Efficient Beam Sweeping Algorithms and Initial Access Protocols for Millimeter-wave Networks

Irmak Aykin, *Student Member, IEEE,* and Marwan Krunz, *Fellow, IEEE*

***Abstract*—5G millimeter-wave (mmW) systems rely on electronically steerable antenna arrays to support directional communications. Directionality complicates the initial access (IA) process, whereby a base station (BS) announces itself to nearby user equipments (UEs), giving them the opportunity to associate with this BS. Existing approaches for IA suffer from long discovery time and/or nonnegligable probability of missing UEs. In this paper, we propose FastLink, an efficient IA protocol for mmW systems, in which discovery beacons are transmitted/received using the narrowest possible beams, allowing for high beamforming gains and low misdetection rate, while maintaining low discovery time. Fastlink executes a unique algorithm, called 3-dimensional peak finding (3DPF), to find the best beam in logarithmic time. We formulate the beam-finding process as a sparse problem and use compressive sensing to determine the minimum number of measurements needed for this process. We first study FastLink for the discovery of a single UE and then extend our analysis to a multi-user scenario. Both simulations and over-the-air experiments based on a custom mmW testbed are used to evaluate FastLink. Our results verify its efficiency, and show that it can reduce the search time by 90% compared to the scanning approach used in 802.11ad systems.**

***Index Terms*—Millimeter-wave, beam finding, initial access, compressive sensing, analog beamforming.**

I. Introduction

Millimeter-wave (mmW) communications have recently attracted considerable interest as a key element of next-generation wireless systems, e.g., 5G New Radio (NR) cellular systems [1] and WiGig [2]. Wireless communications at mmW frequencies are inherently challenging, due to high propagation losses and poor penetration[3], [4]. At the same time, the smaller wavelengths allow many antenna elements to be packed into a single device without increasing its form factor. With proper processing of signals fed into these antennas, transmissions can be beamed along a desired direction. The severe signal attenuation in the mmW bands can then be compensated for by the resulting beamforming gain [5].

Directional communications, however, comes at the cost of more complex initial access (IA) procedure [6], the process by which a user equipment (UE) establishes a connection with a base station (BS). In 4G LTE systems, IA is performed in an omnidirectional fashion, which alleviates the burden of beam alignment. In contrast, the IA procedure in 5G NR is done directionally (see Fig. 1) to reach more users and support subsequent directional data transmissions. Specifically,

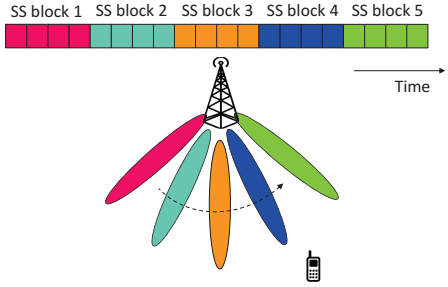


Fig. 1. Directional IA as done in 5G NR. SS blocks are transmitted over different directions to establish communications with the UEs in range.

synchronization signal blocks (SSBs), each consisting of 4 OFDM symbols, are transmitted over different directions at different times to establish communications with the UEs in range. This directional search incurs significant delay, which lowers the link throughput and spectral efficiency [7].

In this paper, we propose FastLink, an efficient mmW-band IA protocol that allows the BS and the UE to transmit/receive using the narrowest possible beams, hence providing the highest possible beamforming gain. At the same time, beam sweeping is done in such a way that the discovery process is much faster than beam scanning mechanisms used in the 802.11ad standard and 5G NR. FastLink exploits compressive sensing (CS) to determine the number of measurements (beam probes) needed to find the 'dominant' channel cluster[1]. Based on this insight, we design a search algorithm, called 3-dimensional peak finding (3DPF), to find the best beam in the 3D space. 3DPF divides the set of beam directions into equally spaced subsets and then finds the beam that achieves the maximum received power in each subset. The number of subsets is a design parameter, determined using CS. We first analyze a single-user scenario and then extend our treatment into a more realistic multi-user scenario, studying the impact of the number of users on FastLink's performance. A higher number of users results in more clusters from the BS perspective. Thus, the BS should search more regions to achieve a low probability of UE misdetection. First, we find the distribution of the total number of clusters in the multi-user case and then optimize the number of subsets searched in 3DPF to fit the multi-user system model.

Main contributions of this paper can be summarized as:

- Utilizing CS methods, we determine the required number of probed directions to identify the channel clusters with

I. Aykin and M. Krunz are with the Department of Electrical and Computer Engineering, University of Arizona, Tucson, AZ, 85721, USA (email: aykin@email.arizona.edu, krunz@email.arizona.edu).

An abridged version of this paper was presented at the MSWiM '18 Conference, Oct. 28 - Nov. 2, 2018, Montreal, Canada.

[1]A cluster is a collection of paths, whose angle-of-arrivals (AoA), angle-of-departures (AoD), and path gains are quite close to each other.

high probability. We show that this number is logarithmically proportional to the total number of narrow beams.
- We develop a novel beam scanning method called 3DPF, which is based on beam partitioning and gradient ascent search. 3DPF collects the required number of beam probes suggested by the CS analysis and allows mmW devices to discover each other directionally, while maintaining low discovery time and misdetection probability.
- We integrate 3DPF into the FastLink protocol and explain the required control messages that need to be exchanged between the BS and the UE in support of this protocol.
- We extend our single-user analysis into multiple users. For this scenario, we compute the total number of clusters observed by the BS and show how the discovery time scales with the number of users.
- We verify the efficiency of 3DPF by conducting MATLAB simulations as well as hardware experiments on a custom mmW testbed. Our simulations implement a slightly modified version of the NYU channel model [5]. Our hardware setup is comprised of a mmW signal generator, vector signal analyzer, and 4×4 uniform planar arrays. We run the experiments in the 29 GHz band, collecting received signal strength (RSS) measurements and timing results. Our results indicate that FastLink reduces the search time by more than 90% compared to 802.11ad-like beam search, and yet achieves almost the same misdetection probability.

II. Related Work

For mmW systems that rely on analog beamforming, several approaches for IA have been investigated in the literature, including exhaustive search [1], two-stage hierarchical search [2], and context-information-based (CI-based) search [8]. Exhaustive search is a brute-force sequential beam searching technique proposed for 5G NR [9]. In this technique, the BS sequentially transmits SSBs along different directions, allowing UEs to detect one of them [1]. This exhaustive search comes at the cost of significant discovery time. The two-stage scanning used in the 802.11ad standard [2] employs a hierarchical multi-resolution beamforming codebook to expedite IA. In the first stage, the access point (AP) sequentially transmits synchronization signals over wide (quasi-omnidirectional) sectors, aiming to determine the best coarse direction. In the second stage (beam refinement), the AP refines its search within the best coarse sector by switching to narrow beams [2]. Although this approach reduces the IA delay, in the worst case, the search time scales linearly with the number of narrow beams. Finally, in CI-based search, the UE simply selects the optimal beam direction using GPS information [7].

Other IA approaches have also been proposed. The authors in [10] used hashing functions to identify the best beam. This reduces the search time, but it also lowers the resulting beamforming gain, leading to a higher misdetection probability. In [11] and [12], the authors utilized multi-lobe beams to identify several channel clusters for combating blockages. However, the resulting beams suffer again from the low beamforming gains. In [13] and [14], the authors proposed a beamforming scheme that exploits the correlation between the sub-6 GHz and the mmW interfaces to provide efficient beam alignment. Their scheme hinges on the availability of a sub-6 GHz channel and the feasibility of performing digital beamforming on the sub-6 GHz channel. The authors in [15] proposed an online learning based channel estimation technique to be used during IA. Nevertheless, this scheme adds a significant training overhead to the system. In [16], the authors studied contextual bandits for efficient beam alignment in mmW systems, but considered only a single channel cluster. Finally, the authors in [17] proposed a model-driven beam-steering scheme called OScan, which reduces the search latency. However, their design is based on mmW channel measurements obtained using horn antennas. These antennas have significantly different characteristics than phased-array antennas, which may result in their algorithm returning a suboptimal beam.

The aforementioned techniques offer different tradeoffs between *discovery time* (time to establish directional communication between the BS and the UE) and *misdetection probability*; the probability that a UE is not detected by the BS. For instance, in its first stage, the 802.11ad scheme scans the space using wide beams, and thus achieves a low discovery time. However, it also has a high chance of missing users due to the low beamforming gain. With the beam refinement in phase 2, the 802.11ad scheme attains better coverage, but this comes at the expense of higher discovery time.

Recently, compressive sensing has gained attention as a means to exploit the sparsity of mmW channels [18], [19]. Essentially, CS is a signal processing technique that can be used to efficiently reconstruct a signal by solving an underdetermined set of linear equations. This idea was applied in [18] for mmW channel estimation and designing analog beamformers. Similarly, the authors in [20] described a CS-based approach to estimate multipath channels (not necessarily mmW) that have a sparse representation, without imposing analog beamforming constraints. In [19], the authors utilized CS in a multiuser MIMO system. While these analyses corroborate the significance of CS for channel estimation, they do not specify a practical way for collecting the required measurements.

