Scaling Multi-User MIMO WLANs: the Case for Concurrent Uplink Control Messages

Oscar Bejarano*, Sadia Quadri*, Omer Gurewitz†, and Edward W. Knightly*
*ECE Department, Rice University, Houston, TX
{obejarano, sq4, knightly}@rice.edu
†CSE Department, Ben Gurion University, Beer Sheva, Israel
{gurewitz}@cse.bgu.ac.il

***Abstract*—Downlink Multi-User MIMO (MU-MIMO) enables the simultaneous spatial sharing of the channel by multiple users to achieve a capacity gain over Single-Input Single-Output (SISO) systems. Unfortunately, the overhead required to enable multi-user MIMO is much higher than the overhead required for single-stream systems. Namely, for K users, collection of channel state information requires K transmission exchanges (i.e., $O(K)$) between the AP and users. Likewise, the MU-MIMO acknowledgement process also requires the same amount of exchanges, thus reducing the performance gains attained via simultaneous downlink transmission. In this paper, we design, implement, and experimentally evaluate Concurrent Uplink Control Messages (CUiC) to scale the MU-MIMO control information exchange process and improve the efficiency of 802.11ac-based MU-MIMO networks. Our key technique is the design of new channel sounding and acknowledgement mechanisms that enable multiple users to transmit their reverse-direction *control messages* (i.e., beamforming reports and acknowledgments) concurrently to the AP, in $O(1)$ transmission slots. We implement CUiC and perform an extensive set of experiments and demonstrate throughput gains of more than 100% compared to 802.11ac.**

I. Introduction

Downlink Multi-User MIMO (MU-MIMO) enables capacity gains via simultaneous transmission from the access point (AP) to multiple users [1], [2]. Consequently, downlink MU-MIMO has been incorporated into the Wi-Fi standard via the IEEE 802.11ac amendment, which promises multi-Gb/sec data rates [3], [4]. Unfortunately, despite such high physical layer rates, a significant amount of airtime is devoted to providing the access point (transmitter) with the required Channel State Information (CSI or CSIT). Indeed, in 802.11ac, the number of required CSI feedback messages increases linearly with the number of users simultaneously served. Similarly, after every downlink MU-MIMO transmission, the number of acknowledgements (ACKs) and ACK-request exchanges between the AP and the served users also increases linearly with the number of users. We will show that the time-resources devoted to these exchanges severely reduce MU-MIMO throughput gains.

In this paper, we design, implement, and experimentally validate a protocol for concurrent uplink transmission of control messages from multiple users. Namely, we present CUiC (Concurrent Uplink Control) to replace the aforementioned *sequential* CSI feedback and ACK messaging system with a design that achieves *parallel* transmission. In particular, we present the following contributions.

First, we demonstrate how to scale uplink feedback by employing only a single transmission slot to simultaneously send the control messages from all users (independently of the number of users). In order for the AP to decode multiple concurrent transmissions, it needs to train its receiver to perform per-user channel estimation, trigger automatic gain control (AGC), and estimate and correct carrier frequency offsets as well as other timing offsets due to digital and RF mismatches between users and AP. Thus, to realize concurrent single-slot feedback, we demonstrate that preamble staggering of user-to-AP packets allows the AP to obtain "clean" (interference-free), per-user measurements to enable this training.

Unlike random access uplink MU-MIMO MAC protocols which require *(i)* additional user-AP frame exchanges to train the AP [5], [6], or *(ii)* special-purpose space-division multiple access (SDMA) techniques to counteract the effects of multi-stream signal overlap (on both preamble and data portions of the packet) [7], we design CUiC such that it emulates a single user (SU) MIMO beamforming system. Namely, in CUiC the AP pre-determines the timing required to align the transmissions from all concurrent users to within 800 ns, and acquires the training needed for stream demultiplexing without additional signaling or frame exchanges. Consequently, this enables the use of legacy signal detection and correction techniques by relying solely on packet preambles, and SU-MIMO spatial demultiplexing techniques [2]. Our mechanism only applies to uplink *control* messages in which the users are pre-selected (the set of users selected by the AP for downlink data is a subset of the users sending CSI to the AP) and timing can be tightly controlled by the AP (see Section IV for further discussion of related work). Furthermore, we demonstrate that as a desirable side effect, CUiC also increases robustness of the sounding process by reducing the time between CSIT estimation and downlink data transmission, which is critical in fast fading channels or in the presence of highly mobile users.

Second, we implement CUiC in an 802.11 OFDM MU-MIMO platform prototype based on WARP and WARPLab [8], and demonstrate the feasibility of decoding multiple simultaneous wide-band 802.11ac control messages in uplink MU-MIMO random access systems via preamble staggering

This research was supported by Cisco Systems, Intel, the Keck Foundation, and by NSF grants CNS-1444056, CNS-1126478 and CNS-1012831.

978-1-4673-7331-9/15/$31.00 © 2015 IEEE

and traditional spatial demultiplexing techniques. Moreover, we propose a frequency offset correction method comprised of a combination of pre-compensation and iterative correction. This allows the AP to apply the offset of a single user to the composite stream at each interference cancellation iteration while treating the rest of the signal components as noise. We demonstrate that in a system comprised of four simultaneously transmitting single-antenna users and a 4-antenna AP, CUiC can achieve a frame error rate of only 6% even at low SINR regimes, i.e., -5 dB.

Finally, we propose and evaluate a suite of policies that incorporate the sounding mechanisms of CUiC into a complete downlink MU-MIMO system. More specifically, these policies combine our concurrent uplink technique with user grouping strategies needed to enable downlink MU-MIMO. We design these policies such that they can be tailored to target user diversity maximization, sounding air-time minimization, or simply to collect feedback from as many users as the number of degrees of freedom (DoF) available at the transmitter (802.11ac-like). We evaluate CUiC via a combination of over-the-air, and channel emulator based experiments. Moreover, we compare against a benchmark based on the sequential feedback in 802.11ac. Our evaluation reveals that all CUiC policies outperform the benchmark for a wide range of Aggregate MAC Protocol Data Unit (A-MPDU) lengths. In particular, we demonstrate the importance of adjusting the system by selecting a different policy for different frame aggregation regimes. More importantly, we show that even when large frame aggregation is utilized to amortize overhead (e.g., 96 kB), our scheme can attain close to 40% throughput gains compared to the benchmark. Similarly, for short packets (e.g., 1.5 kB), CUiC can achieve more than 1.3x throughput gains.

