Fast Beam Alignment for Millimeter Wave Communications: A Sparse Encoding and Phaseless Decoding Approach

Xingjian Li, Jun Fang, Senior Member, IEEE, Huiping Duan, Zhi Chen, and Hongbin Li, Fellow, IEEE

Abstract—In this paper, we study the problem of beam alignment for millimeter wave (mmWave) communications, where a hybrid analog and digital beamforming structure is employed at the transmitter (i.e., base station), and an omni-directional antenna or an antenna array is used at the receiver (i.e., user). By exploiting the sparse scattering nature of mmWave channels, the beam alignment problem is formulated as a sparse encoding and phaseless decoding problem. More specifically, the problem of interest involves finding a sparse sensing matrix and an efficient recovery algorithm to recover the support and magnitude of a sparse signal from compressive phaseless measurements. A sparse bipartite graph code algorithm is developed for sparse encoding and phaseless decoding. Our theoretical analysis shows that, in the noiseless case, our proposed algorithm can perfectly recover the support and magnitude of the sparse signal with a probability exceeding a pre-specified value from $O(K^2)$ measurements, where $K$ is the number of nonzero entries of the sparse signal. The proposed algorithm has a simple decoding procedure which is computationally efficient and noise robust. Simulation results show that our proposed method renders reliable beam alignment even in the low signal-to-noise ratio regime and presents a clear performance advantage over existing methods.

Index Terms—Millimeter wave (mmWave) communications, beam alignment, sparse encoding, phaseless decoding.

I. INTRODUCTION

MILLIMETER wave (mmWave) communication is a promising technology for future cellular networks [1]–[3]. It has the potential to offer gigabits-per-second communication data rates by exploiting the large bandwidth available at mmWave frequencies. Nevertheless, communication at the mmWave frequency bands suffers high attenuation and signal absorption [4], [5]. To address this issue, large antenna arrays should be used to provide sufficient beamforming gain for mmWave communications [6].

Although directional beamforming helps compensate for the significant path loss experienced by mmWave signals, it requires a complicated beamforming training procedure to find the best beamformer-combiner pair that gives the highest beamforming gain [7] because, due to the narrow beam of the antenna array, communication between the transmitter and the receiver is possible only when the transmitter’s and receiver’s beams are well-aligned. A natural approach to perform beamforming training is to exhaustively search all possible beam pairs, which requires the receiver to scan the entire space for each choice of beam direction on the transmitter side. This exhaustive search has a sample complexity of $O(N^2)$ ($N$ denotes the number of possible beam directions) and usually takes a long time (up to several seconds) to converge [8].

To address this issue, many efforts were directed toward reducing the time required for beamforming training. Specifically, the IEEE 802.11ad standard utilizes an exhaustive search at the receiver, with the transmitter adopting a quasi-omnidirectional beam pattern. This process is then reversed to have the transmitter sequentially scan the entire space while the receiver uses a quasi-omnidirectional beam shape. This protocol reduces the sample complexity from $O(N^2)$ to $O(N)$. To further reduce the training time, adaptive beam alignment algorithms, e.g. [9]–[13], were proposed. In these works, a hierarchical multi-resolution beamforming codebook set is employed to avoid the costly exhaustive sampling of all pairs of transmit and receive beams. The basic idea is to use a coarse codebook to first identify the range of the beam direction, and then use high-resolution subcodebooks to find a refined beam direction. This adaptive beam alignment adaptively chooses a subcodebook at each stage based on the output of earlier stages. A drawback of this scheme is that the base station has to interact with each user individually, which may not be feasible at the initial channel acquisition stage.

In addition to the above beam steering techniques, another approach [14]–[21] directly estimates the mmWave channel or its associated parameters, e.g. angles of arrival/departure, without the need of scanning the entire space. This class of methods exploits the sparse scattering nature of mmWave channels and formulate the channel estimation into a compressed sensing problem. The compressed sensing-based channel estimation approach, however, suffers several drawbacks. Firstly,
although very simple linear-time algorithms based on matching pursuit [22], [23] exist for compressed channel estimation, most compressed sensing methods, e.g. [15]–[20], involve a computational complexity that might be excessive for practical systems. Secondly, compressed sensing methods require the knowledge of the phase of the measurements. While in mmWave communications, due to the carrier frequency offset (CFO) and the jitter in the frequency of the oscillators, the phase of the measurements might be corrupted by a random noise that varies across time. As a result, only the magnitude information of the measurements is useful for beam alignment. Recently, a new beam steering scheme called “Agile-Link” [8], [24] was proposed for beam alignment. The proposed algorithm only uses the magnitude of the measurements for recovery of the signal directions and achieves a sample complexity of $O(K \log N)$, where $K$ denotes the number of signal paths. In [25], a two-stage approach was proposed for mmWave channel estimation, which cascades phase retrieval with coherent compressed sensing to reconstruct the sparse channel vector from phaseless measurements. In some other works, [26]–[29], to cope with the unknown phase issue, the CFO is modeled as an unknown parameter that is common to all measurements and jointly estimated with the channel.

In this paper, we continue the efforts towards developing a fast and efficient beam alignment scheme for mmWave communications. Similar to [8], [24], we rely on the magnitude of the measurements. By exploiting the sparse scattering nature of mmWave channels, we show that the beam alignment problem can be formulated as a sparse encoding and phaseless decoding problem. More specifically, the problem of interest is to devise a sparse sensing matrix $A$ (referred to as sparse encoding) and develop a fast and efficient recovery algorithm (referred to as phaseless decoding) to recover the support and magnitude of the sparse signal $x$ from compressive phaseless measurements:

$$y = |Ax|$$

Note that the estimation of sparse signals from compressive phaseless measurements, termed as “compressive phase retrieval (CPR)”, has been extensively studied over the past few years, e.g. [30]–[34]. Nevertheless, there are two important distinctions between our problem and the standard CPR problem. First, standard CPR assumes a random measurement matrix which satisfies the restricted isometry property, whereas for the reason that will be elaborated in Section II, a sparse measurement matrix is considered in our problem. Second, standard CPR aims to retrieve the magnitude and phase of the sparse signal $x$, while for the beam alignment purpose, only the support and magnitude of those nonzero entries in $x$ need to be recovered.

To our best knowledge, in existing literature, the CPR algorithm PhaseCode proposed in [35], whose measurement matrix is devised based on a sparse-graph coding framework, is closest to our sparse encoding and phaseless decoding method. It was shown that PhaseCode can recover a $K$-sparse signal using slightly more than $4K$ measurements with a high probability. Nevertheless, PhaseCode (even its robust version) involves a delicate decoding procedure sensitive to noise and measurement errors, and may suffer severe performance degradation in the presence of noise. To overcome this difficulty, in this work, we propose a sparse bipartite graph code (SBG-Code) algorithm for sparse encoding and phaseless decoding. Different from PhaseCode, our proposed method uses a set of sparse bipartite graphs, instead of a single bipartite graph, to encode the sparse signal. The proposed algorithm involves a simple decoding procedure which has a very low computational complexity (of order $O(N)$) and is robust against noise. Also, it can recover the support and magnitude of a $K$-sparse signal with a sample complexity of $O(K^2)$, thus providing a competitive solution for practical mmWave beam alignment systems.

The rest of the paper is organized as follows. In Section II, the system model is discussed and the beam alignment is formulated as a sparse encoding and phaseless decoding problem. In Section III, an overview of PhaseCode is provided. Our SBG-Code method is developed in Section IV, along with a theoretical analysis in Section V. A robust version of SBG-Code is introduced in Section VI and an extension to the case involving an antenna array receiver is discussed in Section VII. Simulation results are presented in Section VIII, followed by concluding remarks in Section IX.

II. SYSTEM MODEL

Consider a mmWave communication system which consists of a transmitter (base station) and a receiver (user). We assume that a hybrid analog and digital beamforming structure is employed at the transmitter, while the receiver has an omni-directional antenna. In this paper, we consider a single user scenario that receives in all directions (see Fig. 1). The extension to an antenna array receiver is discussed in Section VII. Simulation results are presented in Section VIII, followed by concluding remarks in Section IX.
The mmWave channel is characterized by a geometric channel model [11]

\[ h = \sum_{p=1}^{P} c_p a_t(\theta_p) \]  

(2)

where \( P \) is the number of paths, \( c_p \) is the complex gain associated with the \( p \)th path, \( \theta_p \in [0, 2\pi] \) is the associated azimuth AoD, and \( a_t \in \mathbb{C}^N \) is the transmitter array response vector. Suppose a uniform linear array (ULA) is used. Then the steering vector at the transmitter can be written as

\[ a_t(\theta_p) = \frac{1}{\sqrt{N}} \begin{bmatrix} e^{j\frac{2\pi}{\lambda}d \sin(\theta_p)} & \cdots & e^{j(N-1)\frac{2\pi}{\lambda}d \sin(\theta_p)} \end{bmatrix}^T \]  

(3)

where \( \lambda \) is the signal wavelength, and \( d \) is the distance between neighboring antenna elements. Due to the sparse scattering nature of mmWave channels, \( h \) has a sparse representation in the beam space (angle) domain:

\[ h = Dx \]  

(4)

where \( D \in \mathbb{C}^{N \times N} \) is the discrete Fourier transform (DFT) matrix, and \( x \in \mathbb{C}^N \) is a \( K \)-sparse vector. If the true AoD parameters \( \{\theta_p\} \) lie on the discretized grid specified by the DFT matrix, then the number of nonzero entries in the beam space domain equals the number of signal paths, i.e. \( K = P \). The objective of beam alignment is to estimate the AoD and the attenuation (in magnitude) of each path, which is equivalent to recovering the location indices and the magnitudes of the nonzero entries in \( x \). Note that in mmWave communications, the power of the line-of-sight (LOS) path is usually much higher (about 13 dB higher) than the sum of power of NLOS paths, as reported in [37]. Therefore, in most cases we only need to report the AoD of the strongest path back to the base station. On the other hand, if the strategy of maximal ratio transmission (MRT) is desired, a two-stage estimation scheme [38] can be employed. In the first stage, we employ a beam alignment scheme to identify paths with significant channel gains. Then, in the second stage, the base station steers its beams to directions of those significant paths and each user can estimate the "effective channel" based on the training sequence. Since the effective channel has a substantially reduced dimension as it only contains those significant paths, the training overhead required by the second stage would be considerably lower. After the effective channel is obtained, the optimal precoding vector at the transmitter can be devised.

