Adaptive Multi-Task Compressive Sensing for Localization in Wireless LANs

Rongpeng Li*, Zhifeng Zhao*, Yuan Zhang*, Jacques Palicot† and Honggang Zhang*‡

*Department of Information Science and Electronic Engineering
Zhejiang University, Zheda Road 38, Hangzhou 310027, China
Email: {lirongpeng, zhaozf, yuanzhang}@zju.edu.cn

†Supélec/IETR, Avenue de la Boulaie CS 47601, Cesson-Sévigné Cedex, France
Email: jacques.palicot@supelec.fr

‡Université Européenne de Bretagne (UEB) & Supélec, Avenue de la Boulaie CS 47601, Cesson-Sévigné Cedex, France
Email: honggang.zhang@supelec.fr

Abstract

The spatially distributed sparsity of the mobile devices (MDs) in indoor wireless local area networks (WLANs) makes compressive sensing (CS) based localization algorithms feasible and desirable. In this paper, we exploit the most recent developments in CS to efficiently perform localization in WLANs and design an accurate indoor localization scheme by taking advantage of the theory of multi-task Bayesian compressive sensing (MBCS). The proposed scheme assembles the strength measurements of signals from the MDs to distinct access points (APs) and jointly utilizes them at a central unit or a specific AP to achieve localization, thus being able to alleviate the burden of MDs while simultaneously giving a precise estimation of the locations. Afterwards, we give a deeper insight into the localization problem in more practical scenarios with varying number of MDs and investigate two different adaptive algorithms to meet the satisfactory localization error requirement. Compared to the conventional MBCS algorithms, simulation results validate that both adaptive algorithms could provide superior localization accuracy and exhibit stronger resilience to the changes in the number of MDs.

Index Terms

Bayesian compressive sensing, Localization, WLANs, Multi-task, Received signal strength, Adaptivity

I. INTRODUCTION

A. Motivation

Accompanying with the explosive advancement of multimedia-rich networking data services, location-based services (LBSs) have emerged in recent years [1], [2] and attracted considerable interest and wide applications in various fields, such as on-line social network [3], remote healthcare [4], e-commerce [5], personalized information delivery [6] and so on. Furthermore, location and mobility information is also being exploited to reduce the energy consumption in energy-constrained wireless communication [2], [7] and becomes essential for mobile user management in seamless and ubiquitous communications. As a result, researchers come up with numerous techniques to achieve localization, with the aid of global position system (GPS) [8] and base stations [9], etc. Unfortunately, the cellular-based methods can not provide adequate accuracy required in indoor applications [10]. On the other hand, though widely used in outdoor environment, GPS is often energy-consumable and not suitable for central urban or indoor area with heavy building shadowing [11]. In this light, owing to the ever-growing universal existence of WLANs, some novel localization schemes have been proposed by relying on the received signal strength (RSS)[1] from wireless local area networks (WLANs) to estimate the locations of mobile devices (MDs). But, due to the effect of noise and channel impairments, there exists a challenge to cope with the uncertainty in RSS measurements [16]. So, a Bayesian approach [17] could be exploited by firstly imposing a probabilistic model on the RSS measurements, and then trying to obtain the posterior distribution.

In this paper, we assume that the indoor positioning system can collect the strength measurements of signals from MDs to access points (APs) in the area of interest and assemble them at a network central unit (CU) or a specific AP to perform the localization in WLANs. This methodology could provide several advantages. Firstly, the algorithms running on a CU could

[1]In spite of the possibility to adopt other positioning metrics like time of arrival (ToA) and angle of arrival (AoA) [8], [10], RSS is generally the featured choice in the context of localization [9], [12]–[16].

alleviate the burden of MDs, which usually have limited processing power, short battery lifetime and small memory [18]. Secondly, the location information could be more conveniently applied, since APs or LBS servers could directly utilize it to optimize the whole system. Besides, since the MDs are only located at few places of one large physical space, the locations of MDs can be regarded as sparse signals if we view them as a whole. Therefore, the centralized localization methodology could benefit from a plethora of existing CS algorithms [19]–[21], which explore the fact that a small collection of an originally sparse or compressible signal's linear projections contains sufficient information for signal recovery and thus require far fewer measurements than the Nyquist sampling theorem to accurately reconstruct the signal.

Motivated by the discussions above, we design a multi-task Bayesian compressive sensing (multi-task BCS, MBCS) [22] based localization scheme. The term "multi-task" implies that this scheme would take advantage of the intra- and inter- signal correlations of the RSS measurements at different APs. As a result, it could decrease the total amount of data required for accurate localization. Besides, BCS algorithms are able to provide a criterion named "*error bars*" [23] to gauge the accuracy of the reconstruction vector and make the adaptation of measurement number possible [24]. Following our previous works [25], we propose two adaptive multi-task BCS algorithms to guarantee the effectiveness and accuracy when the number of MDs vary.

B. Related Works

Thus far, there exists a substantial body of works towards the localization problem in WLANs. In [12], Rizos *et al*. collected the RSS measurements and compared them to a pre-built radio environment map. The estimated position would be determined by the k-nearest neighbors (KNN) method, namely the point with the smallest Euclidean distance to the centroid of k nearest neighbors in the map. In [10], Kushki *et al*. modified the KNN scheme and proposed a kernel-based technique to improve the positioning accuracy. As for CS-based localization schemes, Zhang *et al*. [14] showed the corresponding feasibility theoretically. Meanwhile, researchers implemented their CS-based localization schemes on experimental networks [15], [16]. For example, [16] utilized CS to localize the position of one MD and exploited Kalman filtering to track the corresponding movement. However, these schemes [14]–[16] suffer from the uncertainty in RSS measurements and could not adapt to the varying number of MDs. Besides, they are still carried out on the MDs by averaging the RSS values, incurring indispensable yet energy inefficient pre-processing procedures such as AP selection [14] and orthogonalization [15]. [18] intended to bypass these procedures by addressing the localization problem on the AP side. However, the authors did not consider the intra- and inter- signal correlations in the RSS measurements and provided no criterion to determine the sufficient number of measurements as well.

The rest of this paper is structured as follows. In Section II, we give an introduction to the fundamentals of CS theory and describe the relevant MBCS framework. Section III presents the system model and formulates our compressed sensing based localization scheme. Section IV and Section V provide two adaptive algorithms and present the corresponding simulation results. Finally, Section VI concludes our works and offers our future research direction.

II. Basics of Bayesian Compressive Sensing

A. The Compressive Sampling Process

The Nyquist sampling theorem demonstrates that signals, images, videos, and other data can be exactly recovered from a set of uniformly spaced samples taken at the so-called Nyquist rate, which is twice the highest frequency in the signal of interest [26]. But the resulting Nyquist rate in most situations is redundant such that we end up with far more necessary samples. The recently developed theory of compressive sensing (CS) states that if a real-world signal has a sparse representation in a certain transform bases, then it's possible to recover the signal with significantly fewer samples or measurements as required by the Nyquist rate. In this section, we give a brief review of the theory of CS.

Consider a signal $\boldsymbol{x} \in \mathbb{R}^{N\times 1}$, which can be represented in terms of $N \times N$ transform basis $\boldsymbol{\Psi}$ such that

$$\boldsymbol{x} = \sum_{i=1}^{N} s_i \boldsymbol{\psi}_i = \boldsymbol{\Psi}\boldsymbol{s}. \tag{1}$$

The signal $\boldsymbol{x}$ is called K-sparse if its sparse representation $\boldsymbol{s}$ has at most K non-zero entries ($\|\boldsymbol{s}\|_{l_0} \leq K$), where $K \ll N$ and the l_0-norm $\|\boldsymbol{s}\|_{l_0}$ of the vector $\boldsymbol{s}$ is defined as the number of its non-zero components.

Consider a linear projection process that computes $M < N$ inner products between $\boldsymbol{x}$ and a set of $N \times 1$ vectors $\{\boldsymbol{\phi}_j\}_{j=1}^{M}$ as in $y_j = \boldsymbol{\phi}_j^T \boldsymbol{x}$, where $\{\cdot\}^T$ represents the transpose operation. Collect the measurements y_j and form an $M \times 1$ vector $\boldsymbol{y}$, by arranging the projection vectors

$\{\phi_j^T\}_{j=1}^M$ as rows of an $M \times N$ measurement matrix $\mathbf{\Phi}$. After substituting $\boldsymbol{x}$ with Eq. (1), the whole projection process can be represented as follows.

$$\boldsymbol{y} = \mathbf{\Phi}\boldsymbol{x} + \boldsymbol{\varepsilon} = \mathbf{\Phi}\mathbf{\Psi}\boldsymbol{s} + \boldsymbol{\varepsilon} = \mathbf{\Theta}\boldsymbol{s} + \boldsymbol{\varepsilon}, \tag{2}$$

where $\mathbf{\Theta} = \mathbf{\Phi}\mathbf{\Psi}$ is an $M \times N$ matrix, and $\boldsymbol{\varepsilon} = [\varepsilon_1, \varepsilon_2, ...\varepsilon_N]^T$ represents the noisy environment effects with each entry ε_i being a zero mean Gaussian random variable with variance σ^2. The CS theory aims to solve this under-determined problem and find an appropriate $\boldsymbol{x}$ satisfying Eq. (2).

B. Recovery Solutions in Compressive Sensing

In this section, we detail the recovery solutions in CS. Actually, traditional techniques like least square algorithms [27] could not readily handle the under-determined problem in Eq. (2), since the knowledge of the non-zero element positions of $\boldsymbol{x}$ is usually unknown beforehand. Hence, rather than requiring the knowledge of zeros element positions, the recovery algorithms in CS solve this under-determined problem by the means of greedy algorithms [28], [29].

