BUSH: Empowering Large-Scale MU-MIMO in WLANs With Hybrid Beamforming

Zhe Chen[1,4], Xu Zhang[1,4], Sulei Wang[2], Yuedong Xu[2], Jie Xiong[3], Xin Wang[1,4]
[1]School of Computer Science, Fudan University, Shanghai, China
[2]School of Information Science and Technology, Fudan University, Shanghai, China
[3]School of Information Systems, Singapore Management University, Singapore
[4]Engineering Research Center of Cyber Security Auditing and Monitoring, Ministry of Education
{zhechen13, xuzhang09, wangsl16, ydxu, xinw}@fudan.edu.cn, jxiong@smu.edu.sg

***Abstract*—Large-scale MU-MIMO is a promising technology to scale network capacity and the capacity gain grows linearly with the numbers of antennas and users in theory. However, its practical deployment faces three critical challenges in the state-of-the-art WLANs: i) the demand of a large number of expensive RF chains; ii) the linear growth of feedback overheads with the number of antennas; iii) the lack of scalable user selection scheme for a large user population.**

In this paper, we design BUSH, a large-scale MU-MIMO prototype that performs scalable beam user selection with hybrid beamforming for phased-array antennas in legacy WLANs. The architecture of BUSH consists of three components. Firstly, a low complexity algorithm assigns each pair of RF chain and analog beam to the users to effectively reduce channel correlation and cross-talk interference without instantaneous CSI feedbacks. Secondly, as a prerequisite of user selection, BUSH presents a low-overhead probing scheme in multi-carrier WLANs, and designs a highly accurate blind Power Azimuth Spectrum (PAS) estimation algorithm using a single RF chain. Thirdly, the phased-array antennas use analog beamforming to steer spatial beams toward each selected downlink user, and the finite number of RF chains use beamforming to further mitigate the interference among users. We implement BUSH on the WARPv3 boards and evaluate its performance in more than 30 indoor scenarios. The experimental results show that in terms of total throughput BUSH outperforms the legacy 802.11ac by 2.08×, and an alternative benchmark system by 1.22× on average.

I. Introduction

Recently, wireless LAN (WLAN) is undergoing a paradigm shift from single-user to multi-user communication patterns with the introduction of the latest 802.11ac standard [6]. One key reason behind this move is the fact that devices are usually limited by one or two antennas even if the Access point (AP) has many more antennas available. At the same time, more and more wearables and portable devices equipped with WiFi are crowded around us nowadays, causing a much higher data transmission load on the existing WiFi infrastructure. It's predicted 3x more personal wearables [1] will be attached to the human bodies or deployed in the surrounding areas of us in the next few years. While 802.11ac has been widely adopted in enterprises and universities to mitigate the higher demand of connectivity problem, many other issues such as user selection remain challenging and unsolved.

This work was supported in part by 863 program of China (No. 2015AA016106), Natural Science Foundation of China (No. 61402114), Shanghai Municipal R&D Foundation (No. 14511101000), and Shenzhen Municipal Grant (No. KQCX20140519103756206). Yuedong Xu is the corresponding author.

In 802.11ac, multi-user MIMO (MU-MIMO) is employed and allows the AP to simultaneously use multiple antennas to transmit independent data streams to different users in the same frequency band. Users measure the channel to the AP and feed this *channel state information* (CSI) back to the AP sequentially. The AP then sends out different data streams such that each user receives only the data stream it needs, while other unwanted (interfering) data streams are suppressed (this process is called *beamforming*). A popular beamforming scheme is *zero-forcing beamforming* [21] (ZFBF), wherein undesired data streams are forced to zero at a user. MU-MIMO thereby allows the AP to better leverage its antenna resources and achieve multiplicative throughput increases.

Theoretically, downlink capacity grows linearly with the number of antennas [2]. To exploit the full potential of a MIMO WLAN, a recent trend is to deploy large antenna arrays at the AP side. Large-scale MU-MIMO is put forward to further enhance network capacity by equipping the transmitter with more than ten antennas [3], [4]. In MU-MIMO, beamforming is performed at the baseband of a transmitter, which requires one dedicated radio frequency (RF) chain at each antenna. An entire RF chain consists of many PHY hardware components such as the baseband processor, the de/modulator, power amplifier and the ADC/DAC [17]. Although large-scale MU-MIMO has been successfully prototyped in Argos [4] and in BigStation [3], it still faces three major challenges for real life deployment.

- Firstly, RF chains are expensive so that associating each antenna with a RF chain is economically inefficient in a large-scale MU-MIMO system.
- Secondly, the overhead grows linearly with the number of antennas. However, sequential feedback of CSI information from all users brings in a large latency and thus degrades the throughput performance.
- Thirdly, user selection policy is not specified in 802.11ac standard and it plays an important role in the transmission throughput. In 802.11ac, the number of antennas attached to a single AP is limited to 8 while the number of users is usually large so the AP needs to select a group of users

978-1-5090-5336-0/17/$31.00 ©2017 IEEE

for reception at each time slot.

To address the first challenge, our solution is to adopt a digital/analog hybrid beamforming that uses several RF chains to steer a large number of phased-array antennas (e.g. [10]). Figure 1. illustrates the key difference of hybrid beamforming and digital beamforming architectures. Hybrid beamforming is a two-level beamforming architecture: one is analog beamforming that uses low cost phase shifters to control the phase of transmitted signals from each antenna; the other is the conventional digital beamforming via RF chains. The analog beamforming provides beamforming gain for each RF chain, and the digital beamforming suppresses the interference between each user to achieve spatial multiplexing gain. Note that purely digital beamforming requires the same number of RF chains as the antennas which is not realistic in many scenarios.

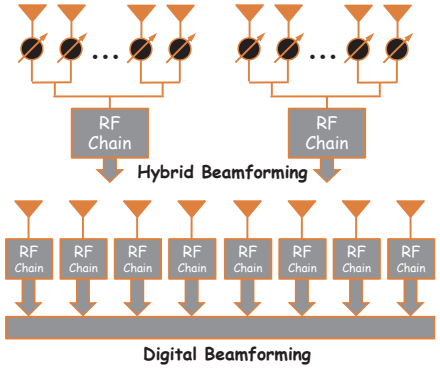

Fig. 1. Architecture of hybrid beamforming and digital beamforming

To address the next two challenges, we propose a novel scheme to simultaneously obtain the *angle-of-arrival* (AoA) and power information from all the users with just one single probing packet and then select the users with the spatial information obtained. Thus, we only need to acquire CSI feedbacks from the selected users, significantly reducing the overhead and latency.

In this paper, we design, implement and evaluate BUSH, a practical large-scale MU-MIMO system compatible with the latest 802.11ac. BUSH performs a joint beam and user selection with hybrid beamforming. The basic principle is to select users with very different strongest AoA peaks in one group to reduce the cross-talk interference. We also note that the widely used MUSIC AoA is not a power function but a probability spectrum so we propose a novel blind received power estimation scheme to extract the effective received power at each AoA direction.

