COMPRESSED SENSING AND ENERGY-AWARE INDEPENDENT COMPONENT ANALYSIS FOR COMPRESSION OF EEG SIGNALS

Simon Fauvel[⋆] *Abhinav Agarwal*[†] *Rabab Ward*[⋆ ∗]

[⋆] Electrical and Computer Engineering, University of British Columbia, Vancouver, Canada
[†] Department of Electronics and Electrical Engineering, Indian Institute of Technology Guwahati, India

ABSTRACT

In this paper, we propose the use of compressed sensing (CS) that is preceded by an energy-efficient, cross-product based independent component analysis (ICA) preprocessing method to efficiently compress electroencephalogram (EEG) signals in the context of a wireless body sensor network (WBSN). In WBSNs, the battery life puts a strict energy constraint at each sensor node. By providing a simple, nonadaptive compression scheme at the sensor nodes, CS offers an efficient solution to compress EEG signals in WBSNs. Through simulations, we demonstrate that our method requires less energy than other state-of-the-art methods using ICA, with a reduction in computations that can reach up to 94%. We also demonstrate that for a fixed compression ratio, the achievable reconstruction error is similar to the state-of-the-art method using ICA, and is much lower than when CS is used alone.

Index Terms— Compressed sensing (CS), Independent component analysis (ICA), Wireless body sensor network, Electroencephalogram (EEG)

1. INTRODUCTION

In recent years, telemedecine has been gaining popularity around the world. Solutions are sought to mitigate the current trend of increasing healthcare costs, and telemedecine is one such solution [1]. Wireless body sensor networks (WBSNs) are at the heart of many telemedicine applications. Through various wireless sensor nodes located on the patient's body, different physiological signals are acquired. The aim is to utilize these signals to assess different medical conditions [2]. One such signal of interest is the electroencephalogram (EEG), which non-invasively measures the electrical activity from the brain. Such signals can then be used for different purposes such as detecting and predicting seizures [3], assessing sleep patterns [4], controlling a brain computer interface [5], and many others.

The main advantages of WBSNs are that they are non-invasive and are less cumbersome for the patient to wear (as compared to their wired counterparts). However, the main challenge in these applications is that the energy supply at the sensor nodes is limited. This forms a particularly important problem in the case of EEG due to the huge amount of data. It is not uncommon to have 10 EEG channels or more, each sampling at 100Hz at the very least. This significant amount of raw data must then be transmitted wirelessly. Therefore, finding an appropriate compression algorithm that could subsample the acquired EEG data at the sensor nodes while still maintaining the relevant features is an important challenge. This has to be done using a minimal amount of computations at the sensor node level due to the sensor's energy limitation.

In recent years, compressed sensing (CS) has gained popularity as a lossy compression scheme that can non-adaptively sample and compress signals at a rate much lower than the Nyquist frequency. In a nutshell, if a signal can be sparsely represented by some basis (i.e., if it can be well represented by a small number of coefficients), then a small number of random projections, roughly proportional to the information rate of the signal, is sufficient to recover the signal exactly. The theory extends to compressible signals (where the signal has many very small coefficients in some basis), although in this case the reconstruction is not exact [6].

CS is an interesting paradigm in the case of WBSN since it requires very simple computations at the sensor nodes (non-adaptive random projections) where the signals are acquired and compressed. In WBSNs, we generally place no limitation on the energy consumption (and computational power) of the server, whereas the sensor nodes are heavily constrained both in terms of energy available and computational power.

Aviyente was the first to investigate the potential of CS as applied to EEG signals [7]. While she obtained good results through the use of a distributed CS reconstruction algorithm (where all channels are reconstructed at once), these results were based on the compression of EEG signals coming from multiple trials (asking a patient to repeat the same task many times and recording the EEG signals each time). This setting is of no interest in telemedicine applications, where the patient is usually not prompted to act in a certain way or to repeat the same task multiple times. To the best of our knowledge, Abdulghani et al. were the first to look at CS for the compression of EEG signals in telemedicine ([8]-[9]).

∗This work was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC), and by the Qatar National Research Fund (QNRF No. NPRP 09-310-1-058).