[30] states that if $M \geq 2K$, the recovery problem, under specific conditions, could be treated as follows.

$$\hat{\boldsymbol{s}} = \arg\min_{\boldsymbol{s}} \|\boldsymbol{s}\|_{l_0} \quad s.t. \quad \|\boldsymbol{y} - \mathbf{\Phi}\mathbf{\Psi}\boldsymbol{s}\|_{l_2} \leq \epsilon. \tag{3}$$

However, Eq. (3) requires to solve a non-convex problem which is usually *NP-hard*. In [31], it has been proved that we can approximate the aforementioned l_0 optimization problem by its l_1 relaxation[2] with a bounded error only if the matrix $\mathbf{\Theta} = \mathbf{\Phi}\mathbf{\Psi}$ satisfies the restricted isometry property (RIP).

$$\hat{\boldsymbol{s}} = \arg\min_{\boldsymbol{s}} \|\boldsymbol{s}\|_{l_1} \quad s.t. \quad \|\boldsymbol{y} - \mathbf{\Phi}\mathbf{\Psi}\boldsymbol{s}\|_{l_2} \leq \epsilon. \tag{4}$$

Here, a matrix $\mathbf{\Theta}$ is said to conform to the RIP for order K with constant $\delta_K \in (0, 1)$ if

$$(1 - \delta_K) \|\boldsymbol{c}\|_{l_2}^2 \leq \|\mathbf{\Theta}\boldsymbol{c}\|_{l_2}^2 \leq (1 + \delta_K) \|\boldsymbol{c}\|_{l_2}^2 \tag{5}$$

for any N-length vector $\boldsymbol{c}$ satisfying $\|\boldsymbol{c}\|_{l_0} \leq K$. While designing and evaluating a suitable measurement matrix $\mathbf{\Phi}$ to ensure the RIP for $\mathbf{\Theta} = \mathbf{\Phi}\mathbf{\Psi}$ is challenging, there fortunately exists a

[2]For an integer real number $p \geq 1$ and a vector $\boldsymbol{z}$, the corresponding l_p-norm $\|\boldsymbol{z}\|_{l_p}$ equals $\left(\sum_i |z_i|^p\right)^{\frac{1}{p}}$.

large set of matrices that have proven to obey the RIP. Generally, if $\mathbf{\Phi}$ is a random matrix with identical independently distributed (i.i.d.) Gaussian or Bernoulli entries or a matrix made up of randomly selected rows of an orthogonal matrix (e.g. the DFT), the RIP would be satisfied [32]. Typically, conditioned on the RIP, the recovery problem in Eq. (4) could be solved by techniques such as basis pursuit (BP) [29], orthogonal matching pursuit (OMP) [28] and OMP's variants [14]. While recovery through l_1 optimization proves to provide highly accurate solutions, the algorithms above often incur high computational complexity as well.

On the other hand, the problem could be solved from a Bayesian prospective as well. Compared with the aforementioned greedy algorithms, Bayesian compressed sensing (BCS) [23] recovers the signal based on the posterior probability instead of a single value and could yield better performance in reconstruction with noisy measurements $\boldsymbol{y}$ in terms of l_0-norm [33]. Therefore, we apply the BCS algorithm to our localization problem. Moreover, in order to combat the negative fading and shadowing effect on measurements and achieve even better recovery performance, we take into account the intra- and inter- signal correlations and adopt a multi-task BCS (MBCS) framework [22]. Specifically, let's assume there are P sets of CS measurements $\{\boldsymbol{y}^i\}_{i=1}^P$, projected from P sets of original compressive signals $\{\boldsymbol{x}^i\}_{i=1}^P$, namely

$$\boldsymbol{y}^i = \mathbf{\Phi}^i \boldsymbol{x}^i + \boldsymbol{\varepsilon}^i = \mathbf{\Phi}^i \mathbf{\Psi}^i \boldsymbol{s}^i + \boldsymbol{\varepsilon}^i = \mathbf{\Theta}^i \boldsymbol{s}^i + \boldsymbol{\varepsilon}^i, \forall i \in \{1, \cdots, P\}, \tag{6}$$

where each $\boldsymbol{x}^i \in \mathbb{R}^N$ exploits a disparate random measurement matrix $\mathbf{\Phi}^i \in \mathbb{R}^{M^i \times N}$ to derive $\boldsymbol{y}^i \in \mathbb{R}^{M^i}$, with every sparse vector in $\{\boldsymbol{s}^i\}_{i=1}^P$ similar or equal to each other. $\boldsymbol{\varepsilon}^i \in \mathbb{R}^{M^i}$ denotes Gaussian noise and if $P = 1$, Eq. (6) would be simplified into Eq. (2), namely a single task case. Obtained from repeated P experiments on similar scenarios or the same type of tasks, MBCS algorithm could collect a subset of highly correlated measurements, which is exactly suitable for the case of RSS-gathering process in our localization scheme and contributes to the information-sharing between tasks.

Commonly, $\boldsymbol{\varepsilon}^i$ can be modeled as an M^i i.i.d. zero mean multivariate Gaussian random variable with variance σ_0^2. As a result, conditioned on $\boldsymbol{s}^i$ and $\lambda_0 = 1/\sigma_0^2$, the likelihood function for $\boldsymbol{y}^i$ in Eq. (6) could be formulated as

$$p(\boldsymbol{y}^i|\boldsymbol{s}^i, \lambda_0) = (2\pi/\lambda_0)^{-M^i/2} \exp(-\frac{\lambda_0}{2} \left\|\boldsymbol{y}^i - \mathbf{\Theta}^i \boldsymbol{s}^i\right\|_{l_2}^2), \forall i \in \{1, \cdots, P\}. \tag{7}$$

The difficulty towards applying BCS lies in the typically intractable computation of the posterior probability function [34]. Hence, we are forced to accept some approximations and establish a hierarchical prior to the sparse vectors [23]. In MBCS, we just assign a common prior with common hyperparameters $\boldsymbol{\lambda} = [\lambda_1, \cdots, \lambda_N]$ to $\boldsymbol{s}^i$ and connect several recovery tasks together [23]:

$$p(\boldsymbol{s}^i|\boldsymbol{\lambda}) = \prod_{j=1}^{N} \mathcal{N}(\boldsymbol{s}_j^i|0, \lambda_j^{-1}), \forall i \in \{1, \cdots, P\}, \tag{8}$$

where $\mathcal{N}(\cdot|0, \lambda_j^{-1})$ represents the Gaussian distribution with zero mean and variance λ_j^{-1}. By the Bayes' rule, the posterior probability function for $\{\boldsymbol{s}^i\}_{i=1}^P$ can be derived by using Eq. (7) and Eq. (8), namely

$$\begin{aligned} p(\boldsymbol{s}^i|\boldsymbol{y}^i, \boldsymbol{\lambda}, \lambda_0) &= \frac{p(\boldsymbol{y}^i|\boldsymbol{s}^i, \lambda_0)p(\boldsymbol{s}^i|\boldsymbol{\lambda})}{p(\boldsymbol{y}^i|\boldsymbol{\lambda}, \lambda_0)} \\ &= (2\pi)^{-\frac{N+1}{2}} \left|\boldsymbol{\Sigma}^i\right|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(\boldsymbol{s}^i - \boldsymbol{\mu}^i)^T(\boldsymbol{\Sigma}^i)^{-1}(\boldsymbol{s}^i - \boldsymbol{\mu}^i)\right) \end{aligned} \tag{9}$$

with mean and covariance given by

$$\boldsymbol{\mu}^i = \lambda_0 \boldsymbol{\Sigma}^i \left(\boldsymbol{\Theta}^i\right)^T \boldsymbol{y}^i, \quad \boldsymbol{\Sigma}^i = (\lambda_0 \left(\boldsymbol{\Theta}^i\right)^T \left(\boldsymbol{\Theta}^i\right) + \boldsymbol{\Lambda})^{-1}, \quad \boldsymbol{\Lambda} = \mathrm{diag}(\lambda_1, \lambda_2, ..., \lambda_N). \tag{10}$$

Subsequently, we have to estimate the hyperparameters $\boldsymbol{\lambda}$ and λ_0 in order to obtain the posterior probability function. The estimates for $\boldsymbol{\lambda}$ and λ_0 can be derived by maximizing the logarithm marginal likelihood

$$\begin{aligned} L(\boldsymbol{\lambda}, \lambda_0) &= \sum_{i=1}^{P} \log p(\boldsymbol{y}^i|\boldsymbol{\lambda}, \lambda_0) = \sum_{i=1}^{P} \log \int p(\boldsymbol{y}^i|\boldsymbol{s}^i, \lambda_0)p(\boldsymbol{s}^i|\boldsymbol{\lambda})d\boldsymbol{s}^i \\ &= -\frac{1}{2}\sum_{i=1}^{P} \left[M^i \log 2\pi + \log\left|\boldsymbol{C}^i\right| + (\boldsymbol{s}^i)^T(\boldsymbol{C}^i)^{-1}\boldsymbol{s}^i\right] \end{aligned} \tag{11}$$

with the unit matrix $\boldsymbol{I}$ and

$$\boldsymbol{C}^i = \lambda_0^{-1}\boldsymbol{I} + \boldsymbol{\Theta}^i\boldsymbol{\Lambda}^{-1}\left(\boldsymbol{\Theta}^i\right)^T, \forall i \in \{1, \cdots, P\}. \tag{12}$$