Suppose the transmitter sends a constant signal \( s(t) = 1 \) to the receiver. The signal received at the \( t \)th time instant can be expressed as

\[ r(t) = h^T b(t) s(t) + w(t) = x^T D^T b(t) + w(t) \]  

(5)

where \( b(t) \in \mathbb{C}^N \) is the precoding/beamingforming vector used by the transmitter at the \( t \)th time instant, and \( w(t) \) denotes the additive complex Gaussian noise with zero mean and variance \( \sigma^2 \). Since a hybrid analog and digital beamforming structure is employed at the transmitter, the precoding vector can be expressed as

\[ b(t) = F_{\text{RF}}(t) f_{\text{BB}}(t) \]  

(6)

in which \( F_{\text{RF}}(t) \in \mathbb{C}^{N \times R} \) and \( f_{\text{BB}}(t) \in \mathbb{C}^R \) represent the radio frequency (RF) precoding matrix and the baseband (BB) precoding vector, respectively. Specifically, to efficiently probe the channel, the transmitter forms multiple beams simultaneously and steers them towards different directions. To this objective, the RF precoding matrix is chosen to be a submatrix of the complex conjugate of the DFT matrix, \( D^* \)

\[ F_{\text{RF}}(t) = D^* S(t) \]  

(7)

where \( S(t) \in \{0, 1\}^{N \times R} \) is a column selection matrix containing only one nonzero entry per column. Note that each column of the DFT matrix can be considered as a beamforming vector steering a beam to a certain direction. Hence, the RF precoding matrix defined in (7) forms \( R \) beams towards different directions simultaneously.

Substituting (6)–(7) into (5), we obtain

\[ r(t) = x^T a(t) + w(t) = a^T(t)x + w(t) \]  

(8)

where \( a(t) \triangleq S(t)f_{\text{BB}}(t) \) is an \( N \)-dimensional sparse vector with at most \( R \) nonzero elements. It should be noted (8) is an ideal model without taking the CFO effect and phase noise into account. CFO is caused by mismatch in the carrier frequencies at the transmitter and the receiver, which is typically in the order of several parts per million (ppm) of the carrier frequency. Since the carrier frequency is high, CFO cannot be neglected in mmWave communications. Besides the CFO, mmWave communication systems also suffer random phase noise due to the jitter of the oscillators. As pointed out in [28], the phase noise \( \phi_t \) can be modeled as a Wiener process, and the increment \( \phi_{t+1} - \phi_t \) follows a Gaussian distribution with zero mean and variance \( \beta \), where the variance \( \beta \) is proportional to the square of the carrier frequency, and thus is significant for mmWave systems. The phase noise, together with the carrier frequency offset, lead to an unknown phase shift to measurements \( r(t) \) that varies across time. In this case, only the magnitude of the measurements \( r(t), t = 1, \ldots, T \) is reliable.

Our objective is to devise a measurement matrix \( A \triangleq [a(1) \ldots a(T)]^T \in \mathbb{C}^{T \times N} \) and develop a fast and efficient recovery algorithm to recover \( z = |x| \), i.e. the support and magnitude of the sparse signal \( x \), from compressive phaseless measurements:

\[ y \triangleq |r| = |Ax + w| \]  

(9)

where \( r \triangleq [r(1) \ldots r(T)]^T \), and \( w \triangleq [w(1) \ldots w(T)]^T \). Note that \( a(t) \) is a sparse vector with at most \( R \) nonzero entries. Therefore

**C1** \( A \) is a sparse matrix with each row of \( A \) containing at most \( R \) nonzero elements.

For this reason, the design of the measurement matrix \( A \) is referred to as sparse encoding. Also, since the period of time for beamforming training is proportional to the number of measurement \( T \), we wish \( A \) is properly devised such that a reliable estimate of \( z = |x| \) can be obtained by using as few measurements as possible.

**Remark 1:** Note that constraint C1 is not a general guideline for devising the measurement matrix \( A \). Instead, this constraint is a result of the use of directional beams for beam alignment. One may wonder why we use directional beams, instead of
pseudorandom beams (corresponding to randomly generated precoding vectors \(\{b(t)\}\)) that uniformly illuminate the entire angular space, for beam alignment. One reason, as to be shown in our paper, is that the use of directional beams enables us to develop an efficient beam alignment scheme which relies on those “prominent” measurements (measurements collected when one of the directional beams coincides with one path of the channel) to identify the beam directions. Also, in the initial access stage, instead of simply transmitting a constant signal \(s(t) = 1\) as described in (5), the base station may need to broadcast a common message which includes the index number of the base station to each user. In this case, directional beams are preferable to pseudorandom beams in order to provide a sufficient (instantaneous) SNR for each user to successfully decode the message.

Remark 2: In practice, the true AoD parameters do not necessarily lie on the discretized grid, which is referred to as grid or basis mismatch. In the presence of grid mismatch, the number of nonzero entries in \(x\) will become larger than the number of signal paths \(P\) due to the power leakage caused by grid mismatch. We note that a variety of super-resolution or off-grid compressed sensing methods, e.g. [23], [25], [39]–[41], have been recently proposed to overcome the grid mismatch issue.

III. REVIEW OF PHASECODE

PhaseCode [35] is a CPR algorithm that is most relevant to our sparse encoding and phaseless decoding method. Here we first provide a brief review on PhaseCode. PhaseCode is a computationally efficient (with the order-optimal time-complexity) algorithm developed in a sparse-graph coding framework. It consists of an encoding step and a decoding step. In the encoding step, the measurement matrix \(A \in \mathbb{C}^{4M \times N}\) is devised according to

\[
A \triangleq H \odot \bar{T}
\]

(10)

where \(\odot\) denotes the Khatri-Rao product, \(H \in \{0, 1\}^{M \times N}\) is a binary code matrix constructed using a random bipartite graph \(G\) with \(N\) left nodes and \(M\) right nodes, with its \((i, j)\)th entry \(H(i, j) = 1\) if and only if left node \(j\) is connected to right node \(i\), otherwise \(H(i, j) = 0\). \(\bar{T} \in \mathbb{C}^{4 \times N}\) is the so-called “trigonometric modulation” matrix that provides 4 measurements for each row of \(H\), and \(\bar{T}\) is given by

\[
\bar{T} \triangleq \begin{bmatrix}
\omega j \omega & \omega j 2 \omega & \cdots & \omega j N \omega \\
-j\omega & -j 2 \omega & \cdots & -j N \omega \\
2 \cos(\omega) & 2 \cos(2 \omega) & \cdots & 2 \cos(N \omega)
\end{bmatrix}
\]

(11)

where \(\omega \in (0, 2\pi/N]\), and \(\omega'\) is a random phase between 0 and \(2\pi\). In the decoding stage, a delicate procedure is employed to recover \(x\). It was shown in [35] that, in the noiseless case, PhaseCode can recover a \(K\)-sparse signal with a high probability using only slightly more than \(4K\) measurements. Nevertheless, there are two major issues when applying PhaseCode to the beam alignment problem. Firstly, in PhaseCode, the bipartite graph \(G\) used to determine the binary code matrix \(H\) is randomly generated. There is no guarantee that the resulting measurement matrix \(A\) satisfies constraint C1. Secondly, PhaseCode, which involves a delicate decoding procedure requiring highly accurate measurements, may suffer severe performance degradation in the presence of noise. This makes PhaseCode unsuitable for beam alignment problems where measurements are inevitably contaminated by noise.

IV. PROPOSED SBG-CODE ALGORITHM

To overcome the drawbacks of existing solutions, we propose a sparse bipartite graph-Code (SBG-Code) algorithm for sparse encoding and phaseless decoding.

A. Sparse Encoding

Different from PhaseCode which employs a single bipartite graph, the proposed SBG-Code uses a set of bipartite graphs \(\{G_l\}_{l=1}^L\) to encode the sparse signal. Let \(H_l \in \{0, 1\}^{M \times N}\) denote the binary code matrix associated with the graph \(G_l\) with \(N\) left nodes and \(M\) right nodes. The \((i, j)\)th entry of \(H_l\) is given by

\[
H_l(i, j) = \begin{cases} 
1 & \text{if left node } j \text{ of } G_l \text{ is connected} \\
& \text{to right node } i \text{ of } G_l \\
0 & \text{otherwise}
\end{cases}
\]

(12)

Given \(\{H_l\}\), the measurement matrix \(A \in \mathbb{R}^{2ML \times N}\) is devised as

\[
A \triangleq \begin{bmatrix}
H_1 \odot T \\
H_2 \odot T \\
\vdots \\
H_L \odot T
\end{bmatrix}
\]

(13)

where \(T \in \mathbb{R}^{2 \times N}\) is a simplified trigonometric modulation matrix defined as

\[
T \triangleq \begin{bmatrix}
\cos(\omega) & 1 & \cdots & 1 \\
2 \cos(\omega) & 2 \cos(2 \omega) & \cdots & 2 \cos(N \omega)
\end{bmatrix}
\]

(14)

in which \(\omega \in (0, \pi/(2N)]\) such that \(\cos(\omega l) \in [0, 1]\). We will show later the trigonometric function \(\cos(n \omega l)\) can be replaced by a general function.

For each graph \(G_l\), each of its left nodes can be deemed as a component of the sparse signal \(x\), and each right node of \(G_l\) refers to a set of 2 measurements obtained as (see Fig. 2)

\[
y_{l,m} = |(H_l(m, :) \odot T)x| \quad \forall m = 1, \ldots, M
\]

(15)
where $H_i[m, \cdot]$ denotes the $m$th row of $H_i$. A left node, say node $n$, is called an active left node if the $n$th signal component, $x_n$, is nonzero. For a $K$-sparse signal $x$, there are $K$ active left nodes in total. A right node may be classified as a nullton, a singleton or a multiton by the following definitions

- Nullton: A right node is a nullton if it is not connected to any active left node.
- Singleton: A right node is a singleton if it is connected to exactly one active left node.
- Multiton: A right node is a multiton if it is connected to more than one active left node.

A bipartite graph which does not contain any multiton right nodes is a no-multiton-graph.

- No-Multiton-graph (NM-graph): A bipartite graph whose right nodes are either singletons or nulltons.

For our proposed SBG-Code, the purpose of employing multiple bipartite graphs is to ensure the existence of an NM-graph among the bipartite graphs with an overwhelming probability.