By exploiting mmW channel sparsity, CS was utilized in [21], [22] for efficient alignment of transmit and receive beams. In these works, probing beams are generated by assigning random phase shifts to each antenna element, so as to achieve incoherence between measurements. Although the resulting beams satisfy the CS constraints, they exhibit quasi-omnidirectional patterns and low beamforming gains. Therefore, these beams are not likely to satisfy the link budget when used for IA and/or data communication [7]. Adaptive CS analysis was used in [23] and [24] to obtain a multi-resolution hiearchical codebook to discover UEs in range. This type of hierarchical codebook iteratively narrows down the spatial search region, similar to divide-and-conquer algorithms, resulting in logarithmic time complexity [25]. However, similar to the random sensing matrices, the wide sectors used during the initial stages of the beam search process may not have enough gain to satisfy the link budget [7]. Consequently, such an approach results in a higher misdetection probability compared

to always using narrow beams. Finally, the authors in [26] utilized CS to design transmit and receive beamformers, but considered only a single-user scenario.

To obviate the above shortcomings, in this paper, we develop a practical algorithm based on CS, which reduces the beam searching delay while achieving a very low misdetection probability. Our algorithm executes the IA process completely over mmW bands. It exploits the unimodality of the received power as a function of beam ID, similar to gradient ascent, eventually maximizing the received power. We first consider the case of a single-user and then extend our treatment into multiple users.

III. Beamforming in Antenna Arrays

We first consider the IA process between a BS and a single UE. Electronically steerable phased-array antennas are assumed at both devices. In addition, we consider analog beamforming for both the BS and the UE. Without loss of generality, we let the BS be the transmitter (Tx) and UE be the receiver (Rx). Before we invoke on the algorithmic details of our IA design, we first explain how a phased-array antenna can be steered towards a desired direction. In this section, we study the beamforming on the UE side (Rx). Extension to the BS is straightforward.

To electronically steer a beam, complex weights should be applied to each antenna element in the array. These weights are determined by calculating the array factors (AFs) of the BS and UE antenna arrays. The AF is the factor by which the element factor of an individual antenna is multiplied to get the firing pattern of the entire array. Without loss of generality, consider a uniform planar array (UPA) with a horizontal inter-element distance d_x and a vertical inter-element distance d_y. Denote the AF by F_{UPA} and let R be the individual gain of each antenna element. Suppose that the incident wave of the received signal arrives at a polar angle θ and azimuth angle ϕ, and that the antennas are placed on an $M \times N$ grid, as shown in Fig. 2(a). Let $\kappa \triangleq d_x m \cos\phi \sin\theta + d_y n \sin\phi \sin\theta$. We can write the received signal at antenna element (m, n), $s_{m,n}$ as:

$$s_{m,n} = R\, e^{j\frac{2\pi}{\lambda}\kappa} \tag{1}$$

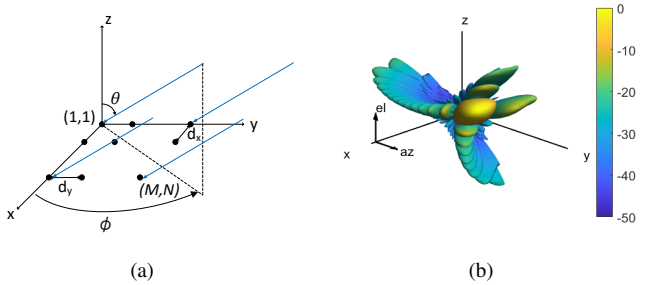
where λ is the wavelength of the signal, $m \in \{1, \cdots, M\}$ and $n \in \{1, \cdots, N\}$. Let s be the output of the antenna array:

$$s = \sum_{m=1}^{M}\sum_{n=1}^{N} s_{m,n} w_{m,n} = R \sum_{m=1}^{M}\sum_{n=1}^{N} w_{m,n} e^{j\frac{2\pi}{\lambda}\kappa} = R\, F_{\text{UPA}}. \tag{2}$$

The signal power at the UPA can then be maximized by maximizing $|R\ F_{\text{UPA}}|$. Assuming the same signal amplitude at each antenna, $|F_{\text{UPA}}|$ is maximized when $w_{m,n}$ is selected in a way to ensure that the received signals are in phase, i.e., by setting $w_{n_x,n_y} = e^{-j\frac{2\pi}{\lambda}\kappa}$. This way, the beam can be steered along the direction (θ, ϕ), as shown in Fig. 2(b).

IV. CS-based Beam Finding

Considering the limited scattering of a mmW channel, in this section, we formulate the problem of finding the best beam as a sparse problem, in line with previous works [5], [18]. We use the notation $x \times y$ to denote a matrix of x rows and y columns. Let the total number of antennas at the BS and the UE be $A_{\text{BS}} = M_{\text{BS}} N_{\text{BS}}$ and $A_{\text{UE}} = M_{\text{UE}} N_{\text{UE}}$, respectively. Here, M_i and N_i, $i \in \{\text{BS}, \text{UE}\}$, refer to the number of rows and columns in the UPA. Then, the sparse channel between the BS and the UE can be denoted by the $A_{\text{UE}} \times A_{\text{BS}}$ matrix $\mathbf{H}$.



(a) (b)

Fig. 2. (a) Illustration of analog Rx beamforming on a UPA (angle of the Rx signal is (θ, ϕ)), (b) normalized 3D directivity pattern of a 16×16 UPA, where $\theta = 45°$ and $\phi = 15°$ (UPA placed on the Y-Z plane).

To express the received signal, BS and UE beamforming should be applied to channel $\mathbf{H}$. In general, beamforming vectors are computed offline for a set of directions (as shown in Section III) and stored in the codebooks at the BS and UE [9]. Note that in addition to phase control, it is also possible to adjust the amplitude of a signal arriving at an antenna element of a phased array [27], and the codebook-based beamforming explained here can also be used with simultaneous amplitude- and phase-controlled antennas. This codebook-based beamforming is employed both in 5G NR [1] and WiGig [2] standards, as it is much faster than designing coherent precoder and combiner vectors on-the-fly based on channel measurements. Let $\mathbb{Q} = \{\mathbf{q}_1, \mathbf{q}_2, \cdots, \mathbf{q}_{D_{\text{UE}}}\}$ denote the codebook of the UE beamformer and $\mathbb{F} = \{\mathbf{f}_1, \mathbf{f}_2, \cdots, \mathbf{f}_{D_{\text{BS}}}\}$ denote the codebook of the BS beamformer, where D_{UE} and D_{BS} are the maximum number of narrow beams that can be generated at the UE and BS, respectively[2]. During IA, if the BS uses a transmit beamforming vector $\mathbf{f}_i \in \mathbb{F}$ and UE uses a receive beamforming vector $\mathbf{q}_j \in \mathbb{Q}$, the received signal y_{ij} can be expressed as:

$$y_{ij} = \mathbf{q}_j^* \mathbf{H} \mathbf{f}_i x + \mathbf{q}_j^* \mathbf{n} \tag{3}$$

where x is the transmitted signal and $\mathbf{n} \in \mathbb{C}^{A_{\text{UE}} \times 1}$ is a matrix whose entries are complex circularly symmetric white Gaussian noise.

For UE beam training, we fix the BS beamformer and define the precoded channel vector $\mathbf{g}_i \triangleq \mathbf{H}\mathbf{f}_i$, where $\mathbf{g}_i \in \mathbb{C}^{A_{\text{UE}} \times 1}$. Then, the problem of finding the optimal UE beamformer given the precoded channel vector can be expressed as:

$$\begin{aligned} \underset{j}{\text{argmax}} \quad & |\mathbf{q}_j^* \mathbf{g}_i|^2 \\ \text{s.t.} \quad & \mathbf{q}_j \in \mathbb{Q}. \end{aligned} \tag{4}$$

A brute-force solution to (4) can be found by first forming an exhaustive dictionary matrix $\mathbf{Q}$ whose jth column is given by

[2] D_{UE} and D_{BS} depend on the resolution of the phase shifters of the UPAs.

$\mathbf{q}_j$, $j = 1, \cdots, D_{\text{UE}}$. Then, D_{UE} measurements corresponding to different Rx beamformers can be collected as $\mathbf{v} = \mathbf{Q}^*\mathbf{g}_i$, where $\mathbf{v}$ is a vector of length D_{UE}, containing the channel measurement results [23]. Note that these results are collected in D_{UE} successive time slots, i.e., the resulting $\mathbf{v}$ is found by concatenating D_{UE} successive results. Our aim here is to exploit channel sparsity and find a solution to (4) that is faster than a brute-force search. For notational convenience, we omit the subscript from D_{UE} for the rest of the paper.

