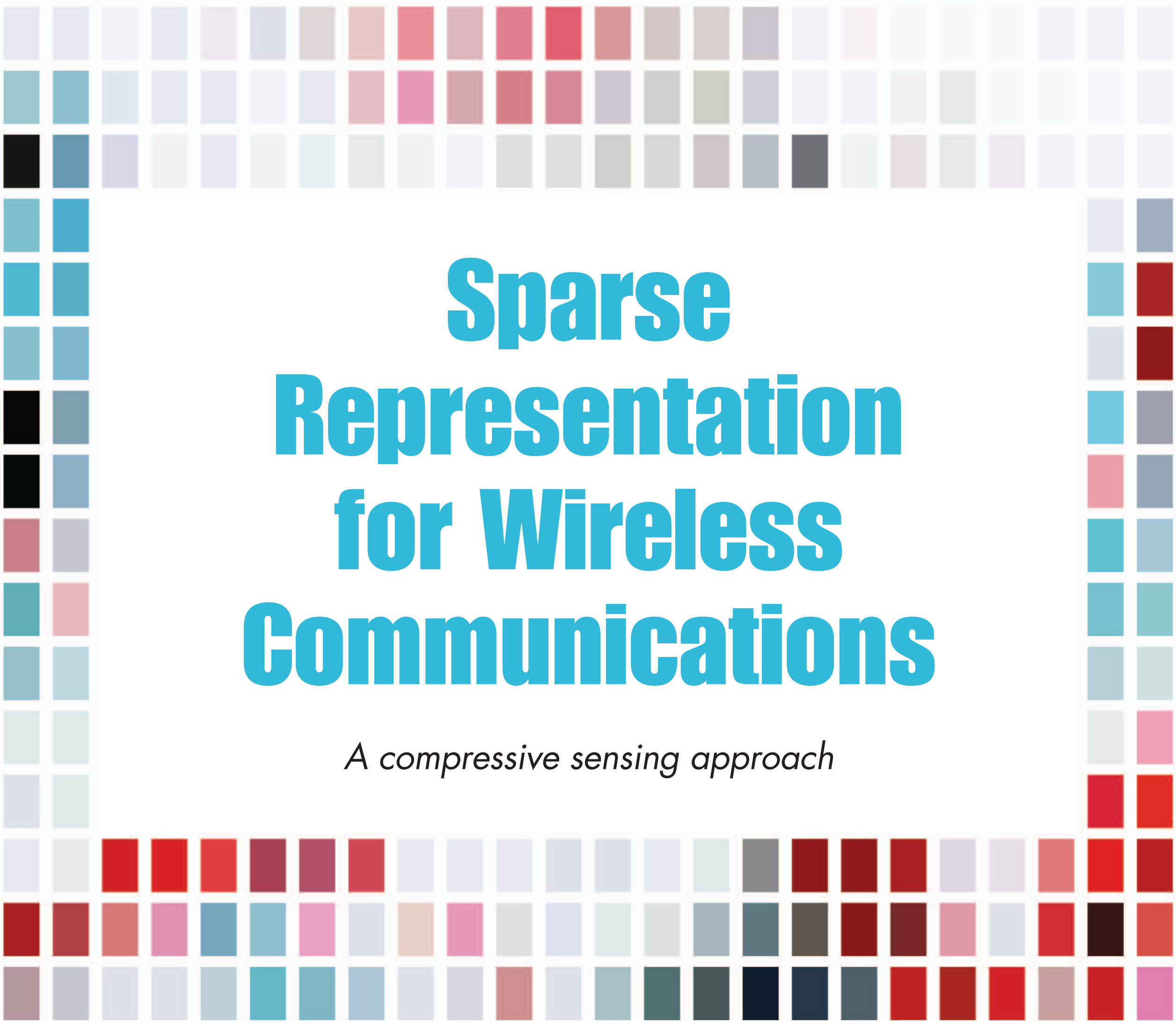
Zhijin Qin, Jiancun Fan, Yuanwei Liu, Yue Gao, and Geoffrey Ye Li



Sparse Representation for Wireless Communications

A compressive sensing approach

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 a 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 (SE) and energy efficiency (EE) of fifth-generation (5G) 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 we introduce recent research progress in applying CS to address the major opportunities and challenges in these networks, including wide-band spectrum sensing in cognitive radio networks (CRNs), data collection in IoT networks, and channel estimation and feedback in massive multiple-input, multiple-output (MIMO) systems. Moreover, we identify other potential applications and research challenges for sparse representation for 5G and IoT networks. This article will provide readers a clear picture of

Digital Object Identifier 10.1109/MSP.2018.2789521
Date of publication: 26 April 2018

how to exploit sparsity properties to process wireless signals in different applications.

Introduction

Sparse representation expresses some signals as a linear combination of a few atoms from a prespecified and overcomplete dictionary [1]. This form of sparse (or compressible) structure arises naturally in many applications [2]. For example, audio signals are sparse in the frequency domain, especially for the sounds representing tones. Image processing can exploit a sparsity property in the discrete cosine domain, i.e., many images' discrete cosine transform (DCT) coefficients 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, wide-band sensing in 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 5G and IoT networks, the amount of generated data becomes huge. Therefore, signal processing has been confronted with challenges regarding high sampling rates for data acquisition and large amounts 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 achieving signal sampling and compression simultaneously.

As an appealing approach to employing sparse representation, Candès proposed CS for reducing data acquisition costs by enabling sub-Nyquist sampling [3]. Based on his advanced theory [4], CS has been widely applied in many areas. The key idea in CS is to enable exact signal reconstruction from far fewer samples than 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 nonadaptive, 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 existing CS research contributions 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 essential in many applications that invoke CS. In 5G and IoT networks, the identified sparse domains mainly include the frequency, spatial, wavelet, and DCT domains. CS can be used to improve the SE and EE for these networks. By enabling the unlicensed use of spectrum, CRNs exploit spectral opportunities over a wide frequency range to enhance the network SE. In wide-band spectrum sensing, spectral signals naturally exploit a sparsity property in the frequency domain because of the 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 meters monitoring infrastructure in IoT applications. In particular, the monitoring readings usually have a sparse representation in the DCT domain because of 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 because of a hidden joint sparsity property from 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 [9] provided an overview on the theory of CS as a novel sampling paradigm that goes against the common wisdom in data acquisition. The study in [10] summarized CS-enabled sparse channel estimation. In [11], the authors provided a comprehensive review of the application of CS in CRNs. Other researchers presented a more specific survey on compressive covariance sensing [12] that included the reconstruction of second-order statistics even in the absence of prior sparsity information. These existing surveys serve different purposes. Some cover the basic principles for beginners, and others focus on specific aspects of CS. In contrast to the existing literature, our article provides a comprehensive overview of the recent contributions to CS-enabled wireless communications from the perspective of adopting different sparse domain projections.

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

Sparse representation

Sparse representation of signals has received extensive attention because of its capacity for efficient signal modeling and related applications. The method solves a problem by searching for the most compact representation of a signal in terms of a linear combination of the atoms in an overcomplete dictionary. The literature has focused on three aspects of sparse representation research:

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 sparse representation, such as wide-band 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. Research has 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, using 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 CS principles and the common sparse domains potentially used in 5G and IoT scenarios.

Principles of standard CS

The principles of standard CS, when it is performed at a single node, can be divided into three parts [3]: sparse representation, projection, and signal reconstruction.

Sparse representation

Generally speaking, sparse signals contain much less information than their ambient dimension suggests. The sparsity of a signal is defined as the number of nonzero 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 let $\mathbf{s}$ 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 nonzero. 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 nonadaptive measurements.

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, i.e.,

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

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

Figure 1 illustrates how the different sensing matrices $\boldsymbol{\Theta}$ influence the projection of a signal from a high dimension to its space, i.e., mapping $\mathbf{s} \in \mathbb{R}^3$ to $\mathbf{x} \in \mathbb{R}^2$. As shown in Figure 1, $\mathbf{s} = (s \;\; s \;\; 0)$ is a three-dimensional signal. When $\mathbf{s}$ is mapped into a two-dimensional space by taking

$$\boldsymbol{\Theta}_1 = \begin{pmatrix} 1 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

as the sensing matrix, the original signal $\mathbf{s}$ cannot be recorded based on the projection under $\boldsymbol{\Theta}_1$. This is because the plane spanned by the two row vectors of $\boldsymbol{\Theta}_1$ is orthogonal to signal $\mathbf{s}$, as shown in Figure 1(a). Therefore, $\boldsymbol{\Theta}_1$ corresponds to the worst projection. As shown in Figure 1(b), we can also observe that the projection by taking

$$\boldsymbol{\Theta}_2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

is not a good one. We note that the plane spanned by the two row vectors of $\boldsymbol{\Theta}_2$ can contain only part of the 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 (2-D) space. When the sensing matrix is set to

$$\boldsymbol{\Theta}_3 = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

as shown in Figure 1(c), the signal $\mathbf{s}$ can be fully recorded, as it falls into the plane spanned by the two row vectors of $\boldsymbol{\Theta}_3$. Therefore, $\boldsymbol{\Theta}_3$ results in a good projection, and $\mathbf{s}$ can be exactly recovered by its projection $\mathbf{x}$ in the 2-D space. Then, it is natural to ask what type of projection is good enough to guarantee exact signal recovery.

