



MAMBAS: Maneuvering Analog Multi-User Beamforming using an Array of Subarrays in mmWave Networks

Zhihui Gao, Zhenzhou Qi, Tingjun Chen

Department of Electrical and Computer Engineering, Duke University

ABSTRACT

Beyond-5G and 6G wireless networks exploit the millimeter-wave (mmWave) frequency bands to achieve significantly improved data rates, and existing mmWave systems rely on analog single-user beamforming (SUBF) or hybrid multi-user beamforming (MUBF). In this work, we focus on improving the performance of multi-user communication in mmWave networks by exploring analog MUBF using an array of subarrays (ASA) with reduced system overhead and hardware complexity as it eliminates digital beamforming and the need for estimating the channel state information (CSI). We present MAMBAS, a novel system that maneuvers analog MUBF using an ASA to support simultaneous communication with multiple users located in close proximity, e.g., within the half-power beamwidth of the ASA. In essence, MAMBAS effectively decouples the user selection, subarray allocation, and beamforming optimization based on a comprehensive understanding of the multi-user support determined by the ASA. We evaluate MAMBAS using a 28 GHz software-defined radio testbed and show that, compared to existing methods, MAMBAS can effectively support users that are $2\times$ more closely spaced while achieving an improved sum rate of up to $2\times$, using only two subarrays. Large-scale ray tracing-based simulations also show that MAMBAS can achieve a sum rate gain of $1.92\text{--}3.86\times$ and is able to maintain consistent performance with significantly increased user density.

CCS CONCEPTS

• **Networks** → **Network architectures; Network components; Wireless access points, base stations and infrastructure.**

KEYWORDS

Millimeter-wave; Multi-user beamforming; Optimization

ACM Reference Format:

Zhihui Gao, Zhenzhou Qi, Tingjun Chen. 2024. **MAMBAS**: Maneuvering Analog Multi-User Beamforming using an Array of Subarrays in mmWave Networks. In *The 30th Annual International Conference on Mobile Computing and Networking (ACM MobiCom '24)*, September 30–October 4, 2024, Washington D.C., DC, USA. ACM, New York, NY, USA, 15 pages. <https://doi.org/10.1145/3636534.3649390>

1 INTRODUCTION

Beyond-5G and 6G wireless systems exploit the millimeter-wave (mmWave) band to achieve significantly improved data rates using the largely available spectrum resources [28, 55]. For example, 5G new radio (NR) supports a single channel with up to 400 MHz bandwidth in the frequency range 2 (FR2) [3, 42], and IEEE 802.11ad supports a bandwidth of 2.16 GHz in 60 GHz unlicensed band [29]. In addition, modern communication systems employ analog and/or digital beamforming techniques using large antenna arrays to support efficient multiple-input and multiple-output (MIMO) operations, as depicted in Fig. 1. For example, single-user beamforming (SUBF) aims to enhance the signal-to-noise ratio (SNR) toward the intended user [7, 61], and is usually realized using a phased array. Multi-user beamforming (MUBF) extends this concept to simultaneously serve multiple users in an all-digital or hybrid beamforming system [17, 24, 40, 54].

The majority of mmWave radios have adopted analog SUBF with efficient beam training schemes to deliver high data rates over mmWave channels with relatively high signal propagation loss [48, 49]. While digital and hybrid MUBF systems can improve the network capacity, their realizations at mmWave frequencies face numerous challenges due to the complex hardware and signal processing algorithms associated with large arrays [12, 59, 61]. Moreover, the feasibility and performance of MUBF systems with high user densities are constrained by the geometry, spatial resolution, and beamforming degrees of freedom (DoF) of the antenna array [61]. These challenges need to be addressed to enable concurrent multi-user transmission in mmWave networks, particularly in scenarios with *closely spaced users*, e.g., within the half-power beamwidth (HPBW) of the array.



This work is licensed under a Creative Commons Attribution International 4.0 License.
ACM MobiCom '24, November 18–22, 2024, Washington D.C., DC, USA
© 2024 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-0489-5/24/09.
<https://doi.org/10.1145/3636534.3649390>

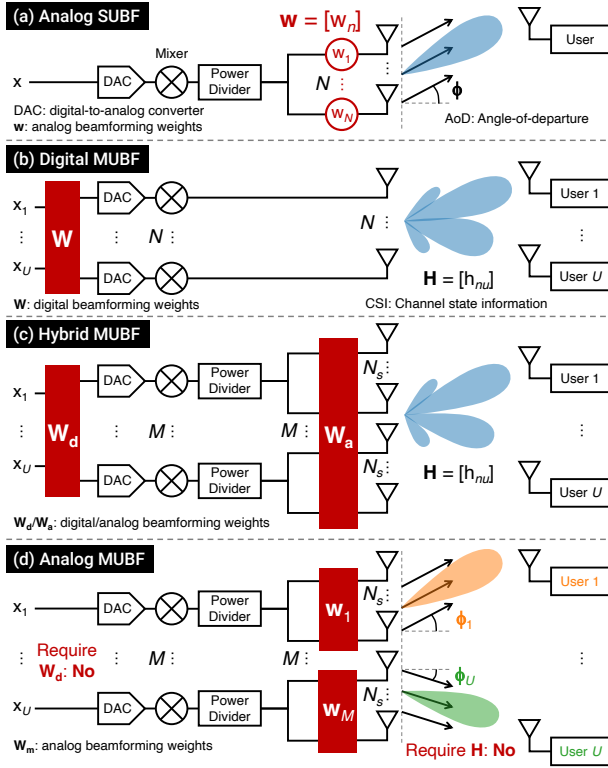


Figure 1: Different SUBF and MUBF systems: (a) Analog SUBF using a single phased array, (b) Digital MUBF using an all-digital array, (c) Hybrid MUBF using an array of subarrays (ASA), (d) Analog MUBF using an ASA, employed by MAMBAS.

In this work, we aim to improve multi-user communication in mmWave networks by exploring analog MUBF using an array of subarrays (ASA), as shown in Fig. 1(d). Compared to analog SUBF where users need to be served based on time division multiple access (TDMA), analog MUBF supports *simultaneous communications with multiple users* and can achieve higher data rates. Compared to digital MUBF, analog MUBF using an ASA has *reduced hardware complexity and energy consumption* with a smaller number of RF chains. Compared to hybrid MUBF (e.g., used in 802.11ay [8, 22]), analog MUBF *eliminates the digital beamforming and the need for channel state information (CSI) estimation*, which is usually required to design the analog/digital beamforming weights in an ASA; instead, we only utilize the angle-of-departure (AoD) of the path, either line-of-sight (LOS) or non-line-of-sight (NLOS), with the highest gain for each user. Moreover, the flexibility of dynamically allocating the subarray(s) to different users presents another DoF for system optimization.

To achieve this goal, we design **MAMBAS**, a novel system for Maneuvering Analog MUBF using an ASA in mmWave networks, where the users can be located in very close proximity, *even within the HPBW of the ASA*. The key insight toward the design of MAMBAS is to effectively decouple the

user selection, subarray allocation, and beamforming optimization in analog MUBF systems employing an ASA while guaranteeing sufficient signal-to-interference-plus-noise ratio (SINR) at each user, based on a comprehensive understanding of the multi-user support determined by the ASA. MAMBAS consists of (i) an offline ASA preprocessing step (§6.1), which establishes the feasibility of an ASA simultaneously serving multiple (closely spaced) users using a simple criterion as a function of the ASA geometry and user locations, and (ii) three functional modules: user selection based on dynamic programming (DP) (§6.2), subarray allocation (§6.3), and optimized beamforming (§6.4).

We experimentally evaluate the performance of MAMBAS using the IBM 28 GHz phased array antenna module (PAAM) boards [25, 52] and software-defined radios (SDRs) deployed in the open-access PAWR COSMOS testbed [16, 51]. This setup allows for prototyping an ASA with two 8-element subarrays arranged in the azimuth plane, with a 5G NR FR2 Physical (PHY) layer and a baseband sampling rate of 122.88 MHz. We show that MAMBAS can effectively support two users with an angular spacing of only 5° , which is much smaller than the ASA's HPBW, while achieving improved data rates compared to existing methods. Through a case study where an ASA serves five closely spaced users, we show that MAMBAS achieves a sum rate gain of 1.4–1.6 $\times$ compared to analog SUBF while maintaining similar fairness across users. We also evaluate the performance of MAMBAS using large-scale simulation based on 28 GHz ray tracing propagation models, including scenarios where users lack LOS paths toward the BS and MAMBAS effectively leverages the strongest NLOS paths available. The simulation results show that for an ASA with eight 32-element subarrays serving eight users, MAMBAS achieves a median sum rate gain of 3.86 $\times$ and 1.92 $\times$ compared to analog SUBF and digital MUBF, respectively. More importantly, in network scenarios with increased user density, MAMBAS is able to achieve consistent performance while other digital MUBF methods completely fail.

To summarize, the main contributions of this paper are:

- We leverage the geometry of ASA to determine the criterion about the minimum angular spacing between nearby users (within the ASA's HPBW) in analog MUBF systems;
- We present the design of MAMBAS, a novel system that efficiently optimizes the user selection, subarray allocation, and beamforming in analog MUBF systems using an ASA;
- We extensively evaluate the performance of MAMBAS using an experimental testbed with 28 GHz SDRs and large-scale simulations based on ray tracing propagation models.

To the best of our knowledge, this is the first work exploring and optimizing analog MUBF to support concurrent multi-user transmissions in mmWave networks with closely spaced users, experimentally validated using advanced 28 GHz SDRs.

2 RELATED WORK

Digital MUBF at sub-6 GHz/mmWave frequencies. Massive MIMO systems employing all-digital MUBF have been well studied in FR1, including Argos [54], which built a 64-antenna BS serving up to 15 users, Arena [10], LuMaMi [41], and other many-antenna systems integrated in open-access testbeds, e.g., COSMOS [51] and POWDER-RENEW [11]. Recently, all-digital beamforming systems at mmWave frequencies have been demonstrated by industry and academia, but with only a small number of RF chains [18]. Building wideband, all-digital mmWave beamforming systems supporting a larger number of RF chains is still challenging due to the complex hardware and signal processing algorithms.

Analog beamforming at mmWave frequencies. Compared to all-digital beamforming, analog beamforming [5, 9] presents a more energy- and cost-effective technique for mmWave systems. In addition to conventional analog beamforming, adaptive phased arrays supporting programmable antenna weights provide the DoF for “manipulating” the beam patterns arbitrary beamforming to enhance received signal power [30], suppress/null interference [39, 44], extend the array FoV and coverage [4, 65], enable full-duplex operation [14, 36], or support radar/imaging and high-definition content delivery applications [26, 71, 73]. Most relevant to our work is Nulli-Fi [39], which uses a phased array to form deep nulls toward other users to mitigate the inter-beam interference from side lobes, but does not support concurrent multi-user transmissions with closely spaced users. In addition, lens antenna arrays [19, 69] present another effective method to support path division multiplexing [19, 69], facilitating directional communication in mmWave networks.

Multi-user support and user selection. Argos showed that the performance of digital ZF beamforming suffers from poor performance with a larger number of users and increased user density [54], which calls for the design of efficient user selection methods in sub-6 GHz [17, 56] and mmWave networks [24]. Recent works have also developed time/frequency domain scheduling algorithms to guarantee fairness across users [20, 21, 32, 34, 62, 67], and more relevantly, including hybrid MUBF co-designs [15, 50]. Most relevant to our work is S²-MAS [24], which proposed a joint user selection and beam steering algorithm for mmWave networks, where the users are selected only based on their spacing compared to the array’s HPBW and thus not optimized for improved SINR or networks with high user density.

Beam training and tracking. Beam training and tracking are two key components that obtain user location information (e.g., AoD) for establishing mmWave links and maintaining link quality in both cellular and Wi-Fi networks. There have been extensive recent works focusing on efficient and adaptive beam training and tracking in mmWave networks

for overcoming mobility and environmental changes (e.g., blockage) [27, 43, 58, 64, 66, 70]. These works are considered to be orthogonal to our work since we focus on multi-user support for achieving improved communication rates in mmWave networks, which can be incorporated with the existing beam training and tracking methods [23, 34, 45, 57, 64].

mmWave SDRs and testbeds. A number of mmWave SDRs and testbeds have emerged in the past years, mostly supporting the unlicensed 60 GHz band, including Open-Mili [72], X60 [53], MillimeTera [46], M-Cube [73], and MIMORPH [37]. In addition, a few SDRs have been developed supporting 5G NR FR2 at 28 GHz such as mMobile [31], which is a 28 GHz SDR with a single RF chain and a 64-element phased array. One of the most advanced programmable antenna arrays supporting MIMO capability is the IBM 28 GHz PAAM [52] that has recently been integrated with USRP SDRs in the open-access PAWR COSMOS testbed [16, 51], which we leverage in this work for the experimental evaluation of MAMBAS.

