Cost-Efficient Indoor White Space Exploration Through Compressive Sensing

Fan Wu, *Member, IEEE,* Dongxin Liu, Zhihao Wu, *Student Member, IEEE,* Yuan Zhang, *Member, IEEE,* and Guihai Chen, *Senior Member, IEEE*

Abstract—Exploring the utilization of white spaces (vacant VHF and UHF TV channels) is a promising way to satisfy the rapid growth of the radio frequency (RF) demand. Although a few outdoor white space exploration methods have been proposed in the past few years, researches that focus on indoor white space exploration just emerged recently. In this paper, we propose a novel cost-efficient indoor white space exploration method by exploiting the location dependence and channel dependence of TV channels' signal strength in indoor environments. We measure the UHF TV channels in a building, and study the temporal and spatial features of indoor white spaces. Based on the extracted features, we design a cost-e**F**ficient **I**ndoor **W**hite space **EX**ploration mechanism, namely FIWEX. Furthermore, we build a prototype of FIWEX and extensively evaluate its performance in real-world environments. The evaluation results show that FIWEX can identify 30.0% more indoor white spaces with 51.2% less false alarms compared with the best known existing solution.

Index Terms—White Space, Indoor, Compressive Sensing

1 Introduction

THE, growth of wireless networks are currently facing an increasing shortage of available radio frequency spectrums, since the number of mobile devices and their related applications is raising rapidly. However, the amount of unlicensed spectrums that are free to use is very limited. To deal with this problem, the concept of Dynamic Spectrum Access (DSA) that aims at exploring the opportunity of sharing and utilization of licensed spectrums among both licensed users and unlicensed users has been proposed and applied to several applications [5], [39].

In 2008, FCC (Federal Communications Commission) issued a historic ruling that allowed unlicensed devices to use the TV spectrum that is not locally occupied by licensed devices. (The unoccupied TV spectrum is often referred to as TV white spaces or simply white spaces). After that, the TV white spaces received more and more attention from DSA developers. Although white spaces are open for unlicensed use now, FCC also required that unlicensed white space devices should not interfere with the licensed devices (TV broadcasts). Therefore, it is essential for all user devices (especially the unlicensed ones) to find out whether a spectrum band is available before using it for communications.

To explore white spaces, there are mainly two approaches: spectrum sensing approach and geo-location database approach. The spectrum sensing approach, which is less widely used, relies on the user devices to perform spectrum sensing. Thus, it requires the devices to be equipped with proper sensing hardware and to have enough power for sensing and signal analysis. On the contrary, the commonly used geo-location database approach does not need the user devices to sense and thus saves their power. A user device gets to know the availability of white spaces by querying an online geo-location database that stores a "white space availability map", which indicates spectrum's availability information at different locations.

For the geo-location database approach, the map of white space availability is vital. In this paper, we study how to efficiently construct a "white space availability map", which is used by the geo-location database approach in indoor environments. Most prior works [4], [7], [19], [27], [40], [43] on white space exploration only focus on outdoor scenarios. Since there is relatively few obstructions in outdoor environments, existing works use signal propagation models to infer whether a spectrum band at a specific location is available or not. However, due to the blocking of the indoor man-made obstacles (e.g., walls), if we directly apply these approaches to indoor environments, it is likely that we would get results that are overly conservative.

Existing works [12], [21] have shown that people stay indoor most of the time, which makes 70% of the spectrum demands coming from the indoor environments. Hence, it is highly necessary to explore the indoor white spaces. To solve the problem of indoor white space exploration, we propose a cost-eFficient Indoor White space EXploration mechanism, called FIWEX. Intuitively, one can always solve the problem by deploying a large number of RF-sensors that cover the entire building. However, this could incur a large amount of expense

D. Liu, Z. Wu, F. Wu, and G. Chen are with the Shanghai Key Laboratory of Scalable Computing and Systems, Department of Computer Science and Engineering, Shanghai Jiao Tong University, China.
Y. Zhang is with the Department of Computer Science and Technology, Nanjing University, China.
E-mails: liu-dx@sjtu.edu.cn, sjtu_wuzhihao@sjtu.edu.cn, fwu@cs.sjtu.edu.cn, zhangyuan@nju.edu.cn, gchen@cs.sjtu.edu.cn
F. Wu is the corresponding author.

on the sensors. By strategically deploying only a limited number of sensors, FIWEX is able to reconstruct the indoor white space availability map at a high accuracy, based on a small number of RF sensors. Due to the innovative utilizations of 1) the existence of *strong channels*, whose signal strength is at least 5dB greater than the white space threshold, 2) the *location dependence* as well as *channel dependence* of channels' signal strengths in indoor environments, 3) the compressive sensing technique, and 4) a well designed *k-medoids* based sensor deployment method, FIWEX features a promising efficiency on the total number of sensors and reconstruction accuracy.

The main contributions of this paper are as follows.

- We perform indoor white space measurements in a building for two weeks. The measurement results confirm that (1) there exist strong channels, (2) indoor white spaces' signal strengths have both location dependence and channel dependence. These two key characteristics allow us to efficiently explore indoor white spaces without making a very dense deployment of RF-sensors over the entire building.
- We propose FIWEX, a cost-efficient indoor white space exploration mechanism that does not require user devices to sense the spectrum by themselves. By taking strong channels, location dependence, and channel dependence into consideration, we propose a compressive sensing based data reconstruction algorithm to reduce the number of sensors needed for detection. In addition, we design an innovative sensor deployment method to improve the reconstruction accuracy.
- We have built a prototype of FIWEX, and evaluated its performance. We evaluate FIWEX's performance compared to WISER, the state-of-the-art indoor white space identification system. On an average, FIWEX can identify 30.0% more indoor white spaces with 51.2% less false alarms.

The remainder of the paper is organized as follows. We introduce our indoor white space measurement experiments in Section 2. Section 3 describes the system model of FIWEX. In Section 4, we provide the detailed system design. Section 5 shows the evaluation results. The related works are discussed in Section 6. At last, we conclude the paper in Section 7.

2 INDOOR WHITE SPACE AVAILABILITY MEASUREMENT

In this section, we introduce our indoor white space measurement. The measurement gives us a deep understanding about the characteristics of indoor white spaces, which can be utilized in designing FIWEX.

2.1 Measurement Setup

The measurement device we used consists of a USRP N210 [1], an log periodic PCB antenna (400-1000 MHz) and a laptop. The daughterboard of our USRP is SBX with 5-10 dBm noise figure. We calibrate the device using a RF signal generator to get the accurate signal strength. As suggested by FCC [2], the gain of the antenna is also taken into consideration during the calibration process.

A number of different methods were proposed for identifying the presence of signal transmissions, such as energy detection, waveform-based sensing, and matched-filtering [42]. Here, we choose energy detection, since it is the most common way with low computation and implementation complexities. In our measurements, we judge whether a TV channel is vacant by comparing the channel's signal strength with a threshold, which depends on the noise floor [34]. If the signal strength of a TV channel is greater than the threshold, we label this channel as locally occupied, otherwise the channel is labeled as vacant. We measure the digital TV channels between 470 MHz - 566 MHZ and 606 MHz - 870 MHz with 8 MHz channel bandwidth, and use the same threshold -84.5 dBm/8 MHz as [41]. Due to the hardware limitations, the vacant channels determined using the above mentioned threshold may be not safe to use, but the observations drawn from the measurement are general, and our mechanism is not limited to any specific threshold. We believe that if the sensitivity of the measurement hardware (e.g., WSA 5000 [3]) can support a threshold of -114 dBm as suggested by FCC [2], our mechanism can be safely used in practice.

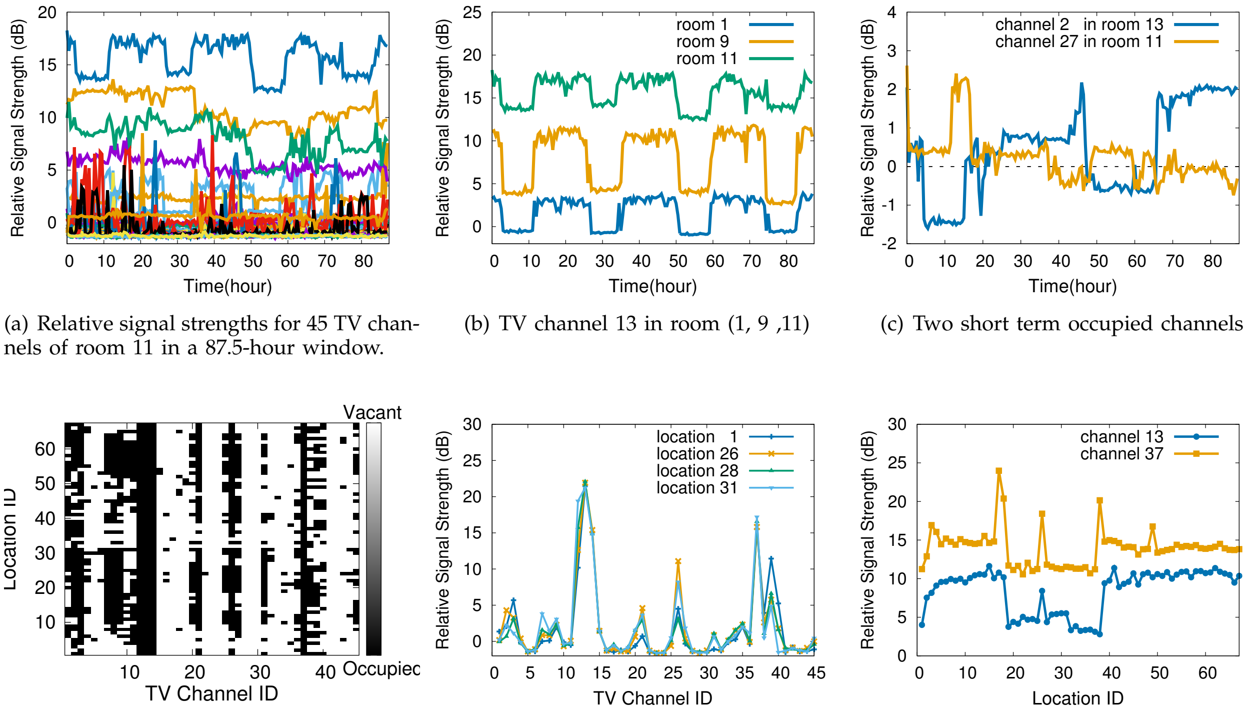
We implement the energy detector based on a GNU Radio FFT program with a bin size of 1024 and sampling rate 4 MHz. We calculate the power of a channel using the average value of all corresponding bins. During the indoor white space measurement experiment, we measure 45 channels in total.

Our measurement is divided into two parts, namely *long-time sensing* and *short-time sensing*.

2.2 Long-Time Sensing

In long-time sensing, we randomly choose 20 rooms and deploy a USRP coupled with a laptop in every room. The absolute signal strengths for all 45 TV channels are measured for contiguous 87.5 hours (20:30, July 7, 2014 - 12:00, July 11, 2014). For the convenience of comparisons, we convert absolute signal strengths to the relative ones by subtracting the white space threshold from them. Devices in different rooms are synchronized using "crontab" command of Unbuntu 12.04. The system time of all laptops are synchronized with the same time server. Then the spectrum sensing programs on our laptops run simultaneously every half an hour.

