Applications of Compressed Sensing in Communications Networks

Hong Huang\textsuperscript{1}, Satyajayant Misra\textsuperscript{2}, Wei Tang\textsuperscript{1}, Hajar Barani\textsuperscript{1}, and Hussein Al-Azzawi\textsuperscript{1}

\textsuperscript{1}Klipsch School of Electrical and Computer Engineering, New Mexico State University, NM, USA
\textsuperscript{2}Department of Computer Science, New Mexico State University, NM, USA

Abstract—This paper presents a tutorial for CS applications in communications networks. The Shannon’s sampling theorem states that to recover a signal, the sampling rate must be as least the Nyquist rate. Compressed sensing (CS) is based on the surprising fact that to recover a signal that is sparse in certain representations, one can sample at the rate far below the Nyquist rate. Since its inception in 2006, CS attracted much interest in the research community and found wide-ranging applications from astronomy, biology, communications, image and video processing, medicine, to radar. CS also found successful applications in communications networks. CS was applied in the detection and estimation of wireless signals, source coding, multi-access channels, data collection in sensor networks, and network monitoring, etc. In many cases, CS was shown to bring performance gains on the order of 10X. We believe this is just the beginning of CS applications in communications networks, and the future will see even more fruitful applications of CS in our field.

Index Terms—Wireless network, performance analysis and modeling.

I. INTRODUCTION

PROCESSING data is a big part of modern life. Interesting data typically is sparse in certain representations. An example is an image, which is sparse in, say, the wavelet representation. The conventional way to handle such signal is to acquire all the data first and then compress it, as is done in image processing. The problem with this kind of processing is, as Donoho puts it in his seminal paper \cite{21}: "Why go to so much effort to acquire all the data when most of what we get will be thrown away? Can we not just directly measure the part that will not end up being thrown away?"

Indeed, compressed sensing (CS) does just that—measuring only the part of data that is not thrown away. The well-known Shannon’s sampling theorem states that to recover a signal exactly, the sampling rate must be as least the Nyquist rate, which is twice the maximum frequency of the signal. In contrast, using CS, far fewer samples or measurements at far below the Nyquist rate are required to recover the signal as long as the signal is sparse and the measurement is incoherent, the exact meanings of which will be revealed later. In addiction, CS have other attractive attributes, such as universality, fault-tolerance, robustness to noise, graceful degradation, etc.

Since its introduction in \cite{7} and \cite{21} in 2006, CS has received much attention in the research community. Thousands of papers have been published on topics related to CS, and hundreds of conferences, workshops and special sessions have been devoted to CS \cite{24}. CS has been showed to bring significant performance gains in wide-ranging applications from astronomy, biology, communications, image and video processing, medicine, to radar. Although there are good survey and tutorial papers on CS itself \cite{3}, \cite{8}, \cite{9}, \cite{17}, there is no good tutorial paper on the application of CS in communications networks to the best of authors’ knowledge. We hope this paper will close the gap.

CS can be applied in all layers of communications networks. At the physical layer, CS can be used in detecting and estimating sparse physical signals such as ultra-wide-band (UWB) signals, wide-band cognitive radio signals, and MIMO signals. Also, CS can be used as erasure code. At the MAC layer, CS can be used to implement multi-access channels. At the network and transport layer, CS can be used for data collection in wireless sensor networks, where the sensory signals are usually sparse in certain representations. At the application level, CS can be used to monitor the network itself, where network performance metrics are sparse in some transform domains. In many cases, CS was shown to bring performance gains on the order of 10X. We believe this is just the beginning of CS applications in communications networks, and the future will see even more fruitful applications of CS in our field.

The rest of the paper is organized as follows. In Section II, we provide an overview of the CS theory. In Section III, we describe the algorithms that implement CS. In Section IV, we describe some extensions of the CS theory. In section V, we provide a tutorial on CS applications in the physical layer. In Section VI, we describe how CS is used in the network and transport layer. In Section VII, we describe CS applications in the MAC and application layers. We conclude in Section VIII. The particular order of Sections V, VI, and VII reflects the level of research activity in each area, with that in the physical layer being the highest.

II. THE THEORY OF COMPRESSED SENSING

A. Sparse Signal

For the vector $x = (x_i), i = 1, 2, n$ we define its $l_p$ norm as the following

$$
\|x\|_p = (\sum_{i=1}^{n} |x|^p)^{1/p}.
$$

(1)
Note that for \( p = 0, \|x\|_0 \) is the number of nonzero elements in \( x \); for \( p = 1, \|x\|_1 \) is the summation of the absolute values of elements in \( x \); for \( p = 2, \|x\|_2 \) is usual Euclidean norm; and for \( p = \infty, \|x\|_\infty \) is the maximum of the absolute values in \( x \). We call a signal \( k \)-sparse if \( \|x\|_0 = k \), namely \( x \) has only \( k \) nonzero elements. We also say the sparsity of the signal is \( k \).

In fact, few real-world signals are exactly \( k \)-sparse. Rather, they are compressible in the sense that they can be well-approximated by a \( k \)-sparse signal. Denote \( \Sigma_k \) is the set of all \( k \)-sparse signals. The error incurred by approximating a compressible signal \( x \) by a \( k \)-sparse signal is given by

\[
\sigma_k(x)_p = \min_{\hat{x} \in \Sigma_k} \|x - \hat{x}\|_p.
\]

(2)

If \( x \) is \( k \)-sparse, \( \sigma_k(x)_p = 0 \). Otherwise, the optimal \( k \)-sparse approximation of \( x \) is the vector that keeps the \( k \) largest elements of \( x \) with the rest elements setting to zero.

Consider a compressible signal \( x \) and order its elements in descending order so that \( |x_1| \geq |x_2| \geq \ldots \geq |x_n| \). We say the elements follow a power-law decay if

\[
|x_i| \leq c i^{-\alpha}
\]

(3)

where \( c \) is a constant and \( \alpha \) is the decay exponent. If \( x \) follows a power-law decay, there exist some constants \( c, r > 0 \), such that

\[
\sigma_k(x)_2 \leq ck^{-\alpha-1/2}.
\]

(4)

In other words, the error also follows a power-law decay.

### B. The Basic Framework of CS

Although CS applies to both finite and infinite dimensional signals, we focus on finite dimensional signals here for the ease of exposition. Consider a \( n \)-dimensional signal \( f \), which has a representation in some orthonormal basis \( \Psi = \{\psi_1, \psi_2, \ldots, \psi_n\} \) where

\[
f = \sum_{i=1}^{n} x_i \psi_i = \Psi x.
\]

(5)

In the above, \( x_i \) and \( \psi_i \) are the \( i \)th coefficient and \( i \)th basis, respectively. The CS theory says that if \( f \) is sparse in the basis \( \Psi \), which needs not be known \textit{a priori}, then, under certain conditions, taking \( m \) nonadaptive measurements of \( f \), where \( m \ll n \), suffices to recover the signal exactly. Each measurement \( y_j \) is a projection of the original signal, i.e., the \( m \)-dimensional measurement can be represented by

\[
y = \Phi x
\]

(6)

where \( y \) is the measurement vector and \( \Phi \) is a \( m \times n \) sensing matrix. Since \( m \ll n \), to recover \( x \) from \( y \) is an ill-posed inverse problem. In other words, there might be multiple \( x \)'s that satisfy \( y \). However, we can take advantage of the fact that \( x \) is sparse in a certain representation \( \Psi \), and formulate the following optimization problem:

\[
(P_0) \min_x \|\Psi x\|_0 \quad \text{subject to} \quad y = \Phi x.
\]

(7)

It is known that solving \( P_0 \) is NP-hard. Fortunately, it was shown that one can replace the \( l_0 \) norm by \( l_1 \) norm, and formulate the following optimization problem instead

\[
(P_1) \min_x \|\Psi x\|_1 \quad \text{subject to} \quad y = \Phi x.
\]

(8)

It can be shown if the signal is sufficiently sparse, the solutions to \( P_0 \) and \( P_1 \) are the same. \( P_1 \) is a convex linear-programming problem with efficient solution techniques.

In the rest of the section, we will describe what kind of sensing matrix will enable signal recover, and how many measurements are need. The most well-known answers to the above question are expressed in terms of spark, mutual coherence, and restricted isometry property.

