Millimeter wave communications channel estimation via Bayesian group sparse recovery

Suryaprakash, R.T.; Pajovic, M.; Kim, K.J.; Orlik, P.V.

TR2016-012   March 2016

Abstract

We consider the problem of channel estimation for millimeter wave communications (mmWave). We formulate channel estimation as a structured sparse signal recovery problem, in which the signal structure is governed by a priori knowledge of the channel characteristics. We develop a Bayesian group sparse recovery algorithm which takes into account for several features unique to mmWave channels, such as spatial (angular) spreads of received signals and power profile of rays impinging on the receiver array. We validate the developed method via numerical simulations and demonstrate an improved estimation performance relative to the existing methods.

MILLIMETER WAVE COMMUNICATIONS CHANNEL ESTIMATION VIA BAYESIAN GROUP SPARSE RECOVERY

Raj Tejas Suryaprakash *

Electrical Engineering and Computer Science
University of Michigan, Ann Arbor, MI 48105
rajtejas@umich.edu

Milutin Pajovic, Kyeong Jin Kim and Phil Orlik

Mitsubishi Electric Research Labs
Cambridge, MA, 02139
{pajovic, kkim, porlik}@merl.com

ABSTRACT

We consider the problem of channel estimation for millimeter wave communications (mmWave). We formulate channel estimation as a structured sparse signal recovery problem, in which the signal structure is governed by a priori knowledge of the channel characteristics. We develop a Bayesian group sparse recovery algorithm which takes into account for several features unique to mmWave channels, such as spatial (angular) spreads of received signals and power profile of rays impinging on the receiver array. We validate the developed method via numerical simulations and demonstrate an improved estimation performance relative to the existing methods.

Index Terms— Bayesian sparse recovery, structured sparsity, mmWave, channel estimation

1. INTRODUCTION

There has recently been an increased interest in mmWave communications (mmWave) as a result of high throughput capabilities offered by mmWave frequency bands [1]. Advances in hardware design have made it possible to design large arrays and use beamforming techniques to overcome the effects of increased path loss at higher frequencies. There is also an increased burden on system hardware in mmWave. Power consumption due to RF chains is significantly increased due to higher frequencies of operation. Moreover, one must account for Analog-to-Digital (ADC) converter cost considerations during system design and analysis. Consequently, beamforming is carried out in the RF regime by using analog phase shifters, or by using hybrid schemes [2].

Previous work on channel estimation in the mmWave region has exploited sparsity in the number of multipath components, and accounted for ADC related constraints. In [2], the authors cast the channel estimation problem as a sparse recovery problem. In this formulation, the measurement matrix consists of vectors $a(\theta_{tx}, \theta_{rx})$, which are parameterized by Direction-of-Departure (DoD) θ_{tx} and Direction-of-Arrival

(DoA) θ_{rx}. The unknown, sparse signal vector contains non-zero channel coefficients at coordinates corresponding to $(\theta_{tx}, \theta_{rx})$. However, such an approach does not account for the structured sparsity of the signal, which is inherent to mmWave. Alternatively, recent work in structured sparse recovery using Bayesian approaches [3], [4] enforce graphical model priors on signal sparsity to recover structured sparse signals, but these models do not accurately model the mmWave channel properties of interest.

Some features unique to mmWave are as follows. There is a low number of multipath components due to high path loss. However, due to high reflectivity, there is a spread in Direction of Arrival (DoA) of rays within each multipath, called a 'cluster'. In convential communication systems, arrivals at the receiver array are localized to a small angular region in space. In contrast, in the mmWave regime, rays impinging on the receiver array are spatially clustered in a continuous band of angles. Although previous work has explored several distributions to model DoA spread [5], and tried to quantify the extent of this spread using Fourier Series based statistics [6], these properties have not yet been exploited for mmWave system design. Furthermore, the powers of these clustered rays follow a power profile, governed by the channel model. The methods in this paper may also be extended to exploit spread in Directon of Departure (DoD), but we leave this for future work.

