# Fast Channel Estimation for Millimetre Wave Wireless Systems Using Overlapped Beam Patterns

Matthew Kokshoorn, Peng Wang, Yonghui Li, and Branka Vucetic

School of Electrical and Information Engineering, The University of Sydney, Australia Email: {matthew.kokshoorn, peng.wang, yonghui.li, branka.vucetic}@sydney.edu.au

*Abstract*—This paper is concerned with the channel estimation problem in millimetre wave (MMW) wireless systems with large antenna arrays. By exploiting the sparse nature of the MMW channel, we present an efficient estimation algorithm based on a novel overlapped beam pattern design. The performance of the algorithm is analyzed and an upper bound on the probability of channel estimation failure is derived. Results show that the algorithm can significantly reduce the number of required measurements in channel estimation (e.g., by 225% when a single overlap is used) when compared to the existing channel estimation algorithm based on non-overlapped beam patterns.

## I. INTRODUCTION

Millimetre wave (MMW) communication has been shown to be a promising technique for next generation wireless systems due to the large expanse of available spectrum in the MMW frequency band, ranging from 30GHz to 300GHz. [1][2]. However, a critical challenge in exploiting the MMW frequency band is its severe signal propagation loss compared to that over conventional microwave frequencies [3][4]. To compensate such a loss, large antenna arrays can be employed to achieve a high power gain. Fortunately, owing to the small wavelength of MMW signals, antennas can be packed into a small area at the transmitter and receiver [5].

Channel state information (CSI) is essential for effective communication and precoder design in MMW systems. The use of large antenna arrays results in a large multiple-input multiple-output (MIMO) channel matrix. This makes the channel estimation a very challenging issue due to the large number of channel parameters to be estimated. On the other side, it was shown in recent channel measurements [6] that the MMW channel exhibits strong sparse propagation due to the high path loss. As a result, the MMW MIMO channel matrix can be represented as a sparse one after being converted into the angular space [7]. Leveraging sparse geometric channel characteristics, [3] developed a "divide and conquer" type multi-stage algorithm for estimating sparse MMW 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 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) angular 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 Kbeam patterns at the transmitter, and then received by one



Fig. 1. Illustration of the beam patterns adopted in the first (a) and second (b) stages of the channel estimation algorithm of [3] when K = 3. The three sub-ranges in the first stage are,  $[0, \pi/3)$ ,  $[\pi/3, 2\pi/3)$  and  $[2\pi/3, \pi)$ , respectively. By assuming that the possible AOAs/AODs are reduced to the sub-range  $[0, \pi/3)$  in the first stage, this sub-range is further divided into  $[0, \pi/9)$ ,  $[\pi/9, 2\pi/9)$  and  $[2\pi/9, \pi/3)$ , respectively, in the second stage.

of the K beam patterns at the receiver, the corresponding channel output for this transmit and receive beam pattern 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. 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 subranges for the channel estimation in next stage. This process continues until the smallest beam width resolution is reached. It is shown in [3] that the algorithm requires estimation time proportional to  $K^2 \lceil \log_K(\max(N_t, N_r)) \rceil$  per path where  $N_t$  and  $N_r$  are, respectively, the numbers of transmit and receive antennas. In rapidly varying channels, such a channel estimation algorithm may not be quick enough to track the fast channel variations. Therefore it is desirable to develop a fast and efficient algorithm to further reduce the estimation time.

In this paper, we propose a fast channel estimation algorithm by designing a set of novel overlapped beam patterns that can significantly reduce the number of measurement time slots required for channel estimation. The overlap between the beam patterns increases the amount of information carried by the channel outputs, and thus a path can then be identified by using a combination of multiple channel outputs, instead of using just a single one. In this way, the proposed algorithm can reduce the number of time slots required for channel estimation and thus speed up the estimation process. We also develop a minimum mean squared error (MMSE) channel estimator to estimate the channel coefficient by optimally combining the selected measurements in all stages. The performance of the proposed algorithm is analyzed and an upper bound on the probability of channel estimation failure is derived. The analysis is then validated by Monte Carlo simulations.

## II. SYSTEM MODEL

Consider a MMW MIMO system where both the transmitter and receiver are equipped with a half-wavelength spacing uniform linear antenna array (ULA). For simplicity, in this paper we assume the same number of antennas, denoted by N, at both the transmitter and receiver. The proposed algorithm can be easily extended to a general asymetric system. Assume that there is a single-path channel between the transmitter and receiver with AOD,  $\phi \in [0, \pi)$ , and AOA,  $\theta \in [0, \pi)$ . Then the corresponding channel matrix can be represented as<sup>1</sup>

$$\boldsymbol{H} = \alpha \boldsymbol{a}_r(\theta) \boldsymbol{a}_t^H(\phi) \tag{1}$$

where  $\alpha$  is the channel fading coefficient,  $(\cdot)^H$  denotes the conjugate transpose operation and  $a_r(\theta)$  and  $a_t(\phi)$  are, respectively, the receive and transmit spatial signatures of the single path. They are defined as  $a_r(\theta) = u(\theta)$  and  $a_t(\phi) = u(\phi)$ , respectively, where

$$\boldsymbol{u}(\epsilon) \triangleq \frac{1}{\sqrt{N}} [1, e^{j\pi \sin(\epsilon)}, \cdots, e^{j\pi (N-1)\sin(\epsilon)}]^T \qquad (2)$$

and  $(\cdot)^T$  is the transpose operation. Based on (1), the overall channel state information only includes three parameters, i.e., the AOA  $\theta$ , the AOD  $\phi$ , and the fading coefficient  $\alpha$ . We assume that both  $\theta$  and  $\phi$  can only take some discrete values from the set  $\{0, \frac{\pi}{N}, \cdots, \frac{\pi(N-1)}{N}\}$ . We aim to find an efficient way to estimate these three parameters.

