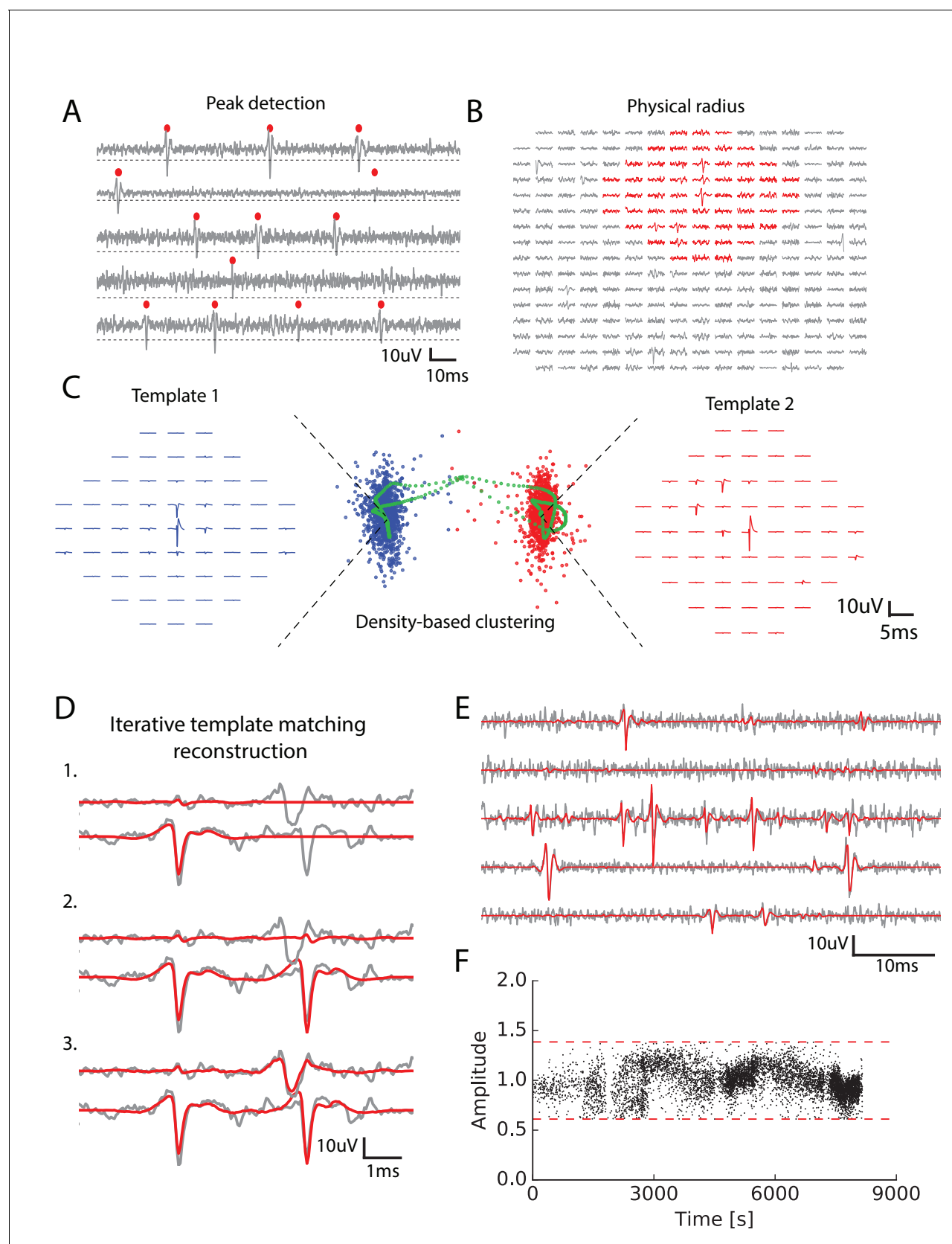


---

## Figures and figure supplements

A spike sorting toolbox for up to thousands of electrodes validated with ground truth recordings in vitro and in vivo

