

# Joint Optimization of Transport Cost and Reconstruction for Spatially-Localized Compressed Sensing in Multi-Hop Sensor Networks

Sungwon Lee and Antonio Ortega

University of Southern California, Los Angeles, USA

E-mail: sungwonl@usc.edu, antonio.ortega@usc.edu

**Abstract**—In sensor networks, energy efficient data manipulation / transmission is very important for data gathering, due to significant power constraints on the sensors. As a potential solution, Compressed Sensing (CS) has been proposed, because it requires capturing a smaller number of samples for successful reconstruction of sparse data. Traditional CS does not take explicitly into consideration the cost of each measurement (it simply tries to minimize the number of measurements), and this ignores the need to transport measurements over the sensor network. In this paper, we study CS approaches for sensor networks that are spatially-localized, thus reducing the cost of data gathering. In particular, we study the reconstruction accuracy properties of a spatially-localized distributed CS system. We introduce the concept of maximum energy overlap between clusters and basis functions, and show that the corresponding metric can be used to estimate the minimum number of measurements needed to achieve accurate reconstruction. Based on this metric, we propose a centralized iterative algorithm for joint optimization of the energy overlap and distance between nodes in each cluster. Our simulation results show that we can achieve significant savings in transport cost with small reconstruction error.

## I. INTRODUCTION

Sensor networks consist of numerous tiny and cheap sensors deployed on any physical region to monitor and report physical phenomena. This makes it possible to directly utilize space/time localized information for a wide range of applications. To collect information from sensors spread over space and deliver it to a destination (called the fusion center (FC)), energy-aware data manipulation / transmission is required because power is a scarce resource in the sensors. For the purpose of efficient data gathering, joint routing and compression has been studied for locally correlated sensor network data. Most of the early work was theoretical in nature and, while providing important insights, did not fully consider practical details of how compression is to be achieved [1], [2], [3]. More recently, it has been shown how practical compression schemes such as distributed wavelets can be adapted to work efficiently with various routing strategies [4], [5], [6], [7].

Transform-based techniques, e.g., wavelet based approaches [4], [5], [8] or the distributed KLT [9], can reduce the number of bits to be transmitted to the FC, so as to reduce transport cost. These transform techniques are inherently critically-sampled, which means that the number of samples (transform coefficients) transmitted to the FC is equal to the number of sensors. Thus, their cost of gathering scales up with

the number of sensors, which could be undesirable for large-scale sensor networks. As a potential alternative, Compressed Sensing (CS) has been proposed because it requires capturing a smaller number of measurements for successful reconstruction; specifically, the number of measurements carried to the FC depends on the characteristics (sparseness) of the signal rather than on the dimension of the signal [10], [11], [12] (which in our case corresponds to the number of sensors in the network.)

However, while the potential advantages of CS have been acknowledged [13], [14], obstacles remain for it to become competitive with more established (e.g., transform-based) data gathering and compression techniques. A primary reason is that CS theoretical developments have focused on *minimizing the number of measurements* (i.e., the number of values relayed to the FC and obtained as linear combinations of samples obtained by the sensors), rather than on *minimizing the cost of each measurement* (i.e., transport cost for each aggregate). In fact, in many CS applications (e.g., [15] [16]), each measurement is a linear combination of *many (or all) samples of the signal*. Clearly, this kind of “dense” measurement system is not efficient for sensor networks, since each final measurement would require aggregating samples from many sensors that are potentially far from each other, so that the total cost can potentially be higher than that of a raw data gathering scheme.

To address this problem, *sparse measurement* approaches (where each measurement requires information from just a few active sensors) have been proposed for both single hop [17] and multi-hop [13], [14] sensor networks. While reducing the number of samples used to compute each CS measurement reduces overall cost, it does not guarantee that the resulting system will be efficient, as the cost also depends on the relative positions of sensors providing information for each measurement. If sensors contributing to a given measurement are far apart, the cost will still be significant even with a sparse measurement approach. This is why sparse random projection (SRP in [17]) does not perform well in terms of transport cost [18].

The key observation in this work is that an efficient measurement system needs to be *both sparse* (few sensors contribute samples to each measurement) and *spatially-localized* (the sensors that contribute to each measurement are close to each other) in order to be competitive in terms of transport cost

and reconstruction accuracy. In this paper we extend our work in [18], where we first proposed the use of a cluster-based technique for CS. In our cluster-based approach each measurement is a linear combination of samples captured *within a single cluster*, and clusters selected to contain sensor nodes that are close to each other. Here, we extend our previous work [18] by analyzing how the choice of spatial clusters affects the reconstruction accuracy, for a given spatially-localized sparsity basis. Moreover, we propose novel clustering techniques that take into consideration both transport cost and reconstruction quality.

More specifically, we have two main contributions in this paper. First, we introduce the concept of *maximum* energy overlap between clusters and basis functions, which we denote  $\beta$ . If basis functions and clusters have similar spatial localization, most of energy of a given basis function is likely to be concentrated in a few clusters, which means that only measurements taken from those clusters are likely to contribute to reconstructing a signal that contains that specific basis function. Since the measurement system is not aware *a priori* of where signals will be localized, it needs to gather enough measurements to reconstruct signals with any spatial localization (refer to the example in Section II-B for details), and since each cluster overlaps only a few such basis functions, it will need to have a larger number of measurements. Conversely, for the same number of measurements, as the energy of the basis functions is more evenly distributed over clusters (smaller  $\beta$ ), this could lead to better reconstruction. To verify this, we provide a proof that the minimum number of measurements for perfect reconstruction is proportional to  $\beta$ . Therefore, for given basis functions, we can estimate performance of different clustering schemes by computing  $\beta$ .

Second, we propose a centralized iterative algorithm with a design parameter,  $\lambda$ , for joint optimization of the energy overlap and distance between nodes in each cluster. A joint optimization is required because there exists a *tradeoff* between  $\beta$  and the distance. To achieve smaller  $\beta$  (which leads to a reduction of number of measurements), each basis function should be overlapped with more clusters. This means that the nodes within a cluster tend to be farther from each other because basis functions are localized in space. Since total transport cost is a function of the number of measurements and transport cost per measurement, the trade-off allows reducing the number of measurements at the cost of increasing transport cost per measurement. By joint optimization using an appropriately chosen  $\lambda$ , we can achieve a good trade-off between transport cost and reconstruction accuracy.

In this paper, after formulating problem in Section 2, we provide our main theoretical results in Section 3. Based on these, we provide an iterative algorithm in Section 4, then verify the performance by simulation in Section 5.

## II. PROBLEM FORMULATION

Before going into details of problem formulation, we briefly present about compressed sensing (CS). A key observation in

CS is that an  $N$ -sample signal ( $\mathbf{x}$ ) having a sparse representation in one basis can be recovered from a small number of measurements (smaller than  $N$ ) onto a second basis that is incoherent with the first [10], [11]. More formally, if a signal,  $\mathbf{x} \in \mathbb{R}^N$ , is sparse in a given basis  $\tilde{\Psi}$  (the sparsity inducing basis or sparsifying basis), then  $\mathbf{x} = \tilde{\Psi}\mathbf{a}$ ,  $|\mathbf{a}|_0 = K$ , where  $K \ll N$ . The original  $K$ -sparse signal can be reconstructed with  $O(K \log N)$  *dense* measurements by finding the sparsest solution to an under-determined linear system of equations.

In this paper, we consider a  $K$ -sparse 2D signal,  $\mathbf{x} \in \mathbb{R}^N$ , in a given sparsifying basis  $\tilde{\Psi}$ . The signal is measured by  $N$  sensors assumed to be positioned on a square 2D grid, i.e.,  $\mathbf{x}$  is a snapshot of 2D data at a given time stamp. For efficient data-gathering from sensors spread over space to the FC located at the center of the network, we consider distributed measurement strategies that are both sparse *and* spatially localized.

