Distributed Compressed Sensing of Jointly Sparse Signals

Marco F. Duarte, Shriram Sarvotham, Dror Baron,
Michael B. Wakin and Richard G. Baraniuk
Department of Electrical and Computer Engineering, Rice University

***Abstract*—Compressed sensing is an emerging field based on the revelation that a small collection of linear projections of a sparse signal contains enough information for reconstruction. In this paper we expand our theory for *distributed compressed sensing* (DCS) that enables new distributed coding algorithms for multi-signal ensembles that exploit both intra- and inter-signal correlation structures. The DCS theory rests on a new concept that we term the *joint sparsity* of a signal ensemble. We present a second new model for jointly sparse signals that allows for joint recovery of multiple signals from incoherent projections through simultaneous greedy pursuit algorithms. We also characterize theoretically and empirically the number of measurements per sensor required for accurate reconstruction.**

I. Introduction

A core tenet of signal processing and information theory is that signals, images, and other data often contain some type of *structure* that enables intelligent representation and processing. Current state-of-the-art compression algorithms employ a decorrelating transform such as an exact or approximate Karhunen-Loève transform (KLT) to compact a correlated signal's energy into just a few essential coefficients. Such *transform coders* [1] exploit the fact that many signals have a *sparse* representation in terms of some basis, meaning that a small number K of adaptively chosen transform coefficients can be transmitted or stored rather than $N \gg K$ signal samples. For example, smooth signals are sparse in the Fourier basis, and piecewise smooth signals are sparse in a wavelet basis [1]; the coding standards MP3, JPEG, and JPEG2000 directly exploit this sparsity.

A. Compressed sensing (CS)

A new framework for single-signal sensing and compression has developed recently under the rubric of *Compressed Sensing* (CS) [2, 3]. CS builds on the surprising revelation that a signal having a sparse representation in one basis can be recovered from a small number of projections onto a second basis that is *incoherent* with the first.[1] In fact, for an N-sample signal that is K-sparse,[2] roughly cK projections of the signal onto the incoherent basis are required to reconstruct the signal with high probability (typically $c \approx 3$). This has promising implications for applications involving sparse signal acquisition. Instead of sampling a K-sparse signal N times, only cK incoherent measurements suffice, where K can be orders of magnitude less than N. Moreover, the cK measurements need not be manipulated in any way before being transmitted, except possibly for some quantization. Interestingly, independent and identically distributed (i.i.d.) Gaussian or Rademacher (random ± 1) vectors provide a useful *universal* measurement basis that is incoherent with *any* given basis with high probability.

B. Distributed compressed sensing (DCS)

In this paper, we introduce new theory and algorithms for *distributed compressed sensing* (DCS) that exploit both intra- *and* inter-signal correlation structures. In a typical DCS scenario, a number of sensors measure signals (of any dimension) that are each individually sparse in some basis and also correlated from sensor to sensor. Each sensor *independently* encodes its signal by projecting it onto another, incoherent basis (such as a random one) and then transmits just a few of the resulting coefficients to a collection point. Under the right conditions, a decoder at the collection point can *jointly* reconstruct all of the signals precisely.

The DCS theory rests on a concept that we term the *joint sparsity* of a signal ensemble. We have introduced a first model for jointly sparse signals and proposed corresponding joint reconstruction algorithms [4]. We have also derived results on the required measurement rates for signals that have sparse representations under each of the models: while the sensors operate entirely *without collaboration*, we see dramatic savings relative to the number measurements required for separate CS decoding.

This work was supported by NSF-CCF, NSF-NeTS, ONR, and AFOSR. Web: dsp.rice.edu/cs.
Email: {duarte, shri, baron, wakin, richb}@rice.edu

[1] Roughly speaking, *incoherence* means that no element of one basis has a sparse representation in terms of the other basis.
[2] By K-sparse, we mean that the signal can be written as a sum of K basis functions.

In this paper, we extend our previous work by introducing a new joint sparsity model that is applicable to many real world scenarios, including sensor networks for smoothly varying signal fields. We also propose also specially tailored recovery algorithms that achieve performance similar or better to that of our previous model. The paper is organized as follows: Section II reviews the CS theory. Section III presents our new joint sparsity model and reconstruction algorithms, and Section IV analysis and simulations. Finally, Section V concludes the paper.

