz oscillation in the lower-FR case.

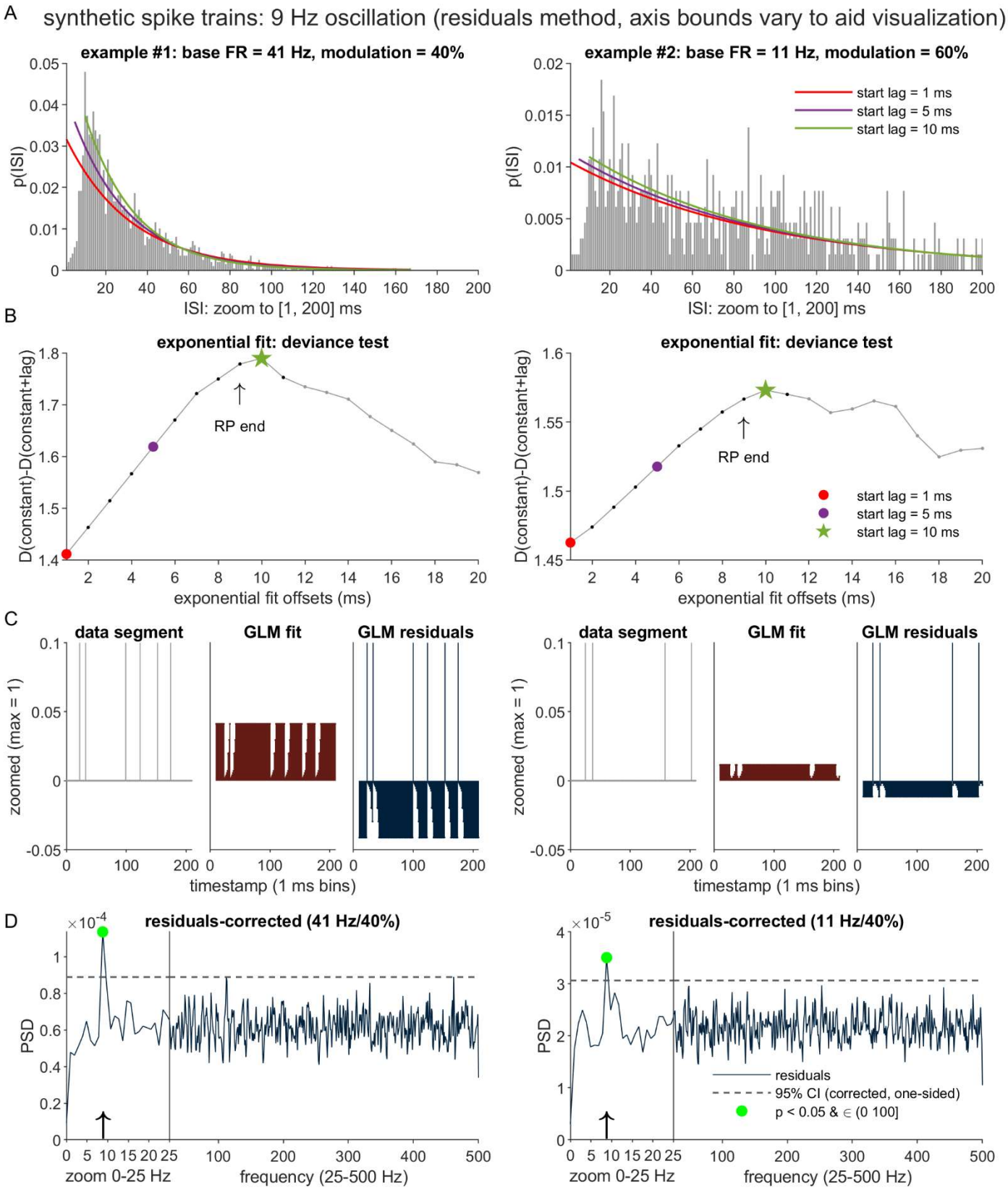
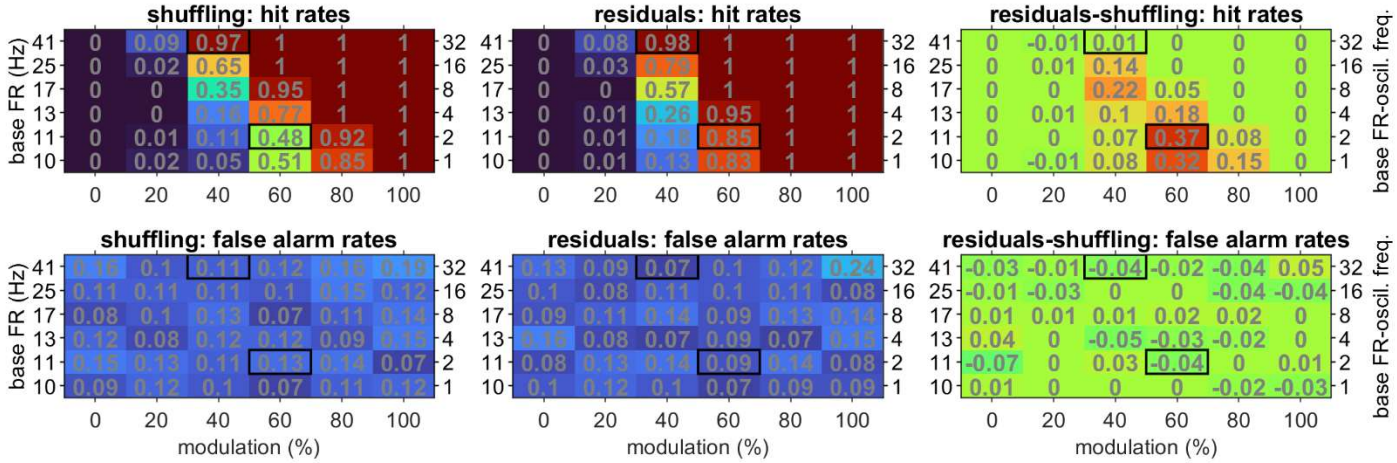

A
synthetic spike trains: 9 Hz oscillation (residuals method, axis bounds vary to aid visualization)
example #1: base FR = 41 Hz, modulation = 40%
example #2: base FR = 11 Hz, modulation = 60%
p(ISI)
ISI: zoom to [1, 200] ms
start lag = 1 ms
start lag = 5 ms
start lag = 10 ms
B
exponential fit: deviance test
D(constant)-D(constant+lag)
RP end
exponential fit offsets (ms)
C
data segment
GLM fit
GLM residuals
zoomed (max = 1)
timestamp (1 ms bins)
D
residuals-corrected (41 Hz/40%)
residuals-corrected (11 Hz/40%)
PSD
zoom 0-25 Hz
frequency (25-500 Hz)
residuals
95% CI (corrected, one-sided)
p < 0.05 & ∈ (0 100]

**Fig. 3. Distortion correction by the proposed residuals method.** (A) Illustration of the procedure for obtaining an estimate of the RP duration ($\hat{n}_r$), as applied to the two example spike trains from Fig. 2. A series of right-shifted exponential curves are fit to each ISI distribution, left-anchored to starting positions advanced in 1 ms steps (with 3 sample iterations highlighted in the figures). (B) Plots of the deviance difference statistic, ΔD, as a function of the first 20 starting positions of the exponential fits. D(constant), D(constant+lag) = deviance measures for the intercept-only and intercept+exponential curve models, respectively. ΔD tracks the goodness of fit contributed by the exponential curve. Each $\hat{n}_r$ estimate is set equal to the post-spike lag immediately preceding the first local maximum in the corresponding ΔD plot. (C) Illustration of the Poisson Generalized Linear Model (GLM) accounting for the effects on spike likelihood of the time elapsed since the most recent of any spikes occurring within the preceding $\hat{n}_r$ ms. The three panels (which highlight the first 209 ms of data) depict the GLM input (left panels), model fit (middle panels), and the model residuals (right panels). (D) PSDs of the GLM residuals. Plotting and statistical conventions follow those from Fig. 2.

synthetic spike trains: method evaluation (see caption for hit/FA defintion; "hit" is FA if modulation = 0%)

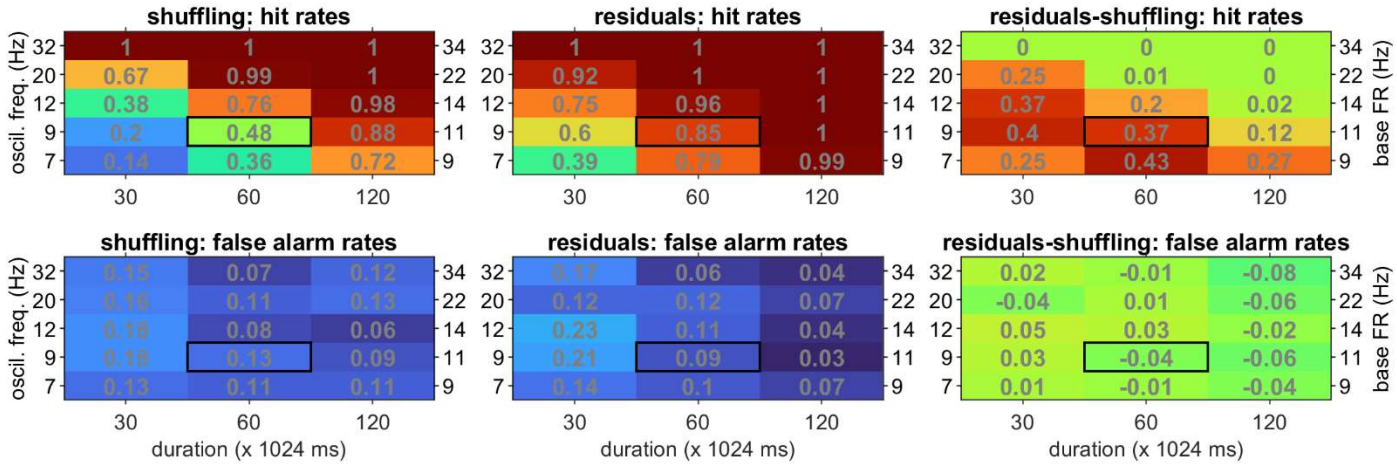
A

results sample #1: duration = 60 x 1024 ms, 9 Hz oscillation, 100 simulation iterations per cell

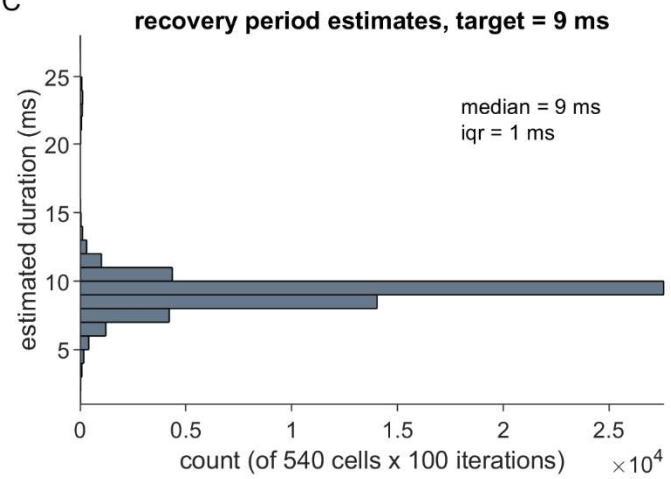



B

results sample #2: base FR-oscil. freq. = 2 Hz, moduation = 60%, 100 simulation iterations per cell




C




partial ROC curves: bootstrapped pAUC comparison

D

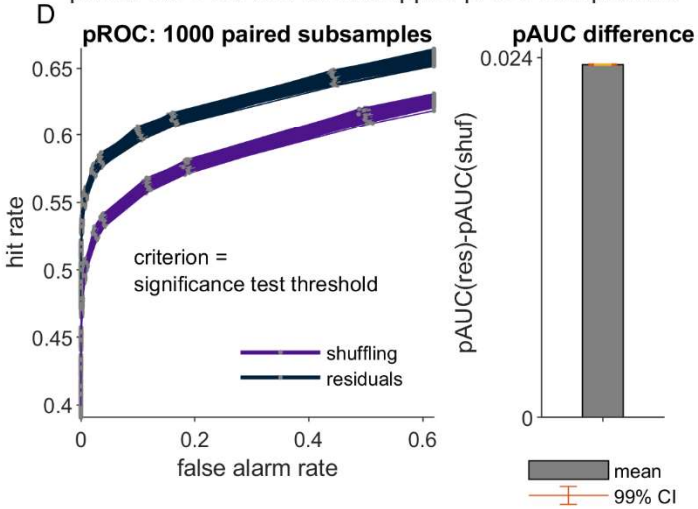


**Fig. 4. Performance of the shuffling and residuals methods over a diverse synthetic dataset.** One hundred synthetic spike trains were generated for each of 540 unique parameter setting combinations (resulting from the crossing of the varied duration ($T$), modulation strength ($m$), oscillation frequency ($f_{osc}$), and [base FR - oscillation frequency] offset ($p_{base_offset}$) parameters; see Materials and Methods). For each spike train, corrected PSDs were generated by the two correction methods. When $m > 0$, a PSD represented a "hit" if it contained ≥ 1 point of significant power within the three nearest neighbors of the ground truth oscillation frequency (= $f_{osc}$, when $m > 0$) on the sampled frequency axis (resolution = 0.9766 Hz). A false alarm (FA)-containing PSD included ≥ 1 point of significant power at frequency points > 5 Hz from $f_{osc}$ (when $m > 0$) or at any frequency point (when $m = 0$). (A) A subset of the methods' hit and FA rates, shown for all combinations of $m$ and $p_{base_offset}$, for a fixed $T$ = 60 x 1024 ms and $f_{osc}$ = 9 Hz. Bold outlines highlight the parameter combinations used for the two example spike trains depicted in Figs. 2-3. (B) Hit and FA rates, shown for all combinations of $T$ and $f_{osc}$, for a fixed $p_{base_offset}$ = 2 Hz and $m$ = 60%. Bold outlines highlight the parameter combinations used for the second, lower FR spike train depicted in Figs. 2-3. (C) Histogram of the residuals method's estimates of the recovery period duration (ground truth $n_r$ = 9 ms), collapsed over all of the synthetic spike trains. (D) Partial receiver operator characteristic (pROC) analysis of the residuals (res) and shuffling (shuf) methods' performance. Left: Hit versus FA rates are plotted over varying statistical test thresholds; individual curves represent spike train subsamples (matched for *res* and *shuf*) that were uniformly drawn from the 540 conditions (to obtain bootstrapped estimates of variability; see Materials and Methods). The "partial" designation refers to the truncating of individual curves to the FA range bounded by the minimum and maximum FAs observed over all curves. Right: Mean and 99% CI corresponding to the $t$-test (df = 999) contrasting the residuals-shuffling partial area under the curve (pAUC) difference against zero.