### C. Spark

The spark of a matrix \( \Phi \) is the smallest number of columns of \( \Phi \) that are linearly dependent [20]. To see why spark is relevant, consider the \( k \)-sparse solution to the linear equation \( y = \Phi x \). The solution is not unique if there are two different \( k \)-sparse vectors \( x \) and \( x' \) such that \( y = \Phi x = \Phi x' \), which implies \( \Phi(x - x') = 0 \). Equivalently, \( \langle x - x', \Phi \rangle = 0 \) is in the null space of \( \Phi \). Since the sparsity of \( \langle x - x', \Phi \rangle \) is at most \( 2k \), so to ensure the uniqueness of the solution, the null space of \( \Phi \) cannot have vectors whose sparsity is less or equal to \( 2k \). In other words, to guarantee the uniqueness of the \( k \)-sparse solution, the smallest number of linear dependent columns of \( \Phi \) must be no less than \( 2k \), or

\[
\text{spark}(\Phi) \geq 2k.
\]

(9)

For the \( m \times n \) matrix \( \Phi \), \( \text{spark}(\Phi) \in [2, m+1] \). So, to make the solution unique, the number of measurements taken must satisfy

\[
m \geq 2k.
\]

(10)

It turns out that equation (9) is also the sufficient condition to guarantee the error bound given by

\[
\|\hat{x} - x\|_2 \leq c \sigma_k(x)_1 / \sqrt{k}
\]

(11)

where \( x \) and \( \hat{x} \) are the original and the recovered signals, \( c \) is a constant, and \( \sigma_k(x)_1 \) is defined in (4). So, if \( x \) is exactly \( k \)-sparse, then \( \sigma_k(x)_1 = 0 \), and the signal recovery is exact.

### D. Mutual Coherence Property (MIP)

According to CS theory [19], to enable signal recovery, the sensing matrix \( \Phi \) must be incoherent with the sparse representation \( \Psi \). For the ease of exposition, we assume both matrices are orthonormal, which is not required by CS theory. Formally, the incoherence between two matrices are defined by

\[
\mu(\Phi, \Psi) = \sqrt{n} \max_{1 \leq k, j \leq n} |\langle \phi_k, \psi_j \rangle|
\]

(12)

where \( \langle \phi_k, \psi_j \rangle \) is the inner product of \( k \)th column in \( \Phi \) and \( j \)th column in \( \Psi \). In other words, the coherence measures the maximum correlation between the columns of \( \Phi \) and \( \Psi \), and it has a range of \([1, \sqrt{n}]\). The coherence is large when the two matrices are closely correlated, and is small otherwise.
With probability $1 - \delta$, a $k$-sparse signal can be recovered exactly if $m$ (the number of measurements) satisfy the following condition for some positive constant $c$

$$m \geq c\mu_k^2(\Phi, \Psi)k \log(n/\delta).$$  

(13)

Thus when $\delta$ approaches 1, the probability of recovery approaches 1. Note that the number of measurements required increases linearly with the sparsity of the signal and quadratically with the coherence of the matrices.

E. Restricted Isometry Property (RIP)

In CS, an important concept is the so-called restricted isometry property [6, 56]. We define the isometry constant $\delta_k$ of a matrix $\Phi$ as the smallest number such that the following holds for all $k$-sparse vectors $x$

$$1 - \delta_k \|x\|_2^2 \leq \|\Phi x\|_2^2 \leq (1 + \delta_k) \|x\|_2^2$$

(14)

Informally, we say a matrix $\Phi$ holds RIP of order $k$ if $\delta_k$ is not too close to one. Three remarks follow:

• A transformation of $k$-sparse signal by a matrix holding RIP approximately preserves the signal’s $l_2$ norm. The degree of preservation is indicated by $\delta_k$, with $\delta_k = 0$ being exactly preserving and $\delta_k = 1$ being non-preserving. Because RIP preserves $l_2$ norm, for a matrix $\Phi$ holding RIP, its null space does not contain $k$-sparse vector. As mentioned earlier, this is important because for any solution $x$ of (8), $x + x_0$ is also a solution, where $x_0$ is a vector in the null space of $\Phi$, making the solution not unique. RIP precludes such situation from arising.

• RIP also implies that all subsets of $k$ columns of $\Phi$ are nearly orthogonal.

• A further implication of RIP is that if $\delta_{2k}$ is sufficiently small, then the measurement matrix $\Phi$ preserves the distance between any pair of $k$-sparse signals, i.e.,

$$(1 - \delta_{2k}) \|x_1 - x_2\|_2^2 \leq \|\Phi(x_1 - x_2)\|_2^2 \leq (1 + \delta_{2k}) \|x_1 - x_2\|_2^2$$

When the measurement involves noise, we can revise $P_1$ by relaxing the measurement constraint as below

$$P_2 \quad \min_x \|\Psi x\|_1 \quad \text{subject to } \|y - \Phi x\|_2 \leq \epsilon$$

(15)

where $\epsilon$ bounds the amount of noise. Using RIP, we can bound the error of signal recovery when noise is present. In fact, if $\delta_{2k} < \sqrt{2} - 1$, then for some constants $c_1$ and $c_2$ the solution $x^*$ to $P_1$ satisfies

$$\|x - x^*\|_2 \leq c_1 \frac{\|x - x_k\|_1}{\sqrt{k}} + c_2 \epsilon$$

(16)

where $x_k$ is the vector $x$ with all but the largest $k$ elements set to zero. Two remarks follow:

• Consider the case where there is no noise. In such case, if $x$ is $k$-sparse, i.e., $x = x_k$, then the recovery is exact. On the other hand, if $x$ is not $k$-sparse, then the recovery is as good as if we know beforehand the largest $k$ elements and use them as an approximation.

• The contribution of the noise is a linear term. In other words, CS handles noise gracefully.

F. Relationships among Spark, MIP, and RIP

Spark and MIP has the following relationship

$$\text{spark}(\Phi) \geq 1 + 1/\mu(\Phi).$$

(17)

RIP is strictly stronger than the spark condition. Specifically, if a matrix $\Phi$ that satisfies the RIP of order $2k$, then $\text{spark}(\Phi) \geq 2k$.

RIP and MIP is related as follows. If $\Phi$ has unit-norm columns and coherence $\mu(\Phi)$, then $\Phi$ satisfies the RIP of order $k$ for all $k < 1/\mu(\Phi)$.

For general matrix, it is hard to verify whether the matrix satisfies the spark and RIP conditions. Verifying MIP is much easier, since it involves calculating the inner products between columns of two matrices and taking the maximum of products, referring to [12].

G. Measurement Matrices

Random matrices are commonly used for measurement matrices, though non-random matrices can also be used as long as they satisfy spark, MIP, or RIP requirements. By definition, random matrices are those that have elements following independent identical distributions. Examples of random matrices include those, whose elements follow Bernoulli distribution ($\text{Prob}(\phi_{i,j} = \pm 1/\sqrt{m}) = 1/2$) or Gaussian distribution with zero mean and variance of $1/m$.

With high probability, random matrices are incoherent with any fixed basis $\Psi$. Further, most random matrices obey the RIP and can serve as the measurement matrix $\Phi_{m \times n}$ as long as the following condition holds for some constant $c$ that varies with the particular matrix used

$$m \geq c k \log(n/k).$$

(18)

In fact, using random matrices is near-optimal in the sense that it is impossible to recover signal using substantially fewer measurements than the left-hand side of (18).

III. THE ALGORITHMS FOR IMPLEMENTING CS

There three main types of algorithms for implementing CS: convex optimization algorithms, combinatorial algorithms, and greedy algorithms. Convex optimization algorithms require fewer measurements but are more computationally complex than those of combinatorial algorithms. Those two types of algorithms represent two extremes in the spectra of the number of measurements and the computational complexity. Greedy algorithms provide a good compromise between the two extremes. Table I summarizes the computational complexity of various algorithms for implementing CS.

A. Convex Optimization Algorithms

With convex optimization algorithms, we use an unconstrained version of $P_2$ given by the following

$$P_4 \quad \min_x \frac{1}{2} \|y - \Phi x\|_2^2 + \lambda \|\Psi x\|_1$$

(19)

where $\lambda$ can be selected based on the how much weight we want put on the fidelity to measurement and sparsity of the
signal. Several ways to select \( \lambda \) were discussed in \cite{25, 28}. Efficient algorithms such as interior-point methods \cite{7} and projected gradient methods \cite{26}, and basis pursuit \cite{23} can be used to solve the convex optimization problem.

### B. Greedy Algorithms

Greedy algorithms iteratively approximate the original signal and its support (the index set of nonzero elements). Representative examples include iterative thresholding \cite{15}, Matching Pursuit (MP) \cite{43}, and its improved version Orthogonal Matching Pursuit (OMP) \cite{46}, which is listed below as a representative algorithm.

