Structured Sampling and Recovery of iEEG Signals

Luca Baldassarre[1], Cosimo Aprile[1,2], Mahsa Shoaran[3], Yusuf Leblebici[2] and Volkan Cevher[1]

[1]Laboratory for Information and Inference Systems (LIONS), EPFL, Lausanne, Switzerland
[2]Microelectronic Systems Laboratory (LSM), EPFL, Lausanne, Switzerland
[3]Department of Electrical Engineering, California Institute of Technology, Pasadena, CA, USA

Abstract—Wireless implantable devices capable of monitoring the electrical activity of the brain are becoming an important tool for understanding, and potentially treating, mental diseases such as epilepsy and depression. Compressive sensing (CS) is emerging as a promising approach to directly acquire compressed signals, allowing to reduce the power consumption associated with data transmission. To this end, we propose an efficient CS scheme which exploits the structure of the intracranial EEG signals, both in sampling and recovery. Our structure-aware approach is conceptually simple to implement in hardware and yields state-of-the-art compression rates up to 32x with high reconstruction quality, as illustrated on two human iEEG datasets.

I. INTRODUCTION

According to the US National Institute of Neurological Disorders and Stroke, more than 50 million people worldwide [1] are affected by epilepsy with 25% of them being pharmaco-resistant. Since 1997, the usage of prostheses has been approved to provide medical treatments for epileptic patients and in 2005 for people with severe depression [2]. However, wireless implantable devices capable of monitoring the brain activity face the challenge of limited area and power resources.

In order to reduce the power consumption of data transmission, many recent approaches (e.g., [3]–[5] and references therein) exploit compressive sensing (CS) [6], [7] to directly sample a compressed version of the signal. In a nutshell, CS consists in taking fewer linear samples than dictated by the Shannon-Nyquist theorem, while still allowing for robust offline signal reconstruction. This is possible by exploiting the fact that the information content of a signal is often much lower than its raw data content.

In this work, we propose an efficient CS sampling scheme adapted to the structure of intracranial electroencephalogram (iEEG) signals which is conceptually simple to implement within a small circuit area. The method is based on the new CS theory developed in [8], which explains the fact that sampling schemes aware of the signal's structure perform much better in practice than the standard random Gaussian or Bernoulli schemes. Our approach consists in taking random components of the Hadamard transform of the input signal, where the randomness is controlled by a probability function that favors lower frequencies.

Furthermore, we also exploit signal structure in the reconstruction phase, by imposing that the wavelet coefficients of the reconstructed signal adhere to the hierarchical configuration often found in natural signals [9]. The ensuing optimization problem is still very efficient to solve, thanks to the fast

computation of the Hadamard transform, the tractability of the structure-promoting function [10] and the convergence rate of the primal-dual optimization algorithm of [11].

In a few words, we reap the benefits of both structured sampling and structured recovery to yield state-of-the-art compression of up to 32x, while maintaining a high signal reconstruction quality as quantitatively demonstrated on two clinical iEEG datasets.

This paper is organized as follows. In Section II, we introduce the main concepts of compressive sensing and describe the structured sampling scheme, while structured recovery and optimization details are provided in Section III. Numerical experiments are reported in Section IV, while Section V concludes the paper with a discussion on the trade-offs between hardware requirements and signal reconstruction quality.

II. COMPRESSIVE SENSING AND STRUCTURED SAMPLING

The main tenet of compressive sensing states that a signal $\mathbf{x} \in \mathbb{R}^n$ which has K non-zero coefficients can be robustly recovered from only $m = \mathcal{O}(K \log \frac{n}{K})$ samples $\mathbf{y} \in \mathbb{R}^m$,

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{w}\,, \tag{1}$$

where $\mathbf{A}$ is a linear operator that either satisfies the Restricted Isometry Property (RIP) or is incoherent [12], and $\mathbf{w}$ accounts for measurement noise. Obviously, $\mathbf{y}$ offers a compressed version of $\mathbf{x}$. If we are able to directly sample $\mathbf{y}$, we save both on storage and communication power. Recovering $\mathbf{x}$, though, requires to solve a non-linear optimization problem. Nonetheless, recent advances in optimization have provided efficient algorithms that can scale to very large signals [11].

