



## **Genetic Algorithm-Based Beam Refinement for Initial Access in Millimeter Wave Mobile Networks**

Downloaded from: <https://research.chalmers.se>, 2024-03-13 10:04 UTC

Citation for the original published paper (version of record):

Guo, H., Makki, B., Svensson, T. (2018). Genetic Algorithm-Based Beam Refinement for Initial Access in Millimeter Wave Mobile Networks. *Wireless Communications and Mobile Computing*, 2018. <http://dx.doi.org/10.1155/2018/5817120>

N.B. When citing this work, cite the original published paper.

## Research Article

# Genetic Algorithm-Based Beam Refinement for Initial Access in Millimeter Wave Mobile Networks

Hao Guo , Behrooz Makki, and Tommy Svensson 

*Department of Electrical Engineering, Chalmers University of Technology, Gothenburg 41258, Sweden*

Correspondence should be addressed to Hao Guo; hao.guo@chalmers.se

Received 29 December 2017; Accepted 12 April 2018; Published 4 June 2018

Academic Editor: Shao-Yu Lien

Copyright © 2018 Hao Guo et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Initial access (IA) is identified as a key challenge for the upcoming 5G mobile communication system operating at high carrier frequencies, and several techniques are currently being proposed. In this paper, we extend our previously proposed efficient genetic algorithm- (GA-) based beam refinement scheme to include beamforming at both the transmitter and the receiver and compare the performance with alternative approaches in the millimeter wave multiuser multiple-input-multiple-output (MU-MIMO) networks. Taking the millimeter wave communications characteristics and various metrics into account, we investigate the effect of different parameters such as the number of transmit antennas/users/per-user receive antennas, beamforming resolutions, and hardware impairments on the system performance employing different beam refinement algorithms. As shown, our proposed GA-based approach performs well in delay-constrained networks with multiantenna users. Compared to the considered state-of-the-art schemes, our method reaches the highest service outage-constrained end-to-end throughput with considerably less implementation complexity. Moreover, taking the users' mobility into account, our GA-based approach can remarkably reduce the beam refinement delay at low/moderate speeds when the spatial correlation is taken into account. Finally, we compare the cases of collaborative users and noncollaborative users and evaluate their difference in system performance.

## 1. Introduction

The next generation of cellular systems (5G) requires both higher data rates (in the order of 10-100 Gbps) and lower end-to-end latencies (down to 1 ms) than previous generations [1]. For this reason, it is aimed at utilizing frequency bands in the 30-300 GHz range in order to obtain sufficiently large bandwidths/data rates. Due to power limitation and high path loss at these frequencies, the coverage range is typically small so that highly directional transmissions are required for such millimeter wave (MMW) communications. On the other hand, the physical size of antennas at the MMW band is relatively small, such that large-scale beamforming can be performed in practice [2, 3]. Employing large-scale beamforming during the initial access (IA) procedure can be a good way to overcome the increased path loss experienced at higher frequencies (see Section 2 for literature review of the IA systems).

One of the most challenging tasks of IA is that the base stations (BSs) make omnidirectional cell searches with directional beams and at the receiver side the users choose

their best beam direction to detect the BSs. Successful access means, e.g., that the received power or the signal-to-noise ratio (SNR) is beyond certain thresholds. After a basic connection is established, the BSs and the users can begin exchanging messages and implement a beam refinement procedure to further improve the beam directions and do additional control actions [4].

For example, the user mobility can be handled by beam refinement. With 5G, it is expected to access wireless networks not only at home or in the office, but also at moving speeds such as in a vehicle. In the moving scenario, the beam refinement process can keep tracking the beams by exploiting spacial correlations so that the computational delay can be remarkably reduced. Furthermore, for vehicular user equipment (VUEs), the system-level performance is improved if we allow a scheme using device-to-device (D2D) communications to enhance the links [5].

IA beamforming at MMW is different from the conventional one since it is hard to acquire the channel state information (CSI) at these frequencies. For this reason, codebook-based beamforming has been recently proposed

as an efficient method to reduce the dependency on CSI estimation/feedback [6, 7]. Also, several works have been presented on both physical layer and procedural algorithms of IA beamforming [8–15]. However, in these works either the algorithms are designed for special metrics, precoding/combining schemes, and channel models or the implementation complexity grows significantly by an increasing number of BSs/users. Moreover, the running delay of the algorithm has been rarely considered in the performance evaluation. On the other hand, generic machine learning-based schemes have been recently proposed which can be effectively applied for different channel models with acceptable implementation complexity [6, 7, 16–18].

In this paper, we study the effect of beam refinement on the performance of MMW networks. In our previous work, we proposed an efficient genetic algorithm- (GA-) based beamforming approach [18] which reaches almost the same performance compared to the exhaustive search with low complexity. Based on [18], the contributions of this paper are as follows. (1) We include the GA-based beam refinement at both the transmitter and the receiver side. Also, (2) we compare different machine learning-based analog beamforming approaches for the beam refinement during IA, including GA-based beamforming [18], Tabu search beamforming [16], link-by-link beamforming [17], and two-level codebook beamforming [6, 7] in large-but-finite multiuser multiple-input-multiple-output (MU-MIMO) MMW communication systems. Moreover, (3) we analyze the effect of various parameters such as the number of transmit/receive antennas, total power budget, and the power amplifier (PA) efficiency on the network performance. As opposed to the literature, we take the algorithm running delay into account. Thus, there is a trade-off between finding the optimal beamforming matrices and reducing the data transmission time slot, and the highest throughput may be achieved by few iterations. We study the system performance in terms of the end-to-end throughput with service outage constraints as well as the implementation complexity. (4) Furthermore, we evaluate and compare the performance of the considered algorithms under various mobile speed of the users. (5) Finally, we consider the case of collaborative users and compare the system performance in the cases with and without information exchanges among users.

Our results demonstrate that the running delay of the algorithms and power amplifier inefficiency affect the system performance remarkably, which should be carefully considered in the system design. Moreover, our proposed GA-based approach outperforms the considered state-of-the-art schemes, in terms of throughput, and reaches (almost) the same results as in the exhaustive search-based approach with fewer number of iterations. Furthermore, when taking the user mobility into account, the GA-based approach can remarkably reduce the algorithm running delay based on the beamforming results in the previous time slots. With collaborative users, the end-to-end throughput can be improved due to the data exchange by D2D links. Thus, the GA-based beamforming approach can be an appropriate candidate for IA in future wireless networks.

## 2. Literature Review

In this part, we present some related research work on IA. The reader familiar with the research area can skip this section and go to Sections 3–5 where we present the system model, the algorithm descriptions, and the simulation results, respectively.

Beamforming techniques at MMW bands have been considered in standard developments IEEE 802.15.3c (TG3c) [19], IEEE 802.11ad (TGad) [20], and ECMA-387 [21]. The problem formulation for IA beamforming at MMW frequencies is introduced in [8] where a fast-discovery hierarchical search method is proposed. Moreover, several design options for MMW IA are presented in [22], where the basic steps in the 3rd-Generation Partnership Project (3GPP) Long-Term Evolution (LTE) standard are used as references, and the overall delay of each design option as a function of the system overhead is evaluated. Then, [11] compares three approaches, namely, exhaustive search, two-step, and context information-based, in terms of miss-detection probability and discovery time. Another comparison work is presented in [12], where it is shown that different IA protocols have a trade-off between delay and average user-perceived throughput.