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

- *Restricted isometry property (RIP)*: The measurement matrix $\boldsymbol{\Phi}$ has the RIP of order K if

$$1 - \delta_K \leq \frac{\| \boldsymbol{\Phi}\mathbf{f} \|_{\ell_2}^2}{\| \mathbf{f} \|_{\ell_2}^2} \leq 1 + \delta_K \tag{3}$$

holds for all of the K-sparse signal $\mathbf{f}$, where δ_K is the restricted isometry constant of a matrix $\mathbf{\Phi}$.

- *Incoherence property*: This requires that the rows of measurement matrix $\mathbf{\Phi}$ cannot sparsely represent the columns of the sparsifying matrix $\mathbf{\Psi}$ and vice versa. More specifically, a good measurement will pick up a little bit of information from 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.

Research has pointed out [71] that verifying both the RIP condition and the incoherence property is computationally complicated but could be achieved with a high probability simply by selecting $\mathbf{\Phi}$ as a random matrix. The common random matrices include the Gaussian matrix, Bernoulli matrix, and almost all other 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 randomly 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. Previous work has demonstrated [72] 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 exact signal recovery is 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]. In contrast 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, researchers have proposed more structured sensing architectures, i.e., structured CS, to implement CS on feasible acquisition hardware. So far, many efforts have focused on designing structured CS matrices, e.g., the random demodulator [17], to make CS implementable at the expense of performance degradation. In particular, the main principle of the 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 antialiasing filter is applied, and the signal is captured by sampling it at a relatively low rate. With additional digital processing to reduce the burden on the analog hardware, the random demodulator bypasses the need for a high-rate analog-to-digital converter (ADC) [17]. A comparison of the Gaussian sampling matrix and the random demodulator is provided in Figure 2 in terms of detection probability, with different compression P/N ratios. As shown in Figure 2, the Gaussian sampling matrix performs better than the random demodulator.

Signal reconstruction

After the compressed measurements are collected, the original signal should be reconstructed. Since most of the basis coefficients in $\mathbf{s}$ are negligible, the original signal can be reconstructed by finding out the minimal set of coefficients that matches the set of compressed measurements $\mathbf{x}$, i.e., by solving

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

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

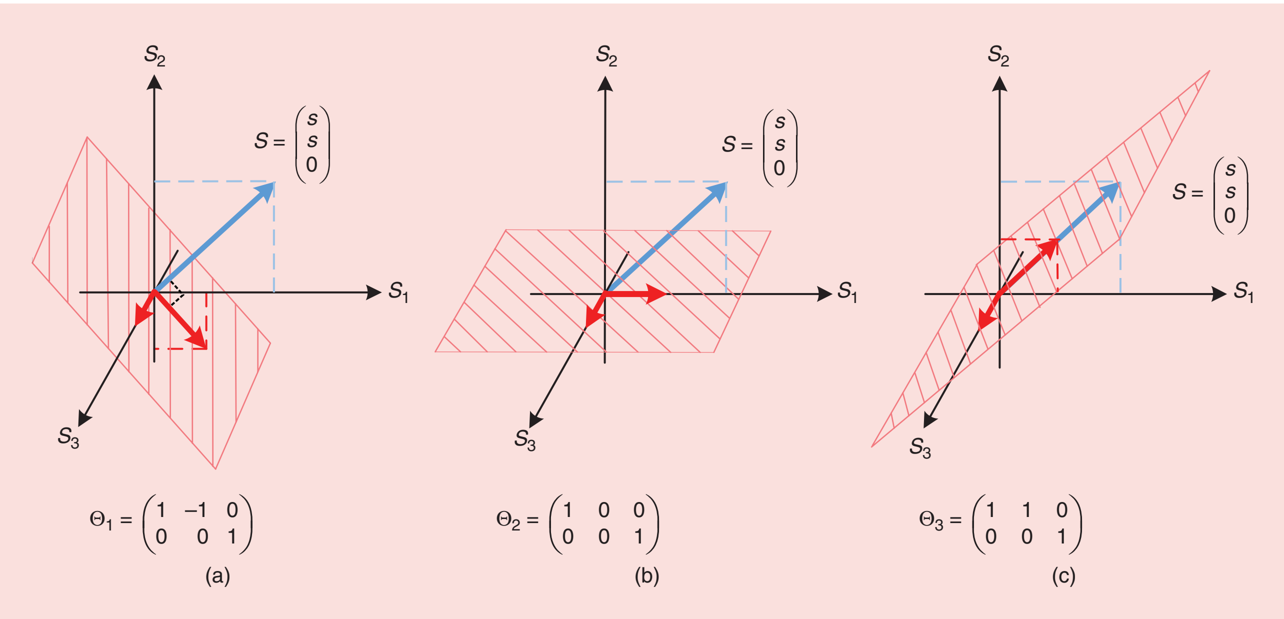


FIGURE 1. The projection of a sparse signal with one nonzero component with different sensing matrices: (a) worst projection, (b) bad projection, and (c) good projection.

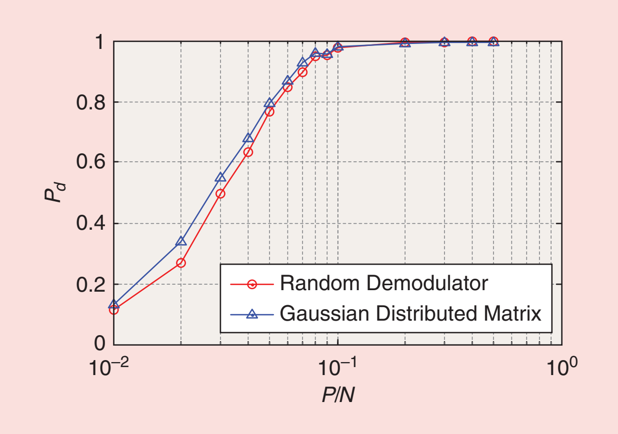


FIGURE 2. The detection probability versus the compression ratio with different measurement matrices. In this case, the signal is one-sparse and the simulation iteration is 1,000 [73].

reconstruction problem in (4) is both numerically unstable and NP-hard [3] when the 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 the ℓ_1-norm is used to substitute for the ℓ_0-norm in (4). Then, (4) can be solved by standard convex solvers, e.g., CVX. A study has proved that the ℓ_1-norm results in the same solution as the ℓ_0-norm when the 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 comparison with convex relaxation, the greedy algorithm usually requires a lower computational complexity and time cost, which makes it more practical for wireless communication systems. Furthermore, recent results have shown that the recovery accuracy achieved by some greedy algorithms is comparable to convex relaxation but requires much lower computational cost [20].

Reweighted CS

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

$$\hat{\mathbf{s}} = \arg\min_{\mathbf{s}} \|\mathbf{Ws}\|_{\ell_1} \quad \text{subject to} \quad \boldsymbol{\Theta}\mathbf{s} = \mathbf{x}, \tag{5}$$

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

Moreover, we can utilize the ℓ_p-norm, e.g., $0 < p < 1$, to lower the computational complexity of the signal recovery process caused by the ℓ_1-norm optimization problem. The authors of [21] proposed an iterative reweighted least-square (LS)-based CS approach to solve (4) in a nonconvex approach as

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

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

Equation (4) becomes nonconvex when $p < 1$. The existing algorithms cannot guarantee reaching a global optimum and may produce only local minima. However, other studies [22], [23] have proven 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.

Distributed CS

Distributed CS (DCS) [25] is an extension of the standard version that considers networks with M nodes. At the mth node, measurement x_m can be given by

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

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

In applications of standard CS, the signal received at the same node has its sparsity property because of its intracorrelation, while, for networks with multiple nodes, signals received at different nodes exhibit strong intercorrelation. The intracorrelation and intercorrelation 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 where standard CS is performed at each single node independently.

In DCS, there are two closely related concepts: distributed networks and DCS solvers. The distributed networks refer to those in which different nodes perform data acquisition in a distributed way and where standard CS can be applied at each node individually to perform signal recovery. In contrast, for DCS solvers, 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 [25]. In DCS, it is generally of interest to minimize both the computational cost and communication overhead. DCS's most popular application scenario is that all signals share the common sparse support but with different nonzero coefficients.

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, wavelet, discrete cosine, and angular domains, to name a few.