We assume that both the transmitter and receiver are equipped with a limited number of RF chains. Following [3], we further assume that these RF chains, at one end, can only be jointly adopted to form a single beam pattern, indicating that only one pilot signal can be transmitted and received at one time. Define f and w ( $||f||_2 = ||w||_2 = 1$ ), respectively, as the  $N \times 1$  unit beamforming vector at the transmitter and  $N \times 1$  unit combining vector at the receiver. The corresponding channel output can be represented as

$$y = \sqrt{p} \boldsymbol{w}^H \boldsymbol{H} \boldsymbol{f} \boldsymbol{x} + \boldsymbol{w}^H \boldsymbol{n}$$
(3)

where x is the transmitted pilot signal with unit power, p is the transmit power and n is a length-N vector of independent and identically distributed (i.i.d.) complex additive white Gaussian noise (AWGN) samples with mean zero and variance  $N_0$ . The key challenge here is how to design a sequence of f and



Fig. 2. Illustration of the overlapped beam patterns adopted in the first (a) and second (b) stages of the proposed algorithm when K = 3.

w in such a way that the channel parameters can be quickly estimated without an exhaustive search of all beam angles.

# III. PROPOSED CHANNEL ESTIMATION WITH OVERLAPPED BEAM PATTERNS

In this section, we propose a new set of beam patterns that are overlapped with one another, based on which a fast channel estimation algorithm is designed to accurately retrieve the channel state information.

#### A. An Example

To better understand our proposed overlapped beam pattern design, here we first consider an example. Following Fig.1, we still divide the AOD/AOA angular spaces into K = 3sub-ranges in the first stage, denoted by  $S_1 = [0, \pi/3)$ ,  $S_2 = [\pi/3, 2\pi/3)$  and  $S_3 = [2\pi/3, \pi)$ , respectively. Instead of using 3 beam patterns to cover them at each link end as in Fig.1(a), we aim to use only M = 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  $S_2$ . It is also seen that each beam pattern can have different amplitudes in different subranges. The beam patterns can be described using a matrix, referred to as the beam pattern description matrix, as

$$\boldsymbol{B} = \begin{bmatrix} b_{1,1} & b_{1,2} & b_{1,3} \\ b_{2,1} & b_{2,2} & b_{2,3} \end{bmatrix}$$
(4)

where  $b_{m,k}$  denotes the amplitude of the  $m^{th}$  beam pattern in sub-range  $S_k$ . 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  $3^{rd}$  and  $1^{st}$  sub-ranges. Due to the symmetry between the two beam patterns, we further have  $b_{1,1} = b_{2,3} = \beta_1$  and  $b_{1,2} = b_{2,2} = \beta_2$ . Furthermore, it is always desirable that the same quantity of signal power is transmitted/received via different sub-ranges during one stage. In order to achieve this, we should normalize each column of **B**. This leads to  $\beta_1 = 1$ and  $\beta_2 = \frac{1}{\sqrt{2}}$ , and the matrix **B** reduces to

$$\boldsymbol{B} = \begin{bmatrix} \beta_1 & \beta_2 & 0\\ 0 & \beta_2 & \beta_1 \end{bmatrix} = \begin{bmatrix} 1 & \frac{1}{\sqrt{2}} & 0\\ 0 & \frac{1}{\sqrt{2}} & 1 \end{bmatrix}.$$
 (5)

<sup>&</sup>lt;sup>1</sup>Note that the use of ULA results in no distinguishable difference between AOAs  $\theta$  and  $-\theta$  or between AODs  $\phi$  and  $-\phi$ . Hence, only AODs and AOAs in the range  $[0, \pi)$  need to be considered.