In [18], we introduce a genetic algorithm-based initial beamforming approach and evaluate the effect of the algorithm running delay on the network performance. There are also previous works using the GA-based selection approach in different communication networks. For instance, in [23] an efficient scheduling scheme is designed based on the genetic algorithm in the return-link of a multibeam satellite system. A turbo-like beamforming scheme based on the Tabu search algorithm is proposed in [16] to reduce both searching complexity and system overhead. A concurrent beamforming protocol, which we refer to as link-by-link beamforming, is presented in [17] to achieve high capacity in indoor MMW networks. Finally, for multistage beamforming, a tree-structured multilevel beamforming codebook is designed for MMW wireless backhaul systems in [6]. Also, in [7], a low-complexity multistage codebook is designed to support the IEEE 802.15.3c protocol. In [9], an exhaustive beam search method is proposed. Two beamforming schemes, namely, random-phase beamforming and directional beamforming, have been tested in [10] under the line-of-sight (LOS) channel conditions. A low-complexity beamforming scheme for initial user discovery is proposed in [13] where limited feedback-type codebooks are used. In [14], an accurate analytical framework for MMW system performance has been developed. Impact of obstacles on the cell search process is considered in [15] for the first time, and a geo-located context database is proposed to speed up the cellular attachment operations by storing and processing the information about the previous cell discovery attempts.

## 3. System Model

We consider a MU-MIMO setup with  $M$  transmit antennas at a BS and  $\tau$  multi-antenna VUEs, each with  $\beta$  antennas. As a result, there are  $N = \tau \times \beta$  total antennas at the receiver side (see Figure 1). This is an extension of our work [18] with single

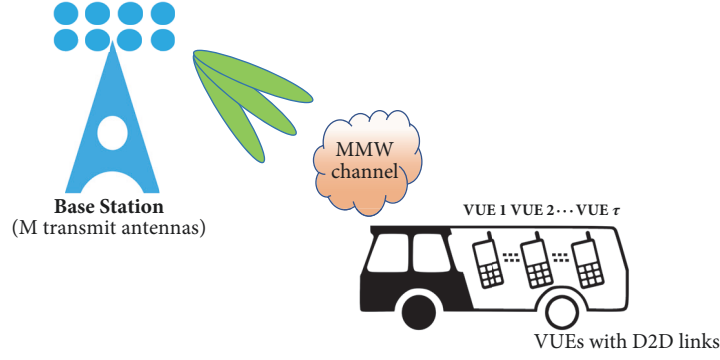


FIGURE 1: An illustration of MMW cooperative downlink communication of VUEs.

receive antennas and allows for beamforming at the receiver side. We assume that each user has perfect CSI. Also, as a more explicit model compared with [24], VUEs are allowed to exchange data with each other by using D2D links which is similar to the model in, e.g., [5]. We set  $M > N$ . At each time slot  $t$ , the aggregated received signal vector  $\mathbf{y}(t)$  at time  $t$  over the users after receive beamforming can be described as

$$\mathbf{y}(t) = \sqrt{\frac{P}{M}} \mathbf{U}(t)^H \mathbf{H}(t) \mathbf{V}(t) \mathbf{x}(t) + \mathbf{z}(t), \quad (1)$$

where  $P$  is the total power budget,  $\mathbf{H}(t) \in \mathbb{C}^{N \times M}$  is the channel matrix with the  $(i, j)$ th element given by  $H_{i,j}(t) = d_{i,j}^\gamma h_{i,j}(t)$ , where  $d_{i,j}$  is the distance between the receiver antenna  $i$  and the transmitter antenna  $j$  and  $\gamma$  is a path loss parameter, and  $h_{i,j}(t)$  denotes the small scale fading.  $\mathbf{x}(t) \in \mathbb{C}^{M \times 1}$  is the intended message signal,  $\mathbf{V}(t) \in \mathbb{C}^{M \times M}$  is the precoding matrix at the BS,  $\mathbf{U}(t) \in \mathbb{C}^{N \times N}$  is the aggregated combining matrix at the users' side, and  $\mathbf{z}(t) \in \mathbb{C}^{N \times 1}$  denotes the independent and identically distributed (IID) Gaussian noise matrix. We assume channels remain the same during the whole algorithm running procedure. In this way, we can drop the time index  $t$  in the following. In our algorithm we assume that each user can share their received signal in order to reach the optimal performance; i.e.,  $y_i$  is known by user  $j$  with  $j \neq i$ . However, we also compare this user-collaborate scheme with the case that users have no collaborations; i.e.,  $y_i$  is not known by user  $j$  with  $j \neq i$ .

Furthermore, the channel model  $\mathbf{H}$  is described as

$$\mathbf{H} = \sqrt{\frac{k}{k+1}} \mathbf{H}_{\text{LOS}} + \sqrt{\frac{1}{k+1}} \mathbf{H}_{\text{NLOS}}, \quad (2)$$

where  $\mathbf{H}_{\text{LOS}}$  and  $\mathbf{H}_{\text{NLOS}}$  denote the line-of-sight and the non-line-of-sight (NLOS) components of the channel, respectively, and the NLOS component is assumed to follow a complex Gaussian distribution. Also,  $k$  controls the relative strength of the LOS and the NLOS components. In (2), setting  $k = 0$  represents an NLOS condition while  $k \rightarrow \infty$  gives a LOS channel. We use this model because most cases in MMW systems have the LOS channel.

**3.1. Initial Beam Refinement Procedure.** Unlike a conventional beamforming procedure acquiring CSI, in MMW

systems we suggest to perform codebook-based beam refinement, which means selecting a precoding matrix  $\mathbf{V}$  out of a predefined codebook  $\mathbf{W}_T$  at the BS while selecting a combining matrix  $\mathbf{U}$  out of a predefined codebook  $\mathbf{W}_R$  at the receiver side, sending test signal, and finally making decisions on transmit/receive beam patterns based on the users' feedback about their performance metrics. As the final step of IA [4], the beam refinement procedure can obtain a refined beam alignment at the cost of computational delay. The time structure for a packet transmission can be seen in Figure 2, where part of the packet period is dedicated to design appropriate beams in the IA procedure (mainly the beam refinement part) and the rest is used for data transmission. Thus, we need to find a balance between the beam design delay and the data transmission period by choosing an efficient approach.

Here, we use discrete Fourier transform- (DFT-) based codebooks [25] at both sides which are defined as

$$\mathbf{W}_T = \{w(m, u)\} = \left\{ e^{-j2\pi(m-1)(u-1)/N_{\text{vec}}} \right\}, \quad (3)$$

$$m = 1, 2, \dots, M, \quad u = 1, 2, \dots, N_{\text{vec}},$$

for the BS, while

$$\mathbf{W}_R = \{w(n, u)\} = \left\{ e^{-j2\pi(n-1)(u-1)/N_{\text{vec}}} \right\}, \quad (4)$$

$$n = 1, 2, \dots, N, \quad u = 1, 2, \dots, N_{\text{vec}},$$

for the users, where  $N_{\text{vec}} \geq \max(M, N)$  is the number of codebook vectors. Note that since our algorithm is generic, one can apply our proposed algorithm for different kinds of codebooks.