Afterwards, differentiate Eq. (11) with respect to $\boldsymbol{\lambda}$ and λ_0 respectively and then set the results to zero, yielding

$$\lambda_j^{\text{new}} = \frac{P - \lambda_j \sum_{i=1}^{P} \Sigma_{(j,j)}^i}{\sum_{i=1}^{P} \left(\mu_j^i\right)^2}, \forall j \in \{1, 2, \cdots N\} \tag{13}$$

and

$$\lambda_0^{\text{new}} = \frac{\sum_{i=1}^{P} \left(M^i - N + \sum_{j=1}^{N} \lambda_j \Sigma_{(j,j)}^i\right)}{\sum_{i=1}^{P} \left\|\boldsymbol{y}^i - \boldsymbol{\Theta}^i\boldsymbol{\mu}^i\right\|_{l_2}^2}. \tag{14}$$

Note that $\boldsymbol{\lambda}^{\text{new}}$ and λ_0^{new} are functions of $\{\boldsymbol{\mu}^i\}_{i=1}^P$ and $\{\boldsymbol{\Sigma}^i\}_{i=1}^P$, while $\{\boldsymbol{\mu}^i\}_{i=1}^P$ and $\{\boldsymbol{\Sigma}^i\}_{i=1}^P$ are pertinent to $\boldsymbol{\lambda}$ and λ_0 in Eq. (10). So we can alternatively iterate between Eq. (10), Eq. (13) and

Eq. (14) until convergence. In this process, [34][3] notes that in order to maximize the likelihood in Eq. (11), some entries of $\boldsymbol{\lambda}$ tend to infinity, which in turn means the corresponding entries in $\boldsymbol{\mu}^i$ tightly approach zero. In other words, the algorithm finally find sparse vectors $\{\boldsymbol{s}^i\}_{i=1}^P$ and could estimate the location vector in terms of the average of $\{\boldsymbol{\mu}^i\}_{i=1}^P$. Besides, since the update of $\boldsymbol{\lambda}^{\text{new}}$ requires the knowledge of the whole set $\{\boldsymbol{\mu}^i\}_{i=1}^P$ and $\{\boldsymbol{\Sigma}^i\}_{i=1}^P$, MBCS algorithm implicitly exploits the intra- and inter- correlations in the RSS signal measurements.

III. Bayesian Compressive Sensing Based Localization Scheme in WLANs

A. System Model and Problem Description

In this paper, we primarily focus on a localization scenario with a set of N non-overlapping grids. Meanwhile, the area of interest is covered by P APs and K MDs equipped with WLAN adapters. As mentioned in Section I, the locations of the MDs can be estimated by comparing the current RSS measurements with a pre-set radio environment map of this area by the central unit (CU) in the "backbone" network. Here, the radio environment map, which is assumed to be readily generated, refers to a table of pre-measured RSS readings of a similar device corresponding to every grid of the area.

For clarity of representation, let an N-length vector $[0, 0, 1 \cdots]^T \in \mathbb{R}^{N \times 1}$ denote the logical location of an MD. To be specific, if the MD is located in the i-th ($i \in \{1, \cdots, N\}$) grid, the corresponding i-th element in the location vector would equal "1". Therefore, if there exist $K \ll N$ MDs in the area of interest, the location vector $\boldsymbol{s} \in \mathbb{R}^{N \times 1}$, which needs to be estimated, contains K entries equalling 1. Thus the location estimation problem could be reformulated into a sparse signal recovery problem, due to the facts that an MD can be reasonably located in exactly one of the grids within the whole area of interest and the number of MDs is very comparatively smaller than that of the grids in aggregate.

Although a number of available methods can be utilized to carry out the comparison between the RSS measurements and the radio environment map, we choose the BCS-based scheme to incorporate the inherent sparsity feature of the localization problem and seek for the best matching, which will be verified to be more accurate and more efficient, lately.

[3]In [34], the authors details why this sparseness could be automatically determined from a variational approximation perspective.

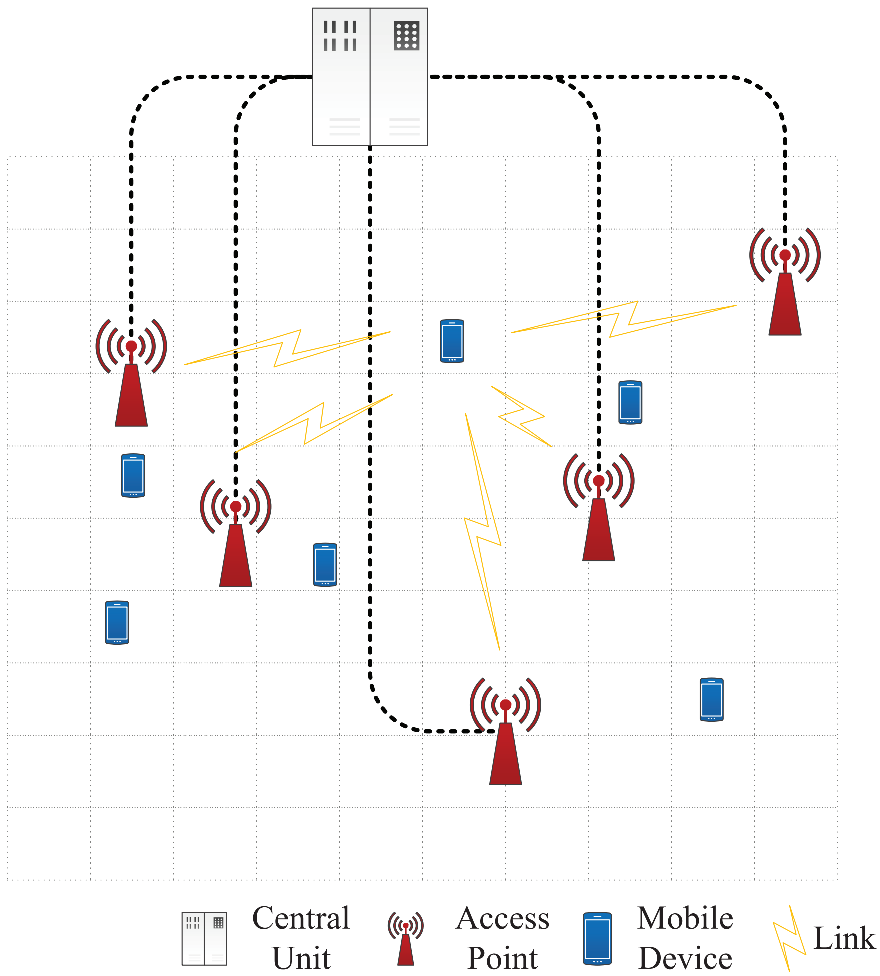


Fig. 1. Localization scenario in wireless local area networking environment.

B. Bayesian Compressive Sensing Based Localization

Given the property of sparseness in the localization scenario, the CS based algorithms or their variants can be applied here. In this Section, we would demonstrate how our BCS based localization scheme could accurately locate the mobile devices in a noisy environment with an optimal number of measurements.

In our BCS based scheme, instead of the common transform bases such as discrete Fourier transform (DFT) basis and discrete wavelet transform (DWT) basis, we finally decide to choose the radio environment map matrices $\{\boldsymbol{\Psi}^i\}_{i=1}^P$ as the projection matrices, after taking into account the obvious sparse structure of the corresponding location vector $\boldsymbol{s}$. For each column vector of the radio environment map matrix $\boldsymbol{\Psi}^i$, $\boldsymbol{\psi}_j^i \in \mathbb{R}^N$ denote the vector of collected RSS samples at grid j from the i-th AP, wherein N indicates the sampling length and entries of $\boldsymbol{\psi}_j^i$ should be different due to the fading and shadowing effect. By means of concatenating the vectors for all the N grids, a CU will construct a single matrix $\boldsymbol{\Psi}^i \in \mathbb{R}^{N\times N}$ for the i-th AP. Therefore, the actual signals at the AP side can be expressed as a linear combination of several columns of $\boldsymbol{\Psi}^i$. In other words, the received signals $\boldsymbol{x}^i$ are the multiplication product of the radio map matrix $\boldsymbol{\Psi}^i$ and the $N \times 1$ location vector $\boldsymbol{s}$, which means that $\{\boldsymbol{s}^i\}_{i=1}^P$ in MBCS are the same and equal $\boldsymbol{s}$.

Now, assume that every AP has collected a train of RSS measurements from MDs at unknown locations. Then, considering the tremendous amount of measurements and the corresponding inherent sparsity property, each AP, say AP i would sample M^i measurements out of $\boldsymbol{x}^i \in \mathbb{R}^N$ from every MD according to the measurement matrix $\boldsymbol{\Phi}^i \in \mathbb{R}^{M^i\times N}$ and feedback them to the CU via backhaul link. When the incoherent measurement process completes, the overall RSS measurements for MDs associated with the i-th AP can be described using Eq. (6) at the CU side, where P is the number of APs in this area and $\boldsymbol{\varepsilon}^i$ represents the Gaussian noise. $\boldsymbol{\Phi}^i$ takes the standard Gaussian matrix with its columns normalized to unit norm, thus guaranteeing the required incoherence between $\boldsymbol{\Phi}^i$ and $\boldsymbol{\Psi}^i$ to meet the RIP.

Finally, the P sets of measurements $\boldsymbol{y}^i$ can be used jointly to recover the location vector. Therefore, we can simply take advantage of MBCS in Section II to solve this reconstruction problem. After completing reconstruction process, the locations of MDs can be precisely determined based on the average of $\{\boldsymbol{\mu}^i\}_{i=1}^P$.