To generate the beam patterns illustrated in Fig.2(a), the beamforming/combining vectors should be designed as follows. Denote by  $f_m$  and  $w_m$ , respectively, the beamforming vector and combining vector corresponding to the  $m^{th}$  beam pattern in Fig.2(a). We then have

$$\boldsymbol{u}^{H}(\boldsymbol{\epsilon})\boldsymbol{f}_{m} = \boldsymbol{u}^{H}(\boldsymbol{\epsilon})\boldsymbol{w}_{m} = Cb_{m,k}, \text{if } \exists \ k \in \{1, 2, 3\}, \boldsymbol{\epsilon} \in \mathcal{S}_{k}, \ (6)$$

where  $u(\epsilon)$  has been defined in (2) and C is a scalar constant that ensures  $||f_m||_2 = ||w_m||_2 = 1$ . Equation (6) can be expressed in a matrix form as

$$\boldsymbol{U}^{H}\boldsymbol{f}_{m} = \boldsymbol{U}^{H}\boldsymbol{w}_{m} = \begin{bmatrix} Cb_{m,1}\boldsymbol{1}_{|S_{1}|} \\ Cb_{m,2}\boldsymbol{1}_{|S_{2}|} \\ Cb_{m,3}\boldsymbol{1}_{|S_{3}|} \end{bmatrix} \triangleq \boldsymbol{g}_{m}$$
(7)

where  $\boldsymbol{U} = \left[\boldsymbol{u}(0), \boldsymbol{u}\left(\frac{\pi}{N}\right), \cdots, \boldsymbol{u}\left(\frac{\pi(N-1)}{N}\right)\right], \mathbf{1}_{|S_k|}$  is a length- $|S_k|$  all-one column vector and  $|S_k|$  returns the size of set  $S_k$ . Therefore  $\boldsymbol{f}_m$  and  $\boldsymbol{w}_m$  can be designed as

$$\boldsymbol{f}_m = \boldsymbol{w}_m = (\boldsymbol{U}^{\boldsymbol{H}})^{-1} \boldsymbol{g}_m. \tag{8}$$

We now perform channel estimation in the first stage using the above designed beamforming/combining vectors  $\{f_m\}$ and  $\{w_m\}$ . In each time slot, we select one beamforming vector  $f_m$  at the transmitter and one combining vector  $w_n$ at the receiver to transmit/receive the pilot signal x. The corresponding channel output, denoted by  $y_{n,m}$ , is given by

$$y_{n,m} = \sqrt{p} \boldsymbol{w}_n^H \boldsymbol{H} \boldsymbol{f}_m \boldsymbol{x} + q_{n,m} \tag{9}$$

where  $q_{n,m}$  is the corresponding noise term in  $y_{n,m}$ . By using  $M^2$  time slots to span all the combinations of  $\{f_m\}$  and  $\{w_n\}$ , we generate a total number of  $M^2$  channel outputs that can be represented in a matrix form as

$$Y = \sqrt{p} W^{H} H F x + Q$$
  
=  $\alpha \sqrt{p} (u^{H}(\theta) W)^{H} (u^{H}(\phi) F)^{H} x + Q$  (10)

where  $F = [f_1, \dots, f_M]$ ,  $W = [w_1, \dots, w_M]$  and  $Q = \{q_{m,n}\}$  is an  $M \times M$  matrix.

Since we only have M = 2 beam patterns at each link end, a total number of  $M^2 = 4$  channel outputs, as described in (10), can be obtained. However, there are in total  $K^2 = 9$  transmitreceive sub-range combinations denoted by  $\{(k_t, k_r)|k_t, k_r =$  $1, 2, 3\}$  where  $k_t$  and  $k_r$  are respectively, the transmit and receive sub-range indices. To accurately extract the AOA/AOD information from Y, let us assume that AOD,  $\phi \in S_{k_t}$  and AOA,  $\theta \in S_{k_r}$ , without loss of generality. In this case, by recalling (6) we have

$$\boldsymbol{u}^{H}(\boldsymbol{\phi})\boldsymbol{f}_{m} = Cb_{m,k_{t}} \text{ and } \boldsymbol{u}^{H}(\boldsymbol{\theta})\boldsymbol{w}_{m} = Cb_{m,k_{r}}.$$
 (11)

Hence (10) can be re-written as

$$\boldsymbol{Y} = \alpha \sqrt{p} C^2 \boldsymbol{b}_{k_r} \boldsymbol{b}_{k_t}^T \boldsymbol{x} + \boldsymbol{Q}$$
(12)

where  $b_i$  denotes the  $i^{th}$  column of B in (5). By further vectorizing (12), we obtain

$$\operatorname{vec}(\boldsymbol{Y}) = \alpha \sqrt{p} C^2 \operatorname{vec}(\boldsymbol{b}_{k_r} \boldsymbol{b}_{k_t}^T) x + \operatorname{vec}(\boldsymbol{Q}) \\ = \alpha \sqrt{p} C^2(\boldsymbol{b}_{k_t} \otimes \boldsymbol{b}_{k_r}) x + \operatorname{vec}(\boldsymbol{Q}) \quad (13)$$

where  $b_{k_t} \otimes b_{k_r}$  denotes the Kronecker product between  $b_{k_t}$ and  $b_{k_r}$ . It is worth noting that the vector  $b_{k_t} \otimes b_{k_r}$  has a unit norm as  $b_{k_t}$  and  $b_{k_r}$  are both unit vectors.