The bipartite graphs $\{G_l\}$ with $N$ left nodes and $M$ ($M > K$) right nodes are designed as follows. First, for simplicity, we assume $r = N/M$ to be an integer. For each graph, we randomly divide $N$ left nodes into $M$ equal-size, disjoint sets (i.e., each set has $r$ left nodes) and establish a one-to-one correspondence between $M$ sets of left nodes and $M$ right nodes. If $N$ is not an integer multiple of $M$, we can still divide $N$ left nodes into $M$ disjoint sets, with all sets except the last one, consisting of $r$ left nodes. Clearly, a right node is a singleton (nullton) if its corresponding set of left nodes contains only one (zero) active left node. As to be shown later, such a design helps maximize the probability that a bipartite graph is an NM-graph, i.e., its right nodes are either singletons or nulltons.

Clearly, for each bipartite graph $G_l$ designed as described, its corresponding binary code matrix $H_l$ has only one nonzero element per column, and at most $r$ nonzero elements per row. As a result, each row of the resulting measurement matrix $A$ contains at most $r$ nonzero elements. We can therefore choose $r \leq R$, which is equivalent to $M \geq N/R$, such that $A$ satisfies constraint C1. Once $A$ is given, the RF precoding matrices $\{F_{RF}(t)\}$ and baseband precoding vectors $\{f_{BB}(t)\}$ can be accordingly determined.

### B. Phaseless Decoding

Next, we discuss how to retrieve the support and magnitude information of $x$ from compressive phaseless measurement $y$. We first ignore the observation noise for ease of exposition and analysis, i.e.,

$$y = |Ax|$$

Let

$$A_l \triangleq H_l \odot T$$

denote the $l$th measurement sub-matrix associated with the bipartite graph $G_l$, and

$$y_l \triangleq |A_l x|$$

denote the corresponding measurements. Suppose $G_l$ is an NM-graph. If a right node is a nullton, it does not connect to any active left nodes and thus we have $y_{l,m} = 0$. Therefore we only need to consider singleton right nodes. A singleton right node means that only one nonzero component of $x$, say $x_n$, contributes to the value of $y_{l,m}$. More precisely, we can write

$$y_{l,m} = \begin{bmatrix} |x_n| \\ 2 \cos(n\omega)x_n \end{bmatrix}$$

Clearly, the magnitude and location index of $x_n$ can be readily estimated as

$$z_n = y_{l,m}^{(1)}$$

$$\hat{n} = \frac{1}{\omega} \arccos \left( \frac{y_{l,m}^{(2)}}{2y_{l,m}^{(1)}} \right)$$

where $y_{l,m}^{(1)}$ and $y_{l,m}^{(2)}$ denote the first and second entry of $y_{l,m}$, respectively. Note that graph $G_l$ is designed such that each left node is connected to one and only one right node. Therefore, by performing the estimation (20) for all singleton right nodes, we are guaranteed to find the location indices and magnitudes of all active left nodes. From the above discussion, we see that if a bipartite graph, say graph $G_l$, is an NM-graph, then $z = |x|$ can be recovered from the corresponding phaseless measurements $y_l$.

The problem is that since we do not have the support information of the sparse signal in advance, there is no guarantee that a designed graph is an NM-graph which only contains singleton and nullton right nodes. To address this issue, we employ multiple bipartite graphs for sparse encoding to increase the probability that there exists at least one NM-graph. Note that in our algorithm, we do not need to know which bipartite graph is an NM-graph. We just perform the decoding as if each graph were an NM-graph. To see this, suppose graph $G_l$ is not an NM-graph and contains a multiton. The multiton right node is a superposition of multiple active left nodes, say, $x_{n_1}$ and $x_{n_2}$, i.e.,

$$y_{l,m} = \begin{bmatrix} |x_{n_1} + x_{n_2}| \\ 2(\cos(n_1\omega)x_{n_1} + \cos(n_2\omega)x_{n_2}) \end{bmatrix}$$

Clearly, performing (20) by treating $y_{l,m}$ as a singleton yields an incorrect estimate of the location index and magnitude. Nevertheless, in this case, it is clear that the estimate of $z = |x|$ based on $y_l$, denoted as $\tilde{z}^{(l)}$, contains fewer than $K$ nonzero components. This is an important observation which helps us differentiate the correct estimate from incorrect ones. Due to the fact that $K$ is unknown in practice, given $L$ estimates $\{\tilde{z}^{(l)}\}_{l=1}^L$, the final estimate can be chosen as the one which has the largest number of nonzero entries. For clarity, our proposed algorithm is summarized in Algorithm 1.

We see that, through the use of multiple bipartite graphs, the proposed SBG-Code circumvents the complicated decoding procedure required by PhaseCode to check whether a right node is a singleton, a mergeable multiton or a resolvable multiton. Although the use of multiple bipartite graphs may result in a higher sample complexity, the simplified decoding procedure can help improve the robustness against noise.
Algorithm 1: Proposed SBG-Code Algorithm.

Given $A_l = H_l \odot \tilde{T}$ and $y_l$ for each bipartite graph $G_l$, $l = 1, \ldots, L$

for $l = 1, \ldots, L$

for $m = 1, \ldots, M$

if $y_{l,m} \neq 0$

Assume the $m$th right node is a singleton.

Estimate the magnitude and the location index of the active left node connected to the $m$th right node via (20)

end if

end for

Obtain an estimate of $z$, denoted as $\hat{z}^{(l)}$.

end for

Given the $L$ estimates $\{\hat{z}^{(l)}\}_{l=1}^L$, choose the estimate that has the largest number of nonzero entries as the final estimate.

C. Discussions

It should be noted that the cosine function used in (14) to encode the sparse signal can be replaced by a general function. For example, a linear function $f(n) = n/N$ can be employed to encode the sparse signal, in which case the trigonometric modulation matrix $T$ is replaced by

$$T = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1/N & 2/N & \cdots & 1 \\ \end{bmatrix}$$

Correspondingly, the $m$th singleton right node can be written as

$$y_{l,m} = \begin{bmatrix} |x_n| \\ \frac{n|x_n|}{N} \\ \end{bmatrix}$$

and the magnitude and location index of $x_n$ can be readily estimated as

$$\hat{z}_n = y_{l,m}^{(1)} \quad \hat{n} = \frac{Ny_{l,m}^{(2)}}{y_{l,m}^{(4)}}$$

V. THEORETICAL ANALYSIS FOR SBG-CODE

We now provide theoretical guarantees for our proposed SBG-Code scheme. We first analyze the probability of a bipartite graph being an NM-graph. To simplify our analysis, we assume $r \leq N/M$ is an integer. The results are summarized as follows.

A. Main Results

Proposition 1: Suppose we have

$$y_l = |A_l x|$$

where $x \in \mathbb{C}^N$ is a $K$-sparse signal, and the location indices of its nonzero components are uniformly distributed. $A_l$ is defined in (17), in which $H_l \in \{0, 1\}^{M \times N}$ is a binary code matrix constructed according to a given bipartite graph $G_l$. Specifically, $H_l$ (i.e., $G_l$) is designed such that each column of $H_l$ has at least one nonzero element, and the $m$th row of $H_l$ has $r_m$ nonzero elements. If $M \geq K$, then the probability that all right nodes of $G_l$ are either singletons or nulltons is upper bounded by

$$P(G_l \text{ is an NM-graph}) \leq r^K C_M^K / C_N^K \triangleq \lambda$$

where $C_N^K$ denotes the number of $K$-combinations from a set with $N$ elements. Also, the inequality (26) becomes an equality if and only if

$$r_1 = \cdots = r_M = r$$

Proof: See Appendix A.

From Proposition 1, we know that the probability of a bipartite graph being an NM-graph is maximized when $r_m = r, \forall m$, in which case each column of $H_l$ has only one nonzero element, and each row of $H_l$ has exactly $r$ nonzero elements. This result explains why we devise the bipartite graphs $\{G_l\}$ as discussed in Section IV.A. Based on this result, our proposed phaseless decoding scheme can recover $z = |x|$ from compressive phaseless measurements with a probability given as follows.

Theorem 1: Consider the phaseless decoding problem described in (16), where the measurement matrix $A \in \mathbb{R}^{2ML \times N} \lambda$ is generated according to our proposed sparse encoding scheme. If $M \geq K$, then our proposed algorithm can recover $z = |x|$ from phaseless measurements (16) with a probability exceeding

$$p = 1 - (1 - \lambda)^L$$

V. THEORETICAL ANALYSIS FOR SBG-CODE

We now provide theoretical guarantees for our proposed SBG-Code scheme. We first analyze the probability of a bipartite graph being an NM-graph. To simplify our analysis, we assume $r \leq N/M$ is an integer. The results are summarized as follows.

A. Main Results

Proposition 1: Suppose we have

$$y_l = |A_l x|$$

where $x \in \mathbb{C}^N$ is a $K$-sparse signal, and the location indices of its nonzero components are uniformly distributed. $A_l$ is defined in (17), in which $H_l \in \{0, 1\}^{M \times N}$ is a binary code matrix constructed according to a given bipartite graph $G_l$. Specifically, $H_l$ (i.e., $G_l$) is designed such that each column of $H_l$ has at least one nonzero element, and the $m$th row of $H_l$ has $r_m$ nonzero elements. If $M \geq K$, then the probability that all right nodes of $G_l$ are either singletons or nulltons is upper bounded by

$$P(G_l \text{ is an NM-graph}) \leq r^K C_M^K / C_N^K \triangleq \lambda$$

where $C_N^K$ denotes the number of $K$-combinations from a set with $N$ elements. Also, the inequality (26) becomes an equality if and only if

$$r_1 = \cdots = r_M = r$$

Proof: See Appendix A.

From Proposition 1, we know that the probability of a bipartite graph being an NM-graph is maximized when $r_m = r, \forall m$, in which case each column of $H_l$ has only one nonzero element, and each row of $H_l$ has exactly $r$ nonzero elements. This result explains why we devise the bipartite graphs $\{G_l\}$ as discussed in Section IV.A. Based on this result, our proposed phaseless decoding scheme can recover $z = |x|$ from compressive phaseless measurements with a probability given as follows.

Theorem 1: Consider the phaseless decoding problem described in (16), where the measurement matrix $A \in \mathbb{R}^{2ML \times N} \lambda$ is generated according to our proposed sparse encoding scheme. If $M \geq K$, then our proposed algorithm can recover $z = |x|$ from phaseless measurements (16) with a probability exceeding

$$p = 1 - (1 - \lambda)^L$$

where $\lambda$ is defined in (26).

Proof: See Appendix B.