II. Concurrent Uplink Control Messages

Today, uplink control messages in MU-MIMO WLANs are transmitted in the same way as other control frames, i.e., clients transmit feedback sequentially, one at a time. Consequently, both the channel state acquisition process (sounding) and the acknowledgment process yield a time-resource-intensive frame exchange between the AP and each individual user before and after a MU-MIMO transmission. In contrast, CUiC realizes *concurrent* uplink transmission of beamforming reports and acknowledgements from a set of users, thereby achieving the same control message acquisition goal in only a fraction of the time.

A. Receiver-Side Beamforming

To reduce the amount of time spent on sounding feedback and data acknowledgement acquisition CUiC implements a receiver-side beamforming strategy, i.e., decoding of spatially multiplexed streams. We dedicate this subsection to provide a brief overview of such technique.

Notation. Let M denote the number of antennas at the AP and N_k represent the number of antennas at user k ($k = 1, \cdots, K$), where K is the number of users served concurrently in a MU-MIMO transmission. Thus, K corresponds to the number of users transmitting concurrent messages to the AP. Although users can have more than one antenna, for simplicity we consider one data stream per user at any time. $\mathbf{H} \in \mathbb{C}^{K \times M}$ represents the channel gain matrix between the AP and all users served simultaneously. We use uppercase (lowercase) boldface to denote matrices (vectors). Thus, $\mathbf{h}_k \in \mathbb{C}^{1 \times M}$ represents the channel vector to user k. Also, $\mathbf{H}^*$ ($\mathbf{h}^*$) denotes the complex conjugate transpose of matrix $\mathbf{H}$ (vector $\mathbf{h}$).

Space-Division Multiple Access. In an SDMA transmission, the AP receives a linear combination of K signals on each antenna. Thus, if the AP receives M linear combinations, for $M = K$ unknown transmissions, each of these multiplexed streams can be estimated. To accomplish this, low-complexity linear schemes such as Minimum-Mean Square Error (MMSE), and Zero-Forcing (ZF) have been proposed [2], [11], [12].

With full CSI at the receiver, ZF can be used to completely suppress inter-stream interference. Figure 1(a) illustrates the operation of ZF. Notice that $\mathbf{h}_1 x_1$ is projected onto the space orthogonal to the interferer (null space). This projection is represented by $\mathbf{h}^*_{proj}\mathbf{h}_1 x_1$. The main drawback of ZF is that it leads to significant noise enhancement if the channel matrix $\mathbf{H}$ is ill-conditioned. Observe that if the angle between the two vectors is small, the projection becomes very small as well, which means that the resulting SINR is significantly decreased. In contrast to ZF, Figure 1(b) illustrates that MMSE does not suppress the interference completely but provides the *optimal compromise* between interference suppression and maximizing the signal strength from the intended user [2].

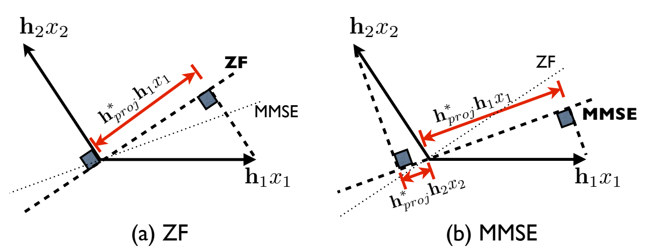


Fig. 1: Geometric representation of ZF and MMSE. In this 2x2 system, stream 1 with channel $\mathbf{h}_1$ is the intended stream (stream to decode), whereas stream 2 with channel $\mathbf{h}_2$ is the interferer.

MMSE minimizes the expected value of the mean square error between the transmitted signal vector $\mathbf{x}$ and the received vector $\mathbf{W}^*\mathbf{y}$. $\mathbf{y}$ is the linear combination of all received signals and $\mathbf{W}^*$ is given by $\mathbf{W}^* = (\mathbf{H}^*\mathbf{H} + \sigma^2\mathbf{I})^{-1}\mathbf{H}^*$ where σ^2 represents the noise variance. At high SNR, the term $\sigma^2\mathbf{I}$ becomes negligible, therefore MMSE achieves the asymptotic performance of ZF [12]. Additionally, a combination of MMSE and a non-linear decoding technique termed Successive Interference Cancellation (SIC) achieves the capacity of the MIMO channel when fading is *i.i.d.* Rayleigh [2]. SIC consists of an iterative receiver, that is, it decodes one stream at a time. More specifically, when decoding a stream, it considers the remaining signal as interference. Then, after decoding a particular stream, its contribution to the overall signal is reconstructed and removed from it. This procedure continues until all streams are decoded.

B. Enabling Spatially Multiplexed Feedback

Decoding multiple concurrent streams requires the receiver to accurately measure the CSI from each stream. Therefore, the individual signal components needed for such estimation (packet preambles) need to be interference-free.

Feedback Timing Structure. CUiC introduces an explicit channel estimation feedback technique that allows the multiple users sounded to reply with their channel information simultaneously. In addition, it also introduces an acknowledgement process where users reply simultaneously with their corresponding block acknowledgements (BAs) after a downlink MU-MIMO transmission. Achieving this requires the AP to detect and decode several signals with no inherent orthogonality among them in time, or code. Consequently, we design a scheme that implements a multiplexed signal decoder and modifies the structure of the feedback and ACK processes in 802.11ac to coordinate transmissions from all users.

To perform MMSE-SIC decoding on the uplink, the AP requires channel information and timing offset estimates, e.g, frequency offsets, from all concurrently transmitting users. CUiC exploits the fact that all packet transmissions require a preamble for signal detection, channel estimation, etc., and modifies the preambles of the CSI feedback reports and block ACKs to provide the AP with the information necessary to decode the composite signal. However, to obtain an accurate representation of the channel to each user and of the different RF mismatches, the preambles need to be *clean*, that is, without interference from other streams. For that reason, CUiC introduces a preamble time-based staggering technique that prevents these preambles from overlapping with one another.

Structure of the Sounding and Acknowledgement Processes in 802.11ac. 802.11ac defines a two-step explicit sounding process to support the downlink MU-MIMO data transmission [3], [4]: *(i) User sounding*, in which the AP broadcasts one training pilot signal from each of its transmit antennas; upon reception of the pilots, the users calculate their channel vector from each AP antenna. *(ii) Beamforming report feedback*, in which the AP polls each user in order to retrieve their reports containing the computed channel estimates. Figure 2 depicts the 802.11ac sounding timeline (not to scale). First, a Null Data Packet Announcement (NDPA) is sent to indicate which users are required to prepare beamforming reports. Second, the AP sends a Null Data Packet (NDP) containing the pilots required by the users to estimate their channels to the AP. Third, all users reply with their reports.