**Pierre Yger *et al***

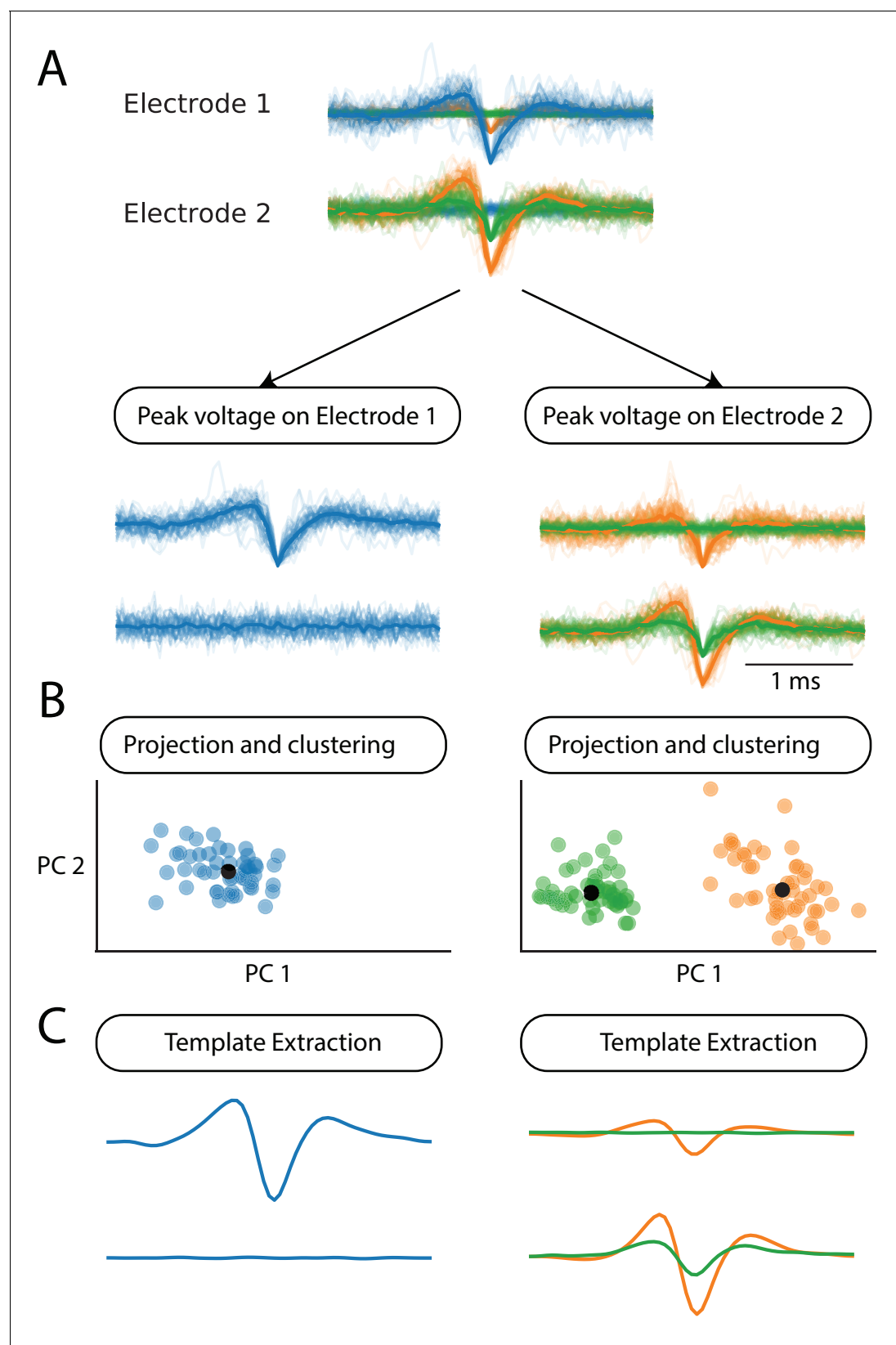


**Figure 1.** Main steps of the spike sorting algorithm. (A) Five randomly chosen electrodes, each of them with its own detection threshold (dash dotted line). Detected spikes, as threshold crossings, are indicated with red markers (B) Example of a spike in the raw data. Red: electrodes that can be  
 Figure 1 continued on next page