After transmit and receive beamforming, $\mathbf{v}$ will contain mostly zero entries. Specifically, assuming that the channel $\mathbf{H}$ has P clusters, when $\mathbf{Q}$ is applied to the precoded channel $\mathbf{g}_i$, the results display at most P peak values. That is, $\mathbf{v}$ is P-sparse. In general, the notion of P-sparsity is used for vectors with at most P nonzero entries. As shown in [28], CS can be used to reconstruct noisy vectors with P peaks. Therefore, for notational convenience, we refer to vectors exhibiting at most P peaks as P-sparse. Insights from the CS theory can thus be used to identify the largest entry of $\mathbf{v}$ by taking only r measurements from $\mathbf{g}_i$, such that $D > r \geq P$.

Define $\mathbf{\Theta}$ as the $A_{\text{UE}} \times r$ compressed measurement matrix and $\mathbf{\Phi}$ as the $r \times D$ dimensionality reduction matrix. Then, the output $\mathbf{\Psi}$ of the compressed measurement process can be represented as:

$$\mathbf{\Psi} = \mathbf{\Phi}\mathbf{v} = \mathbf{\Phi}\mathbf{Q}^*\mathbf{g}_i = \mathbf{\Theta}^*\mathbf{g}_i. \tag{5}$$

Note that $\mathbf{\Psi}$ is a column vector of length r. The problem is then reduced to designing a stable measurement matrix $\mathbf{\Theta}$ such that the key information in any P-sparse compressible channel is not lost through the dimensionality reduction. In other words, we would like to identify the highest peak of $\mathbf{v}$ by using the available measurement results $\mathbf{\Psi}$ and the known dimensionality reduction matrix $\mathbf{\Phi}$. A necessary and sufficient condition to find a solution to our problem for the P-sparse $\mathbf{v}$ is the restricted isometry property (RIP), which is satisfied by a given matrix $\mathbf{\Phi}$ if:

$$(1-\epsilon) \leq \frac{\|\mathbf{\Phi}\mathbf{z}\|_2^2}{\|\mathbf{z}\|_2^2} \leq (1+\epsilon) \tag{6}$$

holds for any arbitrary P-sparse $\mathbf{z}$ and for some isometry constant $\epsilon > 0$ [29]. Specifically, the matrix $\mathbf{\Phi}$ must preserve the length of an arbitrary P-sparse $\mathbf{z}$. RIP can be achieved with high probability simply by selecting $\mathbf{\Phi}$ as a random matrix [30]. That is, when the entries of $\mathbf{\Phi}$ are drawn independently from a $\mathcal{N}(0, 1/r)$ distribution, $\mathbf{\Phi}$ can satisfy RIP with a probability close to 1 if $r \geq cP\log(D/P)$, where c is a small constant that depends on the desired probability of success. This result is summarized in the following theorem [31]:

Theorem 1: Let $r \geq cP\log(D/P)$. Construct $\mathbf{\Phi}$ by drawing its entries independently from a Gaussian distribution $\mathcal{N}(0, 1/r)$. Then, there exists a $c' > 0$ that depends on c such that with probability greater than $1 - e^{-c'r}$, it is possible to reconstruct every P-sparse signal $\mathbf{v}$ of length D using $\mathbf{\Psi}$.

Proof: Follows from [31, Thm. 5.2]. ∎

Accordingly, the highest peak of the P-sparse $\mathbf{v}$ can be identified from only $r \geq cP\log(D/P)$ random measurements. However, if the entries of $\mathbf{\Phi}$ are selected randomly, the resulting measurement matrix $\mathbf{\Theta}$ consists of random phase shifts. Therefore, the beam patterns used in the channel measurements will have low beamforming gains, which will likely not satisfy the link budget [21]. On the other hand, by selecting the columns of $\mathbf{\Theta}$ from the codebook $\mathbb{Q}$, the beamforming gains can be maximized for a specified set of directions. This corresponds to constructing $\mathbf{\Phi}$ as a row selection matrix, which consists of a single 1 at each row (total of r rows) and each column (total of D columns). The remaining entries of the matrix consist of 0s and the column indices of 1s indicate the rows to be selected from $\mathbf{Q}^*$. Fortunately, this is possible, given that different UE beamformers are spatially orthogonal. As long as $\mathbf{\Theta}$ is constructed based on a subset of measurement vectors that are selected from an orthogonal basis ($\mathbb{Q}$), RIP is still satisfied [32]. Note that the orthogonality of the UE beamformers require sufficient beam separation.

Using the above insights, we next describe three algorithms that aim at determining the best beams for the BS and UE. The algorithms aim at collecting $cP\log(D/P)$ measurements in an intelligent way, so as to ensure a low misdetection probability. Because the measurements are taken sequentially, the proposed algorithms use the measurement results from previous steps when deciding on the next UE beam.

V. Peak-Finding Algorithms

The algorithms proposed in this section seek to identify the Rx beam j that yields the largest Rx power for the given precoded channel. In Section V-A, we design an algorithm to find the best beam in the case of a single channel cluster. We extend this design to multiple clusters in Section V-B. In Section V-C, we modify both algorithms to work in a 3D environment. Finally, in Section V-D, we discuss how to select the optimal design parameter K (number of angular regions).

A. Single-peak Finding (SPF) Algorithm

For now, we assume that the Tx is pointing towards a fixed, arbitrarily chosen direction using its beam i and beamformer $\mathbf{f}_i$. Rx tries to find the optimal beam w.r.t the given Tx direction. Extension to finding the optimal beam on the Tx side will be explained in Section VI. While pointing along a given direction i, the Tx sends synchronization signals (see Section VI) and Rx measures the received power by steering its receive beam. For brevity, in this section, we omit the subscript i from $\mathbf{g}_i$. For now, we consider a fixed elevation angle for both Tx and the Rx, and aim at finding the best azimuthal angle for the Rx.

Due to the channel clustering phenomenon, a typical plot of beam ID vs. SNR displays multiple peaks. For now, we consider a single-peak scenario; see Fig. 3. Note that the graph in this figure cannot be perfectly reconstructed without measuring the SNR for each Rx beam. However, our aim is to find the peak by using only a subset of Rx beams, exploiting the previously presented CS analysis. Specifically, we aim at taking r measurements ($r < D$) such that the optimal Rx beamformer for a given Tx beamformer can be found with high probability (see Thm. 1).

Our proposed algorithm, called SPF, is similar to a gradient ascent search. First, the precoded channel $\mathbf{g}$ is measured using

Algorithm 1 Single-peak Finding (SPF) Algorithm

procedure SINGLE_PEAK($\mathbb{Q}$, $\mathbf{g}$)
Initialize:
  $i \leftarrow D/2$
  $v_i \leftarrow \mathbf{q}_i^* \mathbf{g}, \quad v_{i-1} \leftarrow \mathbf{q}_{i-1}^* \mathbf{g}, \quad v_{i+1} \leftarrow \mathbf{q}_{i+1}^* \mathbf{g}$
Recursion:
  if $|v_i|^2 > |v_{i-1}|^2$ **and** $|v_i|^2 > |v_{i+1}|^2$ **then**
    return $|v_i|^2$
  else
    if $|v_{i+1}|^2 - |v_i|^2 > |v_{i-1}|^2 - |v_i|^2$ **then**
      return SINGLE_PEAK($\mathbb{Q}\{i+1:D\}, \mathbf{g}$)
    else
      return SINGLE_PEAK($\mathbb{Q}\{1:i-1\}, \mathbf{g}$)

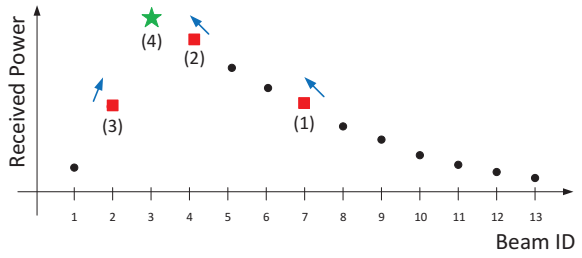


Fig. 3. Received power vs. beam ID (single peak). Numbers below the dots represent the order by which the beams are scanned. Red squares correspond to intermediate beams and the green star corresponds to the target peak.

an Rx beamformer $\mathbf{q}_i$ that generates a starting beam i in the middle of the Rx field-of-view (FoV). From $\mathbf{q}_i$, we obtain $v_i = \mathbf{q}_i^* \mathbf{g}$. Then, i's left and right adjacent beams in the azimuth domain, $i-1$ and $i+1$, are scanned and $v_{i-1} = \mathbf{q}_{i-1}^* \mathbf{g}$ and $v_{i+1} = \mathbf{q}_{i+1}^* \mathbf{g}$ are obtained. If $|v_i|^2 > |v_{i-1}|^2$ and $|v_i|^2 > |v_{i+1}|^2$, where $|\cdot|$ indicates the modulus of a complex number, then v_i is the peak. Otherwise, $\mathbf{v}$ is "rising" towards the left or the right of beam i. Thus, the directions that $\mathbf{v}$ is not rising towards no longer need to be scanned. If both directions are rising, the steepest ascent direction is selected. This leaves us with the subset of the codebook $\mathbb{Q}^{'}$, where $\mathbb{Q}^{'} = \mathbb{Q}\{1, \cdots, i-1\}$ or $\mathbb{Q}^{'} = \mathbb{Q}\{i+1, \cdots, D\}$, eliminating the need to probe the directions in $\mathbb{Q}/\{\mathbb{Q}^{'} \cup \{i\}\}$. The procedure is repeated for the remaining directions in $\mathbb{Q}^{'}$ until the peak is reached (see Algorithm 1). An example is given in Fig. 3.

