

## SYNAPTIC MECHANISMS

## TECHNICAL SPOTLIGHT

# Accurate spike sorting for multi-unit recordings

Takashi Takekawa,<sup>1</sup> Yoshikazu Isomura<sup>1</sup> and Tomoki Fukai<sup>1,2</sup><sup>1</sup>Laboratory for Neural Circuit Theory, RIKEN Brain Science Institute, Hirosawa 2-1, Wako, Saitama 351-0198, Japan<sup>2</sup>Department of Complexity Science and Engineering, University of Tokyo, Kashiwa, Chiba, Japan

**Keywords:** cell assembly, clustering, Efficient Technology of Spike sorting open-source software, robust variational Bayes, wavelet transform

### Abstract

Simultaneous recordings with multi-channel electrodes are widely used for studying how multiple neurons are recruited for information processing. The recorded signals contain the spike events of a number of adjacent or distant neurons and must be sorted correctly into spike trains of individual neurons. Several mathematical methods have been proposed for spike sorting but the process is difficult in practice, as extracellularly recorded signals are corrupted by biological noise. Moreover, spike sorting is often time-consuming, as it usually requires corrections by human operators. Methods are needed to obtain reliable spike clusters without heavy manual operation. Here, we introduce several methods of spike sorting and compare the accuracy and robustness of their performance by using publicized data of simultaneous extracellular and intracellular recordings of neuronal activity. The best and excellent performance was obtained when a newly proposed filter for spike detection was combined with the wavelet transform and variational Bayes for a finite mixture of Student's *t*-distributions, namely, robust variational Bayes. Wavelet transform extracts features that are characteristic of the detected spike waveforms and the robust variational Bayes categorizes the extracted features into clusters corresponding to spikes of the individual neurons. The use of Student's *t*-distributions makes this categorization robust against noisy data points. Some other new methods also exhibited reasonably good performance. We implemented all of the proposed methods in a C++ code named 'EToS' (Efficient Technology of Spike sorting), which is freely available on the Internet.

### Introduction

Clarifying how the brain processes information requires the simultaneous observation of the activities of multiple neurons. Extracellular recording with multi-channel electrodes is a commonly used technique to record the activities of tens or hundreds of neurons simultaneously, with a high temporal resolution (O'Keefe & Recce, 1993; Wilson & McNaughton, 1993; Fynh *et al.*, 2007). Each channel of such an electrode detects a superposition of signals from many neurons, and spike trains of the individual neurons can be sorted from these signals by some mathematical techniques. The fact that different channels sense spikes from the same neuron with varying degrees of attenuation, depending on the distances between the channels and the neuron, makes this sorting a little easier (Lewicki, 1998; Brown *et al.*, 2004; Buzsáki, 2004). Similar mathematical techniques can be applied to data recorded with an array of single electrodes, in which different electrodes detect signals mainly from different neurons.

Spike sorting requires three steps of analysis: (i) detecting spikes from extracellularly recorded data, (ii) extracting features characteristic of the spikes, and (iii) clustering the spikes of individual neurons based on the extracted features. In a standard method of spike sorting, the recorded signals undergo a linear band-pass filter and those with

amplitudes larger than a prescribed threshold are identified as spikes. Principal component analysis (PCA) is then used for extracting the features of spike waveforms and the expectation maximization (EM) method is used for clustering the extracted features (Abeles & Goldstein, 1977; Wilson & McNaughton, 1993; Csicsvari *et al.*, 1998; Wood *et al.*, 2004).

Other methods have also been proposed. Wavelet transform (WT) decomposes a spike waveform into a combination of time–frequency components (Mallat, 1998), among which the features can be searched (Halata *et al.*, 2000; Letelier & Weber, 2000). WT was combined with 'superparamagnetic clustering', which classifies the data without strong assumptions on their distributions (Quiroga *et al.*, 2004). A method was proposed to trace bursting spikes (Pouzat *et al.*, 2004), which can be sorted correctly as bursting spikes of the same neurons. The Markov Chain Monte Carlo algorithm was utilized to estimate the number of source neurons in spike clustering (Nguyen *et al.*, 2003) and to trace a bursting state (Delescluse & Pouzat, 2006). Spike clustering was solved with the EM method for a mixture model of Student's *t*-distributions (Shoham *et al.*, 2003) or with Bayesian inference (Wood & Black, 2008). Spike correlation analysis was shown to require careful treatment of overlapping spikes (Bar-Gad *et al.*, 2001). The detection of submillisecond-range spike coincidences was attempted with massively-parallel multi-channel electrodes and independent-component analysis (Takahashi *et al.*, 2003).

Correspondence: Dr Tomoki Fukai, <sup>1</sup>Laboratory for Neural Circuit Theory, as above.  
E-mail: tfukai@riken.jp

Received 11 September 2009, revised 27 October 2009, accepted 19 November 2009

Multi-unit data, however, are corrupted by biological noise and accurate sorting is generally difficult. In particular, the previous methods of spike sorting suffer from convergence to local minima and selection of an inappropriate model (i.e. the number of clusters). The errors left in a computer-aided sorting must be corrected by human eyes but this procedure is time-consuming and inherently suffers from subjective bias (Harris *et al.*, 2000). In the present study, we explore a method for accurate and robust spike sorting to reduce the load of manual operation. We compare several methods of spike sorting by using the data of simultaneous extracellular and intracellular recordings of neuronal activity (Harris *et al.*, 2000; Henze *et al.*, 2000). These methods include newly devised methods as well as improved versions of conventional methods. In particular, we developed robust variational Bayes (RVB) for spike clustering and a novel filter for spike detection. Variational Bayes (VB) has been used with a mixture of normal distributions (Attias, 1999), whereas RVB employs a mixture model of Student's *t*-distributions. At each stage of spike sorting, we tested known and newly developed mathematical tools, and found that an RVB-based method exhibits an excellent overall sorting performance. All of the sorting methods were solved with deterministic annealing. Neither the EM algorithm nor the variational Bayesian algorithm employs annealing in their usual descriptions. These algorithms, however, are sometimes trapped by local minima that do not correspond to optimal solutions. The deterministic annealing introduces a phenomenological 'temperature parameter' to avoid the convergence to non-optimal solutions (Ueda & Nakano, 1998; Katahira *et al.*, 2008).

We implemented all of the sorting methods tested in this study into an open-source code named 'EToS' (Efficient Technology of Spike sorting) that runs at a high speed. The preliminary results of this study were presented in Takekawa *et al.* (2008).

## Materials and methods

Below, we explain the algorithm used in the present study. The framework of our sorting method is schematically illustrated in Fig. 1.

### Spike detection

The signals were recorded with multi-channel electrodes at the sampling frequency  $\omega_s$  of 20 kHz. They first underwent a band-pass filter to remove slowly changing local field potential and high-frequency fluctuations. In this study, we compared two types of band-pass filters. The classical window method (CWM) employed a finite impulse response filter that was derived by taking a difference between two sampling functions with different frequencies. We used finite impulse response filters rather than infinite impulse response filters. The latter filters are generally faster than the former but they show frequency-dependent phase responses that make the accurate detection of spike peaks difficult. Figure 2A shows the CWM filter for the sampling rate  $\omega_s$  (inset) and its frequency-response property. The band-pass range, order and window function of the filter are 800 Hz–3 kHz, 50 and Hamming type, respectively. Figure 2B displays the frequency-response property of our finite impulse response filter constructed from a Mexican hat (MXH)-type wavelet for the same sampling frequency (inset). The filter has band-pass frequencies around  $\omega_p = 2$  kHz and the order is only 26. The wavelet is given as  $F(l) = (1 - (l/s)^2) \exp(-(l/s)^2) / (\pi^{1/4} \sqrt{3s})$  with  $s = 0.25 \times \omega_s / \omega_p$ , where  $s$  is the time length normalized by  $\omega_s$  and  $l$  is the sampling index (integer). As the two filters are symmetrical with respect to time 0, they do not show phase delays. We note that the

MXH filter with 27 sampled values (including the origin) is computationally less costly than the CWM filter with 51 sampled values. Nevertheless, the MXH filter works as efficiently as the CWM filter in low-cut filtering.