IV. Adaptive Multi-task Compressive Sensing Algorithm for Localization in WLANs

The previous sections have addressed the means of BCS based localization scheme in WLANs. Yet, the varying number of MDs in mobility oriented networks makes it challenging to accurately determine the required measurement number in prior and thus suffering to directly adopt the BCS based localization scheme. Meanwhile, the criterion, which measures the localization error and balances whether the current measurement is sufficiently precise, is worthy to know. But it is still unavailable in the previous localization researches. In this section, we take into the aforementioned considerations and aim to tackle how to achieve the optimal measurement number for location reconstruction, by deriving an adaptive multi-task BCS algorithm, or the AMBCS algorithm.

Recall that, the location vector $\boldsymbol{s}$ is finally estimated as the average of $\{\boldsymbol{\mu}^i\}_{i=1}^P$ or the mean of a multivariate Gaussian distribution in Eq. (9). Indeed, the distribution also provides a byproduct, namely the covariance $\boldsymbol{\Sigma}^i$ in Eq. (10), which can be interpreted as "*error bars*" in [23], [35]. As the name suggests, the "error bars" could be applied as a metric on the accuracy of the recovered location vector $\hat{\boldsymbol{s}}$ and furthermore as a criterion to decide how many measurements are necessary. Intuitively, if the value of "error bars" is larger than anticipation, it implies that the recovered vector still remains quite uncertain and it needs more RSS measurements to meet the accuracy requirement. On the opposite side, if the "error bars" are smaller than expected, which means the measurements are already sufficient, then we could attempt to decrease the number to reduce the possibly unnecessary redundant measurements.

These insights inspire us to understand the impact of measurement number on the uncertainty for localizations or the "error bars". Beforehand, we need to define the formal definition of "error bars".

Definition 1. For each task $i \in \{1, \cdots, P\}$ in an MBCS algorithm, the "error bars" is defined as the determination of the corresponding covariance $\boldsymbol{\Sigma}_{\mathrm{v}}^i$ in Lemma 2.

Indeed, we observe that the increase in measurement number could theoretically contribute to reduce the "error bars".

Theorem 1. Given a set of hyperparameters $\boldsymbol{\lambda}$ and λ_0, a radio map matrix $\{\boldsymbol{\Psi}^i\}_{i=1}^P$ and M^i measurements $\{\boldsymbol{y}^i\}_{i=1}^P$ generated based on a current measurement matrix $\{\boldsymbol{\Phi}^i\}_{i=1}^P$, for any $i \in \{1, \cdots, P\}$, if another $(M^i+1)^{th}$ measurement $y_{M^i+1}^i$ is sampled out of $\boldsymbol{x}^i$ according to a new measurement vector $\{\boldsymbol{r}^i\}^T$, the uncertainty for recovering the location vector $\boldsymbol{s}$, which is measured by the "error bars" $\det(\boldsymbol{\Sigma}_{\mathrm{v}}^i)$, would certainly decrease.

We leave the proof of Theorem 1 in Appendix. Furthermore, the corresponding proof offers the following corollary.

Corollary 1. In order to decrease the "error bars" to the possibly least, for any $i \in \{1, \cdots, P\}$, the newly added measurement vector $\boldsymbol{r}^i$, on which a new sample is based, should be $\boldsymbol{r}_o^i = \arg\max_{\boldsymbol{r}} \boldsymbol{r}^T \boldsymbol{\Psi}^i (\boldsymbol{\Sigma}^i)(\boldsymbol{\Psi}^i)^T \boldsymbol{r}$.

Remark. Theorem 1 tells that according to the "error bars", it's feasible to exploit the AMBCS algorithm to obtain a satisfactory localization accuracy by dynamically increasing or decreasing the number of measurement. Besides, based on Corollary 1, the AMBCS algorithm could be further modified to decrease the "error bars" at a faster pace. Specifically, in order to achieve the maximum of $(\boldsymbol{r}^i)^T \boldsymbol{\Psi}^i (\boldsymbol{\Sigma}^i)(\boldsymbol{\Psi}^i)^T \boldsymbol{r}$, we can design the new measurement vector $\boldsymbol{r}^T$, by performing an eigen-decomposition of the matrix $\boldsymbol{Q} = \boldsymbol{\Psi}^i (\boldsymbol{\Sigma}^i)(\boldsymbol{\Psi}^i)^T$ and assigning the eigenvector with the largest eigenvalue to $\boldsymbol{r}$. Hence, the rate, at which the uncertainty of the recovered location vector diminishes, could be optimized. By merging this adaptive idea to the localization problem, we could design a greedily adaptive multi-task Bayesian compressive sensing (GAMBCS) algorithm, which could save the computational cost of repetitively adding the measurements and performing the MBCS operations.

Finally, we summarize the AMBCS algorithm for localization in Algorithm 1.

V. Simulation Analysis and Numerical Results

In this section, we verify the localization performance of our proposed BCS-based algorithm and its variants, by simulating in an area of 20m $\times$ 20m with discretized grid size of 1.0m $\times$ 1.0m. Correspondingly, the length of the location vector would be $N = 400$. Moreover, we assume there exist $P = 6$ APs altogether in this Gaussian noise-corrupted environment. For ease of comparison, we assume the number of CS measurements M^i for each AP i (each task i) is equivalent. Meanwhile, we adopt the indoor signal propagation model in [36] to calculate the collected RSS at the AP side. Besides, all the simulation results presented hereafter are an average of 100 independent Monte-Carlo runs.

On the other hand, we evaluate the performance based on the mean localization error (MLE), which is defined between the recovered and corresponding original location vectors (e.g. $\{\hat{\boldsymbol{s}}^i\}_{i=1}^P$ and $\{\boldsymbol{s}^i\}_{i=1}^P$) as follows

$$\text{MLE} = \frac{1}{P} \sum_{i=1}^{P} \frac{\|\hat{\boldsymbol{s}}^i - \boldsymbol{s}^i\|_{l_2}}{\|\boldsymbol{s}^i\|_{l_2}}. \tag{15}$$

Meanwhile, we calculate the "normalized error bars" (NEB) for multi-task BCS algorithm by

Algorithm 1 Adaptive multi-task compressive sensing (AMBCS) algorithm for localization in WLANs

Initialization: For $i \in \{1, 2, \cdots, P\}$, initialize the initial measurement number $M^i = M_0^i$ and generate the radio map matrix by $\boldsymbol{\Psi}^i$ by collecting RSS measurements from MDs at the CU side. Initialize the hyperparameters $\boldsymbol{\lambda}$ and λ_0.

Repeat:

1) APs are assigned a $\boldsymbol{\Phi}^i \in \mathbb{R}^{M^i \times N}$, for $\forall i \in \{1, 2, \cdots, P\}$;
2) APs obtain the measurements vector $\{\boldsymbol{y}^i\}_{i=1}^P$ according to $\boldsymbol{\Phi}^i$ for $\forall i \in \{1, 2, \cdots, P\}$ and feed them back to the CU side via the backhaul link;
3) For $\forall i \in \{1, 2, \cdots, P\}$, CU calculates the mean $\boldsymbol{\mu}^i$ and variance $\boldsymbol{\Sigma}^i$ of posterior probability function $\boldsymbol{s}^i$ by Eq. (9) and Eq. (10) and updates the hyperparameters $\boldsymbol{\lambda}$ and λ_0 by Eq. (13) and Eq. (14);
4) CU notifies AP $i \in \{1, 2, \cdots, P\}$ to adaptively adjust the measurement number:
   a) If the "error bars" det $(\boldsymbol{\Sigma}_{\mathrm{v}}^i)$ exceeds a predefined accuracy requirement E_b, CU notifies AP i to perform a new measurement. Return to Step 2.
   b) Otherwise, CU notifies AP to decrease the measurement number by 1 and remove the last row vector from $\boldsymbol{\Phi}^i$. CU calculates the average of $\{\boldsymbol{\mu}^i\}_{i=1}^P$ and assigns it as the recovered location vector $\hat{\boldsymbol{s}}$.

End Repeat

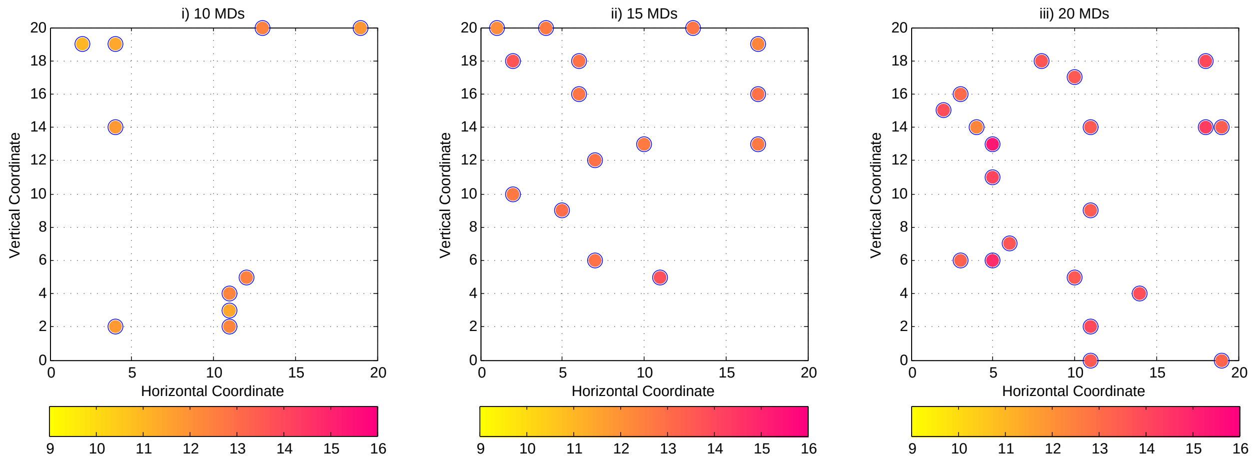
Output: the recovered location vector $\hat{\boldsymbol{s}}$.

