

Full length article

Scalable spectrum database construction mechanisms for efficient wideband spectrum access management[☆]Bassem Khalfi^{a,*}, Bechir Hamdaoui^a, Mohsen Guizani^b, Abdurrahman Elmaghrib^a^a Oregon State University, United States of America^b Qatar University, Qatar

ARTICLE INFO

Article history:

Received 14 April 2020

Received in revised form 1 March 2021

Accepted 1 March 2021

Available online 16 March 2021

Keywords:

Heterogeneous wideband access

Distributed compressive sampling

Local low-rank matrix completion

ABSTRACT

We propose a novel framework for enabling scalable database-driven dynamic spectrum access and sharing of heterogeneous wideband spectrum. The proposed framework consists of two complementary approaches that exploit the merits of compressive sensing theory, low-rank matrix theory, and user cooperation to build an accurate heterogeneous wideband spectrum map by overcoming the time-variability of the number of occupied bands, the need for a high number of measurements per sensing node (SN), the inherent wireless channels' impairments, and the high reporting network overhead. First, exploiting the fact that close-by SNs have a highly correlated spectrum observation, we leverage distributed compressive sensing to enable cooperative heterogeneous wideband spectrum sensing only from a small number of measurements per each SN. Second, to reduce the network overhead due to the high width of the spectrum of interest, we propose a two-step approach that performs spectrum occupancy recovery using the local low-rank property of occupancy sub-matrices. Then, we combine the completed sub-matrices entries to produce the whole spectrum occupancy matrix. Through simulations, we show that the proposed framework efficiently achieves high detection in the sensing step and minimizes the spectrum occupancy matrix recovery error while reducing the overall network overhead.

© 2021 Elsevier B.V. All rights reserved.

1. Introduction

The proliferation of wireless devices and applications during the last decade (e.g., IoT devices and 5G) has increased the demand for wireless bandwidth, and hence a serious shortage problem in the radio spectrum supply has come to the surface. Dynamic spectrum access is a potential solution to address the need for high data-rates and support the connectivity of the growing number of wireless devices in the era of 5G by overcoming the spectrum scarcity and enhancing the spectrum utilization [3]. It achieves that by allowing opportunistic access to temporarily unused portions of the spectrum without interfering with the primary users or other protected wireless systems through spectrum awareness techniques, which can be classified into two families: sensing-based approaches [4–7] and database-driven approaches [8–13]. Sensing-based sensing approaches enable secondary users (SUs) to detect the empty channels in the spectrum on their own by locally scanning it. In

contrast, database-driven approaches provide SUs with a spectrum occupancy database, which they can query for temporarily available spectrum resources. These spectrum databases are constructed by combining all the local measurement reports from different sensing nodes (SNs) via a fusion center that manages the spectrum access policies. These SNs are dedicated to sensing the local environment, which have also been referred to as Environment Sensing Capability (ESC) [9]. ESC plays a critical role in the Citizen Broadband Radio Service (CBRS) spectrum sharing system in the United States. They are deployed in the coastal area to detect incumbent users' transmission activities and then update the spectrum sharing system (SAS) to adjust their database. While sensing-based approaches have been considered in 4G and 5G systems for re-using Wi-Fi 5 GHz bands through License Assisted Access (LAA) feature, database-driven approaches have been embraced and promoted by various companies (e.g., Google [12], RadioSoft [13] and few others), standards organizations (e.g., 5G and CBRS), and independent federal agencies (e.g., FCC [14]).

The current focus of database-driven approaches has been on the TV spectrum band [8] and, more recently, on the DoD's 3.5 GHz radar band [9]. Many other bands can be shared in the near future, which will bear significant fruits to the wireless spectrum industry. Additionally, TV signals propagate long distances, which can be sensed with a small number of SNs in a

[☆] Parts of this paper (Khalfi et al., 2018 [1,2]) have been presented in proc. of IEEE Global Telecommunications Conference (GlobeCom) 2018.

* Corresponding author.

E-mail addresses: khalfib.oregonstate@gmail.com (B. Khalfi), hamdaoui@eecs.oregonstate.edu (B. Hamdaoui), mguizani@ieee.org (M. Guizani), elmaghba@eecs.oregonstate.edu (A. Elmaghrib).

given region. Therefore, extending spectrum databases to cover wider bandwidth, around 30 GHz bandwidth or more, faces two main challenges: (i) *the high sampling rate requirement* resulting from the wideband spectrum at hand, and (ii) *the wireless signal propagation loss and attenuation* resulting from the transmission at high frequencies.

To overcome the first challenge, sub-Nyquist sampling approaches relying on the theory of compressive sensing (CS) have been proposed [4,15], allowing wideband spectrum occupancy recovery from small numbers of samples by taking advantage of the spectrum occupancy sparsity. Despite overcoming the need for high sampling rates, existing CS-based approaches have two main shortcomings. First, practical receiver hardware designs allow only a much smaller number of measurements than what these CS-based approaches necessitate to operate, which typically addressed through either sequential sensing (a device performs multiple sequential sensing scans) or by sensing only a smaller number of bands [16–18]. However, these approaches incur either excessive recovery delays or limit the number of bands that can be exploited. Second, the number of implemented hardware branches in the CS-based approaches is fixed and it depends on the band occupancy – or sparsity – level. Practically, a SU does not have a prior knowledge about this parameter, and it changes over time, making it challenging to achieve accurate recovery [15,16]. Hence, if the number of present signals exceeds the pre-specified sparsity, the signal recovery of the system will not be reliable. On the other hand, implementing more hardware branches than what is needed would increase the system's complexity and results in an excessive Silicon area.

As for overcoming the second challenge—i.e., signal propagation decay and attenuation, some research efforts relying on denser SN deployment and on leveraging low-rank matrix theory to exploit the spatial correlation between SNs' observations have been proposed [7,19]. The basic idea behind these approaches is exploiting the fact that SNs deployed within the same vicinity are very likely to observe the same (roughly) wideband spectrum occupancy, thereby resulting in a spectrum occupancy matrix, whose columns represent the either occupancy decisions or the spectrum measurement reports from the close-by SNs, possesses low-rank matrix property. The low-rank matrix property enables the system to reconstruct the spectrum occupancy matrix based on a small number of its entries [7,19]. Although these approaches reduce the network overhead substantially, they fail when the number of bands increases and the bandwidth of interest is at high frequencies, due to the propagation decay that signals experience at different frequencies (e.g., millimeter wave frequencies adopted by 5G systems [20]). This is because SNs located at different locations tend to observe utterly different spectrum occupancies (especially at high frequencies), which violates the low-rank property of the spectrum occupancy matrix and precludes exploiting in the case of wideband access.

In this paper, we propose complementary methods that exploit the wideband spectrum occupancy heterogeneity and characteristics to build an accurate database for wideband spectrum access systems with minimum overhead. First, we leverage the distributed compressive sensing theory to overcome the need for a high number of measurements per SN while accounting for the spectrum occupancy heterogeneity [1]. Second, we propose a *local low-rank matrix approximation* framework that builds on the low-rank matrix theory by adding the *spatial* dimension to allow a precise global occupancy recovery while reducing sensing overhead [2]. Our approach focuses on exploiting the low-rank matrix property that is present at different sub-matrices of the overall spectrum occupancy matrix, with each sub-matrix corresponds to one neighborhood.

The contributions of this paper can be summarized as follows:

- We propose an efficient wideband sensing framework that enables scalable construction of the spectrum occupancy matrix for wideband spectrum access.
- We prove that taking a small number of measurements per SN is enough to decide on the spectrum occupancy in the neighborhood of each SN through distributed compressive sensing while exploiting spectrum occupancy heterogeneity.
- To the best of our knowledge, we are the first to use local low-rank matrix approximation theory to reduce network overhead and exploit that to build a global and scalable wideband spectrum occupancy database.

The rest of this paper is organized as follows. Section 2 presents the related works. Section 3 presents our system model. Section 4 presents our cooperative approach for efficient wideband spectrum recovery. Section 5 presents our approach for reducing database construction overhead through local low-rank matrix approximation. Section 6 presents the performance evaluation and result analysis. Section 7 concludes the paper.