To detect the pilot signal x from vec(Y), here we adopt the maximum ratio combining (MRC) principle and correlate vec(Y) using the unit vector  $\boldsymbol{b}_{k_t} \otimes \boldsymbol{b}_{k_r}$ . The resultant correlation output, referred to as a channel measurement, is given by

$$r_{k_r,k_t} = (\boldsymbol{b}_{k_t} \otimes \boldsymbol{b}_{k_r})^T \operatorname{vec}(\boldsymbol{Y}^{(s)}) = \boldsymbol{b}_{k_r}^T \boldsymbol{Y} \boldsymbol{b}_{k_t}.$$
 (14)

Note that the above discussions (11)-(14) are based on the assumption of  $\phi \in S_{k_t}$  and  $\theta \in S_{k_r}$ . If this assumption is correct, (14) reduces to

$$r_{k_r,k_t} = \alpha \sqrt{p} C^2 x + \boldsymbol{b}_{k_r}^T \boldsymbol{Q} \boldsymbol{b}_{k_t}, \qquad (15)$$

yielding an SNR of  $\frac{|\alpha|^2 p C^4}{N_0}$  for the pilot signal x. Otherwise if the assumption is incorrect, i.e.,  $\phi \in S_{k'_t}$  and  $\theta \in S_{k'_r}$  for some  $k'_t \neq k_t$  and/or  $k'_r \neq k_r$ , (14) reduces to

$$r_{k_r,k_t} = \alpha \sqrt{p} C^2 \boldsymbol{b}_{k_r}^T \boldsymbol{b}_{k_t'} \boldsymbol{b}_{k_t'}^T \boldsymbol{b}_{k_t} x + \boldsymbol{b}_{k_r}^T \boldsymbol{Q} \boldsymbol{b}_{k_t}$$
(16)

in which the SNR is given by  $(\boldsymbol{b}_{k_r}^T \boldsymbol{b}_{k_r} \boldsymbol{b}_{k_t}^T \boldsymbol{b}_{k_t})^2 \frac{|\alpha|^2 p C^4}{N_0}$  that is always no larger than half of the SNR in (15). By considering all possible values of  $k_r$  and  $k_t$ , we can obtain a total number of  $K^2 = 9$  channel measurements  $\{r_{k_r,k_t}\}$  that can form a  $K \times K$  matrix  $\boldsymbol{R}$ , i.e.,

$$\boldsymbol{R} = \boldsymbol{B}^T \boldsymbol{Y} \boldsymbol{B}.$$
 (17)

Finally, by finding,

$$(\hat{k}_r, \hat{k}_t) = \underset{k_r, k_t=1, \cdots, K}{\operatorname{argmax}} |r_{k_r, k_t}|,$$
(18)

we can reduce the ranges of possible AOAs and AODs to, respectively, the  $\hat{k}_t^{th}$  and  $\hat{k}_r^{th}$  transmit and 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.

#### B. Extension to General Cases

In general, for a single-path MMW channel, our proposed channel estimation algorithm also works in a similar multistage manner as that in [3]. In each stage, denoted s, we divide the possible AOA angular space into  $K = 2^M - 1$  (where M is an integer) non-overlapped sub-ranges  $S_{r,1}^{(s)}, S_{r,2}^{(s)}, \ldots, S_{r,K}^{(s)}$ and divide similarly the possible AOD angular space into  $S_{t,1}^{(s)}, S_{t,2}^{(s)}, \cdots, S_{t,K}^{(s)}$ . Then only M overlapped beam patterns

**Algorithm 1:** Single-path channel estimation algorithm for MMW channels.

$$\begin{array}{l} \mbox{Initialization}: \mathcal{S}_{t,k}^{(1)}, \mathcal{S}_{r,k}^{(1)} \forall k = 1, \cdots, K \\ \mbox{for } s \leq S \ \mbox{do} \\ \mbox{Calculate } {\pmb{F}}^{(s)} \ \mbox{based on } \mathcal{S}_{t,k}^{(s)} \forall k = 1, \cdots, K \ \mbox{and } {\pmb{B}} \\ \mbox{Calculate } {\pmb{W}}^{(s)} \ \mbox{based on } \mathcal{S}_{r,k}^{(s)} \forall k = 1, \cdots, K \ \mbox{and } {\pmb{B}} \\ \mbox{Calculate } {\pmb{W}}^{(s)} \ \mbox{based on } \mathcal{S}_{r,k}^{(s)} \forall k = 1, \cdots, K \ \mbox{and } {\pmb{B}} \\ \mbox{Calculate } {\pmb{W}}^{(s)} \ \mbox{based on } \mathcal{S}_{r,k}^{(s)} \forall k = 1, \cdots, K \ \mbox{and } {\pmb{B}} \\ \mbox{for } m = 1 \ to \ M \ \mbox{do} \\ \mbox{for } n = 1 \ to \ M \ \mbox{do} \\ \mbox{for } n = 1 \ to \ M \ \mbox{do} \\ \mbox{for } n = 1 \ to \ M \ \mbox{do} \\ \mbox{length end} \\ \mbox{end} \\ \mbox{for } n = 1 \ \mbox{for } M \ \mbox{do} \\ \mbox{length end} \\ \mbox{After } M^2 \ \mbox{measurements:} \\ \mbox{y}^{(s)} = \sqrt{p_s} ({\pmb{W}}^{(s)})^H {\pmb{H}} {\pmb{F}}^{(s)} x + {\pmb{Q}} \\ \mbox{R}_{(k_r}^{(s)}, \hat{k}_t^{(s)}) = \mbox{argmax} \\ \mbox{k}_{r,k_t=1,\cdots,K} |r_{k_r,k_r}| \\ \mbox{end} \\ \mbox{end} \\ \mbox{$\hat{\phi} = \frac{\pi}{N} \sum_{s=1}^{S} (\hat{k}_t^{(s)} - 1) K^{S-s}, \ \hat{\theta} = \frac{\pi}{N} \sum_{s=1}^{S} (\hat{k}_r^{(s)} - 1) K^{S-s} \\ \mbox{$\hat{\alpha} = P_R {\pmb{h}}^H ({\pmb{h}} P_R {\pmb{h}}^H + N_0 {\pmb{I}}_S)^{-1} {\pmb{r}} \\ \end{array}$$

