Sparse Representation for Wireless Communications

Zhijin Qin[1], Jiancun Fan[2], Yuanwei Liu[3], Yue Gao[3], and Geoffrey Ye Li[4]

[1]Lancaster University, Lancaster, UK
[2]Xi'an Jiaotong University, Xi'an, China
[3]Queen Mary University of London, London, UK
[4]Georgia Institute of Technology, Atlanta, GA, USA

Abstract—**Sparse representation can efficiently model signals in different applications to facilitate processing. In this article, we will discuss various applications of sparse representation in wireless communications, with focus on the most recent compressive sensing (CS) enabled approaches. With the help of the sparsity property, CS is able to enhance the spectrum efficiency and energy efficiency for the fifth generation (5G) networks and Internet of Things (IoT) networks. This article starts from a comprehensive overview of CS principles and different sparse domains potentially used in 5G and IoT networks. Then recent research progress on applying CS to address the major opportunities and challenges in 5G and IoT networks is introduced, including wideband spectrum sensing in cognitive radio networks, data collection in IoT networks, and channel estimation and feedback in massive MIMO systems. Moreover, other potential applications and research challenges on sparse representation for 5G and IoT networks are identified. This article will provide readers a clear picture of how to exploit the sparsity properties to process wireless signals in different applications.**

Keywords: Wireless communications, compressive sensing, sparsity property, 5G, Internet of Things.

I. INTRODUCTION

Sparse representation expresses some signals as a linear combination of a few atoms from a prespecified and over-complete dictionary [1]. This form of sparse (or compressible) structure arises naturally in many applications [2]. For example, audio signals are sparse in frequency domain, especially for the sounds representing tones. Image processing can exploit a sparsity property in the discrete cosine domain, i.e. many discrete cosine transform (DCT) coefficients of images are zero or small enough to be regarded as zero. This type of sparsity property has enabled intensive research on signal and data processing, such as dimension reduction in data science, wideband sensing in cognitive radio networks (CRNs), data collection in large-scale wireless sensor networks (WSNs), and channel estimation and feedback in massive MIMO.

Traditionally, signal acquisition and transmission adopt the procedure with sampling and compression. As massive connectivity is expected to be supported in the fifth generation (5G) networks and Internet of Things (IoT) networks, the amount of generated data becomes huge. Therefore, signal processing has been confronted with challenges on high sampling rates for data acquisition and large amount of data for storage and transmission, especially in IoT applications with power-constrained sensor nodes. Except for developing more advanced sampling and compression techniques, it is natural to ask whether there is an approach to achieve signal sampling and compression simultaneously.

As an appealing approach employing sparse representations, compressive sensing (CS) technique [3] has been proposed to reduce data acquisition costs by enabling sub-Nyqusit sampling. Based on the advanced theory [4], CS has been widely applied in many areas. The

key idea of CS is to enable exact signal reconstruction from far fewer samples than that is required by the Nyquist-Shannon sampling theorem provided that the signal admits a sparse representation in a certain domain. In CS, compressed samples are acquired via a small set of non-adaptive, linear, and usually randomized measurements, and signal recovery is usually formulated as an l_0-norm minimization problem to find the sparsest solution satisfying the constraints. Since l_0-norm minimization is an NP-hard problem, most of the exiting research contributions on CS solve it by either approximating it to a convex l_1-norm minimization problem [4] or adopting greedy algorithms, such as orthogonal match pursuit (OMP).

It is often the case that the sparsifying transformation is unknown or difficult to determine. Therefore, projecting a signal to its proper sparse domain is quite essential in many applications that invoke CS. In 5G and IoT networks, the identified sparse domains mainly include frequency domain, spatial domain, wavelet domain, DCT domain, etc. CS can be used to improve spectrum efficiency (SE) and energy efficiency (EE) for these networks. By enabling the unlicensed usage of spectrum, CRNs exploit spectral opportunities over a wide frequency range to enhance the network SE. In wideband spectrum sensing, spectral signals naturally exploit a sparsity property in frequency domain due to low utilization of spectrum [5], [6], which enables sub-Nyquist sampling on cognitive devices. Another interesting scenario is a small amount of data collection in large-scale WSNs with power-constrained sensor nodes, such as smart meter monitoring infrastructure in IoT applications. In particular, the monitoring readings usually have a sparse representation in DCT domain due to the temporal and spatial correlations [7]. CS can be applied to enhance the EE of WSNs and to extend the lifetime of sensor nodes. Moreover, massive MIMO is a critical technique for 5G networks. In massive MIMO systems, channels corresponding to different antennas are correlated. Furthermore, a huge number of channel coefficients can be represented by only a few parameters due to

a hidden joint sparsity property caused by the shared local scatterers in the radio propagation environment. Therefore, CS can be potentially used in massive MIMO systems to reduce the overhead for channel estimation and feedback and facilitate precoding [8]. Even though various applications have different characters, it is worth noting that the signals in different scenarios share a common sparsity property even though the sparse domains can be different, which enables CS to enhance the SE and EE of wireless communications networks.

There have been some interesting surveys on CS [9] and its applications [10]–[12]. One of the most popular articles on CS [9] has provided an overview on the theory of CS as a novel sampling paradigm that goes against the common wisdom in data acquisition. CS-enabled sparse channel estimation has been summarized in [10]. In [11], a comprehensive review of the application of CS in CRNs has been provided. A more specific survey on compressive covariance sensing has been presented in [12] that includes the reconstruction of second-order statistics even in the absence of prior sparsity information. These existing surveys serves different purposes. Some cover the basic principles for beginners and others focus on specific aspects of CS. Different from the existing literature, our article provides a comprehensive overview of the recent contributions on CS-enabled wireless communications from the perspective of adopting different sparse domain projections.

In this article, we will first introduce the basic principles of CS briefly. Then we will present the different sparse domains for signals in wireless communications. Subsequently, we will provide CS-enabled frameworks in various wireless communications scenarios, including wideband spectrum sensing in CRNs, data collection in large-scale WSNs, and channel estimation and feedback for massive MIMO, as they have been identified to be critical to 5G and IoT networks and share the same spirit by exploiting the sparse domains aforementioned. Within each identified scenario, we start with projecting a signal to a sparse domain, then in-

troduce the CS-enabled framework, and finally illustrate how to exploit joint sparsity in the CS-enabled framework. Moreover, the reweighted CS approaches for each scenario will be discussed, where the weights are constructed by prior information depending on specific application scenarios. The other potential applications and research challenges on applying CS in wireless networks will also be discussed and followed by conclusions.

This article gives readers a clear picture on the research and development of the applications of CS in different scenarios. By identifying the different sparse domains, this article illustrates the benefits and challenges on applying CS in wireless communication networks.

II. SPARSE REPRESENTATION

Sparse representation of signals has received extensive attention due to its capacity for efficient signal modelling and related applications. The problem solved by the sparse representation is to search for the most compact representation of a signal in terms of a linear combination of the atoms in an overcomplete dictionary. In the literature, three aspects of research on the sparse representation have been focused:

1) Pursuit methods for solving the optimization problem, such as matching pursuit and basis pursuit;
2) Design of the dictionary, such as the K-SVD method;
3) Applications of the sparse representation, such as wideband spectrum sensing, channel estimation of massive MIMO, and data collection in WSNs.

General sparse representation methods, such as principal component analysis (PCA) and independent component analysis (ICA), aim to obtain a representation that enables sufficient reconstruction. It has been demonstrated that PCA and ICA are able to deal with signal corruption, such as noise, missing data, and outliers. For sparse signals without measurement noise, CS can recover the sparse signals exactly with random measurements. Furthermore, the random measurements significantly outperform

measurements based on PCA and ICA for the sparse signals without corruption [13]–[15]. In the following, we will focus on the principles of CS and the common sparse domains potentially used in 5G and IoT scenarios.

A. Principles of Standard Compressive Sensing

The principles of standard CS, such as to be performed at a single node, can be summarized into the following three parts [3]:

1) Sparse Representation: Generally speaking, sparse signals contain much less information than their ambient dimension suggests. Sparsity of a signal is defined as the number of non-zero elements in the signal under a certain domain. Let $\mathbf{f}$ be an N-dimensional signal of interest, which is sparse over the orthonormal transformation basis matrix $\boldsymbol{\Psi} \in \mathbb{R}^{N \times N}$, and $\mathbf{s}$ be the sparse representation of $\mathbf{f}$ over the basis $\boldsymbol{\Psi}$. Then $\mathbf{f}$ can be given by

$$\mathbf{f} = \boldsymbol{\Psi}\mathbf{s}. \tag{1}$$

Apparently, $\mathbf{f}$ can be the time or space domain representation of a signal, and $\mathbf{s}$ is the equivalent representation of $\mathbf{f}$ in the $\boldsymbol{\Psi}$ domain. For example, if $\boldsymbol{\Psi}$ is the inverse Fourier transform (FT) matrix, then $\mathbf{s}$ can be regarded as the frequency domain representation of the time domain signal, $\mathbf{f}$. Signal $\mathbf{f}$ is said to be K-sparse in the $\boldsymbol{\Psi}$ domain if there are only K $(K \ll N)$ out of the N coefficients in $\mathbf{s}$ that are non-zero. If a signal is able to be sparsely represented in a certain domain, the CS technique can be invoked to take only a few linear and non-adaptive measurements.

2) Projection: When the original signal $\mathbf{f}$ arrives at the receiver, it is processed by the measurement matrix $\boldsymbol{\Phi} \in \mathbb{R}^{P \times N}$ with $P < N$, to get the compressed version of the signal, that is,

$$\mathbf{x} = \boldsymbol{\Phi}\mathbf{f} = \boldsymbol{\Phi}\boldsymbol{\Psi}\mathbf{s} = \boldsymbol{\Theta}\mathbf{s}, \tag{2}$$

where $\boldsymbol{\Theta} = \boldsymbol{\Phi}\boldsymbol{\Psi}$ is an $P \times N$ matrix, called the sensing matrix. As $\boldsymbol{\Phi}$ is independent of signal $\mathbf{f}$, the projection process is non-adaptive.

Fig. 1 illustrates how the different sensing matrices $\boldsymbol{\Theta}$ influence the projection of a signal from high dimension to its space, i.e., mapping

$\mathbf{s} \in \mathbb{R}^3$ to $\mathbf{x} \in \mathbb{R}^2$. As shown in Fig. 1, $\mathbf{s} = \begin{pmatrix} s & s & 0 \end{pmatrix}$ is a three-dimensional signal. When $\mathbf{s}$ is mapped into a two-dimensional space by taking $\mathbf{\Theta_1} = \begin{pmatrix} 1 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$ as the sensing matrix, the original signal $\mathbf{s}$ cannot be recorded based on the projection under $\mathbf{\Theta_1}$. This is because that the plane spanned by the two row vectors of $\mathbf{\Theta_1}$ is orthogonal to signal $\mathbf{s}$ as shown in Fig. 1(a). Therefore, $\mathbf{\Theta_1}$ corresponds to the worst projection. As shown in Fig. 1(b), we can also observe that the projection by taking $\mathbf{\Theta_2} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$ is not a good one. It is noted that the plane spanned by the two row vectors of $\mathbf{\Theta_2}$ can only contain part of information of the sparse signal $\mathbf{s}$, and the sparse component in the direction of $\mathbf{s_2}$ is missed when the signal $\mathbf{s}$ is projected into the two-dimensional space. When the sensing matrix is set to $\mathbf{\Theta_3} = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$, as shown in Fig. 1(c), the signal $\mathbf{s}$ can be fully recorded as it falls into the plane spanned by the two row vectors of $\mathbf{\Theta_3}$. Therefore, $\mathbf{\Theta_3}$ results in a good projection and $\mathbf{s}$ can be exactly recovered by its projection $\mathbf{x}$ in the two-dimensional space. Then it is natural to ask what type of projection is good enough to guarantee the exact signal recovery?