The MU-MIMO acknowledgment process (also shown in Figure 2) has a similar structure to the sounding process. In general, it consists of an exchange of Block ACK Requests (BAR) and Block ACKs (BA) between the AP and the different users. Notice that in 802.11ac all MPDUs are aggregate MPDUs or A-MPDUs. Therefore, all replies take the form of block ACKs. After an MU-MIMO downlink data transmission, one of the users replies immediately with its corresponding BA, whereas the rest of them are polled for their BAs via a BAR sent by the AP. To determine which user has to reply first, the ACK policy for the data MPDU of one specific user is set to implicit BA.

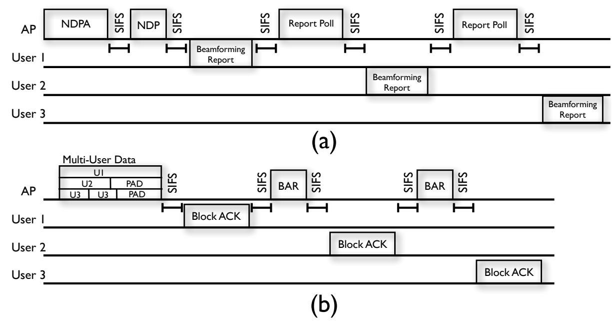


Fig. 2: 802.11ac Timeline: (Top) Sounding and feedback process and (Bottom) acknowledgement process.

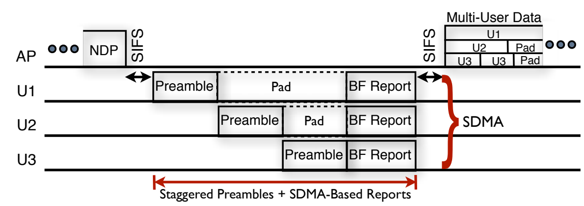


Fig. 3: CUiC Staggered Feedback Format (not to scale).

Scalable Sounding and Acknowledgement Structure. CUiC leverages the basic sounding and acknowledgement structures in 802.11ac to align all transmissions such that the preambles from all users are time-staggered. That is, for the case of the sounding process, upon reception of the NDP all users that were addressed in the NDPA packet compute their channel vectors to the AP. Then, based on the order specified in the NDPA, they align their preambles and beamforming reports. Figure 3 shows the proposed alignment for the case of three different users. By zero-padding between preambles and the payload carrying the beamforming reports, we can align the reports from all users. This allows the AP to receive each preamble corresponding to a different stream in a clean manner which in turn allows it to estimate the carrier frequency offset, symbol timing, and channel coefficients necessary to decode the different frames. Likewise, for concurrent ACKs, the end of the downlink MU-MIMO A-MPDU serves as the trigger for the multiple users to reply simultaneously. This leads to the suppression of all BARs and inter-frame spacings required to coordinate channel access between successive replies.

Synchronization for Spatial Demultiplexing. In contrast to single-user MIMO architectures in which all transmitted streams originate from the same device with RF interfaces sharing a common clock and a fully synchronized transmission trigger, in uplink MU-MIMO (SDMA), transmissions from different users do not have these characteristics. Three major synchronization challenges arise when attempting to decode a composite signal [13]: First, a combination of Doppler shifts and discrepancies between the oscillators used to generate the carrier frequency at the AP and users can lead to several different *carrier frequency offsets (CFOs)* thus causing irreducible inter-carrier interference (ICI). Second, *symbol synchronization* is necessary to align the signal reception from the multiple users with the observation window at the AP.

That is, if the symbols transmitted from all the individual users do not fall within the timing window of the AP, inter-symbol interference (ISI) and ICI can be introduced. Third, the *sampling clock frequency* is obtained from a local oscillator, consequently, any mismatch between the oscillators of the different devices can lead to a misalignment of the digital sampling clock frequency of all users with respect to that of the AP. Therefore to prevent these issues, the AP or users need to compensate for these offsets.

Although CFO estimation is greatly simplified by using clean preambles, compensation and correction becomes more difficult since the received signal consists of a linear combination of multiple independent streams, where each can have a different offset. CUiC relies on a combination of pre-compensation [13] and residual CFO correction at the AP. That is, in CUiC, each user utilizes both the NDPA and NDP to obtain an accurate CFO estimate and then applies a pre-compensation factor on the next uplink transmission. Nonetheless, all residual CFO components that remain after pre-compensation need to be removed by the AP. CUiC introduces a method for correcting such remaining offset from each stream. In general, the technique treats the signal components of all users except the one currently being decoded, as noise, and iteratively removes each component. More specifically, after CFO estimation for all streams, the composite signal is passed through the MMSE-SIC receiver. Once the AP knows which stream to decode first, it applies CFO correction for that specific stream to the entire compound signal. Next, the MMSE-SIC algorithm decodes this first stream, removes the applied CFO correction, and removes the decoded signal component from the original compound signal. The process is repeated until all streams are decoded. This processes can be represented as follows: Let $x_i(n)$ be the time domain symbol n transmitted by the ith user, and $h_i(n)$ be the impulse response of the ith channel (for each AP antenna). Thus, after passing the signal of the ith user through the channel we obtain $s_i(n) = x_i(n) * h_i(n)$. Ignoring noise terms, the composite received baseband signal is given by

$$r(n) = \sum_{i=1}^{K} s_i(n) e^{j2\pi\Delta f_i n}$$

where K is the number of users transmitting simultaneously to the AP, and Δf_i denotes the ith user's CFO normalized by symbol period. Assuming that the AP attempts to decode the signal from user 1 first, it will correct the CFO in time domain by multiplying $r(n)$ with the term $e^{-j2\pi\Delta f_1 n}$. Thus, in the first MMSE-SIC iteration the compound signal obtained after correcting the CFO of the first stream is given by

$$\begin{aligned} y(n) &= r(n) e^{-j2\pi\Delta f_1 n} \\ &= s_1(n)(1) + s_2(n) e^{j2\pi n(\Delta f_2 - \Delta f_1)} + \cdots \\ &+ s_K(n) e^{j2\pi n(\Delta f_K - \Delta f_1)} \end{aligned}$$

Notice that the MMSE-SIC receiver can now start decoding the component of the first stream. After the first stream has been decoded, the compound signal is multiplied by $e^{j2\pi\Delta f_1 n}$ to remove the CFO component from this first stream. This yields a compound signal comprised of all original streams except the first one, where each of them can be decoded in the same way. In our system, this method allowed us to remove an offset of up to several kHz.

