Efficient Measurement Generation and Pervasive Sparsity for Compressive Data Gathering

Chong Luo, *Member, IEEE,* Feng Wu, *Senior Member, IEEE,* Jun Sun, *Member, IEEE,* and Chang Wen Chen, *Fellow, IEEE*

***Abstract*—We proposed compressive data gathering (CDG) that leverages compressive sampling (CS) principle to efficiently reduce communication cost and prolong network lifetime for large scale monitoring sensor networks. The network capacity has been proven to increase proportionally to the sparsity of sensor readings. In this paper, we further address two key problems in the CDG framework. First, we investigate how to generate RIP (restricted isometry property) preserving measurements of sensor readings by taking multi-hop communication cost into account. Excitingly, we discover that a simple form of measurement matrix $[I\ R]$ has good RIP, and the data gathering scheme that realizes this measurement matrix can further reduce the communication cost of CDG for both chain-type and tree-type topology. Second, although the sparsity of sensor readings is pervasive, it might be rather complicated to fully exploit it. Owing to the inherent flexibility of CS principle, the proposed CDG framework is able to utilize various sparsity patterns despite of a simple and unified data gathering process. In particular, we present approaches for adapting CS decoder to utilize cross-domain sparsity (e.g. temporal-frequency and spatial-frequency). We carry out simulation experiments over both synthesized and real sensor data. The results confirm that CDG can preserve sensor data fidelity at a reduced communication cost.**

***Index Terms*—Compressive sensing, restricted isometry property (RIP), wireless sensor networks.**

I. Introduction

CLIMATE, habitat, and infrastructure monitoring [13][36] are among the most important applications of wireless sensor networks. In these monitoring networks, sensor nodes follow the routine to periodically collect readings and transmit them to the data sink. The successful deployment of such networks is facing two main challenges. First, monitoring sensor networks are typically composed of hundreds to thousands of sensors, generating tremendous amount of sensor readings to be delivered to data sink. Second, data transmissions are generally accomplished through multi-hop routing from individual sensor nodes to the data sink. Nodes close to the sink will transmit more data and consume more energy than those at the peripheral of the network. The unbalanced energy consumption has a major impact on network lifetime, as it is most commonly defined as the time to the first node failure.

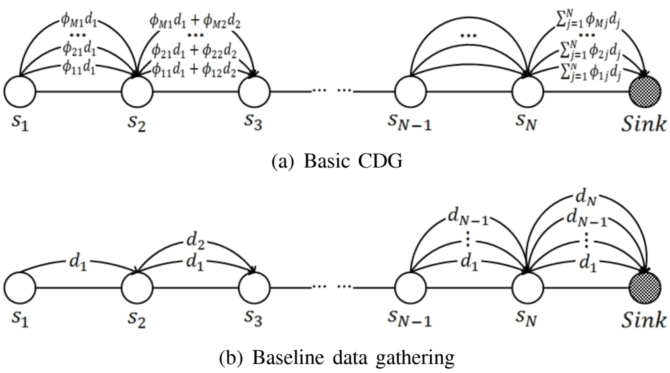

Fig. 1. Comparing CDG and baseline data gathering in a chain-type topology.

We have proposed CDG [30] for efficient data gathering in large scale monitoring sensor networks. By applying compressive sampling theory [20][4][12] to sensor data gathering, CDG achieves substantial communication cost reduction without introducing excessive computation or control overheads. Moreover, CDG elegantly disperses both communication and computation costs to all sensor nodes, which results in a natural load balancing. CDG has thus been shown to efficiently extend the lifetime of monitoring sensor networks, and to increase the network capacity.

The data gathering process of CDG is depicted in Fig. 1(a) through a simple chain-type topology. Comparing with the baseline data gathering scheme (Fig. 1(b)), CDG delivers weighted sums (or *measurements*) of sensor readings, instead of individual readings, to the data sink. To transmit the i^{th} measurement to the sink, s_1 multiplies its reading d_1 with a random coefficient ϕ_{i1} and sends the product to s_2. Then s_2 multiplies its reading d_2 with a random coefficient ϕ_{i2} and sends the sum $\phi_{i1}d_1 + \phi_{i2}d_2$ to s_3. Similarly, each node s_j contributes to the relayed message by adding its own product. Finally, the sink receives $\sum_{j=1}^{N} \phi_{ij}d_j$, a weighted sum of all the readings. This process is repeated using M sets of different weights so that the sink will receive M measurements. According to the compressive sampling theory, when the sensor readings are compressible, the sink will be able to recover N sensor readings from M random measurements even when $M < N$. Let us consider the global

Manuscript received January 14, 2010; revised May 20, 2010 and August 11, 2010; accepted August 31, 2010. The associate editor coordinating the review of this paper and approving it for publication was C.-F. Chiasserini.

C. Luo is pursuing Ph.D. degree at the Institute of Image Communication and Information Processing, Department of Electronic Engineering, Shanghai Jiao Tong University, Shanghai, China. C. Luo is also with Microsoft Research Asia, Beijing, China (e-mail: chong.luo@microsoft.com).

F. Wu is with Microsoft Research Asia, Beijing, China.

J. Sun is with the Institute of Image Communication and Information Processing, Department of Electronic Engineering, Shanghai Jiao Tong University, Shanghai, China, and with the Shanghai Key Laboratory of Digital Media Processing and Transmissions, Shanghai Jiao Tong University, Shanghai, China.

C. W. Chen is with the University at Buffalo, State University of New York.

Digital Object Identifier 10.1109/TWC.2010.092810.100063

and bottleneck communication costs for one round of data gathering and quantify them with the num-messages metric [38]. The global communication cost is defined as the total number of message transmissions in the network, and the bottleneck communication cost is defined as the maximum number of message transmissions of any single node. It is clear from Fig. 1 that when $M << N$, CDG can significantly reduce both global and bottleneck communication cost.

The success of CDG framework depends on how we tackle two critical problems, namely how to efficiently collect measurements and how to recover data from the least number of measurements. The first problem arises from the following observation. In the basic CDG scheme, nodes $s_1, s_2 ... s_{M-1}$ transmit redundant messages, and their energy consumption is even higher than in baseline data gathering. As a result, when the data are not highly compressible, and the required number of measurements (i.e. M) is large, a great deal of redundant messages will waste sensors' precious energy resources and unnecessarily occupy the shared wireless medium.

Second, successful decoding of N sensor readings from M ($M < N$) random measurements is based on the assumption that sensor readings are K-sparse in a certain domain, and the size of M is proportional to the sparsity index K. An N-dimensional signal $\mathbf{d}$ is called a K-sparse signal if there exists a known domain Ψ in which $\mathbf{d}$ can be represented by $\mathbf{d} = \Psi\mathbf{x}$ and $\mathbf{x}$ contains only K non-zero entries. Although sparsity is pervasive in sensor readings, it is sometimes hard to find the most proper domain Ψ in which sensor readings have the sparsest representation. The difficulties are three-fold. First, although a large portion of monitoring sensor networks capture natural signals which should be sparse in frequency domain, the sparsity index varies greatly when different representation bases are used. Second, sensor networks may collect compound signals, part of which is sparse in one domain and the rest of which is sparse in another domain. Last but not least, it is often assumed that sensor readings have strong correlations in adjacent neighborhood, or *adjacent correlation* as we name it. However, this assumption is not always true in reality.

This paper extends the original CDG framework and successfully addresses the two challenges outlined above. First, we design effective measurement generation matrix to avoid redundant message transmission. The matrix should also comply with the restricted isometry property (RIP) in order for successful CS recovery. Excitingly, we discover a simple form of measurement generation matrix $[I \;\; R]$ that satisfies both requirements. Second, although CDG generates random measurements regardless of data correlation patterns, we demonstrate how CDG is able to exploit data sparsity during the decoding process. In particular, we present CS decoding approaches that exploit cross-domain data sparsity, including spatial-frequency and temporal-frequency sparsity.

The rest of this paper is organized as follows: Section II reviews related work on sensor data gathering. Section III briefs the background on compressive sampling theory and its applications in wireless sensor networks. Section IV presents the design of efficient measurement generation matrix and the corresponding transmission scheme. Section V demonstrates how CDG is able to exploit data sparsity across domains. Section VI presents the experimental results on both synthesized data and real sensor data. Section VII concludes this paper with some discussions.

II. Related Work

In sensor data gathering, in-network data suppression and compression are the primary means to reduce communication cost and prolong network lifetime. The fundamental assumption is that sensor readings have spatial or spatial-temporal correlations.

