Millimetre Wave MIMO Channel Estimation using Overlapped Beam Patterns and Rate Adaptation

Matthew Kokshoorn, He (Henry) Chen, Peng Wang, Yonghui Li, and Branka Vucetic

Abstract—This paper is concerned with the channel estimation problem in millimetre wave (mmWave) wireless systems with large antenna arrays. By exploiting the inherent sparse nature of the mmWave channel, we first propose a fast channel estimation (FCE) algorithm based on a novel overlapped beam pattern design. Compared to the existing non-overlapped beam pattern design, the overlap among multiple pilot beam patterns can increase the amount of information carried by each channel measurement and thus reduce the required channel estimation time. We develop a maximum likelihood (ML) estimator to optimally extract the path information from the channel measurements. Then, we propose a novel rate-adaptive channel estimation (RACE) algorithm, which can dynamically adjust the number of channel measurements based on the expected probability of estimation error (PEE). The performance of both proposed algorithms is analyzed. For the FCE algorithm, an approximate closed-form expression for the PEE is derived. For the RACE algorithm, a lower bound for the minimum signal energy-to-noise ratio required for a given number of channel measurements is developed based on the Shannon-Hartley theorem. Simulation results show that the FCE algorithm significantly reduces the number of channel estimation measurements compared to the existing algorithms using non-overlapped beam patterns. By adopting the RACE algorithm, we can achieve up to a 6dB gain in signal energy-to-noise ratio for the same PEE compared to the existing algorithms.

I. INTRODUCTION

Millimetre wave (mmWave) communication has been shown to be a promising technique for next generation wireless systems due to a large expanse of available spectrum [1]–[3]. This spectrum, ranging from 30GHz to 300GHz, offers a potential 200 times the bandwidth currently allocated in today’s mobile systems [4]. However, a critical challenge in exploiting the mmWave frequency band is its severe signal propagation loss compared to that over conventional microwave frequencies [5]–[7]. To compensate for such a loss, large antenna arrays can be employed to achieve a high power gain. Fortunately, owing to the small wavelength of mmWave signals, these arrays can be packed into small areas at the transmitter and receiver [8], [9]. For such mmWave systems, channel state information (CSI) is essential for effective communication and precoder design. However, the use of large antenna arrays results in a large multiple-input multiple-output (MIMO) channel matrix. This makes the channel estimation of mmWave systems very challenging due to the large number of channel parameters to be estimated. Moreover, owing to the high frequency, it is often not feasible to obtain digital samples from each antenna [1]. To resolve this high frequency sampling problem, analog beamforming techniques have been proposed and widely adopted in open literature (see [8], [10]–[12] and references therein). The main idea of analog beamforming is to control the phase of the signal transmitted or received by each antenna via a network of analog phase shifters.

Using analog beamforming techniques, the most straightforward channel estimation method is to exhaustively search in all possible angular directions. Specifically, consider a system with $N$ transmit antennas and $N$ receive antennas. If we aim at achieving a minimum angular resolution of $\pi/N$, an exhaustive search-based channel estimation would then require a set of $N$ transmit beamforming vectors at the transmitter designed to span all possible beam directions and likewise with $N$ receive beamforming vectors at the receiver. By searching all possible $N^2$ combinations, an $N \times N$ matrix can be formed whose entries represent the channel gains between the $N$ transmit and the $N$ receive beams. This matrix is commonly referred to as the virtual channel matrix [13]–[17]. Despite the large number of entries expected for the mmWave MIMO channel matrix, it has been shown in recent measurements [18], [19] that the mmWave channel exhibits sparse propagation characteristics in the angular domain. That is, there are only a few dominant propagation paths in mmWave channels. This sparsity can be seen in the virtual channel matrix, as only a limited number of transmit and receive direction combinations have a non-zero gain [13]. Therefore, the key objective of mmWave channel estimation is to identify these paths so that the transceiver can align the transmit and receive beams along these paths.

Recently, some compressed sensing based channel estimation algorithms have been proposed to explore the channel sparsity in mmWave systems, e.g., [5], [8], [20], [21]. The fundamental idea in some of these approaches is to search multiple transmit/receive directions in each measurement, by creating initial beam patterns that span a wider angular range than those used by the exhaustive search. Similar adaptive beamforming algorithms and multi-stage codebooks were proposed in [12], [22]–[24]. By initially using wider beam patterns, these approaches are able to reduce the number of measurements required for channel estimation. However, this introduces a loss of directionality gain, leading to a lower signal-to-noise ratio (SNR) at the receiver and a higher probability of estimation error (PEE). In this sense, there exists a challenging trade-off between estimation time and accuracy for mmWave channel estimation.

As one of the seminal works on multi-stage codebook approaches, [5] developed a “divide and conquer” search type algorithm to estimate sparse mmWave channels. As shown in Fig. 1, in each stage of this algorithm, the possible ranges
of angles of departure (AODs) and angles of arrival (AOAs) are both divided into $K << N$ non-overlapped angular sub-ranges. Correspondingly, $K$ non-overlapped beam patterns are designed at both the transmitter and receiver such that each transmit (receive) beam pattern exactly covers one AOD (AOA) sub-range. The channel estimation carried out in each stage consists of $K^2$ time slots. In each time slot, the pilot signal is transmitted using one of the $K$ beam patterns at the transmitter, and then collected by one of the $K$ beam patterns at the receiver. The corresponding channel output for each pair of transmit-receive beam patterns can then be obtained. These $K^2$ time slots span all the combinations of transmit-receive beam patterns. By comparing the magnitudes of the corresponding $K^2$ channel outputs, the transmit/receive sub-ranges that the AOD/AOA most likely belong to are determined. The receiver can then feed back the AOD information for use at the transmitter. Afterwards, the algorithm will limit the estimation to the angular sub-range identified at each link end in the previous stage and further divide it into $K$ sub-ranges for the channel estimation in the next stage. This process continues until the smallest beam width resolution is reached. It is shown in [5] that the algorithm requires estimation time proportional to $K^2 \log_2(N)$ per path. Despite the significant improvement when compared to the exhaustive search approach, such a channel estimation algorithm still might not be quick enough to track the fast channel variations, especially for mmWave mobile channels with rapidly changing parameters. Furthermore, at high SNR it may not be necessary to perform so many measurements, which would consume some unnecessary energy.

Motivated by this, in this paper we develop a fast mmWave MIMO channel estimation framework by designing a set of novel overlapped beam patterns that can significantly reduce the number of channel estimation measurements. In order to improve estimation accuracy, we then introduce a novel rate-adaptive channel estimation approach, where the average number of channel measurements is adapted to channel conditions. The main contributions of this paper are summarized as follows:

- Relying on novel overlapped beam pattern design, we first present a fast channel estimation (FCE) algorithm for mmWave systems. In this algorithm, we develop a maximum likelihood (ML) detector to optimally extract the channel AOD/AOA information from the measurements. We also design a linear minimum mean squared error (LMMSE) channel estimator to estimate the channel coefficients by optimally combining the selected measurements in all stages.

- We then develop a rate-adaptive channel estimation (RACE) algorithm, in which additional measurements are permitted when the current measurements are found to have an inadequate probability of success. In this way, the number of measurements can adapt to the channel conditions and the channel estimation accuracy can be significantly improved with minimal measurements.

- We analyze the probability of channel estimation error for the FCE algorithm. In particular, we derive a closed-form approximation, lower bound and upper bound for the PEE. Based on the Shannon-Hartley theorem, we also provide some theoretical analysis for the minimum energy required to estimate the channel using the RACE algorithm.

- Finally, we compare the performance of the proposed algorithms to that of the algorithm in [5] with non-overlapped beam patterns. Simulation results show that both of the proposed algorithms can significantly reduce the number of channel measurements compared to [5]. We show that the FCE algorithm achieves a guaranteed reduction of channel measurements, at the expense of estimation accuracy. On the other hand, the RACE algorithm can achieve the same average reduction of channel measurements as the FCE algorithm, but using up to 6dB less signal energy compared to the algorithm in [5] under the same PEE.

**Notation:** $A$ is a matrix, $a$ is a vector, $\alpha$ is a scalar, and $A$ is a set. $\|A\|_2$ is the 2-norm of $A$, det$(A)$ is the determinant of $A$ and $|A|$ represents the cardinality of $A$. $A^T$, $A^H$ and $A^*$ are the transpose, conjugate transpose and conjugate of $A$, respectively. For a square matrix $A$, $A^{-T}$ represents its inverse. We use diag$(a)$ to denote a diagonal matrix with entries of $a$ on its diagonal. $I_N$ is the $N \times N$ identity matrix, $1_N$ is an $N \times 1$ all 1 column vector and $\lceil \cdot \rceil$ denotes the ceiling function. $CN(m, R)$ is a complex Gaussian random vector with mean $m$ and covariance matrix $R$. $E[a]$ and Cov$(a)$ denote the expected value and covariance of $a$, respectively. $A \otimes B$ denotes the Kronecker product of $A$ and $B$ whereas $A \odot B$ denotes the row-wise Kronecker product of $A$ and $B$.