Pre-synchronization by each user has also been shown to be effective for symbol synchronization and sampling clock frequency misalignment correction [13]. Moreover, in terms of symbol synchronization the use of the OFDM cyclic prefix allows the AP to successfully decode frames for offsets that are less than 0.8 μs (length of the cyclic prefix). Nevertheless, the implementation of longer prefixes has been shown to enhance robustness [14]. Finally, in addition to pre-synchronization, the correction of the residual sampling clock frequency misalignment can be achieved by performing oversampling, i.e., via a fractional spaced equalizer (FSE) [7].

Avoiding Signal Saturation via Power Control. Although SIC is known to perform better when the multiple streams have different magnitudes, if the difference between these concurrent signals is not within the dynamic range of the Analog-to-Digital Converter (ADC), the AGC can erroneously tune the system therefore leading to frame losses, e.g., due to ADC saturation. CUiC allows the AP to do AGC at each individual preamble, and set the RF and baseband gains accordingly. That is, instead of relying on the gains set for the first arriving stream, the AP can tune the gains to the value that maximizes the likelihood of successful reception of all streams. While the development of an *adaptive* AGC and power control methods is out of the scope of this paper, we implement a simple AGC scheme in which we tune the system at each preamble and re-tune again during the arrival of the concurrent signals. Nonetheless, we expect that implementing an adaptive AGC would further increase the performance of the system. In addition, power control techniques in prior work [15] can be incorporated into our scheme.

C. User Selection and Retransmissions

User selection in downlink MU-MIMO is tightly coupled to the sounding process. Selecting the best subset of users to serve simultaneously can occur before [9] or during [10] sounding. In the context of CUiC, we propose a set of policies for combining sounding and user selection, each with a different objective: *(i) Basic operation (802.11ac-like):* The AP collects CSIT from as many users as it has transmit antennas (maximum DoF). Regardless of the user selection algorithm employed prior to sounding, the AP collects M beamforming reports and then serves the subset of those M users that maximize the aggregate rate (two-round user selection). *(ii) Maximize user diversity (mDiv):* Increased user diversity leads to increased rate performance of downlink MU-MIMO [16], [17]. With more information about the channels to different users, the AP can select combinations of them that satisfy a certain rate or fairness criteria. CUiC increases the number of CSIT reports provided at each feedback slot by M-fold, thus giving the AP M times more information compared to the single-user per feedback-slot case. If we assume that the AP can spend the same amount of time performing sounding as 802.11ac, then it can potentially acquire CSIT from M^2

users (vs. M in 802.11ac). In case there are fewer than M^2 associated users, this scheme could lead to maximum achievable rate. *(iii) Minimize sounding air-time utilization (mSo):* In this policy, only one feedback slot is allowed regardless of the number of successfully decoded (concurrent) streams. Similarly to the basic operation, the AP can implement a two-round user selection. Notice that retransmissions are allowed in the first two policies in case the AP fails to decode at least one stream. A retransmission would allow other users to "piggyback" and join the uplink transmission on the next feedback slot (thus increasing user diversity). While in our evaluation we let the AP chooses a random set of users to join the retransmission, the user selection algorithm presented in [10] could improve the process by enabling the AP to report vectors orthogonal to the already sounded users, and allowing best candidates to reply over that next slot. Nonetheless, notice that this would lead to a short contention period thus leading to an increase in sounding air-time.

D. Overhead Analysis

Feedback Overhead. The 802.11ac sounding procedure requires one beamforming report and one report poll per each user (except for the first user which replies after receiving the NDP, without poll). In contrast, CUiC can use only a single feedback slot for the different sounded users to reply with their beamforming reports. In the case of 802.11ac, if a 4-antenna AP sounds only two single-antenna users in a 20 MHz channel, the sounding procedure takes approximately between 631 μs and 727 μs, depending on the number of bits used for matrix quantization (we assume the beamforming reports are sent at QPSK with $\frac{1}{2}$ coding). Evidently, this number is much higher when considering more than a single user. Considering that an 802.11ac device can transmit at 433 Mbps, during that same amount of time the AP could transmit an extra 13 kB, thus making this process very expensive [18]. The following expression represents the amount of time needed for sounding, given that K users will be served in the next MU-MIMO transmission.

$$T_{802.11ac} = t_{NDPA} + (2K+1) \cdot t_{SIFS} + t_{NDP} + K \cdot t_{report} + (K-1) \cdot t_{poll}$$

CUiC removes the need for multiple SIFS, reports, and polls. Moreover, it relies on the 802.11ac Long Training Field (VHT-LTF) from each of the staggered preambles to do channel estimation, where each VHT-LTF is 4 μs long. Thus, to minimize additional overhead in CUiC, instead of staggering the entire preambles of all received reports, we could stagger only a smaller portion of them, e.g., the VHT-LTF of each user, which allows a 4 μs gap between every two non-overlapping symbols, at the expense of triggering AGC only once. Therefore, the overhead increase incurred in CUiC could be as low as $4K$ μs, where K is the number of users to be served concurrently. Consequently, the overall sounding overhead required by CUiC reduces to

$$T_{CUiC} = t_{NDPA} + 2 \cdot t_{SIFS} + t_{NDP} + t_{report} + K \cdot LTF$$

Acknowledgements Overhead. With respect to the acknowledgment process in MU-MIMO, in the best case CUiC removes the need for any block ACK request and any subsequent SIFS. That is, it reduces the transmission time required for the acknowledgment process by about 120 μs and 60 μs for each BAR (at 6 and 24 Mbps, respectively) and by 16 μs for each SIFS. Additionally, it can reduce the number of transmission slots used to send compressed block ACKs (BA) from one per user to a total of only one, i.e., reduces the total time spent transmitting BA's from 512 μs to 128 μs at 6 Mbps, and from 248 μs to 62 μs at 24 Mbps (parameters used: 40 μs PLCP, 34 bytes for MAC header and FCS, 26 bytes for BAR, and 32 bytes for BA). By saving that amount of time, in a 4-user system (at 20 MHz, 64-QAM $\frac{5}{6}$), the AP could otherwise transmit up to an additional 13 kB. Notice that CUiC incurs in a minor overhead of at least 4 μs for each staggered preamble (or VHT-LTF) transmitted, i.e., one for each user.

In summary, for both the feedback and acknowledgment processes, instead of requiring K transmission slots for all K feedback reports/ACKs and $K-1$ polling frames, plus $2K+1$ SIFS (16μs each), our scheme only requires *one* transmission slot and an almost negligible overhead of $4K$ μs.