- *Frequency domain*: Because of low spectrum utilization, the wide-band spectrum signal shows a sparse property when it is converted into the frequency domain.
- *Discrete cosine domain*: Because of temporal correlation, signals in some applications, such as environmental information monitoring, show a sparse property in the discrete cosine domain, as the readings normally do not change too much within a short period.
- *Spatial domain*: As the number of paths and 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 multinode cases, because of the spatial correlation, 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 of how joint sparsity is utilized in different scenarios in 5G and IoT networks:

- *Spatial-frequency domain*: We provide an illustration of a DCS-enabled cooperative CRN in Figure 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 the frequency domain, as only a few channels are occupied. At different locations, the same frequency bands may be occupied, but the signal powers for each frequency band are various, because of 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 the "CS-Enabled CRNs" section.
- *Spatial-temporal domain*: In WSNs, sensor nodes are deployed to periodically monitor data and send the compressed data to the sink. Then, the sink is responsible for recovering the original reading using CS algorithms because the readings across all of the sensor nodes exhibit both spatial and temporal correlations, as we can see from the discussion in the "CS-Enabled Large-Scale WSNs" section.
- *Angular-time domain*: Massive MIMO channels between some users and the massive base station (BS) antennas appear to have spatial common sparsity in both the time domain and the angular domain, as we will discuss in detail in the "Channel Acquisition and Precoding in Massive MIMO" section.

In Table 1, we summarize the different sparse domains and their application in 5G and IoT networks. In addition to the list, it is worth noting that the core of applying CS is to identify how to exploit the sparse property in 5G and IoT networks.

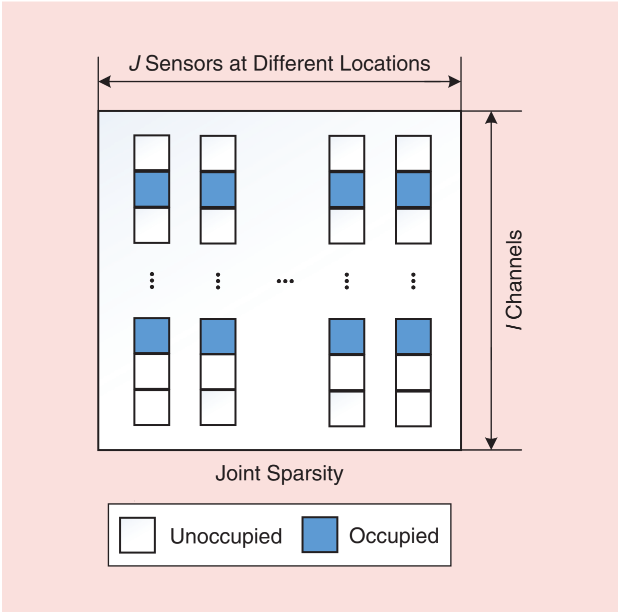


FIGURE 3. Spatial-frequency correlation.

Table 1. Some common sparse domains and their applications in 5G and IoT networks.

Sparse Domain	Sparsifying Transform	Applications	Why Sparse?	Sparsity Property
Frequency domain	FT	Wide-band 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 wide-band 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

WT: wavelet transform.

In the following, we discuss three major applications of CS in 5G and IoT networks.

CS-enabled CRNs

In this section, we introduce the applications of CS in CRNs. Radio-frequency (RF) spectrum is a valuable but tightly regulated resource, because of its unique and important role in wireless communications. The demand for RF spectrum is increasing because of a rapidly expanding market for multimedia wireless services, while the usable spectrum is becoming scarce because of the current rigid spectrum allocation policies. However, according to reports from the Federal Communications Commission and the U.K.'s Office of Communications, more commonly known as *Ofcom*, localized temporal and geographic spectrum utilization is, in reality, extremely low and unbalanced. CR has become a promising approach to solving the spectrum scarcity problem, by allowing unlicensed secondary users (SUs) to opportunistically access a licensed band when the licensed primary user (PU) is absent. To avoid any harmful interference with the PUs, SUs in CRNs should be aware of the occupancy of 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 adopt 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 that the channel frequency response can be considered 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 to several gigahertz, to improve SE and achieve higher opportunistic throughput. As driven by Nyquist–Shannon sampling theory, a simple approach is to acquire the wide-band signal directly by a high-speed ADC, which is particularly challenging or even unaffordable, especially for energy-constrained devices, such as smartphones or even battery-free devices. Therefore, revolutionary wide-band spectrum-sensing techniques are greatly desired to relieve the burden on high-speed ADCs.

Standard compressive spectrum sensing

Recent developments in CS theory inspire 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 wide-band spectrum sensing.

Energy detection-based compressive spectrum sensing

The work in [28] applied CS theory to wide-band spectrum sensing, achieving sub-Nyquist sampling without any loss of information. In Figure 4, we summarize a general framework for compressive spectrum sensing with the energy detection method, where the analog signal at the receiver $r(t)$ has a sparse representation s_f in the frequency domain. The received

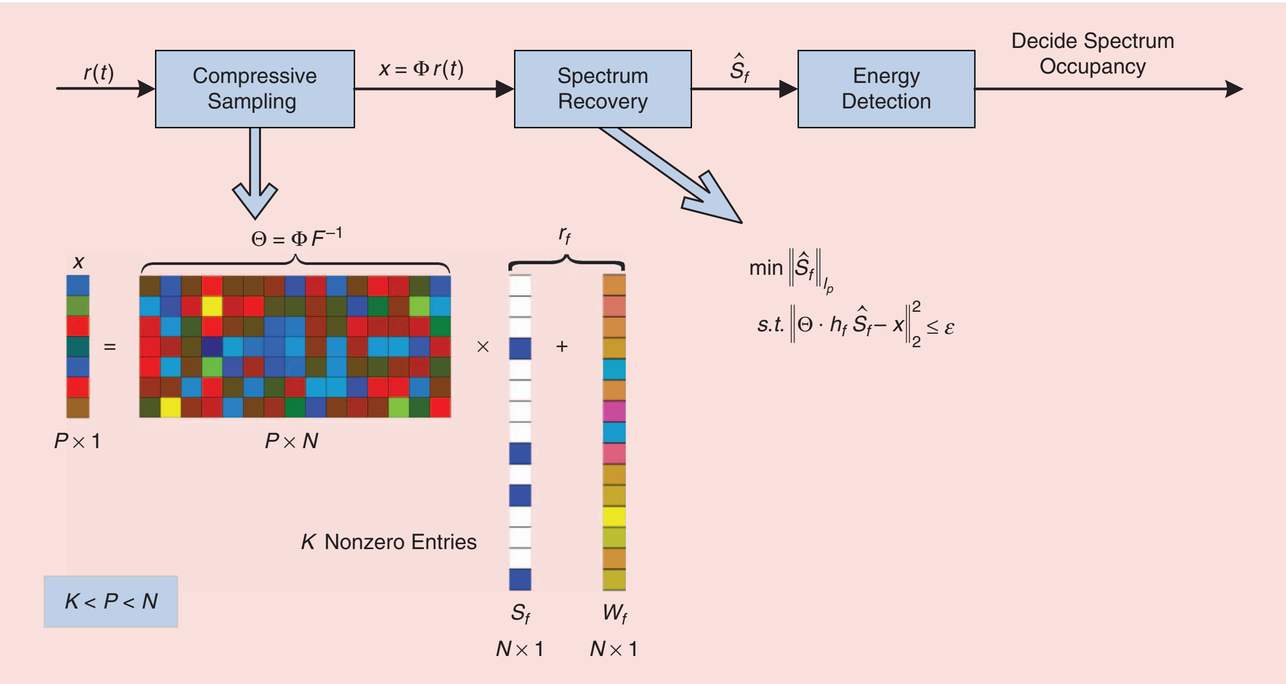


FIGURE 4. The framework of compressive spectrum sensing with energy detection.

signals are then sampled at a sub-Nyquist rate. Because of low spectrum utilization, s_f can be recovered from the undersampled measurements. Then, the energy of each channel can be calculated, and so the spectrum occupancy can be determined.

Recently, the authors of [29] identified that the CS-enabled system is somewhat sensitive to noise, exhibiting a 3-dB signal-to-noise ratio loss per octave of subsampling, which parallels the classic noise-folding phenomenon. To improve robustness to noise, [6] proposed a denoised compressive spectrum-sensing algorithm. It requires the sparsity level to determine in advance the lower sampling rate locally at SUs without loss of any information. However, the sparsity level is dependent on spectrum occupancy, which is usually unavailable in dynamic CR networks. To solve this problem, the researchers in [30] proposed a two-step CS scheme to minimize the sampling rates when the sparsity level is changing. This approach estimates the actual sparsity level first, along with the number of compressed measurements to be collected, and then makes adjustments before sampling. But this algorithm introduces extra computational complexity by performing the sparsity-level estimation.