For the spatially-localized sparse projections, as proposed in [18], we first divide the network into  $N_c$  clusters of sensors close to each other and force each of the  $M$  measurements to be obtained from nodes within one of the clusters. Sensors in the same cluster can create a measurement by a linear combination of data samples captured within the cluster with some (random) coefficients. Since each cluster consists of the localized sensors (which contributes to spatially-localized projections) and the number of sensors in each cluster is smaller than the total number of sensors (which contributes to sparse projections), this can lead to a energy-efficient data gathering. We will show how this procedure can be represented by a series of CS matrices and a matrix associated with the cluster formation in following section.

### A. Spatially-Localized Projections in CS

An aggregation path in a sensor network can be represented by a row of the measurement matrix,  $\Phi$ . We place non-zero random coefficients in the positions corresponding to active sensors that provide their data for a specific measurement and the other positions are set to zero, which means that the sparsity of a particular measurement in the matrix depends on the number of nodes participating in each aggregation. To express  $M$  measurements in matrix formulation, we consider a down-sampling matrix,  $Q$ , that chooses  $M$  measurements with equal probability out of  $N$ . This can be expressed as:

$$\mathbf{y}_{M \times 1} = \mathbf{Q}_{M \times N} \Phi_{N \times N} \mathbf{x}_{N \times 1}$$

Similarly, the aggregations within a cluster can be expressed as a set of rows of  $\Phi$ . Since  $N_c$  non-overlapped clusters are considered, we can express the measurement system as a block diagonal matrix that contains  $N_c$  square sub-matrices,  $\Phi_i$  on its diagonal, so that  $\Phi_i$  represents an aggregation scheme of the  $i^{th}$  cluster in the network. Therefore, the dimension of  $\Phi_i$  is determined by the number of sensors contained in the  $i^{th}$  cluster. To associate  $\Phi_i$  with data,  $\mathbf{x}_i \subset \mathbf{x}$ , measured by sensors in the  $i^{th}$  cluster, we consider a permutation matrix,  $P$ , by multiplying  $\Phi$  with the output of  $P\mathbf{x}$ , that is

$$y = Q\Phi Px, \text{ where } \Phi = \begin{bmatrix} \Phi_1 & & & \\ & \Phi_2 & & \\ & & \ddots & \\ & & & \Phi_{N_c} \end{bmatrix}$$

A rectangular matrix,  $Q\Phi P$ , conveys information about locations and values of the random coefficients. Their locations and values reveal the structure of the clusters and the aggregation coefficients, respectively. Now, we separate it into two matrices:  $\Phi$  for the coefficients and  $P$  for the cluster formation. From now on, we will call the square block-diagonal  $\Phi$  a measurement matrix and  $P$  a clustering matrix.

Here, we need to see how the clustering matrix is related to the sparsifying basis matrix. Since a  $K$ -sparse signal is represented by  $K$  non-zero coefficients in a given basis  $\tilde{\Psi}$ ,  $x = \tilde{\Psi}a$ , the measurements are obtained by

$$y = Q\Phi Px = Q\Phi (P\tilde{\Psi}) a = Q\Phi\Psi a$$

Multiplied with  $\tilde{\Psi}$ , the clustering matrix  $P$  generates  $\Psi$ , whose rows are permuted rows in  $\tilde{\Psi}$ , e.g., if  $P(i, j) = 1$ , the  $i^{\text{th}}$  row of  $\Psi$  is replaced by the  $j^{\text{th}}$  row of  $\tilde{\Psi}$ :  $\psi_i^T = \sum_{j=1}^N P(i, j)\tilde{\psi}_j^T$ , where  $\psi_i^T$  and  $\tilde{\psi}_j^T$  are row vectors of  $\Psi$  and  $\tilde{\Psi}$ , respectively. After permutation by the clustering matrix, the measurement matrices for each cluster,  $\Phi_i$ , are correctly associated with data measured by sensors in the corresponding clusters,  $x_i$ .

In summary, a cluster-based measurement system leads to a block diagonal measurement matrix with appropriate permutation related to the physical positions of sensors. Note that recent work [21] [22], seeking to achieve fast CS computation, has also proposed measurement matrices with a block-diagonal structure, with results comparable to those of dense random projections. Our work, however, is motivated by achieving spatially localized projections so that our choice of block-diagonal structure will be constrained by the *deterministic* positions of the sensors instead of uniformly *random* permutation considered in [21] [22].

### B. Maximum Energy Overlap

A clustering matrix,  $P$ , can be associated to any clustering scheme so that our goal is to design an appropriate  $P$  to achieve both efficient routing and adequate reconstruction accuracy. Though it is clear that gathering within spatially-localized clusters leads to lower costs, it is not obvious how it may affect reconstruction accuracy. Thus, an important goal of this paper is to study the interaction between localized gathering and reconstruction. A key observation is that in order to achieve both efficient routing and adequate reconstruction accuracy, the relationship between sparsifying basis and clusters should be considered [18].

The relationship can be explained by the overlap between basis functions and clusters. As illustrated in Fig. 1 (a), three basis functions ( $B_1, B_2, B_3$ ) have overlaps with four square clusters ( $C_1, C_2, C_3, C_4$ ). In the example, the energy overlap

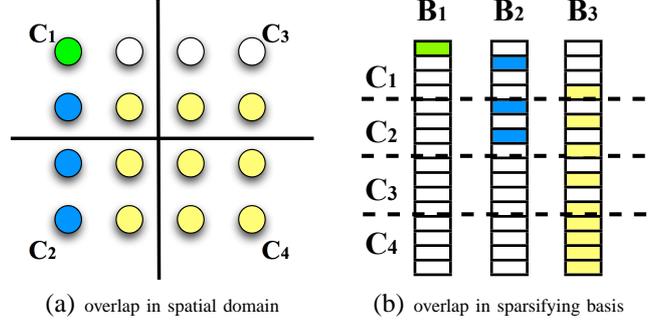


Fig. 1. Illustration of energy overlap (a) in  $4 \times 4$  grid network of 16 sensors; 4 square clusters and 3 basis functions with different spatial resolution. Assume that all the basis functions are normalized to 1 and their coefficients are uniform and inversely proportional to extent of basis functions. (b) in permuted sparsifying basis matrix,  $\Psi = P\tilde{\Psi}$ . The entries of each basis function (column vector of  $\Psi$ ) is filled with colors if non-zero coefficients exist and white otherwise. Note that 13 more basis functions exist but omitted.

of a basis function is the area of region overlapped by clusters.  $B_2$  is overlapped with two clusters ( $C_1$  and  $C_2$ ) and the value of energy overlap with  $C_2$  is larger than with  $C_1$ . In this case, the maximum energy overlap is 1 because  $B_1$  is completely contained within  $C_1$ .

Intuitively, measurements taken from a cluster can also convey information about data in other clusters when basis functions overlap with more than one cluster, e.g.,  $B_2$  in Fig. 1 (a) can be identified with measurements from those clusters ( $C_1$  and  $C_2$ ). If a specific basis function is completely contained within a cluster, e.g.  $B_1$ , then only measurements from  $C_1$  are likely to contribute to reconstructing a signal that contains  $B_1$ .

For example, for a  $K$ -sparse signal, the worst case scenario is when all  $K$  basis functions supporting data are completely contained in a single cluster, e.g.,  $B_1$  in Fig. 1 (a). To achieve a good reconstruction,  $O(K \log N)$  projections would be required from each cluster, leading to a total of  $O(KN_c \log N)$  projections. There would be two reasons for this poor performance. First, the identity of this cluster is not known *a priori*. So it is not possible to concentrate projections within that cluster without measuring information in the others. Second, projections from other clusters not overlapped with those basis vectors do not contribute to reconstruction performance as much as projections from the overlapped cluster.

