Scaling mmWave WLANs With Single RF Chain Multiuser Beamforming

Keerthi Priya Dasala, Josep M. Jornet, *Senior Member, IEEE*, and Edward W. Knightly, *Fellow, IEEE*

***Abstract*— Multi-user transmission in 60 GHz Wi-Fi can achieve data rates up to 100 Gbps by multiplexing multiple user data streams. However, a fundamental limit in the approach is that each RF chain is limited to supporting one stream or one user. In this paper, we scale multi-user 60 GHz WLAN data rate by overcoming this limit and propose *SIngle RF chain Multi-user BeAmforming (SIMBA)*, a novel framework for multi-stream multi-user downlink transmission via a single RF chain. We build on single beamformed transmission via overlayed constellations to multiplex multiple users' modulated symbols such that grouped users at different locations can share the same transmit beam from the AP. For this, we introduce user grouping and beam selection policies that span tradeoffs in data rate, training, and computation overhead. We implement a programmable WLAN testbed using software-defined radios and commercial 60 GHz transceivers and collect over-the-air measurements for different indoor WLAN deployments using a 12-element phased antenna array as well as horn antennas with varying beamwidth. We show that in comparison to single-user transmissions, *SIMBA* achieves 2× improvement in aggregate rate and two-fold delay reduction for simultaneous transmission to four users.**

***Index Terms*— MU-MIMO, 60 GHz, user selection, analog and digital beamforming, IEEE 802.11ad, IEEE 802.11ay.**

I. Introduction

RECENT advances in WLANs have employed *multi-user* MIMO to realize high aggregate data rate. For example, at 5 GHz, the IEEE 802.11ac standard supports multi-user transmission for up to 4 clients and 8 spatial streams yielding a peak aggregate data rate of nearly 7 Gb/s [1]. Likewise, at 60 GHz, IEEE 802.11ay [2] enhances single-user IEEE 802.11ad [3] and supports up to 8 multi-user streams to 8 clients yielding 100 Gb/sec aggregate rate. However in all cases, the number of simultaneously supported streams is limited by the number of baseband RF chains at the AP.

In this paper, we propose for the first time a 60 GHz WLAN architecture in which the number of supported simultaneous users and streams exceeds the number of RF chains. In particular, we introduce ***SI**ngle RF chain **M**ulti-user **B**e**A**mforming (SIMBA)* as a framework for realizing multi-stream multi-user downlink transmission via a single RF chain in 60 GHz WLANs. In this way, we remove the RF-chain limitation on multi-user scaling, enabling more users to be served concurrently. While *SIMBA* can enhance performance for any number of RF chains at the AP, for ease of exposition and to demonstrate the most extreme case, we focus on a single RF chain with simultaneous transmission to multiple clients.

SIMBA's key building block is to multiplex multiple users' modulation constellations into a single beam-formed transmission via *overlayed constellations*. We encode different data for different users into a single modulation structure while ensuring that grouped users can share the same transmit beam from the AP, despite being at different locations. For example, *SIMBA* can transmit 64 QAM to two users that can share a transmit beam in which a low SNR user only decodes 1 bit per symbol by detecting whether the received symbol is in phase or out of phase (i.e., BPSK) whereas the high SNR user decodes the remaining 5 bits per symbol. Therefore, unlike multi-RF chain solutions such as [2], [4], *SIMBA* with a single RF chain does not require methods to assess and manage inter-stream interference from use of different simultaneous beams.

The key challenges for *SIMBA* are to determine for each transmission: which users should be grouped, which transmit beam should be used, and which modulation level of the overlayed constellation each user should be assigned. We propose three policies to span the design space: First, we introduce a policy that targets to maximize the aggregate group rate without regard to the training or computational overhead incurred (*SIMBA-mr*). In particular, using training data comprising each user's SNR matrix of all AP-user beam pairs, *SIMBA-mr* calculates which combinations of users, beams, and modulations, yield the highest aggregate rate. At the other extreme, we propose *SIMBA-opp* as an opportunistic scheme that requires no additional training beyond single-user IEEE 802.11ad, i.e., an SNR training matrix is not required. *SIMBA-opp* serves users in the same order as a reference 802.11ad system, but for each transmission, searches its queue for users (viewed as "free riders") that can share the same beam while increasing aggregate rate, and if so, opportunistically adds them to the transmission. Finally, as a compromise between the sole focus on rate maximization of *SIMBA-mr* and the focus on computational and training simplicity of *SIMBA-opp*, we introduce *SIMBA* with SNR partitioning. The core idea is to exploit that the key source of data rate gain for *SIMBA* is the SNR spread among users.

Manuscript received 9 November 2020; revised 11 November 2021 and 22 April 2022; accepted 14 May 2022; approved by IEEE/ACM TRANSACTIONS ON NETWORKING Editor G. Zussman. Date of publication 23 June 2022; date of current version 20 December 2022. This work was supported in part by Cisco; in part by Intel; in part by NSF under Grant CNS-1923782, Grant 1824529, and Grant 1801857; and in part by DOD: Army Research Laboratory under Grant W911NF-19-2-0269. *(Corresponding author: Keerthi Priya Dasala.)*

Keerthi Priya Dasala and Edward W. Knightly are with the Department of Electrical and Computer Engineering, Rice University, Houston, TX 77005 USA (e-mail: keerthi.dasala@rice.edu).

Josep M. Jornet is with the Department of Electrical and Computer Engineering, Northeastern University, Boston, MA 02115 USA.

Digital Object Identifier 10.1109/TNET.2022.3182976

Indeed, if all users have identical SNR, *SIMBA* provides no data rate gain compared to a single-user system. Thus, *SIMBA-sp* partitions users according to their SNRs and forms groups across partitions, thereby vastly reducing the search space compared to *SIMBA-mr*.

We realize the key components of *SIMBA* on two testbeds. First, we employ X60, a fully programmable cross-layer configurable testbed for 60 GHz WLANs [5]. Next, we deploy WARP-60 as a variable beamwidth testbed utilizing WARP for baseband processing, VubIQ for RF functions at 60 GHz, and mechanically steerable horn antennas of different widths. With these two testbeds, we perform over 49,000 measurements spanning multiple scenarios and topologies.

Our key experimental findings are as follows. We begin with a simple yet important baseline case of two simultaneous users and experimentally study the effect of grouping a high SNR user that is close to the AP with a radially aligned user whose distance from the AP is varied from close (high SNR) to farther away (low SNR). The experiments reveal the critical role of receiver SNR spread in realizing gains with *SIMBA*. Namely, both individual and aggregate data rate increase with increasing SNR difference between the two users' links to the AP, yielding aggregate data rate gains of up to 64% compared to *Single User* transmission when the two users are separated by 22.5 meters.

Second, we study the case that a Line-Of-Sight (LOS) path is unavailable due to blockage, and hence the *SIMBA* AP must connect via a reflected path. While extensive studies [4], [6], [7] prove that Non-LOS (NLOS) paths offer reduced SNR due to a lack of LOS path and could be detrimental to multi-user performance, we have shown that *SIMBA* exploits these NLOS paths to the advantage of increasing performance, with our *SIMBA-mr* and *SIMBA-opp* policies achieving more than $1.41\times$ and $1.26\times$ respective multi-user gains over *Single User* under blockage.

Third, we vary transmit beamwidth from wide to narrow and find that adapting beamwidth at the AP acts as a knob in controlling the SNR spread and grouping efficiency and thereby the aggregate rate for *SIMBA*. We explore the trade-off that wider beams, which can create NLOS paths even when a LOS path exists, can provide better channel grouping opportunities, as more users will be able to share a beam. This advantage must be balanced against the lower directivity gain of wider beams. Thus, despite its low training overhead and complexity, *SIMBA-opp* using 80° beamwidth outperforms *Single User* by 58% and achieves about 75% of the aggregate rate of *SIMBA-mr*.

Next, we explore scaling the number of simultaneous users from 2 to 5 clients. We experimentally show that beamforming to four simultaneous users using *SIMBA* results in $2\times$ aggregate rate improvement over *Single User*. With full training information, the performance of both *SIMBA-sp* and *SIMBA-mr* increases until the number of grouped users is below five, but for *SIMBA-opp* with restricted training information, the performance saturates when more than two users are grouped. We find that as group size increases, the number of quantification levels of modulation (and coding) becomes the limiting factor for gains provided sufficient SNR spread exists. Indeed, no data rate gains of *SIMBA* over *Single User* are possible if only a single Modulation and Coding Scheme (MCS) is supported.

Lastly, we study the latency of *SIMBA* and find that the *SIMBA-sp* policy, although limited in search space to only 0.005% computation overhead compared to *SIMBA-mr*, has approximately two-fold reduction in total data transmission time compared to *Single User*, indicating that latency performance gains exploit *SIMBA-sp*'s policy of grouping users with high SNR diversity.

The remainder of this article is organized as follows: Section II provides *SIMBA*'s multi-user architecture and the framework of policies for user grouping and beam selection. Section III presents our implementation setup which includes the testbeds and measurement methodologies used for data collection. Section IV describes experimental investigations of the performance of *SIMBA* and factors including separation of grouped receivers, beamwidth adaptation, and user group size, all compared to single-user transmission as a baseline. Finally, we discuss related work in Section V and conclude the paper in Section VI.

II. *SIMBA* Framework

In this section, we propose *SIMBA*, **SI**ngle RF chain **M**ulti-user **B**e**A**mforming, a framework for realizing downlink multi-user multi-stream transmissions using a single RF chain. The *SIMBA* framework can support multiple policies in beam forming and user grouping that represent different tradeoffs in data rate, training overhead, and computational overhead. We present three key points in the design space, one that solely maximizes data rate without regard to overhead, one that has negligible additional overhead compared to today's standard, and one that strikes a balance between these two end-points of the design spectrum.

A. System Architecture

SIMBA employs the same baseband and antenna architecture as commercial products employing IEEE 802.11ad. Yet in contrast to 802.11ad which supports a single user at a time with a single RF chain, and in contrast to 802.11ay which supports multiple users with multiple RF chains, the objective of *SIMBA* is to transmit to a number of users greater than the number of RF chains. Below, we describe the case of a single RF chain with *SIMBA* simultaneously transmitting to more than one user at a time.

Figure 1 depicts an example architecture to support *SIMBA*. As shown, the AP is equipped with one RF chain capable of transmitting a data stream to multiple users. After modulation, the data stream can be steered with the depicted phase shifters and the resulting signal is mapped to an antenna array. The antenna array can generate a fixed set of beams typically via a predefined codebook in which each beam corresponds to a set of per-antenna phases. The beam direction and beamwidth are controlled by selection of the codebook entry for each transmission. We consider a receiver that likewise has a single RF chain and is enhanced with the ability to decode overlayed constellations defined as follows.