will be designed at each end to cover these K sub-ranges. The designed M beam patterns are characterized by an  $M \times K$  beam pattern description matrix **B**, which consists of the normalized version of all the  $K = 2^M - 1$  non-zero length-M binary vectors as its columns. Besides the example of **B** with M = 2 and K = 3 in (5), another example with M = 3 and K = 7 is given by<sup>2</sup>

$$\boldsymbol{B} = \begin{bmatrix} 1 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} & 0 & 0 & 0\\ 0 & 0 & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} & 1 & \frac{1}{\sqrt{2}} & 0\\ 0 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} & 0 & 0 & \frac{1}{\sqrt{2}} & 1 \end{bmatrix}.$$
 (19)

Given B, we can then generate both the beamforming vectors  $\{f_m^{(s)}\}$  and combining vectors  $\{w_n^{(s)}\}$  in the same way as (8). For example, to generate  $f_m^{(s)}$ , the corresponding vector  $g_m$  in (8), which is redefined as  $g_{m,t}^{(s)}$  for rigorousness, should be designed such that its  $i^{th}$  entry, denoted by  $g_{m,t}^{(s)}(i)$ , satisfies

$$g_{m,t}^{(s)}(i) = \begin{cases} C_s b_{m,k}, & \text{if } \frac{\pi i}{N} \in \mathcal{S}_{t,k_t}^{(s)}, \exists k_t \in \{1, \cdots, K\}; \\ 0, & \text{if } \frac{\pi i}{N} \notin \mathcal{S}_{t,k_t}^{(s)}, \forall k_t \in \{1, \cdots, K\} \end{cases}$$
(20)

where  $C_s$  is a scalar constant that satisfies  $||\boldsymbol{f}_m^{(s)}||_2 = 1$ . Physically,  $g_{m,t}^{(s)}(i)$  describes the desired beam pattern amplitude at angle,  $\frac{\pi i}{N}$  when  $\boldsymbol{f}_m^{(s)}$  is used. Each combining vector  $\boldsymbol{w}_n^{(s)}$  can be designed in the same way.

The channel output on the  $s^{th}$  estimation stage can then be obtained via  $M^2$  time slots by

$$\boldsymbol{Y}^{(s)} = \sqrt{p_s} (\boldsymbol{W}^{(s)})^H \boldsymbol{H} \boldsymbol{F}^{(s)} \boldsymbol{x} + \boldsymbol{Q}^{(s)}$$
(21)

where  $p_s$  denotes the transmit power of the pilot signal in the  $s^{th}$  stage. Similar to that in [3], we prefer that all the stages have an equal probability of failure, indicating that we can allocate power among stages inverse proportionally to the beamforming gains of these beam patterns, i.e.,

$$p_s = \frac{P_T}{C_s^4} \forall s = 1, 2, \cdots$$
(22)

where  $P_T$  is a constant. Similarly to (17) we convert  $Y^{(s)}$  into the measurement matrix

$$\boldsymbol{R}^{(s)} = \boldsymbol{B}^T \boldsymbol{Y}^{(s)} \boldsymbol{B}.$$
 (23)

The most likely AOD/AOA combination is determined using

$$(\hat{k}_{r}^{(s)}, \hat{k}_{t}^{(s)}) = \operatorname*{argmax}_{k_{r}, k_{t}=1, \cdots, K} |r_{k_{r}, k_{t}}^{(s)}|$$
(24)

where  $r_{k_r,k_t}^{(s)}$  is the  $(k_r^{(s)}, k_t^{(s)})^{th}$  entry of  $\mathbf{R}^{(s)}$ . The selected sub-ranges,  $\mathcal{S}_{t,\hat{k}_t^{(s)}}^{(s)}$  and  $\mathcal{S}_{r,\hat{k}_r^{(s)}}^{(s)}$  are used for the channel estimation on the next stage.