*Figure 1 continued*

affected by the spike, that is the ones close enough to the electrode where the voltage peak is the highest; gray: other electrodes that should not be affected. (C) Example of two clusters (red and blue) with associated templates. Green points show possible combinations of two overlapping spikes from the two cells for various time delays. (D) Graphical illustration of the template matching for in vitro data (see Materials and methods). Every line is an electrode. Grey: real data. Red: sum of the templates added by the template matching algorithm; top to bottom: successive steps of the template matching algorithm. E. Final result of the template matching. Same legend as (D, F) Examples of the fitted amplitudes for the first component of a given template as a function of time. Each dot corresponds to a spike time at which this particular template was fitted to the data. Dashed dotted lines represent the amplitude thresholds (see Materials and methods).

DOI: <https://doi.org/10.7554/eLife.34518.002>



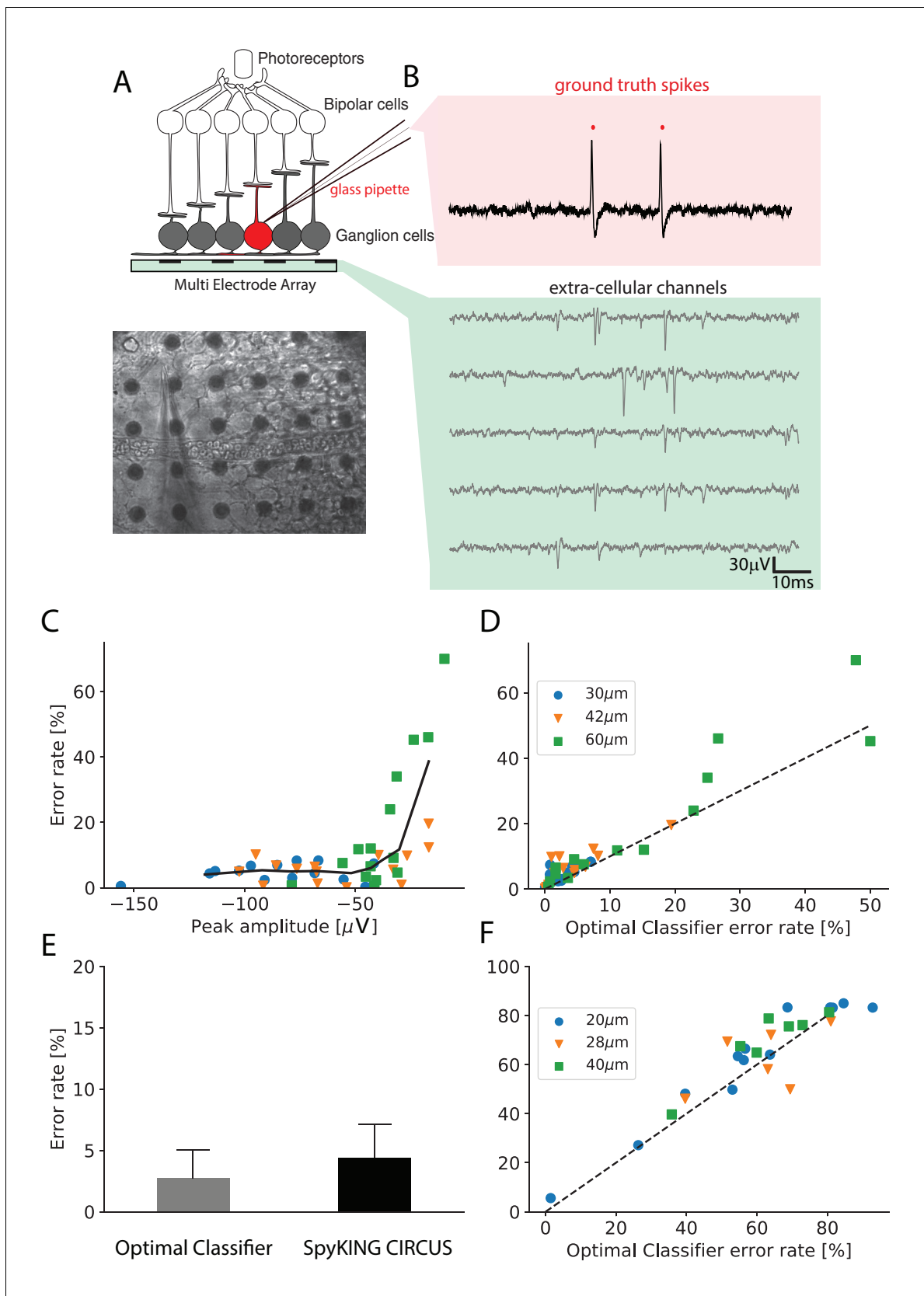
**Figure 1—figure supplement 1.** Schematic of the parallel clustering of the spikes, in a toy example with two electrodes. (A) Pre-clustering step. The different snippets are sorted according to the electrode where they peak. This divides a set of snippets in  $N_{\text{elec}}$  groups. Each of these groups is then

