DEEP HIGH-FREQUENCY EXTRAPOLATION FOR NEURONAL SPIKE RESTORATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Recording neuronal activity using multiple electrodes has been widely used for studying functional mechanisms of the brain. However, handling massive amounts of data is still a challenge. In this paper, we propose a novel strategy to restore high-frequency neuronal spikes from small-volume and low-frequency band signals. Inspired by the fact that high-frequency extrapolation is equivalent to super-resolution problems in 2D signals, we applied a Swin transformer to extrapolate high-frequency information from downsampled neuronal signals both *in vitro* and *in vivo*. We found that aliasing components of input signals and the spike jittering-based selection of the training batch improved the performance of reconstructing accurate neuronal spikes. As a result, we observed reasonably restored neuronal spiking activity, including the spike timing, waveforms, and network connectivity, even with the $\times 8$ subsampled dataset.

1 INTRODUCTION

Multichannel neuronal signal recordings are the key to the brain-machine interfaces (BMIs) because the network analysis allows decoding motor intentions or functional connectivity of the brain [1]. Recent advancements in the multichannel recording hardware have particularly focused on increasing the number of simultaneous recording electrodes to achieve more data for detailed network analysis [2], [3]. For example, in the BMIs, the more electrode data we record, the wider variety of functions we can classify for precise associated operations. Simultaneously, there have been numerous efforts to implement untethered, wireless data transfer for efficient long-term implantation of the recording systems [4], [5]. Despite these efforts, higher numbers of electrodes require either large storage memory or induce higher power consumption in the recording and wireless communication hardware, resulting in significant heat dissipation, which must be avoided in the implantable BMIs.

To decrease the size of recording data to ease the aforementioned constraints, reduced data sampling approaches such as adaptive sampling [6], compressed sensing [7], on-chip spike detection [8], using spiking band power [9], and downgrading raw signal qualities [10] have been suggested. While these approaches have shown promising results in significant data compression, there are fundamental limitations for the applications in advanced BMIs. It has been well studied that broader bandwidths of neuronal signals, including low-frequency data such as local field potentials (LFPs), are strongly correlated with brain functions [11]. In that sense, aforementioned approaches lack acquiring the variety of signals. Previous algorithms are designed to only focus on the sparsity, abrupt changes in spikes. However, raw recording signals including LFPs are not sparse by nature, and are strongly affected by external sources such as signal drift or noise. Thus, the previous compression algorithms are either unsuitable for reconstructing the entire information of neuronal signals or vulnerable to the unavoidable changes in raw signals. Moreover, all these approaches require custom-designed recording hardware for on-chip signal pre-processing, which also limits the universal applicability to the state-of-the-art BMI technologies. Recently, deep learning techniques have been proposed for frequency extrapolation problems [12, 13], but the demonstrations are limited to seismic waveforms and microwave engineering, not to the spectrum of neuronal signals.

In this work, we introduce two methods of high-frequency extrapolation to restore neuronal spikes from downsampled low-frequency band signals both *in vitro* and *in vivo* (Fig. 1). Our models utilize a Swin transformer [15], [14]: *Ours-D* has an upsampling block within the network, whereas *Ours* has a pre-interpolation process instead of the upsampling block. We demonstrated our system on

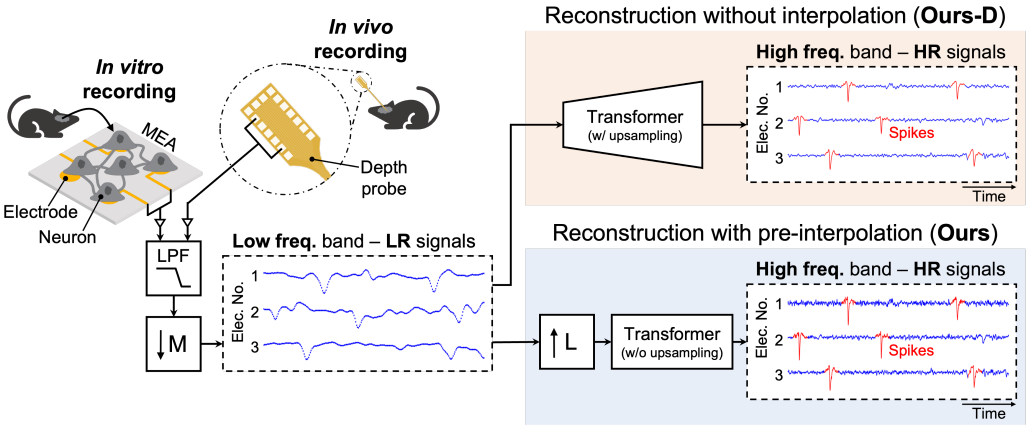

Figure 1: Overview of high-frequency extrapolation using transformer-based neural networks. The transformer used is SwinIR [14]. The low-pass filter used is a Butterworth filter that is a realistic and non-ideal one.

multielectrode recording data from *in vitro* hippocampal cultures and *in vivo* mouse brains, preprocessed by a low-pass filter and downsampled. We found that the aliasing component from the input, which inevitably occurs due to the non-ideal nature of the low-pass filter, is essential for the high frequency extrapolation. The accuracy of the extrapolation was further improved by selecting the jittered spike window as training batches. Our system is compatible with common neuronal signal acquisition hardwares.

Our main contributions are as follows:

- We demonstrated that our transformer-based method predicts accurate spikes from significantly downsampled neuronal signals.
- We showed that the restored multichannel spikes maintained the spatiotemporal information, including the spike timing, waveforms, and network connectivity.
- We found that high-frequency aliasing components of input signals are crucial to extrapolate the high frequencies of the original spikes. In addition, our spike jittering-based batch selection improves the reconstruction performances of the proposed Swin transformer.
- We showed that our pre-trained models could estimate accurate spikes on neural recordings not only *in vitro* but also *in vivo*, empirically implying the generality of our approach.

2 Problem Formulation

We formulate a reconstruction problem of a band-limited signal with sub-Nyquist samples. Let us assume that our input signal is convolved with an anti-aliasing low-pass filter, $h_{lpf}(t)$, of cut-off frequency f_c and then downsampled. f_c is set smaller than the Nyquist frequency $(1/2T_s)$ where T_s is a sampling period. This is written as follows:

$$\phi_{LPF}[n] = \sum_n \phi_{LPF}(t)\delta(t - nT_s) \tag{1}$$

$$\phi_{LPF}(t) = \mathcal{F}^{-1}\{\hat{\phi} \odot \hat{H}_{lpf}\}(t) = \frac{1}{2\pi}\int_{\mathbb{R}} \hat{\phi}(\omega)\hat{H}_{lpf}(\omega)\, \mathrm{e}^{\mathrm{j}\,\omega\, t}\, \mathrm{d}\omega. \tag{2}$$

where

$$\hat{H}_{lpf}(\omega) = \mathcal{F}(h_{lpf}) = \begin{cases} 1, & |\omega| < 2\pi f_c \\ < \epsilon, & |\omega| \geq 2\pi f_c \end{cases} \tag{3}$$

where $\mathcal{F}(\cdot)$ and $\mathcal{F}^{-1}(\cdot)$ denote the Fourier transform and inverse Fourier transform, respectively.

