



HAL
open science

A Bayesian approach for simultaneous spike/LFP separation and spike sorting

Steven Le Cam, Pauline Jurczynski, Jacques Jonas, Laurent Koessler, Sophie Colnat-Coulbois, Radu Ranta

► **To cite this version:**

Steven Le Cam, Pauline Jurczynski, Jacques Jonas, Laurent Koessler, Sophie Colnat-Coulbois, et al.. A Bayesian approach for simultaneous spike/LFP separation and spike sorting. *Journal of Neural Engineering*, 2023, 20 (2), pp.026027. 10.1088/1741-2552/acc210 . hal-04036522

HAL Id: hal-04036522

<https://hal.science/hal-04036522>

Submitted on 9 Nov 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



PAPER

A Bayesian approach for simultaneous spike/LFP separation and spike sorting

Steven Le Cam^{1,*} , Pauline Jurczynski¹, Jacques Jonas^{1,2}, Laurent Koessler¹, Sophie Colnat-Coulbois^{1,2} and Radu Ranta¹ ¹ Université de Lorraine, CNRS, CRAN, F-54000 Nancy, France² Université de Lorraine, CHRU-Nancy, Service de neurologie, F-54000 Nancy, France

* Author to whom any correspondence should be addressed.

E-mail: steven.le-cam@univ-lorraine.fr

Keywords: spike sorting, spike removal, Bayesian approach, variational optimization

RECEIVED
21 October 2022REVISED
22 February 2023ACCEPTED FOR PUBLICATION
7 March 2023PUBLISHED
31 March 2023**Abstract**

Objective. The aim of this paper is to present a novel method for simultaneous spike waveforms extraction and sorting from the raw recorded signal. The objective is twofold: on the one hand, to enhance spike sorting performance by extracting the spike waveforms of each spike and, on the other hand, to improve the analysis of the multi-scale relationships between spikes and local field potentials (LFP) by offering an accurate separation of these two components constitutive of the raw micro recordings. *Approach.* The method, based on a Bayesian approach, is fully automated and provides a mean spike shape for each cluster, but also an estimate for each singular spike waveform, as well as the LFP signal cleaned of spiking activity. *Main results.* The performance of the algorithm is evaluated on simulated and real data, for which both the clustering and spike removal aspects are analyzed. Clustering performance significantly increases when compared to state-of-the-art methods, taking benefit from the separation of the spikes from the LFP handled by our model. Our method also performs better in removing the spikes from the LFP when compared to previously proposed methodologies, especially in the high frequency bands. The method is finally applied on real data (ClinicalTrials.gov Identifier: NCT02877576) and confirm the results obtained on benchmark signals. *Significance.* By separating more efficiently the spikes from the LFP background, our method allows both a better spike sorting and a more accurate estimate of the LFP, facilitating further analysis such as spike-LFP relationships.

1. Introduction

Extra-cellular micro-recordings provide unique opportunity to study the neuronal responses to sensory-motor or cognitive process at the level of the neuronal cell. While the analysis of spikes alone is a fruitful field of investigation for analyzing the function of neuronal circuits [6, 35], the joint analysis of the local field potentials (LFP) provides complementary information and has recently gained more interest [8, 14]. In particular, studying the relationships between spikes and LFP originating from a given neuronal population or from distant ones brings key information about the mechanism of cortical networks involved in motor and sensory processing [26, 40], or higher level processes such as memory [24]. The number and nature of the responding neurons to a given paradigm, and the comprehension of the

relation between the spiking activities and the LFP are expected to hold precious clues for understanding neural information processing. The objectives are then twofold: separating the LFP from the spiking activities and automatically classifying the spikes to identify single unit activities, without human intervention.

The spikes originating from a given neuron can be seen on micro-electrodes placed within a distance of roughly 50–100 μm [25]. The temporal shape of the recorded waveforms depends on the type of neuron (e.g. inhibitory vs excitatory), on its morphology, as well as on the position (i.e. its distance as well as its orientation) of the neuron with respect to the electrode. It is then possible to distinguish between the spikes of different neurons recorded by the same electrode, giving rise to an active field of research for the development of spike sorting methods [5, 37].

To separate the spike waveforms from the LFP and extract the features needed for the classification task, standard approaches first apply a band-pass filter (generally in the 300–3000 Hz band): the activity below 300 Hz is generally assumed to stand mainly for the LFP (i.e. the activity of the neural population surrounding the electrode within a distance range of up to several millimeters given the spatial extent of the surrounding correlated synaptic activities [25]), while the higher frequency bands are considered to hold the main part of the spiking activities. The time instant of the spikes are identified by thresholding the band-passed signal. Discriminant features such as morphological parameters [23] or wavelet coefficients [18, 30, 34] are extracted from the detected spike shapes, considering about 1.5–3 ms time window around the maximum peak of the spike shape. A clustering method is then applied on these features to identify the units from which the spikes originate. Clustering approaches based on mixture models [17, 23, 38, 42] or superparamagnetic clustering (SPC) [10, 27, 34] have been considered. Several recent research efforts have addressed the spike sorting challenge in the context of high dimensional data (arrays of electrodes) [12, 16, 21, 31, 38, 47], however we focus in this paper on the mono-channel issue, as we target the processing of signals recorded using Ad-Tech Behnke-Fried micro-electrodes setup [15], for which the distance between the micro-wires cannot be controlled and can reach up to several millimeters, thus disqualifying methods based on a multichannel strategy.

The LFP and the spiking activities overlap in frequency, and some significant features of the spikes helpful for the classification process may lie in the low frequency range under 300 Hz [50]. Classical standard (non-causal band-pass) filters are commonly considered to extract the spike waveforms, but such approach cannot differentiate well between the components of the spikes and those of the LFP signal respectively [14, 49], and are known to corrupt the spike shapes [33]. For these reasons it is preferable to first separate properly the LFP and the spiking activity before proceeding to the classification of the spikes. On the other hand, the spike activities are likely to contaminate the underlying LFP activity of interest [7]. This can be for example the case in higher frequency (gamma) band, where broadband event-related gamma responses over 60 Hz, and extending up to 200 Hz [41], have been observed in various functional brain systems [29, 45]. Their origins, and in particular in which proportion the spiking activity itself generates these activities, is still a subject of debate [39]. All components due to spikes then need to be removed appropriately before analyzing such activities [1, 41, 49]. Alternatives to the simple low-pass filtering strategy are necessary to separate the LFP from the spikes, which can be formulated as an underdetermined inverse problem where

two major contributive activities must be demixed from a single measured signal. To the best of our knowledge only two efforts can be cited: Zanos *et al* [49] have proposed a Bayesian approach to remove the spike from the LFP after the spike classification step, based on a colored Gaussian prior over the distribution of the LFP. The waveform of the spikes for each unit are reestimated in the process, under the assumption that this waveform is fixed for each event. The second method, called Adaptive Spike-Artifact Removal (ASR), has been recently proposed by Bouroujeni *et al* [3], in which the frequency components of each individual spikes are first identified before being removed from the wideband signal.

In this work we address the problem by proposing an iterative Bayesian approach to separate the LFP from the spiking activities as well as classifying the spikes simultaneously. As in Zanos *et al* [49], we use a prior on the power spectral density of the LFP to proceed to its separation from the spikes. We rely on a classical spike detection step in the higher frequency band (between 300 and 3000 Hz) to identify the spike instant, and we then process the rough unfiltered data to proceed to the separation and to the classification in an iterative manner. A first version of this method working in the time domain has been published in conference [20], in which only the classification performance of the algorithm were presented. The method is here developed in the wavelet domain, providing highly discriminant features of the spike waveforms originating from distinct units and further enhancing the classification results. The wavelet coefficients are modeled as mixture of Gaussians, and a maximum *a posteriori* (MAP) decision strategy is used for the classification task.

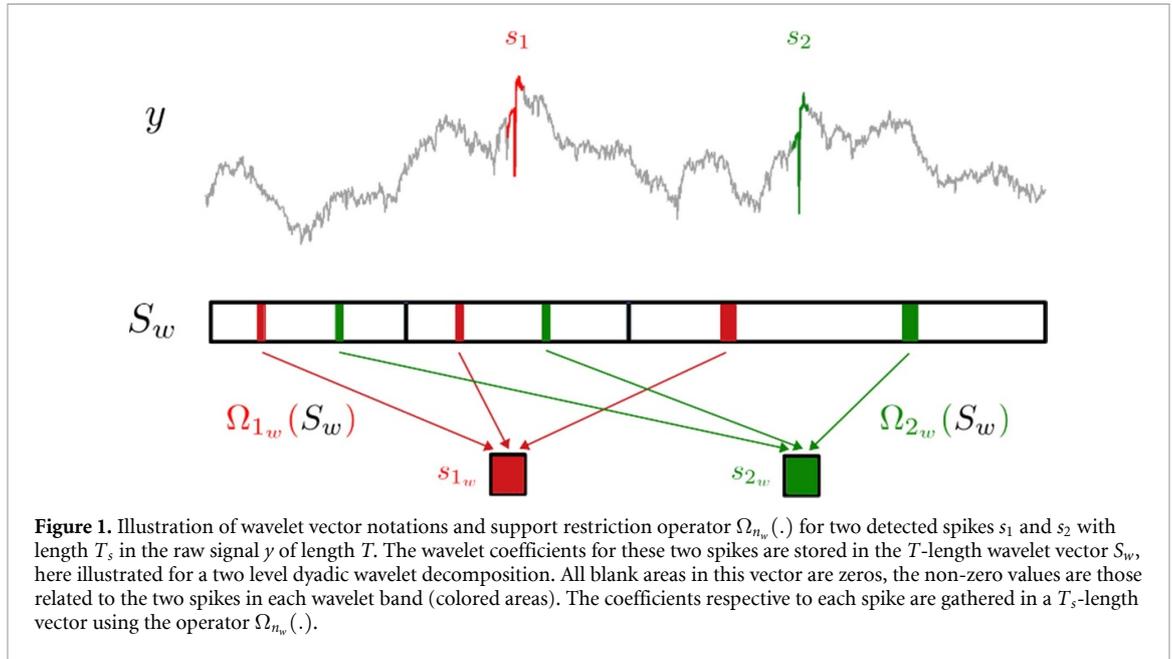
The approach is evaluated based on a realistic set of single-channel simulated data available online [9], and we compare the performance with fully automatic state-of-the-art methods that are relevant in this context of mono-channel recordings, as detailed in the dedicated result section 3.2.1. We also illustrate and compare the performance of the method on real data recorded during frequency tagging or fast periodic visual stimulation (FPVS) [19, 28]. The quality of the spike decontamination from the LFP at least reproduce or even exceed the performance reported in [49], especially in high frequency bands.

2. Methods

2.1. Generative model

In this section we present the general Bayesian formalization of the problem. Let suppose that the recorded channel y of length T can be decomposed as the addition of the LFP x and the spiking signal S , up to ϵ modelling the additive noise:

$$y = x + S + \epsilon. \quad (1)$$



The spiking signal S can be written as a sum of N spike waveforms with a time support T_s of few milliseconds (typically 1–3 ms), or alternatively as a sum of $N = \sum_{k=1}^K N_k$ waveforms originating from K different units (neurons):

$$S = \sum_{n=1}^N s_n = \sum_{k=1}^K \sum_{n=1}^{N_k} s_n^k. \quad (2)$$

The problem can be indifferently considered in the wavelet domain:

$$y_w = x_w + S_w + \epsilon_w \quad (3)$$

with $y_w, x_w, S_w, \epsilon_w$ vectors of concatenated wavelet coefficients with length T , considering any L level dyadic wavelet decomposition of y, x, S and ϵ respectively. S_w contains the wavelet coefficients s_{n_w} related to the decompositions of the N spike waveforms s_n , such that s_{n_w} are vectors with size T_s . We introduce the operator $\Omega_{n_w}(\cdot)$ restricting any wavelet vector of length T to the T_s -length wavelet support of the n th spike, such as $\Omega_{n_w}(S_w) = s_{n_w}$ (see figure 1).

The waveforms s_n^k (and thus the wavelet coefficients $s_{n_w}^k$) originating from a given neuron (class) are not imposed to be equal, variations are expected accounting for the uncertainties inherent to such biological process (e.g. exhaustion effect) and measurements setup (movement of the tissues as well as drift of the electrodes [5, 27]). We model these uncertainties by use of a Gaussian model over the distribution of the wavelet coefficients $s_{n_w}^k$ for each class. Let introduce the hidden variable Z , taking its values in the set $\{\lambda_1, \dots, \lambda_K\}$, λ_k being the class of spike waveforms originating from the k th neuron. The waveforms s_{n_w} then follow a Gaussian mixture model (GMM):

$$p(s_{n_w}) = \sum_{k=1}^K \pi_k p(s_{n_w} | z_n = \lambda_k)$$

$$p(s_{n_w} | z_n = \lambda_k) = \mathcal{N}(s_{n_w} | s_w^k, \Sigma_w^k)$$

with $\pi_k = p(z = \lambda_k)$ the priors on Z , and $\{s_w^k, \Sigma_w^k\}$ the mean and covariance of the wavelet distributions for the class λ_k .

