Disturbance Rejection with Rakeness-based Compressed Sensing: Method and Application to Baseline/Powerline Mitigation in ECGs

Rakeness-Based Compressed Sensing of Multiple-graph Signals for IoT Applications

Mauro Mangia*, Fabio Pareschi†, Rohan Varma°, Riccardo Rovatti*, Jelena Kovačević°, Gianluca Setti†
* DEI University of Bologna, † ENDIF University of Ferrara, ° ECE Carnegie-Mellon University Pittsburgh

***Abstract*—Signals on multiple graphs may model an IoT scenarios consisting of local WSN performing sets of acquisitions that must be sent to a central hub that may be far from the measurement field. Rakeness-based design of Compressed Sensing is exploited to allow the administration of the trade off between local communication and the long range transmission needed to reach the hub. Extensive Montecarlo simulations incorporating real world figures in terms of communication consumption show a potential power saving from 25% to almost 50% with respect to a direct approach not exploiting local communication and rakeness.**

***Index Terms*—Signals on graphs, compressed sensing, rakeness, internet of things**

I. Introduction

Instead of being supported by a sequence of time instants, signals on graphs are supported by a set of vertices between which edges may be drawn and weighted to obtain a graph. More formally [1], [2], a signal x is defined on a set of n vertices V if $x : V \mapsto \mathbb{R}$. For simplicity's sake we will assume $V = \{0, 1, \dots, n-1\}$ to use vertices also as indexes when needed. The relationship between vertices is modeled by possibly weighted edges between the nodes. Tolerating a slight loss of generality in edge weighting, we may model these connection with the so called incidence matrix, i.e., by a $n \times n$ matrix A such that $A_{j,k}$ is the weight associated with the edge from k and j ($A_{j,k} = 0$ means no edge from k to j).

As a link with time-domain quantities, note that a discrete-time periodic signal of period n can be modeled by identifying the vertex j with the j-th time instant and considering $A_{j+1,j} = 1$ for $j = 0, \dots, n-2$, $A_{0,n-1} = 1$ and $A_{j,k} = 0$ otherwise, to encode the periodic sequence of time instants. In this setting the eigendecomposition of the incidence matrix is $A = UDU^{-1}$ with $U_{j,k} = e^{-2\pi \mathbf{i}\, jk/n}$ for $j, k = 0, \dots, n-1$ and D a diagonal matrix with $D_{j,j} = e^{2\pi \mathbf{i}\, j/n}$ for $j = 0, \dots, n-1$. It is then most natural to extend the label *Fourier basis* to the matrix U in the decomposition $A = UDU^{-1}$ of the incidence matrix of a generic graph supporting a signal [3], [4]. If the underlying graph is not oriented, A is symmetric and U is an orthonormal matrix.

Signal processing on graphs exploits this generalization and often assumes that the representation $\xi = U^{-1}x$ of the signal in the Fourier basis of the supporting graph has some special properties, in analogy with a time-domain signal that has some frequency-domain feature.

Signals on graphs fit into a number of scenarios where the relationship between samples is not a simple ordering in time. In unstructured frameworks, the locations at which samples are acquired imply some relationship between them (like the temperature at different spots that are thermically connected in different ways or the consumption of computers in an inhomogeneous company local network) that can be modeled by a generic graph. Moreover, the sensors themselves may belong to a Wireless Sensor Network (WSN) whose nodes have local communication capabilities (that can also be modeled by a graph) and finally deliver their acquisitions to a central hub by means of long range transmissions in some Wide Area Network (WAN).

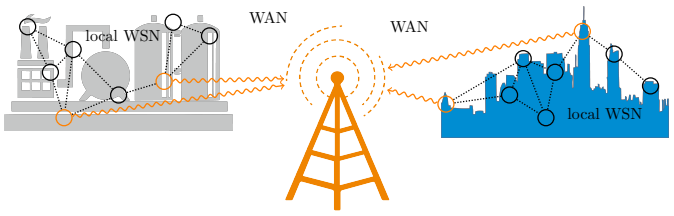

Fig. 1. A grand view of systems made of local Wireless Sensor Networks that communicate their readings to a geographically separated hub.