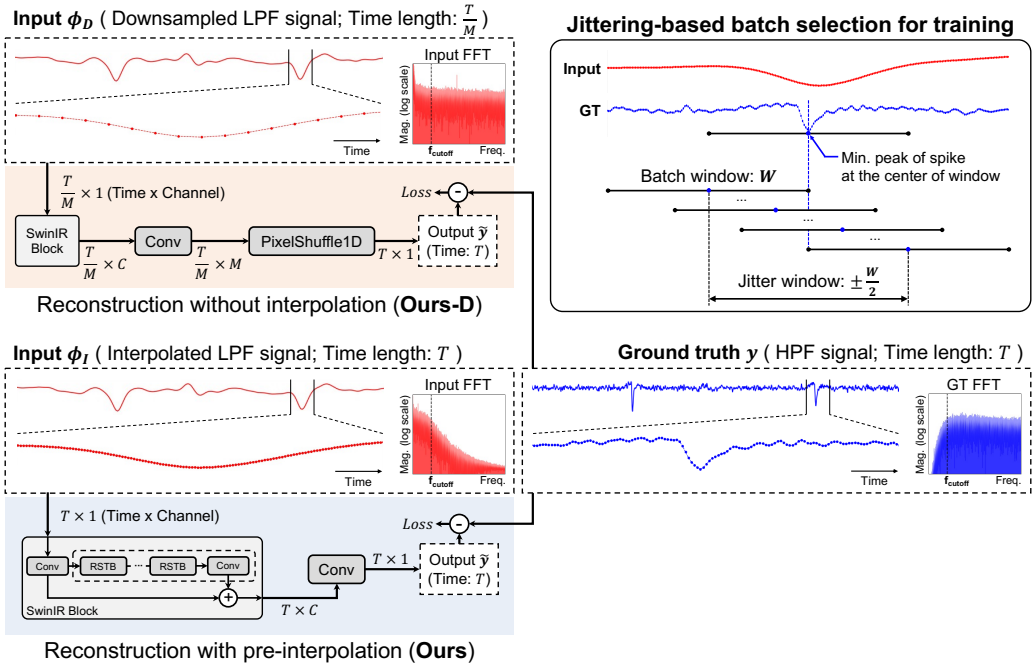

Figure 2: Neural network architectures and illustration of spike jittering-based batch selection.

We use two different inputs: $\phi_D[m]$ and $\phi_I[n]$. The input $\phi_D[m]$ is a downsampled signal of the low-pass filtered (LPF) signal $\phi_{LPF}[n]$, by a factor of M, which is represented by:

$$\phi_D[m] = \phi_{LPF}[mM]. \tag{4}$$

The other input $\phi_I[n]$ is an interpolated signal by re-upsampling $\phi_D[m]$ through the Fourier method by a factor of L equal to the downsampling factor M,

$$\hat{\phi}_I(\omega) = \frac{L}{2\pi M T_s} \sum_l \hat{\phi}\left(\omega - 2\pi l \frac{L}{M T_s}\right), \qquad \phi_I[n] = \frac{1}{2\,\pi} \int_{\mathbb{R}} \hat{\phi}_I(\omega)\, \mathrm{e}^{\mathrm{j}\,\omega\, n}\, \mathrm{d}\omega. \tag{5}$$

We aim to extrapolate frequencies above f_c with supervised training.

$$\theta^* = \arg\min_{\theta} \|f_\theta(\boldsymbol{\phi_D}) - \mathbf{y}\|_2, \quad \text{for } \textit{Ours-D} \tag{6}$$

$$\varphi^* = \arg\min_{\varphi} \|g_\varphi(\boldsymbol{\phi_I}) - \mathbf{y}\|_2, \quad \text{for } \textit{Ours} \tag{7}$$

where $\mathbf{y}$ is a ground truth (GT), high-pass filtered (HPF) signal having high-frequency components with a high temporal resolution. It is convolved with a high-pass filter, $h_{hpf}(t)$, of the same cut-off frequency f_c. The frequency response of the filter is given by:

$$\hat{H}_{hpf}(\omega) = \mathcal{F}(h_{hpf}) = \begin{cases} < \epsilon, & |\omega| < 2\pi f_c \\ 1, & |\omega| \geq 2\pi f_c \end{cases} \tag{8}$$

After terminating the supervised learning, we can obtain high-frequency and high-resolution signals from the trained network.

3 Methods

3.1 Network architecture

As shown in Fig. 2, both neural networks for signal reconstruction without (Eq. 6) and with pre-interpolation (Eq. 7) include a common SwinIR block [14], consisting of the multi-head self-attention-based Swin transformer layers [15]. The key differences between the networks are the input resolution and the presence of an upsampling block in the last part for signal reconstruction.

Common SwinIR block Given an input signal with a time length of T, a shallow feature of the same length T with C channels is extracted by convolving the input signal with 1D kernels ($k_{SF}(\cdot) : \mathbb{R}^{T\times 1} \rightarrow \mathbb{R}^{T\times C}$). By passing the feature through several consecutive residual Swin transformer blocks (RSTB) followed by an additional convolution layer, a deep feature with the same size as the input feature is obtained ($k_{DF}(\cdot) : \mathbb{R}^{T\times C} \rightarrow \mathbb{R}^{T\times C}$). The shallow and deep features are combined with a skip connection.

Reconstruction without interpolation (*Ours-D*) Input of *Ours-D* is a downsampled LPF signal by a factor of M, $\boldsymbol{\phi_D} \in \mathbb{R}^{\frac{T}{M}\times 1}$. Through the SwinIR block, the same length of the output feature with C channels is generated, and a 1D convolution layer adjusts the number of feature channels from C to M ($k_{CM}(\cdot) : \mathbb{R}^{\frac{T}{M}\times C} \rightarrow \mathbb{R}^{\frac{T}{M}\times M}$). The following upsampling block is implemented by efficient sub-pixel convolution [16]. By a 1D pixel shuffle operator, a one-channel output signal with the length of T, $\tilde{\mathbf{y}} = f_\theta(\boldsymbol{\phi_D})$, is obtained ($k_{PS}(\cdot) : \mathbb{R}^{\frac{T}{M}\times M} \rightarrow \mathbb{R}^{T\times 1}$).

$$\tilde{\mathbf{y}} = k_{PS}(k_{CM}(\mathbf{F_{SF}} + \mathbf{F_{DF}})) \tag{9}$$

where $\mathbf{F_{SF}} = k_{SF}(\boldsymbol{\phi_D}),\ \mathbf{F_{DF}} = k_{DF}(\mathbf{F_{SF}})$.