**3.2. Performance Metrics.** The machine learning-based schemes of [6, 7, 16–18] are generic, in the sense that they can be implemented for different metrics. For the simulations, however, we consider the service outage-constrained end-to-end throughput, the complexity and the average number of required iterations as the system performance metric. In some scenarios, it may be required to serve the users with some minimum required rates; otherwise *service outage* occurs. In the  $K$ -th iteration round of the algorithm,

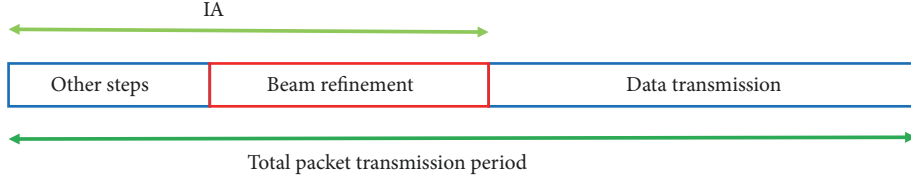


FIGURE 2: Schematic of a packet transmission period.

the service outage-constrained end-to-end throughput in bit-per-channel-use (bpcu) is defined as

$$R(K) = (1 - \alpha K) \sum_{i=1}^{\tau} r_i^K U(r_i^K, \log_2(1 + \theta)), \quad (5)$$

where

$$r_i^K = \log_2(1 + \text{SINR}_i^K), \quad (6)$$

$$U(r_i^K, \log_2(1 + \theta)) = \begin{cases} 1 & r_i^K \geq \log_2(1 + \theta) \\ 0 & r_i^K < \log_2(1 + \theta). \end{cases} \quad (7)$$

Here,  $r_i^K$  denotes the achievable rate of the user  $i$  at the end of the  $K$ -th iteration. Also, parameter  $\alpha$  is the relative delay cost for running each iteration of the algorithm which fulfills  $\alpha N_{\text{it}} < 1$  with  $N_{\text{it}}$  being the maximum possible number of iterations. Then,  $\log_2(1 + \theta)$  is the minimum per-user rate while  $\theta$  represents the minimum required signal-to-interference-plus-noise ratio (SINR) of each user. Also,

$$\text{SINR}_i^K = \frac{(P/M) g_{i,i}^K}{BN_0 + (P/M) \sum_{i \neq j}^N g_{i,j}^K} \quad (8)$$

is the SINR at the receiver of user  $i$  in the iteration round  $K$ . Hence, we define the satisfied user as  $\text{SINR}_i^K \geq \theta$ . Here,  $g_{i,j}$  is the  $(i, j)$ -th element of the matrix  $\mathbf{G}_K = |\mathbf{U}_K^H \mathbf{H} \mathbf{V}_K|^2$  which is referred to as the channel gain throughout the paper. Moreover,  $B$  is the system bandwidth and  $N_0$  is the power spectral density of the noise. We set  $BN_0 = 1$  to simplify the system so that the power  $P$  (in dB,  $10 \log_{10} P$ ) denotes the receiver side SNR as well.

The optimization problem of (5) is formulated as

$$\begin{aligned} \max_{K, \mathbf{U}, \mathbf{V}} \quad & R(K) \\ \text{s.t.} \quad & \forall K \in \{1, 2, 3, \dots, N_{\text{it}}\} \\ & \forall \mathbf{V} \subseteq \mathbf{W}_T \\ & \forall \mathbf{U} \subseteq \mathbf{W}_R. \end{aligned} \quad (9)$$

As opposed to, e.g., [17, Eq. 3], [22, Eq. 1], [26, Eq. 43], [27, Eq. 3], [28, Eq. 5] and [29, Eq. 5], we consider the algorithm running delay in the performance analysis. As seen in the following, there is a trade-off between optimizing beamforming matrices and reducing the data transmission period. In this case the optimal solution may be achieved by running the algorithms for a limited number of iterations.

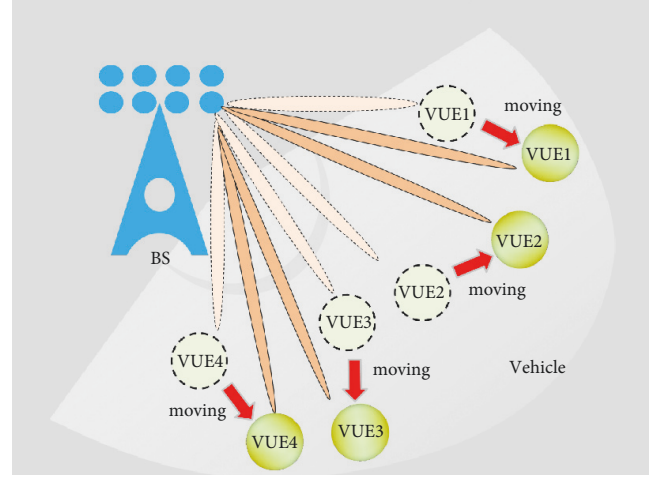


FIGURE 3: Mobility model, assuming that we know the moving distance for each user. The spatial correlation can be exploited by setting the queen of the previous time slot as one of the initial guesses of the next time slot.

**3.3. On the Effect of Power Amplifier Efficiency.** The efficiency of the radio-frequency high power amplifier (PA) should be taken into consideration in the multiantenna systems. Here, we consider the state-of-the-art PA efficiency model [30, Eq. 13], [31, Eq. 3]:

$$\rho_{\text{cons}} = \frac{\rho_{\text{max}}^\mu}{\epsilon \times \rho_{\text{out}}^{\mu-1}} \quad (10)$$

where  $\rho_{\text{cons}}$ ,  $\rho_{\text{out}}$ ,  $\rho_{\text{max}}$  refer to as the consumed power, the output power, and the maximum output power of the PA, respectively. Also,  $\epsilon \in [0, 1]$  represents the power efficiency and  $\mu \in [0, 1]$  is a parameter depending on the PA class. Setting  $\epsilon = 1$ ,  $\rho_{\text{max}} = \infty$  and  $\mu = 0$  in (10) represents the special case (with an ideal PA).

**3.4. On the Effect of User Mobility.** Beamforming solutions for mobile users at high carrier frequencies are important in 5G wireless mobile communications. Here, we use the following mobility model to evaluate the performance of our proposed GA-based beamforming approach and compare the results with those of the considered state-of-the-art schemes. Consider Figure 3 with  $\tau = 4$  multiple-antenna VUEs with data exchange D2D links. Here, we have two cases during the users' mobility.

*Case 1.* This case includes beam refinement with a random queen as initial guess (dash-line VUEs in Figure 3).