III. Implementation and Experimental Evaluation

In this section we present an experimental validation and evaluation of CUiC under a wide variety of indoor WLAN scenarios. Our investigation focuses on *(i)* reliability of our CUiC demultiplexer implementation, *(ii)* overhead reduction of CUiC compared to the sounding and ACK procedures in 802.11ac, and *(iii)* throughput performance of CUiC.

Experimental Methodology. We implement CUiC in the FPGA-based WARP platform [8] and perform over-the-air (OTA) experiments, channel emulator based experiments, and trace-driven emulation, to validate and evaluate our system implementation. First, to validate our schemes we perform controlled experiments using a channel emulator and trace-driven emulation thus allowing us to independently tune different variables at a time. Next, we perform OTA transmissions to investigate performance gains in real channel conditions. We implement an OFDM 802.11-based physical layer and use the 2.484 GHz band, i.e., channel 14, which is not in use in our experimental region, therefore limiting the amount of out-of-network interference affecting our measurements.

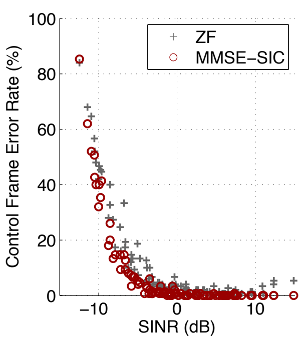
WARP, WARPLab, and Channel Emulator. The WARP platform consists of an FPGA-based software defined radio, interfaced with custom designed radios based on the MAX2829 chipset. This platform allows for the implementation of clean-slate PHY and MAC protocols. WARPLab is a programming environment that integrates MATLAB tools with WARP to control the platform via a host computer for running experiments and collecting data. To perform experiments under controlled and repeatable channel conditions we utilize the Azimuth ACE 400WB channel emulator, which supports 4x4 channel configurations [19].

A. Decoding Reliability of CUiC

Relative Signal Strength. The performance of spatial demultiplexers is highly dependent on the difference between the power of the signal to be decoded next, and the interference plus noise components of the composite signal. Notice that at each stage in the CUiC decoder, only one stream is of interest, whereas the superposition of the rest of the streams is considered interference. Therefore, to determine the SINR at which the intended stream can be reliably decoded with our 4x4 CUiC decoder, we investigate the system's performance at different SINR values.

To this end, we perform a controlled experiment using the Azimuth channel emulator. WARP boards connect to each 4-input/output RF port, and to the host PC that manages the emulator and experimental settings, as well as data collection. We follow the channel model employed in [20], which consists of a 9-tap Rayleigh fading channel with a delay per path profile going from 0 to 80 ns and a path loss per path having a range from 0 to 22 dB. We vary the attenuation of the intended user's signal and the three other interferers. We transmit one thousand 802.11 frames and determine if each of them was successfully received after our decoding process.

Figure 4 (left) depicts the *control frame error rate* (CF-ER), defined as the error rate for sounding feedback control frames transmitted on the uplink, for different SINR levels. We compare the performance of our MMSE-SIC demultiplexer to a ZF-based demultiplexer. First, the control frame error rate of our system decreases rapidly with increasing SINR. Notice that as the SINR reaches 0 dB, performance can reach near zero CF-ER. Additionally, MMSE-SIC outperforms ZF specially for low SNR regimes, as expected. Therefore, we conclude that our MMSE-SIC based CUiC decoder, can reliably decode at least one stream in a 4x4 system even below 0 dB SINR.



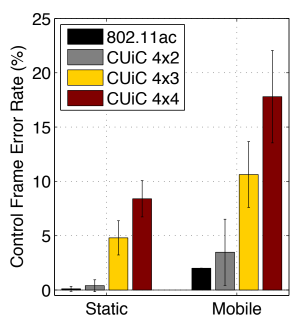


Fig. 4: (Left) CUiC Decoder: Control Frame Error Rate (CF-ER) vs. SINR for MMSE-SIC CUiC and ZF decoder. (Right) Control frame error rate of CUiC and 802.11ac in real WLANs.

Decoding Performance in Real Indoor WLAN Scenarios. Next, we examine the ability of CUiC to decode concurrent streams under different scenarios in a realistic indoor WLAN environment. Additionally, we investigate the benefit of leveraging additional antennas at the AP for increased receive diversity, and robustness. We deploy four single-antenna users and one AP with four antennas in a medium size conference room. Then we evaluate the control frame error rate performance of CUiC in static and mobile conditions. We perform five transmission runs for a different network configuration or mobility pattern. For the static case, the position of each node at each run is different, but remains fixed for the duration of each run. For the mobile case, at each run we arbitrarily move the position of the AP following a different pattern, thus varying the speed, acceleration, and distance to the users. Nonetheless, the maximum speed reached is walking pedestrian speed. In both cases, i.e., static and mobile, environmental mobility due to dynamic scatterers is present.

Figure 4 (right) depicts the average control frame error rate and standard deviation in concurrent uplink transmissions, over each different configuration (topology), for both static and mobile conditions. 802.11ac feedback exhibits the most reliable behavior due to the fact that it transmits each packet individually and sequentially via base-rate SIMO. Namely, to ensure a fair comparison, we enhance the 802.11ac feedback implementation with a linear combining technique to exploit diversity gains by leveraging all AP antennas. However, for static conditions, the control frame error rate for CUiC is below 10%. In a following section we demonstrate that even considering the higher reliability of sequential SIMO feedback as in 802.11ac, CUiC is superior when overhead is considered. More importantly, observe that there is a fast improvement in performance as the number of user transmitting simultaneously decreases compared to the number of receiving antennas at the AP. The reason for this is a combination of increased receiver diversity and a reduction in sources of errors due to timing misalignments and synchronization mismatches.

Channel Correlation and Relative Node Positioning. Channel vector correlation and the spread in signal strengths among concurrently transmitting users affect successful decoding of a composite stream. More specifically, low correlation between the channel vectors of the different concurrently transmitting users allows the AP to better separate the streams after computing and applying either the MMSE or ZF weights. On the other hand, the relative signal strength of all users (influenced by the diverse node positions) affects the ordered SIC process by allowing the AP to first decode the stronger and more robust user therefore reducing the amount of errors propagated from one stage to the other. At the beginning of this section, via a controlled experiment we observed that the difference in power between the signal to be decoded next and the interfering streams is of high importance for reliable frame decoding. However, the system's performance in a real deployment where the AP continuously computes the post processing SINR of each user and determines which one of them should be decoded next, remains unknown.