Reconstruction with pre-interpolation (*Ours*) Input of *Ours* is an interpolated LPF signal with the length of T. Because the length of the input signal is the same as that of the output signal $\tilde{\mathbf{y}} = g_\varphi(\boldsymbol{\phi_I})$, there is no layer for upsampling. Thus, the extracted feature from the SwinIR block is sent into a 1D convolution to generate a one-channel output ($k_C(\cdot) : \mathbb{R}^{T\times C} \rightarrow \mathbb{R}^{T\times 1}$).

$$\tilde{\mathbf{y}} = k_C(\mathbf{F_{SF}} + \mathbf{F_{DF}}) \tag{10}$$

where $\mathbf{F_{SF}} = k_{SF}(\boldsymbol{\phi_I}),\ \mathbf{F_{DF}} = k_{DF}(\mathbf{F_{SF}})$.

3.2 Frequency band separation of neuronal signals using realistic and ideal filters

The neuronal signal of each electrode is separated into two components of a pair by using realistic filters, which are n-th order Butterworth low-pass and high-pass filters, with the same cutoff frequency of f_c. The frequency responses of the filters is given by:

$$\hat{H}_{lpf}(\omega) = \frac{1}{\sqrt{1 + \left(\frac{\omega}{2\pi f_c}\right)^{2n}}}, \qquad \hat{H}_{hpf}(\omega) = \frac{1}{\sqrt{1 + \left(\frac{2\pi f_c}{\omega}\right)^{2n}}}. \tag{11}$$

For the ablation study, input and GT signals are perfectly segregated through ideal filters (a.k.a. brick wall filters), whose frequency responses are given by:

$$\hat{H}_{bw-lpf}(\omega) = \begin{cases} 1, & |\omega| < 2\pi f_c \\ 0, & |\omega| \geq 2\pi f_c \end{cases}, \qquad \hat{H}_{bw-hpf}(\omega) = \begin{cases} 0, & |\omega| < 2\pi f_c \\ 1, & |\omega| \geq 2\pi f_c \end{cases} \tag{12}$$

3.3 Spike jittering-based batch selection for training

A spike is a rapidly changing voltage that occurs within a short time (about a few msec) compared to the time duration in which only noise without spikes is present. For this reason, if a training batch is randomly selected, it is unlikely that the batch will contain spikes. This can be a factor that increases the inefficiency in learning the spike features. To achieve a more accurate reconstruction of spike information, we construct a minibatch with a batch size of B and a window size of W: at least one spike is included in the window for the first half of the batch, and their minimum peak is placed on a random position within the window by jittering the spike timing as follows. Let us assume that a time series data $\mathbf{y^k} \in \mathbb{R}^N$, which is high-frequency and high-resolution signal from electrode k, has s^k spikes ($k = 1, \ldots, K$). For the i-th spike of the electrode, n_i^k denotes the time point where the minimum peak of the spike waveform is located ($i = 1, \ldots, s_k$). To select windows of the first half batch, we choose $\frac{B}{2}$ peaks $n_{i_j}^{k_j}$, by picking out the electrodes and their corresponding spikes $\{k_j, i_j\}$ ($j = 1, \ldots, \frac{B}{2}$) and the equal number of jitters τ_j in the interval $(-\frac{W}{2}, \frac{W}{2}]$ at random. With the chosen variables, the data within the time interval $[n_{i_j}^{k_j} + \tau_j - \frac{W}{2}, n_{i_j}^{k_j} + \tau_j + \frac{W}{2})$ is sampled as the j-th batch window. The other windows in the second half batch are randomly sampled using the time series data of N time points from K electrodes.

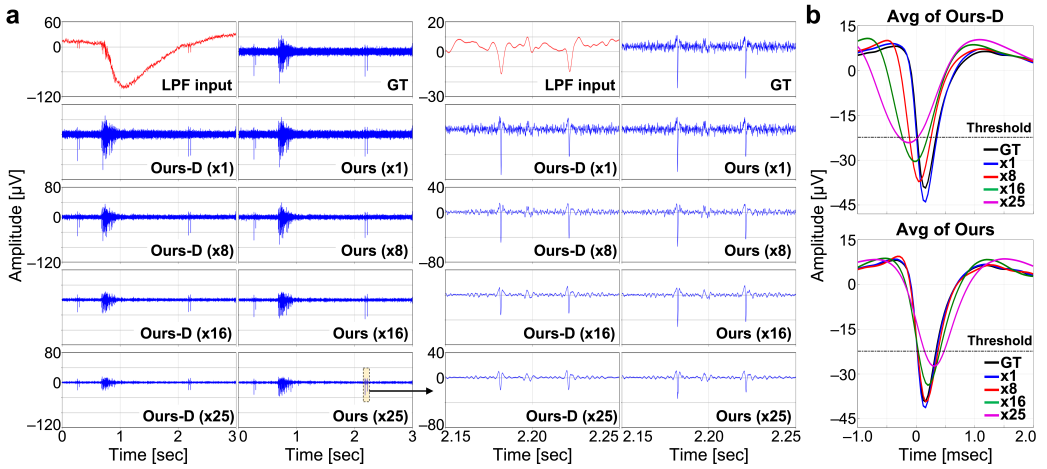

Figure 3: Representative reconstruction results of a electrode of MEA1. (a) Raw traces of LPF input, GT, and restored signals from *Ours-D* and *Ours* models with different scale factors. Except for the LPF input signals, all others are plotted using the same scale. (b) Average waveforms of correctly rebuilt spikes aligned to the GT timestamps.

4 RESULTS

We used two different neuronal datasets: *in vitro* and *in vivo*. The *in vitro* datasets were acquired from independent neuronal cultures on two microelectrode arrays (MEA1 and MEA2). Signals recorded from 100 electrodes of the MEA1 were utilized for training, and those from the other 13 electrodes of the MEA1 and 16 electrodes of the MEA2 were applied for evaluation. The *in vivo* datasets were obtained in two brain regions (Cortex and hippocampus) of two different mice. Signals from 12 electrodes in the cortex were used for training, and those from 4 electrodes in both regions were employed for evaluation. The details of dataset acquisition are provided in Sec A.1.

4.1 NEURONAL DATASET PROCESSING

We separated the neuronal data, measured with a sampling frequency of 25 kHz, into LPF input and HPF GT signals using zero-phase fourth-order Butterworth filters with a cutoff frequency of 200 Hz. The data pair were normalized to the maximum absolute value of the background noise of the HPF signals. To obtain downsampled and interpolated inputs, the LPF signals were subsampled by factors of 1, 8, 16, and 25 and then re-upsampled using the Fourier method by the same factors. Neuronal spikes were detected by setting the threshold at -6 standard deviation of the noise level of the GT signal with a detection dead time of 3 ms. Among the reconstructed spikes, the timestamps within $\pm$ 500 μs of GT timestamps were considered to be correctly restored outcomes (true positive, TP).

4.2 IMPLEMENTATION DETAILS

For the SwinIR model, we set the number of RSTB, STL, feature channels, and a kernel size of 1D convolution to 6, 6, 180, and 3, which are the same as the previous study [14]. In the training phase, input sequences of 128 data points were randomly sampled from the downsampled LPF signals with the corresponding GT sequences of $128r$ data points, where r is the upscaling factor. Different networks were trained in individual scale factors ($r = \times 1$, $\times 8$, $\times 16$, and $\times 25$) for 200 epochs with a batch size of 16, and a mean squared error loss and an Adam optimizer with a fixed learning rate of 1e-4 were used for optimization. In the evaluation, 128 data points were sequentially presented to the network by sliding the window by 64 data points. As baseline models for comparison, we used CNN-based models, TCN [17] and EDSR-Baseline [18]. To make the number of parameters similar to our models, hyperparameters of the baseline models were set as described in Sec A.1.