Such as in [49], x follows a Gaussian distribution $\mathcal{N}(0, \gamma \Sigma_x)$, with γ a scaling parameter. This can be transposed in the wavelet domain, where x_w follows a Gaussian distribution $\mathcal{N}(0, \gamma \Sigma_{x_w})$.³ Finally, the noise ϵ (or equivalently ϵ_w in the wavelet domain) follows a i.i.d Gaussian distribution with zero mean and variance σ_ϵ^2 . The method will adjust between γ (accounting for the power of the LFP) and σ_ϵ^2 (accounting for the noise power). The likelihood of the observations y_w finally writes:

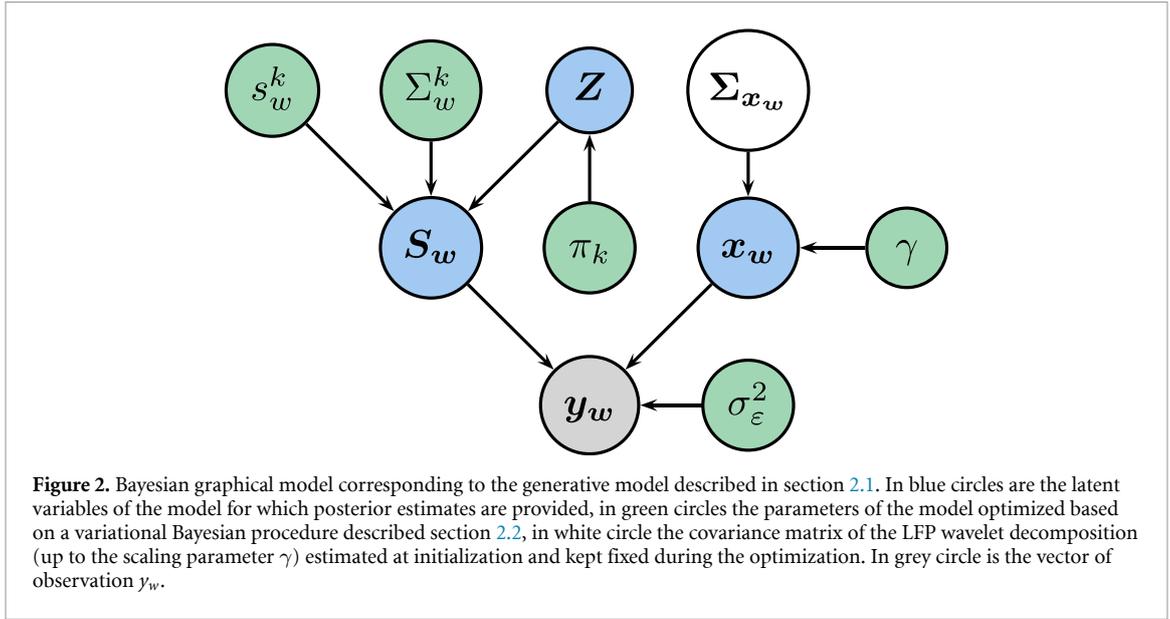
$$p(y_w | x_w, S_w) = \frac{1}{(2\pi)^{T/2} \sigma_\epsilon^T} \exp\left(-\frac{1}{2\sigma_\epsilon^2} \|y_w - x_w - S_w\|_2^2\right) \quad (4)$$

and the log-likelihood \mathcal{L} of the full model is given by:

$$\begin{aligned} \mathcal{L} &= \log p(y_w, x_w, S_w, Z) \\ &= \log p(y_w | x_w, S_w) + \log p(x_w) + \log p(S_w | Z) \\ &\quad + \log p(Z). \end{aligned} \quad (5)$$

The generative model is given on figure 2. The optimization of the model parameters $\Theta = \{\gamma, \{s_w^k, \Sigma_w^k\}_{k=1}^K, \sigma_\epsilon^2\}$ is carried out through a variational Bayesian (VB) procedure detailed in the next section.

³ Let W be the wavelet decomposition matrix, then $\Sigma_{x_w} = W^T \Sigma_x W$.



2.2. Optimization

A VB approach is used to optimize the model (maximization of the marginal log-likelihood of equation (5)). The VB approach consists in introducing the probability density $q(x_w, S_w, Z) = q(x_w) \prod_{n=1}^N q(s_{n_w}) q(z_n)$, which is optimized to approximate the true posterior density $p(x_w, S_w, Z | y_w)$. The first step of the optimization, which can be assimilated to the E-step of the classical expectation-maximization (EM) algorithm, is the updates of each of the $(2N + 1)$ $q(\cdot)$ factors constitutive of this approximate posterior probability: a given factor $q(\cdot)$ is updated as the expectation of the full model log-likelihood \mathcal{L} with respect to the $2N$ remaining $q(\cdot)$ factors [2]. The densities $q(x_w)$ and $q(s_{n_w})$ are identified as multivariate Gaussian distributions, with means \hat{x}_w and \hat{s}_{n_w} and covariance $\hat{\Sigma}_{x_w}$ and $\hat{\Sigma}_{n_w}$ respectively:

$$\hat{x}_w = \gamma \Sigma_{x_w} (\gamma \Sigma_{x_w} + \sigma_\epsilon^2 I_T)^{-1} (y_w - \hat{S}_w) \quad (6)$$

$$\hat{\Sigma}_{x_w} = \sigma_\epsilon^2 \gamma \Sigma_{x_w} (\gamma \Sigma_{x_w} + \sigma_\epsilon^2 I_T)^{-1} \quad (7)$$

$$\hat{s}_{n_w} = \hat{\Sigma}_{n_w} \left(\frac{1}{\sigma_\epsilon^2} \Omega_{n_w} (y_w - \hat{x}_w) + \sum_{k=1}^K \hat{\pi}_{nk} \Sigma_w^{k-1} s_w^k \right) \quad (8)$$

$$\hat{\Sigma}_{n_w}^{-1} = \frac{1}{\sigma_\epsilon^2} I_{T_s} + \sum_{k=1}^K \hat{\pi}_{nk} \Sigma_w^{k-1} \quad (9)$$

with \hat{S}_w a vector containing the N posterior mean for the wavelet coefficients of the spikes \hat{s}_{n_w} at their respective position in the wavelet decomposition of y , such that $\Omega_{n_w}(\hat{S}_w) = \hat{s}_{n_w}$. Hence $\Omega_{n_w}(y_w - \hat{x}_w)$ are the wavelet coefficients of the raw signal y minus those of the estimated (posterior mean) LFP signal \hat{x} restricted

to the wavelet support of the n^{th} spike. $\hat{\pi}_{nk} = q(z_n = \lambda_k)$ are the (approximate) posteriors for each label z_n .

Defining the quadratic form $\|s\|_\Sigma^2 = s^T \Sigma^{-1} s$, the factors $q(z_n)$ are given by:

$$\hat{\pi}_{nk} \propto \pi_k |\Sigma_w^{k-1}| \exp \left(-\frac{1}{2} \|\hat{s}_{n_w} - s_w^k\|_{\Sigma_w^k}^2 - \frac{1}{2} \text{Tr}[\hat{\Sigma}_{n_w} \Sigma_w^{k-1}] \right) \quad (10)$$

up to a normalization constant summing these probabilities to 1. At convergence, a MAP decision strategy based on these probabilities is used to decide for the cluster label of each spike waveforms.

As done in [10, 34], leaning on the benefit of (appropriate) dimensionality reduction for clustering as well as on the property of wavelet decompositions to concentrate the energy of the signal over few coefficients, this clustering part can also be carried out considering a selection of wavelet coefficients assumed to facilitate the discrimination between the clusters.

The M-step provides an update of the model parameters $\Theta = \{\gamma, \{s_w^k, \Sigma_w^k\}_{k=1}^K, \sigma_\epsilon^2\}$, by deriving with respect to each parameter the posterior expectation of the full model log-likelihood:

$$\gamma = \frac{1}{T} \left(\|\hat{x}_w\|_{\Sigma_{x_w}}^2 + \text{Tr}[\hat{\Sigma}_{x_w} \Sigma_{x_w}^{-1}] \right) \quad (11)$$

$$\sigma_\epsilon^2 = \frac{1}{T} \left\| y_w - \hat{x}_w - \sum_{n=1}^N \hat{s}_{n_w} \right\|_2^2 + \text{Tr}[\hat{\Sigma}_{x_w}] + \sum_{n=1}^N \text{Tr}[\hat{\Sigma}_{n_w}] \quad (12)$$

$$s_w^k = \frac{\sum_{n=1}^N \hat{\pi}_{nk} \hat{s}_{n_w}}{\sum_{n=1}^N \hat{\pi}_{nk}} \quad (13)$$

$$\Sigma_w^k = \frac{\sum_{n=1}^N \hat{\pi}_{nk} \left(\|\hat{s}_{n_w} - s_w^k\|_2^2 + \hat{\Sigma}_{n_w} \right)}{\sum_{n=1}^N \hat{\pi}_{nk}} \quad (14)$$

Details for these E-step and M-step update formulas are provided in the appendix A.

2.3. Initialization

2.3.1. Filtering and spike detection

Following [27, 34, 37], the detection of the spikes is carried out by thresholding the high frequency part of the signal in the range 300–3000 Hz (four-poles Butterworth band-pass filter). The threshold is computed as a multiple of the background noise standard-deviation σ_n within this frequency band. The estimation of this parameter must be as robust as possible to artifact and detect the relevant spiking activity, and the following robust estimation is used [13, 34]:

$$\sigma_n = \text{median} \left(\frac{|x - \bar{x}|}{0.6745} \right) \quad (15)$$

with \bar{x} the mean of the signal x . Every event in the band-passed signal with absolute amplitudes over a threshold set as a multiple (usually from 4 to 5) of this robust σ_n estimate are considered as spiking events (events with positive and negative peaks are considered together in the procedure). Note that we do not consider overlapping spikes in this work: within a window of 2 ms only the highest peak (in absolute value) is considered.

2.3.2. LFP covariance matrix

The prior over the wavelet coefficients of the LFP x_w is a Gaussian distribution with zero mean and covariance $\gamma \Sigma_{x_w}$. The matrix Σ_{x_w} might be estimated directly from the wavelet decomposition of the raw signal y , however this $T \times T$ matrix is very large as T can quickly reach several millions up to billions given the high frequency sampling of micro-recording data. Its estimation as well as the computation of the update equations (6) and (7) in which product of this matrix is involved becomes infeasible in practice. In [49], a circulant approximation is proposed for Σ_x , the covariance matrix of the LFP x in the time domain, based on a weak stationarity hypothesis on the LFP signal x .⁴ This approximation does not hold for Σ_{x_w} , which consists in a block matrix of inner wavelet band covariance matrix on its diagonal and inter-band covariance matrices elsewhere. We then propose to carry out the computation of these equations (6) and (7) in the time domain as follows:

$$\hat{x} = \gamma \Sigma_x (\gamma \Sigma_x + \sigma_\epsilon^2 I_T)^{-1} W T^{-1} (y_w - \hat{S}_w) \quad (16)$$

$$\hat{\Sigma}_x = \sigma_\epsilon^2 \gamma \Sigma_x (\gamma \Sigma_x + \sigma_\epsilon^2 I_T)^{-1} \quad (17)$$

with $W T^{-1} (y_w - \hat{S}_w)$ the inverse wavelet transform (temporal reconstruction) of the raw signal wavelet

⁴ this approximation introduces side effects in the estimation of the signal for the few first and last samples, which can be cutted out and ignored in the end.

coefficients y_w free of the (posterior mean of the) spikes wavelet coefficients. The LFP posterior mean \hat{x} is then decomposed back in the wavelet domain, as needed in equation (8).

To sum up, the spike sorting part of the algorithm is carried out in the wavelet domain, while the LFP estimation is done in the time domain (note that equations (11) and (12) can be invariably handled in both time and wavelet domain). This gives the modified generative Bayesian graphical model of figure 3 which is used in practice.

Σ_x is built by circulation of the power spectral density (PSD) g of the LFP x on its lines. g is estimated at the initialization of the algorithm and kept fixed, while the scaling factor γ is estimated at each iteration by the algorithm (see equation (11)) and balance the respective contributions of the LFP and the background noise to the observed signal y . We recall in the appendix B how the circulant property of Σ_x along with the Fast Fourier Transform (FFT) algorithm allow fast computations of such huge matrix product.

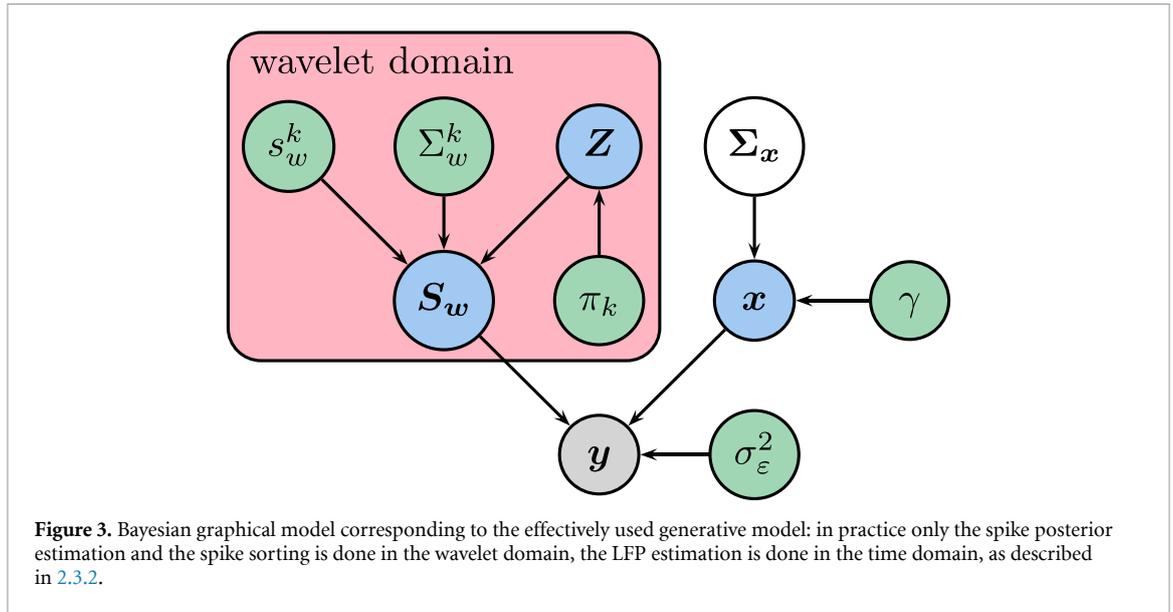
The PSD of the LFP is estimated from the raw signal y , which necessarily implies bias due to the presence of noise and spiking activities superimposed with the LFP in y . Zanos et al [49] propose to fit the exponential function $-\exp(1 + \log(x))$ to the PSD of the raw signal y in the range 1–150 Hz and to extrapolate the value of the PSD in the higher frequency range. We adopt this approach in this paper. Alternative methods such as autoregressive models could have been considered such as in [20], but it did not reveal to significantly change the final results.