Theoretically, i.i.d. sub-Gaussian matrices are incoherent and also satisfy the RIP. Furthermore, they are universal, i.e., given an ortho-normal basis $\mathbf{\Phi}$ which allows for a sparser representation of a signal $\mathbf{x}$, the RIP or the incoherence of $\mathbf{A}\mathbf{\Phi}$ is the same as of the original $\mathbf{A}$ [12]. However, sub-Gaussian matrices are prohibitively expensive to use in practice, since they require $\mathcal{O}(mn)$ space and time.

More efficient types of sampling are being successfully used in real applications, such as subsampled fast transforms, like the Fast Fourier (FFT), the Discrete Cosine (DCT) or the Fast Walsh-Hadamard (FWHT) Transforms, which can be computed in $\mathcal{O}(n \log n)$ time. Despite not being universal, it has been recently shown that they are asymptotically incoherent [8] with some bases such as wavelets, see Figure 1 (left). Furthermore, via multi-level sampling they can easily adapt to the asymptotic sparsity or structure present in many natural signals [8], when represented in sparsifying bases such as wavelets [9].
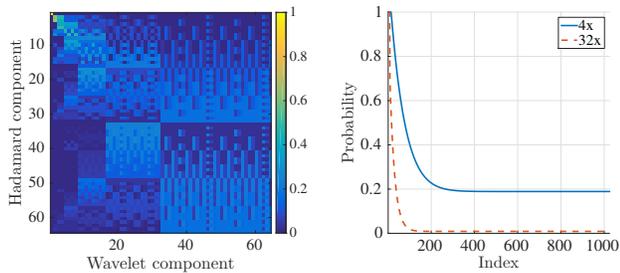
Fig. 1. (left) Coherence between the Hadamard and the Wavelet bases. The coherence decreases for higher frequencies (higher coefficients). (right) Probability functions used for sampling the indices of the Fast Walsh-Hadamard Transform for 4x and 32x compression factors.

For the purpose of circuit implementation, the FWHT has the advantage of only requiring binary operations, and has already been used for parallel sampling of multi-channel data [13]. Following [8], we sample the indices of the FWHT according to a probability function which favors the low frequencies of the signal, which carry most of its energy. The probability function is adapted to the compression factor, so that the lowest frequencies are always sampled, while still allowing to sample some of the higher frequencies, albeit with fast decreasing probability (see Figure 1 (right)). We name this approach Structured Hadamard Sampling (SHS).

III. STRUCTURED RECOVERY AND OPTIMIZATION

Additional structure in the signal $\mathbf{x}$, such as interdependencies between its non-zero coefficients or constraints on its support, allows to reduce the number of samples required for exact or stable recovery (see for instance [14] and [15]). Many of these structures can be encoded via linear inequalities that admit tight and tractable convex relaxations [16]. Interestingly, natural signals are often characterized by sparse and structured representations in time-frequency (or space-frequency) domains, such as provided by wavelets [9].

In order to reconstruct the original signal $\mathbf{x}$ from its compressive samples $\mathbf{y}$, we therefore resort to solving the following optimization problem on the wavelet coefficients $\boldsymbol{\alpha}$,

$$\begin{array}{cl} \underset{\boldsymbol{\alpha} \in \mathcal{A}}{\text{minimize}} & f(\boldsymbol{\alpha}) \\ \text{subject to} & \mathbf{A}\boldsymbol{\Phi}\boldsymbol{\alpha} - \mathbf{y} \in \mathcal{K} \end{array} \quad (2)$$

where f is the Gauge function that promotes the structure we expect in $\boldsymbol{\alpha}$, $\mathcal{K}$ encodes our information about the noise $\mathbf{w}$ in (1) and $\mathcal{A}$ is a constraint set that specifies further assumptions about the signal, e.g. boundedness. We reconstruct the signal as $\widehat{\mathbf{x}} = \boldsymbol{\Phi}\widetilde{\boldsymbol{\alpha}}$, where $\widetilde{\boldsymbol{\alpha}}$ is the solution to (2).

For sparse signals, it is common to use the ℓ_1 norm, $\|\mathbf{x}\|_1 := \sum_{i=1}^n |x_i|$, leading to the Basis Pursuit (BP) optimization problem.

