



# Compressed Sensing of Extracellular Neurophysiology Signals: A Review

Biao Sun<sup>1</sup> and Wenfeng Zhao<sup>2\*</sup>

<sup>1</sup> School of Electrical and Information Engineering, Tianjin University, Tianjin, China, <sup>2</sup> Department of Electrical and Computer Engineering, Binghamton University, State University of New York, Binghamton, NY, United States

This article presents a comprehensive survey of literature on the compressed sensing (CS) of neurophysiology signals. CS is a promising technique to achieve high-fidelity, low-rate, and hardware-efficient neural signal compression tasks for wireless streaming of massively parallel neural recording channels in next-generation neural interface technologies. The main objective is to provide a timely retrospective on applying the CS theory to the extracellular brain signals in the past decade. We will present a comprehensive review on the CS-based neural recording system architecture, the CS encoder hardware exploration and implementation, the sparse representation of neural signals, and the signal reconstruction algorithms. Deep learning-based CS methods are also discussed and compared with the traditional CS-based approaches. We will also extend our discussion to cover the technical challenges and prospects in this emerging field.

## OPEN ACCESS

### Edited by:

Esin Ozturk Isik,  
Boğaziçi University, Turkey

### Reviewed by:

Anil Kumar Sao,  
Indian Institute of Technology Mandi,  
India  
Sreeraman Rajan,  
Carleton University, Canada

### \*Correspondence:

Wenfeng Zhao  
wzhao@binghamton.edu

### Specialty section:

This article was submitted to  
Brain Imaging Methods,  
a section of the journal  
Frontiers in Neuroscience

Received: 17 March 2021

Accepted: 08 July 2021

Published: 26 August 2021

### Citation:

Sun B and Zhao W (2021)  
Compressed Sensing of Extracellular  
Neurophysiology Signals: A Review.  
Front. Neurosci. 15:682063.  
doi: 10.3389/fnins.2021.682063

**Keywords:** compressed sensing, electrophysiology, wireless neural recording, sparse representation (coding), sparse recovery

## 1. INTRODUCTION

Extracellular neural recording has been established for decades for monitoring the neuronal ensemble activities with better temporal resolution (Hubel, 1957; Buzsáki, 2004; Stevenson and Kording, 2011). Compared with other neural activity monitoring approaches, extracellular neural recording offers a broad recording spectrum of electrophysiology signals generated by the living brains, which spans from the slow-varying local field potentials (LFPs) to the transient spiking activities (action potentials [APs]) (Buzsáki et al., 2012). Both neural signal modalities carry essential brain processing information and are crucial to the understanding of brain functions. APs play a key role in neuron-to-neuron communication across the entire nervous system and have been widely studied for their functional representation in neural coding, while LFPs reflect the highly dynamic information flows beyond the reach of observing spiking activities from a few neurons, and have been studied for motor decoding in brain-machine interfaces (Andersen et al., 2004), sleep states (Vyazovskiy et al., 2011), sensory processing (Haslinger et al., 2006) as well as higher cognitive processes such as attention, memory, and perception.

To improve the yield of isolated neurons and the spatial coverage of field potentials from extracellular recordings, novel large-scale neural recording technologies with flexible, densely packed microelectrode tips (a few  $\mu\text{m}$  in diameter) are highly desired (Berényi et al., 2014; Hong and Lieber, 2019). Such needs motivate the engineering endeavors to create novel large-scale neurotechnologies to meet the scientific and clinical queries for investigating the brain-wide cortical dynamics. Recently, neurotechnologies have emerged with unprecedented recording densities toward cellular resolution, distributed brain regions, and depth structures (Jun et al., 2017; Allen et al., 2019; Musk and Neuralink, 2019; Stringer et al., 2019), which hold tremendous

potentials for neuroscientific discoveries. Neuronal population activities can therefore be recorded from diverse brain regions with a high temporal resolution, leading to a better understanding of the functionality of the neural circuits, the information flow, the network connections across different brain regions, and the relationship to the behaviors.

The steep recording density increase leads to the “*Big Neural Data Challenge*.” With tens of kilohertz sampling frequency and more than 10-bit analog-to-digital conversion resolution, several hundreds of megabits per second digitized neural data are generated in real time when incorporating the emerging high-density neural recording probes. The drastically increased neural data pose serious challenges for data processing, storage, and transmission tasks, all of which are of particular relevance to bandwidth-scarce and resource-constrained headstage designs. In general, cable electronics are used to stream an overwhelming amount of neural data online. The power consumption of the data link technologies (Lopez et al., 2017; Intan, 2021) can be non-trivial, and the tethered wire configuration could significantly restrain the envisioned frontier neuroscience research paradigms in many realistic and naturalistic settings. A parallel effort in the community is to augment wireless acquisition capabilities into the neurophysiology experiments, which is of great interest to neuroscientists and brain-machine interface technologists, as chronic brain recording of behaving animals could be achieved in an untethered fashion (Schwarz et al., 2014; Yin et al., 2014; Capogrosso et al., 2016; Zhou et al., 2019; Testard et al., 2021). Novel experimental paradigms can be defined to test hypotheses on the brain functions associated with space coding, foraging, social interaction, and cognitive behaviors for both small (e.g., rodents), large (e.g., swine, non-human primates), and flying animal (e.g., bats) subjects. Bi-directional wireless neural interface devices will also be crucial to the rehabilitation and prostheses for human subjects. Nevertheless, the low data bandwidth and high energy consumption make it challenging to employ the wireless communication components during an electrophysiology study (Larson and Nurmikko, 2016; Nurmikko, 2020).

To ease the challenge, high-fidelity on-chip and on-device neural signal compression schemes (Chae et al., 2008; Gagnon-Turcotte et al., 2016; Wu et al., 2017, 2018; Xu et al., 2018) become essential to relax the bandwidth and energy constraints by reducing the amount of data to be wirelessly transmitted at the system level. For the scope of neural signal compression, several promising approaches have been proposed in the past decades, such as on-chip spike detection and sorting (Lewicki, 1998; Gibson et al., 2011), sparse coding (Kamboh et al., 2007; Gagnon-Turcotte et al., 2016), feature extraction (Wu et al., 2017), and adaptive quantization (Martinez et al., 2018). Moreover, the on-chip hardware overhead for data compression and excessive power consumption cannot be neglected.

Compressed sensing (CS, or compressive sensing) is an emerging signal processing technique for sub-Nyquist sampling and the reconstruction of sparse signals (Donoho, 2006; Candes, 2008). The signal acquisition process of CS is achieved through random/incoherent sampling, while the signal recovery can be achieved using algorithms like convex relaxation or greedy

algorithms (Tropp and Gilbert, 2007; van den Berg and Friedlander, 2007; Grant et al., 2008; Zhang, 2009). By its notion, CS performs signal acquisition and compression simultaneously. CS aims to break the Nyquist–Shannon sampling limits, stating that the minimal sampling rate should be at least twice of the signal bandwidth (Oppenheim, 1999). Built upon the assumption of signal sparsity, CS can lead to a much-lowered sampling frequency compared to the Nyquist rate. Since its inception, CS has been widely investigated in many application domains that are sampling speed limited, such as high-speed analog-to-digital converter (ADCs), radio-frequency receivers (Chen et al., 2010; Mishali and Eldar, 2011; Yoo et al., 2012), and magnetic-resonance-imaging (MRI) (Lustig et al., 2008). For emerging wearable and implantable biomedical signal processing applications with limited hardware, power, and data bandwidth, CS has also been demonstrated beneficial for data compression tasks (Chen et al., 2012; Zhao et al., 2018), for its efficient, low-complexity encoder design, while the signal reconstruction can be accomplished offline.

In the past decade, CS has also been actively studied for neural signal compression tasks (Charbiwala et al., 2011; Schmale et al., 2013; Zhang et al., 2015; Sun et al., 2017; Zhao et al., 2018, 2019). CS has several encouraging features that suit neural recording applications. Nevertheless, there are many associated challenges as the neural signals have unique signal characteristics compared to other biomedical signal modalities. This paper aims to provide the audiences a comprehensive review of the past and the current status of CS-based neural recording systems. There are several reviews and surveys on the CS theories and their broad application domains (Craven et al., 2014; Jaspan et al., 2015; Gurve et al., 2020). Nevertheless, there is no review dedicated to the application of CS to the extracellular neural recording scenarios, which will be the primary focus of this paper.