3 PRELIMINARIES

Let f_c and $\lambda = c/f_c$ (c : speed of light) respectively denote the signal carrier frequency and wavelength. Let $(\cdot)^T$ and $(\cdot)^*$ denote the transpose and complex conjugate operation.

Analog SUBF. Fig. 1(a) shows an all-analog SUBF system consisting of a phased array of N antennas connected to a single RF chain. Analog beamforming allows for achieving directional signal transmission/reception with enhanced signal power and/or spatial interference filtering by adjusting the analog beamforming weight of each antenna. Consider a uniform linear array (ULA) with N antennas arranged in the azimuth plane and an inter-antenna spacing of d . Let $s_n(\phi)$ be the phase delay experienced by a plane wave as it departs the n -th antenna relative to the first antenna in the direction of ϕ , given by $s_n(\phi) = e^{j\frac{2\pi(n-1)d}{\lambda} \sin \phi}$. The *steering vector* of the ULA in the direction of ϕ is given by

$$\mathbf{s}(\phi) = [s_n(\phi)] = [1, e^{j\frac{2\pi d}{\lambda} \sin \phi}, \dots, e^{j\frac{2\pi(N-1)d}{\lambda} \sin \phi}]^T \in \mathbb{C}^{N \times 1}.$$

Note that $\mathbf{s}(\phi)$ is determined by the phased array geometry.

Let $w_n \in \mathbb{C}$ be the weight applied to the n -th antenna. The beamforming weights for a phased array, $\mathbf{w} = [w_n]$, can be configured to alter the beam patterns to enhance signal power and/or suppress interference in the spatial directions. In practice, w_n can be controlled using variable gain attenuators and phase shifters, respectively. The far-field beam pattern in the direction of ϕ is given by $BP(\phi) = |\mathbf{s}^T(\phi) \cdot \mathbf{w}|^2$, and the *beamforming gain* is defined as the power gain of radiated signal normalized to the array input power, i.e.,

$$g(\phi) = BP(\phi)/N = |\mathbf{s}^T(\phi) \cdot \mathbf{w}|^2 / N. \quad (1)$$

Given a desired beamforming direction ϕ_0 (e.g., the AoD toward the user), it is known that (analog) *conjugate beamforming* can achieve a maximum beamforming gain that is

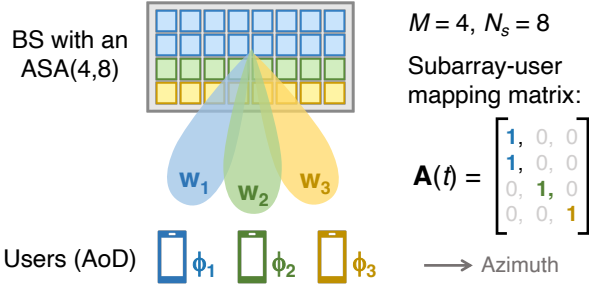


Figure 2: Network model: A BS equipped with an ASA simultaneously serves multiple spatially distributed users.

equal to the number of antennas in the phased array, i.e.,

$$\mathbf{w}_{\text{conj}} = \mathbf{s}^*(\phi_0) \Rightarrow g(\phi_0) = |\mathbf{s}^\top(\phi_0) \cdot \mathbf{s}^*(\phi_0)|^2 / N = N. \quad (2)$$

Digital MUBF. Fig. 1(b) shows an all-digital beamforming system consisting of an N -element antenna array connected to N RF chains, which serves U ($U \leq N$) single-antenna users via all-digital MUBF. In order to support concurrent transmissions of multiple data streams to different users, one needs to design the digital precoder, $\mathbf{W} \in \mathbb{C}^{N \times U}$, which precodes the U signal streams onto the N antennas. This signal model is given by $\mathbf{y} = \mathbf{H}^\top \mathbf{W} \mathbf{x} + \mathbf{n}$, where: (i) $\mathbf{x} \in \mathbb{C}^{U \times 1}$ is the signal intended for the U users, (ii) $\mathbf{y} \in \mathbb{C}^{U \times 1}$ and $\mathbf{n} \in \mathbb{C}^{U \times 1}$ are the received signal and noise at the users, and (iii) $\mathbf{H} = [h_{nu}] \in \mathbb{C}^{N \times U}$ is the CSI matrix, where h_{nu} denotes the complex-valued channel gain from the n -th antenna to the u -th user. Using $\mathbf{H}$ obtained via a channel sounding procedure [35], the linear *digital zero-forcing (ZF) beamforming* is commonly used in practical systems, which aims to minimize the cross-interference across the users by forming “nulls” toward unintended users, i.e.,

$$\mathbf{W}_{\text{zf}} = c_0 \cdot \mathbf{H}^* (\mathbf{H}^\top \mathbf{H}^*)^{-1}. \quad (3)$$

MAMBAS based on All-Analog MUBF. In this work, we explore all-analog MUBF and design MAMBAS, which uses an ASA to simultaneously serve multiple closely spaced users. Compared to digital MUBF, all-analog MUBF sacrifices (partial) beamforming flexibility to achieve lower energy consumption through a reduced number of RF chains, each of which serves an entire subarray instead of an individual antenna. Fig. 1(d) shows the diagram of an all-analog MUBF system. Since mmWave channels exhibit less significant multi-path effects compared to sub-6 GHz channels and the signal energy is usually focused in very few (LOS and/or NLOS) paths, an analog MUBF system can be designed to operate using *only* the AoD information toward the users.

4 NETWORK MODEL AND OBJECTIVE

We consider a BS equipped with an ASA(M, N_s) consisting of M identical subarrays, each of which is a phased array with N_s antennas (see Fig. 2). For simplicity, we assume that

each subarray is a ULA arranged along the azimuth plane with an inter-antenna spacing of d , and all the $N = MN_s$ antennas of the ASA form an $M \times N_s$ planar array. We assume that all antennas are homogeneous with the same maximum radiating power and each subarray has the same total input power. The steering vector of each subarray in the azimuth direction of ϕ , denoted by $\mathbf{s}_m(\phi) \in \mathbb{C}^{N_s \times 1}$, is

$$\mathbf{s}_m(\phi) = [1, e^{j\frac{2\pi d}{\lambda} \sin \phi}, \dots, e^{j\frac{2\pi(N_s-1)d}{\lambda} \sin \phi}]^\top. \quad (4)$$

Let $w_{m,n} = |w_{m,n}| \cdot e^{j\angle w_{m,n}}$ denote the beamforming weight for the n -th antenna in the m -th subarray. Let $\mathbf{w}_m = [w_{m,n}] \in \mathbb{C}^{N_s \times 1}$ and $\mathbf{w} = [\mathbf{w}_m] \in \mathbb{C}^{N_s \times M}$ be the beamforming weights for the m -th subarray and the entire ASA, respectively. The beam pattern determined by the ASA geometry can impact the received signal and interference power at each user. Such impact is more significant when the ASA simultaneously serves *multiple closely spaced users*, such as when the users are located within the HPBW of the ASA.

4.1 Signal, Interference, and Rate Models

We focus on the downlink transmission from a BS equipped with an ASA(M, N_s) to a set of U users, denoted by $\mathcal{U} = \{1, \dots, U\}$. Without loss of generality, we assume that all users reside in the same azimuth plane as the BS, and each user is equipped with a single RF chain. We use $\phi_{\mathcal{U}} = \{\phi_1, \dots, \phi_U\}$ to denote the azimuth AoD corresponding to the strongest (LOS or NLOS) path of individual users toward the BS. Note that our model can also be extended to scenarios where the BS and users reside in a 3-dimensional space.

Subarray allocation and beamforming. We consider *all-analog MUBF* using an ASA, whose subarray(s) can be assigned to serve each user, as shown in Figs. 1(d) and 2. Let $\mathcal{A}_u$ denote the set of subarray(s) allocated to user u with $N_u = |\mathcal{A}_u| \cdot N_s$ antennas. Let $\mathbf{A} = [a_{m,u}] \in \{0, 1\}^{M \times U}$ denote the *subarray-user mapping matrix*, where $a_{m,u} = 1$ if subarray m is allocated to user u , and is 0 otherwise. We further use $\mathbb{A}$ to denote all possible subarray-user mapping matrices, i.e., $\mathbf{A} \in \mathbb{A}$. With a little abuse of notation, let $\mathbf{s}_u = [\mathbf{s}_i]_{i \in \mathcal{A}_u} \in \mathbb{C}^{N_u \times 1}$ denote the concatenated steering vector associated with the set of subarray(s) assigned to user u , and $\mathbf{w}_u = [\mathbf{w}_i]_{i \in \mathcal{A}_u} \in \mathbb{C}^{N_u \times 1}$ denote the beamforming weights for the subarrays in $\mathcal{A}_u$. The *subarray beamforming gain* associated with subarrays $\mathcal{A}_u$ in the direction of ϕ , normalized to the single subarray input power, is given by

$$g_u(\mathbf{w}_u, \phi) = |\mathcal{A}_u| \cdot |\mathbf{s}_u^\top(\phi) \cdot \mathbf{w}_u|^2 / N_u = |\mathbf{s}_u^\top(\phi) \cdot \mathbf{w}_u|^2 / N_s, \quad (5)$$

where the factor of $|\mathcal{A}_u|$ is due to the fact that $|\mathcal{A}_u|$ subarrays can support a $|\mathcal{A}_u|$ times higher total input power in an ASA.

Signal, interference, and SNR/SINR. Assume that with a single antenna BS (without beamforming) and no interference, user u has a received signal power of $P_{\text{rx},u}$, which

corresponds to a *base SNR* of $\gamma_u = P_{\text{rx},u}/P_{\text{nf},u}$, where $P_{\text{nf},u}$ denotes the noise floor of user u . When the BS employs transmit beamforming using $|\mathcal{A}_u|$ subarrays with a beamforming gain of g_u (5), the SNR of user u can be written as

$$\text{SNR}_u(\mathbf{w}_u) = N_s \cdot g_u(\mathbf{w}_u, \phi_u) \cdot \gamma_u. \quad (6)$$

When multiple users are simultaneously served by different subarray(s), user u 's SINR as a function of the ASA beamforming weights $\mathbf{w}$ is given by

$$\begin{aligned} \text{SINR}_u(\mathbf{w}) &= \frac{N_s \cdot g_u^{\text{sig}}(\mathbf{w}_u, \phi_u) \cdot P_{\text{rx},u}}{P_{\text{nf},u} + \sum_{u' \neq u} N_s \cdot g_{u' \rightarrow u}^{\text{int}}(\mathbf{w}_{u'}, \phi_{u'}) \cdot P_{\text{rx},u}} \\ &= \frac{g_u^{\text{sig}}(\mathbf{w}_u, \phi_u) \cdot \gamma_u}{1/N_s + \sum_{u' \neq u} g_{u' \rightarrow u}^{\text{int}}(\mathbf{w}_{u'}, \phi_{u'}) \cdot \gamma_{u'}}, \end{aligned} \quad (7)$$

where the beamforming gains corresponding to the signal and interference toward user u , are given by

$$g_u^{\text{sig}}(\mathbf{w}_u, \phi_u) = |\mathbf{s}_u^\top(\phi_u) \cdot \mathbf{w}_u|^2 / N_s, \quad (8)$$

$$g_{u' \rightarrow u}^{\text{int}}(\mathbf{w}_{u'}, \phi_{u'}) = |\mathbf{s}_{u'}^\top(\phi_{u'}) \cdot \mathbf{w}_{u'}|^2 / N_s. \quad (9)$$

Consider $|w_{n,m}| \in [0, 1]$, the upper bound on user u 's SINR when allocated a single subarray (i.e., $N_u = N_s$) is equal to the maximum SNR that can be achieved by user u with conjugate beamforming and no interference, denoted by $\text{SNR}_u^{\text{max}}$, i.e.,

$$\text{SNR}_u(\mathbf{w}_u) \leq \frac{g_u^{\text{sig}}(\mathbf{w}_u, \phi_u) \cdot \gamma_u}{1/N_s} \leq N_s^2 \cdot \gamma_u := \text{SNR}_u^{\text{max}}. \quad (10)$$