After the band-pass filtering, spikes were detected by amplitude thresholding. As the recorded spikes have negative peaks, the threshold was set to  $-4\sigma$  unless otherwise stated, where the SD of noise was estimated to be  $\sigma = \text{median}(|x|/0.6745)$  from the band-passed signal  $x$  (Hoaglin *et al.*, 1983; Quiroga *et al.*, 2004). The discrete spike waveform detected by each channel was interpolated with quadratic splines and the precise spike-firing time was defined as the time of the greatest negative peak among all detected spikes in all channels. A spike in general exhibits slightly different peak times at different channels. To avoid detecting the same spike more than once, the waveforms detected within a time window of 0.5 ms were regarded as the same spike.

Spike detection is the first step in spike sorting and is considered to affect the quantity of sorted spikes. Lowering the detection threshold enables the detection of more spikes. However, most of the detected spikes with small amplitudes are finally grouped into a contaminated cluster, hence adding no valid spike trains. Therefore, detecting more spikes does not necessarily increase the number of spikes that are suitable for further analysis.

### Feature extraction

At first glance, spike clustering is more efficient in a greater dimension, as more information on the spike waveforms is available. In practice, however, the number of clusters is underestimated, as the dimension is increased beyond a certain value. The difficulty arising from the high-dimensionality of the data space is called 'the curse of dimensionality' (Bishop, 2006) and it should be mitigated by eliminating redundant data information.

In this study, we reduced the dimension of the feature space by either extracting the principal components or selecting the coefficients of WT of spike waveforms. In the PCA, the raw data were first filtered by a 300th order 200 Hz high-pass finite impulse response filter with Hamming window function. The high order of filtering effectively eliminated the DC component from the filtered signals, which becomes a potential obstacle in spike clustering, at a relatively small cost of computations. The filtered data were resampled at 20 kHz, from  $-0.5$  ms ahead to 1.05 ms behind each detected peak time (equivalently, sampling points in the interval  $[-10 : 21]$ ), such that point 0 may coincide with the peak time. Thus, 128-dimensional (four electrodes of 32 points) data were available for each spike. We then extracted 12 principal components from these 128-dimensional data by using PCA.

The PCA, however, is not necessarily useful for clustering, as PCA merely extracts the dimension exhibiting a large variance in data distribution, whereas clustering is most effectively executed in the dimensions in which the data distribution exhibits multiple sharp peaks rather than a single broad peak. Therefore, another spike-sorting algorithm employed WT for extracting the characteristic features of spike waveforms. The raw unfiltered data were resampled at 20 kHz, from  $-0.5$  ms ahead to 1.05 ms behind each detected peak time (equivalently, sampling points in the interval  $[-10 : 21]$ ), such that point 0 may coincide with the peak time. Note that WT requires no preparatory filtering that depends on an empirical choice of cut-off frequency. We then applied the multi-resolution analysis to the spike waveform (Halata *et al.*, 2000; Quiroga *et al.*, 2004) obtained from each channel and derived its time-frequency coefficients. We used

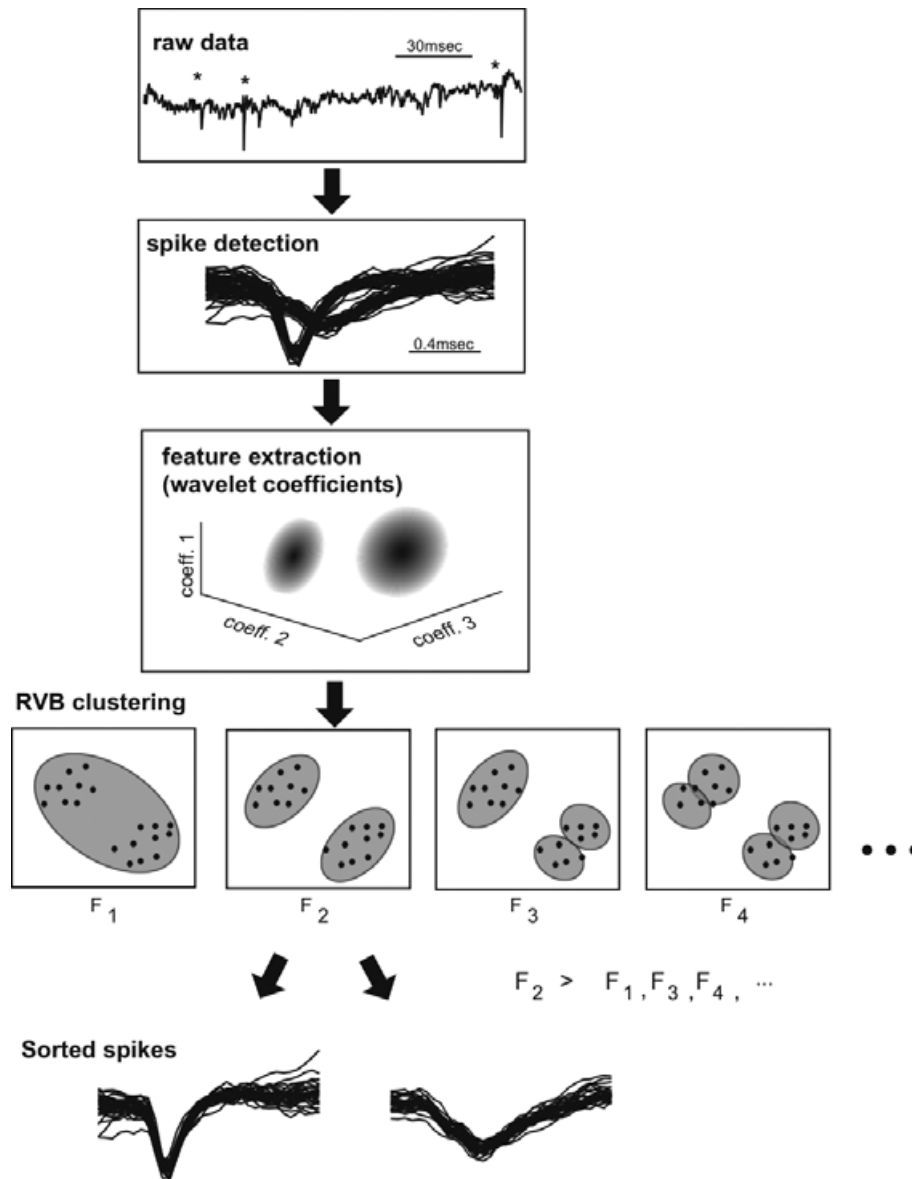


FIG. 1. Schematic illustration of our spike-sorting method. Spikes are detected by the amplitude thresholding of raw data. In the feature extraction, the detected spikes (\* marked with asterisks) were transformed into wavelets and the coefficients of WT displaying distributions with multiple peaks were selected for spike clustering by using RVB. Finally, the data in the multi-dimensional feature space were clustered by RVB into groups of spikes belonging to different neurons.

Harr's wavelet (Harr, 1910; Mallat, 1998) and the Cohen-Daubechies-Feauveau 9/7 (CDF97) wavelet (Cohen *et al.*, 1992; Daubechies, 1992). After the multi-resolution analysis, we obtained a one-dimensional distribution of each coefficient over the ensemble of spikes recorded with each channel.