If $\mathbf{v}$ exhibits a single peak, the SPF algorithm finds it in $\mathcal{O}(\log D)$ time. However, if $\mathbf{v}$ has multiple peaks, SPF finds only one of the peaks, which is not necessarily the optimal one. To cope with that, we next propose the multiple-peak finding (MPF) algorithm.

B. Multiple-peak Finding (MPF) Algorithm

To reduce the chances of getting stuck at a local maximum, in the MPF algorithm, the codebook $\mathbb{Q}$ is divided into K equal-size subsets, each representing an angular region. Here, K is selected in a way that each region is likely to contain at most one peak. Then, the SPF algorithm is executed within each region. Finally, the local maxima that are found at different regions are compared, and the largest of them is selected (see Algorithm 2). The complexity is $\mathcal{O}(K \log \frac{D}{K})$, as the SPF algorithm is executed in K regions, each of size $\frac{D}{K}$. Clearly, as K approaches D, the MPF algorithm converges to exhaustive search and the complexity approaches $\mathcal{O}(D)$. On the other hand, if K is very small (e.g., 1), the complexity is $\mathcal{O}(\log D)$, but there is a high chance that the global maximum will be missed. If the beam direction returned by the algorithm returns cannot support the link budget, a misdetection is declared; otherwise the beam is declared "suboptimal", since communications along that beam can still be established. With proper selection of K, we can increase the chances of having at most one peak in each region, which the algorithm is guaranteed to find.

Algorithm 2 Multi-peak Finding (MPF) Algorithm

procedure MULTI_PEAK($\mathbb{Q}$, $\mathbf{g}$, K)
Divide Regions:
  Divide $\mathbb{Q}$ into K equal regions R_i
Finding Peaks:
  Candidates $\Omega \leftarrow \{\ \}$
  for Each R_i, $i \in \{1, \cdots K\}$ **do**
    localMax $\leftarrow$ SINGLE_PEAK($R_i, \mathbf{g}$)
    $\Omega \leftarrow \Omega \cup$ localMax
  return $\max(\Omega)$

C. 3DPF Algorithm

We now describe how the MPF algorithm can be extended to 3D beam search. In the 3D case, each time beam direction i is scanned, four adjacent beams are also scanned: up and down in the elevation domain, and left and right in the azimuth domain. The main difference between MPF and 3DPF is that when finding the rising directions in 3D, we take measurements from these four neighbors of beam i and determine the rising quadrant (instead of the rising half of the array). Afterwards, the algorithm is run recursively in that quadrant. Similar to MPF, in 3DPF the space is divided into K angular regions to account for multiple peaks. Specifically, the horizontal domain is divided into K_x and the vertical domain is divided into K_y regions, where $K = K_x K_y$. As the algorithm still runs in logarithmic time within each region, the complexity is still $\mathcal{O}(K \log \frac{D}{K})$.

A pseudocode of 3DPF is given in Algorithm 3. In Section IV, $\mathbf{q}$, $\mathbf{f}$ and $\mathbf{v}$ were taken as vectors. In reality, when UPAs are used, these quantities are flattened matrices. For consistency, we keep using the same notation to refer to these matrices. We use $v_{i,up}$, to refer to the measurement result obtained from i's upper neighboring beam in the elevation domain. Similar definitions apply to $v_{i,left}$, $v_{i,right}$, and $v_{i,down}$. The notation for $\mathbf{q}_i$ is extended the same way to account for four different neighboring beam directions.

D. Selecting the Optimal Number of Regions

Here, we aim at identifying the value of K to be used by 3DPF so as to achieve a low misdetection probability. Recall from Theorem 1 that the peaks of the P-sparse measurement vector $\mathbf{v}$ can be found with high probability by scanning $r \geq$

Algorithm 3 3DPF Algorithm

procedure 3D_SPF($\mathbb{Q}$, $\mathbf{g}$)
Initialize:
  $i \leftarrow$ middle beam of $\mathbb{Q}$
  $v_i \leftarrow \mathbf{q}_i^* \mathbf{g}$
  $v_{i,left} \leftarrow \mathbf{q}_{i,left}^* \mathbf{g}$, $v_{i,right} \leftarrow \mathbf{q}_{i,right}^* \mathbf{g}$
  $v_{i,up} \leftarrow \mathbf{q}_{i,up}^* \mathbf{g}$, $v_{i,down} \leftarrow \mathbf{q}_{i,down}^* \mathbf{g}$
Recursion:
  if $|v_i|^2 > |v_{i,up}|^2$ **and** $|v_i|^2 > |v_{i,down}|^2$ **and** $|v_i|^2 > |v_{i,left}|^2$ **and** $|v_i|^2 > |v_{i,right}|^2$ **then**
    return $|v_i|^2$
  else if $|v_{i,down}|^2 - |v_i|^2 < |v_{i,up}|^2 - |v_i|^2$ **and** $|v_{i,left}|^2 - |v_i|^2 < |v_{i,right}|^2 - |v_i|^2$ **then**
    return 3D_SPF($\mathbb{Q}${upper-right quadrant}, $\mathbf{g}$)
  else if $|v_{i,down}|^2 - |v_i|^2 < |v_{i,up}|^2 - |v_i|^2$ **and** $|v_{i,right}|^2 - |v_i|^2 < |v_{i,left}|^2 - |v_i|^2$ **then**
    return 3D_SPF($\mathbb{Q}${upper-left quadrant}, $\mathbf{g}$)
  else if $|v_{i,up}|^2 - |v_i|^2 < |v_{i,down}|^2 - |v_i|^2$ **and** $|v_{i,left}|^2 - |v_i|^2 < |v_{i,right}|^2 - |v_i|^2$ **then**
    return 3D_SPF($\mathbb{Q}${lower-right quadrant}, $\mathbf{g}$)
  else
    return 3D_SPF($\mathbb{Q}${lower-left quadrant}, $\mathbf{g}$)
procedure 3DPF($\mathbb{Q}$, $\mathbf{g}$, K_x, K_y)
Divide Regions:
  Divide $\mathbb{Q}$ into K ($K = K_x K_y$) equal regions R_i
Finding Peaks:
  Candidates $\Omega \leftarrow \{\ \}$
  for Each R_i, $i \in \{1, \cdots K\}$ **do**
    localMax $\leftarrow$ 3D_SPF(R_i, $\mathbf{g}$)
    $\Omega \leftarrow \Omega \cup$ localMax
  return $\max(\Omega)$

$cP\log(\frac{D}{P})$ directions. Setting $K = P$ in the 3DPF algorithm yields $r = 5P\log\frac{D}{P}$ measurements, including the neighbors of a selected beam. Therefore, by letting $K = P$, the 3DPF algorithm achieves an arbitrarily low misdetection probability.

Unfortunately, there is no way to know in advance how many clusters a given environment exhibits at a specific operating frequency. However, there exists several works addressing the distribution of the number of clusters in various mmW bands (e.g., [5], [2]). In this paper, we aim at accounting for a relatively large number of clusters to lower the chances of misdetection. Suppose that P is a discrete random variable and let P^* be the 95th-percentile of P. In [5], P is randomly distributed with pmf $\max\{\text{Poisson}(\mu), 1\}$, where $\mu = 1.8$, for the 28 GHz band. Solving numerically, $P^* \approx 3.72$ when the operating frequency is 28 GHz. We can then select $K = \lceil P^* \rceil$, where $\lceil . \rceil$ is the ceiling function. Note that in [5], P is modeled for a 2D space. In our experiments, we observed that for 2D beam searching, selecting $K = 4$ achieves good performance both in terms of discovery time and misdetection probability. For 3DPF, we report the results for several K values, as discussed in Section VIII.