In their work, they surveyed different sparsifying bases and reconstruction algorithms, coming to the conclusion that the applicability of single-channel CS for EEG signals depends on the final application and the tolerable reconstruction error. More recently, Mijovic et al. proposed to apply Independent Component Analysis (ICA) as a preprocessing step prior to using CS for the compression of EEG signals in newborn babies [10]. They obtained compression results superior to what others not using ICA preprocessing previously obtained. Their system however is not suitable for telemedicine applications since it is energy hungry at the sensor nodes (because of the selected CS measurement matrix and the ICA algorithm). They also applied it on neonatal EEG signals, which are arguably simpler than adult EEG. In this paper, we propose a framework that is suitable for a low-energy, power-efficient telemedicine application. Through the development of a power-efficient ICA algorithm and architecture and by adapting the CS scheme to reduce the load at the sensor nodes, we are able to improve the energy efficiency at these nodes and at the same time achieve similar EEG compression results as those achieved by Mijovic et al.

This paper is organized as follows. Section 2 presents the different building blocks of the framework. Section 3 shows the performance of the system through different experiments. Section 4 discusses the obtained results, while Section 5 concludes the paper.

2. METHODS

The proposed framework for compressing EEG signals is composed of two blocks: the ICA preprocessing block, followed by the compressed sensing one. In our framework, compressed sensing is applied to the derived independent components directly.

2.1. Energy Efficient ICA

ICA has been around for close to two decades now and was proposed as a solution to the blind source separation problem. The solution to the problem is found by enforcing statistical independence of the source signals, commonly through maximizing the non-Gaussianity of the signals or through minimizing the mutual information of the signals [11]. ICA has also been successfully applied to EEG signals (see, for example, the study conducted by Makeig et al. [12]). The underlying basis as to why ICA works in the EEG case is that the electrical scalp potential measured by an EEG electrode can be seen as a mixture of a smaller number of 'sources' located in the brain that give rise to these potentials.

The ICA problem can be formulated as $\boldsymbol{X} = \boldsymbol{AS}$, where $\boldsymbol{X}$ is a matrix containing the measured mixed signals (each column containing one mixed signal), $\boldsymbol{A}$ is the mixing matrix, and $\boldsymbol{S}$ is a matrix containing the independent components (one source per column). The task is to find $\boldsymbol{A}$ and $\boldsymbol{S}$ from the observable measurements $\boldsymbol{X}$.

FastICA (FICA) is a popular algorithm that can solve this problem [12]. Summarizing the FICA algorithm:

1. *Preprocessing*: Whiten (decorrelate) the mixed signals.
2. *Iteration*: In a deflationary manner (i.e. one at a time), estimate each independent component (IC) by maximizing its non-Gaussianity through a contrast function. Using Gram-Schmidt, orthogonalize the found IC with respect to the previously found ICs, and normalize it. Repeat this stage until the component converges.

However, this algorithm is computationally intensive and thus not suitable for WBSN applications. Acharyya et al. proposed an algorithm and an energy-efficient architecture to calculate n-dimensional (nD) cross-products, and mentioned that one potential application is FICA [13]. After identifying $n-1$ components with FICA, the n^{th} component can simply be identified by taking the cross-product of the first $n-1$ components since at that point, the direction for maximal independence has already been determined. We call this method xFICA$^{\text{nD}}$. This allows the saving of one full iteration cycle, which is not negligible.