2. Related works

Spectrum awareness. Our framework combines advances in wideband spectrum sensing [4–8,21–30] and recent advances in recommendation systems [31,32]. Authors in [8] propose a trustworthy spectrum availability database for only TV band spectrum. Similarly, authors in [33] formulate the spatial reuse of the TV White Space (TVWS) spectrum as a schematic optimization approach in which they used a fast matrix completion, a nonlinear support vector machine, and opportunistic spatial reuse algorithms. This approach aims to build a location-specific TVWS database that can be used by a D2D link to determine the maximum permitted power level based on hardware constraints and the acceptable level of interference. Spectrum occupancy database is one of the main entities in a CBRS spectrum Access Sharing (SAS). It provides all band-related information required in the access policy-making process, such as occupancies, spectral locations of CBRS and incumbent devices, operation duration, and transmission power levels. In [27], a wideband sensing-based framework is proposed to achieve high spatial reuse of the unlicensed spectrum in crowded IoT networks. In [9], authors discuss spectrum occupancy databases for CBRS. To illustrate the practical feasibility of 1 GHz bandwidth sensing, authors in [4] provide a proof of concept. More recently, some efforts have been dedicated to applying machine learning and compressive sampling theories for spectrum sensing [6,25,34,35]. For instance, authors in [34] propose a framework called Rxminer that relies on mixed Gaussian and Rayleigh models to identify the spectrum occupancy. Also, authors in [36] combine between compressive sensing theory and cyclostationarity pf signals to estimate the wideband spectrum signal parameters. Their technique outperforms the energy detection-based methods in low SNR regime. In [6], authors exploit the low-rank property of the measurement matrix to recover unreported measurements in the context of collaborative spectrum sensing. This modeling fails to capture frequency reuse, a property of high-frequency bands. More recently, authors extended this approach to detect malicious users in [24]. Authors in [37] propose a joint tensor completion and prediction scheme that combines prediction models with a tensor completion algorithm to retrieve the incomplete measurements. However, to the best of our knowledge none of the related works considered the wideband spectrum sensing from a minimal number of measurements per SN and exploited the heterogeneous spectrum occupancy the way we did.

Collaborative filtering. To predict consumers choices, collaborative filtering was introduced in recommendation systems by relying on the similarity between users and ratings. The two main

challenges for collaborative systems are sparsity and scalability. It uses matrix factorization to address scalability issues [38], which stems from the fact that the users' preference for a particular item is only controlled by a small number or a subset of latent factors, which translates to a low-rank rating matrix. However, this assumption does not hold in real-world applications as investigated in [31], which shows that when the global matrix can be split into several low-rank sub-matrices, more splendid performances can be achieved. The main shortcoming of this proposed approach is the construction of the sub-matrices, which is done by first randomly selecting a number of anchor points, and then, by using distance metrics, points are connected to their closest anchor points. Besides, it suffers from high computation and storage costs. Recently, authors in [32] have proposed SLOMA to overcome previous weaknesses by incorporating the social connections among users. We tailored this approach to make efficient wideband spectrum sensing in a spectrum database.

3. Background and system model

3.1. Wideband spectrum occupancy model

We consider a heterogeneous wideband spectrum sensing (WSS) system with $\mathcal{I}$ frequency bands. Since, in practice, the spectrum accommodates different types of applications, we assume the spectrum is divided into multiple blocks, with each block containing multiple frequency bands that are allocated to the same application. Formally, the $\mathcal{I}$ narrow bands are grouped into $\mathcal{L}$ disjoint contiguous blocks, with each block, $\mathcal{G}_n$, consisting of $\mathcal{L}_n$ contiguous bands with $n = 1 \dots \mathcal{L}$, being assigned to one application type as illustrated in Fig. 1. We model the state of each band i using a Bernoulli model with parameter $p_i \in [0, 1]$ where p_i represents the probability that some primary user (PU) occupies the band i . For simplicity, we assume that a PU cannot occupy more than one band and the band occupancy is independent. Hence, the estimated number of occupied bands within a block $\mathcal{G}_n$ equals to the sum of states of all the bands p_i within that block, $\bar{\kappa}_n = \sum_{i \in \mathcal{G}_n} p_i$. According to real measurement studies [39], the band occupancy statistics (e.g., $\bar{\kappa}_n$) vary from one block to another; that is, the spectrum occupancy is heterogeneous with average occupancies varying significantly from one block to another.

3.2. Spectrum occupancy database model

A database-driven spectrum access system has two main components: the spectrum database, which contains the spectrum occupancy information or map, and multiple SNs, which are to be deployed in the region to perform sensing and report their measurements to the database. We assume that the wideband spectrum of interest is very wide and that each SN can only sense $g \ll \mathcal{L}$ blocks among the $\mathcal{L}$ blocks of the wideband spectrum containing $\mathcal{N}$ bands using sub-Nyquist sampling. We assume the average number of occupied bands in the g blocks to be $\bar{\kappa}$.

The time-domain signal, $r_j(t)$, received by a SN j can be expressed as

$$\begin{aligned} r_j(t) &= \sum_{i=1}^{\kappa(t)} h_{ij}(t) \otimes s_i(t) + w_j(t) \\ &= h_j(t) \otimes s(t) + w_j(t), \end{aligned} \quad (1)$$

where $h_{ij}(t)$ is the channel impulse response between the i th PU and the j th SN, $\mathbf{h}_j(t) = [h_{ij}(t)]_{i=1}^{\kappa(t)}$, $s_i(t)$ is the i th PU's signal with zero mean and power P , $\mathbf{s}(t) = [s_i(t)]_{i=1}^{\kappa(t)}$, $w_j(t)$ is an Additive White Gaussian Noise with variance $\mathcal{N} \times N_0$ (assuming a normalized bandwidth; N_0 is the noise variance), $\otimes$ is the convolution

operator, and $\kappa(t)$ is the number of instantaneous active PUs within the sensed g blocks; for simplicity $\kappa(t)$ is assumed to be equal to the number of occupied bands.

Consider $r_j[p] = r_j(t)|_{t=pT_0}$ with $p = 1, \dots, \sum_{k=z}^{z+g-1} \mathcal{L}_k$ and T_0 is the Nyquist sampling period and z is the first block of the g contiguous frequency blocks sensed by SN j . The discrete Fourier transform of SN j 's received signal can be expressed as $R_j = H_j S + W_j = \mathbf{x}_j + W_j$, where H_j , S , and W_j are the Fourier transforms of $\mathbf{h}_j(t)$, $\mathbf{s}(t)$, and $w_j(t)$, respectively. The vector $\mathbf{x}_j$ represents the faded version of the PUs' signals being sent on the different bands. From the CS theory, it follows that the compressed measurements taken by SN j are [15]:

$$\mathbf{y}_j = \Phi F^{-1}(\mathbf{x}_j + W_j) = \Psi \mathbf{x}_j + \boldsymbol{\eta}_j, \quad (2)$$

where $\mathbf{y}_j \in \mathbb{R}^{\mathcal{M}}$ is the measurement vector, F^{-1} is the inverse discrete Fourier transform, and Φ is the $\mathcal{M} \times \mathcal{N}$ sensing matrix assumed to be full rank, i.e. $\text{rank}(\Phi) = \mathcal{M}$. Unlike previous works [6,24], we exploit the heterogeneity feature of wideband spectrum occupancy by considering a non-uniform sensing matrix, Φ , whose coefficients are drawn from a nonuniform Bernoulli distribution as follows. Each coefficient in the n th block of Φ is $\{\frac{\pm \bar{\kappa}_n}{\sqrt{\mathcal{M}}}\}$. The sensing noise $\boldsymbol{\eta}_j$ is defined as $\Phi F^{-1} W_j$. From a practical viewpoint, one approach of acquiring the average occupancy of each block, $\bar{\kappa}_n$, is by monitoring the occupancy of each band within the block and averaging them over time, as already been proposed in [40,41]. Other machine learning-based prediction approaches can also be used to provide good estimates of the average occupancy.

4. Wideband spectrum occupancy recovery through cooperative compressed sensing

4.1. Limitations of conventional recovery approaches

Broadly speaking, CS-based signal reconstruction approaches that can be exploited to recover the spectrum occupancy vector $\mathbf{x}$ from the measurement vector $\mathbf{y}$ (Eq. (2)) (subscript j is omitted for simplicity) can be divided into two categories: (i) greedy algorithm approaches, such as Matching Pursuit (MP), and Orthogonal Matching Pursuit (OMP) [42], which are fast, easy to implement, and have a low computational cost, but not as much as accurate, and (ii) convex optimization approaches such as Basis Pursuit (BP) [43] and LASSO which has a better accuracy and robustness, but requires more computational cost. A well known example of the convex optimization class is ℓ_1 -minimization [44,45], which recovers the occupancy decision vector $\mathbf{x}$ by solving

$$\mathcal{P} : \min_z \|\mathbf{z}\|_{\ell_1} \quad \text{s.t.} \quad \|\Psi \mathbf{z} - \mathbf{y}\|_{\ell_2} \leq \epsilon \quad (3)$$

where ϵ is a pre-specified error threshold parameter.

Recall that the number of required measurements for enabling successful spectrum occupancy recovery using the CS-based sensing approaches is $\mathcal{M} = \mathcal{O}(\kappa \log(\mathcal{N}/\kappa))$ [17,46] or a recent tighter bound as in [47] or [48], which is a function of the total number of bands, $\mathcal{N}$, and the sparsity level of spectrum, κ . This gives rise to key challenges, which we illustrate next.