VI. FastLink Protocol

In this section, we present our FastLink protocol for IA in mmW systems. FastLink integrates the 3DPF algorithm as part of the message exchange between the BS and the UE. To find a suitable directional link, recent 5G specifications require that the BS covers the whole spatial area with a pre-configured number of beams, using periodically transmitted SSBs [1]. These SSBs carry a primary synchronization signal (PSS), a secondary synchronization signal (SSS), and physical broadcast channel (PBCH) information. PSS is mainly used for initial symbol boundary synchronization to the NR cell and the SSS is used for detection of cell and beam IDs. When the UE enters the coverage area of a BS, it listens to an SS burst (consisting of multiple SSBs) and measures the signal quality of different beams. It then determines the beam for which the received power is maximum (and above a predefined threshold). This beam is chosen for subsequent transmissions/receptions. After determining the best BS beam, the UE has to wait for the BS to schedule the random access channel (RACH) opportunity for the beam direction that the UE has selected. For each SSB, the BS specifies one or more RACH opportunities to occur in certain times, frequencies, and directions, so that the UE knows when to transmit the RACH preamble [9]. During a RACH opportunity, UE performs random access, implicitly informing the BS of its selected beam direction. Note that current 5G NR standards do not specify how beam sweeping will be performed at the UE.

Our 3DPF algorithm can be directly applied at the UE without changing the default 5G IA process. This reduces the search time at the UE side, but the BS still sweeps its beams exhaustively. With some small changes in the 5G IA structure, 3DPF can be employed at both the BS and the UE. In FastLink, the BS and the UE both sequentially scan a small subset of their beam directions. Specifically, the BS steers its beam towards a direction and sequentially transmits multiple copies of the same SS block. During this time, the UE constantly measures the received power, steering its receive beam according to the 3DPF algorithm. At the end of its transmission, the BS switches to Rx mode, and listens to a REPLY message from the UE along the same beam direction. UE sends its REPLY along the best beam direction it found for the given BS beam. After measuring the Rx power of the REPLY message, for the next SS block, the BS steers its beam towards another direction suggested by 3DPF, and the UE runs a new search for the new BS beam. Power measurements of REPLY messages are stored in a table at the BS. In the last stage of 3DPF, BS selects the beam with the highest power.

For a time division duplex (TDD) system, the BS and UE operate on the same frequency, and so the number of clusters they experience are the same. Thus, the optimal K for the BS and the UE are the same, i.e., $K_{\text{BS}} = K_{\text{UE}} = \lceil P^* \rceil$, as discussed in Section V-D. This allows the BS to know how long each SS block should last for (for UE to run 3DPF), without prior communication with the UE.

The transmission block structure for FastLink is shown in Fig. 4. BS sends PSS, SSS and PBCH for τ_{UE} consecutive mini-slots through a selected beam i, where τ_{UE} is the maximum required number of mini-slots for the UE to run 3DPF algorithm with the pre-selected K. Specifically, $\tau_{\text{UE}} = 5\lceil P^* \rceil \log \frac{D_{\text{UE}}}{\lceil P^* \rceil}$, where D_{UE} is the maximum number of narrow beams at the UE. Then, the UE determines the best receive beam for BS

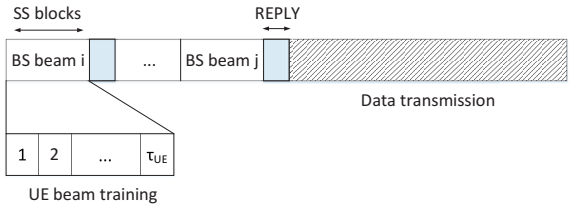


Fig. 4. Proposed transmission block structure for FastLink.

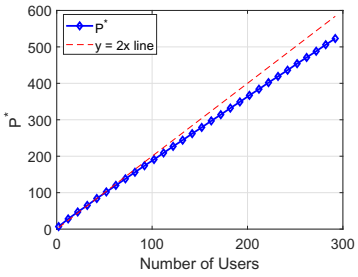


Fig. 5. Optimum K vs. the number of users. The dashed line corresponds to $y = 2x$ function, which is shown here as a baseline.

transmit beam i and sends a REPLY message to the BS. However, the received power of the REPLY signal at the BS can be lower than its sensitivity. If the BS does not receive a REPLY message from the UE after τ_{UE} mini-slots, it stores a minimum power value, ζ, in its table for beam i. Then, the BS continues executing 3DPF, selecting the next transmit beam suggested by the algorithm according to the received power value. In total, BS needs to scan at most $\tau_{\text{BS}} = 5\lceil P^* \rceil \log \frac{D_{\text{BS}}}{\lceil P^* \rceil}$ beams, where D_{BS} is the maximum number of narrow beams at the BS. After collecting all REPLY messages from the UE for the selected beams, the BS finally selects the best transmit beam from the table and the UE selects the best receive beam for the given transmit beam. This way, the number of slots required to establish a directional link can be reduced from $D_{\text{BS}}D_{\text{UE}}$ to $\tau_{\text{BS}}\tau_{\text{UE}}$ ($\tau_{\text{BS}} \ll D_{\text{BS}}$ and $\tau_{\text{UE}} \ll D_{\text{UE}}$). To account for the transmission delay of the REPLY messages, we introduce a parameter δ, which represents the time required for the BS and UE to switch from Rx to Tx mode, as well as the time spent over-the-air. Accordingly, the time to establish a directional link using FastLink can be written as $\tau_{\text{BS}}\tau_{\text{UE}} + \delta\tau_{\text{BS}} = \tau_{\text{BS}}(\tau_{\text{UE}} + \delta)$. Since δ is typically less than 10 μs [33], [34], the feedback delay is rather small compared to the overall duration of the IA.

VII. Extension to Multiple UEs

In this section, we analyze the effect of multiple UEs on the network discovery time. Note that in 5G IA, discovery time scales with $D_{\text{BS}}D_{\text{UE}}$, regardless of the number of users, as the parties perform an exhaustive beam sweep. Specifically, each UE sweeps through its set of beams for each BS beam, and finds the BS beam through which it experienced the highest receive power. Then, the UE waits for the BS to schedule the RACH opportunity for the beam direction that it selected. Eventually, multiple UEs can be discovered by the BS with a single exhaustive sweep. In contrast, in FastLink, as the channel between the BS and each user is different, the BS needs to probe a different set of beams for each user. Thus, in the worst case, the discovery time scales with $D_{\text{BS}}\tau_{\text{UE}}$, as the BS performs an exhaustive scan to discover all users within a single scan cycle. We analyze the effect of the number of users on the network discovery time. Note that the time for a UE to find the best Rx beam for the given BS beam remains the same in the multi-user case, as the UE still aims at discovering a single BS.

Assume that the network consists of U UEs and a single BS, and denote the set of users as $\mathcal{U}$. Let p_i denote the number of clusters between the BS and UE i, and let P be the total number of clusters; $\sum_{i=1}^{U} p_i = P$. Here, we use the statistical channel model in [5], where $p_i \sim \max\{\text{Poisson}(\mu), 1\}, \forall i \in \mathcal{U}$, with $\mu = 1.8$ for the 28 GHz band. For each $i \in \mathcal{U}$, let $\chi_{1,i}, \chi_{2,i}, \cdots, \chi_{U,i}$ be Poisson random variables with means $\mu, 2\mu, U\mu$, respectively. Then, $p_i = \max\{\chi_{1,i}, 1\}, \forall i \in \mathcal{U}$. Define E_0 as the event of having $\chi_{1,i} < 1, \forall i \in \mathcal{U}$, E_1 as the event of having one user with $\chi_{1,i} \geq 1$ and other users with $\chi_{1,j} < 1$, and so on. In general, E_k is defined as the event of having k users with $\chi_{1,i} \geq 1$ and $U - k$ users with $\chi_{1,j} < 1$. We can then express P by the function:

$$P = \begin{cases} U, & \text{if } E_0 \text{ occurs} \\ \chi_{1,i} + U - 1, & \text{if } E_1 \text{ occurs} \\ \chi_{2,i} + U - 2, & \text{if } E_2 \text{ occurs} \\ \vdots & \\ \chi_{U,i}, & \text{if } E_U \text{ occurs.} \end{cases} \tag{7}$$

Now, let $x_0, x_1, \cdots, x_U$ denote the probabilities of the occurance of events $E_0, E_1, \cdots, E_U$, respectively. Naturally, $\sum_{k=0}^{U} x_k = 1$. These probabilities can simply be written as:

$$x_k = \binom{U}{k} \Pr(\chi_{1,i} = 0)^{U-k} (1 - \Pr(\chi_{1,i} = 0))^k. \tag{8}$$

To numerically find an optimal K value, we follow a similar approach to that in SectionV-D. Let P^* be the 95th-percentile of P. Then:

$$\Pr(P > P^*) = \sum_{k=0}^{U} x_k \Pr(P > P^* | E_k \text{ occurs}) = 0.05 \tag{9}$$

As the system consists of U users, each of which having at least one cluster, $P^* \geq U$. Then, the terms in (9) can be simplified as follows:

$$\begin{aligned} \Pr(P > P^* | E_0 \text{ occurs}) &= \Pr(P^* = U) \\ \Pr(P > P^* | E_1 \text{ occurs}) &= \Pr(\chi_{1,i} > P^* + 1 - U) \\ \Pr(P > P^* | E_2 \text{ occurs}) &= \Pr(\chi_{2,i} > P^* + 2 - U) \\ &\vdots \\ \Pr(P > P^* | E_U \text{ occurs}) &= \Pr(\chi_{U,i} > P^*) \end{aligned} \tag{10}$$

where the probabilities in the right hand side can be computed easily as $\chi_{1,i}, \chi_{2,i}, \cdots, \chi_{U,i}$ have generic Poisson distributions $\forall i \in \{1, \cdots, U\}$. Note that because P^* and U are both constants, $\Pr(P^* = U)$ is either 1 or 0. There is no closed-form solution for the P^*. However, for a given U, one can calculate P^* using (8), (9) and (10) (see Fig. 5).