## II. System Model

Consider a mmWave MIMO system composed of $N_t$ transmit antennas and $N_r$ receive antennas. We consider that both the transmitter and receiver are equipped with a limited number of radio frequency (RF) chains. Following [5], we further assume that these RF chains, at one end, can only be combined to form a single beam pattern, indicating that only one pilot signal can be transmitted and received at one time. To estimate the channel matrix, the transmitter sends a pilot signal $x$, with unit energy $||x||_2 = 1$, to the receiver. Denote by $f$ and $w$ (\(||f||_2 = ||w||_2 = 1\)), respectively, the $N_t \times 1$
beamforming vector at the transmitter and \( N_r \times 1 \) beamforming vector at the receiver. The corresponding channel output can be represented as
\[
y = \sqrt{P}w^H H f x + w^H q,
\]
where \( H \) denotes the \( N_t \times N_t \) MIMO channel matrix, \( P \) is the transmit power and \( q \) is an \( N_r \times 1 \) complex additive white Gaussian noise (AWGN) vector following distribution \( \mathcal{C}N(0, N_0 I_{N_r}) \).

In this paper we follow [25] and adopt a two-dimensional (2D) sparse geometric-based channel model. Specifically, we consider an \( L \) path channel between the transceiver, with the \( l \)th path having steering AOD, \( \phi^t_l \), and AOA, \( \phi^r_l \) where \( l = 1, \ldots, L \). Then the corresponding channel matrix can be expressed in terms of the physical propagation path parameters as
\[
H = \sqrt{N_t N_r} \sum_{l=1}^{L} \alpha_l \mathbf{a}_t(\phi^t_l) \mathbf{a}_r^H(\phi^r_l)
\]
where \( \alpha_l \) is the fading coefficient of the \( l \)th propagation path, and \( \mathbf{a}_t(\phi^t_l) \) and \( \mathbf{a}_r(\phi^r_l) \) respectively denote the transmit and receive spatial signatures of the \( l \)th path.

Here, the steering angle, \( \phi^t_l \), is related to the physical angle \( \theta^t_l \in [0, \pi) \) by \( \phi^t_l = \frac{d \sin(\theta^t_l)}{\lambda} \) with \( \lambda \) denoting the signal wavelength\(^1\). A similar expression can be written for \( \phi^r_l \) at the receiver. With half-wavelength spacing, the distance between antenna elements becomes \( d = \lambda/2 \).

From (2), we can see that the overall channel state information of each path includes three parameters, i.e., the AOD \( \phi^t_l \), the AOA \( \phi^r_l \), and the fading coefficient \( \alpha_l \). We assume that the fading coefficient of each path follows a complex Gaussian distribution with zero mean and variance \( P_R \) and that both \( \phi^t_l \) and \( \phi^r_l \) can only take some discrete values from the set \( \mathcal{U}_N = \{0, \frac{\pi}{N}, \ldots, \frac{N(1-1)}{N}\} \). We aim to find an efficient way to estimate these three parameters for each path. The key challenge here is how to design a sequence of \( f \)'s and \( w \)'s in such a way that the channel parameters can be quickly and accurately estimated.

We consider \( M \) pairs of beam patterns that are designed to span all possible transmit-receive combinations. Denote by \( f_m \) and \( w_m \), respectively, the transmit and receive beamforming vectors adopted in the \( m \)th channel measurement time slot such that \( ||f_m||_2 = ||w_m||_2 = 1, \forall m = 1, \ldots, M \). Similarly to [5], we assume the same pilot symbol \( x \) is transmitted during the \( M \) time slots. Then, after \( M \) time slots, we can obtain a sequence of \( M \) measurements represented as
\[
y = \sqrt{P}x h + n,
\]
where \( h \) describes the channel input-output relationship for a given set of transmit and receive beamforming vectors defined by
\[
h_v = \begin{bmatrix} w_1^H H f_1 \\ w_2^H H f_2 \\ \vdots \\ w_M^H H f_M \end{bmatrix}
\]
and
\[
n = \begin{bmatrix} w_1^H q_1 \\ w_2^H q_2 \\ \vdots \\ w_M^H q_M \end{bmatrix}
\]
is an \( M \times 1 \) vector of the corresponding noise terms. Note that since \( ||w_m||_2 = 1, \forall m \), the vector \( n \) follows the same distribution as that of \( q_m, i.e., n \sim \mathcal{C}N(0, N_0 I_M) \).

Motivated by the geometric sparsity of the mmWave channel, in the following two sections, we propose to use a set of overlapped beam patterns that are able to estimate the AOD/AOA information very quickly. We then extend the algorithm to use a rate-adaptive estimation approach. The adaptive nature of the algorithm permits additional measurements to be performed under poor channel conditions. This allows the fast channel estimation to be carried out with significant accuracy and energy efficiency.

III. FAST CHANNEL ESTIMATION WITH OVERLAPPED BEAM PATTERNS

In this section, we develop a fast channel estimation framework for mmWave systems using overlapped beam patterns. Specifically, we design a set of beam patterns that, in different measurement intervals, are overlapped with one another in the angular domain. A maximum likelihood based estimation algorithm is then proposed to accurately retrieve the channel state information from the set of measurements. The proposed channel estimation algorithm also works in a similar multi-stage manner as that in [5] where each stage reduces the possible sub-ranges in which the AOD/AOA are expected to be found.

A. An Example of Overlapped Beam Pattern Design

We will first explain the design principle of overlapped beam patterns using a simple example. Following Fig. 1, we divide the AOD/IAOA angular spaces into \( K = 3 \) sub-ranges in the first stage, denoted by \( S_1 = \{e \in \mathcal{U}_N | 0 \leq e < \pi/3\}, S_2 = \{e \in \mathcal{U}_N | \pi/3 \leq e < 2\pi/3\} \) and \( S_3 = \{e \in \mathcal{U}_N | 2\pi/3 \leq e < \pi\} \), respectively. However, instead of using 3 beam patterns to cover them at each transceiver end as in Fig. 1(a), we propose to use only 2 overlapped beam patterns to achieve this. Fig. 2(a) illustrates our designed beam patterns in the first stage. We can see that the first and second beam patterns cover \( S_1, S_2 \) and \( S_2, S_3 \), respectively, and are overlapped in...
the whole range of $S_2$. Intuitively, if a path is observed in two measurements using adjacent beam patterns, the AOD or AOA must belong to the overlapped sub-range of these two beam patterns. It is also seen that each beam pattern can have different amplitudes in different sub-ranges. We represent the amplitudes of each beam pattern in different sub-ranges by a vector. For beam pattern 1 and 2 in Fig. 2(a), these vectors are respectively defined as

$$b_1 = [b_{1,1}, b_{1,2}, b_{1,3}], \quad b_2 = [b_{2,1}, b_{2,2}, b_{2,3}]$$

(7)

where $b_i$ corresponds to the $i$th beam pattern with $b_{i,k}$ denoting the amplitude of the $i$th beam pattern in sub-range $S_k$, $\forall k = 1, ..., K$. By using $M = 4$ measurement time slots, we can then span all beam pattern combinations between the transceiver. We denote the sequential set of beam patterns respectively adopted at the transmitter and receiver by

$$B_T^{(M)} = \begin{bmatrix}
    b_1 \\
    b_2 \\
    b_2 \\
    b_2
\end{bmatrix} = \begin{bmatrix}
    b_{1,1}, b_{1,2}, b_{1,3} \\
    b_{1,1}, b_{1,2}, b_{1,3} \\
    b_{2,1}, b_{2,2}, b_{2,3} \\
    b_{2,1}, b_{2,2}, b_{2,3}
\end{bmatrix},$$

(8)

$$B_R^{(M)} = \begin{bmatrix}
    b_1 \\
    b_2 \\
    b_2 \\
    b_2
\end{bmatrix} = \begin{bmatrix}
    b_{1,1}, b_{1,2}, b_{1,3} \\
    b_{1,1}, b_{1,2}, b_{1,3} \\
    b_{2,1}, b_{2,2}, b_{2,3} \\
    b_{2,1}, b_{2,2}, b_{2,3}
\end{bmatrix}.$$  

(9)

We refer to these as the beam pattern design matrices, with their $m$th row denoting the beam pattern adopted in the $m$th measurement time slot, where $m = 1, ..., M$. The efficient design of these beam patterns can lead to many solutions. However, one desirable property is that the same quantity of signal energy is transmitted/received via each sub-range over all measurements, i.e., the energy of each column of (8)-(9) should have the same Euclidean norm. This provides the same accuracy for each possible sub-range combination. Another desirable property is that the transmit/receive beamforming gains of all measurement patterns are equal, i.e., the energy of each rows of (8)-(9) should be the same.