A. Spatial Correlation

Since sensors are usually densely deployed in the region of interest, it is commonly assumed that sensor readings have spatial correlations or adjacent correlations. We may classify existing in-network data compression techniques into two categories, according to where the correlation information is utilized.

1) Conventional Compression: Conventional compression techniques utilize the correlation at the encoding side and require explicit data communication among sensors. The simplest form of conventional data compression is quantization and sampling. The clustered aggregation (CAG) technique [40] forms clusters based on sensing values. By grouping sensors with similar readings, CAG only transmits one reading per group to achieve a predefined error threshold. Gupta et al. [24] propose to sample only a subset of sensor nodes in each round of data gathering, and the sink is believed to be able to reconstruct data from partial readings.

More complex compression techniques involve all the sensor nodes at the encoder side and adopt entropy coding or transform coding to reduce data redundancy. Cristescu et al. [18] propose a joint entropy coding approach, where nodes use relayed data as side information to encode their readings. It is obvious that jointly encoded messages cost fewer bits than independently encoded messages. However, this approach utilizes correlations only unidirectionally. If data are allowed to be communicated back and forth during encoding, nodes may cooperatively perform transform to better utilize the correlation. Ciancio et al. [16] and Aćimović et al. [2] propose to compress sensor data through distributed wavelet transform. After the transform, nodes transmit significant coefficients to the sink and discard the small ones. Dang et al. [19] propose to exchange data within a cluster before transform coding so that correlations can be better utilized even if they are not observed between neighboring nodes. However, this work still has a basic assumption that sensors with correlated readings can communicate with each other within one hop.

The limitations of conventional compression are two-fold. First, most processing is performed at energy constrained sensor nodes. The computation complexity of entropy coding is pretty high, and transform based compression requires large amount of data exchange. Second, the correlation pattern needs to be known a priori by all sensor nodes and to be jointly considered with data routing. If the correlation pattern changes or there are abnormal readings, the compression efficiency, transmission efficiency and data fidelity will be largely affected.

2) Distributed Source Coding: Distributed source coding techniques [14][17][26] intend to reduce complexity at sensor nodes and utilize correlation at the sink. The theoretical foundation is the Slepian-Wolf coding theory [34], which claims that compression of correlated readings, when separately encoded, can achieve same efficiency as if they are jointly encoded, provided that messages are jointly decoded. This important conclusion allows sensor nodes to encode their correlated readings independently without data exchanges, and decouples data compression from routing. After encoding, each node simply sends the compressed message along the shortest path to the sink.

A prerequisite of Slepian-Wolf coding is that the global correlation structure needs to be known in order to allocate appropriate number of bits to be used by each node. This is hard to fulfill in a large-scale wireless sensor network. Yuen et al. [41] then proposes a localized Slepian-Wolf coding scheme based on the assumption of adjacent correlation. The scheme determines message size for each node based on its data correlation with one-hop neighbors. However, both global and localized distributed source coding schemes deal with static correlation patterns and are not effective in dynamic settings.

B. Spatial-Temporal Correlation

In continuous monitoring applications, sensors report their readings at short intervals. Therefore, temporal correlations can be utilized to reduce communication cost in data gathering. Some techniques suppress data based on simple comparison, and some others are aided by prediction models. Silberstein et al. [33] form the spatial-temporal suppression problem as one of monitoring node and edge constraints. A monitored node triggers a report if its value changes. A monitored edge triggers a report if the difference between its nodes' values changes. Actually, these two constraints correspond to temporal and spatial correlations respectively.

Goel and Imielinski [23] make an analogy between evolving sensor readings and MPEG videos. They create predictions for sensor nodes based on their past and surrounding readings. These predictions are represented as a prediction-model and sent to the sensor. The sensor suppresses its transmission unless its reading differs from the prediction by more than a pre-specified threshold. Chu et al. [15] adopt a joint approach. They build prediction models only based on temporal correlations. A sensor node triggers transmission if it has anomalous reading which is not correctly predicted. Spatial correlation is utilized to further suppress data when nodes in a neighborhood have similar anomalous readings.

Techniques in this category do not deal with complex correlation models. Similar to conventional compression techniques, they all assume adjacent correlation both temporally and spatially. If strongly correlated readings are not collected in consequent time slots and in immediate neighborhood, most of them cannot achieve effective data reduction.

III. COMPRESSIVE SAMPLING BACKGROUND

A. Compressive sampling theory

Compressive sampling (CS) [20][7] is an emerging research field in digital data acquisition and processing. In the conventional paradigm, natural signals are first acquired at Nyquist-Shannon sampling rate, and then compressed for efficient storage or transmission. CS shifts this paradigm by combining the two processes into a single *compressive sampling* process, greatly reducing the complexities in data acquisition. The key concept in CS theory is *sparse signals*, and the core subject matters of CS research are efficient representation and loyal recovery of sparse signals.

Definition 1 (Sparse signal): Let $\mathbf{d} = (d_1, d_2, ...d_N)^T$ be an N-dimensional signal. We say $\mathbf{d}$ is a K-sparse signal if there are only K ($K \ll N$) non-zero entries in d_is'. Further, we say $\mathbf{d}$ is a K-sparse signal in Ψ domain, if there exists a set of orthonormal basis, denote as $\Psi = [\psi_1\psi_2...\psi_n]$, $\psi_i \in \mathbb{R}^n$, in which $\mathbf{d}$ can be represented by a K-sparse vector $\mathbf{x}$:

$$\mathbf{d} = \sum_{i=1}^{n} x_i\psi_i, \text{ or } \mathbf{d} = \Psi\mathbf{x} \tag{1}$$

Compressive sampling theory states that an N-dimensional K-sparse signal can be efficiently represented by M ($M < N$) linearly projected *measurements*. In particular, let Φ be an $M \times N$ ($M < N$) matrix, then the measurements of $\mathbf{d}$ can be obtained by:

$$\mathbf{y} = \Phi\mathbf{d} \tag{2}$$

where $\mathbf{y} = (y_1, y_2, ...y_M)^T$, and y_is' are called measurements. Matrix Φ is referred to as projection matrix or *measurement matrix* in CS theory.

A question to be answered is whether it is possible and how to recover the N-dimensional signal $\mathbf{d}$ from the M-dimensional measurements $\mathbf{y}$. Candès et al. [7] have shown that when $K \leq \frac{1}{2}M$, exact recovery of $\mathbf{d}$ can be achieved through solving a combinatorial optimization problem:

$$(P0) \qquad \min_{\mathbf{x}\in\mathbb{R}^N} \|\mathbf{x}\|_{l_0} \quad s.t. \quad \mathbf{y} = \Phi\mathbf{d},\ \mathbf{d} = \Psi\mathbf{x} \tag{3}$$

This is an NP-complete problem, and numerically unstable to solve. Fortunately, $P0$ is equivalent to the following l_1-minimization problem $P1$ under certain conditions. It is known that l_1-minimization problem is more tractable, and can be solved with linear programming (LP) techniques [9][20].

$$(P1) \qquad \min_{\mathbf{x}\in\mathbb{R}^N} \|\mathbf{x}\|_{l_1} \quad s.t. \quad \mathbf{y} = \Phi\mathbf{d},\ \mathbf{d} = \Psi\mathbf{x} \tag{4}$$

The equivalence between problems $P0$ and $P1$, a.k.a. l_1/l_0 equivalence [22], relies on the incoherence property [12] between Φ and Ψ, or the restricted isometry property (RIP) [9] [8] of matrix $A = \Phi\Psi$.

Definition 2 (Restricted isometry property): Let A be an $M \times N$ matrix and let $K < N$ be an integer. Define the K-restricted isometry constant δ_K for matrix A as the smallest constant that satisfies:

$$(1 - \delta_K)\|f\|_{l_2}^2 \leq \|Af\|_{l_2}^2 \leq (1 + \delta_K)\|f\|_{l_2}^2 \tag{5}$$

for any sparse vector f with support size no larger than K. Matrix A is said to satisfy the K-restricted isometry property with δ_K.

It is obvious that an orthonormal matrix has RIP constant $\delta_K = 0$ for all $K \leq N$, since $\|Af\|_{l_2}^2 = \|f\|_{l_2}^2$ when A is orthonormal. However, when $M < N$, δ_K will be greater than zero. A small δ_K indicates that matrix A is *almost orthogonal*, and suggests a good chance that the original signal can be exactly recovered. E. J. Candès [10] [8] has established several theorems about the relationship between RIP and CS recovery.

Theorem 1: [10] Assume that the restricted isometry constants of matrix $A = \Phi\Psi$ satisfy

$$\delta_{2K} < \sqrt{2} - 1 \quad (6)$$

then solving problem $P1$ (4) recovers any sparse signal $\mathbf{d}$ with support size no larger than K.