User data rate. Consider a wireless bandwidth of B , the data rate for user u , denoted by $R_u(\mathbf{w})$, is given by

$$R_u(\mathbf{w}) = \begin{cases} B \cdot \log_2(1 + \text{SNR}_u(\mathbf{w})), & \text{w/o interference,} \\ B \cdot \log_2(1 + \text{SINR}_u(\mathbf{w})), & \text{w/ interference.} \end{cases} \quad (11)$$

Similarly, we can derive the upper-bound data rate R_u^{max} from the upper-bound SINR $\text{SINR}_u^{\text{max}}$, which is

$$R_u^{\text{max}} = B \cdot \log_2(1 + \text{SNR}_u^{\text{max}}). \quad (12)$$

4.2 Objective and Formulation

We assume that time is slotted ($t = 0, 1, \dots$) and let $\bar{R}_u$ denote user u 's long-term average data rate. We consider the utility function given by $\sum_{u \in \mathcal{U}} \log \bar{R}_u$, which aims to balance the overall network throughput and fairness between the users. It is well-known that under the *proportional fair (PF)* criteria, this utility function can be maximized by solving the following optimization problem in each slot, t [60]:

$$\max : \sum_{u \in \mathcal{U}} \frac{R_u(t)}{\bar{R}_u(t-1)}, \quad (13)$$

where $R_u(t)$ denotes user u 's achievable rate in slot t , and $\bar{R}_u(t-1)$ denotes user u 's (discounted) average rate up to slot $t-1$. Essentially, $\frac{R_u(t)}{\bar{R}_u(t-1)}$ represents the "priority" of user u and in each slot t , the PF scheduler selects the users that maximize (13) under some constraints (e.g., available

resources). We consider a $\bar{R}_u(t)$ is updated according to

$$\bar{R}_u(t) = (1 - \beta) \cdot R_u(t) + \beta \cdot \bar{R}_u(t-1), \quad (14)$$

where β is a coefficient that determines how fast the averaged rate decays over time and is set to $\beta = 0.5$ [42] in this work.

JUSB-PF in slot t

$$\{\mathcal{K}^\star(t), \mathbf{A}^\star(t), \mathbf{w}^\star(t)\} = \arg \max_{\mathcal{K}, \mathbf{A}, \mathbf{w}} \sum_{k \in \mathcal{K}} \frac{R_k(\mathbf{w}(t))}{\bar{R}_k(t-1)} \quad (15)$$

$$\text{s.t.} : \sum_{u \in \mathcal{U}} a_{m,u} \leq 1, \forall m, \mathcal{K} = \{k : \exists m, a_{m,k} = 1\}, \quad (16)$$

$$R_k(\mathbf{w}(t)) = \log_2(1 + \text{SINR}_k(\mathbf{w}(t))), \forall k \in \mathcal{K} \quad (17)$$

$$|w_{m,n}| \in [0, 1], \angle w_{m,n} \in [0, 2\pi], \forall m, n. \quad (18)$$

We adopt this PF objective and consider the *joint user selection, subarray allocation, and beamforming problem* (JUSB-PF) in slot t , whose inputs includes the ASA geometry (M, N_s, d) and user information ($\{\phi_u\}, \{\gamma_u\}, \{\bar{R}_u(t-1)\}$). In particular, (16) indicates that each subarray can be allocated to at most one user; (17) represents the achievable rate for each selected user given by (7); and (18) denotes the amplitude/phase constraint on the beamforming weights.

In general, solving JUSB-PF in each slot t is challenging and analytically intractable due to: (i) the exponentially large state space of user selection and subarray allocation, and (ii) the non-convex nature of the user SINR function (7). We refer to $\{\mathcal{K}^\star(t), \mathbf{A}^\star(t), \mathbf{w}^\star(t)\}$ as the (numerical) optimal solution to JUSB-PF in slot t returned by a solver. In particular, we enumerate through all possible (exponentially many) combinations of $\{\mathcal{K} \subseteq \mathcal{U}, \mathbf{A} \in \mathbb{A}\}$ and, for each $\{\mathcal{K}, \mathbf{A}\}$, we apply the CasADI optimization framework [6] to solve for the ASA beamforming weights $\mathbf{w}$ that maximize $\sum_{k \in \mathcal{K}} \frac{R_k(\mathbf{w}(t))}{\bar{R}_k(t-1)}$ in (15). Finally, the combination of $\{\mathcal{K}^\star(t), \mathbf{A}^\star(t), \mathbf{w}^\star(t)\}$ that yields the highest value of the objective (15) is recorded. Note that due to the exponentially large state space to be explored, JUSB-PF can be numerically solved within a practical timeframe only for small values of U and M , and it becomes computationally infeasible when $U \geq 4$ or $M \geq 4$.

5 MOTIVATION AND CHALLENGES

As a motivating example, consider an ASA(3, 8) that simultaneously serves 3 users at $f_c = 28$ GHz. The subarray antenna spacing is $d = 0.55 \lambda$, which matches with that of the IBM 28 GHz PAAM [52] used in our experiments (for detailed experimental setup see §7.1). We consider three *closely spaced* users with the same value of γ located at $\phi = \{-5^\circ, 0^\circ, +5^\circ\}$. In this scenario, all three users are located within the ASA's HPBW in the azimuth plane (12.3°). We consider four user selection and beamforming methods as illustrated in Fig. 3:

M1 *Analog SUBF with TDMA-PF*: The ASA serves a single user in each slot, selected based on the PF criteria [60], using analog beamforming with *all* antennas in the ASA;

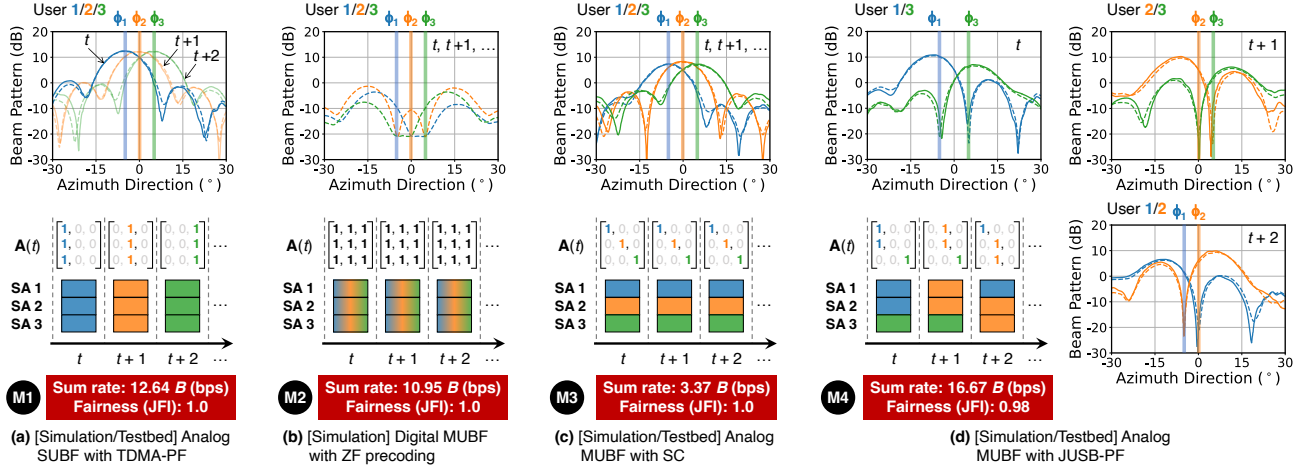


Figure 3: Motivating examples where an ASA(3, 8) communicates with three closely spaced users at $\phi = \{-5^\circ, 0^\circ, 5^\circ\}$. Compared to conventional analog and digital beamforming methods (a)–(c), the proposed joint user selection, subarray allocation, beamforming optimization employing an ASA (d) can achieve improved sum rate while maintaining fairness across users, at minimum system overhead. Simulated and measured beam patterns are indicated by dashed and solid lines, respectively.

- M2** *Digital MUBF with ZF precoding:* The ASA simultaneously serves all U users based on the ZF precoder obtained from the CSI matrix, $\mathbf{H}$ (see §3);
- M3** *Analog MUBF with subarray conjugate (SC) beamforming:* The ASA simultaneously serves U users, where each subarray applies conjugate beamforming toward one user;
- M4** *Analog MUBF with JUSB-PF (proposed):* The ASA simultaneously serves U users based on $\{\mathcal{K}^*(t), \mathbf{A}^*(t), \mathbf{w}^*(t)\}$, jointly optimized via numerically solving JUSB-PF.

For illustration purposes, we consider only a single line-of-sight (LOS) channel between each antenna of the ASA and each user. Note that **M1**, **M3**, and **M4** only requires the AoD information of the users (ϕ), while **M2** requires the CSI matrix ($\mathbf{H}$). For each method, the per-user data rate $R_u(t)$ is calculated using each user's SNR/SINR based on (11), from which the sum rate can be obtained. In §7.1, we will show the measured data rates using 28 GHz SDRs employing a 5G NR PHY layer. The fairness across all users over T slots is measured using Jain's fairness index (JFI) [33], given by

$$\text{JFI} = \left[\sum_u \left(\frac{1}{T} \sum_{t=1}^T R_u(t) \right) \right]^2 / \left[U \cdot \sum_u \left(\frac{1}{T} \sum_{t=1}^T R_u(t) \right)^2 \right]. \quad (19)$$

Fig. 3 shows the example subarray-user mapping matrix, $\mathbf{A}(t)$, which includes the user scheduling in three time slots. For the analog SUBF and MUBF methods, the simulated beam patterns of each subarray shown in Figs. 3(a), 3(c), and 3(d) match very well with that measured using the 28 GHz PAAM. Note that for digital MUBF with ZF, only the simulated beam patterns toward each user are illustrated in Fig. 3(b) since the 28 GHz PAAM does not support all-digital beamforming. The results show that **M1** achieves a higher sum rate of 12.64B bps compared to **M2** and **M3**, since the latter two methods suffer from significant inter-beam interference, *especially when the*

ASA serves multiple closely spaced users. For example, when applying digital MUBF with ZF precoding, the peak gain directions of the formed (digital) beams are “shifted” to other azimuth directions in order to preserve the nulls toward the users. This results in a sum rate of 10.95B bps with an equal SINR of 10.6 dB for each user and a JFI of 1.0. Interestingly, the proposed **M4** achieves a sum rate of 16.67B bps, a 1.32× improvement compared to **M1**, while maintaining a JFI of very close to 1. This is because the optimal solution returned by JUSB-PF (i) balances the selected users and subarray(s) allocated to each user based on historical rates, and (ii) guarantees that the SINR for each selected user is maximized via optimized beamforming at each subarray, which can be reflected by the subarray beam patterns shown in Fig. 3(d).

The motivating example demonstrates the benefits of jointly optimizing user selection, subarray allocation, and beamforming to achieve improved rates and fairness across users. However, solving JUSB-PF presents numerous challenges.

C1 Balancing rates and fairness. In MUBF systems, it is important that users are selected and served to maintain not only sufficient data rates but also fairness (e.g., based on the PF criteria). This becomes more critical for scenarios with closely spaced users, as their achievable rates highly depend on each other and the beamforming capability of the ASA.

C2 User selection and subarray allocation. First, for $\forall K \in \{1, \dots, U\}$, there are $\binom{U}{K}$, $\mathcal{K} \subseteq \mathcal{U}$ possible user selections with $|\mathcal{K}| = K$. As a result, there are an exponential number of possible user selections, i.e., $\sum_{K=1}^U \binom{U}{K} = O(2^U)$. Second, when the number of selected users is smaller than the number of subarrays, some user(s) can be allocated with

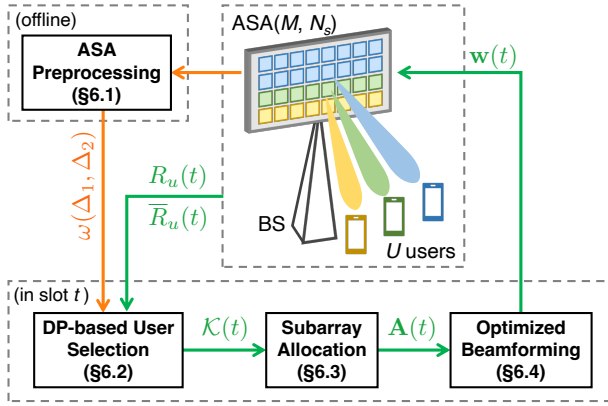


Figure 4: System overview and workflow of MAMBAS.

more subarrays to achieve improved SINR, and such allocation needs to be jointly optimized with the selected users.