Thus, for the same number of measurements, as the energy of basis functions is more evenly distributed over clusters, it could lead to better reconstruction performance because it is less likely that information in only one cluster is required for reconstruction. To quantify the distribution of energy overlap over clusters, we define the maximum energy overlap  $\beta$  as follows:

*Definition 2.1:* Maximum energy overlap,  $\beta(\Psi)$

$$\beta(\Psi) = \beta(P\tilde{\Psi}) = \max_{i,j} \sum_l \Psi_i^2(l, j), \beta(\Psi) \in [0, 1]$$

$\beta$  shows the maximum amount of energy of a basis functions captured by a single cluster. The matrix  $\Psi_i$  is a

rectangular sub-matrix corresponding to the  $i^{\text{th}}$  cluster. For example, as depicted in Fig. 1 (b), we first compute the sum of squared entries (colored cells) for each pair of  $(B_i, C_j)$ ; For  $B_1$ , energy overlap is 1 with  $C_1$  and zero with the other clusters. Then take the maximum value of the computed sums. If  $\beta$  is 1 (maximum value), it indicates that there exists at least one basis function completely covered by a cluster in space such as the overlap between  $B_1$  and  $C_1$  in Fig. 1 (a). In contrast, small  $\beta$  means that most of basis functions are overlapped with multiple clusters in space.

As basis functions are overlapped with more clusters, we will have potentially higher chance to reconstruct signal correctly. To further improve localized CS performance, a clustering scheme that minimizes overlap should be chosen. The degree of overlapping between basis functions and clusters can be measured in many different ways;  $\beta$  is one possible approach to measure worst case energy overlap between basis functions and clusters. We will show how  $\beta$  affects reconstruction accuracy with respect to the minimum number of measurements for perfect reconstruction in following section.

### III. THEORETICAL RESULT

As shown in previous sections, the maximum energy overlap,  $\beta$ , is determined for given sparsifying basis,  $\tilde{\Psi}$ , and a clustering scheme,  $\mathbf{P}$ . Here, we show how  $\beta$  affects reconstruction accuracy by deriving the minimum number of measurements for perfect reconstruction as a function of  $\beta$ .

#### A. Definitions and Assumptions

We consider a clustering scheme where  $N$  sensors are separated into  $N_c$  non-overlapped clusters. For simplicity, we assume these clusters contain the same number of sensors, so that  $\Phi$  has  $N_c$  square sub-matrices with size of  $N/N_c \times N/N_c$  along its diagonal. Therefore, if each sub-matrix is orthogonal, then  $\Phi$  is also orthogonal, and vice versa.

Based on the problem formulation, our main result is based on three assumptions. First, the sparsifying basis,  $\tilde{\Psi}$ , is orthogonal. Second, the maximum absolute value of entries in the sparsifying basis is bounded,  $\max_{i,j} |\tilde{\Psi}(i,j)| \leq 1/\sqrt{\log N}$ , in order to prevent the degenerate cases such as the canonical basis in spatial domain ( $\tilde{\Psi} = \mathbf{I}$ ). This assumption is satisfied by bases such as DCT and Daubechies wavelet with a certain level of decomposition. Lastly, the measurement matrix,  $\Phi$ , is an orthogonalized i.i.d. Gaussian matrix;  $\Phi(i,j) \sim N(0, N_c/N)$  and  $\Phi^T \Phi = \mathbf{I}_N$ . In order to evaluate the coherence between measurement matrix,  $\Phi$ , and permuted sparsifying basis matrix,  $\Psi (= \mathbf{P}\tilde{\Psi})$ , we define a  $N \times N$  matrix  $\mathbf{U}$ :

$$\begin{aligned} \mathbf{U}_{N \times N} &= \begin{bmatrix} \Phi_1 & & & \\ & \Phi_2 & & \\ & & \ddots & \\ & & & \Phi_{N_c} \end{bmatrix} \begin{bmatrix} \Psi_1 \\ \Psi_2 \\ \vdots \\ \Psi_{N_c} \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{U}_1 = \Phi_1 \Psi_1 \\ \vdots \\ \mathbf{U}_{N_c} = \Phi_{N_c} \Psi_{N_c} \end{bmatrix} \end{aligned} \quad (1)$$

By assumption, each square sub-matrix,  $\Phi_i$ , has the same size and so does each  $\mathbf{U}_i$ . And  $\mathbf{U}$  is an orthogonal matrix because  $\Phi$  and  $\tilde{\Psi}$  are orthogonal by assumption and the clustering matrix,  $\mathbf{P}$ , is a permutation matrix, so that

$$\mathbf{U}^T \mathbf{U} = (\Phi \mathbf{P} \tilde{\Psi})^T (\Phi \mathbf{P} \tilde{\Psi}) = (\tilde{\Psi}^T \mathbf{P}^T \Phi^T) (\Phi \mathbf{P} \tilde{\Psi}) = \mathbf{I}_N$$

Since  $\mathbf{U}$  is orthogonal, mutual coherence is defined as in [12].

*Definition 3.1:* Mutual coherence [12]

For two orthogonal matrices ( $\Phi^T \Phi = \mathbf{I}_N, \Psi^T \Psi = \mathbf{I}_N$ ), mutual coherence is defined as  $\mu(\mathbf{U}) = \max_{i,j} |U(i,j)| = \max_{i,j} |\phi(i)\psi(j)|$ , where  $\phi(i)$  is a row vector of  $\Phi$  and  $\psi(j)$  is a column vector of  $\Psi$ . Note that  $\mu(\mathbf{U}) \in [0, 1]$

The coherence is a measure of similarity between  $\Phi$  and  $\tilde{\Psi}$ . A small value of  $\mu(\mathbf{U})$  indicates that  $\Phi$  and  $\tilde{\Psi}$  are incoherent with each other, i.e., no element of one basis ( $\tilde{\Psi}$ ) has a sparse representation in terms of the other basis ( $\Phi$ ). The minimum number of measurements for perfect reconstruction can be computed as follows.

*Theorem 3.2:* Minimum number of measurements [12]

Let  $\mathbf{U} = \Phi \Psi$  be an  $N \times N$  orthogonal matrix ( $\mathbf{U}^T \mathbf{U} = \mathbf{I}$ ) with  $|U(i,j)| \leq \mu(\mathbf{U})$ . For a given signal  $\mathbf{x} = \Psi \mathbf{a}$ , if  $\mathbf{a}$  is supported on a fixed (arbitrary) set  $T$  with  $K$  non-zero entries, the  $l_1$  optimizer can recover  $\mathbf{x}$  exactly with high probability if the number of measurements  $M$  satisfies

$$M = O(K\mu^2(\mathbf{U})N \log N) \quad (2)$$

The bound of measurements by Theorem 3.2 decreases as  $\Phi$  and  $\tilde{\Psi}$  are more incoherent, i.e., the minimum number of measurements for perfect reconstruction is determined by  $\mu$  for given  $K$  and  $N$ . Thus, if we can derive how  $\mu$  changes as a function of  $\beta$ , then we can also compute the bound of measurements.

#### B. Main Result

To get the bound on the number of measurements, we first derive an asymptotic upper bound on mutual coherence. With this bound, we can attain the minimum number of measurements for perfect reconstruction by using Theorem 3.2 because all the matrices are orthogonal.