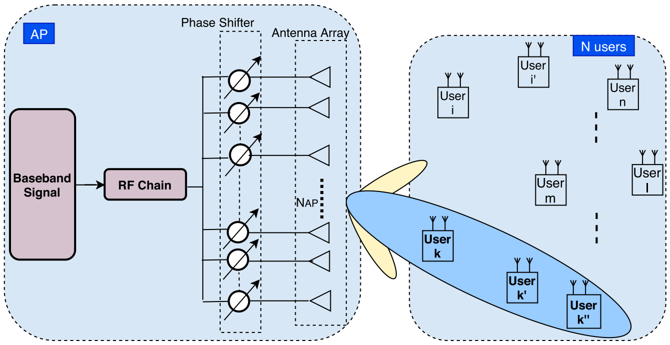


Fig. 1. AP with single RF chain system supporting multiple users.

In *SIMBA*, we use *overlayed constellations* in which each symbol represents information for multiple users. For example, in a two user case, symbols can be considered to be an overlay of two different data streams, one for each user. This can be illustrated using 16-QAM as shown in Fig. 2: The *SIMBA* AP transmits the modulated signal samples (1010 in this example) in Fig. 2. Because *SIMBA* will select a beam that can be received by both users, both receive this signal but with different SNRs. For example, in Fig. 2(b), the receiver with better SNR can distinguish among all of the 16 symbols with minimal error. However, the receiver with low SNR can identify only the quadrant of the transmitted signal sample, and can decode only the two most significant bits of the transmitted sample. Therefore, the *SIMBA* AP in this example delivers the two most significant bits of the symbol to the low SNR user (as it is easiest to decode the quadrant) and the two least significant bits of the symbol to the user requiring high SNR to decode. The AP must also inform the receivers of this encoding via a special header, analogous to how an OFDMA header must specify the group of receivers and allocations of subcarriers to receivers.

The aggregate rate of *SIMBA* depends on the respective rate of each receiver being grouped. For this, *SIMBA* needs to select the Modulation and Coding Scheme (MCS) in order to account for the fact that from the receiver's perspective, there is additional distortion in the received symbols due to modulation for other users in the overlayed constellation and this distortion/noise imposes a penalty on the performance of receivers especially with low SNR. As illustrated in Fig. 2(c), the received symbols at the lower SNR user have a noise sphere of larger radius implying larger error probability compared to that of high SNR user in Fig. 2(b) with smaller noise sphere. To avoid a higher BER for the low SNR user, prior work [8]–[13] reallocates transmit power and bandwidth with unequal error protection to each data stream being overlayed, resulting in non-uniform constellations which are not compatible with conventional modulation schemes in 802.11 standards [2], [3], [14]. Therefore, without violating the equal power constraint to each data stream and to maintain backward compatibility with the single-user MCS schemes mentioned in the standard, we instead select a MCS based on SNR but with modified multi-user predefined tables that target a fixed BER and is sufficiently robust to this additional symbol distortion. We use the standard practical MCS library with uniform constellation shapes from IEEE 802.11ay EDMG SC mode reference PHY implementation [2], [14] with modulation and coding combinations (with LDPC code rates) from $1/2$ BPSK to $7/8$ 64-QAM to select a suitable MCS for *SIMBA* data transmission. Note that all the results report only the PHY data rate and do not refer to throughput as the throughput should consider overhead in training, user grouping, and beam selection computation, thus requiring us to define the transmission timeline of SIMBA, which is out of scope of the paper.

Since SIMBA is a method for modulation, it is independent of the selected coding strategy and can be applied to both coded and uncoded symbol streams, provided that (i) each user's bitstream is coded independently before the overlayed constellation mapper, and (ii) at the frame level, the combining of user streams is done such that the correct number of bits maps to the number of symbols required by the multi-user frame.

Lastly, we comment that while standard multi-user protocols such as 802.11ac [1] provide mechanisms for simultaneous transmission, the specific policy of user selection, scheduling, fairness, etc. are not defined by the standard to enable flexible implementations. Likewise, *SIMBA* provides a multi-user mechanism as discussed in this section and numerous multi-user policies can be realized on top of it. We take the 802.11 design philosophy to define *SIMBA* as opposed to following a clean slate approach. Nonetheless, our design lies in the framework adopted for the state-of-the-art Wi-Fi systems. For instance, we can recognize this design change with 802.11ad [3] vs. 802.11ay [14] systems where multi-user 11ay enhances the single-user 11ad, and that transition from single-user single RF chain to multi-user multi RF chain is a big difference. Likewise, to transition from a Multi-User Multi-RF chain system like 802.11ay to a Multi-User Single RF chain like *SIMBA* would also be a significant difference. So *SIMBA* would be a new and non-trivial modification to the standard.

B. SIMBA With Maximum Rate

Here we present *SIMBA-mr* as a *SIMBA* policy that targets to maximize the aggregate data rate to a group of backlogged users without constraints on training and computational overhead (we do analyze overhead, but do not consider it as a policy design factor). For each transmission, the policy specifies the set of users to be grouped for simultaneous transmission, the beam (codebook entry) to be used, and the modulation and coding scheme for each stream.

We consider that the network consists of U users backlogged for downlink transmission and let $j \in \mathrm{C_{AP}}$ denote the transmit beam index in the codebook of the AP and $k_u \in \mathrm{C_u}$ denote the receive beam index in the u^{th} user's corresponding codebook. The training information consists of the training matrix for each user that comprises the measured signal-to-noise-ratio for each beam pair $SNR_u(j, k_u)$. The maximum instantaneous data rate $R_u^{\max}$ achievable by user u when it is served by itself depends on the best TX-RX beam pair (j_u^*, k_u^*)

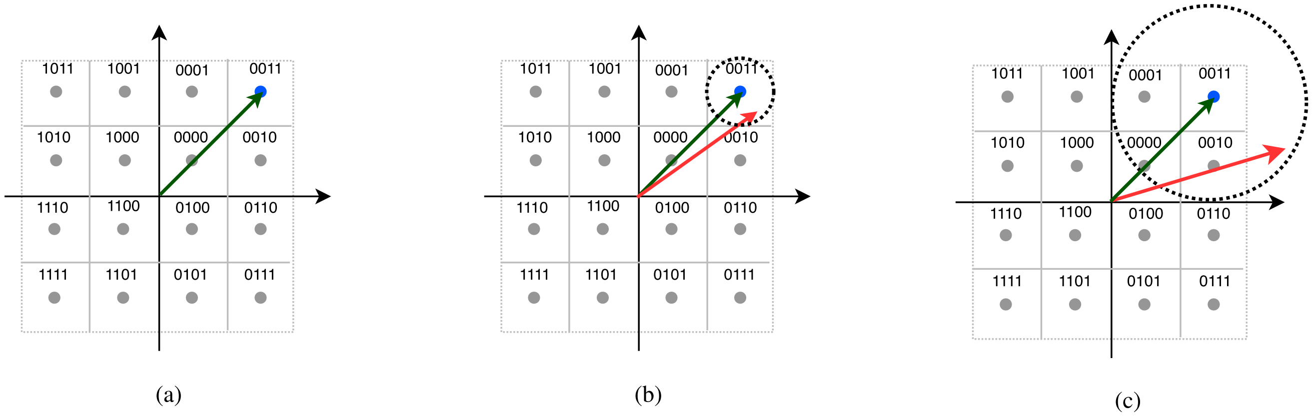


Fig. 2. The figure shows the transmitted sample in green, the received samples in red, and noise in black with signal reception at users with different channel quality. (a) The AP transmits the 16-QAM *SIMBA* modulated signal. (b) A high SNR user experiences less noise and can estimate the transmitted sample up to the small square, i.e., up to 4 bits. (c) A low SNR user sees more noise and hence knows only the quadrant of the transmitted sample, i.e., it knows only 2 bits of the transmitted sample.

associated with it and can be expressed as

$$(j_u^*, k_u^*) = \arg\max_{(j,k)} SNR_u(j,k) \tag{1a}$$
$$R_u^{\max} = MCS(\ SNR_u(j_u^*, k_u^*)\ ), \tag{1b}$$

where MCS$(\cdot)$ gives the data rate achievable for a particular SNR via the minimum SNR tables from [3]. When *Single User* transmission is utilized as an experimental baseline, we always use this best beam pair and rate $R_u^{\max}$ for each user.

Define group G as a candidate subset of backlogged users. If the AP transmits to this group with transmit beam j, each user would select its best receive beam $\{k_u\}_{u\in G}$ that has the greatest SNR for transmit beam j, which is not necessarily (j_u^*, k_u^*) as above. The total available rate of received symbols at user u for transmit beam j is then $MCS(\ SNR(j,k_u)\ )$. However, user u no longer obtains all modulated bits for itself, and hence its individual rate depends on the resulting MCS of the overlayed constellation selected for the group. In particular, the allocation of the number of symbol bits (and the corresponding PHY data rate) to each user in the group is constrained by the available rates at these users.

Let B_{OC} denote the set of bits b_u allocated to users in group G in the final overlayed constellation. Then

$$B_{OC} = (b_1, b_2, b_3, \cdots, b_u, \cdots, b_{|G|}\ ) \tag{2}$$
$$F_{OC} = \sum_{u\in G} B_{OC} \tag{3}$$

where F_{OC} denotes the number of bits in the final overlayed constellation and $|G|$ is the cardinality of user set G. As a result, the final MCS for group G is chosen following the modulation scheme resulting from F_{OC} with a fixed code rate compatible with all the grouped users decoding limit. The resulting aggregate group rate R_G is given by the protocol specific MCS to data rate conversion [3].

SIMBA-mr selects the aggregate-rate maximizing set among candidates of user groups, MCS for overlayed constellation, and beams. Defining $R_u(G, j, k_u)$ as the received symbol rate for the user u in group G using transmit beam j and receive beam k_u, we define the objective of *SIMBA-mr* as follows:

$$\{G^*, j^*, \{k_u^*\}_{u\in G}\} = \arg\max \sum_{u=1}^{U} R_u(G, j, k_u) \tag{4a}$$
$$s.t \quad \frac{L}{R_G} \le \sum_{u=1}^{|G|} \frac{L}{R_u^{\max}} \tag{4b}$$
$$j \in C_{AP}, \tag{4c}$$
$$k \in C_u, u \in U \tag{4d}$$

Equation (4a) targets to maximize the aggregate rate over all users by finding a user group G^* from a set of backlogged users U and selecting a shared transmit beam j^* for all users in the group, with each group member having their best receive beam k_u^* for the selected transmit beam. The first constraint restricts the time it takes to transmit data frame of length L to a user group relative to the time that would be required to transmit independently using *Single User* to all set of users in the group. This ensures that *SIMBA-mr* will never underperform *Single User* by excessively grouping in cases where there is no gain over *Single User*. The last two constraints ensure that the transmit and receive beams are selected from the predetermined AP and user codebooks respectively. *SIMBA-mr* solves the above problem by searching over all possible user and beam combinations. *SIMBA-mr* subsequently repeats this process to find another user group until all backlogged users are served. To realize fairness objectives, an AP can further restrict selections accordingly, a consideration beyond the scope of *SIMBA-mr*.