In each time slot with instantaneous channel realization  $\mathbf{H} \in \mathbb{C}^{N \times M}$ , do the followings:

- (I) Initialization: Consider  $L$ , e.g.,  $L = 10$ , sets of precoding matrices  $\mathbf{V}_l$  and combining matrices  $\mathbf{U}_l$ ,  $l = 1, \dots, L$ , randomly selected from the pre-defined codebook  $\mathbf{W}_T$  and  $\mathbf{W}_R$ .
- (II) Selection: For each  $\mathbf{V}_l$  and  $\mathbf{U}_l$ , evaluate the instantaneous value of the objective metric  $R_l$ ,  $l = 1, \dots, L$ , for example end-to-end throughput (5). Find the best beamforming matrix which results in the best value of the considered metric, named as the *Queen*, e.g.,  $\mathbf{V}_q$  and  $\mathbf{U}_q$  satisfies  $R(\mathbf{V}_l, \mathbf{U}_l) \leq R(\mathbf{V}_q, \mathbf{U}_q)$ ,  $\forall l = 1, \dots, L$  if the end-to-end throughput is the objective function.
- (III) Save the Queen:  $\mathbf{V}_1 \leftarrow \mathbf{V}_q, \mathbf{U}_1 \leftarrow \mathbf{U}_q$
- (IV) Genetic operation I-Crossover: Create  $S < L$ , e.g.,  $S = 5$ , beamforming matrices  $\mathbf{V}_s^{\text{new}}$  and  $\mathbf{U}_s^{\text{new}}$ ,  $s = 1, \dots, S$ , around the Queen  $\mathbf{V}_1$  and  $\mathbf{U}_1$ . These sets are generated by making small changes in the Queen  $\mathbf{V}_q$  and  $\mathbf{U}_q$ .
- (V)  $\mathbf{V}_{s+1} \leftarrow \mathbf{V}_s^{\text{new}}, \mathbf{U}_{s+1} \leftarrow \mathbf{U}_s^{\text{new}}, s = 1, \dots, S$ .
- (VI) Genetic operation II-Mutation: Regenerate the remaining sets  $\mathbf{V}_s$  and  $\mathbf{U}_s$ ,  $s = S + 2, \dots, L$ , randomly with the same procedure as in Step (I).
- (VII) Go back to Step (II) and run for  $N_{\text{it}}$  iterations,  $N_{\text{it}}$  is a fixed number decided by designer. Return the final Queen as the beam selection rule for the current time slot.

ALGORITHM 1: GA-based beam refinement algorithm.

*Case 2.* This case includes beam refinement using the queen in Case 1 as initial guess (full-line VUEs in Figure 3).

By mobility we exploit the spatial correlation by setting the queen of the previous time slot as one of the initial guesses of the next time slot. For  $d_{i,j}$  in (1) we assume that we know the moving speed  $v$  and the time duration of mobility  $\Delta t$ . In this way, we can get an estimate of the user position in Case 2 in a circle whose radius is found by  $v \cdot \Delta t$  with the user position at the previous time slot being the center.

#### 4. Algorithm Description

In this study, we compare the performance of different IA beamforming methods as follows.

**Extended GA-based search [18]:** the algorithm starts by making  $L$  possible beam selection sets at both transmitter and receiver, i.e., submatrices of each codebook. During each iteration, we choose the best set, named as the *Queen*, based on the performance metrics (for example, (5)). Next, we keep the queen and regenerate  $S < L$  similar sets around the Queen by making small changes to the Queen (in the simulations, we replace 10% of the Queen columns randomly without loss of generality). Finally, the other  $L - S - 1$  beamforming matrices are selected randomly to avoid the algorithm from being trapped in a local minima. Note that reducing  $S$  for a given  $L$  can increase the chance of being trapped. After  $N_{\text{it}}$  iterations (set by the designer), the queen is returned as the beam selection result in the current time slot. In this way, this is an extended version of our GA-based approach with beamforming at both the transmitter and the receiver, the basic principles of which can be found in Algorithm 1.

**Tabu search [16]:** The Tabu search approach follows the basic idea as in the GA-based scheme [16] where we choose and update the queen by iterations. The only difference is the evolution method of the queen in successive iterations. With Tabu, we use the definition of *neighborhood* in [16]: one matrix  $\mathbf{A}$  is defined as another matrix  $\mathbf{B}$ 's neighborhood if (1)  $\mathbf{A}$  has only one different column compared with  $\mathbf{B}$

or (2) the index difference between the two corresponding columns in  $\mathbf{A}$  and  $\mathbf{B}$  is equal to one. To make  $S$  beam selection sets, we change the queen from previous round to its neighbors.

**Link-by-link search [17]:** in this strategy, the beam design of  $\tau$  users is not optimized simultaneously. Instead, with a greedy approach, the beamforming solution is settled user-by-user by considering the interference from the other  $\tau - 1$  links. The system performance improves in successive iterations until it converges to some (sub)optimal beamforming rules.

**Two-level search [6, 7]:** being inspired by multistage beamforming techniques, e.g., [6, 7], we design a two-level-codebook search scheme for our system. In the first level, the BS transmits messages over wider sectors using the codebook with  $N_{\text{vec}}/2$  columns, while in the second level it searches the optimal solution within the best such sector by steering narrower beams with an  $N_{\text{vec}}$ -column codebook.

**4.1. On the Implementation Complexity.** To compare different methods, it is necessary that we consider the implementation complexity of each algorithm. For this reason, we derive the per-iteration complexity of different algorithms based on the fact that the product of matrices of size  $N \times M$  and  $M \times M$  has the complexity  $\mathcal{O}(NM^2)$  in MATLAB. In this way, the per-iteration complexity for the GA-based approach is given by

$$C_{\text{GA}} = L \left( 2\mathcal{O}(N^2M) + \mathcal{O}(NM^2) + \mathcal{O}(NM) \right), \quad (11)$$

and  $C_{\text{Tabu}} = C_{\text{GA}}$ ,  $C_{\text{link-by-link}} = \tau \times C_{\text{GA}}$ ,  $C_{\text{two-level}} = 2 \times C_{\text{GA}}$ .  $L$  is the number of beam selection sets within each iteration.

**4.2. On the Effect of User Collaboration.** In order to optimize the end-to-end system throughput (5), each user needs to share its received signal with the other users via the D2D links as mentioned in Section 2. Note that we do not consider the overhead of building up the D2D links in this work. We compare two cases regarding the user collaboration.

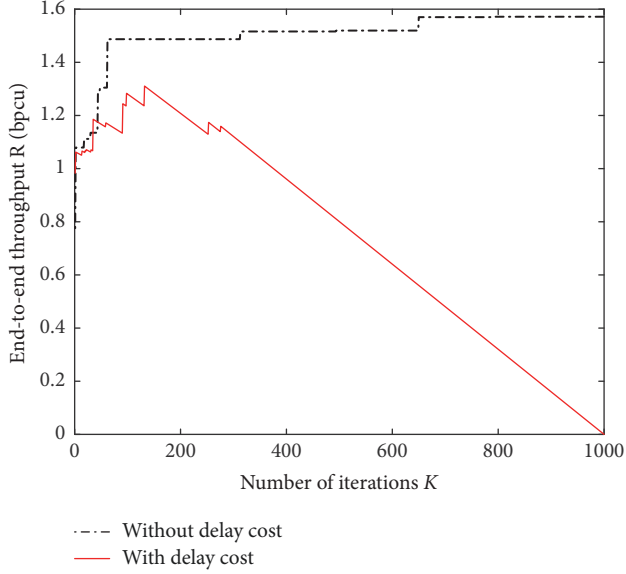


FIGURE 4: An example of the convergence process of the extended GA-based beamforming for systems with (subplot a) and without (subplot b) delay costs of the algorithm.  $M = 32$ ,  $\tau = 4$ ,  $N = 12$ ,  $P = -10$  dB,  $k = 0$ .

*Case 1* (collaborative users (CUs)). Each user knows the received signals of the other users and the system throughput is optimal.

*Case 2* (noncollaborative users (NCUs)). Each user only knows their own received message and the system throughput is suboptimal.

In Section 5, we evaluate the performance of the GA and the Tabu methods in these two cases and investigate the potential gains of collaboration.