To avoid this estimation, Sun et al. [31] proposed adjusting 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. The work in [5] proposed a low-complexity compressive spectrum-sensing algorithm by alleviating the iterative signal recovery process. More specifically, the study showed that geolocation data provided a rough estimation of the sparsity level to minimize the sampling rates. Subsequently, the authors utilized data from a geolocation database as the prior information for signal recovery. By doing so, signal recovery performance was improved, with a significant reduction in the computational complexity and a minimal number of measurements.

Compressive power spectral density estimation

In contrast to 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 the 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 wide-band 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] proposed reconstructing 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, [32] assumed the compressive measurements to be wide-sense stationary, which is not true for some compressive measurement matrices. Subsequently, Lexa et al. [33] proposed a multicoset sampling-based power spectrum estimation method to exploit the fact that a wide-sense stationary signal corresponds with a diagonal covariance matrix of the frequency domain representation of the signal. Additionally, Leus et al. [34] solved the power spectrum blind-sampling problem based on a periodic sampling procedure and further proposed a simple LS reconstruction method for power spectrum recovery.

Beyond sparsity

For spectrum blind sampling, the goal is to perfectly reconstruct the spectrum, and sub-Nyquist rate sampling is possible only if the spectrum is sparse. However, sub-Nyquist rate sampling could be achieved in [34] without making any constraints on the power spectrum, though the LS reconstruction required some rank conditions to be satisfied. Leus et al. [35] 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] developed a compressive power spectrum estimation framework for both sparse and nonsparse signals as well as blind and nonblind detection in the sparse case. For each of those scenarios, the researchers derived the minimal sampling rate allowing perfect reconstruction of the signal's power spectrum in a noise-free environment.

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) was proposed to improve sensing performance by exploiting the collaboration among all of the participating nodes. In CRNs, a CSS network constructs a multinode network. As mentioned previously, the joint sparsity and low-rank properties can be utilized to recover the original signals simultaneously with fewer measurements; DCS is employed, as it fits the CSS model perfectly. In the existing literature, cooperative compressive spectrum sensing mainly includes two categories: centralized and decentralized.

A centralized approach involves a fusion center (FC) performing signal recovery by utilizing the compressed measurements contributed by the spatially distributed SUs. The work in [6] proposed a robust wide-band spectrum-sensing algorithm for centralized CSS. Specifically, each SU senses a segment of the spectrum at a sub-Nyquist rate to reduce the sensing burden. With the collected compressive measurements, CS recovery algorithms are performed at the FC to recover the original signals by exploiting the sparse nature of spectrum occupancy. Note 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, which exploits the joint sparsity or low-rank property to recover the complete matrix.

Zeng et al. proposed a typical decentralized approach in [37], in which the decentralized consensus optimization algorithm could attain high sensing performance at a reasonable computational cost and power overhead by utilizing the joint sparsity property. In decentralized approaches, signal recovery or matrix completion is performed at each individual node. Compared with centralized approaches, decentralized ones are more robust, as they adopt an FC-free network structure. Another advantage of 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 of the participating nodes.

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

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 the performance of wide-band 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 and bandwidths, can also be utilized to enhance 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 the "Reweighted CS" section, reweighted CS normally introduces weights to provide different penalties on large and small coefficients, which naturally inspires the application of reweighted CS in wide-band spectrum sensing with available prior information. In [39], Liu and Wan divided the whole spectrum into different segments, as the bounds between different types of primary radios were known in advance. Within each segment, they proposed an iteratively reweighted ℓ_1/ℓ_2 formulation to recover the original signals. In [5], the researchers suggested a low-complexity wide-band spectrum-sensing algorithm for the TVWS spectrum to improve signal recovery performance, in which they constructed the weights by utilizing the prior information from the geolocation database. For example, in the TVWS spectrum, there are 40 TV channels, each spanning 8 MHz, which can be either occupied or not. Hence, the TV signals show a group sparsity property in the frequency domain, as the nonzero coefficients show up in clusters.

The authors of [40] developed a more efficient approach by utilizing such a group sparsity property. Moreover, the signals in wide-band spectrum sensing have the following two characteristics:
1) The input signals are stationary so that their covariance matrices are redundant.
2) Most information in signals is, in practice, concentrated on the first few lags of the autocorrelation.

Inspired by these characteristics, [41] proposed a spectral prior information-assisted structured covariance estimation algorithm with low computational complexity that especially fits with applications on low-end devices.

Potential research
We have reviewed some research results in CS-enabled CRNs, but 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 points of focus.

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 imperfections, such as noise uncertainty, channel uncertainty, dynamic spectrum occupancy, and transceiver hardware imperfection [11]. For example, 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 real life. Another example comes from the measurement matrix design. As shown in Figure 2, the Gaussian distributed matrix achieves better performance but with a higher implementation cost. Even though investigators have proposed some structured measurement matrices, such as the random demodulator, with a lower cost and acceptable recovery performance degradation to enable the implementation of CS as a replacement for high-speed ADCs, the nonlinear recovery process limits their implementation. Therefore, a big challenge exists in further exploring compressive spectrum sensing in the presence of practical imperfections and to develop a common framework to combat their aggregate effects in CS-enabled CRNs.

Generalized platform for compressive spectrum sensing
The existing hardware implementation of sub-Nyquist sampling systems follows the procedure that the theoretic algorithm is specifically designed for currently available hardware devices. However, it is very difficult or sometimes even impossible to extend current hardware architectures to implement other existing compressive spectrum-sensing algorithms. Thus, it is desired to have a generalized hardware platform that can be easily adjusted to implement different compressive spectrum-sensing algorithms with different types of measurement matrices and recovery algorithms.

CS-enabled large-scale WSNs
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 that are especially important to various IoT applications. In the typical WSN setup, a large number of inexpensive and maybe individually unreliable sensor nodes with limited energy storage and low computational capability are distributed in a smart environment to perform a variety of data processing tasks, such as sensing, data collection, classification, modeling, and tracking. Cyberphysical systems (CPSs) merge wireless communication technologies and environmental

dynamics for efficient data acquisition and smart environmental control. Typically, a CPS consists of a large number of sensor nodes and actuator nodes that, respectively, monitor and control a physical plant by transmitting data to an elaboration node called the *local controller* (*LC*)/*FC*.

Traditional environmental information-monitoring approaches take sensing samples at a predefined speed uniformly at power-constrained sensor nodes and then report the data to an LC/FC, which is normally powerful and capable of handling complex computations. The data transmitted to the LC/FC usually have redundancies that can be exploited to reduce power consumption for data transmission. A common and efficient method is to compress data at each individual sensor node and then transmit them. However, data compression introduces additional power consumption for individual sensor nodes, although the power consumption for 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 in executing a compression algorithm at the power-constrained sensor node.

We note that most natural signals can be transformed to a sparse domain, such as a 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 them to the LC/FC or the neighbor nodes. As a result, fewer measurements are sampled and transmitted, and the corresponding power consumption is significantly reduced.

Sensor node power consumption is mainly from sensing, data processing, and communications with the LC/FC. In a large-scale WSN, low-power sensor nodes seek to take samples at a lower speed or even revert to sleep mode to extend their life span. As the signal received at each sensor node shows temporal correlation and neighboring nodes display spatial correlation, the joint sparsity can be exploited to recover signals from all of the sensor nodes, even though samples from a portion of the participating nodes are missed. The active sensor nodes can be preselected according to their power levels, so the life span of the sensor nodes and of the entire network can be extended by using CS in WSNs. The existing work on CS-enabled WSNs falls mainly into the aforementioned two categories: data gathering and active node selection.

Data gathering

In the traditional data-gathering setup, there is 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. As the sensor nodes are generally limited in computational ability and energy storage, the WSN data-gathering process should be energy efficient, with low overhead, which becomes quite challenging in IoT scenarios where a huge number of sensor nodes is deployed. Taking temperature monitoring as an example, sensors will generate readings similar to those at nearby locations. Furthermore, for each sensor node, the readings from time-adjacent snapshots will be close to each other. These two important observations indicate the temporal–spatial correlations among temperature readings, which enable the application of CS to reduce the network overhead and 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 to the data after they are 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 intrasignal correlation. Moreover, it is also possible to obtain compressed measurements that are each a combination of all of the signals from the cooperative sensor nodes in a WSN. Subsequently, signals can be recovered simultaneously by exploiting both the intersignal and intrasignal correlations at the LC/FC.

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 has only P number of nonzero items, as shown by

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

A sensor node will take a sample at the moment when the corresponding item in Φ_i is 1.