*Figure 1—figure supplement 1 continued on next page*

Figure 1—figure supplement 1 continued

processed independently. (B) Each group of snippets is projected in a low-dimension space, where clustering is performed using a density-based approach (see text and Materials and methods). (C) A template is extracted from each cluster and used for the template matching step.

DOI: <https://doi.org/10.7554/eLife.34518.003>

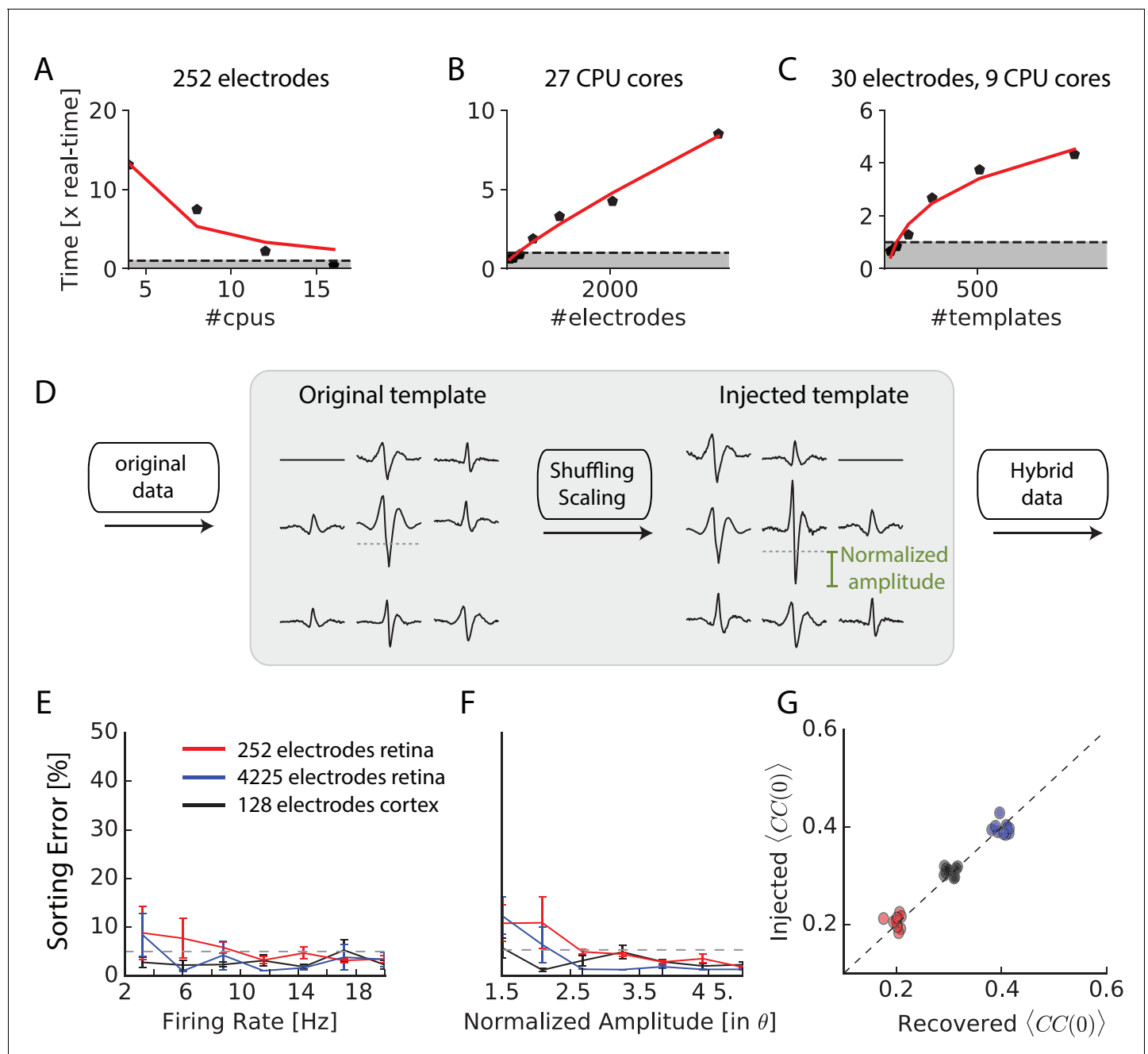


**Figure 2.** Performance of the algorithm on ground truth datasets. (A) Top: Schematic of the experimental protocol in vitro. A neuron close to the multi-electrode array (MEA) recording is recorded in loose patch. Bottom: Image of the patch electrode on top of a 252 electrodes MEA, recording a Figure 2 continued on next page

*Figure 2 continued*