One possible way to make the beam pattern sub-range amplitudes in (7) follow the aforementioned properties, is to normalize $B_T^{(M)}$ and $B_R^{(M)}$ to have unit energy in each row and have equal energy in each column. For the beam patterns in Fig. 2, we have $b_{1,3} = b_{2,1} = 0$, as beam patterns 1 and 2 do not cover, respectively, the third and first sub-ranges. Due to the symmetry between the two beam patterns, we further have $b_{1,1} = b_{2,3} = \beta_1$ and $b_{1,2} = b_{2,2} = \beta_2$. This leads to $\beta_1 = \frac{1}{\sqrt{3}}$ and $\beta_2 = \frac{\sqrt{2}}{\sqrt{3}}$, and the matrices in (8)-(9) become

$$B_T^{(M)} = \begin{bmatrix}
    \frac{\sqrt{2}}{\sqrt{3}} & \frac{1}{\sqrt{3}} & 0 \\
    0 & \frac{1}{\sqrt{3}} & \frac{\sqrt{2}}{\sqrt{3}} \\
    0 & \frac{1}{\sqrt{3}} & \frac{\sqrt{2}}{\sqrt{3}}
\end{bmatrix}, \quad B_R^{(M)} = \begin{bmatrix}
    \frac{\sqrt{2}}{\sqrt{3}} & \frac{1}{\sqrt{3}} & 0 \\
    0 & \frac{1}{\sqrt{3}} & \frac{\sqrt{2}}{\sqrt{3}} \\
    0 & \frac{1}{\sqrt{3}} & \frac{\sqrt{2}}{\sqrt{3}}
\end{bmatrix}.$$  

(10)

In order to observe the resultant amplitude gains (i.e., the transceiver gain) over each of the $K^2 = 9$ sub-range combinations, we introduce another matrix referred to as the generator matrix. We define this as the row-wise Kronecker product between the transmit beam pattern design matrix and the receive beam pattern design matrix such that

$$G^{(M)} = B_T^{(M)} \otimes B_R^{(M)} = \begin{bmatrix}
    b_1 \otimes b_1 \\
    b_1 \otimes b_2 \\
    b_2 \otimes b_1 \\
    b_2 \otimes b_2
\end{bmatrix}$$

(11)

$$= \frac{1}{3} \begin{bmatrix}
    2 & \sqrt{2} & 0 & \sqrt{2} & 0 & 0 & 0 & 0 & 0 \\
    0 & \sqrt{2} & 2 & 0 & 1 & \sqrt{2} & 0 & 0 & 0 \\
    0 & 0 & 0 & \sqrt{2} & 1 & 0 & 1 & \sqrt{2} & 0 \\
    0 & 0 & 0 & 0 & 1 & \frac{\sqrt{2}}{2} & 0 & \sqrt{2} & 1
\end{bmatrix},$$

(12)

recalling that $\otimes$ and $\odot$ represent the row-wise Kronecker and Kronecker product operations respectively. By denoting $b_T^{(m,k)}$ and $b_R^{(m,k')}$ as the entry on the $k$th and $k'$th column on the $m$th row in $B_T^{(M)}$ and $B_R^{(M)}$, respectively, we can express the $d$th column of $G^{(M)}$ as

$$G_d^{(M)} = \begin{bmatrix}
    b_T^{(1,k)} \odot b_R^{(1,k')} \\
    \vdots \\
    b_T^{(M,k')} \odot b_R^{(M,k')}
\end{bmatrix}$$

(13)

where the relationship $d = K(k^t - 1) + k'$ is a result of the Kronecker product operation.

The columns of generator matrix describe the $M = 4$ received measurement gains over each of the $K^2 = 9$ sub-range combinations. For example, if a path were present between the transmit sub-range $k^t = 3$ and the receive sub-range $k' = 1$, the measurement vector $y$ would be expected to be a scalar multiple of column $d = 7$ of $G^{(M)}$, expressed as $G_7^{(M)}$. It is important to note that the sets of beam patterns used in this paper are not unique. We later show that their performance, in terms of PEE, depends only on the Euclidean distance between each column of $G^{(M)}$, which directly determines the probability that one column corresponding to a certain sub-range is to be mistaken for another. In the example set shown in (11), each column has the same equal minimum Euclidean distance when compared with all other columns.
although some have more spatial neighbours at this minimum distance than others.

B. Beamforming Vector Design

To generate the beam patterns illustrated in Fig. 2(a) and described in (10), the transmit and receive beamforming vectors should be designed as follows. Denote by \( f_m \) and \( w_m \), respectively, the transmit beamforming vector and receive beamforming vector corresponding to the \( m \)th pair of beam patterns in \( B_T^{(M)} \) and \( B_R^{(M)} \). Then we design the product of the transmit array response and transmit beamforming vector to have

\[
u^H(x)f_m = Cb_T^{(m,k)}, \quad \exists k \in \{1, 2, ..., K\}, \epsilon \in S_k,
\]

and the product of the receiver array response and receive beamforming vector to have

\[
u^H(x)w_m = Cb_R^{(m,k)}, \quad \exists k \in \{1, 2, ..., K\}, \epsilon \in S_k,
\]

where \( u(\epsilon) \) has been defined in (3) and \( C \) is a scalar constant that ensures \( \|f_m\|_2 = \|w_m\|_2 = 1 \). Physically, \( C \) corresponds to the average directivity gain of each beam pattern and is the same for all \( m = 1, ..., M \) due to the normalization of the rows in (10). Eqs. (14) and (15) can be expressed in a matrix form as

\[
U^H f_m = \left[ \begin{array}{c}
Cb_T^{(m,1)}|S_1| \\
\vdots \\
Cb_T^{(m,K)}|S_K|
\end{array} \right] \triangleq z_T^{(m)}
\]

and

\[
U^H w_m = \left[ \begin{array}{c}
Cb_R^{(m,1)}|S_1| \\
\vdots \\
Cb_R^{(m,K)}|S_K|
\end{array} \right] \triangleq z_R^{(m)}
\]

where \( |S| \) denotes the cardinality of set \( S \) and \( U = [u(0), u(\frac{\pi}{N}), \ldots, u(\frac{(N-1)\pi}{N})] \) is a matrix whose columns describe the antenna array response at each angle. Therefore \( f_m \) and \( w_m \) can be designed as

\[
f_m = (UU^H)^{-1}Uz_T^{(m)}
\]

and

\[
w_m = (UU^H)^{-1}Uz_R^{(m)}
\]

where \( (UU^H)^{-1}U \) is the pseudo inverse of \( U \).

C. Channel Measurements

We now perform channel estimation in the first stage using the previously designed transmit and receive beamforming vectors \( \{f_m\} \) and \( \{w_m\} \). In each time slot, the beamforming vectors \( f_m \) and \( w_m \) are adopted to transmit/receive the pilot signal \( x \). If we substitute the channel in (2) into (5), we get

\[
h_v^{(M)} = xN \sum_{l=1}^{L} \alpha_l \begin{bmatrix}
(a_l^H(\phi_l^r)f_1)w_1^H & a_l^H(\phi_l^r)f_1 \\
\vdots & \vdots \\
a_l^H(\phi_l^r)f_M^H & a_l^H(\phi_l^r)f_M
\end{bmatrix}.
\]

Without loss of generality, let us consider the case with AOD, \( \phi_l^r \in S_{k_l} \) and AOA, \( \phi_l^t \in S_{k_l} \), where \( k_l \) and \( k_l^t \) are respectively, the transmit and receive sub-range indices of the \( l \)th propagation path. By recalling (14)-(15) we write

\[
u^H(\phi_l^t)f_m = Cb_T^{(m,k_l)} \quad \text{and} \quad \nu^H(\phi_l^r)w_m = Cb_R^{(m,k_l^t)}
\]

which leads to

\[
h_v^{(M)} = xNC^2 \sum_{l=1}^{L} \alpha_l \begin{bmatrix}
b_l^{(1,k_l)} & b_l^{(1,k_l^t)} \\
\vdots & \vdots \\
b_l^{(M,k_l)} & b_l^{(M,k_l^t)}
\end{bmatrix}.
\]

We can see from (13) that the vector term in (22) is the weighted sum of the columns of \( G^{(M)} \), i.e., the weighted sum of columns \( d_l = K(k_l^t - 1) + k_l^t, \forall l = 1, ..., L \), in \( G^{(M)} \). Therefore we can express \( h_v \) by the generator matrix as

\[
h_v^{(M)} = xNC^2G^{(M)}v^T
\]

where \( v \) is a \( 1 \times K^2 \) sparse row vector that describes the channel gain at each of the \( K^2 \) sub-range combinations by

\[
v_{d_l} = \alpha_l, \forall l = 1, ..., L
\]

and zero otherwise. For example with \( K = 3 \), a single path (i.e., \( L = 1 \)) with coefficient \( \alpha_1 \), exists on the first transmit sub-range (i.e., \( k_1^t = 1 \)) and second receive sub-range (i.e., \( k_2^t = 2 \)) leads to \( d_1 = K(k_1^t - 1) + k_1^t = 2 \) and

\[
v = [0, \alpha_1, 0, 0, 0, 0, 0, 0, 0, 0].
\]