*Proposition 3.3:* If the measurement matrix,  $\Phi$ , is orthogonalized i.i.d. Gaussian,  $N(0, \frac{N_c}{N})$  and orthogonal sparsifying basis,  $\tilde{\Psi}$  and clustering matrix,  $\mathbf{P}$  are known *a priori*, then  $\mu(\mathbf{U})$  is asymptotically bounded by

$$\Pr \left[ \mu(\mathbf{U}) \leq O\left(\sqrt{\beta \frac{N_c}{N} \log N}\right) \right] = 1 - O\left(\frac{1}{N}\right) \quad (3)$$

Proposition 3.3 quantifies the probability that coherence exceeds a certain bound. The probability that coherence is not

bounded by  $O(\sqrt{\beta \frac{N_c}{N} \log N})$  is close to 0 as  $N$  increases. For the proof, the main technical tools are large deviation inequalities of sum of independent random variables. Specifically, the result is derived from Bernsteins deviation inequality [23] and a union bound for the supremum of a random process. Refer to Appendix A for details of proof.

With aforementioned assumptions and Proposition 3.3, we present the impact of  $\beta$  on reconstruction accuracy in terms of the number of measurements.

**Theorem 3.4:** For a given signal  $\mathbf{x} = \Psi \mathbf{a}$  with  $|\mathbf{a}|_0 = K$  and a clustering (permutation) scheme with parameter  $\beta \in [0, 1]$ , the  $l_1$  optimizer can recover  $\mathbf{x}$  exactly with high probability if the number of measurements  $M$  satisfies

$$M = O(K\beta N_c \log^2 N) \quad (4)$$

Based on the bound of coherence by Proposition 3.3, we can derive Theorem 3.4, the minimum number of measurements for perfect reconstruction, by using Theorem 3.2 because, by assumption,  $\mathbf{U}$  is orthogonal.

Theorem 3.4 implies that if the measurement matrix is dense ( $\beta = 1$  and  $N_c = 1$ ), the number of measurements is nearly minimal (except for additional  $\log N$  factor), regardless of the sparsifying matrix. The bound for dense measurement matrix is identical with the bound,  $O(KN_c \log^2 N)$  by SRM [22]. However, for sparse measurement matrix, the bound in [22] does not involve a deterministic term,  $\beta$ , describing relationship between clusters and basis functions because SRM approach assumes uniformly-random clustering.

In general, the number of measurements is proportional to the maximum energy overlap because basis functions with more uniformly distributed energy increase the probability of correct reconstruction. Also, the number of measurements is proportional to the number of clusters,  $N_c$ . This implies that a sparser measurement matrix (larger  $N_c$ ) requires more measurements for the same level of reconstruction as shown in previous work [17], [22].

Note that there exists a *tradeoff* between  $\beta$  and distance between nodes belonging to the same cluster. The decrease of  $\beta$  can be achieved when each basis function should be overlapped with more clusters, which means that the nodes within a cluster tend to be farther from each other (which leads to a increase of transport cost per measurement) because basis functions are localized in space. Since total transport cost is a function of number of measurements and transport cost per measurement, the trade-off allows reducing the number of measurements at the cost of increasing transport cost per measurement. To construct clusters jointly aware of them, we propose a centralized iterative algorithm.

#### IV. CENTRALIZED ITERATIVE CLUSTERING ALGORITHM

To achieve a good tradeoff between transport cost and reconstruction accuracy, we need to jointly optimize  $\beta$  and distance between nodes in a cluster. This motivated us to design a centralized iterative algorithm that can generate

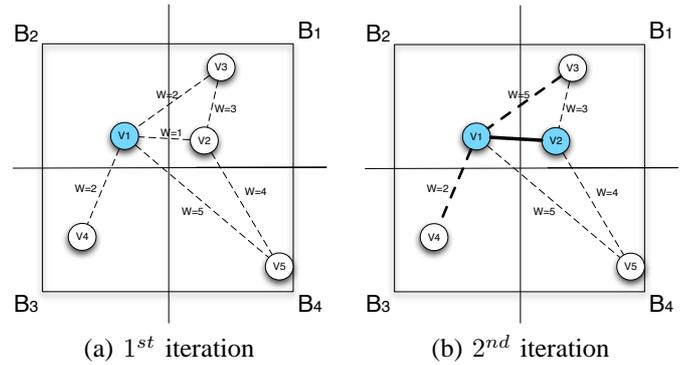


Fig. 2. Illustration of update of edge weights from (a) to (b). There exist 4 square basis functions ( $B_i$ ) and 5 nodes ( $v_i$ ) connected by edges, ( $v_i, v_j$ ), with different weights. Assume that the initial node for clustering is  $v_1$ . At the 1<sup>st</sup> step, a cluster is formed by  $\{v_1, v_2\}$  because ( $v_1, v_2$ ) has the minimum weight of 1. Then, without update of weights, the weight of an edge, ( $v_1, v_3$ ), remains the same but, with update, it increases from 3 to 5 because  $v_3$  and the cluster,  $\{v_1, v_2\}$  is overlapped with the same basis function,  $B_1$ .

optimal clusters that minimize both  $\beta$  and distance between nodes in clusters (transport cost per measurement).

#### A. Algorithm details

For a given undirected graph  $G = (V, E)$ , we assume that sparsifying basis,  $\Psi$ , is known *a priori* and all the basis functions (columns of  $\Psi$ ) are normalized to 1 so that  $\beta \in [0, 1]$ . Also,  $N$  nodes are placed along a square grid in a field with size of  $F$  by  $F$ . To quantify transport cost, we assume that the cost depends on distance between nodes and define the distance *in hops* as  $D(e)$  for an edge,  $e \in E$ , connecting two nodes, i.e., the smallest number of hops between two nodes in a multi-hop network. In addition, unlike what was assumed for Theorem 3.4, we consider clusters with different number of nodes in order to maximize performance.

---

#### Algorithm 1 Joint Optimization of $\beta$ and $D$

---

Given an undirected graph,  $G(V, E)$ , such that  $|V| = N$ .  
Assign  $N_c$  nodes to clusters; one for each cluster,  $V_{C_i}$ .  
 $E_{C_i} = \emptyset, \forall i$ .  
**for**  $k = 1$  to  $N - N_c$  **do**  
  Find  $E_n = \{(v_1, v_2) \mid v_1 \in V, v_2 \in V_{C_i}, \forall i\}$   
  Compute  $W(e) = D(e) + \lambda\beta(e), \forall e \in E_n$   
   $e_{min} = \arg \min_{e \in E_n} W(e)$   
   $v_{min} = \{v_1 \mid e_{min} = (v_1, v_2), v_2 \in V_{C_j}\}$   
  Add  $e_{min}$  to  $E_{C_i}$  and  $v_{min}$  to  $V_{C_i}$ .  
  Remove edges  $\in \{e \mid e = (v_{min}, v), \forall v \in V_{C_j}\}$  from  $E$ .  
  Remove  $v_{min}$  from  $V$ .  
**end for**

---

The goal of the algorithm is to construct  $N_c$  clusters that minimize transport costs assuming that reconstruction is guaranteed to be perfect. Transport costs depend on the distance between nodes and the number of measurements transmitted which, in turn, depends on  $\beta$ . Thus we need joint optimization of  $\beta$  and the distance between nodes. For joint optimization, we designed an algorithm that iteratively finds an edges with minimum weight associated with those two quantities then add

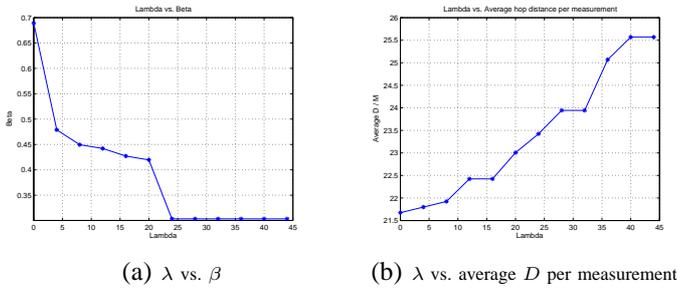


Fig. 3. Joint optimization of different  $\lambda$ . By running the algorithm with Daubechies 4 basis with  $2^{nd}$  level of decomposition, 256 sensors are separated into 16 clusters. Different choices of  $\lambda$  generate different results; as  $\lambda$  increases, the edge weights are more sensitive to the change of  $\beta$  so that  $\beta$  decreases at the cost of increasing  $D$ .

it to one of clusters. We first define the weight for an edge,  $e$ , connected to a cluster as

$$W(e) = D(e) + \lambda\beta(e), \quad \lambda > 0, \quad (5)$$

where  $\beta(e)$  is the maximum energy overlap between a partial cluster with the edge and given basis functions.