We have the following observations from long-time sensing results. First, as shown in Fig. 1(a), there exist strong channels whose relative signal strengths are obviously greater than 0 during most of the time. For example, we observe 3 channels with >5dB relative signal strengths during the whole 87.5 hours interval. This means that these three channels are long time occupied



(a) Relative signal strengths for 45 TV channels of room 11 in a 87.5-hour window.

(b) TV channel 13 in room (1, 9 ,11)

(c) Two short term occupied channels

(d) Indoor white space availability map

(e) Location dependence

(f) Channel dependence

Fig. 1. Results of indoor white space measurement: long-time sensing and short-time sensing

in room 11. Actually, these long time occupied channels may be utilized by the licensed signal transmissions with high power.

Second, we observe that a channel may have different signal strengths at different locations. As shown in Fig. 1(b), the relative signal strength of channel 13 in room 11 is greater than 5 dB during the whole measurement period; in room 9, the channel's relative signal strength moves up and down around 5 dB; in room 1 it is always below 5 dB.

Finally, short term occupied channels exist. We refer a channel, which is occupied for some periods of time and vacant for the rest of time, as a short term occupied channel. Fig. 1(c) shows two short term occupied channels (channel 2 in room 13 and channel 27 in room 11) whose relative signal strengths move up and down around 0. The short term occupied channels are actually occupied by some licensed users. But the base stations, which utilize these channels, are far away from the building we perform the measurement. Due to the attenuation caused by the indoor and outdoor obstacles, the signal strengths of the short term occupied channels are near to the white space threshold, and vary up and down around the threshold due to signal fluctuation. Hence, short term occupied channels can be safely accessed by unlicensed users when they are vacant.

Long-time sensing experiment results show the existence of strong channels and short term occupied channels. If we find strong channels and consider them always strong in the future, then we can pay our attention to the short term occupied channels. In this way, the resources (e.g., RF-sensors, energy) can be more efficiently utilized. Actually, there exist weak channels whose signal strengths are lower than the white space threshold in our measurement experiments, but we don't consider them as always weak in order to protect the licensed users.

2.3 Short-Time Sensing

Indoor short-time sensing experiments collect the spatial-channel features of indoor white space. Our indoor white space exploration mechanism was designed based on the data collected during this process.

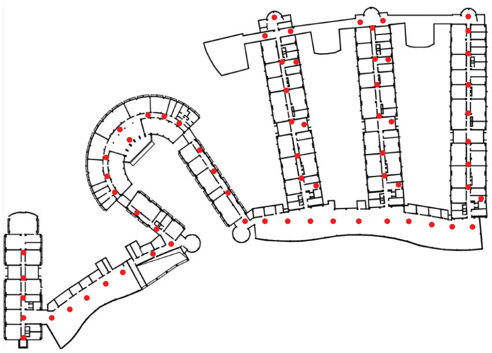


Fig. 2. The map showing the 67 measurement locations on the 3rd floor of the SEIEE building.

The short-time sensing is performed on the third floor of the SEIEE building at Shanghai Jiao Tong University, as shown in Fig. 2. We choose 67 typical locations labeled as location 1 to location 67 (red dots in Fig. 2), and

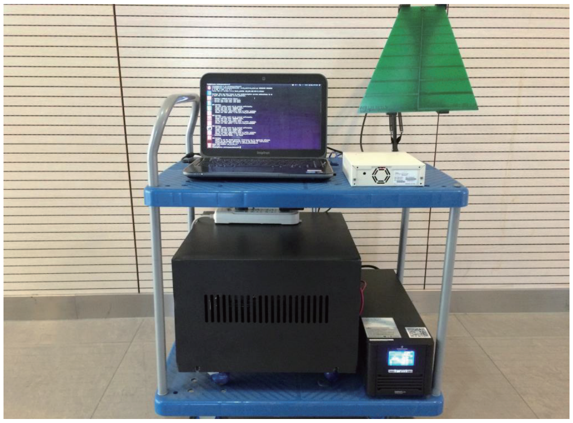


Fig. 3. The measurement device.

measure all 45 TV channels at each location. We mount the measurement device on a movable cart, and use the uninterrupted power supply (UPS) to provide power. The movable measurement device is shown in Fig. 3. In each day, we measure the signal strengths at the 67 locations one by one, and get a *Measurement Matrix (MM)*, which is a 67×45 matrix denoted by M. The measurement process needs about one and a half hours. M contains the absolute signal strengths for 45 channels at 67 locations, each row/column of which represents a location/channel. The measurement lasts for a period of two weeks (April 14, 2014 - April 27, 2014). Accordingly, we get 14 measurement matrices. From the short-time sensing, we have the following observations.

First, Fig. 1(d) shows the indoor white space availability map, where white blocks represent white spaces and black blocks represent occupied channels. We observe that many channels are vacant. On average, 63.9% of channels are vacant according to the short-time sensing results. This means that there are considerable spectrums not being fully utilized. Hence, efficiently utilizing the vacant TV channels is helpful to increase available wireless spectrums. Additionally, we observe that the signal strengths of a channel at different locations could be different (a channel could be occupied at some locations while be vacant at others). This is mainly caused by the complex indoor signal attenuation patterns due to the indoor obstacles (e.g., walls). The outdoor geo-location database based white space exploration mechanisms calculate the signal strengths of TV channels at different outdoor locations through the signal propagation model. If we directly use these approaches to indoor scenarios, we will lose a lot of white spaces.

Second, in Fig. 1(e), the relative signal strengths of all channels are similar at four different locations (1, 26, 28, 31). Prior work [41] described this kind of correlation between different locations according to locations' similarity. They treated the correlated locations as a group and represented the white space availability of all locations in this group using only one of them. It is indeed a creative way to utilize the similarity between locations. However, in our mechanism, we consider not only the similarity but also the linear dependence between locations. *Location dependence* represents the linear dependence between different locations, which means any row of M can be approximately represented as a linear combination of some other rows. Assume that M_i is ith row of M (M_i is a 1×45 row vector containing the signal strengths of 45 channels at location i), we can approximate M_i as

$$M_i \approx a_0 M_0 + a_1 M_{i_1} + a_2 M_{i_2} + ... + a_k M_{i_k},$$

where M_0 is a 1×45 row vector equals to $(1, 1, ..., 1)$ and $M_{i_1}, M_{i_2}, ..., M_{i_k}$ $(i_1, i_2, ..., i_k \neq i)$ are the i_1th, i_2th, ..., i_kth rows of M, and $a_0, ..., a_k$ are the weight parameters. Similar to location dependence, we define *channel dependence* as the linear dependence of channels. As shown in Fig. 1(f), although the similarity criterion is not suitable to describe the correlation between channel 13 and 37, as their signal strengths are not close, our linear dependence criterion works well since the differences of them are almost fixed at different locations. According to the definition of relative location-channel matrix X (Section 3.1), we got that location-channel dependence also exists in X, since $M - X$ is a constant matrix. In Section 4.3, we introduce a novel way to draw the location-channel dependence.

2.4 Summary

The indoor white space measurements give us a better understanding about the characteristics of indoor white spaces. Below we summarize guidelines for designing FIWEX.

- Strong channels are occupied for most of time, thus we can neglect these channels once we spot them.
- Location dependence and channel dependence allow us to infer a channel's signal strength or status at some location based on its correlated channels or locations.

In the following sections, we will show how to use these guidelines to make FIWEX accurate and cost-efficient.

3 System Model

FIWEX aims to accurately identify indoor white spaces with a small number of RF sensors deployed. As shown in Fig. 4, FIWEX is mainly composed of two parts: a central server and a real time sensing module. We first select a set of profiled locations, which cover a number of rooms and corridors of the target building. Just like [41], we assume that white space availability at any indoor location is almost the same as that at its nearest profiled location. Hence, our mechanism mainly focuses on how to get the white space availabilities at all profiled locations from a part of them. In order to make the mechanism efficient in terms of sensor cost and energy consumption, we only deploy sensors at a part of the profiled locations, and recover the complete information based on the correlations among indoor white spaces. The central server receives the partial sensing results reported by the indoor sensors, and then calculates the indoor white space availability map using a reconstruction algorithm. Upon receiving a query for availability of the white spaces at a certain location from a user, the central server makes a response according to the up-to-date indoor white space availability map.

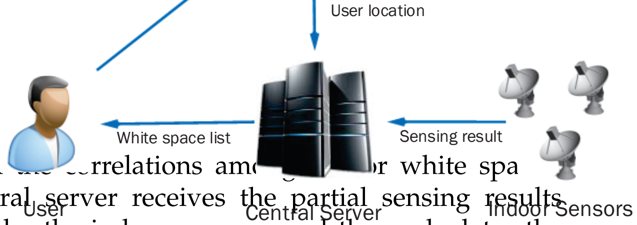


Fig. 4. System architecture of FIWEX.

3.1 Real Time Sensing Module

We describe the real time sensing module using our test building with 67 locations as an example. Although the 67 locations we select cannot cover every room and corridor of the building due to the limitations of room accessibility, we just use them to describe how the real sensing module works. In practice, if the profiled locations are correctly selected, our mechanism can achieve good performance.

The real time sensing module's job is choosing a part of these 67 locations, placing a sensor at each chosen location, performing a "partial sensing" (since not all 67 locations have sensors deployed), and sending the data to the central server.

If the real time sensing module deploys sensors at all 67 locations, clearly the central server can get complete short-time sensing data. However, due to the cost considerations, real time sensing module aims to use less number of sensors to get complete sensing data.

In order to facilitate presentation, we define the following matrices:

- **Relative Location-Channel Matrix (RLCM)**: is a 67×45 matrix recording channels' relative signal strengths. RLCM is denoted by X, where $X(i,j)$ refers to the relative signal strength of channel j at location i,
$$X(i,j) = M(i,j) - TH,$$
where TH is the white space threshold. If $X(i,j) < 0$, then channel j is vacant at location i, otherwise occupied.
- **Binary Index Matrix (BM)**: is a 67×45 matrix, which indicates where the sensors are deployed. BM is denoted by B. For any channel j
$$B(i,j) = \begin{cases} 0 & \text{if no sensor deployed at location } i, \\ 1 & \text{otherwise.} \end{cases}$$
When we deploy a sensor at location i, all the TV channels at location i can be scanned. This means that $B(i,j) = 1$ for $j = 1, 2, \ldots, 45$.
- **Direct Sensory Matrix (DM)**: is a 67×45 matrix, which records the relative signal strengths at locations with sensors deployed. For locations without a sensor, the corresponding rows in DM contain 45 0s. DM is denoted by D:
$$D(i,j) = \begin{cases} X(i,j) & \text{if } B(i,j) = 1, \\ 0 & \text{if } B(i,j) = 0. \end{cases}$$
This means $D = B \circ X$, where $\circ$ refers to the Hadamard product. ($D = B \circ X$ means $D(i,j) = B(i,j)X(i,j)$).

Given a specific number of sensors, different deployments of these sensors lead to different performance of FIWEX. (We will discuss our indoor sensor deployment method in Section 4.5). Once a sensor deployment is determined, then B is fixed. Sensors deployed at different locations collect absolute signal strengths in real time and the corresponding relative ones are recorded in D. After that, real time sensing module submits matrix D to the central server at regular time intervals.

3.2 Central Server

The central server of FIWEX consists of two parts: *data reconstruction* and *white space database*.

In data reconstruction part, a complete relative location-channel matrix is reconstructed based on the direct sensory matrix (D). We define the reconstructed matrix as follows:

- **Reconstructed Matrix (RM)** : is a 67×45 matrix generated by interpolating the missing values in D. RM is denoted by $\tilde{X}$.

