Digital Signal Processing 55 (2016) 64–77

Contents lists available at ScienceDirect

Digital Signal Processing

www.elsevier.com/locate/dsp

A digital multichannel neural signal processing system using compressed sensing

Nan Li [a,*], Morgan Osborn [a], Guoxing Wang [b], Mohamad Sawan [a]

[a] Polystim Neurotechnologies Lab., Electrical Engineering Dept., Polytechnique Montreal, 2900 Edouard-Monpetit, H3T 1J4, Montreal (QC), Canada
[b] Department of Micro/Nano Electronics, School of Microelectronics, Shanghai Jiao Tong University, Dongchuan Road #800, Minhang District, Shanghai, 200240 China

ARTICLE INFO

Article history:
Available online 9 May 2016

Keywords:
Multichannel neural recording
Neural signal processing
Data compression
Compressed sensing
DSP
VLSI

ABSTRACT

This paper concerns a wireless multichannel neural recording system using a compressed sensing technique to compress the recorded data. We put forth a single and a multichannel system applying a Minimum Euclidean or Manhattan Distance Cluster-based (MDC) deterministic compressed sensing matrix. The single-channel signal processing system is composed of spike detection and data compression blocks. For the construction of the MDC matrix, the distance σ is an important parameter, which can take a value of 4 or 5. In addition, the sharing strategy is used to construct a multichannel system, and we analyze the influence of the number of the channels; scan rate on the reconstruction error, compression rate and power consumption; the influence of the signal-to-noise ratio; and reconstruction performance on neural signals. Based on the results, a 256-channel digital signal processing system, implemented in a 130-nm CMOS process, is proposed. This system has power consumption per channel of 12.5 µW and silicon area per channel of 0.03 mm^2, and provides data reduction of around 90% while enabling accurate reconstruction of the original signals.

1. Introduction

Wireless monitoring of neural activity through implantable devices is an important technology that enables advanced diagnosis and treatment of brain disorders such as Parkinson's disease, major depressive disorder and epilepsy [1–3]. Fig. 1 shows a typical wireless neural recording system. However, designing such a wireless neural recording device faces numerous challenges. These include integrating high-density recording electrodes [4,5], avoiding the heating of tissues due to energy transfer to power the implants (the maximum power density should be 0.8 mW/mm^2 for the exposed tissue area [6]), maximizing the device lifetime [7,8], and minimizing the device size [9]. The conflict between huge data size and limited energy available for implantable recording devices is one of the principal challenges; specifically, integrating the necessary wireless transmission component in an implantable device exacerbates the problem of stringent energy constraints [10]. Therefore, data reduction or compression strategies should be employed to minimize the power consumption of the dedicated implantable devices.

Several neural signal reduction or compression techniques are already in use. Signal reduction is widely used to implement data reduction under certain constraints; methods include neural spike detection [11–13] and data feature extraction [14,15]. Both methods involve locating important information and eliminating the remaining parts of the signals. However, signal reduction methods distort or lose some necessary information. For instance, a spike-detection-based neural recording device usually obtains data as the time series or the impulse, which cannot provide the details (shape or amplitude) of the original signal or spikes [16]; feature extraction methods are usually computationally complex, which conflicts with the design of a low-power device [17]. Therefore, it is necessary to find a new method that does not cause significant loss of features when recording neural signals.

Data compression methods avoid these drawbacks by preserving maximum information during the compression phase, which allows recovery of the original signal. Recently introduced compressed sensing (CS) technique shows great potential in compressing neural signals [18]. CS has low-encoder complexity and universality for different kinds of signals. It is a revolution for the traditional Nyquist sampling frequency (Shannon theory), which has attracted considerable attention in the areas of computer science, applied mathematics and electrical engineering [19]. CS preserves the temporal and morphological information of the signal, which

* Corresponding author.
E-mail address: nan.li@polymtl.ca (N. Li).

http://dx.doi.org/10.1016/j.dsp.2016.04.013

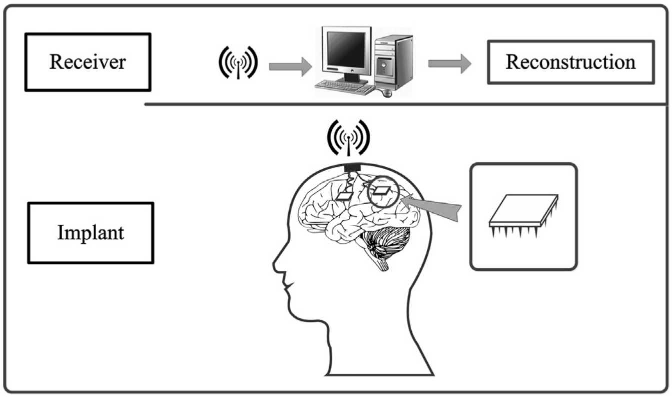

Fig. 1. Simplified diagram of a typical wireless neural monitoring system.

is much better than spike detection or feature extraction methods [20].

1.1. Introduction of the CS technique

In this section, we briefly introduce basic concepts in CS theory. First, the sparsity of the signal is an important concept. A sparse signal can be compressed through a sensing matrix. Suppose a vector (or signal) $x(x_1, x_2, \ldots, x_n) \in \mathbb{R}^N$ and some items of x are zero or close to zero, so this vector can be called a sparse vector (or signal). If x is not sparse in the current basis, but it is sparse under some bases, then it still can be regarded as a sparse signal. For example, suppose a basis $\Psi_{N\times T}$, in which $x = \Psi z$ can be sparsely represented, so x is sparse under basis Ψ.