Optimized beamforming. For given user selection $\mathcal{K}$ and subarray allocation $\mathbf{A}$, one needs to solve for the ASA's beamforming weights $\mathbf{w}$ in (15). However, this is a daunting task since $R_k(\mathbf{w}(t))$ is a non-convex function of many inter-dependent variables. For example, if K users are selected, a total number of $K \cdot (K - 1)$ inter-beam interference signals exist between each pair of users $(k, k'), \forall k' \neq k$ that need to be taken into account when calculating the SINR for each user (7). Essentially, the user SNR/SINR is traded off for placing deep nulls using the beamforming DoF.

The benefits and challenges outlined above motivate the design of MAMBAS, described next.

6 SYSTEM DESIGN

Fig. 4 shows the block diagram and workflow of MAMBAS including an offline ASA preprocessing step and three main modules: *user selection*, *subarray allocation*, and *optimized beamforming*. The key insight behind the design of MAMBAS is to *effectively decouple and independently optimize the user selection, subarray allocation, and beamforming* in JUSB-PF.

In the *offline* stage, the ASA preprocessing step explores the limit of an ASA and develops a simple criterion to determine if it can simultaneously serve a set of users. In the *online* stage, the *user selection* module leverages the ASA preprocessing step and dynamic programming (DP) to efficiently select a set of users, $\mathcal{K}(t)$, that the ASA can simultaneously serve with guaranteed data rates. The *subarray allocation* module then allocates the subarrays to the selected users based on their priority, i.e., determining $\mathbf{A}(t)$. Next, given $\{\mathcal{K}(t), \mathbf{A}(t)\}$, the *optimized beamforming* module efficiently solves for the optimized ASA beamforming weights, $\mathbf{w}$, by relaxing the non-convex objective in (15) to a convex optimization problem. Finally, the obtained $\{\mathcal{K}(t), \mathbf{A}(t), \mathbf{w}(t)\}$ is used for communication between the ASA and selected users in slot t , and the user priority is updated based on (14).

6.1 Offline ASA Preprocessing

Although the beamforming weights for each subarray, $\mathbf{w} \in \mathbb{C}^{N_s \times 1}$ present $(N_s - 1)$ DoF, there are still limitations in the beam pattern that can be generated given the subarray geometry as a function of N_s and d . For example, with small values of N_s , a null cannot be placed arbitrarily close to the main beam-pointing direction. Therefore, it is critical to understand how close the users can be when simultaneously served by an ASA with guaranteed sufficient data rates.

Can M subarrays simultaneously serve M users? We first conduct Monte Carlo simulations to answer this question considering an $\text{ASA}(M, N_s)$ and M users, where the m -th subarray is allocated to serve user m . Note that this is the maximum number of users that the ASA can simultaneously serve (see §4.1). We consider scenarios with varying user SNR between 10–30 dB and user density by randomly placing the M users within the azimuth range of $[-15^\circ, 15^\circ]$. For each scenario, the following optimization that maximizes the sum rate across all users is solved to obtain the optimal beamforming weights for the ASA, i.e.,

$$\begin{aligned} (\text{Max-Sum-Rate}) \mathbf{w}^* &= \max_{\mathbf{w}} : \sum_{m=1}^M R_m(\mathbf{w}) \quad (20) \\ \text{s.t.: } R_m(\mathbf{w}) &= \log_2(1 + \text{SINR}_m(\mathbf{w})), \forall m, \\ |w_{m,n}| &\in [0, 1], \angle w_{m,n} \in [0, 2\pi], \forall m, n. \end{aligned}$$

Fig. 5 shows the distribution of the data rate achieved by each user normalized to the maximum achievable data rate, i.e., $R_m(\mathbf{w}^*)/R_m^{\max}$. The results show that “blindly” using M subarrays to serve M users could yield poor performance: some users, especially those located in close proximity, achieve data rates that are very close to zero. This is because, for closely spaced users, it is extremely challenging to force the ASA to form deep nulls very close to the main beam-pointing direction while guaranteeing sufficient beamforming gain. This is due to the limited beamwidth and angular resolution of the ASA as a function of its geometry, including M , which determines the maximum number of users that can be simultaneously served, and N_s , which determines the number of DoF in analog beamforming. In addition, it is time-consuming to solve Max-Sum-Rate due to the non-convex expression of $R_m(\mathbf{w})$ as a function of $\text{SINR}_m(\mathbf{w})$, which includes $M \times (M - 1)$ pairs of inter-beam interference for each subarray-user pair.

Determine if a set of users can be simultaneously served by an ASA. Our goal is to develop a simple criterion that can determine if it is effective for the ASA to simultaneously serve M users. If not, it could be more beneficial for the ASA to serve only a subset of the users and allocate more subarrays to each user to achieve improved data rates, while the remaining users can be served in future time slots. Based on this intuition, we define a boolean function, Ω_M , which is

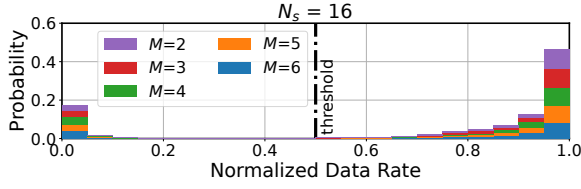


Figure 5: Normalized per-user data rate for M randomly distributed users and subarray size $N_s = 8$.

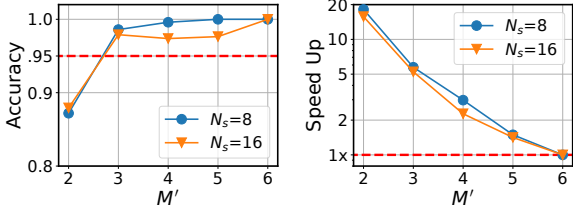


Figure 6: Accuracy and computational time speed up of the approximation $\Omega_M \approx \bigwedge \Omega_{M'}$ given by (22).

equal to 1 if each of the M users can achieve at least half of the maximum achievable data rate, i.e.,

$$\Omega_M(\phi_1, \dots, \phi_M) = \begin{cases} 1, & R_m(\mathbf{w}^*) \geq 0.5 \cdot R_m^{\max}, \forall m, \\ 0, & \text{otherwise.} \end{cases} \quad (21)$$

where $\mathbf{w}^*$ is the optimal solution to (20). However, Ω_M is a high-dimensional function that is computationally expensive to solve, as it involves solving (20) for each given set of users. Interestingly, Ω_M can be approximated based on two insights.

Insight 1: If M users can be simultaneously served by the ASA, then any M' ($M' < M$) consecutive users can also be simultaneously served by the ASA. Therefore, Ω_M can be approximated by the AND operation ‘ $\wedge$ ’ across $(M - M' + 1)$ possible $\Omega_{M'}$ for M' consecutive users, i.e.,

$$\begin{aligned} \Omega_M(\phi_1, \dots, \phi_M) &\approx \bigwedge_{i=0}^{M-M'} \Omega_{M'}(\phi_{i+1}, \dots, \phi_{i+M'}) \\ &= \Omega_{M'}(\phi_1, \dots, \phi_{M'}) \wedge \dots \wedge \Omega_{M'}(\phi_{M-M'+1}, \dots, \phi_M), \end{aligned} \quad (22)$$

where $\Omega_{M'}$ is the solution to (20) for the corresponding subset of M' consecutive users. To evaluate the approximation (22), we consider randomly placed $M = 6$ users and varying subarray sizes $N_s \in \{8, 16\}$, and conduct Monte Carlo simulations for $M' \in \{2, \dots, M-1\}$. Fig. 6 shows the accuracy of the approximation (22), i.e., $\Pr\{\Omega_M = \bigwedge \Omega_{M'}\}$. We empirically select $M' = 3$ and use Ω_3 to approximate Ω_M since it yields an accuracy of $>95\%$, which is comparable to that achieved by $M' \geq 4$, while providing an average runtime speed up of $5.5\times$. Note that for $M \geq 7$, it becomes computationally impractical to obtain Ω_M by numerically solving (20).

Insight 2: For any $M' = 3$ consecutive users, whether they should be simultaneously served by the ASA depends only on their relative locations and the subarray geometry. Intuitively, subarrays with a small number of antennas (N_s) cannot generate fine-grained beam patterns that yield sufficient SINR for each of the three users located in proximity. Our goal is

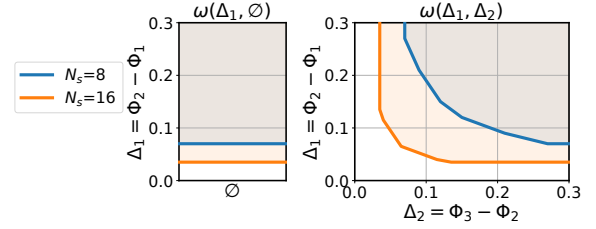


Figure 7: Illustration of $\omega(\Delta_1, \Delta_2)$ for a ULA with $N_s \in \{8, 16\}$, where the shaded areas indicate $\omega(\cdot, \cdot) = 1$.

to obtain the minimum user separation for three consecutive users $\{k-1, k, k+1\}$ such that $\Omega_3(\phi_{k-1}, \phi_k, \phi_{k+1})$ can be directly inferred based on the angular spacing between adjacent users without solving (20). Let $\Phi_k = \sin \phi_k$ denote the directional sine [61] associated to user k . We define a 2-dimensional function $\omega(\Delta_1, \Delta_2)$ which is the solution to (20) with $M = 3$ users located at $\{\phi_1, \phi_2, \phi_3\}$, where Δ_1 and Δ_2 denote the difference between the directional sine of two adjacent users ($\Delta_1 = \Phi_2 - \Phi_1$, $\Delta_2 = \Phi_3 - \Phi_2$), i.e.,

$$\omega(\Delta_1, \Delta_2) = \begin{cases} 1, & \text{if } \Omega_3(\phi_1, \phi_2, \phi_3) = 1, \\ 0, & \text{otherwise.} \end{cases} \quad (23)$$

Specifically, whether three users can be simultaneously served by the ASA or not can be immediately determined based on the pre-computed $\omega(\Delta_1, \Delta_2)$. Then, whether M users can be simultaneously served by the ASA can be determined via the approximation (22) with $M' = 3$. In the case with $M' = 2$, we can use $\omega(\Delta_1, \Delta_2)$ to determine if two users can be simultaneously served by the ASA. Fig. 7 illustrates $\omega(\Delta_1, \Delta_2)$ with different subarray sizes of $N_s \in \{8, 16\}$, where the shaded area corresponds to $\omega(\Delta_1, \Delta_2) = 1$. It can be seen that with a large value of N_s , the ASA can effectively support users with a smaller spacing as expected.

Remark. The offline ASA preprocessing module only needs to be pre-computed once by sweeping the values of Δ_1 and Δ_2 to obtain $\omega(\Delta_1, \Delta_2)$ for a given ASA. Here we present the ASA preprocessing module assuming that each subarray is a ULA, which is a commonly used antenna array geometry. However, our analysis can be adapted and extended to a more general and broader set of ASA and subarray geometries.