Training and Computation Overhead. In advance of grouping and transmission, beam training is required to obtain the information used to select beams and groups to maximize aggregate rate. *SIMBA-mr* can be trained one client at a time, in which the AP sends training frames sequentially across all the beams in the predetermined RF codebook. For each transmit beam, the user measures the SNR for each of its receive beams. Thus, for each user, for the AP with $\mathrm{C_u}$ beams and each user with $\mathrm{C_u}$ beams, this training comprises collecting SNR measurements over $\mathrm{C_{AP}} \times \mathrm{C_u}$ training transmissions.

Computationally, *SIMBA-mr* records the achievable rate of all possible user groups and picks the one with maximum aggregate rate. For this, *SIMBA-mr* yields a search space of $\binom{U}{M}(\mathrm{C_{AP}} \times \mathrm{C_u})^M$ distinct combinations to transmit to M users simultaneously.

If a user is mobile, then the analog beam configuration found by *SIMBA* would fail for that user and would not be able to use the same beams the next time AP wins contention for transmission to the target user group. This would lead to re-training and selection of a new beam for the mobile user before the next data transmission period, thereby addressing mobility. Additionally, existing beamwidth and rate adaptation protocols [7], [15], [16] for mobility can be applied on top of *SIMBA* to address mobility of high mobile clients.

C. Opportunistic SIMBA

While *SIMBA-mr* maximizes aggregate rate, its training and computational overhead may be too high for practical implementation for large numbers of mobile clients. Here, we introduce *SIMBA-opp* as a policy in the other end of the design spectrum. Namely, *SIMBA-opp* repurposes the existing training of IEEE 802.11ad to limit training overhead. Nonetheless, as the name implies, *SIMBA-opp* targets to opportunistically increase aggregate data rate compared to single-user transmission, by adding multiple users to a transmission whenever it is viable. More specifically, we define the policy as follows.

As with 802.11, the underlying training mechanism has the AP send training frames sequentially across all beams in the predetermined RF codebook, while the client employs quasi-omni reception to find the AP's transmit beam providing the highest SNR. Subsequently, the client sweeps its transmit beams while the AP receives quasi-omni such that the outcome yields the single best transmit-receive beam pair via a total of $\mathrm{C_{AP}} + \mathrm{C_u}$ training transmissions. Unlike the required training of *SIMBA-mr*, this procedure does not provide the SNRs of all beam pairs. Hence, the inputs to *SIMBA-opp* do not comprise the entire SNR matrix for each user $SNR_u(j,k)$, but rather, only the maximum SNR SNR_u^{max} that corresponds to the best TX-RX beam pair (j^*, k^*).

The strategy of *SIMBA-opp* is to serve users in the same order that they would have been served by Wi-Fi (presumably, first come first serve, although variations can be supported) with the following modification: We denote the user that is at the head of the queue as the reference user and this user will be served next by *SIMBA-opp* without exception. Yet, in contrast to a single-user system, *SIMBA-opp* will search into the queue attempting to find another user (or users) that happen to share the same transmit beam as the reference user. *SIMBA-opp* will then form a multi-user transmission with these additional users *only if* it will increase the aggregate data rate. Hence, the computational complexity of *SIMBA-opp* can be controlled by limiting how deep it searches the queue, with maximum overhead if it searches all users in the queue to find the best ones to add. As such, given knowledge of the number of MCS levels N_{MCS} in the supported rate specific SNR tables in [17], we denote the number of distinct modulation levels and code rates as N_{mod} and N_{cr} respectively. Depending on the reference user's modulation level and for a queue depth of Q_d, the maximum computational cost to find the best user in the search space is $O(N_{mod} \times Q_d)$. In addition, instead of searching all users in the queue, we may restrict the search to users that share the reference beam which have higher decoding SNR thresholds as only these users have the potential to establish higher levels of modulation than the reference user modulation level in the overlayed constellation.

Thus, we expect that whenever the reference user has relatively low SNR (implying low MCS), the AP will have the opportunity to find a user with higher SNR to superimpose on the low SNR user. The high SNR user can be viewed as a free-rider on the transmission by sharing the same beam. In contrast, if the selected user already has peak MCS by itself, no gain can be realized by grouping it, and *SIMBA-opp* will simply transmit the user by itself.

More formally, let S_i be the user group at the start of the selection process and let $n, n \in U$ be the index of the reference user selected in step i. Denote the MCS level associated with the reference user n as $l_{MCS}(n)$ whose modulation and code rate index levels are respectively denoted as $l_{mod}(n)$ and $l_{cr}(n)$, $\forall\ l_{mod} = 1, 2, \cdots N_{mod},\ l_{cr} = 1, 2, \cdots N_{cr}$. The AP proceeds to the next step $(i{+}1)$ by searching for users $m, m \neq n, (n,m) \in U$ in the transmit beam direction of user n, such that $R_{sum}(S_i \cup \{m\}) > R_{sum}(S_i)$. For this, the requirement is that the modulation of user m is at least one level above the modulation of reference user such that $l_{mod}(m) > l_{mod}(n)$. Intuitively, this is true only when the user m has a significantly higher SNR on the reference transmit beam compared to user n implying that $SNR_m \gg SNR_n$. As an outcome of this selection process, each sequentially added user has SNR greater than the threshold of the reference user and is at a higher modulation level such that the final overlayed constellation has the base modulation level belonging to the reference user. In addition, although each user can use a different MCS, the coding rate for all the grouped users is the same and is compliant with each user's SNR decoding threshold.

The AP ends the user selection process when the sum rate achieved reaches the maximum supported MCS or when all candidate users have their SNR below the reference user's SNR threshold in the reference user's transmit beam direction.

D. SIMBA With SNR Partitioning

Here, we present a final *SIMBA* strategy that represents a balance between *SIMBA-mr* and *SIMBA-opp*. Namely, we present *SIMBA-sp* as a policy that significantly reduces the search space of *SIMBA-mr* by exploiting the fact that transmission groups with high aggregate rate are typically composed of streams having high SNR spread. That is, if all users have the same SNR, *SIMBA* cannot realize a gain over time sharing, as there is no SNR margin for additional users to "free ride." Consequently, *SIMBA-sp* avoids testing all possible beam and group combinations by dividing backlogged users into similar-SNR partitions and attempting to find transmission

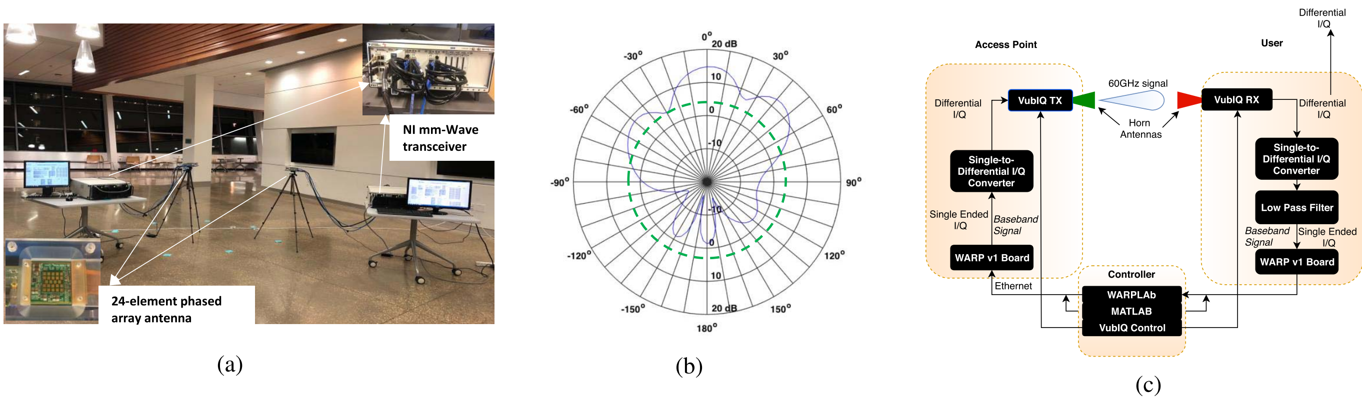


Fig. 3. (a) Experimental setup showing X60 platform with NI mm-Wave transceivers and SiBeam's phased array antenna. (b) Irregular azimuth beam pattern and its average directivity(green) (c) Experimental setup showing WARP-60 platform with 60 GHz signal and control flow.

groups among partitions and not within partitions. *SIMBA-sp* differs from *SIMBA-opp* in that *SIMBA-opp* is forced to serve the reference users in order. In contrast, like *SIMBA-mr*, *SIMBA-sp* can change the service order to realize a gain in aggregate data rate, albeit with far less overhead than *SIMBA-mr*.

More formally, we define *SIMBA-sp* as follows. Like *SIMBA-mr*, the inputs to *SIMBA-sp* are the training information of the SNR matrix for each beam pair $SNR_u(j, k)$. For a given set of backlogged users U, *SIMBA-sp* first sorts users in decreasing order of their maximum (best beam pair) SNR. Let M denote the group size that the AP intends to serve simultaneously. *SIMBA-sp* divides the SNR-sorted users into M partitions labelled 1 to M, such that partition 1 includes $\lceil * \rceil \frac{U}{M}$ users having the highest SNR and partition M has the lowest SNR users. *SIMBA-sp* begins with a "prime user," and we denote the partition of this user by p. While any user can be the prime user, we select the prime user from partition one as these users will dominate the contribution to the aggregate rate in comparison to lower SNR users from other partitions.