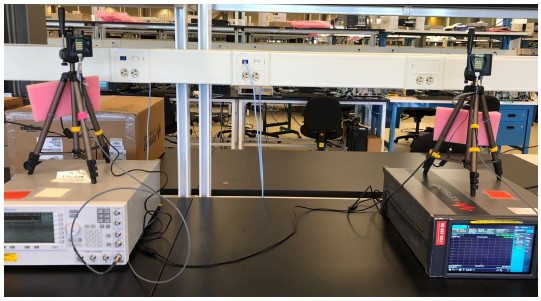
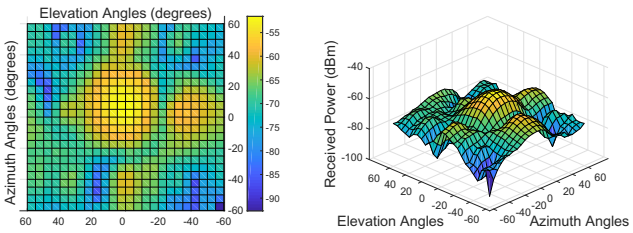


Fig. 6. Test setup with 4×4 UPAs at the Tx and the Rx, signal generator and vector signal analyzer (operating on the 29 GHz band).



(a) (b)

Fig. 7. Rx powers for different beam directions at 29 GHz frequency and a Tx-Rx separation of 3 m. Measured noise power is -95 dBm. (a) Flattened, (b) 3D representation.

VIII. Performance Evaluation

In this section, we evaluate the performance of 3DPF through over-the-air hardware experiments and computer simulations. We compare 3DPF with the 802.11ad beam search approach, in which the worst-case search time scales linearly with D. Note that current 5G NR beam search also scales linearly with D, attaining similar results to 802.11ad.

A. Experimental Results

We first conduct extensive experiments to verify the efficiency of 3DPF. 4×4 UPAs are used in our experiments with $d_x = 0.5\lambda$ and $d_y = 0.6\lambda$. The number of antenna elements and their spacing determine the beamwidth of the patterns generated by the UPAs. However, the beam steering resolution and the maximum number of narrow beams that can be generated (i.e., the codebook size) depend on the resolution of the phase shifters used in the antenna chip. The UPAs used in our experiments are based on 6-bit phase shifters, which allow us to sweep the beam with up to 1° resolution. In this case, neighboring beams overlap, since the beamwidth of the broadside pattern generated by our antennas is 26.5°. For the sake of taking RSS measurements, a continuous wave (CW) signal with 5 dBm amplitude is transmitted over the 29 GHz band, one of the candidate band for 5G NR. To generate the CW signal, Keysight E8257D-ATO-8384 PSG signal generator is used. At the Rx side, the array is connected to Keysight PXA-550-MY55002004 vector signal analyzer (VSA). The PSG and the VSA are connected to a host PC, and the RSS results are obtained via a TCP connection. To steer the transmit/receive beams to desired directions, antenna arrays are connected to microcontrollers, which are interfaced with the host PC through the serial port. The test setup with the Tx, Rx, PSG, and the VSA is shown in Fig. 6.

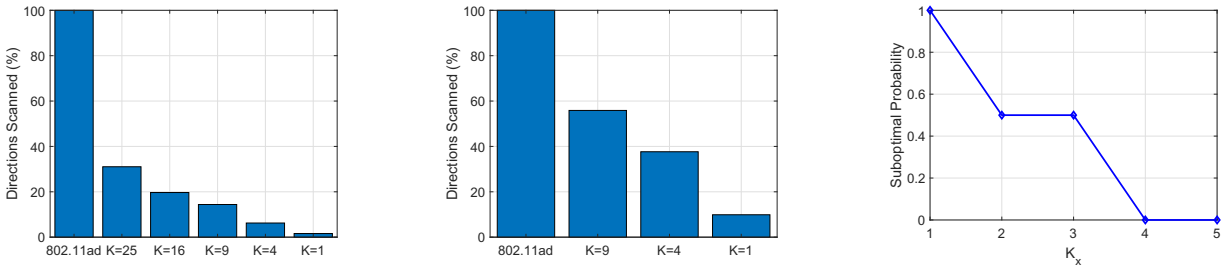
In our experiments, Tx direction is fixed and the best Rx direction for that fixed Tx direction is determined using either 802.11ad beam scan or the 3DPF algorithm. We test several LOS and NLOS scenarios with a Tx-Rx separation of 3 m, where the NLOS path is created by a 1.2 m $\times$ 1.2 m metal reflector. The effective beam scanning range of the single-panel UPAs in our experiments is $\pm 60^\circ$ from broadside in both azimuth and elevation. This region can be considered as one quasi-omni beam for the sector-level scanning phase of 802.11ad. Thus, in the beam refinement phase, 802.11ad protocol exhaustively scans all narrow beams in this region.

To obtain the AoA profile, we exhaustively scanned the 3D space under different scenarios. Using the above setup, we collected RSS measurements within the effective beam scanning range of the UPAs (see Fig. 7). In addition to the exhaustive search, we also implemented our 3DPF algorithm on the Rx side. Next, we compare 3DPF with 802.11ad/exhaustive beam search approach and plot the percentage of the scanned beams. We varied K to study its effect on the search time overhead. In Fig. 8(a), we show the results for an angular step size of 5° ($D = 625$). Fig. 8(b) depicts the results for a larger step size of 15° ($D = 81$). Note that in Fig. 8(a), the search space is large, so the relative overhead of scanning the neighbors (upper, lower, left and right neighboring beams) under 3DPF is negligible. Overall, 3DPF scans less than 30% of all the beams, even when K is as high as 25. However, when the search space is smaller, the overhead is more pronounced. Nevertheless, 3DPF scans only 57% of the space, as shown in Fig. 8(a), for $K = 9$. Beyond this K value, the performance of 3DPF converges to the exhaustive search.

As explained in Section V-B, if the beam that 3DPF returns cannot support the link budget, a misdetection is declared; otherwise, if the beam does not yield the highest Rx power but it can support communications, the beam is declared "suboptimal". Due to short Tx-Rx separation, no misdetection was observed in any of our experiments. However, the channel between the Tx and the Rx changes slightly between different runs of the same scenario. We assumed that deviations that are within 5° from the optimal beam do not cause a suboptimal selection. However, larger deviations are declared as suboptimal. The probability of finding a suboptimal beam vs. K_x ($= K_y = \sqrt{K}$) is shown in Fig. 8(c).

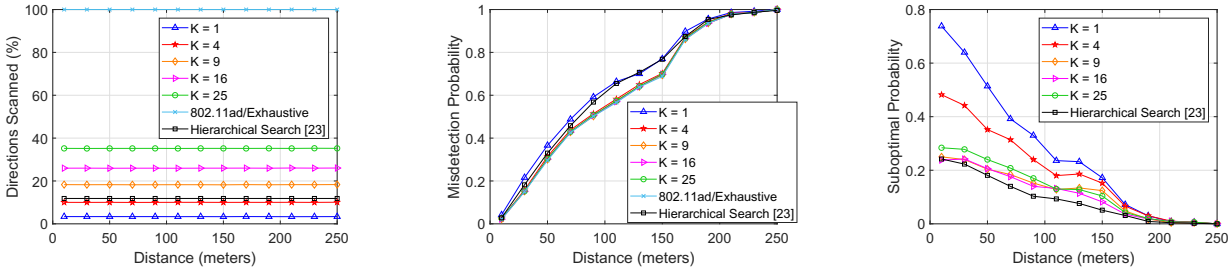
B. Simulation Results

We use simulations to study the impact of Tx-Rx distance. In addition, implementing a hierarchical codebook in our experiments was not possible using our antenna controllers. We have, however, simulated the hierarchical codebook of [23] and compared its performance with our proposed approach. In our simulations, Tx power is set to $P_{\text{Tx}} = 30$ dBm and $A_{\text{BS}} = A_{\text{UE}} = 16$. We run the simulations for the 28 GHz band with UPAs ($d_x = d_y = \lambda/2$) using [5] to model large-scale effects. For small-scale effects, we place random scatterers on an ellipsoid between the Tx and the Rx. The Tx beam is kept the same, and the Rx sweeps its beam in 5° steps to



(a) (b) (c)

Fig. 8. Experimental comparison of the 3DPF algorithm and 802.11ad/exhaustive beam scan. (a) Percentage of scanned beams for 5° resolution, (b) percentage of scanned beams for 15° resolution, (c) suboptimal probability vs K_x.