- **Challenge 1: Hardware limitation.** Existing CS-based hardware architectures consist of physical hardware branches correspond to each required incoherent measurement of $\mathcal{M}$. The relationship mentioned above between the number of required incoherent measurements (hardware branches), total number of bands, and sparsity level shows that the number of hardware branches required for CS-based recovery approaches is still high and impractical. For example, for a wideband with a total number of bands $\mathcal{N} = 100$ with a low number of active bands $\kappa = 8$, the number

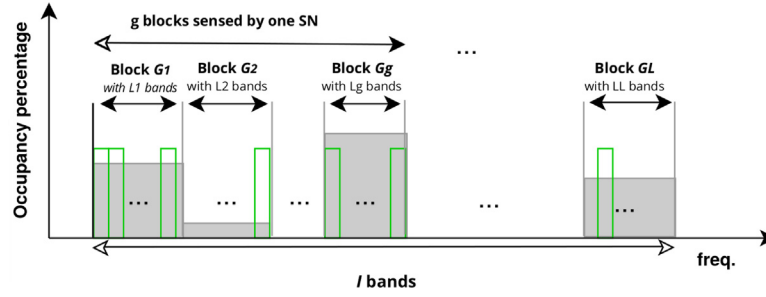


Fig. 1. Wideband spectrum occupancy model. Only small number of bands (the green ones) are occupied at each sensing period.

of required incoherent measurements (hardware branches) still can be as high as $\mathcal{M} = 26$ [17]. In practice, however, the number of branches that reasonable receiver designs have is typically in the order of 4 to 8 [16], which is still much lower than the required number of hardware branches, $\mathcal{M}$, in the CS-based approaches. Hardware challenges limit CS-based approaches' ability to accurately recover a wide range of bands in the entire spectrum, and hence, each SN can only operate on a small portion of the spectrum.

- **Challenge 2: Uncertain and time-varying sparsity.** Most CS approaches assume that the sparsity level, $\mathcal{K}$, is known to the receiver beforehand and fixed, which equals to the overall estimated spectrum sparsity such as in the works [4, 49] and the references therein or over-estimated [50]. In practice, however, a receiver does not have prior information about the sparsity level of a spectrum, and it depends on the current state of the channels, which is time-variant. Therefore, the time variability of the wideband occupancy sparsity and the lack of a priori knowledge makes existing approaches either non-reliable or creating an unnecessarily high overhead.

To address the above two challenges, cooperative spectrum sensing approaches have several advantages as they are, in general, expected to have a higher detection rate and overall performance. Cooperative spectrum sensing approaches are proposed to share the local spectrum availability information of one SN to other SNs, often located on different geographic locations. This property allows the cooperative approaches to reduce the information acquisition and the computation recourse costs since they scan only a portion of the spectrum and then share it with other SNs, which do the same thing to provide the global spectrum information for every close-by SN. Hence, we avoid the redundant scanning of the same area from different SNs and eliminate the unnecessary costs. Nevertheless, wireless channels' fading nature introduces a pivotal challenge to cooperative sensing approaches that will be discussed in challenge 3.

- **Challenge 3: Inconsistent observations.** In practice, different SNs may observe different spectrum occupancy due to wireless channel impairments (e.g., fading, shadowing, etc.), leading to inconsistent measurements across the different SNs. This presents a challenge when it comes to using CS-based sensed measurements to collaboratively recover the spectrum occupancy information because the fusion center will receive contradicting measurement reports about the same portion of the spectrum.

4.2. The proposed wideband spectrum recovery approach

Fading channels generate inconsistency in the measurement reports from the cooperative SNs in the same vicinity. Each SN j reports a different spectrum occupancy vector to the fusion center. However, most of the reports share the same support, the

occupied bands, of the (nearly) sparse occupancy vector. Hence, to be able to detect the support Ω , we propose to compute, for each SN j , the contribution $\xi_{j,n}$ of every column of SN j 's sensing matrix, Ψ_j , to $\mathbf{y}_j$ on each band n ; i.e., $\xi_{j,n} = \langle \mathbf{y}_j, \psi_{j,n} \rangle^2 = (\mathbf{y}_j^T \psi_{j,n})^2$ for $n = 1..N$. For this, we define the sample mean ξ_n as

$$\xi_n = \frac{1}{\mathcal{J}} \sum_{j=1}^{\mathcal{J}} \xi_{j,n} = \frac{1}{\mathcal{J}} \sum_{j=1}^{\mathcal{J}} \langle \mathbf{y}_j, \psi_{j,n} \rangle^2 \text{ for } n = 1..N \quad (4)$$

Once ξ_n is computed, the indices corresponding to the $\mathcal{K}$ highest values among the N statistics are selected iteratively. The selected indices represent the support of the spectrum occupancy. Although inspired by the approach proposed in [51], our proposed recovery approach differs in the following aspects: in our work, (i) we model the signals at the occupied bands as mixed Gaussian and Rayleigh distributions instead of just Gaussian signals, while modeling the unoccupied bands as Gaussian signals with a zero mean and variance N_0 . The signal distribution on the occupied bands is affected by the distance between each SU and the active PU. (ii) we model the sensing matrices as non-uniform Bernoulli distribution, with a mean of 0 and variance $\frac{1}{\omega_i^2}$ with $\omega_i = 1/\tilde{\mathcal{K}}_n$ for each column i if i belongs to the n th block; and (iii) the columns of the sensing matrices are highly correlated due to the low number of measurements, $\mathcal{M}$ in each matrix. Due to a large number of channels, N , it is hard to maintain the orthogonality between columns. Our iterative method for recovering the spectrum support of each SN is presented in Algorithm 1. Recall that the goal is to recover the support vector's indices, not their actual signal values.

Algorithm 1: Spectrum occupancy recovery

Input : $\mathbf{y}_j, \Psi_j, \mathbf{r}_{j,0} = \mathbf{y}_j, j = 1..\mathcal{J}, k = 1$

1 begin

2 while $\|\mathbf{r}_{j,k}\|_{\ell_2} \geq \epsilon \|\mathbf{y}_j\|_{\ell_2}, j = 1..\mathcal{J}$ **do**

3 $n_k = \arg \max_{n \in \{1..N\}} \frac{1}{\mathcal{J}} \sum_{j=1}^{\mathcal{J}} |\langle \mathbf{r}_{j,k-1}, \psi_{j,n} \rangle|^2$

4 $\Omega = \Omega \cup \{n_k\}$

5 $\mathbf{r}_{j,k} = \mathbf{r}_{j,k-1} - \frac{\langle \mathbf{r}_{j,k-1}, \psi_{j,n_k} \rangle}{\|\psi_{j,n_k}\|_{\ell_2}^2} \psi_{j,n_k}$

6 $k = k + 1$

7 return Ω_i

Now that we presented our proposed algorithm, which leverages cooperation to lower the number of required measurements per SN for recovering the support (non-empty channels) of the wideband spectrum, the algorithm's correctness is examined in the following section.

4.3. Correctness of the proposed spectrum recovery approach

The following theorem states the support set, Ω , is almost guaranteed to be covered from only a small number of measurements per SN, when we consider a fairly large number of SNs.

Theorem 1. Consider $\mathcal{J}$ SNs, and let the measurement matrix Ψ_j of SN j contains independent Bernoulli elements, with column i 's elements being set to $\{\frac{\pm 1}{\omega_i}\}$. The vector $\mathbf{x}$ is nearly sparse such that $\mathbf{x}_\ell$ is i.i.d. Gaussian with zero mean and variance N_0 if $\ell \notin \Omega$ and zero mean and variance $\mathbb{E}(x_\ell^2) > N_0$ if $\ell \in \Omega$. With $\mathcal{M} > 1$ measurements per SN, Algorithm 1 recovers Ω with a probability approaching one as $\mathcal{J} \rightarrow \infty$.

Remark 1. Recall that we propose a non-uniformly Bernoulli distributed sensing matrix, and hence, we can improve the detection probability by exploiting any prior knowledge about the spectrum occupancy statistics.

Proof. The proof is based on Kolmogorov's Strong Law of Large Numbers (SLLN) [52], following the same line of argument as in [51]. The main point is to show that the values of $\xi_{j,n}$ of a band n are adequately distinguishable when it is occupied and when it is not. SLLN [52] states that the sample mean $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ of n independent random variables, $X_1, X_2, \dots, X_n$, with finite expectations ($\mathbb{E}(X_n) < \infty$ for $n \geq 1$) converges almost surely to $\mathbb{E}(X_n)$; i.e., $\mathbb{P}(\lim_{n \rightarrow \infty} \bar{X}_n = \mathbb{E}(X_n)) = 1$, and that SLLN holds if one of the following conditions is satisfied:

- (1) $X_1, X_2, \dots, X_n$ are identically distributed.
- (2) $\text{Var}[X_n] < \infty$ and $\sum_{n=1}^{\infty} \frac{\text{Var}[X_n]}{n^2} < \infty$ for all n .

Considering $\xi_{j,n} = \langle \mathbf{y}_j, \boldsymbol{\psi}_{j,n} \rangle^2$, first we need to prove that the expectations of these $\xi_{j,n}$ are finite. Then, since $\xi_{j,n}$ are not identically distributed (due to the presence of fading), we have to prove the second part of Kolmogorov's theorem. Therefore, we start by computing the mean and variance of $\xi_{j,n}$ for every band n to show that both are finite. Without loss of generality, we will assume the occupied bands reside in the first $\mathcal{K}$ bands while the other bands are empty since they contain energy that is less than the specified threshold. The means and variances are given by the following proposition.