II. Compressed Sensing

We briefly explain the Compressed Sensing (CS) framework proposed in [2, 3] to make the paper self-contained. Suppose that x is a signal, and let $\Psi = \{\psi_1, \psi_2, \dots\}$ be a basis or *dictionary* of vectors. When we say that x is *sparse*, we mean that x is well approximated by a linear combination of a small set of vectors from Ψ. That is, $x \approx \sum_{i=1}^{K} \theta_{n_i} \psi_{n_i}$ where $K \ll N$; we say that x is K*-sparse* in Ψ and call Ψ the sparse basis. The CS theory states that it is possible to construct an $M \times N$ *measurement* matrix Φ, where $M \ll N$, yet the measurements $y = \Phi x$ preserve the essential information about x. For example, let Φ be a $cK \times N$ random matrix with i.i.d. Gaussian entries, where $c = c(N, K)$ is an *oversampling factor*. Using such a matrix it is possible, with high probability, to recover any signal that is K-sparse in the basis Ψ from its image under Φ. For signals that are not K-sparse but *compressible*, meaning that their coefficient magnitudes decay exponentially, there are tractable algorithms that achieve not more than a multiple of the error of the best K-term approximation of the signal.

Several algorithms have been proposed for recovering x from the measurements y, each requiring a slightly different constant c. The canonical approach [2, 3] uses linear programming to solve the ℓ_1 minimization problem

$$\widehat{\theta} = \arg\min_{\theta} \|\theta\|_1 \quad \text{subject to} \quad \Phi\Psi\theta = y.$$

This problem requires $c \approx \log_2(1 + N/K)$ [4] but has somewhat high computational complexity. Additional methods have been proposed involving greedy pursuit methods. Examples include Matching Pursuit (MP) and Orthogonal Matching Pursuit (OMP), which tend to require fewer computations but at the expense of slightly more measurements [5].

III. Joint Sparsity Models

In this section, we generalize the notion of a signal being sparse in some basis to the notion of an ensemble of signals being *jointly sparse*. We consider two different *joint sparsity models* (JSMs) that apply in different situations. In these models, each signal is itself sparse, and so we could use the CS framework from above to encode and decode each one separately. However, there also exists a framework wherein a *joint representation* for the ensemble uses fewer total vectors.

We will use the following notation for signal ensembles and our measurement model. Denote the *signals* in the ensemble by x_j, $j \in \{1, 2, \dots, J\}$, and assume that each signal $x_j \in \mathbb{R}^N$. We use $x_j(n)$ to denote sample n in signal j, and we assume that there exists a known *sparse basis* Ψ for $\mathbb{R}^N$ in which the x_j can be sparsely represented. The coefficients of this sparse representation can take arbitrary real values (both positive and negative). Denote by Φ_j the *measurement matrix* for signal j; Φ_j is $M_j \times N$ and, in general, the entries of Φ_j are different for each j. Thus, $y_j = \Phi_j x_j$ consists of $M_j < N$ *incoherent measurements* of x_j. We will emphasize random i.i.d. Gaussian matrices Φ_j in the following, but other schemes are possible, including random ± 1 Bernoulli/Rademacher matrices, and so on.

A. JSM-1: Sparse common component + innovations

In this model, all signals share a *common* sparse component while each individual signal contains a sparse *innovation* component; that is,

$$x_j = z + z_j, \quad j \in \{1, 2, \dots, J\}$$

with $z = \Psi\theta_z$, $\|\theta_z\|_0 = K$ and $z_j = \Psi\theta_j$, $\|\theta_j\|_0 = K_j$. Thus, the signal z is common to all of the x_j and has sparsity K in basis Ψ.[3] The signals z_j are the unique portions of the x_j and have sparsity K_j in the same basis.

A practical situation well-modeled by JSM-1 is a group of sensors measuring temperatures at a number of outdoor locations throughout the day. The temperature readings x_j have both temporal (intra-signal) and spatial (inter-signal) correlations. Global factors, such as the sun and prevailing winds, could have an effect z that is both common to all sensors and structured enough to permit sparse representation. More local factors, such as shade, water, or animals, could contribute localized innovations z_j that are also structured (and hence sparse). A similar scenario could be imagined for a network of sensors recording light intensities, air pressure, or other phenomena. All of these scenarios correspond to measuring properties of physical processes that change smoothly in time and in space and thus are highly correlated.