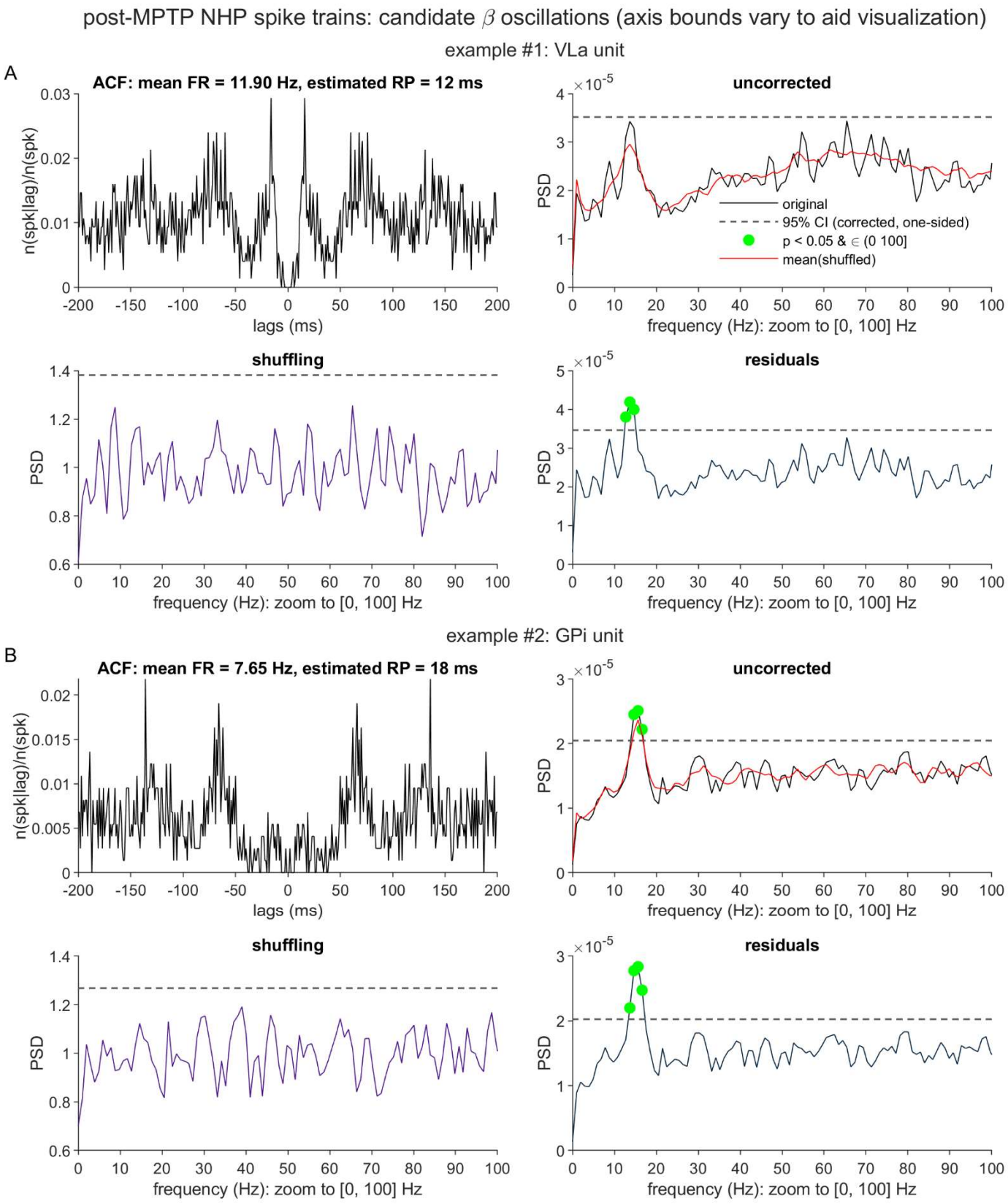

post-MPTP NHP spike trains: candidate β oscillations (axis bounds vary to aid visualization)
example #1: VLa unit
A
ACF: mean FR = 11.90 Hz, estimated RP = 12 ms
n(spk|lag)/n(spk)
lags (ms)
uncorrected
PSD
original
95% CI (corrected, one-sided)
p < 0.05 & ∈ (0 100]
mean(shuffled)
frequency (Hz): zoom to [0, 100] Hz
shuffling
residuals
example #2: GPi unit
B
ACF: mean FR = 7.65 Hz, estimated RP = 18 ms

**Fig. 5. Comparison of the shuffling and residuals output for two highlighted units acquired from a parkinsonian non-human primate (NHP).** Spike trains were sampled from one monkey that had been injected with the MPTP (1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine) neurotoxin, and restricted to resting periods of a reaching task. Example units exhibit autocorrelation functions (ACF) consistent with a possible β (beta band; 12-30 Hz) oscillation. (A) Example unit from the ventrolateral anterior thalamus (VLa). Top, left: ACF, normalized by total spike count; spk = spike. Top right, bottom left, and bottom right: Uncorrected, shuffling-corrected, and residuals-corrected PSDs for the VLa unit, following the formatting and statistical conventions from Fig. 2-3. In the uncorrected spectrum panel, the ISI-attributed PSD (red line) was estimated via an adaptation of the global shuffling method to temporally separated task windows (see Materials and Methods). (B) The same sequence of plots as described in (A), as applied to a unit from the internal globus pallidus (GPi).

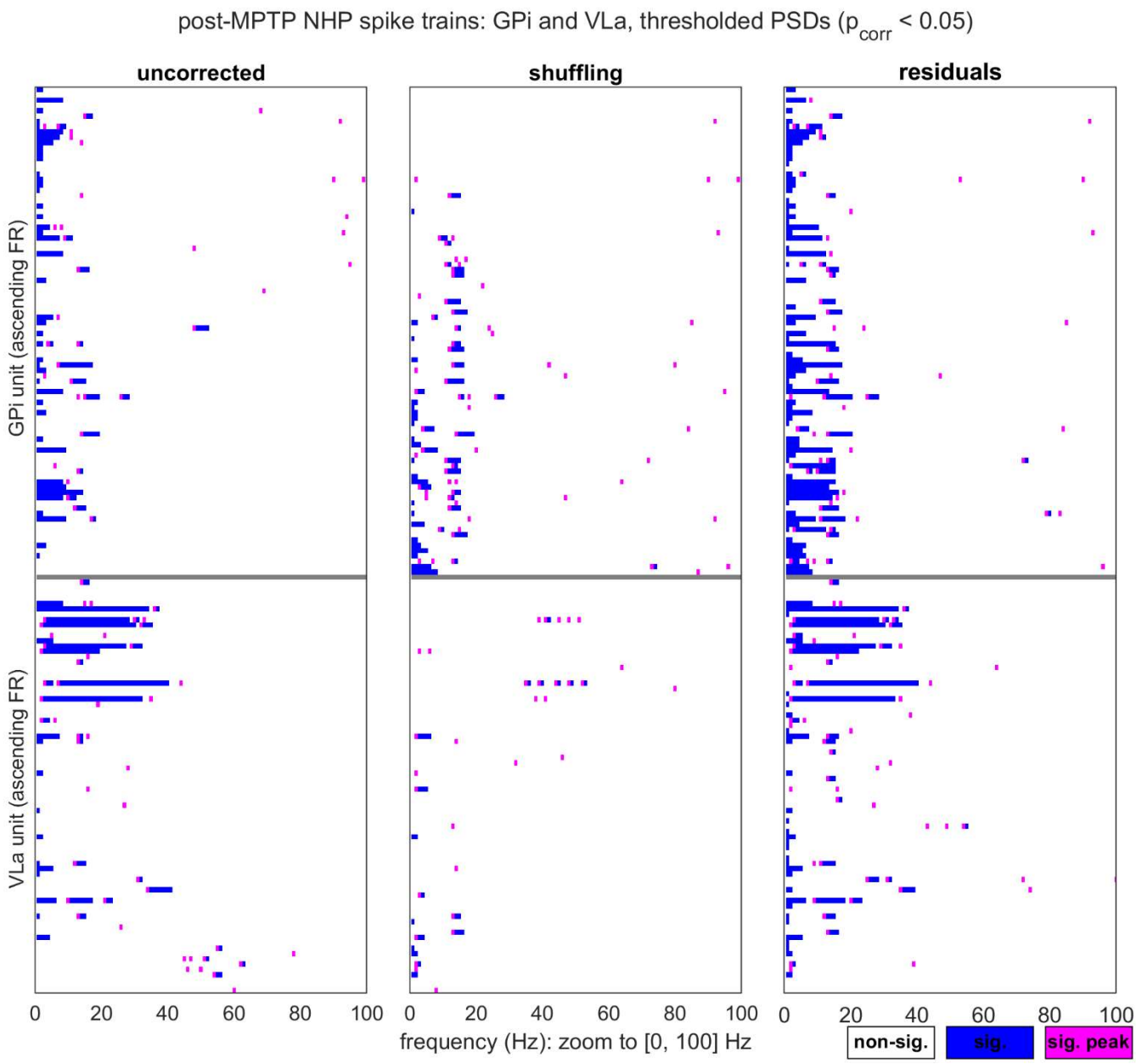



**Fig. 6. Points of significant spectral power identified by the shuffling and residuals methods across the population of analyzed spike trains acquired from the parkinsonian non-human primate (NHP).** Spike data were sampled and statistically tested according to the same conventions as applied to the two example units in Fig. 5; $p_{corr}$ = corrected $p$ value. Left panel: Points of statistically significant (sig.) power for the full set of GPi (top) and VLa (bottom) uncorrected power spectra. Each heatmap row represents a single power spectrum, bounded within [0, 100] Hz, with each significant point labeled as either a magenta peak (local maximum within a set of neighboring significant points) or a blue non-peak. For each region, rows are sorted in order of ascending firing rate (slowest FRs in the highest-positioned rows). Middle panel, right panel: Points of significant power found in the shuffling- and residuals-corrected PSDs.

# Supporting Information

## Supporting Figures

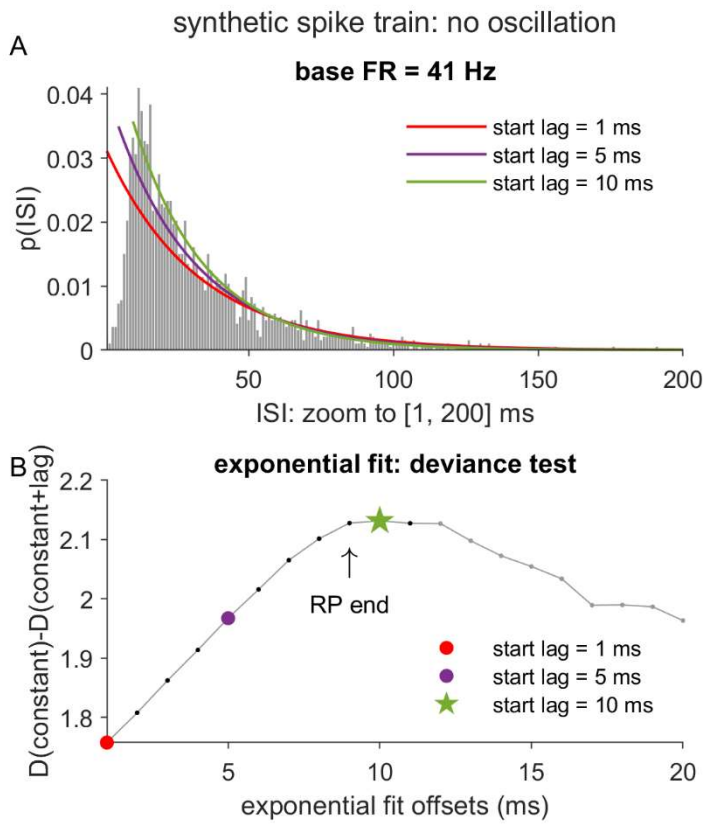



**Fig. S1. Estimation of recovery period duration: No oscillation example.** (A) Illustration of the procedure for obtaining an estimate of the RP duration ($\hat{n}_r$), as applied to a synthetic spike train with no oscillation (modulation strength $m$ = 0). A series of right-shifted exponential curves are fit to the ISI distribution, left-anchored to starting positions advanced in 1 ms steps (with 3 sample iterations highlighted in the figure). (B) Plot of the deviance difference statistic, ΔD, as a function of the first 20 starting positions of the exponential fits. D(constant), D(constant+lag) = deviance measures for the intercept-only and intercept+exponential curve models, respectively. ΔD tracks the goodness of fit contributed by the exponential curve. The $\hat{n}_r$ estimate is set equal to the post-spike lag immediately preceding the first local maximum in the ΔD plot.

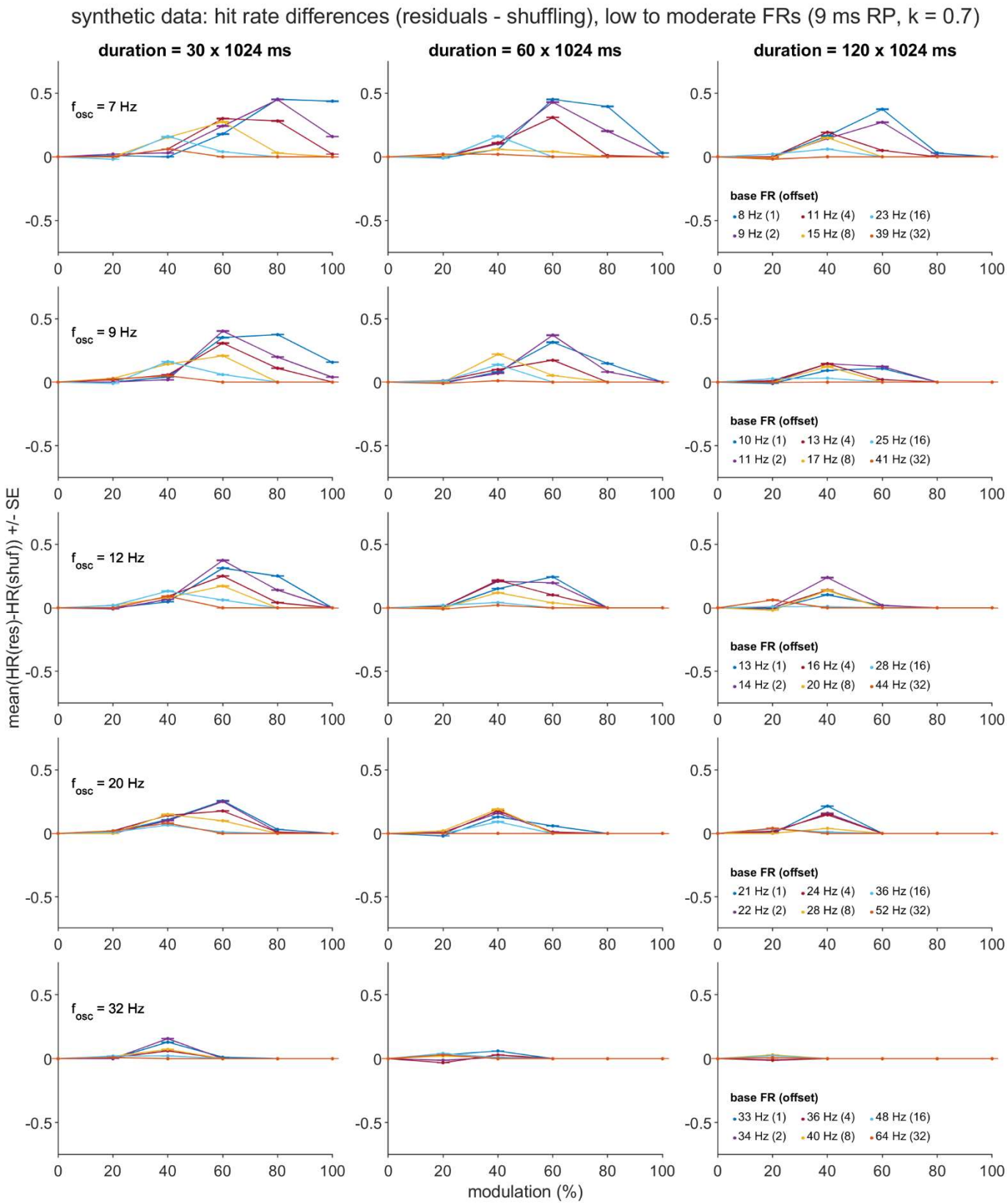

synthetic data: hit rate differences (residuals - shuffling), low to moderate FRs (9 ms RP, k = 0.7)
duration = 30 x 1024 ms
duration = 60 x 1024 ms
duration = 120 x 1024 ms
f_osc = 7 Hz
f_osc = 9 Hz
f_osc = 12 Hz
f_osc = 20 Hz
f_osc = 32 Hz
base FR (offset)
8 Hz (1)
9 Hz (2)
11 Hz (4)
15 Hz (8)
23 Hz (16)
39 Hz (32)
10 Hz (1)
11 Hz (2)
13 Hz (4)
17 Hz (8)
25 Hz (16)
41 Hz (32)
13 Hz (1)
14 Hz (2)
16 Hz (4)
20 Hz (8)
28 Hz (16)
44 Hz (32)
21 Hz (1)
22 Hz (2)
24 Hz (4)
28 Hz (8)
36 Hz (16)
52 Hz (32)
33 Hz (1)
34 Hz (2)
36 Hz (4)
40 Hz (8)
48 Hz (16)
64 Hz (32)
mean(HR(res)-HR(shuf)) +/- SE
modulation (%)

**Fig. S2. Residuals (res) - shuffling (shuf) difference in hit rates ($D_{HR}$) over the varied parameters of the primary synthetic dataset.** Means and standard errors (SE) reflect summaries over the hit rates (HR) computed for each of 1000 subsamples of the original, primary synthetic dataset, which was generated using low-to-moderate firing rates (FR) and a 9 ms relative recovery period (RP; $k$ = steepness parameter). See Materials and Methods and the Figure 4 caption for details regarding the definition of a hit and the subsampling procedure. Plots depict all 540 unique combinations of oscillation frequency ($f_{osc}$, rows), simulation duration ($T$, columns), oscillation modulation strength ($m$, $x$ axes) and base FR - oscillation frequency offset ($p_{base_offset}$, lines).