Theorem 2: [10] Consider the situation where measurements are contaminated with noise, and ϵ be the noise level. Assume that the restricted isometry constants of matrix $A = \Phi\Psi$ satisfy

$$\delta_{2K} < \sqrt{2} - 1 \quad (7)$$

then the solution to the following relaxed l_1-minimization problem recovers the original signal with an error at most proportional to the noise level:

$$\min_{\mathbf{x}\in\mathbb{R}^N} \|\mathbf{x}\|_{l_1} \quad s.t. \quad \|\mathbf{y} - \Phi\mathbf{d}\|_{l_2} < \epsilon,\ \mathbf{d} = \Psi\mathbf{x} \quad (8)$$

It is clear that the RIP of matrix A is crucial to CS recovery. When Ψ is an orthonormal matrix, as it generally is, the restricted isometry constants of A and Φ are the same. Therefore, CS requires the measurement matrix Φ to have good RIP. It has been shown that a random matrix whose entries are i.i.d. Gaussian variables complying to $\mathcal{N}(1, 1/M)$ has good RIP [11]. This is indeed a convenient choice in distributed systems, such as monitoring sensor networks.

B. CS applications in wireless sensor networks

The emergence of CS theory has opened up a new research avenue to distributed data compression. Compared with the conventional paradigm, CS based data compression shifts most computations from the encoder to the decoder, making it a perfect fit for in-network data processing in wireless sensor networks (WSNs). Recently, CS based methods have been developed to address two classical problems in WSNs, namely data persistence and data gathering.

Data persistence problem studies how to preserve cached data in a wireless sensor network consisting of unreliable sensors. Rabbat et al. [32] leverage CS principle and use random gossiping to achieve decentralized data compression and distribution. Wang et al. [39] exploit the use of sparse measurement matrices in generating CS measurements and show that they can also achieve reliable and refinable access to data approximations.

Data gathering problem studies immediate data transmission from sensor nodes to a distant base station after each round of data collection. In a single-hop network, compressive wireless sensing (CWS) [3] is shown to be able to reduce the latency of data gathering by delivering linear projections of sensor readings through synchronized amplitude-modulated analog transmissions. Baron et al. [6] study joint sparsity models and joint data recovery algorithms of CS without considering multi-hop communication and in-network data processing. In an overview paper, Haupt et al. [25] speculate the potential of using CS principle for data aggregation in a multi-hop sensor network. Recently, Lee et al. [29][28] and Quer et al. [31] investigate CS based data gathering in a multi-hop wireless sensor network and attempt to minimize the energy consumption by jointly designing routing and *sparse* random projection. However, the optimization is highly dependent on the data correlation pattern and is application specific. Only a centralized greedy algorithm is presented by Lee et al. [29].

The proposed CDG framework considers dense CS projections. Although using dense CS measurement matrices may not achieve as much energy reduction as using its sparse counterparts, it allows the routing selection to be decoupled from CS projection. As a result, CDG framework can achieve CS based data gathering without centralized control or complicated routing design. With the introduction of CS theory in Section III-A, the two core problems of CDG that we shall address become very clear. First, at the encoder side, or the sensor nodes, how to design the measurement matrix Φ by taking both communication cost and matrix RIP into consideration. Second, at the decoder side, or the data sink, how to choose or design the representation basis Ψ so that sensor readings could be recovered from the least number of measurements.

IV. Efficient Measurement Generation for CDG

This section studies efficient measurement generation in the CDG framework. We will first define the target communication cost for data gathering in a chain-type topology, and then introduce two candidate measurement generation schemes which both achieve this target cost. However, by analyzing their corresponding measurement matrices, we find that one matrix has more favorable RIP than the other. Excitingly, this RIP-preserving matrix has a simply form, and can be easily extended to more complex topologies.

A. CDG in chain-type topology

Figure 1 has illustrated the baseline transmission scheme and the basic CDG scheme in a chain-type topology. The schematic drawing in Fig. 2 compares the communication costs of the two schemes. The x-axis is sensors' distances to the sink counted by number of hops, and the y-axis indicates the number of messages sent by each node. It is obvious that the basic CDG scheme always has a smaller bottleneck load than the baseline transmission because M is smaller than N. However, when the required number of measurements increases (e.g. from M to M'), the global communication cost of the basic CDG scheme could be even higher than that of the baseline transmission. This drawback motivates us to design more efficient measurement generation schemes. In our design, the target communication cost of each sensor node is the smaller number of message transmissions in the basic CDG scheme and in the baseline scheme. It is indicated by solid black curve in Fig. 2.

In the basic CDG scheme, the measurement matrix Φ is a full random matrix, with its entries being i.i.d. Gaussian random numbers drawn according to $\mathcal{N}(0, 1/M)$. It has been shown that a full random matrix, denoted by R in the rest of this paper, has sufficiently good RIP for CS recovery [5]. The proof of matrix R's RIP is based on a well-known property about the concentration of its extreme singular values. Let

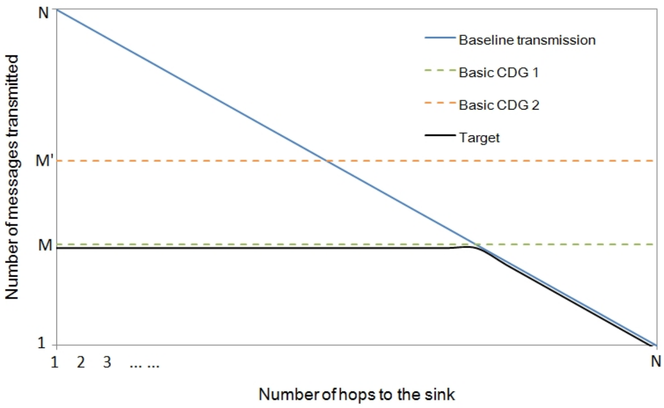

Fig. 2. Sensor transmission loads in a chain-type topology.

$\sigma_{max}(R)$ and $\sigma_{min}(R)$ be the largest and smallest singular values of an $M \times N$ ($M < N$) matrix R, it has been proved by Szarek [35] that:

$$P(\sigma_{max}(R) > 1 + \sqrt{M/N} + \tau) < e^{-\frac{N\tau^2}{2}} \quad (9)$$
$$P(\sigma_{min}(R) < 1 - \sqrt{M/N} - \tau) < e^{-\frac{N\tau^2}{2}} \quad (10)$$

Since the following inequality holds for any matrix A and vector f of matching dimension, the RIP of a full random matrix R can be easily deduced.

$$(\sigma_{min}(A))^2\|f\|_{l_2}^2 \le \|Af\|_{l_2}^2 \le (\sigma_{max}(A))^2\|f\|_{l_2}^2 \quad (11)$$

However, in order to achieve the target communication cost, a full random measurement matrix cannot be used. For the sake of simple notation, we split Φ into two parts, denoted as $\Phi = [\Phi_1\ \Phi_2]$, where Φ_1 is an $M \times M$ sub-matrix, and Φ_2 is an $M \times (N-M)$ sub-matrix. The entries in Φ_2 can still be drawn according to $\mathcal{N}(0, 1/M)$, i.e. $\Phi_2 = R$, but the entries in Φ_1 need to be re-designed. Next, we provide two candidates of Φ_1 matrix, analyze their RIP, and decide which one is better for sensor data gathering.

1) TR-CDG: A natural choice of Φ_1 is an upper triangular matrix whose entries are i.i.d. Gaussian random numbers. The measurement matrix $\Phi = [T\ R]$ can be written as:

$$\Phi = \left(\begin{array}{cccc|ccc} \phi_{11} & \phi_{12} & \cdots & \phi_{1\,M} & \phi_{1\,M+1} & \cdots & \phi_{1\,N} \\ & \phi_{22} & \cdots & \phi_{2\,M} & & & \\ & & & \phi_{3\,M} & \vdots & \vdots & \vdots \\ & \mathbf{0} & \ddots & \vdots & & & \\ & & & \phi_{M\,M} & \phi_{M\,M+1} & \cdots & \phi_{M\,N} \end{array}\right) \quad (12)$$

The transmission process to generate $[T\ \ R]$ matrix is straightforward. Node s_1 transmits $\phi_{11}d_1$ to s_2, and s_2 adds its own product to generate measurement $y_1^2 = \phi_{11}d_1 + \phi_{12}d_2$. Node s_2 also produces measurement $y_2^2 = \phi_{22}d_2$, and transmits both measurements to s_3. Obviously, TR-CDG achieves the target transmission load for all sensor nodes.