In this work, we focus on the problem of channel estimation for mmWave. We utilize the approach of [2] in sensing the channel which accounts for hardware constraints. Our novel contribution may be summarized as follows.

1. We formulate channel estimation as a *group sparse* recovery problem.

2. We develop a Bayesian algorithm which recovers the group sparse channel, thereby accounting for an important feature of mmWave - DoA spread. Our algorithm also allows one to exploit knowledge of the power profile of each cluster arrival.

Thus, the estimation algorithm is intimately matched with the channel model, and exploits the wealth of knowledge about

---

the channel available to us from various measurement campaigns.

The channel model relevant to mmWave is IEEE 802.11ad, whose complete statistical characterization is still underway. Fortunately, it has been discovered that the IEEE 802.11ad channel is similar to the well studied IEEE 802.15.3c channel model [1]. In this work, we employ a channel model similar to the IEEE 802.15.3c, and highlight distributions corresponding to properties of interest. However, the analysis presented here can be easily adapted to other channel models, so long as a statistical description of the channel model is available.

2. CHANNEL MODEL

Table 1 summarizes important parameters of the channel. The received energy arrives in clusters, whose number n_c is uniformly distributed between 1 and c_{max}, which is environment dependent. Cluster i has DoA spread of $\Delta\theta_i$ at the receiver, which is centered at, and symmetric about the cursor. The cursor powers decay exponentially with time of arrival, and intracluster ray powers decay exponentially with angular distance from the cluster. The corresponding virtual channel matrix, illustrated in Fig. 1, is given by

$$H = A(\Theta_{rx})H_s A(\Theta_{tx})^H, \qquad (1)$$

where $\Theta_{tx} = [\theta_{tx,1}, ..., \theta_{tx,n_{tx}}]$ and $\Theta_{rx} = [\theta_{rx,1}, ..., \theta_{rx,n_{rx}}]$ are the DoDs and DoAs of the existing multipath components in the channel, and $A(\Theta_{rx})$, $A(\Theta_{tx})$ are matrices whose columns $\{a(\theta_{tx,i})\}_{i=1}^{n_{tx}}$, $\{a(\theta_{rx,i})\}_{i=1}^{n_{rx}}$ are manifold vectors corresponding to the respective DoD and DoA.

3. CHANNEL SENSING METHOD

The transmitter applies precoders $\{p_i\}_{i=1}^m$ to the symbol $t = 1$ in m successive snapshots. The receiver employs corresponding mixers $\{q_i\}_{i=1}^m$. In this work, we assume random precoding and mixing vectors with $\pm 1, \pm j$ elements, as in [2]. The

Parameter	Variable	Expr./Distrib.
Max. number of clusters	c_{max}	fixed
Realized number of clusters	n_c	$\text{Unif}(\{1, \ldots, c_{max}\})$
DoA spread	$\Delta\theta$	$\frac{|\mathcal{N}(0,\sigma_{sp}^2)|}{\Phi([-\pi/2,\pi/2],0,\sigma_{sp}^2)}$
Cursor arrival times	t_c	$\Gamma e^{-t/\Gamma}$
Cursor power	p_c	$\gamma e^{-t/\gamma}$
Intracluster power decay	$p_{\theta,c}$	$e^{|\theta-\theta_0|/\delta}$
Ray amplitudes	$h_{\theta,c}$	$\mathcal{CN}(0, p_{\theta,c})$