A feature is only useful for separating units if it has a multi-modal distribution, i.e. a distribution with more than one peak. We reduced the dimensionality of the data by selecting the wavelet coefficients with multi-modal distributions. We evaluated each coefficient by applying the RVB clustering algorithm to the distribution of that coefficient. We computed  $F_2 - F_1$  for each coefficient, where  $F_1$  and  $F_2$  are objective functions that rate the goodness-of-fit of the model with one or two clusters, respectively (see Fig. 3). We then selected the 22 coefficients with the largest values of  $F_2 - F_1$ . Note that knowing the explicit number of peaks is not necessary for the purpose discussed here, even if the distribution is better modeled with more

than two peaks. To remove the redundancy of the extracted features, we further reduced the number of the coefficients by using PCA. Our analysis of simultaneous extracellular/intracellular recording data suggested that the present spike clustering is most accurate in the feature dimension of about 8–20 (data not shown). In this study, the dimension was fixed at 12. On the electrophysiological datasets that we analyzed, these coefficients accounted for 98% of the variance of the selected wavelet coefficients. The above reduction was crucial for suppressing the computational load and the error rate in spike clustering. Thus, spikes of the individual neurons were represented in the 12-dimensional feature space spanned by these coefficients.

The mixture of factor analyzer is known to be a powerful method of solving the curse of dimensionality. This method enables feature extraction and clustering in the original data dimension (Görür *et al.*, 2004). In our preliminary studies, however, solving the mixture of factor analyzer was time consuming and required accurate estimation

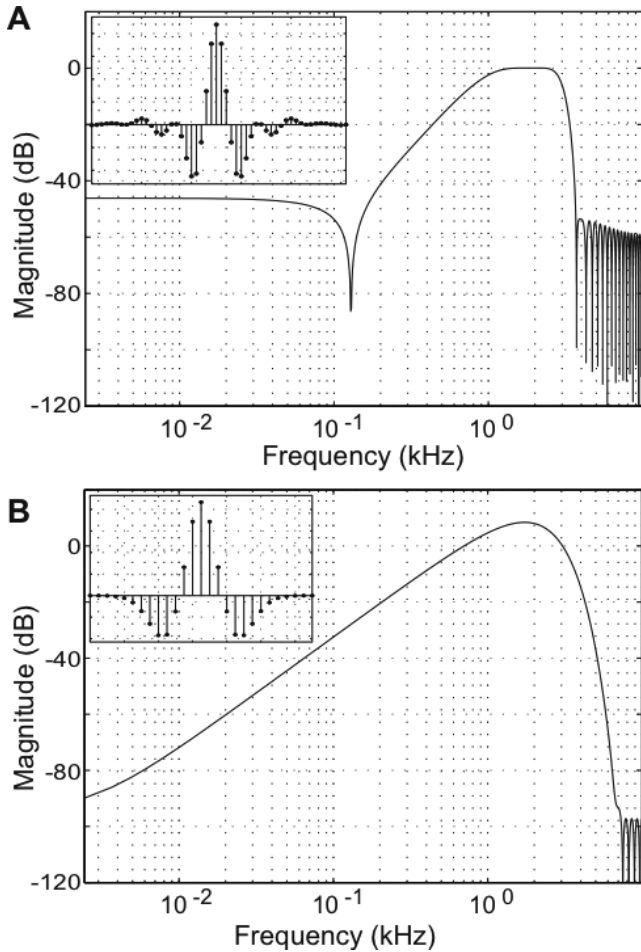


Fig. 2. Filter functions used for spike detection. (A) The frequency–response property is displayed for the 50th order CWM filter constructed through the subtraction of two sampling functions with frequencies of 800 Hz and 3 kHz. The values at both ends of the filter were set to 0 by a Hamming window function. The filter function was sampled at 51 time-points (inset). (B) The frequency–response property is shown for the filter constructed by sampling a MXH-type wavelet with a peak at 2 kHz. The Mexican hat function was sampled at 27 time-points (inset). After the band-pass filtering of recorded signals with the above filters, spikes were detected by amplitude thresholding. The two filter functions were sampled at 20 kHz.

of many parameters, which often deteriorated reliable convergence to a reasonably good solution. Therefore, we do not consider the mixture of factor analyzer in the present study. Our open software ‘EToS’, however, provides the mixture of factor analyzer as an option so that users can test it with their data.

**Clustering**

Let  $p(x_n, z_n = k|\theta, m)$  be the conditional probability that the  $n$ -th data takes a value  $x_n$  and belongs to the  $k$ -th cluster with probability  $\alpha_k$ , where  $\theta = \{\alpha_1, \dots, \alpha_m, \beta_1, \dots, \beta_m\}$  represents the set of parameters characterizing the clusters and  $m$  is the number of clusters. In this study, we fit the clusters with a normal mixture model  $p(x_n, z_n = k|\theta, m) = \alpha_k N(x|\beta_k)$  and Student’s  $t$  mixture model  $p(x_n, z_n = k|\theta, m) = \alpha_k T(x|\beta_k)$ , where  $N(x|\beta_k)$  and  $T(x|\beta_k)$  represent normal and Student’s  $t$ -distributions, respectively, and the normalized cluster size  $\alpha_k$  should satisfy  $\sum_{k=1}^m \alpha_k = 1$ . For the normal distribution,  $\beta_k = \{\mu_k, \sum_k\}$ , where  $\nu_k$  and  $\mu_k$  are the mean and variance of the distribution to fit cluster  $k$ , respectively. For the Student’s

$t$ -distribution,  $\beta_k = \{\nu_k, \mu_k, \sum_k\}$ , where  $\nu_k$  is the number of degrees of freedom of the distribution. EM and VB methods were tested in parameter estimation. Thus, we compared the performance of the following four combined algorithms: normal EM (NEM), Student’s  $t$  EM [robust EM (REM)], normal VB (NVB) and Student’s  $t$  VB (RVB).

Basic algorithms of NEM, REM, NVB and RVB were described in Dempster *et al.* (1977), Peel & McLachlan (2000), Attias (1999) and Archambeau & Verleysen (2007), respectively. The correct number of clusters is usually unknown. In the conventional EM method, we first calculate  $\rho_{nk}^{EM} = p(x_n, z_n = k|\theta^{(t)}, m)$  and  $z_{nk}^{EM} \equiv p(z_n = k|x_n, \theta^{(t)}, m) = \rho_{nk}^{EM} / \sum_m \rho_{nm}^{EM}$  for fixed  $\theta^{(t)}$  at step  $t$ , and then determine the revised parameter  $\theta^{(t+1)}$  by maximizing  $Q(\theta|\theta^{(t)}) = \sum_n \sum_k z_{nk}^{EM} \log p(x_n, z_n = k|\theta, m)$  for given data set  $\{x_1, \dots, x_N\}$ , where  $N$  is the number of data points.  $\theta^{(t+1)}$  is then set to this value and the above procedure is repeated until a stable solution is obtained for a given value of  $m$ . Data  $x_n$  is classified into the cluster that has the largest value of  $z_{nk}^{EM}$ . If, however, this value is smaller than a critical value  $z_{th}$ , the spike is regarded as not belonging to any cluster and is discarded. The solutions obtained for various values of  $m$  are examined with the minimum message length (MML) criterion (Wallace & Freeman, 1987; Figueiredo & Jain, 2000; Shoham *et al.*, 2003). Namely, we calculate the following penalized log-likelihood for different values of  $m$

$$F_m = \sum_n \log \sum_k \rho_{nk}^{EM} - \left[ \frac{N_p}{2} \sum_k \log \frac{\alpha_k N}{12} + \frac{m}{2} \log \frac{N}{12} + \frac{m(N_p + 1)}{2} \right], \tag{1}$$

where  $N_p$  is the number of parameters per component distribution (see Supporting information, Appendix S1). The second term penalizes solutions with large  $m$ , i.e. many clusters. The value of  $m$  that maximizes  $F_m$  is chosen.