B. JSM-2: Common sparse supports model

In this model all signals are constructed from the same sparse set of basis vectors, but with different coefficients:

$$x_j = \Psi\theta_j, \quad j \in \{1, 2, \dots, J\},$$

[3] The ℓ_0 norm $\|\theta\|_0$ merely counts the number of nonzero entries in the vector θ.

where each θ_j is supported only on the same $\Omega \subset \{1, 2, \ldots, N\}$ with $|\Omega| = K$. Hence, all signals are K-sparse and are constructed from the same K elements of Ψ, but with arbitrarily different coefficients.

A practical situation well-modeled by JSM-2 is where multiple sensors acquire the same signal but with phase shifts and attenuations caused by signal propagation. In many cases it is critical to recover each one of the sensed signals, such as in many acoustic localization and array processing algorithms. Another useful application for JSM-2 is MIMO communication [6].

IV. Reconstruction Algorithms

We have studied the JSM-1 model and proposed reconstruction algorithms in [4]. In this paper we focus on the analysis of the JSM-2 model.

Under the JSM-2 signal ensemble model, separate recovery via ℓ_1 minimization would require cK measurements per signal. As we now demonstrate, the total number of measurements can be reduced substantially by employing specially tailored joint reconstruction algorithms that exploit the common structure among the signals, in particular the common coefficient support set Ω.

The algorithms we propose are inspired by conventional greedy pursuit algorithms for CS (such as OMP [5]). In the single-signal case, OMP iteratively constructs the sparse support set Ω; decisions are based on inner products between the columns of $\Phi\Psi$ and a residual. In the multi-signal case, more clues are available for determining the elements of Ω.

A. Recovery via One-Step Greedy Algorithm (OSGA)

When there are many correlated signals in the ensemble, a simple non-iterative greedy algorithm based on inner products will suffice to recover the signals jointly. For simplicity but without loss of generality, we assume that $\Psi = I_N$ (it can be absorbed by the measurement matrix) and that an equal number of measurements $M_j = M$ are taken of each signal. We write Φ_j in terms of its columns: $\Phi_j = [\phi_{j,1}, \phi_{j,2}, \ldots, \phi_{j,N}]$. The algorithm follows:

1) **Get greedy:** Given all of the measurements, compute the test statistics

$$\xi_n = \frac{1}{J} \sum_{j=1}^{J} \langle y_j, \phi_{j,n} \rangle^2 \tag{1}$$

for $n \in \{1, 2, \ldots, N\}$ and estimate the common coefficient support set by

$$\widehat{\Omega} = \{n \text{ having one of the } K \text{ largest } \xi_n\}.$$

When the sparse, nonzero coefficients are sufficiently generic (as defined below), we have the following surprising result, which is proved in [7] .

Theorem 1: Let Ψ be an orthonormal basis for $\mathbb{R}^N$, let the measurement matrices Φ_j contain i.i.d. Gaussian entries, and assume that the nonzero coefficients in the θ_j are i.i.d. Gaussian random variables. Then with $M \geq 1$ measurements per signal, OSGA recovers Ω with probability approaching one as $J \to \infty$.

In words, with *fewer* than K measurements per sensor, it is possible to recover the sparse support set Ω under the JSM-2 model. Of course, this approach does not recover the K coefficient values for each signal; that requires K measurements per sensor.

Theorem 2: Assume that the nonzero coefficients in the θ_j are i.i.d. Gaussian random variables. Then the following statements hold:

1) Let the measurement matrices Φ_j contain i.i.d. Gaussian entries, with each matrix having an oversampling factor of $c = 1$ (that is, $M_j = K$ for each measurement matrix Φ_j). Then OSGA recovers all signals from the ensemble $\{x_j\}$ with probability approaching one as $J \to \infty$.
2) Let Φ_j be a measurement matrix with oversampling factor $c < 1$ (that is, $M_j < K$), for some $j \in \{1, 2, \ldots, J\}$. Then with probability one, the signal x_j cannot be uniquely recovered by any algorithm for any value of J.

The first statement is an immediate corollary of Theorem 1; the second statement follows because each equation $y_j = \Phi_j x_j$ would be underdetermined even if the nonzero indices were known. Thus, under the JSM-2 model, the one-step greedy algorithm asymptotically performs as well as an oracle decoder that has prior knowledge of the locations of the sparse coefficients.