The remainder of this paper is organized as follows. Section 2 covers the background of CS and summarizes the performance metrics that are used throughout the paper. Section 3 reviews the literature on the compressed sensing of the extracellular neural signals. Section 4 reviews deep learning (DL) based CS for neural signals. Section 5 discusses the future challenges and section 6 concludes the paper.

## 2. BACKGROUND

In this section, we first describe the fundamental theories of CS, including the sensing procedure, the sparsity priors, and the reconstruction algorithms. We then show typical CS-based neural recording system architectures. Finally, we present several quality evaluation criteria for neural signals.

### 2.1. Compressed Sensing

Compressed sensing is an emerging low-rate sampling scheme for the signals known to be sparse or compressible on some basis. The basic CS framework (Donoho, 2006), also called the single measurement vector (SMV) model, can be expressed as

$$\mathbf{y} = \Phi\mathbf{x} + \mathbf{e}, \quad (1)$$

where  $\mathbf{x} \in \mathbb{R}^n$  is a single-channel signal,  $\Phi \in \mathbb{R}^{m \times n}$  is the sensing matrix,  $\mathbf{e} \in \mathbb{R}^m$  is the measurement noise, and  $\mathbf{y} \in \mathbb{R}^m$  is the compressed measurement vector. Usually, Equation (1) is underdetermined, i.e.,  $m < n$ , and the ratio  $n/m$  is called the compression ratio (CR).

### 2.1.1. RIP

To ensure accuracy and robustness for signal recovery using the convex  $\ell_1$ -norm method, the sensing matrix  $\Phi$  should satisfy the restricted isometric property (RIP) (Candes, 2008) defined as follows:

$$(1 - \delta_k)\|\mathbf{x}\|_2 \leq \|\Phi\mathbf{x}\|_2 \leq (1 + \delta_k)\|\mathbf{x}\|_2, \quad (2)$$

where  $\delta_k$ , the isometry constant of  $\Phi$ , must be smaller than 1. The smaller the value of  $\delta_k$ , the higher the probability of an exact reconstruction.

### 2.1.2. Incoherent Sampling

In practice, it is difficult to verify the RIP property, while an alternative approach is to quantify the coherence between the sensing matrix and the sparse dictionary (Candes et al., 2011). The coherence  $\mu$  between the sensing matrix and the dictionary measures the correlation between any two columns of  $\Phi$  and  $\Psi$ . Ideally, the coherence will be small, i.e., the two matrices are incoherent, as the value for  $\mu$  is proportional to the number of the required measurements,

$$\mu(\Phi, \Psi) = \sqrt{N} \cdot \max_{1 \leq k, j \leq N} |\langle \Phi_k, \Psi_j \rangle|. \quad (3)$$

### 2.1.3. Synthesis Prior

Since the CS system (1) is underdetermined, the signal  $\mathbf{x}$  cannot be uniquely recovered from the sensing matrix  $\Phi$  and measurements  $\mathbf{y}$ . However, if  $\mathbf{x}$  has a *synthesis sparse prior* (SSP) (Bruckstein et al., 2009), i.e.,

$$\mathbf{x} = \Psi\boldsymbol{\theta}, \quad (4)$$

where  $\Psi \in \mathbb{R}^{n \times n}$  is a pre-defined dictionary, and the signal's representation  $\boldsymbol{\theta} \in \mathbb{R}^n$  is assumed to be  $s$ -sparse, i.e.,

$$\|\boldsymbol{\theta}\|_0 \triangleq |\text{supp}(\boldsymbol{\theta})| = s \ll n, \quad (5)$$

or is well approximated by an  $s$ -sparse vector, then it is possible to estimate  $\mathbf{x}$  via

$$\hat{\mathbf{x}} = \Psi\hat{\boldsymbol{\theta}} \quad \text{with} \quad \hat{\boldsymbol{\theta}} = \arg \min_{\boldsymbol{\theta}} \|\mathbf{y} - \Phi\Psi\boldsymbol{\theta}\|_2^2 + \lambda f(\boldsymbol{\theta}), \quad (6)$$

where  $f(\cdot)$  is a regularization function, and  $\lambda$  is a regularization parameter. A widely used penalty is the  $\ell_1$ -norm penalty, namely  $f(\boldsymbol{\theta}) = \|\boldsymbol{\theta}\|_1$  (Becker et al., 2011).

Motivated by many applications such as EEG/MEG source localization and DOA (direction of arrival) estimation, where multi-channel signals are measured simultaneously, the SMV model (1) has been extended to the multiple measurement vector (MMV) model in Cotter et al. (2005), given by

$$\mathbf{Y} = \Phi\mathbf{X} + \mathbf{E}, \quad (7)$$

where  $\mathbf{X} \in \mathbb{R}^{n \times l}$  is an  $l$ -channel signal,  $\Phi \in \mathbb{R}^{m \times n}$  is the sensing matrix,  $\mathbf{E} \in \mathbb{R}^{m \times l}$  is the measurement noise, and  $\mathbf{Y} \in \mathbb{R}^{m \times l}$  is the compressed data. An essential assumption in the MMV model is that the support (i.e., indexes of nonzero entries) of every column in  $\mathbf{X}$  is identical. Therefore,  $\mathbf{X}$  is simultaneous sparse (also referred to as row-sparse), i.e., a few rows of  $\mathbf{X}$  are nonzero rows. Similar to (6), the estimate of  $\mathbf{X}$  is given by

$$\hat{\mathbf{X}} = \Psi\hat{\boldsymbol{\Theta}} \quad \text{with} \quad \hat{\boldsymbol{\Theta}} = \arg \min_{\boldsymbol{\Theta}} \|\mathbf{Y} - \Phi\Psi\boldsymbol{\Theta}\|_F^2 + \lambda g(\boldsymbol{\Theta}), \quad (8)$$

where  $\|\cdot\|_F$  denotes the Frobenius norm ( $\ell_2$ -norm of all the elements) of the matrix and  $g(\cdot)$  is a regularization function encouraging the simultaneous sparsity. One popular penalty is the  $\ell_{2,1}$ -norm penalty (Ding et al., 2006), namely  $g(\boldsymbol{\Theta}) = \sum_{i=1}^n \|\boldsymbol{\Theta}_{i,\cdot}\|_2$ . In (8),  $\boldsymbol{\Theta}$  is assumed to be row-sparse.

### 2.1.4. Analysis Prior

While the signal reconstruction with SSP has been extensively studied, constructing an appropriate sparse dictionary remains challenging when using CS for neural signal compression. The neural signal segments are non-sparse on widely used dictionaries such as discrete Fourier transform (DFT) basis and discrete cosine transform (DCT) basis. Signal reconstruction using these dictionaries will severely reduce the accuracy. Moreover, the non-stationary behaviors of neural signals pose practical issues over the synthesis model-based methods, and the joint sparsity assumption is valid for only a minor portion of the MMV model. To overcome this limitation, an *analysis sparse prior* (ASP) that takes an analysis point of view has been proposed by Elad et al. (2007). For a signal of interest, ASP assumes that the analysis coefficient vector

$$\mathbf{z} = \boldsymbol{\Omega}\mathbf{x} \quad (9)$$

is expected to be sparse, where  $\boldsymbol{\Omega} \in \mathbb{R}^{d \times n}$  denotes a redundant analysis operator ( $d \geq n$ ), and  $\rho = d/n$  is the redundant ratio. Note that for an invertible square dictionary, SSP and ASP are the same with  $\Psi = \boldsymbol{\Omega}^{-1}$  (Elad et al., 2007). While ASP seems similar to the synthesis counterpart, it is very different when dealing with a redundant operator  $d > n$  (Nam et al., 2013). With ASP, the optimization problem for SMV signal recovery can be formulated as

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x}} \|\mathbf{y} - \Phi\mathbf{x}\|_2^2 + \lambda f(\boldsymbol{\Omega}\mathbf{x}). \quad (10)$$

Similarly, the optimization problem for MMV signal recovery is given by

$$\hat{\mathbf{X}} = \arg \min_{\mathbf{X}} \|\mathbf{Y} - \Phi\mathbf{X}\|_F^2 + \lambda g(\boldsymbol{\Omega}\mathbf{X}). \quad (11)$$

The regularization functions  $f(\cdot)$  and  $g(\cdot)$  are defined in (6) and (8). The successful recovery of the original signals from the compressed measurements using (10) and (11) has been theoretically guaranteed under the restricted isometry property adapted to the dictionary (D-RIP) and restricted orthogonal projection property (ROPP) (Candes et al., 2011; Nam et al., 2013; Peleg and Elad, 2013).