Table 1: Hit rate of spike detection (TP / (TP + FN)) and normalized root mean square error (NRMSE) between the restored and actual waveform on test datasets. The best results are in red.

Trainset: ***in vitro* MEA1** (n = 100)		Hit rate				NRMSE			
Testset	Method	×1	×8	×16	×25	×1	×8	×16	×25
MEA1 (n = 13)	TCN [17]	0.58	0.58	0.58	0.39	0.16	0.15	0.16	0.18
	EDSR [18]	0.95	0.64	0.34	0.16	0.04	0.15	0.21	0.24
	Ours-D	0.97	0.67	0.40	0.15	0.05	0.15	0.21	0.24
	Ours	0.99	0.78	0.65	0.44	0.02	0.07	0.12	0.16
MEA2 (n = 16)	TCN [17]	0.76	0.75	0.72	0.46	0.15	0.14	0.15	0.18
	EDSR [18]	0.97	0.82	0.40	0.09	0.03	0.15	0.21	0.24
	Ours-D	0.99	0.84	0.51	0.08	0.04	0.15	0.21	0.24
	Ours	1.00	0.91	0.80	0.51	0.02	0.06	0.11	0.16

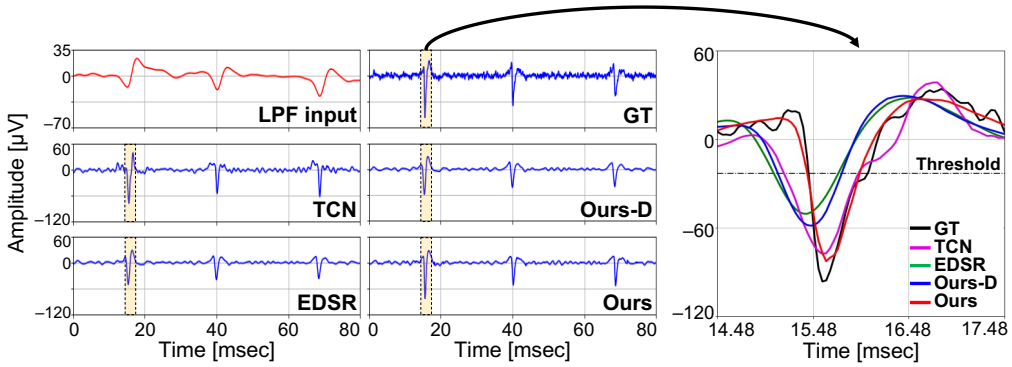

Figure 4: Restored signals of MEA2 in ×16 task (left) and enlarged spike waveforms of the time window highlighted in the raw traces (right).

4.3 Evaluation

Signal reconstruction with different scale factors We restored high-frequency and high-resolution signals from LPF input signals with different scale factors. As shown in the representative raw traces of a single electrode (Fig. 3a), the voltage fluctuations, especially spiking events, were successfully reconstructed in both time windows of burst behavior and tonic firing while the overall signal amplitudes tend to decrease as the scale factor increased. Figure 3b present the average waveforms of the correctly rebuilt spikes for each scale factor. The waveforms of *Ours-D* are shifted toward a negative direction in time, and the degree of the shift is proportional to the scale factor. In the case of the mean waveforms in *Ours*, there are no time delays of spike timestamps up to ×16. The mean time delays of the electrodes for scales factors ×1, ×8, and ×16 are 0 μs, 6.15 $\pm$ 15.02 μs, and 16.67 $\pm$ 26.74 μs (mean $\pm$ standard deviation, n = 13 electrodes), which are smaller than the sampling period (40 μs) of the high-resolution signal.

Validation on an unseen dataset and comparison with CNN-based methods Next, we tested our models on the MEA2 dataset, which was not used for network training, and compared the performance against other CNN-based methods, TCN [17] and EDSR-Baseline [18]. Table 1 provides the quantitative comparisons between TCN, EDSR, and our models on two MEA datasets. For the hit rate of detection and normalized root mean square error (NRMSE) of waveforms (windows from -1 ms to 2 ms of GT timestamps), *Ours* achieves the best performance on both two MEA datasets in all scale factors. Figure 4 shows output signals with the same scale factor of 16 for each model. Among the models, *Ours* restores the most accurate spikes.

Analysis of restored spikes: functional connectivity and spike sorting Using the reconstructed multichannel spikes, we first assessed how the spatiotemporal information for network connectivity

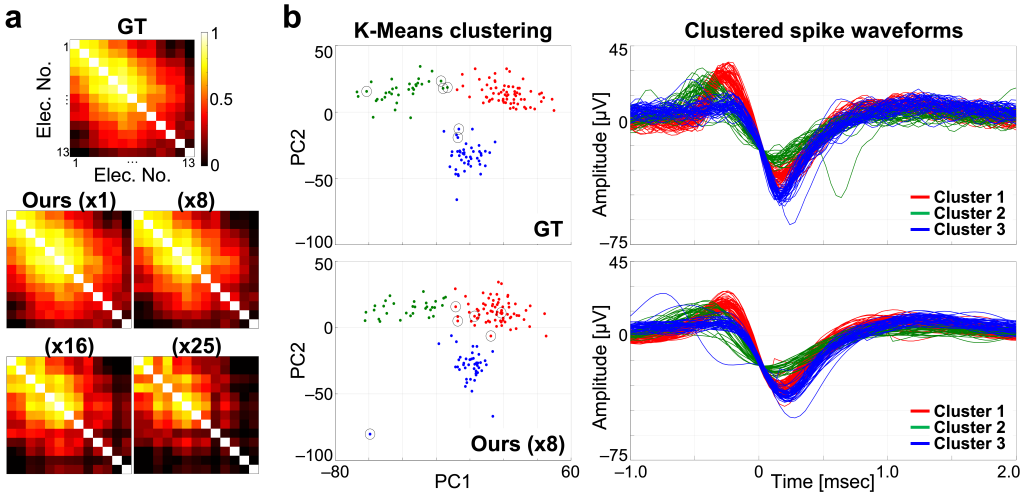

Figure 5: Neuronal spike train analysis. (a) Correlation coefficient matrices between spike trains from multiple electrodes. (b) Spike sorting by K-Means clustering (3-clusters case). Black circles in the principal component space (PC1 vs. PC2) show incorrectly classified spikes (6/150 spikes). Clustering accuracy of the reconstructed spikes: 96%.

is restored. Figure 5a shows correlation matrices for spike trains from multiple electrodes with different scale factors. Each element represents a Pearson correlation coefficient between a pair of rate histograms (50 ms-bin width), and a higher value means a stronger network connection. Despite some missing spikes and the shifted waveforms, the spatial network connectivity was reasonably well reestablished for $\times 8$ in *Ours*. Next, we performed spike sorting to compare how the spike waveform features are reconstructed. Figure 5b presents clustering results of the GT and rebuilt spikes (*Ours*, $\times 8$) from a single electrode that has three clusters. By the K-Means clustering, the spikes were sorted into different clusters, as visualized with different colors in the principal component (PC) space and waveform plots. It shows high clustering accuracy (96%) on the restored spike waveforms, with only a few spikes incorrectly grouped that were indicated as black circles in PC space (6/150 spikes).