Theorem 2 provides tight achievable and converse bounds for JSM-2 signals, in the sense that the number of measurements needed for success is only one greater than the number that yields reconstruction failure. Our numerical experiments show that OSGA works well even when M is small, as long as J is sufficiently large. However, in the case of fewer signals (small J), OSGA performs poorly; see Figure 1. We propose next an alternative recovery technique based on simultaneous greedy pursuit that performs well for small J.

B. Recovery via iterative greedy pursuit

In practice, the common sparse support among the J signals enables a fast iterative algorithm to recover all of the signals jointly. Tropp and Gilbert have proposed one such algorithm, called *Simultaneous Orthogonal Matching Pursuit* (SOMP) [6], which can be readily applied in our DCS framework. SOMP is a variant of OMP that seeks to identify Ω one element at a time. We dub the DCS-tailored SOMP algorithm *DCS-SOMP* .

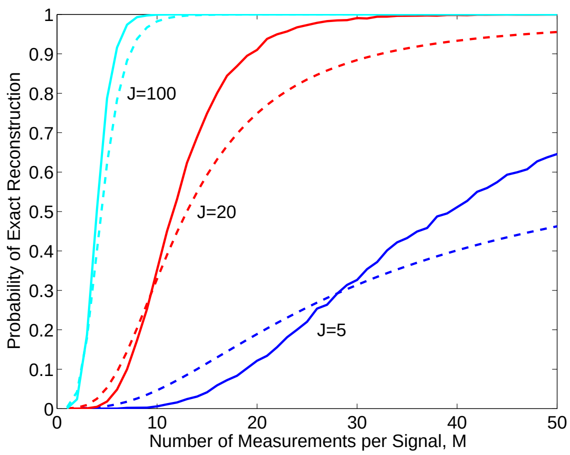

Fig. 1. *Reconstruction using OSGA for JSM-2. Theoretical (dashed) versus experimental (solid) probability of error in recovering the support set Ω in JSM-2 using OSGA for fixed $N = 50$, $K = 5$ and varying number of measurements M and number of signals $J = 5$, $J = 20$, and $J = 100$. The theoretical formulation is presented in [8].*

To adapt the original SOMP algorithm to our setting, we first extend it to cover a different measurement basis Φ_j for each signal x_j. Then, in each DCS-SOMP iteration, we select the column index $n \in \{1, 2, \ldots, N\}$ that accounts for the greatest amount of residual energy across *all* signals. As in SOMP, we orthogonalize the remaining columns (in each measurement basis) after each step; after convergence we obtain an expansion of the measurement vector on an orthogonalized subset of the holographic basis vectors. To obtain the expansion coefficients in the sparse basis, we then reverse the orthogonalization process using the QR matrix factorization. The algorithm is as follows:

1) **Initialize:** Set the iteration counter $\ell = 1$. For each signal index $j \in \{1, 2, \ldots, J\}$, initialize the orthogonalized coefficient vectors $\widehat{\beta}_j = 0$, $\widehat{\beta}_j \in \mathbb{R}^M$; also initialize the set of selected indices $\widehat{\Omega} = \emptyset$. Let $r_{j,\ell}$ denote the residual of the measurement y_j remaining after the first ℓ iterations, and initialize $r_{j,0} = y_j$.
2) **Select** the dictionary vector that maximizes the value of the sum of the magnitudes of the projections of the residual, and add its index to the set of selected indices

$$n_\ell = \arg\max_{n=1,2,\ldots,N} \sum_{j=1}^{J} \frac{|\langle r_{j,\ell-1}, \phi_{j,n}\rangle|}{\|\phi_{j,n}\|_2},$$
$$\widehat{\Omega} = [\widehat{\Omega} \;\; n_\ell].$$

3) **Orthogonalize** the selected basis vector against the orthogonalized set of previously selected dictionary vectors

$$\gamma_{j,\ell} = \phi_{j,n_\ell} - \sum_{t=0}^{\ell-1} \frac{\langle \phi_{j,n_\ell}, \gamma_{j,t}\rangle}{\|\gamma_{j,t}\|_2^2}\gamma_{j,t}.$$

4) **Iterate:** Update the estimate of the coefficients for the selected vector and residuals

$$\widehat{\beta}_j(\ell) = \frac{\langle r_{j,\ell-1}, \gamma_{j,\ell}\rangle}{\|\gamma_{j,\ell}\|_2^2},$$
$$r_{j,\ell} = r_{j,\ell-1} - \frac{\langle r_{j,\ell-1}, \gamma_{j,\ell}\rangle}{\|\gamma_{j,\ell}\|_2^2}\gamma_{j,\ell}.$$