6.2 DP-based User Selection Algorithm

In the online stage, the *user selection module* selects the set of users to be served by the ASA(M, N_s) based on the preprocessing step and users’ historical rates. Let $\mathcal{K}(t) \subseteq \mathcal{U}$ denote the set of selected users in slot t , this module selects the set of users such that the sum of their priority is maximized, i.e.,

$$\mathcal{K}^*(t) = \arg \max_{\mathcal{K} \subseteq \mathcal{U}, |\mathcal{K}| \leq M} : \sum_{k \in \mathcal{K}} \frac{R_k^{\max}}{R_k(t-1)}, \text{ s.t.: } \Omega(\mathcal{K}) = 1. \quad (24)$$

In particular, we set the numerator in (24) to be $R_k^{\max}$ instead of $R_k(\mathbf{w}(t))$ as in (15), since $\mathbf{w}(t)$ is not readily available. This is also intuitive as (24) sets a higher priority on the users

Algorithm 1: DP-based user selection in slot t .

```

Input:  $\omega(\Delta_1, \Delta_2)$ ,  $\{\phi_u\}$ ,  $\{\bar{R}_u(t-1)\}$ 
1 for  $r = 1, \dots, U$  do
2   for  $q = 1, \dots, r-1$  do
3     if  $\omega(\Phi_r - \Phi_q, \text{null}) = 0$  then
4        $\text{DP}_{[q,r]}(t) \leftarrow \frac{R_r^{\max}}{\bar{R}_r(t-1)}$ ,  $\mathcal{K}_{[q,r]}(t) \leftarrow \{r\}$ 
5     else
6       for  $p = 1, \dots, q-1$  do
7          $\text{DP}_{[q,r]}(t) \leftarrow \frac{R_q^{\max}}{\bar{R}_q(t-1)} + \frac{R_r^{\max}}{\bar{R}_r(t-1)}$ 
8          $\mathcal{K}_{[q,r]}(t) \leftarrow \{q, r\}$ 
9         if  $\omega(\Phi_q - \Phi_p, \Phi_r - \Phi_q) = 1$  then
10           $\text{DP}_{[q,r]}(t) \leftarrow \text{DP}_{[p,q]}(t) + \frac{R_r^{\max}}{\bar{R}_r(t-1)}$ 
11           $\mathcal{K}_{[q,r]}(t) \leftarrow \mathcal{K}_{[p,q]}(t) \cup \{r\}$ 
12  $\{q^*, r^*\} = \arg \max_{1 \leq q < r \leq U} : \text{DP}_{[q,r]}, \mathcal{K}(t) \leftarrow \mathcal{K}_{[q^*, r^*]}(t)$ 
13 Sort  $\mathcal{K}(t)$  based on  $\{\frac{R_k^{\max}}{\bar{R}_k(t-1)}\}$  in the descending order
14 if  $|\mathcal{K}(t)| > M$  then
15    $\mathcal{K}(t) \leftarrow \mathcal{K}_{[1:M]}(t)$  (select the first  $M$  users in  $\mathcal{K}(t)$ )
16 return  $\mathcal{K}(t)$ 

```

that can “potentially” achieve a higher rate. In addition, the number of selected users should be no larger than the number of subarrays, i.e., $|\mathcal{K}| = M$. Based on the ASA preprocessing, $\Omega(\mathcal{K})$ can be approximated by the AND operation of multiple $\omega(\Delta_1, \Delta_2)$ for *any three consecutive users* $\{k-1, k, k+1\} \subseteq \mathcal{K}$. Thus, we can find an approximate solution to (24) by solving:

$$\tilde{\mathcal{K}}(t) = \arg \max_{\mathcal{K} \subseteq \mathcal{U}, |\mathcal{K}| \leq M} : \sum_{k \in \mathcal{K}} \frac{R_k^{\max}}{\bar{R}_k(t-1)}, \quad (25)$$

$$\text{s.t.} : \omega(\Phi_k - \Phi_{k-1}, \Phi_{k+1} - \Phi_k) = 1, \forall \{k-1, k, k+1\} \subseteq \mathcal{K}.$$

Although the constraint in (25) can be effectively solved given the ASA preprocessing step with $O(1)$ complexity (see Fig. 7), there are still a total number of $O(2^U)$ possible user selections since $|\mathcal{K}(t)|$ can vary from 1 to U .

We exploit the use of DP to efficiently solve (25) and Algorithm 1 presents the pseudocode for the DP-based user selection algorithm. Consider the sub-problem in DP that selects the set of users from $\mathcal{U}_r = \{1, 2, \dots, r\}$, $\forall r \leq U$. The key insight is that, based on the constraint in (25) determined by $\omega(\Delta_1, \Delta_2)$, if the optimal user selection for the sub-problem over $\mathcal{U}_r$ includes users q and r , where q is the last selected user prior to r , then it must include the optimal user selection for the sub-problem over $\mathcal{U}_q$, if both users p and q are selected with p being the last selected user prior to q .

We define the state $\text{DP}_{[q,r]}(t)$ of DP as the maximum sum user priority over the sub-problem defined on $\mathcal{U}_r$, where: (i) both users q and r are *simultaneously* selected, and (ii) user q is the *last* selected user prior to user r . This is given by

$$\text{DP}_{[q,r]}(t) = \begin{cases} \max_{\mathcal{K} \subseteq \mathcal{U}_r, |\mathcal{K}| \leq M} : \sum_{k \in \mathcal{K}} \frac{R_k^{\max}}{\bar{R}_k(t-1)}, & \text{if } \{q, r\} \subseteq \mathcal{K} \text{ and} \\ & \forall q < q' < r, q' \notin \mathcal{K}, \\ \frac{R_r^{\max}}{\bar{R}_r(t-1)}, & \text{otherwise.} \end{cases}$$

For the boundary case where users q and r cannot be simultaneously selected, we define $\text{DP}_{[q,r]}(t) = \frac{R_r^{\max}}{\bar{R}_r(t-1)}$, i.e., only user r is selected. We further define $\mathcal{K}_{[q,r]}(t)$ as the optimal user selection set associated with the DP state $\text{DP}_{[q,r]}(t)$ of the sub-problem. Now, consider another user p with $1 \leq p < q$. If users $\{p, q, r\}$ can be simultaneously selected, then $\text{DP}_{[q,r]}(t)$ can be recursively obtained by $\text{DP}_{[p,q]}(t)$ based on the DP state update given by

$$\text{DP}_{[q,r]}(t) = \max_{\substack{1 \leq p < q \\ \omega(\Phi_q - \Phi_p, \Phi_r - \Phi_q) = 1}} : \text{DP}_{[p,q]}(t) + \frac{R_r^{\max}}{\bar{R}_r(t-1)}, \quad (26)$$

$\mathcal{K}_{[q,r]}(t)$ and $\text{DP}_{[q,r]}(t)$ can be recursively obtained by (26). At termination, the selected users in $\mathcal{K}(t)$ are sorted based on their priority, $\{\frac{R_k^{\max}}{\bar{R}_k(t-1)}\}$ in the descending order. If $|\mathcal{K}(t)| > M$, the first M users with the highest priority will be selected.

Remark. In each slot, the DP-based user selection algorithm has a polynomial time complexity of $O(U^3)$, which is significantly lower than the exponential time complexity of $O(2^U)$ for the user selection in JUSB-PF (see §4.2).

6.3 Subarray Allocation

Next, the *subarray allocation module* assigns the M subarrays in $\text{ASA}(M, N_s)$ to the selected users in $\mathcal{K}(t)$. If $|\mathcal{K}(t)| < M$, the M subarrays are allocated to the $|\mathcal{K}(t)|$ selected users in a round-robin fashion based on their priority. This subarray allocation process is given by $\mathbf{A}(t) = [a_{m,k}(t)]$, where

$$a_{m,k}(t) = 1 \{k = m \bmod |\mathcal{K}(t)|, \forall k \in \mathcal{K}(t)\}, \quad (27)$$

$$\text{with } \mathcal{K}(t) = \{k \in \mathcal{U} : \exists m \text{ s.t. } a_{m,k}(t) = 1\}.$$

6.4 Optimized Beamforming

Given $\mathcal{K}(t)$ and $\mathbf{A}(t)$, the *optimized beamforming module* solves for the beamforming weights for all subarrays so that the sum priority across the selected users is maximized, i.e.,

$$\max_{\mathbf{w}} : \sum_{k \in \mathcal{K}(t)} \frac{R_k(\mathbf{w}(t))}{\bar{R}_k(t-1)} = \max_{\mathbf{w}} : \sum_{k \in \mathcal{K}(t)} \frac{\log_2(1 + \text{SINR}_k(\mathbf{w}(t)))}{\bar{R}_k(t-1)}. \quad (28)$$

However, the achievable rate for user k as a function of its SINR depends on not only the subarray(s) allocated to k , but also the subarray(s) allocated to other users $k' \neq k$, $k' \in \mathcal{K}(t)$ that cause interference (see (7)). Therefore, instead of directly solving the non-convex optimization (28), which is computationally expensive, we solve the following *convex optimization-based beamforming for each selected user*:

Max-Gain-ZF in slot t , $\forall k \in \mathcal{K}(t)$

$$\mathbf{w}_k = \arg \max_{\mathbf{w}_k} : g_k^{\text{sig}}(\mathbf{w}_k) = \arg \max_{\mathbf{w}_k} : |\mathbf{s}_k^\top(\phi_k) \cdot \mathbf{w}_k|^2 \quad (29)$$

$$\text{s.t.} : g_{k \rightarrow k'}^{\text{int}}(\mathbf{w}_k) = |\mathbf{s}_k^\top(\phi_{k'}) \cdot \mathbf{w}_k|^2 = 0, \forall k' \neq k, \quad (30)$$

$$\mathbf{w}_k = [\mathbf{w}_{m,n}], \forall m \text{ s.t. } a_{m,k} = 1, \forall n = 1, \dots, N_s. \quad (31)$$

Essentially, (31) represents the beamforming weights corresponding to the subarray(s) allocated to user k , which are optimized to maximize the beamforming gain toward user k (29) while simultaneously placing a null toward each of the other users $k' \neq k$ (30), e.g., analog ZF beamforming. Note that Max-Gain-ZF always has non-zero feasible solutions that satisfy (30) if the subarray size is larger than or equal to the number of selected users, i.e., $N_s \geq K$.

Remark. Max-Gain-ZF can be solved with guaranteed convergence due to its convex nature, more efficiently than the non-convex optimization-based beamforming employed in JUSB-PF. Furthermore, Max-Gain-ZF can be disjointed into K independent optimizations per selected user k , which can be solved in parallel for further acceleration. We employ CasADI [6] to solve Max-Gain-ZF, and the obtained beamforming weights $\mathbf{w}(t) = [\mathbf{w}_k(t)]$ are applied to the ASA.

7 PERFORMANCE EVALUATION

We extensively evaluate the performance of MAMBAS using a 28 GHz SDR testbed and large-scale simulations.

7.1 Experimental Setup and Scenarios

IBM 28 GHz PAAM board. We conduct experiments using the IBM 28 GHz PAAM boards deployed in the open-access PAWR COSMOS testbed [16]. To the best of our knowledge, this is the only front-end hardware available to the research community that supports experimentation with ASA in 5G NR FR2. The IBM 28 GHz PAAM board features an 8×8 antenna array with an inter-element spacing of 0.55λ , which includes four 4×4 subarrays, each connected to and controlled by one beamforming integrated circuit (IC) [25]. Each subarray has individual input/output ports at an intermediate frequency (IF) of 3 GHz. It supports programmable beamforming weights, $\mathbf{w}$, on individual antennas via a Python-based application programming interface (API), with (i) 5-bit amplitude control ($|w_n|$) between $[-12 \text{ dB}, 0 \text{ dB}]$, and (ii) phase control ($\angle w_n$) between $[-180^\circ, 180^\circ]$ with a resolution of 4.87° . To experiment with larger subarrays, two subarrays are tiled to form a 4×8 subarray with 8 antennas in the azimuth plane, corresponding to an azimuth HPBW of 12.3° .

Beam pattern measurements. We conduct beam pattern and beamforming gain measurements using an anechoic chamber, which houses a 28 GHz PAAM board and a horn antenna spaced by a link distance of 1.5 meters. The 28 GHz PAAM board is mounted on a positioner that allows for rotation in the azimuth plane with 0.1° resolution. In particular, a continuous-wave signal (@3 GHz) is generated using a signal generator and sent into the IF TX input port of the PAAM board. The signal received by the horn antenna (@28 GHz) is recorded by a spectrum analyzer and normalized to obtain the beamforming gain. This procedure is repeated as

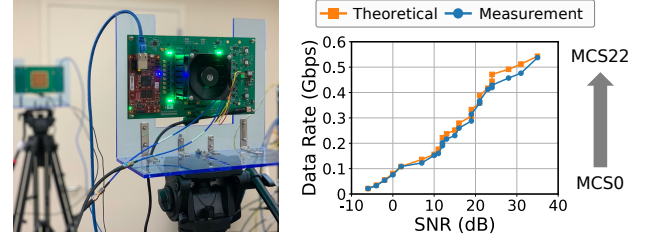


Figure 8: [Testbed] (Left) Two IBM 28 GHz PAAM boards deployed in the COSMOS testbed. (Right) Theoretical and measured data rate with varying link SNR and MCS.

the PAAM board rotates in the azimuth plane to obtain the transmit beam pattern, $BP(\phi)$, and beamforming gain, $g(\phi)$. For analog MUBF, the user SINR is calculated based on the measured $g(\phi)$ and (7), where γ is measured when only a single element is turned on. All our measurement results reveal a close matching between the measured and simulated beam patterns, stemming from the calibration-free nature and excellent beamforming capability of the 28 GHz PAAM.