BUSH has two folds of advantages. By employing phased-array antennas, the AP can use multiple antenna elements to achieve a high digital beamforming gain, while the multiplexing gains are not compromised since each user is assigned with a RF chain. On the other hand, when the number of users exceeds the number of RF chains, the agile beam-user paring can effectively reduce the cross-talk interference in the analog beamforming, and the RF chains can further utilize the fine-grained CSI feedbacks from the receivers to cancel interference between each receiver. Practical implementation of BUSH in WLANs entails multiple unique challenges. Estimating AoAs of multiple users simultaneously is nontrivial in 802.11 WLANs. WiFi systems are not centrally controlled so that users do not always send pilots to the AP as opposed to LTE systems, and the sounding procedure in WiFi systems is much longer. What is worse, when the number of antennas is larger than the number of RF chains, the signal received at each antenna is combined at the RF chain, making the traditional MUSIC algorithm not applicable to obtain the AoA estimations. In MU-MIMO with hybrid beamforming, the analog beamforming selects among a finite number of pre-determined beam patterns, and digital beamforming chooses from a finite number of RF chains. Hence, the beam-user selection becomes a combinatorial decision making problem. To summarize, we make the following contributions in this paper:

- We design BUSH, an MIMO system enabling digital-analog hybrid beamforming, to address the mismatching between the limited RF chains and the large number of antennas.
- We creatively utilize different sub-carriers for different users to simultaneously obtain the AoA information from all the users in just one round of probing, significantly reducing the latency.
- As conventional MUSIC algorithm fails when signals from multiple antennas are mixed, we propose a novel blind AoA estimation method to obtain accurate AoA information and we adopt the power azimuth spectrum (PAS) method to obtain the power information.
- We design an optimization framework for joint beamforming and user selection to achieve a higher throughput.
- We implement and evaluate BUSH on WARP platform and extensive experiments demonstrate the effectiveness of BUSH, outperforming the state-of-the-art systems.

The remainder of this paper is organized as follows: Section II presents the motivation example and key observations followed by the system overview of BUSH in III. We describe the detailed system design in Section IV and Section V. Implementation and evaluations are presented in Section VI. We review the related work in Section VII and conclude our work in Section VIII.

II. Motivation and Observation

A. Motivating Scenarios

Owing to limited number of RF chains in large-scale MIMO systems, only a few users will be served simultaneously in each transmission slot. Previous work in [24] has shown that the user selection policy plays an important role in achieving a high capacity in digital beamforming systems. In hybrid beamforming MIMO systems, user selection is also essential. When the number of users becomes large, their total throughput heavily relies on the user selection algorithm. In what follows, we will show that the users' relative locations greatly affect the throughput.

Figure 2 illustrates a motivating example with a single AP serving three users, $U1$, $U2$ and $U3$. Here, $U1$ and $U2$ are

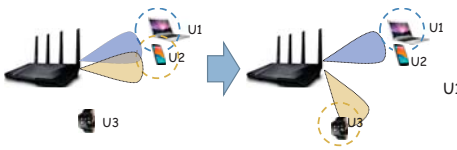

Fig. 2. An example for user and beam selection

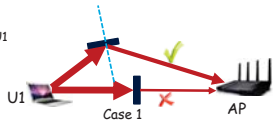

Fig. 3. Beamforming direction selection with AoA and power information.

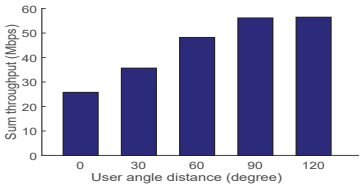

Fig. 4. Hybrid beamforming throughput in different user spatial separations

closely located, while $U3$ is far away from $U1$. Without a carefully designed user selection, the AP may use two analog beams to schedule data streams to $U1$ and $U2$ simultaneously. Although the beams can be adjusted to match $U1$ and $U2$'s locations and digital beamforming (e.g. zero-forcing) can be further applied to mitigate the cross-talk interference, their received SNRs are still limited by the large mutual interference between two beams. A better strategy is to select $U1$ and $U3$ for transmission in which the cross-talk interference is effectively reduced. One can clearly see that a satisfactory user selection strategy can be achieved if the AP is aware of the users' directions with respect to the AP. In realistic wireless systems, the angle information of a user is not known as a priori and it may even get changed over time if the user is not static. Further, the direct path between the user and AP may not always be the strongest if there is an object in-between as shown in Fig. 3. In Case 1, we will employ the reflection path with a stronger power to calculate the angle distance between two users. When another device has a similar AoA as the stronger reflection path shown in Case 2, we may choose the weaker path as less interference will be caused.

B. Validation with benchmark experiments

1) Necessity of User Selection: We implement the hybrid beamforming system described in [10] and measure the spatial angle distance between the two users. The AP has two RF chains, each is equipped with eight antennas. Then, we vary the angle distance between two users from 0 to 120 degrees. From Figure 4, we can see that the throughput performance of hybrid beamforming system is affected by the angle distance. When the angle distance gets larger, the sum throughput also increases. Hence, in hybrid beamforming systems, user selection strategy clearly affects the throughput performance.

Location	1	2	3
1 ms	(-15.00, 0.0192)	(-49.50, 0.0354)	(37.23, 0.0570)
10 ms	(-15.30, 0.0188)	(-49.90, 0.0322)	(36.80, 0.0540)
100 ms	(-15.45, 0.0191)	(-44.87, 0.0305)	(36.76, 0.0499)

TABLE I
The stability of PAS over time

2) PAS Stability: In order to validate the PAS stability, we let each transmitter keeps sending packets for one second, and calculate the PASs in three intervals: $0 \sim 1$ ms, $10 \sim 11$ ms and $100 \sim 101$ ms respectively. We randomly choose three users at different locations and show the strongest dominant PAS for each user in Table I. Each 2-tuple represents the received AoA (degree) and relative power that are averaged over 1 ms. One can observe that the dominant PAS of each user varies very slowly compared with the fast changing channel state information (CSI). No matter whether the users are static or mobile, PASs stay almost unchanged within the maximum physical frame duration (i.e. 5.464ms) of 802.11ac. This implies that we can adopt the stable PAS information to design downlink beamforming for users effectively.

III. Overview of System

BUSH focuses on exploring large-scale MU-MIMO in current 802.11 networks. Thus, BUSH can be implemented via modifying firmware, and assembling the phased-array antennas in the off-the-shelf AP with a little modification. Figure 5. shows the components of BUSH, where the RF chains of AP are equipped with phased-array antennas. The blocks with red dashed lines are the key components of BUSH.

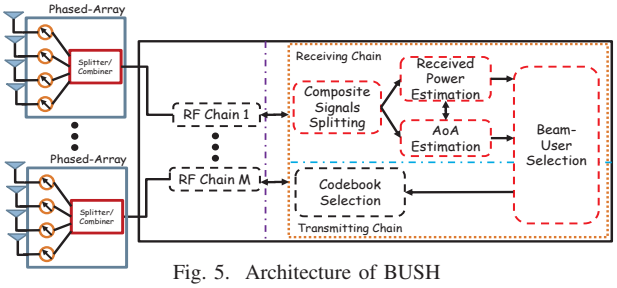

Fig. 5. Architecture of BUSH

The three-key steps of BUSH operations are as follows:

Step 1: AoA estimation. With the novel probing method, BUSH will receive the required information from different users at assigned different sub-carriers with just one probing packet. Once a RF-chain receives this probing packet, the AP will estimate the dominant AoA of each user with the proposed blind AoA estimation approach P-MUSIC (Sec. IV-B).