The key of CS theory is to find out a stable basis $\mathbf{\Psi}$ or measurement matrix $\mathbf{\Phi}$ to achieve exact recovery of the signal with length N from P measurements. It seems an undetermined problem as $P < N$. However, it has been proved in [4] that exact recovery can be guaranteed under the following conditions:

- Restricted isometry property (RIP): Measurement matrix $\mathbf{\Phi}$ has the RIP of order K if

$$1 - \delta_K \leq \frac{\|\mathbf{\Phi f}\|_{\ell_2}^2}{\|\mathbf{f}\|_{\ell_2}^2} \leq 1 + \delta_K \qquad (3)$$

holds for all K-sparse signal $\mathbf{f}$, where δ_K is the restricted isometry constant of a matrix $\mathbf{\Phi}$.
- Incoherence property: Incoherence property requires that the rows of measurement matrix $\mathbf{\Phi}$ cannot sparsely represent the columns of the sparsifying matrix $\mathbf{\Psi}$ and vice verse. More specifically, a good measurement will pick up a little bit information of each component in $\mathbf{s}$ based on the condition that $\mathbf{\Phi}$ is incoherent with $\mathbf{\Psi}$. As a result, the extracted information can be maximized by using the minimal number of measurements.

It has been pointed out that verifying both the RIP condition and incoherence property is computationally complicated but they could be achieved with a high probability simply by selecting $\mathbf{\Phi}$ as a random matrix. The common random matrices include Gaussian matrix, Bernoulli matrix, or almost all others matrices with independent and identically distributed (i.i.d.) entries. Besides, with the properties of the matrix with i.i.d. entries $\mathbf{\Phi}$, the matrix $\mathbf{\Theta} = \mathbf{\Phi\Psi}$ is also random i.i.d., regardless of the choice of $\mathbf{\Psi}$. Therefore, the random matrices are universal as they are random enough to be incoherent with any fixed basis. It has been demonstrated that random measurements can universally capture the information relevant for many compressive signal processing applications without any prior knowledge of either the signal class and its sparse domain or the ultimate signal processing task.

Moreover, for Gaussian matrices the number of measurements required to guarantee the exact signal recovery is almost minimal. However, random matrices inherently have two major drawbacks in practical applications: huge memory buffering for storage of matrix elements, and high computational complexity due to their completely unstructured nature [16]. Compared to the standard CS that limits its scope to standard discrete-to-discrete measurement architectures using random measurement matrices and signal models based on standard sparsity, more structured sensing architectures, named structured CS, have been proposed to implement CS on feasible acquisition hardware. So far, many efforts have been put on the design of structured CS matrices, i.e., random demodulator [17], to make CS implementable with expense of performance degradation. Particularly,
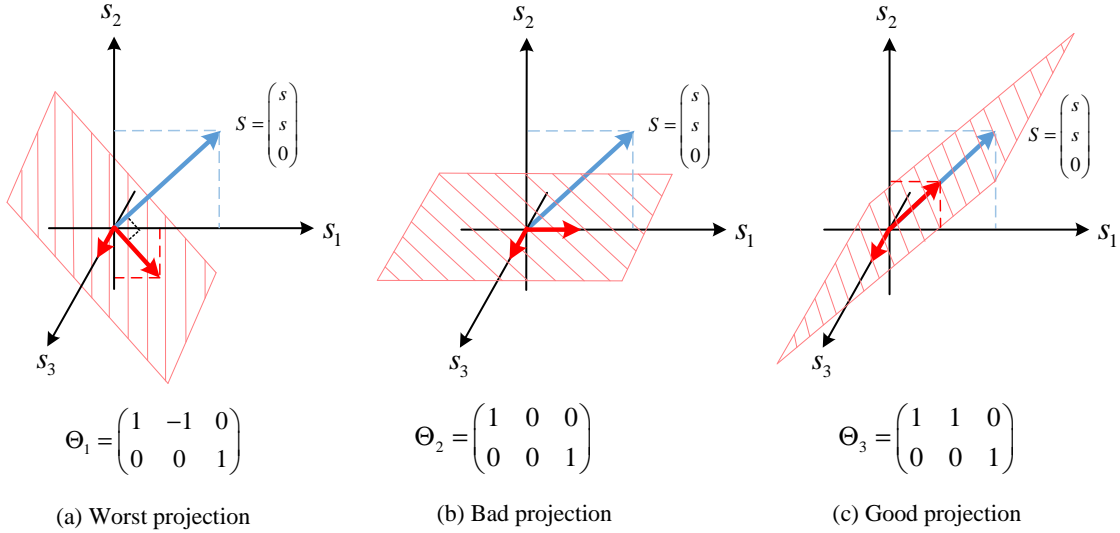
Fig. 1: Projection of a sparse signal with one non-zero component with different sensing matrices.

the main principle of random demodulator is to multiply the input signal with a high-rate pseudonoise sequence, which spreads the signal across the entire spectrum. Then a low-pass anti-aliasing filter is applied and the signal is captured by sampling it at a relatively low rate. With the additional digital processing to reduce the burden on the analog hardware, random demodulator bypasses the need for a high-rate analogue-to-digital converter (ADC) [17]. A comparison of Gaussian sampling matrix and random demodulator is provided in Fig. 2 in terms of detection probability with different compression ratios P/N. From the figure, the Gaussian sampling matrix performs better than the random demodulator.

3) Signal Reconstruction: After the compressed measurements are collected, the original signal should be reconstructed. Since most of the basis coefficients in **s** are negligible, the original signal can be reconstructed by finding out the minimal set of coefficients that matches the set of compressed measurements **x**, that is, by solving

$$\hat{\mathbf{s}} = \arg\min_{\mathbf{s}} \|\mathbf{s}\|_{\ell_p} \text{ subject to } \boldsymbol{\Theta}\mathbf{s} = \mathbf{x}, \quad (4)$$

where $\|\cdot\|_{\ell_p}$ is the ℓ_p-norm and $p = 0$ corresponds to counting the number of non-zero elements in **s**. However, the reconstruction
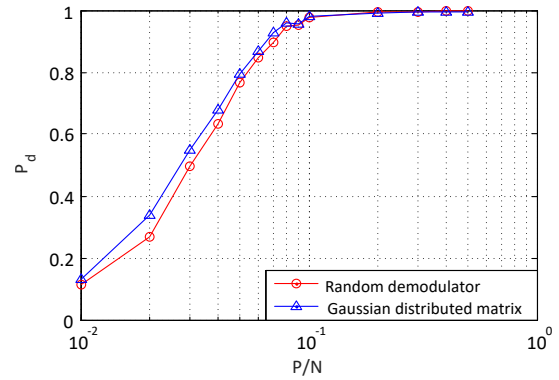

Fig. 2: Detection probability versus compression ratio with different measurement matrices. In this case, the signal is one-sparse and the simulation iteration is 1000.

problem in (4) is both numerically unstable and NP-hard [3] when ℓ_0-norm is used.

So far, there are mainly two types of relaxations to problem (4) to find a sparse solution. The first type is convex relaxation, where ℓ_1-norm is used to substitute ℓ_0-norm in (4). Then (4) can be solved by standard convex solvers, e.g., cvx. It has been proved that ℓ_1 norm results in the same solution as ℓ_0 norm when RIP is satisfied with the constant $\delta_{2k} < \sqrt{2} - 1$ [18]. Another type of solution is to use a greedy algorithm, such as OMP [19], to find a local optimum in each iteration. In com-

parison with the convex relaxation, the greedy algorithm usually requires lower computational complexity and time cost, which makes it more practical for wireless communication systems. Furthermore, the recent result has shown that the recovery accuracy achieved by some greedy algorithms is comparable to the convex relaxation but requiring much lower computational cost [20].

B. Reweighted Compressive Sensing

As aforementioned, ℓ_1-norm is a good approximation for the NP-hard ℓ_0-norm problem when RIP holds. However, the large coefficients are penalized more heavily than the small ones in ℓ_1-norm minimization, which leads to performance degradation on signal recovery. To balance the penalty on the large and the small coefficients, reweighted CS is introduced by providing different penalties on those large and small coefficients. A reweighted ℓ_1-norm minimization framework [3] has been developed to enhance the signal recovery performance with fewer compressed measurements by solving

$$\hat{\mathbf{s}} = \arg\min_{\mathbf{s}} \|\mathbf{W}\mathbf{s}\|_{\ell_1} \quad \text{subject to} \quad \mathbf{\Theta}\mathbf{s} = \mathbf{x},$$
(5)

where $\mathbf{W}$ is a diagonal matrix with $w_1, \ldots, w_n$ on the diagonal and zeros elsewhere.

Moreover, ℓ_p-norm, e.g. $0 < p < 1$, is utilized to lower the computational complexity of signal recovery process caused by the ℓ_1-norm optimization problem. Iterative reweighted least-square (IRLS) based CS approach has been proposed in [21] to solve (4) in a non-convex approach as

$$\hat{\mathbf{s}} = \arg\min_{\mathbf{s}} \sum_{i=1}^{N} w_i s_i \quad \text{subject to} \quad \mathbf{\Theta}\mathbf{s} = \mathbf{x},$$
(6)

where $w_i = \left| s_i^{(l-1)} \right|^{p-2}$ is computed based on the result of the last iteration, $s_i^{(l-1)}$.

It is worth noting that (4) becomes non-convex when $p < 1$. The existing algorithms cannot guarantee to reach a global optimum and may only produce local minima. However, it

has been proved [22], [23] that under some circumstances the reconstruction in (4) will reach a *unique* and *global* minimizer [24], which is exactly $\hat{\mathbf{s}} = \mathbf{s}$. Therefore, we can still exactly recover the signal in practice.

C. Distributed Compressive Sensing

The distributed compressive sensing (DCS) [25] is an extension of the standard one by considering networks with M nodes. At the m-th node, measurement x_m can be given by

$$x_m = \Theta_m s_m, \quad \forall m \in \mathcal{M},$$
(7)

where $\mathcal{M}$ is the the set of nodes in the network. As stated in (2), Θ_m is the sensing matrix deployed at the m-th node, and s_m is a sparse signal of interest. DCS becomes a standard CS when $M = 1$.

In the applications of standard CS, the signal received at the same node has its sparsity property due to its intra-correlation. While for the networks with multiple nodes, signals received at different nodes exhibit strong inter-correlation. The intra-correlation and inter-correlation of signals from the multiple nodes lead to a joint sparsity property. The joint sparsity level is usually smaller than the aggregate over the individual signal's sparsity level. As a result, the number of compressed measurements required for exact recovery in DCS can be reduced significantly compared to the case performing standard CS at each single node independently.

In DCS, there are two closely related concepts: distributed networks and distributed CS solvers. The distributed networks refer to networks that different nodes perform data acquisition in a distributed way and the standard CS can be applied at each node individually to perform signal recovery. While for DCS solver as proposed in [25], the data acquisition process requires no collaboration among sensors and the signal recovery process is performed at several computational nodes, which can be distributed in a network or locally placed within a multiple core processor. Generally, it is of interest to minimize both computation cost and communication overhead in DCS. The most

popular application scenario of DCS is that all signals share the common sparse support but with different non-zero coefficients.

D. Common Sparse Domains for CS-enabled 5G and IoT Networks

CS-enabled sub-Nyquist sampling is possible only if the signal is sparse in a certain domain. The common sparse domains utilized in CS-enabled 5G and IoT networks include frequency domain, wavelet domain, discrete cosine domain, angular domain, to name a few.