## 5. Simulation Results

In the simulations, we use the channel model in (2) in the cases with  $k = 0, 3$ . We set  $\mathbf{H}_{\text{LOS}} = \mathbf{I}_{N \times M}$  where  $\mathbf{I}_{a \times b}$  refers to the normalized all-ones complex matrix. Except for Figure 4 which shows an example of the GA-based procedure, for each point in the curves the results are obtained by averaging over  $10^4$  different channel realizations. In all figures, we set  $N_{\text{it}} = 1000$  since it is a sufficiently large number of iterations after which no performance improvement is observed. Also, in all figures except for Figure 11, we use the normalized distance  $d_{i,j} = d = 1$ . Moreover, we set  $L = 10$ ,  $S = 5$  and  $N_{\text{vec}} = 128$ . In all figures, except for Figure 9, we use the ideal PA; i.e., set  $P_{\text{max}} = \infty$ ,  $\mu = 0$ ,  $\epsilon = 1$  in (10). In Figure 9 we study the effect of imperfect PAs. In Figures 4, 7, 9, and 10, we consider the service outage-constrained end-to-end throughput (5) as the performance metric with  $\theta = -4$  dB. Finally, Table 1 shows the average number of required iterations in each algorithm to reach the (sub)optimal solution.

*On the convergence behavior:* Figure 4 gives an example of the GA performance in the cases with ( $\alpha = 0.001$ ) and without costs of running the algorithm ( $\alpha = 0$ ), respectively

TABLE 1: Average number of required iterations  $\bar{N}$  in different situations.

$M/N$	GA	Tabu	link-by-link	two-level
32/12	502	498	307	501
32/8	500	501	288	498
32/4	488	502	261	500

(see (5)). Here, example means we run our algorithm within one single channel realization. From Figure 4 we observe that very few iterations are required to reach the maximum throughput for the cases with delay cost, which is around  $K = 130$ . That is, considering the cost of running the algorithm, the maximum throughput is obtained by finding a suboptimal beamforming matrix and leaving the rest of the time slot for data transmission (see Figure 2). As a result, as the number of iterations increases, the cost of running the algorithm reduces the end-to-end throughput converging to zero at  $K = 1/\alpha$  (see (5)). Note that the top value of the delay case is less than the other one due to the delay cost.

If there is no running delay, on the other hand, the system performance improves with the number of iterations monotonically. However, the developed algorithm leads to (almost) the same performance as the exhaustive search-based scheme with very limited number of iterations. For example, with the parameter settings of Figure 4, our algorithm reaches more than 90% of the maximum achievable throughput with less than 100 iterations. On the other hand, with the parameter settings of Figure 4, exhaustive search implies testing in the order of  $10^{30}$  possible beamforming matrices. Note that we cannot guarantee that the results are exactly the same with the optimal but because of the “random” part of the algorithm they become very close with large number of  $N_{\text{it}}$ . The trade-off between the performance and the delay cost is the concern here instead of the exact throughput value.

Finally, all considered schemes follow the same ladder-type convergence behavior as in Figure 4. This is because with the considered algorithms the system performance is not necessarily improved in each iteration and may be trapped into local minima. However, considering a couple of random solution checks in each iteration helps to avoid the local minima as the number of iterations increases.

*On the effect of service outage:* Figure 5 demonstrates the service outage-constrained end-to-end throughput (5) for different values of the required received SNR thresholds  $\theta$  in (5). Also, Fig. 6 studies the service outage probability in the cases optimizing (5). Here, the results are presented for  $N = 8$ ,  $\tau = 8$ ,  $M = 32$ ,  $k = 0$ ,  $N_{\text{vec}} = 128$ , which means single-antenna user at the receiver side. As demonstrated in Figures 5 and 6, the service outage constraint affects the end-to-end and the per-user throughput significantly at low SNRs/severe service outage constraints. However, the effect of the service outage probability decreases as the SNR increases or  $\theta$  decreases (Figures 5 and 6).

*Comparison of schemes:* in Figure 7, we compare the throughput (5) reached by the considered algorithms. It can be seen from the figure that for a broad range of SNRs the GA-based beamforming [18] leads to the best system throughput,

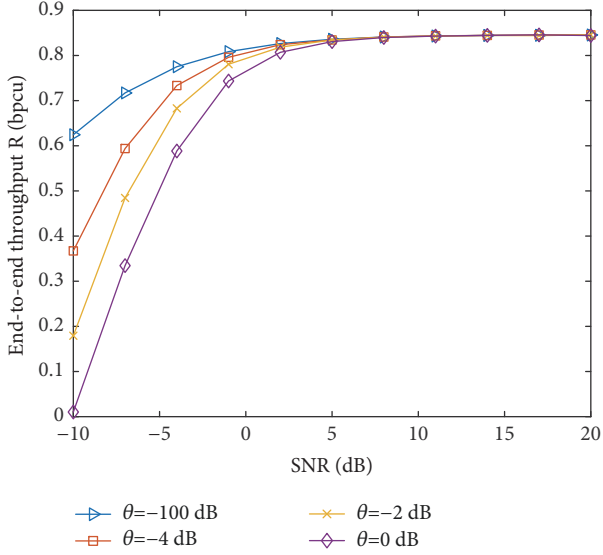


FIGURE 5: Service outage-constrained end-to-end throughput of the GA method with different  $\theta$ 's.  $M = 32$ ,  $\tau = 8$ ,  $N = 8$ ,  $k = 0$ ,  $\alpha = 0$ .

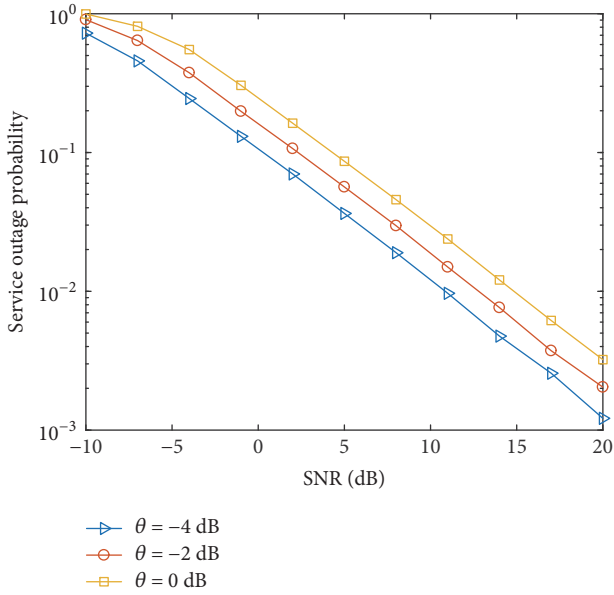


FIGURE 6: Service outage probability of the GA method with different  $\theta$ 's.  $M = 32$ ,  $\tau = 8$ ,  $N = 8$ ,  $k = 0$ ,  $\alpha = 0$ .

followed by the link-by-link search [17], Tabu search [16], and two-level search [6, 7].

Moreover, using the same parameter settings of Figure 7, in Figure 8 we compare the cumulative distribution function (CDF) of the per-user throughput (5) reached by the considered algorithms. From the figure we can see that the GA-based beamforming [18] leads to the best per-user throughput distribution, which means more users can be served by higher throughput, followed by the link-by-link search [17], Tabu search [16], and two-level search [6, 7].

Table 1 shows the average number of iterations  $\bar{N}$  that is required in each scheme to reach a (sub)optimal solution.