5) **Check for convergence:** If $\|r_{j,\ell}\|_2 > \epsilon\|y_j\|_2$ for all j, then increment ℓ and go to Step 2; otherwise, continue to Step 6. The parameter ϵ determines the target error power level allowed for algorithm convergence. Note that due to Step 3 the algorithm can only run for up to M iterations.
6) **De-orthogonalize:** Apply QR factorization on the mutilated basis[4] $\Phi_{j,\widehat{\Omega}} = Q_j R_j = \Gamma_j R_j$. Since $y_j = \Gamma_j \beta_j = \Phi_{j,\widehat{\Omega}} x_{j,\widehat{\Omega}} = \Gamma_j R_j x_{j,\widehat{\Omega}}$, where $x_{j,\widehat{\Omega}}$ is the mutilated coefficient vector, we can compute the signal estimates $\{\widehat{x}_j\}$ as

$$\widehat{\theta}_{j,\widehat{\Omega}} = R_j^{-1}\widehat{\beta}_j,$$
$$\widehat{x}_j = \Psi\widehat{\theta}_j,$$

where $\widehat{\theta}_{j,\widehat{\Omega}}$ is the mutilated version of the sparse coefficient vector $\widehat{\theta}_j$.

In practice, each sensor projects its signal x_j via $\Phi_j x_j$ to produce $\widehat{c}K$ measurements for some $\widehat{c}$. The decoder then applies DCS-SOMP to reconstruct the J signals jointly. We orthogonalize because as the number of iterations approaches M the norms of the residues of an orthogonal pursuit decrease faster than for a non-orthogonal pursuit.

Thanks to the common sparsity structure among the signals, we believe that DCS-SOMP will succeed with $\widehat{c} < c(S)$. Empirically, we have observed that a small number of measurements proportional to K suffices for a moderate number of sensors J. We conjecture that $K+1$ measurements per sensor suffice as $J \to \infty$. Thus, in practice, this efficient greedy algorithm enables an oversampling factor $\widehat{c} = (K+1)/K$ that approaches 1 as J, K, and N increase.

C. Simulations

We now present a simulation comparing separate CS reconstruction versus joint DCS-SOMP reconstruction

[4] We define a *mutilated basis* Φ_Ω as a subset of the basis vectors from $\Phi = [\phi_1, \phi_2, \ldots, \phi_N]$ corresponding to the indices given by the set $\Omega = \{n_1, n_2, \ldots, n_M\}$, that is, $\Phi_\Omega = [\phi_{n_1}, \phi_{n_2}, \ldots, \phi_{n_M}]$. This concept can be extended to vectors in the same manner.

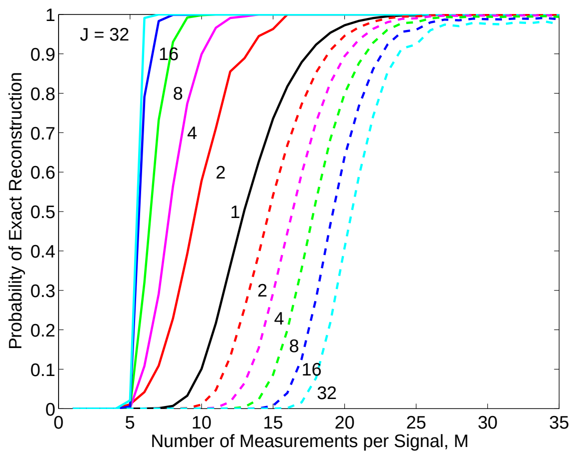

Fig. 2. *Reconstructing a signal ensemble with common sparse supports (JSM-2). We plot the probability of perfect reconstruction via DCS-SOMP (solid lines) and independent CS reconstruction (dashed lines) as a function of the number of measurements per signal M and the number of signals J. We fix the signal length to $N = 50$, the sparsity to $K = 5$, and average over 1000 simulation runs. An oracle encoder that knows the positions of the large signal expansion coefficients would use 5 measurements per signal.*

for a JSM-2 signal ensemble. Figure 2 plots the probability of perfect reconstruction corresponding to various numbers of measurements M as the number of sensors varies from $J = 1$ to 32. We fix the signal lengths at $N = 50$ and the sparsity of each signal to $K = 5$.