Eq. (16) to take into account the localization variance of recovered location vectors

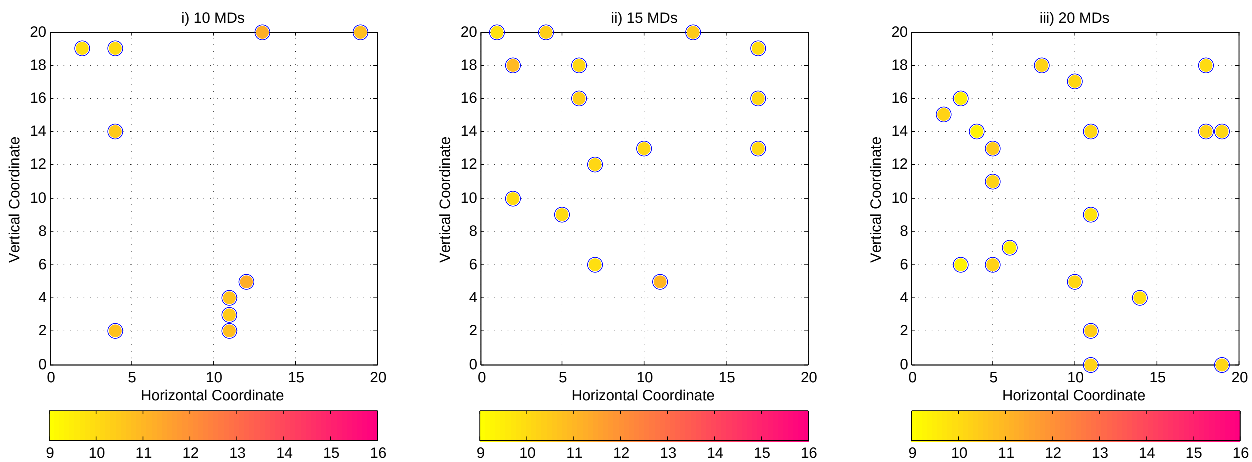
$$\mathrm{NEB} = \frac{1}{P} \sum_{i=1}^{P} \left(\det\left(\boldsymbol{\Sigma}_{\mathrm{v}}^i\right)\right)^{\frac{1}{\|\hat{\boldsymbol{s}}^i\|_{l_0}}}. \tag{16}$$

Here, the $\frac{1}{\|\hat{\boldsymbol{s}}^i\|_{l_0}}$ order of $\det\left(\boldsymbol{\Sigma}^i\right)$ makes it possible to avoid the negative influence incurred by the varying number of MDs.

Firstly, Fig. 2 presents an illustration of localization performance for the MBCS algorithm and AMBCS algorithm. Therein, every AP collects 60 RSS measurements interfered with zero-mean Gaussian noise power ($\sigma_0 = 0.005$) and feedbacks them to the CU. It can be observed that the MBCS algorithm could provide almost precise results. However, due to the increase



(a) MBCS algorithm



(b) AMBCS algorithm

Fig. 2. Localization performance in three scenarios for the MBCS algorithm (top) and the AMBCS algorithm (bottom) . The circles represent the real locations of MDs while the dotted ones represent the estimated locations and the corresponding colour reflects the value of localization variance by Eq. (16).

in localization variance (heavier colour) at the static number of measurements in the MBCS algorithm, the localization confidence would be impaired when more MDs emerge. But, the proposed AMBCS algorithm could offer the result with nearly the same confidence when the number of MDs vary.

Fig. 3 compares the MLE metric-based localization performance among algorithms such as

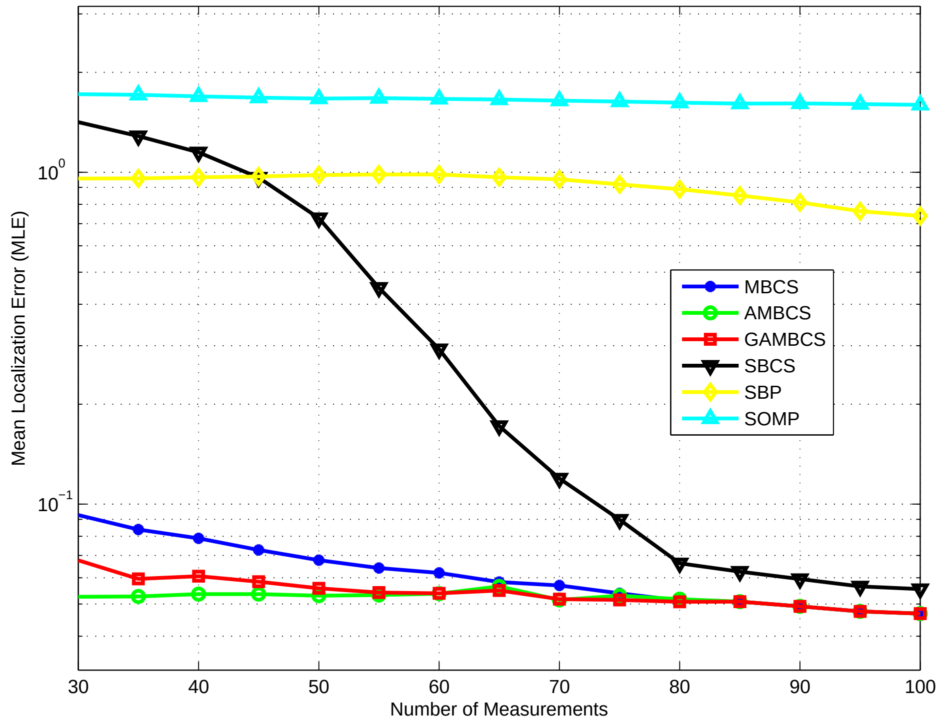


Fig. 3. Localization performance versus varying number of measurements for the algorithms when there exist 10 MDs and noise power $\sigma_0 = 0.005$. For AMBCS and GAMBCS, the initial number of measurements is varying and the NEB is set to be 70.

MBCS, AMBCS, GAMBCS, single-task[4] BCS (SBCS), single-task BP (SBP) and single-task OMP (SOMP), respectively. Fig. 3 shows the localization error decreases by collecting more measurements. Besides, MBCS algorithm could yield a significantly better performance using the same number of measurements by taking advantage of the shared knowledge in inter-AP measurements. Moreover, we also examine the localization performance of both adaptive algorithms when the initial number of measurements is varying and the threshold for NEB is set to be 70. We can find when the number of measurements is small, the adaptive algorithms could provide superior performance than MBCS by dynamically adding more measurements.

Fig. 4 continues the evaluation under the same scenario assumptions using MBCS algorithm

[4]It's worthwhile to note here that the single-task algorithm means that every AP aims to separately recover the location vectors, depending on its own measurements, and the algorithm finally obtains the final estimation result by combining and averaging them.

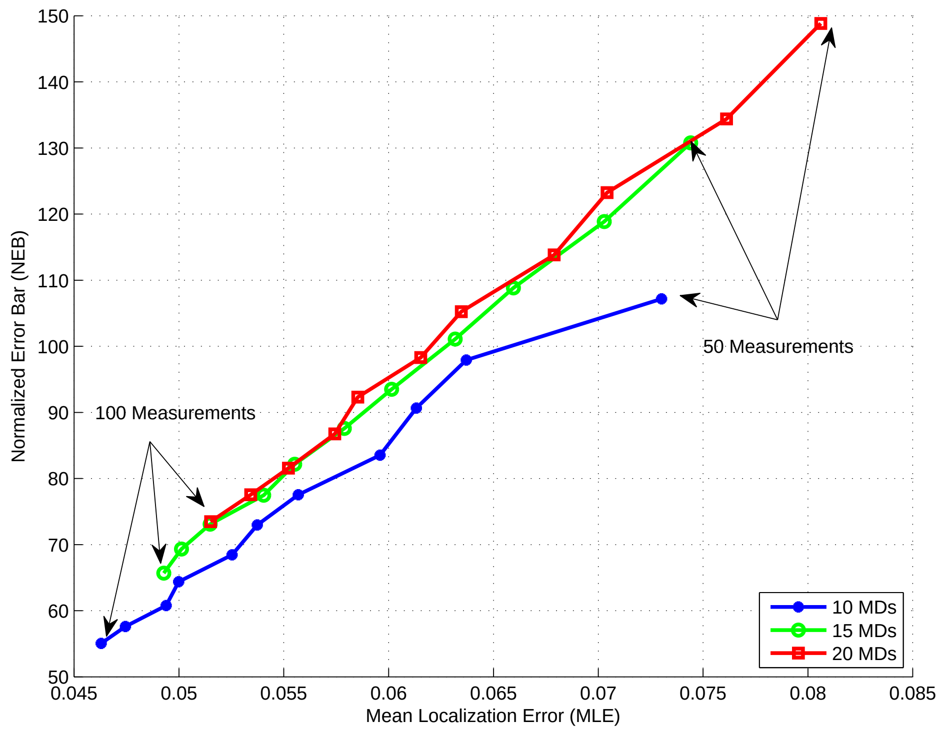


Fig. 4. The relationship between mean location error and "normalized error bars" when the MBCS algorithm is used and noise power $\sigma_0 = 0.005$.

and shows that the MLE performance monotonically varies along with the NEB performance. Besides, it can be observed that the curves under different number of MDs are tightly close while exhibiting the same varying trend. Consequently, given that MLE could not be obtained in practical scenarios, NEB could be a strong alternative and help to determine the necessary number of measurements for a reliable localization performance.