2.2. Compressed Sensing

Consider a signal $\boldsymbol{f}$ of length N, and a dictionary $\boldsymbol{\Psi}_{N\times S} = [\boldsymbol{\psi}_1\boldsymbol{\psi}_2\ldots\boldsymbol{\psi}_S]$, where each $\boldsymbol{\psi}_i, i = 1\ldots S$ is a column vector of length N and corresponds to one basis vector (or atom) of $\boldsymbol{\Psi}$. Therefore, there are S atoms in the dictionary. Vector $\boldsymbol{f}$ can be represented as a linear combination of the atoms of $\boldsymbol{\Psi}$: $\boldsymbol{f} = \sum_{j=1}^{S}\alpha_j\boldsymbol{\psi}_j$, where the α_j's are the weights for each dictionary atom $\boldsymbol{\psi}_j$. Alternatively, we can represent it in vector matrix form as follows: $\boldsymbol{f} = \boldsymbol{\Psi\alpha}$, where $\boldsymbol{\alpha}_{S\times 1}$ is a column vector containing the coefficients α_j. Vector $\boldsymbol{f}$ is said to be K-sparse in basis $\boldsymbol{\Psi}$ if only K coefficients (α_j) are non-zero, where $K \ll N$ (i.e. it takes very few atoms to perfectly represent the original signal in that basis).

CS exploits the fact that most signals have a sparse representation in some basis to directly acquire a small number of samples M ($M \ll N$), roughly in proportion to the information rate (the sparsity, K) of $\boldsymbol{f}$. This acquisition is done through a measurement matrix $\boldsymbol{\Phi}_{M\times N}$ applied to $\boldsymbol{f}$, yielding a measurement vector $\boldsymbol{y}_{M\times 1}$ as follows:

$$\boldsymbol{y} = \boldsymbol{\Phi f} = \boldsymbol{\Phi\Psi\alpha}. \tag{1}$$

The number of measurements M is linked to the sparsity K of the signal but also to the degree of coherence between $\boldsymbol{\Phi}$ and $\boldsymbol{\Psi}$, where coherence is measured by

$$\mu(\Phi,\Psi) = \sqrt{N}\cdot\max_{1\le l,j\le N}|\langle\phi_l,\psi_j\rangle|. \tag{2}$$

The number of required measurements M is then bounded by

$$M \ge C\cdot\mu^2(\Phi,\Psi)\cdot S\cdot\log N, \tag{3}$$

where C is a positive constant [14]. What this means is that it is beneficial to pick a matrix $\boldsymbol{\Phi}$ which is maximally incoherent with $\boldsymbol{\Psi}$. In an ideal scenario, it would be best to pick a matrix $\boldsymbol{\Phi}$ non-adaptively, i.e. it does not depend on the sparsifying matrix $\boldsymbol{\Psi}$. It can be shown that random matrices are maximally incoherent with any sparsifying basis with overwhelming probability [15].

Typically, CS uses a matrix built with entries that are independent and identically distributed Gaussian random variables. However, such a matrix is not convenient for WBSN because it is not easy to generate on cheap hardware and is computationally intensive to construct. Instead, we pick a sub-optimal matrix: a sparse binary sensing matrix. This matrix contains d 1's in each column (where d is determined experimentally, and the location of the 1's is chosen randomly), and the rest of the entries are zeros. While it is not possible to prove that such a matrix is maximally incoherent with any sparsifying basis, it can be shown experimentally that the degradation in performance (as compared to using an optimal matrix) is negligible, while the computational gains are important [16]. In this work, we selected $d = 8$, as it was verified experimentally that 1) this matrix behaves very closely to an optimal matrix, and 2) using a higher number of nonzero elements does not lead to better reconstruction.

Because $M \ll N$, (1) is largely underdetermined, implying that there is an infinite number of solutions for $\boldsymbol{f}$ (or $\boldsymbol{\alpha}$, equivalently). However, because we know that the signal to recover is sparse, the correct solution is often the sparsest one. This can be formulated as the following optimization problem:

$$\text{argmin}_{\boldsymbol{\alpha}} \|\boldsymbol{\alpha}\|_0 \quad \text{such that } \boldsymbol{y} = \boldsymbol{\Phi}\boldsymbol{\Psi}\boldsymbol{\alpha} \tag{4}$$

where $\|\cdot\|_0$ is the seminorm which counts the number of nonzero entries in a given vector. However, (4) is known to be NP-hard. Instead, we replace the ℓ_0 norm by the ℓ_1 norm. This problem was shown to be equivalent to (4) under some conditions (refer to [6]), and it has the advantage of being convex (and thus tractable). In this work, we chose to use the SPGL1 algorithm, for its convergence guarantees and since it requires fewer measurements than greedy algorithms to achieve perfect reconstruction (which implies higher compression) [17].