However, random sampling is not proper for WSNs in practice, since two samples may be too close to each other, which becomes very challenging for cheap sensor nodes. To solve this issue, Chen and Wassell [7] proposed a random-sampling scheme by utilizing the temporal correlation of signals received at a sensor node. In the proposed scheme, the sensor node sends a pseudorandom generator seed to the FC and then sends 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 extends to the required range, the sensor node gradually increases its sampling rate until the recovery error becomes acceptable. By adopting such a scheme, the sensor node adjusts its sampling rate adaptively without knowledge of the sparsity level. To further reduce the sampling rate at sensor nodes, spatial correlation is exploited in combination with temporal correlation. Therefore, the joint sparsity property is utilized at the FC to reduce the number of required measurements.

More recently, Kong et al. proposed a big-data-enabled WSN framework [42] that invokes CS for data completion with a minimal number of samples. The proposed data collection framework consists of two core components.

1) At the cloud, an online learning component predicts the minimal data to be collected to reduce the amount for transmission; these data are considered the principal ones, and their amount is constrained by CS.
2) At each individual node, a local control component tunes the collection strategy according to the dynamics that are present and any unexpected environmental variation.

Combining these two components, this framework can reduce power consumption and guarantee data quality simultaneously.

Anomaly detection

In CS-enabled data-gathering processing, abnormal sensor readings may still lead to severe degradation in 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 abnormal readings will damage signals' sparsity property, as shown in Figure 5. Therefore, it is critical to discover the abnormal sensors to guarantee the security of WSNs and make them 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, at least, be removed. But the abnormal readings caused by external events should be preserved, as they reflect actual WSN scenarios.

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

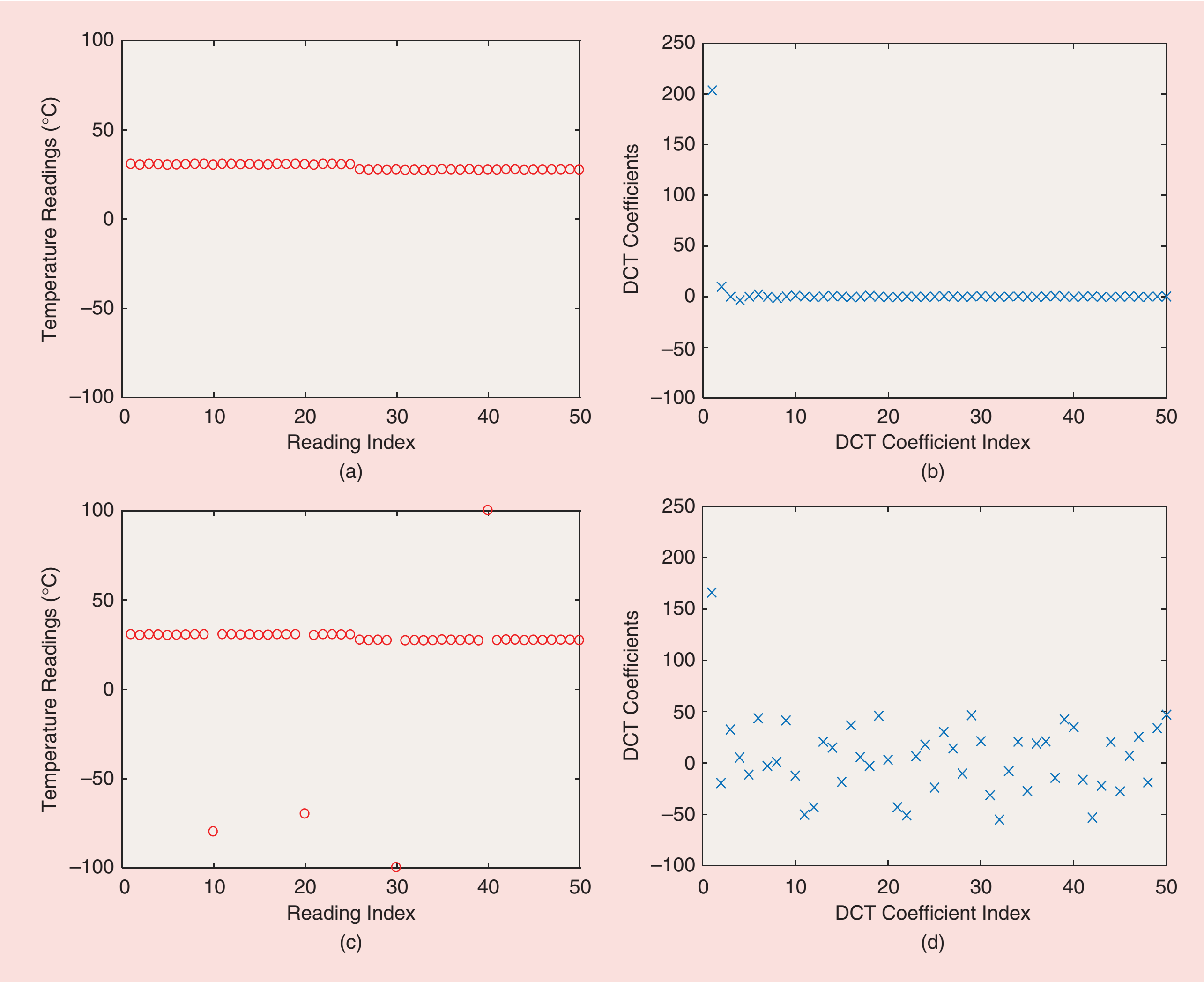


FIGURE 5. The effect of abnormal readings on the sparsity level of temperature readings in the discrete cosine domain: (a) the original signal, (b) the DCT of the original signal, (c) an abnormal signal, and (d) the DCT of the abnormal signal.

Active node selection

In large-scale WSNs, the events are relatively sparse in comparison with the number of sensor nodes. Because of the power constraint, it is unnecessary to activate all of the sensor nodes all of 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 harming performance. Taking a smart monitoring system as an example, as shown in Figure 6, the number of source nodes is N, and there are K $(K \ll N)$ sparse events that are generated by the N source nodes. By invoking CS, only M $(M \le N)$ active sensor nodes are required to capture the K sparse events.

Centralized node selection

The researchers in [43] proposed a centralized node selection approach by applying CS and matrix completion at the LC/FC with the purpose of optimizing the network throughput and extending sensor lifetime. 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 focuses on only the spatial correlation of sensor nodes, Chen and Wassell [44] 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, because of the different physical conditions at different sensor nodes. For example, to extend the WSN lifetime, it is preferable to activate sensor nodes with adequate energy rather than those almost out of energy. Therefore, Chen and Wassell proposed a cost-aware node selection approach in [45] to minimize the sampling cost of the whole WSN, with constraints on the reconstruction accuracy.

Decentralized node selection

In contrast to 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 aim to achieve in-network processing for node selection. Long and Tian proposed a decentralized approach in [46] to perform node selection by allowing each active sensor node to monitor and recover only its local data 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 optimizes not only for itself but also for its inactive neighbors. Moreover, to extend the network lifetime, Caione et al. developed an in-network CS framework [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 the distributed node selection approach are the following: 1) The optimized node selection requires an iterative process, which may require a long period of time. 2) 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. But, in the centralized approach, extra bandwidth resources and power consumption are required to coordinate the active sensor nodes.

Potential research

Even though researchers have carried out extensive studies 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 the CS recovery accuracy in addition to further reducing the power consumption.

Machine learning-aided adaptive measurement matrix design

Even though different applications entail different constraints, the core concept for sparse representation remains the same. Therefore, it is natural to ask if there is a general framework for the sparse representation of the data in different 5G and IoT applications for urban scenarios. 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 [74] that machine learning can be an efficient tool to aid the measurement matrix design so that the lifetime of the whole network as well as of each individual sensor node can be extended to the utmost. Furthermore, we should note that, when designing a measurement matrix, the possibility of its implementation in a

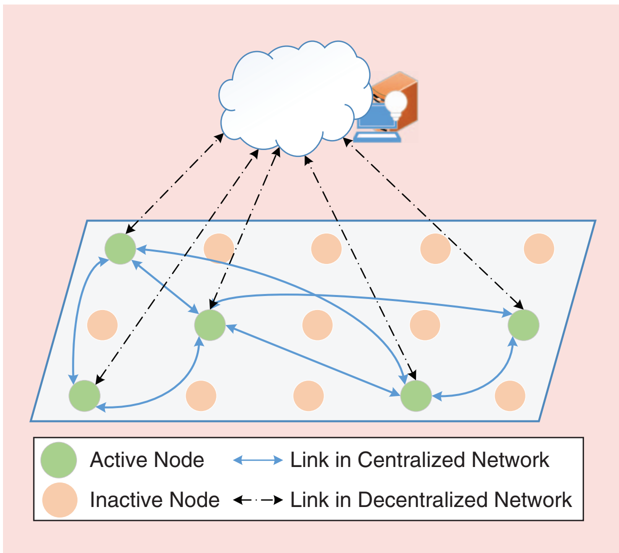


FIGURE 6. The node selection in CS-based WSNs.

real network is one of the most critical factors to be considered. We believe that extensive research work in this direction is highly desirable.