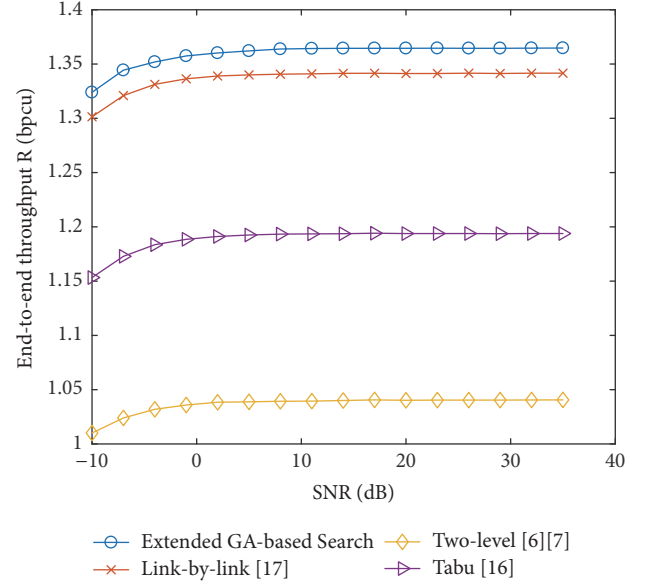


FIGURE 7: Service outage-constrained end-to-end throughput of different methods.  $M = 32$ ,  $\tau = 4$ ,  $N = 12$ ,  $k = 0$ ,  $\alpha = 0$ .

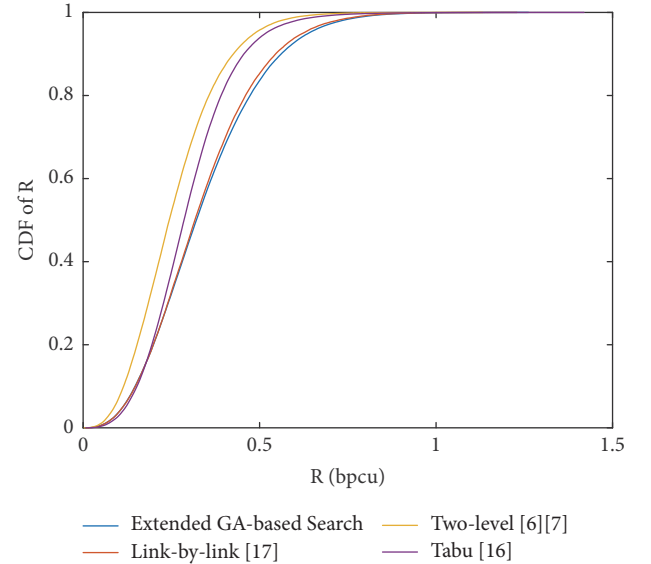


FIGURE 8: CDF of per-user throughput with different methods.  $M = 32$ ,  $\tau = 4$ ,  $N = 12$ ,  $k = 0$ ,  $\alpha = 0$ .

Here, the results are presented for  $k = 0$ ,  $M = 32$ ,  $N = 4, 8, 12$ . We can see that, in all methods, except for the link-by-link approach, the required number of iterations is almost insensitive to the number of receive antennas for the considered parameter setting of Table 1.

*On the effect of imperfect power amplifier:* Figure 9 evaluates the effect of the power amplifier on the throughput (5). We can see that the inefficiency of the PA affects the performance remarkably but this effect decreases with the SNR. This is reasonable because the effective efficiency of the PAs  $\epsilon^{\text{effective}} = \epsilon(p_{\text{out}}/p_{\text{max}})^\mu$  increases with SNR.



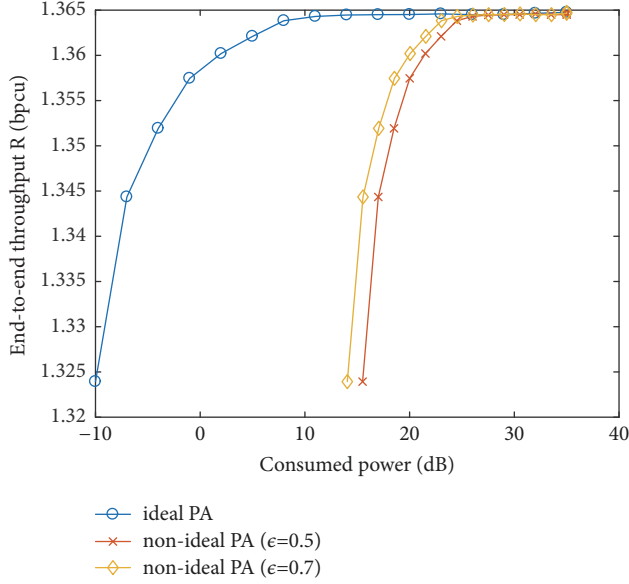


FIGURE 9: The effect of power budget and PAs efficiency on the end-to-end throughput (5).  $M = 32$ ,  $\tau = 4$ ,  $N = 12$ ,  $k = 0$ ,  $\alpha = 0$ .

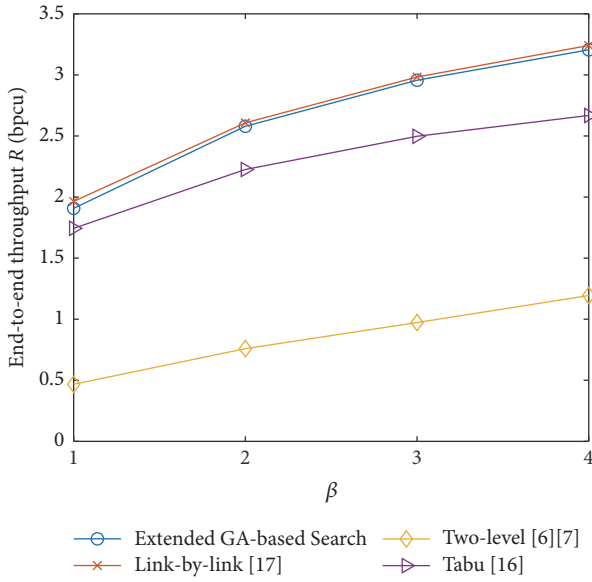


FIGURE 10: Throughput (5) with different number of receive antennas at the user side  $\beta$ .  $M = 32$ ,  $\tau = 4$ ,  $\beta = 1, 2, 3, 4$ ,  $k = 3$ ,  $\alpha = 0$ ,  $P = 2$  dB.

*On the effect of the number of receive antennas:* Figure 10 shows the effect of number of receive antennas per-user  $\beta$  on the throughput (5). As seen in the figure, the end-to-end throughput increases with the number of per-user antennas as expected, since multiantenna techniques can improve the data rate remarkably. Moreover, the relative performance gain of the GA-based and the link-by-link scheme, compared to the other considered schemes, increases with the number of receive antennas, which is an interesting point when designing large-scale networks.