1) Frequency domain: due to low spectrum utilization, the wideband spectrum signal shows a sparse property when it is converted into the frequency domain.
2) Discrete cosine domain: due to temporal correlation, signals in some applications, such as environment information monitoring, show a sparse property in the discrete cosine domain as the readings normally do not change too much within a short period.
3) Spatial domain: as the number of paths as well as the angle-of-arrival are much smaller than the number of antennas in massive MIMO systems, the channel conditions can be represented by a limited number of parameters. In this case, the spatial domain turns into the angular domain.

For multi-node cases, due to the spatical correlation, the joint sparsity is exploited to apply DCS in spatial-x domains, where 'x' can be any of the aforementioned domains. Here, we give some examples on how the joint sparsity is utilized in different scenarios in 5G and IoT networks:

1) Spatial-frequency domain: An illustration of the DCS-enabled cooperative CRN is given in Fig. 3, where the joint sparsity in the spatial-frequency domain is utilized. Specifically, each column represents the signal received at each location, which is sparse in frequency domain as only a few channels are occupied. At different locations, the same frequency bands
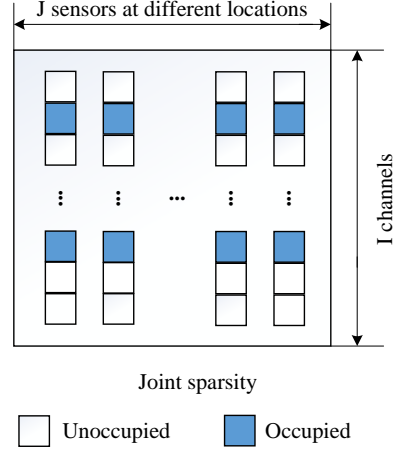

Fig. 3: Spatial-frequency correlation.

may be occupied, but the signal powers for each frequency band are various due to fading and shadowing. Therefore, different columns of the matrix share the common sparse support though each node operates without cooperation. With the DCS framework, each node performs sub-Nyquist sampling individually first and then the original signals can be recovered simultaneously. More details on this issue will be discussed in Section III.

2) Spatial-temporal domain: In WSNs, sensor nodes are deployed to periodically monitor data and to send the compressed data to the sink. Then the sink is responsible for recovering the original reading by CS algorithms as the readings across all sensor nodes exhibit both spatial and temporal correlations, as we can see from the discussion in Section IV.
3) Angular-time domain: Massive MIMO channels between some users and the massive base station (BS) antennas appear the spatial common sparsity in both the time domain and the angular domain, as we will discuss in detail in Section V.

Different sparse domains and their applications in 5G and IoT networks are summarized in Table I, where 'WT' is short for wavelet transform. In addition to the listed sparse domains and their applications, it is worth noting that the core of applying CS is to identify

how to exploit the sparse property in 5G and IoT networks. In the following, we discuss three major applications of CS in 5G and IoT networks.

III. Compressive Sensing enabled Cognitive Radio Networks

In this section, we introduce the applications of CS in CRNs. Radio frequency (RF) spectrum is a valuable but tightly regulated resource due to its unique and important role in wireless communications. The demand for RF spectrum is increasing due to a rapidly expanding market of multimedia wireless services while the usable spectrum is becoming scarce due to the current rigid spectrum allocation policies. However, according to the reports from the Federal Communications Commission (FCC) and the Office of Communications (Ofcom), localized temporal and geographic spectrum utilization is extremely low and unbalanced in reality. Cognitive radio (CR) has become a promising solution to solve the spectrum scarcity problem, by allowing unlicensed secondary users (SUs) to opportunistically access a licensed band when the licensed primary user (PU) is absent. In order to avoid any harmful interference to the PUs, SUs in CRNs should be aware of the spectrum occupancy over the spectrum of interest. Spectrum sensing, which detects the spectrum holes, the frequency bands that are not utilized by PUs, is one of the most challenging tasks in CR. As the radio environment changes over time and space, an efficient spectrum sensing technique should be capable of tracking these fast changes [26]. A good approach for detecting PUs is to adopting the traditional narrowband sensing algorithms, which include energy detection, matched-filtering, and cyclostationary feature detection. Here, the term narrowband implies that the frequency range is sufficiently narrow such that the channel frequency response can be considered as flat. In other words, the bandwidth of interest is less than the coherence bandwidth of the channel [27].

While the present spectrum sensing algorithms have focused on exploiting spectral opportunities over a narrow frequency band, future CRNs will eventually be required to exploit spectral opportunities over a wide frequency range from hundreds of megahertz (MHz) to several gigahertz (GHz), in order to improve spectrum efficiency and achieve higher opportunistic throughput. As driven by the Nyquist-Shannon sampling theory, a simple approach is to acquire the wideband signal directly by a high-speed ADC, which is particularly challenging or even unaffordable, especially for energy-constrained devices, such as smart phones or even battery-free devices. Therefore, revolutionary wideband spectrum sensing techniques become more than desired to release the burden on high-speed ADCs.

A. Standard Compressive Spectrum Sensing

Recent development on CS theory inspires sub-Nyquist sampling, by utilizing the sparse nature of signals [3]. Driven by the inborn nature of the sparsity property of signals in wireless communications, e.g., the sparse utilization of spectrum, CS theory is capable of enabling sub-Nyquist sampling for wideband spectrum sensing.

1) Energy Detection based Compressive Spectrum Sensing: CS theory has been applied to wideband spectrum sensing in [28], where sub-Nyquist sampling is achieved without loss of any information. A general framework for compressive spectrum sensing with the energy detection method is summarized as shown in Fig. 4, where the analog signal at the receiver, $r(t)$ has a sparse representation s_f in the frequency domain. The received signals are then sampled at a sub-Nyquist rate. Due to low spectrum utilization, s_f can be recovered from the under-sampled measurements. Then the energy of each channel can be calculated, and therefore the spectrum occupancy can be determined.

Lately, it has been identified in [29] that the CS-enabled system is somewhat sensitive to noise, exhibiting a 3 dB SNR loss per octave of subsampling, which parallels the classic noise-folding phenomenon. In order to improve robustness to noise, a denoised compressive spectrum sensing algorithm has been proposed

TABLE I: Common sparse domains and their applications in 5G and IoT networks.

Sparse domain	Sparsifying transform	Applications	Why sparse?	Sparsity property
Frequency domain	FT	Wideband spectrum sensing	Low spectrum utilization in practice	Single sparsity
Spatial domain	-	Channel estimation in massive MIMO	Number of paths is much fewer than the number of antennas	Single sparsity
Discrete cosine domain	DCT	Sensor data gathering	Temporal correlation	Single sparsity
Wavelet domain	WT	Sensor data gathering	Temporal correlation	Single sparsity
Spatial-frequency domain	FT	Cooperative wideband spectrum sensing	Spatial correlation and low spectrum utilization	Joint sparsity
Spatial-discrete cosine/wavelet domain	DCT/WT	Active node detection/data gathering	Spatial and temporal correlation	Joint sparsity
Angular-time domain	-	Channel estimation in massive MIMO	Number of paths and degrees of arrival are much fewer than the number of antennas	Joint sparsity

in [6]. The sparsity level is required in advance in order to determine the lower sampling rate at SUs locally without loss of any information. However, sparsity level is dependent on spectrum occupancy, which is usually unavailable in dynamic CR networks. In order to solve this problem, a have proposed a two-step CS scheme has been proposed in [30] to minimize the sampling rates when the sparsity level is changing. In this approach, the actual sparsity level is estimated first and the number of compressed measurements to be collected is then adjusted before sampling. However, this algorithm introduces extra computational complexity by performing the sparsity level estimation. In order to avoid sparsity level estimation, Sun *et al.* [31] have proposed to adjust the number of compressed measurements adaptively by acquiring compressed measurements step by step in continuous sensing slots. However, this iterative process incurs higher computational complexities at the SU as (4) has to be solved several times until the exact signal recovery is achieved. A low-complexity compressive spectrum sensing algorithm has been proposed in [5] by alleviating the iterative process of signal recovery. More specifically, geolocation data can provide a rough estimation of the sparsity level to minimize the sampling rates. Subsequently, data from geolocation database is utilized as the prior

information for signal recovery. By doing so, signal recovery performance is improved with significant reduction on the computational complexity and minimal number of measurements.

2) Compressive Power Spectral Density Estimation: Different from the aforementioned approaches that concentrate on spectral estimation with perfect reconstruction of the original signals, compressive power spectral density (PSD) estimation provides another approach for spectrum detection without requiring complete recovery of the original signals. Compressive PSD estimation has been widely applied as the original signals are not actually required in many signal processing applications. For wideband spectrum sensing applications with spectrally sparse signals, only the PSD, or equivalently the autocorrelation function, needs to be recovered as only the spectrum occupancy status is required for each channel.

Polo *et al.* [32] have proposed to reconstruct the autocorrelation of the compressed signal to provide an estimate of the signal spectrum by utilizing the sparsity property of the edge spectrum, in which the CS is performed directly on the wide-band analog signal. Nevertheless, the compressive measurements are assumed to be wide-sense stationary in [32], which is not true for some compressive measurement matrices. Subsequently, Lexa *et al.* [33] have proposed a multicoset sampling based power spectrum
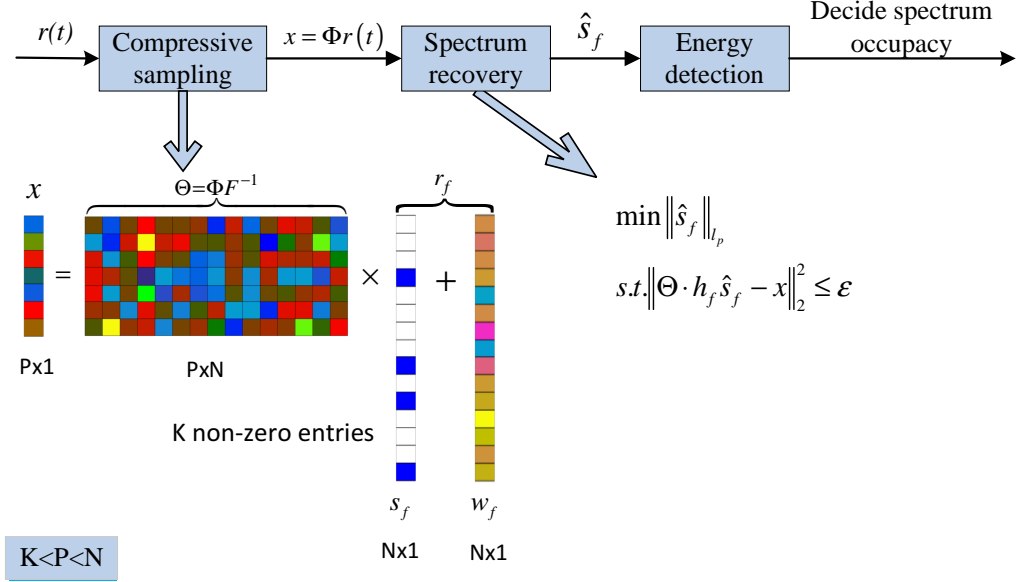
Fig. 4: Framework of compressive spectrum sensing with energy detection.

estimation method, by exploiting the fact that a wide-sense stationary signal corresponds to a diagonal covariance matrix of the frequency domain representation of the signal. Additionally, Leus *et al.* [34] have solved the power spectrum blind sampling problem based on a periodic sampling procedure and have further proposed a simple least-square (LS) reconstruction method for power spectrum recovery.

3) Beyond Sparsity: For spectrum blind sampling, the goal is to perfectly reconstruct the spectrum and sub-Nyquist rate sampling is only possible if the spectrum is sparse. However, sub-Nyquist rate sampling can be achieved in [34] without making any constraints on the power spectrum, but the LS reconstruction requires some rank conditions to be satisfied. Leus *et al.* [35] have further proposed an efficient power spectrum reconstruction and a novel multicoset sampling implementation by exploiting the spectral correlation properties, without requiring any sparsity constraints on the power spectrum. More recently, Cohen and Eldar [36] have developed a compressive power spectrum estimation framework for both sparse and non-sparse signals as well as blind and non-blind detection in the sparse case. For each one of those scenarios, the minimal sampling rate

allowing perfect reconstruction of the signal's power spectrum is derived in a noise-free environment.