Data privacy in CS-enabled IoT networks
The data collected from a variety of IoT network sensors, including data about our daily activities, surroundings, and even our personal physical information, can be recorded and analyzed, which at the same time greatly intensifies the risk of compromising privacy. There are few effective privacy-preserving mechanisms in mobile sensing systems. To provide privacy protection, then, some investigators have proposed adding noise to the original data [75]. The added noise would conceal the data when the original sparse data points are zero or near zero, thus reducing the data sparsity. However, CS aims to achieve this kind of efficiency for sparse data processing. This creates a conflict between privacy and efficiency involving big data processing in CS-enabled IoT networks. Therefore, a great deal of research is expected in this area.

Channel acquisition and precoding in massive MIMO
To satisfy the high data rate requirements stemming from increasing mobile applications, many efforts have been made to improve transmission SE. An effective way is to exploit the spatial degrees of freedom provided by large-scale antennas at the transmitter and the receiver to form massive MIMO systems [48]. The work in [48] showed 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 antennas at the BS is very large. Consequently, linear processing is good enough to make the system performance approach the optimum if the channel state information (CSI) is known at the BS.

Accurate CSI at the BS is essential for massive MIMO to obtain the aforementioned advantage. Because of 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 duplex (TDD) systems, the downlink CSI can be easily obtained by exploiting channel reciprocity. However, most actually deployed systems mainly employ frequency-division duplex (FDD), where channel reciprocity no longer holds. In this situation, the downlink channel has to be estimated directly and then fed back to the BS, which results in extremely high overhead.

To address the CSI estimation and feedback issue in FDD systems, the sparsity of massive MIMO channels must be exploited. Researchers have proposed some CS-enabled CSI acquisition methods for FDD massive MIMO, where they have successfully exploited the correlation in massive MIMO channels to reduce the number of training symbols and the amount of feedback overhead [50].

In this section, we focus on CS-enabled channel acquisition and its related applications. We will first introduce the channel sparsity feature, then discuss channel estimation and feedback, and after that explore precoding and detection. We should note that millimeter-wave (mm-wave) communications are often used with massive MIMO techniques, since the short wavelength makes it very easy to pack a large number of antennas in a small area. In the following discussion, we will also include channel acquisition and precoding based on CS in mm-wave massive MIMO, even if mm-wave channels are slightly different from the traditional wireless channels.

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 modeled as independent random variables. However, this assumption is no longer true in massive MIMO systems. It has been shown that the massive MIMO channel is spatially correlated and has a sparse structure in the spatial domain. This correlation and sparsity are due to the exploitation of high RF and the deployment of large-scale antenna arrays in future wireless communications. In the high-frequency band, the channels have fewer propagation paths, while more transmit and receive antennas make the number of distinguished paths much fewer than the number of channel coefficients; the rich scatters then become limited or sparse.

As shown in Figure 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 because of the shared scatterers and have an independent sparsity support caused by the individual scatterers in the propagation environments. Using this sparsity structure, investigators have proposed many CS-enabled channel acquisition and precoding schemes, as we will discuss in the following.

Compressive pilot design
To obtain a good channel estimation, the length of the orthogonal training sequence must be at least the same as the number of transmit antenna elements. Because of 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. Research has 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. Basing their work on this principle, Lau et al. proposed CS-enabled pilot design schemes [50].

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

Besides the general massive MIMO systems, Alkhateeb et al. designed a compressed hierarchical multiresolution codebook specially for mm-wave massive MIMO systems to construct training beamforming vectors [51]. Based on this idea, others have proposed many different hierarchical multiresolution codebooks from different angles. For example, the research in [52] produced a design of a compressive beacon codebook with different sets of pseudorandom phases. The authors of [53] proposed a common codebook satisfying the conflicting design requirements as well as validating practical mm-wave systems through utilizing the strong directivity of mm-wave channels. Wang et al. [54] proposed a multiresolution uniform-weighting-based codebook, similar to a normalized discrete FT matrix, to reduce the implementation complexity, where the estimation of the angle of arrival and angle of departure 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 practical systems.

Compressive channel estimation and feedback

CS can be used to reduce the cost of 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 the exploiting of sparsity in different domains.

With time-domain sparsity

Work has shown that the channel slowly varies in a number of applications so that the prior channel estimation result can still be used to reconstruct subsequent channels, as illustrated in Figure 8. By using the structure in the figure, investigators have proposed several CS algorithms [8], [55], [56] to recover massive MIMO channels.

Exploiting the common time-domain support, Shen et al. proposed a CS-enabled differential CSI estimation and feedback scheme [55]. The approach in [56] further combines LS and CS techniques to improve estimation performance. Here, the LS and CS techniques can be used to, respectively, 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. Since the current channel vector has only a small number of nonzero 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,

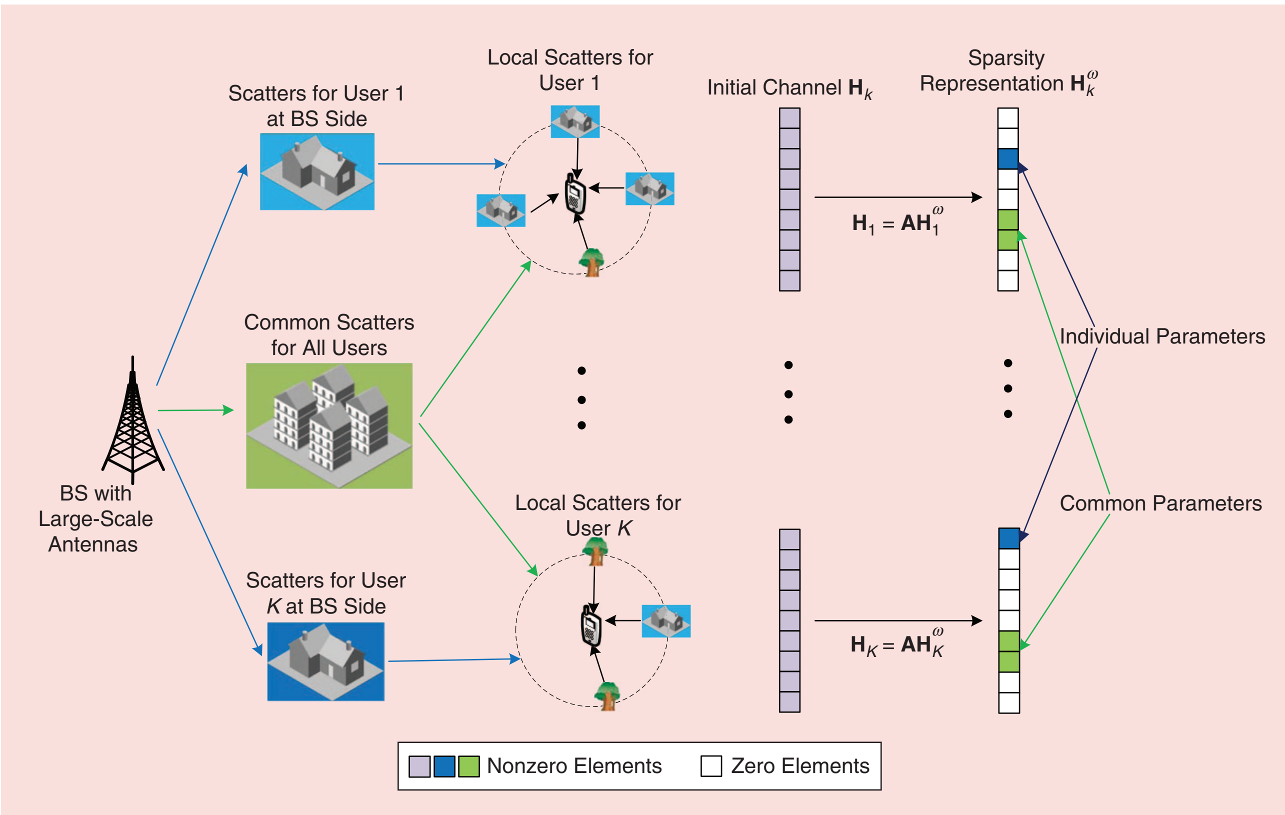


FIGURE 7. A channel model with limited scatters, where **A** is the spatial correlation matrix.