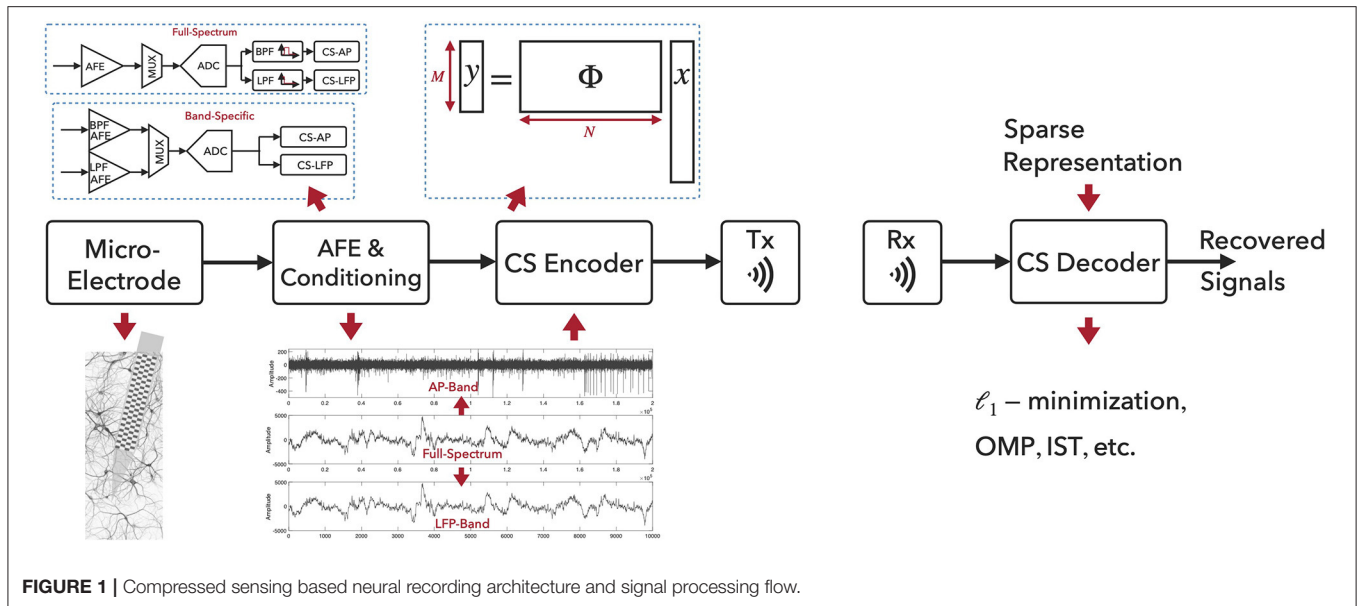


FIGURE 1 | Compressed sensing based neural recording architecture and signal processing flow.

## 2.2. CS-Based Neural Recording System Architecture

Figure 1 illustrates the CS-based neural recording system architecture and the signal processing flow. The aggregated extracellular neural signals are first picked up by the micro-machined neural probes. The recorded signals from brain structures contain broadband components up to tens of kilohertz, while the signals of general interest lie in two distinct frequency bands, i.e., the local field potentials located in the low-frequency band (below 500 Hz) and the action potentials in the high-frequency band (a few kilohertz). However, full-spectrum neural signals are not sparse on commonly used bases, and the sampling and compression of raw neural signals using CS would severely degrade the reconstruction performance. One widely used method to improve the neural signal reconstruction performance is to filter the full-spectrum signals into APs and LFPs for the separate compression tasks. Two types of neural recording architectures exist, as depicted in Figure 1. In the *Full-Spectrum* type, the broad-band neural signals are acquired and conditioned via analog front-end (AFE; e.g., neural amplifiers) with appropriate signal amplitude and bandwidth, and then digitized via Nyquist-rate ADCs. Digital band-pass filters (BPF) and low-pass filters (LPF) are employed and CS will then be applied on the filtered AP and LFP band signals. In the *Band-Specific* type, the neural signals are first filtered in the analog domain and then digitized. CS will be applied after the digitization. When applying CS to AP signals, CS is often combined with spike detection approach (Zhang et al., 2015; Liu et al., 2016; Sun et al., 2017; Wu et al., 2018; Zhao et al., 2018, 2019) for further compression. For AP signals, the neural spike events are detected and aligned temporally to their absolute peaks. The aligned segments (e.g., a segment of 64 samples) containing the spikes are then compressed via the CS technique. For LFP signals, CS can be directly applied to the time-series LFP data (e.g., with a segment of 256/512 samples).

## 2.3. Quality Evaluation Criterion

To evaluate the performance of CS recovery, the most widely used metric is the signal to noise and distortion ratio (SNDR), which quantifies the error between the original neural signal  $\mathbf{x}$  and the reconstructed signal  $\hat{\mathbf{x}}$ :

$$\text{SNDR} = 20 \log \frac{\|\mathbf{x}\|_2}{\|\mathbf{x} - \hat{\mathbf{x}}\|_2}. \quad (12)$$

SNDR is often used to evaluate the segment-to-segment error. For LFP reconstruction, SNDR can also be used to evaluate the spectral feature reconstruction capability across widely adopted frequency bands including  $\delta$  and  $\theta$  (1–8 Hz),  $\alpha$  (8–12 Hz),  $\beta$  (12–30 Hz),  $\gamma$  and high- $\gamma$  bands (30–70, 70–150 Hz), respectively. The other two widely used metrics are to evaluate the reconstruction accuracy in the frequency domain. The first metric is the average absolute error of spectral power ( $\epsilon_{\text{SP}}$ ), defined as

$$\epsilon_{\text{SP}} = \frac{1}{\#\text{Seg}} \sum_{i=1}^{\#\text{Seg}} |\text{SP}_{\text{rec}}(i) - \text{SP}_{\text{ori}}(i)|, \quad (13)$$

where  $\#\text{Seg}$  is the total number of LFP segments,  $\text{SP}_{\text{ori}}(i)$  is the ground-truth of spectral power in the  $i$ th segment calculated from the original LFPs, and  $\text{SP}_{\text{rec}}(i)$  is the spectral power calculated from the recovered LFPs. The second metric is the average absolute error percentage of spectral power ( $\epsilon_{\text{PSP}}$ ), defined as:

$$\epsilon_{\text{PSP}} = \frac{1}{\#\text{Seg}} \sum_{i=1}^{\#\text{Seg}} \frac{|\text{SP}_{\text{rec}}(i) - \text{SP}_{\text{ori}}(i)|}{\text{SP}_{\text{ori}}(i)}, \quad (14)$$

Besides the aforementioned metrics, classification accuracy (CA) is also widely used to evaluate the spike sorting performance of

neural action potentials. The CA was defined as:

$$CA = \frac{\text{\#correctly classified APs}}{\text{\#total APs}} \times 100\%. \quad (15)$$

To compute CA for a given method, all APs were compressed and reconstructed, then principal component analysis (PCA) was used to extract features from reconstructed APs. Finally, the first three principal components of each AP were used by clustering algorithms such as superparamagnetic clustering (SPC) (Quiroga et al., 2004), and the classification results were compared with the ground truth labels.

### 3. COMPRESSED SENSING OF EXTRACELLULAR NEURAL SIGNALS

In this section, we review the application of CS to extracellular neural signal compression. We first introduce the CS encoding part, which includes the sensing matrix design and hardware architecture. We then describe the sparse representations of both APs and LFPs. **Table 1** summarizes the reviewed literature on the compressed sensing of neural signals in a chronological order, with implementation details on CS encoding and decoding processes.

#### 3.1. CS Encoding: Sensing Matrices and Hardware Architecture

One unique advantage of CS is that the sensing matrices can be efficiently constructed and implemented in hardware. In practice, sensing matrices with binary entries have been widely investigated with algorithmic performance guarantees and efficient hardware implementation for both analog and digital CS encoders.