2.3.3. Number of clusters

The number of classes (spike clusters) K is also a parameter to be estimated in such unsupervised classification algorithm. The proposed approach based on VB optimization has to be informed with the number of classes to be estimated, we then propose a statistical procedure to overcome this problem and fully automatize the method: a range of values for K is explored around a putative number of class K_{init} given by any standard clustering algorithm. For each K value in this range, a Bayesian information criterion (BIC) [2] is computed by considering the marginal joint log-likelihood of (\hat{s}_{w_n}, Z) :

$$\text{BIC}_{\text{VB}}(K) = \sum_{n=1}^N \log \left(\sum_{k=1}^K p(\hat{s}_{w_n} | z_n = \lambda_k) \pi_k \right) - 0.5 * (K * L - 1) * \log(N * T_s) \quad (18)$$

with $L = K(1 + T_s + T_s(T_s - 1)/2) - 1$ the number of continuously valued real numbers required to specify the K cluster parameters $\{\pi_k, s_w^k, \Sigma_w^k\}_{k \in [1..K]}$. The classification result for K yielding the maximum BIC value is chosen as the final solution, thus also providing an estimate of the number of classes.



3. Results

3.1. Spectral content of LFP and spike waveforms

Before addressing the evaluation of the proposed method, we briefly analyze the spectral content of a typical spike waveform, which is the basis for the justification of our approach and for the motivation of spike removal in general [48, 49]. Spikes are fast and transient activities with duration of about 2–3 ms, with high positive or negative peaks whose amplitudes and shapes in the signal depend on the position of the recording site with respect to the spiking neuron. It is widely accepted that the main spectral features of the spikes lie in high frequency band, roughly over 300 Hz, while the LFP part of the signal is below this value, as illustrated in figure 4, left column. Spectrum overlap can still however be seen between these components (figure 4, bottom left).

When low-pass filtering both the spike and the LFP below 150 Hz, the residual power of the spike is indeed much lower than the residual power of the LFP (middle column), but not negligible. Keeping in mind that the coherent activity between the LFP and the spike (coherent synaptic activity synchronized with the spikes, usually estimated by spike trigger averaging) is also of much lower amplitude compared to the LFP, such analysis are likely to be severely corrupted by spike residuals. This suggests that a simple low-pass filtering might not be sufficient to separate the LFP from the spikes, thus biasing ulterior relationship analysis.

Common spike sorting methods estimate the spike waveforms to be classified by cutting out a band-pass filtered version of the raw signal over each spike time support. On figure 4 band-pass filtered versions of the LFP epochs and of the spike waveform are given, where the residual amplitude of the LFP with respect to this of the spike is still noticeable. One can

also notice the effect of the filtering on the shape of the spike, especially on both sides of the central peak, where the shape is modified and side bumps are introduced. Both the addition of LFP high frequency components residuals and the filtering effect can mislead the classification, motivating a methodology based on an estimation of the actual spike shapes.

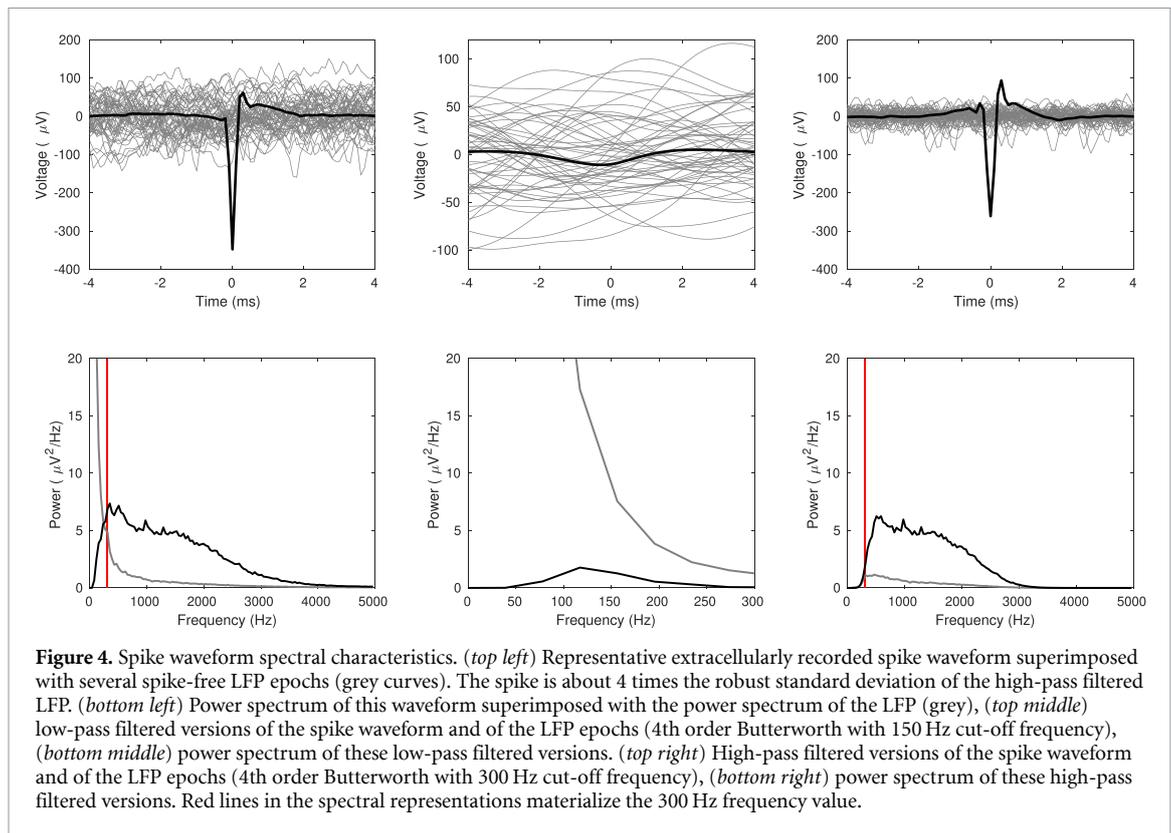
3.2. Simulated data

First, in section 3.2.1 the spike-sorting performances are tested on the synthetic data set of Camuñas-Mesa and Quiroga [9], which has been recently used by several authors for evaluating their spike sorting methods [11, 27, 32].

We then evaluate in section 3.2.2 the ability of the method to accurately remove the spikes from the LFP, based on surrogate signals.

3.2.1. Spike sorting

We evaluate the spike sorting performance of our algorithm on the data set from Camuñas-Mesa and Quiroga [9], which provide realistic simulation of micro recordings containing single units (close neurons) as well as multi units originating from more distant neurons. In these simulations, the presence of 2 up to 20 units are simulated to challenge the algorithms. For each number of unit 5 signals of 10 min length each (24 Khz sampling rate) are available (<https://www135.lamp.le.ac.uk/hgr3/>). This provides a common ground-truth for comparison. This data base has been considered in several recent studies for evaluating spike sorting approaches [11, 27, 32]. The method is evaluated using the same criteria used in these references: a cluster is counted as a hit if it contains more than 50% of a given simulated single unit, and if at least 50% of the spikes originating from this unit are classified in this cluster. Based on this criterion, if a cluster does not match any unit



and is not formed of at least 50% of multi-unit spikes, it is counted as a false positive.

We compare our algorithm, which we will refer to as VBDS in the following (variational Bayesian despiking and sorting method), with four recent algorithms of the literature: Klusta [38], Combinato [27], Waveclus in its automatic version [10] and Mountainsort4 [12]. We run the four selected methods through Spikeinterface [4], a freely available platform including several recent spike sorters as well as various tools for spike sorting evaluation and comparison. All of these four methods consider the classification of the spike waveforms extracted from the signal after band-pass filtering. For this simulated data set, a fourth-order filter in the band [300–3000] Hz has been considered [10]. Klusta is based on a GMM of features obtained through PCA of the high-passed and re-aligned spike waveforms, and use a ‘Masked EM’ strategy for the optimization/classification task. After several tests of the method on the simulated data set, we found out that the parameters used in [10] were indeed the most appropriate, i.e. 10 principal components, with a spike support from -24 ms to 40 ms around the maximum instant of the spikes. Both Combinato [27] and the new implementation of Waveclus [10] are based on SPC, a methodology which suffers from the high number of parameters to be tuned in its original version [34]. Combinato [27] as well as the new version of Waveclus [10] proposes semi automatic (for the former) to fully automatic (for the latter) spike sorting algorithm, and convincing sorting performance have been reported by the

authors when applied on this particular data set of Camunas *et al*. For Combinato, we have retained the default set of parameters as described in the original paper [27], except for the maximum number of clusters at one temperature that we set to 7, the minimum number of spikes in a cluster set to 50, the minimum cluster size for iterative clustering set to 1000, and the number of clustering iterations set to 2. The new version of WaveClus (simply referred as WaveClus in the following) being automatic, no parameter had to be fixed before launching the algorithm (apart of the detection threshold, see below). Finally, we also consider the Mountainsort algorithm of Chung *et al* [12], an efficient and fully automatic spike sorter which is originally dedicated to perform spike sorting in the context of dense microelectrode recordings, but which can also be applied on single channel recording. The method is based on a clustering approach called ISO-SPLIT, which identifies unimodal distributions in the feature space, defined either by features extracted from the spikes identified on a single channel or on several neighboring channels. The method is fully automatic and parameter settings is not needed apart of the detection threshold. We have used the latest version of the algorithm (Mountainsort4).

Note that we have also considered other recent spike sorting methods, such as Kilosort [38], SpyKING CIRCUS [47] and YASS [21]. Kilosort [38] is dedicated to multichannel recordings setup and is then not eligible in this context of single channel recordings. SpyKING CIRCUS [47] has also been developed to handle multichannel data, and while it

can be applied on single channel recordings the results were not convincing in comparison to the other tested methods. The same remark can be made for YASS [21] for which a dedicated training step is necessary for this method to be used at its full potential.

For VBDS, a range of K values is explored as described in section 2.3.3. A standard GMM with EM optimization (GMM-EM [2]) used to set the K_{init} value (with BIC criterion), as well as for the initialization of the set of clustering parameters $\{\pi_k, s_w^k, \Sigma_w^k\}_{k \in [1..K]}$ for each explored value of K . We have explored K values ranging from K_{init} to $1.5 K_{\text{init}}$, with 10 initialization for each value of K to avoid being trapped in local minima. A seven level dyadic decomposition is used (approximation level ranging up to 97.5 Hz). The 10 coefficients with the highest variance have been selected in the detail levels for the clustering part⁵. Mother wavelets from popular wavelet families such as Haar, Daubechies, Symlets or Coiflets have been tested, and Sym6 from the Symlet family was finally chosen as it proves to provide higher discriminative power between spike classes when applied on the used simulated database, although the results were not significantly different when using other wavelets. The selection of the adequate mother wavelet may be left to the experimenters with respect to the dataset to be treated. We also include the results of a previous version of the algorithm in which the spike shapes were directly clustered from their time samples [20], demonstrating the benefit of working in the wavelet domain providing enhanced discrimination power.

For all the algorithms, the spike detection is carried out using a robust thresholding (see equation (15)) in the band [300–3000] Hz. The threshold factor has been set to 4.5 because it was found to give the best compromise between detection of the true spike events and false detections: on average $78.3\% \pm 6.8$ of the total number of spikes are detected (MU spikes included), and $97.2\% \pm 1.7$ of the SU unit spikes are detected, while the number of false positives (spikes that are neither generated by the SU nor the MU) ranges from 1.2% to 3.4% of the total number of detected spikes (ranging from 5 to 123 in number over all 95 simulations).