Note that our proposed algorithm requires a total number of $T = 2ML$ phaseless measurements, in which $M$ is the number of right nodes per bipartite graph and $L$ is the number of bipartite graphs. From (26), we see that increasing $M$ helps achieve a higher $\lambda$, which in turn leads to a higher recovery probability for our algorithm. On the other hand, increasing $L$ improves the probability that there exists at least one NM-graph among those $L$ bipartite graphs, and thus can also enhance the recovery probability. Therefore, given a fixed total number of measurements $T$, there is a tradeoff between the choice of $M$ and $L$. Here we provide an example to illustrate this tradeoff. Suppose $N = 128$, $K = 2$ and $T = 32$. The parameters $M$ and $L$ can be chosen as one of the following cases, and the exact recovery probability of our proposed algorithm can be accordingly calculated:

- $M = 16, L = 1: p = 94.4882\%$
- $M = 8, L = 2: p = 98.6050\%$
- $M = 4, L = 4: p = 99.6454\%$
- $M = 2, L = 8: p = 99.6333\%$

From this example, we see that choosing a moderate value for $M$ and $L$ provides the best performance.

B. Analysis of Sample Complexity

Let $M = \delta K$, where $\delta > 1$ is a parameter whose choice will be discussed later. From Theorem 1, we can derive the number of bipartite graphs required for perfectly recovering $|x|$ with a probability exceeding a prescribed threshold $p_0$:

$$L \geq \frac{\log(1 - p_0)}{\log(1 - \lambda)} = \frac{\log[(1 - p_0)^{-1}]}{\log[(1 - \lambda)^{-1}]}$$

(29)
As a result, the total number of measurements required for exact recovery with a probability exceeding $p_0$ is given by

$$T = 2ML = 28KL \geq \frac{c\delta K}{\log[(1-\lambda)^{-1}]} \quad (30)$$

where $c \triangleq 2 \log[(1-p_0)^{-1}] > 0$ is a constant determined by $p_0$. Note that $\lambda$ defined in (26) can be lower bounded by

$$\lambda = \frac{M !}{M^K(K-M)!} \frac{N^K(N-K)!}{N!} \geq \frac{M !}{M^K(M-K)!} \geq \frac{(M-K+1)^K}{M^K} = \left(1 - \frac{1 - K^{-1}}{\delta}\right)^K \triangleq f(K, \delta) \quad (31)$$

Define

$$h(K, \delta) \triangleq \frac{1}{\log[(1-f(K, \delta))^{-1}]} \quad (32)$$

The term on the right-hand side of (30) can be upper bounded by

$$\frac{c\delta K}{\log[(1-\lambda)^{-1}]} \leq c\delta Kh(K, \delta) \quad (33)$$

To facilitate analyzing the sample complexity of our proposed algorithm, we choose $\delta = K$, i.e. $M = K^2$, which is a choice usually offering satisfactory performance. In this case, it can be easily proved that the function $f(K, \delta)$ decreases with an increasing $K$, and

$$\lim_{K \to +\infty} f(K, \delta) |_{\delta=K} = e^{-1} \quad (34)$$

Therefore $h(K, \delta) |_{\delta=K}$ can be upper bounded by

$$h(K, \delta) |_{\delta=K} \leq \frac{1}{\log[(1-e^{-1})^{-1}]} \approx 1.51 \quad (35)$$

Combining (33) and (35), we can reach that, when $\delta = K$, the term on the right-hand side of (30) is upper bounded by

$$\frac{c\delta K}{\log[(1-\lambda)^{-1}]} \leq 1.51cK^2 \quad (36)$$

In other words, if the total number of phaseless measurements $T$ satisfies

$$T \geq 1.51cK^2 \quad (37)$$

then our proposed algorithm can perfectly recover $\{x\}$ with a probability exceeding $p_0$. From (37), we see that the sample complexity for our proposed algorithm is of order $O(K^2)$, which, surprisingly, is independent of $N$, the dimension of the sparse signal. Such a result can be well explained because for the typical choice of $\delta = K$, the probability of a bipartite graph being an NM-graph is lower bounded by $e^{-1}$ (cf. (34)) even for an arbitrarily large $N$. But notice that the irrelevance of the sample complexity to $N$ is achieved by increasing $r$ since we have $r = N/M$ and $M$ is kept fixed as $K^2$ as $N$ grows. In the beam alignment application, $r$ cannot become arbitrarily large due to the limited number of RF chains.

Although a typical choice of $M = K^2$ is used to analyze the sample complexity, it is not difficult to reach a similar conclusion for a general choice of $M$ with $M = O(K^2)$. As a comparison, note that the sample complexity attained by most compressive phase retrieval methods [32], [33] and the AgileLink beam steering scheme [8], [24] is of order $O(K \log(N))$.

VI. ROBUST SBG-CODE ALGORITHM

The basic idea of our proposed SBG-Code algorithm is to divide the $N$ components of $x$ (i.e., $N$ left nodes) into $M$ disjoint sets, and each set of left nodes is connected to a distinctive right node. If a right node is a singleton, it means that its corresponding set of left nodes contains only one active left node whose location and magnitude can be easily estimated via (20) or (24), depending on which modulation matrix is used. Such an idea works perfectly for the noiseless case. Nevertheless, when the measurements are corrupted by noise, a perfect estimate of the magnitude of the active left node is impossible. Besides, the location index of the active left node may be incorrectly estimated as well. In the following, we develop a robust scheme for sparse encoding and phaseless decoding in the presence of noise.

A. Robust Sparse Encoding

To facilitate our following exposition, the trigonometric modulation matrix (14) or (22) is expressed as a general form as

$$T \triangleq \begin{bmatrix} 1 & 1 & \cdots & 1 \\ t_1 & t_2 & \cdots & t_N \end{bmatrix} \quad (38)$$

where $t_i \neq t_j$ for $i \neq j$, and $t_n > 0, \forall n = 1, \ldots, N$. Let $\{m_1(\ell), \ldots, m_r(\ell)\}$ denote the set of indices of the left nodes connected to the $m$th right node of the graph $G_l$. Note that the index set $\{m_1(\ell), \ldots, m_r(\ell)\}$ determined once the corresponding bipartite graph $G_l$, i.e., $H_l$, is given. Here we assume $r = N/M$ is an integer. The extension to the non-integer case is straightforward, as discussed earlier in Section IV. Also, for simplicity, the superscript $l$ used to denote the index of the bipartite graph is omitted, and in the following, $\{m_1(\ell), \ldots, m_r(\ell)\}$ is simplified as $\{m_1, \ldots, m_r\}$.

Suppose the $m$th right node is a singleton and $x_{m_i}$ is the active left node connected to the $m$th right node, in which $m_i \in \{m_1, \ldots, m_r\}$. When noise is present, the measurements corresponding to the $m$th right node of the graph $G_l$ can be expressed as

$$y_{l,m} = \begin{bmatrix} |x_{m_1} + w_{l,m}^{(1)}| \\ t_{m_1} x_{m_1} + w_{l,m}^{(2)} \end{bmatrix} \triangleq \begin{bmatrix} y_{l,m}^{(1)} \\ y_{l,m}^{(2)} \end{bmatrix} \quad (39)$$

where $w_{l,m}^{(1)}$ and $w_{l,m}^{(2)}$ denote the observation noise added to the first and the second entry of the $m$th right node, respectively. In this case, the location index of the active left node can be estimated as

$$\hat{m}_i = \arg \min_{m_i \in \{m_1, \ldots, m_r\}} \left| t_{m_i} - y_{l,m}^{(2)} / y_{l,m}^{(1)} \right| \quad (40)$$

The problem lies in that, if the index set $\{m_1, \ldots, m_r\}$ contains an element $m_j$ such that $t_{m_j}$ is close to $t_{m_i}$, then it is likely that the location index of the active left node is misidentified as $m_j$ since when noise is present, we may have

$$\left| t_{m_i} - y_{l,m}^{(2)} / y_{l,m}^{(1)} \right| > \left| t_{m_j} - y_{l,m}^{(2)} / y_{l,m}^{(1)} \right| \quad (41)$$
To improve robustness against noise, it is clear that the absolute difference $|t_{m_i} - t_{m_j}|$ should be as large as possible for any pair of indices $\{m_i, m_j\}$ chosen from the set $\{m_1, \ldots, m_r\}$. Inspired by this insight, we propose to use a different modulation matrix for each bipartite graph. Specifically, the modulation matrices are column-permuted versions of the original modulation matrix, i.e.,

$$T_l = TP_l \quad \forall l$$

(42)

where $T_l$ denotes the modulation matrix for graph $G_l$, and $P_l$ is a permutation matrix to be devised. Write

$$T_l \triangleq \begin{bmatrix} 1 & 1 & \cdots & 1 \end{bmatrix}$$

(43)

Following a similar deduction, the location index of the active left node associated with the $m$th right node can be estimated as

$$\hat{m}_i = \arg \min_{m_i \in \{m_1, \ldots, m_r\}} t^{(1)}_{m_i} = \min_{1 \leq i < j \leq r} d^{(l)}_{i,j}$$

(44)

(45)

Therefore, if the permutation matrix $P_l$ is devised such that for each right node $m$, the elements in the corresponding set $\{t^{(1)}_{m_1}, \ldots, t^{(1)}_{m_r}\}$ are sufficiently separated, then the robustness against noise can be improved. Mathematically, let us define the pairwise distance associated with the $m$th right node as

$$d^{(l)}_{m} \triangleq \min_{1 \leq i < j \leq r} \left| t^{(l)}_{m_i} - t^{(l)}_{m_j} \right|$$

Then the design of $P_l$ can be formulated as a Max-Min problem that maximizes the minimum pairwise distance among the pairwise distances associated with $M$ right nodes, i.e.,

$$\max_{P_l} \min_m d^{(l)}_{m}$$

(46)

Such an optimization can be solved by searching for all possible permutation matrices.