The VB is a general technique to solve for the posterior probability distribution of continuous variables. It calculates an approximate distribution of the posterior, assuming that the probability variables are mutually independent. This assumption significantly reduces the cost of computations. Thus, in VB, we alternately renew the probability distributions of parameters  $z$  and  $\theta$  independently according to

$$q(z_n = k)^{(t)} \propto \int \exp q(\theta)^{(t)} \log p(x_n, z_n = k|\theta, m) d\theta = \rho_{nk}^{VB} \tag{2}$$

$$q(\theta)^{(t+1)} \propto p(\theta|m) \exp \sum_n \sum_k q(z_n = k)^{(t)} \log p(x_n, z_n = k|\theta, m) \tag{3}$$

for a given prior distribution  $p(\theta|m)$ . Here,  $q(z)$  and  $q(\theta)$  are estimates of the probability distributions  $p(z|x, m)$  and  $p(\theta|x, m)$  that we are iteratively improving. The most adequate model is the one with the number of clusters that maximizes the lower bound of the log-evidence, which can be approximated by a penalized log-likelihood as (Takekawa & Fukai, 2009)

$$F_m = \sum_n \log \sum_k \rho_{nk}^{VB} - KL[q(\theta), p(\theta|m)], \tag{4}$$

where  $\rho_{nk}^{VB}$  defined in Equation (2) is the generalized likelihood that the  $n$ -th data point  $x_n$  is likely to belong to cluster  $k$ , and the Kullback-Leibler divergence of the prior distribution and the test function is defined as



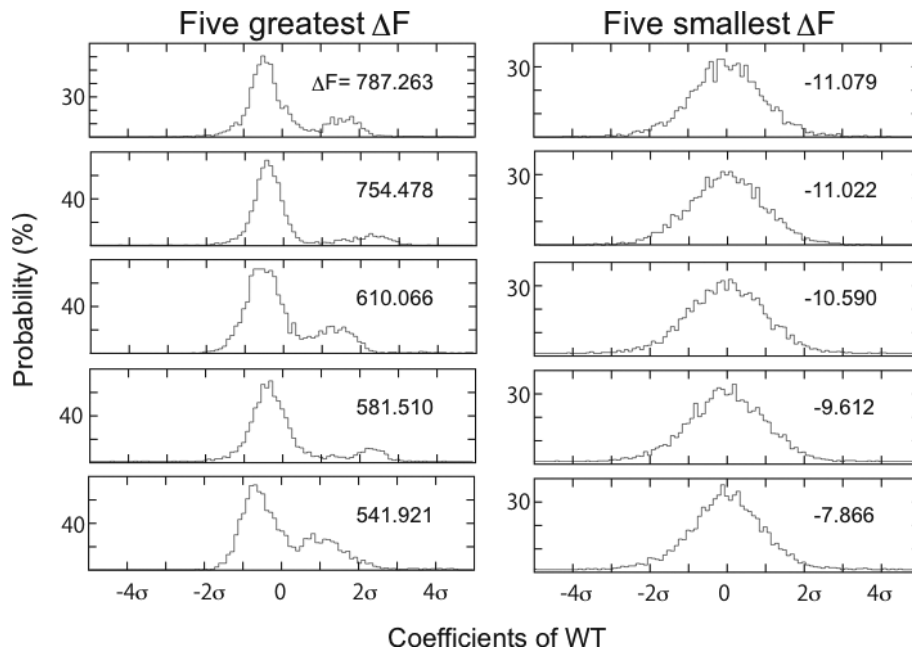


FIG. 3. The distributions of the coefficients of WT. Features of all detected spike waveforms were represented by the coefficients of WT and the coefficient distributions showing more than one peak were searched by RVB and selected for spike clustering. To this end, the difference between the optimization functions of the two models ( $\Delta F = F_2 - F_1$ ) were calculated for each distribution. If the difference is positive (negative), the distribution is better modeled with two peaks (a single peak). Examples of the coefficient distributions, which yielded the greatest five positive or negative values of  $\Delta F$ , are shown in the left and right columns, respectively.

$$KL[q(\theta|x, m), p(\theta|m)] = \int d\theta q(\theta|x, m) \log \left( \frac{q(\theta|x, m)}{p(\theta|m)} \right) (\geq 0) \quad (5)$$

To avoid the convergence to local maxima in solving NEM, REM, NVB and RVB, we introduced another trick, i.e. the deterministic annealing method (Ueda & Nakano, 1998; Katahira *et al.*, 2008). In this method, we introduce the ‘temperature parameter’  $\beta$  to replace  $\rho_{nk}$  with  $\rho_{nk}^\beta$  in the above calculations. Initially,  $\beta < 1$ . Small values of  $\beta$  eliminate valleys of the local minima that may trap the iterative solution and thus make the convergence to the global minimum easier. The value of  $\beta$  was renewed at each step  $t$  according to  $\beta = 0.01 \times 1.05^t$  until  $\beta > 1$  and thereafter was kept as  $\beta = 1$ . Initially, the number of components  $m$  should be sufficiently large and the algorithm may subsequently eliminate redundant components until this number converges. To ensure the convergence to an optimal solution, we erased the smallest cluster and compared the penalized log-likelihood between the eliminated and previous models. Calculations for a given  $m$  were repeated until  $(F_m^{(t+1)} - F_m^{(t)})/N < 10^{-6}$  was satisfied and the model with a larger log-likelihood was employed. This process was repeated until the eliminated model was rejected. The details of the algorithm and the prior for VB are described in the supporting Appendix S1.

### Implementation

We implemented our spike-sorting algorithm in a C++ code and executed it on a GNU/Linux 64-bit environment (Sun Fire X4600 M2; Quad core AMP Opteron 8384 x 8). The program code used a double-precision single-instruction-multiple-data-oriented fast Mersenne Twister pseudo-random number-generating algorithm (Saito & Matsumoto, 2008a,b). The algorithm was optimized for parallel computations in an OpenMP environment. The performance of the

program remained stable without customizing to individual data sets. Unless otherwise stated, the results shown in this article were obtained with the same set of parameter values.

### Results

We compared the performance of the following 24 ( $= 2 \times 3 \times 4$ ) combinations: the CWM filter or MXH filter for spike detection, PCA, Harr wavelet or CDF97 wavelet for feature extraction, and EM or VB for the normal mixture model or Student’s  $t$  mixture model (NEM, REM, NVB and RVB) for spike clustering. We first clarified the excellent performance of our RVB clustering methods using artificial data. The performance of the spike-sorting methods was then tested using the data obtained by simultaneous extracellular and intracellular recordings (Harris *et al.*, 2000; Henze *et al.*, 2000; data are available at <http://crcns.org/data-sets/hc/hc-1>). In these data, we knew the correct sequence of spikes, at least for a single neuron recorded intracellularly and therefore the correct answers for spike sorting were already partially known. Using this information, we examined the accuracy and robustness of the different methods.

#### *Excellent model selection performance of robust variational Bayes*

Conventional methods for spike clustering adopt the maximum likelihood estimation with a mixture of normal distributions. Normal mixtures with different numbers of peaks (i.e. clusters) represent different models of given spike data and the most suitable model (i.e. values of parameters) should be selected by a certain method, such as Akaike’s information criteria (Akaike, 1974), Bayes information criteria (Schwarz, 1978) or MML. These are different ways of penalizing complex models, i.e. models with more clusters. The virtue

of MML is that it determines a precise penalty term by taking the normalized size  $\alpha_k$  of each cluster into account. In this study, we employed MML to improve the performance of the EM method.