When analyzing the RIP of $[T\ R]$, let us also split an N-dimensional signal f into two vectors f_1 and f_2 of dimension M and $N-M$ respectively. Then $\|\Phi f\|_{l_2}^2$ can be written into:

$$\begin{aligned} \|\Phi f\|_{l_2}^2 &= \left\| [T\,R] \begin{bmatrix} f_1 \\ f_2 \end{bmatrix} \right\|_{l_2}^2 = \|Tf_1 + Rf_2\|_{l_2}^2 \\ &= \|Tf_1\|_{l_2}^2 + \|Rf_2\|_{l_2}^2 + 2\langle Tf_1, Rf_2\rangle \quad (13) \end{aligned}$$

According to the properties of full random matrices, the second term on the right hand side of (13) is bounded by $(1-\delta_R)\|f_2\|_{l_2}^2$ and $(1+\delta_R)\|f_2\|_{l_2}^2$ with high probability [5]. Unfortunately, we are not able to find a reasonable restricted isometry constant for a triangular random matrix T, and as a result, leaving the first term unbounded.

Lemma 3: Denote T_M as an $M \times M$ upper triangular matrix with non-zero entries drawn from $\mathcal{N}(0, 1/M)$. Denote $\sigma_{max}(T_M)$ and $\sigma_{min}(T_M)$ as the largest and smallest singular values of T_M. Then, the 2-norm condition number of T_M, denoted by $\kappa_M(T_M) = \frac{\sigma_{max}(T_M)}{\sigma_{min}(T_M)}$, goes to 2^M almost surely as $M \to \infty$.

Proof: Matrix T_M has the same set of singular values as its transpose T'_M, which is a lower triangular matrix. In addition, matrix T'_M has the same condition number as $M \cdot T'_M$, whose non-zero entries are drawn from $\mathcal{N}(0, 1)$. Viswanath and Trefethen [37] have proved that for a lower triangular matrix whose non-zero entries drawn from $\mathcal{N}(0, 1)$, its 2-norm condition number goes to 2^M almost surely as $M \to \infty$. ■

This exponential growth of κ_M with M is in striking contrast to the linear growth of the condition numbers of a full random Gaussian matrix with M. Consequently, a triangular random matrix does not satisfy RIP.

Theorem 4: Let T be an $M \times M$ ($M > 2$) upper triangular matrix with non-zero entries drawn according to $\mathcal{N}(0, 1/M)$, and let f be an M-dimensional sparse vector with support size no larger than K. Matrix T does not satisfy the K-restricted isometry property with $\delta_K \in [0, 1-\epsilon]$ for any $\epsilon > 0$.

Proof: This theorem can be proved by contradiction. Suppose there exists a $\delta_K \in [0, 1-\epsilon]$ with which matrix T satisfies the K-restricted isometry property. Then, according to the RIP definition, we have:

$$(1-\delta_K)\|f\|_{l_2}^2 \le \|Tf\|_{l_2}^2 \le (1+\delta_K)\|f\|_{l_2}^2 \quad (14)$$

It is equivalent to:

$$\begin{aligned} (1-\delta_K)\|f\|_{l_2}^2 \le (\sigma_{\min}(T))^2\|f\|_{l_2}^2 \le (\sigma_{\max}(T))^2\|f\|_{l_2}^2 \\ \le (1+\delta_K)\|f\|_{l_2}^2 \quad (15) \end{aligned}$$

This suggests that:

$$\kappa_M = \frac{\sigma_{\max}(T)}{\sigma_{\min}(T)} \le \sqrt{\frac{1+\delta_K}{1-\delta_K}} \le \sqrt{\frac{2-\epsilon}{\epsilon}} \quad (16)$$

which contradicts with Lemma 3 when M is sufficiently large. ■

The unfavorable RIP of matrix T casts a shadow over the RIP of the $[T\ \ R]$ matrix. Later, we will see from the experimental results that such a matrix has poor CS recovery performance indeed.

2) IR-CDG: The second choice for Φ_1 is the identity matrix. Let us denote $\Phi = [I\,R]$, and name the transmission scheme as IR-CDG. The Φ matrix can be written as:

$$\Phi = \left(\begin{array}{cccc|ccc} 1 & 0 & \cdots & 0 & \phi_{1\,M+1} & \cdots & \phi_{1\,N} \\ 0 & 1 & \cdots & 0 & \phi_{2\,M+1} & \cdots & \phi_{2\,N} \\ & & \ddots & & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & \phi_{M\,M+1} & \cdots & \phi_{M\,N} \end{array}\right) \quad (17)$$

By using $[I\,R]$ as the measurement matrix, the first M sensor nodes simply transmit their original sensor readings to node s_{M+1}. Upon receiving the reading from sensor s_i, s_{M+1} computes the i^{th} product and transmits $d_i + \phi_{i\,M+1} d_{M+1}$ to the next node. In IR-CDG, the first M nodes do not have any computation load, and the rest of nodes have the same computation and communication load as in the basic CDG scheme.

Intuitively, matrix $[I\ R]$ does not carry as much information as matrix $[T\ R]$. However, matrix $[I\ R]$ does have a good RIP and is a better choice than $[T\ R]$ for CS projection. The experiments in Section VI-A will show that using matrix $[I\ R]$ as the measurement matrix can achieve similar CS reconstruction performance as using a full random matrix.

Theorem 5: Let R be an $M \times (N - M)$ matrix with elements drawn according to $\mathcal{N}(0, 1/M)$ and let I be an $M \times M$ identity matrix. If

$$M \geq C_1 K \log\left(\frac{N}{K}\right) \quad (18)$$

then $[I\ R]$ satisfies the RIP of order K with probability exceeding $1 - 3e^{-C_2 M}$, where C_1 and C_2 are constants that depend only on the desired RIP constant δ.

The proof of this theorem is included in Appendix. In order to conceptually understand this theorem, recall that RIP is a metric that describes how close to orthogonal a matrix is. It is obvious that the columns in the identity matrix I is orthogonal to each other. Adding these orthogonal columns does not affect the RIP of full random matrix.

B. Extension to tree-type topology

In many wireless sensor networks, sensors spread out in a two-dimensional area, and the shortest paths from sensors to the data sink present a tree structure. In our previous work [30], we have discussed how to apply the basic CDG scheme to homogeneous networks with tree-type routing structure. In particular, CDG is performed based on subtrees, each of which is led by a direct neighbor of the sink. Figure 3 shows a subtree led by node s_1. The sink solves the set of linear equations from each subtree separately. Assume the i^{th} subtree contains N_i sensor nodes, and the readings can be recovered from M_i measurements. Then in the basic CDG scheme, every node in this subtree transmits M_i messages.

Similar to the chain-type topology, matrix $[I\,R]$ can be used for measurement generation in subtrees. In the i^{th} subtree, at most M_i nodes can send one original reading instead of M_i weighted sums. In the example given in Fig. 3, assume $M_i = 4$, then nodes s_2, s_6, s_7 and s_8 can transmit their original readings. The measurements received by the sink can

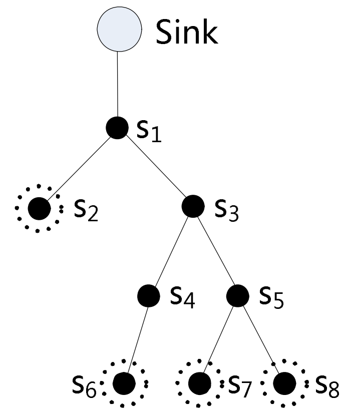

Fig. 3. A subtree led by s_1 for compressive data gathering.

be represented by:

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix} = \begin{pmatrix} \phi_{11} & 1 & \phi_{13} & \phi_{14} & \phi_{15} & 0 & 0 & 0 \\ \phi_{21} & 0 & \phi_{23} & \phi_{24} & \phi_{25} & 1 & 0 & 0 \\ \phi_{31} & 0 & \phi_{33} & \phi_{34} & \phi_{35} & 0 & 1 & 0 \\ \phi_{41} & 0 & \phi_{43} & \phi_{44} & \phi_{45} & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} d_1 \\ d_2 \\ \vdots \\ \vdots \\ d_8 \end{pmatrix} \quad (19)$$

Different from chain-type topology, the nodes which send original readings do not have to be the first M_i nodes in vector $\mathbf{d}$. Since shuffling the columns of the measurement matrix does not change its RIP, we can freely choose these nodes, and assign them with a different sequence number between 1 to M_i. For the sake of communication cost reduction, it is preferred that these nodes are leaf nodes or close to the peripheral of the routing tree. Ideally, IR-CDG can reduce up to $M_i(M_i-1)$ transmissions in the i^{th} subtree when compared to the basic CDG scheme.