Gangopadhyay et al. (2011) presented an analog-domain CS front-end for sub-Nyquist sampling sparse brain signals like electroencephalogram (EEG) and electrocorticogram (ECoG). A one-bit switched-capacitor multiplying digital-to-analog (MDAC)/summation circuit was proposed for CS acquisition, which adopts a Bernoulli random matrix as the sensing matrix. Shoaran et al. (2014) introduced an area- and power-efficient multichannel analog CS encoder architecture for iEEG/ECoG signals for seizure prediction. This work exploits the spatial sparsity of the signals recorded from an ECoG electrode array. The benefits of employing a multichannel CS scheme were validated analytically and experimentally in a 0.18  $\mu\text{m}$  CMOS process. The results of simulations and subsequent reconstructions show the possibility of recovering fourfold compressed intracranial EEG signals with an SNDR up to 21.8 dB while consuming 10.5  $\mu\text{W}$  of power within an effective area of  $250 \times 250 \mu\text{m}$  per channel.

Chen et al. (2012) presented a comparative analysis between the analog and digital implementations of CS encoder designs. For biomedical signals that are not sampling frequency limited, digital CS encoders are more advantageous compared to the analog counterpart at the system level even when a Nyquist-rate ADC is used before the CS encoding stage. Binary entry sensing matrices like Bernoulli ( $\pm 1$ ) or random binary (0/1) matrices are

adopted for digital CS encoder implementations, which avoids the hardware-demanding multipliers. As such, for the scope of extracellular neural signals, the majority of CS encoding stages are implemented in digital styles. A parallel CS encoder with Bernoulli random matrix was presented by Zhang et al. (2014), where a 25-channel CS encoder occupies 0.06  $\text{mm}^2$  silicon area in a 0.18  $\mu\text{m}$  CMOS process and dissipates 270 nW power consumption with a 20 kHz sampling frequency and a supply voltage of 0.6 V. To further reduce the CS encoder hardware toward high-density on-chip neural signal compression tasks, the optimization of sensing matrices has been considered using both deterministic and random sparse sensing matrices (Zhao et al., 2016, 2018). A class of deterministic measurement matrices, namely a Quasi-Cyclic Array Code (QCAC) based matrix and a class of random measurement matrices, termed (1,s)-sparse random binary matrix [(1,s)-SRBM] were constructed. Both types of matrices are highly sparse and constructed with binary entries (0/1), which are exploited to realize both area- and energy-efficient CS encoder VLSI architectures. Note that both types of sensing matrices have sparse column features and can be efficiently generated online, thereby considerably reducing the needed amount of adders in traditional parallel CS encoder designs (Chen et al., 2012).

#### 3.2. CS Decoding: Sparse Representation and Recovery for Neural Signals

To achieve successful CS recovery, the signal of interest should satisfy the sparsity assumption. A major task of successful CS recovery is the proper design of sparsifying dictionaries/bases. Nevertheless, it is often challenging to find the proper sparse representations of the real-world neural signals on the decoder side. Moreover, there is no consensus and understanding of which sparsifying dictionary will be suitable for neural signals as they often exhibit non-sparse structures in most known bases like wavelet, Gabor, or Fourier types (Charbiwala et al., 2011; Gangopadhyay et al., 2011; Schmale et al., 2013; Shoaran et al., 2014). Moreover, neural signals are non-stationary and highly dynamic. In a multichannel neural recording setting, spike events are temporally sparse and the firing rates can vary significantly across different recording channels. Additionally, spike shape variation is common during chronic *in vivo* recording experiments, which might be caused by electrode drifts, oxidation from electrode-electrolyte interaction, sparse-firing neurons, or time-varying variations from the same neuron. These facts further complicate the discoveries of sparse representation of neural signals. In this section, we will separately discuss the sparse representation of different types of neural signals.

##### 3.2.1. Sparse Representation and Recovery for APs

A large body of research has focused on the compression of AP signals, which are the most informative biomarkers of interest for distinguishing the firing events from different neurons via spike sorting (Lewicki, 1998; Gibson et al., 2011). Bulach et al. (2012) considered applying CS for the compression of running AP band signals on both synthetic and recorded neural spike signals. Both wavelet (level-6, db-8) and learned dictionaries are used for signal

**TABLE 1** | Comparison of compressed sensing (CS) methods for neural signals.

References	CS encoder	CS decoder	Dataset	SNDR/CR
Gangopadhyay et al. (2011)	Analog	Wavelet	Physiobank	NA
	Bernoulli	$\ell_1$ -min	ECoG, EEG	16x
Charbiwala et al. (2011)	NA	Learned supports	Private	$\approx 20$ dB
		BPDN	EEG	2x
Chen et al. (2012)	Digital	NA	Private	$\approx 10$ dB
	Bernoulli/Binary	$\ell_1$ -min	EEG	10x
Schmale et al. (2013)	NA	DWT/DCT/DFT/WHT	MT dataset	$\approx 20$ dB
		NA	Spikes, LFPs	2x
Zhou et al. (2013)	NA	K-SVD	Hippocampus	$\approx 13.7$ dB
		Circulant and Toeplitz	Spikes	60x
Suo et al. (2014)	Digital	Spike data	Leicester	$\approx 20$ dB
		Bernoulli	Spikes	3x
Shoaran et al. (2014)	Analog	Gabor	Private	$\approx 14$ dB
		Bernoulli	ECoG	16x
Zhang et al. (2014)	Digital	K-SVD	Leicester	$\approx 13$ dB
		Bernoulli	Spikes	$\approx 5x$
Zhang et al. (2015)	Digital	K-SVD	Private	$\approx 13$ dB
		Bernoulli	Spikes	16x
Zamani et al. (2016)	Digital	BK-SVD	Leicester	$\approx 18$ dB
		Bernoulli	Spikes	9.8x
Sun et al. (2017)	Digital	Analysis prior	Leicester	$\approx 12$ dB
		Bernoulli	Spikes	8x
Zhao et al. (2018)	Digital	K-SVD	HC-1, CHB-MIT	$\approx 12$ dB
		Sparse Binary	Spikes, EEG	4x
Zhao et al. (2019)	Digital	DCT	HC-1	$\approx 28/25$ dB
		Binary	Spikes, LFPs	2x
Sun et al. (2021a)	Digital	Untrained DNN	Leicester, Neuropixels	$\approx 16$ dB
		Gaussian	Spikes	5x
Sun et al. (2021b)	Digital	ADMM	PFC-2, Neuropixels	$\approx 18$ dB
			Sparse Binary	LFPs

recovery. Nevertheless, the authors argued that CS is generally *not* recommended for compressing AP band neural signals, whereas CS is still applicable to the extracted and aligned spikes for further data rate reduction. Schmale et al. (2013) studied the common sparsity bases (Discrete Fourier/Cosine/Wavelet and Walsh-Hadamard Transforms) for both APs and LFPs. The authors argued that discrete cosine transform (DCT) turns out to be best suited for both neural signal modalities.

Although there is a limited success on the compression rate and the recovery performance using the known sparse basis with standard CS recovery methods, a data-dependent sparse basis is further investigated to improve the signal recovery performance (Charbiwala et al., 2011; Suo et al., 2013, 2014; Zhou et al., 2013; Zhang et al., 2014, 2015). Charbiwala et al.

(2011) investigates the compression of detected neural spikes, where the authors assumed empirically that spikes from different neurons are compressible in the wavelet domain (i.e., Daubechies wavelet) and exhibit nearly identical sparsity supports. These identical wavelet supports were learned over time to form a union of supports for spike recovery. The reconstruction algorithm is a combination of both BPDN (Chen and Donoho, 1994) (Basis Pursuit DeNoising) and modified BPDN (Lu and Vaswani, 2010) to leverage the union of supports. A 20-dB spike recovery SNDR and over 90% spike CA can be obtained for a compression ratio of 2. Zhou et al. (2013) demonstrated a sparsifying basis design via dictionary learning approach K-SVD (Aharon et al., 2006) to construct an overcomplete dictionary for the detected and aligned neural action potentials. Iterative

shrinkage thresholding (IST) algorithm was adopted for signal reconstruction. The learned sparsifying dictionary outperforms the DWT-based one with a compression ratio of 5 for an SNDR of 13.7 dB. Suo et al. (2014) introduced a multi-mode compressed sensing system for implantable neural recordings. Exploiting the *self expressiveness*, the authors proposed to use the spike data directly as the sparsifying dictionary. This choice of sparsifying dictionary shows comparable performance to the trained signal-dependent dictionary without extra computation for dictionary learning. Zhang et al. (2014) proposed a compact compressed sensing system for implantable neural recording applications, which also exploits K-SVD for sparse dictionary learning. The signal recovery can be achieved with a coarse estimation of spike shapes based on the learned sparse dictionary, and a fine detail estimation using a wavelet dictionary. Band-limited, inter-spike signals can be recovered via a standard CS recovery method using DWT as sparsifying bases. In Zamani et al. (2016), the block K-SVD (BK-SVD) algorithm was employed to train a block-sparsifying dictionary for neural spikes, then the block sparse Bayesian learning (BSBL) algorithm was adopted to reconstruct the original spikes. The proposed method achieved an overall CR of  $\sim 11.6$  and an SNDR that was up to 8 dB higher than that obtained without spike detection.