To find a set of edges to form  $N_c$  clusters such that total weight of the edges,  $W(e)$ , is minimized, we design an algorithm based on a greedy local heuristics. The algorithm starts from  $N_c$  initial nodes for clusters, one for each cluster; we deterministically chose  $N_c$  nodes located on the grid with equal distance to the adjacent starting nodes. At every iteration, we find edges connected to any of clusters and compute the weights,  $W(e)$ . Then, an edge with the minimum weight is added to the cluster. This procedure continues until every node is assigned to one of the  $N_c$  clusters. For details of the algorithm, refer to Algorithm 1.

The algorithm is similar to Prim's algorithm [24] for finding Minimum Spanning Tree (MST). Given weights of edges, we choose an edge with minimum weight at every step like Prim's algorithm. However, we have additional requirements as compared to Prim's algorithm. First, our algorithm finds  $N_c$  clusters with minimum total edge weights instead of a MST. Thus, an edge with the minimum weight is added to one of clusters to which the edge is connected rather than to a tree. Second, Prim's algorithm runs under the assumption that weight of edges do not change but, in our problem, the edge weights should be updated at every step.

Once a node with the minimum weight is added to a cluster, energy overlap of the edges connected to the cluster changes so that the edge weights should also change. For example, as shown in Fig. 2, a cluster,  $C_1 = \{v_1, v_2, v_3\}$ , is constructed after two iterations without update so that the total weight will be 6 by adding an edge weight of  $(v_1, v_2)$  to that of  $(v_1, v_3)$ . With update of edge weights, however, the edge weight of  $(v_1, v_3)$  increases at the  $2^{nd}$  iteration because  $v_2$  and  $v_3$  are overlapped with the same cluster,  $B_1$ . Thus,  $v_4$  will be added to the cluster so that the total weights of  $C_2 = \{v_1, v_2, v_4\}$  is 3, smaller than the weight of  $C_1$ .

Given a sparsifying basis and positions of sensors, Algorithm 1 generates a set of  $N_c$  clusters by minimizing edge

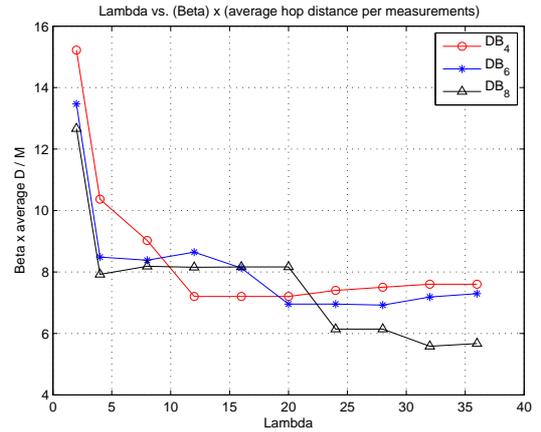


Fig. 4.  $\lambda$  vs.  $\beta \times$  average  $D$  per measurement for three different 2D Daubechies bases with a single level decomposition. 256 sensors with 16 clusters are considered. The y-axis indicates the estimated total transport cost because the number of measurements is proportional to  $\beta$  by Theorem 3.4. The optimal  $\lambda^*$  minimizes the total transport cost for a given basis. As  $E_B$  increases, the results show that optimal  $\lambda^*$  also increases.

weights associated with a  $\lambda$ . Since edge weights,  $W(e)$ , are a linear combination of  $D(e)$  and  $\beta(e)$  with a design parameter  $\lambda$ , different choice of  $\lambda$  can affect performance of the algorithm.

### B. Optimal Choice of $\lambda$

The design parameter  $\lambda$  controls the balance between two competing terms:  $\beta(e)$  and  $D(e)$ . As  $\lambda$  increases,  $\beta(e)$  is more dominant factor on  $W(e)$  than  $D(e)$  so that the edges with smaller  $\beta(e)$  have higher chance to be added to a cluster, which means that the spatial extent of clusters increases. As shown in Fig. 3, as  $\lambda$  increases,  $\beta$  decreases but  $D$  increases. However, it is not clear how to determine the best  $\lambda$  with respect to both reconstruction accuracy and transport cost.

For the same level of reconstruction quality, with larger  $\lambda$ , the minimum number of measurements,  $M$ , decreases thanks to the decrease of  $\beta$  while the distance between nodes within the same cluster,  $D$ , increases. Since transport cost is determined by  $C = MD$ , different  $\lambda$  affects the overall transport cost and there could exist an (or a range of) optimal  $\lambda$  that achieves the largest energy savings with the same level of reconstruction accuracy. To show the existence of optimal  $\lambda^*$ , for simplicity, we will derive  $\lambda^*$  in a toy example.

In the example, we consider  $K$ -sparse signal and  $N_c$  clusters from  $N$  sensors deployed with grid topology in the field with size of  $F$  by  $F$ . The distance between two nodes is defined as the number of hops in a multi-hop network. Given sparsifying basis and positions of sensors, we make two assumptions. First, we assume that all the basis functions have the same spatial extent, e.g. as would be the same with Haar basis with a single level of decomposition. We define the number sensors covered by each basis function as  $E_B$  so that the hop distance between basis functions is  $\sqrt{E_B}$  on average. Second, the energy of basis functions is uniformly spread over space. If each basis function covers  $E_B$  sensors in space, the energy corresponding to any one sensor is equal to  $1/E_B$ . Thus, if a

cluster is overlapped with a basis function by one sensor,  $\beta$  can increase by  $1/E_B$ .

Our derivation of  $\lambda^*$  in this example starts from a situation such that  $\beta$  is maximally minimized; all clusters are overlapped with basis functions by one sensor so that  $\beta$  is minimized to  $1/E_B$ . By Theorem 3.4, the minimum number of measurements for perfect reconstruction,  $M_0 = \frac{KN_c \log^2 N}{E_B}$ . Now, if we exchange two sensors that belong to different clusters, then  $\beta$  will increase by  $1/E_B$ , so that  $M_0$  increases by  $\Delta M$ . Meanwhile, the hop distance,  $D$ , will decrease by  $\sqrt{E_B}$  on average. Note that the exact amount of decrease of  $D$ ,  $\Delta D$  depends on the positions of two switching sensors. Thus, we examine the change of total cost with respect to different locations of two sensors in a switch.

We first define the distance *in hops* between two sensors in the same basis function after the switch as  $\alpha$ . For the switch, we can compute the change of transport cost,  $\Delta C = M_0 D_0 - (M_0 + \Delta M)(D_0 - \Delta D)$ , and derive a condition that transport cost can decrease,  $\Delta C \geq 0$ . Algebraically, it can be seen that if  $\alpha < \alpha^* = \sqrt{E_B}/2$ , we can achieve transport cost savings. Therefore,  $\lambda^*$  should promote the cost-saving switch ( $\alpha \leq \alpha^* - 1$ ) but prevent the others ( $\alpha \geq \alpha^*$ ). By comparing edge weights in each situation, we can derive a bound of  $\lambda^*$ ,

$$\frac{1}{2}E_B^{\frac{3}{2}} < \lambda^* < \frac{1}{2}E_B^{\frac{3}{2}} + 2E_B \quad (6)$$

The derivation in this example shows there exists a range of optimal  $\lambda^*$  achieving savings of total transport cost at the same level of reconstruction quality. As shown in (6), the bound is proportional to the spatial extent of basis functions,  $E_B$ . As  $E_B$  increases,  $\beta(e)$  increases but  $D(e)$  does not change, which means  $\beta$  is over-emphasized. Therefore,  $\lambda^*$  should increase to balance out the dominance of  $\beta(e)$ . But, the bound does not depend on physical characteristics of the network described by  $F$  and  $N$  because the number of hops between nodes remains the same irrespective of them. As shown in Fig. 4, the increase of  $E_B$  also increases  $\lambda^*$  that minimizes the estimated total transport cost. But,  $\lambda^*$  in (6) does not perfectly match with that by simulation because the energy of basis functions in Fig. 4 is not uniformly spread over space. Note that since the bound is derived under some unrealistic assumptions for simplification, we need to generalize it to find  $\lambda^*$  in practical situation. But, this is beyond of scope of this paper thus we will remain the generalization as future work.