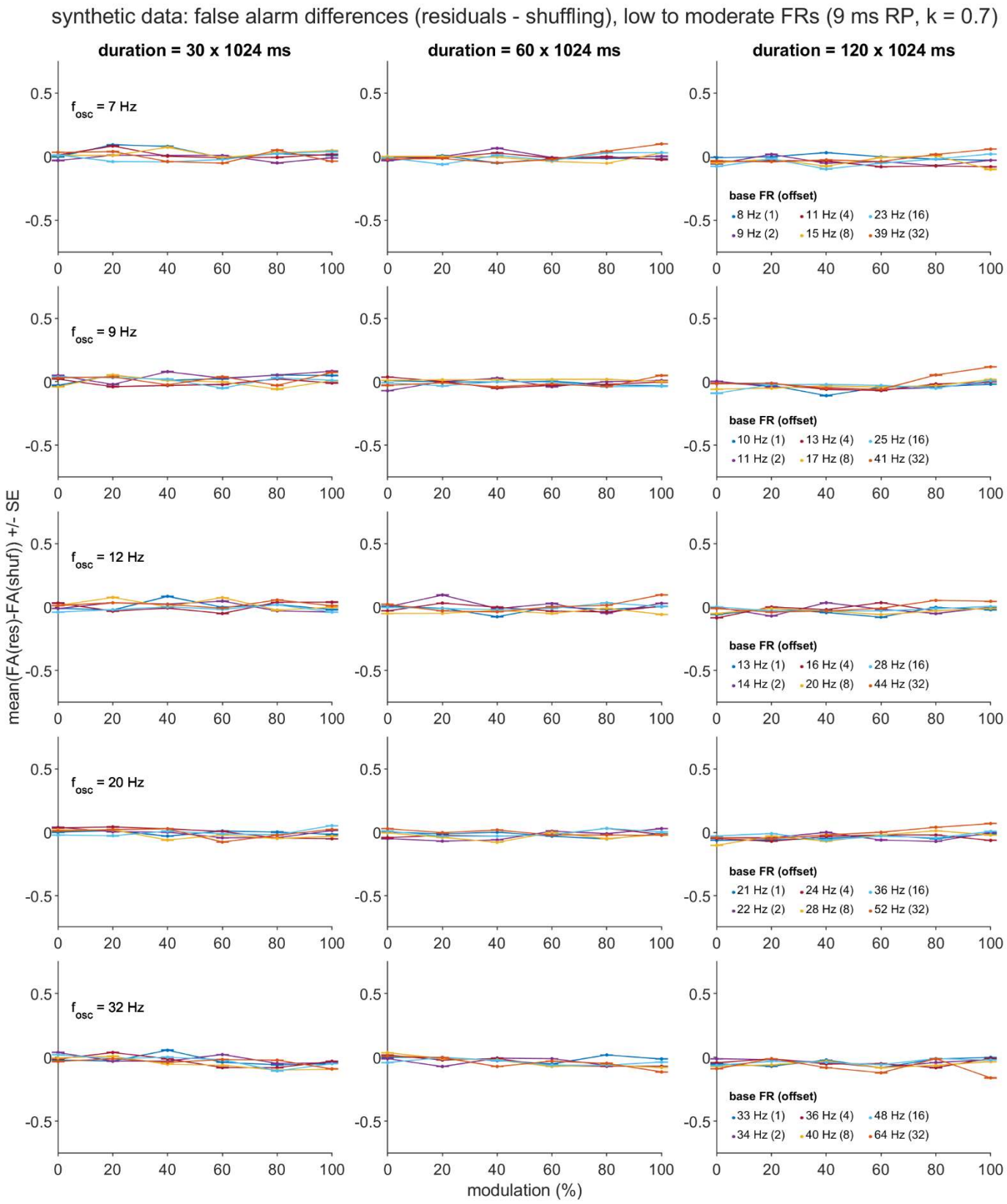

synthetic data: false alarm differences (residuals - shuffling), low to moderate FRs (9 ms RP, k = 0.7)
duration = 30 x 1024 ms
duration = 60 x 1024 ms
duration = 120 x 1024 ms
f_osc = 7 Hz
f_osc = 9 Hz
f_osc = 12 Hz
f_osc = 20 Hz
f_osc = 32 Hz
base FR (offset)
8 Hz (1)
11 Hz (4)
23 Hz (16)
9 Hz (2)
15 Hz (8)
39 Hz (32)
10 Hz (1)
13 Hz (4)
25 Hz (16)
11 Hz (2)
17 Hz (8)
41 Hz (32)
13 Hz (1)
16 Hz (4)
28 Hz (16)
14 Hz (2)
20 Hz (8)
44 Hz (32)
21 Hz (1)
24 Hz (4)
36 Hz (16)
22 Hz (2)
28 Hz (8)
52 Hz (32)
33 Hz (1)
36 Hz (4)
48 Hz (16)
34 Hz (2)
40 Hz (8)
64 Hz (32)
mean(FA(res)-FA(shuf)) +/- SE
modulation (%)

**Fig. S3. Residuals - shuffling difference in false alarm rates ($D_{FA}$) over the varied parameters of the primary synthetic dataset.** Means and standard errors reflect summaries over the false alarm rates (FA) computed for each of 1000 subsamples of the original dataset. See the Figure 4 caption for details regarding the definition of a false alarm. Abbreviations, plotting conventions, and the subsampling procedure follow from those described for Fig. S2 and Fig. 4.

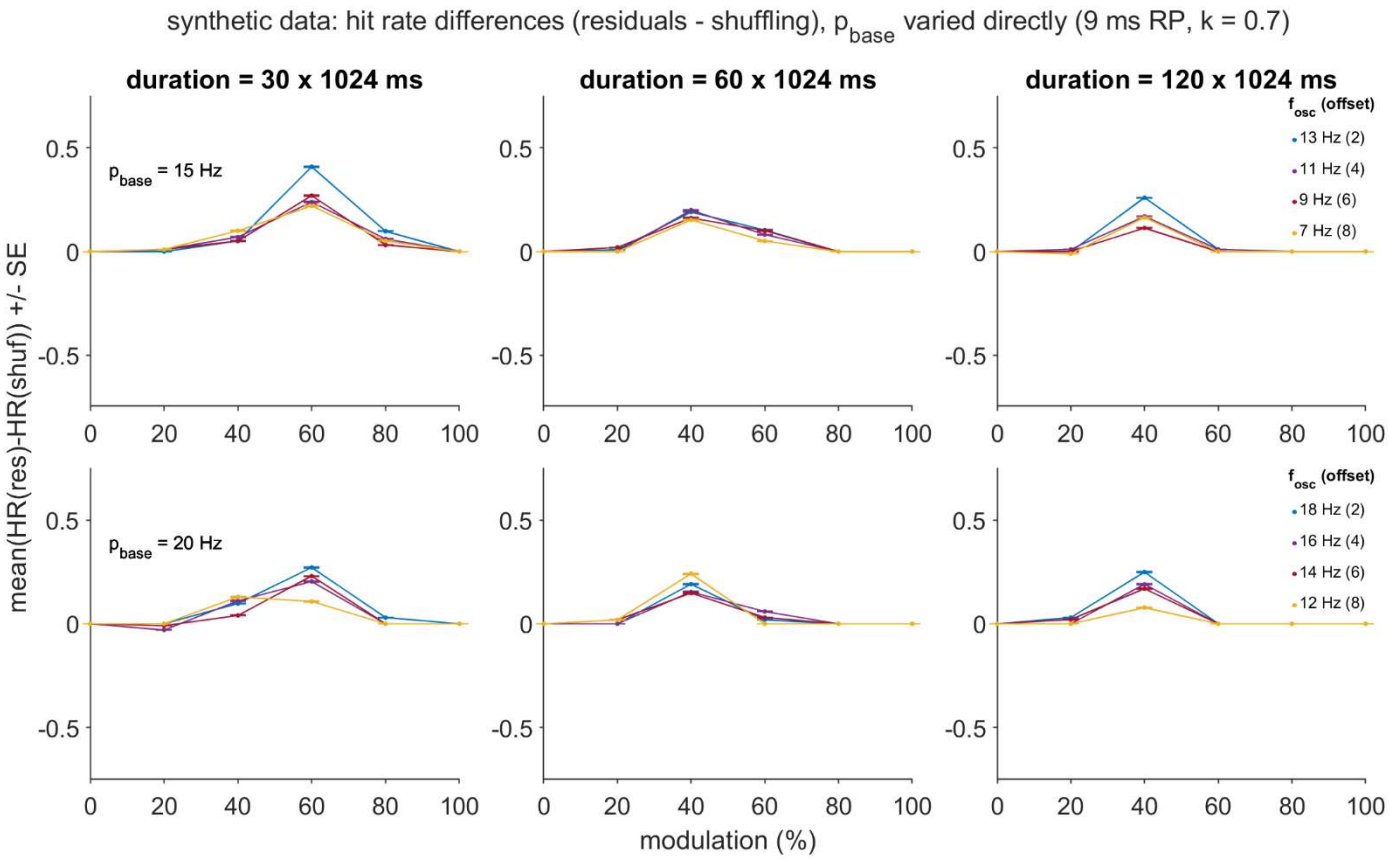



**Fig. S4. Residuals - shuffling difference in hit rates over the varied parameters of the dataset formed with direct base firing rate manipulation.** Means and standard errors reflect summaries over the hit rates computed for each of 1000 subsamples of a dataset in which the $p_{base}$ (i.e., base FR) and $p_{base_offset}$ parameters were varied directly (as opposed to $p_{base_offset}$ and $f_{osc}$). Plots depicts all 144 unique combinations of $p_{base}$ (rows), simulation duration (*T*, columns), oscillation modulation strength (*m*, *x* axes), and oscillation frequency ($f_{osc}$, lines). All other abbreviations follow from those described for Fig. S2. See the Materials and Methods for a description of the subsampling procedure for this dataset.

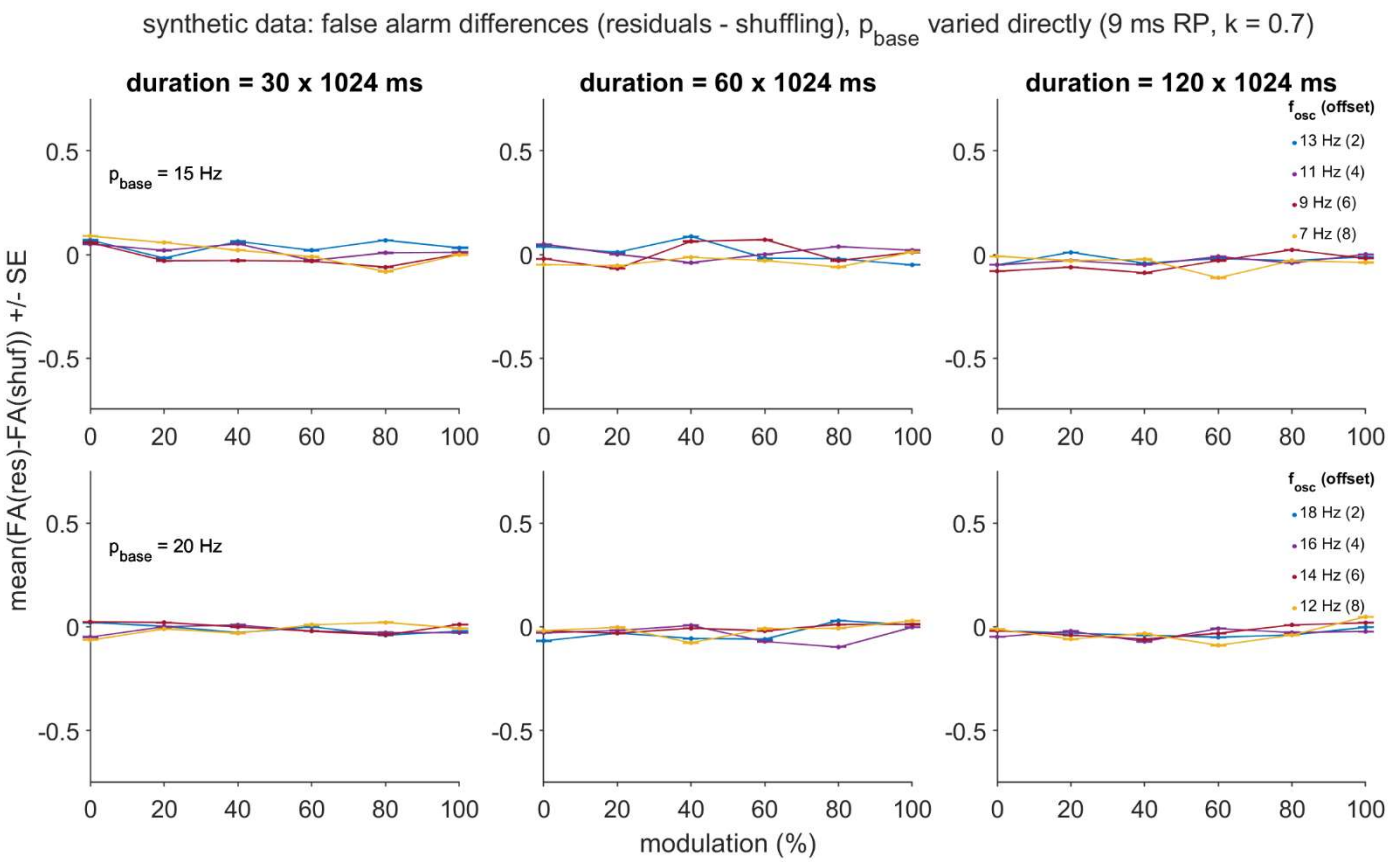



**Fig. S5. Residuals – shuffling difference in false alarm rates over the varied parameters of the dataset formed with direct base firing rate manipulation.** Means and standard errors reflect summaries over the false alarm rates computed for each of 1000 subsamples of a dataset in which the $p_{base}$ and $p_{base_offset}$ parameters were varied directly. All other abbreviations, the plotting conventions, and the subsampling procedure follow from those described for Fig. S3-S4.