Finally by using (23), we can re-write the received channel output vector defined in (1), after \( M \) measurements, as

\[
y^{(M)} = \sqrt{P}NC^2xG^{(M)}v^T + n^{(M)}
\]

D. Maximum Likelihood Detection of AOD/AOA Information

We now require an efficient means of detecting \( v \) given that a generator matrix \( G^{(M)} \) has been used to obtain the channel outputs in (26). Due to its optimal detection properties, this subsection elaborates how to implement a maximum likelihood detection [26] method to extract the AOD/aoa information from the received measurements. We begin by considering the distribution of \( y^{(M)} \). From (26), this can be expressed as

\[
y^{(M)} \sim \mathcal{CN} (\sqrt{PN^2C^2}x, E[G^{(M)}v^T] + \text{E}[n^{(M)}],
\]

\[
PNC^2||x||^2 \text{Cov}[G^{(M)}v^T] + \text{Cov}[n^{(M)}]).
\]

Recall that \( n^{(M)} \sim \mathcal{CN}(0, N^{-1}J_M), \) where \( J_M \) is the \( M \times M \) identity matrix. Also recall that the pilot signal \( x \) has unit energy, i.e., \( ||x||_2 = 1 \). For the signal component, as each of the path coefficients \( \alpha_l \) have zero mean we can write \( E[G^{(M)}v^T] = 0 \) and

\[
\text{Cov}[G^{(M)}v^T] = E[G^{(M)}v^T(G^{(M)}v^T)^H]
\]

\[
= N^2C^4G^{(M)}E[v^Tv](G^{(M)})^H.
\]

By defining a binary version of \( v \), denoted by \( \bar{v} \), with elements defined by

\[
\bar{v}_d = \begin{cases}
1, & \text{if } ||v_d||_2 > 0, \forall d = 1, ..., K^2 \\
0, & \text{otherwise}
\end{cases}
\]

we can separate the AOD/aoa information in \( \bar{v} \) from each of the path coefficient variances, \( P_R \). As each path coefficient
has variance $E[\alpha_i\alpha_i^T] = P_R, \forall i$, this then gives $E[u^T v] = P_Rv^T\bar{v}$. We can then re-write the distribution of $y^{(M)}$ as

$$y^{(M)} \sim \mathcal{CN}(0, \Sigma_u)$$

(31)

where

$$\Sigma_u = PN^2C^AP_RG^{(M)}v^T(vG^{(M)})^H + N_0I_M.$$  

(32)

It can now be seen that $y^{(M)}$ follows a zero mean, circularly symmetric complex Gaussian (CSCG) distribution with corresponding probability density function (PDF) defined as [27]

$$f(y^{(M)}|v, G^{(M)}) = \frac{1}{\pi^M \det(\Sigma_u)^{1/2}} \exp\left(-\frac{(y^{(M)})^H \Sigma_u^{-1} y^{(M)}}{2}\right).$$

(33)

Now let us find the conditional probability of $\bar{v}$, given the receive measurement vector $y^{(M)}$ and knowledge of $G^{(M)}$, denoted by $p(\bar{v}|y^{(M)}, G^{(M)})$. Define $\mathcal{V}$ as the set of all possible binary channel realizations such that $\bar{v} \in \mathcal{V}$. We also define $|\mathcal{V}|$ to represent the cardinality of this set. Following the principle of maximum likelihood detection and based on Bayes rule [28], we can express the probability of $\bar{v}$ for all possible $\bar{v} \in \mathcal{V}$ as

$$p(\bar{v}|y^{(M)}, G^{(M)}) = \frac{f(y^{(M)}|\bar{v}, G^{(M)})p(\bar{v})}{p(y^{(M)}|G^{(M)})}$$

(34)

where

$$p(y^{(M)}|G^{(M)}) = \sum_{\bar{v} \in \mathcal{V}} f(y^{(M)}|\bar{v}, G^{(M)})p(\bar{v}).$$

(35)

is independent of a particular channel realization. We assume that each channel realization is equiprobable, therefore

$$p(\bar{v}) = \frac{1}{|\mathcal{V}|} \quad \forall \bar{v} \in \mathcal{V}. \quad \text{(36)}$$

We then denote the probability that the $d$th element of $\bar{v}$ has a path by $p(\bar{v}_d = 1), \forall d = 1, \ldots, K^2$. We can express this probability as the sum of all $p(\bar{v}|y^{(M)}, G^{(M)})$ in (34) in which $\bar{v}_d = 1$ by

$$p(\bar{v}_d = 1|y^{(M)}, G^{(M)}) = \sum_{\bar{v} \in \mathcal{V}} p(\bar{v}|y^{(M)}, G^{(M)}) \quad \text{for } \bar{v}_d = 1.$$ 

(37)

Following the maximum likelihood approach we then find the most likely sub-range combination by

$$\hat{d} = \arg\max_d [p(\bar{v}_d = 1|y^{(M)}, G^{(M)})].$$

(38)

Finally by finding the most likely transmitter and receiver subranges through

$$\hat{k}_t = \left\lceil \frac{\hat{d}}{K} \right\rceil, \quad \hat{k}_r = \hat{d} - K(\hat{k}_t - 1) \quad \text{(39)}$$

we can reduce the ranges of possible AOD and AOA to, respectively, the $\hat{k}_t$th transmit and $\hat{k}_r$th receive angular subranges. Each of these two sub-ranges will be further divided into another $K$ sub-ranges for the channel estimation in the next stage.

---

**Algorithm 1: Fast channel estimation (FCE) algorithm for mmWave channels using overlapped beam patterns**

**Input:** Transmitter and receiver both know, $N$, $K$ and have $B_T^{(s,M)}$, $B_R^{(s,M)}$

**Initialization:** Set initial sub-ranges, $S^{(1)}_{t,k}, S^{(1)}_{r,k}, \forall k = 1, \ldots, K$

**for** $s \leq S$ **do**

// Calculate:

- $\{f_m^{(s)}\}$ based on $\theta_{t,k}^{(s)}$, $\forall k = 1, \ldots, K$ and $B_T^{(s,M)}$
- $\{w_m^{(s)}\}$ based on $S^{(s)}_{r,k}$, $\forall k = 1, \ldots, K$ and $B_R^{(s,M)}$

**for** $m = 1$ to $M$ **do**

- Transmitter transmits using $f_m^{(s)}$
- Receiver measures using $w_m^{(s)}$

end

// After $M$ measurements:

$$y^{(s,M)} = \sqrt{\rho_s}h^{(s,M)} + n^{(s,M)}$$

$$d^{(s)} = \arg\max_d [p(\bar{v}_d = 1|y^{(s,M)}, G^{(s,M)})]$$

$$\hat{k}_t^{(s)} = \left\lceil \frac{d^{(s)}}{K} \right\rceil, \quad \hat{k}_r^{(s)} = d^{(s)} - K(\hat{k}_t^{(s)} - 1)$$

// Refine sub-ranges based on $\hat{k}_t^{(s)}$ and $\hat{k}_r^{(s)}$ $S_{t,k}^{(s+1)}, S_{r,k}^{(s+1)}, \forall k = 1, \ldots, K.$

end

**Output:**

$$\hat{\phi} = \frac{\pi}{N} \sum_{s=1}^N (\hat{k}_t^{(s)} - 1)K^{-s}, \quad \hat{\alpha} = \frac{\pi}{N} \sum_{s=1}^N (\hat{k}_r^{(s)} - 1)K^{-s}$$

**E. Multi-stage Generalization**

In general the proposed channel estimation algorithm works in a similar multi-stage manner as that in [5], requiring $S = [\log_K N]$ stages. In the $t$th stage, we initially divide the possible AOA angular space into $K$ non-overlapped subranges $S_{t,1}, S_{t,2}, \ldots, S_{t,K}$ and divide the possible AOD angular space into $S^{(t)}_{t,1}, S^{(t)}_{t,2}, \ldots, S^{(t)}_{t,K}$. Then only $M$ overlapped beam pattern pairs will be designed at the transmitter and receiver to cover these $K$ sub-ranges. The designed $M$ beam patterns are characterized by the $M \times K$ beam pattern design matrices $B_T^{(s,M)}$ and $B_R^{(s,M)}$. These should be generated to maximize the minimum Euclidean distance between the columns of the corresponding generator matrix $G^{(s,M)}$.

Given $B_T^{(s,M)}$ and $B_R^{(s,M)}$, we can then generate both the transmit beamforming vectors $\{f_m^{(s)}\}$ and receive beamforming vectors $\{w_m^{(s)}\}$ respectively in the same way as in (18)-(19). For example, to generate $f_m^{(s)}$, the corresponding vector $\pi_T^{(m)}$ in (16), which is redefined as $z_T^{(s,m)}$ for rigorousness, should be designed such that its $i$th entry, denoted by $z_T^{(s,m)}$,$\forall i = 1, \ldots, N$, satisfies

$$z_T^{(s,m)} = \begin{cases} C_m \pi_T^{(m,k_i)}, & \text{if } \frac{m}{N} \in S^{(s)}_{t,k_i}, \exists k_i \in [1, \ldots, K]; \\ 0, & \text{if } \frac{m}{N} \notin S^{(s)}_{t,k_i}, \forall k_i \in [1, \ldots, K] \end{cases} \quad \text{(40)}$$

where $C_m$ is a scalar constant for the $s$th stage to guarantee that $f_m^{(s)}$ satisfies $||f_m^{(s)}||_2 = 1$. Physically, $z_T^{(s,m)}$ describes
the desired beam pattern amplitude at angle \( \frac{\pi}{N} \) when \( f_{(s)}^{(s)} \) is used. Each receive beamforming vector \( u_{(s)}^{(s)} \) can be designed in the same way.