To pick $\boldsymbol{\Psi}$, we rely on the literature. In a previous work, EEG signals were shown to be sparse in a redundant Gabor dictionary [7]. Therefore, an optimal, stochastic Gabor dictionary was built as per [18] to be used as the sparsifying basis. Each atom has length $N = 512$ (4 seconds of data sampled at 128Hz).

2.3. Data Used

We randomly selected 50 epochs of 512 samples from dataset # 1 of BCI Competition IV. This data was recorded from healthy adults performing a motor imagery task [19]. We selected 12 channels in the sensorimotor area of the cortex: F1, FZ, F2, FC3, FC1, FCZ, FC2, FC4, C3, C1, CZ, C2.

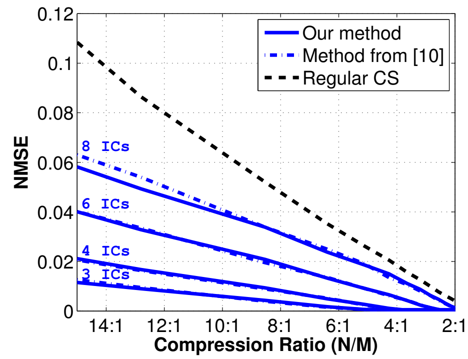

Fig. 1: Normalized mean square error (NMSE) as a function of the compression ratio for regular compressed sensing (CS), state-of-the-art from [10], and our method.

3. RESULTS

3.1. Experiment 1: Compression and Reconstruction Performance

In this experiment, we wish to test the compression and reconstruction performance of 1) CS alone (i.e. the raw EEG data is processed by CS only), 2) our method (i.e. the raw EEG data is first preprocessed by xFICA before applying CS), and 3) the method presented in [10] (i.e. the raw EEG data is first preprocessed by SOBI before applying CS). We thus vary the number of independent components retained as well as the number of measurements M and we compute the reconstruction error, in terms of the normalized mean square error (NMSE), against the compression ratio (defined as CR $= N/M$) where N is the number of raw EEG samples in each epoch (512 in our case). To keep the comparison fair, we must include the mixing matrix entries in the total number of measurements when ICA is used, and we must ensure that the same total number of measurements is used for all methods (thus ensuring a constant compression ratio).

The results are shown in Figure 1. In Figure 2, we show a slice from Figure 1 by selecting a compression ratio of 5:1 and showing the reconstruction error for each block when we vary the number of independent components.

3.2. Experiment 2: Energy Analysis

We now wish to compare the energy performance of the state-of-the-art ICA used in [10] and xFICA. To do so, we modify the numerical complexity models of [20] and then calculate

Table 1: FLOPS comparison between xFICA and [10]

Number of ICs	IOC (xFICA)	IOC ([10])	FLOPS (xFICA)	FLOPS ([10])	% FLOPS saved
3	22.36	3.56	$\mathbf{2.18 \cdot 10^5}$	$3.82 \cdot 10^6$	94.29%
4	38.26	4.62	$\mathbf{6.23 \cdot 10^5}$	$3.90 \cdot 10^6$	84.02%
6	77.74	6.42	$\mathbf{2.97 \cdot 10^6}$	$4.25 \cdot 10^6$	30.28%
8	118.68	7.63	$8.40 \cdot 10^6$	$\mathbf{4.98 \cdot 10^6}$	-68.45%

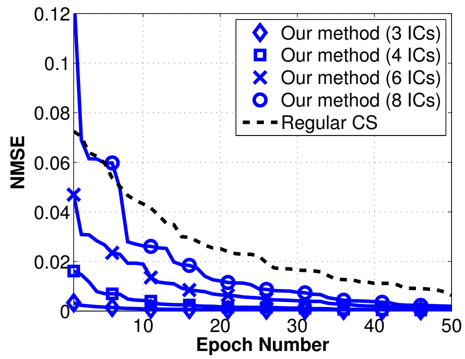

Fig. 2: NMSE for the 50 epochs (in decreasing order of magnitude) when the compression ratio is 5:1.