B. Cooperative Spectrum Sensing with Joint Sparsity

In spectrum sensing, the performance is degraded by noise uncertainty, deep fading, shadowing and hidden nodes. Cooperative spectrum sensing (CSS) has been proposed to improve sensing performance by exploiting the collaboration among all the participating nodes. In CRNs, a CSS network constructs a multi-node network. As aforementioned, joint sparsity property and low-rank property can be utilized to recover the original signals simultaneously with fewer measurements and the DCS is utilized as it fits the CSS model perfectly. In the existing literature, cooperative compressive spectrum sensing mainly include two categories: i) centralized approaches; ii) decentralized approaches.

A centralized approach involves a fusion center (FC) performing signal recovery by utilizing the compressed measurements contributed by the spatially distributed SUs. In [6], a robust wideband spectrum sensing algorithm has been proposed for centralized CSS. Specifically, each

SU senses a segment of the spectrum at sub-Nyquist rate to reduce the sensing burden. With the collected compressive measurements, CS recovery algorithms are performed at the FC in order to recover the original signals by exploiting the sparse nature of spectrum occupancy. It is worth noting that the sparse property of signals received at the SUs can be transformed into the low-rank property of the matrix constructed at the FC. In the case of CSS with sub-Nyquist sampling, the measurements collected by the participating SUs are sent to the FC, where the joint sparsity or low-rank property is exploited to recover the complete matrix.

A typical decentralized approach has been proposed in [37], in which the decentralized consensus optimization algorithm can attain high sensing performance at a reasonable computational cost and power overhead by utilizing the joint sparsity property. Different from the centralized approach, signal recovery or matrix completion is performed at each individual node in the decentralized approaches. Compared with the centralized approaches, the decentralized ones are more robust as it adopts a FC-free network structure. Another advantage of the decentralized approaches is that they allow the recovery of individual sparse components at each node as well as the common sparse components shared by all participating nodes.

Furthermore, the privacy and security issues in CSS have been investigated in [38] by exploiting joint sparsity in the frequency and spatial domains. In [38], measurements corrupted by malicious users are removed during the signal recovery process at the FC so that the recovery accuracy and security of the considered networks can be improved.

C. Compressive Spectrum Sensing with Prior Information

In conventional compressive spectrum sensing, only the sparsity property is utilized. Certain prior information is available in some scenarios and can be exploited to improve performance of wideband spectrum sensing in CRNs.

For example, in the case of spectrum sensing over TV white space (TVWS), where the PUs are TV signals and the transmitted waveforms are determined by the standard, this prior information together with the specifications dictated by the spectrum regulatory bodies, i.e., carrier frequencies, bandwidths, can be also utilized to enhance the signal recovery performance. Thus, it is reasonable to assume that the PSD of the individual transmission is known up to a scaling factor.

As discussed in II-B, reweighted CS normally introduces weights to provide different penalties on large and small coefficients, which naturally inspires the application of reweighted CS in wideband spectrum sensing with available prior information. In [39], the whole spectrum is divided into different segments as the bounds between different types of primary radios are known in advance. Within each segment, an iteratively reweighted ℓ_1/ℓ_2 formulation has been proposed to recover the original signals. In [5], a low-complexity wideband spectrum sensing algorithm for the TVWS spectrum has been proposed to improve the signal recovery performance, in which the weights are constructed by utilizing the prior information from the geolocation database. For example, in the TVWS spectrum, there are 40 TV channels and each channel spans over 8 MHz that can be either occupied or not. Hence, the TV signals show a group sparsity property in the frequency domain as the non-zero coefficients show up in clusters. A more efficient approach has been developed in [40] by utilizing such group sparsity property. Moreover, the signals in wideband spectrum sensing have the following two characteristics: i) the input signals are stationary so that their covariance matrices are redundant; ii) most information in practical signals is concentrated on the first few lags of the autocorrelation. Inspired by these characteristics, a spectral prior information assisted structured covariance estimation algorithm has been proposed in [41] with low-computational complexity, which especially fits in application on low-end devices.

D. Potential Research

We have reviewed some research results in CS-enabled CRNs. There are still many open research issues in the area, especially when practical constraints are considered. In this section, we will introduce a couple of important ones.

1) Performance Limitations under Practical Constraints: Although there exist many research contributions in the field of compressive spectrum sensing, most of them have assumed some ideal operating conditions. In practice, there may exist various imperfection, such as noise uncertainty, channel uncertainty, dynamic spectrum occupancy, and transceiver hardware imperfection [11]. For example, the centralized compressive spectrum sensing normally considers ideal reporting channels, which is not the case in practice. This imperfection may lead to significant performance degradation in practice. Another example comes from the measurement matrix design. As shown in Fig. 2, the Gaussian distributed matrix achieves better performance but with a higher implementation cost. Even through some structured measurement matrices, such as random demodulator, with a lower cost and acceptable recovery performance degradation, have been proposed to enable the implementation of CS as a replacement of high-speed ADCs, the nonlinear recovery process limits its implementation. Therefore, it is a big challenge to further investigate compressive spectrum sensing in the presence of practical imperfection and to develop a common framework to combat their aggregate effects in CS-enabled CRNs.

2) Generalized Platform for Compressive Spectrum Sensing: The existing hardware implementation of sub-Nyquits sampling system follows the procedure that the theoretic algorithm is specifically designed for the current available hardware devices. However, it is very difficult or sometimes even impossible to extend the current hardware architectures to implement other existing compressive spectrum sensing algorithms. Thus, it is desired to have a generalized hardware platform, which can be easily adjusted to implement different compres-sive spectrum sensing algorithms with different types of measurement matrices and recovery algorithms.

IV. COMPRESSIVE SENSING ENABLED LARGE-SCALE WIRELESS SENSOR NETWORKS

WSNs provide the ability to monitor diverse physical characteristics of the real world, such as sound, temperature, and humidity, by incorporating information and communication technologies (ICT), which are especially important to various IoT applications. In the typical setup of WSNs, a large number of inexpensive and maybe individually unreliable sensor nodes with limited energy storage and low computational capability are distributed in the smart environment to perform a variety of data processing tasks, such as sensing, data collection, classification, modeling, and tracking. Cyber-physical systems (CPSs) merge wireless communication technologies and environment dynamics for efficient data acquisition and smart environments control. Typically, a CPS consists of a large number of sensor nodes and actuator nodes, which monitor and control a physical plant, respectively, by transmitting data to an elaboration node, named local controller (LC) or FC.

Traditional environment information monitoring approaches take sensing samples at a pre-defined speed uniformly at power-constrained sensor nodes and then report the data to a LC/FC, which is normally powerful and is capable of handling complex computations. The data transmitted to the LC/FC usually have redundancies, which can be exploited to reduce power consumption for data transmission. A common and efficient method is to compress at each individual sensor node and then transmit. However, data compression introduces additional power consumption for individual sensor node although the power consumption on data transmission is reduced. Furthermore, this approach is unsuitable for real-time applications owing to the high latency in data collection and the high computational complexity to ex-

ecute a compression algorithm at the power-constrained sensor node.

It is noted that most natural signals can be transformed to a sparse domain, such as discrete cosine domain or wavelet domain, where a small number of the coefficients can represent most of the power of the signals that used to be represented by a large number of samples in their original domains. In fact, the data collected at each sensor node show a sparsity property in the discrete cosine domain or the wavelet domain due to the temporal correlation. Inspired by this, CS can be applied at each sensor node to collect the compressed measurements directly and then send to the LC/FC or the neighbour nodes. As a result, fewer measurements are sampled and transmitted and the corresponding power consumption is reduced significantly.

Power consumption of sensor nodes mainly comes from sensing, data processing, and communications with the LC/FC. In a large-scale WSN, the sensor nodes with low power level would wish to take samples at a lower speed or even turn themselves into the sleep mode in order to extend lifetime. As the signal received at each sensor node shows temporal correlation and neighboring nodes show spatial correlation, the joint sparsity can be exploited to recover signals from all sensor nodes even though samples from part of the participating sensor nodes are missed. The active sensor nodes can be pre-selected according to their power levels, therefore, with invoking of CS in WSNs, lifetime of sensor nodes and the whole network can be extended.

The existing work on CS-enabled WSNs mainly falls into the aforementioned two categories: data gathering and active node selection, which will be introduced subsequently.

A. Data Gathering

In the traditional setting up of data gathering, there are a large number of sensor nodes deployed in a WSN to collect monitoring data. Each sensor node generates a reading periodically and then sends it to the LC/FC, which is normally powerful and capable of conducting complex computation. As the sensor nodes are generally limited in computation and energy storage, the data gathering process in WSNs should be energy efficient with low overhead, which becomes very challenging in IoT scenarios where a huge number of sensor nodes are deployed. Taking the temperature monitoring as an example, the adjacent sensors will generate the similar readings as the temperatures of nearby locations are similar. Furthermore, for each sensor node, the readings from adjacent snapshots are close to each other. The above two important observations indicate the temporal-spatial correlations among temperature readings, which enables the application of CS to reduce the network overhead as well as to extend the network lifetime. Moreover, such a joint sparsity is smaller than the aggregate over the individual signal sparsity, which results in a further reduction in the number of required measurements to exactly recover the original signals.

Instead of applying compression on the data after it is sampled and buffered, each sensor node collects the compressed measurements directly by projecting the signal to its sparse domain. At each individual sensor node, one can naively obtain separate measurements of its signal and then recover the signal for each sensor separately at the LC/FC by utilizing the intra-signal correlation. Moreover, it is also possible to obtain compressed measurements that each of them is a combination of all signals from the cooperative sensor nodes in a WSN. Subsequently, signals can be recovered simultaneously by exploiting both the inter-signal and intra-signal correlations at the LC/FC.

1) Measurement Matrix Design: When adopting CS techniques for data gathering in WSNs, sampling at uniformly distributed random moments satisfies the RIP if the sparse basis Ψ is orthogonal. For an arbitrary sensor node i, the $P \times N$ measurement matrix can be a spike one that only has P number of non-zero

items, as shown by

$$\Phi_i = \begin{bmatrix} 0 & 1 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & 1 & \dots & 0 \\ & & & & \ddots & \\ 0 & 0 & 0 & 0 & \dots & 1 \end{bmatrix}. \quad (8)$$

Sensor node will take a sample at the moment when the corresponding item in Φ_i is one.

However, random sampling is not proper for practical WSNs since two samples may be too close to each other, which becomes very challenging for the cheap sensor nodes. In order to solve this issue, Chen and Wassell [7] have proposed a random sampling scheme by utilizing the temporal correlation of signals received at a sensor node. In the proposed scheme, the sensor node will send a pseudorandom generator seed to the FC and then send out the samples that are obtained at an affordable highest rate until a sampling rate indicator (SRI) is received from the FC. Here, the SRI is decided based on the recovery accuracy calculated at the FC. Once the recovery accuracy goes out the required range, the sensor node will gradually increase its sampling rate until the recovery error becomes acceptable. By adopting such a scheme, the sensor node can adjust its sampling rate adaptively without the knowledge of the sparsity level. In order to further reduce the sampling rate at sensor nodes, spatial correlation is exploited in combination with temporal correlation. Therefore, joint sparsity property can be exploited at the FC to reduce the number of required measurements.