Application to *in vivo* datasets Finally, we employed our models in *in vivo* datasets collected from mice, recorded in the auditory cortex (Ctx) and the hippocampus (Hippo). First, we applied pre-trained networks on the *in vitro* trainset to *in vivo* testsets (Trainset: *in vitro* MEA1 in Table 2). The reconstruction performance on the Ctx testset was similar to the results on *in vitro* testsets that were presented in Table 1, whereas the hit rate of the Hippo testset was much lower than those of *in vitro* testsets. This degradation seems to be caused by a significant error, especially in the time window where the LPF input signal fluctuated greatly ($3.5 - 4.0$ sec of Hippo in Fig. 6). Next, we trained our models on *in vivo* Ctx trainset and evaluated it on the other Ctx and Hippo testsets (Trainset: *in vivo* Ctx in Table 2). In this case, it was possible to achieve a slight advance for the Ctx testset (from 0.87 to 0.97) and a large improved restoration for the Hippo testset (from 0.55 to 0.81), particularly with the scale factor of 8.

4.4 Ablation Study

Aliasing components of input signals We separated neuronal data into low and high-frequency band signals using a Butterworth filter, a realistic and non-ideal filter, as described in Section 3.2; this inevitably made a frequency overlap of the two signals. To ablate this effect, we used a dataset of input and GT signals, whose frequency bands are completely split through ideal filtering (IF, Eq. 12). As shown in Fig. 7, the amplitude of signals in the IF case with spike jittering-based training (+JT, Section 3.3) is much smaller than *Ours* (+JT) (77.67% decline), with a large time delay. In the quantitative comparison in Table 3, the results of the ideal filter (IF (+JT)) show an extensive reduction of the hit rate and greater error of the signals for all scale factors.

Table 2: Quantitative results on *in vivo* datasets using two different trained networks (Trained on *in vitro* MEA1 (n = 100) or *in vivo* Ctx (n = 12)). The best and the second-best results are in red and blue, respectively.

			Hit rate		NRMSE	
Testset	Trainset	Method	×1	×8	×1	×8
Ctx (n = 4)	*in vitro* MEA1	Ours-D	0.90	0.71	0.06	0.15
		Ours	0.98	0.87	0.04	0.07
	in vivo Ctx	Ours-D	0.96	0.84	0.04	0.15
		Ours	0.95	0.97	0.04	0.07
Hippo (n = 4)	*in vitro* MEA1	Ours-D	0.82	0.38	0.09	0.16
		Ours	0.89	0.55	0.06	0.10
	in vivo Ctx	Ours-D	0.86	0.49	0.09	0.16
		Ours	0.86	0.81	0.05	0.09

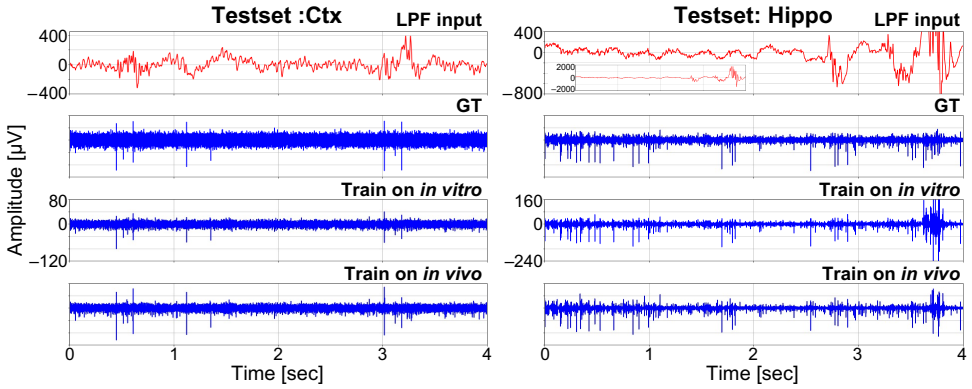

Figure 6: Representative restoration signals of *in vivo* datasets using the pre-trained networks on *in vitro* MEA1 or *in vivo* Ctx. *Ours* model with the scale factor of 8 was used for the training. Except for the LPF input signals, all other signals are plotted using the same scale.

Spike jittering-based window selection for training batches We set the minibatch for the training so that half of them include at least one spike within their batch window, as described in Section 3.3. To remove this effect, we chose the batch by randomly picking out the windows across the entire time series (-JT in Fig. 7). In our model with the non-ideal filter (*Ours* (-JT)), there is a slight amplitude decrease (25.97% decrease from *Ours* (+JT) case), thereby reducing the performance for all scale factors (Table 3). Moreover, the training without the jittering method on the dataset produced by ideal filtering (IF (-JT)) causes a dramatic failure to recover signals. Taking together, we concluded that the overlapping frequency components highly enhance signal reconstruction performance, and the spike jittering-based training improves the restoration capability of accurate spike waveform.

5 Discussion

Two remaining issues need to be addressed in further studies. The first one is the restoration failure at higher scale factors. Here, we set the measurement sampling frequency of 25 kHz and the cutoff frequency of 200 Hz to obtain the low-frequency input signals. Since the Butterworth filter is a non-ideal anti-aliasing filter, it is assumed that any given subsampling causes aliasing. In this sense, we can say that we always encounter aliasing subsampled signals even beyond the Nyquist sampling rate. As shown in the result of ×60 in Fig. 8, however, there were only too tiny signal fluctuations to be detected as spikes. The other issue is computation time for real-time applicability. Table 4 shows the computation time for each model on NVIDIA RTX 3090 24 GB with a scale factor of

Table 3: Quantitative results of the ablation study. The best results are in red.

	Hit rate				NRMSE			
	×1	×8	×16	×25	×1	×8	×16	×25
Ours (+JT)	0.99	0.78	0.65	0.44	0.02	0.07	0.12	0.16
Ours (-JT)	0.73	0.68	0.46	0.37	0.11	0.10	0.14	0.17
IF (+JT)	0.08	0.04	0.05	0.02	0.23	0.23	0.23	0.23
IF (-JT)	0.00	0.00	0.00	0.00	0.24	0.24	0.24	0.25

Table 4: The number of parameters (in M), memory consumption (in MB), and computation time (in ms) for ×16.