Table 1. Parameters of the channel model. Γ, γ and δ are environment dependent model parameters. The term $\Phi([-\pi/2, \pi/2], 0, \sigma_{sp}^2)$ denotes the area between $[-\pi/2, \pi/2]$ under Gaussian density of mean 0 and variance σ^2.
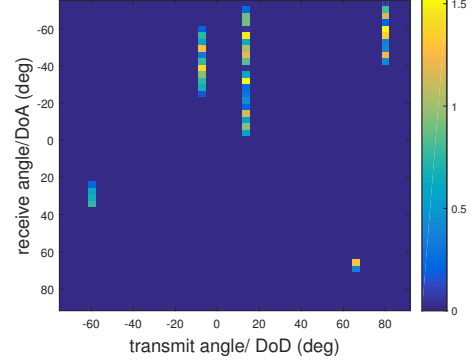

Fig. 1. Plot of $|H_s|$ (single realization). Rows correspond to DoD and columns to DoA. Rays arrive clustered spatially at the receiver, so that non-zero components are group sparse.

i^{th} observed data snapshot is

$$y_i = \sqrt{\rho}q_i^H H p_i t + q_i^H z$$
$$= \sqrt{\rho}q_i^H A(\Theta_{rx})H_s A(\Theta_{tx})^H p_i t + e_i, \qquad (2)$$

where ρ is the SNR, where $e_i \sim \mathcal{CN}(0, \sigma_n^2)$ is the measurement noise, and H, $A(\Theta_{rx})$, $A(\Theta_{rx})$ and H_s are as defined in (1). From properties of the Kronecker product, we know that for any matrices A, B, C, [7, Theorem 13.26]

$$\text{vec}(ABC) = (C^T \otimes A)\text{vec}(B), \qquad (3)$$

where the $\text{vec}(\cdot)$ operation rearranges the elements of its operand columnwise into a vector. Using (3) in (2), we get

$$y_i = \sqrt{\rho}(p_i^T \otimes q_i^H)(A(\Theta_{tx})^* \otimes A(\Theta_{rx}))\text{vec}(H_s) + e \quad (4)$$

Stacking up the m data snapshots into a vector, we get:

$$\underbrace{\begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix}}_{:=y} = \sqrt{\rho}\underbrace{\begin{bmatrix} p_1^T \otimes q_1^H \\ \vdots \\ p_m^T \otimes q_m^H \end{bmatrix}(A(\Theta_{tx})^* \otimes A(\Theta_{rx}))}_{:=A}\underbrace{\text{vec}(H_s)}_{:=x} + e. \qquad (5)$$

In (5), A is a measurement matrix, which is unknown. Therefore, we generate a dictionary of possible DoDs, $\Theta_{tx} = \{\theta_{tx,1}, \theta_{tx,2}, \ldots\}$ and DoAs $\Theta_{rx} = \{\theta_{rx,1}, \theta_{rx,2}, \ldots\}$. The angles Θ_{tx} and Θ_{rx} may not contain the true DoD and DoA, which may lead to grid mismatch. A complete analysis of the effects of grid mismatch is beyond the scope of this work; see, for e.g. [8]. We proceed by selecting Θ_{tx} and Θ_{rx} such that $\sin\Theta =$

$$\begin{cases} \{-\frac{n-2}{n}, -\frac{n-4}{n}, \ldots, \frac{-2}{n}, 0, \frac{2}{n}, \frac{4}{n}, \ldots, 1\} & \text{if } n \text{ even} \\ \{-\frac{n-1}{n}, -\frac{n-3}{n}, \ldots, \frac{-2}{n}, 0, \frac{2}{n}, \frac{4}{n}, \ldots, \frac{n-1}{n}\} & \text{if } n \text{ odd} \end{cases}.$$

These angles correspond to the peak of the mainlobe, and nulls of the uniform linear array beampattern. With some abuse of notation, henceforth we denote the resulting dictionary matrix also by A. Due to our choice of Θ_{tx} and Θ_{rx}, A is completely incoherent [9], i.e., given its ith and j^{th} columns, a_i and a_j, we have that $a_i^H a_j = 0$ for $i \neq j$. Vector x in (5) contains 0 at entries corresponding angles not present in the

multipath departure/arrival structure. A vectorized version of the matrix in Fig. 1 is representative of the sparsity structure of x.