The channel output on the \( s \)th estimation stage can then be obtained after \( M \) time slots by

\[
y^{(s,M)} = \sqrt{P_s} x h^{(s,M)}_{w} + n^{(s,M)}
\]

where

\[
h^{(s,M)}_{w} = \begin{bmatrix}
(w_1^{(s)})^H H f^{(s)}_1 \\
(w_2^{(s)})^H H f^{(s)}_2 \\
\vdots \\
(w_M^{(s)})^H H f^{(s)}_M
\end{bmatrix}
\]

and \( P_s \) denotes the transmit power of the pilot signal in the \( s \)th stage. Similar to that in [5], we prefer that all the stages have an equal probability of failure, indicating that we should allocate power among stages inversely proportional to the beamforming gains of these beam patterns, i.e.,

\[
P_s = \frac{P_T}{C_s}, \quad \forall s = 1, 2, \cdots, S
\]

where \( P_T \) is a constant. Similar to (38) we then find the most likely sub-range combination of the \( s \)th stage by

\[
\hat{d}_{(s)} = \arg\max_d [p(\hat{d}_d = 1|y^{(s,M)}, G^{(s,M)})]
\]

with the corresponding most likely transmitter and receiver sub-ranges given by

\[
\hat{k}_{(s)} = \left\lfloor \frac{\hat{d}_{(s)}}{K} \right\rfloor, \quad \hat{r}_{(s)} = \hat{d}_{(s)} - K(\hat{k}_{(s)} - 1).
\]

The selected sub-ranges, \( S_{t,k_{(s)}} \) and \( S_{r,k_{(s)}} \) are then used for the channel estimation in the next stage. This process continues until the minimum angle resolution \( \frac{\pi}{N} \) is reached requiring \( S = \lceil \log_{K} (N) \rceil \) stages. It is worth pointing out that although the proposed algorithms are elaborated based on the estimation process of a single path, their implementation in multi-path scenario is actually feasible by following the same procedure as in [5, Algorithm 2]. More specifically, multi paths are estimated sequentially, with the first path being estimated using the multi-stage algorithms described above. Subsequent paths can then be found by returning to the first stage and repeating the estimation. Moreover, in each stage’s measurements, the expected contributions from all previously estimated paths can be subtracted to reveal new paths.

**F. Estimation of the Fading Coefficient**

Once all estimation stages described in the previous subsection have been performed, we estimate the identified path fading coefficient \( \alpha \). In [5], the value of \( \alpha \) was estimated based on the measurement of the final stage only. To improve the estimation accuracy, we estimate \( \alpha \) by using all measurements in all stages of the algorithm. Denote by \( r \) and \( \hat{r} \), respectively, the vector of all received measurements and the vector of their estimates such that

\[
r = \begin{bmatrix}
y^{(1,M)} \\
\vdots \\
y^{(S,M)}
\end{bmatrix}, \quad \hat{r} = \begin{bmatrix}
x \sqrt{P_T} NC_{d(t)} G_{d(t)}^{(1,M)} \\
\vdots \\
x \sqrt{P_T} NC_{d(s)} G_{d(s)}^{(S,M)}
\end{bmatrix}
\]

where \( G_{(s,M)}^{(s,M)} \) denotes the \( s \)th column of \( G^{(s,M)} \). Provided that the AOD/DOA estimation is correct in each stage, we can write

\[
r = r \hat{\alpha} + w
\]

where \( w \) is the \( SM \times 1 \) vector of corresponding noise terms. In the case where the AOD/DOA estimation is incorrect, the estimation of the fading coefficient is not important as there will be a beam misalignment between the transmitter and receiver. Following the LMMSE principle [29], we can then estimate the fading coefficient \( \alpha \) as

\[
\hat{\alpha} = P_R \hat{H}^H (\hat{r} P_T \hat{H} + N_0 I_{|\hat{r}|})^{-1} r
\]

where \( I_{|\hat{r}|} \) is an \( |\hat{r}| \times |\hat{r}| \) identity matrix. Now we can formally describe the proposed fast channel estimation algorithm using overlapped beam patterns in Algorithm 1.

**Remark 1.** It can be seen that, compared with the channel estimation algorithm in [5] with the same value of \( K \), our proposed algorithm also requires \( S = \lceil \log_{K} (N) \rceil \) stages, but the number of measurement time slots required in each stage reduces to \( M \), instead of \( K^2 \). In general, this yields a \( \frac{K^2}{M} \) reduction in measurement time slots. For the example of \( K = 3 \) discussed earlier with \( M = 4 \), a \( \frac{K^2}{M} = 225\% \) increase in estimation rate can be achieved.

**IV. RATE ADAPTIVE CHANNEL ESTIMATION ALGORITHM**

The proposed channel estimation scheme explained in the previous section uses a fixed \( G^{(s,M)} \) where the detector is forced to make a decision after \( M \) measurements, irrespective of what the computed probability \( p(\hat{d}_d = 1|y^{(s,M)}, G^{(s,M)}) \) may be. Leveraging the detection method developed in the previous section, we now propose a novel rate-adaptive channel estimation (RACE) algorithm.

We first introduce a target maximum probability of estimation error (PFE), denoted by \( \Gamma \). The basic principle of the RACE algorithm is that after the \( M \) initial measurements are completed in any given stage, if the most likely sub-range combination probability does not satisfy \( p(\hat{d}_d = 1|y^{(s,M)}, G^{(s,M)}) > (1 - \Gamma) \), then additional measurements will be performed. To this end, the receiver will feedback the current most likely transmit sub-range, \( \hat{k}_{(s)} \), and also the information indicating whether more measurements are required or not.

Note that the non-overlapped algorithm proposed in [5] also feeds back a sub-range number to the transmitter, however we include an additional bit corresponding to whether more measurements in this stage are required or not. This feedback only requires \( \lceil \log_{K} (K) \rceil + 1 \) bits, which will be shown to be negligible at high signal-to-noise ratio (SNR) as the average number of additional measurements required is close to zero.

If the specified probability threshold was not met after the \( M \)th measurement, instead of the transmitter and receiver further dividing sub-ranges corresponding to \( \hat{k}_{(s)} \) and \( \hat{r}_{(s)} \), an additional measurement is taken on their sub-range combination. The associated beamforming vectors are designed to correspond to a newly added row to each of \( B_{d(t)}^{(s,M)} \) and \( B_{d(s)}^{(s,M)} \) yielding \( B_{d(t)}^{(s,M+1)} \) and \( B_{d(s)}^{(s,M+1)} \) respectively. We define \( b_{d(s)}^{(s,M+1)} \) as a \( 1 \times K \) sparse binary row vector with a single 1 in its
Algorithm 2: Rate-adaptive channel estimation (RACE) algorithm for mmWave channels

**Input:** Transmitter and receiver both know, $N$, $K$ and have $B_T^{(s,M)}$, $B_R^{(s,M)}$.

**Initialization:** Set initial sub-ranges, $S_{t,k}^{(1)}, S_{r,k}^{(1)}, \forall k = 1, \ldots, K$.

for $s \leq S$ do

// Calculate:

$\{f^{(s)}_{m}\}$ based on $S_{t,k}^{(s)}$, $\forall k = 1, \ldots, K$ and $B_T^{(s,M)}$.

$\{w^{(s)}_m\}$ based on $S_{r,k}^{(s)}$, $\forall k = 1, \ldots, K$ and $B_R^{(s,M)}$.

for $m = 1 \rightarrow M$ do

Transmitter transmits using $f^{(s)}_m$.

Receiver measures using $w^{(s)}_m$.

end

// After $M$ measurements:

$y^{(s,M)} = \sqrt{P} s h_y^{(s,M)} + n^{(s,M)}$.

$\ddot{d}^{(s)} = \text{argmax} [p(\ddot{v}_d = 1 | y^{(s,M)}, G^{(s,M)})]$.

$\ddot{k}_t^{(s)} = \left[ \frac{\ddot{d}^{(s)}}{K} \right]$, $\ddot{k}_r^{(s)} = \ddot{d}^{(s)} - K(\ddot{k}_t^{(s)} - 1)$

// Carry out additional measurements if required:

$R \geq 0$ while $p(\ddot{v}_d,0) = 1 | y^{(s,M+R)}, G^{(s,M+R)} < (1 - \Gamma)$ and $(M + R < M_{\text{max}})$ do

$R = R + 1$ // Increment re-measurement index

Transmitter transmits with:

$f^{(s)}_{M+R}$ based on $S_{t,k}^{(s)}$, $\forall k$ and $B_T^{(s,M)}$.