**Algorithm 1 OMP Algorithm**

Input: \( m \times n \) measurement matrix \( \Phi = (\phi_i) \), where \( \phi_i \) is the \( i \)th column of \( \Phi \), \( n \)-dimensional initial signal \( x \), and error threshold \( \epsilon \).

1. Set \( j = 0 \)
2. Set the initial solution \( x_0 = 0 \)
3. Set the initial residual \( r_0 = y - \Phi x_0 = y \)
4. Set the initial support \( S_0 = \emptyset \)
5. **repeat**
6. \( j = j + 1 \)
7. Select index \( i \) so that max\( _i \| \phi_i^T r_j \|_2 \), update \( S_j = S_{j-1} \cup i \)
8. Update \( x_j = \arg\min_x \| \phi x - y \|_2 \), subject to \( \text{supp}(x) = S_j \)
9. Update \( r_j = (1 - P_j) y \), where \( P_j \) denote the projection onto the space spanned by the columns in \( S_j \)
10. **until** \( \| r_j \|_2 \leq \epsilon \)
11. Output: approximate solution \( x_j \)

In the algorithm listed above, we have set \( \Psi \) to identity matrix for the ease of exposition. In the algorithm, we iteratively select the column of \( \Phi \) that is most correlated with current residual \( r_j \) and add it to the current support \( S_j \) in step 7. Then we update the current signal \( x_j \) so that it is conforming to the measurements \( y \) and current support \( S_j \). Finally we update the current residual \( r_j \) so that it contains measurements excluding those included by the current support. Denote \( \Phi_j \) as the submatrix containing only columns in the current support \( S_j \), then \( P_j = \Phi_j (\Phi_j^T \Phi_j)^{-1} \Phi_j^T \). It can be shown that if the the signal \( x \) is sufficiently sparse, i.e., \( \| x \|_0 \leq \frac{1}{2}(1 + \mu(\Phi)^{-1}) \), OMP can exactly recover \( x \) when \( \epsilon \) is set to zero \cite{17}. For noisy measurements, \( \epsilon \) can be set to a nonzero value, for details of which the reader is referred to \cite{46}. Compared with MP, OMP converges faster with no more than \( n \) iterations.

### C. Combinatorial Algorithm

These algorithms are mostly developed by the theoretical computer science community, in many cases predating the compressed sensing literature. There are two major types of algorithms: group testing and data stream sketches. In group testing \cite{30, 39}, there are \( n \) items represented by a \( n \)-dimensional vector \( x \), of which there are \( k \) anomalous items we would like to identify. The values of elements of \( x \) are nonzero if they correspond to anomalous items and zero otherwise. The problem is to design a collection of tests, resulting in the measurement \( y = \Phi x \), where the element of matrix \( \Phi_{i,j} \) indicates \( j \)th item is included in the \( i \)th test. The goal is to recover \( k \)-sparse vector \( x \) using least number of tests, which is essentially a sparse signal recover problem in CS.

An example of date stream sketches \cite{13, 44} is to identify most frequency source or destination IP addresses passing through a network device. Denote \( x_i \) as the number of times IP address \( i \) is encountered. The vector \( x \) is sparse since the network device is likely to see a small portion of IP address space. Directly storing \( x \) is infeasible since the index set \( \{ i \} \) is too large (\( 2^{32} \), since IP address has 32 bits). Instead, we store a sketch defined by \( y = \Phi x \), where \( \Phi \) is a \( m \times n \)-dimensional matrix with \( m \ll n \). Since \( y \) has a linear relationship with \( x \), \( y \) is incrementally updated each time a packet arrives. The goal is to recover \( x \) from the sketch \( y \), which again is a sparse signal recover problem in CS.

### IV. Extensions of CS

In this section, we describe two extensions to CS: 1) CS for multiple measurement vector signals and 2) CS for analogy signals.

#### A. Multiple Measurement Vectors (MMV)

In MMV problems, we deal with \( l \) correlated sparse signals, which share the same support (the index set of nonzero coefficients). Instead of recovering each signal \( x_i \) independently, we would like to recover the signals jointly by exploiting their common sparse support. Let \( X = (x_1, x_2, ... x_l) \) denote the \( n \times l \) matrix representing the signals, and \( \Lambda = \text{supp}(X) \) denote the the sparse support. The CS problem for MMV can be formulated as

\[
(P_3) \quad \min_{X} \| X \|_1 \text{ subject to } Y = \Phi X
\]

where \( Y \) is the \( m \times l \) measurement matrix. In the above, we assume \( X \) is already in a sparse representation to simplify the presentation. It is straightforward to extend the results to the cases where \( X \) is not in a sparse representation.
A necessary and sufficient condition to recover signal using CS is given by

\[ \text{spark}(A) > 2|\text{supp}(X)| - \text{rank}(X) + 1. \] (21)

Assume the signals are \( k \) sparse, then \( |\text{supp}(X)| = \text{rank}(X) = k \). For a \( m \times l \) matrix \( A \), \( \text{spark}(A) \leq m + 1 \). So, the necessary and sufficient condition for signal recovery becomes

\[ m > k. \] (22)

Recall that for a single signal, the necessary and sufficient condition for signal recovery is \( m \geq 2k \), referring to \cite{9}. Thus MMV CS reduces the number of measurements by half compared with single measurement vector (SMV) CS.

\[
x(t) \in \{-1, 1\}
\]

\[
\begin{matrix}
\text{Analog Filter} \\
\h(t)
\end{matrix}
\]

\[
y[m]
\]

\[
\begin{matrix}
p_c(t) \in \{1, \}
\end{matrix}
\]

Fig. 1. Pseudo-random demodulation for AIC.

### B. Analog-to-Information Conversion (AIC)

Today, digital signal processing is prevalent. Analog-to-digital converters (ADC) are used to convert analog signals to digital signals. ADC requires sampling at Nyquist rate. The ever-increasing demands from applications, such as ultra-wideband communications and radar signal detection, are pushing the performance of ADCs toward their physical limits. In \cite{6}, \cite{7}, CS was used to implement an analog-to-information converter. AIC was shown to be particularly effective for wideband signals that are sparse in frequency domain. We provide an overview of AIC below.

Assume analog sign \( x(t) \) is sparse in an orthogonal basis \( (\Psi = \{\psi_1(t), \psi_2(t), ..., \psi_n(t)\}) \) and is expressed by

\[
x(t) = \sum_{i=1}^{n} \theta_i \psi_i(t). \] (23)

AIC is composed of three components, a wide-band pseudo-random signal demululator \( p_c(t) \), a filter \( h(t) \) (typically a low-pass filter), and a low-rate analog-to-digital converter (ADC), referring to Figure 1. The input of AIC is the signal \( x(t) \). The output is the low-rate measurement \( y_i \), given by

\[
y_i = \int_{-\infty}^{\infty} x(\tau)p_c(\tau)h(t - \tau) \, d\tau |_{t=m\Delta} \] (24)

where \( \Delta \) is the sampling period. Substituting (23) into (24), we have

\[
y_i = \sum_{j=1}^{n} \theta_j \int_{-\infty}^{\infty} \psi_j(\tau)p_c(\tau)h(i\Delta - \tau) \, d\tau. \] (25)

The equivalent measurement matrix is given by

\[
\phi_{i,j} = \int_{-\infty}^{\infty} \psi_j(\tau)p_c(\tau)h(i\Delta - \tau) \, d\tau. \] (26)

So the compressed sensing with analog signals can be formulated as

\[
(P_3) \min_{\theta} \|\theta\|_1 \text{ subject to } y = \Phi \theta \] (27)

which is exactly the same formulation as \( P_1 \) with \( \Psi \) set to identity matrix and \( x \) vector replaced by the \( \theta \) vector.

In \cite{7}, simulation was performed using a 200 MHz carrier modulated by a 100 MHz signal. AIC was shown to able to successfully recover the signal at a sampling rate of one sixth of the Nyquist rate.

### V. Applications of CS in the Physical Layer

CS can be used in detecting and estimating sparse physical signals such as ultra-wide-band (UWB) signals, wide-band cognitive radio signals, and MIMO signals, etc. In addition, CS can be used as erasure code. The details are provided below.

#### A. Ultra-Wide-Band (UWB) Signals

UWB communications is a promising technology for low-power, high-bandwidth wireless communications. In UWB, an ultra-short pulse, on the order of nanoseconds, is used as the elementary signal to carry information. The advantages of UWB are: 1) The implementation of the transmitter is simple because of the use of base-band signaling; 2) UWB has little impact on other narrow-band signals on the same frequency range, since its power spreads out on the broad frequency range. However, one of the challenges for UWB is that it requires extremely high sampling rate (several GHz) to digitize UWB signals based on the Nyquist rate, leading to high cost in hardware. CS provides an effective solution to this problem by requiring much lower sampling rate. There are a number of recent papers in this area, and we provide two examples below.