Once  $S = \lceil \log_K N \rceil$  estimation stages of Algorithm 1 have been completed, the minimum angle resolution  $\frac{\pi}{N}$  is reached and the fading co-efficient,  $\alpha$ , can be estimated. [3] estimated  $\alpha$  based on only the measurement of the final stage. To improve the estimate accuracy, in this paper we propose to estimate  $\alpha$  by using the selected measurement in all stages of the algorithm. Denote by  $r_{\hat{k}_r^{(s)}, \hat{k}_t^{(s)}}^{(s)}$  the selected measurement from  $\mathbf{R}^{(s)}$  in the  $s^{th}$  stage. Define

$$\boldsymbol{r} = [r_{\hat{k}_{r}^{(1)}, \hat{k}_{t}^{(1)}}^{(1)}, r_{\hat{k}_{r}^{(2)}, \hat{k}_{t}^{(2)}}^{(2)}, \cdots, r_{\hat{k}_{r}^{(S)}, \hat{k}_{t}^{(S)}}^{(S)}]^{T}.$$
 (25)

Then from (15) and (22), we can see that, provided that the selection is correct in each stage, (25) can be rewritten as

$$\boldsymbol{r} = \sqrt{P_T} \boldsymbol{x} \cdot \boldsymbol{1}_S \boldsymbol{\alpha} + \boldsymbol{n} \tag{26}$$

where n is the  $S \times 1$  vector of corresponding noise terms. Following the MMSE principle, we can then estimate the fading coefficient  $\alpha$  as

$$\hat{\alpha} = \operatorname{Var}[\alpha] \sqrt{P_T} x^H \mathbf{1}_S^H (\operatorname{Var}[\alpha] P_T \cdot \mathbf{1}_S \mathbf{1}_S^H + N_0 \boldsymbol{I}_S)^{-1} \boldsymbol{r}$$
(27)

where  $Var[\alpha]$  is the variance of  $\alpha$ , and  $I_S$  is an  $S \times S$  identity matrix.

It can be seen that, compared with the channel estimation algorithm in [3] with the same value of K, our proposed algorithm also requires  $S = \lceil \log_K N \rceil$  stages, but the number of time slots required in each stage reduces to  $M^2 = \log_2^2(K+1)$ , instead of  $K^2$ . In general, this yields a  $\frac{K^2}{\log_2^2(K+1)}$  reduction in measurement time slots. For the example of K = 3 discussed earlier, a 225% reduction of measurement time slots can be achieved. Table 1 tabulates the total numbers of time slots required of the proposed algorithm and the algorithm in [3] over a range of antenna array sizes. It is seen a significant reduction in required time slots can be achieved with the proposed algorithm with both K = 3 and K = 7.

<sup>&</sup>lt;sup>2</sup>Note that the columns of **B** can be arbitrarily permuted, which does not affect the performance of our proposed channel estimation algorithm. However, a better column-permutation of **B**, e.g., a gray-coded based permutation, may facilitate the realization of the corresponding  $f_m$  and  $w_m$  when the hardware constraints are considered.

 TABLE I

 COMPARISON OF THE NUMBER OF TIME SLOTS REQUIRED FOR THE

 PROPOSED OVERLAPPED ALGORITHM AND THE NON-OVERLAPPED

 ALGORITHM IN [3].

	K=3				K=7			
N	3	9	27	81	7	49	343	2401
Overlapped	4	8	12	16	9	18	27	36
Non-overlapped	9	18	27	36	49	98	147	196

#### **IV. PERFORMANCE ANALYSIS**

In this section, we analyze the performance of our proposed channel estimation algorithm. We say that the channel estimation fails if the selected transmitter-receiver sub-range pair at the final stage S does not contain the correct AOD/AOA of the channel path. The corresponding probability of channel estimation failure (PCEF) can be expressed as

$$P_{fail}(S) = \Pr\left(\phi \notin \mathcal{S}_{t,\hat{k}_t^{(S)}}^{(S)} \text{ or } \theta \notin \mathcal{S}_{r,\hat{k}_r^{(S)}}^{(S)}\right).$$
(28)

Define  $P_{fail}(s|s-1)$  as the probability that the channel estimation succeeds at stage s-1 but fails at stage s, i.e.,

$$P_{fail}(s|s-1) \triangleq \Pr\left(\phi \notin \mathcal{S}_{t,\hat{k}_{t}^{(s)}}^{(s)} \text{ or } \theta \notin \mathcal{S}_{r,\hat{k}_{r}^{(s)}}^{(s)} \right. \\ \left|\phi \in \mathcal{S}_{t,\hat{k}_{t}^{(s-1)}}^{(s-1)} \text{ and } \theta \in \mathcal{S}_{r,\hat{k}_{r}^{(s-1)}}^{(s-1)}\right).$$
(29)