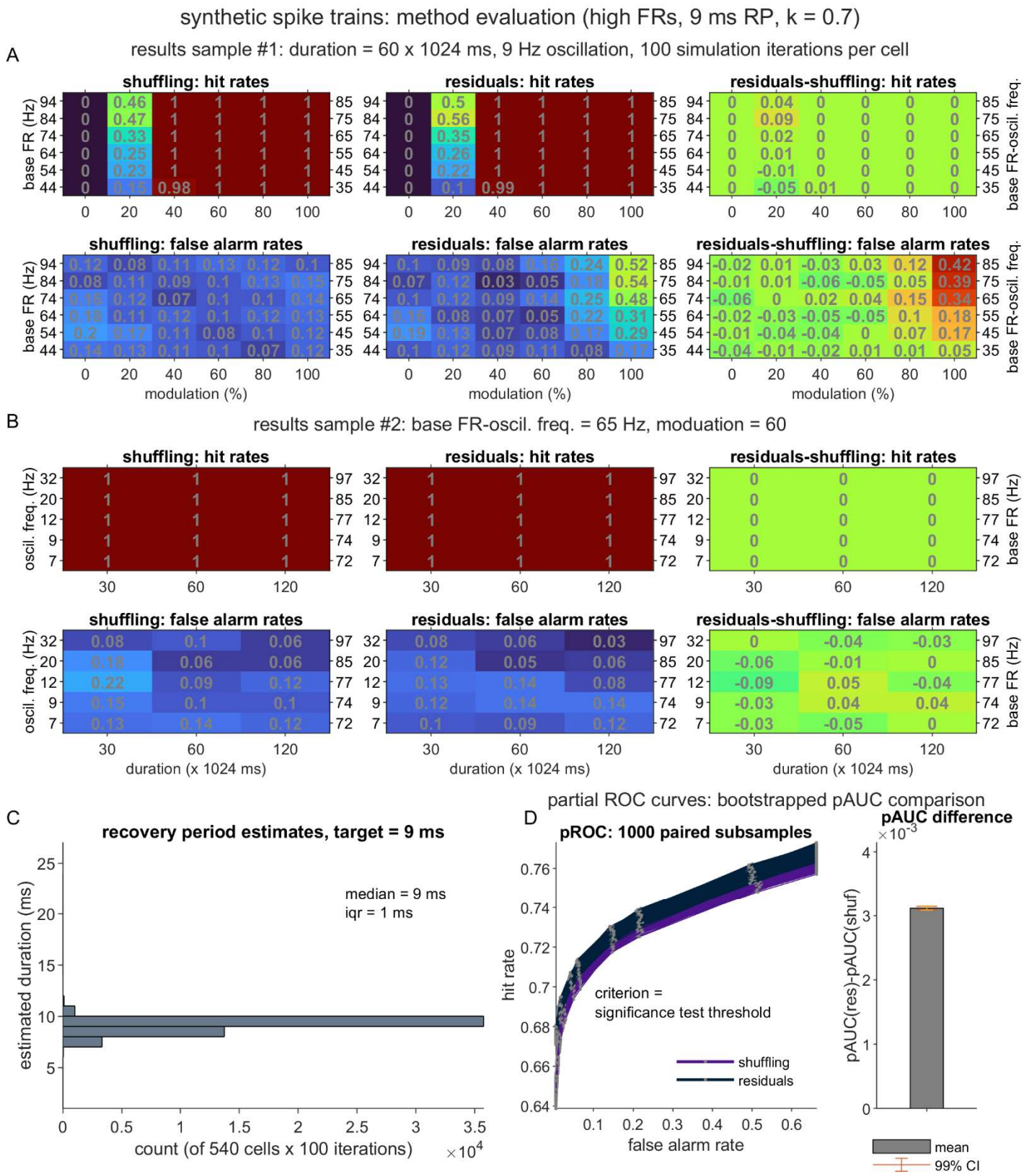

synthetic spike trains: method evaluation (high FRs, 9 ms RP, k = 0.7)
A
results sample #1: duration = 60 x 1024 ms, 9 Hz oscillation, 100 simulation iterations per cell
shuffling: hit rates
residuals: hit rates
residuals-shuffling: hit rates
shuffling: false alarm rates
residuals: false alarm rates
residuals-shuffling: false alarm rates
base FR (Hz)
base FR-oscil. freq.
modulation (%)
B
results sample #2: base FR-oscil. freq. = 65 Hz, moduation = 60
oscil. freq. (Hz)
base FR (Hz)
duration (x 1024 ms)
C
recovery period estimates, target = 9 ms
median = 9 ms
iqr = 1 ms
estimated duration (ms)
count (of 540 cells x 100 iterations)
D
partial ROC curves: bootstrapped pAUC comparison
pROC: 1000 paired subsamples
hit rate
false alarm rate
criterion =
significance test threshold
shuffling
residuals
pAUC difference
pAUC(res)-pAUC(shuf)
mean
99% CI

**Fig. S6. Performance of the shuffling and residuals methods over a synthetic dataset of high firing rate (FR) spike trains.** Panels (A)-(D): Plotting conventions, hit and false alarm definitions, and analysis procedures are identical to those described for the primary dataset depicted in Fig. 4. Relative to the primary dataset, this high FR dataset differed in the use of greater $p_{base_offset}$ values (see "base FR - oscil. freq." tick labels in Panel (A)).

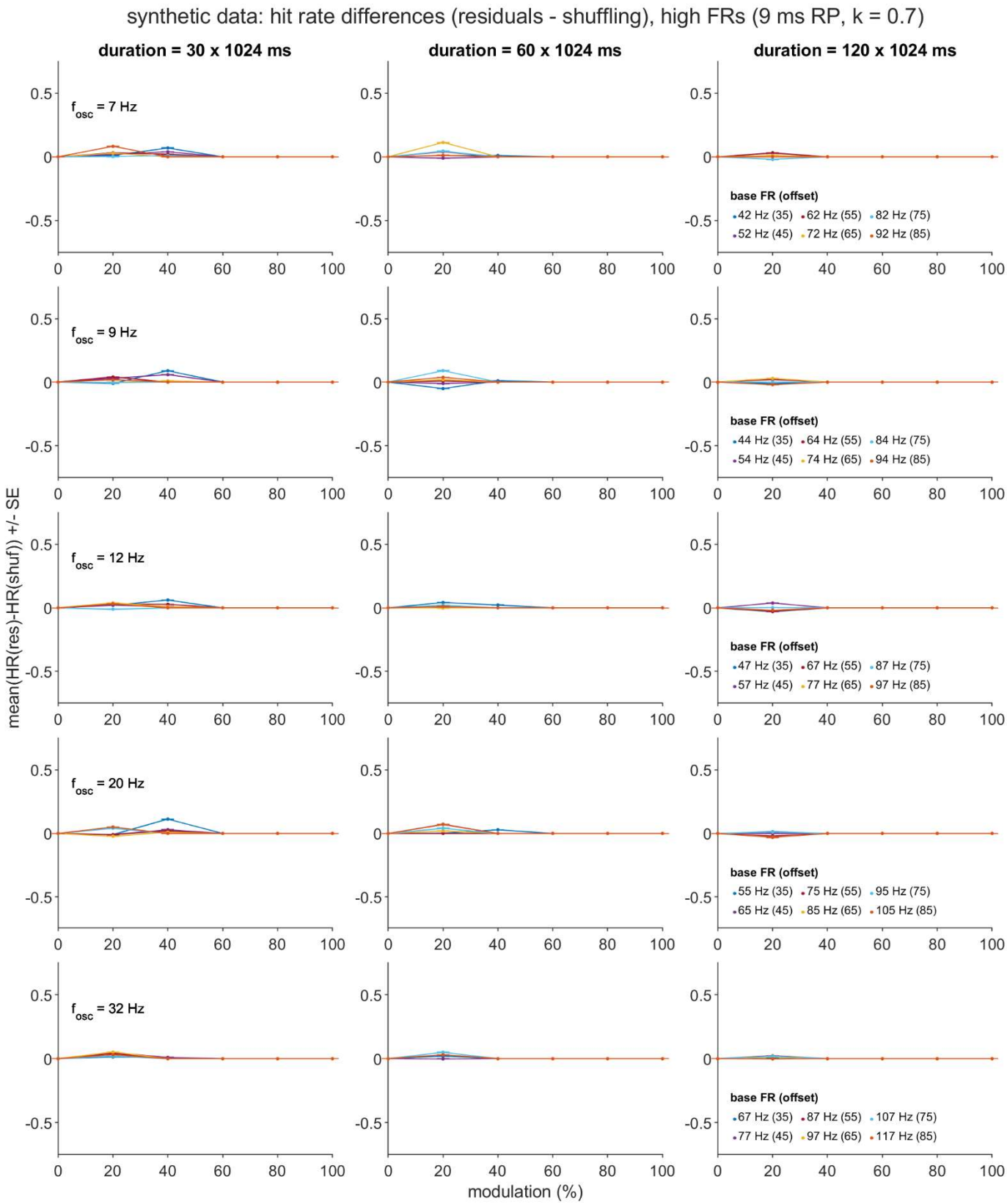

synthetic data: hit rate differences (residuals - shuffling), high FRs (9 ms RP, k = 0.7)
duration = 30 x 1024 ms
duration = 60 x 1024 ms
duration = 120 x 1024 ms
f_osc = 7 Hz
f_osc = 9 Hz
f_osc = 12 Hz
f_osc = 20 Hz
f_osc = 32 Hz
base FR (offset)
42 Hz (35) 62 Hz (55) 82 Hz (75)
52 Hz (45) 72 Hz (65) 92 Hz (85)
base FR (offset)
44 Hz (35) 64 Hz (55) 84 Hz (75)
54 Hz (45) 74 Hz (65) 94 Hz (85)
base FR (offset)
47 Hz (35) 67 Hz (55) 87 Hz (75)
57 Hz (45) 77 Hz (65) 97 Hz (85)
base FR (offset)
55 Hz (35) 75 Hz (55) 95 Hz (75)
65 Hz (45) 85 Hz (65) 105 Hz (85)
base FR (offset)
67 Hz (35) 87 Hz (55) 107 Hz (75)
77 Hz (45) 97 Hz (65) 117 Hz (85)
mean(HR(res)-HR(shuf)) +/- SE
modulation (%)

**Fig. S7. Residuals - shuffling difference in hit rates over the varied parameters of the dataset of high firing rate spike trains.** Means and standard errors reflect summaries over the hit rates computed for each of 1000 subsamples of the original, high FR dataset (see Fig. S6 for details). Abbreviations and plotting conventions follow from those described for Fig. S2.

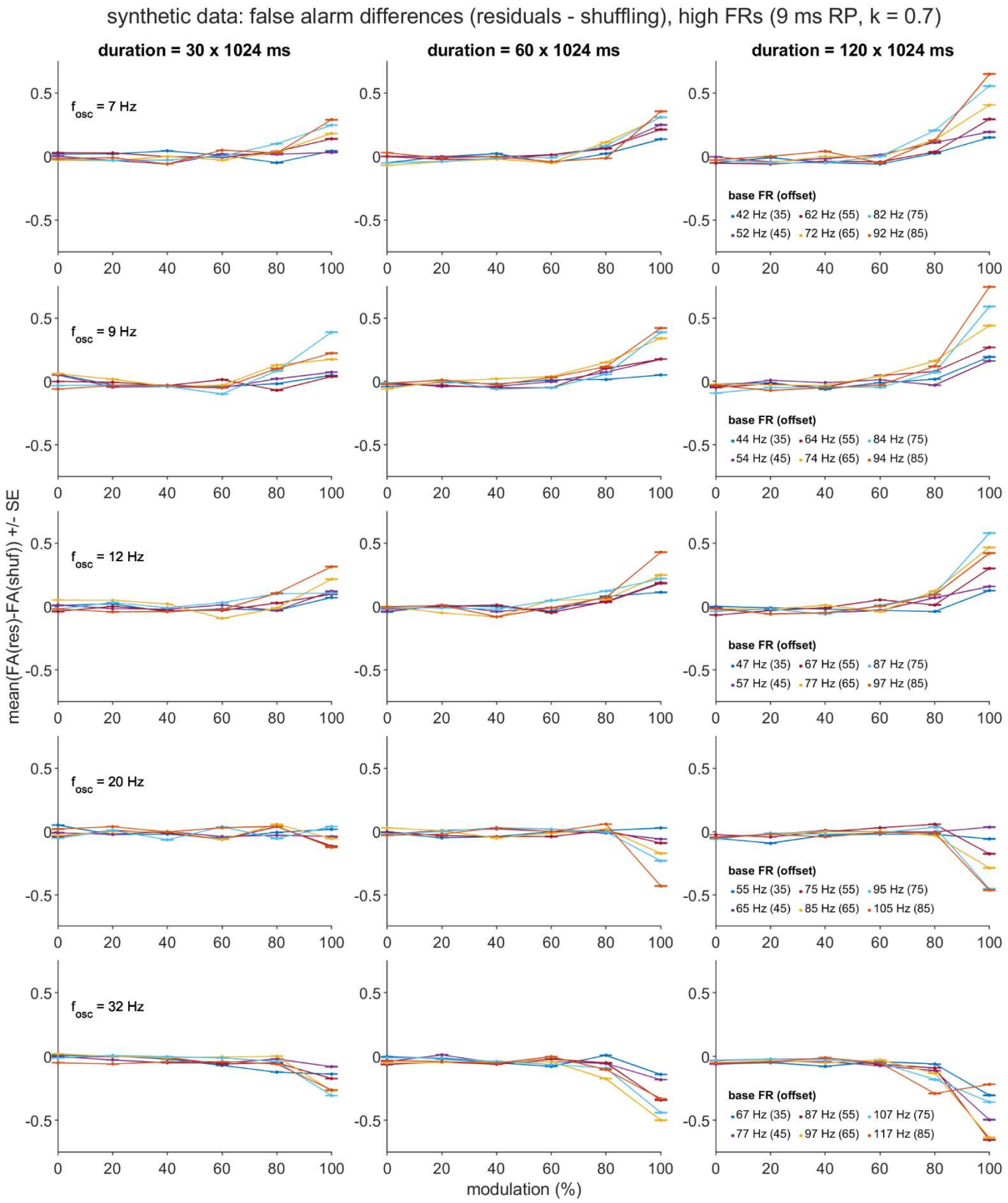

synthetic data: false alarm differences (residuals - shuffling), high FRs (9 ms RP, k = 0.7)
duration = 30 x 1024 ms
duration = 60 x 1024 ms
duration = 120 x 1024 ms
f_osc = 7 Hz
f_osc = 9 Hz
f_osc = 12 Hz
f_osc = 20 Hz
f_osc = 32 Hz
base FR (offset)
42 Hz (35) 62 Hz (55) 82 Hz (75)
52 Hz (45) 72 Hz (65) 92 Hz (85)
44 Hz (35) 64 Hz (55) 84 Hz (75)
54 Hz (45) 74 Hz (65) 94 Hz (85)
47 Hz (35) 67 Hz (55) 87 Hz (75)
57 Hz (45) 77 Hz (65) 97 Hz (85)
55 Hz (35) 75 Hz (55) 95 Hz (75)
65 Hz (45) 85 Hz (65) 105 Hz (85)
67 Hz (35) 87 Hz (55) 107 Hz (75)
77 Hz (45) 97 Hz (65) 117 Hz (85)
mean(FA(res)-FA(shuf)) +/- SE
modulation (%)

**Fig. S8. Residuals - shuffling difference in false alarms over the varied parameters of the dataset of high firing rate spike trains.** Means and standard errors reflect summaries over the false alarm rates computed for each of 1000 subsamples of the original, high FR dataset (see Fig. S6 for details). Abbreviations and plotting conventions follow from those described for Fig. S3.

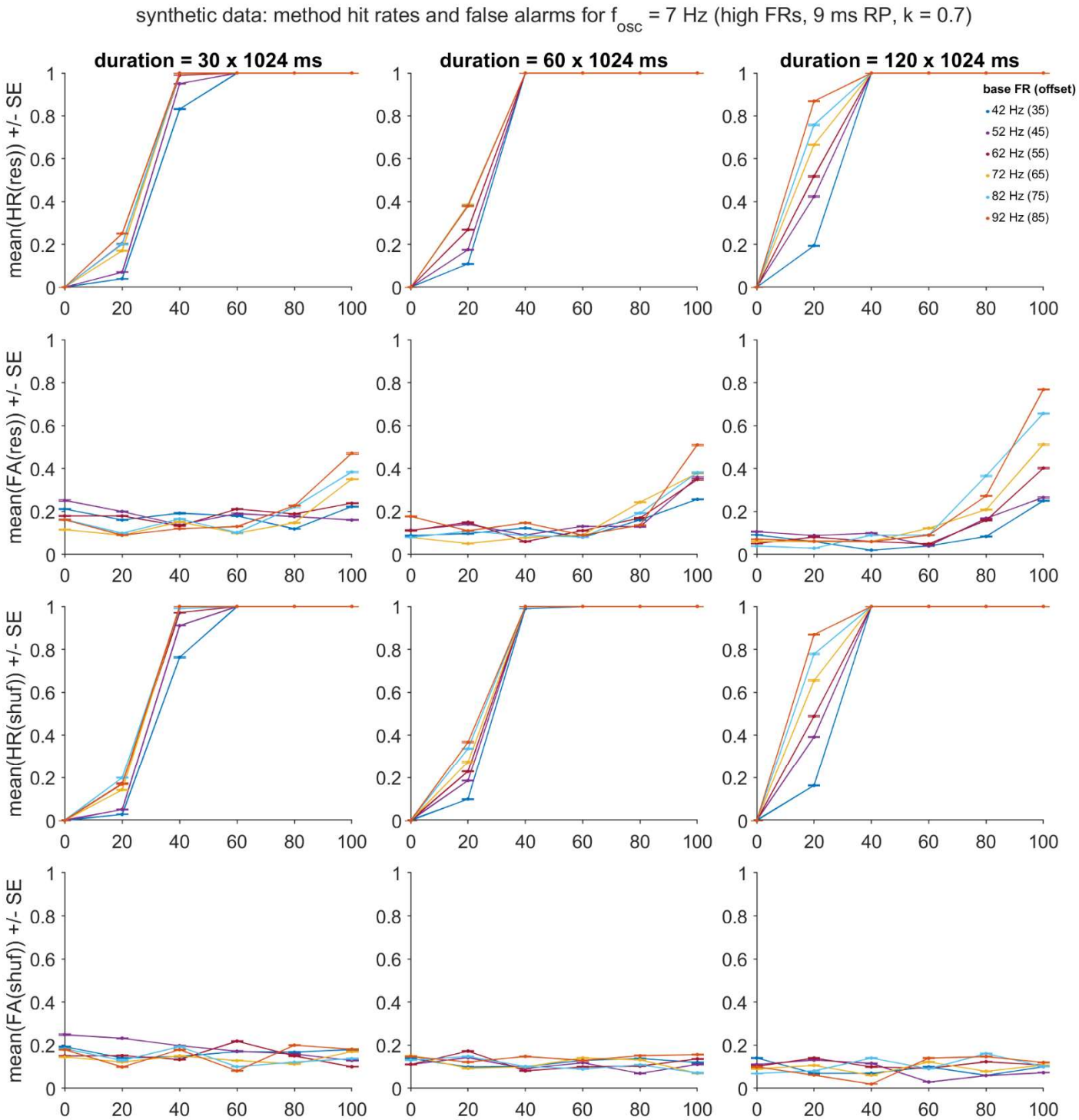



**Fig. S9. Residuals and shuffling hit and false alarm rates for the dataset of high firing rate spike trains, with the oscillation frequency ($f_{osc}$) fixed at 7 Hz.** Means and standard errors reflect summaries over the hit and false alarm rates computed for each of 1000 subsamples of the original, high FR dataset. Plots depict all 108 unique combinations of simulation duration (*T*, columns), oscillation modulation strength (*m*, *x* axes) and base FR - oscillation frequency offset ($p_{base_offset}$, lines) for the cases in which $f_{osc}$ = 7 Hz. Abbreviations follow from those described for Figs. S2-S3.