To study both the impact of channel correlation and signal power variation among users on the ability of the AP to decode the different signals, we deploy a 4x4 system and investigate the channel orthogonality among the four users and the group's signal power variance. In particular, to determine channel orthogonality of the users, we compute the infinity norm condition number of the channel matrix $\mathbf{H}$, i.e., $C_\infty(\mathbf{H})$ [10]. For every uplink SDMA transmission we record the number of failed frames and the infinity norm condition number averaged over all 48 data subcarriers, as well as the variance in signal power among all users. We evaluate the system for a total of 24 indoor topologies with arbitrary node positioning.

Figure 5 (left) shows the cumulative distribution function of the matrix condition number for all five different cases of frame failures. Surprisingly, there is no clear trend showing a higher likelihood of failure with higher condition number. Nonetheless, the case of no failures shows lower condition number with a higher probability compared to the rest of the cases. Clearly, it is very important to group users such that the channel matrix is well-conditioned in order to get zero failed users, otherwise the system could collapse and lead to the failure of all four. In terms of the variance in signal power, there is no evident distinction between all the cases. Therefore, in a real indoor environment, matrix conditioning seems to have a stronger impact on decoding performance at the AP than the variance in the strength of the multiple signals.

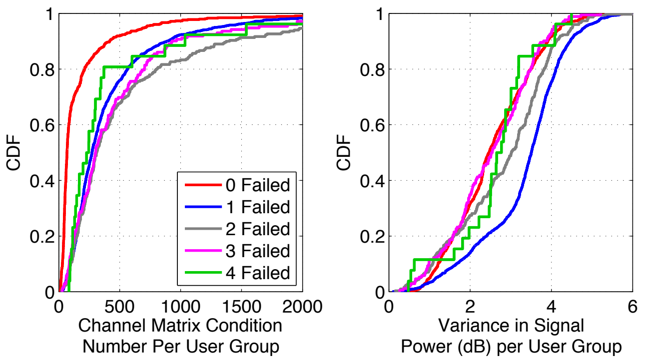


Fig. 5: (Left) CDF of channel matrix infinity condition number. (Right) CDF of variance in signal power among concurrent users.

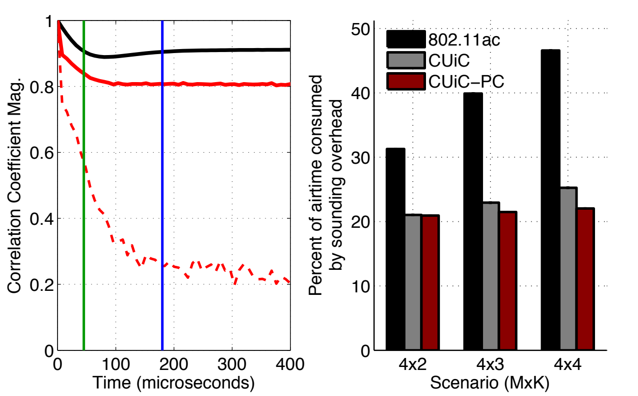


Fig. 6: (Left) Channel autocorrelation and TX trigger for CUiC and 802.11ac-based SIMO feedback. Figure's legend in text below. (Right) Over-the-air experiment showing overhead impact on packet transmissions for 64-QAM 5/6 coding, 20 MHz channels with 400 ns guard intervals (basic CUiC policy).

B. Improving Sounding Resilience in Dynamic Scenarios

By shortening the amount of time between sending the sounding frames and triggering a downlink MU-MIMO transmission, CUiC can also increase the robustness in the presence of rapidly changing channels. In this section we demonstrate how much protection our scheme can provide compared to the sequential SIMO feedback of 802.11ac-based sounding. To this end, we plot the channel correlation as a function of time for different channel profiles and identify the difference in channel variation between the sounding time required in CUiC vs. that required in 802.11ac. Figure 6 (left) depicts the channel correlation coefficient as computed in [21] for static (black) and mobile (red) nodes. Solid lines represent the averages over thousands of channel measurements. Additionally, for illustration purposes we show the channel we measured for the individual worst-case user (dotted curve). Assuming that sounding occurs at time $t = 0$, the first (green) vertical line shows when a downlink transmission is triggered with CUiC. On the other hand, the second vertical line (blue) shows when the downlink transmission is triggered in 802.11ac. That is, the time between $t = 0$ and the green and blue lines depict the amount of time spent doing sounding in CUiC and in 802.11ac, respectively. These sounding times were computed for 20 MHz channels, maximum subcarrier grouping of four, minimum quantization bits, and base rate. Notice however that for other parameters, the 802.11ac sounding time can reach close to 1 millisecond [4], [16], [18].

For several dynamic profiles the difference in channel correlation is significant. For some users, the coefficient drops from about 0.6 to a value below 0.4. Therefore, for many channel profiles, the amount of time saved via CUiC can make the difference between receiving a subset of the frames transmitted vs. not being able to receive anything.

C. Achieving Constant Sounding Overhead with CUiC

CUiC's objective is to maintain constant overhead even as the number of users to be sounded increases (in contrast to current implementations, e.g., 802.11ac). We investigate the amount of overhead that CUiC can suppress compared to 802.11ac for two variations of our scheme, i.e., with and without power control.

Power Control. Without an adaptive AGC algorithm, large differences in the power of the multiple signals can lead to reception issues such as ADC saturation. Although we leave the design of such adaptive system for future work, here we evaluate an idealized protocol in which receive gains are tuned manually to guarantee that all signals fall within the ADC's dynamic range. We denote such scheme as CUiC-PC (Power Control), where retransmissions due to saturation are mostly non-existent. The idealistic scheme provides with a notion of the impact of power control on the overhead reduction that CUiC can attain.

We perform an extensive set of over-the-air (OTA) experiments in a rich scattering indoor office environment at daytime in which all sounded users transmit simultaneously to the AP with their channel estimation feedback. Based on the number of users sounded, and the total number of retransmissions we calculate the amount of overhead involved in the sounding procedure using the 802.11ac timings for each sounding packet (i.e, NDPA, NDP, beamforming report, report poll, and packet duration) [3]. Using these timings we compute the fraction of airtime consumed by sounding overhead, out of the total transmission time, and compare to the SIMO based feedback in 802.11ac. We deploy thirty different topologies and for each topology we transmit over one thousand 802.11 frames.