1) **UWB Channel Estimation:** In reference \cite{45}, CS was applied to UWB multi-path channel estimation. The received multi-path UWB signal can be expressed as

\[
x(t) = p(t) * h(t) = \sum_{i=1}^{L} \theta_i p(t - \tau_i) \] (28)

where \( p(t) \) is the ultra-short pulse used to carry information, and \( h(t) \) is the impulse response of the UWB channel. Typically a Gaussian pulse or its derivatives can be used as UWB pulse, e.g., \( p(t) = p_n(t) e^{-t^2/2\sigma^2} \), where \( p_n(t) \) is a polynomial of degree \( n \), and \( \sigma \) represents the width of the signal. The impulse response of the channel can be expressed as

\[
h(t) = \sum_{i=1}^{L} \theta_i \delta(t - \tau_i) \] (29)

where \( \delta() \) is the dirac delta function, \( \theta_i \) and \( \tau_i \) are the gain and the delay of the \( i \)th path received signal, and \( L \) is the total number of propagation paths.

Typically, the number of the channel parameters \( (\theta_i \text{ and } \tau_i) \) is on the order of \( 10^2 \). However, most of paths carry negligible energy and can be ignored. In other words, the channel parameters are sparse and CS can be used to estimate them.
Two approaches were proposed in [45] for UWB channel estimation using CS. One is correlator-based and the other is rake-receiver-based. It was shown that CS-based approaches outperform the traditional detector using only 30% of ADC resources.

2) UWB Echo Signal Detection: In reference [52], the authors proposed to use CS to detect UWB radar echo signals. Let \( s(t) \) denote the transmitted signal and \( \tau \) as the Nyquist sampling interval of the echo signal. All time-shifted versions of \( s(t) \) constitute a redundant dictionary as follows

\[
\Psi = \{ \psi_i(t) = s(t - i\tau), i \in [1, 2, ..., n] \}. \tag{30}
\]

The received signal is sparse in such dictionary and can be expressed as

\[
x(t) = \sum_{i=1}^{k} \theta_i \psi_i(t) \tag{31}
\]

where \( k \) is the number of target echoes and represents the sparsity of the signal, and \( \theta_i \) indicates the amplitude of \( i \)th target echo. Equation (31) has the exact the same form as (23) and standard methods of CS can be used to detect the UWB echo signals. It was shown in [52] that the sampling rate can be reduced to only about 10% of the Nyquist rate.

B. Wide-Band Cognitive Radio Signals

Dynamic spectrum access (DSA) is an emerging approach to solve today’s radio spectrum scarcity problem. Key to DSA is the cognitive radio (CR) that can sense the environment and adjust its transmitting behavior accordingly to not cause interference to other primary users of the frequency. Thus, spectrum sensing is a critical function of CR. However, wide-band spectrum sensing faces hard challenges. There are two major approaches to do wide-band spectrum sensing. First, we can use a bank of tunable narrow-band filters to search narrow-bands one by one. The challenge of wide-band spectrum sensing is that a large number of filters need to be used, leading to high hardware cost and complexity. Second, we can use a single RF front-end and use DSP to search the narrow-bands. The challenge in this approach is that very high sampling rate and processing speed are required for wide-band signals. CS can be used to overcome the challenges mentioned above. In this subsection, we introduce three approaches of applying CS to the spectrum sensing problem.

1) Spectrum Sensing: A Digital Approach: In this approach, the signal is first converted to the digital domain, and then CS is performed on the digital signal. Let \( x(t) \) denote the signal sensed by CR, \( B \) the frequency range of the wide-band, \( F \) the set of frequency bands currently used by other users. Typically, \( |F| \ll B \) [49], indicating that \( x(t) \) is sparse in the frequency domain and thus CS is applicable. So, instead of sampling at Nyquist rate \( f_N \), we can sample at a much slower rate roughly around \( |F|f_N/B \). In [54], the CS formulation of spectrum sensing is proposed as

\[
f = \arg \min_{f} \|f\|_1 \text{ subject to } x(t) = Sx(t) = SF^{-1}f \tag{32}
\]

where \( f \) is signal representation in the frequency domain, \( F^{-1} \) is the inverse Fourier transformation, and \( S \) is a reduced-rate sampling matrix operating at a rate close to \( |F|f_N/B \), and \( x(t) \) is the reduced rate measurements. Simulation results show that decent signal recovery can be achieved at 50% Nyquist rate [49].

2) Spectrum Sensing: An Analog Approach: In this approach, CS is directly performed on the analog signal [58], which has the advantage of saving the ADC resources, especially in cases where the sampling rate is high. The implementation is similar to that described in Section 11.1. A parallel bank of filters are to acquire measurements \( y_j \). To reduce the number of filters required, which is equal to the number of measurements \( M \) and can be potentially large, each filter samples time-windowed segments of signal. Let \( N_F \) denote the number of filters required, and \( N_S \) denote the number of segments each filter acquire. As long as \( N_FN_S = M \), the measurement is sufficient. Simulations show that using this approach, for a OFDM-based CR system with 256 sub-carriers where only 10 carriers are simultaneously active, a CS system with 8-10 filters can perform spectrum sensing at 20/256 of the Nyquist rate.

3) Spectrum Sensing: A Cooperative Approach: The performance of CS-based spectrum sensing can be negatively impacted by the channel fading and the noise. To overcome such problem, a cooperative spectrum sensing scheme based on CS was proposed in [55]. In this scheme, the assumptions are that there are \( J \) CRs and \( I \) active primary users. The entire frequency range is divided into \( F \) non-overlapping narrow-bands \( \{B_i\}_{i=1}^{F} \). In the sensing period, all CRs remain silent and cooperatively perform spectrum sensing. The received signal at \( j \)th CR is

\[
x_j(t) = \sum_{i=1}^{I} h_{i,j}(t) * s_i(t) + n_j(t) \tag{33}
\]

where \( h_{i,j} \) is the channel impulse response, \( s_i(t) \) is transmitted signal from primary user \( i \), \( * \) denotes convolution, and \( n_j(t) \) denotes noise at the receiver \( j \). We take discrete Fourier transform on \( x_j(t) \) and obtain

\[
\tilde{x}_j(f) = \sum_{i=1}^{I} \tilde{h}_{i,j}(f) \tilde{s}_i(f) + \tilde{n}_j(f). \tag{34}
\]

In this scheme, spectrum detection is possible even when the channel impulse responses are unknown. If \( \tilde{x}_j(f) \neq 0 \), then some primary user is using the channel unless the channel suffers from deep fades. Since it is unlikely that all the CR suffer from deep fades at the same time, cooperative spectrum sensing has a distinct advantage over individual sensing.

Cooperative spectrum sensing is carried out in two steps. First, compressed spectrum sensing is carried out at each individual CR using the approach in [54]. In addition, the \( j \)th CR maintains a binary \( F \)-dimensional occupation vector \( u_j \), where \( u_{j,i} = 1 \) if the frequency band is sensed to be occupied and \( u_{j,i} = 0 \) otherwise. Second, CRs in the one-hop neighborhood exchange the occupation vectors and then update their own occupation vectors using the technique of average consensus [59] as follows

\[
u_j(t + 1) = u_j(t) + \sum_{k \in N_j} w_{j,k}(u_k(t) - u_j(t)) \tag{35}
\]
where \( N_j \) is the neighborhood of \( j \)th CR, and \( w_{j,k} \) is the weight associated with edge \((j, k)\), the selection rules of which are described in [59]. With properly selection rules, it can be shown that
\[
\lim_{t \to \infty} u_j(t) = \frac{1}{J} \sum_{k=1}^{J} u_k(0).
\] (36)

In other words, the frequency occupation vectors of the CRs in the neighborhood all reach the same value that is the average of their initial values. The \( j \)th CR can decide the frequency band \( i \) is occupied if \( w_{j,i} \geq 1/J \), or a majority rule can be used, e.g., the frequency band \( i \) is considered occupied if \( w_{j,i} \geq 1/2 \). Simulation results show that the spectrum sensing performance improves with the number of CRs involved in the cooperative sensing and the average consensus converges fast (in a few iterations).

C. MIMO Signals

Channel state information (CSI) is essential for coherent communication over multi-antenna (MIMO) channels. Convention holds that the MIMO channel exhibits rich multi-path behavior and the number of degree of freedom is proportional to the dimension of the signal space. However, in practice the impulse responses of MIMO channel actually are dominated by a relatively small number of dominant paths. This is especially true with large bandwidths, long signaling duration, or large number of antennas [34], [57]. Because of this sparsity in the multi-path signals, CS can be used to improve the performance in channel estimation.