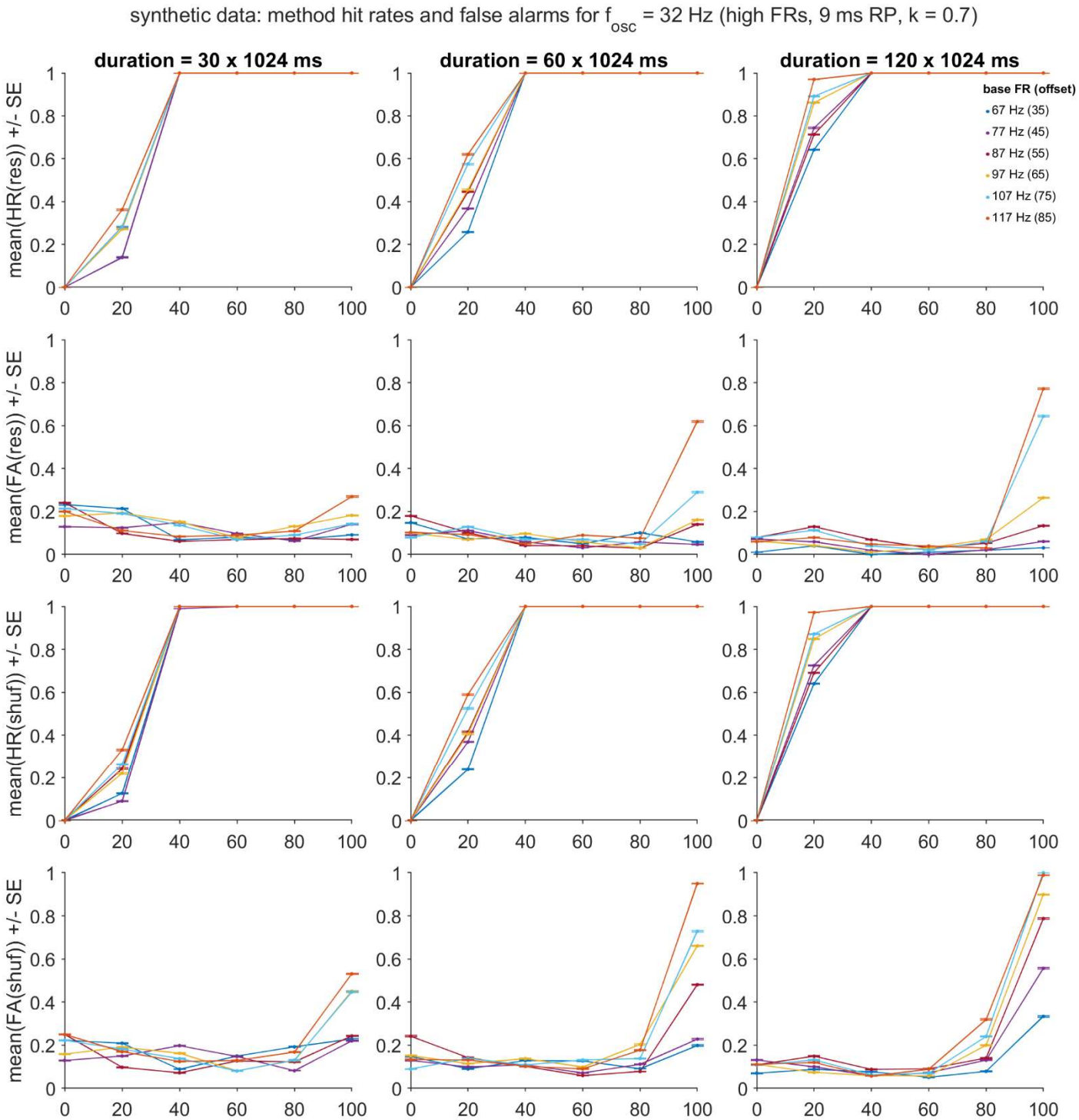



**Fig. S10. Residuals and shuffling hit and false alarm rates for the dataset of high firing rate spike trains, with the oscillation frequency ($f_{osc}$) fixed at 32 Hz.** Abbreviations and plotting conventions follow from those described for Fig. S2-S3 and S9.

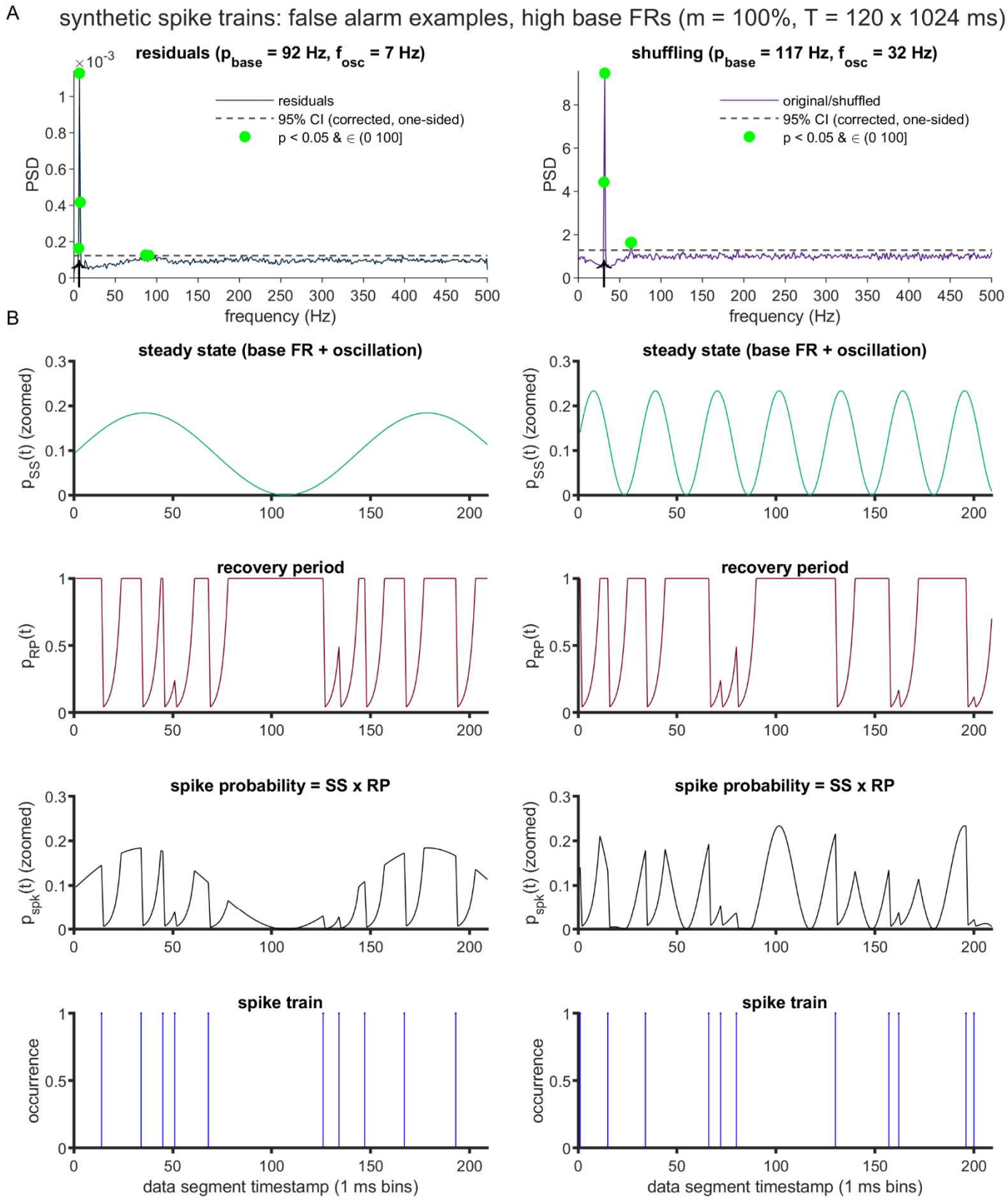
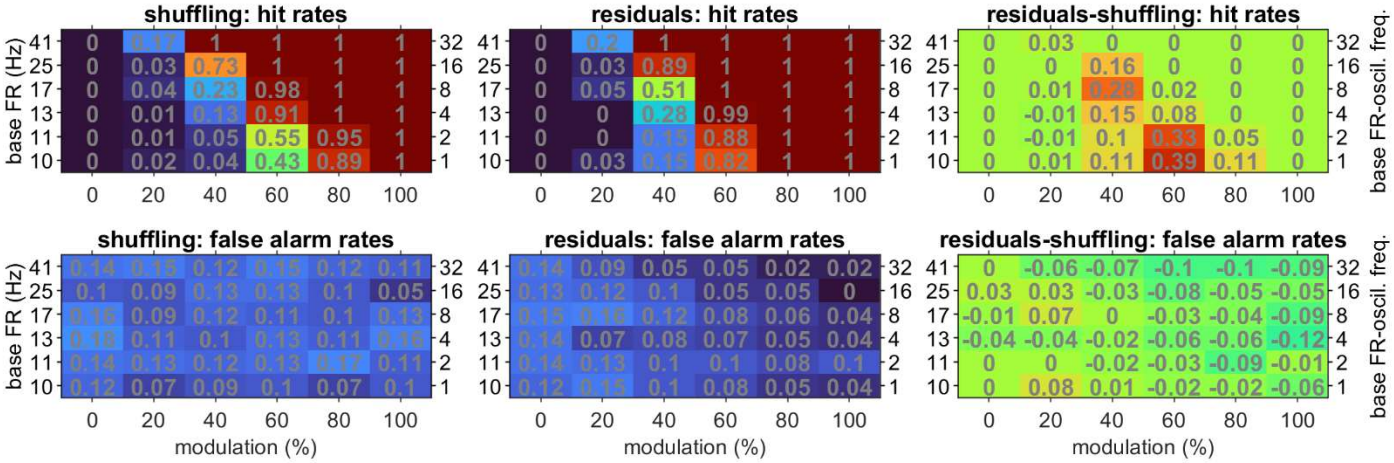

A
synthetic spike trains: false alarm examples, high base FRs (m = 100%, T = 120 x 1024 ms)
residuals ($p_{base}$ = 92 Hz, $f_{osc}$ = 7 Hz)
shuffling ($p_{base}$ = 117 Hz, $f_{osc}$ = 32 Hz)
residuals
95% CI (corrected, one-sided)
p < 0.05 & ∈ (0 100]
original/shuffled
PSD
frequency (Hz)
B
steady state (base FR + oscillation)
$p_{SS}(t)$ (zoomed)
recovery period
$p_{RP}(t)$
spike probability = SS x RP
$p_{spk}(t)$ (zoomed)
spike train
occurrence
data segment timestamp (1 ms bins)

**Fig. S11. Example power spectra and spike train segments for high firing rate cases that produced pronounced false alarms following residuals or shuffling correction.** Two synthetic spike trains were generated to illustrate conditions that yielded especially high false alarm rates under residuals and shuffling correction, respectively. Both spike trains utilized the maximal modulation strength ($m$ = 1.0), duration ($T$ = 120 x 1024 ms), and base FR - oscillation frequency offset ($p_{base_offset}$ = 85 Hz) settings from the original high FR dataset, and the default RP parameters (duration $n_r$ = 9 ms, steepness $k$ = 0.7). Oscillation frequency was set at either the lowest setting from the full high FR dataset (7 Hz, to illustrate residuals FAs) or the highest setting (32 Hz, to illustrate shuffling FAs). (A) Corrected power spectral density (PSD) functions generated by the residuals method (left) and shuffling method (right). Statistical testing and plotting conventions follow from those described for Fig. 2-3. (B) Illustration of the initial 209 ms of the synthetic spike trains (fourth row) from which the PSDs in (A) were computed, and the components of the rate function that governed their generation (first-third rows; see Materials and Methods, Eq. 1 for full details). Term definitions: $p_{ss}$(t) = steady state spiking probability, $p_{RP}$(t) = recovery period spiking probability, $p_{spk}$(t) = spike probability, reflecting the product of the steady state and recovery period components.

synthetic spike trains: method evaluation (low to moderate FRs, 3 ms RP, k = 0.0)

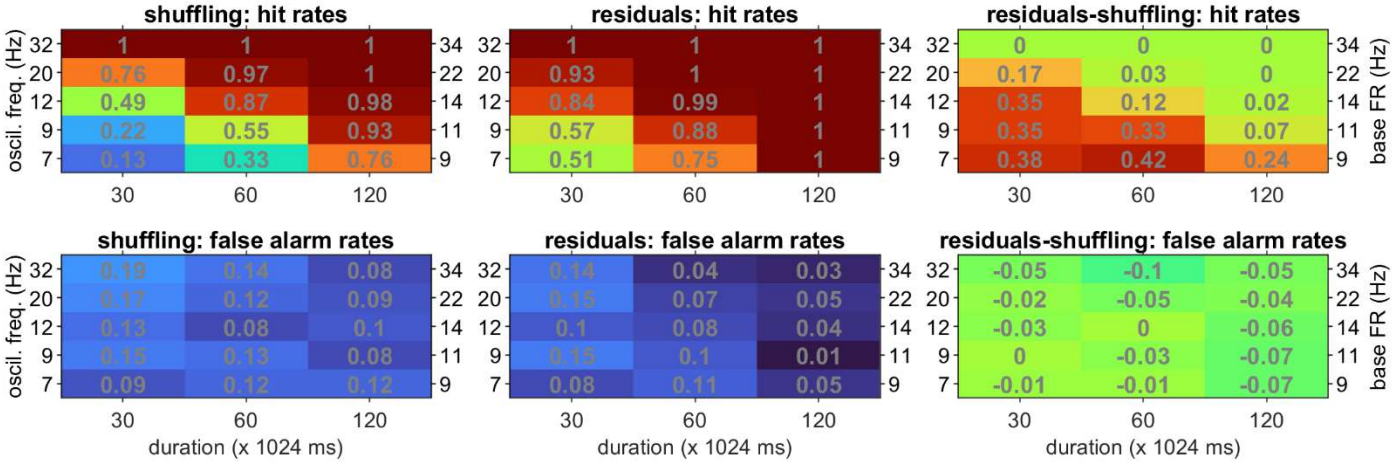
A

results sample #1: duration = 60 x 1024 ms, 9 Hz oscillation, 100 simulation iterations per cell