We constructed artificial data sets to test the clustering ability of various model selection methods. As the features of spike waveforms were suggested to obey a *t*-distribution (Shoham *et al.*, 2003), one data set consisted of artificial data points drawn from 40 Student's *t*-distributions of the degree of freedom  $\nu = 10$  in a 12-dimensional space; the data set therefore contained 40 clusters. The center of each cluster was generated by a normal Gaussian distribution of mean 0 and the variance was given as an identity matrix. The variance matrix of each cluster was generated by a Wishart distribution, with the degree of freedom at 24 and a mean of  $A$  times the identity matrix, where  $A$  takes one of the values determined equidistantly between 0.1 and 0.2. This matrix is a noisy variation of the diagonal matrix, where each diagonal element takes a value between 0.1 and 0.2. The volume of each cluster is proportional to the value of the diagonal element.

Figure 4A displays the number of clusters estimated by NEM, NVB, REM or RVB as a function of the number of data points

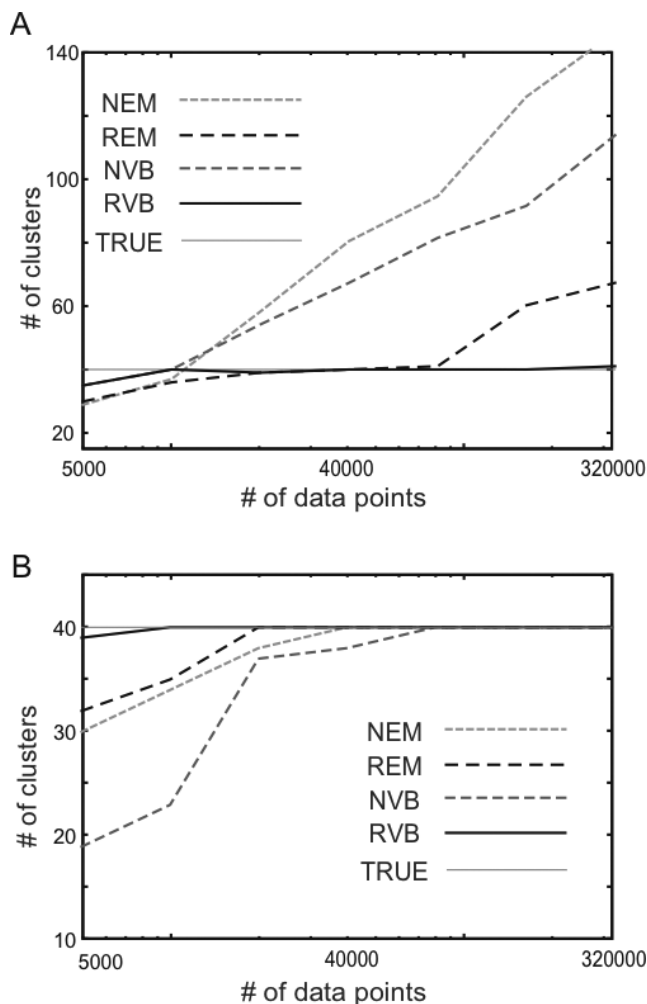


FIG. 4. Comparison of the data-size dependence between the different clustering methods. The abscissa represents the size of data and the ordinate represents the number of the resultant clusters. (A and B) Artificial data were constructed from a mixture of *t*-distributions or a normal mixture model, respectively, to contain 40 true clusters. In both data sets, only RVB was able to estimate the true number of clusters in a broad range of the data size, whereas the other methods often overestimated or underestimated the number of clusters.

sampled from data generated by a mixture of 40 *t*-distributions. NEM and NVB underestimated or overestimated the number of clusters when the data size was small or large, respectively. The methods tend to group sparse data points together in a small data set, whereas they tend to separate data points originating from a single cluster in a large data set. Thus, these methods rarely selected the correct model. REM could select the correct model if the data size was in an appropriate range. However, this method also yielded underestimation or overestimation when the data size was small or large, respectively. In contrast, RVB could estimate the correct, or a nearly correct, number of clusters in a wide range of the data size tested. The performance of the different methods was further compared on another artificial data set generated by a normal mixture model. Similarly, RVB exhibited an excellent performance for this data set (Fig. 4B).

#### Sorting of simultaneous extracellular and intracellular recording data

We then compared the performance of all of the 24 combinations of methods for spike detection, feature extraction and spike clustering by using extracellular/intracellular recording data (Harris *et al.*, 2000; Henze *et al.*, 2000). Generally, the neurons recorded with an intracellular electrode exhibited broadened spike waveforms. This fact seemed to make the separation of these spikes from the narrow spikes that were extracellularly recorded from other neurons somewhat easier. Nevertheless, the different methods exhibited rather different performance levels.

The extracellular/intracellular recording study provided data sets sampled at two different frequencies (i.e. 10 and 20 kHz). The 31 data sets sampled at 20 kHz generally have better quality and hence are more suitable for the present comparison than the 72 data sets sampled at 10 kHz. We sorted only those data sets in which the peak and width of the cross-correlogram between the intracellular and nearest extracellular spikes were less than 0.5 ms. Among the 31 data sets, five data sets satisfied these criteria (d11221.002, d11222.001, d12821.001, d14521.001 and d14521.002). On each data set, we tested each clustering method with 100 different initial conditions. The same data sets were also analyzed by KlustaKwik (K.D. Harris, <http://klustakwik.sourceforge.net>; Hazan *et al.*, 2006), a conventional spike-clustering method employing classification EM for a normal mixture model (Celeux & Govaert, 1992) with Bayes information criteria (or Akaike's information criteria, if the users choose).

Figure 5 summarizes the error counts of the different spike-sorting methods in all trials for a data set (d14521.001) that is the smallest and most difficult among the five. The data contain 181 intracellular spikes, of which both CWM and MXH detected 180 spikes. MXH-CDF97-RVB yielded, on average, 0.35 (0.19%) false-positive and 5.16 (2.85%) false-negative spikes. Without annealing, the scores were 5.17 and 4.19%, respectively (see supporting Table S1 for the results of other data sets). Results with other data sets (including those sampled at 10 kHz) are shown in supporting Figs S1 and S2. The results in each panel are arranged from left to right in an ascending order of the values of score functions as they are the only sources of information to judge the validity of sorted spikes in extracellular recordings with multi-channel electrodes. The figure displays several interesting features. For both CWM and MXH filters, the CDF97 wavelet generally yielded smaller error counts than the PCA and Harr wavelet. When, however, REM was used for spike clustering, the Harr wavelet was better than the CDF97 wavelet, implying that the overall performance of spike sorting depends on the compatibility between the methods used at the three stages. In all of the methods tested,