Consider a MIMO channel with \( N_T \) transmitters and \( N_R \) receivers. Assume that channel has two-sided bandwidth of \( W \) and the signaling has a duration of \( T \). Let \( s(t) \) denote the \( N_T \)-dimensional transmitted signal, \( \hat{s}(f) \) its element-wise Fourier transform, \( h(t, f) \) the time-varying frequency response matrix, which has a dimension of \( N_R \times N_T \). Without the noise, the received signal is given by [2]
\[
x(t) = \int_{-W/2}^{W/2} h(t, f) \hat{s}(f) e^{j2\pi ft} df.
\] (37)

For a multi-path channel, the frequency response is the summation of the contributions from all the paths
\[
h(t, f) = \sum_{n=1}^{N_P} \beta_n a_R(\theta_{R,n}) a_T^H(\theta_{T,n}) e^{j2\pi \nu_n t} e^{j2\pi \tau_n f}
\] (38)
where \( N_P \) denotes the number of paths. For path \( n \), \( \beta_n \) denotes the complex path gain, \( \theta_{R,n} \) the angle of arrival (AoA) at the receiver, \( \theta_{T,n} \) the angle of departure (AoD) at the transmitter, \( \nu_n \) the Doppler shift, and \( \tau_n \) the relative delay. The \( N_T \)-dimensional vector \( a_R(\theta_{R,n}) \) is array response vector at the receiver. The \( N_T \)-dimensional vector \( a_T(\theta_{T,n}) \) is array steering vector at the transmitter, and the superscript \( H \) denotes matrix conjugate transpose. We assume that the maximum delay is \( \tau_{max} \), then \( \tau \in [0, \tau_{max}] \). Also, the two-sided Doppler spread is in the range \( \nu \in [-\nu_{max}/2, \nu_{max}/2] \). The maximum antenna angular spread is assumed at the critical antenna spacing \((d = \lambda/2)\), thus \( (\theta_{R,n, \theta_{T,n}}) \in [-1/2, 1/2] \times [-1/2, 1/2] \) in the normalized unit. It is also assumed that the channel is both time-selective \((T\nu_{max} \leq 1)\) and frequency-selective \((W\tau_{max} \leq 1)\).

The physical model expressed by (38) is nonlinear and hard to analyze. However, it can be well-approximated by a linear model known as a virtual channel model [50], [51]. The virtual model approximates the physical model by uniformly sampling the physical parameter space \([\beta_n, \theta_{R,n, \theta_{T,n}, \tau_n, \nu_n}]\) at a resolution of \((\Delta \theta_{R,n}, \Delta \theta_{T,n}, \Delta \tau_n, \Delta \nu_n) = (1/N_R, 1/N_T, 1/W, 1/T)\). The approximate channel response is given by
\[
h(t, f) \simeq \sum_{i=1}^{N_R} \sum_{k=1}^{N_T} \sum_{l=0}^{L-1} \sum_{m=-M}^{M} h_v(i, k, l, m) a_R(i/N_R)a_T^H(k/N_T)e^{j2\pi \nu_{i,k,l,m} t}e^{j2\pi \tau_{i,k,l,m} f}
\] (39)
\[
h_v(i, k, l, m) \simeq \sum_{n \in \{\text{sampling point}\}} \beta_n
\] (40)
where the summation in (40) is over all paths that contribute to the sampling point, and a phase and attenuation factor has been absorbed in \( \beta_n \). In [59], \( N_R, N_T, L = [W\tau_{max}] + 1, \) and \( M = [T\nu_{max}/2] \) denote the maximum numbers of resolvable AoA, AOD, delays, and one-sided Doppler shifts. Basically, the virtual model characterizes the physical channel using the matrix \( h_v \), which has a dimension of \( D = N_R \times N_T \times L \times (2M + 1) \).

For sparse MIMO channels, the number of nonzero elements \( d \) in the matrix \( h_v \) is far fewer than \( D \). We call such channel \( d \)-sparse. Since the virtual model is linear, we can express the received training signal as
\[
y_{tr} = \Phi H_v.
\] (41)
where \( H_v \) is a \( D \)-dimensional column vector that contains all the elements in \( h_v \) ordered according to the index set \((i, k, l, m)\), and \( \Phi \) is a \( M \times D \)-dimensional measurement matrix, which is a function of transmitted training signal and array steering and response vectors. We can formulate a CS problem as follows
\[
H_v = \arg \min_{H_v} ||H_v||_1 \ \text{subject to} \ \ y_{tr} = \Phi H_v.
\] (42)

The requirement for the measurement is that it is a uniformly random sampling of the signal in the domain of \((i, k, l, m)\). It is shown in [2] that with high probability of success, roughly \( d \) instead of \( D \) measurements are enough to recover the \( D \)-dimensional channel vector \( H_v \), which provides significant savings in training resources consumed, especially for sparse-MIMO channels. The details are given in [2]. Similar results are also reported in [34], [53] for RF signals and in [4] for underwater acoustic signals.

D. CS as Erasure Code

Many physical phenomena are compressible or sparse in some domain. For example, virtual images are sparse in the wavelet domain and sound signals are sparse in the frequency domain. The conventional approach is to use source coding to compress the signal first and then use erasure coding to protect against the missing data caused by the noisy wireless channel.
Let \( x \) denote the \( n \)-dimensional signal. Assume \( S \) and \( E \) be the \( m_{cp} \times n \) source coding matrix and \( l \times m \) erasure coding matrix, respectively. If the signal is \( k \)-sparse in some domain, \( m_{cp} \) is close to \( k \). If the expected probability of missing data is \( p \), then \( l = m/(1-p) \). The transmitted signal is \( ESx \), and the received signal is \( CESx \), where the linear operator \( C \) models the channel. \( C \) is sub-matrix of the identity matrix \( I_l \) with \( e \) rows deleted, \( e \) being the number of erasures. If \( e \leq l - m \), the decoding at the receiver will be successful; otherwise the data can not be decoded and is discarded.

CS can be used as an effective erasure coding method. In [10], CS was applied in wireless sensor networks as erasure code. At the source \( l \) measurements of the signal are generated by random projections \( y = \Phi x \) and sent out, where \( \Phi \) is a \( l \times n \)-dimensional random matrix. The received signal is \( y' = Cy \), with \( e \) measurements erased. Suppose each measurement carries its serial number, we know where the erasure occurred and therefore the matrix \( C \). At the receiver, the standard CS procedure is carried out

\[
  x = \arg \min_x ||\Psi x||_1 \text{ subject to } y' = C\Phi x. \tag{43}
\]

where \( \Psi \) is the representation, under which \( x \) is sparse. Since erasure occurs randomly, \( C\Phi \) is still a random matrix. CS performs data compression and erasure coding in one stroke. Information about the signal is spread out among \( l \) measurements, of which \( m \) measurements are expected to be correctly received, with \( m \) on the same order of \( k \), the sparsity of the data.

Compared with conventional erasure coding methods, CS has similar compression performance but two outstanding advantages: 1) CS allows graceful degradation of the reconstruction error when the amount of missing data exceeds the designed redundancy, whereas the conventional coding methods does not. Specifically, if \( e \leq l - m \), conventional decoding can not recover the data at all. However, if RIP holds, according to (16), CS can still recover partial data, with an error no larger than that of the approximate signal, which keeps the largest \( l - e \) elements of the sparse signal and sets the rest of the elements to zero. 2) In terms of energy consumed in processing, performing CS erasure coding is 2.5 times better than performing local source coding and 3 times better than sending raw samples. [10].

VI. APPLICATIONS OF CS IN NETWORK AND TRANSPORT LAYERS

As mentioned before, most natural phenomena are sparse in some domain, and CS could be effective in wireless sensor networks (WSN) that monitor such phenomena. In [1], CS was used to gather data in a single-hop WSN. Sensors transmit random projections of their data simultaneously in a phase-synchronized channel. The base station receives the summation of the randomly projected data, which constitutes a CS measurement. The \( l_1 \)-norm minimization is performed at the base station to recover the sensory data. An overview of CS’ potential applications in WSN was provided in [33]. In the following, we provide a few detailed examples. We first present a representative CS-based sensory data collection scheme and then present three approaches to reduce measurement cost: 1) using the joint-sparsity in data to reduce the number of measurements, 2) using sparse random project to reduce number of measurements, and 3) using data routing to reduce the cost of data transport. We compare the performance of various data collection schemes at the end of the section.