With DCS-SOMP, for perfect reconstruction of all signals the average number of measurements per signal decreases as a function of J. The trend suggests that, for very large J, close to K measurements per signal should suffice. In contrast, with separate CS reconstruction, for perfect reconstruction of all signals the number of measurements per sensor *increases* as a function of J. This surprise is due to the fact that each signal will experience an independent probability $p \leq 1$ of successful reconstruction; therefore the overall probability of complete success is p^J. Each sensor must then compensate by making additional measurements. This phenomenon further motivates joint reconstruction under JSM-2.

V. CONCLUSIONS

In this paper we have developed a new joint sparsity model that allows us to apply the DCS framework to a wider class of real world settings. For this new model (JSM-2), we have developed an efficient greedy algorithm for joint signal recovery and observed that important collective signal properties can be learned from as little as one measurement per signal. We note that our theoretical results are best-possible in the CS setting and cannot be improved upon. There are many opportunities for applications and extensions of these ideas.

Application to sensor networks: The area that appears most likely to benefit immediately from the new DCS theory is low-powered sensor networks, where energy and communication bandwidth limitations require that we perform data compression while minimizing inter-sensor communications [9]. DCS encoders work completely independently; therefore inter-sensor communication is required in a DCS-enabled sensor network only to support multi-hop networking to the data collection point. Moreover, the fact that typical sensor networks are designed to measure physical phenomena suggests that their data will possess considerable joint structure in the form of inter- and intra-signal correlations.

Compressible signals: In practice natural signals are not exactly ℓ_0 sparse but rather can be better modeled as ℓ_p sparse with $0 < p \leq 1$. Roughly speaking, a signal in a *weak*-ℓ_p ball has coefficients that decay as $n^{-1/p}$ once sorted according to magnitude [3]. The key concept is that the ordering of these coefficients is important. For our new model, we can extend the notion of simultaneous sparsity for ℓ_p-sparse signals whose sorted coefficients obey roughly the same ordering. This condition could perhaps be enforced as an ℓ_p constraint on the composite signal $\left\{\sum_j |x_j(1)|, \ \sum_j |x_j(2)|, \ \ldots, \ \sum_j |x_j(N)|\right\}$.

Acknowledgments. Thanks to Emmanuel Candès, Hyeokho Choi, Albert Cohen, Ron DeVore, Anna Gilbert, Rob Nowak, Jared Tanner, and Joel Tropp for informative and inspiring conversations. Thanks to Mark Davenport for reviewing the manuscript and to Ryan King for supercharging our computational capabilities.

REFERENCES

[1] S. Mallat, *A Wavelet Tour of Signal Processing*. San Diego: Academic Press, 1999.

[2] E. Candès and T. Tao, "Near optimal signal recovery from random projections and universal encoding strategies," *IEEE Trans. Inform. Theory*, 2004, submitted.

[3] D. Donoho, "Compressed sensing," 2004, preprint.

[4] D. Baron, M. F. Duarte, S. Sarvotham, M. B. Wakin, and R. G. Baraniuk, "An information-theoretic approach to distributed compressed sensing," in *Proc. 43rd Allerton Conf. Communication, Control, and Computing*, Monticello, IL, Sept. 2005.

[5] J. Tropp and A. C. Gilbert, "Signal recovery from partial information via orthogonal matching pursuit," Apr. 2005, preprint.

[6] J. Tropp, A. C. Gilbert, and M. J. Strauss, "Simulataneous sparse approximation via greedy pursuit," in *IEEE 2005 Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, Philadelphia, Mar. 2005.

[7] D. Baron, M. Duarte, S. Sarvotham, M. B. Wakin, and R. G. Baraniuk, "Distributed compressed sensing," preprint.

[8] S. Sarvotham, M. B. Wakin, D. Baron, M. F. Duarte, and R. G. Baraniuk, "Analysis of the DCS one-stage greedy algoritm for common sparse supports," Rice University ECE Department, Tech. Rep., Oct. 2005, available at http://www.ece.rice.edu/˜shri/docs/TR0503.pdf.

[9] D. Estrin, D. Culler, K. Pister, and G. Sukhatme, "Connecting the physical world with pervasive networks," *IEEE Pervasive Computing*, vol. 1, no. 1, pp. 59–69, 2002.