Proposition 1. Consider the n th band. The mean of $\xi_{j,n}$ is

$$\mathbb{E}(\xi_{j,n}) = \begin{cases} \sum_{\ell=1}^{\mathcal{K}} \frac{\mathbb{E}(x_\ell^2) \mathcal{M}}{\omega_\ell^2 \omega_n^2} + \sum_{\ell=\mathcal{K}+1}^{\mathcal{N}} \frac{N_0 \mathcal{M}}{\omega_\ell^2 \omega_n^2} + \frac{N_0 \mathcal{M}^2}{\omega_n^4}, & \text{if } n \notin \Omega \\ \frac{\mathbb{E}(x_n^2) \mathcal{M}^2}{\omega_n^4} + \sum_{\ell=1}^{\mathcal{K}} \frac{\mathbb{E}(x_\ell^2) \mathcal{M}}{\omega_\ell^2 \omega_n^2} + \sum_{\ell=\mathcal{K}+1}^{\mathcal{N}} \frac{N_0 \mathcal{M}}{\omega_\ell^2 \omega_n^2}, & \text{if } n \in \Omega \end{cases} \quad (5)$$

and the variance of $\xi_{j,n}$, $\text{Var}(\xi_{j,n})$, is given by Eq. (12).

To prove Proposition 1, we use the definitions of mean and variance and the following Lemma whose proof follows straightforwardly from the definition.

Lemma 1. Let $\boldsymbol{\psi}_n$ be the n th column of the sensing matrix Ψ whose elements are Bernoulli with zero mean and variance $\frac{1}{\omega_n^2}$. Then, we have the following results.

$$\mathbb{E}(\langle \boldsymbol{\psi}_n, \boldsymbol{\psi}_\ell \rangle^2) = \frac{\mathcal{M}}{\omega_n^2 \omega_\ell^2} \quad (6)$$

$$\mathbb{E}(\langle \boldsymbol{\psi}_n, \boldsymbol{\psi}_\ell \rangle^4) = \frac{\mathcal{M}(3\mathcal{M}-2)}{\omega_n^4 \omega_\ell^4} \quad (7)$$

$$\mathbb{E}(\langle \boldsymbol{\psi}_n, \boldsymbol{\psi}_\ell \rangle^2 \langle \boldsymbol{\psi}_n, \boldsymbol{\psi}_p \rangle^2) = \frac{\mathcal{M}^2}{\omega_n^4 \omega_\ell^2 \omega_p^2} \quad (8)$$

$$\mathbb{E}(\|\boldsymbol{\psi}_\ell\|^4 \langle \boldsymbol{\psi}_n, \boldsymbol{\psi}_\ell \rangle^2) = \frac{\mathcal{M}^3}{\omega_n^2 \omega_\ell^6} \quad (9)$$

$$\mathbb{E}(\|\boldsymbol{\psi}_\ell\|^4) = \frac{\mathcal{M}^2}{\omega_\ell^4} \quad (10)$$

$$\mathbb{E}(\|\boldsymbol{\psi}_\ell\|^8) = \frac{\mathcal{M}^4}{\omega_\ell^8} \quad (11)$$

$$\begin{aligned} \text{Var}(\xi_{j,n}) &= \left\{ \begin{aligned} &\sum_{\ell=1}^{\mathcal{K}} \frac{\mathbb{E}(x_\ell^4) \mathcal{M}(3\mathcal{M}-2)}{\omega_\ell^4 \omega_n^4} + 2 \sum_{\ell=1}^{\mathcal{K}} \sum_{m=1, m \neq \ell}^{\mathcal{K}} \frac{\mathbb{E}(x_\ell^2) \mathbb{E}(x_m^2)}{\omega_\ell^2 \omega_m^2 \omega_n^4} + \frac{N_0^2 \mathcal{M}^4}{\omega_n^8} \\ &+ 6 \left[\sum_{\ell=1}^{\mathcal{K}} \frac{\mathbb{E}(x_\ell^2) \mathcal{M}}{\omega_\ell^2 \omega_n^2} \right] \left[\sum_{\ell=\mathcal{K}+1}^{\mathcal{N}} \frac{N_0 \mathcal{M}}{\omega_\ell^2 \omega_n^2} + \frac{N_0 \mathcal{M}^2}{\omega_n^4} \right] \\ &+ \sum_{\ell=\mathcal{K}+1}^{\mathcal{N}} \frac{N_0^2 \mathcal{M}(3\mathcal{M}-2)}{\omega_\ell^4 \omega_n^4} \\ &+ 2 \sum_{\ell=\mathcal{K}+1}^{\mathcal{N}} \sum_{m=\mathcal{K}+1, m \neq \ell}^{\mathcal{N}} \frac{N_0^2 \mathcal{M}^2}{\omega_\ell^2 \omega_m^2 \omega_n^4} \\ &+ 6 \frac{N_0 \mathcal{M}^2}{\omega_n^4} \left[\sum_{\ell=\mathcal{K}+1}^{\mathcal{N}} \frac{N_0 \mathcal{M}}{\omega_\ell^2 \omega_n^2} + \frac{N_0 \mathcal{M}^2}{\omega_n^4} \right], & \text{if } n \notin \Omega \\ &= \left\{ \begin{aligned} &\sum_{\ell=1, \ell \neq n}^{\mathcal{K}} \frac{\mathbb{E}(x_\ell^4) \mathcal{M}(3\mathcal{M}-2)}{\omega_n^4 \omega_\ell^4} + 2 \sum_{\ell=1, \ell \neq n}^{\mathcal{K}} \sum_{p=1, p \neq \ell}^{\mathcal{K}} \frac{\mathbb{E}(x_\ell^2) \mathbb{E}(x_p^2) \mathcal{M}^2}{\omega_n^4 \omega_\ell^4 \omega_p^2} \\ &+ 6 \left[\sum_{\ell=1, \ell \neq n}^{\mathcal{K}} \frac{\mathbb{E}(x_\ell^2) \mathcal{M}}{\omega_n^2 \omega_\ell^2} + \frac{\mathbb{E}(x_n^2) \mathcal{M}^2}{\omega_n^4} \right] \left[\sum_{\ell=\mathcal{K}+1}^{\mathcal{N}} \frac{\mathbb{E}(x_\ell^2) \mathcal{M}}{\omega_n^2 \omega_\ell^2} \right] \\ &+ 4 \sum_{\ell=1, \ell \neq n}^{\mathcal{K}} \frac{\mathbb{E}(x_\ell^2) \mathbb{E}(x_n^2) \mathcal{M}^3}{\omega_n^6 \omega_\ell^2} + \frac{\mathbb{E}(x_n^4)}{\omega_n^8} \\ &+ \sum_{\ell=\mathcal{K}+1}^{\mathcal{N}} \frac{\mathbb{E}(x_\ell^4) \mathcal{M}(3\mathcal{M}-2)}{\omega_n^4 \omega_\ell^4} \\ &+ 2 \sum_{\ell=\mathcal{K}+1}^{\mathcal{N}} \sum_{p=\mathcal{K}+1, p \neq \ell}^{\mathcal{N}} \frac{\mathbb{E}(x_\ell^2) \mathbb{E}(x_p^2) \mathcal{M}^2}{\omega_n^4 \omega_\ell^2 \omega_p^2}, & \text{if } n \in \Omega \end{aligned} \right. \end{aligned} \right. \quad (12)$$

In order to prove that the expectations of $\xi_{j,n}$ are finite, it is sufficient to show that $\mathbb{E}(x_\ell^2)$ and $\mathbb{E}(x_\ell^4)$ are finite. By exploiting the fact that PUs use finite powers in their transmissions, and $\mathbb{E}(x_\ell^2)$ and $\mathbb{E}(x_\ell^4)$ are upper bounded by the transmit power P and P^2 , we can say that $\mathbb{E}(x_\ell^2)$ and $\mathbb{E}(x_\ell^4)$ are finite which also means that expectations of $\xi_{j,n}$ are finite. $\sum_{j=1}^{\infty} \frac{\text{Var}(\xi_{j,n})}{j^2}$ is finite

(upper bounded by $(\max_j \text{Var}(\xi_{j,n})) \sum_{k=1}^{\infty} \frac{1}{k^2}$) which according to

Kolmogorov's theorem is sufficient to prove that ξ_n almost surely converges to the mean given by Proposition 1. Finally, we have $\frac{1}{\mathcal{J}} \sum_{j=1}^{\mathcal{J}} \xi_{j,n}$ converge to $\mathbb{E}(\xi_{j,n})$ for $n = 1..N$. To finish the proof, we still need to show that the mean of an occupied band n is adequately distinguishable from its mean when it is unoccupied. The clear distinction between the two cases is preserved even in the case of uniformly distributed sensing matrices. Yet, it has more importance in our case, non-uniformly distributed sensing matrices. Fig. 2 shows the ratio of the means of the two aforementioned cases for different SNRs and number of measurements $\mathcal{M}$. ■