SIMBA-sp uses the best-SNR transmit beam (codebook entry) of the prime user for transmission to all group members. Hence, *SIMBA-sp* will use this beam for further user grouping and will attempt to add users from other partitions that can share this beam while increasing the aggregate rate, even if the selected transmit beam is not the best selection for added users. In other words, using the transmit beam direction of this prime user from partition p, *SIMBA-sp* will subsequently search users in partition $(p{+}1)$ $(mod\ M)$ to find another user that can share the transmit beam with the prime user. *SIMBA-sp* iterates over the partitions with the same procedure of searching for users that can form a higher-aggregate rate multi-user transmission with already grouped users. Finally at the end of the user grouping process, the users within the group are assigned to different levels of overlayed constellation with the prime user typically embedded at the highest level and the user from the M^{th} partition typically at the lowest. *SIMBA-sp* then repeats the entire process with a new prime user and group until all backlogged users are served. Thus, the effectiveness of this policy depends on the respective signal strengths of the users that have sufficient link budget to share the common beam to realize a distinguished subset of bit assignments in the overlayed constellation.

Computational Complexity. The transmit beam selection of *SIMBA-sp* involves a search space of at most $U \times C_{AP} \times C_u$. For user grouping, *SIMBA-sp* checks up to $(M-1)$ partitions and searches $\lceil * \rceil \frac{U}{M}$ users in each partition. Therefore, the total search space for an M user group in *SIMBA-sp* involves $(U \times C_{AP} \times C_u) + \lceil * \rceil \frac{U}{M}(M-1)$ tests of aggregate rate. *SIMBA-sp* has reduced complexity compared to *SIMBA-mr* which exhaustively tests all users and beams for each user being grouped. The overhead ratio can be computed for *SIMBA-sp* and *SIMBA-mr* using an example: when $U = 15$, $M = 3$, $C_{AP} = 25$, $C_u = 10$, this ratio is approximately 0.005% implying that *SIMBA-sp* significantly reduces computational overhead compared to *SIMBA-mr*.

III. Evaluation Setup: Testbeds and Over-the-Air Experiments

We implement the key components of *SIMBA* and collect over-the-air data to evaluate its performance. In order to study the impact of various parameters such as beamwidth, antenna array beam patterns and multi-user capability, we employ two different platforms, X60 and WARP enhanced with a 60 GHz front end.

A. X60 Phased Array Platform

We perform over-the-air experiments with X60, a configurable testbed for 60 GHz WLANs [5]. X60 features a fully programmable cross-layer architecture for PHY, MAC and Network layers. Fig. 3(a) shows the X60 platform where each X60 node is built with National Instruments' (NI) millimeter-wave transceiver system and employs a user configurable SiBeam phased array antenna module with 24 elements, 12 for TX and 12 for RX. Communication is established over wideband 2 GHz channels that can reach multi-gigabit data rates using real-time electronically steerable (switching time of $1\mu s$) TX and RX beams from a predetermined codebook that has a dictionary of 25 beams which are spaced roughly 5°

apart along the mainlobe direction, thereby covering a sector of $-60°$ (beam index -12) to $60°$ (beam index $+12$) in the azimuthal plane centered around the antenna's broadside direction (beam index 0). Each beam has a 3 dB bandwidth of $25°$ to $35°$ causing each main-lobe to overlap with other neighboring beams. Therefore, it is evident as in Fig. 3(b) that X60 provides beam patterns with predominant main lobes overlapping and strong side-lobes.

We collect channel samples from over-the-air measurements and subsequently perform trace-driven emulation to study *SIMBA*. More details on this methodology are presented in Sec IV.

B. WARP-60 for Variable Beamwidth

Because the X60 beam patterns are fixed, we cannot use that platform to explore beamwidth. Hence, we equip WARP with a 60 GHz front end and horn antennas from [7], [16] in order to vary beamwidth. Moreover, the horn antennas' regularly shaped beam patterns can emulate beam forming of a many-antenna phased array. In particular, we use the testbed setup in Fig. 3(c) which consists of commercial mm-wave transmitter and receiver modules from the VubIQ 60 GHz development system. These mm-wave transceivers can communicate in the 57-64 GHz unlicensed band with up to 1.8 GHz signal bandwidth and can accept and output I/Q baseband signals. We generate I/Q baseband signals at different modulations and rates using WARP baseband [18], and use WARPLab (a framework for rapid physical layer prototyping) to generate BPSK and QPSK baseband signals with 20 MHz bandwidth. This is due to the hardware limitation of WARP board which is capable of data transmission bandwidth of 20 MHz and hence cannot operate in full 1.8 GHz channel. The VubIQ controller upconverts the baseband samples to a 60 GHz signal for over-the-air transmission. Directional transmission and reception is enabled by horn antennas, with beamwidth varied by making use of $7°$, $20°$ and $80°$ horn antennas. The VubIQ module downconverts the client's received signal to analog I/Q which are then sampled by the client's WARP board and demodulated in WARPLab.

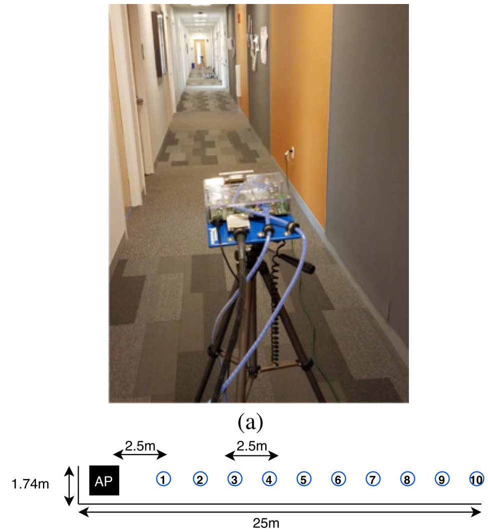
To collect the received signal strength at different client locations and for various receive antenna orientations, a mechanical motor, DC microstrip driver and a commercial motion control setup connected to the VubIQ modules is used to steer the beams with sub-degree accuracy. Using this 60 GHz system, we measure the signal strength of a point to point transmission as this system does not allow for multi-user transmission due to the provision of only a single RF chain. We perform numerous measurements varying the receiver location, antenna beamwidth, and use this data to study the performance of *SIMBA*.

IV. Experimental Evaluation

In this section, we perform over-the-air measurements to evaluate the performance of *SIMBA* and compare to baseline schemes.

A. Distance Between Grouped Receivers

We first experimentally characterize the multi-user gains of *SIMBA* via a simplified setting comprising of two aligned receivers with varying distance between them. This enables us to study the impact of the SNR difference between the two AP-client links.

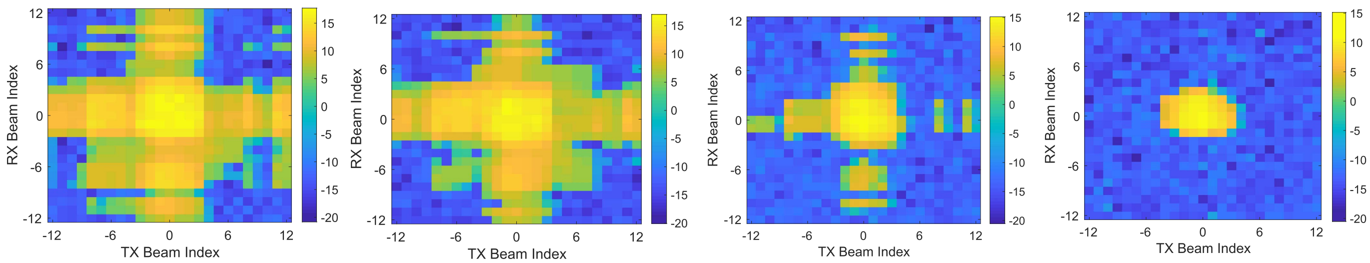


(b)

Fig. 4. (a) Indoor corridor measurement scenario deployed using X60 nodes to study user separation. (b) Map of the corridor measurement locations included in our study *(Not drawn to scale)*.

Setup. We deploy X60 nodes and conduct experiments in an indoor corridor as depicted in Fig. 4. The AP is mounted on a tripod at a height of 1.23m and is fixed at one end of a 1.74 wide corridor. We consider 10 receiver positions (represented by circles) varying the AP-user distance in a straight line from 2.5m to 25m in steps of 2.5m. In all locations, the receiver always faces the AP and is at the same height as the AP. The presence of side walls and windows and metal coating (not shown) create reflections. For each AP-user setting, we collect the received signal strength measurements for all beam-pair combinations. We assume that Single User beam training is performed initially or as needed according to the client or environmental mobility, but typically at a slower time scale than packet transmission such that the overhead is negligible. In our evaluation, nodal and environmental mobility is negligible such that beam training information is reliable, even when used later. We use the Single Carrier (SC)-PHY (MCS 1-20) defined in 802.11ay EDMG-MCS table in [14] to map the SNR to the data rate using the protocol-specific minimum SNR thresholds [17]. We denote the first user as U_1 and it is always at location 1, while the second user U_2 is placed in one of the locations 2-10 as shown in Fig. 4.

SNR Matrix for *SIMBA-mr* . Recall that all *SIMBA* policies (as well as 802.11 and the baseline *Single User*) require measured SNRs of different transmit and receive beam configurations. *SIMBA-mr* uses as its input the SNRs of all possible transmit/receive beam combinations. We denote the transmit and receive indexes of X60's codebook of 25 beam patterns as $i_{tx}(= 1 : 25)$ and $i_{rx}(= 1 : 25)$. Thus, the measurements yield a 25×25 matrix of SNR values for each user, which we depict as a beam-pair heatmap in Fig. 5, with results shown for four different distances from the AP: $2.5m$, $7.5m$, $17.5m$ and $25m$. Due to overlap between neighboring



(a) User location 1 (b) User location 3 (c) User location 7 (d) User location 10

Fig. 5. SNR heatmaps for all beam pair combinations for user location indices 1, 3, 7 and 10 in the corridor.

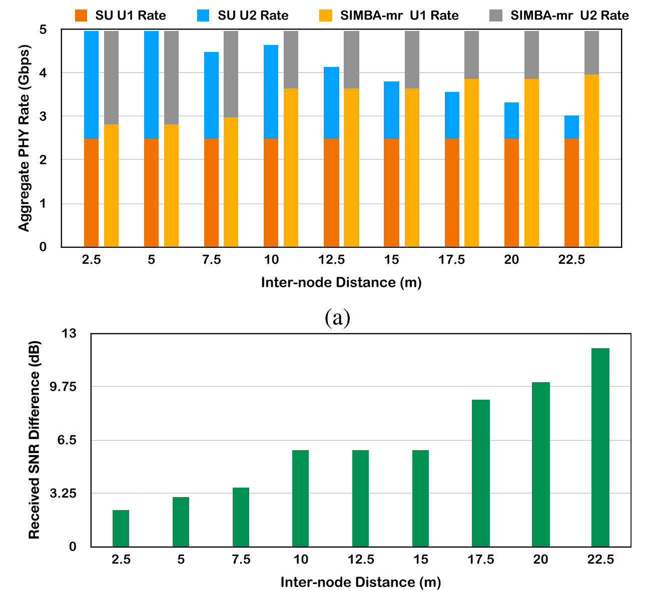
beams in X60 (Sec. III-A), multiple beams may include the LoS component, albeit with different directivity gain. Hence we observe a cluster of high SNR beam pairs close to the central pair (13, 13). Besides the LoS central high SNR region, there are smaller clusters of beam pairs with moderate to high SNR, resulting from reflections and side-lobes. As the TX-RX distance increases, SNR decreases more rapidly, indicating a greater degradation in relative strength of neighboring beam pairs. Furthermore, for greater inter node distances, fewer beam pairs on average achieve the maximum possible signal strength. These characteristics play a key role while selecting beams for *SIMBA* which is studied next.