Method	# Param.	Mem.	Time
TCN [17]	10.14	101.51	7.76
EDSR [18]	10.11	125.57	8.29
Ours-D	10.14	136.58	60.54
Ours	10.13	136.55	59.64

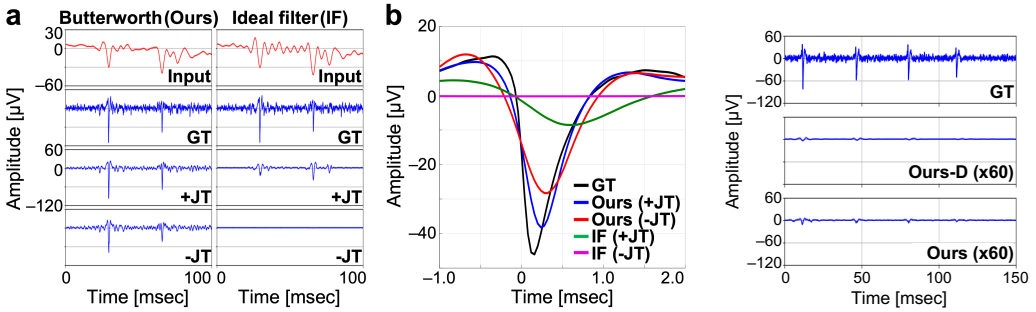

Figure 7: Qualitative ablation study with *Ours* model in ×16 task. (a) Raw traces. (b) Average signal profiles.

Figure 8: Reconstruction failure cases (×60).

16. Although *Ours* showed the best accuracy, there was a long time delay (59.64 ms) in extracting the output signal from the interpolated input (time length of the output signal: 5.12 ms). On the other hand, *Ours-D* required less computation time (60.54 ms) than the time length of the output signal (81.92 ms). Further study on optimizing a pass filter and model structure would be required to improve the restoration capability even with higher scale factors and to resolve the trade-off between accuracy and computation time.

Despite these issues, our proposed approach provides crucial advantages. First of all, our method enables the acquisition of neuronal signals with high-frequency bandwidths through low-frequency and small-volume data recording. The restored signals preserve precise spiking activity and network connectivity for comprehensive and in-depth analysis. Moreover, our models can also work in *in vivo* recording data, typically noisier environment. It showed good feasibility even for applying the trained networks from *in vitro* to *in vivo* datasets and from one region to another regions of the brain (Ctx to Hippo). In addition, our approach has a high universality in that it employs conventional downsampling and interpolation methods for data processing and utilizes both signal regions with and without spikes, compared to previous studies that dealt with only sampled spike windows [19], [20], [21].

6 Conclusion

In this work, we presented a method for restoring high-frequency band multichannel neuronal signals both *in vitro* and *in vivo* with high resolution by recording low-frequency downsampled signals. Based on the Swin transformer, we demonstrated that our proposed model outperforms other CNN-based methods for all scale factors, and well reconstructs the spiking activity and connectivity information of neuronal networks. We also confirmed that higher restoration performance can be achieved by the aliasing components of input and output signals and spike-focused batch selection. We believe that our proposed framework opens a new direction in obtaining high-quality neural information for neuroscience applications while reducing the data size being handled.

7 REPRODUCIBILITY AND ETHICS STATEMENT

A complete description of data acquisition and processing are provided in Sec A.1 and Sec 4.1 for reproducibility. All experiment procedures were approved by Institutional Animal Care and Use Committee (IACUC), and all experiments were preformed in accordance with the guidance of the IACUC.

REFERENCES

[1] J. P. Dmochowski, A. Datta, M. Bikson, Y. Su, and L. C. Parra, "Optimized multi-electrode stimulation increases focality and intensity at target," *Journal of Neural Engineering*, vol. 8, no. 4, p. 046011, 2011.

[2] M. E. J. Obien, K. Deligkaris, T. Bullmann, D. J. Bakkum, and U. Frey, "Revealing neuronal function through microelectrode array recordings," *Frontiers in Neuroscience*, vol. 8, p. 423, 2015.

[3] N. A. Steinmetz, C. Aydin, A. Lebedeva, M. Okun, M. Pachitariu, M. Bauza, M. Beau, J. Bhagat, C. Böhm, M. Broux, *et al.*, "Neuropixels 2.0: A miniaturized high-density probe for stable, long-term brain recordings," *Science*, vol. 372, no. 6539, p. eabf4588, 2021.

[4] B. Lee, Y. Jia, S. A. Mirbozorgi, M. Connolly, X. Tong, Z. Zeng, B. Mahmoudi, and M. Ghovanloo, "An inductively-powered wireless neural recording and stimulation system for freely-behaving animals," *IEEE Transactions on Biomedical Circuits and Systems*, vol. 13, no. 2, pp. 413–424, 2019.

[5] M. M. Ghanbari, D. K. Piech, K. Shen, S. F. Alamouti, C. Yalcin, B. C. Johnson, J. M. Carmena, M. M. Maharbiz, and R. Muller, "17.5 A 0.8 mm 3 ultrasonic implantable wireless neural recording system with linear am backscattering," in *2019 IEEE International Solid-State Circuits Conference-(ISSCC)*, pp. 284–286, IEEE, 2019.

[6] L. Mesin, "A neural algorithm for the non-uniform and adaptive sampling of biomedical data," *Computers in Biology and Medicine*, vol. 71, pp. 223–230, 2016.

[7] B. Sun and W. Zhao, "Compressed sensing of extracellular neurophysiology signals: A review," *Frontiers in Neuroscience*, p. 948, 2021.

[8] D.-Y. Yoon, S. Pinto, S. Chung, P. Merolla, T.-W. Koh, and D. Seo, "A 1024-channel simultaneous recording neural SoC with stimulation and real-time spike detection," in *2021 Symposium on VLSI Circuits*, pp. 1–2, IEEE, 2021.

[9] S. R. Nason, A. K. Vaskov, M. S. Willsey, E. J. Welle, H. An, P. P. Vu, A. J. Bullard, C. S. Nu, J. C. Kao, K. V. Shenoy, *et al.*, "A low-power band of neuronal spiking activity dominated by local single units improves the performance of brain–machine interfaces," *Nature Biomedical Engineering*, vol. 4, no. 10, pp. 973–983, 2020.

[10] N. Even-Chen, D. G. Muratore, S. D. Stavisky, L. R. Hochberg, J. M. Henderson, B. Murmann, and K. V. Shenoy, "Power-saving design opportunities for wireless intracortical brain–computer interfaces," *Nature Biomedical Engineering*, vol. 4, no. 10, pp. 984–996, 2020.

[11] G. Buzsáki, C. A. Anastassiou, and C. Koch, "The origin of extracellular fields and currents—EEG, ECoG, LFP and spikes," *Nature Reviews Neuroscience*, vol. 13, no. 6, pp. 407–420, 2012.

[12] O. W. Bhatti, N. Ambasana, and M. Swaminathan, "Design space and frequency extrapolation: Using neural networks," *IEEE Microwave Magazine*, vol. 22, no. 10, pp. 22–36, 2021.

[13] H. Sun and L. Demanet, "Low frequency extrapolation with deep learning," in *SEG Technical Program Expanded Abstracts 2018*, pp. 2011–2015, Society of Exploration Geophysicists, 2018.