Data reconstruction part aims to find $\tilde{X}$ that approximates X as accurate as possible.

Strong channels can be utilized to improve the accuracy of data reconstruction. In contrast to existing works, observations in Section 2.2 (Fig. 1(b)) show that different locations may have different strong channels. Based on this observation, we define the strong channel matrix as:

- **Strong Channel Matrix (SCM)** : is a 67×45 matrix that indicates strong channels at different locations. SCM is denoted by
$$S(i,j) = \begin{cases} 1 & j \text{ is a strong channel at location } i, \\ 0 & \text{otherwise.} \end{cases}$$

Given D, B and S as inputs, the central server performs compressive sensing based data reconstruction. We will present the details of this process in Section 4.

White space database receives the result of data reconstruction $\tilde{X}$, and calculates the up-to-date indoor white space availability map according to $\tilde{X}$. Intuitively, we should compare $\tilde{X}(i,j)$ with 0 to decide whether channel j is vacant at location i. However, FCC requires that the unlicensed access to TV channels should not interfere with the licensed signal transmissions. This means that a white space exploration system should avoid misidentifying an occupied TV channel as vacant. Hence, we compare $\tilde{X}(i,j)$ with the protection range PR, which is less than 0, when calculating the indoor white space availability map. In this way, the licensed signal transmissions are largely protected. We define the reconstructed indoor white space availability map as

$$MAP(i,j) = \begin{cases} 1 & \text{if } \tilde{X}(i,j) < PR, \\ 0 & \text{if } \tilde{X}(i,j) \geq PR \end{cases}$$

where $MAP(i,j) = 1$ means that channel j is vacant at location i and $MAP(i,j) = 0$ otherwise.

Users submit their indoor positions to the central server through an indoor localization system [23], [35]. Given the indoor position of a user, the database first finds one of the profiled locations which is the nearest to the user, and then returns the white space list of this profiled location to the user after considering the interference with neighbors. Here for simplicity, FIWEX assumes the white space availability of a non-profiled location is the same as the availability of the nearest profiled one. We leave the case where these two availability results are not equal to our future work. We have to note that our approach proposed in the paper is to efficiently identify indoor white spaces. A closely related problem on how to utilize the identified indoor white spaces without interfering is another future work.

Before introducing the detailed system design, we summarize notations we employ in TABLE 1.

TABLE 1
Notations used in this paper

Symbols	Description
M	Measurement matrix (MM)
X	Relative Location-channel matrix (RLCM)
$\tilde{X}$	Reconstructed matrix (RM)
B	Binary index matrix (BM)
B_s	BM with strong channels
D	Direct sensory matrix (DM)
D_s	DM with strong channels
S	Strong channel matrix (SCM)
C, C_0	Channel dependence constraint matrix
P, P_0	Location dependence constraint matrix
$\lambda_1, \lambda_2, \lambda_3$	Lagrange multipliers
L, R	SVD of $\tilde{X} : \tilde{X} = LR^T$
r	Rank of L and R
N	The number of sensors

4 DATA RECONSTRUCTION AND SENSOR DEPLOYMENT

In this section, we present our compressive sensing based indoor white space reconstruction algorithm, and propose a cost-efficient sensor deployment method.

Compressive sensing is a generic data reconstruction technique based on the structure and redundancy of real-world signals or datasets [9], [11], [13], [16], [30]. So far, compressive sensing has been widely applied to different realms [22], [29], [37], [44], [45]. The traditional compressive sensing based matrix completion techniques consider the general low rank feature of the matrices, and do not need any prior knowledge. However, we do have prior knowledge about the strong channels and location-channel dependence. Considering the extra knowledge in the compressive sensing based matrix completion will lead to a better performance. Hence, we introduce strong channels and location-channel dependence to compressive sensing, and solve the problem using the alternative least square method. Compressive sensing based data reconstruction algorithms perform very differently given different data loss patterns [22], and thus different sensor deployments may lead to different performance of FIWEX. In order to find appropriate locations to deploy sensors, we propose a *k-medoids* [28] based sensor deployment method.

4.1 Compressive Sensing

Given D, B and S, compressive sensing based matrix completion can help us to approximately reconstruct X. According to the theory of compressive sensing, matrices with low rank feature can be reconstructed with a high accuracy. When the vector containing all singular values of a matrix is sparse, the matrix is low rank. In Fig. 5, we illustrate the distribution of singular values in 7 of the 14 relative location-channel matrices we got in the short time sensing experiments. The X-axis presents the index of singular values, while the Y-axis presents the normalized values of them. This figure shows that the energy is always contributed by the top several singular values in X. On an average, the top 25% singular values contribute most of the energy in this graph. This phenomenon reveals that X exhibits approximate low rank structure. The followings are the details about our compressive sensing algorithm.

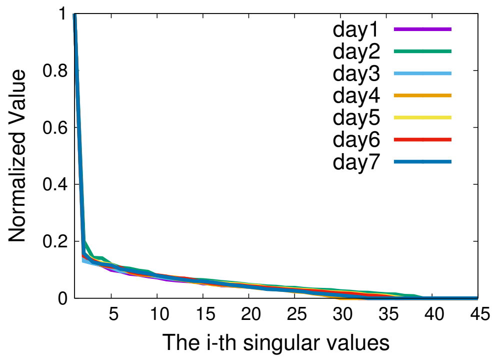


Fig. 5. Low rank feature of X.

We begin with the Singular Value Decomposition (SVD) using the similar methodology as [44]. SVD is a kind of factorization of a matrix, which is usually used for creating low rank matrix approximation. Here, we generalize the discussion of the relative location-channel matrix (X) to an $m \times n$ matrix. For the $m \times n$ matrix X, there exists a factorization of the form

$$X = U\Sigma V^T = \sum_{i=1}^{min(m,n)} \sigma_i \mathbf{u}_i \mathbf{v}_i^T, \quad (1)$$

where U is an $m \times m$ unitary matrix ($i.e., UU^T = U^TU = I$), V^T is the transpose of an $n \times n$ unitary matrix, Σ is an $m \times n$ diagonal matrix with the singular value σ_i of X on the main diagonal, where $\sigma_i \geq \sigma_{i+1}$. Here $\mathbf{u}_i$ and $\mathbf{v}_i$ are the ith columns of U and V, respectively. As we mentioned before, the top 25% singular values of X contribute most of the sum of all singular values, and this means that

$$\sum_{i=1}^{r} \sigma_i \approx \sum_{i=1}^{min(m,n)} \sigma_i \quad (2)$$

and

$$\sum_{i=1}^{r} \sigma_i \mathbf{u}_i \mathbf{v}_i^T \approx \sum_{i=1}^{min(m,n)} \sigma_i \mathbf{u}_i \mathbf{v}_i^T, \quad (3)$$

where $r \ll min(m,n)$. Hence we can approximately represent X as

$$\tilde{X} = \sum_{i=1}^{r} \sigma_i \mathbf{u}_i \mathbf{v}_i^T, \quad (4)$$

where r is the rank of $\tilde{X}$. Actually, $\tilde{X}$ is the best r-rank approximation that minimize the Frobenius norm $\| \cdot \|_F$ between X and $\tilde{X}$. That is, $\tilde{X}$ is the solution to:

$$\begin{aligned} &\text{Minimize} \quad \|X - \tilde{X}\|_F, \\ &\text{Subject to} \quad \mathrm{rank}(\tilde{X}) \leq \mathrm{r}, \end{aligned} \quad (5)$$

where the Frobenius norm $\|X\|_F = \sqrt{\sum_{i,j} X_{ij}^2}$.

In the indoor white space exploration, we are given D and required to estimate X. Then, we can judge whether a channel is busy or not at a given location. It is impossible to directly solve (5), since we do not know the value of original matrix and the proper rank. Instead, we use $B \circ X = D$ as the constraint. Considering the low rank property of X, we can alternatively solve the following rank minimization problem:

$$\begin{aligned} &\text{Minimize} \quad rank(\tilde{X}), \\ &\text{Subject to} \quad B \circ \tilde{X} = D. \end{aligned} \quad (6)$$

However, it is also difficult to solve this rank minimization problem as it is non-convex. So, we transform it to the nuclear norm minimization problem in help of the SVD-like factorization of $\tilde{X}$:

$$\tilde{X} = \tilde{U}\tilde{\Sigma}\tilde{V}^T = LR^T, \quad (7)$$

where $\tilde{\Sigma}$ is an $r \times r$ diagonal matrix containing the maximum r singular value σ_i, $i = 1, 2..., r$, and $L = \tilde{U}\tilde{\Sigma}^{1/2}$, $R = \tilde{V}\tilde{\Sigma}^{1/2}$.

According to the compressive sensing theory [16], [30], we can perform the rank minimization by solving the nuclear norm minimization problem [10], [31] for the low rank matrix LR^T, if the isometry property [30] holds on binary index matrix B. Therefore, we just need to minimize the sum of L's and R's Frobenius norms:

$$\begin{aligned} &\text{Minimize} \quad \|L\|_F^2 + \|R\|_F^2, \\ &\text{Subject to} \quad B \circ (LR^T) = D. \end{aligned} \quad (8)$$

It is usually difficult to find L and R that strictly satisfy (8), because (i) the direct sensory matrix D contains noise, (ii) X is just approximately low rank in the real world, and the rank of X may be largely different in different indoor environments. Thus, we use Lagrange multiplier method to solve (8):

$$\text{Minimize} \quad \|B \circ (LR^T) - D\|_F^2 + \lambda_1(\|L\|_F^2 + \|R\|_F^2). \quad (9)$$

In this way, the constraint $B \circ (LR^T) = D$ is not strictly enforced, and we use the Lagrange multiplier λ_1 to control the tradeoff between the precise fit to the measurement and the rank minimization.

The compressive sensing approach (9) finds the global low rank structure in matrix X. We can further improve the accuracy in two ways: strong channel improvement and location-channel dependence improvement.

4.2 Introducing Strong Channel

Performance of the compressive sensing based indoor white space reconstruction can be improved by taking strong channels into consideration. In contrast to most prior works, which just define a single set of strong channels shared by all indoor locations, we argue that different locations have different strong channels as shown in Section 2. Based on this observation, we define strong channel matrix S in Section 3.2 to describe the spatial variance of strong channels.

In short-time sensing experiment, we obtain 14 short-time sensing data (M) in total. We use 7 of them as the training set while the other 7 are used for evaluation. If the signal strength of channel j at location i is at least 5dB higher than the white space threshold, then channel j is considered as a strong channel at location i and $S(i,j) = 1$; otherwise, $S(i,j) = 0$. We simply consider strong channels as always busy because channels that carry strong signals are normally used for TV broadcast, which is stable in long term (e.g., years). Hence, S can be used in FIWEX.