More recently, a big data enabled WSNs framework has been proposed in [42], which invokes CS for data completion with a minimal number of samples. The proposed data collection framework consists of two core components: i) at the cloud, an online learning component predicts the minimal amount of data to be collected to reduce the amount of data for transmission, and these data are considered as the principal data and their amount is constrained by CS; ii) at each individual node, a local control component tunes the collection strategy according to the dynamics and unexpected environment variation. Combining these two components, this framework can reduce power consumption and guarantee data quality simultaneously.

2) Abnormal Sensor Detection: In the CS-enabled data gathering processing, abnormal sensor readings may still lead to severe degradation on signal recovery at the LC/FC even if CS shows robustness to abnormal sensor readings as it does not rely on the statistical distribution of data to be preserved during runtime. This is because that the abnormal readings will damage the sparsity property of signals, as shown in Fig. 5. Therefore, it is critical to find out those abnormal sensors to guarantee the security of WSNs and make it abnormality-free.

Generally, abnormal readings are caused by either internal errors or external events according to their specific patterns. Abnormal readings due to internal errors fail to represent the sensed physical data, thus they should be removed at least. But the abnormal readings caused by the external events should be preserved as they reflect the actual scenarios of WSNs.

Inspired by recovering data from overcomplete dictionary, an abnormal detection mechanism has been proposed to enhance the compressibility of the signals. First, the abnormal values are detected with the help of recovering signal from an over-complete dictionary. Second, the failing sensor nodes are categorized into different types according to their patterns. Thirdly, the failing nodes caused by the internal errors are removed and then the data recovery is carried out to obtain the ordinal data.

B. Active Node Selection

In large-scale WSNs, the events are relatively sparse in comparison with the number of sensor nodes. Due to the power constraint, it is unnecessary to activate all sensor nodes at all the time. By utilizing the sparsity property constructed by the spatial correlation, the number of active sensor nodes in each time slot can be significantly reduced without scarifying performance. Taking the smart monitoring system as an example as shown in Fig. 6, the number of source nodes is N, and there are K

(a) Original signal

(b) Discrete cosine transform of original signal

(c) Abnormal signal
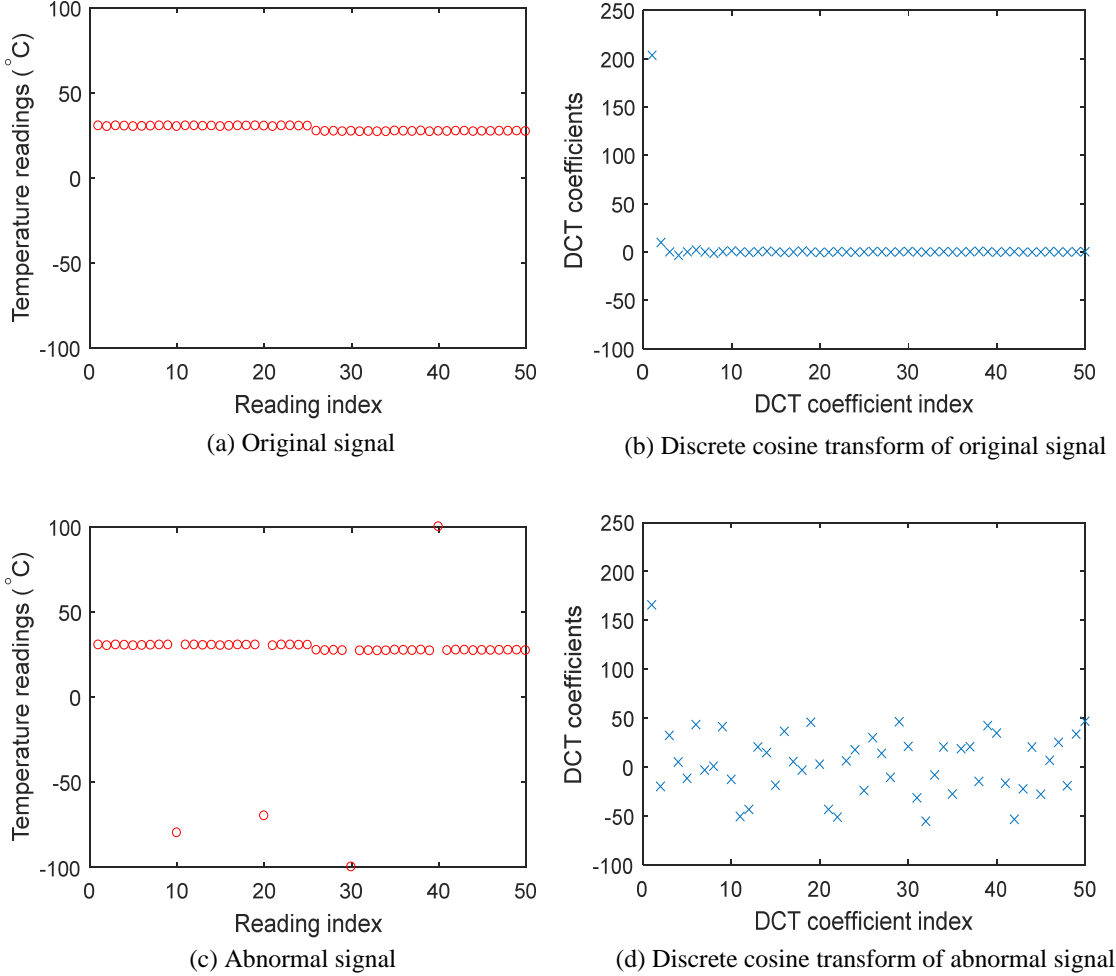
(d) Discrete cosine transform of abnormal signal

Fig. 5: Effect of abnormal readings on the sparsity level of temperature readings in discrete cosine domain.

$(K \ll N)$ sparse events that are generated by the N source nodes. By invoking CS, only M $(M \leq N)$ active sensor nodes are required to capture the K sparse events.

1) Centralized Node Selection: A centralized node selection approach has been proposed in [43] by applying CS and matrix completion at the LC/FC with the purpose of optimizing the network throughput and extending the lifetime of sensors. As a node is either active or sleeping, the state index for a node becomes binary, i.e., $X \in \{0, 1\}$. While conventional node selection in WSNs only exploits the spatial correlation of sensor nodes, Chen and Wassell [44] have exploited the temporal correlation by using the support of the data reconstructed in the previous recovery period to select the

active nodes. Specifically, the FC performs an optimized node selection, which is formulated as the design of a specialized measurement matrix, where the sensing matrix, Φ, consists of selected rows of an identity matrix as shown in (8).

The sensing costs of taking samples from different sensor nodes are assumed to be equal in most of the node selection approaches. However, in WSNs with power-constrained sensor nodes, this assumption does not hold due to the different physical conditions at different sensor nodes. For example, it is preferred to activate sensor nodes with adequate energy rather than those almost running out of energy to extend the lifetime of WSNs. Therefore, a cost-aware node selection approach has been
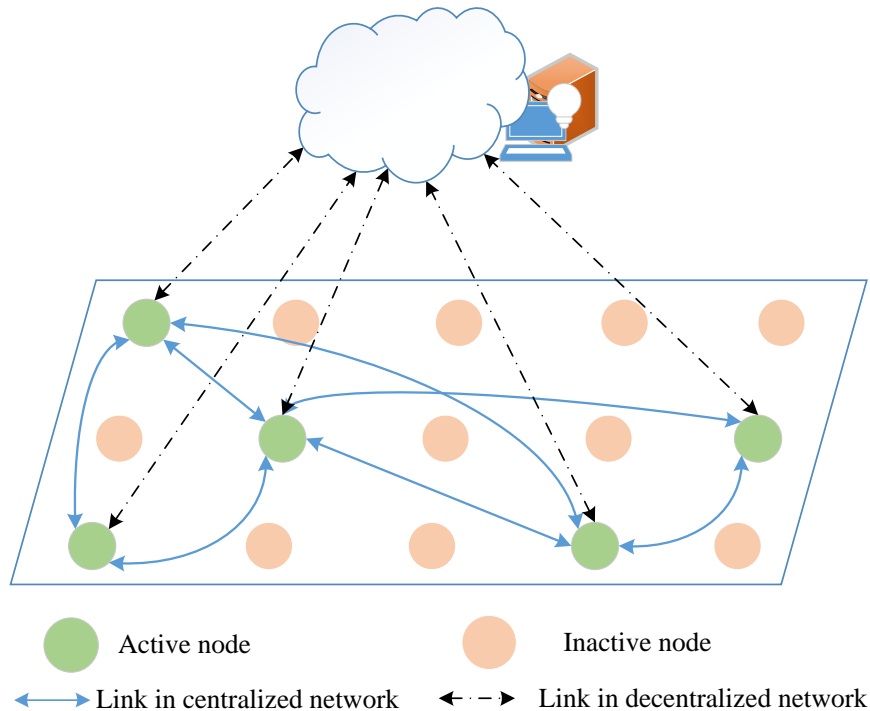
Fig. 6: Node selection in compressive sensing based wireless sensor networks.

proposed in [45] to minimize the sampling cost of the whole WSNs with constraints on the reconstruction accuracy.

2) Decentralized Node Selection: Different from the DCS, which normally conducts signal recovery at the FC by utilizing the data collected from the distributed sensing sources via exploiting the joint sparsity, decentralized CS-enabled WSNs approach aims to achieve in-network processing for node selection. A decentralized approach has been proposed in [46] to perform node selection by allowing each active sensor node to monitor and recover its local data only, by collaborating with its neighboring active sensor nodes through one-hop communication, and iteratively improving the local estimation until reaching the global optimum. It should be noted that an active sensor node not only optimizes for itself, but also for its inactive neighbors. Moreover, in order to extend the network lifetime, an in-network CS framework has been developed in [47] by enabling each sensor node to make an autonomous decision on the data compression and forward strategy with the purpose of minimizing the number of packets to be transmitted.

Generally speaking, the drawbacks for distributed node selection approach come from the following two aspects: i) the optimized node selection requires an iterative process, which may require a long period; ii) the flexibility to vary the number of active sensor nodes is limited, especially according to the dynamic sparsity levels or the channel conditions, which could be time-varying. While for the centralized approach, extra bandwidth resource and power consumption are required to coordinate active sensor nodes.

C. Potential Research

Even though extensive research has been carried out to investigate the application of CS in WSNs, most of them have focused on reducing power consumption at sensor nodes and extending the network lifetime. However, in large-scale WSNs for different IoT applications, big data should be exploited to enhance CS recovery accuracy in addition to further reduce the power consumption.

1) Machine Learning aided Adaptive Measurement Matrix Design: The core concept for sparse representation is the same even though different applications bring different constraints. Therefore, it is straightforward to ask if there is a general framework for the sparse representation of the data in different 5G and IoT applications for urban scenarios. In order to take the minimal number of samples from the set of sensor nodes with the best capability, i.e., highest power levels, the measurement matrix should be properly designed. It has been demonstrated that machine learning can be an efficient tool to aid the measurement matrix design so that the lifetime of the whole network as well each individual sensor node can be extended to the most. Furthermore, it should be noted that when designing a measurement matrix, the possibility of implementation in real network is one of the most critical factors to be considered. We believe that extensive research work in this direction is highly desired.

2) Data Privacy in CS-enabled IoT Networks: The sensing data collected from a variety of featured sensors in IoT networks, our daily activities, surroundings, and even the physical information, can be recorded and analyzed, which at the same time greatly intensifies the risk of privacy exposure. There are few effective privacy-preserving mechanisms in mobile sensing systems. In order to provide privacy protection, privacy-preserving noise has been proposed to be added to the original data to guarantee effective privacy. The added noise will dominate the data when the original sparse data points are zero or near zero, thus reducing the data sparsity. However, CS aims to achieve kind of efficiency for sparse data processing. Then it comes into a conflict countable between privacy and efficiency during big data processing in CS-enabled IoT networks. Therefore, extensive research work is expected in this area.