It is well-known that biological signals are not only sparse in the wavelet domain, but their wavelet coefficients can be naturally arranged on a dyadic tree with the coefficients decaying from root to leaves, as shown in Figure 2 (top-left). This type of structure can be promoted by a tree regularizer that gradually penalizes the coefficients closer to the leaves. In order to do so, we define a group structure $\mathcal{T} = \{\mathcal{G}_1, \ldots, \mathcal{G}_n\}$
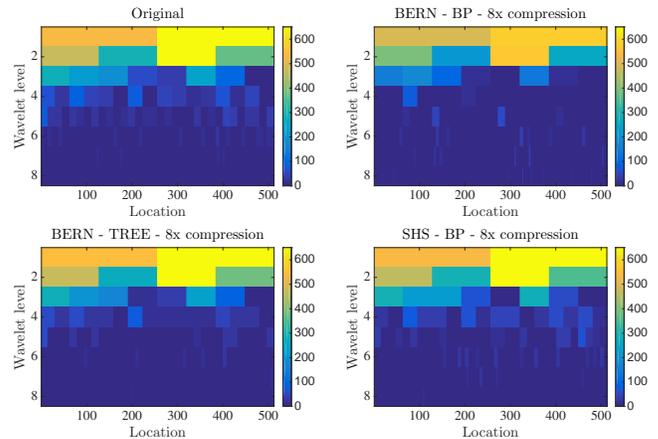

Fig. 2. Tree structure in one signal from `iEEG.org` dataset `I001 P034 D01` (channel 6, first annotated seizure, first 1024 samples window) and in three reconstructions obtained via Bernoulli sampling (BERN) and structured Hadamard sampling (SHS). The tree structure can be enforced via a specific tree regularizer or mostly captured via structured sampling.
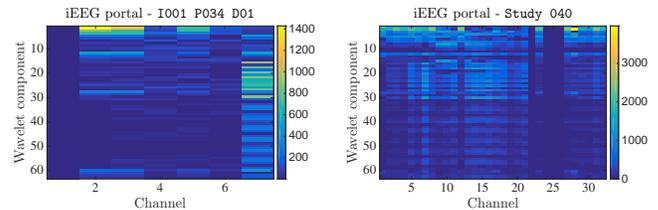

Fig. 3. First 64 Wavelet coefficients of the micro-electrode signals from two datasets from the `iEEG.org` portal. (left) 7 channels from dataset `I001 P034 D01`. (right) 32 channels from dataset `Study 040`. The group structure is evident among the correlated channels in both datasets, however, there remain outlier channels which do not abide to the group structure.

where each group $\mathcal{G}_i \subseteq \{1, \ldots, n\}$ contains the node i in the tree and all its descendants. Let $\mathbf{x}_{|\mathcal{G}}$ be the restriction of the vector $\mathbf{x}$ to the coefficients indexed by $\mathcal{G}$. The tree norm is then defined as $\|\mathbf{x}\|_{\mathcal{T}} := \sum_{\mathcal{G} \in \mathcal{T}} \|\mathbf{x}_{|\mathcal{G}}\|$ [10]. Given a sampling strategy that is not aware of the signal structure (e.g., MCS, discussed later), in Figure 2, we show how this structure emerges in the reconstructed signals when using the tree norm, but not when using the ℓ_1 norm. In the same figure, it can also be noted that when using SHS, the hierarchical structure emerges even if it is not imposed during reconstruction, because it is already mostly captured during sampling.

When the micro-electrodes are very close to each other, due to the high correlation among the signals, their time-frequency domain coefficients tend to be group sparse. That is, when a certain coefficient is zero for a signal, it is likely to be zero also for the correlated signals and vice-versa. Let $\mathbf{X} \in \mathbb{R}^{N \times n}$ be the signal matrix, each row containing the signal for one of the N channels. In order to promote group-sparsity, [5] proposed to use the $\ell_{2,1}$ mixed norm, $\|\mathbf{X}\|_{2,1} := \sum_{i=1}^n \sqrt{\sum_{j=1}^N X_{i,j}^2}$. In Figure 3 (top), we report the first 64 wavelet coefficients for the signals from two datasets, exhibiting the group structure among correlated channels.