To combine compressive sensing and strong channels' feature, we define new Binary Index Matrix B_s and Direct Sensory Matrix D_s after considering the influence of S:

$$B_s(i,j) = \begin{cases} 1 & \text{if } S(i,j) = 1, \\ B(i,j) & \text{otherwise.} \end{cases}$$

$$D_s(i,j) = \begin{cases} ave_{ij} & \text{if } S(i,j) = 1, \\ D(i,j) & \text{otherwise.} \end{cases}$$

Here ave_{ij} is the average relative signal strength of channel j at location i in the training set. In this way, formula (9) can be rewritten as

$$\text{Minimize} \quad \|B_s \circ (LR^T) - D_s\|_F^2 + \lambda_1(\|L\|_F^2 + \|R\|_F^2). \tag{10}$$

4.3 Introducing Location-Channel Dependence

The location dependence and channel dependence (Section 2) represent the location/channel structure of indoor white spaces, and can be utilized to improve the performance of FIWEX. Taking compressive sensing, strong channels, location dependence and channel dependence into consideration, we expand (10) as follows:

$$\begin{aligned} \text{Minimize } & \|B_s \circ (LR^T) - D_s\|_F^2 + \lambda_1(\|L\|_F^2 + \|R\|_F^2) \\ & + \lambda_2\|P(LR^T) - P_0\|_F^2 + \lambda_3\|(LR^T)C - C_0\|_F^2, \end{aligned} \tag{11}$$

where P and P_0 are the location dependence constraint matrices, C and C_0 represent the channel dependence constraint matrices. Lagrange multipliers λ_2 and λ_3 are used as the scaling of $\|P(LR^T) - P_0\|_F^2$ and $\|(LR^T)C - C_0\|_F^2$, respectively. Since different choices of P, P_0, C, and C_0 yield different performance of FIWEX, we discuss how to choose these matrices in the following section.

4.3.1 Choice of location-channel dependence constraint matrices

Choice of $P_{(67\times67)}$ and $P_{0(67\times45)}$: Matrix P and P_0 represent the location dependence in relative location-channel matrix (X) and express the linear relationship between different rows of $X_{(67\times45)}$. We propose an innovative way to find appropriate P and P_0. For each row X_i of X, location dependence means that we can approximate X_i as a linear function of other K most correlated rows $X_{i_k} (k = 1, 2, ..., K$ and $i_k \neq i)$ of X_i. We will discuss the proper value of K in Section 4.3.2. The correlation between different rows is measured by the sum of *Pearson product-moment correlation coefficient* [24]. Then we perform multivariate linear regression to find a set of weights w_{i_k}, such that X_i can be best approximated by a linear combination of X_{i_k}:

$$X_i \approx w_{i_0}X_0 + \sum_{k=1}^{K} w_{i_k}X_{i_k}, \tag{12}$$

where X_0 is a row vector (1×45) equals to $(1, 1, 1, ..., 1)$. Initially, every element of P is set to be 0. For the i-th row of P, we set $P(i, i) = 1$, $P(i, i_k) = -w_{i_k}$, for $k = 1, 2, ..., K$. We also set $P_0(i, j) = w_{i_0}$ for $j = 1, 2, ..., 45$. In this way,

$$P(LR^T) - P_0 = \begin{bmatrix} X_1 - \sum_{k=1}^{K} w_{1_k}X_{1_k} - w_{1_0}X_0 \\ X_2 - \sum_{k=1}^{K} w_{2_k}X_{2_k} - w_{2_0}X_0 \\ ... \\ ... \\ X_{67} - \sum_{k=1}^{K} w_{67_k}X_{67_k} - w_{67_0}X_0 \end{bmatrix},$$

and it describes the difference between each row and its linear representation. Since $X_i \approx w_{i_0}X_0 + \sum_{k=1}^{K} w_{i_k}X_{i_k}$, the value of $\|P(LR^T) - P_0\|_F^2$ is expected to be small. Therefore, the existence of $\|P(LR^T) - P_0\|_F^2$ in minimizing (11) guarantees the location dependence in $\tilde{X}$ and this really improve the reconstruction accuracy.

Choice of $C_{(45\times45)}$ and $C_{0(67\times45)}$: Matrix C and C_0 represent the channel dependence in relative location-channel matrix (X), and express the linear relationship between different columns of $X_{(67\times45)}$. Actually, the way to find C and C_0 is almost the same as that of finding P and P_0, except that C and C_0 consider the columns of X instead of rows. In this way, the channel dependence in $\tilde{X}$ is guaranteed by $\|(LR^T)C - C_0\|_F^2$ in (11).

4.3.2 Stability of location-channel dependence

In this part, we study the stability of the location-channel dependence based on the short-time indoor sensing results, and then find the proper value of K. In Section 2.3, a total of 14 short-time sensing data sets were collected and we can get 14 relative location-channel matrices (RLCM) $X^{(1)}$, $X^{(2)}$, ..., $X^{(14)}$. We calculate P, P_0, C, and C_0 based on one of these 14 RLCMs. If the location dependence and channel dependence are stable over time, then P, P_0, C, and C_0 we get can approximately represent the location dependence and channel dependence of other 13 RLCMs.

We calculate P, P_0, C, and C_0 based on $X^{(1)}$ in the way described in Section 4.3.1. As we mentioned before, the value of $\|PX^{(1)} - P_0\|_F^2$ and $\|X^{(1)}C - C_0\|_F^2$ is expected to be small. If the value of $\|PX^{(i)} - P_0\|_F^2 (\|X^{(i)}C - C_0\|_F^2)$ $(i = 2, 3, ..., 14)$ is close to $\|PX^{(1)} - P_0\|_F^2 (\|X^{(1)}C - C_0\|_F^2)$, then the four matrices (P, P_0, C, C_0) can approximately represent the location-channel dependence in $X^{(i)}$ $(i = 2, 3, ..., 14)$. As a result, the location dependence and channel dependence are stable over time.

We compare $\|PX^{(1)} - P_0\|_F^2$ and the average of $\|PX^{(i)} - P_0\|_F^2$ $(i = 2, 3, ..., 14)$, and set

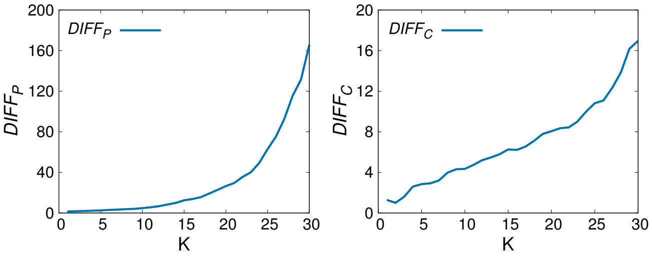
$$DIFF_P = \frac{[\frac{1}{13}\sum_{i=2}^{14} \|PX^{(i)} - P_0\|_F^2] - \|PX^{(1)} - P_0\|_F^2}{\|PX^{(1)} - P_0\|_F^2},$$

where $DIFF_P$ stands for the relative difference between $\|PX^{(1)} - P_0\|_F^2$ and the average of $\|PX^{(i)} - P_0\|_F^2$ $(i = 2, 3, ..., 14)$. Similarly, we use $DIFF_C$ to represent the relative difference between $\|X^{(1)}C - C_0\|_F^2$ and the average of $\|X^{(i)}C - C_0\|_F^2$ $(i = 2, 3, ..., 14)$,

$$DIFF_C = \frac{[\frac{1}{13}\sum_{i=2}^{14} \|X^{(i)}C - C_0\|_F^2] - \|X^{(1)}C - C_0\|_F^2}{\|X^{(1)}C - C_0\|_F^2}.$$

If the values of $DIFF_P$ and $DIFF_C$ are small, then the location dependence and channel dependence are stable. Fig. 6 shows the value of $DIFF_P$ and $DIFF_C$, when K varies from 1 to 30. $DIFF_P$ gets its minimum value (1.54) when $K = 1$ and $DIFF_C$ gets its minimum value (1.28) when $K = 2$. It means that we would obtain very similar P, P_0, C, and C_0 using different $X^{(i)}$ as input, indicating that the location-channel dependence is consistent across different days. Since $DIFF_P$ and $DIFF_C$ get their minimum value when $K = 1$ and $K = 2$, respectively, we set $K = 1$ when calculating P

and P_0 and $K = 2$ when calculating C and C_0. This means that when calculating P and P_0, we just consider each row's most correlated row. When calculating C and C_0, we consider each column's top 2 correlated columns.



(a) Location dependence (b) Channel dependence

Fig. 6. Stability of location-channel dependence

4.4 Data Reconstruction Algorithm

In this part, we present the design of the data reconstruction algorithm. The aim of the data reconstruction algorithm is to find $\tilde{X} = LR^T$ that optimize (11). We let

$$f(L,R) = \|B_s \circ (LR^T) - D_s\|_F^2 + \lambda_1(\|L\|_F^2 + \|R\|_F^2) \\ +\lambda_2\|P(LR^T) - P_0\|_F^2 + \lambda_3\|(LR^T)C - C_0\|_F^2, \tag{13}$$

as the objective function. It is clear that $f(L,R)$ is not a convex function. However, if we fix L or R, the function on the other variable is convex. Hence, we use the alternating steepest descent algorithm [32], which is commonly utilized in the low rank matrix completion, to do the minimization.

In order to minimize $f(L,R)$, we alternatively applies steepest gradient descent to $f(L,R)$ with respect to L and R. As shown in Algorithm 1, L and R are randomly initialized. We first fix L, and update the value of R using a single step of simple line search along gradient descent direction. Then, we fix R, and update L in a similar way. We repeat this process until convergence. Here, we consider that the minimization process is convergent when the relative change of the function value is less than a threshold, which is denoted by

$$\frac{v_{be} - v_{af}}{v_{be}} < ratio,$$

where v_{be} is the function value before updating L and R in each iteration, v_{af} is the function value after that, and $ratio$ is the threshold we set.

We now explain the line search along the gradient descent directions in detail. We denote

$$f(L,R) = f_1(L,R) + f_2(L,R) + f_3(L,R) + f_4(L,R), \tag{14}$$

where

$$\begin{aligned} f_1(L,R) &= \|B_s \circ (LR^T) - D_s\|_F^2, \\ f_2(L,R) &= \lambda_1(\|L\|_F^2 + \|R\|_F^2), \\ f_3(L,R) &= \lambda_2\|P(LR^T) - P_0\|_F^2, \end{aligned}$$

Algorithm 1: Alternating Steepest Descent: $ASD()$

```
Input : $B_s, D_s, P, P_0, C, C_0, \lambda_1, \lambda_2, \lambda_3, r$: Parameters in
        function $f(L,R)$,
        $ratio$: Ratio threshold.
Output: $\tilde{X}$: Reconstructed Matrix
$[m,n] \leftarrow size(B_s)$;
$L \leftarrow random\_matrix(m,r)$;
$R \leftarrow random\_matrix(n,r)$;
repeat
    $v_{be} \leftarrow f(L,R)$;
    $R \leftarrow GD_R(B_s, D_s, P, P_0, C, C_0, \lambda_1, \lambda_2, \lambda_3, L, R)$;
    $L \leftarrow GD_L(B_s, D_s, P, P_0, C, C_0, \lambda_1, \lambda_2, \lambda_3, L, R)$;
    $v_{af} \leftarrow f(L,R)$;
until $\frac{v_{be} - v_{af}}{v_{be}} < ratio$;
$\tilde{X} \leftarrow LR^T$;
return $\tilde{X}$;
```

Algorithm 2: Gradient descent: $GD_R()$

```
Input : $B_s, D_s, P, P_0, C, C_0, \lambda_1, \lambda_2, \lambda_3$: Parameters in
        function $f(L,R)$,
        $L, R$: Inputs of $f(L,R)$.
Output: $R$: The value of $R$ is updated.
$\nabla_r^1 \leftarrow 2(B^T \circ (RL^T) - D^T)L$;
$\nabla_r^2 \leftarrow 2\lambda_1 R$;
$\nabla_r^3 \leftarrow 2\lambda_2(RL^TP^TPL - P_0^TPL)$;
$\nabla_r^4 \leftarrow 2\lambda_3(CC^TRL^TL - CC_0^TL)$;
$\nabla_r \leftarrow \nabla_r^1 + \nabla_r^2 + \nabla_r^3 + \nabla_r^4$;
$t_r \leftarrow \frac{tr(\nabla_r^1\nabla_r^T + \nabla_r^2\nabla_r^T + \nabla_r^3\nabla_r^T + 2\lambda_3(LR^TC - C_0)^TL\nabla_r^TC)}{2(\|B\circ(L\nabla_r^T)\|_F^2 + \lambda_1\|\nabla_r\|_F^2 + \lambda_2\|PL\nabla_r^T\|_F^2 + \lambda_3\|L\nabla_r^TC\|_F^2)}$;
  $R \leftarrow R - t_r\nabla_r$;
return $R$;
```