## V. SIMULATION

The simulation consists of two parts. First, we verify Theorem 3.4 by examining the correlation between the estimated  $M_{est}$  by Theorem 3.4 and the minimum  $M_{sim}$  measured by simulation. Second, we evaluate the performance of the joint optimization in terms of transport cost and reconstruction quality. Each part of the simulation is based on different clustering scheme and different evaluation for reconstruction accuracy. The details of different simulation environments will be described in following subsections.

TABLE I  
CORRELATION COEFFICIENT,  $r \in [-1, 1]$ , BETWEEN  $M_{est}$  AND  $M_{sim}$

$N = 1024, N_c = 16$			
	$DB_4$	$DB_6$	$DB_8$
K=20	0.468	0.429	0.513
K=38	0.567	0.503	0.621
K=55	0.694	0.589	0.617

We first introduce common simulation environment in both parts of simulation. We consider 3 different 2-D Daubechies basis with  $2^{nd}$  level decomposition:  $DB_4$ ,  $DB_6$ , and  $DB_8$ . For each of the basis, we reconstruct 100 synthesized data with three different sparsity ( $K$ ): 20, 38, and 55. Each data is generated by  $K$  random coefficients for each basis. In the network, 1024 nodes are deployed on the square grid in a region of interest and FC located at center of the field collects measurements from sensors with error free communication. For given clusters in a network, we do not assume any priority is given to specific clusters for measurements, i.e., we collect the same number of localized measurements for each cluster. With  $M$  measurements, data is jointly reconstructed with gradient pursuit for sparse reconstruction (GPSR) [25].

### A. Reconstruction accuracy and $\beta$

To verify Theorem 3.4, we first measure the minimum number of measurements,  $M_{sim}$ , required for perfect reconstruction in our simulation. To measure  $M_{sim}$ , we evaluate reconstruction accuracy by perfect reconstruction rate ( $Prr$ ). We consider that each data is perfectly reconstructed if  $\max|x - \hat{x}| < 10^{-3}$  then  $Prr$  is computed over 100 synthesized sparse data. The minimum  $M_{sim}$  for perfect reconstruction is the smallest  $M$  that satisfies the perfect reconstruction rate larger than 0.99.

To collect spatially-localized projections,  $N$  sensors are separated into  $N_c$  non-overlapped clusters with the same size; every cluster contains  $N/N_c$  sensors. We consider 20 different clustering schemes. For each clustering scheme, we divide  $N$  sensors in the field into 16 localized clusters with radial shape going from FC to the boundary of the network. To generate 20 different clustering schemes, we rotate 16 clusters with a certain angle for each clustering scheme.

To estimate  $M_{est}$ , for given  $\Psi$  and  $K$ , we first compute  $\beta$  for each clustering scheme so that we have 20 different values of  $\beta$ , one for each clustering scheme. Then,  $M_{est}$  is computed by Theorem 3.4. To check if  $M_{est}$  and  $M_{sim}$  are correlated to each other, we use Pearson's linear correlation coefficient,  $r \in [-1, 1]$  [26]. Note that we need to consider a

TABLE II  
CORRELATION COEFFICIENT,  $r \in [-1, 1]$ , BETWEEN MAXIMUM ENERGY OVERLAP AND MSE FOR EACH BASIS

$N = 1024, N_c = 16, M = 410$			
	$DB_4$	$DB_6$	$DB_8$
K=20	0.334	0.560	0.499
K=38	0.458	0.633	0.564
K=55	0.448	0.614	0.664

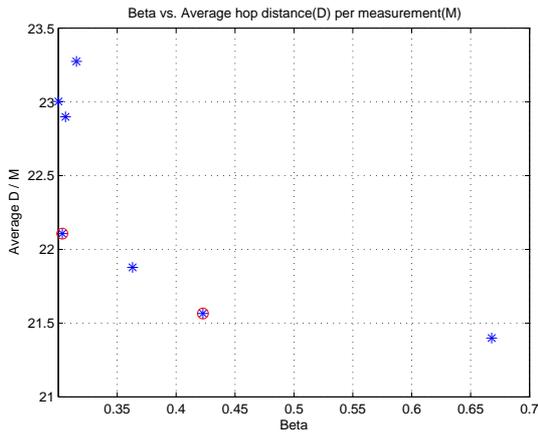


Fig. 5.  $\beta$  vs. average number of hops per measurement. Each point corresponds to the result of Algorithm 1 with different  $\lambda$ . The points with red circle are chosen for the evaluation of transport cost and MSE in Fig. 6

constant value to explain  $O$  operator for  $M_{est}$  as 1 because the correlation coefficient is a measure of the linear dependence between two variables so that does not change regardless of the constant.

As shown in Table I, the correlation value is around 0.55 for different  $K$  and  $\Psi$ , which shows that  $\beta$  affects reconstruction accuracy in terms of the minimum number of measurements for perfect reconstruction. However, our bound is not tight enough to estimate the exact number of measurements because there exists a gap between  $M_{est}$  and  $M_{sim}$ . The reason is that  $\beta$  measures the worst energy overlap so that  $\beta$  can lose accuracy when a few large energy overlap exist. However,  $\beta$  is a useful metric because, based on  $\beta$ , we can compare different clustering schemes and also design a clustering scheme by optimizing  $\beta$  as discussed in Section IV.

Since the maximum energy overlap is to measure worst case energy overlap,  $\beta$  can be misleading in a situation where a small number of large energy overlap dominate the others. For example, suppose there exists an energy overlap with value of 1 (perfect overlap) but the others are relatively small, that is,  $\beta = 1$ . However, successful reconstruction is possible with high quality if the basis function associated to the large energy overlap is not in the data support. Therefore, it would be meaningful to examine the impact of maximum energy overlap of each basis function and clusters on the error occurred by the basis function.

For each basis function, we define  $\beta_i$  measuring the maximum energy overlap between the  $i^{th}$  basis function and clusters. Thus,  $\beta_i$ 's show distribution of energy overlap with respect to basis functions. Table II shows the average correlation between  $\beta_i$  and error occurred by corresponding basis function. The simulation results show that basis functions with larger  $\beta_i$  generate more errors. This implies that basis functions with concentrated energy on fewer clusters have lower probability to be identified by spatially-localized measurements.