B. Robust Phaseless Decoding

We next devise a robust decoding scheme to estimate $z = |x|$ from noisy measurements $y$. In the noisy case, the measurements $y$ are written as

$$y = |Ax + w|$$

(47)

where the measurement matrix $A$ is expressed as

$$A \triangleq \begin{bmatrix} A_1 \\ A_2 \\ \vdots \\ A_L \end{bmatrix} = \begin{bmatrix} H_1 \otimes T_1 \\ H_2 \otimes T_2 \\ \vdots \\ H_L \otimes T_L \end{bmatrix}$$

(48)

and the modulation matrix $T_l$ for graph $G_l$ is given by (42). The measurements associated with the bipartite graph $G_l$ are given by

$$y_l = |A_l x + w_l|$$

(49)

and the measurements, $y_{l,m} \in \mathbb{R}$, corresponding to the $m$th right node of $G_l$ are expressed as

$$y_{l,m} = \left| (H_l(m, :) \otimes T_l) x + w_{l,m} \right| \quad \forall m = 1, \ldots, M$$

(50)

where $w_{l,m}$ denotes the noise added to the $m$th right node of $G_l$. Due to the presence of noise, we usually have $y_{l,m} \neq 0$ even if the $m$th right node is a nullton. Hence we first need to decide whether a right node of $G_l$ is a nullton or not. Such a problem can be formulated as a binary hypothesis test problem:

$$H_0: y_{l,m}^{(1)} = w_{l,m}^{(1)}$$

$$H_1: y_{l,m}^{(1)} = \sum_{m \in S} x_m + w_{l,m}^{(1)}$$

(51)

where $w_{l,m}^{(1)}$ is the additive complex Gaussian noise with zero mean and variance $\sigma^2$, and $S$ denotes the set of indices of those active left nodes that are connected to the $m$th right node. A simple energy detector can be used to perform the detection:

$$y_{l,m}^{(1)} \geq \epsilon$$

(52)

It is clear that $y_{l,m}^{(1)}$ under $H_0$ follows a Rayleigh distribution. Given a prescribed false alarm probability, the threshold $\epsilon > 0$ can be easily determined from the distribution of $y_{l,m}^{(1)}$ under $H_0$. Such an energy detector is able to yield satisfactory detection performance for a moderate and high signal-to-noise ratio. Note that the second row’s measurement, $y_{l,m}^{(2)}$, can be readily incorporated into this energy detector to help improve the nullton detection performance.

To proceed with our decoding scheme, we assume all nullton right nodes of $G_l$ are correctly identified. In this case, we are able to determine whether $G_l$ is an NM-graph or not. Specifically, if $G_l$ is an NM-graph, then it contains $M - K$ nullton right nodes; otherwise the number of nullton right nodes is greater than $M - K$. Although the number of active left nodes, $K$, is unknown a priori, those graphs which have the smallest number of nullton right nodes can be considered as NM-graphs and $K$ can be simply estimated as

$$\bar{K} = M - J$$

(53)

where $J$ denotes the smallest number of nullton right nodes among all graphs.

We now perform decoding on those NM-graphs. Suppose $G_l$ is an NM-graph and its $m$th right node is a singleton. Also, $x_{m_i}$ is the active left node connected to the $m$th right node. From the discussion in the previous subsection, it is clear that the magnitude and location index of this active left node can be estimated as

$$z_{m_i} = y_{l,m_i}^{(1)} \quad \hat{m}_i = \arg \min_{m_i \in \{m_1, \ldots, m_r\}} \frac{y_{l,m_i}^{(1)}}{y_{l,m_i}^{(2)}}$$

(54)

where $\{m_1, \ldots, m_r\}$ denotes the set of indices of those left nodes connected to the $m$th right node. After performing (54) for all singleton right nodes, we are able to obtain an estimate of $z = |x|$. Let $\zeta^{(l)}$ denote an estimate of $z$ obtained from the measurements associated with $G_l$. Since we may have more than one NM-graphs, we are able to collect multiple estimates of $z$. The problem lies in that, due to the existence of noise, these multiple estimates, denoted as $\{\zeta^{(1)}, \zeta^{(2)}, \ldots\}$, are not exactly the same. In the following, we propose a set-intersection scheme to combine these multiple estimates into a more accurate estimate.

To better illustrate our idea, suppose there are two NM-graphs, say $G_1$ and $G_2$, and $x_n$ is the only active left node in $x$. Recall that for each bipartite graph, the $N$ left nodes are divided into $M$ disjoint sets, with each set of left nodes connected to a different
right node. Let $S_n^{(i)}$ denote the set of left nodes to which $x_n$ belongs in graph $G_i$, and $S_n^{(j)}$ denote the set of left nodes to which $x_n$ belongs in graph $G_j$. Suppose the singleton right nodes in both $G_i$ and $G_j$ are correctly identified. Then we know that $x_n$ belongs to both $S_n^{(i)}$ and $S_n^{(j)}$. If the intersection of the two sets $S_n^{(i)}$ and $S_n^{(j)}$, $S_n^{(i)} \cap S_n^{(j)}$, contains only one element, then it must be $x_n$ and the location of $x_n$ can be uniquely determined. Such an idea can be easily extended to the scenario where there are more than two NM-graphs, and for such a case, the set-intersection scheme is more likely to succeed because the more sets are used, the higher the probability of the intersection of these sets containing only one element.

There, however, is a problem for the general case where $x$ contains multiple nonzero components (i.e., multiple active left nodes). In this case, we have no idea which set of left nodes a certain active node belongs to for each NM-graph. As a result, it is impossible to determine which sets should be put together to perform the intersection operation. To overcome this difficulty, we note that the magnitudes of those active left nodes are generally different. Hence the estimated magnitude can be used to identify a certain active left node. Without loss of generality, let $x_1, \ldots, x_K$ denote the nonzero components of $x$ in decreasing order in terms of magnitude, i.e., $|x_1| > \cdots > |x_K| > 0$. For each NM-graph, say graph $G_i$, we can obtain an estimate of $|x_i|$, denoted as $|\hat{x}_i|$. Specifically, let $\hat{z}_i, \cdots, \hat{z}_K > 0$ represent the nonzero components of $|\hat{x}_i|$, then the $k$th largest element $\hat{z}_k$ can be regarded as an estimate of $|x_k|$. For each NM-graph, say $G_i$, the set of left nodes containing $x_k$ can therefore be determined as the set of left nodes containing $\hat{z}_k$. A set intersection operation can then be performed to yield the final estimate of the location index of $x_k$. On the other hand, the magnitude of the $k$th largest component of $x$ can be estimated as the average of all estimates, i.e.,

$$|\hat{x}_k| = \frac{1}{l} \sum_{i=1}^{l} |\hat{z}_{i_k}|$$  \hspace{1cm} (55)

Note that if the intersection of the sets contains more than one element, then we randomly pick an element in the intersection set as the estimate of the location index of $x_k$. In addition, in case the intersection is an empty set, which is possible due to the incorrect association between $\{x_1, \ldots, x_K\}$ and $\{\hat{z}_1, \ldots, \hat{z}_K\}$, we randomly select an estimate from $\{\hat{x}_1^{(1)}, \ldots, \hat{x}_l^{(1)}\}$ as the final estimate. For clarity, our proposed robust SBG-Code algorithm is summarized in Algorithm 2.

We see the proposed decoding algorithm involves very simple addition and multiplication calculations. It can be readily verified that Algorithm 2 has a computational complexity $O(K N/M)$, which is even lower than those greedy-based compressed sensing algorithms, e.g. [22], [23]. Such an extremely low complexity makes it amiable for practical implementation.

VII. EXTENSIONS

A. Extension to Other Array Geometries

In Section II, we assume a ULA is employed at the transmitter. Our proposed scheme can also be extended to other array geometries. Suppose we have an antenna array with an arbitrary geometry. The channel can be written in a form similar to (2), except with its transmitter array response vector expressed as

$$\tilde{\theta}_l = \frac{1}{\sqrt{N}} \begin{bmatrix} e^{j \pi (u_1 \cos(\theta_1) + v_1 \sin(\theta_1))} \\ \vdots \\ e^{j \pi (u_N \cos(\theta_1) + v_N \sin(\theta_1))} \end{bmatrix}^{T}$$  \hspace{1cm} (56)

where $(u_n, v_n)$ is the coordinate of the $n$th antenna element. Due to the sparse scattering nature of mmWave channels, $h$ has a sparse (or an approximately sparse) representation over the dictionary $\tilde{D}$

$$h = \tilde{D}x$$  \hspace{1cm} (57)

where $\tilde{D} \in \mathbb{C}^{N \times N_1}$ is a matrix consisting of a number of steering vectors

$$\tilde{D} = [\tilde{\theta}_1 \cdots \tilde{\theta}_{N_1}]$$  \hspace{1cm} (58)

in which $\{\theta_1, \ldots, \theta_{N_1}\}$ denotes the discretized grid and $N_1 \leq N$ is the number of grid points. The parameter $N_1$ is chosen such that columns of $\tilde{D}$ are approximately orthogonal to each other, i.e. $\tilde{D}^H \tilde{D} \approx I$. For example, if a linear array with randomly spaced elements is employed, the beamwidth of the array is narrow. In this case, we have $\tilde{D}^H \tilde{D} \approx I$ even when $N_1 = N$. On the other hand, if a planar array is employed, the beamwidth of the array becomes wider. In this case, we have to choose a
smaller $N_1$ (say $N_1 = \sqrt{N}$) to ensure that $\hat{D}^H \hat{D} \approx I$ holds valid.

The signal received at the $t$th time instant can be expressed as

$$r(t) = h^T b(t) s(t) + w(t) = b^T (t) h + w(t)$$  \hspace{1cm} (59)

where $b(t) = F_{RF}(t) f_{BB}(t)$. Here the RF precoding matrix $F_{RF}(t)$ is chosen to be a submatrix of $\hat{D}^*$, i.e.,

$$F_{RF}(t) = \hat{D}^* \tilde{S}(t)$$  \hspace{1cm} (60)

where $S(t) \in \mathbb{R}^{N_1 \times R}$ is a column selection matrix. Substitute (60) into (59), we have

$$r(t) = (S(t) f_{BB}(t))^T \tilde{D}^H \tilde{D} x + w(t)$$
$$\approx (S(t) f_{BB}(t))^T x + w(t) \approx a^T(t) x + w(t)$$  \hspace{1cm} (61)

We see that the problem of beam alignment has now been converted to the same problem discussed in Section II.

Note that when an antenna array with a wide beamwidth is employed, a smaller $N_1$ has to be chosen to satisfy the quasi-orthogonality condition. Accordingly, the discretization error becomes larger. Nevertheless, the estimated beam direction, although deviates from the true direction, is still effective due to the wide beamwidth of the antenna array.