On figure 5 is given the spike sorting performance, in the same manner than done in [10]. Figure 5(a) clearly show that the proposed method is significantly more accurate than the other tested algorithms in retrieving the clusters for most of the number of units, and in particular when this number is high. When it comes to false positives outcomes, our proposed algorithm (figure 5(b)) is also very performant and produce significantly less spurious clusters. The performance positively compared with the previous

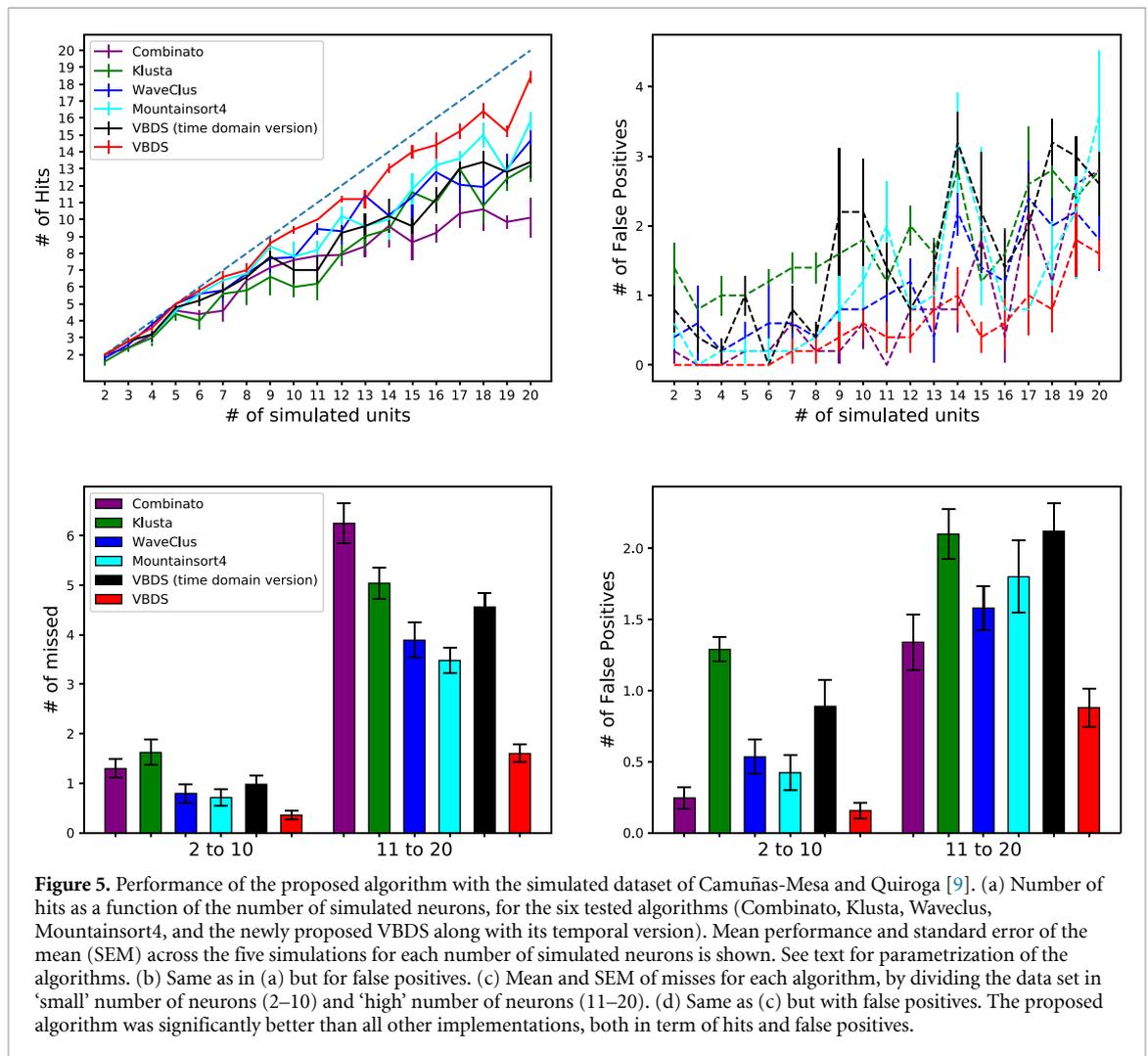
version of the method in the time domain, both for hits and false positive rates. Note that despite our efforts, we did not manage to reproduce the same level of performance in term of hits for Combinato when compared to these reported in [10], illustrating that the method is highly sensitive to its parametrization. However we report a lower number of false positives. Besides, the performance found for Klusta, WaveClus are very similar, whereas Mountainsort4 demonstrates better performance as the number of units to estimate increases, while still keeping the number of false positives limited. To give a closer look to these results, we follow the methodology proposed in [10] by grouping the data in two sets: those with 2–10 units on one side ('low' number of units) and those with 11–20 units ('high' number of units). This synthesis confirms the superiority of our method, especially when the number of cluster to be retrieved is high.

These high performance might be tempered by the computational cost and time necessary to run VBDS. Let remind that this method provides both the sorting of the spikes as well as a cleaned (free of spikes) version of the LFP, which partly explains its higher computational burden. About five to ten iterations are necessary for the algorithm to converge, each iteration involving a filtering of the full LFP signal (see appendix B). If the use of FFT indeed reduces the computational cost, it requires nevertheless several seconds to be performed on a signal containing $24\,000 \times 600 = 14\,400\,000$ samples. In addition, as the number of clusters K is not known, the algorithm had to be launched several times (with the final solution being chosen as this maximizing the BIC criterion as defined in equation (18)). On a recent standard computer (Linux OS, Ubuntu 18.04.5 LTS, 32Go of RAM, intel Core-i7 8x2.90 Ghz, no GPU), it took 3 min 30 on average to process for simulated recordings with 2 units, and up to 20 min for the signals containing 20 units, figures to be compared with this of the other tested algorithms which roughly range between 1 to 2 min as experimented on the same computer, similarly as reported in [10].

3.2.2. Spikes/LFP separation

The data set of Camuñas-Mesa and Quiroga [9] roughly simulate the background LFP as white Gaussian noise and does not hold the frequency property of a realistic LFP. In order to test the spike removal (or despiking) aspect of the algorithm, we simulate a data set of 200 signal examples as follows: we considered real signals recorded at the epilepsy unit of the CHRU (protocol REUNIE), from which we selected 15 min length of 10 000 Hz sampled signal with no spiking events, and we randomly extracted 200 1 min length epochs from this signal. For each epoch, a spike shape is randomly chosen from the data set of Camunas *et al* and 500 events with this shape are distributed on the signal, hence one single unit can be detected within

⁵ Equations (10) and (18) are then computed by considering restrictions of the wavelet vectors and covariances to these selected coefficients.



each 1 min simulated epoch. The amplitude of the spikes are set so that it reaches 6 times the median absolute deviation (MAD) of the LFP epoch. The spike shapes provided by the Camunas data set are sampled with a 96 000 Hz frequency rate, the shapes are decimated to a 10 000 Hz sample rate, and we consider different sub-sampling shifts during the decimation, leading to variabilities in the spike shapes superimposed with the LFP signal. A minimal interval of 10 ms is ensured between two successive spikes, so that no overlapping spiking events are considered. At each spike instant, we finally add an identical Gabor waveform at each spike location to simulate the presence of a coherent LFP activity synchronized with the spiking activity. The Gabor waveform is about 60 ms long, its spectral content lies within the frequency range [0 – 150] Hz, and its maximum amplitude is set to the tenth of the MAD of the real extracted LFP epoch.

Our method VBDS, as well as the Bayesian spike removal method of Zanos *et al* (called BSR in the following) and the Automatic Spike Removal (ASR) method of Boroujeni *et al* [3] have been applied on this synthetic data set. Figure 6(a) illustrates the

despiking result for an epoch of 250 ms involving 4 spiking events to be removed. The resulting traces estimated by the three methods (shifted in amplitude for better visualization) shows that the spikes are indeed successfully removed by all three methods, with no noticeable difference between our method and this of Zanos *et al* for this rather large visualization scale. The ASR method however seem to flatten the amplitude of the signal around the spiking events, in particular around the two last spike instant. This impression is confirmed when computing the spike-triggered average of the signal (mean of the resulting signals within 20 ms windows around each removed spiking events), as can be seen on figure 6(b), where the ASR approach fails to estimate the ground-truth signal. VBDS as well as BSR are much closer from the original signal. A residual artifact remains nevertheless visible when considering the results for the latter, due to the assumption of fixed spike waveforms made by this method: the same mean waveforms is finally removed from the signal for each event. These residuals are also visible on the figure 6(c) where we have zoomed on small epochs of about 5 ms around 15 removed spiking

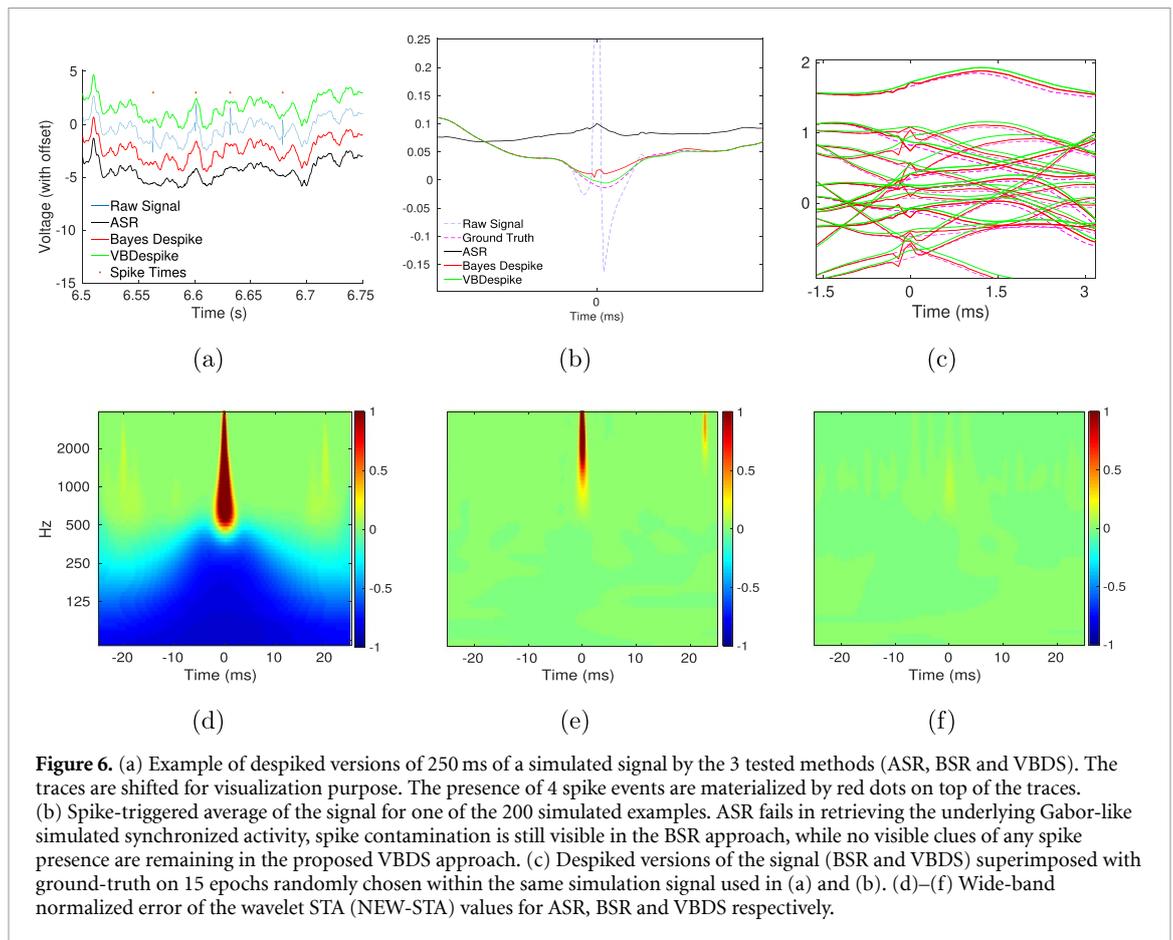


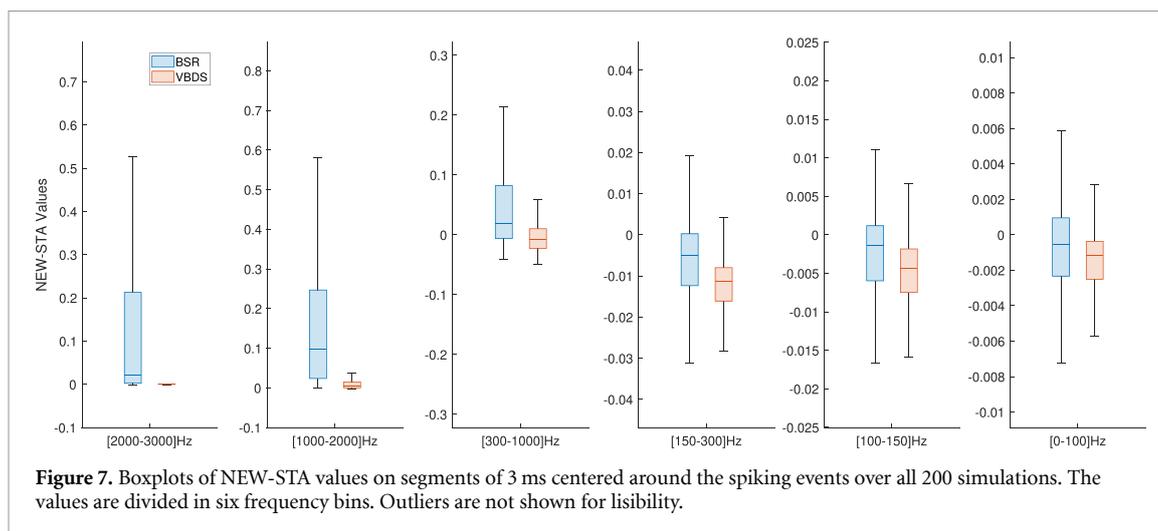
Figure 6. (a) Example of despiked versions of 250 ms of a simulated signal by the 3 tested methods (ASR, BSR and VBDS). The traces are shifted for visualization purpose. The presence of 4 spike events are materialized by red dots on top of the traces. (b) Spike-triggered average of the signal for one of the 200 simulated examples. ASR fails in retrieving the underlying Gabor-like simulated synchronized activity, spike contamination is still visible in the BSR approach, while no visible clues of any spike presence are remaining in the proposed VBDS approach. (c) Despiked versions of the signal (BSR and VBDS) superimposed with ground-truth on 15 epochs randomly chosen within the same simulation signal used in (a) and (b). (d)–(f) Wide-band normalized error of the wavelet STA (NEW-STA) values for ASR, BSR and VBDS respectively.

events selected randomly, and where residual artifacts leaved by the BSR approach are systematically visible, while our method provides smoother versions of the signal very close to the ground-truth. The probabilistic approach used in our method allows to deal with variations in the spike waveforms from an event to another. We remind that these variations are exclusively due to sampling effect in this simulation context, but in real data they might also arise from other instrumental or biological effects in real recordings [5]. Our method provides an individual posterior estimate for each spike and leaves no residual in the wide-band signal.