4. CHANNEL ESTIMATION VIA BAYESIAN INFERENCE

Recollect that we wish to find $\widehat{x}$, an estimate of x observed according to (5). Let s be the indicator for the support of x, given by

$$s_i = \begin{cases} 1 & \text{if } x_i \neq 0 \\ 0 & \text{if } x_i = 0 \end{cases} \quad \text{for } i = 1, \ldots, n. \quad (6)$$

Let A_s and x_s denote the subset of the corresponding matrix or vector for which $s_i = 1$. The conditional density of data, due to Gaussian noise statistics is

$$p(y|x_s, s) = \frac{1}{(\pi \sigma_n^2)^n} \exp\left\{-\frac{||y - A_s x_s||_2^2}{\sigma_n^2}\right\}. \quad (7)$$

The density of x_s given the support s is

$$p(x_s|s) = \frac{1}{|\pi \Sigma_s|} \exp\left\{-x_s^H \Sigma_s^{-1} x_s\right\} \quad (8)$$

where Σ_s is signal covariance matrix. The density of data given the signal support configuration is given by $p(y|s) = \int_{x_s} p(y|x_s, s) p(x_s|s) dx_s$. Define

$$Q_s := A_s^H A_s + \sigma_n^2 \Sigma_s^{-1}.$$

Then computing $p(y|s)$ gives us

$$p(y|s) = \frac{(\pi \sigma_n^2)^{(s-n)}}{|\pi \Sigma_s||Q_s|^{\frac{1}{2}}} \exp\left\{-\frac{||y||_2^2 - y^H A_s Q_s^{-1} A_s^H y}{\sigma_n^2}\right\}.$$

Let $\mathbb{P}(s)$ be prior probability of support s. From (4), the logarithm of the posterior probability of s given data y, $\mathbb{P}(s|y)$ is

$$\log\mathbb{P}(s|y) \equiv \underbrace{\frac{y^H A_s Q_s^{-1} A_s^H y}{\sigma_n^2}}_{\text{data term}} - \underbrace{\log\det(\pi\Sigma_s)}_{\text{signal power term}} -$$

$$(n-s)\log(\pi\sigma_n^2) - \frac{1}{2}\log|\det(Q_s)| + \underbrace{\log\mathbb{P}(s)}_{\text{prior}}. \quad (9)$$

We propose a method to compute $\mathbb{P}(s)$ below.

4.1. Computation of Support Prior $\mathbb{P}(s)$

We may partition s defined in (6) into blocks of length n_{rx}, as $s = \begin{bmatrix} s_1 & \cdots & s_{n_{\text{tx}}} \end{bmatrix}^T$, i.e., support indicator corresponding to columns of H_s. Let c_i be the indicator of cluster presence in column i, for $i = 1, \ldots, n_{\text{tx}}$, and $c = \begin{bmatrix} c_1 & \cdots & c_{n_{\text{tx}}} \end{bmatrix}^T$. We now compute $\mathbb{P}(s)$. We have that

$$\mathbb{P}(s) = \mathbb{P}(s|c, n_{\text{c}})\mathbb{P}(c|n_{\text{c}})\mathbb{P}(n_{\text{c}}). \quad (10)$$

From the distribution of n_{c} specified in Table 1, we have $\mathbb{P}(n_{\text{c}}) = \frac{1}{c_{\max}}$. Since the clusters are distributed randomly from among the n_{tx} columns of H_s, $\mathbb{P}(c|n_{\text{c}}) = 1/\binom{n_{\text{tx}}}{n_{\text{c}}}$, where $\binom{n_{\text{tx}}}{n_{\text{c}}}$ denotes "n_{tx} choose n_{c}". Finally, since we

have assumed that there is no spread in DoD, $\mathbb{P}(s|c, n_{\text{c}}) = \prod_{i:c_i=1} \mathbb{P}(s_i|c_i, n_{\text{c}})$. Since signal energy is present in a continuous band of angles around the cursor, locations in which $s_i = 1$ occur contiguously. Therefore, we only need to compute probabilities of the form

$$\mathbb{P}(s_i = [0, ..., 0, \underbrace{1}_{j}, ..., \underbrace{1}_{j+k}, 0, ..., 0]|c_i, n_{\text{c}}). \quad (11)$$