Previous simulations demonstrate the superior localization performance of AMBCS and GAMBCS in the scenarios where the number of MDs is steady. Here, we examine the response of our adaptive algorithms to varying number of MDs. We set the initial number of measurement and NEB to be 50 and 70 respectively. Hence, according to Fig. 4, the default value (e.g. 70) for "error bars" implies that if we adopt the MBCS algorithm the MLE will be below 0.055 and the final number of measurements will require 70 to 90, depending on the exact number of MDs. During the simulation process, we assume another 5 MDs emerge at the 6^{th} slot and disappear at the 11^{th} slot. The corresponding localization performance is plotted with both adaptive algorithms

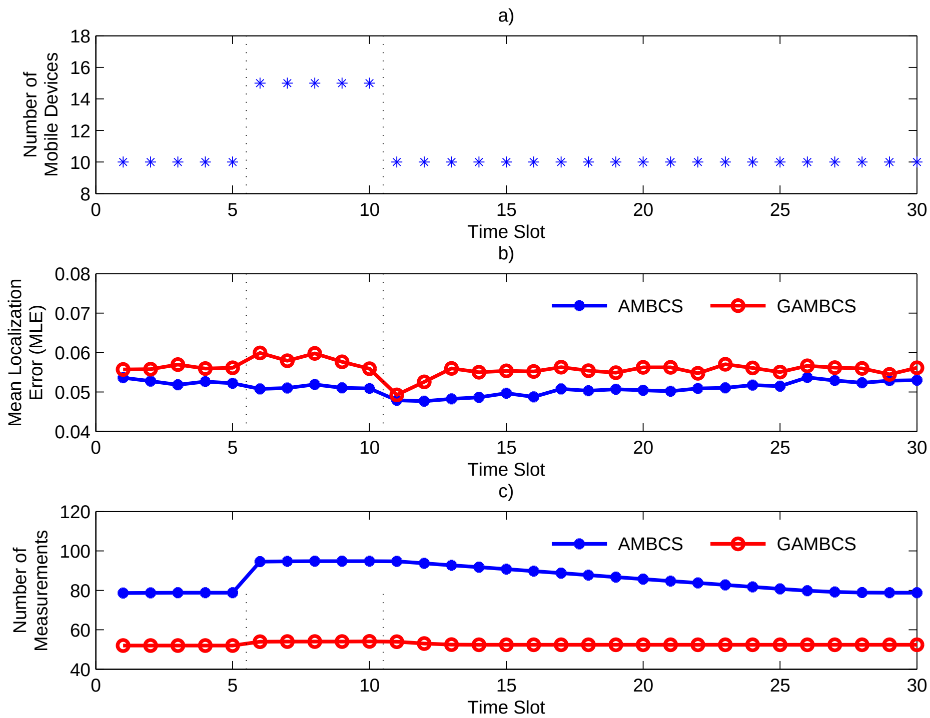


Fig. 5. Localization performance as a function of varying number of mobile devices (MDs) for AMBCS and GAMBCS when initial number of measurements is set to be 50, NEB is set to be 70 and noise power $\sigma_0 = 0.005$.

in Fig. 5. It can be observed in Fig. 5, the adaptivity of both algorithms could be well accustomed to the sudden appearance of MDs without deteriorating the performance. Specifically speaking, the NEB in the localization process would exceed the threshold when more MDs emerge. As a result, every AP will automatically sample more measurements to the CU according the method inside. For example, the AMBCS algorithm requires another 16 measurements to meet the NEB requirement (e.g. 70) and gradually reduce the measurements once they become unnecessary. Meanwhile, due to the greed choice of added observation vector, GAMBCS algorithm only requires another 2 measurements but it offers slightly inferior performance.

In Fig. 6, we examine the impact of working APs on the performance of the AMBCS algorithm. We can find the loss of APs only causes limited negative impact on the corresponding performance. Moreover, AMBCS with fewer APs could still offer superior performance than MBCS, benefiting from the adaptation inside. Fig. 7 presents the localization error under different SNR levels. Each measurement is corrupted by Gaussian noise. The proposed MBCS algorithm

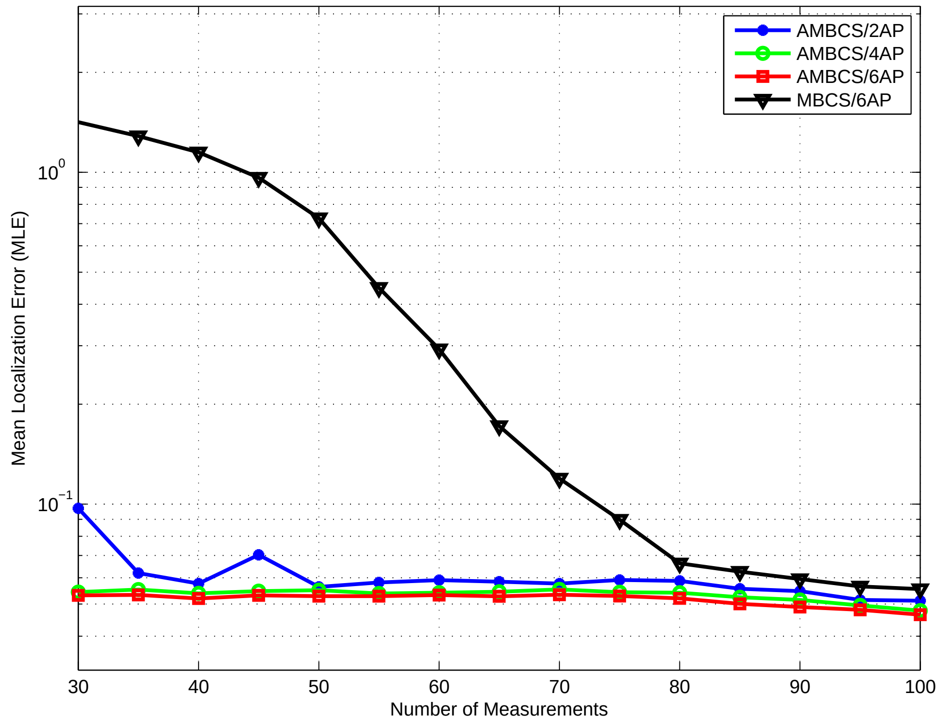


Fig. 6. Localization performance of the AMBCS algorithm versus the number of working APs when there exist 10 MDs and 60 measurements are sampled for the CU.

outperforms much better than the single-task algorithms, especially in the low SNR environment. Moreover, both adaptive algorithms achieve even better results.

VI. Conclusion and Future Works

In this paper, we designed a multi-task BCS localization scheme to estimate the locations of mobile devices more accurately. Specifically we explored the intra- and inter- signal correlations of the RSS measurements and exploited the common structures between different tasks to achieve more precise localization. Moreover, we proposed two adaptive algorithms to dynamically determine the necessary number of RSS measurements, by virtue of the theoretically validated criterion "error bars". The simulation results further verified that both proposed algorithms could achieve superior localization performance, compared to conventional MBCS algorithms.

Though our BCS-based scheme has showed the accuracy and effectiveness by simulations, it's still worthwhile to validate it in practical experimental networks. Therefore, we are dedicated to handle the related meaningful works in the future.

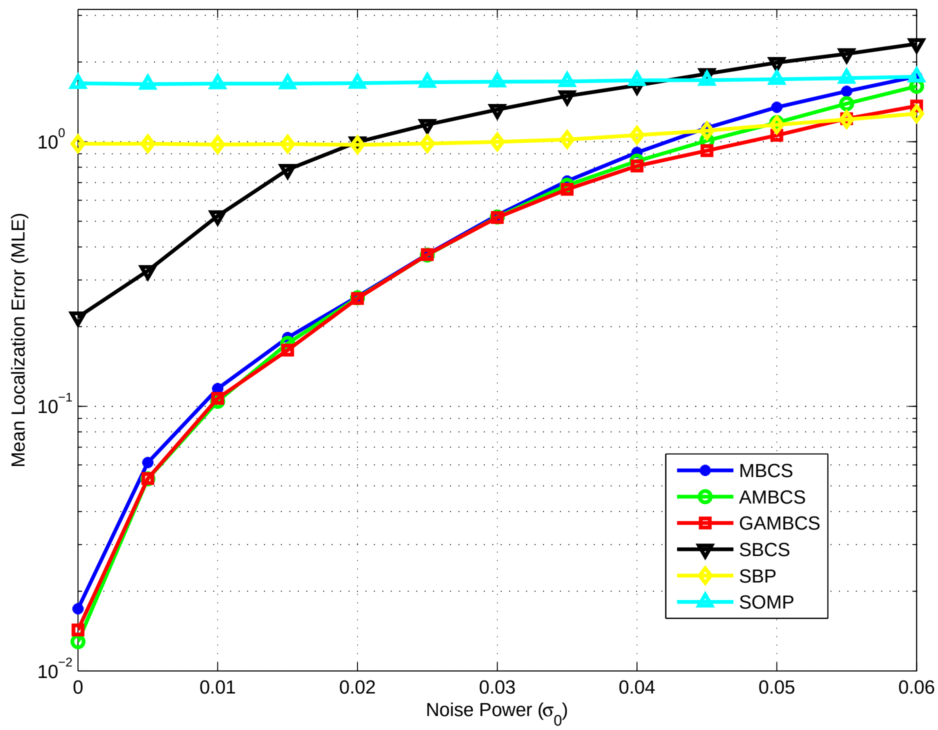


Fig. 7. Localization performance as a function of increasing noise power (σ_0) for MBCS, AMBCS, GAMBCS, SBCS, SBP and SOMP when there exist 10 MDs and 60 measurement are sampled at the CU side.