B. Extension to Antenna Array Receiver

In Section II, we assume the receiver employs an omni-directional antenna that receives in all directions. In this section, we extend to the case where both the transmitter and the receiver have antenna arrays for beam alignment. With a slight abuse of notation, we let $N_r$ and $N_t$ denote the number of antennas at the transmitter and the receiver, respectively. The mmWave channel is characterized by a geometric channel model

$$G = \sum_{p=1}^{P} \alpha_p a_r(\theta_p) a_t^H(\phi_p)$$  \hspace{1cm} (62)

where $P$ is the number of paths, $\alpha_p$ is the complex gain associated with the $p$th path, $\theta_p \in [0, 2\pi]$ and $\phi_p \in [0, 2\pi]$ are the associated azimuth angle of arrival (AoA) and angle of departure (AoD), respectively, and $a_r \in \mathbb{C}^{N_r}$, $a_t \in \mathbb{C}^{N_t}$ denotes the receiver (transmitter) array response vector. We assume that uniform linear arrays are used at both the transmitter and receiver. Since there are only a few paths between the transmitter and the receiver, the channel matrix in the beam space domain has a sparse representation

$$G = \hat{D}_t \hat{G} \hat{D}_r^H$$  \hspace{1cm} (63)

where $\hat{D}_t \in \mathbb{C}^{N_r \times N_r}$ and $\hat{D}_r \in \mathbb{C}^{N_t \times N_t}$ are the DFT matrices, and $\hat{G} \in \mathbb{C}^{N_r \times N_t}$ is a sparse matrix. Suppose the transmitter sends a constant signal $s(t) = 1$ to the receiver. The phaseless measurement received at the $t$th time instant can be expressed as

$$y(t) = |c^H(t) G b(t) s(t) + w(t)|$$
$$= |c^H(t) D_t \hat{G} D_r^H b(t) + w(t)|$$  \hspace{1cm} (64)

where $c(t)$ denotes the combining vector used at the receiver. To perform beam alignment, we can let the receiver steer its beam to a fixed direction over a period of time (or multiple beams towards different directions if multiple RF chains at the receiver are available), and let the transmitter send the codewords devised according to our proposed sparse encoding scheme. Specifically, the receiver uses a certain column of $D_t$, as its combining vector, i.e., $c(t) = D_t[\cdot, i]$, over a period of time, say $t = 1, \ldots, T$. The beamforming vector employed by the transmitter is the same as discussed in Section II, i.e., $b(t) = D_r S(t) f_{BB}(t) \equiv D_r a(t)$. Thus we have

$$y(t) = |a^H(t) \tilde{g}_i + w^*(t)| \quad t = 1, \ldots, T$$  \hspace{1cm} (65)

where $\tilde{g}_i$ denotes the $i$th column of $\hat{G}^H$. We see that the problem is now converted to the sparse encoding and phaseless decoding problem discussed in this paper, and our proposed scheme can be used to recover $|\tilde{g}_i|$. After the receiver has scanned all possible $N_r$ beam directions, we are able to obtain the full knowledge of $|\hat{G}|$, based on which the best beamformer-combiner pair can be obtained. Such a beam alignment scheme has a sample complexity of $O(N_r K^2 / R_s)$, where $R_s$ represents the number of RF chains at the receiver, and $K = \max\{K_1, \ldots, K_{N_r}\}$, with $K_i$ denoting the number of nonzero entries in the $i$th column of $\hat{G}^H$, i.e. $\tilde{g}_i$. Clearly, $K$ is much smaller than the total number of paths $P$.

VIII. SIMULATION RESULTS

We now present simulation results to illustrate the performance of our proposed SBG-Code algorithm.\(^1\) In our simulations, the transmitter employs a ULA with $N$ antennas and $R$ RF chains, where $N$ is set to $N = 128$ throughout our simulations, unless otherwise specified. The receiver uses an omni-directional antenna. The distance between neighboring antenna elements is assumed to be $d = \lambda / 2$. The linear function $f(n) = n / N$ is employed to encode the sparse signal, i.e., the trigonometric modulation matrix is given by (22), for the noiseless case. For the noisy case, the modulation matrix for the $i$th graph is given by (42), where the permutation matrix is devised according to (46).

A. Noiseless Case

We first examine the performance of our proposed algorithm in the noiseless case. The mmWave channel is assumed to have a form of (4) with $K$ paths. The nonzero components of $x$ are assumed to be random variables following a circularly symmetric complex Gaussian distribution $\mathcal{C}\mathcal{N}(0, 1)$, and the locations of nonzero entries of $x$ are uniformly chosen at random. All results are averaged over $10^4$ independent runs. The performance is evaluated via the success rate, which is computed as the ratio of the number of successful trials to the total number of independent runs. A trial is considered successful if $\| \tilde{z} - z \|_2^2 / \| z \|_2^2 < 10^{-8}$, where $\tilde{z}$ denotes the estimate of $z$.

Fig. 3(a) depicts the success rates as a function of the number of measurements $T = 2ML$, where the number of RF chains is set to $R = 8$, and the number of right nodes in each bipartite graph is set to $M = 16$. In the figure, solid lines represent the theoretical performance given in (28), while the circle marks represent the performance obtained via the Monte Carlo experiments. From Fig. 3, we see that our theoretical result matches

\(^1\)Codes are available at http://www.junfang-uestc.net/codes/SBGCode.rar
the empirical result very well. Also, when the number of paths $K$ is small, our proposed scheme can perfectly recover the AoD and the attenuation (in magnitude) of each path with a decent probability even using a small number of measurements, say $T = 32$, thus achieving a substantial overhead reduction for beam alignment. Fig. 3(b) plots the success rates as a function of the number of antennas (i.e. dimension of the sparse signal) $N$, where we set $T = 64$, $R = 8$, and $M = 16$. From Fig. 3(b), we observe that the success rate of our proposed algorithm remains almost unaltered as $N$ grows. This result corroborates our theoretical claim that our proposed algorithm has a sample complexity independent of $N$. It is also interesting to examine the impact of the choice of the number of right nodes per bipartite graph, $M$, on the performance of our proposed algorithm, given the total number of measurements $T$ fixed. Fig. 3(c) plots the success rates as a function of $M$, where we set $T = 64$ and $N = 128$. Note that since the parameter $M$ must be chosen such that $R \geq \text{ceil}(N/M)$, the number of required RF chains changes as $M$ varies. From Fig. 3(c), we see that the best performance is achieved when $M \approx K^2$.

**B. Noisy Case**

Next, a noisy case is considered. To examine the beam alignment performance with more practically relevant mmWave channels, similar to [10], we assume a Rician channel $h$ which consists of $K$ multipath components, i.e. a LOS path and $K - 1$ NLOS paths. The AoD associated with each path is assumed to lie on the discretized grid specified by the DFT matrix, and is uniformly chosen at random. The complex gain of the LOS path is given by $\sqrt{\varrho}e^{j\phi_{\text{LOS}}}$, where the Rician factor $\varrho$, defined as the ratio of the energy in the LOS path to the sum of the energy in other non-LOS paths, is set to $\varrho = 13.2$ dB according to [10], and $\phi_{\text{LOS}} \sim U[0, 2\pi]$ is the random phase of the LOS component.

We compare our method with the robust PhaseCode algorithm. As mentioned earlier in our paper, PhaseCode uses a single randomly generated bipartite graph to encode the sparse signal. The resulting measurement matrix $A$ may not satisfy constraint C1. In our simulation, we allow the constraint C1 to be violated by PhaseCode, i.e. PhaseCode may choose more than $R$ columns of $D^\dagger$ to construct the RF precoding matrix $F_{\text{RF}}(t)$. For our proposed method, the threshold in the energy detector (52) is set to $\epsilon = 3\sigma$. For a fair comparison, the beamforming vectors $b(t)$ (cf. (6)) used by different schemes are normalized to unit norm throughout our simulations. The performance is evaluated via the normalized mean squared error (NMSE) calculated as

$$\text{NMSE} = E \left[ \frac{\|z - \hat{z}\|^2}{\|z\|^2} \right]$$  (66)

Note that PhaseCode is able to retrieve the complete information of $x$. But the accuracy of the estimate of $z = |x|$ is of most concern for beam alignment. Fig. 4(a) shows the NMSEs of respective schemes as a function of $T$, where we set $M = 16$ for our proposed method, and the SNR is set to 5 dB. Here the SNR is defined as

$$\text{SNR} = 10\log(\|h\|^2/(N\sigma^2))$$  (67)

From Fig. 4(a), we see that our proposed method outperforms the robust PhaseCode method by a big margin for difference choices of $K$. The performance improvement is primarily due to...
the fact that our proposed method circumvents the complicated decoding procedure that is needed for PhaseCode and thus gains substantially improved robustness against noise. Fig. 4(b) depicts the NMSEs of respective schemes as a function of SNR, where we set $T = 64$. It can be observed that our proposed method attains a decent accuracy even in the low and moderate SNR regimes, whereas the robust PhaseCode almost fails in this case.

Next, we compare our proposed algorithm with the Agile-Link [8], a beam steering scheme which also relies on the magnitude of measurements for recovery of signal directions. Different from our work, AgileLink employs an analog beamforming structure. To transmit beams towards multiple directions simultaneously, Agile-Link divides the antenna array into a number of subarrays, with each subarray steering its beam towards a certain direction. In our simulations, the number of subarrays for Agile-Link is set to 4, a value which helps achieve the best performance. It should be noted that Agile-Link only recovers signal directions, but not $z$. The beamforming gain defined below is used as a metric to evaluate the performance of respective schemes

$$G_{BF} = E \left[ |a_t^H(\hat{\theta}_{opt})h|^2 / \|h\|^2 \right]$$

in which $\hat{\theta}_{opt}$ denotes the estimated AoD of the path that delivers the maximum energy. For the Agile-Link, $\hat{\theta}_{opt}$ is estimated as the direction with the highest probability. Fig. 5 depicts the beamforming gains of respective algorithms, where we set $K = 2$, and SNR = 5 dB for Fig. 5(a) and $T = 64$, $K = 2$, for Fig. 5(b). For our proposed method, $M$ is set to $M = 16$. We see that our proposed method generally yields a higher beamforming gain than the Agile-Link and the PhaseCode. Also, with $T = 64$ measurements, our proposed method can render a decent beamforming gain (about 0.85) even in the low SNR regime (SNR = 0 dB). Fig. 6 depicts the performance of respective schemes versus the number of antennas $N$, where we set $T = 64$, $K = 2$, and SNR = 5 dB. The parameter $M$ is set $M = 8$ for our proposed method. We see that our proposed method incurs only a mild performance loss as $N$ increases. Such a result, again, corroborates our analysis that our proposed algorithm has a sample complexity independent of $N$, and hence its performance is less susceptible to the increase of the signal dimension $N$. In contrast, the performance of the other two methods drops rapidly as $N$ grows. In Fig. 7, we plot the NMSEs and beamforming gains of respective algorithms versus the number of signal paths $K$, where we set $T = 64$ and SNR = 5 dB. Simulation results show that our proposed method consistently outperforms the other two methods for different values of $K$.