4.4. Exploiting user closeness

Theorem 1 confirms that relying on multiple SNs allows to overcome the two challenges related to hardware limitation and

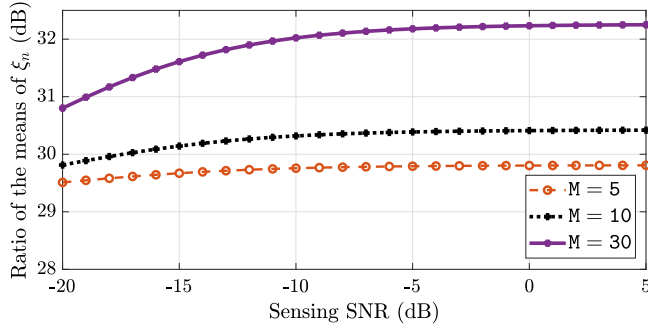


Fig. 2. $\mathbb{E}(\xi_n)/\mathbb{E}(\xi_{n'})$ with n is an occupied band and n' is an unoccupied band for different SNRs: $\mathcal{N} = 256$, $\kappa_i = 29$, weights in the occupied bands $\omega_{in} = 1/\kappa_i$, weights in the unoccupied bands $\omega_{out} = 1$, $N_0 = -120$ dBm.

fading environments. Nevertheless, a large number of SNs is needed to be able to overcome these challenges. However, we can show that the number of required SNs can be largely reduced by leveraging the closeness between SNs. To illustrate this further, consider two SNs with measurement vectors $\mathbf{y}_1 = \Psi_1 \mathbf{x}_1 + \boldsymbol{\eta}_1$ and $\mathbf{y}_2 = \Psi_2 \mathbf{x}_2 + \boldsymbol{\eta}_2$. When the received signals at the SNs are quite similar, say $\mathbf{x}_2 = \mathbf{x}_1 + \delta \mathbf{x}$, $\mathbf{y}_2$ can be expressed as $\mathbf{y}_2 = \Psi_2 \mathbf{x}_1 + \boldsymbol{\eta}_2 + \Psi_2 \delta \mathbf{x}$. This is equivalent to a sequential spectrum sensing where one SN senses the spectrum twice, i.e., $\mathbf{y}_c = [\mathbf{y}_1^T \mathbf{y}_2^T]^T$, $\Psi_c = [\Psi_1^T \Psi_2^T]^T$, and $\boldsymbol{\eta} = [\boldsymbol{\eta}_1^T \boldsymbol{\eta}_2^T + (\Psi_2 \delta \mathbf{x})^T]^T$. With a higher number of measurements, conventional recovery approaches such as LASSO [43] and OMP [42] can be used. Clearly, as one of the signals starts to deviate from the other due to having a higher noise variance, the detection probability of the recovery gradually decreases. We will show the effectiveness of exploiting the closeness of the SNs in the simulations results section.

4.5. Limitations of the proposed recovery approach

So far, we have presented techniques that allow each SN to recover the occupancies of a subset of spectrum blocks within its vicinity. Although these techniques can be useful for individual SNs, they do not provide occupancy information for all blocks at every location. Hence, they cannot be used for constructing wideband spectrum databases, which require acquiring knowledge about spectrum occupancy of all blocks at all locations. To be able to do this, (1) each SN is required to sense and report spectrum occupancies in all bands of the spectrum and (2) each physical location must be covered by some SNs. However, this would incur substantial overhead in terms of the number of measurements needed per SN (and hence the amount of traffic), as well as the sensing delay due to sequential sensing. In addition, this becomes more difficult to realize when considering a large wideband spectrum. This is because large wideband spectrum may contain high-frequency components that are susceptible to high signal attenuations, thus calling for higher numbers of SNs to be able to overcome such high attenuations.

In the next section, we present an efficient approach that exploits the low-rank matrix theory to address these aforementioned limitations.

5. Wideband database construction through local low-rank matrix approximation

Once the vectors $\mathbf{x}_j$ s are recovered by the SNs from the compressed measurements $\mathbf{y}_j$ s, such vectors will be used to construct the spectrum occupancy matrix (spectrum database) $\mathbf{R}$ whose columns correspond to the SNs and rows correspond to the bands. Since each SN senses only a small portion of the wideband spectrum, the occupancy matrix will only be partially filled; i.e., most

of its entries will be missing. For instance, the i th column of the matrix, $\mathbf{R}$, contains the occupancy decisions deduced from $\mathbf{x}_j$ and the rest of the entries are empty. One approach that has been used in the literature to fill in the rest of the entries of $\mathbf{R}$ is the use of collaborative filtering, which is known to work as long as the number of observed decisions in $\mathbf{R}$, Δ , is at least $\xi = O(m^{5/4} r \log m)$ with r is the rank of $\mathbf{R}$ and $m = \max(\mathcal{I}, \mathcal{J})$ [19, Theorem 1.1]. That is, the recovery of the missing entries of $\mathbf{R}$ can be formulated as a convex optimization

$$\mathcal{P}_{\text{rank}} : \min_{\mathbf{X}} \text{rank}(\mathbf{X}) \text{ s.t. } \sum_{(i,j) \in \Delta} (\mathbf{R}_{ij} - \mathbf{X}_{ij})^2 \leq \epsilon \quad (13)$$

or

$$\mathcal{P}_{\text{nuc}} : \min_{\mathbf{X}} \|\mathbf{X}\|_* \text{ s.t. } \sum_{(i,j) \in \Delta} (\mathbf{R}_{ij} - \mathbf{X}_{ij})^2 \leq \epsilon \quad (14)$$

where $\|\cdot\|_*$ is the nuclear norm. Note that the main difference between both approaches is that $\mathcal{P}_{\text{nuc}}$ does not depend on the value of the rank of $\mathbf{R}$.

One of the failing points of this approach is the inconsistent observations of the cooperative SNs that are located in different locations when dealing with relatively high frequencies. In high-frequency bands, the chances that different SNs receiving a totally different spectrum occupancy is very high. Hence, the low-rank matrix property is violated and therefore the aforementioned convex optimization equations are no longer valid. Nevertheless, the low-rank property is still preserved at the sub-matrix levels. Hence, in our framework, we propose an approach that leverages this fact to use the convex optimization equations to construct the sub-matrices and therefore be able to construct the global occupancy matrix. Observations showed that even when the overall spectrum occupancy matrix violates the low-rank property, the sets of the matrix (sub-matrix) representing information from close-by SNs still preserve that property. Each sub-matrix can be considered as a separate matrix that can be efficiently completed/constructed using (14), as described in the next section. The assembly of all sub-matrices can form the global occupancy matrix.

5.1. Spectrum matrix construction

The occupancy matrix of the spectrum can be conceived as a flag matrix filled with zeros and ones, where zero flags represent empty bands while the one flags represent the occupied bands. First, we normalize the matrix entries to have a zero mean value by subtracting 0.5 from each entry of the observed entries in the matrix. The normalization is critical to differentiate between the observed occupancies (part of Δ) and the entries to be recovered (containing zeros) using the low-rank matrix theory. Depending on how many sub-matrices are to be considered, we determine the number of anchor points, q , to be deployed. Now, the spectrum observations of SNs within a given distance from each anchor point are arranged in each sub-matrix. This distance is determined by considering the range of detection for the highest carrier frequency. The width of each sub-region is decided based on the detection range of the highest carrier frequency, which can be estimated using practical propagation models for high frequencies [20]. Overlapping between the sub-regions is desired to help decide on the occupancy of the SNs in the sub-regions boundaries.

5.1.1. Local low-rank spectrum sub-matrices recovery

The occupancy of each spectrum sub-matrix, $\mathcal{M}^k$, for $k = 1, \dots, q$ is the solution to the optimization problem

$$\mathcal{P} : \min_{\mathbf{X}} \|\mathbf{X}\|_* \text{ s.t. } \sum_{(i,j) \in \Delta^k} (\mathbf{O}_{ij}^k - \mathbf{X}_{ij})^2 \leq \epsilon \quad (15)$$

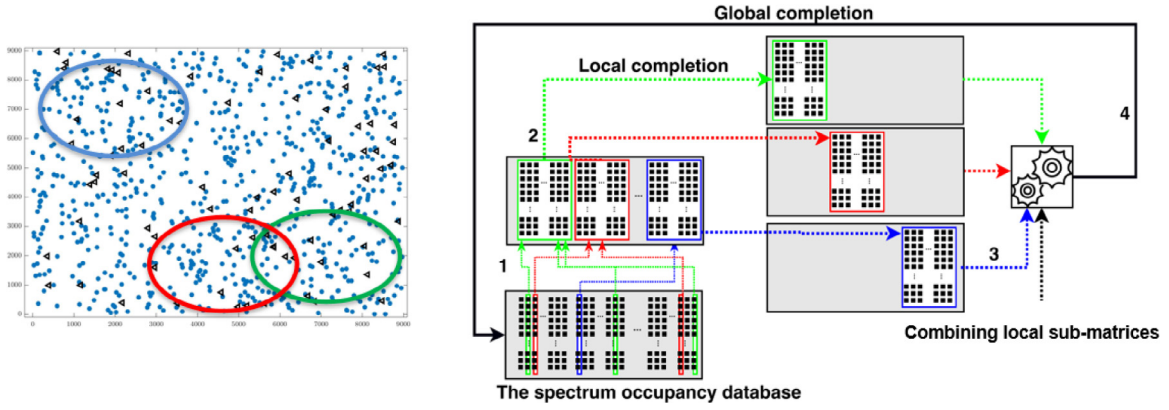


Fig. 3. The different steps of the local low-rank matrix based recovery. (1) Spectrum sub-matrices construction, (2) Local low-rank matrix completion of each sub-matrix, (3) and (4) Global matrix completion.

where Δ^k is a subset of Δ containing observations used to complete the matrix $\mathcal{M}^k$. Note that $\mathcal{P}$ is similar to (14) except that it does not consider the entire observed matrix spectrum occupancies Δ .