A. Compressive Data Gathering (CDG) in WSN

In [41], CS was used for data collection in WSN. In the following, we describe the proposed data collection scheme, and how it handles the abnormal sensor readings.

1) The CS-Based Data Collection Scheme: Consider a large-scale sensor network with \( n \) nodes, each of which holds a value \( x_i \). Assume that a shortest-path spanning tree routed at the base station is built, and that the base station and sensors agree to the seeds for random number generation. To collect the data, \( m \) measurements are taken. For measurement \( j \), the data transmissions start from the leaf-nodes of the spanning tree and work in rounds. In each round, the children send their data to the parent in the tree. Specifically, in the first round, each leaf-node sends the random projection of its data \( (y_{j,i} = \Phi_{j,i}x_i) \) to its parent in the tree. The parent node \( l \) collects data sent by all its children and compute an partial measurement as follows

\[
  y_{j,l} = \sum_{i \in N_l} y_{j,i} + \Phi_{j,l}x_l \tag{44}
\]

where \( N_l \) denote the children of node \( l \). The parent node in turn sends its partial measurement to its own parent. This continues until the base station receives the partial measurements from its children and compute the complete measurement \( y_j = \sum_{i=1}^{n} \Phi_{j,i}x_i \). The base station is able to recover the data by solving the CS problem below

\[
  x = \arg \min_x ||\Psi x||_1 \text{ subject to } ||y - \Phi x||_2 < \epsilon. \tag{45}
\]

where \( \Psi \) is the basis under which the data is sparse, and \( \epsilon \) is the error tolerance.

Using convention methods, the energy consumption in data transmission is very unevenly distributed. The children of the base station are responsible to relaying \( O(n) \) pieces of data, causing them to die quickly due to battery depletion. On the other hand, in CDG each node transmit exactly \( m \) times, and the energy consumption is perfectly balanced.

One shortcoming of CDG is the nodes close to the leaves of the network are required to send more pieces of data than that of the conventional method. For example, a leaf-node is required to send one piece of data (its own sensory reading) in the conventional method, but with CS it is required to send \( m \) pieces of data. To solve this problem, a hybrid CS scheme is proposed, where if a node sends less that \( m \) pieces of data in the conventional method, the conventional method is used, otherwise CS is used [42].

2) Data Recovery with Abnormal Readings: CS can also used to recover data with abnormal readings. For example, if a signal is smooth in the time domain, then its representation in the Fourier domain is sparse. If spikes are injected into the time domain signal, then the signal’s representation in the frequency
domain is no longer sparse. However, we can decompose the signal into two parts

\[ x = x_0 + x_1 \]  

(46)

where \( x_0 \) and \( x_1 \) represent normal and abnormal parts, respectively. Since the spikes are sparse in the time domain and the normal signal is sparse in the frequency domain, the \( l_1 \)-norm minimization is altered to

\[ x = \text{arg min}_x \| \Psi x_0 + I x_1 \|_1 \text{ subject to } \| y - \Phi x \|_2 < \varepsilon. \]  

(47)

where \( I \) is the identity matrix. A representation like \( \Psi x_0 + I x_1 \) is called a over-complete representation. Donoho et al. showed it is feasible to achieve stable recovery of the signal in over-complete representations [22].

\[ \text{Fig. 2. SRP: the node in the center receives data from a randomly selected subset of nodes (dark ones) in the network.} \]

B. Distributed Compressed Sensing (DCS) in WSN

Consider a WSN monitoring an natural phenomenon. Among the signals obtained by the sensors, there are likely both intra-signal and inter-signal correlations. Leveraging these correlations, DCS uses the joint sparsity of the signals to reduce the number of measurements for signal recovery [5], [23], in a fashion similar to MMV described in Section IV.A but with some differences in problem formulation. DCS requires no collaboration among the sensors and provides universal encoder for any jointly sparse signal ensemble. In the following, we first introduce three joint sparsity models and then describe signal recovery algorithms.

1) Joint Sparsity Models (JSM): Assume the signals obtained by the sensors are \( x_j, j = 1, 2, \ldots, J \), and they are sparse in basis \( \Psi \).

**JSM1**: In this model, each signal is composed of a common sparse part and an individual sparse part as follows

\[ x_j = z_0 + z_j \text{ with } z_0 = \Psi \theta_0, z_j = \Psi \theta_j \]  

(48)

where the coefficients \( \theta_0 \) and \( \theta_j \) are \( k_0 \) and \( k_j \)-sparse, respectively. An example of this model is the temperature signals, which can be decomposed into a global average value plus a value reflecting local variations.

**JSM2**: In this model, all signals share the same index set of nonzero coefficients in sparse representation, and different signals have different individual coefficients as follows

\[ x_j = \Psi \theta_j \]  

(49)

where the coefficients \( \theta_j \) are \( k_j \)-sparse. An example of this model is the image observed by multiple sensors, which has the same sparse wavelet-representation but each sensor senses a different value due to different levels of phase-shift and attenuation.

**JSM3**: This model is an extension of JSM1 in that the common part is no longer sparse in any basis

\[ x_j = z_0 + z_j \text{ with } z_0 = \Psi \theta_0, z_j = \Psi \theta_j \]  

(50)

where the coefficients \( \theta_j \) are \( k_j \)-sparse. An example of this model is the signals detected by sensors in the presence of strong noise, which is not sparse in any representation.

2) Joint Reconstruction: To reconstruct the signal, sensor \( j \) acquires its \( n \)-dimensional signal \( x_j \), takes random projections of the signal \( y_j = \Phi x_j \), where \( \Phi \) is the \( m_j \times n \)-dimensional measurement matrix, and sends \( y_j \) to the base station. After receiving the measurements from all the sensors, the base station starts reconstruction, which is different for each of the sparsity models and is described separately below.

**JSM1**: To recover the signal, the following linear program is solved

\[ \{ \theta_0, \theta_1, \theta_2, \ldots, \theta_J \} = \text{arg min}_{\theta_0, \theta_1, \theta_2, \ldots, \theta_J} \sum_{j=0}^{J} \| \theta_j \|_1 \text{ subject to } y_j = \Phi \Psi \theta_j \forall j = 0, 1, 2, \ldots, J \]  

(51)

**JSM2**: In this model, conventional greedy pursuit algorithms (such as OMP) are modified. Specifically, step 7 in Algorithm 1 is modified so that the index set inserted to \( S_j \) includes only the indices of the common support, i.e., the nonzero items having the same indices among all sensors.

**JSM3**: In this model, each sensor’s signal is the addition of a common part \( z_0 \) that is not sparse and a sparse signal \( z_j \) called innovation. The alternating common and innovation estimation (ACIE) scheme was proposed to recover signals [23]. ACIE alternates between two steps: 1) Estimate \( z_0 \) by treating \( z_j \) as noise that can be averaged out. 2) Estimate \( z_j \) by subtracting \( z_0 \) from the signal and using conventional CS recovery techniques.

Extensive simulations have been carried out that demonstrate CS leveraging joint sparsity models can significantly reduce the number of measurements required. For detail, the reader is refer to [5].

C. Sparse Random Projection (SRP) in WSN

The measurement matrix \( \Phi \) used in conventual CS is dense in the sense that there are few zero elements in each row of \( \Phi \). This means each measurement requires data from \( O(n) \) sensors, which is expensive to acquire. Fortunately, it was shown in [60] that sparse random projections (SRP) can be
used to reduce the cost of measurement, where the measurement matrix is given by
\[
\phi_{j,i} = \begin{cases} 
+1 & \text{with prob. } p = 1/2s \\
-1 & \text{with prob. } p = 1/2s \\
0 & \text{with prob. } p = 1 - 1/s
\end{cases} \tag{52}
\]
where \(s\) is a parameter that determines the sparseness of the measurement. In other orders, SRP requires data from only \(O(n/s)\) instead of \(O(n)\) sensors. In order to recover the a \(k\)-sparse signal with high probability, i.e. with probability \(1 - n^{-\gamma}\) for some positive \(\gamma\), the number of measurements required is given by
\[
m = O(sM^2k^2 \log n) \tag{53}
\]
where \(M\) is the peak-to-total energy ratio and is given by
\[
M = \frac{\|x\|_\infty}{\|x\|_2}. \tag{54}
\]
\(M\) bounds the largest element of the signal. SRP works well only if the signal is not too concentrated in a few elements.