Let $\phi := \sin(\theta)$, and $\Delta\phi_w$ be the resolution of the dictionary $\sin(\Theta_{\text{rx}})$. For our choice of Θ_{rx} in Section 3, $\Delta\phi_w = 2/n$. Consider (11), in which a cluster spans from index j to $j+k$. Since the DoA spread is centered at the cursor, sine-DoA of the cursor ϕ_0 corresponds to index $(j+k)/2$, pre-cursor spread $\Delta\phi_-$ to indices j to $(j+k)/2 - 1$, and post-cursor spread $\Delta\phi_+$ to indices $(j+k)/2 + 1$ to k. This computation is illustrated in Fig. 2. Therefore, probability of this cluster is given by (12).
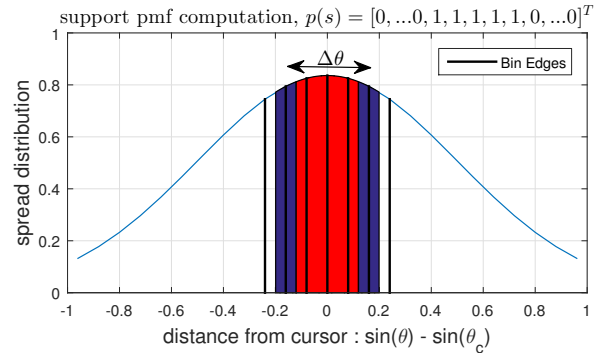

Fig. 2. Support prior computation. The plot shows a cluster with cursor at 0° and DoA spread $\Delta\theta$. The cluster corresponds to 5 non-zero support locations, so that $k = 4$. We see that DoA spreads are in $[\pm(3/2)\Delta\phi_w, \pm(5/2)\Delta\phi_w]$ as given by (13).

Since DoA spread is symmetric about the cursor, and independent of cursor location from Table 1,

$$\mathbb{P}(s_i|c_i, n_{\text{c}}) = 2\mathbb{P}\left(\Delta\phi_+ \in \left[\frac{k-1}{2}\Delta\phi_w, \frac{k}{2}\Delta\phi_w\right]\right). \quad (13)$$

This probability can be computed directly from the density of $\Delta\theta$ in Table 1, which completes the computation of $\mathbb{P}(s)$. Furthermore, notice that (9) involves a signal power term. The channel model specifies the ray power decays exponentially with angular distance from cursor, which can be used to adaptively assign powers to estimated channel coefficients. The overall MAP inference procedure is described in Algroithm 1. The exact MAP estimate requires searching for $\widehat{s}_*$ over the $\{0,1\}^n$ space, which is exponentially complex [4, 10]. Therefore, we resort to a suboptimal search to find the MAP estimate. We add one new support location in each iteration, using the channel model to guide the search, by not considering locations corresponding to 0 prior probability, as in Steps $5 - 11$ of the algorithm.

$$\mathbb{P}\left(\Delta\phi_{-}\in\left[\phi_0-\frac{k}{2}\Delta\phi_w,\ \phi_0-\frac{k-1}{2}\Delta\phi_w\right],\Delta\phi_{+}\in\left[\phi_0+\frac{k-1}{2}\Delta\phi_w,\ \phi_0+\frac{k}{2}\Delta\phi_w\right]\Bigg|\phi_0=\phi_{\frac{j+k}{2}}\right) \quad (12)$$

Algorithm 1 Channel Model Based Inference.