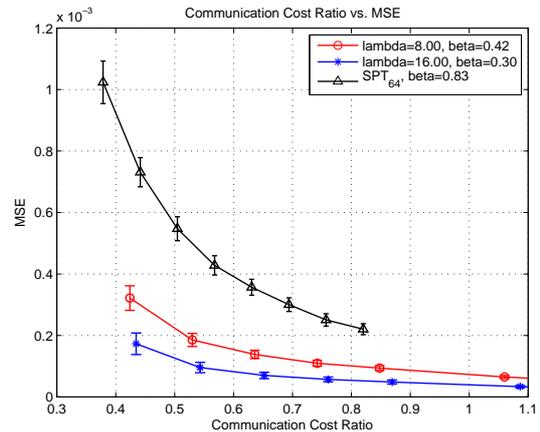


Fig. 6. Transport cost ratio vs. MSE. The x-axis is the ratio of total transport cost of spatially-localized CS to the cost for raw data gathering without any manipulation. We compare performance of results by joint optimization with two different  $\lambda$ 's with that of  $SPT_{64}$  in [18]

### B. Joint optimization

To evaluate the joint optimization, we use mean squared error (MSE) as a metric for reconstruction accuracy. This is because, in practice, we are more interested in the level of error associated to transport cost. For the cost evaluation, transmission cost is computed by  $\sum (\text{bit}) \times (\text{number of hops})$  but the work could be extended to use more realistic cost metrics. Since we allocated the same number of bits for each measurement, transmission cost depends on the product of the number of measurements with the distance in hops. The cost ratio in our simulation is the ratio to the cost for raw data gathering without compression.

In our simulation, we consider a signal with sparsity of 38 in 2D Daubechies-4 basis with the  $2^{nd}$  level of decomposition. Located at center of the field, FC collects  $M$  measurements from 64 clusters. For energy efficiency, measurements from each cluster are routed to FC along shortest path for energy efficiency. For comparison with other CS approaches, we consider a clustering scheme based on shortest path tree (SPT) that showed the best performance in [18].

As discussed in Section IV, in general, smaller  $\beta$  can be achieved by the increase of distance between nodes in the same cluster. The tradeoff can also be observed in Fig. 5; as  $\lambda$  increases, we can achieve smaller  $\beta$  but result in larger hop distance per measurements as discussed in Section IV. In addition, Fig. 5 shows that, as  $\lambda$  increases,  $\beta$  decreases quickly but it is saturated at some point. After that, transport cost increases without improvement of  $\beta$ . Thus, we can expect that one of  $\lambda$ 's around the saturation will correspond to optimal  $\lambda^*$  showing the best performance in terms of total transport cost and reconstruction accuracy.

Figure 6 shows the overall performance with different  $\lambda$ . Each curve shows the average MSE over 100 synthesized data and the variation of reconstruction accuracy is expressed by three times of standard deviation. As expected, a  $\lambda$  with the

value of 16 located at the sharp transition in Fig. 5 shows the best performance. With the best  $\lambda$  and 64 clusters of 1024 sensors, we can achieve 40% cost saving with respect to raw data-gathering with small mean squared error ( $\leq 1 \times 10^{-4}$ ). Compared with SPT-based clustering scheme, our clustering scheme with joint optimization performs better. The clusters in SPT-based clustering scheme consume less energy to construct a measurement. However, the savings in transport cost is compensated by larger number of measurements required for the same level of reconstruction quality which can be explained by large value of  $\beta (= 0.83)$ .

## VI. CONCLUSION

To achieve energy efficient data gathering in WSN, we exploit a sparse and spatially-localized CS measurement system that is aware of transport cost per measurement by constructing measurements within spatially-localized clusters. However, while the spatially-localized measurement system leads to lower transport cost, it is not obvious how it affects reconstruction accuracy. Thus, we first introduced a metric to measure the maximum energy overlap between clusters and basis functions,  $\beta$ . Then we showed that the metric has an impact on reconstruction accuracy with respect to the number of measurements for perfect reconstruction. By exploiting the tradeoff between  $\beta$  and distance between sensors in clusters, we propose a centralized iterative algorithm with a design parameter,  $\lambda$ , to construct clusters that are jointly aware of energy efficiency and reconstruction quality. Our simulation results show that, with an appropriately chosen  $\lambda$ , we can achieve significant savings in transport cost with small reconstruction error.

## APPENDIX

### A. Proof of Proposition 3.3

In this section, we present the details of the proof of Proposition 3.3. The goal is to bound the coherence in terms of  $\beta$  with high probability. The sketch of proof is similar to that of SRM in [22] but the details are different. First, in our problem, the randomness comes from being due to the coefficients of the measurement matrix,  $\Phi(i, j)$ , instead of the uniform permutation as in SRM. Second, we consider an additional quantity,  $\beta$ , as well as the number of clusters,  $N_c$ .

Before going into the details of proof, we approximate a bound related to  $E[U_i^2(j, k)]$  for simplicity. Basically,  $U_i(j, k)$  is the inner product between the  $j^{\text{th}}$  row of  $\Phi_i$  and  $k^{\text{th}}$  column of  $\Psi$  as shown in (1). Suppose that  $X_l = \Phi_i(j, l)\Psi_i(l, k)$  so that  $U_i(j, k) = \sum_l \Phi_i(j, l)\Psi_i(l, k) = \sum_l X_l$ . Note that  $\gamma = \max_{i,j} |\Psi(i, j)| \leq 1/\sqrt{\log N}$  by assumption and  $\Phi(j, k) \sim N(0, N_c/N)$ . Therefore,  $|X_l|$  can be bounded by the product of  $\gamma$  and three times of the standard deviation of  $\Phi(j, k)$  with high probability.

$$|X_l| = |\Phi_i(j, l)\Psi_i(l, k)| \leq 3\sigma\gamma \leq 3\sqrt{\frac{N_c}{N \log N}}$$

The next approximation is concerned with  $E[U_i^2(j, k)]$ . Note that  $\text{Var}(U_i(j, k))$  is equal to  $E[U_i(j, k)^2]$  because  $E(U_i(j, k)) = 0$ .  $\text{Var}(U_i(j, k))$  can be approximated with respect to  $\beta$ .

$$\begin{aligned} \text{Var}(U_i(j, k)) &= E[U_i(j, k)^2] \\ &= E \left[ \left( \sum_{l=1}^{N/N_c} \Phi_i(j, l)\Psi_i(l, k) \right)^2 \right] \\ &= E \left[ \sum_{l=1}^{N/N_c} \Phi_i^2(j, l)\Psi_i^2(l, k) \right] \\ &= \sum_{l=1}^{N/N_c} E[\Phi_i^2(j, l)]\Psi_i^2(l, k) = \sum_{l=1}^{N/N_c} \frac{N_c}{N} \Psi_i^2(l, k) \\ &\leq \frac{N_c}{N} \beta \quad (\text{by definition of } \beta) \end{aligned} \quad (7)$$

In (7), the cross terms are zero since  $E(\Phi_i(j, l)) = 0$  and  $\Phi_i(j, l)$ 's are independent. Now, we present the details of the proof of Theorem 3.3. The main technical tools are large deviation inequalities of sum of independent random variables. Specifically, this is derived from Bernsteins deviation inequality and a union bound for the supremum of a random process.

$$\begin{aligned} \Pr[\mu \leq \alpha] &= \Pr[\max_{j,k} |\mathbf{U}(j, k)| \leq \alpha] \\ &= 1 - \Pr[\max_{j,k} |\mathbf{U}(j, k)| > \alpha] \\ &= 1 - \Pr\left(\bigcup_i \max_{j,k} |\mathbf{U}_i(j, k)| > \alpha\right) \\ &\geq 1 - \sum_{i=1}^{N_c} \sum_{j=1}^{N/N_c} \sum_{k=1}^N \Pr(|\mathbf{U}_i(j, k)| > \alpha) \quad (\text{by Union Bound}) \end{aligned}$$

The tail probability of  $U_i(j, k)$  is obtained by Theorem A.1. Note that  $U_i(j, k)$ 's are independent random variables because, by assumption,  $\Phi(i, j)$ 's are i.i.d. Gaussian random variables with zero mean.