5.1.2. Global recovery via weighted decisions

After completing all the sub-matrices, a global decision is made by assembling all of them into their original global matrix. This is illustrated in Fig. 3. Since the recovered spectrum decision of each SN depends on the neighborhood the SN is situated at, the final spectrum decision is made by combining all the decisions made by neighboring nodes. The entries of the global spectrum occupancy matrix $\hat{\mathbf{R}}$ is then expressed as

$$\hat{R}_{ij} = \sum_{k=1}^q \frac{\mathcal{K}_{ij}^k}{\sum_{s=1}^q \mathcal{K}_{ij}^s} \mathcal{M}_{ij}^k \quad (16)$$

where $\mathcal{K}_{ij}^k$ is a kernel function that depends on the distance, d_{ik} , between SN i and an anchor point $\mathbf{c}_k$. We adopt the following similarity function (Kernel)

$$\mathcal{K}_{ij}^k(d_{ik}) = \begin{cases} 1, & \text{if } d_{ik} < d^{th} \\ e^{-\beta d_{ik}}, & \text{otherwise} \end{cases} \quad (17)$$

where d^{th} is the maximum distance over which the similarity is preserved, and β is a decay parameter. This similarity function preserves a unity gain within the range of the distance threshold d^{th} , and as the SNs get further, the similarity diminishes exponentially to zero.

Finally, the sign of each entry of the matrix $\mathbf{R}$ represents the final binary matrix.

5.2. Computational and communication overhead reduction

Conventionally, the occupancy decisions of spectrum in the vicinity of a SN incurs a communication overhead that is linear in $\mathcal{I}$ and $\mathcal{J}$. When using compressive sensing without low-rank matrix recovery, the incurred communication overhead is linear in $\mathcal{J}$, the number of compressed samples $\mathcal{M}$, and $\lfloor \mathcal{I}/\mathcal{N} \rfloor$. Our proposed framework manages to reduce the complexity by eliminating the linearly dependent overhead on $\lfloor \mathcal{I}/\mathcal{N} \rfloor$. Hence, an overall network overhead reduction and lesser reporting energy are achieved with our proposed scheme. Specifically, the global occupancy matrix recovery complexity equals to q times the complexity of the optimization problem in (15), or even smaller since this can be executed in parallel.

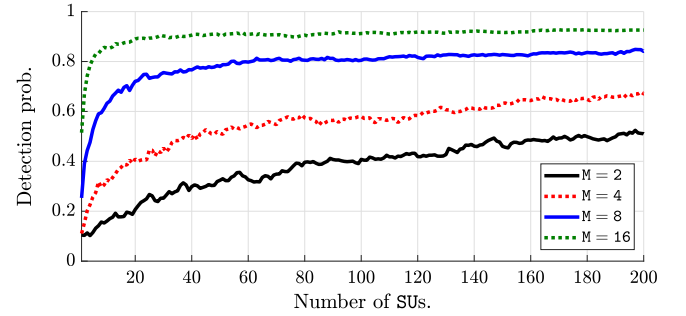


Fig. 4. The detection probability for $N = 128$.

6. Performance evaluation results

We start our performance evaluation by the proposed spectrum occupancy recovery approach discussed in Section 4. We consider a primary system operating over a wideband consisting of $\mathcal{N} = 128$ bands grouped into $g = 4$ blocks with equal sizes. The average probabilities of occupancy in each block are as follows: $\bar{\mathcal{K}}_1 = p_1 \times 32$, $\bar{\mathcal{K}}_2 = p_2 \times 32$, $\bar{\mathcal{K}}_3 = p_3 \times 32$, $\bar{\mathcal{K}}_4 = p_4 \times 32$, where $p_1 = p_3 = 0.1$ and $p_2 = p_4 = 0.001$. The PUs are randomly deployed in a cell. We assume all PUs are transmitting with a constant power $P = 10$ W. We also model the spectrum channels as a Rayleigh distributed channels with mean $1/d^{3/2}$. We model the noise in the unoccupied bands as Gaussian with zero mean and variance $N_0 = -120$ dBm.

Fig. 4 shows the detection probability as a function of the number of cooperating SNs, $\mathcal{J}$. First, we observe that as the number of cooperating SNs increases, a high detection probability is achieved regardless of the number of measurements each SN is taking, thus confirming our main theorem result. This is mainly because as $\mathcal{J}$ increases, $\xi_{j,n}$ converges to its expectation $\mathbb{E}(\xi_{j,n})$, and hence, a better distinction between the bands is achieved. Second, we also observe that a high detection probability is achieved by considering a higher number of measurements, $\mathcal{M}$, for the same number of sensing nodes, $\mathcal{J}$. We notice that with $\mathcal{M} = 16$, only a small number of SNs (4 or 5) is needed to get a very high detection probability. If SNs cannot afford this number of measurements, then it can be reduced with the price of increasing the number of SNs.

Instead of requiring a large number of SNs at different geographic locations, we rather consider to exploit SNs' closeness to each other to reduce the required number of SNs. We examine the potential of this method by applying OMP and LASSO on the data captured by 6 close-by SNs. Fig. 5 clearly shows that the detection

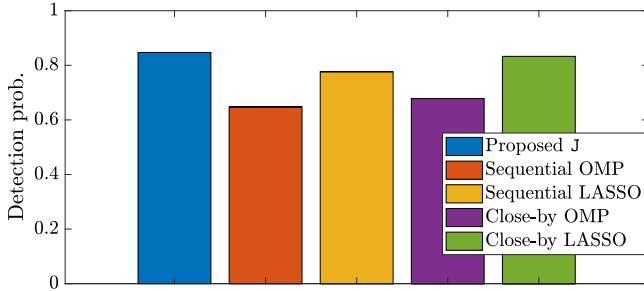


Fig. 5. The detection probability for $M = 8$ and $N = 128$.

rate of our method is very comparable with the results obtained from the approach that requires a high number of SNs, which confirms our hypothesis. Furthermore, our proposed approaches outperform the sequential sensing approach proposed in [16], mainly because of their ability to overcome the hidden terminal problem due to signal propagation decay.

Now that we validated the spectrum sensing techniques, we examine the efficiency of the local low rank spectrum occupancy matrix completion that aims at reducing the network overhead. We assume the presence of multiple PUs operating in some of $\mathcal{I} = 250$ bands (this can be in the 5 – 15 GHz range with 20 MHz bandwidth each). The deployment of the active users follows a Poisson Point Process (PPP) with a density of $2/\text{Km}^2$ deployed in the 2D plane. To mimic real-world scenarios, we assume that high-frequency bands are reused more frequently than low-frequency bands. We also assume that each SN senses $1/5$ of the bands. Sub-matrices are defined based on how far a signal sent over a frequency f_c . We adopted the 3GPP TR 38.901 UMa LOS path loss model [20] given by

$$PL_{dB} = 32.4 + 20 \log_{10}(d(m)) + 30 \log_{10}(f_c(\text{GHz})) \quad (18)$$

for $0.5 < f_c < 100$ GHz and the shadow fading standard deviation equal to 7.8 dB. We consider the sensitivity to be -120 dBm, below which a signal is considered absent. The SNs are deployed according to a uniform PPP with density $10/\text{Km}^2$ deployed in the 2D plane and are linked to the closest anchor point forming the sub-matrices.

To assess the performance of our scheme, we generate the entire spectrum occupancy matrix to compare the final recovery matrix with it. Since our focus is on completing the construction of the spectrum occupancy matrix, we consider the wideband spectrum recovery of the observed portion from each SN to be error free. The spectrum sub-matrix completion is done using [53].

First, we observed from the generated spectrum occupancy matrix that the low-rank property for the sub-matrices is confirmed while the global matrix has no low-rank property ($\text{rank} > 50$ for the case of having 250 bands).

Fig. 6 shows the recovery error (computed as the Frobenius norm) as a function of the number of frequency bands. First, observe that our proposed framework allows achieving a high reduction gain in error (about 10X gain) compared to the classical approach. This is thanks to the observation of the local low rank property (confirmed through simulations). Second, we observe that the recovery error decays as we increase the number of frequency bands for both the proposed and classical methods with a steeper fall in the proposed approach. This trend is because the property of global low-rank is strengthened as we increase the number of frequency bands, and therefore, the recovery error is decreased.