Step 2: Effective received power estimation. After AoA estimation of each user, the effective received power of dominant AoA can be obtained by the proposed blind received power estimation method (Sec. IV-C). With both angle and power information estimated, the PAS of each user is obtained.

Step 3: Beam-User selection. When the number of users are larger than RF chains in the hybrid beamforming system, random user selection or one-to-one mapping scheme may degrade the overall throughput performance. To optimize the throughput performance, the BUSH AP employs a novel beam-user selection algorithm (Sec. V) to pick up the users for downlink service.

IV. Blind PAS Estimation

Challenges: Estimating PAS (i.e. AoA and received power) in a phased-array with less number of RF chains is very challenging due to the fact that the CSI on each antenna element is not available. The RF chain only obtains a composite CSI that

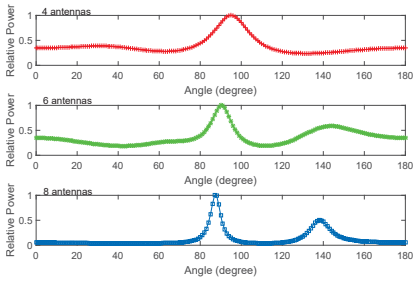

Fig. 6. Impact of antenna numbers on the accuracy of AoA Estimation

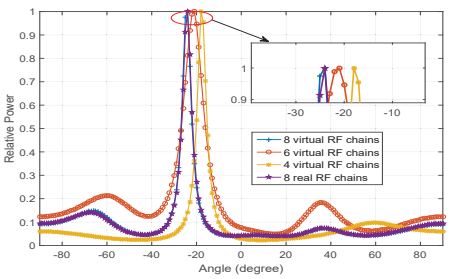

Fig. 7. AoA estimation with different virtual RF chains

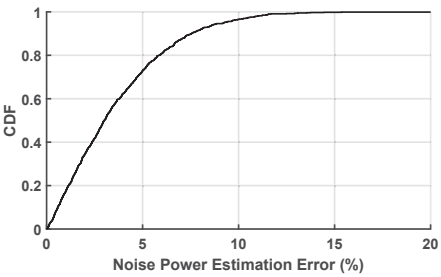

Fig. 8. Noise power estimation error

is the sum of the CSIs on all the antenna elements multiplied by the phase shifter weights. Hence, classical algorithms such as MUSIC cannot be directly applied to calculate the AoAs [9]. Recently, Hekaton [10] presented a compressive sensing based AoA estimation scheme for phased-array antenna using the sounding reference signal of LTE. Probeam [14] employed a beam sweep method to estimate the received SINR but incurred a high probing overhead.

BUSH presents a novel blind PAS estimation method which works with only one *single* RF chain without any reference signal. Our design is based on two observations: (1) the phased-array antenna has a switching latency between codebook entries as small as $1.2\mu s$ [10]; (2) because only coarse AoA resolution is needed for beamforming purposes, a small number of antennas is enough to achieve satisfactory AoA estimation resolution. We measure AoA in Figure 6 with different numbers of antennas. The experimental results clearly show that six to eight antennas are good enough to achieve highly accurate AoA estimations.

A. PAS-orthogonal probing

We propose a novel PAS-orthogonal probing scheme for quick PAS measurements from all the users with just one probing packet. This probing scheme should fulfill three requirements: (1) able to accommodate a large number of users, (2) low feedback overhead/latency and (3) high accuracy. To obtain the required information with just one probing packet, we creatively allocate different sub-carriers to different users. In this way, many users can send pilots simultaneously to the AP in just one probing packet as there are 64 sub-carriers in WiFi systems. However, although OFDM is widely used in cellular networks, it is hard to directly use it in WiFi networks because the users in LTE maintain a tight timing synchronization with the base station. In order to enable OFDMA in WiFi network, we employ the cyclic prefix (CP) to protect the transmissions against slight time misalignment [12]. We also allocate each user multiple continuous sub-carriers to avoid the frequency selective fading. Note that in 802.11ac, with a higher bandwidth (160MHz) supported, more sub-carriers are available and thus more users can be accommodated in one single probing packet.

BUSH does not need to decode the payload of the probing packet as we only need the preamble part to estimate PAS. BUSH is thus robust against noise as the preamble part is usually transmitted at the base rate. Another advantage is that BUSH needs very few samples for AoA signal processing. How many sub-carriers does one user need for accurate AoA estimation? The results are shown in Figure 9. When we use only one sub-carrier to estimate the AoA spectrum, the result is unstable. When we increase the number of sub-carriers and average the correlation matrix for estimation, the AoA spectrum becomes more stable. We can see that with 4-6 sub-carriers, stable AoA spectrum can be well achieved in our experiments. In our implementation, we maintain a set that is initialized as the set of all sub-carriers. Each user is assigned with 4 sub-carriers in this set with the lowest frequency and the assigned sub-carriers are removed subsequently.

The operation of probing is illustrated in Figure 10 and briefly described as follow. First, each user will be assigned corresponding sub-carriers and this assignment information will be broadcasted in a polling packet to all the users. Second, when the AP broadcasts an NDPA frame, each uthe-orem:approximationser replies back simultaneously with data only on the assigned sub-carriers to itself. Note that the strict time synchronization like LTE is not needed here because we do not decode the data but just estimate the PAS information.

With the probing replies received from all the users in one round, BUSH AP proceeds to estimate the power azimuth spectrum from the received combined CSIs. In the next section, we first introduce a novel approach, namely phase-array MUSIC (P-MUSIC) to estimate the AoAs in BUSH.

B. Blind AoA Estimation

We begin with the description of spatial channel model in 802.11 MU-MIMO networks. Consider the AP equipped with a single RF chain and a uniform linear phased-array antenna (ULA) with multiple antenna elements, and a user equipped with a single antenna. The wireless channel consists of P dominant propagation paths, each of which is characterised by its AoA in the real spatial domain. Denote by $\rho_p(t)$ the complex path loss for the p^{th} dominant path, and denote by $\mathbf{n}(t)$ the additive white Gaussian noise. Then, the uplink signal received by the AP is given by

$$\mathbf{x}(t) = \sum_{p=1}^{P} \mathbf{a}(\theta_p)\rho_p(t)s(t) + \mathbf{n}(t), \quad (1)$$

where $s(t)$ is the transmitted signal by the user at time t. Here, $\mathbf{a}(\theta_p)$ represents the array steering vector, that is, $\mathbf{a}(\theta_p) = [1, e^{-j2\pi f_c d \sin\theta_p/C}, \cdots, e^{-j2\pi f_c (N-1) d \sin\theta_p/C}]^T$ where f_c is the carrier frequency, C is the speed of light, and θ_p is the AoA of the p^{th} path.