28 GHz SDRs and data links. Each 28 GHz PAAM board is connected to a USRP N310 SDR to form a programmable 28 GHz radio. We use the C++-based USRP hardware driver (UHD) to configure the USRP and support streaming of baseband I/Q samples to/from the host server via a 10 Gbps SFP+ interface. The USRP is configured to operate at $f_c = 3 \text{ GHz}$ and is connected to the IF ports on the 28 GHz PAAM board. The host server also runs the API to initialize and configure the PAAM board, i.e., setting the beamforming weights $\mathbf{w}$.

We implemented OTA transmission between two 28 GHz SDRs using MATLAB's 5G toolbox [42]. In particular, we consider the 5G NR downlink physical data shared channel (PDSCH) with numerology 3 ($\mu = 3$ in FR2) and a subcarrier spacing of 120 kHz. We use a sampling rate of 122.88 MHz and an FFT size of 1,024, with 768 subcarriers (64 resource blocks) used to carry pilot/data, resulting in an occupied bandwidth of 92.16 MHz. The PHY layer supports 23 modulation and coding schemes (MCSs) for PDSCH [2] with LDPC encoding/decoding [13], from QPSK- $(\frac{120}{1,024})$ (MCS0, 21.6 Mbps) to 256QAM- $(\frac{754}{1,024})$ (MCS22, 0.54 Gbps). For each OTA experiment, the received signals captured by the USRP are analyzed to extract the link SNR and SINR for analog SUBF (without interference) and MUBF (with interference), respectively, and the corresponding block error rate (BLER) after LDPC decoding. For each SNR/SINR value, we empirically select the MCS that yields the highest data rate while guaranteeing a BLER of lower than 0.1 [63]. The measured link data rate as a function of the link SNR is shown in Fig. 8.

Performance metrics. We define the *SNR degradation* as the difference between the user SNR with a single subarray employing conjugate beamforming, $\text{SNR}_u^{\max}$ (10), and the SINR when the ASA employs analog MUBF (7). The *per-user*

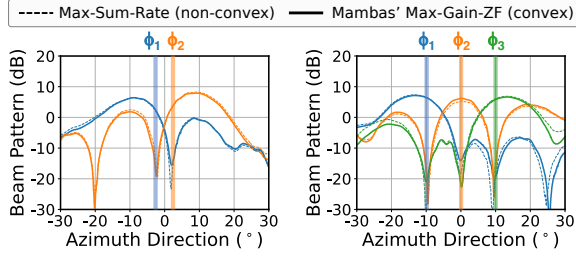


Figure 9: [Testbed] Measured beam patterns with beamforming weights obtained by MAMBAS' convex optimization.

rate and sum rate are measured using the 28 GHz SDRs with OTA transmissions, described above. We use JFI given by (19) to measure the fairness across users.

7.2 Benchmarking MAMBAS' Analog MUBF

ASA preprocessing and optimized analog MUBF. Fig. 9 shows the measured beam patterns of each subarray with two users spaced by 5° ($\phi = \{-2.5^\circ, 2.5^\circ\}$, $\Delta_1 \approx 0.09$) and three users spaced by 10° ($\phi = \{-10^\circ, 0^\circ, 10^\circ\}$, $\Delta_1 = \Delta_2 \approx 0.17$), where the beamforming weights are obtained by solving non-convex Max-Sum-Rate (dashed lines) and convex Max-Gain-ZF (solid lines), respectively. The results show an SINR difference of only <1.3 dB for each user between the measured beam patterns generated by the two methods. The measured SINR for the two users based on MAMBAS' optimized beamforming is 19.2 dB/18.6 dB, corresponding to an SNR degradation of 10.8 dB/11.4 dB repurposed to generate deep nulls, and a sum rate gain of $1.40\times$ compared to TDMA-PF. The null depth is also affected by hardware imperfection (e.g., quantized beamforming weights) and the slight angular deviation between the users' AoD and the actual 28 GHz PAAM's orientation controlled by the motor. Since the HPBW of an 8-element ULA is 12.3° , the two users would experience a much worse SINR of only 1.5 dB/3.8 dB under analog SC beamforming. These results match with our analysis in §6.1 and illustration of $\omega(\cdot, \cdot)$ in Fig. 7, demonstrating the effectiveness of MAMBAS' ASA preprocessing step. Note that the granularity of the formed beam patterns also relies on the quantization resolution of the amplitude/phase control of individual antennas in the 28 GHz PAAM.

Comparison to Nulli-Fi and SC beamforming. Fig. 10 shows the measured sum rates across two users located in the ASA broadside as a function of their angular spacing, with a per-user $\text{SNR}_u^{\max}$ of 20 dB and 30 dB. We consider four different beamforming methods at the ASA: (i) analog SUBF with TDMA, (ii) analog MUBF with SC, (iii) Nulli-Fi [39], and (iv) analog MUBF based on the solution to Max-Gain-ZF (29). In particular, Nulli-Fi is a method that assigns nulls by perturbing the beamforming weights from $\mathbf{w}_{\text{conj}}$ in (2), while preserving the beamforming gain. The horizontal dashed lines indicate the measured data rate achieved by analog

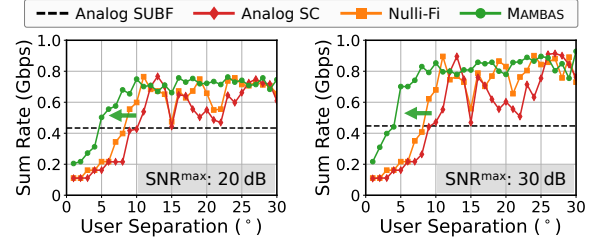


Figure 10: [Testbed] Measured sum rate with varying user angular separation under different beamforming methods.

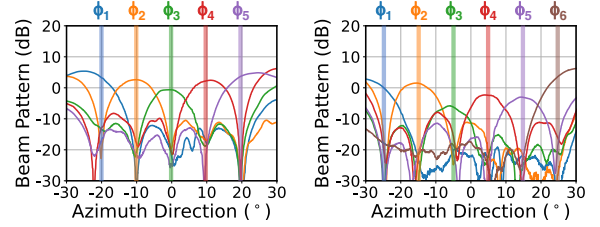


Figure 11: [Testbed] Measured beam patterns when an ASA simultaneously serves five users (left) and six users (right).

SUBF, which is independent of the user spacing since the ASA alternatively serves each user. The results show that MAMBAS starts to outperform analog SUBF when the user spacing is larger than 5° , which matches the threshold for $N_s = 8$ in Fig. 7. Note that Nulli-Fi only starts to outperform analog SUBF when the user spacing is larger than 9° since the null placement is not jointly optimized with the user SINR and the beamforming weights perturbation is limited to within a small range. The performance of analog MUBF with SC follows the conjugate beamforming pattern as expected: the maximum data rate is achieved when one user is located in the null direction of the subarray serving the other user. Overall, MAMBAS can achieve significantly improved data rates, especially in scenarios with very closely spaced users.

Scaling analog MUBF to support more users. Fig. 11 shows the measured beam pattern for each subarray when an ASA simultaneously serves $M = 5/6$ users with 10° spacing, using beamforming weights obtained via solving the convex Max-Gain-ZF. The results show that MAMBAS is capable of optimizing and balancing the SINR across all users. For example, the sum rate gain for $M = 6$ users compared to that achieved by the analog SUBF method (with TDMA) is $2.04\times$ with an average SNR degradation of 14.6 dB for each user, which is traded for placing five deep nulls toward other users. This is also a significant improvement in per-user SINR compared to that achieved by Nulli-Fi or analog SC.

7.3 End-to-End Evaluation: A Case Study

We use a case study with the 28 GHz PAAM boards to evaluate the end-to-end performance of MAMBAS. We consider a BS equipped with a 28 GHz PAAM as an ASA(2, 8) (limited by the 28 GHz PAAM IF ports, see §7.1) and five closely

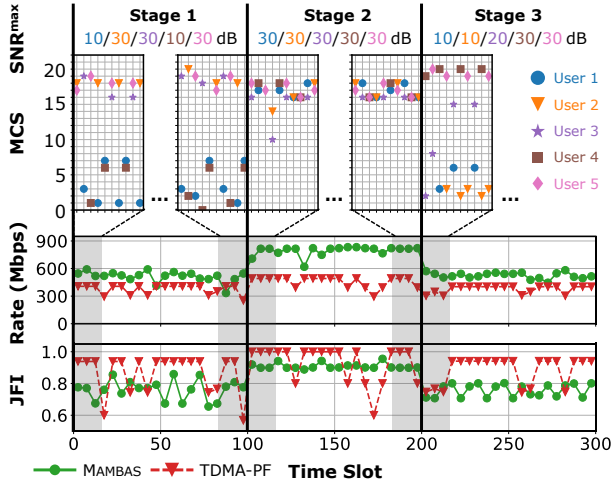


Figure 12: [Testbed] An ASA(2, 8) running MAMBAS serves five closely spaced users: detailed user selection and achieved MCS, as well as the sum rate and JFI across users.

spaced users with only 5° spacing between adjacent users, i.e., $\phi = \{-10^\circ, -5^\circ, 0^\circ, 5^\circ, 10^\circ\}$. In this setup, MAMBAS selects at most two users in each slot, whose data rates are measured when the 28 GHz PAAM simultaneously serves two users. We consider user $\text{SNR}_u^{\max}$ between 10–30 dB and run MAMBAS for 300 slots divided into three 100-slot stages, where the user $\text{SNR}_u^{\max}$ is changed during the transition from one stage to the next, as illustrated in Fig. 12, emulating changes of environmental factors and potential blockage.

Fig. 12 illustrates the user selection and the corresponding MCS for each user as a function of the SINR for slots 1–10, 91–110, and 191–210, during which the user $\text{SNR}_u^{\max}$ is changed. The results showcase MAMBAS’ capability in balancing the data rates achieved by different users based on their $\text{SNR}_u^{\max}$ and location with respect to each other. For example, when the $\text{SNR}_{1/2}^{\max}$ of users 1 and 2 drops from 30 dB to 10 dB in slot 200, MAMBAS selects these two users to be served by the ASA more frequently though they are only spaced by 5° . Fig. 12 also shows the sum rate and JFI across five users achieved by MAMBAS and TDMA-PF over time. The results show that, compared to TDMA-PF, MAMBAS achieves a sum rate gain of $1.44\times$, $1.58\times$, and $1.57\times$ across the three stages, with a JFI that is only 12.99%, 3.49%, and 17.07% lower. This demonstrates the effectiveness and benefits of MAMBAS via analog MUBF with optimized user selection.

7.4 Large-Scale Simulations

We consider an ASA with varying $M \in \{4, 8, 16\}$ and $N_s \in \{8, 16, 32\}$. We consider 112 $0.1 \text{ km} \times 0.1 \text{ km}$ areas in downtown Chicago and implement ray tracing model in MATLAB [68] with the OpenStreetMap geographic database [1, 38] to generate propagation environments at 28 GHz. For each area, we deploy a BS with an ASA at 30 m height above the ground,

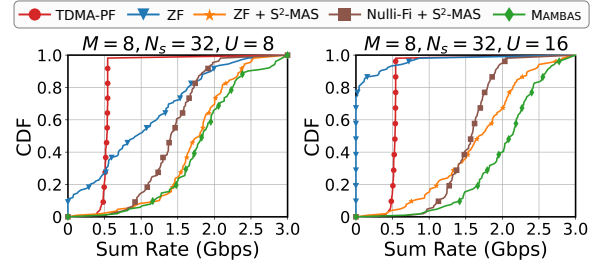


Figure 13: [Simulation] CDFs of sum rate achieved by MAMBAS and four baselines with an ASA(8, 32) serving 8/16 users.

and randomly place U users within the $\pm 30^\circ$ FoV of the BS; the distance between the BS and each user ranges within 20–100 m, resulting in $\text{SNR}_u^{\max}$ between 5–35 dB. In more than 84.2% of the cases, the angular spacing between adjacent users is less than the HPBW of the subarrays when $N_s = U$. For each BS-user pair, the 28 GHz ray tracing propagation model includes the AoD and path loss information as well as the multi-path effects with up to 1-hop reflection and/or diffraction. On average, there are 2.6 paths between each BS and each user, and 12.5% of the considered areas contain at least one user without a LOS path.

We compare the performance of MAMBAS with four baselines: (i) analog SUBF with TDMA-PF, (ii) digital MUBF with ZF precoding using full CSI provided by ray tracing. Note that traditional digital MUBF serves all users by default, which may result in degraded performance with high user density. Therefore, we also implemented: (iii) digital MUBF with ZF precoding and user selection based on S^2 -MAS [24], which selects users with a minimum separation of half of the HPBW, and (iv) Nulli-Fi [39] combined with user selection based on S^2 -MAS. Note that MAMBAS and Nulli-Fi rely on only the AoD information, and do not require CSI or employ digital beamforming (see §6). For the analog SUBF/MUBF schemes, the obtained beamforming weights are quantized to the amplitude/phase resolution of the 28 GHz PAAM. The per-user data rate is calculated based on Fig. 8 and the user SNR/SINR.