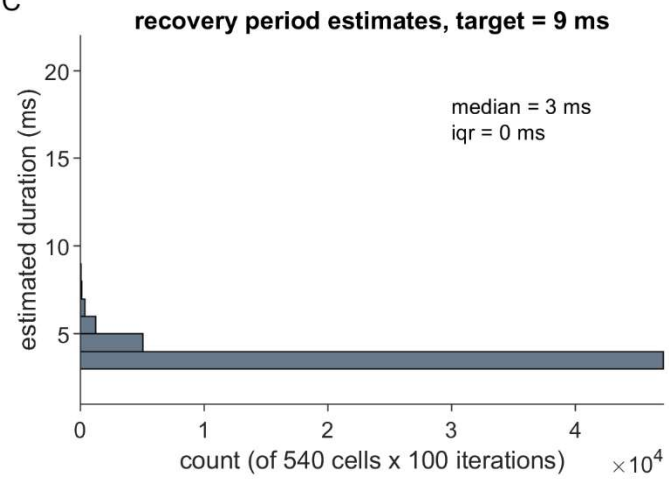



B

results sample #2: base FR-oscil. freq. = 2 Hz, moduation = 60%, 100 simulation iterations per cell




C




D

partial ROC curves: bootstrapped pAUC comparison

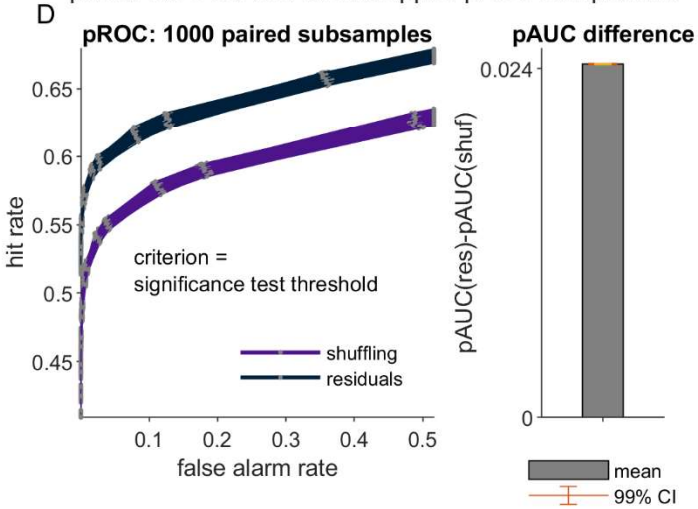


**Fig. S12. Performance of the shuffling and residuals methods over a synthetic dataset generated using a short, absolute recovery period.** Panels (A)-(D): Plotting conventions, hit and false alarm definitions, and analysis procedures are identical to those described for the primary dataset depicted in Fig. 4. Relative to the primary dataset, this secondary dataset differed in the use of an absolute, as opposed to relative RP ($k$ = 0), and a shorter RP duration ($n_r$ = 3 ms).

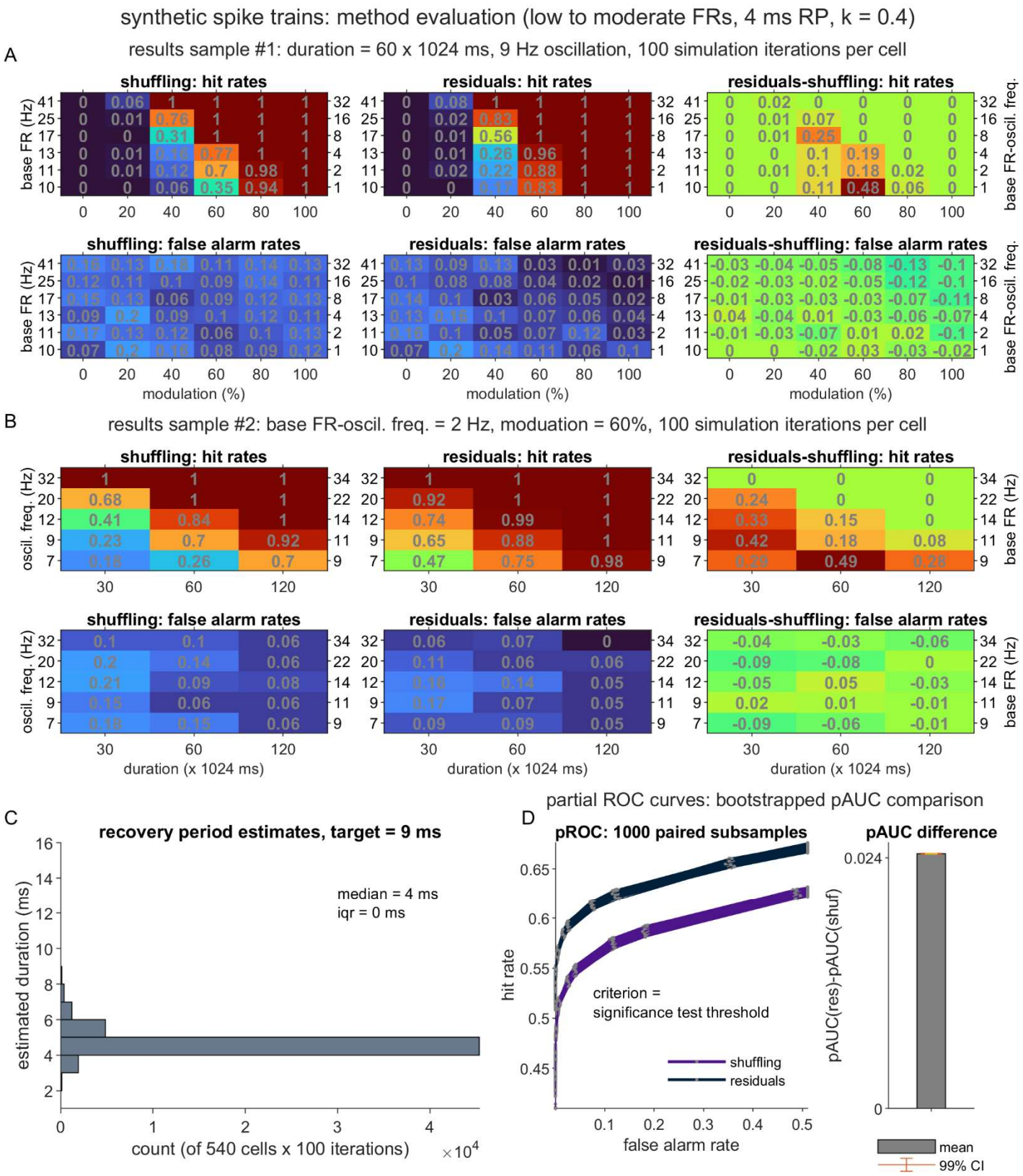
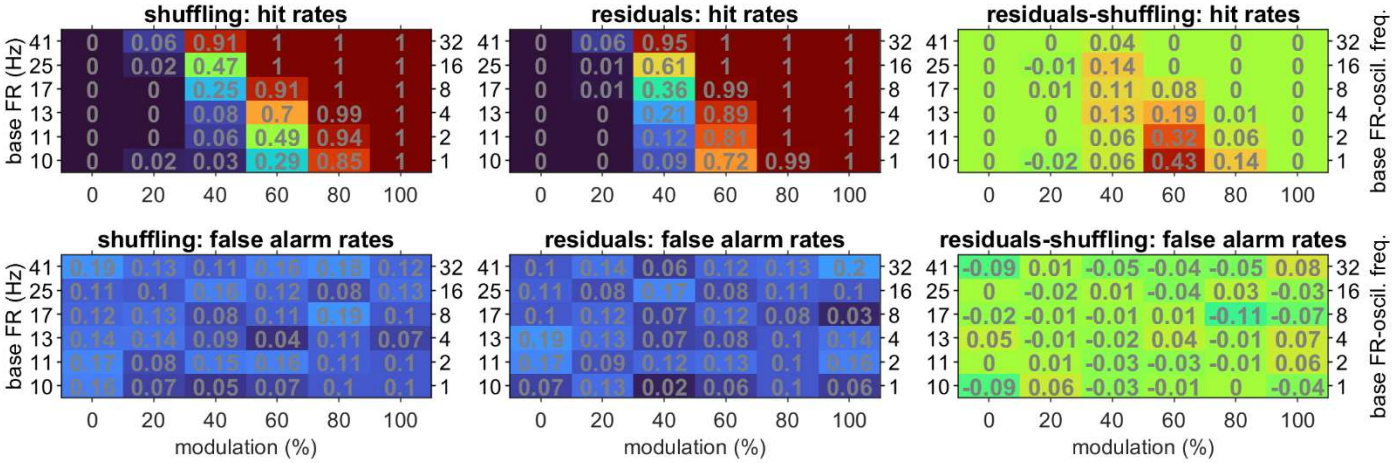

synthetic spike trains: method evaluation (low to moderate FRs, 4 ms RP, k = 0.4)
A
results sample #1: duration = 60 x 1024 ms, 9 Hz oscillation, 100 simulation iterations per cell
shuffling: hit rates
residuals: hit rates
residuals-shuffling: hit rates
shuffling: false alarm rates
residuals: false alarm rates
residuals-shuffling: false alarm rates
base FR (Hz)
base FR-oscil. freq.
modulation (%)
B
results sample #2: base FR-oscil. freq. = 2 Hz, moduation = 60%, 100 simulation iterations per cell
oscil. freq. (Hz)
base FR (Hz)
duration (x 1024 ms)
C
recovery period estimates, target = 9 ms
median = 4 ms
iqr = 0 ms
estimated duration (ms)
count (of 540 cells x 100 iterations)
×10^4
D
partial ROC curves: bootstrapped pAUC comparison
pROC: 1000 paired subsamples
hit rate
false alarm rate
criterion =
significance test threshold
shuffling
residuals
pAUC difference
pAUC(res)-pAUC(shuf)
mean
99% CI

**Fig. S13. Performance of the shuffling and residuals methods over a synthetic dataset generated using a shortened relative recovery period.** Panels (A)-(D): Plotting conventions, hit and false alarm definitions, and analysis procedures are identical to those described for the primary dataset depicted in Fig. 4. Relative to the primary dataset, this secondary dataset differed in the use of a shortened and steeper relative RP ($n_r$ = 4 ms, $k$ = 0.4).

synthetic spike trains: method evaluation (low to moderate FRs, 18 ms RP, k = 0.7)

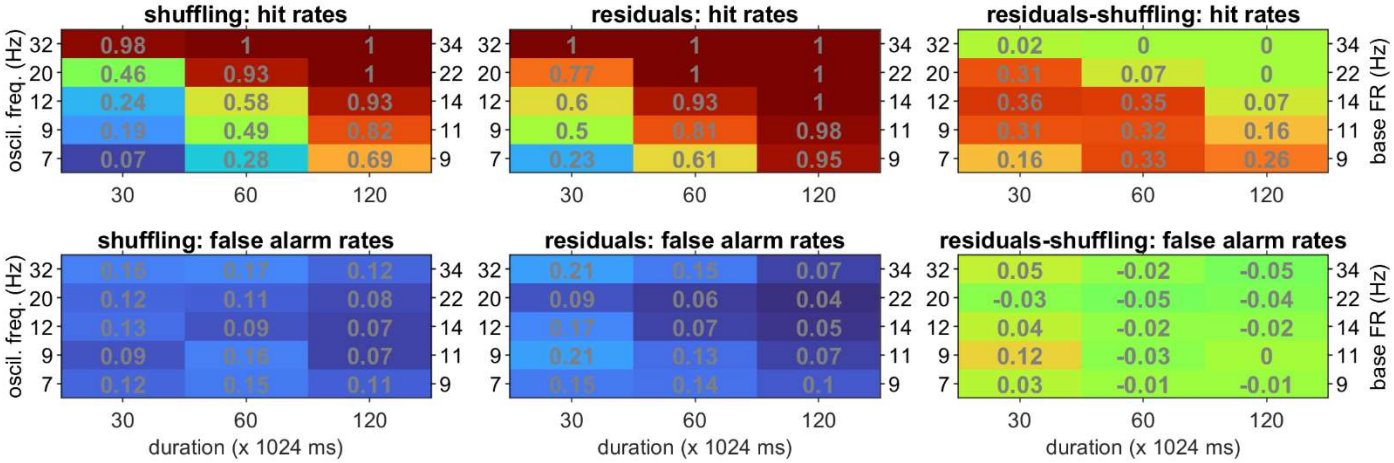
A

results sample #1: duration = 60 x 1024 ms, 9 Hz oscillation, 100 simulation iterations per cell

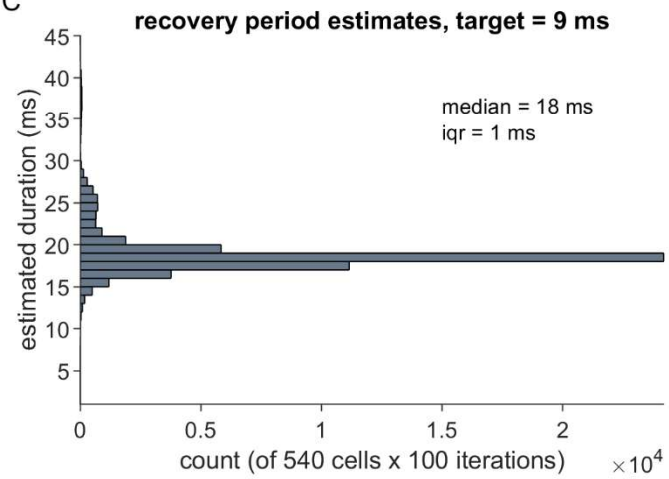



B

results sample #2: base FR-oscil. freq. = 2 Hz, moduation = 60%, 100 simulation iterations per cell




C




partial ROC curves: bootstrapped pAUC comparison

D

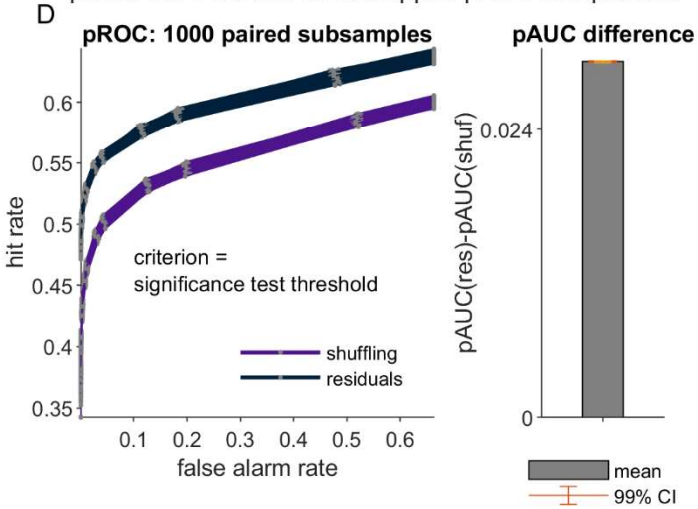


**Fig. S14. Performance of the shuffling and residuals methods over a synthetic dataset generated using a lengthened relative recovery period.** Panels (A)-(D): Plotting conventions, hit and false alarm definitions, and analysis procedures are identical to those described for the primary dataset depicted in Fig. 4. Relative to the primary dataset, this secondary dataset differed in the use of a longer relative RP ($n_r$ = 18 ms, $k$ = 0.7).