the number of FLOPS (F) required for both algorithms:

$$\begin{aligned}
\mathrm{F}_{\mathrm{SOBI}} &= DNC^2/2 + 4C^3/3 + (D-1)C^3/2 + \mathrm{IOC}_{\mathrm{SOBI}} \cdot \\
&\quad I^2[4I(D-1) + 17(D-1) + 4I + 75]/2 \\
\mathrm{F}_{\mathrm{xFICA}} &= NC^2/2 + 4C^3/3 + ICN + \mathrm{IOC}_{\mathrm{xFICA}} \cdot \\
&\quad [2(I-1)(I-1+N) + 5N(I-1)^2/2] + \\
&\quad I + I(I-1)^3
\end{aligned}$$

where N is the epoch length (512), C the number of channels (12), I the number of independent components used, D the number of delay lags used for SOBI (we used $D = 100$, as recommended by the authors of [21]), and $\mathrm{IOC}_{\{\mathrm{SOBI,\ xFICA}\}}$ are the number of iterations for convergence of each algorithm, respectively. To find the IOC values of both algorithms, we took the mean number of iterations over the 50 signals. The results are shown in Table 1.

4. DISCUSSION

As can be seen from Figure 1, adding an ICA preprocessing step decreases the NMSE for a fixed compression ratio (or, alternatively, it allows for an increase in the compression ratio for a fixed RMSE), with our method and the state-of-the-art yielding statistically similar results. In Figure 2, we can see that apart for two epochs for eight ICs, using an ICA preprocessing step systematically yields better results than using CS alone. Of course, selecting fewer independent components has advantages both in terms of compression ratio and speed (since fewer sources need to be reconstructed). Our experiments also suggest that even a small number of ICs (e.g. three or four) are able to accurately represent the original signal. However, one must be careful to keep enough ICs so that the variance of the data is preserved. Indeed, keeping less than three ICs makes it impossible to reconstruct the original signal faithfully.

Table 1 demonstrates that when the number of ICs is six or less, our proposed xFICA requires less FLOPS than the ICA used in [10], with the savings being more important as the number of ICs decreases. This measure is analogous to energy consumption (it is roughly proportional) since the number of FLOPS required is a measure of the dynamic power consumption. Reducing that value has the effect of reducing the overall power consumption.

5. CONCLUSION

This paper addresses the problem of efficiently compressing EEG signals in wireless body sensor network applications, where efficiency is measured in terms of compression ratio, reconstruction accuracy, and energy consumption. It proposes the use of compressed sensing after preprocessing the raw data with an energy-efficient independent component analysis method. It was demonstrated that our system provides significant energy savings as compared to the state-of-the-art method, which also uses an ICA preprocessing block. Depending on the number of independent components selected, our system can achieve up to 94% fewer computations than the state-of-the-art. As well, for a fixed compression ratio, our system achieves similar RMSE performance as the state-of-the-art method, which is much better than that achieved by CS only. In the future, it would be interesting to explore more powerful reconstruction algorithms that would fully exploit the intra- and inter-correlation in EEG signals, so as to further increase the compression ratio.

6. ACKNOWLEDGEMENTS

The authors would like to thank Dr. Amit Acharyya for his insights and guidance regarding the cross-product based ICA.

7. REFERENCES

[1] H. Cao, V. Leung, C. Chow, and H. Chan, "Enabling technologies for wireless body area networks: A survey and outlook," *Communications Magazine, IEEE*, vol. 47, no. 12, pp. 84 –93, dec. 2009.

[2] A. Milenković, C. Otto, and E. Jovanov, "Wireless sensor networks for personal health monitoring: Issues and an implementation," *Computer communications*, vol. 29, no. 13, pp. 2521–2533, 2006.

[3] H.R. Mohseni, A. Maghsoudi, and M.B. Shamsollahi, "Seizure detection in eeg signals: A comparison of different approaches," in *Conf Proc IEEE Eng Med Biol Soc*, 2006, pp. 6724–7.