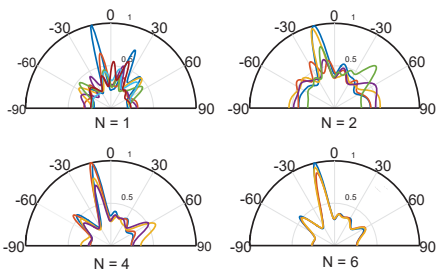

Fig. 9. The effect of number of sub-carriers on AoA spectrum.

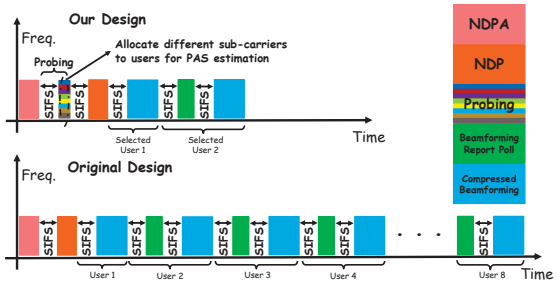

Fig. 10. Procedure of PAS-orthogonal probing

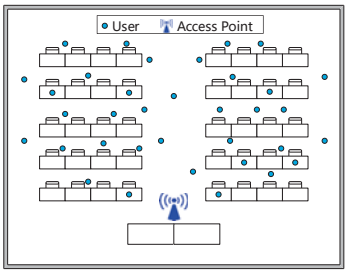

Fig. 11. Evaluation topologies

In a phased-array antenna, although the CSI from each antenna element is not available, we can still apply the orthogonal beam vector to collect a bulk of samples during a fixed instants denoted T_p. With the phased-array antenna, we can only acquire the composite CSIs at the RF chain.

The received signal at the RF chain, denoted by $\mathbf{y}$, is expressed as:

$$\mathbf{y}(t) = \mathbf{w}_{RF}^H \mathbf{A}(\mathbf{\Theta})\boldsymbol{\rho}(t)s(t) + \mathbf{w}_{RF}^H \mathbf{n}(t) \quad (2)$$

where $\mathbf{w}_{RF}$ is the analog beamforming vector implemented using analog phase shifters, and the modulus of all the elements in $\mathbf{w}_{RF}$ is 1 uniformly. Here, we have $\mathbf{A}(\mathbf{\Theta}) = [\mathbf{a}(\theta_1), \mathbf{a}(\theta_2), \cdots, \mathbf{a}(\theta_P)]$ and $\boldsymbol{\rho}(t) = [\rho_1(t), \rho_2(t), \cdots, \rho_P]^T(t)$ as the set of complex channel gains. Therefore, in order to obtain more degrees of freedoms (DoFs), we change the array steering vectors at different time slots where T is the maximum number of slots in our system. Note that T can not exceed the coherent time. Let $\tilde{\mathbf{y}}$ (resp. $\tilde{\mathbf{x}}$) be a set of composite CSIs, i.e. $\tilde{\mathbf{y}} = [\mathbf{y}(t), \mathbf{y}(t+1), \cdots, \mathbf{y}(t+T-1)]^T$ (resp. $\tilde{\mathbf{x}} = [\mathbf{x}(t), \mathbf{x}(t+1), \cdots, \mathbf{x}(t+T-1)]^T$). Then, we can obtain

$$\tilde{\mathbf{y}} = \mathbf{W}_r \tilde{\mathbf{x}} = [\mathbf{w}_{RF}(t), \cdots, \mathbf{w}_{RF}(t+T-1)]^T \tilde{\mathbf{x}}. \quad (3)$$

It is worth noting that to maximize the isolation between different beam patterns, the weight vectors are always orthogonal in our design [7]. Hence, if $T = N$ and for each $\mathbf{w}_i^H \mathbf{w}_j = 0$ when $i \neq j$, we can use $\mathbf{W}_r^{-1}$ to multiply $\mathbf{y}(t)$ to obtain $\mathbf{x}(t)$.

For a phased-array with N antennas, the autocorrelation matrix can also derive according to Eq. (3):

$$\mathbf{R_{yy}} = \mathbf{W}_r \mathbf{A}[\boldsymbol{\rho}\boldsymbol{\rho}^H]\mathbf{A}^H \mathbf{W}_r^H s^2(t) + \sigma^2 \mathbf{I}. \quad (4)$$

Although $\mathbf{W}_r$ is multiplied with $\mathbf{y}_0$, the standard MUSIC algorithm [9] can be extended to the phased-array by us

$$Peak(\theta) = \frac{1}{\bar{\mathbf{a}}^H(\theta)\mathbf{E}_n\mathbf{E}_n^H\bar{\mathbf{a}}(\theta)}. \quad (5)$$

where $\bar{\mathbf{a}}(\theta) = \mathbf{W}_r\mathbf{a}(\theta)$, and $\mathbf{E}_n$ is the noise subspace.

When the number of antenna elements N is large, P-MUSIC requires a longer T for AoA estimation. The tradeoff between accurate estimation and switching cost of phased-shifter codebook needs to be considered. Fortunately, for an array of eight antenna elements, the smallest beamwidth is 22.5 degrees. The limited beamforming resolution facilitates us to use less antennas for AoA estimation. Figure 7 illustrates the accuracy of blind AoA estimation using P-MUSIC. We compare P-MUSIC using different virtual RF chains with that using real RF chains. The AoA peak in P-MUSIC is close to that of real RF chains with eight antenna elements.

C. Blind Received Power Estimation

We proceed to estimate the received power so as to infer the channel gain between the AP and the user. Once the AoAs are obtained, the average received power of each path on the corresponding AoA direction can be estimated subsequently. Denote by α_p the average received power on the p^{th} path. There has

$$\begin{aligned}\alpha_p &= E[|\hat{\rho}_p(t)|^2] \\ &= (\bar{\mathbf{a}}^H(\hat{\theta}_p)\bar{\mathbf{a}}(\hat{\theta}_p))^{-1}\bar{\mathbf{a}}^H(\hat{\theta}_p)\mathbf{R}_{yy}(\bar{\mathbf{a}}(\hat{\theta}_p)\bar{\mathbf{a}}^H(\hat{\theta}_p))^{-1}\bar{\mathbf{a}}(\hat{\theta}_p) \quad (6)\end{aligned}$$

The received power estimated in Eq.(6) is a mixture of the actual received power, denoted by β_p, and the AWGN noise σ^2. We are more interesting in the β_p that reflects the real channel condition on the p^{th} dominant path. Here, σ^2 can be estimated from a long term statistics, and is irrelevant to the paths. This means that we can utilize all kinds of uplink frames to estimate σ^2.