V. Channel Acquisition and Precoding in Massive MIMO

To satisfy high data rate requirements caused by increasing mobile applications, a lot of efforts have been made to improve transmission SE. An effective way is to exploit the spatial degrees of freedom (DoF) provided by large-scale antennas at the transmitter and the receiver to form massive MIMO systems [48]. It has been shown in [48] that the spatial resolution of a large-scale antenna array will be very high and the channels corresponding to different users are approximately orthogonal when the number of the antennas at the BS are very large. Consequently, linear processing is good enough to make the system performance to approach optimum if the CSI is known at the BS.

Accurate CSI at the BS is essential for massive MIMO to obtain the above advantage. Due to the large channel dimension, downlink CSI acquisition in massive MIMO systems sometimes becomes challenging even if uplink CSI estimation is relatively simple. In time-division duplexing (TDD) systems, the downlink CSI can be easily obtained by exploiting channel reciprocity. However, most of the practical deployed systems mainly employ frequency-division duplex (FDD), where the channel reciprocity does not hold any more. In this situation, the downlink channel has to be estimated directly and then fed back to the BS, which will result in the extremely high overhead.

To address CSI estimation and feedback issue in FDD systems, sparsity of massive MIMO channels must be exploited. Some CS-enabled CSI acquisition methods for FDD massive MIMO have been proposed, where the correlation in massive MIMO channels has been successfully exploited to reduce the amount of training symbols and feedback overhead.

In this section, we focus on the CS-enabled channel acquisition and its related applications. We will first introduce the channel sparsity feature and then discuss channel estimation and feedback, and precoding and detection subsequently. It should be noted that mmWave communications are often used with massive MIMO techniques since short wavelength here makes it very easy to pack a large number of antennas in a small area. The channel acquisition and precoding based on CS in mmWave

massive MIMO will be also included in the following discussion even if mmWave channels are slightly different from the traditional wireless channels.

A. Sparsity of Channels

In the channel acquisition and precoding schemes based on CS, the key idea is to use the channel sparsity. Although the channel sparsity in massive MIMO generally exists in the time domain, the frequency domain, and the spatial domain, we mainly focus on the spatial domain channel sparsity in this section.

In conventional MIMO systems, a rich-scattering multipath channel model is often assumed so that the channel coefficients can be modelled as independent random variables. However, this assumption is not true any more in massive MIMO systems. It has been shown that the massive MIMO channel is spatial correlated and has a sparse structure in the spatial domain. This correlation and sparsity is due to two reasons: the exploitation of high radio frequency and the deployment of large-scale antenna arrays in future wireless communications. In high frequency band, the channels have fewer propagation paths while more transmit and receive antennas will make the distinguished paths to be much fewer than the number of channel coefficients, which makes the rich scatters to become limited or sparse.

As shown in Fig. 7, a classical channel model with limited scatterers at the BS is often used in the literature [49]. In this model, different user channels have a partially common sparsity support due to the shared scatterers and an independent sparsity support caused by the individual scatterers in the propagation environments. By using this sparsity structure, many CS-enabled channel acquisition and precoding schemes have been proposed, as we will discuss in the following.

B. Compressive Pilot Design

In order to obtain a good channel estimation, the length of the orthogonal training sequence must be at least same as the number of transmit antenna elements. Due to a huge number of antennas at the BS in massive MIMO systems, the downlink pilots occupy a high proportion of the resources. Consequently, the traditional pilot design is not applicable here. It is necessary to design specific pilots to reduce the training overhead in FDD massive MIMO systems. It has been shown that channel spatial correlation or sparsity can be used to shrink the original channel to an effective one with a much lower dimension so that low-overhead training is enough in massive MIMO systems. Based on this principle, CS-enabled pilot design schemes have been proposed in [50].

When exploiting the CS theory to acquire the CSI in the correlated FDD MIMO systems, how much training should be sent is a very important question. Once the amount of the training is determined, the training symbols can be designed by using of the channel common or/and individual sparsity in the correlated massive MIMO systems. Besides using the common support, a new pilot structure with joint common and dedicated support has been proposed in [50], where the common one is used to estimate the common channel parameters of the related users in multi-user massive MIMO systems while the dedicated one is used to estimate the individual channel parameters of each user.

Besides the general massive MIMO systems, a compressed hierarchical multiresolution codebook has also been designed specially for mmWave massive MIMO systems to construct training beamforming vectors [51]. Based on this idea, a lot of different hierarchical multiresolution codebooks have been proposed from different angles. For example, in [52], a compressive beacon codebook with different sets of pseudorandom phases has been designed. In [53], a common codebook satisfying the conflicting design requirements as well as validating practical mmWave systems has been proposed through utilizing the strong directivity of mmWave channels. In [54], a multi-resolution uniform-weighting based codebook, with similar to an normalized DFT matrix, has been proposed to reduce the implementation complexity, where the estimation of the angle
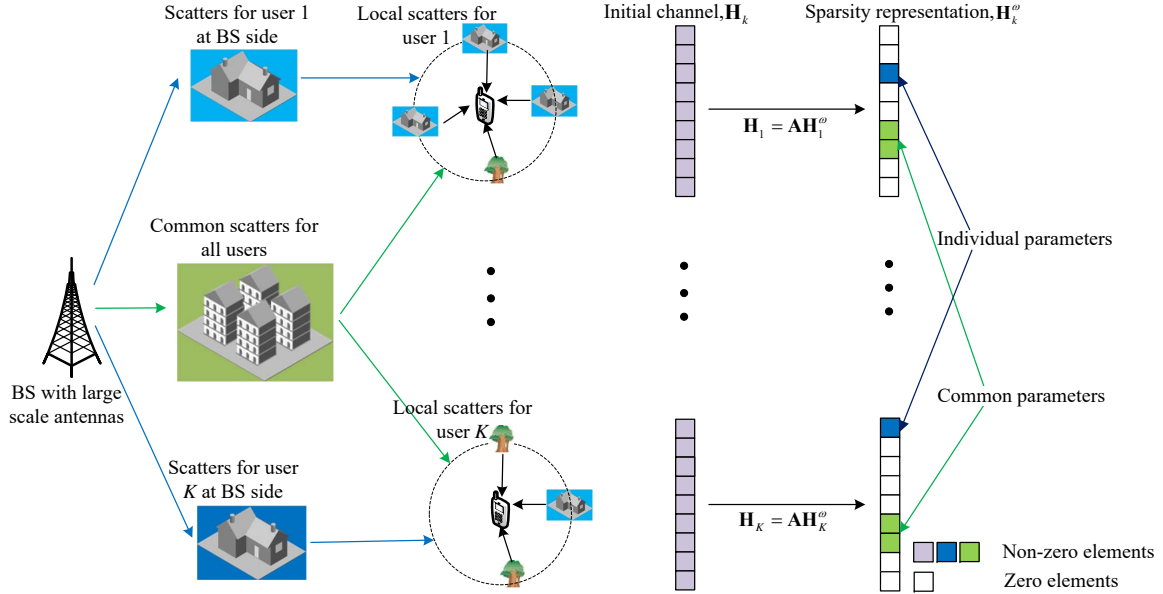
Fig. 7: Channel model with limited scatters, where $\mathbf{A}$ is the spatial correlation matrix.

of arrival (AoA) and angle of departure (AoD) will have a much lower training overhead. Therefore, the codebook based beamforming training procedure can achieve a good balance between complexity and high performance for the practical systems.

C. Compressive Channel Estimation and Feedback

CS can be used to reduce the cost of the channel estimation and feedback by exploiting the channel sparsity. The existing CS based channel estimation and feedback schemes can be divided into the following three categories according to exploiting sparsity in different domains.

1) With Time Domain Sparsity: It has been shown that the channel is slowly-varying in various applications so that the prior channel estimation result can be still used to reconstruct the subsequent channels, as illustrated in Fig. 8. By using the structure in the figure, several CS algorithms have been proposed in [8], [55], [56] to recover massive MIMO channels.

Exploiting the common time-domain support, a CS-enabled differential CSI estimation and feedback scheme has been proposed in [55]. The scheme in [56] further combines

LS and CS techniques to improve estimation performance. In this scheme, the LS and CS techniques can be used to estimate a dense vector obtained by projecting into the previous support and a sparse vector obtained by projecting into the null space of the previous support, respectively. Since the current channel vector has only a small number of non-zero elements outside the support of the previous channel, the proposed scheme can reduce the pilot overhead and improve the tracking performance of the channel subspace. Since the quality of the prior support information will affect the estimation accuracy, a greedy pursuit-based approach with the prior support information and its quality information has been developed in [8], where the prior support information is adaptively exploited based on its quality information to further improve the channel estimation performance.

Besides channel estimation, several channel feedback schemes have been also proposed in [55] and [49]. To exploit the common support, a CS-enabled differential CSI feedback scheme has been developed in [55] by using the temporal correlation of MIMO channels. The proposed scheme can reduce the feedback overhead by about 20% compared with the

(a) Support from prior and current signals

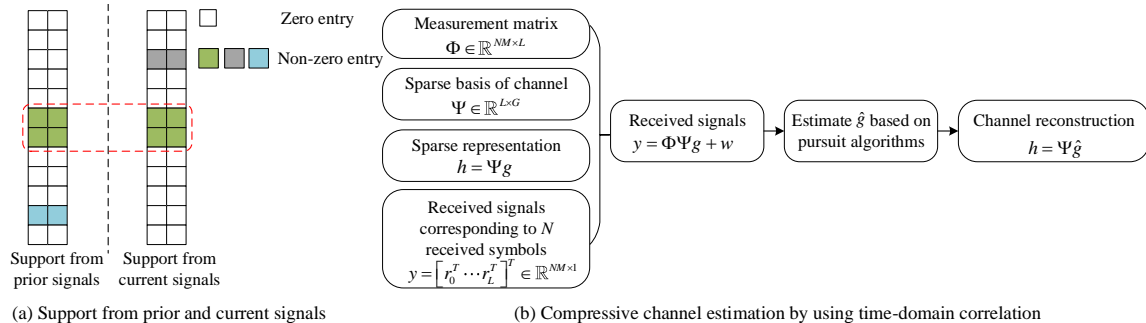(b) Compressive channel estimation by using time-domain correlation

Fig. 8: Illustration of compressive channel estimation with time-domain correlation signal support.

direct CS-enabled channel feedback. In [49], a robust closed-loop pilot and channel state information at the transmitter (CSIT) feedback resource adaptation framework has been developed by using the temporal correlation of multiuser massive MIMO channels. In this framework, the pilot and feedback resources can be adaptively adjusted for successful CSIT recovery under unknown and time-varying channel sparsity levels.

2) With Spatial Domain Sparsity: In practice, the antenna spacing in massive MIMO is usually set to be half-wave length to keep the array aperture compact. Furthermore, the BS with a large-scale antenna array is generally deployed at the top of high buildings such that there are only limited local scatters [57]. In this case, the large-scale MIMO channels exhibit strong angular-domain sparsity or spatial sparsity. This channel sparsity property can be used to reduce the channel estimation and feedback overhead in FDD massive MIMO systems.

In [58], a block optimization algorithm is employed to extract the common angular support information from the channel matrices. The extracted common support information is then used to form the weighted factors and design a weighted block optimization algorithm to estimate the channel matrix. In [59], a spatial sparsity-based compression mechanism has been proposed to reduce the load of the channel feedback. In this mechanism, random projection with unknown sparsity basis and direct compression based on known sparsity basis are used. Since the spatial sparsity will reduce

channel rank, a dictionary learning method, which captures the communication environment and the antennas property, has been proposed to obtain the compressed channel estimation.