ganglion cell. (B) Top, pink box: loose patch recording showing the spikes of the recorded neuron. Bottom, green box: Extra-cellular recordings next to the loose patched soma. Each line is a different electrode (C) Error rate of the algorithm as function of the largest peak amplitude of the ground-truth neuron, recorded extracellularly in vitro. (D) Comparison between the error rates produced by the algorithm on different ground truth datasets and the error rates of nonlinear classifiers (Best Ellipsoidal Error Rate) trained to detect the spikes for in vitro data (**Spampinato et al., 2018**). (E) Comparison of average performance for all neurons detected by the Optimal Classifier with an error less than 10% ( $n = 37$ ). F. Same as D. but for in vivo data (**Neto et al., 2016**) (see Materials and methods).

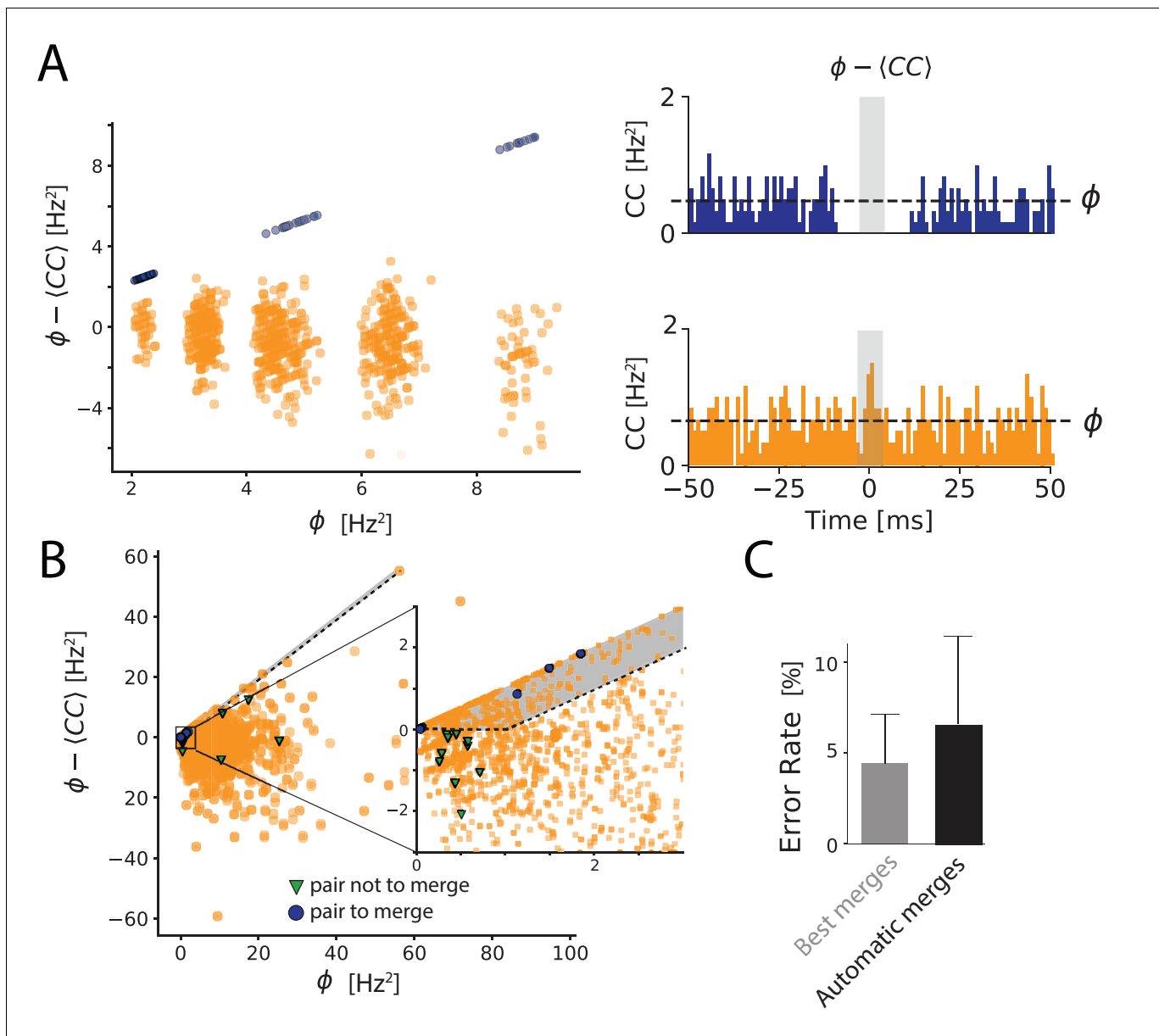
DOI: <https://doi.org/10.7554/eLife.34518.004>