Figure 1 gives an intuitive representation of these structures that suggest exploring the trade-off between local communication/processing and direct transmission to the hub. For example, assuming that the ratio between the typical distance covered by long-range and short-range communications is 10^2 (tens of meters to kilometers) and that no particular directivity can be provided by sensor nodes antennas, one expects that the ratio between entailed powers is of the order of 10^4. This is matched by actual consumption of current implementations. For example, Bluetooth Low Energy modules come with energy-per-bit efficiencies in the range from 31 nJ/bit [5] to 46 nJ/bit [6] while LoRaWAN implementations exhibit energy efficiencies in the range 19 μJ/bit [7] to 220 μJ/bit [8] so that one may expect a ratio ϵ between short- and long-range efficiencies between $\epsilon_{\min} = 1.4 \times 10^{-4}$ and $\epsilon_{\max} = 2.4 \times 10^{-3}$. This is more than enough to allow substantial local data exchange before a single long-range transmission is attempted. An additional cost in the power budget is due to the needed local processing that is in general dominate by local communication cost [9], [10], [11], [12].

We address such a trade-off by exploiting a further prior that is commonly valid for real-world signals, i.e., the fact that they have non-white second-order statistics that can be modeled as a further weighted graph connecting the same vertices.

Hence, the signal is ultimately characterized by three graphs: the one representing the structure of its support, the

one describing the connectivity of the WSN acquiring it, and the one expressing its second-order statistics. From this point of view we are dealing with a *signal on multiple graphs*.

II. Acquisition of multiple graph signals

Acquisition largely benefits from priors on the signal, the most obvious example being Nyquist sampling in which frequency domain information allows us to sample signals in a subset of the time instants. Here we address the efficient acquisition of graph signals exploiting the prior that they are known to be *sparse* in their Fourier domain, i.e, that ξ has at most $\kappa \ll n$ non-zero components. The graph providing the Fourier basis will be named the *sparsity graph* of the signal.

This is the natural setting in which Compressed Sensing (CS) [13], [14] may be employed. In fact, for certain $m < n$ one may find $m \times n$ matrices S such that the measurements in the vector $y = (y_0, \dots, y_{m-1})^\top = Sx = SU\xi$ can be post-processed to yield the original x despite the fact that S (and thus SU) is rectangular. In the graph framework, the easiest case is when $y_j = x_{v_j}$ for certain vertices $v_0, \dots, v_{m-1} \in V$, i.e., when the signal is subsampled and the matrix S is made of m rows of the $n \times n$ identity matrix [15].

Instead, we consider measurements of the form $y_j = \sum_{u \in W_j} S_{j,u} x_u$ for certain $W_j \subseteq V$, assuming that one may use local communication to collect the signal values at the vertices $w \in W_j$, compute y_j and send it to the hub. This is precisely the scenario sketched in the introduction, where acquired values can be propagated locally by the WSN with an energy cost per individual communication (a *hop*) that is only ϵ-times the cost of transmitting y_j to the hub.

Usually, one cannot arbitrarily choose the vertices in W_j since, for example, they must correspond to nodes that are geometrically close. We model this with a *sampling graph* that connects two vertices of V if one of them can communicate a value to the other.

The sampling strategy is a generalization of single-vertex sampling scheme that takes into account the sampling graph constraint. To compute the j-th measurement y_j we randomly select a vertex $v_j \in V$. Assuming that the sampling graph is connected, a distance $h(v_j, u)$ is defined from every vertex $u \in V$. Given a *hop budget* H we select a subset $W_j \subseteq V$ such that $\sum_{u \in W_j} h(v_j, u) \leqslant H$. This can be effectively done by modifying the classical Dijkstra algorithm for the shortest path to a given root (v_j), so that it adds a new vertex to the tree only if there are enough hops left to go from that vertex to the root.