If $x(x_1, x_2, \ldots, x_n) \in \mathbb{R}^N$ is sparse, then it can be compressed through a sensing matrix $\Phi_{N\times M}$ to $y \in \mathbb{R}^M$. When the sparsity of the signal is large, x can be largely compressed, that is, $M \ll N$, which can be described as in (1).

$$y = \Phi_{N\times M} x \tag{1}$$

If x is sparse under basis Ψ, then (1) can change to (2).

$$y = \Phi \Psi z \tag{2}$$

Second, the original signal can be reconstructed by ℓ_1 minimization. Given the original sparse signals and the measurement y, the best way to reconstruct the signal is through ℓ_0 minimization [19]. But finding a solution that approximates ℓ_0 minimization is NP (non-deterministic polynomial-time) hard; therefore, ℓ_1 minimization is widely used in signal reconstruction for CS application [19]. The form of ℓ_1 minimization is shown in (3). Based on the signal reconstruction via ℓ_1 minimization, many signal reconstruction algorithms exist. ℓ_1 minimization reconstruction algorithms, which directly use framework shown in (3), are powerful methods for computing sparse representations [21]; basis pursuit algorithms (BP) belong to this category [22]. Greedy algorithms are another category, which includes match pursuit algorithm (MP), orthogonal matching pursuit algorithm (OMP) [23], and iterative hard or soft thresholding algorithm [24,25]. Greedy algorithms are computationally efficient, but they are usually sensitive to noise especially when the original signals are not exactly sparse. By comparison, ℓ_1 minimization reconstruction algorithms are more robust to noise but at the price of a higher computational cost [26]. In addition, other kinds of algorithms can be used to reconstruct the original signals; for example, a Bayesian-based reconstruction method, called Block Sparse Bayesian Learning (BSBL) algorithm, uses the maximum likelihood to reconstruct the signal, and can reconstruct non-sparse signals [27].

$$x' = \underset{z}{\operatorname{argmin}} \|z\|_1 \qquad \text{subject to} \quad z \in \mathrm{B}(y) \tag{3}$$

where $\mathrm{B}(y) = \{z : \mathrm{A}z = y\}$.

Third, the design of the sensing matrix is another important topic. The sensing matrix strongly influences the amount of reconstruction error and also transmission of compressed signals [28]. In CS theory, the sensing matrix Φ can be a random matrix, such as a sub-Gaussian matrix [29], a random discrete Fourier transmission matrix [30], or a deterministic matrix, such as the Discrete Chirp matrix [31], the Reed Muller matrix [32], low-density parity-check (LDPC) matrix [33]. To correctly reconstruct x, the sensing matrix $\Phi_{N\times M}$ should obey the Restricted Isometry Property, which is described as follows.

Restricted Isometry Property An $M \times N$ sensing matrix Φ is said to satisfy an Restricted Isometry Property (RIP) of *k order*, if it satisfies (4),

$$(1-\varepsilon_k)\|X\|_2^2 \le \|\Phi X\|_2^2 \le (1+\varepsilon_k)\|X\|_2^2 \tag{4}$$

for all the *k-sparse* vectors X. The restricted isometry constant ε_k of matrix Φ lies between 0 and 1. The process of CS compression is shown in Fig. 2. In this diagram, a sparse signal is firstly compressed by a sensing matrix. Then the signal is recovered through the ℓ_1 norm-based reconstruction. After the reconstruction, if x is sparse under the basis Ψ, it still needs to recover the signal in the current basis.

Finally, the research in the field of compressed sensing is not just in the theoretical concept but also in the design of underlying circuitry. There are several articles about the application of the CS technique [34–36]. Also, some designers used the CS technique to design the neural recording circuit [10,20,37]. Fig. 3 shows the principles of use of the CS technique in neural recording circuit design. Figs. 3(a) and 3(b) depict analog and digital single-channel designs that apply the CS technique. The designs have two common parts: a sensing matrix generator and a multiplication block. In Fig. 3, the sensing matrix generator could be a random or deterministic matrix (vector) generator, but most current designs use a random sensing matrix to design the circuit. The multiplication block does the matrix multiplication of the sensing matrix and the signal vector.

1.2. Contribution of this article

In a recent article, we introduced a sensing matrix construction method called a minimum Euclidean or Manhattan distance cluster-based deterministic (MDC) sensing matrix [38]. In this article, we proved that neural signals were not sparse, but contained many identical (or similar) points. We researched these identical or highly similar non-zero points, i.e., the similarity of the points in a signal, to compress neural signals, and we proposed the MDC matrix. From the simulation results, we proved that neural signals can be largely compressed with unit MDC (UMDC) matrix and also can be well reconstructed by basis pursuit algorithm for sparse signals or non-sparse signals which contain lots of similar points. In addition, we proved that the MDC matrix obeys the RIP under two prerequisites, and we concluded that the MDC matrix can compress a signal with a relatively large compression rate (CR) and small reconstruction error rate (RER).

In this article, our contribution is using the MDC matrix to implement single and multichannel digital systems. It can be seen from Fig. 3 that the process of sensing matrix generation does not include any information from the signals that need to be compressed, but the MDC matrix can use the information of the signal. According to [38], we design a digital signal processing system which applies the MDC matrix; the principle of the circuit is shown in Fig. 3(c). The difference between our design and the ones in Figs. 3(a) and 3(b) is that our design uses the information of the signal itself to generate the sensing matrix. In later sections, we give details of construction of a digital circuit using the MDC