Algorithm 3: Gradient descent: $GD_L()$

```
Input : $B_s, D_s, P, P_0, C, C_0, \lambda_1, \lambda_2, \lambda_3$: Parameters in
        function $f(L,R)$,
        $L, R$: Inputs of $f(L,R)$.
Output: $L$: The value of $L$ is updated.
$\nabla_l^1 \leftarrow 2(B \circ (LR^T) - D)R$;
$\nabla_l^2 \leftarrow 2\lambda_1 L$;
$\nabla_l^3 \leftarrow 2\lambda_2(P^TPLR^TR - P^TP_0R)$;
$\nabla_l^4 \leftarrow 2\lambda_3(LR^TCC^TR - C_0C^TR)$;
$\nabla_l \leftarrow \nabla_l^1 + \nabla_l^2 + \nabla_l^3 + \nabla_l^4$;
$t_l \leftarrow \frac{tr\left(\nabla_l^1\nabla_l^T + \nabla_l^2\nabla_l^T + 2\lambda_2(PLR^T - P_0)R\nabla_l^TP^T + \nabla_l^4\nabla_l^T\right)}{2(\|B\circ(\nabla_lR^T)\|_F^2 + \lambda_1\|\nabla_l\|_F^2 + \lambda_2\|P\nabla_lR^T\|_F^2 + \lambda_3\|\nabla_lR^TC\|_F^2)}$;
  $L \leftarrow L - t_l\nabla_l$;
return $L$;
```

$$f_4(L,R) \quad = \quad \lambda_3\|(LR^T)C - C_0\|_F^2,$$

and let

$$\nabla_l^i = \frac{\partial f_i(L,R)}{\partial L}, \nabla_r^i = \frac{\partial f_i(L,R)}{\partial R}, \text{ for } i = 1,2,3,4.$$

Then, the directions of gradient ascent are

$$\nabla_l = \frac{\partial f(L,R)}{\partial L} = \nabla_l^1 + \nabla_l^2 + \nabla_l^3 + \nabla_l^4,$$

$$\nabla_r = \frac{\partial f(L,R)}{\partial R} = \nabla_r^1 + \nabla_r^2 + \nabla_r^3 + \nabla_r^4,$$

where

$$\begin{aligned}
\nabla_l^1 &= 2(B \circ (LR^T) - D)R,\\
\nabla_l^2 &= 2\lambda_1 L,\\
\nabla_l^3 &= 2\lambda_2(P^T PLR^T R - P^T P_0 R),\\
\nabla_l^4 &= 2\lambda_3(LR^T CC^T R - C_0 C^T R),\\
\nabla_r^1 &= 2(B^T \circ (RL^T) - D^T)L,\\
\nabla_r^2 &= 2\lambda_1 R,\\
\nabla_r^3 &= 2\lambda_2(RL^T P^T PL - P_0^T PL),\\
\nabla_r^4 &= 2\lambda_3(CC^T RL^T L - CC_0^T L).
\end{aligned}$$

To minimize $f(L, R)$, we update L and R along the gradient descent directions $-\nabla_l$ and $-\nabla_r$, respectively. If we denote t_l and t_r to be the steepest descent stepsize along the gradient descent directions $-\nabla_l$ and $-\nabla_r$, then the values of L and R are updated as:

$$\begin{aligned}
R &\leftarrow R - t_r \nabla_r,\\
L &\leftarrow L - t_l \nabla_l.
\end{aligned}$$

t_l and t_r are selected to minimize the function value at directions $-\nabla_l$ and ∇_r, which means that

$$t_l = \arg\min_t g_L(t), \quad \text{where} \quad g_L(t) = f(L - t\nabla_l, R),$$
$$t_r = \arg\min_t g_R(t), \quad \text{where} \quad g_R(t) = f(L, R - t\nabla_r).$$

We differentiate $g_L(t), g_R(t)$, and set $g'_L(t) = g'_R(t) = 0$. Then, we get

$$t_l = \frac{tr\left(\nabla_l^1 \nabla_l^T + \nabla_l^2 \nabla_l^T + 2\lambda_2 (PLR^T - P_0) R \nabla_l^T P^T + \nabla_l^4 \nabla_l^T\right)}{2(\|B \circ (\nabla_l R^T)\|_F^2 + \lambda_1 \|\nabla_l\|_F^2 + \lambda_2 \|P \nabla_l R^T\|_F^2 + \lambda_3 \|\nabla_l R^T C\|_F^2)},$$
$$t_r = \frac{tr(\nabla_r^1 \nabla_r^T + \nabla_r^2 \nabla_r^T + \nabla_r^3 \nabla_r^T + 2\lambda_3 (LR^T C - C_0)^T L \nabla_r^T C)}{2(\|B \circ (L \nabla_r^T)\|_F^2 + \lambda_1 \|\nabla_r\|_F^2 + \lambda_2 \|PL \nabla_r^T\|_F^2 + \lambda_3 \|L \nabla_r^T C\|_F^2)}.$$

Functions $GD_R()$ (Algorithm 2) and GD_L (Algorithm 3) show the process of updating R and L, respectively.

Just as shown in [32], the alternating steepest descent method can converge to a stationary point. Yet, considering that $f(L, R)$ is not a convex function, our algorithm may converge to a local optimal point. In practice, we can repeat Algorithm 1 for several times, and find the best result. Moreover, our alternating steepest descent method only contains the addition, multiplication, and Frobenius norm operations of matrix, which gives the algorithm a low time complexity. Hence, it is also feasible, when we meet very large matrices.

4.5 Sensor Deployment

As mentioned before, different deployments of sensors lead to different performance of FIWEX. Although brute-force enumeration of indoor locations can find the best deployment solution, the computation overhead is unacceptable ($C_{67}^N - 1$ comparisons are needed given N sensors). In this section, we propose a novel sensor deployment method based on location dependence and the clustering technique.

According to the low rank matrix completion theory [10], [30], when the matrix corresponding to the sensor deployment satisfies the restricted isometry property (RIP), the low rank matrix X can be recovered with a high accuracy. Usually, some random matrices are used. Considering the truth that deploying one sensor means that we can get one row's data of X, it is not possible for us to construct a random sampling matrix B. However, the reconstruction algorithm is general. We propose a heuristic sensor deployment method, which is shown to be effective in the evaluation part.

Algorithm 4: Sensor deployment

```
Function: $X^{(1)}, X^{(2)}, \dots, X^{(ts)}$: relative location-channel
              matrices used for training,
          $ts$: size of training set,
          $N$: the number of given sensors,
          $v_{opt}$: initially a sufficiently large number.
Output  : $SL_{opt}$: a sensor locations list.
1  $[m, n] = size(X^{(1)})$; $PEA = 0^{m \times m}$;
2  for $i = 1$ to $m$ do
3      for $j = 1$ to $m$ do
4          for $k = 1$ to $ts$ do
5              $PEA(i, j) = PEA(i, j) + \frac{Cov(X_i^{(k)}, X_j^{(k)})}{\sigma_{X_i^{(k)}} \sigma_{X_j^{(k)}}}$;
6          $PEA(i, j) = |PEA(i, j)|$;
7  $PEA = -PEA$;
8  for 1 to 100 do
9      $[SL, v] =$ kmedoids($PEA, N$);
10     if $v < v_{opt}$ then
11         $SL_{opt} = SL$; $v_{opt} = v$;
12 return $SL_{opt}$;
```

We have shown that there exists linear dependence between different locations, which means that the signal strengths at one location can be approximated by the linear combination of those at other locations. Given N sensors, FIWEX can collect the signal strengths of TV channels at N indoor locations, and reconstruct the signal strengths of the whole indoor environment based on them. We hope that the redundancy among the information collected by the sensors should be as little as possible. Hence, we try to deploy sensors at "independent" locations as every one of them can represents a set of correlated locations. And we can maximize the information we get in this way.

Since the brute-force enumeration is unacceptable here, we propose a heuristic method based on the clustering technique. Our sensor deployment algorithm can be briefly divided into two steps: (i) Cluster all locations into N groups (N is the number of sensors here). (ii) Deploy one sensor at each center of these N groups. Algorithm 4 shows the detailed sensor deployment process. We deploy sensors based on a training set with ts relative location-channel matrices ($X^{(1)}, X^{(2)}, \dots, X^{(ts)}$). Since we try to deploy sensors at "independent" locations, we treat every location (row) as a variable and use the sum of Pearson product-moment correlation coefficients in the training set as the similarity metric. In Algorithm 4, line 1-6 show the process of Pearson product-moment correlation coefficients calculation. Matrix PEA stores the Pearson product-moment correlation coefficients among different pairs of locations. The value of Pearson product-moment correlation coefficients are between -1 and 1, where positive value means positive correlation and negative value means negative correlation. In line 6, we calculate the absolute values of Pearson correlation coefficients, since we do not care whether the correlation is positive or not. Different from the Euclidean distance, where a smaller value means two variables are closer, a larger value in PEA means the corresponding two locations are more dependent. Hence, we use the opposite number of Pearson correlation coefficients as the distance metric in the clustering algorithm (line 7).

We choose k-*mediod* [28] technology to do the clustering. In Algorithm 4 line 9, kmedoids(PEA, N) divides the m locations into N clusters, the centers of which are stored in SL. v stores the sum of all distances between every row and the center of cluster it belongs to. A smaller v means a better clustering result. Considering the truth that k-*medoids* clustering algorithm may find the local optimization, we run kmedoids for 100 times and select the best result. Then we deploy N sensors in the N cluster centers.

We also consider the stability of the clustering results. In Section 4.3.2, we show that the location dependence is stable over time, as a result, the clustering algorithm based on location dependence is also stable. That means, if we deploy sensors according to some training data, these sensors will also perform well in the future.

At last, we study the problem of choosing the proper value of N. It is clear that there is a tradeoff between the number of sensors and the system performance. Deploying more sensors gives less reconstruction errors, but leads to a higher cost. We determine the value of N based on the measurement data we get. We first vary the number of sensors from 1 to 67, and get the locations of sensors using the above mentioned sensor deployment method. Then, we calculate the reconstruction errors at different sensors based on the measurement data. Finally, we determine the proper number of sensors according to our reconstruction error threshold. We choose the smallest N, which yields a reconstruction error less than the threshold. Actually, there are also other approaches to determine a proper N. For example, we can choose N by determining the proper number of clusters [33]. In this paper, we mainly focus on the relation between the system performance and the number of sensors. Hence, we only study the approach based on the measurement data in the evaluation.

5 PERFORMANCE EVALUATION

In this section, we perform experiments to evaluate the performance of FIWEX. We first evaluate the cost-efficiency of FIWEX by comparing it to the state-of-the-art indoor white space exploration system, WISER. Then, we study the performance of the data reconstruction method. Next, we show the feasibility of our sensor deployment method. At last, we evaluate the method of determining the proper number of sensors.