This is exemplified in Figure 2 where the largest red disk represents the randomly chosen root v_j and we are given a hop budget $H = 16$. The 3 nearest neighbors of v_j are included in W_j and consume a total of 1 hop each to communicate their values to the root along the red solid edges. Four nodes can connect to the nearest neighbors of the root by means of red dashed edges and thus can communicate their value with 2 hops each. Since the budget is not exhausted by these 11 hops we may add further vertices. Yet, not all the vertices that can communicate to the root in 3 hops can be accommodated. In this case, the budget allows only one node to be selected and linked to the other by a red dotted edge.

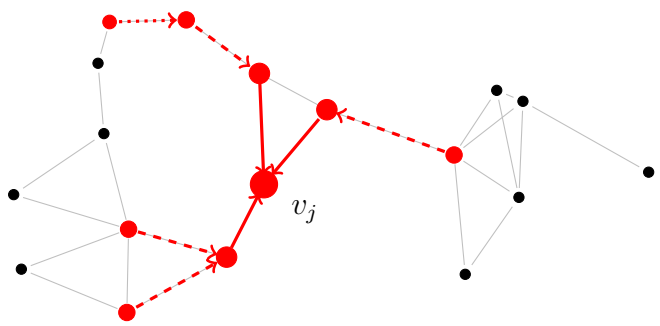

Fig. 2. A generalization of single-vertex sampling. In the graph nodes are connected only if their are closer than a certain threshold.

Once that signal values are collected, the root may combine them in multiple ways by adopting different coefficients, thus producing more than one measurement. This sample reuse saves communication costs but limits the diversity that can be exploited in computing the measurements. Hence, if we say that the same subset of samples may be used at most M times we may define $\Delta m = \lceil m/M \rceil$, and accumulate samples in independently drawn roots v_j and from independently defined neighborhoods W_j for $j = 0, \dots, \Delta m - 1$. Then, we assume $v_j = v_{j \pmod{\Delta m}}$ and $W_j = W_{j \pmod{\Delta m}}$ for $j \geqslant \Delta m$.

As far as coefficients are concerned, the most trivial, CS-inspired, option is to take each non-null entry of S to be the realization of an independent normal random variable. We will denote this classical choice as the *random* option.

III. Correlation graph and Rakeness-Based CS

Independently of their sparsity, most signals feature some sort of energy *localization* that can be detected by considering their correlation matrix $\mathcal{X} = \mathbf{E}[xx^\top]$ and verifying that its eigenvalues are not identical and, thus, there are subspaces along which most of the energy of x concentrates. Localization and sparsity are different priors since the subspaces along which energy concentrates do not need to be κ-dimensional canonical subspaces in the sparsity reference system.

It is a graph prior since the matrix $\mathcal{X}$ is a symmetric matrix that can be interpreted as the incidence matrix of a complete, graph where the edge between v' and v'' has a weight $\mathbf{E}[x_{v'} x_{v''}]$.

The exploitation of such a prior to optimize CS for time-domain signals has been investigated based on the *rakeness* concept [16]. The basic observation is that it is convenient to design the statistics of the coefficients $S_{j,u}$ such that y_j is, on the average, able to *rake* as much energy as possible from the signal.

In its simplest form, this corresponds to draw the coefficients $S_{j,u}$ corresponding to a vertex subset W_j as zero-mean Gaussian random variables with a correlation matrix $\Sigma_{|W_j}$ that solves the optimization problem

$$\max \operatorname{tr}\left(\Sigma_{|W_j} \mathcal{X}_{|W_j}\right) \quad \text{s.t.} \quad \begin{array}{l} \Sigma_{|W_j} \succ 0 \\ \Sigma_{|W_j} = \Sigma_{|W_j}^\top \\ \operatorname{tr}\left(\Sigma_{|W_j}\right) = n_j \\ \operatorname{tr}\left(\Sigma_{|W_j}^2\right) \leqslant r\, {n_j}^2 \end{array} \tag{1}$$