Fig. 7 studies the effect of the number of anchor points. Overall, we observe that as the number of anchor points increases, a reduction in the error is achieved, which confirms the same observation made in Fig. 6.

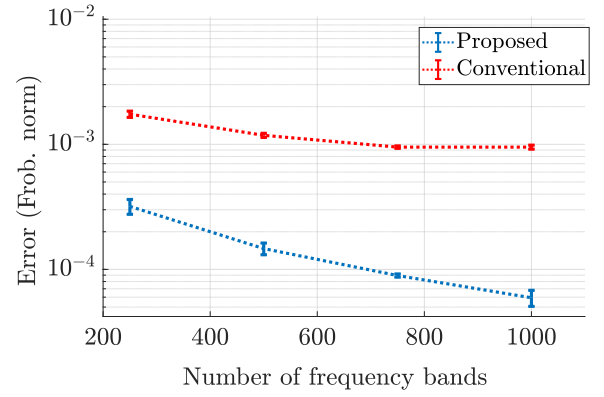


Fig. 6. Error: proposed approach vs traditional approach.

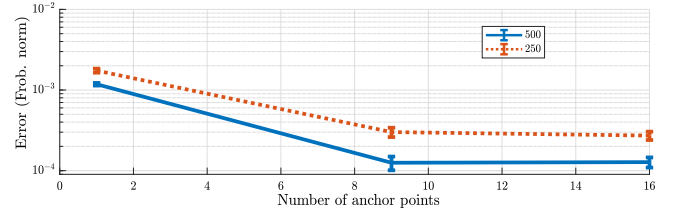


Fig. 7. Effect of the number of submatrices.

7. Conclusions

A framework that builds an accurate spectrum occupancy map for wideband spectrum sharing is proposed. We exploited user cooperation to cope with the SNs' hardware limitations, the time variability in the wideband spectrum occupancy, and the network overhead to improve sensing for each block of the spectrum. Also, we discussed and showed the potential of distributed compressive sampling-based spectrum sensing to overcome sensing overhead by reducing the number of measurements an SU needs to take. Moreover, we showed that the impact of fading could be overcome by considering close-by SNs and exploited the spatial correlation between sensing nodes to achieve scalable decisions for the spectrum occupancy at minimum communication overhead.

CRedit authorship contribution statement

Bassem Khalfi: Conceptualization, Methodology, Software, Formal analysis, Writing - original draft. **Bechir Hamdaoui:** Conceptualization, Validation, Writing - review & editing, Funding acquisition. **Mohsen Guizani:** Conceptualization, Validation. **Abdurrahman Elmaghub:** Software, Writing - review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

This work was supported in part by the US National Science Foundation, USA under NSF awards No. 1162296 and No. 1923884.

References

- [1] B. Khalfi, A. Elmaghub, H. Bechir, Distributed wideband sensing for faded dynamic spectrum access with changing occupancy, in: Global Telecommunications Conference, IEEE GLOBECOM, IEEE, 2018, pp. 1–6.
- [2] B. Khalfi, B. Hamdaoui, M. Guizani, AIRMAP: Scalable spectrum occupancy recovery using local low-rank matrix approximation, in: Global Telecommunications Conference, IEEE GLOBECOM, IEEE, 2018, pp. 1–6.
- [3] Q. Zhao, B.M. Sadler, A survey of dynamic spectrum access, *IEEE Signal Process. Mag.* 24 (3) (2007) 79–89.
- [4] T. Haque, R.T. Yazicigil, K.J.-L. Pan, J. Wright, P.R. Kinget, Theory and design of a quadrature analog-to-information converter for energy-efficient wideband spectrum sensing, *IEEE Trans. Circuits Syst. I. Regul. Pap.* 62 (2) (2015) 527–535.
- [5] B. Khalfi, B. Hamdaoui, M. Guizani, N. Zorba, Exploiting wideband spectrum occupancy heterogeneity for weighted compressive spectrum sensing, in: Proc. of IEEE INFOCOM WKSHPS, 2017, pp. 1–6.
- [6] J.J. Meng, W. Yin, H. Li, E. Hossain, Z. Han, Collaborative spectrum sensing from sparse observations in cognitive radio networks, *IEEE J. Sel. Areas Commun.* 29 (2) (2011) 327–337.
- [7] Z. Qin, Y. Gao, M.D. Plumbley, 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.* 64 (12) (2016) 3106–3117.
- [8] R. Murty, R. Chandra, T. Moscibroda, P. Bahl, Senseless: A database-driven white spaces network, *IEEE Trans. Mob. Comput.* 11 (2) (2012) 189–203.
- [9] M.M. Sohel, M. Yao, T. Yang, J.H. Reed, Spectrum access system for the citizen broadband radio service, *IEEE Commun. Mag.* 53 (7) (2015) 18–25.
- [10] M. Grissa, A.A. Yavuz, B. Hamdaoui, When the hammer meets the nail: Multi-server PIR for database-driven CRN with location privacy assurance, 2017, arXiv preprint arXiv:1705.01085.
- [11] A. Mancuso, S. Probasco, B. Patil, Protocol to access white-space (PAWS) databases: Use cases and requirements, RFC 6953 (2013) 1–23.
- [12] I. Google, Google Spectrum Database, <https://www.google.com/get/spectrumdatabase/>.
- [13] RadioSoft, White Space Database, <http://www.radiosoft.com/>.
- [14] Federal Communications Commission, TV White Space, <https://www.fcc.gov/general/white-space-database-administration>.
- [15] B. Hamdaoui, B. Khalfi, M. Guizani, Compressed wideband spectrum sensing: Concept, challenges and enablers, *IEEE Commun. Mag.* 56 (4) (2018).
- [16] R.T. Yazicigil, T. Haque, M. Kumar, J. Yuan, J. Wright, P.R. Kinget, How to make analog-to-information converters work in dynamic spectrum environments with changing sparsity conditions, *IEEE Trans. Circuits Syst. I. Regul. Pap.* 24 (5) (2018) 1–10.
- [17] R.T. Yazicigil, T. Haque, M.R. Whalen, J. Yuan, J. Wright, P.R. Kinget, Wideband rapid interferer detector exploiting compressed sampling with a quadrature analog-to-information converter, *IEEE J. Solid-State Circuits* 50 (12) (2015) 3047–3064.
- [18] R.T. Yazicigil, T. Haque, P.R. Kinget, J. Wright, Taking compressive sensing to the hardware level: Breaking fundamental radio-frequency hardware performance tradeoffs, *IEEE Signal Process. Mag.* 36 (2) (2019) 81–100.
- [19] E.J. Candès, B. Recht, Exact matrix completion via convex optimization, *Found. Comput. Math.* 9 (6) (2009) 717.
- [20] T.S. Rappaport, Y. Xing, G.R. MacCartney, A.F. Molisch, E. Mellios, J. Zhang, Overview of millimeter wave communications for fifth-generation (5G) wireless networks—with a focus on propagation models, *IEEE Trans. Antennas and Propagation* (2017).
- [21] G. Aswathy, K. Gopakumar, Sub-Nyquist wideband spectrum sensing techniques for cognitive radio: A review and proposed techniques, *AEU-Int. J. Electron. Commun.* 104 (2019) 44–57.
- [22] A. Mariani, A. Giorgetti, M. Chiani, On oversampling-based signal detection, *Int. J. Wirel. Inf. Netw.* 26 (4) (2019) 272–284.
- [23] P. Paysarvi-Hoseini, N.C. Beaulieu, Optimal wideband spectrum sensing framework for cognitive radio systems, *IEEE Trans. Signal Process.* 59 (3) (2010) 1170–1182.
- [24] Z. Qin, Y. Gao, M. Plumbley, Malicious user detection based on low-rank matrix completion in wideband spectrum sensing, *IEEE Trans. Signal Process.* (2017).
- [25] L. Shi, P. Bahl, D. Katabi, Beyond sensing: multi-GHz realtime spectrum analytics, in: Proc. of USENIX Symposium on Networked Systems Design and Implementation, NSDI, 2015, pp. 159–172.
- [26] A. Chakraborty, M.S. Rahman, H. Gupta, S.R. Das, SpecSense: crowdsensing for efficient querying of spectrum occupancy, in: Proc. of IEEE INFOCOM, 2017, pp. 1–9.
- [27] G. Hattab, D. Cabric, Distributed wideband sensing-based architecture for unlicensed massive IoT communications, *IEEE Trans. Cogn. Commun. Netw.* 5 (3) (2019) 819–834.
- [28] A. Mariani, A. Giorgetti, M. Chiani, Wideband spectrum sensing by model order selection, *IEEE Trans. Wireless Commun.* 12 (14) (2015) 6710–6721.
- [29] H. Sun, A. Nallanathan, C.-X. Wang, Y. Chen, Wideband spectrum sensing for cognitive radio networks: a survey, *IEEE Wirel. Commun.* 20 (2) (2013) 74–81.
- [30] G. Zhang, X. Fu, J. Wang, M. Hong, Coupled block-term tensor decomposition based blind spectrum cartography, in: 2019 53rd Asilomar Conference on Signals, Systems, and Computers, IEEE, 2019, pp. 1644–1648.
- [31] J. Lee, S. Kim, G. Lebanon, Y. Singer, S. Bengio, LLORMA: Local low-rank matrix approximation, *J. Mach. Learn. Res.* 17 (1) (2016) 442–465.
- [32] H. Zhao, Q. Yao, J.T. Kwok, D.L. Lee, Collaborative filtering with social local models, in: Proc. of IEEE Int. Conf. on Data Mining, ICDM, 2017, pp. 1–10.
- [33] G. Ding, J. Wang, Q. Wu, Y.-D. Yao, F. Song, T.A. Tsiftsis, Cellular-base-station-assisted device-to-device communications in tv white space, *IEEE J. Sel. Areas Commun.* 34 (1) (2015) 107–121.
- [34] M. Zheleva, R. Chandra, A. Chowdhery, A. Kapoor, P. Garnett, Txminer: Identifying transmitters in real-world spectrum measurements, in: Proc. of IEEE Int. Symposium on Dynamic Spectrum Access Networks, DySPAN, 2015, pp. 94–105.
- [35] H. Hassanieh, L. Shi, O. Abari, E. Hamed, D. Katabi, Ghz-wide sensing and decoding using the sparse fourier transform, in: Proc. of IEEE INFOCOM, 2014, pp. 2256–2264.
- [36] D. Cohen, Y. Eldar, Sub-Nyquist cyclostationary detection for cognitive radio, *IEEE Trans. Signal Process.* (2017).
- [37] M. Tang, G. Ding, Q. Wu, Z. Xue, T.A. Tsiftsis, A joint tensor completion and prediction scheme for multi-dimensional spectrum map construction, *IEEE Access* 4 (2016) 8044–8052.
- [38] Y. Koren, R. Bell, C. Volinsky, Matrix factorization techniques for recommender systems, *Computer* 42 (8) (2009).
- [39] D. Cohen, S. Tsiper, Y.C. Eldar, Analog-to-digital cognitive radio: Sampling, detection, and hardware, *IEEE Signal Process. Mag.* 35 (1) (2018) 137–166.
- [40] M. Yilmaz, D.G. Kuntalp, A. Fidan, Determination of spectrum utilization profiles for 30 MHz–3 GHz frequency band, in: Proc. of the Int. Conf. on Communications, COMM, 2016, pp. 499–502.
- [41] M. Mehdawi, N. Riley, M. Ammar, A. Fanan, M. Zolfaghari, Spectrum occupancy measurements and lessons learned in the context of cognitive radio, in: Proc. of IEEE Telecommunications Forum Telfor, TELFOR, 2015, pp. 196–199.
- [42] J.A. Tropp, A.C. Gilbert, Signal recovery from random measurements via orthogonal matching pursuit, *IEEE Trans. Inf. Theory* 53 (12) (2007) 4655–4666.
- [43] E.J. Candès, J. Romberg, T. Tao, Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information, *IEEE Trans. Inf. Theory* 52 (2) (2006) 489–509.
- [44] E.J. Candès, J.K. Romberg, T. Tao, Stable signal recovery from incomplete and inaccurate measurements, *Comm. Pure Appl. Math.* 59 (8) (2006) 1207–1223.
- [45] D.L. Donoho, Compressed sensing, *IEEE Trans. Inf. Theory* 52 (4) (2006) 1289–1306.
- [46] M. Mishali, Y.C. Eldar, From theory to practice: Sub-Nyquist sampling of sparse wideband analog signals, *IEEE J. Sel. Top. Sign. Proces.* 4 (2) (2010) 375–391.
- [47] A. Elzanaty, A. Giorgetti, M. Chiani, Limits on sparse data acquisition: RIC analysis of finite Gaussian matrices, *IEEE Trans. Inform. Theory* 65 (3) (2018) 1578–1588.
- [48] S. Foucart, H. Rauhut, An invitation to compressive sensing, in: A Mathematical Introduction To Compressive Sensing, Springer, 2013, pp. 1–39.
- [49] S.K. Sharma, E. Lagunas, S. Chatzinotas, B. Ottersten, Application of compressive sensing in cognitive radio communications: A survey, *IEEE Comm. Surv. Tutor.* 18 (3) (2016) 1838–1860.
- [50] B. Khalfi, B. Hamdaoui, M. Guizani, N. Zorba, Efficient spectrum availability information recovery for wideband DSA networks: A weighted compressive sampling approach, *IEEE Trans. Wirel. Commun.* 17 (4) (2018).
- [51] D. Baron, M.F. Duarte, M.B. Wakin, S. Sarvotham, R.G. Baraniuk, Distributed compressive sensing, 2009, arXiv preprint arXiv:0901.3403.