5.1 Methodology

The evaluation is based on the data collected in the short-time sensing. Although due to limitations of room accessibility, we can only consider 67 locations in the experiment, which may not cover all the rooms and corridors in our complex, the evaluation results can still give a good indication of FIWEX's performance. In practice, if a sufficient number of profiled locations are considered, FIWEX is expected to have satisfactory performance. Furthermore, FIWEX is not limited to any specific indoor environment, because it utilizes the general correlations of indoor white spaces [41].

For real applications, every time before deploying FIWEX to a new building, we need to perform the training process in order to find the values of B_s, D_s, P, P_0, C, C_0 and to deploy sensors. We use the measurement results in the short-time sensing to perform the evaluation. In the short-time sensing, we collect a total of 14 data sets (April 14, 2014 - April 27, 2014). We divide them into two parts: the first week's data is used as training set while the data of the second week is used for testing the performance. From the training set, we get 7 relative location-channel matrices, and use the average of them to train WISER and FIWEX. We set $ratio = 0.001,\ \lambda_1 = 10,\ \lambda_2 = 1,\ \lambda_3 = 0.01,\ r = 10$.

We choose FA Rate, WS Loss Rate, and Reconstruction Error as the metrics of system performance. Their definitions are as follows.

- **False Alarm Rate (FA Rate)**: the ratio between the number of channels that a system mis-identifies as vacant and the total number of vacant channels identified by the system.
- **White Space Loss Rate (WS Loss Rate)**: the ratio between the number of channels that a system mis-identifies as occupied and the total number of actually-vacant channels.
- **Reconstruction Error**:

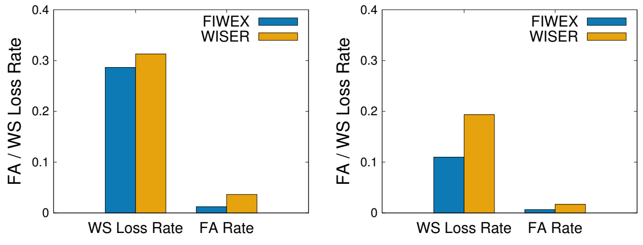
$$\|\tilde{X} - X\|_2 / \|X\|_2$$

where X is the real relative location-channel matrix and $\tilde{X}$ is the reconstructed one.

The protection range, PR, controls the tradeoff between FA Rate and WS Loss Rate. Users can choose a proper value of PR according to their needs. Here, we set $PR = -0.7$. We use 7 datasets to do the evaluation. For each dataset, we run FIWEX for 500 times, and calculate the average result. Then, we average the results of the 7 datasets as the final result.

5.2 Performance Comparison With WISER

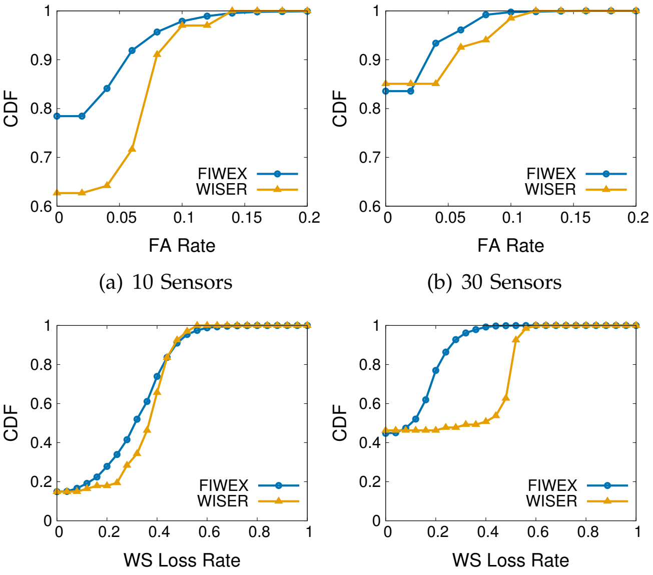
We first run FIWEX and WISER using two different numbers of sensors: 10 and 30, and compare their FA



(a) 10 Sensors

(b) 30 Sensors

Fig. 7. Average FA rate, WS Loss Rate.



(a) 10 Sensors

(b) 30 Sensors

(c) 10 Sensors

(d) 30 Sensors

Fig. 8. CDF curves of FA Rate and WS Loss Rate.

Rates and WS Loss Rates. As shown in Fig. 7(a), when there are 10 sensors, the FA Rate and WS Loss Rate of FIWEX are 1.2% and 28.6% while WISER performs an FA rate 3.6% and a WS Loss Rate 31.3%. The result shows that FIWEX has lower or better FA Rate and WS Loss Rate compared with WISER when 10 sensors are used. Fig. 7(b) illustrates that FIWEX also performs better than WISER when 30 sensors are used.

In these two scenarios, FIWEX performs better than WISER. This is because although WISER and FIWEX consider the correlations among channels and locations, WISER just clusters channels and locations into groups and considers the white space availabilities of the same group are also the same, FIWEX further considers the general linear dependence among channels and locations, which contains more information.

We also study the FA Rates and WS Loss Rates at different locations. We calculate the FA Rates and WS Loss Rates at different locations, and draw their Cumulative Distribution Function (CDF) curves. In the CDF curves, we include the locations that we have already deployed sensors. Since the FA Rates are much smaller than the WS Loss Rates, we change the range of axis to [0:0.2] and [0.6:1], when drawing the CDF curves of FA Rate (Fig. 8(a) and Fig. 8(b)). Fig. 8(a) shows the CDF curves of FA Rate with 10 sensors. It is shown that the CDF curve of FIWEX is "higher" than that of WISER. This means that in FIWEX, less locations suffer the high FA Rates. For example, in FIWEX, only 8.1% locations suffer an FA Rate that is higher than 6%, while the percentage is 29.4% in WISER. When there are 30 sensors, we get similar results about the CDF curves of FA Rate. The details are shown in Fig. 8(b).

The CDF curves of WS Loss Rate are shown in Fig. 8(c) and Fig. 8(d). We observe that FIWEX also performs better than WISER in terms of WS Loss Rate.

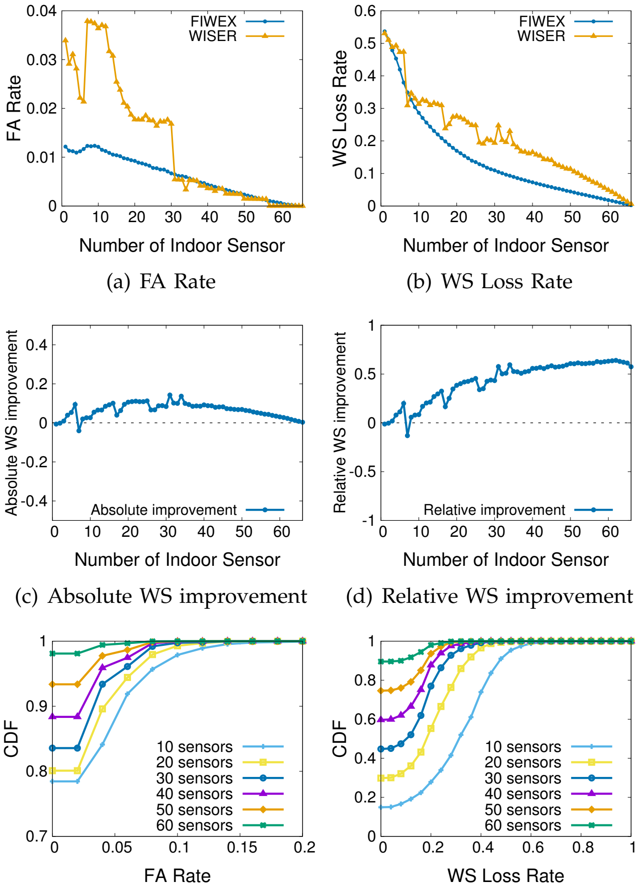
5.3 Performance on The Number of Indoor Sensors

In a general indoor scenario, there exists a tradeoff between system performance and the number of indoor sensors that are deployed. To understand this tradeoff, we vary the number of sensors from 1 to 66 (when there are 67 sensors, FA Rate and WS Loss Rate of FIWEX and WISER are both 0), and evaluate the performance of FIWEX and WISER. In this experiment, we define the following four metrics to describe the FIWEX's performance improvement compared to WISER.

- **Absolute FA improvement:** the difference between FA Rates of WISER and FIWEX.
- **Relative FA improvement:** the ratio between absolute FA improvement and the corresponding FA rate of WISER. (If the FA Rate of WISER is 0 while the absolute FA improvement is not 0, we set the relative FA improvement to be -1.)
- **Absolute WS improvement:** the difference between WS Loss Rates of WISER and FIWEX.
- **Relative WS improvement:** the ratio between absolute WS improvement and the corresponding WS Loss Rate of WISER. (If the WS Loss Rate of WISER is 0 while the absolute WS improvement is not 0, we set the relative WS improvement to be -1.)

In Fig. 9(a), we compare the FA Rates. When the number of sensors is less than 30, the FA Rates of FIWEX are obviously smaller than those of WISER. When the number of sensors is larger than 30, FIWEX and WISER give comparative FA Rates. Actually, the average FA rate (from 1 sensor to 66 sensors) of WISER is 1.25%, while this number is 0.61% for FIWEX. FIWEX has an average absolute FA improvement of 0.64% and an average relative FA improvement of 51.2% compared with WISER. We also observe that the FA Rate of FIWEX changes slowly as the number of sensors increases, this is consistent to the evaluation results in Fig. 8(a) and Fig. 8(b).

Fig. 9(b) shows the WS Loss Rate of FIWEX and WISER with different number of indoor sensors. The average WS Loss Rate (from 1 sensor to 66 sensors) of WISER is 20.3% and for FIWEX, this number is 14.2%. FIWEX has an average absolute WS improvement of 6.1% and an average relative WS improvement of



(a) FA Rate

(b) WS Loss Rate

(c) Absolute WS improvement

(d) Relative WS improvement

(e) FA Rate CDF

(f) WS Loss Rate CDF

Fig. 9. Evaluation results on the number of sensors

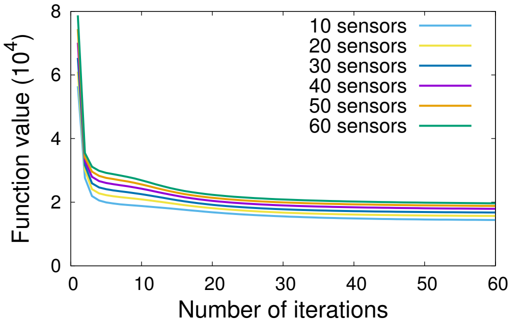
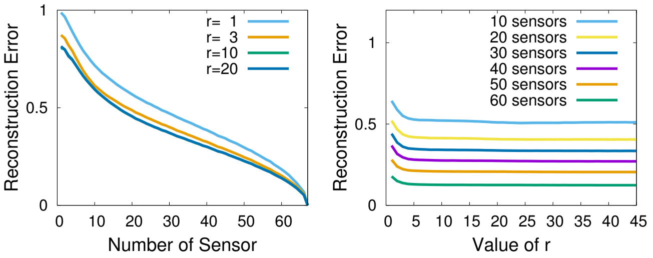


Fig. 10. Convergence of alternating steepest descent.