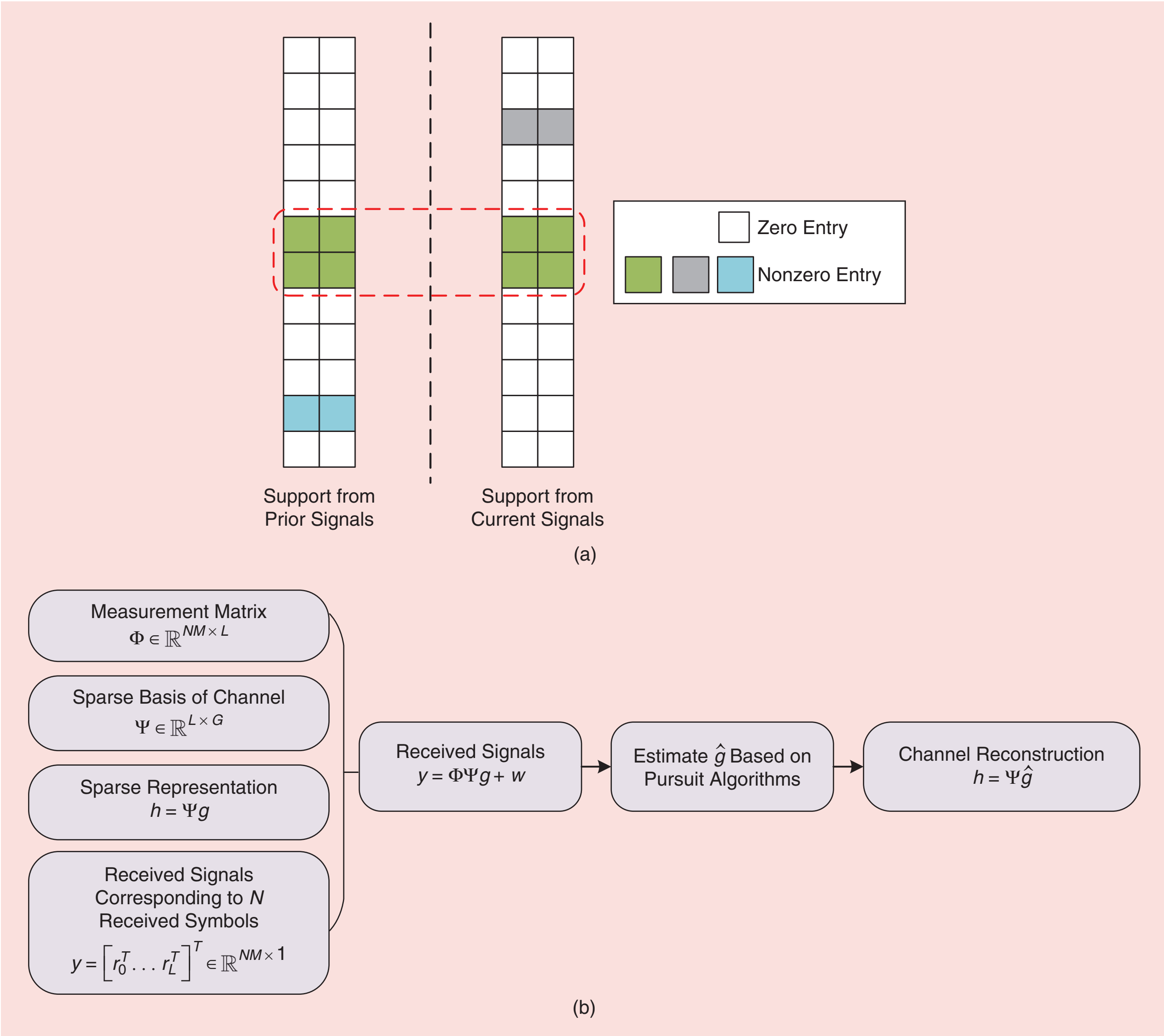


FIGURE 8. An illustration of compressive channel estimation with time-domain correlation signal support: (a) support from prior and current signals and (b) compressive channel estimation by using time-domain correlation.

Rao and Lau developed a greedy pursuit-based approach with the prior support information and its quality information [8], where the prior support information is adaptively exploited based on its quality information to further improve the channel estimation performance.

Besides channel estimation, [55] and [49] also proposed several channel feedback schemes. To exploit the common support, Shen et al. developed a CS-enabled differential CSI feedback scheme [55] by using the temporal correlation of MIMO channels. The proposed scheme can reduce the feedback overhead by about 20% compared with the direct CS-enabled channel feedback. In [49], Liu et al. developed a robust closed-loop pilot and CSI at the transmitter (CSIT) feedback resource adaptation framework 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.

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], Tseng et al. employed a block optimization algorithm 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. The authors of [59] proposed a spatial sparsity-based compression mechanism to reduce the load of the channel feedback. In this mechanism, they used random projection with an unknown sparsity basis and direct compression based on known sparsity. Since the spatial sparsity will reduce channel rank, to obtain the compressed channel estimation, they proposed a dictionary-learning method that captures the communication environment and the antenna property.

After researchers designed the compressed hierarchical multiresolution codebook [51]–[54], the corresponding channel estimation schemes could be developed. In these approaches, the hierarchical multiresolution codebook was capable of generating variable beamwidth radiation patterns to facilitate the use of robust adaptive multipath channel estimation algorithms. Meanwhile, the exploitation of adaptive CS algorithms would also reduce implementation complexity and estimation error.

With spatial-temporal sparsity

In the aforementioned schemes, the investigators independently exploited both the temporal- and spatial-domain sparsities. In practice, they can be jointly used to further reduce the cost of channel estimation and feedback.

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

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 researchers first exploited the sparsity in the spatial-frequency domain and proposed an adaptive CS-enabled feedback scheme to reduce the feedback overhead. In this approach, the feedback can be dynamically configured based on the channel conditions to improve the efficiency. Because of sharing sparse common support for the adjacent subcarriers in orthogonal frequency-division modulation (OFDM), the investigators developed an approximate message passing with the nearest-neighbor sparsity pattern-learning algorithm to adaptively learn the underlying structure to obtain better performance.

Precoding and detection

The precoder design based on the estimated CSI is a very important problem in massive MIMO systems, especially in the mm-wave wide-band systems. Since the wide-band mm-wave massive MIMO channel generally exhibits frequency-selective fading, the precoder design based on the estimated CSI becomes challenging. Generally, the sparse structure of mm-wave massive MIMO channels in angle domain or beam space can be used to simplify the precoder design [63], [64]. In [63], Venugopal et al. used compressive subspace estimation to get the full channel information and then design the precoder to maximize the system SE. To reduce the CSI acquisition cost and address the mismatch between few RF chains and many antennas, the authors used the hybrid precoding design with baseband and RF precoders in mm-wave massive MIMO systems. However, this hybrid precoding introduces performance loss. To mitigate this, [64] utilized an iterative OMP to refine the quality of the hybrid precoders. Meanwhile, a limited feedback technique was also 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 detector has a prohibitively high complexity. Thanks to the structured sparsity of multiple SM signals, a low-complexity signal detector based on CS was introduced to improve signal detection performance. In [65], Gao et al. developed a joint SM transmission scheme for user equipment and a structured CS-enabled multiuser detector for the BS. The proposed detector can reliably detect the resultant SM signals with low complexity by exploiting the intrinsic sparse features.

CSI acquisition and precoding based on CS theory in FDD massive MIMO systems have been mainly discussed so far. In practice, the CS theory can also be applied to TDD massive MIMO systems. Investigators developed a channel estimation approach based on block-structured CS [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 significantly reduced.

Potential research

Researchers have successfully used CS in massive MIMO to improve the performance of channel estimation and precoding. However, there are still many open topics that need to be explored before CS can be implemented in massive MIMO systems. We will discuss some of them in this section.

Effect of antenna deployment

Because of space limitations, large-scale antennas may be deployed in various topologies, i.e., centralized or distributed. Since the different antenna topologies correspond to different channel sparsities, the effect of antenna configuration on the performance of CS-enabled channel estimation is still an open research question.

Measure of channel sparsity

As mentioned previously, channel sparsity is very important for CS-enabled channel estimation. In current research, investigators have assumed and exploited various sparsity models in channel acquisition. Many of these assumed sparsity models lack verification by measured results. Thus, either the sparsity model or the CS-enabled approach needs to be confirmed by some measured results under different propagation conditions.

Channel estimation and feedback with joint support

In the aforementioned 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. Future research needs to explore whether performance can be further improved if all three types of channel sparsity are used and, if so, how much performance gain might be achieved.

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 studying how channel estimation and feedback approaches should be designed without channel sparsity support.

Other CS applications

CS-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 the targets, far fewer numbers 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.

CS-aided impulse noise cancelation

In certain applications, such as OFDM, impulsive noise will severely degrade the system performance. The OFDM signal is often processed in the frequency domain. Even if the impulse noise lasts only a short time, it affects a wide frequency range. By regarding impulse noise as a sparse vector, one study exploited CS to mitigate such noise [67].

CS-aided cloud radio access networks

Cloud radio access networks (C-RANs) have been proposed as a promising technology to support massive connectivity in 5G networks. In C-RANs, 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 in green C-RANs, provided the quality of service is guaranteed. More specifically, one RRH will be switched off only when all the coefficients in its beamformer are set to zero. Such a group sparsity property inspires us to apply CS to active RRH selection in green C-RANs to minimize network power consumption [68], [69]. Additionally, in C-RANs' uplink, the channel estimation from the active users to the RRHs is the key to achieving 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 exhibits a joint sparsity property, which can further facilitate the active user detection and channel estimation in C-RANs [70].