V. Exploiting Pervasive Sparsity

This section studies how to exploit various sparsity of sensor readings in the CDG framework. Compressive sampling performs the same random projection operation for any sparse signal. Correlation information is utilized only at the decoder, and reflected in the representation basis Ψ. As we know, the number of measurements needed for CS recovery is proportional to the sparsity index of the signal. Therefore, if we are able to find the most proper Ψ, on which $\mathbf{d}$ is represented by the sparsest vector $\mathbf{x}$, the number of measurements being sent to the sink can be minimized.

For spatially smooth sensor readings, performing wavelet transform can generate a sparse representation. Usually, multi-level wavelet de-correlation can generate a sparser representation than single level de-correlation. The CDG framework allows us to exploit the data sparsity after multi-level wavelet de-correlation without additional cost. The sink can pre-compute the representation basis Ψ for any reasonable level of wavelet de-correlation and pick a proper one for each round of data collection.

In addition, the CDG framework offers great flexibility in exploiting cross-domain sparsity patterns, some of which are hard to be utilized by conventional in-network compression schemes such as random sampling or transform based compression. Next, we will present the approaches to utilize three typical types of cross-domain sparsity that are observed in monitoring sensor networks.

A. Temporal-Frequency Sparsity

In continuous monitoring applications, sensor readings have spatial-temporal correlations, i.e. the change of sensor readings does not vary much in a close neighborhood. Let d_i^t and d_i^{t+1} be two continuous readings obtained by node i at time instance t and $t+1$. Let $\Delta d_i^t = d_i^{t+1} - d_i^t$ be the difference of the two values. If two nodes i and j are adjacent to each other, then Δd_i^t and Δd_j^t should have similar values. In another word, if Δd_i^ts' are properly ordered into vector $\Delta \mathbf{d}^t$, $\Delta \mathbf{d}^t$ should be sparse in frequency domain. This is why we name this type of sparsity as temporal-frequency sparsity. Let $\mathbf{d}^t$ and $\mathbf{d}^{t+1}$ be the vector of all sensor readings collected at time instance t and $t+1$, then we have:

$$\Delta \mathbf{d}^t = \mathbf{d}^{t+1} - \mathbf{d}^t \tag{20}$$

$$\Delta \mathbf{d}^t = \Psi \mathbf{x}, \qquad \mathbf{x} \text{ is sparse} \tag{21}$$

Assume that $\mathbf{d}^t$ has been recovered by the sink. At time instance $t+1$, the sink collects measurements $\mathbf{y}^{t+1}$ for $\mathbf{d}^{t+1}$. With $\mathbf{y}^{t+1}$ and $\mathbf{d}^t$, we can compute the measurements for $\Delta \mathbf{d}^t$.

$$\mathbf{y}^{t+1} = \Phi \mathbf{d}^{t+1} \tag{22}$$

$$\Delta \mathbf{y}^t = \mathbf{y}^{t+1} - \Phi \mathbf{d}^t = \Phi \Delta \mathbf{d}^t \tag{23}$$

$\Delta \mathbf{d}^t$ can be solved by l_1-minimization with (21) and (23) as constraints. The sensor readings at time instance $t+1$ can then be computed as $\mathbf{d}^{t+1} = \mathbf{d}^t + \Psi \mathbf{x}$. Please be noted that the number of measurements needed is determined not by the sparsity of $\mathbf{d}^t$, but by the sparsity of $\Delta \mathbf{d}^t$.

B. Spatial-Frequency Sparsity

Under normal circumstances, sensor readings are spatially smooth and sparse in frequency domain. However, one of the main purposes of sensor network is to detect abnormal events. When abnormal events are captured, sensor data sparsity in frequency domain will be compromised. In CDG, we tackle this problem by designing an overcomplete representation basis. In compressive sampling theory, Donoho et al. [21] have shown the possibility of stable recovery under a combination of sufficient sparsity and favorable structure of the overcomplete system.

In this case, since the appearance of abnormal readings is usually sporadic, sensor data with abnormal readings can be viewed as a sparse signal in spatial-frequency domain. The vector of sensor readings $\mathbf{d}$ can be conceptually decomposed into two vectors:

$$\mathbf{d} = \mathbf{d_f} + \mathbf{d_s} \tag{24}$$

where $\mathbf{d_f}$ contains the normal part of sensor readings which are sparse in a frequency domain, and $\mathbf{d_s}$ contains the deviated values of abnormal readings which is sparse in spatial domain. Let Ψ be a proper transform matrix for frequency analysis, e.g. a wavelet transform matrix, then (24) can be rewritten into:

$$\mathbf{d} = \Psi \mathbf{x_f} + I \mathbf{x_s} = [\Psi \ I] \begin{bmatrix} \mathbf{x_f} \\ \mathbf{x_s} \end{bmatrix} = \Psi' \mathbf{x}' \tag{25}$$

where I is the identity matrix. Since both $\mathbf{x_f}$ and $\mathbf{x_s}$ are sparse vectors, the $2N$-dimensional vector $\mathbf{x}'$ is sparse too. Now $\mathbf{d}$ has a sparse representation in Ψ' domain, so it can be reconstructed through l_1-minimization. Suppose $\tilde{\mathbf{x}}$ is the solution to $\mathbf{x}'$, then the original sensor readings can be computed by $\tilde{\mathbf{d}} = \Psi' \tilde{\mathbf{x}}$. $\mathbf{x}'$ can also be written into $\mathbf{x}' = [\tilde{\mathbf{x_f}} \ \tilde{\mathbf{x_s}}]^T$. The large non-zero values in $\tilde{\mathbf{x_s}}$ indicate the positions of abnormal readings.

The technique of designing overcomplete representation basis can be used to handle other types of cross-domain sparsity. It can also be jointly adopted with the substitution technique presented in the previous subsection. For example, when the change of sensor readings is similar in close neighborhood except for a few outliers, the combination of the two techniques can deal with the sparsity in temporal-spatial-frequency domain.

C. Sparsity at Reshuffled Ordering and Beyond

In chain-type topology, it is straightforward to compose sensor readings into signal $\mathbf{d}$ according to their hop distances to the sink. When sensors spread in a two-dimensional area, it is not so obvious how to organize all sensor readings into the $\mathbf{d}$ vector. When sensor readings have spatial correlations, they can be organized by certain spatial traversing rules in the sensing area. As a matter of fact, there are many choices to organize $d_i's$, and one of them will result in the sparsest representation in a certain transform domain.

The CDG framework offers the flexibility to exploit data sparsity at reshuffled ordering. Again, this ordering does not need to be known by sensor nodes during the data gathering process. Consider a set of smoothly changing sensor readings which are sparse in wavelet domain. The best ordering for wavelet de-correlation would be the ascending (or descending) order. Suppose the measurements $\mathbf{y}$ are collected for vector $\mathbf{d}$ in which d_is' are in an arbitrary order. Denote O_{ij} as the element operator to shuffle the i^{th} and j^{th} elements in vector $\mathbf{d}$. This operator can be achieved by left-multiplying matrix I_{ij} as following.

$$O_{ij}(\mathbf{d}) = I_{ij}\mathbf{d} = \begin{pmatrix} 1 & & & \cdots & & 0 \\ & 0 & & & 1 & \\ & & 1 & & & \\ \vdots & & & \ddots & & \vdots \\ & & & & 1 & \\ & 1 & & & 0 & \\ 0 & & & \cdots & & 1 \end{pmatrix} \begin{pmatrix} d_1 \\ d_2 \\ \\ \vdots \\ \vdots \\ \\ d_N \end{pmatrix} \tag{26}$$

d_is' can be sort into ascending order through a series of such element operators. Let us denote the new vector as $\mathbf{d}'$. $\mathbf{d}'$ can be represented by:

$$\mathbf{d}' = T_P \cdot \cdots \cdot T_2 \cdot T_1 \cdot \mathbf{d} \tag{27}$$

$$\mathbf{d}' = \Psi \mathbf{x}, \qquad \mathbf{x} \text{ is sparse} \tag{28}$$

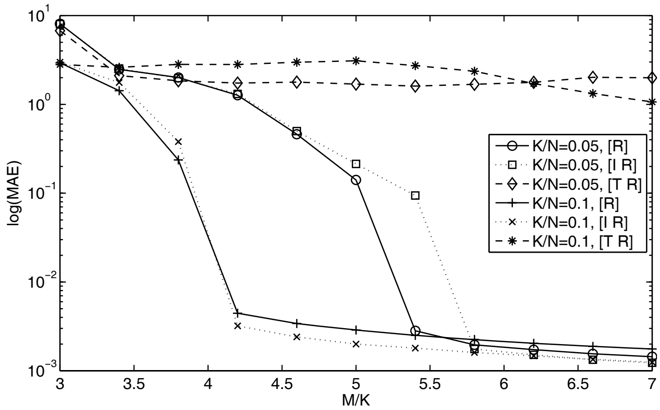

Fig. 4. Data recovery with different measurement matrix.