[14] J. Liang, J. Cao, G. Sun, K. Zhang, L. Van Gool, and R. Timofte, "SwinIR: Image restoration using swin transformer," in *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops*, pp. 1833–1844, 2021.

[15] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo, "Swin transformer: Hierarchical vision transformer using shifted windows," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 10012–10022, 2021.

[16] W. Shi, J. Caballero, F. Huszár, J. Totz, A. P. Aitken, R. Bishop, D. Rueckert, and Z. Wang, "Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1874–1883, 2016.

[17] S. Bai, J. Z. Kolter, and V. Koltun, "An empirical evaluation of generic convolutional and recurrent networks for sequence modeling," *arXiv preprint arXiv:1803.01271*, 2018.

[18] B. Lim, S. Son, H. Kim, S. Nah, and K. Mu Lee, "Enhanced deep residual networks for single image super-resolution," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 136–144, 2017.

[19] B. Sun, H. Feng, K. Chen, and X. Zhu, "A deep learning framework of quantized compressed sensing for wireless neural recording," *IEEE Access*, vol. 4, pp. 5169–5178, 2016.

[20] T. Xiong, J. Zhang, C. Martinez-Rubio, C. S. Thakur, E. N. Eskandar, S. P. Chin, R. Etienne-Cummings, and T. D. Tran, "An unsupervised compressed sensing algorithm for multi-channel neural recording and spike sorting," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 26, no. 6, pp. 1121–1130, 2018.

[21] B. Sun, C. Mu, Z. Wu, and X. Zhu, "Training-free deep generative networks for compressed sensing of neural action potentials," *IEEE Transactions on Neural Networks and Learning Systems*, 2021.

A APPENDIX

A.1 EXPERIMENTAL DETAILS

Neuronal dataset acquisition We acquired biological neuronal signals from *in vitro* neuronal cultures using a microelectrode array (MEA). Two MEAs used as a culture substrate and a read-out interface have 120 microelectrodes (MEA1: 120MEA100/30iR-ITO, MEA2: 120MEA200/30iR-Ti; Multi Channel Systems). Before cell cultivation, their surface was coated with 0.05 mg/mL of poly-D-lysine (A3890401, Gibco) to make them cell-adhesive. Rat hippocampi isolated from 18-day-old Sprague Dawley rats were dissociated, and the cell pellet was obtained by centrifugation. Following resuspending the pellet in a medium, composed of Neurobasal medium (21103049, Gibco), B-27 supplement (17504044, Gibco), GlutaMAX supplement (35050061, Gibco), penicillin-streptomycin (15140122, Gibco), and L-glutamine (25030081, Gibco), the neurons were seeded on the MEA with the density of 1000 cells/mm^2. Two weeks after the cultivation, spontaneous activities of cultured neurons were measured from multiple electrodes and sampled at 25 kHz by a DAQ card (Hardware filter: DC $-$ 10 kHz, Firmware filter: 0.1 Hz $-$ 3.5 kHz; MEA2100-Mini-Systems, Multi Channel Systems).

We also collected *in vivo* neuronal datasets from two C57BL/6J mice using 16-channels neural probes (A1x16-3mm-50-703, NeuroNexus). Mice were born and reared in standard mouse cages ($16 \times 36 \times 12.5$ cm^3) with food and water available ad libitum, and weaned at $3 - 4$ weeks of age and housed together with sex-matched siblings with up to four animals per cage. Mice were maintained at a 12:12-h light/dark cycle at 22 ± 1 °C. Surgery was aseptically carried out at $11 - 12$ weeks. First, mice were anesthetized through intraperitoneal injection of urethane (1.5g/kg) and placed in a stereotaxic apparatus (RWD Life Science, Shenzhen, China) for acute recording. Then, neural probes were implanted in the auditory cortex (AP -3.0 mm ML $+3.83$ mm DV -2.5 mm) or hippocampus (AP -1.6 mm ML $+1.6$ mm DV -1.7 mm). Reference and ground wires were inserted into the cerebellum. Before the recording in the hippocampus, kainic acid (10 mg/kg, K0250-10MG, Sigma Aldrich) was treated for induction of seizures through intraperitoneal injection. Using a DAQ system (M4016, M4200, Intan Technologies), signals were recorded at 25 kHz (Hardware filter: 0.98 Hz $-$ 7.60 kHz, Software notch filter: 60 Hz).

Specifications of baseline models As baseline models for comparison, we used CNN-based models, TCN [17] and EDSR-Baseline [18]. For the TCN model, the kernel size and the number of stacked blocks were set to 3 and 6. The input of TCN was interpolated low-pass filtered (LPF) signal same as *Ours*, and the length was 127, which is the same with the receptive field. Because the output length of the TCN is a single data point, the evaluation was performed by moving the input window point by point. In the case of the EDSR model, the number of residual blocks was 16 with a kernel size of 3. The input for EDSR was downsampled LPF signal, and its length and sliding window were set to 128 and 64, identical to *Ours-D*. In order to make the number of parameters similar to that of our models, the numbers of channels are determined to be 554 and 262 for TCN and EDSR, respectively.

A.2 ADDITIONAL RESULTS

Table A.1: Precision of spike detection on test datasets (TP / (TP + FP)). The values are high in most conditions, implying that most detected spikes are correct.

Trainset: MEA1 (n = 100)		Precision (mean ± SD)			
Testset	Method	$\times 1$	$\times 8$	$\times 16$	$\times 25$
MEA1 (n = 13)	TCN [17]	0.91 ± 0.07	0.91 ± 0.07	0.88 ± 0.06	0.91 ± 0.08
	EDSR [18]	0.98 ± 0.01	0.96 ± 0.04	0.94 ± 0.07	0.82 ± 0.15
	Ours-D	0.85 ± 0.07	0.96 ± 0.04	0.94 ± 0.07	0.77 ± 0.15
	Ours	0.93 ± 0.04	0.95 ± 0.03	0.89 ± 0.07	0.91 ± 0.08
MEA2 (n = 16)	TCN [17]	0.95 ± 0.05	0.94 ± 0.06	0.91 ± 0.06	0.94 ± 0.06
	EDSR [18]	0.99 ± 0.01	0.98 ± 0.02	0.97 ± 0.04	0.91 ± 0.09
	Ours-D	0.90 ± 0.07	0.98 ± 0.02	0.98 ± 0.03	0.85 ± 0.12
	Ours	0.96 ± 0.03	0.97 ± 0.02	0.92 ± 0.07	0.92 ± 0.10

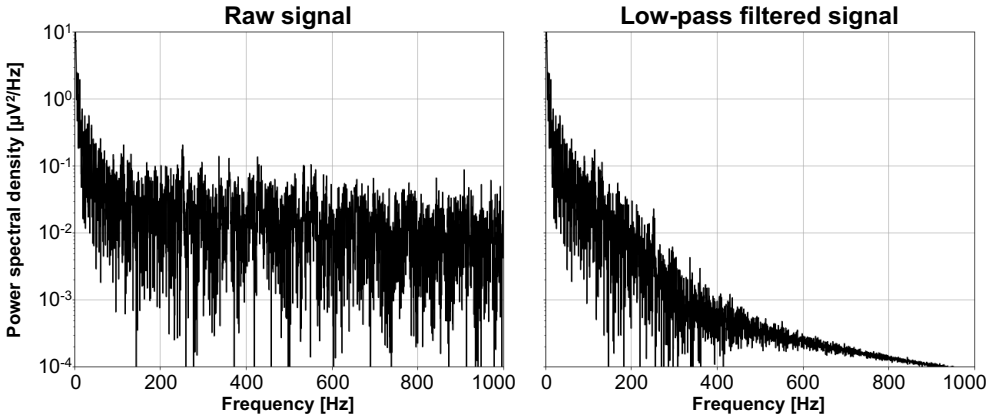

Figure A.1: Power spectral density of raw and low-pass filtered signals.