*On the effect of the user mobility:* Figure 11 shows the effect of the users' mobility on the beam refinement delay for

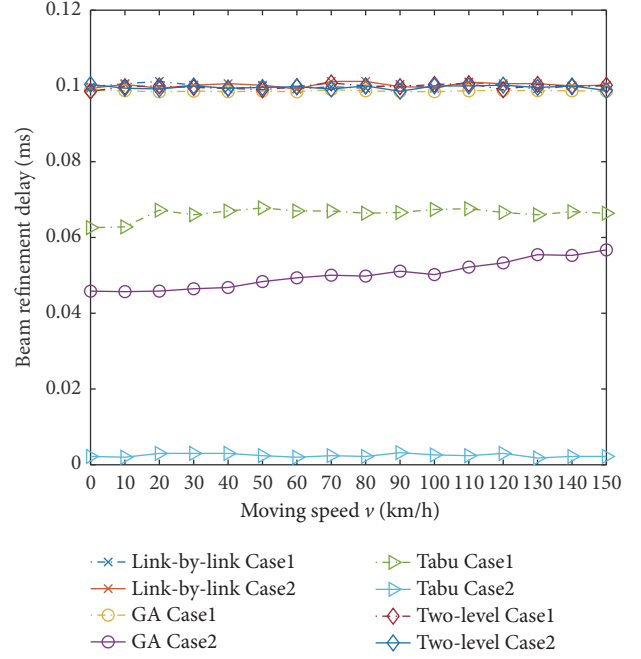


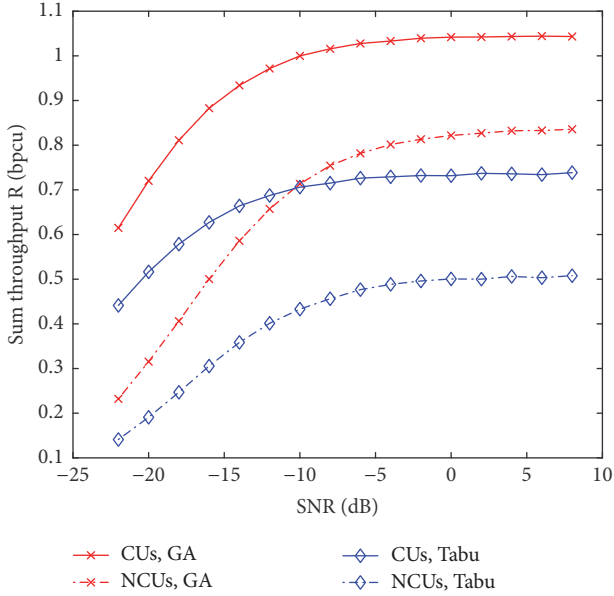
FIGURE 11: Beam refinement delay with different moving speed  $v$ .  $M = 32$ ,  $\tau = 4$ ,  $\beta = 2$ ,  $k = 0$ ,  $\alpha = 0$ ,  $P = 32$  dB, moving time  $\Delta t = 1$  ms,  $\gamma = -3.5$ .

the considered algorithms. Inspired by [11], we evaluate the beam refinement delay (we assume that each iteration takes  $10^{-4}$  overhead of  $\Delta t$ ) of each algorithm in Cases 1 and 2 to check how well these algorithms are suitable for the mobile users. The algorithm running delays in Cases 1 and 2 of each method are all presented in the plot. Here, the results are presented with  $M = 32$ ,  $\tau = 4$ ,  $\beta = 2$ ,  $k = 0$ ,  $\alpha = 0$ ,  $P = 32$  dB, moving time  $\Delta t = 1$  ms,  $\gamma = -3.5$ . As seen in the figure, both the GA-based algorithm and the Tabu-based algorithm can remarkably reduce the beam refinement delay for a broad range of users speeds, since they can use the beam refinement solution in Case 1 as the initial guess in Case 2 when the moving distance is not large. Note that Tabu search has the lowest delay in both cases since it simply changes the queen to its neighbors which takes full advantage of the spatial correlations. However, for GA-based scheme as the users speed increases the beam refinement delay increases slightly, intuitively because the spatial correlation between the positions in successive time slots decreases. Moreover, both the link-by-link search and the two-level-based search do not show noticeable performance gain.

*On the effect of collaborative users:* Figure 12 shows the effect of the users' collaboration on the end-to-end throughput for the GA and Tabu algorithms. Also, Table 2 presents the average number of required iterations for both the GA and the Tabu search in the cases with the CUs and the NCUs. Here, the results are presented with  $M = 32$ ,  $\tau = 4$ ,  $\beta = 2$ ,  $k = 0$ ,  $\alpha = 0$ . As seen in the figure, the performance of both the GA-based algorithm and the Tabu-based algorithm are reduced in the case of NCU. Also, these reductions decrease as the SNR increases. On the other hand, in Table 2 it can be seen that the NCU case requires

TABLE 2: Average number of required iterations  $\bar{N}$  in different situations.

$M/N$	GA, CUs	GA, NCUs	Tabu, CUs	Tabu, NCUs
32/12	502	1	498	1
32/8	500	1	501	1
32/4	488	1	502	1

FIGURE 12: End-to-end throughput for different SNR with the CUs and NCUs cases for GA and Tabu.  $M = 32$ ,  $\tau = 4$ ,  $\beta = 2$ ,  $k = 0$ ,  $\alpha = 0$ .

much smaller iteration time compared with the CUs case for different system configurations. Only one iteration is required for the case with  $\beta \leq 3$ .

## 6. Conclusion

We extended our previously proposed genetic algorithm-(GA-) based beam refinement scheme to include beamforming at both the transmitter and the receiver, and we compared the performance with alternative beam refinement algorithms in an MU-MIMO system, in terms of the service outage-constrained end-to-end throughput and the implementation complexity. Particularly, our extended genetic algorithm-based scheme can reach almost the same throughput as in the exhaustive search-based approach with relatively few iterations in delay-constrained systems. Also, compared to the considered state-of-the-art schemes, our scheme leads to the highest throughput/per-user throughput and the lowest per-iteration implementation complexity, and the relative performance gain increases with the number of receive antennas. Moreover, non-ideal power amplifiers affect the system performance remarkably, which should be carefully considered during the system design. Furthermore, the GA-based approach can exploit the spatial correlation and remarkably reduce the beam refinement delay for a broad range of users speeds, which means it is an appropriate

approach for mobile users. Finally, collaborative users can improve the system-level performance at the expense of computational complexity. For future work, we will investigate our proposed algorithm with more realistic parameter settings/scenarios and compare the result with other structured beamforming methods.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

This work is partly based on [32], and it has been partly performed in the framework of the H2020 Project 5GCAR cofunded by the EU. It has also been supported in part by VINNOVA (Swedish Government Agency for Innovation Systems) within the VINN Excellence Center ChaseOn. The authors would like to acknowledge the contributions of their colleagues.