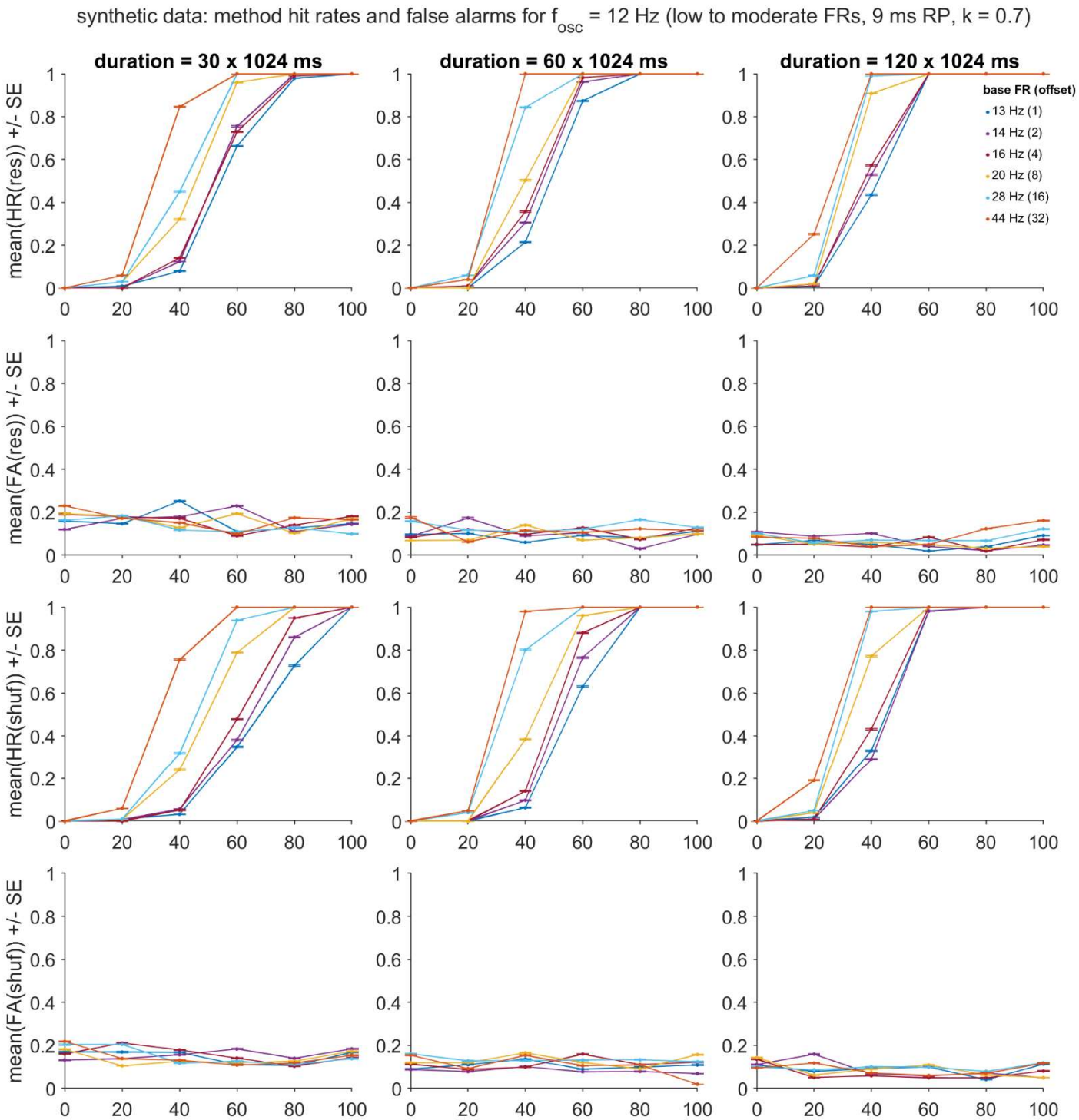



**Fig. S15. Residuals and shuffling hit and false alarm rates for the primary synthetic dataset, with the oscillation frequency ($f_{osc}$) fixed at 12 Hz.** Abbreviations and plotting conventions follow from those described for Fig. S9-S10.

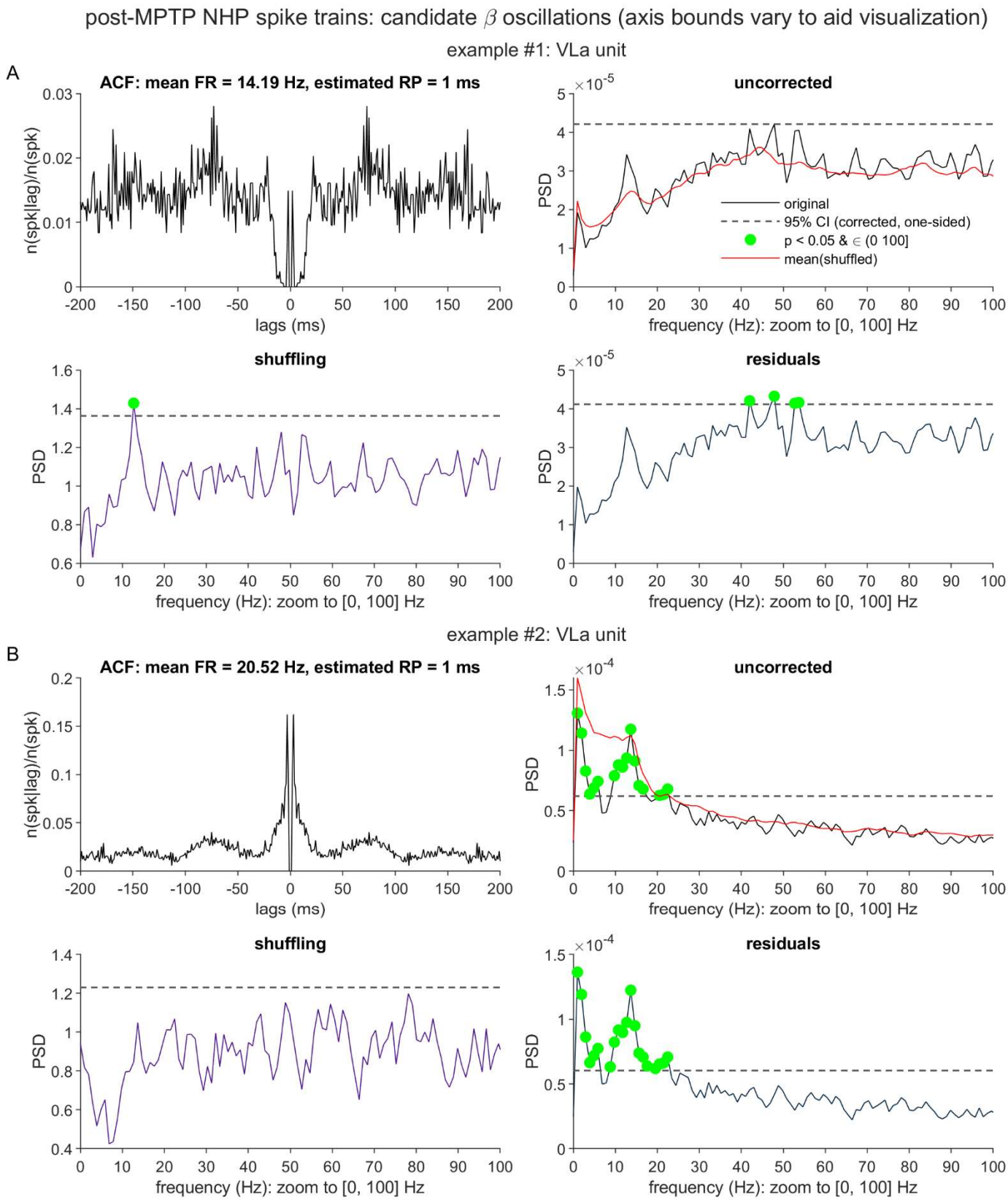

post-MPTP NHP spike trains: candidate β oscillations (axis bounds vary to aid visualization)
example #1: VLa unit
A
ACF: mean FR = 14.19 Hz, estimated RP = 1 ms
n(spk|lag)/n(spk)
lags (ms)
uncorrected
PSD
original
95% CI (corrected, one-sided)
p < 0.05 & ∈ (0 100]
mean(shuffled)
frequency (Hz): zoom to [0, 100] Hz
shuffling
residuals
example #2: VLa unit
B
ACF: mean FR = 20.52 Hz, estimated RP = 1 ms

**Fig. S16. Comparison of the shuffling and residuals output for two units with putative beta oscillations and non-oscillatory features.** Spike trains originated from two ventrolateral anterior thalamus (VLa) units, which had been recorded from the same parkinsonian non-human primate (NHP) that contributed to Fig. 5-6. Panels (A)-(B): Abbreviations and the statistical and plotting conventions follow from those described for Fig. 5.

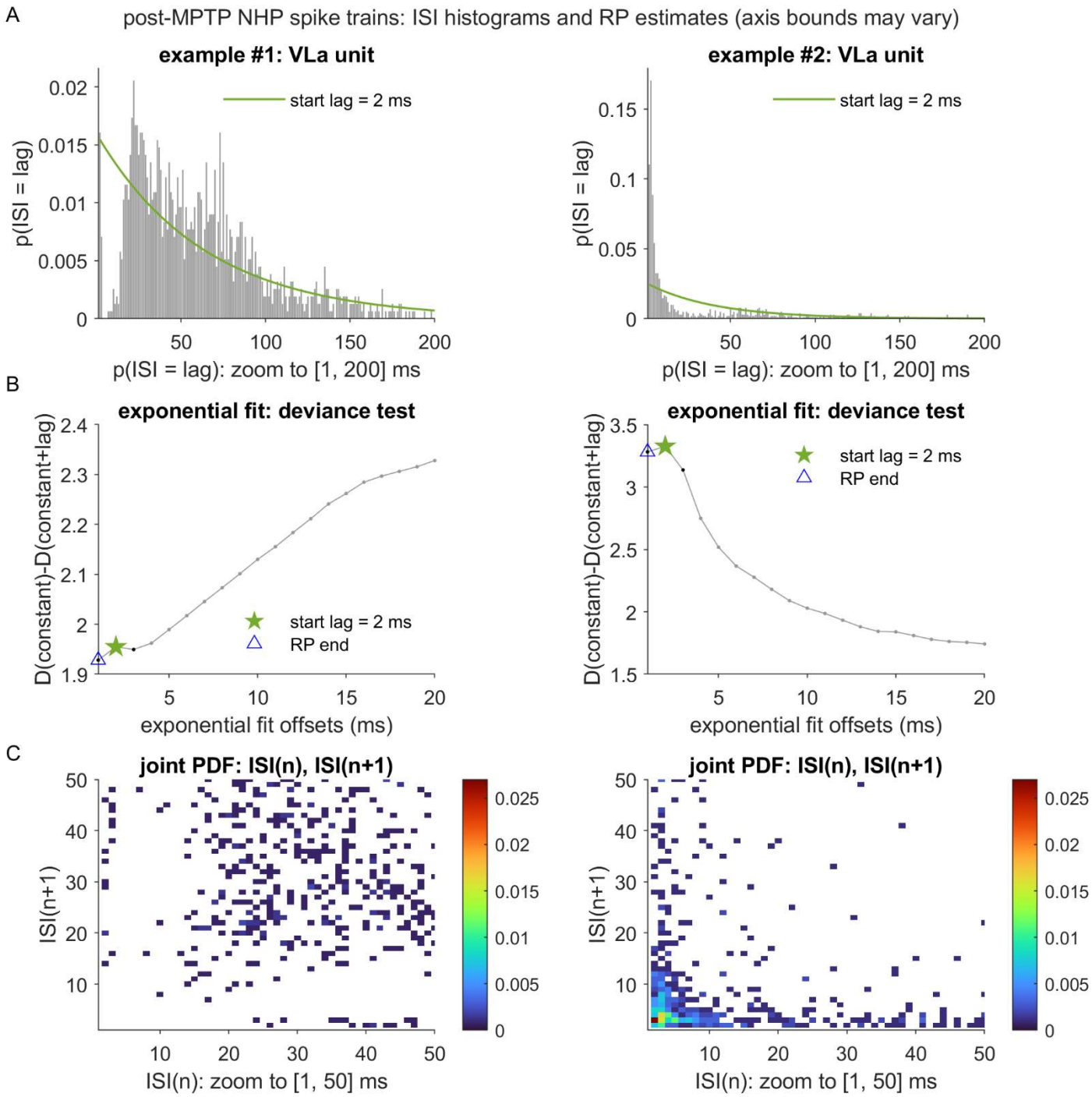



**Fig. S17. Inter-spike interval distributions and RP duration estimation for the two VLa units with suspected oscillatory and non-oscillatory features.** The left and right columns depict the same two empirical spike trains as were presented in Fig. S16A-B. (A) Illustration of the procedure for obtaining the estimated RP duration ($\hat{n}_r$) for the two spike trains. See Fig. 3A-B and the Materials and Methods for full details. For each unit, the figure highlights the exponential curve with the left anchor position ("start lag") that returned the first local maximum in the curve goodness-of-fit plot. (B) Plots of the goodness-

of-fit measure (the deviance difference statistic, ΔD) as a function of the first 20 starting positions of the exponential fits. Each $\hat{n}_r$ estimate (blue triangles) was set equal to the post-spike lag immediately preceding the first local maximum in the corresponding ΔD plot (green stars). (C) Joint probability density function (PDF) for all consecutive [$ISI_n$, $ISI_{n+1}$] inter-spike interval pairs. All remaining abbreviations follow from those described in Fig. 3 and Fig. 5.

## Supporting Tables

Table S1. Parametric Effects on Residuals-Shuffling Hit Rates

Low-to-Moderate Firing Rates (9 ms RP, k = 0.7)

| Effect | Beta | SE | t | p |
|---|---|---|---|---|
| Intercept | 0.084 | 0.00017 | 494.17 | -- |
| T | -0.00086 | 4.6e-06 | -188.48 | -- |
| $f_{osc}$ | -0.0049 | 1.9e-05 | -263.45 | -- |
| $p_{base_offset}$ | -0.0046 | 1.6e-05 | -292.25 | -- |
| m | -0.17 | 0.0012 | -143.68 | -- |
| $m^2$ | -0.48 | 0.0016 | -296.41 | -- |
| T x $f_{osc}$ | 3.4e-05 | 5e-07 | 68.25 | -- |
| T x $p_{base_offset}$ | 3.5e-05 | 4.2e-07 | 82.42 | -- |
| $f_{osc}$ x $p_{base_offset}$ | 0.00027 | 1.7e-06 | 155.21 | -- |
| T x m | -0.0024 | 3.1e-05 | -77.28 | -- |
| $f_{osc}$ x m | -0.0028 | 0.00013 | -21.75 | 7.6e-105 |
| $p_{base_offset}$ x m | 0.00057 | 0.00011 | 5.29 | 1.2e-07 |
| T x $m^2$ | 0.0049 | 4.4e-05 | 112.53 | -- |
| $f_{osc}$ x $m^2$ | 0.03 | 0.00018 | 168.36 | -- |
| $p_{base_offset}$ x $m^2$ | 0.029 | 0.00015 | 191.37 | -- |
| m x $m^2$ | 1.1 | 0.0081 | 132.67 | -- |
| T x $f_{osc}$ x $p_{base_offset}$ | -9.9e-07 | 3e-08 | -33.48 | 2e-245 |
| T x $f_{osc}$ x m | 8.4e-05 | 1.1e-06 | 73.92 | -- |
| T x $p_{base_offset}$ x m | 6.4e-05 | 9.5e-07 | 66.88 | -- |
| $f_{osc}$ x $p_{base_offset}$ x m | 0.00017 | 3.9e-06 | 44.01 | -- |
| T x $f_{osc}$ x $m^2$ | -0.00015 | 4.8e-06 | -32.06 | 3.3e-225 |
| T x $p_{base_offset}$ x $m^2$ | -0.00016 | 4e-06 | -40.27 | -- |
| $f_{osc}$ x $p_{base_offset}$ x $m^2$ | -0.0015 | 1.7e-05 | -90.54 | -- |
| T x m x $m^2$ | 0.013 | 0.00022 | 60.29 | -- |
| $f_{osc}$ x m x $m^2$ | 0.0075 | 0.00088 | 8.54 | 1.3e-17 |
| $p_{base_offset}$ x m x $m^2$ | -0.012 | 0.00074 | -16.60 | 7e-62 |

RP = recovery period; T = duration;

$f_{osc}$ = oscillation frequency;

$p_{base_offset}$ = offset of base firing rate from $f_{osc}$;

m = oscillation modulation strength;