TABLE 2
Convergence iterations and times

Sensors	10	20	30	40	50	60
Iterations	57.8	55.1	54.5	55.8	56.4	57.7
Time (s)	0.156	0.152	0.149	0.147	0.150	0.158



(a) Different r

(b) Different number of sensors

Fig. 11. Influence of r.

30.0% compared to WISER. Fig. 9(c) and Fig. 9(d) show the absolute and relative WS improvement versus the number of indoor sensors. They demonstrate the amount of extra white spaces FIWEX can identify compared with WISER.

We depict the CDF curves of FIWEX regarding to the FA Rates and WS Loss Rates of all locations. Fig. 9(e) shows the 6 CDF curves (10, 20, ..., 60 sensors) about FA rate. Similar to Fig. 8(a) and Fig. 8(b), we change the range of axis to [0:0.2] and [0.7:1]. The CDF curve is "higher" when more sensors are used. This means that the number of locations with high FA Rates decreases as the number of sensors increases. The CDF curves of WS Loss Rate of different number of sensors shown in Fig. 9(f) illustrate similar result.

5.4 Performance of Data Reconstruction Method

In this part, we study the performance of our data reconstruction method. We first verify the convergence of the alternating steepest descent algorithm (Algorithm 1). Then, we study the way to find the proper value of r. At last, we study the influence of strong channels and location-channel dependence.

We study the convergence of Algorithm 1 with 6 different numbers of sensors: 10, 20, ..., 60. In Algorithm 1, the steepest descent method stops when the relative change of the function value is less than $ratio$ ($\frac{v_{be}-v_{af}}{v_{be}} < ratio$). Since we repeat the evaluation for 500 times, the number of iterations needed for convergence may be different. Hence, for the ease of drawing figures, we perform the descent process (lines 5-8) for 60 iterations, which is enough for convergence, in each run of the evaluation. The results are shown in Fig. 10. We can observe that the function values decrease quickly in the first 10 iterations, and then gradually converge.

We also evaluate the average convergence iterations and times on different number of sensors. We run the code on a PC with an Intel Core i5 CPU and 4 GB RAM. TABLE 2 illustrates the results. For example, when 10 sensors are used, the steepest descent algorithm should run 57.8 iterations on average before convergence. This process needs an average of 0.156 seconds. The above results demonstrate that our data reconstruction algorithm converges quickly.

We now focus on the influence of the variable r. r refers to the size of L and R ($L : m \times r$, $R : n \times r$). A larger r leads to less reconstruction errors, but a higher computational cost. A proper r should be as small as possible, while giving a satisfactory reconstruction error. In order to study the relation between r and the

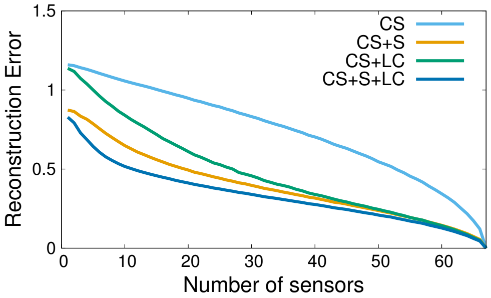


Fig. 12. Influence of strong channels and location-channel dependence.

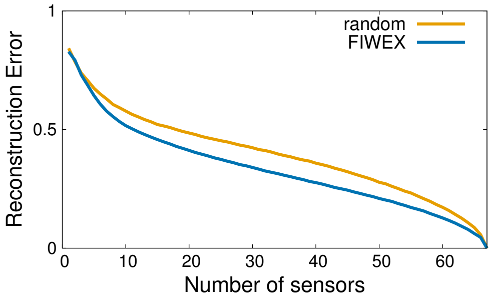


Fig. 13. Comparison with random sensor deployment.

reconstruction error, we vary the number of sensors from 1 to 67, and calculate the reconstruction errors at four different values of r: 1, 3, 10, and 20. The results are shown in Fig. 11(a). The reconstruction errors of $r = 1$ are obviously larger than the others. When r becomes larger, the difference of reconstruction errors between different r becomes smaller. For example, $r = 10$ and $r = 20$ almost share the same reconstruction errors. In Fig. 11(b), we vary the value of r from 1 to 45, and calculate the reconstruction errors with 6 numbers of sensors: 10, 20, ... , 60. We observe that when r is small, the reconstruction error decreases as r increases. Then, when the value of r is large, the reconstruction errors are almost stable at different r.

According to the definition of r in equation (3), r refers to the rank of the matrix $\tilde{X} = LR^T$ after reconstruction. Since matrix X is low rank, $\tilde{X} = LR^T$ can approximates X with high accuracy, when r is larger than the rank of X. Hence, as shown in Fig. 11(b), as the increment of r, the reconstruction errors first decrease, and then become stable. We can choose the proper value of r as a small value, where the reconstruction error is stable. In this work, we use $r = 10$.

At last, we study the efficiency of our strong channels and location-channel dependence aware compressive sensing based matrix completion algorithm. We first run the data reconstruction algorithm using only the compressive sensing based matrix completion. Then, we separately consider the improvements with extra strong channels or location-channel dependence. Next, we add the improvements with strong channels and location-channel dependence at the same time. The objective functions in these four cases are as follows:

- Compressive sensing based matrix completion (CS):

$$\|B \circ (LR^T) - D\|_F^2 + \lambda_1(\|L\|_F^2 + \|R\|_F^2).$$

- With strong channels (CS+S):

$$\|B_s \circ (LR^T) - D_s\|_F^2 + \lambda_1(\|L\|_F^2 + \|R\|_F^2).$$

- With location-channel dependence (CS+LC):

$$\|B \circ (LR^T) - D\|_F^2 + \lambda_1(\|L\|_F^2 + \|R\|_F^2)$$
$$+\lambda_2\|P(LR^T) - P_0\|_F^2 + \lambda_3\|(LR^T)C - C_0\|_F^2.$$

- With both strong channels and location-channel dependence (CS+S+LC):

$$\|B_s \circ (LR^T) - D_s\|_F^2 + \lambda_1(\|L\|_F^2 + \|R\|_F^2)$$
$$+\lambda_2\|P(LR^T) - P_0\|_F^2 + \lambda_3\|(LR^T)C - C_0\|_F^2.$$

We run the data reconstruction algorithm based on the above four objective functions, and calculate the corresponding reconstruction errors with different numbers of sensors. The results are shown in Fig. 12. We observe that the reconstruction errors are the largest, when we only use the compressive sensing based matrix completion method. After considering the extra strong channels or location-channel dependence, the reconstruction errors become smaller. When we consider both of the strong channels and location-channel dependence, the reconstruction errors get the smallest values.

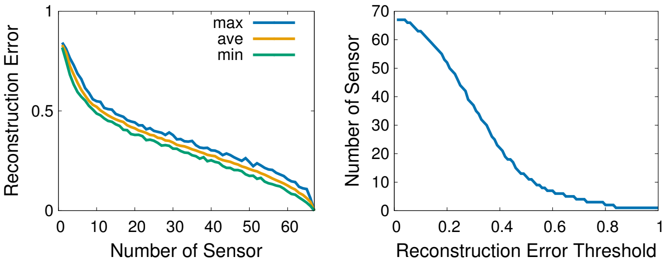
5.5 Performance of Sensor Deployment Method

In this part, we study the performance of our sensor deployment method. We first evaluate the feasibility of our sensor deployment algorithm. Then, we evaluate the method of determining the proper number of sensors.

As we mentioned in Section 4.5, the location dependence of indoor white spaces is considered when deploying indoor sensors. During the deployment process, we place sensors at “independent” locations so as to maximize the white space information we get. To show the efficiency of our sensor deployment method, we run a basic random sensor deployment method from 1 to 67 sensors, and compare their reconstruction errors.

Fig. 13 shows the comparison results. We observe that the reconstruction errors of the random sensor deployment are larger than those of our deployment method. On average, our sensor deployment method gives 15.6% lower reconstruction error compared to the random deployment.

An important problem in indoor white space exploration is determining the proper number of indoor sensors (N). In real world, there is a tradeoff between the number of indoor sensors and the system performance. A large number of sensors means that we can explore indoor white spaces with a high accuracy, while a small number of sensors leads to a low system cost. Hence it is an essential work to determine a proper N.



(a) Reconstruction Error (b) Number of sensors

Fig. 14. Reconstruction error and the number of sensors

In our work, we propose a method to determine the proper number of indoor sensors based on the reconstruction error. We run FWIEX based on the training set for 500 times, and calculate the maximum, average, and minimum reconstruction errors, which is shown in Fig. 14(a). With the increment of the indoor sensors, the reconstruction error decreases. Users can determine the proper number of sensors based on the average reconstruction error they can suffer. Given a reconstruction error threshold, we find the smallest number of sensors N, which gives an average reconstruction error less than the threshold, and then deploy N sensors using Algorithm 4. Fig. 14(b) shows the relation between the reconstruction error threshold and the number of needed indoor sensors. If we want to get a smaller reconstruction error, we should deploy more indoor sensors.

6 RELATED WORK

Most of the existing works on TV white space focused on outdoor scenarios. For example, in [6], Bahl et al. studied the white space characteristics (e.g., spatial variation, spectrum fragmentation, and temporal variation), and proposed the first white space Wi-Fi like wireless network, called WhiteFi. In [15], the challenges of designing short range unlicensed access systems are studied, and the authors proposed efficient algorithms for a dynamic spectrum allocation. A lot of outdoor white space measurements have also been made in Singapore [19], Chicago [27], and Guangzhou [40]. Please see [8] for a survey of both indoor and outdoor white space measurement.

FCC proposed two ways to explore white spaces [2]: spectrum sensing and geo-location database, which have been widely studied in the past few years. In [17], Ghasemi and Sousa studied the regulatory requirements and major challenges of spectrum sensing in cognitive radio systems. A TV white space spectrum sensing prototype was proposed in [7]. Gurney et al. [18] discussed the use of geo-location databases for efficient white space spectrum utilization, and described a detailed geo-location database application for TV band incumbents. In 2014, Zhang et al. [43] proposed a measurement framework for TV white spaces leveraging spectrum sensors deployed on public vehicles. This really improve the efficiency of geo-location database in exploring outdoor white spaces.

The outdoor white space exploration methods cannot be directly applied to indoor scenarios because of the complicated indoor environment. In [41], Ying et al. performed the white space measurement in both indoor and outdoor environments, and studied the strong channel, channel correlation, and location correlation of indoor white spaces characteristics. Based on these characteristics, they proposed the first indoor white space identification system, called WISER. WISER utilizes the channel-location clustering based algorithm to explore indoor white spaces, and obtained great improvement compared to the baseline approach. However, the property that different locations have different strong channels and the linear dependence of locations and channels are not considered by the authors. Based on these observations, we present FIWEX, a compressive sensing based cost-efficient indoor white space exploration mechanism.