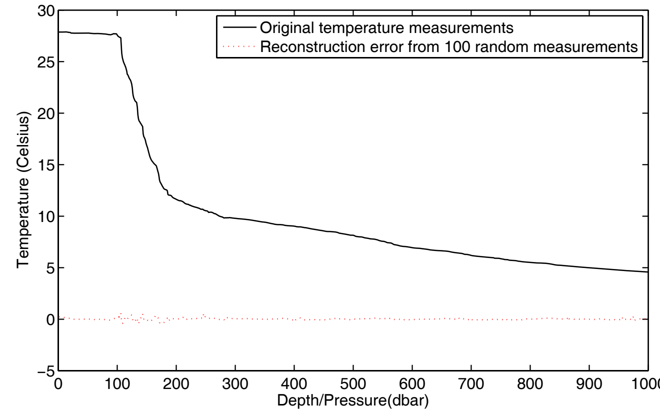

Fig. 5. Temperature data from the Pacific Sea.

where T_i is a general representation for the transform matrix of an element operator, including but not limited to order shuffling. The measurements can be rewritten into:

$$\begin{aligned} \mathbf{y} = \Phi \mathbf{d} &= \Phi \left(T_1^{-1} \cdot T_2^{-1} \cdots T_P^{-1} \cdot \mathbf{d}' \right) \\ &= \left(\Phi \cdot T_1^{-1} \cdot T_2^{-1} \cdots T_P^{-1} \right) \mathbf{d}' \end{aligned} \tag{29}$$

This means if we reshuffle the columns of the measurement matrix Φ according to the order of d_i's, then $\mathbf{x}$ can be solved from the l_1-minimization problem with (28) and (29) as constraints.

Sparsity at reshuffled ordering is useful when the ordering is known a priori or can be learned at low cost. In data gathering sensor networks, if sensor readings have strong temporal correlations, we may reorder $d_i's$ according to their values at time instance t_0. Then, sensor readings collected at time $t_0 + \Delta t$ (where Δt is a small interval), when organized in the same order, can be assumed to be a sparse signal in frequency domain. In next section, we will show how the sparsity at reshuffled ordering is used in a set of real sensor data.

We would like to point out that, among the existing works, only the CDG framework is able to exploit this type of sparsity at very low cost. As a matter of fact, when the data have good spatial correlations, it has been reported that CS based data gathering can hardly outperform randomized downsampling [29][31][28]. However, when data sparsity exhibit a compound pattern, CS based data gathering may still achieve notable cost reduction while other simple mechanisms fail completely.

VI. Experimental Results

We first evaluate the efficient measurement matrices with synthesized data, and then present the results on exploiting pervasive sparsity over real sensor data. Through two sets of real sensor data, we will show that sensor data are indeed sparse in reality. Further, data reconstruction is highly robust and efficient although real data are contaminated with noise.

A. *Measurement Matrix*

We have proved in Section IV that matrix $[I\, R]$ satisfies RIP, and matrix $[T\, R]$ does not. We verify the results with a set of synthesized data. The dimension of the signal is 500 ($N = 500$), and the non-zero entries in the signal are drawn from uniform distribution $U(10, 20)$. We experimented with two sparsity settings (K/N): 0.05 and 0.1. For each sparsity setting, we tried to recover the signal from different numbers of measurements.

Figure 4 shows the mean absolute error (MAE) of CS recovery (in logarithmic scale) when three types of measurement matrices are used. The x-axis is the ratio of the number of measurements to the number of non-zero entries in the signal. Each indicated value in the figure is averaged over 1000 test runs to avoid fluctuations. When $K/N = 0.1$, there is subtle difference between the reconstruction performance of the full random matrix and the $[I\, R]$ matrix. When the number of measurements is sufficient ($M/K > 4.2$), using $[I\, R]$ measurement matrix even has a slightly better performance. When the signal is very sparse ($K/N = 0.05$), using $[I\, R]$ as the measurement matrix demands a slight larger number of measurements to achieve perfect reconstruction than using a full random matrix. Besides, it is clear from the figure that using $[T\, R]$ measurement matrix results in much larger reconstruction error.

B. *Representation Basis*

We use two sets of real sensor data to show how CDG framework exploits pervasive sparsity.

1) CTD data: The set of CTD (Conductivity, Temperature, and Depth) data come from National Oceanic and Atmospheric Administration's (NOAA) National Data Buoy Center (NDBC). Figure 5 shows the temperature data collected in the Pacific Sea at (7.0N, 180W) on March 29, 2008 [1]. The data set contains 1000 readings obtained at different depth of sea, ranging from $4.579°\text{C}$ to $27.875°\text{C}$.

It is clear that the readings are piece-wise smooth, and should be sparse in wavelet domain. However, using different layers of wavelet transform matrices in CS recovery will result in dramatically different performance. Figure 6 shows the recovery performance when different Ψ matrices are used. The performance metric is the peak signal-to-noise ratio (PSNR), which is defined as

$$PSNR = 10 \cdot \log_{10} \left(\frac{(DynamicRange)^2}{MSE} \right) \tag{30}$$

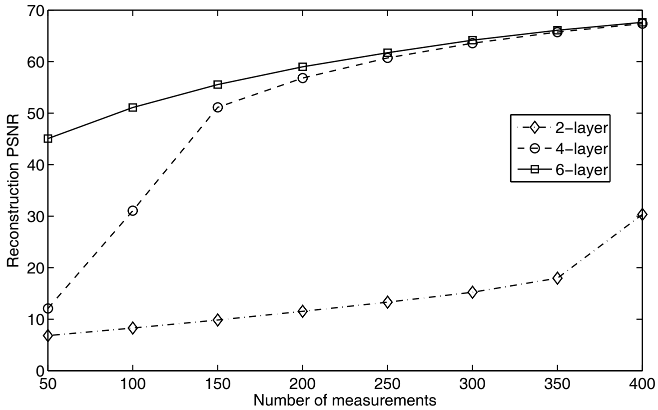

Fig. 6. Data recovery with different representation basis.

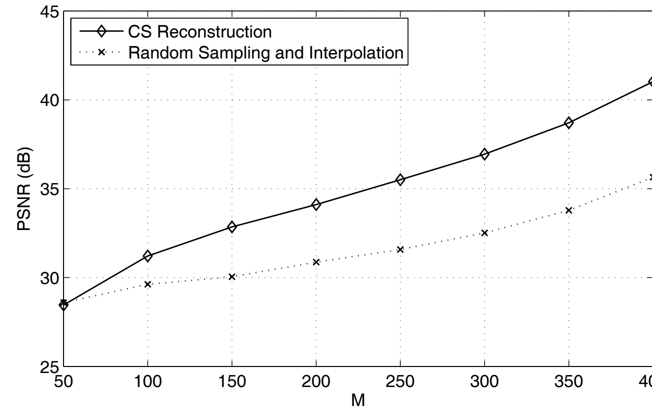

Fig. 7. Comparing reconstruction PSNR of CDG and the anchor scheme at $t_0 + 30$.

where MSE is the mean squared error. When 2-level wavelet transform matrix is used, we cannot get satisfactory results even when we have 400 ($M/N = 0.4$) measurements. In contrast, when 6-level wavelet transform matrix is used, the reconstruction precision is pretty high when there are only 50 ($M/N = 0.05$) measurements, and the precision gradually increases when more measurements are available. The red dotted line in Fig. 5 plots the reconstruction error at $M = 100$. When $M = 100$, CDG reduces the global communication cost by five times and reduces the bottleneck cost by ten times.

The reason why using 2-layer and 4-layer wavelet representation basis cannot achieve satisfactory recovery performance becomes apparent if we check the data sparsity after different levels of wavelet transform. If we consider a non-zero coefficient as one whose absolute value is larger than 0.1, then the temperature data are 263-sparse after 2-layer wavelet transform, 90-sparse after 4-layer transform, and 51-sparse after 6-layer wavelet transform. We would like to point out that, the proposed CDG can utilize the data sparsity after multi-level de-correlation without any additional communication or control cost. The only thing that needs to take care is to use a proper Ψ matrix during CS decoding at the sink. In contrast, conventional in-network compression techniques, such as the one proposed by Ciancio et al. [16], incur additional data exchange costs for every additional level of wavelet de-correlation. Due to this overhead, Ciancio et al. [16] only perform one to two level wavelet de-correlation for sensor readings, although multi-level de-correlation would have reduced more redundancy and produced fewer coefficients.