-- = outside range of double representation

Table S2. Parametric Effects on Residuals-Shuffling False Alarms

Low-to-Moderate Firing Rates (9 ms RP, k = 0.7)

| Effect | Beta | SE | t | p |
|---|---|---|---|---|
| Intercept | -0.023 | 0.00017 | -138.04 | -- |
| T | -0.00034 | 4.5e-06 | -75.60 | -- |
| $f_{osc}$ | -0.0012 | 1.8e-05 | -64.95 | -- |
| $p_{base_offset}$ | -4.3e-05 | 1.5e-05 | -2.81 | 0.0049 |
| m | -0.0026 | 0.0011 | -2.28 | 0.023 |
| $m^2$ | 0.095 | 0.0016 | 59.43 | -- |
| T x $f_{osc}$ | 1.2e-05 | 4.9e-07 | 24.87 | 1.9e-136 |
| T x $p_{base_offset}$ | 1.1e-05 | 4.1e-07 | 27.19 | 1.2e-162 |
| $f_{osc}$ x $p_{base_offset}$ | 4.3e-06 | 1.7e-06 | 2.54 | 0.011 |
| T x m | 0.001 | 3e-05 | 34.38 | 9.8e-259 |
| $f_{osc}$ x m | -0.0031 | 0.00012 | -25.25 | 1.4e-140 |
| $p_{base_offset}$ x m | 0.0055 | 0.0001 | 52.64 | -- |
| T x $m^2$ | 0.00053 | 4.3e-05 | 12.32 | 6.9e-35 |
| $f_{osc}$ x $m^2$ | -0.0027 | 0.00017 | -15.26 | 1.4e-52 |
| $p_{base_offset}$ x $m^2$ | 0.0031 | 0.00015 | 21.28 | 2.1e-100 |
| m x $m^2$ | 0.031 | 0.0079 | 3.93 | 8.6e-05 |
| T x $f_{osc}$ x $p_{base_offset}$ | -1e-06 | 2.9e-08 | -34.66 | 7.9e-263 |
| T x $f_{osc}$ x m | 2.3e-05 | 1.1e-06 | 20.65 | 1e-94 |
| T x $p_{base_offset}$ x m | 2.1e-05 | 9.3e-07 | 22.26 | 9.5e-110 |
| $f_{osc}$ x $p_{base_offset}$ x m | -0.00025 | 3.8e-06 | -65.86 | -- |
| T x $f_{osc}$ x $m^2$ | 2.1e-05 | 4.7e-06 | 4.39 | 1.1e-05 |
| T x $p_{base_offset}$ x $m^2$ | -1.5e-05 | 3.9e-06 | -3.72 | 0.0002 |
| $f_{osc}$ x $p_{base_offset}$ x $m^2$ | -0.00043 | 1.6e-05 | -26.54 | 4.8e-155 |
| T x m x $m^2$ | -0.0034 | 0.00021 | -16.24 | 2.8e-59 |
| $f_{osc}$ x m x $m^2$ | 0.0067 | 0.00086 | 7.73 | 1e-14 |
| $p_{base_offset}$ x m x $m^2$ | -0.027 | 0.00073 | -37.01 | 2.5e-299 |

RP = recovery period; T = duration;

$f_{osc}$ = oscillation frequency;

$p_{base_offset}$ = offset of base firing rate from $f_{osc}$;

m = oscillation modulation strength;

-- = outside range of double representation

Table S3. Parametric Effects on Residuals-Shuffling Hit Rates

Low-to-Moderate Firing Rates (9 ms RP, k = 0.7)

| Effect | Beta | SE | t | p |
|---|---|---|---|---|
| Intercept | 0.1 | 0.0003 | 341.42 | -- |
| T | -0.00099 | 8e-06 | -124.45 | -- |
| $p_{base}$ | -0.0052 | 0.00012 | -43.21 | -- |
| $p_{base_offset}$ | -0.008 | 0.00013 | -59.81 | -- |
| m | -0.44 | 0.002 | -217.82 | -- |
| $m^2$ | -0.62 | 0.0029 | -217.13 | -- |
| T x $p_{base}$ | 8.9e-05 | 3.2e-06 | 27.85 | 3.5e-170 |
| T x $p_{base_offset}$ | 4e-05 | 3.6e-06 | 11.29 | 1.5e-29 |
| $p_{base}$ x $p_{base_offset}$ | 0.00036 | 5.3e-05 | 6.66 | 2.7e-11 |
| T x m | -0.004 | 5.4e-05 | -72.62 | -- |
| $p_{base}$ x m | -0.018 | 0.00082 | -22.29 | 7.4e-110 |
| $p_{base_offset}$ x m | 0.013 | 0.00091 | 14.57 | 5e-48 |
| T x $m^2$ | 0.0079 | 7.7e-05 | 103.32 | -- |
| $p_{base}$ x $m^2$ | 0.036 | 0.0011 | 31.79 | 6.4e-221 |
| $p_{base_offset}$ x $m^2$ | 0.053 | 0.0013 | 41.28 | -- |
| m x $m^2$ | 2.7 | 0.014 | 193.84 | -- |
| T x $p_{base}$ x $p_{base_offset}$ | -7.2e-06 | 9.2e-07 | -7.81 | 5.8e-15 |
| T x $p_{base}$ x m | 2.2e-05 | 7.3e-06 | 3.03 | 0.0025 |
| T x $p_{base_offset}$ x m | 0.00023 | 8.1e-06 | 28.19 | 3e-174 |
| $p_{base}$ x $p_{base_offset}$ x m | -4.1e-05 | 0.00012 | -0.34 | 0.74 |
| T x $p_{base}$ x $m^2$ | -0.00045 | 3.1e-05 | -14.64 | 1.8e-48 |
| T x $p_{base_offset}$ x $m^2$ | -0.00062 | 3.4e-05 | -17.96 | 5.3e-72 |
| $p_{base}$ x $p_{base_offset}$ x $m^2$ | -0.0019 | 0.00051 | -3.77 | 0.00016 |
| T x m x $m^2$ | 0.024 | 0.00038 | 64.06 | -- |
| $p_{base}$ x m x $m^2$ | 0.11 | 0.0057 | 20.26 | 4e-91 |
| $p_{base_offset}$ x m x $m^2$ | -0.084 | 0.0063 | -13.21 | 7.7e-40 |

RP = recovery period; T = duration;

$p_{base}$ = base firing rate;

$p_{base_offset}$ = offset of base firing rate from $f_{osc}$;

m = oscillation modulation strength;

-- = outside range of double representation

Table S4. Parametric Effects on Residuals-Shuffling False Alarms

Low-to-Moderate Firing Rates (9 ms RP, k = 0.7)

| Effect | Beta | SE | t | p |
|---|---|---|---|---|
| Intercept | -0.022 | 0.00033 | -67.70 | -- |
| T | -0.00039 | 8.8e-06 | -44.08 | -- |
| $p_{base}$ | -0.0039 | 0.00013 | -29.52 | 8.9e-191 |
| $p_{base_offset}$ | -0.0036 | 0.00015 | -24.56 | 6.4e-133 |
| m | -0.018 | 0.0022 | -8.17 | 3e-16 |
| $m^2$ | 0.089 | 0.0032 | 28.06 | 1.2e-172 |
| T x $p_{base}$ | 3.1e-05 | 3.5e-06 | 8.67 | 4.4e-18 |
| T x $p_{base_offset}$ | 1.7e-05 | 3.9e-06 | 4.23 | 2.3e-05 |
| $p_{base}$ x $p_{base_offset}$ | 0.0019 | 5.9e-05 | 32.70 | 1.8e-233 |
| T x m | 0.0019 | 6e-05 | 32.31 | 4.4e-228 |
| $p_{base}$ x m | 0.019 | 0.0009 | 21.40 | 1.9e-101 |
| $p_{base_offset}$ x m | -0.0023 | 0.001 | -2.30 | 0.021 |
| T x $m^2$ | 0.0011 | 8.4e-05 | 12.82 | 1.3e-37 |
| $p_{base}$ x $m^2$ | 0.026 | 0.0013 | 20.63 | 2e-94 |
| $p_{base_offset}$ x $m^2$ | 0.022 | 0.0014 | 15.36 | 3.3e-53 |
| m x $m^2$ | 0.25 | 0.016 | 16.26 | 2.2e-59 |
| T x $p_{base}$ x $p_{base_offset}$ | -1.1e-05 | 1e-06 | -10.86 | 1.9e-27 |
| T x $p_{base}$ x m | 6.7e-05 | 8e-06 | 8.43 | 3.5e-17 |
| T x $p_{base_offset}$ x m | 0.00014 | 8.9e-06 | 15.81 | 3.2e-56 |
| $p_{base}$ x $p_{base_offset}$ x m | 0.0015 | 0.00013 | 11.19 | 4.6e-29 |
| T x $p_{base}$ x $m^2$ | 9.3e-05 | 3.4e-05 | 2.74 | 0.0061 |
| T x $p_{base_offset}$ x $m^2$ | -0.00022 | 3.8e-05 | -5.86 | 4.5e-09 |
| $p_{base}$ x $p_{base_offset}$ x $m^2$ | -0.0058 | 0.00056 | -10.21 | 1.8e-24 |
| T x m x $m^2$ | -0.0095 | 0.00042 | -22.76 | 2.2e-114 |
| $p_{base}$ x m x $m^2$ | -0.11 | 0.0062 | -17.11 | 1.4e-65 |
| $p_{base_offset}$ x m x $m^2$ | 0.054 | 0.007 | 7.78 | 7e-15 |

RP = recovery period; T = duration;

$p_{base}$ = base firing rate;

$p_{base_offset}$ = offset of base firing rate from $f_{osc}$;

m = oscillation modulation strength;

-- = outside range of double representation

Table S5. Parametric Effects on Residuals-Shuffling Hit Rates

High Firing Rates (9 ms RP, k = 0.7)

| Effect | Beta | SE | t | p |
|---|---|---|---|---|
| Intercept | 0.0011 | 6.2e-05 | 18.48 | 3.4e-76 |
| T | -4.8e-05 | 1.7e-06 | -28.85 | 6.6e-183 |
| $f_{osc}$ | -0.00011 | 6.8e-06 | -15.83 | 1.9e-56 |
| $p_{base_offset}$ | -0.00019 | 3.6e-06 | -51.56 | -- |
| m | -0.018 | 0.00042 | -42.45 | -- |
| $m^2$ | 0.045 | 0.00059 | 75.82 | -- |
| T x $f_{osc}$ | 2.6e-06 | 1.8e-07 | 14.60 | 2.8e-48 |
| T x $p_{base_offset}$ | 5.1e-06 | 9.7e-08 | 52.69 | -- |
| $f_{osc}$ x $p_{base_offset}$ | 5.1e-06 | 4e-07 | 12.96 | 2.2e-38 |
| T x m | 0.00056 | 1.1e-05 | 49.35 | -- |
| $f_{osc}$ x m | 0.00094 | 4.6e-05 | 20.44 | 8.3e-93 |
| $p_{base_offset}$ x m | 0.0017 | 2.5e-05 | 69.09 | -- |
| T x $m^2$ | -0.00045 | 1.6e-05 | -28.05 | 5.7e-173 |
| $f_{osc}$ x $m^2$ | 0.0013 | 6.5e-05 | 20.03 | 3e-89 |
| $p_{base_offset}$ x $m^2$ | 0.0018 | 3.5e-05 | 51.76 | -- |
| m x $m^2$ | -0.015 | 0.0029 | -5.27 | 1.4e-07 |
| T x $f_{osc}$ x $p_{base_offset}$ | -1.5e-08 | 6.8e-09 | -2.22 | 0.026 |
| T x $f_{osc}$ x m | 5.5e-08 | 4.1e-07 | 0.13 | 0.89 |
| T x $p_{base_offset}$ x m | 4.5e-06 | 2.2e-07 | 20.60 | 3.3e-94 |
| $f_{osc}$ x $p_{base_offset}$ x m | -7.2e-06 | 9e-07 | -8.00 | 1.3e-15 |
| T x $f_{osc}$ x $m^2$ | -2.5e-05 | 1.7e-06 | -14.19 | 1.1e-45 |
| T x $p_{base_offset}$ x $m^2$ | -6e-05 | 9.3e-07 | -64.42 | -- |
| $f_{osc}$ x $p_{base_offset}$ x $m^2$ | -2.8e-05 | 3.8e-06 | -7.26 | 4e-13 |
| T x m x $m^2$ | -0.0017 | 7.8e-05 | -22.31 | 3.2e-110 |
| $f_{osc}$ x m x $m^2$ | -0.0078 | 0.00032 | -24.22 | 1.6e-129 |
| $p_{base_offset}$ x m x $m^2$ | -0.013 | 0.00017 | -74.32 | -- |

RP = recovery period; T = duration;

$f_{osc}$ = oscillation frequency;

$p_{base_offset}$ = offset of base firing rate from $f_{osc}$;

m = oscillation modulation strength;

-- = outside range of double representation

Table S6. Parametric Effects on Residuals-Shuffling False Alarms

High Firing Rates (9 ms RP, k = 0.7)

| Effect | Beta | SE | t | p |
|---|---|---|---|---|
| Intercept | -0.014 | 0.00022 | -63.65 | -- |
| T | 1.5e-05 | 6e-06 | 2.53 | 0.011 |
| $f_{osc}$ | -0.00052 | 2.4e-05 | -21.37 | 3e-101 |
| $p_{base_offset}$ | 0.00024 | 1.3e-05 | 18.24 | 2.5e-74 |
| m | 0.12 | 0.0015 | 76.88 | -- |
| $m^2$ | 0.23 | 0.0021 | 107.57 | -- |
| T x $f_{osc}$ | -5.4e-06 | 6.5e-07 | -8.25 | 1.6e-16 |
| T x $p_{base_offset}$ | -6.7e-06 | 3.5e-07 | -19.16 | 8.6e-82 |
| $f_{osc}$ x $p_{base_offset}$ | 1.3e-05 | 1.4e-06 | 9.33 | 1e-20 |
| T x m | 4.2e-05 | 4.1e-05 | 1.04 | 0.3 |
| $f_{osc}$ x m | -0.0089 | 0.00017 | -53.48 | -- |
| $p_{base_offset}$ x m | 0.002 | 8.9e-05 | 22.75 | 1.6e-114 |
| T x $m^2$ | 0.0005 | 5.7e-05 | 8.83 | 1e-18 |
| $f_{osc}$ x $m^2$ | -0.079 | 0.00023 | -336.62 | -- |
| $p_{base_offset}$ x $m^2$ | 0.0057 | 0.00013 | 45.72 | -- |
| m x $m^2$ | -0.11 | 0.011 | -10.32 | 5.9e-25 |
| T x $f_{osc}$ x $p_{base_offset}$ | -6.5e-07 | 2.4e-08 | -26.43 | 7.2e-154 |
| T x $f_{osc}$ x m | -0.00026 | 1.5e-06 | -178.70 | -- |
| T x $p_{base_offset}$ x m | 2.8e-05 | 7.9e-07 | 35.75 | 1.9e-279 |
| $f_{osc}$ x $p_{base_offset}$ x m | -0.00059 | 3.2e-06 | -181.70 | -- |
| T x $f_{osc}$ x $m^2$ | -0.00069 | 6.2e-06 | -110.94 | -- |
| T x $p_{base_offset}$ x $m^2$ | 0.00015 | 3.3e-06 | 45.24 | -- |
| $f_{osc}$ x $p_{base_offset}$ x $m^2$ | -0.0016 | 1.4e-05 | -116.58 | -- |
| T x m x $m^2$ | 0.006 | 0.00028 | 21.19 | 1.3e-99 |
| $f_{osc}$ x m x $m^2$ | -0.15 | 0.0012 | -129.27 | -- |
| $p_{base_offset}$ x m x $m^2$ | 0.0062 | 0.00062 | 10.05 | 9.2e-24 |

RP = recovery period; T = duration;

$f_{osc}$ = oscillation frequency;

$p_{base_offset}$ = offset of base firing rate from $f_{osc}$;

m = oscillation modulation strength;

-- = outside range of double representation

## Data Availability Statement

Both the experimentally-acquired single-unit data and the analyzed simulation outputs are available on Zenodo at https://doi.org/10.5281/zenodo.8313070. Code sufficient to reproduce all figures and tables, and the simulations and analyses that informed them, is available on Github: https://github.com/kc13/residuals_spectral_analysis. The most recent Github release (1.0.0) has been issued the DOI 10.5281/zenodo.8323322 by Zenodo. The Github repository also stores copies of the single-unit data, and synthetic datasets that are represented in a postprocessed state.

## Financial Disclosure Statement

This work was supported by the National Institute of Neurological Disorders and Stroke at the National Institutes of Health (grant number R01NS117058-01 to RST; https://www.ninds.nih.gov/). Additionally, this research was funded in part by Aligning Science Across Parkinson's through the Michael J. Fox Foundation for Parkinson's Research (ASAP-020519 to RST; https://parkinsonsroadmap.org/, https://www.michaeljfox.org/). For the purpose of open access, the authors have applied a Creative Commons Attribution 4.0 International (CC BY) public copyright license to all Author Accepted Manuscripts arising from this submission. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.