Figures 6(d)–(f) bring quantitative results under the form of normalized wavelet spike-triggered average (W-STA). W-STA, as defined in [49], averages the wavelet coefficients power of spike-triggered epochs of the signal. We introduce here a normalized error version of W-STA as NEW-STA by taking the difference between the wavelet STA of the despiked signal and this of the original spike-free signal (i.e. the real LFP epoch with additive Gabor-like synchronized activity, free of spike), divided by the wavelet STA of this original spike-free signal. The averages of these values over the 200 simulated signals are given on figures (d)–(f) for the three tested despiking algorithms (wide-band frequency range up to 5000 Hz). The ASR method fails in eliminating

high frequency components of the spikes with positive averaged normalized difference well above 1 (sub-figure (d), colorbar values clipped between 1 and -1), and in preserving the lower frequency components of the LFP with averaged normalized wavelet STA values going down to -0.94 around the instant of the spike within the frequency band [50–100] Hz. In the case of the Bayesian despiking method of Zanos *et al*, spike high frequency residuals are also visible over 500 Hz with averaged normalized error wavelet STA values going up to 1.2 at the time instant of the spikes, but the method leaves unchanged the power of the signal outside the time support of the spikes as well as in the lower frequency band. The proposed VBDS approach is very efficient in eliminating the high power components of the spikes in the high frequency band, the NEW-STA remains below 0.13 in average, and also preserves the low frequencies where the main features of the LFP are lying.

On figure 7 is given a focus on the NEW-STA values over epoch of 3 ms around the spiking times, the values being divided in 6 frequency bins. High relative error NEW-STA values are found for the BSR method in the higher frequency band over 1000 Hz as some spiking artefacts are left in the signal, while our VBDS approach produces very low relative error values. In the [300–1000] Hz frequency band where most of the energy of the spikes is lying (see figure 4), both



methods are producing relative errors with medians close to 0 and similar interquartile distance, however BSR tends to leave some spike energy in the signal while VBDS tends to eliminate more components, with however quartile absolute values below 0.1. This remark holds in lower frequency bands where BSR and VBDS are both producing low relative error values which become negligible ranging between 0%–1% in absolute values.

These results confirm that most of the spike features are indeed kept by our spike/LFP separation method, which is a consistent result with regard to the spiking classification performance described in the previous section. In the following, we compare the performance of the methods in the light of real micro recording data.

3.3. Real data

We now illustrate the sorting and separation performance on a set of real data recorded at the CHU of Nancy during FPVS protocol. We consider eight Ad-Tech Behnke-Fried micro-electrodes implanted in the left fusiform gyrus of a 23 year old drug-resistant epileptic patient (left mesial temporal lobe epilepsy) during presurgical evaluation. The patient gave written informed consent and the study (REUNIE, trial N° 2015-A01951-48, ClinicalTrials.gov identifier NCT02877576) was approved by the local ethical committee (CPP Est III, N°16.02.01). The protocol consists in presenting object images on a screen at the base frequency of 6 Hz, with one oddball every five images being a Human face [44]. The patient followed two consecutive sessions of length 70 s with resting state interval of 2 min between them. Each session has been cut out so that the first time instant corresponds to the first face trigger of the session, and lasts up to $1/1.2 = 833$ ms (i.e. one oddball interval) after the last face trigger of the session. The two sessions are concatenated and fed to the spike sorting and despiking algorithms. We first provide spike sorting results for each of the four evaluated spike sorters, and

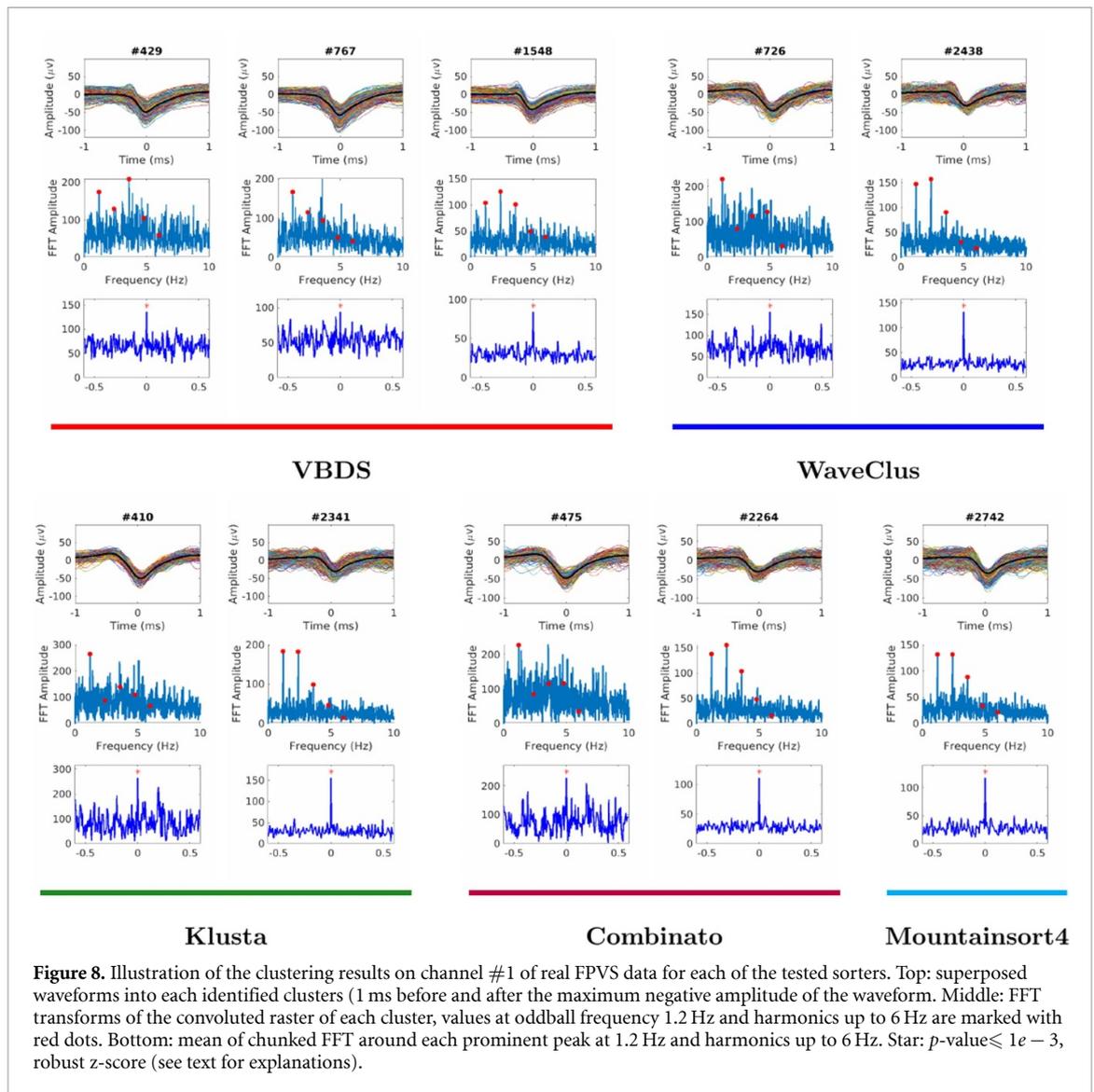
we then illustrate the despiking performance of the method by considering the depollution of one channel with respect to a 1.2 Hz responding spike cluster.

3.3.1. Spike sorting

FPVS leads to periodic activity in brain structures sensitive to the specific visual stimulation [19, 44]. Frequency analysis of the response of individual neurons as identified by spike sorting algorithm should exhibit high peaks at the frequency of the stimulus as well as these of its harmonics. The frequency analysis of the discrete spiking events is carried as follows: the spike train of the considered cluster is convolved with a 20 ms-length square window, yielding an estimate of the firing rate, on which the Fourier transform is performed. Two stimulus are superimposed in this protocol, the base stimulation at 6 Hz and the oddball at 1.2 Hz. Responding clusters at 1.2 Hz have been identified by most of the methods on six of the eight channels, while our investigation did not allow to identify neurons responding at 6 Hz (the implanted region—fusiform gyrus—being known for its specificity for face selectivity [44]). We provide in detail the sorting results for two channels (denoted *Channel#1* and *Channel#2* in the following), on which several responding clusters have been identified.

To determine whether a cluster responds or not, we use the methodology as described in [36]: the FFT of the estimated firing rate (convolved raster) is epoched around 1.2 Hz as well as around each of its harmonics 2.4, 3.6 and 4.8 Hz (0.3 Hz width epoching, 40 frequency bins). These frequency epochs are summed together by considering the combinations between them yielding the highest central peak. A z-score with respect to the background spectrum is computed. If the central peak has a *p*-value below 0.01, then this is considered as a significant cluster.

We compare our method with Klusta, WaveClus, Combinato as well as Mountainsort4, as done in section 3.2.1 using the spikeinterface toolbox [4]. The

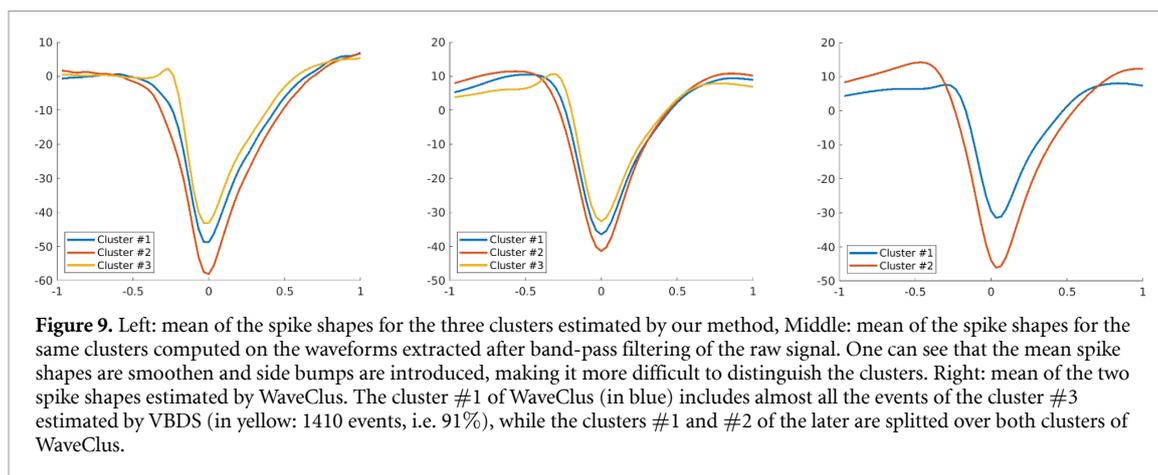


detection thresholds for each sorter has been set to 4.5 times the mad of the band-pass filtered signal, the set of spiking events to be sorted by the four methods are nearly the same up to the specific detection parametrization and event rejection strategies embedded in each method. For each of them, we start by setting the parameters recommended by the authors, and we tune them over the six channels in order to get meaningful results overall, i.e. no obvious over or under-clustering, and visually homogeneous clusters. In particular for Klusta, 9 principal components were used, with a spike support from -30 ms to 30 ms around the maximum peak of the spiking events. For Combinato, we use the default set of parameters as described in [27], except for the maximum number of clusters at one temperature parameter that we set to 7, and the minimum cluster size for iterative clustering set to 1000.

VBDS has been applied in the same way as on the simulated data, based on GMM-EM initialization. The selection of 8 wavelet coefficients (using mother

wavelet Sym6) was found to be more relevant for these data (instead of 10 on simulated data).

Note that in this context of real recordings, it is preferable to consider mechanism to reject artifact events that might pollute the classification process. In this work, we first discard the events with too high amplitudes (in absolute values), this second high (outlier) threshold being set to about five times the value of the robust detection threshold as defined in equation (15). Another rejection procedure is applied during GMM-EM initialization, as Gaussian cluster parameters are initialized based on the wavelet decomposition of the spike waveforms extracted from the band-passed raw signal. Based on this Gaussian characterization of the cluster, it is then possible to reject detected events that are too far from their cluster centroid, similarly as proposed in recent studies [22, 43]. A multivariate Hotelling statistical test has been applied, with p -value set to 0.01. The same procedure is also applied after final convergence of the algorithm.



Inter-spike interval histograms were inspected for each cluster to check for refractory time interval violations. The significance of the difference between clusters within each sorter outcomes have been confirmed by multivariate Hotelling statistical test (p -value < 0.01). Overall, the results produce by the sorters are consistent one from the other. Our method as well as Combinato tend to identify more clusters (15 and 16 respectively over the 6 considered channels), while WaveClus produced the lower number of class (11). This is consistent with the ability of our clustering method to identify more true positive clusters as demonstrated on the simulated data set of section 3.2.1. In the following, we qualitatively analyze the clustering results for the channels #1 and #2.