Table I shows the percent reduction in overhead for each of the evaluated schemes. For the 4x4 case, we achieve nearly 70% reduction, however, as we consider systems with fewer users these gains decrease due to the lower amount of overhead incurred in 802.11ac-based systems. Furthermore, we explore how this overhead reduction translates into performance gains when considering data packet transmissions. Figure 6 (right) depicts the fraction of airtime consumed by sounding overhead for the cases where two, three, or four users reply simultaneously with their beamforming reports. We plot the averages over the thirty different topologies for 802.11ac, and both CUiC and CUiC-PC. Observe that while the overhead in 802.11ac increases rapidly going from 31% to 47% (two users vs. four users), for the concurrent uplink schemes this overhead stays almost constant by going only from 21 to 22%. In conclusion, even without user transmission power control, CUiC can maintain an almost constant overhead as the number of users increases from two to four.

TABLE I: Sounding Overhead Reduction (%)

	4x2	4x3	4x4
CUiC	41.4	54.5	60.5
CUiC-PC	41.8	58.7	67.5

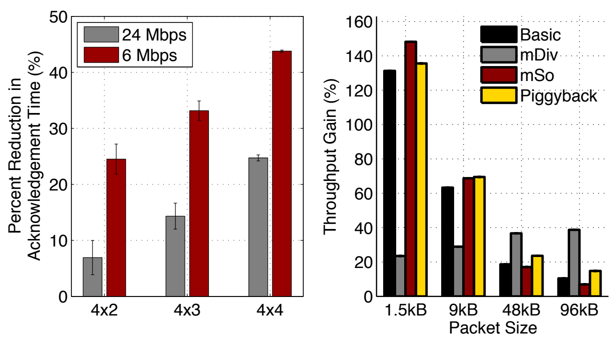


Fig. 7: (Left) CUiC's achieved time reduction in the MU-MIMO acknowledgement process. (Right) Throughput gain of CUiC policies compared to 802.11ac (20 MHz channel, subcarrier grouping of 2, and feedback quant. of $\phi = 9$ and $\psi = 7$ bits).

D. Efficient Acknowledgement Process With CUiC

To investigate the potential gains attained via concurrent block ACK transmissions, we implement a complete acknowledgement process and evaluate its performance in different topologies. We use similar procedure and methodology to that employed in Section III-C. However, in this case we evaluate the performance using two different base rates for both BAs and BARs, i.e., 24 and 6 Mbps. Figure 7 (left) depicts the percent reduction in acknowledgement process time compared to 802.11n/ac systems utilizing these two base rates. Similarly to the sounding procedure, we allow BA retransmissions whenever there is a failed transmission. Observe that even with retransmissions, 4x4 CUiC is able to achieve close to 44% gains and about 25% gains for 6 and 24 Mbps, respectively. Therefore, in a similar way to the sounding process, these gains contribute to an overall system-level efficiency improvement of a complete MU-MIMO system compared to legacy 802.11 SISO ones.

E. Throughput of CUiC Policies

Finally, we evaluate the throughput performance of a complete downlink MU-MIMO system incorporating each of the proposed concurrent feedback and acknowledgement policies, and report their throughput gains over 802.11ac. We consider a system comprised of a 4-antenna AP and 8 single-antenna users (deployed across more than 30 topologies to emulate a larger number of users). Due to hardware constraints, our system cannot meet the fast 802.11ac timing constraints. Consequently, to ensure fair comparison of our policies and the benchmark we rely on a combination of uplink OTA transmissions and downlink MU-MIMO emulation to perform this evaluation.

For the uplink we implement all four CUiC policies on WARP and record all channel estimates for all transmissions, as well as the number of successful/unsuccessful user transmissions, and retransmissions at each instance. Then, using the collected channel estimates we compute the achievable rate [16] at each transmission, for each policy, which we use to compute packet and sounding durations. Thus, our throughput evaluation considers accurate 802.11ac timings, real uplink performance of CUiC, and theoretical downlink ZFBF MU-MIMO rate based on collected (OTA) channels. Moreover, channels are assumed to remain static between sounding and data transmission, for all schemes. This emulation allows us to replay the same channels thus enabling experiment repeatability.

Figure 7 (right) shows that CUiC significantly and consistently outperforms 802.11ac for a wide range of packet sizes, i.e., from single packet to frame aggregation of 64 frames. For shorter packet lengths, the minimum sounding (mSo) policy performs best because it only requires one transmission slot to sound all users and it can serve users much faster. Therefore, as long as the control frame error rate remains low, the best strategy is to truncate the sounding process to the minimum number of slots to maintain constant overhead. Nonetheless, notice that as the number of aggregate frames increases, the maximum diversity (mDiv) strategy outperforms the rest. That is, with higher overhead amortization it is best to utilize the same number of sounding feedback slots as 802.11ac but collect channels from as many users as possible; thus improving user diversity by allowing the AP to serve users with higher-quality downlink channels and consequently increase the overall throughput. From the perspective of absolute aggregate throughput, in the 1.5 kB case we obtained 17 Mbps and 40 Mbps for 802.11ac and the basic CUiC scheme, respectively. Similarly, for the case of 48 kB we obtained 147 Mbps and 174 Mbps for the two respective schemes.

IV. Related Work

Prior work can be broadly classified as follows.

Concurrent uplink transmissions. Uplink multi-user MIMO has been considered in both distributed random access systems and scheduled cellular systems, e.g. LTE-A. Work within the former category focuses on the design of user contention and rate adaptation algorithms [5]–[7], as well as scalable real-time signal processing architectures [22]. On the other hand, work within the cellular category focuses on optimizing the allocation of physical resource blocks (PRBs) to multiple concurrent users [23]. In uplink MU-MIMO data transmissions, the contention and PRB allocation processes allow for the selection of a set of users with low channel correlation. However, in control messages the scheduler (at the AP) determines the set of users from which channel information is needed; therefore, such grouping does not consider their relative uplink channels. This can potentially lead to decoding failure, particularly in the case of an ill-conditioned channel matrix. Thus, in contrast to all prior work that enables user contention based on channel characteristics, CUiC investigates policies for recovery of failed uplink control messages via retransmissions.

Similarly, in contrast to protocols where both the packet preambles and data of multiple users overlap (e.g., [7]), CUiC aligns and synchronizes transmissions of concurrent users to enable the use of multi-stream detection techniques that have been commonly used in SU-MIMO [2] but that are considered impractical for random access MU-MIMO systems [7]. Moreover, unlike random access schemes that rely on past uplink transmissions for estimating channel parameters [6], CUiC does not require any additional signaling and relies fully on the preamble of the current transmissions, thus leading to more accurate estimation of the current channel realizations. Also, CUiC complements the aforementioned *uplink* MU-MIMO schemes by providing a scalable solution for efficient, low-overhead transmission of control messages necessary to enable *downlink* MU-MIMO.