Acknowledgement

This paper is supported by the National Basic Research Program of China (973Green, No. 2012CB316000), the Key (Key grant) Project of Chinese Ministry of Education (No. 313053), the Key Technologies R&D Program of China (No. 2012BAH75F01), and the grant of "Investing for the Future" program of France ANR to the CominLabs excellence laboratory (ANR-10-LABX-07-01).

Appendix

Proof of Theorem 1 and Corollary 1.

Firstly, we give a lemma in the field of information theory, which bridges the link between covariance and differential entropy [37] and implies why the "error bars" could offer the informational uncertainty of our localization problem.

Lemma 1. For a multivariate Gaussian distribution $\mathcal{N}(\boldsymbol{f})$ with mean value $\boldsymbol{\mu_f}$ and covariance $\boldsymbol{\Sigma_f}$, the differential entropy for a continuous distribution $p(\boldsymbol{f})$, which is defined as $H(\boldsymbol{f}) = -\int p(\boldsymbol{f}) \log p(\boldsymbol{f}) d\boldsymbol{f}$, would equal $H(\boldsymbol{f}) = \frac{1}{2}\log|\boldsymbol{\Sigma_f}| + C$. Here, C denotes a constant independent with the distribution $\mathcal{N}(\boldsymbol{f})$.

Moreover, as detailed in the following lemma, $\{\boldsymbol{\Sigma}^i\}_{i=1}^P$ in Eq. (10) could turn into a positive definitive matrix $\{\boldsymbol{\Sigma}_{\text{v}}^i\}_{i=1}^P$ after limited number of elementary algebraic transforms [38].

Lemma 2. After limited number of elementary algebraic transforms, for any $i \in \{1, \cdots, P\}$, $\boldsymbol{\Sigma}^i$ in Eq. (10) could be reformulated as

$$\boldsymbol{\Sigma}^i \rightarrow \begin{bmatrix} \boldsymbol{\Sigma}_{\text{v}}^i & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}. \tag{17}$$

Proof. Without loss of generality, we omit the superindex i in $\boldsymbol{\Sigma}^i$ during the proof. As discussed in the end of Section II, certain elements in $\boldsymbol{\lambda}$ tend to be infinity [34]. Consequently, after limited elementary transforms $\mathscr{T}$, $\boldsymbol{\Lambda}$ in Eq. (10) could be reformulated as

$$\boldsymbol{\Lambda} \rightarrow \begin{bmatrix} \boldsymbol{\Lambda}_{\text{v}} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Lambda}_{\text{inf}} \end{bmatrix}. \tag{18}$$

Meanwhile, there only exists a set $\boldsymbol{s_v}$ of nonzero recovered entries in $\boldsymbol{s}$. Thus, we could only consider the $|\boldsymbol{s_v}|$ column vectors of transform basis $\boldsymbol{\Psi}$, one of which exactly corresponds to the RSS measurements from one estimated MD location to the AP. By the same elementary transforms $\mathscr{T}$, $\boldsymbol{\Psi}$ can be represented as

$$\boldsymbol{\Psi} \rightarrow \begin{bmatrix} \boldsymbol{\Psi}_{\text{v}} & \boldsymbol{\Psi}_{\text{inf}} \end{bmatrix}. \tag{19}$$

Therefore, after the same transformation $\mathscr{T}$, $\boldsymbol{\Sigma}$ can turn into

$$\begin{aligned}
\boldsymbol{\Sigma} &\rightarrow \left(\lambda_0 \begin{bmatrix} \boldsymbol{\Psi}_{\text{v}}^T \\ \boldsymbol{\Psi}_{\text{inf}}^T \end{bmatrix} \boldsymbol{\Phi}^T\boldsymbol{\Phi} \begin{bmatrix} \boldsymbol{\Psi}_{\text{v}} & \boldsymbol{\Psi}_{\text{inf}} \end{bmatrix} + \begin{bmatrix} \boldsymbol{\Lambda}_{\text{v}} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Lambda}_{\text{inf}} \end{bmatrix} \right)^{-1} \\
&= \left(\begin{bmatrix} \boldsymbol{A} \triangleq \lambda_0 \boldsymbol{\Psi}_{\text{v}}^T \boldsymbol{\Phi}^T\boldsymbol{\Phi}\boldsymbol{\Psi}_{\text{v}} + \boldsymbol{\Lambda}_{\text{v}} & \boldsymbol{B} \triangleq \lambda_0 \boldsymbol{\Psi}_{\text{v}}^T \boldsymbol{\Phi}^T\boldsymbol{\Phi}\boldsymbol{\Psi}_{\text{inf}} \\ \boldsymbol{C} \triangleq \lambda_0 \boldsymbol{\Psi}_{\text{inf}}^T \boldsymbol{\Phi}^T\boldsymbol{\Phi}\boldsymbol{\Psi}_{\text{v}} & \boldsymbol{\Lambda}_{\text{inf}} \end{bmatrix} \right)^{-1} \\
&\overset{\text{(a)}}{=} \begin{bmatrix} \left(\boldsymbol{A} - \boldsymbol{B}\boldsymbol{\Lambda}_{\text{inf}}^{-1}\boldsymbol{C}\right)^{-1} & \boldsymbol{A}^{-1}\boldsymbol{B}\left(\boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{B} - \boldsymbol{\Lambda}_{\text{inf}}\right)^{-1} \\ \left(\boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{B} - \boldsymbol{\Lambda}_{\text{inf}}\right)^{-1}\boldsymbol{C}\boldsymbol{A}^{-1} & \left(\boldsymbol{\Lambda}_{\text{inf}} - \boldsymbol{C}\boldsymbol{A}^{-1}\boldsymbol{B}\right)^{-1} \end{bmatrix} = \begin{bmatrix} \boldsymbol{A}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}.
\end{aligned} \tag{20}$$

where the equality (a) comes from Formula 0.7.3 in [38]. Moreover, $\boldsymbol{A}$ is a positive definitive matrix, since $\boldsymbol{A}$ is a combination of a Hermitian matrix and a positive diagonal matrix. Hence, $\boldsymbol{\Sigma}_{\text{v}} = \boldsymbol{A}^{-1}$ is a positive definitive matrix. Now, the lemma comes. □

Then, we can give the proof of Theorem 1.

Proof. Without loss of generality, we omit the superindex i in $\boldsymbol{x}^i$, $\boldsymbol{y}^i$, $\boldsymbol{s}^i$, $\boldsymbol{\Phi}^i$ and $\boldsymbol{\Psi}^i$ during the proof. Lemma 2 implies that only the nonzero recovered entries $\boldsymbol{s}_{\text{v}}$ and corresponding transform basis $\boldsymbol{\Psi}_{\text{v}}$ take effect. By Lemma 1, the differential entropy for $\boldsymbol{s}$ can be directly derived based on the corresponding posterior probability in Eq. (9), namely

$$\begin{aligned} &H(p(\boldsymbol{s}_{\text{v}}|\boldsymbol{y},\boldsymbol{\lambda},\lambda_0)) \\ &= -\int p(\boldsymbol{s}_{\text{v}}|\boldsymbol{y},\boldsymbol{\lambda},\lambda_0)\log p(\boldsymbol{s}_{\text{v}}|\boldsymbol{y},\boldsymbol{\lambda},\lambda_0)dp \\ &= \frac{1}{2}\log|\boldsymbol{\Sigma}_{\text{v}}| + C = -\frac{1}{2}\log|\lambda_0\boldsymbol{\Theta}_{\text{v}}^T\boldsymbol{\Theta}_{\text{v}} + \boldsymbol{\Lambda}_{\text{v}}| + C, \end{aligned} \tag{21}$$

where $\boldsymbol{\Theta}_{\text{v}} = \boldsymbol{\Phi}\boldsymbol{\Psi}_{\text{v}}$ and $\boldsymbol{\Lambda}_{\text{v}}$ is defined in Lemma 2. Moreover, the dependence of differential entropy on the measurement $\boldsymbol{y}$ is reflected by the related covariance and hyperparameters.

If we add another measurement according to the new measurement vector $\boldsymbol{r}^T$, we could obtain a new measurement matrix $\boldsymbol{\Phi}^i_{\text{new}} = \left[\boldsymbol{\Phi}^T \ \boldsymbol{r}\right]^T$ and a new $\boldsymbol{\Theta}_{\text{new}} = \boldsymbol{\Phi}_{\text{new}}\boldsymbol{\Psi}$. Meanwhile, Eq. (21) would be updated as

$$\begin{aligned} &H(p(\boldsymbol{s}_{\text{v}}|\boldsymbol{y}_{\text{new}},\boldsymbol{\lambda},\lambda_0)) = -\frac{1}{2}\log|\lambda_0(\boldsymbol{\Theta}_{\text{new}})^T(\boldsymbol{\Theta}_{\text{new}}) + \boldsymbol{\Lambda}| + C \\ &= -\frac{1}{2}\log\left|\lambda_0\left(\boldsymbol{\Psi}_{\text{v}}^T\boldsymbol{\Phi}^T\boldsymbol{\Phi}\boldsymbol{\Psi}_{\text{v}} + \boldsymbol{\Psi}_{\text{v}}^T\boldsymbol{r}\boldsymbol{r}^T\boldsymbol{\Psi}_{\text{v}}\right) + \boldsymbol{\Lambda}_{\text{v}}\right| + C \\ &= -\frac{1}{2}\log\left|\boldsymbol{\Sigma}_{\text{v}}^{-1} + \lambda_0\boldsymbol{\Psi}_{\text{v}}^T\boldsymbol{r}\boldsymbol{r}^T\boldsymbol{\Psi}_{\text{v}}\right| + C \\ &\overset{(a)}{=} -\frac{1}{2}\log\left(\left|\boldsymbol{\Sigma}_{\text{v}}^{-1}\right|\left|1 + \lambda_0\boldsymbol{r}^T\boldsymbol{\Psi}_{\text{v}}\boldsymbol{\Sigma}_{\text{v}}\boldsymbol{\Psi}_{\text{v}}^T\boldsymbol{r}\right|\right) + C \\ &= H(p(\boldsymbol{s}_{\text{v}}|\boldsymbol{y},\boldsymbol{\lambda},\lambda_0)) - \frac{1}{2}\log\left|1 + \lambda_0\boldsymbol{r}^T\boldsymbol{\Psi}_{\text{v}}\boldsymbol{\Sigma}_{\text{v}}\boldsymbol{\Psi}_{\text{v}}^T\boldsymbol{r}\right|. \end{aligned} \tag{22}$$