3.3.1.1. Channel #1

WaveClus, Klusta as well as Combinato agreed upon two responding clusters, while VBDS finds three responding clusters and Mountainsort4 renders a unique cluster. A careful examination of the extracted spike shapes with high-pass filtering show that these shapes are corrupted, as can be seen on figure 9. We emphasized on this issue in section 3.1, the applied band-pass (non-causal) filtering introduces side bumps on the detected spike shapes and can mask out the discriminant characteristics of the real spikes, resulting in different clustering results between VBDS and the three other sorters (illustrated for WaveClus in figure 9, but similar for the two other sorters).

3.3.1.2. Channel #2

is highly responding and yields three to five clusters depending on the sorters, all of them responding to the stimulation protocol (see figure 10). In particular, our method identifies a responding cluster with 76 events, not identified by the four other methods. This is here again very likely due to the effect of the filtering diminishing the amplitude of the spikes for this cluster, added to the fact that clusters with few events are more difficult to identify. On figure 11

it can be seen that this cluster is easily identifiable when the spikes are extracted from the raw signal by our method, while it is less obvious (but still visible) when considering the band-pass filtered version of the spikes.

A careful inspection of the mean spike shapes for cluster #2 produced by VBDS reveals the presence of a slight but visible positive wave on the left side of the central peak, while cluster #1 remains flat, with a lower amplitude of its negative peak on average. Similarly as for channel #1, the spike shapes as produced by WaveClus (see figure 11, right column, similar observations for Klusta and Combinato) are corrupted by filtering (introduction of side bumps) and are diminished in amplitude, making the differentiation of these two clusters more challenging.

Besides, these results based on four different sorters confirm the difficulties to find consistent results from a sorter to another, as pointed out and illustrated in [4]. Validation and comparison on real data remain a difficult task, however we provide evidences indicating that our method provides more confident outcomes by extracting more accurately the spike shapes from the data.

3.3.2. Spike removal

We now analyze the despiking performance on the channel #1 of the same data set, and we use the clustering results of our algorithm to construct the spike trains necessary for applying the Zanos et al BSR method. The results are illustrated on figure 12 on a 250 ms epoch of this signal, containing 10 clear spiking activities detected by thresholding. This example confirms that the ASR approach highly impact the content of the (low frequency) LFP signal, as can be seen on sub-figures (a) and (d). On the despiked versions of a 250 ms epoch of the signal (sub-figure (a)), BSR clearly leaves some residuals for some spikes (particularly visible by zooming over the last 2 spikes within this epoch), while VBDS do not leaves any apparent artifact. On sub-figure (b), we give the temporal STA by considering the spike times of the

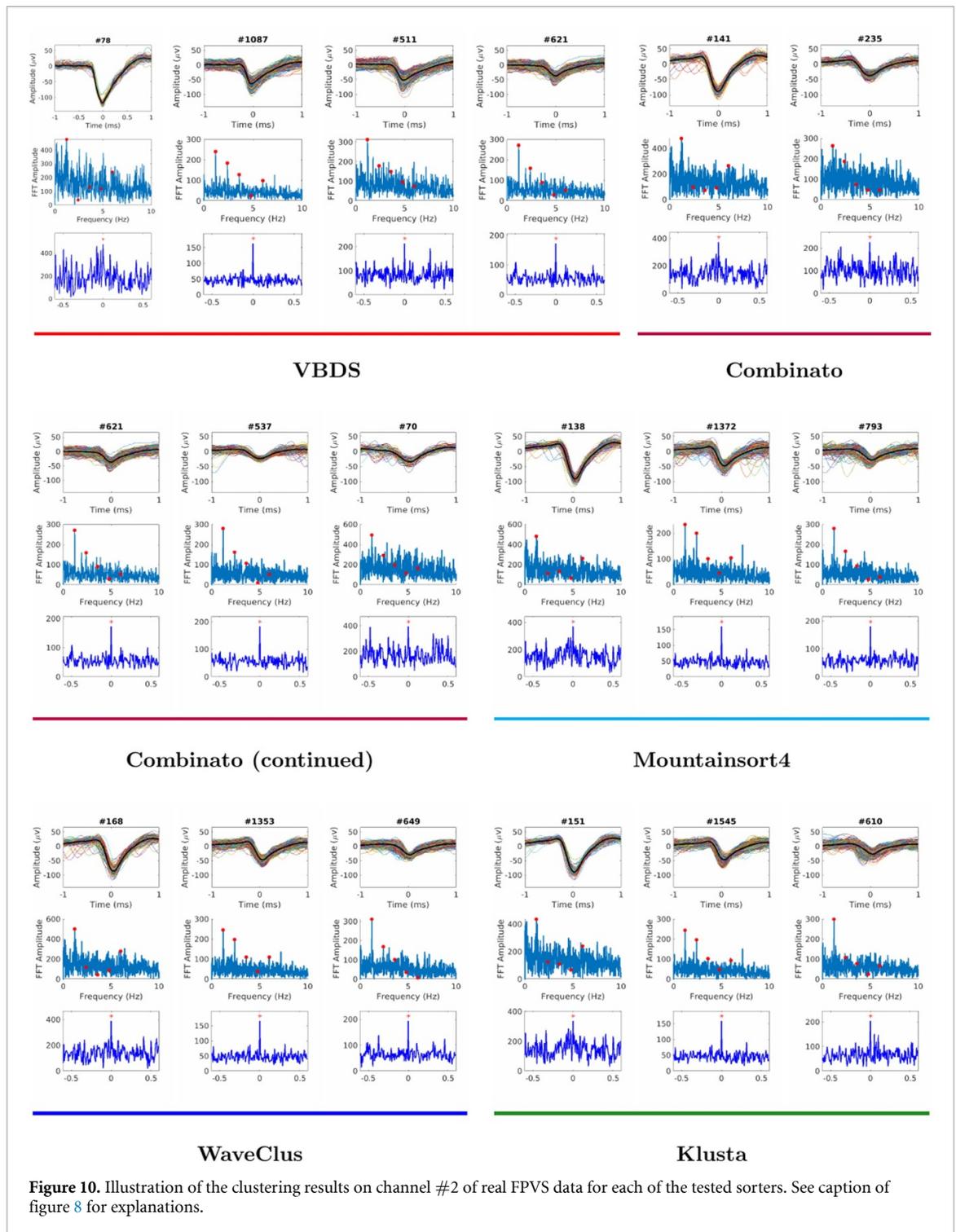


Figure 10. Illustration of the clustering results on channel #2 of real FPVS data for each of the tested sorters. See caption of figure 8 for explanations.

cluster #1 as estimated by our algorithm. The STA has been computed based on the 429 events of this cluster, resulting in a quasi-linear estimation by the BSR algorithm, while our algorithm provides a much less stereotyped version which is more likely to represent the real underlying activity.

The wavelet STA has been computed on the raw signal on subfigure (c), with power values limited to 1000 (clipped colorbar). Prominent power in high frequency bands at the instant of the maximum peak of the spikes is visible. On subfigures (d)–(f) are given

the wavelet STA computed on the despiked versions of the signal using ASR, BSR and VBDS respectively. The range of the colorbar (power values) for these figures are kept to the ones of the raw wavelet-STA of figure (c), in order to appreciate the relative difference between the spike removal results and the raw version. The ASR method tends to suppress all activities including these in the low frequency band related to the LFP. The BSR approach (sub-figure (e)) leaves high frequency components related to the spikes over 500 Hz, while our method suppress efficiently all

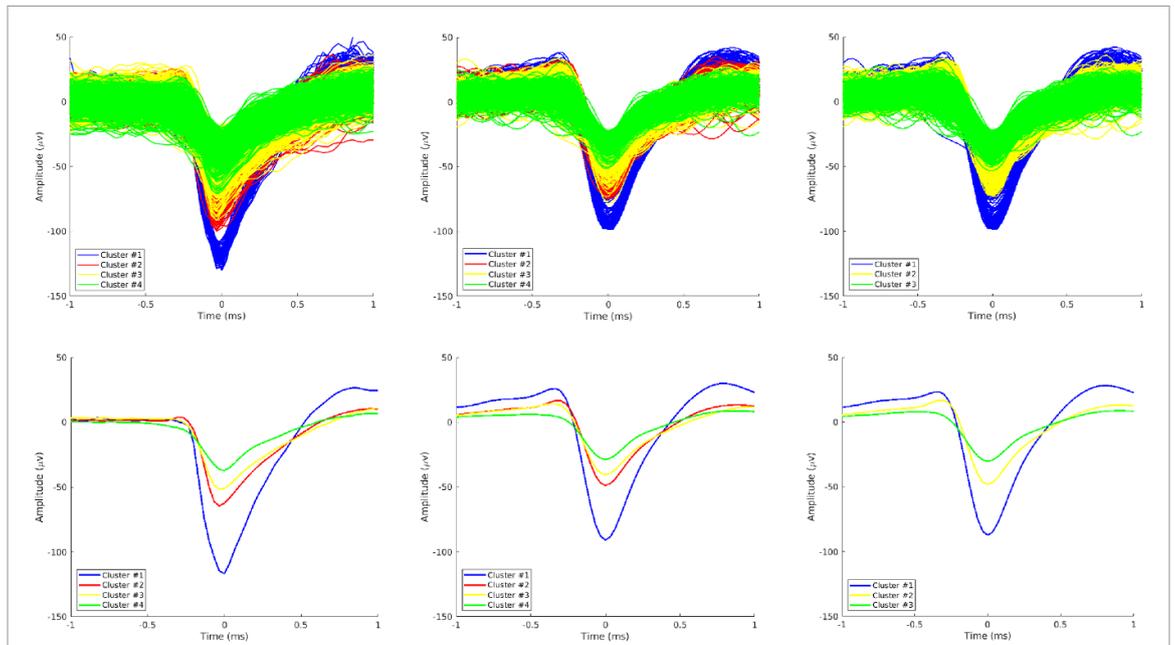


Figure 11. Top: superimposed waveforms for each cluster identified on channel #2, bottom: mean spike shape for each cluster. Left: waveforms as extracted from the LFP and clustered by our method, in particular the cluster #1 with 78 events is clearly visible and easily distinguishable from the other events. Middle: spike waveforms extracted by band-pass filtering of the signal and represented as clustered by VBDS. Right: spike waveforms as estimated and clustered by WaveClus. Cluster #1, which include all the events of the VBDS Cluster #1, also includes spikes of lower amplitude. Clusters #2 and #3 of VBDS are redistributed within clusters #2 and #3 of WaveClus, while cluster #3 of WaveClus include all events classified in the cluster #4 of VBDS.

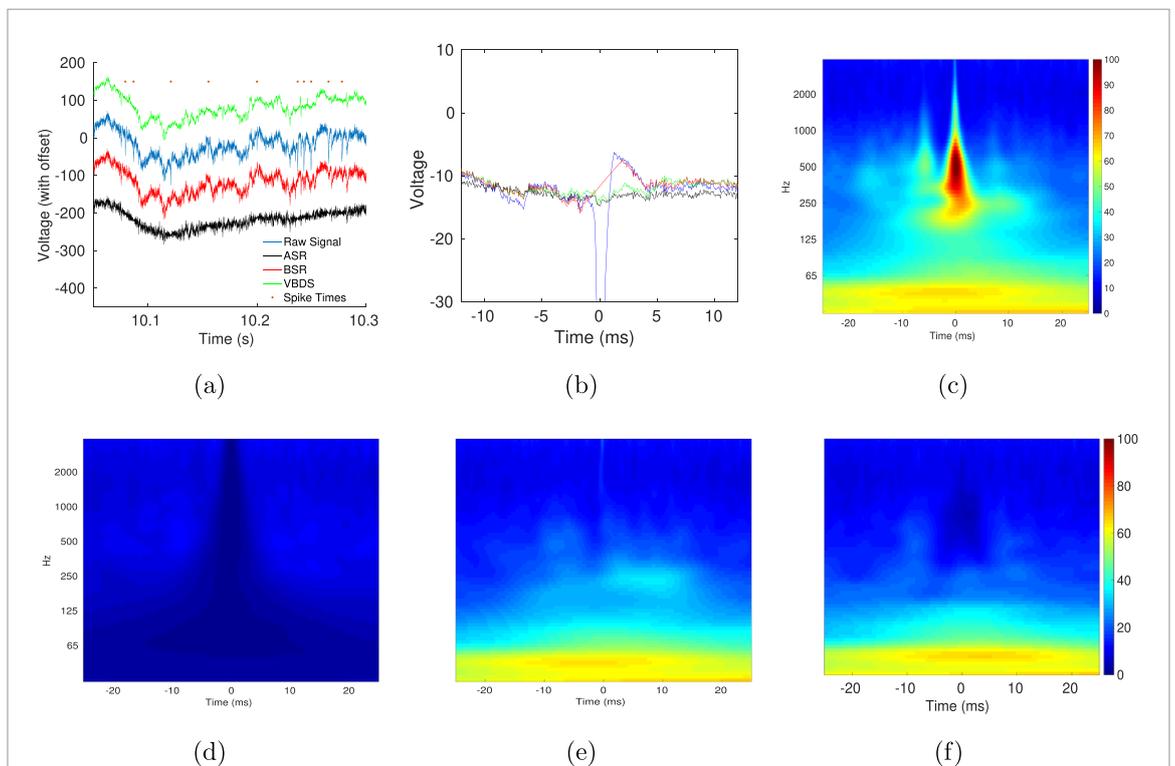


Figure 12. (a) Illustration of the despiking results on a segment of a real signal recorded during FPVS: raw signal (blue), Automatic Spike removal (ASR, black), BSR (red), and VBDS (green). Traces are shifted to appreciate the difference between the methods. Example of despiked versions of 250 ms of a simulated signal by the three tested methods (ASR, BSR and VBDS). The traces are shifted for visualization purpose. The presence of 10 spike events are materialized by orange dots on top of the traces. (b) Spike-triggered average of the raw signal over 200 spiking events randomly selected among the 1566 ones, super-imposed with the STA versions produced by ASR (black), BSR (red) and VBDS (green). (c) W-STA of the raw signal, where the power of the spiking events is clearly visible around above 100 Hz and particularly in the band [300–5000] Hz. (d)–(f) W-STA of the despiked version of the signal as produced by ASR (d), BSR (e) and VBDS (f).

patterns of the spiking activities. In lower frequency band under 300 Hz both approaches produce very similar W-STA maps, and seems to preserve the components of the LFP. Below 150 Hz, both methods preserve the components of the signal, with relative difference values (difference of W-STA divided by the W-STA of the raw signal) ranging from -0.23 to 0.04 for BSR and -0.21 to 0.03 for VBDS.