- [52] V.V. Petrov, Limit theorems of probability theory: sequences of independent random variables, Tech. Rep., Oxford, New York, 1995.
- [53] S. Becker, E. Candes, M. Grant, TFOCS: Templates for first-order conic solvers, version 1.3. 1, 2014.



Bassem Khalfi Bassem Khalfi received his Ph.D. in ECE from Oregon State University, Corvallis, Oregon, USA in 2018. Prior to that he received his BS in telecommunications Engineering from Ecole Supérieure des Communications de Tunis, Carthage University, Ariana, Tunisia in 2012 and MS from National Engineering School of Tunis, Tunis-EL Manar University, Tunisia in 2014. He joined Qualcomm Technologies Inc, CA, USA in 2018 where he is working on designing and implementing 4G and 5G physical layer algorithms. He won the IEEE IWCMC 2017 best paper award. He

has been actively serving in the reviewing process for various IEEE journals, magazines, and conferences. His research focuses on various topics in wireless communication and networks, including dynamic spectrum access and sensing, resource allocation for communication systems and content centric networking.



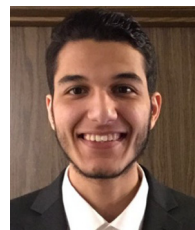
Bechir Hamdaoui (S'02-M'05-SM'12) is a professor in the School of EECS at Oregon State University. He received M.S. degrees in both ECE (2002) and CS (2004), and the Ph.D. degree in ECE (2005), all from the University of Wisconsin-Madison. His research interests are in the general areas of computer networks and wireless communications. He has won several awards, including the ICC 2017 and IWCMC 2017 Best Paper Awards, the 2016 EECS Outstanding Research Award, and the 2009 NSF CAREER Award. He serves/has served as an associate editor for several journals, including

IEEE Transactions on Mobile Computing, IEEE Transactions on Wireless Communications, IEEE Network, and IEEE Transactions on Vehicular Technology. He has also chaired/co-chaired many IEEE conference programs/symposia, including the 2017 INFOCOM Demo/Posters program, the 2016 IEEE GLOBECOM Mobile and Wireless Networks symposium, and many others. He served as a Distinguished Lecturer for the IEEE Communication Society for 2016 and 2017. He is a Senior Member of IEEE.



Mohsen Guizani (S'85-M'89-SM'99-F'09) received the B.S. (with distinction) and M.S. degrees in electrical engineering, the M.S. and Ph.D. degrees in computer engineering from Syracuse University, Syracuse, NY, USA, in 1984, 1986, 1987, and 1990, respectively. He is currently a Professor at the CSE Department in Qatar University, Qatar. Previously, he served in different academic and administrative positions at the University of Idaho, Western Michigan University, University of West Florida, University of Missouri-Kansas City, University of Colorado-Boulder, and Syracuse University.

His research interests include wireless communications and mobile computing, computer networks, mobile cloud computing, security, and smart grid. He is currently the Editor-in Chief of the IEEE Network Magazine, serves on the editorial boards of several international technical journals and the Founder and Editor-in-Chief of Wireless Communications and Mobile Computing journal (Wiley). He is the author of nine books and more than 500 publications in refereed journals and conferences. He guest edited a number of special issues in IEEE journals and magazines. He also served as a member, Chair, and General Chair of a number of international conferences. Throughout his career, he received three teaching awards and four research awards. He also received the 2017 IEEE Communications Society WTC Recognition Award as well as the 2018 AdHoc Technical Committee Recognition Award for his contribution to outstanding research in wireless communications and Ad-Hoc Sensor networks. He was the Chair of the IEEE Communications Society Wireless Technical Committee and the Chair of the TAOS Technical Committee. He served as the IEEE Computer Society Distinguished Speaker and is currently the IEEE ComSoc Distinguished Lecturer. He is a Fellow of IEEE and a Senior Member of ACM.



Abdurrahman Elmaghoub received the B.S. degree from Oregon State University in 2019, and is currently pursuing his Ph.D. degree in the School of Electrical Engineering and Computer Science at Oregon State University. His research interests are in the area of wireless communication and networking with a current focus on applying machine learning to wireless device classification.