Then (28) can be upper bounded by

$$P_{fail}(S) = 1 - \prod_{s=1}^{S} (1 - P_{fail}(s|s-1)) \le \sum_{s=1}^{S} P_{fail}(s|s-1),$$
(30)

in which each additive term  $P_{fail}(s|s-1)$  can be further elaborated as follows. Denote by  $k_t^{\prime(s)}$  and  $k_r^{\prime(s)}$ , respectively, the indices of the sub-ranges where the AOD and AOA of the single path falls in at the  $s^{th}$  stage, i.e.,  $\phi \in \mathcal{S}_{t,k_t^{\prime(s)}}^{(s)}$  and  $\theta \in \mathcal{S}_{r,k_r^{\prime(s)}}^{(s)}$ . Then physically,  $P_{fail}(s|s-1)$  corresponds to the event that the correct measurement  $r_{k_r^{\prime(s)},k_t^{\prime(s)}}^{(s)}$  does not have the largest magnitude among all entries of  $\mathbf{R}^{(s)}$ . Mathematically, we have

$$P_{fail}(s|s-1) = \frac{1}{K^2} \sum_{k_r^{\prime(s)}, k_t^{\prime(s)}=1}^{K} \Pr\left(\max_{k_r, k_t = 1, \cdots, K} |r_{k_r, k_t}^{(s)}| > |r_{k_r^{\prime(s)}, k_t^{\prime(s)}}^{(s)}|\right) \\ \leq \frac{1}{K^2} \sum_{k_r^{\prime(s)}, k_t^{\prime(s)}=1}^{K} \sum_{k_r, k_t=1}^{K} \Pr(|r_{k_r, k_t}^{(s)}| > |r_{k_r^{\prime(s)}, k_t^{\prime(s)}}^{(s)}|).$$
(31)

In order to derive an explicit expression for each probability term in (31), we need to find the joint distributions of  $r_{k_r,k_t}^{(s)}$  and  $r_{k_r'^{(s)},k_t'^{(s)}}^{(s)}$ . For a given  $\alpha$ , by recalling (16), (5) and (22) we

can see that  $r_{k_r,k_t}^{(s)}$  and  $r_{k'_r^{(s)},k'_t^{(s)}}^{(s)}$  can form a jointly Gaussian distributed length-2 vector with mean

$$\mathbf{E}\begin{bmatrix} r_{k_r,k_t}^{(s)} \\ r_{k_r^{(s)},k_t^{(s)}}^{(s)} \end{bmatrix} = \begin{bmatrix} \alpha \sqrt{P_T} \boldsymbol{b}_{k_r}^T \boldsymbol{b}_{k_r^{(s)}} \boldsymbol{b}_{k_t}^T \boldsymbol{b}_{k_t} x \\ \alpha \sqrt{P_T} x \end{bmatrix} \triangleq \begin{bmatrix} a \\ b \end{bmatrix}. \quad (32)$$

and covariance matrix

$$\operatorname{Cov}\left[\begin{array}{c} r_{k_r,k_t}^{(s)} \\ r_{k_r^{(s)},k_t^{(s)}}^{(s)} \end{array}\right] = \left[\begin{array}{c} N_0 & \Sigma \\ \Sigma & N_0 \end{array}\right]$$
(33)

where

$$\Sigma = \mathbf{E}[(r_{k_{r},k_{t}}^{(s)} - \mathbf{E}[r_{k_{r},k_{t}}^{(s)}])^{H}(r_{k_{r}^{(s)},k_{t}^{\prime(s)}}^{(s)} - \mathbf{E}[r_{k_{r}^{\prime(s)},k_{t}^{\prime(s)}}^{(s)}])]$$
  

$$= \mathbf{E}[(\boldsymbol{b}_{k_{r}}^{T}\boldsymbol{Q}^{(s)}\boldsymbol{b}_{k_{t}})^{H}(\boldsymbol{b}_{k_{r}^{\prime(s)}}^{T}\boldsymbol{Q}^{(s)}\boldsymbol{b}_{k_{t}^{\prime(s)}}^{(s)})]$$
  

$$= \mathbf{E}[\boldsymbol{b}_{k_{t}}^{T}(\boldsymbol{Q}^{(s)})^{H}\boldsymbol{b}_{k_{r}}\boldsymbol{b}_{k_{r}^{\prime(s)}}^{T}\boldsymbol{Q}^{(s)}\boldsymbol{b}_{k_{t}^{\prime(s)}}^{T}]$$
  

$$= N_{0}(\boldsymbol{b}_{k_{r}}^{T}\boldsymbol{b}_{k_{r}^{\prime(s)}}\boldsymbol{b}_{k_{t}^{\prime}}^{T}\boldsymbol{b}_{k_{t}}).$$
(34)

Then following equation (4B.21) of [8], we have

$$\Pr\left(|r_{k_r,k_t}^{(s)}| > |r_{k_r'^{(s)},k_t'^{(s)}}^{(s)}| \left| |\alpha| \right) = Q_1(A,B) - \frac{1}{2}I_0(AB)\exp(-\frac{1}{2}(A^2 + B^2)) \quad (35)$$

where  $Q_1$  is the first order generalized Marcum Q function,  $I_0$  is the 0th order modified Bessel function of the first kind,  $A = |a|/\sqrt{N_0 - \Sigma}$  and  $B = |b|/\sqrt{N_0 - \Sigma}$ . By further assuming that the magnitude of the channel coefficient,  $|\alpha|$ , follows distribution  $f_{|\alpha|}(x) = \frac{x}{\operatorname{Var}[\alpha]} \exp(-\frac{x}{2\operatorname{Var}[\alpha]})$ , we can write each additive term in (35) as [9]