Two User Near-Far Grouping with *SIMBA*. We begin with *SIMBA-mr* in which the AP selects its beam via analysis of both users' 25×25 SNR matrices. We consider that there are only two users in the system at a time, so there is no group selection in this experiment. In particular U_1 and U_2 are grouped for simultaneous transmission and *SIMBA-mr* selects a beam such that these two users can simultaneously receive data on the same transmit beam with maximum aggregate rate.

As a baseline, the *Single User* aggregate rate is measured by considering that the two users U_1 and U_2 each get half of the air time. The AP uses the best (rate maximizing) transmit beam for U_1, and U_1 uses its best receive beam. The best transmit-receive beam pair is likewise used for U_2. Note that unlike the multi-user case, the AP can select different transmit beams for U_1 and U_2 for *Single User*.

Two Close Users. Fig. 6(a) depicts the aggregate rate and per user rates of both users for different location of U_2 for *SIMBA-mr* and *Single User*. First, observe that when U_2 is close to the AP and hence also close to U_1, i.e., for locations 2 and 3, *Single User* rates for both users are the maximum supported MCS. Thus, *Single User* can achieve the highest aggregate rate even when the two users share air-time, with each getting half. Likewise, when the users are served with multi-user transmission and beam selection from *SIMBA-mr*, the aggregate rate is equal to *Single User* due to lack of sufficient SNR difference for beams selected for U_2 user locations 2 and 3 (see also Fig. 6).

Increasing Inter-User Distance. In contrast, the trend changes when U_2 is farther away from U_1. For instance, the performance of both schemes is quite different when U_2 is at locations 4-10. Although *Single User* finds beams with the best MCS possible to each user independently, U_2 can't contribute enough to the aggregate rate as the distance grows



(a)

(b)

Fig. 6. (a) Aggregate rate as a function of receiver separation. (b) Received SNR difference arising from joint beam selection of *SIMBA-mr* for users U_1 and U_2.

and U_2 utilizes half the air time. In contrast, for the same locations, *SIMBA-mr* obtains a better sum rate than *Single User*. As shown in Fig. 6(a), *SIMBA-mr* beam selection indicates that U_1 obtains higher rates as the location of U_2 changes from 4-10. For example, in location 7, U_1 obtains a rate of 3.65 Gbps while *Single User* provides a rate of 2.49 Gbps. This is because *SIMBA-mr* leverages SNR diversity which increases with distance between the users as shown in Fig. 6. The sum rate for this scheme is always maximized as it exploits the SNR spread resulting from the near-far location of users.

Finally, Fig. 6(a) suggests that joint selection of beams can successfully provide a rate above 1 Gbps for U_2 regardless of its distance to the AP and spatial separation from U_1. In most cases, gains from maximizing the SNR spread using *SIMBA-mr* are high due to the large ($\sim 30°$) beamwidth and presence of strong side-lobes giving rise to more flexibility of choosing beams. In addition, the high deviation in sum-rate for *SIMBA-mr* and *Single User* also indicates the high performance dependency on the location of available users and the SNR spread among users.

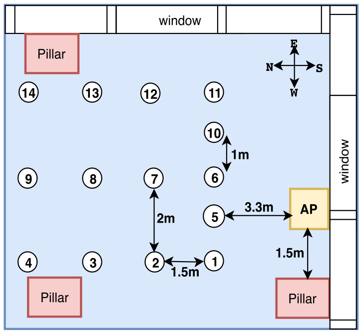


Fig. 7. Experimental layout consisting of AP and 14 users (shown in circles).

Fig. 8. Experimental setup showing LOS blockage with a wooden table.

Finding: As the distance between two receivers increases, multi-user transmission using SIMBA-mr can exploit the disparity in received SNRs to share a beam while maximizing aggregate rate, whereas the same ungrouped users served by Single User are vulnerable to performance degradation.

B. Limitations of LOS Blockage on SIMBA

Prior work has shown that mmWave channels are sparse, i.e., only a few reflected NLOS paths characterize the channel between any two nodes [19], [20]. This implies that when there are blockages in the environment there might be a few beams that can provide sufficient SNR for multi-Gbps communications. This means that there might only be a few beams that can potentially support *SIMBA* as most beams would not even provide a sufficient link budget for a *Single User* transmission. Namely, there would only be a few beams that can provide sufficient SNR at the users thus impacting the required SNR spread for *SIMBA* to have performance gains. Here, we experimentally explore a scenario in which LOS path is unavailable due to blockage and analyze the performance of *SIMBA* under blockage.

Setup. We use X60 with the topology shown in Fig. 7 which includes the AP and 14 different user locations. The floor plan portrays the lobby area, where the AP is fixed at one end of the wide wall at a height of 1.23 m facing North. All users are at the same height as the AP and face South. The presence of pillars, windows, and metal surfaces can create multiple signal reflections. We measure the received SNR for all 25×25 beam-pair combinations under two scenarios: *(i)* the AP has LOS path to the client; *(ii)* LOS path is blocked with a table as shown in Fig. 8. We represent each beam sweep as a heatmap of corresponding SNR values with TX beam indices along the x-axis and RX beam indices along the y-axis.

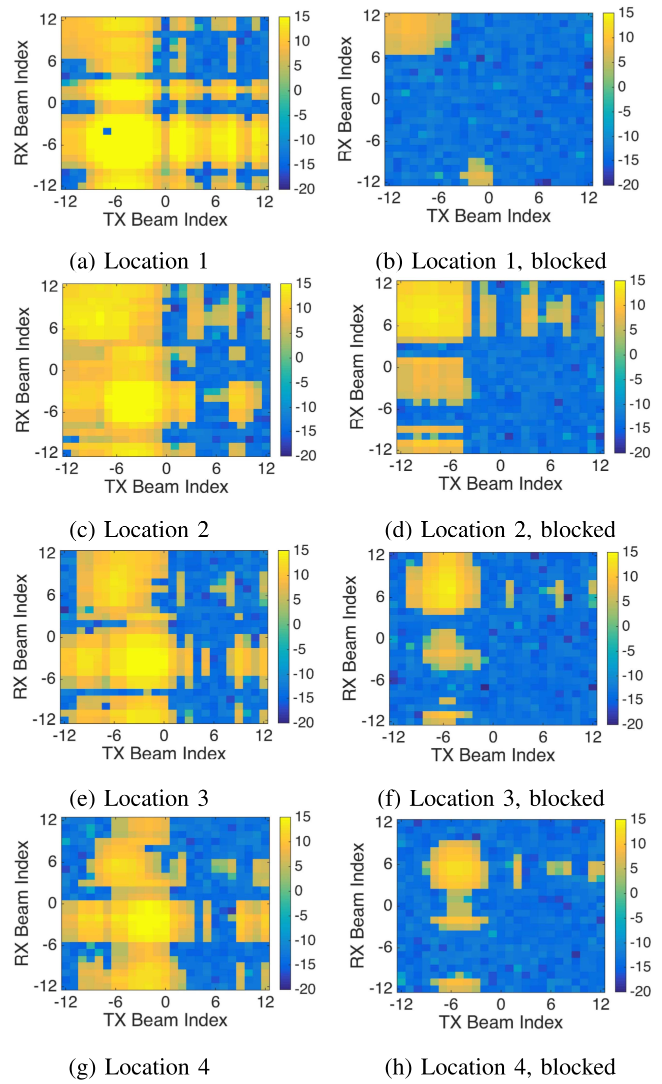


Fig. 9. SNR heatmaps for all beam pair combinations for user location indices 1, 2, 3 and 4 in the lobby under LOS connectivity (left column) and blockage (right column).

SNR Heatmaps. Figure 9 presents the SNR heatmaps for user location indices 1, 2, 3 and 4 under LOS connectivity (first column) and blockage (second column). We do not show the SNR heatmap for all locations due to space limit. The SNR range as seen in the heatmaps is between -20 dB to 15 dB with yellow colored regions indicating beam pairs above 10 dB whereas blue regions indicate beam pairs with negative SNR.

We observe several beam-pairs providing above 10 dB SNR that corresponds to 1 Gbps data rate in our platform. The received SNR for each beam-pair is dependent on the physical paths and the directivity gain along them. Imperfect beam patterns can cause LOS/NLOS paths to be captured by multiple beams albeit with different directivity gains. The beam pair with the highest SNR corresponds to the physical LOS path in Fig. 9(a), 9(c), 9(e) and 9(h). For instance, the beam index -5 at TX and -5 at RX provides the highest directivity gain along the LOS path at position 1. Fig. 9(a) confirms that while beam-pair $(-5, -5)$ is within the high SNR region but due to overlap between neighboring

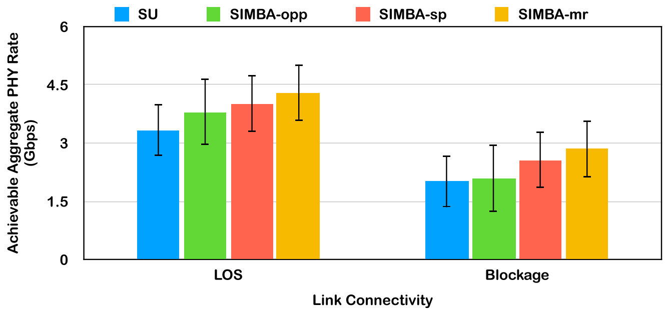


Fig. 10. Achievable Aggregate Rate under LOS Connectivity and Blockage.