1: **Inputs**: σ_n, $\mathrm{iter}_{\max}$
2: $\widehat{s}_{(0)}\leftarrow\mathbf{0}\in\mathbb{R}^n$ ▷ support indicator
3: **for** $i\in\{1,2,\ldots,\mathrm{iter}_{\max}\}$ **do**
4:      $J\leftarrow\{j:\widehat{s}_{(i-1),j}=0\}$
5:      Discard $j\in J$ such that $\mathbb{P}(\widehat{s}_{(i-1)}\cup s_j=1)=0$
6:      **for** $j\in J$ **do** ▷ loop over unselected indices
7:          $\widehat{s}_{(i)}^{(j)}=\widehat{s}_{(i-1)}\leftarrow1,\ s_j\leftarrow1$
8:          Set Σ_s of according to power profile in Table 1
9:          $\mathrm{val}_{(i)}^{(j)}\leftarrow\log\mathbb{P}(s_{(i)}^{(j)}|y)$ ▷ compute from (9)
10:      **end for**
11:      $j_*\leftarrow\arg\max_{j\in J_{\mathrm{cand}}}\{\mathrm{val}_{(i)}^{(j)}\}$
12:      $\mathrm{max\_post}_{(i)}\leftarrow\mathrm{val}_{(i)}^{(j_*)}$
13:      $\widehat{s}_{(i)}\leftarrow s_{(i)}^{(j_*)}$
14: **end for**
15: $i_*\leftarrow\arg\max_i\{\mathrm{max\_post}_{(i)}\}$
16: $\widehat{s}_*\leftarrow\arg\max_{(i)}\widehat{s}_{(i)}$ ▷ MAP over all iterations
17: $\widehat{x}\leftarrow A_{s_*}^{\dagger}y$

5. NUMERICAL SIMULATIONS AND RESULTS

We quantify channel estimation performance using normalized mean squared error, defined as $\mathrm{NMSE}=\frac{||\widehat{x}-x||_2^2}{||x||_2^2}$. For each value of SNR, we generated 35 channel and noise realizations, computed average NMSE over all realizations. Fig. 3 plots the NMSE vs SNR for Bayesian algorithm and channel estimation using OMP [11]. The expected sparsity of the signal is 0.0098. We see a performance gain of around 2 dB at lower SNRs. We highlight here that measurement matrix A was completely incoherent, which along with the low sparsity of x, is a regime in which OMP is proved to perform well. Furthermore, Bayesian approach requires no knowledge of signal sparsity. Fig. 4 plots the estimated channel employing OMP and our algorithm for a single noise realization, at SNR = 3dB, for the same setting as in Fig. 3. This plot highlights the reason underlying the performance gain of the Bayesian algorithm at low SNRs. OMP does not favor grouped signal support over distributed signal support, and hence selects few erroneous support locations. However, the channel based prior $\mathbb{P}(s)$ in the Bayesian algorithm encourages it to select contiguous support locations, since arrivals are spatially clustered. At high SNRs, both OMP and the Bayesian algorithm pick the correct support locations, and therefore, yield similar performance.
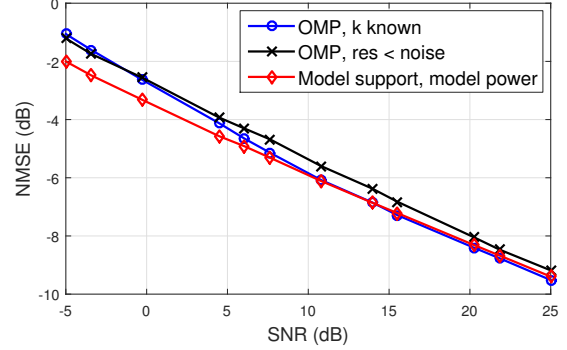