Receiver measures with:

$w^{(s)}_{M+R}$ based on $S_{r,k}^{(s)}$, $\forall k$ and $B_R^{(s,M)}$.

Update:

$\ddot{d}^{(s)} = \text{argmax} [p(\ddot{v}_d = 1 | y^{(s,M+R)}, G^{(s,M+R)})]$.

$\ddot{k}_t^{(s)} = \left[ \frac{\ddot{d}^{(s)}}{K} \right]$, $\ddot{k}_r^{(s)} = \ddot{d}^{(s)} - K(\ddot{k}_t^{(s)} - 1)$

end

// Refine sub-ranges based on $\ddot{k}_t^{(s)}$ and $\ddot{k}_r^{(s)}$:

$S_{t,k}^{(s+1)}$, $S_{r,k}^{(s+1)}, \forall k = 1, \ldots, K$.

end

**Output:**

$\ddot{d} = \frac{1}{N} \sum_{s=1}^{S} (\ddot{k}_t^{(s)} - 1)K^{S-s}$, $\ddot{d} = \frac{1}{N} \sum_{s=1}^{S} (\ddot{k}_r^{(s)} - 1)K^{S-s}$.

$\hat{\alpha} = P_B^H(\ddot{d}F_B^H + N_0 I_{|\ddot{d}|})^{-1} \ddot{r}$.

The $(M + 1)$th transmit and receive beamforming vectors can then be calculated from the new row in these matrices and be used to measure the channel, obtaining $y^{(s,M+1)}$. The updated generator matrix corresponding to (49) can be described by

$$G^{(s,M+1)} = B_T^{(s,M+1)} \odot B_R^{(s,M+1)} = \left[ \begin{array}{c} G^{(s,M)} \\ \bar{y}_K^{(d)} \end{array} \right].$$ (50)

The estimation parameters can then be updated based on the updated $G^{(s,M+1)}$ and $y^{(s,M+1)}$ and the ML detector developed in the previous section. We can generalize this to $G^{(s,M+R)}$ where $R = 0, 1, \ldots$ indexes the additional measurements.

The proposed RACE algorithm is formally described in Algorithm 2. As can be seen from the algorithm description, this process then repeats until either the threshold condition has been met (i.e., $p(\ddot{v}_d = 1 | y^{(s,M)}, G^{(s,M)}) > (1 - \Gamma)$) or a maximum number of measurements, denoted by $M_{\text{max}}$, has been reached (i.e., $M + R = M_{\text{max}}$). Under fading channel conditions, there always exists a non-zero probability of an ‘outage’ occurring when the path coefficient is close to zero. By imposing the upper limit to the number of measurements we reduce the time and energy expended in this case. As the RACE algorithm is able to compute the probability of successful estimation during the estimation process, it can minimize the number of measurements required for successful estimation and therefore reduce the associated energy.

**V. PERFORMANCE ANALYSIS**

In this section, we focus on the performance analysis of the proposed algorithms when a single path is present between the transmitter and receiver. Specifically, we derive two analytical expressions. The first is an expression for the PEE (i.e., an incorrect estimation of the AOA or AOD), whereas the second expression is for the minimum energy-to-noise ratio required by the RACE algorithm for a specified average number of measurements, $M + R$.

A. Probability of AOD/AOA Estimation Error

As the RACE algorithm seeks to achieve a fixed error rate by means of an adaptive $G^{(s,M+R)}$, we first need to find the PEE for a fixed generator matrix, $G^{(s,M+R)}$, $\forall s = 1, \ldots, S$. Note that this is also the case in Algorithm 1 with a fixed generator matrix, $G^{(s,M)}$, $\forall s = 1, \ldots, S$. We begin by defining the PEE in a given stage $s$, assuming all previous stages have been correct, as

$$p_{\text{EE}}(G^{(s,M+R)}, s) = p(\ddot{v}^{(s)} \neq \hat{v}^{(s)}) = 1 - p(\ddot{v}^{(s)} = \hat{v}^{(s)})$$ (51)

where the term $p(\ddot{v}^{(s)} \neq \hat{v}^{(s)})$ represents the probability that the estimated AOD/AOA information $\ddot{v}^{(s)}$ is not equal to the physical channel, $\hat{v}^{(s)}$.

If we treat the channel vector $\ddot{v}^{(s)}$ as information being encoded by the generator matrix $G^{(s,M+R)}$, we can model this estimation error event as a maximum likelihood decoding error. We can then say that $\ddot{v}^{(s)} \neq \hat{v}^{(s)}$ occurs when the received measurement vector $y^{(s)} = \alpha x \sqrt{P_s N_c} C G^{(s,M+R)} (\ddot{v}^{(s)})^T + n^{(s)}$, is incorrectly 'decoded' to another received measurement vector $y^{(s)} = \alpha x \sqrt{P_s N_c} C G^{(s,M+R)} (\hat{v}^{(s)})^T + n^{(s)}$, where $\alpha x \sqrt{P_s N_c}$ is the received energy, $C$ is the channel matrix, $G^{(s,M+R)}$ is the generator matrix, $\ddot{v}^{(s)}$ is the estimated AOD/AOA, $\hat{v}^{(s)}$ is the physical AOD/AOA, $n^{(s)}$ is the noise vector, and $\alpha$ is the received signal-to-noise ratio. This is equivalent to finding the bit error rate of the maximum likelihood decoder.
\[ p(\hat{v}(s) \neq \tilde{v}(s)) = \sum_{\hat{v}' \in V} \sum_{\tilde{v}' \in V, \tilde{v}' \neq \tilde{v}} p(\hat{v}') p(\tilde{v}') \left(1 - \frac{\gamma^2}{\gamma^2 + 2}\right) \]  

where  
\[ \gamma = \sqrt{\frac{P_c N_0^2}{2N_0}} \left\| G^{(s,M+R)}(\tilde{v}(s) - \tilde{v}')^T \right\| \]  

By substituting these terms into (51) and averaging over each \( \tilde{v}(s) \) we get  
\[ P_{EE}(s,M+R) = \sum_{\hat{v} \in V} p(\hat{v}) p_{EE}(s,M+R, \hat{v}(s)) \]  

\[ = \sum_{\hat{v} \in V} \sum_{\tilde{v}' \in V, \tilde{v}' \neq \tilde{v}} p(\hat{v}) \left(1 - \frac{\gamma^2}{\gamma^2 + 2}\right) \]  

Expressing the exact probability of the union term in (56) is quite difficult due to complicated boundaries between each column in \( G^{(s,M+R)} \). This is a direct result of the generator matrix not containing a full set of normalized \( (M + R) \times 1 \) vectors. For example, (11) does not contain the scaled versions of the vectors \([0,1,1,0]^T, [1,1,0,1]^T, [1,0,1,1]^T, [1,1,1,0]^T, [0,1,1,1]^T, [0,0,0,0]^T\). This is an inherent property of the generator matrix being constructed from two smaller matrices \( B_T^{(s,M+1)} \) and \( B_B^{(s,M+1)} \) that do not contain the all-zero column vector. We can, however, find an upper bound of expression in (56) by replacing the union term with a summation to get  
\[ P_{EE}(s,M+R) < \sum_{\hat{v} \in V} \sum_{\tilde{v}' \in V, \tilde{v}' \neq \tilde{v}} p(\hat{v}) \left(1 - \frac{\gamma^2}{\gamma^2 + 2}\right) \]  

The resulting upper bound in (57) is found to be quite loose in most cases. Motivated by this, we find a much tighter approximation by only considering \( \tilde{v}' \) that minimizes \( \left\| G^{(s,M+R)}(\tilde{v}(s) - \tilde{v}')^T \right\| \) (i.e., only considering spatially adjacent columns of \( G^{(s,M+R)} \)). We define this minimum distance for each \( \tilde{v}(s) \) by  
\[ d_{min}(\tilde{v}(s)) = \min_{\tilde{v}' \in V, \tilde{v}' \neq \tilde{v}} \left\| G^{(s,M+R)}(\tilde{v}(s) - \tilde{v}')^T \right\| \]  

and the set of vectors causing it as  
\[ V_{min}(\tilde{v}(s)) = \left\{ \arg \min_{\tilde{v}' \neq \tilde{v}} \left\| G^{(s,M+R)}(\tilde{v}(s) - \tilde{v}')^T \right\| \right\} \]  

We therefore approximate \( P_{EE}(s,M+R) \) by  
\[ P_{EE}(s,M+R) \approx \sum_{\hat{v} \in V} \sum_{\tilde{v}' \in V_{min}, \tilde{v}' \neq \tilde{v}} p(\hat{v}) \left(1 - \frac{\gamma^2}{\gamma^2 + 2}\right) \]  