Fig. 13 plots the CDF of the sum rate achieved by MAMBAS and all four baselines, where an ASA(8, 32) serves $U = 8/16$ users. The results show that digital MUBF with ZF suffers from poor performance with relatively high user density, which can be mitigated by incorporating S^2 -MAS’ user selection algorithm. With this user selection algorithm, Nulli-Fi also significantly outperforms SUBF, yet it achieves lower sum rates in comparison to MUBF methods due to the lack of CSI and the inherent constraints associated with generating wide nulls. On the other hand, with reduced overhead, MAMBAS is still comparable to digital MUBF since the analog MUBF is optimized to achieve sufficient SINR for the selected users (see §6.4). In particular, for an ASA(8, 32) serving 16 users, MAMBAS achieves a median sum rate gain of $3.99\times$,

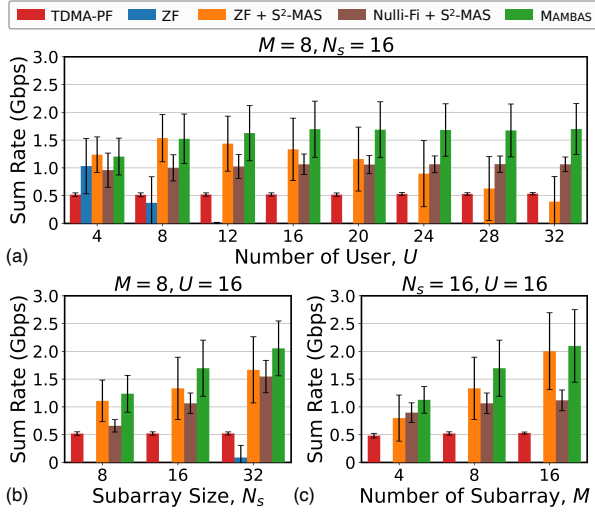


Figure 14: [Simulation] Sum rate achieved by MAMBAS and four baselines with varying ASA geometry and user density.

1.25 $\times$, and 1.34 $\times$ compared to analog SUBF, digital MUBF with S²-MAS, and Nulli-Fi with S²-MAS, respectively.

Fig. 14(a) shows the average sum rate achieved by MAMBAS and four baselines with an ASA(8, 16) serving $U \in \{4, \dots, 32\}$ users, all located within the $\pm 30^\circ$ FoV of the ASA. The performance of analog SUBF is independent of U as it serves a single user at a time. MAMBAS achieves consistently better performance than all four baselines, especially with increased user density. Note that even with the help of S²-MAS' user selection, digital MUBF experiences degraded performance with increased user density since digital ZF beamforming does not optimize for the SINR toward intended users, while Nulli-Fi remains effective over a larger number of users with lower sum rate compared to MAMBAS. Overall, in comparison to the best-performing baseline without the knowledge of CSI, i.e., Nulli-Fi, MAMBAS achieves a sum rate gain of 1.25–1.59 $\times$ with an ASA(8, 16) across scenarios with the varying number of users. Figs. 14(b) and 14(c) show the average sum rate achieved by MAMBAS and four baselines with 16 users served by an ASA with the varying (M, N_s) . It can be seen that an increased subarray size (N_s) always leads to improved sum rates, due to the narrower beamwidth and a larger number of beamforming DoF. Similarly, a larger number of subarrays (M) leads to improved sum rates since the ASA can simultaneously serve more users and assign more subarrays to each selected user.

8 DISCUSSIONS AND LIMITATIONS

Computation overhead. 5G NR FR2 with higher numerology exhibits a much shorter time slot compared to FR1, therefore requiring efficient beam training and tracking as well as real-time baseband processing [47]. MAMBAS integrates three

online modules that operate sequentially that introduce additional computation overhead. Notably, the system employs DP for user selection and a round-robin method for subarray allocation, both of which are with minimal computational complexity, as well as convex optimization-based beamforming, which can be independently processed for each selected user. It is important to note that the term “time slot” in our model does not directly correspond to the conventional definition of a slot in 5G NR terminology. Instead, MAMBAS offers flexibility in adjusting its operation at different time scales based on the rate at which the proportional fairness among users needs to be updated.

AoD estimation error. MAMBAS assumes perfect user AoD information obtained through the beam training and tracking phase. However, this process may incur AoD estimation errors, leading to potential misalignments between the beamforming/nulling and user directions. As a result, the interference experienced by unintended users, for whom nulls are strategically placed, can be impacted. To mitigate these challenges, one approach is to incorporate a constraint in the beamforming optimization (Max-Gain-ZF) that forces a certain null width to compensate for angle estimation errors, similar to that employed by Nulli-Fi [39]. Another approach involves a local fine-tuning procedure that slightly perturbs the null direction to achieve optimized SINR at each user.

9 CONCLUSION

We presented MAMBAS, an analog MUBF system supporting efficient multi-user communication in mmWave networks using an ASA, where users can be located in very close proximity. Specifically, MAMBAS consists of an offline ASA preprocessing step and three key modules responsible for efficient user selection, subarray allocation, and optimized beamforming. We extensively evaluated the performance of MAMBAS using both a 28 GHz SDR testbed and large-scale simulations, and show that it achieves significantly improved sum rates across users while maintaining good fairness, especially in scenarios with high user densities.

ACKNOWLEDGMENTS

This work was supported in part by NSF grants CNS-1827923, CNS-2211944, AST-2232458, and an IBM Academic Award. This work was also supported in part by the Center for Ubiquitous Connectivity (CUBiC), sponsored by Semiconductor Research Corporation (SRC) and Defense Advanced Research Projects Agency (DARPA) under the JUMP 2.0 program. We thank Arun Paidimarri, Bodhisatwa Sadhu, and Alberto Valdes-Garcia (IBM Research), and Ivan Seskar (Rutgers University) for their contributions to this work. We also thank Abhi Adhikari, Kevin Hermstein, and Yixin Liang for their help with the measurements.

REFERENCES

- [1] 2023. Open Street Map. <https://www.openstreetmap.org/>.
- [2] 2023. Share Technote. www.sharetechnote.com.
- [3] 3GPP. 2020. *5G; NR; Base Station (BS) Radio Transmission and Reception*. Technical Specification (TS) 38.211. 3rd Generation Partnership Project (3GPP). https://www.etsi.org/deliver/etsi_ts/138100_138199/138104/16.04.00_60/ts_138104v160400p.pdf Version 16.4.0.
- [4] Omid Abari, Dinesh Bharadia, Austin Duffield, and Dina Katabi. 2017. Enabling high-quality untethered virtual reality. In *Proc. USENIX NSDI'17*.
- [5] Ben Allen and Mohammad Ghavami. 2005. *Adaptive array systems: fundamentals and applications*. John Wiley & Sons.
- [6] Joel A E Andersson, Joris Gillis, Greg Horn, James B Rawlings, and Moritz Diehl. 2019. CasADi – A software framework for nonlinear optimization and optimal control. *Springer MPC* 11, 1 (2019), 1–36.
- [7] Ehsan Aryafar, Narendra Anand, Theodoros Salonidis, and Edward W Knightly. 2010. Design and experimental evaluation of multi-user beamforming in wireless LANs. In *Proc. ACM MobiCom'10*.
- [8] Hany Assasa, Joerg Widmer, Jian Wang, Tanguy Ropitault, and Nada Golmie. 2019. An Implementation Proposal for IEEE 802.11 ay SU/MU-MIMO Communication in ns-3. In *Proc. ACM WNS3'19*.
- [9] Constantine A Balanis. 2016. *Antenna theory: Analysis and design*. John Wiley & Sons.
- [10] Lorenzo Bertizzolo, Leonardo Bonati, Emrehan Demirors, Amani Al-Shawabka, Salvatore D'Oro, Francesco Restuccia, and Tommaso Melodia. 2020. Arena: A 64-antenna SDR-based ceiling grid testing platform for sub-6 GHz 5G-and-beyond radio spectrum research. *Computer Networks* 181 (2020), 107436.
- [11] Joe Breen, Andrew Buffmire, Jonathon Duerig, Kevin Dutt, Eric Eide, Mike Hibler, David Johnson, Sneha Kumar Kasera, Earl Lewis, Dustin Maas, et al. 2020. POWDER: Platform for open wireless data-driven experimental research. In *Proc. ACM WiNTECH'20*.
- [12] Sherif Adeshina Busari, Kazi Mohammed Saidul Huq, Shahid Mumtaz, Linglong Dai, and Jonathan Rodriguez. 2017. Millimeter-wave massive MIMO communication for future wireless systems: A survey. *IEEE Communications Surveys & Tutorials* 20, 2 (2017), 836–869.
- [13] Jinghu Chen, R Michael Tanner, Christopher Jones, and Yan Li. 2005. Improved min-sum decoding algorithms for irregular LDPC codes. In *Proc. IEEE ISIT'05*.
- [14] Tingjun Chen, Mahmood Baraani Dastjerdi, Harish Krishnaswamy, and Gil Zussman. 2019. Wideband full-duplex phased array with joint transmit and receive beamforming: Optimization and rate gains. In *Proc. ACM MobiHoc'19*.
- [15] Tingjun Chen, Jelena Diakonikolas, Javad Ghaderi, and Gil Zussman. 2020. Hybrid scheduling in heterogeneous half-and full-duplex wireless networks. *IEEE/ACM Trans. Netw.* 28, 2 (2020), 764–777.
- [16] Tingjun Chen, Prasanthi Maddala, Panagiotis Skrimponis, Jakub Kolodziej, Abhishek Adhikari, Hang Hu, Zhihui Gao, Arun Paidimarri, Alberto Valdes-Garcia, Myung Lee, et al. 2023. Open-access millimeter-wave software-defined radios in the PAWR COSMOS testbed: Design, deployment, and experimentation. *Computer Networks* 234 (2023), 109922.
- [17] Zhe Chen, Xu Zhang, Sulei Wang, Yuedong Xu, Jie Xiong, and Xin Wang. 2017. BUSH: Empowering large-scale MU-MIMO in WLANs with hybrid beamforming. In *Proc. IEEE INFOCOM'17*.
- [18] Aditya Dhananjay, Kai Zheng, Marco Mezzavilla, Lorenzo Iotti, Dennis Shasha, and Sundeep Rangan. 2021. Pi-Radio v1: Calibration techniques to enable fully-digital beamforming at 60 GHz. *Computer Networks* 196 (2021), 108220.
- [19] Xinyu Gao, Linglong Dai, Shuangfeng Han, I Chih-Lin, and Xiaodong Wang. 2017. Reliable beamspace channel estimation for millimeter-wave massive MIMO systems with lens antenna array. *IEEE Trans. Wireless Commun.* 16, 9 (2017), 6010–6021.
- [20] Zhihui Gao, Ang Li, Yunfan Gao, Bing Li, Yu Wang, and Yiran Chen. 2021. FedSwap: A federated learning based 5G decentralized dynamic spectrum access system. In *Proc. IEEE/ACM ICCAD'21*.
- [21] Zhihui Gao, Ang Li, Yunfan Gao, Yu Wang, and Yiran Chen. 2021. Hermes: Decentralized dynamic spectrum access system for massive devices deployment in 5G. In *Proc. ACM EWSN'21*.
- [22] Yasaman Ghasempour, Claudio RCM Da Silva, Carlos Cordeiro, and Edward W Knightly. 2017. IEEE 802.11 ay: Next-generation 60 GHz communication for 100 Gb/s Wi-Fi. *IEEE Commun. Mag.* 55, 12 (2017), 186–192.
- [23] Yasaman Ghasempour, Muhammad K Haider, Carlos Cordeiro, Dimitrios Koutsonikolas, and Edward Knightly. 2018. Multi-stream beam-training for mmWave MIMO networks. In *Proc. ACM MobiCom'18*.
- [24] Yasaman Ghasempour, Muhammad Kumail Haider, and Edward W Knightly. 2018. Decoupling beam steering and user selection for MU-MIMO 60-GHz WLANs. *IEEE/ACM Trans. Netw.* 26, 5 (2018), 2390–2403.
- [25] Xiaoxiong Gu, Arun Paidimarri, Bodhisatwa Sadhu, Christian Baks, Stanislav Lukashov, Mark Yeck, Young Kwark, Tingjun Chen, Gil Zussman, Ivan Seskar, et al. 2021. Development of a compact 28-GHz software-defined phased array for a city-scale wireless research testbed. In *Proc. IEEE IMS'21*.
- [26] Junfeng Guan, Arun Paidimarri, Alberto Valdes-Garcia, and Bodhisatwa Sadhu. 2021. 3-D imaging using millimeter-wave 5G signal reflections. *IEEE Trans. Microw. Theory Tech.* 69, 6 (2021), 2936–2948.
- [27] Haitham Hassanieh, Omid Abari, Michael Rodriguez, Mohammed Abdelghany, Dina Katabi, and Piotr Indyk. 2018. Fast millimeter wave beam alignment. In *Proc. ACM SIGCOMM'18*.
- [28] Wei Hong, Zhi Hao Jiang, Chao Yu, Debin Hou, Haiming Wang, Chong Guo, Yun Hu, Le Kuai, Yingrui Yu, Zhengbo Jiang, et al. 2021. The role of millimeter-wave technologies in 5G/6G wireless communications. *IEEE J. Microw.* 1, 1 (2021), 101–122.
- [29] IEEE Computer Society. 2012. IEEE Standard for Information technology – 802.11ad/ay. *IEEE Std 802.11* (2012).
- [30] Ish Kumar Jain, Raghav Subbaraman, and Dinesh Bharadia. 2021. Two beams are better than one: Towards reliable and high throughput mmWave links. In *Proc. ACM SIGCOMM'21*.
- [31] Ish Kumar Jain, Raghav Subbaraman, Tejas Harekrishna Sadarahlali, Xiangwei Shao, Hou-Wei Lin, and Dinesh Bharadia. 2020. mMobile: Building a mmWave testbed to evaluate and address mobility effects. In *Proc. ACM mmNets'20*.
- [32] Ish Kumar Jain, Rohith Reddy Vennam, Raghav Subbaraman, and Dinesh Bharadia. 2023. mmFlexible: Flexible Directional Frequency Multiplexing for Multi-user mmWave Networks. (2023).
- [33] Raj Jain, Arjan Durrresi, and Gojko Babic. 1999. Throughput fairness index: An explanation. In *ATM Forum Contribution*.
- [34] Suraj Jog, Jiaming Wang, Junfeng Guan, Thomas Moon, Haitham Hassanieh, and Romit Roy Choudhury. 2019. Many-to-many beam alignment in millimeter wave networks. In *Proc. USENIX NSDI'19*.
- [35] Florian Kaltenberger, Marios Kountouris, David Gesbert, and Raymond Knopp. 2009. On the trade-off between feedback and capacity in measured MU-MIMO channels. *IEEE Trans. Wireless Commun.* 8, 9 (2009), 4866–4875.
- [36] Kenneth E Kolodziej, Brian A Janice, Adrienne I Sands, and Bradley T Perry. 2023. Scalable in-band full-duplex phased arrays: Complexity reduction and distributed processing. *IEEE J. Sel. Areas Commun.* (2023).