Fig. 3. NMSE estimation performance of our algorithm and OMP vs SNR, for a system with 48 transmitter and receiver elements, and $m=750$, and avg. channel sparsity 0.0098. ("Model support, model power" refers to proposed algorithm.)
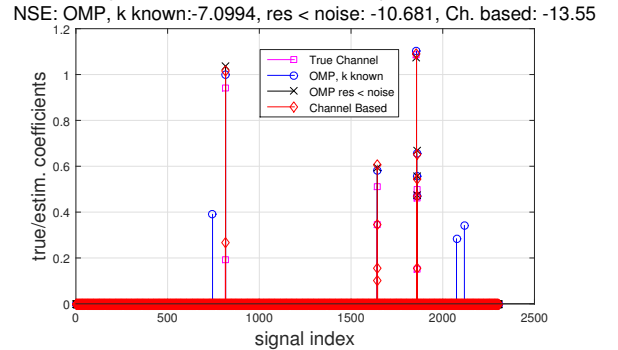

Fig. 4. Stem plot of recovered channel coefficients for a single trial in the setting of Fig. 3 at SNR $=3dB$.

6. CONCLUSION

We have considered the problem of channel estimation for mmWave, which we formulated as a group sparse recovery problem. We have designed a Bayesian estimation algorithm which accounts for several distinguishing properties of mmWave, particularly the DoA spread and power profile of spatially clustered signal arrivals. Our algorithm improves channel estimation performance compared to state of the art by 2 dB at low SNRs, which is the regime of interest in most communication systems. Additionally, the proposed Bayesian framework does not require knowledge of signal sparsity, but utilizes channel model for structured signal recovery.

7. REFERENCES

[1] Theodore S Rappaport, Robert W Heath Jr, Robert C Daniels, and James N Murdock, *Millimeter wave wireless communications*, Pearson Education, 2014.

[2] Roi Méndez-Rial, Cristian Rusu, Ahmed Alkhateeb, Nuria González-Prelcic, and Robert W Heath Jr, "Channel estimation and hybrid combining for mmwave: Phase shifters or switches?," in *Proc. of Information Theory and Applications Workshop (ITA)*, 2015.

[3] Volkan Cevher, Marco F Duarte, Chinmay Hegde, and Richard Baraniuk, "Sparse signal recovery using markov random fields," in *Advances in Neural Information Processing Systems*, 2009, pp. 257–264.

[4] T. Peleg, Y.C. Eldar, and M. Elad, "Exploiting statistical dependencies in sparse representations for signal recovery," *IEEE Transactions on Signal Processing*, vol. 60, no. 5, pp. 2286–2303, May 2012.

[5] G Durgin and TS Rappaport, "Basic relationship between multipath angular spread and narrowband fading in wireless channels," *Electronics Letters*, vol. 34, no. 25, pp. 2431–2432, 1998.

[6] Gregory D Durgin and Theodore S Rappaport, "Effects of multipath angular spread on the spatial cross-correlation of received voltage envelopes," in *IEEE Vehicular Technology Conference*, 1999, vol. 2, pp. 996–1000.

[7] Alan J Laub, *Matrix analysis for scientists and engineers*, Siam, 2005.

[8] Yuejie Chi, Louis L Scharf, Ali Pezeshki, et al., "Sensitivity to basis mismatch in compressed sensing," *IEEE Transactions on Signal Processing*, vol. 59, no. 5, pp. 2182–2195, 2011.

[9] T Tony Cai and Lie Wang, "Orthogonal matching pursuit for sparse signal recovery with noise," *IEEE Transactions on Information Theory*, vol. 57, no. 7, pp. 4680–4688, 2011.

[10] Philip Schniter, Lee C Potter, and Justin Ziniel, "Fast bayesian matching pursuit," in *Information Theory and Applications Workshop, 2008*. IEEE, 2008, pp. 326–333.

[11] Joel Tropp, Anna C Gilbert, et al., "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Transactions on Information Theory*, vol. 53, no. 12, pp. 4655–4666, 2007.