Especially, we find that the 802.11 legacy preamble consists of a legacy short training field (L-STF) and a legacy long training field (L-LTF) modulated with PSK. This means that that the amplitude does not change in the modulation. Therefore, we employ a blind stochastic estimator, namely $M2M4$, for SNR estimation. The SNR can be approximately estimated using the second and fourth order moments of signal strength

$$\text{SNR}_{M2M4} = \frac{\sqrt{2M_2^2 - M_4}}{M_2 - \sqrt{2M_2^2 - M_4}}$$

where $M_2 \approx \sum_{n=1}^{N_s} |x_n|^2/N_s$ and $M_4 \approx \sum_1^{N_s} |x_n|^4/N_s$. Here, N_s represents the size of observation window for SNR estimation, and x_n is the sample of observation window. Authors in [23] demonstrated that while N_s was larger, the SNR estimation became more accurate. When the SNR has been estimated, and the received total power is P_r, σ^2 can be obtained via $\mathbb{E}[P_r/(\text{SNR}_{M2M4} + 1)]$. Here, inaccurate noise power estimation will cause the imperfect estimation of effective received power. To verify the accuracy of our noise power estimation, we compare the proposed method with an oracle data-aided counterpart (preamble-based SNR estimator) [23]. Figure 8 illustrates the CDF of noise power estimation

error. One can see that 95 percent of noise power estimation has an error below 10%. Hence, the effective received power denoted $\beta_{k,p}$ is given by $\beta_{k,p} = \alpha_{k,p} - \sigma^2$.

V. Joint Beamforming and User Selection

In this section, we model the assignment of the RF chains and beams as an integer programming problem, and describe the greedy beam-user selection algorithm in BUSH.

A. Problem Formulation

The purpose of beam-user selection is to assign RF chains and beams to proper users so that the total network throughput is maximized in a scheduling round. Let $\mathcal{R}$, $\mathcal{B}$ and $\mathcal{K}$ denote the sets of RF chains, beam patterns and users respectively. Denote by $U_{r,b}^{(k)}$ the throughput of user k using the r^{th} RF chain and the b^{th} analog beam. Formally, there has $U_{r,b}^{(k)} = \log(1 + \text{SINR}_{r,b}^{(k)})$ where the signal to interference and noise ratio (SINR) is given by

$$\text{SINR}_{r,b}^{(k)} = \frac{\beta_{r,b}^{(k)}}{\sum_{i\in\mathcal{R}}\sum_{j\in\mathcal{B}}\sum_{z\in\mathcal{K}\setminus\{k\}} x_{i,j,z}\gamma_{i,j}^{(b)} + \sigma^2} \tag{7}$$

under the normalized transmission power. Here, $\beta_{r,b}^{(k)}$ is the effective received power (Sec. IV-C) from the r^{th} RF chain to user k on the b^{th} beam. In the denominator, σ^2 denotes the Gaussian noise. $\gamma_{i,j}^{(b)}$ represents the cross-talk interference on the b^{th} beam caused by the concurrent transmission using the i^{th} RF chain and the j^{th} beam, and the summation refers to the total interference.

Instead of randomly selecting RF chains and beams, or fixing the beam-user pairs all the time, we formulate the beam-user selection as a nonlinear inter programming problem ($\mathbf{P}$):

$$\max_{\mathbf{x}} \quad \sum_{r=1}^{|\mathcal{R}|}\sum_{b=1}^{|\mathcal{B}|}\sum_{k=1}^{|\mathcal{K}|} x_{r,b,k}U_{r,b}^{(k)} \tag{8}$$

$$\text{s.t} \quad \sum_{b=1}^{|\mathcal{B}|}\sum_{k=1}^{|\mathcal{K}|} x_{r,b,k} \le 1, \quad \forall r \in \mathcal{R}, \tag{9}$$

$$x_{r,b,k} \in \{0,1\}. \tag{10}$$

The binary variable $x_{r,b,k}$ indicates whether the r^{th} RF chain and the b^{th} beam are assigned to user k or not. Our objective is to maximize the total throughput, given the finite numbers of RF chains, beam patterns and users. Eq.(9) means that each RF chain can be assigned to at most one user, and each RF chain uses at most one analog beam. It is worth noting that the optimization problem ($\mathbf{P}$) does not restrain one user from using more than one RF chains, and does not restrict multiple RF chains to the same direction in theory. The coupling of different beam-user mappings are captured by the SINRs in a complicated way under realistic indoor environments.

Finding the optimal beam-user selection strategy entails a prohibitive computational complexity. Although the numbers of RF chains, beams and users are finite, an exhaustive search may take K^{RB} operations, which cannot be achieved in real-time. In what follows, we present a greedy beam-user selection algorithm with provable performance guarantee.

B. Algorithm Design

We adopt a greedy algorithm to perform the joint RF chain and beam assignment. The basic idea is to select a RF chain-beam-user mapping from the resource pool that leads to the largest throughput increment in each time. Our greedy strategy is specified as Algorithm 1. It starts from an empty set of RF chain-beam-user mapping, and selects the RF chain with the largest network capacity increment, i.e., $f(\mathcal{R}\cup\{r\}, \mathcal{B}\cup\{b\}, \mathcal{K}\cup\{k\}) - f(\mathcal{R},\mathcal{B},\mathcal{K})$ at line 3, in each loop from line 2 to 9. The function $f(\cdot,\cdot,\cdot)$ is calculated as Algorithm 2. Algorithms 1 and 2 run in $O(|\mathcal{R}|^3|\mathcal{K}||\mathcal{B}|)$ time.

Algorithm 1: (RF Chain-) Beam-User Selection

Input: β, γ
Output: $\mathbf{x}$

1: $\tilde{\mathcal{B}} \leftarrow \emptyset$, $\tilde{\mathcal{R}} \leftarrow \emptyset$, $\tilde{\mathcal{K}} \leftarrow \emptyset$, initialize $\mathbf{x}$
2: **while** true **do**
3: $r, b, k \leftarrow \arg\max_{r\in\mathcal{R}\setminus\tilde{\mathcal{R}}, b\in\mathcal{B}, k\in\mathcal{K}} f(\tilde{\mathcal{R}}\cup\{r\}, \tilde{\mathcal{B}}\cup\{b\}, \tilde{\mathcal{K}}\cup\{k\}) - f(\tilde{\mathcal{R}},\tilde{\mathcal{B}},\tilde{\mathcal{K}})$
4: **if** $f(\tilde{\mathcal{R}}\cup\{r\}, \tilde{\mathcal{B}}\cup\{b\}, \tilde{\mathcal{K}}\cup\{k\}) - f(\tilde{\mathcal{R}},\tilde{\mathcal{B}},\tilde{\mathcal{K}}) \le 0$ **then**
5: break
6: **else**
7: $\tilde{\mathcal{R}} \leftarrow \tilde{\mathcal{R}}\cup\{r\}$, $\tilde{\mathcal{B}} \leftarrow \tilde{\mathcal{B}}\cup\{b\}$, $\tilde{\mathcal{K}} \leftarrow \tilde{\mathcal{K}}\cup\{k\}$
8: **end if**
9: **end while**
10: **for** $i \leftarrow 1$ **to** $|\mathcal{B}|$ **do**
11: $x_{\tilde{\mathcal{R}}_i,\tilde{\mathcal{B}}_i,\tilde{\mathcal{K}}_i} \leftarrow 1$
12: **end for**

Algorithm 2: $f(\cdot,\cdot,\cdot)$ Calculation

Input: $\tilde{\mathcal{R}}$, $\tilde{\mathcal{B}}$, $\tilde{\mathcal{K}}$
Output: total capacity c

1: initialize $\mathbf{x}$
2: **for** $i =\leftarrow 1$ to $|\tilde{\mathcal{R}}|$ **do**
3: $x_{\tilde{\mathcal{R}}_i,\tilde{\mathcal{B}}_i,\tilde{\mathcal{K}}_i} \leftarrow 1$
4: **end for**
5: $c \leftarrow \sum_{i=1}^{|\mathcal{R}|}\sum_{j=1}^{|\mathcal{B}|}\sum_{k=1}^{|\mathcal{K}|}(x_{i,j,k}U_{i,j}^{(k)})$

In consideration of hardness of the joint RF-chain, beam and user selection problem, it is hard to find an approximation for the entire algorithm. But we can establish a performance guarantee when the RF chains and their corresponding beam directions are given, i.e., the user selection part in our system. Theorem 1 shows the performance bound with the proof in the appendix.