where, since $y_j = \sum_{u \in W_j} S_{j,u} x_u$ then $\mathrm{tr}\left(\Sigma_{|W_j} \mathcal{X}_{|W_j}\right) = \mathbf{E}[y_j^2]$ is the average energy of the resulting measurements and the first two constraints ensure that $\Sigma_{|W_j}$ is positive definite, symmetric and with a total energy proportional to the number of coefficients $n_j = |W_j|$. As far as the last constraint is concerned, note that, due to the random nature of the signal, observing only its maximum-energy component (the so-called principal component) is not enough to reconstruct it, and energy maximization should be tempered by the need to span the whole signal space. This is obtained by suitably bounding the sum of the squares of the eigenvalues of $\Sigma_{|W_j}$ to prevent them to concentrate only on the principal components [9], [16]. With the value of r suggested in [9], (1) has the analytical solution

$$\Sigma_{|W_j} = \frac{1}{2}\left(\frac{n_j \mathcal{X}_{|W_j}}{\mathrm{tr}\left(\mathcal{X}_{|W_j}\right)} + I_{n_j}\right) \tag{2}$$

where $n_j = |W_j|$ is the cardinality of W_j and I_{n_j} is the $n_j \times n_j$ identity matrix.

Hence, as a second option, instead of drawing the coefficients as random independent normals, for every measurement y_j depending on the vertices in W_j we generate random jointly-Gaussian coefficients with correlation (2).

IV. Empirical evidence

To assess the effectiveness of the proposed approach we perform a Montecarlo analysis of a few configurations. In all trials $n = 128$ while the sparsity level is taken as $\kappa \in \{6, 12, 24\}$ to explore priors with different strengths.

In each trial the *sampling graph* is a realization of a Geometric random graph with n nodes uniformly distributed in $[0,1]^2$ with connections if their distance is less than 0.15 (label Geo-0.15). Hop budgets $H \in \{64, 128, 256\}$ are considered.

The *sparsity graph* can either be the same as the *sampling graph* or the realization of one of the following random graphs for which we adopt the definitions in [18])

- Erdös-Rényi graph with probability of connection equal to 0.1 (label ER-0.1)
- Barabasi-Albert graph whose construction starts from a 10-vertices ER-0.1 and connects every new vertex to 5 previous vertices (label BA-10-5)
- Watts-Strogatz graph with 6 neighbors in the initial ring and with a rewiring probability equal to 0.3 (label WS-6-0.3)

In all cases possibly non-connected realizations are discarded.

To simulate localization, the κ non-zero components in ξ are selected with a non-uniform probability. This probability is communicated to neither the sampling mechanism nor the reconstruction algorithm. What is known by the sampling stage is only the correlation matrix $\mathcal{X} = \mathbf{E}[xx^\top]$ from which the various correlation submatrices $\mathcal{X}_{|W_j}$ are taken to compute (2).

White Gaussian noise is added to the samples giving them an Intrinsic Signal-to-Noise-Ratio $\mathrm{ISNR} = 60\,\mathrm{dB}$. Reconstruction is obtained by Basis Pursuit with De-noising (BPDN) [19] as implemented by SPGL1 [20].

Performance is evaluated as the Probability of Correct Reconstruction (PCR) defined as the probability that the relative error in the reconstruction corresponds to a loss of not more than 6 dB with respect to the ISNR, i.e., $\mathtt{PCR} = \Pr\{\|x\|/\|x-\hat{x}\| \geqslant 54\,\mathrm{dB}\}$.

The qualitative features of all the observed trends coincide. Figure 3 reports how the PCR depends on the number of measurements in three cases that correspond to $\kappa = 6, 12, 24$, i.e., to progressively weakening sparsity priors. The vertex-only option (black dotted track) is taken as a reference.

In all those plots as well as in all tested cases, the position of the continuous tracks shows that if the samples collected by local communication are combined with purely random coefficients no gain is obtained with respect to not using local communication ($H = 0$).

Local communication can be traded for long-range one only if we exploit the correlation graph by means of rakeness-based CS. An optimized choice of the coefficients leverages the availability of multiple samples to compute more informative measurements. Hence, the same reconstruction quality can be obtained at the hub even if less measurements are sent to it through long-range transmission.