After the compressed hierarchical multiresolution codebook is designed in [51]–[54], the corresponding channel estimation schemes can be developed. In these schemes, the hierarchical multi-resolution codebook can be capable of generating variable beam width radiation patterns to facilate the usage of robust adaptive multipath channel estimation algorithms. Meanwhile, the exploitation of adaptive compressive sensing algorithm will also reduce implementation complexity and estimation error.

3) With Spatial-Temporal Sparsity: In the above schemes, the temporal-domain sparsity and the spatial-domain one are independently exploited. In practice, they can be jointly used to further reduce the cost of channel estimation and feedback.

In [60], a structured-CS enabled differential joint channel training and feedback scheme has been proposed, where a structured compressive sampling matching pursuit (S-CoSaMP) algorithm uses the structured spatial-time sparsity of wireless MIMO channels to reduce the training and feedback overhead. In [61], a Bayesian CS based feedback mechanism has been proposed for time-varying spatially and temporally correlated vector auto-regression wireless channels, where the feedback rate can be adaptively adjusted.

4) With Spatial-Frequency Sparsity: Besides the spatial-temporal sparsity, the spatial-

frequency one can also be exploited to reduce the cost of channel estimation and feedback. In [62], the sparsity in the spatial-frequency domain is first exploited and an adaptive CS-enabled feedback scheme is correspondingly proposed to reduce the feedback overhead. In this scheme, the feedback can be dynamically configured based on the channel conditions to improve the efficiency. Due to sharing sparse common support for the adjacent subcarriers in orthogonal frequency division modulation (OFDM), an approximate message passing with the nearest neighbor sparsity pattern learning (AMP-NNSPL) algorithm has been developed to adaptively learn the underlying structure to obtain better performance.

D. Precoding and Detection

The precoder design based on the estimated CSI is a very important problem in massive MIMO systems, especially in the mmWave wideband systems. Since wideband mmWave massive MIMO channel generally exhibits frequency selective fading, the precoder design based on the estimated CSI will become challenging. Generally, the sparse structure of mmWave massive MIMO channels in angle domain or beam space can be used to simplify the precoder design [63], [64]. In [63], compressive subspace estimation is used to get the full channel information and then design the precoder to maximize the system SE. In order to reduce the CSI acquisition cost and address the mismatch between few radio frequency chains and many antennas, the hybrid precoding design with baseband and radio frequency precoders has been used in mmWave massive MIMO systems. However, this hybrid precoding will introduce performance loss. In order to mitigate performance loss, an iterative OMP has been utilized to refine the quality of hybrid precoders [64]. Meanwhile, a limited feedback technique has also been proposed for hybrid precoding to reduce the feedback cost.

CS can be also used in signal detection of massive MIMO with spatial modulation (SM). In massive SM-MIMO, the maximum likelihood (ML) detector has a prohibitively high complexity. Thanks to the structured sparsity of multiple SM signals, a low-complexity signal detector based on CS has been introduced to improve signal detection performance. In [65], a joint SM transmission scheme for user equipment and a structured CS-enabled multi-user detector for the BS have been developed, the proposed detector can reliably detect the resultant SM signals with low complexity by exploiting the intrinsical sparse features.

In the above, we mainly discuss CSI acquisition and precoding based on CS theory in FDD massive MIMO systems. In practice, the CS theory can be also applied to TDD massive MIMO systems. A channel estimation approach based on block-structured CS has been developed in [66], where the common support in sparse channels and the channel reciprocity in TDD mode are used simultaneously so that the computational complexity and pilot overhead can be reduced significantly.

E. Potential Research

As we can see from the above discussion, CS has been successfully used in massive MIMO to improve the performance of channel estimation and precoding. However, there are still many open topics before implementing CS in massive MIMO systems. We will discuss some of them in this section.

1) Effect of Antenna Deployment: Due to space limitation, large-scale antenna may be deployed at various topologies, i.e., centralized or distributed. Since the different antenna topologies corresponding to different channel sparsity, the effect of antenna configuration on the performance of CS-enabled channel estimation is still open.

2) Measure of Channel Sparsity: As mentioned before, the channel sparsity is very important to CS-enabled channel estimation. In the current research works, various sparsity models have been assumed and exploited in channel acquisition. Many of these assumed sparsity models lack of verification by measurement results. It is desired that either the sparsity model or the CS-enabled approach can

be confirmed by some measurement results under different propagation conditions.

3) Channel Estimation and Feedback with Joint Support: In the above channel estimation and feedback schemes, one or two of the three types of channel sparsity are used to improve the channel estimation performance and reduce the feedback overhead. If all three types of channel sparsity are used, can we further improve performance? How much performance gain can be achieved if possible?

4) Channel Estimation and Feedback with Unknown Support: Channel sparsity directly affects CS-enabled channel estimation performance. However, it is still a challenging problem how to get the information on channel sparsity support. At the same time, it is also worth to study how channel estimation and feedback schemes should be designed without channel sparsity support.

VI. Other Compressive Sensing Applications

In addition to the aforementioned applications, CS has been applied in various other areas in wireless communications. Some of them are introduced in this section.

A. Compressive Sensing Aided Localization

In a multiple target localization network, the multiple target locations can be formulated as a sparse matrix in the discrete spatial domain. By exploiting the spatial sparsity, instead of recording all received signal strengths (RSSs) over the spatial grid to construct a radio map from targets, only far fewer number of RSS measurements need to be collected at runtime. Subsequently, the target locations can be recovered from the collected measurements through solving an ℓ_1 minimization problem. As a result, the multiple target locations can be recovered accordingly.

B. Compressive Sensing Aided Impulse Noise Cancellation

In certain applications, such as OFDM, impulsive noise will degrade the system perfor-

mance severely. OFDM signal is often processed in the frequency domain. Even if impulse noise only lasts a short period of time, it affects a wide frequency range. By regarding impulse noise as a sparse vector, CS technique has been exploited to mitigate such type of impulse noise [67].

C. Compressive Sensing Aided Cloud Radio Access Networks

Cloud radio access networks (C-RANs) has been proposed as a promising technology to support massive connectivity in 5G networks. In C-RAN, the BSs are replaced by remote radio heads (RRHs) and connected to a central processor via digital backhaul links. Thanks to the spatial and temporal variation of the mobile traffic, it is feasible to switch off some RRHs with guarantee on the quality of service in green C-RANs. More specifically, one RRH will be switched off only when all the coefficients in its beamformer are set to zeros. Such a group sparsity property inspires us to apply CS to active RRH selection in green C-RANs to minimize the network power consumption [68], [69]. Additionally, in the uplink of C-RANs, the channel estimation from the active users to the RRHs is the key to achieve the spatial multiplexing gain. Generally, the number of active users is low in C-RANs, which makes it possible to apply CS to reduce the uplink training overhead for channel estimation. Moreover, the correlation among active users at different RRHs exhibit a joint sparsity property, which can further facilitate the active user detection and channel estimation in C-RANs [70].

VII. Concluding Remarks

This article has provided a comprehensive overview of sparse representation with applications in wireless communications. Specifically, after introducing the basic principles of CS, the common sparse domains in 5G and IoT networks have been identified. Subsequently, three CS-enabled networks, including wideband spectrum sensing in CRNs, data collection in IoT networks, and channel estimation and feedback in massive MIMO systems, have been

discussed by exploiting different sparsity properties. From the above discussions, it has been concluded that by invoking of CS, the SE and EE of 5G and IoT networks can be enhanced from different perspectives. Furthermore, potential research challenges have been identified to provide a guide for researchers interested in the sparse representation in 5G and IoT networks.

ACKNOWLEDGMENT

REFERENCES

[1] M. Zibulevsky and M. Elad, "L1-L2 optimization in signal and image processing," *IEEE Signal. Proc. Mag.*, vol. 27, no. 3, pp. 76–88, May 2010.

[2] A. M. Bruckstein, D. L. Donoho, and M. Elad, "From sparse solutions of systems of equations to sparse modeling of signals and images," *SIAM review*, vol. 51, no. 1, pp. 34–81, Feb. 2009.

[3] E. Candes, "Compressive sampling," in *Proc. Int. Congress Math.*, vol. 3, Madrid, Spain, Oct. 2006, pp. 1433–1452.

[4] E. J. Candes, 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.

[5] Z. Qin, Y. Gao, and C. G. Parini, "Data-assisted low complexity compressive spectrum sensing on real-time signals under sub-Nyquist rate," *IEEE Trans. Wireless Commun.*, vol. 15, no. 2, pp. 1174–1185, Feb. 2016.

[6] Z. Qin, Y. Gao, M. D. Plumbley, and C. G. Parini, "Wideband spectrum sensing on real-time signals at sub-Nyquist sampling rates in single and cooperative multiple nodes," *IEEE Trans. Signal Process.*, vol. 64, no. 12, pp. 3106–3117, Jun. 2016.

[7] W. Chen and I. J. Wassell, "Energy-efficient signal acquisition in wireless sensor networks: a compressive sensing framework," *IET Wireless Sensor Systems*, vol. 2, no. 1, pp. 1–8, Mar. 2012.

[8] X. Rao and V. K. Lau, "Compressive sensing with prior support quality information and application to massive MIMO channel estimation with temporal correlation," *IEEE Trans. Signal Process.*, vol. 63, no. 18, pp. 4914–4924, Sep. 2015.

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

[10] C. R. Berger, Z. Wang, J. Huang, and S. Zhou, "Application of compressive sensing to sparse channel estimation," *IEEE Commun. Mag.*, vol. 48, no. 11, pp. 164–174, Nov. 2010.

[11] S. K. Sharma, E. Lagunas, S. Chatzinotas, and B. Ottersten, "Application of compressive sensing in cognitive radio communications: A survey," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 3, pp. 1838–1860, thirdquarter 2016.

[12] D. Romero, D. D. Ariananda, Z. Tian, and G. Leus, "Compressive covariance sensing: Structure-based compressive sensing beyond sparsity," *IEEE Signal. Proc. Mag.*, vol. 33, no. 1, pp. 78–93, Jan. 2016.

[13] H. S. Chang, Y. Weiss, and W. T. Freeman, "Informative sensing of natural images," in *IEEE International Conference on Image Processing (ICIP)*, Cairo, Egypt, Nov. 2009, pp. 3025–3028.

[14] J. Wright, Y. Ma, J. Mairal, G. Sapiro, T. S. Huang, and S. Yan, "Sparse representation for computer vision and pattern recognition," *Proc. IEEE*, vol. 98, no. 6, pp. 1031–1044, Jun. 2010.

[15] H. S. Chang, Y. Weiss, and W. T. Freeman, "Informative sensing," *CoRR*, vol. abs/0901.4275, 2009. [Online]. Available: http://arxiv.org/abs/0901.4275

[16] E. Candes and J. Romberg, "Sparsity and incoherence in compressive sampling," *Inverse problems*, vol. 23, no. 3, p. 969, Apr. 2007.

[17] J. Tropp, J. Laska, M. Duarte, J. Romberg, and R. Baraniuk, "Beyond nyquist: Efficient sampling of sparse bandlimited signals," *IEEE Trans. Inf. Theory*, vol. 56, no. 1, pp. 520–544, Jan. 2010.

[18] E. J. Cands, "The restricted isometry property and its implications for compressed sensing," *Comptes Rendus Mathematique*, vol. 346, no. 9, pp. 589 – 592, 2008. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1631073X08000964

[19] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655–4666, Dec. 2007.

[20] J. W. Choi, B. Shim, Y. Ding, B. Rao, and D. I. Kim, "Compressed sensing for wireless communications: Useful tips and tricks," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 3, pp. 1527–1550, thirdquarter 2017.

[21] B. D. Rao and K. Kreutz-Delgado, "An affine scaling methodology for best basis selection," *IEEE Trans. Signal Process.*, vol. 47, no. 1, pp. 187–200, Jan. 1999.