Sounding feedback overhead reduction. Sounding feedback in MU-MIMO has been extensively studied from a theoretical perspective with a focus on quantization mechanisms used to compress feedback [24]. Likewise, practical feedback compression techniques based on 802.11ac have been proposed to reduce feedback in time, frequency, and quantization domains [16], [25]. In contrast, we exploit a fourth dimension along the spatial domain to provide scalability via constant overhead. Lastly, an alternative overhead reduction method is the use of implicit feedback, e.g., see [26]. Implicit feedback relies on the AP to indirectly estimate downlink channels by measuring the reverse links from the users instead of the forward links. Unfortunately, implicit feedback has been shown to also reduce the accuracy of CSIT estimation due to lack of reciprocity between the RF transmit and receive chains in a transceiver's hardware which can severely reduce throughput [27]. While RF calibration counters this effect [26], its accuracy has not yet been validated. Likewise, asymmetric interference at the sender vs. receiver has not yet been considered. Thus, by relying on explicit feedback, CUiC provides better accuracy compared to implicit feedback, while also providing scalable overhead.

V. Conclusion

In this paper we present the design, implementation, and evaluation of a novel MU-MIMO scheme that enables the use of concurrent uplink transmissions by multiple users in order to simultaneously send control and management frames to the AP. More specifically, we present CUiC, a scheme that addresses the high sounding overhead of 802.11ac systems and the inefficiencies of the acknowledgement process in MU-MIMO networks. CUiC strives to reduce the amount of time spent feeding back channel estimates from the users to the AP as well as block ACKs to acknowledge frame reception. We demonstrate that CUiC can achieve near 67% sounding overhead reduction as well as close to 45% reduction in the amount required for 4 users to reply with their ACKs.

References

[1] D. Gesbert, M. Kountouris, R. Heath, C.-B. Ch., and T. Salzer, "Shifting the MIMO paradigm," *IEEE Signal Processing Mag.*, vol. 24, no. 5, pp. 36–46, Sept 2007.

[2] D. Tse and P. Viswanath, *Fundamentals of wireless communication*. Cambridge University Press, 2005.

[3] "IEEE 802.11ac/D7.0, Enhancements for Very High Throughput for Operation in Bands Below 6 GHz," 2013.

[4] O. Bejarano, E. Knightly, and M. Park, "IEEE802.11ac: from channelization to multi-user MIMO," *IEEE Comm. Mag.*, vol. 51, no. 10, pp. 84–90, 2013.

[5] W. Shen, Y. Tung, K. Lee, K. Lin, S. Gollakota, D. Katabi, and M. Chen, "Rate adaptation for 802.11 multiuser MIMO networks," in *Proc. of ACM MobiCom*, 2012.

[6] T.-W. Kuo, K.-C. Lee, K.-J. Lin, and M.-J. Tsai, "Leader-contention-based user matching for 802.11 multiuser mimo networks," *IEEE Trans. W. Comm.*, vol. 13, no. 8, pp. 4389–4400, Aug 2014.

[7] K. Tan, H. Liu, J. Fang, W. Wang, J. Zhang, M. Chen, and G. M. Voelker, "SAM: Enabling practical spatial multiple access in wireless LAN," in *Proc. of ACM MobiCom*, 2009.

[8] "Rice University WARP project," available at: http://warp.rice.edu.

[9] N. Anand, J.-K. Lee, S.-J. Lee, and E. Knightly, "Mode and user selection for multi-user MIMO WLANs without CSI," in *Proc. of IEEE INFOCOM*, 2014.

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

[11] A. Sibille, C. Oestges, and A. Zanella, *MIMO: from theory to implementation*. Elsevier, 2010.

[12] T. Brown, P. Kyritsi, and E. De Carvalho, *Practical Guide to MIMO Radio Channel*. Wiley, 2012.

[13] P. Vandenameele, *Space division multiple access for wireless local area networks*. Springer, 2001.

[14] S. Yun, L. Qiu, and A. Bhartia, "Multi-point to multi-point MIMO in wireless LANs," in *Proc. of IEEE INFOCOM*, 2013.

[15] F. Rashid-Farrokhi, L. Tassiulas, and K. Liu, "Joint optimal power control and beamforming in wireless networks using antenna arrays," *IEEE Trans. on Comm.*, vol. 46, no. 10, pp. 1313–1324, Oct 1998.

[16] O. Bejarano, E. Magistretti, O. Gurewitz, and E. Knightly, "MUTE: Sounding inhibition for MU-MIMO WLANs," in *IEEE SECON*, 2014.

[17] Y. Taesang and A. Goldsmith, "On the optimality of multiantenna broadcast scheduling using zero-forcing beamforming," *IEEE JSAC*, vol. 24, no. 3, pp. 528–541, 2006.

[18] "Cisco inc." *802.11ac: The Fifth Generation of Wi-Fi, Technical White Paper*, August 2012.

[19] "Azimuth systems inc." available at: http://www.azimuthsystems.com/.

[20] E. Aryafar, N. Anand, T. Salonidis, and E. W. Knightly, "Design and experimental evaluation of multi-user beamforming in wireless LANs," in *Proc. of ACM MobiCom*, 2010.

[21] J. W. Wallace, M. A. Jensen, A. L. Swindlehurst, and B. D. Jeffs, "Experimental characterization of the MIMO wireless channel: Data acquisition and analysis," *IEEE Trans. W. Comm.*, vol. 2, no. 2, pp. 335–343, 2003.

[22] 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," in *Proc. of ACM SIGCOMM*, 2013.

[23] E. Dahlman, S. Parkvall, and J. Skold, *4G: LTE / LTE-advanced for mobile broadband*. Academic Press, 2013.

[24] D. Love, R. Heath, V. Lau, D. Gesbert, B. Rao, and M. Andrews, "An overview of limited feedback in wireless communication systems," *IEEE JSAC*, vol. 26, no. 8, pp. 1341–1365, October 2008.

[25] X. Xie, X. Zhang, and K. Sundaresan, "Adaptive feedback compression for MIMO networks," in *Proc. of ACM MobiCom*, 2013.

[26] C. Shepard, H. Yu, N. Anand, E. Li, T. Marzetta, R. Yang, and L. Zhong, "Argos: practical many-antenna base stations," in *ACM MobiCom*, 2012.

[27] S. Haile, "Investigation of channel reciprocity for OFDM TDD systems," Master's thesis, University of Waterloo, 2009.