This points towards a possible power saving. To quantify this, we normalize to 1 J the energy needed by a long-range transmission so that the cost of a short-range transmission gets normalized to the ratio ϵ J discussed in the Introduction. With this, the energy needed by the collection of samples and transmission of the measurements is $E_{\mathrm{CS}} = \left(m^{\mathrm{CS}} + \epsilon H \left\lceil m^{\mathrm{CS}}/M \right\rceil\right)$ J, where m^{CS} is the number of measurement needed to achieve the prescribed performance, M is the maximum number of measurement that each node can compute with the samples it collected, and H is the hop budget constraining sample collection. This compares with $E_{\mathrm{VS}} = m^{\mathrm{VS}}$J, i.e., with the energy (equal to the number of measurements) needed to achieve the same performance level by simple vertex-sampling.

Figure 4 reports the ratio $E_{\mathrm{CS}}/E_{\mathrm{VS}}$ when the desired PCR is set to 95% and in all the cases we tested in an extensive Monte-Carlo simulation.

Though it is evident that as κ increases, our framework looses its ability of allowing any real subsampling and thus power saving, rakeness-based CS is almost always able to yield substantial power saving. Actual reduction depends on the relationship between the sparsity graph and the sampling graph and on the value of ϵ, but in most of the non-extreme cases, at least 25% of the power is unnecessary if rakeness-based CS is adopted.

V. Conclusion

Rakeness-based CS applied to multiple-graph signals is an effective way to administer the trade-off between short- and long-range communication in a quite common IoT scenario that sees the interplay of local WSN and geographic information hubs. It is estimated that its exploitation may yield not less than 25% of power saving.

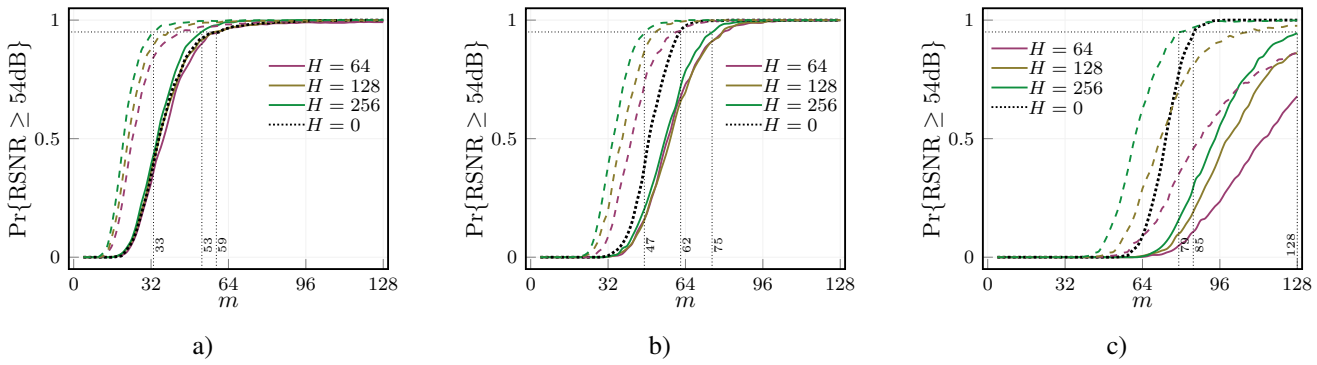

Fig. 3. PCR plotted against m for different configurations. Track color indicates the available hop budget ($H = 0$ signifying vertex-only sampling). Solid lines correspond to random CS, dashed lines correspond to rakeness-based CS. The number of measurements needed to guarantee a PCR of 95% is highlighted for vertex-only sampling ($H = 0$) and for the best random and rakeness-based options. In a) $\kappa = 6$, the sparsity graph is the same Geo-0.15 used for sampling, and each vertex contributes not more than $M = 4$ measurements. In b) $\kappa = 14$, the sparsity graph is WS-6-0.3, and each vertex contributes not more than $M = 8$ measurements. In c) $\kappa = 24$, the sparsity graph is ER-0.1, and each vertex contributes not more than $M = 16$ measurements.