Here, the equality (a) follows from the Schur complements and determinantal formula [38]. By Lemma 2, the last term $\frac{1}{2}\log\left|1 + \lambda_0\boldsymbol{r}^T\boldsymbol{\Psi}_{\text{v}}\boldsymbol{\Sigma}_{\text{v}}\boldsymbol{\Psi}_{\text{v}}^T\boldsymbol{r}\right|$ of Eq. (22) is positive. Consequently, a newly added measurement could contribute to decrease the "error bars", since the measurement vector could be selected intentionally. The claim follows. □

Next, Corollary 1 can be easily proved, after taking into account that the equation $\boldsymbol{\Psi}^i(\boldsymbol{\Sigma}^i)(\boldsymbol{\Psi}^i)^T$ $= \boldsymbol{\Psi}^i_{\text{v}}(\boldsymbol{\Sigma}^i_{\text{v}})(\boldsymbol{\Psi}^i_{\text{v}})^T$ holds for any $i \in \{1, \cdots, P\}$.

REFERENCES

[1] J. Schiller and A. Voisard, *Location-based services*. Morgan Kaufmann, 2004. [Online]. Available: http://www.amazon.com/Location-Based-Services-Kaufmann-Management-Systems/dp/1558609296

[2] A. Kpper, *Location-Based Services: Fundamentals and Operation*. John Wiley & Sons, Ltd, 2005. [Online]. Available: http://www.amazon.com/Location-Based-Services-Fundamentals-Axel-pper/dp/0470092319

[3] S. Scellato, A. Noulas, R. Lambiotte, and C. Mascolo, "Socio-spatial properties of online location-based social networks," in *Proc. ICWSM*, Barcelona, Spain, Jul. 2011.

[4] M. N. Boulos, A. Rocha, A. Martins, M. E. Vicente, A. Bolz, R. Feld, I. Tchoudovski, M. Braecklein, J. Nelson, and G. . Laighin, "CAALYX: a new generation of location-based services in healthcare," *Int. J. Health Geogr.*, vol. 6, no. 1, p. 9, Mar. 2007.

[5] A. Tsalgatidou, J. Veijalainen, J. Markkula, A. Katasonov, and S. Hadjiefthymiades, "Mobile e-commerce and location-based services: Technology and requirements," in *Proc. ScanGIS*, Espoo, Finland, Apr. 2003.

[6] H. Harroud, M. Ahmed, and A. Karmouch, "Policy-driven personalized multimedia services for mobile users," *IEEE Trans. Mob. Comput.*, vol. 2, no. 1, pp. 16–24, Apr. 2003.

[7] F. Gustafsson and F. Gunnarsson, "Mobile positioning using wireless networks: possibilities and fundamental limitations based on available wireless network measurements," *IEEE Signal Process. Mag.*, vol. 22, no. 4, pp. 41–53, Jul. 2005.

[8] N. Patwari, J. N. Ash, S. Kyperountas, A. O. Hero III, R. L. Moses, and N. S. Correal, "Locating the nodes: cooperative localization in wireless sensor networks," *IEEE Signal Process. Mag.*, vol. 22, no. 4, pp. 54–69, 2005.

[9] P. Bahl and V. Padmanabhan, "RADAR: an in-building RF-based user location and tracking system," in *Proc. IEEE INFOCOM*, Tel Aviv, Israel, Mar. 2000.

[10] A. Kushki, K. Plataniotis, and A. Venetsanopoulos, "Kernel-based positioning in wireless local area networks," *IEEE Trans. Mob. Comput.*, vol. 6, no. 6, pp. 689–705, Jun. 2007.

[11] C. A. Patterson, R. R. Muntz, and C. M. Pancake, "Challenges in location-aware computing," *IEEE Pervasive Computing*, vol. 2, no. 2, pp. 80–89, Jun. 2003.

[12] C. Rizos, A. G. Dempster, B. Li, and J. Salter, "Indoor positioning techniques based on wireless LAN," in *Auswireless Conference*, Sydney, Australia, Mar. 2006.

[13] T. He, C. Huang, B. M. Blum, J. A. Stankovic, and T. Abdelzaher, "Range-free localization schemes for large scale sensor networks," in *Proc. ACM Mobicom*, San Diego, CA, USA, Sep. 2003.

[14] B. Zhang, X. Cheng, N. Zhang, Y. Li, Y. Li, and Q. Liang, "Sparse target counting and localization in sensor networks based on compressive sensing," in *Proc. IEEE INFOCOM*, Shanghai, China, Apr. 2011.

[15] C. Feng, W. S. A. Au, S. Valaee, and Z. Tan, "Compressive sensing based positioning using RSS of WLAN access points," in *Proc. IEEE INFOCOM*, San Diego, CA, USA, Mar. 2010.

[16] A. Au, C. Feng, S. Valaee, S. Reyes, S. Sorour, S. Markowitz, D. Gold, K. Gordon, and M. Eizenman, "Indoor tracking and navigation using received signal strength and compressive sensing on a mobile device," *IEEE Trans. Mob. Comput.*, vol. 12, no. 10, pp. 2050–2062, Aug. 2013.

[17] T. Roos, P. Myllymki, H. Tirri, P. Misikangas, and J. Sievnen, "A probabilistic approach to WLAN user location estimation," *International Journal of Wireless Information Networks*, vol. 9, no. 3, pp. 155–164, Jul. 2002.

[18] S. Nikitaki and P. Tsakalides, "Localization in wireless networks via spatial sparsity," in *Conference Record of the Forty Fourth Asilomar Conference on Signals, Systems and Computers (ASILOMAR) 2010*, Monterey, CA, USA, Nov. 2010.

[19] D. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 4036–4048, Apr. 2006.

[20] E. J. Candes and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?" *IEEE Trans. Inf. Theory*, vol. 52, no. 12, pp. 5406–5425, Dec. 2006.

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

[22] S. Ji, D. Dunson, and L. Carin, "Multitask compressive sensing," *IEEE Trans. Signal Process.*, vol. 57, no. 1, pp. 92–106, Jan. 2009.

[23] S. Ji, Y. Xue, and L. Carin, "Bayesian compressive sensing," *IEEE Trans. Signal Process.*, vol. 56, no. 6, pp. 2346–2356, Jun. 2008.

[24] X. Li, S. Hong, Z. Han, and Z. Wu, "Bayesian compressed sensing based dynamic joint spectrum sensing and primary user localization for dynamic spectrum access," in *Proc. IEEE Globecom*, Houston, Texas, USA, Dec. 2011.

[25] Y. Zhang, Z. Zhao, and H. Zhang, "Adaptive bayesian compressed sensing based localization in wireless networks," in *Proc. Chinacom*, Kunming, China, Aug. 2012.

[26] J. Romberg and M. Wakin, "Compressed sensing: A tutorial," in *Proc. IEEE SSP Workshop*, Madison, Wisconsin, Aug. 2007. [Online]. Available: http://people.ee.duke.edu/~willett/SSP//Tutorials/ssp07-cs-tutorial.pdf

[27] C. L. Lawson and R. J. Hanson, *Solving least squares problems*. SIAM, 1974, vol. 161. [Online]. Available: http://epubs.siam.org/doi/pdf/10.1137/1.9781611971217.fm

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

[29] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM J. Sci. Comput.*, vol. 20, no. 1, pp. 33–61, Aug. 1998.

[30] Y. Tsaig and D. L. Donoho, "Extensions of compressed sensing," *Signal Processing*, vol. 86, no. 3, pp. 549–571, Mar. 2006.

[31] E. J. Cands, "The restricted isometry property and its implications for compressed sensing," *Comptes Rendus Mathematique*, vol. 346, no. 9, pp. 589–592, Feb. 2008.

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

[33] D. Wipf and B. Rao, "Comparing the effects of different weight distributions on finding sparse representations," in *Proc. NIPS*, Vancouver, Canada, Dec. 2006.

[34] D. Wipf, J. Palmer, and B. Rao, "Perspectives on sparse bayesian learning," *Advances in Neural Information Processing Systems*, vol. 16, pp. 249–256, 2004.

[35] M. E. Tipping, "Sparse bayesian learning and the relevance vector machine," *J. Mach. Learn. Res.*, vol. 1, no. 1, pp. 211–244, Jun. 2001.

[36] S. Seidel and T. Rappaport, "914 MHz path loss prediction models for indoor wireless communications in multifloored buildings," *IEEE Trans. Antennas Propagat.*, vol. 40, no. 2, pp. 207–217, Feb. 1992.

[37] T. M. Cover and J. A. Thomas, "Differential entropy," in *Elements of information theory*. Wiley-interscience, 2012.

[38] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge University Press, Feb. 1990. [Online]. Available: http://www.amazon.com/Matrix-Analysis-Roger-A-Horn/dp/0521386322