4. Discussion

We present in this paper a novel method for separating the LFP signal from the spikes and simultaneously sorting the spikes in mono-channel micro-electrode recordings. We demonstrate that the method outperformed recent methods of the literature on single channel spike sorting, using a freely available simulated data-base. We also provide evidence that the method accurately removes the spike shapes from the LFP on home made simulated data, and outperform previous work [49], especially in high frequency band over 200 Hz. These results give evidence that both aspects of the algorithm (despiking and sorting) benefit one from each other: accurate extraction of the spike waveforms from the LFP channels enhance the classification accuracy, while the spike decontamination is more accurate when the waveforms are properly removed by considering individual posterior estimates of each individual waveform.

The main flaw that may limit the use of the algorithm is related to its computational cost, many trials being necessary to avoid local minima in the optimization process. Running time can be relatively high when compared to other sorters of the literature, in particular if the number of units in the signal is high and/or if large range of class numbers are to be explored. This algorithm brings however the advantage to offer a despiked version of the LFP as an output, together with the spike clustering. Offline studies of brain responses to a particular protocol (e.g. auditory or visual evoked potential) or condition (sleep, epileptic state) might find interest in this methodology which allows analysis of multi-scale relationships analysis unbiased by the presence of prominent spike residuals that are left on the LFP after standard low-pass filtering.

Some important aspects and current issues in the field of spike sorting are not here directly addressed. The first one is related to the drift of the spike shapes [5, 27], in particular during long-term recording, mainly due to relative movement between electrodes and the recorded tissues, or exhaustion of the recorded neuron. Our approach being probabilistic (GMM of the spike wavelet coefficients), it can handle such uncertainties on the spike waveforms if not excessive, but it cannot deal with continuous drift over time. As an alternative, and based on the hypothesis that the waveforms are rather stable through time windows with few minutes length, the signal can

be divided into shorter epochs [27]. In this case, the cluster parameters obtained on a given time window can be used to initialize the algorithm in the next time window, however leaving the algorithm to re-estimate the possible evolution of the spike statistics.

The second aspect is related to the overlapping of spikes. Very few works are currently addressing specifically this problem in their model (e.g. see [46]), most of the methods are based on post-processing of the spiking events that do not match any clusters (based e.g. on multivariate statistical tests). A template matching strategy can be applied to identify overlapping events belonging to two (or higher number of) units, such as done in Spiking Circus [47] or Tridesclous [16] for example.

The scaling of the method to high dimensional data is also an important aspect to be considered, as the recording technologies are now providing hundreds of recording sites within limited tissue volume. While this is beyond the scope of the present work, we acknowledge from a data analysis point of view that such recordings provide a wider vision of a given phenomenon, and should enhanced significantly the quality of the sorting results.

Data availability statement

The data cannot be made publicly available upon publication because they contain sensitive personal information. The data that support the findings of this study are available upon reasonable request from the authors.

Acknowledgment

We thank Bruno Rossion (DR CNRS, CRAN, Université de Lorraine) for fruitful discussions and for comments that greatly improved the manuscript.

Appendix A. Variational Bayesian updates

A.1. Latent variable posteriors—E-step

We detail here the approximate posterior estimation (or *factor*) $q(\cdot)$ of the latent variables $\mathcal{X} = \{x_w, S_w, Z\}$. The factor $q(\chi_i)$ for a given latent variable χ_i is evaluated as the expectation of the full model log-likelihood \mathcal{L} with respect to all the other independent factors $\chi_{j \neq i}$. Let remind the expression of \mathcal{L} :

$$\mathcal{L} = \log p(y_w | x_w, S_w) + \log p(x_w) + \log p(S_w | Z) + \log p(Z).$$

As Gaussian priors have been chosen for each of the latent variables, each approximate posterior $q(\cdot)$ will also be a Gaussian distribution. Let remind the terms appearing in this quantity:

A.1.1. $q(x_w)$

By successively preserving only the terms in x_w :

$$\begin{aligned} \log q(x_w) &= \langle \mathcal{L} \rangle_{S_w, Z} \\ \log q(x_w) &\propto \langle \log p(y_w | x_w, S_w) \rangle_{S_w} + \log p(x_w) \\ &\propto -\frac{1}{2\sigma_\epsilon^2} \|y_w - x_w - \hat{S}_w\|_2^2 - \frac{1}{2\gamma} x_w^T \Sigma_{x_w}^{-1} x_w \\ &\propto -\frac{1}{2} x_w^T \left(\frac{1}{\gamma} \Sigma_{x_w}^{-1} + \frac{1}{\sigma_\epsilon^2} I_T \right) x_w + \frac{1}{\sigma_\epsilon^2} x_w^T (y_w - \hat{S}_w). \end{aligned}$$

By completion of the square over the variable x_w , we identify the posterior mean and variance for x_w :

$$\hat{x}_w = \frac{1}{\sigma_\epsilon^2} \hat{\Sigma}_{x_w} (y_w - \hat{S}_w) \quad (19)$$

$$\hat{\Sigma}_{x_w} = \left(\frac{1}{\gamma} \Sigma_{x_w}^{-1} + \frac{1}{\sigma_\epsilon^2} I_T \right)^{-1} = \sigma_\epsilon^2 \gamma \Sigma_{x_w} (\gamma \Sigma_{x_w} + \sigma_\epsilon^2 I_T)^{-1}. \quad (20)$$

A.1.2. $q(S_w) = \prod_{n=1}^N q(s_{n_w})$

The posterior over each variables s_{n_w} can be considered separately, as overlapping spikes are ignored at this stage of the algorithm. Similarly as for x_w , we get:

$$\begin{aligned} \log q(s_{n_w}) &= \langle \mathcal{L} \rangle_{x_w, z_n} \\ \log q(s_{n_w}) &\propto \langle \log p(y_w | x_w, S_w) \rangle_{x_w} + \langle \log p(s_{n_w} | z_n) \rangle_{z_n} \\ &\propto -\frac{1}{2\sigma_\epsilon^2} \|\Omega_{n_w} (y_w - \hat{x}_w - S_w)\|_2^2 \\ &\quad - \frac{1}{2} \sum_{k=1}^K \hat{\pi}_{nk} (s_{n_w} - s_w^k)^T \Sigma_w^{k-1} (s_{n_w} - s_w^k) \\ &\propto -\frac{1}{2} s_{n_w}^T \left(\sum_{k=1}^K \pi_k \Sigma_w^{k-1} + \frac{1}{\sigma_\epsilon^2} I_{T_s} \right) s_{n_w} \\ &\quad + s_{n_w}^T \left(\frac{1}{\sigma_\epsilon^2} \Omega_{n_w} (y_w - \hat{x}_w) + \sum_{k=1}^K \hat{\pi}_{nk} \Sigma_w^{k-1} s_w^k \right) \end{aligned}$$

with $\Omega_{n_w}(S_w) = s_{n_w}$. Again by completion of the square, we identify the posterior mean and variance for s_{n_w} :

$$\hat{s}_{n_w} = \hat{\Sigma}_{n_w} \left(\frac{1}{\sigma_\epsilon^2} \Omega_{n_w} (y_w - \hat{x}_w) + \sum_{k=1}^K \hat{\pi}_{nk} \Sigma_w^{k-1} s_w^k \right) \quad (21)$$

$$\hat{\Sigma}_{n_w} = \left(\sum_{k=1}^K \hat{\pi}_{nk} \Sigma_w^{k-1} + \frac{1}{\sigma_\epsilon^2} I_{T_s} \right)^{-1}. \quad (22)$$

A.1.3. $q(Z) = \prod_{n=1}^N q(z_n)$

$$\log q(z_n) = \langle \mathcal{L} \rangle_{x_w, s_{n_w}}.$$

For a given label λ_k we get:

$$\log \hat{\pi}_{nk} \propto \langle \log p(s_{n_w} | z_n = \lambda_k) \rangle_{s_{n_w}} + \log \pi_k$$

which gives under exponential form:

$$\hat{\pi}_{nk} \propto \pi_k |\Sigma_w^{k-1}| \exp \left(-\frac{1}{2} \|\hat{s}_{n_w} - s_w^k\|_{\Sigma_w^k}^2 - \frac{1}{2} \text{Tr}[\hat{\Sigma}_{n_w} \Sigma_w^{k-1}] \right) \quad (23)$$

$$\propto \pi_k \mathcal{N}(\hat{s}_{n_w} | s_w^k, \Sigma_w^k) \exp \left(-\frac{1}{2} \text{Tr}[\hat{\Sigma}_{n_w} \Sigma_w^{k-1}] \right) \quad (24)$$

under the constraint $\sum_{k=1}^K \hat{\pi}_{nk} = 1$ for each n .

A.2. Parameters—M-step

Re-estimation of the parameters $\Theta = \{\gamma, \{s_w^k, \Sigma_w^k\}_{k=1}^K, \sigma_\epsilon^2\}$ are estimated by setting the derivative of the expectation of the complete-data log likelihood \mathcal{L} with respect to each of these parameters to 0.

A.2.1. γ

$$\begin{aligned} \frac{\partial \langle \mathcal{L} \rangle_{x_w, S_w, Z}}{\partial \gamma} &= \frac{\partial \langle \log p(x_w) \rangle_{x_w}}{\partial \gamma} \\ &= -\frac{T}{2\gamma} + \frac{1}{2\gamma^2} \left(\|\hat{x}_w\|_{\Sigma_{x_w}^{-1}}^2 + \text{Tr}[\hat{\Sigma}_{x_w} \Sigma_{x_w}^{-1}] \right) \end{aligned}$$

by setting this derivative to 0:

$$\gamma = \frac{1}{T} \left(\|\hat{x}_w\|_{\Sigma_{x_w}^{-1}}^2 + \text{Tr}[\hat{\Sigma}_{x_w} \Sigma_{x_w}^{-1}] \right). \quad (25)$$

A.2.2. $\{s_w^k, \Sigma_w^k\}_{k=1}^K$

$$\begin{aligned} \frac{\partial \langle \mathcal{L} \rangle_{x_w, S_w, Z}}{\partial s_w^k} &= \frac{\partial \langle \log p(S_w | Z) \rangle_{S_w, Z}}{\partial s_w^k} \\ &= \sum_{n=1}^N \frac{\partial \langle \log p(s_{n_w} | z_n) \rangle_{s_{n_w}, z_n}}{\partial s_w^k} \\ &= -\sum_{n=1}^N \hat{\pi}_{nk} s_w^{kT} \Sigma_w^{k-1} + \sum_{n=1}^N \hat{\pi}_{nk} \hat{s}_{n_w}^T \Sigma_w^{k-1} \end{aligned} \quad (26)$$

$$\begin{aligned} \frac{\partial \langle \mathcal{L} \rangle_{x_w, S_w, Z}}{\partial \Sigma_w^{k-1}} &= \sum_{n=1}^N \frac{\partial \langle \log p(s_{n_w} | z_n) \rangle_{s_{n_w}, z_n}}{\partial \Sigma_w^{k-1}} \\ &= \frac{1}{2} \sum_{n=1}^N \hat{\pi}_{nk} \Sigma_w^k - \frac{1}{2} \sum_{n=1}^N \hat{\pi}_{nk} \\ &\quad \left(\|\hat{s}_{n_w} - s_w^k\|_2^2 + \frac{\partial \text{Tr}[\hat{\Sigma}_{n_w} \Sigma_w^{k-1}]}{\partial \Sigma_w^{k-1}} \right) \\ &= \frac{1}{2} \sum_{n=1}^N \hat{\pi}_{nk} \Sigma_w^k - \frac{1}{2} \sum_{n=1}^N \hat{\pi}_{nk} \\ &\quad \left(\|\hat{s}_{n_w} - s_w^k\|_2^2 + \hat{\Sigma}_{n_w} \right). \end{aligned} \quad (27)$$

Setting equations (26) and (27) to 0 gives the re-estimates formula for the K means and variances of the spike classes:

$$s_w^k = \frac{\sum_{n=1}^N \hat{\pi}_{nk} \hat{s}_{n_w}}{\sum_{n=1}^N \hat{\pi}_{nk}} \quad (28)$$

$$\Sigma_w^k = \frac{\sum_{n=1}^N \hat{\pi}_{nk} \left(\|\hat{s}_{n_w} - s_w^k\|_2^2 + \hat{\Sigma}_{n_w} \right)}{\sum_{n=1}^N \hat{\pi}_{nk}}. \quad (29)$$

A.2.3. $\sigma_{\epsilon_w}^2$

$$\begin{aligned} \frac{\partial \langle \mathcal{L} \rangle_{x_w, S_w, Z}}{\partial \sigma_{\epsilon_w}^{-2}} &= \frac{\partial \langle \log p(y_w | x_w, S_w) \rangle_{x_w, S_w}}{\partial \sigma_{\epsilon_w}^{-2}} \\ &= \frac{T}{2} \sigma_{\epsilon_w}^2 - \frac{1}{2} \|y_w - \hat{x}_w - \hat{S}_w\|_2^2 \\ &\quad - \frac{1}{2} \text{Tr}[\hat{\Sigma}_{x_w}] - \frac{1}{2} \sum_{n=1}^N \text{Tr}[\hat{\Sigma}_{n_w}]. \end{aligned} \quad (30)$$

Setting this equation to 0 gives the re-estimates formula for the noise variance:

$$\sigma_{\epsilon_w}^2 = \frac{1}{T} \left(\|y_w - \hat{x}_w - \hat{S}_w\|_2^2 + \text{Tr}[\hat{\Sigma}_{x_w}] + \sum_{n=1}^N \text{Tr}[\hat{\Sigma}_{n_w}] \right). \quad (31)$$

Appendix B. Computational issues and circulant matrix properties

Σ_{x_w} cannot be reasonably assumed circulant as it is the case for Σ_x in the time domain, the computation of the posterior covariance matrix of x_w (requiring inversion of large matrix) cannot be efficiently computed with fast Fourier transform such as in [49]. The main interest in working with wavelet are for the clustering part, as such representation provide more relevant representation of the data for clustering purposes. The part of the algorithm dealing with the estimation of the LFP x and of its prior and posterior parameters can be equivalently carried out in the temporal domain. We benefit from readily available algorithm for fast computations of wavelet decomposition and inversion. Back in the temporal domain, circulant properties of Σ_x can be exploited to obtain efficient computation performance [49].