**Figure 3.** Scaling to thousands of electrodes. (A) Execution time as function of the number of processors for a 90 min dataset in vitro with 252 electrodes, expressed as a real-time ratio, that is the number of hours necessary to process one hour of data. (B) Execution time as function of the number of electrodes for a 30 min dataset recorded in vitro with 4225 electrodes. (C) Execution time as function of the number of templates for a 10-min synthetic dataset with 30 electrodes. (D) Creation of 'hybrid' datasets: chosen templates are injected elsewhere in the data as new templates. Dashed-dotted lines shows the detection threshold on the main electrode for a given template, and normalized amplitude is expressed relative to this threshold (see Materials and methods). (E) Mean error rate as function of the firing rate of injected templates, in various datasets. Errors bars show standard error over eight templates (F) Error rate as function of the normalized amplitude of injected templates, in various datasets. Errors bars show standard error over nine different templates (G) Injected and recovered cross-correlation value between pairs of neurons for five templates injected at 10 Hz, with a normalized amplitude of 2 (see Materials and methods).

DOI: <https://doi.org/10.7554/eLife.34518.005>





**Figure 4.** Automated merging. (A) Dip estimation (y-axis) compared to the geometrical mean of the firing rate (x-axis) for all pairs of units and artificially generated and split spike trains (see Materials and methods). Blue: pairs of templates originating from the same neuron that have to be merged. Orange: pairs of templates corresponding to different cells. Panels on the right: for two chosen pairs, one that needs to be merged (in blue, top panel) and one should not be merged (orange, bottom panel) the full cross-correlogram and the geometrical mean of the firing rate (dashed line). The average correlation is estimated in the temporal window defined by the gray area. (B) Same as A, for a ground truth dataset. Blue points: points that need to be merged. Green points: pairs that should not be merged. Orange points: pairs where our ground truth data does not allow us to tell if the pair should be merged or not. The gray area corresponds to the region where pairs are merged by the algorithm. Inset: zoom on one region of the graph. (C) Average error rate in the case where the decision of merging units was guided by the ground truth data (left) against the automated strategy designed here (right).

DOI: <https://doi.org/10.7554/eLife.34518.006>