(a) (b) (c)

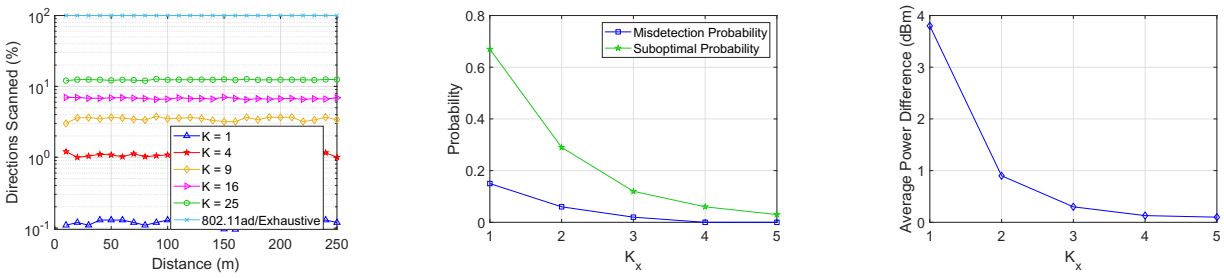
Fig. 9. Simulation-based comparison between 3DPF, 802.11ad, and the hiearchical search in [23]. (a) Directions scanned vs. distance, (b) misdetection probability vs. distance, (c) suboptimal probability vs. distance.

find the best receive direction for the given Tx beam. The beam scanning range is kept the same as in the hardware experiments. The Tx-Rx separation varies between 10 m and 250 m and the results are averaged over 1000 runs.

As shown in Fig. 9(a), for all the considered values of K, the number of scanned directions in 3DPF is significantly smaller than that of 802.11ad. Even when K is as large as 25, 3DPF scans less than 35% of the directions scanned by the exhaustive approach. In Fig. 9(b), we depict the misdetection probability vs. Tx-Rx distance. Except when $K = 1$, the difference in the misdetection probability between 3DPF and 802.11ad is negligible. Note that the discovery time of the hiearchical search scheme in [23] is slightly higher than 3DPF with $K = 4$. However, the misdetection proability of the 3DPF is also about 10% lower than the scheme in [23] for distances between 75 m and 175 m. Therefore, 3DPF can discover UEs more accurately than the hierarchical search approach, especially near the cell edge. Finally in 9(c), suboptimal probability is plotted vs. the distance. Although 3DPF exhibits a high suboptimal probability when K is small, this probability decreases significantly when $K \geq 9$, as a result of the misdetection probability becoming more dominant.

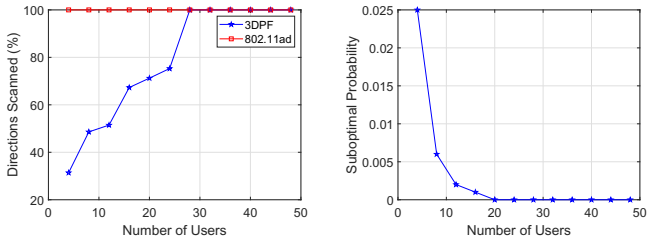
We also implemented the FastLink protocol in our simulation environment. Because FastLink executes 3DPF at both the BS and the UE, its benefit in terms of discovery time are even more pronounced. Specifically, the values observed in Fig. 10(a) are simply the squares of the values in Fig. 9(a) for the corresponding K. In Fig. 10(b), we depict the misdetection and suboptimal probabilities, both of which decrease with increasing K. As K increases, the search space is divided into more sub-regions, meaning that it is less likely to miss one or more peaks. In Fig. 10(c), the average Rx power difference between FastLink and the 802.11ad search scheme is shown for the cases when FastLink returns a suboptimal peak. For all K values, this difference is less than 4 dBm, and the gap decays exponentially fast as K increases.

Next, we compare 802.11ad and 3DPF under multiple UEs and depict the results in Fig. 11. The misdetection probabilities (not shown) are independent of the number of users and remain almost the same. Fig. 11(a) shows the percentage of scanned directions, which increases slowly with the number of users, eventually hitting 100% as the number of users reaches 30. The rate of increase is indicative of the scalability of the beam searching algorithm, and is an important design factor [35]. For 3DPF, even when 20 users are simultaneously trying to connect to a BS within the same IA cycle, the algorithm still offers 30% reduction in search time compared to exhaustive search. Note that in several 5G NR and WiGig use cases (e.g., indoor and outdoor hotspots), coverage radius in tens of meters [36]. Thus, the number of users simultaneously trying to connect to a BS within the same IA cycle is typically small [37]. When multiple UEs are to be found, the performance of hierarchical beam search also approaches that of the exhaustive search. This is due to the fact that multiple quasi-omni sectors that receive relatively strong signals in the first stage need to be scanned with narrow beams in the second stage. Fig. 11(b) shows the suboptimal probability which decays exponentially fast as the number of users increase, since the more directions scanned the less likely it is to miss the global maximum.



(a) (b) (c)

Fig. 10. Simulations of FastLink and 802.11ad search. (a) Directions scanned vs. distance, (b) misdetection and suboptimal probabilities vs. $K_x (= K_y)$, (c) average Rx power difference between FastLink and 802.11ad.



(a) (b)

Fig. 11. Simulation-based comparison between 3DPF and 802.11ad search in the multi-user case. (a) directions scanned vs. the number of users, (b) suboptimal probability vs. the number of users.

IX. Discussion

In this section, we discuss some applications of the 3DPF to other beamforming architectures and network scenarios. Specifically, we explain how we can exploit the availability of multiple RF chains (e.g., hybrid beamforming). We also examine how a multi-BS network such as coordinated multi-point (CoMP) can benefit from the 3DPF algorithm.

The main objective of hybrid beamforming is to properly partition the signal processing between the RF and digital domains so as to attain similar performance to fully digital beamforming. FastLink does not require the availability of multiple RF chains. However, if the Rx has multiple RF chains, 3DPF can simultaneously take measurements from multiple directions. Specifically, if the Rx contains B RF chains, it can simultaneously run the SPF algorithm in parallel in B mutually-exclusive regions, achieving a B-fold decrease in the search time. 3DPF can also be applied to a network that utilizes CoMP, one of the promising concepts to improve cell edge user data rate and spectral efficiency. The basic idea is that transmissions from one or more geographically separated BSs to a single UE are dynamically coordinated in a way to improve average and cell edge throughput. 3GPP defines two different categories of downlink CoMP: coordinated scheduling/beamforming and joint processing CoMP [38]. A joint processing CoMP architecture is very similar to the system model we considered in Section VII. Instead of having a single receiving BS and multiple transmitting UEs, CoMP scenarios mainly deal with a single receiving UE and multiple transmitting BSs. If the UE knows in advance how many BSs will transmit SSBs to itself, it can choose a suitable K to run the 3DPF algorithm with, as discussed in Section VII. In reality, a UE can be in the coverage area of only a few BSs. Therefore, it can still select a relatively small K and run 3DPF to establish communications with multiple BSs.

X. Summary

In this paper, we proposed FastLink, an efficient beam finding protocol for mmW systems. FastLink always transmits using the narrowest possible beams, allowing high beamforming gains and low misdetection rate. We first formulated the beam finding problem as a sparse problem considering a single BS-UE pair, and used CS to determine the minimum number of measurements needed to reconstruct the peaks of the sparse channel. Using CS analysis, we designed the 3DPF algorithm and showed that it significantly reduces the search latency compared to existing search schemes. 3DPF was incorporated into an IA protocol called FastLink to allow the BS and the UE to establish a directional link. We then extended the design into multiple users and showed how the discovery time scales with the number of users. Finally, experimental and simulation results were provided to verify the efficiency of FastLink.

Acknowledgments

The authors would like to thank Keysight Technologies for making their mmW experimental setup available. We would also like to thank Dean Gienger and Andrew Smail from Keysight for providing technical expertise regarding RF measurements and help building the setup. This research was supported by NSF (grants IIP-1822071, CNS-1513649, CNS-1409172, and CNS- 1731164). Any opinions, findings, conclusions, or recommendations expressed in this paper are those of the author(s) and do not necessarily reflect the views of NSF.