[4] I. Feinberg, R.L. Koresko, and N. Heller, "Eeg sleep patterns as a function of normal and pathological aging in man," *Journal of psychiatric research*, vol. 5, no. 2, pp. 107–144, 1967.

[5] B. Blankertz, G. Dornhege, M. Krauledat, K.R. Muller, V. Kunzmann, F. Losch, and G. Curio, "The berlin brain-computer interface: Eeg-based communication without subject training," *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, vol. 14, no. 2, pp. 147–152, 2006.

[6] E.J. Candes and T. Tao, "Decoding by linear programming," *Information Theory, IEEE Transactions on*, vol. 51, no. 12, pp. 4203 – 4215, dec. 2005.

[7] S. Aviyente, "Compressed sensing framework for eeg compression," in *Statistical Signal Processing, 2007. SSP '07. IEEE/SP 14th Workshop on*, aug. 2007, pp. 181 –184.

[8] A.M. Abdulghani, A.J. Casson, and E. Rodriguez-Villegas, "Compressive sensing scalp eeg signals: implementations and practical performance," *Medical and Biological Engineering and Computing*, pp. 1–9, 2011.

[9] A.M. Abdulghani, A.J. Casson, and E. Rodriguez-Villegas, "Quantifying the performance of compressive sensing on scalp eeg signals," in *Applied Sciences in Biomedical and Communication Technologies (ISABEL), 2010 3rd International Symposium on*, nov. 2010, pp. 1 –5.

[10] B. Mijovic, V. Matic, M. De Vos, and S. Van Huffel, "Independent component analysis as a preprocessing step for data compression of neonatal eeg," in *Engineering in Medicine and Biology Society,EMBC, 2011 Annual International Conference of the IEEE*, 30 2011-sept. 3 2011, pp. 7316 –7319.

[11] A. Hyvarinen, "Fast and robust fixed-point algorithms for independent component analysis," *Neural Networks, IEEE Transactions on*, vol. 10, no. 3, pp. 626–634, 1999.

[12] S. Makeig, A.J. Bell, T.P. Jung, T.J. Sejnowski, et al., "Independent component analysis of electroencephalographic data," *Advances in neural information processing systems*, pp. 145–151, 1996.

[13] A. Acharyya, K. Maharatna, and B.M. Al-Hashimi, "Algorithm and architecture for n-d vector cross-product computation," *Signal Processing, IEEE Transactions on*, vol. 59, no. 2, pp. 812 –826, feb. 2011.

[14] E.J. Candes and M.B. Wakin, "An introduction to compressive sampling," *Signal Processing Magazine, IEEE*, vol. 25, no. 2, pp. 21 –30, march 2008.

[15] D.L. Donoho, "Compressed sensing," *Information Theory, IEEE Transactions on*, vol. 52, no. 4, pp. 1289 – 1306, april 2006.

[16] A. Gilbert and P. Indyk, "Sparse recovery using sparse matrices," *Proceedings of the IEEE*, vol. 98, no. 6, pp. 937–947, 2010.

[17] E. van den Berg and M.P. Friedlander, "Probing the pareto frontier for basis pursuit solutions," *SIAM Journal on Scientific Computing*, vol. 31, no. 2, pp. 890–912, 2008.

[18] M. Barwiński, "Product-based metric for gabor functions and its implications for the matching pursuit analysis," M.S. thesis, University of Warsaw, Warsaw, Poland, 2004.

[19] B. Blankertz, G. Dornhege, M. Krauledat, K. Mller, and G. Curio, "The non-invasive berlin braincomputer interface: Fast acquisition of effective performance in untrained subjects," *NeuroImage*, vol. 37, no. 2, pp. 539 – 550, 2007.

[20] A. Kachenoura, L. Albera, L. Senhadji, and P. Comon, "Ica: a potential tool for bci systems," *Signal Processing Magazine, IEEE*, vol. 25, no. 1, pp. 57 –68, 2008.

[21] A. Belouchrani, K. Abed-Meraim, JF Cardoso, and E. Moulines, "Second-order blind separation of temporally correlated sources," in *Proc. Int. Conf. Digital Signal Processing*. Citeseer, 1993, pp. 346–351.