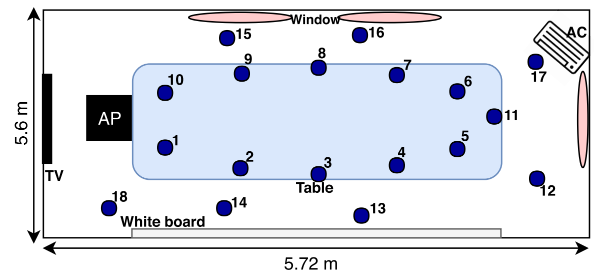


Fig. 11. Experimental floor plan used for measurement of data using WARP-60 testbed.

beams, multiple beams include the LOS paths and we see a cluster of high SNR beam pairs around $(-5, -5)$. Under LOS blockage, the yellow region corresponding to LOS component disappears as seen in second column of Fig. 9. This conforms that the high SNR region has experienced significant SNR reduction after blockage. Lastly, the highest SNR region after the blockage achieves similar SNR under LOS conditions i.e., LOS blockage has not degraded their SNR; indicating that the irregularities and imperfections as well as the side lobes present in the beam pattern must be capturing a reflected path at these beam pairs.

Aggregate Rate. Figure 10 depicts the achievable aggregate rate of all multi-user schemes of *SIMBA* in comparison with *Single User* under scenarios where LOS path is available and unavailable due to blockage. For each *SIMBA* policy, we average the performance over all two user groups in the setup. First, as expected the aggregate rate achieved by *Single User* is higher in LOS case in comparison to the case with blockage. Although *Single User* finds the best beam with the best MCS possible for both users using the NLOS paths, each user cannot contribute enough to the aggregate rate. As discussed earlier, although the higher SNR region after blockage achieves similar SNR under LOS conditions as seen in Fig. 9(b), 9(d), 9(f) and 9(g); note that these positions indicate extreme example positions in which two users are co-located and may see strong reflected paths (i.e., off two walls or windows) from the AP but not all users have the benefit of exploiting reflected paths and therefore achieve degraded SNR under blockage leading to reduction in per-user rate and aggregate rate in these user groups. This would not be an issue in LOS scenario since the beams chosen are mostly the LOS paths which is significantly stronger than any NLOS paths. Therefore the beam selection performed by *Single User* scheme under LOS connectivity results in significantly higher SNR and per-user data rate so that the achievable aggregate rate in LOS scenario would be almost $1.67\times$ of the scenario with blockage.

As evident from Fig. 10, with blockage, *SIMBA-sp*, and *SIMBA-mr* achieve about 28% and 43% multi-user gains over *Single User* scheme. By design, these schemes select analog beam configuration that results in sufficient SINR at the users while not compromising on the aggregate rate. This result implies that although the blockage might have reduced user's SNR in comparison to the LOS case, there is often at least one receive beam at each of the user that could share a transmit beam at the AP while maximizing the SNR spread. While this is true for these both schemes, unfortunately *SIMBA-opp* policy yields only a marginal improvement over *Single User* under blockage. Considering the lower SNR of NLOS users and decrease in overlap of high SNR regions for users across most of the user groups, it becomes increasingly likely that *SIMBA-opp* which uses just the maximum SNR beam-pair information for each user, cannot find rate maximizing beams at the users that can share a common transmit beam at the AP and hence the performance is very close to *Single User*.

Finding. With blockages in the environment, joint beam selection by SIMBA effectively exploits the resulting lower SNR inherent to NLOS paths to the advantage of providing sufficient SNR spread and increasing multi-user gains, whereas transmission to the same ungrouped users using beam separation based policies like Single User perform poorly in NLOS scenarios.

C. Adapting Beamwidth

In this experiment, we explore how beamwidth adaptation can be used to improve the performance of *SIMBA*. For instantaneous data rate maximization, narrow beamwidth helps maximize directivity gain. However, in the presence of mobility, the repeated training overhead for the narrowest beams may overwhelm the advantage of high directivity gain such that wider beams can offer greater resilience to mobility and reduce training overhead [7]. For *SIMBA*, a wider transmit beam covers a larger set of clients, thereby increasing the grouping possibilities and yielding more opportunities to achieve maximum aggregate rate. Yet, the drawback to widening beamwidth is reduced directivity gain, which can reduce SNR and data rate. To explore these tradeoffs, we employ the WARP-60 testbed which generates directional beam patterns with configurable beamwidth using horn antennas.

Setup. We employ the topology shown in Fig. 11 and fix the AP position at one end of a table in an indoor room and vary the user locations in 18 different positions. The room has a white board, AC unit, TV and glass windows which act as reflectors. The receiver orientation is chosen so that it provides a LOS path to the AP from each user position. For each client position, fixed receiver antenna orientation and for a fixed receive antenna beamwidth of $20°$, we perform a $360°$ sweep of the AP in steps of $5°$. The sweep is repeated for different AP horn antenna beamwidths of $7°, 20°$ and $80°$.

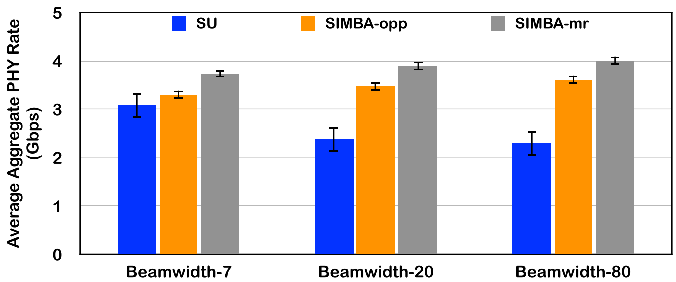


Fig. 12. Average of the aggregate rate for all the users in the setup.

The received signal strength for each point in the AP's sweep is collected.

Beamwidth and Aggregate Rate. For different transmit beamwidths at the AP, Fig. 12 shows the aggregate rate averaged over all users of the multi-user schemes *SIMBA-opp* and *SIMBA-mr* in comparison to *Single User*. First, for the narrowest beamwidth of 7°, the AP's more focused beams yield sufficiently high individual SNRs, that most users can be individually served at peak MCS, yielding modest gains for *SIMBA*.[1] Nonetheless, even in this case of 7° beams, *SIMBA* improves average rate from 3.08 Gbps for *Single User* to 3.31 Gbps for *SIMBA-opp* and 3.73 Gbps for *SIMBA-mr*, gains of 7% and 21% respectively.

Second, Fig. 12 also depicts the achievable rate for a transmit beamwidth of 20°. The figure indicates that as the beamwidth increases, there is a significant drop in the *Single User* average data rate due to the reduction in the directivity gain arising from beamwidth-MCS tradeoff. Nonetheless, this creates the opportunity to group additional users as the probability of covering more users increases with increased beamwidth. Therefore, the achievable aggregate rate of *SIMBA-opp* and *SIMBA-mr* approximately 45% and 63% greater than *Single User*, respectively.

This effect is more pronounced for a transmit beamwidth of 80° as shown in Fig. 12 which reflects more than 1.5× multi-user rate gains of *SIMBA-opp* and *SIMBA-mr* over *Single User*. This is attributed to the fact that increased beamwidth increases the number of possible beam-sharing opportunities and also leads to increased SNR spread which is well exploited by *SIMBA*. The origin of increased SNR spread for wider beamwidths can be assessed by studying the highest received SNR for each user and this is explored below.

Lastly, we observe that despite its computational simplicity and minimal training overhead, *SIMBA-opp* realizes performance quite close to *SIMBA-mr* across all beamwidths.

MCS Distribution. The origin of increased SNR spread for wider beamwidths can be assessed by studying the highest received SNR for each user. Fig. 13 depicts the histogram of the *Single User* MCS index (which increases with PHY rate) for different beamwidth. First, for a transmit beamwidth of 7°, the figure indicates that the majority of the users' rates are concentrated at the highest MCS levels with 60% of the users at MCS index 11 which is equal to PHY rate of 3.85 Gbps.

[1]While in principle, an extremely narrow beam would preclude multi-user grouping entirely, this effect was not observed in our experiments.

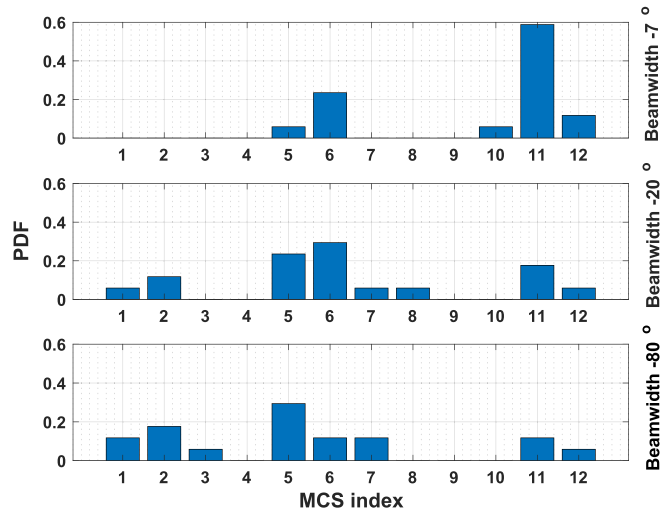


Fig. 13. Rate distribution of all the users for different transmit beamwidths.

This indicates that users are in high SNR regime and each user attains the best rate with an average rate of 3.05 Gbps across all locations. The lower multi-user gains in *SIMBA* for the 7° beamwidth case is because the transmit beam arising from the best (AP-user) beam pair cannot accommodate a second user that could share the beam and jointly maximize aggregate rate for the users.

Second, as beamwidth increases, there is a significant drop in the *Single User* data rate due to the reduction in directivity gain. This can be observed from Fig. 13 which shows that PHY rates for 20° and 80° beamwidth are distributed mostly in the low-moderate MCS index range of 1-8 with fewer users in the highest MCS index levels 10-12. Interestingly, in these wider beamwidth cases, *SIMBA* exploits this SNR (and MCS) spread to group users with net gain.

Finding: (i) Increasing transmit beamwidth leads to higher SNR spread among users providing opportunities to group additional users and hence increase aggregate rate for multi-user transmission. (ii) Wider beamwidth can also exploit multiple paths in addition to a much wider reception signal along the LOS path, improving user grouping opportunities. (iii) SIMBA-opp and SIMBA-mr strategies outperform Single User by 58% and 75% with 80° beamwidth.

D. Scaling Group Size

Thus far, we have considered multi-user transmissions having two users per transmission. Here, we increase group size beyond two to study the viability of further multi-user data rate gains.

Setup. We use the same node placement as depicted in Fig. 7. To ensure a fair comparison between different schemes of *SIMBA* for a fixed group size, we fix the service order (or the reference user) to follow the location index. For each group size and each *SIMBA* policy, we average the performance over all user groups and compare it with the baseline scheme *Single User*.