Compressive sensing is a generic data reconstruction technique based on the structure and redundancy of real-world signals or datasets. These years, compressive sensing theory has been widely studied and utilized in a lot of fields. For instance, Zhang et al. [44] proposed a compressive sensing based data reconstruction algorithm to deal with the missing values in the network traffic matrices. They combined the spatio-temporal features with the compressive sensing method and greatly improve the reconstruction accuracy. Luo et al. [26] were the first to apply compressive sensing technique to sensor data gathering for large-scale wireless sensor networks, and proposed an energy efficient data gathering algorithm. Rallapalli et al. [29] studied the temporal stability and low rank structure in mobile networks, and proposed three compressive sensing based localization schemes: low rank based localization, temporal stability based localization, and temporal stability and low rank localization. According to the extensive evaluation results, their schemes significantly outperform state-of-the-art localization schemes. Kong et al. [22] investigated the data loss pattern in sensor networks and studied the low rank structure, temporal stability feature, and spatial correlation feature from extensive real-world datasets. Based on these observations, they proposed a compressive sensing based sensor network data reconstruction algorithm with high reconstruction accuracy. When we apply compressive sensing technique to the real-world network scenarios, a major challenge is the existence of missing data, measurement errors, and anomalies. In order to address these issues, Chen et al. [13] developed a novel technique to accurately decompose a network matrix into a low rank matrix, an error matrix, an anomaly matrix, and a noise matrix. According to their experiments based on a wide range of real-world network matrices, this kind of decomposition really outperforms state-of-the-art compressive sensing schemes. Moreover, compressive sensing has been utilized for soil moisture sensing [37], latency analysis in wireless sensor networks [45], data gathering in lossy wireless sensor networks [38], and urban traffic sensing [25]. In FIWEX, we combine compressive sensing with indoor white space exploration in an innovative way and efficiently explore indoor white spaces with a high accuracy.

The projection design is an important part of the compressive sensing technique. The original compressive sensing works [9], [11], [16], [30] use some kind of random matrix as the projection, which is not practical in some real-world applications. In [20], Ji et al. presented the concept of Bayesian compressive sensing and designed an adaptive projection selection method using differential entropy. Adaptive compressive sensing, which has been widely utilized in wireless sensor network data gathering [14], [36] and other realms, can help us to design a more efficient sensor deployment algorithm, and this is one of our future work.

7 CONCLUSIONS

In this paper, we performed indoor white space measurements in a real building to study the characteristics of indoor white spaces. The measurement results confirmed the existence of strong channels and location-channel dependence. Motivated by these observations, we proposed a cost-efficient indoor white space exploration mechanism, called FIWEX. Given the same number of sensors, FIWEX can identify more indoor white spaces with an average less false alarms compared to the existing indoor white space exploration systems.

REFERENCES

[1] Universal software radio peripheral. `https://www.ettus.com`.

[2] Second Memorandum Option And Order. `https://apps.fcc.gov/edocs_public/attachmatch/FCC-10-174A1.pdf`.

[3] ThinkRf. The thinkrf wireless signals intelligence platform. `http://www.thinkrf.com/pages/wsa5000`.

[4] R. Ahuja, R. Corke, and A. Bok, "Cognitive radio system using IEEE 802.11 a over UHF TVWS," in *Proceedings of DySPAN*. IEEE, 2008, pp. 1–9.

[5] I. F. Akyildiz, W.-Y. Lee, M. C. Vuran, and S. Mohanty, "NeXt generation/dynamic spectrum access/cognitive radio wireless networks: a survey," *Computer Networks*, vol. 50, no. 13, pp. 2127–2159, 2006.

[6] P. Bahl, R. Chandra, T. Moscibroda, R. Murty, and M. Welsh, "White space networking with wi-fi like connectivity," in *Proceedings of SIGCOMM*. ACM, 2009, pp. 27–38.

[7] R. Balamurthi, H. Joshi, C. Nguyen, A. K. Sadek, S. J. Shellhammer, and C. Shen, "A TV white space spectrum sensing prototype," in *Proceedings of DySPAN*. IEEE, 2011, pp. 297–307.

[8] T. X. Brown, E. Pietrosemoli, M. Zennaro, A. Bagula, H. Mauwa, and S. M. Nleya, "A survey of TV white space measurements," in *e-Infrastructure and e-Services for Developing Countries*. Springer, 2014, pp. 164–172.

[9] E. J. Candès, "Compressive sampling," in *Proceedings of the international congress of mathematicians*, vol. 3. Madrid, Spain, 2006, pp. 1433–1452.

[10] E. J. Candès and B. Recht, "Exact matrix completion via convex optimization," *Foundations of Computational mathematics*, vol. 9, no. 6, pp. 717–772, 2009.

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

[12] V. Chandrasekhar, J. G. Andrews, and A. Gatherer, "Femtocell networks: a survey," *Communications Magazine, IEEE*, vol. 46, no. 9, pp. 59–67, 2008.

[13] Y. Chen, L. Qiu, Y. Zhang, G. Xue, and Z. Hu, "Robust network compressive sensing," in *Proceedings of MOBICOM*. ACM, 2014, pp. 545–556.

[14] C. Chou, R. Rana, and W. Hu, "Energy efficient information collection in wireless sensor networks using adaptive compressive sensing," in *Proceedings of LCM*. IEEE, 2009, pp. 443–450.

[15] S. Deb, V. Srinivasan, and R. Maheshwari, "Dynamic spectrum access in DTV white spaces: design rules, architecture and algorithms," in *Proceedings of MOBICOM*. ACM, 2009, pp. 1–12.

[16] D. L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[17] A. Ghasemi and E. S. Sousa, "Spectrum sensing in cognitive radio networks: requirements, challenges and design trade-offs," *Communications Magazine, IEEE*, vol. 46, no. 4, pp. 32–39, 2008.

[18] D. Gurney, G. Buchwald, L. Ecklund, S. Kuffner, and J. Grosspietsch, "Geo-location database techniques for incumbent protection in the tv white space," in *Proceedings of DySPAN*. IEEE, 2008, pp. 1–9.

[19] M. H. Islam, C. L. Koh, S. W. Oh, X. Qing, Y. Y. Lai, C. Wang, Y. Liang, B. E. Toh, F. Chin, G. L. Tan, and W. Toh, "Spectrum survey in singapore: Occupancy measurements and analyses," in *Proceeding of CrownCom*. IEEE, 2008, pp. 1–7.

[20] S. Ji, Y. Xue, and L. Carin, "Bayesian compressive sensing," *IEEE Transactions on Signal Processing*, vol. 56, no. 6, pp. 2346–2356, 2008.

[21] N. E. Klepeis, W. C. Nelson, W. R. Ott, J. P. Robinson, A. M. Tsang, P. Switzer, J. V. Behar, S. C. Hern, W. H. Engelmann *et al.*, "The national human activity pattern survey (NHAPS): a resource for assessing exposure to environmental pollutants," *Journal of exposure analysis and environmental epidemiology*, vol. 11, no. 3, pp. 231–252, 2001.

[22] L. Kong, M. Xia, X. Liu, M. Wu, and X. Liu, "Data loss and reconstruction in sensor networks," in *Proceedings of INFOCOM*. IEEE, 2013, pp. 1654–1662.

[23] S. Kumar, S. Gil, D. Katabi, and D. Rus, "Accurate indoor localization with zero start-up cost," in *Proceedings of MOBICOM*. ACM, 2014, pp. 483–494.

[24] J. Lee Rodgers and W. A. Nicewander, "Thirteen ways to look at the correlation coefficient," *The American Statistician*, vol. 42, no. 1, pp. 59–66, 1988.

[25] Z. Li, Y. Zhu, H. Zhu, and M. Li, "Compressive sensing approach to urban traffic sensing," in *Processings of ICDCS*. IEEE, 2011, pp. 889–898.

[26] C. Luo, F. Wu, J. Sun, and C. Chen, "Compressive data gathering for large-scale wireless sensor networks," in *Proceedings of MOBICOM*. ACM, 2009, pp. 145–156.

[27] M. A. McHenry, P. A. Tenhula, D. McCloskey, D. A. Roberson, and C. S. Hood, "Chicago spectrum occupancy measurements & analysis and a long-term studies proposal," in *Proceedings of the first international workshop on Technology and policy for accessing spectrum*. ACM, 2006, p. 1.

[28] H. S. Park and C. H. Jun, "A simple and fast algorithm for k-medoids clustering," *Expert Systems with Applications*, vol. 36, no. 2, pp. 3336–3341, 2009.

[29] S. Rallapalli, L. Qiu, Y. Zhang, and Y. Chen, "Exploiting temporal stability and low-rank structure for localization in mobile networks," in *Proceedings of MOBICOM*. ACM, 2010, pp. 161–172.

[30] B. Recht, M. Fazel, and P. A. Parrilo, "Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization," *SIAM review*, vol. 52, no. 3, pp. 471–501, 2010.

[31] B. Recht, W. Xu, and B. Hassibi, "Necessary and sufficient conditions for success of the nuclear norm heuristic for rank minimization," in *Proceedings of CDC*. IEEE, 2008, pp. 3065–3070.

[32] J. Tanner and K. Wei, "Low rank matrix completion by alternating steepest descent methods," *Applied and Computational Harmonic Analysis*, vol. 40, no. 2, pp. 417–429, 2016.

[33] R. Tibshirani, G. Walther, and T. Hastie, "Estimating the number of clusters in a data set via the gap statistic," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 63, no. 2, pp. 411–423, 2001.

[34] H. Urkowitz, "Energy detection of unknown deterministic signals," *Proceedings of the IEEE*, vol. 55, no. 4, pp. 523–531, 1967.

[35] M. Vossiek, L. Wiebking, P. Gulden, J. Wieghardt, C. Hoffmann, and P. Heide, "Wireless local positioning," *Microwave Magazine, IEEE*, vol. 4, no. 4, pp. 77–86, 2003.

[36] J. Wang, S. Tang, B. Yin, and X. Li, "Data gathering in wireless sensor networks through intelligent compressive sensing," in *Proceedings of INFOCOM*. IEEE, 2012, pp. 603–611.

[37] X. Wu and M. Liu, "In-situ soil moisture sensing: measurement scheduling and estimation using compressive sensing," in *Proceedings of IPSN*. ACM, 2012, pp. 1–12.

[38] X. Wu, P. Yang, T. Jung, Y. Xiong, and X. Zheng, "Compressive sensing meets unreliable link: sparsest random scheduling for compressive data gathering in lossy WSNs," in *Proceedings of MOBIHOC*. ACM, 2014, pp. 13–22.

[39] Y. Xing, R. Chandramouli, and S. Mangold, "Dynamic spectrum access in open spectrum wireless networks," *IEEE Journal on Selected Areas in Communications*, vol. 24, no. 3, pp. 626–637, 2006.

[40] S. Yin, D. Chen, Q. Zhang, M. Liu, and S. Li, "Mining spectrum usage data: a large-scale spectrum measurement study," *IEEE Transactions on Mobile Computing*, vol. 11, no. 6, pp. 1033–1046, 2012.

[41] X. Ying, J. Zhang, L. Yan, G. Zhang, M. Chen, and R. Chandra, "Exploring indoor white spaces in metropolises," in *Proceedings of MOBICOM*. ACM, 2013, pp. 255–266.

[42] T. Yücek and H. Arslan, "A survey of spectrum sensing algorithms for cognitive radio applications," *Communications Surveys & Tutorials, IEEE*, vol. 11, no. 1, pp. 116–130, 2009.

[43] T. Zhang, N. Leng, and S. Banerjee, "A vehicle-based measurement framework for enhancing whitespace spectrum databases," in *Proceedings of MOBICOM*. ACM, 2014, pp. 17–28.

[44] Y. Zhang, M. Roughan, W. Willinger, and L. Qiu, "Spatio-temporal compressive sensing and internet traffic matrices," in *Proceedings of SIGCOMM*, vol. 39, no. 4. ACM, 2009, pp. 267–278.

[45] H. Zheng, S. Xiao, X. Wang, and X. Tian, "Energy and latency analysis for in-network computation with compressive sensing in wireless sensor networks," in *Proceedings of INFOCOM*. IEEE, 2012, pp. 2811–2815.