$$\Pr(|r_{k_r,k_t}^{(s)}| > |r_{k_r'^{(s)},k_t'^{(s)}}^{(s)}|) = \int_0^\infty \Pr(|r_{k_r,k_t}^{(s)}| > |r_{k_r'^{(s)},k_t'^{(s)}}^{(s)}| ||\alpha|) f_{|\alpha|}(x) dx = \frac{1}{2} - \frac{\bar{B}^2 - \bar{A}^2}{4\sqrt{1 + \bar{B}^2 + \bar{A}^2 + (\frac{\bar{B}^2}{2} - \frac{\bar{A}^2}{2})^2}}$$
(36)

where  $\bar{A} = E[A] = Var[\alpha] \sqrt{P_T} \boldsymbol{b}_{k_r}^T \boldsymbol{b}_{k_r'^{(s)}} \boldsymbol{b}_{k_t'^{(s)}}^T \boldsymbol{b}_{k_t} |x| / \sqrt{N_0 - \Sigma}$ and  $\bar{B} = E[B] = Var[\alpha] \sqrt{P_T} |x| / \sqrt{N_0 - \Sigma}$ . By substituting (36) into (31) we have

$$P_{fail}(s|s-1) \leq \frac{1}{K^2} \sum_{k_r^{\prime(s)}, k_t^{\prime(s)} = 1}^{K} \sum_{k_r, k_t = 1}^{K} \left[ \frac{1}{2} - \frac{\bar{B}^2 - \bar{A}^2}{4\sqrt{1 + \bar{B}^2 + \bar{A}^2 + (\frac{\bar{B}^2}{2} - \frac{\bar{A}^2}{2})^2}} \right].$$
 (37)

Finally the PCEF can be upper bounded by substituting (37) into (30) such that



Fig. 3. Comparison of the probability of failed channel estimation between the proposed algorithm and [3]'s non-overlapped algorithm.

$$P_{fail}(S) \leq \frac{S}{K^2} \sum_{k'_r, k'_t = 1}^{K} \sum_{k_r, k_t = 1}^{K} \left[ \frac{1}{2} - \frac{\bar{B}^2 - \bar{A}^2}{4\sqrt{1 + \bar{B}^2 + \bar{A}^2 + (\frac{\bar{B}^2}{2} - \frac{\bar{A}^2}{2})^2}} \right].$$
 (38)

### V. NUMERICAL RESULTS

We now provide some numerical examples to verify the performance of our proposed algorithm. We consider a singlepath MMW system with N = 27 antennas at both the transmitter and receiver. The fading coefficient is assumed to follow a complex Gaussian distribution with zero mean and variance  $N^2$ . We set K = 3 for both our algorithm and the algorithm in [3]. Thus,  $K^2 = 9$  time slots are required in each stage of [3]. However, in our proposed algorithm, only  $M^2 = 4$  time slots are involved in in each stage. This significantly reduces the total time slots required for channel estimation. Power allocation among the S stages is applied to both algorithms as (22).

Fig.3 shows the PCEF as described in (28), where the total energy required in the overall channel estimation process is calculated by  $E_T = M^2 \sum_{s=1}^{S} p_s$ . We can see that, to achieve the same PCEF as the algorithm in [3], our proposed algorithm requires 2.5dB more energy in order to reduce the channel estimation time by 225%. The gap between the simulation results and derived upper bound is similar to that derived in [3]. This gap originates from the relaxations in (30) and (31).

Fig.4 shows the relative estimation error of the fading coefficient  $\alpha$ , i.e.,  $\frac{|\hat{\alpha}-\alpha|}{|\alpha|}$ . As can be seen, the proposed algorithm performs better than that in [3] at the same PCEF as more energy is required in the former. Fig.4 also shows that by combining the selected measurements from all stages with MMSE estimator, we significantly improve the estimation accuracy for both algorithms.



Fig. 4. Comparison of relative error in fading coefficient estimation between the non-overlapped algorithm from [3] and the proposed algorithm.

# VI. CONCLUSION

In this paper we have proposed a new channel estimation algorithm for MMW channels based on a novel overlapped beamform pattern design. In general, our proposed algorithm can hasten the channel estimation process by a factor of  $\frac{K^2}{\log_2^2(K+1)}$  when compared to the existing method developed in [3] with the same value of K. For example, with K = 3 the proposed algorithm reduces the number of measurement time slots by 225% and will achieve the same PCEF as that in [3] for a cost of 2.5dB more energy. For channels with rapidly changing channel information, this cost can be justified in order to improve the estimation speed.

#### REFERENCES

- 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] A. Alkhateeb, O. El 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.
- [4] 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.
- [5] 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.
- [6] 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.
- [7] A. Sayeed, "Deconstructing multiantenna fading channels," *IEEE Trans. Sig. Process.*, vol. 50, no. 10, pp. 2563–2579, Oct. 2002.
- [8] J. Proakis, *Digital Communications*, ser. McGraw-Hill series in electrical engineering: Communications and information theory. McGraw-Hill, 1983.
- [9] M. K. Simon and M.-S. Alouini, "A unified approach to the probability of error for noncoherent and differentially coherent modulations over generalized fading channels," *IEEE Trans. Commun.*, vol. 46, no. 12.