A distributed algorithm for data collection in WSN was proposed in [60], which works in two steps. In the first step, assuming the base station and sensors agree to the seeds for random number generation, each sensor \(i\) generates a random projection \(\phi_{j,i}x_i\). If \(\phi_{j,i}\) is zero, nothing needs to done, otherwise the sensor sends the random projection to another sensor chosen randomly, referring to Figure 3. Since \(\Phi\) is \(s\)-sparse, each sensor sends \(n/s\) pieces of data. In the second step, sensor \(j\) waits until it receives \(n/s\) pieces of data and then computes a measurement \(y_j = \sum_{i=1}^{n} \phi_{j,i}x_i\). The base station sends query to \(m\) sensors, and gets \(m\) measurements back. As long as \(m\) satisfies [53], the signal is recovered successfully with high probability.

There is a tradeoff involved in SRP. The larger the value of \(s\), the less the cost of data spreading among the sensors, but the higher the cost of the query from the base station. Suppose the average number of hops between two nodes is \(h\). The data spreading in SRP requires \(hn^2/s\) transmissions. The data query requires \(O(hsM^2k^2 \log n)\) transmissions. The optimum value of \(s\) is given by
\[
s = O\left(\frac{n}{Mk\sqrt{\log n}}\right) \tag{55}
\]

D. Optimizing Data Routing for CS in WSN

Conventional CS assumes each measurement costs the same. In WSN, the assumption is no longer true. Each measurement is a linear combination of data from a number of different sensors, which entails data transport cost. The design of the data collection scheme has implications on data transport cost. SRP reduces the cost of measurement by reducing the number of sensors required for each measurement. Another approach to reduce the measurement cost is to optimize data routing. Using this approach, a spatially localized compressed sensing and routing scheme was proposed [38], [40], where measurements were formed among adjacent sensors. A shortest-distance spanning tree is used to collect data to the base station.

A cautionary note was raised in [47], where the authors studied the interplay between routing and signal representation for compressed sensing in WSN. In WSN, each row of the measurement matrix \(\Phi\) actually represents a path or route, where the nonzero elements in the row represent the nodes in the path. The condition for good reconstruction quality is that \(\Phi\) and \(\Psi\) are incoherent, where \(\Psi\) is the basis on which the signal is sparse. The authors of [47] showed that the condition is not necessarily always met in practice. They used both synthetic and real data sets and considered a number of popular sparsifying transformations such as DCT, Haar Wavelet, etc.

E. Transport Cost of Data Collection Schemes

To collect all the values using conventional methods, \(O(n^{3/2})\) transmissions are required, since there are \(n\) node and each node has an average distance of \(O(n^{1/2})\) hops to the base station. In [48], a randomized gossiping scheme was proposed to collect sensory data using CS in WSN. The technique of average consensus similar to (36) is used. The basic gossip scheme incurs a data transport cost of \(O(n^2)\) transmissions for a network of \(n\) nodes, since it requires broadcasting data from each sensor to all other sensors.

Next, we consider CS data collection schemes based on the spanning tree. Suppose the signal \(x\) is \(k\)-sparse in some domain. According to [18], \(m\) measurements are sufficient to recover the signal, where \(m = ck\log(n/k)\). Using CS to collect data, \(O(nm) = O(nk\log n/k)\) transmissions are required, since \(n\) transmissions are required to collect one measurement in a spanning tree, and there are \(m\) measurements. Furthermore, if SRP is used, the data transportation cost is reduced by a factor \(s\), where \(s\) is the sparsity of the measurement as mentioned in the previous section. When \(k \ll n\), using CS can save transmission cost significantly. The transport costs of various data collection schemes are summarized in Table II.

VII. APPLICATIONS OF CS IN MAC AND TRANSPORT LAYERS

A. CS in On-Off Random Access Channels

In [29], a connection was made between CS and on-off random access channels. In an on-off random multiple access channel, there are \(N\) users communicating simultaneously to a single receiver through a channel with \(n\) degrees of freedom. Each user transmits with probability \(\lambda\). Typically, \(\lambda N < n \ll N\). User \(i\) is assigned as codeword a \(n\)-dimensional vector \(\phi_i\). The signal at the receiver from user \(i\) is \(\phi_i x_i\), where \(x_i\) is a nonzero complex scalar if the user is active and zero otherwise. The total signal at the receiver is given by
\[
y = \sum_{i=1}^{N} \phi_i x_i + w = \Phi x + w \tag{56}
\]
The goal of the receiver is to estimate \( \Omega \) by
\[
\hat{\Omega} = \arg\min_x \|x\|_1 + \|y - \Phi x\|_2^2
\]
where \( w \) is the noise, \( x = [x_1, x_2, ..., x_N] \), and \( \Phi = [\phi_1, \phi_2, ..., \phi_N] \) is the codebook. The active user set is defined by
\[
\Omega = \{i : x_i \neq 0\}. \tag{57}
\]
The goal of the receiver is to estimate \( \Omega \). Since \( |\Omega| \ll n \), we have a sparse signal detection problem. A formulation in terms of CS is as follows
\[
\hat{x} = \arg\min_x \mu \|x\|_1 + \|y - \Phi x\|_2^2 \tag{58}
\]
where \( \mu > 0 \) is an algorithm parameter that weights the importance of sparsity in \( \hat{x} \).

It was shown in [29] that the CS-based algorithms perform better than single-user detection in terms of the number of measurements required to recover signal, and have some near-far resistance. At high signal-to-noise ratio (SNR), CS-based algorithms perform worse than the optimal maximum likelihood detection. However, CS-based algorithms are computationally efficient, whereas optimal maximum likelihood detection is not computationally feasible.

B. CS in Network Performance Monitoring

Effective performance monitoring is essential for the operation of large-scale networks. The challenge that conventional monitoring techniques face is that they do not scale well, since only a small portion of a large-scale network can be monitored. It turns out CS can be used to overcome this challenge. In the following we describe two approaches of using CS for network performance monitoring: 1) directly applying CS for sparse signal monitoring, and 2) applying CS in the transform domain for network performance monitoring.

1) Applying CS Directly: To monitor the performance of the network, monitors are placed at the nodes to measure the end-to-end performance, such as delay, packet loss, etc. In a network of \( n \) links, the path-oriented performance measurements can be expressed by
\[
y = \Phi x \tag{59}
\]
where \( y \) is the \( m \)-dimensional path performance metrics vector, \( x \) is the \( n \)-dimensional link performance metrics vector, and \( \Phi \) is \( m \times n \)-dimensional binary matrix called routing matrix. If path \( i \) includes link \( j \), then \( \phi_{i,j} = 1 \), otherwise, \( \phi_{i,j} = 0 \). The above model applies to some metrics such as true delay, but not to others such as packet loss (actually, it can be made to be applicable after a log transformation).

Network administrators are generally interested in identifying a few severely congested links with large delays, compared to which delays of other links are negligible. In this sense, \( x \) is sparse. A CS-based scheme was proposed to recover \( x \) [27]. The problem remaining to be addressed is whether the routing matrix \( \Phi \) is a good measurement matrix. The authors of the paper leveraged the fact that routing matrices of bipartite expander graphs are good measurement matrices for CS [31].

A bipartite graph \( G(L, R, E) \) consists two sets of nodes: the left set \( L \) and the right set \( R \). There are edges of the graph only between nodes in \( L \) and \( R \). The routing matrix can be transformed into a bipartite graph by making \( L \) be composed the links in the network and \( R \) be composed of paths in the network. There is a link between node \( i \) in \( L \) and node \( j \) in \( R \), only if link \( i \) is in path \( j \).

A \((s, d, \epsilon)\)-expander graph is a bipartite graph \( G(L, R, E) \) with left degree \( d \) (all nodes in \( L \) having \( d \) edges), and for any subset \( S \subseteq L \) with \( |S| \leq s \), the following holds
\[
|N(S)| \geq (1 - \epsilon)d|S| \tag{60}
\]
where \( N(S) \) is the set of neighbors of \( S \). Parameters \( s \) and \( \epsilon \) are called expansion factor and error parameter. In other words, in expander graphs \( G(L, R, E) \), \( L \) is expansive in the sense that any subset \( S \subseteq L \) has a neighborhood size proportional to the size of \( S \).

It was shown that the bipartite matrix of a \((2s, d, \epsilon)\) expander graph can be used as a good measurement matrix of a \( s \)-sparse signal [31], [35]. Reference [27] considered the special case where \( x \) is 1-sparse. The following CS problem was formulated
\[
x = \arg\min_x \|x\|_1 \quad \text{subject to} \quad y = \Phi x. \tag{61}
\]
Let \( x^* \) denote the true delay vector. It was shown that if the network graph is a \((2, d, \epsilon)\)-expander with \( \epsilon \leq 1/4 \), then the following holds [27]
\[
\|x - x^*\|_1 \leq c(\epsilon)\|x_c\|_1 \tag{62}
\]
where \( c(\epsilon) \) is a constant dependent on \( \epsilon \), and \( x_c \) is \( x^* \) with the \( k \)-largest element removed \((k = 1 \text{ in this case})\). So, if the true delay vector is 1-sparse, then \( x_c = 0 \), i.e., the estimation error is zero.