***Theorem* 1.** (Lower Bound) *The performance of Algorithm 1 can achieve at least a 1/2-approximation of the global optimum when the RF chain-beam configurations are fixed.*

VI. Evaluation

A. Implementation and Experimental Setup

We implement BUSH on the software defined radio platform WARPv3 [22]. Four WARPv3 boards with 16 antennas are

used as AP, and three 4-antenna WARPv3 boards act as 12 users (each antenna as a user). Phased-array antennas are also implemented on the AP according to [10] except for the phase calibration. Thanks to the imparity of each antenna, and hardware, phase offsets are intrinsic in an antenna array. Albeit the phase offsets are different on each antenna, luckily, they remain constant once the radios are powered on. Thus, we only need to calibrate the phase offset at the beginning. The authors in [10] use wireless phase calibration every few hours, but that method may interrupt normal communication. We prefer that local wired calibration using SMA splitters and cables [9]. However, [9] only calibrates the receiving RF chains for AoA estimation. We still need to calibrate transmitting RF chains for beamforming. We reverse the calibration way in [9] to calculate the initial transmitting phase offsets of antennas.

The physical layer of BUSH follows 802.11ac specification including the OFDM PHY with 64 sub-carriers, and all kinds of modulations. We adopt ZFBF for digital beamforming.

In our experiments, generally, the AP equips with two RF chains which each RF chain has eight antenna elements, and each user has one omnidirectional antenna. Hence the AP has two degree of freedom served two users concurrently. All the experiments are configured as 20MHz bandwidth and modulation and coding scheme (MCS) table [6].

B. Micro-Benchmarks Evaluation

1) P-MUSIC AoA Estimation: We use a single RF chain with eight antenna elements to evaluate P-MUSIC in BUSH for 35 different topologies including line-of-sight (LOS) and non-line-of-sight (NLOS) channels as illustrated in Figure 11. In our experiments, the path of the strongest incoming is used to study the error incurred due to the composite CSIs. The errors are calculated by comparing P-MUSIC with MUSIC using conventional antennas [8].

Figure 12 illustrates the estimation errors of the strongest path in all topologies. We vary different virtual RF chains to study their errors. For four virtual RF chains, the average, and median errors are 7.104 and 5 degree, respectively. Moreover, the narrowest beam width is 22.5 degree in the ULA array with eight antennas. In this case, we can use four virtual RF chains for beam selection with a marginal performance loss. When the antenna elements of phased-array increases, we can see that six virtual RF chains with three median error is enough for finer beam selection.

2) AoA Errors Impact on Throughput: We also study the impact of AoA errors on BUSH, and the experiment setup is the same as previous section. We use the conjugate beamforming [4] as the performance upper bound of analog beamforming. Note that conjugate beamforming can not be applied to phased-array antennas due to its hardware constraints. We use the AP with a single RF chain and phased-array antenna to transmit packets to one user at each transmission slot, and record the throughput. In Figure 13, we find that the average throughput of four virtual RF chain is only 14.5% less compared with eight virtual RF chains. These results verify the claim that four virtual RF chains for beam selection only causes a marginal performance loss in Section IV.

3) Overhead of PAS-Orthogonal Probing: In this micro-benchmark, we compare the PAS-orthogonal probing and that in the state-of-the-art 802.11ac protocol. In order to isolate the impact of PHY on fair comparison, we bypass the PHY interface and emulate the overhead of PAS probing in MAC layer. In this emulation, we consider that six virtual RF chains of each RF chain are applied, and one RF chain serves one user. From Figure 14, when the number of users increases, the overhead of PAS-orthogonal probing can be negligible. The reason is that our system serves the selected user according to PAS firstly, and then, sounding the served users as vanilla 802.11ac systems. When there are only two users, the extra overhead is the duration of uplink probing, which is higher in BUSH. However, the probing overheads with a few users are very gentle, and a large number of users share the wireless spectrum in reality.

C. System Evaluation

In this section, we evaluate our system in 35 different network topologies. Hence, each data point of our result is averaged over multiple topologies. Albeit Hekaton [10] is designed for LTE networks, we implement and modify it as our reference scheme. Note that we only replace the part of composite CSI measurement depended on LTE sounding reference signals by ours. In addition, we also compare our system with the state-of-the-art 802.11ac MU-MIMO system with two omnidirectional antennas. Moreover, in our experiments, we create a queue for all the users, and remove the user after service in a scheduling round. Hence, the fairness control has been employed in our systems. We use three WARPv3 boards with total ten antennas as the users, and consider 35 different topologies to test our system.

System throughput. We first study the case with two users and two RF chains. Different from Hekaton, a flexible beam and user selection algorithm is designed. It means that our system is not a one-to-one mapping solution such as Hekaton. In our system, two beams can even serve one user if this yields a larger throughput than serving two different users. From Figure 15, we can see that our system achieves an average 8% improvement than Hekaton, because in extremely low SNR cases, BUSH will rule out a user for the optimal total throughput in each transmission, but Hekaton will always serve two users in any cases. Hence, in this way, BUSH is very helpful for cell edge users with extremely low SNRs.

Next, we examine the scalability of multiple users. We increase the total users from two to ten, and record the sum throughput of the whole network. BUSH always outperforms Hekaton and vanilla 802.11ac with different number of users. Figure 17 plots the comparison of average total throughput under three systems. In addition, the CDFs of throughput in Figure 16 illustrates that BUSH achieves an average 22.14% improvement over Hekaton and 108.6% improvement over the vanilla 802.11ac. Compared with the other two schemes, these