Note that signal-dependent sparse representation for neural signal compression cannot adapt to spike shape changes and thereby suffer from reconstruction quality loss when neural signals show varied sparsity supports over time. To accommodate this issue, Zhang et al. (2015) presented a low-power, closed-loop CS neural recording system, where a quality evaluation (QE) block was augmented to provide closed-loop feedback for reconstruction quality estimation and retraining. At the decoder side, the authors argued to adopt measurement SNDR<sub>y</sub> ( $20 \log_2 \frac{\|\mathbf{y}\|_2}{\|\mathbf{y}-\hat{\mathbf{y}}\|_2}$ ) as a quality estimation metric, where  $\mathbf{y}$  is the received measurement, and  $\hat{\mathbf{y}}$  is the reconstructed measurement obtained after the signal recovery ( $\hat{\mathbf{y}} = \Phi\hat{\mathbf{x}}$ ). A linear correlation is observed between SNDR<sub>y</sub> and SNDR<sub>x</sub>. Therefore, when SNDR<sub>y</sub> is lower than a threshold, a reduced recovery quality is observed and re-training is needed to improve the signal recovery. This quality estimation is performed on the decoder side, thereby posing no extra cost over the on-chip encoder resources.

Recently, a few online methods have been proposed for data-independent CS recovery of neural signals, which can eliminate the training and updating procedures required in the data-dependent approaches. The analysis sparse model has been employed for CS-based neural recording (Sun et al., 2017). The analysis model was adopted to enforce sparsity of the neural signals, overcoming the drawbacks of conventional synthesis models and enhancing the recovery performance. A multi-fractional-order difference matrix was constructed as the analysis operator and a group weighting analysis  $\ell_1$ -minimization was proposed for signal recovery. Experimental results on both the synthetic and real datasets revealed that the proposed approach outperforms data-dependent CS-based methods in terms of both spike recovery quality and CA. The synthesis sparse model has also been re-visited to enhance the CS reconstruction

performance (Zhao et al., 2019) by exploiting additional neural signal structure priors, such as BPF pole and the corner frequency information. A binary-weighted  $\ell_1$ -minimization algorithm was proposed to enforce the block structure by adding extra penalties to the sparse solvers when projections were made to high-frequency atoms, and improved the recovery performance for both AP and LFP signals, respectively.

### 3.2.2. Sparse Representation and Recovery for LFPs

Unlike AP signals, the compressed sensing of LFP signals only started to be addressed recently. Schmale et al. (2013) conducted a preliminary sparsity level study for the application of CS to the joint-compression of both LFPs and APs. Four well-known linear transformations (discrete Fourier/cosine/wavelet transform and Walsh–Hadamard transform) have been investigated and compared. Based on the sparsity analysis approach, the authors argued that the discrete cosine transform (DCT) turned out to be best suited for both neural signal modalities. Nevertheless, detailed CS performance has not been performed. Zhao et al. (2019) introduced a coarse-grained approach for LFP signals, where a high- $\gamma$  corner frequency is used as the block boundary for using a binary-weighted  $\ell_1$ -minimization algorithm. Based on the analysis model, CS reconstruction performance for LFPs was further improved by adopting the simultaneous sparse prior (Sun et al., 2021b) for joint-sparse LFP signals. The proposed method reinforces the sparsity with an optimal continuous order difference matrix as the analysis operator. A non-convex optimizer with an alternating direction method of multipliers (ADMM)-based solver was proposed to recover LFPs accurately and fast, which is promising for large-scale neural recording systems.

## 3.3. Discussion

As illustrated in **Table 1**, Bernoulli and binary sensing matrices are the dominant hardware implementation choices in most previous reported literature. Another observation is that the analog-domain CS encoder has limited applications to the extracellular neural signals. This is in part caused by the hardware implementation overhead as compared to the digital implementation, and the lack of knowledge on the sparse representation of complex neural signal modalities, including full-spectrum, LFP-, and AP-band signals in the previous literature. Note that the hardware cost of CS encoder is proportional to the segment size  $n$ , therefore the current CS encoder designs may not scale well toward large-scale neural recording applications. In fact, it is still possible to apply the analog CS encoding scheme after the analog filters, as long as the proper sparse priors are used during the signal reconstruction. Both synthesis (SSP) and analysis (ASP) sparse priors have been studied to improve the neural signal reconstruction quality. Early efforts have primarily focused on SSP and the corresponding recovery algorithms with rigorous theoretical guarantees (e.g.,  $\ell_1$ -minimization and OMP). The common drawback, however, is the poor reconstruction performance caused by non-sparse nature of the neural signals. As a result, learning sparse representations for neural signals has been proposed to improve reconstruction quality. However, this causes new issues regarding

the model robustness and practical deployment. Recent efforts on the analysis model and the spectral-sparse synthesis model have provided high-quality and training-free CS schemes well suited for both LFPs and APs. Finally, current CS methods for AP compression is still limited. First, overlapping spikes are precluded in the existing literature. It is also worth noting that spike alignment after detection is pre-assumed when applying CS to neural spike signals, which may cause additional hardware overheads to ensure successful CS recovery.

#### 4. DEEP LEARNING AS A FRAMEWORK TO SOLVE INVERSE PROBLEMS IN CS

There are a number of efforts employing the DL methods for CS-based neural signal compression tasks. Here, we provide a brief overview on how deep neural networks (DNNs) can be used to estimate  $\hat{x}$  for a given measurement  $y$  by learning a statistical signal transformation. DL can leverage large amounts of neural data to learn statistical transformations in certain high-dimensional spaces to improve the conventional CS algorithms and achieve superior results. The compression and reconstruction of APs (Sun et al., 2016; Sun and Feng, 2017; Wu et al., 2018) have been studied using DL. Nevertheless, there has been little literature on the LFPs compression using DL.

As a regression task, DL learns to solve the inverse problems in CS through supervised learning (Hecht-Nielsen, 1992). After training, no additional hyperparameter tuning is required. This stands in contrast to the traditional reconstruction algorithms in CS, where hyperparameters (e.g., sparsity in OMP and noise level in BPDN) need to be carefully hand-tuned, often through an iterative process to achieve the optimal convergence. By performing the desired transformation in a single feedforward manner, DL outperforms traditional iterative methods in terms of inference speed in general. Different types of DL models have been studied. Considering that AP is a time series signal, early works used fully connected (FC) layers to compress and reconstruct AP. Sun et al. (2016) and Sun and Feng (2017) presented a multilayer perceptron for CS, in which a reconstruction network consisting of multiple FC layers was optimized. Traditional CS used random sensing matrices (whose entries were randomly drawn from Gaussian or Bernoulli distributions), which are often sub-optimal. The DL-based approach still retains the hardware benefits from the linear sensing process of CS. In contrast, most DL-based CS methods considered sensing matrix optimization during the training process. To further increase the compression ratio, various quantization approaches were adopted in DL-based CS methods. Sun et al. (2016) and Sun and Feng (2017) jointly optimized a binary sensing matrix and a reconstruction network simultaneously, outperforming traditional sensing matrices in terms of both recovery quality and computation time. Moreover, a non-uniform multi-bit quantizer for compressed measurements quantization was proposed, which outperformed the uniform quantizers for wireless neural recording, especially with low quantization bit-depth (Sun et al., 2016). In Sun and Feng (2017), the entries of the sensing matrix were quantized

into one bit, thus reduced its storage requirement and boosted the sensing process. More recently, the training-free deep generative model has been successfully applied to AP compression and outperformed model-based and data-driven methods (Sun et al., 2021a).