Aggregate Data Rate. Figure 14 shows the achievable aggregate rate averaged over all possible user groups for

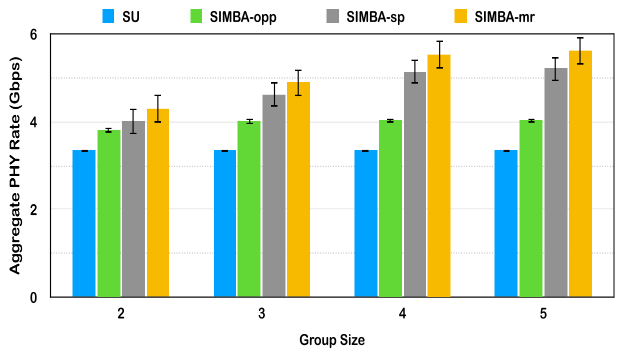


Fig. 14. Achievable aggregate rate vs. number of simultaneous users.

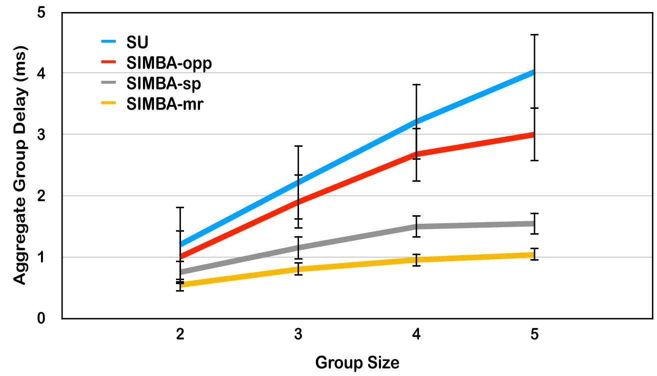


Fig. 15. Aggregate group delay vs. number of grouped users.

different group sizes. First, observe that *SIMBA-mr* achieves approximately 1.5× and 1.7× rate gains via simultaneous transmission of 3 and 4 users respectively. However, the improvement in aggregate rate has diminishing returns when increasing the group size from 4 to 5. Although each user may use a different MCS, we restrict to use the same coding rate for simplicity. Thus, when five users are grouped, one user may not contribute enough to the aggregate rate due to lack of sufficient SNR spread with other users. The limiting factor responsible for such a case is related to the number of modulation quantification levels used, four in this case (BPSK, QPSK, 16-QAM and 64-QAM). Thus, if each user in a group has been assigned to each distinct modulation, the number of simultaneous users that can be served by the AP while having a gain effectively saturates for a group size of four.

Second, *SIMBA-sp*'s performance is quite close to *SIMBA-mr*, despite its significantly reduced search space. This is because *SIMBA-sp* always selects the transmit beam in favor of the user with the maximum SNR which contributes the most to the sum rate and uses this beam subsequently for other users such that it can efficiently exploit the SNR disparity among the users. In addition, the users receive data during the entire transmission owing to which each user in the group attains a better rate in comparison to *Single User*, in which each user receives data only in its allocated time slot.

Third, although *SIMBA-opp* achieves 14% aggregate rate gain over *Single User* for a simultaneous transmission to two users, it has only marginal aggregate rate increase as the group size increases beyond 2. As *SIMBA-opp* finds users that share the reference user's transmit beam in each user group, unfortunately, for larger groups, there is often no user in a position which is sufficiently angularly located with the reference user that would also have sufficient SNR spread.

Finding: (i) Beamforming to 4 different users using SIMBA can improve aggregate rate approximately two-fold over Single User transmission. (ii) SIMBA-sp, although limited in search space, has performance within 93% of SIMBA-mr and outperforms Single User by two-fold. (iii) SIMBA-sp and SIMBA-mr gains over SIMBA-opp points out that the flexibility of having training information on all TX-RX beams in contrast to only having knowledge of the maximum SNR beam pair, helps find users with larger SNR spread and therefore attain higher rate gains. (iv) As the group size increases, the gains in comparison to Single User do not increase linearly and are constrained by the MCS quantification levels and the need for sufficient SNR spread among the grouped users.

E. Aggregate Group Delay

We define aggregate group delay as the total time required to serve a fixed number of bytes for all members of the group. In particular, we define total transmission time, or group delay, as follows: Consider a test user group G, where the number of users varies from 2 to 5. In *Single User*, transmissions are performed sequentially until the all members of the group have been served. In *SIMBA*, the transmission time for the shared beamformed transmissions for users in G is determined by the user operating at the lowest rate. For each sub-topology, we consider a 1 MB frame transmitted by each beam to the users in G during the data transmission period. We use the same setup as in Section IV-D.

Figure 15 shows total delay vs. number of grouped users. First, observe that the total group delay for *Single User* increases nearly linearly with group size as it is the sum of the transmission times for all users in the group. While *Single User* takes about 2.2× more time than *SIMBA-mr* for a two-user transmission, its performance gap increases compared to *SIMBA-mr* such that it takes 3.3× more time for five users.

Second, for a group size of 3, the total group delay for *SIMBA-opp* is 1.89 ms, which improves *Single User* transmission time by 15%. However, comparing the performance of *SIMBA-opp* with *SIMBA-mr*, although marginally close for a group size of 2, the gap increases beyond 2. This is because *SIMBA-opp*, which attempts to choose users that share a beam with the maximum SNR beam of the reference user, most often resulted in low SNR spread cases due to the unavailability of a sufficient number of users sharing the same best transmit beam.

On the other hand, *SIMBA-sp* and *SIMBA-mr* utilize the diversity of the available beams among the selected users in finding the beam grouping solution which together reduce the data transmission time almost two-fold compared to *Single User*. The beam grouping solution of *SIMBA-sp*, although inferior to *SIMBA-mr*, selects beams such that the time taken by the bottleneck-rate user in the shared beam is minimized. As a result, with an increasing number of grouped users, *SIMBA-sp*'s gain remains close (with marginal loss) to the performance of *SIMBA-mr*.

Finding: As group size increases, SIMBA-sp yields a two-fold reduction in group delay compared to Single User with approximately the same performance SIMBA-mr, despite having reduced search space in comparison to SIMBA-mr.

F. Comparison of SIMBA and OFDMA

Here we briefly compare *SIMBA* against *Single User* as an OFDMA scheme from the perspective of their design principles. Although the *Single User* does TDMA, it holds a broader significance as a baseline in this paper. It is a representative of all classes of systems employing Orthogonal Multiple Access (OMA) schemes [21], [22] in which multiple users are allocated to orthogonal resources in the time, frequency, code, or to their combinations.

First, we see that OFDMA shares frequency resources or follows spectrum sharing where the data for each user is assigned to a subset of subcarriers. Spectral efficiency is low when some subcarrier channels, are allocated to users with lower SNR. On the other hand, *SIMBA* allows all of the subcarriers to be used by each user in the group. Hence, the subcarriers allocated to the users with low SNR can still be accessed by those with high SNR, which significantly improves the spectral efficiency. Theoretically, as a benefit of orthogonal subcarrier allocation schemes like OFDMA, there is no interference among users. However, the maximum number of supportable users is rigidly restricted by the number of orthogonal subcarriers which becomes a hard limit in dense scenarios requiring massive connectivity.

Second, according to the downlink multi-user capacity analysis found in [23], one can see that the orthogonal schemes like OFDMA are in general sub-optimal, except for one point: when the amount of degrees of freedom (time and bandwidth) allocated to each user is proportional to its receive power. However, when there is a large disparity between the received powers of two users, this operating point is highly unfair since most of the degrees of freedom is given to the strong user, and the weak user has hardly any rate. In contrast, for any rate pair achieved by OFDMA, *SIMBA* supports more equitable user fairness. That is *SIMBA* finds a beam configuration that achieves rate pairs that are strictly larger as the asymmetry of received SNRs between the two users deepens, whereas OFDMA has to allocate a significant fraction of the subcarriers to the weak user to achieve the near single-user performance and this causes a large degradation in the performance of the strong user.

V. Related Work

Multi-user 60 GHz WLANs with Multiple RF Chains. Multi-user multi-stream transmission is specified in the downlink for the next 60 GHz WLAN standard IEEE 802.11ay with at least one RF chain per stream [2]. Recent work in this context has shown that multiple users falling into the same transmit beam experience significant inter-user interference and cannot separate and decode their data streams even when zero-forcing is applied [6]. Prior work has also shown how to set digital and analog weights to maximize sum capacity [4], [24]–[26]. In contrast, we realize multi-user transmission in which the number of users exceeds the number of RF chains.

Multi-user Below 6 GHz with Multiple RF Chains. Work on MU-MIMO below 6 GHz includes user grouping based on channel state and/or expected transmission time [27], [28]. Likewise, other work targeted user grouping without channel state information by exploiting the rich scattering propagation in indoor environments below 6 GHz [29]–[31]. Unfortunately, such techniques cannot be applied to our scenario as 60 GHz channels lacks the rich scattering propagation environment prominent below 6 GHz [32]. Moreover, we consider only a single RF chain.

Multi-user with a Single RF Chain. Extensive prior work in both uplink and downlink has realized multiple transmission or reception while also requiring only a single RF chain, including Non-Orthogonal Multiple Access (NOMA) [8]–[13], [33]–[37] and hierarchical modulation [38], [39], which employs the same philosophy of overlayed constellations by not relying on a different space, time, or frequency resource for simultaneous transmission. In contrast, our design employs standard-compliant constellations as our focus is not the physical layer design itself, but rather the WLAN architecture that can exploit such features. In this context, we realize the first 60 GHz WLAN design that realizes multi-stream multi-user communication on a single RF chain as well as the first experimental study of such functionality.

Lastly, our recent conference paper in [40] studies the problem of multi-user beamforming on a Single RF Chain in 60 GHz MIMO WLANs. This article is the extended version of [40] and has the same core idea. However, here, we have included more discussions, over-the-air experiments and analyses.

VI. Conclusion

We presented *SIMBA*, a novel system that enables multi-user multi-stream transmission via a single RF chain in 60 GHz WLANs. We proposed and evaluated *(i) SIMBA-mr* which offers the maximum data rate with a relaxed focus on overhead, *(ii) SIMBA-opp* to opportunistically add users that can share the beam of the next user in the queue, and *(iii) SIMBA-sp* which exploits SNR differences among clients to bridge the tradeoffs of *SIMBA-mr* and *SIMBA-opp*. We evaluated the multi-user aggregate rate gains as a function of receiver distance and showed that exploting SNR heterogeneity among clients is a key source of gain. We further showed how widening beamwidth, traditionally a source of rate loss in single-user systems, provides a source of gain for *SIMBA* via new grouping opportunities with high SNR diversity. We increased group size and demonstrated the link between SNR diversity and the fundamental limit imposed by MCS quantization. We also compared the delay incurred for data transmission of users in *SIMBA* over *Single User* and observe a two-fold net reduction in total delay. Lastly, we have demonstrated advantages of SIMBA, a Non-Orthogonal Multiple Access Scheme over TDMA, an Orthogonal Multiple Access scheme. A promising avenue for future work is to experimentally compare with SIMBA with alternate orthogonal schemes such as OFDMA and to study SIMBA as an enhancement to OFDMA, i.e., utilizing SIMBA in preferential time-frequency blocks.