A $N \times N$ circulant matrix M is constructed by rotating a vector $c = [c_0, c_1, \dots, c_{N-1}]$ cyclically on its successive rows: $M = \text{circ}(c)$. It can be easily demonstrated that the eigenvectors of M are the N roots of unity, which form the basis for discrete Fourier transform (DFT), and that the eigenvalues of M are the DFT coefficient of the circulant vector c . Application for covariance matrices follows, as in this case the DFT of c corresponds to the PSD g of the signal x : $\text{fft}(c) = g$. Hence, given an estimate of g , the product of Σ_x with any T -length vector y can be

approximated using efficient algorithmic implementation of the DFT (namely the FFT) as $\Sigma_x y \approx c * y = \text{ifft}(g \otimes \text{fft}(y))$, with $*$ the convolution product, \otimes the Hadamard (element-wise) product and $\text{fft}(\cdot)$ and $\text{ifft}(\cdot)$ respectively the fast Fourier transform and the inverse fast Fourier transform. Equations (11) and (12) can then be efficiently computed by using this approximation, as well as the followings which can be easily derived in the same manner:

$$(\gamma \Sigma_x + \sigma_{\epsilon}^2 I_T)^{-1} y \approx \text{ifft}(\text{fft}(y) \oslash (\gamma g + \sigma_{\epsilon}^2 \mathbf{1}_T)) \quad (32)$$

$$\text{Tr}(\gamma \Sigma_x + \sigma_{\epsilon}^2 I_T)^{-1} \approx \sum_{f=1}^T 1 / (\gamma g(f) + \sigma_{\epsilon}^2) \quad (33)$$

$$\text{Tr}(\Sigma_x^{-1} (\gamma \Sigma_x + \sigma_{\epsilon}^2 I_T)) \approx \sigma_{\epsilon}^2 \gamma \sum_{f=1}^T 1 / (\gamma g(f) + \sigma_{\epsilon}^2) \quad (34)$$

with $\mathbf{1}_T$ a T -length ones vector, and \oslash the element-wise division.

ORCID iDs

Steven Le Cam  <https://orcid.org/0000-0001-9815-5382>

Radu Ranta  <https://orcid.org/0000-0002-0893-8998>

References

- [1] Ardid S, Vinck M, Kaping D, Marquez S, Everling S and Womelsdorf T 2015 Mapping of functionally characterized cell classes onto canonical circuit operations in primate prefrontal cortex *J. Neurosci.* **35** 2975–91
- [2] Bishop C M 2006 *Pattern Recognition and Machine Learning* (Berlin: springer)
- [3] Boroujeni K B, Tiesinga P and Womelsdorf T 2020 Adaptive spike-artifact removal from local field potentials uncovers prominent beta and gamma band neuronal synchronization *J. Neurosci. Methods* **330** 108485
- [4] Buccino A P, Hurwitz C L, Garcia S, Magland J, Siegle J H, Hurwitz R and Hennig M H 2020 Spikeinterface, a unified framework for spike sorting *eLife* **9** e61834
- [5] Buccino A P, Garcia S and Yger P 2022 Spike sorting: new trends and challenges of the era of high-density probes *Prog. Biomed. Eng.* **4** 022005
- [6] Buzsáki G 2004 Large-scale recording of neuronal ensembles *Nat. Neurosci.* **7** 446
- [7] Buzsáki G, Anastassiou C A and Koch C 2012 The origin of extracellular fields and currents—EEG, ECoG, LFP and spikes *Nat. Rev. Neurosci.* **13** 407–20
- [8] Buzsáki G and Schomburg E W 2015 What does gamma coherence tell us about inter-regional neural communication? *Nat. Neurosci.* **18** 484
- [9] Camuñas-Mesa L A and Quiroga R Q 2013 A detailed and fast model of extracellular recordings *Neural Comput.* **25** 1191–212
- [10] Chaure F J, Rey H G and Quiroga R Q 2018 A novel and fully automatic spike-sorting implementation with variable number of features *J. Neurophysiol.* **120** 1859–71
- [11] Chaure F J, Quián Q R, Kochen S S and Rey H G 2017 A novel unsupervised spike sorting implementation with variable number of features *Information Processing and Control (RPIC), 2017 XVII Workshop on (IEEE)* pp 1–7

- [12] Chung J E, Magland J F, Barnett A H, Tolosa V M, Tooker A C, Lee K Y, Shah K G, Felix S H, Frank L M and Greengard L F 2017 A fully automated approach to spike sorting *Neuron* **95** 1381–94
- [13] Donoho D L and Johnstone I M 1994 Ideal spatial adaptation by wavelet shrinkage *Biometrika* **81** 425–55
- [14] Einevoll G T, Kayser C, Logothetis N K and Panzeri S 2013 Modelling and analysis of local field potentials for studying the function of cortical circuits *Nat. Rev. Neurosci.* **14** 770
- [15] Fried I, MacDonald K A and Wilson C L 1997 Single neuron activity in human hippocampus and amygdala during recognition of faces and objects *Neuron* **18** 753–65
- [16] Garcia S and Pouzat C 2015 Tridesclous (software) (available at: <https://github.com/tridesclous/tridesclous>)
- [17] Harris K D, Henze D A, Csicsvari J, Hirase H and Buzsáki G 2000 Accuracy of tetrode spike separation as determined by simultaneous intracellular and extracellular measurements *J. Neurophysiol.* **84** 401–14
- [18] Hulata E, Segev R and Ben-Jacob E 2002 A method for spike sorting and detection based on wavelet packets and Shannon's mutual information *J. Neurosci. Methods* **117** 1–12
- [19] Jonas J, Jacques C, Liu-Shuang J, Brissart Helène, Colnat-Coulbois S, Maillard L and Rossion B 2016 A face-selective ventral occipito-temporal map of the human brain with intracerebral potentials *Proc. Natl Acad. Sci.* **113** E4088–97
- [20] Le Cam S, Tran H, Ranta R and Louis-Dorr V 2019 A Bayesian approach for simultaneous spike extraction and sorting 2019 9th Int. IEEE/EMBS Conf. on Neural Engineering (NER) (IEEE) pp 993–6
- [21] Lee H J, Carlson D E, Razaghi S H, Yao W, Goetz G A, Hagen E, Batty E, Chichilnisky E J, Einevoll G T and Paninski L 2017 YASS: yet another spike sorter *Advances in Neural Information Processing Systems* vol 30 pp 4005–15
- [22] Levi R, Valderhaug V D, Castelbuono S, Sandvig A, Sandvig I and Barbieri R 2021 Bayesian supervised machine learning classification of neural networks with pathological perturbations *Biomed. Phys. Eng. Express* **7** 065021
- [23] Lewicki M S 1998 A review of methods for spike sorting: the detection and classification of neural action potentials *Netw. Comput. Neural Syst.* **9** R53–R78
- [24] Liebe S, Hoerzer G M, Logothetis N K and Rainer G 2012 Theta coupling between v4 and prefrontal cortex predicts visual short-term memory performance *Nat. Neurosci.* **15** 456
- [25] Lindén H, Tetzlaff T, Potjans T C, Pettersen K H, Grün S, Diesmann M and Einevoll G T 2011 Modeling the spatial reach of the LFP *Neuron* **72** 859–72
- [26] Mazzoni A, Brunel N, Cavallari S, Logothetis N K and Panzeri S 2011 Cortical dynamics during naturalistic sensory stimulations: experiments and models *J. Physiol.* **105** 2–15
- [27] Niediek J, Boström J, Elger C E and Mormann F 2016 Reliable analysis of single-unit recordings from the human brain under noisy conditions: tracking neurons over hours *PLoS One* **11** e0166598
- [28] Norcia A M, Gregory Appelbaum L, Ales J M, Cottareau B R and Rossion B 2015 The steady-state visual evoked potential in vision research: a review *J. Vis.* **15** 4–4
- [29] Norman Y, Yeagle E M, Khuvis S, Harel M, Mehta A D and Malach R 2019 Hippocampal sharp-wave ripples linked to visual episodic recollection in humans *Science* **365** eaax1030
- [30] Ortiz-Rosario A, Adeli H and Buford J A 2015 Wavelet methodology to improve single unit isolation in primary motor cortex cells *J. Neurosci. Methods* **246** 106–18
- [31] Pachitariu M, Steinmetz N, Kadir S, Carandini M and Harris K 2016 Fast and accurate spike sorting of high-channel count probes with KiloSort *NIPS Proc. (Neural Information Systems Foundation, Inc.)*
- [32] Pedreira C, Martinez J, Ison M J and Quiroga R Q 2012 How many neurons can we see with current spike sorting algorithms? *J. Neurosci. Methods* **211** 58–65
- [33] Quiroga R Q 2009 What is the real shape of extracellular spikes? *J. Neurosci. Methods* **177** 194–8
- [34] Quiroga R Q, Nadasdy Z and Ben-Shaul Y 2004 Unsupervised spike detection and sorting with wavelets and superparamagnetic clustering *Neural Comput.* **16** 1661–87
- [35] Quiroga R Q and Panzeri S 2009 Extracting information from neuronal populations: information theory and decoding approaches *Nat. Rev. Neurosci.* **10** 173
- [36] Retter T L, Rossion B and Schiltz C 2021 Harmonic amplitude summation for frequency-tagging analysis *J. Cogn. Neurosci.* **33** 2372–93
- [37] Rey H G, Pedreira C and Quiroga R Q 2015 Past, present and future of spike sorting techniques *Brain Res. Bull.* **119** 106–17
- [38] Rossion B et al 2016 Spike sorting for large, dense electrode arrays *Nat. Neurosci.* **19** 634–41
- [39] Scheffer-Teixeira R, Belchior H, Leao R N, Ribeiro S and Tort A B L 2013 On high-frequency field oscillations (>100 Hz) and the spectral leakage of spiking activity *J. Neurosci.* **33** 1535–9
- [40] Scherberger H, Jarvis M R and Andersen R A 2005 Cortical local field potential encodes movement intentions in the posterior parietal cortex *Neuron* **46** 347–54
- [41] Schomburg E W, Anastassiou C A, Buzsáki G and Koch C 2012 The spiking component of oscillatory extracellular potentials in the rat hippocampus *J. Neurosci.* **32** 11798–811
- [42] Shoham S, Fellows M R and Normann R A 2003 Robust, automatic spike sorting using mixtures of multivariate t-distributions *J. Neurosci. Methods* **127** 111–22
- [43] Souza B C, Lopes-dos Santos V, Bancelo J and Tort A B L 2019 Spike sorting with Gaussian mixture models *Sci. Rep.* **9** 1–14
- [44] Voltart A, Yan X, Maillard L, Colnat-Coulbois S, Hossu G, Rossion B and Jonas J 2022 Intracerebral electrical stimulation of the right anterior fusiform gyrus impairs human face identity recognition *NeuroImage* **250** 118932
- [45] Watson B O, Ding M and Buzsáki G 2018 Temporal coupling of field potentials and action potentials in the neocortex *Eur. J. Neurosci.* **48** 2482–97
- [46] Wouters J, Kloosterman F and Bertrand A 2018 Towards online spike sorting for high-density neural probes using discriminative template matching with suppression of interfering spikes *J. Neural Eng.* **15** 056005
- [47] Yger P et al 2018 A spike sorting toolbox for up to thousands of electrodes validated with ground truth recordings *in vitro* and *in vivo* *elife* **7** e34518
- [48] Zanos S, Zanos T P, Marmarelis V Z, Ojemann G A and Fetzi E E 2012 Relationships between spike-free local field potentials and spike timing in human temporal cortex *J. Neurophysiol.* **107** 1808–21
- [49] Zanos T P, Mineault P J and Pack C C 2010 Removal of spurious correlations between spikes and local field potentials *J. Neurophysiol.* **105** 474–86
- [50] Zur G and Joshua M 2019 Using extracellular low frequency signals to improve the spike sorting of cerebellar complex spikes *J. Neurosci. Methods* **328** 108423