2) Temperature in Data Center: A contemporary practical application of WSNs is to monitor server temperatures in data centers. The sensor data used in this research are collected from a fraction of a data center, where three sensors are placed at the top, middle, and bottom of each computer rack. There are 498 sensors in total, and the temperature are measured every 30 seconds. We analyze these data offline to see how much traffic would be reduced if CDG was used. For simplicity, we assume that all 498 sensors form one subtree to the sink, and each node only communicates with adjacent nodes.

An important observation on this set of data is that sensor readings exhibit little spatial correlations. Although racks are physically close to each other, temperature readings are dominated by server loads instead of ambient temperature. If we stack up the sensor readings in spatial order, the resulting vector is not sparse in any intuitively known domain. In this case, conventional in-network compression schemes will fail completely. However, CDG can effectively reduce the traffic at bottleneck nodes (by a ratio of two to five) through the flexibility that the framework offers. We observe that temperature readings do not change abruptly within short time intervals. Such temporal correlations can be converted into frequency sparsity at reshuffled ordering. In particular, at time instance t_0, the sink obtains exact values of all d_i's through either naive multi-hop forwarding or acquiring $M = N$ random measurements. Then, d_i's can be sorted in ascending order to form signal $\mathbf{d}$. This reordered signal is piece-wise smooth. Because of the temporal correlation, sensor readings collected at $t_0 + \Delta t$ can also be regarded as piece-wise smooth when organized in the same order. To cope with the situation that temporal correlation becomes weak when the time interval increases, we can refresh the ordering of d_i periodically (e.g. every one or two hours).

The proposed CDG allows the sink to reconstruct such reordered sparse signals from $M (M < N)$ random measurements without changing the data gathering process. In the following experiments, we compare the reconstruction performance of CDG with an anchor scheme based on random sampling and interpolation. A round of data gathering at time $t_0 + 30$, i.e., 30 minutes after the reordering moment, is considered. Figure 7 compares the reconstruction PSNR (peak signal-to-noise ratio) of the two methods at different M. To avoid fluctuation, the MSE (mean squared error) of each setting is averaged over 100 test runs. Results show that CDG significantly outperforms the anchor scheme in most cases, and the performance gain is around 4dB when $M = 250$, which corresponds to a compression ratio of two.

Figure 8 shows the original sensor readings and the reconstructed readings by the two methods when $M = 250$. The sequences shown are representative ones whose MSE is the closest to the average MSE of each method. For easy observation of the differences, the reconstructed readings by CDG and the anchor scheme are shifted by 10 and 20 degrees centigrade respectively. It can be seen that the original sensor

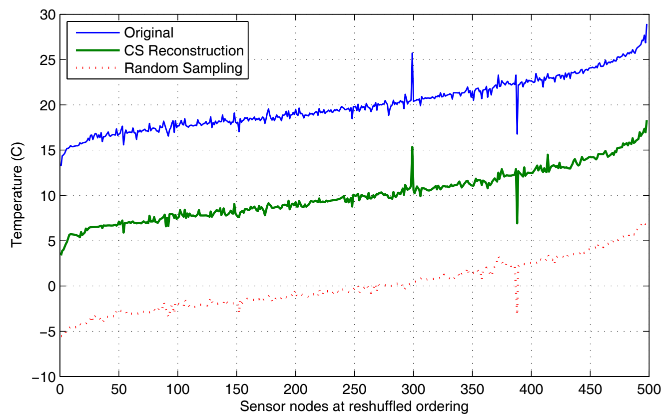

Fig. 8. Comparing reconstructed readings by CDG and the anchor scheme when $M \approx 0.5N$.

readings are generally in ascending order. However, there are two conspicuous spikes in the curve indicating abrupt temperature changes. CS based data reconstruction is capable of capturing all the spikes in all the test runs while random sampling method only opportunistically captures them. It can be concluded that while both methods have similar performance when the target signal is smooth, CDG has significant advantage over random sampling when the sensor data contain large fluctuations or abnormal readings.

VII. Conclusion and Future Work

We have successfully applied CS theory to wireless sensor networks to address the large-scale data gathering problem. When CS theory is used to solve communication problems, we need to consider the design of Φ and Ψ matrices, from both CS and communication perspectives. In this paper, we have discovered that measurement matrices of form $[I\ R]$ not only preserve the restricted isometry property but incur the minimum communication cost in multi-hop networks, making it a favorable choice for data gathering applications. We have also demonstrated through three typical cases how CDG exploits cross-domain sparsity of sensor readings. By choosing the proper Ψ matrix, the data sink is able to recover sensor readings with the minimum number of measurements which are collected by a unified data gathering process. Simulation results based on real sensor data have demonstrated the efficiency of CDG in exploiting the pervasive sparsity.

Although the measurement generation based on $[I\ R]$ matrix extends the application of CDG to smaller scale WSNs, we would like to mention that CDG is more suitable for large scale sensor networks. CDG is also more effective for networks with stable routing structure. The extension of CDG to more challenging networking scenarios will be our future work.

Appendix A
Proof of Theorem 5

Proof: Theorem 5 can be proved similarly as Theorem 1 in Laska et al. [27]. Let Φ be an $M \times N$ matrix with form $[I\ R]$, where elements in R are drawn according to $\mathcal{N}(0, 1/M)$. Let f be an N-dimensional signal. Write f into $f = [f_1^T\ f_2^T]^T$, where f_1 has length M and f_2 has length $N-M$. Then we have:

$$\|\Phi f\|_{l_2}^2 = \|[I\ R]f\|_{l_2}^2 = \|f_1\|_{l_2}^2 + \|Rf_2\|_{l_2}^2 + 2\langle f_1, Rf_2\rangle \quad (31)$$

According to previous results on the RIP of a random matrix [27], the second term on the right-hand side of (31) is bounded by:

$$(1-\delta)\|f_2\|_{l_2}^2 \le \|Rf_2\|_{l_2}^2 \le (1+\delta)\|f_2\|_{l_2}^2 \quad (32)$$

with probability exceeding $1 - 2e^{-M\delta^2/8}$. The third term on the right-hand side of (31) can be written as:

$$2\langle f_1, Rf_2\rangle = 2f_1^T Rf_2 \quad (33)$$

Since the elements in R are drawn according to $\mathcal{N}(0, 1/M)$, it is not hard to deduce that $2f_1^T Rf_2 \sim \mathcal{N}(0, 4\|f_1\|_{l_2}^2\|f_2\|_{l_2}^2/M)$. According to the property of Gaussian variable, the absolute value of this term is bounded by:

$$|2f_1^T Rf_2| \le \delta\|f_1\|_{l_2}\|f_2\|_{l_2} \le \delta\|f\|_{l_2}^2 \quad (34)$$

with probability exceeding $1 - e^{-M\delta^2/8}$. Combining (32) and (34) into (31), we have the following bound

$$(1-2\delta)\|f\|_{l_2}^2 \le \|\Phi f\|_{l_2}^2 \le (1+2\delta)\|f\|_{l_2}^2 \quad (35)$$

with probability exceeding $1 - 3e^{-M\delta^2/8}$ ■

References

[1] NBDC CTD data, http://tao.noaa.gov/refreshed/ctd_delivery.php.

[2] J. Aćimović, B. Beferull-Lozano, and R. Cristescu, "Adaptive distributed algorithms for power-efficient data gathering in sensor networks," in *Proc. Intl. Conf. on Wireless Networks, Comm. and Mobile Computing*, pp. 946–951, June 2005.

[3] W. Bajwa, J. Haupt, A. Sayeed, and R. Nowak, "Compressive wireless sensing," in *Proc. IPSN*, pp. 134–142, 2006.

[4] R. Baraniuk, "Compressive sensing," *IEEE Signal Process. Mag.*, vol. 24, no. 4, pp. 118–121, July 2007.

[5] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin, "A simple proof of the restricted isometry property for random matrices," vol. 28, no. 3, pp. 253–263, Dec. 2008.

[6] D. Baron, M. F. Duarte, S. Sarvotham, M. B. Wakin, and R. G. Baraniuk, "An information-theoretic approach to distributed compressed sensing," in *43rd Allerton Conference on Communication, Control, and Computing*, Sep. 2005.

[7] E. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.

[8] E. Candès, J. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Comm. Pure Appl. Math.*, vol. 59, no. 8, pp. 1207–1223, Aug. 2006.