*Theorem A.1:* Bernstein's inequality [23]

If  $X_1, X_2, \dots, X_n$  are independent (*not necessarily identical*) and zero-mean random variables, and  $|X_i| \leq M, \forall i$ , then

$$\Pr \left[ \left| \sum_{i=1}^n X_i \right| \geq \alpha \right] \leq 2 \exp \left( - \frac{\alpha^2/2}{\sum_{i=1}^n E[X_i^2] + M\alpha/3} \right) \quad (8)$$

Suppose that  $X_l = \Phi_i(j, l)\Psi_i(l, k)$ . Then, by the approximation, we have  $M = 3\sqrt{\frac{N_c}{N \log N}}$  and  $\sum E[X_i^2] = \text{Var}(U_i(j, k)) \leq \frac{N_c}{N} \beta$ . By substituting  $M$  and  $\sum E[X_i^2]$  to Bernstein's inequality,

$$\begin{aligned}
& Pr[\mu \leq \alpha] \\
& \geq 1 - \sum_{i=1}^{N_c} \sum_{j=1}^{N/N_c} \sum_{k=1}^N 2 \exp\left(-\frac{\alpha^2/2}{\text{Var}(\mathbf{U}_i(j,k)) + M\alpha/3}\right) \\
& \hspace{15em} \text{(by Bernstein inequality)} \\
& \geq 1 - 2N^2 \exp\left(-\frac{\alpha^2/2}{(\beta N_c/N) + (3\sqrt{N_c/N \log N})\alpha/3}\right) \\
& \hspace{15em} \text{(by approximation)} \\
& = 1 - \exp\left(\log 2N^2 - \frac{\alpha^2/2}{(\beta N_c/N) + (\sqrt{N_c/N \log N})\alpha}\right)
\end{aligned}$$

We need to find the minimum  $\alpha$ ,  $\alpha^*$ , such that probability,  $Pr[\mu \leq \alpha^*]$ , asymptotically goes to '1'. To achieve the asymptotic behavior, the  $2^{nd}$  term ( $exp$  term) in the last inequality should be zero with large  $N$ , which means that the  $2^{nd}$  order polynomial inside the  $exp$  term is negative. The polynomial can be expressed as a function of  $\alpha$ .

$$f(\alpha) = -\alpha^2/2 + \sqrt{\frac{N_c}{N \log N}} \log 2N^2 \alpha + \frac{\beta N_c}{N} \log 2N^2$$

To find  $\alpha^*$  for large  $N$ , we first check the characteristics of  $f(\alpha)$ . Since  $f(0) > 0$ ,  $f'(0) > 0$ , and  $f''(0) < 0$ , the larger root of  $f(\alpha)$  is the minimum  $\alpha$  such that  $Pr[\mu \leq \alpha^*] = 1$  for large  $N$ . Algebraically, the larger root of  $f(\alpha)$  is  $O(\sqrt{\beta \frac{N_c}{N} \log N})$ . Thus, we can conclude that

$$\begin{aligned}
Pr[\mu \leq \alpha^*] &= 1 - O(1/N), \\
\text{where } \alpha^* &= O\left(\sqrt{\beta \frac{N_c}{N} \log N}\right)
\end{aligned}$$

□

With the asymptotic bound of coherence above, we can derive Theorem 3.4, the minimum number of measurements for perfect reconstruction, by Theorem 3.2.

## REFERENCES

- [1] R. Cristescu, B. Beferull-Lozano, and M. Vetterli, "On network correlated data gathering," in *INFOCOM*, Mar. 2004.
- [2] S. Patten, B. Krishnamachari, and R. Govindan, "The impact of spatial correlation on routing with compression in wireless sensor networks," in *IPSN*, Apr. 2004.
- [3] P. von Rickenbach and R. Wattenhofer, "Gathering correlated data in sensor networks," in *DIALM-POMC*. ACM, Oct. 2004.
- [4] A. Ciancio, S. Patten, A. Ortega, and B. Krishnamachari, "Energy-efficient data representation and routing for wireless sensor networks based on a distributed wavelet compression algorithm," in *IPSN*, Apr. 2006.
- [5] G. Shen and A. Ortega, "Joint routing and 2D transform optimization for irregular sensor network grids using wavelet lifting," in *IPSN*, Apr. 2008.
- [6] S. Patten, G. Shen, Y. Chen, B. Krishnamachari, and A. Ortega, "Senzip: An architecture for distributed en-route compression in wireless sensor networks," in *ESSA*, Apr. 2009.

- [7] G. Shen and A. Ortega, "Transform-based distributed data gathering," in *To appear in IEEE Transactions on Signal Processing*.
- [8] R. Wagner, H. Choi, R. Baraniuk, and V. Delouille, "Distributed wavelet transform for irregular sensor network grids," in *SSP*, Jul. 2005.
- [9] M. Gastpar, P. Dragotti, and M. Vetterli, "The distributed karhunen-loeve transform," in *MMSP*, Dec. 2002.
- [10] D. L. Donoho, "Compressed sensing," in *IEEE Transactions on Information Theory*, Apr. 2006.
- [11] E. Candes, J. Romberg, and T. Tao, "Robust uncertainty principles : exact signal reconstruction from highly incomplete frequency information," in *IEEE Transactions on Information Theory*, Feb. 2006.
- [12] E. Candes and J. Romberg, "Sparsity and incoherence in compressive sampling," in *Inverse Problems*, Jun. 2007.
- [13] S. Lee, S. Patten, M. Sathiamoorthy, B. Krishnamachari, and A. Ortega, "Compressed sensing and routing in sensor networks," in *USC CENG Technical Report*, Apr. 2009.
- [14] G. Quer, R. Masierto, D. Munaretto, M. Rossi, J. Widmer, and M. Zorzi, "On the interplay between routing and signal representation for compressive sensing in wireless sensor network," in *ITA*, Feb. 2009.
- [15] M. Lustig, D. Donoho, and J. M. Pauly, "Sparse MRI: The application of compressed sensing for rapid mr imaging," in *MRM*, Dec. 2007.
- [16] M. F. Duarte, M. A. Davenport, D. Takhar, J. N. Laska, T. Sun, K. F. Kelly, and R. G. Baraniuk, "Single pixel imaging via compressive sampling," in *IEEE Signal Processing Magazine*, Mar. 2008.
- [17] W. Wang, M. Garofalakis, and K. Ramchandran, "Distributed sparse random projections for refinable approximation," in *IPSN*, Apr. 2007.
- [18] S. Lee, S. Patten, M. Sathiamoorthy, B. Krishnamachari, and A. Ortega, "Spatially-localized compressed sensing and routing in multi-hop sensor networks," in *GSN*, Jun. 2009.
- [19] M. F. Duarte, M. B. Wakin, D. Baron, and R. G. Baraniuk, "Universal distributed sensing via random projections," in *IPSN*, Apr. 2006.
- [20] D. Baron, M. B. Wakin, M. F. Duarte, S. Sarvotham, and R. G. Baraniuk, "Distributed compressed sensing," in *PrePrint*, 2005.
- [21] L. Gan, T. T. Do, and T. D. Tran, "Fast compressive imaging using scrambled block hadamard ensemble," in *EUSIPCO*, sep 2008.
- [22] T. Do, T. Tran, and L. Gan, "Fast compressive sampling with structurally random matrices," in *ICASSP*, Apr. 2008.
- [23] G. Lugosi, "Concentration-of-measure inequalities," in *Lecture notes*, Feb. 2006.
- [24] J. Kruskal, "On the shortest spanning subtree of a graph and the traveling salesman problem," in *Proceedings of the American Mathematical Society*, vol. 7, no. 1, Feb. 1956, pp. 48–50.
- [25] M. Figueiredo, R. Nowak, and S. Wright, "Gradient projection for sparse reconstruction: application to compressed sensing and other inverse problems," in *IEEE Journal of Selected Topics in Signal Processing*, 2007.
- [26] K. Pearson, "Mathematical contributions to the theory of evolution. iii. regression, heredity and panmixia," in *Philos. Trans. Royal Soc. London Ser.*, vol. 187, no. A, 1896, p. 253318.