[22] R. Chartrand, "Exact reconstruction of sparse signals via nonconvex minimization," *IEEE Signal Process. Lett.*, vol. 14, no. 10, pp. 707–710, Oct. 2007.

[23] R. Chartrand and V. Staneva, "Restricted isometry properties and nonconvex compressive sensing," *Inverse Problems*, vol. 24, no. 3, p. 035020, May. 2008.

[24] R. Chartrand and W. Yin, "Iteratively reweighted algorithms for compressive sensing," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.(ICASSP)*, Las Vegas, NV, Mar. 2008, pp. 3869–3872.

[25] D. Baron, M. F. Duarte, M. B. Wakin, S. Sarvotham, and R. G. Baraniuk, "Distributed compressive sensing," 2009.

[26] Y.-C. Liang, Y. Zeng, E. Peh, and A. T. Hoang, "Sensing-throughput tradeoff for cognitive radio networks," *IEEE Trans. Wireless Commun.*, vol. 7, no. 4, pp. 1326–1337, Apr. 2008.

[27] H. Sun, A. Nallanathan, C.-X. Wang, and Y. Chen, "Wideband spectrum sensing for cognitive radio networks: a survey," *IEEE Wireless Commun.*, vol. 20, no. 2, pp. 74–81, Mar. 2013.

[28] Z. Tian and G. Giannakis, "Compressed sensing for wideband cognitive radios," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.(ICASSP)*, Honolulu, HI, Apr. 2007, pp. 1357–1360.

[29] M. A. Davenport, J. N. Laska, J. R. Treichler, and R. G. Baraniuk, "The pros and cons of compressive sensing for wideband signal acquisition: Noise folding versus dynamic range," *IEEE Trans. Signal Process.*, vol. 60, no. 9, pp. 4628–4642, Sep. 2012.

[30] Y. Wang, Z. Tian, and C. Feng, "Sparsity order estimation and its application in compressive spectrum sensing for cognitive radios," *IEEE Trans. Wireless Commun.*, vol. 11, no. 6, pp. 2116–2125, Jun. 2012.

[31] H. Sun, W.-Y. Chiu, and A. Nallanathan, "Adaptive compressive spectrum sensing for wideband cognitive radios," *IEEE Commun. Lett.*, vol. 16, no. 11, pp. 1812–1815, Nov. 2012.

[32] Y. L. Polo, Y. Wang, A. Pandharipande, and G. Leus, "Compressive wide-band spectrum sensing," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.(ICASSP)*, Taipei, Taiwan, Apr. 2009, pp. 2337–2340.

[33] M. A. Lexa, M. E. Davies, J. S. Thompson, and J. Nikolic, "Compressive power spectral density estimation," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.(ICASSP)*, Prague, Czech Republic, May 2011, pp. 3884–3887.

[34] G. Leus and D. D. Ariananda, "Power spectrum blind sampling," *IEEE Signal Process. Lett.*, vol. 18, no. 8, pp. 443–446, Aug. 2011.

[35] D. D. Ariananda and G. Leus, "Compressive wideband power spectrum estimation," *IEEE Trans. Signal Process.*, vol. 60, no. 9, pp. 4775–4789, Sep. 2012.

[36] D. Cohen and Y. C. Eldar, "Sub-Nyquist sampling for power spectrum sensing in cognitive radios: A unified approach," *IEEE Trans. Signal Process.*, vol. 62, no. 15, pp. 3897–3910, Aug. 2014.

[37] F. Zeng, C. Li, and Z. Tian, "Distributed compressive spectrum sensing in cooperative multihop cognitive networks," *IEEE J. Sel. Areas Signal. Proc.*, vol. 5, no. 1, pp. 37 –48, Feb. 2011.

[38] Z. Qin, Y. Gao, and M. D. Plumbley, "Malicious user detection based on low-rank matrix completion in wideband spectrum sensing," *IEEE Trans. Signal Process.*, vol. 66, no. 1, pp. 5–17, Jan. 2018.

[39] Y. Liu and Q. Wan, "Enhanced compressive wideband frequency spectrum sensing for dynamic spectrum access," *EURASIP J. Adv. Signal Process.*, vol. 2012, no. 1, p. 177, Dec. 2012.

[40] Y. C. Eldar, P. Kuppinger, and H. Bolcskei, "Block-sparse signals: Uncertainty relations and efficient recovery," *IEEE Trans. Signal Process.*, vol. 58, no. 6, pp. 3042–3054, Jun. 2010.

[41] D. Romero and G. Leus, "Wideband spectrum sensing from compressed measurements using spectral prior information," *IEEE Trans. Signal Process.*, vol. 61, no. 24, pp. 6232–6246, Dec. 2013.

[42] L. Kong, D. Zhang, Z. He, Q. Xiang, J. Wan, and M. Tao, "Embracing big data with compressive sensing: A green approach in industrial wireless networks," *IEEE Commun. Mag.*, vol. 54, no. 10, pp. 53–59, Oct. 2016.

[43] Z. Qin, Y. Liu, Y. Gao, M. Elkashlan, and A. Nallanathan, "Wireless powered cognitive radio networks with compressive sensing and matrix completion," *IEEE Trans. Commun.*, vol. 65, no. 4, pp. 1464–1476, Apr. 2017.

[44] W. Chen and I. J. Wassell, "Optimized node selection for compressive sleeping wireless sensor networks," *IEEE Trans. Veh. Technol.*, vol. 65, no. 2, pp. 827–836, Feb. 2016.

[45] ——, "Cost-aware activity scheduling for compressive sleeping wireless sensor networks," *IEEE Trans. Signal Process.*, vol. 64, no. 9, pp. 2314–2323, May 2016.

[46] Q. Ling and Z. Tian, "Decentralized sparse signal recovery for compressive sleeping wireless sensor networks," *IEEE Trans. Signal Process.*, vol. 58, no. 7, pp. 3816–3827, Jul. 2010.

[47] C. Caione, D. Brunelli, and L. Benini, "Distributed compressive sampling for lifetime optimization in dense wireless sensor networks," *IEEE Trans. Ind. Informat.*, vol. 8, no. 1, pp. 30–40, Feb. 2012.

[48] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590–3600, Nov. 2010.

[49] A. Liu, F. Zhu, and V. K. Lau, "Closed-loop autonomous pilot and compressive CSIT feedback resource adaptation in multi-user FDD massive MIMO systems," *IEEE Trans. Signal Process.*, vol. 65, no. 1, pp. 173–183, Jan. 2017.

[50] V. K. Lau, S. Cai, and A. Liu, "Closed-loop compressive CSIT estimation in FDD massive MIMO systems with 1 bit feedback," *IEEE Trans. Signal Process.*, vol. 64, no. 8, pp. 2146–2155, Apr. 2016.

[51] A. Alkhateeb, O. El Ayach, G. Leus, and R. W. Heath, "Channel estimation and hybrid precoding for millimeter wave cellular systems," *IEEE J. Sel. Areas Signal. Proc.*, vol. 8, no. 5, pp. 831–846, 2014.

[52] Z. Marzi, D. Ramasamy, and U. Madhow, "Compressive channel estimation and tracking for large arrays in mm-wave picocells," *IEEE J. Sel. Areas Signal. Proc.*, vol. 10, no. 3, pp. 514–527, 2016.

[53] J. Song, J. Choi, and D. J. Love, "Common codebook millimeter wave beam design: Design-

ing beams for both sounding and communication with uniform planar arrays," *IEEE Trans. Commun.*, vol. 65, no. 4, pp. 1859–1872, 2017.

[54] G. Wang, S. A. Naeimi, and G. Ascheid, "Low complexity channel estimation based on dft for short range communication," in *Proc. IEEE Int. Conf. Commun. (ICC).* IEEE, 2017, pp. 1–7.

[55] W. Shen, L. Dai, Y. Shi, X. Zhu, and Z. Wang, "Compressive sensing-based differential channel feedback for massive MIMO," *Electr. Lett.*, vol. 51, no. 22, pp. 1824–1826, Oct. 2015.

[56] Y. Han, J. Lee, and D. J. Love, "Compressed sensing-aided downlink channel training for FDD massive MIMO systems," *IEEE Trans. Commun.*, vol. 65, no. 7, pp. 2852–2862, Jul. 2017.

[57] L. Lu, G. Li, A. Swindlehurst, A. Ashikhmin, and R. Zhang, "An overview of massive MIMO: Benefits and challenges," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 742–758, Oct. 2014.

[58] C. C. Tseng, J. Y. Wu, and T. S. Lee, "Enhanced compressive downlink CSI recovery for FDD massive MIMO systems using weighted block-minimization," *IEEE Trans. Commun.*, vol. 64, no. 3, pp. 1055–1067, Mar. 2016.

[59] M. S. Sim, J. Park, C.-B. Chae, and R. W. Heath, "Compressed channel feedback for correlated massive MIMO systems," *J. of Commun. and Netw.*, vol. 18, no. 1, pp. 95–104, Feb. 2016.

[60] W. Shen, L. Dai, Y. Shi, B. Shim, and Z. Wang, "Joint channel training and feedback for FDD massive MIMO systems," *IEEE Trans. Veh. Technol.*, vol. 65, no. 10, pp. 8762–8767, Oct. 2016.

[61] X.-L. Huang, J. Wu, Y. Wen, F. Hu, Y. Wang, and T. Jiang, "Rate-adaptive feedback with Bayesian compressive sensing in multiuser MIMO beamforming systems," *IEEE Trans. on Wireless Commun.*, vol. 15, no. 7, pp. 4839–4851, Jul. 2016.

[62] P.-H. Kuo, H. Kung, and P.-A. Ting, "Compressive sensing based channel feedback protocols for spatially-correlated massive antenna arrays," in *Proc. Wireless Commun. Netw. Conf. (WCNC)*, Paris, France, Apr. 2012, pp. 492–497.

[63] K. Venugopal, N. Gonzlez-Prelcic, and R. W. Heath, "Optimality of frequency flat precoding in frequency selective millimeter wave channels," *IEEE Wireless Commun. Lett.*, vol. 6, no. 3, pp. 330–333, Jun. 2017.

[64] J. Mirza, B. Ali, S. S. Naqvi, and S. Saleem, "Hybrid precoding via successive refinement for millimeter wave MIMO communication systems," *IEEE Commun. Lett.*, vol. 21, no. 5, May 2017.

[65] Z. Gao, L. Dai, Z. Wang, S. Chen, and L. Hanzo, "Compressive-sensing-based multiuser detector for the large-scale SM-MIMO uplink," *IEEE Trans. Veh. Technol.*, vol. 65, no. 10, pp. 8725–8730, Oct. 2016.

[66] Y. Nan, L. Zhang, and X. Sun, "Efficient downlink channel estimation scheme based on block-structured compressive sensing for tdd massive MU-MIMO systems," *IEEE Wireless Commun. Lett.*, vol. 4, no. 4, pp. 345–348, Aug. 2015.

[67] A. B. Ramirez, R. E. Carrillo, G. Arce, K. E. Barner, and B. Sadler, "An overview of robust compressive sensing of sparse signals in impulsive noise," in *European Signal Process. Conf. (EUSIPCO)*, Aug. 2015, pp. 2859–2863.

[68] Y. Shi, J. Zhang, and K. B. Letaief, "Group sparse beamforming for green cloud-RAN," vol. 13, no. 5, pp. 2809–2823, May 2014.

[69] B. Dai and W. Yu, "Sparse beamforming and user-centric clustering for downlink cloud radio access network," *IEEE Access*, vol. 2, pp. 1326–1339, Oct. 2014.

[70] X. Xu, X. Rao, and V. K. N. Lau, "Active user detection and channel estimation in uplink cran systems," in *Proc. IEEE Int. Conf. Commun. (ICC)*, London, UK, Jun. 2015, pp. 2727–2732.