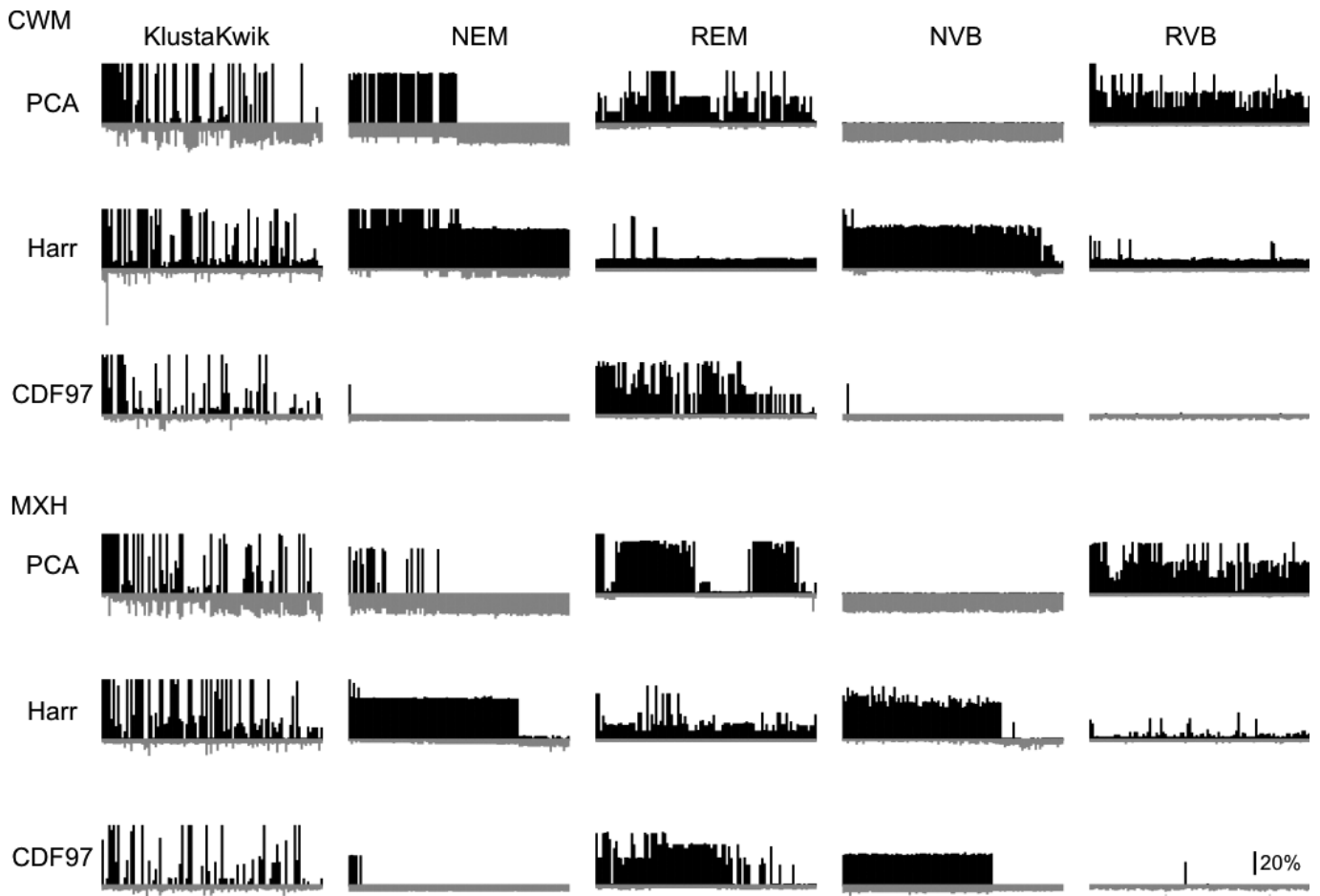


FIG. 5. Comparison between several spike-sorting methods with an extracellular/intracellular data set. Each method was tested for 100 different initial conditions, and the percentages of false positive (upward plots) and negative (downward plots) errors are displayed (from left to right) in increasing order of the corresponding score function. Upper and lower traces are for CWM and MXH filters, respectively. The leftmost panels show results with a KlustaKwik with Bayes information criteria.

MXH and CWM filters exhibited a similar quality of the overall performance. As MXH is simpler (it has only a single parameter) and computationally less costly than CWM, the use of MXH is recommended. Comparison between KlustaKwik and our NEM reveals that replacing Bayes information criteria with MML significantly improved the performance of NEM.

In Fig. 6, we show an example of the spike trains sorted by the combination of MXH (spike detection), CDF97 wavelet (feature extraction) and RVB (spike clustering) for an extracellular/intracellular recording data set. In this data set, the intracellular data contained 1260 spikes of a neuron and our spike-sorting algorithm detected a total of 3125 spikes in the extracellular data and categorized them into eight clusters, among which three clusters were contaminated (data not shown). Figure 6A displays the spike waveforms and auto-correlograms and cross-correlograms of the five valid clusters, as well as the spike distributions in the feature space. The reconstructed spike train is displayed in Fig. 6B, together with the local field potentials recorded by four extracellular channels and the intracellularly recorded membrane potential. The sorted spikes coincided well with the intracellularly recorded action potentials.

In summary, the combination of the CDF97 wavelet yielded excellent performance with NEM and NVB (several percent of false-negative and false-positive errors), and the best performance was obtained by the combination of the same wavelet with RVB (a few

percent of total errors). Unlike in clustering artificial data (Fig. 4), the performances of NEM, NVB and RVB were equally good at clustering extracellular/intracellular data. This was partly because intracellularly recorded spikes were broad and easily distinguished from the spikes of other neurons.

#### Computation time

On a single core (eight core) of central processing unit, 100 trials of spike sorting of an extracellular/intracellular data set containing about 14 000 spikes were estimated to take about 9.6 (1.6), 11.8 (1.9), 9.4 (1.5) and 9.0 (1.5) h with NEM, REM, NVB and RVB, respectively (MXH/CDF97 wavelet for spike detection/feature extraction). Our sorting program was paralleled by OpenMP and the computation time was reduced roughly in inverse proportion to the number of cores. The reduction worked more effectively for large data size.

#### Discussion

Spike sorting consists of three steps of analysis, namely spike detection, feature extraction and spike clustering. We have developed various methods for spike sorting and studied how the overall performance of spike sorting depends on different methods employed