- [37] Jesus O Lacruz, Rafael Ruiz Ortiz, and Joerg Widmer. 2021. A real-time experimentation platform for sub-6 GHz and millimeter-wave MIMO systems. In *Proc. ACM MobiSys'21*.
- [38] Yiming Li, Zeyu Li, Zhihui Gao, and Tingjun Chen. 2023. Geo2SigMap: High-fidelity RF signal mapping using geographic databases. *arXiv preprint arXiv:2312.14303* (2023).
- [39] Sohrab Madani, Suraj Jog, Jesús Omar Lacruz, Joerg Widmer, Haitham Hassanieh, et al. 2021. Practical null steering in millimeter wave networks. In *Proc. USENIX NSDI'21*.
- [40] Mohammad Majidzadeh, Aleksi Moilanen, Nuutti Tervo, Harri Pennanen, Antti Tolli, and Matti Latva-aho. 2017. Partially connected hybrid beamforming for large antenna arrays in multi-user MISO systems. In *Proc. IEEE PIMRC'17*.
- [41] Steffen Malkowsky, Joao Vieira, Liang Liu, Paul Harris, Karl Nieman, Nikhil Kundargi, Ian C Wong, Fredrik Tufvesson, Viktor Öwall, and Ove Edfors. 2017. The world's first real-time testbed for massive MIMO: Design, implementation, and validation. *IEEE Access* 5 (2017), 9073–9088.
- [42] MATLAB 5G Toolbox 2020. MATLAB 5G Toolbox.
- [43] Mohammad Hossein Mazaheri, Alex Chen, and Omid Abari. 2021. mmTag: A millimeter wave backscatter network. In *Proc. ACM SIGCOMM'21*.
- [44] Arun Paidimarri and Bodhisatwa Sadhu. 2020. Spatio-temporal filtering: Precise beam control using fast beam switching. In *Proc. IEEE RFIC'20*.
- [45] Joan Palacios, Daniel Steinmetzer, Adrian Loch, Matthias Hollick, and Joerg Widmer. 2018. Adaptive codebook optimization for beam training on off-the-shelf IEEE 802.11 ad devices. In *Proc. ACM MobiCom'18*.
- [46] Michele Polese, Francesco Restuccia, Abhimanyu Gosain, Josep Jornet, Shubhendu Bhardwaj, Viduneth Ariyaratna, Soumyajit Mandal, Kai Zheng, Aditya Dhananjay, Marco Mezzavilla, et al. 2019. MillimeTera: Toward a large-scale open-source mmWave and terahertz experimental testbed. In *Proc. ACM mmNets'19*.
- [47] Zhenzhou Qi, Zhihui Gao, Chung-Hsuan Tung, and Tingjun Chen. 2023. Programmable millimeter-wave MIMO radios with real-time baseband processing. In *ACM WiNTECH'23*.
- [48] Sundeep Rangan, Theodore S Rappaport, and Elza Erkip. 2014. Millimeter-wave cellular wireless networks: Potentials and challenges. *Proc. IEEE* 102, 3 (2014), 366–385.
- [49] Theodore S Rappaport, Shu Sun, Rimma Mayzus, Hang Zhao, Yaniv Azar, Kevin Wang, George N Wong, Jocelyn K Schulz, Mathew Samimi, and Felix Gutierrez. 2013. Millimeter wave mobile communications for 5G cellular: It will work! *IEEE Access* 1 (2013), 335–349.
- [50] Vishnu V Ratnam, Andreas F Molisch, Ozgun Y Bursalioglu, and Haralabos C Papadopoulos. 2018. Hybrid beamforming with selection for multiuser massive MIMO systems. *IEEE Trans. Signal Process.* 66, 15 (2018), 4105–4120.
- [51] Dipankar Raychaudhuri, Ivan Seskar, Gil Zussman, Thanasis Korakis, Dan Kilper, Tingjun Chen, Jakub Kolodziejski, Michael Sherman, Zoran Kostic, Xiaoxiong Gu, et al. 2020. Challenge: COSMOS: A city-scale programmable testbed for experimentation with advanced wireless. In *Proc. ACM MobiCom'20*.
- [52] Bodhisatwa Sadhu, Yahya Tousi, Joakim Hallin, Stefan Sahl, Scott K Reynolds, Örfan Renström, Kristoffer Sjögren, Olov Haapalahti, Nadav Mazar, Bo Bokinge, et al. 2017. A 28-GHz 32-element TRX phased-array IC with concurrent dual-polarized operation and orthogonal phase and gain control for 5G communications. *IEEE J. Solid-State Circuits* 52, 12 (2017), 3373–3391.
- [53] Swetank Kumar Saha, Yasaman Ghasempour, Muhammad Kumail Haider, Tariq Siddiqui, Paulo De Melo, Neerad Somanchi, Luke Zakrajsek, Arjun Singh, Owen Torres, Daniel Uvaydov, et al. 2017. X60: A programmable testbed for wideband 60 GHz WLANs with phased arrays. In *Proc. ACM WiNTECH'17*.
- [54] Clayton Shepard, Hang Yu, Narendra Anand, Erran Li, Thomas Marzetta, Richard Yang, and Lin Zhong. 2012. Argos: Practical many-antenna base stations. In *Proc. ACM MobiCom'12*.
- [55] Ali Sufyan, Khan Bahadar Khan, Osama A Khashan, Talha Mir, and Usama Mir. 2023. From 5G to beyond 5G: A comprehensive survey of wireless network evolution, challenges, and promising technologies. *Electronics* 12, 10 (2023), 2200.
- [56] Sanjib Sur, Ioannis Pefkianakis, Xinyu Zhang, and Kyu-Han Kim. 2016. Practical MU-MIMO user selection on 802.11 ac commodity networks. In *Proc. ACM MobiCom'16*.
- [57] Sanjib Sur, Ioannis Pefkianakis, Xinyu Zhang, and Kyu-Han Kim. 2018. Towards scalable and ubiquitous millimeter-wave wireless networks. In *Proc. ACM MobiCom'18*.
- [58] Sanjib Sur, Xinyu Zhang, Parmesh Ramanathan, and Ranveer Chandra. 2016. BeamSpy: Enabling robust 60 GHz links under blockage. In *Proc. USENIX NSDI'16*.
- [59] A Lee Swindlehurst, Ender Ayanoglu, Payam Heydari, and Filippo Capolino. 2014. Millimeter-wave massive MIMO: The next wireless revolution? *IEEE Commun. Mag.* 52, 9 (2014), 56–62.
- [60] SSA Tiwari, S Suman, and P Viswanathan. 2014. Long term evolution (LTE) protocol verification of MAC scheduling algorithms in NetSim. *Tetcos, Bengaluru, India, Tetcos White Paper* (2014).
- [61] David Tse and Pramod Viswanath. 2005. *Fundamentals of wireless communication*. Cambridge University Press.
- [62] Kaidong Wang and Konstantinos Psounis. 2018. Scheduling and resource allocation in 802.11 ax. In *Proc. IEEE INFOCOM'18*.
- [63] Lifang Wang. 2021. Implementation of Low-Density Parity-Check codes for 5G NR shared channels.
- [64] Song Wang, Jingqi Huang, and Xinyu Zhang. 2020. Demystifying millimeter-wave V2X: Towards robust and efficient directional connectivity under high mobility. In *Proc. ACM MobiCom'20*.
- [65] Song Wang, Jingqi Huang, Xinyu Zhang, Hyoil Kim, and Sujit Dey. 2020. X-array: Approximating omnidirectional millimeter-wave coverage using an array of phased arrays. In *Proc. ACM MobiCom'20*.
- [66] Timothy Woodford, Xinyu Zhang, Eugene Chai, Karthikeyan Sundaresan, and Amir Khojastepour. 2021. Spacebeam: Lidar-driven one-shot mmwave beam management. In *Proc. ACM MobiSys'21*.
- [67] Dingwen Yuan, Hsuan-Yin Lin, Jörg Widmer, and Matthias Hollick. 2020. Optimal and approximation algorithms for joint routing and scheduling in millimeter-wave cellular networks. *IEEE/ACM Trans. Netw.* 28, 5 (2020), 2188–2202.
- [68] Zhengqing Yun and Magdy F Iskander. 2015. Ray tracing for radio propagation modeling: Principles and applications. *IEEE Access* 3 (2015), 1089–1100.
- [69] Yong Zeng and Rui Zhang. 2016. Millimeter wave MIMO with lens antenna array: A new path division multiplexing paradigm. *IEEE Trans. Commun.* 64, 4 (2016), 1557–1571.
- [70] Ding Zhang, Panneer Selvam Santhalingam, Parth Pathak, and Zizhan Zheng. 2023. CoBF: Coordinated beamforming in dense mmWave networks. *Proc. ACM Meas. Anal. Comput. Syst.* 7, 2 (2023), 1–26.
- [71] Ding Zhang, Puqi Zhou, Bo Han, and Parth Pathak. 2022. M5: Facilitating multi-user volumetric content delivery with multi-lobe multicast over mmWave. In *Proc. ACM SenSys'22*.
- [72] Jialiang Zhang, Xinyu Zhang, Pushkar Kulkarni, and Parameswaran Ramanathan. 2016. OpenMili: A 60 GHz software radio platform with a reconfigurable phased-array antenna. In *Proc. ACM MobiCom'16*.
- [73] Renjie Zhao, Timothy Woodford, Teng Wei, Kun Qian, and Xinyu Zhang. 2020. M-Cube: A millimeter-wave massive MIMO software radio. In *Proc. ACM MobiCom'20*.