The advantages of DL-based methods are the deep features extracted from the neural data, thereby avoiding manual designs of sparse dictionaries and yielding better reconstruction performance. However, there is still a lack of comprehensive understanding on the robustness and generalization capabilities of DL-based approaches. Also, the network may need to be retrained when recording and application scenarios are changed.

#### 5. REMAINING CHALLENGES

Here we list the remaining challenges for CS-based wireless neural recording systems:

**Hardware Efficiency Toward Large-Scale Neural Recording:** Despite the initial success, the hardware cost of CS can still be costly for high-density wireless neural recording devices. Aggressive approaches on the construction of hardware-efficient sensing matrices and circuit-level efforts to further minimize the computational hardware and measurement storage elements are required.

**Sparse Representation for Raw Neural Signals:** So far, CS cannot be applied to raw neural signals. This mainly originates from the fact that it lacks the knowledge of sparse representation of raw neural signals (Pagin and Ortmanns, 2018), which causes reduced signal reconstruction accuracy and eventually low performance in spike detection and sorting accuracy. Current CS approaches require the knowledge of spike location and only perform well on spike segments. This also limits the analog CS front-end designs for raw neural signals to reduce the sampling rate. If such sparse representation exists, this could bring extra opportunities for the CS-based neural signal compression tasks.

**Overlapped Spikes and Misalignment:** Neural recording based on multielectrode technology measures the activities from thousands of neurons simultaneously; it is, therefore, highly likely to observe multiple spikes within the same spike window, which is also referred to as the overlapping spikes. Currently, overlapping spikes are excluded and there is limited success for the signal reconstruction of overlapping spikes using CS methods. If successful, state-of-the-art spikes sorting approaches (e.g., KiloSort, Pachitariu et al., 2016) can therefore be incorporated for accurate neural coding tasks. Moreover, accurate on-chip spike alignment is difficult and can be hardware demanding due to low signal-to-noise ratio, sampling jitter, and noise effects (Gibson et al., 2011). Therefore, CS algorithms should be robust in the presence of spike overlapping as well as misalignment.

**Real-Time Signal Reconstruction:** The increasing number of channels is an irreversible trend in neural recording applications, which also brings huge computational burdens on the real-time performance of CS reconstruction. Most multi-channel CS algorithms exploit the intra- and inter-channel correlations to improve the signal reconstruction speed (Sun et al., 2021b).



Although the reconstruction efficiency is somewhat improved, the overall reconstruction accuracy decreases compared with the channel-by-channel reconstruction approach. Moreover, the channel correlation assumption may not always hold valid. Therefore, how to balance the reconstruction accuracy and speed remains a major challenge for CS-based large-scale neural recording applications.

## 6. CONCLUSION

This paper reviews the CS-based extracellular neural recording systems reported in the research literature in the past decade. CS has been widely considered to be a promising technique and actively studied for neurophysiology signal compression purposes. The key aspects of CS, i.e., the sensing matrix (both analog and digital CS encoder design), the sparse representation of neural signals (both APs and LFPs), and the corresponding

signal reconstruction algorithms, are covered. In particular, the associated challenges are discussed in detail at different CS stages for different neural signal modalities. Despite its current progress, there remain several challenging topics needed to be resolved toward practical CS-based neural recording systems in the future.

## AUTHOR CONTRIBUTIONS

Both authors contributed to manuscript planning, literature review and writing, and read and approved the submitted version.

## FUNDING

BS was supported by the National Natural Science Foundation of China under Grant 61971303. WZ was supported by the Binghamton University IEEC under Grant 1145629.

## REFERENCES

- Aharon, M., Elad, M., and Bruckstein, A. (2006). K-SVD: an algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Trans. Signal Process.* 54, 4311–4322. doi: 10.1109/TSP.2006.881199
- Allen, W. E., Chen, M. Z., Pichamoorthy, N., Tien, R. H., Pachitariu, M., Luo, L., et al. (2019). Thirst regulates motivated behavior through modulation of brainwide neural population dynamics. *Science* 364:eav3932. doi: 10.1126/science.aav3932
- Andersen, R. A., Musallam, S., and Pesaran, B. (2004). Selecting the signals for a brain-machine interface. *Curr. Opin. Neurobiol.* 14, 720–726. doi: 10.1016/j.conb.2004.10.005
- Becker, S., Bobin, J., and Candès, E. J. (2011). NESTA: A fast and accurate first-order method for sparse recovery. *SIAM J. Imaging Sci.* 4, 1–39. doi: 10.1137/090756855
- Berényi, A., Somogyvári, Z., Nagy, A. J., Roux, L., Long, J. D., Fujisawa, S., et al. (2014). Large-scale, high-density (up to 512 channels) recording of local circuits in behaving animals. *J. Neurophysiol.* 111, 1132–1149. doi: 10.1152/jn.00785.2013
- Bruckstein, A. M., Donoho, D. L., and Elad, M. (2009). From sparse solutions of systems of equations to sparse modeling of signals and images. *SIAM Rev.* 51, 34–81. doi: 10.1137/060657704
- Bulach, C., Bihl, U., and Ortmanns, M. (2012). “Evaluation study of compressed sensing for neural spike recordings,” in *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (San Diego, CA), 3507–3510. doi: 10.1109/EMBC.2012.6346722
- Buzsáki, G. (2004). Large-scale recording of neuronal ensembles. *Nat. Neurosci.* 7, 446–451. doi: 10.1038/nn1233
- Buzsáki, G., Anastassiou, C. A., and Koch, C. (2012). The origin of extracellular fields and currents-EEG, ECOG, LFP and spikes. *Nat. Rev. Neurosci.* 13, 407–420. doi: 10.1038/nrn3241
- Candès, E. J. (2008). The restricted isometry property and its implications for compressed sensing. *Comptes Rendus Math.* 346, 589–592. doi: 10.1016/j.crma.2008.03.014
- Candès, E. J., Eldar, Y. C., Needell, D., and Randall, P. (2011). Compressed sensing with coherent and redundant dictionaries. *Appl. Comput. Harmon. Anal.* 31, 59–73. doi: 10.1016/j.acha.2010.10.002
- Capogrosso, M., Milekovic, T., Borton, D., Wagner, F., Moraud, E. M., Mignardot, J.-B., et al. (2016). A brain-spine interface alleviating gait deficits after spinal cord injury in primates. *Nature* 539, 284–288. doi: 10.1038/nature20118
- Chae, M., Liu, W., Yang, Z., Chen, T., Kim, J., Sivaprakasam, M., et al. (2008). “A 128-channel 6 mw wireless neural recording IC with on-the-fly spike sorting and UWB transmitter,” in *2008 IEEE International Solid-State Circuits Conference-Digest of Technical Papers* (San Francisco, CA), 146–603. doi: 10.1109/ISSCC.2008.4523099
- Charbiwala, Z., Karkare, V., Gibson, S., Markovic, D., and Srivastava, M. B. (2011). “Compressive sensing of neural action potentials using a learned union of supports,” in *2011 International Conference on Body Sensor Networks* (Dallas, TX), 53–58. doi: 10.1109/BSN.2011.28
- Chen, F., Chandrakasan, A. P., and Stojanovic, V. M. (2012). Design and analysis of a hardware-efficient compressed sensing architecture for data compression in wireless sensors. *IEEE J. Solid-State Circ.* 47, 744–756. doi: 10.1109/JSSC.2011.2179451
- Chen, S., and Donoho, D. (1994). “Basis pursuit,” in *Proceedings of 1994 28th Asilomar Conference on Signals, Systems and Computers, Vol. 1* (Pacific Grove, CA), 41–44. doi: 10.1109/ACSSC.1994.471413
- Chen, X., Yu, Z., Hoyos, S., Sadler, B. M., and Silva-Martinez, J. (2010). A subnyquist rate sampling receiver exploiting compressive sensing. *IEEE Trans. Circ. Syst.* 58, 507–520. doi: 10.1109/TCSI.2010.2072430
- Cotter, S. F., Rao, B. D., Engan, K., and Kreutz-Delgado, K. (2005). Sparse solutions to linear inverse problems with multiple measurement vectors. *IEEE Trans. Signal Process.* 53, 2477–2488. doi: 10.1109/TSP.2005.849172
- Craven, D., McGinley, B., Kilmartin, L., Glavin, M., and Jones, E. (2014). Compressed sensing for bioelectric signals: a review. *IEEE J. Biomed. Health Inform.* 19, 529–540. doi: 10.1109/JBHI.2014.2327194
- Ding, C., Zhou, D., He, X., and Zha, H. (2006). “R 1-PCA: rotational invariant l 1-norm principal component analysis for robust subspace factorization,” in *Proceedings of the 23rd International Conference on Machine Learning* (Pittsburgh, PA), 281–288. doi: 10.1145/1143844.1143880
- Donoho, D. L. (2006). Compressed sensing. *IEEE Trans. Inform. Theory* 52, 1289–1306. doi: 10.1109/TIT.2006.871582
- Elad, M., Milanfar, P., and Rubinstein, R. (2007). Analysis versus synthesis in signal priors. *Inverse Probl.* 23:947. doi: 10.1088/0266-5611/23/3/007
- Gagnon-Turcotte, G., LeChasseur, Y., Bories, C., Messaddeq, Y., De Koninck, Y., and Gosselin, B. (2016). A wireless headstage for combined optogenetics and multichannel electrophysiological recording. *IEEE Trans. Biomed. Circuits Syst.* 11, 1–14. doi: 10.1109/TBCAS.2016.2547864
- Gangopadhyay, D., Allstot, E. G., Dixon, A. M., and Allstot, D. J. (2011). “System considerations for the compressive sampling of EEG and ECOG bio-signals,” in *2011 IEEE Biomedical Circuits and Systems Conference (BioCAS)* (San Diego, CA), 129–132. doi: 10.1109/BioCAS.2011.6107744
- Gibson, S., Judy, J. W., and Marković, D. (2011). Spike sorting: The first step in decoding the brain: the first step in decoding the brain. *IEEE Signal Process. Mag.* 29, 124–143. doi: 10.1109/MSP.2011.941880