**C. Noisy Case With Grid Mismatch**

Lastly, we evaluate the performance of our proposed method in the presence of grid mismatch. The channel $h$ is generated according to the Rician model as discussed earlier, except that
the AoD of each path is uniformly chosen from $[-\pi/2, \pi/2]$ and does not necessarily lie on the discretized grid. The number of signal paths is set to 2 in this experiment. In addition to AgileLink, we also compare our proposed method with the GAMP-based compressed phase retrieval (CPR-GAMP) method [34] and the concatenated phase retrieval and NOMP (PR-NOMP) method [25]. For the CPR-GAMP, its precoding vectors $\{b(t)\}$ are randomly generated with entries uniformly chosen from the unit circle. On the other hand, to apply PR-NOMP to our problem, its precoding vector $b(t)$ is devised according to (6), with entries of $F_{RF}(t)$ uniformly chosen from unit circle and entries of $f_{BB}(t)$ randomly generated from a normal distribution. Again, to achieve a fair comparison, beamforming vectors $\{b(t)\}$ used by respective schemes are normalized to unit norm.

Fig. 8 and Fig. 9 depict the NMSEs and beamforming gains of respective schemes, where we set $M = 32$ for our proposed method. From Fig. 8 and Fig. 9, we see that, in the meaningful region where $T < N$ ($N = 128$), our proposed method presents a substantial performance improvement over the CPR-GAMP and the PR-NOMP in the low SNR regime, say SNR = 0 dB.
Note that in mmWave communications, the SNR before beam alignment is usually low, with the range of interest from −5 dB to 10 dB at most. Therefore such a performance advantage in the low SNR regime is highly desirable. The performance improvement is probably due to the fact that our proposed method relies on the prominent measurements to identify the beam directions, and thus is more resilient to low SNRs. We also observe that PR-NOMP achieves the highest beamforming gain when SNR = 10 dB. This is probably because PR-NOMP which iteratively updates AoD parameters using Newton refinements is more resilient against grid mismatch. We would like to emphasize that both CPR-GAMP and PR-NOMP have a much higher computational complexity than our proposed method. Comparing Fig. 5 with Fig. 9, we see that our proposed method suffers from a certain amount of performance loss in the presence of grid mismatch. Nonetheless, it can still provide an acceptable beamforming gain (about 0.6) even when SNR = 0 dB.

IX. CONCLUSIONS

The problem of mmWave beam alignment was examined in this paper. By exploiting the sparse scattering nature of mmWave channels, we showed that the problem of beam alignment can be formulated as a sparse encoding and phaseless decoding problem. A SBG-Code method was developed to encode the sparse signal and retrieve the support and magnitude of the sparse elements, and presents a clear advantage over existing mmWave beam alignment algorithms.

APPENDIX A

PROOF OF PROPOSITION 1

Before preceding, we first show that the probability that all right nodes of $G_l$ are either singletons or nulltons is maximized when each column of $H_l$ has only one nonzero element, i.e., each left node is connected to only one right node. Such a fact can be easily verified via an edge-deletion operation performed on $G_l$. Specifically, for each left node of $G_l$, if it has more than one edge, that is, it is connected to more than one right node, then we reserve only one edge and delete all the other edges. It is clear that after the edge-deletion operation, the number of singletons and nulltons of $G_l$ either increases or remains unchanged. Therefore, the probability that all right nodes of $G_l$ are either singletons or nulltons is maximized when each column of $H_l$ has only one nonzero element. Note that in this case, we have

$$\sum_{m=1}^{M} r_m = rM = N$$  \hspace{1cm} (69)

We now calculate the probability that $G_l$ is an NM-graph when each left node is connected to only one right node. More precisely, we divide $N$ left nodes into $M$ disjoint sets, where the $m$th set consisting of $r_m$ left nodes is connected to the $m$th right node. There are $K$ active left nodes in total. We need to calculate the probability that each set of left nodes, denoted as $S_m$, contains at most one active left node. Define

$$\mathcal{M} \triangleq \{1, \ldots, M\}$$  \hspace{1cm} (70)

Let $K \triangleq \{i_1, \ldots, i_K\}$ be a subset of $\mathcal{M}$ consisting of $K$ elements, and $\{i_{K+1}, \ldots, i_M\} = \mathcal{M} - K$ be the difference set between $\mathcal{M}$ and $K$. It can be easily verified that the number of ways of dividing $N$ left nodes into $M$ disjoint sets such that each set $S_m, m \in K$, contains only one active left node is given as

$$K! C_{N-K}^{r_{i_1}-1} C_{N-R_1-K+1}^{r_{i_2}-1} \cdots C_{N-R_K-1}^{r_{i_K}-1} C_{N-R_{K+1}-1}^{r_{i_{K+1}}-1} \cdots C_{N-R_{M-1}}^{r_{i_M}-1}$$

\begin{equation}
= \frac{K!(N-K)!}{\prod_{i=1}^{K} (r_{i_1} - 1)! \prod_{i=K+1}^{M} r_{i_1}!} \tag{71}
\end{equation}

where $R_m \triangleq \sum_{i=1}^{N} r_{i_1}$. Thus, the number of ways of dividing $N$ left nodes into $M$ disjoint sets such that each set contains at most one active left node is given by

$$n_1 \triangleq \sum_{\{i_1, \ldots, i_K\} \subseteq \mathcal{M}} \frac{K!(N-K)!}{\prod_{i=1}^{K} (r_{i_1} - 1)! \prod_{i=K+1}^{M} r_{i_1}!} \tag{72}$$

On the other hand, the total number of ways of assigning $N$ left nodes to $M$ disjoint sets is given as

$$n_2 \triangleq C_N^{r_{1}} C_{N-r_{1}}^{r_{2}} \cdots C_{N-r_{M-1}}^{r_{M}} \tag{73}$$

Therefore the probability that $G_l$ is an NM graph can be calculated as

$$P(G_l \text{ is an NM-graph}) = \frac{n_1}{n_2} = \frac{\eta(K)}{C_N^K} \tag{74}$$

where

$$\eta(K) \triangleq \sum_{\{i_1, \ldots, i_K\} \subseteq \mathcal{M}} \left( \prod_{i=1}^{K} r_{i_1} \right) \tag{75}$$

Next, we prove

$$\eta(K) \leq r^K C_M^K$$  \hspace{1cm} (76)

holds for all $1 \leq K \leq M$ when $\sum_{i=1}^{M} r_{i_1} = rM$. The inequality (76) is proved by mathematical induction. First, we prove the base case: $K = 1$. It is easy to verify that

$$\eta(1) = \sum_{i=1}^{M} r_{i_1} = rM = rC_M^1 \tag{77}$$

We then proceed to the inductive step. Suppose the following inequality holds for $K' - 1$

$$\eta(K' - 1) \leq r^{K'-1} C_{M}^{K'-1} \tag{78}$$

We need to prove

$$\eta(K') \leq r^K C_K^K \tag{79}$$

To this goal, we multiply both sides of (78) by $\sum_{i=1}^{M} r_{i_1}$, which yields

$$\eta(K' - 1) \left( \sum_{i=1}^{M} r_{i_1} \right) \leq r^{K'-1} C_{M}^{K'-1} M r$$

\begin{equation}
= M r^{K'} C_M^{K'-1} \tag{80}
\end{equation}
The left-hand side of (80) can be further written as

\[
\eta(K' - 1) = \sum_{i=1}^{M} \left( \sum_{j=1}^{r_i} \right)
\]

\[
= \sum_{i=1}^{M-K} \sum_{\{i_1, \ldots, i_{K'}\} \subseteq M - \{i\}} r_i^2 r_{i_1} \cdots r_{i_{K'}-2} + K' \eta(K')
\]

\[
= \frac{1}{M - K + 1} \sum_{\{i_1, \ldots, i_{K'}\} \subseteq M} \left( \prod_{t=1}^{K'} \frac{r_{i_t}}{r_{j_t}} \right) \sum_{j \neq k} \frac{r_{i_k}}{r_{j_k}} + K' \eta(K')
\]

\[
+ K' \eta(K') \tag{81}
\]

For any \(\{i_1, \ldots, i_{K'}\} \subseteq M\), using the inequality of arithmetic and geometric means (also referred to as the AM-GM inequality), we have

\[
\sum_{j, k} \frac{r_{i_k}}{r_{j_k}} \geq K'(K' - 1) \left( \prod_{j \neq k} \frac{r_{i_k}}{r_{j_k}} \right) \left( \frac{1}{K'(K' - 1)} \right)
\]

\[
= K'(K' - 1) \tag{82}
\]

in which the inequality becomes an equality if and only if \(r_{i_1} = \cdots = r_{i_{K'}}\). Hence, we have

\[
\sum_{\{i_1, \ldots, i_{K'}\} \subseteq M} \left( \prod_{t=1}^{K'} \frac{r_{i_t}}{r_{j_t}} \right) \geq K'(K' - 1) \eta(K')
\]

\[
\tag{83}
\]

in which the inequality (83) becomes equality if and only if \(r_1 = \cdots = r_M = r\). Combining (80), (81) and (83), we arrive at

\[
M r^{K' - 1} C_M^{K'-1} \geq \eta(K' - 1) \left( \sum_{i=1}^{M} r_i \right)
\]

\[
\geq \left( \frac{K'(K' - 1)}{M - K + 1} + K' \right) \eta(K') = \frac{MK'}{M - K + 1} \eta(K') \tag{84}
\]

From (84), we have

\[
\eta(K') \leq \frac{M - K' + 1}{MK'} M r^{K' - 1} C_M^{K'-1} = r^{K' - 1} C_M^{K'}
\]

\[
\tag{85}
\]

Thus the inductive step is proved. This completes our proof.

**APPENDIX B**

**PROOF OF THEOREM 1**

According to our proposed algorithm, we see that the support and magnitude of \(x\) can be perfectly recovered when there is at least one NM-graph in all bipartite graphs \(\{G_l\}_{l=1}^L\). Therefore, the probability that our proposed algorithm succeeds to recover the support and magnitude of \(x\) equals the probability that there is at least one NM-graph in \(\{G_l\}_{l=1}^L\), which is equivalent to

\[
p = 1 - (1 - P(G_l \text{ is an NM-graph})^L = 1 - (1 - \lambda)^L
\]

\[
\tag{86}
\]

**REFERENCES**

[1] T. S. Rappaport, J. N. Murdock, and F. Gutierrez, “State of the art in 60-GHz integrated circuits and systems for wireless communications,” Proc. IEEE, vol. 99, no. 8, pp. 1390–1436, Aug. 2011.