Conclusions

This article provided a comprehensive overview of sparse representation, with applications in wireless communications. Specifically, after introducing the basic CS principles, we identified the common sparse domains in 5G and IoT networks. Subsequently, we discussed the exploitation of different sparsity properties in three CS-enabled networks: wide-band spectrum sensing in CRNs, data collection in IoT networks, and channel estimation and feedback in massive MIMO systems. In the previous discussion, we concluded that, by invoking CS, the SE and EE of 5G and IoT networks can be enhanced from different perspectives. Furthermore, we identified potential research challenges to provide a guide for researchers interested in sparse representation in 5G and IoT networks.

Acknowledgments

Jiancun Fan is supported by the National Natural Science Foundation of China under grant 61671367, and by the China Postdoctoral Science Foundations under grants 2014M560780 and 2015T81031. Yue Gao is supported by funding from the Physical Sciences Research Council in the United Kingdom through grant EP/R00711X/1.

Authors

Zhijin Qin (zhijin.qin@lancaster.ac.uk) received her B.S. degrees from Queen Mary University of London (QMUL) and Beijing University of Posts and Telecommunications in 2012, and her Ph.D. degree in electronic engineering from QMUL in 2016. She was with Imperial College London as a research associate. She has been a lecturer (assistant professor) with Lancaster University, United Kingdom, since August 2017. Her research interests include compressive sensing, the Internet of Things, and nonorthogonal multiple access. She received the Best Paper Award at the 2012 Wireless Technology Symposium and the 2017 IEEE Global Communications Conference.

Jiancun Fan (fanjc@xjtu.edu.cn) received his B.S. and Ph.D. degrees in electrical engineering from Xi'an Jiaotong University, China, in 2004 and 2012, respectively. From 2009 to 2011, he was a visiting scholar at the Georgia Institute of Technology, Atlanta. He is currently an associate professor at Xi'an Jiaotong University. His general research interests include signal processing and wireless communications, with an emphasis on multiple-input, multiple-output communication systems, cross-layer optimization for spectral- and energy-efficient networks, practical issues in long-term

evolution and fifth-generation systems, and machine learning. He was a recipient of the Best Paper Award at the 2017 20th International Symposium on Wireless Personal Multimedia Communications.

Yuanwei Liu (yuanwei.liu@qmul.ac.uk) received his B.S. degree from Queen Mary University of London (QMUL) and Beijing University of Posts and Telecommunications (BUPT) in 2011, his M.Sc. degree from BUPT in 2014, and his Ph.D. degree in electrical engineering in 2016 from QMUL, where he is now a lecturer (assistant professor). His research interests include fifth-generation wireless networks, the Internet of Things, and stochastic geometry. He received the Exemplary Reviewer Certificate from *IEEE Wireless Communication Letters* and *IEEE Transactions on Communication*. He serves as an editor of *IEEE Communications Letters* and *IEEE Access*.

Yue Gao (yue.gao@qmul.ac.uk) received his B.S. degree from Beijing University of Posts and Telecommunications in China in 2002 and his M.Sc. and Ph.D. degrees in telecommunications and microwave antennas from Queen Mary University of London (QMUL) in the United Kingdom in 2003 and 2007, respectively. He is a reader and director of the Whitespace Machine Communication Laboratory at QMUL. His laboratory is conducting theoretical research into practical aspects of the interdisciplinary area that spans antennas, signal processing, spectrum sharing, and the Internet of Things. He is the secretary of the IEEE Technical Committee on Cognitive Networks and serves as an editor of *IEEE Transactions on Vehicular Technology* and *IEEE Wireless Communication Letters*. He was a corecipient of the EU Horizon Prize Award on Collaborative Spectrum Sharing in 2016, and of the Research Performance Award from QMUL in 2017. He is an Engineering and Physical Sciences Research Council Fellow (2018–2023). He is a Senior Member of the IEEE.

Geoffrey Ye Li (liye@ece.gatech.edu) received his B.S.E. and M.S.E. degrees in 1983 and 1986, respectively, from the Department of Wireless Engineering, Nanjing Institute of Technology, China, and his Ph.D. degree in 1994 from the Department of Electrical Engineering, Auburn University, Alabama. Since 2000, he has been with the School of Electrical and Computer Engineering at the Georgia Institute of Technology, Atlanta, as an associate professor and then a full professor. His general research is in signal processing and machine learning for wireless communications. He has won the IEEE Communications Society (ComSoc) Stephen O. Rice Prize Paper Award, the IEEE ComSoc Award for Advances in Communication, the IEEE Vehicular Technology Society (VTS) James Evans Avant Garde Award, the IEEE VTS Jack Neubauer Memorial Award, the IEEE Signal Processing Society Donald G. Fink Overview Paper Award, and the Distinguished Electrical and Computer Engineering Faculty Achievement Award from the Georgia Institute of Technology. He is a Fellow of the IEEE.

References

[1] M. Zibulevsky and M. Elad, "L1-L2 optimization in signal and image processing," *IEEE Signal Process. 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 Rev.*, vol. 51, no. 1, pp. 34–81, Feb. 2009.

[3] E. Candès, "Compressive sampling," in *Proc. Int. Congr. Mathematics*, vol. 3, Madrid, Spain, Oct. 2006, pp. 1433–1452.

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

[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, June 2016.

[7] W. Chen and I. J. Wassell, "Energy-efficient signal acquisition in wireless sensor networks: A compressive sensing framework," *IET Wireless Sensor Syst.*, 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, Sept. 2015.

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

[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," *Commun. Surveys Tuts.*, vol. 18, no. 3, pp. 1838–1860, 2016.

[12] D. Romero, D. D. Ariananda, Z. Tian, and G. Leus, "Compressive covariance sensing: Structure-based compressive sensing beyond sparsity," *IEEE Signal Process. 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 *Proc. IEEE Int. Conf. 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, June 2010.

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

[16] E. Candès and J. Romberg, "Sparsity and incoherence in compressive sampling," *Inverse Problems*, vol. 23, no. 3, pp. 969–985, 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. Candès. (2008). The restricted isometry property and its implications for compressed sensing. *Comptes Rendus Mathematique*. [Online]. *346(9)*, pp. 589–592. 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," *Commun. Surveys Tuts.*, vol. 19, no. 3, pp. 1527–1550, 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, May 2008. doi: 10.1088/0266-5611/24/3/035020.

[24] R. Chartrand and W. Yin, "Iteratively reweighted algorithms for compressive sensing," in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing (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," *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. Acousics, Speech, and Signal Processing (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, Sept. 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, June 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. Acoustics, Speech, and Signal Processing (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. Acoustics, Speech, and Signal Processing (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, Sept. 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. Topics Signal Process.*, 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. Advances Signal Process.*, vol. 2012, no. 177, Dec. 2012. doi: 10.1186/1687-6180-2012-177.

[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, June 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] W. Chen and I. J. Wassell, "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, July 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. Topics Signal. Process.*, 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. Topics Signal. Process.*, vol. 10, no. 3, pp. 514–527, 2016.

[53] J. Song, J. Choi, and D. J. Love, "Common codebook millimeter wave beam design: Designing 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. Communication (ICC)*, 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, July 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. Commun. 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. Wireless Commun.*, vol. 15, no. 7, pp. 4839–4851, July 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 Communication Network 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, June 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 *Proc. European Signal Processing Conf. (EUSIPCO)*, Aug. 2015, pp. 2859–2863.

[68] Y. Shi, J. Zhang, and K. B. Letaief, "*Group sparse beamforming green cloud-RAN*," *IEEE Trans. Wireless Commun.*, 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. Communication (ICC)*, London, June 2015, pp. 2727–2732.

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

[72] B. Bah and J. Tanner, "Improved bounds on restricted isometry constants for gaussian matrices," *SIAM J. Matrix Anal. Appl.*, vol. 31, no. 5, pp. 2882–2898, 2010.

[73] Y. Gao and Z. Qin, "Implementation of compressive sensing with real-time signals over TV white space spectrum in cognitive radio," in *Proc. IEEE Vehicle Technology Conf. (VTC-Fall)*, Montréal, Canada, Sept. 2016, pp. 1–5.

[74] B. Sun and H. Feng, "Efficient compressed sensing for wireless neural recording: A deep learning approach," *IEEE Signal Process. Lett.*, vol. 24, no. 6, pp. 863–867, June 2017.

[75] C. Dwork, F. McSherry, K. Nissim, and A. Smith, "Calibrating noise to sensitivity in private data analysis," in *Proc. Theory Cryptography*, Springer, 2006, pp. 265–284.