References

[1] 3GPP TR 38.802 v14.2.0, "Study on new radio access technology-physical layer aspects (release 14)," Sep. 2017. [Online]. Available: http://www.3gpp.org/ftp//Specs/archive/38-series/38.802/38802-e20.zip

[2] IEEE Computer Society, "IEEE Standard-part 11: Wireless LAN medium access control (MAC) and physical layer (PHY) specifications amendment 3: Enhancements for very high throughput in the 60 GHz band (adoption of IEEE std 802.11ad-2012)," 2014. [Online]. Available: https://standards.ieee.org/findstds/standard/802.11ad-2012.html

[3] J. G. Andrews, S. Buzzi, W. Choi, S. V. Hanly, A. Lozano, A. C. Soong, and J. C. Zhang, "What will 5G be?" *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1065–1082, 2014.

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

[5] M. R. Akdeniz, Y. Liu, M. K. Samimi, S. Sun, S. Rangan, T. S. Rappaport, and E. Erkip, "Millimeter wave channel modeling and cellular capacity evaluation," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1164–1179, 2014.

[6] 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, 2013.

[7] M. Giordani, M. Mezzavilla, C. N. Barati, S. Rangan, and M. Zorzi, "Comparative analysis of initial access techniques in 5G mmWave cellular networks," in *Proc. of the IEEE CISS Conference*, Princeton, NJ, Mar. 2016, pp. 268–273.

[8] A. Abdelreheem, E. M. Mohamed, and H. Esmaiel, "Location-based millimeter wave multi-level beamforming using compressive sensing," *IEEE Commun. Lett.*, vol. 22, no. 1, pp. 185–188, 2018.

[9] M. Giordani, M. Polese, A. Roy, D. Castor, and M. Zorzi, "A tutorial on beam management for 3GPP NR at mmWave frequencies," *IEEE Communications Surveys & Tutorials*, 2018.

[10] O. Abari, H. Hassanieh, M. Rodriguez, and D. Katabi, "Millimeter wave communications: From point-to-point links to agile network connections," in *Proc. of the 15th ACM Workshop on Hot Topics in Networks*, Atlanta, GA, Nov. 2016, pp. 169–175.

[11] I. Aykin, B. Akgun, and M. Krunz, "Smartlink: Exploiting channel clustering effects for reliable millimeter wave communications," in *Proc. of the IEEE INFOCOM Conference*, Paris, France, Apr. 2019, pp. 1117–1125.

[12] ——, "Multi-beam transmissions for blockage resilience and reliability in millimeter-wave systems," *IEEE Journal on Selected Areas in Communications*, 2019. [Online]. Available: http://dx.doi.org/10.1109/JSAC.2019.2947938

[13] M. Hashemi, C. E. Koksal, and N. B. Shroff, "Out-of-band millimeter wave beamforming and communications to achieve low latency and high energy efficiency in 5G systems," *IEEE Trans. Commun.*, vol. 66, no. 2, pp. 875–888, 2018.

[14] A. Ali, N. González-Prelcic, and R. W. Heath, "Millimeter wave beam-selection using out-of-band spatial information," *IEEE Trans. Wireless Commun.*, vol. 17, no. 2, pp. 1038–1052, 2018.

[15] X. Lin, S. Wu, C. Jiang, L. Kuang, J. Yan, and L. Hanzo, "Estimation of broadband multiuser millimeter wave massive MIMO-OFDM channels by exploiting their sparse structure," *IEEE Trans. Wireless Commun.*, vol. 17, no. 6, pp. 3959–3973, 2018.

[16] M. Hashemi, A. Sabharwal, C. E. Koksal, and N. B. Shroff, "Efficient beam alignment in millimeter wave systems using contextual bandits," in *Proc. of the IEEE INFOCOM Conference*, Honolulu, HI, Apr. 2018, pp. 2393–2401.

[17] A. Zhou, L. Wu, S. Xu, H. Ma, T. Wei, and X. Zhang, "Following the shadow: Agile 3-D beam-steering for 60 GHz wireless networks," in *Proc. of the IEEE INFOCOM Conference*, Honolulu, HI, Apr. 2018, pp. 2375–2383.

[18] A. Alkhateeb, G. Leusz, and R. W. Heath, "Compressed sensing based multi-user millimeter wave systems: How many measurements are needed?" in *Proc. of the IEEE ICASSP Conference*, South Brisbane, Australia, Apr. 2015, pp. 2909–2913.

[19] J. Choi, "Beam selection in mm-Wave multiuser MIMO systems using compressive sensing," *IEEE Trans. Commun.*, vol. 63, no. 8, pp. 2936–2947, 2015.

[20] W. U. Bajwa, J. Haupt, A. M. Sayeed, and R. Nowak, "Compressed channel sensing: A new approach to estimating sparse multipath channels," *Proc. of the IEEE*, vol. 98, no. 6, pp. 1058–1076, 2010.

[21] D. Steinmetzer, D. Wegemer, M. Schulz, J. Widmer, and M. Hollick, "Compressive millimeter-wave sector selection in off-the-shelf IEEE 802.11ad devices," in *Proc. of the 13th ACM CoNEXT Conference*, Seoul, South Korea, Dec. 2017, pp. 414–425.

[22] X. Cheng, M. Wang, and S. Li, "Compressive sensing-based beamforming for millimeter-wave OFDM systems," *IEEE Trans. Commun.*, vol. 65, no. 1, pp. 371–386, 2016.

[23] A. Alkhateeb, O. El Ayach, G. Leus, and R. W. Heath, "Channel estimation and hybrid precoding for millimeter wave cellular systems," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 831–846, 2014.

[24] Z. Xiao, H. Dong, L. Bai, P. Xia, and X.-G. Xia, "Enhanced channel estimation and codebook design for millimeter-wave communication," *IEEE Trans. Veh. Technol.*, vol. 67, no. 10, pp. 9393–9405, 2018.

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

[26] I. Aykin and M. Krunz, "Fastlink: An efficient initial access protocol for millimeter wave systems," in *Proc. of the 21st ACM MSWiM Conference*, Montreal, CA, Oct. 2018, pp. 109–117.

[27] X. Song, T. Kühne, and G. Caire, "Fully-/partially-connected hybrid beamforming architectures for mmWave MU-MIMO," *arXiv preprint arXiv:1904.10276*, 2019.

[28] E. J. Candes, "The restricted isometry property and its implications for compressed sensing," *Comptes rendus mathematique*, vol. 346, no. 9-10, pp. 589–592, 2008.

[29] E. J. Candès and M. B. Wakin, "An introduction to compressive sampling [a sensing/sampling paradigm that goes against the common knowledge in data acquisition]," *IEEE Signal Processing Mag.*, vol. 25, no. 2, pp. 21–30, 2008.

[30] E. J. Candes and T. Tao, "Decoding by linear programming," *IEEE Trans. Inf. Theory*, vol. 51, no. 12, pp. 4203–4215, 2005.

[31] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin, "A simple proof of the restricted isometry property for random matrices," *Constructive Approximation*, vol. 28, no. 3, pp. 253–263, 2008.

[32] E. Candes and J. Romberg, "Sparsity and incoherence in compressive sampling," *Inverse problems*, vol. 23, no. 3, p. 969, 2007.

[33] M. O. Al-Kadri, A. Aijaz, and A. Nallanathan, "An energy-efficient full-duplex MAC protocol for distributed wireless networks," *IEEE Wireless Communications Letters*, vol. 5, no. 1, pp. 44–47, 2015.

[34] K. Medepalli, P. Gopalakrishnan, D. Famolari, and T. Kodama, "Voice capacity of IEEE 802.11 b, 802.11 a and 802.11 g wireless LANs," in *Proc. of the IEEE GLOBECOM Conference*, vol. 3, Dallas, TX, Nov. 2004, pp. 1549–1553.

[35] X. Song, S. Haghighatshoar, and G. Caire, "A scalable and statistically robust beam alignment technique for millimeter-wave systems," *IEEE Trans. on Wireless Commun.*, vol. 17, no. 7, pp. 4792–4805, 2018.

[36] M. Shafi, A. F. Molisch, P. J. Smith, T. Haustein, P. Zhu, P. De Silva, F. Tufvesson, A. Benjebbour, and G. Wunder, "5G: A tutorial overview of standards, trials, challenges, deployment, and practice," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 6, pp. 1201–1221, 2017.

[37] C. N. Barati, S. A. Hosseini, M. Mezzavilla, T. Korakis, S. S. Panwar, S. Rangan, and M. Zorzi, "Initial access in millimeter wave cellular systems," *IEEE Trans. Wireless Commun.*, vol. 15, no. 12, pp. 7926–7940, 2016.

[38] 3GPP TR 36.814 v9.2.0, "Further advancements for E-UTRA physical layer aspects (release 9)," Mar. 2017. [Online]. Available: http://www.3gpp.org/ftp//Specs/archive/36-series/36.814/36814-920.zip