A. Optimization algorithm

In order to solve the constrained convex optimization problem (2), we use the primal-dual algorithm proposed in [11],

named `DecOpt`. This algorithm is very flexible in handling different problem types, scalable and guaranteed to converge at an optimal rate [11]. Its iterations require to compute the proximity operator of f and to apply $\mathbf{A\Phi}$ or its adjoint, which for the FWHT and wavelet transform require only $\mathcal{O}(n \log n)$ time. The proximity operator of f is defined as

$$\text{prox}_f(\mathbf{z}) = \underset{\mathbf{x} \in \mathbb{R}^n}{\text{argmin}} \frac{1}{2} \|\mathbf{x} - \mathbf{z}\|_2^2 + f(\mathbf{x}) \ .$$

The proximity operators of the ℓ_1 norm and of the $\ell_{2,1}$ mixed norm can be computed in closed form via soft-thresholding or group soft-thresholding, while the proximity operator for the tree norm can be computed in a finite number of steps via an active set algorithm [10]. In practice, there is almost no computational difference between the three approaches, thus we can take advantage of additional structure at almost no increase in computational cost.

IV. NUMERICAL EXPERIMENTS

We demonstrate the advantages of structured sampling and structured recovery on two datasets of clinical iEEG signals obtained from the iEEG portal[1]. The first is dataset `I001 P034 D01` and contains more than 1 day of iEEG recordings. We selected the first annotated seizure, which comprises 161475 samples for each of the 7 micro-electrodes. We split the signals in 157 non-overlapping windows of 1024 samples each. During this seizure, the first channel was inactive and the seventh was picking up the AC signal, while channels 2 to 6 were recording normally and are lightly correlated, as shown in Figure 4. The second dataset is from `Study 040` and contains recordings from 32 micro-electrodes, which are highly correlated, even though 6 channels are not properly recording, see Figure 5. We used 20 non-overlapping windows of size 1024 samples from the first annotated seizure.

We compare structured Hadamard sampling (SHS) applied to each channel independently, with the same subsampling for all the channels, to two other sampling approaches. The first, named BERN, uses the same random Bernoulli $\{\pm 1\}$ matrix to sample each channel independently [4]. The second, named Multi-Channel Sampling (MCS) [5], designed to be highly power-efficient, uses a Bernoulli $\{0, 1\}$ matrix to sample across the channels at each time step. The compression achievable by this method depends on the number of samples taken at each time step, with a minimum of one sample, yielding a compression factor equal to the number of channels. SHS and BERN sampling strategies are limited only by the length of the considered time window.

We also compare the three structured recovery methods described in the previous section. Namely, Basis Pursuit (BP) using the ℓ_1 norm, L2L1 using the $\ell_{2,1}$ mixed norm and the TREE method which uses the tree norm, $\|\mathbf{x}\|_\mathcal{T}$. As sparsifying basis, we use the Daubechies-4 Wavelet basis as provided by the Rice Wavelet Toolbox[2]. We pursue the following protocol for the experiments:

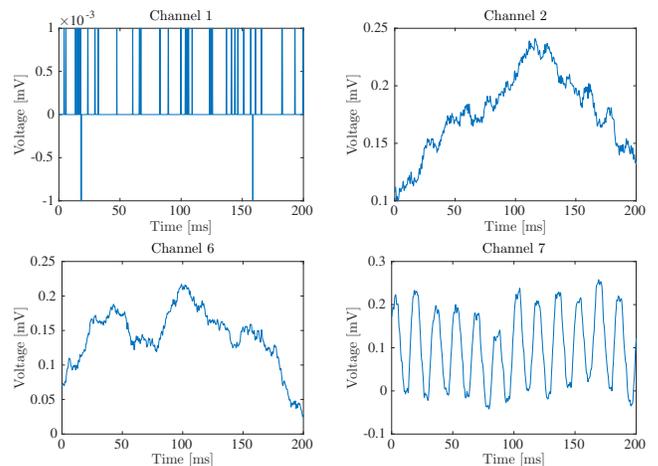1) Sample all channels in each window according to the sampling method chosen: MCS, SHS or BERN.

Fig. 4. Example of micro-electrode signals from `iEEG.org` dataset `I001 P034 D01` (first seizure, first 1024 samples window). Channel 1 is inactive, since it simply jumps between $-1\mu V$ and $1\mu V$. Channel 2 to 6 record normal activity, which is not much correlated. Channel 7 exhibits strong AC components, possibly picked up from the power sources.

2) Reconstruct using `DecOpt` [11] via BP, L2L1 or TREE.
3) Compute the signal-to-noise ratio (SNR) of the reconstructed signals.
4) Average over 20 different randomizations of the sampling scheme.