[2] S. Rangan, T. S. Rappaport, and E. Erkip, “Millimeter-wave cellular wireless networks: Potentials and challenges,” Proc. IEEE, vol. 102, no. 3, pp. 366–385, Mar. 2014.

[3] S. Chen and J. Zhao, “The requirements, challenges, and technologies for 5G of terrestrial mobile telecommunication,” IEEE Commun. Mag., vol. 52, no. 5, pp. 36–43, May 2014.

[4] A. L. Swindlehurst, E. Ayanoglu, P. Heydari, and F. Capolino, “Millimeter-wave massive MIMO: The next wireless revolution?” IEEE Commun. Mag., vol. 52, no. 9, pp. 56–62, Sep. 2014.

[5] M. R. Akdneniz et al., “Millimeter wave channel modeling and cellular capacity evaluation,” IEEE J. Sel. Areas Commun., vol. 32, no. 6, pp. 1164–1179, Jun. 2014.

[6] A. Alkhateeb, J. Mo, N. Gonzalez-Prelcic, and R. Heath, “MIMO precoding and combining solutions for millimeter-wave systems,” IEEE Commun. Mag., vol. 52, no. 12, pp. 122–131, Dec. 2014.

[7] C. Liu, M. Li, S. V. Hanly, P. Whiting, and I. B. Collings, “Millimeter-wave small cells: Base station discovery, beam alignment, and system design,” IEEE Wireless Commun., vol. 25, no. 4, Apr. 2018.

[8] O. Abari, H. Hassanieh, M. Rodriguez, and D. Katabi, “Millimeter wave communications: From point-to-point links to agile network connections,” in Proc. 15th ACM Workshop Hot Topics Netw., Atlanta, GA, USA, Nov. 9–10, 2016, pp. 169–175.

[9] J. Song, J. Choi, S. G. Lawer, D. J. Love, T. A. Thomas, and A. A. Ghosh, “Adaptive millimeter wave beam alignment for dual-polarized MIMO systems,” IEEE Trans. Wireless Commun., vol. 14, no. 11, pp. 6283–6296, Nov. 2015.

[10] S. Hur, T. Kim, D. J. Love, J. V. Krogmeier, T. A. Thomas, and A. Ghosh, “Millimeter wave beamforming for wireless backhaul and access in small cell networks,” IEEE Trans. Commun., vol. 61, no. 10, pp. 4391–4403, Oct. 2013.

[11] A. Alkhateeb, O. E. Ayach, G. Leus, and R. Heath, “Channel estimation and hybrid precoding for millimeter wave cellular systems,” IEEE J. Sel. Topics Signal Process., vol. 8, no. 5, pp. 831–846, Oct. 2014.

[12] S. Noh, M. D. Zoltowski, and D. J. Love, “Multi-resolution codebook and adaptive beamforming sequence design for millimeter wave beam alignment,” IEEE Trans. Wireless Commun., vol. 16, no. 9, pp. 5689–5701, Sep. 2017.

[13] M. Hussain and N. Michelusi, “Energy efficient beam-alignment in millimeter wave networks,” in Proc. 51st Asilomar Conf. Signals, Syst., Comput., Pacific Grove, CA, USA, Oct. 29–Nov. 1, 2017, pp. 1219–1223.

[14] A. Alkhateeb, G. Leus, and R. Heath, “Compressed sensing based multi-user millimeter wave systems: How many measurements are needed?” in Proc. 40th IEEE Int’l Conf. Acoust., Speech Signal Process., Brisbane, Australia, Apr. 19–24, 2015, pp. 2909–2913.

[15] P. Schniter and A. Sayeed, “Channel estimation and precoder design for millimeter-wave communications: The sparse way,” in Proc. 48th Asilomar Conf. Signals, Syst., Comput., Pacific Grove, CA, USA, Nov. 2–5, 2014, pp. 2027–2031.

[16] T. Kim and D. J. Love, “Virtual AoA and AoD estimation for sparse millimeter wave MIMO channels,” in Proc. 16th IEEE Int’l Workshop Signal Process. Advance Wireless Commun., Stockholm, Sweden, Jun. 28–Jul. 1, 2015, pp. 146–150.

[17] Z. Marzi, D. Ramasamy, and U. Madhow, “Compressive channel estimation and tracking for large arrays in mm-wave picocells,” IEEE J. Sel. Topics Signal Process., vol. 10, no. 3, pp. 514–527, Apr. 2016.

[18] Z. Zhou, J. Fang, L. Yang, H. Li, Z. Chen, and R. S. Blum, “Low-rank tensor decomposition-aided channel estimation for millimeter wave MIMO-OFDM systems,” IEEE J. Sel. Areas Commun., vol. 35, no. 7, pp. 1524–1538, Jul. 2017.

[19] X. Li, J. Fang, H. Li, and P. Wang, “Millimeter wave channel estimation via exploiting joint sparse and low-rank structures,” IEEE Trans. Wireless Commun., vol. 17, no. 2, pp. 1123–1133, Feb. 2018.

[20] Z. Gao, L. Dai, Z. Wang, and S. Chen, “Spatially common sparsity based adaptive channel estimation and feedback for FDD massive MIMO,” IEEE Trans. Signal Process., vol. 63, no. 23, pp. 6169–6183, Dec. 2015.

[21] B. Gao, C. Zhang, D. Jin, and L. Zeng, “Compressed SNR-and-channel estimation for beam tracking in 60-GHz WLAN,” China Commun., vol. 12, no. 6, pp. 46–58, Dec. 2015.

[22] J. A. Tropp and A. C. Gilbert, “Signal recovery from random measurements via orthogonal matching pursuit,” IEEE Trans. Inform. Theory, vol. 53, no. 12, pp. 4655–4666, Dec. 2007.
[23] B. Mandandipoor, D. Ramasamy, and U. Madhow, “Newtonized orthogonal matching pursuit: Frequency estimation over the continuum,” IEEE Trans. Signal Process., vol. 64, no. 19, pp. 5066–5081, Oct. 2016.

[24] H. Hassanieh, O. Abari, M. Rodriguez, M. Abdelghany, D. Katabi, and P. Indyk, “Fast millimeter wave beam alignment,” in Proc. ACM Special Interest Group Data Commun., Budapest, Hungary, Aug. 20–25, 2018, pp. 432–445.

[25] M. E. Rasekh and U. Madhow, “Noncoherent channel estimation for mm-wave massive MIMO,” in Proc. 52nd Asilomar Conf. Signals, Syst., Comput., Pacific Grove, CA, USA, Oct. 28–31, 2018, pp. 889–894.

[26] N. J. Myers and R. W. Heath, “A compressive channel estimation technique robust to synchronization impairments,” in Proc. 18th IEEE Int. Workshop Signal Process. Advances Wireless Commun., Sapporo, Japan, Jul. 3–6, 2017, pp. 1–5.

[27] N. J. Myers and R. W. Heath, “Joint CFO and channel estimation in millimeter wave systems with one-bit ADCs,” in Proc. IEEE 7th Int. Workshop Comput. Advances Multi-Sensor Adaptive Process., Curacao, Netherlands Antilles, Dec. 10–13, 2017, pp. 1–5.

[28] N. J. Myers and R. W. Heath Jr., “Message passing-based joint CFO and channel estimation in millimeter wave systems with one-bit ADCs,” 2018, arXiv:1806.06425.

[29] J. Zhu, H. Cao, and Z. Xu, “A two-stage approach to estimate CFO and channel with one-bit ADCs,” 2018, arXiv:1805.05069.

[30] M. L. Moreveca, J. K. Romberg, and R. G. Baraniuk, “Compressive phase retrieval,” Proc. SPIE, vol. 6701, 2007, Art. no. 670102.

[31] H. Ohlsson, A. Y. Yang, R. Dong, and S. Sastry, “CPRL—An extension of compressive sensing to the phase retrieval problem,” in Advances Neural Inform. Process. Syst. 25, Lake Tahoe, NV, USA, Dec. 3–6, 2012, pp. 1367–1375.

[32] Y. Shechtman, A. Beck, and Y. C. Eldar, “GESPAR: Efficient phase retrieval of sparse signals,” IEEE Trans. Signal Process., vol. 62, no. 4, pp. 928–938, Feb. 2014.

[33] S. Bahmani and J. Romberg, “Efficient compressive phase retrieval with constrained sensing vectors,” in Proc. Advances Neural Inform. Process. Syst., vol. 28, Montreal, QC, Canada, Dec. 7–12, 2015, pp. 523–531.

[34] P. Schniter and S. Rangan, “Compressive phase retrieval via generalized approximate message passing,” IEEE Trans. Signal Process., vol. 63, no. 4, pp. 1043–1055, Feb. 2015.

[35] R. Pedarsani, D. Yin, K. Lee, and K. Ramchandran, “PhaseCode: Fast and efficient compressive phase retrieval based on sparse-graph codes,” IEEE Trans. Inform. Theory, vol. 63, no. 6, pp. 3663–3691, Jun. 2017.

[36] X. Song, S. Haghhighatshoar, and G. Caire, “Efficient beam alignment for mmWave single-carrier systems with hybrid MIMO transceivers,” 2018, arXiv:1806.06425.

[37] Z. Muhi-Eldeen, L. Ivrissimtzis, and M. Al-Nuaimi, “Modelling and measurements of millimetre wavelength propagation in urban environments,” IET Microwaves, Antennas Propag., vol. 4, no. 9, pp. 1300–1309, Mar. 2010.

[38] A. Alkhateeb, G. Leus, and R. Heath, “Limited feedback hybrid precoding for multi-user millimeter wave systems,” IEEE Trans. Wireless Commun., vol. 14, no. 11, pp. 6481–6494, Nov. 2015.

[39] Z. Yang, L. Xie, and C. Zhang, “Off-grid direction of arrival estimation using sparse Bayesian inference,” IEEE Trans. Signal Process., vol. 61, no. 1, pp. 38–42, Jan. 2013.

[40] E. Candès and C. Fernandez-Granda, “Towards a mathematical theory of super-resolution,” Commun. Pure Appl. Math., vol. 67, no. 6, pp. 906–956, Jun. 2014.

[41] J. Fang, F. Wang, Y. Shen, H. Li, and R. S. Blum, “Super-resolution compressed sensing for line spectral estimation: An iterative reweighted approach,” IEEE Trans. Signal Process., vol. 64, no. 18, pp. 4649–4662, Sep. 2016.