One cautionary note: the above scheme applies only to cases where the network bipartite graphs are expanders. But for some networks, the partite graphs are not expanders, and such scheme is not applicable.

2) Applying CS in the Transform Domain: A CS-based network monitoring scheme was proposed in [11]. The CS-based scheme provides a scalable monitoring technique that require measurements on only a few end-to-end paths. Diffusion wavelet [12] was used as the sparsity-inducing basis. Diffusion wavelet is the generalization of wavelets that provides

| Data collection schemes | Data transport cost (\( n \): number of nodes, \( k \): data sparsity, \( s \): measurement sparsity) |
|------------------------|-------------------------------------------------|
| Conventional data collection | \( O(n^{3/2}) \) |
| CS Gossip [48] | \( O(n^2) \) |
| CS Spanning tree [38], [40] | \( O(nk \log n/k) \) |
| CS Spanning tree + SRP [38], [40], [60] | \( O(nk \log n/k/s) \) |
The reconstructed path metrics is given by

\[ y_s = Ay \] (63)

where \( A \) is the identity matrix with \( n_p - n_s \) rows deleted, retaining \( n_s \) rows corresponding to measurements \( y_s \).

A measurement graph \( G(V, E) \) can be constructed as follows. The vertices \( V \) of \( G \) corresponds to the paths of the physical network. There is an edge between two vertices \( v_i \) and \( v_j \), if \( i \) and \( j \) share a link in the physical network. A link is assigned a weight \( w_{i,j} \) to indicated the degree of correlation of performance metrics between two vertices (two physical paths). We can define the weight such that it is proportional to the fraction of shared physical links in the two paths. Let \( L_i \) denote the set of physical links in path \( i \). Thus, weight \( w_{i,j} \) is given by

\[ w_{i,j} = \left| \frac{L_i \cap L_j}{L_i \cup L_j} \right| \] (64)

We assign each vertex \( v_i \) with the value of performance metric of path \( i \). Thus, we obtain a performance metric function \( y(V) \) defined on the vertices \( V \) of \( G \). This function can be represented in diffusion wavelet basis as follows

\[ y = \sum_{i=1}^{n} \beta_i b_i = B\beta \] (65)

where \( b_i \)'s are orthonormal diffusion wavelets defined on the \( V \), \( \beta_i = y^T b_i \) is the \( i \)th wavelet coefficient, \( B \) is a \( n \times n \) matrix composed of \( b_i \)'s, and \( \beta \) is a \( n \)-dimensional vector composed of \( \beta_i \)'s.

By proper selection of the diffusion wavelet basis (the reader is refer to [11] for details), we can make \( \beta \)'s representation to be sparse, where CS is effective. What is given is \( y_s \), which is a subset of path metrics. The goal is to reconstruct \( \beta \), which is entire set of path metrics. We can formulate a CS problem as follows

\[ \hat{\beta} = \arg \min \| \beta \|_1 \quad \text{subject to} \quad y_s = Ay = AB\beta. \] (66)

The reconstructed path metrics is given by

\[ \hat{y} = B\hat{\beta}. \] (67)

The above method was applied to two case studies [12]. The first case study is monitoring the end-to-end delay in the Abilene network consisted of 11 nodes and 30 unidirectional links. It was shown that it takes only 3 measurements per time step to estimate the mean network end-to-end delay with an error of less than 10%. The error decreases with more time steps and eventually approaches zero. The second case study is monitoring the bit-error rate (BER) in all-optical networks.
[20] D. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via $l_1$ minimization," Proc. Natl. Acad. Sci., vol. 100, no. 5, pp. 2197-2202, 2003.

[21] D.L. Donoho, "Compressed sensing," IEEE Trans. Information Theory, vol. 52, no. 4 pp. 1289-1306, 2006.

[22] D. Donoho, M. Elad, and V. Temlyakov, "Stable recovery of sparse overcomplete representations in the presence of noise," IEEE Trans. Inform. Theory, vol. 52, no. 1, pp. 618, Jan. 2006.

[23] M. F. Duarte, M.B. Wakin, D. Baron, and R.G. Baraniuk, "Universal distributed sensing via random projections," in ACM Proceedings of the 5th international conference on Information processing in sensor networks (IPSN), pp. 177-185, 2006.

[24] M.F. Duarte and Y.C. Eldar, "Structured compressed sensing: From theory to applications," IEEE Transactions on Signal Processing, vol. 58, no. 9, pp. 4083-4098, 2010.

[25] Y.C. Eldar, "Generalized SURE for exponential families: Applications to regularization," IEEE Trans. Signal Processing, vol. 57, no. 2, pp. 471-481, 2009.

[26] M.A.T. Figueiredo, R.D. Nowak, and S. J. Wright, "Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems," IEEE J. Sel. Top. Signa., vol. 1, pp. 586597, 2007.

[27] J.H. Firooz and S. Roy, "Network tomography via compressed sensing," in IEEE Global Telecommunications Conference (GLOBECOM), 2010.

[28] J. Friedman, T. Hastie, and R. Tibshirani, "Regularization paths for logistic regression," Journal of Computational and Graphical Statistics, vol. 17, pp. 1-22, 2008.

[29] A.K. Fletcher, S. Rangan, and V.K. Goyal, "On-off random access channels: A compressed sensing framework," arXiv preprint arXiv:0903.1022, 2009.

[30] A.C. Gilbert, M.J. Strauss, and R. Vershynin, "One sketch for all: Fast algorithms for Compressed Sensing," in Proc. 39th ACM Sym. Theory of Computing (STOC), San Diego, CA, June 2007.

[31] A. Gilbert and P. Indyk, "Sparse recovery using sparse matrices," Proceedings of the IEEE, vol. 98, no. 6, pp. 937-947, 2010.

[32] M. Grant et al., CVX:MATLAB Software for Disciplined Convex Programming, version 1.21, 2010 [Online]. Available: http://cvxr.com/cvx

[33] J. Haupt, W.U. Bajwa, M. Rabbat, R. Nowak, "Compressed Sensing for Networked Data," IEEE Signal Processing Magazine, vol.25, no.2, pp.92,101, March 2008.

[34] J. Haupt, W.U. Bajwa, G. Raz, and R. Nowak, "Toeplitz Compressed Sensing Matrices With Applications to Sparse Channel Estimation," IEEE Transactions on Information Theory, vol.56, no.11, pp.5862,5875, Nov. 2010.

[35] P. Indyk and M. Ruzic, Near-optimal sparse recovery in the l1 norm, in Proc. 49th Annual IEEE Symposium on Foundations of Computer Science, pp. 199207, 2008.

[36] S. Kirollos, et al., Analog-to-information conversion via random demodulation, in Proc. of the IEEE Dallas Circuits and Systems Workshop (DCAS), 2006.

[37] J.N. Laska, et al., "Theory and implementation of an analog-to-information converter using random demodulation," in Proc. of the IEEE International Symposium on Circuits and Systems (ISCAS), pp. 1959-1962, 2007.

[38] S. Lee, S. Pattern, M. Satiamoorthy, B. Krishnamachari, and A. Ortega, "Spatially-localized compressed sensing and routing in multi-hop sensor networks," in Geosensor Networks, pp. 11-20, Springer Berlin Heidelberg, 2009.

[39] M.A. Iwen, "Combinatorial sublinear-time fourier algorithms," Found. of Comput. Math., 10:303338, 2010.

[40] S. Lee, S. Pattern, M. Satiamoorthy, B. Krishnamachari, and A. Ortega, "Compressed sensing and routing in multi-hop networks," University of Southern California CENG Technical Report, 2009.

[41] C. Luo, F. Wu, J. Sun, C.W. Chen, "Compressive data gathering for large-scale wireless sensor networks," in ACM International Conference on Mobile Computing and Networking (MobiCom), pp. 145-156, 2009.

[42] J. Luo; L. Xiang,C. Rosenberg, "Does Compressed Sensing Improve the Throughput of Wireless Sensor Networks?;" IEEE International Conference on Communications (ICC), 2010 , vol., no., pp.1.6, 23-27 May 2010.

[43] S.G. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," IEEE Trans. Signal Process., 41:3393415, 1993.

[44] S. Muthukrishnan, "Data Streams: Algorithms and Applications," Found. Trends in Theoretical Comput. Science, vol.1, Now Publishers, Boston, MA, 2005.