We also provide a greatly simplified lower bound for Algorithm 1 when \( R = 0 \), by only considering a single one
of these $\vartheta'$ that cause a minimum distance. Recall that the generator matrix is designed such that all paths have the same minimum Euclidean distance, which we define here as $\sqrt{\min} = d_{\min}(\vartheta^{(s)})$, $\forall \vartheta^{(s)}$. By assuming each channel realization is equiprobable, this leads to

$$p_{EE}(\vartheta^{(s)}, R) > \frac{1}{2} \left[ 1 - \frac{P(4N^2PR_{\min})}{\sqrt{P(4N^2PR_{\min}) + 4N_0}} \right]$$

(61)

Finally, we upper bound the probability of an error occurring in any of the $S$ stages as

$$p_{EE} = 1 - \sum_{s=1}^{S} (1 - p_{EE}(\vartheta^{(s)}, R))$$

(62)

$$< \sum_{s=1}^{S} \sum_{t=1}^{R} \frac{p(t)}{2} \left[ 1 - \sqrt{\frac{1}{2} \frac{\gamma_{s,M}^2}{\gamma_{s,M}^2 + 2}} \right]$$

(63)

(64)

In order to validate our analysis, Figs. 3 (a) and (b) compare the derived analytical expressions with their corresponding Monte Carlo simulations for a single stage system and three stage system, respectively. As we can observe from Fig. 3(a), the given in (60) is actually quite tight. Focusing on the three stage system in Fig. 3(b), we see that the substitution of the approximation in (60) into the upper bound (63) causes some disparity, however still provides a tighter approximation than (64). A lower bound is not presented in Fig. 3(b) as the use of (61) in the upper bound (63) would not be mathematically correct.

B. Minimum Energy and Time Requirement for the RACE Algorithm

We now present an expression for the minimum signal energy-to-noise ratio required by the RACE algorithm for an average number of measurements in a given stage. From an information theory standpoint, we can describe the RACE algorithm as the information process that a $1 \times K^2$ binary vector $\vartheta^{(s)}$ is encoded into $M + R$ symbols by the generator matrix $G^{(s,M+R)}$. This information transfer has an equivalent approximation in (60) into the upper bound (63) causes some disparity, however still provides a tighter approximation than (64). A lower bound is not presented in Fig. 3(b) as the use of (61) in the upper bound (63) would not be mathematically correct.

By denoting the total energy used in stage $s$ by $E_s = P_x(M + R)$, we can re-write (67) as

$$\frac{E_s}{N_0} \geq \frac{(M + R)^2(2\frac{K^2}{M^2} - 1)}{C_s^4N^2\|G^{(s,M+R)}(\vartheta^{(s)})^T\|_2^2}$$

(68)

The expression in (68) gives a good insight into how to minimize the amount of energy required for successful channel estimation. For example, one intuitive approach to reduce the energy requirement is to maximize the term, $C_s^4N^2\|G^{(s,M+R)}(\vartheta^{(s)})^T\|_2^2$ which directly determines the amount of energy that arrives at the receiver. Furthermore, as we have a minimum number of $N$ initial measurements that have been designed such that they propagate non-zero energy over each sub-range combination (i.e., each column of $G^{(s,M+R)}$ has a non-zero norm), we ensure that $\|G^{(s,M+R)}(\vartheta^{(s)})^T\|_2^2 > 0$, $\forall \vartheta^{(s)}$. This then guarantees that $E_s/N_0$ always takes on a finite value.

As $G^{(s,M+R)}$ is adaptive and based on feedback, it becomes quite difficult to determine the exact energy requirement for various $R > 0$. However, if we let $g_i^{(s,M+R)}$ denote the $i$th row of $G^{(s,M+R)}$, we can expand the term $\|G^{(s,M+R)}(\vartheta^{(s)})^T\|_2^2$ by splitting it into the sum of received energy in each measurement time slots. The row/column energy normalization, seen in the design of (10), leads to the generator matrix having $\|G^{(s,M)}(\vartheta^{(s)})^T\|_2^2 = \frac{K^2}{N} \forall \vartheta^{(s)}$. Expanding this expression we then get

$$\|G^{(s,M+R)}(\vartheta^{(s)})^T\|_2^2 = \sum_{i=1}^{M+R} (g_i^{(s,M+R)}(\vartheta^{(s)})^T)^2$$

$$= \|G^{(s,M)}(\vartheta^{(s)})^T\|^2 + \sum_{i=M+1}^{M+R} (g_i^{(s,M+R)}(\vartheta^{(s)})^T)^2$$

$$= \frac{M}{K^2} + \sum_{i=M+1}^{M+R} (\vartheta^{(s)})^T)^2.$$  

(69)

Following the RACE algorithm, we can also deduce that the term $g_i^{(s,M+R)}(\vartheta^{(s)})^T$ in (69) results in either unit energy or no energy, depending on whether the most likely sub-range combination from the previous measurement timeslot was correct or not. If we denote the probability of this information being incorrect as $p_{EE}(G^{(s,M+R)})$, we can then write the average energy contribution from this term as

$$\|G^{(s,M+R)}(\vartheta^{(s)})^T\|_2^2 = \frac{M}{K^2} + \sum_{R=1}^{R} (1 - p_{EE}(G^{(s,M+R)})),$$  

(70)

Substituting (70) back into (68), we have

$$\frac{E_s}{N_0} \geq \frac{(M + R)^2(2\frac{K^2}{M^2} - 1)}{C_s^4N^2\|G^{(s,M+R)}(\vartheta^{(s)})^T\|_2^2}.$$  

(71)

Unfortunately, the probability in (71) still depends on previous sub-range estimates determined from the previous measurement and is thus difficult to obtain a closed-form. Therefore, we can...
however, consider a few extreme cases where we assume a fixed probability of correct feedback information. That is, we use $p_{FB} = 1 - p_{EE}G_{s,M,R-1}$, $\forall R > 0$. Finally, we can express the average minimum energy to noise ratio required for successful channel estimation in a given stage with a certain number of measurements to be

$$E_s \geq \frac{(M + R)^2(2\pi^2 - 1)}{C^2N^2||\alpha||^2\left[\frac{M}{K^2} + R \times p_{FB}\right]}.$$ 

(72)

To validate this expression, Fig. 4 shows the average number of measurements required to estimate the AOD/AOA information in a single stage for a range of different signal energy to noise ratios. Here we also plot (72) for two special cases. The first is when $p_{FB} = 1$, representing the best case scenario where the previous sub-range estimate is always correct. The second is the worst case scenario where the previous sub-range estimate is random (i.e., no feedback) giving $p_{FB} = 1/K^2$.

As we can see from Fig. 4, (72) provides a bound for the best case scenario $p_{FB} = 1$, and for the worst case scenario $p_{FB} = 1/K^2$. To make the simulation results match the bounds, here we do not limit the number of measurements (i.e., $M_{max} = \infty$). Although $p_{FB} = 1/K^2$ still gives a mathematical lower bound for the energy requirement, the condition that $p_{FB} = 1/K^2$ describes a system that performs far worse than the RACE algorithm. As such, we see the two lines intercept at high SNR. We can also observe from this figure that the RACE algorithm falls between these two curves and converges to $R = 0$ at high SNR. Finally, the asymptotic nature of the results in Fig. 4 supports the need to set $M_{max} << \infty$ so that measurements do not continue indefinitely.

C. Discussion of System Performance Parameters

We end this section by briefly discussing the performance parameter selection for the proposed algorithms for a generalized mmWave MIMO system with $N_t$ transmit antenna and $N_r$ receive antenna, each equipped with a limited number of RF chains such that they can only transmit and receive with a single beam pattern.

1) Selection of number of sub-ranges in each stage, $K$:

In general, by using a smaller value of $K$, a faster channel estimation can be achieved at a greater energy efficiency and a more computationally efficient ML detection. The main drawback of setting a smaller $K$ lies in the wider beam patterns and the resultant loss of directionality gain. From an energy efficiency point of view, this loss is outweighed by the speed advantage. However, if a peak power constraint were imposed, more directionality may be required in the early stages and this can be achieved by increasing $K$. Selection of $K$ may also depend on the number of antennas being equipped at each link end. For example, it is generally accepted that a mmWave mobile link will have more antennas at a base station (BS) than at the mobile station (MS) [3], [32]. For such a general asymmetric system, it is possible to use different $K$ in each stage and at each transceiver end. This grants flexibility to achieve the same number of estimation stages for each system. Then it is only important that the product of the $K$’s over all stages equals to the antenna number. An alternative approach is to simply use more stages at the BS than at the MS. In this case, once the MS has reached its final stage before the transmitter, it can continue its role in the estimation procedure by utilizing the already estimated AOA.