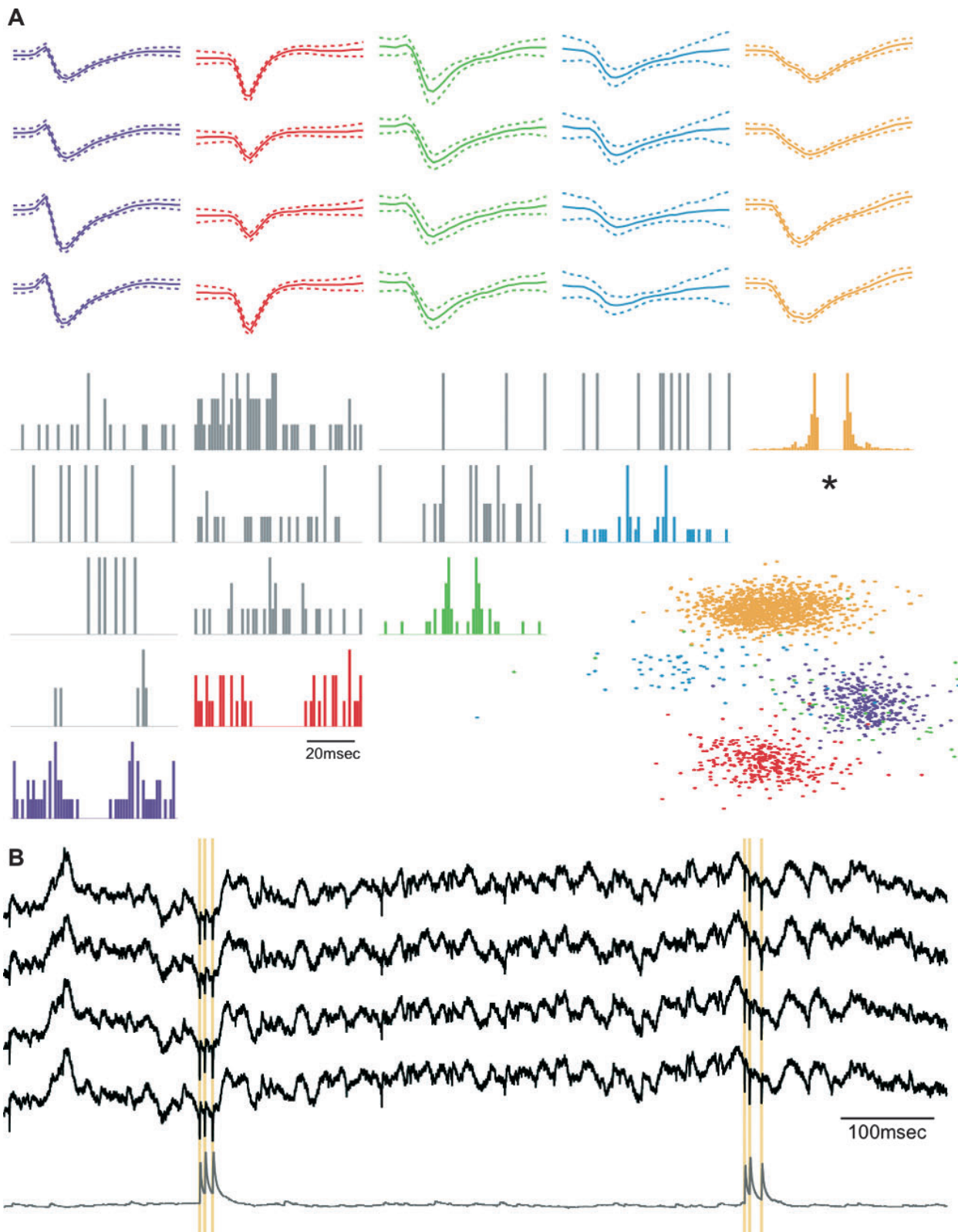


FIG. 6. An example of spike sorting, tested with the data of simultaneous extracellular/intracellular recordings (d11222.001). (A) Our RVB method identified five valid clusters. The waveforms of the sorted spikes are shown for four channels of the recording tetraode. Dashed curves indicate the SD of the recorded signals. Note that our method could separate small-amplitude spikes. The auto-correlations and cross-correlations are shown for the valid spike trains. The bin width is 1 ms. Blobs represent the clusters of spikes identified by our method in the feature space. (B) One of the valid spike trains (yellow vertical lines, marked by an asterisk in A) was superimposed on the membrane potential recorded intracellularly (lower). The upper four traces represent the local field potentials recorded extracellularly.



at each step by using simultaneous extracellular/intracellular recording data. A simple MXH filter works as efficiently as a conventional CWM filter for spike detection. The use of the CDF97 wavelet for feature extraction generally yielded much better results than the Harr wavelet. The RVB-based method that combines the MXH filter, CDF97 wavelet and RVB spike clustering showed the best accuracy and robustness in overall spike sorting. The RVB clustering method was also used to search the distributions of the wavelet coefficients useful for spike clustering, namely those coefficients distributed with more than one peak were searched and supplied to spike clustering.

The RVB, i.e. VB for a mixture of Student's  $t$ -distributions, also showed excellent performance in clustering the artificial data generated by Student's  $t$ -distributions (Fig. 4A) or normal distributions (Fig. 4B). Compared with other methods that are strongly affected by data size, our RVB method was robust against large variations in data size. The REM method also worked well in relatively broad ranges of the data size. In contrast, NEM and NVB for normal distributions showed no such robustness. In particular, the number of normal distributions (i.e. clusters) increased proportionally to data size when the data was generated by  $t$ -distributions (Fig. 4A). The Student's  $t$ -distribution possesses longer tails than the Gaussian and produced outliers, which could be covered only by an excessive number of normal distributions (Ripley, 1996; Svensén & Bishop, 2005; Archambeau & Verleysen, 2007). The better performance of RVB and REM is consistent with the fact that a  $t$ -distribution can be written by an infinite sum of Gaussian distributions (Student, 1908; Lange *et al.*, 1989; Peel & McLachlan, 2000). Although some methods for normal distributions were reasonably good in the analysis of extracellular/intracellular recording data, the above results encourage us to use the RVB method.

The primary purpose of the present study was to develop a method to accurately perform spike sorting that requires minimal manual operation. Two types of error in manual operations were previously considered in detail (Harris *et al.*, 2000). Commission errors (or false-positive errors) occur when spikes belonging to different neurons are grouped together, whereas omission errors (or false-negative errors) occur when not all spikes emitted by a single neuron are grouped together. Some human operators made false-negative errors more often than false-positive errors, whereas others exhibited the opposite tendency. The manual-operation results were significantly impinged by the subjective bias and level of experience of each operator. The RVB-based method could accurately sort simultaneous extracellular/intracellular recording data, generating just a few percent of false-positive and false-negative errors (Fig. 5). We found that smaller values of  $z_{th}$  tend to suppress the percentage of false-negative errors at the cost of a small increase in the total error (data not shown). As false negatives can affect spike coincidence analysis more strongly than false positives (Pazienti & Gruen, 2006), a  $z_{th}$  value of 0.5 to 0.8 is recommended (here,  $z_{th} = 0.8$ ).

In summary, we developed an accurate and efficient method to spike-sort multi-unit data, based on the WT and RVB. This sorting method significantly improved the reliability of spike sorting to reduce the labor and bias of manual operations. The developed software, EToS, is freely available at <http://etos.sourceforge.net/>.

## Supporting Information

Additional supporting information may be found in the online version of this article:

Fig. S1. Comparison between several spike-sorting methods with the remaining extra-intra data sets sampled at 20 kHz.

Fig. S2. Comparison between several spike-sorting methods with the remaining extra-intra data sets sampled at 10 kHz.

Table S1. Numerical results of RVB.

Appendix S1. Expectation Maximization (EM) and Variational Bayes (VB) methods for mixture normal distribution model and mixture Student's  $t$ -distribution model.

Please note: As a service to our authors and readers, this journal provides supporting information supplied by the authors. Such materials are peer-reviewed and may be re-organized for online delivery, but are not copy-edited or typeset by Wiley-Blackwell. Technical support issues arising from supporting information (other than missing files) should be addressed to the authors.

## Acknowledgements

This work was partially supported by a Grant-in-Aid for Scientific Research on Priority Areas from MEXT (nos. 17022036 and 20019035). T.T. was supported by the RIKEN Special Postdoctoral Researchers Program.

## Abbreviations

CDF97, Cohen-Daubechies-Feauveau 9/7; CWM, classical window method; EM, expectation maximization; EToS, Efficient Technology of Spike sorting; MML, minimal message length; MXH, Mexican hat; NEM, normal expectation maximization; NVB, normal variational Bayes; PCA, principal component analysis; REM, Student's  $t$  (robust) expectation maximization; RVB, robust variational Bayes; VB, variational Bayes; WT, wavelet transform.

## References

- Abeles, M. & Goldstein, M.H. (1977) Quantification, smoothing, and confidence limits for single-units' histograms. *J. Neurosci. Methods*, **5**, 317–325.
- Akaike, H. (1974) A new look at the statistical model identification. *IEEE Trans. Automat. Contr.*, **19**, 716–723.
- Archambeau, C. & Verleysen, M. (2007) Robust Bayesian clustering. *Neural Netw.*, **20**, 129–138.
- Attias, H. (1999) Inferring parameters and structure of latent variable models by variational Bayes. In Laskey, K.B. & Prade, H. (Eds), *Proceedings of 15th Conference on Uncertainty in Artificial Intelligence*. Morgan Kaufmann Publishers, San Francisco, pp. 21–30.
- Bar-Gad, I., Ritov, Y., Vaadia, E. & Bergman, H. (2001) Failure in identification of overlapping spikes from multiple neuron activity causes artificial correlations. *J. Neurosci. Methods*, **107**, 1–13.
- Bishop, C.M. (2006) *Pattern recognition and machine learning*. Springer Science + Business Media, LLC, New York.
- Brown, E.N., Kass, R.E. & Mitra, P.P. (2004) Multiple neural spike train data analysis: state-of-the-art and future challenges. *Nat. Neurosci.*, **7**, 456–461.
- Buzsáki, G. (2004) Large-scale recording of neuronal ensembles. *Nat. Neurosci.*, **7**, 446–451.
- Celeux, G. & Govaert, G. (1992) A classification EM algorithm for clustering and two stochastic versions. *Comput Stat Data Anal*, **14**, 315–332.
- Cohen, A., Daubechies, I. & Feauveau, J.C. (1992) Biorthogonal bases of compactly supported wavelets. *Commun. Pure Appl. Math.*, **45**, 485–500.
- Csicsvari, J., Hirase, H., Czurko, A. & Buzsáki, G. (1998) Reliability and state dependence of pyramidal cell-interneuron synapses in the hippocampus: an ensemble approach in the behaving rat. *Neuron*, **21**, 179–189.
- Daubechies, I. (1992) *Ten Lectures on Wavelets*. CBMS-NSF Regional Conference Series in Applied Mathematics, No. 61, SIAM ed., Cambridge Univ. Press, Cambridge.
- Delescluse, M. & Pouzat, C. (2006) Efficient spike-sorting of multi-state neurons using inter-spike intervals information. *J. Neurosci. Methods*, **150**, 16–29.
- Dempster, A.P., Laird, N.M. & Rubin, D.B. (1977) Maximum-likelihood from incomplete data via the EM algorithm. *J. R. Stat. Soc. B*, **39**, 1–38.
- Figueiredo, M. & Jain, A.K. (2000) Unsupervised selection and estimation of finite mixture models. *Proceedings of the 15th International Conference on Pattern Recognition (ICPR2000)*, Vol. 2. Barcelona, pp. 87–90.