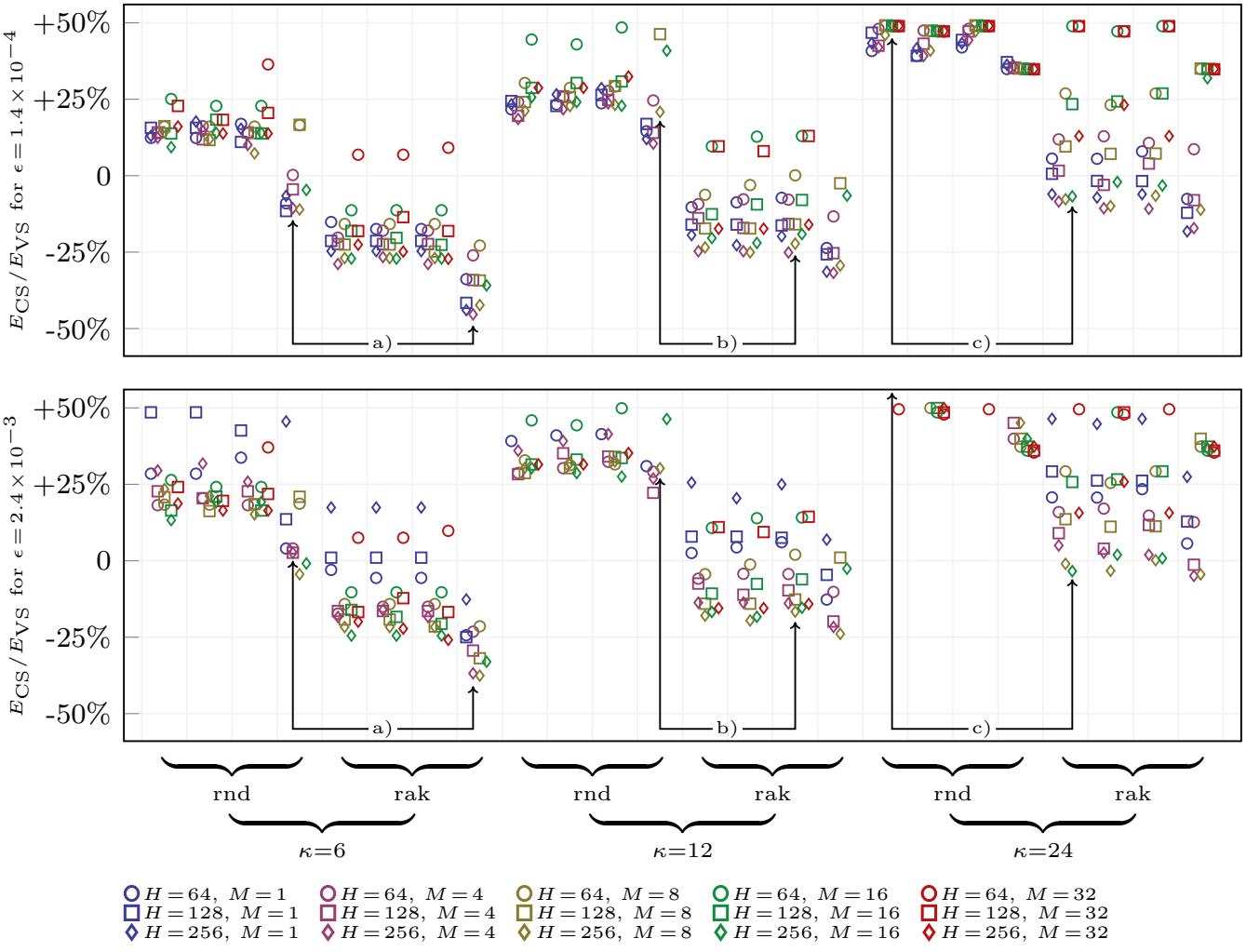

Fig. 4. Power saving with respect to vertex-only sampling in all the tested configurations. Each group of 4 points with the same shape and color correspond to the 4 sparsity graph (ER-01, BA-10-2, WE-10-0.6, and Geo-0.51). The color of a point indicates the available hop budget H, while its shape indicates the maximum number of measurements M provided by each vertex. Different sparsities κ are shown and for each sparsity, random and rakeness-based CS is considered. The upper plot considers a ratio between the energy needed by short-range and long-range communication equal to $\epsilon = \epsilon_{\min} = 1.4 \times 10^{-4}$. The lower plot considers $\epsilon = \epsilon_{\max} = 2.4 \times 10^{-3}$. Highlighted points correspond to the a), b), and c) plots of Figure 3.