- Grant, M., Boyd, S., and Ye, Y. (2008). Cvx: Matlab software for disciplined convex programming.
- Guurve, D., Delisle-Rodriguez, D., Bastos-Filho, T., and Krishnan, S. (2020). Trends in compressive sensing for EEG signal processing applications. *Sensors* 20:3703. doi: 10.3390/s20133703
- Haslinger, R., Ulbert, I., Moore, C. I., Brown, E. N., and Devor, A. (2006). Analysis of LFP phase predicts sensory response of barrel cortex. *J. Neurophysiol.* 96, 1658–1663. doi: 10.1152/jn.01288.2005
- Hecht-Nielsen, R. (1992). “Theory of the backpropagation neural network,” in *Neural Networks for Perception*, ed H. Wechsler (Washington, DC: Elsevier), 65–93. doi: 10.1016/B978-0-12-741252-8.50010-8
- Hong, G., and Lieber, C. M. (2019). Novel electrode technologies for neural recordings. *Nat. Rev. Neurosci.* 20, 330–345. doi: 10.1038/s41583-019-0140-6
- Hubel, D. H. (1957). Tungsten microelectrode for recording from single units. *Science* 125, 549–550. doi: 10.1126/science.125.3247.549
- Intan, R. (2021). *Intan*. Available online at: [https://intantech.com/files/Intan\\_RHD2164\\_datasheet.pdf](https://intantech.com/files/Intan_RHD2164_datasheet.pdf)
- Jaspan, O. N., Fleysler, R., and Lipton, M. L. (2015). Compressed sensing MRI: a review of the clinical literature. *Br. J. Radiol.* 88:20150487. doi: 10.1259/bjr.20150487
- Jun, J. J., Steinmetz, N. A., Siegle, J. H., Denman, D. J., Bauza, M., Barbarits, B., et al. (2017). Fully integrated silicon probes for high-density recording of neural activity. *Nature* 551, 232–236. doi: 10.1038/nature24636
- Kamboh, A. M., Raetz, M., Oweiss, K. G., and Mason, A. (2007). Area-power efficient VLSI implementation of multichannel dwt for data compression in implantable neuroprosthetics. *IEEE Trans. Biomed. Circuits Syst.* 1, 128–135. doi: 10.1109/TBCAS.2007.907557
- Larson, L., and Nurmikko, A. (2016). “Microwave communication links for brain interface applications,” in *2016 IEEE 16th Topical Meeting on Silicon Monolithic Integrated Circuits in RF Systems (SiRF)* (Austin, TX), 73–76. doi: 10.1109/SiRF.2016.7445472
- Lewicki, M. S. (1998). A review of methods for spike sorting: the detection and classification of neural action potentials. *Network* 9, R53–R78. doi: 10.1088/0954-898X\_9\_4\_001
- Liu, X., Zhang, M., Xiong, T., Richardson, A. G., Lucas, T. H., Chin, P. S., et al. (2016). A fully integrated wireless compressed sensing neural signal acquisition system for chronic recording and brain machine interface. *IEEE Trans. Biomed. Circ. Syst.* 10, 874–883. doi: 10.1109/TBCAS.2016.2574362
- Lopez, C. M., Putzeys, J., Raducanu, B. C., Ballini, M., Wang, S., Andrei, A., et al. (2017). A neural probe with up to 966 electrodes and up to 384 configurable channels in 0.13μm SOI CMOS. *IEEE Trans. Biomed. Circuits Syst.* 11, 510–522. doi: 10.1109/TBCAS.2016.2646901
- Lu, W., and Vaswani, N. (2010). “Modified basis pursuit denoising (modified-BPDN) for noisy compressive sensing with partially known support,” in *2010 IEEE International Conference on Acoustics, Speech and Signal Processing* (Dallas, TX), 3926–3929. doi: 10.1109/ICASSP.2010.5495799
- Lustig, M., Donoho, D. L., Santos, J. M., and Pauly, J. M. (2008). Compressed sensing MRI. *IEEE Signal Process. Mag.* 25, 72–82. doi: 10.1109/MSP.2007.914728
- Martinez, D., Clément, M., Messaoudi, B., Gervasoni, D., Litaudon, P., and Buonviso, N. (2018). Adaptive quantization of local field potentials for wireless implants in freely moving animals: an open-source neural recording device. *J. Neural Eng.* 15:025001. doi: 10.1088/1741-2552/aaa041
- Mishali, M., and Eldar, Y. C. (2011). Sub-nyquist sampling. *IEEE Signal Process. Mag.* 28, 98–124. doi: 10.1109/MSP.2011.942308
- Musk, E., and Neuralink (2019). An integrated brain-machine interface platform with thousands of channels. *J. Med. Intern. Res.* 21:e16194. doi: 10.2196/16194
- Nam, S., Davies, M. E., Elad, M., and Gribonval, R. (2013). The cosparsity analysis model and algorithms. *Appl. Comput. Harmon. Anal.* 34, 30–56. doi: 10.1016/j.acha.2012.03.006
- Nurmikko, A. (2020). Challenges for large-scale cortical interfaces. *Neuron* 108, 259–269. doi: 10.1016/j.neuron.2020.10.015
- Oppenheim, A. V. (1999). *Discrete-Time Signal Processing*. Pearson Education India.
- Pachitariu, M., Steinmetz, N., Kadir, S., Carandini, M., and Harris, K. (2016). “Fast and accurate spike sorting of high-channel count probes with kilosort,” in *NIPS Proceedings* (Barcelona: Neural Information Systems Foundation, Inc).
- Pagin, M., and Ortmanns, M. (2018). “Study of compressed sensing and predictor techniques for the compression of neural signals under the influence of noise,” in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* (Honolulu, HI), 1102–1105. doi: 10.1109/EMBC.2018.8512469
- Peleg, T., and Elad, M. (2013). Performance guarantees of the thresholding algorithm for the cosparsity analysis model. *IEEE Trans. Inform. Theory* 59, 1832–1845. doi: 10.1109/TIT.2012.2226924
- Quiroga, R. Q., Nadasdy, Z., and Ben-Shaul, Y. (2004). Unsupervised spike detection and sorting with wavelets and superparamagnetic clustering. *Neural Comput.* 16, 1661–1687. doi: 10.1162/089976604774201631
- Schmale, S., Knoop, B., Hoeffmann, J., Peters-Drolshagen, D., and Paul, S. (2013). “Joint compression of neural action potentials and local field potentials,” in *2013 Asilomar Conference on Signals, Systems and Computers* (Pacific Grove, CA), 1823–1827. doi: 10.1109/ACSSC.2013.6810617
- Schwarz, D. A., Lebedev, M. A., Hanson, T. L., Dimitrov, D. F., Lehew, G., Meloy, J., et al. (2014). Chronic, wireless recordings of large-scale brain activity in freely moving rhesus monkeys. *Nat. Methods* 11, 670–676. doi: 10.1038/nmeth.2936
- Shoaran, M., Kamal, M. H., Pollo, C., Vanderghenst, P., and Schmid, A. (2014). Compact low-power cortical recording architecture for compressive multichannel data acquisition. *IEEE Trans. Biomed. Circ. Syst.* 8, 857–870. doi: 10.1109/TBCAS.2014.2304582
- Stevenson, I. H., and Kording, K. P. (2011). How advances in neural recording affect data analysis. *Nat. Neurosci.* 14, 139–142. doi: 10.1038/nn.2731
- Stringer, C., Pachitariu, M., Steinmetz, N., Carandini, M., and Harris, K. D. (2019). High-dimensional geometry of population responses in visual cortex. *Nature* 571, 361–365. doi: 10.1038/s41586-019-1346-5
- Sun, B., and Feng, H. (2017). Efficient compressed sensing for wireless neural recording: a deep learning approach. *IEEE Signal Process. Lett.* 24, 863–867. doi: 10.1109/LSP.2017.2697970
- Sun, B., Feng, H., Chen, K., and Zhu, X. (2016). A deep learning framework of quantized compressed sensing for wireless neural recording. *IEEE Access* 4, 5169–5178. doi: 10.1109/ACCESS.2016.2604397
- Sun, B., Mu, C., Wu, Z., and Zhu, X. (2021a). Training-free deep generative networks for compressed sensing of neural action potentials. *IEEE Trans. Neural Netw. Learn. Syst.* doi: 10.1109/TNNLS.2021.3069436. [Epub ahead of print].
- Sun, B., Zhang, H., Zhang, Y., Wu, Z., Bao, B., Hu, Y., et al. (2021b). Compressed sensing of large-scale local field potentials using adaptive sparsity analysis and non-convex optimization. *J. Neural Eng.* 18:026007. doi: 10.1088/1741-2552/abd578
- Sun, B., Zhao, W., and Zhu, X. (2017). Training-free compressed sensing for wireless neural recording using analysis model and group weighted-minimization. *J. Neural Eng.* 14:036018. doi: 10.1088/1741-2552/aa630e
- Suo, Y., Zhang, J., Etienne-Cummings, R., Tran, T. D., and Chin, S. (2013). “Energy-efficient two-stage compressed sensing method for implantable neural recordings,” in *2013 IEEE Biomedical Circuits and Systems Conference (BioCAS)* (Rotterdam), 150–153. doi: 10.1109/BioCAS.2013.6679661
- Suo, Y., Zhang, J., Xiong, T., Chin, P. S., Etienne-Cummings, R., and Tran, T. D. (2014). Energy-efficient multi-mode compressed sensing system for implantable neural recordings. *IEEE Trans. Biomed. Circuits Syst.* 8, doi: 10.1109/TBCAS.2014.2359180
- Testard, C., Tremblay, S., and Platt, M. (2021). From the field to the lab and back: neuroethology of primate social behavior. *Curr. Opin. Neurobiol.* 68, 76–83. doi: 10.1016/j.conb.2021.01.005
- Tropp, J. A., and Gilbert, A. C. (2007). Signal recovery from random measurements via orthogonal matching pursuit. *IEEE Trans. Inform. Theory* 53, 4655–4666. doi: 10.1109/TIT.2007.909108
- van den Berg, E., and Friedlander, M. P. (2007). SPGL1: A solver for large-scale sparse reconstruction.
- Vyazovskiy, V. V., Olcese, U., Hanlon, E. C., Nir, Y., Cirelli, C., and Tononi, G. (2011). Local sleep in awake rats. *Nature* 472, 443–447. doi: 10.1038/nature10009
- Wu, T., Zhao, W., Guo, H., Lim, H. H., and Yang, Z. (2017). A streaming PCA VLSI chip for neural data compression. *IEEE Trans. Biomed. Circ. Syst.* 11, 1290–1302. doi: 10.1109/TBCAS.2017.2717281