- Fynh, M., Hafing, T., Treves, A., Moser, M.B. & Moser, E.I. (2007) Hippocampal remapping and grid realignment in entorhinal cortex. *Nature*, **466**, 190–194.
- Görür, D., Rasmussen, C.E., Tolias, A.S., Sinz, F. & Logothetis, N.K. (2004) Modelling spikes with mixtures of factor analysers. *Lect. Notes Comput. Sci.*, **3157**, 391–398.
- Halata, E., Segev, R., Shapira, Y., Benveniste, M. & Ben-Jacob, E. (2000) Detections and sorting of neural spikes using wavelet packets. *Phys. Rev. Lett.*, **85**, 4637–4640.
- Harr, A. (1910) Zur Theorie der orthogonalen Funktionensysteme. *Math. Ann.*, **69**, 331–371.
- Harris, K.D., Henze, D.A., Csicsvari, J., Hirase, H. & Buzsáki, G. (2000) Accuracy of tetrode spike separation as determined by simultaneous intracellular and extracellular measurements. *J. Neurophysiol.*, **84**, 401–414.
- Hazan, L., Zugaro, M. & Buzsáki, G. (2006) Klusters, NeuroScope, NDManager: a free software suite for neurophysiological data processing and visualization. *J. Neurosci. Methods*, **155**, 207–216.
- Henze, D.A., Borhegyi, Z., Csicsvari, J., Mamiya, A., Harris, K.D. & Buzsáki, G. (2000) Intracellular features predicted by extracellular recordings in the hippocampus in vivo. *J. Neurophysiol.*, **84**, 390–400.
- Hoaglin, D.C., Mosteller, F. & Tukey, J.W. (1983) *Understanding robust and exploratory data analysis*. Wiley-Interscience, Malden.
- Katahira, K., Watanabe, K. & Okada, M. (2008) Deterministic annealing variant of variational Bayes method. *J. Phys.: Conf. Ser.*, **95**, 012015.
- Lange, K.L., Little, R.J.A. & Taylor, J.M.G. (1989) Robust statistical modeling using t-distribution. *J. Am. Stat. Assoc.*, **84**, 881–896.
- Letelier, J.C. & Weber, P.P. (2000) Spike sorting based on discrete wavelet transform coefficients. *J. Neurosci. Methods*, **101**, 93–106.
- Lewicki, M. (1998) A review of methods for spike sorting: the detection and classification of neural action potentials. *Netw.-Comput. Neural Syst.*, **9**, R53–R78.
- Mallat, S. (1998) *A wavelet tour of signal processing*. Academic Press, San Diego.
- Nguyen, D., Frank, L. & Brown, E. (2003) An application of reversible-jump Markov chain Monte Carlo to spike classification of multi-unit extracellular recordings. *Network*, **14**, 61–82.
- O'Keefe, J. & Recce, M.L. (1993) Phase relationship between hippocampal place units and the EEG theta rhythm. *Hippocampus*, **3**, 317–330.
- Pazienti, A. & Gruen, S. (2006) Robustness of the significance of spike synchrony with respect to sorting errors. *J. Comput. Neurosci.*, **21**, 329–342.
- Peel, D. & McLachlan, G.J. (2000) Robust mixture modeling using the t-distribution. *Stat. Comput.*, **33**, 1065–1076.
- Pouzat, C., Delescluse, M., Viot, P. & Diebold, J. (2004) Improved spike-sorting by modeling firing statistics and burst-dependent spike amplitude attenuation: a Markov chain Monte Carlo approach. *J. Neurophysiol.*, **91**, 2910–2928.
- Quiroga, R.Q., Nadasady, Z. & Ben-Shaul, Y. (2004) Unsupervised spike sorting with wavelets and superparamagnetic clustering. *Neural Comput.*, **16**, 1661–1687.
- Ripley, B.D. (1996) *Pattern Recognition and Neural Networks*. Cambridge University Press, Cambridge.
- Saito, M. & Matsumoto, M. (2008a) SIMD-oriented fast Mersenne twister: a 128 bit pseudorandom number generator. In Keller, A., Heinrich, S. & Niederreiter, H. (Eds), *Monte Carlo and Quasi Monte Carlo Methods 2006*. Springer Verlag, Berlin, pp. 607–610.
- Saito, M. & Matsumoto, M. (2008b) A uniform real random number generator obeying the IEEE 754 format using an affine transition. In *Abstracts of 8th International Conference on Monte Carlo and Quasi-Monte Carlo Methods in Scientific Computing (MCQMC '08)*, p. 151. [The software is available at <http://www.math.sci.hiroshima-u.ac.jp/~m-mat/MT/SFMT/index.html>].
- Schwarz, G. (1978) Estimating the dimension of a model. *Ann. Stat.*, **6**, 461–464.
- Shoham, S., Fellows, M.R. & Normann, R.A. (2003) Robust, automatic spike sorting using mixtures of multivariate t-distributions. *J. Neurosci. Methods*, **127**, 111–122.
- Student. (1908) The probable error of a mean. *Biometrika*, **6**, 1–25.
- Svensén, M. & Bishop, C.M. (2005) Robust Bayesian mixture modeling. *Neurocomputing*, **64**, 235–252.
- Takahashi, S., Anzai, Y. & Sakurai, Y. (2003) Automatic sorting for multi-neuronal activity recorded with tetrodes in the presence of overlapping spikes. *J. Neurophysiol.*, **89**, 2245–2258.
- Takekawa, T. & Fukai, T. (2009) A novel view of the variational Bayesian clustering. *Neurocomputing*, **72**, 3366–3369.
- Takekawa, T., Isomura, Y. & Fukai, T. (2008) Accurate spike sorting of multiunit recording data based on the robust variational Bayesian clustering. *Soc. Neurosci. Abstr.*, **38**, 673–613.
- Ueda, N. & Nakano, R. (1998) Deterministic annealing EM algorithm. *Neural Netw.*, **11**, 271–282.
- Wallace, C. & Freeman, P. (1987) Estimation and inference via compact coding. *J. R. Stat. Soc. B*, **49**, 241–252.
- Wilson, M.A. & McNaughton, B.L. (1993) Dynamics of the hippocampal ensemble code for space. *Science*, **261**, 1055–1058.
- Wood, F. & Black, M.J. (2008) A nonparametric Bayesian alternative to spike sorting. *J. Neurosci. Methods*, **173**, 1–12.
- Wood, F., Fellows, M., Donoghue, J.P. & Black, M.J. (2004) Automatic spike sorting for neural decoding. *Proc 26th Ann Int Conf IEEE EMBS*, San Francisco, 4009–4012.