Tables I and II report the results on the first dataset averaged over 157 windows and channels 2 to 6. A posteriori, we excluded channels 1 and 7 from the analysis of the performance because these channels are either inactive or not recording the neurological signal. An advantage of channel-wise sampling (SHS or BERN) against the MCS sampling is that the former does not suffer from mixing "noisy" channels with "clean" ones. In an embedded system, a sub-circuit may be required for MCS in order to detect if a channel is recording properly. Furthermore, the compression factor of MCS is limited by the number of channels, while SHS and BERN can achieve much higher compression rates. However, only SHS yields an acceptable reconstruction performance above 16dB at $32\times$ compression, with 10dB considered as the minimum required performance in order to retain diagnostically relevant information [17].

Table III contains the reconstruction SNR for the second dataset. In this case, the SNRs for all methods are very high, with an advantage for SHS with structured recovery, either L2L1 or TREE. The high SNRs can be explained by the fact that the signals in this dataset are quite regular, whose wavelet coefficients are then very sparse, therefore requiring much fewer linear measurements for robust recovery.

The running time of the optimization algorithm is, in general, less than 10 seconds per time window for recovering the signals from all the channels simultaneously on an Intel Xeon E5-2630 @ 2.40GHz.

V. DISCUSSION

The numerical results show that structured sampling, in combination with structured recovery, allows to more faithfully

---

[1] www.ieeg.org
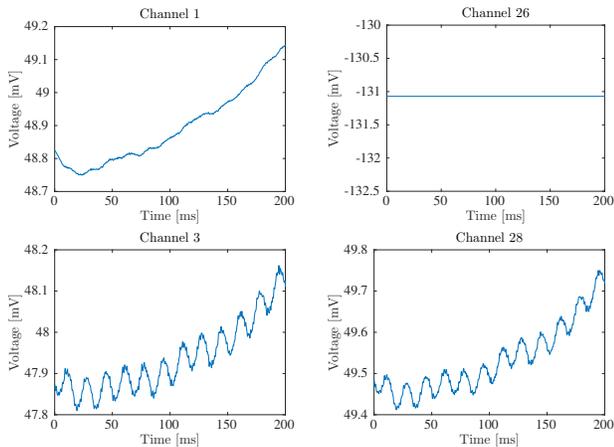[2] http://dsp.rice.edu/software/rice-wavelet-toolbox

Fig. 5. Example of micro-electrode signals from iEEG.org dataset Study 040 (first seizure, first 1024 samples window). Channel 26 seems completely inactive, it sends a constant signal of approximately $-131mV$. Channels 3 and 28, among others, are highly correlated. Channel 1 is an example of a channel which does not exhibit the smaller oscillations of channels 3 and 28.

TABLE I. IEEG.ORG PORTAL DATASET I001 P034 D01. MEAN SNR OVER CHANNELS 2-6

Sampling	Recovery	Compression factor			
		7/4	7/3	7/2	7
	BP	29.3	25.7	21.4	15.2
MCS	L2L1	31.3	28.1	24.3	17.3
	TREE	34.1	30.6	26.8	21.2

TABLE II. IEEG.ORG PORTAL DATASET I001 P034 D01. MEAN SNR OVER CHANNELS 2-6

Sampling	Recovery	Compression factor				
		2	4	8	16	32
SHS	BP	34.1	27.6	23.7	21.0	16.7
	L2L1	35.3	28.4	24.0	21.2	16.8
(this work)	TREE	35.6	28.8	24.6	22.2	17.6
BERN	BP	33.1	24.4	16.7	10.8	5.7
	L2L1	35.8	27.3	18.5	11.8	6.6
	TREE	36.9	29.5	23.0	17.8	13.5

TABLE III. IEEG.ORG PORTAL DATASET STUDY 040. MEAN SNR OVER ALL CHANNELS.

Sampling	Recovery	Compression factor				
		2	4	8	16	32
MCS	BP	79.6	72.4	66.5	61.2	55.6
	L2L1	77.4	71.6	64.6	60.4	58.9
	TREE	86.6	77.4	70.8	59.6	12.3
SHS	BP	91.1	84.0	80.0	77.5	73.8
	L2L1	93.6	86.3	82.1	79.5	74.8
(this work)	TREE	92.8	85.3	81.3	78.7	74.7
BERN	BP	104	85.9	70.0	63.4	60.9
	L2L1	73.4	69.4	63.7	59.4	57.1
	TREE	83.3	76.4	61.7	54.3	32.4

reconstruct the original signals, as compared with the traditional Bernoulli [4] or the multi-channel sampling [5] schemes.