## References

- [1] M. Cudak, A. Ghosh, T. Kovarik et al., "Moving Towards Mmwave-Based Beyond-4G (B-4G) Technology," in *Proceedings of the 2013 IEEE 77th Vehicular Technology Conference (VTC Spring)*, pp. 1–5, Dresden, Germany, June 2013.
- [2] Z. Pi and F. Khan, "An introduction to millimeter-wave mobile broadband systems," *IEEE Communications Magazine*, vol. 49, no. 6, pp. 101–107, 2011.
- [3] S. Sun, G. R. Maccartney, M. K. Samimi, S. Nie, and T. S. Rappaport, "Millimeter wave multi-beam antenna combining for 5G cellular link improvement in New York City," in *Proceedings of the 1st IEEE International Conference on Communications (ICC '14)*, pp. 5468–5473, IEEE, Sydney, Australia, June 2014.
- [4] mmMagic Project D4.2, Final radio interface concepts and evaluations for mm-wave mobile communications, [https://bscw.5g-mmmagic.eu/pub/bscw.cgi/d214055/mmMAGIC\\_D4.2.pdf](https://bscw.5g-mmmagic.eu/pub/bscw.cgi/d214055/mmMAGIC_D4.2.pdf).
- [5] Y. Sui and T. Svensson, "Uplink enhancement of vehicular users by using D2D communications," in *Proceedings of the 2013 IEEE Globecom Workshops (GC Wkshps)*, pp. 649–653, Atlanta, GA, USA, December 2013.
- [6] S. Hur, T. Kim, D. J. Love, J. V. Krogmeier, T. A. Thomas, and A. Ghosh, "Multilevel millimeter wave beamforming for wireless backhaul," in *Proceedings of the 2011 IEEE Globecom Workshops*, pp. 253–257, Houston, TX, USA, December 2011.
- [7] . Li Chen, . Ying Yang, . Xiaohui Chen, and . Weidong Wang, "Multi-stage beamforming codebook for 60GHz WPAN," in *Proceedings of the 2011 6th International ICST Conference on Communications and Networking in China (CHINACOM)*, pp. 361–365, Harbin, China, August 2011.
- [8] V. Desai, L. Krzymien, P. Sartori, W. Xiao, A. Soong, and A. Alkhateeb, "Initial beamforming for mmWave communications," in *Proceedings of the 2014 48th Asilomar Conference on Signals, Systems and Computers*, pp. 1926–1930, Pacific Grove, CA, USA, November 2014.
- [9] C. Jeong, J. Park, and H. Yu, "Random access in millimeter-wave beamforming cellular networks: issues and approaches," *IEEE Communications Magazine*, vol. 53, no. 1, pp. 180–185, 2015.

- [10] Z. Abu-Shaban, H. Wymeersch, X. Zhou, G. Seco-Granados, and T. Abhayapala, "Random-phase beamforming for initial access in millimeter-wave cellular networks," in *Proceedings of the 59th IEEE Global Communications Conference, GLOBECOM 2016*, usa, December 2016.
- [11] M. Giordani, M. Mezzavilla, and M. Zorzi, "Initial Access in 5G mmWave Cellular Networks," *IEEE Communications Magazine*, vol. 54, no. 11, pp. 40–47, 2016.
- [12] Y. Li, J. G. Andrews, F. Baccelli, T. D. Novlan, and J. Zhang, "On the Initial Access Design in Millimeter Wave Cellular Networks," in *Proceedings of the 2016 IEEE Globecom Workshops (GC Wkshps)*, pp. 1–6, Washington, DC, USA, December 2016.
- [13] V. Raghavan, J. Cezanne, S. Subramanian, A. Sampath, and O. Koymen, "Beamforming Tradeoffs for Initial UE Discovery in Millimeter-Wave MIMO Systems," *IEEE Journal of Selected Topics in Signal Processing*, vol. 10, no. 3, pp. 543–559, 2016.
- [14] Y. Li, J. G. Andrews, F. Baccelli, T. D. Novlan, and C. J. Zhang, "Design and Analysis of Initial Access in Millimeter Wave Cellular Networks," *IEEE Transactions on Wireless Communications*, vol. 16, no. 10, pp. 6409–6425, 2017.
- [15] I. Filippini, V. Sciancalepore, F. Devoti, and A. Capone, "Fast Cell Discovery in mm-wave 5G Networks with Context Information," *IEEE Transactions on Mobile Computing*, 2017.
- [16] X. Gao, L. Dai, C. Yuen, and Z. Wang, "Turbo-like beamforming based on tabu search algorithm for millimeter-wave massive mimo systems," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 7, pp. 5731–5737, 2016.
- [17] J. Qiao, X. Shen, J. W. Mark, and Y. He, "MAC-layer concurrent beamforming protocol for indoor millimeter-wave networks," *IEEE Transactions on Vehicular Technology*, vol. 64, no. 1, pp. 327–338, 2015.
- [18] H. Guo, B. Makki, and T. Svensson, "A genetic algorithm-based beamforming approach for delay-constrained networks," in *Proceedings of the 2017 15th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt)*, pp. 1–7, Paris, France, May 2017.
- [19] J. P. Gilb, *IEEE standards 802.15. 3cpart 15.3: wireless medium access control (MAC) and physical layer (PHY) specifications for high rate wireless personal area networks (WPANs) amendment 2: millimeter-wave-based alternative physical layer extension [s]*, IEEE Computer Society, New York, NY, USA, 2009.
- [20] C. Cordeiro et al., *IEEE P802. 11 Wireless LANs, PHY/MAC Complete Proposal Specification (IEEE 802.11-10/0433r2)*, 2010.
- [21] H. Rate, "GHz PHY, MAC and PALs, Standard ECMA-387, ser," 2010, <https://www.ecma-international.org/publications/files/ECMA-ST/ECMA-387.pdf>.
- [22] C. N. Barati, S. A. Hosseini, M. Mezzavilla et al., "Initial Access in Millimeter Wave Cellular Systems," *IEEE Transactions on Wireless Communications*, vol. 15, no. 12, pp. 7926–7940, 2016.
- [23] B. Makki, T. Svensson, G. Cocco, T. De Cola, and S. Erl, "On the throughput of the return-link multi-beam satellite systems using genetic algorithm-based schedulers," in *Proceedings of the IEEE International Conference on Communications, ICC 2015*, pp. 838–843, gbr, June 2015.
- [24] H. Guo, B. Makki, and T. Svensson, "A comparison of beam refinement algorithms for millimeter wave initial access," in *Proceedings of the 2017 IEEE 28th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, pp. 1–7, Montreal, Canada, October 2017.
- [25] L. Wan, X. Zhong, Y. Zheng, and S. Mei, "Adaptive codebook for limited feedback MIMO system," in *Proceedings of the 2009 IFIP International Conference on Wireless and Optical Communications Networks (WOCN)*, pp. 1–5, Cairo, Egypt, April 2009.
- [26] J. Choi, "Beam Selection in mm-Wave Multiuser MIMO Systems Using Compressive Sensing," *IEEE Transactions on Communications*, vol. 63, no. 8, pp. 2936–2947, 2015.
- [27] O. E. Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. W. Heath, "Spatially sparse precoding in millimeter wave MIMO systems," *IEEE Transactions on Wireless Communications*, vol. 13, no. 3, pp. 1499–1513, 2014.
- [28] B. Li, Z. Zhou, W. Zou, X. Sun, and G. Du, "On the efficient beam-forming training for 60GHz wireless personal area networks," *IEEE Transactions on Wireless Communications*, vol. 12, no. 2, pp. 504–515, 2013.
- [29] H.-H. Lee and Y.-C. Ko, "Low complexity codebook-based beamforming for MIMO-OFDM systems in millimeter-wave WPAN," *IEEE Transactions on Wireless Communications*, vol. 10, no. 11, pp. 3607–3612, 2011.
- [30] B. Makki, T. Svensson, T. Eriksson, and M.-S. Alouini, "On the Required Number of Antennas in a Point-To-Point Large-but-Finite MIMO System: Outage-Limited Scenario," *IEEE Transactions on Communications*, vol. 64, no. 5, pp. 1968–1983, 2016.
- [31] D. Persson, T. Eriksson, and E. G. Larsson, "Amplifier-aware multiple-input single-output capacity," *IEEE Transactions on Communications*, vol. 62, no. 3, pp. 913–919, 2014.
- [32] H. Guo, *Initial Access in mm-wave 5G Mobile Communications [Master thesis]*, Chalmers University of Technology, 2017.