2) Selection of $M$ and resultant design of $G_{s,M}$:

For a given number of sub-ranges $K$, $M$ pairs of initial beam patterns can then be designed to span them. The use of a lower $M$ value leads to a faster estimation of the stage, however, generally requires a higher SNR. As mentioned in subsection III-A, there exist many combinations of $B_{T}^{(s,M)}$ and $B_{R}^{(s,M)}$ that can be used for a given $M$ and $K$. Effective design should consider the resulting generator matrix $G_{s,M}$, and ensure that all columns have the same minimum Euclidean distance. In addition to this, each column of $G_{s,M}$ should have the same energy. The equal-power criterion for each row/column is not as important as the Euclidean distance property, however, may simplify any power requirements by proving uniform transmit/receive power within a stage.

3) Selection of $M_{max}$ and $\Gamma$:

Finally, the selection of the maximum number of measurements per stage $M_{max}$ should be selected based upon any maximum timing constraints. Increasing $M_{max}$ will significantly increase the energy efficiency at medium to high SNR regime, although the average number of measurements per stage will still converge to $M$ at high SNR. We find that, for a fading channel with Gaussian distributed channel coefficients, the selection of $\Gamma$ is not as important as $M_{max}$. This is due to the non-zero probability of an outage condition. In general, increasing $\Gamma$ will reduce the PEE by increasing the average number of measurements and therefore expending more energy. In contrast, at high SNR, increasing $M_{max}$ will allow more measurements to be carried out in the unlikely event that it needs them. Therefore the later uses less average energy.

VI. NUMERICAL RESULTS

We now provide some numerical examples to verify the performance of our proposed algorithms. We consider a mmWave
In this paper we have proposed a novel fast channel estimation algorithm for mmWave communication systems based on a novel overlapped beam pattern design. In general, our proposed algorithm can speed up the channel estimation process by a factor of $K^2/M$ when compared to existing algorithms
with non-overlapped beam patterns. Using a fixed number of measurements, we show that this reduction of measurements comes at an energy-to-noise ratio expense of 2.5dB in order to achieve the same probability of estimation error (PEE). For channels with rapidly changing channel information, this cost may be justified in order to improve the estimation speed. We have also proposed a novel rate-adaptive channel estimation algorithm, in which additional measurements are carried out when a high probability of estimation error is expected. We show that by taking this approach, the channel can be estimated more efficiently, yielding significant gains of up to 6dB when compared to the algorithm in [5], while still converging to the same average number of measurements as the fast channel estimation (FCE) algorithm at high SNR.

REFERENCES

[1] Z. Pi and F. Khan, “An introduction to millimeter-wave mobile broadband systems,” IEEE Commun. Mag., vol. 49, no. 6, pp. 101–107, June 2011.

[2] T. S. Rappaport, S. Sun, R. Mayzus, H. Zhao, Y. Azar, K. Wang, G. N. Wong, J. K. Schulz, M. Samimi, and F. Gutierrez, “Millimeter wave mobile communications for 5G cellular: It will work!” IEEE Access, vol. 1, pp. 335–349, May 2013.

[3] W. Hong, K.-H. Baek, Y. Lee, Y. Kim, and S.-T. Ko, “Study and prototyping of practically large-scale mmwave antenna systems for 5g cellular devices,” Communications Magazine, IEEE, vol. 52, no. 9, pp. 63–69, 2014.

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

[5] A. Alkhateeb, O. B. Aysal, 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.

[6] H. Zhang, S. Venkateswaran, and U. Madhow, “Channel modeling and MIMO capacity for outdoor millimeter wave links,” in Proc. Wireless Commun. Netw. Conf. (WCNC), Apr. 2010, pp. 1–6.

[7] E. Torkildson, H. Zhang, and U. Madhow, “Channel modeling for millimeter wave mmimo,” in Proc. Inf. Theory Appl. Workshop. IEEE, 2010, pp. 1–8.

[8] S. Hur, T. Kim, D. Love, J. Krogmeier, T. 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.

[9] B. Biglarbegian, M. Fakharzadeh, D. Busuioc, M.-R. Nezhad-Ahmadi, and S. Safavi-Naeini, “Optimized microstrip antenna arrays for emerging millimeter-wave wireless applications,” IEEE Trans. Antennas Propag., vol. 59, no. 5, pp. 1742–1747, 2011.

[10] “Wireless medium access control (MAC) and physical layer (PHY) specifications for high rate wireless personal area networks (WPANs),” amendment 2: Millimeter-wave-based alternative physical layer extension, “IEEE Std 802.15.3c,” pp. cl–187, Oct 2009.

[11] J. Wang, Z. Lan, C.-W. Pyo, T. Baykas, C.-S. Sum, M. Azzur Rahman, R. Funada, F. Kojima, I. Lakki, H. Harada, and S. Kato, “Beam codebook based beamforming protocol for multi-gbps millimeter-wave WPAN systems,” in IEEE J. Select. Areas Commun., 2009, pp. 1390–1399.

[12] L. Chen, Y. Yang, X. Chen, and W. Wang, “Multi-stage beamforming codebook for 60GHz WPAN,” in Proc. 6th Int. ICST Conf. Commun. Network. China, Aug. 2011, pp. 361–365.

[13] A. Sayeed, “Deconstructing multiantenna fading channels,” IEEE Trans. Sig. Process., vol. 50, no. 10, pp. 2563–2579, Oct. 2002.

[14] Z. Hong, K. Liu, R. W. Heath Jr, and A. M. Sayeed, “Spatial multiplexing in correlated fading via the virtual channel representation,” IEEE J. Sel. Areas Commun., vol. 21, no. 3, pp. 485–496, 2003.

[15] J. C. Lagarias, J. A. Reeds, M. H. Wright, and P. E. Wright, “Convergence properties of the nelder–mead simplex method in low dimensions,” SIAM J. Optim., vol. 9, no. 1, pp. 112–147, 1998.

[16] J. Mo, P. Schniter, N. Gonzalez Prelic, and R. Heath, “Channel estimation in millimeter wave mmimo systems with one-bit quantization,” in Proc. Conf. Signals, Syst. Comput., Nov 2014, pp. 957–961.

[17] J. Seo, Y. Sung, G. Lee, and D. Kim, “Pilot beam sequence design for channel estimation in millimeter-wave MIMO systems: A POMDP framework,” in Proc. IEEE Signal Process. Adv. Wireless Commun., June 2015, pp. 236–240.

[18] G. MacCartney and T. Rappaport, “73 GHz millimeter wave propagation measurements for outdoor urban mobile and backhaul communications in New York City,” in IEEE Int. Conf. on Commun. (ICC), June 2014, pp. 4862–4867.

[19] M. Akdeniz, Y. Liu, M. Samimi, S. Sun, S. Rangan, T. Rappaport, and E. Erkip, “Millimeter wave channel modeling and cellular capacity evaluation,” IEEE J. Select. Areas Commun., vol. 32, no. 6, pp. 1164–1179, June 2014.

[20] M. Kokshoorin, P. Wang, Y. Li, and B. Vucetic, “Fast channel estimation for millimetre wave wireless systems using overlapped beam patterns,” in IEEE Int. Conf. on Commun. (ICC), June 2015, pp. 1304–1309.

[21] W. Baywa, J. Haupt, A. Sayeed, and R. Nowak, “Compressed channel sensing: A new approach to estimating sparse multipath channels,” Proc. IEEE, vol. 98, no. 6, pp. 1058–1076, June 2010.

[22] “Wireless medium access control (mac) and physical layer (phy) specifications for high rate wireless personal area networks (WPANs),” amendment 2: Millimeter-wave-based alternative physical layer extension, IEEE standard 802.15.3c, Oct. 2009.

[23] Y. Tsang, A. Poon, and S. Addepalli, “Coding the beams: Improving beamforming training in mmwave communication system,” in Proc. IEEE Global Telecomm. Conf., Dec 2011, pp. 1–6.

[24] X. Zhang, A. Molisch, and S.-Y. Kung, “Variable-phase-shift-shifted RF-baseband codec for mimo antenna selection,” IEEE Trans. Signal Process., vol. 53, no. 11, pp. 4091–4103, Nov 2005.

[25] A. Sayeed and V. Raghavan, “The ideal MIMO channel: Maximizing capacity in sparse multipath with reconfigurable arrays,” in Proc. ISIT, July 2006, pp. 1036–1040.

[26] J. G. Proakis, Digital communications. New York, NY, USA: McGraw-Hill, 1995.

[27] R. G. Gallager, “Circularly-symmetric gaussian random vectors,” preprint, pp. 1–9, 2008.

[28] R. D. Bock and M. Atikin, “Marginal maximum likelihood estimation of item parameters: Application of an em algorithm,” Psychometrika, vol. 46, no. 4, pp. 443–459, 1981.

[29] L. L. Scharf, Statistical signal processing. Addison-Wesley Reading, MA, 1991, vol. 98.

[30] G. L. Stüber, Principles of mobile communication. Springer Science & Business Media, 2011.

[31] C. E. Shannon, “A mathematical theory of communication,” Bell Syst. Tech. J., vol. 27, pp. 379–423, 1948.

[32] T. Bai, A. Alkhateeb, and R. Heath, “Coverage and capacity of millimeter-wave cellular networks,” IEEE Commun. Mag., vol. 52, no. 9, pp. 70–77, 2014.