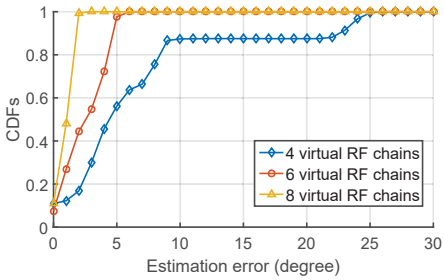

Fig. 12. Accuracy of P-MUSIC estimation

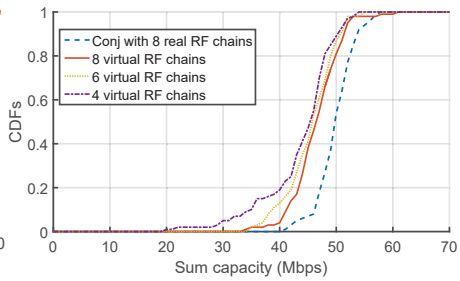

Fig. 13. Data rate of PAS-based beamforming

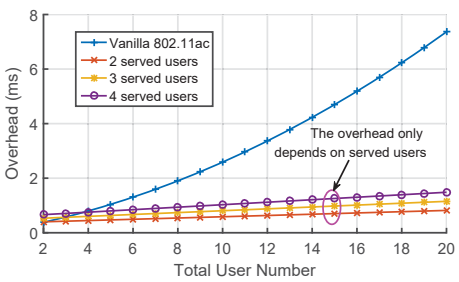

Fig. 14. The overhead of PAS-Orthogonal probing

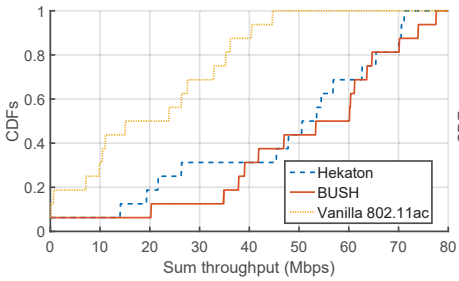

Fig. 15. The CDFs of two RF chains and two users

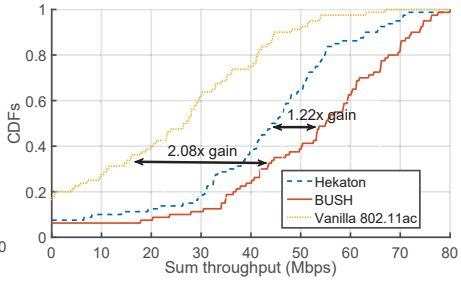

Fig. 16. The CDFs of throughput comparison

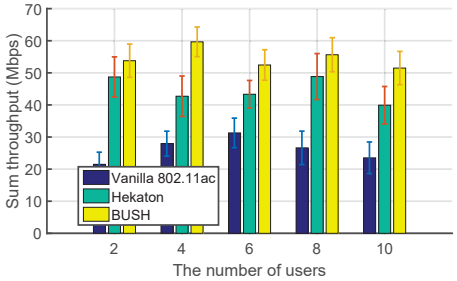

Fig. 17. The sum throughput of three systems

improvements mainly stem from the joint user-beam selection with hybrid beamforming.

VII. Related Work

1) Large-Scale MU-MIMO System: Argos [4] is the first reported base station architecture that can serve multiple users simultaneously through multi-user beamforming with a large number of antennas. Equipped with 64 antennas, Argos achieves a ten times higher capacity compared with single antenna systems. However, the complexity of computing digital beamforming grows quickly so that real-time application of Argos becomes less feasible in very large-scale MIMO systems. In view of this challenge, Bigstation [3] presented a parallel computing architecture that enables real-time signal processing in MIMO systems which may have tens or hundreds of antennas. Despite of its excellent performance, Bigstation demands a powerful server, which is not suitable for current WiFi APs that their chipset processing capability is much lower. As a special case with single user but multiple streams, Chen *et al.* designed an energy efficient system to deliver 3D multiview streaming with guaranteed quality [25]. A related 802.11 energy efficient rate adaptation algorithm can be found in [26].

2) Analog beamforming: In [18], phase-shifter was first adopted for analog beamforming in MIMO systems. Probeam [14] used phased-array antennas to design downlink scheduling and user association algorithms for multiple cells in a non-MIMO WiMAX network. The authors in [7] designed a codebook for analog beamforming in millimeter-wave (mmWave) WPAN networks. A common feature of above work is that analog beamforming needs to search all the codebook entries so that the coordination overhead increases with the number of antenna elements in phased-array antennas.

3) Hybrid Beamforming: Hybrid beamforming consists of analog beamforming that uses phase shifters and digital beamforming that uses RF chains for downlink transmission. Authors in [17] and [20] designed hybrid precoding schemes for single user and multiple users respectively based on compressive sensing in mmWave communications. However, existing work mainly focused on the theoretic understanding of the gap between fully digital and hybrid beamforming when the numbers of RF chains and data streams vary. Hekaton [10] was the latest hybrid beamforming architecture for current LTE systems. To enable large-scale MIMO, Hekaton introduced a compressive AoA estimation algorithm for phased-array antennas and a downlink beam selection metric known as *signal-to-leakage-power-ratio (SLR)*. However, Hekaton needed reference signals for AoA estimation. In addition, Hekaton was designed specifically for LTE systems, and may not be an ideal solution for large-scale MIMO in WLANs.

VIII. Conclusion

This paper tackles an important issue in 802.11 WLANs: enabling large-scale MU-MIMO with a finite number of RF chains. We design BUSH, a unified and integrated system that employs a hybrid digital and analog beamforming with phased-array antennas to achieve both beamforming gains and multiplexing gains. BUSH adopts a low-complexity probing scheme to measure composite CSIs from the array of antenna elements and an accurate yet blind PAS estimation algorithm. With the estimated AoAs and signal strength, BUSH presents an efficient beam-user selection algorithm that greatly reduces the channel correlation and cross-talk interference. We implement BUSH on the WARPv3 software radio platform. Extensive experiments demonstrate that when total throughput serves as a metric, BUSH achieves a gain of 108% over 802.11ac and a gain of 22% over other counterparts on average.

REFERENCES

[1] http://www.idtechex.com/research/reports/wearable-technology-2016-2026-000483.asp

[2] Larsson, Erik G., et al. "Massive MIMO for next generation wireless systems," IEEE Commun. Mag. 2014

[3] Q. Yang, X. Li, H. Yao, J. Fang, K. Tan, W. Hu, J. Zhang, and Y. Zhang, "BigStation: enabling scalable real-time signal processingin large mu-mimo systems," Proc. of ACM SIGCOMM 2013.

[4] C. Shepard, H. Yu, N. Anand, E. Li, T. Marzetta, R. Yang, and L. Zhong, "Argos: practical many-antenna base stations," Proc. of ACM Mobicom 2012.

[5] J. Xiong, and K. Jamieson, "SecureAngle: improving wireless security using angle-of-arrival information," Proc. of ACM Hotnets 2010.

[6] Matthew Gast. 2013. 802.11ac: A Survival Guide (1st ed.). O'Reilly Media, Inc..

[7] J. Wang, Z. Lan, C. Pyo, T. Baykas, C. Sum, M. A. Rahman, J. Gao, R. Funada, F. Kojima, H. Harada, and S. Kato, 2009. "Beam codebook based beamforming protocol for multi-Gbps millimeter-wave WPAN systems," IEEE J. Select. Areas Commun. 2009.

[8] J. Xiong and K. Jamieson, "Towards fine-grained radio-based indoor location," Proc. of ACM HotMobile 2012.

[9] J. Xiong, and K. Jamieson, "ArrayTrack: a fine-grained indoor location system," Proc. of USENIX NSDI 2013.

[10] X. Xie, E. Chai, X. Zhang, K. Sundaresan, A. Khojastepour, and S. Rangarajan, "Hekaton: Efficient and Practical Large-Scale MIMO," Proc. of ACM MobiCom 2015.