REFERENCES

[1] A. Sandryhaila, J.M.F. Moura, "Discrete Signal Processing on Graphs," IEEE Transactions on Signal Processing, vol. 61, no. 7, pp. 1644–1656, 2013

[2] D. I. Shuman, S. K. Narang, P. Frossard, A. Ortega and P. Vandergheynst, "The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains," IEEE Signal Processing Magazine, vol. 30, no. 3, pp. 83-98, May 2013.

[3] A. Sandryhaila and J. M. F. Moura, "Discrete Signal Processing on Graphs: Frequency Analysis," in IEEE Transactions on Signal Processing, vol. 62, no. 12, pp. 3042-3054, June, 2014.

[4] A. Sandryhaila and J. M. F. Moura, "Discrete signal processing on graphs: Graph fourier transform," IEEE International Conference on Acoustics, Speech and Signal Processing 2013, pp. 6167-6170.

[5] http://www.st.com/content/st_com/en/products/wireless-connectivity/bluetooth-bluetooth-low-energy/bluenrg.html

[6] http://www.microchip.com/wwwproducts/en/RN4020

[7] http://www.semtech.com/apps/product.php?pn=SX1272

[8] http://www.nemeus.fr/en/nemeus-mm002-2/

[9] M. Mangia, F. Pareschi, V. Cambareri, R. Rovatti, G. Setti, "Rakeness-Based Design of Low-Complexity Compressed Sensing," IEEE Transactions on Circuits and Systems - I: Regular Papers, vol. 64, no. 5, pp. 1201-1213, May 2017.

[10] G. Yang, V. Y. F. Tan, C. K. Ho, S. H. Ting and Y. L. Guan, "Wireless Compressive Sensing for Energy Harvesting Sensor Nodes," in IEEE Transactions on Signal Processing, vol. 61, no. 18, pp. 4491-4505, Sept. 2013.

[11] J. Y. Park, M. B. Wakin and A. C. Gilbert, "Modal Analysis With Compressive Measurements," in IEEE Transactions on Signal Processing, vol. 62, no. 7, pp. 1655-1670, April 2014.

[12] X. Liu, M. Zhang, T. Xiong, A. G. Richardson, T. H. Lucas, P. S. Chin, R. Etienne-Cummings, T. D. Tran, J. Van der Spiegel, "A Fully Integrated Wireless Compressed Sensing Neural Signal Acquisition System for Chronic Recording and Brain Machine Interface," IEEE Transactions on Biomedical Circuits and Systems, vol. 10, no. 4, pp. 874-883, Aug. 2016.

[13] D. L. Donoho, "Compressed Sensing," IEEE Transactions on Information Theory, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[14] E.J. Candes, M.B. Wakin, "An Introduction To Compressive Sampling," IEEE Signal Processing Magazine, vol. 25, no. 2, pp. 21–30, 2008

[15] S. Chen, R. Varma, A. Sandryhaila, J. Kovačević, "Discrete Signal Processing on Graphs: Sampling Theory," IEEE Transactions on Signal Processing, vol. 63, no. 24, pp. 6510–6523, 2015

[16] M. Mangia, R. Rovatti, G. Setti, "Rakeness in the design of analog-to-information conversion of sparse and localized signals," IEEE Transactions on Circuits and Systems - I: Regular Papers, vol. 59, no. 5, pp. 1001–1014, 2012

[17] V. Cambareri, M. Mangia, F. Pareschi, R. Rovatti, G. Setti, "A case study in low-complexity ECG signal encoding: How compressing is compressed sensing?" IEEE Signal Processing Letters, vol. 22, no. 10, pp. 1743–1747, 2015

[18] en.wikipedia.org/wiki/ErdosRenyi_model|Barabasi-Albert_model|Watts_and_Strogatz_model|Random_geometric_graph

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

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