- Wu, T., Zhao, W., Keefer, E., and Yang, Z. (2018). Deep compressive autoencoder for action potential compression in large-scale neural recording. *J. Neural Eng.* 15:066019. doi: 10.1088/1741-2552/aae18d
- Xu, J., Nguyen, A. T., Zhao, W., Guo, H., Wu, T., Wiggins, H., et al. (2018). A low-noise, wireless, frequency-shaping neural recorder. *IEEE J. Emerg. Sel. Top. Circ. Syst.* 8, 187–200. doi: 10.1109/JETCAS.2018.2812104
- Yin, M., Borton, D. A., Komar, J., Agha, N., Lu, Y., Li, H., et al. (2014). Wireless neurosensor for full-spectrum electrophysiology recordings during free behavior. *Neuron* 84, 1170–1182. doi: 10.1016/j.neuron.2014.11.010
- Yoo, J., Becker, S., Loh, M., Monge, M., Candes, E., and Emami-Neyestanak, A. (2012). “A 100 MHz-2 GHz 12.5 x sub-nyquist rate receiver in 90 nm cmos,” in *2012 IEEE Radio Frequency Integrated Circuits Symposium* (Montreal, QC), 31–34. doi: 10.1109/RFIC.2012.6242225
- Zamani, H., Bahrami, H., and Mohseni, P. (2016). “On the use of compressive sensing (cs) exploiting block sparsity for neural spike recording,” in *2016 IEEE Biomedical Circuits and Systems Conference (BioCAS)* (Shanghai), 228–231.
- Zhang, J., Mitra, S., Suo, Y., Cheng, A., Xiong, T., Michon, F., et al. (2015). A closed-loop compressive-sensing-based neural recording system. *J. Neural Eng.* 12:036005. doi: 10.1088/1741-2550/12/3/036005
- Zhang, J., Suo, Y., Mitra, S., Chin, S., Hsiao, S., Yazicioglu, R. F., et al. (2014). An efficient and compact compressed sensing microsystem for implantable neural recordings. *IEEE Trans. Biomed. Circ. Syst.* 8, 485–496. doi: 10.1109/TBCAS.2013.2284254
- Zhang, Y. (2009). *User's Guide for Yall1: Your Algorithms for l1 Optimization*. Technical report. Rice University.
- Zhao, W., Sun, B., Wu, T., and Yang, Z. (2016). “Hardware efficient, deterministic QCAC matrix based compressed sensing encoder architecture for wireless neural recording application,” in *2016 IEEE Biomedical Circuits and Systems Conference (BioCAS)* (Shanghai), 212–215. doi: 10.1109/BioCAS.2016.7833769
- Zhao, W., Sun, B., Wu, T., and Yang, Z. (2018). On-chip neural data compression based on compressed sensing with sparse sensing matrices. *IEEE Trans. Biomed. Circuits Syst.* 12, 242–254. doi: 10.1109/TBCAS.2017.2779503
- Zhao, W., Wu, T., Xu, J., Zhao, Q., and Yang, Z. (2019). “Block-sparse modeling for compressed sensing of neural action potentials and local field potentials,” in *2019 53rd Asilomar Conference on Signals, Systems, and Computers* (Pacific Grove, CA), 2097–2100. doi: 10.1109/IEEECONF44664.2019.9048841
- Zhou, A., Santacruz, S. R., Johnson, B. C., Alexandrov, G., Moin, A., Burghardt, F. L., et al. (2019). A wireless and artefact-free 128-channel neuromodulation device for closed-loop stimulation and recording in non-human primates. *Nat. Biomed. Eng.* 3, 15–26. doi: 10.1038/s41551-018-0323-x
- Zhou, S., Dai, B., Xiang, Y., Xu, S., Zhang, B., Song, Y., et al. (2013). “Compressive sensing of neural action potentials by designing overcomplete dictionaries,” in *2013 IEEE International Conference on Green Computing and Communications and IEEE Internet of Things and IEEE Cyber, Physical and Social Computing* (Beijing), 1848–1852. doi: 10.1109/GreenCom-iThings-CPSCom.2013.343

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Sun and Zhao. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.