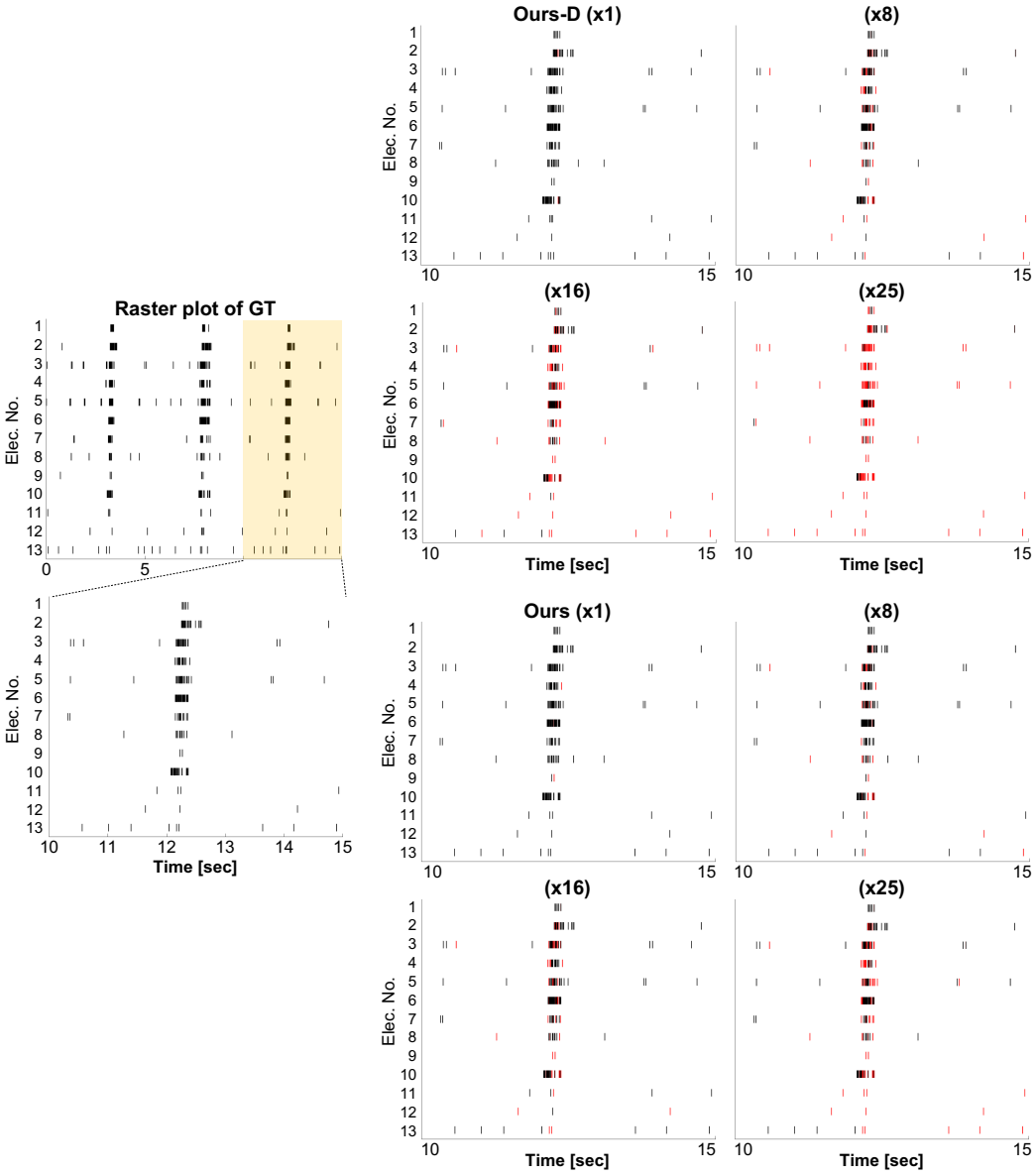

Figure A.2: Raster plots of neuronal spiking activity in $\times 1$, $\times 8$, $\times 16$ and $\times 25$ tasks. Red timestamps represent missing spikes (false negative, FN). Consistent with the reduction in spike amplitude and hit rate, the number of missing spikes increases as the factor rises.

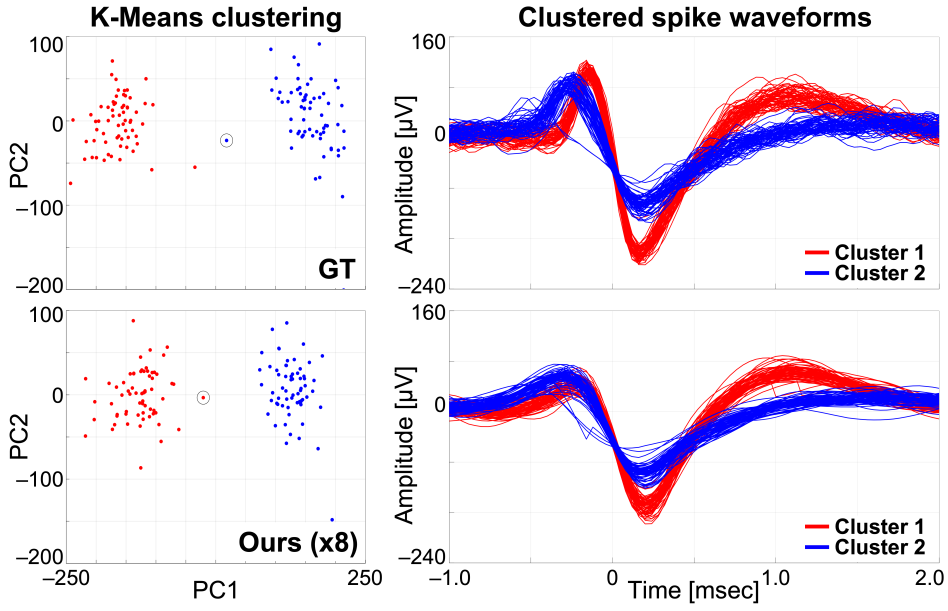

Figure A.3: Spike sorting by K-Means clustering (2-clusters case). A black circle in the principal component space (PC1 vs. PC2) shows an incorrectly classified spike (1/130 spikes). Clustering accuracy of the reconstructed spikes: 99.23%.