[9] E. Candès and T. Tao, "Decoding by linear programming," *IEEE Trans. Inf. Theory*, vol. 51, no. 12, pp. 4203–4215, Dec. 2005.

[10] E. J. Candès, "The restricted isometry property and its implications for compressed sensing," *Compte Rendus de l'Academie des Sciences, Paris*, vol. 1, no. 346, pp. 589–592, 2008.

[11] E. J. Candès and T. Tao, "Near optimal signal recovery from random projections: universal encoding strategies?" *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 5406–5425, Feb. 2006.

[12] E. J. Candès and M. B. Wakin, "An introduction to compressive sampling," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 21–30, Mar. 2008.

[13] C. W. Chen and Y. Wang, "Chain-type wireless sensor network for monitoring long range infrastructures: architecture and protocols," *International J. Distrib. Sensor Networks*, vol. 4, no. 4, Oct. 2008.

[14] J. Chou, D. Petrovic, and K. Ramchandran, "A distributed and adaptive signal processing approach to reducing energy consumption in sensor networks," in *Proc. IEEE Infocom*, pp. 1054–1062, Mar. 2003.

[15] D. Chu, A. Deshpande, J. M. Hellerstein, and W. Hong, "Approximate data collection in sensor networks using probabilistic models," in *Proc. ACM ICDE*, pp. 48–59, 2006.

[16] A. Ciancio, S. Pattem, A. Ortega, and B. Krishnamachari, "Energy-efficient data representation and routing for wireless sensor networks based on a distributed wavelet compression algorithm," in *Proc. IPSN*, pp. 309–316, 2006.
[17] R. Cristescu, B. Beferull-Lozano, and M. Vetterli, "On network correlated data gathering," in *Proc. IEEE Infocom*, vol. 4, pp. 2571–2582, Mar. 2004.
[18] R. Cristescu, B. Beferull-Lozano, M. Vetterli, and R. Wattenhofer, "Network correlated data gathering with explicit communication: Np-completeness and algorithms," *IEEE/ACM Trans. Networking*, vol. 14, no. 1, pp. 41–54, Feb. 2006.
[19] T. Dang, N. Bulusu, and W.-C. Feng, "Rida: a robust information-driven data compression architecture for irregular wireless sensor networks," *Wireless Sensor Networks*, vol. 4373, pp. 133–149, Apr. 2007.
[20] D. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.
[21] D. Donoho, M. Elad, and V. Temlyakov, "Stable recovery of sparse overcomplete representations in the presence of noise," *IEEE Trans. Inf. Theory*, vol. 52, no. 1, pp. 6–18, Jan. 2006.
[22] D. L. Donoho, "Neighborly polytopes and sparse solution of underdetermined linear equations," *Preprint*, 2005.
[23] S. Goel and T. Imielinski, "Prediction-based monitoring in sensor networks: taking lessons from MPEG," in *ACM SIGCOMM CCR*, vol. 31, pp. 82–98, Oct. 2001.
[24] H. Gupta, V. Navda, S. Das, and V. Chowdhary, "Efficient gathering of correlated data in sensor network," *ACM TOSN*, vol. 4, no. 1, Jan. 2008.
[25] J. Haupt, W. U. Bajwa, M. Rabbat, and R. Nowak, "Compressed sensing for networked data," *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 92–101, Mar. 2008.
[26] G. Hua and C. W. Chen, "Correlated data gathering in wireless sensor networks based on distributed source coding," *Intl. J. Sensor Networks*, vol. 4, no. 1/2, pp. 13–22, 2008.
[27] J. Laska, M. Davenport, and R. Baraniuk, "Exact signal recovery from sparsely corrupted measurements through the pursuit of justice," in *Proc. Forty-Third Asilomar Conference on Signals, Systems and Computers*, pp. 1556–1560, Nov. 2009.
[28] S. Lee, S. Pattem, M. Sathiamoorthy, B. Krishnamachari, and A. Ortega, Compressed sensing and routing in multi-hop networks. In *USC CENG Technical Report*, 2009.
[29] S. Lee, S. Pattem, M. Sathiamoorthy, B. Krishnamachari, and A. Ortega, "Spatially-localized compressed sensing and routing in multi-hop sensor networks," in *Proc. Third International Conference on Geosensor Networks*, vol. 5669, pp. 11–20, July 2009.
[30] C. Luo, F. Wu, J. Sun, and C. W. Chen, "Compressive data gathering for large-scale wireless sensor networks," in *Proc. ACM Mobicom'09*, pp. 145–156, Sep. 2009.
[31] G. Quer, R. Masiero, D. Munaretto, M. Rossi, J. Widmer, and M. Zorzi, "On the interplay between routing and signal representation for compressive sensing in wireless sensor networks," in *ITA*, Feb. 2009.
[32] M. Rabbat, J. Haupt, A. Singh, and R. Nowak, "Decentralized compression and predistribution via random gossiping," in *IPSN*, pp. 51–59, Apr. 2006.
[33] A. Silberstein, R. Braynard, and J. Yang, "Constraint chaining: on energy-efficient continuous monitoring in sensor networks," in *Proc. ACM SIGMOD ICMD*, pp. 157–168, 2006.
[34] D. Slepian and J. K. Wolf, "Noiseless encoding of correlated information sources," *IEEE Trans. Inf. Theory*, vol. 19, pp. 471–480, July 1973.
[35] S. J. Szarek, "Condition numbers of random matrices," vol. 7, no. 2, pp. 131–149, June 1991.
[36] R. Szewczyk, A. Mainwaring, J. Polastre, J. Anderson, and D. Culler, "An analysis of a large scale habitat monitoring application," in *Proc. ACM SenSys*, pp. 214–226, Nov. 2004.
[37] D. Viswanath and L. N. Trefethen, "Condition numbers of random triangular matrices," in *SIAM J. Matrix Anal. Appl.*, vol. 19, pp. 564–581.
[38] L. Wang and A. Deshpande, "Predictive modeling-based data collection in wireless sensor networks," *Wireless Sensor Networks*, vol. 4913, pp. 34–51, Jan. 2008.
[39] W. Wang, M. Garofalakis, and K. Ramchandran, "Distributed sparse random projections for refinable approximation," in *Proc. IPSN*, pp. 331–339, Apr. 2007.
[40] S. Yoon and C. Shahabi, "The clustered aggregation (CAG) technique leveraging spatial and temporal correlations in wireless sensor networks," *ACM Trans. Sensor Networks*, vol. 3, no. 1, Mar. 2007.
[41] K. Yuen, B. Liang, and B. Li, "A distributed framework for correlated data gathering in sensor networks," *IEEE Trans. Veh. Technol.*, vol. 57, no. 1, pp. 578–593, Jan. 2008.

Chong Luo (M'05) received her B.S. degree from Fudan University, Shanghai, China in 2000, M.S. degree from National University of Singapore in 2002. She is currently pursuing the Ph.D. degree in the Department of Electronic Engineering, Shanghai Jiao Tong University, Shanghai, China. She has been with Microsoft Research Asia since 2003, where she is a Researcher in the Internet Media group. Her research interests include multimedia communications, wireless sensor networks and media cloud computing.

Feng Wu (M'99-SM'06) received his B.S. degree in Electrical Engineering from XIDIAN University in 1992. He received his M.S. and Ph.D. degrees in Computer Science from Harbin Institute of Technology in 1996 and 1999, respectively. Wu joined in Microsoft Research Asia in 1999. He has been a researcher with Microsoft Research Asia since 2001 and is now a Senior Researcher/Research Manager. His research interests include image and video representation, media compression and communication, as well as, computer vision and graphics. He has authored or co-authored over 200 papers published in International journals and conferences. As a co-author, he got the best paper award in IEEE T-CSVT 2009, PCM 2008 and SPIE VCIP 2007.

Jun Sun received his B.S. and M.S. degrees from the University of Electronic Science and Technology of China and Shanghai University, China, in 1989 and 1992 respectively, and Ph.D. degree from Shanghai Jiao Tong University, Shanghai, China, in 1995. As a professor, he is now with the Institute of Image Communication and Information Processing of Shanghai Jiao Tong University. His general interests include DTV coding and video communication.

Chang Wen Chen (S'86-M'90-SM'97-F'04) received his BS from University of Science and Technology of China in 1983, MSEE from University of Southern California in 1986, and Ph.D. from University of Illinois at Urbana-Champaign in 1992. He was elected an IEEE Fellow for his contributions in digital image and video processing, analysis, and communications and an SPIE Fellow for his contributions in electronic imaging and visual communications.