However, when evaluating a sampling scheme for a wireless implantable device, it is necessary to consider all the performance factors. For example, the MCS strategy has been designed to optimize area and power usage, at the cost of sacrificing the reconstruction quality when the channels are not correlated enough. The random Bernoulli sampling [4] offers excellent reconstruction quality at low compression factors, but requires a larger chip area than MCS.

More studies are required to find the best trade-off between area usage, power consumption and reconstruction fidelity.

The proposed structured Hadamard sampling seems to offer a very promising alternative to the current state-of-the-art. In the future, we will focus on circuit implementation and analysis of such a low-power and compact solution for neural recording. Furthermore, we are currently studying how to learn the best sampling pattern from previously collected signals.

REFERENCES

[1] M. Leonardi and T. B. Ustun, "The global burden of epilepsy," *Epilepsia*, vol. 43, pp. 21–25, 2002.

[2] C. B. Nemeroff, H. S. Mayberg, S. E. Krahl, J. McNamara, A. Frazer, T. R. Henry, M. S. George, D. S. Charney, and S. K. Brannan, "Vns therapy in treatment-resistant depression: clinical evidence and putative neurobiological mechanisms," *Neuropsychopharmacology*, vol. 31, no. 7, pp. 1345–1355, 2006.

[3] J. N. Laska, S. Kirolos, M. F. Duarte, T. S. Ragheb, R. G. Baraniuk, and Y. Massoud, "Theory and implementation of an analog-to-information converter using random demodulation," in *IEEE International Symposium on Circuits and Systems*, 2007, pp. 1959–1962.

[4] F. Chen, A. P. Chandrakasan, and V. M. Stojanovic, "Design and analysis of a hardware-efficient compressed sensing architecture for data compression in wireless sensors," *IEEE Journal of Solid-State Circuits*, vol. 47, no. 3, pp. 744–756, 2012.

[5] M. Shoaran, M. H. Kamal, C. Pollo, P. Vandergheynst, and A. Schmid, "Compact low-power cortical recording architecture for compressive multichannel data acquisition," *IEEE Transactions on Biomedical Circuits and Systems*, vol. 8, no. 6, pp. 857–870, December 2014.

[6] E. J. Candès, "Compressive sampling," in *Proceedings oh the International Congress of Mathematicians: Madrid, August 22-30, 2006: invited lectures*, 2006, pp. 1433–1452.

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

[8] B. Adcock, A. C. Hansen, C. Poon, and B. Roman, "Breaking the coherence barrier: A new theory for compressed sensing," *arXiv preprint:1302.0561*, 2013.

[9] S. Mallat, *A wavelet tour of signal processing*. Academic press, 1999.

[10] R. Jenatton, J. Mairal, G. Obozinski, and F. Bach, "Proximal methods for hierarchical sparse coding," *Journal of Machine Learning Reasearch*, vol. 12, pp. 2297–2334, 2011.

[11] Q. Tran-Dinh and V. Cevher, "A primal-dual algorithmic framework for constrained convex minimization," *arXiv preprint:1406.5403*, 2014.

[12] S. Foucart and H. Rauhut, *A mathematical introduction to compressive sensing*. Springer, 2013.

[13] V. Majidzadeh, A. Schmid, and Y. Leblebici, "A 16-channel, 359 μw, parallel neural recording system using walsh-hadamard coding," in *Custom Integrated Circuits Conference*, 2013.

[14] R. Baraniuk, V. Cevher, M. Duarte, , and C. Hegde, "Model-based compressive sensing," *IEEE Transactions on Information Theory*, vol. 56, no. 4, pp. 1982–2001, 2010.

[15] A. Kyrillidis, L. Baldassarre, M. El Halabi, Q. Tran-Dinh, and V. Cevher, "Structured sparsity: Discrete and convex approaches," in *Compressed Sensing and its Applications*. Springer, 2015.

[16] M. E. Halabi and V. Cevher, "A totally unimodular view of structured sparsity," in *AISTATS*, 2015.

[17] G. Higgins, S. Faul, R. P. McEvoy, B. McGinley, M. Glavin, W. P. Marnane, and E. Jones, "Eeg compression using jpeg2000: How much loss is too much?" in *International Conference of the IEEE Engineering in Medicine and Biology Society*, 2010, pp. 614–617.