[11] X. Xie, X. Zhang, and E. Chai, "Cross-Cell DoF Distribution: Combating Channel Hardening Effect in Multi-Cell MU-MIMO Networks," Proc. of ACM MobiHoc 2015.

[12] J. Fang et al., "Fine-Grained Channel Access in Wireless LAN," IEEE/ACM Trans. Netw., 2013.

[13] "IEEE Draft Standard for IT - Telecommunications and Information Exchange Between Systems - LAN/MAN - Specific Requirements - Part 11: Wireless LAN Medium Access Control and Physical Layer Specifications - Amd 4: Enhancements for Very High Throughput for operation in bands below 6GHz," IEEE P802.11ac/D3.0, June 2012.

[14] J. Yoon, K. Sundaresan, M. A. Khojastepour, S. Rangarajan, and S. Banerjee, "ProBeam: a practical multicell beamforming system for OFDMA small-cell networks," Proc. of ACM MobiHoc 2013.

[15] F. Comtech, "Fci-3740 phased array antenna", http://www.fidelity-comtech.com/products/phased-arrayantennas.

[16] J. Xiong and K. Jamieson, "SecureArray: improving wifi security with fine-grained physical-layer information," Proc. of ACM MobiCom 2013.

[17] O. E. Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi and R. W. Heath, "Spatially Sparse Precoding in Millimeter Wave MIMO Systems," IEEE Wireless Commun., 2014.

[18] V. Venkateswaran and A. J. van der Veen, "Analog Beamforming in MIMO Communications With Phase Shift Networks and Online Channel Estimation," IEEE Trans. Signal Processing, 2010.

[19] AR9331 - Highly-Integrated and Cost Effective IEEE 802.11n 1x1 2.4 GHz SoC for AP and Router Platforms, in Atheros Data Sheet, 2010.

[20] A. Alkhateeb, G. Leus and R. W. Heath, "Limited Feedback Hybrid Precoding for Multi-User Millimeter Wave Systems," IEEE Trans. Wireless Commun., 2015.

[21] J. Xiong, K. Sundaresan, K. Jamieson, M. A. Khojastepour, and S. Rangarajan, "MIDAS: Empowering 802.11ac Networks with Multiple-Input Distributed Antenna Systems," Proc. of ACM CoNEXT 2014.

[22] WARP Project, http://warpproject.org

[23] D. R. Pauluzzi and N. C. Beaulieu, "A comparison of SNR estimation techniques for the AWGN channel," IEEE Trans. Commun., 2000.

[24] X. Xie and X. Zhang, "Scalable user selection for MU-MIMO networks," Proc. of IEEE INFOCOM 2014.

[25] Z. Chen, X. Zhang, Y. Xu, Y. Zhu and X. Wang, "POM: Power Efficient Multi-view Video Streaming over Multi-antenna Wireless System", Proc. of IEEE ICME 2016.

[26] Y. Xu, C. S. Lui, and D. M. Chiu, "Improving energy efficiency via probabilistic rate combination in 802.11 multi-rate wireless networks", Ad Hod Networks 2009.

[27] M. Fisher, G. Nemhauser and G.Wolsey, "An Analysis of Approximations for Maximizing Submodular set Functions-II", Mathematical Programming Study, 1978.

APPENDIX: PROOF OF THEOREM 1

Proof. Before commencing the proof, we introduce two important concepts.

Matroid: A *matroid* is a pair $(V, \mathcal{I})$ such that V is a finite set, and $\mathcal{I} \subseteq 2^V$ is a collection of subsets of V satisfying the following two properties:

- $X \subseteq Y \subseteq V$ and $Y \subseteq \mathcal{I}$ implies $X \in \mathcal{I}$
- $X, Y \in \mathcal{I}$ and $|Y| > |X|$ imply $\exists e \in Y \setminus X$ such that $X \cup \{e\} \in \mathcal{I}$

where X and Y are two finite sets.

Sub-modularity: A function $g : 2^V \to \mathbb{R}$ is *sub-modular* if for every $X \subseteq Y \subseteq V$ and $e \in V \setminus Y$ it holds that

$$g(X \cup \{e\}) - g(X) \geq g(Y \cup \{e\}) - g(Y) \tag{11}$$

To facilitate our analyze, we define a RF chain and beam selection as $\mathbf{y}$, $(y_{ij} \in \{0, 1\})$, where $y_{ij} = 1$ indicates the i^{th} RF chain transmits data stream on the j^{th} beam. Certainly, we obtain

$$\sum\nolimits_{j=1}^{|\mathcal{B}|} y_{ij} \quad \leq \quad 1, \quad \forall i \in \mathcal{R}, \tag{12}$$

where Eq. (12) is the counterpart of Eq. (9). We also define the RF chain and beam interference as $\mathbf{I}$ for given $\mathbf{y}$, where $I_{ij} \in \mathbf{I}$ is the interference caused by $\mathbf{y}$ when the i^{th} RF chain works with the j^{th} beam. Thus, when user k receives data from RF chain i and its j^{th} beam, the utility can be calculated as

$$U_{i,j}^{(k)} \quad = \quad \log(\mathbf{SINR}_{i,j}^{(k)}) \approx \log(I_{i,j}\beta_{i,j}^{(k)}). \tag{13}$$

We define $\tilde{\mathcal{R}}$ as the set of RF chains that have been associated with users. Thus, in each iteration, our algorithm chooses the user $k = \arg\max_{k \in \mathcal{K}} U_{r,b}^{(k)}$ for $r \in \mathcal{R} \setminus \tilde{\mathcal{R}}$ and $b \in \mathcal{B}$. Define $\Phi = \{I_{i,j}\beta_{i,j}^{(k)} | i \in \mathcal{R}, j \in \mathcal{B}, k \in \mathcal{K}\}$ as the ground set and S as a collection of subsets of Φ. Also, for any $S_i \in S$, there must not be $I_{r,j_1}\beta_{r,j_1}^{(k_1)} \in S_i \wedge I_{r,j_2}\beta_{r,j_2}^{(k_2)} \in S_i$ $(j_1 \neq j_2 \vee k_1 \neq k_2)$ for any RF chain r, i.e., one RF chain can not be allocated to more than one users. Obviously, (Φ, S) is a matroid. Denote $g : S \to \mathbb{R}$ as the utility function

$$g(S_i) = \sum_{u_i \in S_i} u_i, \tag{14}$$

where u_i denotes a subset of a ground set. Suppose that X and Y are two intermediate beam-user selection sets that has $X \subseteq Y$. For an alternative beam-user mapping $e \in \Phi$, we obtain $g(X \cup \{e\}) - g(X) = g(\{e\}) = g(B \cup \{e\}) - g(B)$ according to our greedy algorithm used to associate the RF chains and beams. Thus, $g(\cdot)$ is sub-modular. Given the fixed RF chain-beam configurations, our greedy algorithm selects the best RF chain that contributes to the largest throughput increment in each iteration. According to [27], the sub-optimality of maximizing a sub-modular function over a partition matroid using our greedy algorithm was shown to be bounded by $\frac{1}{2}$. Thus, this theorem is established. □