REFERENCES

[1] O. Bejarano, E. Knightly, and M. Park, "IEEE 802.11ac: From channelization to multi-user MIMO," *IEEE Commun. Mag.*, vol. 51, no. 10, pp. 84–90, Oct. 2013.
[2] Y. Ghasempour, C. R. C. M. da Silva, C. Cordeiro, and E. W. Knightly, "IEEE 802.11ay: Next-generation 60 GHz communication for 100 Gb/s Wi-Fi," *IEEE Commun. Mag.*, vol. 55, no. 12, pp. 186–192, Dec. 2017.
[3] *IEEE Standard for Information Technology—Telecommunications and Information Exchange Between Systems—Local and Metropolitan Area Networks—Specific Requirements—Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications Amendment 3: Enhancements for Very High Throughput in the 60 GHz Band*, IEEE Standard 802.11ad-2012 (Amendment to IEEE Standard 802.11-2012, as amended by IEEE Standard 802.11ae-2012 and IEEE Standard 802.11aa-2012), 2012, pp. 1–628.
[4] Y. Ghasempour, M. K. Haider, C. Cordeiro, D. Koutsonikolas, and E. Knightly, "Multi-stream beam-training for mmWave MIMO networks," in *Proc. ACM MobiCom*, Oct. 2018, pp. 225–239.
[5] S. K. Saha *et al.*, "X60: A programmable testbed for wideband 60 GHz WLANs with phased arrays," *Comput. Commun.*, vol. 133, pp. 77–88, Jan. 2019.
[6] Y. Ghasempour and E. W. Knightly, "Decoupling beam steering and user selection for scaling multi-user 60 GHz WLANs," in *Proc. ACM MobiHoc*, Jul. 2017, pp. 1–10.
[7] M. K. Haider and E. W. Knightly, "Mobility resilience and overhead constrained adaptation in directional 60 GHz WLANs: Protocol design and system implementation," in *Proc. ACM MobiHoc*, Jul. 2016, pp. 61–70.
[8] S.-H. Chang, M. Rim, P. C. Cosman, and L. B. Milstein, "Optimized unequal error protection using multiplexed hierarchical modulation," *IEEE Trans. Inf. Theory*, vol. 58, no. 9, pp. 5816–5840, Sep. 2012.
[9] S. Vanka, S. Srinivasa, Z. Gong, P. Vizi, K. Stamatiou, and M. Haenggi, "Superposition coding strategies: Design and experimental evaluation," *IEEE Trans. Wireless Commun.*, vol. 11, no. 7, pp. 2628–2639, Jul. 2012.
[10] B. Masnick and J. Wolf, "On linear unequal error protection codes," *IEEE Trans. Inf. Theory*, vol. IT-13, no. 4, pp. 600–607, Oct. 1967.
[11] S. Jakubczak and D. Katabi, "SoftCast: One-size-fits-all wireless video," in *Proc. ACM SIGCOMM Conf.*, 2010, pp. 449–450.
[12] X. L. Liu, W. Hu, Q. Pu, F. Wu, and Y. Zhang, "ParCast: Soft video delivery in MIMO-OFDM WLANs," in *Proc. 18th Annu. Int. Conf. Mobile Comput. Netw.*, 2012, pp. 233–244.
[13] S. Sen, S. Gilani, S. Srinath, S. Schmitt, and S. Banerjee, "Design and implementation of an 'approximate' communication system for wireless media applications," in *Proc. ACM SIGCOMM Conf.*, 2010, pp. 15–26.
[14] *IEEE 802.11ay-2021—IEEE Standard for Information Technology—Telecommunications and Information Exchange Between Systems Local and Metropolitan Area Networks—Specific Requirements Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications Amendment 2: Enhanced Throughput for Operation in License-exempt Bands Above 45 GHz*, IEEE Standard 802.11, 2021.
[15] S. Sur, V. Venkateswaran, X. Zhang, and P. Ramanathan, "60 GHz indoor networking through flexible beams: A link-level profiling," in *Proc. ACM SIGMETRICS Int. Conf. Meas. Modeling Comput. Syst.*, Jun. 2015, pp. 71–84.
[16] S. Naribole and E. Knightly, "Scalable multicast in highly-directional 60-GHz WLANs," *IEEE/ACM Trans. Netw.*, vol. 25, no. 5, pp. 2844–2857, Oct. 2017.
[17] C. R. C. M. da Silva, A. Lomayev, C. Chen, and C. Cordeiro, "Analysis and simulation of the IEEE 802.11ay single-carrier PHY," in *Proc. IEEE ICC*, May 2018, pp. 1–6.
[18] A. Khattab, J. Camp, C. Hunter, P. Murphy, A. Sabharwal, and E. W. Knightly, "WARP: A flexible platform for clean-slate wireless medium access protocol design," *ACM SIGMOBILE Mobile Comput. Commun. Rev.*, vol. 12, no. 1, pp. 56–58, Jan. 2008.
[19] H. Xu, V. Kukshya, and T. S. Rappaport, "Spatial and temporal characteristics of 60-GHz indoor channels," *IEEE J. Sel. Areas Commun.*, vol. 20, no. 3, pp. 620–630, Apr. 2002.
[20] T. S. Rappaport, G. R. MacCartney, M. K. Samimi, and S. Sun, "Wideband millimeter-wave propagation measurements and channel models for future wireless communication system design," *IEEE Trans. Commun.*, vol. 63, no. 9, pp. 3029–3056, Sep. 2015.
[21] A. W. Scott and R. Frobenius, "Multiple access techniques: FDMA, TDMA and CDMA," in *RF Measurements for Cellular Phones and Wireless Data Systems*. Hoboken, NJ, USA: Wiley, Jan. 2008, pp. 413–429.
[22] H. Li, G. Ru, Y. Kim, and H. Liu, "OFDMA capacity analysis in MIMO channels," *IEEE Trans. Inf. Theory*, vol. 56, no. 9, pp. 4438–4446, Sep. 2010.
[23] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge, U.K.: Cambridge Univ. Press, 2005.
[24] A. Alkhateeb, G. Leus, and R. W. Heath, Jr., "Limited feedback hybrid precoding for multi-user millimeter wave systems," *IEEE Trans. Wireless Commun.*, vol. 14, no. 11, pp. 6481–6494, Nov. 2015.
[25] M. K. Samimi, S. Sun, and T. S. Rappaport, "MIMO channel modeling and capacity analysis for 5G millimeter-wave wireless systems," in *Proc. IEEE 10th EuCAP*, Apr. 2016, pp. 1–5.
[26] R. A. Stirling-Gallacher and M. S. Rahman, "Multi-user MIMO strategies for a millimeter wave communication system using hybrid beamforming," in *Proc. IEEE ICC*, Jun. 2015, pp. 2437–2443.
[27] S. Sur, I. Pefkianakis, X. Zhang, and K.-H. Kim, "Practical MU-MIMO user selection on 802.11ac commodity networks," in *Proc. ACM Mobicom*, Oct. 2016, pp. 122–134.
[28] T. Tandai, H. Mori, and M. Takagi, "Cross-layer-optimized user grouping strategy in downlink multiuser MIMO systems," in *Proc. IEEE VTC*, Apr. 2009, pp. 1–6.
[29] N. Anand, J. Lee, S.-J. Lee, and E. W. Knightly, "Mode and user selection for multi-user MIMO WLANs without CSI," in *Proc. IEEE INFOCOM*, Apr. 2015, pp. 451–459.
[30] X. Xie and X. Zhang, "Scalable user selection for MU-MIMO networks," in *Proc. IEEE INFOCOM*, Apr. 2014, pp. 808–816.
[31] A. B. Flores, S. Quadri, and E. W. Knightly, "A scalable multi-user uplink for Wi-Fi," in *Proc. USENIX NSDI*, 2016, pp. 179–191.
[32] G. R. Maccartney, T. S. Rappaport, S. Sun, and S. Deng, "Indoor office wideband millimeter-wave propagation measurements and channel models at 28 and 73 GHz for ultra-dense 5G wireless networks," *IEEE Access*, vol. 3, pp. 2388–2424, 2015.
[33] Z. Ding, P. Fan, and H. V. Poor, "Random beamforming in millimeter-wave NOMA networks," *IEEE Access*, vol. 5, pp. 7667–7681, 2017.
[34] J. Cui, Y. Liu, Z. Ding, P. Fan, and A. Nallanathan, "Optimal user scheduling and power allocation for millimeter wave NOMA systems," *IEEE Trans. Wireless Commun.*, vol. 17, no. 3, pp. 1502–1517, Mar. 2018.
[35] Z. Wei, L. Zhao, J. Guo, D. W. K. Ng, and J. Yuan, "A multi-beam NOMA framework for hybrid mmWave systems," in *Proc. IEEE ICC*, May 2018, pp. 1–7.
[36] Z. Xiao, L. Zhu, J. Choi, P. Xia, and X.-G. Xia, "Joint power allocation and beamforming for non-orthogonal multiple access (NOMA) in 5G millimeter wave communications," *IEEE Trans. Wireless Commun.*, vol. 17, no. 5, pp. 2961–2974, May 2018.
[37] Z. Wei, D. W. K. Ng, and J. Yuan, "Beamwidth control for NOMA in hybrid mmWave communication systems," 2019, *arXiv:1902.04227*.
[38] H. Jiang and P. A. Wilford, "A hierarchical modulation for upgrading digital broadcast systems," *IEEE Trans. Broadcast.*, vol. 51, no. 2, pp. 223–229, Jun. 2005.
[39] X. Wang and L. Cai, "Proportional fair scheduling in hierarchical modulation aided wireless networks," *IEEE Trans. Wireless Commun.*, vol. 12, no. 4, pp. 1584–1593, Apr. 2013.
[40] K. P. Dasala, J. M. Jornet, and E. W. Knightly, "SIMBA: Single RF chain multi-user beamforming in 60 GHz WLANs," in *Proc. IEEE INFOCOM*, Jul. 2020, pp. 1499–1508.